COMPUTATIONAL STUDIES TO UNDERSTAND THE ROLE OF

ALLOSTERY IN COPPER REGULATION IN *Mycobacterium tuberculosis*

AND IN THE DESIGN OF HPV VACCINES

Abhigna Polavarapu

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Chemistry

Indiana University

April 2015

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____

Mu-Hyun Baik, Ph.D.

_____

David Giedroc, Ph.D.

_____

Charles Dann III, Ph.D.

_____

Peter Ortoleva, Ph.D.

Date of Dissertation Defense

September 27th, 2013

This thesis is dedicated to my parents, husband and my daughter for their love, continuous

support and encouragement throughout my Ph.D

Acknowledgements

I wish to express my deepest gratitude towards my thesis advisor, Prof. Mu-Hyun Baik, for providing me with guidance and encouragement throughout my graduate career. I have grown tremendously as a scientist under his patient (and sometimes not-so-patient) training and the freedom he gave me in the choosing of my projects. I could not have imagined having a better advisor and mentor for my Ph.D. study. I also wish to thank my committee members, Prof. Charles Dann III, Prof. David Giedroc and Prof. Peter Ortoleva, for helpful suggestions and critical evaluation my research and dissertation.

I like to thank all the past and present members of the Baik research group for being excellent colleagues and friends, making it fun to come to work. In particular, I'd like to thank Dr. Soumya Ghosh and Dr. Shivnath Mazumder for many helpful discussions.

Finally, I wish to thank our collaborators for their helpful contributions to my project and intellectual development. These include Prof. David Giedroc (Indiana University), Prof. Peter Ortoleva (Indiana University), Dr. Harshad Joshi, and Dr. Abhishek Singharoy.

Abhigna Polavarapu

<u>Computational Studies to Understand the Role of Allostery in Copper</u>

<u>Regulation in *Mycobacterium tuberculosis* and in the Design of HPV Vaccines</u>

Allostery is defined as the change in the structure, function or activity of a specific site on a protein, due to the binding of a substrate or effecter on a different site of the same protein. This phenomenon has been observed and studied in two different protein systems of therapeutic importance. CsoR protein in *Mycobacterium tuberculosis* adopts classical allostery to regulate the concentration of Cu(I) inside the cell. Cu(I) is speculated to bind in an unusual trigonal planar geometry with two cysteines and one histidine. When CsoR is bound to copper an overall structural change (allostery) is envisioned and its affinity to DNA is lost. In the current computational exploration we focus on the binding mode of Cu(I) and identify different protonation states of copper bound cysteines. MD simulations were performed on the apo and copper bound form with a starting structure from QM/MM calculations to predict the allosteric structural transition. The dynamic properties of the capsid of the human papillomavirus (HPV) type 16 were also examined using classical molecular dynamics simulations. The allostery identified in the components of the HPV is non-classical because the mean structure of the epitope carrying loops remains unchanged as the result of

allosteric effect, but the structural fluctuations are altered significantly, which in turn changes the biochemical reactivity profile of the epitopes. Exploiting this novel insight, a new vaccine design strategy is proposed, where a relatively small virus fragment is deposited on a silica nanoparticle in such a way that the fluctuations of the h4 helix are suppressed. The structural and dynamic properties of the epitope carrying loops on this hybrid nanoparticle match the characteristics of epitopes found on the full virus like particle precisely, suggesting that these nanoparticles may serve as potent, cost-effective and safe alternatives to traditionally developed vaccines.

_____

Mu-Hyun Baik, Ph.D.

_____

David Giedroc, Ph.D.

_____

Charles Dann III, Ph.D.

_____

Peter Ortoleva, Ph.D.

Table of Contents

ABBREVIATIONS

ArsR   Arsenic regulator

B. subtilis, Bs   *Bacillus subtilis*

BmrR   Bacillus multidrug responding
transcriptionalregulator

BxmR   *Oscillatoria brevis* CPx-ATPase metal-regulated

CsoR   Copper-sensitive operon repressor

CueR   Copper efflux regulator

CzrA   Chromosomally encoded zinc-regulated repressor

Fur   Ferric uptake regulator

HPV   Human Papilloma Virus

MD   Molecular Dynamics

MerR   Mercuric ion resistance regulator

M. tuberculosis, Mtb   *Mycobacterium tuberculosis*

NikR   Nickel responsive regulator of the *nik* operon

NMA   Normal Mode Analysis

NMR   Nuclear magnetic resonance

NmtR   Nickel/cobalt-responsive transcriptional repressor

PCA   Principal Component Analysis

SmtB   *Synechococcus* metallothionein locus, gene

SoxR   Redox-sensitive transcriptional activator

S. lividans, Sl   *Streptomyces lividans*

T. thermophillus,Tt   *Thermus thermophilus*

INTRODUCTION

The interaction between biochemical macromolecules and small molecules that serve as messengers or trigger of cellular responses play a pivotal role in biological sensing and signaling, which in turn are critical for the survival of any organism.[1] Proteins are one of the main pillars of these constructs and unlike other biologically important macromolecules, such as DNA, RNA or glycosides, they display a rich regulatory mechanism based on allostery - the coupling of molecular changes of one site to detectable chemical changes at a remote site of the protein.[1-4] The most obvious manifestation of this principle is documented in the lock-and-key mechanism, where the interaction of a small molecule with a specific binding site of a regulatory protein can lead to the activation of the protein that sets in motion a chain of biochemical reactions. For example, the cooperative binding of oxygen to hemoglobin is under allosteric control, wherein the uptake of one oxygen molecule by hemoglobin increases the oxygen binding affinities of remaining unoccupied binding sites by influencing the quaternary structure of the protein.[5-9] Over the years, hemoglobin served as an important model system for investigating the basic principle of allostery and cooperativity in macromolecules.[10-13] Hemoglobin is a tetrameric metalloprotein, and has four binding sites in each monomeric unit that consists of a heme moiety, a porphyrin group carrying a Fe(II) center.[13,14] Due to the presence of these four binding sites,

the oxygen binding gives rise to multiple equilibria.[13-15] Curiously, the rate of oxygen binding that can be visualized for example as a function of partial pressure, as shown in Figure 1.1a, is sigmoidal.[5,16] The equilibrium constants for these four oxygen binding events are dependent on each other i.e., filling the first oxygen binding site in hemoglobin increases the affinity of the remaining sites for oxygen and vice versa. This influence of one oxygen molecule binding on the binding of another oxygen is known as homotropic cooperativity. Overall, this cooperative equilibrium binding makes the binding curve sigmoidal rather than hyperbolic, which is characteristic of single-site oxygen binding proteins, such as myoglobin.[5,13,16] Cooperativity in hemoglobin results from a conformational "switch" from a weak-binding state to a strong-binding state, which is the manifestation of allostery.[8,9,17]
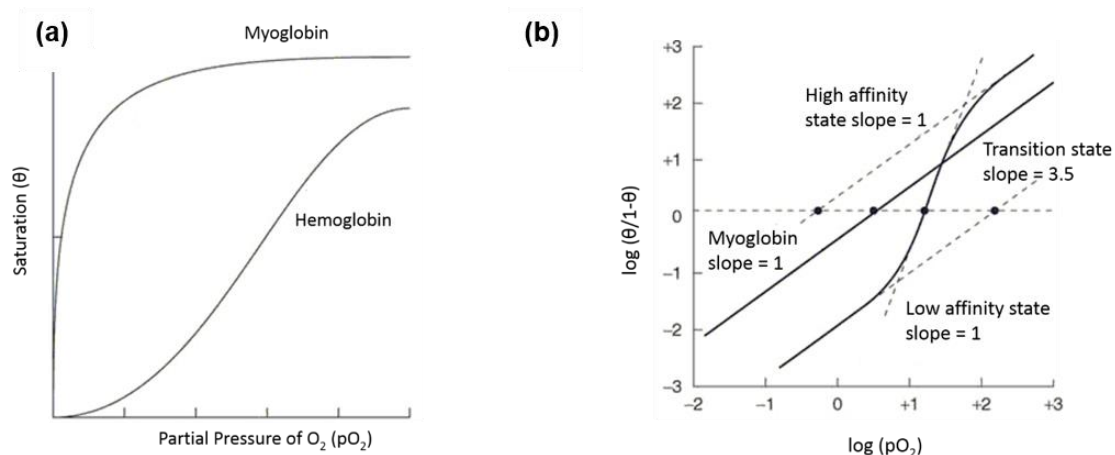
Figure Error! No text of specified style in document..**1.** (a) Oxygen binding curve of hemoglobin and myoglobin (b) Hill plot describing the transition between low affinity, transition and high affinity oxygen binding states of hemoglobin

The O_2 binding equilibrium for myoglobin can be given by the following expression: $Mb + O_2 \leftrightarrow Mb - O_2$ and the equilibrium constant, $K_{eq} = \frac{[Mb-O_2]}{[Mb][O_2]}$.

Oxygen dissociation from myoglobin follows Michaelis-Menten kinetics and can be expressed as fractional saturation, 'θ' and [O_2] in partial pressure, pO_2 where:

$$\theta = \frac{[Mb-O_2]}{[Mb]+[Mb-O_2]}$$

1.1

Substituting K_{eq} and $1/K_{eq} = [O2]_{1/2}$ in equation 1.1, yields equation 1.2 that describes a hyperbolic curve (Figure 1.1a):

$$\theta = \frac{pO_2}{P_{50}+pO_2}$$

1.2

As hemoglobin is a tetramer, the equilibrium involves four binding steps which can be expressed as $Hb + 4\,O_2 \leftrightarrow Hb(O_2)_4$ and in more general terms $Hb + nO_2 \leftrightarrow Hb(O_2)_n$ and does not follow Michaelis-Menten kinetics. Hence, the fractional saturation is described by Hill equation and the binding curve is sigmoidal rather than hyperbolic:

$$\theta = \frac{(pO_2)^n}{(P_{50})^n+(pO_2)^n}$$

1.3

Quantitatively, the binding curve of hemoglobin and the allosteric nature of cooperativity is explained by equation 1.4 and this equation yields Hill plot (Figure 1.1b):

$$\log\frac{\theta}{1-\theta} = n\log(pO_2) - n\log(P_{50}) \qquad\qquad 1.4$$

where 'θ' is the fraction of oxygen-binding sites that are occupied, pO_2 is the partial pressure of oxygen, and P_{50} is the oxygen partial pressure for half-saturation. Hill plots predict the transition between n binding states in cooperativity and give a direct numerical measure of the degree of cooperativity from its maximum slope, n x H, which is known as the Hill coefficient. A Hill coefficient of 1 indicates that the binding of oxygen is non-cooperative as seen in myoglobin, a value greater than 1 indicates positive cooperativity. For example, a maximum value of n x H is 3.5 (Figure 1.1b) for hemoglobin observed for the transition between low affinity and high affinity states, suggesting cooperativity between the four oxygen binding sites. This model explains allostery in hemoglobin as the interconversion between two states: the T (tense) and the R (relaxed) conformations of the molecule.[8,9,18] The R state has higher affinity for oxygen. Under conditions where pO_2 is high (such as in the lungs), the R-state is favored; in conditions where pO_2 is low (as in exercising muscle), the T-state is favored.[8,9,18] In agreement with the above model, the X-ray crystal structures of

oxygen bound and unbound structures of hemoglobin revealed the existence of two quaternary conformations - low-affinity deoxy T-state and the high-affinity oxy R-state.[17] The transition within the two states has been determined based on stereochemical events triggered by O_2 binding to Fe(II), which moves the iron atom by 0.6 Å into the plane of the porphyrin ring.[10-12] This structural change initiates a series of geometrical responses in the F-helix and neighboring helices, due to the pull on the proximal histidine exerted by the movement of iron atom into the plane of the porphyrin ring, illustrated in Figure 1.2.[10-12] These structural variations are transmitted to the subunit interfaces and the salt bridges between the subunits are broken to attain a relaxed 'R' state.[10-12] Upon conversion from the deoxy (T) structure to the oxy (R) structure, the $\alpha1\beta2$ dimer rotates relative to the other by ~15° and opens the Fe(II) binding sites to accommodate oxygen in other subunits.[10-12,17] Similarly, even today allostery in proteins is often explained via structural changes - Structure-function paradigm, open and closed state of a binding site to accommodate and release substrate at a distant site, surface remodeling to allow protein-protein association and inter-domain movements with respect to each other that impact the quaternary structure of the protein.
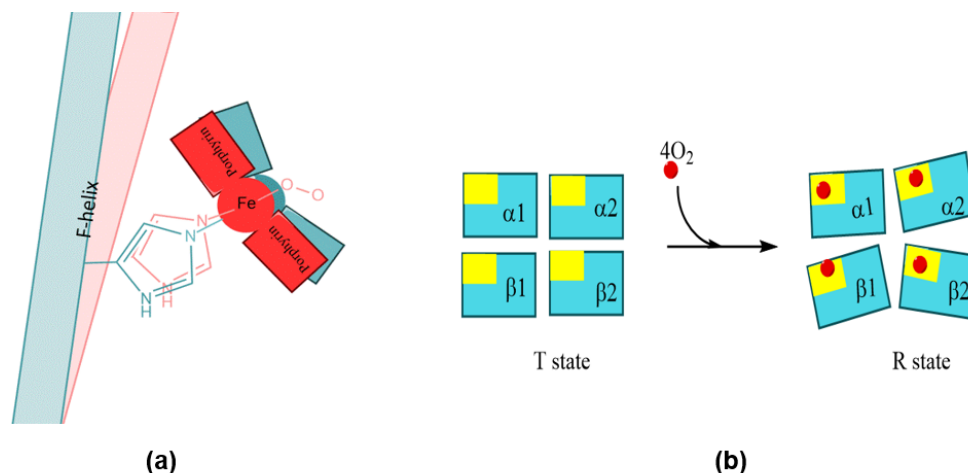
Figure 1.2. (a) Structural change at the heme center of deoxy-hemoglobin 'T' state (blue) on binding to oxygen that initiates the allosteric transition to oxy-hemoglobin 'R' state (red) (b) Cartoon showing the quaternary structure change from 'T' to 'R' state

The structure-function relationship supports the existence of distinct key structures for different functional states of protein. Two conceptual models exist for understanding allostery and explaining the relationship between structure-coupled transitions and discrete functional states of a protein. These models have benefited tremendously from precise molecular structure determinations of tetrameric hemoglobin and they are: The Koshland-Nemethy-Filmer (KNF)[9] and the Monod-Wyman-Changeux (MWC) models.[8] The KNF model describes the structural transition between the T- and R-states as an induced-fit mechanism, where binding of a ligand leads to a subsequent structural change at a remote functional site. This theory suggests that the subunits in hemoglobin switch independently from the binding to an oxygen molecule, but cooperatively facilitate further transitions to the high-affinity R-state. In contrast, the main

concept of the MWC model is the existence of different interconvertible states in equilibrium, e.g. all subunits of hemoglobin must exist in either the T- or R-state. Binding of a ligand leads to a shift in the thermal equilibrium towards one state.

The transition within the two states has been determined based on stereochemical events triggered by O_2 binding to Fe(II), which moves the iron atom by 0.6 Å into the plane of the porphyrin ring.[10-12] This structural change initiates a series of geometrical responses in the F-helix and neighboring helices, due to the pull on the proximal histidine exerted by the movement of iron atom into the plane of the porphyrin ring, illustrated in Figure 1.2.[10-12] Support for the MWC model of allostery comes from nuclear magnetic resonance (NMR) studies that demonstrate the coexistence of different conformations of phosphorylation regulated nitrogen regulatory protein C (NtrC).[19-21] NMR spectroscopy was used to demonstrate a strong correlation between the phosphorylation-driven activation of the signaling protein NtrC and microsecond timescale backbone dynamics in an area remote from the phosphorylation site as shown in Figure 1.3.[19] Solution structures and the motions of the regulatory domain of NtrC were determined in three functional states: unphosphorylated (inactive), phosphorylated (active) and a partially active mutant.[19] The NMR-detected dynamics in the unphosphorylated state were detected exactly at the area of conformational change upon activation, indicative of an equilibrium of states

(Figure 1.3).[19] Chemical shift analysis of NtrC states with different activity revealed that the dynamics detected by NMR relaxation experiments represent an exchange between inactive and active states in unphosphorylated NtrC. The relative amount of shift is a direct measure of the equilibrium constant between the two states. Phosphorylation shifts the equilibrium towards the active state. Additional studies have indicated that the area involved in dynamic exchange is responsible for downstream signal transduction, the activation of ATPase activity in the central domain of NtrC and consequently activation of transcription.[22]
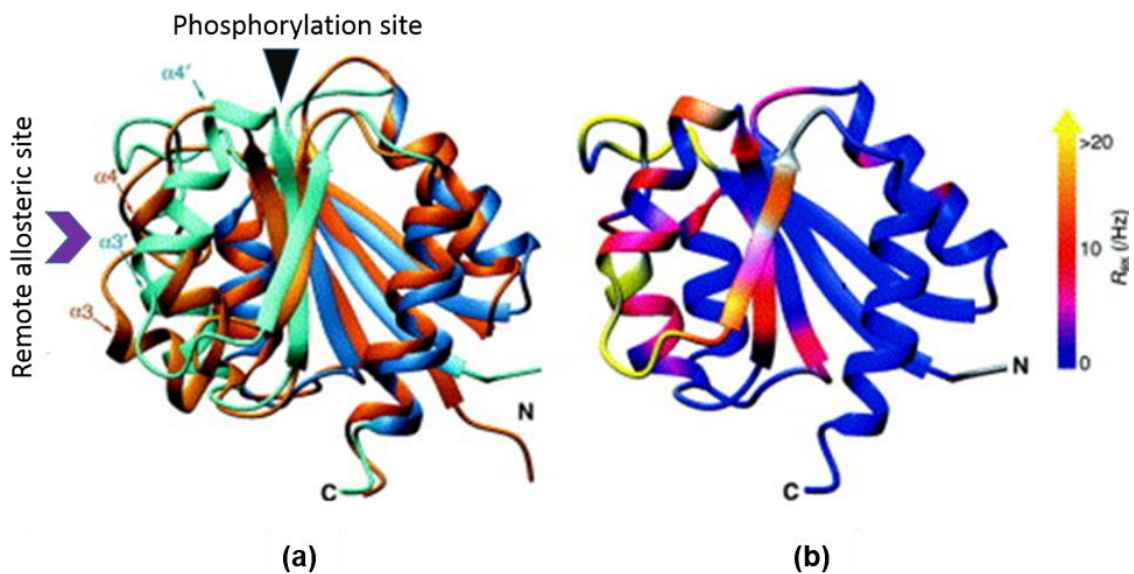
Figure 1.3. (a) Superimposition of inactive (orange/yellow) and active (blue/cyan) of NtrC regulatory (switch) domain. (b) NMR backbone relaxation techniques were applied to detect microsecond timescale motion in region that undergoes conformational changes due to ligand binding. Dynamic regions were identified by the NMR exchange term R_{ex} (shown as a continuous color scale)[19]

NtrC protein, is the first single domain allosteric protein and the presence of allosteric regulation in this protein challenged the traditional MWC and KNF models, which only considered oligomeric proteins as potentially allosteric. The NMR relaxation studies on NtrC protein also provide experimental evidence for dynamic origins of allostery (Figure 1.3).[19] For calmodulin - a calcium binding signaling protein, a similar conformational exchange process was also detected using NMR relaxation experiments.[23-25] Hence, ligand binding favors a pre-existing structure and the allosteric process can be viewed as a shift in the thermal equilibrium in these systems. The different co-existing conformations can be attributed to local minima on an energy landscape[25-28] a model originally developed for protein folding, that are divided by energy barriers. In this theory, the allostery changes the energy landscape by either lowering the end state minima or lowering the energy barriers, thereby allowing transitions to other conformations.[29-32]

DYNAMIC ALLOSTERY

Over the last few decades, advances in experimental and computational tools led to a more quantitative understanding of the role that structural dynamics play in protein function in biochemical processes.[3,4,29,33-37] One consequence of this deeper appreciation was that the classical static perspective

of allosteric interactions has been extended and there is growing awareness in the community now that the modulation of chemical behavior does not necessarily have to be linked to structural changes in the protein quaternary structure.[8,9,18] In this sense, the perspective on allostery can be extended to include entropically dominated state changes, in addition to more classical views that include state variations that are mostly determined by enthalpic changes - which are mostly linked to detectable structural variations.[3,4]

To cast this view in the framework of population shift models that have enjoyed some popularity,[27,38-40] protein fluctuations give rise to changes in the partition of the populated conformational states between which the protein alternates and these protein motions can play a significant role in allosteric control mechanisms.[33] The rates of these structural fluctuations depend on the ligand binding affinity and ligand binding can stabilize certain fluctuations that have functional significance.[41] Different kinds of structural fluctuations exist within a protein, each covering a different time scale that ranges from ~10^{-14} to ~10^5 seconds. These molecular motions include vibrational modes (~10^{-14}s to ~10^{-11} sec) and stochastic conformational transitions (~10^{-11} to ~10^5 sec).[42] Transitions between conformational states in typical biochemical reactions like association and dissociation of ligands or large scale changes like protein folding and unfolding are divided by barriers greater than ~$4k_BT$.[42] The high frequency

modes correspond to localized stretching of N–H or C–H bonds. High frequency vibrational modes (time scales shorter than ~10^{-13} sec) correspond to energies greater than k_BT at physiological temperatures of 300K. Therefore, they are in principle not thermally excited. Frequencies that are thermally excited in the higher frequency region correspond to the vibrations of side chains, as illustrated in Figure 1.4.[42] The low frequency vibrational modes in operation at a time scale of ~10^{-11} sec are over-damped collective modes, which are correlated over whole domains contribute significantly to the entropy of the protein.[42]
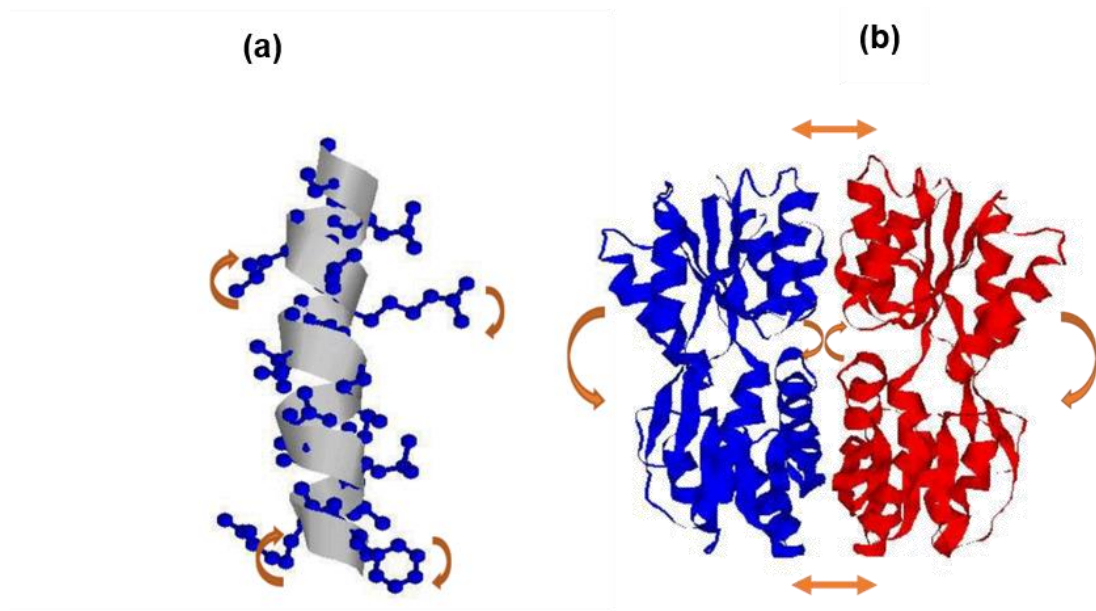
Figure 1.4. (a) High frequency modes (ring, side chain rotations and bond stretches) (b) Low frequency global modes (domain movements)

Though these protein motions play a crucial role in fostering the allosteric conformational change, the concept of allostery without a conformational change

was introduced - dynamic allostery,[33] in which changes in protein dynamics alone produce allostery. By separating the motions into vibrational (normal mode) and conformational effects, the feasibility of an extreme case of no enthalpic term in the allosteric free energy was proposed: where even with the total absence of a conformational change, ligand-induced changes in protein dynamics could produce free energies sufficient for an allosteric communication between distinct binding sites.[33] These energy contributions originate from the changes in frequencies and amplitudes of thermal fluctuations in response to ligand binding and can involve dynamic behavior ranging from highly correlated low-frequency normal mode vibrations to random local anharmonic motions of individual atoms or groups. Dynamic allostery of this form is predominantly an entropic effect.[33,36,43,44] For example, this effect is manifested in the negative cooperative binding of cAMP to the dimeric catabolite activator protein (CAP).[36,45-47] The binding of the first cAMP to one subunit of a CAP dimer does not alter the conformation of the other subunit as evidenced in NMR chemical shift data.[45] However, the system dynamics were modulated by the sequential binding process: the first cAMP partially enhanced and the second cAMP completely quenched protein motions.[36,45] Consequently, the second cAMP binding has significant contribution to conformational entropy penalty, leading to the observed negative cooperative binding of cAMP to CAP.[36,45] To

understand the propagation of dynamic behavior along the cooperative reaction coordinate, the backbone motions of CAP^N as a function of the cAMP ligation state, over a wide range of functionally relevant timescales were measured using NMR relaxation techniques.[36,45] Upon binding of the first cAMP to CAP^N, the backbone residue resonances of both subunits undergo extensive peak broadening which indicates substantial conformational exchange and also suggesting that the cAMP$_1$–CAP^N complex populates an ensemble of alternate conformations that interconvert on the μs–ms timescale. These slow motions are activated by the first cAMP binding but this chemical change cannot induce long-range structural effects in the other subunit. Residues of the unliganded subunit in the cAMP$_1$–CAP^N complex that are located as far as 35 Å from the bound cAMP (for example, Tyr23, Leu29 and Ile30) showed enhanced dynamics and the prominent conformational dynamics on the μs–ms timescale exhibited by almost all residues in cAMP$_1$–CAP^N are completely quenched upon binding to the second cAMP molecule.[36,45,48] Particularly, the slow and fast motions of residues located at distant regions are affected in the absence of a visible connectivity pathway.[48,49] This finding strengthens the mechanical view of allosteric cooperativity, wherein binding effects are assumed to propagate through a series of conformational distortions (or mechanical coupling of protein motions). In this system, the ligand-induced redistribution of the protein's

dynamic fluctuations affects the regions associated with cooperativity, thereby providing a means of propagating the allosteric signal to the distal site even in the absence of structural changes.

Hence, in this broadened view of allostery - proteins are treated as a dynamic ensemble of conformational states where the ligand binding re-distributes the molecular ensemble leading to altered conformations at remote substrate binding sites. The dynamic view of allostery illustrates the possibility of inducing allostery in classically non-allosteric proteins by either mutating certain residues or by introducing disorder in certain regions of protein. These perturbations can redistribute the protein dynamic ensembles and facilitate a population shift that can alter the function of the protein.[38,41] Consistent with this model, all non-fibrous proteins can be considered allosteric and studies focused on identification of dynamically coupled conserved sites in these proteins can aid in modulating the hidden allosteric property.[35,50] In an effort to understand such networks between dynamically coupled sites, multiple V → A mutations were made in the small globular protein eglin c.[50,51] NMR spin relaxation, residual dipolar coupling, and scalar coupling studies, illustrate that the structural architecture of this non-allosteric protein forms a dynamic network and that local perturbations are transmitted as dynamic and structural changes to distal sites as far as 16 Å away.[50,51] Two basic types of propagation mechanism were observed

in eglin c: contiguous pathways of enhanced (attenuated) dynamics with no change in structure; and dispersed (noncontiguous) changes in methyl rotation rates that appear to result from subtle deformation of backbone structure.[50,51] In this thesis, particular emphasis was placed on understanding this inherent dynamic behavior of two distinct protein systems using theoretical methods wherein: a conformational change is observed in one protein while the allostery in the other system propagates through coupling of protein motions at distant sites.

ALLOSTERIC SYSTEMS STUDIED IN THIS THESIS

In this thesis, the goal is to understand the role and mechanism of allostery in functional motions and structural changes in two different protein systems. Computational methods are used to illustrate the mechanism of allostery and exploit the principles of allostery in the design of new drugs and vaccines. We describe two systems, a regulatory metalloprotein in bacteria that exhibits conformational change when bound to a metal and a human papilloma virus (HPV) virus-like particle system in which a conformational change is not observed instead, the protein dynamics contribute to the allosteric control. In Chapter 3, allosteric regulation in metal transcriptional regulator proteins on binding to Cu(I) was investigated. The metal transcriptional regulator protein on

binding to a metal ion alters its affinity towards DNA either by binding or unbinding to DNA.[52-56] The coordination geometry of the metal in the protein plays a key role to drive changes in tertiary and/or quaternary structure and/or dynamics of the metal transcriptional regulator protein. In this research, the coordination geometry of the Cu(I) ion was studied and the correlation between the specific geometry enforced by the protein to bind to Cu(I) and allostery in the protein was thoroughly investigated using high level quantum calculations as well as molecular dynamics simulations. In the following section, an overview of the allosteric regulation in different metal transcriptional regulator proteins in bacteria is provided.

METAL TRANSCRIPTIONAL REGULATORS AND ALLOSTERY

Transcriptional regulators are specialized allosteric proteins that sense cytosolic concentrations of metabolic compounds and other small molecular effectors in order to achieve an appropriate response to changing growth conditions.[57-59] These proteins operate through the interaction with the operator/promoter region of DNA just upstream of the specific operon that is being regulated.[59,60] The ligand or the compound to be regulated binds to the protein–DNA complex, to a site distinct from the DNA binding site and a structural or dynamic change in conformation occurs, that modulates the affinity

or structure of the regulatory protein–DNA complex.[61] Metalloregulatory proteins are a sub-class of transcriptional regulators that have evolved to regulate the expression of cellular metal uptake and detoxification systems.[52-56,62] A number of metal transcriptional regulator families exists and the metal selectivity and specificity of individual proteins can vary significantly even within a single family.[63]

Metal regulatory proteins also known as metal sensor proteins, are involved in metal resistance and their role is to increase the expression of genes involved in metal detoxification, storage and efflux.[53] Metal sensor proteins can regulate the transcription of genes in different ways; one way is by de-repression of regulated genes where the direct binding of a specific ion to a repressor allosterically inhibits DNA binding as is seen in ArsR, CsoR and CopY families,[64-66] or by allosteric activation of transcription initiation by RNA polymerase through the remodeling of the promoter structure as is observed in MerR family.[67] Some metal regulator proteins turn off the expression of uptake systems in response to metal excess, where direct binding of metal ion(s) allosterically activates DNA binding; in this case the metal ion acts as co-repressor of the regulated operon.[68] Consistent with this, the Fur, DtxR and NikR structural scaffolds have evolved to sense only the first row transition elements that are required for biological function; in contrast, the ArsR[64] and MerR families[67]

contain representative members that have evolved to sense a far wider range of metal ions, including both first row transition elements as well as heavy metal ion xenobiotics, and organic As/Sb/Hg metalloid compounds.[56]

Recent structural insights from a wide range of bacterial metal sensor proteins collectively emphasize several common features that characterize these allosteric switches on binding to specific metal ions.[55,56,63,68] One important aspect is that the covalent bond between a metal and its ligand influences the tertiary and/or quaternary structure and/or dynamics but a thorough understanding of the effect of metal coordination geometry on structural and dynamic changes in the protein is not fully established. The degree of metal specificity enforced by these transcriptional regulators suggests that the biological metal selectivity is dictated primarily by the coordination chemistry of the metal ion being regulated.[55]

Chapter 3 focuses on the computational studies that were performed on a Cu(I) sensing transcriptional regulator protein in *Mycobacterium tuberculosis* that belongs to the CsoR family (Mtb-CsoR).[65] The main objective of this study is to critically evaluate the role that protein dynamics play in mediating the overall structural change leading to allostery in these systems as well as to determine

precisely how allostery is linked to coordination geometry of metal ion in these bacterial metal sensor proteins.

Broadly stated, the studies described in Chapter 3 have been motivated and driven by the following question: What is the mechanism of allosteric communication between the metal binding amino acids of CsoR that affords Cu(I) mediated loss of affinity towards DNA? Throughout this chapter an effort is made to map functional information on specific CsoR residues to account for the evolutionary conservation of residues other than the Cu(I) binding residues that are suggested to be important in driving the allosteric switch. Using computational modeling techniques and the insights derived from the previously performed experimental studies,[65,69-73] the results presented identify specific residues that are important for the communication mechanism and lay the groundwork for future efforts to fully elucidate the allosteric mechanism. This chapter also presents what is known about CsoR structure and function and motivates the driving questions behind the studies presented in this thesis.

In addition, computational techniques and methodologies that are developed and applied to study Mtb-CsoR like system, but have more general applicability in the study of other proteins with allosteric control mechanisms or multiple ligand binding sites are also described in Chapter 2. Finally, the results

presented for Mtb-CsoR are used to make connections with other proteins that belong to the CsoR family to develop a more general understanding of these systems.

ALLOSTERY IN HPV VACCINE DESIGN

In Chapter 4, the dynamics of Human papilloma virus (HPV) virus-like particle (VLP) were studied and a new vaccine design strategy was proposed by exploiting the dynamic allosteric coupling between two strongly correlated immunogenic sites in HPV VLP. In general, most currently available vaccines are based on a 'natural' form of the pathogen, which is rendered either non- or weakly pathogenic by killing, inactivating or attenuating the pathogen, or by inclusion in the vaccine of selected antigenic components of the pathogen (subunit vaccines).[74,75] In essence, the vaccine is designed to be as similar as possible to the full or naturally occurring form of the pathogen.[76] Epitope-based vaccines provide a new strategy for therapeutic application of pathogen-specific immunity while decreasing the risks associated with the natural vaccines that contain the viral genetic material.[77-80] In this thesis, a VLP based vaccine design approach was presented where the viral capsid protein structures of HPV mimic the organization and conformation of authentic native viruses but lack the viral genome, potentially yielding safer and cheaper vaccine candidates. A handful of

VLP-based vaccines are currently commercialized worldwide: GlaxoSmithKline's Engerix (hepatitis B virus)[81] and Cervarix (HPC),[82-85] and Merck and Co., Inc.'s Recombivax HB (hepatitis B virus)[86] and Gardasil (human papillomavirus)[87,88] are some examples. The HPV type 16 VLP surface, has outwardly projecting loops containing epitopes that elicit the production of type-specific antibodies and immune response.[89-91] Several factors should be considered in the design of a new vaccine and the most critical factor is the orientation and presentation of the epitopes to the host immune system. Previous experimental work on the HPV VLP's elaborated the implications of deleting certain regions other than the epitopes from the VLP capsid pentamer suggesting a long distance control over epitope conformation within the VLP.[92,93] Recent computational studies on HPV VLP propose fluctuation-immunogenicity hypothesis wherein the immunogenicity is anti-correlated with epitope loop fluctuation intensity[94] suggesting that the immunogenicity/antibody-binding is correlated with the size of VLP. In this research, considering the insights from previous experimental and computational work, we tried to predict the relation between the assembly size of the VLP and epitope dynamics which is anticipated to be linked with the extent of immune response. Molecular dynamics simulations and associated analysis on these systems traced us to the concept of dynamic allostery which is in interplay between the h4 helices that stabilize the higher assemblies of capsid

proteins and epitopes of the VLP. Based on this concept, a vaccine design strategy has been proposed in this research, where the dynamics of the h4 helix are controlled by providing a nanoparticle support and the subsequent changes in the epitope fluctuations are studied. In Chapter 4, we describe how the principles of allostery are applied in the design of hybrid nanoparticle based vaccines and suggest a computational model of a new generation vaccine.

Each of the subsequent chapters of this thesis focuses on different aspects of allostery and understanding as well as exploiting this phenomenon to design new generation drugs and vaccines. Chapter 2 of this thesis, demonstrates the application of MD simulations and discusses about the specific approaches required to understand allosteric mechanisms. The force field development protocol and new force fields that were devised to simulate the two protein systems are also discussed in Chapter 2. These force fields are especially required to simulate the proteins in solution with their corresponding ligands and also enable us to analyze the change in protein conformation and dynamics in the presence and absence of the ligand. The MD simulation and computational analyses presented in Chapters 3 and 4 identify a network of specific residue contacts or correlations between specific epitope domains that connect the perturbation site of these proteins to the active site.

REFERENCES

(1) Kuriyan, J.; Eisenberg, D., *Nature* **2007,** *450*, 983-990.

(2) Goodey, N. M.; Benkovic, S. J., *Nat. Chem. Biol.* **2008,** *4*, 474-482.

(3) Tsai, C. J.; Del Sol, A.; Nussinov, R., *Mol. BioSyst.* **2009,** *5*, 207-216.

(4) Tsai, C.-J.; Del Sol, A.; Nussinov, R., *J. Mol. Biol.* **2008,** *378*, 1-11.

(5) Hill, A. V., *Biochem. J.* **1913,** *7*, 471.

(6) Adair, G. S., *J. Biol. Chem.* **1925,** *63*, 529-545.

(7) Wyman, J.; Allen, D. W., *J. Polym. Sci.* **1951,** *7*, 499-518.

(8) Monod, J.; Wyman, J.; Changeux, J.-P., *J. Mol. Biol.* **1965,** *12*, 88-118.

(9) Koshland Jr, D.; Nemethy, G.; Filmer, D., *Biochemistry* **1966,** *5*, 365-385.

(10) Perutz, M. F., *Q. Rev. Biophys.* **1989,** *22*, 139-237.

(11) Perutz, M. F., *Nature* **1970,** *228*, 726-734.

(12) Perutz, M. F.; Wilkinson, A.; Paoli, M.; Dodson, G., *Annu. Rev. Biophys. Biomol. Struct.* **1998,** *27*, 1-34.

(13) Pauling, L., *Proc. Natl. Acad. Sci. U. S. A.* **1935,** *21*, 186.

(14) Wyman, J., *Adv. Protein Chem* **1964,** *19*, 91.

(15) Edelstein, S. J., *Annu. Rev. Biochem.* **1975,** *44*, 209-232.

(16) Bohr, C.; Hasselbalch, K.; Krogh, A., *Skand. Arch. Physiol.* **1904,** *16*, 402-412.

(17) Perutz, M.; Bolton, W.; Diamond, R.; Muirhead, H.; Watson, H., *Nature* **1964,** *203*, 687-690.

(18) Monod, J.; Changeux, J.-P.; Jacob, F., *J. Mol. Biol.* **1963,** *6*, 306-329.

(19) Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D., *Science* **2001,** *291*, 2429-2433.

(20) Liu, M. S.; Todd, B.; Yao, S.; Feng, Z. P.; Norton, R. S.; Sadus, R. J., *Proteins: Struct., Funct., Bioinf.* **2008,** *73*, 218-227.

(21) Tripathi, S.; Portman, J. J., *J. Phys. Chem. B* **2013,** *117*, 13182-13193.

(22) Otten, R.; Villali, J.; Kern, D.; Mulder, F. A., *J. Am. Chem. Soc.* **2010,** *132*, 17004-17014.

(23) Tjandra, N.; Kuboniwa, H.; Ren, H.; Bax, A., *Eur. J. Biochem.* **1995,** *230*, 1014-1024.

(24) Malmendal, A.; Evenäs, J.; Forsén, S.; Akke, M., *J. Mol. Biol.* **1999,** *293*, 883-899.

(25) Miyashita, O.; Wolynes, P. G.; Onuchic, J. N., *J. Phys. Chem. B* **2005,** *109*, 1959-1969.

(26) Frauenfelder, H.; McMahon, B. H.; Austin, R. H.; Chu, K.; Groves, J. T., *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98*, 2370-2374.

(27) Okazaki, K.-i.; Takada, S., *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105*, 11182-11187.

(28) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G., *Annu. Rev. Phys. Chem.* **1997,** *48*, 545-600.

(29) Zhuravlev, P. I.; Papoian, G. A., *Q. Rev. Biophys.* **2010,** *43*, 295-332.

(30) Tsai, C.-J.; Ma, B.; Nussinov, R., *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96*, 9970-9972.

(31) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G., *J. Mol. Biol.* **2005,** *346*, 1121-1145.

(32) Chung, H. S.; McHale, K.; Louis, J. M.; Eaton, W. A., *Science* **2012,** *335*, 981-984.

(33) Cooper, A.; Dryden, D., *Eur. Biophys. J* **1984,** *11*, 103-109.

(34) Chennubhotla, C.; Yang, Z.; Bahar, I., *Mol. BioSyst.* **2008,** *4*, 287-292.

(35) Gunasekaran, K.; Ma, B.; Nussinov, R., *Proteins: Struct., Funct., Bioinf.* **2004,** *57*, 433-443.

(36) Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G., *Nat. Struct. Mol. Biol.* **2006,** *13*, 831-838.

(37) Swain, J. F.; Gierasch, L. M., *Curr. Opin. Struct. Biol.* **2006,** *16*, 102-108.

(38) Kar, G.; Keskin, O.; Gursoy, A.; Nussinov, R., *Curr. Opin. Pharmacol.* **2010,** *10*, 715-722.

(39) Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R., *Protein Sci.* **2002,** *11*, 184-197.

(40) Ruschak, A. M.; Kay, L. E., *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109*, E3454-E3462.

(41) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R., *Protein Sci.* **2000,** *9*, 10-19.

(42) Kurzyński, M., *Phys. A (Amsterdam, Neth.* **2000,** *285*, 29-47.

(43) Kern, D.; Zuiderweg, E. R., *Curr. Opin. Struct. Biol.* **2003,** *13*, 748-757.

(44) Wand, A. J., *Nat. Struct. Mol. Biol.* **2001,** *8*, 926-931.

(45) Popovych, N.; Tzeng, S.-R.; Tonelli, M.; Ebright, R. H.; Kalodimos, C. G., *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 6927-6932.

(46) Toncrova, H.; McLeish, T. C., *Biophys. J.* **2010,** *98*, 2317-2326.

(47) Yang, J.; Garrod, S. M.; Deal, M. S.; Anand, G. S.; Woods Jr, V. L.; Taylor, S., *J. Mol. Biol.* **2005,** *346*, 191-201.

(48) Hawkins, R. J.; McLeish, T. C., *Biophys. J.* **2006,** *91*, 2055-2062.

(49) Bahar, I.; Chennubhotla, C.; Tobi, D., *Curr. Opin. Struct. Biol.* **2007,** *17*, 633-640.

(50) Clarkson, M. W.; Gilmore, S. A.; Edgell, M. H.; Lee, A. L., *Biochemistry* **2006,** *45*, 7693-7699.

(51) Clarkson, M. W.; Lee, A. L., *Biochemistry* **2004,** *43*, 12448-12458.

(52) Grossoehme, N. E.; Giedroc, D. P., In *Spectroscopic Methods of Analysis*, Springer: 2012; pp 165-192.

(53) Giedroc, D. P.; Arunkumar, A. I., *Dalton Trans.* **2007**, 3107-3120.

(54) Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J., *Nature* **2009,** *460,* 823-830.

(55) Waldron, K. J.; Robinson, N. J., *Nat. Rev. Microbiol.* **2009,** *7*, 25-35.

(56) Reyes-Caballero, H.; Campanello, G. C.; Giedroc, D. P., *Biophys. Chem.* **2011,** *156*, 103-114.

(57) Browning, D. F.; Busby, S. J. W., *Nat. Rev. Microbiol.* **2004,** *2*, 57-65.

(58) Hantke, K., *Curr. Opin. Microbiol.* **2001,** *4*, 172-177.

(59) Lefstin, J. A.; Yamamoto, K. R., *Nature* **1998,** *392*, 885-888.

(60) Ogata, K.; Sato, K.; Tahirov, T., *Curr. Opin. Struct. Biol.* **2003,** *13*, 40-48.

(61) Ptashne, M.; Gann, A., *Nature* **1997,** *386*, 569-577.

(62) Chen, P. R.; He, C., *Curr. Opin. Chem. Biol.* **2008,** *12*, 214-221.

(63) Pennella, M. A.; Giedroc, D. P., *BioMetals* **2005,** *18*, 413-428.

(64) Busenlehner, L. S.; Pennella, M. A.; Giedroc, D. P., *FEMS Microbiol. Rev.* **2003,** *27*, 131-143.

(65) Liu, T.; Ramesh, A.; Ma, Z.; Ward, S. K.; Zhang, L. M.; George, G. N.; Talaat, A. M.; Sacchettini, J. C.; Giedroc, D. P., *Nat. Chem. Biol.* **2007,** *3*, 60-68.

(66) Fu, Y.; Tsui, H. C. T.; Bruce, K. E.; Sham, L. T.; Higgins, K. A.; Lisher, J. P.; Kazmierczak, K. M.; Maroney, M. J.; Dann, C. E.; Winkler, M. E.; Giedroc, D. P., *Nat. Chem. Biol.* **2013,** *9*, 177-183.

(67) Brown, N. L.; Stoyanov, J. V.; Kidd, S. P.; Hobman, J. L., *FEMS Microbiol. Rev.* **2003,** *27*, 145-163.

(68) Ma, Z.; Jacobsen, F. E.; Giedroc, D. P., *Chem. Rev. (Washington, DC, U. S.)* **2009,** *109*, 4644-4681.

(69) Ma, Z.; Cowart, D. M.; Scott, R. A.; Giedroc, D. P., *Biochemistry* **2009,** *48*, 3325-3334.

(70) Sakamoto, K.; Agari, Y.; Agari, K.; Kuramitsu, S.; Shinkai, A., *Microbiology* **2010,** *156*, 1993-2005.

(71) Dwarakanath, S.; Chaplin, A. K.; Hough, M. A.; Rigali, S.; Vijgenboom, E.; Worrall, J. A., *J. Biol. Chem.* **2012,** *287*, 17833-17847.

(72) Higgins, K. A.; Giedroc, D. P., *Chem. Lett.* **2013**.

(73) Coyne III, H. J.; Giedroc, D. P., *Biomol. NMR Assignments* **2013,** *7*, 279-283.

(74) Theiler, M.; Smith, H. H., *J. Exp. Med.* **1937,** *65*, 787-800.

(75) Plotkin, S. L.; Plotkin, S. A., *Vaccines* **2004,** *5*, 1-16.

(76) Plotkin, S. A., *Clin. Infect. Dis.* **2008,** *47*, 401-409.

(77) Correia, B. E.; Bates, J. T.; Loomis, R. J.; Baneyx, G.; Carrico, C.; Jardine, J. G.; Rupert, P.; Correnti, C.; Kalyuzhniy, O.; Vittal, V., *Nature* **2014**.

(78) Correia, B. E.; Ban, Y.-E. A.; Holmes, M. A.; Xu, H.; Ellingson, K.; Kraft, Z.; Carrico, C.; Boni, E.; Sather, D. N.; Zenobia, C., *Structure* **2010,** *18*, 1116-1126.

(79) Sette, A.; Livingston, B.; McKinney, D.; Appella, E.; Fikes, J.; Sidney, J.; Newman, M.; Chesnut, R., *Biologicals* **2001,** *29*, 271-276.

(80) Sette, A.; Fikes, J., *Curr. Opin. Immunol.* **2003,** *15*, 461-470.

(81) Keating, G. M.; Noble, S., *Drugs* **2003,** *63*, 1021-1051.

(82) Crosbie, E. J.; Kitchener, H. C., *Expert Opin. Biol. Ther.* **2007,** *7*, 391-396.

(83) Le Tallec, D.; Doucet, D.; Elouahabi, A.; Harvengt, P.; Deschuyteneer, M.; Deschamps, M., *Hum. Vaccines* **2009,** *5*, 467-474.

(84) Monie, A.; Hung, C. F.; Roden, R.; Wu, T. C., *Biologics* **2008,** *2*, 97-105.

(85) Szarewski, A., *Expert Opin. Biol. Ther.* **2010,** *10*, 477-487.

(86) Venters, C.; Graham, W.; Cassidy, W., *Expert. Rev. Vaccines* **2004,** *3*, 119.

(87) Brown, D. R.; Garland, S.; Ferris, D. G.; Joura, E.; Steben, M.; James, M.; Radley, D.; Vuocolo, S.; Garner, E. I. O.; Haupt, R. M.; Bryan, J. T., *Hum. Vaccines* **2011,** *7*, 230-238.

(88) Smith, J. F.; Brownlow, M.; Brown, M.; Kowalski, R.; Esser, M. T.; Ruiz, W.; Barr, E.; Brown, D. R.; Bryan, J. T., *Hum. Vaccines* **2007,** *3*, 109-115.

(89) Chen, X. S.; Garcea, R. L.; Goldberg, I.; Casini, G.; Harrison, S. C., *Mol. Cell* **2000,** *5*, 557-567.

(90) Nardelli-Haefliger, D.; Roden, R.; Benyacoub, J.; Sahli, R.; Kraehenbuhl, J.-P.; Schiller, J. T.; Lachat, P.; Potts, A.; De Grandi, P., *Infect. Immun.* **1997,** *65*, 3328-3336.

(91) Day, P. M.; Gambhira, R.; Roden, R. B.; Lowy, D. R.; Schiller, J. T., *J. Virol.* **2008,** *82*, 4638-4646.

(92) Ryding, J.; Dahlberg, L.; Wallen-Ohman, M.; Dillner, J., *J. Gen. Virol.* **2007,** *88*, 792-802.

(93) White, W. I.; Wilson, S. D.; Palmer-Hill, F. J.; Woods, R. M.; Ghim, S. J.; Hewitt, L. A.; Goldman, D. M.; Burke, S. J.; Jenson, A. B.; Koenig, S.; Suzich, J. A., *J. Virol.* **1999,** *73*, 4882-4889.

(94) Joshi, H.; Cheluvaraja, S.; Somogyi, E.; Brown, D. R.; Ortoleva, P., *Vaccine* **2011,** *29*, 9423-9430.

COMPUTATIONAL MODELING OF ALLOSTERIC PROTEINS

Allosteric transitions, in which binding of an effector molecule to one site of a protein is coupled to a conformational change at a distant site or the change in the dynamics of specific domains correlate to the fluctuation of other domains, are examined using computational methods in this thesis. Although the concept of allostery was proposed more than 40 years ago,[1,2] developing a mechanistic understanding of different classes of allostery continues to be an active and dynamic area of research. The X-ray crystal structures of ligand bound and unbound states have provided insight into the structural transitions underlying allostery in small number of allosteric proteins.[3-6] However, the static nature of these structures present several important challenges for structural biology that can be approached using computational methods. To study allostery, atomistic level details are required to decipher how the ligand binding or the altered dynamics information is passed between two distant sites through a protein. This transfer of information involves both side chain groups and the polypeptide backbone scaffold. The inherent complexity of the allosteric phenomenon requires multiple approaches to obtain the necessary detail for a complete mechanistic understanding of allosteric transitions.

NMR spectroscopy is a particularly useful tool for investigating protein motions because it is sensitive to molecular motion over a wide variety of timescales.[7-9] Solution NMR experiments can be performed under physiological conditions of solvent, pH, and temperature, thereby making a more direct correlation to the in vivo function however, detailed mechanistic information on allostery is not easily obtained from the aforementioned NMR experiments.[10] This mechanistic detail is essential for a fundamental understanding of the allosteric behavior and, is crucial for aiding in the design of allosteric drugs and vaccines that interact with the target, altering the allostery and in turn impede the function of the protein. The atomistic detail on amino acid motions derived from molecular dynamics (MD) simulations is a powerful complement to solution NMR and crystallographic studies to shed further light on allosteric mechanisms. Various methods employed to study the protein motions of are shown in Figure 1.

In this chapter, the computational methodologies applied in studying the allostery are discussed with the focus on the derivation of all-atom molecular mechanics force fields for modeling metal sensor proteins and silica nanoparticle tethered HPV VLP systems.

MOLECULAR DYNAMICS SIMULATIONS OF ALLOSTERIC PROTEINS

Molecular dynamics simulations emerged as a very useful tool to study time dependent processes/motions in molecular systems.[11-13] In bio-systems each of these dynamic motions have a characteristic time-scale, amplitude and energy range. Macromolecules in general, and proteins in particular, display a broad range of characteristic motions that are either very fast and very localized, such as atomic fluctuation or slow motions that occur on the scale of the whole molecule, such as the folding transition.[14,15] Many of these motions have an important role in the biochemical function of the protein and lead to the allosteric transitions.[14,15] Furthermore, various small- or medium-scale protein motions are coupled to one another that lead to large-scale dynamic transitions.[14,16,17] The time scales of several of these motions are listed in Table Error! No text of specified style in document.**.1**.[18]
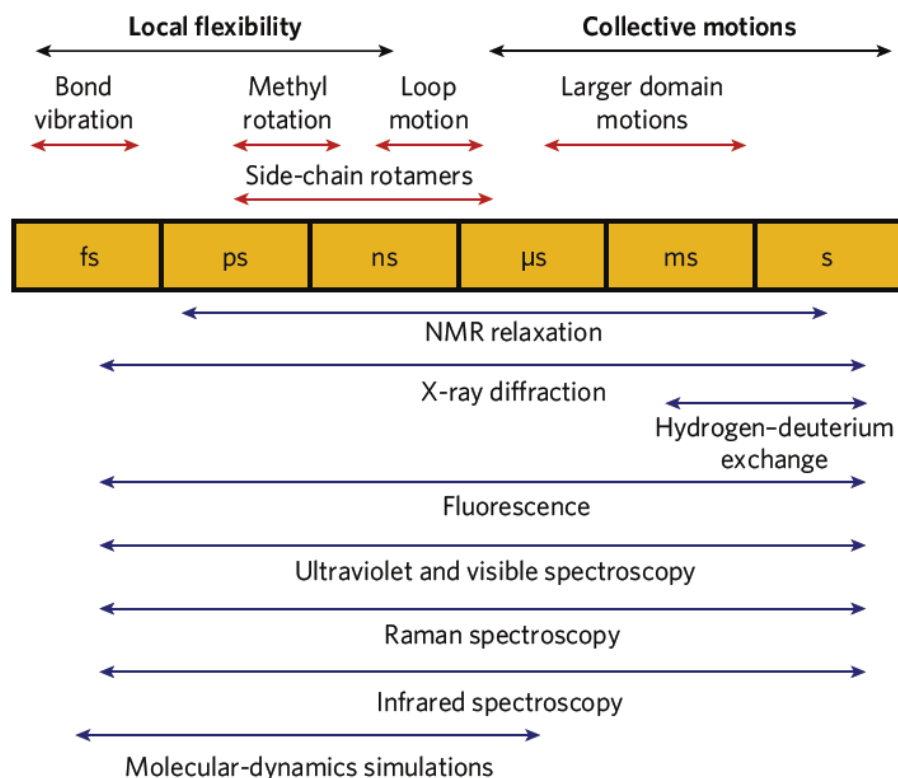
Figure Error! No text of specified style in document..**1.** List of different methods available to study protein dynamic modes based on their time scale[18]

Table Error! No text of specified style in document..**1.** Classification of protein dynamic modes based on their time scale

Local Motions:	fs - ps (10^{-15} - 10^{-12} s)
Atomic fluctuation and Side chain motion	
Medium Scale Motions:	ns - μs (10^{-9} - 10^{-6} s)
Loop motion, Terminal-arm motion and Rigid-body motion (helices)	
Large Scale Motions:	μs - ms (10^{-6} - 10^{-3} s)
Domain motion and Subunit motion	
Global Motions:	ms - h (10^{-3} - 10^{4} s)
Helix-coil transition, Folding/unfolding and Subunit association	

Allosteric transitions can often result in conformational changes of significant magnitude that take place in the μs–ms time scale, a time regime that cannot currently be sampled by means of standard MD techniques.[19-21] However, the binding of the allosteric effector molecule increases the population of active conformations by modulating certain intermediate protein motions at different time scales and accelerating the allosteric transition. Standard MD simulations can be used to analyze the thermal motion of atomic sites in the ps–ns time scale. In this time scale, several protein motions take place, including side chain motion of hindered and unhindered surface residues, methyl group rotations, loop motion, collective motion of a few residues, helix–coil transitions, and folding of small peptides as listed in Table Error! No text of specified style in document.**.1**. Modification of these protein motions caused by effector binding can be captured by correlation

analysis of motions observed along MD trajectories and combined with observations obtained by experimental NMR or crystallographic data.[17,20,21] For the purposes of this work, focus is placed on the characterization of fast (picosecond - 10 nanosecond) time scale motions that may collectively lead to the allosteric transition.

MD BASED ANALYSIS

The collective motions within proteins that lead to allosteric transitions can be derived from atomic interactions from a crystal structure, from the comparison of two molecular conformations or from the analysis of several conformations of the protein in its active and inactive states, as can be generated by the MD simulations. The analysis of these collective motions of the allosteric proteins can be used to investigate the conformational energy landscape, to improve the sampling space and in the refinement of X-ray and NMR data.[22,23] Computational analysis of collective behavior must be based on knowledge of the structural fluctuations that occur as a result of thermal motion in the protein and can be obtained using different approaches as discussed in the following sections.[24-28]

PRINCIPAL COMPONENT ANALYSIS

MD simulations generate an overwhelming amount of information contained in the trajectory of atomic coordinates. In order to extract the concerted fluctuations with large amplitudes from the trajectory, a principal component analysis (PCA) can be carried out on a large number of configurations chosen from the MD trajectory.[29-31] PCA involves diagonalization of the covariance matrix of atomic fluctuations to yield collective variables that are sorted based on their contribution to the total mean-square fluctuation.[32-36] For studies that focus on relating large-scale motions to function, however, the computational task can be reduced by selecting only backbone or Cα atoms for the PCA. This analysis is often termed as essential dynamics analysis (EDA).[37,39]

NORMAL MODE ANALYSIS

Normal mode analysis (NMA) is one another major simulation techniques used to probe the large-scale, shape-changing motions in biological molecules.[40-42] Although it has a connection to the experimental techniques of infrared and Raman spectroscopy, its recent application has been to predict functional motions in proteins or other biological molecules.[43,44] Functional motions are those that relate to function and are often the consequence of binding other molecules. In NMA studies, it is always assumed that the normal modes with the largest

fluctuation (lowest frequency modes) are the ones that are functionally relevant, because, like function, they exist by evolutionary design.[45,46] The ultimate validation of the functionally relevant mode must come from comparisons with experimental data and indeed studies that compare predictions of NMA with transitions derived from multiple X-ray conformers do suggest that the low-frequency normal modes are often functionally relevant.[42,47] A Hessian matrix is created, which is the matrix of second derivatives of the potential energy function with respect to the mass-weighted Cartesian coordinates. This stage determines the eigenvalues and eigenvectors. Because of the large size of this 3N×3N matrix, where N is the number of atoms in the molecule, this stage presents memory problems for large molecules. The process results in a set of 3N eigenvalues and a set of 3N eigenvectors each with 3N components. The eigenvalues are sorted in ascending order and the eigenvectors are sorted accordingly. The first six eigenvalues should have values close to zero because these correspond to the three translational and three rotational degrees of freedom for the whole molecule. The seventh eigenvector is the lowest frequency mode, and it is often predicted to be a functionally relevant mode.[48,49]

MOLECULAR DYNAMICS SIMULATIONS AND MOLECULAR MECHANICS FORCE FIELDS

Molecular dynamics (MD) is a powerful scientific method to study the properties, conformations, internal atomic motions and bio-molecular fluctuations of molecules as a function of time in atomistic level of detail.[50-54] In molecular mechanics methods the smallest unit is the atom, and all the atoms are considered to be spherical with particular radii (derived from experiments or theory) and have a net charge.[50-52] The interactions between these spherical atom-like particles are defined by classical potentials and these interactions are pre-assigned to obtain spatial distribution and energy of atom-like particles.[51,52] A simple molecular mechanics potential energy equation is given by Eq. (1).

Energy = Bond Stretching Energy + Angle bending Energy + Torsional Energy + Non-Bonded Interaction Energy   2.1

Equation 2.1, substituted with the required parameters describe the behavior of different kinds of atoms and bonds, is called a force field. Many different kinds of force fields have been developed over the years which are associated with various MD packages. Some include additional energy terms that describe other kinds of deformations. In this study both GROMOS[55-57] and CHARMM[58] force fields are used and new force fields were parameterized to

model the system of interest. The mathematical form of the energy terms varies from force-field to force-field. Hence the more common and general energy terms are described in this chapter.

BOND STRETCHING

The function for defining bond stretching is given in Eq. (2.2).

$$E = \sum_{bonds} k_b (r - r_o)^2 \qquad 2.2$$

The bond stretching energy equation is based on Hooke's law. The "k_b" force constant controls the stiffness of the bond, while "r_o" defines the equilibrium bond length. Unique "k_b" and "r_o" parameters are assigned to each pair of bonded atoms based on the type of bond and bond length. This equation calculates the energy associated with vibration about the equilibrium bond length.

ANGLE BENDING

The function for determining the angle bending is given in Eq. (3).

$$E = \sum_{angles} k_\theta (\theta - \theta_o)^2 \qquad 2.3$$

The angle bending energy equation is also based on Hooke's law. The "k_θ" is the force constant that controls the stiffness of the angle spring, while "θ_o" defines its

equilibrium angle. Equation 3 estimates the energy associated with vibration about the equilibrium angle.

TORSIONAL ENERGY

Proper Dihedral: The rotation of 'R'-groups around a bond is defined by a periodic equation, as the rotation is periodic every n°. In order to avoid negative terms this equation uses a cosine function as shown in Eq. (2.4).

$$E = \sum_{torsion} A[1 + cos(n\tau - \varphi)] \qquad 2.4$$

The "A" parameter controls the amplitude of the curve, the "n" parameter controls its periodicity, and "φ"shifts the entire curve along the rotation angle axis (τ)."n" reflects the type symmetry in the dihedral angle. For example, a CH3-CH3 bond, repeats its energy every 120°. The cis conformation of a dihedral angle is assumed to be the zero torsional angle by convention.

Improper Dihedral: An improper dihedral is required to preserve the planarity in planar groups (like aromatic rings). This dihedral type can also be used to prevent interconversion of stereocenters. The main difference between dihedral and improper angles is that, traditionally, the order of atoms making up a dihedral angle follows the lineup of covalent bonds. Whereas, in an improper angle the sequence of atoms often permutes the succession of chemical bonds, so

that the axis of an improper rotation does not necessarily coincide with a chemical bond. A virtual torsion (or improper dihedral) angle describes a torsion in terms of geometry, but not in terms of chemical bond and the functional form used is represented in Eq. (2.5). The most important modification when moving to coordination complexes is that the plane of the ligands need to be exactly defined. For example, in square planar complexes it is necessary to define an average plane through the ligands. Alternatives to the regular out-of-plane function for square planar complexes have included deviation of the metal from the coordination plane,[59] functions of the angle of each ligand with the normal to the coordination plane,[60] or dummy atoms placed at the axial positions.[61,62]

$$E = \sum_{improper} k_\delta (\delta - \delta_o)^2 \qquad\qquad 2.5$$

In addition to these terms, the CHARMM force field has one additional term; that is the Urey-Bradley term,[51] which is an interaction based on the distance between atoms separated by two bonds(1,3 interaction).

NON-BONDED ENERGY

The non-bonded energy represents the pair-wise sum of the energies of all possible interacting non-bonded atoms "i" and "j". Non-bonded potential includes van-der-Waals and electrostatic potentials. The van-der-Waals potentials take into

account repulsion between atoms at small separations and weak attraction at larger distances. The common form of this potential for a pair of atoms "i" and "j" is given by a Lennard-Jones function E LJ is represented in Eq. (2.6).

$$E_{LJ} = \sum_i \sum_j \frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} \qquad 2.6$$

Van der Waals attraction occurs at short distances between atoms, and rapidly dies off as the interacting atoms move apart by a few Angstroms. Repulsion occurs when the distance between interacting atoms becomes slightly less than the sum of their contact radii. The "A" and "B" parameters control the depth and position (interatomic distance) of the potential energy well for a given pair of non-bonded interacting atoms. The repulsion at small separations between atoms is associated with the Pauli Exclusion Principle, while weak attraction at larger distances is due to London dispersion interactions.

The electrostatic contribution is modeled using a coulombic potential which is represented in Eq. (2.7).

$$E_{coloumb} = \sum_i \sum_j \frac{q_i q_j}{R_{ij}} \qquad 2.7$$

The electrostatic energy is a function of the charge on the non-bonded atoms, their interatomic distance, and a molecular dielectric expression that accounts for the

attenuation of electrostatic interaction by the environment (e.g. solvent or the molecule itself). Often, the molecular dielectric is set to a constant value between 1.0 and 5.0. A linearly varying distance-dependent dielectric (i.e. 1/R) is sometimes used to account for the increase in environmental bulk as the separation distance between interacting atoms increases. Partial atomic charges can be calculated for small molecules using an *ab-initio* or semi-empirical quantum techniques. The final potential energy function for the force field description is given in Eq. (2.8).

$$E = \sum_{bonds} k_b(r - r_o)^2 + \sum_{angles} k_\theta(\theta - \theta_o)^2 + \sum_{torsion} A[1 + cos(n\tau - \varphi)] + \sum_{improper} k_\delta(\delta - \delta_o)^2 + \sum_i \sum_j \frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \sum_i \sum_j \frac{q_i q_j}{R_{ij}} \qquad 2.8$$

FORCE FIELD PARAMETERIZATION AND OPTIMIZATION

In the previous section, the potential energy function that determines the force field, and the required parameters were described. This chapter focuses on determining the parameter values for previously undefined or poorly defined systems. The choice of force field, the potential energy functions and their parameterization, is solely determined by the system of study. The two possibilities which are considered during parameterizing a new compound are general force fields[63] which cover wide areas of chemistry such as the CGenFF[64-66], UFF[67] etc. or force fields developed for a specific type of compound. The latter

types lack transferability and general applicability, but are supposed to be accurate, as they are built to suit the specific system.[68] For most applications the essential task, and the first step to consider, is to accurately reproduce structures, and these are also the basis for an exact computation of the corresponding spectroscopic and thermodynamic properties of inorganic and bioinorganic compounds.[69-72]

During the process of parameterization of new force field the most important factor to consider is the goal of the force field and accuracy that is needed for structures. For some applications, a reasonable structure is all that is needed. If so, it is even possible that an existing force field with generalized parameters can fulfill the goal.[67] Type of the functional form needed is very dependent upon the goal and also upon the type of complex to be modeled. Reasonable structures of metal complexes can be achieved with relatively simple functional forms, but sterically induced distortions in coordinatively unfilled complexes (e.g. distortions from planarity) require an accurate and flexible definition of metal-centered angles. Especially, when the system of interest involves metal coordinating with a protein, the coordination of the ligands may be unusual. These types of complexes require extra terms in the potential energy function to enforce unusual geometry at the metal center. For example, improper dihedrals and out of plane bending terms are introduced to maintain the planarity
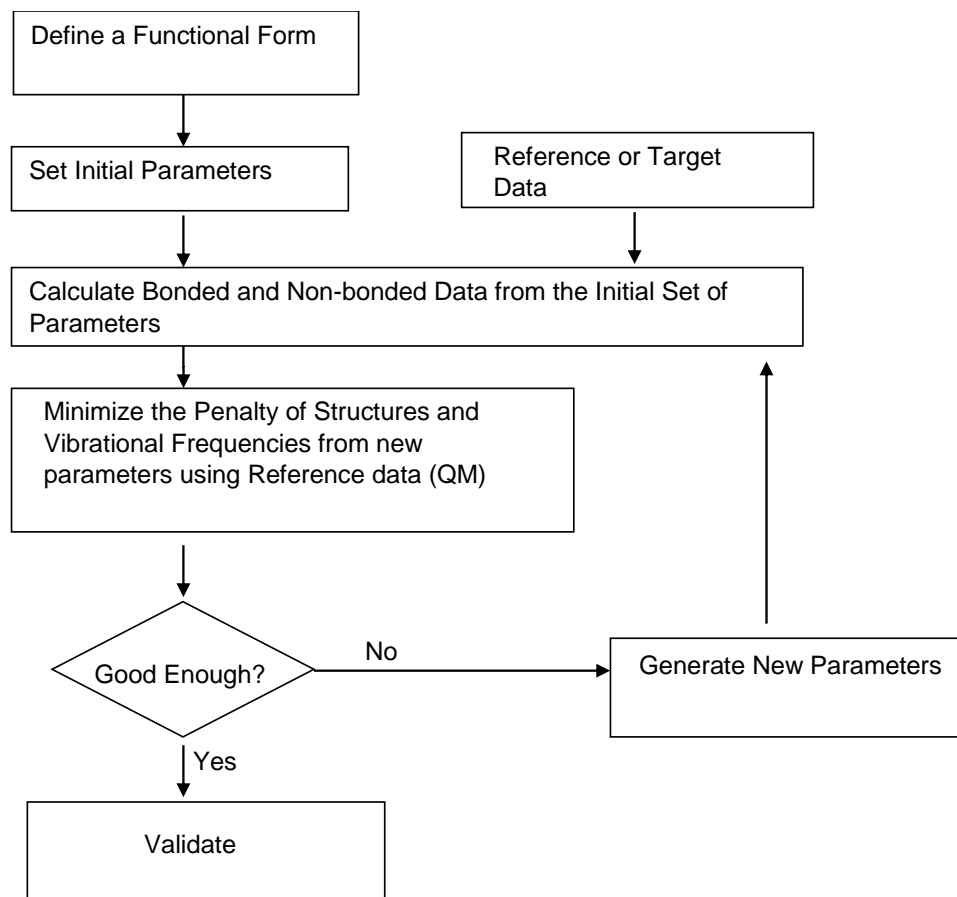
of the metal complex as well as the orientation of the rings while coordinating with the metal.

Assigning an initial set of values for unknown parameters is very important and this step controls the tediousness of this process. Most of the MD simulation programs are provided with modules that will identify atoms with missing parameters and provide an initial first guess of values.[73-75] These programs use a pattern recognition algorithm to guess the approximate parameter, and these parameters should be checked with chemical intuition. But generating an initial set of parameters from these programs is always a better estimate than to start with random numbers. Since the new parameters are usually estimated by one's intuition and then optimized, a second consideration in developing force fields is the selection of reference or target data that is used to improve the new parameters. It must always be remembered that force fields are empirical, and can never be better than the data used to create the force field. The parameters needed to extend the force field for a new system should fit into the different terms of the force field that is chosen,[68] as these parameters are generally optimized using different target/source data. The internal parameters such as equilibrium bond lengths, equilibrium angle values, and dihedral multiplicity are often adjusted to reproduce gas-phase geometric data obtained from quantum mechanical calculations, electron diffraction, crystallographic data or microwave

experiments.[51] The force constants, such as the bond length and angle force constants are usually deduced using vibrational spectra or calculations, which contain individual frequencies and their assignments.[76] The approaches for deriving electrostatic parameters mostly include reproduction of target data from QM calculations.[77,78] One method of obtaining partial atomic charges is to correlate the calculated electrostatic potential (ESP) of the molecule calculated with QM methods.[77,78] A popular variation of this method is called the RESP method,[79] where the charges on atoms that are minimally exposed to solvent are restrained. The goal of both methods is to produce the partial atomic charges that reproduce the electrostatic potential created by the molecule.[80] Automatic parameterization procedures have been attempted in this work; however, a significant amount of manual interference is generally required. A third consideration during parameterization is the procedure used to optimize the force field. When a force field is selected for use, the information that one wishes to extract must be considered.[81] For example, if one is interested in examining the atomic details of water interactions with protein residues, the proper force field to use would be an all-atom force field designed specifically for biomolecules that allows for explicit representation of water molecules (as opposed to an extended atom force field where hydrogens are not explicitly represented but treated as part of the atom to which they are bound).

A flow chart outlining the parameterization and automation procedure is shown in Figure Error! No text of specified style in document..**2**. Though the chart looks linear and simple several cycles of optimization have to be performed to validate the new force fields. For example, the functional form is frequently updated during the process, by adding, removing, or changing specific terms. The quality and accuracy obtained from the simulations with a newly parameterized force field is obviously based on the methods and target data which are used in the optimization of the derived parameters.[68] Any new compounds to be used with the new force field must be tested in order to ensure that they are treated correctly and the required accuracy is obtained.[68] Generally, the extent of transferability is considered to be minimal, and new parameters must be generated each time a new class of molecules is being used with the force field.[68] In this case, the GROMOS force fields[55] were used and extended to model the Cu(I) coordination in the copper sensing proteins and CHARMM27 force fields[58] were adopted to model the silica surface tethered to the virus-like particle. Since silica force fields[82] are developed using CHARMM package, applying CHARMM force fields to model this system is the suitable choice.

```
┌─────────────────────────────┐
│ Define a Functional Form    │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐      ┌──────────────────────┐
│ Set Initial Parameters      │      │ Reference or Target  │
└─────────────────────────────┘      │ Data                 │
              │                      └──────────────────────┘
              │                                │
              ▼                                ▼
┌────────────────────────────────────────────────────────────┐
│ Calculate Bonded and Non-bonded Data from the Initial Set of│
│ Parameters                                                  │
└────────────────────────────────────────────────────────────┘
              │
              ▼
┌─────────────────────────────────────┐
│ Minimize the Penalty of Structures  │
│ and Vibrational Frequencies from new│
│ parameters using Reference data (QM)│
└─────────────────────────────────────┘
              │
              ▼
        ╱╲
       ╱  ╲        No
      ╱ Good╲──────────────────▶ ┌──────────────────────────┐
      ╲Enough?╱                   │ Generate New Parameters  │
       ╲  ╱                       └──────────────────────────┘
        ╲╱
         │ Yes
         ▼
┌─────────────────────────────┐
│        Validate             │
└─────────────────────────────┘
```

Figure Error! No text of specified style in document.**.2.** Process flow used in the derivation of new parameters for developing the new MM force field

REFERENCES

(1) Monod, J.; Wyman, J.; Changeux, J.-P., *J. Mol. Biol.* **1965,** *12*, 88-118.

(2) Monod, J.; Changeux, J.-P.; Jacob, F., *J. Mol. Biol.* **1963,** *6*, 306-329.

(3) Bocquet, N.; Nury, H.; Baaden, M.; Le Poupon, C.; Changeux, J.-P.; Delarue, M.; Corringer, P.-J., *Nature* **2009,** *457*, 111-114.

(4) Schultz, S. C.; Shields, G. C.; Steitz, T. A., *J. Mol. Biol.* **1990,** *213*, 159-166.

(5) Parkinson, G.; Wilson, C.; Gunasekera, A.; Ebright, Y. W.; Ebright, R. E.; Berman, H. M., *J. Mol. Biol.* **1996,** *260*, 395-408.

(6) Schultz, S. C.; Shields, G. C.; Steitz, T. A., *Science* **1991,** *253*, 1001-1007.

(7) Lipari, G.; Szabo, A., *J. Am. Chem. Soc.* **1982,** *104*, 4546-4559.

(8) Lipari, G.; Szabo, A., *J. Am. Chem. Soc.* **1982,** *104*, 4559-4570.

(9) Kay, L. E., *Biochem. Cell Biol.* **1998,** *76*, 145-152.

(10) Manley, G.; Rivalta, I.; Loria, J. P., *J. Phys. Chem. B* **2013,** *117*, 3063-3073.

(11) Adcock, S. A.; McCammon, J. A., *Chem. Rev. (Washington, DC, U. S.)* **2006,** *106*, 1589-1615.

(12) Okazaki, K.-i.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G., *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103*, 11844-11849.

(13) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y., *Science* **2010,** *330*, 341-346.

(14) Berendsen, H. J.; Hayward, S., *Curr. Opin. Struct. Biol.* **2000,** *10,* 165-169.

(15) Kern, D.; Zuiderweg, E. R., *Curr. Opin. Struct. Biol.* **2003,** *13,* 748-757.

(16) Boehr, D. D.; Nussinov, R.; Wright, P. E., *Nat. Chem. Biol.* **2009,** *5,* 789-796.

(17) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R., *Protein Sci.* **2000,** *9,* 10-19.

(18) Henzler-Wildman, K.; Kern, D., *Nature* **2007,** *450,* 964-972.

(19) Elber, R., *Curr. Opin. Struct. Biol.* **2011,** *21,* 167-172.

(20) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E., *Curr. Opin. Struct. Biol.* **2009,** *19,* 120-127.

(21) Kidd, B. A.; Baker, D.; Thomas, W. E., *PLoS Comput. Biol.* **2009,** *5.*

(22) Okazaki, K.-i.; Takada, S., *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 11182-11187.

(23) Mouawad, L.; Perahia, D.; Robert, C. H.; Guilbert, C., *Biophys. J.* **2002,** *82,* 3224-3245.

(24) Abrams, C. F.; Vanden-Eijnden, E., *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107,* 4961-4966.

(25) De Groot, B. L.; Vriend, G.; Berendsen, H. J., *J. Mol. Biol.* **1999,** *286,* 1241-1249.

(26) Sherwood, P.; Brooks, B. R.; Sansom, M. S., *Curr. Opin. Struct. Biol.* **2008,** *18,* 630-640.

(27) Bahar, I.; Rader, A., *Curr. Opin. Struct. Biol.* **2005,** *15,* 586-592.

(28) Schlick, T., *F1000 Biol. Rep.* **2009,** *1*.

(29) Lange, O. F.; Grubmüller, H., *Proteins: Struct., Funct., Bioinf.* **2008,** *70*, 1294-1312.

(30) Skjaerven, L.; Martinez, A.; Reuter, N., *Proteins: Struct., Funct., Bioinf.* **2011,** *79*, 232-243.

(31) Lange, O. F.; Grubmüller, H., *Proteins: Struct., Funct., Bioinf.* **2006,** *62*, 1053-1061.

(32) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K., *J. Phys. Chem.* **1996,** *100*, 2567-2572.

(33) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A., *J. Mol. Biol.* **2009,** *385*, 312-329.

(34) Das, A.; Mukhopadhyay, C., *J. Chem. Phys.* **2007,** *127*, 165103.

(35) Tournier, A. L.; Smith, J. C., *Phys. Rev. Lett.* **2003,** *91*, 208106.

(36) Papaleo, E.; Mereghetti, P.; Fantucci, P.; Grandori, R.; De Gioia, L., *J. Mol. Graphics Modell.* **2009,** *27*, 889-899.

(37) Amadei, A.; Linssen, A.; Berendsen, H. J., *Proteins: Struct., Funct., Bioinf.* **1993,** *17*, 412-425.

(38) Arcangeli, C.; Bizzarri, A. R.; Cannistraro, S., *Biophys. Chem.* **2001,** *90,* 45-56.

(39) Chen, C.; Xiao, Y.; Zhang, L., *Biophys. J.* **2005,** *88*, 3276-3285.

(40) Case, D. A., *Curr. Opin. Struct. Biol.* **1994,** *4*, 285-290.

(41) Levitt, M.; Sander, C.; Stern, P. S., *J. Mol. Biol.* **1985,** *181*, 423-447.

(42) Lindahl, E.; Azuara, C.; Koehl, P.; Delarue, M., *Nucleic Acids Res.* **2006,** *34*, W52-W56.

(43) Bandekar, J., *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1992,** *1120*, 123-143.

(44) Cael, J.; Gardner, K.; Koenig, J.; Blackwell, J., *J. Chem. Phys.* **2008,** *62*, 1145-1153.

(45) Tirion, M. M., *Phys. Rev. Lett.* **1996,** *77*, 1905.

(46) Jääskeläinen, S.; Verma, C. S.; Hubbard, R. E.; Lestko, P.; Caves, L. S., *Protein Sci.* **1998,** *7*, 1359-1367.

(47) Venters, C.; Graham, W.; Cassidy, W., **2004**.

(48) Ichiye, T.; Karplus, M., *Proteins: Struct., Funct., Bioinf.* **1991,** *11*, 205-217.

(49) Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H., *Chem. Rev. (Washington, DC, U. S.)* **2009,** *110*, 1463-1497.

(50) Jensen, F., *Introduction to computational chemistry*. 2nd ed.; John Wiley & Sons: Chichester, England ; Hoboken, NJ, 2007.

(51) Burkert, U.; Allinger, N. L., *Molecular mechanics*. American Chemical Society: Washington, D.C., 1982.

(52) Hay, B. P., *Coord. Chem. Rev.* **1993,** *126*, 177-236.

(53) Hansson, T.; Oostenbrink, C.; van Gunsteren, W., *Curr. Opin. Struct. Biol.* **2002,** *12*, 190-196.

(54) Leach, A. R.; Schomburg, D., *Molecular modelling: principles and applications*. Longman London: 1996.

(55) Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F., *J. Comput. Chem.* **2005,** *26*, 1719-1751.

(56) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., *J. Comput. Chem.* **2005,** *26*, 1701-1718.

(57) Soares, T. A.; Daura, X.; Oostenbrink, C.; Smith, L. J.; van Gunsteren, W. F., *J. Biomol. NMR* **2004,** *30*, 407-422.

(58) Brooks, B. R.; III, C. L. B.; Jr, A. D. M.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., *J. Comput. Chem.* **2009,** *30*, 1545-1614.

(59) Adam, K. R.; Antolovich, M.; Brigden, L. G.; Lindoy, L. F., *J. Am. Chem. Soc.* **1991,** *113*, 3346-3351.

(60) Allured, V. S.; Kelly, C. M.; Landis, C. R., *J. Am. Chem. Soc.* **1991,** *113*, 1-12.

(61) Sabolović, J.; Raos, N., *Polyhedron* **1990,** *9*, 2419-2427.

(62) Norrby, P. O.; Aakermark, B.; Haeffner, F.; Hansson, S.; Blomberg, M., *J. Am. Chem. Soc.* **1993,** *115*, 4859-4867.

(63) Gajewski, J. J.; Gilbert, K. E.; Kreek, T. W., *J. Comput. Chem.* **1998,** *19*, 1167-1178.

(64) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., *J. Comput. Chem.* **2010,** *31*, 671-690.

(65) Vanommeslaeghe, K.; MacKerell, A. D., Jr., *J. Chem. Inf. Model.* **2012,** *52*, 3144-3154.

(66) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr., *J. Chem. Inf. Model.* **2012,** *52*, 3155-3168.

(67) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M., *J. Am. Chem. Soc.* **1992,** *114*, 10024-10035.

(68) Lemkul, J. A.; Allen, W. J.; Bevan, D. R., *J. Chem. Inf. Model.* **2010,** *50*, 2221-2235.

(69) Comba, P., *Coord. Chem. Rev.* **1999,** *182*, 343-371.

(70) Comba, P.; Remenyi, R., *Coord. Chem. Rev.* **2003,** *238*, 9-20.

(71) Wiesemann, F.; Teipel, S.; Krebs, B.; Howeler, U., *Inorg. Chem.* **1994,** *33*, 1891-1898.

(72) Zimmer, M., *Chem. Rev. (Washington, DC, U. S.)* **1995,** *95*, 2629-2649.

(73) Thacher, T. S.; Hagler, A.; Rabitz, H., *J. Am. Chem. Soc.* **1991,** *113*, 2020-2033.

(74) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., *J. Comput. Chem.* **2004,** *25*, 1157-1174.

(75) Maple, J. R.; Dinur, U.; Hagler, A. T., *Proc. Natl. Acad. Sci. U. S. A.* **1988,** *85*, 5350-5354.

(76) Warshel, A.; Lifson, S., *J. Chem. Phys.* **1970,** *53*, 582.

(77) MacKerell Jr, A. D., *Atomistic models and force fields*. Marcel Dekker: New York: 2001.

(78) Halgren, T. A., *J. Comput. Chem.* **1996,** *17*, 490-519.

(79) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., *J. Phys. Chem.* **1993,** *97*, 10269-10280.

(80) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A., *J. Am. Chem. Soc.* **1993,** *115*, 9620-9631.

(81) Guvench, O.; MacKerell Jr, A. D., In *Molecular Modeling of Proteins*, Springer: 2008; pp 63-88.

(82) Lopes, P. E.; Murashov, V.; Tazi, M.; Demchuk, E.; MacKerell, A. D., *J. Phys. Chem. B* **2006,** *110*, 2782-2792.

TRANSITION METALS IN BIOCHEMISTRY

Transition metal ions like manganese, iron, cobalt, copper, zinc and to a lesser degree nickel, molybdenum, tungsten, and vanadium are known to be vital for many biochemical processes in all living organisms.[1-3] The unique properties of transition metals (d-block metals), including access to several oxidation states and varied coordination geometries have been exploited by nature to perform a wide variety of functions. Metal ions are essential cofactors for functional expression of many proteins and for performing various biochemical functions.[4] In cells, several trace elements are needed to activate and stabilize enzymes, such as superoxide dismutase, metalloproteases, protein kinases, and transcriptional factors containing zinc finger proteins.[4] Metal ions also play key roles as catalytic cofactors in reversible oxidation-reduction reactions, hydrolytic reactions, and structural rearrangements of small organic signaling molecules by coordinating to the active site of the respective enzymes. Most of the proteins involved in the electron transfer chemistry consist of metals at the active site e.g. Iron sulfur cluster proteins.[4] Despite of their importance in cellular homeostasis, they can be toxic to the cell at higher concentrations.[5] As a result, all the cells have evolved mechanisms to ensure a balance between each of these transition metal ions. The homeostasis of these essential metal ions is maintained by extensive network of

regulatory proteins that control the uptake, intracellular availability and efflux/storage.[2,3]

METAL HOMEOSTASIS

Metal homeostasis is defined as the process in which an optimal concentration of metal ions and their bioavailability is maintained in the cell, or in intracellular components of the cell.[2,3] The shortage of a metal ion may initiate several cellular events that may increase the uptake and lower the cellular requirement for this particular metal ion by down-regulating the proteins or biochemical processes that necessitate the metal. On the other hand, when there is excess metal ion concentration, cells are required to efflux, store or detoxify metal ions to mitigate the effect of metal toxicity.[2] In bacteria, all of these processes are highly coordinated largely at the level of transcriptional regulation by a panel of metal sensor proteins/metal regulator proteins.

In Gram-negative bacteria, metal ions have to pass through the outer membrane into the periplasmic space and then through the plasma membrane before they reach the cytoplasm where they are complexed with proteins.[3] Trimeric ß-barrel proteins, porins in the outer membrane allow for non-selective diffusion of metal ions; in other cases, specific high affinity outer membrane receptors mediating the transport of specific metal chelates are present.[3,6] The

cytosol effectively collects all metal ions by employing specific metal uptake transporters located in the plasma membrane.[6] The uptake systems are driven either by the hydrolysis of ATP, for example, ATP-binding cassette (ABC) transporters and P-type ATPases, or via coupling to an energetically favorable process, such as co-transport of protons or other small molecules across the lipid bilayer.[7] In gram-positive bacteria, an extra layer of lipopolysaccharide or complex carbohydrates is present and this may potentially affect the rate and mechanism of metal uptake.[7]

In the cytoplasm and periplasmic space in Gram-negative bacteria, copper ions require specific transporting and trafficking mechanisms to target it to correct proteins[1] and this is usually mediated by proteins called metallochaperones.[1,8] The metal transfer from these metallochaperones to a target protein is accomplished by formation of an intermediate protein-protein complex like the mechanism described Cu trafficking via a ligand exchange reaction.[9,10] Cys-rich metallothioneins involved in Zn and Cu detoxification, or ferritin-like bacterioferritins and Dps-type proteins for iron storage, can function to sequester excess metal ions in the cytosol.[8,11-16] Overall, the intracellular concentration of the metal ions is tightly regulated to maintain normal metabolism of the cell. Pathogenic bacteria which colonize in human body require the same micronutrients (metals and other inorganic ions) that are essential for the host cells

to maintain their cellular metabolism. As a result, a competition exists between the host cells and the colonizing pathogenic bacteria leading to metal stress in the immediate environment of the bacteria in the host cell. Hence, both prokaryotes and eukaryotes developed specific metal acquiring protein systems which import, store and export metal ions by distinguishing each metal based on its charge, size, stable oxidation state, coordination geometry etc.[17-19] Each metal has specific role in the cell and an individual metal ion is capable of performing only one or a few of these functions. One such biologically active metal is copper which is required by many enzymes such as oxidases and antioxidant enzymes like superoxide dismutases.[20] The reversible oxidation reaction of copper from Cu(I) to Cu(II) makes copper an essential cofactor for some critical enzymes involved in O_2 binding and reduction like the multicopper oxidases.[21,22] Uncomplexed Cu(I) in the cytosol can also catalyze the formation of peroxides and reactive oxygen species which can create an oxidative stress in the cells.[23-25] Many neurodegenerative diseases like Alzheimer's and Parkinson's are associated with Cu(II) dependent nucleation of amyloid peptides and catalyzing the reduction of di-oxygen to peroxides after reduction of Cu(II) to Cu(I).[26-37] Hence the copper concentration inside the cells should be regulated in a way the bioavailability of Cu(I) should be minimal and kept very low by complexing with cysteine rich proteins such as metallothioneins and copper chaperones.[1,14-17,19,38-40] Studies on copper homeostasis

mechanism in pathogenic organisms have shown that copper export, in addition acquisition is important for virulence and these organisms are very sensitive to excess concentrations.[38,41-44] Pathogenic bacteria regulate the copper concentration by employing a network of proteins which can sense excess copper by binding to them with high specificity and activating the transcription of gene which translates metal exporter proteins like the ATPases. [37,40-43] Hence, for many bacterial organisms maintaining cellular metal ion homeostasis is an essential requirement for survival.

ROLE OF COPPER IN BIOLOGY

Copper is an essential transition metal ion in various biochemical processes and is required by many proteins and enzymes involved in electron-transfer cycle and in the catalytic oxidation of substrates.[22,45-47] The ability of Cu to undergo reversible oxidation and reduction is very detrimental as free cellular Cu ions can undergo Fenton-like reaction, catalyzing the production of activated oxygen radicals, reactive hydroxyl radicals, which can readily react with proteins, lipids and nucleic acids.[23,25,38,48,49] In support of this, recent studies show that copper stress can activate the genes related to oxidative stress defense, suggesting a connection between Cu homeostasis and oxidative stress,[50,51] although the underlying mechanism remains unclear. A more recent report suggests that in *E. coli* and *B.*

subtilis, Cu(I) may directly damage the solvent exposed iron-sulfur clusters in proteins and significantly interrupt their functions in the absence of molecular oxygen; iron release caused by such damage may induce mis-regulation of Fe homeostasis and, in aerobic conditions, Fe mediated oxidative stress.[25,52] On the other hand, free Cu(II) ion is capable of directly oxidizing amino acid side chains as well as inducing amyloid formation, and thus is associated with many amyloid-related diseases in mammals.[53,54] Therefore, the bioavailable copper levels must be strictly controlled inside the cell and is proposed to be buffered at a concentration in the 10^{-18} M range.[6] As such, it is widely accepted that in bacterial and yeast cells, there is no free or bio-available copper in the cell.[40] Control of Cu homeostasis in bacteria is particularly important since most bacteria seem to lack a cytosolic Cu requirement.[55,56] Thus, virtually any Cu(I) in the cytoplasm in a non-photosynthetic bacterium is toxic. This high reactivity of copper is the reason why bacteria try to avoid Cu toxicity through buffering Cu in the cytoplasm very tightly.[1,55,56] Once the buffering capacity of the Cu is exceeded in cytosol, transcriptional regulatory proteins termed Cu sensors bind to the Cu ions, and through allosteric regulation mechanism they up regulate the expression of genes that encode metal efflux systems such as Cu-exporting P1-type ATPases and Cu metallochaperones.[1,8,14,15,21,38,39,50,55] P1-type ATPases efflux the excess copper while

metallochaperones sequester extra Cu ions simultaneously, attaining homeostasis.[14,15,26,57]

CU(I) SENSING PROTEIN FAMILIES IN BACTERIA

As discussed in the previous section, to combat the elevated levels of Cu, bacterial pathogens implement sophisticated mechanisms for Cu sensing and detoxification.[1,3,15,21,38,58] The Cu regulatory mechanisms are generally encoded into an operon with the following basic framework: a Cu transcription factor (commonly a repressor), a Cu-binding protein (chaperone) that delivers Cu to the repressor, and a P-type ATPase (an exporter).[1-3,8] This operon is de-repressed in the presence of Cu, leading to expression of one or more Cu efflux pumps. In different bacterial pathogens, the above basic framework is modified to respond to their unique niche.[16,56,58] Several transcriptional metal regulatory protein families have thus far been structurally and /or functionally characterized to some detail and the major Cu sensing families among these are reviewed in the next subsections. ArsR/SmtB and MerR proteins are the two most extensively studied families of metalloregulatory proteins, and the proteins involved in sensing Cu specifically will be discussed here.[59-65] The CopY family, the other known family of Cu-sensors in prokaryotes, will also be briefly described.[2,57,66] From this discussion, an understanding of metal specificity and mechanisms involved in metal-dependent

allosteric regulation of DNA binding allows us to place our studies on the CsoR family proteins in context. It should be noted that the molecular and atomic level details by which an individual metalloregulatory protein selectively responds to one or more metal ions remain elusive, due to the limited number of high resolution structures of each functionally relevant conformational state within the regulatory network.

MerR FAMILY

CueR belongs to the MerR-family of metal-responsive transcriptional regulators[67-70] is a Cu(I) sensing protein, other members of which include zinc-sensing ZntR,[71-74] mercury-sensing MerR,[65,75,76] cadmium-sensing CadR,[77] lead-sensing PbrR,[78,79] and cobalt-sensing CoaR.[80] All MerR family metal sensors possess very similar N-terminal winged helical domains comprised of a helix-turn-helix-β-hairpin structure, followed by a long dimerization helix, but quite divergent C-terminal effector binding domains.[65,76] The structural diversity in the C-terminal region makes it possible for individual MerR family proteins to sense not only various metal ions as mentioned above but also oxidative stress by SoxR via an [2Fe-2S] cluster,[81-85] and small molecule drugs in the case of BmrR and MtaN.[86-89] This family of metal sensor proteins are shown to be bound to the promoters of their target genes in the presence and absence of their effector

metals.[65,76] In the presence of elevated levels of metal ions, they allosterically activate transcription by realigning unusually spaced consensus RNA polymerase recognition sequences, while in the absence of metal ions these proteins may lead to the repression.[65,76] The operator-promoter regions upstream of *copA* and *cueO* reveal unusual spacing (19-bp, rather than 17-bp) between −10 and −35 RNA polymerase consensus binding sequences, characteristic of genes regulated by MerR-family transcriptional activators.[65,69,70,76] Expression of both *cueO* and *copA* is substantially induced in response to elevated copper levels by the action of the CueR transcriptional regulator which has zeptamolar sensitivity to free Cu(I), less than one atom per cell.[68,69] Mutants lacking *cueR* show reduced copper-tolerance and lack copper-induced expression of *copA* and *cueO*.[66,90,91] The crystal structure of a CueR homodimer reveals a buried metal-receptor site at the dimer interface that contributes to selectivity toward the monovalent metals by restricting binding to a linear two-coordinate geometry involving two cysteines, Cys^{112} and Cys^{120}.[68]

CopY FAMILY

The *copYZAB* operon in *E. hirae* is regulated by the copper-responsive repressor CopY.[92-95] The current model indicates that at low copper concentrations CopY is present as a Zn(II) containing homodimer and binds to the *copYZAB* operator-promoter region repressing transcription.[96,97] In the presence of copper,

the copper-chaperone donates Cu(I) to CopY, displacing Zn(II), alleviating DNA-binding and allowing transcription of *copYZAB* to proceed.[95,96] The carboxyl-terminal region of CopY possesses a C*XC*XXX*X*C*XC* metal-binding motif. Each Zn(II) ion is coordinated to all four cysteines in the repressing form of CopY, and is replaced by two Cu(I) ions in the induced, non DNA-binding form.[97-100] Spectroscopic studies suggest that each CopY protomer within the dimer is capable of binding two equivalents of Cu(I) per monomer to form a highly luminescent binuclear S_4-Cu_2 cluster exactly analogous to Cu(I) formed by the ArsR/SmtB family regulator BxmR.[2,92,96,98] CopY-like regulatory proteins are not widely distributed amongst bacteria but have been associated with the copper-resistance determinants of *Enterococcus faecium*,[101] *Lactococcus lactis* IL1403[51] *Streptococcus mutans*,[102] and *Streptococcus gordonii*.[103] The amino-terminal half of CopY has structural similarity to the BlaI and MecI repressors that belong to the "winged helix" family of proteins and mediate resistance to β-lactam antibiotics.[99] Further biophysical and structural characterization of the C-terminal Cu(I) binding domain in the intact homodimeric repressor, however, will be required to fully understand how Cu(I) is capable of mediating an allosteric or regulatory response upon DNA binding, while Zn(II) binding is not initiating the allosteric response.

CsoR FAMILY

The first copper sensing CsoR family was discovered in *Mycobacterium tuberculosis* (Mtb-CsoR).[104] Under elevated Cu levels the expression of Mtb-CsoR was found to be strongly induced, with the binding of Cu(I) triggering the derepression of the *rv0967-rv0970* operon in which *rv0969* encodes for a P1-type ATPase involved in Cu transport and *rv0967* for Mtb-CsoR.[104] It has subsequently been found that members of the CsoR family are wide spread in bacterial genomes. Examples of CsoR proteins that transcriptionally regulate *copZA* Cu resistance operons have been identified and characterized to varying degrees in *Bacillus subtilis*,[105] *Listeria monocytogenes*,[39] *Thermus thermophillus*,[106] *Streptomyces lividans*,[107] *Geobacillus thermodenitrificans,*[108] and *Staphylococcus aureus*.[109]

CsoR binds to a region of the *cso* operon containing dyad symmetry and DNA-binding is weakened upon Cu(I)-binding, thus alleviating repression in elevated metal.[104] The structure of Cu(I)-CsoR reveals a homodimer with two Cu(I) ions, each Cu(I) bound in a trigonal coordination complex involving two cysteines and a histidine in a subunit bridging site.[104] Although initial studies described CsoR as a homodimer,[104] it is now known that most CsoR proteins exist in dimer of dimers architecture and bind to Cu(I) with a stoichiometry of one ion per protomer.[106-108]

The majority of identified CsoR-related sequences appear to possess all three Cu(I)-binding ligands, consistent with Cu(I)-sensing.[39,104,105,107-109] However, some of the more distantly related sequences lack all three ligands raising possibility that they might be involved in sensing metals other than Cu(I) or other substrates such as organic molecules.[106] Furthermore, a recent report has identified a CsoR member that is not thought to be involved in Cu homeostasis, but instead appears to function in response to sulfur stress.[110]

COPPER(I) AND DNA BINDING SITES IN CsoR PROTEINS

Unlike other Cu(I) sensors such as *E. coli* CueR and *E. hirae* CopY which are largely confined to the Proteobacteria and Firmicutes, respectively, genes encoding CsoR proteins are widely distributed through most other major bacterial species.[104] The 2.6 Å crystallographic structure of Cu(I)-bound CsoR reveals a homodimeric structure with a core antiparallel four-helix bundle (α1, α1′, α2, α2′) and the short C-terminal α3 helix stacked against the base of the molecule, proximate to α2′ helix of the opposite protomer.[104] The Cu(I) ion is coordinated to an inter-subunit metal binding site formed by two conserved cysteines (Cys36 and Cys65′) and one conserved histidine (His61′).[104,107] Crystal structures have been solved for Cu(I)-bound CsoR from *Mycobacterium tuberculosis,[104] Geobacillus thermodenitrificans[108]* and Apo CsoR from *Streptomyces lividans[107]* and *Thermus*

thermophilus.[106] Mtb–CsoR was suggested to form a homodimer, while Sl–CsoR, Gt-CsoR and Tt–CsoR were shown to possess dimer of dimers architecture.[106-108] The core four-helix bundle structure of the CsoR proteins from the above mentioned three families is very similar except that the Mtb-CsoR has a C-terminal tail spanning ~30 amino acids.[104] This C-terminal tail is present in certain pathogenic Mycobacterium species and was not revealed in the crystal structure of Mtb-CsoR possibly due to its higher flexibility.[104] The function of the tail in Mtb-CsoR remains unclear but the role of the tail was proposed to be responsible for maintaining dimer-tetramer equilibrium in the solution.[104] All three regulators bind two Cu(I) ions per dimer. Each copper ion is coordinated by one residue of the first protomer and two residues of the second protomer in a dimer. C-H-C motifs coordinate Cu(I) in Mtb–CsoR (Cys36-His61′-Cys65′) and Sl–CsoR (Cys75-His100′-Cys104′), while copper ion binding involves a C-H-H motif in Tt–CsoR (Cys41, His70′ and His66′).[104,106,108]

Due to the lack of a classical DNA binding motif such as winged helix-turn-helix domain commonly found in other metalloregulatory proteins, how apo-CsoR binds to the associated DNA operator remains unclear, as well as the mechanism by which Cu(I) binding induces allosteric negative regulation of operator DNA binding. Though these transcriptional regulators do not contain any known DNA-binding motif the characteristic antiparallel four-helix bundles

have been suggested to act as a DNA-binding fold.[57,104] Recent experiments directed towards the prediction of the DNA binding motifs in CsoR proteins,[111] suggest that the prominent patch of positive electrostatic surface potential on the α1-α2 face and close to the Cu(I) binding site (R15 (R24 in Bsu-CsoR) and R52 (next to K60 in Bsu-CsoR) in Mtb-CsoR) forms a part of the DNA binding site.[111] Previous studies on Mtb-CsoR have shown that the substitution of both Arg15 and Arg52 with alanine resulted in the loss of DNA binding[104] and the multiple sequence alignments of CsoR in different bacteria suggest that Arg15 is conserved in CsoR orthologs.[104] The mass spectrometry studies on Bsu-CsoR show that K3 and K96 are the most reactive lysines in the apoprotein, with the reactivity of K96, K97, and K100 most strongly attenuated in the CsoR–DNA complex.[111] The two Lys residues, K18 at the N-terminus of the α1 helix and K96 just C-terminal to the α3 helix, are shown to be protected from amidation in the Cu(I)-loaded tetramer versus apo-CsoR.[111] Attenuation of reactivity of K18 relative to apo state (R10 in Mtb-CsoR and R13 in Tt-CsoR structures) may suggest a reorganization of the α1 and α2 helices on Cu(I) binding.[111] Modest protection from amidation of K96 was also observed and was suggested to be a result of change in the structure or remodeling of the tetramer interface, which is required to drive allosteric inhibition of operator DNA binding by Cu(I).[111] A most recent work on the DNA operator binding of *S.lividans* CsoR predicts that the α1-helix RLXR motif

establishes contact at the GTA dyad regions of the type 2 operator site, and this would lead to the positioning of an Arg-rich cluster (RLXR) towards these major or minor grooves of the GTA dyads through hydrogen bonding interactions between Arg and purines (G/A).[112] This work suggests that one Sl-CsoR tetramer binds with the RLXR motif pointed towards the major face and the second CsoR tetramer binds with the RLXR motif towards the minor face where the functional groups of the purines of the GTA dyads are likely to be exposed.[112] The above proposed interactions indicate that the apo CsoR tetramer binds DNA through one face, creating asymmetry on the opposite face of tetramer. On binding to Cu(I) this type of flexes leading to asymmetry are restricted and a 'flat or taut' conformation was shown to exist on binding all four Cu(I) sites.[112]

MOTIVATION TO UNDERSTAND ALLOSTERIC COPPER SENSOR PROTEINS IN PATHOGENIC BACTERIA

Mtb is a pathogen that is well-adapted to respond to anti-microbial Cu in the host, as shown by its success as a human pathogen. The Mtb genome encodes two Cu-responsive regulons, the established Cu responsive operon regulated by CsoR[104] as well as a second Cu-responsive regulon controlled by RicR.[113] The *cso* operon is expressed in mice, implicating an important role during infection, and mutation of the *cso* operon member encoding the Cu transporting ATPase, CtpV,

leads to decreased lung damage in guinea pigs and in mice with compromised immunity.[114] Moreover, loss of an outer membrane channel protein, MctB, which is not regulated by CsoR or RicR, leads to increased cellular copper load and decreased virulence in guinea pig infections.[42] In comparison, MymT, a metallothionein, is protective against Cu stress in vitro, but mutation of *mymT* does not result in a detectable decrease in virulence in the mouse infection model.[11] It is unclear why the *mymT* mutant does not have any discernible phenotype in the mouse model, but perhaps using the guinea pig model, where phenotypes were observed for *ctpV* and *mctB* mutants, may elucidate a role.[55,56,114] Mounting evidence suggests that the anti-microbial properties of copper are used by host immune cells as one tool to defend against microbial pathogens.[55,56] In turn, pathogenic bacteria like Mtb implement tightly controlled copper homeostatic mechanisms to utilize copper yet resist copper toxicity.[18,42] Understanding how the pathogens sense and cope with Cu deficiency or toxicity at distinct times after infection may possibly lead to new drug targets against Mtb for the development of novel therapeutic agents to cure tuberculosis. Attacking bacterial Cu resistance mechanisms and copper responsive regulator proteins like CsoR in conjunction with conventional antibiotics, may prove to be a valuable avenue for the development of anti-bacterial therapies.

COMPUTATIONAL MODEL OF Mtb-CsoR AND STUDIES ON ALLOSTERY

IN METAL SENSOR PROTEINS

As mentioned in the previous section Mtb-CsoR was initially characterized

as a dimer as shown in Figure 3.1,[104] but recent biophysical and structural studies

on CsoR in other bacteria demonstrate the presence of dimer of dimers

architecture.[104,106-108] Therefore, in this computational evaluation of allostery in

CsoR both the dimeric and tetrameric states were considered. In the presence of

copper the CsoR repressor dissociates from the *cso* operon which can then be

expressed. The operon contains the gene for what is believed to be a cellular

copper exporter – CtpV.[114] Binding of Cu(I) to the Mtb-CsoR dimer is believed to

cause a conformational change that has low affinity for DNA, allowing expression

of the *cso* operon, thus providing a protective mechanism against toxic levels of

copper within cells.[104] The binding of copper to Mtb-CsoR is exceptionally strong

($K_{Cu} \geq 10^{-19}$M), which is expected for a system that is critical for removal of toxic

copper from cells. Liu et al., 2007 used a series of techniques to understand the

structure of the Cu(I)-bound CsoR repressor.[104] Protein crystallography of Cu(I)-

bound CsoR indicates that CsoR is an alpha-helical dimer, with each protomer

composed of three helices (Figure 3.1).[104] The copper is bound between the two

protomers, coordinated by the side chains of amino acids from each subunit

(Cys36, Cys65ʹ, and His61ʹ). Copper K-edge X-ray absorption spectroscopy (XAS)

was used to provide accurate bond-length information of the copper site in Cu(I)-bound CsoR.[104] Analysis of the extended X- ray absorption fine structure (EXAFS) oscillations indicated a three-coordinate site with two Cu—S ligands at 2.21 Å and one oxygen or nitrogen at 2.06 Å.[104] Analysis of the Cu XAS of the Cu(I)-bound Mtb-CsoR H61A mutant indicated a two-coordinate site with two sulfurs at 2.14 Å.[104] Cys-Cys-His ligand trigonal binding mode of Cu(I) is unusual and there are only few other examples of such trigonal coordination of Cu(I) that include ScAtx1 metallochaperone[115] and a yeast Sco1 protein.[116] The functional relevance of the atypical binding motif of Cu(I) in allostery is investigated using computational methods in this research. Unnatural amino acid substitution experiments performed on Mtb-CsoR show that atom substitutions on the Nε2 of Cu(I) coordinating His-61 can allosterically uncouple Cu(I) and DNA binding, with no effect on Cu(I) binding affinity and coordination structure.[117] Tyr-35 and Glu-81 amino acids were observed to be in close proximity to the Cu(I) binding region in the crystal structure and a collective interaction between the Nε of His-61, phenol group of Tyr 35 and the carboxylic acid group of Glu-81 was hypothesized to be crucial for the allosteric coupling in Mtb-CsoR.[117] In this chapter, we describe the computational studies that are performed on Mtb-CsoR to answer the following questions: What role does copper binding have in the conformational preference of CsoR? Does copper-binding preferentially stabilize one conformation over

another? What residues are important in promoting these conformational changes? In order to predict the Cu(I) coordination geometry and the residue contacts which lead to the allosteric regulation, a thorough understanding of the electronic structure of Cu(I) and it coordinating ligands is crucial. Hence, we chose to use high level quantum mechanics calculations to predict the geometry around Cu(I) along with protonation state of coordinating cysteines. In this evaluation, we observed that the coordination geometry of Cu(I) is extremely effected by the protonation state of the cysteines, there by influencing the secondary coordination sphere residue contacts that might lead to the conformational change on Cu(I) binding.

Quantum mechanics methods allow us to model the distribution of electrons around atomic nuclei and can explicitly account for the making and breaking of covalent bonds during chemical reactions, while being limited to very short (fs to ps) timescales. But our goal is to interpret the protein dynamics that may cause the conformational change of the protein on binding to Cu(I) and these motions happen in the medium range (ns-us) timescales (Table 2.1). Therefore hybrid QM/MM methods in conjugation with all atom MD simulations were employed to predict the contacts between residues in the first and second coordination sphere and the small motions in the protein regions on binding to Cu(I) that might contribute to the allosteric transition.
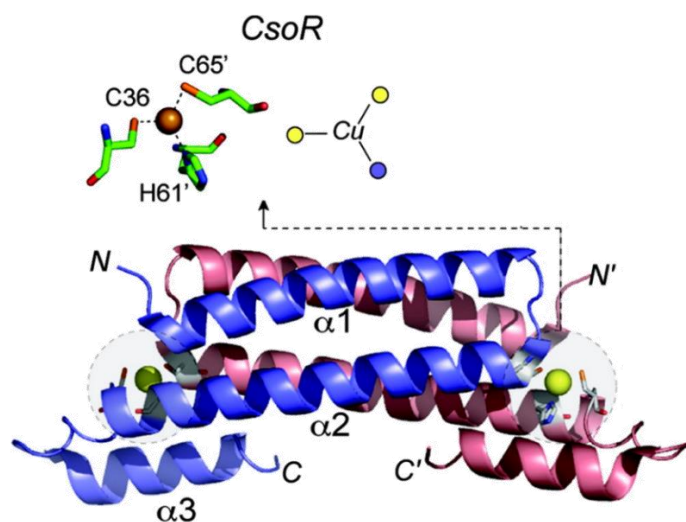
Figure **Error! No text of specified style in document.**.1. X-Ray crystal structure of Mtb-CsoR, showing the copper coordinating amino acids[104]

To elucidate the role of copper binding on conformational states and dynamics of CsoR, additional sampling was carried on various forms of CsoR from *Mycobacterium tuberculosis*, *Thermus thermophillus* and *Streptomyces lividans* using all-atom molecular dynamics (MD) simulations. MD simulations of biological molecules have often been utilized to elucidate biological mechanisms.[118-121] Computational studies and MD simulations of some metal transcriptional regulators have also been reported in literature.[122-129] Bradley, Chivers, & Baker, utilized equilibrium molecular dynamics simulation to explore the conformational dynamics of the *E. coli* NikR tetramer.[122] Correlation analyses and PCA were used to identify residue contacts between Ni(II) and DNA binding domains of NikR. One drawback of this study is the absence of Ni(II) coordination data and simulations were only performed on the apo-NikR.[122] But the residue interactions

that might be important in the allosteric transformation were calculated using contact and position correlation methods, to find clusters of residues that share similar correlation patterns with Ni(II)- and DNA-binding site.[122,130] Sindhikara, Roitberg, & Merz, in a different study evaluated the Ni(II) and apo form of *Pyrococcus horkoshii* species NikR using both quantum mechanics based methods and MD simulations.[127] QM based force-field parameters were developed to accurately represent the four nickel-centers in the nickel-bound form.[127] These simulations provide additional insight into the dynamical preference that different initial conformations and nickel-binding states have on the protein overcoming the limitations of the computational model developed by Bradley et al., in 2008.[122] While the global conformational changes on the computational time-scale are not observed, several key flexible regions were highlighted especially the α2-β2 loop whose hinge-like motion allows functional global conformational change.[127] Sindhikara et al., 2009 suggest that regions with either high correlation relevance or flexibility seem to be correlated with biologically important and evolutionarily conserved regions.[127] Specific mutations were suggested to these residues for the *P. horikoshii* species and also identified analogous regions in *E. coli* and *H. pylori* species of NikR that were expected to have similar character.[122,127] Guo et al., 2010 evaluated the structure and conformational dynamics of the MerR metal sensor on binding to Hg(II).[129] MD simulations were used in the above work,[129] to predict the

interdomain motions on a timescale of approximately 10 ns involving large-amplitude (approximately 20 Å) domain opening-and-closing, coupled to approximately 40° variations of interdomain torsional angle. This correlated domain motion was suggested to be important in propagating the allosteric changes from the metal-binding site to the DNA-binding site while maintaining DNA contacts required to initiate DNA unwinding.[129]

Extensive simulations were also performed on the *Staphylococcus aureus* CzrA that belongs to ArsR/SmtB family of metal sensing transcriptional repressors by Merz and coworkers.[123,126,128] Unrestrained MD and QM/MM MD simulations performed on the apo and zinc-bound form of CzrA reveal the conformational change from "closed" to "open" form on binding to zinc.[123] Chakravorty et al., 2013 successfully determined a structural ensemble for the zinc-bound form of CzrA by effectively using NOE and RDC data in combination with modern MD and QM/MM MD techniques to model metal ion coordination.[128] Crucial second-coordination shell hydrogen bond between the conserved metal-binding ligand His97 and His 67′ of the αR helix,[131,132] that connects the metal binding region to the residues in the DNA-binding region was predicted experimentally in Zn-bound form of CzrA.[132] Computational studies performed by Chakravorty et al., 2012 suggest that this hydrogen bond is strengthened on the order of ~10 kcal/mol on metal ion binding compared to an apo allosteric form of the protein.[126] Similar

study was performed on Mtb apo-NmtR using constrained MD and energy minimization methods to design the Zn(II) binding site in this protein.[124] Several calculations performed at the DFT level of theory to validate the proposed Ni(II) coordination complex in NmtR were also published.[124] Steered MD simulations were also performed on the NmtR system to predict the Ni(II) coordinating ligands and predict the structural change on binding to the metal.[124]

In this computational exploration we show that Cu(I) coordination in Mtb-CsoR can be studied by combining experimentally derived structural information (XAS, X-ray crystallography)[104] with QM/MM methods to derive the first and second coordination sphere geometries and residue interactions. The QM/MM derived bonds, angles, dihedrals and charges were used to approximate and derive the force constants to create coordinate-covalent bonds between metal and protein in MD simulations. The accurate treatment of metal ions will correctly account for first and second coordination-shell effects such as metal ion mediated hydrogen-bonding interactions and will help capture the polarizing influence of metal ion binding on the protein conformation, ultimately leading to a more precise ensemble of structures. Correlation analysis and residue contacts were also analyzed to predict the changes in the local motions of Mtb-CsoR in Cu(I) bound and apo forms.

METHODS AND MATERIALS

QUANTUM MECHANICS

All calculations were carried out using Density Functional Theory(DFT) as implemented in the Jaguar 7.0 suite[133] of *ab-initio* quantum chemistry programs. Geometries were optimized using the B3LYP[134-136] functional with the 6-31G** basis set. Copper was represented by the Los Alamos LACVP basis. The energies were reevaluated by additional single point calculations at each optimized geometry using Dunnings's correlation consistent triple-ζ basis set[137] cc-pVTZ(-f) with the standard double set of polarization functions. In these single-point calculations, Cu was described by a modified version of LACVP, designated as LACV3P, where the exponents were decontracted to match the effective core potential with the triple-ζ quality basis. Vibrational frequency calculation results based on analytical second derivatives at the B3LYP/6-31G**/LACVP level of theory were used to confirm proper convergence to local minima and to derive the zero-point-energy (ZPE) and vibrational entropy corrections at room temperature using un-scaled frequencies. Solvation energies were evaluated by a self-consisted reaction field[138-140] (SCRF) approach with a solvent-excluding surface cavity, based on accurate numerical solutions of the Poisson–Boltzmann equation. In the results reported below, solvation calculations were carried out at the gas phase geometry using the

6–31G**/LACVP basis and employing a dielectric constant of $\varepsilon = 80.37$ for water. Whereas the accurate computation of absolute solvation energies remains a challenge and potentially requires careful inspection of the empirical parameters, the differential solvation energy is expected to be less sensitive owing to significant error cancellation when the same empirical parameters are used. Thus, the differential solvation corrections are most likely more reliable than the absolute energies of solvation. The energy components were computed with the following protocol. The free energy in solution phase G (sol) has been calculated as follows:

$$G(\text{sol}) = G(\text{gas}) + \Delta G(\text{solv}) \qquad\qquad 3.1$$

$$G(\text{gas}) = H(\text{gas}) - TS(\text{gas}) \qquad\qquad 3.2$$

$$H(\text{gas}) = E(\text{SCF}) + ZPE \qquad\qquad 3.3$$

$$\Delta E(\text{SCF}) = \sum E(\text{SCF})_{\text{products}} - \sum E(\text{SCF})_{\text{reactants}} \qquad\qquad 3.4$$

$$\Delta E(\text{SCF}) = \sum E(\text{SCF})_{\text{products}} - \sum E(\text{SCF})_{\text{reactants}} \qquad\qquad 3.5$$

G (gas) is the free energy in gas phase; ΔG (solv) is the free energy of solvation as computed using the continuum solvation model; H (gas) is the enthalpy in gas phase; T is the temperature (298.15K); S (gas) is the entropy in gas phase; E (SCF) is the self-consistent field energy, i.e. "raw" electronic energy as computed from

the SCF procedure and ZPE is the zero point energy. Note that by entropy here we refer specifically to the vibrational/rotational/translational entropy of the solute(s); the entropy of the solvent is incorporated implicitly in the continuum solvation model. Equation 3.**6** was used to calculate the pK$_a$ of deprotonation of cysteines in the copper bound form of CsoR.

$$\Delta G(sol) = -RTlnK_a = 2.303*R*T*pK_a = 1.36*pK_a \text{ (at 298.15 K)} \qquad 3.6$$

QUANTUM MECHANICS/MOLECULAR MECHANICS

We used QSite[141] 4.5 for combined QM and MM calculations. The QM and MM boundary is treated based on a frozen orbital approximation[142]. Quantum calculations in QSite are carried out using Jaguar 7[133] and the molecular mechanics calculations are performed by IMPACT[143] which is developed by Levy and co-workers. The OPLS-AA molecular mechanics force field is used for the treatment of the MM part while the pure QM part was treated at the DFT-B3LYP[134-136] level of theory. The basis set used for the entire QM region is LACVP*, which uses 6-31G* for non-transition metals. This is the default basis set used in the parameterization of the frozen orbital cuts. All geometry optimizations are performed using LACVP* basis set. The methodology has been extensively tested and shown to give reliable results for the relative conformational energies of dipeptides and protonation energies of protein side-chains. The errors resulting

from the QM/MM interface have been tested to be small (~0.5 kcal/mol) and are within the error ranges of DFT methodology.[142] This methodology uses an adiabatic approach, in which the MM region is fully optimized after each QM step, resulting in very large reductions in computational effort as compared to carrying out a QM gradient evaluation at each geometry step.[142,144,145]

The fundamental idea of QM/MM methodology is to divide a large system into two regions, QM and MM. The reactive chemical event is limited within the QM region, while the surrounding region is modeled by MM. If a suitable coupling of the QM and MM energy functions can be achieved, this methodology enables us to study reactions with the accuracy of the QM model, retaining the structural and electrostatic effects of the MM environment with the same computational cost of a quantum chemical treatment of the QM region alone. Four QM/MM models were built, using the Cu-CsoR crystal structure[104] as reference with the QM and MM regions shown in Figure 3.2. The Mtb-CsoR computational model starts with Glu-4 at the N-terminus and residue Phe-88 at the C-terminal due to the lack of coordinates for residues 1-3 and the C-terminal tail in the crystal structure.[104] The number of atoms that can be included in the QM region is limited to 200.[146] An iterative approach was adopted to evaluate the two copper centers in CsoR. The first set of calculations included a single copper coordination sphere in the QM region while the other copper center was treated as the MM region. Positional

restraints were used to restrict the movement of Cu(I) in the MM region. In the second set of calculations the other copper center was optimized using the QM, wherein the previously optimized copper coordination complex was included in the MM region, and the geometry is fixed at the QM optimized geometry. Simultaneous optimizations were performed until a reasonable agreement has been obtained between the two copper centers. In all the models, carboxylic acid amino acids are deprotonated (Glu, Asp) and the basic amino acids are protonated (Lys). The N-Terminal is modeled as $-NH2$ and the C-Terminal as $-COOH$.
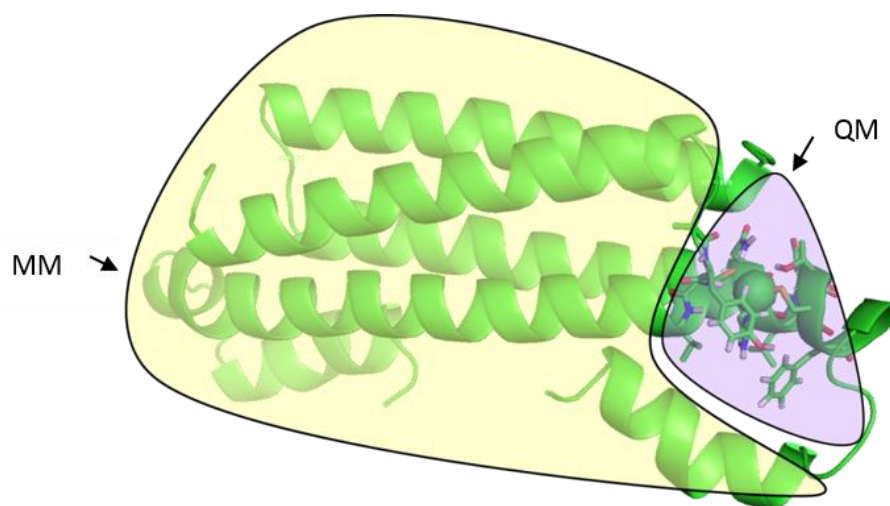
Figure 3.2. QM and MM regions as used in the QM/MM calculations of Mtb-CsoR protein. The small region marked in violet is the QM which includes copper coordination environment, and the large yellow region is the MM region

MOLECULAR DYNAMICS SIMULATIONS

MD simulations were carried out using the gromacs[147] package with constant number, pressure, and temperature (NPT) and periodic boundary

conditions. Standard Gromos force-field[148] for the apo form was used, and force fields for each copper complex that we developed in our laboratory. The linear constraint solver[149] (LINCS) method was used to constrain bond lengths, allowing an integration step of 2 fs. Electrostatic interactions were calculated with the Particle-Mesh Ewald algorithm. All simulations were preceded and followed by minimizations and sufficient equilibrations. Minimizations were run using the steepest descent algorithm with an energy minimization tolerance of 1 kJ mol^{-1} nm^{-1} and maximum force of no greater than 1000 kJ mol^{-1} nm^{-1}, with Lennard-Jones and Coulomb cutoffs set to 1.0 nm. Equilibration was performed on the solvent and ions around the protein to stabilize the system from collapsing during the unrestrained dynamics simulation. The reason is that the solvent is mostly optimized within itself, and not necessarily with the solute. Indeed, the solvent should be brought to the desirable temperature and proper orientation around the solute (the protein). After a desired temperature is reached, pressure is applied to the system until it reaches a certain density. All equilibrations are performed with a position restraining force on the heavy atoms of the protein (anything that is not hydrogen). The motion of the atoms is not restricted to avoid large energy changes in the system. The utility of position restraints is that they allow us to equilibrate the solvent (water) around the protein, without the added variable of structural changes in the protein. Equilibration is conducted in two phases. The first phase is

conducted under an NVT ensemble (constant Number of particles, Volume, and Temperature) over a time frame of 100ps. This ensemble is also referred to as "isothermal-isochoric" or "canonical". V-rescale temperature coupling method is used. NVT equilibration stabilizes the temperature of the system. Equilibration of pressure is conducted under an NPT ensemble, wherein the number of particles, pressure, and temperature are all constant over a time frame of 100ps. The ensemble is also called the "isothermal-isobaric" ensemble, and most closely resembles experimental conditions. For pressure coupling the Parrinello-Rahman[150] barostat is used. Starting velocities were generated with a Maxwell-Boltzmann distribution corresponding to 300 K. All structures reported in this work were found to be stable within these MD runs. The energy minimized structure from the QM/MM calculations was used as the starting structure, by freezing the positions of Cu(I) and heavy atoms coordinating amino acids. CHELPG charges from the Jaguar output file are used for Cu(I) and its coordinating amino acids to maintain a charge of -2 units for the Cu(I)-bound Mtb-CsoR dimer, each Cu(I) coordination contributing a charge of -1((2 Cys-S$^-$) -> -2 (Cu(I)) -> +1 = -1 per Cu(I) site) . The simulation trajectories obtained over 10 ns for copper bound form and 20 ns for apo form were subjected to cluster analyses with the Jarvis-Patrick algorithm[151-154].

RESULTS AND DISCUSSION

SMALL MODEL QM CALCULATIONS

DFT has emerged as the QM method of choice for realistic simulations of systems as large as 150 atoms using a relatively high level of theory. Even so, this limit is still much smaller than the total size of the Cu(I)–CsoR complex. Therefore, smaller models were constructed using the immediate coordination environment of copper for the high-level calculations. A representative minimalist model is shown in Figure 3.3a. Only the copper atom and the amino acid side chains directly bound to copper were included in these calculations. To reduce the computational cost and to avoid irrelevant conformations of N- and C-termini, amino acids were terminated at the Cα position, and the amine and carboxyl groups are replaced by protons. DFT calculations show that the copper coordination geometry is linear, with the two cysteines as coordinating ligands (Figure 3.3b). The histidine ring was observed to be in non-coordinating distance and shifts away from the Cu(I) center. Not surprisingly, the QM calculations predict a linear geometry for Cu(I) which is not in agreement with the EXAFS experiments that reveal a trigonal coordination[104] as shown in Figure 3.3c. Therefore, carrying out exclusive QM calculations will limit our understanding of protein enforced constraints on the Cu(I) binding and prediction of the incurred

electronic structures. Hence, QM/MM methods (discussed in the next section) were utilized to calculate the electronic structure of copper binding geometry with higher level of theory and the protein effect with classical theory.
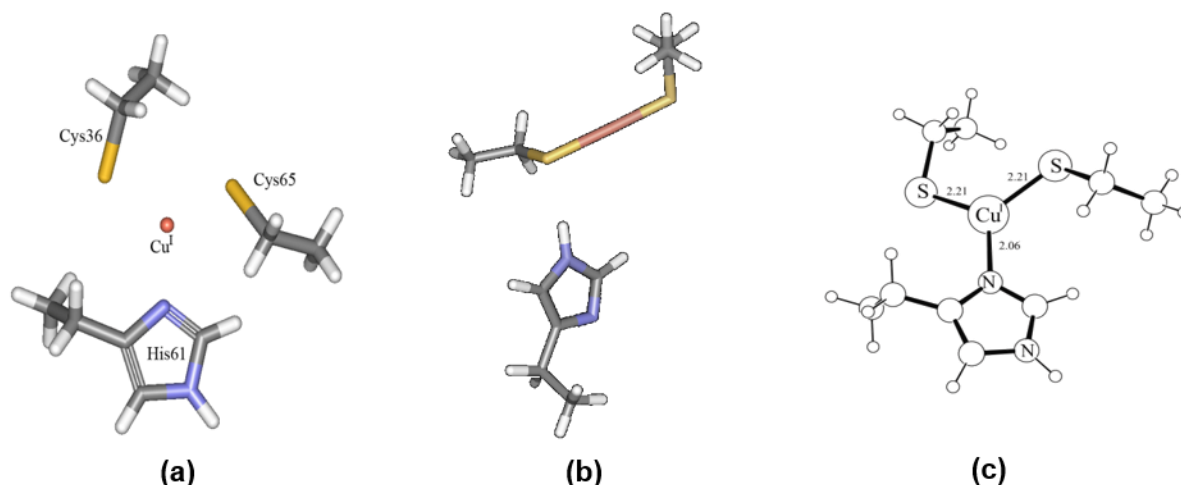
Figure 3.3. (a) Small model representative of Cu coordination used in QM calculations (b) Optimized QM geometry of Cu complex (c) EXAFS data depicted on the Cu complex

**COORDINATION GEOMETRY OF COPPER(I) IN Mtb-CsoR**

To investigate the copper coordination geometry and study the effect of protonation of cysteines on binding to Cu(I), four QM/MM models were built, starting from the crystal structure of copper bound Mtb-CsoR.[104] In all the models, the protonation state of the polar amino acids was adjusted to mimic the biological pH of 7. Different protonation states of the amino acids are assigned using the ProtAssign program at pH of 7. The computational model does not include the 30 amino acid tail in the C-terminal and 3 amino acids in the N-terminal.[104] As a result

of this, the C- and N-termini were protonated (-COOH and –NH2) in all the Mtb-CsoR models. All the backbone cuts were made 5 Å from the metal active site. The boundaries of QM region in this model include Tyr32 and Val33 of chain A, and Asn67 and Phe63 of Chain B. The QM region consists of following amino acids with the copper binding amino acids depicted in red color- TYR-CYS-VAL, ASN-HIS-LEU-GLU-THR-CYS-PHE.

Model A (Cys65-S⁻, Cys36-S⁻, His61- N$_\delta$, Cu(I))

In Model **A** both the copper coordinating Cys65 and Cys36 are deprotonated. The total charge on the QM region is -2, with the copper coordination geometry contributing to -1 charge and glutamate carboxylic acid group contributing a charge of -1. The optimized structure of the QM regions is shown in Figure 3.4. The geometry around the copper is calculated to be tri-coordinate, with Cys65—Cu(I)—Cys35 angle ~130° slightly higher than the angles (120°) in trigonal planar complexes. Thus the geometry around the Cu(I) in this model was inferred to be distorted trigonal. The His61- Nε proton involves in a hydrogen bonding interaction with the OH atom of Tyr35. This hydrogen bonding has been proposed in the previous experimental work on the Mtb-CsoR system and has been shown to be crucial in transmitting the signal involved in the

allosteric regulation on binding to copper.[104] The hydrogen bonding distance is calculated to be ~2.2 Å.
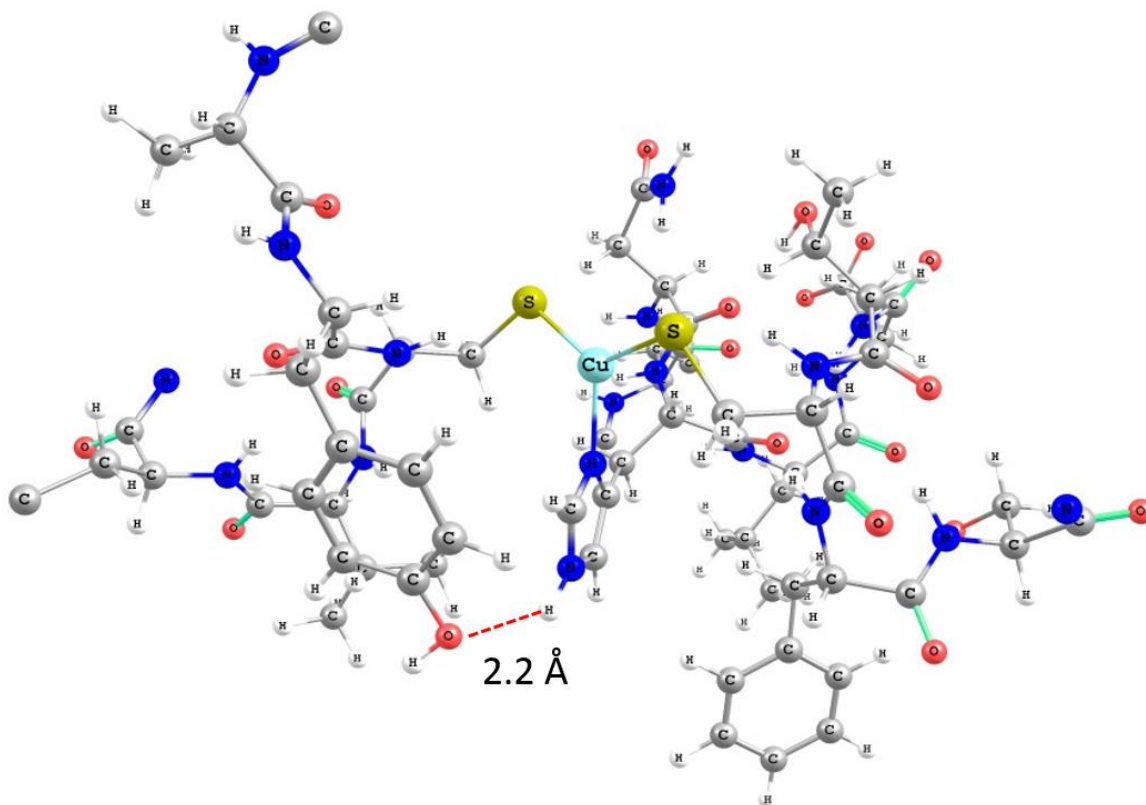
Figure 3.4. QM/MM optimized structure of Model A, Cu (Cyan) coordinating with Cys36, Cys61 and His61. H-bond interaction (red dotted line) between His61-Nε proton and Ty35-OH is shown

Model B (Cys65-S-, Cys36-SH, His61- Nδ, Cu(I))

In Model **B**, Cys65 is deprotonated. On protonating Cys36 (Cys65-S- -> -1, Cys36-SH -> 0, Cu(I) -> +1 = 0 per Cu(I) site) the net charge on Cu(I) site was calculated to be "zero". The overall charge of the QM region is -1, with the glutamate carboxylic acid that is included in the QM region contributing to the -1

charge. The optimized structure of the QM regions is shown in Figure 3.5. The coordination geometry around Cu(I) differs from the geometry predicted in Model **A**, and also from the XAS experiments. The deprotonated Cys36 and His61 ligands coordinate in a linear fashion, while the protonated Cys65 was observed to be at a non-coordinating distance of 4.25 Å from Cu(I). The geometry around Cys36, Cu(I) and His61 is close to linear with an angle of 179°. The hydrogen bonding interaction between His61- Nε proton and Tyr35-OH is not identified in Model **B**.

4.9 Å
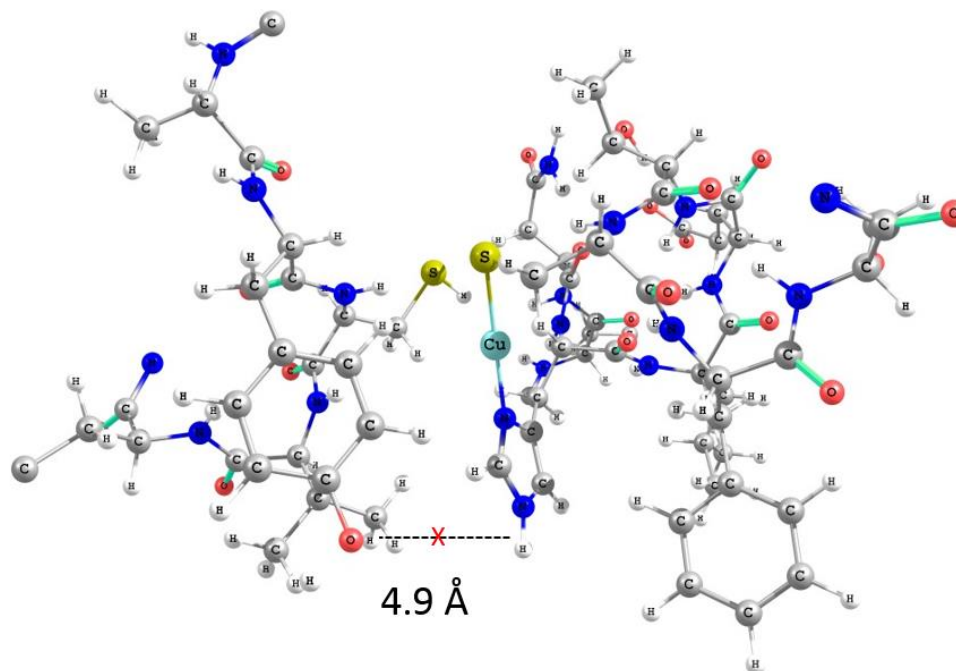
Figure 3.5. QM/MM optimized structure of Model-**B,** Cu (Cyan) coordinating with Cys36-SH, Cys65 and His61. Hydrogen bonding interaction is not observed between His61-Nε proton and Ty35-OH

**Model C (Cys65-SH, Cys36-S⁻, His61- Nδ, Cu(I))**

Cys36 is protonated in Model **C**. The total charge of the QM region is -1(calculation showed in the Model-B), with the deprotonated glutamate carboxylic acid contributing to the -1 charge. The optimized structure of the QM region is shown in Figure 3.6. The geometry around Cu(I) is close to trigonal planar, but this model does not possess the crucial hydrogen bonding involved in the coupling of copper binding to allosteric regulation.
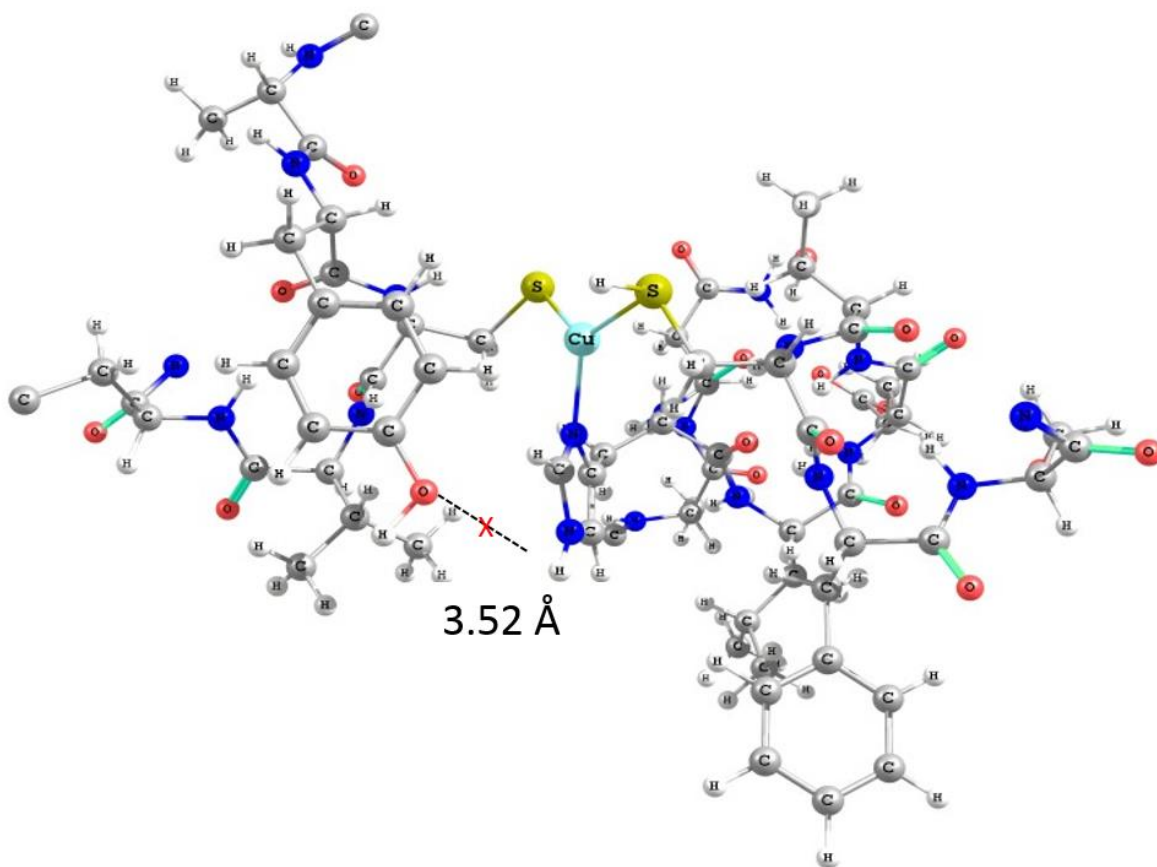
Figure 3.6. QM/MM optimized structure of Model **C,** Cu (Cyan) coordinating with Cys36, Cys65-SH and His61. Hydrogen bonding interaction is not observed between His61-Nε proton and Ty35-OH

Model D (Cys65-SH, Cys36-SH, His61- Nδ, Cu(I))

In this model both Cys65 and Cys36 are protonated. The total charge of the
QM region is 0, +1 charge from the copper coordination(2 Cys-SH -> 0, Cu(I) -> +1
= +1 per Cu(I) site) and -1 charge from the glutamate carboxylic acid group. The
optimized structure of the QM region is shown in Figure 3.7. The calculated
geometry around Cu(I) is close to trigonal planar and the hydrogen bond between
His61- Nε proton and Tyr35-OH is not present in this model.

3.33 Å
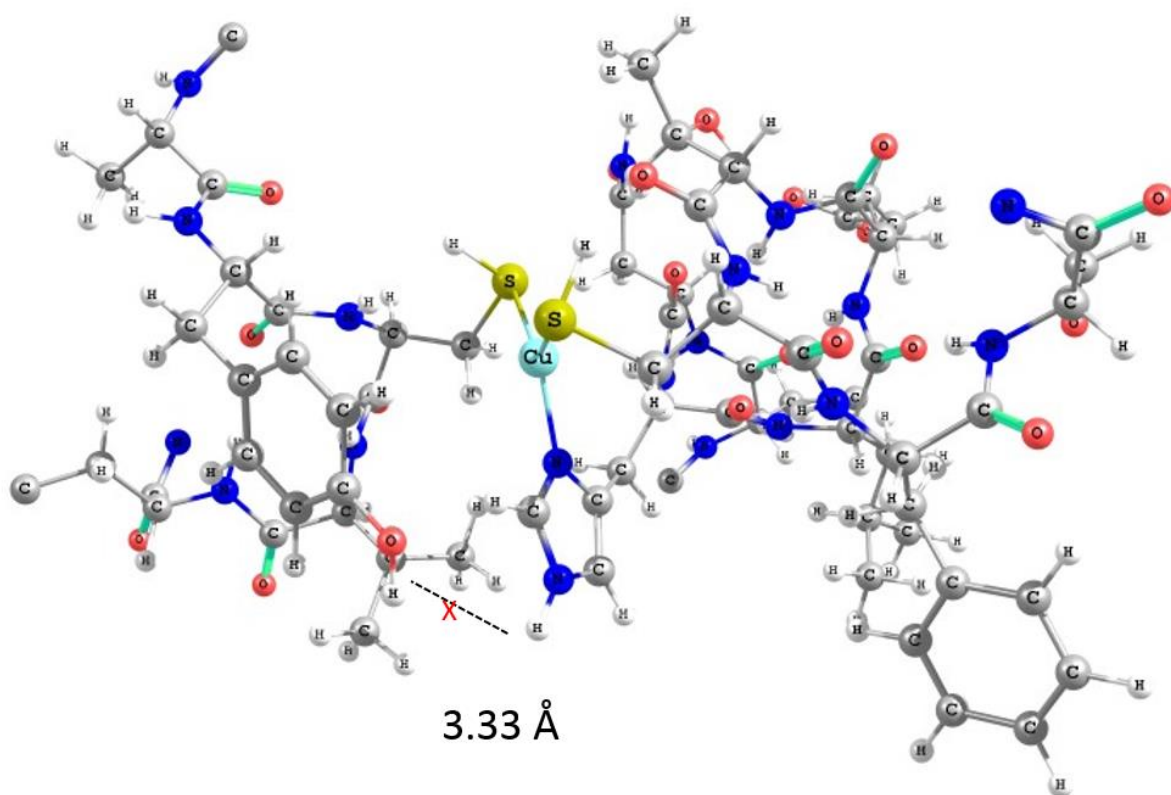
Figure 3.7. QM/MM optimized structure of Model-**D,** Cu (Cyan) coordinating with
Cys36-SH, Cys65-SH and His61. Hydrogen bonding interaction is not observed
between His61-Nε proton and Ty35-OH

As discussed above, the protonation state of Cys36 and Cys65, has a drastic influence on the coordination geometry of Cu(I) as well as the hydrogen bonding network and residue-residue interactions of the second coordination shell amino acids. QM/MM calculations validate the presence of a "bonded model" and reveal that the Cu(I) remains bonded to its amino acid ligands that are shown to involve in trigonal coordination.[104] The hydrogen bond between the His61, Tyr35 and Glu81 (Figure 3.4) is only observed in Model **A**, where both cysteines were deprotonated. This interaction was proposed to have an important role in allosteric regulation of Mtb-CsoR.[104,117] In order to predict the underlying electronic and structural aspects that result in the loss of H-bonding interaction between metal coordinating His61 and second coordination shell Tyr35 and Glu81 on protonating cystienes, the bond lengths, angles and the energetics of the ligand coordination to Cu(I) were evaluated.

The coordination geometry of Cu(I), QM/MM derived bond lengths and angles were (Table 3.1) compared with those obtained from the EXAFS experiments on Cu(I)-bound form of Mtb-CsoR.[104] When one of the cysteine is protonated, His-ND becomes a better ligand to Cu(I) which is indicated by the decrease in the bond distance between Cu(I) and His61 in Model **B** and **C** (Table 3.1). Even shortening of Cu(I)–His bond length is observed on protonating both the coordinating cysteines. The bonding model represented by Model **A** is

consistent with the data derived from EXAFS data with variance of 0.009 Å. In Model **A** the geometry around Cu(I) is observed to be distorted trigonal coordination which is in close agreement with the predicted trigonal planar coordination geometry around Cu(I).[104]

Table Error! No text of specified style in document.**.1.** Calculated Bond lengths of QM/MM models

Model	Description	Cu−N (His61) Å	Cu−S (Cys36) Å	Cu−S (Cys65) Å	Calculated Geometry
A	Both Cys deprotonated	2.12	2.35	2.28	Distorted trigonal
B	Cys65 deprotonated	1.94	4.25	2.20	Linear
C	Cys36 Deprotonated	2.08	2.29	2.43	Distorted trigonal
D	Both Cys protonated	2.04	2.48	2.31	Distorted trigonal
Crystal	EXAFS	2.06	2.21	2.21	Trigonal planar

To evaluate the nature of interaction particularly between His61 and Cu(I), Zeigler Rauk energy decomposition analysis was performed. Several small models were generated from the structural data derived from the QM/MM structures. Two fragments were constructed by performing a cut at the Cu(I)–N$_\delta$(His61) bond in different protonation state combinations of the two cysteines. The various

contributions to the electronic bonding energy of Cu(I)–N_δ(His61) bond are tabulated in Table 3.2. In Model **A,** the electronic interaction between His61 and Cys65–Cu(I)–Cys36 fragment is calculated to be -0.35 kcal/mol suggesting the presence of a weak bond. In the experimental evaluation of binding constant of Mtb-CsoR the log K_{Cu} of wild type was reported to be 18.0, while in the H61A mutant log K_{Cu} is predicted to be 14.9 suggesting that His61 does not contribute majorly to the binding energy of Cu(I) which is in agreement with our computed binding energy.[104,117] However, in other binding models when at least one of the cysteine is protonated the electronic interaction between the histidine and Cu(I) increases and the maximum interaction energy of ~36kcal/mol was observed for Model **D**. According to lewis acid and base theory, cysteine in thiolate form is stronger nucleophile compared to a thiol, hence in the deprotonated form the bond between cysteine and Cu(I) is stronger and the partial negative charge on thiolate stabilizes the positive charge on Cu(I) leading to stronger electrostatic interaction. This renders the histidine imidazole interaction weaker when the two coordinating cysteines are deprotonated. But, when either of the cysteine is protonated, the histidine imidazole is a stronger nucleophile compared to a thiol promoting a stronger bonding interaction between histidine and Cu(I) which is supported by the increase in the interaction energy in Models **B**, **C** and **D** in Table 3.2. In addition to the electronic and structural analysis, pKa analysis was also

performed to further evaluate the energetics of deprotonation of cysteines at physiological pH.

Table 3.2. Ziegler–Rauk energy decomposition of His61-Cu bond in different models, all energies are in kcal mol^{-1}

Ziegler–Rauk energy decomposition	A S$^-$, S$^-$	B SH,S$^-$	C S$^-$,SH	D SH,SH
ΔE_{pauli}	84.11	85.36	83.61	78.25
ΔE_{el-st}	-58.04	-85.81	-68.35	-77.94
$\Delta E_{nonorb-int}$	26.07	-0.50	15.26	0.31
$\Delta E_{orb-int}$	-26.42	-33.41	-30.00	-36.40
ΔE_{total}	-0.35	-33.85	-14.74	-36.09

pKa ANALYSIS

Bonding and structural analysis of Cu(I) coordination geometry in Mtb-CsoR prove that for the formation of hydrogen bond between Nε-His and Tyr-OH and Glu-COO-, both Cys36 and Cys65 should be deprotonated. In addition to the above discussed computational studies, theoretical pKa calculations were performed to validate the shift in pKa's that triggers the deprotonation of Cys36 and Cys65 when bound to Cu(I). Both histidine and cysteine prefer to bind to borderline soft metal cations such as Cu(I).[155,156] At physiological pH, cysteine side chains with typical pKa values between 8 and 9 would be protonated in metal-free proteins.[157] Binding to a metal cation (acting as a Lewis acid) decreases the pKa of

cysteines,[156,158] facilitating the deprotonation of sulfhydryl group under physiological conditions.

Small models of Model **A**, **B**, **C** and **D** were built with different protonation states of cysteines and the free energies of proton loss are calculated using the Equation 3.6. Subsequently, the pKa of each deprotonation step is calculated and are presented in Table 3.3. The first deprotonation of Cys65 sulfhydryl side chain has a pKa of 3.6 and that of Cys36 is calculated to be -1.6, indicating that when bound to Cu(I) Cys36 is likely to deprotonate first. The second deprotonation of Cys65 with Cys36 in deprotonated state is calculated to be 7.7 and for that of Cys36 when Cys65 is deprotonated is 2.5. From these results, we can interpret that the two copper coordinating cysteines are present in the deprotonated form (Model **A**) even considering the error associated with the computed values (+/- 0.5 pH units). The pKa of cysteines in Mtb-CsoR shifts by 9 pH units and these relatively large shifts in pKa of cysteines were previously reported in Cu(I) binding chaperone CopZ in *Bacillus subtilis*.[159] The two pKa's of cysteines coordinating to Cu(I) in CopZ are experimentally predicted to be <4 and ~6.[159] The acid–base properties of thioredoxin-like proteins, which also contain a CXXC motif, have been extensively characterized and very wide range of pK_a values are found in these proteins, from ~3.5 for the first cysteine of the active site motif of *E. coli* DsbA to ~8.8 for the same cysteine of *B. subtilis* ResA.[160-163]

Table 3.3.pK$_a$ values calculated for all the QM/MM models

| Description	pKa
Cys65-SH (first deprotonation)	3.6
Cys36-SH (second deprotonation)	2.5
Cys36-SH (first deprotonation)	-1.6
Cys65-SH (second deprotonation	7.7

From the above presented quantum chemical calculations we were able to understand the Cu(I) first coordination sphere geometry and identify the proposed hydrogen bond between Cu(I) coordinating His61 and second coordination sphere Tyr35 and Glu81.[104,117] QM/MM structural studies and pKa analysis predicted a valid model for the coordination of Cu(I) by Mtb-CsoR protein. The geometry of Cu(I) in Mtb-CsoR is described to be a tricoordinate [2+1] binding mode, with two strong bonds between Cu(I) and cysteines and a weak interaction between Cu(I) and histidine. Due to a stronger interaction between N$_\delta$ of His61 and Cu(I) (Models **B**, **C** and **D**) the pull exerted by the movement of histidine towards Cu(I), increases the distance between histidine and second coordination amino acids Tyr35 and Glu81, eventually leading to loss of hydrogen bonds. Therefore, an optimal bonding distance is necessary between Cu(I) and N$_\delta$ of His61 for the formation of H-bond between Tyr35 and His61. Moreover, the

bond distance is a measure of bond strength which was perceived to be influenced by the protonation state of the cysteines bound to Cu(I) and this phenomenon suggests that changes in the pH may have an underlying effect on the allosteric mechanism in Mtb-CsoR protein. Though further experimental studies are required to validate the above finding. Overall, quantum calculations were successful in providing a clear understanding of electronic structure of Cu(I) binding to Mtb-CsoR. The experimental Cu(I) binding studies and X-ray structures support our predicted structures and binding energies. Hence, Model **A** structural parameters were taken as a reference to build force fields for the Cu(I) bound form of Mtb-CsoR. The details, protocol and validation of the Cu(I) force fields were presented in the next section.

FORCE FIELD DEVELOPMENT

The copper bound CsoR models were simulated using the specifically developed force fields for Cu(I) ion trigonal coordination. The metal–ligand force field derivation is highly nontrivial and tedious. A small model was built to represent the Cu(I) coordination. Model **A** structure was used for the parameterization of new force fields. The geometric parameters, atomic charges (optimized geometries) and force constants (vibrational frequency calculations) for Cu(I)-ligand complex were taken from the small model complexes described

above, which were optimized using DFT at the B3LYP/6-31G**/LACVP level of theory (see QM computational details). The force-constants and ESP charges were iteratively adjusted until the normal modes derived from the quantum mechanical (QM) and molecular mechanical (MM) calculations were in good agreement. Figure 3.8 shows the correlation plot between MM normal modes generated using the new force fields and QM normal modes for Cu(I)—His—2Cys complex, which is used to validate the developed force fields. The Lennard-Jones potentials for Cu(I) non-bonded interactions were taken from corresponding values in gromos45a3 force fields. These parameters were then transferred to the topology files for each full-length CsoR-Cu(I) complex, replacing the original Gromacs-generated parameters for the corresponding atoms. The partial charges on the neighboring carbon atoms were adjusted such that the overall charge after adding the QM-derived charges was the same as in the MM model. The derived force constants for bonds, angles and dihedrals are tabulated in Table 3.4.
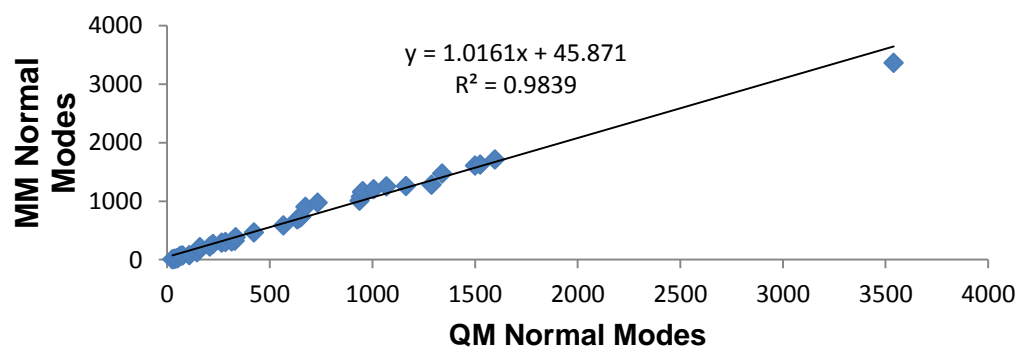
Figure 3.8. Correlation plot between the normal models obtained from MM and QM calculations

Table 3.4. The bond lengths, angles, improper dihedrals and the force constants developed for Cu(I) complex

Bonds	Bond length, nm	Force constant, K_b
Cu−SG$_{Cys36}$	0.235	48×10^4
Cu−SG$_{Cys65}$	0.229	58×10^4
Cu−ND$_{His61}$	0.212	23×10^5

Angles	θ	Force constant, K_θ
SG$_{Cys65}$− Cu−S$_{Cys36}$	129.2	536.99
SG$_{Cys65}$− Cu− ND$_{His61}$	117.1	634
SG$_{Cys36}$− Cu− ND$_{His61}$	113.6	634
Cu−SG$_{Cys65}$−CB$_{Cys36}$	107.0	481
Cu−SG$_{Cys65}$−CB$_{Cys65}$	104.3	481
Cu− ND$_{His61}$−CE$_{His61}$	116.9	574.51
Cu− ND$_{His61}$−CG$_{His61}$	136.9	574.51

Out of Plane Bending / Improper Dihedrals	Planarity
ND$_{His61}$−SG$_{Cys65}$−SG$_{Cys36}$− Cu	Planar
ND$_{His61}$−CE$_{His61}$− NE$_{His61}$− Cu	Planar

After the derivation of parameters for the Cu(I) coordination, the protein

was simulated using MD simulations to predict high resolution structures of both

the Cu(I)-bound and apo Mtb-CsoR to predict the structural change involved in allosteric regulation. Both the dimeric and tetrameric forms of CsoR were evaluated and analyzed for the differences in structure and hydrogen bonding network.

MD SIMULATIONS OF CU(I) AND APO CsoR DIMER

In order to analyze residue interactions between the Cu(I) binding region and allosteric networks involved in transmission of conformational change in Mtb-CsoR protein, equilibrium MD simulations were employed. Molecular dynamics methods have been used with a variety of proteins to recognize key residue interactions and networks that are involved in conformational changes as described in Chapter 2 of this thesis. We performed MD analysis on both the apo and Cu(I)-bound states of Mtb-CsoR dimer as an initial step to help identify functionally important residue interactions. Earlier experimental studies on Mtb-CsoR suggest that Mtb-CsoR is present as dimer in solution,[104] as opposed to the recent experimental work that reveal the evidence for the existence of tetrameric form of CsoR in *Streptomyces lividans*, *Thermus thermophillus* and other bacteria.[106-109] MD simulations were performed on Cu(I) and Apo CsoR dimer to identify the residue interactions in the second coordination sphere of Cu(I) and to analyze the communication pathway initiated on binding to the metal. Additionally, MD

simulations were also performed on the tetrameric form of CsoR, to understand

the residue interactions and conformational changes between and within the two

dimeric units of CsoR that control the overall conformational change in the CsoR

tetramer. The crystal structure of Cu(I)-bound Mtb-CsoR was published[104] and was

used as a starting structure to optimize the Cu(I) binding geometry and to assign

bonds between the metal and the coordinating residues. The crystal structure is

missing a 21 amino acid C-terminal tail and 3 amino acid residues in N-terminal

and hence our computational model starts with Glu4 and ends with Phe88. An

attempt was made to build the C-terminal tail using our simulations and

homology modeling techniques, but due to the unstructured nature of tail and also

the existence of several orientations of the tail with respect to the CsoR protein, the

assignment of inter helix-tail interactions was hampered. Additionally, each of

these orientations will possess a minima and the selection of the relevant

minimum energy structure would be difficult without any supporting

experimental data.

The starting structure for simulation apo CsoR was based on the X-ray

crystal structure of Cu(I)- bound Mtb-CsoR with missing C-terminal tail and the

first three N-terminal residues. Equilibration of the Cu(I)-bound Mtb-CsoR protein

with no Cu(I) ions was performed using MD simulations to obtain a reasonable

structure of the apo form and to evaluate the changes in quaternary structure of

Mtb-CsoR on loosing Cu(I) ions. As the Cu(I) Mtb-CsoR protein structure was

stabilized by the electrostatic potential created by the two Cu(I) ions, the initial

simulation of Cu(I)-bound Mtb-CsoR structure with no Cu(I) ions resulted in the

unfolding of the protein and loss of secondary structure was observed. Several

minimizations and equilibrations were carried out in order to achieve a stable

structure during the simulation by gradually decreasing the forces acting on the

Cu(I) binding residues. Subsequently, gas phase calculations were conducted

employing the simulated annealing procedures that allow the protein to traverse

different conformational landscapes and reform the contacts between the helices.

In the simulated annealing procedure, the temperature of the system was steadily

increased from 270 K to 320 K in 10 steps and subsequently was cooled down to

300 K. The most populated structure from the simulated annealing procedure was

solvated in 20,766 water molecules (solvent). Several cycles of minimizations and

equilibrations were performed with a constraining force on the protein, allowing

the water molecules to orient around the protein. The total charge on the apo Mtb-

CsoR protein was zero. Finally the protein was simulated for an additional time

20 ns of production simulation, with a total of 50 ns simulation including the

equilibration step. As the crystal structure of apo form of Mtb-CsoR was not solved

using the X-ray crystallography, the average structure obtained from the 50 ns of

MD simulation was validated using the X-ray crystal structures of *T.thermophillus*

and *S.lividans* apo CsoR. In order to understand the conformational space traversed by the apo form of *T.thermophillus* CsoR and *S.lividans* CsoR, MD simulations were also performed on the corresponding apo forms using their crystal structures.

MD SIMULATIONS OF APO *Thermus thermophillu*s (Tt) CsoR

The thermophilic bacterium *T. thermophilus* belongs to the phylum *Deinococcus–Thermus*, and thus, is only distantly related to Mtb, Sl etc.[106] *T. thermophilus* constitutes of the *copZ-csoR-copA* operon, which is repressed by CsoR under copper-limiting conditions.[106] *In vitro*, *T. thermophilus* CsoR was shown to be promiscuous in binding various metal ions, including Cu(I), Cu(II), Zn(II), Cd(II), Ag(I) and Ni(II), all of which release CsoR from the *copZ* promoter.[106] *In vivo*, it is reported that copper and zinc ions significantly increase *copZ-csoR-copA* expression, suggesting that the response of CsoR to various metal ions is also physiologically relevant.[106] The copper ion binding motifs of most CsoR family proteins contain C-H-C,[104,105,107,108] whereas the corresponding residues of T.thermophilus are C-H-H.[106] The 3D crystal structure of Tt-CsoR in its tetrameric form was determined at a resolution of 2.1 Å.[106] The tetrameric structure is composed of two dimers, i.e. chains AB and CD.[106] Several disordered main-chain regions that were not included in the X-ray crystal structure (M1–L6, V71–G78 and

Y93–R94 of chain A; M1–S4, T73–E82 and K92–R94 of chain B; M1–L6, V71–D79

and Y93–R94 of chain C; and M1–S4, A72–E82 and Y93–R94 of chain D) were

constructed using protein building and modeling tools. The coordinates of the side

chains of K11, D79, E85 and E86 of chain A; E37, K38, E85 and E86 of chain B; K11,

E14, E82, E85, E86 and K92 of chain C; and E37, K38 and E85 of chain D were also

not determined due to their poor electron densities. Using loop building tool

PrimeX[164] and homology modeling techniques[165] the missing residues and loops

were constructed. The reconstructed crystal structure of Tt-CsoR was simulated

using MD simulations to understand the residue contacts in apo form of CsoR.

Though the Cu(I) ion binding motifs and affinity of Tt-CsoR is different from Mtb-

CsoR, the interactions of the conserved residues were assumed to be homologous.

Simulated annealing was performed to re-establish the α-helix contacts in the

refined X-ray crystal structure. The most populated structure generated by the

simulated annealing techniques was solvated in 40,371 water molecules.

Production run of 20 ns time scale was performed and most populated conformers

were obtained from cluster analysis of the simulation and is shown in Figure 3.9.
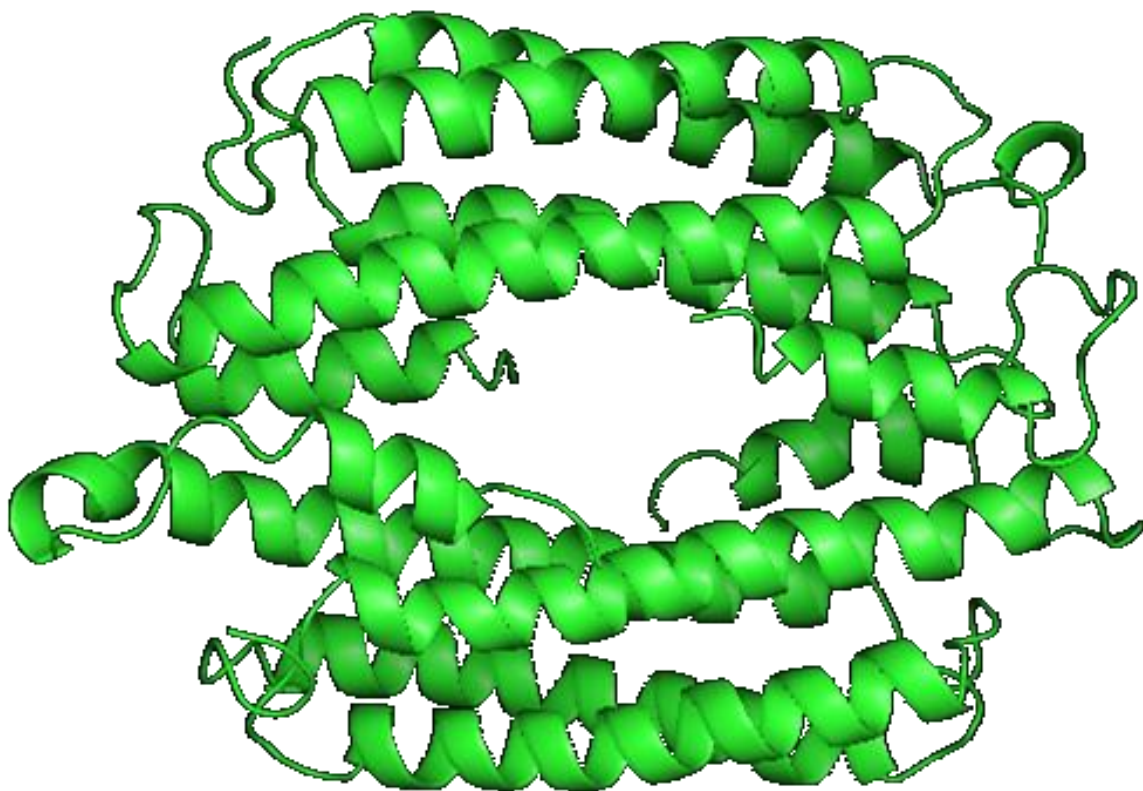
Figure 3.9. Average structure of the apo Tt-CsoR protein from 20 ns MD simulation

MD SIMULATION OF APO STREPTOMYCES LIVIDANS (SL) CSOR

The crystal structure of apo form of Sl-CsoR of 1.7 Å resolution was used as a starting structure in our MD simulations. Simulated annealing and production MD of 20 ns were performed including 42,645 waters. The most populated structure was derived using cluster analysis and is shown in Figure 3.10.
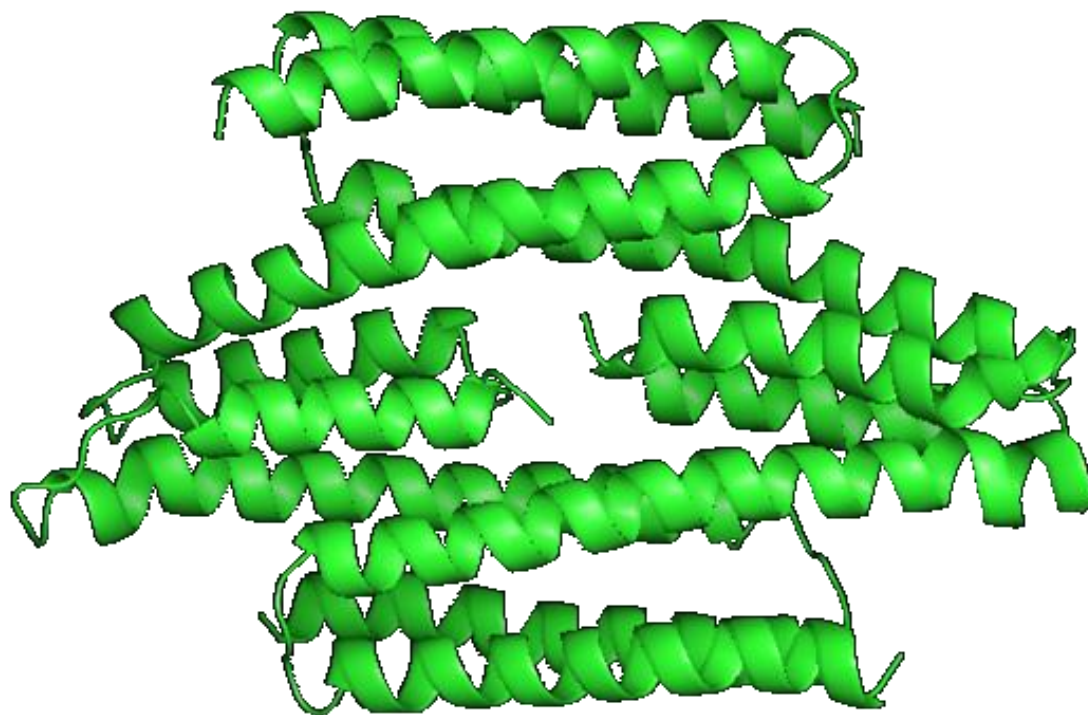
Figure 3.10. Average structure of the apo Sl-CsoR protein from 20 ns MD simulation

MD SIMULATIONS OF CU(I)-BOUND MTB-CSOR

As the founding member of CsoR family, Mtb-CsoR has been characterized using biological, biophysical and structural methods in Cu(I)-bound state. The 2.6 Å crystallographic structure of Cu(I)-bound CsoR was used as a starting structure for the QM/MM calculations. The minimum energy structure obtained from the QM/MM calculations was used in MD simulations. The developed parameters for the Cu(I) coordination geometry were added to gromos 43a2 force fields. The extended force field with the electrostatic charges from the DFT calculations was used in order to account for the effect of change in electrostatic potential of the protein on binding to Cu(I). To validate and test the developed force fields and

proposed bonding model from our QM/MM calculations rigid bond model of Cu(I) binding geometry was simulated with the inclusion of corrections for the electrostatic charges to the force field.

The Cu(I) Mtb-CsoR structure was initially equilibrated by freezing the bonds in the first coordination sphere of the copper at the QM/MM optimized geometry. As Cu(I) remained bound to CsoR protein and the bond distance between the amino acids and copper remained the same throughout the simulation, the constraints were subsequently removed. Charges of the first coordination shell were adjusted using the CHELPG charges such that each copper coordination site has a charge of -1. Hence the resulting charge on the Cu(I)-bound Mtb-CsoR dimer is -2. In the MD simulation, 0.2 M NaCl is added to the solvent to neutralize the protein charge and to simulate and mimic the ionic concentration in the experiments.

STRUCTURAL ANALYSIS OF THE DIMER

In accordance with the initial experiments of Mtb-CsoR,[104] the dimers of both the Cu(I)-bound and apo Mtb-CsoR were compared structurally to examine any changes in the orientation of the helices. The most populated structures were obtained from the simulations in dimeric state. Initial benchmarking was performed by comparing the X-ray crystal structure of Cu(I)-bound Mtb-CsoR and

most populated structure of the protein obtained from the computational simulations using the secondary structure alignment methods. The computed model exhibited several similarities with that of crystal structure, reproducing the experimentally observed hydrogen bonding interaction between the side chains of Tyr35, His61 and Glus81. Superposition of the two structures by secondary structure matching gave a root mean square deviation of 1.15 Å as shown in Figure 3.11. Unnatural amino acid substitution studies[117] on Mtb-CsoR indicate that this H-bond interaction is required to initiate the allosteric switching on binding to Cu(I). In the dimeric Cu(I)-bound Mtb-CsoR, the H-bond distance between $N\varepsilon$ atom of HIS61 and OE1 of Glu81 varied from 1.9 – 2.4 Å and that of OH atom of Tyr35 and OE1 of Glu81 was observed to be between 1.7 -2.5 Å. In some conformational clusters of Cu(I)-bound Mtb-CsoR, Tyr35 was perceived to act as an H-bond bridge between His61 and Glu81 and in other clusters the H-bonding triad between Tyr35-His61-Glu81 was observed as shown in Figure 3.12.
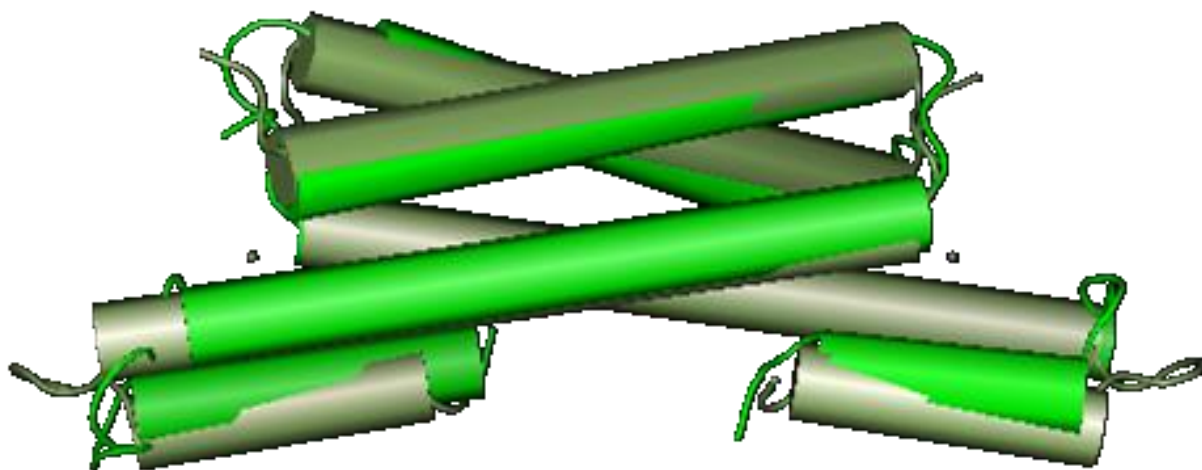
Figure 3.11. Overlay of X-ray crystal structure (silver) and computed structure of Cu(I)-bound Mtb-CsoR (green)

In the apo form of Mtb-CsoR dimer, the above mentioned H-bonds are not detected in the simulations. While, an H-bond interaction (2.6/2.7 Å) involving OE1 atom of Glu-122 and the OH atom of Tyr74 and a bridging water molecule which further H-bonds (2.7/2.8 Å) with OE1 of Glu122 and Nε atom of His100 (2.7/2.6 Å) was reported in apo form of Sl-CsoR in its homotetrameric form.[107] Although, these H-bond interactions were not reported in the apo CsoR of *Thermus thermophilus* in its homotetrameric state.[106] Instead, in the dimeric apo Mtb-CsoR the copper binding amino acid residues were identified to involve in H-bond interactions with the amino acid residues present in the N-terminus and in the unstructured loop region that connects the α2 and α3 helices. Key H-bond interactions observed in the dimeric form of Apo-CsoR were between OH atom of

Tyr35 to the N-terminus of Glu4 and between the backbone of Cys36 and OE1 of Asp72 shown in Figure 3.12. The interaction with the N-terminus could be an artifact of the computational model, due to the lack of first three amino acids that constitute the N-terminal of apo CsoR.

The apo and Cu(I) bound form of Mtb-CsoR obtained from the MD simulation were evaluated for the presence of structural changes by implementing global alignment of secondary structure of two form and is shown in Figure 3.13. The secondary structure alignment of the two structures demonstrates that there is significant structural change in the apo form that manifests in $\alpha3$ and $\alpha3'$ helices and these helices tend to move towards the $\alpha2$ and $\alpha2'$ helices. The unstructured loop connecting the $\alpha2$ and $\alpha3$ helices in the dimeric apo form was perceived to interact with copper binding amino acids through the H-bond interactions described previously and therefore closing the metal binding site shown in Figure 3.12.
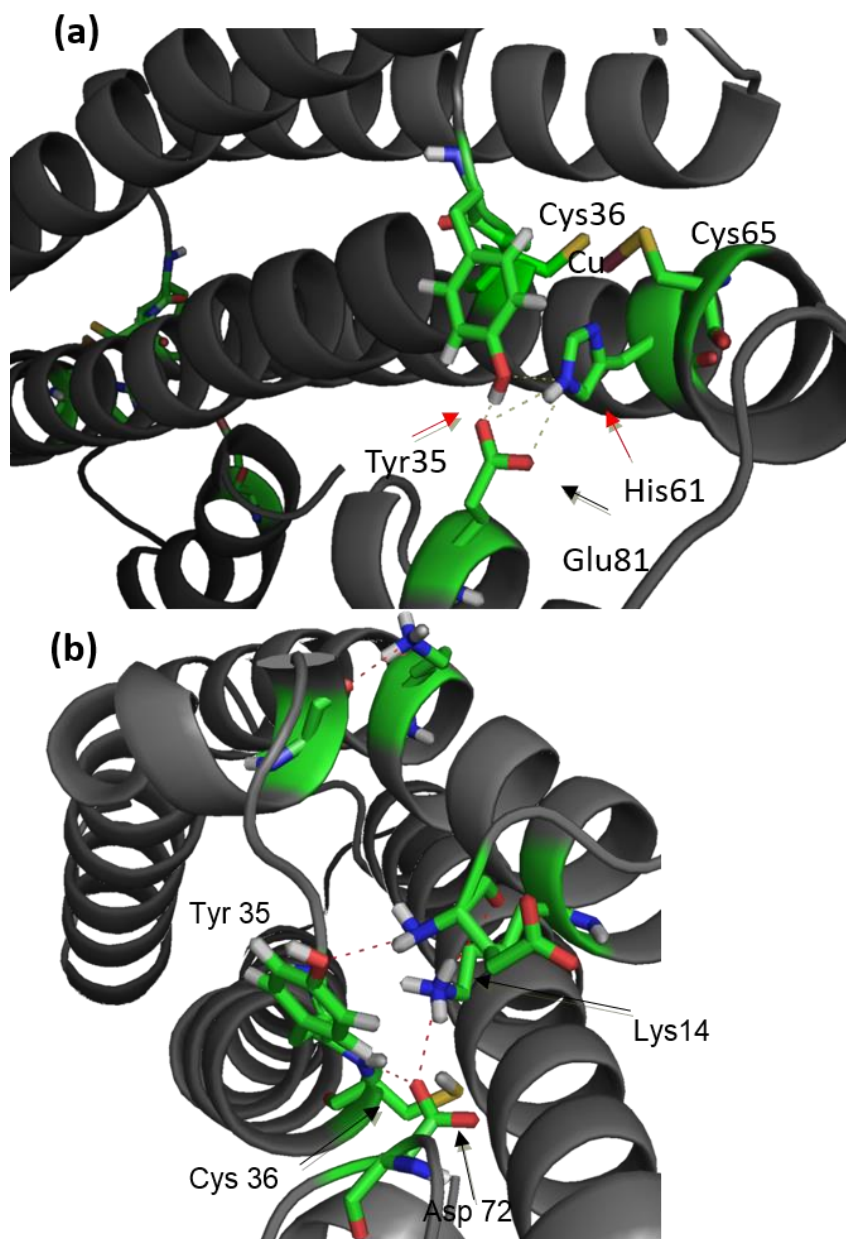
Figure 3.12. (a) Hydrogen bonding network stabilizing the two protomers in Cu(I)-bound Mtb-CsoR dimer (b) Hydrogen bonding network stabilizing the two protomers in apo Mtb-CsoR

These results predict an overall change in the alignment of the α-helices in apo CsoR on binding to Cu(I). Though the simulations of dimeric form of Mtb-CsoR resulted in key findings which are in agreement with the experiments,[104] this

model presents certain drawbacks such as the higher flexibility in the truncated $\alpha 3$ helix (missing C-terminal tail). In the dimer of dimer architecture of Mtb-CsoR, this flexibility will be attenuated by the presence of adjacent dimer that restricts the motion of $\alpha 3$ helix. Henceforth, particular emphasis was placed on dimer of dimers structure of Mtb-CsoR in computational studies which will enable us to understand the allostery in CsoR proteins of other bacteria that do not possess the C-terminal tail. The computed dimeric forms of apo and Cu(I)-bound Mtb-CsoR were also used to develop the homotetrameric structures, using the Sl-CsoR tetramer as a template.
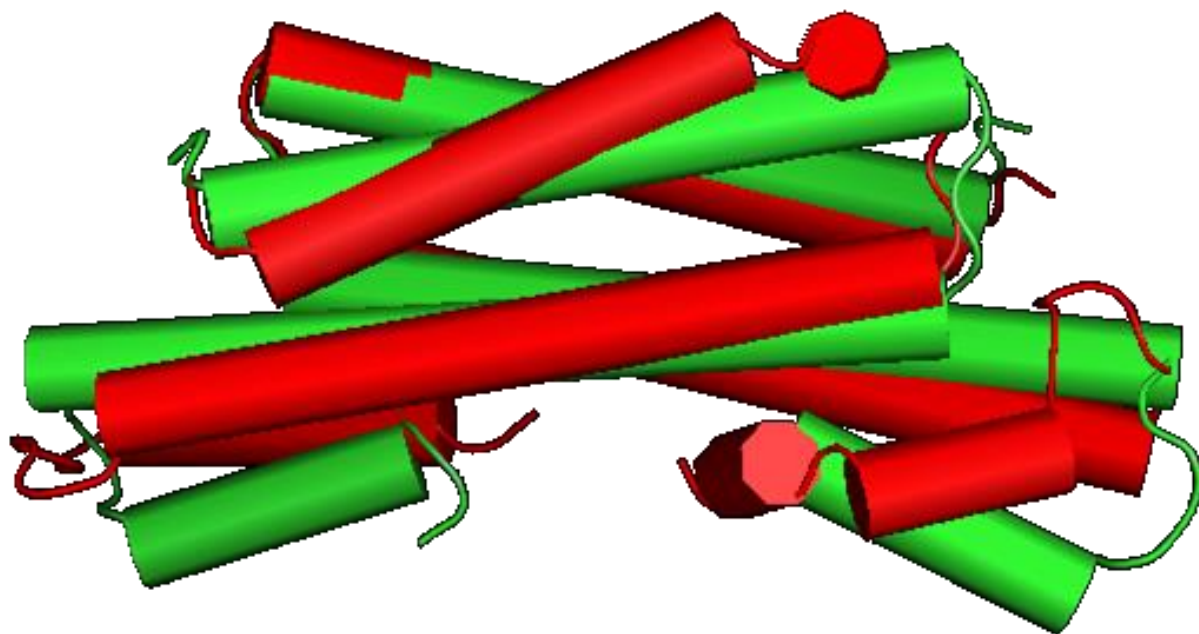
Figure 3.13. Overlay of computed structure of Cu(I)-bound (green) and Apo (red) Mtb-CsoR.

**ANALYSIS OF CU(I) AND APO CsoR TETRAMER**

Hydrogen Bonding Network

The H-bond interactions reported and analyzed in the earlier sections are examined in the simulations of tetrameric form of both apo and Cu(I)-bound Mtb-CsoR. Plots of H-bond interactions between Tyr35, His61 and Glu81 of the adjacent protomer for all the four Cu(I) binding sites were generated for both the apo and Cu(I)-bound Mtb-CsoR shown in Figure 3.14. At each copper binding site, three H-bonds are possible with a maximum of twelve H-bonds for each tetrameric CsoR. In the plots shown in Figure 3.14(b) a maximum of five H-bonds are present in the tetramer between OH atom of Tyr35 and OE1 or OE2 of Glu81, suggesting that this H-bond interaction is present in all the metal binding sites with an exception of five H-bonds when Tyr35 is in H-bonding distance from both OE1 and OE2 atoms of Glu81. From these plots, it can be established that at least two metal binding sites are involved H-bond interactions during the first 0-8 ns time frame. Loss of this H-bond interaction was observed after 8 ns and reestablished around 9.5 ns. In the subsequent plot, mapping the H-bond interaction between Tyr35 – His36 – Glu81 a maximum of six H-bond interactions were observed suggesting the prediction that, these interactions can be disrupted by solvent water and not all the metal binding sites possess these H-bonds concurrently

115

throughout the 10 ns simulation. Surprisingly, none of these H-bond interactions were observed in the apo form of Mtb-CsoR. Instead, new H-bond interactions were identified involving OH atom of Tyr35 and Nε-H atom of His61. These results substantiate that the conformational switch on binding to copper is initiated by this hydrogen bonding network which may be very crucial in propagating allostery. As stated in the discussion of Mtb-CsoR dimer, in apo form of Sl-CsoR an interaction between Tyr74 (Tyr35 in Mtb-CsoR) and Glu122 (Glu81 in Mtb-CsoR) with a bridging water molecule was reported.[107] In our simulations, we did not identify any such bridging water molecule between His61 and Glu 81, in both the dimeric and tetrameric forms of Mtb-CsoR. But in the simulations of apo form of Sl-CsoR the above described interaction (Figure 3.16) was observed in accordance with the experiments. Analysis of H-bond interactions in apo Mtb-CsoR reveals alternative interactions that are established between the Nε-H atom of His61 and OE1 atom of Asp72, backbone C=O of His74 and Gly73 of adjacent protomers shown in Figure 3.15(b). The loop regions connecting the α2 and α3 helices constitutes of Asp72, Gly73, His74 and Gly73 is conserved in the CsoR proteins of bacteria.[104,111] Additional interactions were observed between OH atom of Tyr35 and OE1 of Asp38 and Asp32. Asp 38 is also conserved in the CsoR proteins of bacteria, suggesting a key role of these H-bonds in stabilizing the apo CsoR conformation and in the transmission of allosteric transitions.

**Secondary Structure Analysis and Radius of Gyration (Rg)**

NMR studies conducted on *G. thermodenitrificans* CsoR in both the apo and Cu(I)-bound states revealed that Cu(I) binding results in an discontinuity or kink formations in the long α2-helix located between the Cu(I) binding residues His75 and Cys79.[108] Crystal structures of apo CsoR proteins have more recently been solved from *S. lividans*[107] and *T. thermophilus*.[106] These structures reveal a similar architecture, with the *S. lividans* structure characterized by a continuous α2 helix in the apo form, just as in apo *G. thermodenitrificans* CsoR in solution.[108]
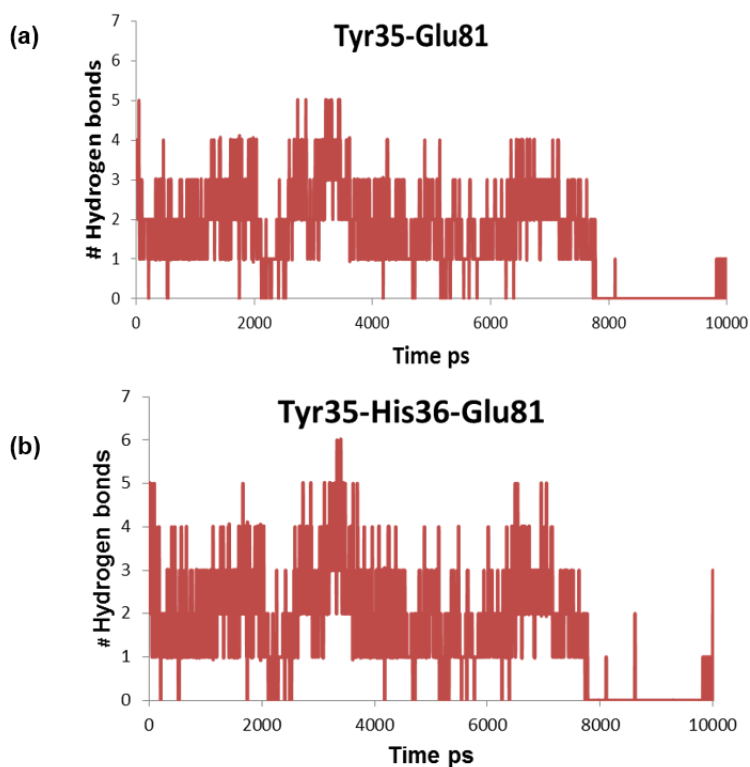
Figure 3.14. (a) Number of hydrogen bonds formed between Ty35-Glu81 in Cu(I) Mtb-CsoR **(b)** Number of hydrogen bonds formed between Ty35-Glu81-His61 during the simulation of Cu(I) Mtb-CsoR
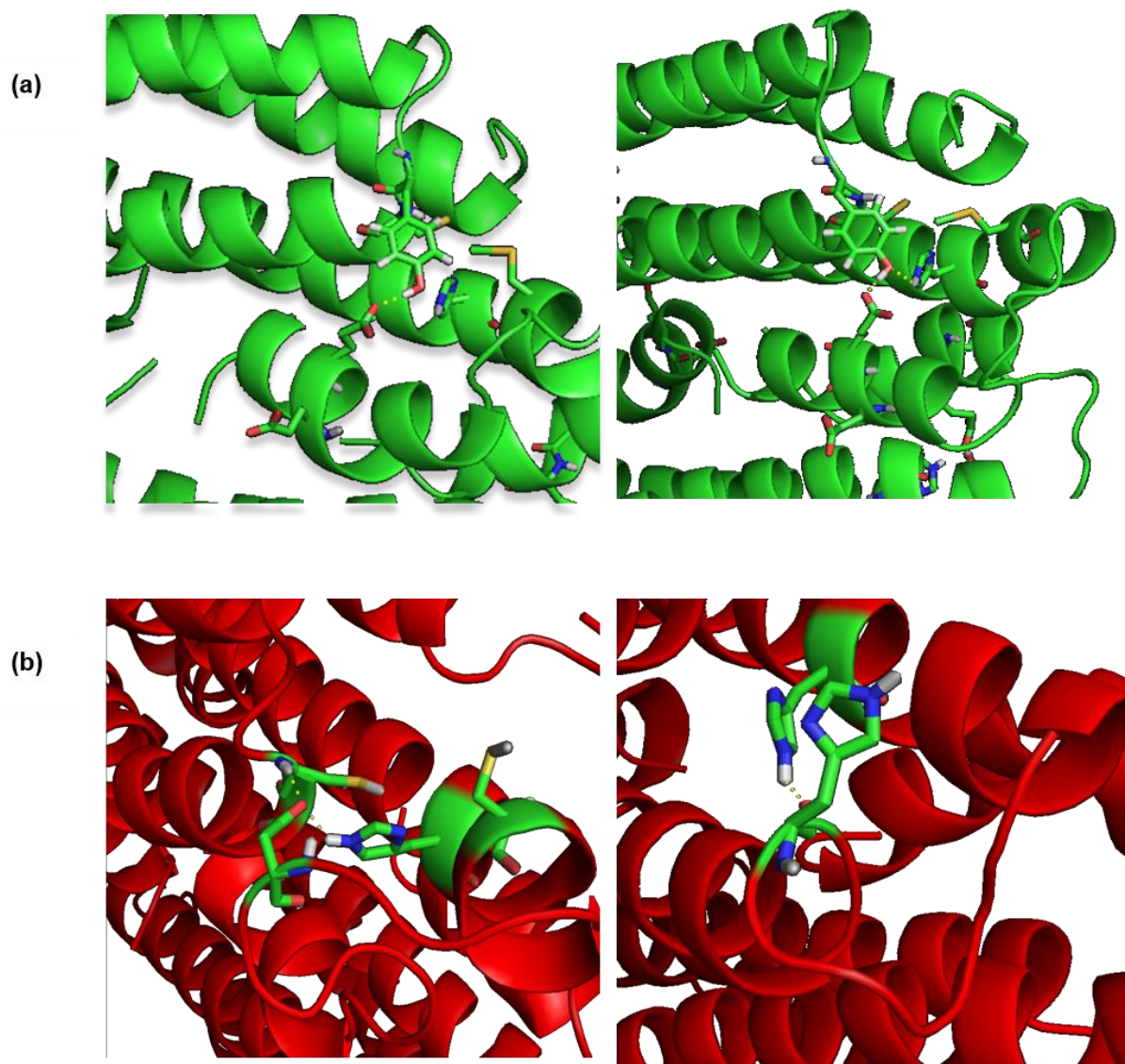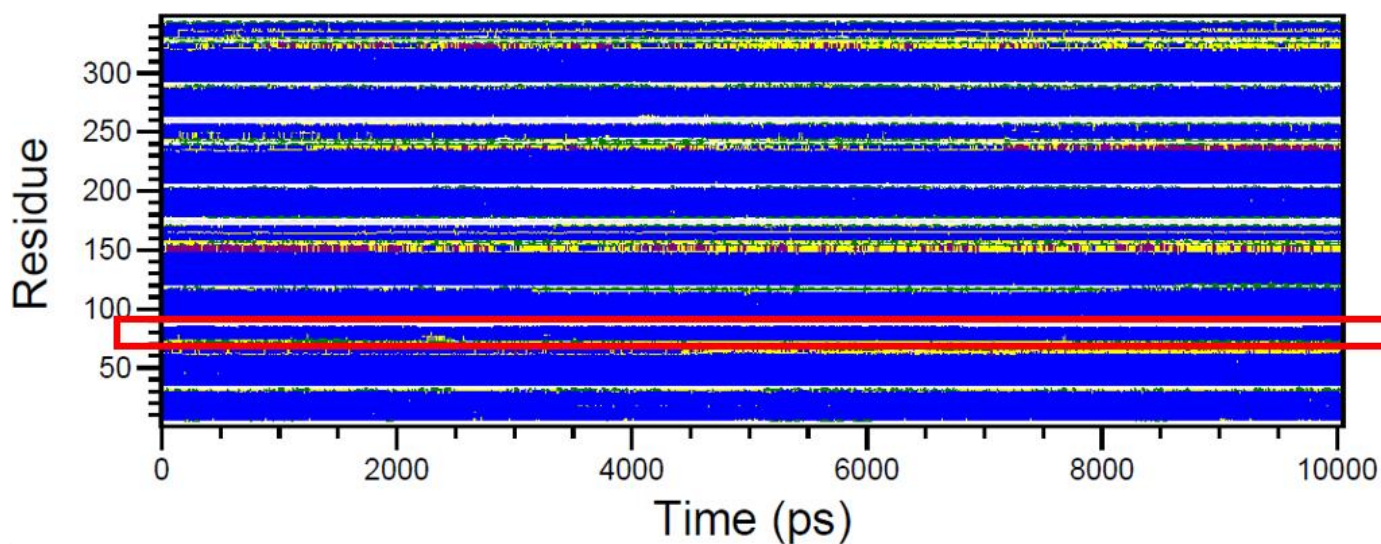
Figure 3.15. The hydrogen bonding network reproduced in the average structures of the most populated clusters (a) Cu(I)-bound Mtb-CsoR (b) Apo Mtb-CsoR
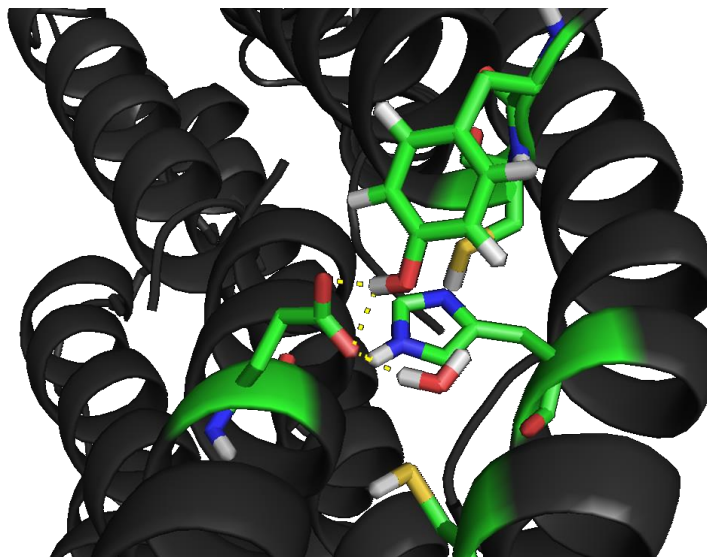
Figure 3.16. Hydrogen bonding network and bridging water molecule stabilizing the two protomers in apo Sl-CsoR tetramer

To identify the nature of change or disorder in the structure of Mtb-CsoR on binding to Cu(I), secondary structure maps were generated from the simulations of the apo and Cu(I)-bound Mtb-CsoR. Interestingly, as seen in Figure 3.17 the α-helical character in the Cu(I) bound Mtb-CsoR was lost and a turn or bend-like structure is initiated at His59 and Asn60. This region accommodates His61 and Cys65 residues which are involved in copper coordination. These results quantify that the α2 helix around the copper coordinating residues starts to unfold and transform into a turn. The loss of secondary structure develops into a kink in the α2-helix, and the α2-helix bends towards the Cu(I) binding site to accommodate the trigonal planar coordination of Cu(I). The secondary structure map is projected on to the CsoR structure to show the formation of a kink and loss of α-helical character in Figure 3.18.

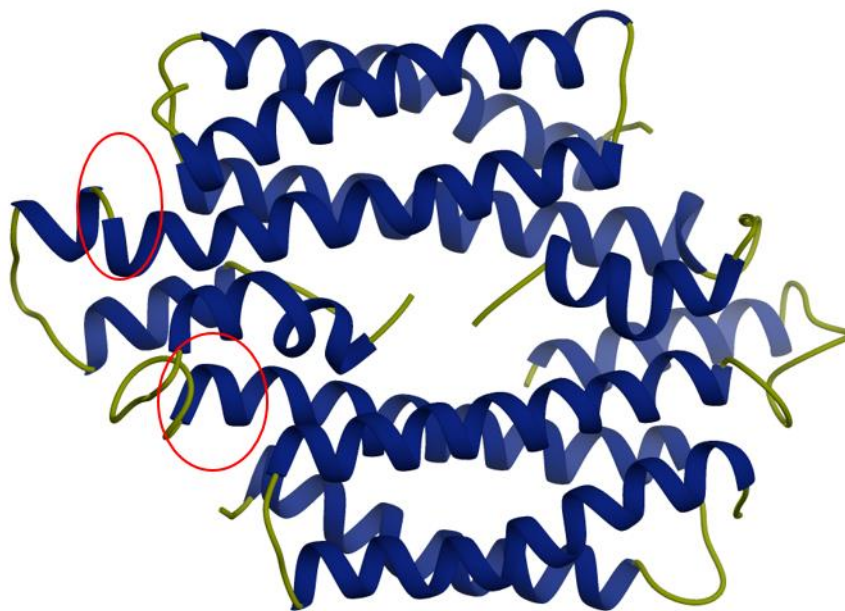**Figure 3.17**. Secondary structure map of Apo Mtb-CsoR (top) and Cu(I) Mtb-CsoR (bottom) tetramer (constituting of 340 residues each).Turn initiation in Cu(I)-bound Mtb-CsoR at residue His59

Figure 3.18. Secondary structure of the Cu(I)-bound CsoR projected on to Mtb-CsoR tetramer. The loss of α-helical character is shown in red circles

This discontinuity in helical geometry of Cu(I)-bound Gt-CsoR was shown to be present in solution and was proposed to be contribute to the allosteric switching mechanism in Gt-CsoR.[108] Although the presence of the kink in the α2 helix was not reported in Mtb-CsoR in earlier studies,[104] this discontinuity in α2 helical geometry was also recently reported for Cu(I)-bound Mtb-CsoR.[108] A prominent aspect of structural transition in Gt-CsoR was described to be an interruption in the α2 helix in Gt-CsoR; which in turn was proposed to drive a global compaction of the tetrameric form. SAXS experiments revealed that CsoR is hydrodynamically smaller in the Cu(I)-bound state with radius of gyration Rg of 26.9 Å for apo CsoR and 25.1 Å for Cu(I)-bound CsoR of *Geobacillus*.[108] The calculated Rg values for apo and Cu(I)-bound Mtb-CsoR are 21.9 Å and 21.7 Å

respectively. Though the Rg value for apo CsoR is slightly higher than Cu(I)-bound Mtb-CsoR, these results were not very conclusive, due to the high flexibility of the unstructured loop connecting the $\alpha 2$ and $\alpha 3$ helices that might indeed affect calculation of Rg value.

Principal Component Analysis

The internal motions in a protein may reveal the transmission of a conformational adjustment to affect the binding or reactivity at an active site, as in allosteric effects. Such functional internal motions may involve complex correlations between atomic motions, and the nature of the motion is inherent in the protein structure. We investigated the correlations between atomic positional fluctuations in both apo and Cu(I)-bound CsoR protein, as derived from (nanosecond) MD simulations in aqueous environment. Applying the essential dynamics method, the covariance matrix was calculated and diagonalized for all atoms and backbone atoms of the protein to obtain the eigenvectors of the protein from these simulations. The RMSF values for the backbone atoms of each amino acid residue obtained from the significant eigenvectors of the protein are plotted in Figure 3.19. Perturbations were more pronounced near the Cu(I)-binding pocket and at the more peripheral loop regions from Ala69 to His74 (Ala66 to His71 in Figure 3.19) of the tetramer with relatively smaller changes in the α-helices

compared to the apo Mtb-CsoR. Large perturbations occur in the C-terminal α3-helix from Gly73 through the C-terminus in the apo form. The largest perturbation is observed in Asp72, which is perceived to originate from the formation of H-bond interaction between the OE1 or OE2 atom of Asp72 and the Nϵ-H atom of His61. The other large fluctuations are localized to the amino acids from Ser29 to Tyr35 (Ser26 to Tyr32 in Figure 3.20) near Cu(I) binding region. The apo-CsoR tetramer is observed to possess greater fluctuations and is globally more dynamic across the structured α-helical regions (red line in Figure 3.19). In contrast, Cu(I)-bound Mtb-CsoR is less globally dynamic, with perturbations observed in residues present in the α3 helix through the C-terminus. To further investigate the dynamic regions in both apo and Cu(I)-bound Mtb-CsoR, B-factors were calculated from the MD simulation data. The B-factors are a measure of fluctuations and quantify the degree of flexibility in a protein molecule. The calculated B-factor values were projected on to the CsoR structure in apo and Cu(I)-bound states shown in Figure 3.20. The fluctuations are color coded from red to blue, red depicting high fluxional regions and blue depicting the least fluxional regions. Apart from the unstructured loop regions which exhibit large fluctuations, Tyr35 was revealed to be highly fluxional in apo Mtb-CsoR. The arginine residues are shown to be perturbed to a larger extent compared to other amino acid residues in the structured α-helical regions of apo Mtb-CsoR. Arg10, Arg15 and Arg52 were identified to be the most

fluctuating arginines in the apo form, suggesting a rearrangement of arginine residues in the apo form to initiate DNA binding on loosing Cu(I) ions.[108] The structured α (1, 2 and 3) helices were observed to be rigid (blue in color) in Cu(I)-bound Mtb-CsoR as opposed to apo CsoR where certain regions of α2 and α3 helices are shown to possess higher flexibility. Essential dynamics and B-factor studies on CsoR reveal that the α-helices of the apo Mtb-CsoR tend to have higher flexibility (higher fluctuation) compared to those of Cu(I)-bound Mtb-CsoR, indicating that on binding to Cu(I) the α-helices assume a 'rigid or taut' structure that may affect the affinity to bind DNA promoter region.[112]
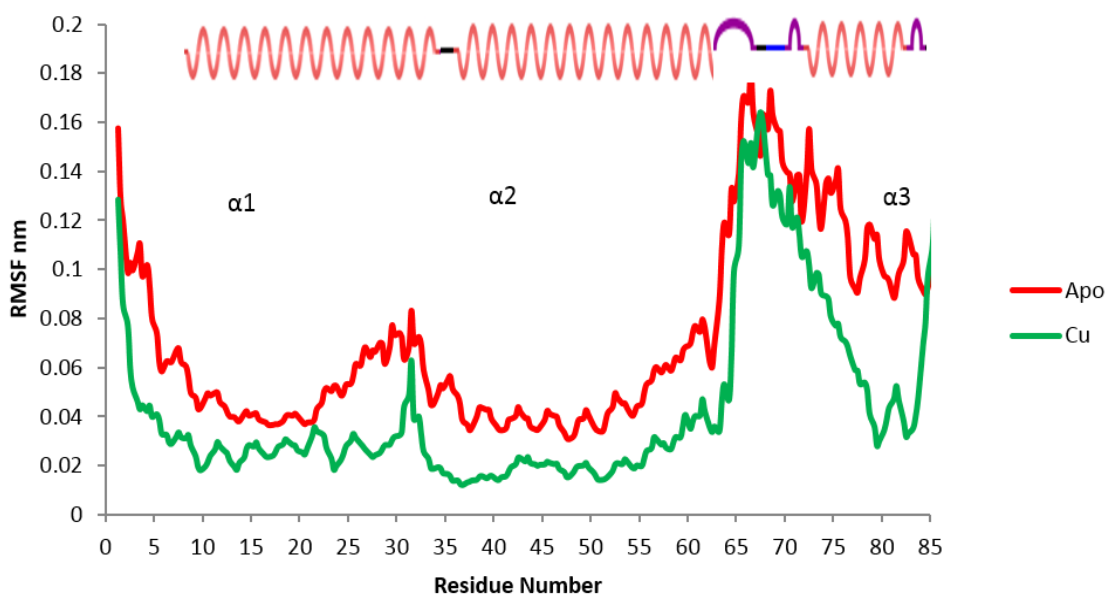
Figure 3.19. RMSF plot of protein backbone derived from the first eigenvector from covariance analysis
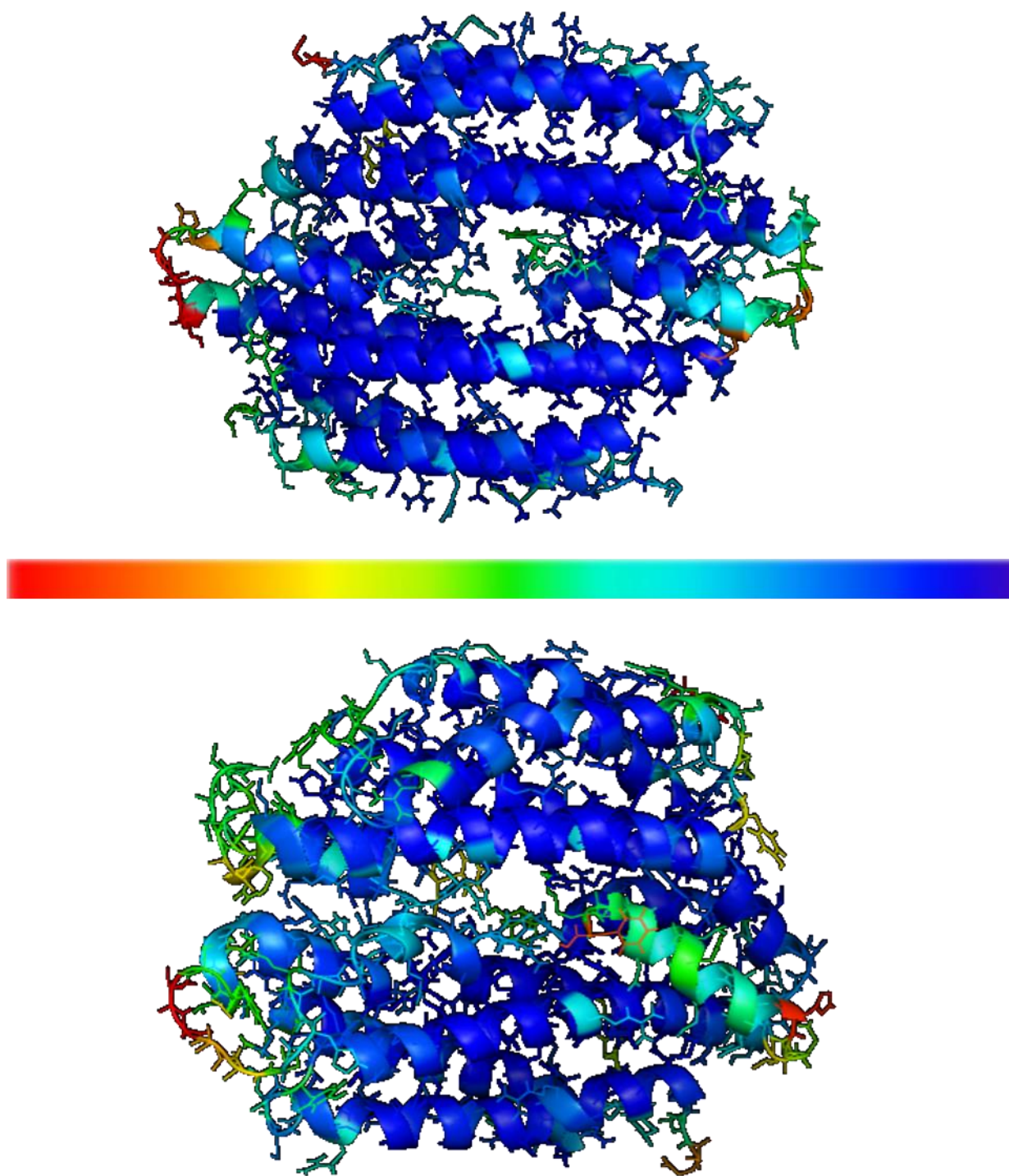
Figure 3.20. Structures of Cu-bound CsoR (top) and Apo-CsoR (bottom) with B-factors projected on to the structures, least fluctuating (blue) - most fluctuating (red)

Mechanism of Allosteric Conformational Switch in Mtb-CsoR

An hypothesis for the mechanism of allosteric Cu(I) regulation in the CsoR family has been put forward based on the crystal structure of Cu(I)-bound Mtb-CsoR[104] and was tested experimentally using unnatural amino acid substitutions of the Cu(I) coordinating His61 residue.[117] The unnatural amino acids substitutions were designed to eliminate the H-bond interactions between the N_ϵ atom of the imidazole ring and the second coordination sphere amino acids Glu81 and Tyr35.[117] Cu(I) binding to the non-native His61-substituted Mtb-CsoR did not significantly affect the Cu(I) binding affinity, but the allosteric coupling free energy (ΔGc) was determined to be close to zero.[117] This observation lead to the proposal that allosteric switching in Mtb-CsoR is initiated upon coordination of Cu(I) to the $N\delta$ atom of the His61 ligand, triggering the formation of a H-bond network through N_ϵ atom that results in dissociation of CsoR from the operator region of DNA.

In our computational models, we were able to reproduce the second coordination sphere H-bonding network initiated upon binding to Cu(I). These interactions were not detected in the simulations of apo Mtb-CsoR. Similarly, the H-bond interactions were not depicted in the apo form of Tt-CsoR as opposed to

apo SL-CsoR, in which these second coordination sphere H-bonds were described

both experimentally and computationally.
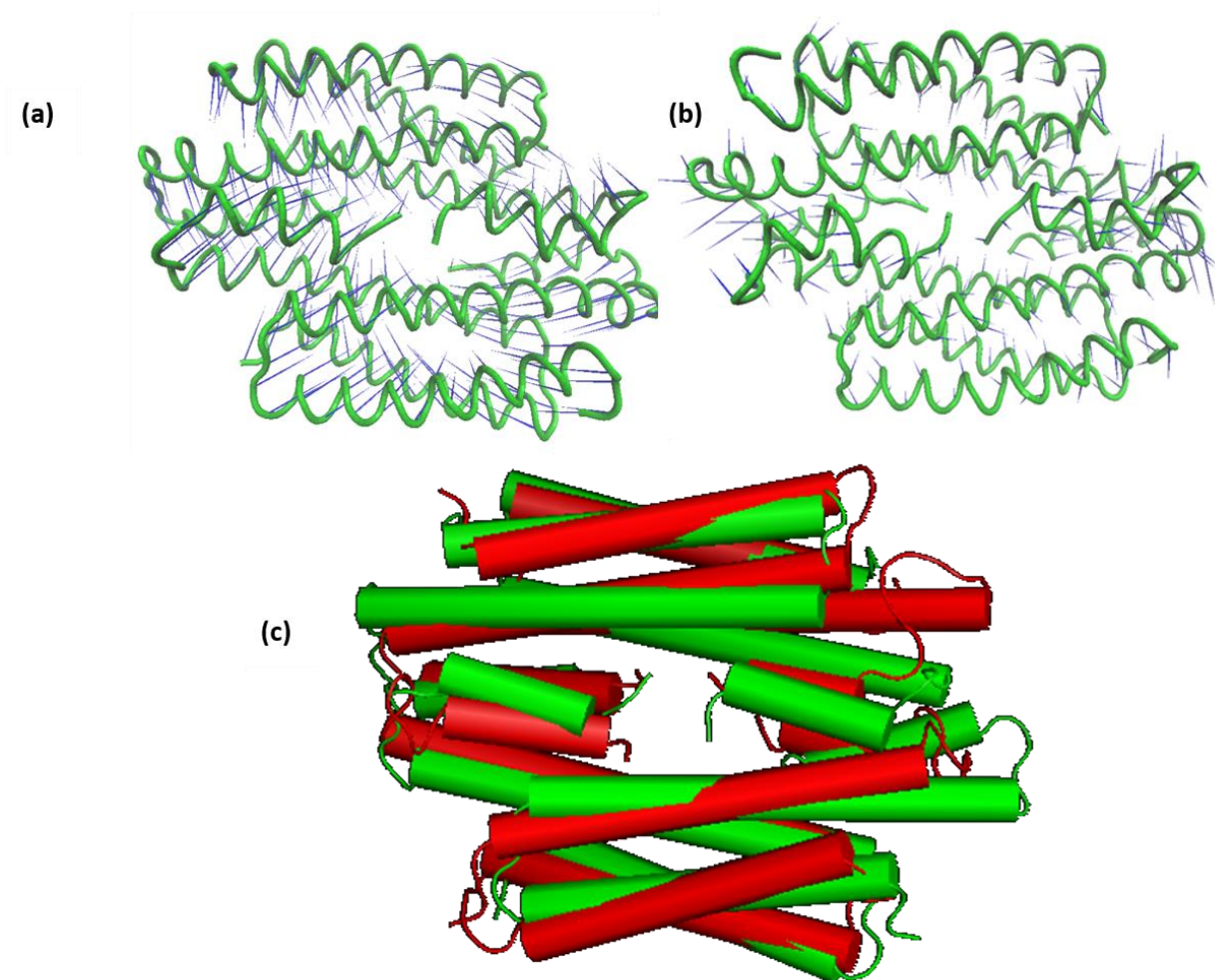
Figure 3.21. The vectors representing the direction of motion in (a) apo Mtb-CsoR (b) Cu(I)-bound Mtb-CsoR and (c) structural alignment showing different conformation of apo(red) and Cu(I)-bound(green) Mtb-CsoR

Comparison of H-bonding interactions, conformations and fast time scale

internal dynamic (ps –ns) motions of Cu(I)-bound and apo CsoR was performed

to demonstrate the allosteric switching in Mtb-CsoR. A global alignment of both

Cu(I) and apo forms of Mtb-CsoR in dimeric form (Figure 3.21(c)) reveals that the two protomers in the apo CsoR reorient to attain a 'X-like' conformation due to the absence of an anchor (which is Cu(I)) that holds the protomers in a 'taut' conformation.[112] The loop connecting the $\alpha 2$ and $\alpha 3$ helix interacts with Cu(I) coordinating amino acids and closes into the metal binding region in the apo form. A RMS value of ~4.42 Å was obtained for the alignment of dimeric form and ~6.20 Å for tetrameric form of Mtb-CsoR. In the tetrameric architecture, similar conformation changes were observed that are consistent with the DNA binding studies performed on Sl-CsoR and Gt-CsoR. The vectors of functional models derived from the essential dynamics analysis were projected on the most populated structure of Cu(I)-bound and apo Mtb-CsoR in tetrameric state in Figure 3.21. Comparison of these two structures unveils that the α-helices of apo form are highly flexible and this can also be noted from the length of the vectors that were scaled based on the degree of displacement from the average structure. The direction of the vectors suggest that the, N-terminal region and the loop connecting $\alpha 2$ and $\alpha 3$ helices, closes into the Cu(I) binding region in located in the periphery of $\alpha 1$, $\alpha 2$ and $\alpha 3$ helices in the apo form. This flexing of the loop in the apo form, drags the $\alpha 3$ helix that may translate into the reorganization of the dimer-dimer interface.
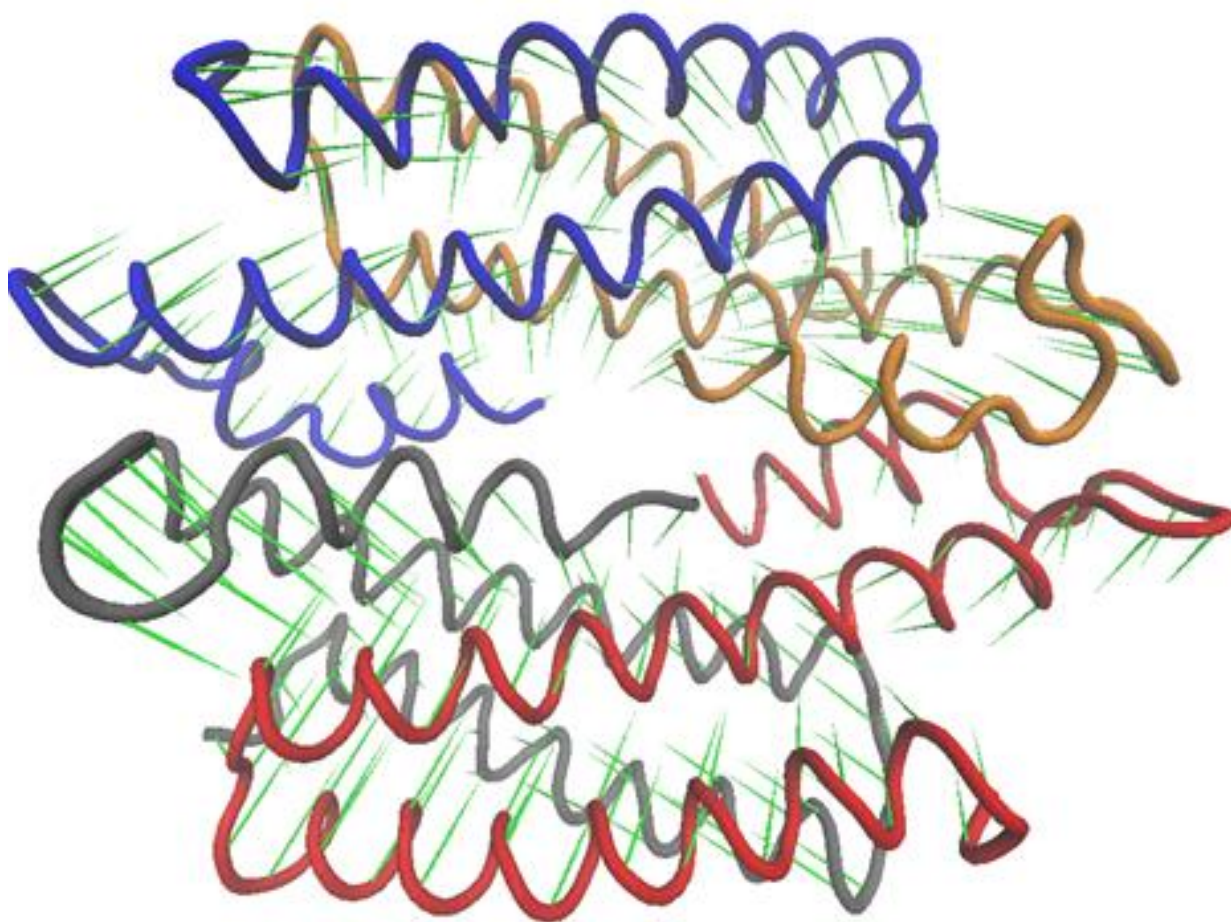
Figure 3.22. The vectors representing the overall motion of helices in Mtb-CsoR on loosing Cu(I) ions

In contrary, the Cu(I)-bound Mtb-CsoR reveals relatively smaller displacements in the α-helical structure and the displacement of the loop between the $\alpha 2$ and $\alpha 3$ helix was observed to be directed away from the copper binding site. The above identified 'hinge like' motion directed away from the copper binding site indicates the formation of the 'kink' in the $\alpha 2$ helix between His59 and Asn60 which is due to the enforcement of trigonal coordination geometry to bind

Cu(I). Hence, on binding to Cu(I) the Mtb-CsoR attains an open and rigid conformation that retains little affinity for DNA binding. This inflexibility of Cu(I)-bound CsoR may result from the H-bond interactions of Tyr35-His61-Glu81 that constrain the movement of the $\alpha1$ helix of one protomer relative to the $\alpha2$ and $\alpha3$ helices of the adjacent protomer in addition to the inflexibility acquired on binding to Cu(I). In the apo form, these inter and intra protomers H-bonds are not present, which leads to the increase in the degree of fluctuation and in turn effecting the flexibility of apo Mtb-CsoR. A similar model was put forward for the DNA promoter binding of apo Sl-CsoR.[112] The displacement vectors shown in Figure 3.22, represents the overall conformational transition of CsoR from the Cu(I) bound form to Apo form in Mtb-CsoR. The electrostatic potential maps of the Cu(I)-bound Mtb-CsoR shown in Figure 3.23, reveal that the positive potential is sparse and is distributed, while in the apo form continuous regions of positive potential are concentrated along each face of the tetramer exposing the arginine residues (R15 and R52 in Mtb-CsoR)[104,111] to initiate DNA binding.
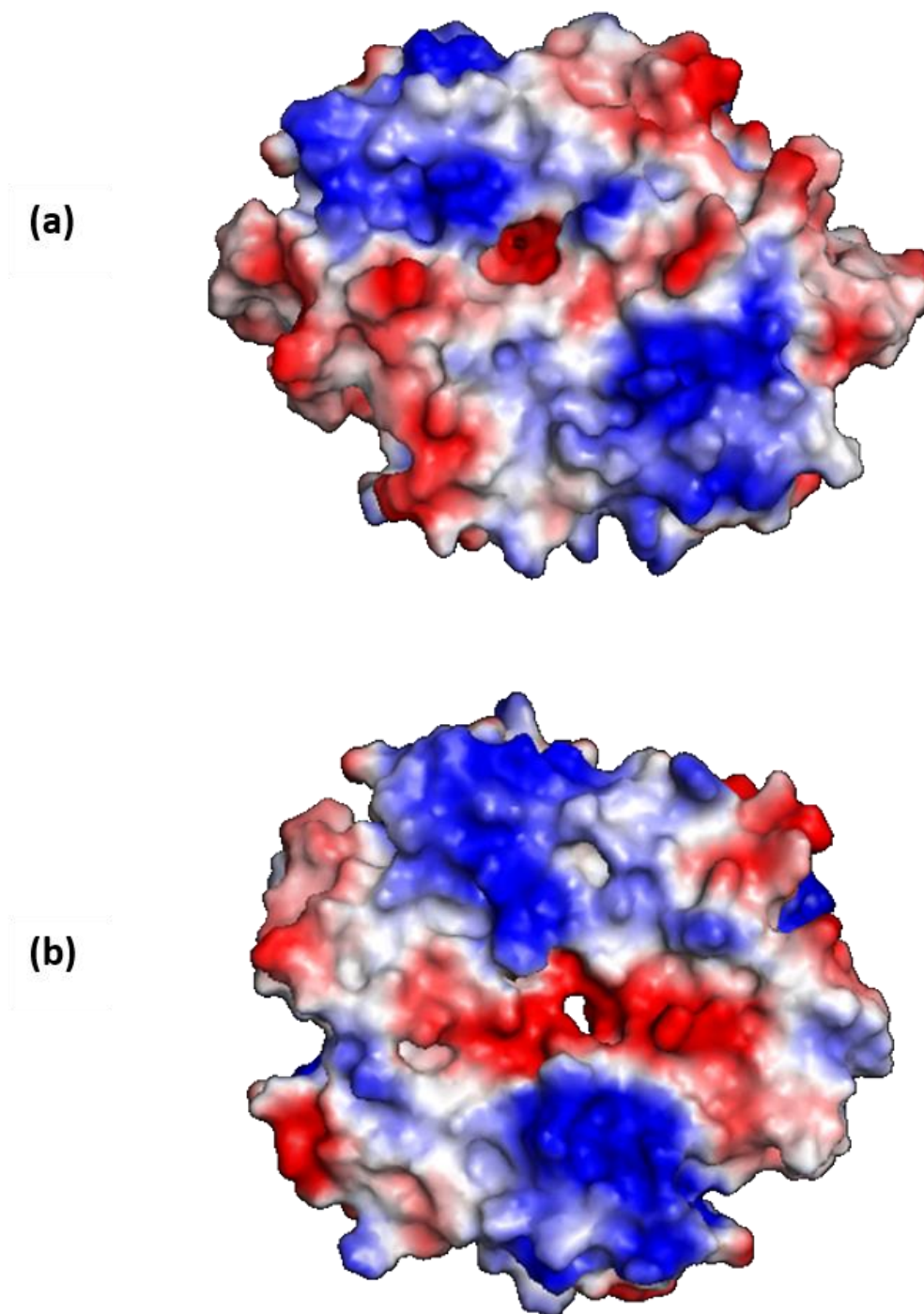
Figure 3.23. The electrostatic potential maps of (a) Cu(I)-bound Mtb-CsoR and (b) apo Mtb-CsoR

CONCLUSIONS

MD simulations along with essential dynamics analysis enabled us to describe the inherent global features of Cu(I)-mediated conformational switch in the tetrameric assembly of Mtb-CsoR. In this work, all-atom simulations in explicit solvent of the allosteric forms of the paradigm Cu(I) sensing protein of CsoR family in Mtb were presented to understand the role of quaternary structural change in allosteric communication. In addition, Cu(I) coordination geometry of CsoR was also predicted and elaborately explained by considering the effect of protonation state of cysteine ligands in allostery. The Cu(I) coordination geometry was calculated as a [2+1] binding mode, with the deprotonated cysteines as strong ligands and histidine coordination is considered to be very weak. The strength of Cu(I)—His61 bonding interaction was anticipated to play a key role in driving allosteric conformational switch rather than copper binding affinity of CsoR. Our simulations provide some key insights into the mechanism of allosteric regulation in Mtb-CsoR and are in close agreement with the extensive experimental findings for this system. We find that large quaternary structural conformational transitions of CsoR play an essential role in the ability of CsoR to function as an allosterically regulated transcriptional repressor. Our simulations of allosteric switching in Mtb-CsoR successfully capture this structural transition between closed to open conformations upon Cu(I) binding to CsoR. We were able to

provide insights on the disordered nature of α helices in the apo Mtb-CsoR that transform into more ordered and rigid structures in response to metal ion binding. We examined the mobility of residues in the two allosteric forms of the protein and observed an increase in the rigidity of the metal binding region and the α-helices in the Cu(I) bound allosteric form. Our simulations support the existence of a functionally important hydrogen-bonding pathway that connects the first coordination shell of the Cu(I) complex to the structural change involved in allosteric regulation. This pathway was stable only in Cu(I)-bound Mtb-CsoR but was destabilized or absent in the apo form. Our data suggest the possibility that the allosteric transition occurs by attaining a "rigid" structure when bound to Cu(I) and ultimately prevents it from adopting a more flexible high DNA binding affinity conformation.

In summary, our simulations provide new perspectives into allosteric switching in a metal sensor proteins. These studies take on added significance when the crystal structures of both the apo- and Cu(I) bound CsoR protein are difficult to crystalize and prevent atomic/molecular level comparison to understand the allosteric communication on binding to an effector (Cu(I)). Further studies have to be conducted, in order to illustrate the mechanism of allosteric regulation with DNA bound to the apo-CsoR.

REFERENCES

(1) Waldron, K. J.; Rutherford, J. C.; Ford, D.; Robinson, N. J., *Nature* **2009,** *460,* 823-830.

(2) Giedroc, D. P.; Arunkumar, A. I., *Dalton Trans.* **2007**, 3107-3120.

(3) Ma, Z.; Jacobsen, F. E.; Giedroc, D. P., *Chem. Rev. (Washington, DC, U. S.)* **2009,** *109,* 4644-4681.

(4) Vella, F., *Biochem. Educ.* **1995,** *23,* 115-115.

(5) Tottey, S.; Waldron, K. J.; Firbank, S. J.; Reale, B.; Bessant, C.; Sato, K.; Cheek, T. R.; Gray, J.; Banfield, M. J.; Dennison, C.; Robinson, N. J., *Nature* **2008,** *455,* 1138-U1117.

(6) Outten, C. E.; O'Halloran, T. V., *Science* **2001,** *292,* 2488-2492.

(7) Davidson, A. L.; Chen, J., *Annu. Rev. Biochem.* **2004,** *73,* 241-268.

(8) Waldron, K. J.; Robinson, N. J., *Nat. Rev. Microbiol.* **2009,** *7,* 25-35.

(9) Banci, L.; Bertini, I.; Cantini, F.; Felli, I. C.; Gonnelli, L.; Hadjiliadis, N.; Pierattelli, R.; Rosato, A.; Voulgaris, P., *Nat. Chem. Biol.* **2006,** *2,* 367-368.

(10) Fu, Y.; Tsui, H.-C. T.; Bruce, K. E.; Sham, L.-T.; Higgins, K. A.; Lisher, J. P.; Kazmierczak, K. M.; Maroney, M. J.; Dann III, C. E.; Winkler, M. E., *Nat. Chem. Biol.* **2013,** *9,* 177-183.

(11) Gold, B.; Deng, H.; Bryk, R.; Vargas, D.; Eliezer, D.; Roberts, J.; Jiang, X.; Nathan, C., *Nat. Chem. Biol.* **2008,** *4,* 609-616.

(12) Blindauer, C. A.; Harrison, M. D.; Robinson, A. K.; Parkinson, J. A.; Bowness, P. W.; Sadler, P. J.; Robinson, N. J., *Mol. Microbiol.* **2002,** *45*, 1421-1432.

(13) Andrews, S. C.; Robinson, A. K.; Rodríguez-Quiñones, F., *FEMS Microbiol. Rev.* **2003,** *27*, 215-237.

(14) Cobine, P. A.; Pierrel, F.; Winge, D. R., *Biochim. Biophys. Acta, Mol. Cell Res.* **2006,** *1763*, 759-772.

(15) Robinson, N. J.; Winge, D. R., *Annu. Rev. Biochem.* **2010,** *79*, 537.

(16) Puig, S.; Thiele, D. J., *Curr. Opin. Chem. Biol.* **2002,** *6*, 171-180.

(17) Solioz, M.; Stoyanov, J. V., *FEMS Microbiol. Rev.* **2003,** *27*, 183-195.

(18) Agranoff, D.; Krishna, S., *Front. Biosci.* **2004,** *9*, 2996-3006.

(19) Rensing, C.; Grass, G., *FEMS Microbiol. Rev.* **2003,** *27*, 197-213.

(20) MacPherson, I. S.; Murphy, M. E. P., *Cell. Mol. Life Sci.* **2007,** *64*, 2887-2899.

(21) Cooksey, D. A., *FEMS Microbiol. Rev.* **1994,** *14*, 381-386.

(22) Kosman, D. J., *JBIC, J. Biol. Inorg. Chem.* **2010,** *15*, 15-28.

(23) Cabiscol, E.; Tamarit, J.; Ros, J., *Int. Microbiol.* **2010,** *3*, 3-8.

(24) Borkow, G.; Gabbay, J., *Curr. Med. Chem.* **2005,** *12*, 2163-2175.

(25) Macomber, L.; Imlay, J. A., *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 8344-8349.

(26) Zheng, Z. Q.; White, C.; Lee, J.; Peterson, T. S.; Bush, A. I.; Sun, G. Y.; Weisman, G. A.; Petris, M. J., *J. Neurochem.* **2010,** *114*, 1630-1638.

(27) Huang, X. D.; Cuajungco, M. P.; Atwood, C. S.; Hartshorn, M. A.; Tyndall, J. D. A.; Hanson, G. R.; Stokes, K. C.; Leopold, M.; Multhaup, G.; Goldstein, L. E.; Scarpa, R. C.; Saunders, A. J.; Lim, J.; Moir, R. D.; Glabe, C.; Bowden, E. F.; Masters, C. L.; Fairlie, D. P.; Tanzi, R. E.; Bush, A. I., *J. Biol. Chem.* **1999,** *274*, 37111-37116.

(28) Atwood, C. S.; Huang, X. D.; Khatri, A.; Scarpa, R. C.; Kim, Y. S.; Moir, R. D.; Tanzi, R. E.; Roher, A. E.; Bush, A. I., *Cell. Mol. Biol.* **2000,** *46*, 777-783.

(29) Bush, A. I., *Alzheimer Dis. Assoc. Disord.* **2003,** *17*, 147-150.

(30) Curtain, C. C.; Ali, F. E.; Smith, D. G.; Bush, A. I.; Masters, C. L.; Barnham, K. J., *J. Biol. Chem.* **2003,** *278*, 2977-2982.

(31) Atwood, C. S.; Perry, G.; Zeng, H.; Kato, Y.; Jones, W. D.; Ling, K. Q.; Huang, X. D.; Moir, R. D.; Wang, D. D.; Sayre, L. M.; Smith, M. A.; Chen, S. G.; Bush, A. I., *Biochemistry* **2004,** *43*, 560-568.

(32) Butterfield, D. A.; Bush, A. I., *Neurobiol. Aging* **2004,** *25*, 563-568.

(33) Huang, X. D.; Moir, R. D.; Tanzi, R. E.; Bush, A. I.; Rogers, J. T., *Ann. N. Y. Acad. Sci.* **2004,** *1012*, 153-163.

(34) Syme, C. D.; Nadal, R. C.; Rigby, S. E. J.; Viles, J. H., *J. Biol. Chem.* **2004,** *279*, 18169-18177.

(35) Gomez-Balderas, R.; Raffa, D. F.; Rickard, G. A.; Brunelle, P.; Rauk, A., *J. Phys. Chem. A* **2005,** *109*, 5498-5508.

(36) Maynard, C. J.; Bush, A. I.; Masters, C. L.; Cappai, R.; Li, Q. X., *Int. J. Exp. Pathol.* **2005,** *86*, 147-159.

(37) Rozga, M.; Bal, W., *Chem. Res. Toxicol.* **2010,** *23*, 298-308.

(38) Dupont, C. L.; Grass, G.; Rensing, C., *Metallomics* **2011,** *3*, 1109-1118.

(39) Corbett, D.; Schuler, S.; Glenn, S.; Andrew, P. W.; Cavet, J. S.; Roberts, I. S., *Mol. Microbiol.* **2011,** *81*, 457-472.

(40) Rae, T.; Schmidt, P.; Pufahl, R.; Culotta, V.; O'halloran, T., *Science* **1999,** *284*, 805-808.

(41) Schwan, W. R.; Warrener, P.; Keunz, E.; Stover, C. K.; Folger, K. R., *Int. J. Med. Microbiol.* **2005,** *295*, 237-242.

(42) Wolschendorf, F.; Ackart, D.; Shrestha, T. B.; Hascall-Dove, L.; Nolan, S.; Lamichhane, G.; Wang, Y.; Bossmann, S. H.; Basaraba, R. J.; Niederweis, M., *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108*, 1621-1626.

(43) Klein, J. S.; Lewinson, O., *Metallomics* **2011,** *3*, 1098-1108.

(44) White, C.; Lee, J.; Kambe, T.; Fritsche, K.; Petris, M. J., *J. Biol. Chem.* **2009,** *284*, 33949-33956.

(45) Holwerda, R. A.; Wherland, S.; Gray, H. B., *Annu. Rev. Biophys. Bioeng.* **1976,** *5*, 363-396.

(46) Gray, H. B.; Winkler, J. R., *Annu. Rev. Biochem.* **1996,** *65*, 537-561.

(47) Klinman, J. P., *Chem. Rev. (Washington, DC, U. S.)* **1996,** *96*, 2541-2562.

(48) Lee, D. H.; O'Connor, T. R.; Pfeifer, G. P., *Nucleic Acids Res.* **2002,** *30*, 3566-3573.

(49) Lloyd, D. R.; Phillips, D. H., *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **1999,** *424*, 23-36.

(50) Ward, S. K.; Hoye, E. A.; Talaat, A. M., *J. Bacteriol.* **2008,** *190*, 2939-2946.

(51) Magnani, D.; Barré, O.; Gerber, S. D.; Solioz, M., *J. Bacteriol.* **2008,** *190*, 536-545.

(52) Chillappagari, S.; Seubert, A.; Trip, H.; Kuipers, O. P.; Marahiel, M. A.; Miethke, M., *J. Bacteriol.* **2010,** *192*, 2512-2524.

(53) Dong, J.; Atwood, C. S.; Anderson, V. E.; Siedlak, S. L.; Smith, M. A.; Perry, G.; Carey, P. R., *Biochemistry* **2003,** *42*, 2768-2773.

(54) Hashimoto, M.; Hsu, L. J.; Xia, Y.; Takeda, A.; Sisk, A.; Sundsmo, M.; Masliah, E., *Neuroreport* **1999,** *10*, 717-721.

(55) Hodgkinson, V.; Petris, M. J., *J. Biol. Chem.* **2012,** *287*, 13549-13555.

(56) Festa, R. A.; Thiele, D. J., *PLoS Pathog.* **2012,** *8*, e1002887.

(57) Solioz, M.; Abicht, H. K.; Mermod, M.; Mancini, S., *J. Biol. Inorg. Chem.* **2010,** *15*, 3-14.

(58) Peña, M. M.; Lee, J.; Thiele, D. J., *Neth. J. Nutr.* **1999,** *129*, 1251-1260.

(59) Busenlehner, L. S.; Pennella, M. A.; Giedroc, D. P., *FEMS Microbiol. Rev.* **2003,** *27*, 131-143.

(60) Osman, D.; Cavet, J. S., *Nat. Prod. Rep.* **2010,** *27*, 668-680.

(61) Pennella, M. A.; Giedroc, D. P., *BioMetals* **2005,** *18*, 413-428.

(62) Chen, P. R.; He, C., *Curr. Opin. Chem. Biol.* **2008,** *12*, 214-221.

(63) Helmann, J. D.; Soonsanga, S.; Gabriel, S., In *Molecular microbiology of heavy metals*, Springer: 2007; pp 37-71.

(64) Arunkumar, A. I.; Campanello, G. C.; Giedroc, D. P., *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 18177-18182.

(65) Brown, N. L.; Stoyanov, J. V.; Kidd, S. P.; Hobman, J. L., *FEMS Microbiol. Rev.* **2003,** *27*, 145-163.

(66) Osman, D.; Cavet, J. S., *Adv. Appl. Microbiol.* **2008,** *65*, 217-247.

(67) Chen, K.; Yuldasheva, S.; Penner-Hahn, J. E.; O'Halloran, T. V., *J. Am. Chem. Soc.* **2003,** *125*, 12088-12089.

(68) Changela, A.; Chen, K.; Xue, Y.; Holschen, J.; Outten, C. E.; O'Halloran, T. V.; Mondragón, A., *Science* **2003,** *301*, 1383-1387.

(69) Stoyanov, J. V.; Hobman, J. L.; Brown, N. L., *Mol. Microbiol.* **2001,** *39*, 502-512.

(70) Outten, F. W.; Outten, C. E.; Hale, J.; O'Halloran, T. V., *J. Biol. Chem.* **2000,** *275*, 31024-31029.

(71) Brocklehurst, K. R.; Hobman, J. L.; Lawley, B.; Blank, L.; Marshall, S. J.; Brown, N. L.; Morby, A. P., *Mol. Microbiol.* **1999,** *31*, 893-902.

(72) Hitomi, Y.; Outten, C. E.; O'Halloran, T. V., *J. Am. Chem. Soc.* **2001,** *123*, 8614-8615.

(73) Outten, C. E.; Outten, F. W.; O'Halloran, T. V., *J. Biol. Chem.* **1999,** *274,* 37517-37524.

(74) Singh, V. K.; Xiong, A.; Usgaard, T. R.; Chakrabarti, S.; Deora, R.; Misra, T. K.; Jayaswal, R. K., *Mol. Microbiol.* **1999,** *33,* 200-207.

(75) Osborn, A. M.; Bruce, K. D.; Strike, P.; Ritchie, D. A., *FEMS Microbiol. Rev.* **1997,** *19,* 239-262.

(76) Hobman, J. L., *Mol. Microbiol.* **2007,** *63,* 1275-1278.

(77) Brocklehurst, K.; Megit, S.; Morby, A., *Biochem. Biophys. Res. Commun.* **2003,** *308,* 234.

(78) Hynninen, A.; Touze, T.; Pitkänen, L.; Mengin-Lecreulx, D.; Virta, M., *Mol. Microbiol.* **2009,** *74,* 384-394.

(79) Chen, P.; Greenberg, B.; Taghavi, S.; Romano, C.; van der Lelie, D.; He, C., *Angew. Chem.* **2005,** *117,* 2775-2779.

(80) Rutherford, J. C.; Cavet, J. S.; Robinson, N. J., *J. Biol. Chem.* **1999,** *274,* 25827-25832.

(81) Gaudu, P.; Weiss, B., *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93,* 10094-10098.

(82) Storz, G.; Imlayt, J. A., *Curr. Opin. Microbiol.* **1999,** *2,* 188-194.

(83) Watanabe, S.; Kita, A.; Kobayashi, K.; Miki, K., *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 4121-4126.

(84) Zheng, M.; Storz, G., *Biochem. Pharmacol.* **2000,** *59,* 1-6.

(85) Pomposiello, P. J.; Demple, B., *Trends Biotechnol.* **2001,** *19,* 109-114.

(86) Godsey, M. H.; Heldwein, E. E. Z.; Brennan, R. G., *J. Biol. Chem.* **2002,** *277,* 40169-40172.

(87) Zheleznova, E. E.; Markham, P. N.; Neyfakh, A. A.; Brennan, R. G., *Cell* **1999,** *96,* 353-362.

(88) Godsey, M. H.; Baranova, N. N.; Neyfakh, A. A.; Brennan, R. G., *J. Biol. Chem.* **2001,** *276,* 47178-47184.

(89) Heldwein, E. E. Z.; Brennan, R. G., *Nature* **2001,** *409,* 378-382.

(90) Osman, D.; Waldron, K. J.; Denton, H.; Taylor, C. M.; Grant, A. J.; Mastroeni, P.; Robinson, N. J.; Cavet, J. S., *J. Biol. Chem.* **2010,** *285,* 25259-25268.

(91) Santo, C. E.; Taudte, N.; Nies, D. H.; Grass, G., *Appl. Environ. Microbiol.* **2008,** *74,* 977-986.

(92) Cobine, P.; Wickramasinghe, W.; Harrison, M.; Weber, T.; Solioz, M.; Dameron, C., *FEBS Lett.* **1999,** *445,* 27-30.

(93) Solioz, M.; Stoyanov, J. V., *FEMS Microbiol. Rev.* **2003,** *27,* 183-195.

(94) Strausak, D.; Solioz, M., *J. Biol. Chem.* **1997,** *272,* 8932-8936.

(95) Portmann, R.; Magnani, D.; Stoyanov, J. V.; Schmechel, A.; Multhaup, G.; Solioz, M., *J. Biol. Inorg. Chem.* **2004,** *9,* 396-402.

(96) Cobine, P. A.; Jones, C. E.; Dameron, C. T., *J. Inorg. Biochem.* **2002,** *88,* 192-196.

(97) Lu, Z. H.; Dameron, C. T.; Solioz, M., *BioMetals* **2003,** *16,* 137-143.

(98) Cobine, P. A.; George, G. N.; Jones, C. E.; Wickramasinghe, W. A.; Solioz, M.; Dameron, C. T., *Biochemistry* **2002,** *41*, 5822-5829.

(99) Portmann, R.; Poulsen, K. R.; Wimmer, R.; Solioz, M., *BioMetals* **2006,** *19*, 61-70.

(100) Wunderli-Ye, H.; Solioz, M., *FEMS Microbiol. Rev.* **2003,** *27*, 183-195.

(101) Hasman, H.; Kempf, I.; Chidaine, B.; Cariolet, R.; Ersbøll, A. K.; Houe, H.; Hansen, H. C. B.; Aarestrup, F. M., *Appl. Environ. Microbiol.* **2006,** *72*, 5784-5789.

(102) Vats, N.; Lee, S. F., *Microbiology* **2001,** *147*, 653-662.

(103) Mitrakul, K.; Loo, C.; Hughes, C.; Ganeshkumar, N., *Oral Microbiol. Immunol.* **2004,** *19*, 395-402.

(104) Liu, T.; Ramesh, A.; Ma, Z.; Ward, S. K.; Zhang, L. M.; George, G. N.; Talaat, A. M.; Sacchettini, J. C.; Giedroc, D. P., *Nat. Chem. Biol.* **2007,** *3*, 60-68.

(105) Smaldone, G. T.; Helmann, J. D., *Microbiology* **2007,** *153*, 4123-4128.

(106) Sakamoto, K.; Agari, Y.; Agari, K.; Kuramitsu, S.; Shinkai, A., *Microbiology* **2010,** *156*, 1993-2005.

(107) Dwarakanath, S.; Chaplin, A. K.; Hough, M. A.; Rigali, S.; Vijgenboom, E.; Worrall, J. A., *J. Biol. Chem.* **2012,** *287*, 17833-17847.

(108) Chang, F.-M. J.; Coyne, H. J.; Ramirez, C. A. C.; Fleischmann, P. V.; Fang, X.; Ma, Z.; Ma, D.; Helmann, J. D.; García-de los Santos, A.; Wang, Y.-X., *J. Biol. Chem.* **2014**, jbc. M114. 556704.

(109) Baker, J.; Sengupta, M.; Jayaswal, R. K.; Morrissey, J. A., *Environ. Microbiol.* **2011,** *13*, 2495-2507.

(110) Grossoehme, N.; Kehl-Fie, T. E.; Ma, Z.; Adams, K. W.; Cowart, D. M.; Scott, R. A.; Skaar, E. P.; Giedroc, D. P., *J. Biol. Chem.* **2011,** *286*, 13522-13531.

(111) Chang, F.-M. J.; Lauber, M. A.; Running, W. E.; Reilly, J. P.; Giedroc, D. P., *Anal. Chem.* **2011,** *83*, 9092-9099.

(112) Tan, B. G.; Vijgenboom, E.; Worrall, J. A., *Nucleic Acids Res.* **2013**, gkt902.

(113) Festa, R. A.; Jones, M. B.; Butler-Wu, S.; Sinsimer, D.; Gerads, R.; Bishai, W. R.; Peterson, S. N.; Darwin, K. H., *Mol. Microbiol.* **2011,** *79*, 133-148.

(114) Ward, S. K.; Abomoelak, B.; Hoye, E. A.; Steinberg, H.; Talaat, A. M., *Mol. Microbiol.* **2010,** *77*, 1096-1110.

(115) Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Su, X.; Borrelly, G.; Robinson, N., *J. Biol. Chem.* **2004,** *279*, 27502-27510.

(116) Nittis, T.; George, G. N.; Winge, D. R., *J. Biol. Chem.* **2001,** *276*, 42520-42526.

(117) Ma, Z.; Cowart, D. M.; Ward, B. P.; Arnold, R. J.; DiMarchi, R. D.; Zhang, L.; George, G. N.; Scott, R. A.; Giedroc, D. P., *J. Am. Chem. Soc.* **2009,** *131*, 18044-18045.

(118) Karplus, M.; McCammon, J. A., *Nat. Struct. Mol. Biol.* **2002,** *9,* 646-652.

(119) Elber, R., *Curr. Opin. Struct. Biol.* **2011,** *21*, 167-172.

(120) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E., *Curr. Opin. Struct. Biol.* **2009,** *19*, 120-127.

(121) Mouawad, L.; Perahia, D.; Robert, C. H.; Guilbert, C., *Biophys. J.* **2002,** *82,* 3224-3245.

(122) Bradley, M. J.; Chivers, P. T.; Baker, N. A., *J. Mol. Biol.* **2008,** *378,* 1155-1173.

(123) Chakravorty, D. K.; Wang, B.; Lee, C. W.; Giedroc, D. P.; Merz Jr, K. M., *J. Am. Chem. Soc.* **2011,** *134,* 3367-3376.

(124) Lee, C. W.; Chakravorty, D. K.; Chang, F.-M. J.; Reyes-Caballero, H.; Ye, Y.; Merz Jr, K. M.; Giedroc, D. P., *Biochemistry* **2012,** *51,* 2619-2629.

(125) Chakravorty, D. K.; Wang, B.; Ucisik, M. N.; Merz Jr, K. M., *J. Am. Chem. Soc.* **2011,** *133,* 19330-19333.

(126) Chakravorty, D. K.; Parker, T. M.; Guerra, A. J.; Sherrill, C. D.; Giedroc, D. P.; Merz Jr, K. M., *J. Am. Chem. Soc.* **2012,** *135,* 30-33.

(127) Sindhikara, D. J.; Roitberg, A. E.; Merz Jr, K. M., *Biochemistry* **2009,** *48,* 12024-12033.

(128) Chakravorty, D. K.; Wang, B.; Lee, C. W.; Guerra, A. J.; Giedroc, D. P.; Merz Jr, K. M., *J. Biomol. NMR* **2013,** *56,* 125-137.

(129) Guo, H.-B.; Johs, A.; Parks, J. M.; Olliff, L.; Miller, S. M.; Summers, A. O.; Liang, L.; Smith, J. C., *J. Mol. Biol.* **2010,** *398,* 555-568.

(130) Leitch, S.; Bradley, M. J.; Rowe, J. L.; Chivers, P. T.; Maroney, M. J., *J. Am. Chem. Soc.* **2007,** *129,* 5085-5095.

(131) Lee, S.; Arunkumar, A. I.; Chen, X.; Giedroc, D. P., *J. Am. Chem. Soc.* **2006,** *128*, 1937-1947.

(132) Grossoehme, N. E.; Giedroc, D. P., *J. Am. Chem. Soc.* **2009,** *131*, 17860-17870.

(133) Jaguar 7.0, Schrödinger, LLC, New York, NY, 2007.

(134) Becke, A. D., *Phys. Rev. A* **1988,** *38*, 3098-3100.

(135) Becke, A. D., *J. Chem. Phys.* **1993,** *98*, 5648-5652.

(136) Vosko, S. H.; Wilk, L.; Nusair, M., *Can. J. Phys.* **1980,** *58*, 1200-1211.

(137) Dunning, T. H., *J. Chem. Phys.* **1989,** *90*, 1007-1023.

(138) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B., *J. Phys. Chem.* **1996,** *100*, 11775-11788.

(139) Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A., *J. Phys. Chem. B* **1997,** *101*, 1190-1197.

(140) Friedrichs, M.; Zhou, R. H.; Edinger, S. R.; Friesner, R. A., *J. Phys. Chem. B* **1999,** *103*, 3057-3061.

(141) QSite, *Schrödinger , Inc., Portland, Oregon, 2001*.

(142) Philipp, D. M.; Friesner, R. A., *J. Comput. Chem.* **1999,** *20*, 1468-1494.

(143) Impact 7.2, *Schrödinger , Inc., Portland, Oregon, 2001*.

(144) Murphy, R. B.; Philipp, D. M.; Friesner, R. A., *J. Comput. Chem.* **2000,** *21*, 1442-1457.

(145) Murphy, R. B.; Philipp, D. M.; Friesner, R. A., *Chem. Phys. Lett.* **2000,** *321*, 113-120.

(146) Friesner, R. A.; Guallar, V., *Annu. Rev. Phys. Chem.* **2005,** *56*, 389-427.

(147) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., *J. Comput. Chem.* **2005,** *26*, 1701-1718.

(148) David Van Der Spoel, E. L., Berk Hess, Gerrit Groenhof, Alan E. Mark, Herman J. C. Berendsen,, *J. Comput. Chem.* **2005,** *26*, 1701-1718.

(149) Hess, B., *J. Chem. Theory Comput.* **2008,** *4*, 116-122.

(150) Focher, P.; Chiarotti, G. L.; Bernasconi, M.; Tosatti, E.; Parrinello, M., *Europhys. Lett.* **1994,** *26*, 345-351.

(151) Barnard, J. M.; Downs, G. M., *J. Chem. Inf. Comput. Sci.* **1992,** *32*, 644-649.

(152) Brown, R. D.; Martin, Y. C., *J. Chem. Inf. Comput. Sci.* **1996,** *36*, 572-584.

(153) Downs, G. M.; Barnard, J. M., *Rev. Comput. Chem.* **2002,** *18*, 1-40.

(154) Wild, D. J., *ACS 36th Central Regional Meeting* **June 2, 2004**.

(155) Cerda, B. A.; Wesdemiotis, C., *J. Am. Chem. Soc.* **1995,** *117*, 9734-9739.

(156) Dudev, T.; Lim, C., *J. Am. Chem. Soc.* **2002,** *124*, 6759-6766.

(157) Silva, J.; Williams, R., *The biological chemistry of the elements: the inorganic chemistry of life*. Oxford University Press Oxford: 2001.

(158) Hightower, K. E.; Huang, C.-c.; Casey, P. J.; Fierke, C. A., *Biochemistry* **1998,** *37*, 15555-15562.

(159) Zhou, L.; Singleton, C.; Le Brun, N., *Biochem. J* **2008,** *413*, 459-465.

(160) Nelson, J. W.; Creighton, T. E., *Biochemistry* **1994,** *33*, 5974-5983.

(161) Lewin, A.; Crow, A.; Oubrie, A.; Le Brun, N. E., *J. Biol. Chem.* **2006,** *281*, 35467-35477.

(162) Mössner, E.; Iwai, H.; Glockshuber, R., *FEBS Lett.* **2000,** *477*, 21-26.

(163) Kortemme, T.; Creighton, T., *J. Mol. Biol.* **1995,** *253*, 799.

(164) Bell, J. A.; Ho, K. L.; Farid, R., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012,** *68*, 935-952.

(165) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T., *Bioinformatics* **2006,** *22*, 195-201.

VIRUS-LIKE PARTICLES (VLPS)

Virus-like particles (VLPs) are assemblies of multiple proteins that mimic the organizational features of viruses including repetitive surface particulate structures such that they may elicit a pathogen-associated molecular pattern recognition response by the innate immune system.[1-7] Because they are devoid of genetic material VLPs provide a safer and more cost effective alternative to traditional vaccine development methods and several high priority viruses[1] have been targeted, namely the human papillomavirus (HPV)[8-10] (Gardasil[11,12] and Cervarix[13,14]), Chikungungya,[15,16] and hepatitis E[17,18] viruses. Despite these promising developments, the impact of VLPs on vaccine design at large remains limited, in part because many technical and fundamental challenges are currently unsolved.[1,7,19-25] For example, Gardasil is a very successful VLP-based vaccine that comprises of a mixture of VLPs derived from the L1 major capsid proteins of four different HPV types, namely 6, 11, 16 and 18. Gardasil induces specific antibody responses against these HPV types.[26-29] However, with more than 40 oncogenic HPV types identified to date,[30,31] it is clear that the capabilities of current vaccine design technology must be expanded to enable a broader spectrum of protection.

Traditional vaccine design approaches rely on cost-intensive, repetitive laboratory procedures and testing protocols. As a result, vaccine development is a

time-consuming and costly undertaking. One possible way of streamlining the discovery process is to utilize computer-aided design strategies to narrow the search and better understand the properties of various VLPs. By exploring the structural and dynamic features of a VLP *in silico*, and correlating them to experimentally observable efficacy data, the most salient molecular features of the VLP that may give rise to the immunogenicity can be identified. Exploiting these properties will enable a rational design approach that may significantly shorten vaccine development time. Due to their enormous size, probing the dynamic structure of a VLP under realistic conditions requires computationally intense molecular dynamics (MD) simulations. The advent of high-performance computing platforms[32-34] and sophisticated modeling algorithms,[35,36] made these daunting simulations within reach.[37-39]

Here, we extend the strategy of applying MD simulations for the *in silico* construction and molecular level analysis of candidates for VLP vaccines.[36,40] Specifically, HPV type 16 is examined, as (i) an all-atom X-ray structure with known epitope regions is readily available[41-43] and (ii) well-defined molecular level properties thought to be responsible for the immunogenicity were previously identified.[44-47] These VLPs can be assembled either from 12 or 72 pentamers of the L1 protein arranged in a T=1 or 7 icosahedron structures, respectively.[42,43,47] Whereas Gardasil is based on T=7 L1 icosahedral structure, Cervarix is an

alternative HPV vaccine that is based on a T=1 L1 structure. The assemblies are stabilized by strong hydrophobic interactions.[41,43] The C-terminus of the L1 protein consists of four helical regions h2, h3, h4 and h5 that are responsible for intra- and inter-molecular stabilization, as illustrated in Figure Error! No text of specified style in document..**2**a. The domains h2, h3 and h5 are responsible for L1 maintaining the structural integrity of the pentamer and h4 preserves the inter-pentamer connectivity, thereby determining the overall structure. The VLP surface has outwardly projecting protein loops containing epitopes that interact with the immune system to elicit production of type-specific antibodies.[48] Antibodies binding to the epitopes often render the associated virus/VLP inactive and neutralized. Neutralization assays of HPV-16 VLPs with human sera were used to identify five epitope bearing loop regions denoted BC (residue 49 to 70), DE (110 to 154), EF (170 to 189), FG (262 to 291) and HI (347 to 360).[41,43,49] These loops are thought to be more flexible than the rest of the L1 monomer and show notable conformational differences across HPV types.[48] Epitope deletion strongly affects the antibody-binding capacity of the VLPs,[44] and reduces their immunogenicity by a factor of at least 10-20 compared to wild-type VLPs.[50] There are several characteristics of VLP epitopes that may influence specific immune response; these properties include peptide sequence, loop conformation, and proximity/orientation relative to neighboring loops.[41,43]

The exact molecular-scale relationship between epitope structure and immune response is difficult to establish. For example, if individual epitope characteristics such as shape or sequence were the only factors relevant for immunogenicity, an L1 protein monomer or L1 pentamer could serve as a vaccine. However, the monomer is essentially not at all immunogenic and the pentamer is only weakly so,[51] in contrast to the highly immunogenic whole VLP. This difference in immunogenicity cannot be explained by the assumption that a VLP contains more epitopes than a pentamer or a monomer; as the increase in the monomer/pentamer dosage (and hence the number of epitopes) does not imply any associated increase in the HPV immunogenicity.[51] The epitope geometry among these structures is also very similar.[49] Furthermore, weakly organized VLP assemblies are found to be less immunogenic than more tightly packed ones.[52] One explanation of these observations comes from the fluctuation-immunogenicity hypothesis: To illicit proper immune response epitope fluctuations must be minimized,[40] such that the epitope structures are better defined and rigid over time - in tightly packed VLPs the epitope fluctuations are less pronounced than those in smaller systems. In other words, simply presenting an epitope to the immune system is not enough, as the structural fluctuations may render them unrecognizable. This hypothesis emphasizes the importance of understanding the dynamics of epitope structure and suggests that the immune response to the

smaller, simpler assemblies could be improved if the structural fluctuations can be reduced. Allosteric effects are widely recognized as central to controlling structure and dynamics of high-order protein assemblies,[53-68] and the tightly packed VLP constitutes a highly illustrative example of such long-range control. In particular, we found that epitope structure and function are strongly affected by allosteric interactions with the h4 helix of the L1 protein. We envisioned that this key interaction may be reproduced by tethering the L1 protein to a silica surface, thus mimicking the structure and conformational dynamics of the epitopes in the much more complex VLP assembly using a bioinorganic hybrid construct consisting of a silica nanoparticle and a much smaller piece of the virus. In this design the L1 protein is tethered to silica surfaces covalently. We examined how the epitope structure and dynamics are modulated by changing the curvature of the silica particle model and surface protein concentration. Interestingly, we found that the proposed bioinorganic mimics have epitope properties of the wild type VLP but do so in a way that (*i*) does not require construction of large T-numbered assemblies, (*ii*) facilitates easy synthesis, (*iii*) are genome-free and more stable than a pure macromolecular assembly. A similar construct was previously considered experimentally in the context of designing a vaccine against Porcine Circovirus Type 2,[69] and is already in trials.[70] Here, we extend such ideas to HPV 16 VLP

vaccines and quantitatively assess the role of the silica nanoparticle on the structure and importantly, the dynamics of the epitopes attached to it.

HYBRID DESIGN PRINCIPLES

The proposed hybrid designs involve L1 protein or pentamers attached to a silica surface using covalent –(CH2)3-NH2- tethers. Using silica in the hybrid design has the following advantages. First, surface properties are sufficiently adjustable to hold small molecules or larger nanostructures.[71] Consequently, silica-based nanoparticles are very useful in bioanalysis once conjugated with biological entities for analyte recognition and/or signal generation. Second, silica nanomaterials are effectively "transparent". They are unlikely to absorb light in the near-infrared, visible and ultraviolet regions or to interfere with magnetic fields, which allows the functional groups inside silica matrix to keep their original optical and magnetic properties; this facilitates use of lighter magnetic fields during medical procedures.[72-74] Third, silica matrices are highly nontoxic and biocompatible. Furthermore, well-established silica surface-chemistry facilitates the modification of silica-based nanohybrids.[75] Finally, to enable computer-aided design, dependable all-atom force fields are available for silica- water, ion and biomolecule simulations.[76,77] Also, silica surface has a negative potential, and therefore electrostatically binds to the inner surface of an L1 protein or pentamer.

In a virus, the protein capsid encloses negatively charged DNA or RNA. The latter electrostatically stabilize the assembly through interactions with the inner capsid surface. As used in applications such as nanotemplating[71,78] and in delivery agents,[79] within hybrid material design, the negatively surface charged silica plays the stabilizing role of the viral genome. However, like most other nanomaterials, including gold or magnetic nanoparticles, and quantum dots, silica particles are difficult to directly and uniformly suspend in aqueous solutions of different salinity. One option to overcome these limitations is to coat these materials with a more stable and physically adaptive material.[75,78] This way, stoichiometrically defined nanoparticles with correctly oriented immobilized biological elements are obtained.[71] For the present simulation, the silica surface is functionalized with aliphatic-amino tethers.

METHODS AND MATERIALS

MODEL PREPARATION

The models presented in this work include L1 monomer, pentamer and T=1 VLP in various arrangements and on different silica surface geometries connected via covalent tethers. All-atom, explicit solvent MD simulations were performed using NAMD[80,81] 2.7 for 10 ns on each of these assemblies. Atomic coordinates of the L1 monomer are obtained from the crystal structure (PDB code: 1DZL[42]). To
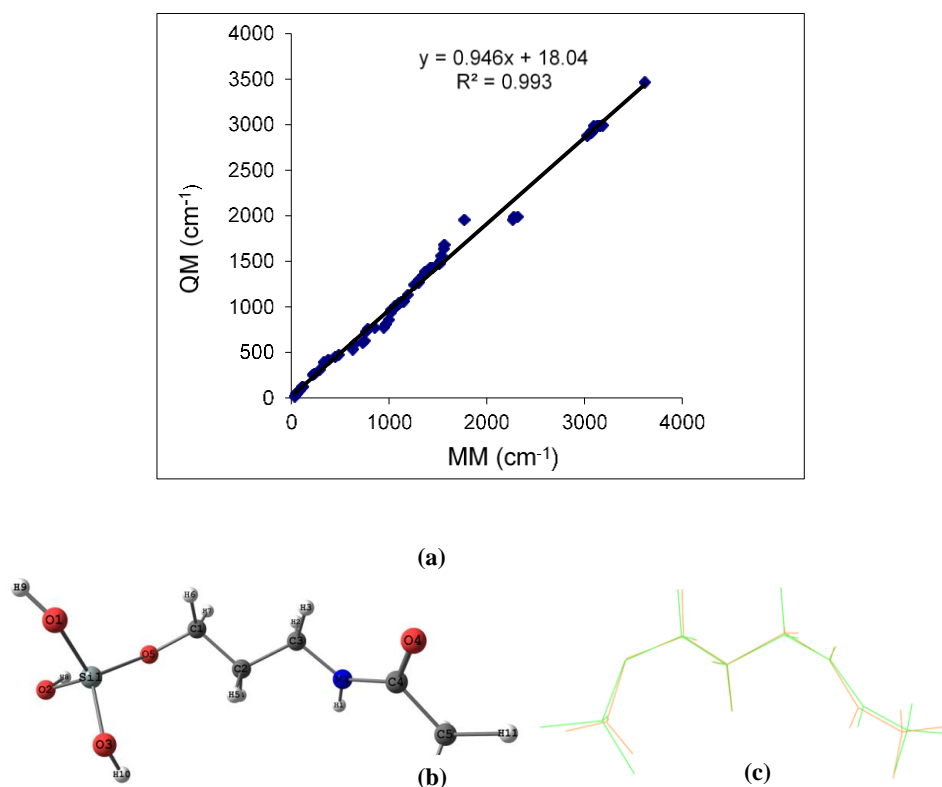
validate the allosteric effect of the h4 helix (residue 414 to 434), test simulations were performed on h4 helices that were truncated and artificially rigidified. The T=1 VLP is constructed from 60 copies of the monomer using icosahedral symmetry transformations (from VIPERDB).[82] The pentamer is extracted from this VLP to maintain structural continuity with the latter. All systems considered are immersed in a box of TIP3P waters,[83] extending at least 20 Å from the surface of the protein models. A 0.3M NaCl buffer solution was introduced to mimic the conditions under which the experiments we conducted, using the VMD[84] auto ionize feature. The resulting solvated system sizes range from ~10^5 to ~10^6 atoms. The 20 nm x 20 nm silica surface is generated by creating the appropriate images of the central unit cell using the IMAGE facility in CHARMM.[85,86] The silica model has a thickness of 5 nm incorporating 2 core layers of tetrahedral silicate moieties, and surface layers including hydrophilic (Si–OH) and hydrophobic (Si–H) groups on opposite ends. Curvature is introduced in the silica surface via steered MD simulations. The silica surface is functionalized with aliphatic-amino tethers that can covalently hold the L1 protein to the silica surface, and have simple electronic properties. CHARMM27[87] force fields are used for protein simulations. For simulating the silica surface, force fields developed by MacKerell were used.[76]

FORCE FIELD DEVELOPMENT

A systematic procedure as demonstrated in recent paper[76] was followed to develop force fields for the organic tether attached to silica surface. For simulating the silica surface, force fields parameters developed by MacKerell's group were used. The parameters of the silicon atom attached to the tether have been modified during the optimization of tether parameters. A small model of the tether attached to silica, represented in Figure Error! No text of specified style in document..**1**b was used to parameterize the force fields. Initially the structure of the tether is optimized using density functional calculations[88] at the B3LYP[89,90] level. Optimized structure of the tether is then used for force field parameterization with CHARMM27[87] force fields. Available CHARMM27 parameters[91,92] for similar organic compounds and silica were introduced as the guess parameters for the first optimization step. The structure was then calculated with the new MM force fields for validating the guess parameters. Charges on the atoms were derived from Merz-Kollman ESP charges[93] obtained from QM calculations with Gaussian 03[94] package. After several cycles of optimizing the parameters, reproducibility of target QM bond lengths, bond angles and vibrational spectra have been tested. Figure Error! No text of specified style in document..**1**c shows the comparison of structures obtained from QM and MM calculations, and Figure Error! No text of specified style in document..**1**a shows the correlation between normal modes for simulations with QM and MM

methods. In Table Error*! No text of specified style in document.*.1 bond lengths and

angles are compared between the QM and MM methods. The new force field

parameters are tabulated in Table Error*! No text of specified style in document.*.1.

(a)

(b)

(c)

Figure Error! No text of specified style in document..**1. (a)** Correlation of vibrational frequencies obtained from QM and MM normal mode calculations. **(b)** QM Model of the tether bound to silica. **(c)** Comparison of QM structure (Red) of tether to that of MM optimized structure (Green) with the new force fields.

Table Error! No text of specified style in document.**.1.** Comparison of bond lengths (Å) and bond angles (°) of QM and MM optimized structure using the newly developed MM force fields.

Bond Length	MM	QM
Si1-O2	1.67	1.67
O2-C1	1.43	1.43
C1-C2	1.51	1.52
C2-C3	1.51	1.53
C3-N1	1.44	1.45
N1-C4	1.37	1.37
C4-C5	1.50	1.51
Bond Angle	**MM**	**QM**
Si1-O2-C1	119.4	121.7
O2-C1-C2	111.0	111.8
C1-C2-C3	108.7	111.8
C2-C3-N1	111.3	112.8
C3-N1-C4	119.9	120.4
N1-C4-C5	114.5	115.5

MOLECULAR DYNAMICS SIMULATION DETAILS AND ASSOCIATED ANALYSIS[1]

All MD simulations were run on PowerPC 970MP processors of the BigRed supercomputer at Indiana University. To distinguish the behavior of the simulated constructs in terms of loop structures and their fluctuations, following molecular scale measures were considered. Here, we focus on the

---

[1] The MD simulation and analysis were performed in collaboration with Prof. Ortoleva's group, those include Dr. Abhishek Singharoy and Dr. Harshad Joshi and is part of published paper "Epitope Fluctuations in the Human Papillomavirus Are Under Dynamic Allosteric Control: A Computational Evaluation of a New Vaccine Design Strategy" Singharoy, A.; Polavarapu, A.; Joshi, H.; Baik, M.-H.; Ortoleva, P., *J. Am. Chem. Soc.* **2013,** *135*, 18458-18468.

analysis of the FG loop (Residue 262-291) as it is found to be most relevant for eliciting immune response; for certain comparison the EF and HI loops are also considered.[49] Similar results hold for the other epitopes.

Dihedral distribution for loops: The distribution of backbone loop dihedral angles is a good indicator for the conformational space explored by the epitope containing loops and provides a measure of the flexibility of the loop. In each case, 10,000 time points were extracted from MD simulations to construct the probability distribution of loop conformations. Since distribution from a randomly selected half of the ensemble is found to accurately reproduce the other half, analyzed structures are representative of the phenomenon of interest.

Power spectra: The power spectrum provides the distribution of atomic vibration intensity across a range of frequencies. Lower frequencies represent slower motions, while high frequencies represent faster modes. These measures were used to discriminate between the energies of different frequency motions of a given loop type as manifested in an assembly of a given size (i.e., from L1 monomer to T=1 VLP). Simulations were designed to assess potential differences in behavior of loops between the L1 protein assemblies and quantify them to serve as a basis of our computer-aided vaccine discovery

strategy. This study focused on loops known to contain critical epitopes, some of which are neutralizing.

Positional **variance**: The overall fluctuations of a particular loop from its average configuration were compared. Loop fluctuation is not easily quantified in X-ray or cryo-EM data. While a structure provides the most likely or average configuration, its fluctuation measures the importance of other configurations away from the average, but which may be functionally relevant within the framework of our immunogenicity hypothesis. Information about the dynamics obtained from MD provides advantages over the inherently averaged experimental data. Positional variance of the loop atoms was quantified as another measure of epitope fluctuation. Positional variance was computed by summing over the deviation of individual backbone atom position and dividing by the number of backbone atoms in the loop. This measure is slightly different from the usual root mean square fluctuation (RMSF). RMSF measures fluctuation from a fixed reference structure by aligning two structures, thus eliminating translational and rotational motions. In contrast, average loop positional variance calculated here contains contributions from overall displacements of the loops and their motions relative to the rotation/translation and internal motions of the assembly. The overall motions potentially affect epitope location and orientation within loops; according to our hypothesis, these overall fluctuations also affect immunogenicity and

binding properties of the monomer or larger assemblies. Thus, including the effect of overall and internal assembly motions on loop fluctuations provides a more complete measure of their potential relevance to immunogenicity.

Energy Analysis, Contact maps and Hydrogen Bonds: We also performed energy, contact map and hydrogen bond analyses. They were carried out on each trajectory using standard tools available in VMD.[84] In particular, energies were computed using the NAMD Energy plugin. Contact between two residues is considered if they were within a cutoff of 10 Å and neglected if the inter-residue distance exceeds the cutoff. Finally, hydrogen bonds were defined solely on the basis of geometric parameters (bond angle: 20°; bond length: 3.8Å) between donors and acceptors. Analysis of interplay between these properties for each of the simulated constructs yielded insights into interactions between the pentamer, tether and silica surface.

Correlation analysis: The pairwise correlations measuring the standard inter-residue three dimensional orientational coupling were computed using the covariance between positions of the ith C_γ atom at time t with respect to its initial value along the computed trajectory. A time averaged covariance matrix was built as:

$$\langle C_{ij} \rangle = \frac{1}{T} \int_1^T \Delta \hat{r}_i(t) \cdot \Delta \hat{r}_j(t) dt \qquad 4.1$$

where $\Delta \hat{r}_i$ is the unit vector of the displacement of the the ith C⊸⊚atom at time t, and T is the length of time over which we calculated the covariations. Positive correlations indicate correlated motion between the two residues, whereas negative values correspond to anti-correlation. Correlations with magnitude lesser than 0.5 were considered statistically insignificant and therefore neglected, where $\Delta \hat{r}_i$ is the unit vector of the displacement of the the ith atom at time t, and T is the length of time over which we calculated the covariations. Positive correlations indicate correlated motion between the two residues, whereas negative values correspond to anti-correlation. Correlations with magnitude lesser than 0.5, were considered statistically insignificant and therefore neglected.

RESULTS AND DISCUSSIONS

DYNAMIC PROPERTIES AND STRUCTURAL FLUCTUATION

Figure Error! No text of specified style in document.**.2**a shows a space filling model and a cartoon representation of the HPV-L1 protein monomer. The epitope bearing loop regions denoted BC (residue 49 to 70), DE (110 to 154), EF (170 to 189), FG (262 to 291) and HI (347 to 360)[41,43,49] are marked in red, yellow, green, blue and cyan, respectively. Five monomers assemble into a HPV-L1 pentamer, as illustrated in Figure Error! No text of specified style in document.**.2**b and twelve pentamers can finally be arranged in a T=1 icosahedral structure to afford the VLP,

shown in Figure Error! No text of specified style in document..**2**c. Of the five loops the FG loop is most important for eliciting immune response, followed by the EF and HI.[49] The BC and DE loops are thought to be less important. Therefore, we concentrate on the FG, EF and HI loops[95] in this work.[49] To examine the structural fluctuations and dynamic properties of the epitopes in this series of increasing complexity, we calculated the backbone dihedral angle distributions, positional variance and power spectra at each stages of assembly. Not surprisingly, the dihedral angles show the narrowest range in the VLP followed by the pentameric structure and display the widest range in the monomeric form, which is simply an indication of the increasing compactness as we move from the monomer to the VLP. Similarly, the positional variance of the loop fluctuations are largest in the monomer with amino acid positions showing variances as large as 12 Å^2 in EF and 7 Å^2 in FG and HI loops, as shown in Figure Error! No text of specified style in document..**3**a-c. In the pentamer these positional variances decrease significantly with the maximum not exceeding 3.5 Å^2 in all cases (Figure Error! No text of specified style in document..**3**d-f). Finally, the fluctuations in the amino acid positions become negligible in the T=1 VLP, as illustrated in Figure Error! No text of specified style in document..**3**g-i. The high-resolution profiles summarized in Figure Error! No text of specified style in document..**3** illustrate that the loops are rigidified substantially as the monomers are assembled into the pentamer. Although further stiffening occurs

163

when the pentamers are combined to give the final VLP, the change in positional

variance is much less pronounced, supporting the idea that the full-scale VLP may

not be necessary to mimic the epitope dynamics sufficiently. As expected, the

power spectra show a successive decrease in the low-frequency region upon

forming the pentamer and the VLP. At higher frequency, the power spectra for the

three constructs are similar because short timescale motions like bond oscillations

are similar for all three assemblies. These results establish a consistent trend of

decreasing structural flexibility of the epitopes as the assembly process progresses

from monomers to the full VLP – the dynamic behavior of an epitope is therefore

heavily influenced by the neighboring regions of same protein, and the presence

of other proteins. Interestingly, the mean structure of the epitopes remain

practically identical in all constructs, as shown in Figure Error! No text of specified

style in document..**3j**, where the mean structures of EF, FG and HI loops in the

monomeric, pentameric and VLP constructs are compared by overlaying them.

This conclusion is somewhat unexpected, as it is intuitively plausible that the

organization of the epitope carrying protein into the higher order constructs

should impact both the structure and dynamics of the protein. The fact that the

mean structure of the epitope in smaller constructs is identical to that of the whole

VLP is critically important, however, as this structural fidelity is a necessary, but

not sufficient, requirement for utilizing smaller virus fragments to elicit the

antiviral immune response. This result is particularly interesting within the context of the aforementioned experimental observation that the monomeric protein and the pentamer are essentially not immunogenic,[51] and suggests that this failure is due to the dynamic flexibility of the epitope in these constructs. To make the simpler constructs immunogenic, we must better understand the origin of the structural fluctuations and find means of inhibiting the positional variance in them.

One important factor for decreasing epitope fluctuations is structural confinement and inertial effects of the L1 assembly as its size increases.[51] A monomer has the lowest weight and the epitopes are least confined; therefore, associated fluctuations are maximum. In the pentamer and subsequently the VLP, both epitope confinement and assembly inertia increase notably, suggesting a significant decrease in structural fluctuations. For example, HI from one monomer is confined by FG from its counterclockwise neighbor (Figure Error! No text of specified style in document..**2**b). Similarly, FG interacts with loops DE and HI from its clockwise neighbor. This structural confinement from inter-epitope interactions is particularly effective for loop HI, as shown in the positional variance diagrams Figure Error! No text of specified style in document..**3**g → Figure Error! No text of specified style in document..**3**h → Figure Error! No text of specified style in document..**3**i. On the other hand, loop EF lacks inter-epitope contacts and is more solvent accessible

than FG or HI. Thus, it preserves some of its structural fluctuations in the pentamer, as illustrated in Figure Error! No text of specified style in document..**3**a → Figure Error! No text of specified style in document..**3**b, and becomes fully confined in the VLP (Figure Error! No text of specified style in document..**3**c). In other words, the neighboring proteins inhibit structural fluctuations of the epitope carrying loops simply by occupying the space that is needed to execute some of the structural fluctuations. Secondly, epitopes are subjected to identical increase in inertia as the system size increases, which will dampen structural fluctuations of all epitopes. Thus, confinement and increased inertia gives a plausible overall trend, but specific inter-epitope interactions are important and must be examined in greater detail to understand the non-uniform changes in fluctuations.
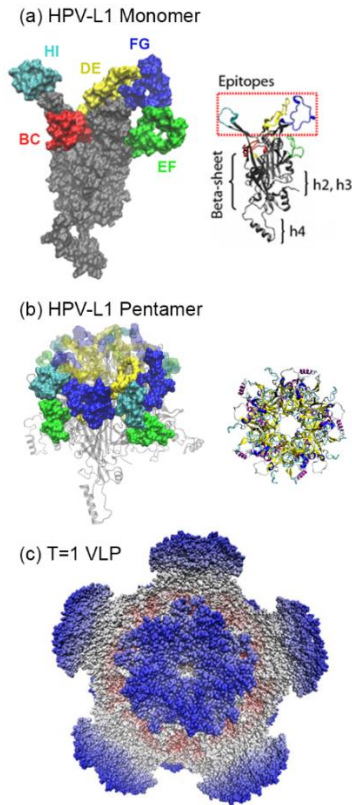
Figure Error! No text of specified style in document.**.2.** The structures the VLP components: (a) the isolated HPV-L1 monomer. The protein is depicted in gray with five different epitopes presented in red (BC), yellow (DE), green (EF), blue (FG), and cyan (HI). (b) HPV-L1 pentamer (c) HPV particle consisting of 12 pentamers arranged in T=1 icosahedral structure.

The tertiary structure of an L1 monomer is composed of β-sheets that carry the epitopes and α-helices on either end, as illustrated in Figure Error! No text of specified style in document.**.2**a. To better understand how these different components interact with each other and modulate the fluctuations, we constructed a covariance matrix from a 10 ns trajectory, shown in Figure Error! No text of specified style in document.**.4**a. In this diagram, strongly correlated structural distortions can be identified by high correlation coefficients that are marked by bright yellow and

bright red spots in the diagram. Correlation coefficients smaller than |0.5| can be considered statistically independent.
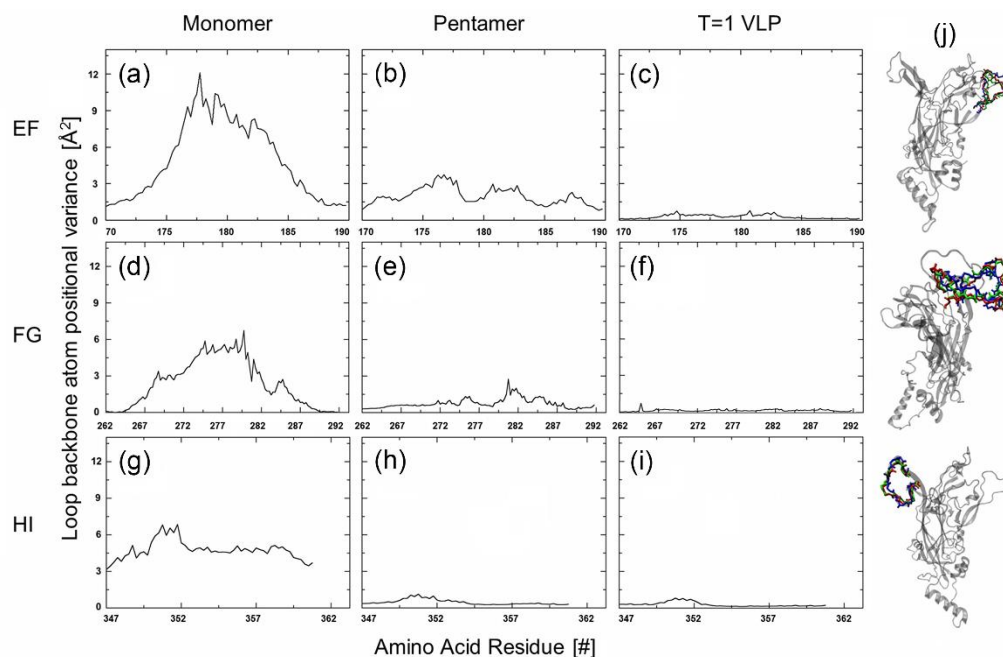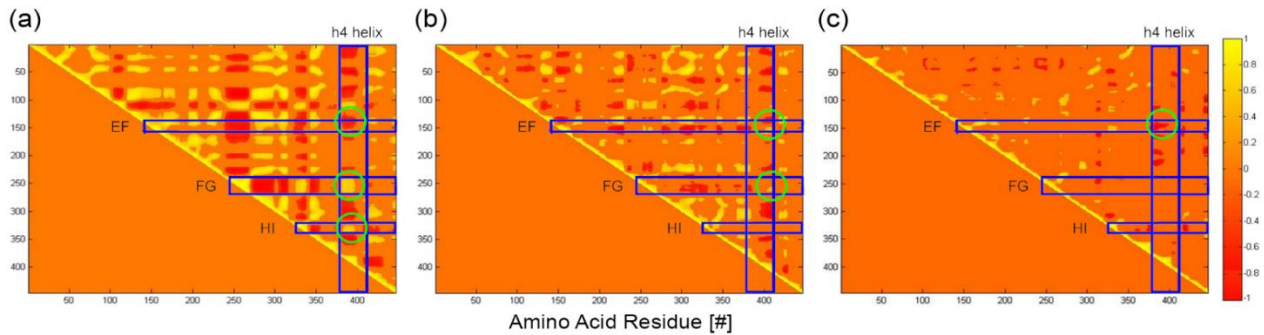

Figure Error! No text of specified style in document..**3.** Backbone atom positional variance for L1 monomer (left), pentamer (middle) and T=1 VLP (center). For each loop X-axis denotes the residue number and Y-axis denotes fluctuation in Å². Like the spread of the dihedral distributions, the positional variance decreases as assembly size increases. The fourth column shows mean orientation of the epitopes suggesting they change minimally between the monomer (blue), pentamer (red) and VLP (green).

Not surprisingly, all epitope carrying loops show significant correlation, as they are spatially close and structural distortions of one loop will cause steric clashes with another loop. Surprisingly, strong correlations are seen between the highly flexible h4 helix and all epitopes with correlation coefficients that are consistently larger than |0.6|, marked in green on Figure Error! No text of specified

style in document.**.4**a for the three loops of interest in this study. This strong communication is unexpected, as h4 is on the opposite side of the protein and the distance between the center of the h4 helix and the center of the FG-loop, for example, is 55.3 Å. It is not obvious how structural fluctuations in the helix will be mediated by the epitope region of the protein over such a long distance.

Figure Error! No text of specified style in document.**.4:** Covariance matrices from

10ns trajectory of (a) an isolated L1 protein, (b) an isolated pentamer and (c) T=1 VLP
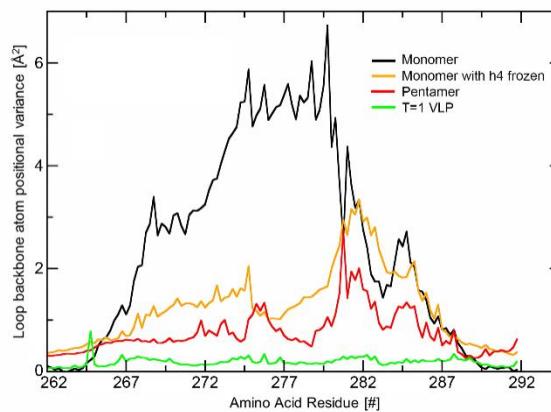
Figure Error! No text of specified style in document.**.5.** Positional variance for the FG loop the fluctuations of free L1 protein, L1 protein with the h4-helix artificially frozen, pentamer and T=1 VLP.

To further investigate this unusual coupling, we designed a computational experiment by manually changing the force fields associated with the amino acids in the h4 helix, as to artificially rigidify the helix and simply freeze the h4 helix structurally. If the fluctuations of the h4 helix and those of the epitope loops are truly coupled in an allosteric fashion, this artificially induced rigidity on the h4 helix should be translated to the loops and our simulations should show a reduced positional variance for the loops. The positional variance plot for the FG loop is given in Figure 4.4 and illustrates that loss of structural fluctuations in the h4 helix are indeed reported to the FG-loop. The positional variance of the residues 267-280 that are part of the FG-loop is greatly reduced from 3-7 Å^2 in the original protein (black line in Figure 4.4) to 1-3 Å^2 (orange line in Figure 4.4) – similar observation can also be made for the other loop regions. Thus, allosteric interactions between the h4 helix and epitopes play a major role in modulating epitope fluctuations. Whereas all epitopes show dynamic allosteric coupling to the h4 helix in the monomer, as indicated by the high correlation coefficients shown in Figure Error! No text of specified style in document..**4**,[96] the variation of this coupling with the assembly into higher order constructs is notably different from each other. In the monomer, the differences in fluctuation between epitopes mostly arise from intra-molecular confinement. For instance, since FG is more confined than HI its fluctuations are dampened: Almost all of the amino acids in the HI-loop can

170

engage in fluctuations giving rise to the positional variances of the backbone atoms of these residues to be higher than 4 Å^2 consistently, with the maximum variance being around 7 Å^2 (Figure Error! No text of specified style in document.**.3**g). Amino acids in the FG-loop display a distinctively different variance profile. Whereas the maximum variance is comparable at ~6 Å^2(Figure Error! No text of specified style in document**..3**d), there are many amino acids that show only small positional variances affording a profile that covers a range of 0-6 Å^2, unlike in the HI loop where the range was 3-7 Å^2. When multiple L1 proteins are brought together, the epitopes arrange on the surface and the h4 helices are either solvent-exposed in the pentamer, or they penetrate neighboring pentamers via hydrophobic interactions which stabilize the VLP core. In the higher order assemblies, additional confinement and strong inter-molecular interactions decrease epitope fluctuation compared to the free monomer, simplifying the covariance matrices significantly, as shown in Figure Error! No text of specified style in document**..4**b and Figure Error! No text of specified style in document**..4**c. Most interestingly, the FG and EF loops remain correlated to the h4 helix in the pentamer, as highlighted in green on Figure Error! No text of specified style in document**..3**b. The positional variance of the amino acids in the pentamer, shown in red in Figure Error! No text of specified style in document**..4**, show a striking similarity to the profile we obtained by simply freezing the h4 helix (orange in Figure Error! No text of specified style in document**..4**).

Both the magnitude and shape of the variance profiles are very comparable, suggesting that the main reason for diminished fluctuation of the FG-loop in the pentamer is the change in chemical environment of the h4 helix, which leads to an allosteric stiffening of the epitope containing loops. In the VLP, only the EF loop maintains a correlation to the h4 helix (Figure Error! No text of specified style in document..**4**c), but this finding must be interpreted with caution, as the loops in the VLP are practically rigid structures showing little to no structural variation. Whether or not these small positional variances are correlated to the h4 helix bears little chemical meaning. Interestingly, the allostery does not invoke any significant changes in the mean structures of the epitopes. Instead, the allosteric connection is expressed in variations of structural fluctuations, which in turn impact the biochemical behavior of the epitopes.

This finding is interesting from a general perspective about allosteric interactions in proteins. Traditionally,[53-55,57,62,97-99] allosteric interactions involve a small molecule binding event at one site of a protein that triggers a structural change at a different site, which is accompanied by reactivity changes at that site. Recently, this classical view of allosteric interactions has been extended and there is growing awareness of the fact that the modulation of chemical behavior does not have to be related to structural changes necessarily.[56,58,59,61-63,100-102] Changes in the entropy profile of molecular fragments can be just as powerful in modulating

the chemical behavior[56,103,104] giving rise to dynamic allosteric effects,[105] the most prominent manifestation of which is the change in structural fluctuation. Our study goes one step further in generalizing the concept of dynamic allostery in that we find that the dynamic coupling between two strongly correlated sites is general and substrate binding is only one of many possible ways of changing the chemical properties of a molecular fragment. In this case, the mean structure of the epitope containing loops remains practically invariant across the sequence of L1 constructs noted above, but the function and biochemical reactivity of these loops are nonetheless altered significantly, as the structural flexibility of the epitopes are modified. Our current work constitutes a rare demonstration of a strong dynamic allosteric effect across a long distance of 55 Å, where the allosteric signal transduction pathway contains standard peptidyl building block – in a previous theoretical study,[106] long distance dynamic allostery was envisioned to require structurally rigid components. Our findings suggest that these long-range correlations may be more common than thus far envisioned and that they do not require specially constructed, exotic entities to establish the dynamic allostery.

Rapid and large structural fluctuations of the epitopes are expected to reduce the antibody binding affinity,[40] and diminish the immune response. Depending upon micro-environmental conditions the population of the L1 assemblies will shift from one form to another; e.g., analytical ultracentrifugation

and light scattering analyses show that at a pH of ~6 and salinity >0.5M NaCl a T=1 or T=7 VLP structure is stable; at pH 8.2 they dissociate into L1 sub assemblies.[52] Larger assemblies exhibit lower levels of epitope fluctuation that facilitates stronger affinity for antibody binding.[40,51] However, some epitope fluctuation is required to allow antibody binding due to entropic enhancement of the epitope-antibody binding free energy.[107] Thus, there is an optimal level of epitope fluctuation intensity at which binding is favored, and beyond which the entropic barrier to binding becomes significant. Simulations suggest that the optimal fluctuation level is at <1Å/residue, which is observed in the VLP illustrated in Figures 2c, 2f and 2i. Finally, the allosteric scheme presented here is positively co-operative, i.e., the association of the h4 helices within the pentamer gives rise to an optimal epitope fluctuation level that, in turn, promotes antibody binding and subsequent immune response. A more quantitative and detailed study of the energetics involving the allosteric signal transduction pathway and epitope-antibody binding is required to decipher the exact mechanism of our suggested allostery. This is beyond the scope of the present study, and is partly addressed elsewhere.[108] In this work, we questioned how the insight discussed above can be exploited and we considered a strategy for silica-based hybrid nano-constructs where HPV substructures are attached to a silica nanoparticle in a way that will mimic the epitope properties of an entire T=1 VLP.
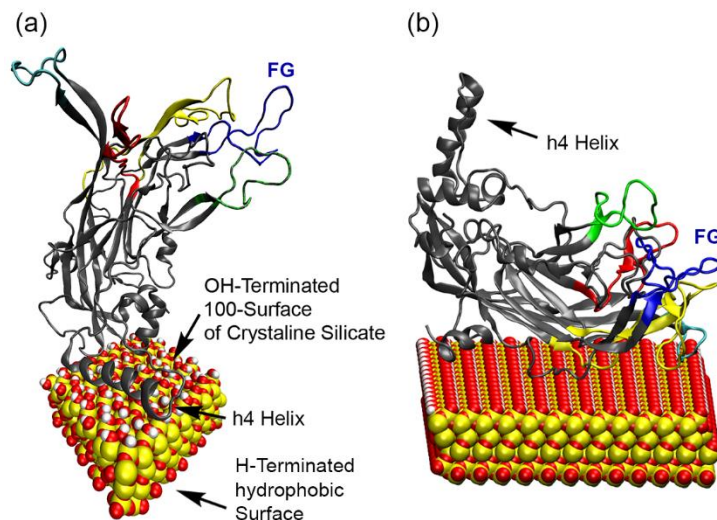
Figure Error! No text of specified style in document.**.6.** Proof-of-principle models of silica-bound L1-proteins where (a) the h4 helix and (b) the β-sheet portions are connected to the silica surface.

HYBRID NANOPARTICLES

Given the strong allosteric correlation between the epitope containing loops

and the h4 helix examined above, one potentially effective way of controlling the

dynamic properties of the loop regions is to modify the chemical environment of

the h4 helix. To allow fluctuation of L1 epitopes in the sub Å range, which

corresponds to the fluctuations seen in the T=1 VLP and which we assume to be

the optimal range of fluctuations for eliciting immune response, we first attached

the L1 protein to a model silica surface only using the innate electrostatic and

hydrogen-bond based attraction between the h4 helix and the silica surface. After

some experimentation, we chose to present the hydrophilic 100-surface of

crystalline silica, where each terminal surface oxygen of the silicate was

protonated, to the h4 helix of a single L1 protein. The model surface was 20 nm x 20 nm large and had a thickness of 5 nm incorporating 2 core layers of tetrahedral SiO_2 moieties. This design is shown in Figure Error! No text of specified style in document..**6**a, where only a small portion of the 400 nm^2 silica surface is shown for illustration. The surface layers were properly terminated with hydrophilic (Si–OH) and hydrophobic (Si–H) groups on each side of the silica sheet. Due to the finite size of the silica sheet chosen to make the simulations computationally feasible, buckling is observed in the silica-water simulations. To avoid such buckling of the surface, harmonic restraints are used on the hydrophobic silica layer that is furthest from the L1-binding surface. However, the surface in contact with the protein is kept unconstrained so that surface fluctuations can affect protein dynamics. This is a reasonable approximation that is commonly used in studies involving silica sheets.[109] The positional variance that results from letting the h4 helix interact with the silica surface is shown in Figure Error! No text of specified style in document..**7** in cyan color. As seen above, all epitope fluctuations decrease significantly and the variance of almost all amino acids in the FG-loop are <2 $Å^2$. The extent of fluctuation dampening and the shape of the positional variance profile is remarkably similar to what we found when we artificially inhibited the h4 fluctuations (Figure Error! No text of specified style in document..**5**, orange line), indicating that our basic design idea is plausible and that restriction

of fluctuations of the h4 helix is faithfully coupled to the FG loop leading to diminished loop fluctuations.

Whereas the model above is encouraging, it is not probable that the L1 monomer will self-assemble into the desired structure where only the h4 helix has contact to the silica surface, while the rest of the L1 protein maintains its overall structure. In fact, we found that a different structure, where the ⦾-sheet portion of the L1 protein gains contact with the surface, is energetically more preferable in good agreement with the notion that β-sheets can bind strongly to silica surfaces.[110] Among the many possible adducts, one is shown in Figure Error! No text of specified style in document..**6**b. The contact between the β-sheet and the silica surface is maximized and the h4 helix points away from the surface and, thus, none of the conceptual design motif initially envisioned is incorporated in this energetically more feasible structure. From a possible vaccine design perspective this protein-silica association is undesirable, as the structure of the protein as a whole and the mean structures of the epitope carrying loops are compromised significantly. As these latter structures are energetically favorable and intuitively plausible, it is safe to conclude that a self-assembly approach to preparing the desired L1-silica hybrid system is not promising to succeed. In addition, all of our monomer simulations suggest that the cooperative confinement that is present in the pentamer is needed to further reduce the positional variance to the desirable range of < 1 Å^2.
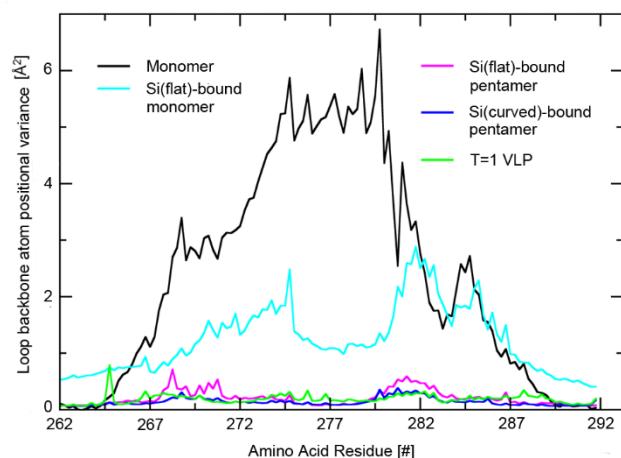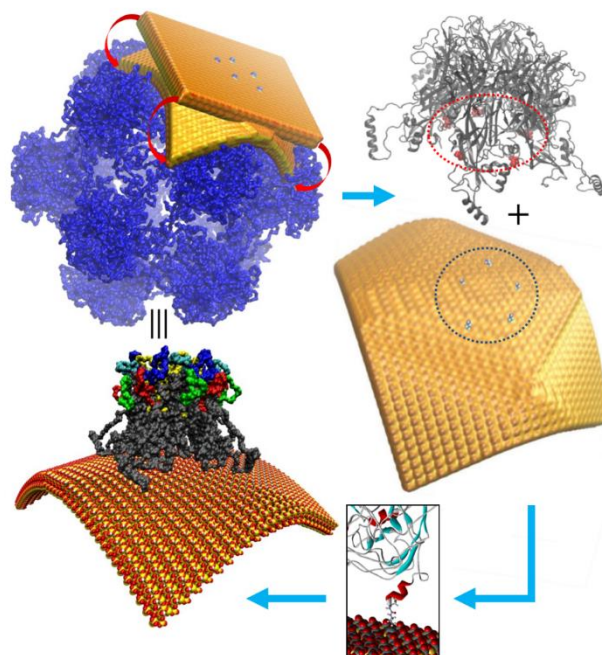
Figure Error! No text of specified style in document.**.7.** The positional variance profiles of the silica-bound protein constructs. The profiles of the monomer and T=1 VLP are shown as references.

A more complex design that may offer a solution to the problems identified above and that may provide a more realistic bioinorganic nanoparticle is to place the L1 pentamer on the silica surface and control the protein-silica contact points by using covalent linkages. Several strategies are readily available for attaching proteins to a solid support.[111] For example, the silica surface can be treated with amino silanes to afford a uniform surface layer of primary amines,[112,113] which we modeled using terminal $-Si-O-(CH_2)_3-NH_2$ moieties on the silica surface. The L1-proteins can be covalently linked to such an activated surface by a peptide coupling reaction to afford a permanent amide tether[112] containing a Silica$-Si-O-(CH_2)_3-NH-CO-L1$ motif, where the tether is anchored at the h4 helix. This design overcomes shortcomings of the monomer-silica structure as (*i*) interactions between epitopes that are located on different L1 proteins in the pentamer are

maintained, (*ii*) the covalent linkage provides control over the silica-protein contact point and suppresses structural degradation that may arise from undesirable interactions between the ⊚⊚sheet portion of L1 and the silica surface, and (*iii*) the h4 helix fluctuations will be inhibited both by the covalent linkage and non-covalent interactions of the h4 helix with the silica surface, as demonstrated in our small model above. Use of silica with aliphatic-amino tethers in the construction of hybrid VLPs has several attractive features: Silica (*i*) has a highly tunable surface chemistry which facilitates conjugation with biological entities,[114] (*ii*) is essentially transparent to light,[71] and (*iii*) is nontoxic and biocompatible.[115] However, like most other nanomaterials, including gold or magnetic nanoparticles and quantum dots, silica particles are difficult to directly and uniformly suspend in aqueous solutions with different salinities.[72,73] Additional details on practical advantages of the present design are discussed in section 3.4.2. The proposed design provides a simple model that includes fundamental features of a hybrid material based vaccine. However, for laboratory preparation of such material amorphous silica are preferred.[75] Studying amorphous silica computationally is difficult due to the large range of surface silicate group densities that can be obtained under various conditions, and the associated range of different interactions with proteins;[77] crystalline silica is more tractable for computer simulations. Our focus is on understanding the effects of factors such as atomic

forces, interaction energetics, friction imposed by neighboring loops, allostery, and inertia on the structure and function of L1 assemblies and, thus, we attached our protein models to a 100-surface of crystalline silica.[116] With these factors well understood, additional complexities in hybrid vaccine design arising from the use of amorphous silica can be addressed in future work; for the purpose of this study the use of crystalline silica is a reasonable compromise. One additional design component that we found to be important is that the silica surface must be curved – ideally, resembling the surface curvature found in the VLP. Our final model design protocol is illustrated in Figure 4.8: First, the flat silica surface is brought in proximity to the VLP. Then the edges of the silica are steered towards atoms on the VLP surface. Since overall stability, geometry and nearest neighbor interactions within silica is maintained using harmonic restraints, steered motion of atoms on the edges of the silica sheet gradually couples to those towards the interior. Consequently, the surface buckles forming a hydrophilic convex face with radius of curvature similar to that of the T=1 VLP. Next, the pentamer is tethered to the alkyl-amino groups across the surface where the silica curvature matches the inherent curvature of VLP-confined pentamer. If the silicon surface is left flat, the match of structural fluctuations between the silica-tethered pentamer and the VLP becomes less ideal, as shown in magenta color in Figure Error! No text of specified style in document..7.

Figure Error! No text of specified style in document..**8.** Hybrid design protocol: first, a flat sheet of silica generated using CHARMM is curved using the T=1 VLP as a template. Then, tethers are planted on the curved hydrophilic silica surface forming a -Si-O-$(CH_2)_3$-NH_2 linkage. Finally an L1 pentamer is attached to the surface via the tethers with the connectivity Silica-Si-O-$(CH_2)_3$-NH-CO-L1.

The L1 pentamer with positive potential surface was attached to the silica surface possessing negative electrostatic surface potential. The epitope fluctuation characteristics of this design are very close to those from the full VLP (Figure Error! No text of specified style in document..**7**) and similarly for helix-epitope correlations. However, a closer look into epitope properties reveals that they are different from those of the VLP. For example, we observe that the number of inter-epitope contacts and hydrogen bonds is lower than those in the T=1 VLP. As the pentamer is computationally extracted from a VLP it expands to a new equilibrium structure. When the expanded pentamer is attached to a flat silica surface, its

inherent curvature and associated inter-L1 contact (which mediate epitope properties) is lost. In particular, a large fraction of the inter-epitope hydrogen bonds connecting the FG and HI loops is lost. Thus, the fluctuations are also marginally higher in specific regions of the epitope. For example, THR residue 266 in the FG loops consistently loses hydrogen bonding interactions with the ASN residue 357 of the HI loop from its clockwise neighbor, thereby fluctuating more than in a complete VLP (Figures 6). Implications of residue-level conformations, such as those of THR 266, on the immunogenicity of associated constructs are discussed in the main text in the light of experimental findings. In conclusion, properties of the pentamer-flat silica design indicate that, in addition to confinement, inter-L1 interactions and helix-epitope allostery, one must consider surface curvature of the silica nanoparticle, as it plays a crucial role in determining epitope dynamics.

To quantify this finding, the effect of silica surface curvature is investigated. Silica surface curvature chosen to be that of the T=1 VLP so that the bound pentamer maintains its inherent curvature. This design reproduces all epitope properties of the VLP, as is reflected in the positional variance (Figure Error! No text of specified style in document..**7**), dihedral distribution, correlation plot, number of hydrogen bonds, contact analysis and epitope energetics.
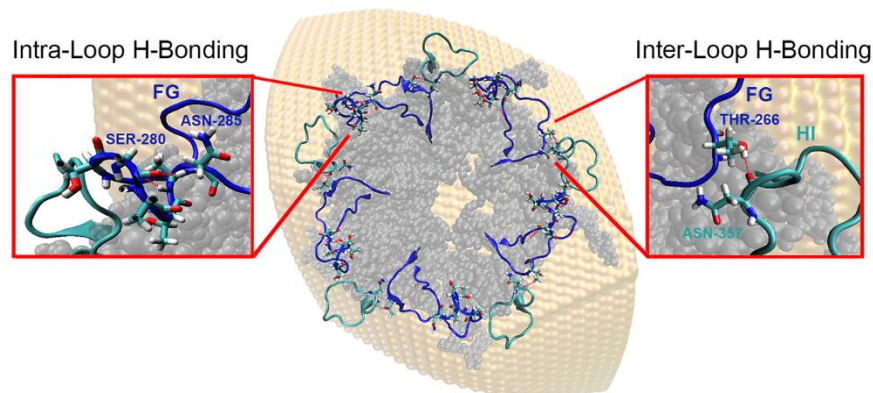
Figure Error! No text of specified style in document.**.9.** Organization of the surface epitopes FG and HI using inter- and intra-loop hydrogen-bonds.

The epitope fluctuation characteristics of the pentamer covalently attached to the curved silica surface is remarkably close to that of the full VLP, as shown in Figure 4.7 in blue compared to the T=1 VLP fluctuations shown in green. Engineering the structural and dynamic properties of the epitope containing loops is a necessary condition for eliciting the desired immune response, but there are additional conditions that must be met to faithfully reproduce the immunogenic properties of the wild type virus and/or the full-scale VLP. Several atomic scale features have previously been identified to be important for eliciting a proper immune response. For example, mouse monoclonal antibody H16.V5 binds to a major part of the FG loop and neutralizes HPV16;[117] mutation of ASN-285 leads to the failure of this antibody binding, suggesting that ASN-285 is directly involved in the H16.V5 binding.[117] Similarly, SER-282 appears to be important for the epitope to bind another antibody, H16.E70.[117] These experimental observations

emphasize that immune response to the L1 epitopes strongly depends on the details of the epitope structure. Fine scale structural details must therefore be carefully accounted for within a design strategy for assembling an artificial hybrid vaccine. Hydrogen-bonds play a particularly important role and we have carefully monitored the hydrogen-bond network that organizes the orientation of the epitopes to each other. Two hydrogen-bonds were especially interesting: The THR-266 residue of the FG-loop forms a strong inter-loop hydrogen-bond with an ASN-357 residue on a neighboring HI loop in the pentamer, as shown in Figure Error! No text of specified style in document..**9**. This is an important structural feature also found in the VLP that must be preserved for proper immune response - it is impossible to form this inter-loop hydrogen-bond in a monomer, which is one of the reasons why the L1-monomer is unlikely to be useful as a vaccine. A second hydrogen-bond of importance involves the residues 280-285, which enforce a relatively consistent conformation in that part of the FG-loop by engaging in a network of mutually exchangeable hydrogen-bonds with each other. In Figure 4.9 we depict one such H-bonded snapshot structure, where SER-280 and ASN-285 formed a hydrogen-bond. As a consequence of these intra-loop hydrogen-bonds, the FG-loop adopts a helix-like secondary structure, as illustrated in the detailed view of this region in Figure Error! No text of specified style in document..**9**. As this portion of the FG-loop is most solvent accessible, the helix-like folding provides

an energetic advantage. This structural detail is present in the VLP and we propose that it is an important recognition motif that must be preserved in a vaccine to elicit a proper immune response.[117] Note that the positional variance profiles (Figure Error! No text of specified style in document..**7**) have consistently indicated a very large change in positional variance around residue 280 when moving from a monomer to higher order assemblies - this hydrogen-bond network is responsible for the significant change in the fluctuation profile, as it is not present in the monomer and the residues around 280 have a much higher degree of structural freedom. These delicate structural details support and amplify the dynamic allosteric effects of the h4 helix discussed above to ultimately generate a structural fluctuation profile of the epitopes in the silica-mounted pentamer that is essentially identical to what is seen in the much larger T=1 VLP of HPV 16.

CONCLUSIONS

The preparation of well-defined small subassemblies, such as the pentamer, from monomeric L1-proteins is much easier[118] than synthesizing the larger assemblies like the T=1 or T=7 icosahedra. Thus, the vaccine development process could be expedited significantly if the smaller particles can be used to elicit a immune response instead of having to prepare the much more elaborate VLP structures.[119] The difficulty of assembling the higher order constructs lies in the

strong contribution that entropy makes to the energetics of the VLP; enthalpically, the interactions between the monomers in the pentamer assembly is much stronger than those between pentamers in the VLP. Thus, the translational entropy penalty associated with the formation of the pentamer from monomers is partially compensated for, allowing the free energy of assembly to be much more favorable for the pentamer than for the VLP.[120] Therefore, the pentamers are ideal targets in a rational vaccine design strategy. Mounting the pentamer on a silica nano-particle is enthalpically highly favorable, i.e. the enthalpy of tether mediated pentamer binding to silica is much higher than the interaction between pentamers to afford the VLP. By providing a minimally invasive chemical modification to install the chemical anchor for coupling the pentamer to the surface-modified silica nanoparticle, we introduce an additional driving force and engineer precise control over the assembly. In comparison, the preparation of the VLP from the pentamers is much more demanding, as the entropic penalties originating from the required precise relative orientation of the pentamers during the final assembly to the VLPs must be overcome.

Computer simulations have become a standard tool of biomedical research over the last few decades, but they are mainly used to rationalize and confirm experimental observations.[116-120] Given the level of sophistication and degree of realism in today's computer models, truly predictive computer modeling is not

only possible, but bears significant advantages over purely experimental approaches, as we demonstrate in this work by deriving a logical and rational vaccine design strategy. To the best of our knowledge, this is the first computational study that used all-atom structures of hybrid silica-protein nano-constructs to provide a novel nanoscale perspective on a long-standing challenge of VLP-based vaccine design. We discovered an intriguing dynamic allosteric coupling between the h4 helix and the epitope containing loops and devised an effective exploitation strategy for rigidifying the epitopes to reproduce the structural and dynamic properties of these epitopes in the VLP using the silica surface mounted pentamer. A silica mounted VLP as a basis of a vaccine against porcine virus[69] was reported previously and served as an inspiration of this work. Silica nanoparticles provide a potentially revolutionary opportunity for developing vaccines and we demonstrate how they can be utilized rationally. Our work highlights an intriguing connection between structure, dynamics and function, while explicitly outlining a strategy for exploiting dynamic allostery which is a relatively new concept that remains poorly understood. Our design has advantages over traditional T=1 VLPs in that they are (a) expected to be thermally stable (b) easy to prepare and (c) genome-free. Furthermore, silica nanoparticles are already FDA approved,[120] but remain an underutilized resource in vaccine development. In ongoing work, these theoretically identified vaccine candidates

will be prepared and characterized in collaboration with experimentalists to test our hypothesis - whereas adjustments and improvements to our initial strategies discussed above are expected, the foundation of the control mechanisms that we identified and explained above is generally valid. In addition to the obvious benefit of having identified a specific vaccine development strategy, the dynamic allosteric control mechanism outlined in this work appears to be generally applicable and more common than previously thought and deserves further investigation.

REFERENCES

(1) Grgacic, E. V. L.; Anderson, D. A., *Methods* **2006**, *40*, 60-65.

(2) Georgens, C.; Weyermann, J.; Zimmer, A., *Curr. Pharm. Biotechnol.* **2005**, *6*, 49-55.

(3) Cubas, R.; Zhang, S.; Kwon, S.; Sevick-Muraca, E. M.; Li, M.; Chen, C.; Yao, Q., *J. Immunother.* **2009**, *32*, 118-128.

(4) Jennings, G. T.; Bachmann, M. F., *Biol. Chem.* **2008**, *389*, 521-536.

(5) Ault, K. A.; Giuliano, A. R.; Edwards, R. P.; Tamms, G.; Kim, L. L.; Smith, J. F.; Jansen, K. U.; Allende, M.; Taddeo, F. J.; Skulsky, D.; Barr, E., *Vaccine* **2004**, *22*, 3004-3007.

(6) Han, J. E.; Wui, S. R.; Park, S. A.; Lee, N. G.; Kim, K. S.; Cho, Y. J.; Kim, H. J., *Vaccine* **2012**, *30*, 4127-4134.

(7) Buonaguro, L.; Tagliamonte, M.; Tornesello, M. L.; Buonaguro, F. M., *Expert Rev. Vaccines* **2011**, *10*, 1569-1583.

(8) Stanley, M.; Pinto, L. A.; Trimble, C., *Vaccine* **2012**, *30 Suppl 5*, F83-87.

(9) Foldvari, M.; Kumar, P., *Ther. Deliv.* **2012**, *3*, 1005-1017.

(10) Jemon, K.; Young, V.; Wilson, M.; McKee, S.; Ward, V.; Baird, M.; Young, S.; Hibma, M., *PLoS One* **2013**, *8*, e66866.

(11) Brown, D. R.; Garland, S.; Ferris, D. G.; Joura, E.; Steben, M.; James, M.; Radley, D.; Vuocolo, S.; Garner, E. I. O.; Haupt, R. M.; Bryan, J. T., *Hum. Vaccines* **2011**, *7*, 230-238.

(12) Smith, J. F.; Brownlow, M.; Brown, M.; Kowalski, R.; Esser, M. T.; Wruiz, W.; Barr, E.; Brown, D. R.; Brian, J. T., *Hum. Vaccines* **2007**, *3*, 109-115.

(13) Chaplin, S.; Nicholas, N., *Prescriber* **2008**, *19*, 46-50.

(14) Einstein, M. H.; Baron, M.; Levin, M. J.; Chatterjee, A.; Edwards, R. P.; Zepp, F.; Carletti, I.; Dessy, F. J.; Trofa, A. F.; Schuind, A.; Dubin, G., *Hum. Vaccines* **2009**, *5*, 705-719.

(15) Metz, S. W; Gardner, J.; Geertsema, C.; Le, T. T.; Goh, L.; Vlak, J. M.; Suhrbier, A.; Gijlman, G. P., *PLoS Negl. Trop. Dis.* **2013**, *7*, e2124.

(16) Akahata, W.; Yang, Z.-Y.; Andersen, H.; Sun, S.; Holdaway, H. A.; Kong, W.-P.; Lewis, M. G.; Higgs, S.; Rossmann, M. G.; Rao, S.; Nabel, G. J., *Nat. Med.* **2010**, *16*, 334-338.

(17) Guu, T. S. Y.; Liu, Z.; Ye, Q.; Mata, D. A.; Li, K.; Yin, C.; Zhang, J.; Tao, Y. J., *Proc. Nat. Acad. Sci. USA* **2009**, *106*, 12992-12997.

(18) Li, T.-C.; Suzaki, Y.; Ami, Y.; Dhole, T. N.; Miyamura, T.; Takeda, N., *Vaccine* **2004**, *22*, 370-377.

(19) Jennings, G. T.; Bachmann, M. F., *Annu. Rev. Pharmacol. Toxicol.* **2009**, *49*, 303-326.

(20) Roldão A.; Mellado M. C. M.; Castilho L. R.; Carrondo M. J. T.; Alves, P. M., *Expert Rev. Vaccines* **2010**, *9*, 1149-1176.

(21) Da Silva, D. M.; Schiller, J. T.; Kast, W. M., *Vaccine* **2003**, *21*, 3219-3227.

(22) Pushko, P.; Pumpens, P.; Grens, E., *Intervirology* **2013**, *56*, 141-165.

(23) Lehtinen, M.; Dillner, J., *Nat. Rev. Clin. Oncol.* **2013**, *10*, 400-410.

(24) Loucq, C., *Clin Exp Vaccine Res.* **2013**, *2*, 4-7.

(25) Bárcena, J.; Blanco, E., *Subcell. Biochem.* **2013**, *68*, 631-665.

(26) Brown, D. R.; Kjaer, S. K.; Sigurdsson, K.; Iverson, O.-E.; Hernandez-Avila, M.; Wheeler, C. M.; Perez, G.; Koutsky, L. A.; Tay, E. H.; Garcia, P.; Ault, K. A.; Garland, S. M.; Leodolter, S.; Olsson, S.-E.; Tang, G. W. K.; Ferris, D. G.; Paavonen, J.; Steben, M.; Bosch, F. X.; Dillner, J.; Joura, E. A.; Kurman, R. J.; Majewski, S.; Muñoz, N.; Myers, E. R.; Villa, L. L.; Taddeo, F. J.; Roberts, C.; Tadesse, A.; Bryan, J.; Lupinacci, L. C.; Giacoletti, K. E. D.; Sings, H. L.; James, M.; Hesley, T. M.; Barr, E., *J. Infect. Dis.* **2009**, *199*, 926-935.

(27) Dillner, J.; Kjaer, S. K.; Wheeler, C. M.; Sigurdsson, K.; Iversen, O. E.; Hernandez-Avila, M.; Perez, G.; Brown, D. R.; Koutsky, L. A.; Tay, E. H.; Garcia, P.; Ault, K. A.; Garland, S. M.; Leodolter, S.; Olsson, S. E.; Tang, G. W.; Ferris, D. G.; Paavonen, J.; Lehtinen, M.; Steben, M.; Bosch, F. X.; Joura, E. A.; Majewski, S.; Munoz, N.; Myers, E. R.; Villa, L. L.; Taddeo, F. J.; Roberts, C.; Tadesse, A.; Bryan,

J. T.; Maansson, R.; Lu, S.; Vuocolo, S.; Hesley, T. M.; Barr, E.; Haupt, R., *BMJ* **2010**, *341*, c3493.

(28) The FUTURE II Study Group, *New Eng. J. Med.* **2007**, *356*.

(29) Wheeler, C. M.; al., e., *J. Infect. Dis.* **2009**, *199*, 936-944.

(30) Tota, J. E.; Ramanakumar, A. V.; Jiang, M.; Dillner, J.; Walter, S. D.; Kaufman, J. S.; Coutlée, F.; Villa, L. L.; Franco, E. L., *Am. J. Epidemiol.* **2013**, electronic publication ahead of print, May/09/2013

(31) de Villiers, E.-M.; Fauquet, C.; Broker, T. R.; Bernard, H.-U.; zur Hausen, H., *Virology* **2004**, *324*, 17-27.

(32) Schulz, R.; Lindner, B.; Petridis, L.; Smith, J. C., *J. Chem. Theor. Comp.* **2009**, *5*, 2798-2808.

(33) Sanbonmatsu, K. Y.; Tung, C. S., *J. Struct. Biol.* **2007**, *157*, 470-480.

(34) Nakano, A., Kalia, R.K., Nomura, K. , Sharma, A., Vashista, P., Shimojo, F., van Duin, A.C.T., Goddard, W.A., Biswas, R. Srivastava, D. and Yang, L.H. *Int. J. High Perf. Comp. Applic.* **2008**, *22*, 113-128.

(35) Saunders, M. G.; Voth, G. A., *Curr. Opin. Struct. Biol.* **2012**, *22*, 144-150.

(36) Joshi, H.; Singharoy, A. B.; Sereda, Y. V.; Cheluvaraja, S. C.; Ortoleva, P. J., *Progr. Biophys. Mol. Bio.* **2011**, *107*, 200-217.

(37) Miao, Y.; Johnson, J. E.; Ortoleva, P. J., *J. Phys. Chem. B* **2010**, *114*, 11181–11195.

(38) Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K., *Structure* **2006**, *14*, 437-449.

(39) Ode H; Nakashima M; Kitamura S; Sugiura W; H, S., *Front. Microbiol.* **2012**, *3*, 258.

(40) Joshi, H.; Cheluvaraja, S.; Somogyi, E.; Brown, D. R.; Ortoleva, P. J., *Vaccine* **2011**, *29*, 9423-9430.

(41) Bishop, B.; Dasgupta, J.; Klein, M.; Garcea, R. L.; Christensen, N. D.; Zhao, R., *J. Biol. Chem.* **2007**, *282*, 31803-31811.

(42) Chen, X. S.; Garcea, R. L.; Goldberg, I.; Casini, G.; Harrison, S. C., *Mol. Cell* **2000**, *5*, 557-567.

(43) Bishop, B.; Dasgupta, J.; Chen, X., *Virol. J.* **2007**, *4*, 3.

(44) Ryding, J.; Dahlberg, L.; Wallen-Ohmann, M.; Dillner, J., *J. Gen. Vir.* **2007**, *88*, 792-802.

(45) Roden, R.; Armstrong, A.; Haderer, P.; Christensen, N.; Hubbert, N.; Lowy, D.; Schiller, J.; Kirnbauer, R., *J. Virol.* **1997**, *71*, 6247-6252.

(46) Nowak, M. A., *Sem. Vir.* **1996**, *7*, 83-92.

(47) Stanley, M.; Lowy, D. R.; Frazer, I., *Vaccine* **2006**, *24, Supplement 3*, S106-S113.

(48) Christensen, N. D.; Cladel, N. M.; Reed, C. A.; Budgeon, L. R.; Embers, M. E.; Skulsky, D. M.; McClements, W. L.; Ludmerer, S. W.; Jansen, K. U., *Virology* **2001**, *291*, 324-334.

(49) Carter, J. J.; Wipf, G. C.; Madeleine, M. M.; Schwartz, S. M.; Koutsky, L. A.; Galloway, D. A., *J. Virol.* **2006**, *80*, 4664-4672.

(50) Thönes, N.; Müller, M., *Virology* **2007**, *369*, 375-388.

(51) Schädlich, L.; Senger, T.; Gerlach, B.; Mücke, N.; Klein, C.; Bravo, I. G.; Müller, M.; Gissmann, L., *J. Virol.* **2009**, *83*, 7690-7705.

(52) Zhao, Q.; Allen, M. J.; Wang, Y.; Wang, B.; Wang, N.; Shi, L.; Sitrin, R. D., *Nanomed.-Nanotechnol.* **2012**, *8*, 1182-1189.

(53) Giedroc, D. P.; Arunkumar, A. I., *Dalton Trans.* **2007**, 3107-3120.

(54) Monod, J.; Changeux, J. P.; Jacob, F., *J. Mol. Biol.* **1963**, *6*, 306.

(55) Monod, J.; Wyman, J.; Changeux, J. P., *J. Mol. Biol.* **1965**, *12*, 88.

(56) Tzeng, S.-R.; Kalodimos, C. G., *Nature* **2009**, *462*, 368-372.

(57) Weinkam, P.; Pons, J.; Sali, A., *Proc. Natl. Acad. Sci. USA.* **2012**, *109*, 4875-4880.

(58) Kalodimos, C. G.; Nyas, *Ann. NY Acad. Sci.* **2012**, *1260*, 81-86.

(59) Kern, D.; Zuiderweg, E. R. P., *Curr. Opin. Struct. Biol.* **2003**, *13*, 748-757.

(60) Brunori, M., *Protein Sci.* **2011**, *20*, 1097-1099.

(61) Changeux, J. P.; Edelstein, S. J., *Science* **2005**, *308*, 1424-1428.

(62) Freire, E., *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11680-11682.

(63) Hilser, V. J., *Science* **2010** *327*, 653-654.

(64) Tsai, C. J.; del Sol, A.; Nussinov, R., *J. Mol. Biol.* **2008**, *378*, 1-11.

(65) Daily, M. D.; Gray, J. J., *Proteins* **2007**, *67*, 385-399.

(66) Swain, J. F.; Gierasch, L. M., *Curr. Opin. Struct. Biol.* **2006**, *16*, 102-108.

(67) Gunasekaran, K.; Ma, B. Y.; Nussinov, R., *Proteins* **2004**, *57*, 433-443.

(68) Hawkins, R. J.; McLeish, T. C. B., *J. R. Soc., Interface* **2006**, *3*, 125-138.

(69) Guo, H. C.; Feng, X. M.; Sun, S. Q.; Wei, Y. Q.; Sun, D. H.; Liu, X. T.; Liu, Z. X.; Luo, J. X.; Hong, Y., *Virol. J.* **2012**, *9*, 108.

(70) Benezra, M.; Penate-Medina, O.; Zanzonico, P. B.; Schaer, D.; Ow, H.; Burns, A.; DeStanchina, E.; Longo, V.; Herz, E.; Iyer, S.; Wolchok, J.; Larson, S. M.; Wiesner, U.; Bradbury, M. S., *J. Clin. Invest.* **2011**, *121*, 2768-2780.

(71) Kumar, S.; Huang, C.; Zheng, G.; Bohm, E.; Bhatele, A.; Phillips, J. C.; Yu, H.; Kale, L. V., *IBM J. Res. Dev.* **2008**, *52*, 177-188.

(72) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., *J. Comput. Chem.* **2005**, *26*, 1781-1802.

(73) Carrillo-Tripp, M.; Shepherd, C. M.; Borelli, I. A.; Venkataraman, S.; Lander, G.; Natarajan, P.; Johnson, J. E.; Brooks III., C. L.; Reddy, V. S., *Nucleic Acids Res.* **2009**, *37*, D436–D442.

(74) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *J. Chem. Phys.* **1983**, *79*, 926-935.

(75) Humphrey, W.; Dalke, A.; Schulten, K., *J. Mol. Graph.* **1996**, *14*, 33-38.

(76) Brooks, B. R.; Brooks III, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; A., B.; Caflisch, S.; Caves, L.; Cui, Q.;

Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; and Karplus, M., *J. Comp. Chem.* **2009**, *30, 1545-1614*.

(77) MacKerell, A. D.; Banavali, N.; Foloppe, N., *Biopolymers* **2000**, *56*, 257-265.

(78) Foloppe, N.; MacKerell, A. D., *J. Comput. Chem.* **2000**, *21*, 86-104.

(79) Lopes, P. E. M.; Murashov, V.; Tazi, M.; Demchuk, E.; MacKerell, A. D., *J. Phys. Chem. B* **2006**, *110*, 2782-2792.

(80) Jaguar 7.0, Schrödinger, Inc., New York, NY, 2013.

(81) Myung, S.; Lorton, K. P.; Merenbloom, S. I.; Fioroni, M.; Koeniger, S. L.; Julian, R. R.; Baik, M.-H.; Clemmer, D. E., *J. Am. Chem. Soc.* **2006**, *128*, 15988-15989.

(82) Mantri, Y.; Fioroni, M.; Baik, M.-H., *J. Biol. Inorg. Chem.* **2008**, *13*, 1197-1204.

(83) Grossoehme, N. E.; Giedroc, D. P., *Methods Mol. Biol.* **2012**, *796*, 31-51.

(84) Guerra, A. J.; Giedroc, D. P., *Arch. Biochem. Biophys.* **2012**, *519*, 210-222.

(85) Reyes-Caballero, H.; Campanello, G. C.; Giedroc, D. P., *Biophys. Chem.* **2011**, *156*, 103-114.

(86) Cooper, A.; Dryden, D. T. F., *Eur. Biophys. J. Biophy.* **1984**, *11*, 103-109.

(87) del Sol, A.; Tsai, C. J.; Ma, B. Y.; Nussinov, R., *Structure* **2009**, *17*, 1042-1050.

(88) Tsai, C. J.; Del Sol, A.; Nussinov, R., *Mol. Biosyst.* **2009**, *5*, 207-216.

(89) Hawkins, R. J.; McLeish, T. C. B., *Phys. Rev. Lett.* **2004**, *93*.

(90) Long, D.; Bruschweiler, R., *J. Phys. Chem. Lett.* **2012**, *3*, 1722-1726.

(91) Tsai, C.-J.; del Sol, A.; Nussinov, R., *J. Mol. Biol.* **2008**, *378*, 1-11.

(92) Hawkins, R. J.; McLeish, T. C. B., *Biophys. J.* **2006**, *91*, 2055-2062.

(93) Rohs, R.; Bloch, I.; Sklenar, H.; Shakked, Z., *Nucleic Acids Res.* **2005**, *33*, 7048-7057.

(94) Joshi, H.; Lewis, K.; Singharoy, A.; Ortoleva, P. J., *Submitted* **2013**.

(95) Vertegel, A. A.; Siegel, R. W.; Dordick, J. S., *Langmuir* **2004**, *20*, 6800-6807.

(96) Sun; al., e., *Proc. Natl. Acad. Sci. USA.* **2007**, *104*, 1354-1359.

(97) Meyers, M. A.; Chen, P.-Y.; Lin, A. Y.-M.; Seki, Y., *Progr. Mat. Sci.* **2008**, *53*, 1-206.

(98) Kneuer, C.; Sameti, M.; Haltner, E. G.; Schiestel, T.; Schirra, H.; Schmidt, H.; Lehr, C.-M., *Int. J. Pharm.* **2000**, *196*, 257-261.

(99) Parida, S. K.; Dash, S.; Patel, S.; Mishra, B., *Adv. Colloid Interfac.* **2006**, *121*, 77-110.

(100) Li, X.; Yang, T.; Gao, Q.; Yuan, J.; Cheng, S., *J. Colloid Interf. Sci.* **2009**, *338*, 99-104.

(101) Knopp, D.; Tang, D.; Niessner, R., *Anal. Chim. Acta* **2009**, *647*, 14-30.

(102) Barandeh F; Nguyen P-L; Kumar R; Iacobucci GJ; Kuznicki ML; Andrew Kosterman; Earl J. Bergey; Paras N. Prasad; Gunawardena, S., *PLoS One* **2012**, *7*, e29424.

(103) Aubin-Tam, M. E.; Hamad-Schifferli, K., *Biomed. Mater.* **2008**, *3*, 034001.

(104) Sperling, R. A.; Parak, W. J., *Philos. Trans. R. Soc., A* **2010**, *368*, 1333-1383.

(105) Jin, Y.; Li, A.; Hazelton, S. G.; Liang, S.; John, C. L.; Selid, P. D.; Pierce, D. T.; Zhao, J. X., *Coord. Chem. Rev.* **2009**, *253*, 2998-3014.

(106) Cruz-Chu, E. R.; Aksimentiev, A.; Schulten, K., *J. Phys. Chem. B* **2006**, *110*, 21497-21508.

(107) Lopes, P. E. M.; Murashov, V.; Tazi, M.; Demchuk, E.; MacKerell, A. D., *J. Phys. Chem. B* **2006**, *110*, 2782-2792.

(108) White, W. I.; Wilson, S. D.; Palmer-Hill, F. J.; Woods, R. M.; Ghim, S.-j.; Hewitt, L. A.; Goldman, D. M.; Burke, S. J.; Jenson, A. B.; Koenig, S.; Suzich, J. A., *J. Virol.* **1999**, *73*, 4882-4889.

(109) Thönes, N.; Herreiner, A.; Schädlich, L.; Piuko, K.; Müller, M., *J. Virol.* **2008**, *82*, 5472-5485.

(110) Zlotnick, A., *J. Mol. Recognit.* **2005**, *18*, 479-490.

(111) Quick, R.; Singharoy, A.; Ortoleva, P., *Chem. Phys. Lett.* **2013**, *571*, 61-65.

(112) Takada, S., *Curr. Opin. Struct. Biol.* **2012**, *22*, 130-137.

(113) Durrant, J.; McCammon, J., *BMC Biol.* **2011**, *9*, 71.

(114) Cheatham III, T. E.; Kollman, P. A., *Ann. Rev. Phys. Chem.* **2000**, *51*, 435-471.

(115) Karplus, M.; McCammon, J. A., *Nat. Struct. Mol. Biol.* **2002**, *9*, 646-652.

(116) Tang, F.; Li, L.; Chen, D., *Adv. Mater.* **2012**, *24*, 1504-1534.

SUMMARY

Allostery is a commonly used mechanism to regulate the activity of proteins, wherein ligand binding to one site of a protein alters the function of another site. Thermodynamically, allostery can take place either by a change in the mean structure of the protein (enthalpy driven) or by a change in the dynamics of the protein (entropy driven) or a combination of both. Structural changes in an allosteric protein could be realized by selection of a conformer from a pre-existing ensemble of conformers by the allosteric ligand (Monod, Wyman and Changeux model),[1] or induction of the structural change on binding of the allosteric ligand (Koshland, Nemethy and Filmer model).[2] The molecular basis of allosteric regulation remains a subject of extraordinary interest in biological systems, governing processes such as signal transduction, enzymatic activity, metabolite flux, and protein degradation. In this dissertation, we investigate the mechanisms of coupling between allosteric sites in two different protein systems, which is a quantitative measure of the manner in which these sites functionally interact (enthalpy or entropy driven). Bacterial metalloregulatory proteins and viral capsid proteins were employed as model systems to gain insights into the nature of allostery by employing density functional theory and all-atom molecular dynamics simulations.

In the first case, the effector ligand is a metal ion, namely Cu(I) ion, and the protein "active site ligand" is DNA. In CsoR proteins, Cu(I) behaves as an allosteric inhibitor of DNA binding. The molecular origins behind Cu(I) mediated allosteric transitions were investigated in CsoR proteins of *M.tuberculosis*, *T.thermophillus* and *S.lividans*. In Mtb, both the Cu(I) bound and apo forms of CsoR were studied to identify key structural features that may determine the affinity to bind to Cu(I) or DNA promoter.[3-5] All-atom MD simulations and subsequent analysis were performed on the apo and Cu(I) bound forms of Mtb-CsoR and apo forms of *Thermus thermophillus* and *Streptomyces lividans* species. The structural comparison of apo- and Cu(I)-bound Mtb-CsoR yielded an RMS value of ~6.20 Å in the tetrameric form. A global alignment of both Cu(I) and apo forms of Mtb-CsoR in dimeric form (Figure 22(c) in Chapter 3) reveals that the two protomers in the apo CsoR reorient to attain a 'X-like' conformation due to the absence of an anchor (which is Cu(I)) that holds the protomers in a 'taut' and straight conformation. The resulting 'X-like' conformation can be more suitable to bind to DNA than a rigid and inflexible form that was determined to be characteristic of Cu(I)-bound Mtb-CsoR. The essential H-bond interactions between the conserved residues in CsoR proteins, that stabilize the apo- and Cu(I)-bound Mtb-CsoR in their conformations were identified and a rationale was drawn to explain the influence of these H-bond interactions on the flexibility of CsoR protein structure.

Differences were also observed in the apo CsoR proteins of different bacterial species. In the apo forms of Mtb and Tt CsoR proteins, the H-bond between His61-Tyr35-Glu81 was not observed, while in apo Sl-CsoR the H-bond interaction between Tyr74-Glu122 and a bridging water molecule that interacts with both Glu122 and His100 were observed.[3] Though the presence of C-terminal tail in Mtb-CsoR differentiates it from the CsoR proteins of other bacterial species, our simulations were performed on the Mtb-CsoR protein without the C-terminal tail in order to compare these structural changes with those of CsoR proteins in other bacteria. Further computational studies and analysis should be performed to gain insight into specific conformations of apo and Cu(I)-bound CsoR proteins of different bacteria. Overall in Mtb-CsoR, our simulations were able to predict differences in orientation of α-helices, H-bond interactions and increase in flexibility of conserved arginine residues involved in DNA binding, on binding to Cu(I).[6] The allostery in CsoR proteins can be explained on the basis of "population shift model"[7] where the equilibrium shifts to the conformation ensemble with a rigid and open structure, that possesses least affinity to DNA (open or taut) and is stabilized by binding to Cu(I). Future studies focused on CsoR bound to DNA should be conducted to understand the DNA bound conformation.

In Chapter 4, a computational study that used all-atom structures of hybrid silica-protein nano-constructs was performed to provide a novel nanoscale perspective on a long-standing challenge of VLP-based HPV vaccine design. Interestingly, in L1 capsid proteins of HPV, the allostery between h4 helix and epitopes does not invoke any significant changes in the mean structures of the epitopes (RMS < 1 Å) unlike that of CsoR proteins in bacteria. Instead, the allosteric connection is expressed in variations of structural fluctuations (dynamics or entropy driven), which in turn impact the biochemical behavior of the epitopes and hence effects the immune response.

In our simulations, we discovered an intriguing dynamic allosteric coupling between the h4 helix and the epitope containing loops of the L1 capsid proteins of HPV. Based on the concepts of dynamic allostery in Chapter 4, we devised an effective exploitation strategy for rigidifying the epitopes in lower assemblies (L1 monomer and pentameric forms) to reproduce the structural and dynamic properties of these epitopes in the VLP or higher assemblies (T=1 or T=7 VLP) using the silica surface mounted pentamer. Silica nanoparticles provide a potentially revolutionary opportunity for developing vaccines and we demonstrate how they can be utilized in the hybrid vaccine design. This work highlights an intriguing connection between structure, dynamics and function of proteins, while explicitly outlining a strategy for exploiting dynamic allostery

which is a relatively new concept that remains poorly understood.[8] The dynamic allosteric control mechanism outlined in this work appears to be generally applicable and more common than previously thought and deserves further investigation in order to develop new rational drug and vaccine design strategies.

PERSPECTIVES

The key implication of understanding allosteric mechanisms in bacterial and viral proteins is to design drugs and vaccines that can modulate the biochemical behavior by switching on/off signaling pathways while impacting the pathogenicity and viability of these microorganisms in human hosts.[9-11] Computational methods such as MD simulations are crucial tools to understand the existence of conformational ensembles of different allosteric states and to identify key residue-residue networks that propagate within proteins when switching from one conformational state to another.[7,12-15] Additionally, these methods help to detect allosteric sites, analyze them for their potential as drug targets and also suggest small molecule drug targets, which can then be tested in experimental screening. Hence, understanding the nature and molecular mechanisms of allostery is required in pathogenic bacterial proteins. Further computational evaluation of CsoR proteins is essential to design allosteric drugs that can bind these proteins and modulate the behavior by switching on/off the

Cu(I) regulatory mechanisms/networks. For an effective vaccine design, different viral capsid proteins should be screened for dynamic allostery and computational design targeted towards a single hybrid nanoparticle based VLP vaccine for different viruses should be established in future studies. Below, future computational studies are proposed to understand these systems and direct the allostery towards rational drug and vaccine design.

Cu(I)-bound forms of Tt-CsoR and Sl-CsoR

Both Tt- and Sl-CsoR proteins are the best structural targets to study different allosteric states due to the absence of long C-terminal tail (Mtb CsoR), like nearly all other CsoRs. Computational models of apo Tt-CsoR and Sl-CsoR were described in this thesis, but the conformation of Cu(I)-bound forms should be modeled to evaluate the allosteric transition and conformational changes. The force fields for the Cu(I) coordination sphere were also obtained for Tt-CsoR in the His-Cys-His binding mode (data not shown). Due to the lack of crystal structure for Cu(I)-bound forms of Tt-CsoR and Sl-CsoR, it is difficult to build these structures using traditional MD programs. Steered MD coupled with QM/MM calculations should be performed in order to enforce the Cu(I) coordination geometry without disruption of the α-helical character. The force fields developed can then be applied and subsequent MD simulation analysis will enable us to

determine different allosteric states and conformational changes in CsoR proteins can be studied at molecular/atomic level of detail.

How does CsoR bind to its DNA operator?

The Cu(I)-bound structure of Mtb CsoR reveals a novel protein architecture with no classical DNA binding motif but the characteristic antiparallel four-helix bundles have been suggested to act as a DNA-binding fold.[4,16,17] High resolution structural studies coupled with a better understanding of the DNA operator will provide further insights into the protein-DNA complex formation and how Cu-binding inhibits the complex. A comparison of all three allosteric states (DNA-bound, apo and Cu(I)-bound) will provide detailed insights as to how this putative hydrogen bonding network drives allosteric communication. A recent report on the DNA operator binding S.lividans CsoR predicts that the α-helix RLXR motif established contact at the GTA dyad regions of the DNA operator site.[17] Using docking techniques, the apo Sl-CsoR can be docked onto the DNA operator. Steered and directed molecular dynamics simulations can be used to further refine the structure of the DNA-CsoR complex of *S.lividans*. Multiscale and coarse grained simulations will be used to simulate the DNA-CsoR protein complex to understand the residue interactions and the conformation of CsoR when bound to DNA. These studies will enable us to determine the structural allostery in Sl-CsoR

proteins. The role of H-bond interactions can then deciphered by comparing all the three allosteric states of CsoR.

Role of the C-terminal tail of Mtb-CsoR

As discussed, Mtb CsoR has a unique C-terminal tail consisting of ~30 amino acids residues, ~18 of which are highly conserved in pathogenic mycobacteria. An unexpected finding was that this region plays an important role in DNA binding since truncation of the tail before residue 106 significantly decreases the DNA binding affinity.[4] The conserved residues in the tail are mostly neutral and hydrophobic in nature, and are therefore unlikely to interact directly with the DNA. One possibility is that the tail influences the oligomerization or the conformational state of CsoR in solution when bound to DNA.[4] The crystallographic structure of Cu(I)-bound CsoR reveals that the tetramer may form as a dimer of dimers, with the a3 helices positioned at the dimer-dimer interface.[4] Although the C-terminal tail is not resolved in the X-ray structure, it is very likely to be found at this interface where it could potentially stabilize the tetramer or alternatively change the distance between positively charged patches on one face of the tetramer.[4] This may also suggest the possibility that the C-terminal tail can wrap around the dimers and might mediate disassembly of the protein-DNA complex by interfering with DNA promoter contact sites. In any case, a systematic

study on the assembly state of the full-length Mtb-CsoR will likely shed considerable light on understanding the fundamental role of the C-terminal tail in Mtb CsoR.

Virtual Screening of Allosteric Drug-like Molecules that Bind CsoR Proteins

Target-based ligand screening will enable us to assess many drug-like compounds for potential ligands to bind CsoR protein in pathogenic bacteria.[18] Experimental assays are time-consuming, expensive, and not always applicable. At the same time, computational approaches using docking and protein pocket prediction methods have limited accuracy.[19-21] Using MD simulations and generating high-resolution structural data of proteins, will not only enable us to target various binding pockets in the protein but also allows us to determine the conformational ensemble of the allosteric proteins when bound to small ligands. This target based drug screening approach combined with the knowledge of various residue interactions and H-bonds stabilizing an allosteric state, would provide us with a potential strategy to screen for small molecules which can modulate the allosteric behavior of CsoR proteins.

Building a Library of Viral Epitopes to Effectively Design VLP-Nanoparticle Based Vaccines

In this dissertation, we described an approach to exploit the dynamic allostery present in HPV L1 capsid protein in the design of hybrid nanoparticle-based vaccines. This method could be applied in screening various epitopes in viral capsids to identify the presence of dynamic allostery and its impact on the immunogenicity of the VLP.[8] Detailed analysis of epitopes is important both for the understanding of immunological events and for the development of more effective vaccine and diagnostic tools for various diseases. Identification and characterization of epitopes is a complex process. Although various methods have been developed in this area, there still lacks a simple common approach which can be applied to all epitopes. X-ray crystallography remains the only suitable method to accurately perform structural analysis of any epitope, understand allostery and the immunogenic influence. However, there have been a large number of computational methods developed for mapping allostery of different proteins which are discussed in this theses. The proposed strategy involves constructing putative hybrid vaccine computationally, simulating these epitope constructs to arrive at quantitative metrics that influence the immunogenic properties, and then using the latter to construct a model of a vaccine with high immunogenicity. The designs scored to be highly immunogenic can then be synthesized and tested using experimental assays. This vaccine discovery workflow integrating computation schemes and laboratory techniques could considerably reduce time and costs

associated with experimental screening assays.[22] A database of all the viral epitopes which are tested to be immunogenic can then be established using bioinformatics approaches that will facilitate us to rationally select viral epitopes with high immunogenicity while constructing a hybrid VLP.[23-25] Different viral epitopes determined using the vaccine discovery workflow suggested above, can be assembled and presented using a nanoparticle (silica nanoparticles) to achieve immunity to several diseases caused by viruses. Though immense amount of computational work is required to accomplish this goal, a streamlined process in epitope identification and discovery to develop vaccines and diagnostics will be beneficial during a pandemic with the available knowledge database of immunogenicity of epitopes and biocompatible nanoparticles.

REFERENCES

(1) Monod, J.; Wyman, J.; Changeux, J.-P., *J. Mol. Biol.* **1965,** *12*, 88-118.

(2) Koshland Jr, D.; Nemethy, G.; Filmer, D., *Biochemistry* **1966,** *5*, 365-385.

(3) Dwarakanath, S.; Chaplin, A. K.; Hough, M. A.; Rigali, S.; Vijgenboom, E.; Worrall, J. A., *J. Biol. Chem.* **2012,** *287*, 17833-17847.

(4) Liu, T.; Ramesh, A.; Ma, Z.; Ward, S. K.; Zhang, L. M.; George, G. N.; Talaat, A. M.; Sacchettini, J. C.; Giedroc, D. P., *Nat. Chem. Biol.* **2007,** *3*, 60-68.

(5) Sakamoto, K.; Agari, Y.; Agari, K.; Kuramitsu, S.; Shinkai, A., *Microbiology* **2010,** *156*, 1993-2005.

(6) Chang, F.-M. J.; Lauber, M. A.; Running, W. E.; Reilly, J. P.; Giedroc, D. P., *Anal. Chem.* **2011,** *83*, 9092-9099.

(7) Kar, G.; Keskin, O.; Gursoy, A.; Nussinov, R., *Curr. Opin. Pharmacol.* **2010,** *10*, 715-722.

(8) Singharoy, A.; Polavarapu, A.; Joshi, H.; Baik, M.-H.; Ortoleva, P., *J. Am. Chem. Soc.* **2013,** *135*, 18458-18468.

(9) Conn, P. J.; Christopoulos, A.; Lindsley, C. W., *Nat. Rev. Drug Discovery* **2009,** *8*, 41-54.

(10) Nussinov, R.; Tsai, C.-J.; Csermely, P., *Trends Pharmacol. Sci.* **2011,** *32*, 686-693.

(11) Nussinov, R.; Tsai, C.-J., *Cell* **2013,** *153*, 293-305.

(12) Durrant, J. D.; McCammon, J. A., *BMC biology* **2011,** *9*, 71.

(13) Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A., *J. Am. Chem. Soc.* **2001,** *123*, 12658-12663.

(14) Perryman, A. L.; Lin, J. H.; McCammon, J. A., *Biopolymers* **2006,** *82*, 272-284.

(15) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E., *J. Am. Chem. Soc.* **2011,** *133*, 9181-9183.

(16) Solioz, M.; Abicht, H. K.; Mermod, M.; Mancini, S., *JBIC, J. Biol. Inorg. Chem.* **2010,** *15*, 3-14.

(17) Tan, B. G.; Vijgenboom, E.; Worrall, J. A., *Nucleic Acids Res.* **2013**, gkt902.

(18) Cavet, J. S., In *Stress Response in Pathogenic Bacteria*, 2011; Vol. 19, p 266.

(19) Liang, S.; Zhang, C.; Liu, S.; Zhou, Y., *Nucleic Acids Res.* **2006,** *34*, 3698-3707.

(20) Brady Jr, G. P.; Stouten, P. F., *J. Comput.-Aided Mol. Des.* **2000,** *14*, 383-401.

(21) Huang, B., *OMICS: J. Integr. Biol.* **2009,** *13*, 325-330.

(22) Joshi, H.; Lewis, K.; Singharoy, A.; Ortoleva, P. J., *Vaccine* **2013,** *31*, 4841-4847.

(23) Wang, L.-F.; Du Plessis, D. H.; White, J. R.; Hyatt, A. D.; Eaton, B. T., *J. Immunol. Methods* **1995,** *178*, 1-12.

(24) Keller, P. M.; Arnold, B. A.; Shaw, A. R.; Tolman, R. L.; Van Middlesworth, F.; Bondy, S.; Rusiecki, V.; Koenig, S.; Zolla-Pazner, S.; Conard, P., *Virology* **1993,** *193*, 709-716.

(25) Georgiou, G.; Stathopoulos, C.; Daugherty, P. S.; Nayak, A. R.; Iverson, B. L.;

Curtiss III, R., *Nat. Biotechnol.* **1997,** *15*, 29-34.

Curriculum Vita

Abhigna Polavarapu

Department of Chemistry, UNC, Chapel Hill, North Carolina. | 317.985.8051 | polavara@unc.edu

EDUCATION & EMPLOYMENT

Post-Doctoral Research Associate, UNC Chapel Hill, North Carolina   Current
Joint post-doc with Eastman Chemicals and UNC Chapel Hill
Advisor: Prof. Cynthia Schauer

Ph.D., Biological Chemistry, Indiana University, Bloomington, Indiana   2015
Thesis Title: Computational Design of New Generation Drugs and Vaccines
Thesis Advisor: Prof. Mu-Hyun Baik

M.S., Cheminformatics, School of Informatics, Indiana University, Bloomington, Indiana 2008
Thesis Title: Validity of Homology Models in Ligand Docking
Thesis Advisor: Prof. David Wild

Bachelor of Engineering, Biotechnology, Andhra University, India   2006
Thesis Title: Production of Ground-Nut based Bio-Fuel

PROFILE

- Computational chemist with experience in studying biological systems and metal catalysts
- Molecular mechanics force field development
- Familiar with cyclic voltammetric techniques and process engineering
- Comfortable with handling statistical analysis of data and programming languages
- Extensively trained in oral and written communication about chemical research
- Worked in many diverse research teams including experts from very diverse backgrounds

COMPUTER SKILLS

Databases: MySQL, MS Access, SQL
Languages: C, PERL, MATLAB, Python, R, Unix and Linux Script
Internet Technologies: PHP, HTML, DHTML, CSS
Software: Schrödinger package for Molecular Modeling, QSITE for QM/MM, GROMACS Molecular Dynamics Tool, CHARMM Molecular Dynamics package, NAMD, CPMD, Gaussian, Materials Studio, ADF, Parallel computing and VMD visualizing tools.

RESEARCH PROJECTS

I.   Design of Effective Rhodium based Hydroformylation Catalysts

II.   Prediction of Allosteric Regulation of Metal Sensing by CsoR Protein

III.   Computational Study of Reaction Mechanisms of Organic Nitrogen Mustards Anti-cancer Drugs and Reduction of Pt(IV)-based Anti-cancer Agents

IV.   Computer Aided Drug Design- Design of Novel Silica and Iron Oxide Nanoparticle-Based Vaccines against Human Cervical Cancer

V.   AutoGeff – Automatic Generator of Force Fields for Metal Active Sites in Proteins

VI.   Validation of Homology Protein Models using Small Molecule Docking and QSAR techniques.

PUBLICATIONS

- McCormick, M. C.; Keijzer, K.; **Polavarapu, Abhigna**.; Schultz, F. A.; Baik, M.-H. Understanding Intrinsically Irreversible, Non-Nernstian, Two-Electron Redox Processes: A Combined Experimental and Computational Study of the Electrochemical Activation of Platinum(IV) Antitumor Prodrugs. *J. Am. Chem. Soc.* **2014,** *136*, 8992-9000.

- Singharoy, A.; **Polavarapu, Abhigna**.; Joshi, H.; Baik, M.-H.; Ortoleva, P. Epitope Fluctuations in the Human Papillomavirus Are Under Dynamic Allosteric Control: A Computational Evaluation of a New Vaccine Design Strategy. *J. Am. Chem. Soc.* **2013,** *135*, 18458-18468.

- **Polavarapu, Abhigna**; Stillabower, J. A.; Stubblefield S. G. W.; Taylor, W. M.; and Baik, M.-H. The Mechanism of Guanine Alkylation by Nitrogen Mustards: A Computational Study. *J.Org.Chem,* **2012***, 77, 5914-5921.*

- **Polavarapu, Abhigna**; Baik, M.-H, Giedroc, David. Copper(I) induced Allosteric Regulation of CsoRs in *Mycobacterium tuberculosis*. *Manuscript in preparation*.

CONFERENCES
- Poster presentations at the PINDU Conference consequently from 2008 -2011
- Poster presentation at the Women in Science Conference, Bloomington, Indiana, 2010
- Contribution to the research presented on behalf of IU at XSEDE Supercomputing Conference 2011
- Poster presentation at the Gordon Research Conference, Bio-inorganic Chemistry 2013

AWARDS
- Summer Research Fellowhip, Andhra University, 2004
- Summer Research Fellowship, Indian Academy of Sciences, 2005
- Gold Medalist from Andhra University for securing top most grade 2002-2006
- ChemGRC Travel Award, Indiana University, Bloomington, IN 2012

PROFESSIONAL ACTIVITIES
- Chair, Chemistry Department Post-Doctoral Association (CDPA), UNC, Chapel Hill (Current)

TEACHING EXPERIENCE
- Associate Instructor, Indiana University, Bloomington, IN, 2008-2012
- Led discussion sections for **general chemistry** and **advanced biochemistry** courses at IU
- Set up and instructed laboratory sections for advanced biochemistry courses