Molecular Dynamics Based Predictions of the Structural and Functional Effects of Disease Causing Cardiac Troponin C Mutations

by

Bairam Lotfalisalmasi

B.Sc., Simon Fraser University, 2012

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

In the
Department of Biomedical Physiology and Kinesiology
Faculty of Science

Bairam Lotfalisalmasi 2014
SIMON FRASER UNIVERSITY
Summer 2014

Approval

Name: **Bairam Lotfalisalamsi**

Degree: **Master of Science**

Title of Thesis: ***Molecular Dynamics Based Predictions of the Structural and Functional Effects of Disease Causing Cardiac Troponin C Mutations***

Examining Committee: **Chair:** Dr. James Wakeling
Associate Professor

Dr. Glen F. Tibbits
Senior Supervisor
Professor _____

Dr. Edgar C. Young
Supervisor
Associate Professor _____

Dr. Thomas Claydon
Supervisor
Associate Professor _____

Dr. Damon Poburko
External Examiner
Assistant Professor
Biomedical Physiology and Kinesiology _____

Date Defended: May 29, 2014

Partial Copyright Licence

Abstract

Human cardiac troponin C (HcTnC), the regulatory calcium-binding component of the troponin complex, is responsible for the regulation of cardiac muscle contraction in response to varying cytosolic calcium levels. Mutations that are shown to increase the cTnC Ca^{2+} affinity are hypothesized to induce hypertrophic cardiomyopathies (HCM).

Several mutations in HcTnC have been selected that are associated with HCM. These mutations include A8V, L29Q, C84Y, E134D, and D145E. The structural effects of these mutations have been modeled through equilibrium molecular dynamics and their functional and structural impacts have been assessed.

In each mutant that was analyzed, the equilibrated structures have shown notable deviations from wild-type in the regions known to be cardiac troponin I (cTnI) interaction sites. There were differences in the conformation dynamics of site II and cTnC/cTnI interaction sites. We anticipate these correlations may contribute to Ca^{2+} affinity either directly or indirectly through cTnI association.

Keywords: Hypertrophic cardiomyopathy; Calcium affinity; Troponin C; Troponin I; Troponin T; Cardiac muscles

Acknowledgements

First, I would like to thank my senior supervisor, Dr. Glen F. Tibbits, for providing the environment so I can carry out my project. Next, I would like to thank Dr. Charles M. Stevens for his mentorship and guidance in all stages of my project. I would also like to thank my committee members, Dr. Thomas Claydon and Dr. Edgar C. Young, for their support all through my project.

I would like to thank our lab manager, Haruyo Kashihara for her support and assistance in acquiring resources required to complete my project.

I would also like to thank my friend Kaveh Rayani for his continued support and friendship which helped me during the most difficult time of my study.

This research was enabled in part by support provided by WestGrid (www.westgrid.ca) and Compute Canada Calcul Canada (www.computecanada.ca).

This work was supported by the Theoretical and Computational Biophysics group, NIH Center for Macromolecular Modeling and Bioinformatics, at the Beckman Institute, University of Illinois at Urbana-Champaign.

At last, I would like to thank my parents, Hamid and Akram, my brother, Pasha, for their unconditional love and support. They sacrificed many things in their life so that I wouldn't have to sacrifice anything in mine. Thanks for being an amazing family.

Table of Contents

List of Tables

List of Figures

List of Acronyms

a	Acceleration
Å	Angstroms, a unit of measurement. 1Å = 10^{-10} meters
atm	Atmosphere; Pressure unit
C-cTnC	C-terminal domain of Cardiac Troponin C
CD	Circular Dichroism
FHC	Familial Hypertrophic Cardiomyopathy
GROMACS	GROningen MAchine for Chemical Simulations
H_A	Helix A of troponin C
H_B	Helix B of troponin C
HCM	Hypertrophic Cardiomyopathy
HcTnC	Human Cardiac Troponin C
HcTnI	Human Cardiac Troponin I
HcTnT	Human Cardiac Troponin T
ID	Inhibitory Domain
ITC	Isothermal Titration Calorimetry
MD	Molecular Dynamics
MTSL	1-oxyl-2,2,5,5-tetramethylpyrro- line-3-methyl-16-methanethiosulfonate)
N-cTnC	N-terminal domain of Cardiac Troponin C

N-TnI	N-terminal domain of TnI
PDB	Protein Data Bank, http://www.rcsb.org
PMF	Potential of Mean Force
r	Position
RD	Regulatory Domain
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SEM	Standard Error of the Mean
t	Time
TM	Tropomyosin
Tn	Troponin Complex
TnC	Troponin C
TnI	Troponin I
TnT	Troponin T
v	Velocity
WHAM	Weighted Histogram Analysis Method
WT	Wild Type

Chapter 1.

Background

1.1. Cardiac Muscles

Cardiac muscle cells contain a dense parallel array of small, cylindrical elements called myofibrils. Every myofibril is made of repeating units called sarcomeres, which are comprised of smaller interdigitating filaments called myofilaments[1]. There are two types of myofilaments which are referred to as thin and thick filaments. The thick filament is comprised of myosin heavy and light chains. The thin filament consists of actin, tropomyosin (Tm), and troponin complex (Tn). The troponin complex is a heterotrimer made of 3 individual proteins: troponin I (TnI), troponin T (TnT), and troponin C (TnC)[1].

1.1.1. Thin filament proteins

Troponin I

The troponin I in humans is expressed in three isoforms: fast and slow skeletal isoforms as well as cardiac isoform[2–5]. The human cardiac troponin I (HcTnI) isoform of TnI comprises 210 amino acids[5–7] is made up of 5 domains: N-terminal, IT-arm, Inhibitory, regulatory, and C-terminal mobile domains (Figure 1-1).

The N-terminal domain of TnI (N-TnI) includes residues 2-32 and it is only present in the cardiac isoform of TnI. It plays a vital role in the interaction of TnI with TnC and in the regulation of muscle contraction. This domain consists of three components: an acidic region (residues 2-11), Xaa-Pro motif (residues 12-18), and a section that can be phosphorylated by protein kinase A (residues 19-32) having two serine residues (S23 and S24)[8,9]. Through dot blotting[10] and NMR[11–14] it was demonstrated that residues 19-33 of the N-terminal domain of dephosphorylated TnI are able to interact with the N-

terminal domain of TnC (Figure 1-1)[10,13]. Upon phosphorylation of S23 and S24 a new α-helix motif forms between residues 21-30 of TnI[15]. This structural change disrupts the interaction of TnC with the N-TnI and dislocates it. It is hypothesized that the acidic N-terminal domain of phosphorylated TnI interacts with the positively charged inhibitory domain of TnI and effects the Ca^{2+} dependent regulation of muscle contraction[15].

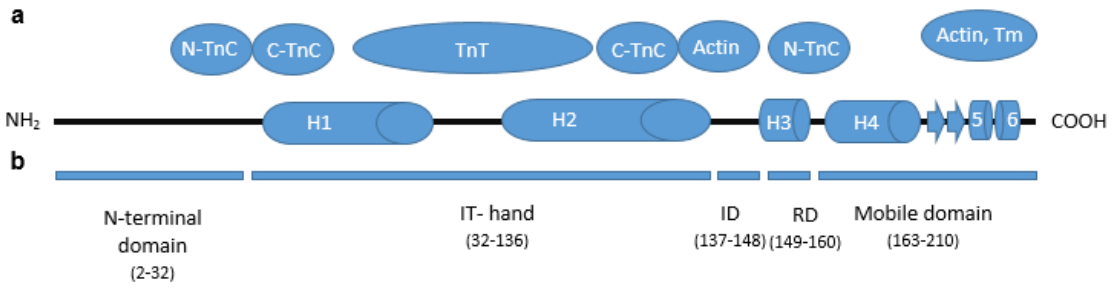


Figure 1-1. Domain structure of human cardiac troponin I.
a) Scheme of HcTnI secondary structure. The structure of the N-terminal domain of TnI, arrangement of α-helices H1-H4 in the TnI molecule, and structure of the mobile domain. The wavy curve represents the Xaa-Pro region of TnI (residues 12-18) forming a proline helix. Short β-strands 1 and 2 are marked by arrows. Proteins of the thin filament that interact with the relevant regions of the TnI molecule are indicated in ovals. N-TnC and C-TnC, N- and C-terminal domains of TnC, respectively; Tm, tropomyosin. b) Domain organization of HcTnI. ID, inhibitory domain; RD, regulatory domain. This figure/caption is a modified version of figure/caption used in the article by Katrukha et al., 2013[5].

As shown in Figure 1-1 and Figure 1-2, the IT-arm, also known as the IT-hand, (residues 32-136), downstream from the N-terminal domain, consists of two α-helices which are referred to as H1 (residues 43-79) and H2 (residues 90-135). H1 and H2 are connected with a short linker[16]. The IT-arm is the least flexible part of the TnI molecule and has a structural role. It provides a point of contact with TnC and TnT and plays role in adjusting the orientation of TnI in the troponin complex. An amphiphilic portion of the H1 α-helix (residues 43-65) interacts with the C-terminal domain of TnC[17,18]. The C-terminal portion of the H1 α-helix (residues 66-79) and H2 contacts with the H2 α-helix of TnT[17,19]. The short flexible linker between H1 and H2 spans from residue 80 to 89. The H2 α-helix of TnI also make a coiled-coil structure with the H2 α-helix of TnT (residues 226-271 of HcTnT)[18,20].

The inhibitory domain in human cardiac TnI, according to Takeda[18] and Kobayashi[21] extends from residues 137 to 148. Other groups (Sykes et al.[22] and Brown

et al.[23]) reported a different amino acid range (residues 129-148) for this region. During low Ca^{2+} concentration period, residues 138-148 of inhibitory domain interact with actin[21,24] which causes a shift in the position of the tropomyosin and prevents formation of the actomyosin complex[25,26].

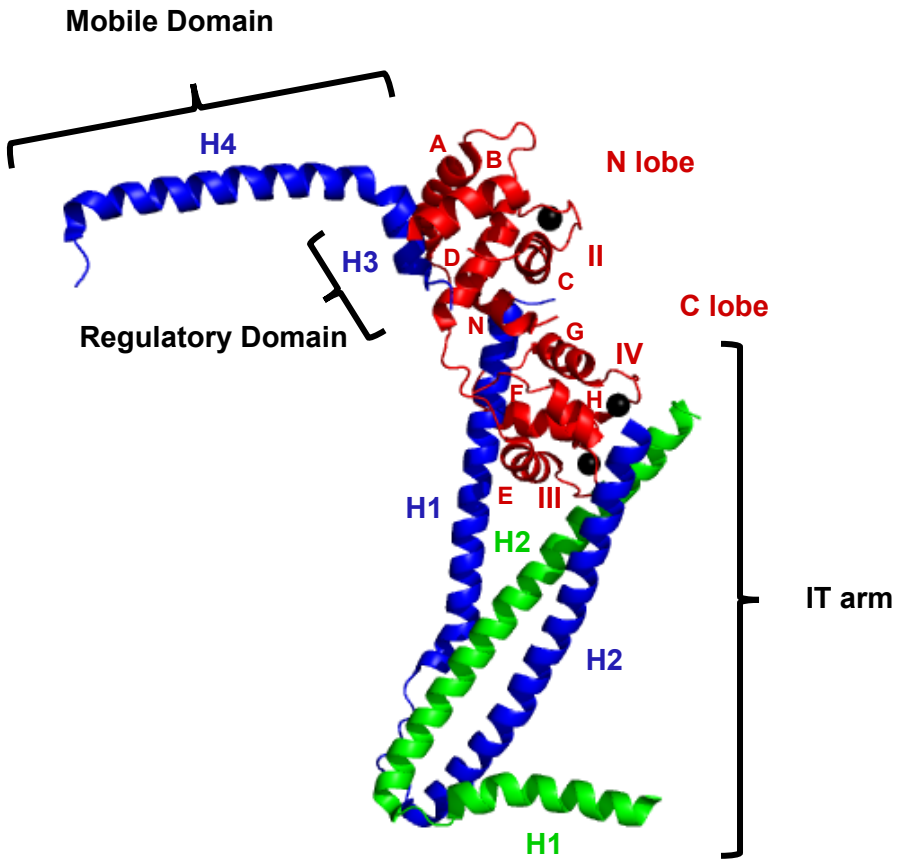


Figure 1-2. Crystal structure of troponin core domain.
TnI, TnC, and TnT are colored in blue, red, and green, respectively. Black sphere represents Ca^{2+}. Each helix within TnI and TnT is indicated by helix number, whereas each helix of TnC is indicated by a capital letter (N and A-H). Crystal structure of the core domain of human cardiac troponin by Takeda et al. (PDB 1J1E)[18].

This regulatory domain extending from residues 149-163 in HcTnI consists of a short α-helix "H3" (residues 150-159)[18]. During periods of high cytosolic Ca^{2+} concentration the regulatory domain binds with hydrophobic patch of the TnC N-terminal domain [27]. This interaction of TnI with TnC leads to the disconnection of the TnI inhibitory domain from actin and the relocation of tropomyosin, which ultimately results in the formation of the actomyosin complex[19,27,28].

The C-terminal domain of HcTnI (residues 163-210) is composed of an α-helix "H4"(residues 164-188) and the C-terminal section of the molecule (residues 190-210). Due to the high motility of the C-terminal region in the presence of Ca^{2+} the structure of this region is not fully understood[18]. Other NMR studies conducted in vitro using troponin complex made from recombinant TnI, TnC and fragments of TnT showed the C-terminal region of TnI to be in an unordered state[29,30]. In a study of different recombinant fragments of human cardiac TnI it has been shown that at low Ca^{2+} concentrations the mobile domain of TnI interacts with tropomyosin and the C-terminal part of actin. These interactions are believed to play a critical role in the regulation of Ca^{2+} dependent contraction and stabilization of the troponin complex on the surface of the thin filament[5,31,32].

Troponin T

Troponin T has multiple functions such as anchoring the troponin complex to the actin filament and participating in the regulation of muscle contraction. It also acts as an organizer of the subunits in the complex[33,34]. In human troponin T is expressed by three genes coding slow and fast skeletal and cardiac isoforms of the protein[5,35,36]. The isoform of troponin T that is mainly expressed in adult human heart (HcTnT) is the 25.9 kDa TnT3 isoform. During the processing of the protein the N-terminal methionine is cleaved so the functional protein product is 287a.a. long. Unfortunately, the only part of the molecule that has been crystalized is the C-terminus (183-288)[5,18].

As illustrated in Figure 1-3, the HcTnT molecule which is extended along the thin filament is composed of the variable N-terminal domain (residues 2-68), conservative central domain (residues 69-200), and the C-terminal domain (residues 201-288)[32,37].

The N-terminal region of the TnT found in different isoforms is shown to have distinct structure[38]. For instance, HcTnT has a unique sequence of ~32 a.a. that is missing from the skeletal isoform of TnT[39]. In this region ~18 out of the 32 a.a. are aspartate or glutamate therefore it is highly polar and negatively charged. Some of the early studies on isolated N-terminal domain claim that this region lacks any type of interaction with proteins of the thin filament[40]. Contrary to this claim others hypothesize that the first 40 a.a. of the HcTnT might interact with the other components of the troponin complex and tropomyosin[41]. Based on studies conducted using different chimeric proteins and deletion mutants of TnT, the N-terminal part of the protein is shown to effect the conformation of the complex, the interaction of troponin complex with actin and tropomyosin, the Ca^{2+} sensitivity of the muscle, and the development of the maximum force of contraction[5,39,42–48].

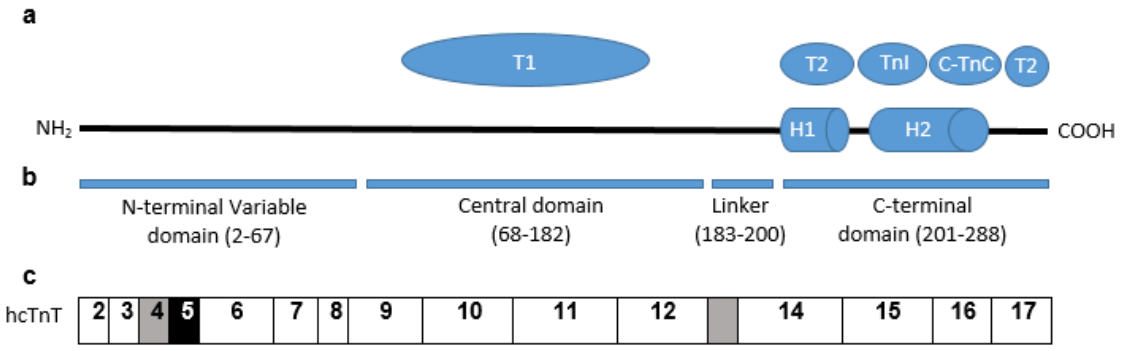


Figure 1-3. Domain structure of human cardiac troponin T.
a) Scheme of the HcTnT secondary structure. Arrangement of α-helices H1 and H2 in the HcTnT molecule. Proteins of the thin filament that interact with the relevant regions of the TnT molecule are indicated in ovals. T1 and T2, regions of interaction with tropomyosin. b) Domain organization of HcTnT molecule. c) Arrangement of the exons coding the relevant parts of the HcTnT molecule. Gray boxes, exons 4 and 13 that are alternatively spliced; black box, exon 5 that is expressed only in the embryonic form of the protein. This figure/caption is a modified version of figure/caption used in article by Katrukha et al., 2013[5].

The central domain includes T1 which is the first tropomyosin interaction site (residues 98-136 of HcTnT)[49]. The force of this interaction is shown to be independent of the concentration of Ca^{2+} [50]. The precise structure of the T1 is not understood but it is believed to be mainly α-helical[51,52]. Following T1 there is a flexible linker expanding from 183 to 200 which connects the central part of the molecule with its C-terminal domain[5,18,52].

The C-terminal domain consists of α-helices H1 (residues 204-220) and H2 (residues 226-271). The H2 α-helix interacts with the C-terminal domain of TnI and the (residues 256-270) of H2 has interactions with the Ca^{2+}-binding loops of the TnC C-terminal domain[18]. The C-terminal domain of HcTnT, besides interacting with TnI and TnC, contains a second site of interaction with tropomyosin (T2). T2 is known to connect to a region around the C190 of tropomyosin and unlike T1 this interaction appears to strengthen in the absence of Ca^{2+} [50,53,54]. The precise location of the T2 region is a subject of a debate. Jin et al. [49] have shown, using mouse slow skeletal TnT, that T2 spans residues 180-204 (homologous to residues 197-239 of HcTnT)[49]. But others claim that the T2 region consists of the last 16 residues of TnT[54–56]. An in vitro experiment conducted by Franklin et al.[56] demonstrated that this C-terminal part of the molecule stabilizes the troponin complex in the non-active state through interaction with tropomyosin and actin.

Troponin C

The human cardiac isoform of TnC (HcTnC), has molecular mass of 18.4 kDa and is composed of 161 a.a.[57]. TnC is a member of the EF protein family. As illustrated in Figure 1-4, the canonical EF-hand is a helix-loop-helix motif which is usually made up of a sequence of 12 residues, commonly known as coordinating or chelating residues. The coordinating residues can be described by their locations in the tertiary geometry (±X, ±Y and ±Z) and the position within the linear loop (1, 3, 5, 7, 9 and 12). The coordinating residues of EF-hand have a pattern X•Y•Z•-Y•X••-Z (the X, Y, Z, -X, -Y and –Z are metal coordination ligands and the dots represent residues). This sequence forms a loop that can contain calcium or magnesium with specific geometries. For Ca^{2+} the 7 ligands located at the vertices of a pentagonal bipyramid are responsible for coordination[58].

The TnC is the Ca^{2+} sensor, which has a dumbbell-shaped structure with two globular lobes connected by a flexible linker situated between α-helices D and E (D–E linker)[59]. Each globular domain has a pair of EF hand Ca^{2+} binding sites (Figure 1-5).

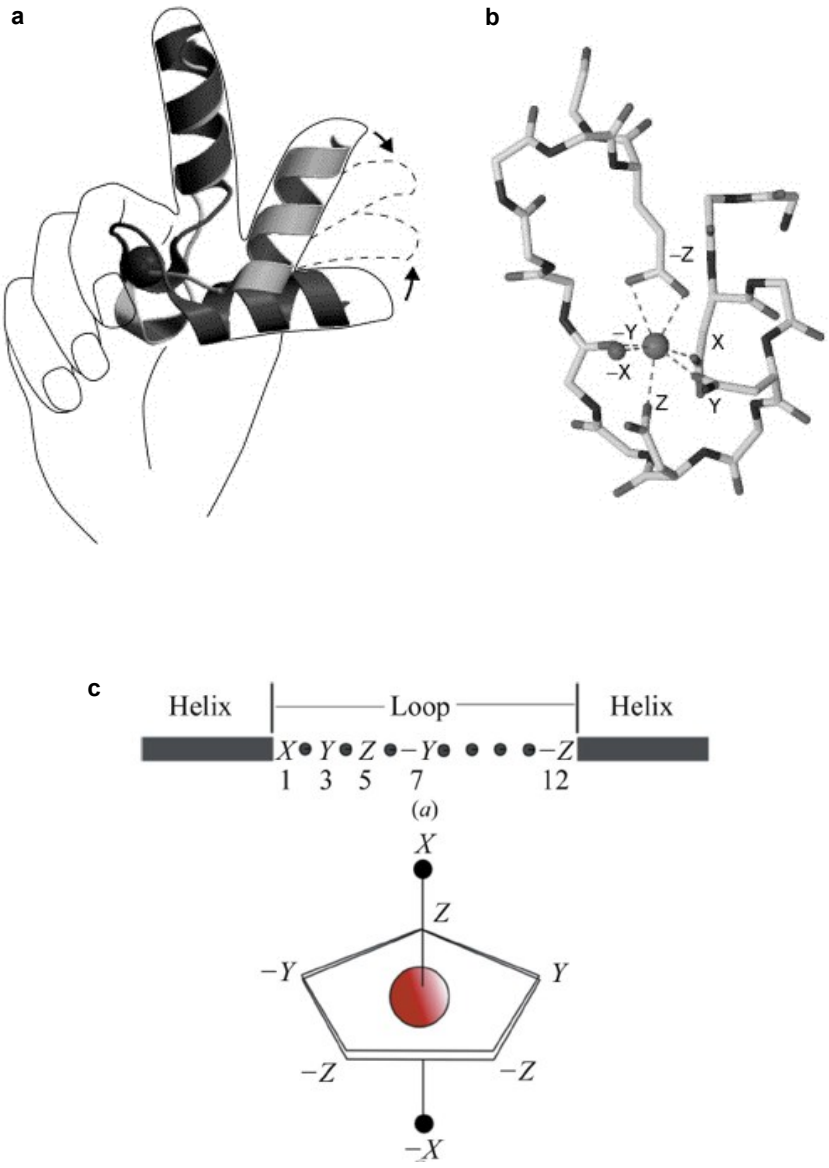


Figure 1-4. The EF-hand motif.
a) A symbolic representation of the EF-hand motif. Helix E winds down the index finger, whereas helix F winds up the thumb of a right hand. When calcium ions bind, the F helix moves from the closed (apoprotein, light grey) to the open (holoprotein, dark grey) conformation. b) and c) The geometry of the calcium ligands. At positions X and Y, aspartic acid or asparagine are usually found; the side chains of aspartic acid, asparagine or serine are found at Z and a peptide carbonyl oxygen lies at –Y. –X is usually a water molecule and –Z is a conserved bidentate ligand, glutamic acid or aspartic acid[58].

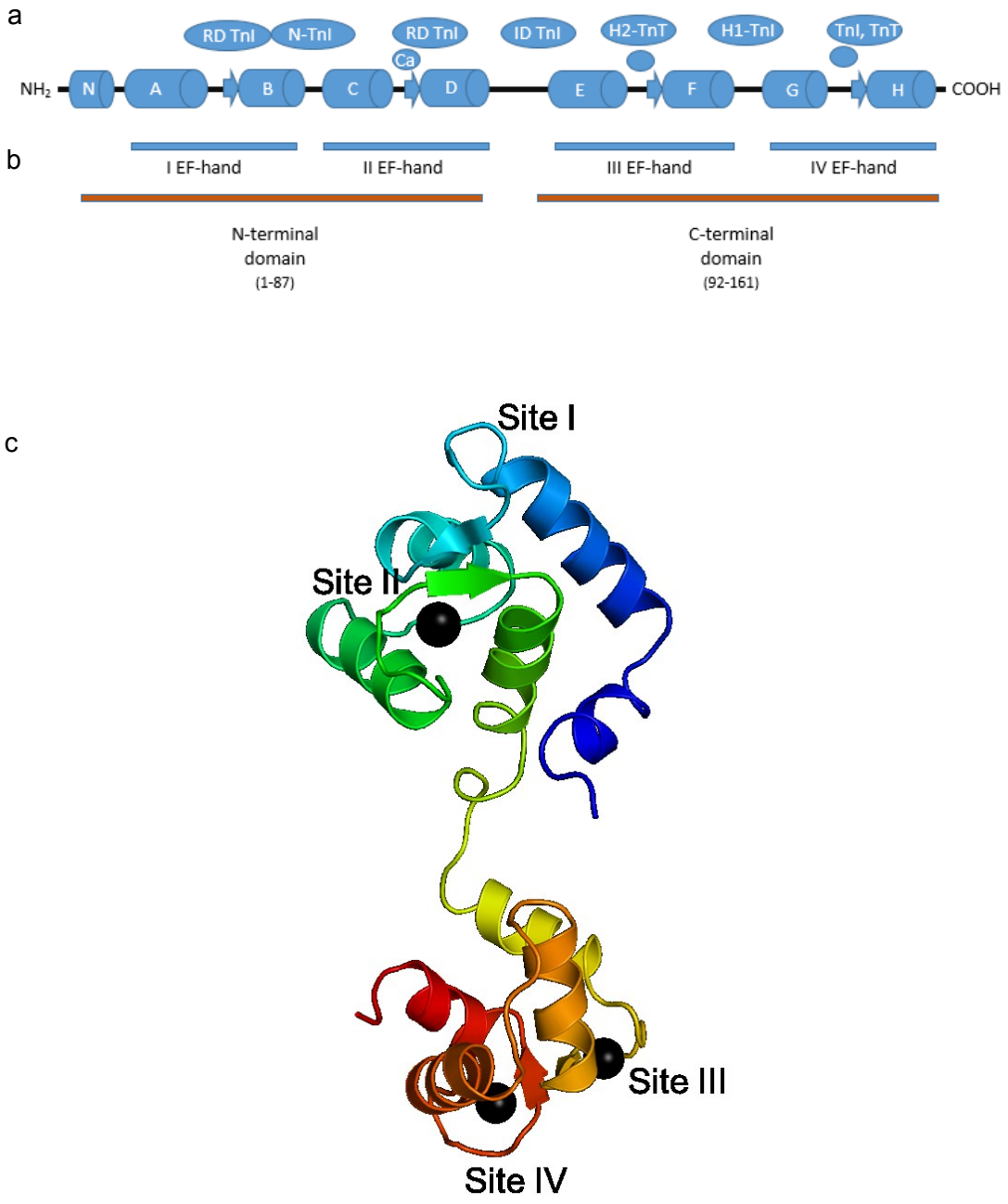


Figure 1-5. NMR structure of the full-length cardiac troponin C along with the domain structure of human cardiac troponin C.

a) Scheme of cTnC secondary structure. α-Helices N and A-H are represented by cylinders, β-strands by arrows. Two circles without designation mark the low-specificity regions that bind both Ca^{2+} and Mg^{2+}. The circle designated with "Ca" represents the high-specificity region that binds Ca^{2+}. Proteins of the thin filament that interact with the relevant regions of the TnT molecule are indicated in ovals. N-TnI, N-terminal domain of TnI; RD, regulatory domain; ID, inhibitory domain. b) Domain structure of c/ssTnC. c) Cartoon rendering color spectrally from the N-terminus (blue) to the C-terminus (red). The calcium ions are presented as black spheres. The N- and C-terminus are labeled (PDB 1AJ4)[59]. This figure/caption is a modified version of figure/caption used in the article by Katrukha et al., 2013[5].

Each EF-hand consists of a Ca^{2+} -binding loop which is positioned between two α-helices (α-helices A-H). Each loop includes a short β-strand that forms an antiparallel β-sheet with the β-strand of the adjacent Ca^{2+} -binding loop[18,59]. The C-terminal domain (residues 92-161) houses two high affinity sites (Site III and IV), which are continuously, under physiological conditions, occupied by either Ca^{2+} or Mg^{2+}. Constant occupation of sites III and IV with metal ions causes the hydrophobic patch of cTnC C-terminal domain (C-cTnC) to be exposed to the N-terminal amphiphilic α-helix of cTnI (N-cTnI). The interaction between C-cTnC and N-cTnI anchors cTnC to the rest of the Tn complex. Therefore C- terminal domain of TnC is thought to have only a structural role[18]. The N-terminal domain (residues 1-87) consists of sites I and II. In the fast skeletal isoform all 4 EF hands are functional and bind Ca^{2+} or Mg^{2+}. In human cardiac TnC due to few non-conservative amino acid substitutions, site I is dysfunctional and does not interact with Ca^{2+} [59,60]. Site II is a low affinity Ca^{2+} specific binding site, which through a series of conformational changes controls muscle contraction. Therefore, the N-terminal domain, specifically site II, has a regulatory role. As shown in Figure 1-6 site II consists of 12 residues. The Site II conserved residues Asp65, Asp67 and Glu76 were important for calcium coordination in isolated TnC. In the case of the whole troponin complex, Gly68 may not take a part in calcium coordination, but was found to be crucial for the maintenance of the structure of site II[42,61,62]. In a recent study, serine 69 was mutated to Cysteine, and troponin was unable to generate force in skeletal muscle fibers[63].

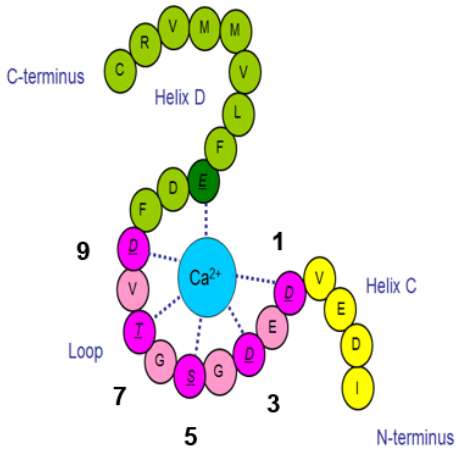


Figure 1-6. Site II EF hand Ca^{2+} coordination.
Schematic diagram of EF-hand illustrating Ca^{2+} coordinating residues. The numbers represent residue numbers in the loop.

1.1.2. Cardiac muscle contraction

A change in myocyte cytosolic Ca^{2+} concentration is responsible for the regulation of cardiac muscle concentration. Contrary to smooth muscle, where contraction is regulated by Ca^{2+} dependent phosphorylation of the myosin light chain, cardiac muscle is regulated by a series of conformational changes in thin filament structure induced by Ca^{2+} binding[64–66]. Currently there are three proposed models of the regulation of cardiac muscle contraction[5]: a three-state model[66–68], a "fly-casting" model[29,69], and a four state model[70,71].

The three state model identifies three structural states for the thin filament: a blocked B-state, a closed C-state, and an open myosin-bound M-state[67]. During diastole, when the cytosolic Ca^{2+} concentration is low (~100 nM), the thin filament is in the B state. In this state the inhibitory domain of TnI is bound to F-actin and the tropomyosin is blocking the myosin-binding site on the surface of the thin filament [72]. The location of tropomyosin in this state is stabilized by the interaction of the T2 region of TnT with tropomyosin, which is enhanced by the low $[Ca^{2+}]$[50]. As the cytosolic Ca^{2+} concentration rises, Ca^{2+} binds to site II of cTnC and causes conformational changes in cTnC. As a result of these conformational changes α-helices B and C move away from α-helices A and D[73–75]. This opens up a hydrophobic patch on the surface of the N-terminal domain of TnC where the regulatory domain of TnI binds[27,76,77]. The interaction of the regulatory domain of cTnI with the N-terminal domain of cTnC causes the TnI inhibitory domain to dissociate from actin and subsequently shifts the tropomyosin molecule to the center of the F-actin groove. Displacement of tropomyosin to the F-actin groove exposes the binding site to myosin and the formation of a weak contact between S1-myosin and actin[27,78,79]. This is known as the closed "C-state". During the C-state tropomyosin is not fully shifted into the actin groove and myosin binding site is not fully exposed. During the M-state the tropomyosin is fully situated in the F-actin groove and the myosin binding site on actin is fully exposed. This condition is optimal for formation of a strong bind between actin and myosin and ultimately muscle contraction[66,80–83]. A drop in cytosolic Ca^{2+} concentration causes Ca^{2+} dissociation from Site II and a subsequent reversion in the cTnC conformation to the "closed" state, along with dissociation of the TnI regulatory domain from N-cTnC. This promotes the association of the TnI inhibitory domain with actin and movement of tropomyosin back to the edge of the groove, which blocks the

myosin-binding site. This model was further advanced by Robinson et al.[84] and Houmeida et al.[85], who showed that for 100 % activation of Ca^{2+}-saturated actin–troponin–tropomyosin complex both S1 myosin and Ca^{2+} are essential.

The fly casting model claims that in skeletal muscle the C-terminal mobile domain of TnI is the main regulator of muscle contraction[69]. It purports that in the absence of Ca^{2+} the mobile domain forms a stable secondary structure which consists of the H4 α-helix that is followed by two short antiparallel β-strands and two back to back short α-helices[5]. In this conformation the mobile domain interacts with actin, and the thin filament maintains its non-active state. According to this model the regulatory domain does not adopt a stabilized conformation and it switches back and forth between actin and the N-terminal domain of TnC. An increase in cytosolic Ca^{2+} concentration stimulates the regulatory domain of TnI to adopt a stable α-helical structure and bind the N-terminal domain of TnC. Following this, the TnI mobile domain disconnects from actin and enters a flexible disordered state. The TnI inhibitory domain then dissociates from actin and interacts with the D–E linker of TnC. Conversely, a drop in cytosolic Ca^{2+} concentration stimulates the interaction of the TnI mobile domain with actin. This provokes the dissociation of TnI regulatory domain from TnC followed by dissociation of the inhibitory domains from the D-E linker of TnC. Finally, the inhibitory domain of TnI binds to actin and the muscle enters the relaxed state[5,69].

The four state model proposed by Lehrer[71] is fundamentally similar to the three state model. Based on this model, the thin filament can adopt four structural states: blocked Mg^{2+}-state, closed Ca^{2+}-state and, open Mg^{2+}-S1 and Ca^{2+}-S1 states[70,71]. In the blocked Mg^{2+}-state, which is equivalent to the B-state in the three state model, the TnI inhibitory domain is bound to actin and the tropomyosin covers the myosin binding site, preventing the formation of actin-myosin complex. The TnI mobile domain also interacts with the surface of actin, stabilizing the troponin complex on the thin filament. Elevation of cytosolic Ca^{2+} concentration promotes the disconnection of the mobile domain from actin and conversion of the TnI mobile domain into a disordered state. Next, the association of TnI regulatory domain with the N-terminal domain of TnC induces the dissociation of the TnI inhibitory domain from actin. This state is identified as the “closed Ca^{2+} state and it is equivalent to the C-state of the three- state model. Finally, the tropomyosin moves away from the myosin-binding site into the actin groove and allows

S1-myosin to bind actin and form a cross-bridge. This is the "open Ca^{2+}-S1- state" which is similar to the M-state of the three-state model. In the Mg^{2+}-S1-state, which is the state that distinguishes the four-state model from the three-state model, S1-myosin interacts with actin in the absence of Ca^{2+}. It is assumed that this interaction leads to relocation of tropomyosin on actin and the disconnection of the TnI mobile domain from actin. This subsequently leads to the association of the regulatory domain with TnC and ultimately increases the sensitivity of TnC to Ca^{2+}. It is necessary to mention that association of S1-myosin with actin alone is not sufficient to form a strong cross bridge and Ca^{2+} is still required for strong contraction. However, this interaction speeds the transition of the thin filament from the inactive to the active state[5,70,71]. According to the four-state model, the relaxation of muscle occurs in two steps, which are distinguished by their rates. In the first step the Ca^{2+} dissociates for site II of TnC, the mobile domain of TnI associates with actin and disconnects from the regulatory domain of TnC. During the second, slower, step, the inhibitory domain of TnI interacts with actin. The S1-myosin resists the interaction between the mobile domain of TnI with actin, thereby stabilizing the open conformation of TnC and preventing muscle relaxation. The interaction of the inhibitory domain with actin is the rate limiting step in regulation of muscle contraction[70,86].

Based on the available models the troponin complex, and in particular troponin C, plays a crucial rule in regulation of cardiac muscle contraction. Muscle contraction is mainly controlled through conformational changes in the troponin complex induced by the fluctuation of Ca^{2+} concentration over the course of the contraction. Any change in the amino acid sequence (mutation) of each protein may disrupt the normal function of the complex and affect the interaction of troponin proteins with the thin filament or ions. Many disease-causing mutations in troponin C, discussed in the next section, are known to influence its affinity for Ca^{2+}.

1.2. Hypertrophic Cardiomyopathy

Hypertrophic Cardiomyopathy (HCM) is a common disease with cases reported in more than 50 countries across 5 continents. Although people diagnosed with HCM have different ethnic and racial origins, they share similar causal mutations, clinical course, and phenotypic expression[87–95]. In regions such as North America[96,97],

Europe[98,99], Asia[91], and east Africa, HCM is recognized as a common genetic disease[100]. The epidemiological studies of HCM has reported a prevalence of 1 in every 500 people[92]. Due to the delay in phenotypic manifestation of the disease, HCM is not commonly recognized in young children; However, reported cases of HCM in children are more prevalent in males [101–103]. HCM is also either under-recognized or clinically diagnosed late in women and blacks [87,89] . Although the mortality rate is similar in both sexes, there tends to be higher rate of heart failure in woman when they are diagnosed later in life. In the athletic fields, sudden cardiac deaths are more prevalent among men[89,90]. The onset of genetic HCM symptoms are age-dependent affecting 50% to 80% of people by age 30 and 95% by age 50-60 years [104].

HCM was first recognized in the 19th century but the familial nature of the disease was only identified after the completion of a detailed study of sudden deaths in young adults [105–107]. HCM is defined morphologically and pathologically by increased thickness of the left ventricular wall associated with non-dilated ventricular chambers, myocyte disarray, interstitial and replacement fibrosis, dysplastic intramyocardial arterioles, and septal hypertrophy. The clinical symptoms of HCM include diastolic dysfunction, atrial and ventricular tachyarrhythmia as well as sudden death[108–110].

Familial hypertrophic cardiomyopathy (FHC), the inherited form of HCM, is caused by autosomal dominant mutations in more than 11 genes that code for proteins that make up the thick and thin contractile myofilaments[98,111–116]. These proteins include: the cardiac β-myosin heavy chain[117], the cardiac regulatory and essential myosin light chains[118], cardiac myosin binding protein C[119,120], and titin[121] from the thick filament. The thin filament proteins involved are: actin[122], α-Tropomyosin (α-Tm)[119,123], cTnT[119,123–126], cTnI[122], and cTnC[107].

In a study conducted by Landstorm et al. (2008), 1025 unrelated patients were evaluated using open reading frame/splice site mutation analyses. Four novel missense mutations in cTnC (Ala8Val, Cys84Tyr, Glu134Asp, and Asp145Glu) were discovered (Figure 1-7)[127]. The effect of these 4 mutations on the myofilament Ca^{2+} regulation was evaluated by Ca^{2+} sensitivity of force development and force recovery. A8V showed the highest increase in Ca^{2+} sensitivity followed by D145E, and C84Y. The E134D substitution did not have an apparent effect. The A8V and D145E mutations are located

in two different functional regions of HcTnC (Figure 1-7), however they demonstrated a nearly equivalent increase in Ca^{2+} sensitivity of force development and force recovery. They propose no structural explanation for the effect observed in A8V or D145E; however, the replacement of cysteine 84 with a bulky tyrosine could modify the angle between N-terminal domain and D-E linker, mimicking an intermediate HcTnC Ca^{2+} bound / open state. This increases the Ca^{2+} sensitivity of force development. Finally, they hypothesize that E134D is not a pathogenic, HCM causing mutation but simply a functionally/clinically insignificant rare variant[127]. These findings were reinforced by Potter et al. (2009), who used actomyosin ATPase assays to evaluate Ca^{2+} sensitivity. Of the same 4 mutants, A8V and C84Y showed the greatest increase in Ca^{2+} sensitivity, D145E also showed an increase in Ca^{2+} sensitivity while E134D showed no significant effect[128].

A8V

Substitution of alanine at residue 8 with valine in cTnC has been shown to increase the Ca^{2+} affinity and affect force recovery in skinned muscle fiber preparations[127,128]. In 2009 Potter et al.,[128] based on their circular dichroism (CD) data in isolated TnC, advised that this mutation might alter the secondary structure content for both the apo (absence of divalent cations) and holo (presence of divalent cations) states. CD was used to observe how cTnC secondary structure changes with the introduction of mutations and upon interaction with other molecules. Previous research has shown that the N-helix makes contacts with helices A and D through hydrophobic and electrostatic interactions[129–131]. Therefore to investigate the effect of A8V on this interaction, researchers used PyMol to model A8V into the crystal structure from Takeda et al.[16] (PDB 1J1E) and claimed that substitution of alanine to valine, with a larger side chain, could change the interaction between the N-helix and D-helix[128]. Since this mutation effects the overall N-terminal structure of cTnC, it can alter TnC and TnI interactions and influence the C-terminal domain[128,132,133].

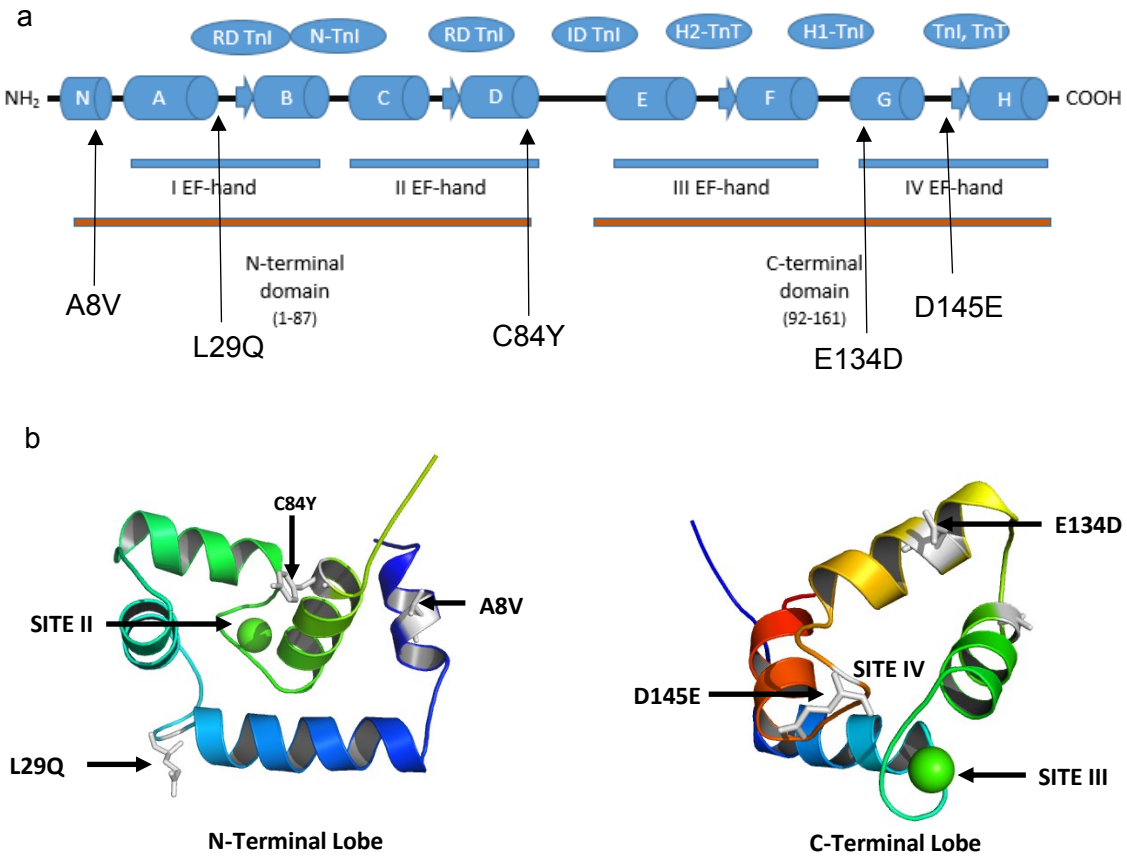


Figure 1-7. Mapping and modeling of HCM-susceptibility mutations in TNNC1-encoded HcTnC

a) The protein linear topology of HcTnC, including functional domains and calcium binding site locations. The location of each of the 5 mutations is noted respectively. b) N-terminal Lobe (left) depicting the location of A8V, L29Q and C84Y mutations in relationship to the Ca^{2+} binding sites and HcTnC helices. Ca^{2+} binding site I is defunct and Ca^{2+} binding site II shows Ca^{2+}bound (green sphere); The Ala8 is located in the first helix of TnC in the beginning of the flexible linker connecting the two domains. C-terminus (right) depicting the location of E134D and D145E mutations in relationship to Ca^{2+} binding sites and helices. Ca^{2+} binding site III pictured with Ca^{2+} bound (green spheres); Glu134 is located in the G-helix between Ca2+ binding sites III and IV and Asp145 is situated at the Z position of Ca^{2+} coordinating residues of site IV (Modified version of Figures from Ref.[5,127]).

In 2013 Cordina et al.[134] used paramagnetic relaxation enhancement (PRE) as well as chemical shift changes obtained from NMR data to study the structural effect of the A8V mutation on the isolated cTnC. They used chemical shift perturbation analysis to highlight regions that may differ structurally and PRE to indicating the direction of change. PRE is a method which incorporates paramagnetic species into a protein and measures distance dependent relaxation rate enhancement effects on nearby nuclei[134].

In this study, as shown in Figure 1-8, to probe the conformation of the N-domain of cTnC, they made four monocysteine constructs (L12C, C35, L48C, and C84) and labeled each of these mutants with the stable nitroxide spin-label MTSL[134]. They determined that the label would be optimally positioned at C84 on the D-helix of the N-terminal domain. It is important to mention that they report no large-scale perturbations due to the introduction of the label on the cysteine residue. They reported large chemical shift perturbations for adjacent residues located in helix N caused by the A8V mutation in the Ca^{2+} bound state. They also reported a large perturbation within site I and helix D.

In the Ca^{2+} free (apo) state they observed notable difference between the A8V and Wild Type (WT). Disorder was observed in the C-domain Ca^{2+} binding sites III and IV, despite the presence of Mg^{2+}. This mutation provoked significant chemical shift changes throughout the entire N-domain particularly the residues located in the N-helix, site I, and site II. A8V seems to cause more severe structural perturbation affecting the entire cTnC molecule in the apo state[134].

The PRE rate data shows that in the holo state, the A8V alters the N−A interhelical angle. They observed site I residues moving toward the spin-label, and helix B residues moving away from spin-label. In the apo state A8V elucidated helix A and B moving away from the spin-label and Helix C moving closer to the label. Overall A8V cTnC seems to favor a more open conformation compared to WT, in both the apo- and holo- states[134].

C84Y

The C84Y mutation, based on structure from Takeda et al.[16] (PDB 1J1E), is located toward the end of N-terminal domain adjacent to the beginning of the flexible linker in the region that makes contact with the cTnI regulatory region (residues 147–162). CD analysis of C84Y, has shown that secondary structure in the apo state is significantly reduced[128], while in the holo state the structure of C84Y is stabilized. The sulfhydryl group of C84 on cTnC forms a hydrogen bond with the backbone amide of Alanine 151 on cTnI. Although hydrogen binding is maintained with the substitution of the bulky tyrosine side group, the space between these two (flexible linker and cTnI regulatory region) interacting regions is reduced[128].

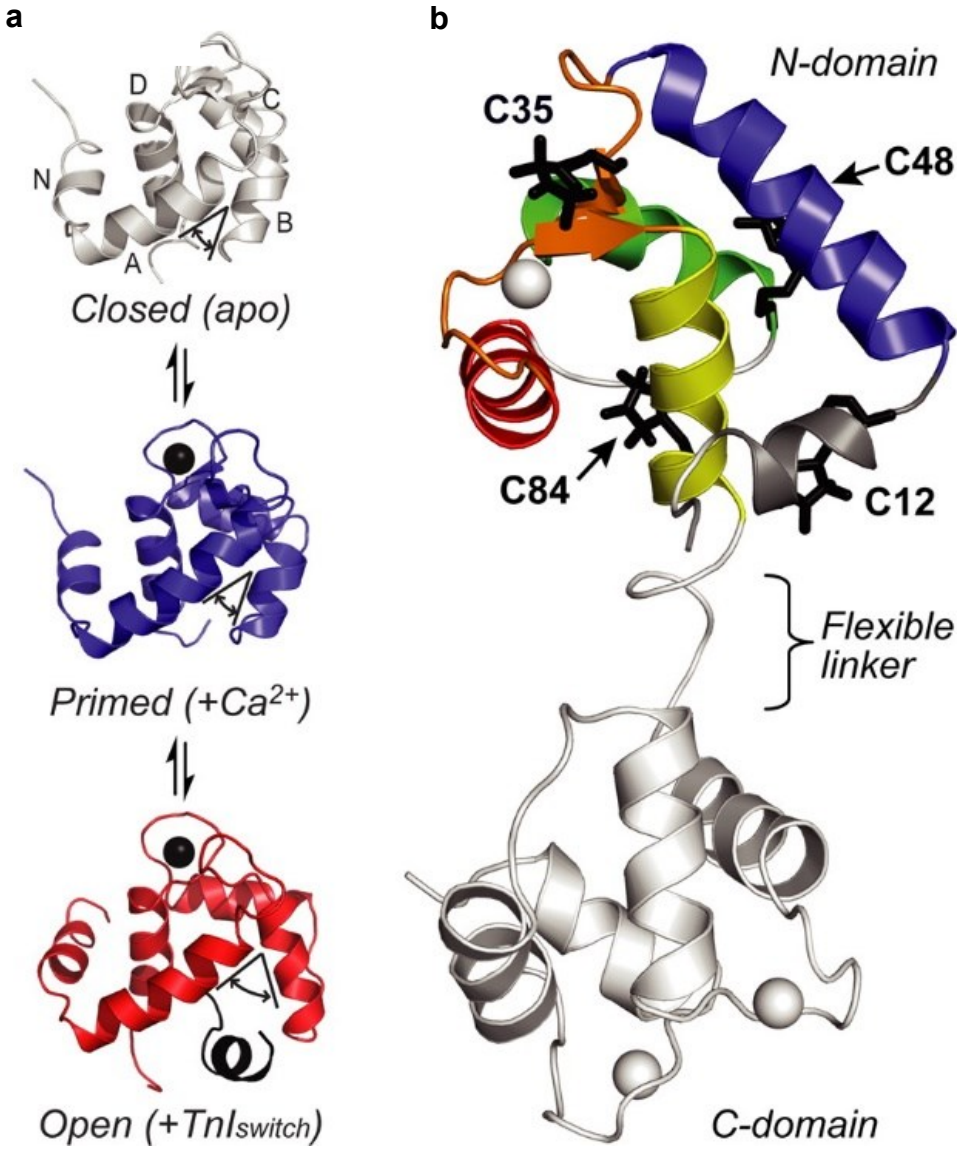


Figure 1-8. NMR structures for the regulatory N-domain of cardiac TnC in various physiological states

a) The N-terminal domain of TnC in the Ca^{2+}-free closed state (gray, PDB: 1SPY[78]), Ca^{2+}-bound primed state (blue, PDB: 1AP4[78]), and Ca^{2+}-bound open state stabilized by the TnIswitch (red, PDB: 1XML[135]). Ca^{2+} and TnIswitch binding results in an increase in the A–B interhelical angle (~40°) and an extension of helix B at the N-terminus. b) Positions of the cysteine mutants created in this study. Helices N and A–D of the N-domain are colored gray, blue, green, red, and yellow, respectively. While there are two Ca^{2+}-binding loops in the N-domain (sites I and II, orange), only site II is capable of binding Ca^{2+} in the cardiac isoform (gray sphere). Four monocysteine constructs (L12C, C35, L48C, and C84) were spin-labeled with MTSL(1-oxyl-2,2,5,5-tetramethylpyrro- line-3-methyl-16-methanethiosulfonate), shown in black, in intact cTnC. This figure is a modified version of figure used in article by Cordinal et al., 2013[134].

E134D

The E134D mutation is located in the C-terminal domain in a region where the N-terminal domain of HcTnC interacts with cTnI. Protein kinase C phosphorylation of residues 43/45 of cTnI induces a conformational change in cTnC helix G centered at residue Glu-134[128,136]. This mutation might change the effect of phosphorylation of cTnI on cTnC and alter the affinity of cTnC for Ca^{2+}. Since PyMol modeling of E134D into 1J1E showed no potential structural changes in the mutation region, the authors of this study concluded that other physiological processes such as phosphorylation of cTnI by protein kinase C or other kinases might be affected by the E134D mutation[128].

D145E

The D145E mutation is located in the +Z position of Ca^{2+} binding site IV and is highly conserved among species and EF-hand proteins. The effect of substitution at the X-coordinate position of the Ca^{2+} binding site on Ca^{2+} binding has been shown in previous studies[137]. Based on CD data the D145E mutation in cTnC causes a decrease in the α-helicity of the holo state of cTnC. Researchers speculate that this might be due to an effect on the helix-loop-helix structure of the EF hand, as Ca^{2+} binding stabilizes the adjacent α-helices[128,129]. The holo state of D145E had shown loss of β-sheet content. This mutation may cause destabilizing effects on the C-terminal β-sheet structure[128].

L29Q

In 2001 Hoffmann et al., found a mutation at position 29 of HcTnC in a 60-year-old male HCM patient[138]. More recently, Gillis et al. (2005) showed that L29, together with neighbouring residues, Asp27, Val28, and Gly30, play an important role in determining the Ca^{2+} affinity of site II[139]. As shown in Figure 1-7, the FHC mutation L29Q is located at the start of the cTnC first EF hand loop. Substitution of a nonpolar leucine to a polar glutamine extends the side chain of residue 29 by one atom. The glutamine side chains point toward the solvent channel between neighbouring protein molecules instead of pointing at cTnC molecule. Zhang el al.[140] had shown that the presence of glutamine's $O^{\varepsilon 1}$ and $N^{\varepsilon 2}$ cause extra hydrogen bonds at the position of residue 29 in the structure. They also found that in comparison to WT the main chain between residues 29 and 32 is shifted towards the solvent channel by up to 0.5 Å[141].

Several groups have separately investigated the molecular mechanism of the L29Q mutation. Schmidtmann et al. (2005) illustrated a decrease in the Ca^{2+} sensitivity measured by ATPase activity and in vitro motility studies using reconstituted fast skeletal muscle myosin, actin, and tropomyosin combined with cardiac troponin[142]. Conversely, Liang et al. (2008) used mouse skinned fibres reconstituted with mouse cTnC, and reported an increase in the Ca^{2+} sensitivity of force development[143]. The differences reported between the two studies are explained by Li et al. (2013), who illustrated that the L29Q mutation alters the Ca^{2+} affinity in isolated cTnC, troponin complexes, reconstituted thin filament preparations, and skinned cardiomyocytes depending on sarcomere length[144]. They also docked the cardiac specific N-terminus of cTnI into binary complex of the N-$cTnC_{L29Q}$ and used simulations to predict structural changes. Their simulated structure suggested that the L29Q mutation causes a significantly more ‘‘open’’ conformation of NcTnC in the presence of pseudo-phosphorylated cTnI[144]. The pseudo-phosphorylation involves mutating Ser23/Ser24, which are the protein kinase A (PKA) phosphorylation sites on cTnI, to aspartate. The replacement of Ser with Asp will introduce a negative charge which mimics the addition of negatively charged phosphate group to the protein.

Despite the effect of the L29Q mutation on Ca^{2+} affinity in mammals, ectothermic species natively have glutamine at position 29 of cTnC[145]. There are 3 other conserved residue differences from human cTnC (D2N, V28I and G30D). These four mutations are collectively called NIQD, and together they increase the Ca^{2+} affinity of cTnC[139,145]. Zhang et al. (2013) solved x-ray crystal structures of Cd^{2+} bound L29Q N-cTnC and NIQD N-cTnC. Through superimposition of WT, L29Q and NIQD NcTnC they showed that there is no significant overall structural difference between the three structures. The all-atom RMSD of L29Q and NIQD from WT were 0.19 Å and 0.12 Å, respectively[140]. They proposed an alternative explanation for the increase in Ca^{2+} affinity wherein substitution of leucine 29 with glutamine and glycine 30 with aspartic acid, which are located on the surface of the protein replace non-polar side chains with polar side chains, this increases the surface hydrophobicity of the protein which leads to the cTnC being more easily activated[146]. These mutations also increase protein solvent interaction and reduce the free energy barrier that has to be overcome for the protein to go through a conformational change and become activated[140,145].

1.3. Molecular dynamics simulation

1.3.1. Equilibrium simulation

Molecular dynamic (MD) is a computer powered simulation of the physical movement of atoms and molecules. Based on this method, atoms and molecules interact for a given time period and provide a view of the motion of the atoms. The molecular system consists of a great number of particles, therefore it is impossible to discover and study the properties of such systems analytically. MD simulations are done to shed light on the structural properties and microscopic interactions of a group of molecules. Molecular Dynamics Simulation solve Newton's equation of motion for a system of N interacting atoms[147]:

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i, i = 1 \dots N.$$

Here r, t and F are position, time, and force, respectively. The forces are the negative derivative of a potential function $U(r_1, r_2, \dots, r_N)$:

$$F_i = -\frac{\partial U}{\partial r_i}$$

In small time steps the equations are solved to calculate the forces acting on the atoms, which are derived from potential function. In MD simulation the forces among the interacting particles depend on particle positions and not on the velocity. Through computing forces the acceleration of an individual atom in the system can be determined easily. By integrating the equation of motion the trajectory, which describes the v (velocity), r and a (acceleration) of the particles varying with time, can be determined. The main idea is that if a system is at state $S(t)$, which is the position r and velocity v of every particle (atom) in the system at time t, then the following states $S(t0 + \Delta t), S(t0 + 2\Delta t)$, are calculated using Newton's law $F = ma$. Here $Fi\ (t0 + n\Delta t)$ is the sum of the forces on i as exerted by the other particles of the system at time $t0 + n\Delta t$. For every particle i the force $Fi\ (t0 + n\Delta t)$ is then integrated to get the new velocity $vi\ (t0 + n\Delta t)$. Using this velocity, for every particle i the new position $ri(t0 + (n + 1)\Delta t)$ can be calculated. The system is run for some time, with temperature and

pressure being kept at required value, and the coordinates are recorded in an output file at regular intervals and collectively referred to as a trajectory[147]. After a period of time depending on the composure of each system, a given system will reach an equilibrium state. For an isolated system a thermodynamic equilibrium is reached when no microscopic change of state are occurring at a measurable rate. Analysis of the output equilibrium trajectory can provide valuable information about macroscopic properties[147]. The MDS software package *GROMACS* (GROningen MAchine for Chemical Simulations) solves Newton's equations of motion iteratively for a system of atoms, providing details about the properties of the system that can then be analyzed or visualized with a number of supplied tools[147].

1.3.2. Free energy

A fundamental thermodynamic equation that describes the free energy for the formation of a protein-ligand complex relates free energy change to changes in enthalpy and entropy ($\Delta Gbind = \Delta H - T\,\Delta S$). The ΔG is the difference in Gibbs free energy of the binding reaction, ΔH and ΔS represent changes in enthalpy and entropy, respectively, and T is the temperature of the system [148]. To calculate this thermodynamic property of cTnC, we used the method below.

Weighted Histogram Analysis Method (WHAM)

A common technique used to calculate the Potential of Mean Force (PMF) along a given reaction coordinate ξ is umbrella sampling. PMF is a type of experiment done in molecular dynamic simulations which computes how a system's energy changes as a function of some specific reaction coordinate parameter. The umbrella sampling is designed to overcome limited sampling at energetically unfavorable configurations by restraining the simulation system with an additional, usually harmonic, potential[149]. The harmonic potential is a dynamic bond between an atom and a point in space. Therefore, a series of N_w separate umbrella simulations are conducted with an umbrella potential:

$$w_i(\xi) = \frac{1}{2} K_i(\xi - \xi_i^c)^2$$

These simulations restrain the system at position $\xi_i^c (i = 1, \dots, N_w)$ with a force constant K_i[149,150]. Umbrella histograms $h_i(\xi)$ are recorded from each N_w umbrella simulation, and are referred to as "umbrella windows". The umbrella histogram $h_i(\xi)$ represents the probability distribution $P_i^b(\xi)$ along the reaction coordinate biased by the umbrella potential $w_i(\xi)$[149]. The most commonly used method to unbias the distribution $P_i^b(\xi)$ and compute the PMF from histograms is the weighted histogram analysis method (WHAM)[149]. Simply, WHAM estimates the statistical uncertainty of the unbiased probability distribution given the umbrella histogram followed by calculating the PMF with the smallest uncertainty[149]. The free energy is the change in PMF of a system along the reaction coordinate.

1.3.3. Molecular dynamics simulation of calcium-sensitizing troponin C mutations

To date none of the FHC associated mutations (A8V, L29Q, C84Y, E134D, and D145E) has been a subject of MDS study. Engineered mutations such as V44Q and L48Q have been explored through simulation, they also have been reported to cause an increase in Ca^{2+} affinity[151,152]. Kekenes et al. [153] studied the structural effects of the V44Q and L48Q mutations on the N-terminal domain of cTnC using MDS. After a 120 ns simulation, through superimposition of the mutated structures with the WT structure they concluded that the mutations preserved the overall arrangement of helices, with relatively minor Cα RMSD differences with respect to wild-type (RMSD ≤ 3.0 Å). While their Cα RMSD did not illustrate a huge change in structure, differences were noted at L_{CD}, the physiologically active Ca^{2+} binding domain. The simulated V44Q mutant structure also showed significant RMSD changes for B-helix, D-helix and AB-loop[153].

The potential of mean force (Section 1.3.2) was calculated along a reaction coordinate, which is the distance from the Ca^{2+} and E76 of site II for the apo wild-type structure form. Kekenes et al.[153] found that the PMF decreases as the Ca^{2+} approaches the binding site until it is 0.4 nm away, at which point the Ca^{2+} encounters a 12.5 KJ/mol energy barrier. Analysis of the trajectory showed that the largest energy barrier is caused by a change in the coordination of Ca^{2+} atom from bi-dentate to mono-dentate and reorganization of the five water molecule coordinating Ca^{2+} [153].

Kekenes et al.[153] hypothesized that tighter Ca^{2+} coordination at site II is necessary for higher Ca^{2+} affinity. Therefore, they measured the mean residue/Ca^{2+} distance for chelating residues (D65, D67, S69, T71 and E76). In general, chelation distances were consistent and well-stabilized across cases, with no distinction between WT and mutated structures[153].

Since previous simulations of WT TnC has shown a significant decrease in site II flexibility upon Ca^{2+} binding[73], they hypothesized that mutations might affect the conformational dynamics of site II. To address this hypothesis, they analyzed the relative flexibility of the N-cTnC backbone, using computational estimates of S^2 based on the N-H bond vector, and illustrated that Sites I and II of WT and mutated structures are more dynamic in the presence of Ca^{2+}. They noted a region of increased dynamics near residue 50 of L_{BC}, which was not present in WT. They believed that destabilization is due to disruption of the hydrophobic packing between B-helix and C-helix induced by mutations[153]. They also illustrated that L_{CD} of the mutated structure in the apo state is considerably less dynamic and Ca^{2+} had no pronounced effect on the simulated order parameters. Lower flexibility in site II residues while the protein is in the apo state means that this region has to go through a smaller change in entropy upon Ca^{2+} binding[153].

1.3.4. Statement of the research project & specific aims

Despite the extensive amount of research focusing on the effect of these disease-causing mutations, little work has been done to discover their exact underlying structural mechanism. The purpose of the study is to provide greater insight into the mechanisms by which these mutations alter the structure and function of cTnC. This research project will seek to extend the previous studies by answering the following questions.

1. What is the effect of HCM-associated mutations on the structure of HcTnC?
    - This will be addressed by examining the overall arrangement, conformational dynamics of site II, and Ca^{2+}/Site II interaction.
2. What is the effect of mutations on the function of HcTnC, which ultimately alter calcium affinity?
    - This will be addressed by calculating the free energy of Ca^{2+} binding of the regulatory site II of HcTnC in WT and mutant proteins.

Chapter 2. Methods

2.1. Structural Models

SWISS-MODEL[154–156] was used to model the initial structures for WT and mutated HcTnC. *SWISS-MODEL* is a fully automated protein structure homology-modelling server, accessible via the ExPASy web server. The WT[57] and mutated structures are based on sequences shown in appendix A. The structure which was used as a template was the X-ray crystal structure of the human cardiac troponin complex Takeda et al.[18] (PDB code: 1J1E chain A). Each of the structures was modeled in the Ca^{2+} bound state.

2.2. Equilibrium simulation

Each of the simulations was performed using *GROMACS Version 4.5.4.* Topology files were created from the SWISS-MODEL derived models through *pdb2gmx*. Command details are described in appendix B. The force field selected for these simulations was *AMBER-03*[157], as it has been demonstrated to be successful for simulating bio molecular systems over a long period of time[157]. The water model used was *SPC/E*, which has been shown to emulate the experimentally derived thermodynamic behavior of water at 300 K[158]. *editconf* was used to create an octahedral box around the model with 0.9 nm distance between the edge of the model and the boundary of the box. Water was added into the system by repeating a small pre-equilibrated system of water coordinates (*spc216.gro*) through the volume of the box, followed by the removal of any water molecules that overlap with the protein. The net charge of the system was set to 0 by replacing randomly selected water molecules with Na^+ or Cl^- ions. This step was followed by energy minimization of the system.

The hydrogens added to the initial x-ray crystal structure, and broken hydrogen bond network in water could lead to structural distortion, but energy minimization mitigates these effects by removing structural clashes or unfavorable interactions between any atoms in the system. Energy minimization was performed in two stages, steepest descents, followed by conjugate gradient minimization to a maximum force of 1000 kJ mol^{-1} nm^{-1}. The steepest descent gradient simply takes a step in the direction of the negative gradient, without any consideration of the history built up in previous steps. The step size is adjusted such that the search is fast, but the motion is always downhill [147]. Conversely, conjugate gradient method uses gradient information from previous steps. Overall, steepest descents will bring the system close to the nearest local minimum very quickly, while conjugate gradients brings it very close to the local minimum, but performs worse far away from the minimum.

The last step before the production molecular dynamics simulation is to equilibrate the water around the restrained protein molecule. This permits a complete solvation of the protein beyond what is possible through energy minimization and prevents distortion of the protein that may occur if an unrestrained molecular dynamics simulation is started with an incompletely solvated protein molecule (Appendix B: 8-9 restrained equilibration)[159].

The production simulation used the same parameters as the position restrained equilibration, with the removal of the restraints on the protein molecule (Appendix B., G). The temperature and pressure of a system tends to drift due to integration errors as well as frictional and external forces. Therefore, there is a weak coupling to an external heat bath and the deviation of system from a reference temperature was corrected through exponential decay of temperature deviation. The coupling temperature was 300 K. Similar to temperature coupling, the system was also coupled to a pressure bath. The reference pressure for coupling was 1.0 bar (0.98 atm). The pressure coupling algorithms act to adjust box vectors and scale the coordinates of the atoms within the unit cell[147]. This intervention, therefore, kept the pressure constant through allowing the volume and shape to fluctuate. This simulation proceeded for 100 ns, and was carried out on the high-performance computing clusters “parallel” and “lattice” provided by WestGrid (www.westgrid.ca) and Compute Canada Calcul Canada (www.computecanada.ca).

The structural coordinates output by the production MD simulation are energy minimized using the steepest descents method to a maximum force of 1000 kJ mol^{-1} nm^{-1} (Appendix B: 12 & 13; Appendix E). This places the HcTnC molecule into the closest local energy minimum. The trajectory was also processed by trjconv to correct any artifacts resulting from the use of periodic boundary conditions during the simulation. Three replicate simulations for each mutant and WT structure were performed.

2.3. Cα Root Mean Square Deviation (RMSD) with respect to wild type

The Cα-RMSD is the measure of the average distance between the corresponding backbone atoms of superimposed proteins. The RMSD equation reads:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}$$

Here, δ is the distance between N pairs of equivalent Cα atoms. To establish the Cα-RMSD (Å) while correcting for the drift between the relative positions of N- and C-terminal domains of TnC, the minimized coordinates produced from the production simulations described in section 2.2 were divided into separate files for the N-terminal domain (residues 1-89) and C-terminal domain of the TnC molecule (Residues 90-161). The N- and C- terminal domains produced from each of the replicated simulations were superimposed using the SuperPose webserver version 1.0 and produced the RMSD for Cα atoms. Similarly, the Cα-RMSD was calculated for each of the domains over the final 50 ns of the production simulations using g_rms (Appendix H: 1 & 2).

2.4. Inter-helical angles

To measure the angle between the helix A (H_A) and helix B (H_B), we used the *Interhlx*[160] program. The residue spans for H_A and H_B were 14-27 and 41-47, respectively.

2.5. Calcium coordinating distances

The tool g_dist was used to measure the distance (Å) between the Ca^{2+} and site II coordinating (chelating) residues (ASP65-OD1, ASP65-OD2, ASP67-OD1, ASP67-OD2, SER69-OG, THR71-O, THR71-OG1, GLU76OE1, and GLU76-OE2). g_dist measures the distance between pairs of specific atoms, this process was automated for each of the chelating atoms using a bash script (Appendix I). This script reports the distances between each coordinating residue and Ca^{2+} over the 100 ns simulation. Distances calculated for the final 10 ns are averaged to account for thermal fluctuation. In the cases in which the coordinating ligand is an aspartic or glutamic acid, the distances between each oxygen atom and the Ca^{2+} were averaged and represented as one distance for each residue.

2.6. Conformational dynamics of HcTnC structures

The Root Mean Square Fluctuation (RMSF) is a measure of the deviation between the position of an atom and some reference position. The RMSF represents the conformational dynamics (flexibility) of HcTnC. This was calculated using the Gromacs g_rmsf tool. g_rmsf, for each frame in the simulation, calculated the RMSD of an atom from its average position and divides the sum of RMSDs by the total number frames being considered.

$$RMSF = \frac{\sum RMSD}{j}$$

Here j is the total number of frames being considered. In general, this tool computes the RMSF of atomic positions in the trajectory (whole-nojump.trr) after fitting to a reference frame (gro.1.tpr) (Appendix J).

2.7. Free Energy Calculations

The change in free energy of binding was calculated using Potential of Mean Force (PMF) in 3 Steps: Pull simulation, Umbrella sampling, and WHAM.

To perform umbrella sampling it is essential to generate a series of configurations along a reaction coordinate, ζ (distance between site II and Ca^{2+}). Some of these configurations will serve as the starting configurations for the umbrella sampling windows, which run in independent simulations[161] (Figure 2-1). In this case a pull simulation was used to define a series of configurations along the reaction coordinate. In Figure 2-1, the dashed arrow represents the Ca^{2+} in configurations extracted from the pull simulation. The green sphere (Ca^{2+}) between the two solid black lines corresponds to the independent simulations conducted within each sampling window, with the center of mass of the free peptide restrained in that window by an umbrella biasing potential. The distance between each configuration was 0.1 nm for the first 2 nm and 0.2 nm for the rest of the pull. The histogram in the bottom of Figure 2-1 shows the ideal result as a histogram of configurations, with neighboring windows overlapping such that a continuous energy function can later be derived from these simulations.

The energy minimized output coordinates from the equilibration simulations were used as inputs for the free energy simulations.

2.7.1. Pull simulations

The initial coordinates were re-parameterized using pdb2gmx to accommodate the use of the *Amber03*[157] force field and TIP3P water model. A triclinic simulation box was defined with dimensions sufficient for the Ca^{2+} ion to travel 5 nm along the Z-axis (Appendix K: 2 & 3).

The TnC molecule was oriented in the simulation box such that an unobstructed path along the Z-axis could be used to remove the calcium from site 2; this orientation was identical for each simulation to prevent energetic variation between mutants due to differences in the reaction coordinate.

The system was solvated and the charged balanced as described for the equilibration simulations (section 2.2) This was followed by steepest decent energy minimization and 1 ns of NPT equilibration (Appendix K: 7-10). The NPT ensemble involves keeping the number of particles (N), system pressure (P) and temperature (T) constant / conserved[147]. Sets of restraints were created for the site II Ca^{2+}, and the Cα

atoms of alpha helical residues. Ca^{2+} movement was restrained in the X and Y direction while the alpha carbons were restrained in all directions with harmonic potentials of 1000 kJ mol^{-1} nm^{-2} (Appendix K: 14-17).

The pull simulation was run until the site II Ca^{2+} atom was separated from the center of mass of the site II atoms by at least 5 nm over 20 ns. Individual frames with Calcium to Site II distances were selected at 0.1 nm intervals for the first 2 nm, or 0.2 nm intervals for the final 3 nm were selected for umbrella sampling (Appendix K: 18-26).

2.7.2. Umbrella Sampling

The preparation for umbrella sampling involves generating a number of input files in order to conduct each of the necessary simulations (Appendix K: 27 & 28). This is followed by a 1 ns NPT equilibration step for each simulation. This process was automated using a bash script shown in Appendix Q. Many of the pulling parameters are the same as those in the sampling simulation, with the exception the pull_rate1, which was then set to zero. This restrains the calcium within a defined window of configurational space. (Appendix K: 29 & 30). Simulations were run on the Lattice and Parallel clusters provided by West Grid. The umbrella sampling on each configuration was used to restrain it within a window corresponding to the chosen COM distance. It outputs a *.xvg (distribution) file which reports the position of the Ca^{2+} as a function of time in each frame of the simulation.

2.7.1. Weighted Histogram Analysis Method (WHAM)

The Weighted Histogram Analysis Method (WHAM) is included in GROMACS as the g_wham utility. The input for g_wham consists of two files, one that lists the names of the *.tpr files of each sampling window, and the other that lists the names of either the pullf.xvg or pullx.xvg files from each sampling window. To automate this step a bash script was used, which renames every umbrella.tpr and pullf-umbrella.xvg file to umbrella-(frame number).tpr and pullf-(frame number).xvg. The script then copies them all into a directory along with a text file listing all the names. Finally the g_wham (Section 1.3.2) tool was used to analyze the output files generated by the umbrella sampling simulations to compute a potential of mean force (PMF)[149] (Appendix U). The G_WHAM

program outputs a histogram file and an energy profile file. The histogram was used to verify that there was sufficient overlap between adjacent windows, and the energy profile *file was used to compute the PMF.*

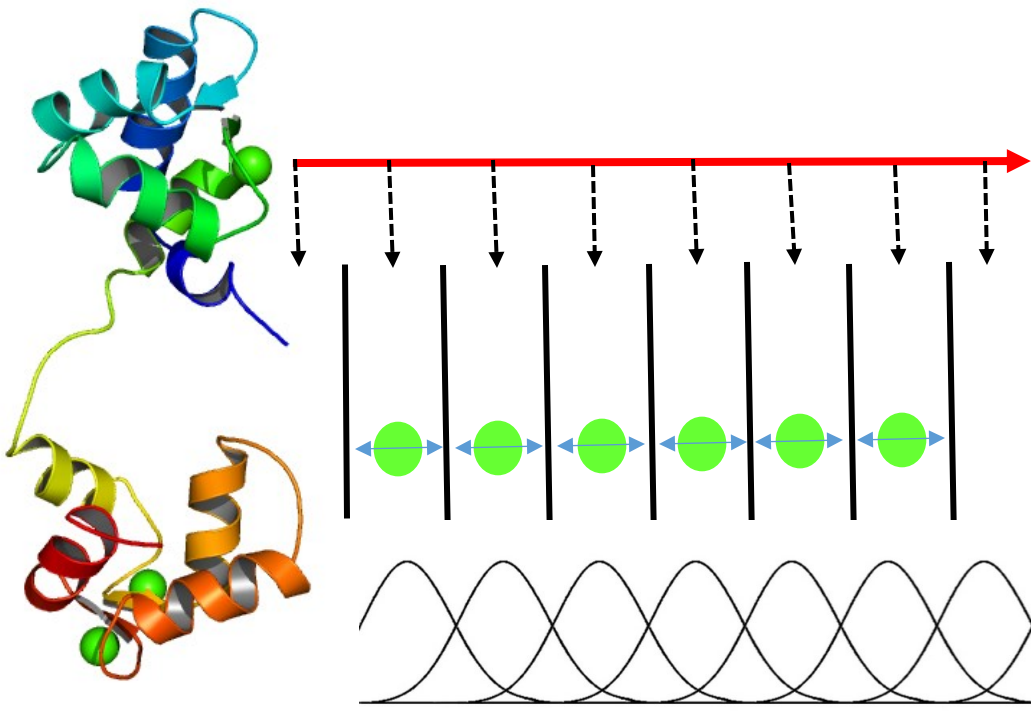


Figure 2-1. Schematic representation of pull simulation and generating configurations

The green sphere represents Ca^{2+}. The red arrow represents the reaction coordinate which the pull simulation runs along. The dashed arrows illustrate the configurations that were extracted after the simulation is complete. The green sphere between the two solid lines represent the independent simulations conducted within each sampling window, with the center of mass of the free peptide restrained in that window by an umbrella biasing potential. The histogram in the bottom shows an ideal histogram of configuration. The NMR structure is HcTnC (PDB code: 1AJ4)[162].

Chapter 3. Results and Discussion

3.1. Quality Assessment

Ramachandran plots were used to analyze the quality of the simulated (100 ns) HcTnC structures. The plots are shown in Figure 3-1. No Ramachandran outliers were detected. The RMSD (Å) of the output from 3 replicated simulations were compared to the RMSD (Å) of the final 50 ns of each of the single simulations and showed no significant difference (Table 3-1). The RMSD (Å) of output structures of the 3 replicates represents the deviation among the simulated structures. Similarly, RMSD of the last 50 ns of a single structure illustrates the deviation within a structure along the course of simulation. The lack of significant difference between the two shows that deviation among the structures is similar to deviation within the structure. Therefore, any deviation of mutated structures from the WT (shown in next section) is not due to variation among replicated runs.

Table 3-1. Cα RMSD (Å) of HcTnC structures.

Cα RMSD (Å) of 3 simulation replicates shown for each HcTnC structure as well as Cα RMSD (Å) of each structure over the last 50 (ns) of the simulation.

Mutations	RMSD of final minimized output structures from 3 replicates (Å)	RMSD over the final 50 ns of single replicate (Å)
A8V	1.53	1.55
L29Q	1.71	1.59
C84Y	1.37	1.57
E134D	1.55	1.54
D145E	1.69	1.54
WT	1.37	1.54

3.2. Calcium coordination distance

To test the hypotheses that tighter Ca^{2+} / site II interaction is essential for increased binding affinity, we measured the mean coordination distances between Ca^{2+} and the coordinating residues A65, Asp67, Ser69, Thr71, and Glu 76. The measured distances are reported in Table 3-2. The Standard Error of the Mean (SEM) is ± 1% of the reported values. We found no significant differences and concluded that coordination distances are consistent across WT and mutant structures. This is consistent with a previous study[153] that showed no association between Ca^{2+} affinity and coordinating distances.

Table 3-2. Mean residue/ Ca^{2+} distance (Å) for site II coordinating residues of final minimized WT and mutated HcTnC.

·	A8V	L29Q	C84Y	E134D	D145E	WT
Asp65 (+X)	3.1	3.2	3.4	3.3	2.8	2.9
Asp67 (-Y)	3.3	3.4	3.3	3.4	3.2	3.2
Ser69 (Z)	3.3	3.6	2.8	4.0	4.0	2.8
Thr71 (Y)	2.7	2.7	2.7	2.7	2.7	2.7
Glu76 (-Z)	2.7	2.7	2.6	2.7	2.6	2.6

The SEM is ± 10% of the reported value.

3.3. Superimposition of mutant and WT

To understand the impact of mutations on the HcTnC structures, the N terminal domain (residues 1-89) and C terminal domain (residues 90-161) of the final minimized mutant and WT structures were superimposed. The Cα-RMSD (single replicate) of mutant N and C terminal domains with respect to WT HcTnC are reported Table 3-3. When these are compared to the RMSD of the variation within a single mutant HcTnC simulation, which are 1.1 Å and 0.9 Å for N and C terminal domains, respectively, over the last 50 ns of simulation, it demonstrates that the deviation is greater than what we would expect along the course of a mutant structure simulation. To visualize the

superimposition we used PyMol to model the N and C terminal domains of simulated mutant into simulated WT structure.

Through superimposition of N-terminal domain final minimized structures, differences were noted at L_{NA}, L_{AB}, $L_{BC,}$ helix A (H_A), and helix B (H_B) of all mutant structures, but most predominantly in A8V, E134D and D145E. The L_{AB} and L_{BC} are the interaction sites for cTnI regulatory and N-terminal domains (Figure 1-5), respectively. In physiological systems, cTnI binding is shown to enhance Ca^{2+} affinity of cTnC. This is mainly due to activation of cross bridges and subsequent alteration of myofilament lattice properties during rigor[153,163]. However, a requirement for the formation of the open state facilitating TnI regulatory domain binding is the exposure of a hydrophobic patch between residues L29, A31, K39, and E66[153,164]. The transition to the open state, involves the progression of helices B and C (BC) away from helices N, A, and D (NAD). This is shown to be associated with enhanced Ca^{2+} affinity observed for multiple mutations along the BC/NAD interface[165]. Our simulation results show increase in exposure of this region for mutant HcTnC structures, which is evident from increase in the angle between Helices A and B (H_A and H_B) (Table 3-4). As illustrated in Figure 3-2, H_A and H_B of A8V and E134D structures present a larger opening (Table 3-4.

Calculated angle between Helix A and B compared favorably to WT. The larger opening can facilitate an enhanced interaction between the hydrophobic patch of cTnC and the cTnI regulatory domain and ultimately affect the Ca^{2+} affinity[166,167]. This finding is consistent with the Huskey et al.[153] results which showed similar increases in H_A/H_B angle for engineered HcTnC mutations (V44Q) that have been reported to exhibit increase in Ca^{2+} affinity in skinned cardiac fibers[151,152]. The same group, illustrated in a microsecond timescale simulations of TnC that the V44Q mutation reduces the free energy barrier to increasing the H_A/H_B angle, which may increase the rate of forming the open state[168]. Unfortunately, we do not have the computer resources available to test the former result.

Table 3-3. Cα RMSD (Å) of final minimized mutant N and C terminal domains with respect to WT HcTnC.

	RMSD (Å) of N-terminal domain (1-89)		RMSD (Å) of C-terminal domain (90-161)	
Mutations	**Compared to WT**	**Over the final 50 ns**	**Compared to WT**	**Over the final 50 ns**
A8V	3.3	1.1	2.6	0.9
L29Q	2.6	1.1	2.7	0.9
C84Y	3.1	1.1	2.6	0.9
E134D	5.1	1.1	2.1	0.9
D145E	3.9	1.1	4.6	0.9

Table 3-4. Calculated angle between Helix A and B of final minimized structure

HcTnC structure	H_A/H_B angle (°)
A8V	68
L29Q	43
C84Y	47
E134D	79
D145E	50
WT	32

For the L29Q N-terminal domain, we observed deviations in helix C and L_{CD} (Figure 3-3). The L_{CD} is the Ca^{2+} binding loop and part of interaction site for cTnI regulatory domain (Figure 1-5). The overall structure of the N-terminal domain seems to have minimal deviation compared to other mutations. This is consistent with the prior studies reporting minimal overall structural changes[141]. Although similar in size, the substitution of a non-polar leucine to a polar glutamine alters the hydrogen bonding properties of the side chain. It has been proposed that EF1 (L_{AB}) and EF2 loops (L_{CD}) are structurally linked and mutations in L_{AB} can alter the structure of the L_{CD} in cNTnC and thus may affect the Ca^{2+} binding affinity[78,145]. Contrary to results reported by Zhang et al.[141] which through superimposition of cTnC N-terminal showed no L_{CD} structural difference among WT and L29Q, our results indicate deviation at this loop. The

discrepancy in our results might be either due to the fact that Zhang et al. had only used the N-terminal domain or the nature of the two studies. Their study was based on x-ray structures which are static in nature vs. our study which investigates a dynamic model of flexible structure.

The superimposition of the C-terminal domain of mutant TnC with WT HcTnC structures did not strongly reveal differences between WT and A8V, L29Q, C84Y, and E134D structures. In case of D145E, the helix G is shifted 90º and the L_{FG} and L_{GH} are displaced (Figure 3-4). The α-helicity of the F-helix is reduced by one residue and the Ca^{2+} is ejected from the site IV. The ejection of Ca^{2+} is either due to reduced affinity of site IV or simply a simulation error. But since it has been shown previously that the D145E mutation significantly reduces the Ca^{2+} binding affinity of site IV, we are inclined to believe that the observed result is due to reduced affinity and not an error. Previous studies have also illustrated that interruption of binding of Ca^{2+} to site III or IV of skeletal TnC results in the increased Ca^{2+} sensitivity of force development[137,169]. Therefore, we tend to believe that the structural changes observed (helix G, $L_{FG,}$ and L_{GH}) play an indirect role in modulation of Ca^{2+} affinity through alteration of TnC interactions with TnT and TnI. As our current model which is being studied lacks TnI and TnT, further structural studies are required to investigate the mechanism by with D145E effects this interaction.

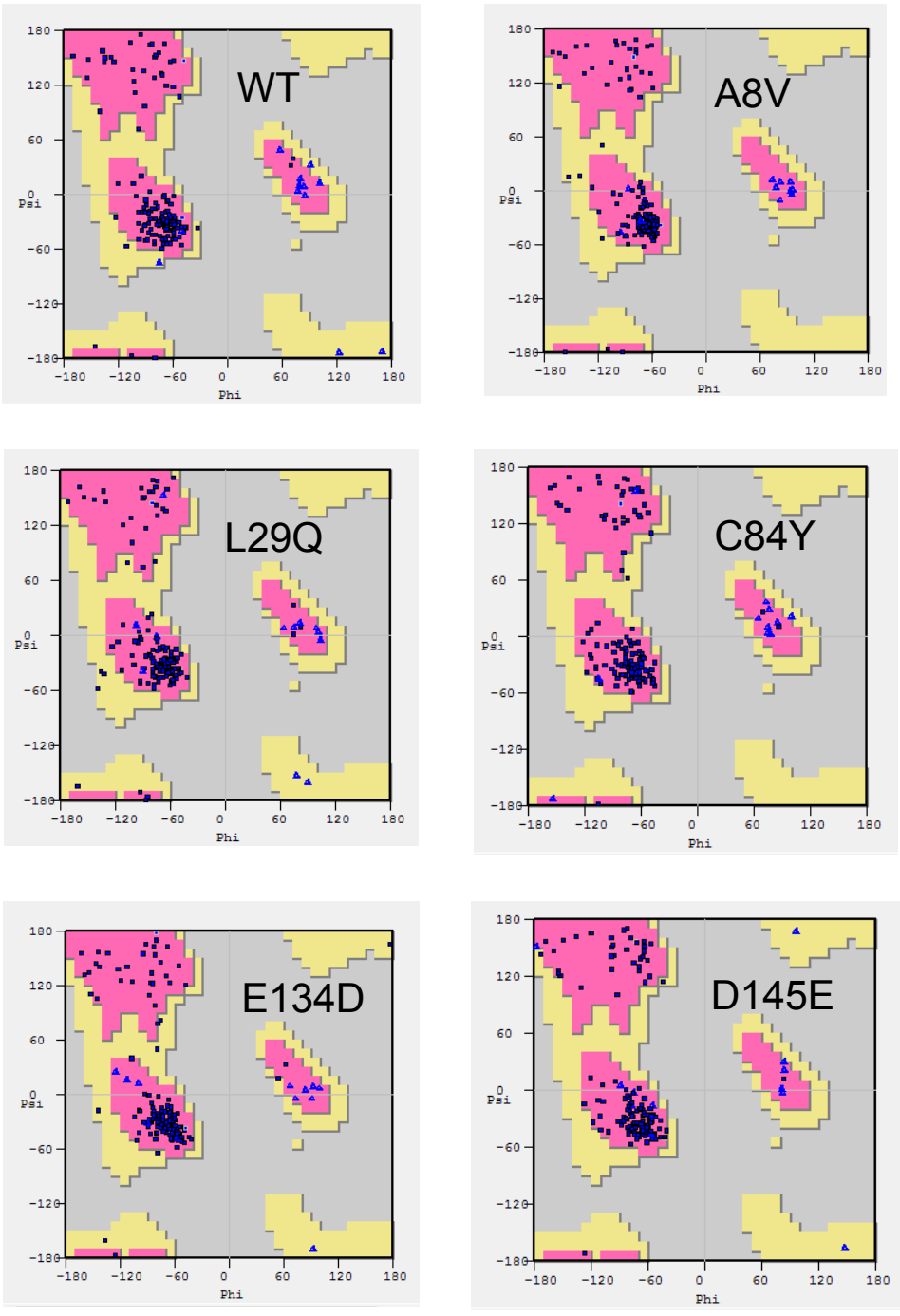


Figure 3-1. Ramachandran plots of WT and mutated HcTnC after 100 ns simulation.

Pink and yellow areas represent preferred and allowed regions respectively. Dark gray represents the disallowed region.

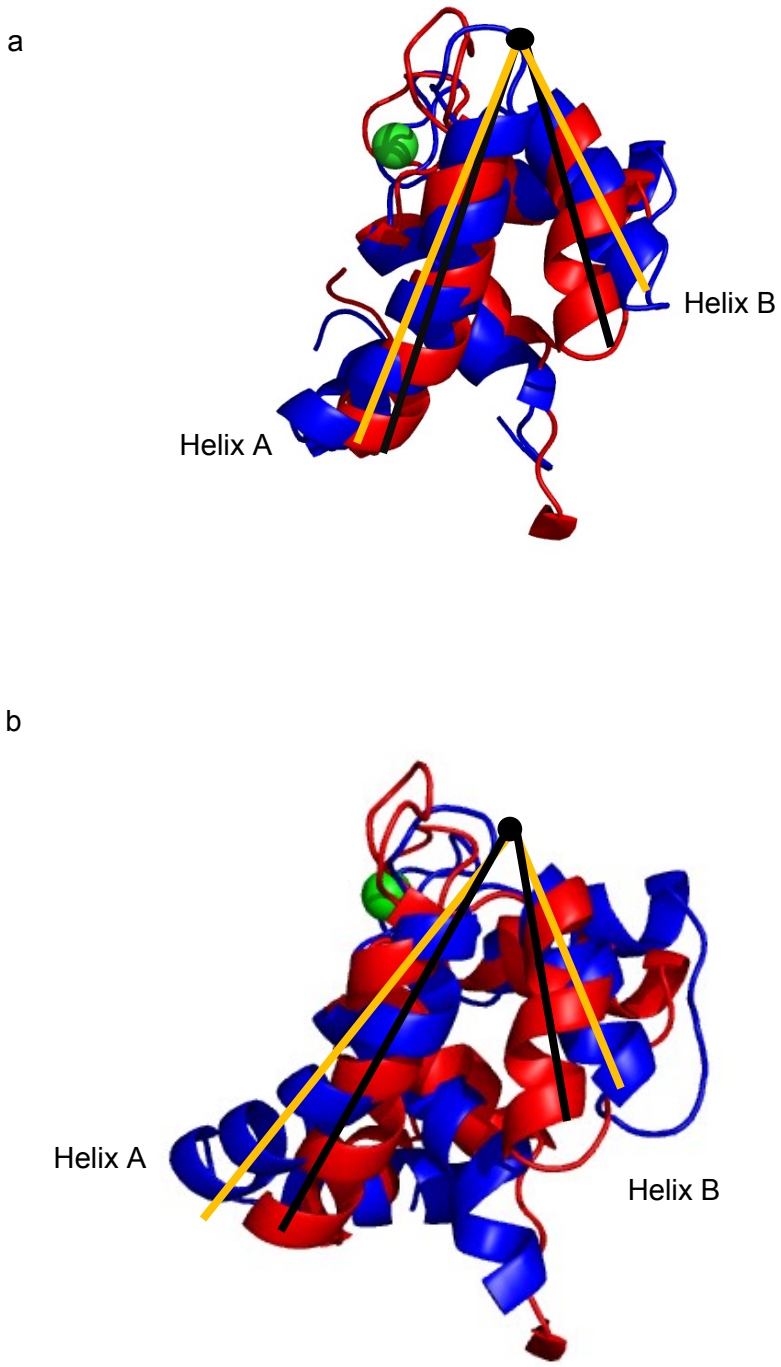


Figure 3-2. N-terminal domain (1-89) of final minimized structures of simulated A8V and E134D superimposed on WT HcTnC structure.

a) Red and blue represent the final minimized structure of simulated N-terminal WT and A8V, respectively. b) Red and blue represent the final minimized structure of simulated N-terminal WT and E134D, respectively. The yellow lines illustrates a more open conformation of mutant compared to WT (black lines). The angles between Helices A and B are reported in Table 3-4.

WT
L29Q

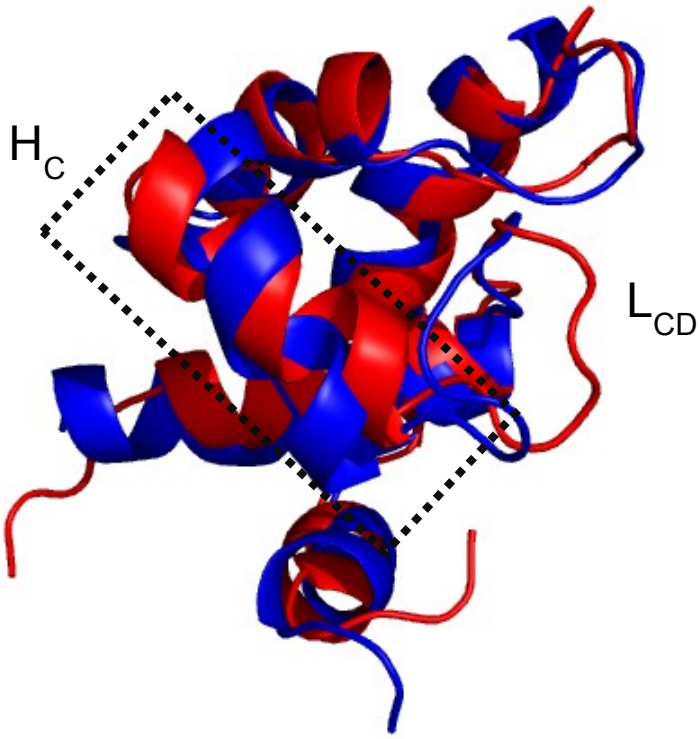


Figure 3-3. N-terminal domain (1-89) of final minimized structure of L29Q superimposed on WT HcTnC structure.

Red and blue represent the final minimized structure of simulated N-terminal WT and L29Q, respectively. Black dotted box marks Helix C (H_C). Loop CD is labeled as L_{CD}.

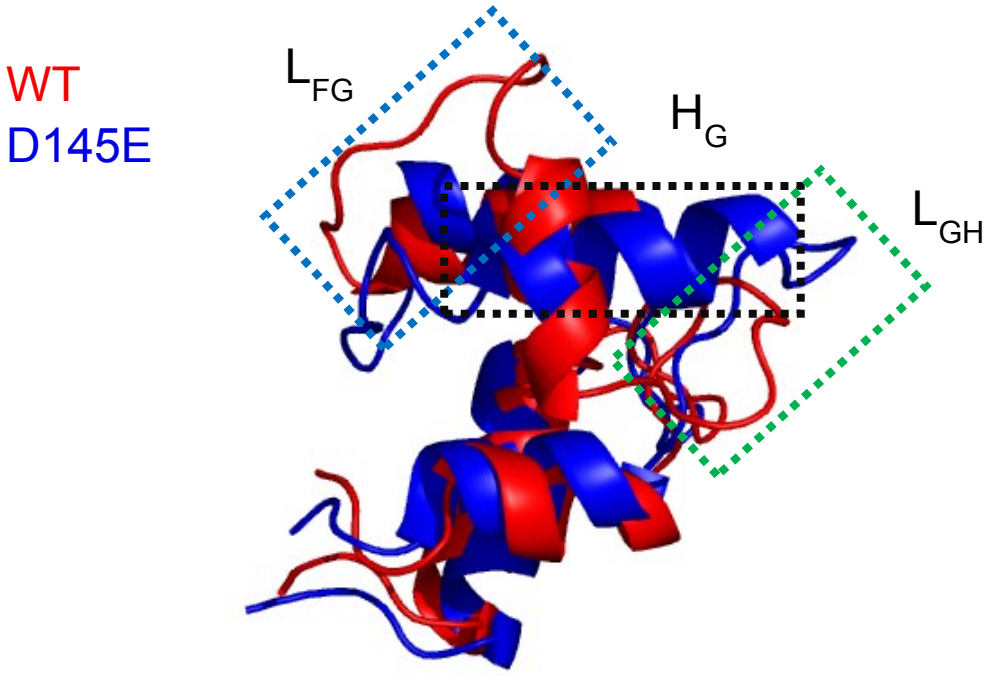


Figure 3-4. C-terminal domain (90-161) of final minimized structure of D145E superimposed on WT HcTnC structure.

Red and blue represent the final minimized structures of simulated C-terminal WT and D145E, respectively. Green, black, and blue boxes represent Loop GH (L_{GH}), Helix G (H_G), and Loop FG (L_{FG}), respectively.

3.4. Conformational dynamics

We measured the RMSF (Å) as an indicator of conformational dynamics. Our result indicates that site II (65-73) of A8V, C84Y and E134D are less dynamic (≈ 0.5-1.0 Å) compared to WT (Figure 3-5). Conversely, L29Q showed higher dynamics (≈ 1.0 Å) compared to WT and D145E which exhibit no change in this region. The change in site II dynamic does not correlate with physical data of these HCM mutations. The reported Ca^{2+} sensitivity of force development for A8V, L29Q, C84Y, and D145E are increased, while E134D does not have an effect[127,128,143,144]. Changes in site II conformational dynamics does not reveal a trend discriminating the HCM mutations which are known to increase the Ca^{2+} affinity. We furthermore noted that the conformational dynamic of L_{FG} for A8V, L29Q, E134D, and D145E is increased (≈ 1.0-2.0 Å) with respect to WT. The L_{FG} is the interaction site for H1-TnI (Figure 1-5). There are no experimental data available to correlate the change in the interaction of H1-TnI/TnC with the Ca^{2+} affinity. Future studies are required to test the effect of increase in L_{FG} conformational dynamics on the H1-TnI/Tnc interaction and the effect of altering the H1-TnI/TnC interaction on Ca^{2+} interaction.

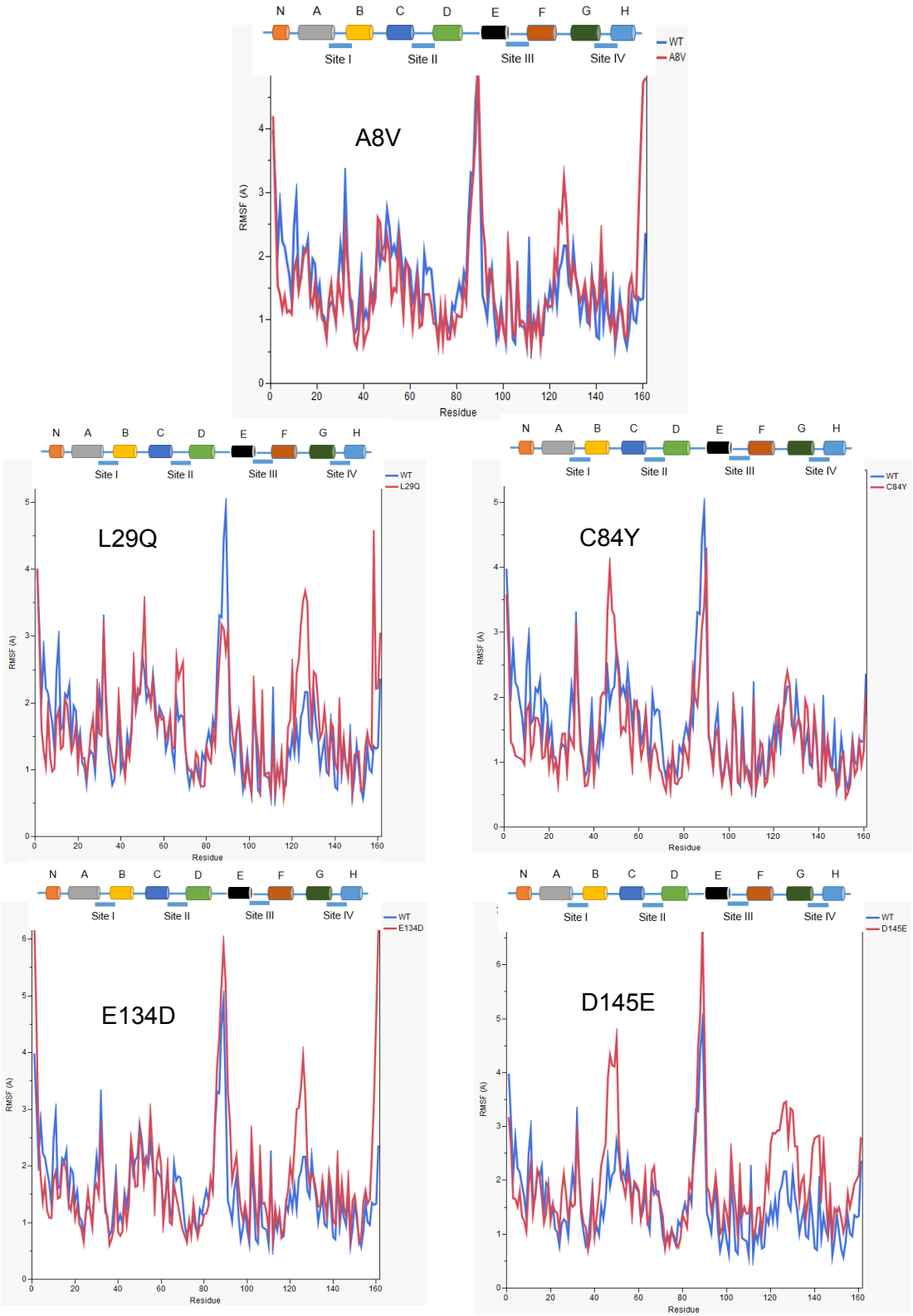


Figure 3-5. RMSF (Å) of simulated mutant and WT HcTnC structure.
Red and blue curves represent mutant and WT structure, respectively. The top panel illustrates HcTnC secondary structure. Helices are shown using cylinders and loops are represented using blue solid line.

3.5. Free Energy calculations

To investigate the effect of FHC-related mutations on the free energy barrier separating the unbound and bound state, we calculated the Potential of Mean Force (PMF) along a reaction coordinate representing the distance from the center of mass of site II and Ca^{2+} (see Figure 2-1). The estimated PMF graphs for mutated and WT HcTnC (single replicate) are shown in Figure 3-6. The value of ΔG (change in Gibbs free energy) is calculated as the difference between the average of the plateau region of the PMF curve (SEM≤ 0.4 KJ/mol) and the energy minimum of the curve (average of first umbrella sampling frame). The ΔG of mutated and WT HcTnC are reported in Table 3-5. The SEM reported in the table is the SEM of the difference between average of the plateau region and the average of first frame. The drops in the PMF function curves are either the energy barrier between Ca^{2+} bound and unbound states or the under sampling along the reaction coordinate. The histogram for mutated and WT HcTnC umbrella sampling (single replicate) is shown in Figure 3-7. The distance ranges associated with drops in the PMF are due to under sampling in those sections. The overlap around areas marked by red rectangles indicates that more sampling windows will improve the continuity of the PMF function (Figure 3-6).

Table 3-5. Change in free energy of binding reported for mutated and WT HcTnC.

HcTnC	Estimated ΔG ± SEM
A8V	-32 ± 3
L29Q	-22 ± 2
C84Y	-25 ± 2
E134D	-18 ± 3
D145E	-42 ± 2
WT	-29 ± 1

Thermodynamic values are reported in kJ per mol.

Because site II is a solvent-exposed binding site and it has numerous negatively charged residues that can coordinate Ca^{2+}, we can anticipate a highly favorable enthalpic interaction as Ca^{2+} binds site II. A recent Isothermal Titration Calorimetry (ITC) study by Skoronsky et al.[170] has shown Ca^{2+} binding to the N-terminal domain to be

endothermic and supported by favorable entropic factors. Their calculated ΔG of Binding of Ca^{2+} to the N-terminal domain of HcTnC is shown to be energetically favorable (≈ -30 kJ/mol) but the condition-dependent enthalpy (ΔH_{ITC}) and entropy (ΔS_{ITC}) of Ca^{2+} binding demonstrate that the process is endothermic and entropically driven[170]. Based on this finding, we hypothesized that Ca^{2+} binding to site II of HCM mutations should be more favorable than WT and increase the Ca^{2+} affinity. The A8V and D145E mutations showed support for the hypothesis with the greatest increase in Ca^{2+} sensitivity of force production[127]. The C84Y, E134D, and L29Q had a lower ΔG, compared to WT. The disagreement between our hypothesis and the observed ΔG results for E134D, L29Q, and C84Y mutations may be due to the energetic cost of restraining the Cα of helical residues of HcTnC to prevent the TnC molecule from moving with the mobile Ca^{2+}. Some of the structural rearrangements that may occur due to binding and release of Ca^{2+} are prevented. This can lower the entropy of the molecule and affect the ΔG. Further work is required to develop a method that rules out this possibility. Additionally, the physical data available for changes in Ca^{2+} affinity are recorded in the presence of TnI and TnT. In our experiment we were investigating these mutations in isolated HcTnC, the absence of TnI can disrupt the transfer of some inter-domain structural changes.

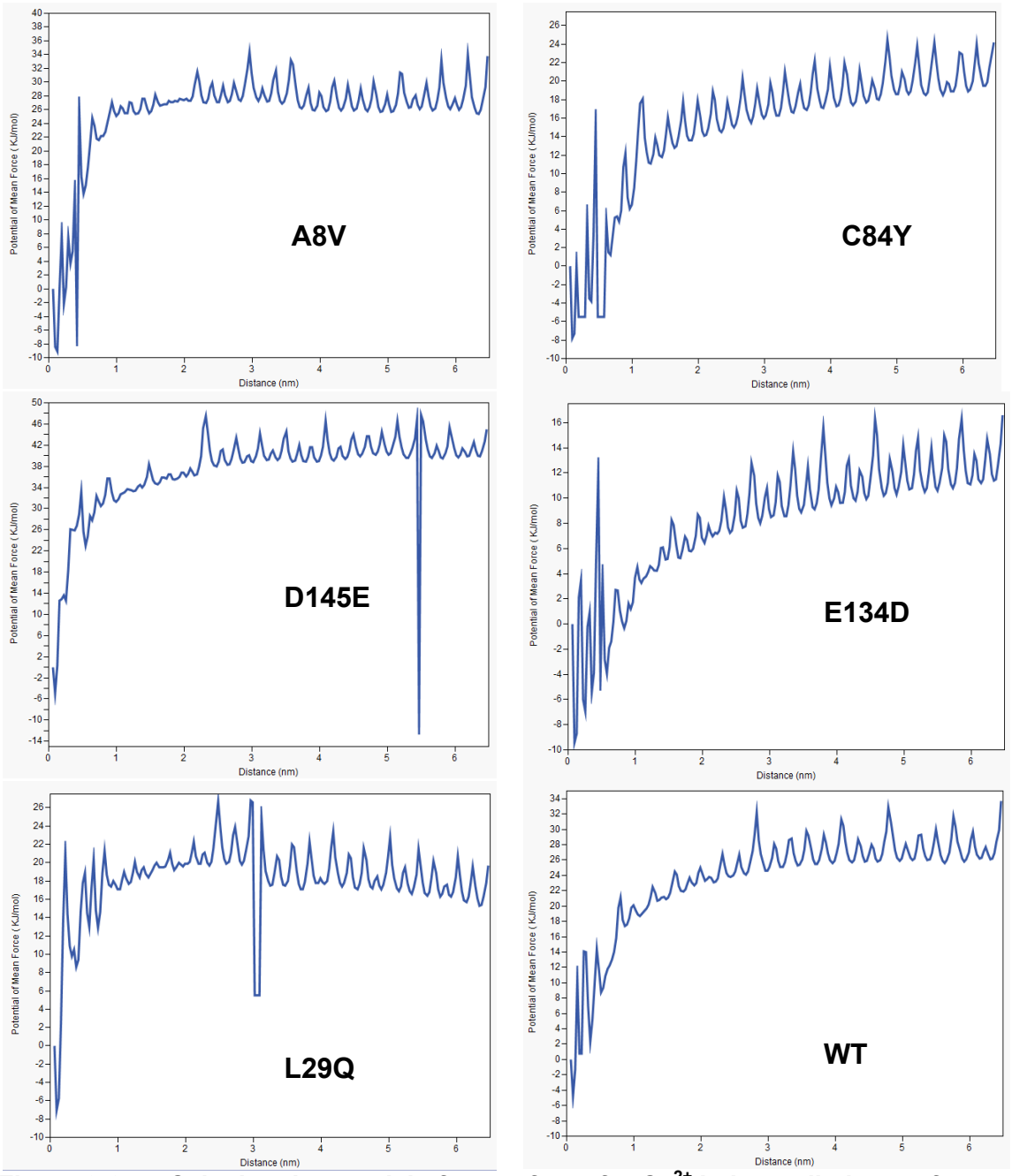
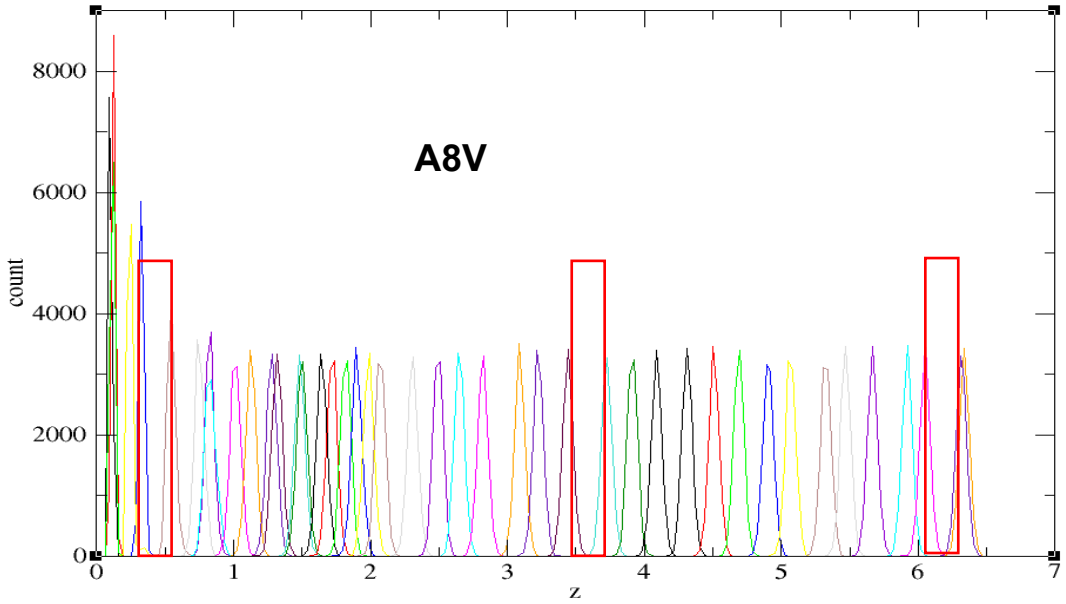
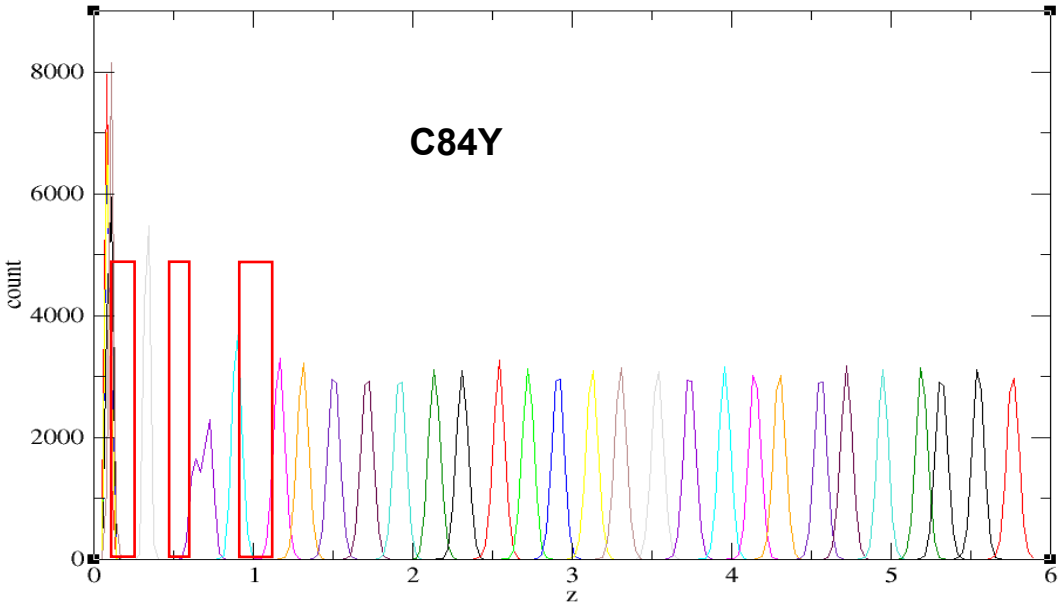
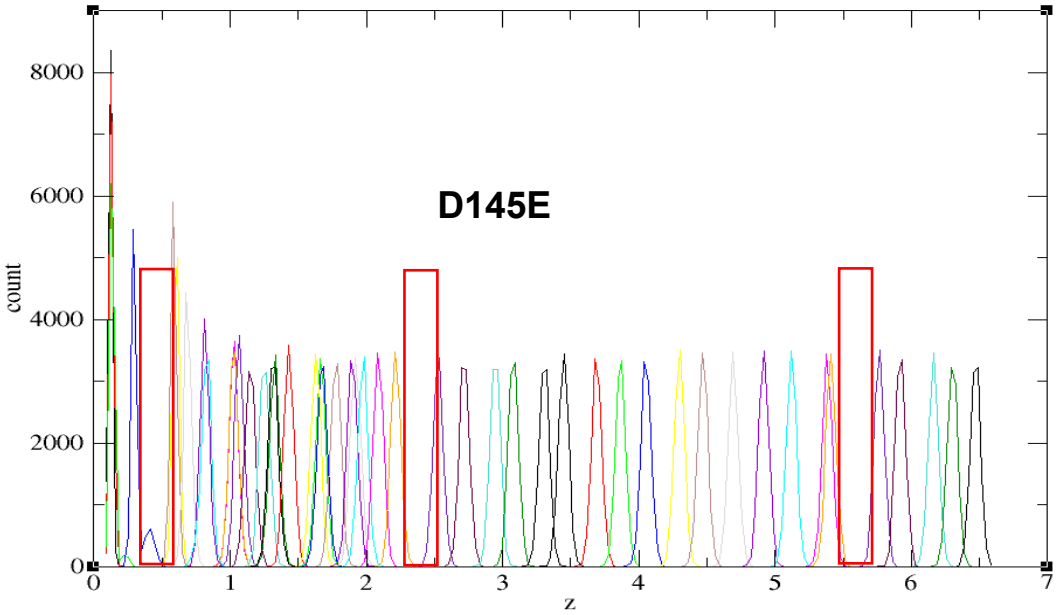


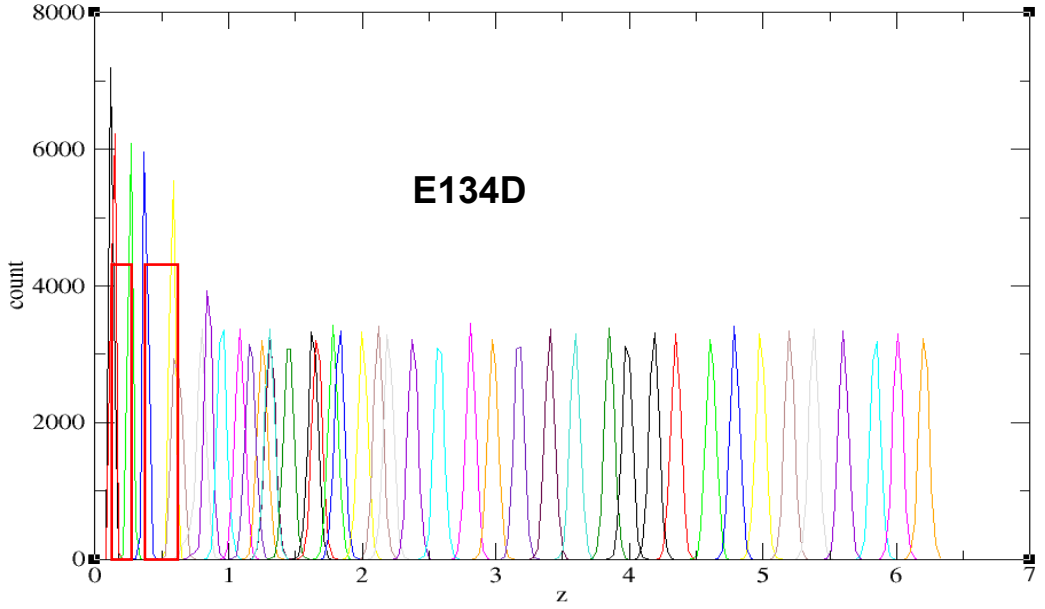
Figure 3-6. Calculated potential of mean force for Ca^{2+} being pulled away from site II along reaction coordinate.

Potential of mean force (KJ/mol) was computed by the distance between Ca^{2+} and center of mass of site II (residue 65-73).


A8V
count
z
0
2000
4000
6000
8000
0
1
2
3
4
5
6
7


C84Y
count
z
0
2000
4000
6000
8000
0
1
2
3
4
5
6


D145E
count
z
8000
6000
4000
2000
0
0
1
2
3
4
5
6
7


E134D
count
z
8000
6000
4000
2000
0
0
1
2
3
4
5
6
7

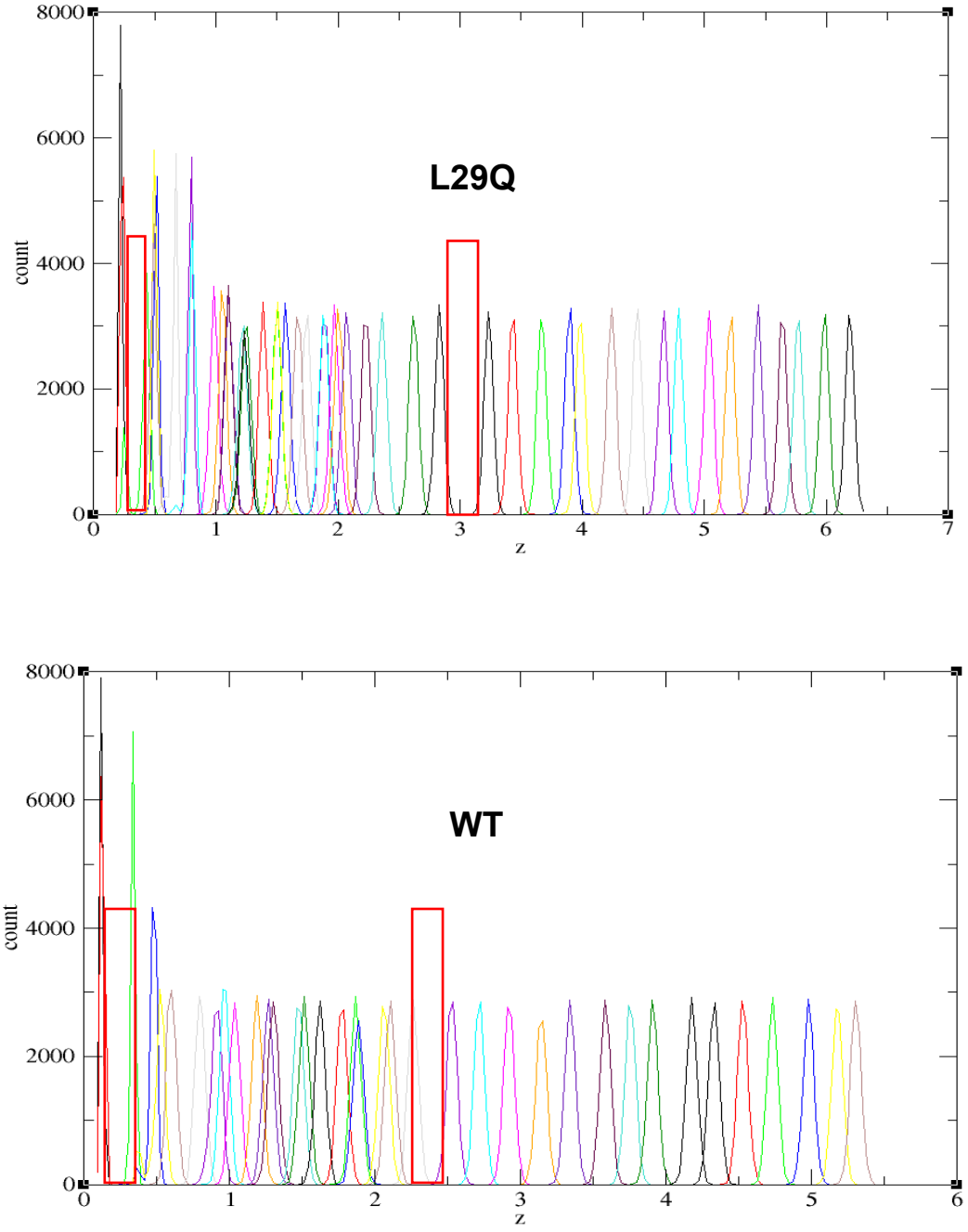


Figure 3-7. Histogram of umbrella sampling.
Histogram of mutated and WT HcTnC umbrella sampling simulation. The Y-axis represents number of counts for the umbrella sampling histogram at each z nm (distance) value (X-axis). Areas marked by red rectangles are determined to have insufficient overlaps.

3.6. Conclusion

We have simulated 5 HCM associated mutations and a wild type structure of a full length human cardiac troponin to identify, through structural and functional analysis, a trend that readily discriminates between mutations and WT. Our results suggest that difference in Ca^{2+} affinity cannot be explained in terms of overall structural changes and Ca^{2+} coordination. Alternatively, the alteration in Ca^{2+} binding affinity might be due to modified dynamics in cTnC/cTnI interaction sites as well as a more open conformation, which facilitate the enhanced interaction. Overall, we observed changes in the arrangement of helices in N-terminal domain. The H_A/H_B angle was increased in all HCM mutations compared to WT. This result is consistent with other studies which investigated engineered mutations such as V44Q[153] and HCM-associated mutations such as L29Q[144]. In relation to the literature, our work corroborates the current paradigm in which a larger H_A/H_B angle is associated with increased in Ca^{2+} affinity. As marking the axis of helix A and B for angle measurement was difficult in some cases, we hope to develop a more reproducible method which can better quantify the open conformation in 3-dimensional space.

We noted Ca^{2+} binding site II (L_{CD}) of the A8V, C84Y, and E134D mutations to exhibit lower and, conversely L29Q to shower higher dynamics. Previous studies[153] had shown the important role of site II dynamics in Ca^{2+} binding. Originally we speculated that lower conformation dynamic reflected more stable interaction of Ca^{2+}/site II. Our results do not completely agree with the original hypothesis and only 3 out of 5 (A8V, C84Y, E134D) mutations showed more stable site II and one (L29Q) showed the opposite effect. The lack of correlation between site II dynamic and affinity might be due to the lack of normally distributed data which can be resolved by increasing the number of simulations. The group[153] that reported a more stable site II for engineered HCM-like mutations limited their simulations to the N-terminal domain and only analyzed N-terminal mutations. Therefore, the discrepancy in these results might be due to the absence of the HcTnC C-terminal domain in their experiment.

The interaction site for H1-TnI (L_{FG}) was more dynamic than WT for A8V, L29Q, E134D, and D145E. This is a novel finding that can be tested experimentally by

examining the effect of this change on the dynamics of interactions and by testing for correlation between changes in H1-TnI/TnC interaction and changes in Ca^{2+} affinity.

Mutations that increase the Ca^{2+} affinity of cTnC are affiliated with hypertrophic cardiomyopathy[127,138,139,141,144,171]. Understanding how the thermodynamics of Ca^{2+} binding changes is an important clue to the management of these disorders. The thermodynamics of Ca^{2+} binding to HcTnC is a cornerstone of future studies that include a comparison with other proteins within the EF hand family. Our results did not completely support our original hypothesis. Originally we hypothesized that higher ΔG of Ca^{2+} binding makes the Ca^{2+}/site II interaction more favorable and ultimately increase the Ca^{2+} affinity. We observed this effect in A8V and D145E but the hypothesis did not hold true for C84Y, E134D and L29Q. Our results indicate that the method of measuring the free energy of Ca^{2+} binding of the isolated HcTnC is not sufficient to describe the effects of mutations on affinity.

References

(1) Pollard, T. D. (2008) Cell biology. Saunders/Elsevier,, Philadelphia :

(2) Wade, R., Eddy, R., Shows, T. B., and Kedes, L. (1990) cDNA sequence, tissue-specific expression, and chromosomal mapping of the human slow-twitch skeletal muscle isoform of troponin I. *Genomics 7*, 346–57.

(3) Tiso, N., Rampoldi, L., Pallavicini, A., Zimbello, R., Pandolfo, D., Valle, G., Lanfranchi, G., and Danieli, G. A. (1997) Fine mapping of five human skeletal muscle genes: alpha-tropomyosin, beta-tropomyosin, troponin-I slow-twitch, troponin-I fast-twitch, and troponin-C fast. *Biochem. Biophys. Res. Commun. 230*, 347–50.

(4) Bhavsar, P. K., Brand, N. J., Yacoub, M. H., and Barton, P. J. (1996) Isolation and characterization of the human cardiac troponin I gene (TNNI3). *Genomics 35*, 11–23.

(5) Katrukha, I. a. (2013) Human cardiac troponin complex. Structure and functions. *Biochem. Biokhimiia 78*, 1447–65.

(6) Vallins, W. J., Brand, N. J., Dabhade, N., Butler-Browne, G., Yacoub, M. H., and Barton, P. J. R. (1990) Molecular cloning of human cardiac troponin I using polymerase chain reaction. *FEBS Lett. 270*, 57–61.

(7) Zabrouskov, V., Ge, Y., Schwartz, J., and Walker, J. W. (2008) Unraveling molecular complexity of phosphorylated human cardiac troponin I by top down electron capture dissociation/electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics 7*, 1838–49.

(8) Mittmann, K., Jaquet, K., and Heilmeyer, L. M. G. (1990) A common motif of two adjacent phosphoserines in bovine, rabbit and human cardiac troponin I. *FEBS Lett. 273*, 41–45.

(9) Howarth, J. W., Meller, J., Solaro, R. J., Trewhella, J., and Rosevear, P. R. (2007) Phosphorylation-dependent conformational transition of the cardiac specific N-extension of troponin I in cardiac troponin. *J. Mol. Biol. 373*, 706–22.

(10) Ferrières, G., Pugnière, M., Mani, J.-C., Villard, S., Laprade, M., Doutre, P., Pau, B., and Granier, C. (2000) Systematic mapping of regions of human cardiac troponin I involved in binding to cardiac troponin C: N- and C-terminal low affinity contributing regions. *FEBS Lett. 479*, 99–105.

(11) Finley, N., Abbott, M. B., Abusamhadneh, E., Gaponenko, V., Dong, W., Gasmi-Seabrook, G., Howarth, J. W., Rance, M., Solaro, R. J., Cheung, H. C., and Rosevear, P. R. (1999) NMR analysis of cardiac troponin C-troponin I complexes: effects of phosphorylation. *FEBS Lett. 453*, 107–112.

(12) Ward, D. G., Brewer, S. M., Calvert, M. J., Gallon, C. E., Gao, Y., and Trayer, I. P. (2004) Characterization of the interaction between the N-terminal extension of human cardiac troponin I and troponin C. *Biochemistry 43*, 4020–7.

(13) Ward, D. G., Brewer, S. M., Gallon, C. E., Gao, Y., Levine, B. A., and Trayer, I. P. (2004) NMR and mutagenesis studies on the phosphorylation region of human cardiac troponin I. *Biochemistry 43*, 5772–81.

(14) Warren, C. M., Kobayashi, T., and Solaro, R. J. (2009) Sites of intra- and intermolecular cross-linking of the N-terminal extension of troponin I in human cardiac whole troponin complex. *J. Biol. Chem. 284*, 14258–66.

(15) Howarth, J. W., Meller, J., Solaro, R. J., Trewhella, J., and Rosevear, P. R. (2007) Phosphorylation-dependent conformational transition of the cardiac specific N-extension of troponin I in cardiac troponin. *J. Mol. Biol. 373*, 706–22.

(16) TAKEDA, S., YAMASHITA, A., MAEDA, K., and MAEDA, Y. Structure of the cote domain of human cardiac troponin in the Ca^{2+}-saturated form. *Nature 424*, 35–41.

(17) Takeda, S. (2005) Crystal structure of troponin and the molecular mechanism of muscle regulation *54*, 35–41.

(18) TAKEDA, S., YAMASHITA, A., MAEDA, K., and MAEDA, Y. Structure of the core domain of human cardiac troponin in the Ca^{2+}-saturated form. *Nature 424*, 35–41.

(19) Vassylyev, D. G., Takeda, S., Wakatsuki, S., Maeda, K., and Maeda, Y. (1998) Crystal structure of troponin C in complex with troponin I fragment at 2.3-A resolution. *Proc. Natl. Acad. Sci. 95*, 4847–4852.

(20) Stefancsik, R., Jha, P. K., and Sarkar, S. (1998) Identification and mutagenesis of a highly conserved domain in troponin T responsible for troponin I binding: Potential role for coiled coil interaction. *Proc. Natl. Acad. Sci. 95*, 957–962.

(21) Kobayashi, T., Patrick, S. E., and Kobayashi, M. (2009) Ala scanning of the inhibitory region of cardiac troponin I. *J. Biol. Chem. 284*, 20052–60.

(22) Lindhout, D. A., and Sykes, B. D. (2003) Structure and dynamics of the C-domain of human cardiac troponin C in complex with the inhibitory region of human cardiac troponin I. *J. Biol. Chem. 278*, 27024–34.

(23) Brown, L. J., Sale, K. L., Hills, R., Rouviere, C., Song, L., Zhang, X., and Fajer, P. G. (2002) Structure of the inhibitory region of troponin by site directed spin

labeling electron paramagnetic resonance. *Proc. Natl. Acad. Sci. U. S. A. 99*, 12765–70.

(24) Dong, W.-J., Robinson, J. M., Stagg, S., Xing, J., and Cheung, H. C. (2003) Ca^{2+}-induced conformational transition in the inhibitory and regulatory regions of cardiac troponin I. *J. Biol. Chem. 278*, 8686–92.

(25) Farah, C. S., Miyamoto, C. A., Ramos, C. H., da Silva, A. C., Quaggio, R. B., Fujimori, K., Smillie, L. B., and Reinach, F. C. (1994) Structural and regulatory functions of the NH2- and COOH-terminal regions of skeletal muscle troponin I. *J. Biol. Chem. 269*, 5230–40.

(26) Talbot, J., and Hodges, R. (1981) Synthetic studies on the inhibitory region of rabbit skeletal troponin I. Relationship of amino acid sequence to biological activity. *J. Biol. Chem. 256*, 2798–2802.

(27) Li, M. X., Spyracopoulos, L., and Sykes, B. D. (1999) Binding of cardiac troponin-I147-163 induces a structural opening in human cardiac troponin-C. *Biochemistry 38*, 8289–98.

(28) Wang, X., Li, M. X., and Sykes, B. D. (2002) Structure of the regulatory N-domain of human cardiac troponin C in complex with human cardiac troponin I147-163 and bepridil. *J. Biol. Chem. 277*, 31124–33.

(29) Blumenschein, T. M. A., Stone, D. B., Fletterick, R. J., Mendelson, R. A., and Sykes, B. D. (2006) Dynamics of the C-terminal region of TnI in the troponin complex in solution. *Biophys. J. 90*, 2436–44.

(30) Julien, O., Mercier, P., Allen, C. N., Fisette, O., Ramos, C. H. I., Lagüe, P., Blumenschein, T. M. A., and Sykes, B. D. (2011) Is there nascent structure in the intrinsically disordered region of troponin I? *Proteins 79*, 1240–50.

(31) Galińska, A., Hatch, V., Craig, R., Murphy, A. M., Van Eyk, J. E., Wang, C.-L. A., Lehman, W., and Foster, D. B. (2010) The C terminus of cardiac troponin I stabilizes the Ca^{2+}-activated state of tropomyosin on actin filaments. *Circ. Res. 106*, 705–11.

(32) Galińska-Rakoczy, A., Engel, P., Xu, C., Jung, H., Craig, R., Tobacman, L. S., and Lehman, W. (2008) Structural basis for the regulation of muscle contraction by troponin and tropomyosin. *J. Mol. Biol. 379*, 929–35.

(33) Tobacman, L. S. (1996) Thin filament-mediated regulation of cardiac contraction. *Annu. Rev. Physiol. 58*, 447–81.

(34) Perry, S. V. (1998) Troponin T: genetics, properties and function. *J. Muscle Res. Cell Motil. 19*, 575–602.

(35) Mesnard, L., Logeart, D., Taviaux, S., Diriong, S., Mercadier, J. J., and Samson, F. (1995) Human cardiac troponin T: cloning and expression of new isoforms in the normal and failing heart. *Circ. Res. 76*, 687–92.

(36) Townsend, P. J., Farza, H., MacGeoch, C., Spurr, N. K., Wade, R., Gahlmann, R., Yacoub, M. H., and Barton, P. J. (1994) Human cardiac troponin T: identification of fetal isoforms and assignment of the TNNT2 locus to chromosome 1q. *Genomics 21*, 311–6.

(37) White, S. P., Cohen, C., and Phillips, G. N. (1987) Structure of co-crystals of tropomyosin and troponin. *Nature 325*, 826–8.

(38) Wei, B., and Jin, J.-P. (2011) Troponin T isoforms and posttranscriptional modifications: Evolution, regulation and function. *Arch. Biochem. Biophys. 505*, 144–54.

(39) Gollapudi, S. K., Mamidi, R., Mallampalli, S. L., and Chandra, M. (2012) The N-terminal extension of cardiac troponin T stabilizes the blocked state of cardiac thin filament. *Biophys. J. 103*, 940–8.

(40) Pearlstone, J. R., and Smillie, L. B. (1982) Binding of troponin-T fragments to several types of tropomyosin. Sensitivity to Ca^{2+} in the presence of troponin-C. *J. Biol. Chem. 257*, 10587–92.

(41) Verin, A. D., and Gusev, N. B. (1988) [The role of the N-terminal fragment of troponin T in the interaction with troponin-tropomyosin complex components of the heart]. *Biokhimiia (Moscow, Russ. 53*, 1235–46.

(42) Wang, S., George, S. E., Davis, J. P., and Johnson, J. D. (1998) Structural determinants of Ca^{2+} exchange and affinity in the C terminal of cardiac troponin C. *Biochemistry 37*, 14539–44.

(43) Jin, J. P., Chen, A., Ogut, O., and Huang, Q. Q. (2000) Conformational modulation of slow skeletal muscle troponin T by an NH(2)-terminal metal-binding extension. *Am. J. Physiol. Cell Physiol. 279*, C1067–77.

(44) Biesiadecki, B. J., Chong, S. M., Nosek, T. M., and Jin, J.-P. (2007) Troponin T core structure and the regulatory NH2-terminal variable region. *Biochemistry 46*, 1368–79.

(45) Chandra, M., Montgomery, D. E., Kim, J. J., and Solaro, R. J. (1999) The N-terminal region of troponin T is essential for the maximal activation of rat cardiac myofilaments. *J. Mol. Cell. Cardiol. 31*, 867–80.

(46) Sumandea, M. P., Vahebi, S., Sumandea, C. A., Garcia-Cazarin, M. L., Staidle, J., and Homsher, E. (2009) Impact of cardiac troponin T N-terminal deletion and phosphorylation on myofilament function. *Biochemistry 48*, 7722–31.

(47) Mamidi, R., Michael, J. J., Muthuchamy, M., and Chandra, M. (2013) Interplay between the overlapping ends of tropomyosin and the N terminus of cardiac troponin T affects tropomyosin states on actin. *FASEB J. 27*, 3848–59.

(48) Gollapudi, S. K., Gallon, C. E., and Chandra, M. (2013) The tropomyosin binding region of cardiac troponin T modulates crossbridge recruitment dynamics in rat cardiac muscle fibers. *J. Mol. Biol. 425*, 1565–81.

(49) Jin, J.-P., and Chong, S. M. (2010) Localization of the two tropomyosin-binding sites of troponin T. *Arch. Biochem. Biophys. 500*, 144–50.

(50) Pearlstone, J. R., and Smillie, L. B. (1983) Effects of troponin-I plus-C on the binding of troponin-T and its fragments to alpha-tropomyosin. Ca^{2+} sensitivity and cooperativity. *J. Biol. Chem. 258*, 2534–42.

(51) Murakami, K., Stewart, M., Nozawa, K., Tomii, K., Kudou, N., Igarashi, N., Shirakihara, Y., Wakatsuki, S., Yasunaga, T., and Wakabayashi, T. (2008) Structural basis for tropomyosin overlap in thin (actin) filaments and the generation of a molecular swivel by troponin-T. *Proc. Natl. Acad. Sci. U. S. A. 105*, 7200–5.

(52) Manning, E. P., Tardiff, J. C., and Schwartz, S. D. (2011) A model of calcium activation of the cardiac thin filament. *Biochemistry 50*, 7405–13.

(53) Chong, P. C., and Hodges, R. S. (1982) Photochemical cross-linking between rabbit skeletal troponin and alpha-tropomyosin. Attachment of the photoaffinity probe N-(4-azidobenzoyl-[2-3H]glycyl)-S-(2-thiopyridyl)-cysteine to cysteine 190 of alpha-tropomyosin. *J. Biol. Chem. 257*, 9152–60.

(54) Morris, E. P., and Lehrer, S. S. (1984) Troponin-tropomyosin interactions. Fluorescence studies of the binding of troponin, troponin T, and chymotryptic troponin T fragments to specifically labeled tropomyosin. *Biochemistry 23*, 2214–20.

(55) Tanokura, M., Tawada, Y., and Ohtsuki, I. (1982) Chymotryptic subfragments of troponin T from rabbit skeletal muscle. I. Determination of the primary structure. *J. Biochem. 91*, 1257–65.

(56) Franklin, A. J., Baxley, T., Kobayashi, T., and Chalovich, J. M. (2012) The C-terminus of troponin T is essential for maintaining the inactive state of regulated actin. *Biophys. J. 102*, 2536–44.

(57) Roher, A., Lieska, N., and Spitz, W. (1986) The amino acid sequence of human cardiac troponin-C. *Muscle Nerve 9*, 73–7.

(58) Lewit-Bentley, A., and Réty, S. (2000) EF-hand calcium-binding proteins. *Curr. Opin. Struct. Biol. 10*, 637–43.

(59) Sia, S. K., Li, M. X., Spyracopoulos, L., Gagné, S. M., Liu, W., Putkey, J. A., and Sykes, B. D. (1997) Structure of cardiac muscle troponin C unexpectedly reveals a closed regulatory domain. *J. Biol. Chem. 272*, 18216–21.

(60) Van Eerd, J. P., and Takahashi, K. (1975) The amino acid sequence of bovine cardiac tamponin-C. Comparison with rabbit skeletal troponin-C. *Biochem. Biophys. Res. Commun. 64*, 122–7.

(61) Gifford, J. L., Walsh, M. P., and Vogel, H. J. (2007) Structures and metal-ion-binding properties of the Ca^{2+}-binding helix-loop-helix EF-hand motifs. *Biochem. J. 405*, 199–221.

(62) Taylor, P., Varughese, J. F., Chalovich, J. M., and Lit, Y. Molecular Dynamics Studies on Troponin (Tnl-TnT- TnC) Complexes : Insight into the Regulation of Muscle Contraction Molecular Dynamics Studies on Troponin (TnI-TnT-TnC) Complexes : Insight into the Regulation of Muscle Contraction 37–41.

(63) Reece, K. L., and Moss, R. L. (2008) Intramolecular interactions in the N-domain of cardiac troponin C are important determinants of calcium sensitivity of force development. *Biochemistry 47*, 5139–46.

(64) Ikebe, M., and Hartshorne, D. J. (1985) Phosphorylation of smooth muscle myosin at two distinct sites by myosin light chain kinase. *J. Biol. Chem. 260*, 10027–31.

(65) Gordon, A. M., Homsher, E., and Regnier, M. (2000) Regulation of contraction in striated muscle. *Physiol. Rev. 80*, 853–924.

(66) Kobayashi, T., and Solaro, R. J. (2005) Calcium, thin filaments, and the integrative biology of cardiac contractility. *Annu. Rev. Physiol. 67*, 39–67.

(67) McKillop, D. F., and Geeves, M. A. (1993) Regulation of the interaction between actin and myosin subfragment 1: evidence for three states of the thin filament. *Biophys. J. 65*, 693–701.

(68) Maytum, R., Lehrer, S. S., and Geeves, M. A. (1999) Cooperativity and switching within the three-state model of muscle regulation. *Biochemistry 38*, 1102–10.

(69) Hoffman, R. M. B., Blumenschein, T. M. A., and Sykes, B. D. (2006) An interplay between protein disorder and structure confers the Ca^{2+} regulation of striated muscle. *J. Mol. Biol. 361*, 625–33.

(70) Zhou, Z., Li, K.-L., Rieck, D., Ouyang, Y., Chandra, M., and Dong, W.-J. (2012) Structural dynamics of C-domain of cardiac troponin I protein in reconstituted thin filament. *J. Biol. Chem. 287*, 7661–74.

(71) Lehrer, S. S. (2011) The 3-state model of muscle regulation revisited: is a fourth state involved? *J. Muscle Res. Cell Motil. 32*, 203–8.

(72) Xu, C., Craig, R., Tobacman, L., Horowitz, R., and Lehman, W. (1999) Tropomyosin positions in regulated thin filaments revealed by cryoelectron microscopy. *Biophys. J. 77*, 985–92.

(73) Lindert, S., Kekenes-Huskey, P. M., Huber, G., Pierce, L., and McCammon, J. A. (2012) Dynamics and calcium association to the N-terminal regulatory domain of

human cardiac troponin C: a multiscale computational study. *J. Phys. Chem. B 116*, 8449–59.

(74) Ovaska, M., and Taskinen, J. (1991) A model for human cardiac troponin C and for modulation of its Ca^{2+} affinity by drugs. *Proteins 11*, 79–94.

(75) Dong, W., Rosenfeld, S. S., Wang, C. K., Gordon, A. M., and Cheung, H. C. (1996) Kinetic studies of calcium binding to the regulatory site of troponin C from cardiac muscle. *J. Biol. Chem. 271*, 688–94.

(76) Gagné, S. M., Tsuda, S., Li, M. X., Smillie, L. B., and Sykes, B. D. (1995) Structures of the troponin C regulatory domains in the apo and calcium-saturated states. *Nat. Struct. Biol. 2*, 784–9.

(77) Slupsky, C. M., and Sykes, B. D. (1995) NMR solution structure of calcium-saturated skeletal muscle troponin C. *Biochemistry 34*, 15953–64.

(78) Spyracopoulos, L., Li, M. X., Sia, S. K., Gagné, S. M., Chandra, M., Solaro, R. J., and Sykes, B. D. (1997) Calcium-induced structural transition in the regulatory domain of human cardiac troponin C. *Biochemistry 36*, 12138–46.

(79) Dong, W. J., Xing, J., Villain, M., Hellinger, M., Robinson, J. M., Chandra, M., Solaro, R. J., Umeda, P. K., and Cheung, H. C. (1999) Conformation of the regulatory domain of cardiac muscle troponin C in its complex with cardiac troponin I. *J. Biol. Chem. 274*, 31382–90.

(80) Vibert, P., Craig, R., and Lehman, W. (1997) Steric-model for activation of muscle thin filaments. *J. Mol. Biol. 266*, 8–14.

(81) Haselgrove, J. C., and Huxley, H. E. (1973) X-ray evidence for radial cross-bridge movement and for the sliding filament model in actively contracting skeletal muscle. *J. Mol. Biol. 77*, 549–68.

(82) Parry, D. A., and Squire, J. M. (1973) Structural role of tropomyosin in muscle regulation: analysis of the x-ray diffraction patterns from relaxed and contracting muscles. *J. Mol. Biol. 75*, 33–55.

(83) Poole, K. J. V, Lorenz, M., Evans, G., Rosenbaum, G., Pirani, A., Craig, R., Tobacman, L. S., Lehman, W., and Holmes, K. C. (2006) A comparison of muscle thin filament models obtained from electron microscopy reconstructions and low-angle X-ray fibre diagrams from non-overlap muscle. *J. Struct. Biol. 155*, 273–84.

(84) Robinson, J. M., Dong, W.-J., Xing, J., and Cheung, H. C. (2004) Switching of troponin I: Ca(2+) and myosin-induced activation of heart muscle. *J. Mol. Biol. 340*, 295–305.

(85) Houmeida, A., Heeley, D. H., Belknap, B., and White, H. D. (2010) Mechanism of regulation of native cardiac muscle thin filaments by rigor cardiac myosin-S1 and calcium. *J. Biol. Chem. 285*, 32760–9.

(86) Xing, J., Jayasundar, J. J., Ouyang, Y., and Dong, W.-J. (2009) Förster resonance energy transfer structural kinetic studies of cardiac thin filament deactivation. *J. Biol. Chem. 284*, 16432–41.

(87) Maron, B. J. (2004) Hypertrophic cardiomyopathy: an important global disease. *Am. J. Med. 116*, 63–5.

(88) Maron, B. J., Casey, S. A., Poliac, L. C., Gohman, T. E., Almquist, A. K., and Aeppli, D. M. (1999) Clinical course of hypertrophic cardiomyopathy in a regional United States cohort. *JAMA 281*, 650–5.

(89) Olivotto, I., Maron, M. S., Adabag, A. S., Casey, S. A., Vargiu, D., Link, M. S., Udelson, J. E., Cecchi, F., and Maron, B. J. (2005) Gender-related differences in the clinical presentation and outcome of hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol. 46*, 480–7.

(90) Maron, B. J., Carney, K. P., Lever, H. M., Lewis, J. F., Barac, I., Casey, S. A., and Sherrid, M. V. (2003) Relationship of race to sudden cardiac death in competitive athletes with hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol. 41*, 974–80.

(91) Kitaoka, H., Doi, Y., Casey, S. A., Hitomi, N., Furuno, T., and Maron, B. J. (2003) Comparison of prevalence of apical hypertrophic cardiomyopathy in Japan and the United States. *Am. J. Cardiol. 92*, 1183–6.

(92) Maron, B. J., Gardin, J. M., Flack, J. M., Gidding, S. S., Kurosaki, T. T., and Bild, D. E. (1995) Prevalence of Hypertrophic Cardiomyopathy in a General Population of Young Adults : Echocardiographic Analysis of 4111 Subjects in the CARDIA Study. *Circulation 92*, 785–789.

(93) Maron, B. J., Olivotto, I., Spirito, P., Casey, S. A., Bellone, P., Gohman, T. E., Graham, K. J., Burton, D. A., and Cecchi, F. (2000) Epidemiology of hypertrophic cardiomyopathy-related death: revisited in a large non-referral-based patient population. *Circulation 102*, 858–64.

(94) Maron, B. J., Niimura, H., Casey, S. A., Soper, M. K., Wright, G. B., Seidman, J. G., and Seidman, C. E. (2001) Development of left ventricular hypertrophy in adults in hypertrophic cardiomyopathy caused by cardiac myosin-binding protein C gene mutations. *J. Am. Coll. Cardiol. 38*, 315–21.

(95) Maron, M. S., Olivotto, I., Zenovich, A. G., Link, M. S., Pandian, N. G., Kuvin, J. T., Nistri, S., Cecchi, F., Udelson, J. E., and Maron, B. J. (2006) Hypertrophic cardiomyopathy is predominantly a disease of left ventricular outflow tract obstruction. *Circulation 114*, 2232–9.

(96) Maron, B. J., McKenna, W. J., Danielson, G. K., Kappenberger, L. J., Kuhn, H. J., Seidman, C. E., Shah, P. M., Spencer, W. H., Spirito, P., Ten Cate, F. J., and Wigle, E. D. (2003) American College of Cardiology/European Society of Cardiology clinical expert consensus document on hypertrophic cardiomyopathy. A report of the American College of Cardiology Foundation Task Force on

Clinical Expert Consensus Documents and the European S. *J. Am. Coll. Cardiol. 42*, 1687–713.

(97) Maron, B. J., Gardin, J. M., Flack, J. M., Gidding, S. S., Kurosaki, T. T., and Bild, D. E. (1995) Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary Artery Risk Development in (Young) Adults. *Circulation 92*, 785–9.

(98) Richard, P., Charron, P., Carrier, L., Ledeuil, C., Cheav, T., Pichereau, C., Benaiche, A., Isnard, R., Dubourg, O., Burban, M., Gueffet, J.-P., Millaire, A., Desnos, M., Schwartz, K., Hainque, B., and Komajda, M. (2003) Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation 107*, 2227–32.

(99) Elliott, P. M., Gimeno Blanes, J. R., Mahon, N. G., Poloniecki, J. D., and McKenna, W. J. (2001) Relation between severity of left-ventricular hypertrophy and prognosis in patients with hypertrophic cardiomyopathy. *Lancet 357*, 420–4.

(100) Maron, B. J., and Maron, M. S. (2013) Hypertrophic cardiomyopathy. *Lancet 381*, 242–55.

(101) Morita, H., Rehm, H. L., Menesses, A., McDonough, B., Roberts, A. E., Kucherlapati, R., Towbin, J. A., Seidman, J. G., and Seidman, C. E. (2008) Shared genetic causes of cardiac hypertrophy in children and adults. *N. Engl. J. Med. 358*, 1899–908.

(102) Lipshultz, S. E., Sleeper, L. A., Towbin, J. A., Lowe, A. M., Orav, E. J., Cox, G. F., Lurie, P. R., McCoy, K. L., McDonald, M. A., Messere, J. E., and Colan, S. D. (2003) The incidence of pediatric cardiomyopathy in two regions of the United States. *N. Engl. J. Med. 348*, 1647–55.

(103) Nugent, A. W., Daubeney, P. E. F., Chondros, P., Carlin, J. B., Cheung, M., Wilkinson, L. C., Davis, A. M., Kahler, S. G., Chow, C. W., Wilkinson, J. L., and Weintraub, R. G. (2003) The epidemiology of childhood cardiomyopathy in Australia. *N. Engl. J. Med. 348*, 1639–46.

(104) Charron, P. (2006) Clinical genetics in cardiology. *Heart 92*, 1172–6.

(105) Maron, B. J., Roberts, W. C., McAllister, H. A., Rosing, D. R., and Epstein, S. E. (1980) Sudden death in young athletes. *Circulation 62*, 218–29.

(106) TEARE, D. (1958) Asymmetrical hypertrophy of the heart in young adults. *Br. Heart J. 20*, 1–8.

(107) Willott, R. H., Gomes, A. V, Chang, A. N., Parvatiyar, M. S., Pinto, J. R., and Potter, J. D. (2010) Mutations in Troponin that cause HCM, DCM AND RCM: what can we learn about thin filament function? *J. Mol. Cell. Cardiol. 48*, 882–92.

(108) Zipes, D. P., Camm, A. J., Borggrefe, M., Buxton, A. E., Chaitman, B., Fromer, M., Gregoratos, G., Klein, G., Moss, A. J., Myerburg, R. J., Priori, S. G., Quinones,

M. A., Roden, D. M., Silka, M. J., Tracy, C., Smith, S. C., Jacobs, A. K., Adams, C. D., Antman, E. M., Anderson, J. L., Hunt, S. A., Halperin, J. L., Nishimura, R., Ornato, J. P., Page, R. L., Riegel, B., Blanc, J.-J., Budaj, A., Dean, V., Deckers, J. W., Despres, C., Dickstein, K., Lekakis, J., McGregor, K., Metra, M., Morais, J., Osterspey, A., Tamargo, J. L., and Zamorano, J. L. (2006) ACC/AHA/ESC 2006 guidelines for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: a report of the American College of Cardiology/American Heart Association Task Force and the European Society of Cardiology Com. *J. Am. Coll. Cardiol. 48*, e247–346.

(109) Miller, M. A., Gomes, J. A., and Fuster, V. (2007) Risk stratification of sudden cardiac death in hypertrophic cardiomyopathy. *Nat. Clin. Pract. Cardiovasc. Med. 4*, 667–76.

(110) Maron, M. S., Olivotto, I., Maron, B. J., Prasad, S. K., Cecchi, F., Udelson, J. E., and Camici, P. G. (2009) The case for myocardial ischemia in hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol. 54*, 866–75.

(111) Alcalai, R., Seidman, J. G., and Seidman, C. E. (2008) Genetic basis of hypertrophic cardiomyopathy: from bench to the clinics. *J. Cardiovasc. Electrophysiol. 19*, 104–10.

(112) Ackerman, M. J., VanDriest, S. L., Ommen, S. R., Will, M. L., Nishimura, R. A., Tajik, A. J., and Gersh, B. J. (2002) Prevalence and age-dependence of malignant mutations in the beta-myosin heavy chain and troponin T genes in hypertrophic cardiomyopathy: a comprehensive outpatient perspective. *J. Am. Coll. Cardiol. 39*, 2042–8.

(113) Niimura, H., Bachinski, L. L., Sangwatanaroj, S., Watkins, H., Chudley, A. E., McKenna, W., Kristinsson, A., Roberts, R., Sole, M., Maron, B. J., Seidman, J. G., and Seidman, C. E. (1998) Mutations in the gene for cardiac myosin-binding protein C and late-onset familial hypertrophic cardiomyopathy. *N. Engl. J. Med. 338*, 1248–57.

(114) Van Driest, S. L., Ackerman, M. J., Ommen, S. R., Shakur, R., Will, M. L., Nishimura, R. A., Tajik, A. J., and Gersh, B. J. (2002) Prevalence and severity of “benign” mutations in the beta-myosin heavy chain, cardiac troponin T, and alpha-tropomyosin genes in hypertrophic cardiomyopathy. *Circulation 106*, 3085–90.

(115) Landstrom, A. P., and Ackerman, M. J. (2010) Mutation type is not clinically useful in predicting prognosis in hypertrophic cardiomyopathy. *Circulation 122*, 2441–9; discussion 2450.

(116) Bos, J. M., Towbin, J. A., and Ackerman, M. J. (2009) Diagnostic, prognostic, and therapeutic implications of genetic testing for hypertrophic cardiomyopathy. *J. Am. Coll. Cardiol. 54*, 201–11.

(117) Geisterfer-Lowrance, A. A., Kass, S., Tanigawa, G., Vosberg, H. P., McKenna, W., Seidman, C. E., and Seidman, J. G. (1990) A molecular basis for familial

hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. *Cell 62*, 999–1006.

(118) Poetter, K., Jiang, H., Hassanzadeh, S., Master, S. R., Chang, A., Dalakas, M. C., Rayment, I., Sellers, J. R., Fananapazir, L., and Epstein, N. D. (1996) Mutations in either the essential or regulatory light chains of myosin are associated with a rare myopathy in human heart and skeletal muscle. *Nat. Genet. 13*, 63–9.

(119) Watkins, H., Conner, D., Thierfelder, L., Jarcho, J. A., MacRae, C., McKenna, W. J., Maron, B. J., Seidman, J. G., and Seidman, C. E. (1995) Mutations in the cardiac myosin binding protein-C gene on chromosome 11 cause familial hypertrophic cardiomyopathy. *Nat. Genet. 11*, 434–7.

(120) Bonne, G., Carrier, L., Bercovici, J., Cruaud, C., Richard, P., Hainque, B., Gautel, M., Labeit, S., James, M., Beckmann, J., Weissenbach, J., Vosberg, H. P., Fiszman, M., Komajda, M., and Schwartz, K. (1995) Cardiac myosin binding protein-C gene splice acceptor site mutation is associated with familial hypertrophic cardiomyopathy. *Nat. Genet. 11*, 438–40.

(121) Satoh, M., Takahashi, M., Sakamoto, T., Hiroe, M., Marumo, F., and Kimura, A. (1999) Structural analysis of the titin gene in hypertrophic cardiomyopathy: identification of a novel disease gene. *Biochem. Biophys. Res. Commun. 262*, 411–7.

(122) Mogensen, J., Klausen, I. C., Pedersen, A. K., Egeblad, H., Bross, P., Kruse, T. A., Gregersen, N., Hansen, P. S., Baandrup, U., and Borglum, A. D. (1999) Alpha-cardiac actin is a novel disease gene in familial hypertrophic cardiomyopathy. *J. Clin. Invest. 103*, R39–43.

(123) Thierfelder, L., Watkins, H., MacRae, C., Lamas, R., McKenna, W., Vosberg, H.-P., Seldman, J. G., and Seidman, C. E. (1994) α-tropomyosin and cardiac troponin T mutations cause familial hypertrophic cardiomyopathy: A disease of the sarcomere. *Cell 77*, 701–712.

(124) Nakajima-Taniguchi, C., Matsui, H., Fujio, Y., Nagata, S., Kishimoto, T., and Yamauchi-Takihara, K. (1997) Novel missense mutation in cardiac troponin T gene found in Japanese patient with hypertrophic cardiomyopathy. *J. Mol. Cell. Cardiol. 29*, 839–43.

(125) Ho, C. Y., Lever, H. M., DeSanctis, R., Farver, C. F., Seidman, J. G., and Seidman, C. E. (2000) Homozygous mutation in cardiac troponin T: implications for hypertrophic cardiomyopathy. *Circulation 102*, 1950–5.

(126) Moolman, J. C., Corfield, V. A., Posen, B., Ngumbela, K., Seidman, C., Brink, P. A., and Watkins, H. (1997) Sudden death due to troponin T mutations. *J. Am. Coll. Cardiol. 29*, 549–55.

(127) Landstrom, A. P., Parvatiyar, M. S., Pinto, J. R., Marquardt, M. L., Bos, J. M., Tester, D. J., Ommen, S. R., Potter, J. D., and Ackerman, M. J. (2008) Molecular and functional characterization of novel hypertrophic cardiomyopathy

susceptibility mutations in TNNC1-encoded troponin C. *J. Mol. Cell. Cardiol. 45*, 281–8.

(128) Pinto, J. R., Parvatiyar, M. S., Jones, M. a, Liang, J., Ackerman, M. J., and Potter, J. D. (2009) A functional and structural study of troponin C mutations related to hypertrophic cardiomyopathy. *J. Biol. Chem. 284*, 19090–100.

(129) Herzberg, O., and James, M. N. Structure of the calcium regulatory muscle protein troponin-C at 2.8 A resolution. *Nature 313*, 653–9.

(130) Satyshur, K. A., Rao, S. T., Pyzalska, D., Drendel, W., Greaser, M., and Sundaralingam, M. (1988) Refined structure of chicken skeletal muscle troponin C in the two-calcium state at 2-A resolution. *J. Biol. Chem. 263*, 1628–47.

(131) Strynadka, N. C., Cherney, M., Sielecki, A. R., Li, M. X., Smillie, L. B., and James, M. N. (1997) Structural details of a calcium-induced molecular switch: X-ray crystallographic analysis of the calcium-saturated N-terminal domain of troponin C at 1.75 A resolution. *J. Mol. Biol. 273*, 238–55.

(132) Chandra, M., da Silva, E. F., Sorenson, M. M., Ferro, J. A., Pearlstone, J. R., Nash, B. E., Borgford, T., Kay, C. M., and Smillie, L. B. (1994) The effects of N helix deletion and mutant F29W on the Ca^{2+} binding and functional properties of chicken skeletal muscle troponin. *J. Biol. Chem. 269*, 14988–94.

(133) Smith, L., Greenfield, N. J., and Hitchcock-DeGregori, S. E. (1999) Mutations in the N- and D-helices of the N-domain of troponin C affect the C-domain and regulatory function. *Biophys. J. 76*, 400–8.

(134) Cordina, N. M., Liew, C. K., Gell, D. A., Fajer, P. G., Mackay, J. P., and Brown, L. J. (2013) Effects of calcium binding and the hypertrophic cardiomyopathy A8V mutation on the dynamic equilibrium between closed and open conformations of the regulatory N-domain of isolated cardiac troponin C. *Biochemistry 52*, 1950–62.

(135) Li, M. X., Spyracopoulos, L., and Sykes, B. D. (1999) Binding of cardiac troponin-I147-163 induces a structural opening in human cardiac troponin-C. *Biochemistry 38*, 8289–98.

(136) Finley, N. L., and Rosevear, P. R. (2004) Introduction of negative charge mimicking protein kinase C phosphorylation of cardiac troponin I. Effects on cardiac troponin C. *J. Biol. Chem. 279*, 54833–40.

(137) Szczesna, D., Guzman, G., Miller, T., Zhao, J., Farokhi, K., Ellemberger, H., and Potter, J. D. (1996) The role of the four Ca^{2+} binding sites of troponin C in the regulation of skeletal muscle contraction. *J. Biol. Chem. 271*, 8381–6.

(138) Hoffmann, B., Schmidt-Traub, H., Perrot, A., Osterziel, K. J., and Gessner, R. (2001) First mutation in cardiac troponin C, L29Q, in a patient with hypertrophic cardiomyopathy. *Hum. Mutat. 17*, 524.

(139) Gillis, T. E., Liang, B., Chung, F., and Tibbits, G. F. (2005) Increasing mammalian cardiomyocyte contractility with residues identified in trout troponin C. *Physiol. Genomics 22*, 1–7.

(140) Zhang, X. L., Tibbits, G. F., and Paetzel, M. (2013) The structure of cardiac troponin C regulatory domain with bound Cd2+ reveals a closed conformation and unique ion coordination. *Acta Crystallogr. D. Biol. Crystallogr. 69*, 722–34.

(141) Zhang, X. L., Tibbits, G. F., and Paetzel, M. (2013) The structure of cardiac troponin C regulatory domain with bound Cd2+ reveals a closed conformation and unique ion coordination. *Acta Crystallogr. D. Biol. Crystallogr. 69*, 722–34.

(142) Schmidtmann, A., Lindow, C., Villard, S., Heuser, A., Mügge, A., Gessner, R., Granier, C., and Jaquet, K. (2005) Cardiac troponin C-L29Q, related to hypertrophic cardiomyopathy, hinders the transduction of the protein kinase A dependent phosphorylation signal from cardiac troponin I to C. *FEBS J. 272*, 6087–97.

(143) Liang, B., Chung, F., Qu, Y., Pavlov, D., Gillis, T. E., Tikunova, S. B., Davis, J. P., and Tibbits, G. F. (2008) Familial hypertrophic cardiomyopathy-related cardiac troponin C mutation L29Q affects Ca^{2+} binding and myofilament contractility. *Physiol. Genomics 33*, 257–66.

(144) Li, A. Y., Stevens, C. M., Liang, B., Rayani, K., Little, S., Davis, J., and Tibbits, G. F. (2013) Familial Hypertrophic Cardiomyopathy Related Cardiac Troponin C L29Q Mutation Alters Length-Dependent Activation and Functional Effects of Phosphomimetic Troponin I*. *PLoS One* (Kimura, A., Ed.) *8*, e79363.

(145) Gillis, T. E., Marshall, C. R., and Tibbits, G. F. (2007) Functional and evolutionary relationships of troponin C. *Physiol. Genomics 32*, 16–27.

(146) Genicot, S., Rentier-Delrue, F., Edwards, D., VanBeeumen, J., and Gerday, C. (1996) Trypsin and trypsinogen from an Antarctic fish: molecular basis of cold adaptation. *Biochim. Biophys. Acta 1298*, 45–57.

(147) D. van der Spoel, E. Lindahl, B. Hess, and the G. development team. (2013) GROMACS USER MANUAL Version 4.6.3.

(148) Ajay, and Murcko, M. A. (2002) Computational Methods to Predict Binding Free Energy in Ligand-Receptor Complexes.

(149) Hub, J. S., de Groot, B. L., and van der Spoel, D. (2010) g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput. 6*, 3713–3720.

(150) Roux, B. (1995) The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun. 91*, 275–282.

(151) Parvatiyar, M. S., Pinto, J. R., Liang, J., and Potter, J. D. (2010) Predicting cardiomyopathic phenotypes by altering Ca^{2+} affinity of cardiac troponin C. *J. Biol. Chem. 285*, 27785–97.

(152) Tikunova, S. B., and Davis, J. P. (2004) Designing calcium-sensitizing mutations in the regulatory domain of cardiac troponin C. *J. Biol. Chem. 279*, 35341–52.

(153) Kekenes-Huskey, P. M., Lindert, S., and McCammon, J. A. (2012) Molecular basis of calcium-sensitizing and desensitizing mutations of the human cardiac troponin C regulatory domain: a multi-scale simulation study. *PLoS Comput. Biol. 8*, e1002777.

(154) Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics 22*, 195–201.

(155) Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res. 37*, D387–92.

(156) Guex, N., Peitsch, M. C., and Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis 30 Suppl 1*, S162–73.

(157) Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem. 24*, 1999–2012.

(158) Shvab, I., and Sadus, R. J. (2013) Intermolecular potentials and the accurate prediction of the thermodynamic properties of water. *J. Chem. Phys. 139*, 194505.

(159) Dynamics, M. (2013) Molecular Simulation Methods with Gromacs.

(160) Yap, K. L., Ames, J. B., Swindells, M. B., and Ikura, M. (2002) Vector geometry mapping. A method to characterize the conformation of helix-loop-helix calcium-binding proteins. *Methods Mol. Biol. 173*, 317–24.

(161) Lemkul, J. A., and Bevan, D. R. (2010) Assessing the stability of Alzheimer's amyloid protofibrils using molecular dynamics. *J. Phys. Chem. B 114*, 1652–60.

(162) Sia, S. K., Li, M. X., Spyracopoulos, L., Gagné, S. M., Liu, W., Putkey, J. A., and Sykes, B. D. (1997) Structure of cardiac muscle troponin C unexpectedly reveals a closed regulatory domain. *J. Biol. Chem. 272*, 18216–21.

(163) Konhilas, J. P., Irving, T. C., and de Tombe, P. P. (2002) Myofilament calcium sensitivity in skinned rat cardiac trabeculae: role of interfilament spacing. *Circ. Res. 90*, 59–65.

(164) Oleszczuk, M., Robertson, I. M., Li, M. X., and Sykes, B. D. (2010) Solution structure of the regulatory domain of human cardiac troponin C in complex with the switch region of cardiac troponin I and W7: the basis of W7 as an inhibitor of cardiac muscle contraction. *J. Mol. Cell. Cardiol. 48*, 925–33.

(165) Tikunova, S. B., Rall, J. A., and Davis, J. P. (2002) Effect of hydrophobic residue substitutions with glutamine on Ca(2+) binding and exchange with the N-domain of troponin C. *Biochemistry 41*, 6697–705.

(166) Albury, A. N. J., Swindle, N., Swartz, D. R., and Tikunova, S. B. (2012) Effect of hypertrophic cardiomyopathy-linked troponin C mutations on the response of reconstituted thin filaments to calcium upon troponin I phosphorylation. *Biochemistry 51*, 3614–21.

(167) Lim, C. C., Yang, H., Yang, M., Wang, C.-K., Shi, J., Berg, E. A., Pimentel, D. R., Gwathmey, J. K., Hajjar, R. J., Helmes, M., Costello, C. E., Huo, S., and Liao, R. (2008) A novel mutant cardiac troponin C disrupts molecular motions critical for calcium binding affinity and cardiomyocyte contractility. *Biophys. J. 94*, 3577–89.

(168) Lindert, S., Kekenes-Huskey, P. M., and McCammon, J. A. (2012) Long-timescale molecular dynamics simulations elucidate the dynamics and kinetics of exposure of the hydrophobic patch in troponin C. *Biophys. J. 103*, 1784–9.

(169) Swindle, N., and Tikunova, S. B. (2010) Hypertrophic cardiomyopathy-linked mutation D145E drastically alters calcium binding by the C-domain of cardiac troponin C. *Biochemistry 49*, 4813–20.

(170) Skowronsky, R. a, Schroeter, M., Baxley, T., Li, Y., Chalovich, J. M., and Spuches, A. M. (2013) Thermodynamics and molecular dynamics simulations of calcium binding to the regulatory site of human cardiac troponin C: evidence for communication with the structural calcium binding sites. *J. Biol. Inorg. Chem. 18*, 49–58.

(171) Pinto, J. R., Parvatiyar, M. S., Jones, M. a, Liang, J., Ackerman, M. J., and Potter, J. D. (2009) A functional and structural study of troponin C mutations related to hypertrophic cardiomyopathy. *J. Biol. Chem. 284*, 19090–100.

Appendix A. Wild type and mutated sequences of Human cardiac troponin C

The highlighted residues are mutated.

Wild Type NNC1_HUMAN Troponin C, Homo sapiens slow skeletal and cardiac muscles[57]

```
        10         20         30         40         50         60
MDDIYKAAVE QLTEEQKNEF KAAFDIFVLG AEDGCISTKE LGKVMRMLGQ NPTPEELQEM

        70         80         90        100        110        120
IDEVDEDGSG TVDFDEFLVM MVRCMKDDSK GKSEEELSDL FRMFDKNADG YIDLDELKIM

       130        140        150        160
LQATGETITE DDIEELMKDG DKNNDGRIDY DEFLEFMKGV E
```

A8V

```
        10         20         30         40         50         60
MDDIYKAVVE QLTEEQKNEF KAAFDIFVLG AEDGCISTKE LGKVMRMLGQ NPTPEELQEM

        70         80         90        100        110        120
IDEVDEDGSG TVDFDEFLVM MVRCMKDDSK GKSEEELSDL FRMFDKNADG YIDLDELKIM

       130        140        150        160
LQATGETITE DDIEELMKDG DKNNDGRIDY DEFLEFMKGV E
```

L29Q

```
        10         20         30         40         50         60
MDDIYKAAVE QLTEEQKNEF KAAFDIFVQG AEDGCISTKE LGKVMRMLGQ NPTPEELQEM

        70         80         90        100        110        120
IDEVDEDGSG TVDFDEFLVM MVRCMKDDSK GKSEEELSDL FRMFDKNADG YIDLDELKIM

       130        140        150        160
LQATGETITE DDIEELMKDG DKNNDGRIDY DEFLEFMKGV E
```

C84Y

```
        10         20         30         40         50         60
MDDIYKAAVE QLTEEQKNEF KAAFDIFVLG AEDGCISTKE LGKVMRMLGQ NPTPEELQEM

        70         80         90        100        110        120
IDEVDEDGSG TVDFDEFLVM MVRYMKDDSK GKSEEELSDL FRMFDKNADG YIDLDELKIM

       130        140        150        160
LQATGETITE DDIEELMKDG DKNNDGRIDY DEFLEFMKGV E
```

E134D

```
        10         20         30         40         50         60
MDDIYKAAVE QLTEEQKNEF KAAFDIFVLG AEDGCISTKE LGKVMRMLGQ NPTPEELQEM

        70         80         90        100        110        120
IDEVDEDGSG TVDFDEFLVM MVRCMKDDSK GKSEEELSDL FRMFDKNADG YIDLDELKIM

       130        140        150        160
LQATGETITE DDIDELMKDG DKNNDGRIDY DEFLEFMKGV E
```

D145E

```
        10         20         30         40         50         60
MDDIYKAAVE QLTEEQKNEF KAAFDIFVLG AEDGCISTKE LGKVMRMLGQ NPTPEELQEM

        70         80         90        100        110        120
IDEVDEDGSG TVDFDEFLVM MVRCMKDDSK GKSEEELSDL FRMFDKNADG YIDLDELKIM

       130        140        150        160
LQATGETITE DDIEELMKDG DKNNEGRIDY DEFLEFMKGV E
```

Appendix B. Equilibrium simulation

1. **pdb2gmx -ignh -f x.pdb -p x.top -o x.gro -water spce**

   Description:

   pdb2gmx: reads a .pdb file and convert it into .gro

   -ignh: Ignore hydrogen atoms that are in the coordinate file

   -f: specifies input structure file

   x.pdb: input structure file

   -p: specifies output topology file

   x.top: output topology file

   -o: output structure file in Gromacs format

   x.**gro**: structure file in Gromacs format

   -water: specifies water model used as a solution

   spce: water model used

2. **editconf -f x.gro -o minimized_box.pdb -d 0.9 -bt octahedron**

   Description:

   editconf: converts generic structure format to .pdb with a defined unit cell

   -f: specifies input structure file

   x.gro: input structure file

   -o: specifies output structure file

   minimized_box.pdb: output structure pdb file with the box around it

   -d: specify distance between the solute and the box

   0.9: distance between the solute and the box

   -bt: specify type of box

   octahedron: type of box

3. **genbox -cp minimized_box.pdb -cs spc216.gro -o min_water.gro -p x.top**

   Description:

   genbox: solvate a solute configuration

   -cp: specifies input structural file

   minimized_box.pdb: input structural file

   -cs: specify input structure of pre-equilibrated water

   spc216.gro: small pre-equilibrated system of water coordinates

   -o: specify output structure file

min_water.gro: output structural file

-p: specify the topology file

x.top: topology file contains information about molecule types and the number of molecules, the preprocessor copies each molecule as needed.

4. **grompp -f minim2.mdp -c min_water.gro –p x.top -o min_sol.tpr**

Description:

grompp: the gromacs preprocessor reads a molecular topology file, checks the validity of the file, expands the topology from a molecular description to an atomic description.

-f : specifies grompp input file with MD parameters

minim2.mdp: grompp input file with MD parameters (Appendix C.)

-c: specifies input structure file

min_water.gro: input structure file

-p: specifies topology files

x.top: topology file

-o: specifies system topology, parameters, coordinates and velocities file

min-sol.tpr: system topology, parameters, coordinates and velocities file

5. **genion -s min_sol.tpr -o with_ions.pdb -nname Cl- -nn (number of ions) -pname Na+ -np (number of ions) -g ions.log -p x.top**

Description:

genion: randomly replaces solvent molecules with monoatomic ions

-s: specifies input system topology, parameters, coordinates and velocities file

min_sol.tpr: input system topology, parameters, coordinates and velocities file

-o: specifies output structure file

with_ions.pdb: output structure file

-nname: specifies name of negative ion

-nn: specifies number of negative ion

Cl-: chloride ion

-pname: specifies name of positive ion

-np: specifies number of positive ion

Na+: sodium ion

-g: specify a log file

Ions.log: log file for added ion

-p: update topology file automatically

x.top: topology file

6. **grompp -f minim3.mdp -c minim_ions.pdb -p x.top -o minimize_cg.tpr**

   Description:

   grompp: The gromacs preprocessor reads a molecular topology file, checks the validity of the file, expands the topology from a molecular description to an atomic description.

   -f : specifies grompp input file with MD parameters

   minim3.mdp: grompp input file with MD parameters (Appendix E.)

   -c: specifies input structure file

   minm_ion.gro: input structure file

   -p: specifies topology files

   x.top: topology file

   -o: specifies MD minimization run input file

   minimize_cg.tpr: MD minimization run input file

7. **mdrun_d -s minimize_cg.tpr -o minim_cg_traj.trr -c minim_cg.pdb -e minim_icg_ener.edr**

   Description:

   mdrun_d: The mdrun program is the main computational chemistry engine within GROMACS. Here it is used for energy minimization.

   -s: specifies the MD minimization run input file

   minimize_cg.tpr: MD minimization run input file

   -o: specifies output full precision trajectory file

   minim_cg_traj.trr: output full precision trajectory file

   -c: specifies output structural file

   minim_cg.pdb: output structural file

   -e: specifies the energy file

   minim_icg_ener.edr: energy file

8. **grompp -f pr.mdp -c minim_cg.pdb -p x.top -o pr.tpr**

   Description:

grompp: The gromacs preprocessor reads a molecular topology file, checks the validity of the file, expands the topology from a molecular description to an atomic description.

-f : specifies grompp input file with MD parameters

pr.mdp: grompp input file with MD parameters (Appendix F.)

-c: specifies input structure file

minim_cg.pdb: input structure file

-p: specifies topology files

x.top: topology file

-o: specifies MD minimization run input file

pr.tpr: MD minimization run input file

9. **mdrun_d -s pr.tpr -o pr_traj.trr -x pr_trax.xtc -c pr.1.gro -e md_ener.edr**

   Description:

   mdrun_d: The mdrun program is the main computational chemistry engine within GROMACS. Here it is used for energy minimization.

   -s: specifies the MD minimization run input file

   pr.tpr: MD minimization run input file

   -o: specifies output full precision trajectory file

   pr_traj.trr: output full precision trajectory file

   -x: specify output compressed trajectory

   pr_trax.xtc: output compressed trajectory

   -c: specifies output structural file

   pr.1.pdb: output structural file

   -e: specifies the energy file

   md_ener.edr: energy file

10. **grompp -f fullmd_sol.mdp -c pr.1.gro -p x.top -o fullmd.tpr**

    Description:

    grompp: The gromacs preprocessor reads a molecular topology file, checks the validity of the file, expands the topology from a molecular description to an atomic description.

    -f : specifies grompp input file with MD parameters

    Fullmd_sol.mdp: grompp input file with MD parameters (Appendix G.)

    -c: specifies input structure file

    Pr.1.gro: input structure file

    -p: specifies topology files

x.top: topology file

-o: specifies MD production run input file

fullmd.tpr: MD production run input file

11. mdrun_d -s fullmd.tpr -o full.1.trr -x full.1.xtc -c full.1.gro -e full.1.edr

Description:

mdrun_d: The mdrun program is the main computational chemistry engine within GROMACS. Here it is used for production run.

-s: specifies the MD production run input file

fullmd.tpr: MD production run input file

-o: specifies output full precision trajectory file

full.1.trr: output full precision trajectory file

-x: specify output compressed trajectory

full.1.xtc: output compressed trajectory

-c: specifies output structural file

full.1.gro: output structural file

-e: specifies the energy file

full.1.edr: energy file

12. grompp -f minim3.mdp -c full.1.gro -p x.top -o minim_final.1.tpr

Description:

grompp: The gromacs preprocessor reads a molecular topology file, checks the validity of the file, expands the topology from a molecular description to an atomic description.

-f : specifies grompp input file with MD parameters

minim3.mdp: grompp input file with MD parameters (Appendix E.)

-c: specifies input structure file

full.1.gro: input structure file

-p: specifies topology files

x.top: topology file

-o: specifies MD minimization run input file

minim_final.1.tpr: MD minimization run input file

13. mdrun_d -s minim_final.1.tpr –o minim_final.trr -c minim_final.pdb -e minim_final.edr

Description:

mdrun_d: The mdrun program is the main computational chemistry engine within GROMACS. Here it is used for minimization run.

-s: specifies the MD minimization run input file

minim_final.1.tpr: MD minimization run input file

-o: specifies output full precision trajectory file

minim_final.trr: output full precision trajectory file

-c: specifies output structural file

minim_final.pdb: output structural file

-e: specifies the energy file

minim_final.edr: energy file

14. trjconv -f full.1.trr -o whole.trr -pbc whole -s full.1.tpr -skip 10

Description:

trjconv: reads the trajectory file and create a new one and make broken molecules whole.

-f: specifies input trajectory file

full.1.trr: input trajectory file

-o: specifies output trajectory file

whole.trr: output trajectory file

-pbc: sets the type of periodic boundary condition treatment

whole: makes broken molecules whole

-s: specifies input structure and mass file

full.1.tpr: input structure and mass file

-skip: specifies how often record trajectory

10: write every 10 frame

15. trjconv -f whole.trr -o whole-nojump.trr -pbc nojump -s full.1.tpr

Description:

trjconv: reads the trajectory file and create a new one and checks if atoms jump across the box and then puts them back

-f: specifies input trajectory file

whole.trr: input trajectory file

-o: specifies output trajectory file

whole-nojump.trr: output trajectory file

-pbc: sets the type of periodic boundary condition treatment

nojump: checks if atoms jump across the box and then puts them back

s: specifies input structure and mass file

full.1.tpr: input structure and mass file

Appendix C. Parameter file detail for energy minimization before adding ions (Equilibrium simulation)

minim2.mdp		Description
define	= -DFLEXIBLE	Will tell ***grompp*** to include flexible water instead of rigid water into your topology.
constraints	= none	No constraints except for those defined explicitly in the topology
Integrator	= steep	A steepest descent algorithm for energy minimization
dt	= 0.002	Time step for integration (picosecond)
nsteps	= -1	Maximum number of steps to integrate or minimize, -1 is no maximum
nstlist	= 10	Frequency to update the neighbor list means it is updated at least every 10 steps
ns-type	= grid	Make a grid in the box and only check atoms in neighboring grid cells when constructing a new neighbor list every nstlist steps. In large systems grid search is much faster than simple search
rlist	= 1.0	Cut-off distance for the short-range neighbor list (nm)
coulombtype	= PME	Electrostatic coulomb type is Fast smooth Particle-Mesh Ewald (SPME) electrostatics
rcoulomb	= 1.0	Distance for the Coulomb cut-off (nm)
vdwtype	= cut-off	Twin range cut-offs with neighbor list cut-off **rlist** and VdW cut-off **rvdw**, where **rvdw ≥ rlist.**
fourierspacing	= 0.12	For ordinary Ewald, the ratio of the box dimensions and the spacing determines a lower bound for the number of wave vectors to use in each (signed) direction

minim2.mdp		Description
fourier_nx fourier_ny fourier_nz	= 0 = 0 = 0	Highest magnitude of wave vectors in reciprocal space when using Ewald.Grid size when using PME or P3M. These values override fourierspacing per direction
pme_order	= 4	Interpolation order for PME. 4 equals cubic interpolation.
ewald_rtol	= 1e-5	The relative strength of the Ewald-shifted direct potential at rcoulomb is given by ewald-rtol
optimize_fft	= yes	Calculate the optimal FFT (Fast Fourier Transform) plan for the grid at startup. This saves a few percent for long simulations, but takes a couple of minutes at start
emtol	= 1000.0	The minimization is converged when the maximum force is smaller than this value
emstep	= 0.01	Initial step-size for energy minimization

Appendix D. Number of ions added to the systems

Troponin C structure	Na^{1+}	Cl^{1-}
Wild Type	24	1
A8V	24	1
L29Q	24	1
C84Y	24	1
E134D	24	1
D145E	26	1

Appendix E. Parameter file detail for energy minimization after adding ions

minim3.mdp		Description
define	= -DFLEXIBLE	Will tell ***grompp*** to include flexible water instead of rigid water into your topology.
constraints	= none	Constraints except for those defined explicitly in the topology
Integrator	= cg	A conjugate gradient algorithm for energy minimization.
dt	= 0.002	Time step for integration (picosecond)
nsteps	= 20000	The maximum number of steps to integrate or minimize.
nstlist	= 10	Frequency to update the neighbor list means it is updated at least every 10 steps
ns-type	= grid	Make a grid in the box and only check atoms in neighboring grid cells when constructing a new neighbor list every nstlist steps. In large systems grid search is much faster than simple search
rlist	= 1.0	Cut-off distance for the short-range neighbor list (nm)
coulombtype	= PME	Electrostatic coulomb type is Fast smooth Particle-Mesh Ewald (SPME) electrostatics
rcoulomb	= 1.0	Distance for the Coulomb cut-off (nm)
vdwtype	= cut-off	Twin range cut-offs with neighbor list cut-off **rlist** and VdW cut-off **rvdw**, where **rvdw** ≥ **rlist.**
fourierspacing	= 0.12	For ordinary Ewald, the ratio of the box dimensions and the spacing determines a lower bound for the number of wave vectors to use in each (signed) direction

minim3.mdp	Description
fourier_nx = 0 fourier_ny = 0 fourier_nz = 0	Highest magnitude of wave vectors in reciprocal space when using Ewald.Grid size when using PME or P3M. These values override fourierspacing per direction
pme_order = 4	Interpolation order for PME. 4 equals cubic interpolation.
ewald_rtol = 1e-5	The relative strength of the Ewald-shifted direct potential at rcoulomb is given by ewald-rtol
optimize_fft = yes	Calculate the optimal FFT (Fast Fourier Transform) plan for the grid at startup. This saves a few percent for long simulations, but takes a couple of minutes at start
emtol = 50.0	Minimization is converged when the maximum force is smaller than this value
emstep = 0.01	Initial step-size for energy minimization

Appendix F. Parameter file detail for position restrained equilibration

pr.mdp		Description
define	= -DPOSRES	Will tell grompp to include posre_Protein_chain_A.itp into your topology, used for position restraints
include	= -I../top	Directories to include in the topology format
Integrator	= md	Run control: A leap-frog algorithm for integrating Newton's equations.
dt	= 0.002	Time step for integration (picosecond)
nsteps	= 1000000	Maximum number of steps to integrate or minimize. Total simulation time: 2 ns
nstxout	= 5000	Frequency to write coordinates to output trajectory file
nstvout	= 5000	Frequency to write velocities to output trajectory file
nstlog	= 5000	Frequency to write energies to log file
nstenergy	= 5000	Frequency to write energies to energy file
nstxtcout	= 25000	Frequency to write coordinates to xtc trajectory
xtc_grps	= protein non-protein	Group(s) to write to xtc trajectory
energygrps	= protein non-protein	Group(s) to write to energy file
nstlist	= 10	Frequency to update the neighbor list (and the long-range forces ;when using twin-range cut-off's)
ns_type	= grid	Make a grid in the box and only check atoms in neighboring grid cells; when constructing a new neighbor list every nstlist steps

pr.mdp	Description
rlist = 0.8	Cut-off distance for the short-range neighbor list
coulombtype = PME	Treatment of electrostatic interactions
rcoulomb = 0.8	Distance for the Coulomb cut-off
fourierspacing = 0.15	For ordinary Ewald, the ratio of the box dimensions and the spacing determines a lower bound for the number of wave vectors to use in each (signed) direction
rvdw = 1.4	Treatment of van der waals interactions
pbc = xyz	Periodic boudary conditions in all the directions
tcoupl = V-rescale	Temperature coupling; temperature coupling using velocity rescaling with a stochastic term
tc-grps = protein non-protein	Temperature coupling; groups to couple separately to temperature bath
tau_t = 0.1 0.1	Temperature coupling; time constant for coupling (one for each group in tc-grps), -1 means no temperature coupling
ref_t = 300 300	Temperature coupling; reference temperature for coupling (one for each group in tc-grps)
Pcoupl = Parrinello-Rahman	Pressure coupling; extended-ensemble pressure coupling where the box vectors are subject to an equation of motion.
Pcoupltype = isotropic	Pressure coupling; isotropic pressure coupling with time constant tau-p (ps)
tau_p = 1.0	Pressure coupling; time constant for coupling (ps)

pr.mdp	Description
compressibility = 4.5e-5	Pressure coupling; the two compressibility [bar-1] values are the compressibility in the x/y and z direction respectively. The value for the z-compressibility should be reasonably accurate since it influences the convergence of the surface-tension, it can also be set to zero to have a box with constant height.
ref_p = 1.0	Pressure coupling; reference pressure for coupling
refcoord_scaling = all	Pressure coupling; the reference coordinates are scaled with the scaling matrix of the pressure coupling.
gen_vel = yes	Pressure coupling; generate velocities in grompp according to a Maxwell distribution at temperature gen-temp [K], with random seed gen-seed. This is only meaningful with integrator md.
gen_temp = 300	Pressure coupling; temperature for Maxwell distribution
gen_seed = 173529	Pressure coupling; used to initialize random generator for random velocities
constraints = all-bonds	Constrain all bonds
nstcomm = 1	Frequency for center of mass motion removal

Appendix G. Parameter file detail for production MD run

pr.mdp		Description
Integrator	= md	Run control: A leap-frog algorithm for integrating Newton's equations.
dt	= 0.002	Time step for integration (ps)
nsteps	= 50000000	Maximum number of steps to integrate or minimize. Total simulation time: 100 ns
nstxout	= 5000	Frequency to write coordinates to output trajectory file
nstvout	= 5000	Frequency to write velocities to output trajectory file
nstlog	= 5000	Frequency to write energies to log file
nstenergy	= 5000	Frequency to write energies to energy file
nstxtcout	= 10000	Frequency to write coordinates to xtc trajectory
xtc_grps	= protein sol	Group(s) to write to xtc trajectory
energygrps	= protein sol	Group(s) to write to energy file
nstlist	= 10	Frequency to update the neighbor list (and the long-range forces ;when using twin-range cut-off's)
ns_type	= grid	Make a grid in the box and only check atoms in neighboring grid cells; when constructing a new neighbor list every nstlist steps
rlist	= 0.9	Cut-off distance for the short-range neighbor list
coulombtype	= PME	Treatment of electrostatic interactions
rcoulomb	= 0.9	Distance for the Coulomb cut-off

pr.mdp	Description
fourierspacing = 0.15	For ordinary Ewald, the ratio of the box dimensions and the spacing determines a lower bound for the number of wave vectors to use in each (signed) direction
pme_order = 4	Interpolation order for PME. 4 equals cubic interpolation
optimize_fft = yes	Calculate the optimal FFT plan for the grid at startup. This saves a few percent for long simulations, but takes a couple of minutes at start.
rvdw = 1.4	Treatment of van der waals interactions
pbc = xyz	Periodic boudary conditions in all the directions
tcoupl = V-rescale	Temperature coupling; temperature coupling using velocity rescaling with a stochastic term
tc-grps = protein non-protein	Temperature coupling; groups to couple separately to temperature bath
tau_t = 0.1 0.1	Temperature coupling; time constant for coupling (one for each group in tc-grps), -1 means no temperature coupling
ref_t = 300 300	Temperature coupling; reference temperature for coupling (one for each group in tc-grps)
Pcoupl = Parrinello-Rahman	Pressure coupling; extended-ensemble pressure coupling where the box vectors are subject to an equation of motion.
Pcouplype = isotropic	Pressure coupling; isotropic pressure coupling with time constant tau-p (ps)
tau_p = 1.0	Pressure coupling; time constant for coupling (ps)

pr.mdp	Description
compressibility = 4.5e-5	Pressure coupling; the two compressibility [bar-1] values are the compressibility in the x/y and z direction respectively. The value for the z-compressibility should be reasonably accurate since it influences the convergence of the surface-tension, it can also be set to zero to have a box with constant height.
ref_p = 1.0	Pressure coupling; reference pressure for coupling
refcoord_scaling = all	Pressure coupling; the reference coordinates are scaled with the scaling matrix of the pressure coupling.
gen_vel = yes	Pressure coupling; generate velocities in grompp according to a Maxwell distribution at temperature gen-temp [K], with random seed gen-seed. This is only meaningful with integrator md.
gen_temp = 300	Pressure coupling; temperature for Maxwell distribution
gen_seed = 173529	Pressure coupling; used to initialize random generator for random velocities
constraints = all-bonds	Constrain all bonds

Appendix H. Measuring RMSD of structures for the last 50 ns using g_rms

1. **make_ndx -f minim_final.pdb -o index.ndx**

   Description:

   make_ndx: will generate special index group. In this case, it is used to create who molecule, N-terminal and C-terminal index group.

   -f: specifies input structure file

   minim_final.pdb: input structure file

   -o: specifies output index file

   index.ndx: output index file

2. **g_rms -s full.1.tpr -f whole-nojump.trr -n index.ndx -o rmsd.xvg -b 50000 -e 100000**

   Description:

   g_rms: compute the RMSD

   -s: specifies the input structure and mass file

   full.1.tpr: input structure and mass file

   -f: specifies input trajectory file

   whole-nojump.trr: input trajectory file

   -n: specifies the input index file

   index.ndx: input index file. Provides the option of measuring RMSD of N or C – terminal domain.

   -o: specifies the output rmsd result

   -b: first frame (ps) to read from trajectory

   -e: last frame (ps) to read from trajectory

Appendix I. Script for measuring chelating distances

```
#!/bin/bash

declare -a RESIDUE=(1CA 65ASP 65ASP 67ASP 67ASP 69SER 71THR 71THR
73ASP 73ASP 76GLU 76GLU)
declare -a ATOM=('CA' 'OD1' 'OD2' 'OD1' 'OD2' 'OG' 'O' 'OG1'
'OD1' 'OD2' 'OE1' 'OE2')
rm index-ca-dist.ndx
touch index-ca-dist.ndx
i=0

for j in {1..11}; do
echo "[ "${RESIDUE[$i]}"-"${ATOM[$i]}" ]" >>index-ca-dist.ndx
cat full.1.gro | grep ${RESIDUE[$i]} | grep
${ATOM[$i]}[[:space:]] | awk '{ print $3 }' >>index-ca-dist.ndx
echo """>>index-ca-dist.ndx
i=$((i + 1))
done

k=1
for l in {1..10}; do
touch temp
echo 0 >>temp
echo "$k">>temp
g_dist -f whole-nojump.trr -s full.1.tpr -n index-ca-dist.ndx -o
Ca-${RESIDUE[$k]}-${ATOM[$k]}.xvg -xvg none<temp
rm temp
k=$((k+1))
done

#collect the distances from each file --> own files. plus one for
the time.
m=1
for filename in *.xvg; do
      # Echo filename to array[0].
      base=$(echo $filename | sed 's/.[^.]*$//')
      echo $base > file-$m.txt
      cat $filename | awk '{print $2}' >>file-$m.txt
      m=$((m + 1))
      done

echo 'Time (ns)' >time.txt

      cat Ca-65ASP-OD1.xvg | awk '{print $1}' >>time.txt
```

Appendix J. measuring the conformational dynamic of HcTnC using g_rmsf

g_rmsf -f whole-nojump.trr -s gro.1.tpr -n index.ndx -o rmsf.xvg -res

Description:

g_rmsf: computes the root mean square fluctuation

-f: specifies input trajectory file

whole-nojump.trr: input trajectory file

-s: specifies input structure and mass file used as a reference

gro.1.tpr: input structure and mass file used as a reference

-n: specifies index file allowing to only compute the RMSF of HcTnC structure

index.ndx: index file allowing to only compute the RMSF of HcTnC structure

-o: specifies output RMSF result

rmsf.xvg: output RMSF result

-res: calculate averages for each residue

Appendix K. Umbrella sampling simulation

1. **pdb2gmx -ignh -ter -f x.pdb -p topol.top -o complex.gro**

   Description:

   Pdb2gmx: reads a .pdb file and convert it into .gro. Generates topology file x.top

   -ignh: ignore hydrogen atoms that are in the coordinate file

   -ter: interactive termini selection, instead of charged (default)

   -f: specifies the input file

   x.pdb: the input structure file from equilibrium simulation

   -p: specifies output topology file

   topol.top: output topology file

   -o: output structure file in Gromacs format

2. **editconf -f complex.gro -o newbox1.gro -d 1**

   Description:

   editconf: defines unit cell

   -f: specifies input structrue file

   complex.gro: input structure file

   -o: specifies output structure file

   newbox1.gro: output structure file

   -d: specify distance between the solute and the box

   1: distance between the solute and the box (nm)

3. **editconf -f newbox1.gro -o newbox.gro -box (copy the distances from new distance and change the Z to 15.0) -center (copy the centers from the new center)**

   Description:

   editconf: modifies the unit cell

   -f: specifies input structure file

   newbox1.gro: input structure file

   -o: specifies output structure file

   newbox.gro: output structure file

   -box: specifies box vector length which is acquired from the box vector provided in the last step (3) and change the Z axis length to 15.0 nm

   -**center:** specifies coordination of geometrical center provided by the last step (3)

4. **genbox -cp newbox.gro -cs spc216.gro -o solv.gro -p topol.top**

   Description:

   genbox: Solvate a solute configuration

   -cp: specifies input structural file

   newbox.gro: input structural file

   -cs: specify input structure of pre-equilibrated water

   spc216.gro: small pre-equilibrated system of water coordinates

   -o: specify output structure file

   solv.gro: output structural file

   -p: specify the topology file

   topol.top: topology file contains information about molecule types and the number of molecules, the preprocessor copies each molecule as needed.

5. **grompp -f ions.mdp -c solv.gro -p topol.top -o ions.tpr**

   Description:

   grompp: The gromacs preprocessor reads a molecular topology file, checks the validity of the file, expands the topology from a molecular description to an atomic description.

   -f : specifies grompp input file with MD parameters

   ions.mdp: grompp input file with MD parameters (Appendix L.)

   -c: specifies input structure file

   solv.gro: input structure file

   -p: update topology files automatically

   topol.top: topology file

   -o: specifies MD minimization run input file

   ions.tpr: MD minimization run input file

6. **genion -s ions.tpr -o solv_ions.gro -p topol.top -pname K -np (number of ions) -nname Cl -nn (number of ions)**

   Description:

   genion: randomly replaces solvent molecules with monoatomic ions

   -s: specifies MD production/minimization run input file

   ions.tpr: MD production/minimization run input file

   -o: specifies output structure file

   solv_ions.gro: output structure file

-p: update topology file automatically

topol.top: topology file

-pname: specifies name of positive ion

K: sodium ion

-np: specifies number of positive ion

-nname: specifies name of negative ion

Cl: chloride ion

-nn: specifies number of negative ion

7. **grompp -f ions.mdp -c solv_ions.gro -p topol.top -o em.tpr**

Description:

grompp: The gromacs preprocessor reads a molecular topology file, checks the validity of the file, expands the topology from a molecular description to an atomic description.

-f: specifies grompp input file with MD parameters

ions.mdp: grompp input file with MD parameters (Appendix L.)

-c: specifies input structure file

solv_ions.gro: input structure file

-p: update topology files automatically

topol.top: topology file

-o: specifies MD minimization run input file

em.tpr: MD minimization run input file

8. **mdrun -v -deffnm em**

Description:

mdrun_d: The mdrun program is the main computational chemistry engine within GROMACS. Here it is used for energy minimization.

-v: be loud and noisy. Reporting the step by step progress.

-deffnm: set the default filename for all file options

em: minimization run input file em.tpr

9. **grompp -f npt.mdp -c em.gro -p topol.top -o npt.tpr**

Description:

grompp: the gromacs preprocessor reads a molecular topology file, checks the validity of the file, expands the topology from a molecular description to an atomic description.

-f: specifies grompp input file with MD parameters

npt.mdp: grompp input file with MD parameters (Appendix M.)

-c: specifies input structure file

em.gro: input structure file

-p: update topology files automatically

topol.top: topology file

-o: specifies MD equilibration run input file

npt.tpr: MD equilibration run input file

10. **mdrun -v -deffnm npt**

Description:

mdrun_d: The mdrun program is the main computational chemistry engine within GROMACS. Here it is used for energy minimization.

-v: be loud and noisy. Reporting the step by step progress.

-deffnm: set the default filename for all file options

npt: minimization/equilibration run input file npt.tpr

11. **dssp -i x.pdb -o x.dssp**

Description:

dssp: program that evaluates the secondary structure of a protein

-i: specifies input structure file

-x.pdb: input structure file

-o: specifies output

x.dssp: output file

12. **tail -160 x.dssp | awk {'print $1" " $4'} | grep H | awk {'print $1'} > helicalresidues.txt**

Description:

This command writes the residue number of all the helical residues into a text file (**helicalresidues.txt**).

13. **make_ndx -f em.gro**

Description:

make_ndx: creates and index file

-f: specifies input structure file

em.gro: input structure file

Note: 3 groups should be created in the index file. First, the Ca^{2+} (**CA162**) which interacts with site II (it is usually the residue 162). Second, Cα of alpha helical residues (**helicalalpha**) of HcTnC mutated or WT. Third, Cα of site II (SiteII) residues 65-73.

14. Duplicate the **posre_Ion2.itp** and rename it to *posre_Ion_2.itp*. Then copy the content below into the file. Atom number 1 is Ca^{2+} that interacts with Site II. It's movement in the X and Y axis is restrained.

```
[ position_restraints ]
; atom  type      fx      fy      fz
    1       1       1000  1000  0
    2       1       1000  1000  1000
    3       1       1000  1000  1000
```

15. In the **topol_Ion2.itp** duplicate the last three lines and change "posre" to "posre_B"

16. **genrestr -f em.gro -n index.ndx -o posre_helix.itp**

    Description:

    genrestr: produces an include file for a topology containing a list of atom numbers and three force constants for the x-, y-, and z-direction. It applies the restrain required for the pull simulation.

    -f: specifies input structure file

    em.gro: input structure file

    -n: specifies input index file

    index.ndx: input index file

    -o: specifies output file include for topology

    posre_helix.itp: output file include for topology

    Note: select "helicalalpha" group when prompts.

17. Add the text shown below to the end of **topol_protein.itp:**

```
#ifdef POSRES_B
#include "posre_Protein.itp"
#endif
```

18. **grompp -f md_pull.mdp -c npt.gro -p topol.top -n index.ndx -o pull.tpr**

    Description:

    grompp: the gromacs preprocessor reads a molecular topology file, checks the validity of the file, expands the topology from a molecular description to an atomic description.

    -f: specifies grompp input file with MD parameters

    md_pull.mdp: grompp input file with MD parameters (Appendix N.)

    -c: specifies input structure file

    npt.gro: input structure file

    -p: update topology files automatically

topol.top: topology file

-n: specifies input index file

index.ndx: input index file containing the site II and CA162

-o: specifies MD pull simulation run input file

pull.tpr: MD pull simulation run input file

19. **mdrun -s pull.tpr -v**

    Description:

    mdrun_d: The mdrun program is the main computational chemistry engine within GROMACS. Here it is used for pulling simulation.

    -s: specifies MD pull simulation run input file

    Pull.tpr: MD pull simulation run input file

    -v: be loud and noisy. Reporting the step by step progress.

20. **trjconv -s pull.tpr -f traj.xtc -o conf.gro -sep**

    Description:

    trjconv: Extracts all the configurations

    -s: specifies input structure and mass file

    pull.tpr: input structure and mass file

    -f: specifies input trajectory file

    traj.xtc: input trajectory file

    -o: specifies output trajectory file

    conf.gro: output trajectory file

    -sep: Write each frame to a separate conf*.gro file

    Note: when prompts select “group 0” to save the whole system

21. Create a text file and name it “groups.txt”. Then find out the group number for Site II and CA162 from the index file created in step 13. Type the group number into the file.
22. Run the Bash script (Appendix O. number 1): **./Bash.sh**
23. Run the Perl script (Appendix O. number 2): **perl perl.pl**
24. Run: **./setupUmbrella.py summary_distances.dat 0.1 run-umbrella.sh &> caught-output-1.txt**

    Description: This will separate the configurations that are 0.1 nm. The detail of the scripts above are shown in Appendix P.

25. Run: **./setupUmbrella.py summary_distances.dat 0.1 &> caught-output.txt**

    Description: This will separate the configurations that are 0.2 nm away**.**

26. Replace all the configuration numbers before 2.0 nm in the caught-output.txt with the configuration number before 2.0 in the caught-output-1.txt.This way you have sampling windows every 0.1 nm for the first 2 nm and for 2-5 nm of pull simulation you will have windows every 0.2 nm.

27. Run the bash script: **./mkdir**

    Note: This script will run a brief NPT equilibration in each window using the npt_umbrella.mdp file shown in Appendix R.

28. Run the bash script: **./mdrun**

    Note: Copies the required file for umbrella sampling simulation into the frames directory. Script detail is shown in Appendix S.

Note: The next two command executed using Westgrid. They need to be executed in the frame directories.

29. **grompp -f md_umbrella.mdp -c npt.gro -t npt.cpt -p topol.top -n index.ndx -o umbrella.tpr**

    Description:

    grompp: the gromacs preprocessor reads a molecular topology file, checks the validity of the file, expands the topology from a molecular description to an atomic description.

    -f: specifies grompp input file with MD parameters

    md_umbrella.mdp: grompp input file with MD parameters (Appendix T.)

    -c: specifies input structure file

    npt.gro: input structure file

    -t: specifies input full precision trajectory

    npt.cpt: input full precision trajectory

    -p: update topology files automatically

    topol.top: topology file

    -n: specifies input index file

    index.ndx: input index file containing the site II and CA162

    -o: specifies MD umbrella sampling simulation run input file

    umbrella.tpr: MD umbrella sampling simulation run input file

30. **mdrun -deffnm umbrella -pf pullf-umbrella.xvg -px pullx-umbrella.xvg**
    Description:

    mdrun: The mdrun program is the main computational chemistry engine within GROMACS. Here it is used for umbrella sampling.

    -dffnm: Set the default filename for all file options.

    Umbrella: default name

    -pf: specifies output umbrella sampling result **xvgr** file.

pullf-umbrella.xvg: Output umbrella sampling result **xvgr** file.

-px: Specifies output umbrella sampling result **xvgr** file.

pullx-umbrella.xvg: Output umbrella sampling result **xvgr** file.

Appendix L. Parameter file detail for energy minimization before/after adding ions (Umbrella sampling)

ions.mdp	Description
integrator = steep	Algorithm (steep = steepest descent minimization)
emtol = 1000.0	Stop minimization when the maximum force < 1000.0 kJ/mol/nm
emstep = 0.01	Energy step size
nsteps = 50000	Maximum number of (minimization) steps to perform
nstlist = 1	Frequency to update the neighbor list and long range forces
ns_type = grid	Method to determine neighbor list (simple, grid)
rlist = 1.4	Cut-off for making neighbor list (short range forces)
coulombtype = PME	Treatment of long range electrostatic interactions
rcoulomb = 1.4	Short-range electrostatic cut-off
rvdw = 1.4	Short-range Van der Waals cut-off
pbc = xyz	Periodic Boundary Conditions

Appendix M. Parameter file detail for equilibration after addition of ions

npt.mdp	Description
define = -DPOSRES	Position restrain the protein
integrator = md	Run parameters; leap-frog integrator
nsteps = 50000	Run parameters; number of steps.
dt = 0.002	Run parameters; time step for integration (picosecond)
nstxout = 1000	Output control; save coordinates every 2 ps
nstvout = 1000	Output control; save velocities every 2 ps
nstenergy = 1000	Output control; save energies every 2 ps
nstlog = 1000	Output control; update log file every 2 ps
continuation = no	Bond parameters; initial simulation
constraint_algorithm = lincs	Bond parameters; holonomic constraints
constraints = all-bonds	Bond parameters; all bonds constrained
lincs_iter = 1	Bond parameters; accuracy of LINCS
lincs_order = 4	Bond parameters; also related to accuracy
ns_type = grid	Neighbor searching; search neighboring grid cels
nstlist = 5	Neighbor searching; 0.01 ps
rlist = 1.4	Neighbor searching; short-range neighbor list cutoff (in nm)
rcoulomb = 1.4	Neighbor searching; short-range electrostatic cutoff (in nm)
rvdw = 1.4	Neighbor searching; short-range van der Waals cutoff (in nm)

npt.mdp	Description
coulombtype = PME	Electrostatics; Particle Mesh Ewald for long-range electrostatics
pme_order = 4	Electrostatics; cubic interpolation
fourierspacing = 0.16	Electrostatics; grid spacing for FFT (Fast Fourier transform)
tcoupl = Berendsen	Temperature coupling is on; Weak coupling for initial equilibration
tc-grps = Protein Non-Protein	Two coupling groups - more accurate
tau_t = 0.1 0.1	Time constant for each group, in ps
ref_t = 310 310	Reference temp, one for each group, in K
pcoupl = Berendsen	Pressure coupling is on; Pressure coupling on in NPT, also weak coupling
pcoupltype = isotropic	Uniform scaling of x-y-z box vectors
tau_p = 2.0	Time constant, in ps
ref_p = 1.0	Reference pressure (in bar)
compressibility = 4.5e-5	Isothermal compressibility, bar^-1
refcoord_scaling = com	Scale the center of mass of the reference coordinates with the scaling matrix of the pressure coupling.
pbc = xyz	Periodic boundary conditions; 3-D PBC
DispCorr = EnerPres	Dispersion correction; account for cut-off vdW scheme
gen_vel = yes	Velocity generation; Velocity generation is on
gen_temp = 310	Temperature for velocity generation
gen_seed = -1	Random seed

npt.mdp	Description
nstcomm = 10 comm-mode = Linear comm-grps = System	Center of mass (COM) motion removal; These options remove COM motion of the system

Appendix N. Parameter file detail for pull simulation

md_pull.mdp	Description
define = -DPOSRES_B	Will tell ***grompp*** to include posre*.itp into your topology, used for position restraints.
integrator = md	Run parameters; A leap-frog algorithm for integrating Newton's equations of motion.
dt = 0.002	Time step for integration (ps)
tinit = 0	Starting time for your run (ps)
nsteps = 2500000	Maximum number of steps to integrate. 5 ns
nstcomm = 10	Frequency for center of mass motion removal. Every 10 steps
nstxout = 5000	Output parameters; Frequency to write coordinates to output trajectory file, the last coordinates are always written. Every 10 ps
nstvout = 5000	Frequency to write velocities to output trajectory, the last velocities are always written
nstfout = 500	Frequency to write forces to output trajectory.
nstxtcout = 500	Frequency to write coordinates to xtc trajectory
nstenergy = 500	Frequency to write energies to energy file
constraint_algorithm = lincs	Bond parameters;
constraints = all-bonds	Convert all bonds to constraints.
continuation = yes	Do not apply constraints to the start configuration and do not reset shells, useful for exact continuation and reruns
nstlist = 5	Single-range cutoff scheme; Frequency to update the neighbor list

md_pull.mdp	Description
ns_type = grid	Make a grid in the box and only check atoms in neighboring grid cells when constructing a new neighbor list every nstlist steps
rlist = 1.4	Cut-off distance for the short-range neighbor list (nm).
rcoulomb = 1.4	Distance for the Coulomb cut-off (nm)
rvdw = 1.4	Distance for the LJ or Buckingham cut-off (nm)
coulombtype = PME	Particle-Mesh Ewald (PME) electrostatics parameters
fourierspacing = 0.12	For ordinary Ewald, the ratio of the box dimensions and the spacing determines a lower bound for the number of wave vectors to use in each (signed) direction.
fourier_nx = 0 fourier_ny = 0 fourier_nz = 0	Highest magnitude of wave vectors in reciprocal space when using Ewald.
pme_order = 4	Interpolation order for PME. 4 equals cubic interpolation.
ewald_rtol = 1e-5	Electrostatic potential; The relative strength of the Ewald-shifted direct potential at rcoulomb is given by ewald-rtol
optimize_fft = yes	Calculate the optimal FFT plan for the grid at startup. This saves a few percent for long simulations, but takes a couple of minutes at start.
Tcoupl = Nose-Hoover	Temperature coupling using a Nose-Hoover extended ensemble.

md_pull.mdp	Description
tc_grps = Protein Non-Protein	Groups to couple separately to temperature bath
tau_t = 0.5 0.5	Time constant for coupling (one for each group in tc-grps) (ps)
ref_t = 310 310	Reference temperature for coupling (one for each group in tc-grps) (K)
Pcoupl = Parrinello-Rahman	Pressure coupling is on; Extended-ensemble pressure coupling where the box vectors are subject to an equation of motion.
pcoupltype = isotropic	Isotropic pressure coupling with time constant tau-p [ps]
tau_p = 1.0	Time constant for coupling (ps)
compressibility = 4.5e-5	Compressibility (NOTE: this is now really in bar-1) For water at 1 atm and 300 K the compressibility is 4.5e-5 (bar^{-1}).
ref_p = 1.0	Reference pressure for coupling (bar)
refcoord_scaling = com	Scale the center of mass of the reference coordinates with the scaling matrix of the pressure coupling.
gen_vel = no	Generate velocities is off
pbc = xyz	Periodic boundary conditions are on in all directions
DispCorr = EnerPres	Long-range dispersion correction; Apply long range dispersion corrections for Energy and Pressure
pull = umbrella	Pull code; Center of mass pulling using an umbrella potential between the reference group and one or more groups.
pull_geometry = distance	Pull along the vector connecting the two groups. Components can be selected with pull-dim

md_pull.mdp	Description
pull_dim = N N Y	The distance components to be used with geometry distance and position, and also sets which components are printed to the output files
pull_start = yes	Add the COM distance of the starting conformation to pull-init
pull_ngroups = 1	The number of pull groups
pull_group0 = SiteII	The name of the reference group
pull_group1 = CA162	The name of the pull group.
pull_rate1 = 0.0025	The rate of change of the reference position. 10 (nm/ps)
pull_k1 = 1000	The force constant. For umbrella pulling this is the harmonic force constant in [kJ mol^{-1} nm^{-2}]. For constant force pulling this is the force constant of the linear potential, and thus minus (!) the constant force in [kJ mol^{-1} nm^{-1}]. Force unit is (kJ mol^{-1} nm^{-2})

Appendix O. Script for measuring distance using (g_dist)

1. Bash Script

```
#!/bin/bash
i=0
for eachFile in conf*.gro; do
   g_dist -s pull.tpr -f conf$i.gro -n index.ndx -o
dist$i.xvg < groups.txt
   i=$((i + 1))
done

      echo $i > filenumber.txt
```

2. Perl script

```
#!/usr/bin/perl -w

use strict;

# loop g_dist command - measure distance in each frame,
write to a file
#for (my $i=0; $i<=500; $i++) {
#  print "Processing configuration $i...\n";
#  system("g_dist -s pull.tpr -f conf${i}.gro -n
index.ndx -o dist${i}.xvg < groups.txt &>/dev/null");
#}

#open filenumber.txt & store the contents (as a number)
in a variable "jmax"
open(FILENUMBER, "filenumber.txt");
   my $jmax = <FILENUMBER>;
   $jmax=$jmax-2;
close(FILENUMBER);


# write output to single file
open(OUT, ">>summary_distances.dat");

for (my $j=1; $j<=$jmax; $j++) {
   open(IN, "<dist${j}.xvg");
   my @array = <IN>;

   my $distance;

   foreach $_ (@array) {
         if ($_ =~ /[#@]/) {
```

```
            # do nothing, it's a comment or formatting
line
        } else {
            my @line = split(" ", $_);
            $distance = $line[1];
        }
   }

   close(IN);
   print OUT "$j\t$distance\n";
}

close(OUT);

# clean up
#print "Cleaning up...\n";

#for (my $k=0; $k<=500; $k++) {
#  unlink "dist${k}.xvg";
#}

exit;
```

Appendix Q. Script for generating umbrella simulation files

1. mkdir

```
#!/bin/bash

cp frames.dat fr
mkdir fr
lambdas=$(`cat frames.dat | wc -l`)
counter=0
for framenumber in `cat frames.dat`; do
     mkdir fr/frame$framenumber

done
for framenumber in `cat frames.dat`;do
     cp conf$framenumber.gro fr/frame$framenumber
done
for framenumber in `cat frames.dat`;do
     cp index.ndx topol.top npt_umbrella.mdp
topol_Protein.itp posre_helix.itp topol_Ion2.itp
posre_Ion_2.itp posre_Ion2.itp posre_Protein.itp
fr/frame$framenumber
done
cp caught-output.txt fr
cd fr
lambdas=$(`cat frames.dat | wc -l`)
counter=0
for framenumber in `cat frames.dat`;do
     cd  frame$framenumber
     if  [ ! -f npt$framenumber.gro ]; then
     grompp -f npt_umbrella.mdp -c
conf$framenumber.gro -p topol.top -n index.ndx -o
npt$framenumber.tpr
     mdrun -deffnm npt$framenumber
     cd ..
     else
     cd ..
fi

        done
```

Appendix P. Script for extracting configurations with specific distance gap

1. **setupUmbrella.py**

```
#!/usr/bin/env python
__description__ = \
"""
Take a file containing frame number vs. distance, then identify
frames that
sample those distances at approximately sample_interval.
Optionally takes an
arbitrary number of template files.  The program searches the
contents of these
files for a search string (by default XXX), replaces the search
string with the
frame number, then writes out each file with a unique, frame-
specific filename.
This last feature means that this script can be used to
automatically generate
input files to run each umbrella sampling simulation.

v. 0.1: fixed bug where the script would choke on distance_files
that went by
        frame intervals besides steps of 1.
"""
__author__ = "Michael J. Harms [harmsm@gmail.com]"
__date__ = "120314"
__usage__ = "setupUmbrella.py distance_file sample_interval
[template_file1 template_file2 ...]"
__version__ = "0.1"

import sys, os, re

def readDistanceFile(distance_file):
    """
    Read a distance file that correlates file number with center-
of-mass
    distance between pulling groups.  Format is:

    0 com0
    1 com1
    2 com2
    ...

    If this data was generated using distances.pl (Justin
Lemkul), it could
    have multiple com calculations in the same file.  This
function starts from
```

```
    the bottom of the file and only takes the last output.
    """

    # Read file
    f = open(distance_file,'r')
    lines = f.readlines()
    f.close()

    # Read the data from the bottom
    out_dict = {}
    for i in range(len(lines)-1,-1,-1):

        # Split on white-space; grab frame/distance
        columns = lines[i].split()
        key = int(columns[0])
        value = float(columns[1])

        # If we've already seen this key, we're done reading the
unique data
        # from the file.
        if key in out_dict:
            break
        else:
           out_dict[key] = value

    # Now put the values into a simple list, sorting to make sure
that they are
    # in the correct order.
    keys = out_dict.keys()
    keys.sort()
    out = [(k,out_dict[k]) for k in keys]

    return out


def sampleDistances(distance_table,sample_interval):
    """
    Go through the distances list and sample frames at
sample_interval.
    Appropriate samples are identified by looking forward through
the
    distances to find the one that is closest to current_distance
+
    sample_interval.
    """

    distances = [d[1] for d in distance_table]

    current_index = 0
    sampled_indexes = [current_index]
```

```
    while current_index < len(distances):

        target_distance = distances[current_index] +
sample_interval

        # Walk through the rest of the distances list and find
the distance
        # that is closest to the target distance
        onward = [abs(target_distance-d) for d in
distances[current_index:]]
        next_index = onward.index(min(onward)) + current_index

        # If we run out of distances to compare, our next_index
will be the
        # current index.  This means we're done.  Otherwise,
record that we
        # found a new index.
        if current_index == next_index:
            break
        else:
            sampled_indexes.append(next_index)
            current_index = next_index

    return sampled_indexes


def
createOutputFile(template_file,frame_number,search_string="XXX"):
    """
    Look for instances of the search string in a template file
and replace with
    the frame number in a new file.
    """

    out_file = "frame-%i_%s" % (frame_number,template_file)

    # Prompt the user before wiping out an existing file
    if os.path.exists(out_file):
        answer = raw_input("%s exists!  Overwrite (y|n)?" %
out_file)
        answer = answer[0].lower()
        if answer != "y":
            return None

    # Read the contents of the template file
    f = open(template_file,'r')
    file_contents = f.read()
    f.close()
```

```
    # Write out the template file contents, replacing all
instances of
    # the search string with the frame number
    f = open(out_file,'w')
    f.write(re.sub(search_string,"%i" %
frame_number,file_contents))
    f.close()


def main(argv=None):
    """
    Parse command line, etc.
    """

    if argv == None:
        argv = sys.argv[1:]

    # Parse required command line arguments
    try:
        distance_file = argv[0]
        sample_interval = float(argv[1])
    except (IndexError,ValueError):
        err = "Incorrect command line arguments!\n\n%s\n\n" %
__usage__
        raise IOError(err)

    # See if a template file has been specified
    try:
        template_files = argv[2:]
    except IndexError:
        template_files = []

    # Figure out which frames to use
    distance_table = readDistanceFile(distance_file)
    sampled_indexes =
sampleDistances(distance_table,sample_interval)

    # If any template files were specified, use them to make
frame-specific
    # output
    if len(template_files) != 0:
        print "Creating frame-specific output for files:"
        print "\n".join(template_files)
        for t in template_files:
            for i in sampled_indexes:
                frame = distance_table[i][0]
                createOutputFile(t,frame,search_string="XXX")

    # Print out summary of the frames we identified
    out = ["%10s%10s%10s\n" % ("frame","dist","d_dist")]
```

```
    for i in range(len(sampled_indexes)):

        frame = distance_table[sampled_indexes[i]][0]
        dist = distance_table[sampled_indexes[i]][1]
        if i == 0:
            delta_dist = "%10s" % "NA"
        else:
            prev_dist = distance_table[sampled_indexes[i-1]][1]
            delta_dist = "%10.3f" % (dist - prev_dist)

        out.append("%10i%10.3f%s\n" % (frame,dist,delta_dist))

    return out


# If called from the command line...
if __name__ == "__main__":
    out = main()
    print "".join(out)
```

Appendix R. Parameter file detail for NPT equilibration in each sampling window

npt_umbrella.mdp	Description
define = -DPOSRES_B	Will tell grompp to include posre*.itp into your topology, used for position restraints.
integrator = md	Run parameters; A leap-frog algorithm for integrating Newton's equations of motion.
dt = 0.002	Time step for integration (ps)
tinit = 0	Starting time for your run (ps)
nsteps = 50000	Maximum number of steps to integrate. 0.1 ns
nstcomm = 10	Frequency for center of mass motion removal. Every 10 steps
nstxout = 5000	Output parameters; Frequency to write coordinates to output trajectory file, the last coordinates are always written. Every 10 ps
nstvout = 5000	Frequency to write velocities to output trajectory, the last velocities are always written
nstfout = 5000	Frequency to write forces to output trajectory.
nstxtcout = 5000	Frequency to write coordinates to xtc trajectory
nstenergy = 5000	Frequency to write energies to energy file
constraint_algorithm = lincs	Bond parameters;
constraints = all-bonds	Convert all bonds to constraints.
continuation = yes	Do not apply constraints to the start configuration and do not reset shells, useful for exact continuation and reruns

npt_umbrella.mdp	Description
nstlist = 5	Single-range cutoff scheme; Frequency to update the neighbor list
ns_type = grid	Make a grid in the box and only check atoms in neighboring grid cells when constructing a new neighbor list every nstlist steps
rlist = 1.4	Cut-off distance for the short-range neighbor list (nm).
rcoulomb = 1.4	Distance for the Coulomb cut-off (nm)
rvdw = 1.4	Distance for the LJ or Buckingham cut-off (nm)
coulombtype = PME	Particle-Mesh Ewald (PME) electrostatics parameters
fourierspacing = 0.12	For ordinary Ewald, the ratio of the box dimensions and the spacing determines a lower bound for the number of wave vectors to use in each (signed) direction.
fourier_nx = 0 fourier_ny = 0 fourier_nz = 0	Highest magnitude of wave vectors in reciprocal space when using Ewald.
pme_order = 4	Interpolation order for PME. 4 equals cubic interpolation.
ewald_rtol = 1e-5	Electrostatic potential; The relative strength of the Ewald-shifted direct potential at rcoulomb is given by ewald-rtol
optimize_fft = yes	Calculate the optimal FFT plan for the grid at startup. This saves a few percent for long simulations, but takes a couple of minutes at start.
Tcoupl = Nose-Hoover	Temperature coupling using a Nose-Hoover extended ensemble.

npt_umbrella.mdp	Description
tc_grps = Protein Non-Protein	Groups to couple separately to temperature bath
tau_t = 0.5 0.5	Time constant for coupling (one for each group in tc-grps) (ps)
ref_t = 310 310	Reference temperature for coupling (one for each group in tc-grps) (K)
Pcoupl = Parrinello-Rahman	Pressure coupling is on; Extended-ensemble pressure coupling where the box vectors are subject to an equation of motion.
pcoupltype = isotropic	Isotropic pressure coupling with time constant tau-p [ps]
tau_p = 1.0	Time constant for coupling (ps)
compressibility = 4.5e-5	Compressibility (NOTE: this is now really in bar-1) For water at 1 atm and 300 K the compressibility is 4.5e-5 (bar-1).
ref_p = 1.0	Reference pressure for coupling (bar)
refcoord_scaling = com	Scale the center of mass of the reference coordinates with the scaling matrix of the pressure coupling.
gen_vel = yes	Generate velocities is on
pbc = xyz	Periodic boundary conditions are on in all directions
DispCorr = EnerPres	Long-range dispersion correction; Apply long range dispersion corrections for Energy and Pressure
pull = umbrella	Pull code; Center of mass pulling using an umbrella potential between the reference group and one or more groups.
pull_geometry = distance	Pull along the vector connecting the two groups. Components can be selected with pull-dim

npt_umbrella.mdp	Description
pull_dim = N N Y	The distance components to be used with geometry distance and position, and also sets which components are printed to the output files
pull_start = yes	Add the COM distance of the starting conformation to pull-init
pull_ngroups = 1	The number of pull groups
pull_group0 = helicalalpha	The name of the reference group
pull_group1 = CA162	The name of the pull group.
pull_init1 = 0	The reference distance at t=0 (nm)
pull_rate1 = 0.0	The rate of change of the reference position. (nm/ps)
pull_k1 = 1000	The force constant. For umbrella pulling this is the harmonic force constant in [kJ mol-1 nm-2]. For constant force pulling this is the force constant of the linear potential, and thus minus (!) the constant force in [kJ mol-1 nm-1]. Force unit is (kJ mol-1 nm-2)
pull_nstxout = 1000	Frequency for writing out the COMs of all the pull group. Every 2 ps
pull_nstfout = 1000	Frequency for writing out the force of all the pulled group. Every 2 ps

Appendix S. Script for preparing frame directories for umbrella sampling simulation

1. **mdrun**

```
#!/bin/bash

cp md_umbrella.mdp Lattice-jobstart.sh /fr
lambdas=$(`cat frames.dat | wc -l`)
counter=0
for framenumber in `cat frames.dat`;do
      cd  frame$framenumber
      mv  npt$framenumber.gro npt.gro
      mv  npt$framenumber.cpt npt.cpt
      mv  conf$framenumber.gro conf.gro
      mv  npt$framenumber.tpr npt.tpr
      cd ..
done

lambdas=$(`cat frames.dat | wc -l`)
counter=0
for framenumber in `cat frames.dat`;do
         cp  md_umbrella.mdp frame$framenumber
      cp  Lattive-jobstart2.sh frame$framenumber
done
```

Appendix T. Parameter file detail for umbrella sample simulation

md_umbrella.mdp	Description
define = -DPOSRES	Will tell ***grompp*** to include posre*.itp into your topology, used for position restraints.
integrator = md	Run parameters; A leap-frog algorithm for integrating Newton's equations of motion.
dt = 0.002	Time step for integration (ps)
tinit = 0	Starting time for your run (ps)
nsteps = 10000000	Maximum number of steps to integrate. 20 ns
nstcomm = 10	Frequency for center of mass motion removal. Every 10 steps
nstxout = 50000	Output parameters; Frequency to write coordinates to output trajectory file, the last coordinates are always written. Every 100 ps
nstvout = 50000	Frequency to write velocities to output trajectory, the last velocities are always written
nstfout = 5000	Frequency to write forces to output trajectory.
nstxtcout = 5000	Frequency to write coordinates to xtc trajectory
nstenergy = 5000	Frequency to write energies to energy file
constraint_algorithm = lincs	Bond parameters;
constraints = all-bonds	Convert all bonds to constraints.
continuation = yes	Do not apply constraints to the start configuration and do not reset shells, useful for exact continuation and reruns

md_umbrella.mdp	Description
nstlist = 5	Single-range cutoff scheme; Frequency to update the neighbor list
ns_type = grid	Make a grid in the box and only check atoms in neighboring grid cells when constructing a new neighbor list every nstlist steps
rlist = 1.4	Cut-off distance for the short-range neighbor list (nm).
rcoulomb = 1.4	Distance for the Coulomb cut-off (nm)
rvdw = 1.4	Distance for the LJ or Buckingham cut-off (nm)
coulombtype = PME	Particle-Mesh Ewald (PME) electrostatics parameters
fourierspacing = 0.12	For ordinary Ewald, the ratio of the box dimensions and the spacing determines a lower bound for the number of wave vectors to use in each (signed) direction.
fourier_nx = 0 fourier_ny = 0 fourier_nz = 0	Highest magnitude of wave vectors in reciprocal space when using Ewald.
pme_order = 4	Interpolation order for PME. 4 equals cubic interpolation.
ewald_rtol = 1e-5	Electrostatic potential; The relative strength of the Ewald-shifted direct potential at rcoulomb is given by ewald-rtol
optimize_fft = yes	Calculate the optimal FFT plan for the grid at startup. This saves a few percent for long simulations, but takes a couple of minutes at start.
Tcoupl = Nose-Hoover	Temperature coupling using a Nose-Hoover extended ensemble.

md_umbrella.mdp	Description
tc_grps = Protein Non-Protein	Groups to couple separately to temperature bath
tau_t = 0.5 0.5	Time constant for coupling (one for each group in tc-grps) (ps)
ref_t = 310 310	Reference temperature for coupling (one for each group in tc-grps) (K)
Pcoupl = Parrinello-Rahman	Pressure coupling is on; Extended-ensemble pressure coupling where the box vectors are subject to an equation of motion.
pcoupltype = isotropic	Isotropic pressure coupling with time constant tau-p [ps]
tau_p = 1.0	Time constant for coupling (ps)
compressibility = 4.5e-5	Compressibility (NOTE: this is now really in bar-1) For water at 1 atm and 300 K the compressibility is 4.5e-5 (bar^{-1}).
ref_p = 1.0	Reference pressure for coupling (bar)
refcoord_scaling = com	Scale the center of mass of the reference coordinates with the scaling matrix of the pressure coupling.
gen_vel = no	Generate velocities is off
pbc = xyz	Periodic boundary conditions are on in all directions
DispCorr = EnerPres	Long-range dispersion correction; Apply long range dispersion corrections for Energy and Pressure
pull = umbrella	Pull code; Center of mass pulling using an umbrella potential between the reference group and one or more groups.
pull_geometry = distance	Pull along the vector connecting the two groups. Components can be selected with pull-dim

md_umbrella.mdp	Description
pull_dim = N N Y	The distance components to be used with geometry distance and position, and also sets which components are printed to the output files
pull_start = yes	Add the COM distance of the starting conformation to pull-init
pull_ngroups = 1	The number of pull groups
pull_group0 = SiteII	The name of the reference group
pull_group1 = CA162	The name of the pull group.
pull_init1 = 0	The reference distance at t=0 (nm)
pull_rate1 = 0.0	The rate of change of the reference position. (nm/ps)
pull_k1 = 1000	The force constant. For umbrella pulling this is the harmonic force constant in [kJ mol^{-1} nm^{-2}]. For constant force pulling this is the force constant of the linear potential, and thus minus (!) the constant force in [kJ mol^{-1} nm^{-1}]. Force unit is (kJ mol^{-1} nm^{-2})
pull_nstxout = 1000	Frequency for writing out the COMs of all the pull group. Every 2 ps
pull_nstfout = 1000	Frequency for writing out the force of all the pulled group. Every 2 ps

Appendix U. Script for running g_wham (Weighted Histogram Analysis Method)

```
#!/bin/bash

mkdir gpullf
cat caught-output.txt | awk '{print $1}' > frames.dat
lambdas=$(`cat frames.dat | wc -l`)
counter=0
for framenumber in `cat frames.dat`;do
     cp frames.dat frame$framenumber
done
for framenumber in `cat frames.dat`;do
     cd frame$framenumber
     mv umbrella.tpr umbrella$framenumber.tpr
     cp umbrella$framenumber.tpr /mnt/Data1/freenrgy-
completed/A8V/gpullf
     cd ..
done
for framenumber in `cat frames.dat`;do
     cd frame$framenumber
     mv pullf-umbrella.xvg pullf$framenumber.xvg
     cp pullf$framenumber.xvg /mnt/Data1/freenrgy-
completed/A8V/gpullf
     cd ..
done
cd gpullf
mkdir output
i=0
for line in `cat ../frames.dat`; do
outputname=`printf "pullf_%04d.xvg" $i`
echo $line
echo $outputname
cp pullf$line.xvg output/$outputname
i=`expr $i + 1`
done
i=0
for line in `cat ../frames.dat`; do
outputname=`printf "umbrella_%04d.tpr" $i`
echo $line
echo $outputname
cp umbrella$line.tpr output/$outputname
i=`expr $i + 1`
done
cd output
ls pull*.xvg >pullf-files.dat
ls umbrella*.tpr >tpr-files.dat
g_wham -it tpr-files.dat -if pullf-files.dat -o -hist -unit
kJ
```