University of Alberta

Process and Troponin

by

Ryan Michael Blau Hoffman 

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

Edmonton, Alberta
Fall, 2008

Your file Votre référence
ISBN: 978-0-494-46334-5
Our file Notre référence
ISBN: 978-0-494-46334-5

"Who am I? What am I?"

—Brian Sykes

To my parents

Abstract

A process-oriented mechanism of troponin is presented. The mechanism differs from previous proposals in two basic ways. First, the kinetic pathways for the activation and inhibition of striated muscle contraction are proposed to be mediated by different regions of the troponin complex. The process of inhibition is not activation in reverse. Second, disorder-to-order transitions (protein folding events) are proposed to have mechanistic significance.

The three tissue isoforms of troponin I were characterized with intrinsic disorder prediction algorithms. This lead to the discovery of isoform-specific variation of the intrinsic disorder in troponin I. The most highly variable region between the isoforms spans residues 57–95 of the fast-skeletal isoform, which is known to form a coiled coil substructure with troponin T. Consideration of the results in light of the structural data motivated the proposal of a novel mechanistic cascade, in which conformational transitions of tropomyosin are communicated to the troponin core domain, with isoform-specific modulation of this communication through the differential stabilization of the coiled coil substructure.

The functional processes of troponin can be modulated through ligand binding. The binding of W7, a drug-like inhibitor of striated muscle contraction, to cardiac troponin C (TnC), is examined through several studies. Binding to the separated domains of TnC is characterized with global analysis, in which all titratable signals are simultaneously fit to the global dissociation constant. This (non-standard) approach facilitates model selection and comprehensive interpretation of the data. I develop a global likelihood analysis that models the error distribution as a Gaussian. I assert (but do not prove) that the fitting of a population of titratable signals improves the robustness of the analysis so that titrations with highly pre-asymptotic endpoints can still be well-determined.

The structure of the calcium-activated regulatory domain of cardiac troponin C (NTnC), bound to W7, is then presented. This unpublished result shows that W7 occupies a location on the surface

of NTnC appropriate to occlude the binding of the target peptide. This provides visually explicit insight into the mechanism of W7 activity in striated muscle.

Acknowledgements

Some think brevity is the goal with an acknowledgments section. I agree it's more dignified than what you're about to read. But in this case *I'm* the one that pays the page charges. I really want to thank everyone that has accompanied me for the ride, or given me a push, or simply pointed and laughed. So here comes the blubbering. (I've sectioned it to make it more skimmable.)

To my parents and my brother

Mom and Dad, you've given me life, love, and love of life. You've supported me in every possible way, as much as you could, unconditionally. Words don't suffice but here they are. Mom, see what has become of the love of reading you instilled? See how happy I am? Dad, see what you started! You're my first science mentor! As for the parenting, I think you did a better job with Charlie than with myself but he *was* a little wimpy... Adam, you've shown me over and over, what it is to succeed without trying, and what it is to continue succeeding after things get really difficult. I love you all so much.

To Brian

Thank you for supervising me, for mentoring me, for letting me contribute to your work, for *really* working on me. I hope I can give enough back. I hope I represent you well, especially your patience and generosity. There were more then a few times when you rescued me from others or from myself.

To Monica

Just ask Brian: we should all be nice like Monica Li! She is a lynchpin member of the laboratory. She coauthored my first papers, and she did all of the experimental work for those papers. Thank you so much, Monica.

To Tharin

Tharin Blumenschein shared her corner of the lab with me. She is my other major collaborator in the Sykes lab, over my PhD studies. With a penetratingly concrete intellect, and a playfully outgoing demeanor, she made the work go more easily.

To Olivier

I can't mangle French on paper, can I? Oh yes I can! Thanks so much OJ, for everything. Managing to live happily with a coworker is no small feat. You always know the right thing to say, especially when its the wrong thing to say.

To Nick Shaw

It happened again! Thanks for being a great roomate Nick. You've been patient with my recent shirking of roomate responsibilities as I freak and fuss over this little book.

To Olga

Olga Baryshnikova really enriched the lab, and although she wasn't constantly chattering (a good thing), I always had the sense that a little songbird occupied the room.

To Ian Robertson

It has been a high pleasure discussing science, art and life with you. Thanks for joining the lab.

To Dave

David Hau was one my earliest grad students with whom I shared lab space, and is an enduring friend. He exemplifies those individuals that have kept me grounded in the clouds.

To Marta

Marta Oleszczuk (shout out to Jacek!) took over Tharin's spot, with me, in the south-west corner of the lab. She's offered poignant input on my most recent paper, spectacular social commentary, coffee consumption camaraderie, besides cracking me up on a regular basis.

To Robert

Robert Boyko taught me so much without trying. His soft-spoken demeanor, patience, and willingness to coach (and encourage) my early forays into programming, have directly contributed to my person.

To Pascal

Pascal Mercier came up with all kinds of programming solutions faster than I can generate problems. Thanks, Pascal, for your extensive help and colorful manner. You always went the extra mile.

To my friends in Edmonton

Many thanks to (in rough order-of-appearance) Alan Sukonnic, Ruth Masliah, Stefania Iacchelli, Dustin Ritchie, Jason Lamoreux, Olga Baryshnikova, Quang Tran, Haley Shandro (née Cleary), Michael Shandro, Shiraz Khan, Ting-Wai Lee, Nancy Sykes, Brian Zaharko (yes, you go in this section!), Chris Smolyk, the other crusty old men at La Tienda, the gang at Avenue Pizza, and John Paul Glaves. You people are great! May you all thrive; I know you've helped me to.

Old friends

My friends have always egged me on. Jared Wolk, Misha Solomon, Aaron Spevakow, you people are practically my family, or at least my depraved uncles. Ryan Angelo, Ryan Robertson, Scott Skagford, Rachel Prasow, thanks for getting me through Wise Wood. James (Sacha) Bye, Laura McNally, Raja Roy-Choudhury, Mike Wilson, Tracey Pennimpede, Brent McCurdy, Maddy Marcus, Chris Lackey, Colin Hassey, you all were pillars of hilarity during my undergraduate studies.

Teachers

A number of professors at Queen's coaxed me along and showed me how much fun this can be. Walter Szarek and Martin Petkovich provided expositions into organic chemistry and molecular biology. Steve Smith introduced me to protein NMR, and took the time to discuss my career with me. Rather than filling pails, those people lit fires. In Edmonton, Michael James, Bart Hazes, Ernst Bergmann, Mark Glover, and Ryan McKay all taught me things in a classroom. Drs. James and Glover also formed my Supervisory Committee and have been available for off-the-cuff discussions. Thank you, Mark and Michael, I know that I have not yet sufficiently reciprocated.

Past Supervisors

I thank my first supervisor, Floyd Snyder, who took me on when I was as green as it gets, and wrote me *way* more letters of reference than I would want to! Jason Maynes also trained me directly. I'm so fortunate to have learned the rudiments of labwork from these very exacting, principled, and rigorous scientists. Siew-Sing Li provided a lot of help with liaising with Floyd, as well as a ton of pleasant conversations.

Michael Nesheim helped me to find my way. His patience and encouragement, and the strength of his intellect and wit, are an ongoing source of inspiration. Here's a representative quote: "If people knew how carcinogenic oxygen was you could probably sell them masks to filter it out." Nicole Brufatto was very helpful during my stay in Michael's lab; I did most of my work with her assistance.

Marion and Sue

My first invitation to graduate studies at U. of A. came from Marion Benedict, who has always done as much as possible to make me feel comfortable and secure. Thanks, Marion, for being so warm

and helpful. Susan Smith was an ongoing source of help, an occasional lunch-buddy, and a constant ally. Thanks so much, Sue!

University of Alberta, Chapter I

My story at the University of Alberta unfolds in chapters. When I worked across the hall, several individuals were instrumental in my learning. Leo Spyracopoulos taught me the rudiments of assigning protein NMR spectra. Sean McKenna was a huge source of camaraderie and helped me in my earliest days in Edmonton. Darrin Lindhout taught me many of the practical aspects of protein NMR with a seemingly endless patience. Paul Scott immediate impressed upon me his enthusiasm for science, and was always willing to discuss any aspect of my work or career. Thank you, Paul.

Ron McElhaney

You performed your role as Graduate Coordinator with complete objectivity and what seemed to be a fierce sense of responsibility. When I was writing my Candidacy Proposal, you made yourself available for very productive conversations. Thank you, Ron.

University of Alberta, Chapter II

Angela Thiessen is one of the most beautiful people around. Simply radiant: you get a suntan standing next to her. Xu Wang was the first person to show me how to debug a program. He always had time for helping me and we had good science talks. Grant Gay and Jeffrey DeVries were also fun to talk with. Thanks, Jan Rainey, for the science and career talks, for the beers. Xingang (Gump) Zhao is yet another excellent scientist I've had the honor of associating with. David Corson and Melissa Crane (née Rakovszky) expressed and purified the raw material for much of my work. Dave also helped me to butcher rabbits. Gerry McQuaid diligently maintained our NMR spectrometer until the time of his retirement. Afterwards he diligently maintained an appearance at the Slupsky Invitational.

Helpful Members of the Department

Leslie Grad, Ruthvin Lewis, David Bannock, Kevin Wilson and lab members (especially Boray Nguyen), Gillian McCuaig, Zhongwei Zhao, Nasim Boroumand, Ramasamy Sankaranarayanan, Cyril Kay, Larry Smilie, Matt Schellenberg, Craig Garen, Grace Garen, Howard Young, Catherine Trieber, Gina Thede, and Dean Schieve all provided helpful discussions and one point or another, or assistance with using equipment. Bernard Lemire, as the current Graduate Coordinator, has been very helpful and I sincerely thank him.

Helpful Members of Other Departments

Mark Miskolzie and Albin Otter of the Department of Chemistry provided helpful input, deuterated solvents, and spectrometer access. I have also benefitted from exchanges with David Wishart and Mark Berjanskii. Gerard J. Kleywegt (University of Uppsala) assisted in the use of HIC-Up and

XPLO2D, begrudging helping me to protonate ligands for NMR structure determination, and also took considerable efforts to explain the XPLOR/CNS representation of target covalent parameters.

Key interactions

John Moult (University of Maryland Biotechnology Institute), Keith Dunker (Indiana University School of Medicine), Zsuzsanna Dosztányi (Hungarian Academy of Sciences), and more recently Vladamir Uversky (Indiana University School of Medicine) have all provided helpful discussions towards my study of intrinsic disorder in troponin. The computational profiling of intrinsic disorder was John Moult's proposal.

At UCSD

Peter Wolynes, Diego Ferreiro, Patrick Weinkam, Joachim Lätzer, Joe Hegler, and Michael Prentiss have all been instrumental in my apprehension of energy landscape theory. I wish I could have included our collaborative work here, but this will happen in due course. Tracy Hogan helped me greatly with liaising with Peter, and has been fun to talk with. Patricio Craig was also fun to talk with. Jose Onuchic patiently listened to one of my presentations and offered critical feedback.

Thesis reviewers

A number of people read and commented on my thesis. Brian Sykes, by far, did the most here. Olga, Marta, Ian, and Peter Holmes all provided helpful feedback at times.

External examiners

I sincerely thank Elizabeth Komives (UCSD) and Lukasz Kurgan (UofA) for their time and attention.

Funding

I would like to thank Floyd Snyder and Brian Sykes for their support during the years in which I lacked funding. I thank the Alberta Heritage Foundation for Medical Research for supporting my PhD work since 2006, and a summer studentship in 2002.

Contents

List of Figures

List of Symbols and Abbreviations

Symbols

Å— "Ångstrom" (unit of 10^{-10} meters)
• — "bound to" (used as an infix operator between two chemical species)
$\leftrightarrow$ — chemical equilibrium
[X]— "concentration of (chemical species) X"
Hz — "Hertz" (unit per second)
K — dissociation constant
kcal — kilocalorie
M — moles per liter
MHz — "megaHertz" (10^6 Hz)
μM — micromoles per liter
mM — millimoles per liter
pM — picomoles per liter
χ^2 — squared residuals

Abbreviations

cCTnC — C-terminal domain of TnC, cardiac isoform
cNTnC — N-terminal (regulatory) domain of TnC, cardiac isoform
cNp — N-terminal, cardiac-specific extension of troponin I
cSp — switch region of troponin I, cardiac isoform
cTn — troponin, cardiac isoform
CaM — calmodulin
COSY — correlation spectroscopy
CTnI — C-terminal domain of troponin I
DQF — double quantum filtered
DMSO — dimethyl sulfoxide
H1 — N-terminal helix of TnI, also called Rp40
H2 — The coiled-coil-forming helix of TnI
HCA — hydrophobic cluster analysis
HCA1 — hydrophobic cluster analysis cluster number 1 (number varies)
HMBC — heteronuclear multiple bond correlation
HSQC — heteronuclear single quantum coherence
ID — instrinsic disorder
Ip — inhibitory region of troponin I
L1 — Loop connecting H1 and H2
Md — mobile domain of troponin I

NMR — nuclear magnetic resonance
NOE — nuclear Overhauser effect
NOESY — NOE spectroscopy
NTnC — N-terminal (regulatory) domain of TnC
NTnI — N-terminal domain of TnI
PDB — Research Collaboratory for Structural Bioinformatics Protein Data Bank
ppm — parts per million
RMS — root mean squared
sCTnC — C-terminal domain of TnC, fast-twitch skeletal isoform
sNTnC — N-terminal (regulatory) domain of TnC, fast-twitch skeletal isoform
sTn — troponin, fast-twitch skeletal isoform
Sp — switch region of troponin I
Tm — tropomyosin
Tn — troponin
TnC — troponin C
TnI — troponin I
TnT — troponin T
TnT-T1 — N-terminal domain of troponin T
TnT-T2 — C-terminal domain of troponin T

Chapter 1

Process and Troponin

The research of my mentor falls under the rubric of structural biology. His laboratory endeavors to understand life processes by visualizing, at the level of individual atoms, the shape and topology of protein molecules in their manifold functional states. This introduction focuses on the logic used in the day-to-day proceedings of structural biologists.

1.1 Process, function, and purpose

A process is a spatially and temporally ordered sequence of events. A function is a process with a purpose. The program of structural biology discloses structure–function relationships. I emphasize that a structure–function relationship is the same thing as a structure–purpose relationship. An implication of this is that structural biologists endeavor to know *the purpose of life processes.*

Life, as a natural phenomenon, does not occur for any purpose in the usual sense of the word. I will avoid a debate over the the true purpose of the organism. Instead, I posit that *the purpose of the organism is to sustain its living condition.* This will seem quite incorrect to many of those trained in evolutionary biology but the statement reflects the biochemist's usual mental proclivity. Each instant in time, a living cell sustains its own living condition through drawing upon available resources.

Life processes fulfill the purpose of life. To the extent that an individual life process contributes to the purpose of an organism, this process itself can be assigned a purpose. In this purview, the aquaporins serve the purpose of conducting the flow of water. The contractile proteins serve the purpose of directed movement. The processes sustaining an organism can be assigned purposes by isolating the machinery that mediates them. Importantly, one needs to dissect-out the correct set of components in order to realize their potential to complete a purposeful task. Once a correct group of components is isolated, purposefulness can be determined.

1.2 Structure and process in simple machines

The program of structural biology can be demonstrated through an analogy to simpler scenarios. The following section illustrates a thought experiment: a researcher (call him Fred) is given the

three dimensional coordinates corresponding to a simple machine in all of its functional states, and is asked to explain how the machine works, towards what purpose.

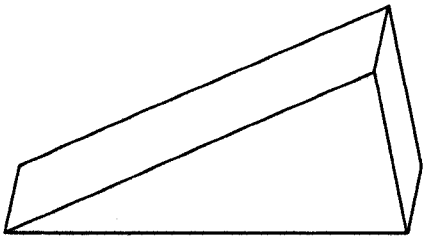


Figure 1.1: What is the function of this machine?

The inclined plane

The inclined plane is one of the simplest machines. But given the structure of an inclined plane (Figure 1.1), Fred hesitates to recognize it as a machine. What orientation does the machine occupy when it functions? Are there other components in the system that also need to be considered? If Fred assumes that the structure of the machine contains all of the components required for purposeful function then he is already at a mental impasse!

Fred cannot divine the function of this machine from only looking at its structure. He begins to imagine scenarios whereby the machine interacts with other objects. He may start by assuming that the orientation of the machine, as shown in Figure 1.1, is arbitrary. Fred then realizes: if the machine is pushed into a softer medium, it will divide the medium. Perhaps the machine is a wedge?

The reader knows the machine is actually an inclined plane and not a wedge. But Fred does not have this information. The wedge and the inclined plane have identical structures but have different processes associated with their functions, different purposes. One can be used for dividing a whole into parts, the other for lifting a mass. Neither function can be excluded or supported based on the structure (Figure 1.1), but Fred frames a hypothesis anyway.

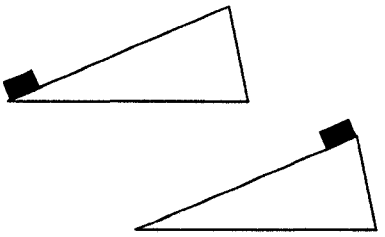


Figure 1.2: Two 'structures' of the inclined plane containing its 'substrate' (the grey mass). The reader can assume these both correspond to functional *states*. Other information needs to be considered in defining the function, namely the *process* that connects the two states. Here, the process is called 'sliding.' With knowledge of what 'sliding' means, can the purpose of the inclined plane be determined?

Suppose the structure of the inclined plane was determined in the presence of some of its functional components, in two of its functional states (as shown in Figure 1.2). Fred looks at these structures and thinks, 'Well, that's a funny way to wedge through a medium — those chunks of media look smaller than the wedge itself! Perhaps it's actually an inclined plane?' Fred is on his way to correctly assigning a structure-function relationship! Are these structures sufficiently informative to allow a structural biologist to describe the process of lifting? Not really — Fred needs to know that the weight is being moved in opposition to the force of gravity. Fred needs to know that the weight can be slid continuously over the surface of the plane, that the frictional coefficient of the plane is sufficiently small. Fred needs to invoke a process to connect the functional states.

Fred's problem is that to call something a machine, one needs to assign to it a function. One potential function of the inclined plane is 'lifting' in which the potential energy of a mass is increased by moving it away from the center of a gravitational field. An inclined plane makes lifting easier but less efficient compared with directly raising the mass. Note the difficulties with arriving at lifting as *the purpose* of the inclined plane. For example, the reverse process — dropping — can also be mediated by an inclined plane. But the process of dropping is hindered by the use of an inclined plane — it makes dropping slower and more controlled. Still, one can imagine scenarios where this would seem purposeful. The context or scenario constrains the possible structure-function relationships; these relationships cannot be meaningfully defined outside of a context.

The knife switch

The electrical knife switch is much closer in purpose to the biochemical machines than the inclined plane (Figure 1.3). The knife switch has only two functional states, called 'off' and 'on,' which regulate an electrical circuit. Knowledge of the switch's structure in the off and on states could seem sufficient to describe its function. In one state, the circuit is completed and functional, in the other state, this path is disrupted so the circuit cannot function. But this level of description fails to capture some very rich features of how the machine functions. Without prior knowledge of how electricity flows, without knowledge that the two states can be interconverted (how to flip the switch), descriptions of the two states will not suffice to describe the function of a simple electrical switch.

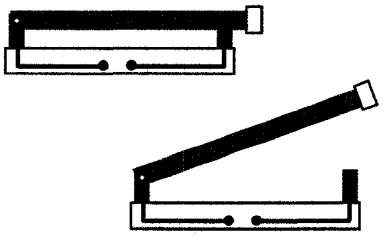


Figure 1.3: Two 'structures' of the electrical knife switch. Darker regions are electrically conductive. The switch has been dissected away from its functional context – an electrical circuit. Can the function of the switch be determined *from* these structures of its states?

The two functional states of a switch are interconverted through a process: 'flipping the switch.' Each functional state modulates a process: 'the flow of electrical current.' In general, knowledge of the functional states of a machine will be insufficient to describe its function without consideration of an underlying process. This would be true of any (real) machine that performs work. All (real) work is associated with a dissipation of energy. The amount of energy dissipated depends on the path taken between the two states, not on the energies of the states themselves.

The examples of the inclined plane, and the knife switch, illustrate that the function of a machine cannot be *directly* disclosed from visualization of the machine in its functional states. What is required is knowledge of the underlying process that the machine undergoes as it functions.

1.3 Troponin as a switch

The central metaphor of troponin research is *troponin as a switch.* Troponin is a molecular switch that regulates striated muscle contraction. Like most biochemical machines, troponin is usually studied in one of its functional states. There are at least two states for troponin, and many research groups worldwide labor to correlate the appearance of the switch in a given state with its function. The above examples show a number of potential problems with this approach. The main problem is the difficulty inherent in isolating the minimal unit of troponin function; this problem will resurface throughout my Thesis.

Figure 1.4: Two 'structures' of large portions of the biochemical switch troponin, corresponding to two of its functional states.[2] The lozenge at the bottom is not actually in the data; it was added so that the function could be more easily visualized. How does troponin work?

Like most biochemical machines, troponin is multipartite and multifunctional. Various observations have lead to (various regions of) troponin being characterized as a switch, rheostat, scaffold, coupler, spring and damper. Any given structural characteristic of troponin could be variously interpreted in terms of these functions. Can a structural biologist correctly correlate the appearance of this machine with its purpose? I think this is possible given knowledge of the types of processes that the machine can undergo; possible but difficult.

Like most biochemical machines, troponin is a protein. Can one understand a proteinaceous

machine through analogies to the macroscopic machines? The answer again hinges on the types of processes that the machine undergoes in the course of function. Many of the types of processes that a biochemical machine undergoes are fundamentally disparate from the simple machines. For example, a knife switch interconverts between functional states through the intervention of a human hand, whereas troponin switches through the statistical-mechanical phenomenon called mass action.

The biochemical machines

Proteins regulate the chemical reactions that sustain life, form structures that coordinate these reactions, and form engines that translocate the structures. In short, proteins are the machines of life. But how apt is this metaphor?

In *What is Life?*, decades before the first observations of protein structure, Erwin Schrödinger compares the machinations of life with a mechanical, macroscopic example — that of clockwork. He explains how a clock at ambient temperatures is effectively at zero absolute temperature because thermal fluctuations do not impact its function. A clock is a purely enthalpically driven machine; both working and stopped states have the same entropy. He argues that the organism also operates mechanistically, and therefore is also enthalpically driven [3, p. 85]:

> " ... few words more are needed to disclose the point of resemblance between a clockwork and an organism. It is simply and solely that the latter also hinges upon a solid — the aperiodic crystal forming the hereditary substance, largely withdrawn from the disorder of heat motion."

Schrödinger correctly anticipates the structure of DNA, and all of the machinations of molecular biology, from this beautiful line of thought. But his arguments are self-limiting in their scope: most of living machinery functions over comparatively narrow temperature ranges compared with a clockwork. The influence of temperature in the function of a machine shows an influence of entropy in the mechanism. An organism's machinations function over a narrow range of temperatures and therefore cannot be accurately modeled as an clock-like assemblage of simple machines.

1.4 Process description of protein structure

My understanding of reality is process-oriented. The physical interactions that stabilize matter into atoms, molecules, and objects are constantly regenerating the entities that we perceive to continuously exist. On an appropriately small spatiotemporal scale everything can be shown to fluctuate. For the simple machines, these fluctuations can have little impact on their function, as Schrödinger argues. But the biochemical machines only manifest their structure-function relationships in the presence of these fluctuations. Protein science shows that *contingently solid entities can result from thermal fluctuations.* These contingently solid entities are ordered (folded) proteins. In this way, Schrödinger's argument re-emerges from its own ashes: the biochemical machines recover a clockwork-like functionality from an underlying disordered, thermally-driven motion. This disor-

dered thermal motion is the process that generates protein structure and function. From the standpoint of protein structure, the process is called *conformational fluctuation on the energy landscape.*

The energy landscape

Proteins are paradigmatic examples of complex systems [4]. This complexity results directly from their being very many possible configurations of the system, with many configurations being similar in energy. This situation is termed *energetic frustration.* The *energy landscape* concept has emerged as a universal descriptor of the configuration space of energetically frustrated polymers such as proteins [4–7]. An energy landscape describes how the energy of a system varies with the configuration of the parts.

Normally, the energy landscape serves as mental device to visualize the stabilization that the protein undergoes as it folds. In this context the landscape is called the *folding funnel.* This stabilization offsets the entropic cost of ordering the protein into a single conformation. Figure 1.5 shows four energy landscapes that, depending on the identity of the protein and the chemical conditions, describe the conformational preferences of proteins. These are qualitative models but they will be useful towards later discussions of the stabilities of regions of troponin. Figure 1.6 shows how the entropic and enthalpic components of the free energy would vary for these four landscapes.

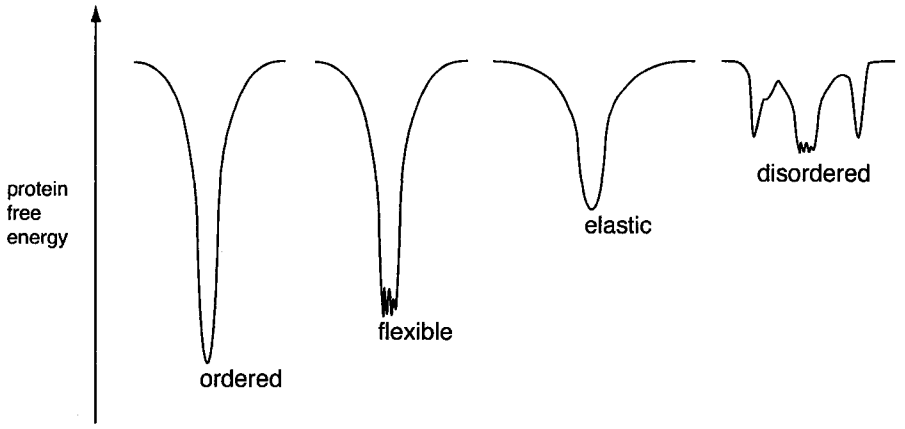


Figure 1.5: Four simplified depictions of energy landscapes. All of these landscapes can apply to proteins. The free energy is depicted on the vertical axis. Within a given landscape, movement in the horizontal direction corresponds to change in conformation. The parabola-like landscape at the left is characteristic of an ordered protein; its landscape is called a folding funnel. A folding funnel is deep compared with the size of a typical conformational fluctuation, in contrast, a disordered landscape (at right) has minima that are readily explored by typical fluctuations.

Manifestations of conformational fluctuation

Figure 1.6 shows four regimes of protein conformational existence, spanned by the protein entropy and enthalpy. The regimes are defined in terms of limiting values of enthalpic stabilization of the

protein and residual entropy in the protein. This model should be considered to apply to regions of proteins; a single protein molecule will usually occupy multiple areas of the ontological space.

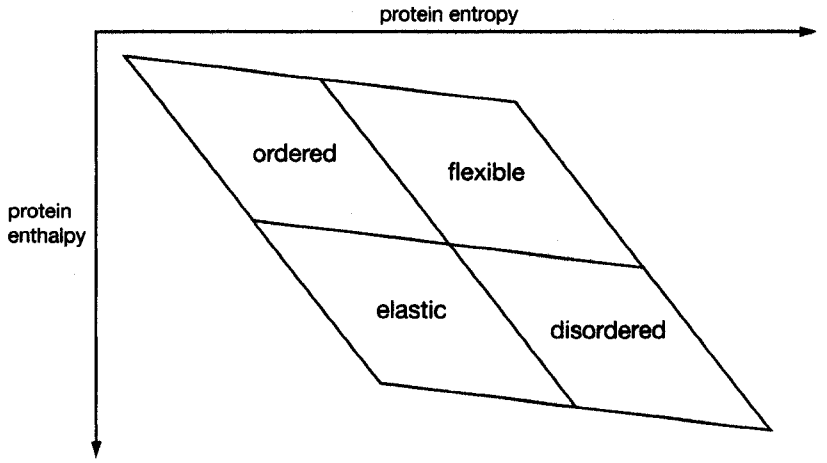


Figure 1.6: A process ontology of protein conformational space. The four regimes are defined by limiting values of stabilization (which increases as one moves upward) and ordering (which increases as one moves leftward) that result from the process of conformational fluctuation. For example, the reversible folding of globular domains can be understood as a transition from the disordered to ordered regimes. Structure-function relationships can be defined in all quadrants of this ontological space, or from interconversions between quadrants.

Ordered and flexible proteins

Ordered proteins have an energy landscape that preferentially stabilizes a single conformation. The amount of stabilization is large enough that the backbone atoms occupy a single conformational state. The stabilization of a single backbone conformation also establishes a persistent internal reference frame, so that ordered proteins can also be described as *rigid* as in the rigid body concept of classical physics. Examples of ordered regions of troponin include those regions assigned a scaffolding role, primarily the divalent-cation-saturated C-terminal domain of troponin C.

The majority of globular proteins studied to date fall into the ordered-flexible continuum. Flexible proteins are stabilized into a distinct structure, as with ordered proteins, but there are multiple stable conformations that interconvert. The energy landscape has multiple minima that are all similarly stabilized, but this stabilization is large enough that the backbone motion is hinge like. These proteins are called *flexible* to emphasize that they contain rigid parts connected by hinges. A flexible portion of troponin is the N-terminal domain of troponin C, which has been shown to undergo conformational exchange in solution. When stabilized into a crystal lattice, or into a particular functional state, flexible proteins are indistinguishable from ordered proteins.

Elastic and disordered proteins

Elastic proteins are the limit obtained when destabilizing the energy landscape of an ordered protein. If the energy landscape has a single attractor, but the stabilization of this conformation is insufficient

to constrain thermal fluctuations to a single conformation, the protein is elastic. The term *elastic* alludes to the plasticity of the conformational state, and its ability to return to a defined equilibrium conformation. Examples of elastic proteins include those described as molten globules.

Intrinsically disordered (or natively unfolded) proteins have only recently begun to divulge their structure-function relationships. Disordered proteins are often dismissed as "unstructured" because their energy landscape does not stabilize a unique conformation. There is a powerful connotation of dismissiveness in the use of the term *unstructured* in a field self-styled as *Structural Biology*. This dismissiveness is unjustified. Chemical structure is intimately linked to the notion of chemical identity [8] so existence of primary structure always implies *something* three-dimensional. The disordered proteins are not *unstructured* but rather are *polystructured*. A single conformation cannot describe their structure-function relationships. The assignment of structure-function relationships for disordered proteins should not seem too implausible following the above discussion. Most (if not all) machines function through the interconversion of functional states.

1.5 Process and troponin

I title my Thesis in homage to Alfred North Whitehead's *Process and Reality* to emphasize my process-oriented description of troponin and troponin function. The chapter you have just read shows that protein structure and function are both manifestations of process. Scientific research is a process. The meaning of written words is not intrinsic: you, the reader, are an essential factor in the process of establishing meaningful and coherent communication. I thank you in advance for your efforts — I hope the process is rewarding.

The next two chapters of this work describe how a process-oriented description of troponin can be used to unify structural and functional observations. I split this presentation into two chapters to demark more clearly the boundary between consensus and speculation. The next chapter summarizes the background knowledge of troponin structure-function, setting the stage for the presentation of my work.

Chapter 3 (p. 18) integrates three of my publications ([9–11]) into a troponin mechanism. First, I discuss the presence of conformational disorder in troponin. All three components of troponin have disordered regions that were poorly incorporated into the description of mechanism. In this chapter I present my favorite research contribution, the discovery of isoform-specific variation in the intrinsic disorder of troponin I. The chapter closes with the presentation of several models of functional disorder in troponin.

As mentioned above, biochemical machines interconvert their functional states through the effects of mass action. The energetics of mass action are examined through binding studies. Chapter 4 (p. 42) summarizes an odyssey that harkens back to my initial project in the Sykes lab. This project focuses on the binding of W7, a calmodulin antagonist and an inhibitor of striated muscle contraction, to cardiac troponin C. Most of the chapter is centered on reconciling local and global descriptions of the process of binding, drawing from two of my papers [12, 13]. This odyssey is still ongoing, as I have not found a statistical procedure for selecting the appropriate subset of local signals for the maximally robust reporting of the global event. I do show that taking the path of max-

imal ignorance — considering all of the local signals — is a more rewarding practise than manually selecting the signals for analysis.

The binding studies directly examine the process, but provide only qualitative descriptions of the functional states. In Chapter 5 (p. 71) I describe the experimentally-derived visualization (structure determination) of the W7–bound state of the regulatory domain of cardiac troponin C. The procedure used to obtain this structure is not completely novel, but it is unconventional, and so I detail this procedure. The structure does provide unique information into the function of W7 and perhaps other ligands for troponin C or calmodulin.

Chapter 2

Switching in Striated Muscle

The introductions of troponin papers are interesting to compare. Each serves as a sort of mini-review of the pertinent literature. Any consensus between these many mini-reviews reflects a true scientific consensus because the same facts are echoed from all over the community, not just from the experts commissioned to write *actual* reviews. One of the first things I learned from delving into the forest of literature on troponin is that the introductions can be boring to read after a while. But when lost in this forest of data the background patterns also become comfortably familiar.

This chapter summarizes the background knowledge of structure-function in troponin.[1] Background knowledge is widely accepted and is not expected to change given new data. Here I survey the reductionist program from intact muscle tissue to recombinantly expressed, purified components. This chapter closes with the widely-accepted *steric blocking mechanism* of troponin regulation, including the knowledge of protein structure in this mechanism. The march of reductionism proceeds throughout the remainder of my Thesis.

2.1 Striated muscle

The regulation of striated muscle contraction is a centerpiece of science. The renaissance scientist Galvani discovered the electrochemical regulation of muscle contraction over 200 years ago [14]. The observation of an inorganic trigger for a physiological event presaged the end of vitalism and the beginning of electrochemistry [15]. Striated muscle is the first biochemical machine to be operated outside of the organism, the first evidence of life's physical underpinnings.

Striated muscle includes the voluntary and cardiac muscle tissues[2]. As the voluntary muscles coordinate motility, breathing, and venous blood circulation, and as the cardiac muscles drive arterial circulation, striated muscle is an integral component of a boatload of organisms. Much effort is directed towards the understanding and manipulation of this class of essential tissue. My mentor and his coworkers have made important contributions to the understanding of the regulation of striated muscle contraction, specifically, the workings of troponin.

[1] I have derived this material from the introductory sections of my publications, primarily references [9–11].

[2] The other major muscle tissue is smooth muscle, which is involuntary, and controls vasoconstriction and peristalsis.

2.2 Regulation by [Ca^{2+}] transients

Inside a striated muscle cell, $Ca^{2+}$is sequestered into a compartment called the sarcoplasmic reticulum. This lowers the free [Ca^{2+}] to sub-μM levels [16]. Electrochemical stimulation of the muscle cell body leads to a cascade of depolarization events inside the cell, akin the propagation of depolarization in a neuron. The depolarization cascade culminates with the release of $Ca^{2+}$from the sarcoplasmic reticulum, raising the cytoplasmic [Ca^{2+}] to μM levels. The Ca^{2+}binds to troponin leading to the movement of tropomyosin and the activation of muscle contraction [16, 17]. The troponin-tropomyosin system is not a force-generating component of muscle; muscle contraction continues in an unregulated fashion in the absence of troponin-tropomyosin. Generations of scientists have, collectively, disassembled and reassembled the regulatory apparatus, each insight providing impetus to the march of reductionism.

The function of a biochemical machine is defined in the context of the organism. As such, striated muscle is best-studied in the living organism itself. As with most biochemical machines, it is not feasible to directly monitor the function of troponin in the living organism. The consequences of troponin function — regulation of muscle contraction — are more readily observed. So troponin is usually characterized in terms of its effect upon striated muscle contraction (or some correlate of contraction, like ATP consumption). There are three broad classes of striated muscle tissue in humans: cardiac muscle, fast-twitch skeletal muscle, and slow-twitch skeletal muscle. The latter two isoform are both voluntary muscle.

2.3 Model tissues: the heart and soleus

Two striated muscle tissues have received most of the attention. One, the soleus, contributes to calf flexion. This muscle is readily dissected from rabbits and has been a favorite model tissue for over 50 years.[3] The heart is another subject of intensive study for multiple reasons, none the least being the prevalence of cardiovascular disease in developed countries. A variety of techniques exist to study the heart within or without the organism, including dissecting individual muscle fibers from the bulk. More recently, a third system has received much attention – cardiomyocytes (cardiac muscle cells) grown in cell culture. Both muscle tissues lend themselves preferentially to different types of inquiry. All of these functional contexts impart a dauntingly rich variety of functional manifestations of the same underlying switch, somewhat akin to the plethora of automobiles employing the same underlying internal combustion engine.

Studies of in-place muscle fibers

The observations of Galvani show that striated muscle can respond to an electrochemical stimulus, even when the organism is long-dead by most accounts. This experiment is still recapitulated under the somewhat obscure naming, *in situ*. The phrase, Latin for *in-place* designates that a muscle

[3] One reason rabbit troponin was favored, historically, is that its TnC ortholog has a single sulfhydryl group, making it amenable to study with sulfhydryl-modifying probes.

fiber is dissected away from an animal and mounted in an apparatus where the force generated from contraction can be measured. The components of the muscle fiber can then be studied *in* their functional *place* albeit outside the organism. Most in-place studies also employ biochemical techniques in order to permeabilize the fiber and to manipulate the intracellular conditions.

Permeabilization

In order to switch an in-place muscle fiber between contraction and relaxation conditions, scientists typically permeabilize the muscle fiber. Solutions of ATP and Ca^{2+} promote contraction, chelator and ATP promote relaxation. But these solutions can only equilibrate with the muscle fiber if the fiber is permeabilized. Permeabilization consists of chemical or mechanical skinning; with chemical skinning a non-ionic detergent is used to permanently disrupt the outer membrane of the muscle fiber [18]. Besides allowing for manipulation of intracellular Ca^{2+}, the permeabilization treatment also facilitates the re-equilibration of the protein components.

Equilibration

A skinned muscle fiber can be exchanged for its troponin, or individual troponin protomers. This technique is routinely used towards the introduction of recombinant troponin (often harboring an experimentally-induced mutation) into a functional context. Substitution of cardiac troponin into a skeletal muscle fiber, and the inverse manipulation, help to clarify which portions of the regulatory apparatus confer tissue-dependent functions. Experiments of this variety have secured the overall assignments of the functional roles of the thin filament. With the functional assignments secured in broad strokes, experimentation can proceed towards the most reductionist levels.

2.4 Structure-function relationships of isolated components

A number of studies have oriented on isolated components of the thin filament, with the historically pivotal discovery of troponin by Ebashi in 1955 being the clearest point of departure from the holistic studies [19]. Subsequent work (including [20–24]) delineated the three components of troponin, and established their functional roles in broad strokes. The steric blocking model of Potter and Gergely was articulated in 1974 [25], establishing the overall functional roles of the troponin proteins. Troponin is composed of three subunits: troponin I (TnI), troponin C (TnC), and troponin T (TnT) [16].

2.4.1 The function of troponin: the steric blocking model

The steric blocking model describes two functionally relevant states of tertiary and quaternary structure, corresponding to high and low sarcoplasmic $[Ca^{2+}]$. In both of these states, TnT is stably bound to tropomyosin (Tm), coupling the position of Tm with the conformation of TnT. In the inhibited, low-Ca^{2+}-state, TnI binds to actin [25], preventing myosin-actin crossbridge cycling and inhibiting muscle contraction [22]. Ca^{2+} binding to TnC initiates conformational changes which

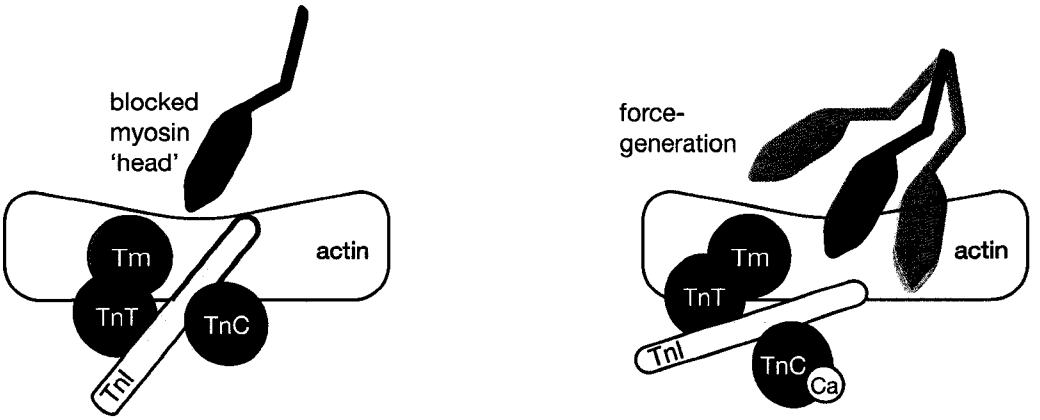


Figure 2.1: The steric blocking model. Note that this diagram illustrates functional relationships, not the actual structures of the proteins. Portions of myosin and tropomyosin (Tm) are both depicted in black; actin is the white, dimpled surface. Actin binding is indicated with magenta outlining. The three troponin proteins are labeled. In the low-Ca^{2+}state (left) the myosin head cannot bind actin, so force is not generated. Ca^{2+}-binding to troponin (specifically, TnC) leads to a series of motions which allow the myosin head to complete its ATP-dependent force generation (right). The motions of the myosin head are indicated with juxtaposed silhouettes.

propagate to the other troponins, releasing the thin filament from inhibition. While enhanced over the years [26–30], this overall description has stood mostly unchallenged [30]. See Figure 2.1 for the troponin mechanism in the broadest possible strokes.

The three troponin protomers have distinct structural and functional roles. Two of the proteins, troponins I and T, are homologous but highly diverged. They are still disclosing new structure function relationships. Troponin C is highly homologous to calmodulin, and is a much more typical globular protein, and is also much better understood in terms of structure-function. As should be evident by the end of this chapter, the best-understood regions of troponins I and T are those that abut troponin C. The following sections briefly introduce the three troponin protomers, so that the reader is aware of the overall structure-function relationships before high-resolution models are reviewed.

2.4.2 Coupling of Tm and Tn: Troponin T

Troponin T (TnT) is named for its tropomyosin-binding capacity [31]. The general layout of the protein has two domains, initially identified through partial proteolysis. The N-terminal (or T1) domain is stably associated with tropomyosin. This proline-rich domain is thought to adopt an extended or fibrous conformation, and makes multiple contacts with tropomyosin. The mode of coupling with tropomyosin is not known in detail. The C-terminal (or T2) domain makes well-known contacts with both troponin I and troponin C.

TnT is one of the most variable of all of the thin filament proteins as all tissue isoforms are also present in splice variants. The functional significance of these isoforms is a subject of ongoing inquiry. This protein has not been extensively studied by myself, nor my colleagues in the Sykes

lab. It gets the least amount of attention in my Thesis.

2.4.3 The inhibitory factor: Troponin I

Troponin I (TnI) is named for its inhibitory properties. It binds a stretch of polymeric actin in a 1:1 fashion, obstructing the binding site for myosin [16]. The stretch of TnI that occludes myosin-binding to actin is called *the inhibitory region*[4]. C-terminal to the inhibitory region, the *switch region* [5] selectively binds to NTnC in the presence of Ca^{2+}. This Ca^{2+}-dependent interaction is the 'main event' of troponin activation.

TnI is thought to make two points of contact with actin [29]. The second actin binding site has been localized to the residues C-terminal to the switch region; this region of TnI is extensively discussed in the next chapter. In its functional context, TnI does not have a 1:1 stoichiometry, and its ability to impart steric blocking necessitates the presence of tropomyosin and TnT [16], and perhaps also the C-terminal actin-binding site.

The cardiac isoform of this protein has a ~30 residue extension that distinguishes it from the fast- and slow-twitch skeletal isoforms. The cardiac-specific extension is a phosphorylation target for protein kinase A [32]. Other functional regions of TnI will be introduced later.

2.4.4 The Ca^{2+} sensor: Troponin C

TnC has two ordered domains separated by a linker. Each domain contains two Ca^{2+}binding sites. Sites 1 and 2 are widely considered to be the actual regulatory sites, while sites 3 and 4 are 'structural' or continuously metal-bound over a contractile cycle [25]. TnC's structure in the low-[Ca^{2+}] state has been known for decades [33, 34]. Ca^{2+}-activation, saturation of sites 1 and 2, leads to exposure of a hydrophobic pocket in the N-terminal domain [35,36]. This exposed hydrophobic site is the binding target for the switch region of TnI.

2.4.5 Three-state models

The steric blocking model originally described two-states of the thin filament, defined in terms of high and low sarcomeric Ca^{2+}-concentrations. The steric blocking model has been extended to three-states. This was proposed in kinetic studies by McKillop and Geeves [37], with important contributions from Lehrer [38]. Later work also resolved three structural states of the thin filament [39–42]. These states can be characterized by the position of Tm relative to actin [43]. As polymerized actin and tropomyosin are both helical proteins, their conformational states can be monitored with fiber diffraction and helical reconstruction techniques.

The contraction-inhibited, low-Ca^{2+}state, is denoted the blocked (B) state [43]. The B state is in equilibrium with the Ca^{2+}-activated (C) state [38,43,44]. The transition from B to C corresponds to Ca^{2+}-activation. The C state is in equilibrium with the open or myosin-induced (M) state, where myosin can tightly bind actin [43]. It should be emphasized that the M state can only be occupied

[4]Cardiac $TnI_{128-147}$ or fast skeletal TnI_{96-115}
[5]Cardiac $TnI_{147-163}$ or fast skeletal $TnI_{115-131}$

when myosin is bound to actin. A depiction of the three-state steric blocking model is presented in Figure 2.2.

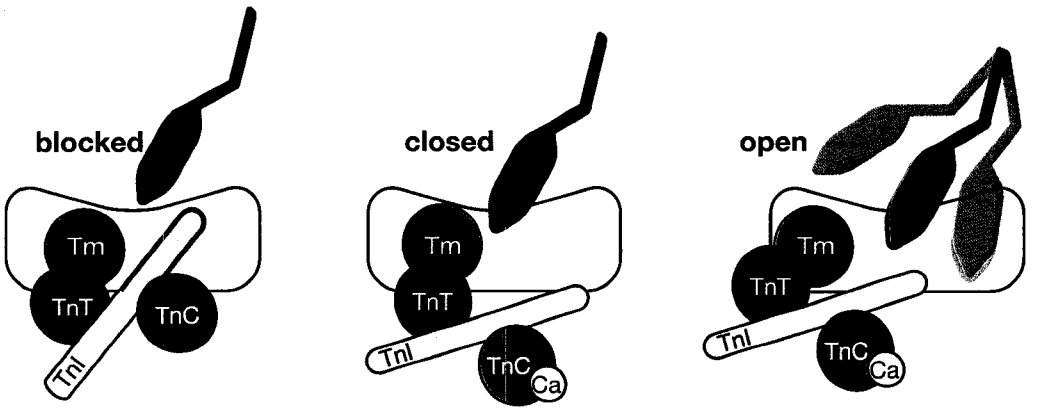


Figure 2.2: The three-state steric blocking model. The progression from the blocked to the closed states (moving from left to center) corresponds to Ca^{2+}activation, as shown in the two-state model (Figure 2.1) but with tropomyosin (Tm) continuing to interfere with force generation even as myosin weakly binds (magenta outlining). A Ca^{2+}-independent, myosin-induced shift in the position of Tm leads to the open state (right) where force is generated.

2.4.6 The troponin core complex structures

Recent crystallographic studies have provided structures of large portions of the troponin complex, in both skeletal and cardiac systems. These structures visualize many protein-protein interactions previously predicted by structural studies on smaller fragments of the Tn complex [30]. As such, they are 'empirical reviews,' incorporating many independently-procured insights into a comprehensive picture.

The Ca^{2+}-activated cardiac troponin core complex

The cardiac troponin (cTn) core complex structure, in the calcium activated state (Figure 2.3, right) was presented by Maeda and coworkers [45]. This X-ray study afforded two models, each derived from distinct crystal forms, and refined to 2.6 and 3.3 Å resolution. Each model had two cTn molecules in the asymmetric unit, so the study determined the structures of four molecules of Tn. The two complexes crystallized lacked the N-terminal domain of $cTnT_{1-182}$, and varied in the length of the TnI component crystallized. One crystal contained $cTnI_{31-163}$, the other contained $cTnI_{31-210}$.

Cardiac TnI and TnT are visualized as two pairs of alpha-helical chopsticks grasping TnC by its C-domain (see Figure 2.3, right). The two chopsticks interdigitate in a coiled-coil fold, forming the *IT arm.* Prior results anticipated the coiled-coil interactions stabilizing the IT arm [28], and the various contacts between TnI and TnC. The cTn crystal structure, however, showed a novel orientation of the N-terminal domain of TnC (NTnC) with respect to the remainder of the IT arm.

The C-terminus of cTnI (CTnI), containing the switching and inhibitory regions, extends off of one chopstick, and interacts with NTnC. The core cTn molecule has two major subdomains, with the regulatory head containing NTnC and CTnI, and the IT arm containing the C-terminus of TnC (CTnC), the N-terminus of TnI (NTnI), and TnT. The structures intimate a cascade of protein-protein interactions upon Ca^{2+} activation of cTnC, with the IT arm's orientation/stability being potentially modulated by cTnC activation.

Crystal structures of the inhibited and activated skeletal troponin core complexes

Fletterick and coworkers recently reported the (chicken fast-twitch) skeletal troponin (sTn) core complex structure, in both Ca^{2+}-activated (resolved to 3 Å , Figure 2.3, center) and inactivated (7 Å , bottom of Figure 2.3, left) states [1]. This pair of structures represents the highest resolution description of the Ca^{2+}-induced switching of troponin. As in the study by Takeda and others [45], all of sTnC, most of sTnI ($sTnI_{1-182}$ [for Ca^{2+}-activated state] or $sTnI_{1-137}$ [for inactive state]), and $sTnT_2$ ($sTnT_{156-262}$) were crystallized. Consistent with previous data, sTnI and sTnT form a pair of chopsticks grasping sCTnC, with the regulatory portion of TnI extending off the end of the C-terminal chopstick.

Electron-microscopy-derived models

Cardiac and skeletal muscle are both called striated muscle for their regularly repeating appearance, visible with light magnification. The striations allude to the similarly regular organization of the constituent molecules. This long-range molecular ordering allows for fruitful imaging and diffraction investigations. The use of electron microscopy and helical reconstruction techniques have allowed for the visualization of the thin filament at the molecular level, defining the relative positions of actin and Tm in the presence and absence of Ca^{2+} [41, 43, 46–48]. Such efforts have documented the rolling movement of Tm upon Ca^{2+} activation [47].

Troponin has largely resisted characterization in its functional context by image reconstruction techniques. The thin filament proteins actin (in its polymeric state) and tropomyosin both exhibit helical symmetry, which is usually exploited in helical reconstructions, but troponin's position around the thin filament axis does not exhibit the same symmetry, and therefore has been recalcitrant (but not unyielding) to helical reconstruction efforts [41, 46]. As the structural and mechanistic results become increasingly better-integrated into a holistic picture, the poorly resolved aspects of troponin fall into sharp relief.

2.5 Foreground results

The description of the background knowledge of structure-function in Tn should help to emphasize the poorly-resolved areas of troponin. These are the tropomyosin-binding domain of TnT, the terminal regions of TnI, and the interaction sites of TnI and actin. Of these areas, there are two structural results that *would* inform the current discussion *if* they were accepted as consensus. As those results are discussed in light of my own work, I introduce them in the next chapter.

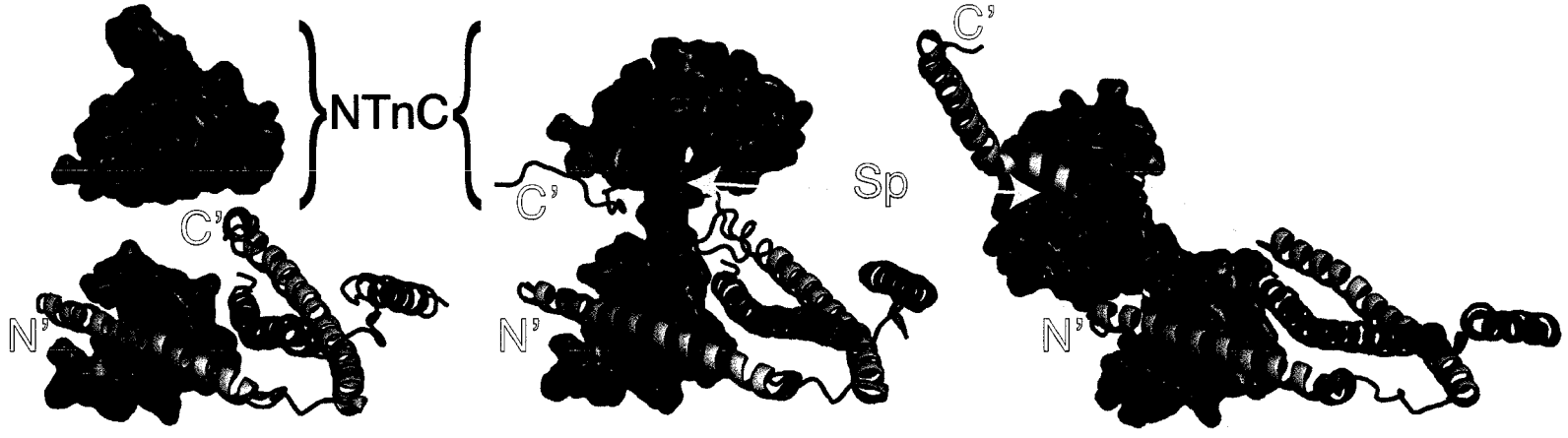


Figure 2.3: Representative coordinates of the troponin core complex structures. At left and center are the chicken fast skeletal core complexes determined in low (left) and high (center) [Ca^{2+}]. The Ca^{2+}-activated human cardiac complex is on the right. All structures are oriented similarly with respect to the C-terminal domain of TnC (red surface) which binds the N-terminal region of TnI (yellow cartoon). To orient the complexes, they were first superimposed over the backbone heavy atoms for residues 91–157 of TnC [less than 1 Å pairwise RMSD], and then were translated so that they can be viewed side-by-side. The core complexes lack the tropomyosin-coupling domain of TnT; only the C-terminal domain is present (blue cartoon). The mechanistically pivotal interaction between the switch region of TnI (Sp) and the N-terminal domain of TnC (NTnC) occurs only in the two Ca^{2+}-activated complexes (center and right). The cardiac structure (right) has a missing region (crystallized but not visualized) of TnI connected with a dashed yellow line to communicate the topology.

Chapter 3

Disorder in Troponin

Shape without form, shade without color,
Paralyzed force, gesture without meaning;

from *The Hollow Men* by T.S. Eliot, 1925

Following from the central metaphor of *troponin as a switch*, both ordered (rigid-body) and disordered (flexible, or hinge-like) portions of the machine can be anticipated. Some of these features have been readily assimilated into popular mechanisms, primarily the events surrounding the Ca^{2+}-activation of TnC. Another proteinaceous mechanism for switching, the disorder-to-order transition (protein folding), was not anticipated by the early exponents of *troponin as a switch.*

Due to a number of practical considerations and theoretical limitations, the ordered states have been much more readily described in the context of a functional mechanism. But without a description of the process of conformational interconversion, a troponin mechanism is as vacuous as T. S. Eliot's hollow men. High resolution structures provide the shape of troponin, but this image is paralyzed into a single mechanistic gesture, and without intrinsic meaning. This chapter integrates all of my[1] interpretations of conformational fluctuation in troponin.

3.1 Mechanistic disorder

Most disorder-function relationships in Tn are either debated or not considered at all. The first reason for this is that the disordered regions of troponin are not described by high-resolution models, and so they are not amenable to interpretation within the dominant structure-function paradigm. The question of *whether these molecules are truly disordered* is frequently raised. This is a reasonable line of inquiry because most reductionist studies are performed very high water activities compared

[1]Others that have contributed include my coauthor Tharin Blumenschein. Tharin is the original proponent of the fly-casting mechanism for the association of TnI and actin, following from her work characterizing TnI dynamics in the troponin core complex by NMR relaxation spectroscopy [49]. I contributed the role of disorder-order transition in the process of activation, crystallized the idea that the kinetic paths for activation and inhibition are structurally distinct, and built the models and diagrams to communicate the mechanism. This collaborative work was published as Reference [9]. Material for this chapter is derived from [9–11] as well as unpublished material.

to intracellular conditions (with protein concentrations on the order of 400 mg/mL). The short answer is, *frequently yes* [50,51]. In fact, crowding could be expected to *promote* intrinsic disorder for some proteins because the folded state can have a higher surface-area-to-volume ratio, and therefore require more solvent, compared with a disordered but collapsed state.

The other frequently raised challenge can be articulated along the lines of, *why would evolution lead to disordered machines, when ordered machines work so much more deterministically?* This is a less reasonable position, reflecting a type of enthalpic-centricism. The *fact* that proteins can fold when driven only by random, thermal fluctuations is *proof* that disordered processes can give rise to some of the most exquisite machines known to the biosphere. A large body of literature describes plausibility, and in some cases, the necessity of conformational interconversion in protein function [51–58].

Disorder and dynamic allostery

Hilser and Thompson have developed theoretical support for a role of intrinsic disorder in the context of allosteric coupling [59]. Allosteric coupling is the 'extent of allostery'; the extent to which ligand binding to one site causes modulation of the affinity of a second site. Their model is based on the assumption that a domain folds prior to binding a ligand, and that only folded domains can interact with each other, contributing an energetic term to the free energy. No assumption is made of the nature of the interdomain interaction (not even the sign of its free energy change). They map the parameter space, and find several optimal regions. They find the allosteric coupling to be maximal between two domains when one or both of the domains is unfolded a substantial fraction of time.

Joshua Wand has also made a strong case for the plausibility of entropically-mediated allostery. He has developed his ideas in the context of calmodulin, which is highly homologous (~50% sequence identity) with troponin C. His position has been expounded over many years, in many publications [60–63] (not an exhaustive list), but the argument finds its most distilled presentation in the context of debate [64]. I cannot improve on his presentation of his position:

> "The time scales of structural transitions and functional events need not correlate with the rates of interconversion of multiple states. Thus, although the transition from one form to another may indeed occur in a given time regime, the origins of allosteric activation may arise elsewhere, i.e., in the conformational entropy underlying the dynamics."

3.2 Disorder in TnC: the DE linker

The core cardiac Tn complex crystal structures (Chapter 2) visualize the complex multiple times, with noteworthy points-of-difference between each observation. Takeda and others interpret the four Tn molecules from two distinct crystal forms as reflecting the flexibility of the DE linker. They write, "...the linker between the two subdomains works as a universal joint." [45] When the four troponin core molecules are superimposed upon cNTnC, the end of the IT arm sweeps out a distance of ~27 Å . This shows the N-domain sampling many orientations with respect to the IT arm. The authors assert the DE linker to be poorly structured "owing to the lack of specific interactions with

the rest of the molecule" [45] which makes sense in the context of the local density model for B factors [65] and NMR order parameters [66]. They cite the conclusions of mutational studies which suggest that deletions to the DE linker are not as deleterious to Tn function as substitutions that rigidify the structure. This interpretation alludes to the findings of Maeda and coworkers in a smaller cTnC•cTnI crystal structure [67]. As the DE linker and Ip regions were crystallized, but not visualized, the DE linker•Ip region likely has multiple conformations with respect to the bulk of the complex. It is possible that the DE linker•Ip region is stabilized by similar interaction in both sTn and cTn, but that this portion of the crystal ensemble is converged in the sTn structure [1] but not in the cTn structure [45].

DE linker rigidity

Should the DE linker exhibit substantial structure in solution, its α helical conformation would ensure a fixed relative orientation between NTnC and CTnC, which would seem mechanistically significant given the quaternary structure of Tn. For example, TnC's interdomain disposition in the sTn complex [1], is accurately predicted from the original crystal structure of sTnC•2Ca^{2+} [33]. This is a consequence of canonical α helicity in the DE linker, in both of these structures. It is parsimonious to consider the intact α helix observed in the Ca^{2+}-activated sTn core complex [1], and other crystal structures [33,34], to be a substantially populated microstate in the native ensemble.

The hypothesis exists: the DE linker is rigid in the Ca^{2+}-activated state but not in the inhibited state [1, 68]. Surely, a structured DE linker presents a stronger mechanical coupling of the head and IT arm, consistent with the linker possessing the property of rigidity. It could be described accurately as "relatively rigid" comparing the two states [68], reflecting that the NTnC and CTnC are more strongly coupled in the Ca^{2+}-activated state. One may question whether this is actually a change in rigidity, however. If the DE linker alternates between being structured and disordered, it cannot be identified as 'rigid' as in the 'rigid body' approximation from mechanics[2], but rather alternates between weakly and tightly constrained environments. For example, a pipe cleaner does not gain intrinsic rigidity upon insertion into a pipe stem. It simply becomes constrained. The DE linker may become more structured when bound to Ip but this does not make it inherently rigid. Although this point seems semantic, such language perpetuates the idea that the DE linker is converged upon a structured helix, in solution-state Tn•2Ca^{2+}. Any increased rigidity in the system requires description of the *mutual* stabilization of TnC and TnI. Appealing to my ontology of protein structure (Section 1.4, p. 6), the DE linker is flexible or elastic but not ordered.

Evidence against DE linker rigidity

Evidence against the rigidity of the DE linker, in both TnC and calmodulin, has been presented previously. A crystallographic study reviewed a number of TnC structures and concluded that the central helix was indeed flexible [69]. The helix was intact in the structures reviewed in that study.

[2] A rigid body's internal frame of reference remains untransformed in the presence of external forces. The definition of an internal reference frame implies that the orientation or location of the rigid body is irrelevant.

Analysis of both Bragg and diffuse scattering of calmodulin crystals concluded that uncorrelated interdomain motions do exist at the expense of the rigidity of the central helix [70].

Backbone amide ^{15}N transverse relaxation rates (R_2) are functions of the rate of rotational reorientation. Recent R_2 measurements on $\{^2H,^{15}N\}$-labeled sTnC in uniformly ^2H-labeled sTn [68] show that Ca^{2+}-free sNTnC (assembled into the sTn core complex) has a higher rate of rotational diffusion than the remainder of the Tn complex [68]. Such dynamics do not preclude the DE linker's structure; rather, this structure is transient, or contingent on additional stabilizing factors. For example, sNTnC tumbles at a similar rate to the rest of the complex in the Ca^{2+}-activated state [68]. This completely corroborates the structure of the sTn core complex [1] and the general notion that the DE linker undergoes disorder-to-order transitions in the context of its function. I depict this feature explicitly in my troponin mechanism (Figure 3.6, p. 36).

3.3 Disorder in TnI: the mobile domain

Due to a slow rate of molecular tumbling, most of the NMR signals from ^{15}N-enriched sTnI in the ~52 kDa sTn complex were too broad to allow for observation [2,49]. A subset of $\{^1H,^{15}N\}$-HSQC crosspeaks (about 40 residues worth of signals) were sharp and dispersed, however, prompting their sequential assignment by Wakabayashi and coworkers [2] and independently by Blumenschein and others [49]. These signals originate in a domain of sTnI C-terminal to sSp, residues 131-182 of chicken sTnI. This domain has higher rotational diffusion than the bulk of sTn and is named the mobile domain (sMd) [2]. The name is intended to denote a stably folded *domain* that is spatially *mobile*, as demonstrated by the structure determination of this region.

Structure in the mobile domain

Murakami and others show how it is possible to define tertiary structure in the highly dynamic mobile domain [2]. The steady-state $\{^1H\}$-^{15}N NOE enhancement for this region averaged to 0.5. Compared with typical globular proteins, such NOE values are indicative of conformational exchange. The presence of interconverting conformers does not, however, preclude the presence of a native conformation. This means that the native state consists of a family of nearly isoenergetic but nonidentical conformers. This region of sTn was structured with NOE-derived distance restraints; secondary structures were further stabilized using hydrogen bond restraints [2]. Hydrogen bonds were assigned in regions of assignable secondary structure as indicated by canonical medium-range NOEs. So this study showcases the ability of the NOE to disclose long-range interactions in fluctuating protein structures. Other NMR observables such as the chemical shift, J-couplings and residual dipolar couplings, are scaled down more extensively due to the averaging of multiple conformations, and would prove less useful when defining the most-populated folded state. Indeed, in an independent study by Blumenschein and others [49], Cα chemical shift analysis of sMd ($sTnI_{131-182}$) indicates sufficient intradomain conformational fluctuations to average the chemical shift to nearly random coil values. The same chemical shifts were reported by Murakami and others [2]. The data of Murakami are highly consistent with those of Blumenschein but the interpretations greatly differ.

The mechanism of Wakabayashi and coworkers: Md-clamping

The successful structure determination of the mobile domain (Md), and the accompanying cryo-EM analysis [2], allow for an elaboration on the mechanism of troponin. Ip and Md are attracted to the N- and C-termini of actin through long-range electrostatic forces. At low [Ca^{2+}], Ip and Md bind to actin, in a conformation reminiscent to a 'C clamp'. The TnI•actin interactions localize the IT arm in a region suitable to displace tropomyosin. Ca^{2+}-activation of sNTnC allows for binding of Sp. This kinetically stable interaction disrupts the Ip•Md•actin interactions enough for Ip to localize to the central helix of sTnC. For reference throughout this work, I call this mechanism Md-clamping. I think this mechanism is incorrect. The affinity of this region for actin is less than that of the inhibitory region [29] so Ip further stabilizes the binding of Md, if anything, not the other way around. This is further discussed below (Section 3.5.1, p. 34).

Previous results describing intrinsic disorder in TnI

Granier and others mapped several immunoreactive epitopes of the human-cardiac TnI isoform, and also predicted secondary structure (and by exclusion, disordered regions) [71]. They found several epitopes that bound to the cardiac-specific N-terminal extension (cNp), which they also predicted to be disordered. Ferreira and coworkers showed chicken muscle TnI and TnT were partially disordered through the spectroscopic monitoring of unfolding equilibria [72]. Cardiac TnI and TnT were also previously implicated to be highly disordered using the PONDR VL-XT algorithm [73]. In cardiac TnI, 156 residues were identified as being disordered (75% of all residues). In cTnT, 235 residues (79%) were identified to be disordered. Finally, I emphasize the results of Sykes and coworkers [49] showing TnI to be highly mobile when incorporated into the troponin core complex through NMR relaxation spectroscopy.

TnI isoform switching

The cardiac isoform of TnI (cTnI) has an N-terminal extension amounting to ~30 more residues than the skeletal isoform (sTnI). This region includes phosphorylation sites for multiple protein kinases including PKA and PKC. As such, the presence of a non-cardiac TnI in heart muscle has readily anticipated functional consequences. In mammals, TnI isoform switching occurs in at least two scenarios. The embryonic/neonatal expression of slow-skeletal TnI in cardiac muscle is thought to improve the robustness of cardiac function in the hypoxic conditions accompanying birth [74]. As summarized in [75], expression of slow-skeletal TnI in cardiac tissue has a number of well-characterized effects. These include an increase in Ca^{2+}sensitivity [75–77], a longer relaxation/relengthening time [78] (although this is not uniformly reported [77]), an increased resilience to pH changes [76], and a decrease in β adrenergic sensitivity. The decrease in β adrenergic sensitivity is entirely attributable to the loss of cardiac TnI's unique N-terminal extension, a known protein kinase A phosphorylation site [32, 79]. The other effects are less easy to explain, although they are mostly attributable to the change in isoform [44, 75]. Cardiac expression of slow-skeletal TnI also occurs in the scenario of heart failure. Proteolysis of cardiac TnI is known to occur in instances of

cardiac stunning[3] [74,80]; the major proteolytic product is cardiac TnI_{1-193} which has been characterized as having normal actin binding equilibrium properties, but an apparent decrease in inhibitory activity, as estimated by reduced myosin ATPase activity and isometric force [81]. Assuming an adaptive significance to the pathological expression of slow-skeletal TnI, the unique properties of this isoform could lead development of novel cardiotonic drugs, or an enhanced understanding of the known drugs.

3.4 Prediction of intrinsic disorder in TnI isoforms

TnI's propensity to intrinsic disorder is well-characterized. But few studies have focused on the isoform-dependence of ordering tendencies in TnI. The following section reproduces an in-depth characterization of intrinsic disorder propensities in the three isoforms of human TnI [11].[4]

3.4.1 Hydrophobic cluster analysis of TnI isoforms

Considerations regarding the generality of hydrophobic stabilization leads to a graphical approach called hydrophobic cluster analysis (HCA) [82,83] which depicts the primary structure in a particular bidimensional representation. Hydrophobic residues that are not clustered in the linear (as normally written) sequence become much more easily detected through this analysis. HCA is useful for the identification of small domains subject to induced structuring and for the clarification of domain boundaries [83, 84]. HCA is sensitive to local structuring because it is just a particular way of collapsing a linearly represented sequence into a two-dimensional array.

The HCA[5] is depicted in Figure 3.1. The HCA algorithm draws lines around the contiguous hydrophobic clusters, making them visually apparent. The clusters are numbered starting from the N-terminus – this annotation was done through alignment of the HCA output so that the hydrophobic clusters were maximally overlapped between isoforms, as shown in Figure 3.1. The major limitation of HCA is that interpretation of the results is not unambiguous [84]. It could be argued, for example, that the conserved sequence element MF at residue 176 of fast-skeletal TnI should be assigned an HCA cluster. Comparing the cardiac isoform with the others (over this region), only a single hydrophobic residue is present (F208) at this position, so I did not assign the cluster to be conserved across the isoforms. For similar reasons, the assignment of HCA2 is tenuous because the cluster spans only two residues, and one of the hydrophobes is absent from the cardiac isoform. Because the missing hydrophobe is substituted with alanine, which can act as a 'placeholder' hydrophobe in this analysis [83,85], I assign HCA2 to be conserved. I also note that clusters 4 and 5 are contiguous for cardiac TnI.

[3] A stunned heart is in a temporarily weakened state following ischemia.

[4] Methods for this chapter are computational, and will be communicated in-line with the discussion. TnI amino acid sequences were obtained from PubMed (http://www.pubmed.gov). Human fast-skeletal TnI, slow-skeletal TnI and cardiac TnI were retrieved from GenBank accessions gi1351297, gi437009 and gi30141910, respectively. The chicken fast-skeletal isoform, accession ga37473768, was also examined as this isoform is the subject of many structural studies. The amino acid sequences were analyzed with the N' Met present.

[5] HCA was performed using a webserver located at http://bioserv.rpbs.jussieu.fr, authored by C. Gaboriaud, V. Bissery, T. Benchetrit, and J.P. Mornon.

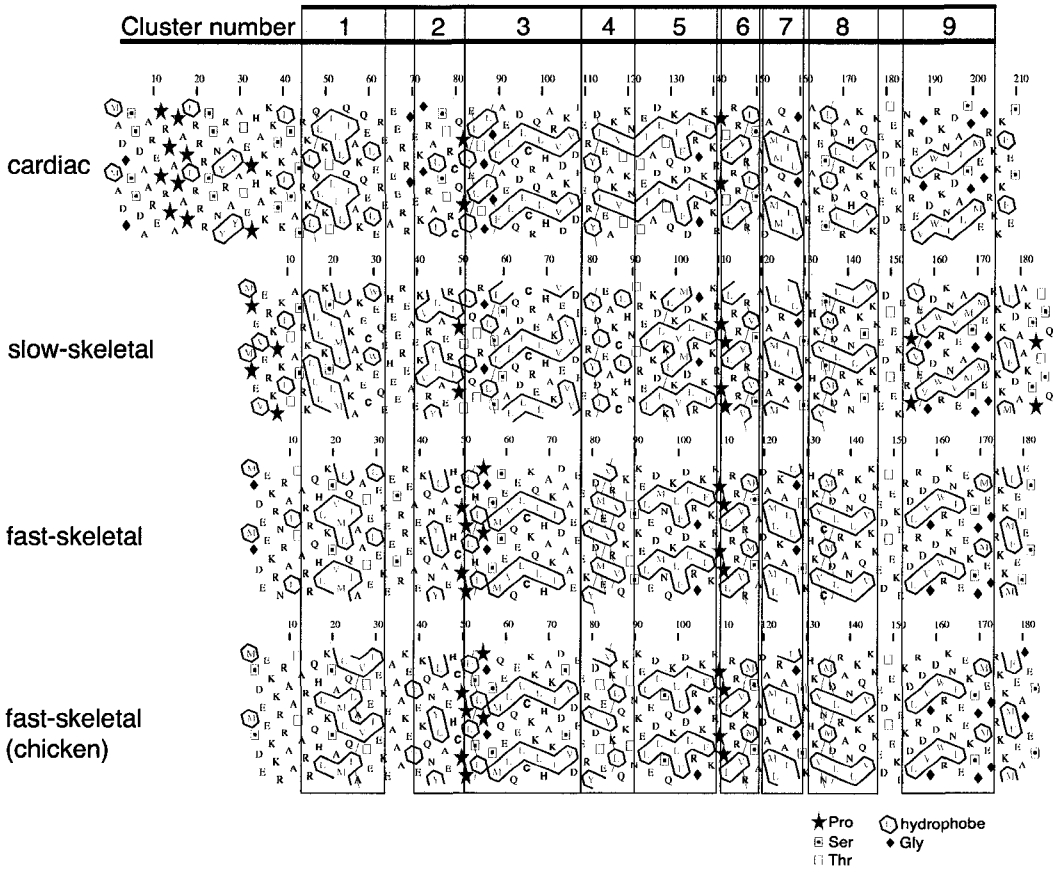


Figure 3.1: Hydrophobic cluster analysis of the three human isoforms of TnI. Given an input sequence the algorithm transforms the data into a two-dimensional representation, and automatically annotates the patterns of hydrophobic clustering. Conserved clusters have been manually annotated. Annotation of the data requires qualitative judgments as to the assignment of alanine or cysteine residues to a hydrophobic cluster and the minimum size of a hydrophobic cluster [83, 85]. This graphic is reproduced with permission from Reference [11].

The assigned HCA clusters (Figure 3.1) enrich the more traditional structure-function analyses, such as protein crystallography (Figure 3.2) and multiple sequence alignment (Figure 3.3). The conserved HCA clusters for TnI are mapped onto its sequence in Figure 3.3; the dashed blue lines under the sequence alignments span individual HCA clusters. Structural-functional annotations are shown beneath the HCA clusters in Figure 3.3. It is evident that there is only an approximate relationship between the precise location of an HCA cluster and the accepted structure-function annotations. The L2 and H2 [6] regions are spanned by HCA3 (Figure 3.3). The H2 and Ip regions are spanned by HCA5; the Ip and Sp regions are spanned by HCA6. This analysis suggests that the functional annotations do not directly correspond to distinct, hydrophobically stabilized structural modules. This could be illustrative of the important role of electrostatics in the stabilization of the Ip•TnC interaction [1, 86].

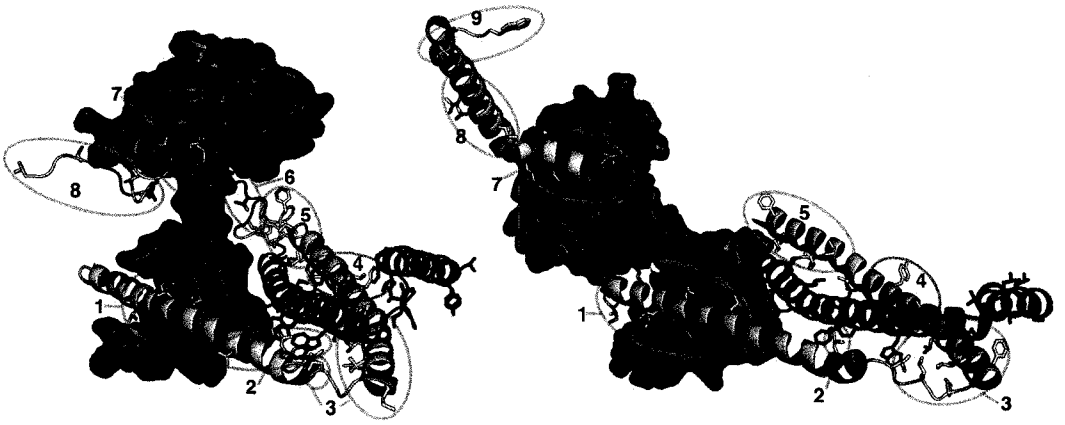


Figure 3.2: The core skeletal (PDB 1YTZ) and cardiac (1J1E, chains D-F) Tn complexes, on left and right, respectively. They are oriented identically with respect to the C-terminal (bottom) domain of TnC (see the explanation in Figure 2.3, p. 17). TnC is depicted as a red surface. TnI is depicted with a yellow cartoon. TnT is depicted with a blue cartoon. All nonhydrogen atoms of hydrophobic sidechains are shown for TnI and TnT. Hydrophobic clusters for TnI are labeled and their sidechains are circled. An interesting difference between the structures, the differential burial of HCA4, is highlighted with a magenta circle. Perhaps the observed HCA6–TnC interaction for the skeletal complex (left) is a related to stabilization propagated C-terminally from the HCA4 cluster, mediated by HCA5 (see Figure 3.7, bottom). This graphic is reproduced with permission from Reference [11].

Most of the clusters can be visualized on the structure of the Ca^{2+}-activated fast-skeletal Tn core complex [1] and the Ca^{2+}-activated cardiac Tn core complex [45], as shown in Figure 3.2. Both structures visualize only a portion of TnI C-terminal to the switch peptide (containing HCA8 and 9). There are only subtle differences in the structural localization of the HCA clusters between the cardiac and skeletal isoforms. Two HCA clusters (HCA6 and HCA9) are not visualized in the cardiac Tn structure (Figure 3.2, right), although just a portion of HCA9 is visualized in the

[6] This nomenclature for TnI's structural elements names regions as helices or loops. The N-terminal helix, H1, interacts with CTnC. A linker (L1) follows, then H2 makes the coiled-coil interactions with TnT, forming the IT arm. C-terminal to H2 is the inhibitory region.

skeletal structure (Figure 3.2, left). The principal difference between the structures is that the central linker of TnC is not observed to be helical in the cardiac Tn complex. This negative observation does not preclude the existence of analogous Ip•TnC interactions in the two protein crystals (see the discussion in Section, 3.2, p. 19). Cluster HCA4 is stabilized through interactions with TnT in only the skeletal structure. The magenta circles in Figure 3.2 highlight this variation between the two structures. Although this difference between the cardiac and skeletal structures has been identified previously [1] the present results suggest a potential functional role for this conformational switching, as developed below (Section 3.5.3, p. 37).

Of the crystallographically-visualizable residues, a subset of residues are seen to fall in conserved hydrophobic clusters that are also solvent accessible (Figure 3.2). For example, HCA clusters 3, 5, and 6 all contain hydrophobic residues that are not buried in either structure (although HCA6 is not visualized in the cardiac structure (Figure 3.2, right). It could be expected that this solvent exposure is dependent on the regulatory state of Tn, however, this does not seem likely given that the low-Ca^{2+}structure (PDB 1YVO, not depicted) shows virtually identical conformations of TnI and TnT. Notably, the TnI/TnT coiled coil is not differently stabilized in the two regulatory states of sTn, with the exception of the presence of the detergent molecule anapoe in the Ca^{2+}-activated skeletal Tn structure. The only substantial change in hydrophobic clustering between the high- and low-Ca^{2+}structures is the well-characterized Sp•NTnC interactions found only in the Ca^{2+}-activated state. Disregarding the complicating factor arising from the greatly different resolutions for these two structures (high-Ca^{2+}– 3Å; low-Ca^{2+}– 7Å), a number of solvent-accessible hydrophobic clusters are conserved without conferring any quaternary contacts. Either these residues stabilize the intrachain native interactions (perhaps promoting main chain hydrogen bonding through bulk solvent exclusion) or they are poorly stabilized, which may be indicative of an underlying structure-function role [87]. Of course, the structural models may not describe all of the relevant protein-protein interactions; for example, the T2 domain of TnT may also make contacts with tropomyosin [31].

3.4.2 Charge-hydropathy distributions

A number of physical properties of ID proteins have been identified [88]. ID proteins have low sequence complexity (entropy) and also have a high ratio of charged residues to hydrophobes [89]. The charge/hydropathy method is a good predictor of extended (or global) disorder, as found in maximally disordered proteins [84,90]. This approach is embodied in the FoldIndex algorithm [91], which calculates the local charge/hydropathy over a user-defined window size.[7]

The FoldIndex analyses for TnI are summarized in Figure 3.3. The software makes assignments of intrinsic structure or disorder based on a sliding window calculation of the charge-to-hydrophobicity ratio. To assess the dependence of the findings on window size, the window size was varied from the default setting of 51 to 25 and 5. All three sets of results are depicted in Figure 3.3: each of the human proteins in the sequence alignment is colored to show the results of the analysis, with red corresponding to residues assigned as having intrinsic stability and green representing disordered segments. The results for a window size of 5 are reflected in the text coloring; the

[7]FoldIndex [91] was run on the proteins with a a step size of 1 residue and window sizes of 51, 25, and 5 residues using the webserver (http://bip.weizmann.ac.il/fldbin/findex).

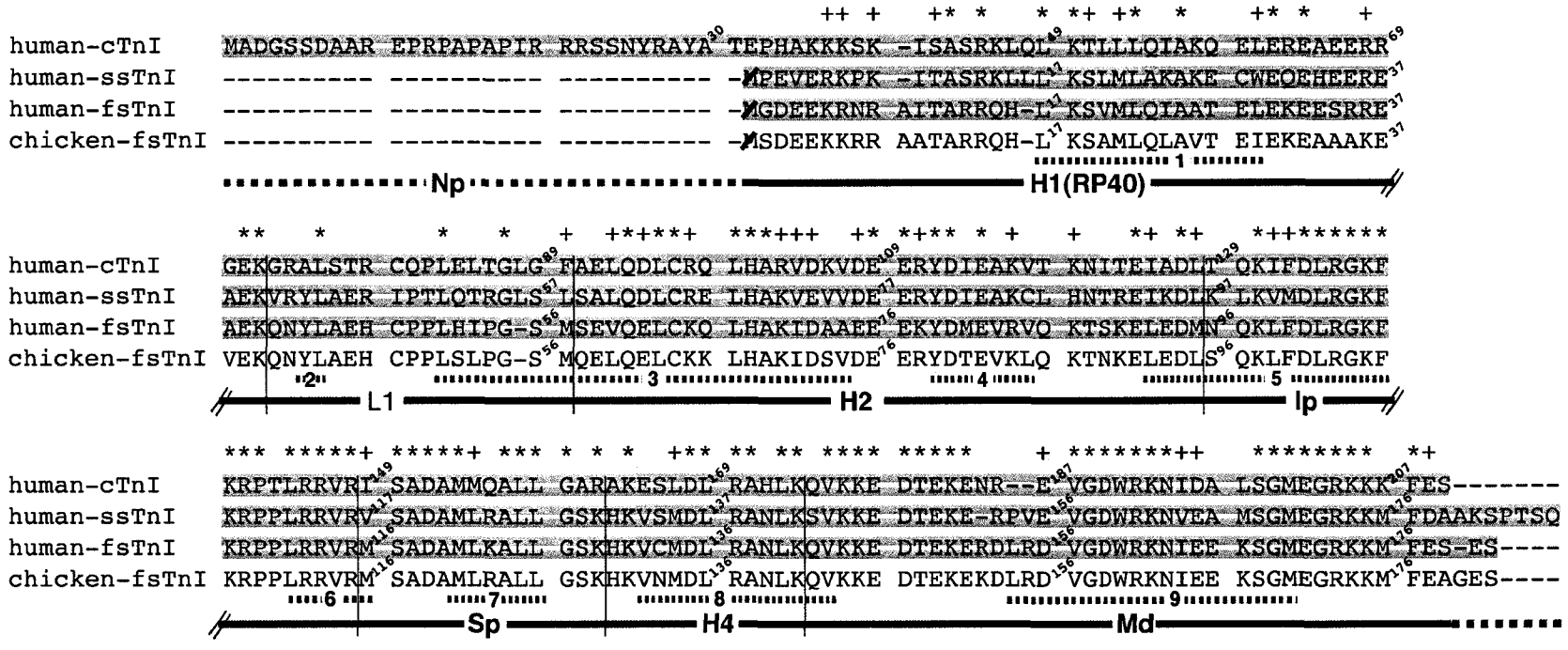


Figure 3.3: FoldIndex output for TnI isoforms. A sequence alignment of the human isoforms is shown; the numbering shown (for non-cardiac sequences) omits the initial Met. Identities (*) and conserved substitutions (+) are shown over the alignment. The text coloring reflects the FoldIndex assignment of folded (colored red) or unfolded (colored green) using a window size of 5. The color of the text overlining and underlining communicates the FoldIndex results for a window size of 25 or 51, respectively. Functional annotations are placed underneath the chicken-fast-skeletal TnI sequence, in black text. Conserved hydrophobic clusters are shown beneath the alignment as blue broken lines that span a cluster. The charge/hydropathy method indicates that there is a predominance of intrinsic disorder, varying with the isoform.This figure is reproduced with permission from Reference [11].

results for window sizes of 25 and 51 are shown as lines running on the top and bottom of the text, respectively. The coloring scheme was chosen so that it is easy to recognize regions where all three window sizes give concurrent predictions. When all three window sizes give matching predictions, the over-line, under-line, and text coloring match each other. The results for the 5 residue window size are the most variable so those results were selected for rendering the text colors.

The FoldIndex results show non-identical sequence–stability profiles for the three isoforms and window sizes. The calculations using window size of 5 (Figure 3.3, text coloring of human TnI sequences) showed more subtle variation between the isoforms, but an overall large degree of agreement. At this window size, FoldIndex gives a similar analysis to the HCA, as shown by the tendency for red-colored letters to lie over the blue dashes in Figure 3.3. Interesting points-of-difference are found in the region of the mobile domain. For example, considering the FoldIndex results for the window size of 5 (Figure 3.3, text coloring) the fast-skeletal TnI sequence 153-DLRDV-157 has the central Arg155 predicted to be ordered. In the slow-skeletal TnI sequence the same stretch is 153-RPVEV-157 with residue Pro154 and Val155 being predicted to be ordered. The cardiac sequence is gapped here, and the surrounding residues are not predicted to be ordered.

With a window size of 25 (Figure 3.3, over-line), the three isoforms had a maximal amount of agreement. Structured regions for all isoforms correspond to HCA1 (part of H1), the N-terminal half of HCA3 (part of H2), and Sp. The cardiac and slow-skeletal isoforms have a structured region predicted to lie just N-terminal to Ip that is not predicted for the fast-skeletal isoform. With the 51 residue window size (Figure 3.3, under-line), both of those isoforms had ordered regions predicted in the N-terminal amphiphilic helices (H1 and H2), and for the cardiac isoform, the Ip region was predicted to be ordered. The fast-skeletal isoform is seen to have the largest extent of predicted intrinsic disorder of the three proteins; with a window size of 51 (Figure 3.3, underline) only a single residue – S34 – is predicted to be ordered (this has no literal meaning). There are few regions sharing high sequence identities (sequence identities are shown in Figure 3.3 as black asterexes over the alignment) as well as large amounts of FoldIndex-predicted structure. These regions are readily rationalized in terms of the known protein-protein interactions and are usually HCA clusters. Based on the default (51 residue) setting (Figure 3.3, underline), the intrinsically structured regions of the human cardiac isoform correspond to residues 65–78, 93–114, and 131–148. Residues 65–77 are between HCA clusters 1–3. Residues 93–114 correspond to HCA clusters 3–5. Residues 131–148 include portions of clusters 5 and 6. The slow-skeletal and cardiac isoforms had more ordered regions than the fast-skeletal isoform, for all window sizes. The most highly conserved and structured region includes Sp.

The charge/hydropathy method is a good predictor of intrinsic disorder although the presence of binding partners (as is the case for coiled-coil-forming proteins) complicates the analysis [84]. In spite of this limitation, the interpretation of a large extent of disorder is strengthened given that the regions with a high local hydrophobic character correspond to the known tertiary/quaternary stabilizing contacts. FoldIndex is seen to only weakly predict ordered/disordered structural boundaries as shown by the sensitivity of those demarcations to window size. The applicability of the concept of domain boundaries to TnI is suspect, as much of the structure of TnI is contingent on binding to other proteins. Any subdomains in TnI (like Md) are small compared with FoldIndex's default win-

dow size (51). The integration of results from multiple ID predictors is advocated, even in routine applications such as domain boundary identification [84, 90].

3.4.3 Machine-learning-based predictions of intrinsic disorder

Intrinsic disorder prediction strategies also include machine learning approaches. In these approaches, a predictor is first trained on protein sequences known to correspond to ordered or disordered solution ensembles. The trained algorithm can then be applied to new data, which will be interpreted in the context of the training set. This biases the predictors towards predicting the flavor [92] of disorder prevalent in the training data. For example, the sequence composition of short disordered regions is different from that of long disordered regions [93, 94].

There are many potential predictors one could employ. Since each publication presenting a novel approach emphasizes its benefits, choosing between the predictors is a daunting task for a nonexpert. The CASP (Critical Assessment of Structure Prediction) competitions allow for objective comparision of the various prediction methods. In the CASP7 competition, intrinsic disorder predictors were queried with the same test set [95]. The test set was derived from proteins with known structures; disordered regions were defined as regions crystallized but not modeled or regions that were not well-restrained by NMR observations. The test set is heavily biased towards short disordered regions (long disordered regions are not amenable to structure determination). Predictors trained on data that is similar to the test set are expected to perform better in this competition.

In this study I used the VSL2B predictor [94], designed to locate both long and short stretches of ID. VSL2B contains three support vector machines (SVM)s arranged into two tiers [94]. The lower tier contains two SVMs that makes predictions based on two different window sizes. The predictors consider a number of features extracted from the sequence, including hydrophobicity and information content. The upper tier is a meta-predictor that makes decisions through consideration of the ouput from the lower tier SVMs. VSL2B is trained on a variety of manually annotated IDPs containing both short and long disordered regions [94]. The disordered regions were long (more than 50 residues) and were identified by various methods including crystallography, spectroscopies, and limited proteolysis. The ordered proteins in VSL2's training data consist of regions of crystal structures with low B-factors.

The neural-network based VL3E predictor [96] considers estimates of evolutionary divergence when assigning probabilities of intrinsic disorder. The training set of VL3E incorporates and extends that of VLSL2 [97]. VL3E is made of two other predictors; it reports the majority consensus between the two. One of the constituent predictors is trained on homologous sequences identified by sequence similarity, "... using homolgous sequences that diversify the training data and thus achieve better coverage of the attribute space of disordered proteins" [96]. The other predictor derives its probabilities of a given amino acid residue from PSI-BLAST profiles (position-specific scoring matrices generated from a multiple sequence alignment).

Another predictor using multiple sequence alignment is DISOPRED2 [98]. DISOPRED2 is a support vector machine that is trained on sequence alignments of proteins with known crystal structures. Disordered regions are defined as those absent from crystal structures. As it is trained on

the missing regions of crystal structures, DISOPRED2 is trained to assume that proteins are ordered with disordered fragments, which makes it predictive of a specific type of disorder.

3.4.4 Sequence-based estimations of local energy content: IUPred

Another predictor, IUPred [56], assigns a measure of local energy content to a region of a protein sequence. IUPred is empirically parameterized to the inverse problem – prediction of structure – an attractive strategy because assumptions as to the nature of ID are not imposed. Due to this theoretical basis [55] IUPred can potentially predict novel flavors of ID [84].

IUPred[8] predicts more subtle disorder profiles than the machine learning algorithms (Figure 3.4C). This predictor located all of the HCA clusters, even the C-terminal clusters 8 and 9. There is a short region in the C-terminus that is predicted to be ordered for all of the isoforms, comprising only residues 167 and 168 of the fast-skeletal sequence.

3.4.5 Comparison of predictions

All of the machine learning-based algorithms predicted extensive disorder in the termini of TnI, especially the N-terminus of cardiac TnI. As well, intrinsic disorder was predicted for some regions containing HCA clusters. The results for VSL2B (Figure 3.4A) are most directly comparable to the (51 residue window) FoldIndex profiles in Figure 3.3. For the slow-skeletal and cardiac isoforms, VSL2B predicts two structured regions corresponding to HCA clusters 3–7 (the IT arm, inhibitory peptide, and the switching region). Interestingly, VSL2B gives no hint of HCA1, which (as a peptide) binds to CTnC with high affinity ($2 \pm 1\mu M$) [99]. Portions of TnI are known to gain (either persistent or transient) stabilization from binding to other proteins, so the question of the length dependence of ID is paramount. The results imply that TnI isoforms can vary in how disorder at the termini is dampened through the formation of structure. I address the potential role of stabilizing binding partners below.

As shown in Figure 3.4B, VSL3E predicted the structured region corresponding to HCA clusters 3–7 for all three isoforms. It also did not indicate the existence of HCA1. The predictions from DISOPRED2 (Figure 3.4D) seem to be the strongest observation favoring the presence of structure. That being said, this program also implicated a large amount of intrinsic disorder in CTnI (all isoforms) and in cNp. The difference in ID between isoforms can vary from large (as in Figure 3.4A) to very small (Figure 3.4D). The VSL2B predictor [94] shows a large amount of variation between the isoforms (Figure 3.4A). The fast-skeletal isoform is predicted to be much more disordered than the other isoforms over the region C-terminal to H2 (fast-skeletal TnI_{58-181}, HCA clusters 3–9). Consistent with the FoldIndex results, cNp (the cardiac-specific region) is predicted to be the most highly disordered region of all the TnI isoforms. As shown in Figure 3.4B, the VL3E predictor [96] presents similar results to VSL2, but places the difference between isoform stabilities to be much smaller. The IUPred algorithm [56] scores most TnI regions as meta-stable (Figure 3.4C). In contrast to VSL2B and VL3E, IUPred attributes the most highly disordered C-terminus to the slow-skeletal isoform. This discrepancy is mirrored in the DISOPRED2 prediction (Figure 3.4D). DISOPRED2

[8] IUPred [56] was executed from http://iupred.enzim.hu/.

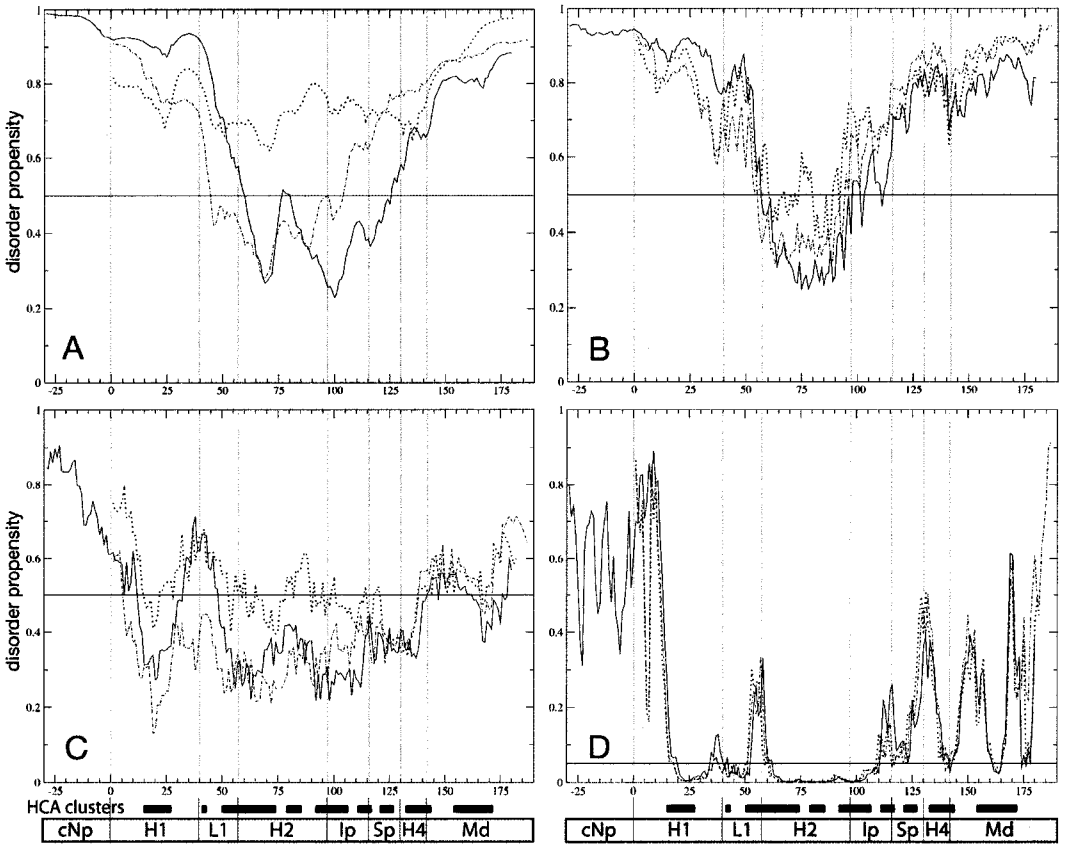


Figure 3.4: Intrinsic disorder (ID) predictions for the human isoforms of TnI. In all plots the solid line is cardiac isoform, the broken line is the slow skeletal isoform, and the dotted line is the fast-skeletal isoform. The horizontal line marks the threshold for the assignment of ID. The cardiac sequence has been shifted by 30 residues to align it with the others. Structural-functional annotations are found beneath the plot; region boundaries are demarked with vertical lines. Note that the alignment of the annotations with the plots is approximate owing to gapping in the actual alignments. See Figure 3.3 (p. 27) for a more accurate alignment of the hydrophobic clusters with the functional annotations. The assigned HCA clusters are demarked with thick black lines over the annotations. (A) The VSL2B predictor predicts a large amount of variation between the isoforms. (B) The VL3E predictor presents similar results, but places the differences between isoforms to be much smaller. (C) IUPred scores most TnI regions as meta-stable. (D) DISOPRED2 predicts the fewest differences between the predictors used, and the largest quantity of intrinsic structure. There is general consensus that the termini of all isoforms are intrinsically disordered. This figure is reproduced with permission from Reference [11].

is also unique in predicting the fewest differences in ID between the three isoforms; VSL2B gave the largest predicted differences between the isoforms.

3.4.6 Predicting relative propensities to intrinsic disorder

All predictors except DISOPRED2 (Figure 3.4D) suggest H2 to have more disorder in the fast-skeletal isoform. In other regions there is less consensus; the relative stabilities for H1 and L1 (comparing the isoforms) are similar comparing VSL2B and VL3E, but dissimilar from IUPred. Also of interest is the large amount of variation between the isoforms in the locale of Ip, with the cardiac isoform having the most ordered regions and the fast-skeletal being the most disordered. Excluding the DISOPRED2 results, the Ip region is predicted to be most disordered for the fast-skeletal isoform, intermediate for the slow-skeletal isoform, and most ordered for the cardiac isoform. There is discordance across the predictors as to whether or not the region is intrinsically ordered, but the N' HCA cluster of Ip (HCA5) is usually predicted to be more ordered than HCA6.

The VSL2B and VL3E (Figures 3.4A, 3.4B) predictors assign greater intrinsic disorder to H1 than to L1 or H2. IUpred (Figure 3.4C) reports a higher ID propensity in L1 than either H1 or H2. The DISOPRED2 (Figure 3.4D), IUPred, and VL3E algorithms predict a sharp transition in disorder propensity between H4 and Md. The VL3E output showed, at the C-terminus of TnI, only substantial differences between the skeletal and cardiac isoforms. Unlike the FoldIndex and VSL2B approaches, slow-skeletal TnI was not strongly implicated as being intermediate in disorder between the other isoforms. The relative ranking of the C-terminal disorder was similar to the VSL2B and VL3E predictors: the skeletal isoforms were more disordered than cardiac TnI. As DISOPRED2 is trained on crystal structures, and considers homologies when assigning ID propensities, this predictor is expected to show minimal differences between the homologous, crystallographically-visualizable Tn complexes.

The use of the disorder prediction algorithms to give relative rankings of the extent of intrinsic order, as attempted above, is somewhat over-ambitious. The algorithms are usually geared towards detection of disorder and not its quantification. Furthermore, there is no explicit theoretical relationship to correlate disorder propensities with ensemble heterogeneity. There is no ID predictor optimized for the prediction of relative disorder between proteins having a similar primary structure[9]. To illustrate, DISOPRED2 – the program giving the most similar disorder propensities for the three isoforms – is a support vector machine and is best-suited for binary categorization. This is reflected in its 5% probability threshold for assigning ID. Since the Md region (of known disorder [2,49]) is also highly conserved (Figure 3.3), the assumptions implicit in the DISOPRED2 do not appear to allow for discerning functional differences in ID between highly related proteins. Qualitative ranking of the extent of disorder in related ID proteins is beyond the scope of DISOPRED2. Furthermore, the training set of DISOPRED2 consists of crystal structures, so the contingent nature of TnI's structure may place it outside the scope of DISOPRED2. Similarly, the neural-net-based VL3E is optimized for the delineation of disordered/structured domain boundaries, and not for quantification of ID. VSL2B is also comprised of support vector machines, however, it is trained on different data and

[9] AK Dunker, personal correspondence

also has a two tier configuration that confers length-dependence to its disorder propensities. The small sequence variations that are detected by FoldIndex (with a small window size, see the text coloring of Figure 3.3) may also cue VSL2B to assign different lengths of a given disordered region. The VSL2B results are seen to be most directly comparable to those of FoldIndex, probably because of the correspondence between long disordered regions (which VSL2B is optimized to detect) and global unfolding (which the charge/hydropathy method detects) [84].

IUPred stands out as the only algorithm describing TnI as meta-stable because the score hovers around the P=0.5 threshold. In light of the physical basis for IUPred's parameterization, this likely reflects TnI's nascent structure that is stabilized by other troponin protomers or by actin. IUPred is not trained on a manually annotated set of ID proteins, unlike the machine learning-based predictors, and therefore has the capacity to predict previously uncharacterized classes of ID [84]. Although the underlying model of IUPred [55] is too coarse grained to allow a high degree of discrimination, the relative degrees of stability shown in Figure 3.4 likely reflect relative degrees of intrinsic stability [10]. The qualitative correspondence between the physically-based (FoldIndex, IUPred) and the machine learning-based (VSL2B, VL3E, DISOPRED2) algorithms substantiate the existence of isoform-dependent variations in the intrinsic disorder of TnI.

3.4.7 Intrinsic disorder and Ca^{2+}-regulation

The main result from this study is that TnI isoforms appear to vary with respect to their intrinsic stabilities, especially in the region spanning L1–Ip. The results have a large degree of consistency considering the broadly different theoretical approaches underlying the alternative predictors used here. I conclude that isoform variations in the intrinsic stability of L1, the IT arm, and Ip may confer some of the tissue-specific properties of TnI.

The pervasive disorder of the C-terminus of TnI is corroborated by all prediction approaches used here. This disorder may be the underlying cause for a poor consensus between existing structural models of this region. The 52 kDa core human cTn complex [45] shows around 30 residues of cardiac TnI after Sp as a single, extended α helix. This structure is stabilized by the presence of inter-unit cell contacts, but the plausibility of this conformation is substantiated by the presence of HCA8. The core skeletal Tn complex [1] shows the final 12 residues of chicken-fast-skeletal TnI in an irregular conformation. Reference to the NMR structure of chicken skeletal Md [2] (PDB 1VDI, 1VDJ) features a large amount of solvent-exposed hydrophobes. HCA clearly identifies the main globular regions, however most of the β bulge is seen to be stabilized exclusively by hydrogen bonding. HCA cluster 9 is likely the folding nucleus for this subdomain, as is substantiated by the previous demonstration of long-range, inter-sidechain NOEs in this cluster [2], despite pervasive structural disorder [2, 49]. As well, the presence of a tryptophan in HCA9 suggests a relatively high degree of intrinsic structuring, as Trp is heavily biased towards appearing in structured regions [84, 90, 100]

[10] Zsuzsanna Dosztányi, personal correspondence

3.5 Mechanistic significance of intrinsic disorder in troponin I

Individual steps in the function of troponin are protein-target associations/dissociations. If the target is also a protein, and this protein has a high propensity to unfolding, the immediate question becomes *does the target fold prior to or subsequent to binding?* The work of Hilser and Thompson (Section 3.1, p. 19) shows that under the assumption that folding preceeds binding, disorder-to-order transitions can underly allostery. Others have shown plausible kinetic benefits from folding subsequent to the formation of the encounter pair.

3.5.1 Disorder and protein-target association

The earliest proposal (that I can identify) that intrinsic disorder promotes protein-target association is due to Brian Pontius [101] following experimental work in collaboration with Paul Berg [102]. He argued that disordered polymers could promote the initial binding of two DNA strands through long-range electrostatic effects. His argument mostly shows how the rate of association is greatly accelerated through the long-range electrostatics and increases in collisional cross-section.

The fly-casting mechanism

This idea was rediscovered in the research group of Peter Wolynes in an analogous context: DNA–protein association. In their hands this phenomenon received its current moniker: fly-casting [103]. As Pontius argued, the collisional cross-section is increased for a disordered protein. But this does not immediately lead to a kinetic enhancement, because the proportion of productive collisions does not necessarily remain constant as the protein becomes disordered. Shoemaker and others [103] inject the energy landscape concept into this scenario, as a mechanism whereby the initial encounter pair of (unfolded) protein and target diffuses through the folding funnel towards the conformation of the bound state, instead of diffusing through space. This randomly directed search is heavily biased towards the native fold; as is the folding of any ordered protein, however the native fold is only realized as a higher-order structure. The fly-casting mechanism affords a kinetic enhancement for binding compared to scenarios where both targets are ordered prior to binding.

Kinetic paths for conformational switching

The fly-casting mechanism for troponin was first proposed by Blumenschein and others [49] to assign a mechanistic function to the mobile domain of TnI. This proposal differs from the Md-clamping mechanism of Wakabayashi and coworkers (Section 3.3, p. 22). In the Md-clamping mechanism the mobile domain binds to actin and further stabilizes the inactivated state. I challenge this idea from consideration of the process of inhibition. In the activated state the switch region of TnI is bound to TnC, leaving the mobile domain in an extremely hydrated environment. When the intracellular [Ca^{2+}] drops, the dissociation of TnI and Ca^{2+} from TnC become more strongly coupled events — the presence of the switch peptide drastically slows the dissociation of Ca^{2+} from TnC. Rather than describing the initiation of inhibition as the reverse of the activation steps, I proposed that fly-casting allowed for a mechanistically necessary event to be more readily promoted [9, 10].

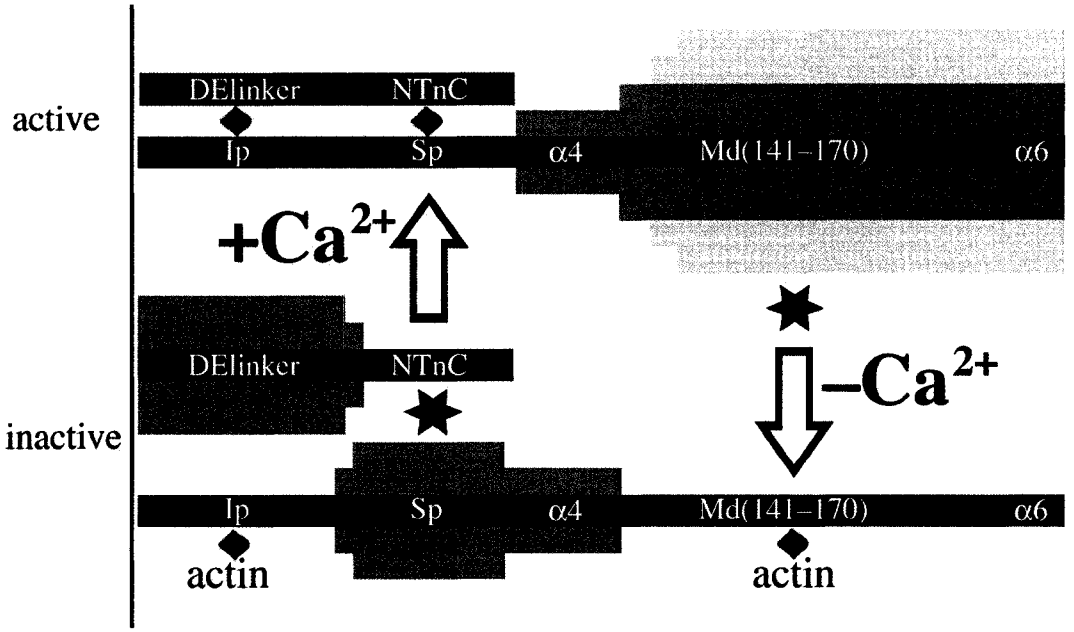


Figure 3.5: Summary of flycasting mechanism for troponin. Disordered regions of CTnI nucleate the interconversion of Tn's two equilibrium states. The primary structure of TnI (or TnC) is depicted in horizontal rectangles. Approximate locations of binding sites are shown with diamonds. Proposed binding nucleation sites are depicted with stars. Lighter shading depicts weaker structure and correspondingly larger equilibrium spatial fluctuations. In the contraction-inhibited state (bottom), Md and Ip both bind actin. Upon Ca^{2+}-activation, Sp nucleates the release of CTnI from the thin filament. In the Ca^{2+}-activated state (top), Md undergoes the largest spatial fluctuations, and initiates the return of Ip to the thin filament upon inhibition. This event is promoted by long-range electrostatic interactions and flycasting. This figure is reproduced with permission from Reference [9].

Fly-casting activity would allow the mobile domain to assist with the dissociation of the switch region, avoiding the unparsimonious notion that a disordered region stabilizes the bound state. My proposal is depicted in schematic in Figure 3.5 and explicitly in Figure 3.6. My specific contribution here is the notion that the structural cascades leading to inhibition are not the reverses of the pathways of activation. Each step *is* reversible from a microscopic point of view, but the functional transitions of troponin do not occur through those pathways.

Isoform-specific variation in fly-casting regions

Under the flycasting mechanism, isoform-specific variations in the ID of CTnI could confer some of the isoform-specific regulatory properties. Crossbridge cycling (the mechanism of force generation) occurs on a time scale approximately 10 times slower than the switching of the Tn-Tm system [104]. This implies that for force kinetics to be modulated by Tn-Tm, the equilibrium distributions of the Ca^{2+}-activated closed and open states (Figure 2.2, p. 15) must vary between isoforms [44, 104, 105] with many of these differences being attributable to variations in Tm [43]. Multiple lines of inquiry converge at this interpretation [106] although Maytum and others find the closed/open ratio to vary

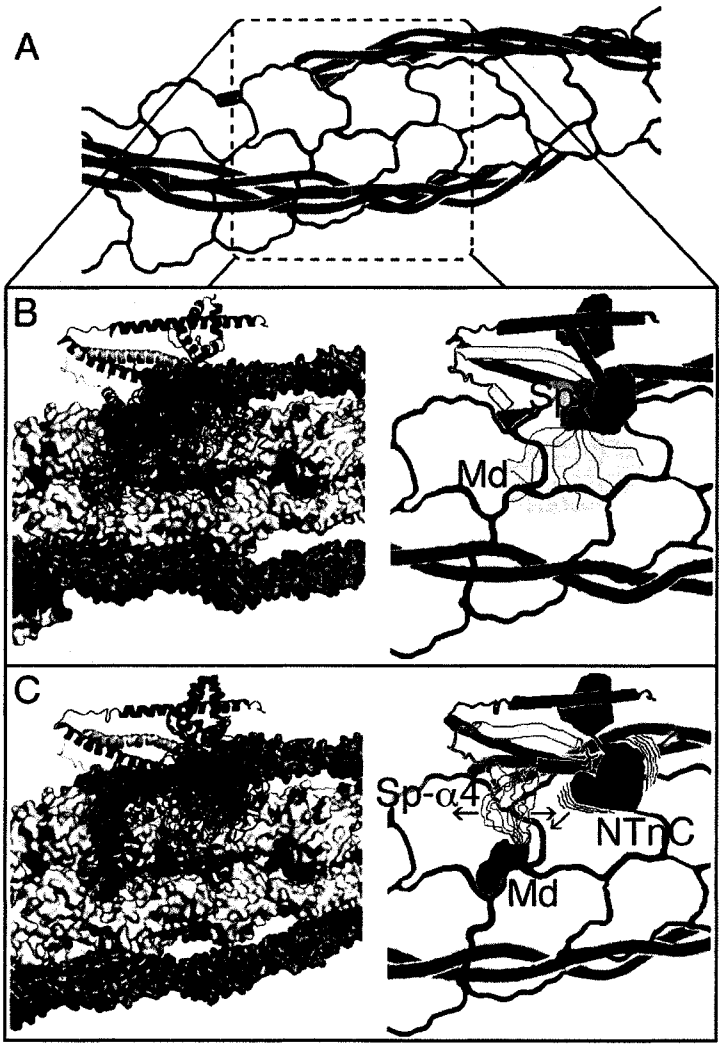


Figure 3.6: Depiction of flycasting mechanism for troponin. The details of how this model was constructed are in Appendix A. **A.** The thin filament, depicting the edges of actin monomers with black lines. The inhibited (purple) and activated (red) states of Tm are illustrated with thick colored lines tracing the backbone coordinates. The approximate location of Tn in both states is demarked with the callout box. **B.** The activated state (excluding TnT-T1) is shown with a combination of schematic and space-filling representations at left; the same orientation is shown in cartoon form at right. In both images actin is white, Tm is grey/black, TnT-T2 is yellow, TnC is red, and TnI is blue. The mobile domain of TnI (Md) extends large distances, 'fishing' for its binding partner. The conformational disorder in Md is under-represented here due to the use of a fairly small (20) family of NMR structures. **C.** The contraction-inhibited state. The motions of NTnC are conveyed by superimposing the NMR structure of sTnC onto CTnC portion of the Tn core domain. Md is bound to actin along with the inhibitory region of TnI, with the Md-Sp region being transiently structured. Approximate coordinates of the switch region of TnI (Sp) are indicated with blue dashes. Ca^{2+} binding to TnC allows for subsequent Sp binding, releasing CTnI from actin. This figure is reproduced with permission from Reference [9].

between tissue isoforms for only the low-Ca^{2+}state [107]. The results weakly support the hypothesis that functional differences between isoforms originate in variations in the intrinsic structure of Md, but the effect seems too subtle to address with the tools used here. Based on the present results, I cannot attribute the longer relaxation time for cardiac muscle expressing slow-skeletal TnI [78] to a lower rate of association between Md and actin as would be expected under the flycasting mechanism.

3.5.2 Intrinsic structure in the switch region

As shown in Figure 3.5 (left), the switch region (Sp), modeled as disordered in the actin-bound state, undergoes a disorder-to-order transition as it binds Ca^{2+}-activated NTnC. The FoldIndex results (Figure 3.3) suggest that Sp is intrinsically structured which is inconsistent with most other ID predictions (Figure 3.4). As such, there is some ambiguity as to the intrinsic structure of Sp. It is evident that any intrinsic disorder of Sp originates from a mechanism other than a local compositional bias towards charged residues (which is the only mechanism considered by FoldIndex). Although Tripet and others concluded that Sp entropically hinders the binding of TnI to actin [29] this does not necessarily indicate that Sp remains disordered when bound to actin. So the actual degree of intrinsic structure in Sp, and its potential to vary across isoforms, remains ambiguous.

Of some interest is a key histidine present in the skeletal isoforms (H130 in human fast-skeletal TnI) that is thought to confer pH tolerance [76,108,109]. Recent work by Westfall and Metzger [109] shows that multiple substitutions in the H4 helix can result in similar phenotypes, adding subtlety to the 'histidine button' model of pH resilience in skeletal TnI. The present results do not clarify the structural-dynamical mechanism underlying the isoform-specific nature of pH sensitivity because the results are equivocal regarding the intrinsic structure of Sp.

3.5.3 Isoform-dependent stabilities of the coiled-coil region

Structural and enthalpic changes can be correlated semiquantitatively [110]. Comparing the cardiac and fast-skeletal isoforms, the large variations in predicted ID for H1 do not appear to impact their three dimensional structures (Figure 3.2). When incorporated into Tn, this region is similarly enthalpically stabilized in the two isoforms. Under the assumption that the two isoforms have similar entropy change of incorporation into the Tn complex, the fast-skeletal isoform is seen to have greater residual entropy in the complex[11]. A lone interaction between TnT and TnI is observed in the skeletal core complex that is not observed in the cardiac structure (Figure 3.2). This interaction is between V83 of TnT (the T2 domain, as numbered in 1YTZ) and Y79 of TnI. The difference between the two structures leads to the TnT-T2 interhelical angle varying by more than 30°. Assuming the difference is functionally significant, and considering the present results, this interaction could enhance the conformational stability of the IT arm in response to structural changes in Tm-TnT. TnT may constrain TnI differently between regulatory states, as depicted at the top of Figure 3.5 with large amplitude conformational fluctuations in H2 in the inactive state, that are dampened in the activated state. As depicted in Figure 3.5, these fluctuations would stabilize the inactive state. If

[11] This argument also makes assumptions about solvation energetics being similar for the two complexes.

this model is accurate, the ID predictions suggest that the skeletal isoform is relatively disordered in the inactivated state, which would be consistent with it being more strongly inhibited, or "more off", than for cardiac troponin [107]. As the fast-skeletal isoform is less strongly activated by myosin binding compared with the cardiac isoform [16, 111], the greater intrinsic disorder of fsTnI reported here directly supports the hypothesis that the cardiac system has a stronger mechanical coupling of Tm•TnT with TnI, as developed in [16].

3.5.4 The inhibitory region in the context of activation

The conventionally defined inhibitory region of TnI (Ip) does not completely span HCA clusters HCA5 and HCA6 (see Figure 3.3). Those two clusters are interrupted by the conserved sequence elements KRPP. Reference to Figure 3.2 shows that the two HCA clusters of Ip have different roles. HCA5 almost exclusively contacts itself and TnT, whereas HCA6 mostly makes contacts with TnC. Assuming the conformation of TnT is influenced by the regulatory state of Tm, this implies that the regulatory transitions of the thin filament can modulate the affinity of HCA5 for TnT. Since HCA5 is proximal to HCA6, this also explains how the apparent Ca^{2+}sensitivity of Tn can be modulated by the regulatory state of the thick filament, as is further developed below.

The N-terminus of TnC is often called 'the regulatory domain' because it reversibly binds $Ca^{2+}$and Sp [30], although various drugs are thought to modulate regulation through interactions with the 'structural' C-terminal domain [79, 112]. The present results suggest a structural mechanism for troponin that can be modulated by such drugs. The disposition of the IT arm in the cardiac structure could lead to weaker HCA5•TnT interactions, compared with the fast-skeletal structure (Figure 3.2). This structural cascade is depicted in the bottom of Figure 3.7 as a series of forces transmitted from Tm to the Ip•TnT•TnC substructure (Figure 3.7, bottom, green arrows on TnT). The destabilization in the TnI-TnT coiled-coil region resulting from cascade would propagate along TnI C-terminally, perhaps facilitated by the rigidity of the conserved diprolyl moiety bridging HCA5 and HCA6, to weaken the HCA6•TnC interactions, thereby destabilizing the central DE helix of TnC. This would increase the rotational mobility of NTnC, promoting the dissociation of Sp (including HCA7) [1], and facilitating the transfer of Ip to the actin surface. I emphasize that Ip has been originally defined in terms of its ability, as a cyanogen bromide cleavage fragment, to *inhibit* muscle contraction; perhaps redefining Ip to span $sTnC_{92-116}$ will allow for more informed functional studies of activation.

3.5.5 Structural transduction of phosphorylation in cardiac TnI

The cardiac-specific N-terminal extension (cNp) contains phosphorylation sites for protein kinases A, C and D [32]. This region's ID is a consensus across all of the predictors used here. Intrinsic disorder is frequently identified in the locale of phosphorylation sites [113]. Signal transduction 'hub proteins' – those that integrate multiple cascades – have been recently spotlighted as having a high propensity towards ID [58]. As cNp is a known point of integration for multiple signalling systems [32, 74], cardiac TnI fits well into that paradigm. Furthermore, TnI has been shown to bind promiscuously, as it can bind to and influence a number of membrane associated proteins [114–116].

ID proteins are often found in moonlighting (functionally promiscuous) roles [57]. Given the antagonistic effects of phosphorylation from protein kinase A versus C [32], it is likely that the downstream effects of phosphorylation do not principally involve a disorder-to-order transition, unlike the SERCA-phosopholamban interaction [117]. It is more likely that the intrinsic disorder in this region facilitates promiscuous molecular recognition by protein kinases. To assess the possibility of a disorder-to-order transition being imparted by phosphorylation, the human cardiac TnI sequence was modified (Ser23Asp and Ser24Asp) to mimic a pseudophosphorylation experiment. No change in predicted ID could be detected using FoldIndex or IUPred. Simply increasing the quantity of negative charge in cNp is insufficient for producing a large change in its propensity towards ID as assessed with the techniques used here.

Recently, an NMR structure of the cNp region was published [118]. As in the case of the mobile domain structure [2], the NMR observations that were used to restrain the structure also directly disclose the high degree of disorder in cNp (for example, near-random-coil secondary chemical shifts) [9]. Our results imply that the structure of cNp is nascent at most. Howarth and others also perform a computational pseudophosphorylation experiment of cNp and indicate that secondary structure predisposition of cNp responds to this manipulation. This finding is inconsistent with our observation that pseudophosphorylation does not induce structuring in cNp. Howarth and others present a docking model for PKA-phosphorylated cNp. I suggest that the cNp-Tn interaction may be driven by cumulative electrostatic interactions, as has been recently demonstrated to have significance in the context of the phosphorylation of intrinsically disordered proteins [119]. Perhaps the ID of cNp, combined with its variable phosphorylation levels, allows it to function more like a rheostat than a toggle switch, as has been demonstrated previously in another biological system [120]. The lack of induced structuring in cNp is consistent with a recently proposed mechanism for the effects of PKA phosphorylation, in which the binding of switch peptide is mediated through electrostatic interactions [121].

3.6 Conclusion

The results demonstrate the prevalence of intrinsic disorder in all tissue isoforms of TnI. Detailed interpretation of the results shows that most, but not all, of the ID can be attributed to a high ratio of charged residues to hydrophobes. There is strong evidence that the isoforms vary in ID but the various predictors used do not concur with respect to the relative extent of ID, or even the rank order of ID between isoforms (considering a conserved region). The results motivate the development of predictors targeted to problems of this type. The cardiac specific N-terminal extension is predicted to be globally disordered using all prediction methods. I note that most regions of TnI have been described as structured to varying degrees, in varying scenarios, and that the results indicate the inherent predisposition of this structure towards dynamics.

The main finding in this chapter is that a 'structural' region of TnI (a region having established structure–function relationships) has a high degree of ID and there is a large predicted difference in ID between the isoforms, as exemplified by (but not limited to) the H2 region. I contextualize this finding in a model of thin filament regulation (Figure 3.7). In this model, forces can be propagated

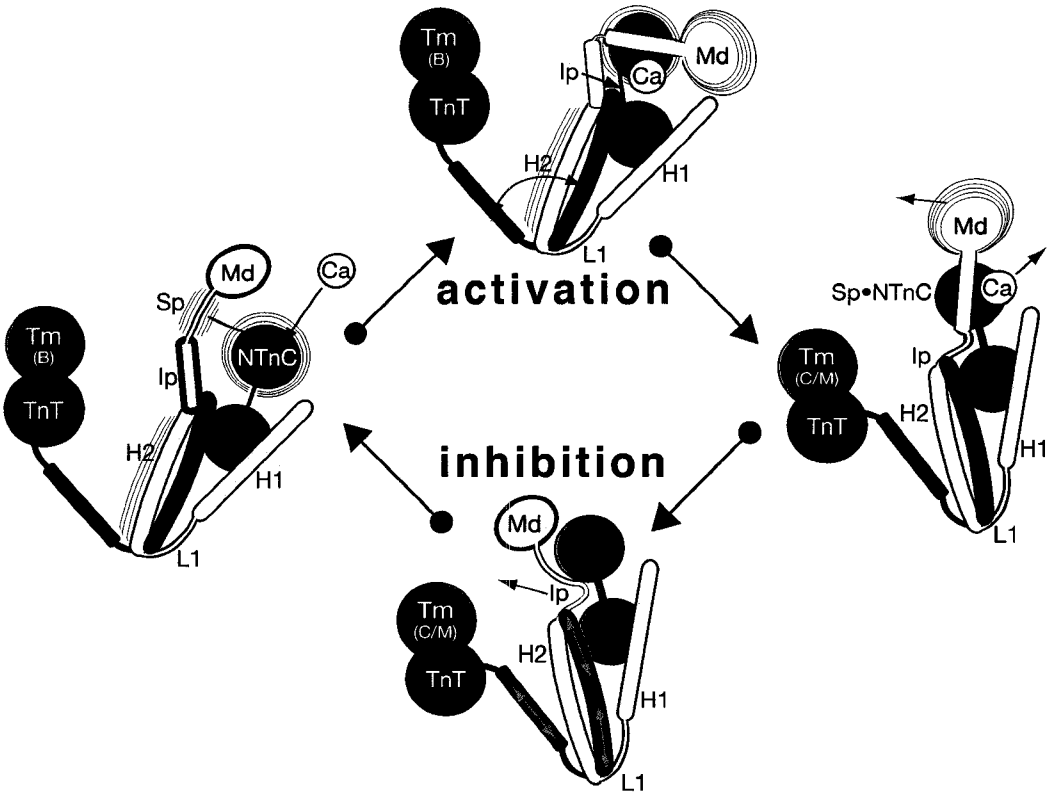


Figure 3.7: A model for troponin regulation of muscle contraction. Troponin is depicted in red (TnC), yellow (TnI) and blue (TnT). Tropomyosin is shown in black, with labels denoting the B or C/M states. Actin-bound states are outlined in magenta (thicker outlines correspond to tighter binding). Small black 'wiggle lines' indicate structural dynamics (only for TnC and TnI). Small arrows depict the structural fluctuations that initiate (nucleate) protein-protein or protein-Ca^{2+}interactions; the magenta arrows depict the binding of TnI to actin. The inhibited, low-Ca^{2+} state (left) is depicted with the low-Ca^{2+}model described in [9], borrowing largely from the model of Pirani and others [41]. The switch peptide (Sp) is shown as disordered prior to Ca^{2+}-activation; NTnC is folded but highly dynamic because the DE helix is flexible. Ca^{2+}-activation, initiated by Ca^{2+}-binding to NTnC and the subsequent binding of Sp, followed by the binding of Ip (progressing clockwise from the left image to the right) produces the activated state (right). Here the Tn complex from [9] is depicted, showing Ip interacting with TnT and TnC, and the mobile domain lacking structure. A Ca^{2+}-independent conformational transition in TnT (cyan arrow, at top) is shown dampening the intrinsic dynamics of TnI, thereby stabilizing the activated state. Deactivation is depicted as in [9] with Md initiating return to the actin surface through flycasting [103]. I suggest that dissociation of Ip is promoted by the structural cascade depicted with green arrows, communicating conformational fluctuations in Tm-TnT to Ip. This figure is reproduced with permission from Reference [11].

from tropomyosin to the 'Ca^{2+}-sensing' substructure through two cascades. One cascade (Figure 3.7, top) begins with a conformational transition in TnT that is demonstrable through comparison of the fast-skeletal and cardiac core Tn complex structures. This transition changes the interhelical angle in TnT, allowing for a new intermolecular contact with TnI. The inherent conformational dynamics of TnI are then dampened – I speculate the dampened state corresponds to the activated state. A second cascade (Figure 3.7, bottom) is a mechanically-mediated destabilization of the TnI-TnT interaction culminating with the dissociation of TnI residues L92–L95 (numbering for the fast-skeletal isoform) from TnT. I agree that, based only on the current results, assigning high amplitude conformational dynamics to TnI, in the troponin complex, remains tentative. The agreement between previous experimental results (like references [29, 49, 68]) and the present results is certainly encouraging. I hope to spur more explicit reconciliation of dynamical and structural aspects of the troponin mechanism, and of the significance of the multiple isoforms of troponin protomers.

Chapter 4

The binding of W7

In this chapter[1] I summarize a number of investigations into the process of ligand binding. The ligand of interest is W7, an inhibitor of striated muscle contraction, and a lead compound for the development of cardiotonic drugs.

4.1 Cardiotonic drugs

Cardiotonic drugs are clinically useful when forestalling heart failure [79, 112]. These agents modulate the contractile force of the heart. Levosimendan, a drug in clinical use, is thought to work through direct sensitization of cTnC to Ca^{2+} [122]. Such a drug mechanism is therapeutically desirable because intracellular $[Ca^{2+}]$ is not perturbed, preserving the regulation of other Ca-based signalling pathways [112, 122].

The calmodulin antagonist W7

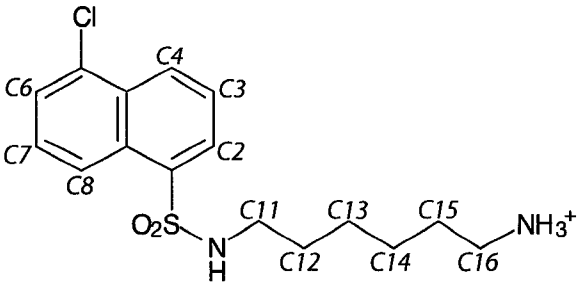


Figure 4.1: The chemical structure of W7 (N-[6-aminohexenyl]-5-chloro-napthalenesulfonamide). Standard numbering for carbon atoms is shown in italics. This figure is reproduced with permission from Reference [12].

[1] Content for this chapter comes from [12] and [13] as well as unpublished material. I will delineate the specific contributions throughout the chapter.

W7 is a well-characterized calmodulin (CaM) antagonist, widely used to dissect Ca^{2+}-mediated signaling pathways. Previous studies have reported that W7 binds CaM with a range of per-domain stoichiometries (1–6), and to sites of weak (K = 200 μM) and strong (11 μM) affinities [123]. The binding of W7 to chicken TnC•4$Ca^{2+}$was also described with mixed binding modes, with the stronger K being 25 μM [123]. This compound inhibits muscle contraction in skinned muscle fibers and in perfused hearts; cTnC has been implicated as the mechanistically-relevant target for W7 in the muscle fiber [124]. W7 and its analogs have two functional moieties: an aromatic 'head' and an amine 'tail' (Figure 4.1). Both moieties are required for activity. NMR spectroscopy has been employed towards the rationalization of W7's mechanism of action. A previously published structure of CaM•4Ca^{2+}•2W7 visualized a single molecule of W7 binding to each domain [125]. An NMR study on J8 [126], a closely related W7 analog, reported multiple orientations of J8 within the binding pocket of CaM (multiple binding 'poses'), with no definable conformation for the tail moiety. The study of the binding of W7 to cTnC, a protein sharing ~50% identity with CaM, may clarify the mechanistic picture as TnC and CaM have structurally analogous drug (and target peptide) binding sites [127].

4.2 NMR characterization of weak binding: a case study

Monica Li wanted to study the binding of W7 to the Ca^{2+}-bound C-terminal domain of cardiac TnC (cCTnC•2Ca^{2+}). She prepared an NMR sample containing 1.76 mM ^{15}N-labeled protein and a stock solution of 130 mM W7 in DMSO-d6. She incrementally added the W7/DMSO and acquired $\{^1H,^{15}N\}$ correlation spectra. Such NMR spectra correlate amide ^{15}N nuclei with covalently bonded ^{1}H. These experiments monitor, at atomic resolution, changes in the chemical environment, and therefore reflect structural changes in a protein due to the binding of a ligand.

The overlayed spectra are shown in Figure 4.2 Many of the protein NMR signals change their frequencies as a function of [W7]; they are *titratable*. As the titratable signals move continuously from initial to final points, without significant line broadening, they are in the NMR *fast exchange limit* which has important implications for the correct modeling of binding.

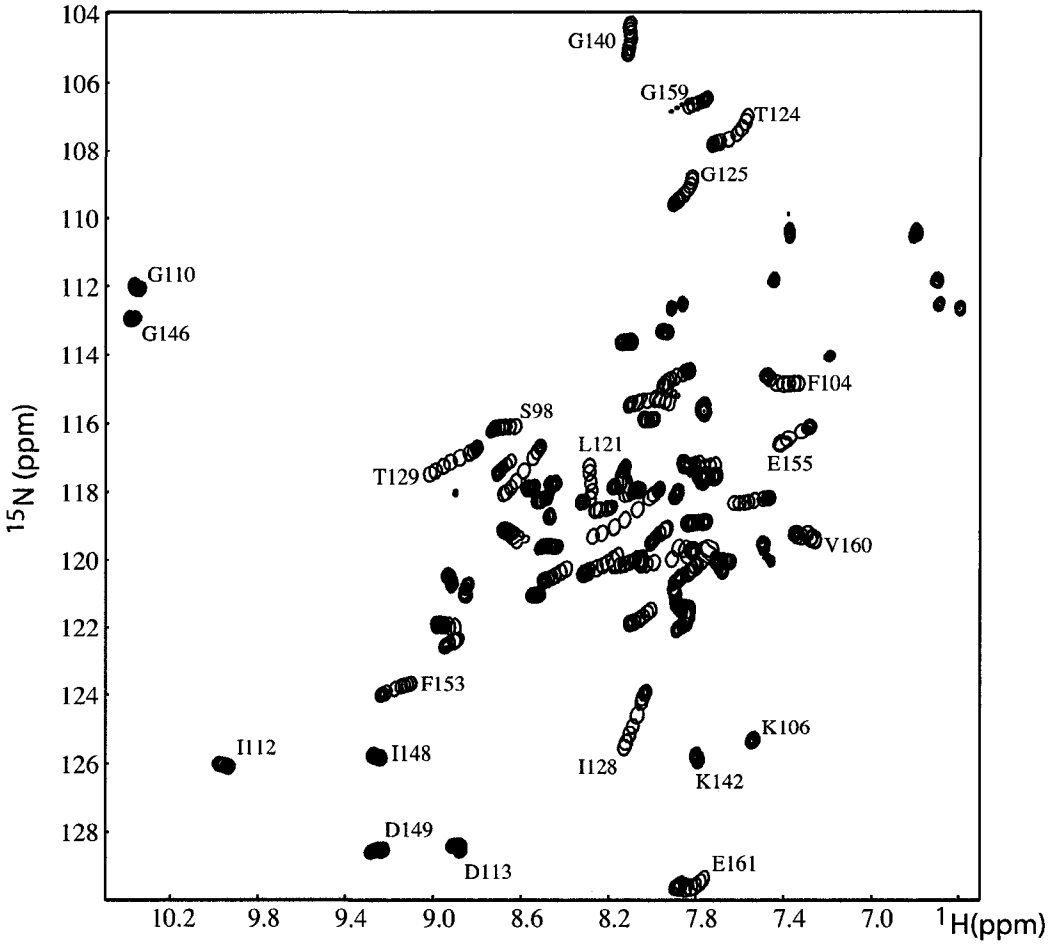


Figure 4.2: The binding of W7 to cCTnC•2Ca^{2+}monitored with $\{^{1}H,^{15}N\}$-correlation spectra. This is Monica Li's data. The initial step in the titration is shown with filled-in contours; subsequent increments are shown as open circles. The peaks move continuously over the titration, without any substantial line broadening being apparent. A subset of $\{^{1}H,^{15}N\}$ crosspeaks shift non-linearly, including F104, L121, T124, G125 and E161. Most of these residues are located in the linker region between the EF and GH subdomains. The non-linearity is pronounced and evident near the midpoint of the titration. This figure is reproduced with permission from Reference [12].

4.2.1 Fast exchange conditions

The term *fast exchange* has a quantitative meaning in NMR due to the possibility of chemical exchange occurring at rates comparable with the spectral frequencies [128]. For example, looking at the titratable signal originating in Thr129 (Figure 4.2) it moves about 0.5 ppm in the ^{1}H dimension over the experiment. The frequency of the unbound protein is readily known (ν_a); the frequency of the bound state (ν_b) can be approximated if we assume that binding goes nearly to saturation. For this discussion assume that Thr129 shifts 0.5 ppm in the ^{1}H dimension. On the spectrometer this data was acquired on (500 MHz ^{1}H), that means that the free and bound states differ by $0.5x50^8/10^6$ Hz, or 250 Hz.

The observed frequency of the NMR signal is a function of the population of the free and bound states. Consider when the free and bound states are equally populated. If the chemical exchange process equilibrating the free and bound states occurs at an exchange rate (k_{ex}) of 250 Hz, a given NMR signal *chemically* changes its peak frequency on the same timescale that is needed to resolve the two signals. In effect, the signal's frequency becomes scrambled. As a result, no signal is resolved. This is called the intermediate exchange limit.

If the exchange rate is much less than 250 Hz, and the free and bound states are equipopulated, then two NMR signals are seen of equal intensity. This called the *slow exchange limit*, where the following condition is satisfied:

$$2\pi|\nu_a - \nu_b| \gg k_{ex} \tag{4.1}$$

In the example of T129's behavior, the peak is not observed at its two frequencies corresponding to the free and bound state, but instead moves smoothly between the free and bound states. This indicates that the exchange rate is fast compared with the timescale required to resolve the two states:

$$2\pi|\nu_a - \nu_b| \ll k_{ex} \tag{4.2}$$

The peak smoothly moves from ν_a to ν_b with its position reflecting the population of the two states. This shows that, in the case of Monica's data, the binding event (when interpreted as a two-state process) occurs with an exchange rate much larger than 250 Hz.

The fast exchange condition has an additional and fortuitous implication: the peak *assignments* of the free state (given in Reference [129]) are easily followed to the bound state. This makes the assignment of the bound state trivial. If the titration increments are sufficiently finely spaced, automated assignment of the bound state is possible using current peak detection algorithms, for protein spectra of sufficient resolution. Weak binding is therefore amenable to semiautomated analyses and parallel structure determination studies.

4.2.2 Linear perturbation model of binding

Binding can be characterized with most spectroscopies, as the primary observable is nearly always weakly coupled to the structural changes induced in the environment of the probe. The following treatment applies to most spectroscopies, including NMR when the chemical exchange occurs in the

fast exchange regime (Section 4.2.1). Binding then produces linear perturbations in the observable, which can then be modeled as:

$$\delta_i = \delta_\infty f_i + \delta_0(1 - f_i) \tag{4.3}$$

Where δ_i is the observable signal at titration increment i, δ_0 is the initial (unperturbed) signal, f_i is the fractional population of the bound state at increment i,[2] and δ_∞ is the signal at complete ligand saturation. The variables δ_0 and δ_i are assigned directly from observations. An *incremental change* in signal intensity can be defined as the difference between the initial signal and the one at increment i:

$$\Delta\delta_i = \delta_i - \delta_0 = \delta_\infty f_i + \delta_0(1 - f_i) - \delta_0 \tag{4.4}$$

For the simple equilibrium $\mathrm{L} + \mathrm{P} \leftrightarrow \mathrm{P} \bullet \mathrm{L}$, an analytical expression can be derived relating the total quantities of protein and ligand to the observables, with the dissociation constant K and δ_∞ being free parameters. This is shown below (Equation 4.6).

4.2.3 Extracting a titration curve from 2D data

At each stage in the titration the ligand induced change in chemical shift, $\Delta\delta$, can be modeled as the vectorial magnitude of the $\Delta\delta$ along each spectral dimension. For a $\{^{15}\mathrm{N},^{1}\mathrm{H}\}$-HSQC-monitored titration, each two-dimensional datum may be interpreted as the scalar product of two one-dimensional displacements:

$$\Delta\delta_i = \sqrt{\Delta\delta_{H,i}{}^2 + (\Delta\delta_{N,i}C_N)^2} \tag{4.5}$$

$\Delta\delta_i$ is the observable at increment i and C_N is sometimes included to scale the $^{15}\mathrm{N}$ chemical shift to a comparable value to that of its attached proton. Parenthetically, the choice of the scaling factor C_N varies in the literature.[3] When the binding is modeled based on one signal, the choice of inclusion of the scaling factor has little impact on the outcome.

4.2.4 The single-site binding model

I distinguish between between binding *sites* and binding *signals* although in some cases it is natural to equate the two. *Signals* are NMR peaks that monitor a binding event whereas *sites* are the actual interfaces that contribute to the stabilization or destabilization of the bound state. Not all NMR signals that monitor binding necessarily form the binding site, nor must all binding sites contain titratable NMR signals. An open question remains as to the best procedure for defining a binding site based on the observation of a population of titratable signals having variable degrees of coupling to binding.

[2] If the total protein concentration, $[P_0]$, is equal to $[P] + [P \bullet L]$, then $f_i = \frac{[P \bullet L]}{[P_0]}$.

[3] Sometimes this takes a value of 1 (no scaling), sometimes 1/5 to correct for differences in the large chemical shift anisotropy of $^{15}\mathrm{N}$, other times it corrected for its different gyromagnetic ratio ($C_N = \gamma_{15_N}/\gamma_{1_H}$) [130]. The choice of C_N largely impacts the site specific variation in the magnitude of $\Delta\delta$. If the chemical shift change in the two dimensions is proportional (shifting linearly) then there is no difference between the alternative choices of C_N.

The dissociation constant for single site binding can be modeled with an analytical expression for the change in the response variable:

$$\Delta\delta_i = \frac{\delta_\infty - \delta_0}{2P_i}\left((P_i + L_i + K) - \sqrt{(P_i + L_i + K)^2 - 4P_iL_i}\right) \tag{4.6}$$

where $\Delta\delta_i$ refers to the total chemical shift perturbation for increment i, δ_∞ and δ_0 are the factors defined in Equation 4.3 (the former is a parameter, the latter an observed value), K is the dissociation constant, L_i is the total concentration of ligand, and P_i is the total concentration of protein. This equation models single-site, reversible binding [131].

4.2.5 Nonlinear least squares

Predicting chemical shifts is a difficult endeavor but the theory described by Equation 4.6 describes *perturbations*, $\Delta\delta$s, of readily observed initial chemical shifts. Given a model that predicts an observation from known (manipulated) variables, the model can be fit to actual observations through the method of least squares. An error function is defined as the sum of the squared residuals over each titration increment[4]:

$$\epsilon(K, \delta_\infty, \{\Delta\delta\}_i) = \sum_i \left(\Delta\delta_{obs,i} - \Delta\delta_{pred,i}\right)^2 \tag{4.7}$$

The experimental data set is represented as $\{\Delta\delta\}_i$ – the set of $\Delta\delta$ values for all i titration increments – which is understood to be accompanied by known protein and ligand concentrations. The parameter values that give a minimum value for the error function are taken as the best fit. The use of the square of the residual, instead of its absolute value, is the main reason why least squares approaches lack robustness to outliers. For a Gaussian distribution of errors (and therefore a low frequency of outliers) the least squares solution is optimal.

It is desirable to obtain a measure of the error in the fitted parameters. For the linear case, the quality of a least squares estimate can be summarized with another parameter, the standard deviation. For the nonlinear case there is no general expression of the error of the fit, and workers most often report a value based on a linearization of the fit.

The least squares method is optimal if the errors in the data are normally distributed (the criterion of *random errors*) and if the magnitude of error is uniform over the data (the criterion of *homoskedacity*). (This is formally limited to linear models, as described by the Gauss-Markov theorem.) Small deviations from homoskedacity are acceptable but should the error be clustered toward the maximum or minimum values the least squares fit becomes highly biased towards the extrema. This lack of robustness is the usually cited reason for the supposed intractability of fitting pre-asymptotic binding data. This criticism will be further addressed below (Section 4.2.6, p. 49).

[4]Numerous statements in this chapter derive from standard statistical textbooks, including [132] and [133].

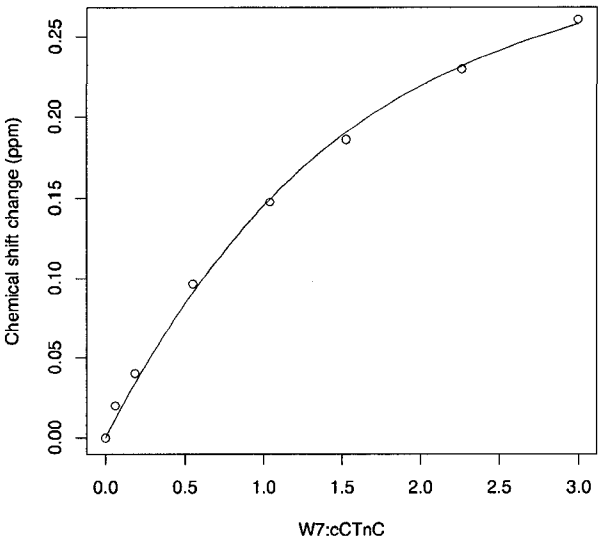


Figure 4.3: The binding signal from Thr129 plotted as a function of the molar ratio of total W7 to total cCTnC. The fitted parameters are: $K = 1.54 \pm 0.30$ mM, $\delta_\infty = 0.362 \pm 0.024$ ppm. The error ranges are determined through the default 'nls' function in R, which uses an iterative linearization procedure.

4.2.6 One-signal analysis

Monica's data show a plethora of NMR signals monitoring the process of binding. Two well-resolved signals originate in residues Ile128 and Thr129. These signals both shift extensively over the titration, and so have high signal-to-noise (the noise resulting from imprecision in the quantification of peak position). Thr129's 2D-chemical shift perturbations were quantified with Equation 4.5 and are plotted in Figure 4.3. A least-squares fit against the single-site binding model was performed[5]; the fitted line is shown with the data in Figure 4.3.

The least squares fit gives two parameter values, K and δ_∞. Both have physical meaning, but only the value of K characterizes the binding *site*. The value of δ_∞ characterizes the binding *signal*; as defined in Equation 4.3, this is chemical shift that would be expected if the titration completed until endpoint. This can be seen by 'zooming out' the plot in Figure 4.3 until the asymptotic part of the curve is visualized. See Figure 4.4. Complete saturation of the binding sites was not accomplished, consistent with other binding studies of naphthalenesulfonamides [123, 125, 126]. W7's limited solubility in water largely excludes the possibility of saturating the binding sites at the observed binding affinities and stoichiometries.

As studies of sparingly soluble ligands are commonly attempted, and weak binding allows for large concentrations of free (solvated) ligand, many investigations are necessarily confined to the pre-asymptotic phase. Accurate fitting of weak binding data is frequently confounded by the limited sampling of the asymptotic phases of the binding curve. This has traditionally been considered to be a basic limitation to the study of weak binding.[6]

[5]The analysis is in Appendix C.

[6]Recently the orthodoxy have been challenged [134]; standard methodologies such as ITC remain tractable when characterizing weak binding, albeit in limiting circumstances.

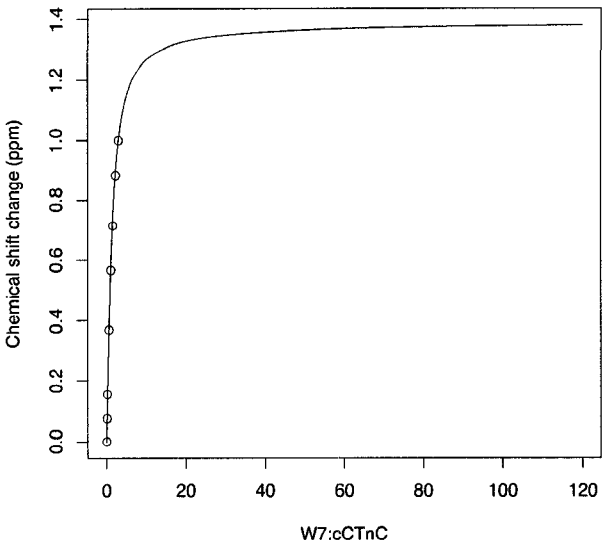


Figure 4.4: The same data and analysis as in Figure 4.3 but with the axes scaled to show the asymptotic part of the curve. The graphical meaning of the two parameters is: K is the X-coordinate of inflection point at 50% of maximum signal change; δ_∞ is the maximal signal change. The data do not adequately sample most of the fitted surface.

Signal heterogeneity

Results due to Cui and coworkers, characterizing an equilibrium between calmodulin and a target peptide, illustrate an attempt to reconcile these issues [130]. Individual analysis of each of the binding signals affords a wide range of Ks. The standard error for the reported Ks, based on the least-squares fits, are approximately $\pm 0.03\text{x}10^{-5}$M; the Ks were reported as a range $1.2 - 8.8\text{x}10^{-5}$M. The error in the fit is therefore much lower than the variation in fitting different signals. Treating the reported Ks as a random sampling, the Student's confidence interval is $3.7-7.4\text{x}10^{-5}$M which is smaller than their reported range. Clearly, the combination of highly precise single-site fits, with a large range of fitted Ks, posed interpretative obstacles. The authors report that although a set of sites have the largest $\Delta\delta$s these signals were not considered because "their flexible motion effect may cause overestimation" [130]. While I applaud the transparency of Cui and others, the theoretical basis for excluding these signals is poorly motivated.

Sorsa and others, studying the binding of levosimendan to cardiac TnC (cTnC), write [135]:

> "All titration data supported our conclusions that the effective binding site for levosimendan is in the cNTnC and that cTnI blocks the C-domain binding of levosimendan. It should be noticed that we did not characterize an exact binding site since we observed chemical shift changes on several amino acids of the N-domain. The affinity constants calculated for residues in different parts of the domain is, therefore, different (from 200 μM upward) and an exact K value for the affinity of levosimendan to the cNTnC cannot be given. In addition, many amino acids could be involved in small structural changes due to eventual non-specific low-affinity binding site of levosimendan on cTnC."

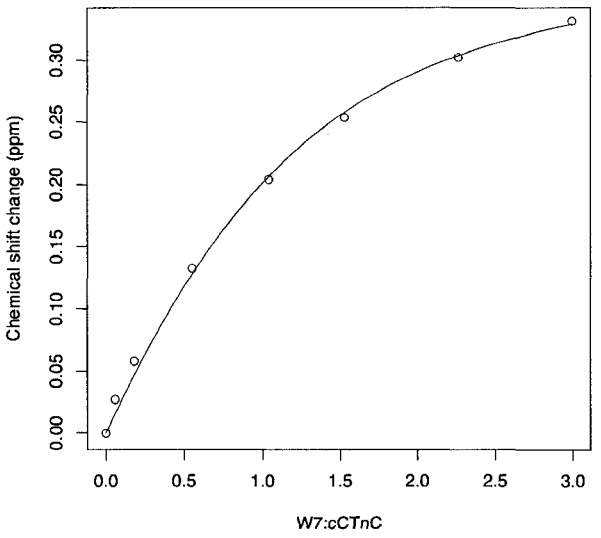


Figure 4.5: The binding signal from Ile128 plotted as a function of the molar ratio of total W7 to total cCTnC. The fitted parameters are: $K = 0.942 \pm 0.19$ mM, $\delta_\infty = 0.414 \pm 0.022$ ppm.

4.2.7 Parallel one-signal analyses

This analysis can be repeated for the other titratable signals. Ile128 has a very similar response to Thr129 in Figure 4.2. For Ile128 I have repeated the same analysis as for Thr129 and plotted the results in Figure 4.5.

The fitted parameter K for the two signals *ought* to agree because they are both reflecting the same binding event to the same site. But the error ranges reported alongside the fitted parameters are insufficiently small for the two parameters to agree to the level of statistical significance. The lower bound for K 'according to' the signal of Thr129 is 1.24 mM, which is still larger than the upper bound of K 'according to' the signal of Ile128, 1.03 mM.

I elaborate upon two aspects of this problem. Firstly, it seems the error ranges are not properly being estimated. That issue is eclipsed by the larger question of *how could the value of K vary so extensively when considering two signals of similar signal-to-noise?* One answer is that there are two fitted parameters, and we expect one parameter to be the same, but the other to be site-specific. Normalization is a procedure which allows one to assign δ_∞ to a global value, so that the two fits can be directly compared. This is explained following a brief discussion of error.

4.2.8 Sources of error

NMR provides many observations that differ only in their incorporation of random error. For example, the binding of a ligand larger than a single atom is likely to influence more than a single chemical shift. The global binding constant is a function of all of the microscopic processes that give rise to it (although these need not be explicitly considered). NMR signals also perturbed by secondary conformational changes (processes not directly coupled to ligand binding) will appear to be cooperatively activated by binding and will show more complicated titration behavior. As such the random error in a chemical shift, in the context of a titration experiment, arises from technical limitations such as the accuracy of peak detection and spectral resolution, as well as structural limi-

tations such as how strongly the multiple signals are coupled to processes that confound the binding signal.

There is necessarily a large degree of error in the X-coordinate of a titration curve owing to the imprecisions in gravitometric and volumetric methods. Considering a popular micropipettor, the Gilson series, even optimal scenarios would have an error of ~5% error in a volume delivery. The error propagates along the titration, so that the cumulative error in the last point is very large. This is unfortunate because the last points are influential in a least-squares fit, and because the last points are therefore considered to be key to estimating the location of the asymptotic phase.

Importantly, titration error cannot play a part in producing *any* of the site-to-site variations in the fitted parameters. All of the titration signals result from the same titration and therefore incorporate *identical* contributions from titration error. The variations in the local K determined from fitting Ile128 and Thr129 could be from imprecisions in the experiments and algorithms used to define the titration curves, but not from differences in how they monitor f_i. For two residues that have a similar signal-to-noise, monitoring the same phenomena, there is no reason for their fitted Ks to vary, save for disagreement with the underlying model.

4.2.9 Normalization of signals

The linear perturbation model of binding does not assign physical significance to the value of δ_{∞}. But, in the context of the single-site binding model, *two signals described by Equation 4.3 should shift the same, in proportion to their total displacement.* I would like to illustrate some insights that can follow from careful analysis of local and globally fitted parameters in this situation.

Starting with Equation 4.4, I add labels to indicate local and global variables:

$$\Delta\delta_{ir} = \delta_{ir} - \delta_{0r} = \delta_{\infty r} f_i + \delta_{0r}(1 - f_i) - \delta_{0r} \tag{4.8}$$

Where the index i denotes *titration increment* as before and r denotes a *residue index* to mark a variable as local (residue-specific — characteristic of a particular signal). The variable f_i does not have a local definition — it is the ratio $\frac{[P \bullet L]}{[P_0]}$ which is a global quantity.

At some later point in the titration, say titration point j, the incremental signal change is modeled as $\Delta\delta_{jr} = \delta_{\infty r} f_j + \delta_{0r}(1 - f_j) - \delta_{0r}$. The ratio of the signal intensities at two titration points, what I meant above by *proportional* change in signal, is given by:

$$\begin{aligned} \frac{\Delta\delta_{ir}}{\Delta\delta_{jr}} &= \frac{\delta_{\infty r} f_i + \delta_{0r}(1 - f_i) - \delta_{0r}}{\delta_{\infty r} f_j + \delta_{0r}(1 - f_j) - \delta_{0r}} \\ &= \frac{\frac{\delta_{\infty r} f_i}{\delta_{0r}} - f_i}{\frac{\delta_{\infty r} f_j}{\delta_{0r}} - f_j} \\ &= \frac{f_i}{f_j} \end{aligned} \tag{4.9}$$

This shows that the normalized chemical shift change for a particular signal (a ratio of two local variables) is a function of globals. So, if one normalizes the data (for a given signal, dividing each

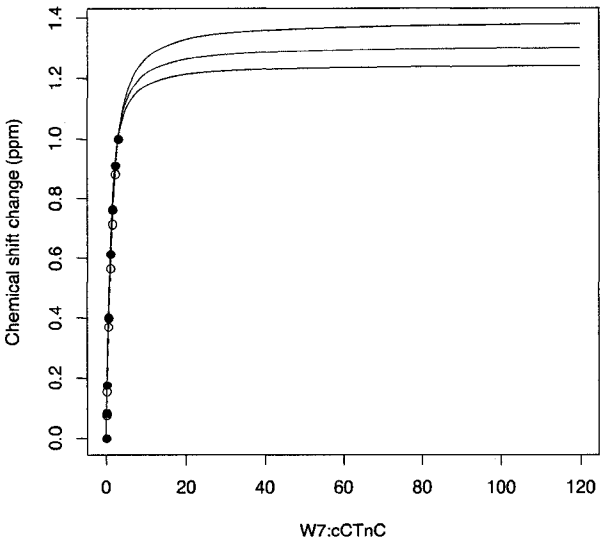


Figure 4.6: The binding signals for the two signals Ile128 and Thr129 are shown with their local fits to the single-site binding model (black curves) showing different asymptotic behaviors after normalization. By globally fitting the two signals (fitting them to the same values of K and δ_∞) the curve in blue is obtained.

incremental δ_i by the last value) they would expect the values of δ_∞ to agree between different residues. By repeating the least-squares fitting of the data after normalization, δ_∞ can be fit to a global parameter. By changing the local δ_∞ to a global one, a degree of freedom for each residue is eliminated from the analysis. As shown in Figure 4.6, the global fit (blue curve) interpolates between the twin local fits (black curves). I propose that the multiple observations afforded by high resolution NMR allow for sufficient sampling power to determine, with accuracy and precision, the dissociation constant from the pre-asymptotic phase of the binding curve. My presentation here does not successfully generalize my position, but I aim to convince the reader of its validity in this situation. An obstacle to generalizing such a position from the examination of real data is that, until this point, the applicability of the single-site binding model (Equation 4.6) has been assumed. This assumption is now shown to be erroneous.

4.2.10 Breakdown of the single-site binding model

Ile128 and Thr129 are arguably the two most concurrent signals in the titration shown in Figure 4.2 (p. 44). These amide groups are proximal in the primary structure, and they are similarly perturbed due to binding. In Figure 4.7 I show how inclusion of one more signal, from Leu121, completely confounds the analysis.

Examining the raw data for Leu121, in Figure 4.2, shows this residue to shift in two different vectors over the titration. At low concentrations of ligand, the residue shifts primarily upfield in the proton coordinate. A sharp inflection point is seen near the midpoint of the titration, over subsequent points the signal shifts primarily upfield in the nitrogen coordinate.

4.2.11 Towards a new binding model

The titration behavior of Leu121 immediately *proves* a breakdown in the linear perturbation model described by Equation 4.4. No one-to-one binding event will produce two distinct spectral perturba-

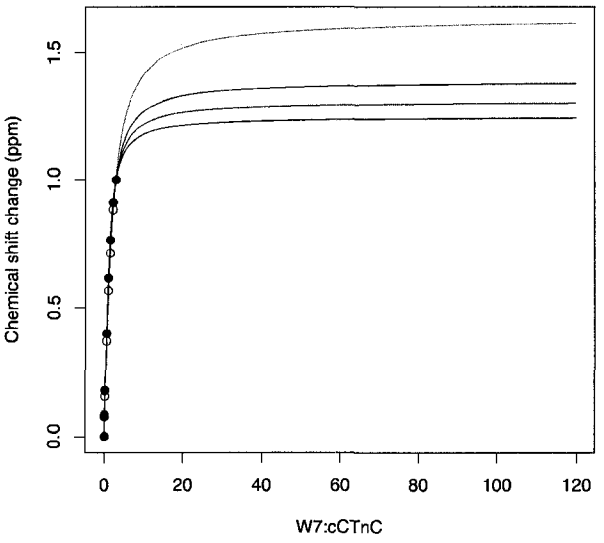


Figure 4.7: Adding the normalized data for Leu121 immediately suggests problems with the current analysis. The new global fit with three residues (green curve) is completely outside the local fits for Ile128 or Thr129 (black curves) or the previous global fit from Figure 4.6 (blue curve). The new fit suggests that binding is proceeding to a completely different endpoint than the previous analysis showed. Re-examining the raw data for Leu121 (Figure 4.2) shows this signal does not shift monotonically over the titration, confirming a breakdown of the single-site binding model (Equation 4.6) or the linear perturbation model (Equation 4.4).

tions at different phases in the titration. Other signals, for example those from Thr124 and Phe104, recapitulate the titration response of Leu121. The signal response changes nonlinearly with increasing concentration of the bound state. A parsimonious interpretation is that there are multiple binding events occurring, one strong binding even being primarily monitored at low concentration of ligand, and a second, weaker binding event dominating the signal at high concentrations. In this section I outline some alternate explanations before opting to continue with the parsimonious analysis in the subsequent section.

As previously discussed, interpreting curved HSQC trajectories due to a titration as two-site binding can present interpretative difficulties [126]. For example, the HSQC resonance for Leu121 shifts predominantly in the ^{1}H dimension over the first increments of the titration ($\sim$ 0.5 ppm upfield), and nearly exclusively in the ^{15}N dimension ($\sim$ 0.75 ppm upfield) over the final increments (see Figure 4.2). These changes are difficult to reconcile with an interpretation of an enhanced local magnetic field at Leu121 due to orientations normal to W7's aromatic moiety. Such a rationale does not account for the apparent separation of the invoked local fields throughout the titration, as detected simultaneously by backbone amide nuclei. When expressed in ppm, the chemical shift should scale only with the perturbations to the local magnetic environment. The two nuclei of an amide group are separated by $\sim$1 Å and are therefore, to a first approximation, in similar magnetic environments. ^{15}N, due to its large chemical shift anisotropy, is more sensitive to different conformations than ^{1}H, so the distribution of curved HSQC trajectories likely maps a [W7]-dependent conformational change rather than a direct ligand-mediated chemical shift perturbation [126].

Other factors could account for the apparent two-state equilibrium implied by the titration data for cCTnC•2Ca^{2+}•W7. The solvation/desolvation of millimolar concentrations of W7 may significantly impact the observed chemical shift changes. The substantive changes in solvent composition over the titration are due to the presence of W7 (accompanied by the DMSO used to solvate it). Naphthalene is extremely hydrophobic, making W7 amphiphilic. This introduces the potential complication of phase equilibria such as aggregation of W7. If aggregated W7 interacts with cCTnC,

this could also account for the inflection points in the trajectories of 2D-NMR crosspeaks.

Enumeration of all of the potential confounding processes is crushingly overambitious! Just considering the solvation/desolvation of W7 requires multiple equilibria in which water participates. For example, the W7 can repartition from its DMSO stock into bulk solvent:

$$\mathrm{W7 \bullet DMSO_d + (p+q)H_2O \leftrightarrow W7 \bullet (H_2O)_p + DMSO_d \bullet (H_2O)_q}$$

Solvated W7 could aggregate:

$$\mathrm{aW7 \bullet (H_2O)_r \leftrightarrow W7_a \bullet (H_2O)_{(r-s)} + sH_2O}$$

Hydrated W7 and hydrated DMSO could also be interacting:

$$\mathrm{bW7 \bullet (H_2O)_t + DMSO_e \bullet (H_2O)_u \leftrightarrow bW7 \bullet (H_2O)_{t-v} \bullet DMSO_{e-f} + vH_2O + fDMSO}$$

There are many potential confounding processes. A detailed description of the solution state equilibria necessitates the consideration of water activity, an effort that is difficult to justify as the rigorous description of mechanism will only be applicable to the NMR titration itself. A binding model with an appropriate underlying form can still provide insights into the underlying phenomenon.

4.2.12 The two-site, sequential binding model

A two-site, sequential binding model (P+L⇌P•L⇌P•L_2) [131] allows for a precise fit to these data. This model features two dissociation constants, $K1$ and $K2$, and two values of the scale parameter, $\delta 1$ and $\delta 2$. A global fit to this binding model was attempted [12]. The global $K1$ and $K2$ were determined as follows. A nonlinear least-squares fit was performed for each site, leaving only the values of $\delta 1$ and $\delta 2$ free, and varying the values of $K1$ and $K2$ over a sampling grid. The residual sum of squares (χ^2) for each site was summed to give a global value ($\Sigma\chi^2$). The values of $K1$ and $K2$ that gave the lowest $\Sigma\chi^2$ are identified as the global parameters.

Figure 4.8 shows the values of $K1$ and $K2$ producing the smallest global $\Sigma\chi^2$. The analysis determines that $K2$ takes minimal values (4 x $K1$ in this model) [131]. This indicates that the microscopic binding constants for the competitive binding site are of comparable magnitude to the primary site. Figure 4.9 shows the site specific fits to the global binding constants. Examining the distribution of nonlinear trajectories (paths through 'HSQC space') in Figure 4.2, one can qualitatively localize the secondary binding events to the region bridging the two EF hands present in cCTnC. This region experiences the largest chemical shift perturbation over all portions of the titration.

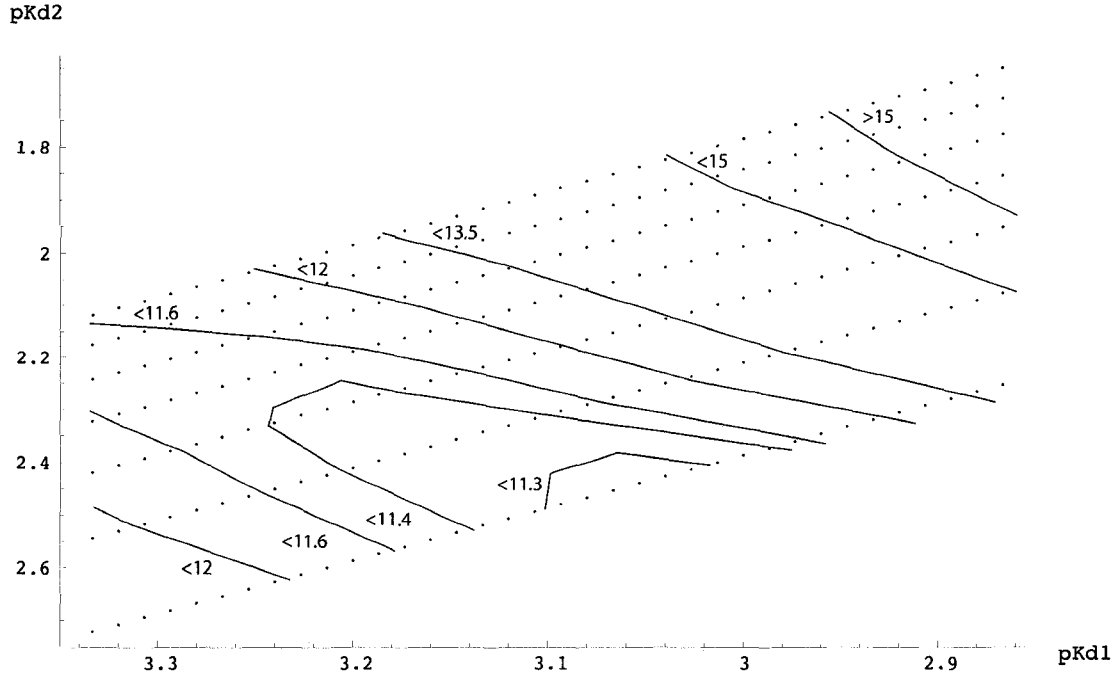


Figure 4.8: Determination of global Ks ($K1$ and $K2$) for a two-site, sequential binding model [131]. The global χ^2 ($\Sigma\chi^2$) is plotted as a function of p$K1$ (horizontal axis) and p$K2$ (vertical axis). The $\Sigma\chi^2$ surface is sampled at 252 points. Contour lines show levels of constant global χ^2; the global χ^2 is multiplied by 1000 to improve readability. The global χ^2 minimum occurs with $K1$ in the range $10^{-3.04}$–$10^{-3.07}$ (0.85–0.91 mM), and $K2$ = 4.001 x $K1$ (3.40–3.65 mM). The group of 6 points in the lowest contours have a $\Sigma\chi^2$ <0.01112. This figure is reproduced with permission from Reference [12].

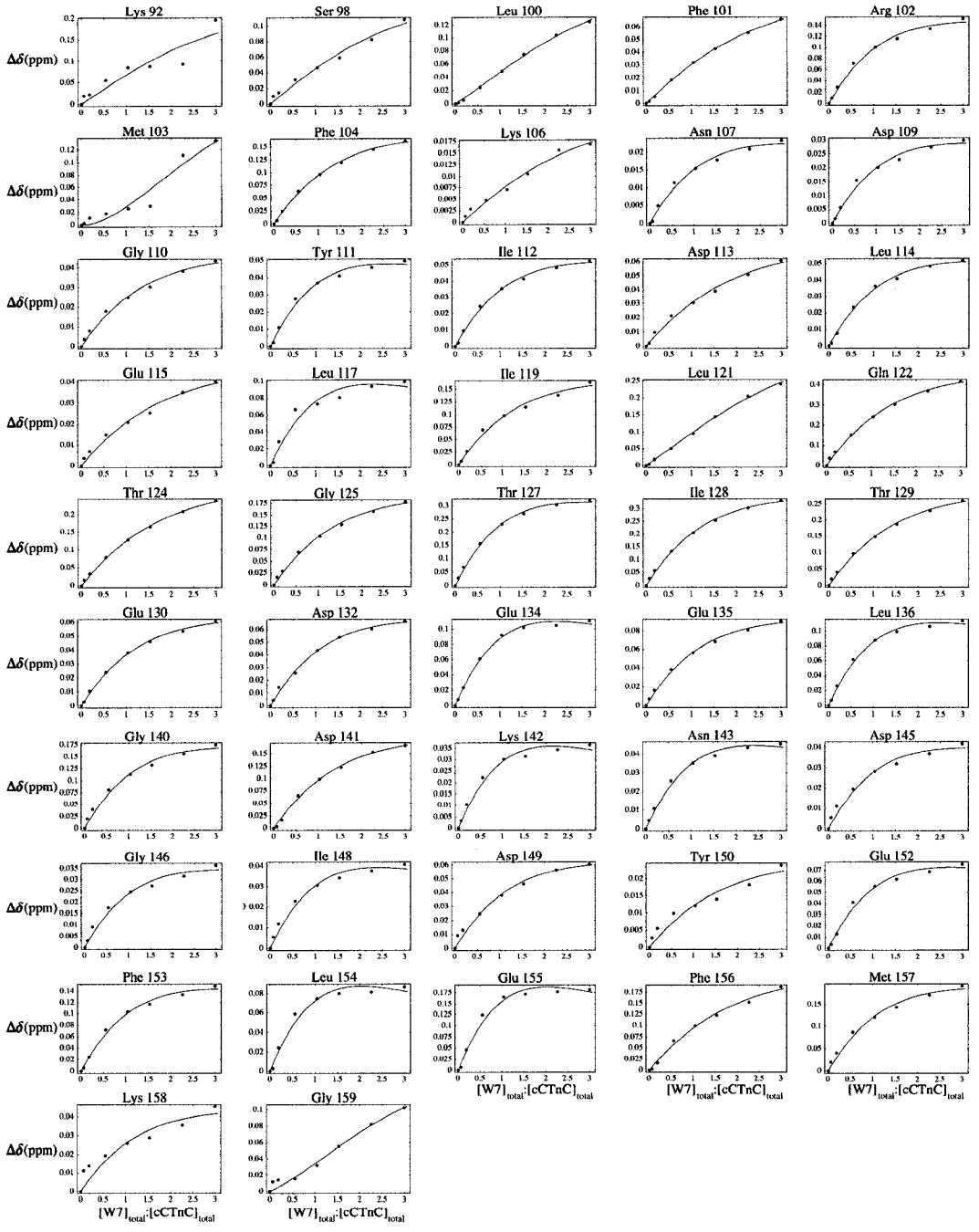


Figure 4.9: 47 of the 2D-{^{1}H,^{15}N}-HSQC crosspeaks from the titration of cCTnC•2Ca^{2+} with W7, fit to a two-site, sequential binding model ($K1$ = 0.88 mM; $K2$= 3.5 mM). The raw titration data analyzed here are shown in Figure 4.2. The quality of the global fit is high, with few residues' fits having substantial outliers. This figure is reproduced with permission from Reference [12].

4.2.13 Global fitting

Protein NMR spectroscopists are provided with a wealth of site-specific titration data, akin to the simultaneous monitoring of many non-perturbing probes (nuclei). Such experimentalists usually wish to report global binding constants. This is most often accomplished by selective consideration of NMR signals that are strongly perturbed by binding. The most strongly shifting residues can be parsimoniously (and naively) assumed to be the most proximal to the binding site. Furthermore, strongly perturbed residues have the largest signal to noise content given that the 'accurate' perturbation in chemical shift needs to dominate the 'random' perturbations in the detection of a peak's maximum. The chief problem with selective characterization of the strongly shifting residues is that, depending on the averaging strategy used, the apparent precision of the result can be somewhat arbitrarily manipulated.

In the case of W7 binding to cCTnC•2Ca^{2+}, the breakdown of the single-site binding model was not at all indicated by examination of the data for Ile128 and Thr129. The general argument for globally fitting data of this sort is that *exceptions disprove the rule* and so one must be aware of self-inconsistencies in the data. Some titration signals are strongly coupled to artifactual processes. Other titration signals could be monitoring an allosteric response rather than the primary binding event. But in the case of weak binding, when only the pre-asymptotic phase of the binding curve is available for analysis, the tremendous site-specific resolution of NMR spectroscopy offers unique opportunities to oversample this phase of the curve.

4.2.14 Conclusions: the binding of W7 to cCTnC•2Ca^{2+}

The above discussion introduces the main experimental approach continued throughout this chapter. Most of the discussion from this point forward will not make reference to these data. The main reason for this is that subsequent experiments showed that in the presence of TnI_{34-71} (H1), W7 no longer binds to cCTnC•2Ca^{2+} [13].[7] The binding of TnI blocks the binding site for W7 so that the interesting solution behavior characterized in the previous section has little-to-no biological relevance. The W7–cCTnC•2Ca^{2+}system does not contain enough of troponin's functional context to allow for the elucidation of structure-function relationships and is therefore presented here as a (problematic!) model system.

4.3 The binding of W7 to cNTnC•Ca^{2+}

Monica Li also studied the binding of W7 to cNTnC•Ca^{2+}. This experiment was expected to more readily inform W7's mode-of-action in the thin filament. For the cardiac isoform of TnC, the Ca^{2+}-activated N-terminal domain is not stabilized into an 'open' conformation but instead undergoes a closed↔open conformational equilibrium. Binding of the TnI switch region (Sp) stabilizes the open conformation (and initiates contraction). A ligand that can modulate the closed↔open conformational equilibrium, or can modulate the cNTnC•Ca^{2+}–Sp binding equilibrium, is of po-

[7] This recapitulates the results of Sorsa and others, with levosimendan as the ligand, as quoted in Section 4.2.6 (p. 49).

tential chemotherapeutic interest. As shown below, W7 has both of these properties, similar to bepridil [136].

4.3.1 NMR-monitored titrations

The methodology here is the same as used for the previous example, and is detailed in Appendix B. Figure 4.10 shows the correlation spectra acquired over incremental increases in [W7]. A subspectrum showing the response of methionine Cϵ methyl correlations is also shown.

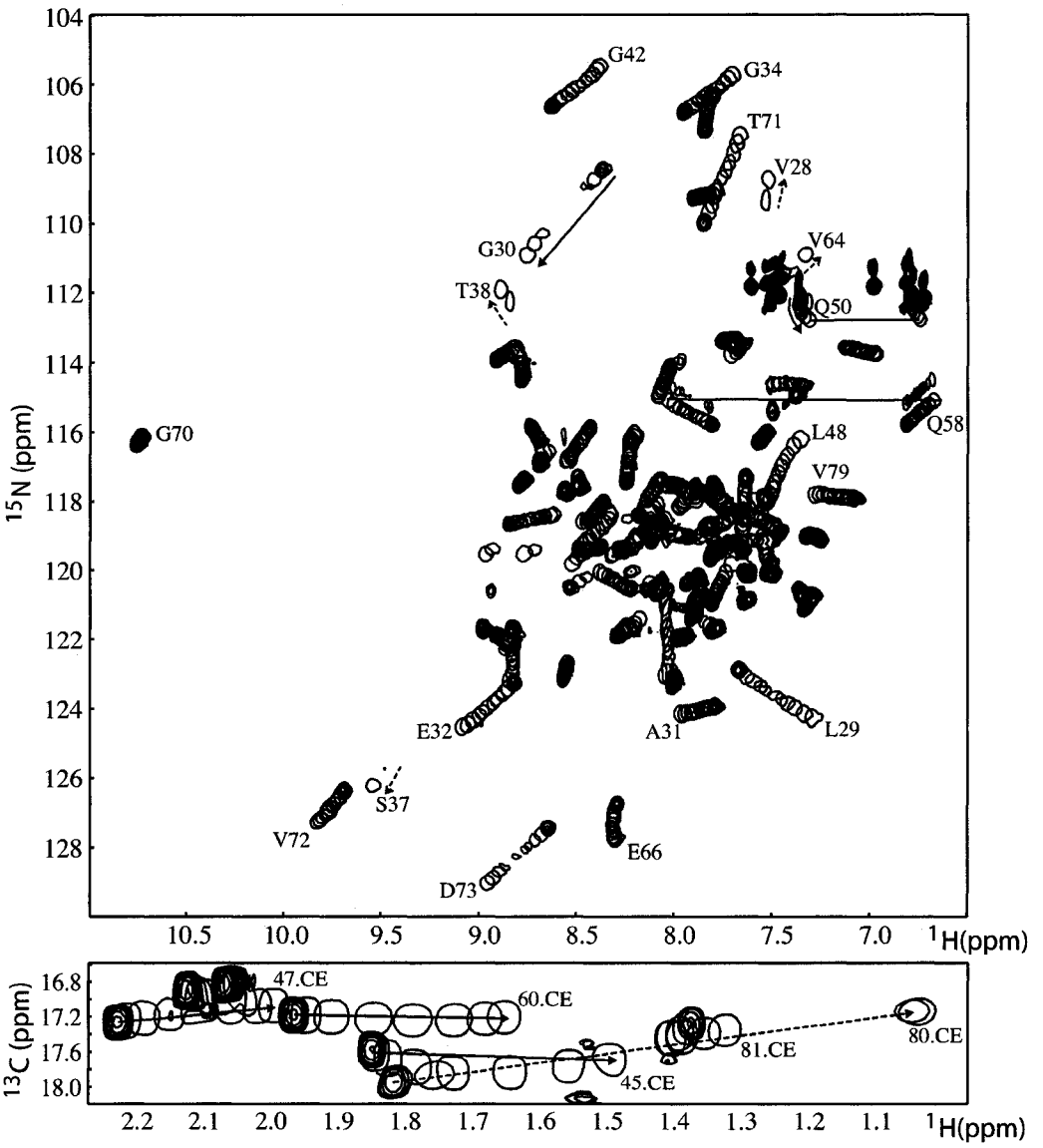


Figure 4.10: Titration of W7 into cNTnC•Ca^{2+}, monitored with $\{^1H,^{15}N\}$-HSQC (top) and $\{^1H,^{13}C\}$-HSQC (bottom) NMR spectra. The initial point of the titration is shown with filled contours. Intermediate and final points are displayed with open contours. The final increments are assigned. Arrows indicate the paths underwent by the crosspeaks; dashed lines indicate the most probable trajectory for extensively broadened peaks. At top: most backbone amide signals shift linearly. Amide sidechain crosspeaks from Q50 and Q58 are titratable (each sidechain generates two crosspeaks, which are connected with lines). Some residues' crosspeaks broaden over the middle points of the titration. At bottom: all methionine Cϵ methyl crosspeaks are titratable. The crosspeaks from M80 and M81 shift in distinct ways. This figure is reproduced with permission from Reference [12].

Most 2D-{^{1}H,^{15}N}-HSQC crosspeaks (Figure 4.10, top) shift 'linearly' (consistent with Equation 4.4) which means the initial, intermediate, and final chemical shifts for a given NMR crosspeak can be connected with a straight line. Exceptions to this pattern include crosspeaks that do not shift, or that deviate from linearity. Figure 4.10, top shows two titratable backbone amide crosspeaks that shift along linear paths until the final increments. One of these, belonging to L48, is located in the linker region connecting cNTnC's first (non-Ca^{2+}-coordinating) EF hand with the second EF hand. The other, corresponding to E66, is localized to the Ca^{2+}-binding site. Many of the amide crosspeaks undergo changes in lineshape over the titration: L29, G30 and D73 are noteworthy examples. This line broadening indicates exchange kinetics on the intermediate NMR time scale for these specific signals. V28, S37, T38, and V64 are extensively broadened to the point of being unobservable near the midpoints of the titration. These signals shift in directions consistent with the 'opening' of cNTnC, usually effected through the binding of $cTnI_{147-163}$ to cNTnC•Ca^{2+} [137].

Nuclei located in Gln and Met sidechains recapitulate the chemical shift changes reported by backbone amide nuclei. The amide sidechain crosspeaks of Q50 and Q58 are titratable (Figure 4.10, top). The sidechain amide group of Q58 presents two chemically non-equivalent protons shifting in opposite directions over the titration (Figure 4.10). This behavior is demonstrated by the same residues under activation by $cTnI_{147-163}$ [137]. It is therefore likely that the behavior of Q58's NMR signals are only indirectly coupled to W7 binding. The sidechain amide protons of Q50, however, shift in the same direction. Electrostatic fields can impinge upon protein atoms upon ligand binding. Such events perturb the chemical shift of the nuclei of polarizable atoms. A modeling study of the mechanism of Ca^{2+} sensitizers in cTnC [138] concluded that Q50 and D88 had long-range electrostatic interactions which impacted Ca^{2+} binding. This interaction was perturbed under binding of TFP, bepridil, or pimobendan. The present results support a functional role for Q50, which is clearly titratable (Figure 4.10). The chemical shift changes for the sidechain of Q58, coupled simultaneously to other spectral changes indicative of the 'opening' conformational transition of cNTnC•Ca^{2+}, show that W7 can directly modulate the functional conformational equilibria in cNTnC.

2D-{^{1}H,^{13}C}-HSQC crosspeaks originating from Met Cϵ methyl groups (Figure 4.10, bottom) show extensive changes over the titration. Crosspeaks from M47, M60, and M80 shift linearly as W7 is added to the sample. M80 experiences notable exchange broadening, with the crosspeaks in the midpoint of the titration being broadened to the point of being unobservable. M81 and M45's crosspeaks shift non-linearly; M81's path is so non-linear as to have a sharply-defined inflection point at the midpoint of the titration. The close proximity of M80 and M81 in the primary structure is not reflected in their Cϵ methyl groups' spectral behavior over the titration.

4.3.2 Global fitting to single-site binding model

As outlined for the fitting of the cCTnC•2Ca^{2+}data (Section 4.2), the raw data were converted into titration curves, and nonlinear least-squares fitting was applied. The global parameter, K was systematically varied, and the local parameter, δ_{∞}, was allowed to vary, reflecting the signal-specific

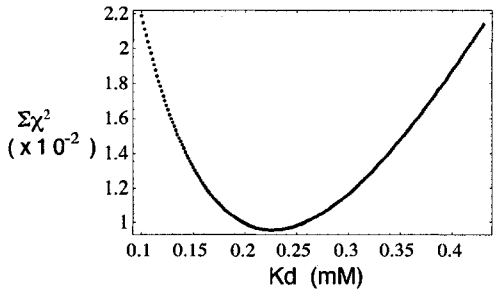


Figure 4.11: Determination of a global K for the cNTnC•Ca^{2+}•W7 complex within a single-site binding model. The vertical axis shows the sum of χ^2s for the single-site fits, given a particular K (horizontal axis). The minimum $\Sigma\chi^2$ occurs over the interval of 0.15–0.30 mM. Over this range, 41 $\{^1H,^{15}N\}$-HSQC signals contribute to the fit. The algorithm used to obtain global fits automatically culls local fits with $\Sigma\chi^2$s over 100. This figure is reproduced with permission from Reference [12].

variation in sensitivity to binding. The one-dimensional[8] search of the $\Sigma\chi^2$ space is shown in Figure 4.11. A confidence interval for K of 0.15–0.30 mM, was selected based on this plot. I emphasize that, as with the previous analysis, the confidence interval is not selected by objective criteria and so this analysis is not completely satisfactory. Nonetheless, the global fit (shown in Figure 4.12) seems to adequately reconcile 41 binding signals.

[8]There are 41 local parameters is this fit because the values of δ_∞ are fitted locally.

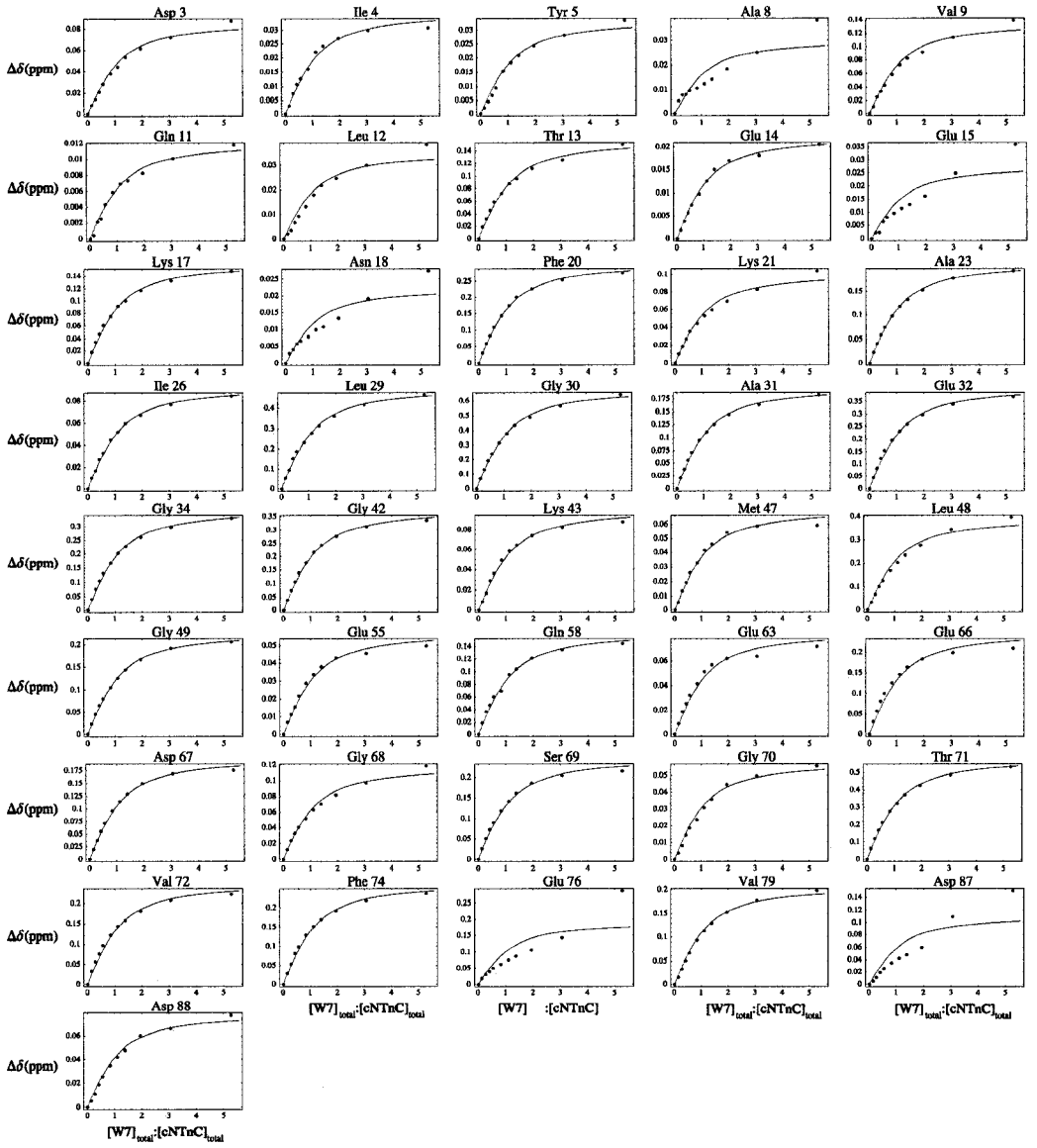


Figure 4.12: 41 of the 2D-$\{^1H,^{15}N\}$-HSQC crosspeaks, from the titration of cNTnC•Ca^{2+} with W7, fit to a single-site binding model (K = 0.22 mM). The quality of the fit, considered on a residue-by-residue basis, is seen to be high. Note how the signals from residues 8,15,18,76, and 87 all have residuals that systematically deviate from the global fit, and that systematically agree with each other. These signals are both weak (Ala 8) and strong (Glu 76) reporters of the binding event, and it is unclear which residues should be excluded based on signal-to-noise arguments. This figure is reproduced with permission from Reference [12].

4.3.3 Likelihood analysis

In general, the more strongly multiple observations corroborate a hypothesis, the more strongly one can assert its validity. The process of scientific induction – perhaps all cases of correct statistical inference [132] – is described by Bayesian statistical theory [132, 133]. The relationship between the plausibility of a hypothesis and the weight of evidence are summarized in Bayes' Theorem:

$$P(K|\{\Delta\delta_i\},\{\sigma_i\},E) = \frac{P(\{\Delta\delta_i\},\{\sigma_i\}|K,E)P(K|E)}{P(\{\Delta\delta_i\},\{\sigma_i\}|E)} \tag{4.10}$$

The set of observations $\{\Delta\delta_i\}$ also denotes the known concentrations of titrant and target, and the set of experimental errors $\{\sigma_i\}$ is also taken as known[9]. The experimental conditions are explicitly represented with the symbol E. Equation 4.10 has the following verbalization: *the probability of* K given the data, errors, and conditions is directly proportional to *the likelihood* of the data and the errors (given the value of K and the conditions), and to the *probability of* K *in the absence of the data* (given the conditions), and inversely proportional to *the evidence* (the probability of the data and errors, given the conditions).

Consider the experimental process: the researcher attempts a titration experiment (hypothesizing the existence of an observable binding event) by NMR (limiting the solubilities of the system components to a realistic range) and observes a binding signal. The observation of a binding signal implies that the change in the signal of the target molecule is larger than the limitations of spectral resolution (which could be estimated beforehand, given the spectrometer field strength, the pulse sequences used, and the intrinsic linewidth of the signal). Application of Equations 4.5 and 4.6 *already implies prior knowledge* of fast exchange kinetics, which constrains the possible values of K given the smallest observable change in the signal.

Furthermore, consideration of multiple titration signals that are highly similar to each other lends much more plausibility to their mutual identification with a common parameter. The ability of the experimenter to selectively include or exclude a given signal, and the inability of orthodox statistics to prescribe the correct decision (akin to the well-known stopping problem in statistical theory), motivates an explicit consideration of the *force of evidence* as written in the denominator of Equation 4.10. A theoretically rigorous expression for this term in the denominator may allow for the selection of an appropriate subset of local binding signals consistent with the global model. I have not yet arrived at such an expression, but the assumption that all signals reporting the same event fall within one standard deviation of each other, as explored below, does lead to a useful analysis.

A Gaussian likelihood function

When there is no prior knowledge of the parameter value (corresponding to a uniform prior probability distribution), and when the probability of a given observation (the rareness or quality of the evidence) need not be considered, the least squares and maximum likelihood approaches are very similar. For a Gaussian distribution of errors the two approaches give formally equivalent solutions.

[9]The index on σ shows that the error varies with the titration increment.

An advantage of likelihood analysis is that the errors are explicitly fitted to a model (instead of being estimated from the quality of the fit) and that it can be extended to situations where prior likelihoods, or varying quantities of evidence, impact the certainty with which one may estimate the parameter values.

In most cases of parameter estimation, Equation 4.10 can be reduced to the proportionality:

$$P(K|\{\Delta\delta_i\},\{\sigma_i\},E) \propto P(\{\Delta\delta_i\},\{\sigma_i\}|K,E) \tag{4.11}$$

so that maximization of the likelihood occurs at the most probable value of K. Modelling the probability distribution of each datum as a Gaussian, and the likelihood of the data as the product of the individual probabilities, the likelihood can be written as:

$$P(\{\Delta\delta_i\},\{\sigma_i\}|K',E) = \prod_i \prod_j \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(K-K')}{2\sigma_i^2}} \tag{4.12}$$

where I have adopted a notational shortcut of denoting a probability distribution function: the variable K' is a point on the distribution, not the parameter value situated at the center of the distribution. The double product shows that the probability of each datum, for each increment i, and each signal j, contributes to the overall likelihood.

Applying the analysis to the binding of W7 to cNTnC•Ca^{2+}

Appendix D.1 shows the Mathematica script I wrote to implement this analysis. The main piece of algebra required is manipulating Equation 4.6 so that a K is predicted given a value of $\Delta\delta_i$, the observable, and σ, the global error in the titration:

$$K = \frac{\Delta\delta_\infty L_i}{\Delta\delta_i} + \frac{\Delta\delta_i P_i}{\Delta\delta_\infty} - L_i - P_i \tag{4.13}$$

The term $\Delta\delta_\infty$ is shorthand for the term $\delta_\infty - \delta_0$ in Equation 4.6. With equations 4.12 and 4.13, I numerically evaluated the global likelihood, varying K' in Equation 4.12 and locating the maximum (after assuming an initial value of the global error σ). Then, I optimized σ in the same fashion (holding the value of K fixed to the previously fitted value). I initially evaluated this with all 41 of the (normalized) titration signals illustrated in Figure 4.12.

The analysis implied that a number of signals, when fitted to the global parameters, had local fits that fell outside one value of the global σ. Excited that I may have a procedure for selecting the correct subset of residues consistent with a global fit, I iterated the analysis, each time excluding signals that fall outside one standard deviation of the global fit. Over six iterations, the global parameters changed from $K = 13.6\mu M$, $\delta_\infty = 0.759$, $\sigma = 238\mu M$ (41 signals considered) to $K = 144\mu M$, $\delta_\infty = 1.06$, $\sigma = 40.0\mu M$ (21 signals considered).

I conclude that likelihood analysis is a plausible procedure for global fitting of these data and for the establishment of appropriate error bounds for the global fit. The analysis suggests a possible solution to the model selection problem when the analysis is applied iteratively, excluding outlier signals. With the error analysis on firmer ground, I return to the consideration of titration error.

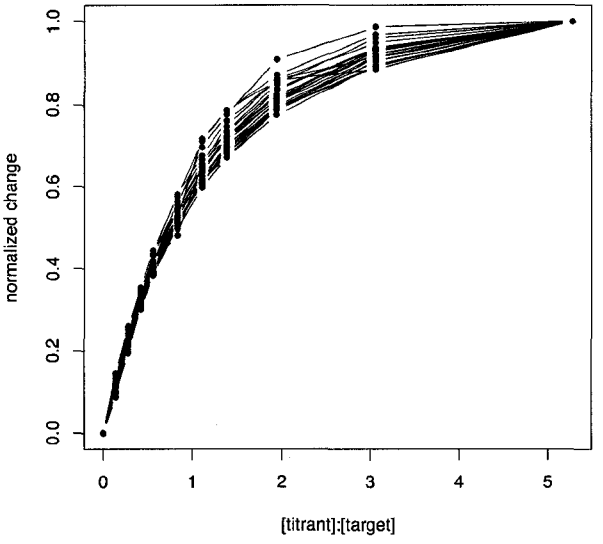


Figure 4.13: The subset of 21 globally-consistent titration signals after normalization. Although the errors in the Y-coordinate are heteroskedastic, the variation does not extend to the extrema, so a least-squares fit could still give an optimal value. Consideration of the cumulative error in the X-coordinate, which is heteroskedastic but *the same for each titration increment* is needed or the fit will be overly precise.

4.3.4 Titration error and normalization revisted

It is illustrative to simultaneously plot the NMR signals (the 21 residue subset selected above) after normalization (Figure 4.13). The first and last data points are fixed at 0 and 1. The divergences between the multiple signals at intermediate points increase monotonically over the titration. Since least squares fits are highly influenced by extrema, the normalization approach ensures that *both the zero point and the final point have infinite precision* so that heteroskedacity in the Y-coordinate has negligible influence.

Unfortunately, this happy situation draws emphasis to the confounding influence of the X-error. The excellent interpolability of the data demands more explicit modelling of titration error. Unlike the global change in the response variable, which is monitored simultaneously on corroborating sites, the titration error can only be directly estimated from performing replica experiments. Just as an astronomer gets a single opportunity to observe the death of a star, the protein NMR spectroscopist has finite financial resources for isotopic labels and spectrometer time, so conducting replica experiments of this kind is often unfeasible.

For the analysis of a single titration, the typical error propagation used in the analytical treatment of gravimetric and volumetric methods can be used to estimate the error. If replica experiments exist the errors can be estimated from the observed variance, or a combination of the two approaches. For example, the random error in the preparation of a stock (titrant) solution becomes a systematic error when propagated through replica titrations (using the same stock of titrant). One way to incorporate estimates of titration error is weighted least squares.

Weighted least squares

Limited heteroskedacity (non-uniform distributions of the error) can be treated with weighted least squares. In this treatment, an increment's residual is inversely scaled by a weighting, W_i. This is commonly approximated as $\frac{1}{E_i^2}$ (for errors distributed normally) or $\frac{1}{E_i}$ (for error distributed accord-

ing to a Poisson distribution). The target function for least-squares fitting (Equation 4.7) is modified to de-emphasize error-prone data:

$$\epsilon(K, \delta_\infty, \{\Delta\delta\}, \{W\}) = \sum_i W_i \left(\Delta\delta_{obs,i} - \Delta\delta_{pred,i}\right)^2 \tag{4.14}$$

As with nonlinear least-squares fits, there is no general expression for the quality of the global fit based on the final value of the target function.

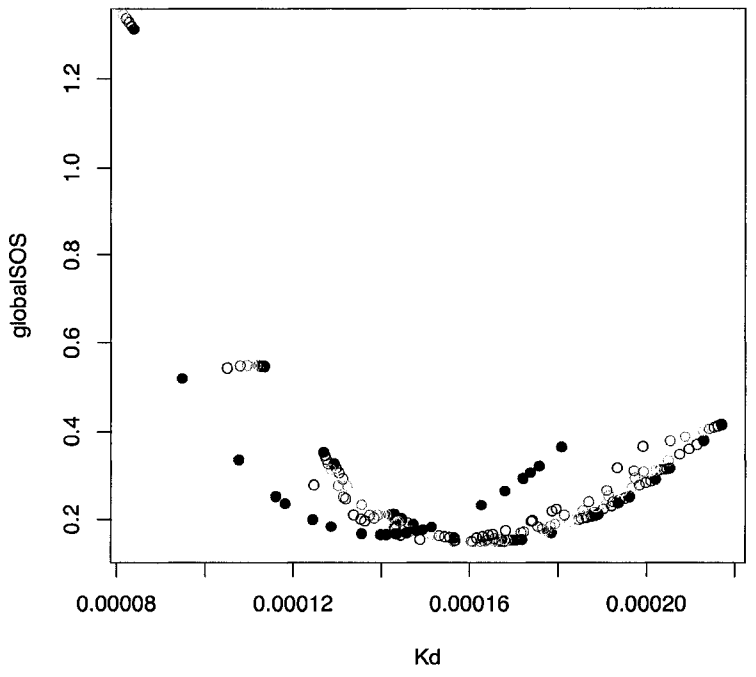


Figure 4.14: Global analysis of the binding of W7 to cNTnC•$Ca^{2+}$with weighted least-squares. Each color in the plot represents a different value of the weighting vector. The solid black circles are the unweighted solution. The open circles colored brown, maroon, red, orang-red, orange, yellow, yellow-green, green, and blue incorporate 1, 2, 3, 4, 5, 6, 7, 8, and 9% cumulative error over the titration, respectively. The solid purple circles show the solution with cumulative 10% error, which leaves the last point completely excluded from the analysis. Still, excluding this point does not greatly perturb the solution, showing that titration error ought to be considered.

I repeated the global least-squares analysis but with weightings. The script for the analysis is given in Appendix C.3.The results are summarized in Figure 4.14. By increasingly de-emphasizing the latter points in the titration, the globally fitted value of K does significantly change. The unweighted analysis (Figure 4.14, solid black points) has a minimum around $K \sim 160\mu M$. If each successive point in the titration is de-emphasized by 10% (which excludes the final point completely) the minimum shifts to $K \sim 140\mu M$.

The weighted least squares analysis motivates two conclusions. First, global analysis is sufficiently robust that explicit consideration of titration error does not confound the analysis. Second, global analysis is sufficiently robust to estimate K from the pre-asymptotic phase of the binding

curve, as even deletion of the final point does not greatly perturb the value of the fitted parameter.

4.3.5 Titration of cNTnC•Ca^{2+} into W7

The usual approach in my mentor's laboratory is to make a stock of ligand and to add that into a sample of protein, as the primary interest is the protein signals' respond to binding. The titrations analyzed above were all executed in this fashion. The binding of W7 to cNTnC•Ca^{2+}(Figure 4.10) showed indications of multiple-site binding, or at least two different processes modulating signals. While this process did not impact the binding event as monitored at the backbone amides, the Met Cϵ methyls did directly evince two processes (especially Met81 and Met45).

As discussed above, there are many phenomena that could be invoked to explain the multistate behavior. Many more mechanisms can be conceived than can be differentiated on the basis of the data (Section 4.2.11, p. 52). An obvious factor is the presence of DMSO, which is the solvent in which stock concentrations of ligand can be performed. Control experiments show that DMSO does not directly perturb the NMR signals of cNTnC•Ca^{2+}(most recently articulated in Reference [139]) but it could introduce – actually it must introduce – solution equilibria that modulate the activity of W7. This raises doubts regarding the validity of the usual mass-action relationships which are strictly applicable to infinitely dilute solutions.

To clarify the solution equilibrium, I performed a titration in which buffered protein is added to a solution of W7. The initial W7 solution is dilute in terms of total W7 concentration because there is no DMSO present to increase its apparent solubility.

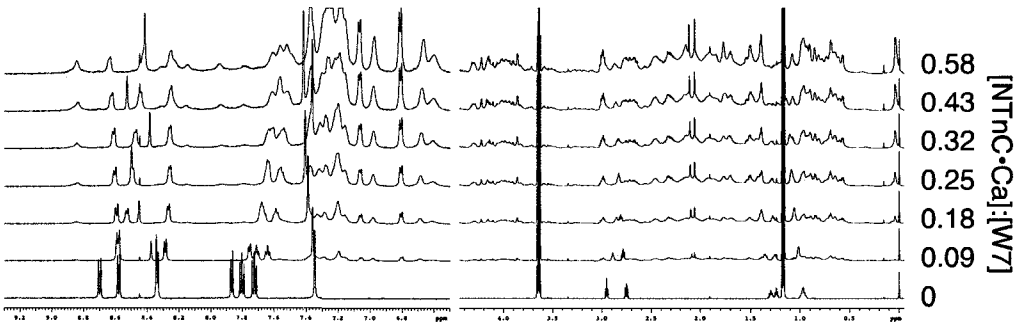


Figure 4.15: The binding of cNTnC•$Ca^{2+}$to W7, monitored with 1D ^{1}H spectra. The protein:W7 ratio is shown to the right of each spectrum. cNTnC•Ca^{2+}-binding induces chemical shift changes. Unexpectedly, the binding signal is more sensitive to titrant than was observed for the binding of W7 to cNTnC•Ca^{2+}, indicating that the binding is much stronger. The pH varied within a unit over this titration, so the experiment was repeated with more controlled pH, shown in Figure 4.16.

The titration is shown in Figure 4.15. The W7 resonances (at the bottom of the plot) can be tracked through the titration to their final points. As such they are in the fast exchange limit and can be analyzed with the same formalism as for the other titrations. These data afford the initial and arresting observation that binding, as monitored through this experiment, is as much as an order of magnitude stronger than the 0.22 mM dissociation constant previously determined (Figures 4.11

and 4.12). This experiment was repeated in order to confirm this finding (Figure 4.16).

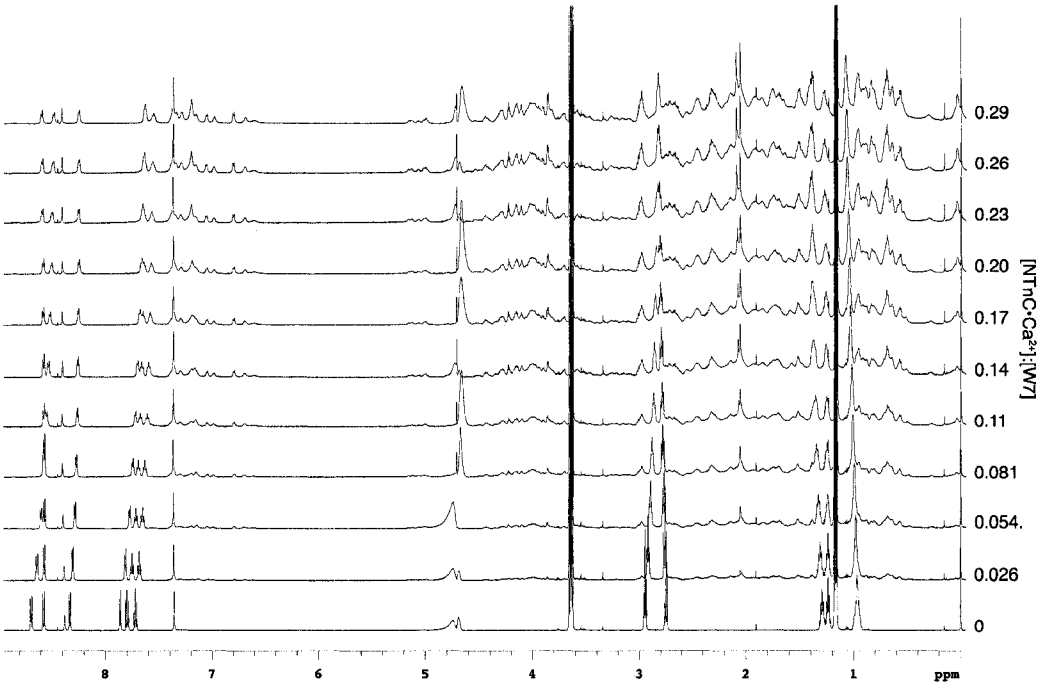


Figure 4.16: The experiment shown in Figure 4.15 was repeated with the pH more tightly regulated. The apparently strong binding was corroborated by this data.

The most-downfield signal strongly responds to cNTnC•Ca^{2+}-binding and was selected for initial analysis. The peak frequencies were extracted from the 1D spectra in VNMRJ and the titration curve was visualized.

The ^{1}H spectra in Figure 4.16 cannot be fit with a single-site binding model. The binding signal responds more quickly than could be anticipated for infinitely strong 1:1 binding. As such, the appropriate form of a binding model describing this equilibrium would allow higher stoichiometries than 1:1 binding. I proposed and Brian Sykes modeled the following solution equilibria to reconcile the findings:

$$\mathrm{P} + \mathrm{L} \leftrightarrow \mathrm{P} \bullet \mathrm{L} \leftrightarrow \mathrm{P} \bullet \mathrm{L} \bullet \mathrm{L_n} \tag{4.15}$$

Where P is the protein (cNTnC•Ca^{2+}) and L is the ligand (W7). The model has two equilibria. The binding constants for the second step, as well as the stoichiometry of the second binding event (n) allow for explicit modeling of secondary nonspecific binding. The simplest form of the model has the ligand-detected binding signal responding 1:1 with the depletion of the free species L (so one mole of $P \bullet L \bullet L_n$ reflects the binding of $n + 1$ ligand signals).

Although this model cannot be solved analytically, numerical simulation (see the spreadsheet in Appendix E) allows for confirmation that the *form* of the model is appropriate. As shown in

Figure 4.17, both the titration of W7 with cNTnC•Ca^{2+}, and the titration of cNTnC•$Ca^{2+}$with W7, are consistent with this model. Appropriate parameter values here are $K1 \sim 40\mu$M, and $K2 \sim 0.3$pM with $n = 4$. These values are not inconsistent with the affinities reported by Hidaka and others [123]; with a global $K2$ of 0.3pM and a stoichiometry of 4, the microscopic binding would be on the order of $\sqrt[4]{3x10^{-13}} \sim 0.7$mM. Examining the global analysis for the binding of W7 to cNTnC•Ca^{2+}(Figure 4.12), one sees that the majority of signals (especially the most sensitive ones) have residuals of a positive sign at early titration increments. This strong binding event may account for these pervasive and systematic residuals.

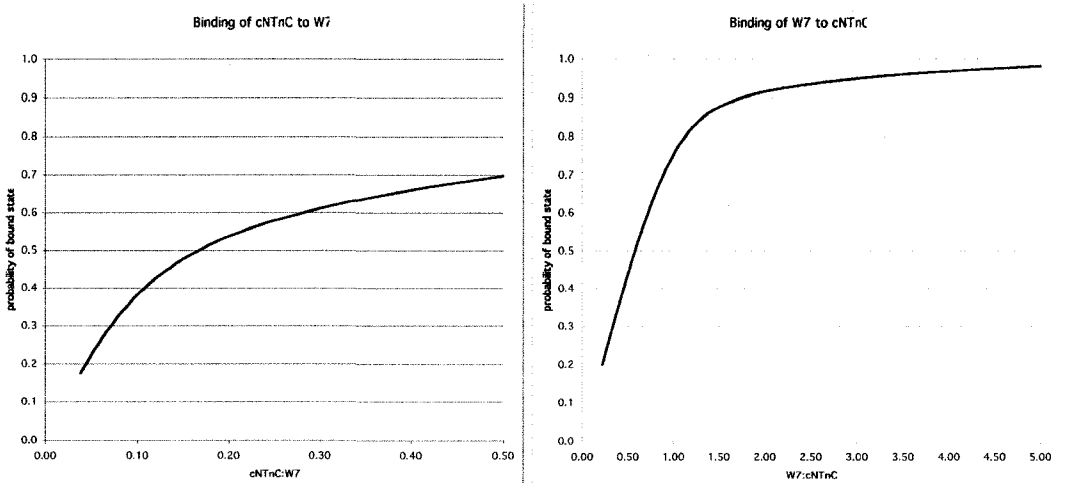


Figure 4.17: Simulation of the W7–cNTnC solution equilibria based on Equation 4.15.

4.4 Biochemical implications

It is evident that W7 binds to the domains of cTnC, perturbing their structures. This interpretation is favored over the alternative explanation of W7 promoting a wholly distinct fold, because a set of the crosspeaks are well-dispersed (reflecting a stable tertiary structure) and this dispersion is insensitive to W7 binding. The constructs used in this work lacked Cys residues (they are substituted with Ser). Previous studies have implied that C84 of cNTnC is required for the binding of levosimendan [140]. Evidently, in the case of cNTnC•Ca^{2+}•W7, binding competency does not necessitate the presence of this residue, as was demonstrated for cNTnC•Ca^{2+}•bepridil. The presence of another ligand-coordinating residue could enhance the affinity of wildtype cNTnC for W7, however.

The Ks obtained in this work disagree (by more than an order of magnitude) with a previous report of the primary (strong) binding of W7 to TnC•4Ca^{2+} [123]. It is difficult to reconcile this discrepancy because the previous study used intact chicken TnC, a distinct homolog, and different conditions. As well, the present work focuses on the domain-specific binding interactions, and attempts to characterize two binding sites for the cCTnC•2Ca^{2+}•W7 equilibrium. Two-site, sequential fits for cCTnC•2Ca^{2+}•W7 yield a K for the primary binding event of ~0.88 mM. It is

important to note that the K for the secondary binding event, 3.5 mM, is the smallest possible K consistent with the model used, which requires $K2$ to be more than 4 times the value of $K1$. This indicates the secondary binding site to be similar in affinity to the primary one. The advantage of the model used is that the solution is analytic; the disadvantage is that because the model is stoichiometric and not explicitly site-specific, the binding constants obtained are difficult to interpret alongside the characteristically site-specific data generated by NMR. The stoichiometric picture presented here supports the theory that W7 is binding in multiple poses to cCTnC•2Ca^{2+}, alluding to the conclusions of Craven and others in the study of CaM•4Ca^{2+}•J8 by NMR spectroscopy [126]. The conclusion regarding J8 was obtained through an automated assignment of the intermolecular NOEs. Weak binding can reflect a larger external (mechanical/configurational) entropy for the ligand in the bound state [141] raising the issue of whether an atomic-resolution description of weak binding is possible without deconvolution of the ensemble-averaged observables.

4.5 Conclusion

This chapter summarizes an experimental and intellectual odyssey that has enriched my doctoral studies. The foremost conclusion I have reached is that NMR can provide more observations than an experimenter may be willing to interpret! I have motivated, but not proved, the conclusion that this interpretative burden is a blessing in disguise. Explicit reconciliation of the local signals into a global model allows for increased precision in the fitted parameters, which is helpful in the study of weak binding, where the asymptotic phase of the binding curve cannot be explicitly sampled. The statistical robustness afforded by global analysis motivates explicit consideration of the titration error. Through this work I have proved that W7 binding to either Ca^{2+}-saturated domain of cardiac TnC is *not* single-site or one-to-one. My titration data for the binding of cNTnC•$Ca^{2+}$to W7 shows that a strong binding event does occur and suggests appropriate conditions for the formation of a 1:1 complex. In the next chapter, I visualize cNTnC•Ca^{2+}•W7, one of the states reached through the process of binding.

Chapter 5

Structure of cNTnC•Ca^{2+}•W7

> "Biological processes are essentially dynamic. ... Close collaboration with [non-crystallographic] disciplines is necessary in order to test hypotheses within the framework of the known static structure. The continuing development of nuclear magnetic resonance techniques appears most promising for the detection of the transient structures not easily accessible to crystallography and may provide important structural information that is at present beyond the limit of resolution of diffraction methods." [142, p. 12]

The term *rigor* denotes both precision and accuracy. There is a continuum of methodological rigor in the field of structural biology. Protein crystallography has, over its relatively long history, evolved towards rigor. Very high resolution crystal structures can have one or more observation per degree of freedom in the model, making the model *actually* determined. Most crystal structures are not of this quality, and rely on iterative model building to resolve the model. Initial models can be built from knowledge of analogous structures, including fragments of those structures. This structure determination strategy is called *molecular replacement.* The manifold strategies for phasing and refining a crystal structure are guided by a widely accepted validation procedure, to guard against over-refinement. The far opposite end of the continuum of rigor are theory-based predictions of protein conformation from knowledge of sequence. This important area attempts to predict the structure with no sequence-specific empirical data, one of the main goals of protein science. As the protein folding problem remains unsolved, no theory-based prediction can rival the rigor of empirical methods, although the gap is closing.

The opening quote of this chapter, from a fairly old (around 1976) reference on protein crystallography, alludes prophetically to the role of NMR spectroscopy in structural biology. Indeed, NMR has shown to be an important complementary method to crystallography. But no NMR-based structure determination has approached the theoretical rigor of the extant crystallographic methods, and this may be impossible. All NMR-based structure determinations have relied on empirical knowledge of target covalent geometry – these parameters come from crystal structures.[1]

[1] An oft-repeated ritual is the indoctrination of new graduate students towards exclusively favoring crystallography or NMR, as if the methods themselves were as interesting as their subjects of study.

To determine a protein conformation by NMR a number of observations are required to exclude most of the conformational space. The conformers consistent with all of the observations form the *NMR ensemble.* An NMR ensemble has no relationship to a thermodynamic ensemble. When the members an NMR ensemble are highly similar to one another, this result is frequently referred to as a *high resolution NMR structure* which is misleading because the concept of spatial resolution does not apply here. Perhaps the term originally denoted a structure determined from high *spectral* resolution observations; if so, the popular usage has tragically diverged. I want to popularize the idiom *high-precision NMR structure* to replace the looser terminology.

High-precision NMR ensembles reflect only a self-consistency in the data. Low-precision NMR ensembles can result from a lack of experimental observations (from practical reasons like limited spectral resolution), or from self-inconsistent observations (from errors or from conformational heterogeneity). Most NMR-based structure determinations rely on the distance dependence of the intensity of the nuclear Overhauser effect (NOE). Importantly, the interpretation of NOEs as originating from a single conformer *imposes* a physically implausibly high precision to the coordinates of the calculated ensemble [143]. I say all of this to emphasize that a structure determination by NMR is a means of describing the most-highly ordered conformer. The accuracy of NMR ensembles cannot be determined without prior knowledge of the structure. As such, the only accepted procedure for validating an NMR structure is its compliance with idealized covalent geometry [144].

As discussed in Chapter 4, the NMR-detected, W7-induced structural changes in cNTnC•Ca^{2+} are consistent with the well-characterized closed-to-open conformational transition. Therefore, the protein coordinates for the bound state are hypothesized to be well-approximated by those of the open conformation. This motivates a structure determination strategy that bootstraps from the past knowledge of the open conformation of cNTnC•Ca^{2+}, akin to the molecular replacement procedure in crystallography.

There are two structure determinations of the protein in this conformation from my mentor's laboratory. One study was historically important for establishing that cNTnC•Ca^{2+} occupies an open conformation when bound to the switch region of TnI [137]. Another study showed the switch-peptide and the cardiotonic drug bepridil bound, together, to cNTnC•Ca^{2+} [136]. This structural difference was assigned to be the mechanistic basis for the weaker affinity of the two ligands for cNTnC•Ca^{2+} when they are both in solution. W7 was introduced as a model cardiotonic ligand in the previous chapter (Section 4.1, p. 42). In this chapter[2], NOE-derived distance restraints are used to dock W7 onto cNTnC•Ca^{2+}. The results illustrate that drug binding to cNTnC•Ca^{2+} can modulate the affinity for switch peptide, perhaps illuminating a cardiotonic mechanism of action.

5.1 Assignment of W7

The NMR spectrum of W7 (Figure 5.1) was difficult to assign. The molecule has two symmetrically arranged aromatic proton spin systems that cannot be readily distinguished from each other. Within

[2]The results in this chapter are my unpublished results, based on data I acquired and analyzed. This work is the latest installment of a long-term collaboration with Monica Li. Some of the discussion derives from previous binding studies [12] which (correctly) anticipated the present structures. Unlike other chapters, the methodology is included in-line with the discussion because it had not yet been peer-reviewed.

each spin system, only the central spin is identifiable by direct inspection of the DQF-COSY (Figure 5.1, left). A number of experimental strategies were employed. Alternate solvents were tested[3] to see if the chemical exchange of the sulfonamide hydrogen could distinguish it from the amine hydrogens. If this signal were unambiguously identified, perhaps it can be correlated to the ring hydrogen H2. No appropriate aprotic solvents were found towards this end; we tried methylene chloride, acetone, and chloroform. W7 was insoluble in all of those solvents.

Unambiguous assignments were eventually accomplished through natural-abundance ^{13}C spectroscopy at high fields (800 MHz ^{1}H). The directly-detected ^{13}C spectrum was predicted to have a $^{35}Cl/^{37}Cl$ isotope shift for ring carbon C5, giving rise to two ^{13}C signals with the expected peak separation (1.0 Hz) [145]. The ramifications of this lynchpin observation are depicted in Figure 5.1. The key HMBC correlation supporting the unambiguous assignment reflects the magnetization transfer shown by the cyan arrows in Figure 5.1 at top left.

[3] With assistance from A. Otter, M. Miskolzie.

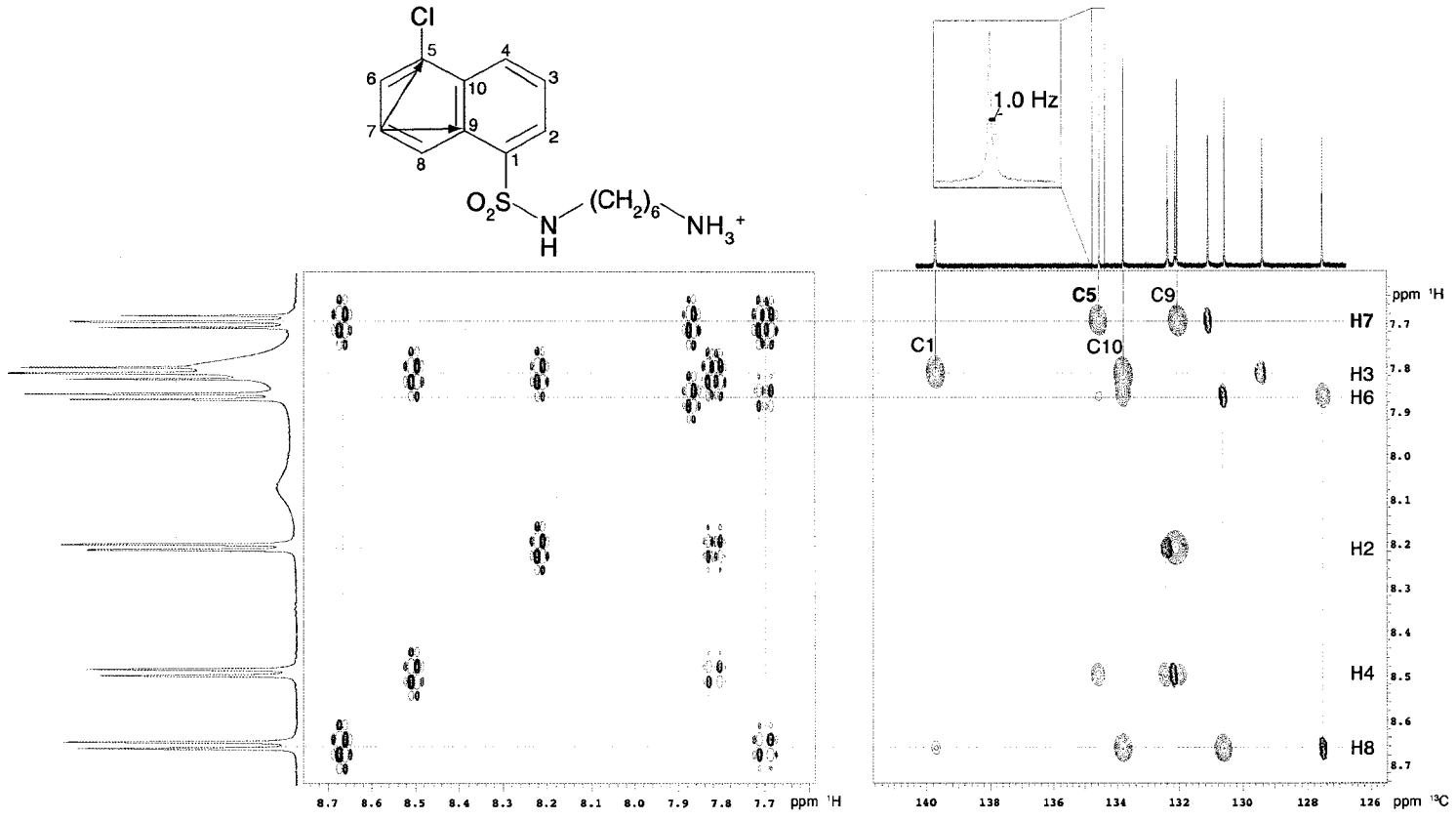


Figure 5.1: The chemical structure of W7 is shown with numbering for the aromatic carbons (and their attached hydrogens), at top left. The molecule has two ^{1}H spin systems, one on carbons 2, 3 and 4, the other on carbons 6, 7 and 8. Correlations in the chlorine-coupled spin system are shown with green lines; correlation in the sulfonamide-coupled spin system are in orange. The cyan arrows illustrate a key HMBC correlation from H7 to C5 and C9, which are both unprotonated. The HMBC spectrum, at bottom right, shows correlations between carbon atoms and protons located three bonds away (because in an aromatic ring, the $^3J_{CH}$ values are usually larger than $^2J_{CH}$ [146, p. 248.]). The natural abundance $\{^1H,^{13}C\}$-HSQC spectrum is overlayed onto the HMBC (blue peaks). The identity of C5 is directly disclosed from the observation of a ^{35}Cl–^{13}C isotope shift in the natural abundance ^{13}C spectrum (above the HMBC). The other correlation to C5, from H4, cannot be confused with the H7 correlation because only the latter is unambiguously coupled to two spins in the DQF-COSY (at bottom left). To the right of the DQF-COSY is the aromatic region of the ^{1}H spectrum. The two broad peaks not represented in the DQF-COSY originate from the amide and amine protons of the W7 tail moiety.

5.2 Assignments of W7 in the bound state

Assignments of free W7 were transferred to the bound state through monitoring the titration of cNTnC•Ca^{2+}into W7 (Section 4.3.5, p. 67) as shown in Figure 5.2. The subspectra shown in Figure 5.2 have the assignments for H3 and H6 lying on the same peak. Examining the titration that proceeded closer to saturation (Figure 4.15, p. 67) shows that the two peaks may be resolved in the bound state. The structure calculations presented here treats those assignments as ambiguous.

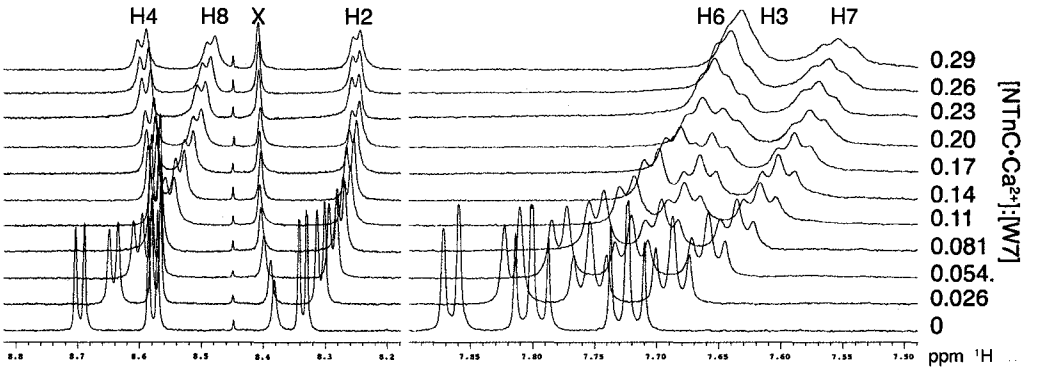


Figure 5.2: Assignments for W7 in the unbound state (Figure 5.1) were transferred to those of the bound state by monitoring the titration with NMR. The peak labeled 'X' is from imidazole and is highly pH dependent. Please note the bipartite axis; the left and right halves are scaled differently; vertical scaling is the same for both halves. This is a subspectrum of the titration presented in Section 4.3.5 (p. 67, Figure 4.16).

5.3 Determination of intermolecular contacts

Protein assignments can be transferred from the free state to the bound state through the same approach as used for W7. The HSQC spectra used to monitor the titration presented in Section 4.3 (p. 59) allowed for a number of protein atoms to be unambiguously assigned.

Intermolecular NOEs were obtained using a three-dimensional, ^{13}C-edited, ^{12}C-filtered, transferred NOESY experiment [147] included in Biopack (Varian Inc.). A chief advantage of this pulse sequence is that the directly detected dimension has no ^{13}C enrichment, and does not require ^{13}C decoupling, allowing long acquisition times and therefore high digital resolution. The spectrum was acquired at 800 MHz. Sweep widths were 8000.000 Hz in the directly detected (W7) dimension, 7998.054 Hz in the indirect (protein) dimension, and 4022.335 in the indirect (protein ^{13}C) dimension. There were 2048, 160, and 64 complex points acquired in the direct ^{1}H, indirect ^{1}H, and indirect ^{13}C dimensions.

Processing was performed with NMRPipe [148]. Residual water was removed from the free induction decay by matching it to a time-domain polynomial. Strong apodization was required because the long acquisition time resulted in too much incorporation of noise into the indirectly detected dimensions. The direct dimension was apodized with a unshifted squared cosine function,

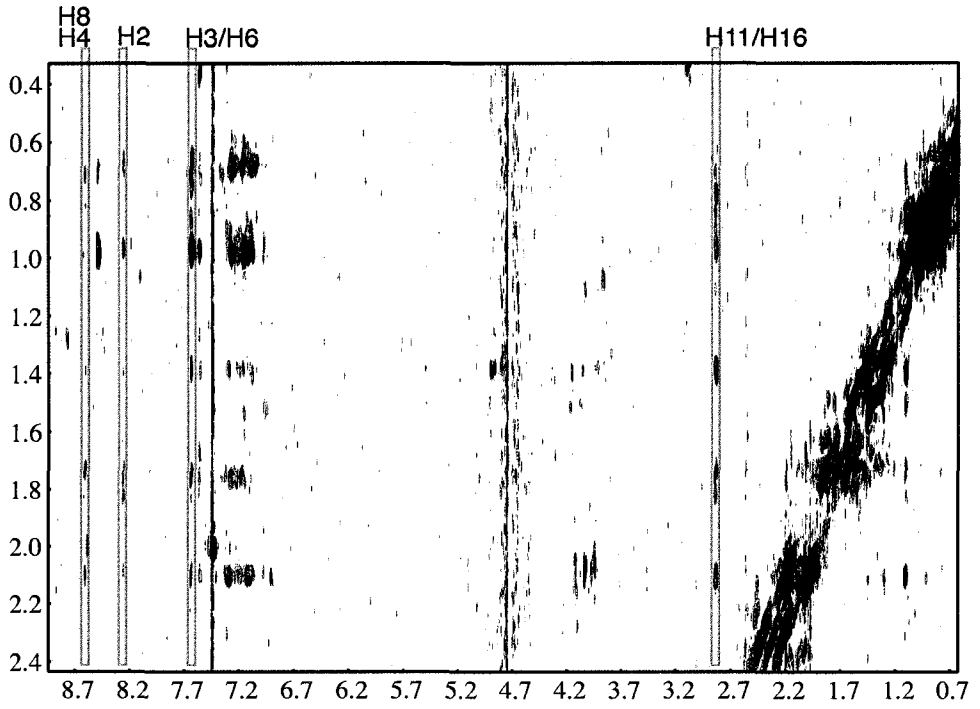


Figure 5.3: Intermolecular NOEs were observed through a $\{^{13}C,^{1}H\}$-edited,$\{^{12}C,^{1}H\}$-filtered NOESY experiment. Here, the spectrum is shown with the ^{13}C dimension collapsed. Signals with positive amplitude (potential information) are black, negative signals (noise or artifacts) are in red. A number of 'bleed through' signals that originate from the protein but survive the ^{12}C filter occur in the aromatic region of the spectrum. To assign this data, peaks at the ligand frequencies were exclusively considered (in green boxes).

and a 5 Hz linebroadening was imposed with an exponential window. Forward-back linear prediction was used in the indirect dimensions.

Spectral assignment was facilitated with NMRView and NMRViewJ. The residues involved in the 22 contacts that were assigned are highlited in Figure 5.7.

5.3.1 Virtualization of W7

The computational representation of W7 was defined with the HIC-Up server (http://xray.bmc.uu.se-/hicup/) [149] and XPLO2D (ftp://xray.bmc.uu.se/pub/gerard/xutil/) [150] and were manually checked. Xplor-NIH does not incorporate the W7 DIHEdral energy terms from the parameter files, leading to a lack of planarity in the naphthalene moiety of W7. I demonstrated this to myself by obtaining identical results with the DIHEdral terms commented out. Consultation with the primary author and maintainer of Xplor-NIH, Charles Schwieters, lead to the decision to restrain the aromatic-ring-atom-planarity directly with a patch to the IVM module. There is a crystallographic structure determination of W7 which does not describe deviations from this target geometry [151]. The coordinate,

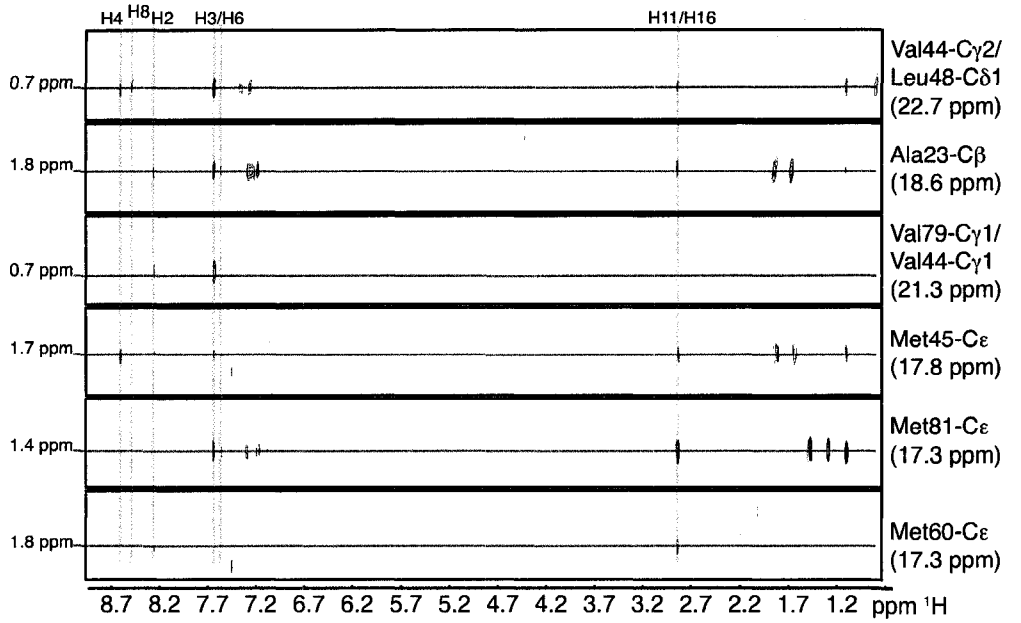


Figure 5.4: Assigned regions of the collapsed spectrum shown in Figure 5.3. The vertical green lines show the W7 assignments. Horizontal cyan lines show the protein ^{1}H chemical shifts for the assigned signals. The carbon chemical shifts are given at right and correspond to the entire subspectrum in each box.

topology, and parameter files are included in Appendix G.2.

The amine group of W7 is expected to be protonated (cationic) in solution. The virtual representation used here is unprotonated. As electrostatic forces are not modeled in the structure determination protocol, the presence or absence of a charged group does not impact the results. The structure presented could benefit from additional refinement, including the consideration of electrostatic, polarization, and solvation effects.

5.4 Empirical target functions

Structure determination by NMR typically consists of a molecular dynamics simulation in the presence of non-physical potentials (pseudopotentials). The pseudopotentials serve to 'steer' the molecular dynamics towards solutions that are consistent with the data. Two types of pseudopotentials were used here. The first type imparts distance restraints, the second type keeps the backbone dihedral angles in the allowed regions of configurational (Ramachandran) space. In Xplor-NIH, the latter potential is called a RAMA potential [152].

NOEs are often specified in terms of groups of atoms (for example, the three protons of a methyl residue are indistinguishable). All protein atoms are explicitly modeled at all points in the simu-

lation. As such, a formula is required to convert the known distances in the simulation into NOE-equivalent distances. Sum averaging is the most popular approach to this because it can treat both unambiguous and ambiguous (more than one pair of atoms being stipulated) restraints.

There are two classes of distance restraints in the calculation. Restraints derived from the cNTnC•Ca^{2+}–W7 NOEs were applied with a 'hard' cutoff so that they would be dominant in the calculation. The reason for this is that, should ligand-binding distort the protein structure, I wanted the results to reflect the the inappropriateness of my choice of initial coordinates. Distance restraints that are taken from other studies (the intraprotein [137] and protein–Ca^{2+}restraints [153]) were applied with a 'soft' cutoff for the same reason. The hard potential has the energy scaling quadratically with the violated distance; the soft potential increases more slowly with increasingly violated distance restraints. The script to generate the restraints is shown in Appendix F.1. The distance restraints are summarized in Appendix F.2.

5.5 Structure determination by restrained docking

A simulated annealing script, appropriate for NOE-driven docking, was derived from the Xplor-NIH example script, gb1-anneal.py. The script is reproduced in Appendix G.1. The protein topology was generated from its sequence using the Protocol module of Xplor-NIH. Initial coordinates were derived from chain 1 of the cNTnC•Ca^{2+}bepridil NMR ensemble (PDB 1LXF) that had been manually edited to conform to the correct topology, changing Cys35 and Cys85 to serines using a text editor.

All molecular dynamics and energy minimizations were performed with Xplor-NIH's IVM module [154] against all terms of the target function.[4] Dynamics were performed in torsion space except for the final minimization. Initial coordinates of W7 and NTnC were separated by more than 80 Å; an initial docking at high temperatures (2500K) for 200 ps (2000 timesteps)[5] Protein backbone atoms were held fixed; all other atoms had random velocities assigned from a Maxwell distribution. Next, the temperature was dropped in 100K steps to 25K. At each step, the system equilibrated for 0.2 ps (100 steps). In this manner, W7 was docked onto cNTnC•Ca^{2+}. The solution was further refined within the protocol. The docked complex was heated to 1500K and cooled to 25K in 12.5K steps to improve convergence of sidechain conformations.

Until this point in the protocol the system has had no opportunity to relax the protein backbone coordinates. The velocities from the previous dynamics runs were inputed, and the system was equilibrated at 25 K for 200 ps (2000 timesteps) with no protein atoms held fixed. The final coordinates were then minimized with all atoms free. A final minimization in Cartesian mode was performed with backbone atoms held fixed. The script is presented in Appendix G.1.

[4]The terms are: ANGLe, DIHEdral, IMPRoper, BOND, VDW (all restraining covalent geometry) and the RAMAchandran and NOE terms (empirical terms to restrain the protein conformation).

[5]The IVM module adjusts timestep length to keep the magnitude of rescaling roughly constant with each step. In the current case, the simulation proceeded for either 200 ps total or 2000 timesteps, whichever came first.

Figure 5.5: Convergent stereogram of the 10 lowest-energy conformers of 19 docked complexes shown as a C_α trace with W7 shown as sticks. Ca^{2+}, the Cl atoms of W7, and the terminal nitrogen atoms of W7 are shown with spheres. Non-carbon atoms are colored blue (nitrogen), red (oxygen), green (chlorine), and orange (sulfur). The two binding poses of W7 can be landmarked using the red oxygens of the sulfonamide moiety. The N-terminal helix (helix N) of cNTnC is at the top of the figure; reading towards the C-terminus in a roughly clockwise fashion the four helices are named A–D. Helix D is pointing out of the page.

5.5.1 Quality of the NMR ensemble

Of the 50 structures calculated, 46 had no violations in the protein–W7 NOE term (I sometimes call this *the docking term*) reflecting completely satisfied dockings but also had wide-ranging energies, some being quite high. Of those, even the highest energy structure had almost no violations of the NOE terms. Rather, deviation from target covalent geometries accounts for all of the high energy structures. This shows that the protocol works very well for docking the ligand so as to satisfy the intermolecular NOEs, but that it can be improved with respect to energy minimizing the protein coordinates. W7 was seen to occupy two distinct orientations within the binding site. At this point, the intermolecular NOEs were set to an upper distance of 6.1 Å. The calculation was repeated after decreasing the upper distance to 5.1 Å. This manipulation raised the stringency of the docking term, and was expected to improve the convergence of the ligand. With the increased stringency, 19 of the 50 structures had no violations in the NOE term. 20 of the structures had one violated restraint, and the remainder had 2-3 restraints violated. The upper bound was again changed to 4 Å and the calculation was repeated. This proved to be too stringent of an upper bound to allow for satisfaction of the restraints; none of the solutions had zero violations in the docking term. The 5.1 Å calculation was selected for presentation here.

The 10 lowest energy structures were further analyzed; below these structures are referred to as *the ensemble.* Figure 5.5 shows the ensemble. The backbone RMS over all 89 residues is 1.32

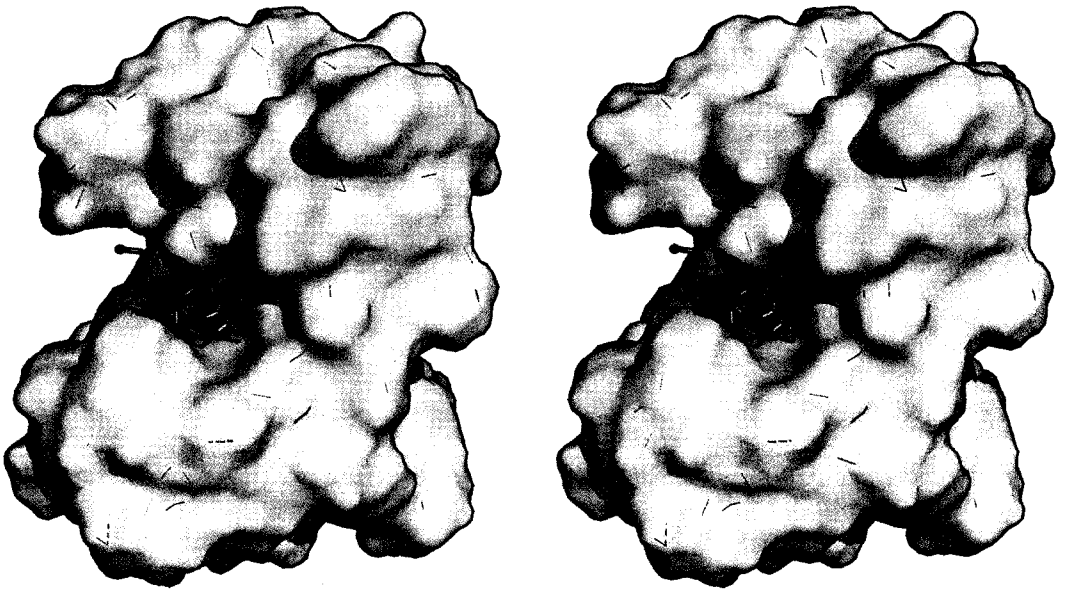


Figure 5.6: The surface of the ensemble is depicted; heteroatoms are colored as in Figure 5.5. The complex is oriented so that helices A and B are facing the reader; A is pointing upwards (See Figure 5.5). W7's binding site is very well-defined despite the presence of two bound poses.

Å (ranging from 0.86–1.76 Å), aligning them to the lowest-energy conformer. This is a similar precision to the deviation from the initial backbone coordinates: 1.42 Å (ranging from 1.16–1.70 Å). Backbone dihedral angles fall mostly into the most favored region of Ramachandran space (92.5% of residues), with 7.1% being in the additionally allowed, 0.4% being in the generously allowed region, and no residues being in the prohibited region, as determined with ProcheckNMR [155].

A feature of this experimental design is that the intra-protein and protein–Ca^{2+}distance restraints have been previously used to determine a very precise NMR ensemble [137], in excellent agreement with a subsequently determined crystal structure [1]. They can be thought of as control restraints. As such, should ligand-binding deform the protein conformation beyond the previously solved conformation, the control restraints will 'report' the discrepancy by increasing the system's energy. The ensemble has energies varying from –360.84 to –514.03 kcal/mol (on average –423.13 kcal/mol), with almost all violations being in the non-covalent target parameters. There were no NOE violations in the Ca^{2+}-coordinating pseudopotential. Of the violations in the intraprotein NOE potential, two occur in only one structure. The other three restraints are frequently violated in the ensemble. One restraint is intra-residue, between the backbone amide hydrogen and the hydrogens of the C_γ of Arg83. The other two violated restraints are inter-residue but are 'medium range' (they involve atom pairs separated by less than three sequence indices). The violated medium range restraints are between the H_α of Ile26 and the H_γ of Leu29, and between the H_α of Val79 and H_β of Val82. None of the distances in the model disagree from the target values by more than 1Å. The intraprotein distance restraints are tightly calibrated, and so the results in general corroborate the accuracy of the docking.

Figure 5.7: Sidechains coordinating W7 are demarked with white surfaces, with each ensemble member represented. Heteroatoms are depicted as in Figure 5.5. Residues are labeled in red. The intermolecular contacts involve residues in helices A–D. The depth of W7 with respect to the hydrophobic pocket seems substantiated by the self-consistency of the contacts. Note that not all surfaces forming the binding pocket are "determined" (compare with Fig. 5.6).

5.5.2 The binding site and poses

Figure 5.7 shows the ensemble with a surface representation for the residue sidechains identified as contributing intermolecular contacts. Most of the binding interface is defined from contributions from these residues. They form a self-consistent description of a binding site disparate from previously determined complexes, as discussed further below. The naphthalene ring of W7 is stabilized into a narrow cleft formed exclusively of sidechain atoms.

The structures define a single binding site with two poses for W7's aromatic ring (Figure 5.5). Without establishing additional distance restraints to further constrain W7's binding pose, I do not know whether the two poses arise from lingering ambiguity in the data or whether they are actually nearly isoenergetic, equally populated microstates. Repeating the structure calculation with tighter distance restraints did improve the convergence of the ensemble, but did not disambiguate the solution.

5.6 Structure-function relationships

Naphthalenesulfonamides, TFP, bepridil, EMD57033, and levosimendan all have polar and hydrophobic groups; they have varying degrees of charge density on the polar groups. Towards defining generalized cardiotonic structure-activity relationships, several cTnC•ligand complexes have been solved to high precision. The structures of cNTnC•Ca^{2+}•bepridil•cTnI$_{147-163}$ [136] and cCTnC•-

$2Ca^{2+}$•EMD57033 [156] both visualize similar protein-drug interactions, albeit in different domains of cTnC. Bepridil and EMD57033 both bind to a homologous hydrophobic pocket nestled between two EF hands. The generality of this binding site is further implicated by the X-ray structure of the Ca^{2+}-activated skeletal troponin complex [1]. In that study, the polyoxyethylene detergent anapoe binds to the N-domain of TnC, in an analogous location to bepridil in cNTnC•Ca^{2+}•-bepridil•$cTnI_{147-163}$. Anapoe is reported to be essential for the production of sufficiently ordered troponin crystals to allow for diffraction to high resolutions. The authors surmise that ligand binding stabilizes unique microstates of cTnC's native ensemble, and report that anapoe enhances the contractility of skinned, permeabilized muscle fibers [1].

Figure 5.8: The two lowest energy ensemble members (which reflect the two binding poses for W7) are shown superimposed, over the backbone atoms of helices A–D (RMSD $\sim$ 1.7 Å), upon the initial coordinates. The ensemble members are colored in grey. Heteroatoms are depicted as in Figure 5.5. The first conformer of the calmodulin•$4Ca^{2+}$•2W7 complex (PDB 1MUX, carbon atoms colored violet) was superimposed upon the initial coordinates (RMSD = 1.64 Å). The first conformer of the cNTnC•Ca^{2+}•bepridil•$TnI_{147-163}$ structure (PDB 1LXF, carbon atoms colored cyan) was aligned similarly (RMSD = 1.45 Å). W7 in the calmodulin complex, and bepridil in the presence of $TnI_{147-163}$, colocalize to a position other than W7 in the ensemble.

The drug binding sites in cTnC have subtle structure-activity relationships. Both computational [138] and experimental [157,158] studies have concluded that multiple binding sites or modes could exist for several cTnC•ligand interactions. A study by Kleerekoper and others suggested that there are two drug binding sites in cNTnC's hydrophobic pocket, with one site (demarked by Met 45, 60, and 80) conferring Ca^{2+}-sensitizing effects, and the other site (containing Met 47, 81 and 85) also being capable of coordinating ligand, but being capable of occluding $cTnI_{147-163}$ and therefore conferring Ca^{2+}desensitizing effects [157]. The present results show that the binding site of W7 in

the absence of the switch region is not localized to an analogous position to that of the CaM complex (Figure 5.8).

As concluded in my earlier work [12], the spatial distribution of binding-perturbed signals directly implies the primary binding sites for W7 in cNTnC•Ca^{2+} to be the functionally important hydrophobic pocket. This is a suitable location to allow W7 to disrupt interactions with cTnI as proposed by Adhikari and Wang [124]. Adhikari and Wang consider (and exclude) the interpretation that another protein, including myosin light chain kinase, is W7's target; this possibility is not addressed by my PhD work. The dissociation constant for $cTnI_{147-163}$ binding cNTnC•Ca^{2+} is 0.154±0.01 mM [137] – similar to the affinity of W7 for cNTnC•Ca^{2+} determined previously (dissociation constant of 0.22 mM) [12].

Figure 5.9: The switch peptide (cardiac $TnI_{147-163}$, in blue) is shown with cNTnC•Ca^{2+}(in red, oriented with helix N towards the reader); both protein chains are from the structure 1MXL. The ligands from Figure 5.8 are shown with the same colorings. In the ensemble (grey carbons) W7 is in a position to sterically clash with the switch peptide. Bepridil (cyan carbons), which is bound in the presence of switch peptide, does not seem to occupy the same binding site as W7 in the ensemble.

W7 binding to cTnC interferes with the binding of cTnI, increasing the dissociation constant for the switch region from 0.14 mM (Section 4.3.3, p. 64) to 0.5 mM [13]. As previously asserted, if most Met residues in cNTnC simultaneously interact with a single ligand, this would likely disallow cTnI binding [157]. The structures certainly suggest that when the switch region binds it would sterically clash with W7 (Figure 5.9).

Chapter 6

Conclusion

I have presented a process-oriented description of troponin, distilling three of my publications [9–11]) into a single account of how troponin functions (Chapter 3, p. 18). First I summarized the now extensive evidence against rigidity in the DE linker (Section 3.2, p. 19). Borrowing from Maeda and others' description of this region as a "universal joint" [45], and with knowledge of the switching-induced change in the stability of this region [68], I incorporated a disorder-to-order transition of this region into a mechanistic role (Figure 3.5, p. 35).

My main contribution to the fly-casting mechanism of troponin was the motivation and proposal (Section 3.5.1, p. 34) and explicit modeling (Figure 3.6, p. 36) of distinct conformational cascades for the processes of activation and inhibition. In general, when I propose mechanisms from data (be they mine or others'), the best outcome is that the work is read, and the derivative hypotheses evaluated. I would like to advertise that others have already considered my work sufficiently meritorious to frame their interpretations within it.

The muscle physiologists Gabriele Pfitzer and Robert Stehle, when characterizing the phenotype of the Lys184 deletion mutant of murine cardiac TnI, write [159]:

> "Considering [the mechanism] by which the initially intrinsically unfolded Md at high [Ca^{2+}] nucleates by binding to actin (flycasting hypothesis) upon Ca^{2+}-removal, Lys184 deletion could interfere with the efficient flycasting activity. As a consequence, a possible scenario is that the kinetics of the complete transfer of [the inhibitory region of troponin I] to actin (switch-off) can be indirectly disturbed by the mutation in Md, but not the reverse transition towards cTnC (switch-on) upon Ca^{2+}-binding to cTnC.
>
> "The hypothesis is supported by considering both, our present and previous functional results...
>
> "... Thus, our findings are in line with the hypothesis that perturbations of folding of the Md primarily affect relaxation kinetics. Following a similar way of thinking and considering previous results ... we assume that a perturbation of the folding-unfolding kinetics of the [inhibitory region] (e.g. by the mutation R146G) would impair both relaxation and Ca^{2+} activation kinetics."

The publication disseminating the most recent electron-micrograph reconstruction of the thin filament from the William Lehman, Larry Tobacman and coworkers states, "Our results support the hypothesis that TnI binding and folding on actin–tropomyosin are intimately involved in the mechanism of muscle regulation" [160] and then cites my mechanism. I think the weight of evidence handsomely favors the existence of a disorder-to-order transition in the function of the C-terminal region of TnI, and the central linker of TnC. But the fly-casting mechanism involves more than the invocation of a disorder-to-order transition, it also denotes a kinetic speedup from this transition. Falsification of this mechanism is nontrivial but possible, and much of my work towards this end is inconclusive at this juncture and unattractive for presentation in my Thesis.

I have discovered and reported the isoform-specific variation in the intrinsic disorder of TnI (Section 3.4, p. 23), and have developed novel conformational cascades to contextualize the findings and to rationalize the known functional properties of the isoforms (Figure 3.7, p. 40). Both of the novel cascades introduce means for the intrinsic disorder of troponin I to modulate the rates of conformational switching. The region of coupling between troponin I and troponin T is identified as intrinsically disordered, and this disorder is implicated to be highly isoform-specific. As troponin T yokes troponin to tropomyosin, and as myosin-induced shifts in tropomyosin are known to modulate the apparent Ca^{2+}-sensitivities of the troponin complex, these cascades break new mechanistic ground. I am unable to correlate isoform-specific dependence of intrinsic disorder with functional properties (Section 3.5.1, p. 35) owing in large part to isoform-specific variations in troponin T and tropomyosin. The intrinsic disorder of the proposed flycasting regions in the C-terminus of troponin I are strongly corroborated by these findings. The disorder of the switching region is less clearly addressed by the results.

I then describe a higher-order troponin process: the drug-induced modulation of troponin activity. Chapter 4 (p. 42) summarizes a number of binding studies in the general context of the study of weak binding, and the system-specific context. The drug-like naphthalenesulfonamide W7 is characterized in a number of solution equilibria with troponin C. Most of the chapter is centered on reconciling local and global descriptions of the process of binding.

I have not found a statistical procedure for selecting the appropriate subset of local signals for the maximally robust reporting of the global event. I show that taking the path of maximal ignorance — considering all of the local signals — is a more rewarding practise than manually selecting the signals for analysis. This is becoming standard practise in my mentor's laboratory, partly as a result of my first publication [12]. I show that normalization of the data is valuable in the context of global fitting as it eliminates a degree of freedom, per signal, in the final model (Section 4.2.9, p. 51 and Section 4.3.4, p. 65). I develop and apply a likelihood analysis that allows for establishing meaningful confidence intervals in the global fit (Section 4.3.3, p. 64). I propose and explore, but do not prove in any general sense, the possibility that the sampling power afforded by multisignal reporting of binding allows for accurate estimation of binding constants from the pre-asymptotic phase of the binding curve.

My Thesis concludes, appropriately enough, with a structure determination of the end state of a binding process. The structure of the N-terminal domain of cardiac troponin C, in its Ca^{2+}-activated state and in complex with W7, is determined with a minimal set of NMR-derived intermolecular

distance restraints (Chapter 5, p. 71). This unpublished result shows W7 localizing to a different binding site than was predicted from analogous structures. The position of W7 in the bound state is favorable for sterically occluding the binding of the switch region of TnI, an illuminating finding considering the previously assigned function of W7 as an inhibitor of striated muscle contraction.

Bibliography

[1] Vinogradova MV, Stone DB, Malanina GG, Karatzaferi C, Cooke R, Mendelson RA, & Fletterick RJ. Ca2+-regulated structural changes in troponin. *Proc Natl Acad Sci U S A*, **102**(14):5038 (2005).

[2] Murakami K, Yumoto F, Ohki SY, Yasunaga T, Tanokura M, & Wakabayashi T. Structural Basis for Ca(2+)-regulated Muscle Relaxation at Interaction Sites of Troponin with Actin and Tropomyosin. *J Mol Biol*, **352**(1):178 (2005).

[3] Schrödinger E. *What is life? : the physical aspect of the living cell ; with, Mind and matter ; & Autobiographical sketches.* Cambridge University Press, Cambridge (1992).

[4] Frauenfelder H. Proteins: paradigms of complexity. *Proc Natl Acad Sci U S A*, **99 Suppl 1**:2479 (2002).

[5] Plotkin SS & Onuchic JN. Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q Rev Biophys*, **35**(2):111 (2002).

[6] Plotkin SS & Onuchic JN. Understanding protein folding with energy landscape theory. Part II: Quantitative aspects. *Q Rev Biophys*, **35**(3):205 (2002).

[7] Wolynes PG. Energy landscapes and solved protein-folding problems. *Philos Transact A Math Phys Eng Sci*, **363**(1827):453 (2005).

[8] Del Re G. Ontological status of molecular structure. *HYLE*, **4**:81 (1998).

[9] Hoffman RMB, Blumenschein TMA, & Sykes BD. An interplay between protein disorder and structure confers the Ca^{2+} regulation of striated muscle. *J Mol Biol*, **361**(4):625 (2006).

[10] Hoffman RMB & Sykes BD. *Regulatory Mechanisms of Striated Muscle Contraction*, vol. 592 of *Adv Exp Med Biol*, chap. Disposition and dynamics: interdomain orientations in troponin., pp. 59–70. Springer-Verlag Press (2007).

[11] Hoffman RMB & Sykes BD. Isoform-specific variation in the intrinsic disorder of troponin I. *Proteins*, **73**(2):338 (2008 Apr 23).

[12] Hoffman RMB, Li MX, & Sykes BD. The binding of W7, an inhibitor of striated muscle contraction, to cardiac troponin C. *Biochemistry*, **44**(48):15750 (2005).

[13] Li MX, Hoffman RMB, & Sykes BD. Interaction of cardiac troponin C with W7 in the presence of three functional regions of cardiac troponin I. *Biochemistry*, **45**(32):9833 (2006).

[14] Martonosi AN. Animal electricity, Ca2+ and muscle contraction. A brief history of muscle research. *Acta Biochim Pol*, **47**(3):493 (2000).

[15] Piccolino M. Luigi Galvani and animal electricity: two centuries after the foundation of electrophysiology. *Trends Neurosci*, **20**(10):443 (1997).

[16] Gordon AM, Homsher E, & Regnier M. Regulation of contraction in striated muscle. *Physiol Rev*, **80**(2):853 (2000).

[17] Tobacman LS. Thin filament-mediated regulation of cardiac contraction. *Annu Rev Physiol*, **58**:447 (1996).

[18] Best PM. Cardiac muscle function: results from skinned fiber preparations. *Am J Physiol*, **244**(2):H167 (1983 Feb).

[19] Endo M. Calcium ion and troponin: Professor S. Ebashi's epoch-making achievement. *Biochem Biophys Res Commun*, **369**(1):30 (2008).

[20] Murray AC & Kay CM. Separation and characterization of the inhibitory factor of the troponin system. *Biochem Biophys Res Commun*, **44**(1):237 (1971).

[21] Murray AC & Kay CM. Hydrodynamic and optical properties of troponin A. Demonstration of a conformational change upon binding calcium ion. *Biochemistry*, **11**(14):2622 (1972).

[22] Greaser ML & Gergely J. Purification and properties of the components from troponin. *J Biol Chem*, **248**(6):2125 (1973).

[23] McCubbin WD, Mani RS, & Kay CM. Physicochemical studies on the interaction of the calcium-binding protein (troponin C) with the inhibitory protein (troponin I) and calcium ions. *Biochemistry*, **13**(13):2689 (1974).

[24] Syska H, Wilkinson JM, Grand RJ, & Perry SV. The relationship between biological activity and primary structure of troponin I from white skeletal muscle of the rabbit. *Biochem J*, **153**(2):375 (1976).

[25] Potter JD & Gergely J. Troponin, tropomyosin, and actin interactions in the Ca2+ regulation of muscle contraction. *Biochemistry*, **13**(13):2697 (1974).

[26] Ebashi S. Regulatory mechanism of muscle contraction with special reference to the Ca-troponin-tropomyosin system. *Essays Biochem*, **10**:1 (1974).

[27] Heeley DH, Golosinska K, & Smillie LB. The effects of troponin T fragments T1 and T2 on the binding of nonpolymerizable tropomyosin to F-actin in the presence and absence of troponin I and troponin C. *J Biol Chem*, **262**(21):9971 (1987).

[28] Pearlstone JR & Smillie LB. The interaction of rabbit skeletal muscle troponin-T fragments with troponin-I. *Can J Biochem Cell Biol*, **63**(3):212 (1985).

[29] Tripet B, Van Eyk JE, & Hodges RS. Mapping of a second actin-tropomyosin and a second troponin C binding site within the C terminus of troponin I, and their importance in the Ca2+-dependent regulation of muscle contraction. *J Mol Biol*, **271**(5):728 (1997).

[30] Sykes BD. Pulling the calcium trigger. *Nat Struct Biol*, **10**(8):588 (2003).

[31] Perry SV. Troponin T: genetics, properties and function. *J Muscle Res Cell Motil*, **19**(6):575 (1998).

[32] Layland J, Solaro RJ, & Shah AM. Regulation of cardiac contractile function by troponin I phosphorylation. *Cardiovasc Res*, **66**(1):12 (2005).

[33] Herzberg O & James MNG. Structure of the calcium regulatory muscle protein troponin-C at 2.8 A resolution. *Nature*, **313**(6004):653 (1985).

[34] Sundaralingam M, Bergstrom R, Strasburg G, Rao ST, Roychowdhury P, Greaser M, & Wang BC. Molecular structure of troponin C from chicken skeletal muscle at 3-angstrom resolution. *Science*, **227**(4689):945 (1985).

[35] Herzberg O, Moult J, & James MNG. A model for the Ca2+-induced conformational transition of troponin C. A trigger for muscle contraction. *J Biol Chem*, **261**(6):2638 (1986).

[36] Gagne SM, Tsuda S, Li MX, Smillie LB, & Sykes BD. Structures of the troponin C regulatory domains in the apo and calcium-saturated states. *Nat Struct Biol*, **2**(9):784 (1995).

[37] McKillop DF & Geeves MA. Regulation of the interaction between actin and myosin subfragment 1: evidence for three states of the thin filament. *Biophys J*, **65**(2):693 (1993).

[38] Geeves MA & Lehrer SS. Dynamics of the muscle thin filament regulatory switch: the size of the cooperative unit. *Biophys J*, **67**(1):273 (1994).

[39] Phillips GNJ, Fillers JP, & Cohen C. Tropomyosin crystal structure and muscle regulation. *J Mol Biol*, **192**(1):111 (1986).

[40] Vibert P, Craig R, & Lehman W. Steric-model for activation of muscle thin filaments. *J Mol Biol*, **266**(1):8 (1997).

[41] Pirani A, Vinogradova M, Curmi P, King W, Fletterick R, Craig R, Tobacman L, Xu C, Hatch V, & Lehman W. An Atomic Model of the Thin Filament in the Relaxed and Ca(2+)-Activated States. *J Mol Biol*, **357**(3):707 (2006).

[42] Poole KJV, Lorenz M, Evans G, Rosenbaum G, Pirani A, Craig R, Tobacman LS, Lehman W, & Holmes KC. 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**(2):273 (2006).

[43] Lehman W, Hatch V, Korman V, Rosol M, Thomas L, Maytum R, Geeves MA, Van Eyk JE, Tobacman LS, & Craig R. Tropomyosin and actin isoforms modulate the localization of tropomyosin strands on actin filaments. *J Mol Biol*, **302**(3):593 (2000).

[44] de Tombe PP, Belus A, Piroddi N, Scellini B, Walker JS, Martin AF, Tesi C, & Poggesi C. Myofilament calcium sensitivity does not affect cross-bridge activation-relaxation kinetics. *Am J Physiol Regul Integr Comp Physiol*, **292**(3):R1129 (2007).

[45] Takeda S, Yamashita A, Maeda K, & Maeda Y. Structure of the core domain of human cardiac troponin in the Ca(2+)-saturated form. *Nature*, **424**(6944):35 (2003).

[46] Lehman W, Rosol M, Tobacman LS, & Craig R. Troponin organization on relaxed and activated thin filaments revealed by electron microscopy and three-dimensional reconstruction. *J Mol Biol*, **307**(3):739 (2001).

[47] Lehman W, Craig R, & Vibert P. Ca(2+)-induced tropomyosin movement in Limulus thin filaments revealed by three-dimensional reconstruction. *Nature*, **368**(6466):65 (1994).

[48] Lehman W, Vibert P, Uman P, & Craig R. Steric-blocking by tropomyosin visualized in relaxed vertebrate muscle thin filaments. *J Mol Biol*, **251**(2):191 (1995).

[49] Blumenschein TMA, Stone DB, Fletterick RJ, Mendelson RA, & Sykes BD. Dynamics of the C-terminal region of TnI in the troponin complex in solution. *Biophysical Journal*, **90**(7):2436 (2006).

[50] Flaugh SL & Lumb KJ. Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27(Kip1). *Biomacromolecules*, **2**(2):538 (2001).

[51] Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett*, **579**(15):3346 (2005).

[52] Welch GR. *The Fluctuating enzyme*, vol. v. 5. Wiley, New York (1986).

[53] Frauenfelder H, Sligar SG, & Wolynes PG. The energy landscapes and motions of proteins. *Science*, **254**(5038):1598 (1991).

[54] Bokor M, Csizmók V, Kovács D, Bánki P, Friedrich P, Tompa P, & Tompa K. NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. *Biophys J*, **88**(3):2030 (2005).

[55] Dosztányi Z, Csizmók V, Tompa P, & Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol*, **347**(4):827 (2005).

[56] Dosztányi Z, Csizmók V, Tompa P, & Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**(16):3433 (2005).

[57] Tompa P, Szász C, & Buday L. Structural disorder throws new light on moonlighting. *Trends Biochem Sci*, **30**(9):484 (2005).

[58] Dosztányi Z, Chen J, Dunker AK, Simon I, & Tompa P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res*, **5**(11):2985 (2006).

[59] Hilser VJ & Thompson EB. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc Natl Acad Sci U S A*, **104**(20):8311 (2007).

[60] Lee AL, Kinnear SA, & Wand AJ. Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat Struct Biol*, **7**(1):72 (2000).

[61] Lee AL & Wand AJ. Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature*, **411**(6836):501 (2001).

[62] Wand AJ. Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nat Struct Biol*, **8**(11):926 (2001).

[63] Igumenova TI, Frederick KK, & Wand AJ. Characterization of the fast dynamics of protein amino Acid side chains using NMR relaxation in solution. *Chem Rev*, **106**(5):1672 (2006).

[64] Wand AJ. On the dynamic origins of allosteric activation. *Science*, **293**(5534):1395 (2001).

[65] Halle B. Flexibility and packing in proteins. *Proc Natl Acad Sci U S A*, **99**(3):1274 (2002).

[66] Zhang F & Bruschweiler R. Contact model for the prediction of NMR N-H order parameters in globular proteins. *J Am Chem Soc*, **124**(43):12654 (2002).

[67] Vassylyev DG, Takeda S, Wakatsuki S, Maeda K, & Maeda Y. The crystal structure of troponin C in complex with N-terminal fragment of troponin I. The mechanism of how the inhibitory action of troponin I is released by Ca(2+)-binding to troponin C. *Adv Exp Med Biol*, **453**:157 (1998).

[68] Blumenschein TMA, Stone DB, Fletterick RJ, Mendelson RA, & Sykes BD. Calcium-dependent changes in the flexibility of the regulatory domain of troponin C in the troponin complex. *J Biol Chem*, **280**(23):21924 (2005).

[69] Soman J, Tao T, & Phillips GNJ. Conformational variaion of calcium-bound troponin C. *Proteins*, **37**(4):510 (1999).

[70] Wall ME, Clarage JB, & Phillips GN. Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. *Structure*, **5**(12):1599 (1997).

[71] Ferrieres G, Calzolari C, Mani JC, Laune D, Trinquier S, Laprade M, Larue C, Pau B, & Granier C. Human cardiac troponin I: precise identification of antigenic epitopes and prediction of secondary structure. *Clin Chem*, **44**(3):487 (1998).

[72] Martins SM, Chapeaurouge A, & Ferreira ST. Equilibrium unfolding and conformational plasticity of troponin I and T. *Eur J Biochem*, **269**(22):5484 (2002).

[73] Cheng Y, LeGall T, Oldfield CJ, Dunker AK, & Uversky VN. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry*, **45**(35):10448 (2006).

[74] Day S, Westfall M, & Metzger J. Tuning cardiac performance in ischemic heart disease and failure by modulating myofilament function. *J Mol Med* (2007).

[75] Davis JP, Norman C, Kobayashi T, Solaro RJ, Swartz DR, & Tikunova SB. Effects of thin and thick filament proteins on calcium binding and exchange with cardiac troponin C. *Biophys J*, **92**(9):3195 (2007).

[76] Westfall MV, Rust EM, & Metzger JM. Slow skeletal troponin I gene transfer, expression, and myofilament incorporation enhances adult cardiac myocyte contractile function. *Proc Natl Acad Sci U S A*, **94**(10):5444 (1997).

[77] Kentish JC, McCloskey DT, Layland J, Palmer S, Leiden JM, Martin AF, & Solaro RJ. Phosphorylation of troponin I by protein kinase A accelerates relaxation and crossbridge cycle kinetics in mouse ventricular muscle. *Circ Res*, **88**(10):1059 (2001).

[78] Fentzke RC, Buck SH, Patel JR, Lin H, Wolska BM, Stojanovic MO, Martin AF, Solaro RJ, Moss RL, & Leiden JM. Impaired cardiomyocyte relaxation and diastolic function in transgenic mice expressing slow skeletal troponin I in the heart. *J Physiol*, **517 (Pt 1)**:143 (1999).

[79] Li MX, Wang X, & Sykes BD. Structural based insights into the role of troponin in cardiac muscle pathophysiology. *J Muscle Res Cell Motil*, **25**(7):559 (2004).

[80] Foster DB & Van Eyk JE. In search of the proteins that cause myocardial stunning. *Circ Res*, **85**(5):470 (1999).

[81] Foster DB, Noguchi T, VanBuren P, Murphy AM, & Van Eyk JE. C-terminal truncation of cardiac troponin I causes divergent effects on ATPase and force: implications for the pathophysiology of myocardial stunning. *Circ Res*, **93**(10):917 (2003).

[82] Gaboriaud C, Bissery V, Benchetrit T, & Mornon JP. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Lett*, **224**(1):149 (1987).

[83] Eudes R, Le Tuan K, Delettre J, Mornon JP, & Callebaut I. A generalized analysis of hydrophobic and loop clusters within globular protein sequences. *BMC Struct Biol*, **7**:2 (2007).

[84] Ferron F, Longhi S, Canard B, & Karlin D. A practical overview of protein disorder prediction methods. *Proteins*, **65**(1):1 (2006).

[85] Eudes R, Le Tuan K, Delettre J, Mornon JP, & Callebaut I. A generalized analysis of hydrophobic and loop clusters within globular protein sequences. *BMC Struct Biol*, **7**:2 (2007).

[86] Lindhout DA, Boyko RF, Corson DC, Li MX, & Sykes BD. The role of electrostatics in the interaction of the inhibitory region of troponin I with troponin C. *Biochemistry*, **44**(45):14750 (2005).

[87] Ferreiro DU, Hegler JA, Komives EA, & Wolynes PG. Localizing frustration in native proteins and protein assemblies. *Proc Natl Acad Sci U S A*, **104**(50):19819 (2007).

[88] Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*, **11**(4):739 (2002).

[89] Uversky VN, Gillespie JR, & Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, **41**(3):415 (2000).

[90] Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, & Dunker AK. Comparing and combining predictors of mostly disordered proteins. *Biochemistry*, **44**(6):1989 (2005).

[91] Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, & Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, **21**(16):3435 (2005).

[92] Vucetic S, Brown CJ, Dunker AK, & Obradovic Z. Flavors of protein disorder. *Proteins*, **52**(4):573 (2003).

[93] Radivojac P, Obradovic Z, Smith DK, Zhu G, Vucetic S, Brown CJ, Lawson JD, & Dunker AK. Protein flexibility and intrinsic disorder. *Protein Sci*, **13**(1):71 (2004).

[94] Peng K, Radivojac P, Vucetic S, Dunker AK, & Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*, **7**:208 (2006).

[95] Bordoli L, Kiefer F, & Schwede T. Assessment of disorder predictions in CASP7. *Proteins*, **69 Suppl 8**:129 (2007).

[96] Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, & Obradovic Z. Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinform Comput Biol*, **3**(1):35 (2005).

[97] Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, & Dunker AK. Predicting intrinsic disorder from amino acid sequence. *Proteins*, **53 Suppl 6**:566 (2003).

[98] Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, & Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*, **337**(3):635 (2004).

[99] Mercier P, Li MX, & Sykes BD. Role of the structural domain of troponin C in muscle regulation: NMR studies of Ca2+ binding and subsequent interactions with regions 1-40 and 96-115 of troponin I. *Biochemistry*, **39**(11):2902 (2000).

[100] Weathers EA, Paulaitis ME, Woolf TB, & Hoh JH. Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein. *FEBS Lett*, **576**(3):348 (2004).

[101] Pontius BW. Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem Sci*, **18**(5):181 (1993).

[102] Pontius BW & Berg P. Rapid renaturation of complementary DNA strands mediated by cationic detergents: a role for high-probability binding domains in enhancing the kinetics of molecular assembly processes. *Proc Natl Acad Sci U S A*, **88**(18):8237 (1991).

[103] Shoemaker BA, Portman JJ, & Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A*, **97**(16):8868 (2000).

[104] Brenner B & Chalovich J. Kinetics of thin filament activation probed by fluorescence of N-((2-(Iodoacetoxy)ethyl)-N-methyl)amino-7-nitrobenz-2-oxa-1, 3-diazole-labeled troponin I incorporated into skinned fibers of rabbit psoas muscle: implications for regulation of muscle contraction. *Biophys J*, **77**(5):2692 (1999).

[105] Stehle R, Iorga B, & Pfitzer G. Calcium regulation of troponin and its role in the dynamics of contraction and relaxation. *Am J Physiol Regul Integr Comp Physiol*, **292**(3):R1125 (2007).

[106] Boussouf SE, Maytum R, Jaquet K, & Geeves MA. Role of tropomyosin isoforms in the calcium sensitivity of striated muscle thin filaments. *J Muscle Res Cell Motil*, **28**(1):49 (2007).

[107] Maytum R, Westerdorf B, Jaquet K, & Geeves MA. Differential regulation of the actomyosin interaction by skeletal and cardiac troponin isoforms. *J Biol Chem*, **278**(9):6696 (2003).

[108] Dargis R, Pearlstone JR, Barrette-Ng I, Edwards H, & Smillie LB. Single mutation (A162H) in human cardiac troponin I corrects acid pH sensitivity of Ca2+-regulated actomyosin S1 ATPase. *J Biol Chem*, **277**(38):34662 (2002).

[109] Westfall M & Metzger J. Single amino acid substitutions define isoform-specific effects of troponin I on myofilament Ca(2+) and pH sensitivity. *J Mol Cell Cardiol* (2007).

[110] Lavigne P, Bagu JR, Boyko R, Willard L, Holmes CF, & Sykes BD. Structure-based thermodynamic analysis of the dissociation of protein phosphatase-1 catalytic subunit and microcystin-LR docked complexes. *Protein Sci*, **9**(2):252 (2000).

[111] Fitzsimons DP & Moss RL. *Regulatory Mechanisms of Striated Muscle Contraction*, vol. 592 of *Adv Exp Med*, chap. Cooperativity in the regulation of force and the kinetics of force development in heart and skeletal muscles: cross-bridge activation of force., pp. 177–189. Springer-Verlag Press (2007).

[112] Endoh M. Mechanisms of action of novel cardiotonic agents. *J Cardiovasc Pharmacol*, **40**(3):323 (2002).

[113] Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, & Dunker AK. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res*, **32**(3):1037 (2004).

[114] Feldman L & Rouleau C. Troponin I inhibits capillary endothelial cell proliferation by interaction with the cell's bFGF receptor. *Microvasc Res*, **63**(1):41 (2002).

[115] Li Q, Liu Y, Shen PY, Dai XQ, Wang S, Smillie LB, Sandford R, & Chen XZ. Troponin I binds polycystin-L and inhibits its calcium-induced channel activation. *Biochemistry*, **42**(24):7618 (2003).

[116] Li Q, Shen PY, Wu G, & Chen XZ. Polycystin-2 interacts with troponin I, an angiogenesis inhibitor. *Biochemistry*, **42**(2):450 (2003).

[117] Karim CB, Zhang Z, Howard EC, Torgersen KD, & Thomas DD. Phosphorylation-dependent conformational switch in spin-labeled phospholamban bound to SERCA. *J Mol Biol*, **358**(4):1032 (2006).

[118] Howarth JW, Meller J, Solaro RJ, Trewhella J, & Rosevear PR. Phosphorylation-dependent conformational transition of the cardiac specific N-extension of troponin I in cardiac troponin. *J Mol Biol*, **373**(3):706 (2007).

[119] Borg M, Mittag T, Pawson T, Tyers M, Forman-Kay JD, & Chan HS. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc Natl Acad Sci U S A*, **104**(23):9650 (2007).

[120] Pufall MA, Lee GM, Nelson ML, Kang HS, Velyvis A, Kay LE, McIntosh LP, & Graves BJ. Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science*, **309**(5731):142 (2005).

[121] Baryshnikova OK, Li MX, & Sykes BD. Modulation of cardiac troponin C function by the cardiac-specific N-terminus of troponin I: influence of PKA phosphorylation and involvement in cardiomyopathies. *J Mol Biol*, **375**(3):735 (2008).

[122] Levijoki J, Pollesello P, Kaivola J, Tilgmann C, Sorsa T, Annila A, Kilpelainen I, & Haikala H. Further evidence for the cardiac troponin C mediated calcium sensitization by levosimendan: structure-response and binding analysis with analogs of levosimendan. *J Mol Cell Cardiol*, **32**(3):479 (2000).

[123] Hidaka H, Yamaki T, Naka M, Tanaka T, Hayashi H, & Kobayashi R. Calcium-regulated modulator protein interacting agents inhibit smooth muscle calcium-stimulated protein kinase and ATPase. *Mol Pharmacol*, **17**(1):66 (1980).

[124] Adhikari BB & Wang K. Interplay of troponin- and Myosin-based pathways of calcium activation in skeletal and cardiac muscle: the use of W7 as an inhibitor of thin filament activation. *Biophys J*, **86**(1 Pt 1):359 (2004).

[125] Osawa M, Swindells MB, Tanikawa J, Tanaka T, Mase T, Furuya T, & Ikura M. Solution structure of calmodulin-W-7 complex: the basis of diversity in molecular recognition. *J Mol Biol*, **276**(1):165 (1998).

[126] Craven CJ, Whitehead B, Jones SK, Thulin E, Blackburn GM, & Waltho JP. Complexes formed between calmodulin and the antagonists J-8 and TFP in solution. *Biochemistry*, **35**(32):10287 (1996).

[127] Gariepy J & Hodges RS. Localization of a trifluoperazine binding site on troponin C. *Biochemistry*, **22**(7):1586 (1983).

[128] Roberts GCK & Lian LY. *NMR of macromolecules: a practical approach*, vol. 134, chap. 6, pp. 153–182. IRL Press at Oxford University Press, Oxford (1993).

[129] Sia SK, Li MX, Spyracopoulos L, Gagne SM, Liu W, Putkey JA, & Sykes BD. Structure of cardiac muscle troponin C unexpectedly reveals a closed regulatory domain. *J Biol Chem*, **272**(29):18216 (1997).

[130] Cui Y, Wen J, Hung Sze K, Man D, Lin D, Liu M, & Zhu G. Interaction between calcium-free calmodulin and IQ motif of neurogranin studied by nuclear magnetic resonance spectroscopy. *Anal Biochem*, **315**(2):175 (2003).

[131] Williams T, Shelling J, & Sykes B. *NMR in the Life Sciences*, pp. 93–103. NATO ASI Series A: Life Sciences. Plenum Publishing Co. (1985).

[132] Jaynes ET & Bretthorst GL. *Probability theory: the logic of science*. Cambridge University Press, Cambridge, UK (2003).

[133] Sivia DS & Skilling J. *Data analysis: a Bayesian tutorial*. Oxford science publications. Oxford University Press, Oxford, 2nd edn. (2006).

[134] Turnbull WB & Daranas AH. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J Am Chem Soc*, **125**(48):14859 (2003).

[135] Sorsa T, Pollesello P, Permi P, Drakenberg T, & Kilpelainen I. Interaction of levosimendan with cardiac troponin C in the presence of cardiac troponin I peptides. *J Mol Cell Cardiol*, **35**(9):1055 (2003).

[136] Wang X, Li MX, & Sykes BD. 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**(34):31124 (2002).

[137] Li MX, Spyracopoulos L, & Sykes BD. Binding of cardiac troponin-I147-163 induces a structural opening in human cardiac troponin-C. *Biochemistry*, **38**(26):8289 (1999).

[138] Ovaska M & Taskinen J. A model for human cardiac troponin C and for modulation of its Ca2+ affinity by drugs. *Proteins*, **11**(2):79 (1991).

[139] Robertson IM, Baryshnikova OK, Li MX, & Sykes BD. Defining the binding site of levosimendan and its analogues in a regulatory cardiac troponin C-troponin I complex. *Biochemistry*, **47**(28):7485 (2008).

[140] Sorsa T, Pollesello P, & Solaro RJ. The contractile apparatus as a target for drugs against heart failure: interaction of levosimendan, a calcium sensitiser, with cardiac troponin c. *Mol Cell Biochem*, **266**(1-2):87 (2004).

[141] Gilson M, Given J, Bush B, & McCammon J. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J*, **72**(3):1047 (1997).

[142] Blundell TL & Johnson LN. *Protein crystallography*. Academic Press, New York (1976).

[143] Bonvin AM & Brunger AT. Conformational variability of solution nuclear magnetic resonance structures. *J Mol Biol*, **250**(1):80 (1995).

[144] Saccenti E & Rosato A. The war of tools: how can NMR spectroscopists detect errors in their structures? *J Biomol NMR*, **40**(4):251 (2008).

[145] Morales-Rìos MS, Garcìa-Martínez C, Joseph-Nathan P, & Zepeda LG. 35Cl/37Cl one-bond isotope effects on 13C chemical shifts in a series of para-substituted chlorobenzenes. *Magnetic Resonance in Chemistry*, **33**(2):149 (1995).

[146] Silverstein RM, Bassler GC, & Morrill TC. *Spectrometric identification of organic compounds*. Wiley, New York, 5th edn. (1991).

[147] Lee W, Revington MJ, Arrowsmith C, & Kay LE. A pulsed field gradient isotope-filtered 3D 13C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS Lett*, **350**(1):87 (1994).

[148] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, & Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, **6**(3):277 (1995).

[149] Kleywegt G & Jones T. Databases in protein crystallography. *Acta Cryst*, **D54**:1119 (1998).

[150] Kleywegt G. Dictionaries for Heteros. *CCP4/ESF-EACBM Newsletter on Protein Crystallography*, **31**:45 (1995).

[151] Hempel A, Camerman N, Mastropaolo D, & Camerman A. Calmodulin antagonists: Structure of N-(6-aminohexyl)-5-chloro-1-naphthalenesulfonamide hydrochloride (W-7) and comparison with trifluoperazine (TFP) - Calmodulin binding. *Canadian Journal of Chemistry-Revue Canadienne De Chimie*, **83**:1141 (2005).

[152] Kuszewski J, Gronenborn AM, & Clore GM. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci*, **5**(6):1067 (1996).

[153] Spyracopoulos L, Li MX, Sia SK, Gagne SM, Chandra M, Solaro RJ, & Sykes BD. Calcium-induced structural transition in the regulatory domain of human cardiac troponin C. *Biochemistry*, **36**(40):12138 (1997).

[154] Schwieters CD, Kuszewski JJ, Tjandra N, & Clore GM. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson*, **160**(1):65 (2003).

[155] Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, & Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*, **8**(4):477 (1996).

[156] Wang X, Li MX, Spyracopoulos L, Beier N, Chandra M, Solaro RJ, & Sykes BD. Structure of the C-domain of human cardiac troponin C in complex with the Ca2+ sensitizing drug EMD 57033. *J Biol Chem*, **276**(27):25456 (2001).

[157] Kleerekoper Q, Liu W, Choi D, & Putkey JA. Identification of binding sites for bepridil and trifluoperazine on cardiac troponin C. *J Biol Chem*, **273**(14):8153 (1998).

[158] Li Y, Love ML, Putkey JA, & Cohen C. Bepridil opens the regulatory N-terminal lobe of cardiac troponin C. *Proc Natl Acad Sci U S A*, **97**(10):5140 (2000).

[159] Iorga B, Blaudeck N, Solzin J, Neulen A, Stehle I, Lopez Davila AJ, Pfitzer G, & Stehle R. Lys184 deletion in troponin I impairs relaxation kinetics and induces hypercontractility in murine cardiac myofibrils. *Cardiovasc Res*, **77**(4):676 (2008).

[160] Galinska-Rakoczy A, Engel P, Xu C, Jung H, Craig R, Tobacman LS, & Lehman W. Structural basis for the regulation of muscle contraction by troponin and tropomyosin. *J Mol Biol*, **379**(5):929 (2008).

[161] Slupsky CM & Sykes BD. NMR solution structure of calcium-saturated skeletal muscle troponin C. *Biochemistry*, **34**(49):15953 (1995).

[162] Narita A, Yasunaga T, Yasunaga T, Ishikawa T, Mayanagi K, & Wakabayashi T. Ca2+-induced switching of troponin and tropomyosin on actin filaments as revealed by electron cryo-microscopy. *Journal of Molecular Biology*, **308**(2):241 (2001).

[163] Chandra M, Dong WJ, Pan BS, Cheung HC, & Solaro RJ. Effects of protein kinase A phosphorylation on signaling between cardiac troponin I and the N-terminal domain of cardiac troponin C. *Biochemistry*, **36**(43):13305 (1997).

[164] Li MX, Gagne SM, Tsuda S, Kay CM, Smillie LB, & Sykes BD. Calcium binding to the regulatory N-domain of skeletal muscle troponin C occurs in a stepwise manner. *Biochemistry*, **34**(26):8330 (1995).

[165] Hart RC, Bates MD, Cormier MJ, Rosen GM, & Conn PM. *Synthesis and characterization of calmodulin antagonistic drugs*, vol. 102, pp. 195–204 (1983).

[166] Johnson BA. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol*, **278**:313 (2004).

[167] Boyko R & Sykes BD. http://www.bionmr.ualberta.ca/bds/software/xcrvfit/.

Appendix A

Integrative model of flycasting mechanism

The following is reproduced from [9]. It describes the building of the model depicted in Figure 3.6. The coordinates are available online as supplementary material.

A.1 Manipulation of molecular coordinates

Recent published cryoelectron micrographic reconstructions of the skeletal muscle thin filament, in the inhibited, low-Ca^{2+} and activated, high-Ca^{2+} states, are accompanied by PDB files of the constituent molecules as downloadable supplementary material [41]. That model contains reconstructions for actin, Tm, and Tn. The extent of Tn modeled by the PDB files is limited to the core complexes recently solved by crystallography [1], although additional density, attributed to CTnI, was observed in the inhibited state [41]. Here we model this additional density using the recently published NMR structures of CTnI(PDB 1VDI [2]), and account for the absence of this density in the activated state using the corresponding structures solved in the presence of 0.1 mM $CaCl_2$ (PDB 1VDJ). Our goal is not to describe in detail the Md-actin interaction, but to provide a complete visualization of the mechanism. The interdomain dynamics of TnC (in the inhibited state) were qualitatively visualized using the solution structure of Ca^{2+}-activated TnC (PDB 1TNW [161]. Many of the conformers are not depicted to simplify the picture. Molecular visualization, superimposition and docking was performed with PyMOL (http://pymol.sourceforge.net/).

A.2 Depicting the inhibited state (Figure 2C)

The previous model [41] presents the core domain of Tn docked onto the thin filament. In this state Md's structure is converged and actin-bound. To visualize this, the 20 conformers of the NMR structure of Md (high Ca^{2+} structure, PDB IVDI) were superimposed along the central globular subdomain ($Md_{141-170}$) and docked into the the cryoelectron model. This was accomplished by docking a single conformer, than superimposing the remaining structures onto it. The N-terminal helix of Md ($\alpha 4$) is proposed to have intermediate structuring, and was oriented away from the hypothesized binding site on actin. The actual binding site was selected visually, using the extra density [41] as a reference. This docking is severely underdetermined and should not be used to divulge high precision structure-function relationships. As previously shownd [2], superimposing Md in this fashion allows the unconverged $\alpha 4$ to trace out a large radius, which corresponds to the C-terminus of Sp.

Our hypothesis is that while Md is actin-bound in this state (corroborated by density observed in [41, 162]) Sp is mostly unstructured. Immediately C-terminal to Ip, we propose Sp to undergo structural fluctuations into the space between NTnC and actin, constrained by Ip•actin and Md•actin. Accordingly, we do not attempt to visualize any aspect of Sp except its mean vector between the two actin-bound regions of TnI, which is represented by the blue dash marks connecting Ip to Md (Figure 2, bottom). These vectors range from 19-38 Å in length and convey no structural information, making them valuable only as qualitative aids towards visualization of Sp dynamics in the inhibited state.

By virtue of the DE linker's weak structure in this state, NTnC has relatively unconstrained rotational diffusion. To visualize this, the NMR structure of chicken skeletal TnC in the Ca^{2+}-activated state [161] was superimposed onto the CTnC component of the inactivated Tn complex. Specifically, a single conformer from the NMR ensemble was superimposed, by the α-carbons, onto chain 'C' of 1YTZ ($\sim$1.4 Å RMSD). The remainder of the NMR ensemble was superimposed along the C-domain. About half of the conformers were removed from the model to simplify the picture and to constrain the amplitude of NTnC's positional fluctuations. The use of the Ca^{2+}-activated state does not detract from the model as the conformational state of NTnC is not (adequately) visualized; the use of the NMR structure allows for qualitative visualization of TnC's interdomain dynamics in this state.

A.3 Depicting the activated state (Figure 2A, top)

Visualization of the Ca^{2+}-activated state required less modelling. A portion of Md is visualized in the crystal structure of activated Tn ($TnI_{131-143}$) and not depicted. The NMR ensemble of Md was superimposed upon the initial residues of α4 ($TnI_{131-137}$). The two structural models were not covalently joined; the model is discontinuous at the Sp-α4 junction. We do not intend Figure 2 to depict the relative orientation of Md to Tn in either inactive or active states, since there is currently not enough information available for that.

Appendix B

Methodology: W7 binding studies

B.1 Protein constructs

The cloning, mutagenesis and expression of human cNTnC (C35S, C84S; residues 1-89) and cCTnC (residues 91-161) has been reported in reference [163].

Two synthetic cTnI peptides, cTnI34-71, acetyl-AKKKSKISASRKLQLKTLLLQIAKQELERE-AEERGEK-amide and cTnI128-163, acetyl-TQKIFDLRGKFKRPTLRRVRISADAMMQALLGA-RAK-amide, were prepared with standard methodologies as described by Tripet and others [29].

Protein expression

^{15}N- and $\{^{15}N,^{13}C\}$-labeled protein was expressed and purified as previously described [156, 164, preps for cNTnC and cCTnC, respectively].

W7

W7 (N-[6-aminohexenyl]-5-chloronapthalenesulfonamide) was purchased from Sigma-Aldrich as a dihydrochloride salt. Stock solutions of W7 in DMSO-d6 (Cambrige Isotope Laboratories) were prepared for titrations. Stocks were stored at -20°C, and were protected from exposure to light. W7 concentrations were determined gravimetrically and corrected with UV absorbance measurements ($\epsilon_{1cm,297nm}$ = 8060 from [165]). Gilson Pipetman P2 and P10 were used to deliver the stock solutions.

NMR samples

NMR samples were prepared through reconstitution of lyophilized protein into 100 mM KCl, 10 mM imidazole, 10% D_2O. Protease inhibitor cocktail I (Calbiochem, lot B48045), DSS (0.17 mM), and NaN_3 (0.01%) were added; $CaCl_2$ was added to 10 mM. Samples were brought to pH 6.7 (uncorrected value) with NaOH (~30 μL of 1M NaOH). Samples were filtered through a Millipore Spin-X column (0.22 μL pore size) and decanted into an 5 mm NMR tube. Sample volumes were 500μL, containing 0.5–2 mM protein. Protein concentrations were determined by amino acid analysis. Titrations were initially performed with ^{15}N-labeled samples (one replica each for cNTnC and cCTnC) and were repeated with $\{^{15}N,^{13}C\}$-labeled samples.

NMR spectroscopy

NMR data were acquired at 30°C on a Varian Inova 500 MHz spectrometer possessing a triple resonance probe and Z-pulsed field gradient. The sample was not spun. The experiments were implemented in Biopack (Varian Inc.). All spectra were processed with NMRPipe/NMRDraw [148]. Processing featured forward-back linear prediction in the indirect dimension, and mild apodization and zero filling prior to Fourier transformation (in both dimensions). Peak picking and analysis was performed with the NANUC distribution of NMRView [166] as maintained by Pascal Mercier (Chenomx Inc.).

B.1.1 Titration of cNTnC•Ca2+ or cCTnC•2Ca2+ with W7

For both cNTnC and cCTnC, $\{^1H,^{15}N\}$-HSQC and $\{^1H,^{13}C\}$-HSQC spectra were acquired at incremental concentrations of W7. The titration of ^{15}N-labeled cNTnC had an initial protein concentration of 0.50 mM; $[W7]_{total}$ was 0, 0.069,0.139,0.208, 0.277, 0.414, 0.551, 0.688, 0.959, 1.495 and 2.542 mM (11 data points). This titration is the source of the 2D-$\{^1H,^{15}N\}$-HSQC data presented here. A second sample of cNTnC, with $\{^{15}N,^{13}C\}$-labeling, was also titrated; the initial protein concentration was 1.25 mM, and $[W7]_{total}$ was varied between 0, 0.120, 0.361, 0.833, 1.426, 2.361, 3.509 and 4.630 mM (8 data points). This sample is the source of the 2D-$\{^1H,^{13}C\}$-HSQC spectra presented here. The titration of $^{15}N,^{13}C$-labeled cCTnC had an initial protein concentration of 1.76 mM, and $[W7]_{total}$ was varied between 0, 0.107, 0.322, 0.960, 1.80, 2.62, 3.84 and 5.02 mM (8 points). The $\{^1H,^{15}N\}$- and $\{^1H,^{13}C\}$-HSQC data presented here are from the same trial.

B.1.2 Addition of cNTnC•Ca2+ to W7

A solution of 1.5 mM W7 was prepared in NMR buffer (above) and transferred to an NMR tube. The pH was adjusted to ~6.8 so the volume was 509 μL. A stock of 2.0 mM cNTnC was prepared as described above was pH adjusted to ~6.8 (pH paper was used, my notes have the pH range as 6.5–7). 10μL increments of protein were added at each titration point. At each point in the titration, small aliquots of acid or base were added to maintain a constant pH. The pH seemed to stabilize after point 7 of the titration.

B.1.3 Software

Extraction of the raw data from the assigned HSQC peak lists was done with Perl scripts written by myself. The software xcrvfit was useful throughout my PhD for the interactive perusal of the data [167]. Some simulations and fitting routines were implemented in the free, open source software package R. The development and testing was performed using R version 2.5.1 (2007-06-27) (Gentoo/PPC) and R 2.6.0 GUI 1.21 (4833) (Mac OSX). Older work was done with Mathematica 5 (Wolfram Research).

Appendix C

R scripts

C.1 Least-squares fits to the single site binding model

This script, as written, generates Figure 4.7. It shows a number of aspects of linear regression, and data visualization, using R's internal functions. The single site fits preceeding Figure 4.7 were made with smaller codeblocks from this script.

```
#-------------------------------------------------------------------------------#
#simultaneous fit

inData <- read.table("121-128-129-combined-normalized.xcAD")
inData.ligandConc <- inData[[1]]
inData.protConc <- inData[[3]]
inData.ratio <- list( inData[[1]] / inData[[3]])[[1]]
inData.shifts.raw <- inData[[2]]
numPoints <- length(inData.shifts.raw)
inData.shifts.raw.lastPoint <- inData.shifts.raw[[numPoints]]

testInp <- list(l0=inData.ligandConc,p0=inData.protConc,delDel=inData.shifts.raw)
testFit <- nls( delDel ~ ( maxShift / (2*p0) ) * ( (p0+l0+Kd)-( (p0+l0+Kd)^2 - 4*p0*l0 )^0.5)
, testInp, start=list(Kd=0.0001,maxShift=1.05))
#the above gives a nonlinear least squares fit


#below: making obs vs predicted plot; predicted function is discretely sampled
#sampling <- 200 #number of points to sample curve at
#extrapolation <- 40 #Percentage of input range to plot over -- 1.1 means 'extrapolate 10%'
tempLig <- seq(inData.ligandConc[1],extrapolation*inData.ligandConc[length(inData.ligandConc)],
                                                                    length=sampling)
tempProt <- seq(inData.protConc[1],inData.protConc[length(inData.protConc)], length=sampling)
plot(tempLig/tempProt, predict(testFit,list(l0=tempLig,p0=tempProt)),type = "l",xlab="W7:cCTnC",
                                          ylab="Chemical shift change (ppm)",main="Overlay")
lines(tempLig/tempProt, predict(testFit,list(l0=tempLig,p0=tempProt)),col="blue")
points(inData.ratio,inData.shifts.raw,pch=19,col="blue")
summary(testFit)
testFit

inData <- read.table("128-129-combined-normalized.xcAD")
inData.ligandConc <- inData[[1]]
inData.protConc <- inData[[3]]
inData.ratio <- list( inData[[1]] / inData[[3]])[[1]]
inData.shifts.raw <- inData[[2]]
numPoints <- length(inData.shifts.raw)
inData.shifts.raw.lastPoint <- inData.shifts.raw[[numPoints]]

testInp <- list(l0=inData.ligandConc,p0=inData.protConc,delDel=inData.shifts.raw)
```

```
testFit <- nls( delDel ~ ( maxShift / (2*p0) ) * ( (p0+l0+Kd)-( (p0+l0+Kd)^2 - 4*p0*l0 )^0.5)
, testInp, start=list(Kd=0.0001,maxShift=1.05))
#the above gives a nonlinear least squares fit


#below: making obs vs predicted plot; predicted function is discretely sampled
#sampling <- 200 #number of points to sample curve at
#extrapolation <- 40 #Percentage of input range to plot over -- 1.1 means 'extrapolate 10%'
tempLig <- seq(inData.ligandConc[1],extrapolation*inData.ligandConc[length(inData.ligandConc)],
                                                                        length=sampling)
tempProt <- seq(inData.protConc[1],inData.protConc[length(inData.protConc)], length=sampling)
#plot(tempLig/tempProt, predict(testFit,list(l0=tempLig,p0=tempProt)),type = "l",xlab="W7:cCTnC",
                                                ylab="Chemical shift change (ppm)",main="Overlay")
lines(tempLig/tempProt, predict(testFit,list(l0=tempLig,p0=tempProt)),col="green")
#points(inData.ratio,inData.shifts.raw,pch=19,col="green")
summary(testFit)
testFit

#normalizing fits for 128, 129 and comparing them
#then normalizing and fitting simultaneously

#setwd("/Users/rmbh/work/THESIS/projects/crossVal-final/crossVal-preThesis/Rscripts")
setwd("/Users/rmbh/work/THESIS/projects/crossVal-final/")

require(stats); require(graphics)

#inputing data
inData <- read.table("129.xcAD")
inData.ligandConc <- inData[[1]]
inData.protConc <- inData[[3]]
inData.ratio <- list( inData[[1]] / inData[[3]])[[1]]
inData.shifts.raw <- inData[[2]]
numPoints <- length(inData.shifts.raw)
inData.shifts.raw.lastPoint <- inData.shifts.raw[[numPoints]]
inData.shifts.norm <- inData.shifts.raw / inData.shifts.raw.lastPoint
#plot(inData.ratio,inData.shifts.norm)

testInp <- list(l0=inData.ligandConc,p0=inData.protConc,delDel=inData.shifts.norm)
testFit <- nls( delDel ~ ( maxShift / (2*p0) ) * ( (p0+l0+Kd)-( (p0+l0+Kd)^2 - 4*p0*l0 )^0.5)
, testInp
, start=list( Kd=0.0001, maxShift=1.05 ))
#the above gives a nonlinear least squares fit


#below: making obs vs predicted plot; predicted function is discretely sampled
sampling <- 400 #number of points to sample curve at
extrapolation <- 40 #Percentage of input range to plot over -- 1.1 means 'extrapolate 10%'
tempLig <- seq(inData.ligandConc[1],extrapolation*inData.ligandConc[length(inData.ligandConc)],
                                                                        length=sampling)
tempProt <- seq(inData.protConc[1],inData.protConc[length(inData.protConc)], length=sampling)

#plot(tempLig/tempProt, predict(testFit,list(l0=tempLig,p0=tempProt)),type = "l",xlab="W7:cCTnC",
ylab="Chemical shift change (ppm)",main="")
lines(tempLig/tempProt, predict(testFit,list(l0=tempLig,p0=tempProt)))
points(inData.ratio,inData.shifts.norm)
summary(testFit)
testFit


#inputing data
inData <- read.table("128.xcAD")
inData.ligandConc <- inData[[1]]
inData.protConc <- inData[[3]]
inData.ratio <- list( inData[[1]] / inData[[3]])[[1]]
```

```
inData.shifts.raw <- inData[[2]]
numPoints <- length(inData.shifts.raw)
inData.shifts.raw.lastPoint <- inData.shifts.raw[[numPoints]]
inData.shifts.norm <- inData.shifts.raw / inData.shifts.raw.lastPoint
#plot(inData.ratio,inData.shifts.norm)

testInp <- list(l0=inData.ligandConc,p0=inData.protConc,delDel=inData.shifts.norm)
testFit <- nls( delDel ~ ( maxShift / (2*p0) ) * ( (p0+l0+Kd)-( (p0+l0+Kd)^2 - 4*p0*l0 )^0.5)
, testInp, start=list(Kd=0.0001,maxShift=1.05))
#the above gives a nonlinear least squares fit


#below: making obs vs predicted plot; predicted function is discretely sampled
#sampling <- 200 #number of points to sample curve at
#extrapolation <- 40 #Percentage of input range to plot over -- 1.1 means 'extrapolate 10%'
tempLig <- seq(inData.ligandConc[1],extrapolation*inData.ligandConc[length(inData.ligandConc)],
                                                                    length=sampling)
tempProt <- seq(inData.protConc[1],inData.protConc[length(inData.protConc)], length=sampling)

#plot(tempLig/tempProt, predict(testFit,list(l0=tempLig,p0=tempProt)),type = "l",xlab="W7:cCTnC",
                                              ylab="Chemical shift change (ppm)",main="Overlay")
lines(tempLig/tempProt, predict(testFit,list(l0=tempLig,p0=tempProt)))
points(inData.ratio,inData.shifts.norm,pch=19)
summary(testFit)
testFit
```

C.2 Global least-squares fit

This repeats the global single-site analysis in [12], but implemented in R. The original Mathematica scripts are available online as Supplementary Material [12].

```
#v. 0.1  (Oct. 23, 2007)
#v. 0.2  (Nov. 4, 2007)
#rmbh

#RUN: plotGlobalNorm() after sourcing this in R

setwd("~/work/THESIS/projects/crossVal-final/crossVal-preThesis/testSets/NTnCW721/")
datasets <- Sys.glob("*.dat")

plotGlobalNorm <- function(allData){
plotSingleNorm(allData)
plotAllNorm(allData)
}

##routine to input a set of coupled titration curves
##data should be represented with labels
##data is stored without normalization
allData <- data.frame() #using data frames allows mixing of characters (say set names)
allSetNames <- list()
iter = 0
for (dataset in datasets) {
iter = iter + 1
setName <- substring(dataset, 1, {nchar(dataset)-4} ) #drop 4-character suffix from filename
allSetNames[{length(allSetNames)+1}] <- c(setName)
inData <- read.table(dataset, header=FALSE,
col.names=list("concTitrant","deltaDelta","concTarget"))
inData <- data.frame(setName,inData)
if( iter==1 ){ allData <- inData }
else{ allData <- rbind(allData,inData)}
}
```

```
##routine to normalize and to simultaneously plot all raw data
plotAllNorm <- function(allData){
iter <- 0
allData.deltaDelta.norm <- data.frame()
for (name in allSetNames){
iter <- iter + 1
set <- subset(allData,allData$setName==name)
deltaDelta.norm <- set$deltaDelta / set$deltaDelta[length(set$deltaDelta)]
concRatio <- set$concTitrant / set$concTarget
plot(concRatio,deltaDelta.norm,type="b",xlab="",ylab="",pch=20)
par(new=TRUE)
if( iter==1 ){ allData.deltaDelta.norm <- deltaDelta.norm }
else{ allData.deltaDelta.norm <- cbind(allData.deltaDelta.norm,deltaDelta.norm)}
}
title(main="All normalized signals", xlab="[titrant]:[target]",
ylab="normalized change",  new=FALSE)
}


##routine to normalize and to individually plot all raw data
plotSingleNorm <- function(allData){
layout(matrix(1:25,nrow=5,ncol=5)) #make window large enough for 25 signals
par(mar=c(2,2,1,1)+0.5)
iter = 0
allData.deltaDelta.norm <- data.frame()
for (name in allSetNames){
iter = iter + 1
set <- subset(allData,allData$setName==name)
deltaDelta.norm <- set$deltaDelta / set$deltaDelta[length(set$deltaDelta)]
concRatio <- set$concTitrant / set$concTarget
   titleString <- c("residue",name)
plot(concRatio,deltaDelta.norm,type="b",main=titleString,xlab="[titrant]:[target]",
                                                                    ylab="normalized change")
par(new=FALSE,ask=TRUE)
if( iter==1 ){ allData.deltaDelta.norm <- deltaDelta.norm }
else{ allData.deltaDelta.norm <- cbind(allData.deltaDelta.norm,deltaDelta.norm)}
}
}


doNLSss <- function( titrationData , signalName){
inData <- subset( titrationData,allData$setName==signalName)
inData.shifts.norm <- inData$deltaDelta / inData$deltaDelta[length(inData$deltaDelta)]
inData2 <- list(l0=inData$concTitrant, p0=inData$concTarget, delDel=inData.shifts.norm)
nlsFit <- nls(
delDel ~ ( maxShift / (2*p0) ) * ( (p0+l0+Kd)-( (p0+l0+Kd)^2 - 4*p0*l0 )^0.5)
, inData2, start=list(Kd=0.0001,maxShift=1.05))
nlsFit
}
#doNLSss(allData, 17)

##routine to return the residuals for a fit
getNLSss <- function( titrationData , signalName){
inData <- subset( titrationData,allData$setName==signalName)
inData.shifts.norm <- inData$deltaDelta / inData$deltaDelta[length(inData$deltaDelta)]
inData2 <- list(l0=inData$concTitrant, p0=inData$concTarget, delDel=inData.shifts.norm)
nlsFit <- nls(
delDel ~ ( maxShift / (2*p0) ) * ( (p0+l0+Kd)-( (p0+l0+Kd)^2 - 4*p0*l0 )^0.5)
, inData2, start=list(Kd=0.0001,maxShift=1.05))
# nlsFit
residualSOS <- sum(summary(nlsFit)$residuals^2)
fitKd <-  coef(nlsFit)[1]
fitMaxShift <- coef(nlsFit)[2]
c(fitKd, fitMaxShift, residualSOS=residualSOS)
```

```
}

##routine to evaluate the residuals for a given query Kd, scale
evalSOSss <- function( titrationData , signalName, Kd, maxShift){
inData <- subset( titrationData, titrationData$setName==signalName)
inData.shifts.norm <- inData$deltaDelta / inData$deltaDelta[length(inData$deltaDelta)]
outputSOS <- 0
iter <- 0
for( delDelObs in inData.shifts.norm ){
iter <- iter + 1
l0 <- inData$concTitrant[iter]
p0 <- inData$concTarget[iter]

#INPUT VARIABLES Kd and maxShift USED HERE
delDelPred <- ( maxShift / (2*p0) ) * ( (p0+l0+Kd)-( (p0+l0+Kd)^2 - 4*p0*l0 )^0.5)
outputSOS <- outputSOS + (delDelPred-delDelObs)^2
}
outputSOS
}



##routine to map the SOS-error surface of the global data set
mapGlobalSOS <- function( titrationData ){
outTable <- data.frame()
#first build a list of all of the local fits
localKds <- list()
localMaxShifts <- list()
iter <- 0
for (name in allSetNames){
fit <- getNLSss( titrationData, name )
localKds <- rbind(localKds, fit[1])
localMaxShifts <- rbind(localMaxShifts, fit[2])
}
ssGrid <- cbind(Set=allSetNames,localKds,localMaxShifts)
print("All local fits")
print(ssGrid)

KdGrid <- sort(as.numeric(ssGrid[,2]))
MaxShiftGrid <- sort(as.numeric(ssGrid[,3]))
startMaxShift <- mean((as.numeric(ssGrid[,3])))

print("Now optimizing Kd with fixed value of maxShift...")
for (Kd in KdGrid){
globalSOS <- 0
maxShift <- startMaxShift
for (name in allSetNames){
localSOS <- evalSOSss(titrationData , name, Kd, maxShift)
globalSOS <- globalSOS + localSOS
}
outTable <- rbind( outTable, cbind( Kd, globalSOS ))
}
print("...Completed optimizing Kd")
outTable
}

#globalSOS <- mapGlobalSOS(allData)
#plot(globalSOS)

doWeightedNLSss <- function( titrationData , signalName, weightsVector){
inData <- subset( titrationData,titrationData$setName==signalName)
inData.shifts.norm <- inData$deltaDelta / inData$deltaDelta[length(inData$deltaDelta)]
inData2 <- list(l0=inData$concTitrant, p0=inData$concTarget, delDel=inData.shifts.norm)
nlsFit <- nls(
```

```
delDel ~ ( maxShift / (2*p0) ) * ( (p0+l0+Kd)-( (p0+l0+Kd)^2 - 4*p0*l0 )^0.5)
, inData2, start=list(Kd=0.0001,maxShift=1.05), weights=weightsVector)
nlsFit
}

getWeightedNLSss <- function( titrationData , signalName, weightsVector){
inData <- subset( titrationData,titrationData$setName==signalName)
inData.shifts.norm <- inData$deltaDelta / inData$deltaDelta[length(inData$deltaDelta)]
inData2 <- list(l0=inData$concTitrant, p0=inData$concTarget, delDel=inData.shifts.norm)
nlsFit <- nls(
delDel ~ ( maxShift / (2*p0) ) * ( (p0+l0+Kd)-( (p0+l0+Kd)^2 - 4*p0*l0 )^0.5)
, inData2, start=list(Kd=0.0001,maxShift=1.05), weights=weightsVector)
# nlsFit
residualSOS <- sum(summary(nlsFit)$residuals^2)
fitKd <-  coef(nlsFit)[1]
fitMaxShift <- coef(nlsFit)[2]
c(fitKd, fitMaxShift, residualSOS=residualSOS)
}


generateSimpleTitrErr <- function( numPoints, percentErr ){
out <- list()
iter <- 0
while (iter < numPoints){
out <- cbind( out, (1-iter*(percentErr/100 )))
iter <- iter + 1}
as.vector(out,mode="numeric")
}


##routine to map the SOS-error surface of the global data set WITH WEIGHTS
mapWeightedGlobalSOS <- function( titrationData, weightsVector ){
outTable <- data.frame()
#first build a list of all of the local fits
localKds <- list()
localMaxShifts <- list()
iter <- 0
for (name in allSetNames){
fit <- getWeightedNLSss( titrationData, name, weightsVector )
localKds <- rbind(localKds, fit[1])
localMaxShifts <- rbind(localMaxShifts, fit[2])
}
ssGrid <- cbind(Set=allSetNames,localKds,localMaxShifts)
print("All local fits")
print(ssGrid)

KdGrid <- sort(as.numeric(ssGrid[,2]))
MaxShiftGrid <- sort(as.numeric(ssGrid[,3]))
startMaxShift <- mean((as.numeric(ssGrid[,3])))

print("Now optimizing Kd with fixed value of maxShift...")
for (Kd in KdGrid){
globalSOS <- 0
maxShift <- startMaxShift
for (name in allSetNames){
localSOS <- evalSOSss(titrationData , name, Kd, maxShift)
globalSOS <- globalSOS + localSOS
}
outTable <- rbind( outTable, cbind( Kd, globalSOS ))
}
print("...Completed optimizing Kd")
outTable
}
```

```
#using the following weights:
# weights3 <- c( 0.00, 0.02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20 )
# interesting results compared with
# weights2 <- c( 1.00 1.00 1.00 1.00, 0.10, 0.10, 0.10, 0.10, 0.10, 0.05, 0.05 )
# minimum SOS is for Kd = 0.00017056 using weights3; used below


##THIS IS A HACK to CONVERGE AROUND THE FITTED KD with a FIXED maxShift
mapWeightedGlobalSOS2 <- function( titrationData, weightsVector ){
outTable <- data.frame()
outTable2 <- data.frame()
#first build a list of all of the local fits
localKds <- list()
localMaxShifts <- list()
iter <- 0
for (name in allSetNames){
fit <- getWeightedNLSss( titrationData, name, weightsVector )
localKds <- rbind(localKds, fit[1])
localMaxShifts <- rbind(localMaxShifts, fit[2])
}
ssGrid <- cbind(Set=allSetNames,localKds,localMaxShifts)
print("All local fits")
print(ssGrid)

KdGrid <- sort(as.numeric(ssGrid[,2]))
MaxShiftGrid <- sort(as.numeric(ssGrid[,3]))
startMaxShift <- mean((as.numeric(ssGrid[,3])))

print("Now optimizing Kd with fixed value of maxShift...")
for (Kd in KdGrid){
globalSOS <- 0
maxShift <- startMaxShift
for (name in allSetNames){
localSOS <- evalSOSss(titrationData , name, Kd, maxShift)
globalSOS <- globalSOS + localSOS
}
outTable <- rbind( outTable, cbind( Kd, globalSOS ))
}
print("...Completed optimizing Kd")

print("VERSION2 (hackery): now converging on optimized Kd...")
for (maxShift in MaxShiftGrid){
globalSOS <- 0
Kd <- 1.705647e-04
for (name in allSetNames){
localSOS <- evalSOSss(titrationData , name, Kd, maxShift)
globalSOS <- globalSOS + localSOS
}
outTable2 <- rbind( outTable2, cbind( maxShift, globalSOS ))
}
print("...END: optimized global value of maxShift for Kd=1.705647e-04")
outTable2
}
```

C.3 Global, single-site model with weighted least-squares

This is the code to produce Figure 4.14 (p. 66).

```
source("ssfit-global-ls.r")

weights1percent <- generateSimpleTitrErr(11,1)
weights2percent <- generateSimpleTitrErr(11,2)
```

```
weights3percent <- generateSimpleTitrErr(11,3)
weights4percent <- generateSimpleTitrErr(11,4)
weights5percent <- generateSimpleTitrErr(11,5)
weights6percent <- generateSimpleTitrErr(11,6)
weights7percent <- generateSimpleTitrErr(11,7)
weights8percent <- generateSimpleTitrErr(11,8)
weights9percent <- generateSimpleTitrErr(11,9)
weights10percent <- generateSimpleTitrErr(11,10)

withweights1percent <- mapWeightedGlobalSOS( allData, weights1percent )
withweights2percent <- mapWeightedGlobalSOS( allData, weights2percent )
withweights3percent <- mapWeightedGlobalSOS( allData, weights3percent )
withweights4percent <- mapWeightedGlobalSOS( allData, weights4percent )
withweights5percent <- mapWeightedGlobalSOS( allData, weights5percent )
withweights6percent <- mapWeightedGlobalSOS( allData, weights6percent )
withweights7percent <- mapWeightedGlobalSOS( allData, weights7percent )
withweights8percent <- mapWeightedGlobalSOS( allData, weights8percent )
withweights9percent <- mapWeightedGlobalSOS( allData, weights9percent )
withweights10percent <-mapWeightedGlobalSOS( allData, weights10percent )



plot(withweights1percent,pch=19)
points(withweights1percent,col="brown")
points(withweights2percent,col="maroon")
points(withweights3percent,col="red")
points(withweights4percent,col="orangered")
points(withweights5percent,col="orange")
points(withweights6percent,col="yellow")
points(withweights7percent,col="yellowgreen")
points(withweights8percent,col="green")
points(withweights9percent,col="blue")
points(withweights10percent,col="purple",pch=19)

#the "ssfit-global-ls.r" script can alter the current working directory
#setwd(''pathToStartingDirectory'')
```

Appendix D

Mathematica Script

D.1 Maximum likelihood analysis of titration of cNTnC•Ca2+

This script is named 'prototype-maxLikelihood.nb' A PDF of the notebook is included following an expansion of the input cell, given below.

```
residues={3,4,5,8,9,11,12,13,14,15,17,18,20,21,23,26,29,30,31,32,34,42,43,47,
    48,49,55,58,63,66,67,68,69,70,71,72,74,76,79,87,88};namelist:={"Met 1",
    "Asp 2","Asp 3","Ile 4","Tyr 5","Lys 6","Ala 7","Ala 8","Val 9","Glu 10",
    "Gln 11","Leu 12","Thr 13","Glu 14","Glu 15","Gln 16","Lys 17","Asn 18",
    "Glu 19","Phe 20","Lys 21","Ala 22","Ala 23","Phe 24","Asp 25","Ile 26",
    "Phe 27","Val 28","Leu 29","Gly 30","Ala 31","Glu 32","Asp 33","Gly 34",
    "Ser 35","Ile 36","Ser 37","Thr 38","Lys 39","Glu 40","Leu 41","Gly 42",
    "Lys 43","Val 44","Met 45","Arg 46","Met 47","Leu 48","Gly 49","Gln 50",
    "Asn 51","Pro 52","Thr 53","Pro 54","Glu 55","Glu 56","Leu 57","Gln 58",
    "Glu 59","Met 60","Ile 61","Asp 62","Glu 63","Val 64","Asp 65","Glu 66",
    "Asp 67","Gly 68","Ser 69","Gly 70","Thr 71","Val 72","Asp 73","Phe 74",
    "Asp 75","Glu 76","Phe 77","Leu 78","Val 79","Met 80","Met 81","Val 82",
    "Arg 83","Ser 84","Met 85","Lys 86","Asp 87","Asp 88","Ser 89"};
(*full pathname must be entered below*)

workingdir (*needs trailing '/'*):=
    "/Users/rmbh/Documents/writing/published/w7_binding_tnc05/original_\
submission/W7_binding_cTnC_supplement/ndom_nh_scale/";
fileglob:=".xcAD";

(*user input stops here*)

(*initialization*)

Off[General::spell1]
(*
  makes master input table from directory full of Xcurvfit input files
  *)
getinput=Module[
      {it,file,concL,concP,vars,obs}
      ,{
         input={};
         Do[
           file=
             ToString[
```

```
                workingdir
                  <>
                  ToString[residues[[i]]]
                  <>
                  fileglob
                ];
            it=Import[file,"Table"];
            concL:=Transpose[it][[1]];
            concP:= Transpose[it][[3]];
            vars=Transpose[{concL,concP}];
            obs = Transpose[it][[2]];
            AppendTo[input,{residues[[i]],vars,obs}];
            ,{i,1,Length[residues]}
          ];
        }];
inputOrig = input;

lastpoints=
    {
        Transpose[input][[1]]
        ,
        Part[
           Transpose[input][[3]]//Transpose
           ,Transpose[input][[3]][[1]]//Length
           ]
        }//Transpose;
```

```
(*applying learnMaxLikelihood.nb to the fitSS notebook*)

(*define ' input' in the closed cell below*)


(*evaluating the next cell defines a normalized data set*)
normalize[data_] := Module[{norm, temp, resi, vars, obs, normObs},
  {
   norm = {};
   Do[
    temp = input[[j]];
    resi = temp[[1]];
    vars = temp[[2]];
    obs = temp[[3]];
    normObs = obs / obs[[-1]];
    AppendTo[norm, {resi, vars, normObs}];
    , {j, Length[data]}
   ];
   norm
  }]
norm = normalize[input];
norm = norm[[1]];
(*undefine input so i don' t accidentally use it*)
input =.

(*make likelihood function' s exponential a difference between pred and obs Δδs*)
(*this time set up correctly - to describe a Gaussian of Kds (not Δδs)*)

lhSSlocal [data_, kd_, scale_, sigma_, residue_] :=
 Module[{vars, obs, δss, evalss, target, b, x, x2, a0, a1}, {
   vars = data[[residue]][[2]]; (*first index here is residue number*)
   obs = data[[residue]][[3]];
   a1 := ((a0 * x) / δss) + ((δss * x2) / a0) - x - x2;
   evalss[{LIG_, PROT_}, SHIFT_] := a1 /. {a0 → scale, x → LIG, x2 → PROT, δss → SHIFT};
   Sum[
    Log[
     (1 / (2 * Pi * sigma^2) ^0.5) *
      Exp[-1 * ((evalss[vars[[j]], obs[[j]]] - kd) ^2) / (2 * sigma^2)]]
    , {j, 2, Length[vars]}]}]

lhSSglobal [data_, kd_, scale_, sigma_] :=
 Module[{vars, obs, δss, evalss, target, b, x, x2, a0, a1}, {
   a1 := ((a0 * x) / δss) + ((δss * x2) / a0) - x - x2;
   evalss[{LIG_, PROT_}, SHIFT_] := a1 /. {a0 → scale, x → LIG, x2 → PROT, δss → SHIFT};
   Sum[
    vars = data[[residue]][[2]];
    obs = data[[residue]][[3]];
    Sum[
     Log[
      (1 / (2 * Pi * sigma^2) ^0.5) *
       Exp[-1 * ((evalss[vars[[j]], obs[[j]]] - kd) ^2) / (2 * sigma^2)]]
     , {j, 2, Length[vars]}]
    , {residue, Length[data]}]}]
```

```
doGlobalSS[data_] := Module[{max}, {
   (*NOTE: I don't understand the order for output terms chosen by NMaximize*)
   max = NMaximize[
     lhSSglobal[data, kd, scale, sigma]
     , {{kd, 0.0001, .00026}, {scale, 1, 1.2}, {sigma, 0.00001, 0.001}}];
   max}]
visualizeGlobal[dat_, {fitKD_, fitERR_, fitSCALE_, maxL_}] :=
 Module[{data, temp, a, b, c, d}, {
   a = fitERR; b = fitKD; c = fitSCALE;
   Do[
    temp =
     Plot[lhSSlocal[dat, kd, c, a, j], {kd, b - 2 * a, b + 2 * a},
      PlotLabel → j, PlotPoints → 10, DisplayFunction → Identity];
    d = maxL / (Length[dat] - 2); (*vertical scale*)
    Show[
     temp
     , Graphics[Line[{{b, 0}, {b, 150}}]]
     , Graphics[Line[{{b + a, 0}, {b + a, d}}]]
     , Graphics[Line[{{b - a, 0}, {b - a, d}}]]
     , DisplayFunction → $DisplayFunction]
    , {j, Length[dat]}]
  }]
getGSSvars[GSSout_] :=
 {GSSout [[1, 2]][[1, 2]], GSSout[[1, 2]][[3, 2]], GSSout[[1, 2]][[2, 2]], GSSout[[1, 1]]}

doGlobalSS[norm]

{{2840.18, {kd → 0.0000136064, scale → 0.759414, sigma → 0.000237294}}}

visualizeGlobal[norm, getGSSvars[%161]]

culled1 = Delete[norm, {{10}, {38}, {40}}];
doGlobalSS[culled1]

{{2790.25, {kd → 0.000146375, scale → 0.975626, sigma → 0.000156618}}}

visualizeGlobal[culled1, getGSSvars[%127]]

culled2 = Delete[culled1, {{4}, {7}, {11}}];
doGlobalSS[culled2]

{{2696.64, {kd → 0.000173391, scale → 1.03787, sigma → 0.000109058}}}

visualizeGlobal[culled2, getGSSvars[%130]]

culled3 = Delete[culled2, {{3}, {4}, {5}, {35}}];
doGlobalSS[culled3]

{{2506.25, {kd → 0.000168152, scale → 1.05314, sigma → 0.0000745819}}}

visualizeGlobal[culled3, getGSSvars[%133]]

culled4 = Delete[culled3, {{1}, {7}, {18}, {23}, {25}, {27}, {31}}];
doGlobalSS[culled4]

{{2063.92, {kd → 0.000151326, scale → 1.059, sigma → 0.0000445629}}}

visualizeGlobal[culled4, getGSSvars[%136]]
```

```
culled5 = Delete[culled4, {{2}, {3}}];
doGlobalSS[culled5]
```

```
{{1911.12, {kd → 0.000146682, scale → 1.05873, sigma → -0.0000408406}}}
```

```
doGlobalSS[culled5]
```

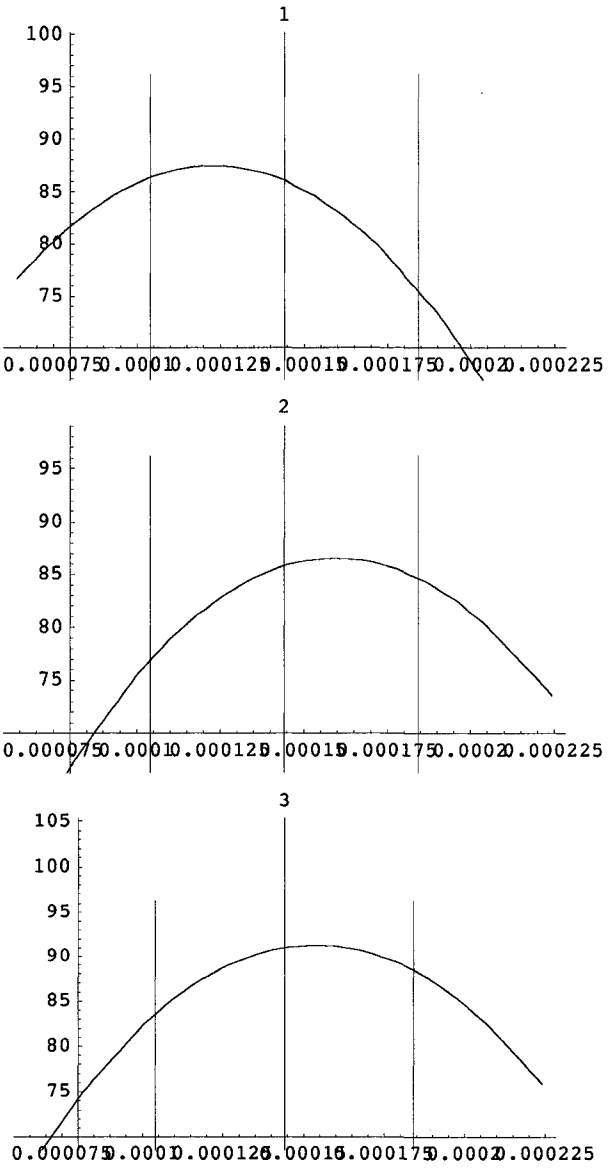
```
{{1911.12, {kd → 0.000146682, scale → 1.05873, sigma → -0.0000408406}}}
```

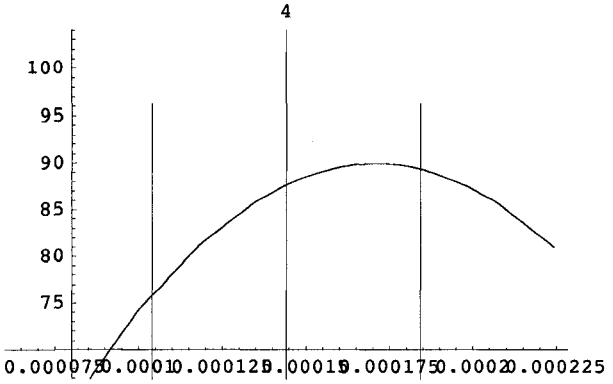
```
visualizeGlobal[culled5, getGSSvars[%140]]
```

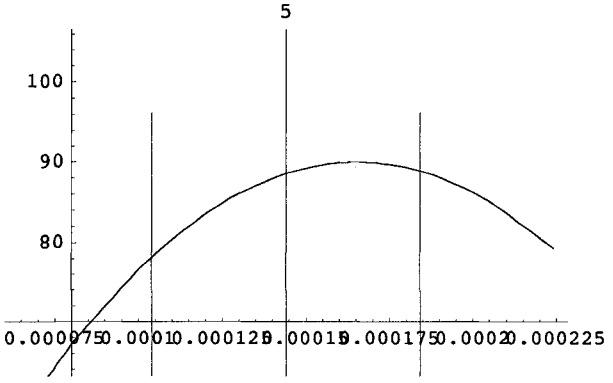
```
culled6 = Delete[culled5, {{7}}];
doGlobalSS[culled6]
```

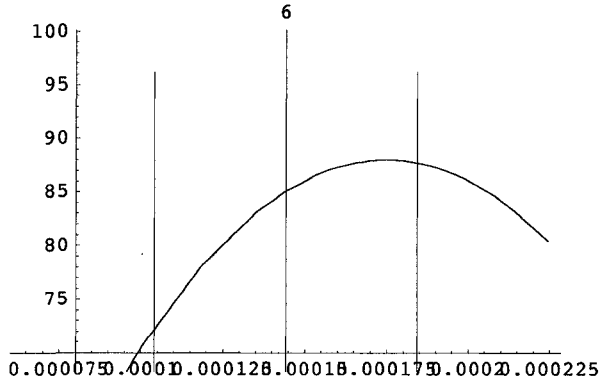
```
{{1828.39, {kd → 0.000144146, scale → 1.05814, sigma → -0.0000400422}}}
```

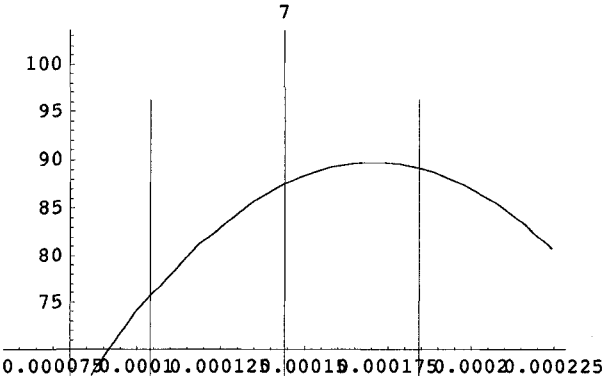
```
visualizeGlobal[culled6, getGSSvars[%146]]
```

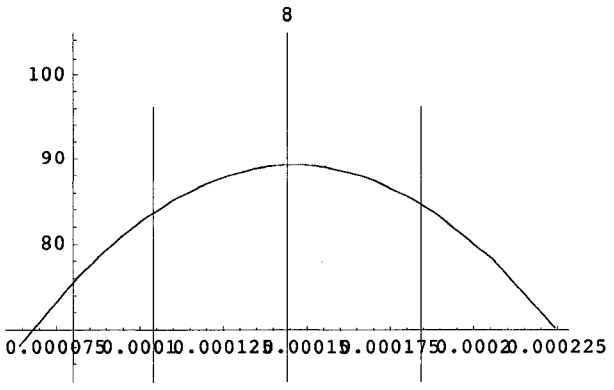


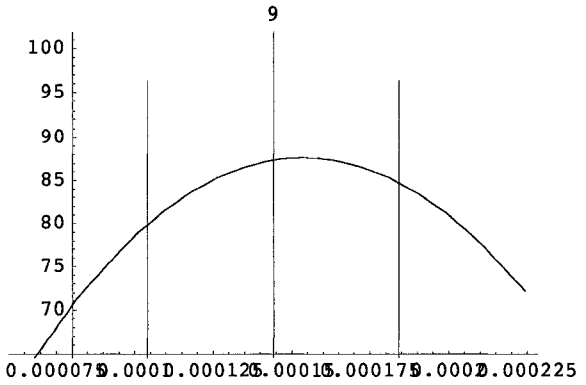

4
100
95
90
85
80
75
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

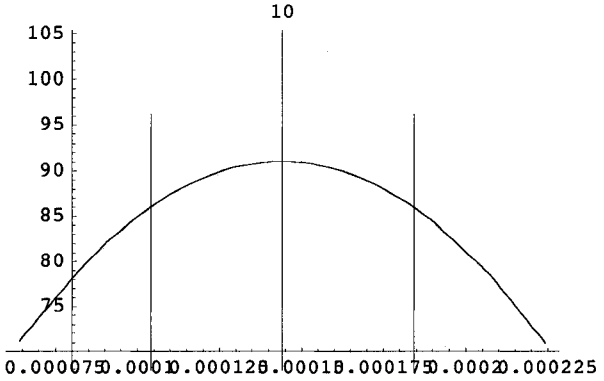

5
100
90
80
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

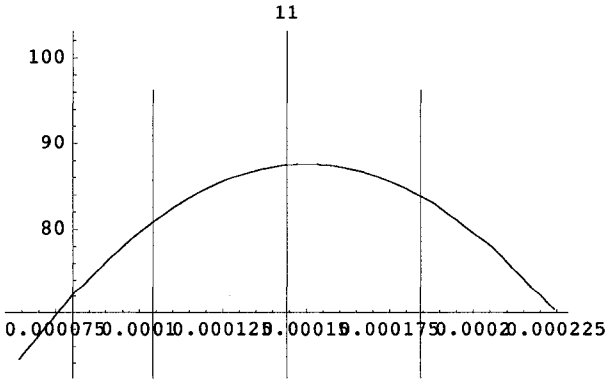

6
100
95
90
85
80
75
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

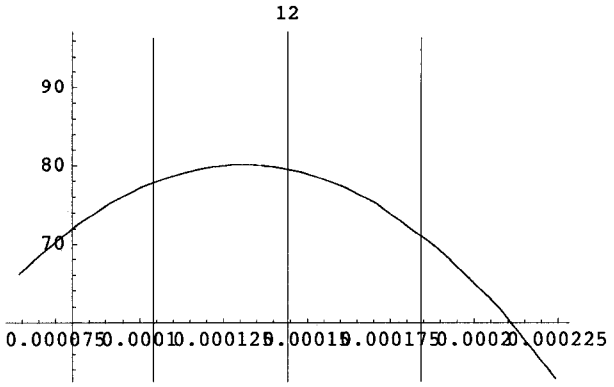

7
100
95
90
85
80
75
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

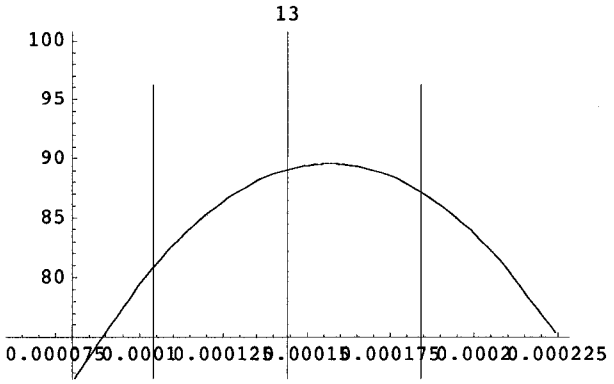

8
100
90
80
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

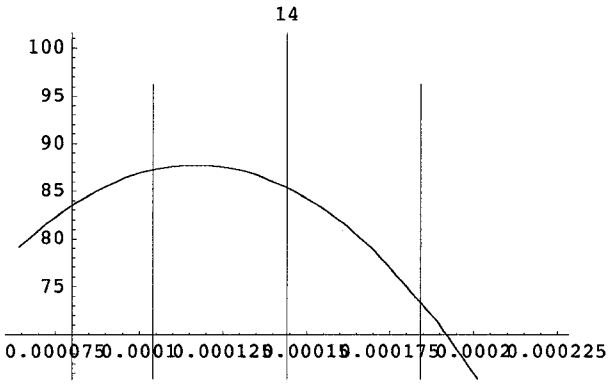

9
100
95
90
85
80
75
70
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

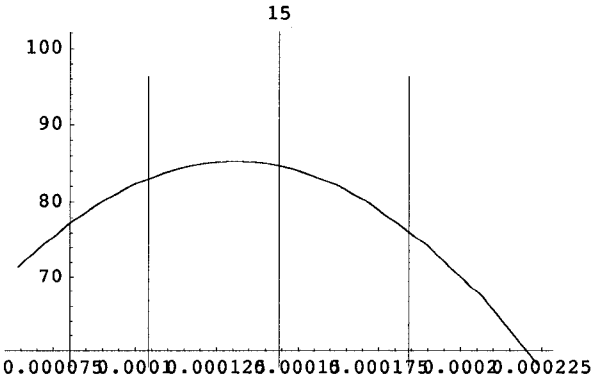

10
105
100
95
90
85
80
75
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

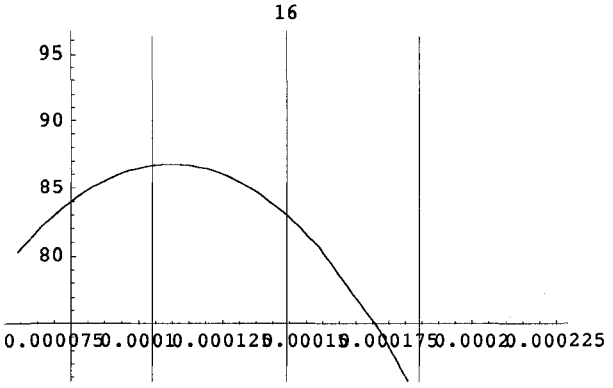

11
100
90
80
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

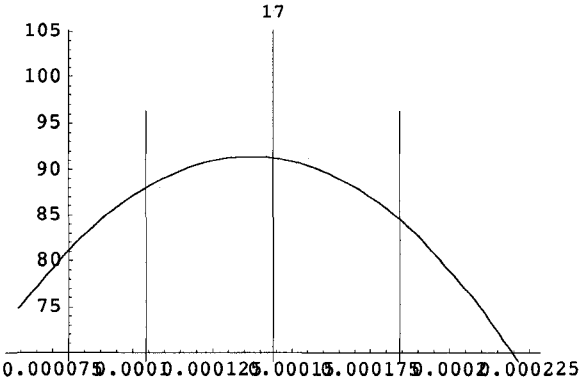

12
90
80
70
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

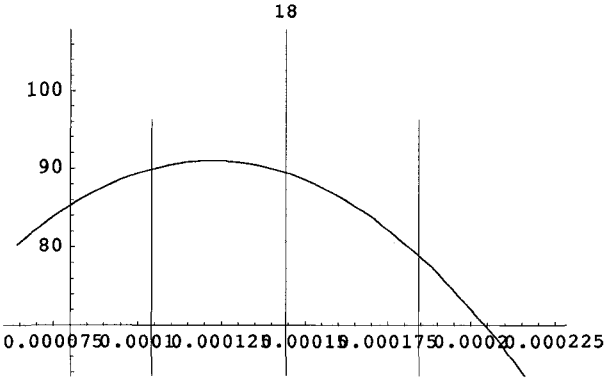

13
100
95
90
85
80
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

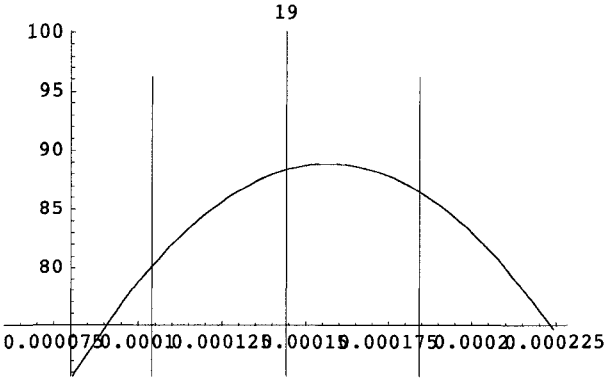

14
100
95
90
85
80
75
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

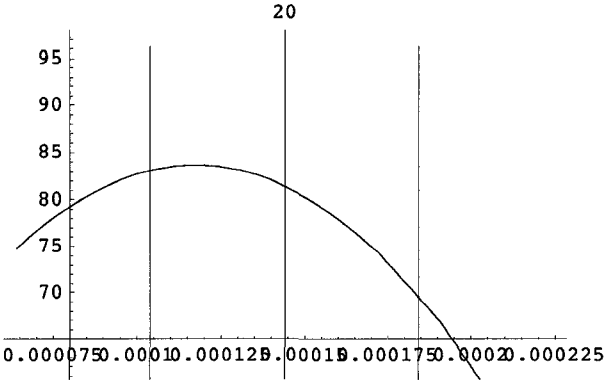

15
100
90
80
70
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225

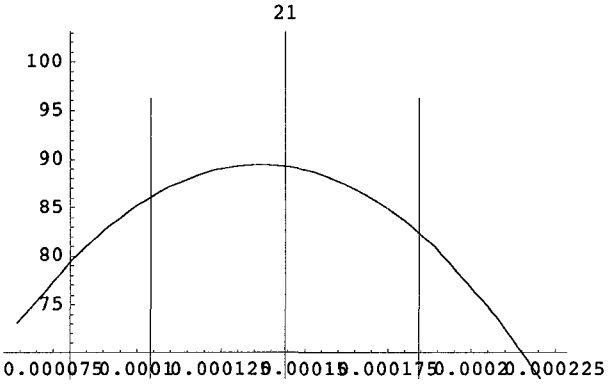

16
95
90
85
80
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225


17
105
100
95
90
85
80
75
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225


18
100
90
80
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225


19
100
95
90
85
80
0.000075
0.0001
0.000125
0.00015
0.000175
0.0002
0.000225





{Null}

Length[culled6]

21

Appendix E

Spreadsheet to Simulate the Association of cNTnC•Ca2+ and W7

		[D]	[N:D]	[N:D:Dn]	[N]	Ntotal	Ntotal/Dtota	[Q]	[D]	[N:D]	[N:D:Dn]	[N]	Dtotal	Dtotal/Ntotal	[Q]
K1 =	4.00E-05	0.0000100	1.49E-03	0.0000000	0.0059600	0.0074500	4.97	0.9933333	0.0000100	1.00E-04	0.0000000	0.0004000	0.0001100	0.22	0.2000000
K2 =	3.00E-13	0.0000250	1.47E-03	0.0000000	0.0023600	0.0038350	2.56	0.9833333	0.0000200	1.67E-04	0.0000000	0.0003333	0.0001867	0.37	0.3333335
		0.0000400	1.46E-03	0.0000000	0.0014599	0.0029199	1.95	0.9733333	0.0000300	2.14E-04	0.0000000	0.0002857	0.0002443	0.49	0.4285721
[D]total =	1.50E-03	0.0000550	1.44E-03	0.0000000	0.0010507	0.0024956	1.66	0.9633333	0.0000400	2.50E-04	0.0000000	0.0002500	0.0002900	0.58	0.5000021
		0.0000700	1.43E-03	0.0000001	0.0008168	0.0022464	1.50	0.9533333	0.0000500	2.78E-04	0.0000000	0.0002222	0.0003278	0.66	0.5555607
n	4.000	0.0000850	1.41E-03	0.0000002	0.0006653	0.0020793	1.39	0.9433333	0.0000600	3.00E-04	0.0000000	0.0002000	0.0003601	0.72	0.6000104
		0.0001000	1.40E-03	0.0000005	0.0005591	0.0019572	1.30	0.9333333	0.0000700	3.18E-04	0.0000000	0.0001818	0.0003883	0.78	0.6363822
		0.0001150	1.38E-03	0.0000008	0.0004803	0.0018621	1.24	0.9233333	0.0000800	3.33E-04	0.0000000	0.0001667	0.0004135	0.83	0.6666970
		0.0001300	1.36E-03	0.0000013	0.0004195	0.0017843	1.19	0.9133333	0.0000900	3.46E-04	0.0000001	0.0001538	0.0004365	0.87	0.6923543
		0.0001450	1.35E-03	0.0000020	0.0003711	0.0017181	1.15	0.9033333	0.0001000	3.57E-04	0.0000001	0.0001428	0.0004577	0.92	0.7143537
		0.0001600	1.33E-03	0.0000029	0.0003314	0.0016598	1.11	0.8933333	0.0001100	3.67E-04	0.0000002	0.0001333	0.0004774	0.95	0.7334287
D+N=N:D		0.0001750	1.30E-03	0.0000041	0.0002982	0.0016069	1.07	0.8833333	0.0001200	3.75E-04	0.0000003	0.0001249	0.0004961	0.99	0.7501295
K1=[N]*[D]/[N:D]		0.0001900	1.28E-03	0.0000056	0.0002699	0.0015576	1.04	0.8733333	0.0001300	3.82E-04	0.0000004	0.0001176	0.0005139	1.03	0.7648771
		0.0002050	1.26E-03	0.0000074	0.0002455	0.0015108	1.01	0.8633333	0.0001400	3.89E-04	0.0000005	0.0001110	0.0005310	1.06	0.7779989
N:D+nD+N:D:Dn		0.0002200	1.23E-03	0.0000096	0.0002240	0.0014655	0.98	0.8533333	0.0001500	3.94E-04	0.0000007	0.0001051	0.0005475	1.10	0.7897538
K2=[N:D][D]n/[N:D:Dn]		0.0002350	1.20E-03	0.0000122	0.0002049	0.0014210	0.95	0.8433333	0.0001600	3.99E-04	0.0000009	0.0000998	0.0005637	1.13	0.8003489
		0.0002500	1.17E-03	0.0000153	0.0001878	0.0013767	0.92	0.8333333	0.0001700	4.04E-04	0.0000011	0.0000950	0.0005795	1.16	0.8099521
		0.0002650	1.14E-03	0.0000188	0.0001723	0.0013322	0.89	0.8233333	0.0001800	4.08E-04	0.0000014	0.0000906	0.0005951	1.19	0.8187009
[N]total	0.00050	0.0002800	1.11E-03	0.0000227	0.0001581	0.0012874	0.86	0.8133333	0.0001900	4.12E-04	0.0000018	0.0000866	0.0006105	1.22	0.8267088
		0.0002950	1.07E-03	0.0000270	0.0001451	0.0012420	0.83	0.8033333	0.0002000	4.15E-04	0.0000022	0.0000830	0.0006259	1.25	0.8340708
		0.0003100	1.03E-03	0.0000317	0.0001331	0.0011961	0.80	0.7933333	0.0002100	4.18E-04	0.0000027	0.0000796	0.0006413	1.28	0.8408666
		0.0003250	9.91E-04	0.0000368	0.0001219	0.0011496	0.77	0.7833333	0.0002200	4.20E-04	0.0000033	0.0000764	0.0006567	1.31	0.8471637
		0.0003400	9.49E-04	0.0000423	0.0001116	0.0011026	0.74	0.7733333	0.0002300	4.23E-04	0.0000039	0.0000735	0.0006723	1.34	0.8530198
		0.0003550	9.05E-04	0.0000479	0.0001020	0.0010553	0.70	0.7633333	0.0002400	4.25E-04	0.0000047	0.0000708	0.0006880	1.38	0.8584843
		0.0003700	8.61E-04	0.0000538	0.0000931	0.0010079	0.67	0.7533333	0.0002500	4.26E-04	0.0000056	0.0000682	0.0007040	1.41	0.8636000
		0.0003850	8.16E-04	0.0000598	0.0000848	0.0009607	0.64	0.7433333	0.0002600	4.28E-04	0.0000065	0.0000658	0.0007203	1.44	0.8684039
		0.0004000	7.71E-04	0.0000658	0.0000771	0.0009139	0.61	0.7333333	0.0002700	4.29E-04	0.0000076	0.0000635	0.0007369	1.47	0.8729283
		0.0004150	7.26E-04	0.0000718	0.0000700	0.0008678	0.58	0.7233333	0.0002800	4.30E-04	0.0000088	0.0000614	0.0007538	1.51	0.8772015
		0.0004300	6.82E-04	0.0000777	0.0000634	0.0008227	0.55	0.7133333	0.0002900	4.30E-04	0.0000101	0.0000594	0.0007712	1.54	0.8812482
		0.0004450	6.38E-04	0.0000834	0.0000573	0.0007788	0.52	0.7033333	0.0003000	4.31E-04	0.0000116	0.0000575	0.0007891	1.58	0.8850905
		0.0004600	5.96E-04	0.0000889	0.0000518	0.0007362	0.49	0.6933333	0.0003100	4.31E-04	0.0000133	0.0000556	0.0008075	1.61	0.8887477
		0.0004750	5.55E-04	0.0000941	0.0000467	0.0006953	0.46	0.6833333	0.0003200	4.31E-04	0.0000151	0.0000539	0.0008264	1.65	0.8922370
		0.0004900	5.15E-04	0.0000990	0.0000420	0.0006561	0.44	0.6733333	0.0003300	4.31E-04	0.0000170	0.0000522	0.0008459	1.69	0.8955737
		0.0005050	4.77E-04	0.0001035	0.0000378	0.0006188	0.41	0.6633333	0.0003400	4.30E-04	0.0000192	0.0000506	0.0008660	1.73	0.8987714
		0.0005200	4.42E-04	0.0001077	0.0000340	0.0005834	0.39	0.6533333	0.0003500	4.29E-04	0.0000215	0.0000491	0.0008868	1.77	0.9018422
		0.0005350	4.08E-04	0.0001114	0.0000305	0.0005499	0.37	0.6433333	0.0003600	4.28E-04	0.0000240	0.0000476	0.0009083	1.82	0.9047971
		0.0005500	3.76E-04	0.0001148	0.0000274	0.0005183	0.35	0.6333333	0.0003700	4.27E-04	0.0000267	0.0000462	0.0009306	1.86	0.9076457
		0.0005650	3.47E-04	0.0001177	0.0000245	0.0004887	0.33	0.6233333	0.0003800	4.26E-04	0.0000296	0.0000448	0.0009535	1.91	0.9103966
		0.0005800	3.19E-04	0.0001202	0.0000220	0.0004610	0.31	0.6133333	0.0003900	4.24E-04	0.0000327	0.0000435	0.0009773	1.95	0.9130576
		0.0005950	2.93E-04	0.0001224	0.0000197	0.0004351	0.29	0.6033333	0.0004000	4.22E-04	0.0000360	0.0000422	0.0010018	2.00	0.9156355
		0.0006100	2.69E-04	0.0001242	0.0000176	0.0004109	0.27	0.5933333	0.0004100	4.20E-04	0.0000395	0.0000409	0.0010271	2.05	0.9181366
		0.0006250	2.47E-04	0.0001256	0.0000158	0.0003884	0.26	0.5833333	0.0004200	4.17E-04	0.0000433	0.0000397	0.0010533	2.11	0.9205662
		0.0006400	2.27E-04	0.0001267	0.0000142	0.0003674	0.24	0.5733333	0.0004300	4.14E-04	0.0000472	0.0000385	0.0010803	2.16	0.9229291
		0.0006550	2.08E-04	0.0001275	0.0000127	0.0003479	0.23	0.5633333	0.0004400	4.11E-04	0.0000514	0.0000374	0.0011081	2.22	0.9252297
		0.0006700	1.90E-04	0.0001279	0.0000114	0.0003297	0.22	0.5533333	0.0004500	4.08E-04	0.0000558	0.0000363	0.0011368	2.27	0.9274718
		0.0006850	1.75E-04	0.0001281	0.0000102	0.0003128	0.21	0.5433333	0.0004600	4.04E-04	0.0000604	0.0000352	0.0011663	2.33	0.9296585
		0.0007000	1.60E-04	0.0001280	0.0000091	0.0002971	0.20	0.5333333	0.0004700	4.01E-04	0.0000652	0.0000341	0.0011966	2.39	0.9317928
		0.0007150	1.47E-04	0.0001277	0.0000082	0.0002825	0.19	0.5233333	0.0004800	3.97E-04	0.0000702	0.0000331	0.0012277	2.46	0.9338772
		0.0007300	1.34E-04	0.0001271	0.0000074	0.0002688	0.18	0.5133333	0.0004900	3.93E-04	0.0000754	0.0000320	0.0012597	2.52	0.9359137
		0.0007450	1.23E-04	0.0001264	0.0000066	0.0002561	0.17	0.5033333	0.0005000	3.88E-04	0.0000809	0.0000310	0.0012924	2.58	0.9379043
		0.0007600	1.13E-04	0.0001254	0.0000059	0.0002442	0.16	0.4933333	0.0005100	3.83E-04	0.0000865	0.0000301	0.0013258	2.65	0.9398504
		0.0007750	1.03E-04	0.0001243	0.0000053	0.0002330	0.16	0.4833333	0.0005200	3.79E-04	0.0000923	0.0000291	0.0013600	2.72	0.9417533
		0.0007900	9.48E-05	0.0001230	0.0000048	0.0002226	0.15	0.4733333	0.0005300	3.74E-04	0.0000983	0.0000282	0.0013948	2.79	0.9436142
		0.0008050	8.69E-05	0.0001216	0.0000043	0.0002128	0.14	0.4633333	0.0005400	3.68E-04	0.0001044	0.0000273	0.0014303	2.86	0.9454338
		0.0008200	7.97E-05	0.0001201	0.0000039	0.0002036	0.14	0.4533333	0.0005500	3.63E-04	0.0001107	0.0000264	0.0014664	2.93	0.9472129
		0.0008350	7.31E-05	0.0001184	0.0000035	0.0001949	0.13	0.4433333	0.0005600	3.57E-04	0.0001171	0.0000255	0.0015030	3.01	0.9489521
		0.0008500	6.70E-05	0.0001166	0.0000032	0.0001868	0.12	0.4333333	0.0005700	3.52E-04	0.0001237	0.0000247	0.0015402	3.08	0.9506516
		0.0008650	6.15E-05	0.0001147	0.0000028	0.0001790	0.12	0.4233333	0.0005800	3.46E-04	0.0001304	0.0000238	0.0015778	3.16	0.9523120
		0.0008800	5.64E-05	0.0001127	0.0000026	0.0001717	0.11	0.4133333	0.0005900	3.40E-04	0.0001372	0.0000230	0.0016159	3.23	0.9539334
		0.0008950	5.17E-05	0.0001107	0.0000023	0.0001647	0.11	0.4033333	0.0006000	3.34E-04	0.0001441	0.0000222	0.0016543	3.31	0.9555160
		0.0009100	4.75E-05	0.0001085	0.0000021	0.0001581	0.11	0.3933333	0.0006100	3.27E-04	0.0001511	0.0000215	0.0016930	3.39	0.9570600
		0.0009250	4.36E-05	0.0001063	0.0000019	0.0001517	0.10	0.3833333	0.0006200	3.21E-04	0.0001582	0.0000207	0.0017319	3.46	0.9585654
		0.0009400	4.00E-05	0.0001040	0.0000017	0.0001457	0.10	0.3733333	0.0006300	3.15E-04	0.0001653	0.0000200	0.0017711	3.54	0.9600325
		0.0009550	3.67E-05	0.0001017	0.0000015	0.0001399	0.09	0.3633333	0.0006400	3.08E-04	0.0001724	0.0000193	0.0018104	3.62	0.9614612
		0.0009700	3.36E-05	0.0000993	0.0000014	0.0001343	0.09	0.3533333	0.0006500	3.02E-04	0.0001796	0.0000186	0.0018498	3.70	0.9628516
		0.0009850	3.09E-05	0.0000968	0.0000013	0.0001289	0.09	0.3433333	0.0006600	2.95E-04	0.0001868	0.0000179	0.0018892	3.78	0.9642040
		0.0010000	2.83E-05	0.0000943	0.0000011	0.0001238	0.08	0.3333333	0.0006700	2.89E-04	0.0001940	0.0000172	0.0019287	3.86	0.9655185
		0.0010150	2.60E-05	0.0000918	0.0000010	0.0001188	0.08	0.3233333	0.0006800	2.82E-04	0.0002012	0.0000166	0.0019680	3.94	0.9667952
		0.0010300	2.38E-05	0.0000892	0.0000009	0.0001140	0.08	0.3133333	0.0006900	2.76E-04	0.0002083	0.0000160	0.0020073	4.01	0.9680344
		0.0010450	2.18E-05	0.0000866	0.0000008	0.0001093	0.07	0.3033333	0.0007000	2.69E-04	0.0002154	0.0000154	0.0020464	4.09	0.9692363
		0.0010600	2.00E-05	0.0000840	0.0000008	0.0001047	0.07	0.2933333	0.0007100	2.63E-04	0.0002225	0.0000148	0.0020852	4.17	0.9704013
		0.0010750	1.83E-05	0.0000813	0.0000007	0.0001003	0.07	0.2833333	0.0007200	2.56E-04	0.0002295	0.0000142	0.0021239	4.25	0.9715297
		0.0010900	1.67E-05	0.0000787	0.0000006	0.0000960	0.06	0.2733333	0.0007300	2.50E-04	0.0002365	0.0000137	0.0021623	4.32	0.9726219
		0.0011050	1.53E-05	0.0000759	0.0000006	0.0000918	0.06	0.2633333	0.0007400	2.43E-04	0.0002434	0.0000132	0.0022003	4.40	0.9736785
		0.0011200	1.40E-05	0.0000732	0.0000005	0.0000877	0.06	0.2533333	0.0007500	2.37E-04	0.0002502	0.0000127	0.0022380	4.48	0.9746998
		0.0011350	1.27E-05	0.0000705	0.0000004	0.0000836	0.06	0.2433333	0.0007600	2.31E-04	0.0002569	0.0000122	0.0022753	4.55	0.9756865
		0.0011500	1.16E-05	0.0000677	0.0000004	0.0000797	0.05	0.2333333	0.0007700	2.25E-04	0.0002635	0.0000117	0.0023122	4.62	0.9766391
		0.0011650	1.06E-05	0.0000649	0.0000004	0.0000758	0.05	0.2233333	0.0007800	2.19E-04	0.0002700	0.0000112	0.0023487	4.70	0.9775582
		0.0011800	9.61E-06	0.0000621	0.0000003	0.0000720	0.05	0.2133333	0.0007900	2.13E-04	0.0002764	0.0000108	0.0023847	4.77	0.9784446
		0.0011950	8.72E-06	0.0000593	0.0000003	0.0000683	0.05	0.2033333	0.0008000	2.07E-04	0.0002826	0.0000104	0.0024202	4.84	0.9792989
		0.0012100	7.90E-06	0.0000564	0.0000003	0.0000646	0.04	0.1933333	0.0008100	2.01E-04	0.0002888	0.0000099	0.0024552	4.91	0.9801219
		0.0012250	7.14E-06	0.0000536	0.0000002	0.0000609	0.04	0.1833333	0.0008200	1.96E-04	0.0002948	0.0000095	0.0024898	4.98	0.9809142
		0.0012400	6.44E-06	0.0000507	0.0000002	0.0000574	0.04	0.1733333	0.0008300	1.90E-04	0.0003007	0.0000092	0.0025238	5.05	0.9816767

Appendix F

Definining distance restraints

F.1 extract_edfilt.pl

This script extracts distance restraints from assignments of intermolecular contacts in an edited, filtered NOE spectrum. The input data is a three-dimensional NMRView peak list with the direct dimension being the ligand frequency. In the NMRView peaklist, protein assignmentes are given explicitly but all ligand resonances are labeled "99.HX" and are assigned within the body of this script. Note that the protein assignments can be ambiguous but only if there are two ambiguous atoms.

Importantly, this script does not do any calibration based on the intensity of the NOEs. All restraints are calibrated between the same range, which is defined in the header of the script.

```
#!/usr/bin/perl -w
#rmbh08
#for the edited, filtered, NOESY, assigned in NMRView
#applies approximate assignments of ring, tail protons
#handling of ambiguous protein atom assignments
#this version: assignments validated by BDS. H6 and H8 assignments switched.

$version=0.7;  #Jul 3, 2008

$dist = 6;
$lower = $dist-1.8;
$upper = 0.1;

my @assignedLines;
my %ambiguousProteinAtoms;

while(<>){
    if(m,99.HX,){                                #all ligand shifts are assigned as the same atom
if(m,(99.HX\s?){2},){                  #treatment for two ambiguous protein protons
    m,(\d+)\s+\{99.HX 99.HX\}(.+?)\{(\d+.\w+) (\d+.\w+)\}(.+?)\{(\d+.\w+) (\d+.\w+)\}(.+),;
    push(@assignedLines, "$1 \{99.HX\}$2\{$3\}$5\{$6\}$8\n");
    push(@assignedLines, "$1 \{99.HX\}$2\{$4\}$5\{$7\}$8\n");
    $ambiguousProteinAtoms{$1}="$3 $4"; #ambiguous peaks are tracked by index
    next}
push(@assignedLines,$_)}}

print "!generated with 'extract_edfilt v.$version'\n";
for(@assignedLines){
    @entries = split;
    $index = $entries[0];
    $ligandShift = $entries[2];
    $protonName = $entries[8];
#    $protonShift = $entries[9];
#    $carbonName = $entries[15];
```

```
#    $carbonShift = $entries[16];
    next if (exists $seen{$index}); #skip entries for ambiguous atoms that have already been 'seen'

#cleaning up atom names
    if( exists $ambiguousProteinAtoms{$index}){
    $protonName = $ambiguousProteinAtoms{$index};
    $protonName =~ s,(\d+)\.(\w+) (\d+)\.(\w+),(resid $1 and name $2) or (resid $3 and name $4),;
    $seen{$index}++;
}else{
    $protonName =~ s,[\{\}],,g;
    $protonName =~ s,(\d+)\.(\w+),resid $1 and name $2,g;}

#making protein protons into pseudoatoms
    $protonName =~ s,HB1,HB\#,g;
    $protonName =~ s,HE1,HE\#,g;
    $protonName =~ s,HD11,HD1\#,g;
    $protonName =~ s,HD21,HD2\#,g;
    $protonName =~ s,HG11,HG1\#,g;
    $protonName =~ s,HG21,HG2\#,g;

#W7 assignments are made here
#    print "DEV: W7 shift is $ligandShift\n";
    if( $ligandShift < 4 ){ $type = 'tail'}
    elsif( $ligandShift > 8.6 ){ $type = 'ringH4'}
    elsif( $ligandShift > 8.48 ){ $type = 'ringH8'}
    elsif( $ligandShift > 8.25 ){ $type = 'ringH2'}
    elsif( $ligandShift > 7.6 ){ $type = 'ambigRing'}
    elsif( $ligandShift > 7.53 ){ $type = 'ringH7'}
    else{ die "no W7 assignment; this shouldn't happen\n" }

#Distance restraint selections are defined here
    if( $type eq 'tail' ){ $W7name = "name  \#H1\#" }
    elsif( $type eq 'ringH2' ){ $W7name = "name \#H2"}
    elsif( $type eq 'ringH4' ){ $W7name = "name \#H4"}
    elsif( $type eq 'ringH8' ){ $W7name = "name \#H8"}
    elsif( $type eq 'ringH7' ){ $W7name = "name \#H7"}
    elsif( $type eq 'ambigRing'){ $W7name = "name \#H3 or name \#H6" }
    else{die "no W7 selection(2); this shouldn't happen(2)\n"}

#distance restraints are writen here
    print "assign ( $protonName ) (segid WW7 and resid 1 and ($W7name) )";
    print " $dist $lower $upper\n";
}
```

F.2 The distance restraint table "intermolecularNOEsFinal1.tbl"

```
!generated with 'extract_edfilt v.0.7'
assign ( (resid 48 and name HD1#) or (resid 44 and name HG2#) ) (segid WW7 and resid 1 and (name #H4) ) 5 3.2 0.1
assign ( (resid 48 and name HD1#) or (resid 44 and name HG2#) ) (segid WW7 and resid 1 and (name #H3 or name #H6) ) 5 3.2 0.1
assign ( resid 82 and name HG1# ) (segid WW7 and resid 1 and (name #H3 or name #H6) ) 5 3.2 0.1
assign ( (resid 79 and name HG1#) or (resid 44 and name HG1#) ) (segid WW7 and resid 1 and (name #H8) ) 5 3.2 0.1
assign ( (resid 79 and name HG1#) or (resid 44 and name HG1#) ) (segid WW7 and resid 1 and (name #H2) ) 5 3.2 0.1
assign ( (resid 79 and name HG1#) or (resid 44 and name HG1#) ) (segid WW7 and resid 1 and (name #H3 or name #H6) ) 5 3.2 0.1
assign ( resid 23 and name HB# ) (segid WW7 and resid 1 and (name #H2) ) 5 3.2 0.1
assign ( resid 23 and name HB# ) (segid WW7 and resid 1 and (name #H3 or name #H6) ) 5 3.2 0.1
assign ( resid 23 and name HB# ) (segid WW7 and resid 1 and (name #H7) ) 5 3.2 0.1
assign ( resid 45 and name HE# ) (segid WW7 and resid 1 and (name #H4) ) 5 3.2 0.1
assign ( resid 45 and name HE# ) (segid WW7 and resid 1 and (name #H2) ) 5 3.2 0.1
assign ( resid 45 and name HE# ) (segid WW7 and resid 1 and (name #H3 or name #H6) ) 5 3.2 0.1
assign ( resid 60 and name HE# ) (segid WW7 and resid 1 and (name #H2) ) 5 3.2 0.1
assign ( resid 81 and name HE# ) (segid WW7 and resid 1 and (name #H3 or name #H6) ) 5 3.2 0.1
assign ( resid 81 and name HE# ) (segid WW7 and resid 1 and (name #H7) ) 5 3.2 0.1
assign ( resid 26 and name HD1# ) (segid WW7 and resid 1 and (name #H3 or name #H6) ) 5 3.2 0.1
assign ( (resid 44 and name HG2#) or (resid 48 and name HD1#) ) (segid WW7 and resid 1 and (name  #H1#) ) 5 3.2 0.1
assign ( resid 23 and name HB# ) (segid WW7 and resid 1 and (name  #H1#) ) 5 3.2 0.1
assign ( resid 45 and name HE# ) (segid WW7 and resid 1 and (name  #H1#) ) 5 3.2 0.1
assign ( resid 60 and name HE# ) (segid WW7 and resid 1 and (name  #H1#) ) 5 3.2 0.1
assign ( resid 81 and name HE# ) (segid WW7 and resid 1 and (name  #H1#) ) 5 3.2 0.1
assign ( resid 26 and name HD1# ) (segid WW7 and resid 1 and (name  #H1#) ) 5 3.2 0.1
```

Appendix G

Structure calculations

G.1 Script for NOE-based docking in Xplor-NIH

```
xplor.requireVersion("2.18")
xplor.parseArguments()

##this script is named 'anneal.py'
##adapted from: gb1-anneal.py, originally by CDS
##modified by RMBH for NOE-based docking of W7 onto NTnC-Ca2+
##former file name: gb1-anneal-clean-singconf-collapse.py
# slow cooling protocol in torsion angle space for protein G. Uses
# NOE, RDC, J-coupling restraints.
# this script performs annealing from an extended structure
# CDS 2005/05/10

proteinTab="noe-tabs/1MXL.tbl"
W7noeTab="noe-tabs/intermolecularNOEsFinal1.tbl"

outFilename = "anneal_STRUCTURE.pdb"
numberOfStructures=50

import protocol
protocol.initRandomSeed()   #set random seed - by time
command = xplor.command
seed=1325 #I don't know whether this is doing anything

mySEQ = "gen-prot/tnc-acys.seq"
from psfGen import seqToPSF
seqToPSF(mySEQ,startResid=1)
protocol.initCoords("gen-prot/1LXFnoTnInoBep_ACYS_nochainID.pdb")
protocol.addUnknownAtoms()

command("parameter @gen-w7/w7-noDIHE.par end")
command("structure @gen-w7/w7x.psf end")
command("coordinates @gen-w7/w7x.pdb end")

command("parameter @gen-ca/calcium.par end")
command("structure @gen-ca/calcium.psf end")
command("coordinates @gen-ca/calcium2.pdb end")
```

```
W7restrainString = """resname WW7 and (name CL1 or name C5 or name C6
                     or name 1H6 or name C7  or name 1H7 or name C8
                     or name 1H8 or name C9  or name C10 or name C4
                     or name 1H4 or name C3  or name 1H3 or name C2
                     or name 1H2 or name C1)"""
BBrestrainString = """name C or name CA or name N or name O"""

from potList import PotList
potList = PotList()

from simulationTools import MultRamp, StaticRamp, InitialParams, FinalParams

rampedParams=[]
highTempParams=[]

# set up NOE potential
noe=PotList('noe')
potList.append(noe)
from noePotTools import create_NOEPot
for (name,scale,file) in [('protein',1,proteinTab),
                          ('calcium',1, "noe-tabs/Ndom_calcium.tbl"),
                          ]:
    pot = create_NOEPot(name,file)
    pot.setPotType("soft") #- if you think there may be bad NOEs
    pot.setScale(scale)
    noe.append(pot)
rampedParams.append( MultRamp(2,30, "noe.setScale( VALUE )") )

noeW7=PotList('noeW7')
potList.append(noeW7)
for (name,scale,file) in [('w7',1,W7noeTab)]:
    pot = create_NOEPot(name,file)
    pot.setPotType("hard")
    pot.setScale(scale)
    noeW7.append(pot)
rampedParams.append( MultRamp(2,30, "noeW7.setScale( VALUE )") )

from xplorPot import XplorPot

#Rama torsion angle database
protocol.initRamaDatabase()
potList.append( XplorPot('RAMA') )
rampedParams.append( MultRamp(.002,1,"potList['RAMA'].setScale(VALUE)") )

# setup parameters for atom-atom repulsive term. (van der Waals-like term)
potList.append( XplorPot('VDW') )
rampedParams.append( StaticRamp("protocol.initNBond()") )
rampedParams.append( MultRamp(0.9,0.8,
                              "command('param nbonds repel VALUE end end')") )
rampedParams.append( MultRamp(.004,4,
                              "command('param nbonds rcon VALUE end end')") )
# nonbonded interaction only between CA atoms
highTempParams.append( StaticRamp("""protocol.initNBond(cutnb=100,
                                                        tolerance=45,
                                                        repel=1.2,
```

```
                                                        onlyCA=1)""") )
potList.append( XplorPot("BOND") )
potList.append( XplorPot("ANGL") )
potList['ANGL'].setThreshold( 5 )
rampedParams.append( MultRamp(0.4,1,"potList['ANGL'].setScale(VALUE)") )
potList.append( XplorPot("IMPR") )
potList['IMPR'].setThreshold( 5 )
rampedParams.append( MultRamp(0.1,1,"potList['IMPR'].setScale(VALUE)") )

# Give atoms uniform weights
from atomAction import SetProperty
AtomSel("all            ").apply( SetProperty("mass",100.) )
AtomSel("all            ").apply( SetProperty("fric",10.) )

# IVM setup
from ivm import IVM

tors = IVM()
tors.group(W7restrainString)
tors.group(BBrestrainString)
print tors.groupList()
protocol.torsionTopology(tors)

tors3 = IVM()
tors3.group(W7restrainString)
print tors3.groupList()
protocol.torsionTopology(tors3)

cart = IVM()
protocol.initMinimize(cart)
cart.group(W7restrainString)
print cart.groupList()
protocol.cartesianTopology(cart)

# objects which perform simulated annealing
from simulationTools import AnnealIVM
init_t  = 2500.     # Need high temp and slow annealing to converge
dockCool = AnnealIVM(initTemp =init_t,
                finalTemp=25,
                tempStep =100,
                ivm=tors, #tors has restrained backbone atoms
                rampedParams = rampedParams)
torCool = AnnealIVM(initTemp =1500,
                finalTemp=25,
                tempStep =12.5,
                ivm=tors,
                rampedParams = rampedParams)

def calcOneStructure(loopInfo):
    """ this function calculates a single structure, performs analysis on the
    structure, and then writes out a pdb file, with remarks.
    """
    # initialize parameters for high temp dynamics.
    InitialParams( rampedParams )
    # high-temp dynamics setup - only need to specify parameters which
```

```
#   differfrom initial values in rampedParams
InitialParams( highTempParams )
# high temp dynamics --  crude docking of W7 (protein coordinates are already minimize
protocol.initDynamics(tors,
                      potList=potList, # potential terms to use
                      bathTemp=init_t,
                      initVelocities=1,
                      finalTime=200,    # stops at 800ps or 8000 steps
                      numSteps=2000,    # whichever comes first
                      printInterval=100/2)
tors.setETolerance( init_t/1000 )  #used to det. stepsize. default: t/1000
tors.run()

# initialize parameters for cooling loop
InitialParams( rampedParams )
# initialize integrator for simulated annealing
protocol.initDynamics(tors,
                      potList=potList,
                      numSteps=100,       #at each temp: 100 steps or
                      finalTime=.2 ,       # .2ps, whichever is less
                      printInterval=100)
# perform simulated annealing
dockCool.run()

# initialize parameters for cooling loop
InitialParams( rampedParams )
# initialize integrator for simulated annealing
protocol.initDynamics(tors,
                      potList=potList,
                      numSteps=100,       #at each temp: 100 steps or
                      finalTime=.2 ,       # .2ps, whichever is less
                      printInterval=100)
# perform simulated annealing
torCool.run()

#low temp dynamics -- relaxing backbone positions
FinalParams( rampedParams )
protocol.initDynamics(tors3,
                      potList=potList, # potential terms to use
                      bathTemp=25,
                      initVelocities=0,
                      finalTime=200,    # stops at 800ps or 8000 steps
                      numSteps=2000,    # whichever comes first
                      printInterval=100/2)
tors3.setETolerance( init_t/1000 )  #used to det. stepsize. default: t/1000
tors3.run()


# final torsion angle minimization
protocol.initMinimize(tors,
                      printInterval=50)
tors.run()

# final all- atom minimization
protocol.initMinimize(cart,
```

```
                                potList=potList,
                                dEPred=10)
    cart.run()
    #do analysis and write structure
    loopInfo.writeStructure(potList)
    pass

from simulationTools import StructureLoop
StructureLoop(numStructures=numberOfStructures,
              pdbTemplate=outFilename,
              structLoopAction=calcOneStructure).run()
```

G.2 Virtualization of W7

G.2.1 Coordinates (.pdb file)

```
The following file is named ``w7x.pdb'':
REMARK FILENAME="w7x.pdb"
REMARK Created by XPLO2D V. 031127/3.2.1 at Mon May 10 13:07:13 2004 for rmh
REMARK XPLO2D pseudo-PDB file
REMARK B-factors   <===> atom types
REMARK Occupancies <===> nr of hydrogens
REMARK DATE:13-Feb-06  15:45:47       created by user: rmh
ATOM      1  CL1 WW7     1       5.624  -0.272   1.534  1.00  1.00      WW7
ATOM      2  C5  WW7     1       4.167  -0.725   0.726  1.00  2.00      WW7
ATOM      3  C6  WW7     1       3.969  -2.139   0.444  1.00  3.00      WW7
ATOM      4  1H6 WW7     1       4.670  -2.802   0.708  1.00  4.00      WW7
ATOM      5  C7  WW7     1       2.749  -2.597  -0.224  1.00  5.00      WW7
ATOM      6  1H7 WW7     1       2.626  -3.571  -0.412  1.00  6.00      WW7
ATOM      7  C8  WW7     1       1.713  -1.653  -0.618  1.00  7.00      WW7
ATOM      8  1H8 WW7     1       0.884  -1.978  -1.073  1.00  8.00      WW7
ATOM      9  C9  WW7     1       1.891  -0.231  -0.345  1.00  9.00      WW7
ATOM     10  C10 WW7     1       3.152   0.249   0.345  1.00 10.00      WW7
ATOM     11  C4  WW7     1       3.327   1.669   0.628  1.00 11.00      WW7
ATOM     12  1H4 WW7     1       4.156   1.993   1.084  1.00 12.00      WW7
ATOM     13  C3  WW7     1       2.288   2.610   0.242  1.00 13.00      WW7
ATOM     14  1H3 WW7     1       2.407   3.584   0.436  1.00 14.00      WW7
ATOM     15  C2  WW7     1       1.070   2.151  -0.426  1.00 15.00      WW7
ATOM     16  1H2 WW7     1       0.369   2.815  -0.688  1.00 16.00      WW7
ATOM     17  C1  WW7     1       0.867   0.737  -0.714  1.00 17.00      WW7
ATOM     18  S1  WW7     1      -0.632   0.233  -1.504  1.00 18.00      WW7
ATOM     19  O1  WW7     1      -1.066  -1.009  -0.941  1.00 19.00      WW7
ATOM     20  O2  WW7     1      -0.523   0.370  -2.927  1.00 20.00      WW7
ATOM     21  N1  WW7     1      -1.851   1.375  -1.022  1.00 21.00      WW7
ATOM     22  HAD WW7     1      -2.739   0.986  -1.267  1.00 22.00      WW7
ATOM     23  C11 WW7     1      -1.908   1.742   0.391  1.00 23.00      WW7
ATOM     24 1H11 WW7     1      -1.276   2.501   0.545  1.00 24.00      WW7
ATOM     25 2H11 WW7     1      -1.621   0.953   0.934  1.00 25.00      WW7
ATOM     26  C12 WW7     1      -3.298   2.169   0.860  1.00 26.00      WW7
ATOM     27 1H12 WW7     1      -3.618   2.933   0.300  1.00 27.00      WW7
ATOM     28 2H12 WW7     1      -3.253   2.454   1.818  1.00 28.00      WW7
ATOM     29  C13 WW7     1      -4.270   0.995   0.723  1.00 29.00      WW7
ATOM     30 1H13 WW7     1      -4.304   0.701  -0.232  1.00 30.00      WW7
ATOM     31 2H13 WW7     1      -5.182   1.282   1.017  1.00 31.00      WW7
ATOM     32  C14 WW7     1      -3.796  -0.167   1.597  1.00 32.00      WW7
ATOM     33 1H14 WW7     1      -3.700   0.144   2.543  1.00 33.00      WW7
ATOM     34 2H14 WW7     1      -2.913  -0.497   1.263  1.00 34.00      WW7
ATOM     35  C15 WW7     1      -4.824  -1.303   1.534  1.00 35.00      WW7
ATOM     36 1H15 WW7     1      -5.713  -0.965   1.843  1.00 36.00      WW7
ATOM     37 2H15 WW7     1      -4.528  -2.053   2.126  1.00 37.00      WW7
ATOM     38  C16 WW7     1      -4.943  -1.808   0.093  1.00 38.00      WW7
ATOM     39 1H16 WW7     1      -5.187  -1.039  -0.497  1.00 39.00      WW7
ATOM     40 2H16 WW7     1      -5.664  -2.500   0.054  1.00 40.00      WW7
ATOM     41  N2  WW7     1      -3.705  -2.396  -0.396  1.00 41.00      WW7
ATOM     42  HAB WW7     1      -2.977  -1.711  -0.369  1.00 42.00      WW7
ATOM     43  HAC WW7     1      -3.453  -3.172   0.182  1.00 43.00      WW7
ATOM     44  HAA WW7     1      -3.833  -2.709  -1.337  1.00 44.00      WW7
END
```

G.2.2 Topology (.psf file)

The following file is named "w7x.psf":

```
PSF

       6 !NTITLE
 REMARKS FILENAME="w7x.psf"
 REMARKS Created by XPLO2D V. 031127/3.2.1 at Mon May 10 13:07:13 2004 for rmh
 REMARKS XPLO2D pseudo-PDB file
 REMARKS B-factors   <===> atom types
 REMARKS Occupancies <===> nr of hydrogens
 REMARKS DATE:13-Feb-06  15:45:47       created by user: rmh
```

```
44 !NATOM
 1 WW7  1    WW7  CL1  CLW1   0.000000E+00   35.4530           0
 2 WW7  1    WW7  C5   CW2    0.000000E+00   12.0110           0
 3 WW7  1    WW7  C6   CW3    0.000000E+00   12.0110           0
 4 WW7  1    WW7  1H6  HW4    0.000000E+00   1.00800           0
 5 WW7  1    WW7  C7   CW5    0.000000E+00   12.0110           0
 6 WW7  1    WW7  1H7  HW6    0.000000E+00   1.00800           0
 7 WW7  1    WW7  C8   CW7    0.000000E+00   12.0110           0
 8 WW7  1    WW7  1H8  HW8    0.000000E+00   1.00800           0
 9 WW7  1    WW7  C9   CW9    0.000000E+00   12.0110           0
10 WW7  1    WW7  C10  CW10   0.000000E+00   12.0110           0
11 WW7  1    WW7  C4   CW11   0.000000E+00   12.0110           0
12 WW7  1    WW7  1H4  HW12   0.000000E+00   1.00800           0
13 WW7  1    WW7  C3   CW13   0.000000E+00   12.0110           0
14 WW7  1    WW7  1H3  HW14   0.000000E+00   1.00800           0
15 WW7  1    WW7  C2   CW15   0.000000E+00   12.0110           0
16 WW7  1    WW7  1H2  HW16   0.000000E+00   1.00800           0
17 WW7  1    WW7  C1   CW17   0.000000E+00   12.0110           0
18 WW7  1    WW7  S1   SW18   0.000000E+00   32.0660           0
19 WW7  1    WW7  O1   OW19   0.000000E+00   15.9990           0
20 WW7  1    WW7  O2   OW20   0.000000E+00   15.9990           0
21 WW7  1    WW7  N1   NW21   0.000000E+00   14.0070           0
22 WW7  1    WW7  HAD  HW22   0.000000E+00   1.00800           0
23 WW7  1    WW7  C11  CW23   0.000000E+00   12.0110           0
24 WW7  1    WW7  1H11 HW24   0.000000E+00   1.00800           0
25 WW7  1    WW7  2H11 HW25   0.000000E+00   1.00800           0
26 WW7  1    WW7  C12  CW26   0.000000E+00   12.0110           0
27 WW7  1    WW7  1H12 HW27   0.000000E+00   1.00800           0
28 WW7  1    WW7  2H12 HW28   0.000000E+00   1.00800           0
29 WW7  1    WW7  C13  CW29   0.000000E+00   12.0110           0
30 WW7  1    WW7  1H13 HW30   0.000000E+00   1.00800           0
31 WW7  1    WW7  2H13 HW31   0.000000E+00   1.00800           0
32 WW7  1    WW7  C14  CW32   0.000000E+00   12.0110           0
33 WW7  1    WW7  1H14 HW33   0.000000E+00   1.00800           0
34 WW7  1    WW7  2H14 HW34   0.000000E+00   1.00800           0
35 WW7  1    WW7  C15  CW35   0.000000E+00   12.0110           0
36 WW7  1    WW7  1H15 HW36   0.000000E+00   1.00800           0
37 WW7  1    WW7  2H15 HW37   0.000000E+00   1.00800           0
38 WW7  1    WW7  C16  CW38   0.000000E+00   12.0110           0
39 WW7  1    WW7  1H16 HW39   0.000000E+00   1.00800           0
40 WW7  1    WW7  2H16 HW40   0.000000E+00   1.00800           0
41 WW7  1    WW7  N2   NW41   0.000000E+00   14.0070           0
42 WW7  1    WW7  HAB  HW42   0.000000E+00   1.00800           0
43 WW7  1    WW7  HAC  HW43   0.000000E+00   1.00800           0
44 WW7  1    WW7  HAA  HW44   0.000000E+00   1.00800           0

45 !NBOND: bonds
 1       2       2       3       2      10       3       4
 3       5       5       6       5       7       7       8
 7       9       9      10       9      17      10      11
11      12      11      13      13      14      13      15
15      16      15      17      17      18      18      19
18      20      18      21      21      22      21      23
23      24      23      25      23      26      26      27
26      28      26      29      29      30      29      31
29      32      32      33      32      34      32      35
35      36      35      37      35      38      38      39
38      40      38      41      41      42      41      43
41      44

81 !NTHETA: angles
 1       2       3       1       2      10       3       2      10
 2       3       4       2       3       5       2      10       9
 2      10      11       4       3       5       3       5       6
 3       5       7       6       5       7       5       7       8
 5       7       9       8       7       9       7       9      10
 7       9      17      10       9      17       9      10      11
 9      17      15       9      17      18      10      11      12
10      11      13      12      11      13      11      13      14
11      13      15      14      13      15      13      15      16
13      15      17      16      15      17      15      17      18
17      18      19      17      18      20      17      18      21
19      18      20      19      18      21      20      18      21
18      21      22      18      21      23      22      21      23
21      23      24      21      23      25      21      23      26
24      23      25      24      23      26      25      23      26
23      26      27      23      26      28      23      26      29
27      26      28      27      26      29      28      26      29
26      29      30      26      29      31      26      29      32
30      29      31      30      29      32      31      29      32
29      32      33      29      32      34      29      32      35
33      32      34      33      32      35      34      32      35
32      35      36      32      35      37      32      35      38
36      35      37      36      35      38      37      35      38
35      38      39      35      38      40      35      38      41
39      38      40      39      38      41      40      38      41
38      41      42      38      41      43      38      41      44
42      41      43      42      41      44      43      41      44

62 !NPHI: dihedrals
 1       2       3       4       1       2       3       5
10       2       3       4      10       2       3       5
 1       2      10       9       1       2      10      11
 3       2      10       9       3       2      10      11
 2       3       5       6       2       3       5       7
 4       3       5       6       4       3       5       7
 3       5       7       8       3       5       7       9
 6       5       7       8       6       5       7       9
 5       7       9      10       5       7       9      17
 8       7       9      10       8       7       9      17
 7       9      10       2       7       9      10      11
```

```
 17       9      10       2      17       9      10      11
  7       9      17      15       7       9      17      18
 10       9      17      15      10       9      17      18
  2      10      11      12       2      10      11      13
  9      10      11      12       9      10      11      13
 10      11      13      14      10      11      13      15
 12      11      13      14      12      11      13      15
 11      13      15      16      11      13      15      17
 14      13      15      16      14      13      15      17
 13      15      17       9      13      15      17      18
 16      15      17       9      16      15      17      18
 21      23      26      28      24      23      26      29
 25      23      26      27      23      26      29      31
 27      26      29      32      28      26      29      30
 26      29      32      35      30      29      32      33
 31      29      32      34      29      32      35      37
 33      32      35      38      34      32      35      36
 32      35      38      40      36      35      38      41
 37      35      38      39      35      38      41      44
 39      38      41      43      40      38      41      42

 19 !NIMPHI: impropers
  2       1       3      10       3       2       4       5
  5       3       6       7       7       5       8       9
  9       7      10      17      10       2       9      11
 11      10      12      13      13      11      14      15
 15      13      16      17      17       9      15      18
 18      17      19      20      21      18      22      23
\item 23      21      24      25      26      23      27      28
 29      26      30      31      32      29      33      34
 35      32      36      37      38      35      39      40
 41      38      42      43

  0 !NDON: donors


  2 !NACC: acceptors
 19      18      20      18

  0 !NNB

  0       0       0       0       0       0       0       0
  0       0       0       0       0       0       0       0
  0       0       0       0       0       0       0       0
  0       0       0       0       0       0       0       0
  0       0       0       0       0       0       0       0
  0       0       0       0

  1       0 !NGRP
  0       0       0
```

G.2.3 Parameters (.par file)

The following file is named "w7-noDIHE.par".

```
Remarks ww7.par
Remarks Created by XPLO2D V. 031127/3.2.1 at Mon May 10 13:07:13 2004 for rmh
Remarks Auto-generated by XPLO2D from file W7_prdrg.pdb
Remarks Parameters for residue type WW7
Remarks IMPRoper terms for naphthalene ring are increased

 set echo=false end

 { edit if necessary }
 BOND CLW1 CW2    1000.0  1.727 ! Nobs =    1
 BOND CW2  CW3    1000.0  1.455 ! Nobs =    1
 BOND CW2  CW10   1000.0  1.457 ! Nobs =    1
 BOND CW3  HW4    1000.0  1.000 ! Nobs =    1
 BOND CW3  CW5    1000.0  1.464 ! Nobs =    1
 BOND CW5  HW6    1000.0  1.000 ! Nobs =    1
 BOND CW5  CW7    1000.0  1.456 ! Nobs =    1
 BOND CW7  HW8    1000.0  1.000 ! Nobs =    1
 BOND CW7  CW9    1000.0  1.459 ! Nobs =    1
 BOND CW9  CW10   1000.0  1.515 ! Nobs =    1
 BOND CW9  CW17   1000.0  1.457 ! Nobs =    1
 BOND CW10 CW11   1000.0  1.458 ! Nobs =    1
 BOND CW11 HW12   1000.0  1.000 ! Nobs =    1
 BOND CW11 CW13   1000.0  1.454 ! Nobs =    1
 BOND CW13 HW14   1000.0  1.000 ! Nobs =    1
 BOND CW13 CW15   1000.0  1.463 ! Nobs =    1
 BOND CW15 HW16   1000.0  1.000 ! Nobs =    1
 BOND CW15 CW17   1000.0  1.457 ! Nobs =    1
 BOND CW17 SW18   1000.0  1.768 ! Nobs =    1
 BOND SW18 OW19   1000.0  1.431 ! Nobs =    1
 BOND SW18 OW20   1000.0  1.434 ! Nobs =    1
 BOND SW18 NW21   1000.0  1.739 ! Nobs =    1
 BOND NW21 HW22   1000.0  1.000 ! Nobs =    1
 BOND NW21 CW23   1000.0  1.461 ! Nobs =    1
 BOND CW23 HW24   1000.0  1.000 ! Nobs =    1
 BOND CW23 HW25   1000.0  1.000 ! Nobs =    1
 BOND CW23 CW26   1000.0  1.528 ! Nobs =    1
 BOND CW26 HW27   1000.0  1.000 ! Nobs =    1
 BOND CW26 HW28   1000.0  1.001 ! Nobs =    1
 BOND CW26 CW29   1000.0  1.530 ! Nobs =    1
 BOND CW29 HW30   1000.0  1.000 ! Nobs =    1
 BOND CW29 HW31   1000.0  1.000 ! Nobs =    1
 BOND CW29 CW32   1000.0  1.529 ! Nobs =    1
 BOND CW32 HW33   1000.0  1.000 ! Nobs =    1
```

```
BOND CW32 HW34    1000.0   1.000 ! Nobs =    1
BOND CW32 CW35    1000.0   1.533 ! Nobs =    1
BOND CW35 HW36    1000.0   1.000 ! Nobs =    1
BOND CW35 HW37    1000.0   1.000 ! Nobs =    1
BOND CW35 CW38    1000.0   1.532 ! Nobs =    1
BOND CW38 HW39    1000.0   0.999 ! Nobs =    1
BOND CW38 HW40    1000.0   1.000 ! Nobs =    1
BOND CW38 NW41    1000.0   1.455 ! Nobs =    1
BOND NW41 HW42    1000.0   1.000 ! Nobs =    1
BOND NW41 HW43    1000.0   1.000 ! Nobs =    1
BOND NW41 HW44    1000.0   1.000 ! Nobs =    1

{ edit if necessary }
ANGLe CLW1 CW2  CW3      500.0   117.41 ! Nobs =    1
ANGLe CLW1 CW2  CW10     500.0   122.32 ! Nobs =    1
ANGLe CW3  CW2  CW10     500.0   120.26 ! Nobs =    1
ANGLe CW2  CW3  HW4      500.0   119.83 ! Nobs =    1
ANGLe CW2  CW3  CW5      500.0   120.37 ! Nobs =    1
ANGLe HW4  CW3  CW5      500.0   119.80 ! Nobs =    1
ANGLe CW3  CW5  HW6      500.0   119.54 ! Nobs =    1
ANGLe CW3  CW5  CW7      500.0   120.90 ! Nobs =    1
ANGLe HW6  CW5  CW7      500.0   119.56 ! Nobs =    1
ANGLe CW5  CW7  HW8      500.0   120.15 ! Nobs =    1
ANGLe CW5  CW7  CW9      500.0   119.64 ! Nobs =    1
ANGLe HW8  CW7  CW9      500.0   120.21 ! Nobs =    1
ANGLe CW7  CW9  CW10     500.0   119.70 ! Nobs =    1
ANGLe CW7  CW9  CW17     500.0   120.97 ! Nobs =    1
ANGLe CW10 CW9  CW17     500.0   119.33 ! Nobs =    1
ANGLe CW2  CW10 CW9      500.0   119.13 ! Nobs =    1
ANGLe CW2  CW10 CW11     500.0   121.09 ! Nobs =    1
ANGLe CW9  CW10 CW11     500.0   119.77 ! Nobs =    1
ANGLe CW10 CW11 HW12     500.0   120.22 ! Nobs =    1
ANGLe CW10 CW11 CW13     500.0   119.53 ! Nobs =    1
ANGLe HW12 CW11 CW13     500.0   120.25 ! Nobs =    1
ANGLe CW11 CW13 HW14     500.0   119.58 ! Nobs =    1
ANGLe CW11 CW13 CW15     500.0   120.87 ! Nobs =    1
ANGLe HW14 CW13 CW15     500.0   119.55 ! Nobs =    1
ANGLe CW13 CW15 HW16     500.0   119.65 ! Nobs =    1
ANGLe CW13 CW15 CW17     500.0   120.71 ! Nobs =    1
ANGLe HW16 CW15 CW17     500.0   119.65 ! Nobs =    1
ANGLe CW9  CW17 CW15     500.0   119.79 ! Nobs =    1
ANGLe CW9  CW17 SW18     500.0   121.32 ! Nobs =    1
ANGLe CW15 CW17 SW18     500.0   118.89 ! Nobs =    1
ANGLe CW17 SW18 OW19     500.0   109.20 ! Nobs =    1
ANGLe CW17 SW18 OW20     500.0   110.60 ! Nobs =    1
ANGLe CW17 SW18 NW21     500.0   106.46 ! Nobs =    1
ANGLe OW19 SW18 OW20     500.0   119.77 ! Nobs =    1
ANGLe OW19 SW18 NW21     500.0   104.38 ! Nobs =    1
ANGLe OW20 SW18 NW21     500.0   105.41 ! Nobs =    1
ANGLe SW18 NW21 HW22     500.0   107.41 ! Nobs =    1
ANGLe SW18 NW21 CW23     500.0   117.42 ! Nobs =    1
ANGLe HW22 NW21 CW23     500.0   107.46 ! Nobs =    1
ANGLe NW21 CW23 HW24     500.0   108.36 ! Nobs =    1
ANGLe NW21 CW23 HW25     500.0   108.41 ! Nobs =    1
ANGLe NW21 CW23 CW26     500.0   113.74 ! Nobs =    1
ANGLe HW24 CW23 HW25     500.0   109.51 ! Nobs =    1
ANGLe HW24 CW23 CW26     500.0   108.40 ! Nobs =    1
ANGLe HW25 CW23 CW26     500.0   108.36 ! Nobs =    1
ANGLe CW23 CW26 HW27     500.0   109.44 ! Nobs =    1
ANGLe CW23 CW26 HW28     500.0   109.43 ! Nobs =    1
ANGLe CW23 CW26 CW29     500.0   109.63 ! Nobs =    1
ANGLe HW27 CW26 HW28     500.0   109.45 ! Nobs =    1
ANGLe HW27 CW26 CW29     500.0   109.44 ! Nobs =    1
ANGLe HW28 CW26 CW29     500.0   109.44 ! Nobs =    1
ANGLe CW26 CW29 HW30     500.0   109.43 ! Nobs =    1
ANGLe CW26 CW29 HW31     500.0   109.43 ! Nobs =    1
ANGLe CW26 CW29 CW32     500.0   109.57 ! Nobs =    1
ANGLe HW30 CW29 HW31     500.0   109.52 ! Nobs =    1
ANGLe HW30 CW29 CW32     500.0   109.45 ! Nobs =    1
ANGLe HW31 CW29 CW32     500.0   109.43 ! Nobs =    1
ANGLe CW29 CW32 HW33     500.0   109.51 ! Nobs =    1
ANGLe CW29 CW32 HW34     500.0   109.48 ! Nobs =    1
ANGLe CW29 CW32 CW35     500.0   109.37 ! Nobs =    1
ANGLe HW33 CW32 HW34     500.0   109.49 ! Nobs =    1
ANGLe HW33 CW32 CW35     500.0   109.48 ! Nobs =    1
ANGLe HW34 CW32 CW35     500.0   109.50 ! Nobs =    1
ANGLe CW32 CW35 HW36     500.0   109.44 ! Nobs =    1
ANGLe CW32 CW35 HW37     500.0   109.44 ! Nobs =    1
ANGLe CW32 CW35 CW38     500.0   109.57 ! Nobs =    1
ANGLe HW36 CW35 HW37     500.0   109.49 ! Nobs =    1
ANGLe HW36 CW35 CW38     500.0   109.46 ! Nobs =    1
ANGLe HW37 CW35 CW38     500.0   109.43 ! Nobs =    1
ANGLe CW35 CW38 HW39     500.0   108.70 ! Nobs =    1
ANGLe CW35 CW38 HW40     500.0   108.71 ! Nobs =    1
ANGLe CW35 CW38 NW41     500.0   112.54 ! Nobs =    1
ANGLe HW39 CW38 HW40     500.0   109.47 ! Nobs =    1
ANGLe HW39 CW38 NW41     500.0   108.68 ! Nobs =    1
ANGLe HW40 CW38 NW41     500.0   108.70 ! Nobs =    1
ANGLe CW38 NW41 HW42     500.0   109.48 ! Nobs =    1
ANGLe CW38 NW41 HW43     500.0   109.50 ! Nobs =    1
ANGLe CW38 NW41 HW44     500.0   109.50 ! Nobs =    1
ANGLe HW42 NW41 HW43     500.0   109.42 ! Nobs =    1
ANGLe HW42 NW41 HW44     500.0   109.45 ! Nobs =    1
ANGLe HW43 NW41 HW44     500.0   109.47 ! Nobs =    1

{ edit if necessary }
! DIHEdral CLW1 CW2  CW3  HW4      750.0 0      0.00 ! Nobs =    1 ... Value =    -0.73
! DIHEdral CLW1 CW2  CW3  CW5      750.0 0    180.00 ! Nobs =    1 ... Value =   179.26
! DIHEdral CW10 CW2  CW3  HW4      750.0 0    180.00 ! Nobs =    1 ... Value =   179.98
! DIHEdral CW10 CW2  CW3  CW5      750.0 0      0.00 ! Nobs =    1 ... Value =    -0.03
! DIHEdral CLW1 CW2  CW10 CW9      750.0 0    180.00 ! Nobs =    1 ... Value =  -179.17
```

```
! DIHEdral CLW1 CW2  CW10 CW11    750.0 0     0.00 ! Nobs =    1 ... Value =    0.27
! DIHEdral CW3  CW2  CW10 CW9     750.0 0     0.00 ! Nobs =    1 ... Value =    0.08
! DIHEdral CW3  CW2  CW10 CW11    750.0 0   180.00 ! Nobs =    1 ... Value =  179.52
! DIHEdral CW2  CW3  CW5  HW6     750.0 0   180.00 ! Nobs =    1 ... Value =  179.97
! DIHEdral CW2  CW3  CW5  CW7     750.0 0     0.00 ! Nobs =    1 ... Value =   -0.02
! DIHEdral HW4  CW3  CW5  HW6     750.0 0     0.00 ! Nobs =    1 ... Value =   -0.04
! DIHEdral HW4  CW3  CW5  CW7     750.0 0   180.00 ! Nobs =    1 ... Value =  179.97
! DIHEdral CW3  CW5  CW7  HW8     750.0 0   180.00 ! Nobs =    1 ... Value = -179.96
! DIHEdral CW3  CW5  CW7  CW9     750.0 0     0.00 ! Nobs =    1 ... Value =    0.02
! DIHEdral HW6  CW5  CW7  HW8     750.0 0     0.00 ! Nobs =    1 ... Value =    0.04
! DIHEdral HW6  CW5  CW7  CW9     750.0 0   180.00 ! Nobs =    1 ... Value = -179.97
! DIHEdral CW5  CW7  CW9  CW10    750.0 0     0.00 ! Nobs =    1 ... Value =    0.03
! DIHEdral CW5  CW7  CW9  CW17    750.0 0   180.00 ! Nobs =    1 ... Value = -179.35
! DIHEdral HW8  CW7  CW9  CW10    750.0 0   180.00 ! Nobs =    1 ... Value = -179.99
! DIHEdral HW8  CW7  CW9  CW17    750.0 0     0.00 ! Nobs =    1 ... Value =    0.63
! DIHEdral CW7  CW9  CW10 CW2     750.0 0     0.00 ! Nobs =    1 ... Value =   -0.08
! DIHEdral CW7  CW9  CW10 CW11    750.0 0   180.00 ! Nobs =    1 ... Value = -179.53
! DIHEdral CW17 CW9  CW10 CW2     750.0 0   180.00 ! Nobs =    1 ... Value =  179.31
! DIHEdral CW17 CW9  CW10 CW11    750.0 0     0.00 ! Nobs =    1 ... Value =   -0.14
! DIHEdral CW7  CW9  CW17 CW15    750.0 0   180.00 ! Nobs =    1 ... Value =  179.84
! DIHEdral CW7  CW9  CW17 SW18    750.0 0     0.00 ! Nobs =    1 ... Value =    0.08
! DIHEdral CW10 CW9  CW17 CW15    750.0 0     0.00 ! Nobs =    1 ... Value =    0.46
! DIHEdral CW10 CW9  CW17 SW18    750.0 0   180.00 ! Nobs =    1 ... Value = -179.30
! DIHEdral CW2  CW10 CW11 HW12    750.0 0     0.00 ! Nobs =    1 ... Value =    0.49
! DIHEdral CW2  CW10 CW11 CW13    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.53
! DIHEdral CW9  CW10 CW11 HW12    750.0 0   180.00 ! Nobs =    1 ... Value =  179.93
! DIHEdral CW9  CW10 CW11 CW13    750.0 0     0.00 ! Nobs =    1 ... Value =   -0.10
! DIHEdral CW10 CW11 CW13 HW14    750.0 0   180.00 ! Nobs =    1 ... Value =  179.98
! DIHEdral CW10 CW11 CW13 CW15    750.0 0     0.00 ! Nobs =    1 ... Value =    0.01
! DIHEdral HW12 CW11 CW13 HW14    750.0 0     0.00 ! Nobs =    1 ... Value =   -0.05
! DIHEdral HW12 CW11 CW13 CW15    750.0 0   180.00 ! Nobs =    1 ... Value =  179.98
! DIHEdral CW11 CW13 CW15 HW16    750.0 0   180.00 ! Nobs =    1 ... Value = -179.67
! DIHEdral CW11 CW13 CW15 CW17    750.0 0     0.00 ! Nobs =    1 ... Value =    0.32
! DIHEdral HW14 CW13 CW15 HW16    750.0 0     0.00 ! Nobs =    1 ... Value =    0.36
! DIHEdral HW14 CW13 CW15 CW17    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.65
! DIHEdral CW13 CW15 CW17 CW9     750.0 0     0.00 ! Nobs =    1 ... Value =   -0.55
! DIHEdral CW13 CW15 CW17 SW18    750.0 0   180.00 ! Nobs =    1 ... Value =  179.21
! DIHEdral HW16 CW15 CW17 CW9     750.0 0   180.00 ! Nobs =    1 ... Value =  179.44
! DIHEdral HW16 CW15 CW17 SW18    750.0 0     0.00 ! Nobs =    1 ... Value =   -0.80
! DIHEdral CW9  CW17 SW18 OW20    750.0 0   -90.00 ! Nobs =    1 ... Value =  -95.64
! DIHEdral CW15 CW17 SW18 OW20    750.0 0    90.00 ! Nobs =    1 ... Value =   84.60
! DIHEdral OW19 SW18 NW21 HW22    750.0 0   -60.00 ! Nobs =    1 ... Value =  -50.74
! DIHEdral SW18 NW21 CW23 HW24    750.0 0    90.00 ! Nobs =    1 ... Value =   86.81
! DIHEdral HW22 NW21 CW23 HW25    750.0 0    90.00 ! Nobs =    1 ... Value =   89.13
! DIHEdral NW21 CW23 CW26 HW27    750.0 0   -60.00 ! Nobs =    1 ... Value =  -56.91
! DIHEdral NW21 CW23 CW26 HW28    750.0 0   180.00 ! Nobs =    1 ... Value =  -176.84
! DIHEdral NW21 CW23 CW26 CW29    750.0 0    60.00 ! Nobs =    1 ... Value =   63.12
! DIHEdral HW24 CW23 CW26 HW27    750.0 0    60.00 ! Nobs =    1 ... Value =   63.69
! DIHEdral HW24 CW23 CW26 HW28    750.0 0   -60.00 ! Nobs =    1 ... Value =  -56.24
! DIHEdral HW24 CW23 CW26 CW29    750.0 0   180.00 ! Nobs =    1 ... Value =  -176.28
! DIHEdral HW25 CW23 CW26 HW27    750.0 0   180.00 ! Nobs =    1 ... Value =  -177.54
! DIHEdral HW25 CW23 CW26 HW28    750.0 0    60.00 ! Nobs =    1 ... Value =   62.53
! DIHEdral HW25 CW23 CW26 CW29    750.0 0   -60.00 ! Nobs =    1 ... Value =  -57.50
! DIHEdral CW23 CW26 CW29 HW30    750.0 0   -60.00 ! Nobs =    1 ... Value =  -59.09
! DIHEdral CW23 CW26 CW29 HW31    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.10
! DIHEdral CW23 CW26 CW29 CW32    750.0 0    60.00 ! Nobs =    1 ... Value =   60.92
! DIHEdral HW27 CW26 CW29 HW30    750.0 0    60.00 ! Nobs =    1 ... Value =   60.94
! DIHEdral HW27 CW26 CW29 HW31    750.0 0   -60.00 ! Nobs =    1 ... Value =  -59.07
! DIHEdral HW27 CW26 CW29 CW32    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.05
! DIHEdral HW28 CW26 CW29 HW30    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.12
! DIHEdral HW28 CW26 CW29 HW31    750.0 0    60.00 ! Nobs =    1 ... Value =   60.87
! DIHEdral HW28 CW26 CW29 CW32    750.0 0   -60.00 ! Nobs =    1 ... Value =  -59.11
! DIHEdral CW26 CW29 CW32 HW33    750.0 0    60.00 ! Nobs =    1 ... Value =   56.06
! DIHEdral CW26 CW29 CW32 HW34    750.0 0   -60.00 ! Nobs =    1 ... Value =  -64.00
! DIHEdral CW26 CW29 CW32 CW35    750.0 0   180.00 ! Nobs =    1 ... Value =  176.03
! DIHEdral HW30 CW29 CW32 HW33    750.0 0   180.00 ! Nobs =    1 ... Value =  176.06
! DIHEdral HW30 CW29 CW32 HW34    750.0 0    60.00 ! Nobs =    1 ... Value =   56.00
! DIHEdral HW30 CW29 CW32 CW35    750.0 0   -60.00 ! Nobs =    1 ... Value =  -63.97
! DIHEdral HW31 CW29 CW32 HW33    750.0 0   -60.00 ! Nobs =    1 ... Value =  -63.93
! DIHEdral HW31 CW29 CW32 HW34    750.0 0   180.00 ! Nobs =    1 ... Value =  176.02
! DIHEdral HW31 CW29 CW32 CW35    750.0 0    60.00 ! Nobs =    1 ... Value =   56.05
! DIHEdral CW29 CW32 CW35 HW36    750.0 0   -60.00 ! Nobs =    1 ... Value =  -58.40
! DIHEdral CW29 CW32 CW35 HW37    750.0 0   180.00 ! Nobs =    1 ... Value =  -178.38
! DIHEdral CW29 CW32 CW35 CW38    750.0 0    60.00 ! Nobs =    1 ... Value =   61.63
! DIHEdral HW33 CW32 CW35 HW36    750.0 0    60.00 ! Nobs =    1 ... Value =   61.59
! DIHEdral HW33 CW32 CW35 HW37    750.0 0   -60.00 ! Nobs =    1 ... Value =  -58.39
! DIHEdral HW33 CW32 CW35 CW38    750.0 0   180.00 ! Nobs =    1 ... Value =  -178.38
! DIHEdral HW34 CW32 CW35 HW36    750.0 0   180.00 ! Nobs =    1 ... Value =  -178.36
! DIHEdral HW34 CW32 CW35 HW37    750.0 0    60.00 ! Nobs =    1 ... Value =   61.65
! DIHEdral HW34 CW32 CW35 CW38    750.0 0   -60.00 ! Nobs =    1 ... Value =  -58.34
! DIHEdral CW32 CW35 CW38 HW39    750.0 0   -60.00 ! Nobs =    1 ... Value =  -56.74
! DIHEdral CW32 CW35 CW38 HW40    750.0 0   180.00 ! Nobs =    1 ... Value =  -175.83
! DIHEdral CW32 CW35 CW38 NW41    750.0 0    60.00 ! Nobs =    1 ... Value =   63.70
! DIHEdral HW36 CW35 CW38 HW39    750.0 0    60.00 ! Nobs =    1 ... Value =   63.29
! DIHEdral HW36 CW35 CW38 HW40    750.0 0   -60.00 ! Nobs =    1 ... Value =  -55.81
! DIHEdral HW36 CW35 CW38 NW41    750.0 0   180.00 ! Nobs =    1 ... Value =  -176.28
! DIHEdral HW37 CW35 CW38 HW39    750.0 0   180.00 ! Nobs =    1 ... Value =  -176.73
! DIHEdral HW37 CW35 CW38 HW40    750.0 0    60.00 ! Nobs =    1 ... Value =   64.18
! DIHEdral HW37 CW35 CW38 NW41    750.0 0   -60.00 ! Nobs =    1 ... Value =  -56.30
! DIHEdral CW35 CW38 NW41 HW42    750.0 0   -60.00 ! Nobs =    1 ... Value =  -59.98
! DIHEdral CW35 CW38 NW41 HW43    750.0 0    60.00 ! Nobs =    1 ... Value =   59.99
! DIHEdral CW35 CW38 NW41 HW44    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.98
! DIHEdral HW39 CW38 NW41 HW42    750.0 0    60.00 ! Nobs =    1 ... Value =   60.47
! DIHEdral HW39 CW38 NW41 HW43    750.0 0   180.00 ! Nobs =    1 ... Value =  -179.56
! DIHEdral HW39 CW38 NW41 HW44    750.0 0   -60.00 ! Nobs =    1 ... Value =  -59.53
! DIHEdral HW40 CW38 NW41 HW42    750.0 0   180.00 ! Nobs =    1 ... Value =   179.54
! DIHEdral HW40 CW38 NW41 HW43    750.0 0   -60.00 ! Nobs =    1 ... Value =  -60.49
! DIHEdral HW40 CW38 NW41 HW44    750.0 0    60.00 ! Nobs =    1 ... Value =   59.54

{ edit if necessary }
IMPRoper CW2  CLW1 CW3  CW10    750.0 0    0.000 ! Nobs =    1 ... Value =  -0.394
```

```
IMPRoper CW3  CW2  HW4  CW5     750.0 0    0.000 ! Nobs =    1 ... Value =   -0.007
IMPRoper CW5  CW3  HW6  CW7     750.0 0    0.000 ! Nobs =    1 ... Value =    0.002
IMPRoper CW7  CW5  HW8  CW9     750.0 0    0.000 ! Nobs =    1 ... Value =   -0.011
IMPRoper CW9  CW7  CW10 CW17    750.0 0    0.000 ! Nobs =    1 ... Value =    0.353
IMPRoper CW10 CW2  CW9  CW11    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.317
IMPRoper CW11 CW10 HW12 CW13    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.015
IMPRoper CW13 CW11 HW14 CW15    750.0 0    0.000 ! Nobs =    1 ... Value =    0.020
IMPRoper CW15 CW13 HW16 CW17    750.0 0    0.000 ! Nobs =    1 ... Value =   -0.005
IMPRoper CW17 CW9  CW15 SW18    750.0 0    0.000 ! Nobs =    1 ... Value =    0.147
! >>> NOTE - unusual value for following improper :    26.17 reset to +35.0
 IMPRoper SW18 CW17 OW19 OW20    750.0 0   35.000 ! Nobs =    1 ... Value =   26.174
 IMPRoper NW21 SW18 HW22 CW23    750.0 0  -35.000 ! Nobs =    1 ... Value =  -34.264
 IMPRoper CW23 NW21 HW24 HW25    750.0 0   35.000 ! Nobs =    1 ... Value =   33.845
 IMPRoper CW26 CW23 HW27 HW28    750.0 0   35.000 ! Nobs =    1 ... Value =   32.962
 IMPRoper CW29 CW26 HW30 HW31    750.0 0   35.000 ! Nobs =    1 ... Value =   32.908
 IMPRoper CW32 CW29 HW33 HW34    750.0 0   35.000 ! Nobs =    1 ... Value =   32.879
 IMPRoper CW35 CW32 HW36 HW37    750.0 0   35.000 ! Nobs =    1 ... Value =   32.912
 IMPRoper CW38 CW35 HW39 HW40    750.0 0   35.000 ! Nobs =    1 ... Value =   33.442
 IMPRoper NW41 CW38 HW42 HW43    750.0 0  -35.000 ! Nobs =    1 ... Value =  -33.143

 { edit if necessary }
 NONBonded CLW1 .2600 3.671       .2600   3.671  ! assuming Chlorine
 NONBonded CW2  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW3  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW4  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW5  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW6  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW7  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW8  0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW9  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW10 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded CW11 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW12 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW13 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW14 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW15 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW16 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW17 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded SW18 0.0430  3.3676    0.0430  3.3676 ! assuming Sulfur
 NONBonded OW19 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded OW20 0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
 NONBonded NW21 0.2384  2.8509    0.2384  2.8509 ! assuming Nitrogen
 NONBonded HW22 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW23 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW24 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HW25 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW26 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW27 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HW28 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW29 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW30 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HW31 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW32 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW33 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HW34 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW35 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW36 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HW37 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded CW38 0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
 NONBonded HW39 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HW40 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded NW41 0.2384  2.8509    0.2384  2.8509 ! assuming Nitrogen
 NONBonded HW42 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HW43 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen
 NONBonded HW44 0.0498  1.4254    0.0498  1.4254 ! assuming Hydrogen

 set echo=true end
```