UNIVERSITY OF CALIFORNIA SAN DIEGO

# Multiscale Computational Models and Analysis to Understand Molecule to Organ Effects of 2'-deoxy-ATP on Cardiovascular Function and Disease

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Bioengineering with a Specialization in Multi-scale Biology

by

Marcus Terrence Hock

Committee in charge:

Professor Andrew D. McCulloch, Chair
Professor James A. McCammon, Co-Chair
Professor Rommie E. Amaro
Professor Kevin R. King
Professor Jeffrey H. Omens

2024

The Dissertation of Marcus Terrence Hock is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

# DEDICATION

This dissertation is dedicated to my family, my parents and siblings, and Miranda. I love you all.

# EPIGRAPH

Do the best you can until you know better.
Then when you know better, do better.

*Maya Angelou*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Andrew McCulloch for his support as my PhD advisor, as well as Professor J. Andrew McCammon as co-chair and co-advisor. They have provided immeasurable expertise and guidance, and have helped me to become my scientist I am today.

I would also like to thank the entire Cardiac Mechanics research group. In particular, thanks to Abby Teitgen, who has been a great friend and worked with me very closely on several of these projects, and provided endless insight and put up with all my questions and random ideas. Additionally, I would like to thank Joe Powers, for his guidance and wisdom throughout the entire PhD process, and being an exceptional mentor and friend. Thank as well to Dr. Regnier and our collaborators at the University of Washington who have provided significant help with validating the computational modeling approaches with experimental measurements.

My family, even being out of the state, has always been there to support me and help me push my scientific career. Their support has been amazing and I would not be where I am today without them. Lastly, I of course I have to thank my girlfriend Miranda, who has been my number one supporter through the entire PhD process. She made the graduate school fun, and helped show me the way first as she defended her PhD two years prior. I don't know where I would be without you.

Chapter 1, in full, is a reprint of the material as it appears in the *Journal of Applied Physics*. Multiscale computational modeling of the effects of 2'-deoxy-ATP on cardiac muscle Ca2+ handling. M. T. Hock*, A. E. Teitgen*, K. J. McCabe*, S. P. Hirakis, G. A. Huber, M. Regnier, R. E. Amaro, J. A. McCammon, and A. D. McCulloch. Multiscale computational modeling of the effects of 2'-deoxy-ATP on cardiac muscle Ca2+ handling. J. Appl. Phys. 134(7):074905, 2023. *Equal contribution. The dissertation author was the co-first author of this publication.

Chapter 2, in part, has been submitted as it may appear in the *Journal of Chemical Information and Modeling* 2024. NetSci: A Library for High Performance Biomolecular

Simulation Network Analysis Computation. A. M. Stokely, L. W. Votapka, M. T. Hock, A. E. Teitgen, J. A. McCammon, A. D. McCulloch, R. E. Amaro. The dissertation author and carried out the biomolecular simulation and analysis of SERCA.

Chapter 3, in part, has been submitted for publication in *PNAS*, 2024. Multiscale modeling shows how 2'-deoxy-ATP rescues ventricular function in heart failure. A. E. Teitgen, M. T. Hock, K. J. McCabe, M. C. Childers, G. A. Huber, B. Marzban, D. A. Beard, J. A. McCammon, M. Regnier, A. D. McCulloch. The dissertation author was the second author and investigator of this paper. The dissertation author carried out the molecular dynamics analysis, Markov State Model construction, Brownian dynamics simulations, and contributed to the scientific process and writing.

## VITA

| | |
|---|---|
| 2019 | Bachelor of Science, Chemical and Biological Engineering, University of Colorado Boulder |
| 2019–2024 | Research Assistant, Bioengineering, University of California San Diego |
| 2020–2024 | Teaching Assistant, Bioengineering, University of California San Diego |
| 2021 | Master of Science, Bioengineering, University of California San Diego |
| 2024 | Doctor of Philosophy, Bioengineering, Specialization in Multi-Scale Biology, University of California San Diego |

## PUBLICATIONS

Randall, E. Benjamin, **Marcus T. Hock**, Rachel Lopez, Bahador Marzban, Collin Marshall, and Daniel A. Beard. "Quantitative Analysis of Mitochondrial ATP Synthesis." *Mathematical Biosciences*, June 17, 2021, 108646.

**Hock, Marcus T.**, Abigail E. Teitgen, Kimberly J. McCabe, Sophia P. Hirakis, Gary A. Huber, Michael Regnier, Rommie E. Amaro, J. Andrew McCammon, and Andrew D. McCulloch. "Multiscale Computational Modeling of the Effects of 2'-Deoxy-ATP on Cardiac Muscle Calcium Handling." *Journal of Applied Physics* 134, no. 7 (August 16, 2023): 074905.

Teitgen, Abigail E., **Marcus T. Hock**, Kimberly J. McCabe, Matthew C. Childers, Gary A. Huber, Bahador Marzban, Daniel A. Beard, J. Andrew McCammon, Michael Regnier, and Andrew D. McCulloch. "Multiscale modeling shows how 2'-deoxy-ATP rescues ventricular function in heart failure" *In review at PNAS*, 2024.

Stokely, A. M., Lane W. Votapka, **Marcus T. Hock**, Abigail E. Teitgen, J. Andrew McCammon, Andrew D. McCulloch, Rommie E. Amaro. "NetSci: A Library for High Performance Biomolecular Simulation Network Analysis Computation" *Submitted to Journal of Chemical Information and Modeling*, 2024.

## FIELDS OF STUDY

Major Field: Bioengineering

Specialization in multi-scale biology

ABSTRACT OF THE DISSERTATION

Multiscale Computational Models and Analysis to Understand Molecule to Organ Effects of 2'-deoxy-ATP on Cardiovascular Function and Disease

by

Marcus Terrence Hock

Doctor of Philosophy in Bioengineering with a Specialization in Multi-scale Biology

University of California San Diego, 2024

Professor Andrew D. McCulloch, Chair
Professor James A. McCammon, Co-Chair

Heart failure is characterized by ventricular weakening, leading to the inability to circulate sufficient blood to the body. Despite clinical advances in recent years, heart failure remains a significant cause of morbidity and mortality within the United States and affects 6.2 million people. Clinical approaches traditionally target pathological symptoms by targeting calcium signaling and neurohormonal responses. Recently, novel therapeutics, known as myosin modulators, have shown promise by targeting the contractile machinery to regulate contractile function. To develop targeted therapeutics, detailed mechanistic models of cardiovascular function are needed which can account for deviations from normal function based on mechanisms of disease

as well as therapeutics. We have developed and integrated a new framework of multiscale models that provide mechanistic insights into the mechanisms of therapeutic molecules for rescue of cardiovascular function. Specifically, we model the effects of deoxy-ATP (dATP), a known myosin activator, on motor function starting at the molecular level of function. Because dATP also has shown experimental improvements in diastolic cardiac function, we modeled and explored the molecular effects of dATP on the SERCA pump, in addition to the effects of dATP on myosin. This work highlights a new framework that captures allosteric SERCA molecular changes that influence calcium sequestration and subsequent cardiac relaxation. Molecular analysis of myosin and dATP demonstrated structural rearrangement in the region of the actin binding surface. We constructed Markov state models to quantify the nature of these changes and helps to reduce the MD simulations to more interpretable changes, which led to observed changes in the actin binding kinetics based on Brownian dynamics simulations. Allosteric changes of SERCA analyzed via generalized correlation analysis led to changes in calcium handling kinetics. The molecular effects, when propagated up in scale to tissue and organ scale, help to demonstrate improvement in cellular and ventricular function, especially in the context of heart failure. This multiscale framework highlights new methods of analysis within the context of dATP and shows promise to guide development of new highly targeted heart failure therapeutics.

# Introduction

## 0.1 Employing computational modeling approaches to solve cardiovascular biological problems

Cardiovascular disease is the leading cause of death in the US, approximated at about one in four deaths [1]. Heart failure specifically affects approximately 6.5 million Americans, of which half exhibit heart failure with reduced ejection fraction in which the heart is unable to contract strongly enough to supply a sufficient oxygen supply to the body [2]. Most treatments prolong patient life by treating the symptoms of heart failure, rather than the underlying mechanisms of actions [3]. As such, there is an urgent need to develop new therapies and treatment approaches which target the underlying mechanisms of the pathophysiology. To develop new therapies in mechanistic approach, in addition to *in vivo*, *in vitro*, and clinical work, detailed computational models are vital as well to understand the full mechanism and nuisances of a particular therapy. Computational models, which can range widely in scale and application, have become an increasingly useful tool for clinicians and well as researchers as computational power and accessibility has expanded in recent years.

The modern era of computing provides an unprecedented opportunity to simultaneously, home desktop computers have become increasing powerful and capable, even on the scale of the smartphone, while high performance computing continues to push the boundaries of computational power and resources [4]. And while Moore's Law, which famously presented the expectation that computing power would double every two years while keeping size, cost and power constant, appears to have finally failed to hold true in recent years, there is by no means any

indication that computational power means will not continue to develop. Future developments will likely arise through new developments in computing architecture and application, rather than just simply packing more transistors into a chip [5, 6]. Both ARM architecture as well as the use of graphical processing units (GPUs) also have provided significant benefits over traditional CPU architectures, by providing more power efficient approaches and computationally efficient instructions and actions, for example for linear algebra and tensor heavy calculations [7, 8].

In parallel with the raw increase in computational potential, machine learning algorithms and approaches have reached widespread application. These machine learning advancements have led to revolutionary new tools in, but not limited to, biological and health sciences and research. These applications range from scenarios that speed up tedious tasks, to providing solutions to seemingly unanswerable questions. For instance, machine learning algorithms have been developed to create 3D meshes from cardiac MRIs, which is a traditionally time consuming and laborious process [9]. Alphafold has provided scientists with the ability to predict any protein structure based purely on the amino acid sequence, with an impressive level of accuracy for most ordered proteins [10, 11]. Though Alphafold is not a replacement for the high-resolution crystal structures, such a tool enables researchers to gain structure based insights on protein mechanism and function, that would otherwise be out of reach. Furthermore, the same algorithms used for Alphafold can be used to predict structures for homology models, using on structure and sequence to predict another slightly different protein, which can provide access to mutated structures, or different species structures.

The ability to make precise modifications and predictions about protein structure directly supports the development of highly personalized medicine for patient treatment [12]. Clinicians and researchers can collect patient genomic data with relative ease due to the accessibility of sequencing techniques and therefore generate large scale population datasets. These patient "atlases" can provide structure to function relationships when combined with tools such as alpha fold. For instance, hundreds of point mutations within the cardiac $\beta$-myosin protein have been associated with divergent forms, hypertrophy and dilated, cardiomyopathies. As such, identifying

a particular protein to target is no longer enough to develop mechanism-based therapies. Instead, the specific pathomechanism of each mutation (or class of mutation) must be elucidated to develop personalized therapies. Ultimately, linking highly detailed structure-function models with systems and biochemical models across a range of scales is necessary to achieve these goals for therapeutic development. Cardiomyopathy severity and prognosis is frequently associated with the structural changes at the organ level that occur as a compensatory mechanism. Therefore, additional patient measurements and data based on imaging and patient history, which are readily available, should also be included in such models which account for and predict growth and remodeling. This paradigm-shift from reductionist models towards detailed multiscale models heralds a new frontier for personalized medicine and improved patient care and outcomes.

## 0.2 Cardiovascular disease and heart failure with reduced ejection fraction

As mentioned in the previous section, heart disease and therapy remains a significant unmet challenge for our healthcare system. Different forms of cardiovascular diseases are estimated to cause $320 billion in annual costs to the healthcare system and lost productive, with projections to triple by 2030 [13, 14]. Among these, heart failure of various forms some extremely poor prognosis, and stems from a range of mechanisms and causes [15]. During systole, the heart contracts to eject and circulate blood and oxygen out to the rest of the body. Systolic heart failure, or heart failure with reduced ejection fraction (HFrEF), occurs when the pumping action fails to circulate sufficient oxygen and blood to the rest of the body due to weakened ventricular function. Clinically, this can be diagnosed by measuring the ejection fraction (EF), which is the volume of blood pumped, relative to the volume of blood in the heart. Normal EFs range from 55%-60%, and mild heart failure with reduced ejection fraction is identified to be an EF from 40-49%, and severe HFrEF is below 40% [16].

Treatments for HFrEF in recent years have improved significantly, prolonging patient life

expectancy by slowing the progression of the disease. Traditionally, a multipronged treatment approach is employed which involves inhibitors of the renin-angiotensin-aldosterone system (RAAS), in combination with $\beta$-blockers, MRAs, and SGLT2 inhibitors. A second, but less commonly employed approach focus on an ionotropic treatments to increase contractile force such as calcium sensitiving agents, Phosphodiesterase inhibitors or even $\beta$-adrenergic receptor agonists. Overall, these approaches represent an upregulated neurohormonal responses; they can help to mitigate dysregulated cardiac function but do not directly target contractile efficiency. Furthermore, these approaches target specific cardiomyocyte pathways to provide a therapeutic benefit, rather than the underlying cause of the heart failure. As such, these treatments can be pro-arrhythmogenic, impair ventricular filling and relaxation, leading to increased heart rate and myocardial oxygen consumption [17], highlighting the need for novel therapeutic approaches. While regenerative medicine approaches based on the application of cell suspension and implantable tissue have appears promising, current trials have generally only demonstrated safety, failing to show significant clinical improvement [18].

Cardiac muscle cells, cardiomyocytes, are a highly organized and structured cell type. Sarcomeres are the contractile unit within the cell and are composed of a highly order lattice structure of overlapping filaments [19]. The myosin motor protein powers this action as the filaments slide past each other during muscle contraction through the cross-bridge cycle [20]. Because the HF is frequently thought of as a disease of the sarcomere, novel therapeutic approaches have been developed to target $\beta$-myosin specifically to either increase or decrease contractility [21, 22]. Clinical trials of some of myosin modulators have demonstrated the efficacy of a directed approach, targeting the underlying mechanisms of action leading to the observed myopathies. A phase III clinical trial for the myosin activator omecamtiv mecarbil (OM) found that in a treatment of heart failure, OM lowered the incidence of composite heart-failure or death compared to the non-treatment group [23]. This primary outcome from the study does not highlight the complications with OM, of which most significant is the reduced filling of the heart, the diastolic function.

Contrary to myosin activators, myosin inhibitors similarly approached treating myopathies by targeting the underlying molecular machinery. Hypertrophic cardiomyopathy (HCM), is characterized by thickening of the ventricular walls, reduced relaxation and filling, leading to reduced cardiac output [24, 25]. Crudely, HCM is often considered to by hyper-contraction of the sarcomere, and therefore myosin inhibitors present an opportunity to reduce the contractile function [26]. Genetic studies have demonstrated that HCM is a familial disease, and mutations within cardiac $\beta$-myosin have been linked to HCM [27, 28]. These studies highlight the connection between malfunction contractile unit and cardiovascular function. Mavacamten, one such myosin inhibitor, progressed through clinical trials and demonstrated significant improvement in patient treatment [29, 30, 31]. Given the positive trials results, and relatively low side effects and additional factors, the FDA granted approval to Mavacamten for treatment of hypertrophic cardiomyopathy. Though mavacamten and OM were two of the earliest myosin modulators to reach the clinic, this new targeted approach has opened to a whole new array of possible therapeutic molecules.

### 0.2.1 2'-deoxy-ATP as a myosin activator a possible therapuetic

2'-deoxy-adenonsine triphosphate (dATP) is another myosin activator that has demonstrated its ability to increase contractile function in cardiomyocytes [32]. dATP, a naturally occurring near analog of ATP, and is formed by the cleavage of the 2' hydroxyl group on the ribose ring of ATP. ATP hydrolysis powers most energy demanding cell processes, muscle cell contraction included [33, 34]. Because dATP contains the same high energy phosphate bonds, its hydrolysis can also be used to provide chemical energy to cell processes. The enzyme Ribonucleotide reductase (R1R2) cleaves the hydroxyl group to convert ATP to dATP [35]. This enzyme is expressed in expressed in cells undergoing cell division because dATP is one of the building blocks for DNA. However, in cardiomyocytes, this protein is not produced and the concentration of dATP is effectively zero in these cells. Interestingly, when the concentration of dATP is increase to 1-2% of the total nucleotide pool, cardiomyocytes demonstrate a significant

increase in force production and cell shortening [36].

To understand the mechanism by which dATP acts as a myosin activator, a range of experimental and computational approaches across have been employed to probe at this question. *In vitro* motility assays demonstrated increased myosin activity as measured by thin filament movement when dATP is present in the solution rather than ATP [37]. These results indicate the direct targeting of dATP on myosin activity, specifically in the cross-bridge cycle. However, intact muscle experiments are necessary to understand how dATP increases contractile function at such low levels relative to ATP. Gene therapy experiments in which R1R2 is overexpressed in cardiomyocytes have demonstrated measured increases in muscle function and overall systolic function [38, 39, 40]. Furthermore, contrary to other myosin activators like OM, dATP does not show a reduction in diastolic function, but rather shows improved muscle relaxing and filling compared to controls. This lusitropric benefit makes dATP a unique candidate to treat HFrEF. The full mechanism by which dATP induces these therapeutic benefits is still yet to be resolved, and therefore additional studies are necessary, in particular to understand the changes to cardiac relaxation.

Computational studies of the effects of dATP on myosin have shown that structural changes are induced by the binding of dATP [41, 42]. Specifically, dATP induces a rearrangement of the actin binding surface that leads greater electrostatic attraction towards the actin thin filament. Increased electrostatics likely directly contribute to increased cross-bridge cycling that is observed in the *in vitro* motility assays as described above. Further molecular dynamics studies have also evaluated the coordinated motion and information flow from the nucleotide binding pocket where ATP or dATP bind to the regulatory domains of myosin [43]. This regulation in and out of the "OFF" state, also known super relaxed state, of the muscle is crucial to both muscle contractile function and metabolic regulation. Important metabolic influences of the super relaxed state have been highlighted in cardiovascular function, and life-long cellular homeostasis [44]. Recent studies have investigated the role of the super relaxed state in skeletal muscle as well. A 2024 study demonstrated that the regulation in and out of the super relax state is important for

thermogenesis during hibernation of small mammals, but found that in large mammals, there is no significant difference in in SRX state, showing the OFF state is not always used to regulate energetics and reduce metabolic demands [45]. Ultimately, these studies show that the super relaxed state provides regulation across a range of muscle types and regulation needs.

### 0.2.2 Additional targets of dATP and heart failure therapeutics

dATP also likely influences other cellular function because it is a near analog of ATP which is widely used across a myriad of cellular processes. Beyond the actomyosin ATP demand, the sarcoendoplasmic ATPase pump (SERCA), Plasma membrane $Ca^{2+}$ ATPase, and sodium potassium ($Na^{+}$/$K^{+}$) pump also have high ATP demands during the active transport of ions [46]. dATP may influence the function of all these pumps, (in addition to other ATP utilizing processes that demand the remaining fractions of ATP). SERCA function is crucial to maintaining cellular calcium homeostasis, and malfunction is heavily implicated in disease. Therefore, SERCA presents another attractive therapeutic target to promote healthy cardiac diastolic function. Phase I/II clinical trials have demonstrated that AAV delivery of SERCA itself is safe to patients with advanced heart failure, though specific clinical endpoint results have been mixed [47, 48, 49]. Despite these outcomes, these studies have paved the way for the future development of safe gene therapies for cardiovascular diseases. The lack of clinical improvement also may be due to the relatively low dose of gene delivery, rather than SERCA being a poorly selected mechanistic target.

## 0.3 Multi scale computational approaches to build more insightful models

In this work, we leverage existing modeling techniques, and combine them with novel approaches and analysis to meet the demands necessary for detailed disease pathology and drug development. Specifically, we will apply computational multiscale biology approaches to understand dATP as a myosin activator. Deoxy-ATP particularly highlights the necessity

for highly detailed models because a single change of two atoms provides, even in 1% of the nucleotides, leads to such pronounced change in muscle behavior. Furthermore, though dATP does show significant promise as a HFrEF therapeutic, it still poses clinical barriers, most notably drug delivery, which likely will require gene therapy approaches, as well as possible metabolic remodeling [39, 50]. Therefore, a detailed mechanistic viewpoint is not only beneficial to move dATP towards the clinic, but to also identify possible therapeutic targets and molecules which may circumvent those limitations specific to dATP, while still providing the same mechanistic advantage.

### 0.3.1 Molecular dynamics simulations

Protein crystallography, which is a robust field that has only grown since its inception with the first crystal structure of myoglobin in 1957 [51], provides a glimpse into the complex machinery of cellular processes. Though some proteins exists in a relatively static nature, all proteins undergo some degree of motion, and have a dynamic behavior, which static crystal structures fail to capture [52]. Molecular dynamics (MD) simulations apply Newton's equations of motion to these static structures and enable the visualization of proteins in dynamic environments [53]. Thanks to recent advances in computational power, as described in the previous sections, we can simulate larger proteins than ever before, and in fact ensembles of proteins, for longer than ever [54]. Because crystallography typically requires some manipulation of the protein environment, MD simulations help to visualize protein structures in their native solvated state, and when simulated for long enough, allow for transitions to be observed [55].

In this work, we use MD simulations as a computational microscope to understand how dATP changes the protein dynamics of the myosin motor protein. We also investigate the influence of dATP on SERCA as it sequesters calcium diastole. Molecular dynamics simulations, though still often too short to see significant conformational or kinetic changes depending on the size of the protein, do still provide a much more detailed picture as to how a protein, such as myosin behave in the presence of a different bound nucleotide, ligand or even mutation. Because

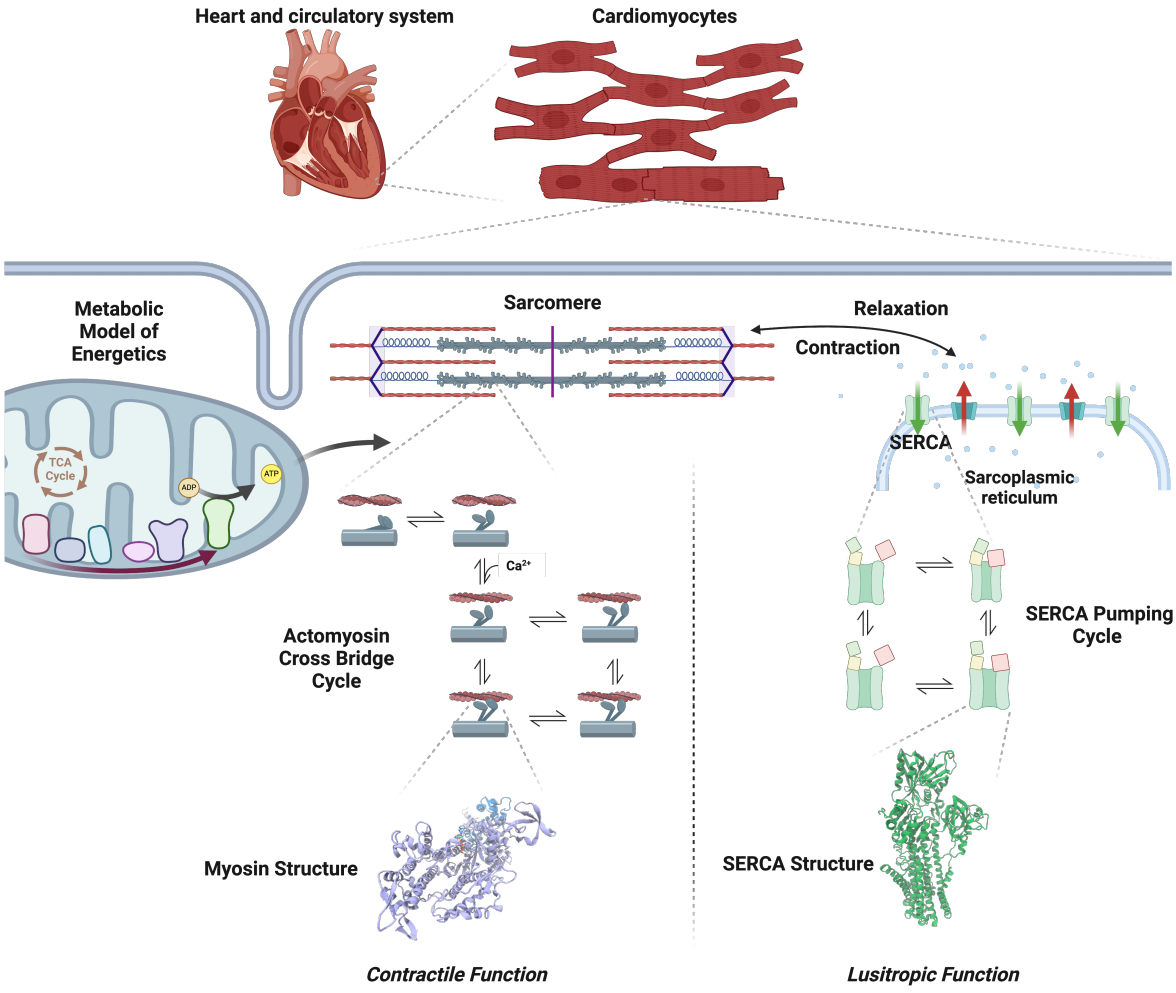


**Figure 0.1.** Multi-scale framework showing the range of systems from molecular levels to organ system model. Specifically highlighting the structure and function of myosin, the major ATPase involved in muscle contraction, and SERCA, another major ATPase involved in calcium handling and relaxation.

of the control by which users can build the simulation, molecular dynamics provide an extremely valuable computational "sandbox" to explore molecular behavior, which we leverage in this dissertation.

However, even with the advances in computational power, the conclusions drawn can often be obscured by the relatively limited simulation durations [56]. For instance, observing the full transition of the myosin protein from the post-powerstroke to pre-powerstroke transition is outside the realm of traditional molecular dynamics simulations. Enhanced sampling and steered molecular dynamics techniques modify the energy surface of the simulation in order to overcome large free energy barries in an attempt to reach a desired state [57]. Gaussian accelerated molecular dynamics changes the energy landscape based on a Gaussian distribution which enables reweighting of the simulation to predict how an unbiased simulation would behave. In doing so, the sampling of the simulation increases in a manner comparable to increasing the simulation duration by an order of magnitude. Despite enhanced sampling methods like these, the challenge of insightful interpretation of MD simulation remains.

**Generalized Correlation Analysis**

The flow of information through a protein is crucial to its function. This information flow is ultimately driven by thermodynamics, or changes in thermodymics based on a catalyst, co-factor or environmental change. However, as mentioned in the section above, because of the sampling limitations of molecular dynamics simulations, we often fail to observe the desired event or transition. Fundamentally, protein function reduces to the coordinated motion of residues relative to one another. Initial attempts to quantify the relationship between residues relied on linear measurements such as the Pearson correlation [58]. A generalized correlation coefficient based on the mutual information of protein motion was proposed by Grubmüller and Lange which has been widely adapted [59]. Generalized correlated (GC) analysis highlights key residue-residue information that obscure via visual or traditional analysis method. The insights provided by GC analysis can help to predict drugable domains of a protein, or estimate the severity of an

identified mutation. In the following work, we apply this approach to understand the allosteric communication within SERCA based on the binding of ATP or dATP to the nucleotide binding site. Further, we use a newly developed software tool, Netsci, which leverages CUDA GPU acceleration in its algorithm which significantly reduces the computational time and increases the accessibility of generalzied correlation analysis [60].

**Markov State Models**

The construction of Markov State models (MSMs) provide a means to reduce the complexity of a molecular dynamics simulation (or frequently several simulations) into a reduced set of kinetically relevant structures [61, 62]. Simply put, a Markov state model is composed of a finite number of discrete states, between which, the protein can transition between based on some estimated probability, though of course MSMs may be employed outside the scope of protein simulation. By definition, the different states of the Markov model are memoryless, meaning that the probability of transition to the next state of the protein system does not depend on its previous state [63]. Therefore, to estimate the transition probability between states, one can simply count the frequency of transitions to adjacent states. Though the transitions can be estimated trivially, identifying the model states is much more nuanced. This selection of states can be broken down into the following steps: featurization, dimensionality reduction, and clustering [64]. After these steps the MSM can be constructed and analyzed. Feature selection requires expert knowledge of the simulation system, and poorly selected features lead to either MSMs that fail to converge, do not behave in a Markovian manner, or do not provide any kinetic insights. Yet, properly constructed models can provide significant insights and help to understand molecular systems in new ways. In this work, we successfully construct a validated Markov model that characterizes the behavior of pre-powerstroke myosin. Specifically, we compare how the dynamics and transition kinetics differ depending on whether ATP is bound to myosin, or dATP.

### 0.3.2 Brownian Dynamics Simulations

Even rapid cellular processes such as cross-bridge cycling between actin and myosin occur over timescales, and length scales, outside of the scope of MD simulation capabilities, which makes predicting drug effects on biological systems so difficult. Yet, estimated changes to kinetics in protein protein binding, based on structural and molecular information is quintessential to mechanistic and targeted drug design. Brownian dynamics (BD) simulations provide a means to compute the rates of protein-protein binding kinetics, but at the expense of reduced structural detail [65, 66]. Brownian dynamics generally use an implicit water model, contrary to MD simulations the explicitly model water molecules, and treat the protein structures as static [67]. Rather than focuses on the interatomic forces from chemical bonds, Brownian dynamics simulations predict protein-protein association based on simplified long range electrostatics, hydrodynamics, and the random motion of a Brownian walk [68]. Traditionally, tens of thousands up to millions of independent BD simulations are carried out to construct and estimate of the second order association rate between molecules. Based on the frequency of successful binding events to unsuccessful, the rate constant can be estimated. To overcome the limitation of rigid simulations structures, in our work, we apply an ensemble based BD simulation approach. Ensemble docking provides a balanced approach of conformational detail without requiring excessive computational resources. Specifically, we use different molecular structures based on the influence of different bound nucleotides (ATP, dATP) or nucleotide free structures to understand the molecular influence of dATP. When using BD simulations to characterize the behavior of actomyosin binding during the attachment step of the cross-bridge cycle, we use distinct conformations identified from MSM analysis. In other BD approaches this this thesis work, we use an ensemble BD approach based on conformations identified via traditional clustering algorithms to predict the kinetics of nucleotide and calcium binding to SERCA.

### 0.3.3 Muscle filament and cellular modeling

Since the original sliding filament theory was put forth by two independent papers [69, 70] and subsequent two and three state models of muscle contraction were postulated, newer and more complex models have continued to blossom. Present day models of the sarcomere, for instance, are spatially explicit, account for cooperative effects of calcium binding, account for metabolite concentrations and even incorporate electrical activation of the calcium release [71, 72, 73]. While the complexity and accuracy of these models are greatly improved in recent years, the tunability of individual model parameters is most crucial. For instance, measured changes in Brownian dynamics 2nd order kinetics can be modified in a model of sarcomere contractility. Alternatively, and estimated change in the free energy of a particular reaction be used to predict a new equilibrium for binding events [74]. We account for changes in kinetics to the cross-bridge cycle, on and "off" state, and calcium handling of SERCA to modify existing models of cardiovascular mechanics in this work. Furthermore, we combine two scales, a Monte Carlo cross-bridge model and differential equation cell model, of muscle activation in order to highlight the increased importance of cooperative muscle activation due the presence of dATP.

### 0.3.4 Whole heart and circulatory modeling

While clinical significance can be observed at each of the previous scales analyses discussed previously, we maintain the perspective that modeling whole heart function and cardiovascular output is essential to truly understand the effects of molecular level therapeutics such as dATP. Beyond the context of dATP, these final scale is essential for understanding possible mechanisms and applications of additional drugs [75]. In this work, we use a relatively simple geometry of the heart coupled to a circulatory model based on restive flow [76, 77]. Though a finite element model may provide greater insight into heterogeneous activation of myosin activators, and how this influences function, it is outside the scope of this work. Additionally, a finite element model would have significantly greater cost. Most importantly, our current organ

and circulation model still provides additional context that helps to resolve differences between predicted effects and existing mechanistic hypotheses with *in vivo* models of cardiovascular function. Ultimately, the models at each scale are only as useful as identifying what they fail to explain.

# Chapter 1

# Multiscale Computational Modeling of the Effects of 2'-deoxy-ATP on Cardiac Muscle Calcium Handling

## 1.1 Abstract

2'-deoxy-ATP (dATP), a naturally occurring near analog of ATP, is a well-documented myosin activator that has been shown to increase contractile force, improve pump function, and enhance lusitropy in the heart. Calcium transients in cardiomyocytes with elevated levels of dATP show faster calcium decay compared with cardiomyocytes with basal levels of dATP, but the mechanisms behind this are unknown. Here we design and utilize a multiscale computational modeling framework to test the hypothesis that dATP acts on the Sarcoendoplasmic Reticulum Calcium-ATPase (SERCA) pump to accelerate calcium re-uptake into the sarcoplasmic reticulum during cardiac relaxation. Gaussian accelerated Molecular Dynamics simulations of human cardiac SERCA2A in the E1 *apo*, ATP-bound and dATP-bound states showed that dATP forms more stable contacts in the nucleotide binding pocket of SERCA and leads to increased closure of cytosolic domains. These structural changes ultimately lead to changes in calcium binding, which we assessed using Brownian Dynamics simulations. We found that dATP increases calcium association rate constants to SERCA and that dATP binds to *apo* SERCA more rapidly than ATP. Using a compartmental ordinary differential equation model of human cardiomyocyte excitation-

contraction coupling, we found that these increased association rate constants contributed to the accelerated rates of calcium transient decay observed experimentally. This study provides clear mechanistic evidence of enhancements in cardiac SERCA2A pump function due to interactions with dATP.

## 1.2 Introduction

The Sarcoendoplasmic Reticulum Calcium-ATPase (SERCA) 2A is a P-Type ATPase [78, 79] which is critical for sequestration of calcium into the Sarcoplasmic Reticulum (SR) during cardiac relaxation and is the dominant SERCA isoform in cardiac muscle [79]. SERCA is a transmembrane protein embedded in the SR lipid membrane which consists of 3 cytosolic domains (Nucleotide binding domain - "N", Phosphorylation domain - "P", Actuator domain - "A") as well as 10 transmembrane (M) helices, M1 through M10 (Fig. 1A) [80, 81]. The nucleotide binding region is located within the N domain. Calcium binding occurs in the transmembrane region between helices M4, M5, M6, and M8, at binding locations known as Site I and Site II (Fig. 1C) [82, 81]. Generally, SERCA transitions between two major states as it pumps calcium into the SR in an ATP-driven manner: E1 and E2. In the first state, E1, the the calcium binding sites face the cytosolic side of the membrane. Binding of ATP and two calcium ions, followed by ATP dephosphorylation and hydrolysis, reconform the protein so that calcium can be released into the SR lumen [83, 81]. The SR-facing conformation is known as E2. Release of ADP, phosphate, and calcium ions into the SR lumen allow the protein to move back into the E1 state [84, 81]. A simplified ordinary differential equation model of SERCA function developed by Tran *et al.* [85] describes several rate-limiting steps within the cycle: (1) $Mg^{2+}$ATP binding, (2) binding of the first calcium ion to site I (site II binding is then considered to occur instantaneously), (3) ADP release, coupled with the E1-E2 transition, (4) Release of calcium ions into the SR lumen, and (5) Pi release, coupled with E2-E1 transition. Here we focus on how the first two of these rate-limiting steps may be affected by a molecular modification to

ATP.

2'-deoxy-ATP (dATP) is a naturally occurring nucleotide that differs in structure to ATP by only a cleaved hydroxyl group on the 2' carbon of the ribose ring (Fig. 1B). In spite of the similar structures of the two molecules, force production has been found to increase dramatically in cardiac cells when ATP is replaced by even a small fraction of dATP [86, 87]. This finding has led to various studies exploring the potential for dATP as a therapeutic myosin activator to treat heart failure [88, 39, 36, 50, 40, 89]. In one such study, upregulation of dATP to $\sim 1\%$ of the overall ATP pool led to increased cell shortening, but also had a marked and beneficial effect on the calcium transient leading to improved lisutropy [38]. These results showed decreased time to 50% and 90% calcium transient decay, suggesting that the rate of intracellular calcium reduction is enhanced during dATP treatment. This may confer additional therapeutic benefit in heart failure which is commonly characterized by decreased function and expression of SERCA2A in failing cardiac myocytes [90, 91, 92]. SERCA is also under investigation as a therapeutic target [90, 93, 92, 94, 95, 49]. To reveal potential therapeutic mechanisms of dATP on SERCA2A, analysis at multiple scales from atomic resolution molecular dynamics to whole cell function is required.

We developed a multiscale modeling approach spanning molecular to whole cell scales. Gaussian accelerated Molecular Dynamics (GaMD) simulations of human cardiac E1 SERCA2A embedded in a lipid bilayer [96, 97, 98, 99] were conducted on 3 separate systems: *apo*, ATP-bound, and dATP-bound. Analysis of the GaMD trajectories allowed us to locate key sites in the cytosolic and transmembrane domains of SERCA that may be modified by dATP binding. Rigid body Brownian Dynamics (BD) simulations [68] were then used to measure association rate constants of ATP, dATP, and calcium ions to SERCA2A. We found that dATP bound to SERCA2A with greater affinity than ATP, and calcium bound with a higher affinity to dATP-bound than ATP-bound SERCA. Finally, in a compartmental ordinary differential equation model of whole cell calcium handling [100], the effects of these molecular differences on the calcium transient were predicted.

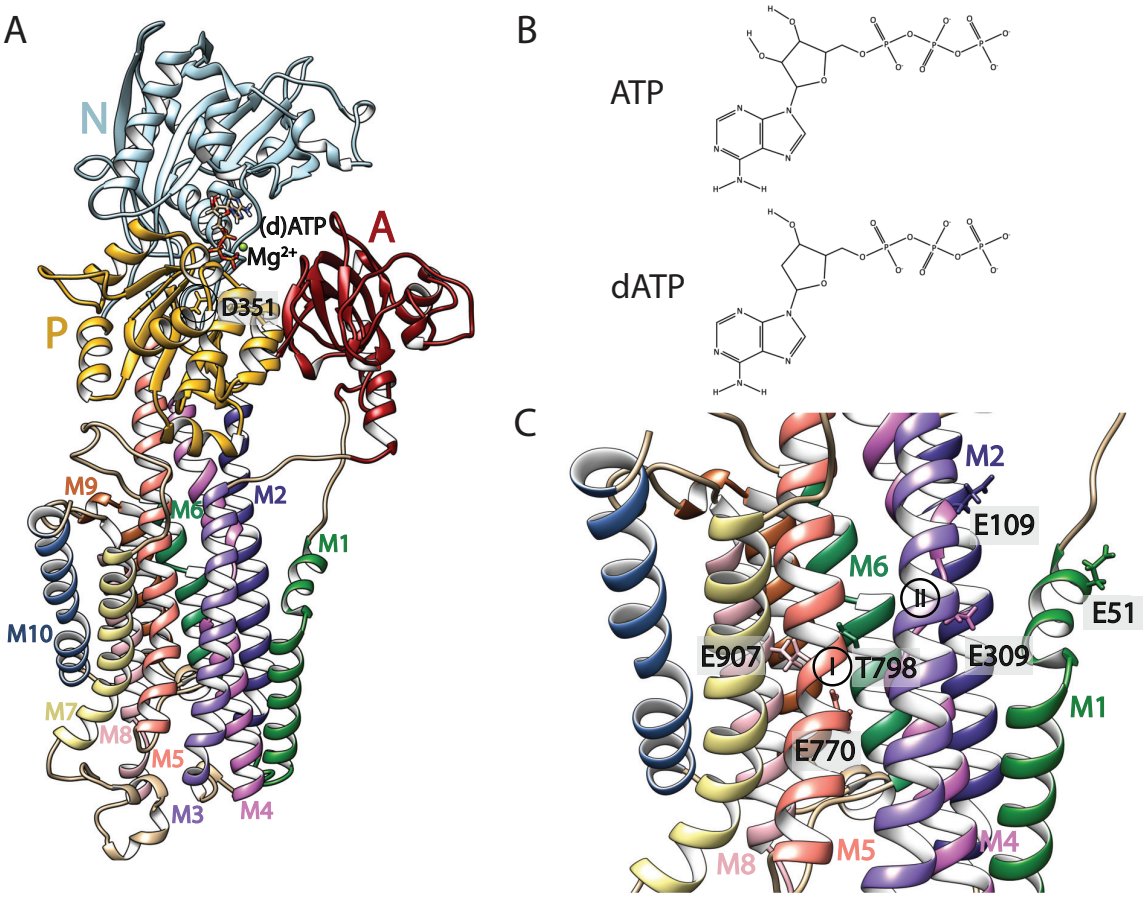


**Figure 1.1.** (A) Overview of SERCA structure. A, N, and P cytosolic domains, as well as M1-M10 transmembrane helices are labeled. D351, the nucleotide phosphorylation site, is also labeled. Nucleotide and $Mg^{2+}$ are shown in the binding site on the N domain. (B) ATP and dATP chemical structures. Note missing hydroxyl group on the ribose ring for dATP. (C) Calcium binding site I and site II, and key residues used for assessing calcium binding path dynamics including E907 on M8, E770 on M5, and T798 on M6 (site I), E309 on M4 (site II), and E109 (M2) and E51 (M1) which comprise part of the calcium entry path.

## 1.3 Methods

### 1.3.1 Gaussian Accelerated Molecular Dynamics

We began with a crystal structure of human cardiac SERCA2A in the E2 state from the Protein Data Bank (PDB) (PDB ID: 7BT2) [101, 99]. Since no human cardiac SERCA2A structures in the E1 state were available, we constructed a homology model using a SERCA 1A crystal structure in the E1 state purified from rabbit fast-twitch skeletal muscle (PDB ID: 3W5A) [98]. Homology modeling was done using SWISS-MODEL [102]. Sequence identity was 83.7%, with 93.1% similarity, determined using Clustal Omega, [103] and the GMQE score for the model was 0.77. Chain A from the 3W5A crystal structure was used for homology modeling. Ligands $K^+$, MPD, and PCW were removed from the 7BT2 structure, and Na+, $Mg^{2+}$, M1, PTY, and sarcolipin were removed from the 3W5A structure. Nucleotide and $Mg^{2+}$ position within the nucleotide binding pocket were taken directly from the 7BT2 structure. CHARMM-GUI was used to prepare all simulation files [104]. Protonation states were determined using the PDB2PQR PropKa tool, at a pH of 7.0 [105, 106, 107, 108]. Protonated residues were: HSE (683), and HSD (5, 32, 38, 190, 278, 284, 868, 872, 880, 882, 944). A disulfide bond was added between CYS 875 and CYS 887. The protein was embedded in a 12.5 by 12.5 nm lipid bilayer, and the position of SERCA within the bilayer was determined using the Orientations of Proteins in Membranes (OPM) database [109]. The lipid bilayer was composed of POPC (51% upper leaflet, 66% lower leaflet), POPE (43% upper leaflet, 17% lower leaflet), and POPS (6% upper leaflet, 17% lower leaflet), with 308 total lipids in the upper leaflet and 297 total lipids in the lower leaflet, based on experimentally determined membrane composition in the cardiac SR [110]. A rectangular water box of thickness 22.5 nm was added, and period boundary conditions were utilized for simulation. $K^+$ and $Cl^-$ ions were added using the Monte Carlo placement method at a 150 mM concentration to neutralize the system [111]. AMBER input files were generated using CHARMM-GUI[112, 104].

To construct the dATP structure, the extra hydroxyl group was removed from the ATP

structure, taken directly from the 7BT2 crystal structure, using Chimera [113]. These ATP and dATP structures were then input into CHARMM-GUI, and antechamber was used to generate force field parameter files using the AMBER GAFF2 force field [114, 115]. The AMBER FF19SB [116] force field was used for protein residues, and the Lipid17 and OPC [117] force fields were used for lipid and water molecules, respectively. The SHAKE algorithm was used to constrain the motion of hydrogen-containing bonds [118].

MD and GaMD simulations were performed using Amber20 [119], and all simulations were run on the Triton Shared Computing Cluster through the San Diego Supercomputer Center [120]. A total of 150 ns of conventional MD and 200 ns of GaMD (3 replicates) were performed for *apo*-SERCA, and ATP- and dATP-bound SERCA. Prior to these simulations, minimization was done over 5000 steps of steepest decent minimization with 10 kcal $mol^{-1}$ $Å^{-2}$ positional restraints on all protein atoms and 2.5 kcal $mol^{-1}$ $Å^{-2}$ positional restraints on all lipid atoms, with NMR restraints. Equilibration was done over 6 steps, for 1.875 ns total. The Langevin temperature equilibration scheme using a collision frequency of 1.0 $ps^{-1}$ was utilized to set the system temperature to 303.15 K using the NVT ensemble over 2 steps (125 ps each). During these heating steps, 10 kcal $mol^{-1}$ $Å^{-2}$ positional restraints were present on all protein atoms for the first step and 5 kcal $mol^{-1}$ $Å^{-2}$ positional restraints were present on all protein atoms for the second step, and 2.5 kcal $mol^{-1}$ $Å^{-2}$ positional restraints with NMR restraints were present on all lipid atoms for both steps. The system was then equilibrated over 4 stages using the semi-isotropic (with constant surface tension) NPT ensemble (constant number of particles, pressure, and temperature), for 125 ps, 500 ps, 500 ps, and 500 ps, respectively, with the system set to 1.0 bar. Positional restraints on all protein atoms were 2.5, 1.0, 0.5, and 0.1 kcal $mol^{-1}$ $Å^{-2}$ for each step, respectively. Positional and restraints on all lipid atoms were 1.0, 0.5, 0.1, and 0 kcal $mol^{-1}$ $Å^{-2}$ for each step, respectively, with NMR restraints.

MD and GaMD simulations were run at 303.15 K using the PMEMD (Particle Mesh Ewald Molecular Dynamics) method with a 9 Å nonbonded cutoff, and 2 fs timestep. Coordinates were saved every 100 ps for MD and 20 ps for GaMD simulations. In GaMD, a Gaussian

distribution is used to provide a boost potential for the system in order to enhance conformational sampling at shorter simulation time scales [96]. A single boost potential was applied to the total potential energy only. The final frame from the 150 ns MD simulations was used as the starting point for the GaMD simulations, and 0.4 ns of conventional MD prep, 2.4 ns of conventional MD (to calculate potential statistics), 0.4 ns of GaMD pre-equilibration (with boost potential applied), and 10.4 ns of GaMD equilibration (with boost potential applied and boost parameters updated) were run before all GaMD production simulations. The three GaMD replicates were averaged for analysis. Chimera, VMD, and PyMol were used for trajectory analysis and visualization [113, 121, 122]. Analysis was performed using the AMBER cpptraj [123] and the MDTraj python libraries [124]. Hydrogen bonding analysis used a 3 Å and 135° cutoff. Energetic reweighting of trajectory data was performed after all simulations using a Gaussian approximation of cumulant expansion to the second order. This step is necessary because a boost potential was applied at each time step in order to flatten the energy landscape during the simulation and increase conformational sampling [96]. Briefly, the potential for mean force (PMF) as a function of reaction coordinate $A_j$ is calculated as:

$$PMF(A_j) = -\frac{1}{\beta} \ln p(A_j) \tag{1.1}$$

Where $\beta = k_B T$ and $p(A_j)$ is the canonical ensemble distribution. Because boost potentials followed a Gaussian distribution, $p(A_j)$ must be calculated from the ensemble distribution of the boosted data set as:

$$p(A_j) = p^*(A_j) \frac{\left\langle e^{\beta \Delta V(r)} \right\rangle_j}{\sum_{j=1}^{M} \left\langle e^{\beta \Delta V(r)} \right\rangle_j} \ln p(A_j), J = 1, ..., M \tag{1.2}$$

Where $\Delta V(r)$ is the boost potential for each frame, M is the number of bins, and $\left\langle e^{\beta \Delta V(r)} \right\rangle_j$ is the ensemble averaged Boltzmann factor for frames in bin $j$. $\left\langle e^{\beta \Delta V(r)} \right\rangle$ is approximated using second

order cumulant expansion and is calculated as:

$$\langle e^{\beta \Delta V(r)} \rangle = exp\{\sum_{k=1}^{\infty} \frac{\beta^k}{k!} \sigma_{\Delta V}^2\} \tag{1.3}$$

Re-weighting of the GaMD simulations was done using scripts and protocols developed by Miao *et al.* [125].

### 1.3.2 Brownian Dynamics

BD simulations with Browndye 2 [68] were used to probe SERCA binding kinetics. Browndye treats molecules as rigid cores and uses an adaptive time step to efficiently simulate binding kinetics. We first carried out BD simulations of ATP and dATP binding to the *apo* SERCA structures derived from our GaMD simulations and compared their respective association rate constants. In order to better capture the conformational dynamics and variability of SERCA, 30 SERCA structures were used independently in 30 separate simulations for ATP and dATP. The 30 conformations were obtained through hierarchical clustering of the *apo* GaMD SERCA simulations, using $C_\alpha$ RMS as the cutoff metric. The lipid bilayer was removed for nucleotide binding simulations to reduce simulation compute cost, given that the bilayer should not substantially affect the relative rates of binding of ATP and dATP. Browndye uses "reaction pairs" as a reaction coordinate to measure progress of binding events. These pairs were defined based on the starting homology model with ATP or dATP present. Pairs were defined by contacts between (d)ATP and residues PHE 487 and ARG 559 with a distance less than 3.5 Å. The full list of binding pairs can be found in Table 1.1, and is further illustrated in Fig. 1.2 in the supplementary material. The encounter complex description in Browndye, which specifies the distance between pairs necessary for a reaction to be considered complete, was left unspecified, allowing for a range of binding probabilities to be observed as a function of reaction distance. For each of the 30 representative structures, BD simulations were carried out to measure the association of ATP, and separately with dATP, with 50,000 individual trajectories simulated per

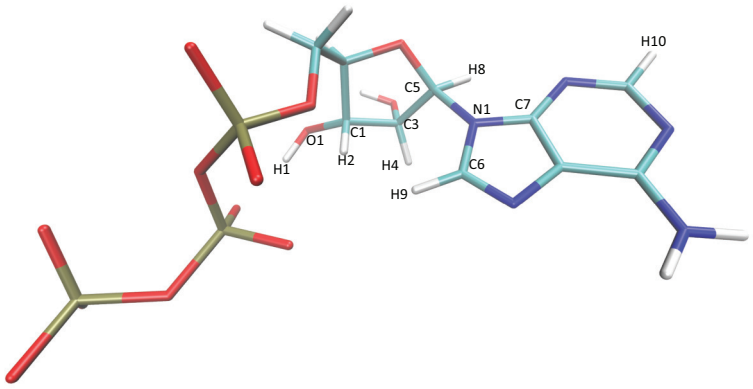


**Figure 1.2.** Labeled diagram for the atom names of ATP used to define Browndye reaction pairs. The same naming scheme is used for dATP, and only atoms conserved in both structures are used in the reaction pair criteria.

conformation. Overall, 1.5 million total ATP trajectories were simulated as well as 1.5 million dATP trajectories. The same AMBER force field used for the GaMD simulations was used to parameterize the protein and nucleotide charges and radii for BD simulations. Binding rate constant curves of the resulting simulations were then averaged based on the cluster weight of each observed representative structure. [h!]

We again employed BD simulations to investigate changes in calcium binding to site I and II in SERCA when ATP or dATP is bound. The starting SERCA structures were clustered from the GaMD ATP and dATP simulations. Thirty representative structures were extracted from each nucleotide condition, with the membrane intact. The same hierarchical clustering approach based on the $C_\alpha$ RMS was again applied to generate representative structures. The membrane was included to ensure that calcium enters the SERCA protein through a realistic entry point, likely via the M1/M2/M4 path, and not through the transmembrane region buried in the lipid bilayer [81]. Reaction pairs for calcium to site I were defined based on atoms forming hydrogen bonds in a previously solved crystal structure [126]. A second set of reaction pairs was also established from the same crystal structure for site II (full list for site I and II can be found in Table 1.2 in the supplementary material). For site I, the 30 structures from the ATP conditions were used to run 50,000 BD trajectories, leading to 1.5 million trajectories. Similarly,

**Table 1.1.** Encounter complex description used for Browndye simulations of (d)ATP binding to SERCA

| SERCA Binding Pairs | |
|---|---|
| Residue-Atom Name | ATP or dATP Atom |
| PHE 486-CB | H10 |
| PHE 486-CG | H10 |
| PHE 486-HB2 | H10 |
| PHE 486-CD2 | H10 |
| ARG 558-NE | C3 |
| ARG 558-NE | H4 |
| ARG 558-HE | C1 |
| ARG 558-HE | O1 |
| ARG 558-HE | H1 |
| ARG 558-HE | H2 |
| ARG 558-HE | C3 |
| ARG 558-HE | H4 |
| ARG 558-HE | C5 |
| ARG 558-CZ | C3 |
| ARG 558-HD2 | O1 |
| ARG 558-HD2 | H1 |
| ARG 558-NH2 | C3 |
| ARG 558-NH2 | H4 |
| ARG 558-HH21 | C3 |
| ARG 558-HH21 | H4 |
| ARG 558-HH21 | C5 |
| ARG 558-HH21 | H8 |
| ARG 558-HH21 | N1 |
| ARG 558-HH21 | C6 |
| ARG 558-HH21 | H9 |
| ARG 558-HH21 | C7 |

the 30 structures from the dATP GaMD simulations were used to run an additional 50,000 BD trajectories per conformation, again leading to 1.5 million trajectories simulated. The binding rate constant curves were averaged based on the frequency of the cluster throughout the GaMD simulations. The same procedure was repeated using the same structures to investigate binding to site II, with the reaction pairs adjusted accordingly for an additional 1.5 million trajectories per nucleotide condition. In total, between the two calcium sites and two nucleotides bound, 6 million calcium binding BD trajectories were simulated.

**Table 1.2.** Encounter complex description used for Browndye simulations of calcium associating to (d)ATP-bound SERCA

| Atom Calcium Interaction Pairs | |
|---|---|
| Site I | Site II |
| ASN 767-OD1 | ASN 795-OD1 |
| GLU 770-OE1 | GLU 309-OE1 |
| GLU 770-OE2 | GLU 309-OE2 |
| GLU 907-OE1 | |
| GLU 907-OE2 | |
| THR 798-OG1 | |

All BD simulations for both the nucleotide and calcium association simulations used 150 mM ionic strength, and a desolvation parameter of 0.025 based on sensitivity analysis to allow for realistic binding distance criteria. Each BD trajectory had a maximum of 1000000 steps. The dielectric coefficient of the solvent was set to 78, while the solute dielectric coefficient was set to 4.

### 1.3.3 Calcium Transient Modeling

To assess how changes in nucleotide and calcium association rate constants to SERCA impact the myocyte calcium transient as a whole, we utilized a whole cell excitation-contraction coupling (ECC) model developed by Himeno *et al.* [100]. This model was chosen because it explicitly includes a three-state model of SERCA [85] which has parameters for ATP binding, calcium binding, and E1-E2 transition. In this model, state $P_1$ is E1 SERCA, which undergoes a

reversible reaction dictated by rates $K_1^+$ and $K_1^-$ to state $P_{2-5}$ (E1 with ATP and 2 calcium ions bound). Calcium binding is considered to be fully cooperative, i.e. binding of the second calcium ion is instantaneous after the first calcium binding event. Reaction rates $K_2^+$ and $K_2^-$ encompass MgADP dissociation as well as the E1-E2 transition, leading the model to state $P_{6-10}$. Finally, reversible reaction rates $K_3^+$ and $K_3^-$ return the pump to state $P_1$ [85]. dATP experimental data were digitized from Korte *et al*. Fig. 1B (GFP and R1R2, respectively) [38]. Fura ratio units were converted to calcium concentration by setting the maximum fluorescence value to 0.45 $\mu$M, and the minimum value to 0.05 $\mu$M, since these are approximately the maximum and minimum calcium values typically seen in the ECC model [100]. The ECC model was optimized to match the ATP experimental calcium transient by varying $Amp_{SERCA}$, $Amp_{NCX}$, $Amp_{NaK}$, and $f_n$, the same parameters which were tuned by Himeno *et al*. in parameterizing their original model [100]. Optimization was conducted using Particle Swarm Optimization in MATLAB [127]. The timescale of the applied current in the ECC model was adjusted to reflect differences between human and rat and to more closely fit control (ATP) calcium transients.

## 1.4 Results

### 1.4.1 dATP is more stable in the nucleotide binding pocket, facilitating E1-ATP to E1-ADP transition via enhanced phosphorylation and movement of cytosolic domains

We first assessed differential interactions of ATP and dATP in the nucleotide binding pocket in the N domain. Computing the overall number of contacts between the nucleotide and residues identified to come within 3 Å for at least one frame of any of the simulations, we found that dATP had a greater number of contacts overall (Fig. 2A). The average number of contacts was 8.7 for ATP and 10.5 for dATP. Further, we found that dATP had a lower RMSD overall (averaged across all 3 GaMD simulations) than ATP (Fig. 2B). This suggests that dATP is interacting more closely with a greater number of residues in the N domain binding pocket, and that it binds more stably, leading to less movement within the binding pocket.

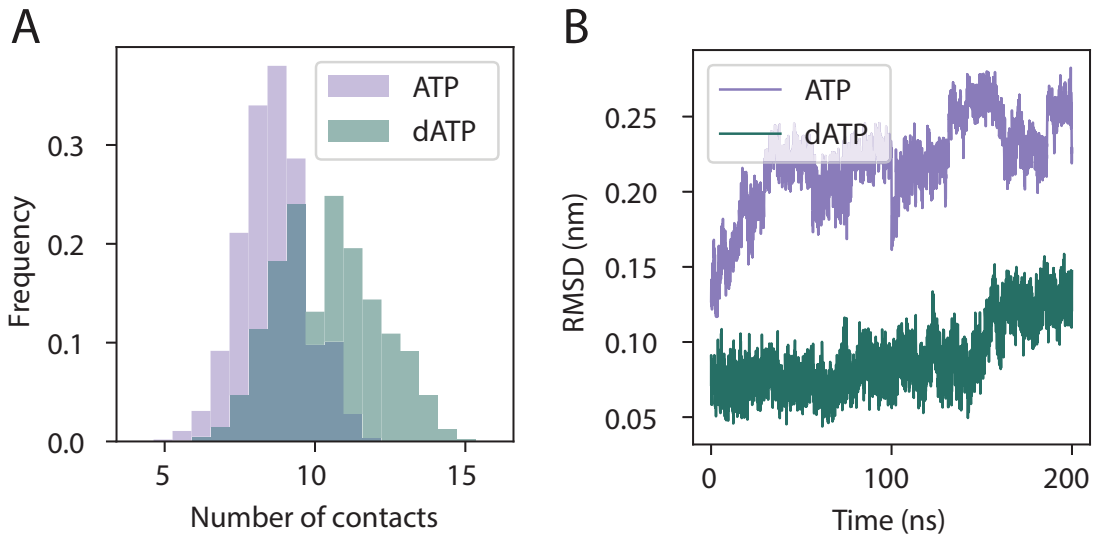


**Figure 1.3.** dATP is more stable in the nucleotide binding pocket. (A) The overall number of contacts in the binding pocket (averaged across 3 replicates) was greater for dATP than for ATP. Only residues identified to come within 3 Å for at least one frame of the simulation were included. Data are displayed as histograms showing the frequency distribution across the GaMD simulations. (B) The RMSD for dATP (averaged across 3 replicates) was lower than for ATP.

We next assessed specific residue interactions with ATP and dATP in the nucleotide binding pocket. The distances across three GaMD replicates between the nucleotide and several residues of interest are shown in Fig. 3A. Distances to known nucleotide interaction residues PHE 487, ARG 559, and LYS 514 [81, 128, 129, 130] were unchanged (the average distances were 0.39 nm, 0.30 nm, and 0.47 nm for ATP and 0.39 nm, 0.26 nm, and 0.45 nm for dATP, respectively). This suggests that interactions with these residues do not explain differences in nucleotide association. However, we found that ATP came in closer contact with several residues towards the top of the nucleotide binding pocket than dATP, including THR 441, LYS 492, ARG 677, and ARG 489 (average distances were 0.50 nm, 0.33 nm, 0.36 nm, and 0.43 nm for ATP and 0.56 nm, 0.39 nm, 0.53 nm, and 0.49 nm for dATP, respectively) (Fig. 3B). The phosphate tail of dATP was shown to come in closer contact with several residues towards the bottom of the nucleotide binding pocket than ATP, including LYS 352, THR 353, THR 624, and ASP 626 (average distances were 0.83 nm, 0.95 nm, 0.91 nm, and 0.56 nm for ATP and

0.45 nm, 0.47 nm, 0.41 nm, and 0.42 nm for dATP, respectively (Fig. 3C). We hypothesize that this is due to the additional hydroxyl group on ATP (Fig. 1B), allowing it to form additional hydrogen bonding interactions with residues at the top of the binding pocket, while dATP does not, causing the phosphate tail to be pulled downwards. The hydrogen bond occupancy of the 2' hydroxyl group on ATP was measured to be 8.21% as a hydrogen bond acceptor, 4.80% as a donor, and overall acting as either a donor or acceptor in 12.4% of the simulation. While this is a minority of the simulation time, the consistent interaction significantly alters its orientation and behavior in the pocket. The preserved 3' hydroxyl shows similar levels of hydrogen bonding occupancy between ATP and dATP, at 30.4% and 34.8% occupancy respectively as either a donor or acceptor. Interestingly, in dATP, the 3' hydroxyl almost never bonds as a donor, at only 0.0312% of frame compared to 4.92% in ATP. These results are further supported by clustering analysis, which confirmed that ATP is most commonly located more horizontally across the top portion of the binding pocket across all 3 replicates, while dATP shows a more vertically aligned conformation in which the phosphate tail is drawn downwards towards the P domain. This serves to position dATP such that the gamma phosphate is located substantially closer to the phosphorylation residue ASP 351 (average distance was 1.2 nm for ATP and 0.74 nm for dATP) (Fig. 3A, C) [81]. This may allow for faster phosphorylation by dATP compared with ATP.

Finally, we assessed the differential effects of ATP and dATP on the movement of cytosolic domains of SERCA2A. Closure of the A and N domains (Fig. 1A) in particular is important for the E1-ATP to E1-ADP transition [131, 79, 131]. We found that the average difference between the center of mass of the A and N domains was smaller for dATP than for ATP (average distance was 3.8 nm for ATP and 3.5 nm for dATP), suggesting that dATP enhances closure of these domains (Fig. 4A). Further, the standard deviation in A-N domain distances was smaller for dATP than ATP (0.09 nm for ATP vs 0.08 nm for dATP), suggesting that dATP may also stabilize these domains in a more closed conformation. This was also shown to correspond to the first principal component from a PCA analysis conduced on SERCA structures from the GaMD simulations (Fig. 4B), suggesting that this is the major motion captured by our

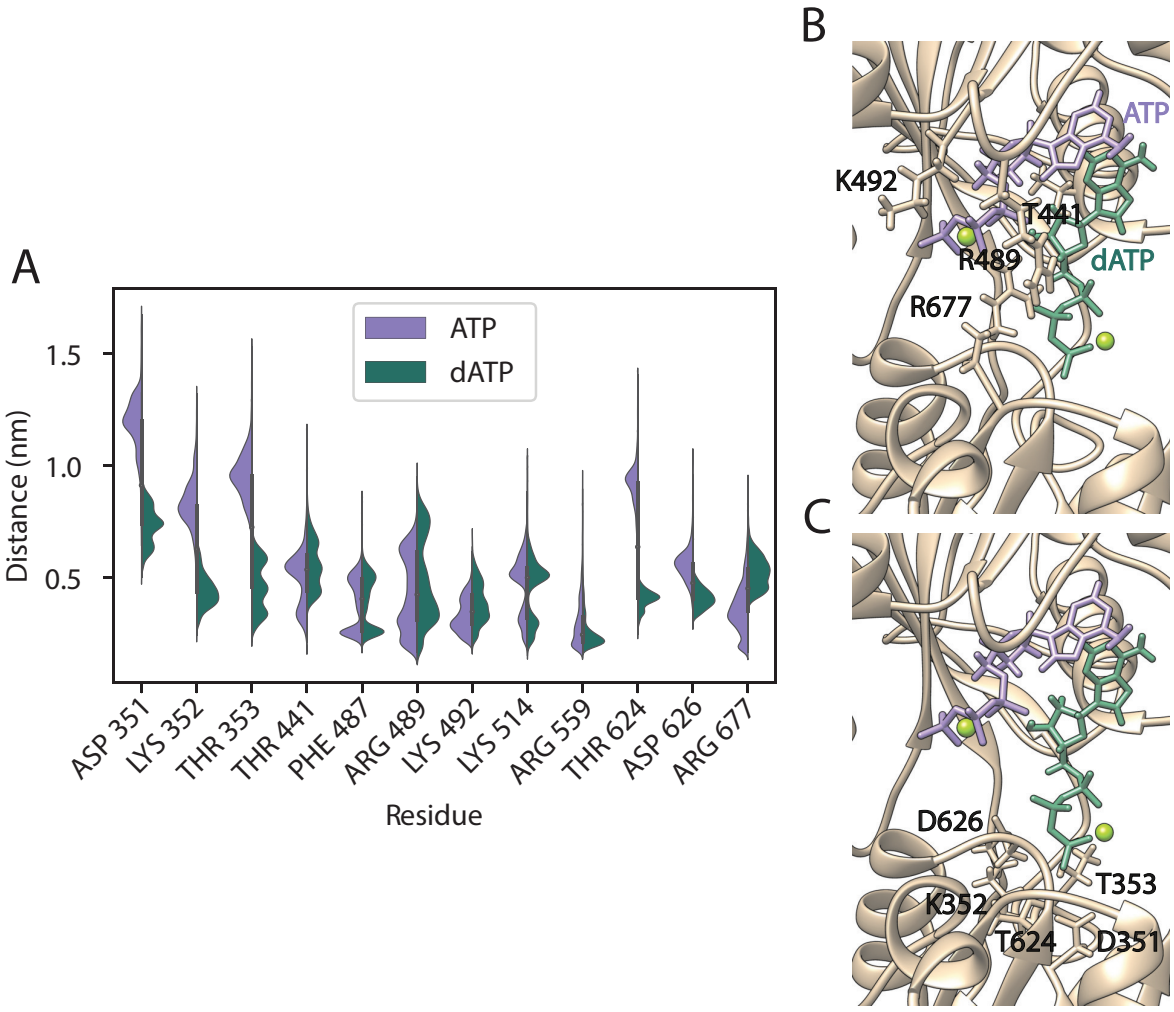


**Figure 1.4.** dATP occupies a more vertically aligned conformation in the binding pocket than ATP. (A) Distance between nucleotide and residues of interest for ATP and dATP (averaged across three replicates). For each residue of interest, distance distributions (between the residue and nucleotide) are shown as violin plots, where the white dot in the center represents the median, and the thick gray bar represents the interquartile range. Density curves are shown for ATP on the left side of the plot and dATP on the right side of the plot, where the width of the curve represents the frequency of the data (nucleotide-residue distance) at that point. ATP came in closer contact with residues towards the top of the binding pocket, including THR 441, LYS 492, ARG 677, and ARG 489 (shown in (B)). dATP was shown to come in closer contact with residues towards the bottom of the binding pocket including LYS 352, THR 353, THR 624, ASP 626, and phosphorylation residue ASP 351 (shown in (C)). This may be explained by the missing hydroxyl group on dATP leading to weaker interactions with residues at the top of the binding pocket. Distances to nucleotide association residues PHE 487 and ARG 559, as well as LYS 514, were unchanged.

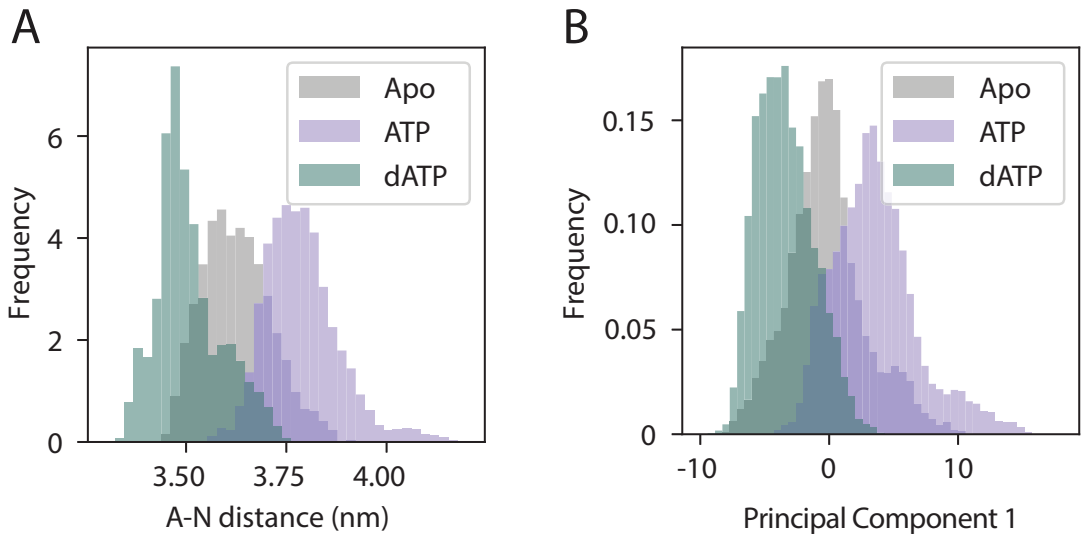


**Figure 1.5.** dATP enhances transition from E1-ATP to E1-ADP via A-N domain closure. Data are displayed as histograms showing the frequency distribution across the GaMD simulations. (A) A-N domain distance (averaged across 3 replicates), measured from the center of mass for each domain, was lower for dATP than ATP. (B) First principal component from PCA analysis of $C_\alpha$ movement performed on GaMD trajectories, which corresponds to A-N domain movement.

simulations. Further, reweighting analysis confirmed that these effects were not due to bias in the GaMD simulations since the decrease in A-N distance with dATP, represented by a minimum in the Potential of Mean Force (PMF), was still present even after recovering the original energy landscape (without the GaMD boost potential) via reweighting. Interestingly, ATP increased A-N domain distances with respect to the *apo* structure, while dATP decreased A-N domain distances with respect to the *apo* structure. This may be due to the fact that our GaMD simulations do not have calcium present; in the absence of calcium, the nucleotide can bind to the N domain and can lead to opening of the cytosolic domains, but cannot lead to subsequent complete closure of these domains or phosphorylation [81, 99]. The effects of dATP on SERCA appear to be pronounced enough to lead to greater closure of the A and N domains. We expect that if calcium was present in the GaMD simulations, ATP A-N distances would be smaller than the *apo* case, and dATP A-N distances would be decreased even further.

### 1.4.2 dATP binding to SERCA leads to opening of calcium binding path

Next, we assessed whether dATP binding affects calcium association in the transmembrane region. The calcium binding sites have been shown to be allosterically linked to the nucleotide binding site via transmembrane helix M5 [81]. Calcium binds first to site I (passing through site II, which is gated by GLU 309) [79, 131, 132, 128]. This is believed to occur through a cooperative mechanism, where binding of calcium to site I increases calcium binding affinity for site II [133, 132, 134]. We computed distances between THR 799 (M6), GLU 770 (M5), and GLU 709 (M8), since these residues make up calcium binding site I [131] (Fig. 1C). We found that the average distances between M5 and M6 were increased (average distances were 0.55 nm for ATP and 0.66 for dATP) (Fig. 5A). Average distances between M5 and M8 were also increased (average distances were 0.45 nm for ATP and 0.52 nm for dATP) (Fig. 5B). Average distances between M6 and M8, on the other hand, were decreased (average distances were 0.26 nm for ATP and 0.25 nm for dATP) (Fig. 5C). This suggests that the net effect of dATP is to cause opening of calcium binding site I, since M5-M6 and M6-M8 distances were increased by a greater magnitude than M6-M8 distances were decreased with dATP. Similar to the effects of dATP on A-N domain distances, here we observed again that ATP and dATP had opposing effects on transmembrane helix distances with respect to the *apo* structure. This may again be due to the fact that our GaMD simulations were carried out in the absence of calcium.

Further, we assessed the effects of dATP on calcium entry. Calcium is proposed to enter SERCA through two different paths [135, 132]. The first is composed of M6, M7, M8, and M9, while the second is composed of M1, M2, and M4. We chose to focus on the second, since there is greater evidence supporting this path [135, 136, 137, 138]. We assessed distances between GLU 51 (M1) and GLU 109 (M2), residues which have been implicated in calcium sensing and compose part of the binding path [135] (Fig. 1C), and found that dATP also led to an increase in this distance (average distance was 0.88 nm for ATP and 0.94 nm for dATP) (Fig. 5D). We again verified that these results were not due to bias in the GaMD simulations by

conducting a reweighting analysis since the changes in residue distances with dATP, represented by a minimum in the Potential of Mean Force (PMF), were still present even after recovering the original energy landscape (without the GaMD boost potential) via reweighting. Further, the standard deviation of the distance distribution was smaller for dATP than ATP (0.11 nm for ATP and 0.06 nm for dATP), suggesting that dATP also acts to stabilize these residues in a more open position. Thus, dATP may lead to opening of both the calcium binding path and calcium binding site I, which could facilitate enhanced calcium association.

### 1.4.3 dATP increases rates of nucleotide and calcium association to SERCA compared with ATP

BD simulations revealed that dATP bound more rapidly to the *apo* structure of SERCA compared with ATP. The weighted average binding curve shows that across a wide range of reaction distances greater than 7 Å, dATP binds more rapidly to the *apo* structure than ATP (Fig. 6A and B). Selecting a reaction distance of 8.11 Å corresponding to the ATP binding rate constant of $2.59\times10^7$ $(\mathrm{M\ s})^{-1}$ used by Tran *et al.* [85], the corresponding dATP association rate constant was 36% higher ($3.52\times10^7$ $(\mathrm{M\ s})^{-1}$). We attribute the lower dATP binding rate constants at low reaction distances to noise, since very few BD simulations reached these small distances. Less than 0.05% of simulations reached a distance of less than 8 Å and less than 0.03% of reactions reached a distance of less than 5 Å. Although ATP does have an additional polar hydroxyl group that may promote additional electrostatic attraction to the nucleotide binding pocket, we anticipate that removal of this functional group reduces steric hindrance, therefore allowing dATP easier access to the binding site, and that this effect dominates the BD simulations. Furthermore, analysis of a structure by structure comparison of the 30 SERCA conformations used in our simulations revealed that although the nucleotide binding rate constant varied considerably from conformation to conformation, dATP consistently associated more rapidly to SERCA compared with ATP.

Further, BD simulations of dATP-bound SERCA showed differences in calcium binding

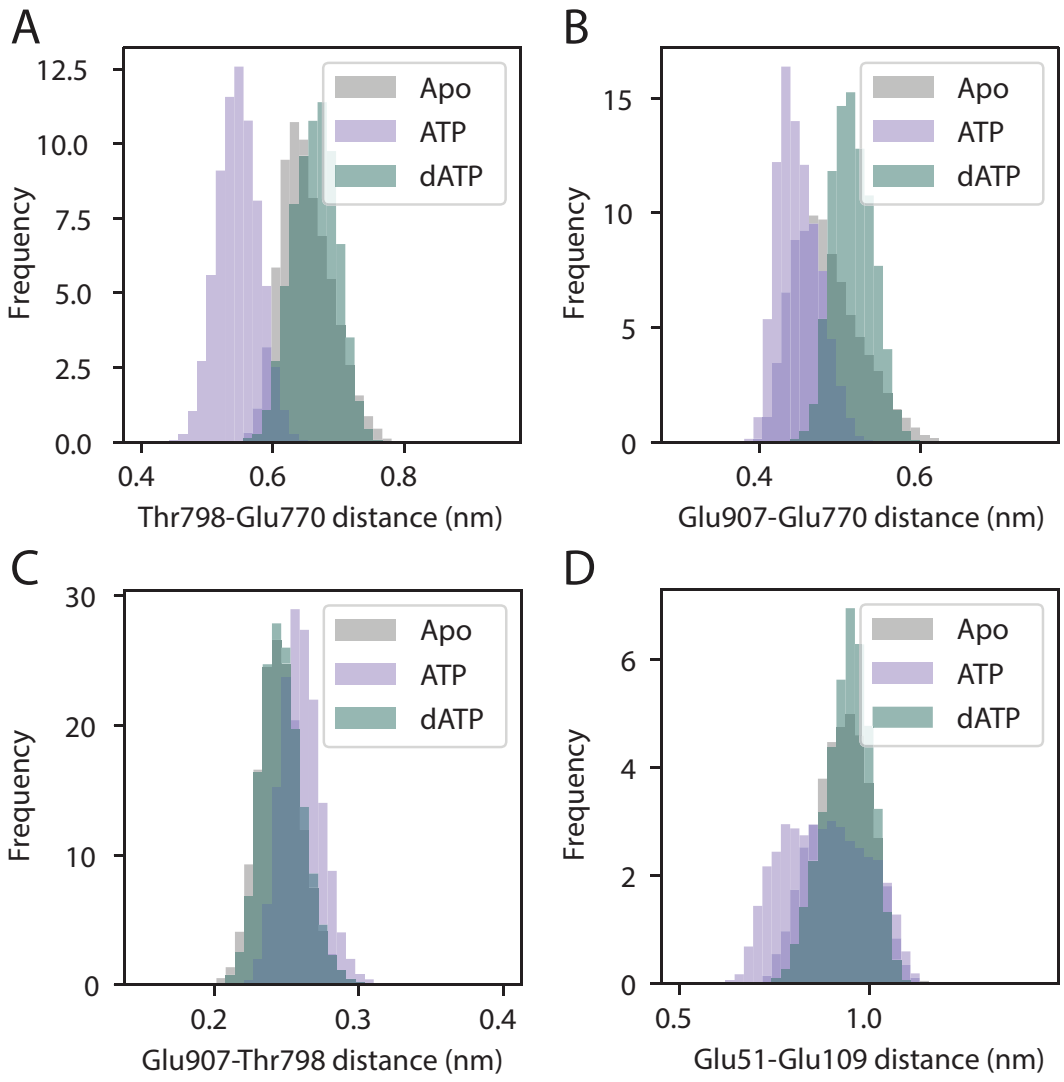


**Figure 1.6.** dATP binding leads to rearrangement of calcium binding path. Distances for key residues on M5 (GLU 770), M6 (THR 798), and M8 (GLU 907) that make up calcium binding site I, as well as calcium sensing residues on M1 (GLU 51) and M2 (GLU 109) that make up the calcium entry path are shown. Distances were averaged across 3 replicates. Data are displayed as histograms showing the frequency distribution across the GaMD simulations. Distances between M5 and M6 (A) and M5 and M8 (B) were increased, while distances between M6 and M8 (C) were decreased with dATP. Distances between M1 and M2 were increased with dATP (D). This suggests that dATP may facilitate calcium binding by opening the calcium entry path to site I.

compared to ATP-bound SERCA (Fig. 1.7C, D). For reaction distances between 4 and 12 Å, calcium bound more rapidly to site I when dATP was bound than ATP. Experimental estimates of calcium association rate constants to SERCA vary by up to three orders of magnitude, making determining an absolute binding distance and rate constant difficult, however a previous BD study estimated a calcium association rate constant of $3.13\times10^9$ (M s)$^{-1}$ to ATP-bound SERCA [135]. In our simulations, this value corresponded to a reaction distance of 10 Å at which calcium binding to site I of dATP-bound SERCA was 23% greater ($3.85\times10^9$ (M s)$^{-1}$). Interestingly, dATP reduced the association rate constant of calcium to site II. It is believed that calcium binds to site I first and initiates a conformational change before a second calcium ion can bind to site II. Therefore, the structures extracted from our GaMD simulations may not be an accurate representation of SERCA when the first calcium ion is bound. This process is also cooperative, so we anticipate that this sequential step significantly alters the binding kinetics of the second calcium ion. As such, we focus our subsequent multiscale analysis on the effects of dATP on calcium binding to SERCA at site I.

### 1.4.4 Enhanced calcium binding to dATP-bound SERCA accelerates myocyte calcium transient decay

We optimized parameters of the Himeno ECC model to match measured calcium transients in the presence of ATP, as described in the methods [100, 127, 38]. In the Tran model of SERCA kinetics[85] (incorporated within the Himeno model), $K_1^+$ (rate of ATP binding to SERCA) and $K_{d,Cai}$ (dissociation constant for calcium binding to SERCA) were then adjusted based on the results of the BD simulations. Thus, $K_1^+$ was increased from $2.59\times10^7$ (M s)$^{-1}$ to $3.52\times10^7$ (M s)$^{-1}$ based on the BD-predicted change in nucleotide association to SERCA, and $K_{d,Cai}$ was was decreased by 23% from 0.0027 mM to 0.0021 mM based on the BD-predicted change in calcium binding to site I. These changes decreased time to 50% decay (DT50) of the calcium transient, but only modestly from 279 ms to 240.5 ms (Fig. 7A). However, decreasing $K_{d,Cai}$ by 41%, which is still within the range of the BD simulation results, shortened DT50 by

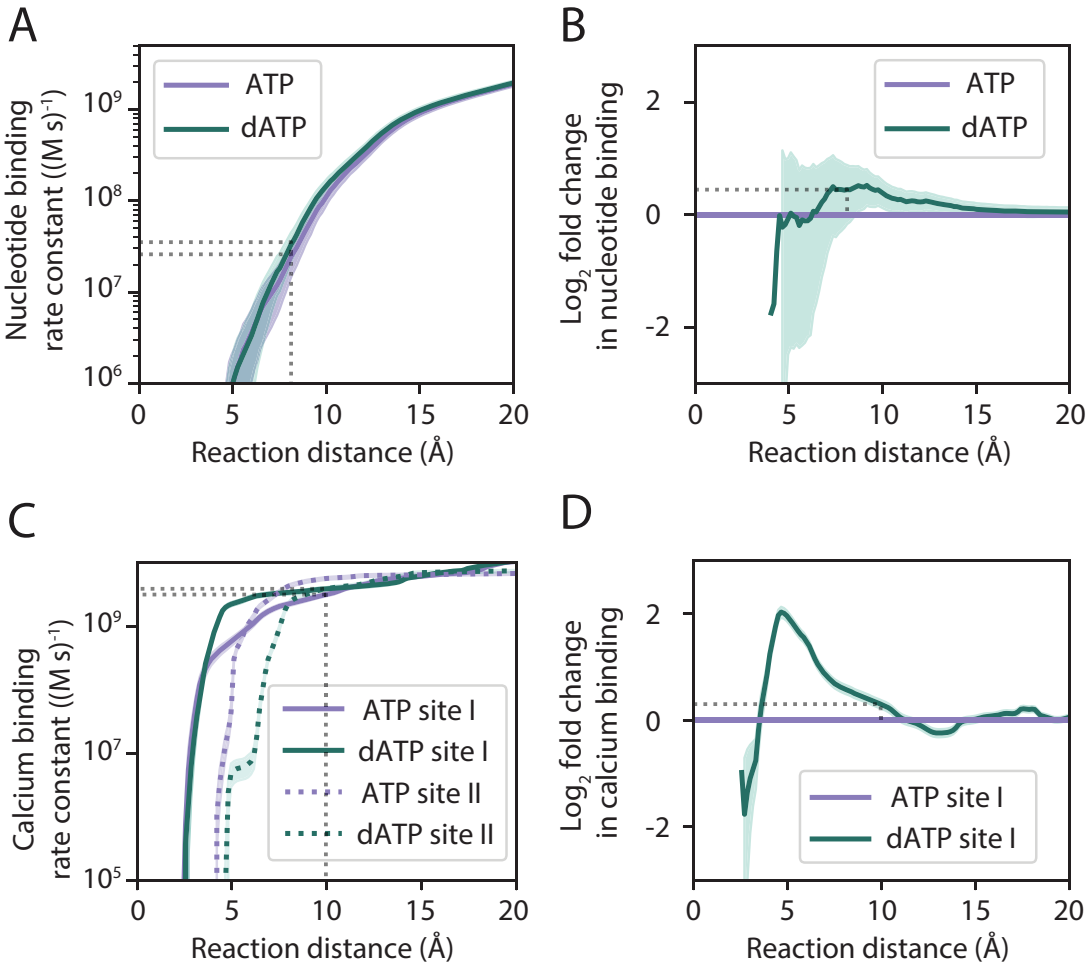


**Figure 1.7.** (A) BD-predicted binding rate constants as a function of reaction distance. Dotted lines represent reaction distance for association rate constant of ATP to SERCA used in the Tran *et al.* model [85]. (B) $Log_2$ fold change in nucleotide binding rate constant as a function of reaction distance. $Log_2$ fold change is used to more easily visualize changes in binding rate constant e.g. a doubling of the binding rate constant is equivalent to a $log_2$ fold change of 1, quadrupling is equivalent to a $log_2$ fold change of 2, etc. Log fold changes of less than 1 are negative while fold changes greater than 1 are positive (e.g. a halving of the binding rate constant is equivalent to a $log_2$ fold change of -1). dATP binds more rapidly to the nucleotide site for reaction distances greater than 7 Å below which sampling errors increase owing to the low number of simulations reaching lower reaction distances. (C) BD-predicted association rate constant of calcium to sites I and II. dATP increases calcium binding to site I compared with ATP. Calcium binds site II more rapidly when ATP is bound to SERCA than dATP. Dotted lines represent reaction distance that corresponds to the calcium association rate constant determined from a previous BD study [135]. (D) $Log_2$ fold change in the association rate constant of calcium to site I when dATP is bound compared with ATP.

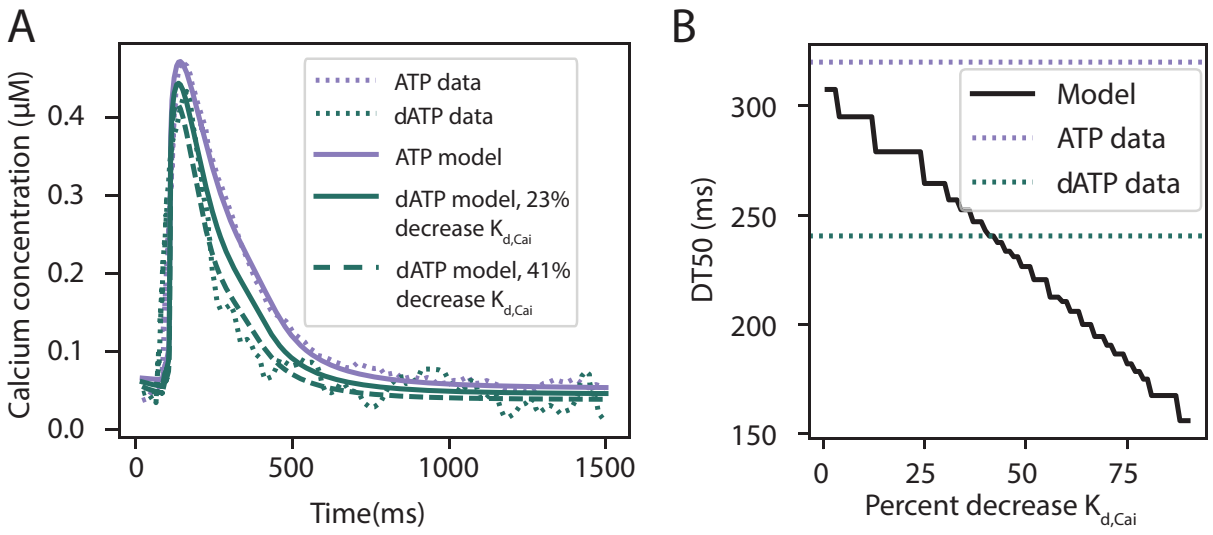


**Figure 1.8.** dATP accelerates the rate of myocyte calcium transient decay via enhanced calcium association. (A) Effects of changing model parameters $K_1^+$ and $K_{d,Cai}$ (nucleotide and calcium association rate constants, respectively) in the Tran model [85]. ATP and dATP experimental data digitized from Korte *et al*. [38] are shown as dotted lines. ATP model calcium transient (optimized to match ATP experimental data), and dATP predicted model calcium transients for 36% increase in $K_1^+$ combined with 23% and 41% decreases in $K_{d,Cai}$ are also shown. (B) Percent decrease in $K_{d,Cai}$ vs DT50. With a 23% decrease in $K_{d,Cai}$ (based on BD results), we are not able to fully match experimental measurements. However, with a 41% decrease in $K_{d,Cai}$, we are able to match experimental measurements of DT50 with dATP. These simulations also include a 36% increase in $K_1^+$ based on BD results, but changes in $K_1^+$ were not shown to substantially affect DT50.

an amount similar to experimental observations in the presence of dATP (Fig. 7B). In contrast, changes in the nucleotide association rate constant had little effect, suggesting that increasing calcium binding to site I may be the primary mechanism by which dATP increases SR calcium reuptake.

## 1.5 Discussion

The results of this study provide new evidence for mechanisms by which dATP treatment may contribute to improved SERCA pump function in cardiac myocytes. Firstly, dATP was shown to be more stable in the binding pocket of SERCA and was positioned to facilitate faster phosphorylation. Increased separation of the A-N cytosolic domains indicates faster pump

function with dATP treatment. Further, we observed separation of transmembrane helices M5 and M6 and M5 and M8, as well as M1 and M2, which may facilitate faster calcium association. dATP demonstrates a 36% higher association rate constant than ATP to the N domain of SERCA. However, inputting the nucleotide and calcium association rate constants determined from BD analysis ($K_1^+$ and $K_{d,Cai}$) into an ECC model did not fully explain experimentally observed differences in the calcium transient due to dATP, but further decreasing the calcium dissociation constant could.

While this study provides valuable insight into the molecular mechanisms by which dATP affects SERCA function, our GaMD simulations did not capture the phosphorylation event or downstream protein conformational changes after calcium binding. Future molecular dynamics or quantum mechanics simulation studies would aid in better elucidating these effects. Our BD simulations also allowed for the entry of the nucleotides and calcium from any orientation from the SERCA pump, not only the cytosolic side, which is a limit of our approach. Moreover, there are also other ATPases that contribute to intracellular calcium dynamics such as the Plasma Membrane Calcium-ATPase (PMCA) and the Sodium-Potassium ATPase exchanger (NCX) which may also be affected by dATP to enhance calcium efflux from the cell. SERCA is regulated by phospholamban and several post-translational modifications that could also affect its pump function [92]. It is likely that both SERCA upregulation and these other pumps together could explain the improved relaxation observed experimentally due to dATP treatment. This could be further explored with computational and experimental studies that specifically focus on the effects of the cooperative binding of $Ca^{2+}$ with relation to dATP and ATP. For instance, additional MD simulations of SERCA with one or both $Ca^{2+}$ binding sites occupied will certainly change the free-energy landscape and thus, yield additional structures and insights. Additionally, since dATP is a candidate therapeutic approach for heart failure with reduced ejection fraction, which is characterized by prolonged twitch relaxation and calcium transient decay associated with downregulated SERCA function, it would be useful to apply this analysis to a model of ECC in the failing cardiac myocyte.

In this study, we analyzed the effects of nucleotide binding on SERCA pump kinetics and the differential effects of ATP and dATP on calcium affinity via changes to the transmembrane domain. This study has demonstrated the power of multiscale modeling for investigating the effects of ATP analogs on cardiac cells, as we integrated knowledge from the atomic to the cellular level to uncover potential mechanisms of dATP which scale up to a significantly altered calcium transient and cardiac function as a whole.

## 1.6 Acknowledgments

Chapter 1, in full, is a reprint of the material as it appears in the *Journal of Applied Physics*. Multiscale computational modeling of the effects of 2'-deoxy-ATP on cardiac muscle Ca2+ handling. M. T. Hock*, A. E. Teitgen*, K. J. McCabe*, S. P. Hirakis, G. A. Huber, M. Regnier, R. E. Amaro, J. A. McCammon, and A. D. McCulloch. Multiscale computational modeling of the effects of 2'-deoxy-ATP on cardiac muscle Ca2+ handling. J. Appl. Phys. 134(7):074905, 2023. *Equal contribution. The dissertation author was the co-first author of this publication.

# Chapter 2

# Correlated Motion Analysis of SERCA Allostery

## 2.1 Abstract

We present the Netsci program - an open-source scientific software package that leverages GPU acceleration and a k-nearest-neighbor algorithm in order to estimate the mutual information (MI) between data in a set. The GPU acceleration presented here, as an improvement upon existing estimators, enables calculation speeds several orders of magnitude faster than CPU-based implementations, all with dataset size limits determined only by the available hardware. To demonstrate the validity and usefulness of Netsci, we show that the MI is correctly computed for the analytically-verifiable two-dimensional Gaussian distribution, and we also reproduce the generalized correlation (GC) analysis performed in an earlier study on the B1 domain of protein G. In addition, we apply Netsci to the analysis of molecular dynamics simulations of the Sarcoendoplasmic Reticulum Calcium-ATPase (SERCA) pump. Specifically, we use Netsci to understand the allosteric mechanisms and pathways of SERCA, and compare the differential effects of the binding of two nucleotides, ATP and 2'-deoxy-ATP (dATP). We determine that ATP binding to SERCA, compared to dATP, induces differential allosteric effects. The most likely information pathways from the bound nucleotide to the calcium binding domain are also predicted using our MI estimator in combination with network analysis tools on the SERCA pump, which differs based on the bound nucleotide. Netsci is shown to be a useful program

for the estimation of MI and GC within general datasets, and for the analysis of intraprotein communication and information transfer, in particular.

## 2.2 Introduction

One important problem within molecular structural biology is the need to quickly and robustly extract meaningful information about the concerted motion between different regions of a biomolecule, or as an effect of other molecules, from the large quantities of data generated within simulation trajectories detailing atomic motion. For instance, allostery occurs when a binding event or conformational change at one part of a protein causes a meaningful structural change at another part of the protein - this can be quantified by analyzing the correlated motion between the different parts of the protein, signifying the strength of their relationship. Many alternative approaches have been developed to solve this problem. For instance, such studies have focused on the communication of allosteric information between any two residues in a protein [58, 139, 59, 140, 141, 142, 143]. Among them, metrics that make use of mutual information (MI), and its extension, generalized correlation (GC), stand out for their ability to handle nonlinear relationships. A number of tools exist which can peform MI or GC for the analysis of biomolecular motion, some of the most prominent being Dynetan[144], and MDiGest[145]. However, these packages only utilize CPU implementations of the MI estimation algorithm.

In this work, we develop a novel computational tool called "Netsci" that uses GPU-accelerated code in order to perform the k-nearest neighbors algorithm for estimating MI, and by extension, GC. We show that Netsci correctly predicts the MI for analytically verifiable systems, reproducing the results of non-GPU-accelerated (CPU) implementations. When compared to CPU implementations, Netsci can perform MI calculations many orders of magnitude faster, and is not subject to space constraints - Netsci is essentially limited only by hard constraints imposed by system hardware. Although we have prepared Netsci to be usable for a wide variety of data

applications in any scientific field, we focus on applying Netsci to perform correlation analyses of real biomolecular systems of interest.

Molecular dynamics (MD) simulation is a widely-used computational technique to provide atomic-resolution details about the behavior of biomolecular systems. By utilizing a classical approximation of the forces between atoms of the system, a molecular trajectory may be propagated forward in time, showing likely conformational motions and both intramolecular and intermolecular interactions. When the trajectory is written, MD simulations generate large quantities of data, encapsulating the details of the atomic motions. A detailed explanation of generalized correlation within the context of MD is provided below in the Theory section. We repeat and regenerate the results of a seminal GC analysis of the B1 domain of protein G. We also perform a new study on the inter- and intra-molecular correlations within the Sarcoendoplasmic Reticulum Calcium-ATPase (SERCA) pump as an additional use case. We also implement a "local alignment" procedure that seems to greatly reduce spurious correlations detected in large or highly flexible structures.

Netsci has been applied to the analysis of allosteric motions within a biomolecule in this study, yet the tool may be used for purposes outside of biophysics, and may be applied to the fast correlation analysis of large data sets, regardless of their field of application.

## 2.3 Theory

### 2.3.1 Information theory and mutual information

The original presentation of information theory by Shannon [146] applied specifically to the transmission of a message of length $N$ from a source to a destination across a noisy channel. In the time since the original publication, information theory has been applied to a wide variety of applications [147, 148, 149, 150]. To generalize beyond the communication problem, we will assume that a bivariate set of data $z_i = (x_i, y_i)$, $i = 1, ..., N$ has been gathered. The data sets $x$ or $y$ can be of any dimension. We will assume that each of the $N$ elements of the data are

independent and identically distributed realizations of the random variables $Z = (X,Y)$, and that the they are distributed according to $\mu(x,y)$, a proper smooth function. The marginal densities are $\mu(x) = \int \mu(x,y)dy$ and $\mu(y) = \int \mu(x,y)dx$.

The Shannon entropy can be defined as:

$$H(X) = -\int \mu(x)\log\mu(x)dx \tag{2.1}$$

where the base of the logarithm depends on the units desired for the information, whether bits ($\log_2$), nats ($\log_e$), decimal digits ($\log_{10}$), or otherwise. In this work, we will use the natural logarithm. The Shannon entropy, in this context, represents the logarithm of the number of "reasonably probable" data outcomes. The mutual information $I(X,Y)$ is defined as:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{2.2}$$

The value of $I(X,Y)$ measures the strength of the connection between the variables $X$ and $Y$; if the two variables were completely independent, then $I(X,Y)$ would be zero.

MI may be exactly obtained only if the distribution of variables is known. In most cases, the underlying distribution of data is not known, and the distribution, and thus the MI, must be estimated approximately. A number of methods to estimate MI have been developed, including approaches that use cumulant expansions [151, 152], kernel density estimators [153, 154], and adaptive binning [155, 156, 157, 158]. One successful algorithm to estimate MI, which we particularly focus on, uses a k-nearest neighbor approach to estimate the densities of the data within the distribution [159]. The k-nearest neighbor approach to estimating MI has been utilized in a number of interesting biophysics applications [160, 161]. The advantages of the k-nearest neighbors approach include data efficiency (it works well for smaller data sets compared to other estimators), adaptive (the algorithm automatically uses higher resolution where the data is dense), and minimal bias.

Under the condition that $\mu$ is a uniform distribution, we may approximate the continuous

integral in eq. 2.1 by a discrete sum.

$$\widehat{H}(X) = -\frac{1}{N}\sum_{i=1}^{N}\widehat{\log(\mu(x_i))} \tag{2.3}$$

Now, an estimate for $\widehat{\log(\mu(x_i))}$ must be defined. In this paper, we use a k-nearest neighbor estimator. In order to rank neighbors of a data point $z_i$ by nearness, we use the max norm,

$$||z - z'|| = \max\{||x - x'||, ||y - y'||\} \tag{2.4}$$

where we choose to use the a similar max norm for $||x - x'||$ and $||y - y'||$, although this is not required - a Euclidean norm could be used, for instance. For each data point $z_i$, let $\varepsilon_x(i)/2$ and $\varepsilon_y(i)/2$ represent the distances from $z_i$ to its $k^{th}$ nearest neighbor projected onto the $X$ and $Y$ subspaces, respectively. The value $p_i$ is the integrated density within a distance $\varepsilon/2$ of the point $x_i$ defined as:

$$p_i(\varepsilon) = \int_{||\xi - x_i|| < \varepsilon/2} \mu(\xi) d\xi \tag{2.5}$$

Note that $\varepsilon(i) = \max\{\varepsilon_x(i), \varepsilon_y(i)\}$. Consider the probability distribution:

$$P_k(\varepsilon_x, \varepsilon_y) = P_k^{(b)}(\varepsilon_x, \varepsilon_y) + P_k^{(c)}(\varepsilon_x, \varepsilon_y) \tag{2.6}$$

Specifically, $P_k^{(b)}(\varepsilon_x, \varepsilon_y)$ represents the probability distribution that there are $k-1$ data points within the rectangle $x_i \pm \varepsilon_x(i)/2$ and $y_i \pm \varepsilon_y(i)/2$, a rectangle defined by the $k$th nearest neighbor within in the $x$ subspace, $N-k-1$ points that are outside a different rectangle defined by $x_i \pm (\varepsilon_x(i) + d\varepsilon_x)/2$ and $y_i \pm (\varepsilon_y(i) + d\varepsilon_y)/2$, and one data point in the space between the two rectangles. The probability distribution $P_k^{(c)}(\varepsilon_x, \varepsilon_y)$ is similar, though the rectangles are defined by the $k$th nearest neighbor within the $y$ subspace, which may be the same, or a different, point

used to define $P_k^{(b)}(\varepsilon_x, \varepsilon_y)$. These quantities are then

$$P_k^{(b)}(\varepsilon_x, \varepsilon_y) = \binom{N-1}{k} \left( \frac{d^2[q_i^k]}{d\varepsilon_x d\varepsilon_y} \right) (1-p_i)^{N-1-k} \tag{2.7}$$

and

$$P_k^{(c)}(\varepsilon_x, \varepsilon_y) = (k-1) \binom{N-1}{k} \left( \frac{d^2[q_i^k]}{d\varepsilon_x d\varepsilon_y} \right) (1-p_i)^{N-1-k} \tag{2.8}$$

Similar to $p_i$, the value $q_i(\varepsilon_x, \varepsilon_y)$ is the integrated density within a tiny rectangle of size $\varepsilon_x \times \varepsilon_y$ centered at $(x_i, y_i)$. As mentioned before, $p_i$ is the integrated density within a tiny square of side length $\varepsilon$ - tiny enough that we may assume that $\mu(x)$ is constant within:

$$p_i(\varepsilon) \approx c_d \varepsilon^d \mu(x_i) \tag{2.9}$$

where $d$ is the dimension of $x$, and $c_d$ is the volume of the $d$-dimensional unit ball. For the maximum norm used in this study, we simply use $c_d = 1$. In this case,

$$I(X_1, X_2) = \psi(k) - 1/k - \langle \psi(n_x) + \psi(n_y) \rangle + \psi(N) \tag{2.10}$$

where $\psi(x)$ is the digamma function, and $n_x(i)$ and $n_y(i)$ are the number of points with distance less than or equal to $\varepsilon_x(i)/2$ and $\varepsilon_y(i)/2$, respectively.

### 2.3.2 Generalized correlation

Many approaches to estimate the correlation of data have been developed. One of the most commonly used correlation measures is the Pearson (linear) correlation coefficient - which quantifies the strength of the relationship between two variables as a value between -1 and 1. A Pearson correlation value of 1 indicates perfect correlation, a value of 0 indicates independence of the variables, and a value of -1 indicates perfect anticorrelation. The Pearson product-moment

correlation is defined as

$$r(X,Y) = \frac{\mathbb{E}(X-\mu_X)(Y-\mu_Y)}{\sigma_X \sigma_Y}, \tag{2.11}$$

where $\mathbb{E}$ indicates expectation value, $\mu_X$ is the mean of random variable $X$, $\mu_Y$ is the mean of random variable $Y$, $\sigma_X$ is the standard deviation of $X$ and $\sigma_Y$ is the standard deviation of $Y$. The Pearson correlation defines the best linear fit between data sets sampled from $X$ and $Y$. As mentioned before, the Pearson correlation suffers a number of insufficiencies when used for data that is related nonlinearly, and sets of related vector data that oscillate in non-parallel directions. The MI can address these shortcomings. In order to provide an equivalent quantity to $r$ in eq. 2.11, researchers have defined the GC coefficient $r_{\mathrm{MI}}$ [59], which makes use of the MI.

$$r_{\mathrm{MI}}(X,Y) = \left(1 - e^{-\frac{2I(X,Y)}{d}}\right)^{\frac{1}{2}} \tag{2.12}$$

### 2.3.3 Local alignments

While performing MI calculations on large structures, such as the SERCA pump, we had initially aligned the structures against a reference structure (for instance, the first frame of the trajectory). We observed that this approach tended to cause the MI algorithm to predict anomalously large correlations at the distant extremities of the protein, as the larger structures would exhibit "breathing" motions upon structural alignment. Since large groups of atoms at the protein extremities would exhibit concerted motions that would disproportionally bias the algorithm, causing it to indicate an unrealistically large correlated motion for these groups of atoms, compared to the relatively stationary atoms at the protein core.

To remedy this problem, we developed an algorithm to perform local structural alignments. For each node of data (each residue center of mass, in this study), we choose all adjacent nodes within a certain cutoff value (10 Å in this study), not including the original node. We then align the entire molecule by this set of adjacent nodes, and then input the trajectory positions of the original node into the MI algorithm for analysis. All molecular MI results reported were

made following such a local alignment - which functioned well to remove the anomalously large correlated motions observed at the protein extremities when such a local alignment was not performed.

### 2.3.4 Network construction and analysis

The full residue-residue matrix of GC measurements provides a robust dataset with a wide range of analyses to consider. Converting the measured correlation between residues into an adjacency matrix first, and then a full network, opens the correlated motion analysis to another myriad of graph theory analyses. In this work, we construct a weighted graph between residues by converting the $r_{MI}$ matrix into a weighted adjacency matrix $A$. Specifically, $A(i,j) = -\ln(r_{MI}(i,j))D(i,j)$ where $r_{MI}(i,j)$ is the GC between residues (or nodes) $i$ and $j$ and $D(i,j)$ is the shortest distance between any heavy atoms of residues $i$ and $j$. In each case, the structure used to calculate the distances was the average structure from each individual MD simulation using a traditional whole protein RMS alignment, not the local alignment approach outlined in the previous section. In this definition, although no cutoff distance is necessary to define whether two residues can be connected, closer residues are prioritized.

With this adjacency matrix, a weighted graph is easily constructed with shorter edge weights corresponding to highly correlated motions in closer proximity. We then use shortest path algorithms to identify pathways between residues of interest and compare the information flow through allosteric pathways.

## 2.4 Results and Discussion

### 2.4.1 Gaussian distribution

We first validate the Netsci implementation of the k-nearest neighbors algorithm using a distribution whose MI value can be exactly computed - a two-dimensional Gaussian distribution with zero mean and unit variance, and a covariance value $r$. Netsci predicts the MI with an

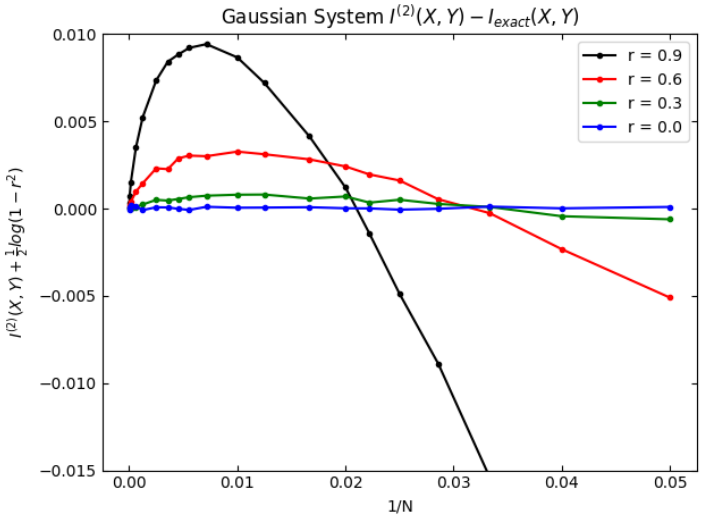


**Figure 2.1.** Absolute error $I^{(2)}(X,Y) - I_{exact}(X,Y)$ for the Gaussian system with zero mean, unit variance, and differing values of covariance $r$. This plot demonstrates the sensitivity of the algorithm to the number of data points, showing the importance of larger N. In all cases, $k = 1$ and the number of trials is $2 \times 10^6$ if $N \leq 100$, $5 \times 10^5$ if $100 < N \leq 1000$, and $1 \times 10^4$ if $N > 1000$.

accuracy typical of previous implementations of the k-nearest-neighbors algorithm (Figure 2.1; compare with Figure 2 of Kraskov *et al.* [159]).

In this case, the exact value of I(x,y) is known (Eq. 2.13).

$$I_{exact}(X,Y) = -\frac{1}{2}\log\left(1 - r^2\right) \tag{2.13}$$

In Figure 2.1, we show the relative error between the exact MI and the MI computed for a few values of $r$, where $r$ is the determinant of the covariance matrix describing the 2D Gaussian. As was also shown by Kraskov *et al.*, the estimates improve with larger N, and the algorithm shows no systematic error when the Gaussians are truly uncorrelated ($r$=0). Because of this systematic error, Kraskov *et al.* recommended a value of $k$ between 2 and 4, which shows reduced noise compared to a $k$ value of 1, but without the large systematic bias from larger values of $k$. However, if one is testing for statistical independence of the datasets, Kraskov *et al.* recommended a value of $k = N/2$, since for independent data, the systematic error will be

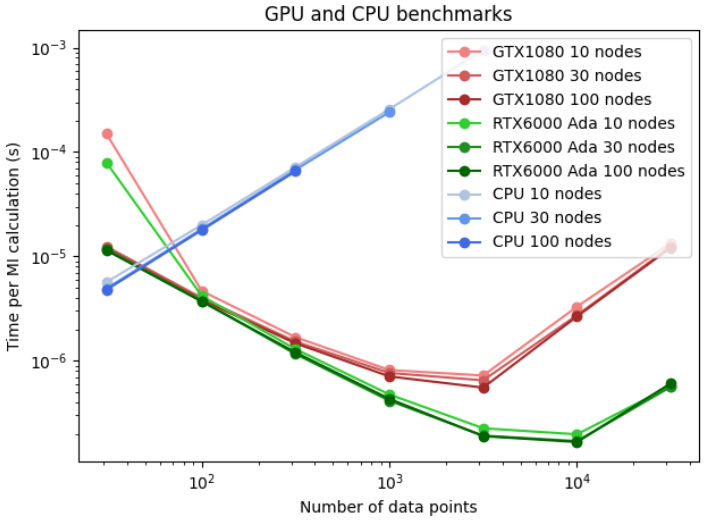


**Figure 2.2.** Benchmarks of Netsci computed for sets of data points sampled from independent Gaussian distributions. For each platform, pairwise MI quantities for 10, 30 and 100 nodes were defined, each with independent Gaussian distributions relating each pair of nodes. MI values were computed pairwise for each distribution. The x-axes indicate the number of data points, for the set of all distributions, that were sampled and fed into the MI algorithm. The time per MI calculation is plotted on the y-axis, representing the time per individual MI calculation between each pair of nodes. For larger inputs, benchmarks seemed to show that the GPU algorithm consistently performed approximately three orders of magnitude faster than the CPU implementation using GTX1080 GPUs, and yet another order of magnitude of performance, at least, was gained by using RTX6000 Ada GPUs. Note that some data points overlap due to close similarities in the benchmarks.

nonexistent, and a larger $k$ value will diminish noise. This toy system shows that our GPU implementation of the algorithm is correct - producing the same results as Kraskov's original publication and the same as a correct CPU implementation. Benchmarks show a performance gain of at least three orders of magnitude for large inputs, compared to the CPU implementation, when using a GTX1080 GPU, and at least another additional order of magnitude increase when using the powerful RTX6000 Ada GPUs (Figure 2.2).

### 2.4.2 Protein G System

In order to demonstrate the utility and correctness of our GPU implementation of the MI and GC estimator, we repeat a previous study by Lange *et al.* [59], where GC estimates

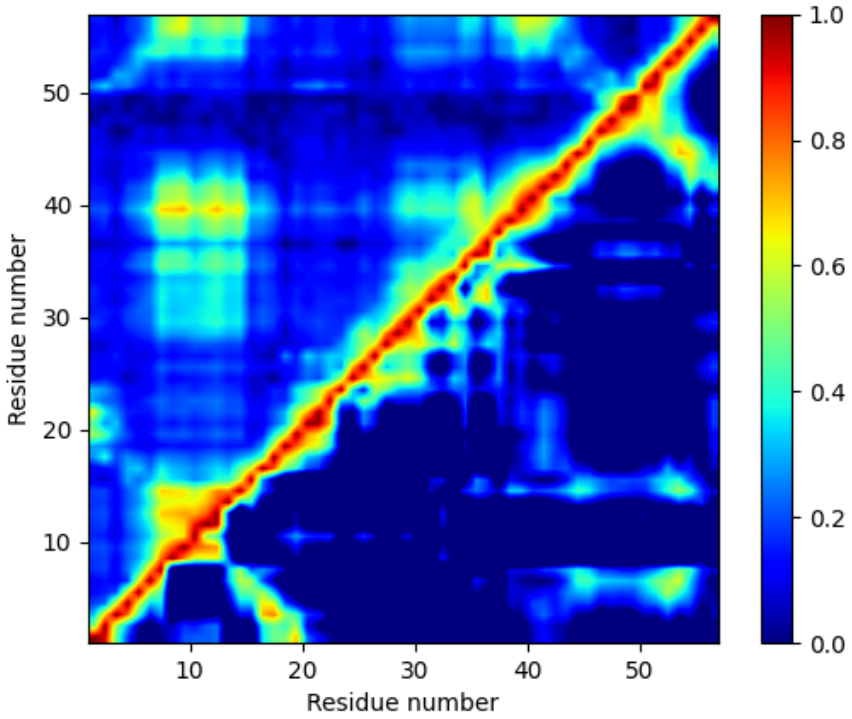


**Figure 2.3.** The GC coefficient (upper echelon) and the Pearson coefficient (lower echelon) correlation matrices for the B1 domain of protein G.

using MI were first applied to a set of biochemical systems. One of these systems was the B1 domain of Protein G. Lange *et al.* examined the correlated motions of protein G using both GC and the linear (Pearson) correlation, showing that linear correlation predicted a number of key relationships within protein G intramolecular motion, but that a GC description revealed all the same relationships and more. Their original results are replicated in this study and displayed in Figure 2.3.

As observed in Figure 2 of Lange *et al.*, strong correlations are detected near the main diagonal, due to close backbone connections, as well as two other diagonals perpendicular to the main. As mentioned by Lange *et al.*, these correlations represent tight hydrogen-bond contacts between beta strands $\beta 1$ and $\beta 2$, as well as between $\beta 3$ and $\beta 4$. A weaker correlation close to the main diagonal between residues 22 and 38 are caused by the less-tight-packing within the

$\alpha$-helix. The new signals that appeared due to the use of GC, but not linear correlation, include a connection between the $\alpha$-helix and the $\beta 2$ sheet - probably due to their close proximity. One interesting result of this analysis implies that, while linear correlations would detect correlations *within* secondary structure elements, GC showed the ability to detect correlations *between* different secondary structure elements, as highlighted by Lange *et al.* [59]. We used the local alignment procedure for this analysis which, due to protein G's relatively small size, had little effect on the outputs, and very similar results to Lange *et al.* were generated, in spite of the fact that Lange *et al.* aligned the entire trajectory using a global least-squares fitting to all alpha carbons.

### 2.4.3 SERCA pump results

The SERCA pump is localized to the Sarcoplasmic Reticulum (SR) membrane and is responsible for pumping calcium ions back into the SR following muscle contraction. SERCA is an ATPase pump and is thus powered by energy from ATP hydrolysis, allowing it to establish a concentration gradient of calcium across the SR membrane. It is typically characterized by its three cytosolic domains, canonically labeled "N", where nucleotide binding occurs, "A", the actuator domain, and "P", where phosphorylation occurs [81]. After nucleotide binding to the N domain, the N and A domains move within a closer proximity and phosphorylation occurs at ASP351, which allows for further conformational changes in the protein, eventually leading to calcium release into the SR [162]. Further, SERCA contains 10 transmembrane helices, typically labeled M1 through M10, where calcium binding occurs [80]. Specifically, calcium from the cytosol binds between helices M4, M5, M6, and M8. Helices M1 and M2 make up part of the calcium entry pathway into these binding sites [81]. Two calcium ions bind per hydrolyzed ATP, at two sites within this region. Calcium binds cooperatively, and binding of calcium to the first binding site increases the affinity of the second binding site for calcium [82].

SERCA 2A is the dominant isoform found in cardiac muscle, and thus plays a large role in cardiac relaxation and calcium handling [163]. SERCA 2A dysfunction has been linked to

pathogenesis in the heart and elsewhere, and SERCA has also been investigated as a therapeutic target [90, 163]. A recent computational study investigated the effects of 2'-deoxy-ATP (dATP), a small molecule analog of ATP, on SERCA [164]. dATP is a potential small molecule therapeutic for treating heart failure, and has been shown to improve contractile function in the heart by acting on myosin, a motor protein largely responsible for cardiac contraction [50, 32, 87, 39, 36, 89, 40, 38, 88]. dATP has also been shown to speed reuptake of calcium into the SR, leading to improved cardiac relaxation [38]. We previously showed that this could be at least partially explained by dATP binding and activating the SERCA pump in place of ATP [164]. This study found that dATP formed more stable contacts in the nucleotide binding pocket than ATP, leading to enhanced A-N domain closure and an increase in the calcium association rate *via* opening of the calcium binding pathway. However, it is unclear how changes in the cytosolic region due to dATP binding lead to conformational and functional changes in the transmembrane calcium binding region. The M5 helix has been shown to mechanically couple the calcium and nucleotide binding sites [81], but it is unknown how dATP affects this or if other connection pathways exist. Therefore, MI provides a useful tool for investigating how conformational changes to the protein following dATP binding propagate to changes in calcium association.

In order to demonstrate the practical utility of Netsci on a real system of biomedical interest, we perform a number of analyses using GC on the SERCA pump. As mentioned previously, the SERCA pump binds ATP (or dATP), which induces an allosteric mechanism that binds and pumps calcium ions across a lipid bilayer. We first examine the correlated motions between the nucleotide (ATP or dATP) bound in the active site and the alpha carbons within the entire protein (Fig. 2.4).

**Overall protein correlation and nucleotide influence**

As can be seen in Fig. 2.4, a wide range of correlations are observed between the motions of the ATP and dATP center of mass and the center of mass of each residue. Most notably, relatively low correlation is observed between ATP and the residues forming the calcium-binding

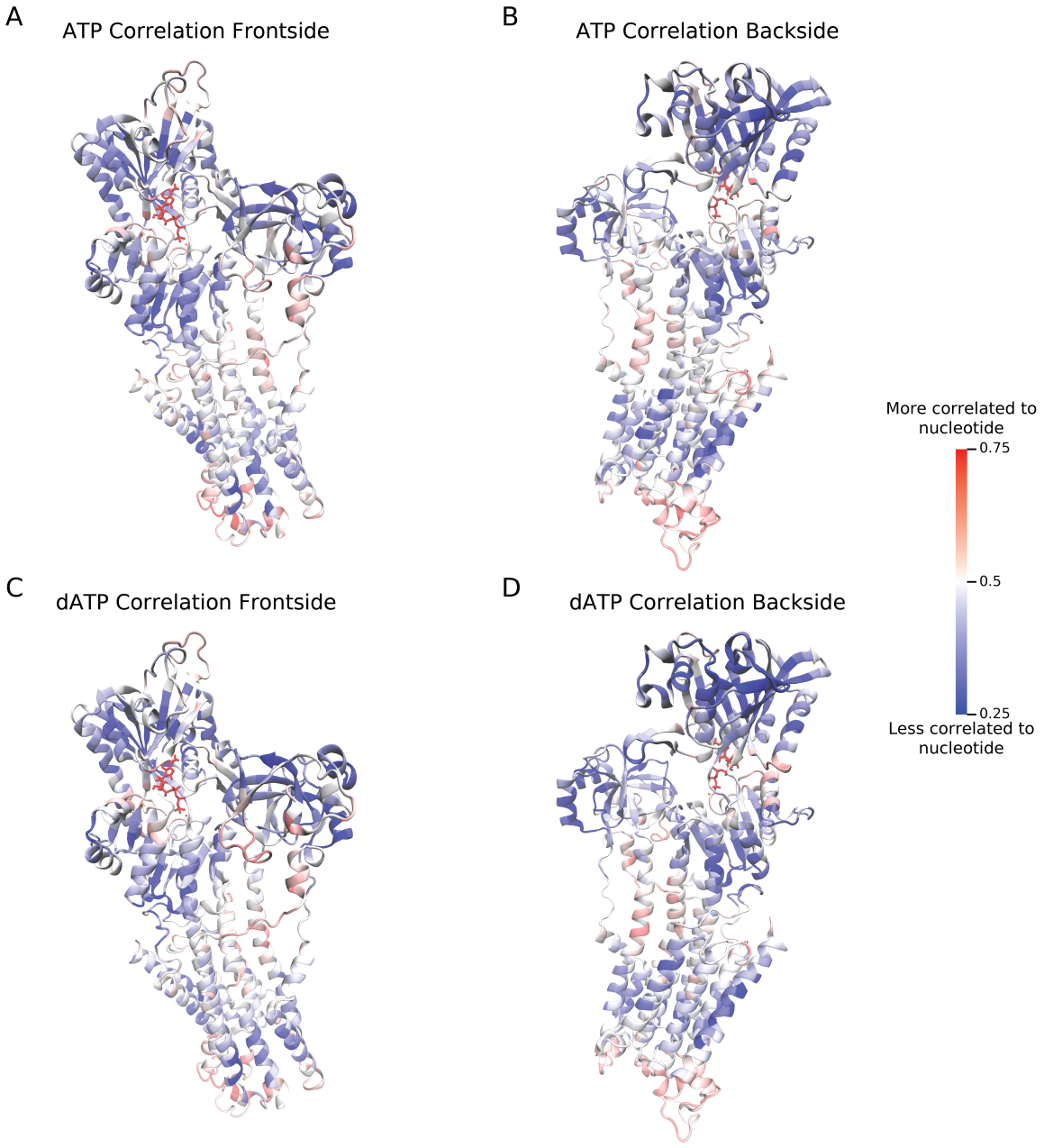


**Figure 2.4.** The SERCA pump structure, depicted in cartoon representation, bound to ATP (A, B) or dATP (C, D). The residues of the protein are colored by GC between the residue's center of mass and the nucleotide center of mass. Correlations within the protein span a continuum from values indicating low correlation (blue) to areas of high correlation (red) to ATP or dATP.

sites Fig. 2.4A and B as denoted by the white and blue coloring of the cartoon ribbon diagram in the lower third of the intermembrane domain. We see a similar low amount of correlation when dATP is bound as well (Fig. 2.4C and D. Relatively high correlated motion to ATP and dATP is observed in the core regions of the transmembrane domain above the calcium binding sites, as colored by the pink regions.

Relatively low correlated motion is observed between ATP and the most of the P, N, and A domains. There is fairly high correlation at the lumenal ends of the transmembrane helices, which may be important to regulate motion in order to prevent or allow ion flow. There is also a conspicuous "pathway" of high correlation that extends down the back side of the SERCA pump (Fig. 2.4B), which may function to propagate the allosteric signal between the nucleotide binding site and the calcium binding site. The correlation between Asp351 and the nucleotides is moderately high, with values of 0.536 for ATP and 0.579 for dATP.This makes the bound nucleotide (ATP or dATP) one of the most correlated residues to this autophosphorylation site. Surprisingly though, the bound $Mg^{2+}$ ion shows a stronger correlation to Asp351 than either nucleotide does to Asp351. Furthermore, there is an increase in correlation between Asp351 and $Mg^{2+}$ when transitioning from the *apo* state to either dATP or ATP bound. We see an increase in $r_{MI}$ from 0.523 to 0.600 and 0.637 for ATP and dATP respectively. This suggests that the binding of either nucleotide increases the strength of interaction to the $Mg^{2+}$ ion to the autophosphorylation site.

Upon the binding of ATP from the *apo* state, we observe a reduction in correlation in some of the calcium binding residues relative to the rest of the protein, specifically residues D800, V304, T799, and E908 (Fig 2.5). The remaining calcium coordinates residues do no show any obvious change in correlation relative to the rest of the protein. Alternatively, upon dATP binding from *apo*, residues N796, A305, and E309 increase in correlated motion. Again, the other calcium residues do not show any visible difference in correlation. When directly comprising the correlation between ATP or dATP and the calcium coordinating residues, we found consistently greater correlation between dATP and the residues compared to dATP. As

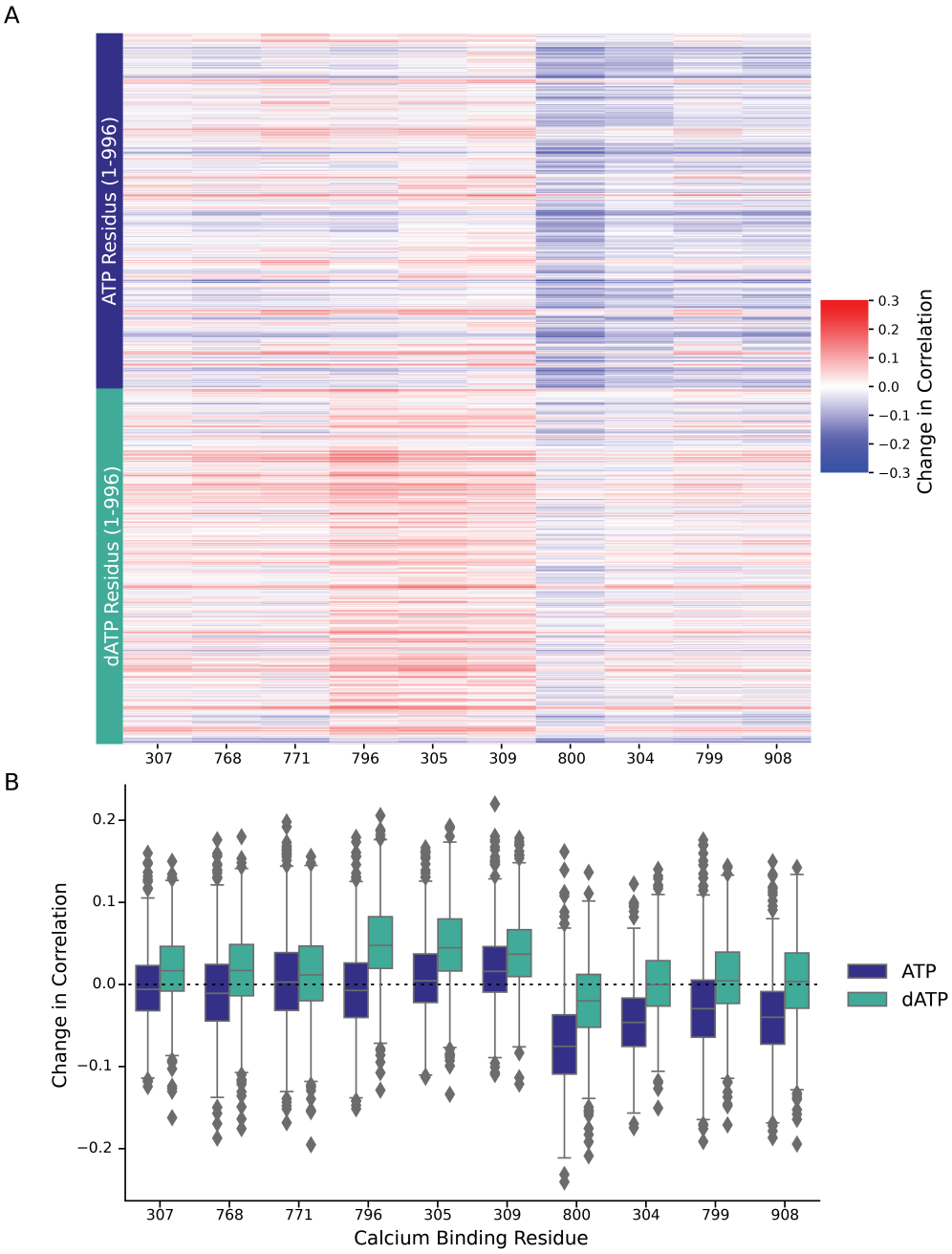


**Figure 2.5.** (A) Change in measured correlation ($r_{MI}$) from apo to nucleotide condition of residues in involved with calcium coordination relative to the rest of the protein shown as a heatmap. Calculated as $r_{MI}$(ATP) - $r_{MI}$(apo) or $r_{MI}$(dATP) - $r_{MI}$(apo). Positive (red) indicates an increase in measured correlation upon the nucleotide being bound, and negative (blue) indicates a decrease in measured correlation upon nucleotide binding. Top portion is for the ATP simulations and bottom portion for dATP simulations. (B) Distributions of the measured correlation separated by simulation type.

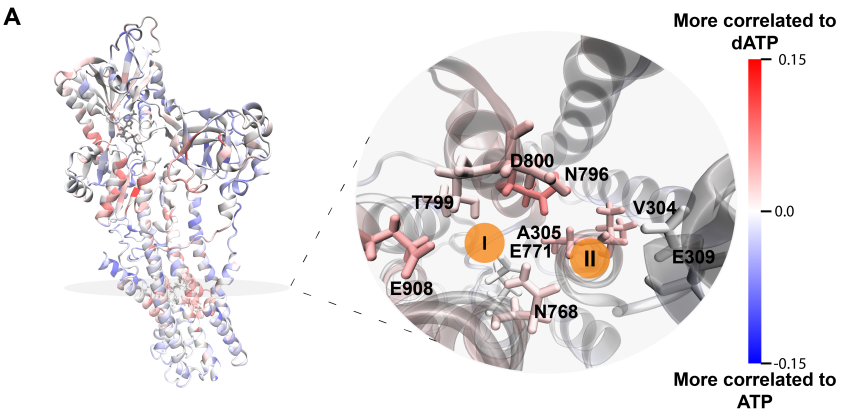


**Figure 2.6.** The net change in correlation to the nucleotide when ATP is replaced by dATP as calculated by $r_{MI,dATP} - r_{MI,ATP}$. Blue indicates greater correlation to ATP and red indicates greater correlation to dATP. The calcium handling residues around site I and II (orange) are shown from a top–down slice across the protein indicating most residues have greater correlation to dATP.

seen in Fig. 2.6 which shows the stronger correlation between the residues in site I and site II to dATP.

In addition to analysis of the correlated motions of every alpha carbon to ATP, we also used Netsci to compute the pairwise correlated motion of residue center of mass to every other residue center of mass. The raw correlated motion plots in Fig. 2.7A-C resemble the raw GC plot for protein G in Fig. 2.3. Several "blocks" or "bands" of high pairwise correlated motions can be found throughout the plot. Most of these, as in the case of protein G, are likely caused by the close proximity of rigid substructures within the protein, such as secondary structure elements. Specifically, in regions of the transmembrane domain, the highly ordered alpha helices appear highly correlated (e.g. residues 800 - 994). The most interesting information that one can obtain from a pairwise correlated motion analysis, however, is the observation of the change of correlated motion upon binding of ATP from *apo*, binding of dATP from *apo*, or the exchange from ATP to dATP (Fig. 2.7D-F).

First, we will consider the effect of binding ATP to the *apo* system. By careful analysis of the "bands" in figure 2.7D, we can see that an overall loss of correlation occurs in many regions of the protein - which implies a sort of "loosening" of motion, or an "unfreezing" of the protein

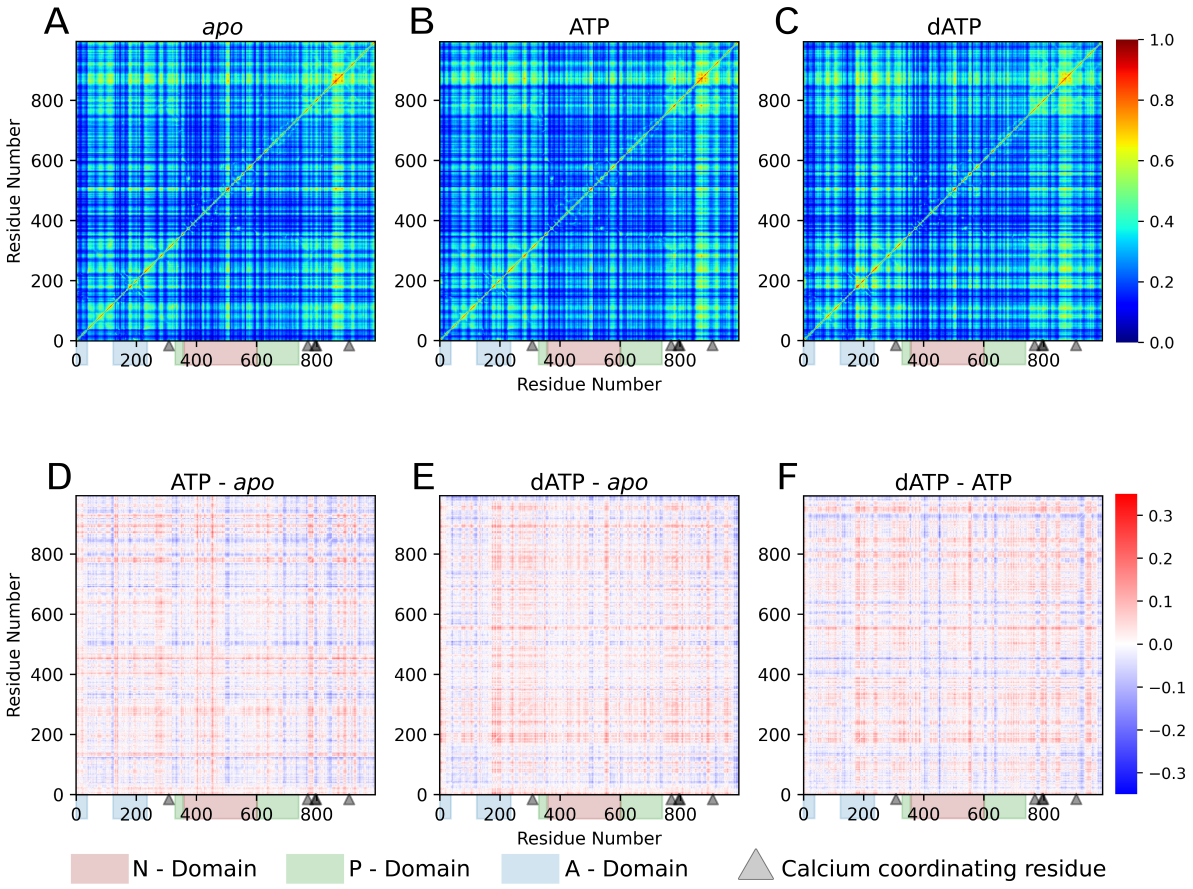


**Figure 2.7.** Absolute and relative pairwise correlated motion between every pair of alpha carbons within the SERCA pump. The top row shows the raw correlated motions for the *apo* system (A), the ATP-bound system (B), and the dATP-bound system (C). The lower row shows the relative change in correlation going from the *apo* to the ATP-bound state (D), from the *apo* to the dATP-bound state (E), and the ATP-bound to dATP-bound state (F). Regions of the cytosolic domain are labeled based on their residue spans.

itself. Among the regions that experienced an overall loss of correlation upon ATP binding to the *apo* structure, the A and N cytosolic domains, which are fairly close to and encompass the ATP binding site itself, experienced such a "loosening". This finding agrees well with previous results showing an increase in A-N domain distance with ATP binding [164]. Additionally, and significantly, the transmembrane region close to the calcium binding domain experienced a "loosening" phenomenon upon ATP binding, a result that corroborates our observation in Fig. 2.4, where a low correlated motion was also observed between the ATP and the calcium-binding residues. In contrast, some of the "bands" indicate that several regions gained correlated motion upon ATP binding, implying a sort of "freezing up" or "stiffening" of regions of the protein. Among the regions that "stiffened", a few parts of the A cytosolic domain, as well as a few residues (556 to 563) near the back of the ATP binding site, were included. Additionally, a few parts of the P cytosolic domain, as well as the lumenal domain of the transmembrane region. In the case of the "stiffening" of the lumenal domain, our data supports the idea of a "disorder gradient" from a high amount of disordered motion within the calcium-binding region, gradually shifting towards a region of low disorder in the lumenal portion, where the calcium is released. We hypothesize that this "disorder gradient" provides the mechanism that promotes the traversal of calcium across the membrane, thereby contributing to the pumping mechanism of SERCA.

**Network and graph theory analysis of SERCA**

Given these global observations of correlated motions, we now employ the graph theory and network analysis to further characterize SERCA behavior. The constructed graphs for each of the simulations have 996 nodes, including the nucleotide and magnesium ion, or 995 nodes for the *apo* simulation case without a bound nucleotide. A shortest distance pathfinding algorithm was then utilized to find the pathways of communication between the nucleotide and calcium binding coordinating residues in the transmembrane helices. The definition of the adjacency matrix used to construct prioritizes both nearby residues as well as high degrees of correlation, leading to a smaller network weight between nodes.

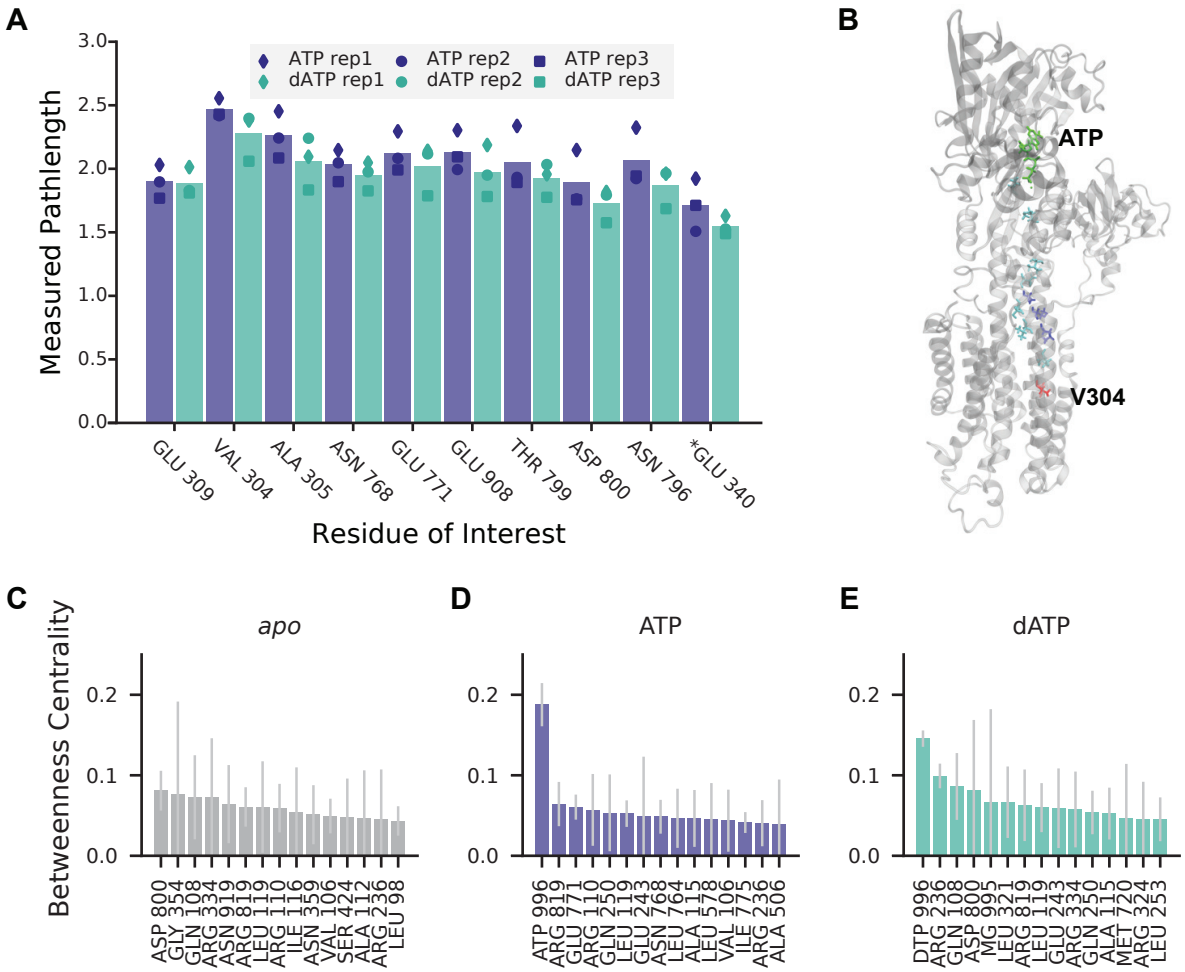


**Figure 2.8.** Network analysis comparing the information flow via correlated motions between ATP bound and dATP bound SERCA. A) The average computed effective path length from the nucleotide (ATP or dATP) to calcium coordinating residues. Each of the triplicate MD simulations and subsequent network analyses are shown as a different symbol (square, diamond and circle). B) An example visualization of the pathway from the nucleotide (colored green) to VAL304 (red) with the residues along the path shown in blue for ATP, and teal for dATP. C-E) Network analysis measuring the betweenness centrality, which represents the frequency that each node is visited if every possible node traversal is computed (top 15 residues shown). Standard deviations of trilplicate analysis plotted as gray error bars.

In order for SERCA to undergo catalytic activity, both calcium and the nucleotide must be bound to the protein [99]. Therefore, understanding the communication pathway between the calcium binding sites and nucleotide is essential to understanding the allosteric mechanisms at work. We computed the shortest path from either ATP or dATP and found that for eight of the nine pathways, the path length is shorter for dATP bound network compared to the ATP bound (Fig. 2.8A). The one exception is GLU309 (which functions as a gating residue [165]), which is extremely close, and likely within the expected error of the path length measurements. However, surprisingly, while almost all of the measured path lengths according to the defined weights were shorter for dATP than ATP, the number of nodes visited along each path was greater for all dATP network paths. Although the individual correlation between ATP and the rest of the SERCA protein tended to be greater than for dATP and the rest of the protein, the intra-protein correlations are higher for dATP, and this is reflected in the shorter measured pathways from nucleotide to residues around Site I and II. We also measured the path length from the nucleotide to GLU340 which has been studied extensively and known to be crucial for interdomain communication and found that dATP also has a shorter measured path than ATP to GLU340.

We also carried out non residue-specific global pathway analysis of protein correlation and information flow. Betweenness centrality of each node, which determines how frequently a node appears in the shortest path between other nodes in the network, estimates the significance of a residue in protein communication (figure 2.8C and D). Unsurprisingly, both ATP and dATP had the greatest betweenness centrality, likely acting as a linker between the N and P domains. However, dATP had a much lesser measure of betweenness compared to ATP. Arg236, which lies in the A domain, just at the start of the linkage to helix 3 had nearly the same measure of betweenness in the dATP simulation. This is contrary to the ATP simulation where the protein residues with the greatest degree of betweenness was only one third of ATP. The global pathway results reveal that there is a more uniform distribution of high visit residues for the dATP network as compared to ATP, which dominates the distribution and drops off sharply thereafter. We propose that as dATP promotes the headpiece closure, and induces a more highly correlated

structure, it allows for more efficient communication overall, therefore relying less on ATP as the primary residue for information transduction.

## 2.5 Conclusions

In this study, we developed and applied Netsci - a fast implementation k-nearest-neighbor-based algorithm that utilizes GPU acceleration for the estimation of MI within sets of data. We also present the use of the ”local alignment”, which was useful for reducing spurious correlations in large or flexible biomolecular systems. We validated Netsci by estimating the MI of a 2D Gaussian distribution - a situation for which results may be directly compared to the analytic solution. We also repeated a MI and GC analysis on the B1 domain of protein G.

We also applied Netsci towards a detailed analysis of allostery within the SERCA pump. A loss of correlated motion within the calcium binding site upon ATP and dATP binding constituted a potentially interesting observation, perhaps suggesting a mechanism for forced calcium traversal across the membrane through the use of a “disorder gradient.” Additionally, pathway analyses of information travel through the protein suggest routes through which the allosteric signal is propagated from the bound nucleotide to the calcium binding domain. The differential effects of ATP vs. dATP binding have also been characterized - where ATP shows a “stiffening” of portions of the A and P domains of the SERCA pump, and pathway analyses indicate that ATP functions as a sort of information bottleneck through which communication of information across the protein tends to flow. In contrast, dATP binding shows more of general “stiffening” of many different domains, facilitating the traversal of information across the protein, but not through a bottleneck. Our novel local alignment approach also served to remove spurious systematic bias in the MI estimation caused by global protein “breathing” motions.

While we specifically applied this algorithm to MD simulation of a biomolecule, to explain the allosteric communication within domains of proteins, Netsci is not limited to this purpose and may be used to estimate MI and GC for essentially any data application.

## 2.6 Materials and Methods

### 2.6.1 Protein G molecular dynamics

Protein G simulations were prepared and run in a manner as similar as possible to a previously published work [59]. A crystal structure for Protein G was obtained from the Protein Data Bank indexed as structure 1PGB. Using Gromacs software, the protein was solvated with a cubic box of TIP4PEW waters with a margin of at least 1 nm in the X, Y, and Z directions. Sodium ions were added to neutralize the system charge. The periodic system was modeled using Particle Mesh Ewald with a nonbonded cutoff of 1 nm, and all bonds were constrained to their equilibrium lengths for simulation. Using OpenMM, the system was minimized, and then simulated at a constant temperature of 300 K, and a constant pressure of 1 ATM with a Langevin integrator with a damping constant of 0.1/ps and a timestep of 2 fs. Equilibration was performed for 5 ns, followed by production simulations of 195 ns. During the production phase, trajectory frames were saved at a frequency to generate 19500 frames. Following the local alignment procedure described above, with a cutoff of 0.75 nm defining all adjacent residues, The MI of all pairs of alpha carbons was then performed using Netsci with a k-nearest-neighbors k value of 6.

### 2.6.2 SERCA pump molecular dynamics

All atom molecular dynamics simulations were performed with AMBER [166] starting with the crystal structure of SERCA1A in the E1.Mg2+ state from *Oryctolagus cuniculus* (PDB: 3W5A) [98]. In preparation for simulation, all ligands and ions were removed from the 3W5A structure, and ATP or dATP were manually docked in the nucleotide binding pocket based on the ATP position in 7BT2, which has the nucleotide bound [99]. Protein protonation state was determined based on PropKa at a pH of 7 [106, 107]. SERCA was oriented for placement into a lipid bilayer according to the Orientations of Proteins in Membranes (OPM) database [109]. CHARMM-GUI was used to build the membrane and solvate the system in a 12.5 by 12.5 nm lipid bilayer composed of POPC (51% upper leaflet, 66% lower leaflet), POPE (43% upper leaflet,

17% lower leaflet), and POPS (6% upper leaflet, 17% lower leaflet) experimentally determined membrane composition in the cardiac SR [110]. The water box padding was set to 22.5 nm and 150 mM $K^+$ $Cl^-$ ions were added according to Monte Carlo placement. The dATP structure was constructed by removing the extra hydroxyl from the ATP structure in 7BT2. Both ligands were parameterized using antechamber to generate AMBER GAFF2 force field parameters [114, 115]. The proteins and lipids were parameterized according to AMBERFF19FSB and Lipid17 with OPC water molecules [116, 117]. All molecular dynamics simulations were carried out on the Triton Shared Computing Cluster (TSCC). The production runs for each SERCA condition (*apo*, ATP, and dATP) were each 312 ns. Prior to the production simulation, minimization was carried out over 5000 steps of steepest descent minimization with 10 kcal $mol^{-1}$ $Å^{-2}$ positional restraints on all protein atoms and 2.5 kcal $mol^{-1}$ $Å^{-2}$ positional restraints on all lipid atoms, with NMR restraints. A six step equilibration approach was carried out for 1.875 ns total. During the first two steps, an NVT Langevin equilibration scheme was applied where 10 kcal $mol^{-1}$ $Å^{-2}$ positional restraints were present on all protein atoms for the first step and 5 kcal $mol^{-1}$ $Å^{-2}$ positional restraints were present on all protein atoms for the second step. Lipids were restrained using 2.5 kcal $mol^{-1}$ $Å^{-2}$ positional restraints with NMR restraints were present on all lipid atoms for both steps. The next four stages used an NPT ensemble for 125 ps, 500 ps, 500 ps, and 500 ps, respectively, with the system set to a pressure of 1.0 bar. All atom protein restraints set decreased through these stages and were specified at 2.5, 1.0, 0.5, and 0.1 kcal $mol^{-1}$ $Å^{-2}$ respectively. Similarly, positional and restraints on all lipid atoms were 1.0, 0.5, 0.1, and 0 kcal $mol^{-1}$ $Å^{-2}$ for each step, respectively, with NMR restraints. Triplicate production MD simulations were run at 303.15 K using the PMEMD (Particle Mesh Ewald Molecular Dynamics) method with a 9 Å nonbonded cutoff, and 2 fs timestep. The first triplicate was ran for 312.25 ns coordinates saved every 5 ps leading to 62,450 frames, while the second and third triplicate were each ran for 300 ns with 60,000 frames per simulation for replicates 2 and 3. The mutual information algorithm was applied to the first triplicate with all 62,450 frames, as well as just the first 60,000 frame with no significant change. Therefore we selected to use all available

data and include the full trajectory. Images were generated using VMD[121, 167].

## 2.7 Acknowledgements

Chapter 2, in part, has been submitted as it may appear in the *Journal of Chemical Information and Modeling* 2024. NetSci: A Library for High Performance Biomolecular Simulation Network Analysis Computation. A. M. Stokely, L. W. Votapka, M. T. Hock, A. E. Teitgen, J. A. McCammon, A. D. McCulloch, R. E. Amaro. The dissertation author and carried out the biomolecular simulation and analysis of SERCA.

# Chapter 3

# Multiscale modeling shows how 2'-deoxy-ATP rescues ventricular function in heart failure

## 3.1 Abstract

2'-deoxy-ATP (dATP) improves cardiac function by increasing the rate of crossbridge cycling and $Ca^{2+}$ transient decay. However, the mechanisms of these effects and how therapeutic responses to dATP are achieved when dATP is only a small fraction of the total ATP pool remain poorly understood. Here, we used a novel multiscale computational modeling approach to analyze the mechanisms by which dATP improves ventricular function. We integrated atomistic simulations of pre-powerstroke myosin and actomyosin association, filament-scale Markov state modeling of sarcomere mechanics, cell-scale analysis of myocyte $Ca^{2+}$ dynamics and contraction, organ-scale modeling of biventricular mechanoenergetics, and systems level modeling of circulatory dynamics. Molecular and Brownian dynamics simulations showed that dATP increases the actomyosin association rate by 1.9 fold *via* stabilization of pre-powerstroke myosin. Markov state models predicted that dATP also increases the pool of myosin heads available for crossbridge cycling, increasing steady state force development at low dATP fractions by 1.3 fold due to mechanosensing and nearest-neighbor cooperativity. This was shown to be the primary mechanism by which dATP improves contractile function at all scales. Together

with faster myocyte $Ca^{2+}$ handling, this led to improved ventricular contractility, especially in a failing heart model in which dATP increased ejection fraction by 16% and the energy efficiency of cardiac contraction by 1%. This work represents a complete multiscale model analysis of a small molecule myosin modulator from single molecule to organ system biophysics, and elucidates how the molecular mechanisms of dATP may improve cardiovascular function in heart failure with reduced ejection fraction.

## 3.2 Introduction

The naturally occurring nucleotide 2'-deoxy-ATP (dATP) is a candidate myosin activator that has shown promise for treating heart failure with reduced ejection fraction (HFrEF) [50]. dATP has been shown to increase force production in skinned myocardium, to increase shortening in isolated cardiomyocytes, and to improve ventricular function in animal models [50, 32, 87, 39, 36, 89, 40, 38, 88]. How levels of dATP as low as 1-2% of the ATP pool significantly improve muscle contraction is not well-understood [38, 88]. In addition to increasing the rate of crossbridge cycling, recent studies in cardiac and skeletal muscle suggest that dATP may also increase the pool of myosin available for crossbridge cycling [168, 169, 37, 170]. We hypothesize that dATP alters the recruitment dynamics of myosin from the thick filament backbone to more active states and that nearest-neighbor cooperativity leads to increased cardiac muscle force development even when dATP is only a small fraction of total adenine nucleotide content in the cell. Further, our recent simulations suggest that dATP increases SERCA pump function, leading to faster $Ca^{2+}$ transient decay [164], which has also been observed previously experimentally [39, 38]. This may contribute to faster relaxation of cardiomyocytes post-contraction. It is unclear how the these distinct molecular and cellular mechanisms of dATP integrate into improved ventricular pump function, especially in HFrEF when energy metabolism is typically impaired [171, 172].

Here, we use a novel combination of multiscale computational models to simulate

the effects of dATP on myosin dynamics, using molecular dynamics (MD) and rigid body Brownian dynamics (BD). We investigate the resulting effects on sarcomere kinetics using a spatially explicit Markov state sarcomere model. We then assess how these mechanisms interact with altered myocyte $Ca^{2+}$ handling to enhance contractility and lusitropy, using a model of cardiomyocyte mechanics and $Ca^{2+}$ dynamics. Finally, we assess how these myocyte responses contribute to observed improvements in left ventricular mechanoenergetics and hemodynamics in the normal and failing heart by incorporating this cardiomyocyte model into a biventricular mechanics and lumped-parameter circulatory system model. This comprehensive multiscale model analysis of the heart can be used to predict organ system scale cardiovascular function from atomic resolution simulations of molecular mechanisms and shows how very low fractions of dATP are able to significantly improve pump function and efficiency in the failing heart. Our modeling approach may additionally be a useful tool to study other sarcomere-targeted small molecule activators and inhibitors.

## 3.3 Results

### 3.3.1 dATP alters pre-powerstroke myosin dynamics, increasing its affinity for actin

A combination of molecular modeling techniques were utilized to assess how dATP and ATP differentially affect the pre-powerstroke conformation of myosin at a molecular level. Three molecular dynamics (MD) simulations of 2 $\mu$s each for ATP-myosin and dATP-myosin were carried out (Fig. 3.1A), and featurization analysis of these MD trajectories was carried out based on structural loops and motifs associated with actin binding based on experimental studies [173, 174, 175] (Fig. 3.2A). Time-lagged Independent Component Analysis (tICA) was employed for further dimensionality reduction, and was utilized to construct three-state Markov state models (MSM) for ATP-myosin and dATP-myosin to capture their major conformational dynamics (Fig. 3.2B,C, Fig. 3.8 and 3.9). Representative conformations from the three states are

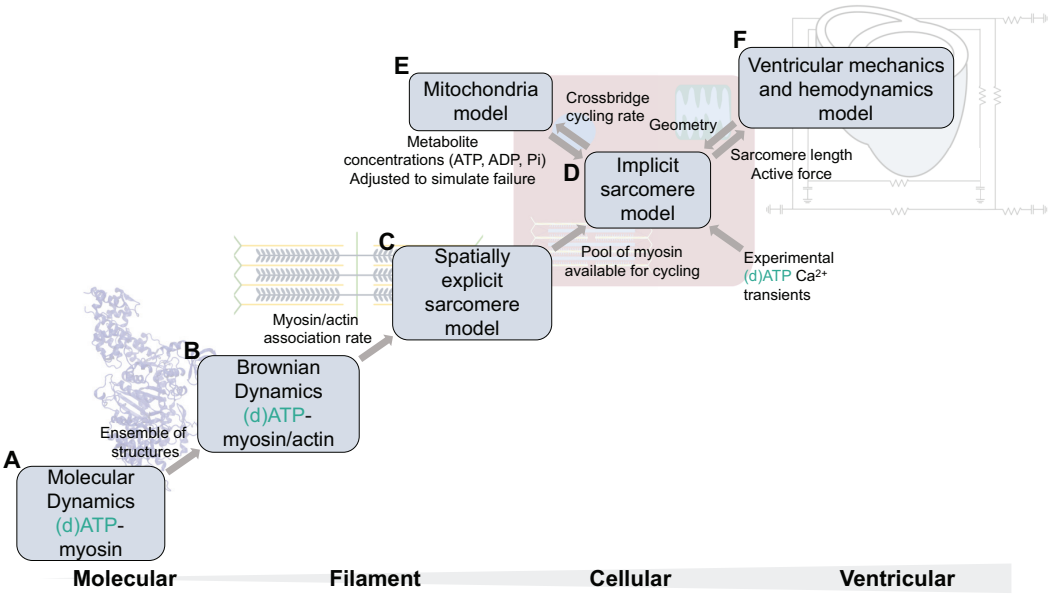


**Figure 3.1.** Multiscale modeling overview. Gray arrows indicate coupling between models. MD simulations of ATP-myosin and dATP-myosin binding to actin (A) in combination with BD simulations (B) were utilized to determine myosin.actin association rate, which was used to constrain a spatially explicit model of cooperative sarcomere mechanics (C). The effects of dATP on myosin predicted by this model were extended to a myocyte model containing an implicit sarcomere mechanics model (D), which is driven by experimental $Ca^{2+}$ data, and is coupled to a mitochondrial energetics model (E). The myocyte model ((D), (E), and experimental $Ca^{2+}$ data) is embedded within a biventricular mechanics and hemodynamics model of the failing heart (F).

shown in Fig. 3.2G, H. The stationary distribution of the ATP MSM for states 0, 1 and 2 was 34.1%, 40.7% 25.2%, respectively. However, for the dATP MSM, the stationary distribution was 5.00% , 8.17%, 86.8% for states 0, 1 and 2, respectively, suggesting that dATP-myosin has a lower probability of transitioning to a new conformational sub-state. Further, based on mean first passage time (MFPT) anaysis of the transition times between metastable states, the dATP model shows much more rapid transitions into state 2 as the dominant state, whereas the ATP model has much more balanced kinetic transitions between all of the states (Fig. 3.2E, F). We further found that the overall root mean square fluctuation (RMSF) of the protein was lower for dATP-myosin than ATP-myosin (Fig. 3.2D). Together, these results suggest that dATP may stabilize pre-powerstroke myosin within the context of our MD simulations. However, it should be noted that this MD-MSM method does not explicitly demonstrate increased stability of the pre-powerstroke biochemical state, since these simulations do not capture transition out of the pre-powerstroke state.

An ensemble-based approach was used to carry out rigid body BD simulations (Fig. 3.1B) of actomyosin association, with conformations sampled from the metstastable states of the MSM. dATP-myosin showed a significantly higher ensemble-averaged crossbridge formation rate than ATP-myosin at all simulated reaction distances, where reaction distance is a parameter in the simulation which defines the distance at which the two molecules are considered to bind (Fig. 3.8). The previously reported ATP-myosin.actin association rate of $2.50\times10^6\ M^{-1}s^{-1}$ based on experimental measurements [42, 176, 177] corresponds to a reaction distance of 7.17 Å (indicated by the dashed vertical line in Fig. 3.2I). Using this reaction distance, the predicted association rate for dATP-myosin.actin was $4.78\times10^6\ M^{-1}s^{-1}$, a 1.9-fold increase over ATP-myosin.actin. For reaction distances between 6.5 Å and 10 Å, association rates were 1.54 to 2.13 fold greater for dATP than ATP.

### 3.3.2 Increased force-dependent recruitment of myosin and nearest-neighbor cooperativity explain significantly increased steady state tension development with low fractions of dATP

We next assessed how the predicted increase in actomyosin association affects sarcomere mechanics. Several studies have shown that fractions of dATP as low as 1-2% of the ATP pool are sufficient to significantly increase contractile force and shortening [38, 88, 169]. We utilized a spatially explicit Markov state sarcomere model to assess whether increased actomyosin association could explain these effects with 1% dATP [42, 178] (Fig. 3.1C). After prescribing the ATP-myosin.actin association rate ($k_f^+ = 2.50\times10^6\ M^{-1}s^{-1}$) based on reported measurements [42, 176, 177], increasing the actomyosin association rate to $4.78\times10^6\ M^{-1}s^{-1}$ as predicted for dATP-myosin.actin by the BD simulations resulted in an 3% increase in overall sarcomere steady state force at maximal $Ca^{2+}$ activation (Fig. 3.3B).

A previous study in our group showed using this sarcomere model that dATP increases the crossbridge cycling rate (the powerstroke and crossbridge detachment rates, specifically) in addition to increasing the actomyosin association rate, and that these parameter changes were sufficient to explain experimental changes for simulations with 100% dATP [42]. However, increasing parameters $k_p^+$ (powerstroke rate) and $k_g^+$ (detachment rate) in addition to $k_f^+$ by the same amount as in [42] resulted in an 8% increase in steady state force with 1% dATP. Experimentally, 100% dATP was shown to increase maximum steady state force by 31% in demembranated rat cardiac trabeculae [87]. Experimental data on changes in steady state force with 1% dATP are not available, but given data showing that small fractions of dATP are sufficient to significantly increase force production, we would expect a larger increase in force with 1% dATP. This suggests that although our previous modeling results were sufficient to explain the effects of 100% dATP, additional mechanisms must be considered for small fractions of dATP (Fig. 3.3B).

We then modified the model to include the active (ON) and inactive (OFF) states of myosin. Transition between these ON and OFF states is governed by parameters $k_m^+$ and $k_m^-$,

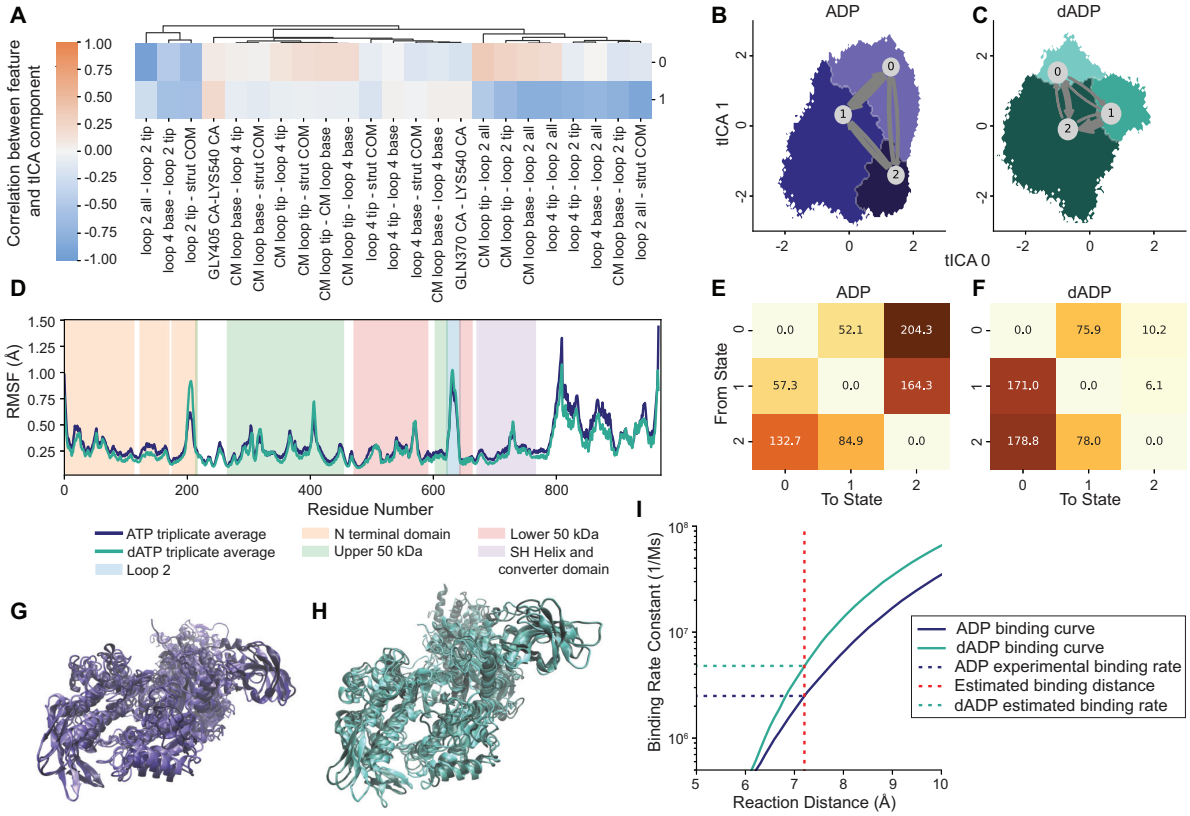


**Figure 3.2.** MD simulations and MSM demonstrate that binding of dATP may stabilize the pre-powerstroke myosin head compared with binding of ATP, increasing its affinity for actin. (A) Correlation analysis between input features from MD simulations (distances between key structural features on myosin) and first (0) and second (1) tICA components. Clustered with 'City-Block' metric. Center of mass is abbreviated as COM, and alpha carbons are abbreviated as CA. (B)-(C) tICA space visualization of MD simulations, with three metastable states shown for each MSM based on first and second tICA components. Arrows represent flux between states. (D) Root mean square fluctuation (RMSF) shown for ATP and dATP, averaged across three MD trajectories for each. Regions of interest on myosin are highlighted. (E) Mean first passage times between metastable states shown in (B) of ATP-bound myosin simulations (ns). (F) Mean first passage times between metastable states shown in (C) of dATP-bound myosin simulations (ns). (G) Representative conformations from three metastable states for ATP-bound myosin. (H) Representative conformations from three metastable states for dATP-bound myosin. (I) Binding rate constant estimates of myosin binding to actin using BD simulations.

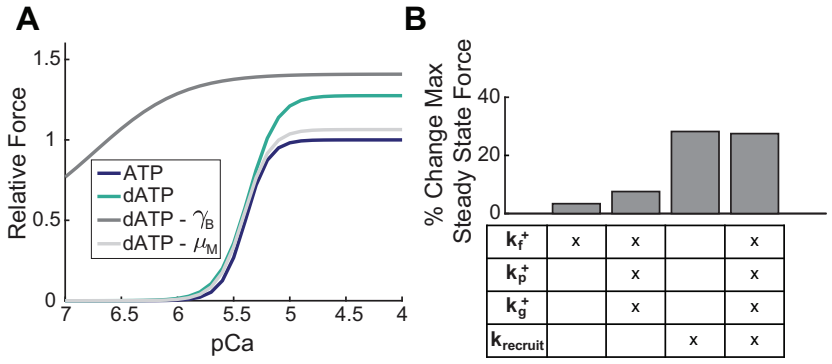


**Figure 3.3.** dATP increases the pool of myosin available for crossbridge cycling, which leads to disproportionate increases in force with 1% dATP. (A) Model-predicted force-pCa curves are shown for ATP (purple) and 1% dATP (teal). ATP curve was fit to experimental steady state force-pCa data from [87]. dATP simulation includes increases in actomyosin association rate ($k_f^+$), powerstroke rate ($k_p^+$), and detachment rate ($k_g^+$), as well as increased force-dependent recruitment of myosin ($k_{recruit}$). Effects of setting cooperative parameters $\gamma_B$ and $\mu_M$ to one, thus removing their effects from the model, are also shown. (B) Relative contributions of increased crossbridge binding and cycling and increased myosin recruitment to increases in maximum steady state force (at pCa 4.0) relative to ATP. Differences are expressed as percentages relative to ATP.

as well as $k_{recruit}$, which describes the force dependence of this transition [179, 33]. We found that increasing parameter $k_{recruit}$ from 0.2 $N^{-1}m^{-2}$ to 779 $N^{-1}m^{-2}$, in addition to $k_f^+$, $k_p^+$, and $k_g^+$, resulted in a 28% increase in steady state force with 1% dATP, which was the maximal increase in force that could be achieved and is close to the observed increase of 31% for 100% dATP (Fig. 3.3A, B, additional details found Table 3.1). $k_{recruit}$ was the only parameter in the model which could be increased to produce such a dramatic increase in steady state force. $k_{recruit}$ determines the force dependence of the recruitment of myosin from the thick filament backbone, and is regarded as a mechanism of thick filament mechanosensing [179]. Previous computational work [37] showed that dATP activates the resting conformation of cardiac myosin, and X-ray diffraction data [169] and fluorescent assays [170] showed decreases in the fraction of myosin heads in an ordered or low ATPase activity state (respectively) with increased dATP [168]. Further, X-ray diffraction data indicates that dATP treatment increases strain in the thick filament backbone, providing support for the idea that dATP leads to recruitment of additional

myosin heads via mechanosensing [169]. Therefore, our results suggest that dATP-mediated recruitment of myosin to state(s) that can contribute to contraction is the dominant mechanism by which it increases steady state force, especially at low dATP fractions.

We further assessed whether nearest-neighbor cooperativity could explain this increase in steady state force with increased myosin recruitment. After increasing $k_f^+$, $k_p^+$, $k_g^+$, and $k_{recruit}$, we set each of the cooperative parameters ($\gamma_B$, $\gamma_M$, and $\mu_M$) to one, thus eliminating their effects from the model, and assessed their relative impacts on maximum steady state force. We found that setting $\gamma_B$ to one resulted in increased steady state force (41% increase relative to ATP) but flattened the force-pCa curve due to loss of cooperativity *via* thin filament (tropomyosin) overlap (Fig. 3.3A). Setting $\mu_M$ to one resulted in reduced steady state force (6% increase relative to ATP) due to a loss of cooperativity between neighboring bound crossbridges (Fig. 3.3A).

These results support the hypothesis that by increasing the pool of myosin available for crossbridge cycling, a small fraction of dATP can have a disproportionate effect on sarcomere mechanics by promoting the formation of ATP-myosin.actin crossbridges via nearest-neighbor cooperative interactions, both from neighboring bound crossbridges and tropomyosin overlap.

### 3.3.3 Increased myosin recruitment and calcium sequestering dynamics are needed to explain improvements in myocyte contractility and lusitropy with elevated dATP

We next utilized an implicit model of sarcomere mechanics and cardiomyocyte $Ca^{2+}$ handling to extend these results to the whole myocyte level (Fig. 3.1D. This implicit model was chosen because it is less computationally expensive than the spatially explicit sarcomere model utilized above (Fig. 3.1C), and is more comprehensive in its inclusion of a viscoelastic model, $Ca^{2+}$ dynamics, and coupling to a mitochondria model (Fig. 3.1E) to allow for cell shortening and whole heart mechanoenergetics simulations (Fig. 3.1F). The implicit myocyte model utilized in this study contains an additional crossbridge cycling state (the weakly-bound state) compared with the spatially explicit model, and we thus adjusted slightly different crossbridge cycling

parameters to achieve the same effect in the implicit model ($k_f^+$, $k_f^-$, and $k_w^+$) as in the spatially explicit model ($k_f^+$, $k_p^+$, and $k_g^+$). More details on these models, including parameter selection, can be found in the Methods and tables 3.1 and 3.2.

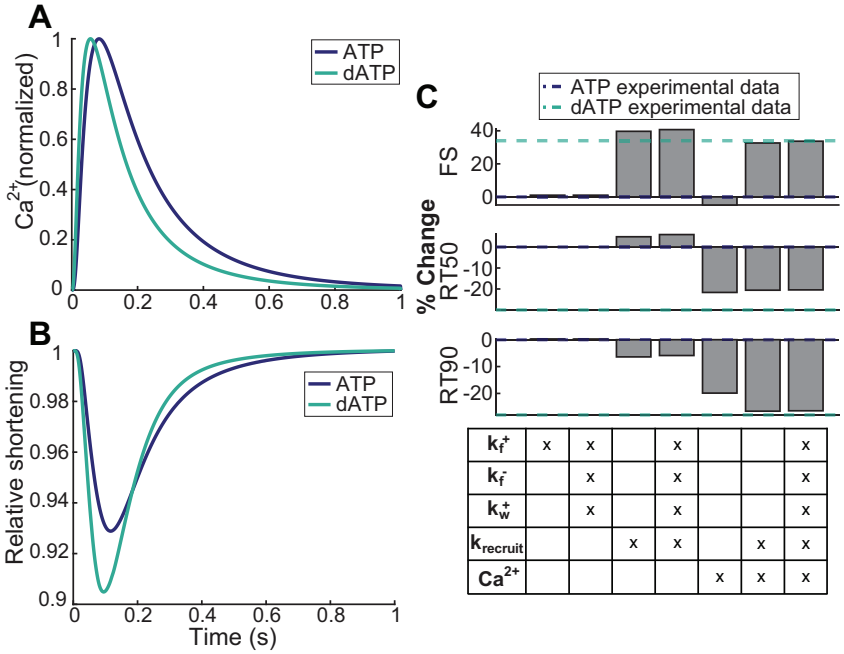


**Figure 3.4.** Increased myosin recruitment and $Ca^{2+}$ sequestering dynamics are needed to explain improvements in myocyte contractility and lusitropy with elevated dATP. (A): Model-simulated $Ca^{2+}$ transients for ATP (purple) and dATP (teal), based on average experimental data from [38, 39]. (B): Cell shortening simulations for ATP (purple) and 1% dATP (teal), including increased crossbridge binding (increasing $k_f^+$) and cycling (increasing $k_f^-$ and $k_w^+$), faster $Ca^{2+}$ dynamics (shown in (A)), and increased myosin recruitment (increasing $k_{recruit}$). (C): Relative contributions of increased crossbridge binding and cycling, faster $Ca^{2+}$ dynamics, and increased myosin recruitment to changes in FS, RT50, and RT90 compared with average experimental data from [38, 39]. Baseline experimental ATP values are shown as purple dashed lines, and experimental dATP values are shown as teal dashed lines. Differences are expressed as percentages relative to ATP.

We found that increasing the ATP-myosin.actin association rate, $k_f^+$, from $2.50\times10^6$ $M^{-1}s^{-1}$ to $4.78\times10^6$ $M^{-1}s^{-1}$ in the myocyte model based on the BD results resulted in only a 1% increase in fractional shortening (FS) with 1% dATP, consistent with our findings using the spatially explicit model (Fig. 3.4C). Similarly, increasing parameters $k_f^-$ (actomyosin detachment rate) and $k_w^+$ (weakly- to strongly-bound transition rate) in addition to $k_f^+$ did not further increase

FS (Fig. 3.4C). However, as was the case at the filament scale, we found that increasing parameter $k_{recruit}$ from 0.2 $N^{-1}m^{-2}$ to 37 $N^{-1}m^{-2}$ in addition to $k_f^+$, $k_f^-$, and $k_w^+$ resulted in a 41% increase in FS with 1% dATP (Fig. 3.4B), greater than the experimentally measured increase of 34%. Again, this was the only parameter which could produce this effect. This further supports the conclusion that dATP treatment leads to disproportionate increases in force with 1% dATP by disrupting the resting states of myosin, which outweighs the effects of increased crossbridge binding and cycling. However, increased recruitment of myosin with elevated dATP resulted in slowed time to 50% relaxation (RT50) and only explained 21% of the experimental change in time to 90% relaxation (RT90) [38, 39]. (Fig. 3.4C), so we next sought to assess additional factors that could explain these changes in relaxation.

When the effects of 1% dATP (99% ATP) on the $Ca^{2+}$ transient were simulated by prescribing the average experimental dATP $Ca^{2+}$ transient, which showed decreased time to 50% and 90% $Ca^{2+}$ transient decay (DT50 and DT90, respectively) relative to the 100% ATP transient [38, 39] (Fig. 3.4A), RT50 was decreased by 22% and RT90 was decreased by 20%, which is closer to the experimental data [38, 39] (Fig. 3.4C). When the more rapid $Ca^{2+}$ dynamics with elevated dATP were combined with increased myosin recruitment from resting states, as described above, these three mechanisms together explained 97% of the average experimental increase in FS, 70% of the experimental increase in RT50, and 96% of the experimental increase in RT90 (Fig. 3.4B, C). Additionally, including increased rates of crossbridge binding and cycling with elevated dATP in addition to increased myosin recruitment and faster $Ca^{2+}$ transient decay did not substantially change FS, RT50, or RT90 (Fig. 3.4C).

Therefore, our model predictions suggest that the integrative mechanisms of dATP on myosin recruitment and $Ca^{2+}$ sequestering dynamics can explain improved contractility and lusitropy with elevated dATP at the myocyte level, as well as the high sensitivity of cardiac muscle to small fractions of dATP.

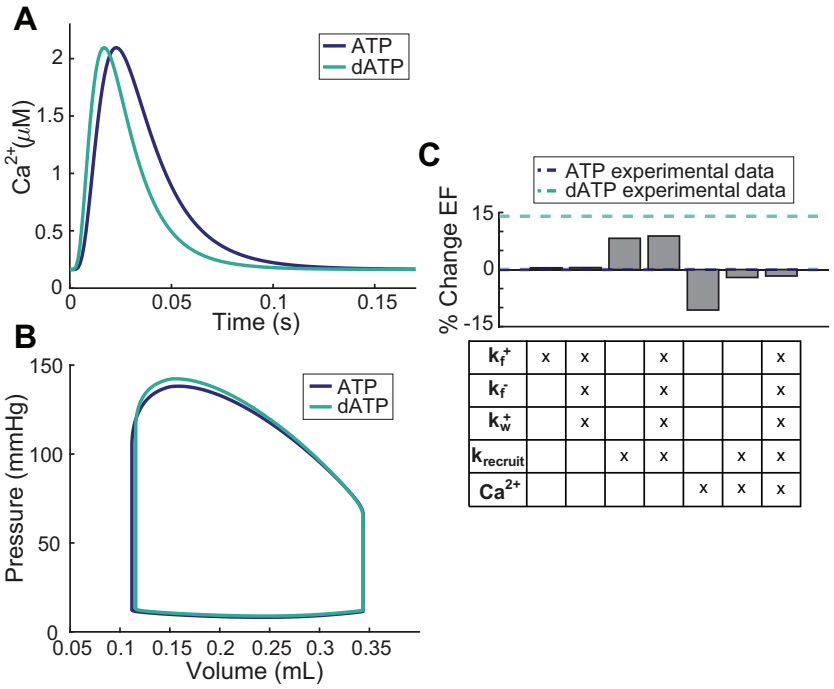


**Figure 3.5.** Increased myosin recruitment leads to improvements in ventricular contractility with elevated dATP. (A): Model-simulated average calcium transients for ATP (purple) and dATP (teal), based on experimental data from [38, 39] (B): Pressure volume loops for ATP (purple) and 1% dATP (teal), including increased crossbridge binding (increasing $k_f^+$) and cycling (increasing $k_f^-$ and $k_w^+$), faster $Ca^{2+}$ dynamics (shown in (A)), and increased myosin recruitment (increasing $k_{recruit}$). (C): Relative contributions of increased crossbridge binding and cycling, faster $Ca^{2+}$ dynamics, and increased myosin recruitment to changes in EF compared with experimental data from [39]. Baseline experimental ATP values are shown as purple dashed lines, and experimental dATP values are shown as teal dashed lines. Differences are expressed as percentages relative to ATP.

### 3.3.4 Increased myosin recruitment with elevated dATP contributes to improved ventricular mechanoenergetics

In animal models, the ventricular concentration of dATP has successfully been increased *via* upregulation of the enzyme ribonucleotide reductase (R1R2), which converts ADP to dADP [50]. dADP is then converted to dATP by the normal cellular rephosphorylation process. Elevated dATP has been observed to significantly increase left ventricular developed pressure (LVdevP), cardiac output (CO), and ejection fraction (EF) in transgenic mice over-expressing R1R2 and infarcted pig hearts treated with R1R2 *via* an adeno-associated viral vector *in vivo* [39, 40]. Therefore, we next utilized our whole heart and circulation model to assess how the predicted effects of dATP on sarcomere and $Ca^{2+}$ dynamics at the myocyte level extend to altered ventricular function. This model contains the same implicit sarcomere and $Ca^{2+}$ handling models utilized for myocyte level simulations (Fig. 3.1D), which are further coupled to a mitochondria model (Fig. 3.1E) and embedded within a biventricular mechanics and hemodynamics model (Fig. 3.1F), as described in the Methods.

After adjusting parameters $K_{SE}$, $k_{passive}$, $\eta$, $k_{on}$, $C_{Ao}$, and $Am_{ref}$ in the baseline ventricular model to match experimentally measured EF in mice [39] (all other parameters were kept the same as in the myocyte model), we simulated dATP treatment in the same way as in the myocyte shortening simulations (Table 3.2). We found that increasing $k_f^+$ from $2.50\times10^6$ $M^{-1}s^{-1}$ to $4.78\times10^6$ $M^{-1}s^{-1}$ in the ventricular model based on the BD results led to a $<1\%$ increase in EF, and increasing $k_f^-$ (actomyosin detachment rate), and $k_w^+$ (weakly- to strongly-bound transition rate) did not further increase in EF (Fig. 3.5C). Consistent with our results at the filament and myocyte scales, we found that increasing $k_{recruit}$ from 0.2 $N^{-1}m^{-2}$ to 37 $N^{-1}m^{-2}$ in addition to $k_f^+$, $k_f^-$, and $k_w^+$ led to a 9% increase in EF, which more closely matches experimental data which showed a 14% increase in EF with dATP [39] (Fig. 3.5C).

Interestingly, when we included faster $Ca^{2+}$ dynamics in the ventricular model (Fig. 3.5A, B, C) we observed reduced EF, contrary to our findings at the myocyte scale (Fig. 3.5C). However,

these findings are consistent with multiscale modeling results from [180], where increasing the $Ca^{2+}$ reuptake rate into the sarcoplasmic reticulum led to reduced EF. The combined effects of dATP on myosin recruitment and $Ca^{2+}$ sequestering dynamics led to an overall decrease in EF with 1% dATP, likely due to these $Ca^{2+}$ transient effects. However, these results support the conclusion that increased recruitment of myosin from the thick filament backbone is the primary mechanism by which dATP improves contractility.

### 3.3.5 Elevated dATP improves ventricular function in the failing heart in part due to improved energetic efficiency

To simulate HFrEF, the metabolite concentrations in the model were adjusted to mean values previously measured experimentally in failing rat hearts [73]. This resulted in reduced EF (Fig. 3.6A, C). Further, ATP and ADP concentrations were decreased and Pi concentrations were increased, consistent with [73] (Fig. 3.6G-I). We found that with 1% dATP in the failing heart model, EF increased by 16%, CO increased by 16%, and LVDevP increased by 13% (Fig. 3.6C-E). EF was returned closer to normal with just 1% dATP (61% vs 67% in the baseline healthy simulation), and was returned to 67% with 7% dATP (Fig. 3.6C). This aligns well with experimental data in pigs which showed a 16% increase in EF with dATP in failing hearts [40]. Therefore, the mechanisms identified at the filament (increased recruitment of myosin), and cellular (faster $Ca^{2+}$ dynamics) scales were sufficient to explain experimentally measured changes in EF with 1% dATP in the failing heart.

Overall, dATP improved ventricular function in a dose-dependent manner (Fig. 3.6). Further, our model was able to predict the effects of varying percentages of dATP on metabolite concentrations and energetic function in the failing heart. ATP levels were unchanged (suggesting that dATP treatment does not substantially deplete ATP pools), while ADP and Pi levels increased with increasing dATP ratio (Fig. 3.6G-I), as in [39]. The creatine phosphate (CrP)/ATP ratio was decreased slightly with increasing dATP ratio (Fig. 3.6K). Further, myocardial oxygen consumption ($MVO_2$) and ATPase rate also increased with increasing dATP ratio (Fig. 3.6J, L).

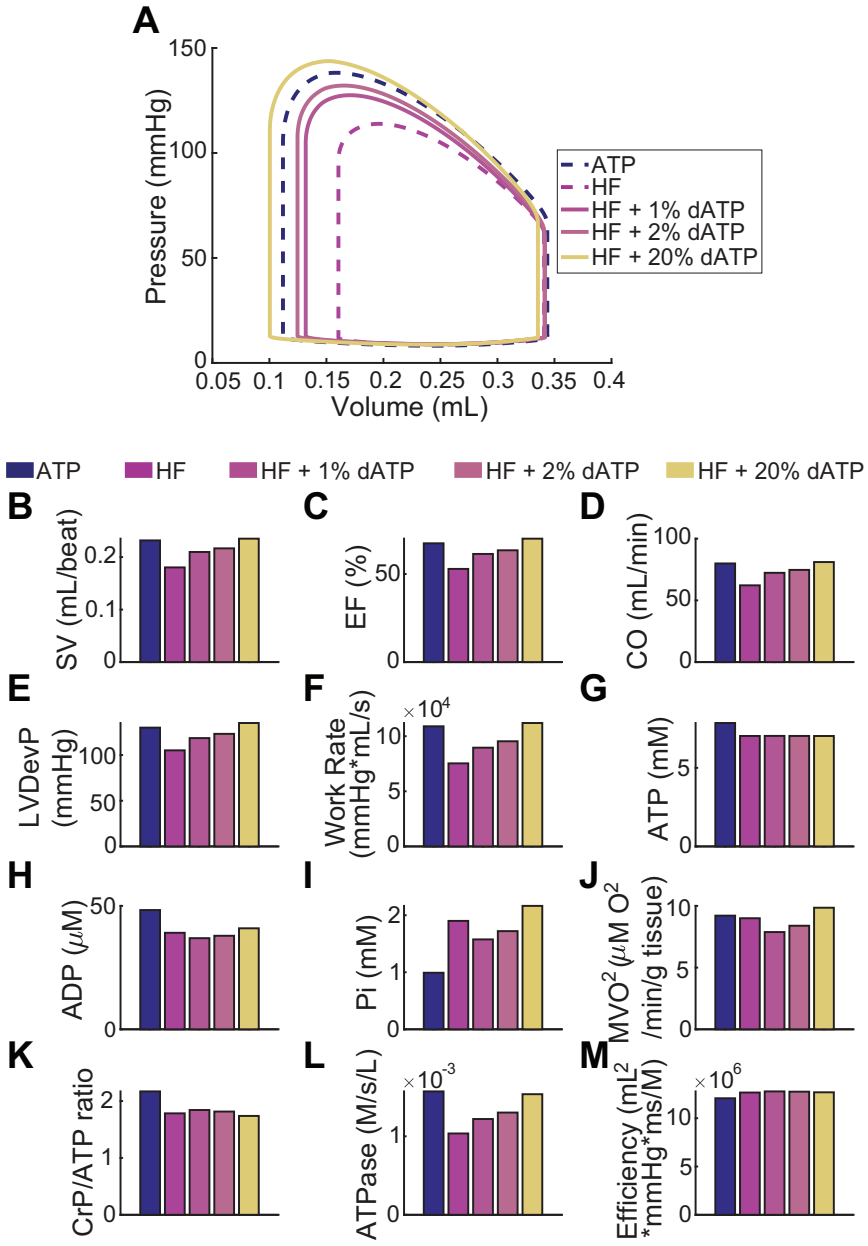


**Figure 3.6.** Elevated dATP leads to improved ventricular function in failing hearts with 1% dATP, and improves energetic efficiency. (A) Pressure-volume loops for varying ratios of dATP, including increased myosin recruitment (increasing $k_{recruit}$), crossbridge binding (increasing $k_f^+$) and cycling (increasing $k_f^-$ and $k_w^+$), and $Ca^{2+}$ sequestering dynamics with elevated dATP. Baseline ATP healthy heart simulation (fit to data from [39]) is shown as a purple dashed line. For HFrEF simulations, pink dashed line is baseline HFrEF simulation, yellow line is 20% dATP, and color gradient represents increasing ratios of dATP (1%, 2%, and 20%). Varying dATP percentages were simulated as described in the Methods. (B)-(M) Metrics of LV mechanical function and energetics *vs.* dATP ratio, for the same dATP percentages as in (A).

However, with 1% dATP these metrics remained below normal, non-failing levels. Interestingly, efficiency, defined as the work per beat divided by ATP hydrolysis rate, increased with increasing dATP ratio, and was increased by 1% with 1% dATP (Fig. 3.6M). This indicates another potential mechanism by which dATP may improve ventricular function in the failing heart, and could explain why dATP treatment does not lead to further metabolic impairment at low dATP fractions.

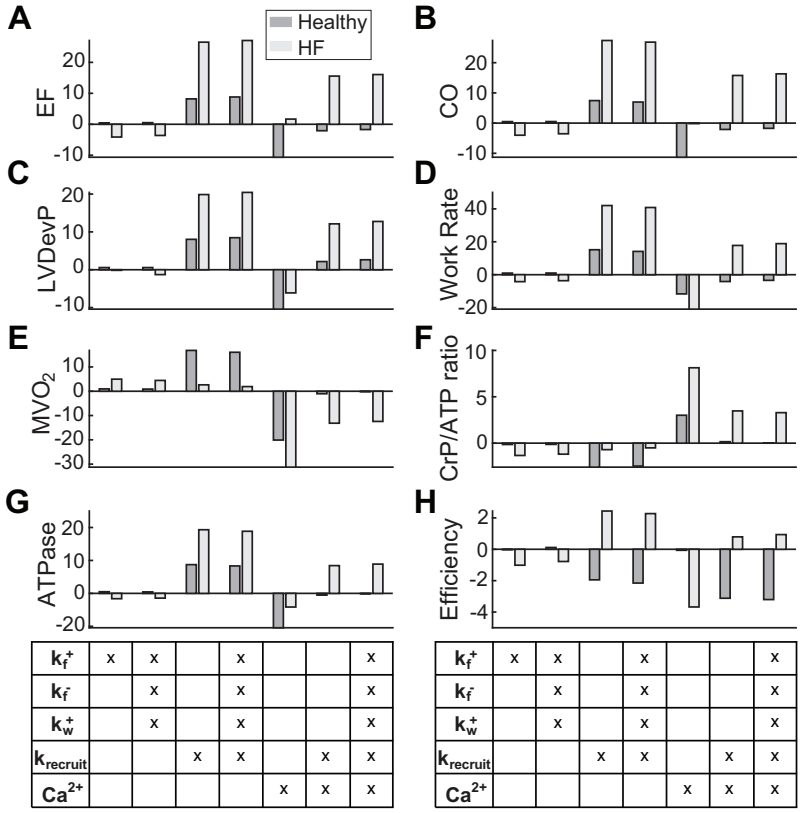


**Figure 3.7.** Elevated dATP improves ventricular function, especially in the failing heart. (A)-(H): Effects of parameter changes on percent change in metrics of LV function and energetics with 1% dATP in healthy and failing heart simulations, compared to ATP. Parameter changes include increased myosin recruitment (increasing $k_{recruit}$), crossbridge binding (increasing $k_f^+$) and cycling (increasing $k_f^-$ and $k_w^+$), and $Ca^{2+}$ sequestering dynamics with elevated dATP.

Finally, we assessed how each of our identified mechanisms of dATP contributed to changes in ventricular function in our normal and failing models. EF, CO, and LVDevP all increased to a greater extent with 1% dATP in failure (16%, 16%, and 13%, respectively) than in the healthy heart simulations, where EF was decreased by 2%, CO was decreased by 2%, and

LVDevP was increased by 3% compared to ATP (Fig.3.7A, B, C). Further, these results show that increased myosin recruitment has a larger impact in failure, leading to greater improvements in function. Interestingly, faster $Ca^{2+}$ dynamics with dATP treatment led to increased EF in the failing heart simulation, despite decreasing it in the healthy heart simulation (Fig. 3.7A). Increased myosin recruitment also led to greater increases in work rate (Fig. 3.7D) and efficiency in the failing heart simulations (Fig. 3.7H) (despite decreasing efficiency in the healthy heart simulations), while faster $Ca^{2+}$ dynamics led to an 8% increase in the CrP/ATP ratio (Fig. 3.7F) and a 31% decrease in $MVO_2$ (Fig. 3.7E) in the failing heart. These findings suggest that the net effect of dATP treatment is to improve contractile function, primarily due to its effects on myosin recruitment, while simultaneously improving energetic efficiency and the overall metabolic state of the failing heart, at least in part due to its effects on $Ca^{2+}$ handling.

## 3.4 Discussion

In this study, we used multiscale computational modeling to integrate therapeutic mechanisms of dATP from the molecular scale to the cardiovascular system in the failing heart. We predicted an increase in the actomyosin association rate with elevated dATP, potentially *via* stabilization of pre-powerstroke myosin. However, we found that this increase, as well as enhanced crossbridge cycling, did not lead to significant changes in contractile function. We found that recruitment of myosin into the crossbridge cycling pool contributed to increases in steady state force at the filament level, as well as increases in myocyte shortening and ventricular ejection fraction. Enhanced recruitment of myosin was shown to dominate contractile behavior at all scales, suggesting that this is the primary mechanism by which dATP improves contractility. This illustrates a particular strength of a multiscale modeling approach, since we were able to assess which mechanisms identified at different scales are most important in translating to changes in physiology, rather than focusing only on parameter changes at a single scale which may or may not be physiologically relevant. Accounting for the faster $Ca^{2+}$ transient decay observed

with elevated dATP along with enhanced myosin recruitment allowed us to also fully explain changes in cellular relaxation, as well as ventricular mechanics in the failing heart. LV function was shown to improve in a dose-dependent manner in simulations of the failing heart, with 1% dATP restoring EF closer to normal levels, in agreement with experimental results. dATP also improved energetic efficiency without further impairing metabolic state in HFrEF simulations. Notably, dATP had a more pronounced impact on ventricular function and energetic efficiency in the failing heart. The multiscale modeling framework developed in this study not only provides a powerful tool for linking molecular effects to changes in ventricular function, but also allowed us to parse the relative effects of several mechanisms of dATP at various scales of function, which would be difficult to accomplish experimentally.

Our modeling approach allowed us to gain new mechanistic insight into the effects of dATP on myosin, which agrees well with previous MD simulations [41, 168, 169, 181]. While previous myosin MSM have been constructed [182], the novel MSM-BD framework utilized in this study allowed us to gain insight into the mechanisms by which dATP specifically increases the actomyosin association rate, suggesting a combination of stabilization of the pre-powerstroke myosin structure and conformational changes in key protein regions, as well as overall changes in electrostatics. Our hypothesized mechanism of stabilization of the pre-powerstroke state of myosin may also decrease the likelihood of transition back into the OFF state, which could be further investigated in future computational studies. Interestingly, our BD simulations showed that dATP increases the actomyosin association rate to a slightly lesser extent than was shown previously [42], although our results were generally consistent (we observed a 1.9 fold change in the association rate compared with a 2.3 fold change in [42]). This is likely because our simulations covered a broader range of possible myosin conformations. Featurization and dimensionality reduction analysis using tICA showed that loop 2 motion was the most important kinetic feature in the MD simulations. This could increase actomyosin association rates and possibly other steps in the crossbridge cycle, and warrants further investigation [183]. This aligns with previous work suggesting that switch 1 provides an allosteric mechanism for transmitting

changes in the nucleotide binding pocket to loop 2 and the actin binding surface, increasing the electrostatic affinity of myosin for actin [37]. Further, the metastable conformation that was most dominant in our analysis, state 2 of the dATP simulation, also had the fastest association rate according to our ensemble BD approach. Generally all of these structures in state 2 can be described by having a more pronounced loop 2 extension. Even within this subsample of conformations found in state 2, the conformation that associated most quickly according to the BD simulations had an even greater extension of loop 2 relative to the other structures in the state, again pointing to loop 2 as a key structural and electrostatic feature. This analysis also matches with experimental assessments of loop 2 function, which highlight how additional positively charged lysines inserted into the loop increases weak binding [173]. Our improved method therefore overcomes a major limitation of BD in which molecules are treated as rigid bodies, by accounting for the conformational variability in several distinct sub-states, and allowing us to analyze the impact of structural features on protein-protein association.

Our model predictions indicate that thick filament mechanosensing largely contributes to the disproportionate effects of dATP on force. This has been proposed as a mechanism underlying length-dependent activation, where force development coincides with myosin heads transitioning from the thick filament backbone towards thin filaments [179, 184, 169, 185]. It is plausible that this may explain how elevated dATP leads to increased recruitment of myosin S1 heads from the thick filament backbone, but the underlying mechanism is still unclear. Furthermore, our model predictions indicated that nearest-neighbor cooperativity was also necessary to explain the large increases in force observed at low dATP percentages, which is consistent with results published in [179, 42]. This may point to a mechanism in which a greater number of myosin heads binding to actin increases the number of exposed myosin binding sites via cooperative mechanisms; this in turn allows more myosin heads to bind and generate force, leading to increased recruitment from the thick filament backbone *via* a strain-dependent positive feedback mechanism [37, 179]. However, it is important to note that myosin can also be recruited from the OFF state via phosphorylation of myosin binding protein C [186] and the myosin light

chain [187], so further investigation is warranted to determine whether these mechanisms may play a role. Further, our results suggest that myosin recruitment is the dominant mechanism by which dATP treatment leads to large increases in force with small amounts (1%) of dATP, while increases in crossbridge binding and cycling were sufficient to explain experimental data with 100% dATP, as shown previously [42]. This may be due to the fact that relatively low amounts of dATP may be sufficient to fully deplete the inactive myosin pool. Indeed, recent work showed that with 100% dATP, most myosin heads in the inactive pool have likely been recruited [169, 170]. Thus, with 100% dATP, increased crossbridge binding and cycling may be the dominant mechanism by which dATP further increases force production, since all available myosin heads have already been recruited. A combination of experimental techniques and molecular modeling would allow us to fully explore how dATP interacts with nearest-neighbor cooperativity and the inactive state of myosin at varying percentages.

We found that the combined mechanisms of increased myosin recruitment and faster $Ca^{2+}$ dynamics led to both improvements in contractile function and energetic efficiency in the failing heart, while efficiency was decreased in healthy heart simulations. This is consistent with previous findings for Omecamtiv Mecarbil, which has some similar mechanisms to dATP [188]. This is likely because in both healthy and failing conditions, dATP improves both work rate and ATPase rate; however, work rate is increased to a greater extent in the failing model, leading to an overall increase in efficiency, while ATPase rate is increased to a greater extent in the healthy model, leading to an overall decrease in efficiency. Interestingly, the large improvements in both ventricular function and energetic efficiency in HFrEF were mainly due to increased myosin recruitment into the crossbridge cycling pool. This could plausibly be due to a larger initial pool of myosin heads in inactive states in failure, but additional data is needed to further explore this. We also observed that increased myosin recruitment led to increases in $MVO_2$ in the healthy heart, but not in the failing heart, which may also contribute to improvements in energetics in failure. Further, we observed that faster $Ca^{2+}$ transient decay led to reduced EF in healthy heart simulations but increased EF in failing heart simulations, which could be due to the fact that

relaxation is likely impaired in failure. Further, we found that dATP increased the CrP/ATP ratio in failure despite not changing it in healthy heart simulations with 1% dATP. Experimental data from healthy transgenic mouse hearts showed a decreased CrP/ATP ratio with elevated dATP [39], but the CrP/ATP ratio has not been measured in HFrEF with dATP, so additional data is needed.

This model provides a powerful tool for assessing the mechanisms of dATP by integrating existing experimental data spanning the molecular to whole organ levels to generate new model predictions which can be tested through further experimentation and modeling.

## 3.5 Limitations

While the modeling framework developed in this study provides valuable insight into the mechanisms of dATP and could be extended in the future to assess additional small molecule therapeutics, additional work is needed to further develop the model to enable modular replacement of model components and parameter adjustment, since this model was developed specifically based on data for dATP. One major limitation of our model framework is propagation of uncertainty across scales, which remains a major challenge in multiscale modeling.

We chose to utilize two different sarcomere models in this study: a spatially explicit model [42] for filament scale force-pCa simulations, and an implicit model [73] for myocyte shortening and ventricular scale pressure-volume loop simulations. We chose to utilize the spatially explicit model to assess how small fractions of dATP interact with nearest-neighbor cooperativity to produce disproportionate amounts of steady state force, because this model contains a sophisticated representation of cooperative mechanisms. However, this model is computationally expensive and does not allow for simulation of myocyte shortening, $Ca^{2+}$ dynamics, or energetics, so we chose to utilize the implicit modeling framework developed by Lopez *et al.* [73] for larger scale simulations. Due to differences in each of these models, the parameter values adjusted in each model differed slightly, mainly due to the addition of

a weakly-bound state in the implicit model. dATP may also increase the transition between the weakly- and strongly-bound states, but additional studies are needed to fully address the mechanisms by which dATP increases the rate of crossbridge cycling. Further, it was necessary to increase $k_{recruit}$ by a greater extent in the spatially explicit model than in the implicit model to achieve the same effect. This may be explained by the fact that overall forces are lower by several orders of magnitude in the spatially explicit model, which represents a single sarcomere, than in the implicit model, which represents a whole cell.

Further, while we were able to estimate an actomyosin association rate from BD simulation results based on the measured ATP-myosin.actin association rate, we observed a range of possible association rates for varying reaction distances. Additionally, the ensemble binding curve from dATP state 2, which has the largest stationary distribution also has the fastest binding rate. It is possible that within dATP state 2, conformational variability leads to greater exploration of states to identify a more favorable conformation. However, the variance of binding rates from each of the dATP states is not substantially different, which suggests that the change in measured association rate is not based on conformational exploration. Further, our BD simulations only included a single myosin head and several actin monomers. The accuracy of these simulations could be further improved in the future by including a more complete representation of the thin filament and multiple myosin heads. Adding additional constraints on myosin head movement, rather than allowing it to freely diffuse around the actin filament, would also improve association rate predictions. Additionally, it is important to emphasize that while our MD and MSM analysis inform our hypothesis that dATP stabilizes pre-powerstroke myosin, leading to increases in the actomyosin association rate, additional simulations are needed to assess transition kinetics of dATP out of the pre-powerstroke state.

Additional work is needed to fully explore the effects of dATP on $Ca^{2+}$ dynamics. The relative changes in DT50 and DT90 of the $Ca^{2+}$ transient with dATP measured in isolated cardiomyocytes (*in vitro*) at 1 Hz were utilized to scale the $Ca^{2+}$ transient for the ventricular simulations (at 7 Hz). This does not take into account possible frequency effects on changes to

the $Ca^{2+}$ transient with dATP, or the effects of the experimental preparation. Further, we assumed that the changes in the $Ca^{2+}$ transient were independent of dATP fraction, but the validity of this assumption warrants further investigation. Additional work is also needed to determine the underlying mechanism by which dATP increases the rate of $Ca^{2+}$ transient decay. We showed in a recent study that dATP acts on the sarcoplasmic reticulum ATPase (SERCA), leading to faster pumping of $Ca^{2+}$ back into the sarcoplasmic reticulum and thus faster $Ca^{2+}$ transient decay, [164, 42] but additional studies are needed to determine whether dATP also acts on other ATPase pumps or mechanisms that regulate $Ca^{2+}$ handling in the cell. Extension of our model to include a model of the SERCA pump and other proteins involved in $Ca^{2+}$ handling would allow us to further investigate the effects of dATP on the $Ca^{2+}$ transient. It is possible that there are additional mechanisms at play at the ventricular level such as regulation by the autonomic nervous system that could be taken into account in future studies [189]. Additionally, we utilized averaged experimental dATP $Ca^{2+}$ transient data for our simulations, but experimental measurements of the extent of the effect of dATP on $Ca^{2+}$ dynamics vary substantially.

In addition to uncertainty in the $Ca^{2+}$ transient, $k_{recruit}$ was increased to a different extent at the filament level than at the cell and ventricular levels, and had to be increased by around three orders of magnitude to match the expected increase with 1% dATP in the spatially explicit model, and by around 2 orders of magnitude in the spatially implicit model. The representation of myosin transition out of the OFF state in this model is fairly simplistic, so it is difficult to determine the exact increase in this parameter caused by elevated dATP without a more structural relevant representation of myosin ON/OFF state dynamics. The myosin OFF state is an ongoing area of investigation, and the mechanisms by which dATP disrupts the resting conformation of myosin are still not fully understood, but additional experimental and computational studies will aid in improving understanding. Regardless, model results showed that $k_{recruit}$ was the only parameter which could be increased to produce disproportionate increases in force with small amounts of dATP, so while the exact magnitude of this change is unknown, this provides support for the conclusion that this is the primary mechanism by which small amounts of dATP improve

contractile function. We also carried out an additional sensitivity analysis where we varied $k_{recruit}$ simultaneously with the $Ca^{2+}$ transient to assess how variability in these parameters affects myocyte and ventricular mechanics. Generally, faster $Ca^{2+}$ transient decay leads to decreased FS, RT50, and RT90, while increased $k_{recruit}$ leads to increased FS, RT50, and RT90. At the ventricular level, faster $Ca^{2+}$ transient decay leads to decreased EF, CO, LVDevP, work rate, $MVO_2$, and ATPase rate, while increasing the CrP/ATP ratio and efficiency. Increasing $k_{recruit}$ has the opposite effect, increasing EF, CO, LVDevP, work rate, $MVO_2$, and ATPase rate, while decreasing the CrP/ATP ratio and efficiency. However, the effects on efficiency are more variable, especially in HFrEF simulations. These patterns are consistent with our overall conclusions. In the future, additional experimental data on $Ca^{2+}$ dynamics, as well as higher resolution data and simulations on myosin OFF state dynamics will aid in better elucidating these mechanisms and the interplay between them.

Finally, our modeling approach utilized a simplified model of the heart which approximates the left and right ventricles as hemispheres. While this approach allowed us to gain valuable insight into the ways in which dATP affects ventricular performance at a relatively low computational cost, in the future this framework could be extended to capture more realistic geometries. A finite element model of the heart would allow us to incorporate patient-specific geometries and to assess regional changes in mechanics, as well as potential growth and remodeling. Further, in our assessment of changes in ventricular function with elevated dATP in healthy and HFrEF simulations, we focused primarily on changes in EF, as well as other global heart metrics such as CO and LVDevP, because these were the only experimental *in vivo* data available for comparison. In the future, it would be valuable to expand our model to include additional mechanisms of heart failure such as myocyte death in myocardial infarction, changes in excitation-contraction coupling, and spatially varying changes in sarcomere dynamics with mutations in sarcomere proteins. Additionally, dATP treatment may lead to functional and/or morphological changes in mitochondria, so further experimental data and expansion of our model to include these mechanisms would allow for a more complete assessment of the effects of dATP

on energetics in failure.

Fig. 3.1 provides an overview of the modeling approach. Code can be found at https://github.com/abbyteitgen11/dATP_multiscale_model, and was developed in part based on a previously published multiscale rat mechanoenergetics model [73, 190, 188, 33], as well as a spatially explicit sarcomere model developed in our group [42].

## 3.6 Methods

### 3.6.1 Molecular Dynamics Simulations of ATP-Myosin and dATP-Myosin

We first conducted Molecular Dynamics (MD) simulations of pre-powerstroke ADP.Pi-myosin and dADP.Pi-myosin. Starting coordinates for Atlantic bay scallop (*Argopecten irradians*) myosin II S1 in the pre-powerstroke state (ADP.Pi) were obtained from an X-ray crystal structure in the Protein Data Bank (PDB, www.rcsb.org [101]) solved by [191] (PDB ID: 1QVI, 2.54 Å resolution). The original structure includes myosin II with ADP and vanadate ($VO_4$) [192], a $Ca^{2+}$-bound essential light chain (ELC), and a $Mg^{2+}$-bound regulatory light chain (RLC). The $VO_4$ ion was replaced with inorganic phosphate (Pi). To increase the computational efficiency of simulations investigating motor domain dynamics, myosin II was truncated after residue 810 and residues in the RLC were removed. Then, missing heavy atoms were built using Modeller [193]. These starting coordinates were used to generate an additional system in which ADP was replaced by dADP *via* removal of the 2' oxygen. These systems will be referred to as ADP.Pi-myosin and dADP.Pi-myosin, respectively. Next, hydrogen atoms were modeled onto the structures using the leap module of AMBER and each protein was solvated with explicit water molecules in a periodic, truncated octahedral box that extended 10 Å beyond any protein atom. Finally, $Na^+$ and $Cl^-$ counterions were added to neutralize the systems and then 120 mM $Na^+$ and $Cl^-$ ions were added.

All simulations were performed with the AMBER package [119, 194] and the ff14SB

force field [195]. Water molecules were treated with the TIP3P force field [196]. Metal ions were modeled using the Li and Merz parameter set [197, 198, 199]. ADP, dADP, and Pi molecules were treated with the GAFF2 force field [200] with custom parameters derived from QM simulations using ORCA [201, 202]. The SHAKE algorithm was used to constrain the motion of hydrogen-containing bonds [203, 118]. Long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) method.

Each system was minimized in three stages. First, hydrogen atoms were minimized for 1000 steps in the presence of 100 kcal $mol^{-1}$ restraints on all heavy atoms. Second, all solvent atoms were minimized for 1000 steps in the presence of 25 kcal $mol^{-1}$ restraints on all protein atoms. Third, all atoms were minimized for 8000 steps in the presence of 25 kcal $mol^{-1}$ restraints on all backbone heavy atoms (N, O, $C_\alpha$ and C atoms) After minimization, systems were heated to 310 K during three successive stages. In each stage, the system temperature is increased by $\sim$100 K over 100 ps (50,000 steps) using the canonical NVT (constant number of particles, volume, and temperature) ensemble. During all heating stages, 25 kcal $mol^{-1}$ restraints were present on the backbone heavy atoms (N, O, $C_\alpha$ and C atoms). After the system temperatures reached 310 K, the systems were equilibrated over 5 successive stages using the isobaric-isothermal NPT (constant number of particles, pressure, and temperature) ensemble. During the first 4 stages, the systems were equilibrated for 0.4 ns in the presence of restraints on backbone atoms. The strength of the restraints was decreased from 25 kcal $mol^{-1}$ during the first stage to 1 kcal $mol^{-1}$ during the fourth stage. During the final equilibration stage, the systems were equilibrated for 5 ns in the absence of restraints.

Production dynamics for conventional MD simulations were then performed using pmemd in the canonical NVT ensemble using an 8 Å nonbonded cutoff, a 2 fs time step, and coordinates were saved every 10 ps. Simulations were run in triplicate for 2,000 ns each (12 $\mu$s net sampling). Unless specified otherwise, simulations were analyzed separately, and the results of replicate simulations were averaged together.

### 3.6.2 Markov State Model Construction

These triplicate simulations of ADP.Pi-bound and dADP.Pi-bound myosin were used to construct two Markov state models (MSM) for each of the nucleotide conditions (ATP and dATP) (Fig. 3.8). Feature selection to inform these MSM was carried out, focusing on the behavior of the actin binding surface. The residue composition and structural dynamics of the cardiomyopathy loop, loop 2, and loop 4, and closure of the cleft have all been associated with myosin binding to actin during the cross-bridge cycle [204, 205, 174, 175]. Therefore, we used a collection of these features, based on the distances between their centers of mass of either the tip, base or entire loops of interest, as well as two different measurements of the cleft opening. Center of mass calculations were used as an initial dimensionality reduction, and provided a set of internal coordinates. To further reduce the number of dimensions, time-lagged Independent Component Analysis (tICA) was used to reduce these 22 features to only two primary features, using a time lag of 200 ps (Fig. 3.8B). tICA is widely used in MSM construction because it effectively identifies the slow kinetics of MD simulations (as compared to principal component analysis or other dimensionality reduction techniques) [206]. All six trajectories were used in featurization and dimensionality reduction steps.

A *k*-means clustering algorithm was applied to the 2-D tICA spaces, to reduce the MD simulations into 500-state space using the two primary features from the tICA analysis, where each frame of the simulation is classified into one of the 500 states, and transitions between states were counted using a Bayesian approach (Fig. 3.8C) [207, 208]. Implied timescales (IT) analysis was utilized to determine an appropriate lag time for MSM construction. This analysis indicated that at a lag time of 900 ps, the timescales plateaued. Therefore, a lag time of 900 ps was used to construct MSM for both ATP-bound myosin as well as dATP-bound myosin (Fig. 3.9). IT analysis also revealed two major motion components for both ATP and dATP myosin simulations, suggesting that these 500-state MSMs could be reduced to three metastable states (Fig. 3.8). Note that states 0, 1 and 2 are used purely for naming convention, and state

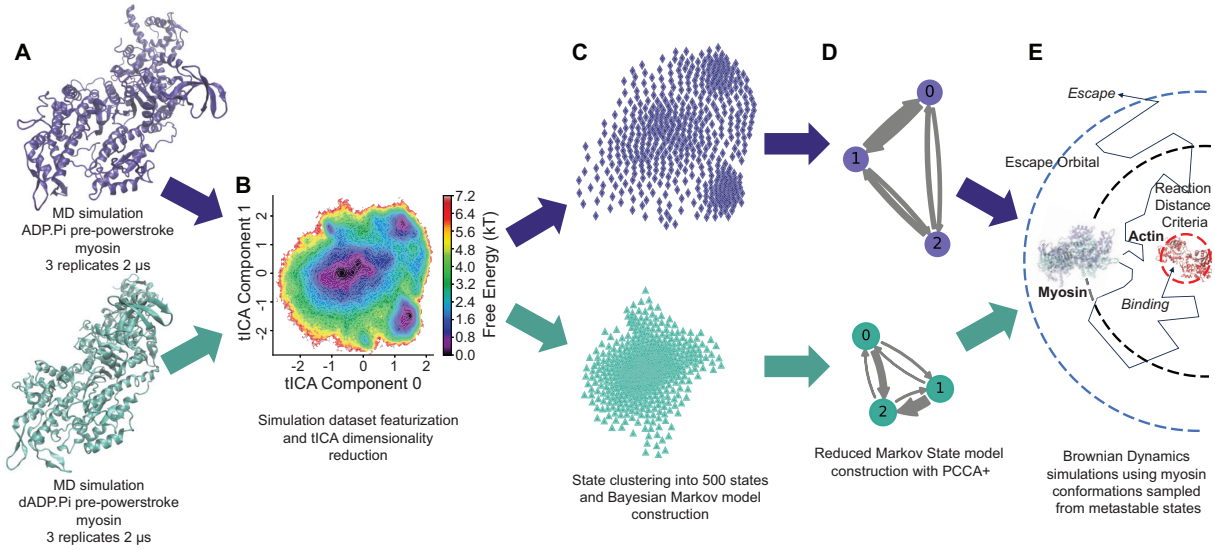


**Figure 3.8.** Workflow of the MD to Markov state model construction used. The ATP and dATP trajectories were first reduced in dimensions to 22 features measured on different loops and surfaces on the actin binding surface of myosin (A). From the input features, time-lagged Independent Component Analysis (tICA) was used to further reduce the simulation to two dimensions (B). A *k*-means clustering algorithm was used to cluster the MD trajectory into a 500 microstate Markov model using a Bayesian estimation approach (C). PCCA+ fuzzy clustering were used to further reduce the Markov state models of ATP and dATP into 3 metastable state models each (D). From each metatstable state, 15 conformations were sampled, and used as inputs to the Brownian dynamics simulations using an ensemble based approach (E). Brownian dynamics simulations were carried out over a range of reaction distances to generate a binding curve.

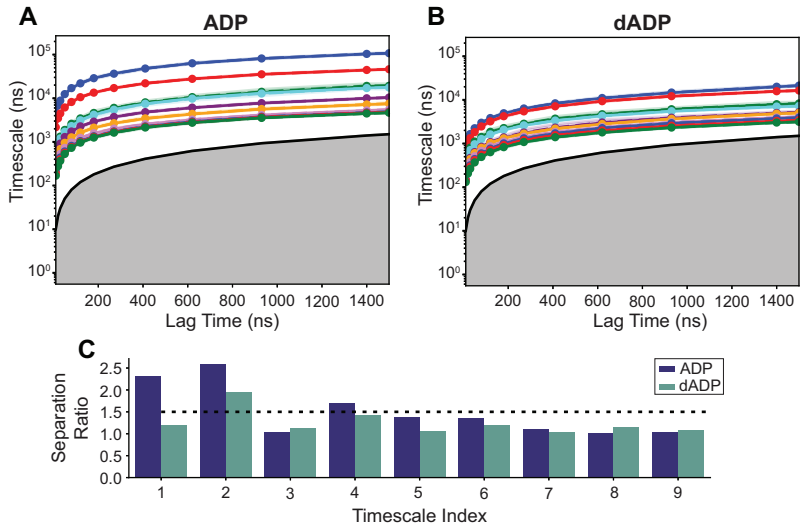


**Figure 3.9.** Implied timescale plots for the construction of the ATP-bound myosin (A) and dATP-bound mysoin MSM (B). First ten timescales are shown. The shaded colored bars represent the statistical uncertainties estimated using a Bayesian approach. The gray shaded region represents the forbidden timescale space where the lag time becomes larger than the timescales of the motion, and the behavior cannot be resolved. A lag time of 900 ns was selected for the model. (C) The ratio of the implied timescales were plotted at a lag time of 900 ns to visualize the timescale separation. Three macro states were selected based on ratio of the 2nd and 3rd timescales being greater than 1.5 for both the ATP and dATP models.

0 in the ATP MSM is not analogous to state 0 in the dATP MSM. A Chapman-Kolmogorov test was used to validate the MSMs, and showed minimal deviation between the observed and predicted transition kinetics within the expected bounds (Fig. 3.10) [209]. From each of the three metastable states, 15 frames were sampled to be used in the Brownian dynamics (BD) simulations, for both the ATP model and dATP model, cumulatively leading to 45 structures per nucleotide condition to be used in BD simulations (Fig. 3.8E). Analysis was carried out in Python using PyEMMA [207].

### 3.6.3 Brownian Dynamics Simulations of Actomyosin Association

We used rigid body BD simulations to estimate the association rate of pre-powerstroke myosin to actin. In this work, we improved upon previous BD simulations [42] by using an ensemble of representative structures from the MD trajectories that have been sampled from

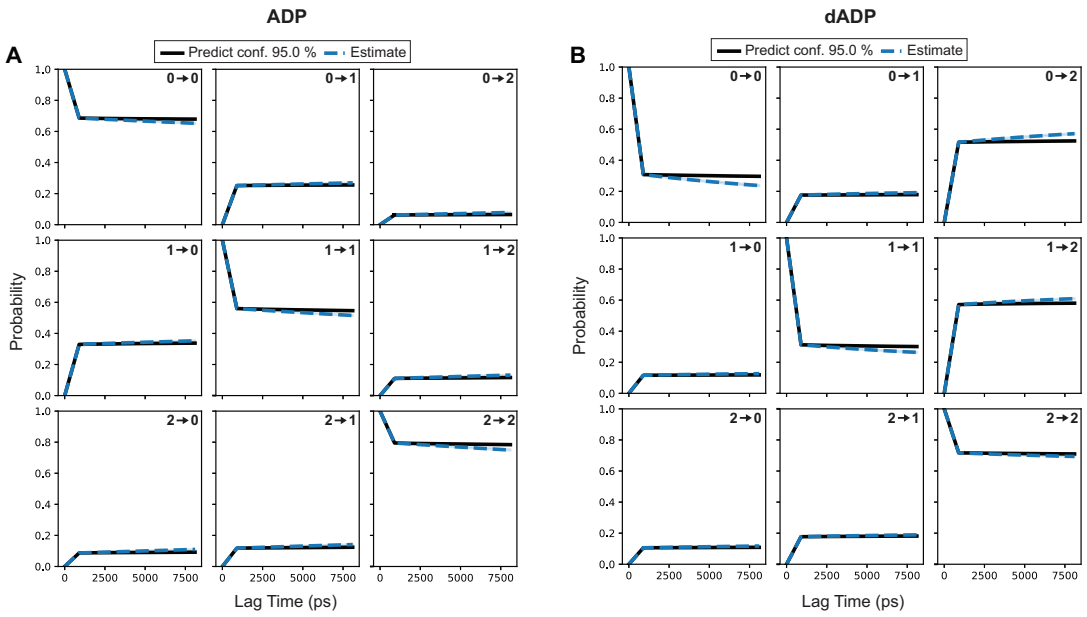


**Figure 3.10.** Chapman–Kolmogorov plot visualization to test the assumption that the model is not sensitive to the lag time selected for the ATP-bound (A) and dATP-bound (B) Markov state models. The solid prediction line shows the transition probabilities from one metastable state to another (i.e. 1 -¿ 2) estimated at integer multiples of the 900 ns selected lag time. The dotted ”estimated” line is from independently creating models at each of the integer multiples of the originally selected lag time, and plotting those transition probabilities. The shading represents the 95% confidence intervals for the estimated transitions. The transitions generally agree between the predicted and estimated transitions, validating the Markovianity of the models.

metastable states in the MSM as described above. Traditional approaches using BD typically require a single frame from the MD simulation to be selected and used for the BD simulations. However, our approach using an ensemble of conformations avoids this frame selection challenge. In BD simulations, the first protein of interest (myosin) is placed randomly on a sphere and allowed to diffuse, either binding to the second protein of interest (actin) if it reaches a set reaction distance criteria, or escaping if it reaches a set escape orbital (Fig. 3.8E).

BD simulations were carried out with Browndye 2.0 [68]. From each independent conformation sampled from the ATP and dATP MSMs, independent BD simulations were carried out to estimate the association rate between the myosin S1 head and a scallop actin dimer homology model. The homology model was built using the source sequence from *p. magellanicus* and template structure (PDB ID: 3J8A) [210]. We defined reaction pairs within Browndye based on possible hydrogen bonding pairs within 3.5 Å from sampled conformations within each metastable state of the MSM (15 per metastable state), defined in the bound actomyosin state. The bound state was constructed by aligning the myosin structure with actin using a crystal structure of bound actin and myosin (PDB ID: 6X5Z) as a reference [205]. Alignment was done in VMD [167, 211, 212]. PQR files were created for all structures using the Amber20 force field with PDB2PQR [105, 108]. Electrostatic fields were generated for both myosin and actin structures using APBS [213]. Rather than define a reaction distance, we simulated BD trajectories without a reaction endpoint and recorded the closest distance between actin and myosin during the trajectory in order to calculate association rates at a range of reaction distance criteria. BD simulations were carried out with 250,000 trajectories per conformation, leading to 3.75 million total trajectories for each metastable state of each MSM, and 11.25 million BD trajectories for ATP and dATP each. The binding curves were then averaged based on the stationary distributions of the metstable states in the three-state MSM for both ATP and dATP. These binding curves were then used for analysis and comparison. The full mechanics and simulation approach of the BD methods are described in more detail in [68].

### 3.6.4 Spatially Explicit Sarcomere Model

We next modeled the effects of dATP on sarcomere mechanics, incorporating the association rates determined from the BD simulations. To do this, we modified a spatially explicit cooperative model of the sarcomere previously published in our group [42, 178]. This model is constructed using a Monte Carlo Markov Chain and contains 26 regulatory units (RUs). Each RU represents seven actin monomers, one myosin S1 head, troponin, and tropomyosin. This model consists of five states: blocked, with no calcium bound to the thin filament ($B_0$), blocked, with calcium bound to the thin filament ($B_1$), closed (C), strongly bound ($M_1$), and post-powerstroke ($M_2$). Parameters $k_{Ca}^{+}$ and $k_{Ca}^{-}$ determine transition between $B_0$ and $B_1$, representing calcium binding to the thin filament; parameters $k_b^{+}$ and $k_b^{-}$ determine the transition between $B_1$ and C, representing the movement of tropomyosin across the actin surface to expose the myosin binding site; parameters $k_f^{+}$ and $k_f^{-}$ determine the transition between C and $M_1$, representing myosin binding to actin; parameters $k_p^{+}$ and $k_p^{-}$ determine the transition between $M_1$ and $M_2$, representing the powerstroke; parameters $k_g^{+}$ and $k_g^{-}$ determine transition from $M_2$ to C, representing detachment.

Transition rates between states $B_1$, C, $M_1$, and $M_2$ for a given RU depend on the states of the two neighboring RUs. Cooperative coefficients $\gamma_B$ and $\gamma_M$ (for a B to C transition), and $\mu_B$ and $\mu_M$ (for a C to M transition) are computed based on the additional free energy contributions of nearest-neighbor RUs compared to a reference state. $\mu_B$ is set equal to $\gamma_M$ to maintain reversibility. These cooperative coefficients incorporate mechanisms of both thin filament (tropomyosin) overlap and strain dependence from neighboring bound myosin heads.

In addition to these five states, we added a sixth state to represent the pool of inactive myosin heads not available for crossbridge cycling (OFF). To do this, we incorporated the transition between the active and inactive states based on [179] and [33], governed by parameters $k_m^{+}$ and $k_m^{-}$, as well as $k_{recruit}$, which describes the force dependence of this transition (Equation 3.1), where $[M_2]$ is the fraction of RUs in the $M_2$ (post-powerstroke, force-producing) state.

$$k_m^+ = k_m^{+ref}(1 + k_{recruit}[M_2]) \tag{3.1}$$

After addition of this state, all other state occupancies were approximately the same as in the original model formulation, except for a lower occupancy in the C state. Further description of this model can be found in [42]. External RUs were clamped in the $B_0$ state, and all other RUs were initially set to the OFF state. 640 simulations were run for steady state force-pCa simulations, and transition probability was determined using a random number from 0 to 1, using a time step of $5\times10^{-4}$ as in [42]. Analysis was carried out using CUDA 11.7 [60] and Python 3.9.13 [214].

Default parameters from the McCabe et al. model [42] were utilized for this study, except those that were optimized: parameters $k_m^+$, $k_m^-$, and $k_{recruit}$ were optimized to match the approximate steady state percentage of myosin heads in the OFF state (30%) based on X-ray diffraction data [170, 169], as well as to match steady state force-pCa data for ATP from [87]. Parameters $k_p^+$ and $k_g^+$ were adjusted to simulate dATP treatment in the same way as in [42], further detailed in Table 3.1. $k_{recruit}$ was further adjusted to maximize steady state force with 1% dATP, and $k_f^+$ was set based on BD results (Table 3.1). Fitting was done to minimize the sum of squared error between data and model and was conducted using parameter sweeps carried out in Python 3.9.13 [214]. Hill curve fits were utilized to smooth model outputs for force-pCa simulations.

### 3.6.5 Myocyte Mechanics Model

We next assessed the effects of dATP on whole myocyte mechanics using a non-spatially explicit sarcomere and myocyte mechanics model developed by [73], which is comprised of a system of differential-algebraic equations (DAEs). Analysis was carried out in MATLAB R2018b [215]. ODE15s was used to solve all differential equations. This model consists of six states: inactive (OFF), nonpermissible (N), permissible (P), weakly-bound ($A_1$), strongly-

bound ($A_2$), and post-ratcheted ($A_3$) [73, 190, 188]. Parameters $k_f^+$ and $k_f^-$ determine transition between P and $A_1$, representing myosin binding to actin; parameters $k_w^+$ and $k_w^-$ determine transition between $A_1$ and $A_2$, representing the transition from weakly-bound to strongly-bound; parameters $k_p^+$ and $k_p^-$ determine transition between $A_2$ and $A_3$, representing the powerstroke; parameter $k_g^+$ determines transition from $A_3$ to P, representing detachment. Parameters $k_{on}$ and $k_{off}$ represent $Ca^{2+}$ association to the thin filament and determine transition from N to P, along with a cooperative parameter $k_{coop}$. This model includes the same model of transition between the active and inactive states as the spatially explicit model, based on [179] (Equation 3.2), where ON represents the active states, OFF represents the inactive states, and $\sigma_{XB}$ represents active contractile force. This model also includes parallel passive and series elastic springs, and a parallel dashpot, to represent sarcomere viscoelasticity. Additional details on this model can be found in [33, 190].

$$\frac{dON}{dt} = k_m^+(1 + k_{recruit}\sigma_{XB})OFF - k_m^- ON \tag{3.2}$$

Default model parameters from the rat model [73] were used for this study, except those that were optimized as described in Table 3.2. Parameters were optimized to match steady state force-pCa rat data from [87] and average unloaded shortening data from [38, 39]. For force-pCa simulations, sarcomere length was fixed at 2.25 $\mu$m based on experimental protocols [87], and $k_{SE}$ was set to $5 \times 10^4$ mmHg/$\mu$m to simulate isometric contraction. $k_{on}$ was set to 50 $s^{-1}$, $k_{passive}$ was set to 0.1 mmHg/$\mu$m, and $\eta$ was set to 1.5 mmHg/$\mu$m to match $EC_{50}$ and steady state force for ATP. For shortening simulations, $k_{SE}$ was set to 35 mmHg/$\mu$m, $k_{passive}$ was set to 0.1 mmHg/$\mu$m, and $\eta$ was set to 0.001 mmHg/$\mu$m to match FS, RT50, and RT90 for ATP average shortening data. For each simulation, all myosin heads were initially set to the OFF state. Shortening simulations were carried out at 1 Hz for comparison to average experimental shortening data from [38, 39], and the model was run for 3 beats to reach steady state.

$$k_x = k_{x,ATP}[1 - dATP(\%)] + k_{x,dATP}[dATP(\%)] \tag{3.3}$$

Since this model is not spatially explicit, parameters were scaled as functions of dATP level, with overall nucleotide concentrations kept constant according to Equation 3.3, where $k_x$ represents the parameter assumed affected by dATP. Parameter $k_f^+$ was set based on BD results, and parameters $k_f^-$ and $k_w^+$ were optimized to mach $EC_{50}$ and steady state force for 100% dATP, further described in Figs. S12-S14. This achieved the same effect as increasing parameters $k_f^+$, $k_p^+$, and $k_g^+$ in the spatially explicit sarcomere model. Parameter $k_{recruit}$ was further optimized to match average shortening data for 1% dATP [38, 39]. Optimization was done using a combination of manual tuning (to determine a reasonable parameter range) and Particle Swarm Optimization in MATLAB [127]. Some parameter and state names were altered from original model: $U_{SR}$ was changed to OFF, $U_{NR}$ was changed to ON, $k_{SR}^+$ was changed to $k_m^+$, $k_{SR}^-$ was changed to $k_m^-$, $k_{force}$ was changed to $k_{recruit}$, $k_a$ was changed to $k_f^+$, $k_d$ was changed to $k_f^-$, $k_1$ was changed to $k_w^+$, $k_{-1}$ was changed to $k_w^-$, $k_2$ was changed to $k_p^+$, $k_{-2}$ was changed to $k_p^-$, and $k_3$ was changed to $k_g^+$.

The $Ca^{2+}$ interpolation function from this model (based on experimental data from [216]) was used to simulate $Ca^{2+}$ transients at a range of frequencies for analysis at the cellular and organ levels. We adjusted parameters a, b, c, and $Ca_0$ in the interpolation function to match DT50 and DT90 for ATP and dATP seen experimentally. Experimental studies have shown varying results on the effects of dATP on $Ca^{2+}$; Korte *et al.* [38] reported a 50% decrease in DT50 and a 49% decrease in DT90 with elevated dATP, while Nowakowski *et al.* did not report a significant decrease in DT50 and DT90 with elevated dATP. Therefore, to account for the range in experimental values, we averaged these two data points to obtain an average $Ca^{2+}$ transient for dATP, where DT50 is decreased by 31% and DT90 is decreased by 25%. This variability in $Ca^{2+}$ data is further addressed in Figs. S22-S26. Corresponding shortening data was also averaged as described above, where average FS for ATP was 7.1 and average FS for dATP was 9.5.. For

**Table 3.1.** Summary of parameter changes made to spatially explicit sarcomere model from [42] compared to implicit sarcomere model from [73]. All original parameters from [42] were utilized in updated model after addition of the OFF state, except for transition parameters $k_m^+$, $k_m^-$, and $k_{recruit}$, which were optimized to match steady state ATP force-pCa data [87] and to maintain a baseline OFF state occupancy of approximately 30%, based on experimental data [169, 170]. In the updated model, $k_f^+$ was increased to 478 $s^{-1}$ compared to 567 $s^{-1}$ in the original model based on updated BD data, but $k_p^+$ and $k_g^+$ were increased by the same amount to simulate dATP. $k_{recruit}$ was also increased to simulate dATP. This differs slightly from the parameters increased to simulate dATP in the spatially implicit model ($k_f^+$, $k_f^-$, $k_w^+$, and $k_{recruit}$, further discussed in the main text). Additionally, $k_{recruit}$ was increased by a greater amount in the spatially explicit model than in the spatially implicit model, also discussed in the main text. Values shown are for 100% dATP.

| | **Original Spatially Explicit Model** | | **Updated Spatially Explicit Model** | | **Spatially Implicit Model** | |
|---|---|---|---|---|---|---|
| | ATP | dATP | ATP | dATP | ATP | dATP |
| $k_{Ca}^+$ ($\mu M^{-1}\,s^{-1}$) | 90 | | 90 | | | |
| $k_{Ca}^-$ ($s^{-1}$) | 570 | | 570 | | | |
| $k_b^+$ ($s^{-1}$) | 13000 | | 13000 | | | |
| $k_b^-$ ($s^{-1}$) | 100 | | 100 | | | |
| $k_f^+$ ($M^{-1}s^{-1}$) | 250 | 567 | 250 | 478 | 250 | 478 |
| $k_f^-$ ($s^{-1}$) | | | | | 304.7 | 460 |
| $k_w^+$ ($N^{-1}m^{-2}$) | | | | | 112.4 | 170 |
| $k_p^+$ ($s^{-1}$) | 50 | 80 | 50 | 80 | 811.7 | |
| $k_g^+$ ($s^{-1}$) | 135 | 230 | 135 | 230 | 144.6 | |
| $\gamma_B$ | 45 | | 45 | | | |
| $\gamma_M$ | 21 | | 21 | | | |
| $\mu_M$ | 2 | | 2 | | | |
| $k_m^+$ ($s^{-1}$) | | | 16 | | 15.5 | |
| $k_m^-$ ($s^{-1}$) | | | 15 | | 50 | |
| $k_{recruit}$ ($N^{-1}m^{-2}$) | | | 0.2 | 779 | 0.2 | 37 |

ventricular simulations, we used the original model $Ca^{2+}$ transient from [73], and applied these same relative changes to DT50 and DT90 for dATP. Maximum and minimum $Ca^{2+}$ transient values were set to be equal for ATP and dATP for all simulations since they were not shown to vary significantly experimentally [38, 39]. For cellular-level simulations, minimum $Ca^{2+}$ was set to 0, and maximum $Ca^{2+}$ was set to 1. For organ-level simulations, minimum and maximum $Ca^{2+}$ values were set to be be the same as in the original model from [73].

### 3.6.6 Crossbridge Energetics and Mitochondrial Metabolism Model

The mitochondria model implemented in [73] (based on [77]) was utilized to simulate myocardial energetics. This model consists of 29 ordinary differential equations describing the membrane potential, metabolite concentrations, and ion concentrations in the mitochondria. The mitochondria is divided into three main compartments: matrix, inter-membrane space, and

**Table 3.2.** Summary of parameter changes made to implicit model from [73] to simulate ATP and dATP compared to original model parameters. ATP and dATP parameters were fit independently. $k_f^+$ was altered based on actomyosin association rate from BD. $k_f^-$, $k_w^+$, and $k_{recruit}$ were optimized to match dATP experimental data [87, 38, 39]. Viscoelastic parameters $k_{passive}$, $\eta$ and $k_{SE}$ were adjusted based on experimental protocol (force-pCa for filament simulations, shortening for myocyte simulations, and PV loops for ventricular simulations), as described in the methods. $Ca^{2+}$ association parameter $k_{on}$ was adjusted to account for differences between *in vitro* and *in vivo* data [217]. Parameter values shown are for 100% ATP and 100% dATP. Parameters $k_f^+$, $k_f^-$, $k_w^+$, and $k_{recruit}$ were increased to simulate dATP treatment.

| | **Original** | **Filament Simulations** | | **Myocyte Simulations** | | **Ventricular Simulations** | |
|---|---|---|---|---|---|---|---|
| | ATP | ATP | dATP | ATP | dATP | ATP | dATP |
| $k_{passive}$ $(mmHg/\mu m)$ | 25 | 0.1 | | 0.1 | | $5.8\times10^4$ | |
| $k_{coop}$ | 9.68 | | | | | | |
| $k_{on}$ $(s^{-1})$ | 101.2 | 50 | | 50 | | 200 | |
| $k_{off}$ $(s^{-1})$ | 723.9 | | | | | | |
| $k_m^+$ $(s^{-1})$ | 15.5 | | | | | | |
| $k_m^-$ $(s^{-1})$ | 50 | | | | | | |
| $k_{recruit}$ $(N^{-1}m^{-2})$ | 0.2 | | 37 | | 37 | | 37 |
| $k_f^+$ $(M^{-1}s^{-1})$ | 559.6 | 250 | 478 | 250 | 478 | 250 | 478 |
| $k_f^-$ $(s^{-1})$ | 304.7 | | 460 | | 460 | | 460 |
| $k_w^+$ $(s^{-1})$ | 112.4 | | 170 | | 170 | | 170 |
| $k_w^-$ $(s^{-1})$ | 21.3 | | | | | | |
| $k_p^+$ $(s^{-1})$ | 811.7 | | | | | | |
| $k_p^-$ $(s^{-1})$ | 43.3 | | | | | | |
| $k_g^+$ $(s^{-1})$ | 144.6 | | | | | | |
| $k_{SE}$ $(mmHg/\mu m)$ | $5\times10^4$ | | | 35 | | $1\times10^4$ | |
| $\eta$ $(mmHg*s/\mu m)$ | 1 | 1.5 | | 0.001 | | $1\times10^{-4}$ | |

cytosol. As in [73], metabolite concentrations in the mitochondria model feed into the myocyte mechanics model. The crossbridge cycling rate from the sarcomere model is used to calculate the ATPase rate, which feeds into the mitochondria model. Coupling to the energetics model was only implemented for ventricular simulations. The model from [73] was altered so that the mitochondria model updates the metabolite pools every three beats to allow for assessment of changes in metabolite concentrations over time. Three beats was chosen as a reasonable timescale to allow for a stable solution to model equations. Analysis was carried out in MATLAB R2018b [215].

### 3.6.7 Ventricular Mechanics and Hemodynamics Model

Ventricular simulations under healthy and failing conditions were carried out using the rat ventricular mechanics and hemodynamics model from [73]. This model is based on [76],

**Table 3.3.** Summary of model metabolite concentrations in healthy and HFrEF simulations. Total adenine nucleotide pool (TAN), total creatine pool (CRtot), total exchangeable phosphate pool (TEP), and oxidative capacity ($Ox_{capacity}$) were adjusted based on average experimental values for healthy and TAC (heart failure) rats from [73].

| | **Healthy** | **HFrEF** |
|---|---|---|
| TAN (M/L cell) | 0.0076 | 0.0070 |
| CRtot (M/L cell) | 0.0303 | 0.0230 |
| TEP (M/L cell) | 0.0263 | 0.0242 |
| $Ox_{capacity}$ (unitless) | 1 | 0.7482 |

where the left and right ventricles are modeled geometrically using thin-walled hemispheres with three segments: left ventricular free wall, right ventricular free wall, and the septum. In each of these segments, the implicit sarcomere and energetics model described above is utilized to calculate tension. Thus, sarcomere length and active force from the implicit myocyte model feed into the ventricular model, and the recomputed geometry is used to update sarcomere length in the implicit sarcomere model. This model is also coupled to a lumped-parameter circulation model, which represents the aorta, arteries, capillaries, and veins.

Mean sham rat data from Lopez *et al.* were used for all analysis in healthy simulations, and mean transverse aortic constriction (TAC) rat metabolite data were swapped in for heart failure simulations [73]. The optimized crossbridge parameters described above for the myocyte mechanics model were also used here (Table 2). We also updated the passive force formulation in the model (Equation 3.4, 3.5) to produce a more realistic end diastolic pressure volume relationship, based on [218], where $\gamma$ was set to 8 to produce a physiologic end-diastolic pressure-volume relationship (Equation 3.6).

$$\sigma_{passive}(SL) = k_{passive}(SL - SL_{rest}) + \sigma_{passive,collagen}(SL) \tag{3.4}$$

$$\sigma_{passive,collagen}(SL) = \left\{ \begin{array}{ll} Pcon_{collagen}[e^{PExp_{collagen}(SL-SL_{collagen})} - 1] & SL > SL_{collagen} \\ 0 & \text{otherwise} \end{array} \right\} \tag{3.5}$$

$$\sigma_{passive}(SL) = k_{passive}(SL - SL_{rest})^{\gamma} \tag{3.6}$$

$Ca^{2+}$ association parameters were adjusted to account for the differences between *in vivo* and *in vitro* data [217]. $K_{SE}$ was adjusted to $1 \times 10^4$ mmHg/$\mu$m, $k_{passive}$ was adjusted to $5.8 \times 10^4$ mmHg/$\mu$m, and $\eta$ was adjusted to $1 \times 10^{-4}$ (Table 3.2). Further, aortic compliance ($C_{Ao}$) was set to 0.0015 mL/mmHg and LV, RV, and septal midwall reference surface areas $Am_{ref}$ were scaled by a factor of 1.28 to match the wild-type experimental EF from [39]. All other parameters were kept unchanged, and dATP was simulated in the same way as in the myocyte model. The model was run for 120 beats to reach steady state as in [73]. Analysis was carried out in MATLAB R2018b [215]. Additional details on this model can be found in [33].

## 3.7 Acknowledgements

Chapter 3, in part, has been submitted for publication in *PNAS*, 2024. Multiscale modeling shows how 2'-deoxy-ATP rescues ventricular function in heart failure. A. E. Teitgen, M. T. Hock, K. J. McCabe, M. C. Childers, G. A. Huber, B. Marzban, D. A. Beard, J. A. McCammon, M. Regnier, A. D. McCulloch. The dissertation author was the second author and investigator of this paper. The dissertation author carried out the molecular dynamics analysis, Markov State Model construction, Brownian dynamics simulations, and contributed to the scientific process and writing.

# Chapter 4

# Conclusions

In this work, we present a cohesive model that uses molecular level simulation and analysis to predict whole organ function. This modeling framework focuses on the application of the myosin activator deoxy-ATP and its therapeutic potential to treat heart failure with reduced ejection fraction. However, the methods and interconnected models presented truly present their application in the ability to be modified for other targeted approaches to treat myopathies. For instance, at the molecular scale, the combination of molecular Markov motors, while not novel on its own, when combined with ensemble docking or Brownian dynamics simulations reveal previously uncovered insights. As such, the approaches used in this dissertation are immediately highly impactful even beyond the context of cardiovascular and musculoskeletal diseases focused on myosin modulators. The development of new computational resources, as well, based around new computing architectures such as GPU acceleration also increase the accessibility of our modeling approaches.

## 4.1 Key Findings

Molecular dynamics simulations demonstrated that dATP influences the structural dynamics of SERCA, which directly influences the calcium handling and subsequently diastolic function. When compared to ATP bound simulations of SERCA, dATP bound in a manner with stronger protein contacts and changes in the cytosolic domains of SERCA. Simultaneously,

dATP promoted changes in the calcium entry pathways to binding sites. Brownian dynamics simulations of calcium binding revealed that dATP increased accessibility to the binding sites results in faster calcium sequestration.

Using our newly developed software tool Netsci, we observed that dATP leads to changes in the correlated motion of SERCA relative to the binding of ATP. Specifically, dATP appears to aid SERCA in more efficient allosteric communication between the nucleotide binding site, the autophosphorylation site and the rest of the protein. Therefore, with a tighter network of communication, the SERCA pump is more readily able to open binding pathways for calcium. This helps to explain the changes in diastolic function based on the increased calcium flux from the cytosol into the sarcoplasmic reticulum.

Molecular dynamics simulations of pre-powerstroke myosin, led to altered conformational dynamics based on whether ATP or deoxy-ATP was bound. The simulations showed that myosin explored different conformational states based on the nucleotide bound. Further, these states were explored with different kinetics as measured using MSM analysis and validation. Matching out expectations, Brownian dynamics simulations showed that the different conformational states from the MSM different based on the state in the model, as well as the nucleotide bound, underscoring the importance of an ensemble-based approach to protein-protein docking.

Incorporating the observed changes in molecular behavior into filamentous and cellular models helped to match and explain changes to muscle mechanics. Crossbridge attachment rates contributed to a modest increase in contractility. The cooperative nature of thin filament activation helped to explain the non-linear increases in contractility based on dATP. However, modifications to the transitions in and out of the OFF state of myosin were required to match experimental measurements that show such significant changes in contractility.

Using an ordinary differential equation model, we recapitulated the cooperative and XB effects to simulate whole organ function and include calcium handling behavior. The changes to calcium sequestration are essential to capture the changes in diastolic function in addition to other clinical measurements of cardiac function. A metabolic model of heart failure also

supported the finding that dATP provides a larger measured increase in cardiovascular output and function than when dATP is applied to a healthy heart. This is compounded by increases in metabolic efficiency due to dATP. A coupled "tri-seg" model of whole heart function coupled to lumped parameter resistance model of circulation completed the molecule to whole organ framework and similarly predicted increases in circulatory function.

## 4.2 Limitations in modeling approaches

This work presents the first analysis, to our knowledge, that specifically models molecular mechanisms all the way up to whole organ and body analysis, and interconnects across each model step. However, like all models, this framework also has limits and assumptions in its approach. On the whole, we used a very targeted, 'depth-first' modeling approach which focused on bridging known and predicted changes across scales.

A breadth-first modeling approach, for instance, may have focused exclusively on molecular simulations of the effects of dATP in a wide range of crystal structures and structural states through cross-bridge cycle, and interacting heads motifs. However, we chose to focus our simulation analysis on crystal structures that were available in high resolution, and would mostly likely result in significant and observable changes. Even though structural resolution has been increasing, thanks in part to advances in cryo-EM approaches, the resolution of the myosin in different states does not each the desired threshold for accurate molecular modeling. Additionally, larger structures reduce the simulation and sampling available based on current computational resources. However, as structural biologists continue to push the boundaries of measuring protein structures in parallel with advances in computing power, we acknowledge such limitations will be more easily overcome.

Within the context of our molecular simulations, we acknowledge there were limitations and assumptions in our approach. Specifically, we assume that the ensemble based approach taken in the Brownian dynamics simulations are adequate to capture the necessary dynamics

involved in binding, both for actomyosin association and SERCA calcium handling. However, combined MD-BD approaches could further improve accuracy.

One area that we believe could use greater analysis is uncertainty quantification, especially as we step up in scales. Our modeling approach was based on limiting the number model parameters necessary in order match or compare against experimental data. However, future modeling would be well suited to focused on quantifying the uncertainty associated with particular model parameter changes. This type of error analysis would provide insights to highlight possible gaps or highly sensitive model parameters that should be further studied with additional experiments, or adjustments to the model framework. Specifically, the use of machine learning or artificial intelligence to direct parameter adjustment and error quantification could advance the development of this type of modeling significantly. For instance, using Bayesian parameter estimation with conditional variational autoencoders [219].

As computational biologists, our approach was to use all available data to help informs and build or multiscale model. However, this required integrating data sources from multiple different species and tissue preparations. Specifically, we integrated dATP modeling data from mouse measurements, skinned myofibrils, intact muscle and rat measurements. Ideally, there would be no inconsistency in the model species and would be from an animal model that is clinically relatively similar to humans, such as swine or bovine. However, these animal models are much more expensive, and take longer to develop compared to smaller rodent models.

Additionally, at the organ scale, when we model full circulatory function, a rat trans-aortic constriction model which focused on energetic remodeling. This is important to consider that the TAC model of heart failure may not be the best model of HFrEF, and is only one particular instances. In this modeling scheme, the metabolic changes were actually used as the underlying causes of the changes in function rather than a filament change or modification of the heart geometry.

## 4.3 Future Directions

The limitations discussed in the previous section all represent additional possible work that I or future researchers could follow. However, the modeling framework is it stands now is still robust and can provide significant and novel insights when applied in new contexts. Specifically, we believe that this framework could be used to model the effects of mysoin inhibitors in the context of hypertrophic cardiomyopathy. Because the key mechanisms of dATP appear to be in opposition to myosin inhibitors, we believe that the framework could be modified intuitively.

This multiscale modeling framework could also be utilized in a reverse mechanism drug discovery approach. Robust sensitivity across the model parameters could be used to identify potential mechanisms with high impact on organ function. This approach could be particularly useful in the context of modeling other cardiovascular diseases beyond a metabolic model of HF, and then identifying sensitive parameters and mechanisms afterwards. For example, using a geometry that reflects dilated cardiomyopathy, and adjusting constitutive properties of the organ model, we may find different mechanisms and model parameter sensitivities than with a health heart model. These mechanisms could then guide and identify specific protein targets.

Ultimately, this multiscale modeling approach, and depth of analysis, not only highlights the current mechanisms of dATP in the context of cardiovasuclar function, but provides a framework for future modeling endeavors of myosin modulators. We believe that developing detailed multiscale models are a necessary step to increasing the accessibility of personalized medicine and therefore leading to widespread improved clinical outcomes.

# Bibliography


[1] CDC, "Heart Disease Facts | cdc.gov," 2020.

[2] "Heart Disease Facts & Statistics | cdc.gov," 2018.

[3] B. A. Steinberg, X. Zhao, P. A. Heidenreich, E. D. Peterson, D. L. Bhatt, C. P. Cannon, A. F. Hernandez, G. C. Fonarow, and Get With the Guidelines Scientific Advisory Committee and Investigators, "Trends in patients hospitalized with heart failure and preserved left ventricular ejection fraction: prevalence, therapies, and outcomes," *Circulation*, vol. 126, pp. 65–75, 2012.

[4] B. Still, "Gearing Up for the Next Challenge in High-Performance Computing," 2015.

[5] J. Shalf, "The future of computing beyond Moore's Law," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 378, p. 20190061, 2020.

[6] G. E. Moore, "Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff.," *IEEE Solid-State Circuits Society Newsletter*, vol. 11, pp. 33–35, 2006.

[7] M. A. Geeves, "Stretching the lever-arm theory," *Nature*, vol. 415, pp. 129–131, 2002.

[8] R. V. Aroca and L. M. G. Gonçalves, "Towards green data centers: A comparison of x86 and ARM architectures power efficiency," *Journal of Parallel and Distributed Computing*, vol. 72, pp. 1770–1780, 2012.

[9] S. Govil, B. T. Crabb, Y. Deng, L. Dal Toso, E. Puyol-Antón, K. Pushparajah, S. Hegde, J. C. Perry, J. H. Omens, A. Hsiao, A. A. Young, and A. D. McCulloch, "A deep learning approach for fully automated cardiac shape modeling in tetralogy of Fallot," *Journal of Cardiovascular Magnetic Resonance*, vol. 25, p. 15, 2023.

[10] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, pp. 583–589, 2021.

[11] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, and S. Velankar, "AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models," *Nucleic Acids Research*, vol. 50, pp. D439–D444, 2022.

[12] H. Tada, N. Fujino, A. Nomura, C. Nakanishi, K. Hayashi, M. Takamura, and M.-a. Kawashiri, "Personalized medicine for cardiovascular diseases," *Journal of Human Genetics*, vol. 66, pp. 67–74, 2021.

[13] M. Birger, A. S. Kaldjian, G. A. Roth, A. E. Moran, J. L. Dieleman, and B. K. Bellows, "Spending on Cardiovascular Disease and Cardiovascular Risk Factors in the United States: 1996 to 2016," *Circulation*, vol. 144, pp. 271–282, 2021.

[14] S. S. Virani, A. Alonso, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, A. R. Chang, S. Cheng, F. N. Delling, L. Djousse, M. S. V. Elkind, J. F. Ferguson, M. Fornage, S. S. Khan, B. M. Kissela, K. L. Knutson, T. W. Kwan, D. T. Lackland, T. T. Lewis, J. H. Lichtman, C. T. Longenecker, M. S. Loop, P. L. Lutsey, S. S. Martin, K. Matsushita, A. E. Moran, M. E. Mussolino, A. M. Perak, W. D. Rosamond, G. A. Roth, U. K. A. Sampson, G. M. Satou, E. B. Schroeder, S. H. Shah, C. M. Shay, N. L. Spartano, A. Stokes, D. L. Tirschwell, L. B. VanWagner, C. W. Tsao, and American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee, "Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association," *Circulation*, vol. 141, pp. e139–e596, 2020.

[15] D. L. Mann and M. R. Bristow, "Mechanisms and Models in Heart Failure," *Circulation*, vol. 111, pp. 2837–2849, 2005.

[16] B. Bozkurt, A. J. S. Coats, H. Tsutsui, C. M. Abdelhamid, S. Adamopoulos, N. Albert, S. D. Anker, J. Atherton, M. Böhm, J. Butler, M. H. Drazner, G. Michael Felker, G. Filippatos, M. Fiuzat, G. C. Fonarow, J.-E. Gomez-Mesa, P. Heidenreich, T. Imamura, E. A. Jankowska, J. Januzzi, P. Khazanie, K. Kinugawa, C. S. P. Lam, Y. Matsue, M. Metra, T. Ohtani, M. Francesco Piepoli, P. Ponikowski, G. M. C. Rosano, Y. Sakata, P. Seferović, R. C. Starling, J. R. Teerlink, O. Vardeny, K. Yamamoto, C. Yancy, J. Zhang, and S. Zieroth, "Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure: Endorsed by the Canadian Heart Failure Society, Heart Failure Association of India, Cardiac Society of Australia and New Zealand, and Chinese Heart Failure Association," *European Journal of Heart Failure*, vol. 23, pp. 352–380, 2021.

[17] F. I. Malik, J. J. Hartman, K. A. Elias, B. P. Morgan, H. Rodriguez, K. Brejc, R. L. Anderson, S. H. Sueoka, K. H. Lee, J. T. Finer, R. Sakowicz, R. Baliga, D. R. Cox,

M. Garard, G. Godinez, R. Kawas, E. Kraynack, D. Lenzi, P. P. Lu, A. Muci, C. Niu, X. Qian, D. W. Pierce, M. Pokrovskii, I. Suehiro, S. Sylvester, T. Tochimoto, C. Valdez, W. Wang, T. Katori, D. A. Kass, Y.-T. Shen, S. F. Vatner, and D. J. Morgans, "Cardiac Myosin Activation: A Potential Therapeutic Approach for Systolic Heart Failure," *Science*, 2011.

[18] E. M. McNally and L. Mestroni, "Dilated Cardiomyopathy," *Circulation Research*, vol. 121, pp. 731–748, 2017.

[19] J. D. Powers, S. A. Malingen, M. Regnier, and T. L. Daniel, "The Sliding Filament Theory Since Andrew Huxley: Multiscale and Multidisciplinary Muscle Research," *Annual Review of Biophysics*, vol. 50, no. 1, p. null, 2021.

[20] A. Huxley, "Muscle Structure and Theories of Contraction," *Progress in Biophysics and Biophysical Chemistry*, vol. 7, pp. 255–318, 1957.

[21] S. M. Day, J. C. Tardiff, and E. M. Ostap, "Myosin modulators: emerging approaches for the treatment of cardiomyopathies and heart failure," *The Journal of Clinical Investigation*, vol. 132, 2022.

[22] Y.-J. Chen, C.-S. Chien, C.-E. Chiang, C.-H. Chen, and H.-M. Cheng, "From Genetic Mutations to Molecular Basis of Heart Failure Treatment: An Overview of the Mechanism and Implication of the Novel Modulators for Cardiac Myosin," *International Journal of Molecular Sciences*, vol. 22, p. 6617, 2021.

[23] J. R. Teerlink, R. Diaz, G. M. Felker, J. J. McMurray, M. Metra, S. D. Solomon, K. F. Adams, I. Anand, A. Arias-Mendoza, T. Biering-Sørensen, M. Böhm, D. Bonderman, J. G. Cleland, R. Corbalan, M. G. Crespo-Leiro, U. Dahlström, L. E. Echeverria, J. C. Fang, G. Filippatos, C. Fonseca, E. Goncalvesova, A. R. Goudev, J. G. Howlett, D. E. Lanfear, J. Li, M. Lund, P. Macdonald, V. Mareev, S.-i. Momomura, E. O'Meara, A. Parkhomenko, P. Ponikowski, F. J. Ramires, P. Serpytis, K. Sliwa, J. Spinar, T. M. Suter, J. Tomcsanyi, H. Vandekerckhove, D. Vinereanu, A. A. Voors, M. B. Yilmaz, F. Zannad, L. Sharpsten, J. C. Legg, C. Varin, N. Honarpour, S. A. Abbasi, F. I. Malik, and C. E. Kurtz, "Cardiac Myosin Activation with Omecamtiv Mecarbil in Systolic Heart Failure," *New England Journal of Medicine*, vol. 384, pp. 105–116, 2021.

[24] J. A. Spudich, "Three perspectives on the molecular basis of hypercontractility caused by hypertrophic cardiomyopathy mutations," *Pflügers Archiv - European Journal of Physiology*, vol. 471, pp. 701–717, 2019.

[25] Maron Barry J., "Clinical Course and Management of Hypertrophic Cardiomyopathy," *New England Journal of Medicine*, vol. 379, pp. 655–668, 2018.

[26] L. Sen-Martín, Fernández-Trasancos, M. López-Unzu, D. Pathak, A. Ferrarini, V. Labrador-Cantarero, D. Sánchez-Ortiz, M. R. Pricolo, N. Vicente, D. Velázquez-Carreras, L. Sánchez-García, J. Nicolás-Ávila, M. Sánchez-Díaz, S. Schlossarek, L. Cussó, M. Desco, M. Villalba-Orero, G. Guzmán-Martínez, E. Calvo, R. Barriales-Villa,

J. Vázquez, F. Sánchez-Cabo, A. Hidalgo, L. Carrier, J. A. Spudich, K. M. Ruppel, and J. Alegre-Cebollada, "Broad therapeutic benefit of myosin inhibition in hypertrophic cardiomyopathy," 2024.

[27] J. A. Spudich, "The myosin mesa and a possible unifying hypothesis for the molecular basis of human hypertrophic cardiomyopathy," *Biochemical Society transactions*, vol. 43, pp. 64–72, 2015.

[28] S. Nag, D. V. Trivedi, S. S. Sarkar, A. S. Adhikari, M. S. Sunitha, S. Sutton, K. M. Ruppel, and J. A. Spudich, "The myosin mesa and the basis of hypercontractility caused by hypertrophic cardiomyopathy mutations," *Nature Structural & Molecular Biology*, vol. 24, pp. 525–533, 2017.

[29] I. Olivotto, A. Oreziak, R. Barriales-Villa, T. P. Abraham, A. Masri, P. Garcia-Pavia, S. Saberi, N. K. Lakdawala, M. T. Wheeler, A. Owens, M. Kubanek, W. Wojakowski, M. K. Jensen, J. Gimeno-Blanes, K. Afshar, J. Myers, S. M. Hegde, S. D. Solomon, A. J. Sehnert, D. Zhang, W. Li, M. Bhattacharya, J. M. Edelberg, C. B. Waldman, S. J. Lester, A. Wang, C. Y. Ho, D. Jacoby, J. Bartunek, A. Bondue, E. V. Craenenbroeck, M. Kubanek, D. Zemanek, M. Jensen, J. Mogensen, J. J. Thune, P. Charron, A. Hagege, O. Lairez, J.-N. Trochu, C. Axthelm, H.-D. Duengen, N. Frey, V. Mitrovic, M. Preusch, J. Schulz-Menger, T. Seidler, M. Arad, M. Halabi, A. Katz, D. Monakier, O. Paz, S. Viskin, D. Zwas, I. Olivotto, H. P. B.-L. Rocca, M. Michels, D. Dudek, Z. Oko-Sarnowska, A. Oreziak, W. Wojakowski, N. Cardim, H. Pereira, R. Barriales-Villa, P. G. Pavia, J. G. Blanes, R. H. Urbano, L. M. R. Diaz, P. Elliott, Z. Yousef, T. Abraham, K. Afshar, P. Alvarez, R. Bach, R. Becker, L. Choudhury, D. Fermin, D. Jacoby, J. Jefferies, C. Kramer, N. Lakdawala, S. Lester, A. Marian, A. Masri, M. Maurer, S. Nagueh, A. Owens, D. Owens, F. Rader, S. Saberi, M. Sherrid, J. Shirani, J. Symanski, A. Turer, A. Wang, O. Wever-Pinzon, M. Wheeler, T. Wong, and M. Yamani, "Mavacamten for treatment of symptomatic obstructive hypertrophic cardiomyopathy (EXPLORER-HCM): a randomised, double-blind, placebo-controlled, phase 3 trial," *The Lancet*, vol. 396, pp. 759–769, 2020.

[30] M. Y. Desai, K. Wolski, A. Owens, S. S. Naidu, J. B. Geske, N. G. Smedira, H. Schaff, K. Lampl, E. McErlean, C. Sewell, D. Zhang, J. M. Edelberg, A. J. Sehnert, and S. E. Nissen, "Study design and rationale of VALOR-HCM: evaluation of mavacamten in adults with symptomatic obstructive hypertrophic cardiomyopathy who are eligible for septal reduction therapy," *American Heart Journal*, vol. 239, pp. 80–89, 2021.

[31] M. Y. Desai, A. Owens, K. Wolski, J. B. Geske, S. Saberi, A. Wang, M. Sherrid, P. C. Cremer, N. K. Lakdawala, A. Tower-Rader, D. Fermin, S. S. Naidu, N. G. Smedira, H. Schaff, E. McErlean, C. Sewell, L. Mudarris, Z. Gong, K. Lampl, A. J. Sehnert, and S. E. Nissen, "Mavacamten in Patients With Hypertrophic Cardiomyopathy Referred for Septal Reduction: Week 56 Results From the VALOR-HCM Randomized Clinical Trial," *JAMA cardiology*, vol. 8, pp. 968–977, 2023.

[32] M. Regnier, A. J. Rivera, Y. Chen, and P. B. Chase, "2-deoxy-ATP enhances contractility of rat cardiac muscle," *Circulation Research*, vol. 86, no. 12, pp. 1211–1217, 2000.

[33] B. Marzban, R. Lopez, and D. A. Beard, "Computational Modeling of Coupled Energetics and Mechanics in the Rat Ventricular Myocardium," 2020.

[34] D. A. Beard, B. Marzban, O. Y. Li, K. S. Campbell, P. M. L. Janssen, N. C. Chesler, and A. J. Baker, "Reduced Cardiac Muscle Power with low ATP simulating heart failure," *Biophysical Journal*, 2022.

[35] M. Kolberg, K. R. Strand, P. Graff, and K. K. Andersson, "Structure, function, and mechanism of ribonucleotide reductases," *Biochimica Et Biophysica Acta*, vol. 1699, pp. 1–34, 2004.

[36] S. D. Lundy, S. A. Murphy, S. K. Dupras, J. Dai, C. E. Murry, M. A. Laflamme, and M. Regnier, "Cell-based delivery of dATP via gap junctions enhances cardiac contractility," *Journal of Molecular and Cellular Cardiology*, vol. 72, pp. 350–359, 2014.

[37] J. D. Powers, C.-C. Yuan, K. J. McCabe, J. D. Murray, M. C. Childers, G. V. Flint, F. Moussavi-Harami, S. Mohran, R. Castillo, C. Zuzek, W. Ma, V. Daggett, A. D. McCulloch, T. C. Irving, and M. Regnier, "Cardiac myosin activation with 2-deoxy-ATP via increased electrostatic interactions with actin," *Proceedings of the National Academy of Sciences*, vol. 116, pp. 11502–11507, 2019.

[38] F. S. Korte, J. Dai, K. Buckley, E. R. Feest, N. Adamek, M. A. Geeves, C. E. Murry, and M. Regnier, "Upregulation of Cardiomyocyte Ribonucleotide Reductase Increases Intracellular 2 deoxy-ATP, Contractility, and Relaxation," *Journal of molecular and cellular cardiology*, vol. 51, pp. 894–901, 2011.

[39] S. G. Nowakowski, S. C. Kolwicz, F. S. Korte, Z. Luo, J. N. Robinson-Hamm, J. L. Page, F. Brozovich, R. S. Weiss, R. Tian, C. E. Murry, and M. Regnier, "Transgenic overexpression of ribonucleotide reductase improves cardiac performance," *Proceedings of the National Academy of Sciences*, 2013.

[40] S. Kadota, J. Carey, H. Reinecke, J. Leggett, S. Teichman, M. A. Laflamme, C. E. Murry, M. Regnier, and G. G. Mahairas, "Ribonucleotide reductase-mediated increase in dATP improves cardiac performance via myosin activation in a large animal model of heart failure," *European Journal of Heart Failure*, vol. 17, no. 8, pp. 772–781, 2015.

[41] S. G. Nowakowski, M. Regnier, and V. Daggett, "Molecular mechanisms underlying deoxy-ADP.Pi activation of pre-powerstroke myosin," *Protein Science*, vol. 26, no. 4, pp. 749–762, 2017.

[42] K. J. McCabe, Y. Aboelkassem, A. E. Teitgen, G. A. Huber, J. A. McCammon, M. Regnier, and A. D. McCulloch, "Predicting the effects of dATP on cardiac contraction using multiscale modeling of the sarcomere," *Archives of Biochemistry and Biophysics*, vol. 695, p. 108582, 2020.

[43] M. C. Childers and M. Regnier, "Using 2'-deoxy-ADP to probe stability of the myosin interacting heads motif at atomic resolution," *Biophysical Journal*, vol. 122, p. 169a, 2023.

[44] C. N. Toepfer, A. C. Garfinkel, G. Venturini, H. Wakimoto, G. Repetti, L. Alamo, A. Sharma, R. Agarwal, J. F. Ewoldt, P. Cloonan, J. Letendre, M. Lun, I. Olivotto, S. Colan, E. Ashley, D. Jacoby, M. Michels, C. S. Redwood, H. C. Watkins, S. M. Day, J. F. Staples, R. Padrón, A. Chopra, C. Y. Ho, C. S. Chen, A. C. Pereira, J. G. Seidman, and C. E. Seidman, "Myosin Sequestration Regulates Sarcomere Function, Cardiomyocyte Energetics, and Metabolism, Informing the Pathogenesis of Hypertrophic Cardiomyopathy," *Circulation*, vol. 141, pp. 828–842, 2020.

[45] C. T. A. Lewis, E. G. Melhedegaard, M. M. Ognjanovic, M. S. Olsen, J. Laitila, R. A. E. Seaborne, M. N. Grønset, C. Zhang, H. Iwamoto, A. L. Hessel, M. N. Kuehn, C. Merino, N. Amigó, O. Fröbert, S. Giroud, J. F. Staples, A. V. Goropashnaya, V. B. Fedorov, B. M. Barnes, Tøien, K. L. Drew, R. J. Sprenger, and J. Ochala, "Remodelling of skeletal muscle myosin metabolic states in hibernating mammals," *eLife*, vol. 13, 2024.

[46] K. Tran, D. S. Loiselle, and E. J. Crampin, "Regulation of cardiac cellular bioenergetics: mechanisms and consequences," *Physiological Reports*, vol. 3, 2015.

[47] B. E. Jaski, M. L. Jessup, D. M. Mancini, T. P. Cappola, D. F. Pauly, B. Greenberg, K. Borow, H. Dittrich, K. M. Zsebo, and R. J. Hajjar, "Calcium Upregulation by Percutaneous Administration of Gene Therapy in Cardiac Disease (CUPID Trial), a First-in-Human Phase 1/2 Clinical Trial," *Journal of Cardiac Failure*, vol. 15, pp. 171–181, 2009.

[48] M. Jessup, B. Greenberg, D. Mancini, T. Cappola, D. F. Pauly, B. Jaski, A. Yaroshinsky, K. M. Zsebo, H. Dittrich, and R. J. Hajjar, "Calcium Upregulation by Percutaneous Administration of Gene Therapy in Cardiac Disease (CUPID)," *Circulation*, vol. 124, pp. 304–313, 2011.

[49] B. Greenberg, J. Butler, G. M. Felker, P. Ponikowski, A. A. Voors, A. S. Desai, D. Barnard, A. Bouchard, B. Jaski, A. R. Lyon, J. M. Pogoda, J. J. Rudy, and K. M. Zsebo, "Calcium upregulation by percutaneous administration of gene therapy in patients with cardiac disease (CUPID 2): a randomised, multinational, double-blind, placebo-controlled, phase 2b trial," *The Lancet*, vol. 387, pp. 1178–1186, 2016.

[50] K. S. Thomson, G. L. Odom, C. E. Murry, G. G. Mahairas, F. Moussavi-Harami, S. L. Teichman, X. Chen, S. D. Hauschka, J. S. Chamberlain, and M. Regnier, "Translation of Cardiac Myosin Activation with 2-deoxy-ATP to Treat Heart Failure via an Experimental Ribonucleotide Reductase-Based Gene Therapy," *JACC. Basic to translational science*, vol. 1, pp. 666–679, 2016.

[51] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis," *Nature*, vol. 181, pp. 662–666, 1958.

[52] Q.-F. Yang and C. Tang, "On the necessity of an integrative approach to understand protein structural dynamics," *Journal of Zhejiang University. Science. B*, vol. 20, pp. 496–502, 2019.

[53] J. D. Durrant and J. A. McCammon, "Molecular dynamics simulations and drug discovery," *BMC Biology*, vol. 9, p. 71, 2011.

[54] S. A. Hollingsworth and R. O. Dror, "Molecular Dynamics Simulation for All," *Neuron*, vol. 99, pp. 1129–1143, 2018.

[55] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, "Long-timescale molecular dynamics simulations of protein structure and function," *Current Opinion in Structural Biology*, vol. 19, pp. 120–127, 2009.

[56] D. Hamelberg, J. Mongan, and J. A. McCammon, "Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules," *The Journal of Chemical Physics*, vol. 120, pp. 11919–11929, 2004.

[57] L. W. Votapka, A. M. Stokely, A. A. Ojha, and R. E. Amaro, "SEEKR2: Versatile Multiscale Milestoning Utilizing the OpenMM Molecular Dynamics Engine," *Journal of Chemical Information and Modeling*, vol. 62, pp. 3253–3262, 2022.

[58] T. Ichiye and M. Karplus, "Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations," *Proteins: Structure, Function, and Bioinformatics*, vol. 11, no. 3, pp. 205–217, 1991.

[59] O. F. Lange and H. Grubmüller, "Generalized correlation for biomolecular dynamics," *Proteins: Structure, Function, and Bioinformatics*, vol. 62, no. 4, pp. 1053–1061, 2006.

[60] NVIDIA, P. Vingelmann, and F. H. Fitzek, "CUDA, release: 11.7," 2022.

[61] V. S. Pande, K. Beauchamp, and G. R. Bowman, "Everything you wanted to know about Markov State Models but were afraid to ask," *Methods (San Diego, Calif.)*, vol. 52, pp. 99–105, 2010.

[62] B. E. Husic and V. S. Pande, "Markov State Models: From an Art to a Science," *Journal of the American Chemical Society*, vol. 140, pp. 2386–2396, 2018.

[63] R. Zwanzig, "From classical dynamics to continuous time random walks," *Journal of Statistical Physics*, vol. 30, pp. 255–262, 1983.

[64] J. D. Chodera and F. Noé, "Markov state models of biomolecular conformational dynamics," *Current opinion in structural biology*, vol. 25, pp. 135–144, 2014.

[65] S. H. Northrup, S. A. Allison, and J. A. McCammon, "Brownian dynamics simulation of diffusion-influenced bimolecular reactions," *The Journal of Chemical Physics*, vol. 80, pp. 1517–1524, 1984.

[66] A. Muñiz-Chicharro, L. W. Votapka, R. E. Amaro, and R. C. Wade, "Brownian dynamics simulations of biomolecular diffusional association processes," *WIREs Computational Molecular Science*, vol. n/a, no. n/a, p. e1649.

[67] F. Fogolari, A. Brigo, and H. Molinari, "The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology," *Journal of molecular recognition: JMR*, vol. 15, pp. 377–392, 2002.

[68] G. A. Huber and J. A. McCammon, "Browndye: A software package for Brownian dynamics," *Computer Physics Communications*, vol. 181, pp. 1896–1905, 2010.

[69] A. F. Huxley and R. Niedergerke, "Structural Changes in Muscle During Contraction: Interference Microscopy of Living Muscle Fibres," *Nature*, vol. 173, pp. 971–973, 1954.

[70] H. Huxley and J. Hanson, "Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation," *Nature*, vol. 173, pp. 973–976, 1954.

[71] J. D. Powers, C. D. Williams, M. Regnier, and T. L. Daniel, "A Spatially Explicit Model Shows How Titin Stiffness Modulates Muscle Mechanics and Energetics," *Integrative and Comparative Biology*, vol. 58, pp. 186–193, 2018.

[72] C. D. Williams, M. Regnier, and T. L. Daniel, "Axial and Radial Forces of Cross-Bridges Depend on Lattice Spacing," *PLOS Computational Biology*, vol. 6, p. e1001018, 2010.

[73] R. Lopez, B. Marzban, X. Gao, E. Lauinger, F. Van den Bergh, S. E. Whitesall, K. Converso-Baran, C. F. Burant, D. E. Michele, and D. A. Beard, "Impaired Myocardial Energetics Causes Mechanical Dysfunction in Decompensated Failing Hearts," *Function*, vol. 1, p. zqaa018, 2020.

[74] E. B. Randall, M. Hock, R. Lopez, B. Marzban, C. Marshall, and D. A. Beard, "Quantitative analysis of mitochondrial ATP synthesis," *Mathematical Biosciences*, p. 108646, 2021.

[75] R. E. Amaro and A. J. Mulholland, "Multiscale methods in drug design bridge chemical and biological complexity in the search for cures," *Nature Reviews Chemistry*, vol. 2, pp. 1–12, 2018.

[76] J. Lumens, T. Delhaas, B. Kirn, and T. Arts, "Three-Wall Segment (TriSeg) Model Describing Mechanics and Hemodynamics of Ventricular Interaction," *Annals of Biomedical Engineering*, vol. 37, pp. 2234–2255, 2009.

[77] J. Bazil, D. Beard, and K. Vinnakota, "Catalytic Coupling of Oxidative Phosphorylation, ATP Demand, and Reactive Oxygen Species Generation," *Biophysical Journal*, vol. 110, pp. 962–971, 2016.

[78] M. G. Palmgren and P. Nissen, "P-Type ATPases," *Annual Review of Biophysics*, vol. 40, pp. 243–266, 2011.

[79] M. Brini and E. Carafoli, "Calcium Pumps in Health and Disease," *Physiological Reviews*, vol. 89, pp. 1341–1378, 2009.

[80] T. L.-M. Sørensen, J. V. Møller, and P. Nissen, "Phosphoryl Transfer and Calcium Ion Occlusion in the Calcium Pump," *Science*, vol. 304, pp. 1672–1675, 2004.

[81] C. Toyoshima, "Structural aspects of ion pumping by Ca2+-ATPase of sarcoplasmic reticulum," *Archives of Biochemistry and Biophysics*, vol. 476, pp. 3–11, 2008.

[82] D. M. Clarke, T. W. Loo, G. Inesi, and D. H. MacLennan, "Location of high affinity Ca2 +-binding sites within the predicted transmembrahe domain of the sarco-plasmic reticulum Ca2+-ATPase," *Nature*, vol. 339, pp. 476–478, 1989.

[83] C. Olesen, T. L.-M. Sørensen, R. C. Nielsen, J. V. Møller, and P. Nissen, "Dephosphorylation of the Calcium Pump Coupled to Counterion Occlusion," *Science*, vol. 306, pp. 2251–2255, 2004.

[84] M. Bublitz, M. Musgaard, H. Poulsen, L. Thøgersen, C. Olesen, B. Schiøtt, J. P. Morth, J. V. Møller, and P. Nissen, "Ion Pathways in the Sarcoplasmic Reticulum Ca2+-ATPase *," *Journal of Biological Chemistry*, vol. 288, pp. 10759–10765, 2013.

[85] K. Tran, N. P. Smith, D. S. Loiselle, and E. J. Crampin, "A Thermodynamic Model of the Cardiac Sarcoplasmic/Endoplasmic Ca2+ (SERCA) Pump," *Biophysical Journal*, vol. 96, pp. 2029–2042, 2009.

[86] Regnier M., Rivera A. J., Chen Y., and Chase P. B., "2-Deoxy-ATP Enhances Contractility of Rat Cardiac Muscle," *Circulation Research*, vol. 86, pp. 1211–1217, 2000.

[87] M. Regnier, H. Martin, R. J. Barsotti, A. J. Rivera, D. A. Martyn, and E. Clemmens, "Cross-bridge versus thin filament contributions to the level and rate of force development in cardiac muscle," *Biophysical Journal*, vol. 87, pp. 1815–1824, 2004.

[88] B. Schoffstall, A. Clark, and P. B. Chase, "Positive Inotropic Effects of Low dATP/ATP Ratios on Mechanics and Kinetics of Porcine Cardiac Muscle," *Biophysical Journal*, vol. 91, pp. 2216–2226, 2006.

[89] S. C. Kolwicz, G. L. Odom, S. G. Nowakowski, F. Moussavi-Harami, X. Chen, H. Reinecke, S. D. Hauschka, C. E. Murry, G. G. Mahairas, and M. Regnier, "AAV6-mediated Cardiac-specific Overexpression of Ribonucleotide Reductase Enhances Myocardial Contractility," *Molecular Therapy*, vol. 24, pp. 240–250, 2016.

[90] M. Periasamy and S. Huke, "SERCA Pump Level is a Critical Determinant of Ca2+Homeostasis and Cardiac Contractility," *Journal of Molecular and Cellular Cardiology*, vol. 33, pp. 1053–1063, 2001.

[91] S. M. Pogwizd, K. Schlotthauer, L. Li, W. Yuan, and D. M. Bers, "Arrhythmogenesis and Contractile Dysfunction in Heart Failure," *Circulation Research*, vol. 88, pp. 1159–1167, 2001.

[92] L. Zhihao, N. Jingyu, L. Lan, S. Michael, G. Rui, B. Xiyun, L. Xiaozhi, and F. Guanwei, "SERCA2a: a key protein in the Ca2+ cycle of the heart failure," *Heart Failure Reviews*, vol. 25, pp. 523–535, 2020.

[93] L. J. Motloch, M. Cacheux, K. Ishikawa, C. Xie, J. Hu, J. Aguero, K. M. Fish, R. J. Hajjar, and F. G. Akar, "Primary Effect of SERCA2a Gene Transfer on Conduction Reserve in Chronic Myocardial Infarction," *Journal of the American Heart Association*, vol. 7, p. e009598, 2018.

[94] C. Hayward, N. R. Banner, A. Morley-Smith, A. R. Lyon, and S. E. Harding, "The Current and Future Landscape of SERCA Gene Therapy for Heart Failure: A Clinical Perspective," *Human Gene Therapy*, vol. 26, pp. 293–304, 2015.

[95] T. Tsuji, F. del Monte, Y. Yoshikawa, T. Abe, J. Shimizu, C. Nakajima-Takenaka, S. Taniguchi, R. J. Hajjar, and M. Takaki, "Rescue of Ca2+ overload-induced left ventriclur dysfunction by targeted ablation of phospholamban," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 296, pp. H310–H317, 2009.

[96] Y. Miao, V. A. Feher, and J. A. McCammon, "Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation," *Journal of Chemical Theory and Computation*, vol. 11, pp. 3584–3595, 2015.

[97] J. Wang, P. R. Arantes, A. Bhattarai, R. V. Hsu, S. Pawnikar, Y.-m. M. Huang, G. Palermo, and Y. Miao, "Gaussian accelerated molecular dynamics: Principles and applications," *WIREs Computational Molecular Science*, vol. 11, no. 5, p. e1521, 2021.

[98] C. Toyoshima, S. Iwasawa, H. Ogawa, A. Hirata, J. Tsueda, and G. Inesi, "Crystal structures of the calcium pump and sarcolipin in the Mg2+-bound E1 state," *Nature*, vol. 495, pp. 260–264, 2013.

[99] Y. Kabashima, H. Ogawa, R. Nakajima, and C. Toyoshima, "What ATP binding does to the Ca2+ pump and how nonproductive phosphoryl transfer is prevented in the absence of Ca2+," *Proceedings of the National Academy of Sciences*, vol. 117, pp. 18448–18458, 2020.

[100] Y. Himeno, K. Asakura, C. Cha, H. Memida, T. Powell, A. Amano, and A. Noma, "A Human Ventricular Myocyte Model with a Refined Representation of Excitation-Contraction Coupling," *Biophysical Journal*, vol. 109, pp. 415–427, 2015.

[101] H. M. Berman, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, pp. 235–242, 2000.

[102] T. Schwede, J. Kopp, N. Guex, and M. C. Peitsch, "SWISS-MODEL: an automated protein homology-modeling server," *Nucleic Acids Research*, vol. 31, pp. 3381–3385, 2003.

[103] F. Sievers and D. G. Higgins, "Clustal Omega for making accurate alignments of many protein sequences," *Protein Science*, vol. 27, no. 1, pp. 135–145, 2018.

[104] J. Lee, D. S. Patel, J. Ståhle, S.-J. Park, N. R. Kern, S. Kim, J. Lee, X. Cheng, M. A. Valvano, O. Holst, Y. A. Knirel, Y. Qi, S. Jo, J. B. Klauda, G. Widmalm, and W. Im, "CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans," *Journal of Chemical Theory and Computation*, vol. 15, pp. 775–786, 2019.

[105] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker, "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations," *Nucleic Acids Research*, vol. 32, pp. W665–667, 2004.

[106] C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski, and J. H. Jensen, "Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values," *Journal of Chemical Theory and Computation*, vol. 7, pp. 2284–2295, 2011.

[107] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen, "PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions," *Journal of Chemical Theory and Computation*, vol. 7, pp. 525–537, 2011.

[108] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker, "PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations," *Nucleic Acids Research*, vol. 35, pp. W522–525, 2007.

[109] M. A. Lomize, A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, "OPM: Orientations of Proteins in Membranes database," *Bioinformatics*, vol. 22, pp. 623–625, 2006.

[110] R. Bick, L. Buja, W. Van Winkle, and G. Taffet, "Membrane Asymmetry in Isolated Canine Cardiac Sarcoplasmic Reticulum: Comparison with Skeletal Muscle Sarcoplasmic Reticulum," *The Journal of Membrane Biology*, vol. 164, pp. 169–175, 1998.

[111] H. E. Autzen and M. Musgaard, "MD Simulations of P-Type ATPases in a Lipid Bilayer System," in *P-Type ATPases: Methods and Protocols* (M. Bublitz, ed.), Methods in Molecular Biology, pp. 459–492, New York, NY: Springer, 2016.

[112] S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: A web-based graphical user interface for CHARMM," *Journal of Computational Chemistry*, vol. 29, no. 11, pp. 1859–1865, 2008.

[113] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF Chimera—A visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[114] X. He, V. H. Man, W. Yang, T.-S. Lee, and J. Wang, "A fast and high-quality charge model for the next generation general AMBER force field," *The Journal of Chemical Physics*, vol. 153, p. 114502, 2020.

[115] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," *Journal of Molecular Graphics and Modelling*, vol. 25, pp. 247–260, 2006.

[116] C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu, and C. Simmerling, "ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution," *Journal of Chemical Theory and Computation*, vol. 16, pp. 528–552, 2020.

[117] S. Izadi, R. Anandakrishnan, and A. V. Onufriev, "Building Water Models: A Different Approach," *The Journal of Physical Chemistry Letters*, vol. 5, pp. 3863–3871, 2014.

[118] S. Miyamoto and P. A. Kollman, "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models," *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 952–962, 1992.

[119] D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York, and P. A. Kollman, "AMBER 20," *University of California, San Francisco.*, 2020.

[120] S. D. S. Center, "Triton Shared Computing Cluster," 2022.

[121] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.

[122] Schrödinger, LLC, "The PyMOL Molecular Graphics System, Version 1.8." 2015.

[123] D. R. Roe and T. E. I. Cheatham, "PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data," *Journal of Chemical Theory and Computation*, vol. 9, pp. 3084–3095, 2013.

[124] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, and V. S. Pande, "MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories," *Biophysical Journal*, vol. 109, pp. 1528–1532, 2015.

[125] Y. Miao, W. Sinko, L. Pierce, D. Bucher, R. C. Walker, and J. A. McCammon, "Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation," *Journal of Chemical Theory and Computation*, vol. 10, pp. 2677–2689, 2014.

[126] C. Toyoshima, M. Nakasako, H. Nomura, and H. Ogawa, "Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution," *Nature*, vol. 405, pp. 647–655, 2000.

[127] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, pp. 1942–1948 vol.4, 1995.

[128] G. Inesi, D. Lewis, C. Toyoshima, A. Hirata, and L. de Meis, "Conformational Fluctuations of the Ca2+-ATPase in the Native Membrane Environment," *Journal of Biological Chemistry*, vol. 283, pp. 1189–1196, 2008.

[129] S. Hua, H. Ma, D. Lewis, G. Inesi, and C. Toyoshima, "Functional Role of "N" (Nucleotide) and "P" (Phosphorylation) Domain Interactions in the Sarcoplasmic Reticulum (SERCA) ATPase," *Biochemistry*, vol. 41, pp. 2264–2272, 2002.

[130] J. D. Clausen, D. B. McIntosh, B. Vilsen, D. G. Woolley, and J. P. Andersen, "Importance of Conserved N-domain Residues Thr441, Glu442, Lys515, Arg560, and Leu562 of Sarcoplasmic Reticulum Ca2+-ATPase for MgATP Binding and Subsequent Catalytic Steps: PLASTICITY OF THE NUCLEOTIDE-BINDING SITE *," *Journal of Biological Chemistry*, vol. 278, pp. 20245–20258, 2003.

[131] B. Mueller, M. Zhao, I. V. Negrashov, R. Bennett, and D. D. Thomas, "SERCA Structural Dynamics Induced by ATP and Calcium," *Biochemistry*, vol. 43, pp. 12846–12854, 2004.

[132] Y. Huang, H. Li, and Y. Bu, "Molecular dynamics simulation exploration of cooperative migration mechanism of calcium ions in sarcoplasmic reticulum Ca2+-ATPase," *Journal of Computational Chemistry*, vol. 30, no. 13, pp. 2136–2145, 2009.

[133] G. Inesi, M. Kurzmack, C. Coan, and D. E. Lewis, "Cooperative calcium binding and ATPase activation in sarcoplasmic reticulum vesicles.," *Journal of Biological Chemistry*, vol. 255, pp. 3025–3031, 1980.

[134] Z. Zhang, D. Lewis, C. Strock, G. Inesi, M. Nakasako, H. Nomura, and C. Toyoshima, "Detailed Characterization of the Cooperative Mechanism of Ca2+ Binding and Catalytic Activation in the Ca2+ Transport (SERCA) ATPase," *Biochemistry*, vol. 39, pp. 8758–8767, 2000.

[135] P. M. Kekenes-Huskey, V. T. Metzger, B. J. Grant, and J. Andrew McCammon, "Calcium binding and allosteric signaling mechanisms for the sarcoplasmic reticulum Ca²+ ATPase," *Protein Science: A Publication of the Protein Society*, vol. 21, pp. 1429–1443, 2012.

[136] A. P. Einholm, B. Vilsen, and J. P. Andersen, "Importance of Transmembrane Segment M1 of the Sarcoplasmic Reticulum Ca2+-ATPase in Ca2+ Occlusion and Phosphoenzyme Processing*," *Journal of Biological Chemistry*, vol. 279, pp. 15888–15896, 2004.

[137] A. G. Lee and J. M. East, "What the structure of a calcium pump tells us about its mechanism," *Biochemical Journal*, vol. 356, pp. 665–683, 2001.

[138] M. Musgaard, L. Thøgersen, B. Schiøtt, and E. Tajkhorshid, "Tracing Cytoplasmic Ca2+ Ion and Water Access Points in the Ca2+-ATPase," *Biophysical Journal*, vol. 102, pp. 268–277, 2012.

[139] P. K. Agarwal, S. R. Billeter, P. R. Rajagopalan, S. J. Benkovic, and S. Hammes-Schiffer, "Network of coupled promoting motions in enzyme catalysis," *Proceedings of the National Academy of Sciences*, vol. 99, no. 5, pp. 2794–2799, 2002.

[140] J. D. Wichard, R. Kuhne, and A. ter Laak, "Binding site detection via mutual information," in *2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*, pp. 1770–1776, IEEE, 2008.

[141] A. Sethi, J. Eargle, A. A. Black, and Z. Luthey-Schulten, "Dynamical networks in tRNA: protein complexes," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6620–6625, 2009.

[142] A. T. VanWart, J. Eargle, Z. Luthey-Schulten, and R. E. Amaro, "Exploring residue component contributions to dynamical network models of allostery," *Journal of chemical theory and computation*, vol. 8, no. 8, pp. 2949–2961, 2012.

[143] P. M. Gasper, B. Fuglestad, E. A. Komives, P. R. L. Markwick, and J. A. McCammon, "Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 21216–21222, 2012.

[144] M. C. R. Melo, R. C. Bernardi, C. de la Fuente-Nunez, and Z. Luthey-Schulten, "Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories," *The Journal of Chemical Physics*, vol. 153, p. 134104, 2020.

[145] F. Maschietto, B. Allen, G. Kyro, and V. Batista, "MDiGest: A Python package for describing allostery from molecular dynamics simulations," *The Journal of Chemical Physics*, vol. 158, 2023.

[146] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[147] H. Akaike, "A new look at the statistical model identification," *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.

[148] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[149] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, "Conditional likelihood maximisation: a unifying framework for information theoretic feature selection," *The journal of machine learning research*, vol. 13, pp. 27–66, 2012.

[150] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: is a correction for chance necessary?," in *Proceedings of the 26th annual international conference on machine learning*, pp. 1073–1080, 2009.

[151] S. Roberts and R. Everson, *Independent component analysis: principles and practice*. Cambridge University Press, 2001.

[152] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[153] Y.-I. Moon, B. Rajagopalan, and U. Lall, "Estimation of mutual information using kernel density estimators," *Physical Review E*, vol. 52, no. 3, p. 2318, 1995.

[154] R. Steuer, J. Kurths, C. O. Daub, J. Weise, and J. Selbig, "The mutual information: detecting and evaluating dependencies between variables," *Bioinformatics*, vol. 18, no. suppl_2, pp. S231–S240, 2002.

[155] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical review A*, vol. 33, no. 2, p. 1134, 1986.

[156] G. A. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1315–1321, 1999.

[157] P. Grassberger, "Finite sample corrections to entropy and dimension estimates," *Physics Letters A*, vol. 128, no. 6-7, pp. 369–373, 1988.

[158] M. S. Roulston, "Estimating the errors on measured entropy and mutual information," *Physica D: Nonlinear Phenomena*, vol. 125, no. 3-4, pp. 285–294, 1999.

[159] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 69, no. 6, p. 16, 2004.

[160] Z. Yao and W. L. Ruzzo, "A regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data," in *BMC bioinformatics*, vol. 7, pp. 1–11, BioMed Central, 2006.

[161] L. I. Shachaf, E. Roberts, P. Cahan, and J. Xiao, "Gene regulation network inference using k-nearest neighbor-based mutual information estimation: revisiting an old DREAM," *BMC bioinformatics*, vol. 24, no. 1, p. 84, 2023.

[162] A. Das, H. Rui, R. Nakamoto, and B. Roux, "Conformational Transitions and Alternating-Access Mechanism in the Sarcoplasmic Reticulum Calcium Pump," *Journal of Molecular Biology*, vol. 429, pp. 647–666, 2017.

[163] G. Inesi, A. M. Prasad, and R. Pilankatta, "The Ca2+ ATPase of cardiac sarcoplasmic reticulum: Physiological role and relevance to diseases," *Biochemical and Biophysical Research Communications*, vol. 369, pp. 182–187, 2008.

[164] M. T. Hock, A. E. Teitgen, K. J. McCabe, S. P. Hirakis, G. A. Huber, M. Regnier, R. E. Amaro, J. A. McCammon, and A. D. McCulloch, "Multiscale computational modeling of the effects of 2'-deoxy-ATP on cardiac muscle calcium handling," *Journal of Applied Physics*, vol. 134, p. 074905, 2023.

[165] B. Vilsen and J. P. Andersen, "CrATP-induced Ca2+ occlusion in mutants of the Ca(2+)-ATPase of sarcoplasmic reticulum," *The Journal of Biological Chemistry*, vol. 267, pp. 25739–25743, 1992.

[166] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, J. T. Berryman, S. R. Brozell, D. S. Cerutti, I. T.E. Cheatham, G. A. Cisneros, V. W. D. Cruzeiro, T. A. Darden, N. Forouzesh, G. Giambaşu, T. Giese, M. K. Gilson, H. Gohlke, A. W. Goetz, J. Harris, Z. Huang, S. Izadi, S. A. Izmailov, K. Kasavajhala, M. C. Kaymak, A. Kovalenko, T. Kurtzman, T. S. Lee, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, M. Manathunga, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, A. Rahnamoun, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, J. Wang, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D. M. York, C. Zhao, Q. Zhu, and P. A. Kollman, "Amber 2024," 2024.

[167] J. Stone, "An Efficient Library for Parallel Ray Tracing and Animation," Master's thesis, Computer Science Department, University of Missouri-Rolla, 1998.

[168] W. Ma, M. Childers, J. Murray, F. Moussavi-Harami, H. Gong, R. Weiss, V. Daggett, T. Irving, and M. Regnier, "Myosin dynamics during relaxation in mouse soleus muscle and modulation by 2-deoxy-ATP," *The Journal of Physiology*, vol. 598, no. 22, pp. 5165–5182, 2020.

[169] W. Ma, T. S. McMillen, M. C. Childers, H. Gong, M. Regnier, and T. Irving, "Structural OFF/ON transitions of myosin in relaxed porcine myocardium predict calcium-activated force," *Proceedings of the National Academy of Sciences*, vol. 120, p. e2207615120, 2023.

[170] J. Walklate, K. Kao, M. Regnier, and M. A. Geeves, "Exploring the super-relaxed state of myosin in myofibrils from fast-twitch, slow-twitch, and cardiac muscle," *Journal of Biological Chemistry*, vol. 298, no. 3, p. 101640, 2022.

[171] F. Wu, E. Y. Zhang, J. Zhang, R. J. Bache, and D. A. Beard, "Phosphate metabolite concentrations and ATP hydrolysis potential in normal and ischaemic hearts," *The Journal of Physiology*, vol. 586, pp. 4193–4208, 2008.

[172] T. Doenst, T. D. Nguyen, and E. D. Abel, "Cardiac metabolism in heart failure: Implications beyond atp production," *Circulation Research*, vol. 113, no. 6, pp. 709–724, 2013.

[173] P. B. Joel, H. L. Sweeney, and K. M. Trybus, "Addition of Lysines to the 50/20 kDa Junction of Myosin Strengthens Weak Binding to Actin without Affecting the Maximum ATPase Activity," *Biochemistry*, vol. 42, pp. 9160–9166, 2003.

[174] M. H. Doran, M. J. Rynkiewicz, E. Pavadai, S. M. Bodt, D. Rasicci, J. R. Moore, C. M. Yengo, E. Bullitt, and W. Lehman, "Myosin loop-4 is critical for optimal tropomyosin repositioning on actin during muscle activation and relaxation," *Journal of General Physiology*, vol. 155, p. e202213274, 2022.

[175] M. Lorenz and K. C. Holmes, "The actin-myosin interface," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 12529–12534, 2010.

[176] M. A. Geeves, "Dynamic interaction between actin and myosin subfragment 1 in the presence of ADP," *Biochemistry*, vol. 28, pp. 5864–5871, 1989.

[177] H. D. White and E. W. Taylor, "Energetics and mechanism of actomyosin adenosine triphosphatase," *Biochemistry*, vol. 15, pp. 5818–5826, 1976.

[178] S. G. Campbell, F. V. Lionetti, K. S. Campbell, and A. D. McCulloch, "Coupling of Adjacent Tropomyosins Enhances Cross-Bridge-Mediated Cooperative Activation in a Markov Model of the Cardiac Thin Filament," *Biophysical Journal*, vol. 98, pp. 2254–2264, 2010.

[179] K. S. Campbell, P. M. L. Janssen, and S. G. Campbell, "Force-Dependent recruitment from the myosin off state contributes to length-dependent activation," *Biophysical Journal*, vol. 115, no. 3, pp. 543–553, 2018.

[180] K. S. Campbell, B. S. Christman, and S. G. Campbell, "Multiscale modeling of cardiovascular function predicts that end-systolic pressure volume relationship can be targeted via multiple therapeutic strategies," *Frontiers in Physiology*, vol. 11, 2020.

[181] M. C. Childers, M. Geeves, V. Daggett, and M. Regnier, "Modulation of post-powerstroke dynamics in myosin II by 2'- deoxy-ADP," *Archives of Biochemistry and Biophysics*, vol. 699, p. 108733, 2021.

[182] J. R. Porter, A. Meller, M. I. Zimmerman, M. J. Greenberg, and G. R. Bowman, "Conformational distributions of isolated myosin motor domains encode their mechanochemical properties," *eLife*, vol. 9, p. e55132, 2020.

[183] R. Clark, M. A. Ansari, S. Dash, M. A. Geeves, and L. M. Coluccio, "Loop 1 of transducer region in mammalian class I myosin, Myo1b, modulates actin affinity, ATPase activity, and nucleotide access," *The Journal of Biological Chemistry*, vol. 280, pp. 30935–30942, 2005.

[184] M. Linari, E. Brunello, M. Reconditi, L. Fusi, M. Caremani, T. Narayanan, G. Piazzesi, L. Vincenzo, and M. Irving, "Force generation by skeletal muscle is controlled by mechanosensing in myosin filaments," *Nature*, vol. 528, pp. 276–279, 2015.

[185] K. S. Campbell, "Impact of myocyte strain on cardiac myofilament activation," *Pflügers Archiv: European Journal of Physiology*, vol. 462, no. 1, pp. 3–14, 2011.

[186] S. Ponnam and T. Kampourakis, "Microscale thermophoresis suggests a new model of regulation of cardiac myosin function via interaction with cardiac myosin-binding protein C," *The Journal of Biological Chemistry*, vol. 298, p. 101485, 2021.

[187] L. Alamo, W. Wriggers, A. Pinto, F. Bártoli, L. Salazar, F.-Q. Zhao, R. Craig, and R. Padrón, "Three-Dimensional Reconstruction of Tarantula Myosin Filaments Suggests How Phosphorylation May Regulate Myosin Activity," *Journal of Molecular Biology*, vol. 384, pp. 780–797, 2008.

[188] S. G. Tewari, S. M. Bugenhagen, K. C. Vinnakota, J. J. Rice, P. M. Janssen, and D. A. Beard, "Influence of metabolic dysfunction on cardiac mechanics in decompensated hypertrophy and heart failure," *Journal of molecular and cellular cardiology*, vol. 94, pp. 162–175, 2016.

[189] R. Gordan, J. K. Gwathmey, and L.-H. Xie, "Autonomic and endocrine control of cardiovascular function," *World Journal of Cardiology*, vol. 7, no. 4, pp. 204–214, 2015.

[190] S. G. Tewari, S. M. Bugenhagen, B. M. Palmer, and D. A. Beard, "Dynamics of cross-bridge cycling, ATP hydrolysis, force generation, and deformation in cardiac muscle," *Journal of Molecular and Cellular Cardiology*, vol. 96, pp. 11–25, 2016.

[191] S. Gourinath, D. M. Himmel, J. H. Brown, L. Reshetnikova, A. G. Szent-Gyorgyi, and C. Cohen, "Crystal structure of scallop myosin S1 in the pre-power stroke state to 2.6 Å resolution: flexibility and function in the head," *Structure*, vol. 11, no. 12, pp. 1621–1627, 2003.

[192] C. C. Goodno, "Inhibition of myosin ATPase by vanadate ion," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, no. 6, pp. 2620–2624, 1979.

[193] B. Webb and A. Sali, "Comparative protein structure modeling using MODELLER," *Current Protocols in Protein Science*, vol. 86, no. 2.9.1–2.9.37, 2016.

[194] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, "An overview of the Amber biomolecular simulation package," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 3, no. 2, pp. 198–210, 2013.

[195] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB," *Journal of Chemical Theory and Computation*, vol. 11, no. 8, pp. 3696–3713, 2015.

[196] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, 1983.

[197] P. Li and K. M. Merz, Jr., "Taking into account the ion-induced dipole interaction in the nonbonded model of ions," *Journal of Chemical Thoery and Computation*, vol. 10, no. 1, pp. 289–297, 2014.

[198] P. Li, L. F. Song, and K. M. Merz, Jr., "Systematic parameterization of monovalent ions employing the nonbonded model," *Journal of Chemical Thoery and Computation*, vol. 11, no. 4, pp. 1645–1657, 2015.

[199] P. Li, L. F. Song, and K. M. Merz, Jr., "Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water," *The Journal of Physical Chemistry B*, vol. 119, no. 3, pp. 883–895, 2015.

[200] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *Journal of Computational Chemistry*, vol. 25, no. 9, pp. 1157–1174, 2004.

[201] F. Neese and J. Wiley, "The ORCA program system," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 1, pp. 73–78, 2012.

[202] F. Neese, "Software update: the ORCA program system, version 4.0," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 8, no. 1, p. e1327, 2018.

[203] K. D. Hammonds and D. M. Heyes, "Shadow Hamiltonian in classical NVE molecular dynamics simulations: A path to long time stability," *The Journal of Chemical Physics*, vol. 152, no. 2, p. 024114, 2020.

[204] J. C. Klein, A. R. Burr, B. Svensson, D. J. Kennedy, J. Allingham, M. A. Titus, I. Rayment, and D. D. Thomas, "Actin-binding cleft closure in myosin II probed by site-directed spin labeling and pulsed EPR," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 12867–12872, 2008.

[205] M. H. Doran, E. Pavadai, M. J. Rynkiewicz, J. Walklate, E. Bullitt, J. R. Moore, M. Regnier, M. A. Geeves, and W. Lehman, "Cryo-EM and Molecular Docking Shows Myosin Loop 4 Contacts Actin and Tropomyosin on Thin Filaments," *Biophysical Journal*, vol. 119, pp. 821–830, 2020.

[206] S. Schultze and H. Grubmüller, "Time-Lagged Independent Component Analysis of Random Walks and Protein Dynamics," *Journal of Chemical Theory and Computation*, vol. 17, pp. 5766–5776, 2021.

[207] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, and F. Noé, "PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models," *Journal of Chemical Theory and Computation*, vol. 11, pp. 5525–5542, 2015.

[208] B. Trendelkamp-Schroer, H. Wu, F. Paul, and F. Noé, "Estimation and uncertainty of reversible Markov models," *The Journal of Chemical Physics*, vol. 143, p. 174101, 2015.

[209] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, "Markov models of molecular kinetics: generation and validation," *The Journal of Chemical Physics*, vol. 134, p. 174105, 2011.

[210] J. von der Ecken, M. Müller, W. Lehman, D. J. Manstein, P. A. Penczek, and S. Raunser, "Structure of the F-actin–tropomyosin complex," *Nature*, vol. 519, pp. 114–117, 2015.

[211] J. E. Stone, J. Gullingsrud, and K. Schulten, "A System for Interactive Molecular Dynamics Simulation," 2001.

[212] J. Eargle, D. Wright, and Z. Luthey-Schulten, "Multiple Alignment of protein structures and sequences for VMD," *Bioinformatics*, vol. 22, no. 4, pp. 504–506, 2006.

[213] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: Application to microtubules and the ribosome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 18, pp. 10037–10041, 2001.

[214] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[215] T. M. Inc, "MATLAB version: 9.5.0 (R2018b)," 2018.

[216] P. M. L. Janssen, L. B. Stull, and E. Marbán, "Myofilament properties comprise the rate-limiting step for cardiac relaxation at body temperature in the rat," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 282, no. 2, pp. H499–H507, 2002.

[217] S. M. Mijailovich, M. Prodanovic, C. Poggesi, M. A. Geeves, and M. Regnier, "Multiscale modeling of twitch contractions in cardiac trabeculae," *Journal of General Physiology*, vol. 153, 2021.

[218] S. M. Kim, B. E. Randall, F. Jezek, D. A. Beard, and N. C. Chesler, "Computational modeling of ventricular-ventricular interactions suggest a role in clinical conditions involving heart failure," *Frontiers in Physiology*, vol. 134, p. 1231688, 2023.

[219] T. Tune, K. B. Kooiker, J. Davis, T. Daniel, and F. Moussavi-Harami, "Identifying Mechanisms and Therapeutic Targets in Muscle using Bayesian Parameter Estimation with Conditional Variational Autoencoders," 2024.