2019

Multi-Scale Computational Studies of Calcium (Ca^{2+}) Signaling

Bin Sun
University of Kentucky, sunbinxod@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0003-2583-4493
Digital Object Identifier: https://doi.org/10.13023/etd.2020.028

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Bin Sun, Student

Dr. Peter M. Kekenes-Huskey, Major Professor

Dr. Yinan Wei, Director of Graduate Studies

Multi-Scale Computational Studies of Calcium (Ca^{2+}) Signaling

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy
in the College of Arts and Sciences
at the University of Kentucky

By
Bin Sun
Lexington, Kentucky

Director: Dr. Peter M. Kekenes-Huskey, Professor of Chemistry
Lexington, Kentucky

https://orcid.org/0000-0003-2583-4493

ABSTRACT OF DISSERTATION

Multi-Scale Computational Studies of Calcium (Ca^{2+}) Signaling

Ca^{2+} is an important messenger that affects almost all cellular processes. Ca^{2+} signaling involves events that happen at various time-scales such as Ca^{2+} diffusion, trans-membrane Ca^{2+}transport and Ca^{2+}-mediated protein-protein interactions. In this work, we utilized multi-scale computational methods to quantitatively characterize Ca^{2+} diffusion efficiency, Ca^{2+} binding thermodynamics and molecular bases of Ca^{2+}-dependent protein-protein interaction. Specifically, we studied 1) the electrokinetic transport of Ca^{2+} in confined sub-μm geometry with complicated surfacial properties. We characterized the effective diffusion constant of Ca^{2+} in a cell-like environment, which helps to understand the spacial distribution of cytoplasmic Ca^{2+}. 2) the association kinetics and activation mechanism of the protein phosphatase calcineurin (CaN) by its activator calmodulin (CaM) in the presence of Ca^{2+}. We found that the association between CaM and CaN peptide is diffusion-limited and the rate could be tuned by charge density/distribution of CaN peptite. Moreover, we proposed an updated CaM/CaN interaction model in which a secondary interaction between CaN's distal helix motif and CaM was highlighted. 3) the roles of Mg^{2+} and K^{+} in the active transport of Ca^{2+} by sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) pump. We found that Mg^{2+} most likely act as inhibitor while K^{+} as agonist in SERCA's transport process of Ca^{2+}. Results reported in this work shed insights into various aspects of Ca^{2+} signaling from molecular to cellular level.

KEYWORDS: Computational Simulation, Calcium Signaling, Diffusion, Protein Conformation, Association Rates, Free energies

Author's signature: Bin Sun

Date: December 20, 2019

Multi-Scale Computational Studies of Calcium (Ca^{2+}) Signaling

By
Bin Sun

Director of Dissertation: Dr. Peter M. Kekenes-Huskey

Director of Graduate Studies: Dr. Yinan Wei

Date: December 20, 2019

Dedicated to the people I love.

ACKNOWLEDGMENTS

I would like to first thank my adviser Dr. Kekenes-Huskey. Under his guidance, I learned not only the experimental skills but also the mindset of viewing scientific problem from multiple angles. I have also been inspired by his love and persistence in science.

Besides, I would like to thank my committee members, Dr. Richards, Dr. Yang and Dr. Zhan for their insightful comments and encouragement.

Next, I would like to thank my labmates for the helpful discussions we have and funny moments we shared in the lab.

Lastly, I would thank my family and my friends for their countless support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Ca^{2+} Signaling is Important.

Ca^{2+} is an ubiquitous intracellular secondary messenger that impacts almost all cellular processes.[1] To name a few, in the heart, the excitation-contraction coupling of heart muscle is tightly related to cytoplasm Ca^{2+} concentration oscillating.[2] In the brain, the Ca^{2+} influx to neurons through different Ca^{2+}-permeable channels can differentially regulate gene expression.[3] In the lymphocytes, the immune response is triggered by Ca^{2+} flux into the cytoplasm from the extracellular space.[4] Severe diseases such as cardiac arrhythmias and Alzheimer's disease (AD) are heavily associated with Ca^{2+} homeostasis disruption and Ca^{2+} signaling dysregulation.[5,6] Thus it is of great importance to understand Ca^{2+} signaling mechanisms.

1.2 Ca^{2+} Signaling is Complicated

The task of deciphering Ca^{2+} signaling is challenging because of the complexity of Ca^{2+} signaling network. The complexity is mainly manifested in the following two aspects:

Multiple paths of Ca^{2+} uptake and release. The spatial and temporal distribution of Ca^{2+} is complicated. Specifically, Ca^{2+} ions are heterogeneously distributed in cell compartments and its concentration changes are precisely and timely controlled through various exchangers/pumps/channels (Fig. 1.1(a)). In the resting cell, the Ca^{2+} concentration in cytoplasm is $\sim$100 nM.[7] This value is $\sim$20,000 fold lower than the extracellular concentration.[8] Upon activation, the membrane-embedded voltage-gated Ca^{2+} channels allow extracellular Ca^{2+} flee into cytoplasm rapidly at the speed of $\sim$1 million Ca^{2+} ions per second per channel.[1] Subsequently, cytoplasmic Ca^{2+}s are sequestered back into sarco/endoplasmic (SR/ER) at the rate about 10 Ca^{2+} ions per second by sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) pump.[9] There also exists other Ca^{2+} channels and exchangers including $Na+/Ca^{2+}$ exchanger (NCX)[10] and ryanodine receptors (RyRs).[11] These exquisite Ca^{2+} uptake/release paths maintain a low and heterogeneous intercellular Ca^{2+} concentration, yet allowing it to change responsively upon stimulation. This exquisiteness, however, challenges the modeling of cytosolic Ca^{2+} concentration as Ca^{2+} fluxes through all these paths are strongly coupled in space and time.[12,13]

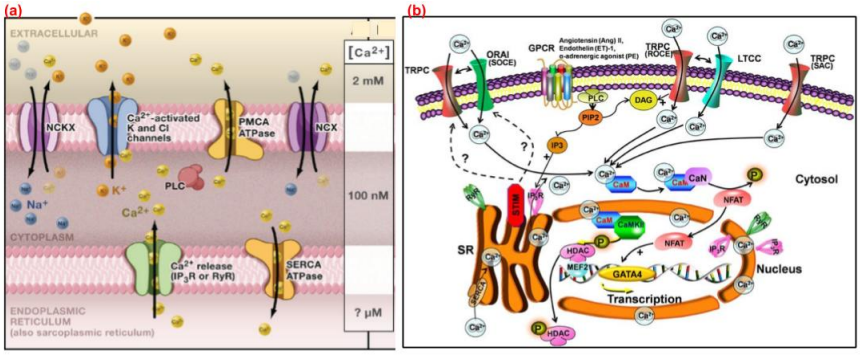


Figure 1.1: (a) Ca^{2+} concentrations in cellular compartments and the main pumps/exchangers that maintain such concentrations. Image is adapted from ref [1]. (b) Detailed Ca^{2+}-triggered protein-protein interaction (PPI)s in cardiac cell. Image is adapted from ref[14].

Ca^{2+} signaling engages countless protein-protein interactions. These inter-actions are initiated by various Ca^{2+} binding proteins (CBP)s. CBPs serve as interpreters that translate signaling information encoded in Ca^{2+} concentrations to altered cell functions. CBPs are capable of binding Ca^{2+} with a wide range of affinities, ranging from nM to mM.[1] Although many bear the commonly shared EF-hand (helix-loop-helix) Ca^{2+} binding motif,[15] CBPs are generally divided into two categories, depending on if they undergo significant conformational changes upon Ca^{2+} binding.[8] The first group are ones that do not undergo significant changes after binding Ca^{2+}, resulting in no protein-protein interaction (PPI)s triggered. The typical examples are parvalbumin (PV) and calbindin, which are also named as "pure" Ca^{2+} buffers.[16] The second group will experience large conformational changes to bind target peptide/proteins after Ca^{2+} binding, including the S100 family of proteins[17] and calmodulin (CaM) that regulates $\sim$ 300 targets.[18] The PPI cascade initiated by this group is complicated. For example, in the cardiac cell (Fig. 1.1(b)), Ca^{2+}-saturated CaM triggers the dephory-sphoration of nuclear factor of activated T-cells (NFAT) via calcineurin (CaN). Following that, NFAT was translocated into nuclueus and exerts regulatory effects on various genes transcription in T-cells.[14,19] Additionally, the CaM-dependent kinase (CaMK) family, which has been reported to be involved in a wide range of cancer-related process[20] and cardiac remodeling,[21] is also activated by Ca^{2+} signaling. The large number of PPIs involved in Ca^{2+} signaling implies a thorough description of Ca^{2+} signaling mechanism is non-trivial.

1.3 Understanding Three Prototypical Aspects of Ca^{2+} Signaling

Toward a quantitative description of Ca^{2+} signaling, we focused on three prototypical phenomena: cellular Ca^{2+} diffusion, Ca^{2+}-mediated protein-protein

interaction (PPI) and Ca^{2+} active transport through the sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) pump (Fig. 1.2).

Celluar Ca^{2+} diffusion. Ca^{2+} signaling efficiency is in part determined by Ca^{2+} diffusion rate. This is because the signaling transduction requires Ca^{2+} ion to diffuse from where it is stored (SR/ER) to where it is utilized by proteins controlling Ca^{2+}-dependent signaling processes. In cellular environment, the compartmentalized Ca^{2+} concentration signifes the importance of Ca^{2+} diffusion in nanodomains.[22] Under such scale, Ca^{2+} diffusion is influenced by intracellular crowders such as proteins, nucleic acids and membranes that can have non-neutral surfaces.[23,24] Ca^{2+} ion diffusion is also coupled with the co-/counter-transport of other cellular ions (mostly K^+, Na^+, Mg^{2+} and Cl^-). In addition, buffer proteins can significantly change Ca^{2+} diffusion efficiency as they selectively bind Ca^{2+}.[16] To understand how Ca^{2+} diffusion is influenced by the interplay of these factors, we simulated Ca^{2+} diffusion in mesoporous silica channels that have comparable scales as cellular nanodomains.[25] Further the controllable surface chemical properties of silica material makes it a good system to study cellular Ca^{2+} diffusion. Specifically, the surfaces of silica materials are covered with functional silanol (Si-OH) groups.[26] Si-OH has been experimentally shown to have the protonated state (Si-OH_2^+) and deprotonated state (Si-O^-)[26] and the latter has the ability to bind Ca^{2+}.[27] By tuning Si-OH group density and equilibrium constants of the (de)protonation/Ca^{2+} binding reactions, precise surface charge densities and degree of Ca^{2+} adsorption could be achieved. The effects of crowders charge and Ca^{2+} adsorption on Ca^{2+} diffusion could be rigorously investigated in this model, which could deepen our understanding of Ca^{2+} diffusion in cellular environment.

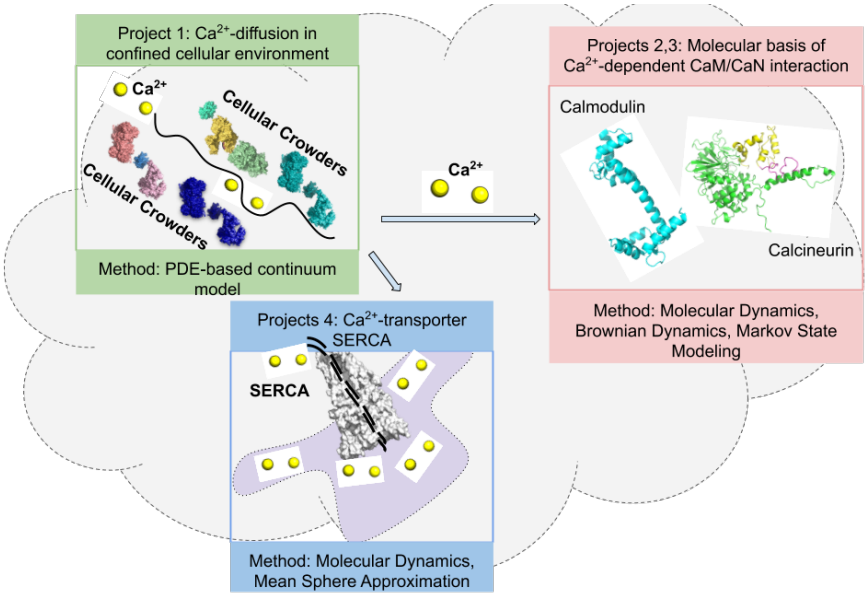


Figure 1.2: Overall view of specific projects in the thesis. In project 1 (Chapter 2), the diffusion of Ca^{2+} through charged crowders in confined cellular geometry was studied via partial differential equation (PDE)-based continuum model. In projects 2 and 3 (Chapter 3 and 4), the molecular basis of Ca^{2+} dependent Calcineurin (CaN)/Calmodulin (CaM) interaction was studied via molecular dynamics (MD), Brownian dynamic (BD) simulations and Markov sate model (MSM). Specifically, the binding rate constants as well as the interaction model between calcineurin (CaN) and CaM was studied. In project 4 (Chapter 5), the binding affinities of Ca^{2+}, Mg^{2+} and K^{+} in sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) were studied via molecular dynamics (MD) and mean sphere approximation (MSA). The roles of Mg^{2+} and K^{+} in the SERCA's transport cycle of Ca^{2+} were also investigated.

Ca^{2+}-mediated PPI interaction. Additionally, we studied the activation mechanism of the protein phosphatase calcineurin (CaN) by its activator calmodulin (CaM) in the presence of Ca^{2+}. The significance of this PPI are two folds: 1) CaM and CaN are key nodes in Ca^{2+} signaling network that control neural and cardiac developments and can be detrimentally activated under pathological stimuli. CaM is a highly conserved Ca^{2+} sensor across all mammalian species[28] that binds as many as ∼300 targets.[18] Meanwhile, CaN is a phosphatase that plays an important role in various processes such as T-cell activation and cardiac development.[29,30] 2) The CaN/CaM pair provides us an appropriate model to study intrinsically disordered peptide (IDP)-involved PPIs, which is a growing interest due to the prevalence of IDP-mediated signaling transduction.[31] The regulatory domain of CaN, at which CaM binds, is intrinsically disordered.[32–35] IDPs are proteins that lack of well-defined secondary structures at physiological conditions. Many IDPs gain structural stability after binding to their globular protein partners.[31,36–38]

We specifically evaluated the binding rate constant of CaN's intrinsically disordered regulatory domain to CaM. The binding rate plays an important role in cell function as many biochemical reactions must occur rapidly sometimes in a diffusion-limited regime.[39] While progress has been made on predicting the association rates of structured proteins,[40] characterizing the association process of IDP is challenging because of the structural flexibility.[41] Part of the complexity is the dual role of electrostatics exerted in the association process. On one hand, intramolecular electrostatic interactions govern IDPs' conformational ensemble compactness,[42,43] which correlates with the accessibility of binding site to protein partners. On the other hand, intermolecular long-range electrostatic interactions drive protein-protein association.[44] We explored this dual role of electrostatics in the CaN/CaM PPI model. The role on conformational properties was investigated by performing extensive molecular dynamics (MD) on CaN regulatory domain constructs bearing different charge densities. The role on driving association was investigated via Brownian dynamics (BD) that simulate the diffusional encounter of representative CaN structures to CaM. At last, the two effects were taken together via a stochastic model reported in ref[45] to give the effective association rate constant of CaN regulatory domain to CaM. We additionally updated the interaction model of CaN activation by CaM, inspired by experimental evidence showing that a secondary interaction between CaN's regulatory domain and CaM is needed to fully activate CaN.[46] By using protein-protein docking and MD simulations, we identified on CaM' surface a potential site for this interaction, which was affirmed by experimental assays. This molecular level model extended our understanding of CaN activation by CaM. More importantly, this model might apply to other CaM-regulated systems, given the large number of CaM-regulated targets and the similar binding patterns they share when binding CaM.

Ca^{2+} active transport through sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) pump. SERCA plays an important physiological role in maintaining Ca^{2+} homeostasis in excitable cells, as it pumps cytoplasmic Ca^{2+} into the sarcoplasmic/endoplasmic reticulum (SR/ER) with energy provided by adenosine triphosphate (ATP) hydrolysis.[47] SERCA has been widely studied for its role in returning intracellular Ca^{2+} to basal levels following stimuli that elevates Ca^{2+} content.[48] SERCA selectively binds Ca^{2+} over $Mg^{2+}$and K^+, despite the latters being in excess. We reported in this sub-project the molecular basis of SERCA's selectivity on Ca^{2+}, with the emphasis on Ca^{2+}'s favorable binding thermodynamics over the competing Mg^{2+} and K^+ cations. We performed extensive all-atomic molecular dynamics simulations to probe molecular determinants of Ca^{2+} binding in the pump. In addition, we related the binding thermodynamics to a state model to assess to what extent Mg^{2+} and K^+ accelerates or hinder SERCA's transport cycle. This study elucidates the impacts of Mg^{2+} and K^+ on SERCA's activity, which provides theoretical bases for developing Ca^{2+} handling models containing Ca^{2+}-uptake through SERCA.

1.4 Methods Used in Present Work

Multi-Scale Computational Methods Serve as an Unique Toolkit to Explore Ca^{2+} Signaling

Due to the significance as well as the complexity of Ca^{2+} signaling, abundant experimental studies have been conducted to explore the signaling process. These experiments generated valuable knowledge such as Ca^{2+}-caused tissue damage,[49,50] Ca^{2+} handling in cytoplasm,[51] identification of Ca^{2+}-binding motif shared in CBPs[15] and discovery of Ca^{2+}-mediated protein-protein interactions.[52,53] However, molecular level descriptions of Ca^{2+} signaling are still poorly understand. These molecular descriptions include determinants of Ca^{2+} binding thermo-dynamics, interaction model of Ca^{2+}-mediated PPI and molecular events orchestrated for pump-facilitated Ca^{2+} transport and so on. Computational methods, in this regard, provide complementary tools to experimental studies to achieve such molecular level descriptions. For example, molecular dynamic (MD) simulations have helped to reveal CaM's conformatonal and dynamic properties[54–58] as well as the coupling between the motions of SERCA's cytosolic domain to Ca^{2+} entering/leaving to the binding sites.[59,60] Despite these achievements via molecular dynamics, multi-scale computational methods are still needed as Ca^{2+} signaling consists of events that happen at time-scales beyond MD's capable scope. For example, in typical experiments, the time-scale of Ca^{2+} reaching equilibria in nanodomain is $\sim$100 μs per μm.[22] Large motions of the cytosolic domain of SERCA were observed in 1 μs time-scale, however, the exchange of Ca^{2+} with non-congtive cation in SERCA binding sites occur at sub-millisecond time scales.[59] Unfortunately, all-atomic molecular simulations can only reveal events of proteins at most for sub-millisecond time-scale with specifically-dedicated hardware/software combination.[61,62] Therefore, multi-scale computational methods besides MD are needed to gain insights into processes that happen at longer time-scale than MD.

Partial differential equation (PDE)-Based Continuum Model to Explore Ca^{2+} Diffusion in Cellular Nanodomains

In cells, Ca^{2+} ions are heterogeneously distributed in nano-scale compartments,[22] in which the continuum model is a perfect tool to study Ca^{2+} diffusion. Partial differential equation (PDE) based continuum model coupled with finite element method has been widely used to study the signaling transmission in synapse[63,64] as well as ion diffusion in lithium-batteries.[65] One such widely used continuum model is the Poisson-Nernst-Planck (PNP) theory which describes the electrokinetic phenomena in nano domains. The Poisson-Nernst-Planck (PNP) equation is a combination of two equations: the Nernst-Planck (NP) equation that describes the ion diffusion flux caused by concentration gradient and electrical force; the Poisson equation that updates the electric field due to

the redistribution of ion species. These two equations are given as:

$$-\nabla \cdot J_i = 0, \tag{1.1}$$

$$J_i = -D_i \left(\nabla c_i + \frac{z_i F c_i}{RT} \nabla \phi \right), \tag{1.2}$$

$$-\epsilon_r \epsilon_0 \nabla^2 \phi = F \sum_{1=1}^{N} z_i c_i, \tag{1.3}$$

where J_i, D_i, c_i, z_i are flux density, diffusion coefficient, molar concentration, and valence electron number of i^{th} ionic species. F is the Faraday constant, ϕ is electric potential, T is absolute temperature, and R is gas constant. ϵ_0 and ϵ_r are vacuum and relative permittivity of the electrolyte solution, respectively. Another important factor that influences Ca^{2+} diffusion is the Ca^{2+} binding or adsorption to proteins.[66] We used the silanol group (SiOH) on silica materials to mimic Ca^{2+} adsorption by proteins. Specifically, we consider the following equation which contains the deprotonation of SiOH and binding of Ca^{2+} to SiO^- reactions:

$$\mathrm{Si{-}OH} + \mathrm{Ca^{2+}} \rightleftharpoons \mathrm{Si{-}OCa^{+}} + \mathrm{H^{+}}, \tag{1.4}$$

The consequence of this reaction is to alter the surface charge density (σ_s), which is ultimately reflected in the Neumann condition that is applied when solving the PNP equation:

$$-\nabla \phi \cdot \boldsymbol{n} = \sigma_s / (\epsilon_o \epsilon_r), \tag{1.5}$$

where $\boldsymbol{n}$ is the unit outer normal vector. This simplified Ca^{2+} adsorption model coupled with PNP equation enables us to understand how Ca^{2+}'s diffusion is shaped by charged crowders in the nano-scale cellular compartments.

Molecular Dynamics Simulations to Study Conformational Dynamics and Thermodynamics of Ca^{2+}-Binding Proteins

The essential nature of MD simulation is to describe the time-dependent movement of atoms in molecules according to Newton's second law.[67] MD serves as a unique tool in exploring microscopic interactions that can be elusive to experiments. More importantly, by using statistical mechanics, macroscopic thermodynamic properties can be accurately calculated from corresponding microscopic components given by MD.[68] Therefore, MD has been widely used to study folding, binding and conformation dynamics of biological molecules such as proteins, nucleic acids and lipids.

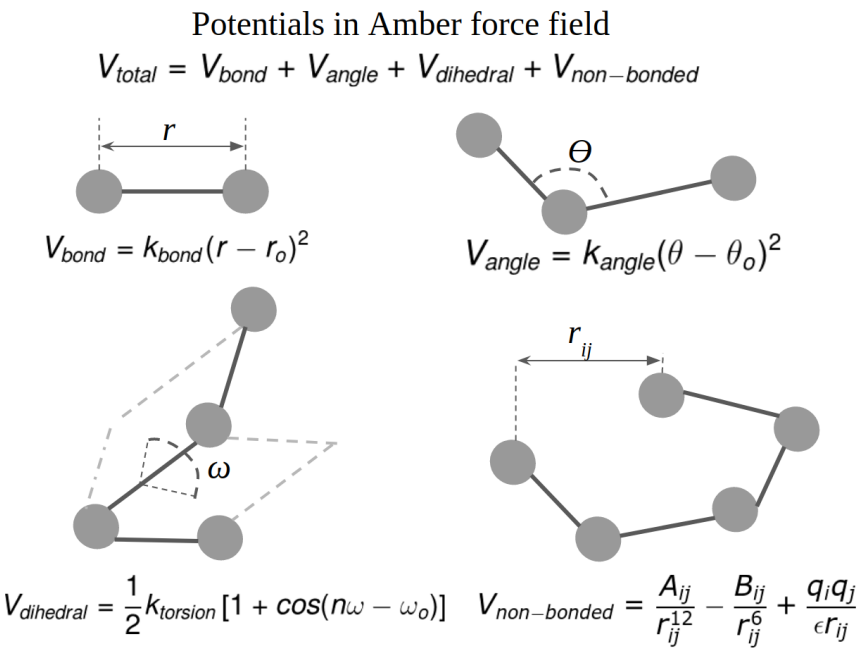


Figure 1.3: Energy function forms of Amber force field potential.[69]

MD consists of solving the equation of motion:

$$m_i \ddot{r}_i = F_i \tag{1.6}$$

$$F_i = \frac{\partial}{\partial r_i} U \tag{1.7}$$

The force F_i acting on atom i is derived from the potential of the system $U = u(r_1, r_2, ...r_N)$ where $N = 3i$ is the total coordinates of atoms. The accuracy of MD is determined by the potential form of $u(r_1, r_2, ...r_N)$ which reflects our understanding of microscopic interactions in the system. In modern empirical MD implication, the energy potential of U consists of four core terms: bond potential (V_{bond}), angle potential (V_{angle}), dihedral potential ($V_{dihedral}$) and non-bonded interaction potential ($V_{non-bond}$) which includes the electrostatic potential ($V_{electro}$) and van der Waals potential (V_{vdW}) (see Fig. 1.3 for the potential terms in the commonly used Amber force field[69]).

Of the four energy terms, the bond and angle terms describe the potential arise from the deviation of bonds and angles from their equilibrium states and have the common expression simplified as an harmonic potential:

$$V_{bond/angle} = k(a_i - a_o)^2 \tag{1.8}$$

where k is called force constant that reflects the 'rigidity' of the bond/angle, r_a and a_o are the instant and equilibrium bond-length/angle-value, respectively. The force constant of bond and angle in modern force field are at the orders of (sub)hundred kcal/mol which are larger than thermo fluctuation $k_b T$,[70]

suggesting that the degree of freedom associated with these two terms are largely frozen during simulation. Unlike the bond and angle potentials that have the quadratic harmonic potential form, the dihedral potential has the form as:

$$V_{dihedral} = \frac{1}{2}k[1 + cos(n\omega_i - \omega_o)] \tag{1.9}$$

where k is the force constant, n is multiplicity and ω_i/ω_o are instant/equilibrium dihedral values. Due to the smaller force constant k, the dihedral potential has the potential comparable to thermo fluctuation (k_bT) and thus contributes most to the conformational thermodynamics of biological molecules. Besides the bond, angle and dihedral potentials that belong to the category of bonded interactions, the non-bonded potential ($V_{non-bonded}$) including electrostatic part and vdW is also considered:

$$V_{non-bonded} = V_{electro} + V_{vdW} \tag{1.10}$$

$$V_{electro} = \frac{q_i q_j}{\epsilon r_{ij}} \tag{1.11}$$

$$V_{vdW} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \tag{1.12}$$

The charge-charge electrostatic interaction between a pair of atoms is described by coulombic potential and the neutral interaction is described via the Lennard-Jones potential (12-6 potential). r_{ij} is the separation between atom i and j. q_i, q_j, A_{ij} and B_{ij} are fitted point charges and vdW constants of atom i and j. $V_{non-bonded}$ plays an vital role in biomolecules' thermodynamic as well as structural properties.[71] Besides the four terms, the solvation effect (V_{solv}) can be optionally added to the system via either incorporating explicit water molecules characterized by water models such as TIP3P[72] or an implicit continuum solvation field described by the generalized Born (GB) approximation.[73]

The potential forms stated above, together with the corresponding parameters, are called *force field*.[71] The applicability and quality of the force field are largely determined by the parameter optimization process. Structural properties of model compounds such as amino acid analogy from the *ab initio* Quantum Mechanics (QM) calculations are preferred target data for parameter optimization because the least amount of guesswork is needed due to the rigorous theory of QM.[67,71] Bases on QM calculations, the equilibrium bond and angle values, as well as the corresponding force constants of bonded potentials can be readily achieved. However, relying solely on QM-calculated properties for parameterization in certain cases leads to inconsistency with experimental data measured in condensed phase, as QM calculations are mostly conducted in non-condensed phase.[74] In such cases where experimental data is available, taking these data into account for parameterization can improve the accuracy of the force field.

One featured application of MD is to study the accompanied conformational changes of proteins when exerting their functions.[75,76] With the advances of computational hardware[77] and non-equilibrium algorithm such as accelerated MD[78] and native-structure based bias potential,[79] large scale of molecular dynamics (MD) sampling of proteins become feasible, with timesacle approaching sub-millisecond.[61,62] The large amount of MD data needs to be physically interpreted to compare with experimental measurements.[79] One such interpretation is to extract the 'slow' transition between metastable states as these transitions are associated with biologically important events.[80] A Markov-based conformat-ional space discretization method has been applied to analyze MD sampling of various proteins..[81–83] These successful applications with robust analyzing methods have demonstrated the substantial potentials of MD in exploring biomolecule functions.

Brownian Dynamics Simulation to Characterize Ca^{2+}-mediated Protein-Protein Association Kinetics

Protein-protein interaction is an indispensable part in Ca^{2+} signaling transduction. Besides the conventional *affinity* concept that reflects the favorable thermodynamic gains upon binding, the process of binding also plays an important role in characterizing the signaling efficiency.[84] It is thus of great importance to understand the rate constants of protein-protein association. Brownian dynamics (BD) was developed to simulate the diffusional encounter of two large biological molecules from which the corresponding association kinetics can be estimated.[85] The underlying equation governing the brownian motion is given as:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + D\frac{\mathbf{F}(t)}{k_b T} + \mathbf{R}(t) \tag{1.13}$$

where $\mathbf{r}$ is position of a molecule, D is translational diffusion coefficient, $\mathbf{F}$ is force acting on the molecule and $\mathbf{R}$ is a random displacement at time t. In the special case where $\mathbf{F} = 0$, Eq. 1.13 is reduced to pure random brownian motion. The acting force $\mathbf{F}$ consists of two parts: electrostatic interactions (long range driving force) and van der Waal's (short range repulsion force) interactions. In implementation, the electrostatic part is first solved via APBS[86] to evaluate the electrostatic potential of binding partners. The APBS tool numerically solves the linearized Poisson-Boltzmann equation under a given ionic strength (usually physiological 0.15 M KCl):

$$-\epsilon\nabla^2\psi = \sum \rho_i q_i - \kappa^2\psi \tag{1.14}$$

where ψ is the electrostatic potential, $\rho_i q_i$ is the charge distribution of fixed charge i, and κ is the inverse of Debye length. The Debye length reflects the scale over which mobile charges could screen out electric potential fields. The vdW uses the standard 12-6 Lennard–Jones as that in Eq. 1.10 or a modified

softer 8-6 potential:[85]

$$V(r) = 4\epsilon[\frac{3}{4}(\frac{r_m}{r})^8 - (\frac{r_m}{r})^6] \tag{1.15}$$

where r is the separation of two atoms in a pair, r_m is the sum of vdW radii of two atoms, and ϵ is the well depth. Both the radii, well depth and point charge parameters can be adapted from commonly used force fields such as AMBER99.[87]

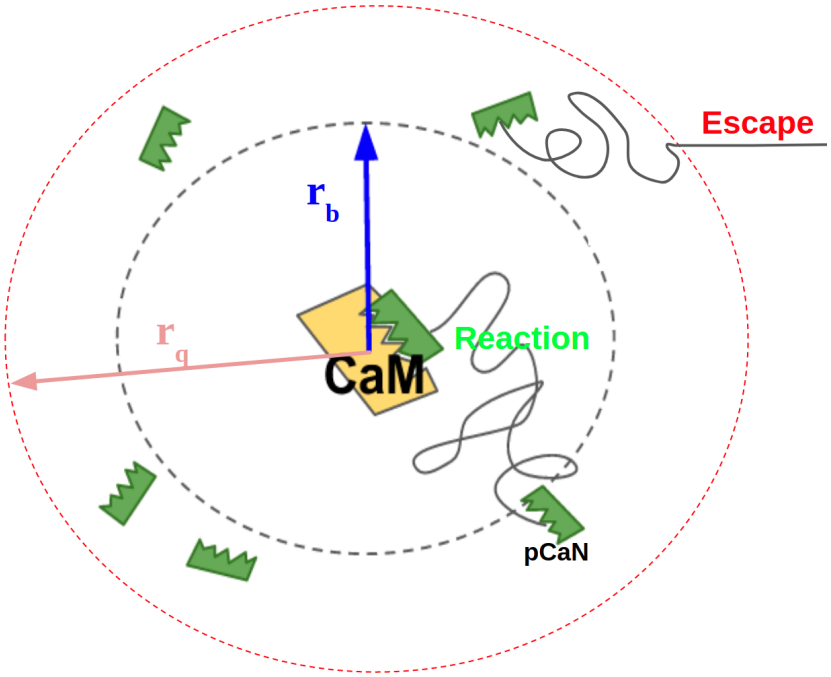


Figure 1.4: Illustration of escape and reaction events in Brownian dynamic (BD) simulations. The larger Calmodulin (CaM) is treated as fixed receptor and the smaller pCaN (a 24-residue peptide from the CaM binding region (CaMBR) of Calcineurin (CaN)) is movable ligand. Enormous copies of pCaN are placed at the b sphere. The association kinetics between pCaN and CaM is estimated from the flux ratio of reaction to escape.

The association rate is estimated as the flux ratio of reaction to escape (Fig. 1.4).[85] Specifically, the ligand (the moving molecule) whose motion is governed by Eq. 1.13 is initially placed far away from the receptor (the fixed molecule) with the distance being r_b such that the potential between these two is radially symmetric. At this distance, the ligand can diffuse either further away from or closer to the receptor, depending on the **F**. For the former, if the ligand-receptor distance exceeds a pre-defined value r_q, the trail was counted as escape. For the latter, there is certain chance that the ligand can proceed to an appropriate distance to form an encounter-complex with receptor and successfully trigger the reaction. It should be noted that an important assumption is made in Brownian dynamic (BD): proteins are treated as rigid bodies. This assumption ignores the dynamic nature of proteins and assumes that the binding site is always accessible. One possible direction to take the conformational flexibility into account is to sample the gating of binding site via MD *a prior*. Based on MD sampled opening/closing events of binding site, a Markov state model can be constructed to achieve the gating kinetics. Lastly, as reported in,[88] the gating kinetics can be incorporated into BD-generated rigid body diffusional encounter rate to give association rate constant with

conformational flexibility taken into account.

Mean Spherical Approximation (MSA) to Estimate the Thermodynamic Properties of Ca^{2+} in Binding Sites

The Ca^{2+} signaling pathway consists of Ca^{2+} binding proteins (CBP)s with various Ca^{2+}-binding affinities.[1] CBPs are capable of selectively binding Ca^{2+} even in the excess of Mg^{2+}, and this selectivity originates from the unique coordination oxygens of CBPs.[89,90] This is afforded through their EF-hand (helix-loop-helix) Ca^{2+} binding motif that highly conserved among more than 800 CBPs.[90] Ca^{2+} ions are coordinated by oxygen atoms from the loop region residues of these EF-hand CBPs. Understanding how the arrangement of coordinating oxygen atoms in CBPs contributes to Ca^{2+} selective binding is of fundamental importance to deciphering Ca^{2+} homeostasis. The mean sphere approximation (MSA) is a proper model for estimating Ca^{2+} binding thermodynamics in CBPs, given the Ca^{2+} binding sites are usually well defined and conservative.

The mean sphere approximation (MSA) is a mean-field model which estimates the chemical potentials of cation in electrolyte solution with finite-sized ions. Specifically, it calculates the partition of ions between bulk electrolyte solution and filters (a finitely sized volume) via minimizing the chemical potential of each ion species.[91] In the Nonner MSA implementation for describing cation affinity in Ca^{2+} channels,[92] for each ion species, an excess chemical potential (μ_i^{ex}) which accounts for electrostatic and hard-sphere interactions between ions inclusive of oxygens, was calculated:

$$\mu_i^{ex} = \mu_i^{ES} + \mu_i^{HS} \tag{1.16}$$

The μ_i^{ES} is electrostatic potential due to interactions with other ion species and coordination oxygens. Since in MSA model, each ion species including oxygens comprising the filter are treated as hard sphere with specified radii, a hard sphere potential μ_i^{HS}, was also calculated to account for repulsion among finite sized ions. This model signifies the favorable environment for cation binding in proteins such as CBPs and calcium channels[92–94] in which the calcium coordinating oxygens are confined in the binding site. In such environment cations can easily have negative (favorable) chemical potentials. The specific expressions of μ_i^{ES} and μ_i^{HS} are given as:[93]

$$\begin{aligned}
\mu_i^{ES} &= -\frac{\lambda_b}{\beta}\left[\frac{\Gamma z_i^2}{1+\Gamma\sigma_i} + \eta\sigma_i\left(\frac{2z_i - \eta\sigma_i^2}{1+\Gamma\sigma_i} + \frac{\eta\sigma_i^2}{3}\right)\right] \\
\mu_i^{HS} &= k_bT\left[\frac{3\xi_2\sigma_i + 3\xi\sigma_i^2}{\Delta} + \frac{9\xi_2^2\sigma_i^2}{2\Delta^2} + \xi_0\sigma_i^3\left(1 + \Delta\left(\frac{\xi_3}{\Delta} + \frac{3\xi_1\xi_2}{\xi_0\Delta^2} + \frac{3\xi_2^3}{\xi_0\Delta^3}\right)\right) - ln\Delta\right]
\end{aligned} \tag{1.17}$$

where z_i, σ_i and $\beta = \frac{1}{k_bT}$ are charge, radius of ion i and thermal energy, respectively. $\lambda_b = \frac{e^2}{4\pi\epsilon_0\epsilon_r k_bT}$ is Bejerrum length which refers to the separation

of two elementary charges at which the electrostatic interaction potential is comparable to thermal energy. ϵ_0 and ϵ_r are vacuum and relative permittivities, respectively. The terms Γ, η and Δ refers to the contribution from electrostatic shielding, unequal sizes of ions and free volume not occupied by hard spheres, respectively (see Eq. S1 for specific expressions). We additionally include the solvation contribution (S_i) as estimated via generalized Born theory of ion hydration energies in MSA:

$$\Delta S_i = \frac{z_i^2}{r_i}\left(\frac{e^2 N_a}{2 \cdot 4\pi\epsilon_0}\right)\left(1 - \frac{1}{\epsilon_r}\right) \tag{1.18}$$

Where N_a is the Avogadro constant. The excess potentials of each ion species in bath solution and in filter were calculated and denoted as $\mu_{o,i}^{ex}$ and μ_i^{ex} respectively. At equilibrium, ion concentration in the filter (ρ_i) is given as:

$$k_b T ln\rho_i = k_b T ln\rho_{o,i} + \mu_{o,i}^{ex} - \mu_i^{ex} - z_i e\Psi - S_i \tag{1.19}$$

Where Ψ is relative Donnan potential of filter to bath and could be solved via a iterative procedure as that described in.[92] In CBPs, the features of filter (Ca^{2+} binding sites), such as oxygen number and filter volume, could be inferred from MD.[94] The ion charge (z_i) and radius (σ_i) can be adapted from widely use force field.[95] Given the known bath concentrations of each ion species ($\rho_{o,i}$) and filter features, the Ca^{2+} concentration in filter as well as excess chemical potential could be achieved by solving Eq. 1.17 and Eq. 1.19 iteratively.[92]

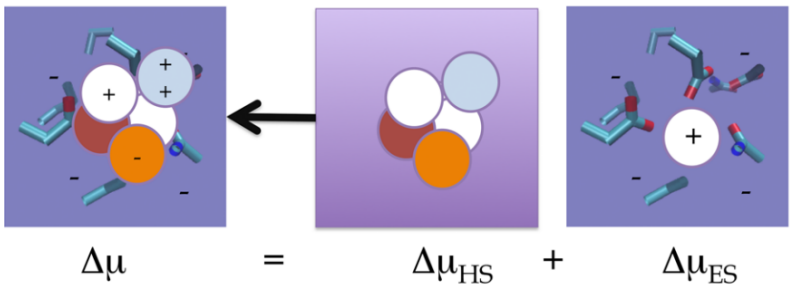


Figure 1.5: Partition of cations from bulk solution into protein binding sites bearing coordinating oxygen atoms via mean sphere approximation (MSA). Image is adapted from[94]

Protein-protein docking

Ca^{2+}-signaling triggered cellular functions heavily rely on various protein-protein interactions. Locating the protein-protein interaction (PPI) surface and further identifying key residues that contributes to the interaction provides substantial theoretical guidelines for designing drugs that can modulate disease-related PPI.[96] However, identifying the PPI interaction surface is challenging because the interaction surface is usually flat and lack specific interaction patterns.[97] Studies aim to extract the features of PPI interaction surface generally focus on the secondary structure characterization. For example, it has been shown that there is greater chance for a helix to reside in the grooves formed between α helices of its binding partner.[98,99] Protein-protein docking methods, in this regard, provides us a toolset to study more detailed residue-residue interaction at PPI interaction surface. Various protein-protein docking engines such as ZDOCK[100] and RosettaDOCK[101] have been used to reveal the structural details of protein-protein interactions.[102–104]

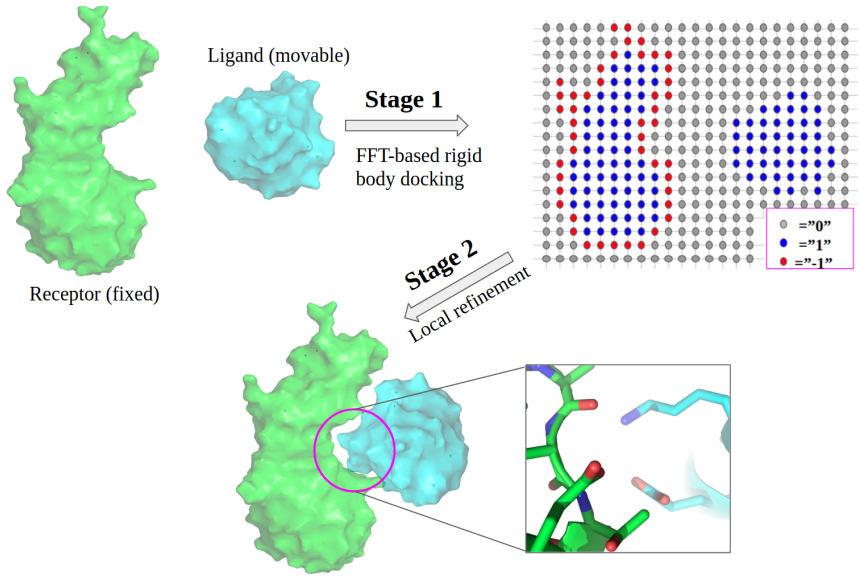


Figure 1.6: General two stages of protein-protein docking. In stage 1, the receptor and ligand are treated as rigid bodies and a fast Fourier transform (FFT)-based searching procedure is performed on grids constructed on receptor and ligand to globally search poses that satisfy shape complementarity. To demonstrate this process, we assume that in the given simple example, there is only atom type interaction ($I = 1$ in Eq. 1.20) and the outlier points of receptor has interaction value as $R_i = -1$, and interior has $R_i = 1$, for ligand $L_i = 1$. The rest points have interaction value equals 0. It was clearly seen that E in Eq. 1.20 reaches minimium when ligand has largest contact with the outlier layer of receptor. In stage 2, local refinement is performed on docked pose to optimize the side chain interactions between receptor and ligand

Currently protein-protein docking methodology consists of two general stages[105] (Fig. 1.6). The first stage is a global search of rigid receptor (the larger protein that is fixed) and ligand (the smaller protein that is movable) poses subjected to a score function consisting of shape and chemical complementarity,[105,106] with the former plays a major role. The goal of this global search is to generate receptor-ligand poses that are nearly close to native binding complex. The importance of shape complementarity is underlined by the idea that conformational change happened upon binding are small compared to the relative position of binding partners[105] and poses satisfy shape complementarity have more energetic interactions due to large surface area excluded from solvent.[107] The chemical complementarity mainly refers to electrostatic interaction,[108] van der Waals interaction and (de)solvation energies[109] at the PPI inter-surface. Including the chemical complementarity is indispensable as the binding is eventually driven by thermodynamic advantage (ΔG) upon binding. In the global search stage, receptor and ligand are represented by 3D-grids with user-defined spacing such that a fast Fourier transform (FFT)-based search procedure can be applied

on the grids. In the FFT implementation, the score function usually has the following expression:[110]

$$E(\alpha,\beta,\gamma,\lambda,\mu,\nu) = \sum_{i=1}^{I} \overline{R_i(x,y,z)}\hat{T}(\lambda,\mu,\nu)\hat{D}(\alpha,\beta,\gamma)L_i(x,y,z) \tag{1.20}$$

Where $\hat{T}$ and $\hat{D}$ are translational and rotational operator, respectively. I is total number of atom type interactions, R_i and L_i are defined on the grid points of receptor and ligand and have the meaning of interaction values of atom type i. The fast Fourier transform (FFT) is commonly used to accelerate the search in translational space $\hat{T}$, for example, the translational operation of ligand in Fourier space $L_i(n,m,l)$ is:

$$\hat{T}(\lambda,\mu,\nu)L_i(n,m,l) = e^{\frac{-2\pi i}{N(n\lambda+m\mu+l\nu)}} L_i(x,y,z) \tag{1.21}$$

Similarly, after applying FFT on the translational space of receptor, the score function can be expressed in terms of receptor and ligand in Fourier space as:[110]

$$E(\alpha,\beta,\gamma,\lambda,\mu,\nu) = \sum_{i=1}^{I}\sum_{nlm} \overline{R_i(n,l,m)}\hat{D}(\alpha,\beta,\gamma)L_i(n,l,m)e^{\frac{-2\pi i}{N(n\lambda+m\mu+l\nu)}} \tag{1.22}$$

The biggest advantage of converting score function in Cartesian space (Eq. 1.20) to Fourier space (Eq. 1.22) is that the latter has significantly less operations ($O(N^3 log N^3)$) than that ($O(N^6)$) in Cartesian space.[110,111] This improved search efficiency makes the exhaust search of ligand-receptor relative conformations feasible. In the second stage, after the nearly native receptor-ligand complex is achieved, the local refinement is performed at PPI interface to further optimize residue-residue interactions. In this stage, two strategies are commonly used to rearrange the sidechains of interacting resides: 1) Energy minimization after parameterized by empirical force fields,[112] and 2) assigning orientations based on sidechain rotatmer library sampled from available structural database.[113] Protein flexibility is to certain extent accounted by this local refinement process, with the aim to achieve more native-like binding poses. It should be noticed that, experimental data, when available, can be incorporated in these stages of protein-protein docking to narrow down the search scope and improve docking accuracy.[114]

Chapter 2 Characterizing Calcium Diffusion in Mesoporous Silica Thin Films- A Model Resemble Cellular Calcium Diffusion

- This chapter is based on "Sun, B.; Blood, R.; Atalay, S.; Colli, D.; Rankin, SE.; Knutson, BL.; Kekenes-Huskey,PM. book chapter of *Computational Materials, Chemistry, and Biochemistry: From Bold Initiatives to the Last Mile*, Springer International Publishing. **2017**"

2.1 Introduction

Ion transport in mesoporous films and the impact of defects

Mesoporous silica films have garnered considerable interest for applications to separation chemistry, drug delivery, and biosensors[115] owing to their tunable control of ion mass transport processes within their highly-charged porous networks. In part, this precise tuning is afforded through the high density of silanol groups on silica surfaces,[27] which can support pH-dependent charge regulation and substrate adsorption in aqueous media.[116,117] In turn, these surface chemistry phenomena have been shown to modulate ion permeation and conductance, beyond estimations based on restricted diffusion or tortuosity alone. Here electrostatic interactions play a profound role in modulating ion transport, for which the ionic-strength determined electric double layer length scales are on the same order as pore diameters. As a result, there is a rich set of physical phenomena that can emerge in these materials. For instance, two extremes in background ionic strength give insight into the complex tug and pull governing ion transport. At low ionic strengths, potentials arising from surface charges can extend significantly into the bulk fluid phase, where they strongly dictate rates of electrodiffusion. At high ionic strengths, surface potentials are largely attenuated, in which case transport is largely controlled by concentration gradients and differences in osmotic pressure.[118–120] These phenomena have been characterized in nanoporous 'slits' and 'channels', but have been less explored in nanoporous films with uniform ordered mesopores generated by liquid hexagonal close packed (HCP) (mesocrystalline) templating.[116,121–124]

Diffusion-limited transport of small molecules in nanoscale media is sensitive to material morphology, surface charges, the solution ionic strength and physiochemical properties of the diffuser such as size and charge. Therefore, a prominent challenge in assessing transport in mesoporous silica films is how these contributions collectively control ion diffusion rates in perfectly-formed porous regions, and the extent to which structural defects perturb these rates. Among these, framework defects are among the most poorly understood factors and are known to perturb substrate diffusivity and permeation in nanoscopic materials.[125] Framework defects in HCP (mesocrystalline)line materials share many similarities with mesoporous films, and are categorized by their size as

micro (<20 Å), macro (>500 Å), and meso (intermediate),[126] which include cracks and holes representing the predominant macro-scale defects,[126] while imperfect mesoHCP (mesocrystalline) formation or intergrowth account for the majority of meso- and micro-scale defects.[127] Below 500 Å, the most common structural defects include HCP (mesocrystalline) stacking faults resulting in merged pores,[128,129] inter-growth of different mesoHCP (mesocrystalline) forms[130,131] and hydroxyl (OH)-terminated surfaces[132] (silanol nests). Thus, critical for evaluating and optimizing chemical processes in real materials is a fundamental understanding of mass transport[132–134] in both ordered (HCP (mesocrystalline)) and defect-containing mesophases.

Techniques for incorporating microscopy data into nanoscale simulations

Transmission and scanning electron microscopy have become the standard imaging modalities for probing the structural integrity of nanoporous media.[135,136] Though considerable effort has been invested in characterizing prominent HCP (mesocrystalline) defects in electron micrographs, less has been done to simulate the impact of representative defects on material transport and performance properties. For instance, studies examining gas/liquid adsorption in nanoporous silica with morphological defects have been reported based on molecular simulations,[137,138] but these were not explicitly linked macroscale transport phenomena. It is our speculation that the abundance of defects and difficulty in translating these structural features to forms amenable to simulations have challenged probing via simulation nanoscale transport phenomena in structurally-imperfect nanoporous media. Here, advances in automated segmentation of electron microscopy data has the potential to ease the burden of manual identification and characterization of material structural features, which could serve as the basis for detailed substrate transport simulations. Recent examples include utilizing image processing such as equalization, segmentation and Shannon entropy to characterize porosity and other features in nanomaterials.[139–142] Conversely, techniques for reconstructing 3D structures from 2D electron microscopy (EM) slices are widely implemented for investigating nanoparticle and nanocatal-yst structures in complex hiearchical arrangements.[143,144]

Modeling approaches for ionic transport in mesoporous media

Computational models for estimating the extent to which defects impact nanoporous material performance could improve material design, yet most approaches assume perfectly ordered mesop-orous materials. For perfect (defect-free) nanoporous silica materials, a variety of simulation approaches have been developed to characterize substrate adsorption and transport, including continuum, stochastic methods such as Monte Carlo and molecular dynamics simulations (reviewed in[145]). While molecular simulations provide a powerful toolset for probing atomic-level physical phenomena including local self-diffusion, hydration and adsorption,[146–148] continuum transport simulations constitute a

multi scale complement to extrapolate nanoscale molecular-scale physiochemical information to macroscopic phenomena occurring in nanoporous material.[149–151] Among the most successful models of ion transport at the continuum level is Poisson-Nernst-Planck (PNP) theory, which is a continuum model of electrostatically driven ion diffusion coupled with ion-dependent electrostatic fields. Recent developments have augmented PNP simulations to include pH-regulated surface charge[151] for modeling ion conductance in nanoporous materials. Despite the wealth of simulation effort targeting nanoscale materials, how ionic transport might vary in heterogeneous or defect-containing nanoporous materials is under explored.

Paper Objectives

In this study, we have developed a workflow as a foundational step toward imaging-informed, computational modeling of ion transport in mesoporous silica films with structural defects. This workflow (Fig. 2.1) enlists a computer vision technique, matched filtering, to 1) discriminate HCP from defect-containing regions from electron microscopy (EM) 2) performs partial differential equation simulations of electrokin-etic transport in 3D models of such regions, and 3) estimates heterogoneous effective transport parameters in a given material. Step 1 utilizes matched filtering to automatically detect mesocrystalline features for EM . In this procedure, kernels representing such features are convolved against the data; the kernel generating the maximal response above a user-defined signal-to-noise ratio is used to annotate the region in the data. Step 2 the kernels developed for step 1 are converted into 3D meshes by projecting the 2D signature perpendicular to plane to create a 3D pore. Step 3 A pH- and surface charge-dependent model of electrolyte transport (PNP) partial differential equation is solved in the 3D meshes using the finite element method, from which effective transport parameters such as conductivity and diffusion are calculated. Step 4 is to interpolate the effective parameters from Step 3 onto the annotated regions determined in Step 1. We applied this to a mesoporous silica film synthesized and characterized by Wooten et al[152] (electron microscopy data in Fig. 2.3) as well as simulated data to demonstrate the algorithm performance.

We demonstrate that our workflow for the first time automates electrokinetic transport simulations in microscopy-derived, defect containing structural data of mesoporous films. Our proposed computer vision method demonstrates reasonable accuracy in discriminating mesocrystalline (bulk) from defect regions in transmission electron microscopy data and simulated data, thereby providing means to characterize the nature and prevalence of defects. Further, our implementation of the pH-/surface-chemistry dependent PNP model is consistent with experimental measurements of KCl and $CaCl_2$ conductance in silica-based nanochannels over a broad range of ionic strengths and pHs, delineating regimes that permit simplified electrokinetic models. Without any additional fitting, the model predicts permeation properties of an anionic dye 5(6)-Carboxy-

fluorescein (CF) in a mesoporous film within experimental uncertainty. With this model, we predict the degree to which defects in a mesoporous film influence permeation properties, the optimal conditions under which to make these variations apparent, and conditions for selectively tuning small molecule permeability based on pore size, charge and buffering. Overall, we envision that this model will provide rigorous means to characterize high resolution microscopy data, from which heterogeneous transport parameters can be estimated. Further, its basis as a finite element model should permit its extension to wide-ranging material types, including hierarchically-structured composite materials.

2.2 Methods

Our workflow for segmentation and PDE-based simulations of ion transport in EM-resolved nanoporous media is shown in Fig. 2.1. Key stages of this workflow include 1) automated unit cell feature detection in EM-characterized nanoporous media (Sect. 2.2), 2) three-dimensional meshes based on the detected unit cells (Sect. 2.2), 3) effective transport parameter estimation based on Poisson-Nernst-Planck (PNP) simulations of ion transport within porous regions of each unit cell (Sect. 2.2) 4) extrapolation of unit cell transport parameter estimates onto the imaged material surface (Sect. 2.2).

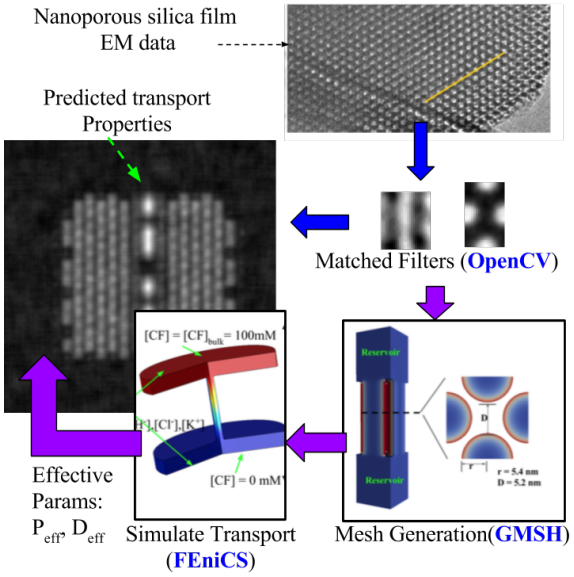


Figure 2.1: Workflow based on structural data from Wooten et al.:[152] 1)Matched filter unit cell detection from EM data. 2) 3D geometries construction with meshing. 3)PNP solved in representative unit cell geometries. 4) Interpolate effective transport parameters for entire film from unit cells

Matched filter unit cell determination from segmentation of bulk mesocrystal and defect EM data

We utilize 'matched filtering' for the first stage of our workflow, in order to assess the likelihood of a kernel representing a feature of interest is present in a given data set. Consider a measurement, $\vec{m}$, that consists of of a signal, $\vec{s}$, embedded in additive noise, $\vec{n}$:

$$\vec{m} = \vec{s} + \vec{n}, \tag{2.1}$$

The goal of matched filtering is to identify a matched filter, $\vec{h}$, that maximizes the signal-to-noise ratio (SNR) for a measurement, $\vec{m}$,[153]

$$y = \vec{h}^T \vec{m} = \vec{h}^T \vec{s} + \vec{h}^T \vec{n}, \tag{2.2}$$

where $\vec{h}^T$ denotes the filter's transpose. It can be shown that the optimal matched filter can be determined via

$$\vec{h} = \frac{1}{\sqrt{\vec{s}^T R_n^{-1} \vec{s}}} R_n^{-1} \vec{s}, \tag{2.3}$$

where $R_n = E\{nn^T\}$ represents the noise covariance matrix. In the event that random variates are drawn from a mean-zero, Gaussian, white noise process

of variance, σ^2, the noise covariance matrix reduces to $R_v = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Commonly, the signal or multiple instances thereof may be embedded within a larger data set (r), such as an image. In which case, determining the location of s within r is commonly performed by convolving the kernel h with the image, r

$$Y = h * r, \tag{2.4}$$

Computationally, this is commonly done via the discrete Fourier transform, given that

$$h * r = \mathscr{F}^{-1}\left[\mathscr{F}[h] \cdot \mathscr{F}[r]\right], \tag{2.5}$$

by the convolution theorem. Probable detections of the signal s within Y are based on identifying positions, at which the SNR is above a user-specified threshold criterion, λ,

$$\text{SNR} \equiv \frac{|\vec{h}^T\vec{s}|}{\vec{h}^T\sigma^2\mathbf{I}\vec{h}} \geq \lambda, \tag{2.6}$$

where σ_n represents the standard deviation of the noise.

In this study, we consider a 90nm thick mesoporous silica film with hexagonally-packed, 5nm radius pores (see Fig. 2.2a) synthesized by Wooten et al,[152] from which the matched filters were determined. These data reveal a preponderance of ordered hexagonally-packed pore regions with a small number of defects evident as small linear features. It is speculated that the linear features represent pores that became fused during the film preparation, thereby yielding a 'channel-like' fused pore shown in Fig. 2.2b. Since the primary goal of the matched filtering is to discriminate bulk regions from defect features, the underlying structure corresponding to the linear feature in the EM is of little significance. Therefore, here we determine the matched filter kernels, or 'filters', based manually identifying representative bulk and 'fused pore' regions in the EM data. We first applied the PYTHON OPENCV 'Contrast Limited Adaptive Histogram Equalization' routine to equalize the pixel intensities across the image. After which, the identified regions were subdivided into unit cells representing an instance of the nanoporous feature. These subdivisions were averaged to attenuate uncorrelated background noise and reveal the consensus structures shown in Fig. 2.2, which served as matched filters for each data feature. In our approach, we consider multiple filters that represent different features in the EM data, as well as rotations thereof in order to detect alternate orientations. Since the orientation of the bulk hexagonally-packed regions and fused pore features vary across the EM field of view, we created a bank of filter rotations for each matched filter that were spaced at 10 degree increments. The correlation response for a given image pixel, y_{ij}, was determined by taking the maximum response across all rotated variants of a given filter. Given that the data considered in this study contains signal (nanopores), noise that is not necessarily additive, and other features in the EM data that are not represented

by a given filter, h, we modified our threshold criterion as follows

$$\log \frac{\exp(\vec{h_i}^T \vec{s})}{\sigma^2 \gamma \exp(\vec{h_{iC}}^T \vec{s})} \geq \lambda_i, \tag{2.7}$$

where $\vec{h_{iC}}$ represents the complement of matched filter i, $\vec{h_i}$, which we define as $\vec{h_{iC}} = 1 - \vec{h_i}$. This complement penalizes signal that falls outside of the signal signature defined in $\vec{h_i}$. We found this term was necessary to discriminate the correlation outputs from the bulk and fused pore matched filters. We tested both filters against two subsections of a 'fused pore-rich' region and a 'bulk-like' region (see bottom left panels of Fig. 2.3). Regions of the test EM data that returned responses below the threshold parameters for either filter are designated as 'uncharacterized.' All aforementioned numerical procedures were conducting using the PYTHON2.7 libraries NUMPY, SCIPY and OPENCV-PYTHON.

Mesh generation from matched filter unit cells

Effective parameter estimation in the second stage of Fig. 2.1 is based on numerical solution of the PNP equation via the finite element method in Sect. 2.2, using 3D meshes informed from the segmented images. Unit cells determined from our segmentation protocol provided a basis for generation of tetrahedralized, finite element meshes via GMSH,[154] for which the pore radii and spacings were approximated from the segmented data. In principle, however, the segmented data could be used directly for mesh generation.[155] From these data, we created 'extruded' unit cells of length 90nm, which assumed the inner pores are perpendicular to the EM-resolved film surface (Fig. 2.2). The MathEval and Box field in GMSH were used for mesh refinement. Specifically, these functions allow one to increase the mesh resolution as a user-defined function of distance from features of interest, such as near the nanopore walls where higher concentration and electric potential gradients are expected. Mesh generation scripts demonstrating these features are provided in the bitbucket repository `https://bitbucket.org/pkhlab/poissonnernstplanck`. The extruded pores interfaced with two identical reservoirs to represent contact with bulk solution. In principle, the reservoir size should be significantly larger than the pore dimensions to minimize artifacts introduced by the reservoir boundaries on the electrostatic potential adjacent to the silica surfaces.**LinChen2016**, [156] In a recent study,[157] it was demonstrated that more modestly sized reservoirs on the order of 20 nm were sufficient to minimize these artifacts; here, we set the reservoir depth(z direction) to be 40nm, while the width and length of reservoir are shown as in Fig. 2.2. Meshes resembling 'nanochannel' and 'nanoslit' geometries were constructed in a similar fashion for the validations described in this study (see Fig. S1). Moreover, we evaluate the conductivities near the midpoint of the silica pores, which is expected to further reduce boundary artifacts. In the Results section, we demonstrate good agreement

with experimental conductivity measurements, which suggests our choices of domain configuration and conductance measurements (see Methods) were appropriate.

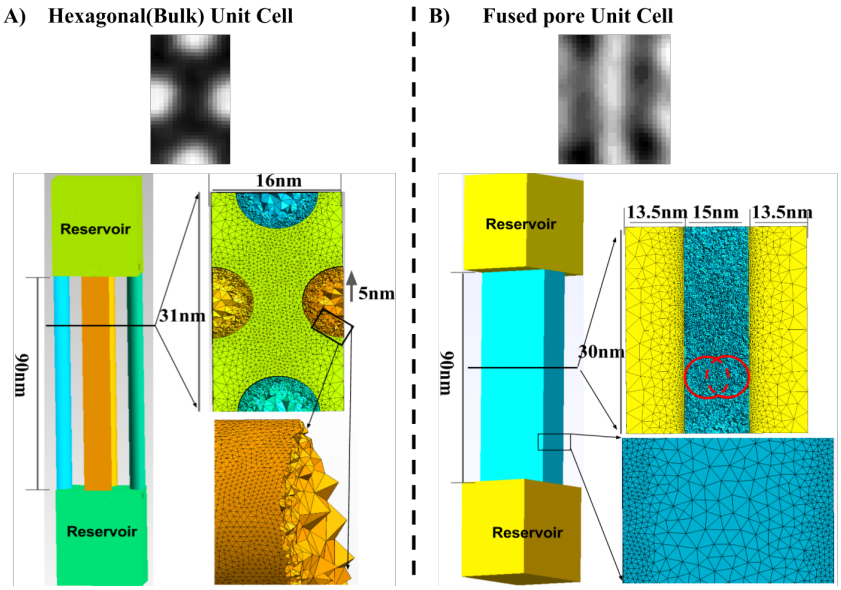


Figure 2.2: Two unit cells detected by our "matched filter" method based on the EM image of silica membrane fabricated by Wooten et al.[152] The dimensions of the unit cells were automatically determined by the segmentation procedure. The corresponding 3D geometries (with meshing) generated by GMSH[154] are also shown. A) Hexagonal(bulk) unit cell. B) Fused pore unit cell

Effective transport parameter determination via finite element solutions of the Poisson-Nernst-Plank transport model

The mass transport of ions in a silica nanopore was described by the Poisson-Nernst-Planck (PNP) equation. The Nernst-Planck (NP) equation describes the ionic mass flux density of each ion species subject to concentration and electric potential gradient. Meanwhile, the electric potential distribution within the domain is determined by Poisson equation. These two equations are given as:

$$-\nabla \cdot J_i = 0, \tag{2.8}$$

$$J_i = -D_i \left(\nabla c_i + \frac{z_i F c_i}{RT} \nabla \phi \right), \tag{2.9}$$

$$-\epsilon_r \epsilon_0 \nabla^2 \phi = F \sum_{1=1}^{N} z_i c_i, \tag{2.10}$$

Here, J_i, D_i, c_i, z_i are flux density, diffusion coefficient, molar concentration, and valence electron number of i^{th} ionic species. F is the Faraday constant, ϕ is electric potential, T is absolute temperature, and R is gas constant. ϵ_0 and ϵ_r are vacuum permittivity and the relative permittivity of the electrolyte solution, respectively. The boundary conditions are given as: 1) When evaluating

ionic conductance: at the both ends of the reservoir, ionic concentrations are maintained at the bulk values (i.e., $c_i = c_{i,bulk}$) while potentials of $\phi = 0$ and $\phi = 0.2V$ are applied. 2) When evaluating effective diffusion constant/permeability: at the both ends of the reservoir, electric potentials are set as 0 (i.e., $\phi = 0V$) while concentrations of $c_i = c_{i,bulk}$ and $c_i = 0$ are applied at either end. 3)Within the nanopore, we apply a reflective boundary condition for the ions ($\boldsymbol{n} \cdot \boldsymbol{J_i} = 0$). 4) A Neumann condition on the potential is also applied based on the silica surface charge density:

$$-\nabla\phi \cdot \boldsymbol{n} = \sigma_s/(\epsilon_o\epsilon_r), \tag{2.11}$$

$\boldsymbol{n}$ is the unit outer normal vector.

Surface protonation and K^+/Ca^{2+} surface adsorption

Recently, the Qian group and their collaborators have conducted a series of numerical electrodiffusion simulations based on silica nanochannel geometries.[116,123,158] These models assumed four ionic species, H^+,OH^-,Cl^- and K^+, as well as silanol (SiOH) protonation:

$$\mathrm{Si{-}OH} \rightleftharpoons \mathrm{Si{-}O^-} + \mathrm{H^+}, \tag{2.12}$$

$$\mathrm{Si{-}OH} + \mathrm{H^+} \rightleftharpoons \mathrm{Si{-}OH_2^+}, \tag{2.13}$$

The inclusion of the pH-dependent regulation of surface charge density in the PNP model was found to give superior agreement with experimental measurements of KCl conduction at non-neutral pH. More importantly, their results indirectly demonstrate that metal adsorption between the monovalent cations (e.g., K^+) and the channel wall silanol groups is negligible,[159,160] as their PNP model was sufficient to recapitulate experimental results without considering K^+ adsorption. For divalent cations such as Ca^{2+}, adsorption onto the silica surface is believed to be significant.[159,161,162] This motivated a site-binding model for divalent cation adsorption that appeared most consistent with experimental data among several reaction possibilities.[159]

$$\mathrm{Si{-}OH} + \mathrm{Ca^{2+}} \rightleftharpoons \mathrm{Si{-}OCa^+} + \mathrm{H^+}, \tag{2.14}$$

The corresponding equilibrium constants of the (de)protonation and Ca^{2+} adsorption reactions are thus given by

$$K_{a1} = \frac{\eta_{SiO^-}[H^+]_s}{\eta_{SiOH}}, \tag{2.15}$$

$$K_{a2} = \frac{\eta_{SiOH_2^+}}{\eta_{SiOH}[H^+]_s}, \tag{2.16}$$

$$K_m = \frac{\eta_{SiOCa^+}[H^+]_s}{\eta_{SiOH}[Ca^{2+}]_s}, \tag{2.17}$$

where η_i is the surface site density of the i^{th} functional group. $[H^+]_s$ and $[Ca^{2+}]_s$ are the surface molar concentrations of corresponding ions. To our knowledge, however, pH regulation and Ca^{2+} adsorption have not been used for modeling electrokinetic phenomena. Thus we combined models for protonation[163] and Ca^{2+} adsorption,[159] for which the total density of sites on the silica surface that can support chemical reactions is given by

$$\eta_{Total} = \eta_{SiOH} + \eta_{SiO^-} + \eta_{SiOH_2^+} + \eta_{SiOCa^+}, \tag{2.18}$$

From this equation, the effective surface charge density can be determined by

$$\sigma_s = -F\eta_{Total}\frac{K_{a1} - K_{a2}[H^+]^2 - K_m[Ca^{2+}]}{K_{a1} + [H^+] + K_{a2}[H^+]^2 + K_m[Ca^{2+}]}, \tag{2.19}$$

Eq. 2.19 serves is used in the Neumann condition definied in Eq. 2.11.

Finite element solution of PNP equations

The Poisson-Nernst-Planck (PNP) equations in the present study were numerically solved via the finite element method (FEM) using the commercial finite-element COMSOL (www.comsol.com) package and the free open-source FEniCS[164] library. For all two dimensional (2D) geometries we considered (see Fig. S1), the PNP equations were solved by COMSOL with full pH-/adsorption regulated surface charge density to ensure consistency with prior studies of ion conductance in nanomaterials.[116,165,166] For reasons of computational expense, we utilized the finite element method (FEM) package FEniCS to solve the PNP equations, assuming first-order Lagrange bases and default solver parameters. To simplify the boundary conditions, the Grahame equation was used to relate the surface charge density to the electric potential at the silica wall, which for a monovalent salt is given by

$$\sigma_s(\phi_0) = \sqrt{8c_0\epsilon_o\epsilon_r k_B T}\sinh\left(\frac{e\phi_0}{2k_B T}\right), \tag{2.20}$$

where ϕ_0 is the electric potential at the pore surface (for divalent salt, e.g., $CaCl_2$, the corresponding Grahame equation is given as Eq. S2).[163]

Conductivities and permeabilities of unit cells and bulk material

For numerical estimation of the KCl ionic conductance in nanoporous media, we evaluated the following conductance relationship proposed by Yeh et al:[116]

$$S = \frac{I}{\delta V} = \frac{F\langle J\rangle_\Gamma}{V_a - V_b}, \tag{2.21}$$

$$\langle J\rangle_\Gamma \equiv \sum_{i=1}^{N}\int_\Gamma J(\Gamma)_i d\Gamma, \tag{2.22}$$

where J_i, the flux density for species i, is computed from steady-state solutions to Eq. 2.8, V_a and V_b are the average electric potentials at the two ends of nanopore ($V_a - V_b$ value is close to applied potential bias $\delta\phi = 0.2V$, see Table S1), N is number of ion species, F is Faraday's constant, Γ is the cross-sectional surface within the nanopore center. The flux density was either provided directly from COMSOL or estimated from FENICS steady-state solutions using PARAVIEW.[167] For the $CaCl_2$ conductance in a nanochannel (length= 5mm, width=30 μm and Height=18nm[161]), we utilized a 2D model represented, given that width$\gg$ height. We assumed $G_{final} = G_{slit} W 10^{-4}$, where G_{slit} is the ionic conductance of the 2D nanoslit, W is the width of the 3D nanochannel. The division by 1.0000×10^4 reflects that our simulated domain was of length 5×10^{-7} m versus the 5×10^{-3} m channel used in Feust et al;[161] despite this approximation, we found reasonable agreement between our predictions and experimental data.

Similarly, the membrane permeability, P_{eff}, and effective diffusion constant, D_{eff}, were evaluated as:[168,169]

$$P_i = \frac{K_i D_{\mathrm{i,eff}}}{L_m} \tag{2.23}$$

$$D_{\mathrm{i,eff}} = \frac{\langle J_i \rangle L_x}{[i]_{bulk}} \tag{2.24}$$

where K_i and $D_{\mathrm{i,eff}}$ are the partition coefficient and effective diffusion constant of each species in membrane, respectively. L_m is the thickness of membrane, $\langle J_i \rangle$ is the average flux density over the cross-section area at the middle of membrane(calculated in the same way as V_a and V_b evaluations mentioned above), L_x is the length along diffusion direction(defined as the distance between the external ends of two reservoirs) and $[i]_{bulk}$ is the bulk concentration. The value of K_{CF} in silica membrane is assumed to be 1×10^{-3} , which lies in the range of small organic molecules.[168]

Extrapolation of effective conductivity estimates on EM-imaging data

In the final step, partial differential equation predictions of effective transport parameters are extrapolated onto the original EM-resolved structure. Given that each filter 'hit' represents a match for an entire unit cell, we the transport parameter estimated for the corresponding filter to a region commensurate in size to the unit cell. In regions that did not contain an obvious filter match, we assigned a permeation value that was intermediate to the fused and bulk pore unit cells, as the unclassified regions still presented porous features that could permit substrate diffusion. In this case, the effective permeation, P_{eff}, for the entire material surface is determined by a surface area-weighted (Γ) average of the parameters determined for the different unit cell types, e.g.

$$P_{\mathrm{eff}} = \frac{1}{\Gamma_{tot}} \sum_i \Gamma_i s_i \tag{2.25}$$

where i corresponds to the fused pore, bulk and unclassified regions. All code written in support of this publication are publicly available at https://bitbucket.org/pkhlab/poissonnernstplanck. Simulation input files and generate data are available upon request.

2.3 Results and Discussion

Automated feature detection and mesh generation for oriented porous films

A key contribution from our workflow presented in Fig. 2.1 is the automated detection of prominent structural features in imaged nanoporous films. In the Methods section, we outlined our procedure for generating matched filters representative of bulk (hexagonal closed packed) regions and fused pore defects. We note that the postulated fused pore is one of several types of defects evident in the data and that additional matched filters would be required to detect those features. However, given the lack of data for adequately training the matched filters for each defect type, we limit our approach to fused pores, which are somewhat prevalent in the EM image (Fig. 2.3B).

Here we tested the performance of these data-derived filters on subsections of the raw EM data that were not used for filter training. Namely, in Fig. 2.3 we present a roughly 100nm by 100nm region that contains diagonal striations that we attribute to fused pore features. In the top two rows of Fig. 2.3, we show the rotated filters (left column) as well as the corresponding matched filter outputs (right column) for the fused pore filter. In the bottom row we denote the raw data used for the feature detection, as well as the above-threshold regions from all pore rotations indicated in red or green for the bulk and defect filters, respectively. The marked results in Fig. 2.3 suggest that the matched filtering protocol is able to detect and classify the bulk and fused pore features, though exact quantification of the accuracy is difficult given the resolution of the EM data. We note that there are several regions in the image that were not classified by either filter. By visual inspection, those regions present surface features that neither appear fused nor adhere to an HCP configuration. In principle, these unclassified regions could be used to train additional filters to facilitate complete characterization of the EM surface. We also found that tuning the threshold parameters was necessary to optimize the matched filtering results. Nevertheless, these data indicate 1) that filter rotation in ten degree increments is sufficient to reliably identify data features independent of their orientation and 2) that bulk and fused pore defect regions can be automatically detected in the raw test data.

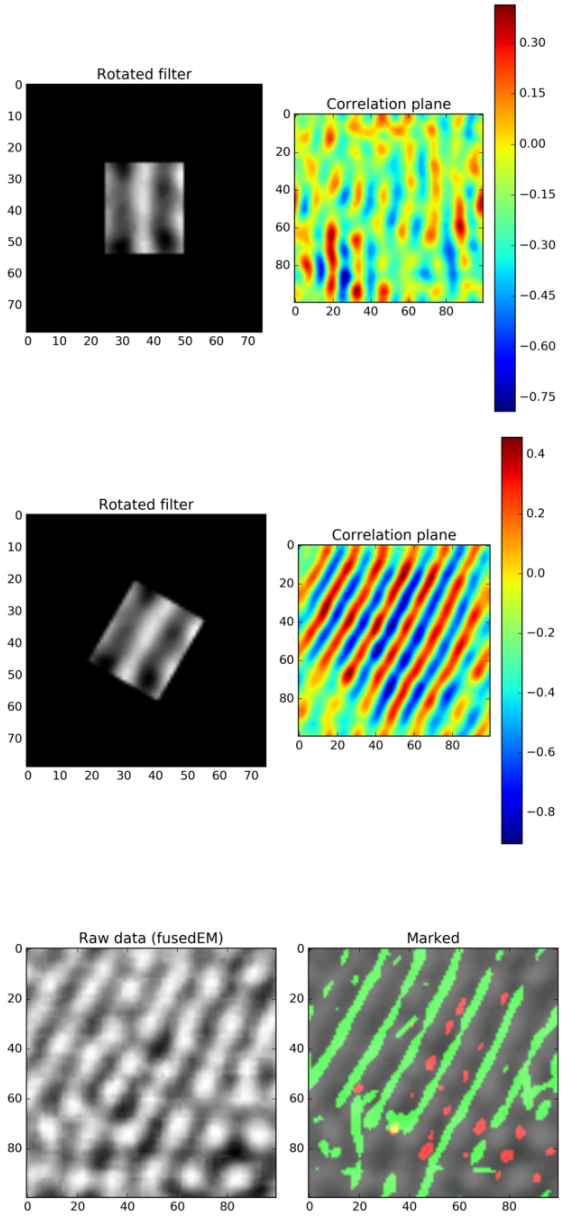


Figure 2.3: Matched filtering results for simulated transmission electron microscopy (TEM) images of a mesoporous silica film from Wooten et al.[152] First and second rows correspond to the filters and corresponding matched filtering result at 0 and 30 degree rotations. Bottom row provides the raw data and an image denoting match filter detected fused pore (green) and bulk unit cell(red) features.

Electrokinetic model of transport in oriented mesoporous films and other porous media

KCl conductance in mesoporous silica.

The intermediate stages of the Fig. 2.1 workflow entails estimation of effective transport parameters including conductivity (S_{eff}) and effective diffusivities (D_{eff}) or permeabilities(P_{eff}) for ionic species, given 3D representations of the porous features identified in the EM film data. These features include a perfectly cylindrical pore corresponding to the HCP bulk unit cell (Fig. 2.2A) and a 'slit-like' geometry representing fused pores that traverse the entire film depth (Fig. 2.2B). The dimensions of the two unit cells were determined by the segmentation procedure. The length and width of hexagonal unit cell were determined to be 31nm and 16nm and four one-half nanopores centered on each edge, for which each pore has a radius of 5nm. The length of fused pore unit cell was set to 30nm, while the width consisting of a 15nm slit formed by the fusion of two pores (represented by two red circles in Fig. 2.2B) centered between two 13.5nm wide impermeable regions. While the fused pore feature in principle could reflect a cylindrical pore oriented parallel to the film surface, it would not conduct ions traversing perpendicular to the film thus we do not explicitly consider this morphology.

A focal point of this section is the validation of our implemented model against several experimental assays of electrolyte conductance in nanochannels or nanoslits,[117,161] in order to establish confidence in its application to a distinctly different morphology: mesoporous films. In this section, we describe the computational modeling of KCl and $CaCl_2$ diffusion in several nanoporous silica morphologies, as well as nanochannels and nanoslits previously characterized in the literature.[117,161] We additionally consider mesoporous silica films synthesized and characterized by Wooten et al.[152] All systems are modeled subject to voltage gradients or concentration gradients for measuring conductance or diffusion properties, respectively, under a broad range of ionic strengths. In concurrence with prior studies,[116,118,165,166] we describe the electrokinetic mass transport using the PNP model under steady-state conditions, whereby the electrostatic field (ϕ) and electro-diffusion of electrolytes are coupled and solved simultaneously (Eq. 2.8). We further include reaction terms reflecting proton and metal equilibria with the silanol-terminated silica surface (Eq. 2.12 and Eq. 2.14), which together determine the surface charge density governing the PNP model (Eq. 2.19). While prior computational studies have characterized aspects of conductance in nanochannels and nanoslits,[116,166] in this study we examine such transport phenomena in defect-containing mesoporous silica films.

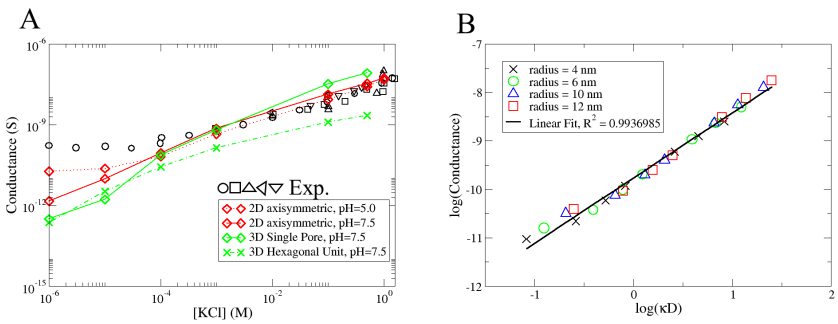


Figure 2.4: KCl ionic conductance in a nanopore. A) Comparison between experimental data and numerical results. For the 2D axisymmetric pore, the full PNP pH/adsorption model was used while for the 3D nanopore, the PNP was solved with fixed electric potential (values are from 2D results) applied at nanopore wall. Black symbols designate experimental KCl conductance data for a 34 nm silica nanopore of radius 5.1 nm (pH=7.5) from.[117] B) Ionic conductance of nanopore (2D axisymmetric, pH=7.5) at varying radiis and bulk [KCl] (expressed as κD, D=10.2nm is the diameter of nanopore)

Using the validated pH-dependent PNP model, we predicted the conductance of a KCl solution (H^+, OH^-, K^+ and Cl^-) through a single 34nm nanopore of radius 5.1nm at pH=7.5 and pH=5, assuming a 2D axially-symmetric domain. These conditions mirror those considered by Yeh et al,[116] although their model additionally included a modified Stokes component to capture electrosmotic flow. Consistent with experimental measurements from Smeet et al[117] and Yeh el al,[116] the model predicts that conductance scales with concentration at higher ionic strengths, as the electric potential of the channel walls are largely shielded by short Debye lengths. We note that while the Smeets et al[117] study reports findings based on a roughly neutral pH, we found that pH=5 was necessary to recapitulate those findings, and further, the conductance data resemble those from a related experiment conducted at pH=5.[118] Regardless of pH, the predicted conductance data show similar declines in conductance from 1×10^{-3} to 1M, which suggests that the contribution of surface charge is negligible at higher ionic strengths. In other words, the channel flux is largely dependent on the applied field parallel to the channel and that electrostatic interactions within the channel are of little consequence under these conditions. At lower ionic strengths, conductance is dictated by surface charge owing to the overlapping of electric double layers, however the pH mitigates this effect by attenuating the net negative charge of the boundary. We found analogous trends for a 3D nanopore domain the HCP unit cell assuming a fixed (pH independent) surface charge.

The change in conductance can be rationalized based on the fluxes of each ionic species, which arise due to concentration gradients (first term in the right hand side of Eq. 2.8) and electro-diffusion (second term of Eq. 2.8). For the latter

of which, the amplitude decreases as a function of concentration (see Fig. S2), which ultimately determines the overall conductance. It is worth noting that the effect of the potential gradient along the pore wall on ion diffusion is akin to widening or constricting the pore radius for counter- or co-ions, respectively.

Interestingly, when these conductance data are represented as a function of the dimensionless parameter κD, the trends scales as log(S) versus log(κD), as shown in Fig. 2.4B. This trend is observed both when ionic strength is varied from 1×10^{-2} to 80mM for fixed pore radii (4-12 nm), as well as when radii are varied for a fixed ionic strength. The latter variation supports the notion that the presence of surface charge on the pore boundary essentially modulates the effective pore diameter. Note that this κD dependence breaks down for variable (e.g. pH-dependent) surface charges at low ionic strength ($\kappa \to 0$).

$CaCl_2$ transport and adsorption

In the preceding section, we demonstrated agreement between numerical predictions of KCl conductance in several nanoporous geometries, for which the surface charge is dependent on pH but independent of K^+. We now consider the extent to which an adsorbable ion can compete with protonation and thereby influence ion conduction. Specifically, we consider Ca^{2+} adsorption to the surface, which has been shown to be significant[118,161] and a necessary factor in conductance models to recapitulate experimental data[117,170] Accordingly, we include in Eq. 2.19 Ca^{2+} surface adsorption, assuming equilibrium constants in the micromolar range. For validating the $CaCl_2$ model, we refer to $CaCl_2$ ionic conductance data collected from a 'nanoslit' with a length of 5mm, width of $30\mu m$ and height of 18nm, respectively (Fuest et al[161]). Here we assume a 2D domain in Cartesian space (length versus height), given that the width is much greater than the height. Further, since at steady state the ionic conductance of a nanochannel is inversely proportional to the length of the nanochannel,[171] we assumed a slit length of 500 nm instead of 5 mm for reasons of computational expense(Fig. S1). The simulation domain in this example contains four ion species: H^+,OH^-,Ca^{2+} and Cl^-.

Simulated and experimentally-measured conductances are reported in Fig. 2.5 for PNP models with and without Ca^{2+} adsorption. Analogous to our simulations for KCl, at low pH (pH = 5), we note that conductance (in log units) decreases linearly with log [$CaCl_2$] for calcium chloride concentrations of 1 mM and higher, in agreement with experiment. Since the conductivities predicted among the surface charge models were comparable, it is expected that the pH and Ca^{2+} do not significantly modulate the surface change density for $CaCl_2$ above 1 mM. As the concentration is decreased below 1 mM, however, the rate of decrease in conductance slows and approaches a minimum at approximately 100 μM, below which conductance accelerates with decreasing ionic strength. This is attributed to the fact that at low ionic strength, protons are more effective in neutralizing wall charge. Both the pH and pH+Ca-dependent surface charge models capture this behavior to an appreciable degree, although they do not capture the

magnitude of the conductance increases exhibited in the experimental data. We attribute this discrepancy in part to the uncertainty in the parameters η_{total}, pK_a, pK_b and pK_m. We attribute the experimentally-observed trends at low ionic strength to the increased involvement of pH in regulating the surface potential, as discussed in.[162,172] Namely, as the ionic strength of the solution is decreased, protons have a higher tendency to be attracted to the negatively charged slit walls, upon which they neutralize the negative wall charge.

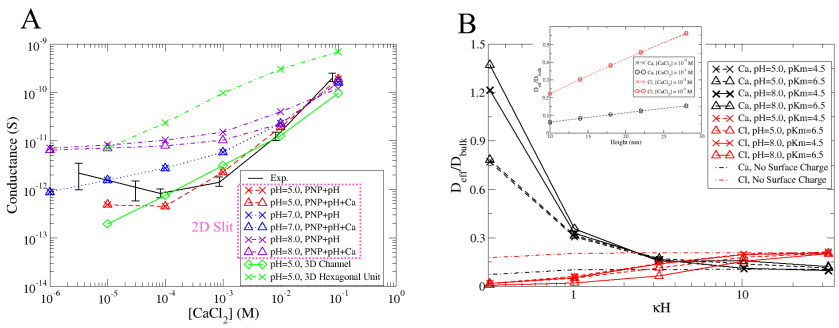


Figure 2.5: Comparison of $CaCl_2$ ionic conductance between numerical and experimental data. (A) Experimental and numerical conductance as a function of [$CaCl_2$]. PNP+pH: pH regulated surface charge density. PNP+pH+Ca: pH regulated surface charge density plus Ca^{2+} adsorption. (B) Effective diffusion constant of Ca^{2+} and Cl^- versus [$CaCl_2$] (expressed as κH, $H = 18$ nm is the height of nanochannel/nanoslit) under different Ca^{2+} adsorption constants and pHs. The inset denotes the effective diffusion constant at varying nanoslit heights when no surface charge density present

Apparent from these data is that Ca^{2+} adsorption has an insignificant contribution to the conductance, except at basic pHs. Accounting for Ca^{2+} adsorption decreases the conductance modestly relative to the basic PNP+pH model. To rationalize this behavior, in Fig. 2.5B we present effective diffusion constants for Ca^{2+} and Cl^- as functions of ionic strength, pH and adsorption equilibrium constants. At high ionic strength, D_{eff}s for both ions approach limits predicted for neutral pores, with Cl^- presenting 20% of its bulk diffusion rate compared to about 10% for Ca^{2+}. As the ionic strength is reduced, the Cl^- D_{eff} declines to nearly negligible values at [$CaCl_2$]=1×10^{-6} M, while that of Ca^{2+} increases by nearly eight-fold, hence in this regime the current predominantly arises due to cation flux. Further, these trends are modestly attenuated as pK_m is reduced at basic pH to reflect increased Ca^{2+} adsorption, as the adsorbed Ca^{2+} ions partially neutralize the attractive, negative silica wall surface charge that would otherwise facilitate Ca^{2+} diffusion. Under acidic conditions, this distinction is lost as there are fewer sites available to accommodate Ca^{2+} relative to protons. Hence, the drop in conductance shown in Fig. 2.5 could be attributed to a reduction in the Ca^{2+} D_{eff} as Ca^{2+} surface binding increases. We emphasize here that our model of Ca^{2+} adsorption only modulates surface charge and therefore neglects changes to the free Ca^{2+} in the channel interior.

Small charged molecule permeation properties of a mesoporous silica film

Validation of the PNP model for 5(6)-Carboxyfluorescein (CF) permeation in a silica film

In the previous sections, implementations of the PNP equations were validated against experiment and simulation studies for several nanoporous silicate geometries. Here we utilize PNP simulations to explore permeation properties of small ligands in a oriented mesoporous film,[152] for which transmission electron microscopy (TEM) provides a distribution of bulk and defect features (Fig. 2.3b) A key difference between these geometries and those considered in the prior sections is that the porous silica membranes here present a multitude of densely packed pores. We first validate our model by simulating permeation properties of 5(6)-Carboxyfluorescein (CF) in the Wooten et al[152] membranes. Based on their experimental setup, the length (thickness of membrane) and radius of the nanopore is set as 90nm and 5nm, respectively. The pH is set as 7.4 (all CF numerical simulations were conducted at pH=7.4 unless otherwise stated) and background [KCl] is set as 0.9M to mimic the ionic strength (0.9M, κD=31.8) in experimental setup. At this basic pH, CF is expected to form the anion CF^- (pK_a=6.5) and thus we do not anticipate adsorption of the small molecule. Our model consists of five ionic species H^+,OH^-,Cl^-,K^+ and CF^-, for which we assume an absorbing boundary condition for CF on one reservoir edge (e.g. [CF]=0).

As shown in Fig. 2.3B, although the nanopores are mainly in hexagonal packing arrangement, structural defects such as fused nanopores are evident. To resemble the real membrane composition, we thus consider two unit cell structures appeared in the TEM image: a hexagonal unit cell and a fused pore unit cell (dimensions are shown as in Fig. 2.2). The permeabilities of CF for these two unit structures were predicted to be 1.12×10^{-6} m/s and 1.185×10^{-6} m/s, respectively (see Fig. S3) and modestly less than the experimentally-determined value of approximately 1.4×10^{-6} m/s.[152] Not unexpectedly, the difference in the unit cell permeabilities can be explained by the relative porosities (see Fig. S4), with the fused pore have a greater cross-sectional area than the HCP cell (0.357 and 0.317, respectively). Unlike in Fig. S3, for which we used a background ionic strength as 0.9M to mimic the experimental setup in Wooten et al,[152] here we used a range of dilute ionic strengths. The motivation for dilute conditions is two-fold: 1) high ionic strength has a strong electrostatic screening effect which undermines the influence of wall electric potential on CF permeation and 2) it has been shown surface charge density/electric potential begins to dominate ionic transport when bulk concentration is smaller that 1 mM.[161] As shown in Fig. 2.6, for a fixed surface potential, both hexagonal and fused pore unit cells present larger CF permeabilities as ionic strength is increased. The increase with $\log(\kappa D)$ plateaus when the Debye length is much small smaller than D, under which case the permeability is nearly completely determined by the pore diameter. These permeabilities indicate

that the higher background ionic strength attenuates the repulsive interaction between anionic CF and the negatively-charged nanopore wall, thus leading a larger CF permeation. Analogously, for a given ionic strength, decreasing the magnitude of the electric potential permits greater CF permeability. These effects become more apparent as ionic strength is reduced, and moreover, the differences in permeabilities for the respective unit cells magnify.

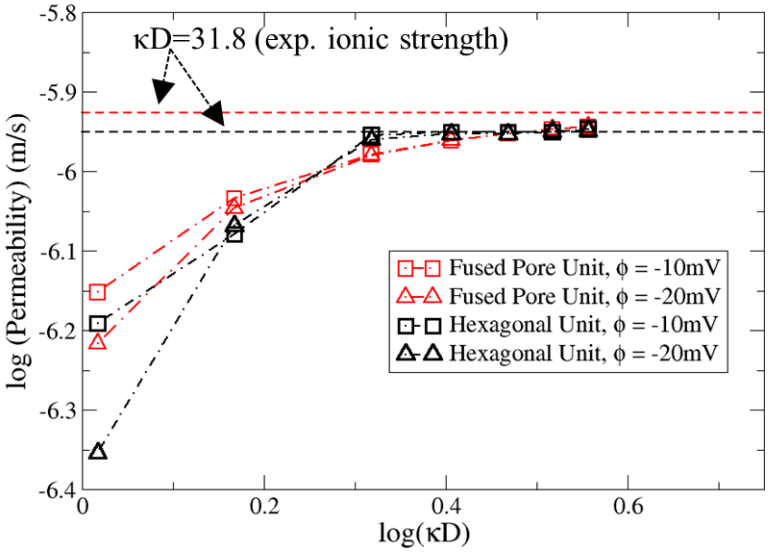


Figure 2.6: Permeability of CF in hexagonal and fused pore unit cells vs. κD (D=10nm is the diameter of pore) under different wall electric potentials when pH=7.4 and [CF]=1mM. The two straight dashed lines denote predicted CF permeabilities under 900mM background ionic strength (κD = 31.18, see Fig. S3 for details)

In support of the final step of the work flow in Fig. 2.1, we interpolate these permeability predictions for the fused pore ($6.2 \times 10^{-1}\,\mu m\,s^{-1}$) and the bulk HCP pore ($4.5 \times 10^{-1}\,\mu m\,s^{-1}$) unit cells onto the matched filtered data from Fig. 2.3. Since the matched filtering provides a quantitative means of estimating the propensity of detected defects relative to bulk, the effective permeability of the entire EM film can be approximated by surface area-weighted conductivities of the corresponding unit cells. While in principle the conductivities between adjacent unit cells could be coupled and thus disfavor this simple extrapolation approach, given the modest surface potentials we assumed that the conductivity of a given unit cell was independent of its neighbors. In Table 2.1 we summarize these data for the fused pore-rich and bulk-like regions. The fused pore-rich region presents roughly 90% fractional surface area attributed to fused pore unit cells, thus the P_{eff} of $6 \times 10^{-1}\,\mu m\,s^{-1}$ approaches P_{eff} predicted for a single fused pore. For the bulk-like region, the surface area was split between bulk-like, fused pore, and uncharacterized unit cells, and thus yielded an intermediate P_{eff} of $5.5\,\mu m\,s^{-1}$. Surprisingly, the bulk-like unit cells did not contribute the majority of the surface, but instead, a significant percentage was unclassified despite resembling the bulk HCP form by visual inspection. Upon investigating the correlation outputs corresponding to the unclassified cells, we found that relative spacing between pores were dilated, which arguably led to a reduced

Table 2.1: Effective permeabilities, P_{eff}, for the fused pore-rich and bulk-like regions, as computed by Eq. 2.25. Predicted conductivities for bulk and fused unit cells are from Fig. 2.6 and their surface area weights are estimated from matched filter detections in Fig. 2.3

Case	S.A. (%)			P_{eff}
	Bulk	Fused	Uncharacterized	
Fused pore-rich	0.05	88.9	0.06	$6.0 \times 10^{-1}\,\mu m\,s^{-1}$
Bulk-like	0.28	0.44	0.28	$5.5 \times 10^{-1}\,\mu m\,s^{-1}$
Exp				$1.4\,\mu m\,s^{-1}$

overlap of the HCP pore filter. We anticipate that augmenting the rotated filter bank with dilations and contractions of the 'typical' unit cell could potentially improve detection for such cases.

Exploiting surface interactions to tune permeation rates

Factors controlling CF permeation In this section, we examine how CF permeation can be controlled by ionic strength, nanopore wall surface charge density, electric potential and selective CF binding. In Fig. 2.7 we demonstrate ion permeabilities as a function of ion size and ion/wall electrostatic interaction energy. Not surprisingly, the permeability decreases with effective pore radius, which represents the difference between the actual pore radius and that radius of a permeant ion (e.g. $r_{\text{eff}} = r_{\text{pore}} - r_{\text{ion}}$). In other words, as the ion size is increased, the effective pore radius decreases and smaller permeabilities result. We note that as the ion size approaches that of the pore, additional factors would likely have to be added to the model to accurately model the ion/wall interactions, including hydrodynamic interactions and potential changes in local diffusivity.[173] Similarly, increasing the repulsive interaction energy decreases permeability, while attractive energies increase permeation. The effects of electrostatic interactions attenuate with increasing ionic strength, as shown for $CaCl_2$ effective diffusion rates in Fig. 2.5b.

As shown in Fig. 2.7B, at modest ionic strengths (100 mM), the capacity to modulate ion transport by charge is diminished, as the predicted permeabilities at 100 mM are constant across the range of electrostatic interaction energies considered, compared to the 1 mM data. If instead one modified the surface chemistry to selectively bind a substrate, the loss in electrostatically-driven selectivity commonly observed at higher loadings (ionic strength)[174] could be circumvented. We illustrate this by assuming there exists a buffer that selectively binds CF, which in principle could arise through adding appropriate functional groups to the silica surface or tethering a binding agent to the pore wall.[175,176] For simplicity, we assume rapid equilibrium for this potential reaction, that is, the binding/dissociation between CF and buffer occur more

rapidly than the timescale of diffusion. Under this limit, the local diffusivity of CF in the buffer region (D_{buffer}) can be described by:[177]

$$D_{buffer} = D_{free}\left(1 + \frac{K_s[B]}{(K_s + [CF])^2}\right)^{-1} \tag{2.26}$$

where K_s is the dissociation constant, $[B]$ is concentration of buffer and D_{free} is the diffusion constant without any buffer. We model the effects of this rapid equilibrium by altering the local diffusion constant used in the PNP equation (Eq. 2.8), which effectively reduces the net flux with increasing buffer concentration and CF binding affinity. As shown in Fig. 2.7B, a buffer concentration [B]= 1mM (comparable to [CF]), the permeability is reduced by 6.27% compared to the absence of buffer. As [B] is increased to 10mM and 100mM, the permeability is further reduced to 27.6% and 39.0%, respectively. Overall, these data suggest that considerable flexibility in tuning permeation properties in these mesoporous films may be realized, though exploiting combinations of altered pore sizes,[178,179] tuning ionic strength/loading rates,[178] or introducing ion-specific surface chemistry.[176,180]

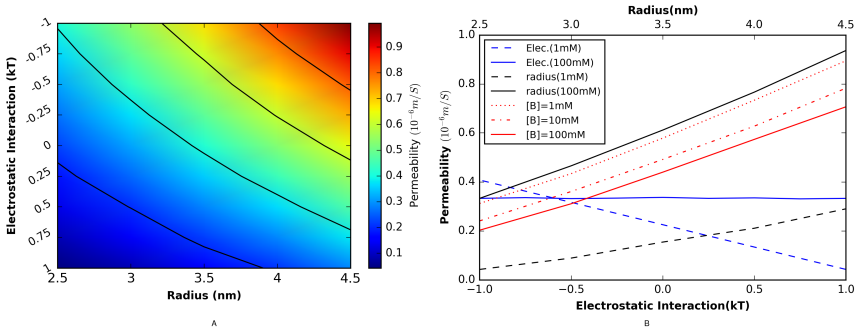


Figure 2.7: A) Numerically simulated CF permeability in hexagonal unit cell as a function of electrostatic interaction energies and pore radiis. Background ionic strength and bulk [CF] are both set as 1mM. B) CF permeability at fixed pore radii (2.5nm) and fixed electrostatic interaction energy (1kT) under different background ionic strength. Red lines denote CF permeability when a 1nm thick buffer layer present along the inner wall of pore, the local diffusion constant of CF is given as Eq. 2.26 where K_s is assumed to be 1. The dots, dashed and solid red lines depict buffer concentration [B]=1mM, 10mM and 100mM, respectively.

2.4 Conclusions

In this study, we have developed and explored the utility of a workflow for automatically characterizing EM microscopy data of mesoporous silica films and performing detailed ion transport simulations of both bulk and defect mesocrystalline features. A primary innovation in this method is the use of a matched filter approach to robustly detect known features in electron microscopy data. The procedure relies of a bank of filters for which the EM signature of a given defect is known, as well as rotated versions of those filters. Our results demonstrate that both bulk and fused features could be determined with high selectivity. We emphasize that our matched filtering approach is rather simple, but can be easily extended to incorporate sundry developments in matched filter theory to improve the accuracy and reliability of detecting diverse signals in EM data.

To simulate ion transport in these silica-based materials, we implemented a PNP model that includes pH and Ca^{2+} adsorption regulation of pore charge density. Our models confirm that electrolyte conductance in mesoporous silica films is dependent on ionic strength, surface charge, pH, and adsorption kinetics, as has been already demonstrated for single nanoporous channels and slits. Under some conditions, the effects of ionic strength on transport can be described as an effective pore radius that reflects the solution Debye length. We extended these simulations to include a small molecule, 5(6)-Carboxyfluorescein (CF), and identified parameters under which CF transport could be optimized

through variation of porosity, surface charge and selective CF binding. In our approach, we assumed a rapidly-equilibrating buffering zone, that notably reduced the apparent diffusion coefficient of the selected-for ion, thus leading to smaller transport properties. Lastly, we leveraged simulation results of representative unit cells derived from the EM microscopy data to estimate effective CF permeation rate for the mesoporous silica membrane surface. Overall, the workflow in Fig. 2.1 that we establish in this study and validate, where possible, has strong potential to benefit the characterization of effective transport properties in increasingly complex composite materials, including those with hierarchical degrees of structures, and especially materials significant incidents of defects that perturb substrate diffusion.

Chapter 3 Electrostatic control of calcineurin's intrinsically-disordered regulatory domain binding to calmodulin

- This chapter is based on "Sun, B.; Cook, EC.; Creamer, TP.; Kekenes-Huskey, PM. *Biochimica et Biophysica Acta - General Subjects* **2018,** *1862,* 2651–2659"

3.1 Introduction

Calcineurin (CaN) is a ubiquitously expressed protein that regulates myriad developmental and signaling processes.[29,30] It is chiefly regulated by Calmodulin (CaM), one of the most prolific proteins in terms of its role in shaping intracellular signal transduction cascades. Despite the fundamental importance of CaM-regulated CaN phosphatase activity in organism physiology, the molecular mechanisms governing this process are incompletely understood. Given that CaM/CaN is a prototypical example of a protein/protein complex involving a globular protein (CaM) and an intrinsically disordered binding domain (CaN),[181,182] structural details of the protein/protein complex are restricted to intact CaM bound to a small fragment of the CaN regulatory peptide. In this regard, the CaM/CaN complex is similar to the tens of CaM/protein target complexes[183] that have resisted structure determination methods beyond the binding of short peptides. Remarkably, despite the CaN regulatory domain presenting little stabilized secondary structure, the CaM/CaN complex binds with picomolar affinity,[184] afforded in part by rapid, diffusion-limited binding.

CaN is heterodimeric protein consisting of two domains: chain A (57-61 kDa) and chain B (19 kDa),[30,185] while CaM (17 kDa) is comprised of two alpha-helix rich domains capable of binding Ca^{2+}. At Ca^{2+} concentrations typical of resting cells (50 to 100 nM),[186] CaN phosphatase activity is negligible, while CaM is believed to be in Ca^{2+}-free state.[187] Under these conditions, the CaN catalytic domain is autoinhibited by the protein's auto-inhibitory domain (AID). At rising Ca^{2+} concentrations, the CaN AID is removed from the catalytic domain. CaM binding to the AID-containing CaN regulatory domain (RD) (Ser373 to Thr468) is a critical determinant of this process.[182,188] Interestingly, like many IDP-containing complexes, well-defined secondary structure is observed only upon binding a protein target.[31,36–38]

In absence of hydrophobic residues[189] that would otherwise promote collapse of protein into a molten apolar core, many IDPs such as the CaN regulatory domain (RD) are polyampholytic.[190] Metrics like the net charge per residue (NCPR) have been proposed to relate charge density in IDPs to ensemble properties including compactness and shape.[42,43] Formally, NCPR$= |f_+ - f_-|$ where f_+ and f_- are fractions of positively and negatively charged residues, respectively, and fraction of charged residues (FCR) is calculated as $\frac{N_c}{N_t}$ where N_c is the number of charged residues and N_t is the total number of residues.

Generally, IDPs with large NCPR values (> 0.25) tend to adopt more extended conformations due to repulsive intra-molecular electrostatic interactions.[42,191] However, the distribution of charged residues in the CaN RD is heterogeneous, thus such sequence-dependent metrics have limited utility in determining localized properties, such as the availability of binding motifs to target proteins.

We hypothesized therefore that RD sequence charge composition (as measured by NCPR) and ionic strength influence the dynamic availability of conformations amenable to CaM binding, while long-range electrostatic interactions drive diffusion-limited association (see Fig. 3.1). To investigate this hypothesis, we utilize long-timescale MD simulations to probe the highly dynamic conformational ensembles comprising the RD constructs, toward delineating the extent to which conformational gating kinetics and long-range electrostatic interactions govern IDP/protein association. A chief outcome of this work is that charge-sensitive 'local' and long range factors, namely IDP conformational dynamics and IDP/target electrostatic interactions, can jointly facilitate diffusion-limited target association.

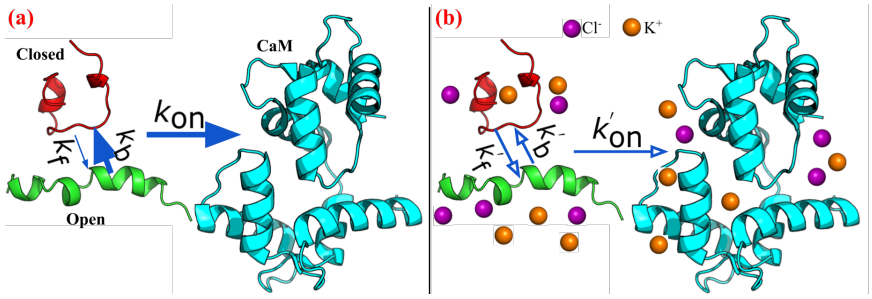


Figure 3.1: Schematic of CaN peptide binding to CaM. The green and red peptides represent CaM conformations which are capable and incapable of binding to CaM (colored in cyan). The association rates between intrinsically-disordered CaN peptides and CaM are controlled by open/closed state gating kinetics (depicted by k_{b} and k_{f}) and the CaM/CaN diffusional encounter rate (depicted by k_{on}). Our study demonstrates that low (a) and high (b) ionic strengths influence both contributions to afford diffusion limited encounters

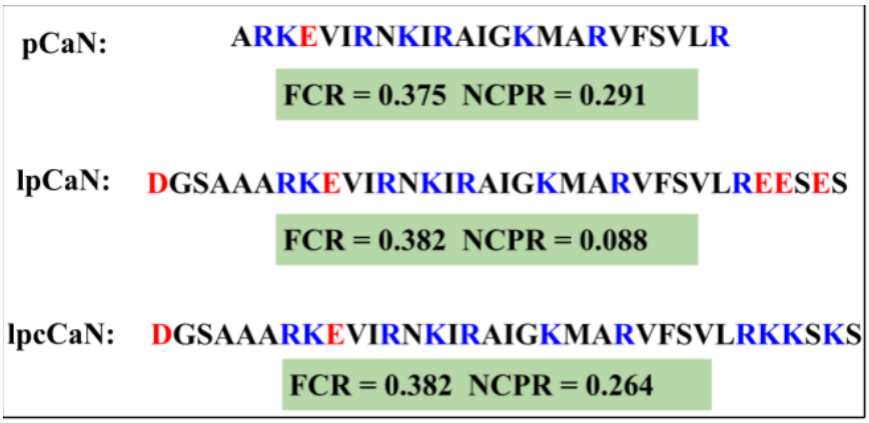


Figure 3.2: Amino acid sequences of three CaN peptide constructs examined here, including their respective FCR and NCPR scores. pCaN: native CaM binding region (CaMBR) of CaN (Ala391-Arg414). lpCaN: five predominantly negatively charged residues affixed to the pCaN termini. lpcCaN: lpCaN construct with three positively-charged substitutions at the C-terminus. In present study, the residue numbers of lpCaN and lpcCaN are counted from 1 to 34 while pCaN has the residue numbers from 6 to 29

3.2 Methods

The initial structures of three CaN RD constructs predicted by Rosetta[192] (see Sect. 3.2) were followed by extensive microsecond scaled MD via Amber14[193] (see Sect. 3.2). The 2D replica-exchange umbrella sampling (REUS) potential of mean force (PMF) calculations of CaN RD constructs were performed by NAMD2.11[194] (see Sect. 3.2). The MD trajectories were used to characterize the conformational dynamic of CaN RD constructs through Markov sate model (MSM) via Aqualab[195] (see Sect. 3.2). The diffusional encounter rates of CaN RD constructs with CaM were simulated by BrownDye package[85] (see Sect. 3.2) and the effective association rates with conformational dynamics taken into account were given in Sect. 3.2.

Structure preparation

The N-domain (residue numbers 3 to 75) and C-domain (residue numbers 76 to 147) of CaN were extracted from the crystal structure (PDB ID: 4Q5U[196]). For CaN peptides, three different peptides with varying lengths and charge distributions were considered: 1) pCaN: native binding region for CaM. 2) lpCaN: elongated pCaN with five additional residues added to two ends of pCaN, respectively. 3) lpcCaN: charge mutated lpCaN having EESE to KKSK mutations at the C-terminal end. Since diffusion limited binding between CaM and CaN are suggested for both intact CaN and its regulatory domain

(arXiv:1611.04080v1), we postulated that these comparatively shorter constructs could capture key factors governing diffusion limited association in a computationally tractable manner. Rosetta[192] was used to model initial conformations for the CaN peptides. The *ab initio* structural prediction was conducted by running the "AbinitioRelax.linuxgccrelease" installed on our local computing resources. The parameters used in present study are similar to that listed in.[197] The example flags set (parameters) of pCaN structural prediction are provided in the supplement (see Sect. 6). The fragment libraries (e.g., frag3 and frag9) were generated via the online server (http://robetta.bakerlab.org/fragmentsubmit.jsp). The number of output conformations was set to ten. No pre-defined secondary structure file was specified. According to the energy score, for each CaN peptide, the conformation with lowest energy was picked out for further extensive MD sampling. Although just one conformations was selected for each CaN peptide, it was expected that the following microsecond MD ensure adequate sampling.

Molecular dynamic simulation

We next performed MD simulations to extensively explore the conformational space of the CaN peptides. The Amber ff99SB-ILDN[198] force field was chosen, given its improved recapitulation of experimentally-observed IDP ensembles, in contrast to common forcefields that tend to predict overly collapsed states for IDPs.[199] The MD was performed by using Amber14.[193] The implicit solvent model (igb = 2 with salt concentration = 0.15 M) was used. The reason for choosing implicit solvent model was as following: 1) According to a recent study,[200] a combination of ff99SB-ILDN with implicit solvent model could achieve reasonably accurate sampling for IDPs; 2) Implicit solvent model would enable more sufficient sampling with affordable time cost. The cutoff value for non-bond interactions was set as 999 Å. The starting structure was first subjected to 50000 steps of energy minimization. The minimized structure was slowly heated from 1 to 298.15 K by using the Berendsen Thermostat within 800 ps. During the MD process, the time interval was set to 2 fs and the SHAKE[201] constraints were applied on bonds involving hydrogen atoms. By setting the initial temperature in the heating stage equal to 1 instead of 0 and ig = -1 would generate different initial velocity distributions for the system, thus independent simulations can be achieved. For each peptide, three independent MDs were performed to ensure the reliability of the sampling (total 15 μs production run for each peptide). To study the effect of ionic strength on sampling, we ran analogous simulation with salt concentration = 1.5 M, resulting in a total 30 μs production run for each peptide (15 μs at 0.15 M and 15 μs at 1.5 M ionic strength). Although 1.5 M ionic strength is non-physiological, however, this high ionic strength would significantly increase the electrostatic-screening effect and facilitates more pronounced changes in CaN peptide conformational properties under a different electrostatic environment as compared to the physiological 0.15 M ionic strength. In RMSF calculations,

for each peptide, we first performed rms-fitting of trajectory frames to the first frame of trajectory. The rms-fitting would eliminate the translational and rotational effect and ensure the RMSF reflect the fluctuation of atoms. Contact map data of heavy atoms was collected via CPPTRAJ in Amber with distance cutoff as 7 Å and only residue pairs which are at least 5 residues apart (i and $i+5$) in sequence are considered.

Two-dimensional replica-exchange umbrella sampling (REUS) PMF calculation

Two-dimensional PMF calculations were performed to characterize the free energy profile associated with conformational space of each peptide. Two reaction coordinate (RC)s were defined: 1) α which describes the α-helical content of the peptide (ranging from 0.1 to 0.9) and 2) radius of gyration (RG) of the peptide (ranging from 5 to 32 Å). Each RC range was divided into nine bins resulting in total 81 windows (with interval being 0.1 and 3Å for α and RG, respectively). The two force constants of the harmonic potentials imposed on these two RCs are $1.000 \times 10^3\,\mathrm{kcal\,mol^{-1}\,U^{-2}}$ for α and $2.5\,\mathrm{kcal\,mol^{-1}\,Å^{-2}}$ for RG. For each peptide, the representative structure from the most populated cluster was chosen as the starting structure. NAMD2.11[194] was chosen to perform the 2D REUS calculations due to it's *colvar* module which supports various user-defined collective variables. The CHARMM36[202,203] force field was used in the 2D REUS calculations. For each window, the simulation length was set to 20 ns and only the last 15 ns data was used to calculate free energy by WHAM.[204]

Markov sate model (MSM) analysis via Aqualab

For each peptide, a 1D kinetic trajectory was created from the 15 μs MD trajectory describing the state change along simulation time. Open states were defined based on examining Browndye-predicted association rates as a function of root mean squared deviations (RMSD); the RMSD value below which the association rate rapidly increased above negligible values was utilized as the open/closed state cutoff criterion. We note the 'closed states' are not necessarily precluded from binding, but we assume that the timescale for assuming a conformation compatible with the complex via induced fit is slow. These criteria differed among the three peptides. For pCaN, the open state was defined as RMSD $<$ 7 Å while for lpCaN and lpcCaN the open state are defined as RMSD $<$ 5 Å. Using these cutoffs, Markovian networks were created based on the 1D trajectory (rates of open to closed and vice versa) via Aqualab,[195] for which conditions such as detailed balance[205] were imposed to define $\boldsymbol{P}$, the equilibrium probability matrix and $\boldsymbol{T}$, the transition probability matrix.

Brownian dynamic (BD) simulations

The binding of CaN peptide and N/C terminal domains of CaM are treated as two independent events and simulated separately by using the BrownDye package.[85] For each peptide, ten conformations for each RMSD cluster were randomly selected to perform BD simulations with N/C-domain of CaM. The PDB2PQR[206] was first used to generate the pqr files for CaM N/C domains and the selected conformations of CaN peptides from MD trajectory with radii and point charge parameters adapted from the AMBER99[87] force field. The generated pqr files were then passed into APBS[86] to evaluate the electrostatic potential of these structures. APBS was used to numerically solve the linearized Poisson-Boltzmann equation assuming an ionic strength of 0.15 M and 1.5 M NaCl:

$$-\epsilon\nabla^2\psi = \sum \rho_i q_i - \kappa^2\psi \tag{3.1}$$

where ψ is the electrostatic potential, $\rho_i q_i$ is the charge distribution of fixed charge i, and κ is the inverse of Debye length. The Debye length reflects the scale over which mobile charges could screen out electric potential fields.

In present BD simulation, the reaction criterion was chosen to be six pairs of contacts with distance of contact being less than 10 Å. The contact list was created via the make_rxn_pairs routine in Browndye package based on the pCaN-CaM complex crystal structure (PDB ID: 4Q5U[196]) with distance cutoff being 5 Å. 10000 single trajectory simulations for each system were conducted on 10 parallel processors using nam-simulation. Thus for each peptide, the total number of BD trajectories was about 1 million. The reaction rate constants were calculated using compute-rate-constant from the BrownDye package.

To estimate the association rate and its sensitivity to ionic strength, we computed association rates for the terminal domains separately, assuming that both components bind independently,

$$\frac{1}{k_{\text{on}}} = \frac{1}{k_{Cterm}} + \frac{1}{k_{Nterm}} \tag{3.2}$$

where the rates in the right hand side correspond to the association rates for the C and N terminal domains, respectively. We anticipate that this expression under-approximates the rate of complex formation, given that tethered binding partners generally exhibit higher efficiencies for forming intact complexes.[207,208]

Effective association rate combined with gating kinetics

The effective association rate constant after taking conformational dynamics into account was given by Szabo et al:[45]

$$k_{\text{eff}} = \frac{K_D K_{eq} k_{\text{b}} Z[k_{\text{f}} + k_{\text{b}}]}{k_{\text{f}}(K_{eq} + K_D Z[k_{\text{f}} + k_{\text{b}}]) + k_{\text{b}} Z[k_{\text{f}} + k_{\text{b}}](K_D + K_{eq})} \tag{3.3}$$

where

$$Z[k_{\text{f}} + k_{\text{b}}] = 1 + ((k_{\text{f}} + k_{\text{b}})R^2/D)^{0.5} \tag{3.4}$$

where K_D is the association rate when the peptide is always in open state and in present study, K_D is the BD simulated association rate constant of the open state CaN peptides with CaM (e.g., $K_D = k_{on,open}$). K_{eq} is characteristic constant indicating the extent to which the association is diffusion-controlled (see[45] for more details). In present study, we set $K_{eq} = 1 \times 10^{20}\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$. k_f and k_b are the conversion rate between the open and closed state determined from MSM. R is the contact distance at which the transient complex formed and in present study we set R equal to the average b-radius values from BD simulations. D is the relative translational diffusional constant and was calculated via.[39,85]

$$D = \frac{K_D}{4\pi R f_\infty} \tag{3.5}$$

where f_∞ is the reaction probability which was at the order of 1×10^{-4} given by BD simulations.

3.3 Results and Discussion

Molecular simulations confirm the intrinsically-disordered structure of the CaN regulatory domain

Several studies using circular dichroism spectroscopy, hydrogen-deuterium exchange mass spectrometry, Fourier transform infrared spectroscopy and X-ray crystallography indicate that the nearly one-hundred amino acids of the CaN RD domain (Ser373 to Thr468[181,182]) form an intrinsically disordered ensemble.[32,34,182,209] Of these, approximately twenty amino acids (Ala391-Arg414) comprising the CaMBR adopt an alpha-helix in the presence of CaM.[196] Here we examine three RD constructs (pCaN, lpCaN and lpcCaN, see Fig. 3.2) that present diffusion-limited association with CaM (arXiv:1611.04080v1). pCaN spans residues A391 to R414 and has been co-crystallized with CaM (PDB ID: 4Q5U[196]). lpCaN includes three additional polar residues (Ser3, Ser32 and Ser34) and four additional charged residues (Asp1, Glu30, Glu31 and Glu33) while lpcCaN is a peptide of the same length with substitutions of three glutamic acids with lysines at the C-termini of lpCaN. The set of constructs considered here span a range of charge densities that we later demonstrate tune CaM/CaN association kinetics. pCaN and lpcCaN have similar NCPR values of 0.291 and 0.264, which are considerably larger than the value for lpCaN (0.088); previous works,[42,191] suggest NCPR scores above 0.25 reflect extended IDP conformations given the propensity for repulsive intramolecular interactions, whereas those below this threshold are comparatively compact. We expected therefore that 1) the CaN peptides lack well-resolved secondary structure characteristic of a folded protein and 2) the ensemble of lpCaN should be modestly more compact than that of lpcCaN, given that latter has higher charge density.

To investigate the hypothesis, we performed 5 μs MD simulations in triplicate (total 15 μs) at 0.15 M and 1.5 M ionic strength, respectively. The choice

of physiological (0.15 M) and high ionic strength was intended to probe the contribution of intra-peptide electrostatic interactions to ensemble properties, as such interactions would be screened at high ionic strength. While simulations of IDPs of up to 100 residues have been reported elsewhere,[191] the breadth of simulations used in this study restricted our construct sizes to 24 to 34 a.a. Our MD simulations indicate that the heavy atom root mean squared fluctuations (RMSF) for each residue in Fig. S5(e-f) are shown to be larger than 5 Å for all three peptides at both ionic strengths, which is consistent with the high mobility loop scores reported in Fig. S5(b). These data suggest that the peptides do not form stable folded structures in solution regardless of ionic strength. The MD-generated structures present a multitude of conformations, ranging from loosely-formed, hairpin-like configurations to extended structures. lpCaN presents perhaps the most hairpin character, as corroborated by intramolecular contacts reported through contact map analyses in Fig. S6. Among these contacts are prominent interactions between Arg12-Glu30, Arg23-Glu30 and Arg13-Glu32, which we attribute to transient salt-bridge formation. For lpcCaN, the mutation of negative residues (Glu30, Glu31 and Glu33) to the positively charged residues (Lys30, Lys31 and Lys33) appears to disrupt these intramolecular contacts, thereby yielding a more extended conformational ensemble relative to lpCaN. Given the similar NCPR values of pCaN and lpcCaN, we expected pCaN would similarly present fewer intramolecular contacts than lpCaN. Surprisingly, pCaN has similar contact map features as lpCaN, that is, both peptides have comparable intra-contacts. Later we will demonstrate that it is in fact the interconversion kinetics, not the average structural properties, of lpCaN and pCaN that dictate binding kinetics. Additionally, we found that increasing ionic strength to 1.5 M screens the electrostatic interactions between residues comprising the reported salt bridges. As a result, we observe for pCaN and lpcCaN that the structures become modestly more extended on average. It is important to acknowledge that while implicit solvent simulations permit greater degrees of conformation sampling compared to their all-atom counterparts,[210] the utilization of an implicit solvent comes with certain limitations. Among these include the propensity to overestimate alpha-helical character and to alter the interaction strengths of salt-bridging amino acid pairs.[211] Similarly, in the event that charged amino acids might chelate counterions in solution, as exemplified by acidic EF-hands in diverse Ca^{2+} binding proteins,[212] implicit solvent may underestimate the strength of such interactions. It would therefore be of interest in future studies to assess the significance of these limitations in IDPs using explicit all-atom simulations of comparable length to implicit solvent trajectories.

RD fragment charge density appears to control compactness independent of ionic strength

To support the formation of the CaN/CaM PPI, the CaN CaMBR must be revealed to the solvent-exposed CaM surface. The exposure of the CaMBR could

occur spontaneously, which would promote binding by presenting mutually compatible conformations independent of the complementary species, or via an induced-fit mechanism when the binding partner, CaM, is present. In the previous section, we indicated that the peptides have considerable structural variability, therefore here we determine whether this variability confers greater access to the CaMBR *independent* of CaM.

In Fig. 3.3(a-c), we report the RMSD of the CaMBR binding region for each configuration from the MD simulations, relative to the extended, alpha-helical pCaN conformation that is compatible with the CaM binding surface. From these simulations, we identify conformations that are amenable for CaM binding ("open" state) and those unsuitable for CaM binding ("closed" state), using a cutoff of pCaN: 7 Å, lpCaN and lpcCaN: 5 Å. We utilized a more restrictive criterion for the longer constructs, as the 7 Å cutoff assumed for pCaN yielded structures that were incompatible with CaM. RMSD values below the cutoff more closely resemble the fully-extended reference structure, whereas values above this cutoff are more compact. As shown in Fig. 3.3(a-c), all three peptides adopt a small percentage of CaM-compatible configurations as measured by RMSD and the percentages appear to be insensitive to ionic strength. These data additionally indicate that lpCaN (NCPR = 0.088) has the smallest percentage of CaM-campatible structures as assessed by RMSD compared with the bound CaN complex, relative to lpcCaN (NCPR = 0.264) and pCaN (NCPR = 0.291).

In Fig. 3.3(d-f), we present the structures of the most probable conformations based on RMSD clustering analysis. Significantly, each peptide was observed to partially fold into an α-helix, indicating that bound-like 'residual' structures can spontaneously form in the absence of the binding partner, as has been reported for other IDPs.[213–216] The open state probabilities we determined represented a small, but significant fraction (pCaN: $\sim$20%; lpCaN: $\sim$1%; lpcCaN: $\sim$45%) of the conformations sampled. Importantly, these data indicate extended/CaM-compatible conformations form spontaneously and in a charge density (NCPR)-dependent manner (see Fig. 3.3(a-c)). We speculate that the tendency for a percentage of the conformational ensemble to assume an extended pose relative to a hairpin fold suggests that intra-molecular repulsion may partially destabilize the formation of loose hairpins. This effect would be exacerbated with charge densities of increasing magnitude, such as those reflected in the NCPR values for pCaN and lpcCaN, and relatively diminished for low NCPR peptides like lpCaN.

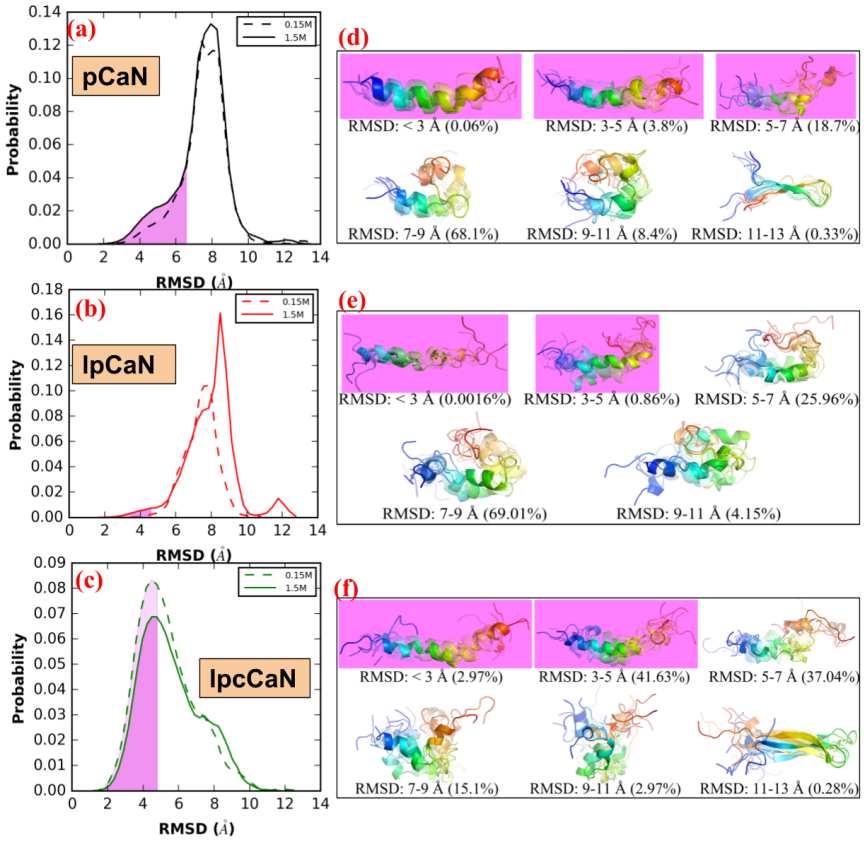


Figure 3.3: Distribution of RMSD (with respect to bound-pCaN crystal structure in PDB 4Q5U) in the MD of each CaN peptide at 0.15 M ionic strength and 1.5M ionic strength, respectively(a-c). The shaded area colored in violet denotes the open state-like conformation. The representative structures (colored in rainbow with N-termini as blue and C-termini as red) for each RMSD range and percentage of conformations within this RMSD range were also shown(d-f)

CaM-compatible conformations are energetically unfavorable in absence of CaM

To establish a thermodynamic basis for the trends of greater conformational diversity for the high NCPR cases (pCaN and lpcCaN) relative to the low NCPR case (lpCaN), we report potential of mean force (PMF) calculations for these peptides as a function of α-helical character, a measure of secondary structure formation, and RG, a measure of compactness (see Fig. 3.4). Such PMFs have been used to characterize the propensity for IDPs to assume specific ensemble characteristics, including IDP compactness.[217,218] Each construct preferentially adopted smaller α-helical character than the 84% reflected in pCaN when bound to CaM. Lacking CaM, unfolded CaN RD states dominate the conformational

distribution, thus CaM is apparently necessary to shift the IDP ensemble toward the bound state via an induced-fit mechanism.

Interestingly, we observe that the range of RG and α-helical values within a few k_bT of the energy minima (0 k_bT) are larger for the high NCPR cases compared to lpCaN. These data mirror our findings for the histogram of RMSD distributions in Fig. 3.3, with the low NCPR case presenting a narrower distribution relative to the high NCPR cases. Further, the PMF data support the observation for the lpCaN and lpcCaN peptides that the former structure assumes a more compact, hairpin-like configuration relative to the latter, as we observed in Fig. 3.3(d-f). This indicates that the high NCPR cases access a larger range of conformations in their IDP ensembles that overlap with the CaM-bound structures, albeit in contrast to the more narrowly peaked distributions presented for the low NCPR (lpCaN) configuration. Our results are consistent with the work done by Mao et al[42] for protamine IDPs demonstrating that globule-to-coil transitions were more favored with increasing of NCPR values.

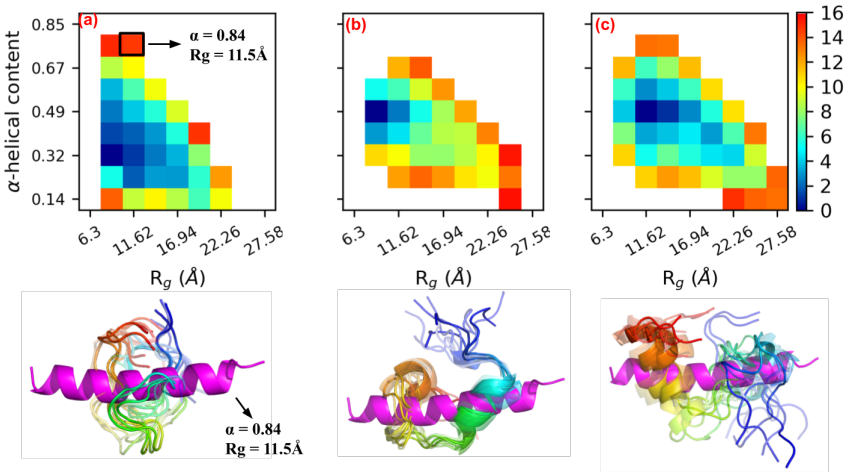


Figure 3.4: Two dimensional PMFs for pCaN (a), lpCaN (b) and lpcCaN (c) at 0.15 M ionic strength. The x and y axis depict α-helical and RG reaction coordinates, respectively. The x axis of (b) and (c) are hidden for clarity. For each peptide, ten randomly selected structures (colored in rainbow with N-termini as blue and C-termini as red) from lowest energy area are compared against bound state pCaN conformation (colored in magenta) from PDB 4Q5U (α = 0.844, RG = 11.54 Å). The unit of color bar is k_bT where k_b is Boltzmann constant and T = 298 K is temperature

CaN regulatory domain ensemble conformational dynamics are rapid and have modest ionic-strength dependence

Our unconstrained MD and PMF calculations both indicate that the CaN RD peptides do not readily assume an open-state compatible with CaM, although there exist some infrequent, CaM-compatible configurations. In this regard, one can view the accessibility of the pre-folded CaMBR domain to CaM as a 'gating' event, which in principle could control the apparent binding rate for this process.[219,220] Given that our previous work (arXiv:1611.04080v1) in which CaN peptides are assumed to have fully CaM-compatible CaMBR conformations demonstrated that all three peptides are capable of binding CaM, albeit at substantially different rates.We hypothesized that the appearance of bound-like structures before binding serves to nucleate loosely-associated CaM-compatible transient encounter states with low alpha-helical character, which permits 'induced folding' in the presence of CaM to access alpha-helix rich bound-states.

CaN peptides present open and closed states with rapid gating kinetics

As a first step towards probing this hypothesis, we first estimated the transition kinetics between open and closed states identified in Sect. 3.3 using Markov sate model (MSM) analysis. Intuitively, we would expect that higher rates of

accessing CaM-compatible open states would maximize the CaM/CaN association rate. We note here that we defined the open state as consisting of conformations below the 7 Å (for pCaN) and 5 Å (for lpCaN and lpcCaN) RMSD cutoff, while all conformations with dissimilar RMSDs were lumped into a single closed state. We verify that the states are essentially Markovian as the correlation times become negligible beyond roughly tens of nanoseconds which is faster than the diffusion encounter time.

Overall, based on this partition of MD data, the transition rates between closed and open states are rapid (the slowest rate is at $1 \times 10^7\,s^{-1}$ order, see Table S2) and lead to the short-lived open states shown in Fig. S7 (average life times of open state for all three peptides under both ionic strengths are around 0.2 ns).

Modulation of electrostatic interactions do not change overall compactness of CaN peptides ensembles, but does influence the gating rates of lpCaN

Ionic strength was shown to have negligible impact on the RMSD of our predicted peptide structures relative to CaM-bound conformation. However, given the pronounced role of electrostatics in facilitating protein/protein association rates and protein folding,[44,221] we sought to determine whether transition kinetics between conformations were influenced by ionic strength. Hence, we compared MSM rate predictions for MD generated structures at low (0.15 M) and high (1.5 M) ionic strength. Here we found that for pCaN and lpcCaN, increasing ionic strength does not affect the gating rates between open and closed states. However, for lpCaN, increasing ionic strength increased k_f from $1 \times 10^7\,s^{-1}$ to $1 \times 10^8\,s^{-1}$ order. As a result the lifetime of its closed states decreased from 12.83 to 4.42 ns, as shown in Fig. S7. Hence, the open and closing kinetics of peptides with high NCPR are appear to be less sensitive to ionic strength, compared to structures with low NCPR. These results concur with findings from Liu et al,[222] for which they demonstrated that the fast-phase structural fluctuations as measured by Fluorescence correlation spectroscopy (FCR) for the IDP Sic1 disappeared with decreasing ionic strength. At first glance, it is surprising that the ionic strength did not appreciably alter the average properties of the conformation ensemble. However, we anticipate that the change in ionic strength, while significant, was insufficient to strongly disfavor the desolvation of the charged groups in order to drive hydrophobic collapse. In contrast, a less hydrophilic solvent, e.g. one with a low dielectric constant, would strongly reduce the desolvation energy and thereby is expected to lead to a hydrophobic collapse.

Long-range interactions promote rapid CaM/CaN association

Our results thus far indicate that the CaN RD peptides adopt CaM-compatible conformations in the absence of CaM frequently, albeit transiently. Here we determine the compatibility of these transient states with the CaM/CaN

binding interface using Brownian dynamic (BD) simulations. Specifically, we sought to evaluate two hypotheses: 1) that frequent presentation of CaN open states promote near diffusion-limited association rates and 2) that long-range electrostatic interactions are exploited in PPIs involving IDPs. Motivating our first hypothesis are recent indications that target-compatible residual structures of the isolated p53 up-regulated modulator of apoptosis (PUMA) IDP form spontaneously as a function of ionic strength and electrolyte composition.[223] For the latter hypothesis, we adopt the paradigm of electrostatically-driven association of globular proteins,[39,224–226] which depends critically on the notion of a transient encounter complex.[227,228] The encounter complex serves as the rate determining step in PPI formation, whereby a protein loosely binds to its protein target, before adopting the fully-formed bound configuration. However, unlike PPIs involving globular partners that typically feature regions of complementarily-charged hydrophilic patches,[229,230] such patches may only be transiently presented in IDPs. Though these results offer specific insights into the ubiquitous Ca^{2+} signaling partners CaM and CaN, these trends may vary in importance depending on the IDP sequence, particularly those with vastly different amino acid charge densities and distributions.

CaN open state conformations compatible with CaM

We tested these hypotheses by assuming each peptide must achieve a minimal number of 'native contacts' with the CaM N-terminal and C-terminal domains to constitute a transient encounter complex. The native contacts are obtained by analyzing the crystal structure of CaM-pCaN complex, in which key interactions between CaM and pCaN were extracted to guide the BD simulations. From this standpoint, the lenient conditions for association are tantamount to the notion of a transient encounter complex,[231,232] which is formed upon association of two binding partners prior to forming the fully-bound complex. Because we test the first hypothesis using conformations generated from the MD simulations *without* CaM, this test bears similarity to the conformational selection paradigm,[233] though we emphasize CaM is likely required to completely form the bound complex from the transient encounter state. However, the abundance of IDP-target protein complexes that exhibit diffusion-limited association suggest that the folding rate from the encounter state to the bound complex is rapid. The MD-generated open states presented in each of the peptide configurations are compatible with both the N- and C-terminal CaM domains to varying degrees, as the open state of each peptide gives BD-simulated k_{on}s in the diffusion-limited regime ($> 1 \times 10^7\,\mathrm{M^{-1}\,s^{-1}}$). Notably, all peptides considered here are capable of forming the transient encounter complex with CaM at rates that decrease with increasing ionic strength. While not explored in this study, binding affinities via methods including Molecular Mechanics-Generalized Born Surface area (MM-GBSA) techniques[234,235] could help rank configurations most compatible with the bound complex.

Rapid IDP conformational ensemble dynamics promote rapid association

Lastly, we investigated the role of conformational gating rates on the effective association rates based on the stochastic gating model postulated by Szabo et al.[45] In this gating model, k_{on} is the expected 'ideal' association rate between CaN peptide and CaM. By 'ideal' it means that the CaN peptide is aways in CaM-compatible open state and the concentration of CaN peptide and CaM are sufficiently large, thus k_{on} here reflect the diffusion timescale of CaN peptide to CaM. By considering this, k_{on} can be compared with k_b/k_f which reflect the conversion time scale between open/closed states of CaN peptide. In the gating model, there are two limits that bound the effective rates: 1) given gating rates that are significantly faster than diffusional encounter rate ($k_f + k_b \gg k_{on}$), the effective association rate k_{eff} is equivalent to the rate associated with the open state, that is, $k_{eff} = k_{on}$. 2) given gating rates significantly smaller than the diffusional encounter rate ($k_f + k_b \ll k_{on}$), k_{eff} is given by the weighted average of the association rates for all accessed states, that is $k_{eff} = \langle k_{on} \rangle$.

Rates associated with intermediate regimes are obtained by evaluating Eq. 3.3 using the MSM-estimated gating rates. Based on the data in Table S3 we show in Fig. 3.5 for pCaN and lpcCaN that k_{eff} and k_{on} are comparable (e.g. $k_{eff}/k_{on} \rightarrow 1$), indicating a marginal effect of conformational gating on the association rate. This arises because the conformational transition rates are of the order $1 \times 10^9\,s^{-1}$, roughly 100 times faster than diffusional encounter rate, based on our BD simulated k_{on}s of $1 \times 10^7\,M^{-1}\,s^{-1}$ order (see Table S3). In contrast, the slower transition kinetics for lpCaN yield a k_{eff} that is about 50% of the maximal k_{on}, albeit it is still in a diffusion-limited regime. Moreover, the rates are strongly attenuated at 1.5 M relative to low ionic strength conditions of 0.15 M, which suggest the strong role of long-range electrostatic interactions in promoting association. These data indicate that diffusion-limited association kinetics are realized in the CaN IDP constructs, though the effective rate depends both on ensemble gating kinetics and long-range electrostatic interactions.

An important factor to consider in this model is that the inability of the closed state to bind as rapid as the open state does not render the former 'unsuitable' for binding. Rather, we anticipate its binding rate will be much slower relative to the open states, as the latter of which requires lesser structural reorganization to assume the correct binding pose. We support this speculation based on Gō model predictions, for which we demonstrated a greater degree of frustration of transition from closed-like conformations to open-like. Namely, for closed state pCaN conformations, the BD-simulated complex presents an average fraction of native contact as $Q_n = 0.24$ and $Q_c = 0.22$, for the percent native contacts of the N and C terminal CaM domains, respectively while the open state pCaN conformations yield, $Q_n = 0.56$ and $Q_c = 0.59$ (see Fig. S13 and Fig. S12). Since lower native contact values will likely encounter greater frustration in folding landscapes compared to near native states, and reduced frustration results in increased rates of folding kinetics,[236] we anticipate that induced-fit binding

occurs more slowly than the binding of open-state conformations.

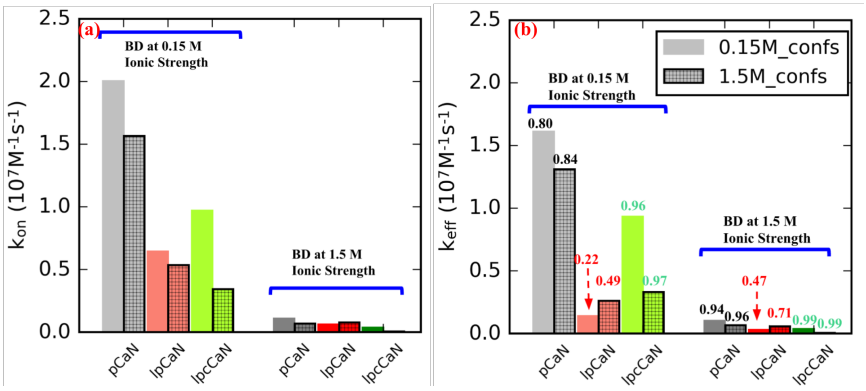


Figure 3.5: Association rate constants between CaN peptide and CaM before (a) and after (b) taking CaN peptide's conformational dynamic (from MSM modeling) into consideration using Eq. 3.3. The bars without grids and with grids depict results in which CaN peptide conformations were sampled at 0.15 M and 1.5 M ionic strength, respectively. In (a), k_{on} was calculated via Eq. 3.2 where k_{Cterm} and K_{Nterm} are the average values of ten randomly selected conformations from each peptide's open state. In (b) the numbers above each bar represent the ratios of k_{eff} to k_{on}

3.4 Conclusions

Modulation of IDP charge density provides molecular basis for tuning IDP ensemble kinetics and protein-protein association kinetics

Our studies of CaN conformational dynamics and CaM/CaN association reveal several interesting features. While the role of charge distribution in IDPs has been shown to be a strong predictor of ensemble structure including compactness,[42,191] our simulations reveal that measures such as NCPR may offer predictive estimates for the ionic strength sensitivity of conformation transition kinetics. Namely, higher NCPR structures are more likely to adopt conformations that complement their binding target, and are less sensitive to changes in ionic strength that may influence gating kinetics. However, it is important to note that this trend may not generalize to necessarily all IDPs, given the wide range of protein/protein association rates ($< 1 \times 10^3$ to $> 1 \times 10^9\,\text{M}^{-1}\,\text{s}^{-1}$[39]) reported in the literature, which hints at the possibility of different assembly mechanisms. Second, we demonstrate that long-range electrostatic interactions can play a paramount role in determining the kinetics of forming PPIs involving intrinsically-disordered partners, while protein-solvent and protein-protein electrostatic interactions govern the kinetics of presenting target-compatible binding motifs. Together, these factors suggest that IDPs can achieve diffusion-limited association by controlling conformational gating, so long as a conformation amenable to association is rapidly sampled. Overall our findings build upon the growing understanding of the roles of conformation

selection and induced fit in dictating PPIs, both identifying how conformational selection can accelerate association, despite potential requirements for induced fitting in order to adopt the final binding pose.

Control of protein-protein association kinetics may support efficiency of CaM regulation despite diverse targets

Our study focused on CaN's binding interaction with CaM, of which the latter regulates a staggering array of eukaryotic signaling cascades through forming PPIs with target protein.[183] What sets CaM apart from other such hubs is the surprisingly diverse variety of targets it regulates, despite presenting a single isoform across all mammalian species[28] . In part, its ability to regulate this diversity is attributed to the conformational heterogeneity of the CaM binding interface[237] it is capable of forming. These findings provide intriguing insight into the interplay between conformational diversity and electrostatically-driven protein-protein association involving CaN, which are likely to extend to wide-ranging processes regulated by intrinsically-disordered proteins. As such, exploiting IDP composition to tune PPI kinetics could offer new tools to probe and modulate important biochemical signal transduction pathways.

Chapter 4 Molecular basis of Calmodulin-dependent Calcineurin Activation: The importance of distal helix/Calmodulin Interaction

- This chapter is based on "Sun, B.; Vaughan, D.; Tikunova, S.; Creamer, TP.; Davis, JP.; Kekenes-Huskey, PM. *Biochemistry* **2019,** *58,* 4070–4085"

4.1 Introduction

Calcineurin (CaN) is a phosphatase that contributes to gene expression in response to changes in Ca^{2+} homeostasis, it plays integral roles in physiological processes including neurological development and maintenance, immune responses and tissue remodeling.[185,238] CaN is heterodimeric protein consisting of two domains: chain A (57-61 kDa) contains the protein's catalytic site, while chain B (19 kDa) contributes to enzyme regulation.[185] CaN is activated by rising intracellular Ca^{2+} levels. While it presents modest catalytic activity in response to Ca^{2+} alone, optimal phosphatase activity occurs upon binding Ca^{2+}-saturated CaM. At depressed Ca^{2+} levels, the enzyme is inhibited by its auto-inhibitory domain (AID) that directly binds to the phosphatase catalytic site. Maximal relief from auto-inhibition occurs upon the binding of CaM to CaN's regulatory domain.

Our current understanding of the protein's activation and enzymatic activity has been shaped by a number of atomic resolution structures of CaN determined by X-ray crystallography[32,239–243] and nuclear magnetic resonance spectroscopy.[244] Of the many CaN structures that have been deposited to the Protein Data Bank are examples that have revealed the protein's auto-inhibited state (PDB ID: 1aui[32]), a potentially non-physiological 2:2 CaM/CaN stoichiometric configuration,[241,245,246] complexes of the enzyme with immunosuppressants[239,242] and transcription factors.[240,243] However, much less is known about the structural basis of CaM-dependent regulation of CaN, as atomic resolution CaM/CaN complexes are limited to intact CaM bound to small peptides comprising the CaM binding region (CaMBR) of the CaN regulatory domain.[33] From those structures, it is clear that the CaMBR assumes α helical secondary structure when bound to CaM. Nevertheless, the paucity of structural information inclusive of complete CaM and CaN proteins leaves critical details of CaM-dependent CaN regulation unresolved.

It is increasingly understood that CaM-dependent CaN activation depends on structural properties of the 95-residue ($\approx$10 kDa) CaN regulatory domain.[35] This segment is intrinsically disordered,[32–35] which signifies that it does not assume a well-defined fold in solution and indirect probes of its conformational properties in the absence and presence of Ca^{2+}-activated CaM have revealed important clues about the mechanism of CaN regulation. It was first observed via circular dichroism (CD) by Rumi-Masante et al that upon CaM binding, nearly fifty residues of the RDs folded into α helices of which only half could

be accounted for by the CaMBR region. By using hydrogen exchange mass spectrometry (HXMS), they further identified a region C-terminal to the CaMBR that formed an α-helix upon CaM binding.[35] Dunlap et al[46] confirmed the observation in a mutagenesis study of that region. Namely, they revealed that single point mutations of three alanines within the distal helix region into glutamic acids disrupted helix formation. Importantly, these mutations reduced CaN's apparent affinity for a substrate, pNPP, that competes with the AID for the CaN active site.[46] This region was coined the 'distal helix' region (DH, residue K441 - I458).

Simulations of CaN have helped bridge experimental probes of its phosphatase activity[185,247,248] with static, atomistic-resolution structural data. Li et al reported conformational changes of the CaN B domain following Ca^{2+} binding via molecular dynamics (MD) simulation and proposed that similarity between the apo- and holo-CaN B-domain conformations enables the former to regulate CaN activity independent of Ca^{2+}.[249] Harish et al utilized virtual screening and MD simulations to design inhibitory peptides of CaN using the native AID peptide as template.[250] Simulations have also been used to study the involvement of CaN residues outside of its catalytic domain in the binding and anchoring of inhibitory immunosuppressant drugs and analogs thereof.[140,251–253] Similarly, computational studies examining structural mechanisms of CaM-dependent regulation of targets including CaN have emerged recently, including myelin basic protein (MBP)[254] and myosin light chain kinase (MLCK).[255,256] In complem-ent to these studies, we have additionally shown via molecular dynamics and Brownian dynamics simulations that the CaMBR is highly dynamic in solution in the absence of CaM, that CaM binding to the CaMBR is diffusion-limited, and that the corresponding association rates are tuned by the charge density of the CaN peptide.[257] Despite these contributions, the sequence of molecular events that follow CaMBR binding and culminate in relief of CaN auto-inhibition remain unresolved.

Observations in[35,46,258] formed the basis of a working model of CaN activation whereby the folding of the intrinsically-disordered distal helix into an α helix-rich structure is coupled to relieving CaN autoinhibition. However, it was still unclear whether the distal helix directly bound to CaM, and if so, where they might share PPI interfaces or how those putative PPIs are stabilized. In large part, the challenge in identifying potential PPI sites arises because such interaction sites generally assume large, flat surfaces lacking specific interaction patterns,[97] such as grooves formed between α helical bundles.[98,99] Computational protein-protein docking engines have begun to address this challenge, including ZDOCK[100] and RosettaDOCK,[101] which have been used to successfully elucidate structural details of intrinsically disordered peptide-involved regulation. For example, Hu et al utilized ZDOCK to successfully predict the binding modes between disordered Yersinia effector protein and its chaperone partner.[102] Schiffer et al explored the molecular mechanism of ubiquitin transfer starting from top-ranked ZDOCK predicted binding pose between ankyrin repeat and SOCS box protein 9 (ASB9) and creatine kinase

(CK).[103] Bui et al reported that phosphorylation of the IDP fragment of transcription factor Ets1 leads to more binding-competent structures to its coactivator as evident by MD and RosettaDOCK.[104] Our studies have therefore used vetted protein-protein docking techniques and extensive MD to uncover and validate plausible sites for the secondary interaction between CaN's distal helix motif and CaM.

In this study, we harmonized a physical model of CaM-dependent CaN activation with the latter's activity. This entailed using computational methods including protein-protein docking, enhanced sampling and classical MD simulations to identify potential interaction sites between the distal helix and CaM. The protein-protein docking yielded several candidate interaction sites that we defined as sites A through D (Fig. 4.2(a)). Of these, site D on the CaM solvent-accessible surface appears to stabilize the distal helix by moderate-affinity intermolecular interactions. Among the intermolecular interactions stabilizing this putative PPI are two residues, lysine (K30) and glycine (G40) found on the 'back-side' of CaM distal to where CaMBR is known to bind. Their mutation to K30E and G40D were found to abolish enzyme activity[259] in another globular CaM target, Myosin Light Chain Kinase (MLCK), that apparently relies on still unresolved secondary interactions to initiate catalysis.[260,261] Analogously, our simulations of CaM K30E and G40D variants indicate that the mutations substantially impair distal helix binding at site D. In complement to these simulations, we demonstrate that the distal helix A454E variant also destabilizes the distal helix/site D interaction in agreement with reduced phosphatase activity shown by Dunlap et al.[46] Our data strongly suggest that the site D and CaN distal helix region are important to CaN activation, as site directed variants at site D residues K30 and G40 reduces CaN-dependent dephosphorylation of pNPP. Based on these results, we provide an updated structural model of CaN activation by CaM that reflects specific CaM/distal helix interaction sites (see Fig. 4.1) beyond the classical CaM-peptide binding motif. We qualitatively rationalize that this mechanism controls the effective concentration of the AID near CaN's catalytic site, and by extension, CaN's catalytic activity.

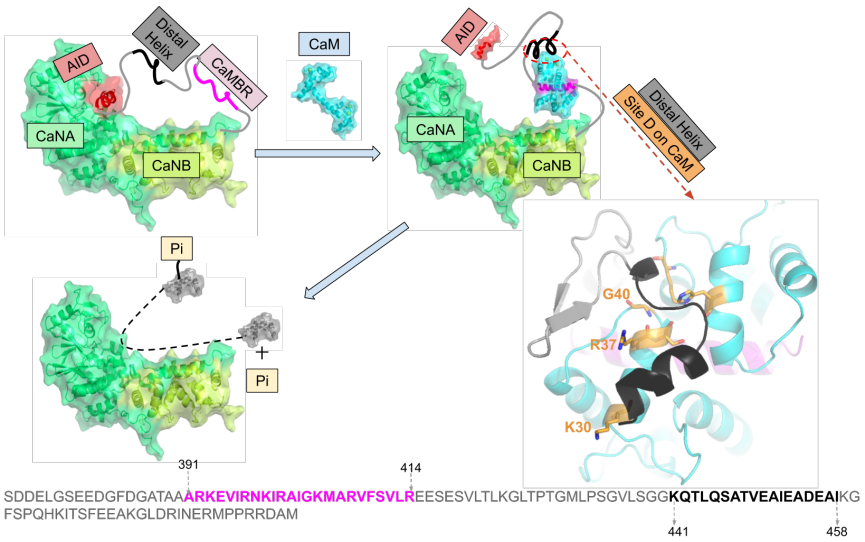


Figure 4.1: Refined model of Calcineurin (CaN) activation by Calmodulin (CaM) through direct binding of the 'distal helix' to CaM, based on the mechanism initially proposed in.[46] The two chains of CaN (CaNA and CaNB) are colored in limegreen and lime, respectively. AID is colored in red. CaM is colored in cyan, CaMBR is colored in magenta. The amino acid sequence of CaN RD is shown at the bottom of the panel with CaMBR and the distal helix region colored in magenta and black, respectively. In the absence of CaM, CaN is inhibited by its auto-inhibitory domain (AID). After CaM binds the CaM binding region (CaMBR) on the CaN regulatory domain, a secondary interaction between CaM and a 'distal helix' ultimately remove the AID from the CaN catalytic domain. The activated CaN enzyme catalyzes the dephosphorylation of target proteins essential to myriad physiological functions.

4.2 Methods

Our simulation protocol consisted of four primary steps. 1) replica exchange molecular dynamics (REMD) simulations to generate trial conformations of the isolated CaN distal helix region, 2) ZDOCK protein/protein docking engine to yield initial poses for putative CaM/CaN interaction sites, 3) refinement of poses using extensive, microsecond-length molecular dynamics simulations and 4) Molecular Mechanics-Generalized Born and Surface Area continuum solvation (MM-GBSA) were used to rank-order distal helix/CaM pose interaction scores. We further challenged the predicted structural models by introducing mutations in the distal helix and putative interaction site D that have been experimental probed in prior works and in this study via pNPP phosphatase assay.

Replica exchange molecular dynamics (REMD) sampling of the isolated distal helix

In accordance with our approach in,[257] we performed replica exchange molecular dynamics (REMD) simulations of the distal helix region (K441-I458) in the absence of CaM to exhaustively sample likely conformational that are in equilibrium. The distal helix peptide was constructed by the auxiliary TLEAP program in Amber16[262] in an extended configuration and parameterized using the Amber ff99SBildn[198] force field. The peptide was then minimized via SANDER[263] *in vacuo* until convergence of the energy gradient (drms ≤ 0.05) or the number of steps 1×10^5 (with first 50 steps of steepest decent and rest steps of conjugate gradients algorithm) was satisfied. The minimized structure was then used as the starting structure for REMD simulations coupled with the Hawkins, Cramer, Truhlar pairwise generalized born implicit solvent model[264] via the *igb = 1* option in Amber. The monovalent 1:1 salt concentration was set to 0.15 M and a non-bound cutoff of 99 Å was chosen. Ten replicas were created with temperature ranges spanning 270-453 K. The temperature of each replica was calculated via the Patriksson et al webserver[265] to ensure the exchange probability between neighbouring replicas was approximately 0.4, as recommended in.[266,267] Each replica was first subjected to 1×10^5 steps of energy minimization via PMEMD with the first 50 steps via steepest decent and remaining via conjugate gradients. The minimized systems were subsequently heated from 0 to their respective target temperatures over an 800 ps interval using a timestep of 2 fs with a Langevin thermostat. The equilibrated replicas were then subjected to 100 ns of production REMD simulations under target temperature with Langevin thermostat. The SHAKE[201] algorithms were used for REMD simulations. Clustering analysis with a hierarchical agglomerative (bottom-up) approach using CPPTRAJ were conducted on the 300 K REMD trajectory to divide the trajectory into ten clusters; the average root mean squared deviations (RMSD) between each cluster was around 6 Å.

Docking of distal helix to CaM/CaMBR complex via ZDOCK

The protein-protein docking webserver ZDOCK3.0.2[100] was used to determine probable binding poses for the REMD-generated distal helix conformations on the CaMBR-bound CaM complex. The CaM/CaMBR complex configuration was obtained from the Protein Databank (PDB ID: 4q5u[258]). It has been reported that 62% percent of experimentally-resolved PPIs are characterized by the binding of an α-helical peptide within grooves formed between adjacent α-helical on the target protein surface;[99] therefore we narrowed the ZDOCK search to four α-helical-containing regions on the CaM solvent-exposed surface. These sites are shown in Fig. 4.2(a), from which we determined a list of probable amino acid contacts as input to ZDOCK (see Table S4). During the ZDOCK calculations, the receptor (CaM/CaMBR complex) was kept fixed while grids were constructed around receptor with dimensions as 80*X*80*X*80 and spacing

as 1.2 Å. The ligand (distal helix) was then docked via the fast fourier transform (FFT) algorithm on the 3D grids. The scoring function consists of interface atomic contact energies (IFACE),[268] shape complementarity and electrostatics with charge adopted from CHARMM19 force field.[269] The initially generated 2×10^3 poses were subjected to a culling process to eliminate those having no contacts with residues we specified in Table S4. After culling, there were zero, two, eighty-eight and three poses left at sites A-D, respectively. The pose with highest score at each site was chosen for further refinement using molecular dynamics.

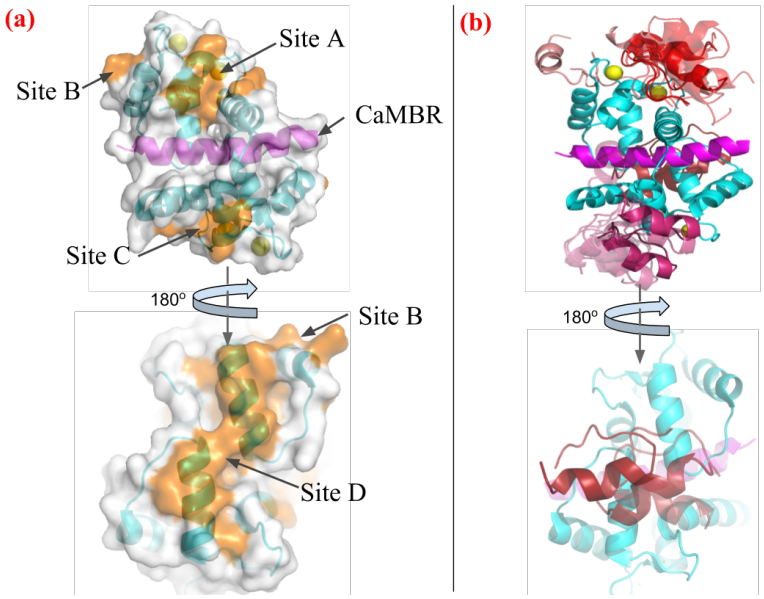


Figure 4.2: (a) Four tentative binding sites (orange) on the surface of CaM-CaMBR complex. CaM is colored in cyan, CaMBR is colored in magenta and Ca^{2+} ions are colored in yellow. (b) ZDOCK predicted conformations of distal helix interacting with CaM/CaMBR complex at each site. Predicted distal helix conformations from site A to D are colored as red, salmon, warmpink and firebrick, respectively.

Conventional molecular dynamics (MD) simulations of ZDOCK-generated distal helix/CaM poses

Explicit-solvent MD simulations were performed on the ZDOCK-predicted distal helix/CaM complexes to further refine the distal helix binding poses. The amino acid sequence from CaMBR to distal helix is shown at the bottom of Fig. 4.1 and the sequence definition of CaMBR and distal helix are the same as.[46] We first inserted peptide linkers for each pose between the CaMBR C-terminus (R414) and the N-terminus (K441) of the distal helix via TLEAP. The initial linker was generated via TLEAP and energy-minimized as done in Sect. 4.2. The minimized structures were subsequently simulated *in vacuo* to heat the systems from 0 to 300 K. The last frame of the short equilibration run was subject to additional energy minimization *in vacuo* to facilitate its compliance with the distal helix and CaMBR termini. The top poses from ZDOCK presented distal helix orientations that were all compatible with the CaMBR and linker configurations. The optimized linker was placed adjacent to the CaMBR and distal helix; TLEAP was then used to link the peptide components. The resulting structures were then subjected to energy minimization, followed by a 100 ps heating process to raise the system temperature to 300 K for which all atoms except the linker were fixed via the ibelly function in SANDER MD engine of

Amber. This minimization and heating was performed *in vacuo* to further relax the linker in the presence of distal helix and CaM/CaMBR complex. The last frame of the heating stage was used as input configurations for explicit-solvent molecular dynamics simulations.

Each *in vacuo* starting configuration was solvated in a TIP3P[72] waterbox with 12 Å boundary margin. K^+ and Cl^- ions were added to neutralize the protein and establish a 0.15 M salt concentrations. After parameterizing the system using the ff14SB force field[270] via TLEAP, the system was subjected to energy minimization, for which all atoms except hydrogens, water and KCl ions were constrained by the *ibelly* function. The cutoff value for non-bond interactions was set to 10 Å. A 2 fs timestep was chosen, as SHAKE[201] constraints were applied on bonds involving hydrogen atoms. Two heating procedures were performed to heat the system from 0 to 300 K using the Amber16 SANDER.MPI engine.[263] In the first heating stage, the *ibelly* function was used to keep the protein fixed and the surrounding solvent unrestrained. The water box was heated to 300 K over a 100 ps interval under the NVT ensemble. For the second heating stage, the entire system was heated from 0 to 300 K over 500 ps under the NPT ensemble, for which the backbone atoms of CaM, CaMBR and distal helix were constrained by harmonic potential (force constants of $3\,\mathrm{kcal\,mol^{-1}\,Å^{-2}}$). Thereafter, an additional 1 ns equilibrium stage was conducted at 300 K under the same constraints, but with a reduced force constant of $1\,\mathrm{kcal\,mol^{-1}\,Å^{-2}}$. These equilibrium simulations were followed by 100 ns production-level MD simulations. The weak-coupling thermostat[271] was used during the simulation. Clustering analysis was performed on the production trajectory using the same strategy in Sect. 4.2. The average RMSD between each cluster was around 6 Å. Based on the rationale that extending simulations using less-frequently sampled structures provides greater overall sampling of the conformational space,[272] we identified 5-6 low-probability states as inputs for subsequent MD simulations. Approximately 1 μs of trajectory data were simulated in total for each site.

MD simulations of CaM (K30E and G40D) and CaN distal helix variants (A454E)

Clustering analyses were performed on the production-level MD trajectories of the distal helix/CaM configurations that yielded the most favorable binding scores by MM/GBSA. The binding free energy between distal helix and CaM was estimated via Molecular Mechanics-Generalized Born and Surface Area continuum solvation (MM-GBSA).[273]

$$\Delta G = \langle G_{DH-CaM} \rangle - \langle G_{CaM} \rangle - \langle G_{DH} \rangle \tag{4.1}$$

Where $\langle G_{DH-CaM} \rangle$, $\langle G_{CaM} \rangle$ and $\langle G_{DH} \rangle$ are ensemble-averaged free energies of distal helix-CaM complex, CaM and distal helix, respectively. A representative structure of the most populated cluster was selected as an input for *in silico* mutagenesis in order to validate the model against experiment. Namely, the

CaM K30E and G40D variants, as well as the CaN A454E variant, were built by replacing and regenerating the amino acid side chains using TLEAP. The resulting structures were energy minimized with a stop criterion of (drms <= 0.05) for the energy, during which all atoms except the mutated residues were fixed via the *ibelly* function in Amber. The energy-minimized structure was then solvated and simulated according to the same procedure in Sect. 4.2. In the MM-GBSA calculations, the trajectories of these three components in Eq. 4.2 were extracted from MD trajectories via CPPTRAJ at a 2 ns frequency. The generated sub-trajectories were used as input of MMPBSA.py script in Amber16 to calculate the free energies of each part. The salt concentration was set as 0.15 M with generalized Born model option setting as *igb = 5*. No quasi-harmonic entropy approximation was made during the calculation.

Structural Analyses

Clustering analysis, root mean squared deviations (RMSD)/RMSF calculations, hydrogen bonds and secondary structure analysis were performed via CPPTRAJ[274] in Amber16. The reference structure used for these analyses was the CaM/CaMBR crystal structure (PDB ID: 4q5u[258]). Secondary structure for each residue was calculated using CPPTRAJ with define secondary structure of proteins (DSSP) algorithm.[275] The COLVAR module[276] within VMD was used to assess the total α-helix content of REMD-generated distal helix and DH_{A454E} conformation. The *hbond* command within CPPTRAJ was used to analyze hydrogen bonds between distal helix and CaM/CaMBR. During the hbond analysis, the angle cutoff for hydrogen bonds was disabled while the default 3 Å cutoff between acceptor and donor heavy atoms was used. Scripts and CPPTRAJ input files used for above analyses will be publicly available at https://bitbucket.org/pkhlab/pkh-lab-analyses/src/default/2018-CaMDH.

Calcineurin phosphatase assay using para-nitrophenyl phosphate (pNPP) substrate

Materials. pNPP was obtained as the bis(tris) salt (Sigma), dithiothreitol reducing agent (Sigma), assay buffer (80 mM Tris pH 8, 200 mM KCl, 2 mM $CaCl_2$), and 50 mM $MnCl_2$ used as a CaNactivating cofactor. *Preparation of Enzymes and Proteins*. The CaM wild-type, K30E and G40D variants were generated, expressed and purified as previously described.[259] CaN was expressed from the pETagHisCN plasmid (from Addgene, Cambridge MA) in *E. coli* BL21 (DE3) CodonPlus RIL cells (Agilent, La Jolla CA). The enzyme was unified via a Ni-NTA column followed by a CaM-sepharose column (GE Healthcare, Piscataway NJ) as described in.[35] *Enzyme Assay*. Phosphatase assays were performed using 30 nM CaN, and 90 nM CaM in 96-well Corning Costar microtiter plates with a reaction volume of 120 μL. Assays proceeded in the manner of[46] with each CaM assayed in triplicate and over three plates to account for technical variation. Control reactions absent CaN were added to the end of each lane

with 200 mM pNPP to determine the rate of enzyme-independent substrate hydrolysis. *Kinetic Analysis*. The pNPP substrate reactions were varied over 11 concentrations, increasing from 0 mM to 200 mM for each column. 60 minute UV-Vis recordings were obtained on a Molecular Devices FlexStation 3 plate reader using Softmax Pro 7 software at 405 nm with 10 minute read intervals. The resulting data were inspected for appropriate Michaelis-Menten kinetics by plotting absorbance against substrate concentration. Readings were linearized to produce the double reciprocal Lineweaver-Burk plot for estimation of V_{max} and K_M based on the following equation:

$$\frac{1}{V} = \frac{K_M}{V_{max}}[pNPP] + \frac{1}{V_{max}} \tag{4.2}$$

4.3 Results

Prior studies[35,277] have indicated that CaM binding to the CaN's canonical CaM-binding region requires secondary interactions beyond this region to fully activate the phosphatase. Rather, CaN activity is likely dependent on a secondary interaction between the CaN regulatory domain and CaM. A study by Dunlap et al[46] suggested that a distal helix region spanning residues K441 to I458 was likely involved in CaM binding. However, it was unclear which region(s) of the CaM solvent-exposed surface would contribute to a potential PPI. We therefore used molecular dynamics and protein-protein docking simulations to identify plausible wild-type CaN interaction sites on CaM, and challenge these predictions with mutagenesis. Our predicted site was validated using a CaN pNPP phosphatase assay.

Regulatory domain (RD)-construct propensity for secondary structure formation in absence of CaM

Circular dichroism (CD) and HXMS analysis in[35] suggest that there exists α-helical structure beyond the canonical CaMBR region after CaM's binding. We therefore sought to assess α helicity in the REMD-simulated distal helix peptides. Previously,[257] we found that extensive MD simulations of the isolated CaMBR yielded a small population of α-helical structures suitable for binding CaM in its canonical binding pose.[183]

We therefore applied a similar REMD procedure (see Sect. 4.2) to the proposed distal helix segment of the CaN regulatory domain to assess the propensity for the spontaneous formation of secondary structure in the absence of CaM. Here, we performed 100 ns of REMD simulations on the wild-type (WT) distal helix as well as a A454E variant. The latter was considered as it has been reported to exhibit reduced α-helical content in the presence of CaM,[46] which is suggestive of abolishing the distal helix/CaM interaction. Following the REMD simulations, we performed clustering analysis to identify the predominant conformations of the two peptide configurations. Interestingly, we observed

that both the WT distal helix and its A454E mutant partially fold into an α helix in the absence of CaM. As shown in Fig. 4.3(a), the representative structure of most populated clusters of the distal helix and A454E mutant (83.8% and 85.3% of the total trajectory, respectively) both contain helical fragments. While the overall α-helix contents ($\approx$ 45%) of these two fragments were statistically indistinguishable, a contiguous helix was formed in the WT distal helix, whereas it was fragmented in the mutant. These helicity features are further quantified as residue's α-helix structural probability shown in Fig. 4.3(b): the distal helix region has the maximum probability present at middle region while the A454E has maximums present near the two terminis. Both the simulated distal helix and its variant therefore could adopt α-helix content in the absence of CaM, but it remains to be determined whether the dominant structures are capable of binding the CaM surface. We note that experimental assays of the complete RD do not detect significant secondary structure; this discrepancy may be a result of using substantially different RD lengths (S374 to Q522 residues in Rumi-Masante et al[35] and K441-I458 in this study). We discuss this difference in further detail in Limitations (Sect. 4.4).

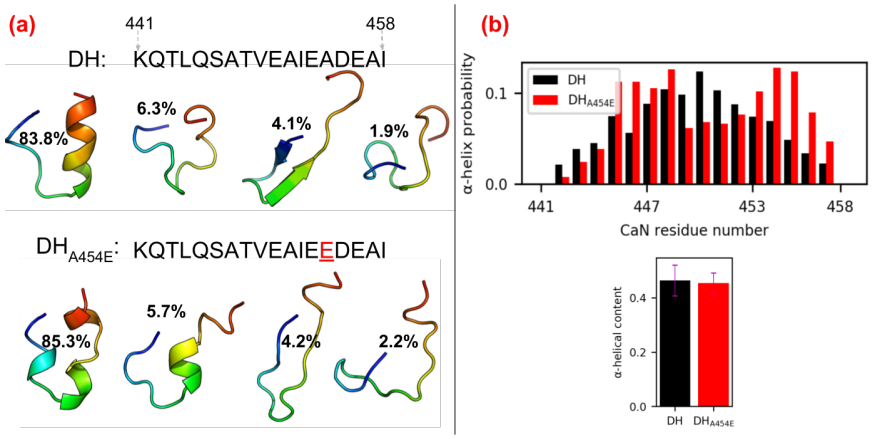


Figure 4.3: (a) Sequence of distal helix/DH$_{A454E}$ and representative structures of four most populated clusters from 100 ns REMD simulations. The structures are colored in rainbow with N-termini as blue and C-termini as red. (b) Secondary structure probability of each residue calculated from REMD trajectory via CPPTRAJ with DSSP algorithm. The lower panel shows the total α-helix contents of two fragments calculated via the COLVAR module of VMD.

Protein-protein interactions between RD-construct and peptide-bound CaM

The overwhelming majority of CaM-containing complex structures resolved to date include only limited fragments of the bound target protein.[183] CaM-bound CaN is no exception, as the mostly likely physiological conformation[258] consists of monomeric CaM in a canonical 'wrapped' conformation about a target region in CaN(A391-R414);[33] however, it is evident that secondary interactions beyond this domain play a role in CaN activity, yet atomistic-level structural details of these interactions have not yet been resolved. Therefore, in order to resolve potential binding regions for the distal helix region, we seeded a protein-protein docking engine, ZDOCK,[100] with candidate α-helical structures identified through REMD simulations. The docking simulations were performed in regions that included grooves formed between α helices we identified at the CaM solvent-accessible surface. We selected these regions, since such secondary structures are believed to nucleate protein-protein interactions.[278] Furthermore, a thorough examination of protein-protein complex structures in the Protein Data Bank in 2011 suggested that α helices contribute to 62% of all PPI interaction surfaces[99] between binding partners. Narrowing the search region on CaM to those containing α-helical regions yielded four candidate sites (A-D) that spanned nearly the entire CaM solvent-exposed surface (see Fig. 4.2(a)).

The most energetically-favorable distal helix/CaM poses predicted via ZDOCK at sites A-D are summarized in Fig. S14. The docked poses reflect significant interactions of at least the distal helix C-terminal loop with loops bridging

adjacent α-helices on the CaM surface. At site A, polar residues near N97, Y99 and D133 from two of the C-terminal CaM domain's loops interact with the distal helix, compared with just one EF-hand motif loop at site B (D129, D133 and D135). The site C poses were primarily stabilized by hydrophobic interactions formed from CaN residues L444/I458 and F16/L4 on CaM, in addition to a loop-loop interaction via CaM D64. The site D poses reflected distal helix C-terminal loop interactions with CaM EF-hand loop residues near N42 and K94. Most poses were unsurprisingly parallel to α-helical/α-helical 'grooves' on the CaM solvent-exposed surface and were evidently anchored through interactions between the proteins' loop regions.

In contrast, we found that the A454E variant docked poorly at sites A-D (see Fig. S15), as assessed by the proximity of docked poses to the designated CaM sites in fact, most predicted poses tended to localize toward site A, albeit with weak interactions. Moreover, we speculate that the impaired binding of DH_{A454E} may arise from its fragmented α helical structure, in contrast to the contiguous regions for the WT variant (see Fig. S14/Fig. S15 for docking poses). Although docking scores were provided by the ZDOCK algorithm to rank order potential poses, we did not analyze these scores in detail as we later refined these structures using more detailed simulations and energy expressions. This refinement corrects for artifacts from the ZDOCK algorithm, which assumes rigid conformations for both proteins that would ordinarily be expected to relax in the bound complex. Hence, in the following section we pursue extensive microsecond-scale all-atom MD simulations to refine and assess the predicted poses.

Molecular dynamics (MD) simulations of docked distal helix/CaM poses

The docked CaN/CaM configurations from the previous section were intended as inputs for MD-based refinement of nearly intact CaN regulatory domain complexes with CaM. Subsequent refinement using microsecond-length MD simulations relax the rigid protein conformations assumed in ZDOCK. To refine these poses, we linked the docked distal helix fragments with the CaMBR fragment resolved in the CaM/CaN complex (PDB ID: 4q5u) from.[258] Each of the four candidate binding sites yielded distal helix orientations that were compatible with the 26 residue-length linker. Following initial optimizations of the linker described in Sect. 4.2, we performed μs-length, explicit solvent simulations with the regulatory domain bound to CaM. Since the predicted A454E distal helix poses appeared to be inferior to those of the WT variant, we refined only the WT poses and thereafter introduced A454E mutations to the refined conformations.

Binding free energy between distal helix and CaM/CaMBR We first assess the integrity of the predicted binding modes based on Molecular Mechanics-Generalized Born and Surface Area continuum solvation (MM-GBSA). MM/GBSA scoring of the MD-generated configurations provides a coarse estimate of

binding affinity without significantly more expensive free energy methods. We reported the binding free energy of distal helix between CaM/CaMBR as well as between CaMBR and CaM in Fig. 4.4. Significantly, we found that binding of WT distal helix at the CaM site D yielded a more pronounced favorable average binding free energy ($\Delta G \approx -27.7$ kcal mol^{-1}) than sites A, B and C (-2.5 kcal mol^{-1}, -17 kcal mol^{-1}, -22.5 kcal mol^{-1}) with P-values (1×10^{-4}, 2.8×10^{-3} and 1.144×10^{-1}, respectively) confirming that the means are significant compared to the null hypothesis. Notably, these thermodynamically favorable scores are suggestive of the potential for the distal helix to bind multiple regions on the CaM surface, although site D is the most favorable site. Similarly, the binding free energies of distal helix interactions were generally substantially weaker (-2.5 to -27.5 kcal mol^{-1}) than those between the CaMBR and CaM ($\Delta G < -1.20 \times 10^2$ kcal mol^{-1}) Although MM-GBSA is a very approximate scoring method for molecular complexes, the consistent trends in numbers of hydrogen bond contacts, RMSF amplitudes and binding scores suggests that the site D is the most likely region for forming stable CaM/distal helix interactions.

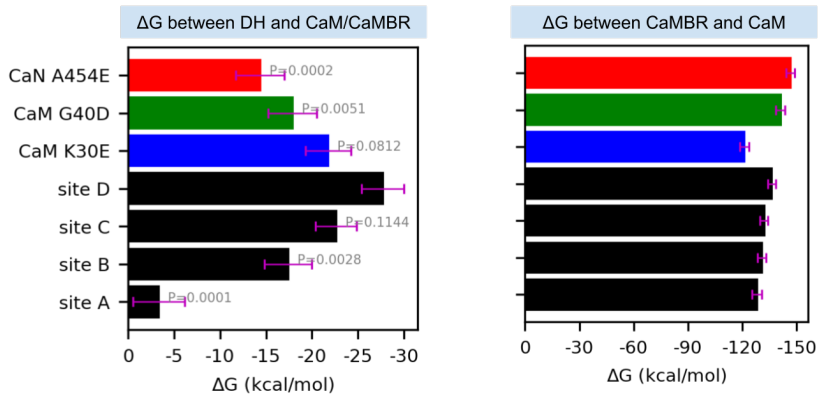


Figure 4.4: Approximate binding free energies between CaM and the distal helix (left) or CaMBR regions (right) via Molecular Mechanics-Generalized Born and Surface Area continuum solvation (MM-GBSA). Black bars correspond to wild-type CaN, whereas colored bars utilize the A454E CaN and CaM variants. The calculation was conducted on frames extracted every 2 ns from MD trajectories. The error bar represents standard error of mean. The values above bars in the left panel are P values of each case with null hypothesis that their mean values are equal to site D.

We supplement the energy scores with structural indicators of stability, namely contacts and RMSF. We report in Fig. 4.5 the corresponding root mean squared deviations (RMSD) and root mean squared fluctuations (RMSF) of the peptide backbone atoms from CaM and CaMBR. We additionally include two CaM variants with mutations at site D, which we rationalize later in Sect. 4.3. We found that the average RMSD values of the MD-predicted conformations relative to the experimentally-determined CaM/CaMBR structure were at or below 2 Å; we attribute these small fluctuations to stable CaM/CaMBR interactions that were insensitive to the distal helix docking pose. Similar to the RMSD data, the CaM and CaMBR RMSF values are comparable in amplitude and nearly indistinguishable between distal helix/CaN docking poses, with most residues presenting values below 1.5 Å. The prominent peaks in excess of 5.0 Å correspond to the CaM termini and the N-terminus of the CaMBR. We additionally observe a variable region midway along the CaM sequence, which corresponds to the labile linker between its globular N- and C- domains that is implicated in allosteric signaling.[279]

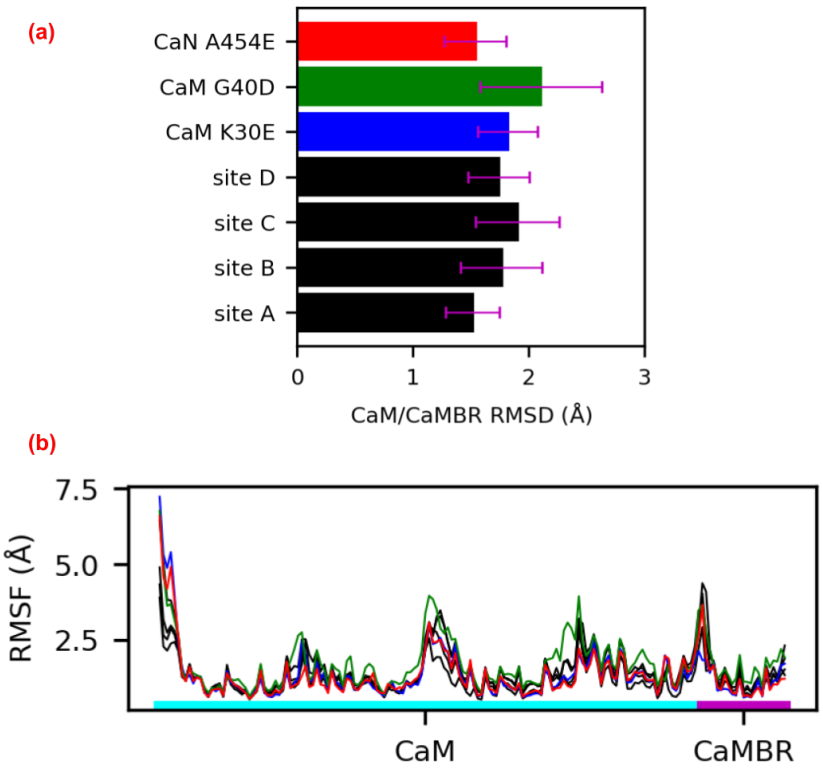


Figure 4.5: (a) Root mean squared deviations (RMSD) of peptide backbone atoms of CaM and CaMBR from μs-length MD simulations. The reference structure for the RMSD calculation was based on the CaM/CaMBR crystal structure (PDB ID: 4q5u). (b) Root mean squared fluctuations (RMSF) of non-hydrogen atoms in CaM and CaMBR.

Distal helix poses The small and statistically indistinguishable RMSF values for the CaM/CaMBR in Fig. 4.5 suggest that distal helix binding had negligible impact on binding the CaM recognition motif. This is an important observation, as viable binding poses for the distal helix are expected to preserve the binding between the CaMBR and CaM. We base this assumption on CD data collected in[280] that indicated substantial alpha helical character in the CaM/CaN complex following dissociation of the distal helix domain. Therefore, we then assessed the integrity of the distal helix poses using RMSF analyses and measurements of inter-protein contacts. In Fig. 4.7 we report representative configurations of the distal helix region (red) in complex with CaM (cyan), as well as their corresponding per-residue RMSF values in Fig. 4.6. To guide interpretation, we hypothesized that RMSF values above 5 Å were indicative of poorly stabilized residues. We later rationalize this value by comparing approximate binding energies as computed by MM-GBSA. At site A, both the distal helix/CaMBR linker and the distal helix reflect RMSF values in excess of ∼10 and ∼15

Å, respectively. These large fluctuations arise from the breadth of binding orientations evident in Fig. 4.7(a), which we interpreted as a poorly-stabilized configuration. Similarly, the site B configurations also appear to be loosely bound, based on linker and distal helix RMSF values beyond 10 Å. In contrast, the distal helix RMSF values at sites C and D were below 5 Å, with the latter site reporting the smallest values among the sites we considered, which is evidence of a stable binding configuration.

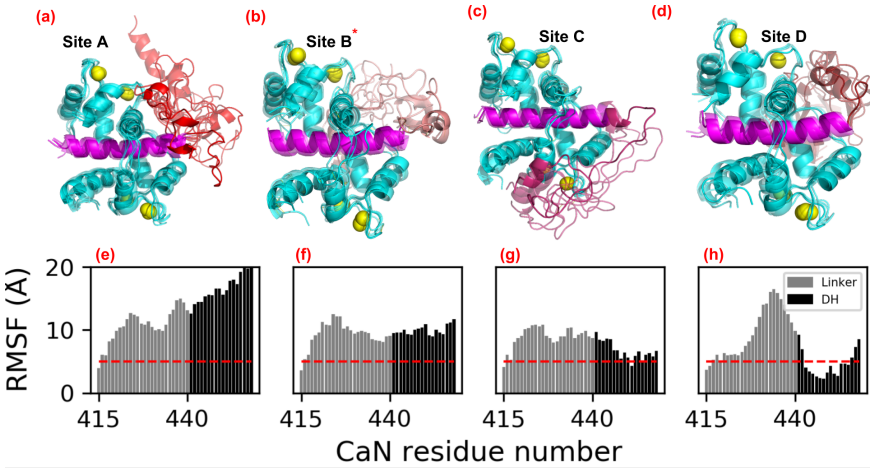


Figure 4.6: (a-d) Representative structures of from the microsecond length MD simulations initialized from ZDOCK-predicted distal helix poses. CaM is colored in cyan, CaMBR is colored in magenta and Ca^{2+} ions are depicted as yellow spheres. The linker and distal helix regions in site A-D are colored as red, salmon, warmpink and firebrick, respectively. (e-h) Non-hydrogen atom RMSFs of linker and distal helix residue calculated from MD simulations of each site, as an indicator of binding stability. The red dash line depicts RMSF value as 5 Å. * During the MD simulations, distal helix structures initiated at site B migrated toward site D (see Fig. S16).

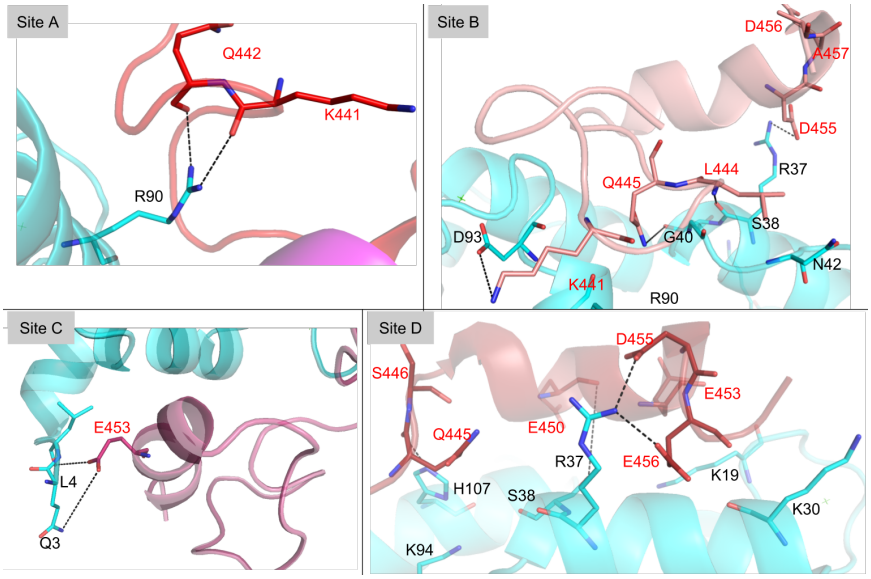


Figure 4.7: Interaction between linker/distal helix of CaN and CaM at site A-D. Key residues at the interaction surface are shown in sticks with black labels for CaM residues and red labels for distal helix residues. See supporting information of[281] for specific values.

As has been shown in other proteins regulated by disordered protein domains,[282–284] there are often multiple poses the contribute to regulation. We therefore assessed the most significant inter-protein contacts contributing to the ensemble of distal helix binding poses at sites A-D. Among these poses, the distal helix configurations at site D presented the lowest distal helix RMSF values among the considered sites. Significantly, the site D distal helix configuration presented several hydrogen bond-facilitated interactions with CaM, including two long-duration (37% and 55% of sampled configurations) interactions between Q445 and CaM residues R37/K94, pairing of CaM K21 with glutamic acids E453 and E450, as well as E456 with CaM residues K30 and R37. Contacts between CaM and CaN, as well as their longevities (as assessed by the percentage of MD frames satisfying a hydrogen bond contact cutoff of 3 Å between oxygen and nitrogen atoms) are additionally quantified in Fig. 4.8. The latter data indicate a modestly greater degree of hydrogen bonding of the distal helix at site D (10 h-bonds were above 10%) versus site B (9), and a significantly greater degree relative to sites A (1) and C (3). Furthermore, the site D pose appears to be stabilized by both the N- and C-domains of CaM (residues D20-S38 and R90-N111, respectively). We speculate that this bi-dentate interaction could improve CaMBR binding by locking CaM into its collapsed configuration and thereby prevent disassembly. Although during the simulations, the distal helix at site D maintained significant α-helix (see Fig. S16), we note that a significant percentage of the predicted structures exhibited beta sheet character in the linker region (see Fig. S17) that was not observed in the CD cpectra collected

by Rumi-Mansante et al.[35] This persistent secondary structure was limited to a few residues (see Fig. S17) thus may be beyond the limits of detection in earlier CD experiments. We comment on this further in the Limitations (see Sect. 4.4). Meanwhile, site B reflected interactions with both CaM terminal domains that were attenuated, while sites A and C were mostly bound by interactions of their linker regions with the CaM N-domain. Interestingly, we observed that the distal helix poses originating at site B migrated toward site D (see Fig. S16), which likely explains the higher hydrogen bonding in site B versus sites A and C.

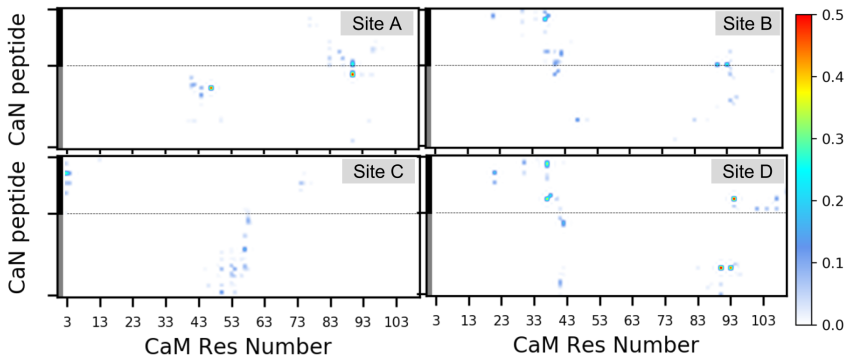


Figure 4.8: Percentage of simulated frames which have hydrogen bonds formed between CaN peptide (linker and distal helix) and CaM. The linker and distal helix are indicated by grey and black bar, respectively.

As a result of HXMS conducted by Rumi-Masante et al[35] of the RD construct CaN in solution with CaM, it is apparent that residues R414 through E456 are within a stretch of residues that are somewhat protected from solvent, which suggest that relief of CaN autoinhibition entails binding at least the distal helix region. We note that the HXMS data could not precisely distinguish which residues were protected, as proteolysis and mass spec was conducted on short peptides. Further, HXMS data detects only bonds involving backbone amide protons, thus we speculate that the CaN site chain interactions with CaM may stabilize the distal helix alpha helical structure. Hence, we suggest that CaM/CaN configurations that stabilize the distal helix region likely contribute to CaN activation. Based on this rationale, the small RMSF values and extensive hydrogen bonding of the CaN distal helix with the CaM site D relative to other ZDOCK identified regions suggest that CaN is most stabilized at site D.

Effects of putative CaN/CaM site D mutagenesis

MD simulations of the WT CaN CaMBR-distal helix sequence suggest that CaM site D is a probable binding region for the CaN regulatory domain. To challenge this hypothesis, we performed MD simulations of CaN distal helix and CaM site D variants that could reduce CaN activity to test whether the distal helix/CaM interaction was impaired. Namely, we introduced the CaN A454E and CaM K30E and G40D mutations into the MD-optimized WT structures. We elected to mutate the WT CaMBR/distal helix complexes with CaM, as the WT complex appeared to have favorable stability, whereas repeating the REMD/zdock steps with the mutants may not have yielded viable configurations. The proposed A454E CaN variant was based on CD data collected by Dunlap et al[46] that demonstrated reduced α-helical content upon binding CaM relative to the WT with impaired CaN activation. The CaM variants we examined in this study were based on experimental mutagenesis studies[259] of CaM-dependent Myosin Light Chain Kinase (MLCK) activation,

for which secondary interactions beyond the canonical CaM binding motif were required for enzyme activation[260,261] (Fig. 4.9(a)). Although these secondary CaM interactions are involved in directly binding the MLCK catalytic domain in contrast to CaN,[260] two residues (K30 and G40) implicated in binding[259] reside within the site D identified in our simulations.

We also reported the MM-GBSA-calculated binding free energies between distal helix and CaM of the mutants in Fig. 4.4. While the WT distal helix at the CaM site D has most stable binding with $\Delta G \approx -27.7$ kcal mol^{-1}, the three mutations K30E, G40D and A454E have less favorable ΔGs as -21.8 kcal mol^{-1}, -17.9 kcal mol^{-1} and -24.4 kcal mol^{-1} with P-values being 8.12×10^{-2}, 5.1×10^{-3} and 2×10^{-4}, respectively. The MM-GBSA-energies clearly shown that mutations would impair the binding affinity between distal helix and CaM. Accordingly, we presented linker and distal helix RMSF data for the WT and mutants in Fig. 4.9(b). The distal helix RMSF values among the two CaM variants were moderately increased compared to the WT case. Specifically, for the WT system, the distal helix residues were entirely within 10 Å and as low as ∼2.5 Å. In contrast, the K30E variant yielded RMSF values no smaller than approximately 5 Å, while the C-terminal half approaches values nearing 15 Å. This trend manifested in fewer long-lived hydrogen bond contacts between the distal helix and both CaM domains (see Fig. 4.9). Similarly, the G40D mutation appeared to significantly disrupt interactions with CaN, as the entire distal helix region was characterized with RMSF values over ∼10 Å in amplitude, with corresponding decreases in hydrogen bond contacts. We reported the MMGBSA calculated binding free energy between the distal helix and the CaM/CaMBR in Fig. 4.4. Among the mutations we considered, the A454E mutant had the most severe impact on RMSF values, as all residues comprising the linker and distal helix regions resulted in fluctuations above 8 Å. We also reported the α-helix probability of distal helix residue for variants in Fig. S18. It was found that all variants preserved a significant degree of overall helicity despite evidence of impaired interactions with CaM. However, the specific residues which formed α-helix were different among the variants: the mutation of A454 to E454 shifted the helicity to the first half of distal helix while the two CaM variants had the second half region being α helical. Altogether, these simulation data suggest that 1) the WT distal helix is stabilized at the site D CaM region, 2) site D residues R37 and K30 are implicated in distal helix binding and 3) disruption of site D binding by CaN A454E is consistent with reduced helicity and enzyme activity measured experimentally.

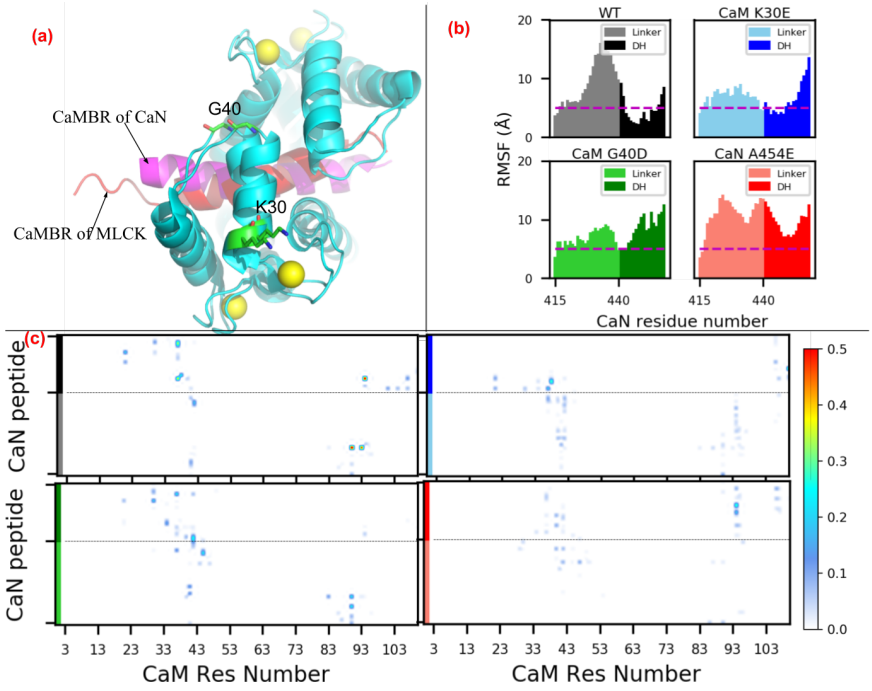


Figure 4.9: (a) Comparison of CaM-petide complex structure from CaN and MLCK (PDB ID: 2lv6[285]). K30 and G40 are labeled (shown as sticks) based on their implication in the activation of the CaM target Myosin Light Chain Kinase (MLCK)[259] and proximity to site D determined by our simulations. (b) Non-hydrogen RMSF of linker and distal helix in WT and mutants. The dash line depicts RMSF value as 5 Å. (c) Percentage of simulated frames which have hydrogen bonds formed between linker/distal helix and CaM

Phosphatase assays of site-directed CaM mutants

To support the simulation results, namely that the distal helix region binding predominantly to site D would impact CaN activity, we analyzed the kinetics of CaN mediated hydrolysis of pNPP. Our hypothesis was that disruption of site D/distal helix binding would reduce the accessibility of the catalytic site for pNPP binding which would reduce the apparent substrate affinity. This reduction would arise from the AID competing for the catalytic site, as a result of compromised site D/distal helix interactions. We therefore conducted pNPP assays using two site D variants, K30E and G40D. We analyzed substrate turnover in a Michaelis-Menten model, as described in the Methods. Phosphatase assays performed on CaM variants strongly suggest a statistically significant reduction (p-values in Table 4.1) in catalytic activity by a substantial increase in K_M for K30E and G40D over the WT (27.6±1.3 mM, 46.0±2.8 mM, and 35.5±2.2 mM, respectively) indirectly indicating weaker binding of the distal helix peptide to the mutated CaM construct.

Table 4.1: Kinetic parameters of pNPP dephosphorylation with WT CaM and two site D variants. P-values given by Welch's t-test for difference of means with unequal variance.

CaM	K_M (mM)	SD	p-value
WT	27.6	1.3	-
K30E	46.0	2.8	0.002
G40D	35.5	2.2	0.008

4.4 Discussion

Summary of Key Findings

We have used computational modeling to elucidate a potential mechanism for CaM-dependent regulation of CaN activity, whereby the binding of a 'distal helix' region of the regulatory domain relieves CaN auto-inhibition. Our microsecond-duration MD simulations indicate that the distal helix region maintains bound to the solvent accessible CaM surface, which could decreases the ability of the AID to bind CaN's catalytic site (see Fig. 4.1). In contrast, we predict that an engineered variant (A454E) disrupts the domain's secondary structure and ability to competently bind CaM. Both predictions are in agreement with experimental probes of CaN regulatory domain structure and phosphatase activity.[46] Namely, among the four potential regions on CaM's surface that were solvent-accessible after binding the CaMBR, our data suggest that an RD region spanning the CaMBR through the distal helix was best stabilized at a site nestled between the CaM N- and C-terminal domains. In silico mutagenesis of two N-terminal CaM residues (K30E and G40D), prevented distal helix binding in our model, which we suggest hinders CaN activation, similar to identical mutations in CaM that were found to inactivate another CaM target, Myosin Light Chain Kinase (MLCK). We confirmed the potential CaM site D binding site for the distal helix through site-directed K30E and G40D variants, which we found to weaken CaN binding as reflected by reduced (weaken) MM-GBSA scores and an increase in K_M (from 27.6 mM to 46.0 and 35.5 mM, respectively) in a pNPP phosphatase assay. Although our REMD simulations suggest that the isolated distal helix region spontaneously assumes significant α-helical in absence of CaM; in contrast to trends observed in the complete RD domain observed experimentally,[35] we do not believe this significantly impacts our suggestions of site D in functional contributions to CaN activation. We discuss this limitation and its implications in Sect. 4.4.

Plausible binding modes for putative CaN distal helix with CaM

Comparison w Creamer assays

Previous studies suggest that binding of regulatory domain residues beyond the CaMBR region are involved in CaM-dependent relief of CaN autoinhibition.[35,46] Increases in regulatory α-helical content were reported upon binding CaMthat could not be accounted for by the CaMBR alone. Alanine to glutamic acid mutations at RD positions (A451E, A454E and A457E) C-terminal to the CaMBR decreased α-helical content and CaN activity. Further, HXMS studies indicate that in a complex of CaM with a regulatory domain/AID/C-terminal domain CaN construct that the CaMBR through distal helix regions had reduced solvent accessibility, suggestive of secondary interactions beyond the CaMBR. We calcula-ted the backbone hydrogen bonds formed within the linker and distal helix region as an indicator of solvent-protection and compared this against experimen-tal HXMS data. As shown in Fig. 4.10, WT site D has 16 hydrogen bonds with 2 dominant hydrogen bonds (red arrow) formed within the β-sheet region (Fig. S17). Also in the distal helix region, two long-lived hydrogen bonds (>40% simulation time) were found. Compared with other sites/mutants, backbone hydrogens at site D would be most protected from HXMS due to the larger number of hydrogen bonds and relatively longer duration. Although A454E has the largest number of hydrogen bonds, most are short-lived and the residue pairs which form hydrogen bonds are well seperated in sequence, indicating these hydrogen bonds do not contribute to α-helix secondary structure. Our computational modeling suggests that the putative distal helix region contains significant α helical character when bound to CaM site D, which qualitatively resemble those of experiment and suggests reduced susceptibility to hydrogen/d-euterium exchange. Nevertheless, compared to experimental HXMS data showing solvent-protected hydrogens are present across the whole linker and distal helix region, our computational backbone hydrogen bonds data indicates a lesser degree of solvent-protection as the majority of hydrogen bonds are present in the N terminus of linker region and C-terminus of distal helix region in site D. This discrepancy could be explained by the different lengths of CaN constructs used in HXMS experiment and our simulations. The construct in HXMS experiment contains the entire RD domain including AID and the C-terminus, while our simulations contain residues of A391 to I458 of RD domain. Additionally, several long-lived hydrogen bonds between the distal helix and CaM site D were found to stabilize the bound configuration, which dampened the fluctuations of peptide position found at other identified sites (A-C) as reported by RMSF and energetic analyses (Fig. 4.8).

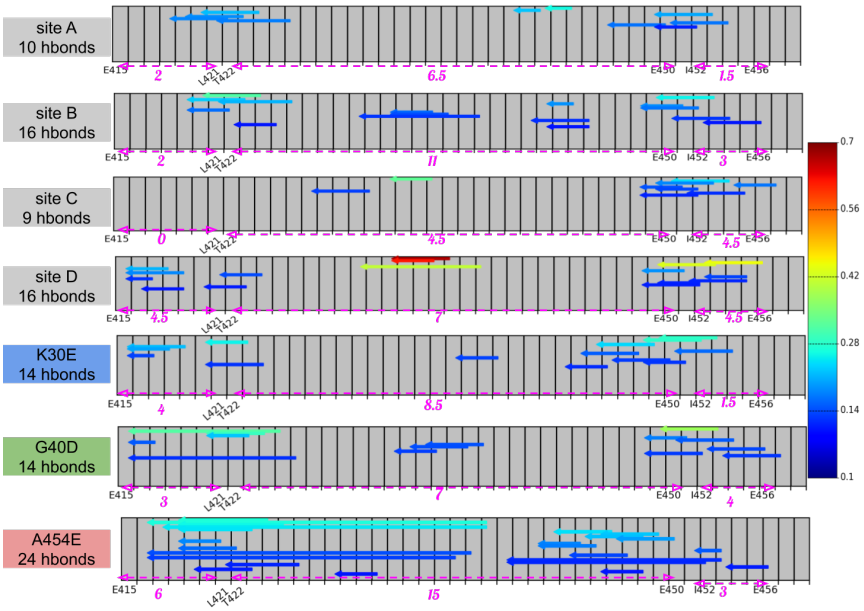


Figure 4.10: Backbone hydrogen bond analysis in the linker and distal helix region (E415 to I458). Each arrow represents one hydrogen bond with color indicating percentage of simulated frames with this hbond existed. Only hydrogen bonds exist >10% of simulation time are shown (we also show hdrogen bonds with >5% in Fig. S19). The whole region was divided into three subregions as indicated by the dashed magenta arrows below each subpanel. The subregion definition is consistent as the experimental HXMS data in Figure 8 in.[35] The number under the magenta arrow depicts the number of hydrogen bond in this subregion (one trans-subregion hydrogen bond contributes 0.5 to each subregion).

While we believe site D is the most probable site for distal helix binding, interactions with other potentially less-favorable sites could occur and contribute to the bound RD conformational ensemble. Such a diverse ensemble of strongly and weakly bound conformations is increasingly evident in complexes involving IDPs and globular targets[257,286] and may be adopted by CaN as well. It is also interesting that CD experiments in[46] suggested that the distal helix contact is abolished at temperatures above 38 degrees Celsius. It is tempting to speculate that the comparatively larger RMSFs of the bound distal helix configurations relative to the CaMBR, in addition to the weaker interaction energies, may render the distal helix interaction susceptible to melting.

Comparison of WT CaM with CaM variants

Strengthening the case for the involvement of the CaM site D in binding the CaN distal helix are our comparisons against two CaM variants with substantially impaired ability to relieve enzyme auto-inhibition in another CaM target,

Myosin Light Chain Kinase (MLCK).[259] CaM appears to relieve MLCK auto-inhibition[287] through binding the kinase's regulatory domain[288] and adopts a similar conformation as the CaN/CaM complex with CaM 'wrapping' around an α helical CaMBR motif (see also Fig. 4.9(a)).[46,285] Importantly, both appear to utilize secondary interactions beyond the CaMBR motif and it was shown by Van Lierop et al for MLCK that K30E and G40D mutations far from its CaMBR-binding domain prevented CaM-dependent kinase activity. These sites are localized to the site D region we identified for the distal helix in our study. Although the secondary interactions in MLCK likely involve CaM binding directly adjacent to the enzyme's catalytic domain,[289] we speculated that mutagenesis of these CaM residues could also impact CaN activation. Namely, we hypothesized that mutations of these residues would destabilize distal helix binding. We confirmed this in our computational model by demonstrating less favorable distal helix binding scores, and validated these predictions via pNPP assay.

Assessment of phosphatase activity

To challenge our hypothesis that impaired distal helix binding to CaM reduce CaN activity, we used kinetic phosphatase assays with the substrate pNPP on WT and the aforementioned CaM mutants. The Michaelis constant, K_M, obtained from these experiments informs on the ability of the catalytic site to bind and dephosphorylate pNPP. This substrate is specific to the catalytic site due to its low molecular weight, which allows for a probe of the extent to which CaM binding removes the AID. We reported significantly higher K_M for both K30E and G40D, thus these mutants evidence weaker distal helix binding that impedes removal of the AID from the CaN catalytic site. As a result, the CaM variants reduce the CaN catalysis of the dephosphorylation reaction, which can be interpreted as the AID competing with pNPP at the catalytic site and yielding a reduced apparent substrate affinity. This loss in affinity coincides with 40% increases in K_M reported for CaN A454E relative to WT CaN,[46] which were attributed to impaired distal helix formation. It should be noted that the small pNPP molecular is a preferable candidate for assessing distal helix binding, as opposed to common peptide-based dephosphorylation targets like RII.[258] Namely, the phospho-peptide binds to a site outside the active site (the LxVP site), therefore its binding, and hence K_M, would be unaffected by mutations in the distal helix region. pNPP, on the other hand, binds directly to the active site. Mutations in the distal helix region that disrupt its folding and allow the AID to bind to the active site would result in reduced pNPP binding (higher K_M). This explanation has been used by earlier authors studying the inhibitory properties of the AID as a peptide.[258]

Tether-model of CaM-dependent CaN activation

We recognize that a shortcoming of our modeling approach is that it is limited to simulations of CaM complexes with fragments of the CaN regulatory domain, whereas distal helix binding's effects on CaN activity are coupled to the entire regulatory domain and specifically, the AID. We therefore discuss a qualitative description of 'linker' dynamics of the regulatory domain appropriate for the AID-dependent inactivation of CaN. Specifically, we speculate that we can describe extents of CaN inactivation based on the AID's effective concentration at the CaN catalytic site as determined by the formation of distal helix/CaM interactions. This effective concentration is controlled by the tethering of the AID to CaN, which effectively confines the AID to a smaller volume (than free diffusion) that results in a higher interaction probability with the active site.[290] We use this effective concentration perspective to qualitatively assess how distal helix interactions with CaM impact CaN activity, as explicit all-atom simulations of the complete RD are prohibitively expensive. Here we leveraged previous theoretical models of protein activation[207,208] by describing AID binding to the CaN catalytic domain as an intra- PPI. This PPI leverages a molecular tether (the regulatory domain) to enhance the *local* effective AID (p) concentration near the catalytic domain.

To illustrate this principle in CaN, we provide a basic extension of a linker-dependent modulation model we recently applied to the calcium-dependent troponin I (TnI) switch domain binding to troponin C (TnC).[290] For this reaction, Ca^{2+} binding to TnC generates a conformation that can facilitate TnI binding:

$$\text{TnC} \overset{\text{Ca}^{2+}}{\rightleftharpoons} \text{Ca}\cdot\text{TnC} \overset{\text{TnI}}{\rightleftharpoons} \text{Ca}\cdot\text{TnC}\cdot\text{TnI}; \tag{4.3}$$

hence, increasing the TnI concentration would promote the generation of TnC·TnI with fewer equivalents of Ca^{2+}. In the tethered state, we estimated that the *effective* switch peptide concentration was an order of magnitude greater near its TnC target than would be expected for a 1:1 stoichiometric ratio of untethered (free) switch peptide to TnC. Accordingly, we experimentally confirmed that formation of the TnC/TnI switch peptide occurred at lower Ca^{2+} concentrations for the TnC-tethered TnI compared to a cleaved system in which both TnC and TnI were untethered.[290]

In a similar vein, we created a hypothetical linker-based model of CaN activation, based on a polymer-theory based model for the probability distribution of the linker spanning the CaMBR and AID domains (see Fig. 4.11). We introduce this model with several assumptions. Firstly, we postulate the CaN inhibition is dependent on the free AID concentration, of which the latter is determined by the RD 'tether' length. This tether length can assume three distributions associated with the CaM-free, CaMBR-bound CaM and CaMBR+distal helix-bound CaM, respectively. Lastly, for simplicity we assume that the distal helix binds CaM independent of the AID's bound state, though in reality we recognize there will be a competition between these two events.

Under these assumptions, we describe the effective [AID] at the CaN catalytic domain, based on the RD linker length in its CaM-free, CaMBR-bound CaM and CaMBR+distal helix-bound CaM states. We based this on an effective concentration model for tethered ligands suggested by Van Valen et al,[207]

$$[\text{AID}]^{eff} = \left(\frac{3}{4\pi\xi L}\right)^{3/2} \exp\left(-\frac{3D^2}{4\xi L}\right) \tag{4.4}$$

, where D is the distance between CaMBR and catalytic site, L is linker length, and ξ is the persistence length. The units of $[\text{AID}]^{eff}$ in Eq. 4.4 was achieved via fitting to existing experimental data. Namely, experimental assays were reported to investigate the competitive inhibitory effect of isolated AID peptide on CaN phosphate activity on substrate peptide.[291,292] In the assays, the reduction of phosphate activity was recorded as isolated AID peptide was added to intact CaN pre-incubated with CaM and substrate RII peptide. According to the experimental setup, there existed three competitive components that could bind the catalytic site of CaN: substrate RII peptide, isolated AID peptide and tethered AID from the intact CaN itself. Similar to P_{on} definition which represents the probability of switch peptide being on under the competitive binding of free ligand and tethered ligand to receptor in,[207] we

also defined a P_{on} that represents the percentage of CaN phosphate activity on substrate RII peptide under competitive binding from isolated AID peptide and tethered AID:

$$P_{on} = \frac{1 + \frac{[\mathrm{RII}]}{K_{d1}}}{1 + \frac{[\mathrm{RII}]}{K_{d1}} + \frac{[\mathrm{AID}]}{K_{d2}} + \frac{[\mathrm{tAID}]}{K_{d2}}} \tag{4.5}$$

, where [RII], [AID] and [tAID] are concentrations of substrate, isolated and tethered AID peptide, respectively. [RII] is set as 5 μM according to exprimental setup and the dissociation constant of substrate K_{d1} is assumed to be 10 μM. Tethered AID peptide is assumed to have same dissociation constant as isolated peptide with an experimentally estimated K_{d2} of 40 μM.[291,292] The fitting of Eq. 4.5 to experimental data in[291] with [tAID] as free pameter is shown in Fig. 4.11(b). [tAID] was fitted as 2.07 μM and this value is corresponding to $[\mathrm{AID}]^{eff}$ of 'CaMBR+distal helix-bound CaM' case in our tether model. In following tether model analysis, the $[\mathrm{AID}]^{eff}$ from Eq. 4.4 were scaled by [tAID] to give meaningful unit of effective AID concentration.

We first provide a rough estimate for the linker length through simulations of residues E415-M490 C-terminal to the CaMBR (see Fig. 4.11(a)). Starting from WT/A454E site D simulations, an optimized fragment (residues K459 to M490) containing AID built by TLEAP was fused to the C-termini of distal helix in the representative structure of first two most populated clusters. The complete structures were resolvated and simulated for $\approx 0.7\mu s$ as that described in Sect. 6. These simulations indicate that the WT AID to CaM distance is approximately 23 Å, versus approximately 40 Å for the A454E variant that precludes distal helix binding.

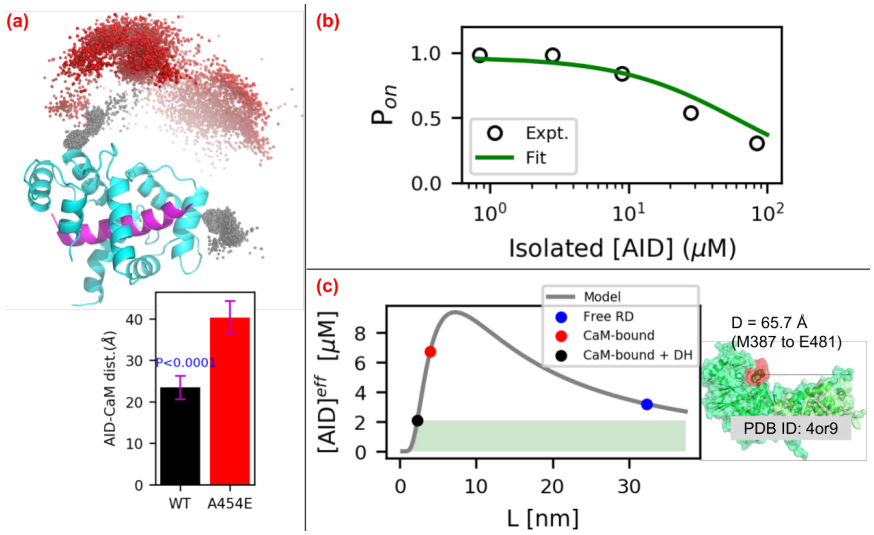


Figure 4.11: (a) Distribution of AID center of mass (COM) relative to the CaM/CaMBR complex. The black and red spheres represent the COMs of AID in WT and A454E cases, respectively. The lower panel depict distance between COMs of AID and CaM. The number above black bar are P values of WT case with null hypothesis that its values are equal to A454E case. (b) Fitting of the competitive-inhibitor model (Eq. 4.5) to experimental data from.[291] (c) Effective AID concentrations calculated via Eq. 4.4. The shaded green area represents effective [AID] that leads to CaN's activation. Right panel illustrates the assumed distance between CaMBR and catalytic site. The value is set as 66 Å in this study.

Based on these data, in Fig. 4.11(c) we demonstrate the effective AID concentration over a range of ligand lengths (L), predicted from Eq. 4.4 assuming $D = 66$ Å for the distance between CaM and the CaN AID binding site and $\xi = 3$ Å.[293] The black dot represents the CaMBR/distal helix (DH)-bound case, which has a tethered ligand length estimated from our simulation of approximately 23 Å or roughly 8 free amino acids. The blue dot represents free RD, which has ligand length of 95 residues (M387 to E481). The red dot represents the CaMBR-bound (no distal helix interaction as for the A454E case, in this case, the tethered ligand length estimated from our simulation as 40 Å). Based on these linker lengths, the corresponding effective [AID] concentrations for CaMBR-bound (A454E) states were 6.76 μM versus 2.07 μM for the CaMBR/distal helix-bound case. For the free RD case, the effective [AID] is 3.20 μM. This approximate model qualitatively captures the experimental trends in activity data reported in the literature,[46,294] namely that maximal CaN activation requires CaM binding.

There are several considerations that could improve the accuracy of this model. These include assumptions that the linker follows a random-walk chain distribution, that the catalytic domain does not attract and thereby bias the AID distribution and that the CaN molecule does not sterically clash with the linker

chain. Further, precise knowledge of the CaM distribution relative to the CaN B-chain would be needed to refine the effective linker lengths. Despite these assumptions, the model provides a qualitative basis for how RD mutations or variations in RD length could influence the efficiency of CaN (in)activation, similar to the model systems with synthetic linkers, as in.[295]

Limitations

We observed appreciable degrees of alpha helical and beta sheet character in the regulatory domain that were not evident in the CD data from.[35] A primary distinction between the modeling and experimental studies is that we used a much smaller regulatory domain fragment (residue A391 to I458) than the full length domain in Rumi-Masante et al,[35] owing to the computational expense. It is possible that there are different tendencies to form secondary structure, based on the regulatory domain length. Since we simulated only a small fragment of the RD domain, this might have increased the peptide's preference for alpha helical structure than would otherwise be observed in measurements of the entire RD. For instance, it has been shown that IDPs have length-dependent preference of residue compositions as longer IDP has more enriched K, E and P than short IDP,[296] implying the conformational properties of IDPs which are determined by sequence charge distribution[297] are also length-dependent. As a concrete example, Lin et al[298] reported that the 40-residue disordered amyloid beta monomer has reduced β-hairpin propensity when compared to the longer 42-residue monomer.

We additionally recognize that differences in ionic strength or solvent composition might influence the percentage of alpha helical character, although this seemed to be a modest effect in our simulations of the CaMBR alone.[257] Importantly, in that study, we reported negligible alpha helical character for that isolated CaMBR peptide, which suggests that our force field was not artificially stabilizing alpha helices, as had been an issue in earlier modeling studies of IDPs.[199,299] Nevertheless, the potential overestimate of alpha helical content for the isolated peptide is probably of little consequence, since the predicted bound distal helix was shown to confirm exhibit significant alpha helical content consistent with experiment.

We utilized REMD to sample the distal helix sequence in the absence of CaM; although REMD has been shown to perform well in terms of qualitatively describing conformational landscape, chemical shifts, α-helix stability for peptides of lengths comparable to the distal helix,[300–302] we did not have the means to experimentally validate the predicted apo ensembles. Nevertheless, the simulations provide testable hypotheses in terms of the α helical content. We additionally limited ourselves to subsets of the CaM surface for the docking search, which represented approximately 38% of the solvent-exposed surface area. However, given that the microsecond-length simulations were sufficient to reorient the site B configurations into the site D site, we anticipate the docked distal helix candidates reasonably sampled the thermodynamically-accessible

regions of the CaM surfaces. Although it has been demonstrated that RD binding to CaM is diffusion-limited, it is also possible that the intermediate complexes could be further optimized to form a final bound state, which would perhaps lead to more accurate assessments of critical intermolecular contacts and energy estimates. For the latter, alchemical methods such as thermodynamic integration may provide more accurate affinity estimates, albeit at a substantially greater computational expense compared to 'end point' methods like MM/GBSA. Lastly, more detailed simulations of the RD ensemble in the presence of the complete CaM and CaN structures are needed to more accurately characterize the effective AID distribution controlling CaN (in)activation.

There are several compelling directions to pursue that would provide essential clues governing CaM-dependent CaN activation. For one, we have predicted several contacts that appear to be involved in stabilizing the distal helix region; mutagenesis of these potential 'hotspots' on the CaM and measurements of subsequent CaN phosphatase could help validate this site. In addition, more detailed characterization of the RD intrinsically-disordered conformation ensemble would benefit future modeling. Given the difficulty in probing ensemble properties of IDPs, it is likely that modeling and experiment, such as fluorescence resonance energy transfer (FRET) labeling, should work in tandem toward this goal. Furthermore, relating these RD ensemble properties to the propensity for AID and CaN catalytic domain interactions would comprise an essential step toward a complete model of CaM-dependent CaN activation.

4.5 Conclusions

We have developed a computational strategy to elucidate potential binding poses for a secondary interaction (the 'distal helix') between the CaN regulatory domain and CaM that is apparently essential for competent CaN activation. We combined REMD simulations of isolated distal helix peptides, protein-protein docking of the distal helix peptides to the CaMBR-bound CaM surface, and microsecond-scale MD simulations of candidate poses to implicate a so-called CaM site D in binding the CaN distal helix. The predicted site D region is in part stabilized through direct interactions with K30 and indirectly through G40, which is consistent with experimental probes of a CaM-activated enzyme, MLCK. We confirmed the predictions via pNPP phosphatase assay in which mutations of K30E and G40D in CaM caused reduced CaN activity compared with WT CaM. With these data, we provide a qualitative model of AID-dependent CaN activation, which can be used to further refine potential molecular mechanisms governing the activation process and susceptibility to missense mutations. Importantly, our data suggest a potentially novel mechanism of CaM-dependent target regulation whereby interactions distal from the canonical CaM-peptide binding motif control target auto-inhibition. Given the broad range of physiological processes mediated by CaM binding to intrinsically disordered target proteins,[183] the mechanistic details of CaN activation in

this study may extend to diverse systems, including channel and cytoskeletal regulations.[183,303]

Chapter 5 Thermodynamic of Cation binding in SERCA and Its impact on Enzyme Functions

- This chapter is based on "Sun, B.; Stewart, BD.; Kucharski, AN.; Kekenes-Huskey, PM. *Journal of Chemical Theory and Computation* **2019,** *15,* 2692–2705"

5.1 Introduction

Sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) is a 110-kDa transmembrane cation pump which actively transports Ca^{2+} ions into the SR/ER by utilizing energy released from adenosine triphosphate (ATP) hydrolysis.[47] SERCA has been widely studied for its role in returning intracellular Ca^{2+} to basal levels following stimuli that elevates Ca^{2+} content.[48] The pump's catalytic cycle is roughly characterized by four states comprising a sequential cycle $E1 \rightarrow E1P.2Ca \rightarrow E2P.2Ca \rightarrow E2$. In E1, the Ca^{2+} binding sites are exposed to the cytosolic space, whereas the E2 conformations expose the low-affinity Ca^{2+} sites toward the SR/ER lumen. The transition between E1 and E2 is driven by ATP hydrolysis at residue Asp 351[47] following Ca^{2+} binding, for which E1P and E2P are the respective phosphorylated states of the enzyme. Accompanying transitions between catalytic states are prominent changes in its ten transmembrane (TM) helices as well as the cytosolic actuator (A) domain, nucleotide-binding domain (N) and phosphorylation domain (P). Many of these changes have been resolved through x-ray crystallography.[304–307]

Based on available structural models of the protein and a wealth of biochemical studies,[308–315] reaction schemes linking the E1 and E2 states are beginning to emerge. Utilizing SERCA vesicles coupled with spin label molecules, Inesi et al observed changed electron spin resonance spectrum upon Ca^{2+} binding that revealed conformational changes in the enzyme.[312] Dupont and co-workers similarly measured changes in intrinsic fluorescence upon Ca^{2+} binding and further proposed a two-step Ca^{2+} binding process to high affinity sites evidenced by the pumps slow rate of fluorescence changes.[310,311] Additionally, conformational changes linking the E1 and E2 states were explored by kinetic studies of intrinsic fluorescence changes upon Ca^{2+} binding and release.[308,309] To probe molecular determinants of Ca^{2+} binding in the pump, an E309Q mutant bound with two Ca^{2+} in its phosphorylated state was determined via x-ray crystallography,[316] which revealed that the altered TM arrangements caused by the mutation leads to impaired pump functionality. With respect to cation binding affinities, Inesi et al measured Ca^{2+} binding and stoichiometry to SERCA vesicles in via chromatography,[317] while others have probed the binding of the non-cognate Mg^{2+} and K^{+} ions via intrinsic fluorescence changes in SERCA.[318,319]

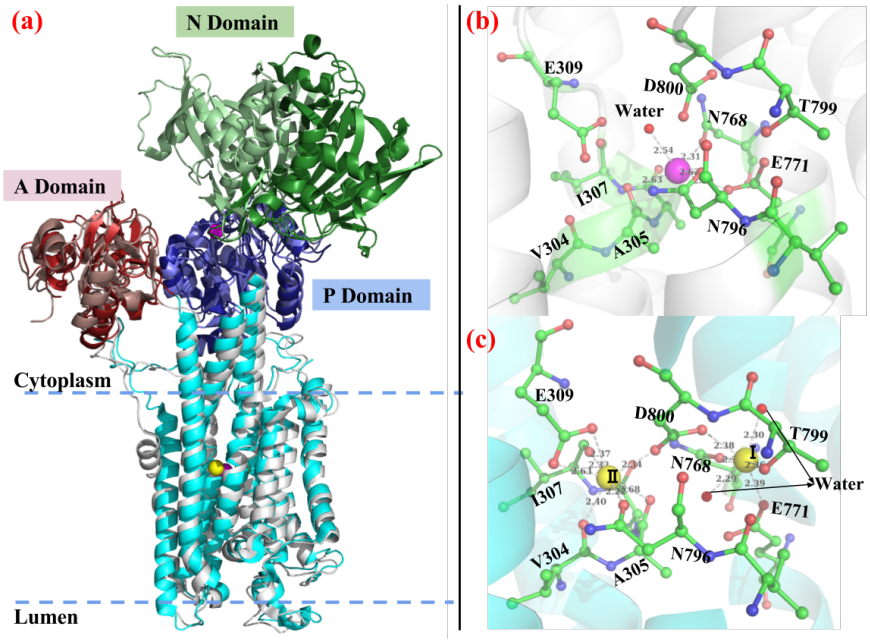


Figure 5.1: Comparison of Ca^{2+}-bound (PDB ID: 1su4) and Mg^{2+}-bound (PDB ID: 3w5b) SERCA crystal structures. (a) Superimpose of two crystal structures with 1su4/3w5b cytosolic domains (A, P and N) colored in red/salmon, blue/slate and green/palegreen. The TM helices are colored in cyan and grey for 1su4 and 3w5b, respectively. (b-c) Mg^{2+} and Ca^{2+} binding site comparison. In this orientation, sites I and II are right and left, respectively. Mg^{2+} and Ca^{2+} are represented as magenta and yellow balls. The distance between coordination oxygen atoms and cation are also shown. The oxygen atoms of crystal water molecules are shown as red balls.

In complement to experiment, extensive molecular dynamics simulations have uncovered many aspects of cation binding and SERCA function that are difficult to probe via experimental approaches. Huang et al first explored the Ca^{2+} binding pathways to SERCA TM sites via MD simulation and observed the cooperative binding of two Ca^{2+} ions.[320] Kekenes-Huskey et al performed MD simulations on SERCA with and in the absence of Ca^{2+} to examine gating of Ca^{2+} binding by E309,[321] in addition to estimating ion binding free energies and kinetics. Espinoza-Fonseca et al evaluated the interaction energy between $Ca^{2+}/Mg^{2+}/K^{+}$ and SERCA binding sites based on MD trajectories and reported that Ca^{2+} has the most negative interaction energy while K^{+} has the least negative value.[59] In addition to these initial studies on cation binding to the SERCA pump, more recent studies have probed mechanisms of SERCA function and its modulation by regulatory proteins and drugs.[60,322–328]

Although studies have been reported on the roles of Mg^{2+}, K^{+} on SERCA binding, less has been done to provide a thermodynamic basis for their effects on SERCA function. This is of particular importance, as intracellular K^{+} and Mg^{2+}

concentrations are orders of magnitude larger than the roughly sub micromolar Ca^{2+} concentrations found in typical cells. Here, simulation studies of cation binding to molecular sites that resemble high affinity, amino acid-based motifs have been informative. Dudev et al for instance constructed cation binding sites using model compounds and calculated cation binding energies via quantum mechanics.[329] Ye et al designed a theoretical framework in combination with MD simulations to calculate cation binding free energies and isolate the energetic contribution from geometric and confinement effect.[330] Implicit models that describe Ca^{2+} binding thermodynamics as via ion density-based formalisms have also been reported, including reference interaction site modeling[331] and density functional theory.[92] Besides these methods relied on explicit binding site configurations, Nonner et al developed the MSA model in which the binding site is treated as confined filter filled with coordination oxygens. A variant of density functional theory called the mean sphere approximation (MSA) approach has proven effective in rank-ordering the binding of cationic species to oxygen-rich binding domains such as EF-hands in β parvalbumin (β-PV)[94] and Ca^{2+} channel selectivity filters,[93,332] through describing the electrostatics and hard-sphere contributions to the chemical potential of partitioning ions into oxygen-rich 'filters'.

Our study has therefore focused on utilizing MD derived data with thermodynamic and MSMs to assess contributions of Mg^{2+} and K^+ binding on the SERCA turnover rate. Here we performed MD simulations of Ca^{2+}, Mg^{2+} and K^+-bound WT SERCA as well as the E309Q and N796A variants. These MD data provided structural information to assess cation binding thermodynamics via MSA to elucidate the molecular basis of SERCA's preference of Ca^{2+} over Mg^{2+} and K^+. Further, we relate these studies of E1 state ion binding to a state-based kinetic model of SERCA pump rate to determine the extents to which Mg^{2+} and K^+ facilitate or inhibit catalysis. With this approach, we provide a multiscale and molecular basis for cation binding to SERCA and impacts on pump function.

5.2 Results

We performed triplicate simulations of wild-type SERCA and its variants E309Q and N768A to probe the binding site coordination of the cations Ca^{2+}, Mg^{2+} and K^+, each replica was at least 100 ns in length. Ca^{2+}- and Mg^{2+}-bound configurations of the protein have been determined through x-ray crystallography, but to our knowledge, the binding of K^+ to the Ca^{2+} binding domain has only been resolved via simulation.[59] In this study, we critically examine the protein atoms and waters that directly coordinate bound cations. However, to ensure that our simulations of the intact protein are consistent with prior studies of the enzyme, we briefly summarize standard analyses of the transmembrane bundles and cytosolic domains in the Supplement (see Sect. S1.4).

Cation coordination in the Ca^{2+} binding region

In this study, we highlight structural and dynamic contributions of the Ca^{2+} and Mg^{2+} binding domains to the thermodynamics of ion binding. This is analogous to our approach for probing Ca^{2+}/Mg^{2+}-binding to the β parvalbumin (β-PV) protein,[94] for which we used MD simulation-derived structural data from the cation-bound configurations to parameterize a statistical mechanical model of fluid thermodynamics called mean sphere approximation (MSA). Specifically, we used the radius of the ions' inner coordination sphere and amino acid oxygens comprising the sphere to estimate the binding site volume and coordinating oxygen density for MSA. We use a similar strategy for SERCA in that we assess the coordination of a given ion based on the number of oxygens within six Å of the bound ion. In contrast to our previous study, we additionally include coordinated waters in the MSA calculation that are directly involved in stabilizing the ion.

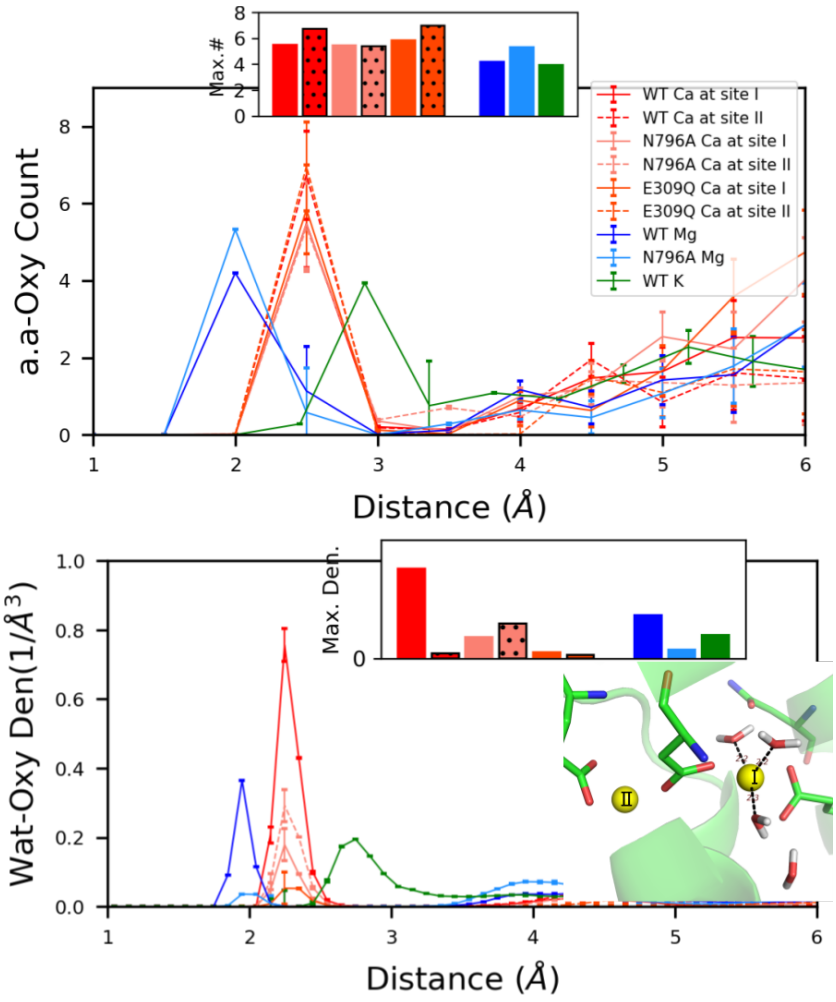


Figure 5.2: Radial distribution function (RDF) of amino acid and water oxygen atoms about bound Ca^{2+}, Mg^{2+}, or K^{+}. The distribution around Ca^{2+} for each individual case is shown in Fig. S9 for clarity. The inset bar graphs show the maximum number of coordinating amino acid oxygen and water density around the cation in each case (the bars with black dots represent site II Ca^{2+}). The coordinating waters with site I Ca^{2+} of WT SERCA is also shown. Ca^{2+} ions bound to WT SERCA tend to reflect the highest degree of coordination among the modeled systems.

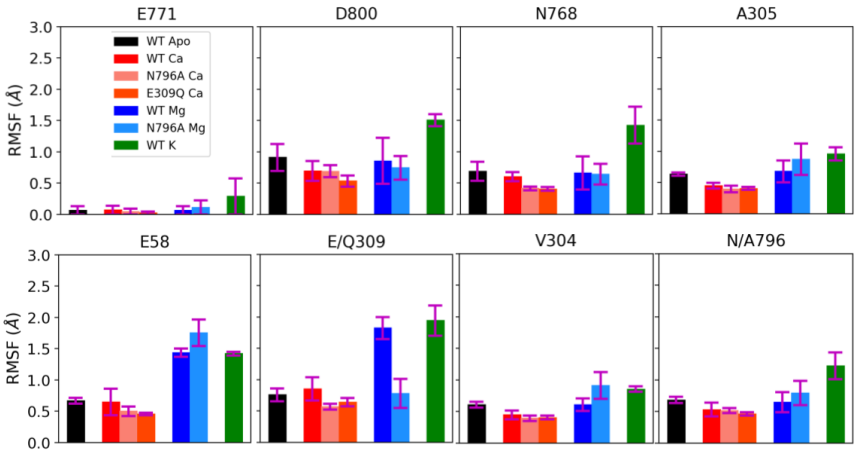


Figure 5.3: Heavy atom RMSF of key coordinating residues at the cation binding sites. The error bars were calculated from the triplicate trajectories for each case. The upper row represents site I residues while lower row were site II residues (E58 was not involved into coordination with cations in the crystal structure, although interactions with bound ions were predicted in our simulations, see Fig. S11). Ca^{2+}-bound configurations generally exhibit lesser mobility compared to the configurations with non-cognate ions Mg^{2+} and K^{+}.

In Fig. 5.2 we report radial distribution function (RDF)s of cation-coordinating oxygens from the simulations. These RDFs demonstrate that Ca^{2+}, Mg^{2+}, and K^{+} have varying degrees of coordination with amino acid oxygens within the binding site. We summarize the identity of the coordinating amino acids in Sect. S1.4, although from the perspective of MSA theory, only the number of contributed amino acid oxygens is important. The Ca^{2+} within site I was optimally coordinated with 8.5 oxygens on average (including three water oxygens). Ca^{2+} within site II maintained six amino acid oxygens pairings, similar to the x-ray crystal structure, but did not directly coordinate waters. By virtue of having more coordinating oxygens, we speculate that the Ca^{2+} ion in site I is bound more tightly than that found within site II. In contrast, for the N796A mutant, the site I Ca^{2+} had a reduced coordination number of six relative to over eight in the WT structure. Interestingly, for the site II Ca^{2+} in the N796A variant, the loss of coordination to site 796 was compensated by interactions formed with E58 and a bound water molecule (see Fig. S11(b)) to yield a greater coordination number than observed in the WT. For the E309Q mutant, two possible side chain rotamers of Q309 were investigated, as both rotamers were viable starting positions (see Fig. S11(c-d) for illustration of rotatmer directions). We found that these rotamers yielded identical Ca^{2+} coordination patterns for the two binding sites: site I had six coordinating oxygens versus seven for site II, while neither included bound waters (see Fig. S11(c-d)). Similar to the N796A site II Ca^{2+} case, E58 in TM1 also participated into the coordination with Ca^{2+} at site II in both rotamers of E309Q. Overall, these simulations reveal

that ion/oxygen pairing is remarkably labile between the sites I and II, and can incorporate bound waters to maximize Ca^{2+} coordination.

In contrast to Ca^{2+}, Mg^{2+} binds in a 'hybrid' site between sites I and II with an average six coordinating oxygens (including one water molecule, see Fig. S13(a)). K^+, on the other hand, binds site I with five coordinating oxygens (including one water molecule, see Fig. S13(c)), while K^+ at site II is highly dynamic and interchanges with water immobilized in the SERCA interior. Given that these ions are both positively charged and not remarkably different in size relative to Ca^{2+}, we had anticipated that the non-cognate ions might adopt higher coordination numbers in the native Ca^{2+} sites. Clearly the molecular simulations did not reflect this expectation; in Sect. 5.2 we provide a thermodynamic rationale via MSA theory for why these non-cognate ions present impaired coordination numbers.

The differences in cation-oxygen coordination patterns for the cases considered are accompanied by variations in the coordinating residues' mobilities relative to WT. These mobilities are measured as RMSF values in Fig. 5.3, for which the upper row represents site I and the lower row, site II. Generally, Ca^{2+}-bound systems have RMSF values for most residues around 0.5 Å, which are the smallest among the ions considered. To a certain extent, the reduced mobility could be interpreted as an indication of tighter and more favorable binding, although this would come at an entropic cost that is not explicitly estimated here. We were, however, surprised to see little change in RMSF for the Ca^{2+}-free (apo) state versus the Ca^{2+}-bound cases. One possible reason is that in apo state, waters fill the binding sites and stabilize residues via a hydrogen bonding network (see Fig. S7) - in this capacity, bound waters might 'prop' open the Ca^{2+}-binding domains to promote rapid incorporation of solvated Ca^{2+} ions from the bulk medium. Additionally, we found that WT SERCA and its variants presented negligible differences in RMSF upon Ca^{2+} binding, whereas the non-cognate Mg^{2+} and K^+ manifest significant RMSF increases across all residues (Mg^{2+} generally above 0.8 Å and K^+ above 1 Å). It is possible that the greater mobility of coordinating residues for the non-cognate ions are indicative of impaired coordination. We had anticipated that waters could be incorporated into the ions' binding domain to suppress fluctuations in amino acids contributing to coordination, much as was observed for the apo state. However, it is apparent the the strong electrostatic affinity for these ions with the coordination residues limited the volume within which waters could be incorporated. At a minimum, these data suggest that ion coordination is dynamic, with fluctuations on a nanosecond timescale (see Fig. S6), which ultimately may play a role in selecting Ca^{2+} over non-cognate ions.

Thermodynamics of ion binding at sites I and II

Molecular dynamics simulations provide qualitative insight to the binding of various ions in the binding sites of SERCA, but alone do not directly predict affinities. Therefore we explored MSA to semi-quantitatively estimate free

energies and selectivity of ion binding. MSA predicts chemical potentials of partitioning solvated ions into the SERCA binding domains, assuming a similar approach that was performed for the β-PV[94] protein. Namely, MSA theory estimates the chemical potential of ion binding, based on the assumption of finite sized ions and chelating oxygens confined to a spherical binding site volume. To utilize this method, we first compute oxygen RDFs about bound ions. These data provide oxygen filter densities and volumes, from which chemical potentials of ion partitioning into the binding site 'filter' from the surrounding bulk solution can be estimated. As shown in Fig. 5.4, Ca^{2+} at site I of WT case presents the most negative and therefore thermodynamically-favorable MSA- predicted chemical potentials across all cases, corresponding to the largest number of coordination oxygens. For the SERCA variants, site I and site II Ca^{2+} ions have modestly less favourable chemical potentials compared with site I Ca^{2+} from WT cases, as the relative values were approximately 0.3 kcal/mol higher. These are consistent with the comparable Ca^{2+} coordination numbers present among variants. Compared with Ca^{2+}-binding, Mg^{2+} binding at WT and mutant SERCA yielded more significantly disfavored chemical potentials at approximately 2 kcal/mol relative to WT site I Ca^{2+}. Among all cases, the K^{+} relative potential was the most positive at 4.2 kcal/mol, indicating that K^{+} is the least thermodynamically favored at TM sites. For Mg^{2+} and K^{+}, the MSA predicted potentials also correlated with the cation-coordination patterns as these two ions had reduced coordination number when compared with Ca^{2+} (5.2/4.7 for Mg^{2+}/K^{+} versus 8.5 for site I Ca^{2+}). These data show that MSA could capture the key factors governing cation affinities such as coordination number and binding site volume. In addition, with inclusion of coordinating waters in MSA, the predicted potential for WT site I Ca^{2+} is most favourable among all cases, which agrees with experimentally measured affinities. Our MSA results indicate that site I confers greater Ca^{2+} affinity due to extensive inclusion of water coordination. In Sect. S1.4, we utilized Grid Inhomogeneous Solvation Method (GIST) to assess the relative thermodynamics of water binding to the Ca^{2+} binding domain. In general, we found that when bound waters are present in the ions' coordination shells, the predicted free energies are on the order of -12 kcal/mol and thus very thermodynamically favorable.

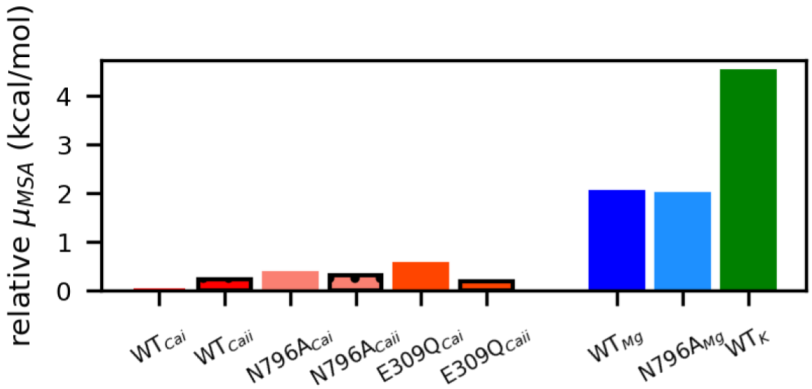


Figure 5.4: Mean sphere approximation (MSA)-predicted chemical potentials for each cation based on oxygen-coordination pattern (waters included) and optimal filter volumes derived from MD simulations. Potentials are given relative to Ca^{2+} bound to wild-type SERCA at site I. Ca^{2+} bound cases exhibit more favorable binding interactions compared to the binding of non-cognate ions, with the WT cases reflecting the most favorable potentials relative to the N796A and E309Q variants.

The membrane into which SERCA is embedded presents a negative surface charge owing to solvent-exposed phosphate head groups.[333] The corresponding negative electrical potential can attract positively charged ions and thereby increase their concentration near the transmembrane-bound SERCA, as we previously observed in.[321] Since the partitioning of cations into the SERCA binding sites depends on the composition of the surrounding electrolyte, we speculated that the local elevation of cations near the membrane surface would subsequently increase the concentration of bound cations. Additionally, the higher ionic strength could also reduce the desolvation energy and thereby further favor cation binding to SERCA, though this effect would likely be offset by screening electrostatic interactions between cations and the oxygens in the SERCA Ca^{2+} binding domains. To investigate this hypothesis, we determined the effective ion concentration near the membrane using a 1D solution of the linearized Poisson-Boltzmann equation,[163] $[i]_{eff} = e^{-\beta Z_i \zeta}[i]_{bath}$ where Z_i is the charge of ion and $\beta = \frac{1}{k_b T}$, $[i]_{bath}$ is ion concentration in the bath and ζ is the membrane potential.

Assuming $\zeta = -25$ mV,[24] we predicted that monovalent cation concentrations would be increased by 2.7 fold, anions would decrease by 0.37 fold, and the divalent Ca^{2+}/Mg^{2+} ions would increase by 7.4 fold. As shown in Fig. 5.5, at low bath [Ca], both Mg^{2+} and K^{+} were present in the SERCA Ca^{2+} binding domain with concentrations of 7.7 M and 1.7 M, respectively. As bath [Ca] was increased, Ca^{2+} partitioned into the binding domain in favor of Mg^{2+}; at roughly 2×10^{-5} M, the ratio of Mg^{2+} to Ca^{2+} was 1:1. The Ca^{2+} concentration

at which Mg^{2+}/Ca^{2+} was 1:1 varied by roughly 0.6 fold under the assumption of charged versus neutral membrane. In other words, according to our model, the local electrostatic environment about the membrane did not significantly impact the Ca^{2+} binding affinity. In contrast, in[321] we demonstrated that the negative surface charge densities of SERCA and the lipid enhanced the association rate of Ca^{2+} to the protein, thus the local electrostatic environment may have a greater contribution to ion binding kinetics than steady-state binding.

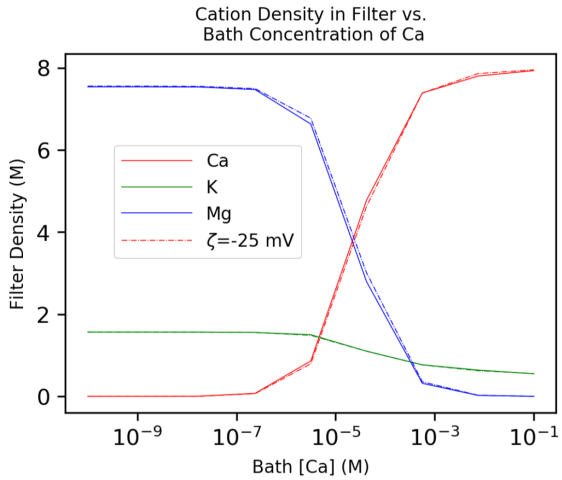


Figure 5.5: Cation concentration in the SERCA binding domains (assuming a filter volume = 0.33 nm^3 from our MD simulations), N_{oxy} = 7 (coordination number of Ca^{2+},[329] this value lies between the coordination number of site I, N_{oxy} = 8.5, and site II, N_{oxy} = 6.7, for Ca^{2+} determined in our simulation of WT SERCA) and with the cation solvation energy from.[92] [KCl] = 150 mM and [$MgCl_2$] = 2 mM. As cytosolic Ca^{2+} is increased, Ca^{2+} displaces Mg^{2+} bound to SERCA and reaches saturation at millimolar Ca^{2+} concentrations. Data are also presented assuming a membrane potential of $\zeta = -25$ mV, which locally increases bath cation concentrations by several fold according to Poisson-Boltzmann theory estimates.

Steady-state catalytic activity

Lastly, we relate our predictions of Ca^{2+} and non-cognate ion binding to the SERCA pumping rate. For the complete pumping cycle, two Ca^{2+} ions in the cytoplasm are transported into the SR/ER by first binding SERCA to its E1 state. This binding process was proposed by Inesi et al[9] to consist of two Ca^{2+} successive binding events via a cooperative mechanism. Subsequent steps include binding of MgATP, a slow conformational transition to the E2 state, release of Ca^{2+} ions into the SR/ER lumen, and a return to the E1 apo state. In practice, by accounting for the transition rates between SERCA conformational states, the time-evolution of each state can be described, which in turn can be related to the pump's cycling rate. However, generally the transitions between 'micro state' conformations within the E1 or E2 'macro' states, such as the Ca^{2+} binding steps in E1, are rapid relative to the slow E1 to E2 transitions. Hence, the micro states comprising the E1 and E2 stages are approximately in steady-state. This allowed us to describe the SERCA pump cycle rate as a two-state model for the E1 and E2 macro states (Eq. 5.8), which we used to relate experimentally and computation-determined binding constants to SERCA function.

In this state-based model, with exception to the undetermined transition rates between E1 and E2 (k_5^+, k_5^-, k_6^+ and k_6^-), the resting rate constants and substrate concentrations were taken from.[9,334] The above unknown rates were left as free parameters that were fitted to reproduce experimentally-determined turnover rates from[9] (see black data points in Fig. 5.6a)). This fitting accuracy was assessed as the difference between model predicted rates and experimental values reached minimum, as defined by

$$\text{Diff} \equiv \sqrt{\left(\frac{\sum_{i=1}^{N}(M_{r,i} - Expt_{r,i})^2}{N}\right)} \tag{5.1}$$

where $M_{r,i}$ is state model predicted pump rate at experimental [Ca], $Expt_{r,i}$ is the experimentally measured rate and N is the total number of data points. As shown in Fig. 5.6(a), the fitted Ca_{only} model (blue) reproduces the experimentally-determined SERCA turnover rates as a function of cytosolic Ca^{2+} concentration, which validates our state-based model.

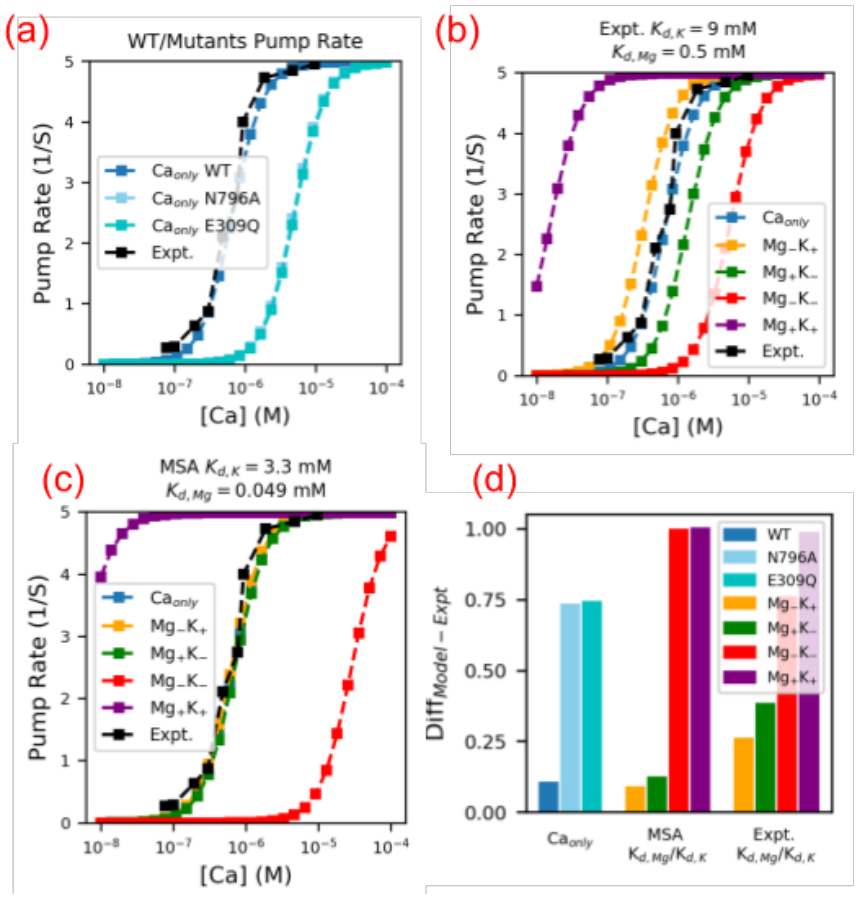


Figure 5.6: (a) Comparison of pump turnover rate between WT and SERCA variants using the Ca_{only} model with MSA predicted Ca^{2+} affinity for N796A and E309Q variants. The experimental data is from Inesi et al.[9] (b-c) Comparison of SERCA pump turnover rate of different state models against experimental data at [Mg] = 2 mM and [K] = 150 mM. In panel (b) the experimentally measured $K_{d,K}$ and $K_{d,Mg}$ in Table S4 were used. In panel (c) the MSA predicted $K_{d,K}$ and $K_{d,Mg}$ were used. (d) Normalized difference between each state model and experimental data for panels (a-c), as evaluated by Eq. 5.1. Our state-based models reproduce steady-state WT SERCA pumping rates and predict impaired rates for the N796A and E309Q variants.

Effects on non-cognate ions on steady-state behavior of SERCA

Since a primary focus of this study was to elucidate to which extent the non-cognate ions Mg^{2+} and K^{+} influenced the SERCA transport cycle, we introduced additional microstates representing the Mg^{2+}- and K^{+}-bound configurations. The resulting representations are summarized in Fig. S4 and differ in terms of whether the ions serve as inhibitors or intermediates. Using the fitted model from Fig. 5.6a, we introduced dissociation constants for the Mg^{2+}

and K^+ species. We considered two strategies for defining those constants: 1) using experimentally-determined values from Table S4 and 2) constants determined from rescaling of the MSA-predicted values. For 2), the MSA predicted Mg^{2+}/K^+ chemical potentials were first converted to dissociation constants via $K_d = e^{\frac{\mu_{MSA}}{RT}}$. Second, the K_ds were multiplied by a scaling factor, λ, that minimizes the difference between the MSA-predicted and experimentally-measured Ca^{2+} dissociation constants,

$$(Cai_{ref} - \lambda \cdot Cai_{MSA})^2 - (Caii_{ref} - \lambda \cdot Caii_{MSA})^2 = 0 \tag{5.2}$$

where $Cai_{ref}/Caii_{ref}$ and $Cai_{MSA}/Caii_{MSA}$ are site I/II Ca^{2+} dissociation constants from[9] (4×10^{-8} M and 4×10^{-6} M) and from MSA calculations (9.55×10^{-6} M and 1.38×10^{-5} M), respectively. Eq. 5.2 was minimized by $\lambda = 0.17$, thus yielding 4.93×10^{-5} M and 3.3×10^{-3} M for the MSA-predicted values of $K_{d,Mg}$ and $K_{d,K}$ for Mg^{2+} and K^+ (see also Table S3). Predictions of the SERCA cycling rate using experimentally-determined dissociation constants are shown in Fig. 5.6(b) and the rescaled MSA constants in panel (c). Both approaches indicate that the Mg_+K_- (green) and Mg_-K_+ (yellow) provide the best agreement with the experimentally-measured turnover rates, as these two models have relative smaller normalized difference values than other models, with the MSA-determined dissociation constants yielding the strongest agreement overall. Hence, the cycling rate data reported by Cantilina et al[9] was sufficient to eliminate two of the four proposed models. To discriminate between the remaining Mg_+K_- and Mg_-K_+ models, we next assessed the abilities of the respective models to reproduce steady-state Ca^{2+} binding data measured at various Mg^{2+} concentrations by Guillain et al.[318] In these experiments, both the E1.Mg and E1.2Ca states contributed a fluorescence signal indicative of Ca^{2+} saturation, therefore we report in Fig. 5.7 the combined probabilities of those states,

$$P_{Cabinding} = P_{E1.Mg} + P_{E1.2Ca} \tag{5.3}$$

where $P_{E1.Mg}$ and $P_{E1.2Ca}$ are the probabilities of the E1.Mg and E1.2Ca states (see Sect. S1.3). The experimental data (black) shown in Fig. 5.7 demonstrate that Ca^{2+} saturation naturally increases with increasing cytosolic Ca^{2+}, but importantly, saturation increases as Mg^{2+} is raised from 0 mM (circles) to 20 mM (solid triangles). These data indicate that Mg^{2+} locks SERCA into an E1 state in absence of Ca^{2+}. We plot in Fig. 5.7(a-b) the predicted $P_{Cabinding}$ data for the Mg_-K_+ and Mg_+K_-, respectively, as well as the fit in Fig. 5.7(c). We find that the Mg_-K_+ model provides the optimal fit with difference of 0.75 normalized to Mg_+K_- model. The Mg_-K_+ model correctly captures the plateau in Ca^{2+} saturation at Ca^{2+} concentrations below 1×10^{-7}, in contrast to the competing model. We note that as Mg^{2+} is increased to unphysiologically high concentrations ($\geq$5 mM), the slope of the experimentally-determined saturation curves decreases, which is indicative of a loss in Ca^{2+}-binding cooperativity. Our model does not directly consider ion-dependent modulation of Ca^{2+}

binding cooperativity, therefore this behavior is not reproduced in our predicted data and accounts for some of the error relative to experiment. Additionally, we predict a greater population of the Ca^{2+}-bound state at increasing Mg^{2+} levels than is experimentally observed, which accounts for the remainder of the error. Nevertheless, we find that $Mg_{-}K_{+}$ model provides the best agreement with experimental data, especially within physiological Mg^{2+} concent-rations. Therefore, Mg^{2+} most likely acts as inhibitor and K^{+} as an intermediate in the SERCA pumping cycle.

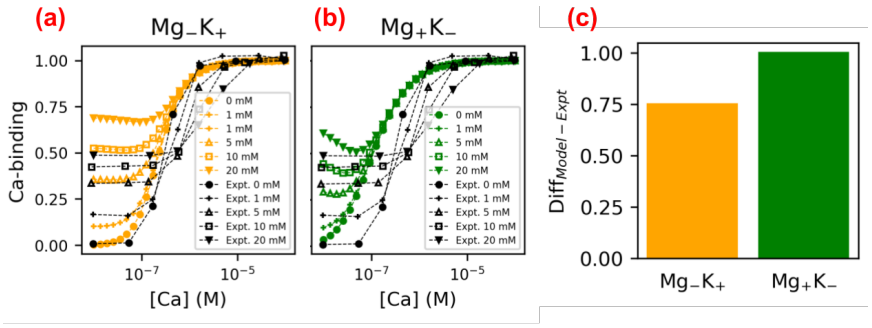


Figure 5.7: (a-b) Comparison of [Ca] dependence of Ca^{2+}-SERCA binding of $Mg_{-}K_{+}$ and $Mg_{+}K_{-}$ models at varying Mg^{2+} concentrations ([K]= 150 mM) against experimental data from.[318] Experimentally measured $K_{d,K}$ and $K_{d,Mg}$ were used. (c) Normalized difference between state models and experimental data as evaluated by Eq. 5.1. The $Mg_{-}K_{+}$ model, which assumes Mg^{2+} and K^{+} act as inhibitors and agonists, respectively, provide the best fit to experimental data.

Lastly, as a demonstration of how MSA-predicted Ca^{2+} affinities could be applied to infer changes in SERCA pumping function, we compared the turnover rates between WT and two variants using relative affinity estimates from Sect. 5.2. To be consistent with the procedure of obtaining Mg^{2+}/K^{+} affinity from the MSA potentials, the same scaling factor $\lambda = 0.17$ was applied to the MSA-predicted affinities for the E309Q and N796A variants (Ca^{2+}affinities at site I/II are 2.98×10^{-6} M/2.81×10^{-6} M for N796A and 4.06×10^{-6} M/2.25×10^{-6} M for E309Q). Relative to WT SERCA for which the half-maximal pump rate (approximately 2.5 per second) occurs at 6×10^{-7} M [Ca], the weaker affinities exhibited by the E309Q and N796A variants right-shift the half-maximal rate to $\sim 1 \times 10^{-5}$ M [Ca]. In other words, the SERCA variants are essentially non-functional within physiological Ca^{2+} concentrations (1×10^{-7} to 1×10^{-6} M[335]), which is consistent with experimentally-observed decreases in SERCA activity for the E309Q and N796A variants.[336] Although we recognize that the experimentally-measured activities arise from a culmination of factors beyond just the Ca^{2+} binding affinity in the binding domains, these data qualitatively indicate that MSA predicted affinities can then be used to rationalize steady-state pump turnover rates estimated from molecular-level simulations.

5.3 Discussion

Ion coordination and contributions to cation binding affinity

A key contribution from this study was our use of MSA theory to evaluate trends in Ca^{2+} and non-cognate binding in the SERCA E1 states. By using molecular dynamics simulations of the wild type and two site-directed mutations, we

could probe shifts in the binding site configurations - relative to the available crystal structures of the pump - that contribute to chelating cations. Overall, the MSA theory, when informed using molecular simulation data including water distributions, appears to be effective in rank-ordering ions by affinity (approximately -7 to -10 kcal/mol, see Fig. 5.8). Further, our state-based model of SERCA pumping function correctly captures cycling rates across physiological Ca^{2+} concentrations and predicts functional effects of site-directed mutations (N796A and E309Q).

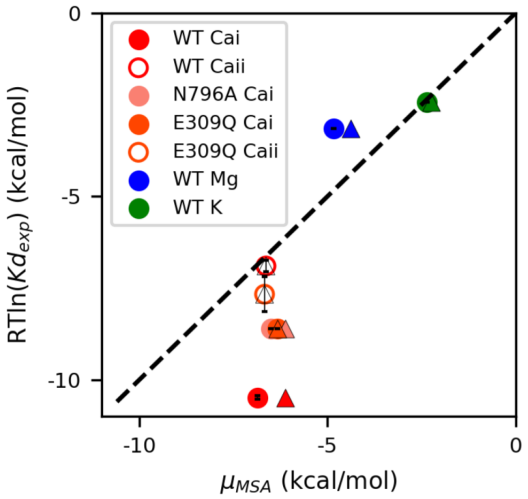


Figure 5.8: Correlation between MSA predicted chemical potential and experimentally measured binding free energy. The circles and triangles represent MSA results with waters and without waters taken into account, respectively. $RT = 0.593$ kcal/mol at room temperature and Kd_{exp} is experimentally measured dissociation constants in Table S4. The MSA chemical potentials correctly rank-order Ca^{2+} versus non-cognate ion binding.

We first applied this combined MSA/MD technique to an EF-hand containing protein called β parvalbumin to investigate factors contributing to its Ca^{2+} affinity and selectivity.[94] In both systems, the MSA/MD approach indicated that high affinity Ca^{2+} binding is afforded through the tight clustering of chelating oxygens with partial negative charges. Importantly, by predicting chemical potentials of ion binding, the MSA theory provided thermodynamic information about the relative competition between ions for the Ca^{2+} binding domains, which is generally unfeasible by MD alone. For instance, we determined that Mg^{2+}, and not K^{+}, is thermodynamically more probable to occupy the pump at low Ca^{2+} levels, but is disfavored as Ca^{2+} rises toward micromolar concentrations typical in eukaryotic cells. In principle, this approach could additionally account for variations in electrolye composition expected near charged lipid bilayers, but we found scant differences in binding assuming a surface charge of 25 mV (see Fig. 5.5).

Consistent with our earlier findings for Ca^{2+}-binding to β parvalbumin,[94] MSA predicts increasing stability of the Ca^{2+} ion as the number of coordinating oxygens is increased. This trend reversed as the oxygen density increased beyond roughly six per 2.2×10^{-1} nm^3, which is presumably when the volume is insufficient to accommodate all coordinating oxygens. For the E309Q and N796A variants, we potentially reduced the number of oxygens that could directly coordinate Ca^{2+}. Based on the MD simulations for E309Q, however, it appeared that the loss of one carboxylic acid oxygen was compensated by

a new interaction with E58 that maintained the same coordination number as the wild-type structure. This suggests that there is some degree of flexibility in coordination partners for the ion, which may facilitate the E309 residue's role in kinetically gating the binding site.[321] In contrast, the N796A mutation was not offset by a nearby available oxygen, thus the predicted chemical potential was less favorable. **Therefore, our data suggest that the MSA could reasonably rank-order Ca^{2+} binding stabilities based on structures of the immediate coordination shell** based on Fig. 5.4, but is less sensitive to broad ranges of binding affinities reported in the literature (see Fig. 5.8).

We note that our predictions of favorable binding stabilities for Ca^{2+} at site II for the N796A and E309Q variants are at odds with Inesi et al,[337,338] who suggested based on SERCA ATPase activity that the site was incapable of binding Ca^{2+}. Surprisingly, a structure of the E309Q variant (PDB ID: 4nab[316]) deposited in the protein databank contains Ca^{2+} at both binding sites. Ostensibly, the E309Q variant has a much lower, but nonzero affinity for Ca^{2+}. Thus, it is likely that the MSA model does not sufficiently penalize nonoptimal Ca^{2+} coordination, or reflect changes in internal strain that might disfavor coordination.[94]

We also found that including water distributions made a modest improvement in chemical potentials, as per Fig. 5.8. Based on Fig. S5, the narrow binding site volumes presented in sites I and II favor Ca^{2+} and Mg^{2+} over K^+, while both divalent ions are increasingly stabilized by greater densities of coordinating oxygens and waters. It is interesting that the water-free MSA calculations indicate Ca^{2+} could be more stable in a binding site volume of $2.2 \times 10^{-1}\,nm^3$, which is smaller than the volume presented in the MD simulations, albeit with a fewer number of oxygens (approximately five versus six). However, when water is considered, the stabilities at the smaller binding volumes are disfavored. It is possible that there is a kinetic advantage to having mobile waters involved in coordination, which could favor more rapid exchange of bound ions with the surrounding solvent. It is also possible that the trend can be explained on a thermodynamic basis, namely that by preserving waters that coordinate in the ion in bulk solvent, the desolvation cost upon binding SERCA are reduced, which should increase the apparent binding affinity. An intriguing possibility is that the Ca^{2+} affinity could be indirectly tuned by controlling the density of binding site waters.

The primary function of the SERCA pump is to transport cytosolic Ca^{2+} into the sarcoplasmic or endoplasmic reticulum, therefore we verified that MSA would indeed predict an unfavorable change in chemical potential based on conformational changes induced in sites I and II upon forming the E2 state. Since the focus of this study was on the E1 Ca^{2+} binding thermodynamics, we did not perform MD simulations of the E2 state. Nevertheless, based on the crystal structure of SERCA/beryllium fluoride complex(PDB ID: 3b9b[307]), which represents the pump's E2P state, we illustrate in Fig. 5.9 that drastic changes in the Ca^{2+} binding site configuration culminate in a significant reduction in oxygens that could potentially coordinate Ca^{2+}. We further evaluate the

binding stability for a hypothetical Ca^{2+} bound between the labeled coordination groups and find that the MSA values are not only more positive than those of the E1 state, but are additionally greater than 0 kcal/mol. The positive values of approximately 1 kcal/mol indicate that Ca^{2+} binding at these position is less thermodynamically favorable than partitioning into the surrounding bulk electrolyte. In other words, when SERCA transitions into the E2 state, it is thermodynamically preferred for Ca^{2+} to vacate the binding site in favor of the reticulum lumen. Along these lines, mutations that alter the free energy difference for the Ca^{2+} sites in the E1 and E2 configurations could affect the efficiency of the ATPase.

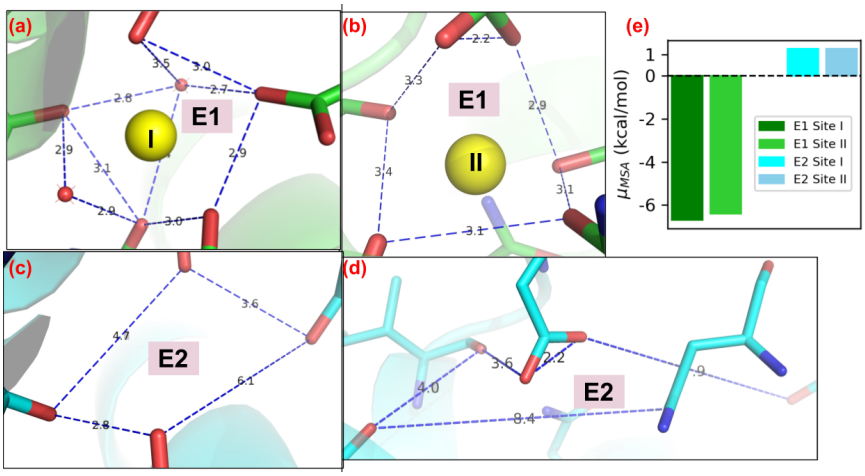


Figure 5.9: Comparison of Ca^{2+} binding sites in E1 state (a-b) and E2 state (c-d) from crystal structures of PDB 1su4 and 3b9b,[307] respectively. Ca^{2+} and water molecules are shown as yellow and red spheres, respectively. The blue dashed lines in (a-b) outline the overall shape of Ca^{2+} coordinating spheres in E1 state while in (c-d) they depict assumed Ca^{2+} binding sites in E2 state. Panel (e) depicts MSA predicted Ca^{2+} chemical potential for these two sites. In E1 state the number of coordinating oxygens are seven and six for site I and II while these values are assumed as four for both sites in E2 state.

It has been suggested that the number of cations (Ca^{2+}, Mg^{2+}, and K^{+}) bound to SERCA is approximately constant across its many conformational states.[339] In other words, Ca^{2+} binding is offset by Mg^{2+} and K^{+} dissociation. ATP is bound to SERCA in complex with Mg^{2+},[340] but there also exists structural[341] and activity[342] data that confirm Mg^{2+} binds in the pump's transmembrane region. Toyoshima et al,[341] for example, obtained the x-ray crystal structure of the pump with a single Mg^{2+} bound at a 'hybrid' transmembrane site, for which the ion is coordinated by ~5 oxygens with distances of approximately 2.0 Å. Since Mg^{2+} bears the same charge as Ca^{2+}, but with a modestly smaller radius, it is somewhat surprising that Mg^{2+} preferrentially binds at an intermediate site between the canonical Ca^{2+} binding sites I and II. We attribute the thermodynamic preference for Ca^{2+} at those sites based on two factors: the higher desolvation energy for Mg^{2+} relative to Ca^{2+} (779.94 kT versus 648.65 kT[92]) and the smaller binding site volume for the intermediate site versus sites I and II. With respect to the latter factor, in principle, Mg^{2+} could reduce the site I and II volumes by pulling the chelating oxygens toward the smaller ion, as we previously observed in.[94] For the EF-hand (helix-loop-helix) containing protein, we found that constricting the binding loop region introduced greater strain for Mg^{2+} relative to Ca^{2+}, which we suggested would reduce the overall binding affinity for Mg^{2+}. Although we did not explicitly evaluate strain that could be introduced upon Mg^{2+}-binding for SERCA, we anticipate an even greater cost for reducing the binding site volume, since the chelating amino acid are firmly tethered

to relatively immobile transmembrane helices. It is important to emphasize, though, that the non-optimal volumes for Mg^{2+} binding at sites I and II does not preclude the ion from binding, as we demonstrate high binding probabilities at very low Ca^{2+} levels (see Fig. 5.7).

Consistent with our earlier findings in β-PV, K^+ is disfavored from binding the native Ca^{2+} binding volume based on its significantly larger radius and smaller charge.[343] However, K^+ and other monovalent cations have been demons-trated to stimulate SERCA function, with K^+ being the most efficient agonist.[319,339,342] Moreover, based on molecular simulations performed in the presence of transm-embrane bound K^+, Espinosa-Fonseca et al suggest that K^+ facilitates the pump's E2 to E1 transition through its stabilization of the E1 state[59] In our simulations, we found that of the two K^+ ions placed in the Ca^{2+} binding domains, the site I K^+ remained bound with approximately four coordinating oxygens with ion/oxygen distances exceeding 4 Å. Based on the MSA predictions, although K^+ has a favorable ($\mu < 0$) binding potential that is consistent with its sub-millimolar binding affinity at site I,[319] Mg^{2+} and Ca^{2+} are considerably more likely to be bound (see Fig. 5.5). Meanwhile, water frequently exchanged with K^+ in site II, which is suggestive of a low affinity for the cation.

Based on our comparison with proteins that selectively bind K^+, such as the KcsA K^+ channel, selective binding of K^+ over competing Ca^{2+} might best be afforded through placement of carbonyl oxygens at sufficiently large distances to accommodate the K^+ ion's larger volume. Namely, in K^+ channels, oxygens from the backbone or side chains of amino acid forming the selectivity filter are exquisitely arranged to achieve precise pore size control and K^+-oxygen interaction strength.[344] In contrast, it is likely that K^+ exerts its agonistic effects on SERCA through binding the cytoplasmic domain, as was evidenced in a crystal structure determined by Sorenson et al, based on a bound K^+ ion in the P-domain.[306] According to the K^+-oxygen coordination pattern shown in Fig. S10, the MSA predicted potential for the P domain K^+ is -2.38 kcal/mol, which is comparable to the values predicted for K^+ bound to site I (-2.35 kcal/mol). Consistent with this structural evidence, the E2P.2Ca dephosphorylation data indicate that K^+ stimulates the Ca^{2+}-release step in this state, whereby the lumenal Ca^{2+} affinity is reduced, rates of exchanging lumenal Ca^{2+} with lumenal solvated Ca^{2+} are accelerated,[345] and Ca^{2+} release is enhanced.[346] Similarly, dephosphorylation of the E2P state is blunted in the absence of K^+.[339]

Although the thermodynamics of K^+ binding are unfavorable relative to Mg^{2+} and Ca^{2+}, there remains the possible role of K^+ in shaping the kinetics of SERCA function. It has been speculated, for instance, that K^+ accelerates Ca^{2+} binding by first transiently occupying site I, after which it exchanges quickly with a Ca^{2+} at site II.[346] K^+-facilitated exchange could therefore permit faster incorporation of Ca^{2+} into site I, as opposed to the direct migration of Ca^{2+} into a site that is potentially only partially-formed.[346] This interpretation is consistent with our observations from molecular dynamics simulations, and was initially reported in simulations of Ca^{2+}-free SERCA by Musgaard et al.[347]

Relating cation affinity to ATPase activity

In our opinion, the bridging of molecular-level simulation and MSA thermodynamic data with state-based models represents a significant contribution toward multi-scale modeling of steady-state SERCA activity. Numerical and computational modeling of SERCA activity has spanned phenomenological frameworks, such as Hill-type models[348–350] to those representing distinct stages of the catalytic process as states.[9,334,351–353] Our implementation is based on the stepwise binding of Ca^{2+} originally proposed by Inesi et al,[9] but additionally considers Mg^{2+}- and K^{+}-bound states, followed by a reduction scheme to lump the E1 states separately from those comprising the E2 configurations. Significantly, the states defined in our model coincide with SERCA structures determined by x-ray crystallography and are ordered in a manner consistent with assays of SERCA function[354] In contrast, the state model proposed by Tran et al[334] assumed that ATP binding precedes the binding of two Ca^{2+} ions, which has not been experimentally verified.

Existing models of SERCA function have faithfully recapitulated the pump's activity. However, an advantage of our state-based model is that its alignment with experimentally-determined structures permits us to directly investigate how structural modifications might impact activity. For example, we performed MSA/MD simulations of two mutated SERCA variants that yielded modified Ca^{2+} binding constants that we used to predict SERCA activity (see Fig. 5.6). While we acknowledge that there are likely myriad changes to the pump's energetics following mutation that could be accounted for in the state-based model, our implementation here is a significant step toward structure-based modeling of SERCA activity. At a minimum, our fitted state model is consistent with steady-state pump rates data collected as a function of Ca^{2+} by Cantilina et al,[9] Ca^{2+}-saturated versus Mg^{2+},[318] as well as molecular simulations that predict Mg^{2+} inhibition of the pump.[59] Furthermore, by considering different linkages of K^{+}- and Mg^{2+}-bound states to the reaction scheme, we were able to determine that assumptions of Mg^{2+} inhibition and K^{+} agonist were most consistent with experimental data collected by Guillain et al.[318] In principle, multiscale models of SERCA activity that include structure-derived thermodynamic information could permit in silico investigations of how disease-associated SERCA mutations,[355,356] post-translational modifications,[357] and binding of regulatory proteins such as phospholamban[351] affect pump function.

Limitations

There are several limitations of our approach that could be addressed in future studies. Our study largely focused on conformational changes and energetics of the cations' immediate coordination sphere. Mg^{2+} and Ca^{2+}-bound structures (PDB ID: 1su4 and 3w5b, respectively) deposited in the Protein Data Bank indicate substantial differences in the conformations of the transmembrane bundle helices and cytosolic domains that will necessarily

contribute to the free energies of binding. For this reason, the free energy differences implied in experimental measurements of cation binding affinities reflect contributions from both ion coordination and the pump's different conformations. It is also important to emphasize that the MSA assumptions of a spherical binding volume within which amino acid oxygens are immobilized crudely approximates the actual binding site. Thus, the predicted thermodynamic quantities are most appropriate for rank-ordering different ion/binding site configurations. Additionally, there is evidence[317,354] that Ca^{2+} binding is cooperative, although here we treat the binding events independently. It may be possible to partially recover some of the cooperative behavior by modeling SERCA with only one bound Ca^{2+} and using end-point methods such as Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA)[358] to estimate energy changes upon binding a second Ca^{2+}. To our knowledge, however, a half-saturated Ca^{2+}-bound SERCA structure has not been determined, which would challenge efforts to validate predictions. Additionally, in the sequential binding model used by Inesi and others[9,307] ($E1 \rightarrow E1.Ca \rightarrow E1'Ca \rightarrow E1'2Ca$), the apparent kinetics governing the transition between the E1.Ca to E1'Ca states can vary depending on a variety of factors, including the presence the regulatory protein phospholamban.[351,359] Thus, extending this model to broader conditions and regulatory proteins would require careful consideration of conformational changes that might accompany the E1.Ca to E1'Ca transition. Lastly, the Ca^{2+}-binding steps in our model are assumed to be in steady state relative to the substantially slower transitions between the E1 and E2 states. In the event that the pump is subject to rapid Ca^{2+} oscillations, such as spontaneous Ca^{2+} spiking in glia,[360] the steady-state approximation may be unreliable.

5.4 Conclusion

We utilized molecular dynamics simulations, mean sphere approximation theory and state-based modeling to probe effects of Ca^{2+}, Mg^{2+} and K^+ binding on the SERCA pump cycle. The MD and MSA approaches indicate that favorable binding of Ca^{2+} in the wild-type SERCA configuration is facilitated through a high degree of coordination by amino acids comprising the binding sites, as well as significant contributions from water coordination. This coordination pattern appears to be impaired in the E309Q and N796A variants; using MSA theory, we found that the chemical potential of Ca^{2+} binding is less favourable relative to wild-type as a result. Hence, mutations near the Ca^{2+} binding domains that alter the coordination number, hydration and binding site volume can be expected to modulate Ca^{2+} affinity in a manner qualitatively described by the MSA theory. Similarly, the coordination patterns exhibited in the Mg^{2+} and K^+-bound structures led to less favorable binding estimates from MSA-theory. These findings were qualitatively consistent with measured affinity data reported in the literature.[316–319,337,346] Additionally, we developed a state-based model of SERCA activity that we used to: 1) relate Ca^{2+} binding affinities to the

SERCA pump rate; and 2) determine whether Mg^{2+} and K^+ are pump agonists or inhibitors. We found that the state model treating Mg^{2+} as an inhibitor and K^+ as an agonist (Mg_K_+) was most consistent with experimental data. Despite the limitations of the assumptions used for the various models, we believe that the approach provides an attractive framework for evaluating allosteric functional effects of ion binding on SERCA, which may be extendable to other Ca^{2+} transporters, such as P2X4.[361]

5.5 Methods

Construction of the SERCA systems

Molecular dynamics (MD) simulations on wild-type (WT) SERCA and two variants of SERCA: E309Q and N796A were performed; the latter variants were chosen given dimished site II Ca^{2+}-binding.[336] For the WT SERCA system, we considered apo (free of bound ions) Ca^{2+}-bound, Mg^{2+}-bound and K^+-bound states. Ca^{2+}-bound (PDB ID: 1su4[304]) and Mg^{2+}-bound (E1.Mg, PDB ID: 3w5b[341]) SERCA X-ray crystal structures were used as the starting structures. The apo SERCA system was constructed based on 1su4 with two Ca^{2+} ions removed as done in.[362] The structure of the site I or II bound K^+ SERCA has not been determined, thus we created the structure based on the Mg^{2+}-bound variant. For the N796A mutant, Ca^{2+}-bound and Mg^{2+}-bound cases were considered, based on mutating N796 to alanine. Similarly, for the E309Q mutant, we obtained two rotamers compatible with the binding site as evaluated through UCSF Chimera.[363] These E309Q rotamers were designated as "E309Q_r1" and "E309Q_r2". All SERCA cases considered in present study are summarized in Table S1.

The cation coordinating acidic residues E309, E771 and D800 were assumed to be deprotonated, while E908 was protonated, consistent with;[59] further a disulfide bond was introduced between residues C876 and C888. The system was inserted into a homogeneous 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipid bilayer of POPC via the Membrane Builder[364] within the CHARMM-GUI software.[365] This system was solvated via TIP3P waters[72] using a 20 Å margin perpendicular to the membrane. Counterions K^+ and Cl^- were added into the system via Monte-Carlo method to neutralize the system and maintain an ionic strength of 0.15 M. Both the solvation and ion-adding were performed via the Solvator module within CHARMM-GUI online-server. The final system contained ∼255,000 atoms including lipid bilayer with 461 lipids, and ∼59,000 TIP3P water molecules and was parameterized by the CHARMM36 force field.[202,203]

Molecular dynamics simulations

MD simulations were performed using NAMD.[366] The system was first subjected to an energy minimization process consisting of 2000 steps' steepest descent

(SD) and 2000 steps' adopted basis Newton rRphson (ABNR) algorithm. For each case, the minimized system was heated from 0 K to 303.14 K over 25 ps with 1 fs timestep via the Langevin thermostat, using randomized velocities for three triplicate preparations. For each replica, harmonic constraints were introduced during minimization and heating on protein side chains, protein backbone atoms, lipid heavy atoms and ions. The force constants of constraints were set to 5, 10, 10 and 1.0×10^1 kcal mol^{-1} Å^{-2}, respectively, and were gradually reduced to zero over five equilibration steps of 20 ps in duration. 100 ns production simulations was then performed on the equilibrated system with the Shake algorithm,[201] using 2 fs timesteps.

Simulation trajectory analysis

Standard trajectory analyses including RMSF and coordination oxygen/Ca^{2+} distances were computed using Lightweight Object-Oriented Structure Library (LOOS).[367] Coordination pattern analysis on each cation with oxygen atoms from nearby amino acids were performed in each binding site. This consisted of extracting residues within 20 Å from D800 (the shared residue of the conventional two Ca^{2+}-binding sites in SERCA) based on the Ca^{2+}-bound SERCA crystal structure (PDB ID: 1su4). The water density around cation was computed via the radial command implemented in CPPTRAJ[274] and analyzed via Grid Inhomogeneous Solvation Method (GIST) (as described in Sect. S1.4).

Mean sphere approximation (MSA)

The MSA is a mean-field model which estimates cation chemical potentials in electrolyte solution with finite-sized ions. In this study, the SERCA cation binding sites were treated as confined filters filled with oxygens from coordinating residues and water molecules. The MSA model calculates cation distributions between bulk electrolyte solution that minimize the chemical potential for partitioning ions into a finite size volume occupied by coordinating oxygens; these volume and number of oxygens were determined by MD as in Kucharski et al.[94] The free energy expression in this model was assuming negative-charged spherical domains was based on a chemical potential accounting for electrostatic and hard-sphere interactions between ions inclusive of oxygens, as described in.[93,368,369] In this representation, which is analogous to the Ca^{2+} binding domain in β-PV and calcium channels,[92–94] ionic species that have a negative chemical potential in the binding filter are thermodynamically favored to bind. We include in this approach a solvation contribution as estimated via generalized Born theory of ion hydration energies:

$$\Delta G_{GB} = \frac{z^2}{r}\left(\frac{e^2 N_a}{2 \cdot 4\pi\epsilon_0}\right)\left(1 - \frac{1}{\epsilon_r}\right) \tag{5.4}$$

where z is charge number, r is radii, e is electron charge, N_a is the Avogadro constant, ϵ_0 and ϵ_r are vacuum dielectric and the relative dielectric constant of

the solvent. Additional details are elaborated in the supplementary material of.[94]

In the SERCA system, the cation/SERCA configurations can present differing numbers of coordination oxygen atoms (see Fig. 5.2 and Table S2) and volumes, as shown previously for β-PV in:[94]

$$V_f = \frac{4}{3}\pi R^3 \tag{5.5}$$

$$R = r_{rdf} + r_{oxy} \tag{5.6}$$

where r_{rdf} is the radii of optimal coordination sphere which can be obtained from the cation-oxygen coordination pattern analysis based on MD simulations (see Fig. 5.2) while r_{oxy} is the radii of oxygen atom. Charges and radii values of all ions are taken from the Li-Merz work.[370] The specific parameters used in the MSA calculations are listed in Table S2.

State-based model of the SERCA pump cycle

SERCA pumping is characterized by two prominent states, E1 and E2, comprised of microstates corresponding to various conformations or bound states of the pump (Fig. S4). In the E1 state, two Ca^{2+} ions bind to SERCA through cooperative mechanism followed by the binding of MgATP.[9] We represent each state as by an ordinary differential equation of the form

$$ds_i/dt = \sum k_{ji} s_j - \sum k_{ij} s_i \tag{5.7}$$

where s_i is state i, and k_{ij} describe rates for transitioning between states i and j. The models tested in this paper are summarized in Fig. S4.While the cooperative Ca^{2+} binding mechanism in the E1 state is well-accepted, the Ca^{2+} release mechanism of the E2 state is not clearly resolved, therefore we use the state representations proposed in Tran et al.[334]

Based on the technique proposed by Smith et al,[371] we can applied rapid equilibrium assumption to fast processes (those within E1 or E2) relative to the slow rates of transitions between the E1 and E2 states to simplify the model. As shown in Fig. S4, states in dashed boxes were assumed to be in quasi-steady-state and thus were lumped together. Hence, the original multi-state models were simplified into a two-state model (see bottom Fig. S4) in which new apparent rate constants were derived based on original rate constants. The resulting equations are summarized in the supporting information in.[372] For the lumped 2-state model, the turn over rate is derived as in Tran's paper:[334]

$$V = \frac{\alpha_1^+ \alpha_2^+ - \alpha_1^- \alpha_2^-}{\alpha_1^+ + \alpha_1^- + \alpha_2^+ + \alpha_2^-} \tag{5.8}$$

where $\alpha_i^{\pm}$ are apparent rate constants between lumped two states (the expressions are given in the supporting information in[372]). The final parameters used are listed in Table S3.

Chapter 6 Conclusion

Ca^{2+} is an important secondary messenger in cellular life. Perturbation of Ca^{2+} signaling can result in dysfunction of heart, brain and immune system. In this thesis, we utilized various computational technologies including partial differential equation (PDE)-based continuum modeling, molecular dynamics (MD), Brownian dynamic (BD), mean sphere approximation (MSA) and protein-protein docking to gain understanding of three prototypical aspects of Ca^{2+} signaling: cellular Ca^{2+} diffusion, Ca^{2+}-mediated protein-protein interaction (PPI) and pump-facilitated Ca^{2+} transport.

Signal transduction requires messenger to diffuse from the place where they are released to target proteins at which they are bound. At the cellular scale, the diffusion rate of Ca^{2+} is affected by charged crowders such as proteins and nucleic acids. In addition, buffer proteins that bind Ca^{2+} ions can hinder or facilitate Ca^{2+} diffusion. To understand how Ca^{2+} diffusion is affected by these factors, we simulated Ca^{2+} diffusion in mesoporous silica channels using partial differential equation (PDE)-based continuum model. These silica channels have comparable dimensionalities as cellular compartments, and more importantly, the silanol function groups (Si-OH) at silica materials surface enable us to imitate the effects of charged crowders and Ca^{2+} adsorption. We found that the presence of negative surface charge facilitates Ca^{2+} diffusion. This effect was attenuated by increasing ionic strength that shields the favorable interactions between negative charges and Ca^{2+}. The Ca^{2+} adsorption hindered Ca^{2+} diffusion at basic pH environment, with an extent less than expected, as a pK_m increased by 2 unit caused $\sim$10% reduction of Ca^{2+} effective diffusion constant. We also characterized the amount of Ca^{2+} ions adsorbed by silica surface. These modelings provided quantitative descriptions of Ca^{2+} diffusion in a cell-like environment, which sheds insights into the spatial distribution of cytoplasmic Ca^{2+}. Moreover, our model shows the potential of probing Ca^{2+} diffusion in real cell in which the membranes have non-neutral surfaces. This is important because certain Ca^{2+} release is Ca^{2+}-diffusion dependent. Accurate Ca^{2+} diffusion characterization in a realistic cell-environment help to understand the dyssynchrony of Ca^{2+} waves in excitable cell (such as cardiac myocytes) that are commonly present in heart failure.

We studied the molecular bases of Ca^{2+}-mediated calmodulin (CaM)/calcineurin (CaN) interaction. This pair of PPI is important as they are key nodes of the Ca^{2+} signaling network. The intrinsically disordered nature of CaN's regulatory domain, at which CaM binds, also makes this PPI pair an appropriate model to study intrinsically disordered peptide (IDP)-involved PPI. IDP-involved PPI has attracted great attention due to the capability of IDP in recognizing wide range of protein targets. In this subproject, we studied the association process between CaN's intrinsically disordered regulatory domain and CaM under varying ionic strengths, with the aim to understand how electrostatic interactions affect the

association rate constants. The motivation comes from the fact that electrostatic interactions play a dual role in determining IDP's ensemble properties and driving protein-protein association. Extensive all-atomic molecular dynamics (MD) simulations were performed on CaN' regulatory domain constructs bearing different charge densities. For each representative structure of CaN constructs, the Brownian dynamic (BD) simulations were performed to estimate its diffusional encounter rates with CaM. During MD and BD simulations, by changing the ionic strengths, we were able to tune the intramolecular electrostatic in CaN's regulatory domain and intermolecular electrostatic interactions that drive CaN/CaM association. We found that CaN construct's ensemble compactness were not significantly affected by intramolecular electrostatic interactions. However, the conversion rates between CaM-capable and incapable conformations are affected by electrostatic interactions, with slower conversion rates for the CaN construct having lower net charge per residue (NCPR). The diffusion encounter rate between CaN's regulatory domain and CaM were significantly enhanced by the intermolecular electrostatic interactions, as evidenced by at least $\sim$30% faster rate constants after intermolecular electrostatic interactions are considered in BD simulations. In addition, we obtained the effective rate constants after taking CaN regulatory domain's conformation dynamics into consideration. This effective rate constant reflects the reality that IDPs assume continuously changing conformations when bind to protein partners. The extent to which conformational dynamics affect association rate constants depends on the relative value between conversion rate and diffusional encounter rate. For CaN constructs with a much slower conversion rate than diffusional encounter rate, their effective rate constants were significantly reduced as CaM-incapable structures occupied considerable percentage of the time course. These studies revealed the dual role of electrostatic interactions that resulted in the diffusion-limited association rate of CaN's intrinsically disordered regulatory to CaM. More generally, protein-protein associations are fundamental events occurring in various biological processes such as enzyme activation/inhibition and immune system response. Understanding the mechanism of disease-related protein-protein association is of great therapeutic importance as binding rates characterization can be used to develop new drug candidates.

We additionally investigated the secondary interaction between a motif in CaN's regulotary domain and CaM, as experimental evidence indicated that this secondary interaction is indispensable for CaN's full activation by CaM. By using protein-protein docking and extensive MD simulations, we were able to identify on CaM's surface a potential site for this secondary interaction. This prediction was affirmed by phosphatase assays as CaM variants with mutation at the site leaded reduced CaN phosphoratase activity. The identification of site D provides a hot spot area on CaM that could be used to design CaM variants with controllable CaN activation abilities. Given the numerous biological processes regulated by CaN, in which some are disease-related, such engineered CaM variants are of great therapeutic importance. Moreover, besides CaN, there are other CaM-regulated enzymes with autoinhibition domains such as

CaM dependent kinases (CaMKI and CaMKII) and myosin light-chain kinase (MLCK). These enzymes may share a similar activation mechanism as CaN in which a secondary interaction with CaM is needed for full activation.

Sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) is a calcium pump that actively transports Ca^{2+} from cytosol into sarcoplasmic/endoplasmic reticulum (SR/ER)) which prepares the excitable cells for the next cycle of Ca^{2+} oscillating. We studied the active transport of Ca^{2+} by SERCA, with the aim to understand SERCA's selectivity of Ca^{2+} over Mg^{2+} and K^+ given the latter two are in much larger excess in cytoplasm. We performed all atomic MD simulations on SERCA with Ca^{2+}, Mg^{2+} and K^+ bound, respectively. It was found that the binding site was optimized for Ca^{2+} binding, as evidenced by the smallest root mean squared fluctuations (RMSF) values of coordinating residues and the largest number of coordinating oxygens. The coordination pattern was less optimal for Mg^{2+} and not favorable for K^+, as the events of K^+ leaving were observed during MD simulation. Using MD-derived binding site information such as number of coordinating oxygen and volume of coordination spheres, we quantified the cation binding thermodynamics via the mean sphere approximation (MSA) method. Our MSA-generated thermodynamics correctly ranked cation binding affinities as validated by experimental values. We further evaluated the impacts of Mg^{2+} and K^+ binding on SERCA's function cycle via state models. We found that Mg^{2+} most likely acts as a inhibitor while K^+ as a intermediate during SERCA's transport of Ca^{2+}. This study deciphered the molecular determinants of SERCA's selectivity of Ca^{2+} and evaluated the impacts of Mg^{2+} and K^+ on SERCA's enzymatic activities. Our results deepen the understanding of SERCA functions in a real cell-like environment in which Mg^{2+}/K^+ ions are inevitably present. In addition, the MD/MSA combination strategy can be applied to other Ca^{2+}-permeable channels such as $Na+/Ca^{2+}$ exchanger (NCX) and polycystin-2 (PC2) to gain molecular bases of their functions.

The future work will focus on refining the simulation models with more realistic conditions taken into consideration. For example, considering the steric effect exerted by protein crowders on cellular Ca^{2+} diffusion is needed to obtain Ca^{2+} flux profiles that can be directly compared with experimental measurements. Our current studies regarding the CaM/CaN PPI were simulated with only the regulatory domain of CaN presented. Building complex structure of CaM with the intact CaN is the prerequisite to simulate the complete activation process, which allows the identification of critical residue-residue interactions. This helps to design CaM variants with controllable abilities of activating CaN. In addition, simulations of SERCA with membrane having different lipid compositions are important because membrane lipids play regulatory role in SERCA function. The goal of these refinements is to push the limit of our understanding of Ca^{2+}-signaling mechanism with quantitative descriptions that can be experimentally validated.

Appendix: Supplementary Material

Terms Γ, η and Δ refers to the contribution from electrostatic shielding, unequal sizes of ions and free volume not occupied by hard spheres in mean sphere approximation (MSA), respectively. Their expressions are as:

$$\begin{aligned}
\Gamma^2 &= \pi\lambda_b \sum_i \rho_i \left[\frac{z_i - \eta\sigma_i^2}{1+\Gamma\sigma_i}\right]^2 \\
\eta &= \frac{\pi}{2\Omega\Delta}\sum_k \frac{\rho_k\sigma_k z_k}{1+\Gamma\sigma_k} \\
\Omega &= 1 + \frac{\pi}{2\Delta}\sum_k \frac{\rho_k\sigma_k^3}{1+\Gamma\sigma_k} \\
\Delta &= 1 - \frac{\pi}{6}\sum_k \rho_k\sigma_k^3
\end{aligned} \tag{S1}$$

where z_i, ρ_i and σ_i are charge, concentration and radius of ion i, respectively. $\lambda_b = \frac{e^2}{4\pi\epsilon_0\epsilon_r k_b T}$ is Bejerrum length which refers to the separation of two elementary charges at which the electrostatic interaction potential is comparable to thermal energy. ϵ_0, ϵ_r are vacuum and relative permittivities, respectively.

Table S1: Average electric potentials calculated at the two ends of nanopore (see Eq. 2.21) in KCl conductance validation(2D-axisymmetric)

[KCl]	$V_a(V)$	$V_b(V)$	$V_a - V_b(V)$	$\frac{V_a - V_b}{\delta\phi}$
1×10^{-6} M	0.192929	-0.002049	0.194978	97.489%
1×10^{-5} M	0.194460	0.000472	0.193988	96.994%
1×10^{-4} M	0.196105	0.002011	0.194094	97.047%
1×10^{-3} M	0.197282	0.001943	0.195339	97.670%
1×10^{-2} M	0.198244	0.001437	0.196807	98.404%
1×10^{-1} M	0.198898	0.001008	0.19789	98.945%

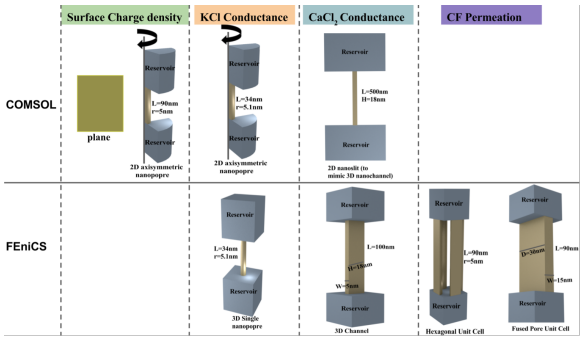


Figure S1: Illustration of all geometries used in simulations.

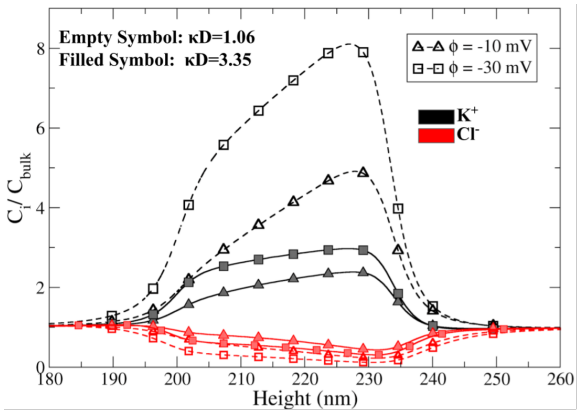


Figure S2: Concentration profile of K^+ and Cl^- along the center line of nanopore under different wall electric potentials and κD (D=10.2 nm is the diameter of nanopore). The height at x axis refers to the z direction: $0 < height < 200$nm is bottom reservoir, $200 < height < 234$ nm is nanopore and $234 < height < 434$nm is top reservoir.

Grahame equation for $CaCl_2$ salt solution

When the solution is composed of divalent cation (e.g., $CaCl_2$ solution), the Grahame equation used to relate wall electric potential (ϕ_0) and surface charge density (σ_s) is given as:

$$\sigma_s(\phi_0) = \sqrt{8\epsilon_o\epsilon_r k_B T}\sinh\left(\frac{e\phi_0}{2k_BT}\right)\left([CaCl_2](2+exp(\frac{-e\phi_0}{k_BT}))\right)^{0.5} \quad \text{(S2)}$$

Effects of porosity on permeability

Besides the electrostatic interaction, we also explored the packing compactness of nanopores on CF permeation. For clarity, we defined porosity as shown in Fig. S4 to characterize the packing compactness of nanopores on the silica membrane.

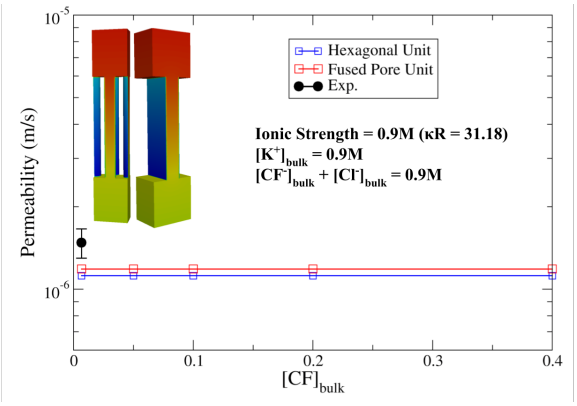


Figure S3: Numerically simulated permeability of CF passing through unit cell structures(dimensions are shown in Fig. 2.2). Total ionic strength is maintained as 0.9 M which results κD=31.18 (κ is inverse of Debye length and R=10 nm is diameter of nanopore). An fixed electric potential $\phi = -69.5mV$(value is from a PNP+pH regulated surface charge model at pH = 7.4) was applied on the nanopore wall

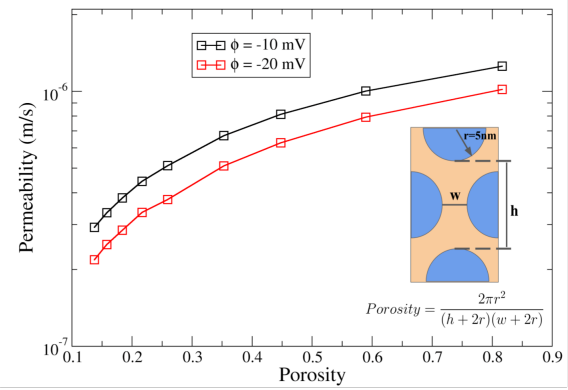


Figure S4: =

1 mM.] Permeability of CF vs. hexagonal unit cell porosity. Background ionic strength is set as 1 mM and [CF] = 1 mM.

While keeping nanopore radius $r = 5nm$ fixed, by changing the value of h and w, we can achieve hexagonal unit cells with different porosities. We then calculated CF permeability in these hexagonal cells at pH=7.4, bulk [CF]=1mM and background ionic strength = 1mM. It can be shown clearly in Fig. S4 that, CF permeability linearly increases with porosity, indicating that a more compacted packing of nanopore will has higher CF permeability. Also at fixed porosity, a less negative electric potential permits a larger permeability, which is consistent with trends observed in Fig. 2.6.

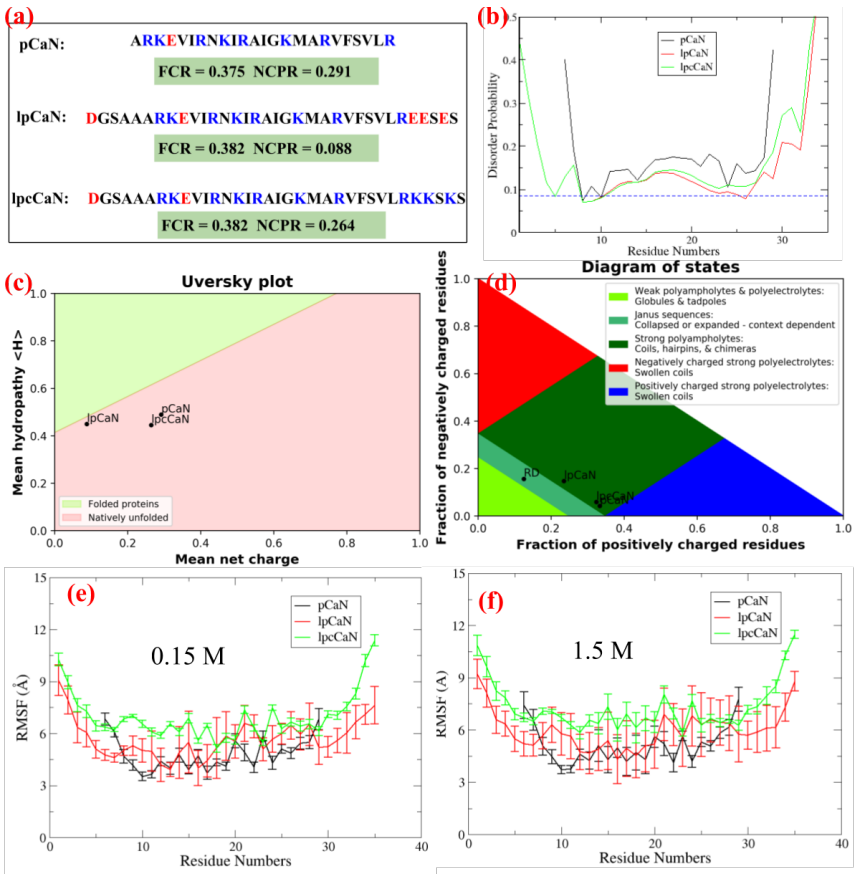


Figure S5: (a) Sequences of the three CaN peptides studied in the present work. The positively charged residues and negatively charged residues are colored in blue and red, respectively, along with FCR and NCPR scores. (b) Disorder probabilities predicted by DisEMBL.[373] The shown curves are scores based on "hot-loops" which is reported to be a good criterion to define disorder.[373,374] The blue dash line denotes random expectation values. (c) Mean hydropathy score and mean net charges of the three peptides and their locations in the Uversky diagram[375] (d) Locations of the three CaN peptides and CaN RD in the Das-Pappu diagram.[297] Figures (c) and (d) are generated by localCIDER.[376] (e-f) RMSF of each residue during the 15 μs MD at 0.15M and 1.5M ionic strength, respectively

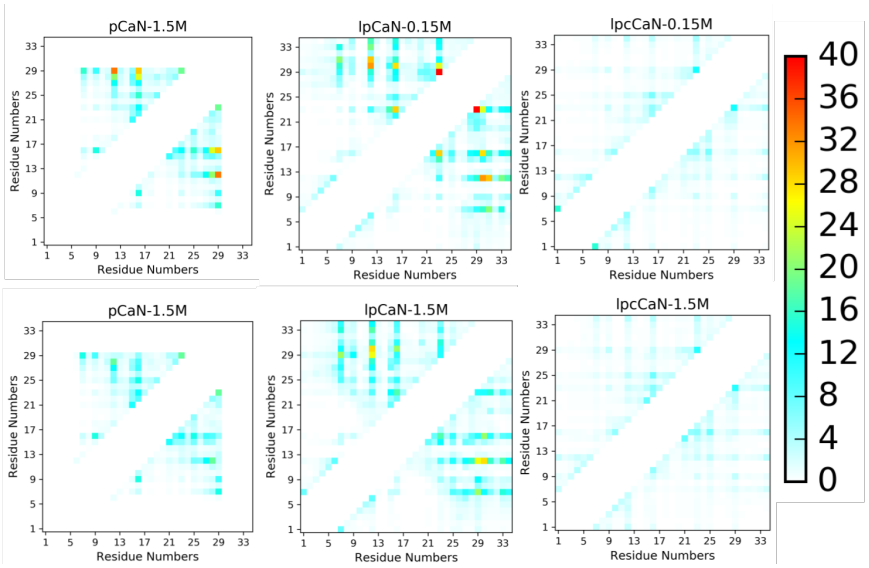


Figure S6: Contact map analysis of 15 μs MD trajectory of three CaN peptide under 0.15 M and 1.5 M ionic strength, respectively. Contact data was collected via CPPTRAJ in Amber with distance cutoff as 7 Å and only residue pairs which are at least 5 residues apart (i and $i+5$) in sequence are considered. The unit of numbers on color bar is number of average contacts for each pair over the simulation time

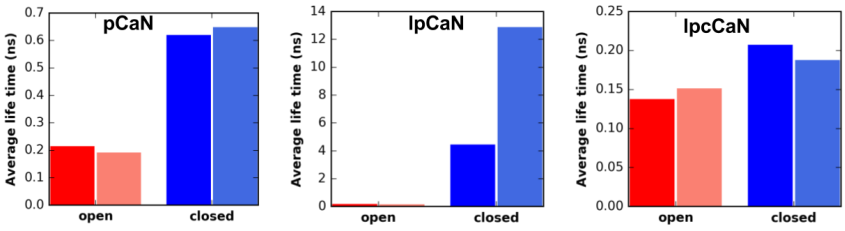


Figure S7: Average life time of open and closed states of CaN peptides determined by MSM at 0.15 M (faded color) and 1.5 M (dark color), respectively. The specific life time and corresponding gating rates are listed in Table S2

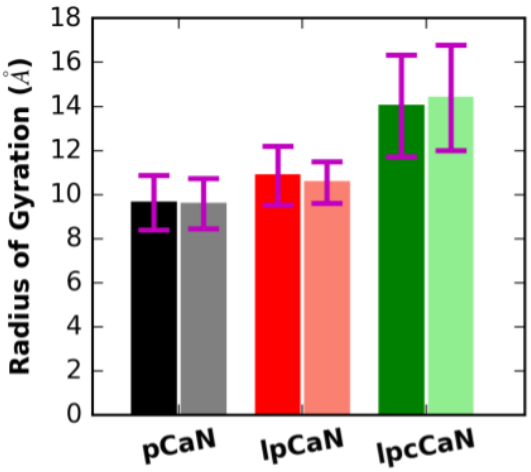


Figure S8: Average RG of CaN peptides at 0.15 M (faded color) and 1.5 M (dark color), respectively

Flags set of Rosetta modeling

The following flags set was used for pCaN structural prediction via Rosetta:

"-in:file:fasta pCaN_fasta
-in:file:frag3 aapCaN03_05.200_v1_3
-in:file:frag9 aapCaN09_05.200_v1_3
-abinitio:relax
-nstruct 10
-out:pdb
-use_filters true
-abinitio::increase_cycles 10
-abinitio::rg_reweight 0.5
-abinitio::rsd_wt_helix 0.5
-abinitio::rsd_wt_loop 0.5
-relax::fast"

Table S2: Average life times and gating rates between CaN peptides' open and closed conformations sampled at 0.15 M and 1.5 M ionic strength.

	Ionic Strength(M)	Average Lifetime (ns)		Gating rate (s^{-1})	
		Γ_{open}	Γ_{closed}	k_b	k_f
pCaN	0.15	0.19	0.65	5.27×10^9	1.55×10^9
	1.5	0.21	0.62	4.70×10^9	1.62×10^9
lpCaN	0.15	0.11	12.83	8.97×10^9	7.79×10^7
	1.5	0.13	4.42	7.48×10^9	2.26×10^8
lpcCaN	0.15	0.15	0.19	6.64×10^9	5.35×10^9
	1.5	0.14	0.21	7.31×10^9	4.85×10^9

Table S3: BD-simulated encounter rates (k_{on}) under 0.15 M and 1.5 M ionic strength for the open state CaN peptides sampled the same two ionic strengths. The effective association rates (k_{eff}) calculated via Eq. 3.3 are also shown.

		k_{on}/k_{eff} ($1 \times 10^7\,M^{-1}\,s^{-1}$)		Gating rate ($1 \times 10^7\,s^{-1}$)	
		0.15_BD	1.5_BD	k_b	k_f
pCaN	0.15_confs	2.00 /1.61	0.11/0.102	527	155
	1.5_confs	1.56/1.31	0.07/0.065	470	162
lpCaN	0.15_confs	0.64/0.14	0.06/0.03	897	7.79
	1.5_confs	0.53/0.26	0.08/0.06	748	22.6
lpcCaN	0.15_confs	0.97/0.93	0.04/0.04	664	535
	1.5_confs	0.34/0.33	0.005/0.005	731	485

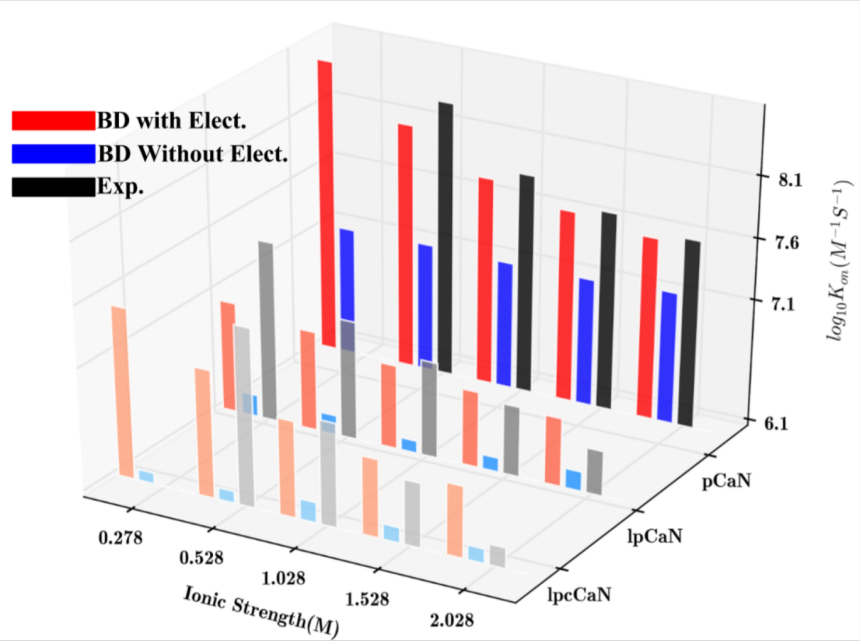


Figure S9: BD calculated association kinetics between rigid CaN peptides and CaM. The electrostatic interaction was turned off by setting CaN peptides charges to be zero. Specifically, after turning of electrostatic interaction, lpCaN retains 50% above association rates while lpCaN and lpcCaN reduce to much smaller k_{on}s, implying that electrostatic interaction has larger impacts on lpCaN and lpcCaN than pCaN

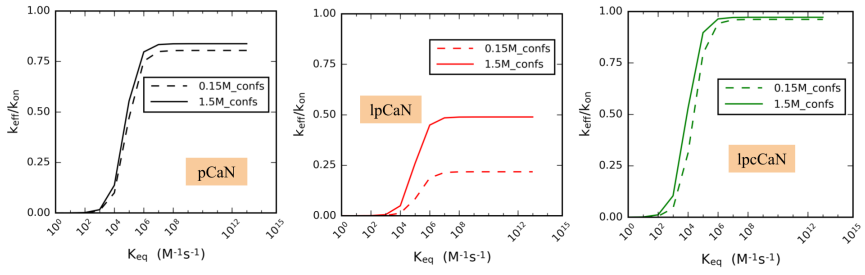


Figure S10: Sensitivity of k_{eff} to K_{eq} for three CaN peptides sampled at 0.15 M and 1.5 M ionic strength, respectively.

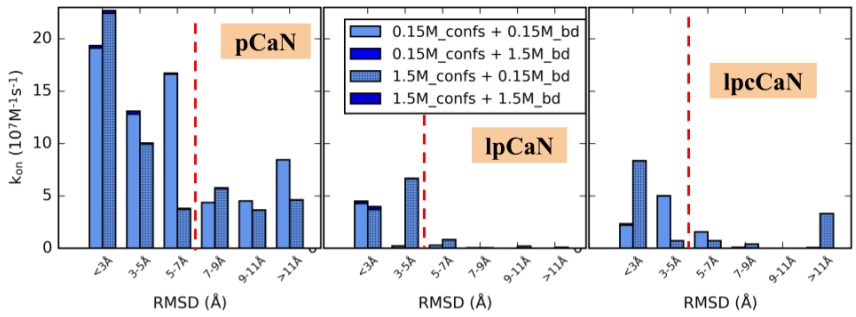


Figure S11: BD-simulated separate association rate constants with CaM C domain vs. RMSD under 0.15 M and 1.5 M ionic strength with CaN peptides conformations sampled at the same two ionic strengths, respectively. The red dash line designates the border of open and closed states based on RMSD

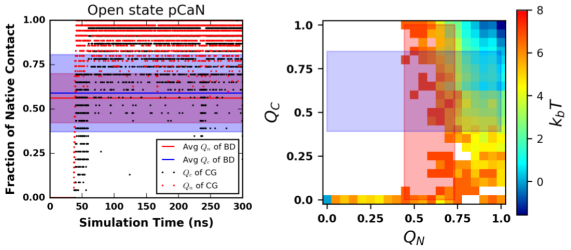


Figure S12: Fraction of native contact (Q) and free energy profile in coarse-grained (CG) simulation with Gō model. A: Fractions of native contact between N/C-domain of CaM and pCaN (denoted as Q_n and Q_c, respectively) along CG simulation time (only the first 300 ns of 1 μs is shown for clarity). B: 2D free energy profile projected along Q_n and Q_c in CG simulation. The shaded areas colored in light red and blue depict the ranges of Q_n and Q_c values in BD simulations, respectively. The last frame of each BD trajectory of the 10 pCaN open conformations (RMSD $<$ 3 Å) with CaM N/C domain were used to calculate average and standard deviation of Q_n and Q_c

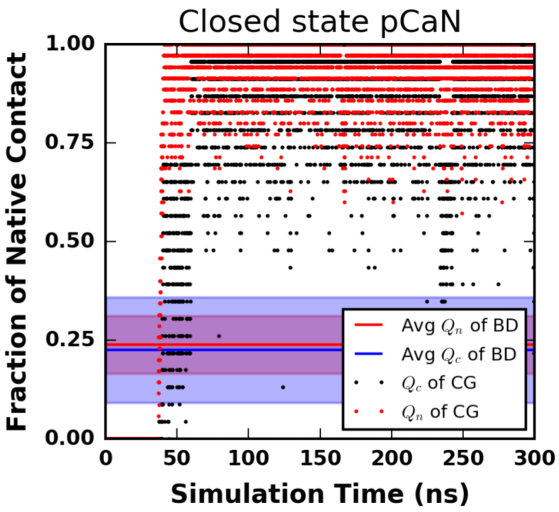


Figure S13: Fractions of native contact between N/C-domain of CaM and closed state pCaN in the BD-simulated pCaN-CaM complex

Methods

Simulation with AID added to site D predicted CaM/DH interaction configuration

The initial structure corresponding to CaN distal helix to AID region (residues 459 to 490) was built from sequence via tleap. The initial structure was subjected to minimization and MD simulation in vacuum according to the procedure described Sect. 4.2. The optimized structure was then appended to the C-terminus of distal helix region from the representative structure of site D simulations via tleap. The representative structures of the first two most populated clusters from site D simulations were selected, making the simulation duplicate. The sytem was then solvated in TIP3P waterbox with 0.15 M KCl ions added. The simulation details are same as previous section in which the tleap built structure was first relaxed while rest part being fixed during the heating and equilibrium stage. After reaching equilibrium, about 0.7 μs production simulation was performed from each replica of the duplicate. The simulations were repeated for the CaN A454E mutant.

Table S4: Residues at each tentative binding site on collapsed CaM used in ZDOCK to predict distal helix interaction at each site

Tentative Site	Residues
Site A	R86, F89, V142, Y138
Site B	R106, I125, D118, D122
Site C	I9, F12, L69, F65
Site D	K30 ,T34, R37, Q49

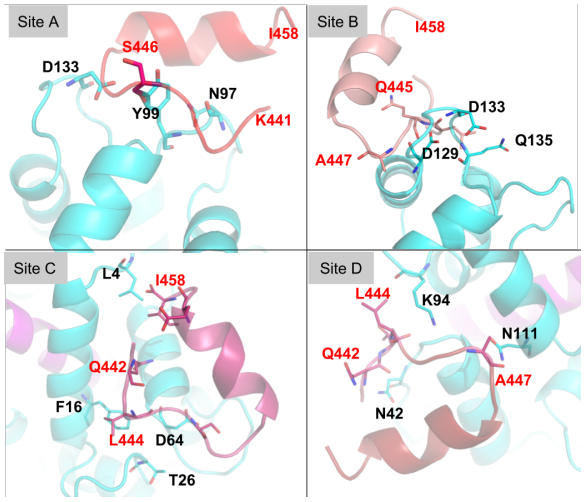


Figure S14: Highest-ranking CaM/distal helix interaction poses predicted by ZDOCK3.0.2[100] webserver at each site. The color scheme is same as Fig. 4.2. Key residues at the interaction surface are shown in sticks with black labels for CaM residues and red labels for distal helix residues. Comparisons of the WT distal helix poses versus predictions for DH_{A454E} are shown in Fig. S15.

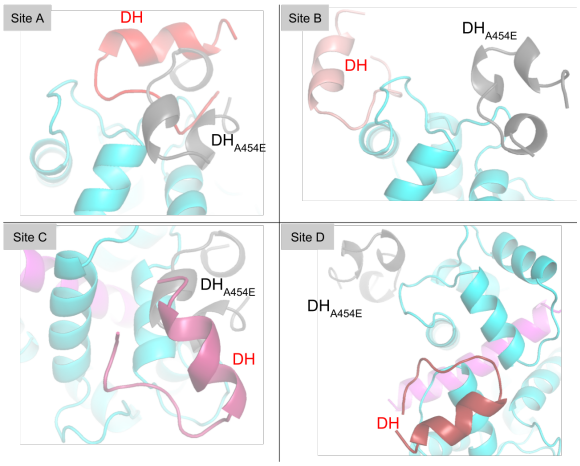


Figure S15: Comparison of Zdock predicted poses of DH and DH_{A454E} mutant at each site. The DH_{A454E} mutant is colored in gray. At site A and C, the poses of DH and mutant are close, while at site B and D, the DH_{A454E} mutant are predicted to be located near site A.

Figure S16: Overlap of MD simulated distal helix conformation starting at site B (colored in salmon) and site D (colored in red). During the simulations, distal helix starting at site B migrated to site D.

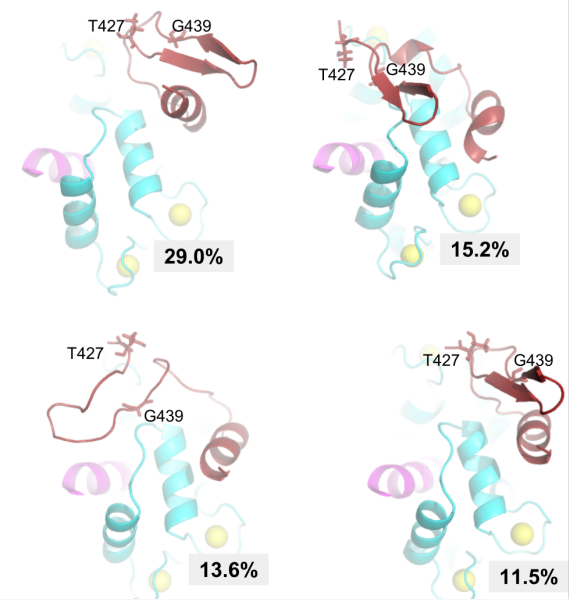


Figure S17: Illustration of β-sheet formed in T427-G439 region from site D simulations. The shown structures are representative structures of first four most populated clusters.

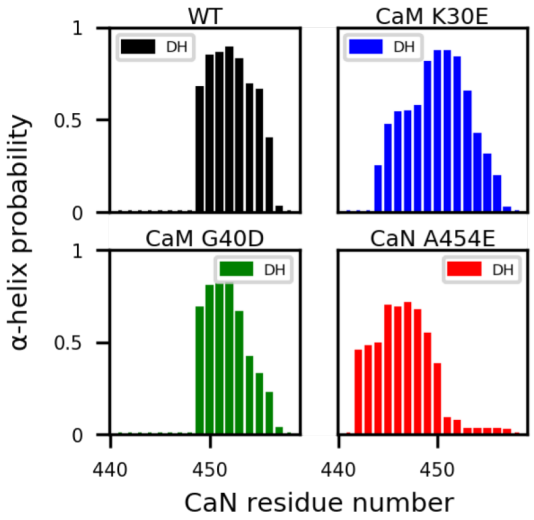


Figure S18: α-helix structural probability of each residue in distal helix region of WT, CaM K30E and G40D mutants and CaN A454E mutant calculated from MD simulations initiated at site D.

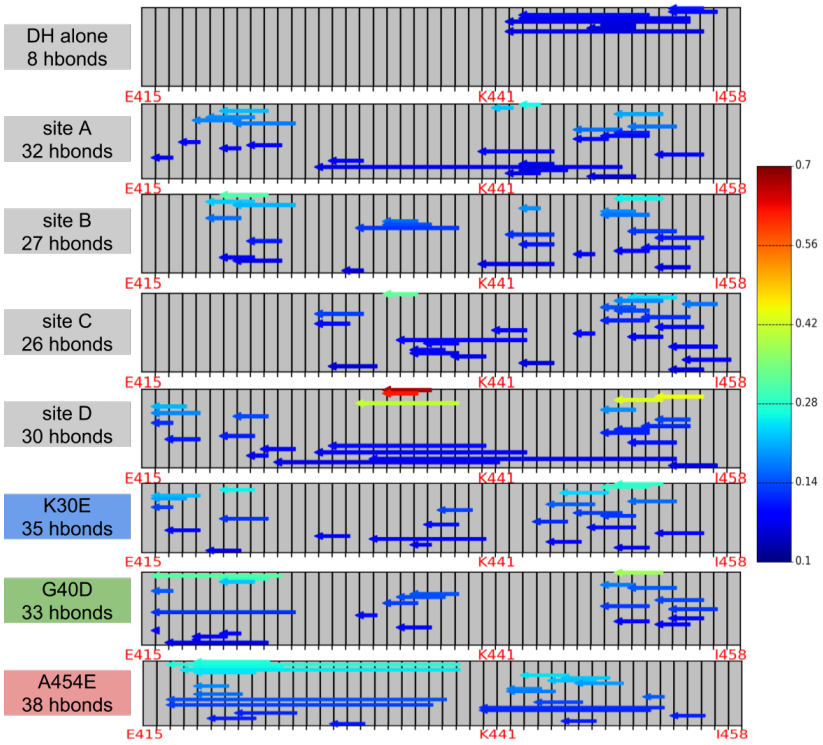


Figure S19: Backbone hydrogen bond analysis in the linker and distal helix region. Each arrow represents one Hbond with color indicating percentage of simulated frames with this hbond existed. hbonds exist >5% of simulation time are shown.

Table S5: MD simulation cases

Cases	Cation occupancy	PDBID of ini. struc	Simulation length (ns)
WT	2 Ca^{2+}s	1su4	300
WT	1 Mg^{2+}	3w5b	300
WT	2 K^{+}s	3w5b	300
N796A	2 Ca^{2+}s	1su4	400
N796A	1 Mg^{2+}s	3w5b	355
E309Q_r1[a]	2 Ca^{2+}s	1su4	240
E309Q_r2[b]	2 Ca^{2+}s	1su4	170

[a,b]Different rotamers of Q309 side chain

Table S6: Parameters used in SERCA MSA model. The diameter and charge of ions were adapted from Li-Merz[370] and the absolute solvation energy for ions were taken from table 1 in Nonner et al.[92] The Nonner et al results were adapted from values relative to solvated H^+ in.[377]

Parameter	Definition	Values
N_{oxy} at Ca^{2+} site I	no. of amino acid oxygen	5.5
N_{oxy} at Ca^{2+} site II	no. of amino acid oxygen	6.7
N_{oxy} at K^+ site I	no. of amino acid oxygen	3.9
N_{oxy} at Mg^{2+} site	no. of amino acid oxygen	4.2
r_{rdf} at Ca^{2+} site I, II	optimal coordination radii	2.5 Å
r_{rdf} at Mg^{2+} site	optimal coordination radii	2.0 Å
r_{rdf} at K^+ site I	optimal coordination radii	2.9 Å
ρ_i	bath [KCl]	150 mM
	bath [$MgCl_2$]	1 mM
	bath [$CaCl_2$]	0.1 nM to 0.1 M
ϵ_f	filter dielectric	25.0
ϵ_b	bath dielectric	78.4
σ_O	O diameter	0.354 nm
σ_{Ca}	Ca^{2+} diameter	0.272 nm
σ_{Mg}	Mg^{2+} diameter	0.236 nm
σ_K	K^+ diameter	0.352 nm
σ_{Cl}	Cl^- diameter	0.454 nm
z_O	O charge	-0.7469 e
z_{Ca}	Ca^{2+} charge	1.77 e
z_{Mg}	Mg^{2+} charge	1.69 e
z_K	K^+ charge	1 e
z_{Cl}	Cl^- charge	-1 e
w_{Ca}	solvation energy	-648.7 kT[92,377]
w_{Mg}	solvation energy	-779.9 kT[92,377]
w_K	solvation energy	-141.6 kT[92,377]
w_{Cl}	solvation energy	-122.6 kT[92,377]
kT/zF		25.6 mV
V_f of Ca^{2+} sites	filter volume	0.326 nm^3
V_f of Mg^{2+} site	filter volume	0.224 nm^3
V_f of K^+ site I	filter volume	0.426 nm^3

Table S7: Parameters used SERCA MSM models in present work.

Parameter	Value	reference
$K_{d,Mg}$	4.93×10^{-5} M/5×10^{-4} M	MSA/Expt.
$K_{d,K}$	3.3×10^{-3} M/9×10^{-3} M	MSA/Expt.
$K_{d,2}$	2.24×10^{-3} M	[334]
Ca(SR)	1×10^{-6} M	[9]
[MgATP]	3×10^{-4} M	[9]
[MgADP]	1×10^{-6} M	[9]
[Pi]	1×10^{-6} M	[9]
k_1^+	1×10^8 s^{-1}	[9]
k_1^-	4×10^2 s^{-1}	[9]
k_2^+	3.0×10^1 s^{-1}	[9]
k_2^-	4.0×10^1 s^{-1}	[9]
k_3^+	4×10^8 s^{-1}	[9]
k_3^-	1.6×10^1 s^{-1}	[9]
k_4^+	2×10^7 s^{-1}	[9]
k_4^-	4.0×10^1 s^{-1}	[9]
k_5^+	1.5×10^1 s^{-1}	fitted
k_5^-	7.5 s^{-1}	fitted
k_6^+	7.5 s^{-1}	fitted
k_6^-	7×10^9 s^{-1}	fitted

Table S8: Experimental dissociation constant of cations with SERCA

Cation	Dissociation constant (K_d)	reference
Ca^{2+} site I	0.018 - 0.023 μM	[317]
Ca^{2+} site II	6.67 - 11.1 μM	[317]
Mg^{2+}	5 mM / 0.5 mM[a]	[318] and [346]
K^+ site I	0.625 mM	[319]
K^+ site II	16.67 mM	[319]
K^+ binds to P-domain[b]	na	[306]
N796A and E309Q Ca^{2+} site I	0.5 μM	[337]
N796A and E309Q Ca^{2+} site II	no binding	[337]
E309Q with 2 Ca^{2+} binding	0.46 - 4.38 μM[c]	[316]

[a] in [346] it was postulated that Mg^{2+} binds to site II with $K_d = 0.5$ mM and this bound-Mg^{2+} would exchange quickly with Ca^{2+} and the consequence is Mg^{2+} accelerates Ca^{2+} bindings to SERCA; [b] a K^+ binding site located in P-domain was identified via X-ray crystallography; [c] This data is apparent affinity for both site I and II obtained by fitting to experimental data using Hill equation.

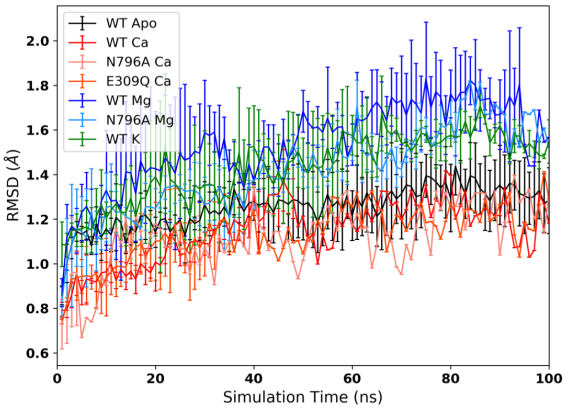


Figure S20: Back-bone atom RMSD of TM helices vs. simulations time. The error bars are calculated from the triplicate trajectories for each case.

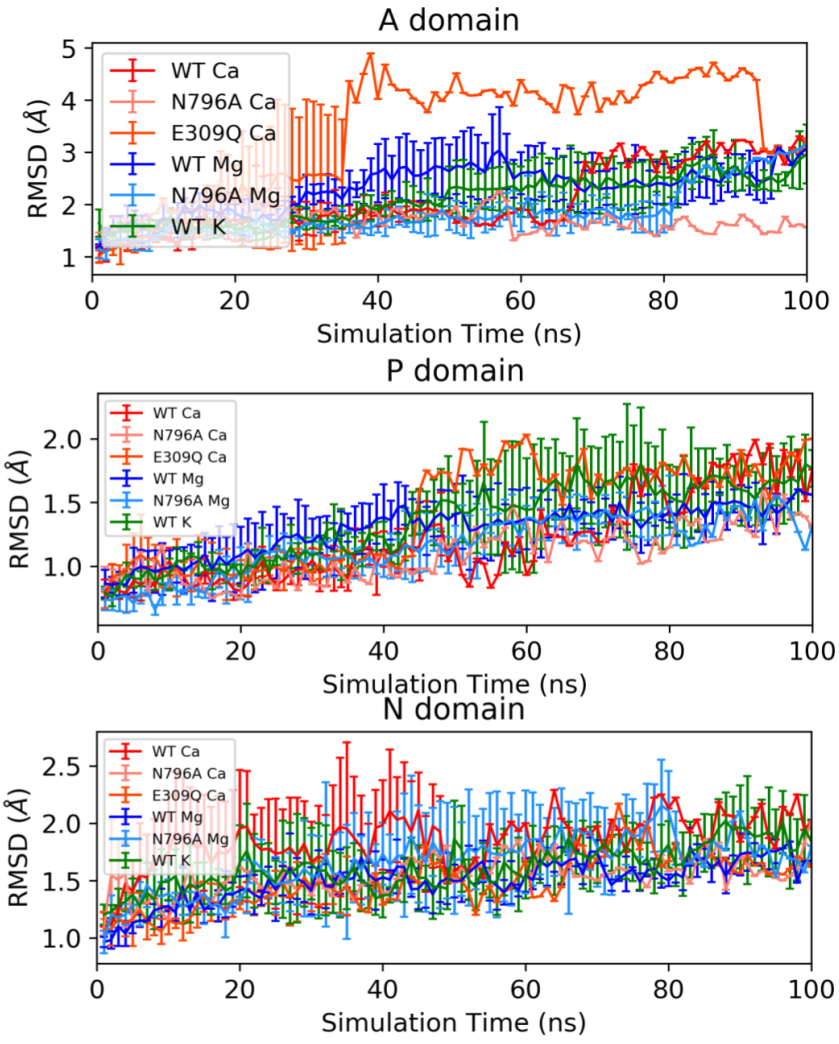


Figure S21: Back-bone atom RMSD of cytosolic domains vs. simulations time. Alignments were conducted using backbone atoms of cytosolic domains. The error bars are calculated from the triplicate trajectories for each case.

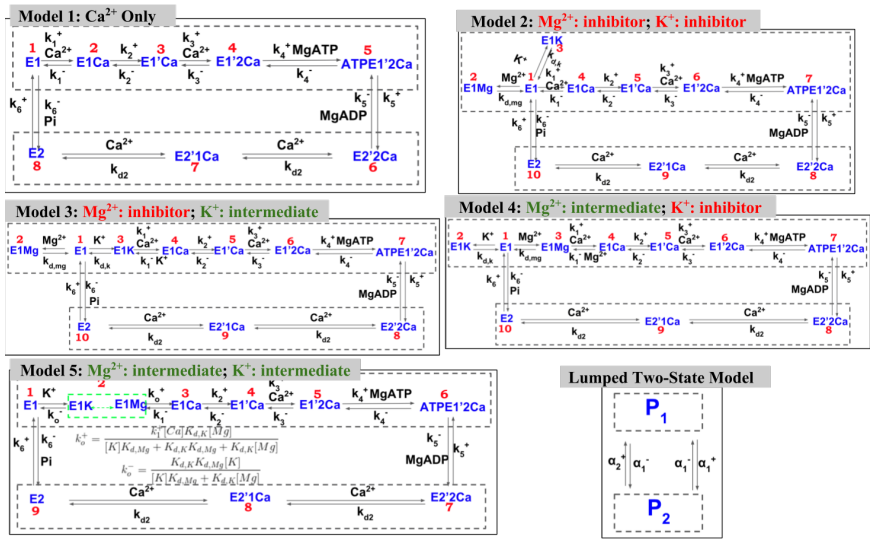


Figure S22: Schema of the five MSM models of SERCA pump considered in present study. In Model 1 (Ca_{only}), only the native Ca^{2+} cation is involved in the pump cycle. In Model 2 ($Mg\_K\_$), both Mg^{2+} and K^+ are assumed to act as inhibitors of SERCA. In Model 3 (Mg_K_+), Mg^{2+} is assumed to act as inhibitors of SERCA while K^+ acts as intermediate. In Model 4 ($Mg_+K\_$), the roles of Mg^{2+} and K^+ are opposite as Model 3, namely, Mg^{2+} is assumed to act as intermediate of SERCA while K^+ acts as inhibitor. In Model 5 (Mg_+K_+), both Mg^{2+} and K^+ are assumed to act as intermediates of SERCA. For all models, the states in dashed-box are lumped together based on the lumping strategy proposed by Smith et al[371] to form a two-state model. The apparent rates between the two-state model are given in supplementary material.

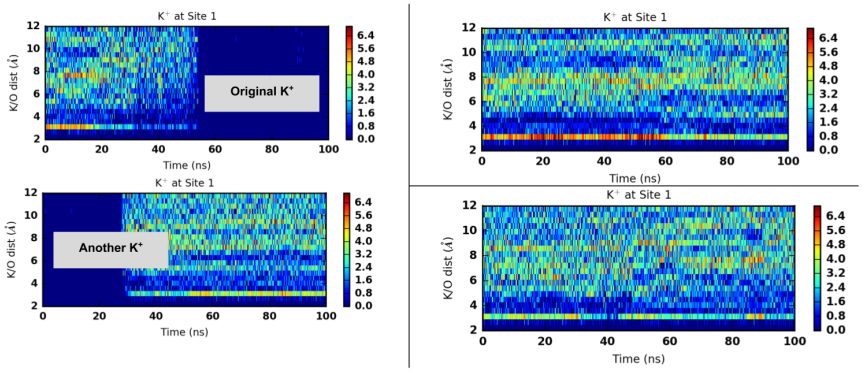


Figure S23: Coordination pattern of K^+ at SERCA site I with residue oxygen atoms in the 1st 100 ns WT simulations. In the 1st 100 ns simulation, the original K^+ at site I flee away after about 40 ns, however, another K^+ will bind at site I afterwards. The radial distribution is calculated based on the second K^+ after 40 ns.

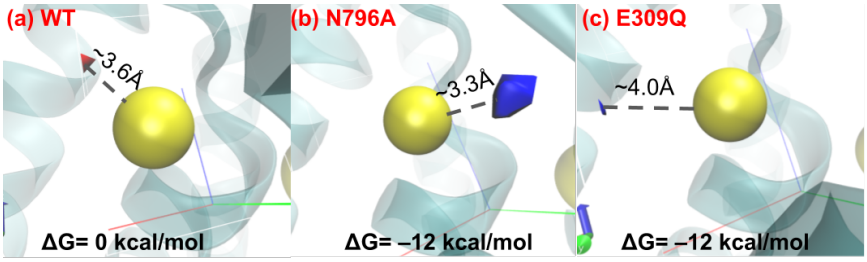


Figure S24: Free energy of water sites around site II Ca^{2+} calculated by GIST. For N796A and E309Q, the blue isosurface corresponds to water sites with $\Delta G = -12$ kcal/mol. For WT the shown red isosurface depicts a water oxygen density as 0.075 while the ΔG is 0 everywhere over the grids. The distance between Ca and isosurface is also shown.

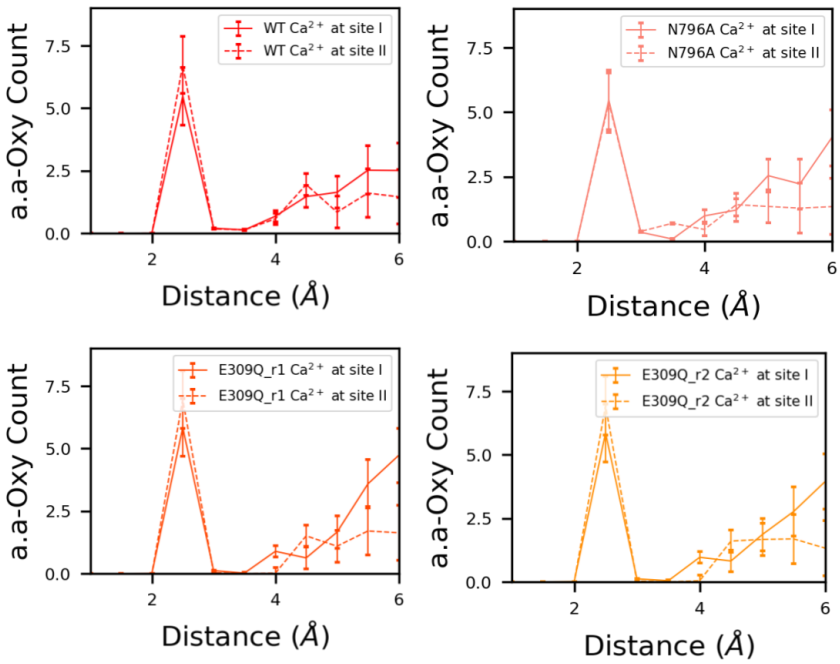


Figure S25: Radial distribution of oxygen around Ca^{2+} for each case.

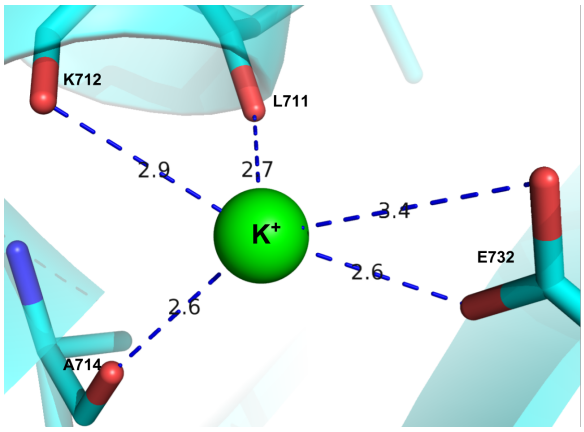


Figure S26: K^+ binding site at P domain from crystal structure PDB 1t5s. Number of coordinating oxygens is five and filter volume is assumed to be 0.43 nm^3.

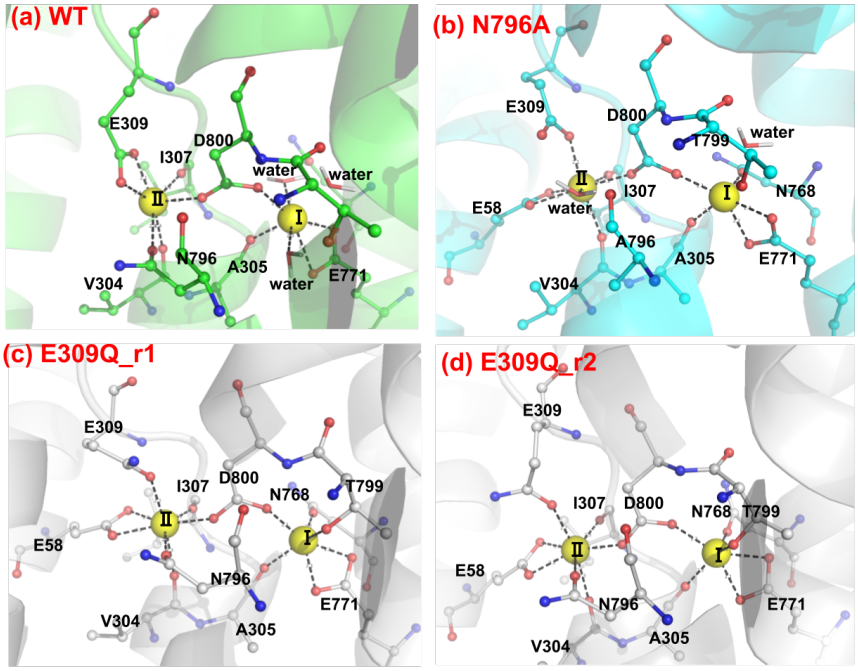


Figure S27: Representative coordination patterns of Ca^{2+} (yellow spheres) in the two binding sites from MD simulation in WT SERCA and mutants. Key residues are shown in sticks and balls. In WT site I, there are 3 waters coordinating with Ca^{2+}, while in N796A/E309Q mutants, site I Ca^{2+} has one/zero coordinating water molecules.

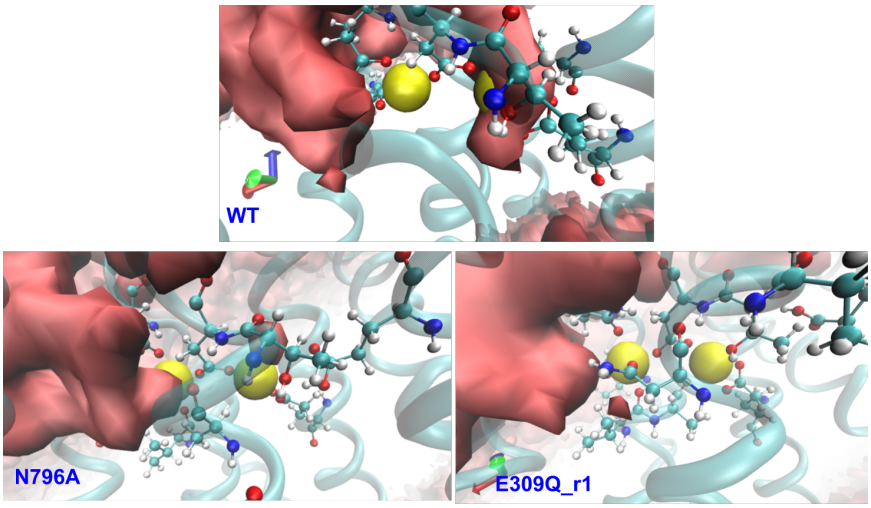


Figure S28: Water density around Ca^{2+} (yellow spheres) calculated from MD simulations via the VolMap module of VMD. The shown red isosurface depicts water density as 0.2.

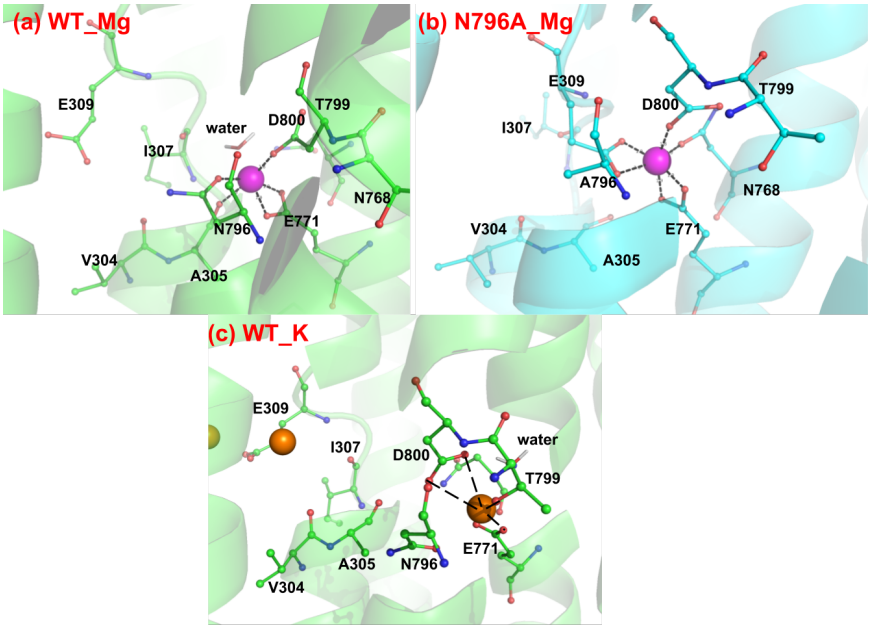


Figure S29: Representative coordination patterns of Mg^{2+} (magenta spheres) and K^+ (orange spheres) in the binding sites from MD simulation in WT SERCA and mutants. In both WT and N796A, Mg^{2+} resides in a 'hybrid' binding site rather than the conventional two Ca^{2+} binding site. K^+ at site I is stable and not stable K^+ binds at site II.

Supplementary Methods

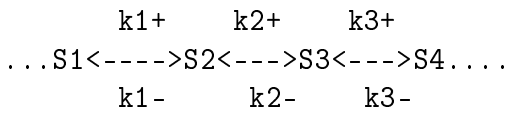
Grid Inhomogeneous Solvation Method (GIST)

The Grid Inhomogeneous Solvation Method (GIST) model[378,379] was used to characterize the theromodynamic properties of coordinating water. GIST calculations were performed via the SSTmap program.[380] Before GIST calculations, the MD trajectory was first subjected to rms fitting on key cation coordinating residues to align cation binding sites. The Cα atoms of residues V304, I307, N768, E771, N796 and D800 were used for this rms fitting. The grid center required in GIST was defined by the cation coordinates obtained from the last frame in the rms-fitted trajectory. Grid dimensions were set to 16, 20 and 20 with grid spacing being 0.5 Å. The free energy of water at each grid point was calculated as $\Delta G = E_{sw} + E_{ww} - dTS_{orient} - dTS_{trans}$ where E_{sw} and E_{ww} are solute-water and water-water interaction energies while dTS_{orient} and dTS_{trans} are orientational and translational entropic energies.

Model reduction strategy

Here we summarize our lumping strategy used to consolidate the SERCA model into fewer (macro) states, by assuming rapidly-exchanging 'micro' states are

in pseudo-equilibrium. The basic approach follows that of Smith et al[371] as illustrated by the generalized example:

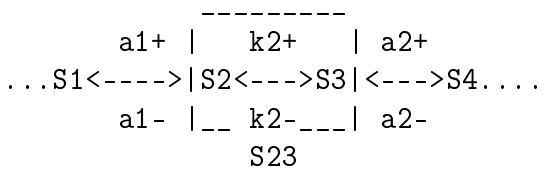


S1 to S4 are four states with forward and reverse rate constants connecting neighboring states. Suppose S2 and S3 are in rapid equilibrium with each other, namely k_2^+,k_2^- $\geq$ k_1^+,k_1^-,k_3^+,k_2^-. At any time:

$$S3(t) = K_2 * S2(t) \tag{S3}$$

$$K_2 = \frac{k_2^+}{k_2^-} \tag{S4}$$

Now we lump S2 and S3 together to form a new pseudo state S23. For the lumped model, we will use new apparent rate constants α_1^+, α_1^-, α_2^+, α_2^- to represent the transition between states.



For the lumped model, we have:

$$S23(t) = S2(t) + S3(t) \tag{S5}$$

$$= (1 + K_2)S2(t) \tag{S6}$$

$$= (1 + \frac{1}{K_2})S3(t) \tag{S7}$$

$$S2(t) = \frac{1}{1 + K_2}S23(t) \tag{S8}$$

$$S3(t) = (1 + \frac{1}{K_2})S23(t) \tag{S9}$$

Now we consider relationship between original rate constants and the new apparent rate constants. In both models, for S4, we have:

$$S4(t) = S3(t)k_3^+ = S23(t)\alpha_2^+ \tag{S10}$$

Now we put Eq(7) into Eq(8), we see clearly that:

$$\frac{k_3^+}{1 + \frac{1}{K_2}}S23(t) = S23(t)\alpha_2^+ \tag{S11}$$

so we have the expression of α_2^+ as:

$$\alpha_2^+ = \frac{k_3^+}{1 + \frac{1}{K_2}} \tag{S12}$$

similarly, we can derive the expression of α_1^+, α_1^-, α_2^-:

$$\alpha_1^+ = k_1^+ \tag{S13}$$

$$\alpha_1^- = \frac{k_1^-}{1 + K_2} \tag{S14}$$

$$\alpha_2^- = k_3^- \tag{S15}$$

$$\frac{dP_{1-5}}{dt} = -(\alpha_2^- + \alpha_1^+)P_{1-5} + (\alpha_2^+ + \alpha_1^-)P_{6-8} \tag{S16}$$

$$V_{max} = \frac{\alpha_1^+\alpha_2^+ - \alpha_1^-\alpha_2^-}{\alpha_1^+ + \alpha_1^- + \alpha_2^+ + \alpha_2^-} \tag{S17}$$

$$I_{Ca2+} \propto N_{SERCA} V_{max} \tag{S18}$$

Bibliography

(1) Clapham, D. E. *Cell* **2007**, *131*, 1047–1058.

(2) Cheng, H.; Lederer, W. J.; Cannell, M. B. *Science* **1993**, *262*, 740–744.

(3) Bading, H.; Ginty, D. D.; Greenberg, M. E. *Science* **1993**, *260*, 181–186.

(4) Vig, M.; Kinet, J.-P. *Nature Immunology* **2009**, *10*, 21–27.

(5) Mattson, M. P.; Chan, S. L. *Cell Calcium* **2003**, *34*, 385–397.

(6) Wehrens, X. H.; Lehnart, S. E.; Marks, A. R. *Annual Review of Physiology* **2005**, *67*, 69–98.

(7) Cannell, M. B.; Berlin; Lederer, W. J. *Science* **1987**, *238*, 1419–1423.

(8) Chin, D.; Means, A. R. *Trends in Cell Biology* **2000**, *10*, 322–328.

(9) Cantilina, T.; Sagara, Y.; Inesi, G.; Jones, L. R. *Journal of Biological Chemistry* **1993**.

(10) Altimimi, H. F.; Schnetkamp, P. P. *Channels* **2007**, *1*, 62–69.

(11) Endo, M. *Physiological Reviews* **1977**, *57*, 71–108.

(12) Dupont, G.; Combettes, L.; Bird, G. S.; Putney, J. W. *Cold Spring Harbor Perspectives in Biology* **2011**, *3*, DOI: 10.1101/cshperspect.a004226.

(13) Wacquier, B.; Combettes, L.; Van Nhieu, G. T.; Dupont, G. *Scientific Reports* **2016**, *6*, 19316.

(14) Domínguez-Rodríguez, A.; Ruiz-Hurtado, G.; Benitah, J. P.; Gómez, A. M. *Frontiers in Physiology* **2012**, *3 NOV*, 452.

(15) Lewit-Bentley, A.; Réty, S. *Current Opinion in Structural Biology* **2000**, *10*, 637–643.

(16) Schwaller, B. *Cellular and Molecular Life Sciences* **2009**, *66*, 275–300.

(17) Donato, R. *The International Journal of Biochemistry & Cell Biology* **2001**, *33*, 637–668.

(18) Yap, K. L.; Kim, J.; Truong, K.; Sherman, M.; Yuan, T.; Ikura, M. *Journal of Structural and Functional Genomics* **2000**, *1*, 8–14.

(19) Molkentin, J. D. *Cardiovascular Research* **2004**, *63*, 467–475.

(20) Brzozowski, J. S.; Skelding, K. A. *Pharmaceuticals* **2019**, *12*, DOI: 10.3390/ph12010008.

(21) Zhong, P.; Quan, D.; Peng, J.; Xiong, X.; Liu, Y.; Kong, B.; Huang, H.; Zhong, P.; Quan, D.; Peng, J.; Xiong, X.; Liu, Y.; Kong, B.; Huang, H. *Journal of Molecular and Cellular Cardiology* **2017**, *109*, 1–16.

(22) Higley, M. J.; Sabatini, B. L. *Neuron* **2008**, *59*, 902–913.

(23) Leung, C.; Kinns, H.; Hoogenboom, B. W.; Howorka, S.; Mesquida, P. *Nano Letters* **2009**, *9*, 2769–2773.

(24) Worley, J. F.; McIntyre, M. S.; Spencer, B.; Mertz, R. J.; Roe, M. W.; Dukes, I. D. *Journal of Biological Chemistry* **1994**, *269*, 14359–14362.

(25) Rühle, B.; Davies, M.; Lebold, T.; Bräuchle, C.; Bein, T. *ACS Nano* **2012**, *6*, 1948–1960.

(26) Andersen, M. B.; Frey, J.; Pennathur, S.; Bruus, H. *Journal of Colloid and Interface Science* **2011**, *353*, 301–310.

(27) Persello, J.; Papirer, E. *Adsorption on Silica Surfaces* **2000**, *10*, 298–339.

(28) Chin, D.; Means, A. R. *Trends in Cell Biology* **2000**, *10*, 322–328.

(29) Jain, J.; McCafffrey, P. G.; Miner, Z.; Kerppola, T. K.; Lambert, J. N.; Verdine, G. L.; Curran, T.; Rao, A. *Nature* **1993**, *365*, 353–355.

(30) Rusnak, F.; Mertz, P. *Physiological Reviews* **2000**, *80*, 1483–1521.

(31) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions, 2005.

(32) Kissinger, C. R. et al. *Nature* **1995**, *378*, 641–644.

(33) Manalan, A. S.; Klee, C. B. *Proceedings of the National Academy of Sciences* **1983**, *80*, 4291–4295.

(34) Shen, X.; Li, H.; Ou, Y.; Tao, W.; Dong, A.; Kong, J.; Ji, C.; Yu, S. *Journal of Biological Chemistry* **2008**, *283*, 11407–11413.

(35) Rumi-Masante, J.; Rusinga, F. I.; Lester, T. E.; Dunlap, T. B.; Williams, T. D.; Dunker, A. K.; Weis, D. D.; Creamer, T. P. *Journal of Molecular Biology* **2012**, *415*, 307–317.

(36) Wright, P. E.; Dyson, H. *Journal of Molecular Biology* **1999**, *293*, 321–331.

(37) Dyson, H.; Wright, P. E. *Current Opinion in Structural Biology* **2002**, *12*, 54–60.

(38) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021–1025.

(39) Schreiber, G.; Haran, G.; Zhou, H.-X. *Chemical Reviews* **2009**, *109*, 839–860.

(40) Qin, S.; Pang, X.; Zhou, H.-X. *Structure* **2011**, *19*, 1744–1751.

(41) Zhou, H.-X.; Pang, X.; Lu, C. *Phys. Chem. Chem. Phys.* **2012**, *14*, 10466–10476.

(42) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. *Proceedings of the National Academy of Sciences* **2010**, *107*, 8183–8188.

(43) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. *Proceedings of the National Academy of Sciences* **2010**, *107*, 14609–14614.

(44) Zhou, H.-X.; Pang, X. *Chemical Reviews* **2018**, *0*, null.

(45) Szabo, A.; Shoup, D.; Northrup, S. H.; McCammon, J. A. *The Journal of Chemical Physics* **1982**, *77*, 4484–4493.

(46) Dunlap, T. B.; Cook, E. C.; Rumi-Masante, J.; Arvin, H. G.; Lester, T. E.; Creamer, T. P. *Biochemistry* **2013**, *52*, 8643–8651.

(47) Ma, H.; Inesi, G.; Toyoshima, C. *Journal of Biological Chemistry* **2003**, *278*, 28938–28943.

(48) Inesi, G.; Lewis, D.; Toyoshima, C.; Hirata, A.; de Meis, L. *Journal of Biological Chemistry* **2008**, *283*, 1189.

(49) Kristián, T.; Siesjö, B. K. *Life Sciences* **1996**, *59*, 357–367.

(50) Pugashetti, R.; Shinkai, K.; Ruben, B. S.; Grossman, M. E.; Maldonado, J.; Fox, L. P. *Journal of the American Academy of Dermatology* **2011**, *64*, 296–301.

(51) Gwathmey, J. K.; Copelas, L.; MacKinnon, R.; Schoen, F. J.; Feldman, M. D.; Grossman, W.; Morgan, J. P. *Circulation Research* **1987**, *61*, 70–76.

(52) Head, J. F.; Perry, S. V. *Biochemical Journal* **1974**, *137*, 145–154.

(53) Awata, H.; Huang, C.; Handlogten, M. E.; Miller, R. T. *Journal of Biological Chemistry* **2001**, *276*, 34871–34879.

(54) Komeiji, Y.; Ueno, Y.; Uebayasi, M. *FEBS Letters* **2002**, *521*, 133–139.

(55) Yang, C.; Jas, G. S.; Kuczera, K. *Biochimica Et Biophysica Acta-Proteins And Proteomics* **2004**, *1697*, 289–300.

(56) Likić, V. A.; Gooley, P. R.; Speed, T. P.; Strehler, E. E. *Protein Science* **2005**, *14*, 2955–2963.

(57) Zhang, S.; Jin, G.; Zhang, X.-S.; Chen, L. *PROTEOMICS* **2007**, *7*, 2856–2869.

(58) Ovchinnikov, V.; Karplus, M. *Journal of Physical Chemistry B* **2012**, *116*, 8584–8603.

(59) Espinoza-Fonseca, L. M.; Autry, J. M.; Thomas, D. D. *PLoS ONE* **2014**, *9*, e95979–e95979.

(60) Fernández-De Gortari, E.; Espinoza-Fonseca, L. M. *Journal of Biological Chemistry* **2018**, *293*, 12405–12414.

(61) Piana, S.; Shaw, D. E. *Journal of Physical Chemistry B* **2018**, *122*, 11440–11449.

(62) Wang, Q.; Pechersky, Y.; Sagawa, S.; Pan, A. C.; Shaw, D. E. *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116*, 9390–9399.

(63) Smart, J. L.; McCammon, J. A. *Biophysical Journal* **1998**, *75*, 1679–1688.

(64) Tai, K.; Bond, S.; MacMillan, H.; Baker, N.; Holst, M. *Biophysical Journal* **2003**.

(65) Dickinson, E. J.; Ekström, H.; Fontes, E. *Electrochemistry Communications* **2014**, *40*, 71–74.

(66) Hess, B.; Van Der Vegt, N. F. A. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 13296–13300.

(67) Allen, M. P. *NIC Series* **2004**, *23*, 1–28.

(68) Coveney, P. V.; Wan, S. *Physical Chemistry Chemical Physics* **2016**, *18*, 30236–30240.

(69) Weiner, S. J.; Kollman, P. A.; Singh, U. C.; Case, D. A.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *Journal of the American Chemical Society* **1984**, *106*, 765–784.

(70) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(71) Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues, 2004.

(72) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(73) Onufriev, A. V.; Case, D. A. *Annual Review of Biophysics* **2019**, *48*, 275–296.

(74) MacKerell, A. D. et al. *The Journal of Physical Chemistry B* **1998**, *102*, 3586–3616.

(75) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. *Current Opinion in Structural Biology* **2015**, *31*, 64–74.

(76) Lee, H. S.; Qi, Y.; Im, W. *Scientific Reports* **2015**, *5*, 8926.

(77) Shaw, D. E. et al. *Commun. ACM* **2008**, *51*, 91–97.

(78) Hamelberg, D.; Mongan, J.; McCammon, J. A. *The Journal of Chemical Physics* **2004**, *120*, 11919–11929.

(79) Shea, J.-E.; Brooks III, C. L. *Annual Review of Physical Chemistry* **2002**, *52*, 499–535.

(80) Best, R. B.; Chen, Y.-G.; Hummer, G. *Structure* **2005**, *13*, 1755–1763.

(81) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. *Journal of Chemical Physics* **2007**, *126*, 155102.

(82) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 19011–19016.

(83) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. *Journal of Chemical Physics* **2013**, *139*, 015102.

(84) Zhou, H. X.; Bates, P. A. *Current Opinion in Structural Biology* **2013**, *23*, 887–893.

(85) Huber, G. A.; McCammon, J. A. *Computer Physics Communications* **2010**, *181*, 1896–1905.

(86) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proceedings of the National Academy of Sciences* **2001**, *98*, 10037–10041.

(87) Wang, J.; Cieplak, P.; Kollman, P. A. *Journal of Computational Chemistry* **2000**, *21*, 1049–1074.

(88) Sun, B.; Cook, E. C.; Creamer, T. P.; Kekenes-Huskey, P. M. *Biochimica et Biophysica Acta - General Subjects* **2018**, *1862*, 2651–2659.

(89) McPhalen, C. A.; Strynadka, N. C. J.; James, M. N. G. In *Metalloproteins: Structural Aspects*, Anfinsen, C. B., Edsall, J. T., Richards, F. M., Eisenberg, D. S., Eds.; Advances in Protein Chemistry, Vol. 42; Academic Press: 1991, pp 77–144.

(90) Carafoli, E.; Krebs, J. *Journal of Biological Chemistry* **2016**, *291*, 20849–20857.

(91) Blum, L. *Molecular Physics* **1975**, *30*, 1529–1535.

(92) Nonner, W.; Gillespie, D.; Henderson, D.; Eisenberg, B. *Journal of Physical Chemistry B* **2001**, *105*, 6427–6436.

(93) Nonner, W.; Catacuzzeno, L.; Eisenberg, B. *Biophysical Journal* **2000**, *79*, 1976–1992.

(94) Kucharski, A. N.; Scott, C. E.; Davis, J. P.; Kekenes-Huskey, P. M. *The Journal of Physical Chemistry B* **2016**, *120*, 8617–8630.

(95) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. *Journal of Chemical Theory and Computation* **2013**, *9*, 2733–2748.

(96) Jochim, A. L.; Arora, P. S. *Mol. BioSyst.* **2009**, *5*, 924–926.

(97) Nevola, L.; Giralt, E. *Chem. Commun.* **2015**, *51*, 3302–3315.

(98) Jochim, A. L.; Arora, P. S. *Molecular BioSystems* **2009**, *5*, 924–926.

(99) Bullock, B. N.; Jochim, A. L.; Arora, P. S. *Journal of the American Chemical Society* **2011**, *133*, 14220–14223.

(100) Pierce, B. G.; Wiehe, K.; Hwang, H.; Kim, B.-H.; Vreven, T.; Weng, Z. *Bioinformatics* **2014**, *30*, 1771–1773.

(101) Chaudhury, S.; Berrondo, M.; Weitzner, B. D.; Muthu, P.; Bergman, H.; Gray, J. J. *PLoS ONE* **2011**, *6*, DOI: 10.1371/journal.pone.0022477.

(102) Hu, X.; Lee, M. S.; Wallqvist, A. *Biochemistry* **2009**, *48*, 11158–11160.

(103) Schiffer, J. M.; Malmstrom, R. D.; Parnell, J.; Ramirez-Sarmiento, C.; Reyes, J.; Amaro, R. E.; Komives, E. A. *Structure* **2016**, *24*, 1248–1256.

(104) Bui, J. M.; Gsponer, J. *Structure* **2014**, *22*, 1196–1203.

(105) Smith, G. R.; Sternberg, M. J. E. *Current Opinion in Structural Biology* **2002**, *12*, 28–35.

(106) Huang, P.-S.; Love, J. J.; Mayo, S. L. *Journal of Computational Chemistry* **2005**, *26*, 1222–1232.

(107) Shentu, Z.; Al Hasan, M.; Bystroff, C.; Zaki, M. J. *Proteins: Structure, Function, and Bioinformatics* **2008**, *70*, 1056–1073.

(108) Sheinerman, F. B.; Norel, R.; Honig, B. *Current Opinion in Structural Biology* **2000**, *10*, 153–159.

(109) Jackson, R. M.; Sternberg, M. J. E. *Journal of Molecular Biology* **1995**, *250*, 258–275.

(110) Padhorny, D.; Kazennov, A.; Zerbe, B. S.; Porter, K. A.; Xia, B.; Mottarella, S. E.; Kholodov, Y.; Ritchie, D. W.; Vajda, S.; Kozakov, D. *Proceedings of the National Academy of Sciences* **2016**, *113*, E4286–E4293.

(111) Harrison, R. W.; Kourinov, I. V.; Andrews, L. C. *Protein Engineering, Design and Selection* **1994**, *7*, 359–369.

(112) Esmaielbeiki, R.; Nebel, J.-C. *BMC Bioinformatics* **2014**, *15*, 171.

(113) Wang, C.; Schueler-Furman, O.; Baker, D. *Protein Science* **2005**, *14*, 1328–1339.

(114) Kar, R. K.; Gazova, Z.; Bednarikova, Z.; Mroue, K. H.; Ghosh, A.; Zhang, R.; Ulicna, K.; Siebert, H.-C.; Nifantiev, N. E.; Bhunia, A. *Biomacromolecules* **2016**, *17*, 1998–2009.

(115) Innocenzi, P.; Malfatti, L. *Chem. Soc. Rev.* **2013**, *42*, 4198–4216.

(116) Yeh, L. H.; Zhang, M.; Qian, S. *Analytical Chemistry* **2013**, *85*, 7527–7534.

(117) Smeets, R. M. M.; Keyser, U. F.; Krapf, D.; Wu, M. Y.; Dekker, N. H.; Dekker, C. *Nano Letters* **2006**, *6*, 89–95.

(118) Duan, C.; Majumdar, A. *Nature Nanotechnology* **2010**, *5*, 848–852.

(119) Singh, K. P.; Kumar, M. *The Journal of Physical Chemistry C* **2011**, *115*, 22917–22924.

(120) Daiguji, H.; Yang, P.; Majumdar, A. *Nano Letters* **2004**, *4*, 137–142.

(121) Schoch, R. B.; Han, J.; Renaud, P. *Rev. Mod. Phys.* **2008**, *80*, 839–883.

(122) Chang, C. C.; Kazoe, Y.; Morikawa, K.; Mawatari, K.; Yang, R. J.; Kitamori, T. *Analytical Chemistry* **2013**, *85*, 4468–4474.

(123) Atalay, S.; Yeh, L. H.; Qian, S. *Langmuir* **2014**, *30*, 13116–13120.

(124) Cheng, L.-J.; Guo, L. J. *Nano Letters* **2007**, *7*, 3165–3171.

(125) Yu, M. M.; Noble, R. D. R. D.; Falconer, J. L. J. L. *Accounts of Chemical Research* **2011**, *44*, 1196–1206.

(126) Yan, Y.; Davis, M. E.; Gavalas, G. R. *Journal of Membrane Science* **1997**, *123*, 95–103.

(127) Bonilla, G.; Tsapatsis, M.; Vlachos, D. G.; Xomeritakis, G. *Journal of Membrane Science* **2001**, *182*, 103–109.

(128) John, N.; Stevens, S.; Terasaki, O.; Anderson, M. *Chemistry - A European Journal* **2010**, *16*, 2220–2230.

(129) Zhou, H.-X.; Rivas, G.; Minton, A. P. *Annual Review of Biophysics* **2008**, *37*, 375–397.

(130) Roeffaers, M. B. J.; Ameloot, R.; Baruah, M.; Uji-i, H.; Bulut, M.; De Cremer, G.; Müller, U.; Jacobs, P. A.; Hofkens, J.; Sels, B. F.; De Vos, D. E. *Journal of the American Chemical Society* **2008**, *130*, 5763–5772.

(131) Thomas, J. M.; Gai, P. L. In *Advances in Catalysis*; Elsevier: 2004; Vol. 48, pp 171–227.

(132) HUNGER, M.; KARGER, J.; PFEIFER, H.; CARO, J.; ZIBROWIUS, B.; BULOW, M.; MOSTOWICZ, R. *Journal of the Chemical Society-Faraday Transactions I* **1987**, *83*, 3459–&.

(133) Valiullin, R.; Kortunov, P.; Kärger, J.; Timoshenko, V. *Journal of Chemical Physics* **2004**, *120*, 11804–11814.

(134) Auerbach, S.; Kärger, J.; Vasenkov, S. **2003**, 97.

(135) Zhou, W.; Wang, Z. *Springer* **2006**.

(136) Zhao, D. Z.; Yang, P.; Melosh, N.; Feng, J.; Chmelka, B. F.; Stucky, G. D. *Advanced Materials* **1998**, *10*, 1380–1385.

(137) Coasne, B.; Galarneau, A.; Di Renzo, F.; Pellenq, R. J. M. *The Journal of Physical Chemistry C* **2007**, *111*, 15759–15770.

(138) Coasne, B.; Pellenq, R. J.-M. *The Journal of Chemical Physics* **2004**, *121*, 3767–3774.

(139) Alexander, S. K.; Azencott, R.; Bodmann, B. G.; Bouamrani, A.; Chiappini, C.; Ferrari, M.; Liu, X.; Tasciotti, E. In; Springer, Berlin, Heidelberg: 2009, pp 590–597.

(140) Zhao, C.; Qiao, Y. *Nanoscale* **2016**, *8*, 17658–17664.

(141) Hamngren Blomqvist, C.; Abrahamsson, C.; Gebäck, T.; Altskär, A.; Hermansson, A.-M.; Nydén, M.; Gustafsson, S.; Lorén, N.; Olsson, E. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2015**, *484*, 288–296.

(142) Wojcik, T. R.; Krapf, D. *IEEE Photonics Journal* **2011**, *3*, 337–343.

(143) Okuda, M.; Takeguchi, M.; Zhu, Y.; Hashimoto, A.; Ogawa, N.; Tagaya, M.; Chen, S.; Hanagata, N.; Ikoma, T. *Surface and Interface Analysis* **2010**, *42*, 1548–1551.

(144) Zečević, J.; de Jong, K. P.; de Jongh, P. E. *Current Opinion in Solid State and Materials Science* **2013**, *17*, 115–125.

(145) Daiguji, H. *Chem. Soc. Rev.* **2010**, *39*, 901–911.

(146) Shirono, K.; Tatsumi, N.; Daiguji, H. *The Journal of Physical Chemistry B* **2009**, *113*, 1041–1047.

(147) Joseph, S.; Aluru, N. R. *Langmuir* **2006**, *22*, 9041–9051.

(148) Lopes, P. E. M.; Murashov, V.; Tazi, M.; Demchuk, E.; MacKerell, A. D. *The Journal of Physical Chemistry B* **2006**, *110*, 2782–2792.

(149) Yeh, L.-H.; Xue, S.; Joo, S. W.; Qian, S.; Hsu, J.-P. *The Journal of Physical Chemistry C* **2012**, *116*, 4209–4216.

(150) Barisik, M.; Atalay, S.; Beskok, A.; Qian, S. *The Journal of Physical Chemistry C* **2014**, *118*, 1836–1842.

(151) Atalay, S.; Yeh, L.-H.; Qian, S. *Langmuir* **2014**, *30*, 13116–13120.

(152) Clark Wooten, M. K.; Koganti, V. R.; Zhou, S.; Rankin, S. E.; Knutson, B. L. *ACS Applied Materials and Interfaces* **2016**, *8*, 21806–21815.

(153) Van Trees, H. L., *Detection, Estimation, and Modulation Theory, Part I*; John Wiley & Sons, Inc.: New York, USA, 2001, p 716.

(154) Geuzaine, C.; Remacle, J.-F. *International Journal for Numerical Methods in Engineering* **2009**, *79*, 1309–1331.

(155) Liao, T.; Zhang, Y.; Kekenes-Huskey, P. M.; Cheng, Y.; Michailova, A.; McCulloch, A. D.; Holst, M.; McCammon, J. A. *Molecular Based Mathematical Biology (Online)* **2013**, *1*, 164–179.

(156) Daiguji, H.; Yang, P.; Majumdar, A. *Nano Letters* **2004**, *4*, 137–142.

(157) Mitscha-Baude, G.; Buttinger-Kreuzhuber, A.; Tulzer, G.; Heitzinger, C. *Journal of Computational Physics* **2017**, *338*, 452–476.

(158) Yeh, L.-H.; Ma, Y.; Xue, S.; Qian, S. *Sensors & Actuators: B. Chemical* **2015**, *215*, 266–271.

(159) Datta, S.; Conlisk, A.; Li, H.; Yoda, M. *Mechanics Research Communications* **2009**, *36*, 65–74.

(160) Kirby, B. J.; Hasselbrink, E. F. *ELECTROPHORESIS* **2004**, *25*, 187–202.

(161) Fuest, M.; Boone, C.; Conlisk, A.; Prakash, S. In *2015 Transducers - 2015 18th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*, IEEE: 2015, pp 247–250.

(162) Van Der Heyden, F. H. J.; Stein, D.; Besteman, K.; Lemay, S. G.; Dekker, C. *Physical Review Letters* **2006**, *96*, DOI: 10.1103/PhysRevLett.96.224502.

(163) Israelachvili, J. N., *No Title*; Academic Press: 2011, p 704.

(164) Alnæs, M.; Blechta, J.; Hake, J.; Johansson, A.; Kehlet, B.; Logg, A.; Richardson, C.; Ring, J.; Rognes, M.; Wells, G. *Archive of Numerical Software* **2015**, *3*, DOI: 10.11588/ans.2015.100.20553.

(165) Ma, Y.; Yeh, L. H.; Lin, C. Y.; Mei, L.; Qian, S. *Analytical Chemistry* **2015**, *87*, 4508–4514.

(166) Ma, Y.; Xue, S.; Hsu, S.-C.; Yeh, L.-H.; Qian, S.; Tan, H. *Physical Chemistry Chemical Physics* **2014**, *16*, 20138.

(167) Ahrens, J.; Geveci, B.; Law, C. *Elsevier* **2005**.

(168) Stein, W. *Academic Press* **1986**.

(169) Kekenes-Huskey, P. M.; Scott, C. E.; Atalay, S. *The Journal of Physical Chemistry B* **2016**, *120*, 8696–8706.

(170) Yeh, L.-H.; Xue, S.; Joo, S. W.; Qian, S.; Hsu, J.-P. *The Journal of Physical Chemistry C* **2012**, *116*, 4209–4216.

(171) Karnik, R.; Castelino, K.; Fan, R.; Yang, P.; Majumdar, A. *Nano Letters* **2005**, *5*, 1638–1642.

(172) Karnik, R.; Castelino, K.; Fan, R.; Yang, P.; Majumdar, A. *Nano Letters* **2005**, *5*, 1638–1642.

(173) Setny, P.; Baron, R.; Kekenes-Huskey, P.; McCammon, J.; Dzubiella, J. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 1197–1202.

(174) Hausman, R.; Digman, B.; Escobar, I. C.; Coleman, M.; Chung, T. S. *Journal of Membrane Science* **2010**, *363*, 195–203.

(175) Wanunu, M.; Meller, A. *Nano Letters* **2007**, *7*, 1580–1585.

(176) Yameen, B.; Ali, M.; Neumann, R.; Ensinger, W.; Knoll, W.; Azzaroni, O. *Journal of the American Chemical Society* **2009**, *131*, 2070–2071.

(177) Wagner, J.; Keizer, J. *Biophysical Journal* **1994**, *67*, 447–456.

(178) Lin, X.; Yang, Q.; Ding, L.; Su, B. *ACS Nano* **2015**, *9*, 11266–11277.

(179) Sun, Z.; Deng, Y.; Wei, J.; Gu, D.; Tu, B.; Zhao, D. *Chemistry of Materials* **2011**, *23*, 2176–2184.

(180) Calvo, A.; Yameen, B.; Williams, F. J.; Soler-Illia, G. J.; Azzaroni, O. *Journal of the American Chemical Society* **2009**, *131*, 10866–10868.

(181) Yang, S.-A.; Klee, C. B. *Biochemistry* **2000**, *39*, 16147–16154.

(182) Rumi-Masante, J.; Rusinga, F. I.; Lester, T. E.; Dunlap, T. B.; Williams, T. D.; Dunker, A. K.; Weis, D. D.; Creamer, T. P. *Journal of Molecular Biology* **2012**, *415*, 307–317.

(183) Kursula, P. *Amino Acids* **2014**, *46*, 2295–2304.

(184) Quintana, A. R.; Wang, D.; Forbes, J. E.; Waxham, M. N. *Biochemical and Biophysical Research Communications* **2005**, *334*, 674–680.

(185) Klee, C. B.; Ren, H.; Wang, X. *Journal of Biological Chemistry* **1998**, *273*, 13367–13370.

(186) Lewis, R. S. *Annual Review of Immunology* **2001**, *19*, 497–521.

(187) Persechini, A.; Cronk, B. *Journal of Biological Chemistry* **1999**, *274*, 6827–6830.

(188) Dunlap, T. B.; Cook, E. C.; Rumi-Masante, J.; Arvin, H. G.; Lester, T. E.; Creamer, T. P. *Biochemistry* **2013**, *52*, 8643–8651.

(189) Marsh, J. A.; Forman-Kay, J. D. *Biophysical Journal* **2010**, *98*, 2383–2390.

(190) Dobrynin, A. V.; Colby, R. H.; Rubinstein, M. *Journal of Polymer Science Part B: Polymer Physics* **2004**, *42*, 3513–3538.

(191) Sherry, K. P.; Das, R. K.; Pappu, R. V.; Barrick, D. *Proceedings of the National Academy of Sciences* **2017**, *114*, E9243–E9252.

(192) Simons, K. T.; Bonneau, R.; Ruczinski, I.; Baker, D. *Proteins: Structure, Function, and Bioinformatics* **1999**, *37*, 171–176.

(193) Case, D. A. et al. Amber 14; University of California, San Francisco, 2014.

(194) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.

(195) Berezovska, G.; Prada-Gracia, D.; Mostarda, S.; Rao, F. *The Journal of Chemical Physics* **2012**, *137*, 194101.

(196) Dunlap, T. B.; Guo, H.-F.; Cook, E. C.; Holbrook, E.; Rumi-Masante, J.; Lester, T. E.; Colbert, C. L.; Vander Kooi, C. W.; Creamer, T. P. *Biochemistry* **2014**, *53*, 5779–5790.

(197) Xiong, P.; Wang, M.; Zhou, X.; Zhang, T.; Zhang, J.; Chen, Q.; Liu, H. *Nature Communications* **2014**, *5*, 5330.

(198) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

(199) Henriques, J.; Cragnell, C.; Skepö, M. *Journal of Chemical Theory and Computation* **2015**, *11*, 3420–3431.

(200) Maffucci, I.; Contini, A. *Journal of Chemical Theory and Computation* **2016**, *12*, 714–727.

(201) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *Journal of Computational Physics* **1977**, *23*, 327–341.

(202) MacKerell, A. D. et al. *The Journal of Physical Chemistry B* **1998**, *102*, 3586–3616.

(203) MacKerell, A. D.; Feig, M.; Brooks, C. L. *Journal of the American Chemical Society* **2004**, *126*, 698–699.

(204) Grossfield, A.

(205) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. *The Journal of chemical physics* **2009**, *131*, 124101.

(206) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Research* **2007**, *35*.

(207) Van Valen, D.; Haataja, M.; Phillips, R. *Biophysical Journal* **2009**, *96*, 1275–1292.

(208) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. In *Proceedings of the ...* National Acad Sciences: 2000; Vol. 97, p 8868.

(209) Manalan, A. S.; Klee, C. B. *Proceedings of the National Academy of Sciences* **1983**, *80*, 4291–4295.

(210) Onufriev, A. In, Wheeler, R. A., Spellmeyer, D. C., Eds.; Annual Reports in Computational Chemistry, Vol. 4; Elsevier: 2008, pp 125–137.

(211) Ruhong, Z. *Proteins: Structure, Function, and Bioinformatics*, *53*, 148–161.

(212) Grabarek, Z. *Journal Of Molecular Biology* **2006**, *359*, 509–525.

(213) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. *Journal of Molecular Biology* **2004**, *338*, 1015–1026.

(214) Sivakolundu, S. G.; Bashford, D.; Kriwacki, R. W. *Journal of Molecular Biology* **2005**, *353*, 1118–1128.

(215) Yoon, M.-K.; Venkatachalam, V.; Huang, A.; Choi, B.-S.; Stultz, C. M.; Chou, J. J. *Protein Science* **2009**, *18*, 337–347.

(216) Kjaergaard, M.; Teilum, K.; Poulsen, F. M. *Proceedings of the National Academy of Sciences* **2010**, *107*, 12535–12540.

(217) Song, D.; Wang, W.; Ye, W.; Ji, D.; Luo, R.; Chen, H.-F. *Chemical Biology and Drug Design* **2017**, *89*.

(218) Guo, X.; Han, J.; Luo, R.; Chen, H.-F. *RSC Adv.* **2017**, *7*, 29713–29721.

(219) Zhou, H.-X.; Wlodek, S. T.; McCammon, J. A. *Proceedings of the National Academy of Sciences* **1998**, *95*, 9280–9283.

(220) Zhou, H.-X. *Biophysical Journal* **2010**, *98*, L15–L17.

(221) Roca, M.; Messer, B.; Warshel, A. *FEBS Letters* **2007**, *581*, 2065–2071.

(222) Liu, B.; Chia, D.; Csizmok, V.; Farber, P.; Forman-Kay, J. D.; Gradinaru, C. C. *The Journal of Physical Chemistry B* **2014**, *118*, 4088–4097.

(223) Wicky, B. I. M.; Shammas, S. L.; Clarke, J. *Proceedings of the National Academy of Sciences* **2017**, *114*, 9882–9887.

(224) Schreiber, G.; Fersht, A. R. *Nature Structural Biology* **1996**, *3*.

(225) Sheinerman, F. B.; Norel, R.; Honig, B. *Current Opinion in Structural Biology* **2000**, *10*, 153–159.

(226) Vijayakumar, M.; Wong, K.-Y.; Schreiber, G.; Fersht, A. R.; Szabo, A.; Zhou, H.-X. *Journal of Molecular Biology* **1998**, *278*, 1015–1024.

(227) Gabdoulline, R.; Wade, R. *Biophysical Journal* **1997**, *72*, 1917–1929.

(228) Alsallaq, R.; Zhou, H.-X. *Proteins: Structure, Function, and Bioinformatics* **2008**, *71*, 320–335.

(229) Lee, L.-P.; Tidor, B. *Protein Science* **2001**, *10*, 362–377.

(230) Green, D. F.; Tidor, B. *Proteins: Structure, Function, and Bioinformatics* **2005**, *60*, 644–657.

(231) Gabdoulline, R. R.; Wade, R. C. *Journal of Molecular Recognition* **1999**, *12*, 226–234.

(232) Tang, C.; Iwahara, J.; Clore, M. G. *Nature* **2006**, *444*, 383–386.

(233) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nature Chemical Biology* **2009**, *5*, 789–796.

(234) Onufriev, A.; Case, D. A.; Bashford, D. *J Comput Chem* **2002**, *23*, 1297–1304.

(235) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins-Structure Function And Bioinformatics* **2004**, *55*, 383–394.

(236) Tzul, F. O.; Schweiker, K. L.; Makhatadze, G. I. *Proceedings of the National Academy of Sciences* **2015**, *112*, 201410424.

(237) Shukla, D.; Peck, A.; Pande, V. S. *Nature Communications* **2016**, *7*.

(238) Rusnak, F.; Mertz, P. *Physiological Reviews* **2000**, *80*, 1483–1521.

(239) Griffith, J. P.; Kim, J. L.; Kim, E. E.; Sintchak, M. D.; Thomson, J. A.; Fitzgibbon, M. J.; Fleming, M. A.; Caron, P. R.; Hsiao, K.; Navia, M. A. *Cell* **1995**, *82*, 507–522.

(240) Li, H.; Zhang, L.; Rao, A.; Harrison, S. C.; Hogan, P. G. *Journal of Molecular Biology* **2007**, *369*, 1296–1306.

(241) Ye, Q.; Wang, H.; Zheng, J.; Wei, Q.; Jia, Z. *Proteins: Structure, Function and Genetics* **2008**, *73*, 19–27.

(242) Jin, L.; Harrison, S. C. *Proceedings of the National Academy of Sciences* **2002**, *99*, 13522–13526.

(243) Sheftic, S. R.; Page, R.; Peti, W. *Scientific Reports* **2016**, *6*, 38920.

(244) Takeuchi, K.; Roehrl, M. H.; Sun, Z. Y. J.; Wagner, G. *Structure* **2007**, *15*, 587–597.

(245) Ye, Q.; Li, X.; Wong, A.; Wei, Q.; Jia, Z. *Biochemistry* **2006**, *45*, 738–745.

(246) Majava, V.; Kursula, P. *PLoS ONE* **2009**, *4*, ed. by Ratna, B., e5402.

(247) Fruman, D. A.; Klee, C. B.; Bierer, B. E.; Burakoff, S. J. *Proceedings of the National Academy of Sciences* **1992**, *89*, 3686–3690.

(248) Parsons, J. N.; Wiederrecht, G. J.; Salowe, S.; Burbaum, J. J.; Rokosz, L. L.; Kincaid, R. L.; O'Keefe, S. J. *Journal of Biological Chemistry* **1994**, *269*, 19610–19616.

(249) Li, J.; Jia, Z. G.; Zhou, W. C.; Wei, Q. *Proteins: Structure, Function and Bioinformatics* **2009**, *77*, 612–623.

(250) Harish, B. M.; Saraswathi, R.; Vinod, D.; Devaraju, K. S. *Journal of Biomolecular Structure and Dynamics* **2016**, *34*, 983–992.

(251) Rodríguez, A.; Roy, J.; Martínez-Martínez, S.; López-Maderuelo, M. D.; Niño-Moreno, P.; Ortí, L.; Pantoja-Uceda, D.; Pineda-Lucena, A.; Cyert, M. S.; Redondo, J. M. *Molecular Cell* **2009**, *33*, 616–626.

(252) Grigoriu, S.; Bond, R.; Cossio, P.; Chen, J. A.; Ly, N.; Hummer, G.; Page, R.; Cyert, M. S.; Peti, W. *PLoS Biology* **2013**, *11*, ed. by Petsko, G. A., e1001492.

(253) Song, R.; Li, J.; Zhang, J.; Wang, L.; Tong, L.; Wang, P.; Yang, H.; Wei, Q.; Cai, H.; Luo, J. *Biochimie* **2017**, *142*, 158–167.

(254) Nagulapalli, M.; Parigi, G.; Yuan, J.; Gsponer, J.; Deraos, G.; Bamm, V. V.; Harauz, G.; Matsoukas, J.; De Planque, M. R.; Gerothanassis, I. P.; Babu, M. M.; Luchinat, C.; Tzakos, A. G. *Structure* **2012**, *20*, 522–533.

(255) Gsponer, J.; Christodoulou, J.; Cavalli, A.; Bui, J. M.; Richter, B.; Dobson, C. M.; Vendruscolo, M. *Structure* **2008**, *16*, 736–746.

(256) Liu, F.; Chu, X.; Lu, H. P.; Wang, J. *Proceedings of the National Academy of Sciences* **2017**, *114*, E3927–E3934.

(257) Sun, B.; Cook, E. C.; Creamer, T. P.; Kekenes-Huskey, P. M. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2018**, *1862*, 2651–2659.

(258) Dunlap, T. B.; Guo, H.-F.; Cook, E. C.; Holbrook, E.; Rumi-Masante, J.; Lester, T. E.; Colbert, C. L.; Vander Kooi, C. W.; Creamer, T. P. *Biochemistry* **2014**, *53*, 5779–5790.

(259) Van Lierop, J. E.; Wilson, D. P.; Davis, J. P.; Tikunova, S.; Sutherland, C.; Walsh, M. P.; David Johnson, J. *Journal of Biological Chemistry* **2002**, *277*, 6550–6558.

(260) Zhi, G.; Abdullah, S. M.; Stull, J. T. *Journal of Biological Chemistry* **1998**, *273*, 8951–8957.

(261) Soderling, T. R.; Stull, J. T. Structure and regulation of calcium/calmodulin-dependent protein kinases, 2001.

(262) Case, D. et al. Amber 16, University of California, San Francisco. 2016.

(263) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *Journal of Computational Chemistry* **2005**, *26*, 1668–1688.

(264) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry* **1996**, *100*, 19824–19839.

(265) Patriksson, A.; van der Spoel, D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073–2077.

(266) Zhang, W.; Wu, C.; Duan, Y. *The Journal of Chemical Physics* **2005**, *123*, 154105.

(267) Patel, S.; Vierling, E.; Tama, F. *Biophysical Journal* **2014**, *106*, 2644–2655.

(268) Mintseris, J.; Pierce, B.; Wiehe, K.; Anderson, R.; Chen, R.; Weng, Z. *Proteins: Structure, Function, and Bioinformatics* **2007**, *69*, 511–520.

(269) Chen, R.; Weng, Z. *Proteins: Structure, Function and Genetics* **2002**, *47*, 281–294.

(270) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.

(271) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

(272) Malmstrom, R. D.; Lee, C. T.; Van Wart, A. T.; Amaro, R. E. *Journal of Chemical Theory and Computation* **2014**, *10*, 2648–2657.

(273) Genheden, S.; Ryde, U. *Expert Opinion on Drug Discovery* **2015**, *10*, 449–461.

(274) Roe, D. R.; Cheatham, T. E. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084–3095.

(275) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.

(276) Fiorin, G.; Klein, M. L.; Hénin, J. *Molecular Physics* **2013**, *111*, 3345–3362.

(277) Cho, M. J.; Vaghy, P. L.; Kondo, R.; Lee, S. H.; Davis, J. P.; Rehl, R.; Heo, W. D.; David Johnson, J. *Biochemistry* **1998**, *37*, 15593–15597.

(278) Guharoy, M.; Chakrabarti, P. *Bioinformatics* **2007**, *23*, 1909–1918.

(279) de A. Marques, M.; Parvatiyar, M. S.; Yang, W.; de Oliveira, G.; Pinto, J. R. *Archives of Biochemistry and Biophysics* **2018**, DOI: 10.1016/J.ABB.2018.12.026.

(280) Dunlap, T. B.; Kirk, J. M.; Pena, E. A.; Yoder, M. S.; Creamer, T. P. *Proteins: Structure, Function and Bioinformatics* **2013**, *81*, 607–612.

(281) Sun, B.; Vaughan, D. M.; Tikunova, S. B.; Creamer, T. P.; Davis, J. P.; Kekenes-Huskey, P. M. *Biochemistry* **2019**, *58*, 4070–4085.

(282) Zamoon, J.; Nitu, F.; Karim, C.; Thomas, D. D.; Veglia, G. *Proceedings of the National Academy of Sciences* **2005**, *102*, 4747.

(283) Zhou, G.; Pantelopulos, G. A.; Mukherjee, S.; Voelz, V. A. *Biophysical Journal* **2017**, *113*, 785–793.

(284) Paul, F.; Noé, F.; Weikl, T. R. *Journal of Physical Chemistry B* **2018**, *122*, 5649–5656.

(285) Grishaev, A.; Anthis, N. J.; Clore, G. M. *Journal of the American Chemical Society* **2012**, *134*, 14686–14689.

(286) Dogan, J.; Gianni, S.; Jemth, P. The binding mechanisms of intrinsically disordered proteins, 2014.

(287) Hoeflich, K. P.; Ikura, M. **2002**, *108*, 739–742.

(288) Blumenthal, D. K.; Takio, K.; Edelman, A. M.; Charbonneau, H.; Titani, K.; Walsh, K. A.; Krebs, E. G. *Proceedings of the National Academy of Sciences* **1985**, *82*, 3187–3191.

(289) Fitzsimons, D. P.; Herring, B. P.; Stull, J. T.; Gallagher, P. J. *Journal of Biological Chemistry* **1992**, *267*, 23903–23909.

(290) Siddiqui, J. K.; Tikunova, S. B.; Walton, S. D.; Liu, B.; Meyer, M.; de Tombe, P. P.; Neilson, N.; Kekenes-Huskey, P. M.; Salhi, H. E.; Janssen, P. M.; Biesiadecki, B. J.; Davis, J. P. *Frontiers in Physiology* **2016**, *7*, 632.

(291) Fruman, D. A.; Pai, S. Y.; Burakoff, S. J.; Bierer, B. E. *Molecular and Cellular Biology* **1995**, *15*, 3857–3863.

(292) Sagoo, J. K.; Fruman, D. A.; Wesselborg, S.; Walsh, C. T.; Bierer, B. E. *The Biochemical journal* **1996**, *320 (Pt 3*, 879–84.

(293) Borcherds, W.; Becker, A.; Chen, L.; Chen, J.; Chemes, L. B.; Daughdrill, G. W. *Biophysical Journal* **2017**, *112*, 2038–2042.

(294) Stemmer, P. M.; Klee, C. B. *Biochemistry* **1994**, *33*, 6859–6866.

(295) Qin, Y.; Liu, J.; Li, X.; Wei, Q. *Biochimica et Biophysica Acta - Proteins and Proteomics* **2005**, *1747*, 171–178.

(296) Peng, K.; Radivojac, P.; Vucetic, S.; Dunker, A. K.; Obradovic, Z. *BMC Bioinformatics* **2006**, *7*, 208.

(297) Das, R. K.; Pappu, R. V. *Proceedings of the National Academy of Sciences* **2013**, *110*, 13392–13397.

(298) Lin, Y. S.; Bowman, G. R.; Beauchamp, K. A.; Pande, V. S. *Biophysical Journal* **2012**, *102*, 315–324.

(299) Fluitt, A. M.; De Pablo, J. J. *Biophysical Journal* **2015**, *109*, 1009–1018.

(300) Guo, Z.; Mohanty, U.; Noehre, J.; Sawyer, T. K.; Sherman, W.; Krilov, G. *Chemical Biology and Drug Design* **2010**, *75*, 348–359.

(301) Miller, Y.; Ma, B.; Nussinov, R. *Proceedings of the National Academy of Sciences* **2010**, *107*, 9490–9495.

(302) Best, R. B.; Mittal, J. *Proteins: Structure, Function and Bioinformatics* **2011**, *79*, 1318–1328.

(303) Zhang, M.; Pascal, J. M.; Zhang, J.-F. *Proceedings of the National Academy of Sciences* **2013**, *110*, 4828–4833.

(304) Toyoshima, C.; Nakasako, M.; Nomura, H.; Ogawa, H. *Nature* **2000**, *405*, 647–655.

(305) Toyoshima, C.; Nomura, H. *Nature* **2002**, *418*, 605–611.

(306) Sørensen, T. L. M.; Clausen, J. D.; Jensen, A. M. L.; Vilsen, B.; Møller, J. V.; Andersen, J. P.; Nissen, P. *Journal of Biological Chemistry* **2004**, *279*, 46355–46358.

(307) Olesen, C.; Picard, M.; Winther, A. M. L.; Gyrup, C.; Morth, J. P.; Oxvig, C.; Møller, J. V.; Nissen, P. *Nature* **2007**, *450*, 1036–1042.

(308) Inesi, G.; Scarpa, A. *Biochemistry* **1972**, *11*, 356–359.

(309) Froehlich, J. P.; Taylor, E. W. *Journal of Biological Chemistry* **1976**, *251*, 2307–2315.

(310) Dupont, Y. *Biochemical and biophysical research communications* **1976**, *71*, 544–50.

(311) Dupont, Y.; Leigh, J. B. *Nature* **1978**, *273*, 396–398.

(312) Coan, C. R.; Inesi, G. *Biochemical and biophysical research communications* **1976**, *71*, 1283–8.

(313) Meis, L.; Vianna, A. L. *Annual Review of Biochemistry* **1979**, *48*, 275–292.

(314) Inesi, G. *Annual Review of Physiology* **1985**, *47*, 573–601.

(315) Mintz, E.; Guillain, F. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1997**, *1318*, 52–70.

(316) Clausen, J. D.; Bublitz, M.; Arnou, B.; Montigny, C.; Jaxel, C.; Møller, J. V.; Nissen, P.; Andersen, J. P.; Le Maire, M. *EMBO Journal* **2013**, *32*, 3231–3243.

(317) Inesi, G.; Kurzmack, M.; Coan, C.; Lewis, D. E. *The Journal of biological chemistry* **1980**, *255*, 3025–3031.

(318) Guillain, F.; Gingold, M. P.; Champeil, P. *Journal of Biological Chemistry* **1982**, *257*, 7366–7371.

(319) Lee, a. G.; Baker, K.; Khan, Y. M.; East, J. M. *The Biochemical journal* **1995**, *305 (Pt 1*, 225–31.

(320) Huang, Y.; Li, H.; Bu, Y. *Journal Of Computational Chemistry* **2009**, *30*, 2136–2145.

(321) Kekenes-Huskey, P. M.; Metzger, V. T.; Grant, B. J.; Andrew McCammon, J. *Protein Science* **2012**, *21*, 1429–1443.

(322) Das, A.; Rui, H.; Nakamoto, R.; Roux, B. *Journal of Molecular Biology* **2017**, *429*, 647–666.

(323) Rui, H.; Das, A.; Nakamoto, R.; Roux, B. *Journal of Molecular Biology* **2018**, *430*, 5050–5065.

(324) Ha, K. N.; Masterson, L. R.; Hou, Z.; Verardi, R.; Walsh, N.; Veglia, G.; Robia, S. L. *Proceedings of the National Academy of Sciences* **2011**, *108*, 2735–2740.

(325) Bidwell, P.; Blackwell, D. J.; Hou, Z.; Zima, A. V. *Journal of Biological ...* **2011**.

(326) Smolin, N.; Robia, S. L. *Journal of Physical Chemistry B* **2015**, *119*, 1407–1415.

(327) Zhao, Y.-T.; Valdivia, C. R.; Gurrola, G. B.; Powers, P. P.; Willis, B. C.; Moss, R. L.; Jalife, J.; Valdivia, H. H. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, E1669–E1677.

(328) Tadini-Buoninsegni, F.; Smeazzetto, S.; Gualdani, R.; Moncelli, M. R. *Frontiers in Molecular Biosciences* **2018**, *5*, 36.

(329) Dudev, T.; Lim, C. *Chemical Reviews* **2014**, *114*, 538–556.

(330) Yu, H.; Rick, S. W. *The Journal of Physical Chemistry B* **2010**, *114*, 11552–11560.

(331) Kiyota, Y.; Takeda-Shitaka, M. *The Journal of Physical Chemistry B* **2014**, *118*, 11496–11503.

(332) Gillespie, D.; Nonner, W.; Henderson, D.; Eisenberg, R. S. *Physical Chemistry Chemical Physics* **2002**, *4*, 4763–4769.

(333) Hughes, E.; Clayton, J. C.; Middleton, D. A. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2009**, *1788*, 559–566.

(334) Tran, K.; Smith, N. P.; Loiselle, D. S.; Crampin, E. J. *Biophysical Journal* **2009**, *96*, 2029–2042.

(335) Berridge, M. J.; Bootman, M. D.; Roderick, H. L. *Nature Reviews Molecular Cell Biology* **2003**, *4*, 517–529.

(336) Clarke, D. M.; Loo, T. W.; Inesi, G.; MacLennan, D. H. *Nature* **1989**, *339*, 476–478.

(337) Inesi, G.; Zhang, Z.; Lewis, D. *Biophysical Journal* **2002**, *83*, 2327–2332.

(338) Inesi, G.; Ma, H.; Lewis, D.; Xu, C. *The Journal of biological chemistry* **2004**, *279*, 31629.

(339) Shigekawa, M.; Pearl, L. J. *The Journal of biological chemistry* **1976**, *251*, 6947–52.

(340) Vianna, A. L. *Biochimica et Biophysica Acta (BBA) - Enzymology* **1975**, *410*, 389–406.

(341) Toyoshima, C.; Iwasawa, S.; Ogawa, H.; Hirata, A.; Tsueda, J.; Inesi, G. *Nature* **2013**, *495*, 260–264.

(342) Jones, L. R. *BBA - Biomembranes* **1979**, *557*, 230–242.

(343) Shannon, R. D.; Prewitt, C. T. *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry* **1969**, *25*, 925–946.

(344) Cordero-Morales, J. F.; Cuello, L. G.; Zhao, Y.; Jogini, V.; Cortes, D. M.; Roux, B.; Perozo, E. *Nature Structural and Molecular Biology* **2006**, *13*, 311–318.

(345) Yamasaki, K.; Wang, G.; Daiho, T.; Danko, S.; Suzuki, H. *Journal of Biological Chemistry* **2008**, *283*, 29144–29155.

(346) Moutin, M. J.; Dupont, Y. *Journal of Biological Chemistry* **1991**, *266*, 5580–5586.

(347) Musgaard, M.; Thøgersen, L.; Schiøtt, B. *Biochemistry* **2011**, *50*, 11109–11120.

(348) Sipido, K. R.; Wier, W. G. *The Journal of Physiology* **1991**, *435*, 605–630.

(349) Luo, C. *Circulation Research* **1994**.

(350) Shannon, T. R.; Wang, F.; Puglisi, J.; Weber, C.; Bers, D. M. *Biophysical Journal* **2004**, *87*, 3351–3371.

(351) Ceholski, D. K.; Trieber, C. A.; Young, H. S. *Journal of Biological Chemistry* **2012**, *287*, 16521–16529.

(352) Higgins, E. R.; Cannell, M. B.; Sneyd, J. *Biophysical Journal* **2006**, *91*, 151–163.

(353) Koivumaki, J. T.; Takalo, J.; Korhonen, T.; Tavi, P.; Weckstrom, M. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2009**, *367*, 2181–2202.

(354) Zhang, Z.; Lewis, D.; Strock, C.; Inesi, G.; Nakasako, M.; Nomura, H.; Toyoshima, C. *Biochemistry* **2000**, *39*, 8758–8767.

(355) Dhitavat, J.; Dode, L.; Leslie, N.; Sakuntabhai, A.; Lorette, G.; Hovnanian, A. *Journal of Investigative Dermatology* **2003**, *121*, 486–489.

(356) Hovnanian, A. *Sub-cellular biochemistry* **2007**, *45*, 337–63.

(357) Xu, A.; Hawkins, C.; Narayanan, N. *Journal of Biological Chemistry* **1993**, *268*, 8394–7.

(358) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Accounts of Chemical Research* **2000**, *33*, 889–897.

(359) Glaves, J. P.; Trieber, C. A.; Ceholski, D. K.; Stokes, D. L.; Young, H. S. *Journal Of Molecular Biology* **2011**, *405*, 707–723.

(360) Schipke, C. G.; Heidemann, A.; Skupin, A.; Peters, O.; Falcke, M.; Kettenmann, H. *Cell Calcium* **2008**, *43*, 285–295.

(361) Chun, B. J.; Stewart, B. D.; Vaughan, D. D.; Bachstetter, A. D.; Kekenes-Huskey, P. M. *The Journal of Physiology*, *0*, DOI: 10.1113/JP277377.

(362) Espinoza-Fonseca, L. M.; Thomas, D. D. *PLoS ONE* **2011**, *6*.

(363) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *Journal of Computational Chemistry*, *25*, 1605–1612.

(364) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. *Journal of Computational Chemistry* **2014**, *35*, 1997–2004.

(365) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *Journal of Computational Chemistry* **2008**, *29*, 1859–1865.

(366) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.

(367) Romo, T. D.; Leioatts, N.; Grossfield, A. *Journal of Computational Chemistry* **2014**, *35*, 2305–2318.

(368) Blum, L. *Journal of Statistical Physics* **1978**, *18*, 451–474.

(369) Blum, L.; Rosenfeld, Y. *Journal of Statistical Physics* **1991**, *63*, 1177–1190.

(370) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. *Journal of Chemical Theory and Computation* **2013**, *9*, 2733–2748.

(371) Smith, N. P.; Crampin, E. J. *Progress in biophysics and molecular biology* **2004**, *85*, 387–405.

(372) Sun, B.; Stewart, B. D.; Kucharski, A. N.; Kekenes-Huskey, P. M. *Journal of Chemical Theory and Computation* **2019**, *15*, 2692–2705.

(373) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B. *Structure* **2003**, *11*, 1453–1459.

(374) Peti, W.; Page, R. *Protein Expression and Purification* **2007**, *51*, 1–10.

(375) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins: Structure, Function, and Bioinformatics* **2000**, *41*, 415–427.

(376) Holehouse, A. S.; Das, R. K.; Ahad, J. N.; Richardson, M. O.; Pappu, R. V. *Biophysical Journal* **2017**, *112*, 16–21.

(377) Fawcett, W. R. *The Journal of Physical Chemistry B* **1999**, *103*, 11181–11185.

(378) Nguyen, C. N.; Gilson, M. K.; Kurtzman Young, T. *arXiv.org* **2011**, *arXiv:1108.4876*.

(379) Nguyen, P. T. M.; Nguyen, V. T.; Annapureddy, H. V. R. *Chemical physics* **2012**.

(380) Haider, K.; Cruz, A.; Ramsey, S.; Gilson, M. K.; Kurtzman, T. *Journal of Chemical Theory and Computation* **2018**, *14*, 418–425.

Vita

Bin Sun was born in Guangshui, Hubei, China. He received his bachelor degree (Biopharmacy) and master degree (Medicinal Chemistry) from Huazhong University of Science and Technology, Wuhan, China. After that he joined Dr. Peter M. Kekenes-Huskey's group at University of Kentucky in August, 2015 for Ph.D study.

Publications

- Bin Sun, Darin Vaughan, Svetlana Tikunova, Trevor P. Creamer, Jonathan P. Davis and Peter M Kekenes-Huskey. Calmodulin-Calcineurin Interaction beyond the Calmodulin-Binding Region Contributes to Calcineurin Activation, **2019**, *Biochemistry* 58, 4070–4085.

- Bin Sun, Bradley D Stewart, Amir N Kucharski and Peter M Kekenes-Huskey. Thermodynamics of cation binding to the sarco/endoplasmic reticulum calcium ATPase pump and impacts on enzyme function, **2019**, *Journal of Chemical Theory and Computation* 15, 2692–2705.

- Bin Sun, Erik C Cook, Trevor P Creamer and Peter M Kekenes-Huskey. Electrostatic control of calcineurin's intrinsically-disordered regulatory domain binding to calmodulin, **2018**, *Biochimica et Biophysica Acta (BBA)-General Subjects* 12, 2651-2659.

- Bin Sun, Ryan Blood, Selcuk Atalay, Dylan Colli, Stephen E. Rankin, Barbara L. Knutson, Peter M Kekenes-Huskey. Simulation-based characterization of electrolyte and small molecule diffusion in oriented mesoporous silica thin films. **2017** *book chapter of Computational Materials, Chemistry, and Biochemistry: From Bold Initiatives to the Last Mile*, Springer International Publishing.

- Jinfeng Huang, Sun Bin, Yuan Yao and Junjun Liu. Fast and Reliable Thermodynamic Approach for Determining the Protonation State of the Asp Dyad, **2017**, *J. Chem. Inf. Model* 57, 2273-2280.

- Heng-Yi Yu, Zu-Yu Chen, Bin Sun, Junjun Liu, Fan-Yu Meng, Ye Liu, Tian Tian, An Jin, Han-Li Ruan. Lignans from the Fruit of Schisandra glaucescens with Antioxidant and Neuroprotective Properties, **2014**. *Journal of natural products 77, 1311-1320*

- Jinfeng Huang, Yali Zhu, Bin Sun, Yuan Yao, Junjun Liu, Determination of the protonation state of the Asp dyad: conventional molecular dynamics versus thermodynamic integration, **2016**. *Journal of molecular modeling 22, 58*

- Chunmei Chen, Jianping Wang, Junjun Liu, Hucheng Zhu, Bin Sun, Jing Wang, Jinwen Zhang, Zengwei Luo, Guangmin Yao, Yongbo Xue, Yonghui Zhang, Armochaetoglobins A-J: Cytochalasan Alkaloids from Chaetomium globosum TW1-1, a Fungus Derived from the Terrestrial Arthropod Armadillidium vulgare **2015**, *Journal of natural products 78, 1193-1201*

- Ying Tang, Yongbo Xue, Guang Du, Jianping Wang, Junjun Liu, Bin Sun, Xiao-Nian Li, Guangmin Yao, Zengwei Luo, Yonghui Zhang, Structural revisions of a class of natural products: scaffolds of aglycon analogues of fusicoccins and cotylenins isolated from fungi **2016**. *Angewandte Chemie International Edition, 55, 4069-4073*

- Hucheng Zhu, Chunmei Chen, Junjun Liu, Bin Sun, Guangzheng Wei, Yan Li, Jinwen Zhang, Guangmin Yao, Zengwei Luo, Yongbo Xue, Yonghui Zhang, Hyperascyrones A–H, polyprenylated spirocyclic acylphloroglucinol derivatives from Hypericum ascyron Linn, **2015**. *Phytochemistry 115, 222-230*

- Guanqun Zhan, Junjun Liu, Junfei Zhou, Bin Sun, Haji Akber Aisa, Guangmin Yao, Amaryllidaceae alkaloids with new framework types from Zephyranthes candida as potent acetylcholinesterase inhibitors **2017**. *European journal of medicinal chemistry 127, 771-780*

- Hucheng Zhu, Chunmei Chen, Jing Yang, Xiao-Nian Li, Junjun Liu, Bin Sun, Sheng-Xiong Huang, Dongyan Li, Guangmin Yao, Zengwei Luo, Yan Li, Jinwen Zhang, Yongbo Xue, Yonghui Zhang, Bioactive acylphloroglucinols with adamantyl skeleton from Hypericum sampsonii **2014**. *Organic Letters*