EXPLORING THE SHIFTS IN THE ENERGY LANDSCAPE OF CALMODULIN UNDER DIFFERENT ENVIRONMENTAL CONDITIONS

by

BÜŞRA TAYHAN

Submitted to the Graduate School of Engineering and Natural Sciences

in partial fulfilment of

the requirements for the degree of Master of Science

Sabancı University

December 2024

ABSTRACT

EXPLORING THE SHIFTS IN THE ENERGY LANDSCAPE OF CALMODULIN UNDER DIFFERENT ENVIRONMENTAL CONDITIONS

Büşra Tayhan

Molecular Biology, Genetics and Bioengineering, MSc. Thesis, 2024

Thesis Supervisor: Canan Atılgan

Thesis Co-supervisor: Ali Rana Atılgan

Keywords: molecular dynamics simulations, calmodulin,

metadynamics simulations, ionic strength, collective variables.

Proteins are flexible structures, and their functions are closely tied to their ability to adopt multiple conformations. Their dynamic nature enables them to shift between different shapes or states, which is crucial for interacting with other biomolecules or catalyzing reactions. This behavior of proteins is strongly influenced by environmental conditions, such as temperature, pH, and ionic strength, which can significantly alter their structural states and functions.In this thesis, we investigate how the conformational landscape of proteins shifts under changing environmental conditions. Our model system is calmodulin (CaM), a calcium-binding protein involved in numerous cellular processes. Known for its adaptability, CaM exhibits a wide array of conformations under various experimental conditions, making it an ideal candidate for studying environmental effects on protein dynamics. Moreover, CaM serves as the sensing domain in genetically encoded fluorescent calcium biosensors, highlighting the importance of understanding its conformational changes for designing efficient and responsive sensors. The regions visited by CaM are well-represented by a pair of essential degrees of freedom (DoF) describing the relative positioning of its two lobes (*cis*/*trans*) and the compactness of the overall CaM structure. Since the time scale of classical molecular dynamics (MD) simulations do not allow for overcoming the high energy barriers that might separate various minima, we resort to well-tempered metadynamics simulations using the essential DoF as collective variables (CVs). We explore four different conditions representing Ca^{2+} bound/unbound CaM at physiological/low ionic strength. We find that

CaM acts like a juggler, adjusting the position of its minima between four stable regions (*cis*-compact, *trans*-compact, *cis*-open, *trans*-open). After identifying the locations of the minima via metadynamics, we select representative structures from these minima and perform equilibration runs. Our work unveils the structural basis of the observed minima and validates the effectiveness of our chosen CVs. Our results reveal that, although the energy surface is much shallower under low ionic strength (IS) conditions, conformers get stuck in artificial minima due to strong transient salt bridges. It appears that ions act as lubricants, helping to release salt bridges that cause the conformers to become trapped. As a result, we find IS of the environment to be a major factor to be considered in protein design problems.

ÖZET

FARKLI ÇEVRESEL KOŞULLARDA KALMODULİNİN ENERJİ YÜZEYİNDEKİ DEĞİŞİMLERİN ARAŞTIRILMASI

Büşra Tayhan

Moleküler Biyoloji, Genetik ve Biyomühendislik, Yüksek Lisans Tezi, 2024

Tez Danışmanı: Canan Atılgan

Tez Eş Danışmanı: Ali Rana Atılgan

Anahtar kelimeler: moleküler dinamik simülasyonları, kalmodulin, iyonik kuvvet, metadinamik benzetimleri, kolektif değişkenler

Proteinler sabit yapılar değildir; fonksiyonları, birden fazla konformasyon edinebilme yetenekleriyle yakından ilişkilidir. Bu dinamik doğaları, proteinlerin diğer biyomoleküllerle etkileşim kurmalarını veya reaksiyonları katalize edebilmeleri için farklı şekiller veya durumlar arasında geçiş yapmalarını mümkün kılar. Proteinlerin dinamik davranışları, sıcaklık, pH ve iyonik kuvvet gibi çevresel koşullara güçlü bir şekilde bağlıdır ve bu koşullar, proteinlerin yapılarını ve fonksiyonlarını önemli ölçüde değiştirebilir.

Bu tezde, proteinlerin konformasyonel çeşitliliğinin değişen çevresel koşullar altında nasıl farklılaştığını araştırıyoruz. Model sistem olarak, çok sayıda hücresel süreçte yer alan ve farklı deneysel koşullarda birçok konformasyonu tespit edilmiş bir kalsiyum bağlayıcı protein olan kalmodulini (CaM) seçtik. CaM, çevresel etkilerin protein dinamikleri üzerindeki etkilerini incelemek için ideal bir adaydır. Dahası, CaM, genetik olarak kodlanmış floresan kalsiyum biyosensörlerinde algılama bölgesi olarak görev yapar; bu da konformasyonel değişimlerini anlamanın, etkili ve duyarlı sensörlerin tasarlanması açısından önemini vurgular.

CaM'in ziyaret ettiği bölgeler, iki temel serbestlik derecesi ile iyi bir şekilde temsil edilir: iki lobununun birbirinden bağımsız dönüş açısı (*cis*/*trans*) ve lobların birbirine olan uzaklığı. Klasik moleküler dinamik (MD) benzetimleriyle ulaşılabileceğimiz zaman ölçekleri çeşitli

minimumları ayıran yüksek enerji bariyerlerini aşmaya izin vermediğinden, temel serbestlik derecelerini kolektif değişkenler olarak kullanarak metadinamik benzetimlerini uyguladık. Fizyolojik/düşük iyonik kuvvet koşullarında kalsiyumlara bağlı/bağlı olmayan CaM'i temsil eden dört farklı koşulu araştırdık. CaM'in, dört kararlı bölge (*cis*-kompakt, *trans*-kompakt, *cis*-açık, *trans*-açık) arasında minimumlarının konumunu ayarlayan bir jonglör gibi davrandığını tespit ettik. Metadinamik benzetimleri yoluyla minimumların konumlarını belirledikten sonra, bu minimumlardan temsilci yapılar seçip minimizasyon ve ardından dengeleme benzetimleri gerçekleştirdik. Bunu, gözlemlenen minimumların yapısal temelini anlamak ve seçilen kolektif değişkenlerin etkinliğini doğrulamak için yaptık. MD benzetim sonuçlarımız, düşük iyonik kuvvet koşulları altında enerji yüzeyinin çok daha sığ olmasına rağmen, konformerlerin yapay minimumlarda sıkışıp kaldığını ortaya koydu. Bu durumu, iyonların, konformerlerin sıkışmasına neden olan tuz köprülerinin serbest kalmasına yardımcı olan faktörler olarak tanımlayarak açıklayabilmekteyiz. Dolayısıyla iyonik kuvveti protein tasarımında hesaba katılması gereken önemli bir aktör olarak belirlemekteyiz.

Nisi speratis, insperatum non invenientis,

Quum inveniri non possit atque inaccessum sit.

ACKNOWLEDGEMENTS

I would like to thank my advisors, Canan Atilgan and Ali Rana Atilgan, for their invaluable support and guidance. Their mentorship extended beyond simply accompanying me on my academic journey; they also played a vital role in helping me grow both intellectually and personally. Through their thought-provoking questions, they enhanced my critical thinking skills, enabling me to approach both my academic pursuits and daily life with greater awareness and insight.

I would also like to express my gratitude to my jury members: Oznur Tastan, Umut Sahin, and Nihan Celebi for their thoughtful comments and questions. Their feedback not only helped me further consolidate my work but also inspired me to consider it from different perspectives. They were very friendly and sincere during my thesis defense, which encouraged me and made the experience much more positive.

During my master's program, I had the opportunity to work with three bright and intelligent undergraduate students: Dilara, Sila, and Dilan. Although they were younger than me, they broadened my perspective and enriched my understanding. I truly enjoyed both learning together and spending time with them. Sıla and Dilara were our PURE students, and together we worked on metadynamics simulations. Throughout this process, we held numerous meetings, both in person and on Zoom. In every meeting, they brought many ideas to the table, making the entire experience exciting and engaging. When it came to my personal life, Dilara and I had so much fun together both in and outside of school. She brought me the cutest gifts from everywhere she went and took care of me when I was sick. I met Dilan when she came for her summer internship, and during the short time she was here, we did so much besides work. We went on a shopping date, hit the gym, and enjoyed a variety of activities together.

Sabancı also brought me together with wonderful people outside the office, including Alize, Armita, and Kadriye, with whom I shared long discussions on various topics, spent countless

enjoyable hours, and received invaluable support. Alize and I met in our machine learning class. She's someone I could talk to for hours about anything, from movies to literature and games. She was also always thoughtful, often reaching out to check in and see how I was doing. Armita has always managed to make me burst into laughter with her jokes since the day we met. Kadriye, with her colorful personality and unique style, was one of the most iconic people I met at this school. Her positive energy and sweet gestures brought me so much happiness.

Now it's time to thank all the MIDSTLab members. Melike, Ata, Isik, Ebru, Goksin and Furkan, who made me feel like this place was my home and I always enjoyed their pleasant conversations. Melike was one of the first people to help me with our field of study when I came here. Whenever I had a question, she answered it immediately and helped me a lot in this way, she also added beautiful songs to my music library during our road trips. Although Ata is new to our team, he quickly adapted to the group with his friendly attitude and added color with his cheerful personality. I cannot forget the beautiful gifts Isik gave me. When I lived at home, she often drove me back, we shared pleasant conversations (and a little gossip) over coffee at Starbucks, and she became my puzzle buddy. When I first came here, Ebru and Goksin welcomed me with warm smiles. Ebru was the one who introduced Sabanci to me and was the first person to give me flowers outside my family. We talked a lot during our short breaks, and she listened to my complaints even when she went to Arizona for her Postdoctorate. Since I met Goksin he has brought me luck and supported me in so many ways. She has helped me a lot with moving into my house, where I have built wonderful memories, and she has opened her doors for me and introduced me to two wonderful young ladies, Irma and Jenny. One of the greatest fortunes she brought into my life was Furkan. The time I spent here with Furkan felt as short as a single day, but I made so many beautiful memories with him that they could not fit into a lifetime. Every moment we shared was so precious that I cannot separate one memory from another or say, 'We also did this together.' He believed in me more than I believed in myself and was always there for me. Even in my saddest moments, he found a way to cheer me up. He worried about me more than I worried about myself, always thinking of everything before I even had a chance to. He contributed greatly to both my personal life and academic journey, helping me discover myself and shaping me into an even better person than I was.

A big thank you goes to my friend Onur from Antalya. Together, we overcame many challenges there while having fun, and we both built bright futures. Although his constant stream of new ideas and boundless energy sometimes exhausted me, we managed to maintain a six-year-long friendship.

And Deniz holds a very special place in my heart; we've been together since our childhood. She witnessed all my happy moments and listened to me pour my heart out countless times. Even though we've only been able to meet 2-3 times in the past few years, she always made me feel as if she were right by my side.

My family never withheld their support and stood by my side in every decision I made. My mom was also my closest friend; she enjoyed listening to everything I shared in great detail and made an effort to understand my work. My brother listened to my worries, helped me make sense of them, and offered advice based on his own life experiences. My dad always wished the best for me and supported me whenever I needed help. And Thor, for seven years, was my adorable companion, joining me on countless adventures.

My time here has been incredibly educational and meaningful for me. I learned new methods, gained fresh perspectives, and found an environment where I could develop myself intellectually and personally. More importantly, this place introduced me to many wonderful people who will always hold a special place in my life. The experiences and relationships I built here have left a lasting impact, and I will carry them with me wherever I go.

This study was supported by the TÜBİTAK 121Z329 project

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVATIONS

Å	Ångström
fs	femtosecond
K	Kelvin
kcal	kilocalorie
MD	Molecular Dynamics
ns	nanosecond
PDB	Protein Data Bank
RDF	Radial Distribution Function
RMSD	Root Mean Square Deviation
wt-MetaD	Well-Tempered Metadynamics
IS	Ionic strength
CaM	Calmodulin
Ca^{2+}	Calcium ion
μs	microsecond

1. INTRODUCTION

1.1 Overview

Proteins are inherently dynamic, continuously shifting between various shapes and conformations [1]. This structural flexibility is crucial for their biological roles, as it enables proteins to adjust their shape to interact with other biomolecules or carry out catalytic functions [2]. Unlike small, rigid molecules, proteins consist of long chains of amino acids that fold into intricate three-dimensional structures [1]. However, these structures are not static; instead, they exist in a dynamic equilibrium, constantly transitioning between multiple conformations. This flexibility allows a single protein to perform a range of functions depending on the specific conformation it adopts [3, 4].

For instance, proteins can be involved in numerous cellular processes, and during these processes, they may need to rapidly switch between different functional states. The ability to shift between these states is determined by the energy landscape of the protein, which describes the different conformational possibilities that a protein can adopt [5]. Each state represents a local energy minimum, and proteins transition between these states depending on changes [6, 7], such as binding to ligands, changes in pH, ionic concentration or temperature in the environment, or post-translational modifications [8]. These shifts allow proteins to alter their surface properties, such as exposing or hiding binding sites, which is crucial for interacting with other molecules [7, 9].

Among the many proteins that change their shape to recognize molecular partners, calmodulin (CaM) stands out as one of the most classic examples. CaM is a well-known Ca^{2+}-binding protein conserved across eukaryotes [10]. It can effectively adapt its shape to fit the biochemical environment of its interaction partners, providing the necessary specificity for a wide range of cellular processes. In addition to regulating the secondary messenger calcium ion, CaM plays a pivotal role in numerous cellular processes, including interacting with transcription factors, modulating gene expression, and binding to protein kinases [11].

The ability to adopt multiple conformations allows CaM to dynamically adapt to its binding partners, ensuring high specificity and functionality in diverse signaling pathways. It is also the main ingredient of the range of genetically encoded fluorescent calcium biosensors, which use various chimeric forms of CaM as the sensing domain. These sensors are known to be affected by the conditions in the immediate environment of where the measurements are made and understanding the conformational changes the sensor domain undergoes is important for the design of efficient sensors [12]. **In this thesis, we aim to explore the conformational flexibilities of CaM under high/low salt conditions, in the presence/absence of calcium ions to shed light on the extent to which these changes may be incorporated into a general design problem.**

The thesis is organized as follows: Chapter 1, **Introduction,** introduces the problem addressed in this study and its biological relevance. It provides an overview of the structural and functional properties of calmodulin (CaM), explains how to navigate its conformation surface, and reviews previous studies that explored the conformation landscape of CaM. This chapter concludes with the motivation behind this research. Chapter 2, **Materials and Methods**, describes the approaches used in this study, including classical molecular dynamics (MD) simulations and well-tempered metadynamics simulations. It details the preparation of systems, the selection of collective variables, and the methods used to analyze simulation trajectories. Chapter 3, **Results and Discussion**, presents findings from classical MD simulations and well-tempered metadynamics simulations, exploring how these methods provide insights into the structural basis of the observed conformational changes and the biological relevance of these findings. Chapter 4, **Conclusions**, summarizes the key findings of this thesis work, providing an overview of the insights gained into the conformational landscape of calmodulin. Finally, Chapter 5, **Epilogue and Future Work**, discusses potential directions for future research that could extend and build upon the results presented here.

1.2 What is Known on the Structure-Function Relationships in CaM

CaM is a small, acidic calcium-binding protein that is highly conserved among eukaryotes and serves as the primary calcium signaling protein in mammals [13]. Although the acidic residues of CaM increase its water solubility, this protein also highly enriched in hydrophobic residues, leading to highly collapsed structures [14]. CaM consists of approximately 148 residues and is structurally composed of two globular domains, each capable of binding two Ca^{2+} ions: the N-terminal domain (NTD) and the C-terminal domain (CTD), which span residues 1-68 and 92-148, respectively. These domains are connected by a long, central helix, comprising residues 69-91, which is essential for the functionality of this protein [15].

Each globular domain contains two unique Ca^{2+}-binding motifs known as EF-hands, which bind two Ca^{2+} ions cooperatively [16]. These motifs have a classic 'helix-loop-helix' configuration, where two α-helices are connected by a flexible loop, allowing Ca^{2+} ions to fit perfectly within the interhelical region [17]. The flexible loop, rich in negatively charged residues like aspartate and glutamate, plays a crucial role in coordinating the binding of these ions [18]. Its negatively charged residues create an electrostatic attraction to the positively charged calcium ions, ensuring strong and specific binding. This interaction stabilizes the protein structure and triggers a reorientation of the helices, inducing conformational changes throughout the domain [19]. This precise arrangement of the negatively charged residues in the EF-hand loop provides the specificity needed for Ca^{2+} binding, which is essential for regulating calcium-dependent processes.

Because CaM contains four unique motifs, each roughly 30 residues long, only a small segment of the 148-residue structure is not involved in Ca^{2+} binding, including 8–10 residues in the central region of CaM [19]. This central helical linker plays a crucial role in the function of this protein in cellular processes, not only holding the N-terminal (NTD) and C-terminal (CTD) domains together, but also enabling structural flexibility of the protein [20]. This flexibility is key to the wide range of conformations CaM can adopt, allowing its domains to move independently in solution [21] as illustrated in Figure 1.1.

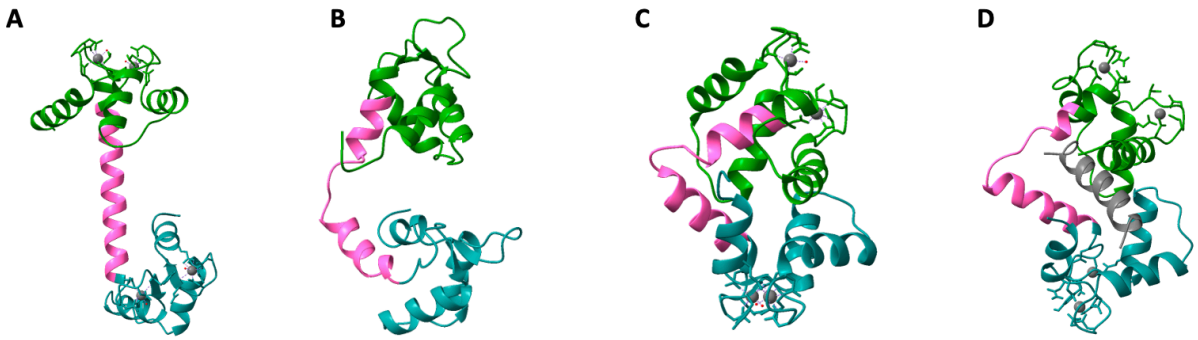

Figure 1.1 Three-dimensional structures of CaM include: **a** an open Ca^{2+}-loaded crystal structure (PDB ID: 3CLN), **b** a closed apo NMR structure (PDB ID: 6Y95), **c** a closed X-ray structure (PDB ID: 1PRW), and **d** an NMR structure of myosin light chain kinase (MLCK)-bound Ca^{2+}-CaM (PDB ID: 2K0F). The N-terminal, linker, and C-terminal domains are colored green, pink, and blue, respectively, while Ca^{2+} ions and the peptide are shown in gray. This color coding is used throughout the thesis, unless otherwise specified.

The central helix also plays a pivotal role in binding partners, enabling CaM to modulate various cellular processes [22]. Upon binding Ca^{2+} ions, each EF-hand loop undergoes significant structural shifts, reorienting the lobes [13]. This reorganization allows the globular domains to come closer together or move further apart [23]. Such conformational flexibility enables the helical linker region to wrap around various diverse binding partners, a crucial feature for the role of CaM in signaling pathways and cellular regulation [19]; see also Figure 1.1.d.

Although interactions with target proteins often depend on Ca^{2+} binding, CaM also associates with proteins in its calcium-free state, frequently recognizing an unique motif that is compatible with both Ca^{2+}-bound and Ca^{2+}-free forms [24, 25]. This dual capability reflects the structural versatility of CaM, which allows it to adopt open, semi-open, or closed conformations [11]. These distinct conformations have been extensively studied through NMR and X-ray crystallography of calcium-bound, target-bound, and unbound forms, as summarized in Table 1.1.

In the calcium-loaded peptide-free state, CaM typically adopts an open, dumbbell-shaped conformation [13], as resolved in an X-ray crystal structure (Figure 1.1.a). The two globular

domains are connected by a flexible central helix and positioned in an extended arrangement. However, a compact, globular conformation has also been observed in a calcium-loaded crystal structure (see Figure 1.1.c), where the lobes are closer together [26].

Additionally, the peptide-unbound, calcium-loaded NMR structures of CaM exhibit a diverse range of conformations, spanning from open to closed states [13]. This variability demonstrates the intrinsic flexibility of CaM, showing that calcium binding alone does not restrict it to a single conformation but prepares it for dynamic transitions.

In peptide-bound forms, such as those determined by NMR, CaM often adopts a closed conformation where the N-terminal and C-terminal lobes clamp tightly around the target peptide, stabilizing the complex and enhancing specificity in signaling pathways (see Figure 1.1.d) [13].

Calcium-free (*apo*) CaM structures, as shown in Figure 1.1.b, exhibit a semi-open conformation, characterized by partially collapsed lobes that retain dynamic flexibility [20]. This structural arrangement balances readiness and adaptability, enabling CaM to swiftly transition to calcium-bound or target-bound forms as needed. Notably, the C-terminal domain of *apo*-CaM pre-forms into a semi-open state, facilitating its interaction with IQ motifs in target proteins [25].

Table 1.1 Calmodulin structures from Protein Data Bank (PDB)

PDB ID	Source Organism	Experimental Resolution (Å)	Method	Residue Number	Conformation	Ligand	Mutation
3CLN	Black rat (*Rattus rattus*)	2.20 Å	X-Ray Diffraction	148	Open	Ca^{2+}	-
1PRW	Cow (*Bos taurus*)	1.70 Å	X-Ray Diffraction	149	Closed	Ca^{2+}	M3L
2K0E	Human (*Homo sapiens*)	-	Solution NMR	148	Open	Ca^{2+}	-
2K0F	Human (*Homo sapiens*)	-	Solution NMR	148	Closed	19-mer peptide from Myosin light chain kinase, Ca^{2+}	-
1DMO	African clawed frog (*Xenopus laevis*)	-	Solution NMR	148	Closed	-	-
1MUX	African clawed frog (*Xenopus laevis*)	-	Solution NMR	148	Closed	WW-7 Complex, Ca^{2+}	-
1LIN	Cow (*Bos taurus*)	2.00 Å	X-Ray Diffraction	148	Closed	Trifluoperazine, Ca^{2+}	-
2BCX	Chicken (*Gallus gallus*)	2.00 Å	X-Ray Diffraction	148	Closed	Ryanodine receptor-1, Ca^{2+}	-
2M55	Human (*Homo sapiens*)	-	Solution NMR	148	Closed	α-synuclein, Ca^{2+}	-
2KNE	Human (*Homo sapiens*)	-	Solution NMR	148	Closed	ATPase, Ca++ transporting, plasma membrane 4, Ca^{2+}	-
6Y95	Human (*Homo sapiens*)	-	Solution NMR	148	Closed	-	N53I
1CDL	Human (*Homo sapiens*)	2.40 Å	X-Ray Diffraction	147	Closed	Calcium/Calmodulin-Dependent Protein Kinase Type II α-Chain, Ca^{2+}	-
2BBM	Fruit fly (*Drosophila melanogaster*)	-	Solution NMR	148	Closed	Myosin light chain kinase (MLCK), Ca^{2+}	-

The structural versatility of CaM enables it to function as a dynamic molecular adaptor, fine-tuning interactions with diverse binding partners and signaling pathways [19]. In its open, calcium-loaded state, CaM undergoes conformational changes that expose its hydrophobic pockets, allowing it to interact with calcium-regulated enzymes such as kinases and phosphatases [13]. This interaction facilitates key processes like muscle contraction, neurotransmitter release, and gene transcription [27]. For example, Ca^{2+} binding to CaM activates Ca^{2+}/CaM-dependent protein kinases (CaMKs) and phosphatases such as calcineurin, which are essential for signal transduction and cellular responses [22]. Additionally, CaM regulates ion channels like the ryanodine receptor and voltage-gated

calcium channels, which play pivotal roles in calcium homeostasis and cellular excitability [19, 28, 29].

Conversely, in its closed, peptide-bound conformation, CaM wraps tightly around hydrophobic motifs in target peptides, such as those in myosin light chain kinase (MLCK), to regulate actomyosin contraction and cytoskeletal dynamics, essential for muscle contraction and cell motility [22, 30]. This structural adaptation ensures stable and specific interactions, highlighting the role of CaM as a molecular switch in calcium-dependent processes [19].

Interestingly, CaM exhibits significant functional adaptability in its calcium-free, semi-open conformation, enabling it to interact with targets independently of calcium [27]. For instance, it binds conserved IQ motifs, characterized by the IQxxxRGxxxR consensus sequence, found in voltage-gated sodium channels and unconventional myosins [25]. This enables calcium-independent regulation of processes such as neuronal signal propagation and intracellular trafficking [29]. This highlights CaM's dual role as a calcium sensor and adaptor, supporting cellular functions even when calcium levels are low.

Beyond its cellular roles, CaM has been widely utilized in biotechnology as a foundation for biosensor development. Leveraging its unique ability to undergo significant conformational changes upon calcium binding, CaM has been fused with fluorescent proteins, such as GFP, and a peptide, to create genetically encoded calcium indicators (GECIs) like GCaMP [31]. These biosensors are among the most successful and widely studied due to CaM's structural versatility, which allows it to respond sensitively to the presence or absence of calcium ions [32]. In GCaMP, the peptide acts as a calcium-sensitive partner that interacts with CaM in a dynamic, calcium-dependent manner. In the absence of calcium, CaM and the peptide remain loosely associated; however, upon calcium binding, CaM undergoes a conformational shift, tightly wrapping around the peptide [33]. This structural change enhances the fluorescent signal of GFP, enabling real-time visualization of calcium dynamics in living cells. Such precision and sensitivity have made GECIs invaluable tools in neuroscience and cell biology [34, 35]. Furthermore, CaM's adaptability has been exploited to develop target-specific biosensors for detecting protein-protein interactions in complex environments, demonstrating its immense utility in both scientific research and technological advancements

[34]. MD simulations on a series of GECIs have proven invaluable to understand how the ON/OFF states are distinguished by conformational changes in the CaM domain upon calcium removal [31].

1.3 How to Navigate the Conformation Surface of a Protein

Protein conformational surfaces describe the potential energy landscape of a protein, consisting of energy minima that represent stable states and energy barriers that proteins must traverse to move between functional states. Investigating these dynamics requires integrating computational and experimental approaches to construct a comprehensive map of the conformational surface [36]. Efficient sampling of stable conformations and transitions is essential to decipher pathways between key conformations [3].

MD simulations are commonly used to explore the dynamic behavior of biomolecules. However, these simulations are often constrained by limited simulation timescales, which may be insufficient to overcome energy barriers and sample rare transitions between distinct conformational states [5, 36]. As a result, the system can become trapped in local energy minima, making it challenging to capture rare events or explore the full conformational space of large, complex systems [37]. These limitations are particularly challenging when investigating processes such as protein folding, ligand binding, or enzymatic activity, where significant energy barriers must be crossed to observe the complete pathway [38].

Several enhanced sampling techniques have been developed to address these challenges. Methods such as replica exchange molecular dynamics (REMD) [39], umbrella sampling [40], and adaptive biasing force (ABF) [41] are commonly employed to overcome energy barriers. However, these approaches often suffer from limitations, such as high computational cost or reliance on predefined reaction coordinates [42].

The metadynamics (MetaD) method dynamically explores the free energy landscape by introducing a bias potential along selected degrees of freedom, referred to as collective variables (CVs) [36, 43] [38]. This method is an advanced version of Adaptive Umbrella Sampling (AUS). Unlike AUS, which may get trapped in initial metastable states, MetaD

iteratively adds small Gaussian potentials that allow the system to escape these states [5] (see section 2.2 for details).

The convergence of MetaD simulations heavily depends on the selection of collective variables, as they define the degrees of freedom along which the system is biased [44]. If the collective variables do not effectively capture the slowest motions, the simulation may fail to sample the free-energy landscape adequately, resulting in incomplete or misleading results [38]. Therefore, poorly chosen CVs can trap the system in metastable states or fail to distinguish between critical conformational states, ultimately hindering the accurate reconstruction of the free energy. To address this, it is crucial to ensure that the CVs fulfill specific criteria. They should describe the slow degrees of freedom associated with the process of interest, clearly differentiate between the relevant metastable states, and be computationally efficient to evaluate on-the-fly during the simulation [5, 42]. Adhering to these principles allows the accurate identification of optimal CVs, enhancing both the efficiency and reliability of metadynamics simulations.

While standard metadynamics promotes barrier crossing effectively, it can sometimes overfill the free energy surface, leading to inaccurate estimates. This limitation is addressed by well-tempered metadynamics (wt-MetaD), which gradually decreases the bias deposition as the simulation progresses [45] (see section 2.2 for details). Through the use of wt-MetaD, one is able to efficiently sample the conformational space of biomolecular systems, capturing rare events and transitions that are not observable by conventional MD simulations. The dynamic biasing mechanism of wt-MetaD enables one to reconstruct the free energy surface with greater accuracy. We have made use of these properties of wt-MetaD to get insights into the complex conformational behavior of CaM.

1.4 What is Known about the Conformation Landscape of CaM

The conformational landscape of CaM has been extensively studied using various computational approaches, each offering unique insights into its structural adaptability and dynamics. It was shown that subtle pH differences trigger conformational changes in CaM, which is mimicked by protonation of residues with upshifted pKa values [46, 47]. In another study, MD simulations were combined with perturbation-response scanning (PRS) [48, 49] to identify the reaction coordinate driving conformational transitions in CaM. Dihedral angles and distances were used as degrees of freedom (DoFs) to project MD trajectories into conformational subspaces, providing a detailed understanding of population shifts among CaM's conformational states [50]. Similarly, another study employed PRS together with steered molecular dynamics (SMD) to facilitate conformational transitions. This study also utilized dihedral angles and distances as DoFs to project molecular dynamics trajectories, further elucidating the pathways and dynamics of functional transitions in CaM [51].

Metadynamics simulations have been used to explore the molecular mechanism underlying CaM’s recognition of target peptides, revealing how peptide binding reshapes its conformational landscape at atomic resolution [23]. The role of calcium ions in modulating CaM's conformational landscape has also been investigated, demonstrating how calcium binding and release affect its structural collapse during peptide binding [52].

Additionally, the choice of the coarse grained dihedral angle and distance pair as CVs, as defined by Aykut et al. [50], have found broader applications in other contexts. For instance, these DoFs were employed in machine learning models to analyze molecular simulations and predict the functional effects of genetic variations in CaM, illustrating their versatility in interpreting this protein’s dynamics [53].

1.5 Motivation

Although extensive experimental studies have been conducted, fully capturing the conformational landscape of CaM remains a challenge due to its inherent flexibility and the intricate effects of environmental factors like ionic strength variations in the cell and calcium levels. Experimental techniques, while powerful, typically provide static snapshots, leaving the dynamic processes connecting these states largely unexplored. Computational methods, such as MD simulations, offer a way to bridge this gap by providing time-resolved insights into protein dynamics. However, the limited timescales of classical MD simulations often prevent them from overcoming high-energy barriers, restricting their ability to sample rare but biologically significant conformational states.

This study is driven by the need to overcome these limitations and explore the conformational landscape of CaM under varying environmental conditions. By employing the enhanced sampling technique of wt-MetaD, we aim to capture rare conformational events and uncover previously inaccessible states. Our focus is on understanding how ionic strength and calcium binding modulate CaM's conformational transitions, which are central to its ability to interact with diverse binding partners and regulate critical cellular processes.

Through this work, we seek to provide a detailed, dynamic view of CaM’s energy landscape, shedding light on the interplay between environmental factors and protein flexibility. This knowledge not only enhances our understanding of CaM’s biological role but also contributes to broader efforts to link protein dynamics with function, with implications for drug design and therapeutic interventions targeting calcium-mediated signaling pathways as well as the design of new and improved GECIs.

2. MATERIALS AND METHODS

2.1 Classical Molecular dynamics (MD) Simulations

2.1.1 System Preparation and MD Simulations

To understand how different environmental conditions influence the dynamic behavior of CaM, we performed two 1 μs-long molecular dynamics (MD) simulations for each system listed in Table 2.1. Calcium loaded and open structure of CaM (PDB Code: 3CLN) was obtained from the Protein Data Bank (PDB) and was used to prepare the different systems. The properties of the systems used in the MD simulations are given in Table 2.1 with labels, equilibrated box sized and ionic strengths (IS). Systems were labeled as $3CLN^{Ca2+}{}_{IS\_P}$ (calcium loaded CaM at physiological IS), $3CLN^{Ca2+}{}_{IS\_L}$ (calcium loaded CaM in low IS), $3CLN_{IS\_P}$ (*apo* CaM at physiological IS) and $3CLN_{IS\_L}$ (*apo* CaM at low IS), respectively. To stimulate the influence of the absence of Ca^{2+} in the environment, *apo* condition, Ca^{2+} ions were removed from the initial structure. MD simulations were performed by using NAMD [54] software, and CHARMM36 force-field was used for topologies and parameters. VMD [55] was utilized for preprocessing of structures, such as protein structure file (PSF) generation, solvation (constructing water-box), ionization and visualization of MD trajectories. By using the solvent plug-in VMD 1.9.4, protein structures were solvated in a rectangular water box with a minimum distance of 10 Å between the protein and the nearest edge of the water box. KCl salt is used for ionization purposes. K^+ and Cl^- ions were added to neutralize charges and maintain a 150 mM physiological ionic strength. (Table 2.1). For the low IS condition, on the other hand, number of ions were adjusted to neutralize the system. The Particle Mesh Ewald method was employed for calculating long-range electrostatic interactions with a cutoff of 12 Å and a switching distance of 10 Å. The RATTLE algorithm was applied to use the Verlet algorithm with a time step of 2 fs. Langevin piston was used for pressure control at 1 atm, and the temperature was maintained at 310 K by the Langevin thermostat. Each system was minimized for 10,000 steps and the equilibrium

simulations were performed for 1μs. Trajectories were stored every 10 ps leading to 10,000 snapshots for each trajectory.

Table 2.1 Summary of system parameters for each condition

Simulation Label	Simulation Condition	Equilibrated Box size (Å)	Ionic strength (mM)	Number of ions
$3CLN^{Ca2+}_{IS\_P}$	Ca^{2+} loaded protein in physiological condition	65.3x82.4x61.8	177 mM	43 K^+ 28 Cl^-
$3CLN^{Ca2+}_{IS\_L}$	Ca^{2+} loaded protein in low ionic strength condition	65.3x82.4x61.7	38 mM	15 K^+
$3CLN_{IS\_P}$	Ca^{2+} unbound protein in physiological condition	65.1x82.3x61.6	199 mM	51 K^+ 28 Cl^-
$3CLN_{IS\_L}$	Ca^{2+} unbound protein in low ionic strength condition	65.3x82.5x61.8	57 mM	23 K^+

2.1.2 Trajectory Analyses

The first frame of each trajectory is used as a reference for RMSD (root-mean-square deviation) calculations. Radial distribution function ($g(r)$) analyses are performed in VMD with the default settings [55]. Additionally, manipulation of trajectory files and RMSD calculations are done by using ProDy [56] package of python programming language. The in-house prepared codes are provided in the Appendices and are shared on GitHub.

2.2 Well-Tempered Metadynamics (wt-MetaD) Simulations

As outlined in the Introduction, MetaD dynamically explores the free energy landscape by introducing a bias potential along selected degrees of freedom, referred to as collective variables (CVs) [36]. It iteratively adds small Gaussian potentials that allow the system to escape these states [5]. Each Gaussian potential, defined by a width (σ) and height (ω), is based on the probability distribution of the CV (see *eq* 2.1) This bias potential accumulates in the explored regions and is approximately the negative of the free energy, gradually filling the minima (see *eq* 2.2). By adding this bias, the system avoids getting trapped and is enabled to thoroughly explore the metastable states [57].

$$V_G(S,t) = \omega \sum_{t\prime<t} \exp\left(-\frac{(S-s(t\prime))^2}{2\delta^2}\right) \tag{2.1}$$

$$\lim_{t\to\infty} V_t(S) = -F(S) + C \tag{2.2}$$

In MetaD, the bias potential can be fine-tuned by adjusting the height and width of the Gaussians [23]. The height sets the intensity of each bias increment, while the width determines the spread of each increment across the landscape.

In wt-MetaD, the Gaussian potential heights are scaled down over time, governed by the bias factor $T+\Delta T$, where T is the system temperature and ΔT is an adjustable parameter (see eq 2.3). Tuning ΔT enhances barrier crossing, allowing broader exploration of the CV space. Moreover, setting a finite ΔT confines the exploration to regions of the free energy surface (FES) that are energetically relevant, ensuring that the focus remains on physically significant areas [38]. This approach improves statistical accuracy within these regions, enabling a more precise reconstruction of the free energy surface while also facilitating the exploration of previously inaccessible energy landscapes.

$$F(s,t) = -\frac{(T+\Delta T)}{\Delta T} V_G(S,t) \tag{2.3}$$

2.2.1 Selection of Collective Variables (CVs)

Two geometry-related CVs were defined to efficiently explore the conformational landscape of CaM (Figure 2.1.b, Figure 2.1.c). To represent the switch between open and closed conformations of CaM and following previous work in our group [50], we defined the linker end-to-end distance (l_{linker}) within the CaM linker region as the distance between the C_α atom of residue 69, marked as the beginning, and the C_α atom of residue 91, marked as the end. We also used the coarse-grained torsion angle (φ) to describe the rotational motions of the two lobes of CaM relative to each other. For this definition, we used the center of mass of the N-terminal domain, the C_α atoms at the beginning and end of the helical linker region, and the center of mass of the C-terminal domain in that order.

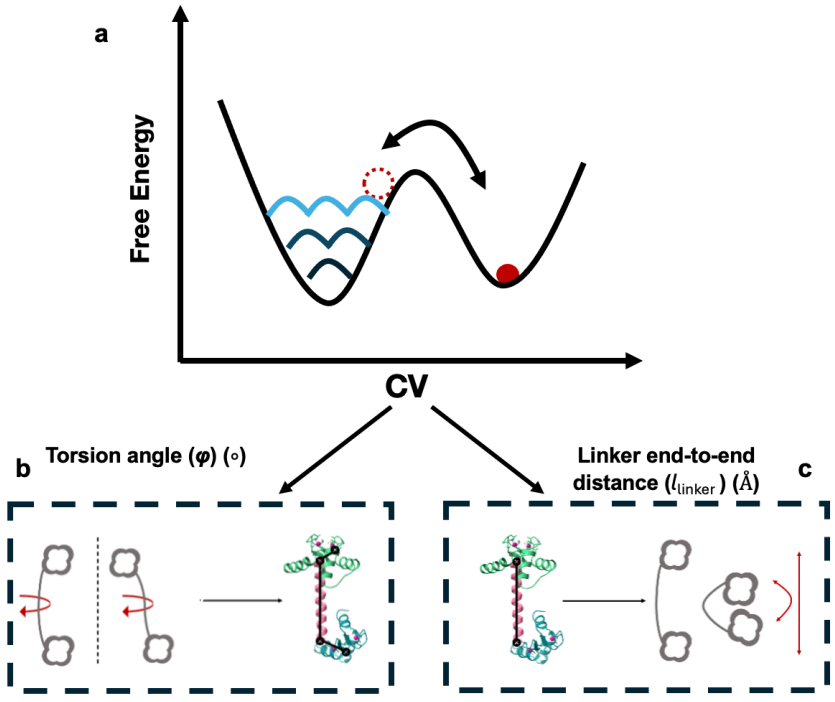

Figure 2.1 **a** Schematic representation of MetaD simulations, and selected collective variables for wt-MetaD simulations; **b** torsion angle to observe rotational motions of CaM, and **c** distance between helical-linker region to describe the transition between open/closed conformations.

2.2.2 System Preparation and wt-MetaD Simulations

We selected snapshots from MD simulation trajectories for each condition as starter structures for wt-MetaD simulations (see Chapter 3.1). During this selection phase, we focused on the RMSD and reduced DoF results. The most stable conformers from each condition were selected, specifically by targeting areas with minimal changes in RMSD (Figure 3.1). Additionally, φ) l_{linker} plots (Figure 3.3) were crucial in confirming that these selected structures accurately represented the most probable torsion and distance configurations throughout the trajectory. After selecting the snapshots, we removed the water molecules and KCl ions. We then immersed the structures in water and salt conditions that correspond to the conditions under which the wt-MetaD will be carried out. We applied the same system preparation protocol used in the initial MD simulations for consistency (section 2.1). For instance, a snapshot from the $3CLN^{Ca2+}{}_{IS\_L}$ simulation, low ionic strength condition was used to set up a new Ca^{2+}-loaded, low IS system. Using VMD, we generated PSF files, soaked the protein in a water box with at least 10 Å of water on all sides, and added KCl ions for ionization according to the protocol in Table 2.1. Following the same MD simulation protocol, each system was minimized for 10,000 steps, followed by a 100 ns equilibrium simulation to reach a stable local minimum. This approach ensured that the starting configurations for the wt-MetaD simulations began from well-defined minima, allowing Gaussian hills to be deposited effectively for enhanced sampling.

Using the atomic coordinates and velocities obtained from the minimization and equilibration simulations, wt-MetaD simulations were performed with the selected CVs detailed in subsection 2.2.1 (also see CV parameter file in Figure A.1). The upper and lower limits of l_{linker} were defined as 10–50 Å and divided into 1.5 Å grid intervals, while φ was divided into 10° grids spanning the range $[-\pi,\pi]$. The parameters of the simulations are listed in Table 2.2. The bias temperature of the wt-MetaD is set to 1490 K.

Table 2.2 WT-MetaD Simulation Parameters and Duration for Each System

	Parameters		Systems			
			$3CLN^{Ca2+}_{IS\_P}$	$3CLN^{Ca2+}_{IS\_L}$	$3CLN_{IS\_P}$	$3CLN_{IS\_L}$
Collective Variables	Linker End-to-End Distance (l_{linker})	Upper Limit (Å)	50	50	50	50
		Lower Limit (Å)	10	10	10	10
		Width	1,5	1,5	1,5	1,5
	Torsion Angle (φ)	Upper Limit (Å)	180	180	180	180
		Lower Limit (Å)	-180	-180	-180	-180
		Width	10	10	10	10
Gaussian Potential Height			0,2	0,2	0,2	0,2
Gaussian Potential Frequency			500	500	500	500
Gaussian Potential Width (kcal/mol)			1,0	1,0	1,0	1,0
Well-Tempered			1490 K	1490 K	1490 K	1490 K
Simulation Time(ns)			400	400	400	450

3. RESULTS AND DISCUSSION

3.1 Classical Molecular Dynamics (MD) Simulations

We performed two independent 1 μs MD simulations for each system to observe how the absence/presence of calcium ions and variations in ionic concentration affect the conformational changes of CaM. To analyze these changes, we conducted RMSD analyses.

The RMSD of the C_α atoms in the NTD, CTD, and linker regions of the protein were compared to their initial starting structures for each MD trajectory (Figure 3.1). In physiological IS, RMSD fluctuated between 3–9 Å in *holo* CaM across both simulations. The NTD, CTD, and linker showed limited movement throughout the trajectory compared to other systems (Figure 3.1.a). In the low IS system, although the first run showed similarity to the physiological case, the RMSD of the CTD increased significantly after 400 ns in the second simulation, leading to an overall RMSD increase due to CTD motion. This may be due to the C-lobe of CaM being located closer to the N-lobe, resulting in CaM adopting a more closed conformation (Figure 3.2).

In *apo* CaM under physiological conditions, the overall RMSD fluctuates between 4 and 10 Å (Figure 3.1.c). This large variation in RMSD is mainly due to the motion of the N-lobe, which fluctuates between 4–8 Å in the first run and 4–6 Å in the second. In contrast, the linker and the C-lobe are more stable. Similarly, in the low IS system (Figure 3.1.d), the RMSD of *apo*-CaM also ranges from 4 to 10 Å, with higher fluctuations in the N-lobe compared to the other parts of the protein. This result indicates that, removal of Ca^{2+} ions increased the motion of N-lobe irrespective of the ionic concentration in the system. Consequently, the overall RMSD fluctuations of CaM also increase due to this enhanced mobility.

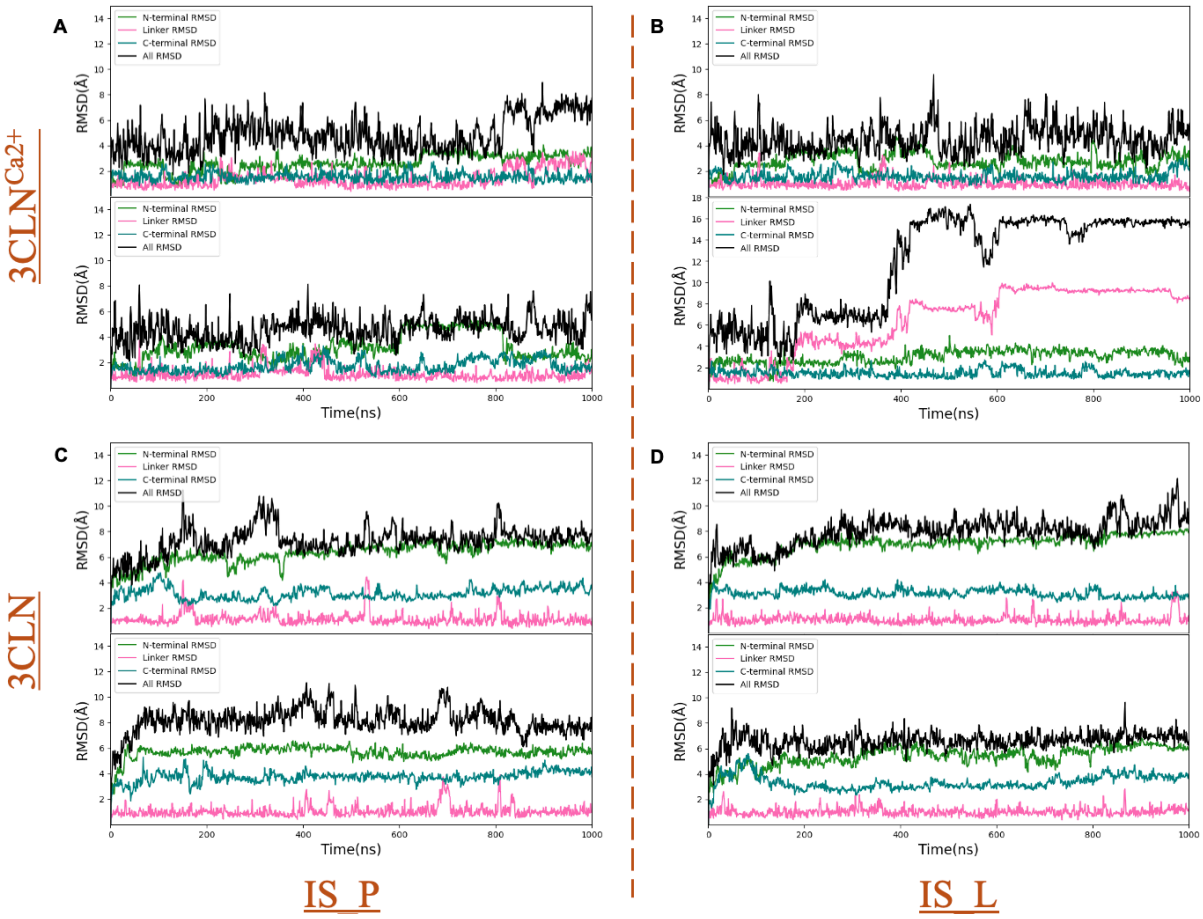

Figure 3.1 RMSD graphs of two independent 1μs long simulations of CaM in different environmental conditions. N-terminal domain, C-terminal domain, flexible linker region and overall RMSD depicted in green, blue, pink and black, respectively

Capturing the full range of protein motions is challenging due to the high dimensionality of the trajectory. Therefore, we reduced the high-dimensional motions of CaM into a two-dimensional projection by defining two degrees of freedom, φ and l_{linker} as described in section 2.2.1. These CVs allow us to follow the relative positioning of the N- and C-lobes, each of which has relatively low internal mobility, with respect to each other [50]. Figure 3.2 illustrates the time-dependent changes in these DoF, showing how the dynamics of the system evolved throughout the simulation. Additionally, in Figure 3.3, we combined the data from these two independent simulations based on their DoF to gain information about the conformational landscape of the protein.

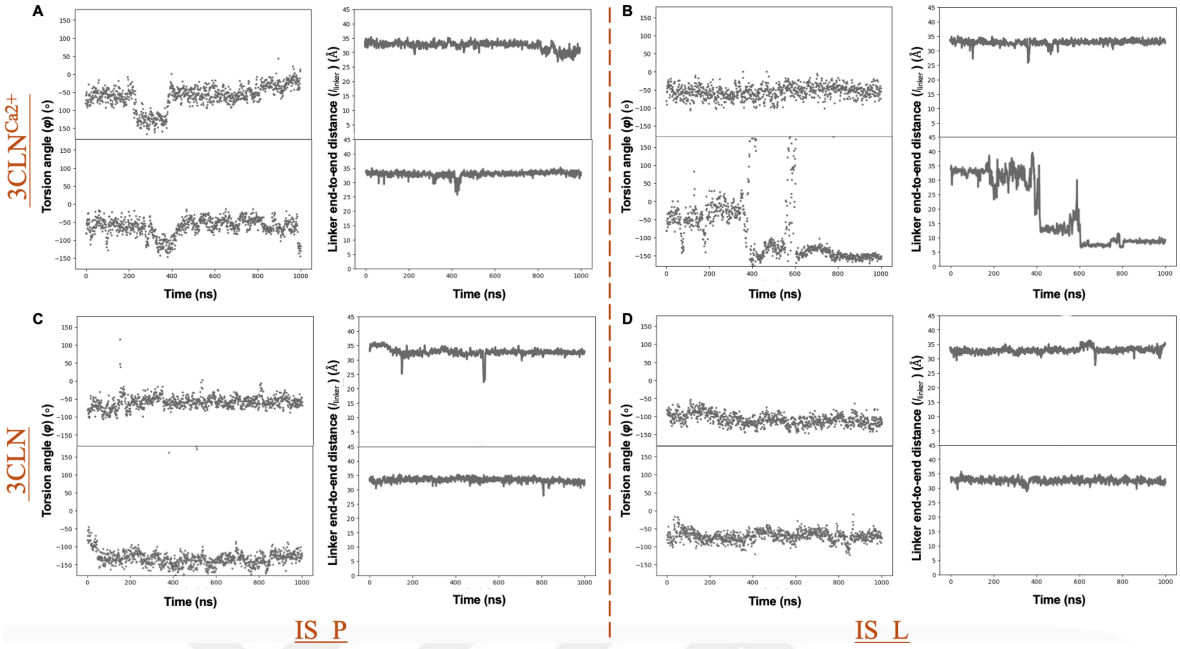

Figure 3.2 Changes in selected DoF over time during the two independent simulations for each system.

In high ion concentration, *holo* CaM (Figure 3.3.a), stuck in its open conformation, exhibited only rotational motions from 0° in the *trans* position to -150° in the *cis* position. In both simulations (Figure 3.2.a), the torsion angle initially stayed around 0° to -50°. Between 200 and 400 ns, it shifted to a lower range, fluctuating between -100° and -150°. After ca. 400 ns, the torsion angle returned to higher values.

At low ionic concentration, calcium-loaded CaM visited two distinct minima (Figure 3.3.b.). The first represents the *trans* open conformation, occurring between 0° and -100° and at distances of 25 to 37 Å, while the second corresponds to a compact form, where the two lobes face each other in the *cis* position, with distances of 10 to 15 Å and angles in the range of [-150°,-100°]. In the second run (Figure 3.2.b), the two lobes of CaM rotate relative to each other, which triggers them to move closer together and assume a compact conformation that is not observed in any of the other systems. This movement results in the formation of two distinct minima.

Under physiological conditions, *apo* CaM predominantly adopts an open conformation with an inter-lobe distance of 30–35 Å (Figure 3.3.c.). Two slightly separated minima are

observed: one between -20° and -70° in the *trans* position, and the other between -100° and -180° in the *cis* position. This separation arises because, in the first run, the torsion angle remained mostly between -20° and -70°, while in the second run, it initially started at -20° but then shifted to a lower range between -100° and -180° (see Figure 3.2.c).

Probability distribution of $3CLN_{IS\_L}$ revealed that lowering the salt concentration results in narrower energy wells (Figure 3.3.d.). The structure predominantly shows an open conformation with an inter-lobe distance between 30 and 37 Å, with two closed minima: one between -20° and -70°, and the other between -70° and -100°.

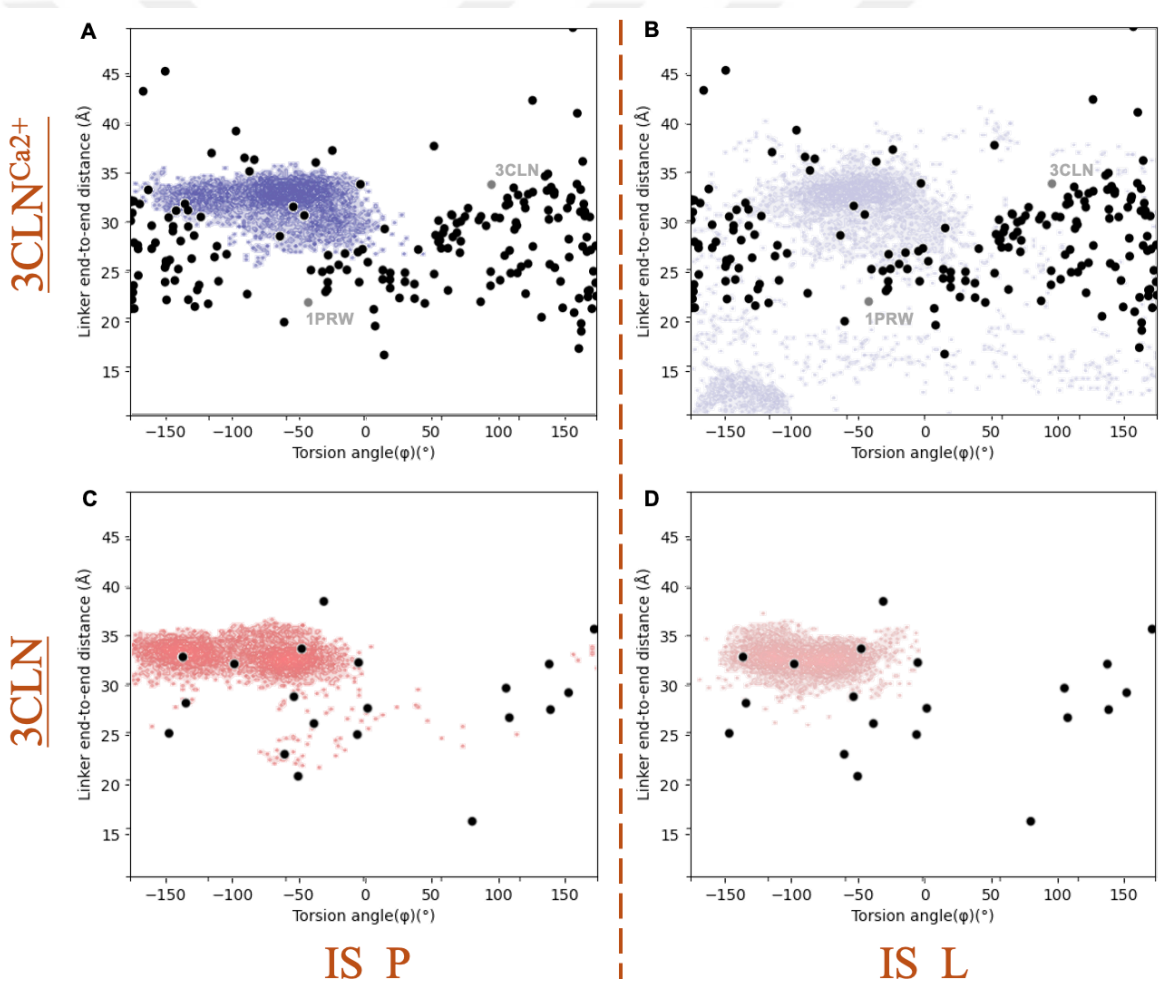

Figure 3.3 Torsion angle (φ) vs linker end-to-end distance (l_{linker}) joint probability plot of classical MD simulation results under the four environmental conditions studied. **a** Calcium-loaded CaM in physiological, and **b** low IS, represented by dark and light lavender, respectively. **c** *apo* CaM in physiological, and **d** low IS, shown with coral pink and peachy pink. Black and gray dots represent NMR and crystal structures respectively.

To compare our MD simulations with experimentally determined structures of CaM, we obtained Ca^{2+} loaded (PDB ID: 2K0E), *holo* WW-7 peptide bound (PDB ID: 1MUX, see Table 1.1) , *apo* CaM (PDB ID: 6Y95, also shown in Figure 1.1.c), solution NMR structures, as well as the crystal structures of Ca^{2+}-loaded CaM in its closed and open forms (PDB IDs: 1PRW and 3CLN, respectively), from the PDB. These structures were projected onto the same plots in their respective experimental conditions in Figure 3.3, with NMR structures shown as black dots and crystal structures as gray dots.

In calcium-loaded CaM under physiological conditions, some projected NMR structures fall within the joint probability distribution, particularly in the open conformation region with linker distances around 30–35 Å. However, other NMR conformations, especially those representing more compact or alternative structural states, are clearly not sampled by the MD trajectories.

Similarly, in calcium-loaded CaM under low ionic strength conditions, while a subset of NMR structures overlaps with the MD-sampled conformational space, the trajectory does not reach all NMR-derived conformations, especially those with different torsion angles or tighter linker distances.

For *apo* CaM in physiological conditions, the MD trajectory samples a broad range of open conformations that overlap with most NMR structures. However, certain NMR conformations observed in more compact or closed forms remain unsampled in the MD simulations, indicating that the full spectrum of experimental conformations is not captured.

Overall, this general approach is not efficient to capture all possible *apo*/*holo* conformations of CaM under different environmental conditions because they often got stuck in one or two energy minima and could not cross high energy barriers (Figure 3.1). To overcome this problem, we used enhanced sampling techniques to explore the molecular configurational space. In the following sections, we discuss our results as obtained from well-tempered simulations.

3.2 Well-Tempered Metadynamics Simulations

We performed 400 ns-long wt-MetaD simulations for each system to efficiently sample the energy landscape of CaM under different environmental conditions, except for $3CLN_{IS\_P}$ (Ca^{2+} -unbound CaM in physiological condition), which we extended to observe how the system behaves at the newly sampled lower linker distance (Figure 3.4.f and Figure 3.4.e).

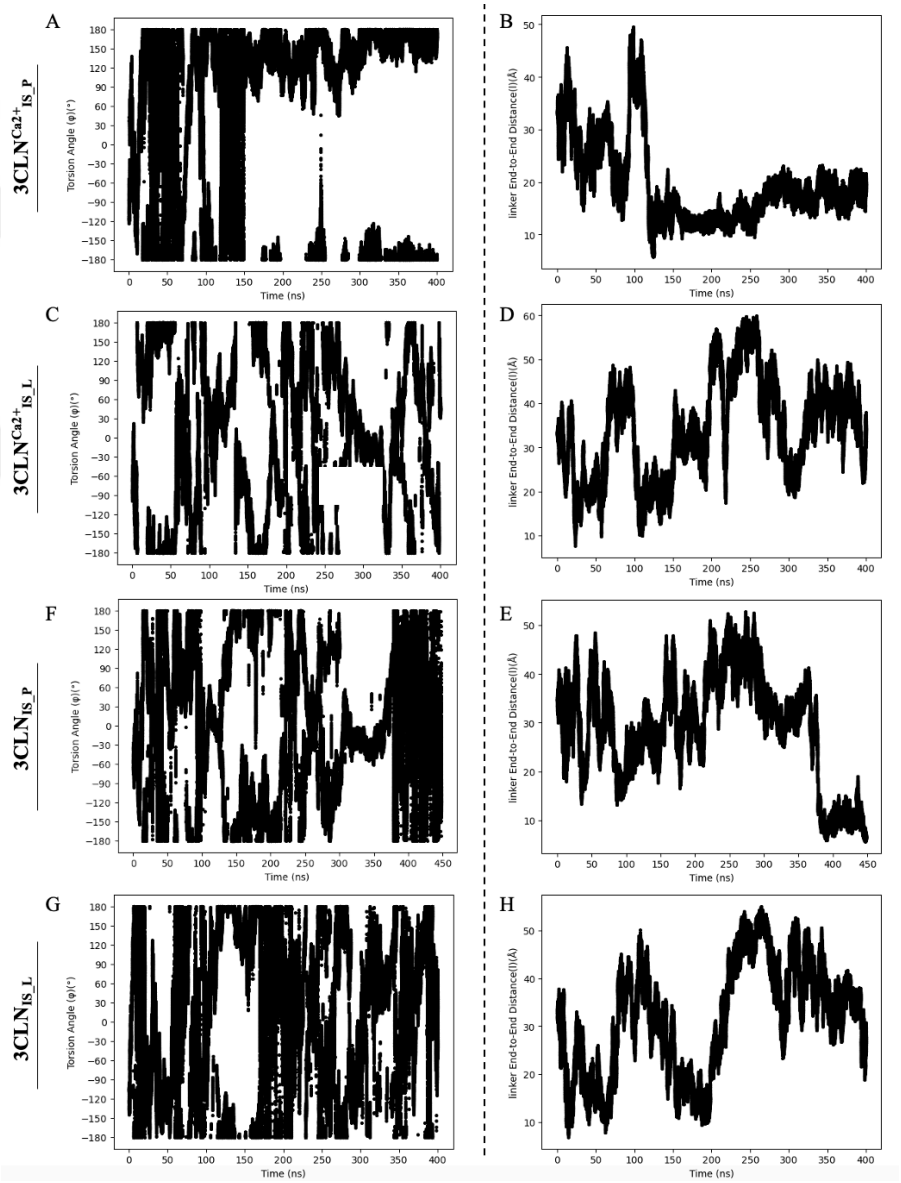

Figure 3.4 wt-MetaD trajectories for two selected CVs: the left panel depicts changes in the torsional angle (°), while the right panel shows the linker end-to-end distance (Å) over time across all systems.

The CVs vs. time plots (Figure 3.4) illustrate that in the $3CLN^{Ca2+}{}_{IS\_P}$ system, initially, the entire range was sampled evenly, and after 150 ns, a very deep minimum was found in the [100°,180°] range (Figure 3.4.a), leading to continued sampling in that region. For the other systems, during the simulation time, all possible torsional angles were sampled (Figure 3.4.c, f and g). Based on the distance vs. time graphs, conformational changes were observed in which the distance between CaM's lobes increased and decreased, allowing various open and closed structures to be observed during the simulations (Figure 3.4.b, d, e and h).

We note that, although upper and lower limits of l_{linker} were defined as 10–50 Å, the system occasionally explores higher or lower distances during simulations. These upper and lower limits create grids representing the anticipated range for efficient sampling but do not restrict the actual values the system may explore. Therefore, fluctuations in distance may occur due to thermal motion, protein flexibility, or external forces.

To gain insights into the free energy landscape of each system, as shown in Figure 3.5, we normalized the energies of each minimum by using the open dumbbell-shaped crystal structure of CaM (PDB ID: 3CLN) as a reference (see Chapter 3.1),which also serves as our starting structure (see Chapter 2.1). To do this, we calculated the degrees of freedom of 3CLN and compared these with the potential mean force (PMF) results for each system. We then extracted the energies corresponding to the same dihedral angles and distances as those in 3CLN, using these values as baselines for normalization; i.e. zero of free energy. Then, we scaled the free energy surface plots based on the minimum and maximum energy values calculated for each system. By determining the range of energies in the PMF data, we adjusted each plot to span a consistent range, enabling standardized comparisons across systems. The color bar in Figure 3.5 indicates the standardized range.

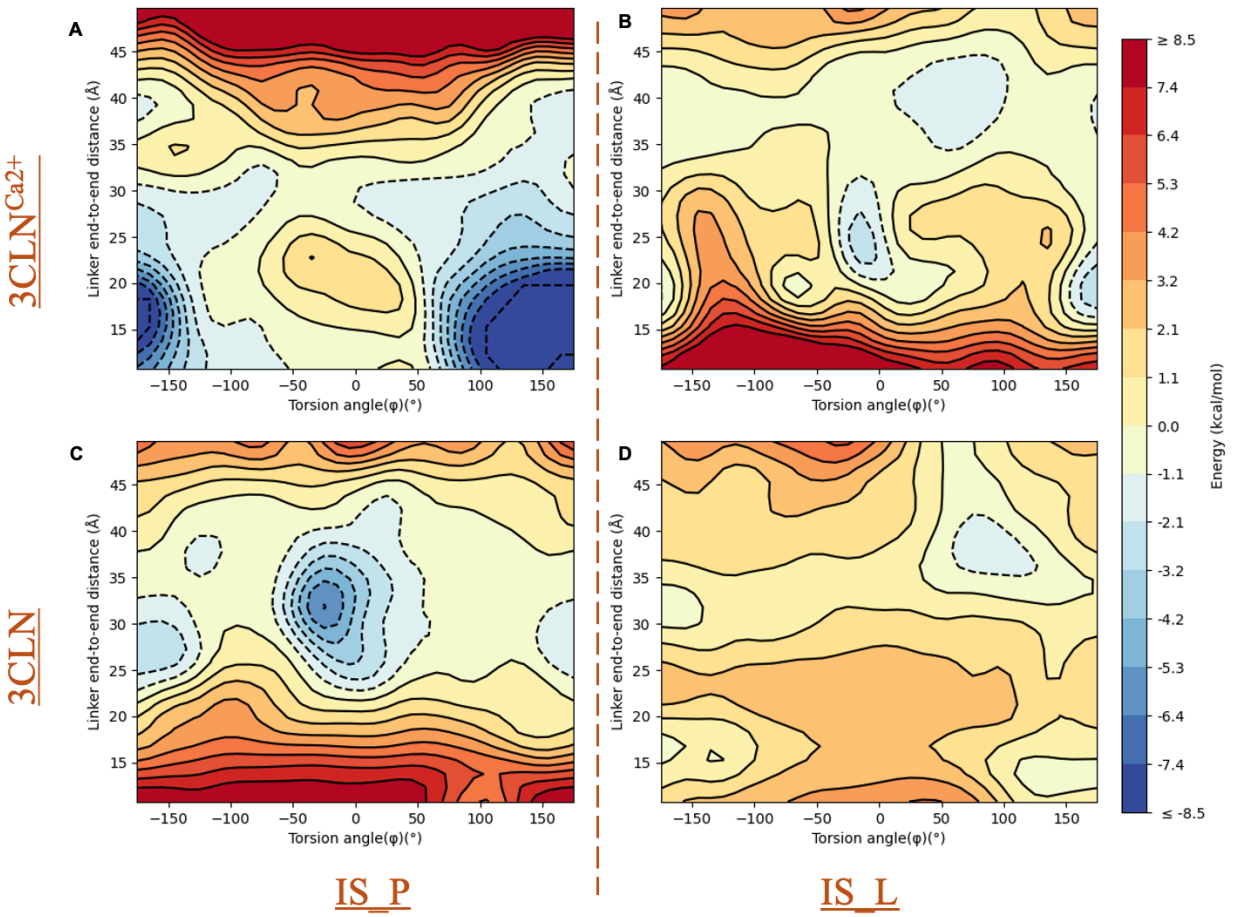

Figure 3.5 Free energy surface of CaM obtained from well-tempered metadynamics in different environmental conditions; **a** *Holo* CaM in physiological and **b** low IS environments, **c** Ca^{2+}-free CaM physiological and **d** low IS environments.

The free energy surface of $3CLN^{Ca2+}{}_{IS\_P}$, obtained from wt-MetaD and shown in Figure 3.5.a, reveals two prominent energy minima. The first minimum is characterized by a torsion angle in the range [100°,180°] between the two lobes of the protein, with a linker distance spanning the range 10–20 Å. This minimum corresponds to a *cis*-closed conformation, which is even more compact than the closed conformation of CaM reported in 1PRW coded PDB structure (see in Figure 1.1.c.). The second minimum lies again in the cis conformation of the torsional angle but with a linker distance of 37–41 Å, representing a *cis*-open conformation.

The energy difference between these minima is 7.3 kcal/mol, indicating that the first minimum is significantly deeper than the second. This suggests that, under physiological conditions, calcium-loaded CaM strongly favors the closed *cis* conformation.

In contrast, the free energy surface of calcium-loaded CaM ($3CLN^{Ca2+}_{IS\_L}$) under low ionic strength conditions reveals four relatively shallow minima, each corresponding to distinct structural conformations (Figure 3.5.b.). The first minimum, with a torsion angle of -20° to 20° and a linker distance of 20–30 Å, suggests a *trans*-semiclosed conformation with an energy at a depth of approximately -6 kcal/mol. The second minimum, at a torsion angle of 40–100° and a distance of 35–45 Å, indicates a *cis*-open conformation with an energy at a depth of -2 kcal/mol. The third minimum, with a torsion angle of 150–180° and a distance of 15–25 Å, represents a *cis*-closed conformation with an energy depth of around -6 kcal/mol. The fourth minimum, at a torsion angle of 170–180° and a distance of 37–43 Å, also represents a *cis*-open conformation, with an energy depth of approximately -1 kcal/mol. Representative structures selected from these minima align with those reported in PDB IDs 1MUX and 2K0E (Table 1.1), where many NMR-derived structures similarly align with these minima in terms of dihedral angles and distances.

For the *apo* structure, $3CLN_{IS\_P}$, two distinct minima are observed (Figure 3.5.c.). The first minimum, characterized by a torsion angle ranging from -180° to -130° and a distance of 20–33 Å, suggests a *trans*-semiclosed conformation with an energy depth of approximately 7.8 kcal/mol. The second minimum features structures with torsion angles between -50° and 50° and distances of 23–40 Å, indicating a mix of both open and closed conformations in *trans* and *cis* orientations, with an energy depth also around 7.8 kcal/mol. Notably, many structures from the NMR dataset of *apo* CaM (PDB ID: 6Y95; see Figure 1.1.b) align closely with these minima in terms of dihedral angles and distances.

The free energy surface of *apo* CaM at low IS shows a single minimum with a dihedral angle between 60° and 150° and linker distances ranging from 34 to 42 Å, indicating an *trans*-open conformation with an energy depth of approximately 3.8 kcal/mol (Figure 3.5.d.). This energy minimum is shallower compared to other systems and closely resembles the dumbbell-shaped open structure of *holo* CaM (PDB ID: 3CLN; see Figure 1.1.a).

Overall, lowering ion concentration leads to shallower energy barriers between conformers, and facilitates easier navigation along the energy landscape, allowing for a variety of accessible conformations. However, under physiological conditions, deeper energy minima are formed, indicating a more stable and specific conformational state, apparently facilitated by the additional ions in the system.

3.3 MD Simulations for Exploring Structural Basis of wt-MetaD Results

To understand why the minima are structured as observed and validate the effectiveness of our selected CVs, we selected representative structures and performed minimization followed by equilibration runs with classical MD. We focused on the $3CLN^{Ca2+}{}_{IS\_L}$ condition, as it displayed energy minima common to those observed under the other conditions, including both *holo* and *apo* CaM at physiological and low ionic concentrations (Figure 3.5). From this condition, we selected four structures representative of the common minima observed (Figure 3.6).

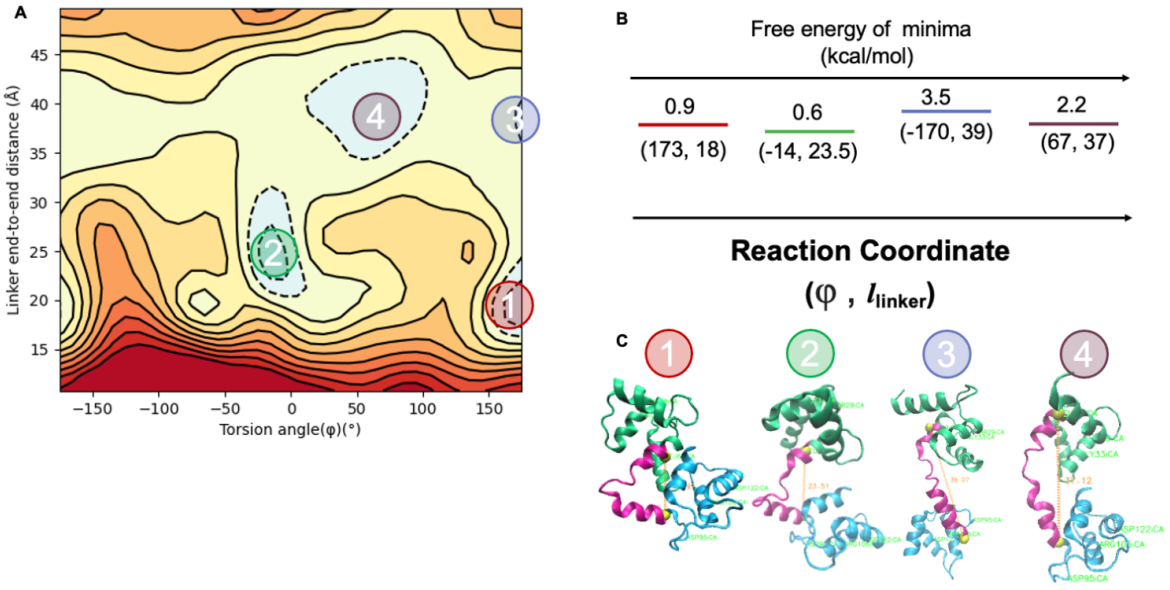

Figure 3.6 **a** Four selected conformers from $3CLN^{Ca2+}{}_{IS\_L}$ metadynamics run depicted with red, green, blue and purple circles respectively, **b** along with dihedral angles, distances, free energies, and **c** three-dimensional structures belonging to these conformers.

The first conformer adopts a *cis*-closed conformation, represented by '1' in Figure 3.6.a. It shares a similar energy minimum with $3CLN^{Ca2+}{}_{IS\_P}$ and resembles the *apo*-CaM structure in physiological conditions, although with a slightly reduced lower distance. The second structure '2' originates from an energy minimum closely matching that of the *apo* CaM at physiological condition, adopting a *trans*-closed conformation in a slightly more compact form. The third conformer, '3', also shares a similar minimum with $3CLN^{Ca2+}{}_{IS\_P}$ and adopts an open *cis* conformation. The final conformer, labeled '4', shares the same minimum as apon CaM in low IS, with an *trans*-open conformation.

After selecting the conformers, we removed water molecules and salts from these snapshots, then performed a 10.000-step minimization (Figure 3.7) followed by a 100 ns equilibration run for each structure, using the same MD simulation protocol as described in Chapter 2.1.1 across all systems. In some cases, the runs were extended to 200 ns to gain deeper insights into the dynamics. We also generated probability distribution graphs, similar to those in Chapter 3.1, for these four conformers under different conditions to effectively project the trajectory of these simulations (Figure 3.8).

During the minimization process, the conformers remained within specific torsion angles and distances, exhibiting minimal shifts from their initial configurations (Figure 3.7). This suggests that each selected conformer was stabilized within the respective energy minimum and did not explore other regions of the conformational landscape, despite them being selected from the $3CLN^{Ca2+}{}_{IS\_L}$ metadynamics run results (Figure 3.6). This result confirms the effectiveness of our selected CVs, as the conformers were trapped within local minima, indicating that these CVs accurately represent stable conformations.

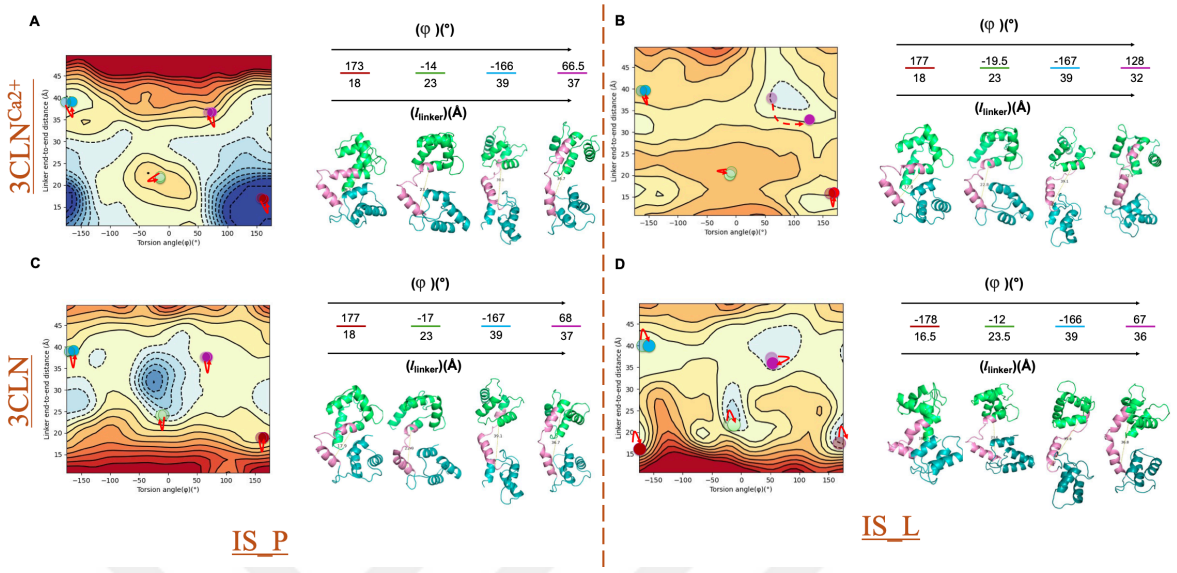

Figure 3.7 Minimization fates of selected conformers. Initial conformers are shown with opaque circles, while the final snapshots by the end of the minimization process are represented by transparent circles in the same colors. Torsion angles and dihedral values are displayed next to the plots, with the three dimensional structures of these final snapshots positioned alongside each system: **a** *holo* CaM in physiological conditions, **b** *holo* CaM at low IS, **c** *apo* CaM in physiological conditions and **d** *apo* CaM at low IS.

In the equilibration runs, both calcium bound and unbound CaM in physiological conditions exhibited behavior consistent with the DOFs observed in the MetaD simulations (Figure 3.8.a and Figure 3.8.c). However, under low IS conditions, despite the much shallower energy surface uncovered by the wt-MetaD simulations, some conformers became trapped in DOFs that do not align with expected conformational states (Figure 3.8.b and Figure 3.8.d). These unexpected states, which were not typically sampled in wt-MetaD simulations, indicate that the systems might be stuck in local minima *en route* to the actual minima. We reasoned that these likely result from the altered electrostatic environment in the low IS conditions.

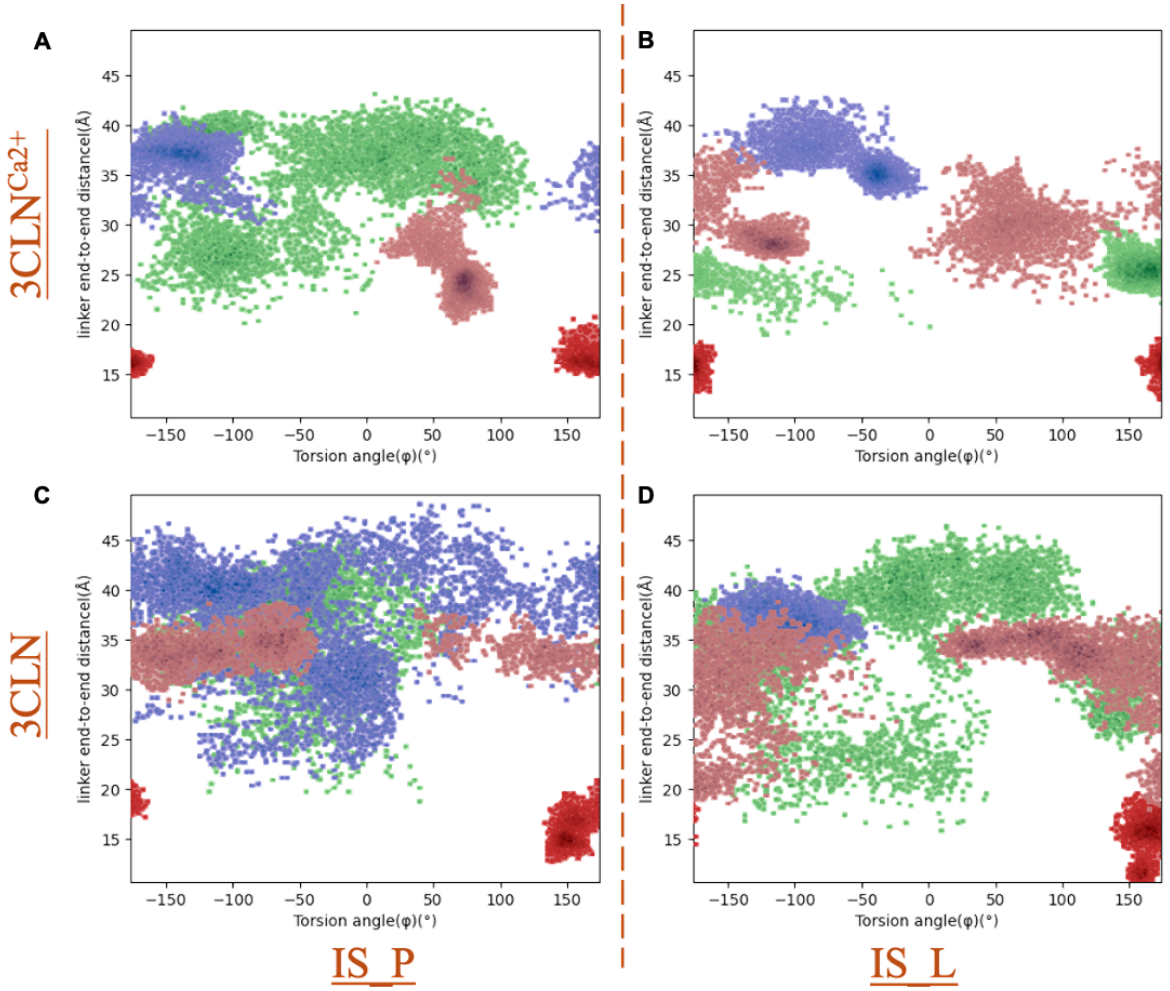

Figure 3.8 Torsion angle/ linker end-to-end distance graphs of the MD simulation results for each conformer across all systems, using a consistent color scale: probability distribution of conformer 1 is shown in red, conformer 2 in green, conformer 3 in blue, and conformer 4 in pink.

Conformer-1 also remained trapped in a single minimum, adopting a *cis*-closed structure, and did not overcome the energy barrier, regardless of the environmental factors (Figure 3.8.c and Figure 3.8.d). In contrast, the other conformers explored a range of torsional angles and distances on the conformational surface. To investigate the factors contributing to this restricted behavior in conformer-1, salt bridge interactions were analyzed across all systems. Salt bridge occupancies were determined from a 100 ns MD simulation of the first conformer for each condition. Unique salt bridges were identified and those observed exclusively for the first conformer were identified. These salt bridges were then filtered based on

occupancies during the trajectories, as calculated by the VMD hydrogen-bond plugin, with particular emphasis on those with substantially increased values. (Table 3.2).

Table 3.2 Salt Bridge occupancies (%) of Conformer 1

	Holo		Apo	
	$3CLN^{Ca2+}_{IS\_P}$	$3CLN^{Ca2+}_{IS\_L}$	$3CLN_{IS\_P}$	$3CLN_{IS\_L}$
E47-K75	54	14	25	46
K77-E139	23	50	82	82
E6-K94	85	90	68	47
Total	162	154	168	175

Salt bridge interactions between E6 and K94 (Figure 3.9) were notably strong, with the salt bridge distance fluctuating between 3.2 and 5 Å throughout the trajectory across all systems. However, in low IS conditions of *apo* CaM, this distance increased to as much as 10 Å.

The salt bridge interactions between K77 and E139 repeatedly break and reconnect throughout the trajectory. In $3CLN^{Ca2+}_{IS\_P}$ (Figure 3.9.a), the distance varies from 5 to 10 Å. In $3CLN^{Ca2+}_{IS\_L}$ (Figure 3.9.b), it initially fluctuates from 5 to 7 Å during the first 50 ns, then shifts to 3–5 Å region before returning to 5–10 Å after 60 ns. In $3CLN_{IS\_P}$ (Figure 3.9.c), the distance alternates from 5–10 Å to 3.2 Å, eventually stabilizing at 3.2 Å after 60 ns. In $3CLN_{IS\_L}$ (Figure 3.9.d), these interactions are mostly stable, fluctuating between 3.2 and 7 Å.

In all systems except $3CLN_{IS\_P}$ (Figure 3.9.c), the salt bridge between E47 and K75 forms only after other interactions are established. However, in *apo* CaM under physiological conditions, this interaction was disrupted after 75 ns.

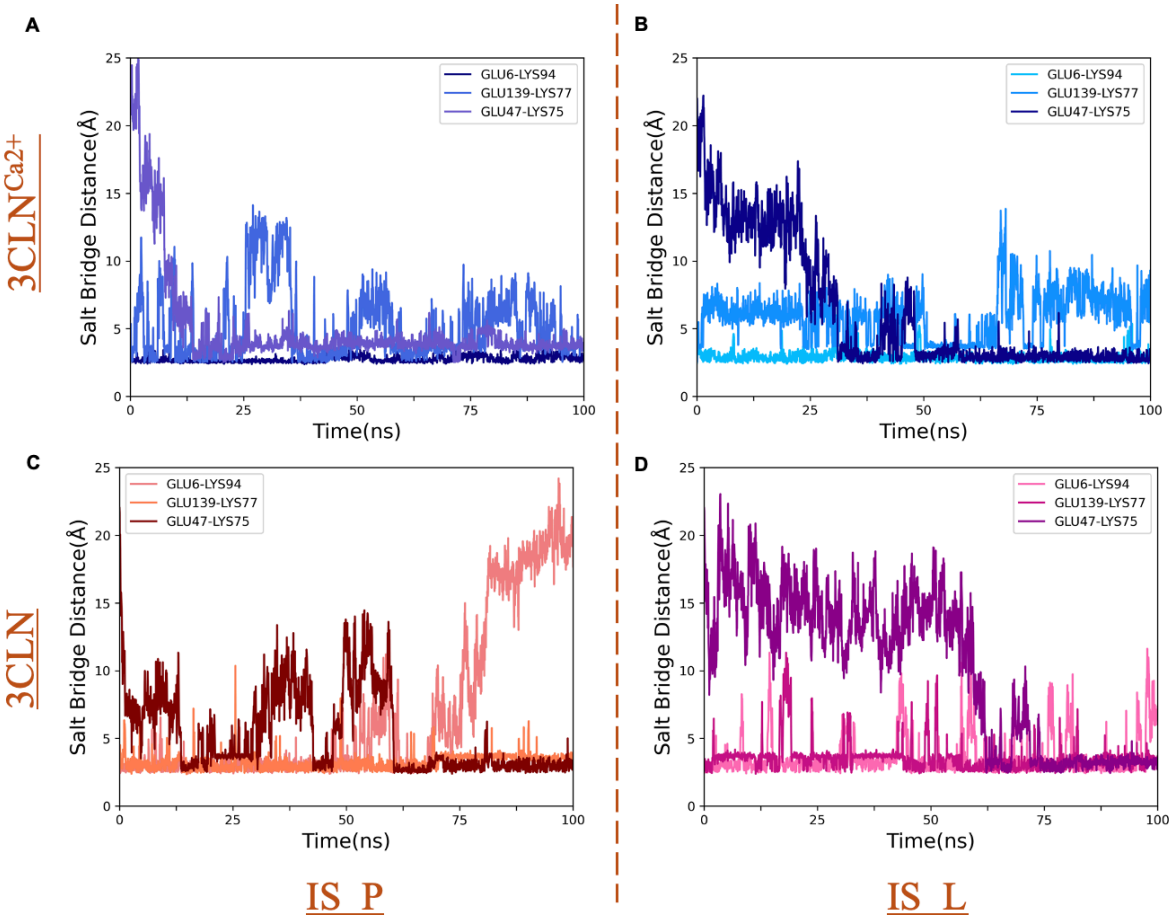

Figure 3.9 Changes in salt bridge interactions with high occupancies in CaM throughout the trajectory for each system

These unique salt bridges have an additional important feature: While the residues involved in forming these salt bridges are located in different regions of CaM, they all work to bridge the linker to the N- and the C-lobes (Figure 3.10). E6 and K94 are positioned at the starting points of the N-terminal and C-terminal domains of CaM, respectively (Figure 3.10.a). This strong salt bridge drawing them closer together and causing the long helical linker connecting them to bend. The salt bridge between the linker region and the CTD (LYS77−GLU139) brings the linker and CTD closer together (Figure 3.10.b). Throughout the trajectories, significant fluctuations suggest that the CTD tends to distance itself from the linker, but the salt bridges between K77-E139 and E6-K94 constrain this separation, limiting further displacement. The hydrogen bond network between the NTD-linker (E47−K75) and

interdomain residues (K75−D80), with E47−K75 also forming a salt bridge, creates a stronger interaction that contributes to the contraction of the linker region, reducing the distance between the N-lobe and the linker (Figure 3.10.c). The hydrogen bond between D80 and K75, where D80 acts as the hydrogen donor, further stabilizing the linker region. The E47−K75 interaction brings these regions closer together, enhancing the compactness of the overall structure.

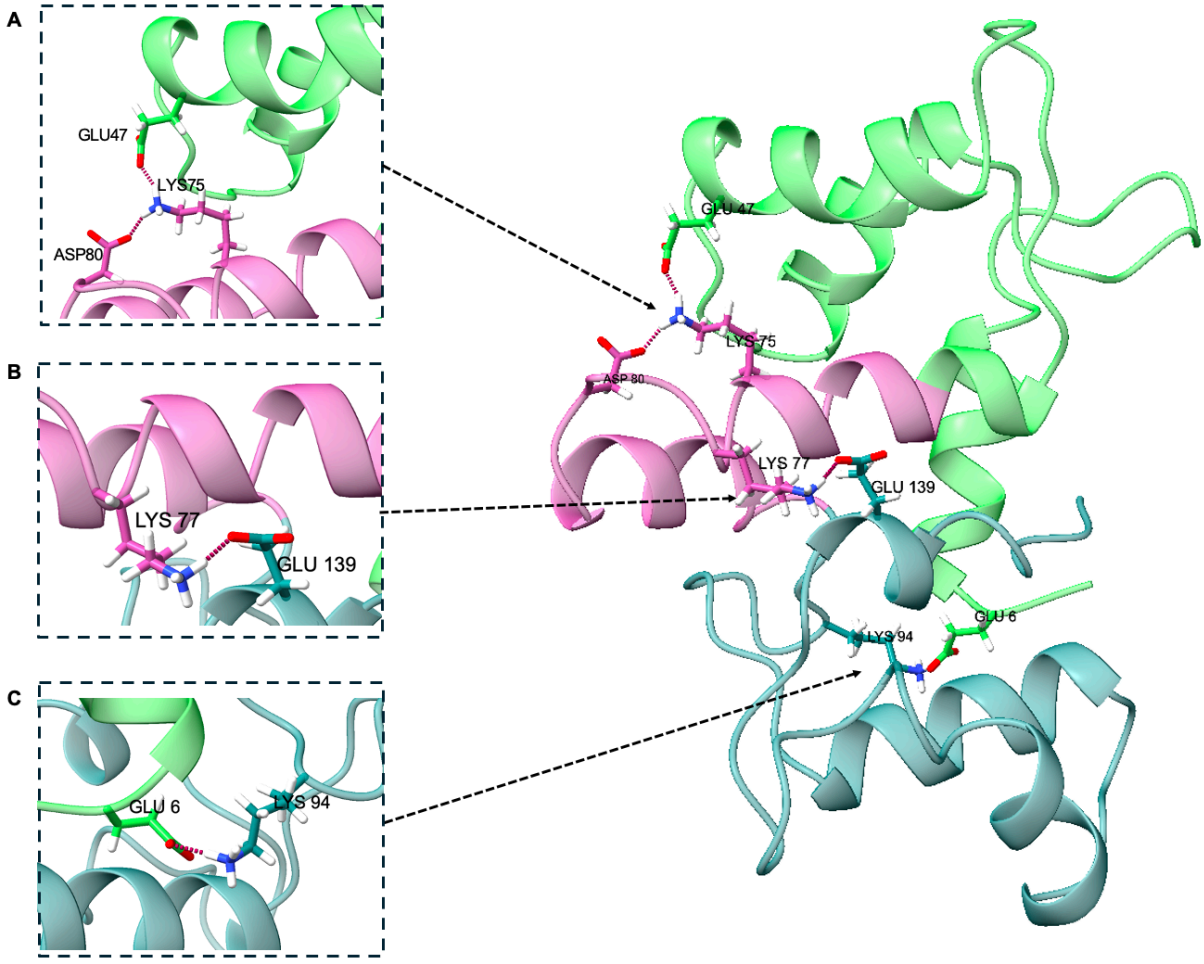

Figure 3.10 Salt bridge interactions which holds conformer 1 in cis-closed conformation thoughout the trajectory for all systems

In conformer 1, CaM appears to be unable to undergo rotational motion, possibly due to the presence of these salt bridges. Additionally, the protein adopts a more compact structure, likely influenced by these stabilizing interactions.

We then investigated the effect of reducing ion concentration, with a particular focus on the compact structure of conformer 3 in low IS holo CaM. To quantify the distribution of charges around the protein, we calculated the radial distribution function between the side chains of negatively charged residues of the protein and potassium ions for all systems (Figure 3.11.b). The first coordination shell between negatively charged side chains and K^+ ions is observed at a distance of 2–4 Å, while the second coordination shell appears at 4–6 Å.

Irrespective of the absence or presence of Ca^{2+} ions, K^+ ions are closer to the negatively charged residues of CaM in low IS, although in physiological conditions, the system has more positively charged ions (Figure 3.11). This can be explained with the electrical double layer formed around the protein [58]. As depicted in Table 2.1, CaM is highly negatively charged, therefore in physiological IS, negatively charged parts of CaM attracts positively charged K^+ ions (in this case, they act as counter-ions) from the surrounding water [59]. These K^+ ions accumulate around CaM and form a layer close to the particle surface. At the same time, Cl^- acts as co-ions, repelled by highly negatively charged CaM. A second layer is formed with the attraction of the other counter-ions and repletion of the co-ions. In low IS conditions, however, only counter-ions are present, so they are more strongly attracted to the negative charges on CaM compared to physiological IS. Due to the absence of co-ions, a diffusive layer (second layer) cannot form, leading to a tighter clustering of ions around CaM.

The balanced ion distribution under physiological conditions leads to a broader, less dense shielding effect. This broader, diffused shielding allows CaM to transition between different conformations more freely than in low IS conditions.

In Figure 3.11 the electrostatic potential distribution for each condition was calculated using the Adaptive Poisson-Boltzmann Solver (APBS) tool [60] to visualize and interpret the electrostatic effects surrounding CaM. The calculations were based on a biomolecular dielectric constant of 2 and a solvent dielectric constant of 78.54, with a solvent molecule radius of 1.4 Å and a temperature of 310 K. For holo CaM, the van der Waals radius and

electrostatic charges of Ca^{2+} were determined using PDB2PQR, and the KCl salt concentration was adjusted according to each system's ionic strength.

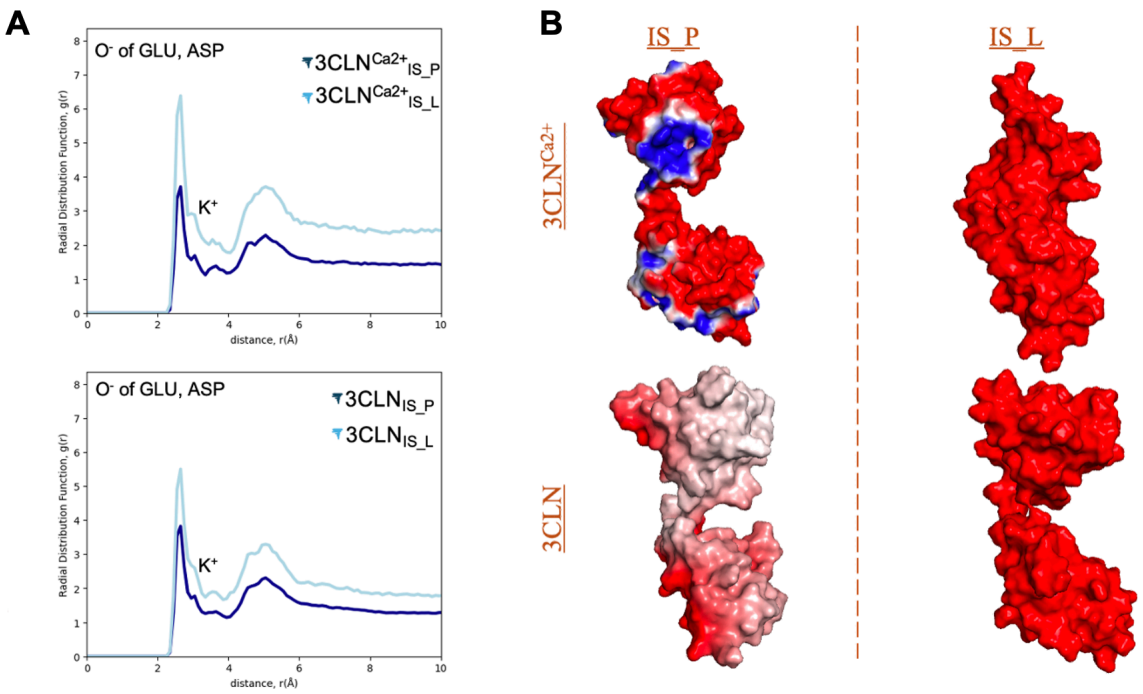

Figure 3.11 **a** Electrostatics isocontours of conformer 3 drawn at ± 1 k_BT/e for the studied as labelled. Blue regions represent positive potentials and red represent negative potentials at those points along the surface of the protein. **b** Radial distribution function, $g(r)$, between C_β atoms of the negatively charged residues and K^+ ions.

The electrostatic charge distribution results align with the findings from the RDF calculations. In high ionic concentration conditions, *holo* CaM displays partially positive and neutral charge regions (indicated in blue in Figure 3.11.a), while *apo* CaM exhibits partially neutral patches (shown in white). In contrast, for both *holo* and *apo* CaM at low IS, the surface of the protein is predominantly negatively charged (depicted in red in Figure 3.11.a). This effect arises because K^+ ions are positioned closer to the surface of the protein in low IS compared to physiological conditions.

To summarize, although the energy landscape is shallower in low IS conditions, ions seem to act as lubricants, facilitating the release of salt bridges that would otherwise trap conformers in specific local minima.

3.4 Subcellular Localization and Regulatory Roles of CaM across Cellular Compartments

Although CaM is primarily an intracellular protein found in the cytosol or nucleus [19], it also binds to proteins in various cellular compartments [61], including the endoplasmic reticulum (ER) and sarcoplasmic reticulum (SR), which serve as primary intracellular calcium storage sites [62]; mitochondria, which buffer cytosolic calcium and generate ATP through calcium-dependent activation of key enzymes in the tricarboxylic acid cycle [63]; and the Golgi apparatus, which maintains localized calcium gradients essential for vesicular trafficking and post-translational modifications [64] (Figure 3.12.a). Beyond regulating calcium levels, CaM also plays a crucial role in calcium signaling by undergoing conformational changes that enable interactions with target proteins in these organelles [19]. Depending on its binding partner and the calcium levels, CaM interacts either by wrapping around its targets or through specific lobes, either the N-lobe or C-lobe.

To explore how CaM adjusts its conformation for binding partners in different organelles and to assess the biological relevance of the conformations sampled in our wt-MetaD simulations, we compared these conformations with experimental structures of CaM bound to kinases, pumps, receptors and channels in terms of the same DoFs. This comparison revealed how CaM adopts distinct conformations to accommodate various biochemical environments, illustrating its versatility in facilitating interactions and regulatory functions across cellular compartments (Figure 3.12.b).

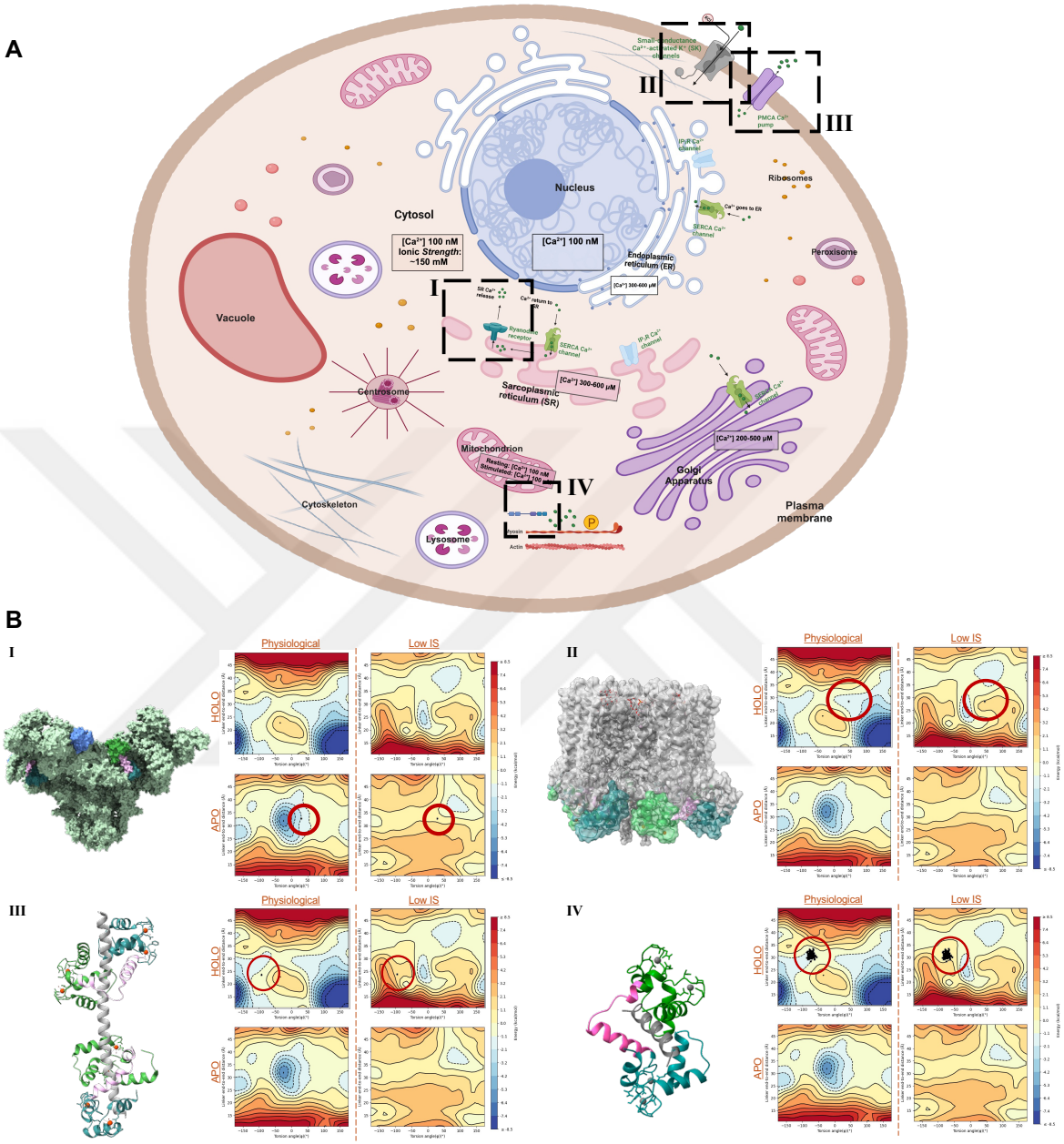

Figure 3.12 Insights into the biological roles of CaM. **a** Diagram of a eukaryotic cell highlighting the main organelles, the typical calcium concentration distribution across compartments, and the presence of proteins located in different parts of the cell, whose functions are regulated by CaM binding; **b** comparative conformational landscapes of calmodulin from wt-MetaD simulations for each system, overlayed with conformations of CaM from experimental structures with binding partners; binding of CaM to RyR at the SR (PDB ID: 6JI8) (I), to SK channel located in the plasma membrane (PDB ID: 6CNN) (II), to PMCA also located in the plasma membrane (PDB ID: 4AQR) (III), and MLCK in the cytoplasm (PDB ID: 2K0F) (IV).

Ryanodine receptors (RyRs) are intracellular calcium release channels located on the SR membrane and mediate the release of Ca^{2+} ions from the SR into the cytoplasm [65] (illustrated in Figure 3.12.b with I). RyRs contain several CaM binding motifs and are regulated by both *apo* and calcium-loaded CaM [66]. When *apo* CaM binds to RyR, it reduces RyR activity to prevent excessive calcium leakage; however, when calcium is bound to CaM, it fine-tunes RyR activity during excitation-contraction coupling, ensuring efficient and controlled calcium release [67]. The cryo-electron microscopy (cryo-EM) structure of RyR bound to apo-CaM reveals how *apo* CaM interacts with the receptor [67] (shown in Figure 3.12.b with I). The N-lobe of CaM is positioned in the upper part of a cleft formed by RyR's helical domain, while the C-lobe is situated at the bottom of the cleft, surrounded by the handle and central domains. When we compared this cryo-EM structure with the regions sampled in our metadynamics simulations, we found that *apo* CaM aligns with energy minima under physiological conditions, adopting a *trans*-semiclosed conformation.

Small-conductance calcium-activated potassium (SK) channels are integral membrane proteins that facilitate the flow of K^+ across the cell membrane [68] (depicted in Figure 3.12.a with II). They are activated by increases in intracellular Ca^{2+} levels, and they are also voltage-independent. Each SK channel tetramer associates with four *holo* CaMs [69] (see Figure 3.12.b.II). Upon an increase in intracellular Ca^{2+} concentration, the Ca^{2+} bound N-lobe of *holo* CaM interacts with the linker of the SK channel in a Ca^{2+} dependent manner [68]. This interaction triggers conformational changes in the SK channel, leading to the opening of the pore and allowing K^+ ions to flow through [69]. This interaction is crucial for SK channel functionality; without CaM, SK channels cannot respond to changes in intracellular Ca^{2+} levels and they become inactive. In the cryo-EM structure of the *holo* CaM-bound SK channel, *holo* CaM adopts a *trans*-semiclosed conformation. This conformation was also well sampled with wt-MetaD in $3CLN^{Ca2+}{}_{IS\_P}$ system.

The plasma membrane Ca^{2+}-ATPase (PMCA) is a transport protein located in the plasma membrane of eukaryotic cells [70, 71] (depicted in Figure 3.12.a with III), and functions to expel calcium ions (Ca^{2+}) from the cytosol into the extracellular space and to maintain low

intracellular Ca^{2+} concentrations. When intracellular Ca^{2+} levels rise, CaM binds to Ca^{2+} ions, undergoing a conformational change and *holo* CaM wraps around the binding domain in C-terminal tail of the PMCA [72]. This interaction with PMCA activates the pump, enhancing its ability to transport Ca^{2+} out of the cell [70]. The X-ray structure of *holo* CaM bound to PMCA reveals that once CaM wraps around the binding domain of PMCA, it adopts a cis-closed conformation, which is also observed in the $3CLN^{Ca2+}{}_{IS\_P}$ system (Figure 3.12.b.III).

Myosin light chain kinase (MLCK) is primarily localized in the cytoplasm and is associated with structures such as stress fibers and cleavage furrow during cell division [73, 74] (Figure 3.12.a with IV). Its main function is to phosphorylate the regulatory light chain of myosin II for the activation of myosin, which facilitates muscle contraction and various cellular movements [73]. The binding of CaM to MLCK significantly increases the activity of the kinase and in the absence of CaM MLCK remains largely inactive, therefore affecting cellular signaling involving muscle contraction and other processes dependent on actomyosin dynamics [73]. When calcium ions bind to CaM, it undergoes a conformational change and interacts efficiently with MLCK [75]. This interaction typically occurs in the C-terminal domain of CaM and stabilizes binding, then the N-lobe of CaM clamps around MLCK and increases the activity of the kinase [75]. In the NMR structure of a peptide derived from MLCK bound to Ca^{2+}-loaded CaM (see Figure 3.12.b.IV, also see Figure 1.1.d) , CaM adopts a *trans*-semiclosed conformation, which is also sampled in the *holo* physiological system.

4. CONCLUSIONS AND FUTURE WORK

CaM adopts diverse conformations, including semi-open, closed, and fully open states. This conformational multiplicity, which has been extensively studied, enables it to regulate key signaling pathways by interacting with a broad range of binding partners. Environmental factors such as ionic strength and calcium availability strongly influence these interactions, emphasizing CaM's adaptability in calcium-dependent processes. Understanding its conformational landscape is essential for uncovering its functional versatility and advancing applications such as genetically encoded calcium indicators (GECIs).

Since the conformational adaptability of CaM directly influences its ability to interact with binding partners and modulate critical cellular processes, this study focuses on its conformational landscape under varying environmental conditions, employing classical MD and wt-MetaD simulations.

Key findings reveal that, while classical MD simulations provide valuable insights into CaM's conformational states, their exploration of the conformational landscape prove to be limited, often becoming confined to one or two minima. Consequently, they could not capture the full range of conformational states observed in experimental studies such as NMR and X-ray crystallography. This limitation is caused by the inability of MD simulations to overcome high-energy barriers within the accessible simulation timescales.

Conversely, wt-MetaD allow us not only to sample the positions of experimental structures in terms of the same DoFs but also to identify energy minima corresponding to conformations that have not been previously reported through experimental methods. This expanded sampling of CaM’s energy landscape provides a deeper and more comprehensive understanding of its dynamic behavior. Furthermore, minimization runs from the selected conformers obtained through wt-MetaD simulations exhibit minimal shifts from their initial configurations. These conformers remain trapped in their respective minima and do not escape to explore the broader conformational surface. This observation suggests the presence of well-defined local minima in the energy landscape. The consistency between the

minimization results and the sampled positions of experimental structures further highlights the suitability of the chosen CVs in accurately capturing CaM's conformational dynamics and the structural basis of its energy minima.

Our wt-MetaD results indicate that, under physiological ionic strength, CaM exhibits deeper energy minima. In contrast, reducing the ionic strength significantly lowers the energy barriers between conformational states, leading to the expectation that this would enable CaM to transition more freely across its conformational landscape. However, equilibration runs from selected conformers demonstrated that the reduced ionic concentration also promoted the formation of local minima with high energy barriers, trapping some conformers in energetically less favorable states. This is prevented under physiological conditions because the elevated ionic concentration helps break transient salt bridges, allowing conformers to explore the entire conformational landscape. In contrast, under low ionic strength, conformers become trapped in some local minima, which are avoided by wt-MetaD, preventing a comprehensive exploration of the conformational landscape.

Equilibration runs also revealed that a conformer adopts a cis-closed conformation, stabilized by a dynamic exchange between three salt bridges. This network staples the compact form, creating a high-energy barrier that restricts further sampling.

The applications and newly sampled conformations identified in this study hold potential for future research, offering a basis for exploring alternative methodologies, validating experimental findings, and designing innovative computational or experimental approaches.

Selection of a good CV is non-trivial, so the selection of alternative CVs such as the radius of gyration or root mean square deviation (RMSD), could provide projections of the conformational landscape onto other relevant degrees of freedom, allowing for a broader understanding of CaM's dynamics. Comparing the results of wt-MetaD simulations with different CVs to those presented in this study could validate and expand our understanding of the energy landscape.

Furthermore, the newly sampled *apo* and *holo* conformers identified in this work provide potential templates for designing improved GECIs while offering valuable insights into the biological function of CaM inside the cell. These conformers, which include semi-open,

closed, and fully open states in both *cis* and *trans* orientations, could be tailored for biosensors optimized for specific experimental conditions. They also enhance our understanding of how CaM adapts its structure to modulate interactions with diverse binding partners during cellular processes.

These conformers could also serve as precursors for other computational techniques. For instance, perturbation response scanning (PRS) [48] could be applied to these structures to identify reaction coordinates that drive transitions between conformational states. Additionally, other enhanced sampling methods such as umbrella sampling or infrequent metaD [38, 76] could offer a detailed exploration of transitions between energy minima, including the regions near the kinetic barriers and validate the free energy surfaces obtained from wt-MetaD simulations.

REFERENCES

[1] J. R. Allison, "Computational methods for exploring protein conformations," *Biochemical Society Transactions,* vol. 48, no. 4, 2020/08/28, doi: 10.1042/BST20200193.

[2] W. Potrzebowski, J. Trewhella, and I. Andre, "Bayesian inference of protein conformational ensembles from limited structural data," *PLoS Computational Biology,* vol. 14, no. 12, 2018 Dec 17, doi: 10.1371/journal.pcbi.1006641.

[3] A. R. Atilgan and C. Atilgan, "Computational strategies for protein conformational ensemble detection," *Current Opinion in Structural Biology,* vol. 72, 2022/02/01, doi: 10.1016/j.sbi.2021.08.007.

[4] C. Atilgan, "Computational Methods for Efficient Sampling of Protein Landscapes and Disclosing Allosteric Regions," *Computational Molecular Modelling in Structural Biology,* vol. 113, 2018/01/01, doi: 10.1016/bs.apcsb.2018.06.001.

[5] G. Bussi, A. Laio, G. Bussi, and A. Laio, "Using metadynamics to explore complex free-energy landscapes," *Nature Reviews Physics 2020 2:4,* vol. 2, no. 4, 2020-03-06, doi: 10.1038/s42254-020-0153-0.

[6] G. Wei, W. Xi, R. Nussinov, and B. Ma, "Protein ensembles: how does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell," *Chemical reviews,* vol. 116, no. 11, 2016 Jan 25, doi: 10.1021/acs.chemrev.5b00562.

[7] S. Lu, M. Ji, D. Ni, and J. Zhang, "Discovery of hidden allosteric sites as novel targets for allosteric drug design," *Drug Discovery Today,* vol. 23, no. 2, 2018/02/01, doi: 10.1016/j.drudis.2017.10.001.

[8] J. Gavalda-Garcia, D. Bickel, J. Roca-Martinez, D. Raimondi, G. Orlando, and W. Vranken, "Data-driven probabilistic definition of the low energy conformational states of protein residues," *NAR Genomics and Bioinformatics,* vol. 6, no. 3, 2024/07/02, doi: 10.1093/nargab/lqae082.

[9] G. R. Bowman *et al.*, "Discovery of multiple hidden allosteric sites by combining Markov state models and experiments," *Proceedings of the National Academy of Sciences,* vol. 112, no. 9, 2015-3-3, doi: 10.1073/pnas.1417811112.

[10] X. R. Liu, M. M. Zhang, D. L. Rempel, and M. L. Gross, "A Single Approach Reveals the Composite Conformational Changes, Order of Binding, and Affinities for Calcium Binding to Calmodulin," *Analytical Chemistry,* April 9, 2019, doi: 10.1021/acs.analchem.9b01062.

[11] M. Zhang *et al.*, "Structural Basis for Calmodulin as a Dynamic Calcium Sensor," *Structure,* vol. 20, no. 5, 2012/05/09, doi: 10.1016/j.str.2012.03.019.

[12] J. Akerboom *et al.*, "Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics," *Frontiers in Molecular Neuroscience,* vol. 6, 2013 Mar 4, doi: 10.3389/fnmol.2013.00002.

[13] J. Gsponer *et al.*, "A Coupled Equilibrium Shift Mechanism in Calmodulin-Mediated Signal Transduction," *Structure(London, England:1993),* vol. 16, no. 5, 2008 May 7, doi: 10.1016/j.str.2008.02.017.

[14] K. L. Kelly *et al.*, "Conformational Ensembles of Calmodulin Revealed by Nonperturbing Site-Specific Vibrational Probe Groups," *The Journal of Physical Chemistry A,* vol. 122, no. 11, February 5, 2018, doi: 10.1021/acs.jpca.8b00475.

[15] Y. S. Babu, C. E. Bugg, and W. J. Cook, "Structure of calmodulin refined at 2.2 Å resolution," *Journal of Molecular Biology,* vol. 204, no. 1, 1988/11/05, doi: 10.1016/0022-2836(88)90608-0.

[16] H. Kawasaki and R. H. Kretsinger, "Conformational landscape mapping the difference between N-lobes and C-lobes of calmodulin," *Journal of Inorganic Biochemistry,* vol. 177, 2017/12/01, doi: 10.1016/j.jinorgbio.2017.08.025.

[17] N. MR and C. WJ, "Structures of EF-hand Ca(2+)-binding proteins: diversity in the organization, packing and response to Ca2+ binding - PubMed," *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine,* vol. 11, no. 4, 1998 Dec, doi: 10.1023/a:1009253808876.

[18] "Novel aspects of calmodulin target recognition and activation," *European Journal of Biochemistry,* vol. 270, no. 3, 2003, doi: 10.1046/j.1432-1033.2003.03414.x.

[19] V. SW and L. E, "Novel aspects of calmodulin target recognition and activation - PubMed," *European journal of biochemistry,* vol. 270, no. 3, 2003 Feb, doi: 10.1046/j.1432-1033.2003.03414.x.

[20] H. Kuboniwa *et al.*, "Solution structure of calcium-free calmodulin," *Nature Structural Biology 1995 2:9,* vol. 2, no. 9, 1995/09, doi: 10.1038/nsb0995-768.

[21] A. Villalobo, H. Ishida, H. J. Vogel, and M. W. Berchtold, "Calmodulin as a protein linker and a regulator of adaptor/scaffold proteins," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* vol. 1865, no. 3, 2018/03/01, doi: 10.1016/j.bbamcr.2017.12.004.

[22] A. M. Westerlund and L. Delemotte, "Effect of Ca2+ on the promiscuous target-protein binding of calmodulin," *PLOS Computational Biology,* vol. 14, no. 4, 3 Nis 2018, doi: 10.1371/journal.pcbi.1006072.

[23] G. Fiorin, A. Pastore, P. Carloni, and M. Parrinello, "Using Metadynamics to Understand the Mechanism of Calmodulin/Target Recognition at Atomic Detail," *Biophysical Journal,* vol. 91, no. 8, 2006 Jul 28, doi: 10.1529/biophysj.106.086611.

[24] B. M. M. Grant, M. Enomoto, M. Ikura, and C. B. Marshall, "A Non-Canonical Calmodulin Target Motif Comprising a Polybasic Region and Lipidated Terminal Residue Regulates Localization," *International Journal of Molecular Sciences,* vol. 21, no. 8, 2020 Apr 15, doi: 10.3390/ijms21082751.

[25] S. MB and I. M, "Pre-formation of the semi-open conformation by the apo-calmodulin C-terminal domain and implications binding IQ-motifs - PubMed," *Nature structural biology,* vol. 3, no. 6, 1996 Jun, doi: 10.1038/nsb0696-501.

[26] J. L. Fallon and F. A. Quiocho, "A Closed Compact Structure of Native Ca2+-Calmodulin," *Structure,* vol. 11, no. 10, 2003/10/01, doi: 10.1016/j.str.2003.09.004.

[27] M. Lai, D. Brun, S. J. Edelstein, and N. L. Novère, "Modulation of Calmodulin Lobes by Different Targets: An Allosteric Model with Hemiconcerted Conformational Transitions," *PLOS Computational Biology,* vol. 11, no. 1, 22 Oca 2015, doi: 10.1371/journal.pcbi.1004063.

[28] J. Jiang *et al.*, "Site-specific modification of calmodulin Ca2+ affinity tunes the skeletal muscle ryanodine receptor activation profile," *Biochemical Journal,* vol. 432, no. 1, 2010/11/15, doi: 10.1042/BJ20100505.

[29] K. P. Hoeflich and M. Ikura, "Calmodulin in Action: Diversity in Target Recognition and Activation Mechanisms," *Cell,* vol. 108, no. 6, 2002/03/22, doi: 10.1016/S0092-8674(02)00682-7.
[30] H. KP and I. M, "Calmodulin in action: diversity in target recognition and activation mechanisms - PubMed," *Cell,* vol. 108, no. 6, 03/22/2002, doi: 10.1016/s0092-8674(02)00682-7.
[31] M. Berksoz and C. Atilgan, "Ranking Single Fluorescent Protein Based Calcium Biosensor Performance by Molecular Dynamics Simulations," *bioRxiv,* 2024-07-30, doi: 10.1101/2024.07.29.605619.
[32] J. Ding *et al.*, "Structural basis of the ultrasensitive calcium indicator GCaMP6," *Science China Life Sciences 2014 57:3,* vol. 57, no. 3, 2014-01-04, doi: 10.1007/s11427-013-4599-5.
[33] Y. Zhang *et al.*, "Fast and sensitive GCaMP calcium indicators for imaging neural populations," *Nature 2023 615:7954,* vol. 615, no. 7954, 2023-03-15, doi: 10.1038/s41586-023-05828-9.
[34] Y. Nasu *et al.*, "Structure- and mechanism-guided design of single fluorescent protein-based biosensors," *Nature Chemical Biology 2021 17:5,* vol. 17, no. 5, 2021-02-08, doi: 10.1038/s41589-020-00718-x.
[35] W. TJ *et al.*, "A neuron-based screening platform for optimizing genetically-encoded calcium indicators - PubMed," *PloS one,* vol. 8, no. 10, 10/14/2013, doi: 10.1371/journal.pone.0077728.
[36] A. Laio and M. Parrinello, "Escaping free-energy minima," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, no. 20, 2002 Sep 23, doi: 10.1073/pnas.202427399.
[37] F. Hooft, A. P. d. A. Ortíz, and B. Ensing, "Discovering Collective Variables of Molecular Transitions via Genetic Algorithms and Neural Networks," *Journal of Chemical Theory and Computation,* vol. 17, no. 4, March 4, 2021, doi: 10.1021/acs.jctc.0c00981.
[38] D. Ray and M. Parrinello, "Kinetics from Metadynamics: Principles, Applications, and Outlook," *Journal of Chemical Theory and Computation,* vol. 19, no. 17, August 16, 2023, doi: 10.1021/acs.jctc.3c00660.
[39] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chemical Physics Letters,* vol. 314, no. 1-2, 1999/11/26, doi: 10.1016/S0009-2614(99)01123-9.
[40] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *Journal of Computational Physics,* vol. 23, no. 2, 1977/02/01, doi: 10.1016/0021-9991(77)90121-8.
[41] J. Comer, J. C. Gumbart, J. Hénin, T. Lelièvre, A. Pohorille, and C. Chipot, "The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask," October 7, 2014, doi: 10.1021/jp506633.
[42] A. Laio, F. L. Gervasio, A. Laio, and F. L. Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Reports on Progress in Physics,* vol. 71, no. 12, 2008-11-26, doi: 10.1088/0034-4885/71/12/126601.
[43] A. Laio, M. Parrinello, A. Laio, and M. Parrinello, "Escaping free-energy minima," *Proceedings of the National Academy of Sciences,* vol. 99, no. 20, 2002-10-1, doi: 10.1073/pnas.202427399.

[44] J. McCarty and M. Parrinello, "A variational conformational dynamics approach to the selection of collective variables in metadynamics," *The Journal of Chemical Physics,* vol. 147, no. 20, 2017/11/28, doi: 10.1063/1.4998598.
[45] A. Barducci, G. Bussi, and M. Parrinello, "Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method," *Physical Review Letters,* vol. 100, no. 2, 2008-01-18, doi: 10.1103/PhysRevLett.100.020603.
[46] A. R. Atilgan, A. O. Aykut, and C. Atilgan, "Subtle pH differences trigger single residue motions for moderating conformations of calmodulin," *The Journal of Chemical Physics,* vol. 135, no. 15, 2011/10/21, doi: 10.1063/1.3651807.
[47] S. Negi, A. O. Aykut, A. R. Atilgan, and C. Atilgan, "Calmodulin Readily Switches Conformation upon Protonating High pKa Acidic Residues," *The Journal of Physical Chemistry B,* vol. 116, no. 24, June 5, 2012, doi: 10.1021/jp3032995.
[48] C. Atilgan and A. R. Atilgan, "Perturbation-Response Scanning Reveals Ligand Entry-Exit Mechanisms of Ferric Binding Protein," *PLOS Computational Biology,* vol. 5, no. 10, 23 Eki 2009, doi: 10.1371/journal.pcbi.1000544.
[49] C. Atilgan, Z. N. Gerek, S. B. Ozkan, and A. R. Atilgan, "Manipulation of Conformational Change in Proteins by Single-Residue Perturbations," *Biophysical Journal,* vol. 99, no. 3, 2010/08/04, doi: 10.1016/j.bpj.2010.05.020.
[50] A. O. Aykut, A. R. Atilgan, and C. Atilgan, "Designing Molecular Dynamics Simulations to Shift Populations of the Conformational States of Calmodulin," *PLOS Computational Biology,* vol. 9, no. 12, 5 Ara 2013, doi: 10.1371/journal.pcbi.1003366.
[51] F. Jalalypour, O. Sensoy, and C. Atilgan, "Perturb–Scan–Pull: A Novel Method Facilitating Conformational Transitions in Proteins," *Journal of Chemical Theory and Computation,* vol. 16, no. 6, April 23, 2020, doi: 10.1021/acs.jctc.9b01222.
[52] C. Ghosh and B. Jana, "Role of Calcium in Modulating the Conformational Landscape and Peptide Binding Induced Closing of Calmodulin," *The Journal of Physical Chemistry B,* vol. 125, no. 9, February 26, 2021, doi: 10.1021/acs.jpcb.1c00783.
[53] M. D. McCoy, I. John Hamre, D. K. Klimov, and M. S. Jafri, "Predicting Genetic Variation Severity Using Machine Learning to Interpret Molecular Simulations," *Biophysical Journal,* vol. 120, no. 2, 2020 Dec 15, doi: 10.1016/j.bpj.2020.12.002.
[54] J. C. Phillips *et al.*, "Scalable molecular dynamics on CPU and GPU architectures with NAMD," *The Journal of Chemical Physics,* vol. 153, no. 4, 2020/07/28, doi: 10.1063/5.0014475.
[55] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics," *J Mol Graph,* vol. 14, no. 1, pp. 33-8, 27-8, Feb 1996, doi: 10.1016/0263-7855(96)00018-5.
[56] A. Bakan, L. M. Meireles, and I. Bahar, "ProDy: protein dynamics inferred from theory and experiments," *Bioinformatics,* vol. 27, no. 11, pp. 1575-7, Jun 1 2011, doi: 10.1093/bioinformatics/btr168.
[57] V. Leone, F. Marinelli, P. Carloni, and M. Parrinello, "Targeting biomolecular flexibility with metadynamics," *Current Opinion in Structural Biology,* vol. 20, no. 2, 2010/04/01, doi: 10.1016/j.sbi.2010.01.011.
[58] J. Wu, "Understanding the Electric Double-Layer Structure, Capacitance, and Charging Dynamics," *Chemical Reviews,* vol. 122, no. 12, May 20, 2022, doi: 10.1021/acs.chemrev.2c00097.

[59] Š. L, T. E, and T. M, "Strong-coupling effective-field theory for asymmetrically charged plates with counterions only - PubMed," *Physical review. E,* vol. 110, no. 1-1, 2024 Jul, doi: 10.1103/PhysRevE.110.014609.
[60] J. E *et al.*, "Improvements to the APBS biomolecular solvation software suite - PubMed," *Protein science : a publication of the Protein Society,* vol. 27, no. 1, 2018 Jan, doi: 10.1002/pro.3280.
[61] D. E. Clapham, "Calcium signaling," *Cell,* vol. 80, no. 2, 1995/01/27, doi: 10.1016/0092-8674(95)90408-5.
[62] B. L and B. I, "Metallomics and the cell: some definitions and general comments - PubMed," *Metal ions in life sciences,* vol. 12, 2013, doi: 10.1007/978-94-007-5561-1_1.
[63] M. R. Duchen, "Mitochondria and calcium: from cell signalling to cell death," *The Journal of Physiology,* vol. 529, no. Pt 1, 2000 Nov 15, doi: 10.1111/j.1469-7793.2000.00057.x.
[64] D. NJ and T. AV, "Calcium gradients and the Golgi - PubMed," *Cell calcium,* vol. 40, no. 5-6, 2006 Nov-Dec, doi: 10.1016/j.ceca.2006.08.012.
[65] L. Xu and G. Meissner, "Mechanism of Calmodulin Inhibition of Cardiac Sarcoplasmic Reticulum Ca2+ Release Channel (Ryanodine Receptor)," *Biophysical Journal,* vol. 86, no. 2, 2004/02/01, doi: 10.1016/S0006-3495(04)74155-7.
[66] D. M. Balshaw, L. Xu, N. Yamaguchi, D. A. Pasek, and G. Meissner, "Calmodulin Binding and Inhibition of Cardiac Muscle Calcium Release Channel (Ryanodine Receptor) *," *Journal of Biological Chemistry,* vol. 276, no. 23, 2001/01/01, doi: 10.1074/jbc.M010771200.
[67] D. Gong *et al.*, "Modulation of cardiac ryanodine receptor 2 by calmodulin," *Nature 2019 572:7769,* vol. 572, no. 7769, 2019-07-05, doi: 10.1038/s41586-019-1377-y.
[68] C.-H. Lee and R. MacKinnon, "Activation mechanism of a human SK-calmodulin channel complex elucidated by cryo-EM structures," *Science (New York, N.Y.),* vol. 360, no. 6388, 2018 May 4, doi: 10.1126/science.aas9466.
[69] M. Gu *et al.*, "Small-conductance Ca2+-activated K+ channels: insights into their roles in cardiovascular disease," *Experimental & Molecular Medicine 2018 50:4,* vol. 50, no. 4, 2018-04-13, doi: 10.1038/s12276-018-0043-z.
[70] H. Tidow *et al.*, "A bimodular mechanism of calcium control in eukaryotes," *Nature 2012 491:7424,* vol. 491, no. 7424, 2012-10-21, doi: 10.1038/nature11539.
[71] M. Berrocal and A. M. Mata, "The Plasma Membrane Ca2+-ATPase, a Molecular Target for Tau-induced Cytosolic Calcium Dysregulation," *Neuroscience,* vol. 518, 2023/05/10, doi: 10.1016/j.neuroscience.2022.04.016.
[72] N. Juranic *et al.*, "Calmodulin Wraps around Its Binding Domain in the Plasma Membrane Ca2+ Pump Anchored by a Novel 18-1 Motif," *Journal of Biological Chemistry,* vol. 285, no. 6, 2010/02/05, doi: 10.1074/jbc.M109.060491.
[73] A. Poperechnaya, O. Varlamova, P.-j. Lin, J. T. Stull, and A. R. Bresnick, "Localization and Activity of Myosin Light Chain Kinase Isoforms during the Cell Cycle," *The Journal of Cell Biology,* vol. 151, no. 3, 2000 Oct 30, doi: 10.1083/jcb.151.3.697.
[74] K. E. Kamm and J. T. Stull, "Dedicated Myosin Light Chain Kinases with Diverse Cellular Functions *," *Journal of Biological Chemistry,* vol. 276, no. 7, 2001/02/16, doi: 10.1074/jbc.R000028200.

[75] J. Gsponer *et al.*, "A Coupled Equilibrium Shift Mechanism in Calmodulin-Mediated Signal Transduction," *Structure,* vol. 16, no. 5, 2008/05/07, doi: 10.1016/j.str.2008.02.017.
[76] P. Tiwary and M. Parrinello, "From Metadynamics to Dynamics," *Physical Review Letters,* vol. 111, no. 23, 2013-12-03, doi: 10.1103/PhysRevLett.111.230602.

APPENDIX

```
colvar {
       name dihedral1

       lowerBoundary -180.0
       upperBoundary 180.0
          width  10
          dihedral {
               # COM of N-Domain
               group1 { atomNumbersRange 1-952 }
               # CA atom of residue 69
               group2 { atomnumbers 955 }
               # CA atom of residue 91
               group3 { atomnumbers 1323 }
               # COM of C-Domain
               group4 { atomnumbersRange 1337-2184 }
  }
}

colvar {
  name distance1
  width 1.5
  lowerBoundary 10
  upperBoundary 50
  distance {
               # CA atom of residue 69
               group1 { atomnumbers 955 }
               # CA atom of residue 91
               group2 { atomnumbers 1323 }
  }
}

metadynamics {
   name     meta-dihedral
   colvars  dihedral1 distance1
   hillWeight              0.2
   newHillFrequency        500
   writeFreeEnergyFile           on
   writeHillsTrajectory    on
   hillwidth               1.0
   wellTempered            on
   biasTemperature         1490
}
```

Figure A.1 Example collective variables script