University of Texas at El Paso

ScholarWorks@UTEP

2023-05-01

Developing And Applying Computational Methods On Biomolecules

Shengjie Sun
University of Texas at El Paso

DEVELOPING AND APPLYING COMPUTATIONAL METHODS ON BIOMOLECULES

SHENGJIE SUN

Doctoral Program in Computational Science

APPROVED:

Lin Li, Ph.D., Chair

Jianjun Sun, Ph.D.

Jorge A. Munoz, Ph.D.

Georgialina Rodriguez, Ph.D.

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

To

all the stubborn dreamers out there,

may you find

the courage and resilience

to pursue your aspirations.

DEVELOPING AND APPLYING COMPUTATIONAL METHODS ON BIOMOLECULES

by

SHENGJIE SUN, MS.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Computational Science Program

THE UNIVERSITY OF TEXAS AT EL PASO

May 2023

ABSTRACT

Computational biophysics is an interdisciplinary subject that uses numerical algorithms to study the physical principles underlying biological phenomena and processes. Electrostatic interactions play an important role in computational molecular biophysics and their potential impact on disease mechanisms. At distances larger than several Angstroms, electrostatic interactions dominate all other forces, while the alteration of short-range electrostatic pairwise interactions can also have significant effects. The dual nature of electrostatic interactions, being dominant at long-range and specific at short-range, underscores their profound implications for wild-type structure and function. Any disruption of the complex electrostatic network of interactions may abolish wild-type functionality and could be the dominant factor contributing to pathogenicity.

During my doctoral research, I utilized a diverse range of multi-scale computational techniques, such as structural modeling, molecular dynamics simulations, and electrostatic analysis, to investigate various biological phenomena. My research involved studying the kinetic cycle of myosin and the activation mechanism of Janus Kinases (JAKs). Additionally, I built upon my prior experience developing the Hybridizing Ions Treatment (HIT) method by implementing machine learning to create an improved and more user-friendly version of HIT, called HIT-2.

Sudden cardiac death is responsible for half of all deaths related to cardiovascular diseases. Understanding the mechanism of the kinetic cycle of cardiac myosin is crucial for developing protective measures and drugs for the heart. The change in state from rigor to post-rigor is key to explaining the binding and dissociation of myosin. Using β-cardiac myosin in rigor and post-rigor states, I modeled actomyosin complexes and found that there are fewer interactions and lower electrostatic binding strength in the post-rigor state compared to the rigor state. In the post-rigor

state, there were higher levels of free binding energy, fewer salt bridges, and fewer hydrogen bonds, resulting in a lower binding affinity compared to the rigor state. This decrease in binding affinity creates important conditions for the dissociation of myosin from the actin filament. The findings of this study on the mechanism of the myosin kinetic cycle provide a novel direction for future research on genetic diseases.

The family of Janus Kinases associated the JAK-STAT signaling pathway plays a vital role in the regulation of various cellular processes. A model of the inhibited full-length JAK1 and the energies of JAK1 with Tyrosine Kinase domain (TK) at different positions were calculated. Also, dynamic programming was applied to find the energetically smoothest path. Through a comparison of the energetically smoothest paths of the different mutations, the reasons why these mutations lead to negative or positive regulation of JAK1 activation are illustrated. Besides, the effects of phosphorylation of tyrosine in TK on activation process and ATP hydrolysis are also investigated and studied.

Hybridizing Ions Treatment-2 (HIT-2) is used to model biomolecule-bound ions using the implicit solvation model. By modeling ions, HIT-2 allows the user to calculate important electrostatic features of the biomolecules. HIT-2 applies an efficient algorithm to calculate the position of bound ions from molecular dynamics simulations. Modeling parameters were optimized by machine learning methods from thousands of datasets. The optimized parameters produced results with errors lower than 0.2 Å. The testing results in molecular dynamics simulations also proved that HIT-2 can effectively identify bound ion types, numbers, and positions. HIT-2 can significantly improve electrostatic calculations for many problems in computational biophysics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: ELECTROSTATICS IN COMPUTATIONAL BIOPHYSICS

1.1 Introduction

In this chapter, we consider the role of electrostatics on the structure and function of biological macromolecules and summarize the specific features that manifest electrostatic interactions. We refer to electrostatic interactions that result from charged atoms, including the atoms of water molecules. This allows us to attribute these specific features and their disruption caused by missense mutations to molecular mechanisms implicated in diseases. We focused on computational findings; however, when possible, these findings are backed up by experimental observations. The review emphasizes recent investigations to assess current interest in this field.

The electrostatic features in biological molecules are described in this review within three functions: long-range effects [1, 2]; short-range effects[3-5]; and pH-dependent effects[6-8]. The long-range electrostatic force is the main force making biomolecules travel and bind to other biomolecules over a long distance[1, 9]. Due to the limitations of current computing, it is challenging to simulate long-range electrostatics in molecular dynamic simulations. Therefore, many efforts have been proposed and developed for simulations involving long-range electrostatic interactions[2, 9]. Long-range effects are introduced in Section 2.1. Such algorithms of long-range electrostatics are widely used in biological problems, such as protein–protein interactions [10-12], protein–DNA/RNA interactions[13], etc. Short-range interactions, including salt bridges and hydrogen bonds, which are favorable interactions in terms of Coulombic energy, are present inside/among the macromolecules, serving as an important feature contributing to macromolecular architecture. Short-range effects frequently contribute to receptor–ligand binding [14, 15] and to the specificity of the binding mode[16, 17]. The pH-optimum is the particular pH at which biological activity, macromolecular stability, and binding are best optimized[18]. It originates

1

from the pKa values of the ionizable groups, indicating the importance of assigning correct protonation states in molecular dynamics simulations [19-23]. Recent contributions to modeling pH-dependent phenomena are shown in Section 2.3. Constant pH MD (CpHMD) approaches are an efficient tool to explore the effects of pH on biomolecular stability[24] and binding[25].

Electrostatics plays a crucial role in biomolecular stability and binding. Thus, mutations causing deviation of electrostatic properties may influence macromolecular thermodynamics, resulting in dysfunctional biomolecules. Mutations can either delete or reverse a charge, and these changes are expected to cause serious effects on both long- and short-range interactions[26, 27]. Recent works on salt-bridge disruption and hydrogen-bond disruption[28, 29] are shown in chapter 1.3. The last section is about drug discovery, and we focus on computational works that indicate the role of salt bridges and hydrogen bonds[30-33] to identify efficient drugs. This includes inhibitor screening methods for hepatitis C virus[34, 35], insulin amyloid fibril[36], and antibiotics against SARS-CoV-2[37, 38]. The goal of this review is to highlight the role of electrostatics in the field of computational biophysics and related disease mechanisms. Electrostatics is related to various molecular mechanisms for wild-type functionality and disease causality [27, 39]. Electrostatic features are the key component in drug de-sign according to short- and long-range effects. Another prominent representation of the role of electrostatics is pH-dependence, which typically originates from both long- and short-range electrostatic interactions. Overall, many biological processes at the molecular level are strongly affected by electrostatics, which should be seriously considered in any atomistic-level studies.

1.2 Electrostatics of wild-type biological macromolecules

The role of electrostatics in wild-type biological macromolecules is considered on three levels: (a) long-range effects, i.e. interactions that do not involve physical contact between

interacting entities; (b) short-range effects, i.e. interaction involving physical contacts as the formation of salt-bridge or hydrogen bond; and (c) pH-dependent phenomena, which involve a mixture of long- and short-range effects (Fig. 1.1)
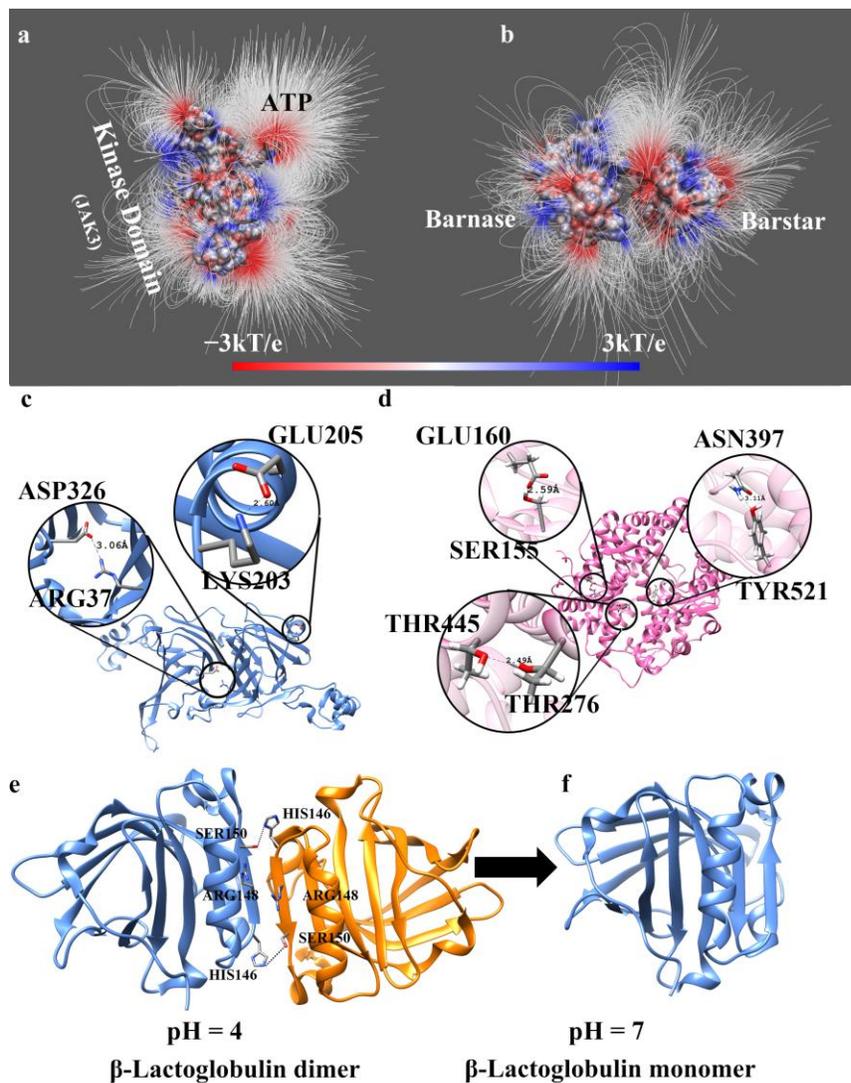
Figure 1.1:   Electrostatic features of wild-type biological macromolecules. a. The electric field lines between JAK3 kinase domain and ATP. The kinase domain and ATP were separated by 15 Å to show the interactions[40] (The structure is constructed by I-TASSER[41] to model the missing loops) b. The electric field lines between barnase and barstar (PDB: 1BRS[42]). The two proteins were separated by 10 Å to show the interactions. c. The salt bridges in 8S α-Globulin (PDB: 2CV6[43]). The missing loops of the 8S α-Globulin are modeled by SWISS-MODEL[44]. The salt bridge pair of ASP326-ARG37 is buried inside the protein while the GLU205-LYS203 is on the surface of the protein. d. The hydrogen bonds in Angiotensin-Converting Enzyme 2 (ACE2)(PDB: 6LZG[45]). The hydrogen bonds of THR445-THR276 and ASN397-TYR421 are inside of the protein while the GLU160-SER155 are surface-exposed. e. The β-Lactoglobulin dimer formed in pH 4. f. The β-Lactoglobulin monomer formed in pH 7 (PDB:6FXB[46]).

1.2.1 Long-range electrostatic effects

The electrostatic force is a long-range force that dominates all other forces when there is no physical contact between the molecules. It is also a universal force in molecular biology since all atoms carry partial charge and thus are subject to electrostatic interactions. Because of that, electrostatics is expected to play a key role in steering interacting partners toward their binding position (Fig. 1.1a, b - electric field lines demonstrate long-range attraction between the partners). However, computational modeling of such a process, especially if the interacting partners are far apart from each other, is quite computationally demanding. Indeed, in typical molecular dynamics (MD) simulations one applies cut-offs of long-range interactions in order to speed up the calculations[1]. This is valid also for studies aiming at the protein folding process, particularly the modeling of unfolded state. Recent work demonstrated that the unfolded state's structural characteristics are more strongly dependent on the cut-offs and Evald sum implementation (a method for computing long-range interactions)[47].

To overcome such limitations, a hybrid method involving DelPhi[48-51] (a popular Poisson-Boltzmann equation solver) and steered MD was developed. The method, termed DelPhi-Force MD (DFMD)[10, 52], was successfully applied to dock substrates to an enzyme[2] and a

biomolecule onto another biomolecule (such as barstar onto barnase)[10]. It was demonstrated that such an approach allows one to take advantage of the MD protocol to sample different conformations while simultaneously adding electrostatic force based guidance of the ligand toward the correct binding pocket/interface[2]. Moreover, it was shown that the DFMD can successfully dock to barnase even if the initial positions and orientations of both are completely different from the correct ones[10].

Of particular interest are cases involving highly charged large biological objects as microtubules. Microtubules are highly negatively charged long objects that serve as highways for various transport (cargo) proteins such as kinesin and dynein. The cargo proteins walk on microtubules, and one could expect that their walking is influenced by the long-range electrostatic interactions. Indeed, computationally it was shown that there are strong electrostatic interactions between microtubule and microtubule binding domains of both kinesin[53] and dynein[54]. The same observation was made for G-actin and myosin (Fig. 1.2c). These computational findings were experimentally confirmed in a study of a series of mutants at the dynein Microtube Binding Domain (MTBD)-microtubule-binding interface to neutral residues. It was discovered that the altered MTBDs' binding affinity for microtubules has significantly increased. Furthermore, it was discovered that the binding and unbinding rates of MTBDs to microtubules were significantly impacted by charge screening by free ions in solution. These findings show the importance of long-range electrostatic interactions in controlling the dynein-microtubule affinity[55].

The role of electrostatics in providing guiding force for macromolecular binding was demonstrated in a series of examples, where one uses the experimental structure of the complex and calculates the electrostatic force generated by one of the partners on the other one. The examples included homodimer, calmodulin, protein-DNA/RNA complexes and quinone in the

reaction center protein[56]. Furthermore, the electrostatic interactions were found to be a crucial factor for viruses' assemblies (Fig. 1.2a). Many studies have proved that the electrostatic interactions among capsomers of the virus provide attractive forces for the viral capsid assembly[57, 58].
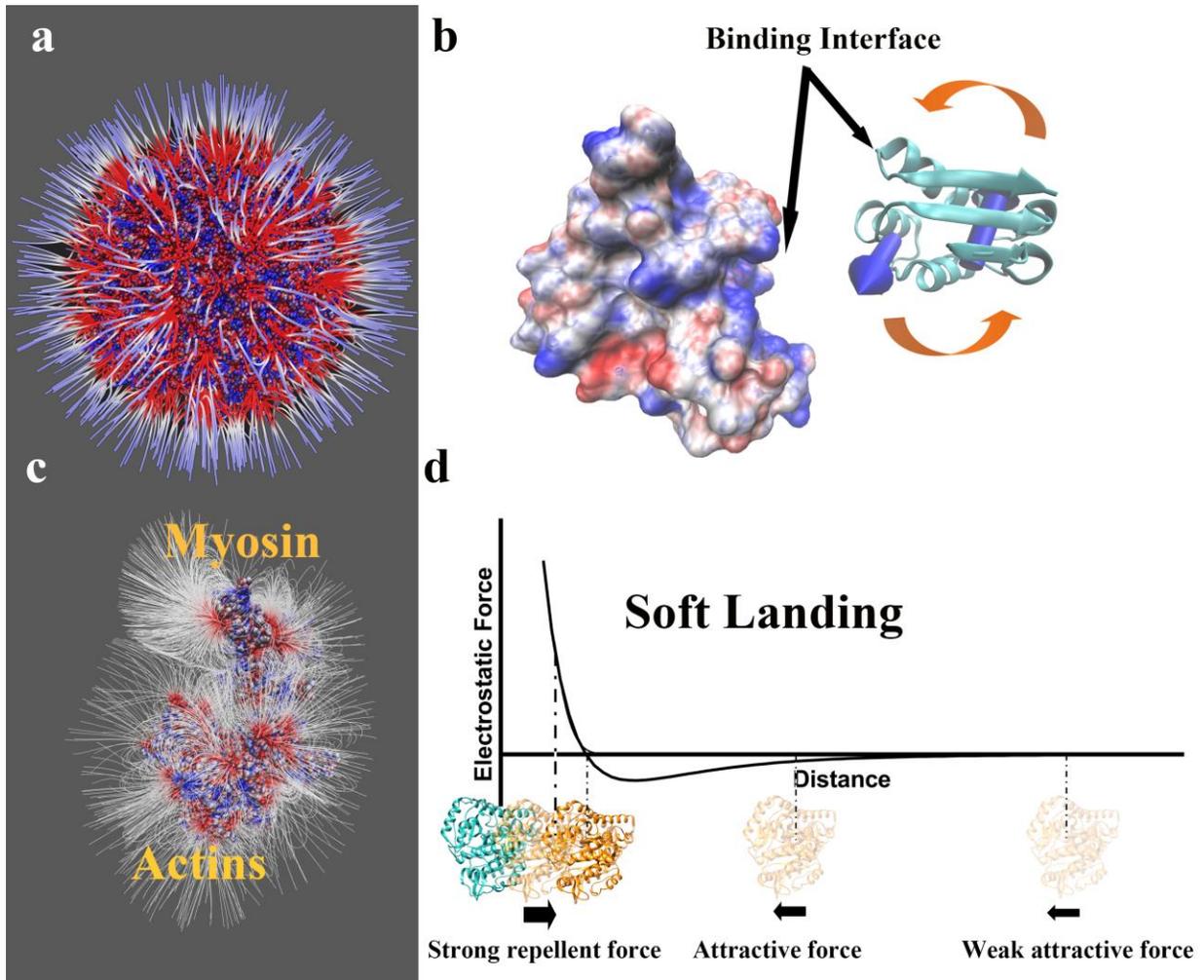
Figure 1.2: Long-range electrostatic effects. a. Electrostatic features surrounding a viral capsid (PDB ID: 5J36[59]). The electrostatic potential and electric field lines on the surface of the viral capsid illustrate the electrostatic interactions among the capsomer proteins. b. The torque in barnase-barstar binding, where the barstar was separated by 20 Å and rotated by 90° to show the torque on the barstar. c. The electric field lines on myosin-actin complex, where the myosin and actin were separated by 15 Å to show the electric field lines. d. The relationship between the electrostatic force and distance in protein-protein interactions. The blue one is alpha-tubulin while the orange one is beta-tubulin.

6

The electrostatic force not only guides the binding partners toward their association, but also assures that their interfaces are properly oriented[52] (Fig. 1.2b). This is demonstrated for cases where the binding partners are deliberately improperly oriented prior to simulations, such that their interfaces are not aligned. The electrostatic force generates torque that reorients the partners in a proper orientation. This was demonstrated in a series of examples including barnase-barstar[10], dynein-microtubule[12], and capsomer-capsomer interactions in viral capsid assembly[58].

As mentioned above, the long-range electrostatic interactions provide a guiding force that drives the partners together, however, this force is not uniform and changes as a function of the distance between the partners. It was demonstrated that the electrostatic force profiles (force as a function of distance) can be grouped into four distinctive categories[11]. In some cases, electrostatics favors the binding, in others it opposes it. The most intriguing case is so termed the "soft landing" case such that at large distances electrostatic force attracts the partners, but just prior physical contact occurs, the electrostatics opposes the binding and just assures soft binding[11] (Fig. 1.2 d). Soft landing may help two binding partners to contact each other in sufficient time, so that they can adjust the configurations at the binding interfaces to avoid atom-atom clash and form hydrogen bonds and salt bridges.

Of particular interest are cases involving intrinsically disordered proteins (IDP). Thus, computational and experimental analysis of the cell-cycle regulator p27 was demonstrated that long-range electrostatic forces acting on enriched charges of IDPs could speed up protein-protein encounter through "electrostatic steering" and simultaneously time, promote "folding-competent" encounter topologies to increase the efficiency of IDP folding upon encounter[60]. Another study

done on three IDPs, forming complexes also outlined the importance of electrostatics. They carried topology-based coarse-grained simulations and projected identical electrostatically accelerated encounter and folding mechanisms for all three complexes. These findings were consistent with earlier research of the charge distributions in known IDP complexes, which indicates that electrostatic interactions play a significant role in facilitating effective coupled binding and folding for quick specific identification[61]. Similarly, a computational alanine scanning has been performed for SARS-CoV-2 Spike and Human ACE2 Proteins. The residues identified with significance in binding and other residues in proximity were studied further through molecular mechanics-based protein binding free energy change prediction methods[62]

1.2.2 Short-range electrostatic effects

While at large distances between interacting atoms the electrostatics manifest its role via direct charge-charge interactions, in the case of short distance when there is contact between interacting partners, the electrostatics contribution is more complex. It still involves direct charge-charge interactions, which could be very strong at Angstrom scale distances, however, another important factor is solvation energy (strictly speaking the short-range effects are results from the electrostatic interactions, orbital relaxation and Pauli-repulsion). In terminology of continuum electrostatics, the balance is maintained by Coulombic interactions and desolvation penalty. Thus, a salt-bridge or a hydrogen bond, which are favorable interactions in terms of Coulombic energy, may not be contributing to the macromolecular stability or binding, due to overcompensation coming from desolvation penalty (Fig. 1.1c-d: the salt bridge pair of ASP326-ARG37 is buried inside the protein while the GLU205-LYS203 is on the surface of the protein. The hydrogen bonds of THR445-THR276 and ASN397-TYR421 are inside of the protein while the GLU160-SER155 are surface-exposed). However, such unfavorable salt-bridges and hydrogen bonds may still be

present in the macromolecular universe and will be an important structural feature contributing to macromolecular architecture. Such short-range interactions are frequently found to contribute to receptor-ligand binding, and the specificity of the binding mode. Below we outline such short-range electrostatic interactions for specific cases found computationally and if possible, backed up with experimental observations.

Polarization of interacting atoms is particularly important when the interactions are short-range interactions. Thus, the impact of electronic polarization on homogeneous and heterogeneous amyloid oligomers using the Drude Oscillator model (a model considering each atom to be a point particle with partial a charge during the simulation) was recently investigated. The system of interest was made of two amyloids i.e., amyloid-β (Aβ) and islet amyloid polypeptide (IAPP), which are involved in the pathology of Alzheimer's disease and Type 2 Diabetes Mellitus. It was found that the most stable system was homogenous A-16-22, which gained stability from salt bridge formation and reduced polarization in hydrophobic residues[62]. Similarly, another study investigated β-strand rich oligomers with Drude force field and found that structural rearrangement occurred causing some loss of β-strand structure in favor of random coil content for all oligomers. It was outlined that low polarization in hydrophobic residues and salt bridge formation contribute to the stability of Homogenous Aβ16−22[16].

Short-range interactions are a key factor that contributes to macromolecular stability and thus received significant interest. The interplay between favorable direct interactions and desolvation penalty. Thus, a recent study reported an investigation of the role of two salt bridges in ubiquitin: surface-exposed salt bridge (SB1:K11-E34) and buried salt bridge (SB2:K27-D52). It was shown that if the SB1 is broken, the mechanical stability of ubiquitin increases slightly, but when the SB2 is broken, the stability reduces significantly[63]. Stability and activity are

typically related and thus a study on a particular G-protein coupled receptor (GPCR), the cannabinoid receptor 1 (CB1), indicated that the switching mechanisms involves a salt bridge, the D263-K328 salt-bridge, that contributes to the stability of CB1[64]. Salt bridge involvement in the activation mechanisms of calcium dependent protein kinase-1 further elucidates the role of short-range interactions in molecular biology. Recent work focused on calcium dependent protein kinase-1 of Plasmodium falciparum (PfCDPK1) and explored the possibility of allosteric inhibition of this kinase. It is shown how on truncation of CAD in PfCDPK1, a conserved salt-bridge required for stabilizing the kinase domain in active state is disrupted and the kinase domain adopts inactive conformation[64]. Similarly, the role of a salt bridge, the His59-Asp103, in human granzyme B (hGzmB) was studied. It was shown that Asp103-Arg216 forms a salt bridge upon activation, breaking the His59-Asp103 hydrogen bond and enlarging the active site to aid in substrate binding[65]. Furthermore, short-range interactions are implicated in protein-protein binding and mutations may be disease-related. This was investigated for SARS-CoV-2's infectivity to figure out if the infectivity is enhanced by naturally occurring mutations. The structural dynamics of the RBD-spike protein mutation enhancing ACE2-binding were computed in silico to achieve this. Due to better interfacial stability of the RBD -strand surrounding the ACE2 across salt bridge hotspots, the ones in the RBD region showed a stronger affinity to ACE2[14]. A related study investigated the mutations (S477N-E484K) in the receptor-binding domain (RBD) of spike protein. The binding affinity of ACE2-RBD was examined using a protein docking and all-atom simulations technique. According to the investigation, the mutant modifies the hydrogen bonding network and binds more strongly than the wild type[14].

Short-range interactions are essential for gating and transport, and frequently are involved in selectivity. This was demonstrated in the case of the potassium channel, where it was observed

that S42 mutations in the pore helix significantly slow down the shutting of this filter gate, an effect that is unrelated to the amino acid's creation of a hydrogen bond at this location[66]. The role of salt bridges in conformational changes needed for the mitochondrial ADP/ATP carrier (AAC), which alternates between cytosol-open (c-) and matrix-open (m-) states to export ATP and import ADP was investigated via molecular dynamics simulations. It was shown that short-range interactions are critical for the functioning of the carrier[67].

The hydrogen bonds are frequently involved in catalysis and in overall function of biological macromolecules. Thus, it was demonstrated that wild type hydrogen bond network is crucial for the activity alpha/beta hydrolase domain-containing 5 (ABHD5), also known as CGI-58, which is the activator of adipose triglyceride lipase (ATGL). Indeed, computational modeling showed that the mutations of E41, R116, and G328 disrupt the hydrogen bonding network with surrounding residues and inhibit membrane targeting or ATGL activation[68]. Another study focused on the T cell receptor's interaction with the peptide major histocompatibility complex and showed that hydrogen bonds and Lennard-Jones contacts, which are physicochemical aspects of the TCR-pMHC dynamic bond strength, are correlated with the immunogenic response brought on by the particular peptide in the MHC groove[69]. Hydrogen bonds are essential in catalytic reactions involving proton transfer. This was demonstrated for the breakdown of uracil, where the nucleoside triphosphate cyclohydrolase (UrcA) catalyzes the two-step hydrolysis of uridine triphosphate (UTP). MD simulations showed that the hydrogen bond interaction helps the reaction intermediate undergo spontaneous conformation overturn in the active site of UrcA[70].

Perhaps the hydrogen bonds are the most important for water molecule arrangement. Since the physiological environment of all biological macromolecules is the water phase, the properties of the water medium affect their stability and functionality. This inspired investigations on

understanding short-range interactions between neighboring water molecules. Since the interactions occur at very short distances, one needs to apply quantum mechanical approaches. Thus, recent work focused on the implications of nuclear quantum fluctuations on equilibrium and dynamical properties relating to bifurcation routes in hydrogen-bond dimers of water and ammonia. It was shown that the classical over-the-hill approach is substituted with a tunneling-controlled mechanism that, from the perspective of the path integral, can be modeled as coordinated inter-basin migrations of polymer beads[71]. The application of quantum mechanical approaches was further extended to probe the development of neighboring hydrogen bonds with water molecules. It was demonstrated that there is a direct mechanism for the emergence of short-range structural fluctuations in the hydrogen bond network of liquid water, which shows that the time development of neighboring hydrogen bonds is closely connected[72, 73].

Hydrogen bonds and salt bridges are important for receptor-ligand interactions, including drug-like molecules binding to the corresponding targets. Recent work explored this topic and applied MD simulations with MMPBSA approaches to study the stability of histone deacetylase inhibitors' effects on the stability of histone deacetylase like proteins (HDLP). The result showed that the stability HDLP-CBHA (m-Carboxycinnamic acid bis-hydroxamide (CBHA) is higher than the free HDLP enzyme. The higher stability was contributed by the increased number of hydrogen bonds[15]. Similarly, the computational modeling of a prospective drug, the dorzagliatin, interactions with human glucokinase activator (GKA). The results showed that dorzagliatin can create the characteristic hydrogen bonds of GKA with Arg63, making a tight binding hydrogen bond network around dorzagliatin[74]. Even in the case of DNA, a computational study indicated that the hydrogen bonds, being the main interaction constraining the variability of the linkers, are enhanced slightly with DNA twist number[75]. Another study carried out MD simulations to study

riboswitches and found that the binding domain in riboswitch is stable for molecule recognition and binding and the switching base pairings are co-evolved in the translation state[76]. Electrostatic interactions were shown to be a crucial factor for reversing the dysfunctional p53 and thus avoiding tumor progression. The N-terminal transactivation domain (TAD) of p53 can regulate cell apoptosis by interacting with the transcriptional adaptor zinc-binding 2 (Taz2) domain of p300. It was demonstrated that electrostatic interactions govern the affinity of the p300 Taz2–p53 TAD2 complex[17].

1.2.3 pKa calculations and pH-dependent phenomena (long- and short-range effects)

Almost all biological processes are pH-dependent, which indicates the crucial role of hydrogen ion concentration for the living systems[77, 78]. This is manifested as pH-dependence of catalysis, stability, binding affinity, and conformational flexibility[78, 79]. Typically there is a particular pH at which the biological activity, macromolecular stability and binding are best optimized, termed pH-optimum[18]. The pH-dependence originates from the pKa's of the ionizable groups, both ionizable amino acids and nucleic acids, and thus indicates the importance of assigning correct protonation states prior to any modeling or introducing pH in molecular dynamics simulations (constant pH MD[80-84]). Below we outline recent contributions on modeling pH-dependent phenomena. It is well understood that pH and conformational changes are coupled and in recent time many researchers apply constant pH MD (CpHMD) approaches to model the effect of pH on protein stability[85], binding[86], dynamics[87] and reactivity[88]. Additional complexity, if one investigates membrane proteins,   comes from the presence of lipid bilayer[89] . Thus, the pKa's of ionizable groups may undergo pKa shift as proteins insert into the membrane [90]. In parallel, significant amount of work was completed using either pre-defined conformational space[91]; or mimicking the conformational stability via Gaussian-based smooth

dielectric function[92] ; Furthermore, there are methods that use coarse-grained lattice-based model of proteins and train the model on existing experimental data[93], and the treecode-accelerated boundary integral solver[94].

In terms of long- and short-range electrostatic interactions discussed above, the pH-dependent phenomena originate from a mixture of both. Thus, a recent computational study of a particular protein (Apolipophorin-III) which associates with a lipid disk demonstrated that the association is strongly pH-dependent, but there are no direct interactions between titratable group of Apolipophorin-III and charged lipid head groups[95]. In contrast, another study on cyclic dipeptides showed that pKa of side chains of lysine increases for cyclic dipeptides compared to the linear ones and this was attributed to short-range[96]. The interplay between short- and long-range electrostatic interactions was demonstrated in study of virus assembly. It was outlined that both the stability and the binding affinity of the E protein are pH independent in pH range 6 to 10, and this is a result of a network of interactions[97]. The short- and long-range interactions are specifically important in treating water phase, either explicitly or implicitly, particularly when one deals with water channels or wires inside biological macromolecules where the water flexibility is highly reduced as compared with bulk water. Thus, water molecules in the Gramicidin A (gA) channel were investigated to probe the effect of different approaches and computational techniques. It was concluded that the MCCE and Drude analysis led to a small net dipole moment as the water molecules change orientation within the channel[98].

The pH-dependence is also pronounced in allosteric regulation. recent work reported combined experimental with computation investigations and showed that allostericity in pH-switching protein is guided by coupling throughout the protein, featuring a large network of hydrophobic interactions that work in concert with key electrostatic interactions[99].

1.3 Electrostatics, disease mechanisms and in silico drug discovery

It was outlined that electrostatics play a crucial role in macromolecular stability and binding. Therefore, any deviation of wild type electrostatic properties would have profound effects on macromolecular thermodynamics, and thus may result in a dysfunctional biological macromolecule. Furthermore, if the biological macromolecule is important for the wild type functioning of the cell, its dysfunction could cause disease. Indeed, as demonstrated in the literature, there is a linkage between the effect of missense mutations on protein stability or interactions and the propensity the mutation to be photogenic[100, 101]. Mutations that either delete a charge, or reverse a charge, are expected to cause dramatical effect on both long- and short-range interactions and thus to have high propensity to be disease-causing[102, 103]. Below we outline recent papers that model the effect of disease-causing mutations on atomic scale, including disruption of salt bridges and hydrogen bonds, and in silico design of drugs that could mitigate the disease-causing effect[28, 104].
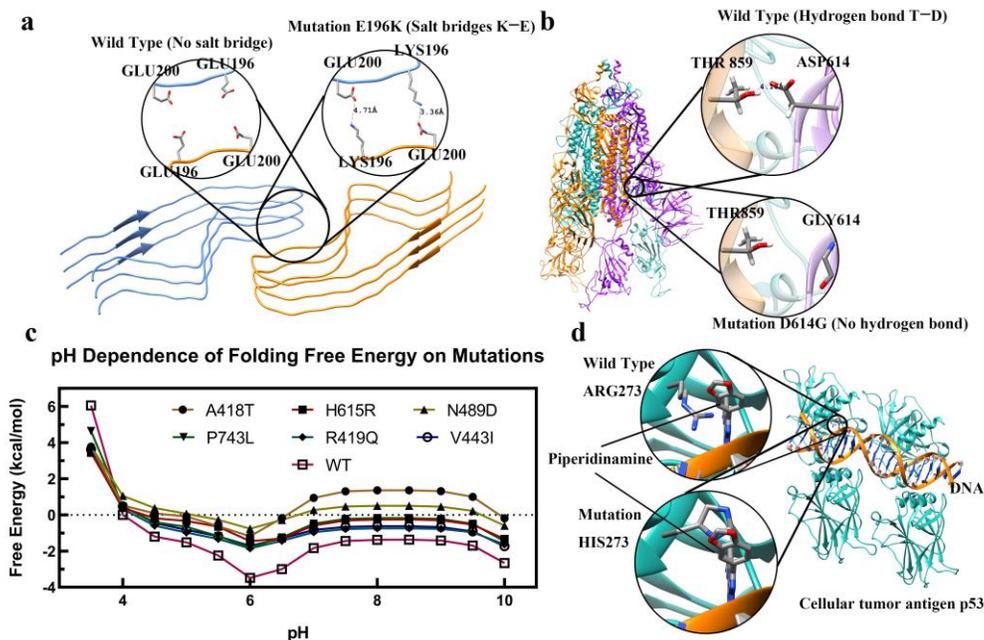
Figure 1.3: The effects of mutations on electrostatic features. a. Mutation of E196K formed a new salt-bridge between GLU and LYS in prion protein (PDB: 7DWV[105]). b. Mutation of D196G (SARS-COV-2 spike protein) caused the loss of hydrogen bond between THR and ASP (PDB: 7KMS[106]). c. The pH dependence of folding free energy for the melanosome of oculocutaneous albinism-2[8]. d. The image shows the p53 protein-DNA and 1-[2-(1,3-benzodioxol-5-yl) pyrazolo[1,5-a] pyrazin-4-yl]-3-piperidinamine. All mutations in a, b, d are achieved using Chimera[107].

1.3.1 Salt-bridge disruption

In the previous section we emphasized on the important role that salt bridges play for stability and interactions of biological macromolecules. However, it should be clarified that the expectation is that disruption of buried salt bridges should be more deleterious than of surface exposed one (unless they participate in binding a partner). Even more, due to the plasticity of biological macromolecules, even a mutation within buried salt bridge may not cause large change of the stability, because the macromolecule can rearrange and accommodate the change[102].

Because electrostatic interactions are long-range interactions, the deleterious effect may involve charged amino acids situated far apart from each other. Thus, a recent paper modeled the electrostatic component of the force acting between a kinesin motor domain and tubulin. The receiver operating characteristic method is used to show that variations in the electrostatic component of the binding force can distinguish between disease-causing and non-disease-causing mutations detected in the human kinesin motor domain. The prediction rate of 0.843 area under the ROC curve owing to a change in the amplitude of the electrostatic force alone is notable because diseases may result from a variety of causes unrelated to kinesin-microtubule binding[108]. Disruption of intermolecular interactions was implicated in another disease, the prion disease. It was proposed that pairs of amino acids from opposing subunits form four salt bridges to stabilize the zigzag interface of the two protofibrils. The results provided structural evidence of the diverse prion strains and highlighted the importance of familial mutations in

inducing different strains. Electrostatics plays an essential role in formation of other complexes such as the virulence factor ESAT-6. This is related to mycobacterium tuberculosis (Mtb), which is a leading death-causing bacterial pathogen. ESAT-6 is hypothesized to form an oligomer for membrane insertion of Mtb as well as rupturing[109].

1.3.2 Hydrogen bond disruptions

Hydrogen bonds are essential for maintaining 3D structure of biological macromolecules and formation of macromolecular complexes. The disruption of wild type hydrogen bonds by mutations may or may not be deleterious. The same considerations apply as outlined for salt bridges above. Below we outline recent works on this subject.

The binding and dissociation between myosin and actin filaments is crucial for heart protection and drug development. It is demonstrated that the loss of hydrogen bonds significantly promotes the detachment of myosin from actin filament, causing the state changes from rigor state to post-rigor state[110]. Polyglutamine tracts are the regions of low sequence complexity frequently found in transcription factors. The length of tracts is related with transcriptional activity and expansion beyond specific thresholds is the cause of polyglutamine disorders. The conformation of the polyQ tract of the androgen receptor is associated with spinobulbar muscular atrophy, depending on its length. This sequence folds into a helical structure, which is stabilized by unconventional hydrogen bonds between glutamine side chains and main chain carbonyl groups. Its helicity directly correlates with tract length. These unusual hydrogen bonds are bifurcated with the conventional hydrogen bonds stabilizing α-helices[111]. TDP-43 is an essential RNA-binding protein forming aggregates in almost all cases of sporadic amyotrophic lateral sclerosis (ALS), frontotemporal lobar dementia (FTLD) and other neurodegenerative diseases. TDP-43 aggregates have a self-templating, amyloid-like structure. This segment adopts a beta-

hairpin structure that forms an amyloid-like structure. This conformer is stabilized by a special class of hyper cooperative hydrogen bonding[112]. Moreover, Amyloid fibrils of α-synuclein are the histological hallmarks of Parkinson's disease, dementia with Lewy bodies and multiple system atrophy. The H50Q mutation results in two novel polymorphs of α-synuclein: narrow and wide fibrils, formed from either one or two protofilaments, respectively. These structures reveal new structural elements, including a hydrogen-bond network and surprising new protofilament arrangements[113].

1.3.3   pH-dependence alteration

Biological macromolecules have evolved to function at a particular pH and any deviation of wild type pH-dependent properties as stability and binding may cause diseases. Such changes could be caused by mutations involving titratable groups and less frequently by mutations causing conformational changes and thus affecting pKa's of titratable residues. Even a mutation that involves the same polarity residues, such as Arg to His, could be disease causing. Indeed, since the intracellular pH of most cancers is constitutively higher than that of normal cells and enhances proliferation and cell survival, substitution of Arg with His, dramatically changes activity of mutant proteins at high pHs[114-116].

The altered pH-dependent properties of biological macromolecules are manifested not only in cancer, but also in many other diseases. Thus, a recent work computationally studied the pH-dependent stability of several melanosome membrane proteins. It was found that disease-causing variants impact the pH dependence of melanosome proteins[117]. Substitution of wild type His with Gln was found to alter the functioning of CLIC2 protein and to cause X-linked channelopathy with cardiomegaly[118].

The Covid-19 pandemic prompted many investigations on the molecular mechanisms of disease and effect of mutations. Recent work on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), causative agent of the Covid-19 pandemic, is thought to release its RNA genome at either the cell surface or within endosomes, the balance being dependent on spike protein stability, and the complement of receptors, co-receptors, and proteases. The investigator performed pKa calculations on a set of structures for spike protein ectodomain and fragments from SARS-CoV-2 and other coronaviruses. It is predicted that a particular aspartic acid contributes to a pH-dependence of the open/closed equilibrium[119]. Another work applied multi-scale computational approaches to study the electrostatic features of spike (S) proteins for SARS-CoV and SARS-CoV-2. It was demonstrated that the complex structures of hACE2 and the S proteins of SARS-CoV/SARS-CoV-2 are stable at pH values ranging from 7.5 to 9[120]. A plausible route which variants can alter viral properties is the transition between receptor binding domain (RBD) up and down forms of the SARS-CoV-2 spike protein trimer. The work predicted that pH-dependence in the mild acidic range, with stabilization of the locked form as pH reduces from 7.5 to 5[121].

1.4 In silico drug discovery

Here we do not attempt to outline the work in the general field of drug discovery, rather we focus on investigations of molecular effects of disease-causing mutations or seeking inhibitors from natural compounds. In the last case, we emphasize on works that manifest the role of salt-bridges and hydrogen bonds to identify efficient inhibitors.

Once the effect of disease-causing mutation is revealed, it is tempting to seek a small molecule, a potential drug, which can mitigate the effect, which frequently involves electrostatic alteration. This is affordable because genotypes aggregate into several phenotypes only. Thus, a particular drug is supposed to be efficient for many genotypes within the same gene. Indeed, in

case of Rett syndrome, most of the top frequency disease-causing mutations were experimentally and computationally shown to result in lowering the affinity of MeCP2 to the cognate DNA, while not affecting protein stability[122, 123]. Similarly, many genotypes, many mutations in spermine synthase gene, which cause Snyder-Robinson syndrome, were found to affect spermine synthase homo-dimerization, and thus the electrostatic funnel that guides the substrate to the active site[32, 124-126]. This resulted in investigations that proposed drugs capable of mitigating disease-causing effect[28, 104, 127]. The most common approach of developing drugs is to design inhibitors. Such a design requires considering long- and short-range electrostatic effects. Recent study carried out residue study to investigate tyrosine kinase enzyme inhibitors. The result showed that a pair of aspartic residues, providing negative potential, plays an important role providing attractive interactions in the binding site of the enzyme[128].

Hepatitis C Virus (HCV) is affecting millions of people worldwide. Existing drugs have different efficiencies against different genotypes of HCV. This prompted a computational investigation to select the most efficient inhibitor against the most prevalent genotype in South Asia and it was shown that short-range hydrogen bonds are the most contributing factor for the binding energy[129]. Molecular docking was utilized in another study to identify inhibitors of insulin amyloid fibril formation. It was elucidated that hydrogen bonds play a crucial role in the interactions between the ligands and insulin[36]. Molecular dynamics and MMPBSA methods were also used to probe the effect of inhibitors. This was done in case of histone deacetylase-like proteins and was shown that the affinity increases with the increase of hydrogen bonds[15]. Similarly, inhibitors against cyclin G associated kinase involved in hepatitis C virus entry into host cells, were studied and short-range interactions analyzed[34]. The importance of hydrogen bonds was demonstrated in a study of amoxicillin, widely known as an antibiotic, to bind to COVID-19

protein in Mpro protease[37]. Another study focused on the observation that SARS-CoV-2 entrance to the host cells occurs via interaction between receptor binding domain (RBD) of the spike (S) protein on the virus with the angiotensin-converting enzyme 2 (ACE2) receptor. Thus, the authors took a peptide consisting of residues 19–48 of ACE2 as the wild-type peptide along with six mutants. It was shown that short-range interactions are essential for selecting the best compound[38]. Other plausible inhibitors of SARS-CoV-2 were also suggested by targeting proteins crucial for SARSCoV-2 infection and cytokine storm[130]. The role of electrostatic interactions was demonstrated in another study where inhibitors against DNA methyltransferase[131]. Short-range interactions were analyzed to identify inhibitors against acetylcholinesterase, is a key enzyme enhancing the cognitive disorder, leading to Alzheimer's disease[132]. Aminoquinolines bind to hemoglobin, and thus prevent the degradation of hemoglobin, while by binding to parasitic tissue, aminoquinolines diminish the parasitic strength of causing malaria. The role of hydrogen bonds in these association events was investigated and shown that the number of hydrogen bonds differ depending on the target[93].

1.5 Conclusion

The review outlined recent contributions emphasizing on the role of electrostatics in the field of computational molecular biophysics and disease mechanism. It was outlined that the electrostatics is still a hot topic, and it is implicated in various molecular mechanisms for wild type functionality and disease causality. Furthermore, the electrostatic considerations are frequently the key component in drug design to mitigate the disease mechanisms. While an attempt was made to group the contributions according to short- and long-range effects, due to the complex nature of electrostatic interactions, frequently both cases are present in the biological systems. A prominent manifestation of the role of electrostatics in molecular biology is the pH-dependence, which

21

typically originates from mixture of long- and short-range electrostatic interactions. Overall, the electrostatics appears to be involved in many processes at molecular level and should be considered in any study dealing with atomistic level of details.

CHAPTER 2: FORCE AND ENERGY ANALYSIS ON MOLECULAR MECHANISMS OF CARDIAC ACTOMYOSIN STATE TRANSFORMING

2.1 Introduction

Sudden cardiac death (SCD) contributes to half of all deaths from cardiovascular diseases[133]. Hypertrophic cardiomyopathy (HCM), the main cause of SCD[134, 135], is a polygenic disease[136], afflict at least one in 500 people[137]. In addition to SCD, HCM causes embolic stroke and heart failure in young and middle-aged people. To date, genetic research has located 33 HCM-related genes; MYBPC3, MYH7, TNNT2, TNNI3, TPM1, ACTC1, MYL2, and MYL3 have been categorized as definitive[138]. These definitive genes encode human cardiac sarcomeres, including myosin-binding protein C (MyBP-C), β-cardiac myosin, cardiac troponin T, troponin I, tropomyosin, cardiac actin, and myosin light chain 2 and 3. β-cardiac myosin, a type of motor protein, is assembled to form myosin filaments (thick filaments). These periodically overlap with actin filaments to form the basic unit of contraction[139]. The actin filament (thin filament) is composed of globular actin (G-actin)[140], tropomyosin (Tpm), troponin, and MyBP-C[141]. Binding between thick and thin filaments is caused by the hydrophobic interactions between the myosin heavy chains with G-actins[142, 143]. Tropomyosins, two pairs of long helix chains on the two sides of the thin filament regulate the exposure of the binding interface of G-actins[144, 145]. Movement of the Tpms is regulated by the troponin complex, a transformable protein including troponin I, troponin T, and Tpm-binding subunit[144]. The complex is regulated by the concentration of Ca^{2+} and moves Tpms into open, close, or blocked states[144, 145].

In the kinetic cycle, the instantaneous increase of Ca^{2+} concentration causes a huge conformational change in the troponin complex with azimuthal movement of the Tpm[144], exposing the binding interface of the G-actins. Then, the myosins approach and bind to the actin

filament. The binding state between the thin filaments and the myosin is called the rigor state. In this state, both Pi and MgADP are released, forming a nucleotide-free state with a strong binding. Once Pi and MgADP have been released, ATP rapidly rebinds to the myosin, causing conformational change and dissociation[146, 147], known as the recovery stroke. Meanwhile, myosin hydrolyzes ATP into ADP and Pi, but does not release them directly[148]. In this state, the pre-powerstroke (PPS), myosin is ready to undergo the powerstroke process. Release of Pi, the initial step in powerstroke, is followed by the release of MgADP, which closes the cleft of the myosin motor domain and forms a strong binding between the myosin and actin filaments. After the release of MgADP, myosin reaches the rigor state. It restarts ATP binding and hydrolysis, waiting for the next kinetic cycle[147].

The powerstroke and recovery stroke are the two fundamental processes in the kinetic cycle. The powerstroke has attracted many investigators, whose studies range from the release of Pi[149]and MgADP[150, 151] to the lever arm movement[152, 153], making the mechanism of the powerstroke much clearer than before. However, far fewer studies on the recovery stroke exist[154]. The powerstroke and recovery stroke are similar mechanisms beneath opposite phenomena. Transformation of myosin from the rigor to post-rigor state is the initial step for the recovery stroke. The necessity of the post-rigor state for the detachment of myosin needs more study. Knowledge of structural differences between the rigor and post-rigor states is the fundamental analysis for understanding the importance of the post-rigor state. The energetic and electrostatic analyses are crucial in understanding binding and dissociation.

Here we used SWISS-MODEL[155] and Chimera[107] to build human β -cardiac actomyosin complexes in rigor and post-rigor states for structural comparison. Afterward, we ran Molecular Dynamic (MD) simulations and calculated the electrostatic features and free binding

energy to explore the differences between the rigor and post-rigor states, revealing the mechanism of the dissociation. The results showed that the actomyosin complex in the post-rigor state possesses a lower binding affinity compared with that in the rigor state. The energy analysis also confirmed that post-rigor state myosin is not suitable for binding compared with that in the rigor state. These changes of actomyosin complex in the post-rigor states provide more suitable conditions for the detachment of myosin. This study shows the importance of the post-rigor state for myosin detachment and sheds light on the mechanism of dissociation of the actomyosin complex.

2.2 Method

2.2.1 Modeling and MD simulations

The model of β-cardiac myosin in rigor state was built using a template of rigor-like squid myosin S1 (PDB: 3I5G[156]) based on the sequence of β-cardiac myosin[157] through the SWISS-MODEL[155]. The sequence identity is 62.41%. The rigor myosin model was then assembled with cardiac actin filament (PDB: 5NOJ[145]) based on the actomyosin complex model of 5JLH[143] to build β-cardiac actomyosin complex in the rigor state through model alignment by Chimera[107]. The model of β-cardiac myosin in the post-rigor state was taken from the Protein Data Bank (PDB: 6FSA[157]). The missing cardiomyopathy loop and strut domain were built according to SWISS-MODEL[155]. Then the β-cardiac myosin in post-rigor state and cardiac actin filament were taken to build β-cardiac actomyosin complex in the post-rigor state, using the actomyosin complex model of 5JLH[143].

The solvation box of the TIP3P water model[158] with 150 mM of NaCl was built by Visual Molecular Dynamics[159] (VMD). The MD simulations were applied by NAMD 2.12[160]. In MD simulations, the end residue (Leu 807) of the myosin and the residues in actin

filament which were 10 Å from the myosin were restrained. The initial temperature and pH were set at 300 K and 7.0, respectively. The periodic boundary conditions and CHARMM 36[161] force field were applied to the systems. 20,000 steps of minimization were applied ahead of the 50 ns simulations. All simulations were repeated twice more.

2.2.2 Electrostatic potential calculation

After simulations, the frames at 50 ns of the actomyosin complexes in rigor and post-rigor states were taken for electrostatic potential calculations. The electrostatic potential calculations were performed using Delphi[48]. The myosin motor domain and actin filament were separated[162] by 20 Å. The charge and the radius of atoms were calculated by the force field CHARMM36 and assigned by pdb2pqr[163]. The resolution was set as 1.5 grid/Å, and the dielectric constants were set at 2.0 and 80.0 for protein and water, respectively. The probe radius for generating the protein surface was 1.4 Å, and the protein filling percentage in the calculation box was 70%. The concentration of NaCl was set at 0.15 M, and the boundary condition for the Poisson Boltzmann equation was set as the dipolar boundary condition. The calculated electrostatic potential on surfaces was visualized by Chimera[107]. VMD[159] was used to show the electrostatic field lines. In both visualizations, the color range was set from red to blue, corresponding to the potential range from -1.0 to 1.0 kT/e.

2.2.3 Electrostatic force

The electrostatic force on the myosin motor domain was calculated by DelphiForce[164] every 0.05 ns in the last 20 ns of MD simulations (30 ns to 50 ns). The parameters for electrostatic force calculation were the same as those for potential calculation. Then the electrostatic force was split into the binding component (Eq. 2.1) and the sliding component (Eq. 2.2).

$$F_{Bi} = F_i \cdot \cos \alpha_i \ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \ (2.1)$$

$$F_{Si} = F_i \cdot \sin \alpha_i \quad \text{...} \quad (2.2)$$

Where the F_{Bi} and F_{Si} are the binding component and sliding component in the ith frame, respectively. F_i is the total electrostatic force on myosin at the ith frame. The α_i is the angle between the total force and the axis, which passed the mass centers of the myosin motor domain and actin filament in the ith frame.

2.2.4 MM/PBSA free binding energy calculation

The binding energy (ΔE_{bind}) is calculated by equation 3.3, which was applied to coulombic binding energy, Van der Waals binding energy, polar and non-polar solvation binding energy.

$$\Delta E_{bind} = E_{actomyosin} - E_{myosin} - E_{actin} \quad \text{....................................} \quad (2.3)$$

where the ΔE_{bind} is the binding energy (electrostatic, Van der Waals, polar or non-polar solvation binding energy). $E_{actomyosin}$ is the corresponding energy of the actomyosin complex while E_{myosin} and E_{actin} are the corresponding energy of the myosin and the actin filament, respectively.

The free binding energy ΔE_{total_bind} was calculated by equation 2.4.

$$\Delta E_{total\_bind} = \Delta E_{cbind} + \Delta E_{vbind} + \Delta E_{pbind} + \Delta E_{npbind} \quad \text{.................} \quad (2.4)$$

ΔE_{total_bind} represents free binding energy. ΔE_{cbind}, ΔE_{vbind}, ΔE_{pbind}, and ΔE_{npbind} represent coulombic binding energy, Van der Waals binding energy, polar and non-polar solvation binding energy, respectively. The coulombic energy and polar solvation energy were calculated by Delphi[48]. The Van der Waals energy was calculated by NAMD 2.12[160]. The non-polar solvation energy was calculated by solvent-accessible surface area (SASA) by equation 2.5.

$$E_{np} = \alpha \cdot SASA + \beta \quad \text{..} \quad (2.5)$$

where E_{np} is the non-polar solvation energy, $\alpha = 0.0054$ and $\beta = 0.92$ kcal/mol. The SASA is the solvent-accessible surface area, which was calculated by NACCESS[165].

2.2.5 Salt bridges and hydrogen bonds.

The salt bridges and hydrogen bonds analyses were performed by VMD based on 1000 frames from the simulations of 30 to 50 ns. The cut-off distance for salt bridges analysis was set as 4 Å[166] while the cut-off distance and angle for hydrogen bonds analysis were set as 3.5 Å and 20° [142, 167].

2.3 Results and Discussion

2.3.1 Structural analysis

In previous works on myosin V[168], the ATP hydrolysis led to the opening of the cleft, which is called the post-rigor state. In the post-rigor state, the myosin detaches from the actin filament[154]. The conformational change in the myosin from rigor to the post-rigor state includes the movements of cardiomyopathy (CM) loop, Loop 3, and Loop 4[143, 154]. The conformational changes caused by ATP hydrolysis between the rigor and post-rigor states are shown in Fig. 2.1a. The cleft is open in the post-rigor state and closed in the rigor state. The simulations for the actomyosin complex in the rigor state reached stability after 10 ns while those in the post-rigor state reached stability after 30 ns (Fig. 2.1b). The separated RMSF (each 10 ns for a stage) were also calculated to indicate the stabilization of the structure. In both complexes, the stages 4 and 5 (Fig. 2.1cd) shows the lowest RMSF which means the structure is stabilized. It is consistent with the result of RMSD. Besides, the slight movement of post-rigor state myosin motor domain was also observed (Fig. 2.1g). It indicates the instability of the post-rigor state.

The RMSF analysis for the myosin is shown in Fig. 2.1e. The residues with high RMSF were GLN 571, THR 211, and GLN 734 in the rigor state myosin. The high RMSF of these residues is because of the formation of loops, which are very flexible (marked in Fig. 2.1h). The GLN 571 is on loop 2, which is located on the interface and possesses high flexibility.

28
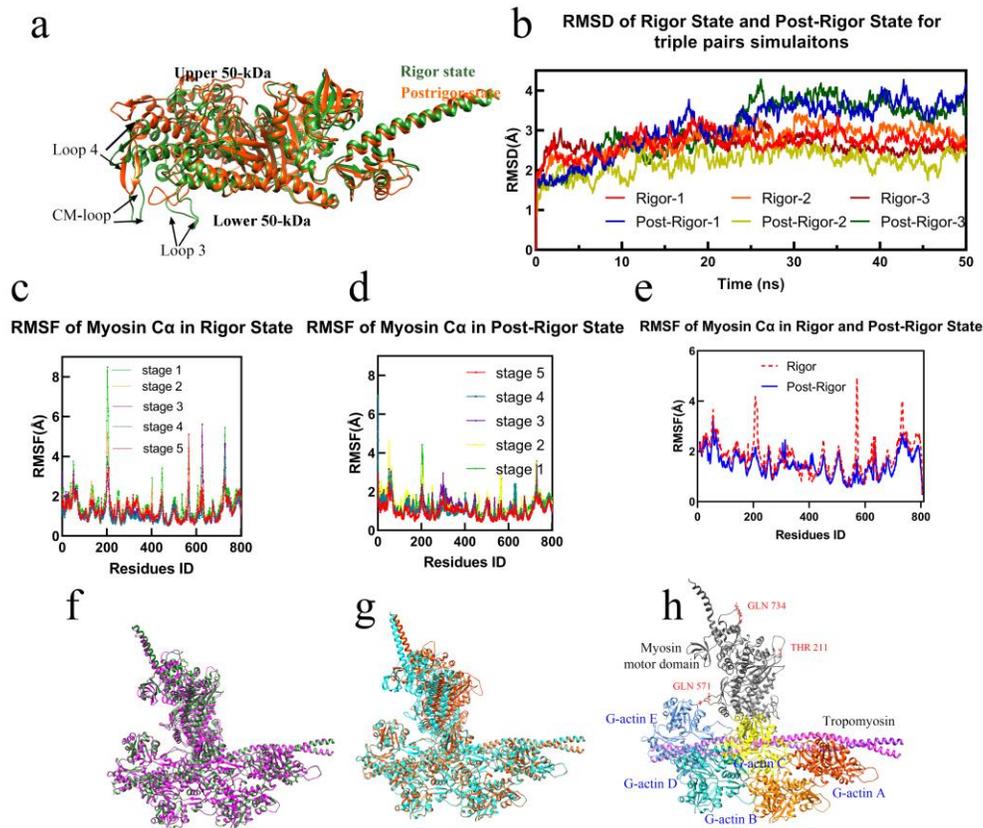
Figure 2.1: The structural analysis for both rigor and post-rigor sate in MD simulations. a. Structural comparison between the myosin motor domains in rigor (forest green) and post-rigor (orange). b. The RMSD (A) of actomyosin in rigor and post-rigor state for triple pairs of simulations. cd. RMSFs of Cα in rigor (c) and post-rigor (d) models were calculated by each 10 ns in 5 stages. e. RMSF of Cα in rigor and post-rigor models. eg. The structural comparison between the structures at the 0 ns (pink and orange) and 50 ns (green and cyan) in rigor (e) and post-rigor (g) models. h. Rigor model with marked residues with high RMSF.

2.3.2 Electrostatic potential surface calculation

We took the structures of actomyosin complexes in the rigor and post-rigor states from the last frame (50 ns) for calculations of electrostatic potential. The electrostatic potential surfaces the electrostatic field lines are shown in Fig. 2.2. The myosin and actin filament in both figures were separated by 20 Å for better visualization of the interfaces. The myosin-binding interface is positively charged while the actin-binding interface is negatively charged (Fig. 2.2abef).

29

Compared with the myosin-binding interface in the rigor stage, the positively charged area decreased in the post-rigor state (Fig. 2.2ab). By contrast, the actin-binding interface was not affected by state changes. In Fig. 2.2ef, tropomyosins are also negatively charged. They can interact with the positively charged myosin, contributing to electrostatic attraction. We further rendered the electrostatic field lines by VMD[169]. The decreased positively charged area in the myosin-binding interface directly caused fewer interactions with actin filament in the post-rigor state. The decrease in interactions is shown in Fig. 2.2dh. Consistent with the electrostatic potential surface, the tropomyosins (Tpms) also interacts with the myosin. This reflects our previous work on interactions between Tpms and myosins[170]. In detail, the electrostatic field lines between the myosin motor domain and tropomyosin are even denser than those with G-actin, especially in the post-rigor state. The electrostatic interactions between myosin and actin filaments are thought to be important contributors to the binding or movements of myosin in the kinetic cycles. The weakened interactions in the post-rigor state contributed to the dissociation of actomyosin complex.
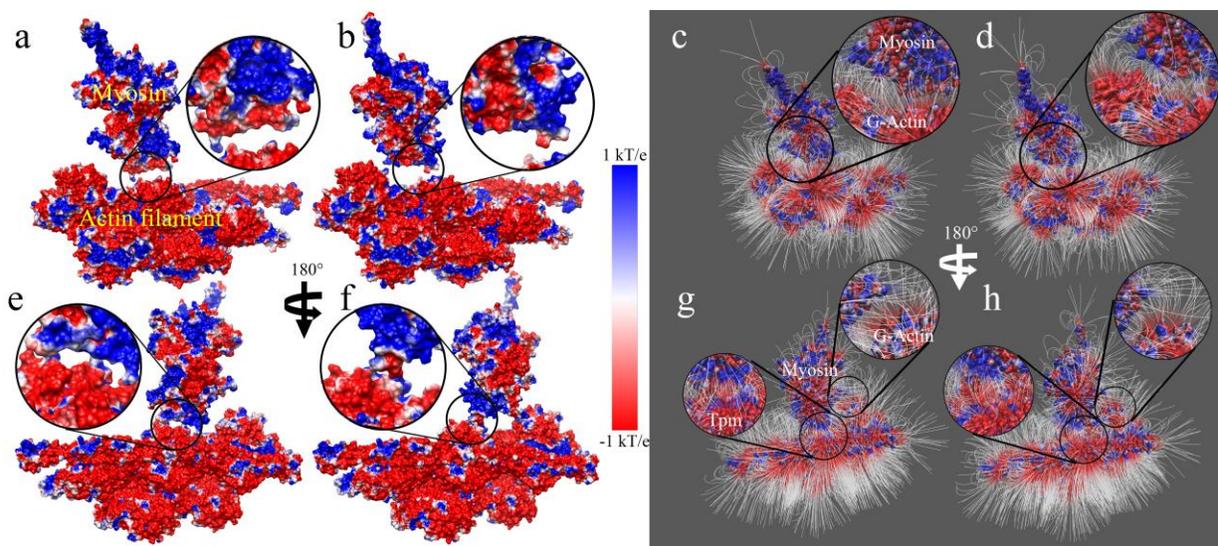
Figure 2.2: The coulombic potential surface and field lines of the actomyosin in states of rigor and post-rigor. aecg. The actomyosin complex in rigor state. bdfh. The actomyosin complex in post-rigor state.

2.3.3 Electrostatic force analysis

In the rigor state, the actin filament acts an electrostatic force of 125.11 kT/e on the myosin; the electrostatic force decreases to 73.52 kT/e in the post-rigor state (Fig. 2.3a,). In Fig. 2.3cf, the forces were analyzed as the binding and sliding components (Eq. 2.1 and 2.2). From rigor to the post-rigor state, the binding component decreases (Fig. 2.3c) (from 120.24 kT/e to 61.62kT/e) along with the decrease of the total force. By contrast, there is no obvious difference for the sliding component (Fig. 2.3f) from rigor to post-rigor state. Combined with analyses of the surfaces of electrostatic potential, the conformational changes in myosin from rigor to post-rigor state shrink the positively charged area and decrease the interactions with actin filament. The decrease of the interaction further suggested the decrease of the electrostatic binding force. The decrease in binding helps detach the myosin from the actin filament.
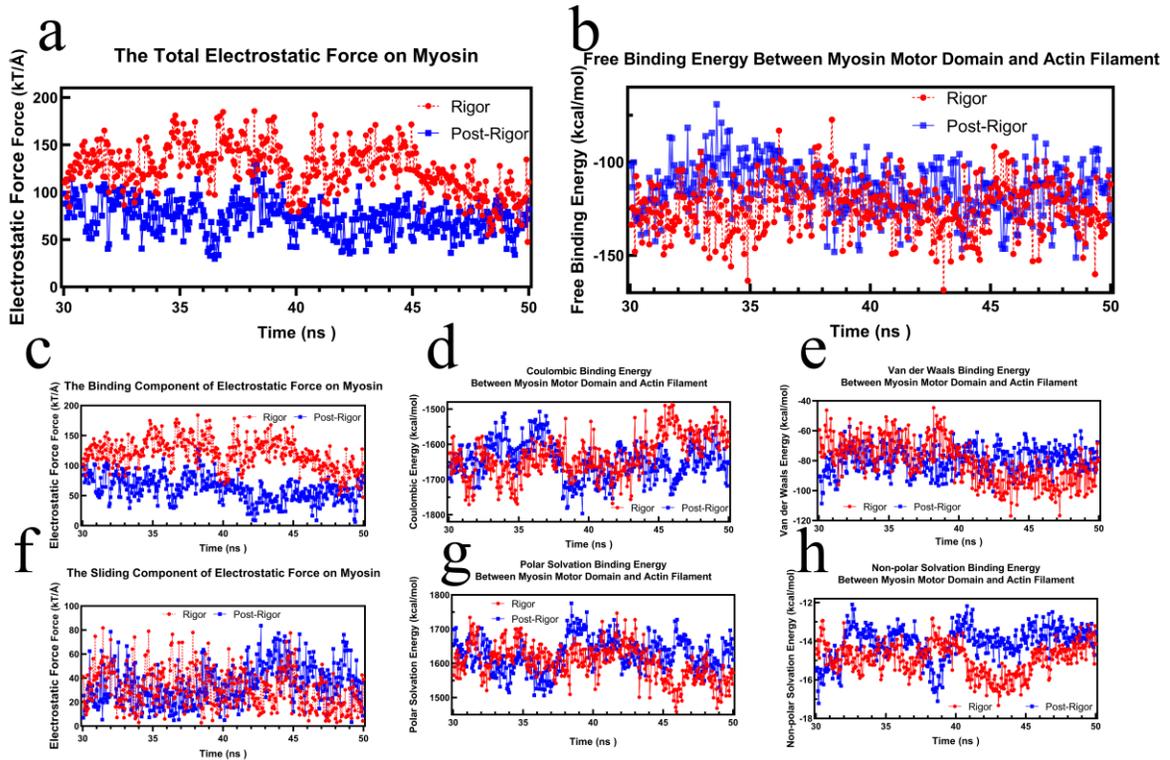
Figure 2.3 The electrostatic force analysis and free energy analysis. a. The total electrostatic force on between myosin motor domain and actin filament. b. The free energy between myosin motor domain and actin filament. c. The binding component of electrostatic force. d. The columbic binding energy. e The Van der waals binding energy. f. The sliding component of electrostatic force. g. The polar solvation binding energy. h. The Non-polar solvation binding energy.

2.3.4 Free binding energy calculation

In addition to testing the electrostatic features, we further tested the free binding energy by MM/PBSA[171, 172]. The MM/PBSA may result in higher binding energy than expectation, but the comparisons of relative binding energy are reliable. In Fig. 2.3b and Table 3.1, the actomyosin complex in rigor state is at a lower state (-124.25 kcal/mol) than that in post-rigor state (-114.78kcal/mol). The difference infers that binding in the rigor state is stronger than that post-rigor state. To further analyze the contributions among each component, the electrostatic, Van der Waals, polar and non-polar solvation binding energy were separately calculated (Eq. 2.4). Results are shown in Table 2.1 and Fig. 2.4degh. Van der Waals and solvation binding energy were lower in rigor state; this is preferable for binding. The polar solvation binding energy contributes most to the rigor state, which is 22.18 kcal/mol lower than in post-rigor state. The difference is due to the conformational change of the post-rigor myosin, which decreases the area of the interface while extending the solvent-accessible surface area.

Table 2.1: Free binding energy (kcal/mol)

	Electrostatic binding energy		Van der Waals binding energy		Polar solvation binding energy		Non-polar solvation binding energy		Free binding energy	
	Rigor	Post-Rigor	Rigor	Post-Rigor	Rigor	Post-Rigor	Rigor	Post-Rigor	Rigor	Post-Rigor
Average	-1631.87	-1646.70	-80.83	-79.56	1603.37	1625.55	-14.91	-14.06	-124.25	-114.78
SD	58.86	52.18	14.12	8.73	51.92	48.26	0.82	0.88	14.28	13.81

3.3.5 Salt bridges and hydrogen bonds.

To better understand the reason for the decrease in binding affinity between rigor and post-rigor, the salt bridges and hydrogen bonds from the last 20 ns were calculated by VMD. The results are shown in Tables 2.2 and 2.3. In the rigor state, there were eight salt bridges with occupancy higher than 0.5 and three salt bridges with occupancy higher than 0.9. By contrast, in the post-rigor state, there were five salt bridges with occupancy higher than 0.5 and one salt bridge with occupancy higher than 0.9. The positions of salt bridges with higher occupancy (> 0.8) are shown in Fig. 2.4. Most salt bridges in rigor state disappeared in post-rigor state. However, the salt bridge (GLU 574-ARG 95) also appeared in the post-rigor state, and the occupancy was increased. Similarly, in the rigor state, a salt bridge formed by LYS 367 and GLU 180 (Tpm X while the LYS 367 in post-rigor formed a salt bridge with GLU 184 (Tpm X). These changes in salt bridges and related residues are related to the transformation from rigor state to post-rigor state. In rigor state (Table 2.3), there were nine hydrogen bonds with occupancy higher than 0.5 and one hydrogen bond with occupancy higher than 0.9. By contrast, there were six hydrogen bonds in the post-rigor state with occupancy higher than 0.5 and no hydrogen bonds with occupancy higher than 0.9. In summary, more salt bridges and hydrogen bonds appear in the rigor state than in the post-rigor state. The fewer salt bridges and fewer hydrogen bonds infer the higher binding energy and instability in the post-rigor state, contributing to the detachment of the myosin from the actin filament.
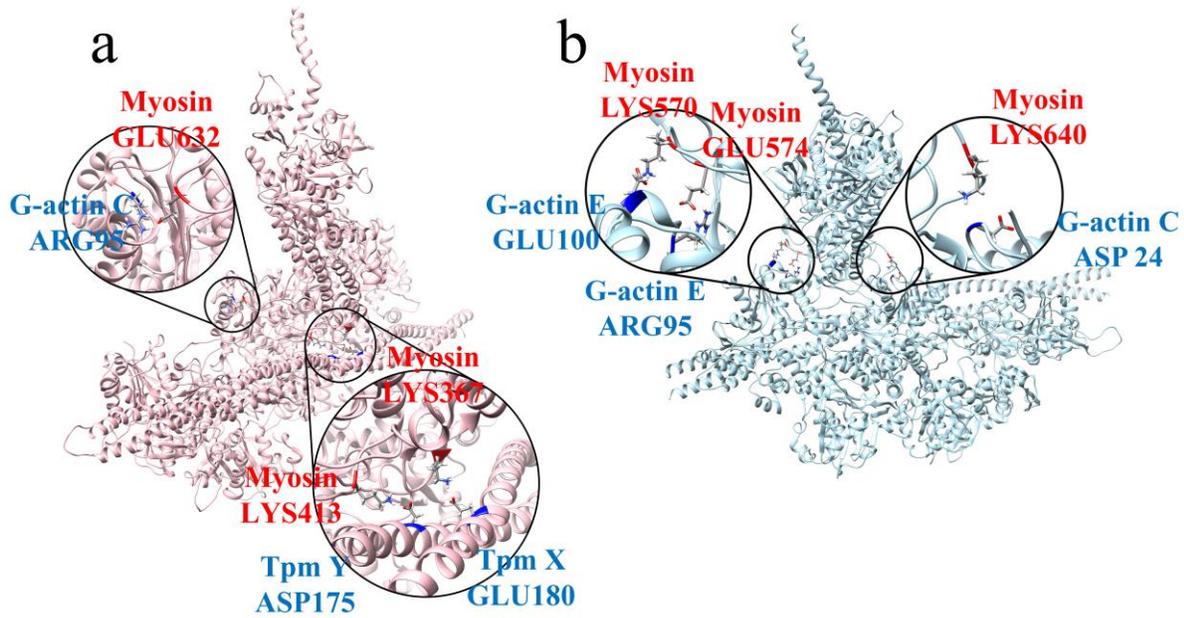
Figure 2.4: The high-occupancy (> 0.8) salt bridges pairs between myosin motor domain and actin filament in rigor (a) and post-rigor (b).

Tables 2.2: The salt bridges between myosin motor domain and actin filament.

Rigor				Post-Rigor			
Myosin	Occupancy	actin filament		Myosin	Occupancy	Actin filament	
LYS367	** 0.994	GLU180	Tpm X	LYS570	** 0.993	GLU100	G-actin E
LYS413	** 0.958	ASP175	Tpm Y	GLU574	* 0.869	ARG95	G-actin E
GLU632	** 0.936	ARG95	G-actin C	LYS640	* 0.836	ASP24	G-actin C
GLU574	* 0.734	ARG95	G-actin E	LYS367	* 0.728	GLU184	Tpm X
LYS639	* 0.721	ASP25	G-actin C	ARG369	* 0.545	GLU192	Tpm X
LYS397	* 0.682	GLU173	Tpm X	LYS635	* 0.441	ASP25	G-actin C
GLU370	* 0.532	ARG178	Tpm Y	LYS542	0.283	GLU167	G-actin C
ARG403	* 0.519	GLU164	Tpm Y	GLU409	0.272	ARG178	Tpm Y
ARG369	0.229	GLU181	Tpm Y	LYS635	0.183	GLU334	G-actin C
GLU371	0.054	LYS328	G-actin C	LYS640	0.009	ASP25	G-actin C
LYS365	0.032	GLU184	Tpm X	ARG369	0.002	GLU195	Tpm X
ASP554	0.024	LYS50	G-actin E				
LYS639	0.021	ASP24	G-actin C				
GLU632	0.02	ARG28	G-actin C				
ARG567	0.015	GLU99	G-actin E				
ARG369	0.013	ASP311	G-actin C				

**: > 0.90
*: > 0.50

Tables 2.3: The high-occupancy (>0.1) hydrogen bonds between myosin motor domain and actin filament.

Rigor	Post-Rigor

Donor	Acceptor	Occupancy	Donor	Acceptor	Occupancy
G-actin C ARG95	Myosin GLU632	** 0.90	G-actin E ARG95	Myosin GLU574	0.83
G-actin C SER350	Myosin GLU536	0.82	Myosin LYS570	G-actin E GLU100	0.79
G-actin C SER350	Myosin GLU536	0.82	G-actin C THR351	Myosin GLU536	0.78
Myosin LYS367	Tpm-X GLU180	0.71	G-actin C THR351	Myosin GLU536	0.77
G-actin E ARG95	Myosin GLU574	0.71	Myosin ARG369	Tpm-X GLU192	0.65
Tpm-Y ARG178	Myosin GLU370	0.68	Myosin LYS640	G-actin C ASP24	0.50
Myosin SER643	G-actin C ASP25	0.65	Myosin LYS367	Tpm-X GLU184	0.48
Myosin ARG369	Tpm-Y GLU181	0.55	Tpm-Y ARG178	Myosin GLU409	0.33
Myosin ARG403	Tpm-Y GLU164	0.53	Myosin ASN568	G-actin E GLU100	0.31
Myosin LYS639	G-actin C ASP25	0.44	Myosin LYS635	G-actin C ASP25	0.27
Myosin TYR410	Tpm-Y GLU163	0.40	G-actin E GLN49	Myosin ASP554	0.27
Myosin LYS413	Tpm-Y ASP175	0.38	G-actin C SER350	Myosin SER532	0.17
Myosin-LYS397	Tpm-X GLU173	0.33	Myosin LYS542	G-actin C GLU167	0.16
G-actin C ARG28	Myosin GLU632	0.33	Myosin LYS635	G-actin C GLU334	0.13
Myosin-LYS640	G-actin C ASP24	0.32	Myosin ARG369	Tpm-X GLU195	0.12
G-actin E GLN49	Myosin ASP554	0.27	G-actin E ARG95	Myosin HSD576	0.10
Myosin-LYS639	G-actin C GLY23	0.14			

2.4 Conclusion

The actomyosin complex includes a myosin and an actin filament. We assembled actomyosin complexes in rigor and post rigor states respectively. Electrostatic features were calculated to analyze the difference between the two states. The MM/PBSA, salt bridges, and hydrogen bonds were also calculated to specify the differences between the two states. The results show that the positively charged area in the myosin interface decreases from rigor to post-rigor state, causing fewer interactions and lower electrostatic binding strength. In the MM/PBSA, free binding energy is higher in the post-rigor state. It infers that the post-rigor state is less conducive than the rigor state to the binding process. The post-rigor state's fewer salt bridges and hydrogen bonds also suggest lower binding affinity between the myosin domain and the actin filament. However, the free binding energy and the electrostatic binding force in post-rigor were still supportive of binding between the myosin and the actin filament. Generally, detachments were not seen in the simulations without the application of external forces. It suggests the detachment between the myosin and actin filament is assisted by other forces. These forces could be from the

lever arms, the light chain, myosin-binding protein C[173], etc. Our findings show that binding affinity is lower in post-rigor state than that in rigor state. It is the important character for the dissociation of actomyosin complexes. In the future, we aim to simulate the whole kinetic circle of myosin and to deeply analyze the functions of key residues.

CHAPTER 3: IMPLEMENT MACHINE LEARNING ALGORITHMS TO TREAT BOUND IONS IN BIOMOLECULES

3.1 Introduction

Electrostatic features are essential for the proper functioning of biomolecules because the electrostatic character of the molecule plays significant roles in biomolecular binding/repelling[109, 174, 175], ion transport[176], and structural stability[177]. Nevertheless, electrostatic calculation *in silico* is a challenging topic in computational biophysics due to the presence of ions. Ions contribute to the complexity of the environment around biomolecules. Highly charged biomolecules, such as nucleic acids, globular proteins, and motor proteins tend to attract ions with opposite charges to balance the net charge in a local environment. The bound ions significantly affect electrostatic potential on biomolecular surfaces, which further influences the interactions of this molecule with other molecules.

At present, there are two models that can handle ions and water surrounding biomolecules: the explicit solvent model and the implicit solvent model. The explicit model is widely applied in all-atom molecular dynamics (MD) simulations when the goal is to simulate ions and H_2O (TIP3P[178], TIP4P[179]) explicitly—that is, with coordinates for each water and ion atom. The explicit model neutralizes highly charged biomolecules by adding unbalanced amounts of cations and anions in the modeled system. However, the electrostatic calculations by the explicit model may consider billions of atoms in hundreds of frames (each frame including the coordinates of the protein, water, and ions). This approach is extremely memory-intensive and time-consuming. Therefore, the implicit solvent model seems to be more suitable to handle ions and water surrounding biomolecules. Implicit solvent models include Poisson-Boltzmann (PB) model[180] and Generalized Born (GB) model[181]. In both models, the ionic environment of biomolecules

is treated homogeneously by setting dielectric constants for biomolecules and solutions. The implicit method avoids the energy calculations for ions and water in solutions[48, 182]. Compared with the explicit method, the implicit model has the obvious advantage of accelerating electrostatic calculations on biomolecules. However, the homogeneous treatment of the solvent ignores local effects of bound ions. The homogenous solvent is not realistic in many situations—especially for highly charged biomolecules. In some cases, even though the net charge of a biomolecule is neutral, the charge distribution on the surface of the biomolecule is not. In detail, the interfaces between biomolecules are often highly charged, for binding or repelling other molecules. Those charged areas in biomolecules may cause nonhomogeneous ionic distribution. Overall, the homogeneous treatment of solutions has the limitation of ignoring this feature of highly charged biomolecules.

To solve the above difficulties when treating highly charged biomolecules, we proposed a Hybridizing Ions Treatment (HIT) method[50, 183] for representing ions in implicit solvation calculations. HIT method adds explicitly bound ions into the implicit solvent model for the electrostatic calculation. The core idea behind this method is using the frames of MD simulations to calculate the position of bound ions via clustering. Compared with prediction methods by coordination numbers[184], geometries[185], or electrostatic potentials[182, 186], HIT calculations are more reliable. This is because HIT uses the results from all-atom MD simulations to do calculations. Most popular MD programs, such as NAMD[160] and GROMACS[187], take the majority of intermolecular forces acting on a system, including Van der Waals (VDW) forces and electrostatic forces, into consideration. Nevertheless, the HIT method has difficulties handling biomolecules with significant movement. It also has trouble treating multi-component systems

simultaneously. Overall, HIT worked well in previous test cases, but sometimes calculations were computationally expensive and program input was unfriendly for users.

Here we report HIT-2, a new version of the Hybridizing Ions Treatment method to combine bound ions and implicit solvent solutions for accurate electrostatic calculations. This method does not require the setting of a cube size or the type and number of bound ions, unlike the first version of HIT[50]. We observed two issues when we developed HIT at first stage: redundancy and incorrect calculation[50]. Redundancy means two or more calculated ion positions are close to one real position of bound ions while an incorrect calculation means the calculated position is far away (over 5 Å) from the real position. Redundancy occurs because HIT cuts a binding site into several pieces equally or approximately equally, and then the two or more segments, which are treated as multiple binding sites. Because HIT only selected bound ions according to ranking of occupancy, redundancy further causes the last correct results to be abandoned. This results in incorrect calculations. In HIT-2, we applied an iteration technique to handle redundancy and incorrect calculations. Iteration ensures a binding site is fully encompassed within a cube by increasing the cube size, so eliminating redundancy and subsequent incorrect calculations. Furthermore, we defined several independent parameters and prepared tools to train and test HIT-2 by machine learning methods.

Several machine learning methods were applied to optimize HIT-2 parameters. We developed a Random Ions Generation Tool (RIGT) to generate 3888 cases for parameter optimization via machine learning. To ensure the accuracy of ion distributions in RIGT, the Maxwell Boltzmann distribution was applied to simulate ionic distributions. After classification and optimization, the best parameters were found and applied to test real datasets (real explicit MD simulations). We tested HIT-2 in proteins with bound Ca^{2+}/Zn^{2+} to further validate the

39

accuracy of the bound ions' positions. Also, we applied HIT-2 on proteins and nucleic acids to validate the broad applicability. Moreover, we improved HIT-2 so that it is also able to handle biomolecules with significant conformational changes. Lastly, we applied HIT-2 to predict the binding position for signal ions (Ca^{2+}) in troponins to exhibit the wide applicability of this method. The results showed it is a very promising tool in computational biophysics and related fields.

3.2 Methods

3.2.1 Dataset

3.2.1.1 Random ions training sets

A water box with ions simulation was prepared by NAMD[160] to observe the velocity distribution of ions. The velocity of ions in the x, y, and z directions was calculated in pm/ns (Å/0.01 ns) and fitted to a Maxwell Boltzmann distribution[188]. The Maxwell Boltzmann distribution (Eq. 3.4) was first defined and used for describing particle speeds in idealized gases. Here we applied it to fit ions moving in solvation boxes.

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ (3.4)}$$

Where the v is the ion's speed and v_x, v_y, and v_z is the speed component in the *x*, *y*, and *z* directions. The v conforms to a Chi-squared distribution while the v_x, v_y, and v_z conform to a normal distribution. Here we took the unit of pm/ns as the ion's speed in simulation. The result was applied for ion generation in the Random Ions Generation Tool (RIGT).

RIGT is the tool designed to generate abundant ionic simulations for HIT-2 testing. It can quickly generate a series of trajectories of ions' simulation for a certain time. The ions trajectories were further combined by RIGT to generate an ionic cloud file (In PDB format). Here in our experiment, the solution was 150 mM NaCl. 10 Na^+ and 10 Cl^- were trapped in a sphere with random diameters of 0 - 5 Å. Maxwell Boltzmann distribution was applied to the velocity of ions

to better simulate each frame of the ionic cloud. Here, the interval is 0.01 ns for saving frames of the ionic cloud. 54 datasets were generated with random simulation times (0.05 ns – 40 ns).

3.2.1.2 Testing sets in MD simulations

A random ions testing set is solvated by a 150 Å × 150 Å × 150 Å solvation box with 150 mM NaCl with 10 K^+ and 10 Cl^- restrained to simulate the bound ions. Cysteine dioxygenase (PDB: 5L0S)[189] and Factor VIIa with bound Ca^{2+} (PDB: 5PB2)[190] were used for Ca^{2+} testing while insulin with bound Zn^{2+} (PDB: 1ZEH)[191] was used for Zn^{2+} testing. Moreover, DNA(PDB: 5J2M)[192] and RNA(PDB: 4TNA)[193] with bound ions are were also tested by HIT-2. The membrane protein BamA (PDB: 4K3B)[194] was simulated to include biomolecules which exhibit marked conformational changes. Additionally, we also conducted actin filament simulations to further explore the potential ability of HIT-2 for the search of signal ions and corresponding binding sites. A piece of actin filament with troponins (PDB: 6KN8)[144] was chosen as the model for simulations. The missing loops in these structures were made up by Swiss-model[195]. The solvation was achieved by VMD[159]/CHARMM-GUI[196] using TIP3P[178]. The membrane of BamA was constructed by the CHAMRMM-GUI membrane builder[197]. MD Simulations were performed on NAMD 2.12[160]. The electrostatic potential was calculated by Delphi[48] using the CHARMM36m[198] forcefield while charges were assigned by pdb2pqr[199]. For all models, the minimization was 10000 steps and followed by 10.2 ns simulations using NAMD 2.12. The simulations include two steps. The first is equilibration (0.2 ns), and the second is production run. NPT was used for the equilibration and NVT was applied in the production run. In equilibration, the constraint was applied to the entire protein/nucleic acids/membrane with the harmonic restraints (no restraints on hydrogens). In the production run, all atoms were no longer restrained in a certain position. In both simulations, the

CHARMM36m[198] was applied as the force field. The temperature was set as 310 K (controlled with a Langevin thermostat with the damping constant as 1.0) and the time step was set as 2fs/step. The PME (full-system periodic electrostatics) was applied in the simulation. The production run of simulations for BamA is 100 ns.

3.2.2 Algorithm

3.2.2.1 Preparation, solvation box cutting, and ion counting

After MD simulations, the relative positions of ions are calculated using the mass center of biomolecules in the corresponding frame. Then ions in all frames are aligned together by their relative positions, forming an ionic cloud (Fig. 3.1ab). The cube size (The side length of each cube) is initialized to 3 Å. Based on the initialized cube size, the ionic cloud is cut into several cubes (Fig. 3.1c). The ions in each cube are counted to calculate the filling ratio (Eq. 4.1 and Fig. 1d).

$$R_f = \frac{n_{ic}}{n_f} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.1)$$

Where the R_f represents the filling ratio, n_{ic} represents the number of ions in the corresponding cube, and the n_f represents the number of frames.

42
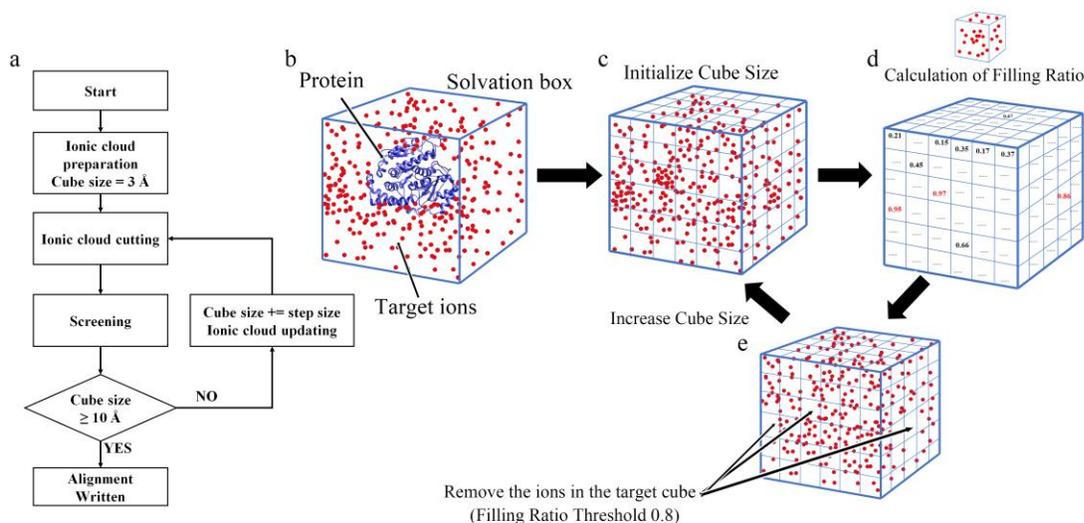
Figure 3.1: The flow chart and schematic graph of the algorithm. a. The flow chart of HIT-2. b. The preparation of the ionic cloud. c. The ionic cloud cutting partition step. d. The calculation of the filling ratio for each cube. e. Removing ions from the ionic cloud after binding site calculations and then increasing the cube size for the next iteration. In this diagram, the filling ratio threshold is set as 0.8. In e, the cubes with filling ratio higher 0.8 are regarded as binding sites and all ions in corresponding cubes will be removed from ionic cloud before the next iteration. The iterations will run until the cube size is bigger than 10 Å.

3.2.2.2 Screening and ionic cloud updating

If the filling ratio is higher than a given threshold (filling ratio threshold), the corresponding cube is selected as a binding site, where the mass center is calculated by equation 3.2, representing the position of the bound ion (Eq. 3.1). The ions in the selected cubes are then removed from the ionic cloud (Fig. 3.1e). Then the cube size increases by a given step size (Å). The step size means the increment of the cube size in iterations. The updated ionic cloud and cube size are further used in the iterations until the cube size is bigger than 10 Å. The filling ratio threshold and step size are important parameters to be optimized by machine learning methods.

$$\vec{P} = \frac{\Sigma \vec{v}_i}{n_{ic}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \text{(3.2)}$$

Where the $\vec{P}$ represents the position of bound ions and the $\vec{v}_i$ represents the positions of all ions in the selected cubes.

3.2.2.3 Structural alignment

After iterations of screening and ionic cloud updating, the positions of bound ions are logged and aligned to the biomolecular structure in the target frame by equation (3.3). The output is the biomolecule with bound ions in PDB format in the target frame (n^{th} frame decided by users).

$$\vec{P}_o = \vec{P}_b + \vec{P} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \text{(3.3)}$$

Where the $\vec{P}_o$ and $\vec{P}_b$ represent the output positions of bound ions and the mass center of the biomolecules.

43

3.2.3 Algorithm Testing

Based on the RIGT, 54 simulation results (0.05 ns to 40 ns) are prepared for testing. In each dataset, HIT-2 applied different step sizes (0.05 to 1 Å) and filling ratio thresholds (0.5 to 0.99) for testing. In total, 3888 cases were generated including the results generated by HIT-2. We define success and failure in the section below and for all successful cases, the error was calculated for further optimization.

3.2.3.1 Classification and optimization

First, we defined strict criteria for success to describe the results from HIT-2. These are two conditions: **1, the number of calculated bound cations/anions should be the same as the real bound cation/anions; 2, the distance between each pair (calculated bound ion and real bound ion) should be smaller than 5 Å.** If these conditions are not met, the result is a failure. To find what parameters (simulation time, step size, and filling ratio thresholds) lead to success, we applied Logistic Regression (LR)[200], Classification and Regression Tree (CART)[201], Random Forest (RF)[202], and Artificial Neural Networks (ANN)[203] to address the binary classification problem (success/failure). methods. The number of cases is 3888 with a train/test split of 0.7/0.3.　For all successful cases with simulation time over 2 ns, we further tested the distance between our calculations and targets. The average distance of all pairs was measured and regarded as an average error for further analysis. The contributions made by simulation time, step size, and filling ratio thresholds to average error are analyzed.　Additionally, the run time of HIT-2 is highly related to step size and simulation time so we also consider these criteria when determining optimal parameters. The run times of HIT-2 for different cases are also measured and compared with the average error for step size and simulation time optimization.

After classifications, quantitative analysis was applied to understand the relationship between error and related parameters. In successful cases, the number and type of calculated bound ions are same as that of real bound ions. Each cases includes 20 pairs of real bound ions and calculated bound ions. We consider the distance between the real and calculated ions as the error and calculate this value for the 20 pairs. The average error was plotted against simulation time, filling ratio and step size for quantitative analysis. Moreover, we also calculated the running time of HIT-2 to further optimize the parameters for users. The PC used for this optimization was a Dell-XPS with Intel I7 processor (i7-11700-2.5GHz) and 16GB memory.

3.3 Results

3.3.1 The workflow of HIT-2

HIT-2 involves 4 steps: preparation, ionic cloud cutting, screening, and position alignment (Fig. 3.1a). In the preparation step, the relative positions of all ions in all frames from MD simulations are calculated and combined into an ionic cloud (Fig. 3.1b). In the ionic cloud cutting step, the ionic cloud is cut into several cubes (Fig. 3.1c). Afterward, the ions are counted and divided by the number of frames to calculate the filling ratio of each cube (Fig. 3.1d). If the filling ratios are larger than the threshold (filling ratio threshold), those cubes are regarded as binding sites. The mass centers of the binding area are the positions of bound ions (Fig. 3.1e). The ions in the binding sites are removed from the ionic cloud before the next iteration. In the next iteration, the size of the cube is increased by the step size. In the alignment step, the positions of bound ions are calculated by the relative positions of bound ions and the mass center of biomolecules (Fig. 3.1a). The simulation time, the filling ratio threshold, and the step size are three crucial parameters affecting the results of HIT-2. We further optimized these parameters by employing several classification machine learning methods with different parameters.

3.3.2 Classification

HIT-2 requires certain conditions to find bound ions from simulations. The three parameters, including filling ratio, step size, and simulation times are set as inputs, and the success/failure is set as an output. Several machine learning methods were applied to obtain a better model to search the parameter space for successful results. Here, we define two very strict criteria for success: 1, the number of calculated bound cations/anions must be the same as the number of real bound cation/anions; 2, the distance between each pair (a calculated bound ion and the corresponding real bound ion) must be shorter than 5 Å.

Fig. 3.2a illustrates the ROC of Logistic Regression (LR)[200], Classification and Regression Tree (CART) [201, 204], Random Forest (RF)[202], and Artificial Neural Network (ANN)[203]. LR has an accuracy of 84% with 0.83 Area Under the Curve (AUC) while the RF, CART, and ANN have high accuracy of around 95%. Among these methods, the ANN possesses the highest AUC (0.97) and the highest accuracy (95%). We further used the ANN model to predict the result from different input parameters (simulation time, filling ratio threshold, and step size). Millions of cases were generated by the ANN model to show the relationship between different parameters and results. Fig. 3.2b shows the predicted success rate by different filling ratio thresholds and simulation times. After the 5 ns simulation, the success rate for different filling ratio thresholds ($\geq$0.55) is nearly 100%. The success rate is highly related to the simulation time. This is because ideal ionic clouds are formed when simulations reach equilibrium. The longer simulations produce more stable equilibria. Among different filling ratio thresholds, the thresholds of 0.95 and 1.00 first reached a 100% success rate after 1 ns. Fig. 3.2c show the success rate predicted by ANN with different step sizes and simulation times. The step size is the increment of cube size in iterations. Intuitively, a smaller step size should lead to more accurate results. The

result shows that when the step size is increased to 1 Å, only 50% of calculations were correct (even when the simulation time is greater than 20 ns). With the decrease in step size, a higher success rate is achieved. When the step size is lower than 0.2 Å, there is no significant difference. Indeed, all simulations with a step size lower than 0.2 Å reached a near 100% success rate after 1 ns. In summary, for most simulations a filling ratio of 0.95 and step size of 0.2 Å is enough to get a nearly 100% success rate after running for 1 ns.
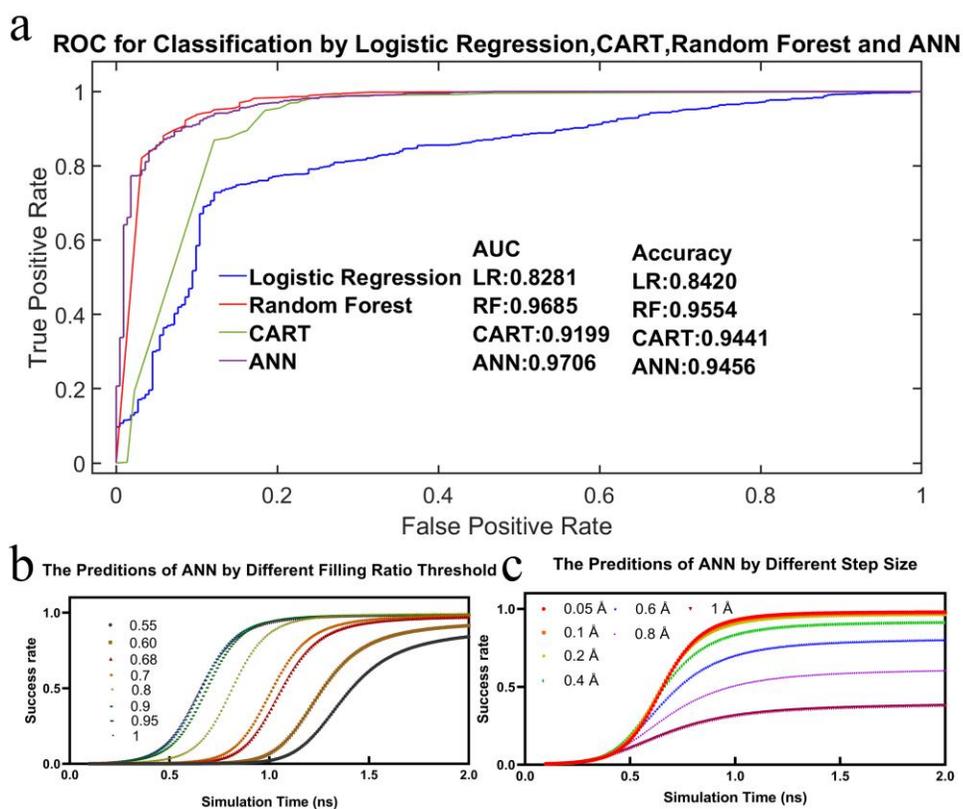
Figure 3.2: The comparison among several machine learning methods and the results generated by ANN. a. The ROC results of Logistic Regression, Random Forest, CART, and ANN. b. The success rate predicted by ANN with different filling ratio thresholds (step size = 0.05). c. The success rate predicted by ANN with different step sizes (The filling ratio is 1.00).

5.3.3 Optimization

Due to the very strict criteria for success, the condition of success is enough for HIT-2 to get reliable results. However, we are interested in improving HIT-2 performance as much as possible and we were not satisfied with "just success". The error (distance between calculated

47

bound ions and real bound ions) should be further quantified to optimize the parameters for users. The average error shown in Fig. 3.3 demonstrates how HIT-2 may still be improved. The relationship between the filling ratio threshold and average error is shown in Figure 3.3a. The average error is decreased when filling ratio threshold increases. Average error reaches a minimum when the filling ratio threshold is 0.95. When the filling ratio is increased to 0.99, error slightly increases. In the generated datasets, there is no escape or rebinding of ions in the binding sites. In this case, the ideal situation (equilibrium system) will cause the lowest error to appear at the point of filling ratio equal to 100%. However, in most cases, an equilibrium simulation cannot be practically achieved because such simulations are time-consuming. In our error analysis, the lowest value appeared when the filling ratio equaled 0.95. The filling ratio is approximately equal to the occurrence frequency of bound ions. In nature, bound ions sometimes are too mobile to be bound tightly, resulting in a frequency of occurrence lower than 100%. In this case, researchers can choose a better filling ratio for their studies. Additionally, testing with a range of filling ratio thresholds is also a good idea to distinguish the strength of different bound ions. With that said, we remark that the filling ratio should always be higher than 0.5; otherwise, a binding site would be treated as two binding sites and this would likely cause incorrect calculations [50].
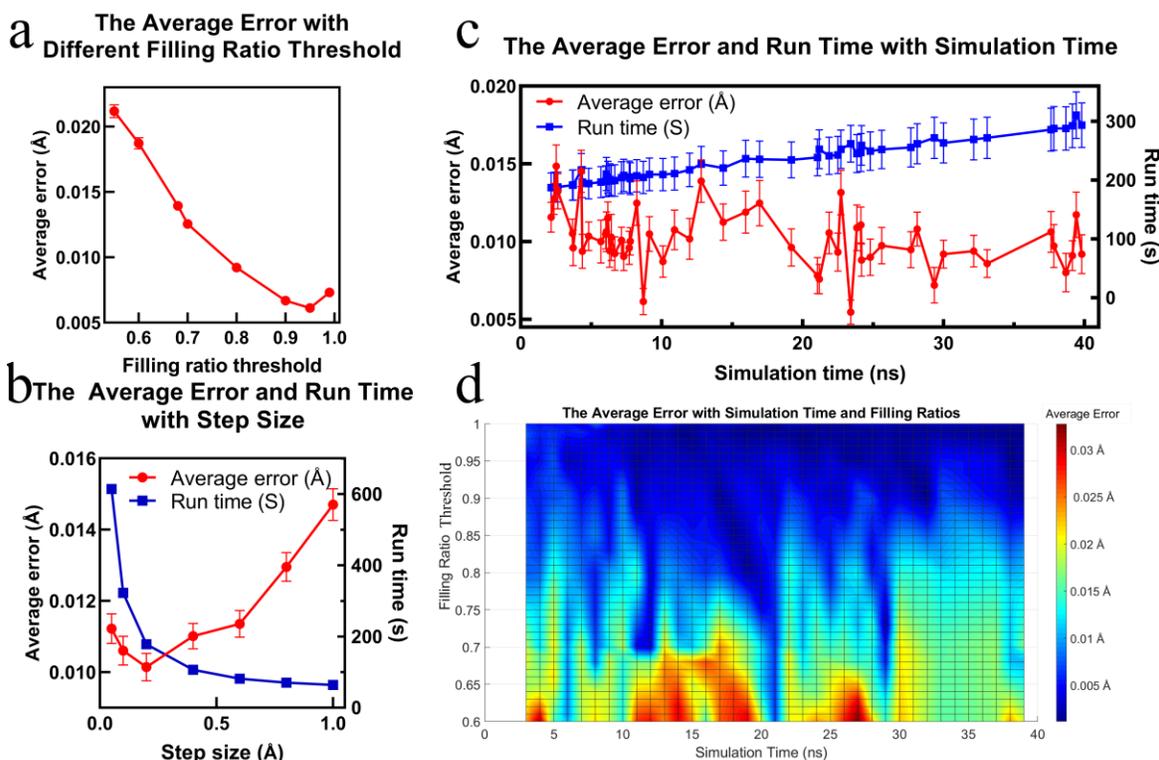
Figure 3.3: The average error and running time plotted against filling ratio, step size, and simulation time. a. The average error against the filling ratio. b. The run time and average error against the step size (the filling ratio is 0.99). c. The run time and average error against simulation time (the step size is 0.2 Å). d. The heat map of average error with different filling ratio thresholds and simulation times (the step size is 0.05 Å).

Step size and simulation time are highly related to the accuracy and the running time of HIT-2. With a decrease in step size, the average error linearly decreased until 0.01 Å at the point of step size = 0.2 Å. When the step size is lower than 0.2 Å, the average error begins to increase (Fig. 3.3b). This is because the screening step (Algorithm section) may count several free ions or miss several bound ions from the ionic cloud. A small step size (<0.2 Å) leads to more screening iterations and may cover more unrelated ions into consideration, causing the bias. This error can be reduced by performing long MD simulations. Moreover, a smaller step size also requires calculations be performed for many small cubes which significantly increases the computational cost of HIT-2. From Fig 5.3b., we see that when step size is lower than 0.2 Å, the run time increase

49

from 3 min to a maximum of 10 min (Fig. 3.3b). Therefore, the 0.2 Å step size is the best choice

for most users. By contrast, after 5 ns, longer simulation contributes less to error minimization.

The average error decreases with an increase of simulation time (Fig. 3.3c). The simulation time

and the HIT-2 run time are linearly related because simulation time is equivalent to the number of

frames, which affects the preparation of the ionic cloud (Algorithm section). Compared to the

filling ratio and step sizes, the simulation time is not that highly related to error minimization.

However, simulation time is highly related to the success rate, which is fundamental to error

minimization. So at least 5 ns simulations with a frequency of less than 10000 fs/frame are

suggested to get accurate results.

The heat map showing average error against simulation time and filling ratio is shown in

Fig. 2d. Simulations longer than 30 ns have an average error lower than 0.02 Å. The average error

decreases dramatically with when the filling ratio increases because the ions in the dataset are

100% bound. Overall, the highest average error is 0.03 Å, which means the average error for any

pair is lower than 0.6 Å (0.03 Å x 20 = 0.6 Å, all error happened at a bound ion) if the calculation

succeeds. Also, we observe that the presence of red peaks diminishes with an increase in simulation

time. Additionally, at higher simulation times, the error decreases more smoothly with an increase

in the filling ratio. Thus, longer simulations can reduce the standard deviations for average errors,

making the results more reliable.

5.3.4 Testing and Application

In the random ions simulations from NAMD, 10 Na^+ and 10 Cl^- are restrained to simulate

bound ions. The calculations by HIT-2 tested the ionic cloud from 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, and

10 ns. Notably, from 0.5 ns onward, the calculations from HIT-2 are successful. We further

analyzed the average errors and the associated standard deviations (Fig. 3.4). We observe the errors

and standard deviations decreased with the increase in simulation time. These results are consistent with the above analysis, further proving the accuracy of HIT-2.
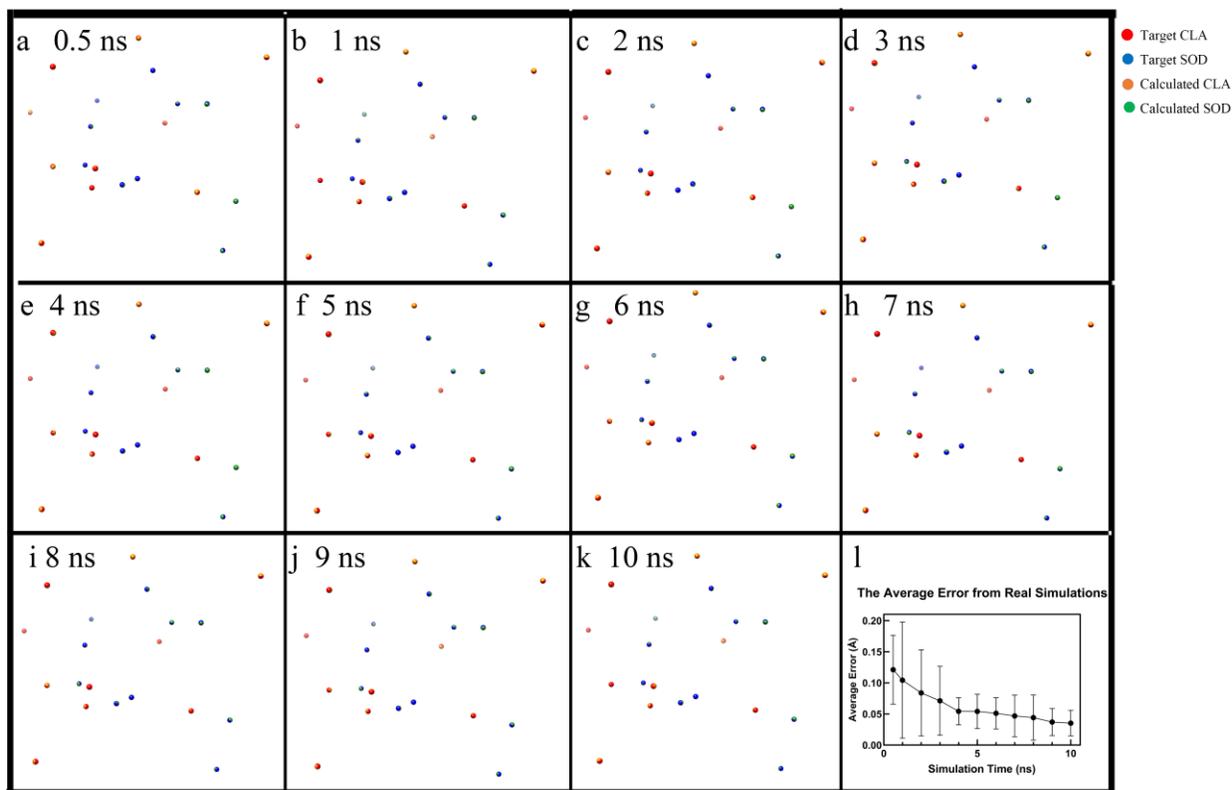
Figure 3.4: The results from NAMD testing set by different simulation time (0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ns), where the target chloride ions, sodium ions, calculated chloride ions, calculated sodium ions are colored red, blue, orange, and green, respectively. All real bound ions are found by HIT-2 and covered by calculated bound ions. l. The results from the NAMD testing set (average error is the average distance among 20 pairs of calculated bound cations/anions).

In X-ray crystallography experiments, ions are co-crystalized with a macromolecule and are thus resolved as a part of the structure. For our tests, we chose the X-ray structure from the PDB bank (5L0S[189], 5PB2[190], and 1ZEH[191]) to do accuracy testing for HIT-2 to determine the positions of Ca^{2+} and Zn^{2+}. The errors associated with our predictions for Ca^{2+} are 3.26 Å and 0.88 Å (Fig. 3.5ef) and those for Zn^{2+} are 0.18 Å and 1.26Å (Fig. 3.4). All calculations are successful (error < 5 Å), but the error exceeds our expectations. The main reason is that the binding area may be bigger than the training sets, causing the bound ions to have a wider active region. In

51

addition to this result, we observe HIT-2 also works for other biomolecules, including DNA/RNA (Fig. 3.5gh) and biomolecules with significant conformational changes (Fig. 3.5jkl). In short, without inputting the number and name of bound ions, HIT-2 directly found the correct number and positions of the ions with an error lower than 4 Å for various *in vitro* systems.
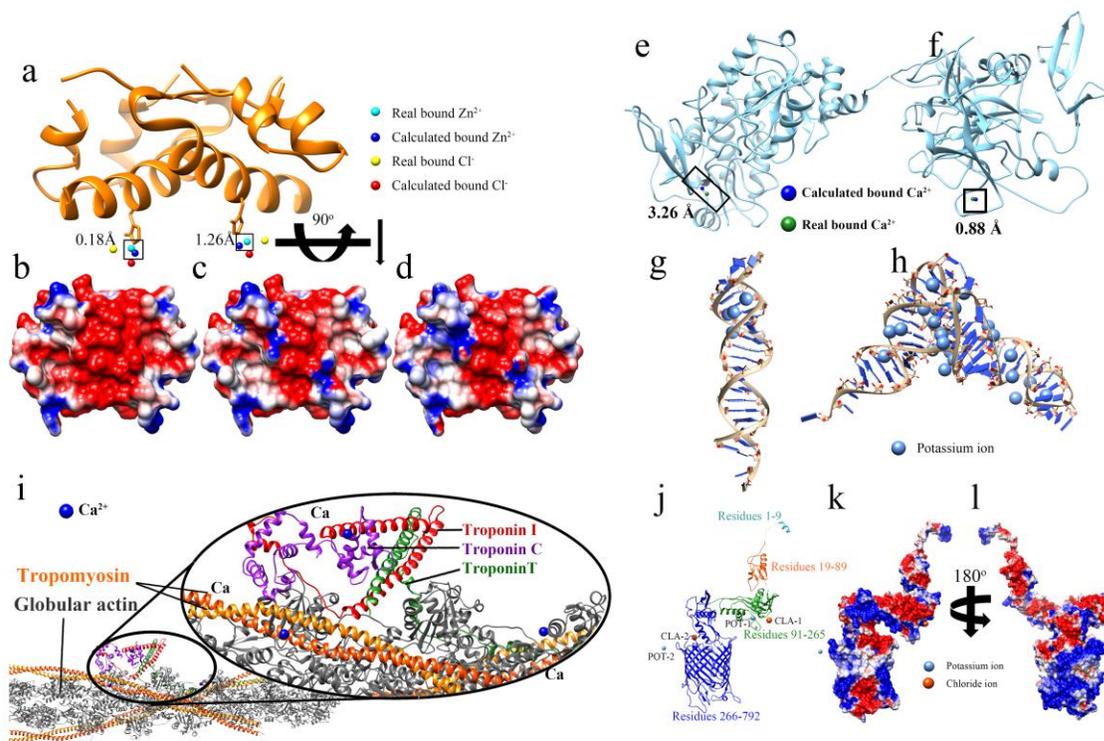
Figure 3.5: Tests and applications of HIT-2. a. The results from bound Zn^{2+} testing for the protein (PDB: 1ZEH) b. The electrostatic surface without bound ions. c. The electrostatic surface with real bound ions. d. The electrostatic surface with calculated bound ions. e. The calculated and real bound Ca^{2+} in protein (PDB: 5L0S). f. The calculated and real bound Ca^{2+} in protein (PDB: 5PB2). g. The structure of HIV-DNA with bound potassium ions. h. The structure of tRNA with bound potassium ions. the light blue balls represent the bound potassium ions. i. The structure of the actin filament with bound calcium ions was produced by HIT-2. jkl. Example of HIT-2 multiple tests for proteins with conformational changes.

We further colored the surface of 1ZEH by electrostatic potential (Fig. 3.5a-d), with and without bound ions. The binding sites without bound Zn^{2+} are negatively charged (Fig. 3.5b). *In vitro*, the bound Zn^{2+} imparted a positively charge to the surface of the protein (Fig. 3.5d). Similarly, with HIT-2, the bound calculated Zn^{2+} also applied a positive charge (Fig. 3.5c).

Some bound ions such as Ca^{2+} also play vital roles in intracellular signaling, such as muscle contractions. Here we simulate with a piece of actin-filament with a troponin complex (including troponin I, troponin C, and troponin T), globular-actins (G-actin), and tropomyosin. Three bound calcium ions (Fig. 3.5i) were found by HIT-2 with two of these being located inside the G-actins and tropomyosin and the other attached to the troponin. These results are consistent with previous studies that show troponin is a receptor for calcium ions. [205]. Moreover, the calcium is not only bound to troponin but also attached to the whole actin filament. The conformational changes of the actin filament could be affected by the coverage of calcium ions on the actin filament. It requires further deep analysis and experimental support.

3.4 Discussion

Bound ions are crucial for the function of highly charged biomolecules. The current implicit solvation model has many limitations when applied to highly charged biomolecules. To solve this problem, our previous work, HIT, was developed, and works by hybridizing bound ions with the implicit solvent model to improve electrostatic calculations.

Because the clustering algorithm in our previous version of HIT is computationally expensive, we sought to make improvements. Here, we abandoned the clustering method in HIT and employed machine learning approaches in HIT-2. Also, the newest HIT-2 tool is more user-friendly by eliminating the need to specify cube size and the type and number of bound ions.

To validate our work, we tested HIT-2 on DNA, RNA, and membrane-bound proteins. Also, we showed HIT-2 can be applied to biomolecules that undergo significant conformational changes, and biomolecules with multiple components. Expectedly, our work shows HIT-2 can also be used for predicting the location of signal ions and associated receptors. Specifically, we demonstrated HIT-2 can locate the position of Ca^{2+} in a troponin complex.

In conclusion, HIT-2 is useful for dealing with difficult biological problems that cannot be resolved by experimentation. However, there are some limitations of HIT-2. First, it performs very well in binary systems with one type of cation and one type of anion, but when multiple types of cations and anions are involved, the problem would require multiple uses of HIT-2. Second, while HIT-2 is able to handle the movement of biomolecules in simulations, it cannot handle rotations. The rotated biomolecules should be aligned first in VMD or other software before the use of HIT-2. In spite of these problems, HIT-2 is a useful tool for users who are studying highly charged biomolecular systems.

CHAPTER 4: PHOSPHORYLATION OF TYROSINE 841 WITHIN THE JAK3 KINASE DOMAIN PROMOTES AN ACTIVATED JH1 STATE

4.1 Introduction

The Janus Kinase (JAK) family are intracellular tyrosine kinases, including JAK1, JAK2, JAK3, and Tyrosine Kinase 2 (TYK2)[206, 207]. They are involved in cytokine signaling via the phosphorylation of Signal Transducers and Activators of Transcription (STATs)[208] through the JAK/STAT pathway. When cytokines bind to their specific receptors, it causes the receptors to dimerize and activate JAK kinases to phosphorylate the receptor. This phosphorylation creates a docking site for Signal Transducers and Activators of Transcription (STATs) to bind. Once the STATs are bound, they are also phosphorylated, dimerized, and translocate into the nucleus where they bind to DNA and regulate the expression of cytokine-responsive genes[209, 210]. The JAK/STAT signaling pathway is critical for the normal functioning of hematopoietic cells, which are responsible for the production and maturation of blood cells. Dysregulation of this pathway has been associated with various hematological malignancies, including leukemia, lymphoma, and myeloproliferative neoplasms. Additionally, mutations in JAK genes have been identified in several types of cancers, such as myelofibrosis and acute lymphoblastic leukemia [211-213]. Understanding the mechanisms underlying the JAK/STAT pathway can provide insights into the development of new therapies for these diseases [214].

JAKs contain 4 fundamental domains, a carboxyl-terminal kinase domain (JH1), an anterior pseudokinase domain (JH2), Src Homology 2 (SH2), and Four-point-one, Ezrin, Radixin, Moesin (FERM). JH1 is the domain with catalytic ability while JH2 lacks catalytic activity but possesses regulatory function[207]. The JAK1, JAK2, and TYK2 are ubiquitously found in different cell lineages[215, 216] while JAK3 is expressed primarily in cells of the immune

system[207, 217-219]. In the JH1 domain, several phosphorylation sites were discovered that regulate catalytic activity positively or negatively. Tyrosines 980 and 981 (Y980 and Y981) in JAK3 were discovered to be autophosphorylated. The kinase activity is positively regulated by phosphorylated tyrosine 980 (pY980) and negatively regulated by pY981[220]. The phosphorylations of Y904 and Y939 were also shown to positively promote JAK3 signal transduction. The effects of phosphorylation of the latter residues were found to be associated with the optimal usage of ATP and promote STAT5 activity[221]. Correspondingly, autophosphorylation of homologous JAK2 sites was also found on Y868, Y972, and Y966. It proved the universality and importance of tyrosine phosphorylation on JAKs. Recently, we found and reported a new tyrosine phosphorylation site, JAK3-Y841. It positively regulates the catalytic activity in the activation of JAK3.

The biological evidence for the regulation of JAK by phosphorylation and dephosphorylation is solid and clear[218, 222]. However, due to technical limitations, previous studies faced difficulties in providing structural details of the effects, including the chemical and physical reasons for the activation of kinases. The challenges of revealing the biological process of activation on kinase are mainly due to the lack of quantitative analyses. Now, based on computational methods and advanced software, we can perform molecular dynamic (MD) simulations[160] to examine the impact on protein structure. The phosphorylation effects on conformational changes, electrostatic forces, and binding energies can be further studied. In this work, we focused on the ATP hydrolysis process and analyzed the binding between ATP/ADP and JAK3-JH1 in Y841/pY841 states. The results show that the binding affinities between JH1-Y841 and ATP/ADP are similar, which may cause difficulties in releasing ADP. They also suggest that the JH1-Y841 is not the optimal stage for ATP hydrolysis. By contrast, JAK3-pY841 has a higher

binding affinity to ATP than ADP, which is essential for the kinase to function normally by binding ATP and releasing ADP. Moreover, based on the full-length JAKs structure[223] and the autoinhibited state of JH1 and JH2[224], we successfully modeled the full-length structure of JAK3 in the autoinhibited state. We found that pY841 also weakened the binding between JH1 and JH2, showing higher sliding force on JH1. They could promote the JH1 flip to the activated position and expose the ATP binding pocket.

4.2 Method

4.2.1 JAK3 Modeling

The autoinhibited/activated full-length JAK3 structure was built through SWISS-MOLDEL by the dimerized full-length JAK1 structure (PDB: 7T6F[223]) and autoinhibited JH1-JH2 dimer (PDB: 4OLI[224]). The phosphorylation of Y841 (pY841) was built by CHARMM-GUI[196]. Autoinhibited full-length JAK3 structure with Y841 (FY841) and pY841(FpY841) simulations were performed to study the effects of pY841 on the JH1 in the autoinhibited state. The JH1-Y841 and JH1-pY841 were separated from whole structure to study the effects of pY841 on ATP/ADP hydrolysis. The JH1 domain with Y841/pY841 was combined with ATP/ADP to build 6 models, including two non-bound models (JH1-Y841 and JH1-pY841) two models bound to ATP (ATP-JH1-Y841 and ATP-JH1-pY841), and two models bound to ADP (ADP-JH1-Y841 and ADP-JH1-pY841). The ATP or ADP was aligned into the ATP-binding pocket of the JH1 domain (The phosphate groups were outside while the adenosine was inside[225]). The model used for alignment of ATP/ADP is JAK3 JH1 with inhibitor (PDB: 5TTV)[226]. We aligned the adenines of ATP/ADP to the inhibitor in 5TTV to form the JH1-ATP/ADP complexes.

4.2.2 Molecular dynamics simulation

For the simulations of each model, the force field applied is CHARMM36m[198]. Models were solvated by TIP3P [178] via CHARMM-GUI[196]. 150 mM KCl was set to ionize the system. Periodic boundary conditions and the long-range electrostatic interactions with particle mesh Ewald[9] were performed during simulations. The temperature was set at 310 K, using a Langevin thermostat with a damping coefficient of 1/ps. The pressure was set to 1 atm using a Nosé–Hoover Langevin piston barostat with a decay period of 25 fs. Simulations were performed after 10,000-step minimization. It includes two steps: NPT equilibrium and NVT production. In equilibrium, the backbones of proteins and ATP/ADP were restrained. In production, all atoms were free except for the two residues in the N- and C- terminals. The simulations were performed by NAMD 2.12[160]. In FY841 and FpY841 simulations, the 0.5 ns NPT equilibrium and 45 ns NVT production were performed. In JH1-ATP/ADP simulations, the 0.5 ns NPT equilibrium and 100 ns NVT production were performed.

4.2.3 Salt bridge and hydrogen bond

Salt bridge and hydrogen bond formation were analyzed by Visual Molecular Dynamics (VMD)[159]. The cutoff distance of salt bridges was set as 4 Å while the cutoff distance and angle for the hydrogen bonds were set as 3.5 Å and 20°[110, 227], respectively.

4.2.4 Topological analysis

In the trajectories of JH1-ATP/ADP simulations, we observed slight conformational changes in the N-lobe of the kinase domain. We calculated the distance between the mass centers of the N-lobe (T815-P906) and C-lobe (S907-S1100) by Structure-Man[40]. The distance represents the cleft size between the N-lobe and C-lobe of the JH1 domain. The average and standard deviation of the distance in 300 frames (70ns-100ns) of 6 models (JH1-Y841, JH1-

pY841, ATP-JH1-Y841, ATP-JH1-pY841, ADP-JH1-Y841, and ADP-JH1-pY841) were calculated to show the topological effects of phosphorylation of Y841.

4.2.5 Electrostatic analysis

The electrostatic potential of the N-lobe of JH1-Y841/pY841 with and without ATP/ADP at 0 ns and 100 ns were calculated by Delphi[48]. Negatively charged (red), positively charged (blue) and neutrally charged (white) surfaces of the N-lobe were depicted by Chimera[107]. Electrostatic forces between JH-1 and ATP/ADP were also calculated from 70 ns to 100 ns by DelphiForce[52]. Electrostatic forces between JH1 and JH2 in FY841 and FpY841 models were calculated from 35 ns to 45 ns. The binding/sliding force was calculated using equations 4.1 or 4.2:

$$F_{binding} = F_T \cdot cos\alpha \quad \text{……………………………….. (4.1)}$$

$$F_{sliding} = F_T \cdot sin\alpha \quad \text{..………………...……….. (4.2)}$$

Where F_T represents the total electrostatic force between two subdomains/ligands and α is the angle between the electrostatic force direction and the mass center connection of two subdomains/ligands (JH1-ATP/ADP or JH1-JH2). The charge and radius were assigned by pdb2pqr[199] from CHARMM36m[198]. The dielectric constant for protein and water were set as 2 and 80, respectively. Salt concentration, probe radius, filling ratio of protein, and resolution were set as 150 mM, 1.4 Å, 0.70, and 2 grid/Å, respectively. The separations of JH1 from JAK3 were visualized by VMD[159].

4.2.6 MM/PBSA Free Binding Energy

The Free energy is calculated by the last 10 ns frames of FY841 and FpY841. We separated JH1 from full-length JAK3 to calculate the binding free energy for JH1 and the FERM-SH2-JH2

complex. The free energy E was calculated by the sum of coulombic, van der Waals, polar, and nonpolar solvation binding energies (Eq. 4.3).

$$E = E_c + E_v + E_p + E_{np} \quad \text{……………………………….... (4.3)}$$

Where E_c and E_p represent coulombic binding energy and polar solvation energy, calculated by Delphi[48]. E_v represents the van der Waals binding energy, calculated by NAMD 2.12[160]. E_{np} represents the nonpolar solvation energy, calculated by the solvent-accessible surface area (SASA) from equation 6.4.

$$E_{np} = \alpha \cdot SASA + \beta \quad \text{………………………………….. (4.4)}$$

Where E_{np} is the nonpolar solvation energy. $\alpha = 0.0054$, and $\beta = 0.92$ kcal/mol. The SASA was calculated using NACCESS[228].

Binding energy (ΔE_{bind}) is calculated by the following equation (Eq. 6.5)[109, 110, 172].

$$\Delta E_{bind} = E_{complex} - E_{JH1} - E_{FERM+SH2+JH2} \text{………………………….. (4.5)}$$

Where the $E_{complex}$, E_{JH1} and $E_{FERM+SH+JH2}$ represent the free energy for the full-length JAK3, JH1 domain, and FERM+SH2+JH2 complex.

4.3 Results

4.3.1 The pY841 effects on autoinhibited full-length JAK3.

The stability of full-length autoinhibited JAK3s with Y841 and pY841 (FY841 and FpY841) models was assessed after 35ns MD simulations, and both models were found to be stable (Fig. 4.1ab). To gain further insights, we performed 10 ns MD simulations to analyze the force and energy involved. Our analysis (Figure 3.2ab) showed that the high-occupancy hydrogen bonds of E601-R840 (86.14%) and S580-L851 (68.32%) disappeared after the phosphorylation of Y841, whereas the occupancy of hydrogen bonds of R657-D846 increased from 46.53% to 52.48%. Moreover, a new hydrogen bond between S826-E542 was formed with 45.54% occupancy. The

overall total occupancy of hydrogen bonds decreased, indicating that the interactions between JH1-pY841 and JH2 were weaker than those between JH1-Y841 and JH2. Notably, key residues such as R841, L851, D846, and S826 are geometrically close to Y841, and the presence of pY841 strongly affects the adjacent residues and hydrogen bond formation. We also examined the binding free energy between JH1-Y841/pY841 and the FERM-SH2-JH2 complex and found that FpY841 had a higher binding free energy (Fig. 4.2c), suggesting that the presence of pY841 promotes the detachment of JH1 from FERM-SH2-JH2.
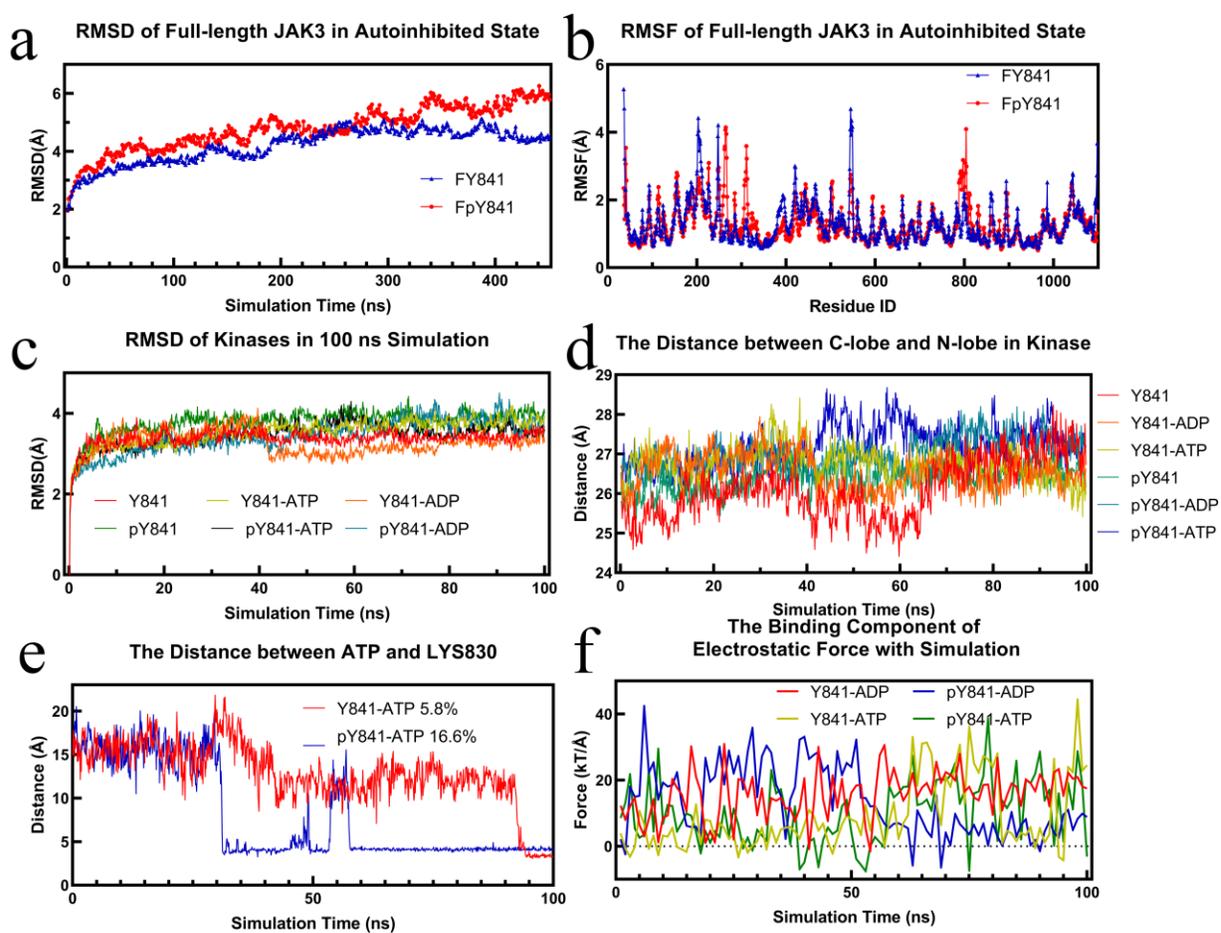
Figure 4.1: The raw data from MD simulations. a. RMSD of full length JAK3 with non- phosphorated and phosphorated Y841 (FY841 and pY841) in auto-inhibited state. b. RMSF of full length JAK3 with non-phosphorated and phosphorated Y841 (FY841 and FpY841) in auto-inhibited state. c. RMSD of the kinase Y841, pY841, and that with ADP or ATP. d. Cleft size of kinase Y841, pY841, and that with ADP or ATP. e. Distances between the salt bridge pairs LYS830/ATP in Y841 and pY841. The occupancy of LYS830/ATP in 5.8% while that in pY841 is 16.6%. f. Binding components of electrostatic force between kinase and ADP/ATP in Y841, pY841.
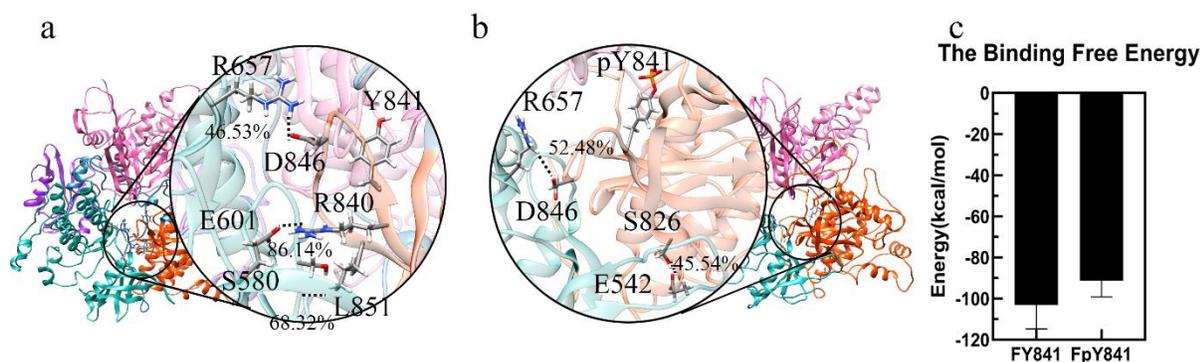
Figure 4.2: The ribbon diagram of high-occupancy (>30%) hydrogen bonds between JH1 and JH2 and binding free energy. The FERM, SH2, JH2 and JH1 are colored pink, purple, blue, and orange respectively. a. The high-occupancy hydrogen bonds in FY841. b. The high-occupancy hydrogen bonds in FpY841. c. The binding free energy between JH1 and other domains (FERM, SH2, JH2) in full-length JAK3 with Y841(FY841) and pY841(FpY841).

The loss of hydrogen bonds and increased binding free energy caused by the pY841 phosphorylation could lead to the detachment of JH1 from autoinhibited JAK3, which is the first step towards activation. To further investigate the potential effects of Y841 phosphorylation on the conformational changes of JAK3, the JH1 and JH2 domains were separated by 10, 15, and 20 Å to simulate the initial steps of activation. The electrostatic forces between JH1 and JH2 were then analyzed to explore the potential continuous effects on JAK3 conformational changes. As shown in Fig. 4.3ad, the sliding force is the main component for sliding JH1 around JH2. Interestingly, the electrostatic force between JH1 and JH2 is nearly perpendicular to the JH1- JH2 mass center connection, as seen in Fig. 4.3bc. This perpendicular alignment indicates that the

sliding force plays a critical role in the sliding of JH1 around JH2, which is a crucial step in the activation process. It is also noteworthy that the JH1 in the activated JAK3 is on the other side of JH2 (Fig. 4.3b), and the increased sliding force on JH1, caused by pY841, may contribute to the flipping of JH1 until the activated state is achieved. Overall, the simulation results suggest that Y841 phosphorylation may affect the detachment of JH1 from the autoinhibited JAK3 by reducing the hydrogen bonds and increasing the binding free energy. The increased sliding force on JH1 caused by pY841 may further contribute to the flipping of JH1 and the activation of JAK3. The electrostatic force between JH1 and JH2 was found to be perpendicular to the JH1- JH2 mass center connection, suggesting that the sliding force is the primary driving force in the JH1 sliding around JH2. These results provide new insights into the conformational changes of JAK3 and its regulation by Y841 phosphorylation, which may have important implications for the development of therapeutic interventions targeting JAK3.
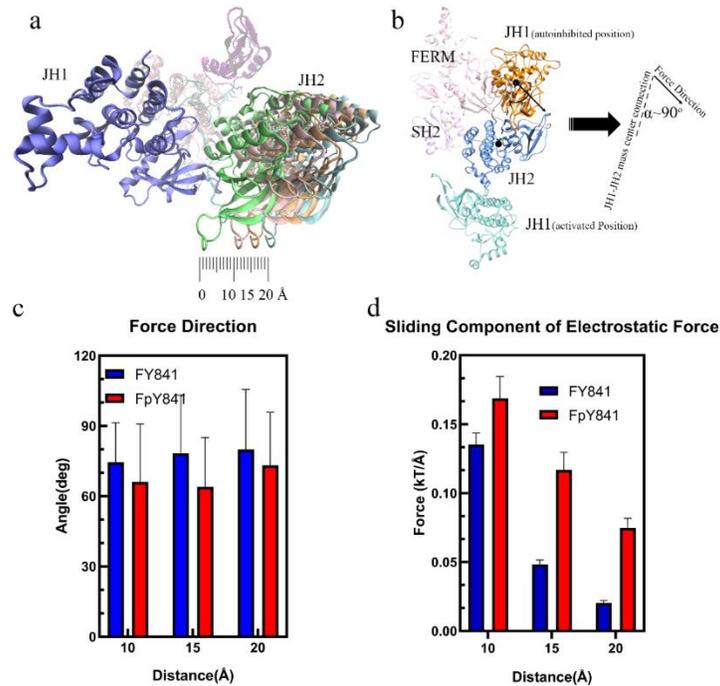
Figure 4.3: The interactions from FY841/FpY841 to JH1. a. The diagram of separation between JH1 and JH2. b. The diagram of the angle (α) between electrostatic force and the JH1-JH2 mass center connection. c. The angle (α) between JH1-JH2 mass center connection and the electrostatic force direction. d. The sliding component of electrostatic force between JH1 and JH2 at 10, 15, 20 Å.

4.3.2 JAK3-JH1 ATP/ADP binding process

A 100 ns simulation was performed for each of the JH1 domain states, including JH1-Y841, JH1-pY841, ATP-JH1-Y841, ATP-JH1-pY841, ADP-JH1-Y841, and ADP-JH1-pY841, and the stability of the conformations (with Root Mean Square Deviation < 4 Å) was observed after 10 ns (Fig. 4.1c). Based on the last 30 ns simulations, the cleft size and electrostatic forces between the JH1 N-lobe and C-lobe were calculated. As shown in Fig. 4.4ab, the phosphorylation of Y841 resulted in a reduction of the JH1 cleft by approximately 0.5 Å, while the JAK3 JH1-pY841 cleft was found to be enlarged by 1 Å after binding with either ATP or ADP. Conversely, the binding of ATP/ADP decreased the cleft size of JAK3-Y841 by about 0.6 Å. Interestingly, despite the decrease in the JH1 cleft caused by pY841, ATP/ADP binding can still enlarge the cleft size, which allows the substrate to enter and facilitates the transfer of gamma phosphate. The phosphorylation of Y841 reverses the conformational change caused by ATP/ADP binding. The wide range of cleft sizes indicates that the phosphorylation of Y841 increases the overall elasticity of the kinase domain. Additionally, the increased cleft size expands the volume of the nucleotide-binding pocket and enhances the binding force of ATP/ADP.
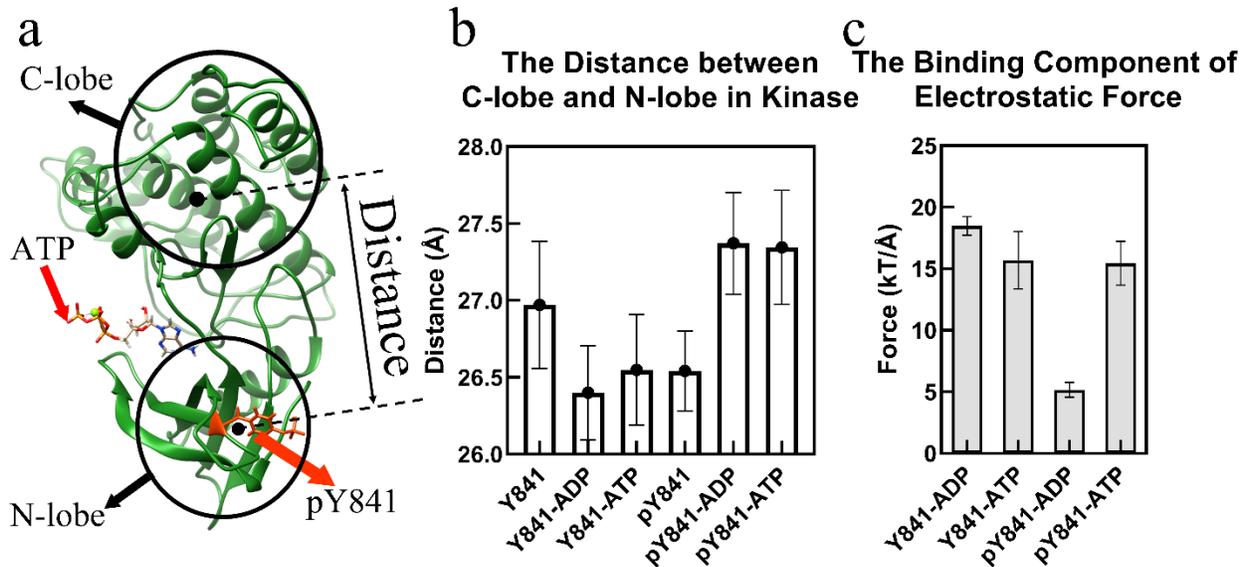
Figure 4.4: Phosphorylation of JH1-Y841 promotes an activated kinase domain state. a. The ribbon diagram of the JAK3 JH1-pY841 domain with ATP. The JH1 cleft is the space between the N-lobe and C-lobe. b. The distance between the JH1 C-lobe and N-lobe in the simulation of JH1-Y841, JH1-pY841, ATP-JH1-Y841, ATP-JH1-pY841, ADP-JH1-Y841, ADP-JH1-pY841. c. The binding component of the electrostatic force between the JH1 domain (Y841/pY841) and ADP/ATP (The error bar is based on the CI of 0.95).

Fig. 4.4c reveals that the binding forces between JH1 with ATP are similar in the presence or absence of Y841 phosphorylation (~15 kT/Å). However, the binding forces between JH1 without and with Y841 phosphorylation and ADP are significantly different. The binding force between JH1 without Y841 phosphorylation and ADP is ~18 kT/Å, which is slightly higher than that when bound to ATP. This suggests that the kinase has difficulty releasing ADP. In contrast, after Y841 phosphorylation, the binding force between JH1 and ADP becomes much weaker, at only ~5 kT/Å, which is significantly lower than that with ATP (ATP-JH1-pY841). The reduced binding force implies that releasing ADP becomes easier, allowing a new ATP to bind and reactivate the catalytic activity. Subsequently, we also analyzed the electrostatic surface of the JH1 domain to visualize the effects of phosphorylation of Y841 (Fig. 4.5). We found that the pY841 state alters the positively charged surface area surrounding the site, as shown in figures 2a and 2b.

In the N-lobe of the JH1 ATP/ADP binding pocket, phosphorylation does not directly cause changes to the electrostatic surface, as shown in Fig. 4.5cd. Nevertheless, after 100 ns simulations with ATP/ADP, the nucleotide-binding pocket shows significant differences. After simulations with ATP, the binding pocket in JH1-Y841 is positively charged (Fig. 4.5e), whereas the JH1-pY841 state is observed to be nearly neutral (Fig. 4.5f). By contrast, the binding pocket is more neutral (Fig. 4.5g) after ADP-JH1-Y841 simulations, but more positive (Fig. 4.5h) after ADP-JH1-pY841 simulations. These electrostatic surface changes are caused by the conformational change of peptide backbones and amino acid side chains. These changes may also contribute to the shifting size of the nucleotide-binding pocket. We observed that the phosphorylation of Y841 initially only affects a small area on the N-lobe, while the simulations amplify the effects on the binding pocket. The conformational change induced by phosphorylation may create new surface charges or alter the distribution of existing charges, affecting the binding pocket's overall charge distribution.
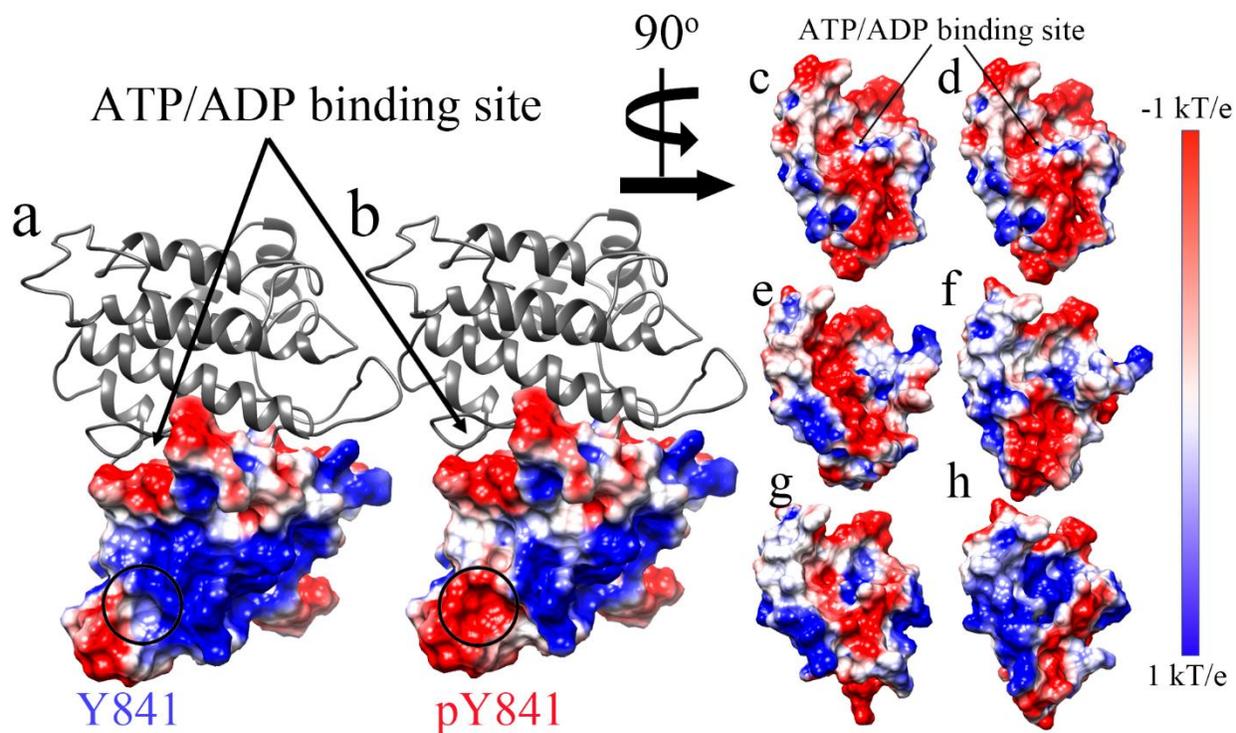
Figure 4.5: The electrostatic surface of the N-lobe of JAK3 JH before and after simulations. ac. JH1-Y841 before simulations. bd. JH1-pY841 before simulations. e. ATP-JH1-Y841 after 100 ns simulation. f. ATP-JH1-pY841 after 100 ns simulation. g. ADP-JH1-Y841 after 100 ns simulation. h. ADP-JH1-pY841 after 100 ns simulation. The surface of electrostatic potential is colored from -1 kT/e (blue) to 1 kT/e (red).

In order to gain a more comprehensive understanding of the effects of Y841 phosphorylation, we analyzed the salt bridges and hydrogen bonds at the intermolecular interfaces. Our analysis revealed that there is only one salt bridge, ATP-LYS830 (Fig. 4.1e), between ATP and JH1 in both states (Y841 and pY841). However, the occupancy of the salt bridge is increased from 5.8% to 16.6% in the pY841 state. Correspondingly, the occupancies of hydrogen bonds between ATP and JH1-Y841 (R916-ATP: 20.7%, L905-ATP: 69.0%) are much lower than that between ATP and JH1-pY841 (R916-ATP: 44.3%, L905-ATP: 82.9%) (Fig. 4.6bc). Specifically, the occupancies of R916-ATP and L905-ATP are 44.3% and 82.9%, respectively, in ATP-JH1-pY841, whereas they are only 20.7% and 69.0% in ATP-JH1-Y841. Notably, a new hydrogen bond between K830-ATP (occupancy of 54.6%) is formed at the edge of the cleft in ATP-JH1-pY841, stabilizing ATP inside the nucleotide-binding pocket. In ADP-JH1 Y841/pY841 simulations, no salt bridge is formed between the kinase domain and ADP. However, the occupancy of hydrogen bonds between ADP and JH1 is generally higher in the pY841 state compared to the Y841 state. Specifically, in ADP-JH1-pY841 simulations, more residues can form hydrogen bonds with ADP, although most of them have very low occurrences. Only L905-ADP has an occupancy of 86.5% (Fig. 4.6f). On the other hand, in ADP-JH1-Y841 simulations, L905-ADP and R916-ADP have occupancies of 97.0% and 78.6%, respectively (Fig. 4.6e). Notably, the hydrogen bond of R916-ADP in the ADP-JH1-Y841 state is located near the edge of the cleft, which may prevent ADP from being released and inhibit the recovery of catalytic activity. Taken together, our data suggest that the binding force between ADP-JH1-Y841 is higher than that of

ADP-JH1-pY841, further supporting the hypothesis that ADP is easily released after ATP hydrolysis in the Y841 phosphorylated kinase. These high-occupancy hydrogen bonds may serve as potential targets for therapeutic interventions.
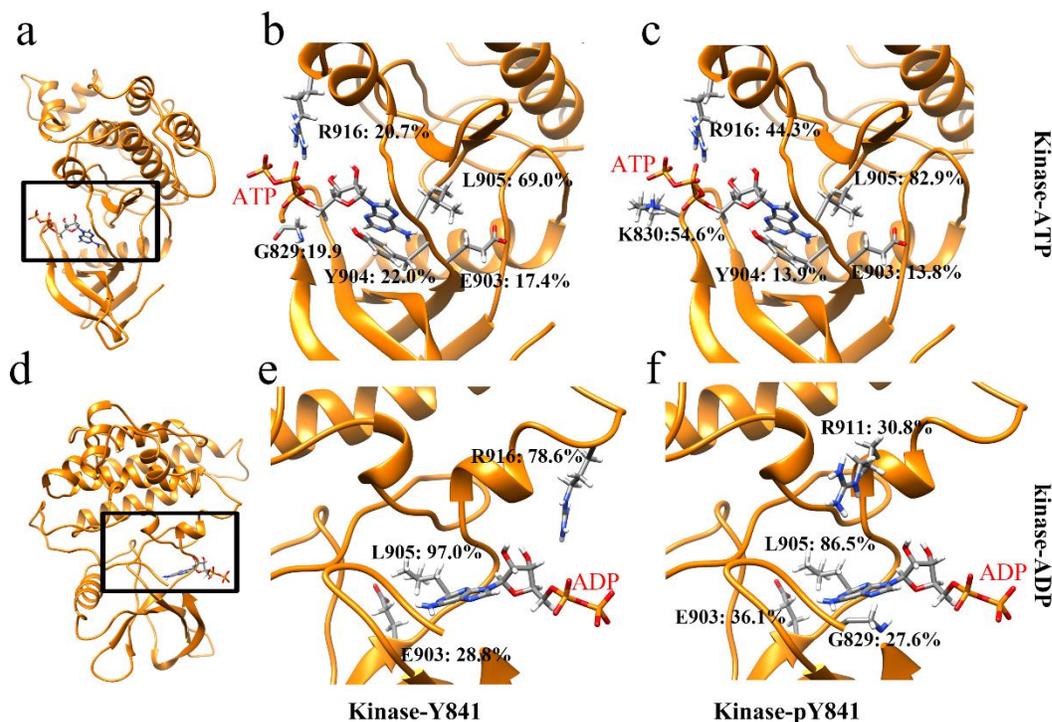
Figure 4.6: The ribbon diagram of high-occupancy (>10%) hydrogen bonds in a. JH1 bound to ATP. d. JH1 bound to ADP. And hydrogen bonds between residues and nucleotides between b. JH1-Y841-ATP. c. JH1-pY841-ATP. e. JH1-Y841-ADP. f. JH1-pY841-ADP.

4.4 Discussion

The pY841 breaks many hydrogen bonds between JH1 and JH2, which are key residues to stabilize the JH1 in the autoinhibited states. The separated JH1-pY841 receives a higher sliding force compared to JH1-Y841, suggesting that pY841 may contribute to the flip of JH1 into the activated position. In following computational study, we conducted topological analysis of the JAK3 JH1 kinase domain with and without ATP/ADP models to investigate the effects of pY841 on ATP hydrolysis. The results show that phosphorylation of Y841 initially decreases the cleft size of the JH1 domain, but dramatically increases the cleft size when ATP/ADP is bound to the kinase.

This finding suggests that phosphorylation of Y841 promotes the activation of JAK3 kinase by facilitating access to the ATP binding pocket. Furthermore, the analysis shows that the electrostatic binding forces between the kinase and ADP/ATP in the JH1-Y841 structure are similar. This similarity makes it difficult to bind ATP and release ADP, hindering the hydrolysis of ATP. However, after phosphorylation of Y841, more salt bridges and hydrogen bonds are formed between JH1 and ATP, which increases the electrostatic binding force on ATP. In contrast, ADP receives a lower electrostatic binding force than ATP, making it easier to release ADP. The increased binding force on ATP and reduced binding force on ADP, induced by phosphorylation of Y841, thus lowers the difficulty for releasing ADP and facilitates the recovery of the kinase catalytic activity. These findings support the idea that pY841 encourages an active JAK3 kinase state by promoting optimal usage of ATP and STAT5 activity. To be noted, phosphorylation of Y904 and Y939 was shown to positively promote optimal usage of ATP and STAT5 activity[221], which suggests that these residues play a critical role in the activation of the JAK3 kinase. One of our key findings was that E903, Y904, and L905 contribute over 100%-occupancy hydrogen bonds and that the phosphorylation of Y841 strongly affects the hydrogen bonds between these key residues and ATP/ADP. This is consistent with previous experimental works and highlights the critical importance of the phosphorylation of Y841 in JAK3 kinase activity. Initially, we observed that pY841 decreased the positively charged surface area surrounding the phosphorylated site without impacting the electrostatic surface of the nucleotide-binding pocket. However, our simulations revealed that over time, more areas in the binding pocket became negatively/positively charged in the pY841 state. This led to the formation of more salt bridges and hydrogen bonds between JH1-ATP, while fewer salt bridges and hydrogen bonds were formed between JH1-ADP.

As a result, the electrostatic binding force between JH1 and ADP was reduced, making it easier for ADP to be released. This likely encourages the kinase to recover catalytic activity.

4.5 Conclusion

Phosphorylation of tyrosine residue Y841 on the Janus kinase (JAK) protein initiates a cascade of events that ultimately leads to the activation of the protein kinase. Specifically, phosphorylation of Y841 breaks the hydrogen bonds between JH1 and JH2 domains, which reduces the binding affinity between them. As a result, JH1 domain detaches from JH2 and flips around it, exposing the ATP binding pocket. During this process, the sliding force acting on JH1-pY841 is increased, promoting its movement to the activated state. The analysis of ATP/ADP binding pockets in the JAK kinase reveals that phosphorylation of Y841 only negatively charges a small portion of the kinase and does not directly affect the ATP/ADP binding pockets. However, long-term simulations of 100 ns show that phosphorylation indirectly affects the binding pockets. Steric effects and negative charges of phosphorylation cause changes in the negatively and positively charged areas of JH1-ATP/ADP that ultimately affect the cleft size and binding force. The highly charged area of JH1-ATP/ADP not only increases the cleft size when ATP/ADP is bound to the kinase but also strengthens the electrostatic binding force on ATP while weakening it on ADP. This reduced force on ADP facilitates the release of ADP, which further facilitates the recovery of kinase catalytic activity. On the other hand, JH1-pY841-ATP has a higher occupancy of hydrogen bonds, contributing to the attraction of ATP. Moreover, JH1-pY841-ADP has fewer hydrogen bonds than JH1-pY841-ATP, facilitating the release of ADP after hydrolysis. The reduced hydrogen bonding in JH1-pY841-ADP leads to a lower binding force between ADP and the JAK kinase, which makes it easier for ADP to be released after hydrolysis. Overall, phosphorylation of Y841 plays a critical role in the activation of JAK kinase. It initiates the

70

separation of the JH1 and JH2 domains, promoting the movement of JH1-pY841 to the activated

state and exposing the ATP binding pocket. Furthermore, phosphorylation indirectly affects the

binding pockets by changing the negatively and positively charged areas of JH1-ATP/ADP, which

ultimately affects the cleft size and binding force. These changes facilitate the release of ADP after

hydrolysis, which further promotes the recovery of the kinase catalytic activity.

CHAPTER 5: USING ENERGY MAP AND DYNAMIC PROGRAMMING TO ANALYZE THE STRUCTURAL TRANSITION OF JANUS KINASE

5.1 Introduction

The family of Janus kinases (JAKs) plays a crucial role in the regulation of various cellular processes, including immune and inflammatory responses[229], hematopoiesis[211, 230], and oncogenesis[231]. The JAK family members, including JAK1, JAK2, JAK3, and TYK2, are noncovalently connected with the intracellular domain of cytokine receptors. When cytokines bind to the extracellular domains of these receptors, it leads to dimerization of the receptors, triggering JAKs to phosphorylate tyrosine residues[232, 233]. Then, the intracellular domains of the receptors serve as the binding sites for Signal Transducers and Activators of Transcription (STATs) and JAKs phosphorylate tyrosine residues[234].   STATs are further phosphorylated by the activated JAKs, dimerized[235], and translocated to the nucleus. This process regulates transcription of genes and is known as the JAK-STAT signaling pathway.

All JAKs are comprised of seven homology domains, including a Four-point-one, Ezrin, Radixin, Moesin (FERM) domain (JH5, JH6, and JH7), an Src Homology 2 (SH2) domain (JH3 and JH4), and two kinase domains (JH2 and JH1), which have a Pseudo Kinase (PK) and a Tyrosine Kinase (TK),   region respectively[236]. The FERM and SH2 domains are tightly associated with juxta-membrane domain cytokine receptor[237]. PK is a noncatalytic domain but regulate TK[238, 239]. Specifically, PK has autoinhibitory domains that keep TK inactive until stimulation from cytokine receptor dimerization[224]. A classic example that illustrates this feature is the VF mutation. V617F is most common mutation in JAK2-PK, which results in constitutive activity[240]. The V617F mutation causes simple steric disruption, disturbing the inhibited state of TK to activate the JAK2[239, 241]. Notably, in the absence of receptor

dimerization, the activated state is likely transient. A V657F mutation causes a JAK1 dimer to form, freeing the TK for constitutive activity and continuous transphosphorylation[223]. Evidence reveals that the PK has two functions: firstly, to bind TK at the inhibited state, preventing TK from releasing; secondly, to stabilize dimerization when a JAK dimer is formed. The P733L mutation, which also occurs on the JAK1-PK domain, results in impaired transphosphorylation and subsequent immunodeficiency[212]. By contrast, S703I, is a gain-of-function mutation in JAK1, and was found to increase JAK1 activity and partnering of JAKs, leading to hyper/neo-morphic cytokine signaling[213].

To fully understand the mechanisms of these mutations, it is necessary to understand the JAK activation process. However, simulating the entire process using all-atom explicit Molecular Dynamics (MD) simulations is challenging due to the resource requirements and the unknown full-length inhibited structures. In 2014, an autoinhibited structure of PK-TK dimer was revealed[224] and in 2022 the full-length JAK1 dimer structure was reported, providing an opportunity to construct the full-length JAK1 in its inhibited state. An energetic and steric analysis of these two states should reveal potential paths of TK from inhibited state to activated state. We aligned two PKs, one from an inhibited PK-TK dimer[224] and the other from an activated full-length JAK1[223] to determine the position of TK in an inhibited state, then constructed the full-length inhibited JAK1 structure. Interestingly, the interfaces of TKs facing PKs in both inhibited and activated states are identical, suggesting a high likelihood that TK rolls around PK to reach the activated position (Figure 1a). However, TK at activated sates is slightly far from PK compared to inhibited states, suggesting rolling and separation both contribute to TK activation. So, based on separational and rotational movement, we separated and rotated[242] TK around PK to build an energy map to find the smoothest path for TK activation. Each dot in the energy map represents a

knot, indicating that the number of possible paths exceeds n^m (where n and m represent the rotation degree and separation distances, respectively). Therefore, computational methods are required to reduce the complexity of these paths. Based on Dijkstra's algorithm[243, 244], we developed a new dynamic programming approach that identifies the path with minimal variance, representing the smoothest path.

5.2 Methods

5.2.1 Modeling and Molecular Dynamics Simulations

The full-length inhibited and activated JAK1 structures were generated using SWISS-MODEL[195] based on the dimerized full-length JAK1 structure (PDB: 7T6F[223]) and inhibited JH1-JH2 dimer (PDB: 4OLI[224]). The JAK1 sequence used for modeling is the *Homo sapiens* tyrosine-protein kinase JAK1 sequence, which can be obtained from UniProt[245]. Models were then solvated with TIP3P [178] and ionized by 150 mM KCl via CHARMM-GUI[196]. The force field applied in MD simulations is CHARMM36m[198]. Periodic boundary conditions and the long-range electrostatic interactions with particle mesh Ewald[9] were performed during simulations. The temperature was set at 310.15 K, using a Langevin thermostat with a damping coefficient of 1/ps. The pressure was set to 1 atm using a Nosé–Hoover Langevin piston barostat with a decay period of 25 fs. Simulations were performed after 10,000-step minimization. Simulation included two steps: NPT equilibrium and NVT production. In the equilibrium phase, the backbones of proteins were restrained. In production, all atoms were freed. The simulations were performed by NAMD 2.12[160]. The 0.5 ns NPT equilibrium and 50 ns NVT production were performed for each MD simulation.

5.2.2 Electrostatic Analysis

The charge and radius were assigned by pdb2pqr[199] from CHARMM36[198]. The dielectric constant for protein and water were set to 2 and 80, respectively. Salt concentration, probe radius, filling ratio of protein, and resolution were set to 150 mM, 1.4 Å, 0.70, and 2 grid/Å, respectively. The electrostatic potential was calculated by Delphi and the electrostatic force by DelphiForce[52]. The hydrogen bond analyses were performed on VMD based on the last 10 ns MD simulations. The cutoff distance and the angle for hydrogen bond analysis were set to 3.5 Å and 20°, respectively.

5.2.3 Simulation of Activation and Analysis of the Smoothest Path

5.2.3.1 Simulation of Activation

Here we separate the JAK1 into 4 domains (FERM: 34-420, SH2: 439-544, PK:583-855, and TK:875-1153) and rotate the TK about the surface of PK to simulate the activation process (Fig. 5.1a). The three mass centers of PK, TK in the inhibited state (TK-I), and TK in the activated state (TK-A) defined a plane. The plane was applied for the rotation and separation of TK in JAK1 activation. First, TK and PK were separated by 25 Å. For each 1 Å increment, 1000 steps of minimization were applied. TK was then rotated along PK at 10° increments (totaling 145°) for each separation. Similarly, for each of these rotations, 1000 steps of minimization were applied. The grid defining this procedure is shown in Fig. 5.1b. Here, the separation and rotation were achieved by StructureMan[242] while the minimization was applied by NAMD2.12[160] with the CHARMM36m force field[198]. After each 1000-step minimization, the total energy (sum of the potential energies) was calculated by NAMD and used to construct the total energy matrix ($M^{30 \times 17}$) Fig. 5.1c.
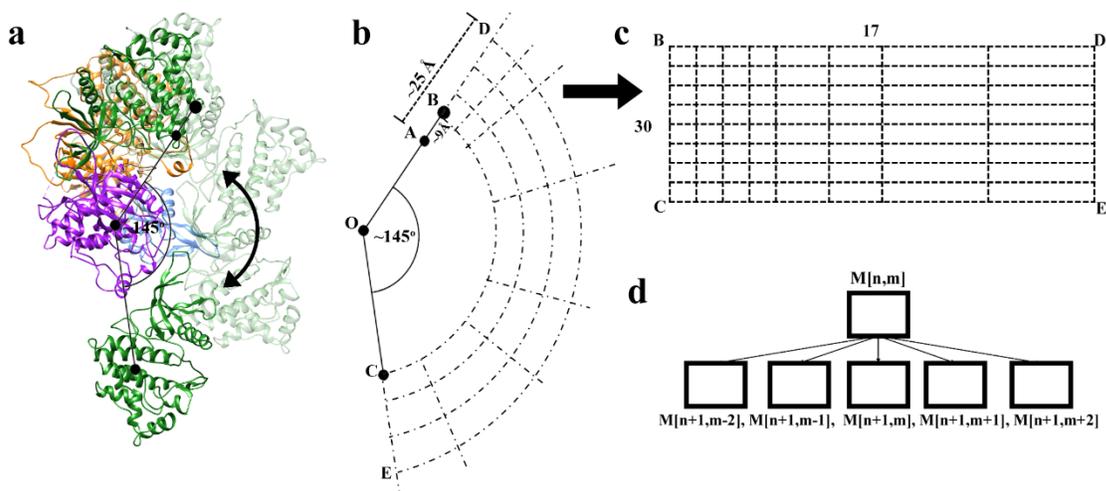
Figure 5.1: The activation process of JAK1 according to the path analysis method. a and b. The diagram of the activation-inhibition process. O represents the mass center of PK, and A and C represent the mass centers of TK in inhibited and activated states respectively. B is the mass center of TK after separating TK from PK by 9 Å to make sure OC=OB. D represents the mass center of TK for 25 Å separation while E represents the corresponding mass center of TK after a 145° rotation from position D. c. The matrix represents the total energy matrix of JAK1 with TK at certain positions, which is obtained by the transformations outlined in b. d. The possible position of TK after each rotation.

5.2.3.2 Analysis of the Smoothest Path

After testing at different separation sites, we can draw the energy map of JAK1 with TK located at different positions, and then calculate the possible path for activation. An energy matrix$_{30X17}$ ($M_{30\times17}$) (Figure 1c) is built based on the separation between 9 to 25 Å and the rotation from 0° to 145°, where rows represent the rotation degree while columns represent the separation distance. We are looking for a path from $M_{[1,1]}$ (After 9 Å separation of TK from the inhibited state) to $M_{[30,1]}$ (activated state). The total energy change from $M_{[1,1]}$ to $M_{[30,1]}$ is known *a priori*, meaning that minimal energy consumption cannot be a criterion. Because of this, we tried to find the smoothest path for TK to move from $M_{[1,1]}$ to $M_{[30,1]}$. The smoothest path means the energy consumption is even among steps during the activation process. Each step represents the TK

76

rotation for 5° in our simulations. Here we used the variance of the differences between the two steps to represent the level of smoothness (Eq. 7.1).

$$\sigma^2 = \frac{\sum_{i=1}^{n}(\Delta E_{i,i+1} - \overline{\Delta E})^2}{n-1} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (7.1)$$

where the σ^2 represents the variance, $\Delta E_{i,i+1}$ represents the energy differences between i^{th} position and $(i+1)^{th}$ position. n equal to 29, because of the rotation is via each 5° rotation/step for 145° rotations to reach the activated state. $\overline{\Delta E}$ represents the mean of all differences (Eq. 7.2).

$$\overline{\Delta E} = \frac{\sum_{i=1}^{n}(\Delta E_{i,i+1})}{n} = \frac{\sum_{i=1}^{n}(E_{i+1} - E_i)}{n} = \frac{(E_{30} - E_1)}{n} \dots\dots\dots\dots\dots\dots\dots\dots (7.2)$$

where the E_i represents the energy at i^{th} position. E_1 is the value of M[1,1] and E_{30} is the value of M[30,1]. Because $\overline{\Delta E}$ is a constant, equation 1 simplifies to equation 7.3.

$$\sigma^2 = \frac{\sum_{i=1}^{n} \Delta E_i^2}{28} + a \sum_{i=1}^{n} \Delta E_i + b \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (7.3)$$

where a and b are constants. Since the ΔE_i is the solo contributor to variance, if use $(\Delta E_i - \overline{\Delta E})^2$ as the "distance" of steps, the variance represents the length of the path and the shortest path can be easily solved by dynamic programming[246]. For each step (rotation), TK can move slightly closer or farther from PK. Here, we set each step to follow two rules:

1. M[n,m] can reach M[n+1,(m-2,m+3)], (n<29, m<=17) (Fig. 5.1d).

2. The initial position is M[1,1] and the last position should be M[30,1].

5.3 Results

5.3.1 The Inhibited Full-Length JAK1.

The surface of inhibited full-length JAK1 has a pentagram-like shape (Fig. 2a), where the TK is tightly bound by both FERM and PK. After separating and rotating of FERM and TK by 90°, we can observe the left and right surface of TK are negatively and positively charged, respectively. Correspondingly, the left end of FERM is positively charged while the middle right is negatively charged (Fig. 5.2b). The electrostatic attraction between the FERM and TK domains plays a vital part in securing their binding. Additionally, an analysis of hydrogen bonds reveals the presence of many formed between the FERM and TK domains. E354 (~160%) and E355 (~180%) contributed the most hydrogen bonding by interacting with K924, Y999, K1026, A1027, and Y1035 (Fig. 5.2c). E354 is located in the fatty acid binding sites for Ruxolitinib[247], an inhibitor used to treat myelofibrosis. Interaction between Ruxolitinib and E354 may prevent the release of TK, making it harder for the JAK to be activated. Additionally, E362 and K390 are also important residues with high occupancies (79.21% and 58.42% respectively) of hydrogen bonds. These bonds play a significant role in strengthening the interaction between the two domains, further contributing to the stability of the protein. They could be the potential targets for inhibitor design.

The interactions between PK and TK are depicted in Fig. 5.2d, where PK and TK have been rotated by 90° to display the distribution of electrostatic potential on their surfaces. The right part of TK is negatively charged and the corresponding left part in PK is positively charged. The inhibition of TK can be explained, at least in part, by these electrostatic attractions. To be more specific, hydrogen bonds are analyzed to uncover the contribution of residues pairs. The residue R724 is the most significant contributor to hydrogen bonds, forming bonds with both E897 and D899. This observation has been previously noted in studies that examine the inhibited interfaces

between the JAK2 PK-TK dimer (R744 in JAK2)[239]. The residue E668 also makes a significant contribution to hydrogen bonding, with an 82.18% occupancy in its hydrogen bond to R893. This observation has been previously seen in previous studies (E688 in JAK2)[239]. A new hydrogen bond was discovered in our MD simulations. The bond involves the residues Y605 and E890, with an occupancy of 61.39%. Phosphorylation is the initial step for JAK activation, where tyrosine phosphorylation has been proven to active most JAKs[218, 248, 249]. Y605 could be a possible phosphorylation site for active JAK1 activation.
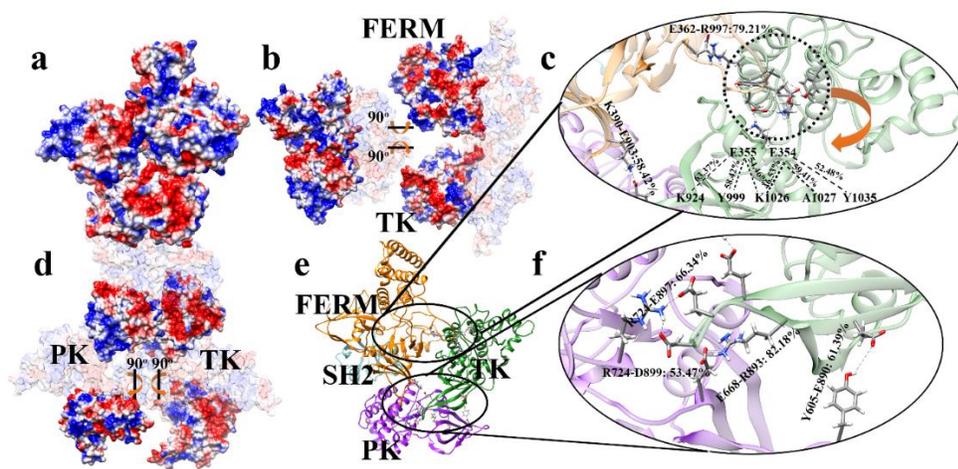
Figure 5.2: The electrostatic surface and hydrogen bonds of inhibited full-length JAK1. a. The electrostatic surface of full-length JAK1 structure. b. The electrostatic surface of the binding interface between FERM and TK (FERM and TK are oppositely rotated 90°) to show the interface. c. The high-occupancy hydrogen bonds between FERM and TK. d. The electrostatic surface of the binding interface between PK and TK (PK and TK are oppositely rotated 90°). e. The structural representation and high-occupancy hydrogen bonds for inhibited full-length JAK1. f. The high-occupancy hydrogen bonds between PK and TK.

5.3.2 The Activated Full-Length JAK1.

In the activated state, TK is no longer connected to FERM, instead, it only interacts with PK. The interfaces of the FERM-SH2-PK complex and TK are displayed in Fig. 5.3a. PK is positioned on the right of the FERM-SH2-PK complex, which is characterized by having mostly negative electrostatic charges. Correspondingly, the center of the TK interface is positively

charged, and it is electrostatically attracted to the negatively charged PK domain. The hydrogen bonds were analyzed to show the individual contributions of residues (Fig. 5.3bc). R930-E801 possess over 82.18% occupancy. Other glutamate residues like E803 and E807 also formed high-occupancy hydrogen bonds with Y933 and K941, respectively. This shows the importance of glutamate in PK at the activated state and provide possible tyrosine residues Y933, phosphorylation on which may activate JAK1. Positive charges provided by R930 and K941 play a role in the electrostatic attraction to the negatively charged surface of PK, which occurs due to the presence of glutamate residues in the protein. The formation of hydrogen bonds between PK and TK is facilitated by close proximity between the two domains—typically when the distance between them is around 3.2 Å[250, 251]. This interaction contributes to the stability of the activated JAK1 structure.
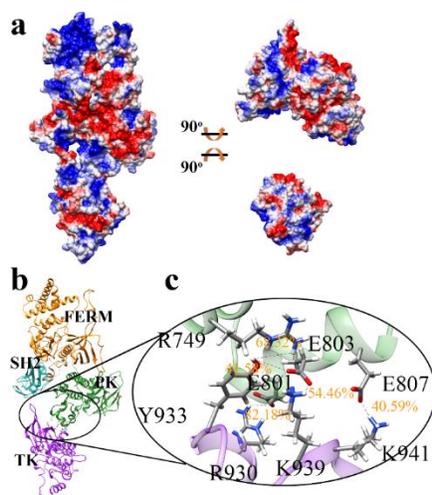
Figure 2: The electrostatic surface and hydrogen bonds of activated full-length JAK1. a. The electrostatic surface of the full-length JAK1 structure. Two regions (FERM-SH2-PK and TK) are oppositely rotated 90º to show the interface. b. The high-occupancy hydrogen bonds and structural representation of activated full-length JAK1. c. The high-occupancy hydrogen bonds between PK and TK.

5.3.3 Using Dynamic Programming to Find the Smoothest Path of TK Translocation in JAK1 Activation.

The activation of JAK1 primarily involves the translocation of the TK from an inhibited position to an activated position. The same interface of TK facing PK at both positions inspired us that JAK1 activation can be achieved by rolling TK around PK. Since efficiency is a fundamental principle for many biological processes, possible minimal paths were initially designed for TK translocation. This includes separations of TK and PK by 9 Å and rotation of TK around PK by 145°. TK was first separated from PK by 9 Å and then rotated 145° to show the total energy change (Fig. 5.4a). TK encounters a high energy barrier around 60°-80°, making it difficult for the protein to continue rotating the full 145°. To better illustrate these energy changes, TK was further separated from PK by increments of 1 Å, up to 25 Å, to observe when the energy barrier would diminish. After a 15 Å separation, the rotation did not encounter a significant energy barrier and after a 20 Å separation, there was no energy barrier at all (Fig. 4a). These energy calculations provide a JAK1 energy map with TK at different locations, shown in Fig. 5.4b. It is evident that a region with high energy is present in the range of 9-12 Å and 60°-80°.

To avoid the energy barrier, we attempt to find the smoothest path for TK translocation. The smoothest path refers to the path with the lowest variance of steps. If the energy of TK translocation fluctuates markedly and frequently, the variance of is large. Each 5° rotation is considered a step. During the rotation, both separation and approach of TK can occur, leading to multiple possible paths. This means that TK moves from the point $M_{(0,9)}$ to $M_{(145,9)}$ in Fig. 5.4b. The problem of finding the minimum variance cannot be solved directly through dynamic programming because the path lacks a recursive structure. To resolve this, a new contributor

"$(\Delta E_{i,i+1} - \overline{\Delta E})^2$" was introduced to calculate each step's contribution to the variance. This allows the creation of a recursive structure for dynamic programming.

After calculations, the smoothest path is found and shown in Fig. 5.4b. Clearly, during the activation, TK first separates from PK for 17 Å before overcoming the 60°-80° rotational energy barrier. Then it approaches PK and continues rolling until reaching the activated site. We also provide the energy heatmaps and smoothest paths of P733L and S703I mutations, shown in Figure 4c and 4d, respectively. The smoothest path analysis suggests that in JAK1-P733L, TK remains near PK until 55° of rotation, after which it suddenly separates by about 6 Å during the next 15° rotation. Subsequently, TK continues to rotate for an additional 40° before abruptly approaching PK, followed by another 20 degrees of rotation, at which point TK ultimately reaches the activated site. While the energy path is smooth, the spatial pathway is not; it is interrupted by sudden separations and approaches. The ideal path should be smooth energetically and spatially— especially for the initial steps. This implies the P733L mutation hampers JAK1 activation and impairs transphosphorylation, resulting in subsequent immunodeficiency[212]. The smoothest path for TK in JAK-S703I involved an initial separation of 16 Å at a 60° rotation, followed by a sudden approach over the next 50° rotation. From there, it smoothly rotated to the activated site (Fig. 5.4d). Compared with WT type JAK1, the initial separation combined with rotations is much slower while the final approach step is much quicker. The mutation of S703I decreases the difficulty of activating JAK1, leading to constitute activation[213]. The smoothest paths of TKs in JAK1s with and without mutations are shown in Fig. 5.4e. The mutation of S703I caused spatially smooth separation and rotation initially but there is quick approach after bypassing the high energy area. The JAK1-S703I mutation created an "energy stair" for TK to navigate around the high energy zone gradually. In contrast, the JAK1-P733L mutation did not allow gradual separation of

TK to bypass the high energy zone. Moreover, these mutations caused different initial total energies (Fig. 5.4g). Compared with WT JAK1, the P733L mutation reduced the initial total energy, making the inhibited state more stable. Conversely, S703I increased the total energy of the inhibited state, reducing the stability of the inhibited JAK1 structure.

The energies of smoothest paths of JAK1s are shown in Fig. 5.4f. The total energy of JAK1 in the activated sate is lower than that of the inhibited state. It infers to that the activation of JAKs could be thermodynamically spontaneous. The reason why it should be activated by cytosine signaling is necessary is because of the initial separation should cross an energy barrier and enter the activated state. This energy barrier is the reason the inhibited state is also known as the "autoinhibited state", and is described in many articles[223, 224, 231, 237, 239]. The VF mutation or the tyrosine phosphorylation which can active the JAKs, could serve as the key to overcome or diminish the energy barrier. For the P733L mutation, which shows a low energy barrier at the initial steps but is also known to weaken transphosphorylation[212] the results could be explained by a spatial lock at the step 17, where TK is at a 9 Å separation and 40° rotation. The next step in the path is to be separated by 6 Å at a 15° rotation, which is difficult to achieve. Thus, the JAK1-P733L structure is locked in this state where the activate site of TK is still covered.
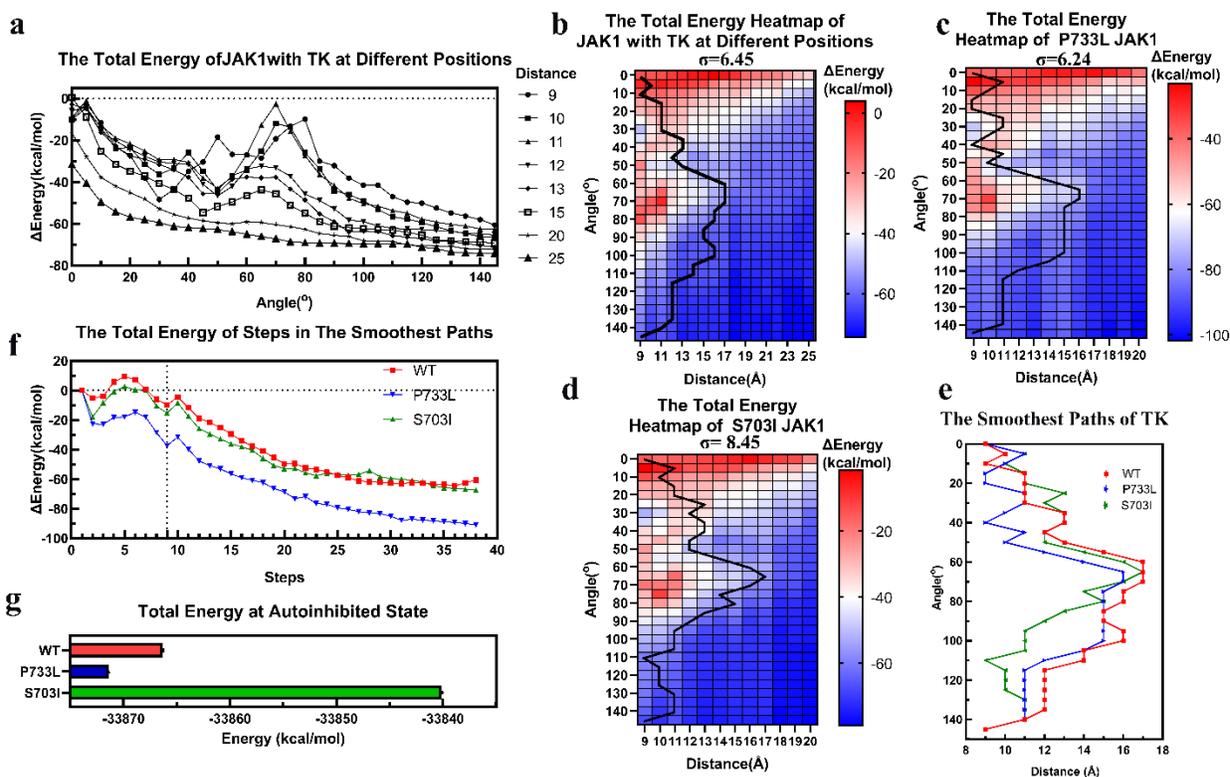
Figure 5.4: The energy analysis of the activation process of JAK1. a. The total energy of JAK1 during the rotation and separation of TK. bcd. The total energy heatmap of JAK1 (WT, P733L, S703I) with TK at different positions, where the black lines represent the smoothest paths for the corresponding energy heatmap. e. A smoothest path comparison among WT, P733L, and S703I. f. The total energy for each step along the smoothest paths for WT, P733L, and S703I JAK1. g. The total energy at the autoinhibited state. The energy shown in the figure represents the relative total energy, which is the difference between the current state and the total energy in the inhibited state (f).

5.4 Discussion and Conclusion

Based on the inhibited PK-TK dimer[224] and activated full-length JAK dimer[223], we successfully built the inhibited full-length JAK1 and activated full-length JAK1 to study the differences between the two states and the structural transitions of JAKs. In the inhibited state, TK is strongly locked in the cavity formed by PK and FERM. The glutamine residues in FERM contributed the most hydrogen bonds, while the arginine residues in PK contributed the most hydrogen bonds. Typically, the FERM interface facing the tyrosine kinase (TK) is negatively charged, while the TK interface facing the FERM domain is positively charged. This charge

84

distribution is reversed in the PK-TK interfaces, with the PK domain interface being positively charged and the TK domain interface being negatively charged. In the activated state, only PK interacts with TK. Less hydrogen bonds occur in this state, contributing to high freedom of movement for TK.

Comparing the inhibited and activated JAK1 structures, the interfaces of TK are distinct, suggesting that TK rotates around PK to reach the activated site. To avoid steric clashes during rotation, initial 9 Å separation of TK is necessary for translocation. After the initial separation, the TK was further separated from PK by increments of 1 Å, up to 25 Å, and then rotated after different separations. The total energies of JAK1 with TK at different positions were calculated and inputted into a matrix suited for dynamic programming where the goal is to find the energetically smoothest path. This method was also applied to two famous mutations, which are the P733L (negative regulatory)[212], and S703I (positive regulatory) mutations in JAK1. We found the energetically smoothest path of P733L is not spatially smooth, especially when bypassing one high energy zone. In contrast, S703I-JAK1 creates an "energy stair" spatially for TK to navigate around the high energy zone gradually. These finding explain the reasons underlying the constitutive activation of S703I-JAK1 and the impaired transphosphorylation of P733L-JAK1. We note, however, the increased initial total energy in S703I-JAK1 may have also contributed to activation.

In this work, the electrostatic features of JAK1 in both inhibited and activated states are analyzed, providing a potential pathway for TK translocation in JAK1 activation. By comparing two mutations, this study reveals the importance of separational and rotational movements, which could be crucial for JAK activation. Furthermore, energy analysis suggests that the inhibited state is in a low energy state where the entire structure is stable. However, the presence of VF mutations[224, 236, 240] or tyrosine phosphorylation[218, 236] increases the energy and forces

TK to leave the cavity and rotate around the PK domain to reach the activated site, ultimately

leading to JAK1 activation.

BIBLIOGRAPHY

1.   Sagui, C. and T.A. Darden, *Molecular dynamics simulations of biomolecules: long-range electrostatic effects.* Annual review of biophysics and biomolecular structure, 1999. **28**(1): p. 155-179.

2.   Peng, Y., et al., *DFMD: fast and effective DelPhiForce steered molecular dynamics approach to model ligand approach toward a receptor: application to spermine synthase enzyme.* Frontiers in molecular biosciences, 2019. **6**: p. 74.

3.   Kumar, S. and R. Nussinov, *Salt bridge stability in monomeric proteins.* Journal of molecular biology, 1999. **293**(5): p. 1241-1255.

4.   Sinha, N. and S.J. Smith-Gill, *Electrostatics in protein binding and function.* Current Protein and Peptide Science, 2002. **3**(6): p. 601-614.

5.   Sun, S., et al., *Molecular mechanisms of cardiac actomyosin transforming from rigor state to post-rigor state.* The Journal of Chemical Physics, 2022. **156**(3): p. 035101.

6.   Xie, Y., et al., *The pH Effects on SARS-CoV and SARS-CoV-2 Spike Proteins in the Process of Binding to hACE2.* Pathogens, 2022. **11**(2): p. 238.

7.   Talley, K. and E. Alexov, *On the pH‐optimum of activity and stability of proteins.* Proteins: Structure, Function, and Bioinformatics, 2010. **78**(12): p. 2699-2706.

8.   Koirala, M., et al., *Computational Investigation of the pH Dependence of Stability of Melanosome Proteins: Implication for Melanosome formation and Disease.* International journal of molecular sciences, 2021. **22**(15): p. 8273.

9.   York, D.M., T.A. Darden, and L.G. Pedersen, *The effect of long‐range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods.* The Journal of chemical physics, 1993. **99**(10): p. 8345-8348.

10.   Koirala, M. and E. Alexov, *Ab-initio binding of barnase–barstar with DelPhiForce steered Molecular Dynamics (DFMD) approach.* Journal of Theoretical and Computational Chemistry, 2020. **19**(04): p. 2050016.

11.   Shashikala, H.M., A. Chakravorty, and E. Alexov, *Modeling electrostatic force in protein-protein recognition.* Frontiers in Molecular Biosciences, 2019. **6**: p. 94.

12.   Li, L., J. Alper, and E. Alexov, *Cytoplasmic dynein binding, run length, and velocity are guided by long-range electrostatic interactions.* Scientific reports, 2016. **6**(1): p. 1-12.

13.   Norberg, J. and L. Nilsson, *On the truncation of long-range electrostatic interactions in DNA.* Biophysical journal, 2000. **79**(3): p. 1537-1553.

14.   Kodchakorn, K., T. Chokepaichitkool, and P. Kongtawelert, *Mutational scanning of spike RBD protein for enhanced ACE2 affinity emerging Southeast Asia in the late transmission phase.* Scientific reports, 2022. **12**(1): p. 1-9.

15.   Dushanan, R., et al., *An In-Silico Approach to Evaluate the Inhibitory Potency of Selected Hydroxamic Acid Derivatives on Zinc-Dependent Histone Deacetylase Enzyme.* Journal of Computational Biophysics and Chemistry, 2021. **20**(06): p. 603-618.

16.   King, K.M., et al., *Impact of electronic polarization on preformed, β-strand rich homogenous and heterogeneous amyloid oligomers.* 2021.

17.   Li, T., et al., *Changes in structure and flexibility of p53 TAD2 upon binding to p300 Taz2.* Journal of Theoretical and Computational Chemistry, 2020. **19**(04): p. 2040007.

18.   Peng, Y. and E. Alexov, *Computational investigation of proton transfer, p K a shifts and p H‐optimum of protein‐DNA and protein‐RNA complexes.* Proteins: Structure, Function, and Bioinformatics, 2017. **85**(2): p. 282-295.

19.   Arthur, E.J. and C.L. Brooks III, *Efficient implementation of constant pH molecular dynamics on modern graphics processors.* Journal of computational chemistry, 2016. **37**(24): p. 2171-2180.

20.   Harris, R.C. and J. Shen, *GPU-accelerated implementation of continuous constant pH molecular dynamics in amber: p K a predictions with single-pH simulations.* Journal of chemical information and modeling, 2019. **59**(11): p. 4821-4832.

21.   Vila-Viçosa, D., et al., *A pH replica exchange scheme in the stochastic titration constant-pH MD method.* Journal of Chemical Theory and Computation, 2019. **15**(5): p. 3108-3116.

22.   Meng, Y. and A.E. Roitberg, *Constant pH replica exchange molecular dynamics in biomolecules using a discrete protonation model.* Journal of chemical theory and computation, 2010. **6**(4): p. 1401-1412.

23.   Oliveira, N.F. and M. Machuqueiro, *Novel US-CpHMD Protocol to Study the Protonation-Dependent Mechanism of the ATP/ADP Carrier.* Journal of Chemical Information and Modeling, 2022. **62**(10): p. 2550-2560.

24.   Reilley, D.J., et al., *Titr-DMD—A Rapid, Coarse-Grained Quasi-All-Atom Constant pH Molecular Dynamics Framework.* Journal of Chemical Theory and Computation, 2021. **17**(7): p. 4538-4549.

25.   Baruah, I. and G. Borgohain, *Binding interaction of a potential statin with β-lactoglobulin: An in silico approach.* Journal of Molecular Graphics and Modelling, 2022. **111**: p. 108077.

26.   Teng, S., et al., *Modeling effects of human single nucleotide polymorphisms on protein-protein interactions.* Biophysical journal, 2009. **96**(6): p. 2178-2188.

27.   Stefl, S., et al., *Molecular mechanisms of disease-causing missense mutations.* Journal of molecular biology, 2013. **425**(21): p. 3919-3936.

28.   Zhang, Z., et al., *Rational design of small-molecule stabilizers of spermine synthase dimer by virtual screening and free energy-based approach.* PLoS One, 2014. **9**(10): p. e110884.

29.   Zhang, Z., et al., *A rational free energy-based approach to understanding and targeting disease-causing missense mutations.* Journal of the American Medical Informatics Association, 2013. **20**(4): p. 643-651.

30.   Peng, Y., et al., *Revealing the effects of missense mutations causing Snyder-Robinson syndrome on the stability and dimerization of spermine synthase.* International Journal of Molecular Sciences, 2016. **17**(1): p. 77.

31.   Zhang, Z., et al., *A Y328C missense mutation in spermine synthase causes a mild form of Snyder–Robinson syndrome.* Human molecular genetics, 2013. **22**(18): p. 3789-3797.

32.   Zhang, Z., et al., *In silico and in vitro investigations of the mutability of disease-causing missense mutation sites in spermine synthase.* PloS one, 2011. **6**(5): p. e20373.

33.   Zhang, Z., et al., *Computational analysis of missense mutations causing Snyder‐Robinson syndrome.* Human mutation, 2010. **31**(9): p. 1043-1049.

34.   Andac, C.A., et al., *In-silico pharmacokinetic and affinity studies of piperazine/morpholine substituted quinolines in complex with GAK as promising anti-HCV agent.* Journal of Computational Biophysics and Chemistry, 2021. **20**(08): p. 869-879.

35.   Xue, W., et al., *Molecular modeling study on the resistance mechanism of HCV NS3/4A serine protease mutants R155K, A156V and D168A to TMC435.* Antiviral research, 2012. **93**(1): p. 126-137.

36.   Sharafdini, R. and H. Mosaddeghi, *Inhibition of Insulin Amyloid Fibrillation by Salvianolic Acids and Calix [n] arenes: Molecular Docking Insight.* Journal of Computational Biophysics and Chemistry, 2021. **20**(05): p. 539-555.

37.   da Fonseca, A.M., et al., *Molecular Docking Study of Antibiotics, Anti-Inflammatory Drugs and [Eu (TTA) 3·AMX] Complex as COVID-19 Biomarker through Interaction of Its Main Protease (Mpro).* Journal of Computational Biophysics and Chemistry, 2021. **20**(04): p. 405-415.

38.   Alibakhshi, A., et al., *Computational design of a potential therapeutic peptide against spike protein of SARS-CoV-2.* Journal of Computational Biophysics and Chemistry, 2021. **20**(04): p. 337-346.

39.   Llamas-García, M.L., et al., *Improved Stability of Human CGI-58 Induced by Phosphomimetic S237E Mutation.* ACS omega, 2022. **7**(15): p. 12643-12653.

40.   Xian, Y., et al., *StructureMan: A Structure Manipulation Tool to Study Large Scale Biomolecular Interactions.* Frontiers in molecular biosciences, 2021. **7**: p. 476.

41.   Yang, J. and Y. Zhang, *I-TASSER server: new development for protein structure and function predictions.* Nucleic acids research, 2015. **43**(W1): p. W174-W181.

42.   Buckle, A.M., G. Schreiber, and A.R. Fersht, *Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-. ANG. resolution.* Biochemistry, 1994. **33**(30): p. 8878-8889.

43.   Itoh, T., et al., *Structure of 8Sα globulin, the major seed storage protein of mung bean.* Acta crystallographica section D: Biological crystallography, 2006. **62**(7): p. 824-832.

44.   Schwede, T., et al., *SWISS-MODEL: an automated protein homology-modeling server.* Nucleic acids research, 2003. **31**(13): p. 3381-3385.

45.   Wang, Q., et al., *Structural and functional basis of SARS-CoV-2 entry by using human ACE2.* Cell, 2020. **181**(4): p. 894-904. e9.

46.   Khan, S., et al., *Revealing the Dimeric Crystal and Solution Structure of β-Lactoglobulin at pH 4 and Its pH and Salt Dependent Monomer–Dimer Equilibrium.* Biomacromolecules, 2018. **19**(7): p. 2905-2912.

47.   Piana, S., et al., *Evaluating the effects of cutoffs and treatment of long-range electrostatics in protein folding simulations.* PLoS One, 2012. **7**(6): p. e39918.

48.   Li, L., et al., *DelPhi: a comprehensive suite for DelPhi software and associated resources.* BMC biophysics, 2012. **5**(1): p. 1-11.

49.   Sun, S., et al., *HIT web server: A hybrid method to improve electrostatic calculations for biomolecules.* Computational and Structural Biotechnology Journal, 2022.

50.   Sun, S., et al., *Hybrid method for representing ions in implicit solvation calculations.* Computational and structural biotechnology journal, 2021. **19**: p. 801-811.

51.   Sarkar, S., et al., *DelPhi Web Server: A comprehensive online suite for electrostatic calculations of biological macromolecules and their complexes.* Communications in computational physics, 2013. **13**(1): p. 269-284.

52.   Li, L., A. Chakravorty, and E. Alexov, *DelPhiForce, a tool for electrostatic force calculations: Applications to macromolecular binding.* Journal of computational chemistry, 2017. **38**(9): p. 584-593.

53.   Li, L., J. Alper, and E. Alexov, *Multiscale method for modeling binding phenomena involving large objects: application to kinesin motor domains motion along microtubules.* Scientific reports, 2016. **6**(1): p. 1-12.

54.   Tajielyato, N., et al., *E-hooks provide guidance and a soft landing for the microtubule binding domain of dynein.* Scientific reports, 2018. **8**(1): p. 1-10.

55.   Pabbathi, A., et al., *Long-range electrostatic interactions significantly modulate the affinity of dynein for microtubules.* Biophysical Journal, 2022. **121**(9): p. 1715-1726.

56.   Li, L., et al., *DelPhiForce web server: electrostatic forces and energy calculations and visualization.* Bioinformatics, 2017. **33**(22): p. 3661-3663.

57.   Salas, G.G.S., et al., *Using computational approaches to study dengue virus capsid assembly.* Computational and Mathematical Biophysics, 2019. **7**(1): p. 64-72.

58.   Xian, Y., et al., *The roles of electrostatic interactions in capsid assembly mechanisms of giant viruses.* International journal of molecular sciences, 2019. **20**(8): p. 1876.

59.   Nath, B.K., et al., *Structural perspectives of beak and feather disease virus and porcine circovirus proteins.* Viral Immunology, 2021. **34**(1): p. 49-59.

60.   Ganguly, D., et al., *Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins.* Journal of molecular biology, 2012. **422**(5): p. 674-684.

61.   Xue, Y., et al., *Role of electrostatic interactions in binding of peptides and intrinsically disordered proteins to their folded targets. 1. NMR and MD characterization of the complex between the c-Crk N-SH3 domain and the peptide Sos.* Biochemistry, 2014. **53**(41): p. 6473-6495.

62.   Mendis, J., E. Kaya, and T.G. Kucukkal, *Identification of Hotspot Residues in Binding of SARS-CoV-2 Spike and Human ACE2 Proteins.* Journal of Computational Biophysics and Chemistry, 2021. **20**(07): p. 729-739.

63.   Nandi, T. and S.R.K. Ainavarapu, *Native Salt Bridges Are a Key Regulator of Ubiquitin's Mechanical Stability.* The Journal of Physical Chemistry B, 2022.

64.   Ji, S., W. Yang, and W. Yu, *Understanding the role of the CB1 toggle switch in interaction networks using molecular dynamics simulation.* Scientific reports, 2021. **11**(1): p. 1-15.

65.   Tripathi, N., et al., *Structural insights into the catalytic mechanism of granzyme B upon substrate and inhibitor binding.* Journal of Molecular Graphics and Modelling, 2022. **114**: p. 108167.

66.   Rauh, O., et al., *Role of Ion Distribution and Energy Barriers for Concerted Motion of Subunits in Selectivity Filter Gating of a K+ Channel.* Journal of Molecular Biology, 2022. **434**(9): p. 167522.

67.   Yao, S., et al., *Investigating the Broad Matrix-Gate Network in the Mitochondrial ADP/ATP Carrier through Molecular Dynamics Simulations.* Molecules, 2022. **27**(3): p. 1071.

68.   Shahoei, R., et al., *Molecular Modeling of ABHD5 Structure and Ligand Recognition.* Frontiers in Molecular Biosciences, 2022: p. 692.

69.   Rollins, Z.A., R. Faller, and S.C. George, *Using molecular dynamics simulations to interrogate T cell receptor non-equilibrium kinetics.* Computational and Structural Biotechnology Journal, 2022. **20**: p. 2124-2133.

70.   Ouyang, Q., et al., *Theoretical investigation on the reaction mechanism of UTP cyclohydrolase.* Physical Chemistry Chemical Physics, 2022.

71.   Méndez, E., P.E. Videla, and D. Laria, *Equilibrium and Dynamical Characteristics of Hydrogen Bond Bifurcations in Water–Water and Water–Ammonia Dimers: A Path Integral Molecular Dynamics Study.* The Journal of Physical Chemistry A, 2022.

72.   Yun, Y., R.Z. Khaliullin, and Y. Jung, *Correlated Local Fluctuations in the Hydrogen Bond Network of Liquid Water.* Journal of the American Chemical Society, 2022.

73.   Koirala, M. and E. Alexov, *Computational chemistry methods to investigate the effects caused by DNA variants linked with disease.* Journal of Theoretical and Computational Chemistry, 2020. **19**(06): p. 1930001.

74.   Liu, W., et al., *Insights into the binding of dorzagliatin with glucokinase: A molecular dynamics simulation.* Journal of Theoretical and Computational Chemistry, 2020. **19**(07): p. 2050027.

75.   Bai, H., et al., *Simulative Analysis of a Family of DNA Tetrahedrons Produced by Changing the Twisting Number of Each Double Helix.* Journal of Computational Biophysics and Chemistry, 2021. **20**(05): p. 529-537.

76.   Zhou, T., et al., *Computational study of switching mechanism in add A-riboswitch.* Journal of Theoretical and Computational Chemistry, 2020. **19**(03): p. 2040001.

77.   Garcia-Moreno, B., *Adaptations of proteins to cellular and subcellular pH.* Journal of biology, 2009. **8**(11): p. 1-4.

78.   Alexov, E., *Special Issue on Computational Chemistry Methods to Predict pKa's of Ionizable Groups in Proteins, RNAs, DNAs and Small Molecules*. 2021, World Scientific. p. 109-110.

79.   Mitra, R.C., Z. Zhang, and E. Alexov, *In silico modeling of pH‐optimum of protein‐protein binding.* Proteins: Structure, Function, and Bioinformatics, 2011. **79**(3): p. 925-936.

80.   Arthur, E.J. and C.L. Brooks, 3rd, *Efficient implementation of constant pH molecular dynamics on modern graphics processors.* J Comput Chem, 2016. **37**(24): p. 2171-80.

81.   Harris, R.C. and J. Shen, *GPU-Accelerated Implementation of Continuous Constant pH Molecular Dynamics in Amber: pK(a) Predictions with Single-pH Simulations.* J Chem Inf Model, 2019. **59**(11): p. 4821-4832.

82.   Vila-Viçosa, D., et al., *A pH Replica Exchange Scheme in the Stochastic Titration Constant-pH MD Method.* J Chem Theory Comput, 2019. **15**(5): p. 3108-3116.

83.   Swails, J.M., D.M. York, and A.E. Roitberg, *Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation.* J Chem Theory Comput, 2014. **10**(3): p. 1341-1352.

84.   Oliveira, N.F.B. and M. Machuqueiro, *Novel US-CpHMD Protocol to Study the Protonation-Dependent Mechanism of the ATP/ADP Carrier.* J Chem Inf Model, 2022. **62**(10): p. 2550-2560.

85.   Reilley, D.J., et al., *Titr-DMD-A Rapid, Coarse-Grained Quasi-All-Atom Constant pH Molecular Dynamics Framework.* J Chem Theory Comput, 2021. **17**(7): p. 4538-4549.

86.   Baruah, I. and G. Borgohain, *Binding interaction of a potential statin with β-lactoglobulin: An in silico approach.* J Mol Graph Model, 2022. **111**: p. 108077.

87.   Ma, S., J.A. Henderson, and J. Shen, *Exploring the pH-Dependent Structure-Dynamics-Function Relationship of Human Renin.* J Chem Inf Model, 2021. **61**(1): p. 400-407.

88.   Liu, R., et al., *Profiling MAP kinase cysteines for targeted covalent inhibitor design.* RSC Med Chem, 2022. **13**(1): p. 54-63.

89.   Santos, H.A., et al., *Constant-pH MD Simulations of DMPA/DMPC Lipid Bilayers.* J Chem Theory Comput, 2015. **11**(12): p. 5973-9.

90.   Teixeira, V.H., et al., *pK(a) Values of Titrable Amino Acids at the Water/Membrane Interface.* J Chem Theory Comput, 2016. **12**(3): p. 930-4.

91.   Khaniya, U., et al., *Characterizing Protein Protonation Microstates Using Monte Carlo Sampling.* J Phys Chem B, 2022. **126**(13): p. 2476-2485.

92.   Pahari, S., et al., *DelPhiPKa: Including salt in the calculations and enabling polar residues to titrate.* Proteins, 2018. **86**(12): p. 1277-1283.

93.   Suhail, M., *The Target Determination and the Mechanism of Action of Chiral-Antimalarial Drugs: A Docking Approach.* Journal of Computational Biophysics and Chemistry, 2021. **20**(05): p. 501-516.

94.   Chen, J., et al., *Computing Protein pKas Using the TABI Poisson–Boltzmann Solver.* Journal of Computational Biophysics and Chemistry, 2021. **20**(02): p. 175-187.

95.   Peng, Y., et al., *pH-dependent interactions of Apolipophorin-III with a lipid disk.* Journal of Computational Biophysics and Chemistry, 2021. **20**(02): p. 153-164.

96.   Sargolzaei, M., H. Nikoofard, and M.N. Jorabchi, *Quantum-chemical ab initio study of side chain pKa of linear and cyclic lysine dipeptides.* Journal of Computational Biophysics and Chemistry, 2021. **20**(02): p. 131-139.

97.   Lopez-Hernandez, A.E., et al., *The electrostatic features of dengue virus capsid assembly.* Journal of Computational Biophysics and Chemistry, 2021. **20**(02): p. 201-207.

98.   Zhang, Y., et al., *Characterizing the water wire in the Gramicidin channel found by Monte Carlo sampling using continuum electrostatics and in molecular dynamics trajectories with conventional or polarizable force fields.* Journal of Computational Biophysics and Chemistry, 2021. **20**(02): p. 111-130.

99.   MacKenzie, D.W.S., et al., *A fine balance of hydrophobic-electrostatic communication pathways in a pH-switching protein.* Proc Natl Acad Sci U S A, 2022. **119**(26): p. e2119686119.

100.   Petukh, M., T.G. Kucukkal, and E. Alexov, *On human disease-causing amino acid variants: statistical study of sequence and structural patterns.* Hum Mutat, 2015. **36**(5): p. 524-534.

101.   Peng, Y. and E. Alexov, *Investigating the linkage between disease-causing amino acid variants and their effect on protein stability and binding.* Proteins, 2016. **84**(2): p. 232-9.

102.   Teng, S., et al., *Modeling effects of human single nucleotide polymorphisms on protein-protein interactions.* Biophys J, 2009. **96**(6): p. 2178-88.

103.   Stefl, S., et al., *Molecular mechanisms of disease-causing missense mutations.* J Mol Biol, 2013. **425**(21): p. 3919-36.

104.   Zhang, Z., et al., *A rational free energy-based approach to understanding and targeting disease-causing missense mutations.* J Am Med Inform Assoc, 2013. **20**(4): p. 643-51.

105.   Wang, L.-Q., et al., *Genetic prion disease–related mutation E196K displays a novel amyloid fibril structure revealed by cryo-EM.* Science advances, 2021. **7**(37): p. eabg9676.

106.   Zhou, T., et al., *Cryo-EM structures of SARS-CoV-2 spike without and with ACE2 reveal a pH-dependent switch to mediate endosomal positioning of receptor-binding domains.* Cell host & microbe, 2020. **28**(6): p. 867-879. e5.

107.   Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis.* Journal of computational chemistry, 2004. **25**(13): p. 1605-1612.

108.   Li, L., et al., *Forces and Disease: Electrostatic force differences caused by mutations in kinesin motor domains can distinguish between disease-causing and non-disease-causing mutations.* Scientific reports, 2017. **7**(1): p. 1-12.

109.   Karki, C., et al., *A computational model of ESAT-6 complex in membrane.* Journal of Theoretical and Computational Chemistry, 2020. **19**(03): p. 2040002.

110.   Sun, S., et al., *Molecular mechanisms of cardiac actomyosin transforming from rigor state to post-rigor state.* The Journal of Chemical Physics, 2021.

111.   Escobedo, A., et al., *Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor.* Nature communications, 2019. **10**(1): p. 1-11.

112. Mompeán, M., et al., *An amyloid-like pathological conformation of TDP-43 is stabilized by hypercooperative hydrogen bonds.* Frontiers in molecular neuroscience, 2016. **9**: p. 125.

113. Boyer, D.R., et al., *Structures of fibrils formed by α-synuclein hereditary disease mutant H50Q reveal new polymorphs.* Nature structural & molecular biology, 2019. **26**(11): p. 1044-1052.

114. White, K.A., et al., *Cancer-associated arginine-to-histidine mutations confer a gain in pH sensing to mutant proteins.* Sci Signal, 2017. **10**(495).

115. Liu, Y., K.A. White, and D.L. Barber, *Intracellular pH Regulates Cancer and Stem Cell Behaviors: A Protein Dynamics Perspective.* Front Oncol, 2020. **10**: p. 1401.

116. White, K.A., K. Kisor, and D.L. Barber, *Intracellular pH dynamics and charge-changing somatic mutations in cancer.* Cancer Metastasis Rev, 2019. **38**(1-2): p. 17-24.

117. Koirala, M., et al., *Computational Investigation of the pH Dependence of Stability of Melanosome Proteins: Implication for Melanosome formation and Disease.* Int J Mol Sci, 2021. **22**(15).

118. Takano, K., et al., *An X-linked channelopathy with cardiomegaly due to a CLIC2 mutation enhancing ryanodine receptor channel activity.* Human molecular genetics, 2012. **21**(20): p. 4497-4507.

119. Warwicker, J., *A model for pH coupling of the SARS-CoV-2 spike protein open/closed equilibrium.* Brief Bioinform, 2021. **22**(2): p. 1499-1507.

120. Xie, Y., et al., *The pH Effects on SARS-CoV and SARS-CoV-2 Spike Proteins in the Process of Binding to hACE2.* Pathogens, 2022. **11**(2).

121. Lobo, V.R. and J. Warwicker, *Predicted pH-dependent stability of SARS-CoV-2 spike protein trimer from interfacial acidic groups.* Comput Struct Biotechnol J, 2021. **19**: p. 5140-5148.

122. Kucukkal, T.G., et al., *Impact of Rett Syndrome Mutations on MeCP2 MBD Stability.* Biochemistry, 2015. **54**(41): p. 6357-68.

123. Yang, Y., et al., *Binding Analysis of Methyl-CpG Binding Domain of MeCP2 and Rett Syndrome Mutations.* ACS Chem Biol, 2016. **11**(10): p. 2706-2715.

124. Peng, Y., et al., *Revealing the Effects of Missense Mutations Causing Snyder-Robinson Syndrome on the Stability and Dimerization of Spermine Synthase.* Int J Mol Sci, 2016. **17**(1).

125. Zhang, Z., et al., *A Y328C missense mutation in spermine synthase causes a mild form of Snyder-Robinson syndrome.* Hum Mol Genet, 2013. **22**(18): p. 3789-97.

126. Zhang, Z., et al., *Computational analysis of missense mutations causing Snyder-Robinson syndrome.* Hum Mutat, 2010. **31**(9): p. 1043-9.

127. Wassman, C.D., et al., *Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53.* Nat Commun, 2013. **4**: p. 1407.

128. İş, Y.S., *Elucidation of Ligand/Protein Interactions between BCR-ABL Tyrosine Kinase and Some Commercial Anticancer Drugs Via DFT Methods.* Journal of Computational Biophysics and Chemistry, 2021. **20**(04): p. 433-447.

129. Hussain, R., H. Khalid, and M.Q. Fatmi, *Molecular Modeling Approach of Serine Protease NS3-4A Genotype 3a as a Potential Drug Target of Hepatitis C Virus: Homology Modeling and Virtual Screening Study.* Journal of Computational Biophysics and Chemistry, 2021. **20**(06): p. 631-639.

130. Panigrahi, D., *Molecular Docking Analysis of the Phytochemicals from Tinospora Cordifolia as Potential Inhibitor Against Multi Targeted SARS-CoV-2 & Cytokine Storm.* Journal of Computational Biophysics and Chemistry, 2021. **20**(06): p. 559-580.

131. Ghamsari, P.A., M. Samadizadeh, and M. Mirzaei, *Halogenated derivatives of cytidine: Structural analysis and binding affinity.* Journal of Theoretical and Computational Chemistry, 2020. **19**(08): p. 2050033.

132. Qader, S.W., et al., *In vitro Acetylcholinesterase inhibitory activity of polyphenolic compounds identified from Matricaria recutita.* Journal of Theoretical and Computational Chemistry, 2020. **19**(08): p. 2050029.

133. Lilly, L.S. and E. Braunwald, *Braunwald's heart disease: a textbook of cardiovascular medicine.* Vol. 2. 2012: Elsevier Health Sciences.

134. Freitas, P., et al., *The amount of late gadolinium enhancement outperforms current guideline-recommended criteria in the identification of patients with hypertrophic cardiomyopathy at risk of sudden cardiac death.* Journal of Cardiovascular Magnetic Resonance, 2019. **21**(1): p. 1-10.

135. Avanesov, M., et al., *Prediction of the estimated 5-year risk of sudden cardiac death and syncope or non-sustained ventricular tachycardia in patients with hypertrophic cardiomyopathy using late gadolinium enhancement and extracellular volume CMR.* European radiology, 2017. **27**(12): p. 5136-5145.

136. Harper, A.R., et al., *Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity.* Nature Genetics, 2021. **53**(2): p. 135-142.

137. Yotti, R., C.E. Seidman, and J.G.J.A.r.o.g. Seidman, *Advances in the genetic basis and pathogenesis of sarcomere cardiomyopathies.* Annual review of genomics human genetics, 2019. **20**: p. 129-153.

138. Ingles, J., et al., *Evaluating the clinical validity of hypertrophic cardiomyopathy genes.* Circulation: Genomic Precision Medicine, 2019. **12**(2): p. e002460.

139. Craig, R. and R. Padrón, *Molecular structure of the sarcomere.* Myology, 2004. **3**: p. 129-144.

140. Al-Khayat, H.A., *Three-dimensional structure of the human myosin thick filament: clinical implications.* Global Cardiology Science, 2013. **2013**(3): p. 36.

141. Luther, P.K., et al., *Direct visualization of myosin-binding protein C bridging myosin and actin filaments in intact muscle.* Proceedings of the National Academy of Sciences, 2011. **108**(28): p. 11423-11428.

142. Li, M. and W. Zheng, *All-Atom Molecular Dynamics Simulations of Actin–Myosin Interactions: A Comparative Study of Cardiac α Myosin, β Myosin, and Fast Skeletal Muscle Myosin.* Biochemistry, 2013. **52**(47): p. 8393-8405.

143. von der Ecken, J., et al., *Cryo-EM structure of a human cytoplasmic actomyosin complex at near-atomic resolution.* Nature, 2016. **534**(7609): p. 724-728.

144. Yamada, Y., K. Namba, and T. Fujii, *Cardiac muscle thin filament structures reveal calcium regulatory mechanism.* Nature communications, 2020. **11**(1): p. 1-9.

145. Risi, C., et al., *Ca2+-induced movement of tropomyosin on native cardiac thin filaments revealed by cryoelectron microscopy.* Proceedings of the National Academy of Sciences, 2017. **114**(26): p. 6782-6787.

146. Sweeney, H.L. and A. Houdusse, *Structural and functional insights into the myosin motor mechanism.* Annual review of biophysics, 2010. **39**: p. 539-557.

147.	Houdusse, A. and H.L. Sweeney, *How myosin generates force on actin filaments.* Trends in biochemical sciences, 2016. **41**(12): p. 989-997.

148.	Rayment, I., et al., *Three-dimensional structure of myosin subfragment-1: a molecular motor.* Science, 1993. **261**(5117): p. 50-58.

149.	Yount, R.G., D. Lawson, and I. Rayment, *Is myosin a" back door" enzyme?* Biophysical journal, 1995. **68**(4 Suppl): p. 44S.

150.	Coureux, P.-D., et al., *A structural state of the myosin V motor without bound nucleotide.* 2003. **425**(6956): p. 419-423.

151.	Reubold, T.F., et al., *A structural model for actin-induced nucleotide release in myosin.* Nature Structural Molecular Biology, 2003. **10**(10): p. 826-830.

152.	Wulf, S.F., et al., *Force-producing ADP state of myosin bound to actin.* Proceedings of the National Academy of Sciences, 2016. **113**(13): p. E1844-E1852.

153.	Takagi, Y., et al., *Force generation in single conventional actomyosin complexes under high dynamic load.* Biophysical journal, 2006. **90**(4): p. 1295-1307.

154.	Tehver, R. and D. Thirumalai, *Rigor to post-rigor transition in myosin V: link between the dynamics and the supporting architecture.* Structure, 2010. **18**(4): p. 471-481.

155.	Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic acids research, 2018. **46**(W1): p. W296-W303.

156.	Yang, Y., et al., *Rigor-like structures from muscle myosins reveal key mechanical elements in the transduction pathways of this allosteric motor.* Structure, 2007. **15**(5): p. 553-564.

157.	Robert-Paganin, J., D. Auguin, and A. Houdusse, *Hypertrophic cardiomyopathy disease results from disparate impairments of cardiac myosin function and auto-inhibition.* Nature communications, 2018. **9**(1): p. 1-13.

158.	Mark, P. and L. Nilsson, *Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K.* The Journal of Physical Chemistry A, 2001. **105**(43): p. 9954-9960.

159.	Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* Journal of molecular graphics, 1996. **14**(1): p. 33-38.

160.	Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* Journal of computational chemistry, 2005. **26**(16): p. 1781-1802.

161.	Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug‐like molecules compatible with the CHARMM all‐atom additive biological force fields.* Journal of computational chemistry, 2010. **31**(4): p. 671-690.

162.	Xian, Y., et al., *Structure manipulation tool structureMan: A structure manipulation tool to study large scale biomolecular interactions.* Frontiers in molecular biosciences, 2020. **7**: p. 476.

163.	Dolinsky, T.J., et al., *PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations.* Nucleic acids research, 2007. **35**(suppl_2): p. W522-W525.

164.	White, K.A., et al., *Cancer-associated arginine-to-histidine mutations confer a gain in pH sensing to mutant proteins.* Science signaling, 2017. **10**(495): p. eaam9931.

165.	Hubbard, S. and J. Thornton, *NACCESS: program for calculating accessibilities.* 1992.

166.	Xie, Y., et al., *Revealing the mechanism of SARS-CoV-2 spike protein binding with ACE2.* Computing in Science Engineering, 2020. **22**(6): p. 21-29.

167.	Chen, C., et al., *The effect of geometrical criteria on hydrogen bonds analysis in aqueous glycerol solutions.* J. Mol. Imag. Dynamics, 2011. **1**: p. 101.

168. Tehver, R. and D.J.S. Thirumalai, *Rigor to post-rigor transition in myosin V: link between the dynamics and the supporting architecture.* 2010. **18**(4): p. 471-481.

169. Humphrey, W., A. Dalke, and K.J.J.o.m.g. Schulten, *VMD: visual molecular dynamics.* 1996. **14**(1): p. 33-38.

170. Sun, S., et al., *Hybrid method for representing ions in implicit solvation calculations.* Computational structural biotechnology journal, 2021. **19**: p. 801-811.

171. Karki, C., et al., *A computational model of ESAT-6 complex in membrane.* Journal of Theoretical Computational Chemistry, 2020. **19**(03): p. 2040002.

172. Genheden, S. and U. Ryde, *The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities.* Expert opinion on drug discovery, 2015. **10**(5): p. 449-461.

173. Dhandapany, P.S., et al., *A common MYBPC3 (cardiac myosin binding protein C) variant associated with cardiomyopathies in South Asia.* Nature genetics, 2009. **41**(2): p. 187-191.

174. Xie, Y., et al., *Spike proteins of SARS-CoV and SARS-CoV-2 utilize different mechanisms to bind with human ACE2.* Frontiers in molecular biosciences, 2020: p. 392.

175. Xie, Y., et al., *Revealing the mechanism of SARS-CoV-2 spike protein binding with ACE2.* Computing in Science & Engineering, 2020. **22**(6): p. 21-29.

176. Sun, S., et al., *Computational study on the function of palmitoylation on the envelope protein in SARS-CoV-2.* Journal of chemical theory and computation, 2021. **17**(10): p. 6483-6490.

177. Matousek, W.M., et al., *Electrostatic contributions to the stability of the GCN4 leucine zipper structure.* Journal of molecular biology, 2007. **374**(1): p. 206-219.

178. Price, D.J. and C.L. Brooks III, *A modified TIP3P water potential for simulation with Ewald summation.* The Journal of chemical physics, 2004. **121**(20): p. 10096-10103.

179. Abascal, J.L. and C. Vega, *A general purpose model for the condensed phases of water: TIP4P/2005.* The Journal of chemical physics, 2005. **123**(23): p. 234505.

180. Nicholls, A. and B. Honig, *A rapid finite difference algorithm, utilizing successive over‐relaxation to solve the Poisson‐Boltzmann equation.* Journal of computational chemistry, 1991. **12**(4): p. 435-445.

181. Jayaram, B., D. Sprous, and D. Beveridge, *Solvation free energy of biomacromolecules: Parameters for a modified generalized Born model consistent with the AMBER force field.* The Journal of Physical Chemistry B, 1998. **102**(47): p. 9571-9576.

182. Petukh, M., et al., *Predicting nonspecific ion binding using DelPhi.* Biophysical journal, 2012. **102**(12): p. 2885-2893.

183. Sun, S., et al., *HIT web server: A hybrid method to improve electrostatic calculations for biomolecules.* Computational and Structural Biotechnology Journal, 2022. **20**: p. 1580-1583.

184. Katz, A.K., et al., *Calcium ion coordination: a comparison with that of beryllium, magnesium, and zinc.* Journal of the American Chemical Society, 1996. **118**(24): p. 5752-5763.

185. Babor, M., et al., *Prediction of transition metal‐binding sites from apo protein structures.* Proteins: Structure, Function, and Bioinformatics, 2008. **70**(1): p. 208-217.

186. Shashikala, H.M., et al., *BION-2: Predicting Positions of Non-Specifically Bound Ions on Protein Surface by a Gaussian-Based Treatment of Electrostatics.* International Journal of Molecular Sciences, 2020. **22**(1): p. 272.

187. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* Journal of computational chemistry, 2005. **26**(16): p. 1701-1718.

188. Condon, E., *A simple derivation of the Maxwell-Boltzmann Law.* Physical Review, 1938. **54**(11): p. 937.

189. Li, Z., et al., *Structural basis of Notch O-glucosylation and O–xylosylation by mammalian protein–O-glucosyltransferase 1 (POGLUT1).* Nature communications, 2017. **8**(1): p. 1-12.

190. Snarski-Adamski, J. and M. Werwiński, *Effect of transition metal doping on magnetic hardness of CeFe12-based compounds.* Journal of Magnetism and Magnetic Materials, 2022. **554**: p. 169309.

191. Whittingham, J.L., et al., *Interactions of phenol and m-cresol in the insulin hexamer, and their effect on the association properties of B28 Pro → Asp insulin analogues.* Biochemistry, 1998. **37**(33): p. 11516-11523.

192. Salie, Z.L., et al., *Structural basis of HIV inhibition by translocation-defective RT inhibitor 4′-ethynyl-2-fluoro-2′-deoxyadenosine (EFdA).* Proceedings of the National Academy of Sciences, 2016. **113**(33): p. 9274-9279.

193. Hingerty, B., R. Brown, and A. Jack, *Further refinement of the structure of yeast tRNAPhe.* Journal of molecular biology, 1978. **124**(3): p. 523-534.

194. Noinaj, N., et al., *Structural insight into the biogenesis of β-barrel membrane proteins.* Nature, 2013. **501**(7467): p. 385-390.

195. Guex, N. and M.C. Peitsch, *SWISS‐MODEL and the Swiss‐Pdb Viewer: an environment for comparative protein modeling.* electrophoresis, 1997. **18**(15): p. 2714-2723.

196. Jo, S., et al., *CHARMM‐GUI: a web‐based graphical user interface for CHARMM.* Journal of computational chemistry, 2008. **29**(11): p. 1859-1865.

197. Wu, E.L., et al., *CHARMM‐GUI membrane builder toward realistic biological membrane simulations.* 2014, Wiley Online Library.

198. Huang, J., et al., *CHARMM36m: an improved force field for folded and intrinsically disordered proteins.* Nature methods, 2017. **14**(1): p. 71-73.

199. Dolinsky, T.J., et al., *PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations.* Nucleic acids research, 2004. **32**(suppl_2): p. W665-W667.

200. Hilbe, J.M., *Logistic regression models.* 2009: Chapman and hall/CRC.

201. Breiman, L., et al., *Classification and regression trees. Belmont, CA: Wadsworth.* International Group, 1984. **432**(151-166): p. 9.

202. Breiman, L., *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.

203. Hopfield, J.J., *Artificial neural networks.* IEEE Circuits and Devices Magazine, 1988. **4**(5): p. 3-10.

204. Breiman, L., et al., *Classification and regression trees.* 2017: Routledge.

205. Vinogradova, M.V., et al., *Ca2+-regulated structural changes in troponin.* Proceedings of the National Academy of Sciences, 2005. **102**(14): p. 5038-5043.

206. Seavey, M.M. and P. Dobrzanski, *The many faces of Janus kinase.* Biochemical pharmacology, 2012. **83**(9): p. 1136-1145.

207. O'shea, J.J., et al., *A new modality for immunosuppression: targeting the JAK/STAT pathway.* Nature reviews Drug discovery, 2004. **3**(7): p. 555-564.

208. Schindler, C., D.E. Levy, and T. Decker, *JAK-STAT signaling: from interferons to cytokines.* Journal of Biological Chemistry, 2007. **282**(28): p. 20059-20063.

209. Xin, P., et al., *The role of JAK/STAT signaling pathway and its inhibitors in diseases.* International immunopharmacology, 2020. **80**: p. 106210.

210. Morgan, E.L. and A. Macdonald, *Manipulation of JAK/STAT signalling by high-risk HPVs: potential therapeutic targets for HPV-associated malignancies.* Viruses, 2020. **12**(9): p. 977.

211. Fasouli, E.S. and E. Katsantoni, *JAK-STAT in early hematopoiesis and leukemia.* Frontiers in Cell and Developmental Biology, 2021. **9**: p. 669363.

212. Eletto, D., et al., *Biallelic JAK1 mutations in immunodeficient patient with mycobacterial infection.* Nature communications, 2016. **7**(1): p. 13992.

213. Gruber, C.N., et al., *Complex autoinflammatory syndrome unveils fundamental principles of JAK1 kinase transcriptional and biochemical function.* Immunity, 2020. **53**(3): p. 672-684. e11.

214. Bousoik, E. and H. Montazeri Aliabadi, *"Do we know jack" about JAK? A closer look at JAK/STAT signaling pathway.* Frontiers in oncology, 2018. **8**: p. 287.

215. Wilks, A.F., *Two putative protein-tyrosine kinases identified by application of the polymerase chain reaction.* Proceedings of the National Academy of Sciences, 1989. **86**(5): p. 1603-1607.

216. Rodig, S.J., et al., *Disruption of the Jak1 gene demonstrates obligatory and nonredundant roles of the Jaks in cytokine-induced biologic responses.* Cell, 1998. **93**(3): p. 373-383.

217. Kawamura, M., et al., *Molecular cloning of L-JAK, a Janus family protein-tyrosine kinase expressed in natural killer cells and activated leukocytes.* Proceedings of the National Academy of Sciences, 1994. **91**(14): p. 6374-6378.

218. Johnston, J.A., et al., *Phosphorylation and activation of the Jak-3 Janus kinase in response to interleukin-2.* Nature, 1994. **370**(6485): p. 151-153.

219. Witthuhn, B.A., et al., *Involvement of the Jak-3 Janus kinase in signalling by interleukins 2 and 4 in lymphoid and myeloid cells.* Nature, 1994. **370**(6485): p. 153-157.

220. Zhou, Y.-J., et al., *Distinct tyrosine phosphorylation sites in JAK3 kinase domain positively and negatively regulate its enzymatic activity.* Proceedings of the National Academy of Sciences, 1997. **94**(25): p. 13850-13855.

221. Cheng, H., et al., *Phosphorylation of human Jak3 at tyrosines 904 and 939 positively regulates its activity.* Mol Cell Biol, 2008. **28**(7): p. 2271-82.

222. Cheng, H., et al., *Phosphorylation of human Jak3 at tyrosines 904 and 939 positively regulates its activity.* Molecular and cellular biology, 2008. **28**(7): p. 2271-2282.

223. Glassman, C.R., et al., *Structure of a Janus kinase cytokine receptor complex reveals the basis for dimeric activation.* Science, 2022. **376**(6589): p. 163-169.

224. Lupardus, P.J., et al., *Structure of the pseudokinase-kinase domains from protein kinase TYK2 reveals a mechanism for Janus kinase (JAK) autoinhibition.* Proc Natl Acad Sci U S A, 2014. **111**(22): p. 8025-30.

225. Caspers, N.L., et al., *Development of a high-throughput crystal structure-determination platform for JAK1 using a novel metal-chelator soaking system.* Acta Crystallographica Section F: Structural Biology Communications, 2016. **72**(11): p. 840-845.

226. Thorarensen, A., et al., *Design of a Janus Kinase 3 (JAK3) Specific Inhibitor 1-((2 S, 5 R)-5-((7 H-Pyrrolo [2, 3-d] pyrimidin-4-yl) amino)-2-methylpiperidin-1-yl) prop-2-en-1-one (PF-06651600) Allowing for the Interrogation of JAK3 Signaling in Humans.* Journal of medicinal chemistry, 2017. **60**(5): p. 1971-1993.

227. Chen, C., et al., *The effect of geometrical criteria on hydrogen bonds analysis in aqueous glycerol solutions.* J. Mol. Imag. Dyn, 2011. **1**(1).

228. Hubbard, S. and J. Thornton, *NACCESS: program for calculating accessibilities.* Department of Biochemistry and Molecular Biology, University College of London, 1992.

229. Chen, L., et al., *Inflammatory responses and inflammation-associated diseases in organs.* Oncotarget, 2018. **9**(6): p. 7204.

230. Rane, S.G. and E.P. Reddy, *JAKs, STATs and Src kinases in hematopoiesis.* Oncogene, 2002. **21**(21): p. 3334-3358.

231. Buchert, M., C. Burns, and M. Ernst, *Targeting JAK kinase in solid tumors: emerging opportunities and challenges.* Oncogene, 2016. **35**(8): p. 939-951.

232. Wilmes, S., et al., *Mechanism of homodimeric cytokine receptor activation and dysregulation by oncogenic mutations.* Science, 2020. **367**(6478): p. 643-652.

233. Watowich, S.S., et al., *Homodimerization and constitutive activation of the erythropoietin receptor.* Proceedings of the National Academy of Sciences, 1992. **89**(6): p. 2140-2144.

234. Stark, G.R. and J.E. Darnell, *The JAK-STAT pathway at twenty.* Immunity, 2012. **36**(4): p. 503-514.

235. Chen, X., et al., *Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA.* Cell, 1998. **93**(5): p. 827-839.

236. Babon, J.J., et al., *The molecular regulation of Janus kinase (JAK) activation.* Biochemical Journal, 2014. **462**(1): p. 1-13.

237. Ferrao, R. and P.J. Lupardus, *The Janus kinase (JAK) FERM and SH2 domains: Bringing specificity to JAK–receptor interactions.* Frontiers in endocrinology, 2017. **8**: p. 71.

238. Saharinen, P., K. Takaluoma, and O. Silvennoinen, *Regulation of the Jak2 tyrosine kinase by its pseudokinase domain.* Molecular and cellular biology, 2000. **20**(10): p. 3387-3395.

239. Toms, A.V., et al., *Structure of a pseudokinase-domain switch that controls oncogenic activation of Jak kinases.* Nature structural & molecular biology, 2013. **20**(10): p. 1221-1223.

240. Baxter, E.J., et al., *Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders.* The Lancet, 2005. **365**(9464): p. 1054-1061.

241. Dusa, A., et al., *Substitution of pseudokinase domain residue Val-617 by large non-polar amino acids causes activation of JAK2.* Journal of Biological Chemistry, 2008. **283**(19): p. 12941-12948.

242. Xian, Y., et al., *StructureMan: A structure manipulation tool to study large scale biomolecular interactions.* Frontiers in molecular biosciences, 2021. **7**: p. 627087.

243. Javaid, A., *Understanding Dijkstra's algorithm.* Available at SSRN 2340905, 2013.

244. Dijkstra, E.W., *A note on two problems in connexion with graphs:(Numerische Mathematik, 1 (1959), p 269-271).* 1959.

245. Consortium, U., *UniProt: a worldwide hub of protein knowledge.* Nucleic acids research, 2019. **47**(D1): p. D506-D515.

246. Bellman, R., *Dynamic programming.* Science, 1966. **153**(3731): p. 34-37.

247. De Marinis, E., et al., *Ruxolitinib binding to human serum albumin: bioinformatics, biochemical and functional characterization in JAK2V617F+ cell models.* Scientific Reports, 2019. **9**(1): p. 1-12.

248. Bright, J.J., L.D. Kerr, and S. Sriram, *TGF-beta inhibits IL-2-induced tyrosine phosphorylation and activation of Jak-1 and Stat 5 in T lymphocytes.* Journal of immunology (Baltimore, Md.: 1950), 1997. **159**(1): p. 175-183.

249. Imada, K. and W.J. Leonard, *The jak-STAT pathway.* Molecular immunology, 2000. **37**(1-2): p. 1-11.

250. Schaeffer, L., *The role of functional groups in drug–receptor interactions*, in *The practice of medicinal chemistry*. 2008, Elsevier. p. 464-480.

251. Steiner, T., *The hydrogen bond in the solid state.* Angewandte Chemie International Edition, 2002. **41**(1): p. 48-76.

CURRICULUM VITA

In August of 1993, from the East emerged a being, named Shengjie Sun. His birth foretold a tale of greatness, one that shall echo through the ages.

In 2012, Shengjie enrolled in The Central South University to study bioengineering. In his third year, he was given the opportunity to conduct research in Dr. Min's laboratory. This turned out to be a significant turning point in his life, as he discovered his passion for research. Over the following two years, Shengjie published two papers, which focused on Schwertmannite. These scientific publications increased his confidence and fueled his interest in the field of science.

After graduating with a bachelor's degree in bioengineering in 2016, Shengjie applied for a Ph.D. program in environmental studies at the University of Queensland (Australia) in 2017. However, he had to stop his studies in 2018 due to financial difficulties. This experience taught him the importance of self-reliance and taking responsibility for one's own career path. He realized that while family support is valuable, ultimately, individuals must be prepared to invest in their own careers and take calculated risks to succeed. This setback did not deter him from his passion for research, and he continued to pursue opportunities to develop his skills and knowledge in the field.

After his disappointment with his Ph.D. program, Shengjie decided to pursue a computational science program at The University of Texas at El Paso. He was fortunate to have Dr. Lin Li as his supervisor, who helped him bridge the gap between different fields of knowledge and come up with innovative ideas that combined biology and chemistry. Despite facing challenges, Shengjie persevered and dedicated two years to developing the Hybridizing Ions Treatment (HIT) method (Versions 1 and 2) for bound searching. In addition, he applied computational methods to various biological problems, such as the kinetic cycle of myosin, palmitoylation on the SARS-CoV-2 Envelope protein, and Janus Kinases. Through his work, Shengjie demonstrated his ability to apply computational methods to a wide range of biological problems and contribute novel insights to the field.