8-2016

Molecular Mechanics Studies of Protein Signaling and Conformation Change

Zhe Jia
Clemson University

MOLECULAR MECHANICS STUDIES OF PROTEIN SIGNALING AND CONFORMATION CHANGE

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Chemistry

---

by
Zhe Jia
August 2016

Accepted by:
Dr. Brian Dominy, Committee Chair
Dr. Steve Stuart
Dr. Emil Alexov
Dr. Julia Brumaghim

ABSTRACT

Physiology signals can be passed by proteins. Many protein signaling starts from ligand binding and undergoes conformation change of the receptors. Many cellular surface receptor proteins contain a Von Willebrand factor (vWF), which is a large multimeric glycoprotein present in blood plasma. This dissertation employed molecular dynamics (MD) simulation to investigate the binding and signaling process of several vWF type A proteins.

Chapter 2 discussed the potential errors modeling and MD sampling methods, and evaluated the accuracy and precision of free energy calculation. An optimized sampling strategy was established to obtain the best computational efficiency. The strategy can be applied to a wide range of protein binding research.

The following chapters investigated the binding and signaling process of anthrax receptors and integrins, which are vWF type A proteins. Binding mechanism, possibility of conformational change, and the role of metal ion in binding process, were analyzed and compared for two structurally highly similar anthrax receptor proteins, tumor endothelial marker 8 (TEM8) and capillary morphogenesis gene 2 (CMG2). The two highly similar proteins are the drug target for distinct diseases. The differences in these

two processes were found can guide the further development of drug specifically targeting one of the proteins. A conformation change between open and closed conformation is known to exist in most vWF type A proteins, but has not been experimentally observed in the anthrax receptors. Chapter 5 investigated the binding and conformation change process of integrins using targeted molecular dynamics simulation, and compared with anthrax receptors. The key residues and correlated motions in conformation change process were revealed, which can serve as a reference to the development in small molecule inhibitors of the signaling process. Results further confirmed the difficulties of observing conformation change in anthrax receptors.

DEDICATION

To my parents
Gongli Jia and Jing Wang
And
To my wife
Yu Fu

This humble work is a sign of my love to you!

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my advisor Prof. Brian Dominy for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to special thanks to Prof. Steve Stuart and his research group for insightful discussions, and thank the rest of my dissertation committee: Prof. Emil Alexov, and Prof Julia Brumaghim, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I thank my fellow labmates Vibhor Agrawal, Tingting Han, Yinling Liu, Richard Overstreet in for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last five years. Also I thank my friends live in Hunters Glen and Heritage Point communities.

TABLE OF CONTENTS

a e of Contents (Continued) Page

a e of Contents (Continued) Page

LIST OF FIGURES

Page

List of Figures (Continued) Page

LIST OF TABLES

Page

CHAPTER 1

INTRODUCTION

Molecular Mechanics Generalized Born Surface Area method is widely used to estimate the binding free energy of protein-ligand interaction and recently for protein-protein interaction. The binding process can be described using the following equation:

where P is the free protein; L is the free ligand; and PL is the protein-ligand complex. In this dissertation, I will first evaluate the energy calculation method, and followed by some applications on the protein-protein binding and protein conformation change I have mainly studied. For instance, in protein-protein complexes, the dissociation constant K_d of Equation 1.1 is measured experimentally using Surface Plasmon Resonance,

The dissociation constant can be further correlated to ΔG, the binding free energy,

where R,T and C^0 are the gas constant, temperature in Kelvin and the standard concentration, respectively. Through this dissertation, ΔG will be used as a general

notation of free energy, for both Gibbs free energy and Helmholtz free energy (Equation 1.3). In this dissertation, constant pressure and temperature are used in protein simulation, as in most condensed system simulation conditions. Thus the pressure volume term in Equation 1.3 is negligible, which should be added for a correct definition of Gibbs free energy [1].

1.1 Thermodynamics basis of protein-ligand binding

The dissociation constant K_d can also be explained by a modern statistical form shown by Sharp[2] and Roux[3]. When ligand (L) molecules are bound to protein (P) molecules, p_1 is the probability of one of the ligand molecules is bound to one of the protein molecules. Conversely, the probability of this ligand molecule is unbound is defined as p_0. And $p_0 + p_1 = 1$. Thus, if we define the total protein concentration in the system is $[P]_{tot}$, then the unbound protein concentration $[P] = p_0[P]_{tot}$, and the concentration of protein-ligand complex $[PL] = p_1[P]_{tot.}$ Hence,

$$K_d = \frac{p_0[P]_{tot}[L]}{p_1[P]_{tot}} = \frac{p_0[L]}{p_1} \tag{1.4}$$

A configurational average is needed to calculate the probabilities. Considering an operator H that gives value 1 for the bound state, and 0 otherwise. The configurational space can be separated into bound and unbound (bulk). Further, p_0 and p_1 can be

expressed as

$$p_0 = \frac{\int_{bulk} dq \int dS e^{-\beta U}}{\int dq \int dS e^{-\beta U}} \tag{1.5}$$

and

$$p_1 = \frac{\int_{bound} dq \int dS e^{-\beta U}}{\int dq \int dS e^{-\beta U}} \tag{1.6}$$

where dq is the coordinate of one of the ligands, dS is the coordinate of the surroundings, $\beta \equiv 1/RT$, and U is the total potential energy. Now, K_d can be written as

$$\mathrm{K_d} = [\mathrm{L}] \frac{\int_{bulk} dq \int dS e^{-\beta U}}{\mathrm{N} \int_{bound} dq \int dS e^{-\beta U}} \tag{1.7}$$

where the indistinguishable ligands are summed up to N. If we assume the bulk part is isotropic and homogenous, the deviation can be further taken to

$$\begin{aligned} \mathrm{K_d} &= [\mathrm{L}] \frac{\mathrm{V} \int_{bulk} dq\, \delta(r_1 - r^*) \int dS e^{-\beta U}}{\mathrm{N} \int_{bound} dq \int dS e^{-\beta U}} \\ &= \frac{\int_{bulk} dq\, \delta(r_1 - r^*) \int dS e^{-\beta U}}{\int_{bound} dq \int dS e^{-\beta U}} \end{aligned} \tag{1.8}$$

where V is the volume of the bulk, δ is the Dirac delta function, r_1-r^* is the distance between the ligand molecule and some arbitrary position in the bulk. The last equality can be performed if we allow [L] = N/V. Equation 1.8 implies that the ratio of the likelihood in bound state to the ratio of likelihood in bulk state is legit to represent K_d. It gives the components we need to calculate to obtain the binding free energy.

Based on Equation 1.8, there are three main components: *U*, the total potential energy of the complex; the configurational integrals and the free energy. Unfortunately, the free energy cannot be evaluated from Equation 1.8 in practice, because it is not practical to sample all possible configurations in a bulk system. Therefore, we need to simplify the equation by making assumptions such that it can be computed. Chapter 2 will focus on evaluating MM/GBSA method to estimate the binding free energy in protein-protein complex.

1.2 Anthrax and protective antigen

Anthrax is a disease known to affect herbivorous animals, such as sheep, cows, goats, and horses. The disease was extremely rare in humans and was not paid much attention to, until 2001, it was used as a bioterrorism weapon resulting in the deaths of 5 people. The attack reinstates the interest in anthrax research[4-6]. I employed computational methods modeled and analyzed the binding mechanism of anthrax toxin (Chapter 3 and 4). The free energy estimation methods were examined using anthrax related proteins in Chapter 2.

Anthrax is caused by an organism called Bacillus anthracis. Although studies have shown

bacterial titer in the bloodstream as high as 10^9 bacteria per mL can be eliminated within hours; once the bacteria appear in the blood, treatment must be initiated within 48 hours. Bacteria can still be killed after 48 hours of the infection, but enough anthrax toxin will remain to be fatal[4,5]. Another important characteristic of Bacillus anthracis is that it is a spore forming organism. The spore is able to survive extremely harsh conditions. The survival time can be up to several decades[7]. In connection with bacterial weapon, this is the most significant attribute of the bacterium. A massive attach using anthrax spores would result in the inhalation and infection of such a large number of people that ample supply and administration of a drug would be very unlikely within 48-hour period.

Bacillus anthracis, is a large rod-shaped, Gram-positive, soil bacterium. The two virulence factors are the capsule and the anthrax toxin[8]. The major cause of death in anthrax is the 3-protein toxin that the bacterium secretes. The three anthrax proteins are designated protective antigen (PA), edema factor (EF), and lethal factor (LF)[9]. PA is nontoxic, but it is responsible for the delivery of the two toxic proteins, EF and LF, which disable the normal immune function of the host[10,11]. PA (83 kDa) binds to one of its two receptors, following cleavage by a protease, forms a heptamer that is able to bind and deliver EF and/or LF into the cell [12]. The X-ray crystal structures of protective antigen, edema factor and lethal factor have all been determined[13,14].

The four domains of PA monomer (Figure 1.1) are mainly made of antiparallel β-sheets[13]. Domain I (residues 1-258) contains two calcium ions and the cleavage site for proteases to activate the protein. An amino terminal fragment of 20kd is cleaved after the protease was activated[15]. Domain II (residues 259-487) participates in the formation of the heptamer. The flexible loop on domain II (β2-β4) aids in membrane insertion. To date, the function of Domain III (residues 488-595) is still not formally reported. Domain IV (residues 596-735) donates the majority of affinity in receptor binding[13].

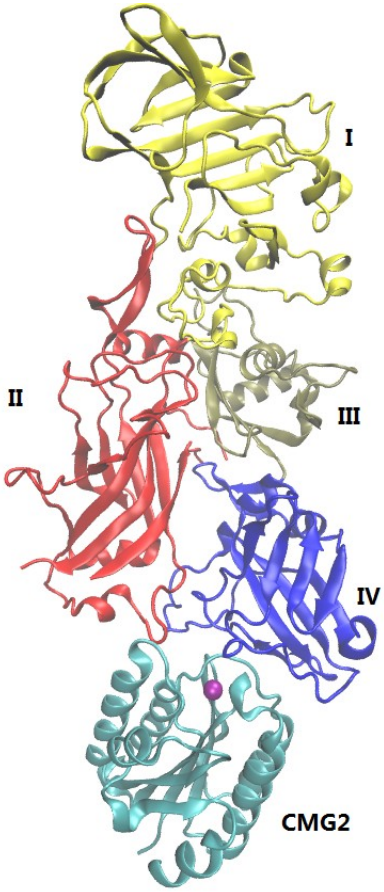


Figure 1.1 Structure of protective antigen binding with CMG2. The upper part shows the four-domain structure of PA83, the lower part shows the VWA domain of CMG2. The four PA domains are colored separately, domain I in yellow, domain II in red, domain III in tan and domain IV in blue.

1.3 Anthrax receptors

Binding to target cells is the entire responsibility of protective antigen. On the surface of mammalian cells, two anthrax receptors, Tumor endothelial marker 8 (TEM8 or ANTXR1) and capillary morphogenesis gene 2 protein (CMG2 or ANTXR2) were found.

TEM8 was the first anthrax toxin receptor described. Originally discovered in endothelia of colon carcinomas by serial analysis of gene expression, TEM8 has been postulated to be important for tumor angiogenesis[16]. To date, three splice variants of TEM8 have been reported: var1 (564 residues), var2 (368 residues), var3 (333 residues), var4 (528 residues), var5 (358 residues). Var1, var2 and var4 are single-pass type-1 integral membrane proteins, whereas var3 and var5, lacking the transmembrane helices, are thought to be a secreted form. The anthrax toxin receptor was first identified as the 368-residue variant 2[17].

The other toxin receptor, CMG2, was first reported in 2001 as a new gene up-regulated during in vitro capillary morphogenesis[18]. In 2003, mutations on CMG2 was identified in infantile systemic hyalinosis and juvenile hyaline fibromatosis patients [19,20]. This protein exhibited elevated expression during umbilical cord endothelial cell tubulation in

collagen matrices[18]. Similar to TEM8, CMG2 also has splice variants: CMG2-488, 489, 386, and 322. The first three are single-pass type-1 integral membrane proteins, and the last one, CMG2-322 is likely a secreted protein as it lacks the transmembrane helix. Excluding the last 13 residues, the primary structures of CMG2-488 and CMG2-489 are identical. CMG2-386 was first formally identified in 2001[18]. In 2003 CMG2-489 was found to be an anthrax toxin receptor, and it is now known that CMG2-488 is also an equally good toxin receptor[10,21]. CMG2-386 differs from CMG2-489 only in that it lacks the membrane-proximal segment spanning residues 218-318. Otherwise the two sequences are the same.

Both receptors are type I transmembrane proteins that possess an integrin-like extracellular I domain. The PA receptors are known to have natural ligands, TEM8 interacts with collagen VI[22], and CMG2 interacts with collagen IV and laminin[18]. Both receptors share highly similar von Willebrand factor type A (vWA) domain. These domains that locate in their extracellular N-terminal are 60% identical in amino acid sequence. The vWA domain binds to PA domains II and IV via its metal ion dependent adhesion site (MIDAS)[23,24]. The interaction of binding mimics the recognition of integrins to the extracellular matrix. Unlike the physiological ligands of vWA domains, a hydrophobic pocket in the receptor holds an insertion of a loop of PA domain II. Due to

the hydrophobic ridge that is built by this insertion, the interaction of PA with its receptor is at least 10-fold (some have reported about 1000-fold) higher than most other ligands of vWA domain[25].

CMG2 is widely distributed in normal adult tissues (e.g., brain, kidney, lung and muscle). Otherwise, TEM8 is seldom observed in normal tissues but abundant in the vasculature of developing embryos and tumor endothelial cells[10,22,26]. The distribution in normal tissues makes CMG2 the dominating receptor that responsible for the lethality of anthrax toxin[27]. Further, the PA binding affinities of CMG2 and TEM8 differ by an order of magnitude. The reported affinity of CMG2-PA (K_d=170 pM, $\Delta G_{binding}$=-56.1 kJ/mol) is much higher than TEM8-PA (K_d=1.1 μM, $\Delta G_{binding}$=-34.2k J/mol) [21,24]. (K_d values are collected from prior experimental works, $\Delta G_{binding}$ are calculated by $\Delta G_{binding}=RTlnK_d$)

A majority of interaction energy comes from the interaction between PA and vWA motif on receptors. MIDAS site of both receptors bind to PA by attracting an aspartic acid side chain (D683) from PA that completes the coordination sphere of the MIDAS metal ion[25]. Toxicity can be completely removed by mutating the aspartic acid to asparagine [23]. There are two possible conformations for the MIDAS motif in vWA domains: closed state (low-affinity ligand bound) and open state (high affinity ligand bound)[28]. Based on the

crystal structure of CMG2-PA binding conformation, CMG2 is reported to adopt the open state, and this open state is considered as a reason that CMG2 has a high affinity towards PA[7,11,25]. Mutational analysis suggests that TEM8 also adopts an open state conformation[23] and was later proved by the structure of TEM8 that was obtained in 2010[29]. To date, the structure of TEM8-PA binding state has not been formally reported.

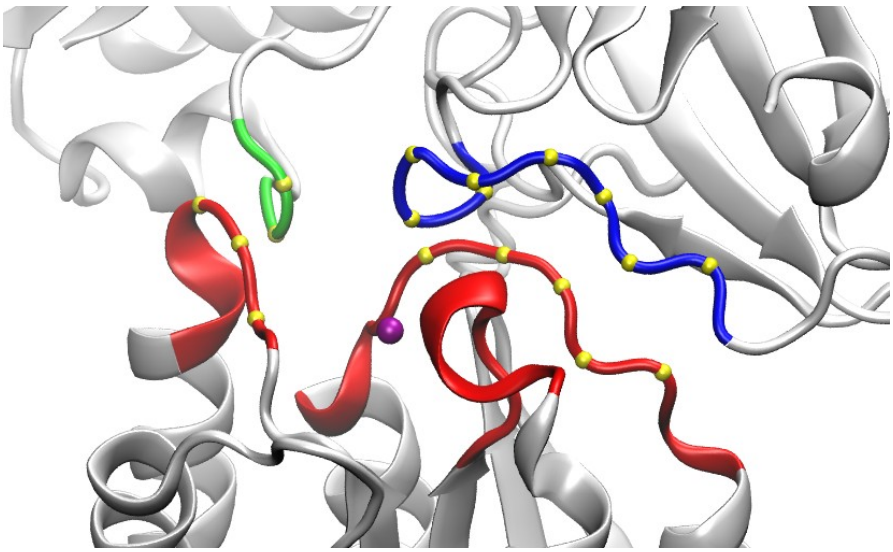


Figure 1.2 Interaction surface between PA and CMG2. Contacting loops are colored red for CMG2. The green part shows PA domain II inserts into the pocket formed by CMG2. The blue loop interacts with the ridge on CMG2 surface. All α-carbon atoms of residues on the loops are colored yellow. Mn^{2+} ion in the MIDAS domain is colored purple.

Large contact surface between PA and receptors is also responsible for the binding

affinity. Comparing to a typical α-integrin-ligand (~1300 Å^2)[11, 25], CMG2 has a much larger contact surface (~2000 Å^2) with its ligand (PA). The contact surface is mainly contributed by the buried surface area between PA domain IV and CMG2 residues around the MIDAS motif. On CMG2, the α2-α3 loops are pushed out from the MIDAS motif and form an outstanding hydrophobic group that inserts into a hydrophobic pocket formed by residues on PA IV. PA domain II contributes a smaller part to the buried surface area between PA and CMG2. On PA II near CMG2, a β-hairpin contains a mixture of hydrophobic and hydrophilic residues inserts into a pocket formed by CMG2 α6-α7 loop and α7 helix. More than 600 Å^2 is buried by this insertion. According to computational structural comparison, the insertion and pockets are also found in TEM8. This implies similar mechanism in the way the two receptors bind to PA.

Although TEM8 and CMG2 are highly identical in sequence, their binding affinities for PA are not similar. Several residues around the vWA domain of both receptors are reported to be responsible for this difference. Residues on the MIDAS domain are highly conserved between TEM8 and CMG2. However, most of the residues on CMG2 contact interface with PA are not conserved TEM8. Mutational analysis proved that single point mutations on the non-conserved surface residues significantly change the binding affinity[29]. These residues are commonly referred to “hot spots”. The study of these

"hotspots" will provide valuable information for designing protein related agents that target the interaction surface of proteins.

Alanine scanning is the most widely used experimental method for identifying binding sites and "hotspots". In recent years, hot spot residues in a protein complex structure can also be predicted by several computational methods. Once the structure of protein is available, it is possible to predict the "hotspots" by computational method, and the efficiency of prediction will be much higher.

Variants of PA and TEM8/CMG2 can be widely used in anthrax or angiogenesis treatment. Modified anthrax toxins are being developed as potential anti-cancer agents [31,32]. PA-binding-affinity-elevated mutants of CMG2 or TEM8 could inhibit anthrax intoxication [33]. However, the exact mechanism of how they bind PA with higher affinity is still not reported. Mutational study of CMG2/TEM8 will find out the important residue(s) and/or conformation(s) that responsible for the binding interaction, hence explain the interaction type and figure out the way to make efficient mutations. PA binds to CMG2 more than 10-fold better than to TEM8. But to serve as a potential anti-cancer agent, the compound is required to bind tighter to TEM8. After knowing the "hotspots" and "hot" region(s), it is possible that PA mutants with additional specificity towards

CMG2 or “reversed” specificity towards TEM8 could be predicted by mutational study of PA. Additional PA variants that bind preferentially to either receptor are potential anti-angiogenesis therapeutic agents target normal tissues or tumor tissues. To obtain the desired models of mutants, I will make point mutations or group mutations on the residues of the interaction surface, both hotspots and non-hotspots. Mutations on hotspots or the highly conserved residues will possibly reduce the binding affinity, whereas, mutations on non-hotspots residues will likely increase the affinity. The binding mechanism of PA-TEM8 binding will be investigated in Chapter 3, followed by a comparison to PA-CMG2 binding in Chapter 4.

CHAPTER 2

EVALUATING THE PERFORMANCE OF MM/GBSA METHOD IN ESTIMATING PROTEIN LIGAND BINDING AFFINITY

ABSTRACT

This chapter evaluated the modeling, simulation and free energy calculation methods used the following chapters, especially the Molecular Mechanics/ Generalized Born Surface Area (MM/GBSA) method. This method is widely used to predict protein-ligand binding affinity, because it is statistically more rigorous than docking and computationally less expensive than FEP/TI. It gives satisfactory prediction results for some protein-ligand binding complexes. But for other systems, simulation can show miserably poor correlation to experiments. The choice of force field, dielectric constant and other simulation parameters can significantly affect the accuracy of the results. The validity of MM/GBSA heavily depends on the sampling method. Thus, the length of MD simulation and number of repeats also play an important role. The chapter evaluated the performance of MM/GBSA in predicting protein-ligand, protein-protein affinities and protein folding energies using several microseconds of MD simulation. I found an anomalous autocorrelation time in energies resulted from the fractal "memory kernel"

like behavior of proteins. The location of ligand molecules, flexibility of ligands and friction parameters have significant impact on the autocorrelation time in energies, thus affect sampling efficiency. Improper sampling method or insufficient equilibration will lead to under estimated errors in energies, further, irreproducible results. MM/GBSA can still serve as a powerful tool in free energy estimation, if the conditions and errors are properly evaluated. Considering all factors above, this chapter gives a solution to obtain statistically rigorous MM/GBSA results efficiently at predetermined computing hours.

2.1 Inroduction

Computer aided drug design (CADD) facilitates and accelerates the identification of new candidates in early state drug discovery. Once the candidates were discovered using high-throughput methods like structural based docking based on empirical scoring function, more accurate low-throughput computational methods are needed to optimize their chemical features. The calculation of protein-protein and protein-ligand interaction free energy is one of the most important protocol in these low-throughput methods. Molecular Mechanics/Poisson Boltzmann Surface Area[34] (MM/PBSA) and Molecular Mechanics/Generalized Born Surface Area[35] (MM/GBSA) are two of the commonly used methods to calculate free energy. Comparing to the statistical mechanically rigorous Free Energy Perturbation[36] (FEP) and Thermodynamics Integration[37] (TI) methods, MM/PBSA and MM/GBSA use reasonable assumptions to improve the computational efficiency. Furthermore, MM/PBSA and MM/GBSA allow for the rigorous and insightful free energy decomposition into interaction type and atom group components. The Linear Interaction Energy[38] (LIE) method is related to MM/PBSA and MM/GBSA, but it requires a large training set to fit the parameters.

There are in fact many approaches to estimate free energy. The main differences are the

force fields and the Generalized Born models applied. The energies can also be calculated from an individual energy-minimized conformation or from an ensemble of conformations extracted from MD/MC simulation trajectories or other sampling methods. Some energy that is known to deteriorate the method accuracy can be included or excluded of binding contributions. For all these methods, comparing to structural based scoring functions such as docking, the physics based methods can provide better correlation with experimental results and enable more detailed analysis [39-43].

However, the assumptions made in the MM/PBSA and MM/GBSA methods limit the accuracy and usage of these methods. For each individual system, the accuracy, efficiency and feasibility vary depending on the size, composition, and research aim. To investigate the proper application of MM/GBSA method, this chapter starts from a brief description of the theoretical backgrounds of sampling and free energy calculation methods, followed by a series of test to access the performance of MM/GBSA method on one of the Von Willebrand Factor A proteins.

2.2 Theoretical Background

2.2.1 Molecular Mechanics

The motion of atoms in a system can be described using Newtonian mechanics in Molecular Mechanics. A force field that composed by some functions and parameters is employed to describe the potential of the system. In most of the force fields for biomolecules, many-body effects are approximated using pair-wise potential, although some many-body effects such as polarization are also included[44,45]. The functions are parameters that were differed chiefly in the most common force fields such as CHARMM[46], Amber [47], OPLS[48], and Gromos[49], although they are based on different theories.

Different force fields treat atoms differently. All atoms are calculated in the modern versions of Amber, CHARMM, and OPLS, whereas the non-polar hydrogen atoms are merged into the carbon atoms in Gromos. Such lightly modified force fields were very popular in the 80's and early 90's when the computing power was not as efficient as today. Recent years, these coarse grained force fields are mainly used to simulate long time scale process such as large protein folding or domain motion.

The molecular potential of CHARMM force field is shown as follows

$$U_{CHARMM} = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\varphi(1 + \cos(n\varphi - \delta)) + \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_u(u - u_0)^2 + \sum_{nonbonded} \varepsilon\left[\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - \left(\frac{r_{0ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_i}{\varepsilon r_{ij}} \quad (2.1)$$

where the first five terms are bonded terms that describe the potential energy from bond stretching, angle bending, dihedral rotation, improper rotation and Urey-Bradley cross-term accounting for angle bending using 1, 3 nonbonded interactions. And the last two terms are nonbonded terms describes repulsion, dispersion, and electrostatics. The Urey-Bradley term is specifically introduced by CHARMM. It is a potential that couples bond and angles motions to improve the description of the torsions of the protein backbone[46,50].

The first five bond terms are a summation of simple harmonic terms, where the k_b, k_θ, k_φ, k_ω, and k_u are the force constant; b, θ, φ, ω and u are the actual value; and the corresponding symbols with "0" subscript are idea values or values at the equilibrium location.

The non-bonded terms contains a Lennard-Jones potential that describes repultion and

dispersion, and a Coulomb term that describes the electrostatics. These potential energies are calculated as pair-wise energies between all pairs of the atoms in the system, with a few exceptions. First, pairs of bonded atoms are excluded because these interactions are included in bond and angle terms. Second, pairs of atoms with three bonds in between (so-called 1-4 interactions) are scaled down because these interactions are partly described by the torsion term. Thirdly, pairs of atoms that have more than a certain cut-off distance between them are excluded or treated using a different function.

The Lennard-Jones potential can be expressed in many slightly different forms depending on the parameter exist in the force field. In all forms, the r^{-6} term describes the dispersion in a system which can in principle be derived from dispersion theory[51]; the r^{-12} term describes the repulsion which is chosen for computational convenience[52]. The repulsion term, physically, can be better described by an exponential term, but with decreased computational efficiency.

The Coulomb potential is described using the partial charges q_i, q_j of the two atoms i, j (Equation 2.1), where ε is the dielectric constant of the media. Comparing to Lennard-Jones potential, the electrostatic energy doesn't decay as fast. Plus, the electrostatic interactions are very important for correct energies and therefore they are

always required to be treated rather accurately even beyond the cut-off distance. In the practice of protein simulation using periodic boundary, the electrostatics are most commonly replaced by a Ewald summation that treats the long-range interaction in reciprocal space using fast Fourier transform[52,53].

2.2.2 Implicit Solvent Models

Implicit solvent models, by their name, treat the solvent molecules implicitly, and this is one of the most common methods to reduce the complexity of the system. The most popular approach is the treat the solvent as a dielectric continuum, which originates from the work of Max Born[54]. In this approach, there are two phases in the solvation process. First, it creates a cavity in the solvent that can tightly accommodate the solute. Second, it introduces the solute into the cavity and turns on the interactions between solute and solvent. The solvation free energy can be expressed by

$$G_{solv} = G_{cav} + E_{rep} + E_{disp} + E_{elec} \qquad (2.2)$$

where G_{cav} is the free energy of creating the cavity in the solvent, and the next three terms are the repulsion, dispersion, and electrostatics interaction energies. Other than the first term on the right hand side of Equation 2.2, the other terms on the right are written as energies rather than free energies. This stems from an assumption that the solvent structure is not perturbed by the introduction of solvent-solute interactions. The

assumption works decently for the repulsion and dispersion terms, but not usually for the electrostatic term[55]. In most continuum solvent models, a non-polar term, G_{np}, takes the first three terms on the right side of Equation 2.2; and a polar term. G_{pol}, takes the electrostatics, such that

$$G_{sol} = G_{np} + G_{pol} \tag{2.3}$$

The most popular method to estimate the non-polar solvation free energy is based on the solvent accessible surface area (SASA)[56]. It assumes a linear relationship between the loss in SASA and binding free energy.

$$G_{np} = \gamma SASA + b \tag{2.4}$$

where γ is an empirical parameter in the unit of surface tension, b is an empirical constant. The parameters are fitted to experimental solvation free energies of hydrocarbons[57].

The treatment of polar solvation energy starts from considering the Poisson equation for the solvent[58]

$$\nabla[\epsilon(r)\nabla\varphi(r)] = -4\pi\rho(r) \tag{2.5}$$

where ε is the dielectric constant, φ is the electrostatic potential, and ρ is the charge density. If ionic strength is considered, an extra Boltzmann factor can be added into Equation 2.5, and results in the Poisson-Boltzmann (PB) equation. This is not common in

biomolecular simulations[59], but approaches based on Equation 2.5 are still refered to PB methods. In most practices, the equation is solved using the finite difference method that distributes the charge distribution and dielectric constant on a grid.

The solvation free energy can be obtained from calculating the reaction field of the system. The reaction field, φ_{react}, is the difference in potentials under the dielectric constant of the actual solvent and the dielectric constant of vacuum, for which 80 and 1 are commonly selected to best reproduce experiment results. The reaction field can be further represented by a set of point charges as in a force field [60] by

$$G_{pol} = \frac{1}{2}\sum q_i \varphi_{reac}(r_i) \tag{2.6}$$

This equation can be reduced in the case of a single ion of radius *a* in pure solvent, which is called the Born formula

$$G_{Born} = -\frac{q^2}{2a}\left(1 - \frac{1}{\epsilon_{solv}}\right) \tag{2.7}$$

where ε_{solv} is the dielectric constant of the solvent.

The generalized Born (GB) model generalizes Equation 2.7 by mapping atoms to spheres with charges and radii. When the distance between two atoms is larger than their radii, Equation 2.6 can be derived into a summation of individual Born terms and pair-wise

Coulomb terms[61].

$$G_{pol} = \sum \frac{q_i^2}{2a_i}\left(\frac{1}{\epsilon_{solv}} - 1\right) + \frac{1}{2}\sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}}\left(\frac{1}{\epsilon_{solv}} - 1\right) \quad (2.8)$$

If we use symbol f_{GB} as an approximation function of the distance between atoms and their effective Born radii, Equation 2.7 can be further reduced to

$$G_{pol} = \frac{1}{2}\left(\frac{1}{\epsilon_{solv}} - 1\right)\sum\sum \frac{q_i q_j}{f_{GB}} \quad (2.9)$$

Among the distinct flavors of GB methods, f_{GB} is approximated following difference theories[61].

2.2.3 Molecular Dynamics Sampling Method

Molecular dynamics employs Newton's second law[62] to iterate the system over time. Each particle *i* in the system has a mass of m_i and a coordinate q_i. The force on particle *i* can be expressed as

$$F_i = m_i \frac{d^2 q_i}{dt^2} \quad (2.10)$$

The force can be calculated from the gradient of force field (potential) using

$$F_i = \frac{dU}{dq_i} \quad (2.11)$$

The coordinate after a finite time step, Δt, can be calculated using a simple Taylor expansion

$$q(t + \Delta t) = q(t) + \frac{dq(t)}{dt}\Delta t + \frac{d^2 q(t)}{dt^2}\frac{\Delta t^2}{2} + \cdots \quad (2.12)$$

At time t, if the coordinate q(t), the velocity dq(t)/dt, and the acceleration $d^2q(t)/dt^2$ (Equation 2.10) are known, the system can be propagated to the next time step $t+\Delta t$. The higher order terms are approximated or omitted.

At the beginning of a Molecular Dynamics simulation, the coordinates and velocities are initialized for all particles in the system. For each consecutive step afterwards, the forces, velocities, and coordinates are updated for each particle based on the force field. There exist numerous algorithms for updating these parameters, such as simple Verlet[63], leapfrog[64], and velocity Verlet algorithms[65].

The time step of iteration should be small enough to cover all the motions in the system, e.g. the fastest motion in biomolecular system is the vibration period of the bonds involve lightest weight atom, hydrogen, which is ~10fs. As a rule of thumb, the time step should be about a tenth of the period of the fastest motion, 1fs. However, the vibration of hydrogen involved bonds is not always valuable to the research, and can sometimes be omitted to increase simulation efficiency. SHAKE[66] is one of algorithms that perform a restraint on bonds involving hydrogen atoms. Once these bonds are restrained, the time step can usually be increased from 1fs to 2fs, to increase the simulation efficiency.

An ensemble of conformations can be collected from an MD simulation. These conformations have different microscopic states, but belong to the same macroscopic or thermodynamic state. In this dissertation, all MD simulations were performed to compare with experimental results at constant pressure and temperature. To keep these parameters comparable in this dissertation, simulations were conducted using constant number of particles, volume, and pressure. Such systems are known as isobaric-isothermal ensembles.

To keep the temperature constant in Molecular Dynamics simulation, an thermostat algorithm[67] is used. All simulations in this dissertation employed a Langevin thermostat[68]. Thus, Langevin dynamics were performed rather than Newtonian dynamics. Comparing to Newtonian dynamics, a velocity related friction term τdq/dt and a random force term R(t) are introduced.

$$m_i \frac{d^2 q_i}{dt^2} = F_i - \tau \frac{dq}{dt} + R(t) \qquad (2.13)$$

The Langevin piston[69] method is used as a barostat to maintain the constant pressure.

In protein binding research, the system can sometimes be very large, but the interesting chemistry only occurs in a small region. A well designed truncation of the protein can increase the simulation efficiency without losing the dynamic details (see 2.4.2).

2.2.4 Sampling Strategy

One most important property of good simulation is being able to reconstruct the prescribed distribution, which, in this dissertation, is a Boltzmann distribution for canonical ensemble[70]. For some quantity A, the average of A can be calculated by averaging the N samples extracted from the simulation as

$$\langle \mathrm{A} \rangle = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=0}^{\mathrm{N}} \mathrm{A_i} \tag{2.14}$$

The precision of this average can be described using the variance. The variance is only valid when the N samples are independent of each other[71]. One approach to measure the independency of the samples is to use the autocorrelation function (ACF). A correlation time τ can be calculated from ACF and be further used to measure the simulation time takes to eliminate the correlation to previous conformations. Once τ is known, the interval between consecutive samples can be determined. Statistical inefficiency can also be estimated based on a given τ[72,73].

In practice, the dynamics of proteins consists a property similar to memory kernel[74,75]. The autocorrelation time is sensitive to protein properties such as location of domains, and the location of the binding site. Therefore, an accurate estimation of τ can be very

challenging. An alternative and effective solution to obtain independent samples is to conduct M independent repeat calculations for a system. The average of property A can be calculated by

$$\langle A \rangle = \frac{1}{M} \sum_{repeats} \frac{1}{N'} \sum_{i=0}^{N'} A_i \qquad (2.15)$$

where N' is the number of samples in each repeat. In ideal perfect sample from infinitely long simulation, the ensemble average of property A calculated from Equation 2.14 and 2.15 should be identical. However, the precisions of A calculated from Equation 2.14 and 2.15 are not identical. The formal one is inversely proportional to$\sqrt{N}$, whereas the latter on is inversely proportional to$\sqrt{M}$.

Independent repeats of simulations are commonly generated by simply assigning different starting velocities to start the simulation[76-78]. Since none of the coordinates of initial conformation are changed, the perturbation to the system can be considered as fairly small.

2.2.5 MM/PBSA and MM/GBSA methods

In MM/GBSA and MM/PBSA methods, the binding free energy between the ligand and the receptor can be calculated using the following equation:

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (2.16)$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{elec} + \Delta E_{vdw} \quad (2.17)$$

$$\Delta G_{sol} = \Delta G_{GB} + \Delta G_{SA} \quad (2.18)$$

$$\Delta S = \Delta S_{trans} + \Delta S_{rot} + \Delta S_{vib} + \Delta S_{conf} \quad (2.19)$$

where ΔG_{bind} is the total binding free energy, E_{MM} is the Molecular Mechanics energy, ΔG_{sol} is the solvation free energy, E_{int} is the internal strength energy, E_{elec} is the Coulombic interaction energy, E_{vDW} is the van der Waals interaction energy, G_{GB} is the generalized Born free energy, E_{SASA} is the solvent accessible surface area interaction energy, S_{trans}, S_{rot}, S_{vib} and S_{conf} are the translocational, rotational, vibrational and conformational entropy, respectively.

One of the main advantages of the MM/GBSA method is that the binding contributions are calculated separately. The separation enables a more elaborated examination of terms that improve or deteriorate the correlation with experiments. Another advantage is that the binding free energy contribution of each individual residue can be decomposed with little additional computational resources. The decomposition allows the identification of binding "hot-spots" at a low cost. However, the method and its implementation are not perfect. The following paragraphs will explain their strengths and weaknesses.

The protein–ligand van der Waals interaction term (E_{vdW}) generally dominates the binding energy differences. Usually, the E_{vdW} term has the best correlation comparing to experimentally measured results.

Force fields sometimes tend to overestimate energy barriers, because the deformation of proteins dislocated the conformation from local minima and yield a non-converged energy result. As a result, the internal strength energy (ΔE_{int}) values tend to be higher (more positive) than the results from a quantum-mechanical method. When using single trajectory method, the internal strength energy will be canceled, and the impact from this problem should be minimized.

The protein–ligand electrostatic interaction term (E_{elec}) is essential but to some degree problematic. The problems are mainly caused by the underestimated screening terms. The use of a fixed protein dielectric constant of 1 and a fixed charged force field omit the motion of proteins and may underestimate the electrostatic screening terms. Adding Generalized Born screening terms from the solvent alleviates this problem. However, the GB term is not always included in the scoring function and could bring extra noise. Hence, extreme caution should be used when calculating protein-ligand interactions

involve hydrogen bonds. The favorable scores for these interactions may result from insufficient Δ*Gsolv* screening term.

Eelec is not always beneficial to the correlation with experimental results. It depends on the composition of the congeneric series being scored. Especially when only a few of the ligands in the series can form hydrogen bonds with the protein, the correlation to experimental data can be negatively affected.

A simple energy minimization for the complexes greatly increases computational efficiency comparing to MD simulations. It can be completed quickly that can be fitted into synthetic chemistry-biological test cycles. However, the protein could not be sufficiently relaxed to accommodate numerous scaffolds in such method, and could in theory raise a serious limitation in free energy calculation. This limitation should be minimized when applied on a congeneris series. On the other hand, the introduction of noise in short MD simulation trajectory sometimes produces even worse correlation to the experimental data.

The electrostatic term E_{elec} for the protein is also affected by the insufficient electrostatic screening terms. This can be more severe when salt bridges within the protein were

disturbed by ligands at different degrees. In consequence, the E_{elec} term can be very noisy for the protein binding to different ligands.

In practice, some entropy contributions for the binding process are ignored for the sake of efficiency. Usually, the changes in translational, rotational and vibrational terms are omitted. When scoring a congeneris series, adding such terms for ligands has little to no effect to the ranking results, especially where the entropy contributions were estimated using rigid rotor harmonic oscillator models. For the same series, the changes in such entropy contributions are believed to be relatively ignorable.

In addition, the entropy term resulting from the restriction of torsional angle in the binding process is also disregarded[79]. The restriction of torsional motions can possibly result in a significant impact to rank ordering, especially in the entropic terms. However, the calculation of torsional terms is extremely expensive and not practical under current technology.

Although the MM/GBSA method can give decent correlation and accurate rank ordering with experimental results, the result show larger dynamic range comparing to the experimental range, typically 3-10 times larger. In the next section, I will build a test

system using protein systems to examine the performance of MM/GBSA method.

2.3. Testing method and system

To validate the sampling and free energy calculation method, a series of studies were performed before conducting the production simulation. These tests were built to discover a good way to perform efficient simulation and obtain statistically converged free energy results.

2.3.1 Construct initial structure

The method studies were conducted on similar systems as the systems in Chapters 3 and 4 to minimize the impact from replacing system components. We built the structure of PA bound to 13 different mutants of TEM8, which are H57A, H57H58NN, H58A, T87S, R88Q, K111R, L113V, D117E, Y119F, E125A, E152A, E152K, E155K, and wide type.

The structures were built in three steps. First, a structure of PA bound to TEM8 was obtained by using the TOPMATCH[80] protein structural precise alignment algorithm. A rigid body structural alignment was applied to replace CMG2 (chain Y in PDB: 1T6B[25]) using TEM8 (chain A in PDB: 3N2N[29]). Missing loops in PA were patched using the

optimized conformation generated by Modeller9.11[81]. Second, PA domain I and III were truncated in the simulation to improve efficiency, because these two domains are not believed to directly contribute to the binding with TEM8[82]. To keep the truncated PA folded, a harmonic restraint were applied on the residues more than 14 Å away from the binding surface in MD simulation. Different strengths of harmonic restraint were tested to obtain minimized potential artificial impact from harmonic restraint. At last, mutations were made on TEM8 using the mutation module in Modeller9.11[81]. To obtain an optimized mutant structure, 1000 steps of energy minimization iteration were conducted in the mutation module.

Some crystallographic water molecules are believed to be important in the protein's biological processes. To investigate the importance of crystallographic water molecules in our research system, three test cases were investigated. The test cases were built based on the structure of PA bound to wild type TEM8. The first one keeps all crystallographic water molecules as the control group. The second one removes all crystallographic water molecules in the structure obtained from protein data bank. The third one only keeps the crystallographic water molecules in the binding site; in this case, the MIDAS domain in TEM8.

To remove energy clashes, the initial structure of PA-TEM8 complex were subjected to an energy minimization in vacuum. The energy minimization has 10 cycles of 1,000 steps, reducing the harmonic restraint each cycle on all protein atoms from 10 kcal·mol^{-1}·Å^{-2} to 1 kcal·mol^{-1}·Å^{-2} in decrements of 1 kcal·mol^{-1}·Å^{-2} under the Steepest Descent method in the CHARMM 35b6 software package[83] and charmm27 force field parameters[46]. The protein complex was then dissolved in a TIP3[84] water and 0.15 mol·L^{-1} NaCl box of 126Å×86Å ×86Å at a mixed density of 0.947 g/cm^3. Extra chloride anions were used to neutralize the positive charge of the protein complex.

2.3.2 Energy minimization

The energy minimizations of the protein solvent were conducted using NAMD 2.10-GPU software package[85] with charmm27 force field. Two cycles of 10000 steps of Conjugated Gradient minimization were run NAMD with fixed protein atoms and without constraints, respectively.

2.3.3 Molecular dynamics

Molecular dynamics simulations were conducted using NAMD 2.10-GPU software package[85] with charmm27 force field at constant temperature at 300 K with a damping

coefficient of $5ps^{-1}$ coupled by Langevin dynamics[86], and constant pressure at 1 bar coupled by Langevin-Noose Hoover piston[87] with a damping time of $50ps^{-1}$. Particle Mesh Ewald[88] was used to calculate long-range electrostatic interactions. The time step was set at 2fs with the use of Rigid Bond (SHAKE) algorithm[89] to restrain hydrogen atoms with the heavy atoms they bound to.

2.3.4 Free energy calculation

The free energy calculations were conducted using CHARMM c35b6. Molecular mechanics energies were calculated using a 999 Å cutoff distance (essentially no cutoff for this system). Non polar solvation energy term $G_{nonpolar}$ was estimated from the solvent accessible surface area (Equation 2.4) using SASA package in CHARMM[90], and γ=0.00542kcal•mol^{-1}•Å^{-2}, β=0.92kcal/mol. The polar solvation energy G_{GB} was calculated with the GBSW approach[91,92] implemented in CHARMM. Dielectric constants of 4 and 80 were used for solute and solvent, respectively. CHARMM default optimized parameters for GB-calculations[93,94] were used. For each PA-TEM8 complex, the binding free energy of MM/GBSA was estimated as follows: $\Delta G_{bind} = G_{complex} - G_{PA} - G_{TEM8}$ where ΔG_{bind} is the binding free energy and $G_{complex}$, G_{PA}, and G_{TEM8} are the free energies of complex, PA, and TEM8, respectively. The binding free energy can be further decomposed into contribution from each individual amino acid. A detailed procedure is

shown is section 3.2.5.

2.4. Result and discussion

2.4.1 The importance of crystallographic water molecules

The crystallographic water molecules are important to maintain a stable conformation, especially those in the binding site. In the control group, where none of the crystallographic water molecules were removed, the structure of protein complex stayed folded after 20ns of MD simulation with an RMSD of 1.92Å, the metal ion in MIDAS binding site shifted 0.25Å from its initial position. In the second case, where all of the crystallographic water molecules were removed, the structure of protein complex also stayed folded after 20ns of MD simulation, but with a larger RMSD at 2.10Å. However, the metal ion in MIDAS binding site shifted 2.21Å from its initial position, which breaking the coordinate bonds in the binding site that are believed to be important to the binding process[9]. In the third system, all crystallographic water molecules were removed, except those in the binding site. As expected, the structure of protein complex stayed folded after 20ns of MD simulation, with a larger RMSD at 1.96Å, and the metal ion in the binding site shifted 0.23Å from its initial position. Although our result may imply that removing those crystallographic water molecules that are not in the binding site would

not impact the results, the deletion may still be dangerous to the accuracy of the results and should be done with extreme caution.

Most crystallographic water molecules are not known to be involved in the binding process. But it is sometimes non trivial to determine if a water molecule is directly or indirectly facilitating the binding process[78]. In some cases, the water molecules are required to be removed to obtain more accurate structure for the pure protein, e.g. protein docking. When simulate proteins in non-water solvents, the decision can be even more difficult to make[95]. We suggest keeping all crystallographic water molecules in the initial structure for a long MD simulation. Thus the fluctuation on structure can be minimized, and the chance of obtaining invalid ensemble average ad variance obtained from the MD simulation can also be minimized.

2.4.2 Harmonics restraint on residues away from the contact surface

To optimize the method of truncation, we compared a series of harmonic restraint parameters to maintain a stable conformation. Since PA domain I and III are known to be less important to the binding process[82], plus they are more than 20Å away from the binding surface, they were truncated in the simulation. When no restraint is applied (black line in Figure 2.1), the truncation destabilized the part of the protein complex that

is close to the truncation site. To prevent unwanted unfolding on the proteins parts far from the binding site, a harmonic restraint were applied on the residues more than 14 Å away from the binding surface in MD simulation. The strength of harmonic restraint was tuned between 0-10kcal•mol^{-1}Å^{-1} in attempt to find the best strength.

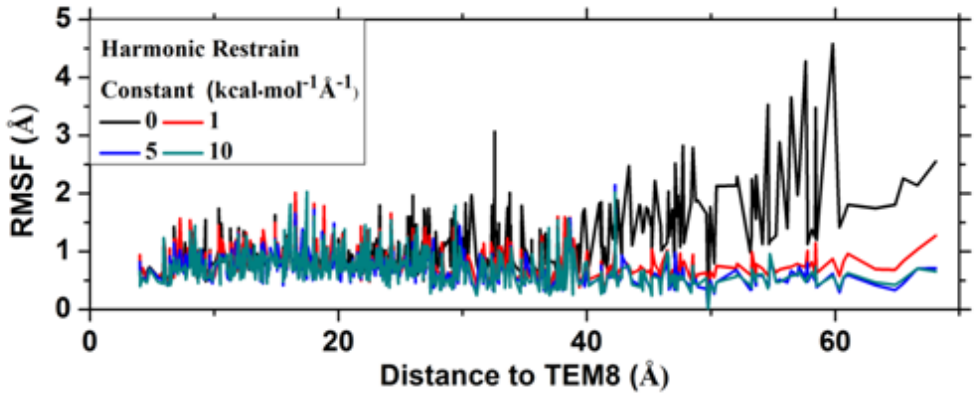


Figure 2.1 Molecular Dynamics simulation stability and equilibration. RMSF of PA residues in 200ps MD simulation versus the shortest distance between the residue and TEM8.

Harmonic restraint of 5, or 10 kcal•mol^{-1}Å^{-1} yielded slightly lower RMSF results than restraint of 1kcal•mol^{-1}Å^{-1}on the residues far away from the binding site (Figure 2.1). Considering the thermal noise at 300K, k_BT~0.6 kcal•mol^{-1}Å^{-1}, a harmonic restraint of 1 kcal•mol^{-1}Å^{-1} should be enough to limit the random drift below 1 Å. 5 or 10kcal•mol^{-1}Å^{-1} of harmonic restraint did not provide significant improvement comparing to the decrease in RMSF when the strength of restraint increase from 0 to 1. Although the stronger

restraint we tested did not interfere the dynamics of residues close to the binding side (measured by RMSF), we picked the lowest restraint that prevented unwanted structural changes resulting from truncation, which is 1kcal•mol^{-1}Å^{-1}.

2.4.3 Anomalous autocorrelation times and memory kernels

To obtain independent samples for property A (Equation 2.15) from MD simulation, we needed to examine the autocorrelation of a series of A_i. The purpose of this dissertation is to calculate the binding free energy; thus the autocorrelation of MM/GBSA interaction free energy between PA and wild type TEM8 is evaluated and shown in Figure 2.2.

Surprisingly, a similar trend of autocorrelation was observed for simulation of 20ns (Figure 2.2a) and 0.2ns (Figure 2.2b). In both cases, the autocorrelation started from 0.3, decreased to -0.3 after ~300 samples and fluctuated. Noticing that the samples in the two cases were extracted at 5ns and 0.05ns time intervals, this autocorrelation result implies the existence of a fractal "memory kernel" like behavior in protein dynamics. A fractal behavior means similar or identical behavior can be observed on different scales of the system. It is usually called a "memory kernel" in MD simulations[75,96,97]. In experiments, the memory kernel is also observed in protein dynamics, but the biological function for such phenomena is still unclear[98].

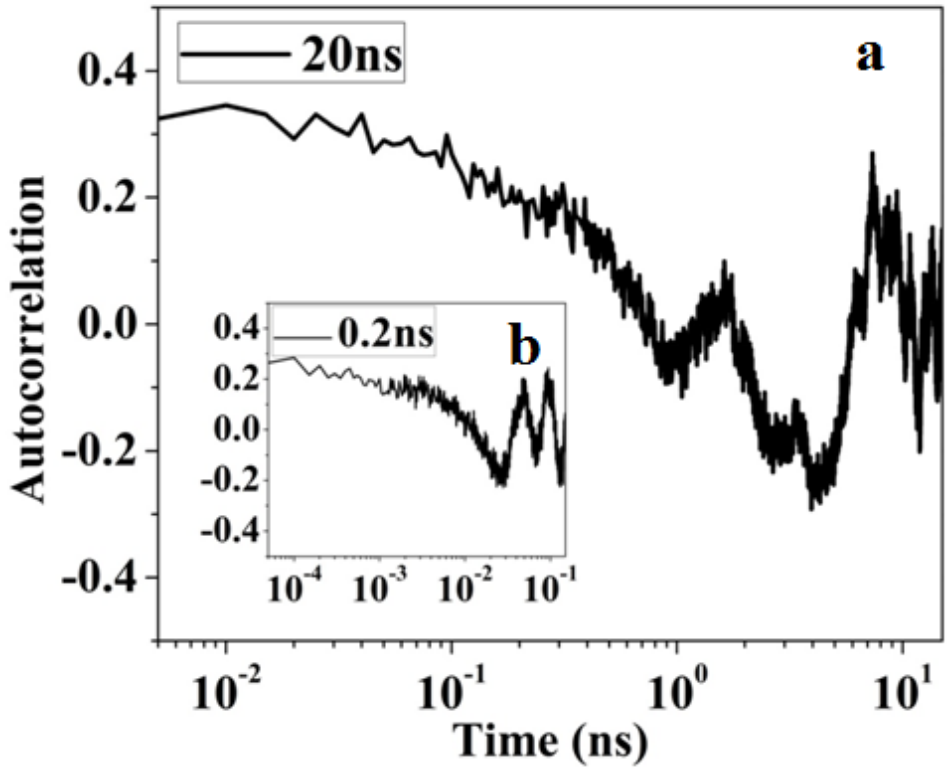


Figure 2.2 Autocorrelation of MM/GBSA interaction energy in 20ns (a) and 0.2ns (b) Protein-Protein interaction, using window sizes of 5ns and 0.05ns, respectively. Each figure has 4000 samples that were both collected after 40ns equilibration run.

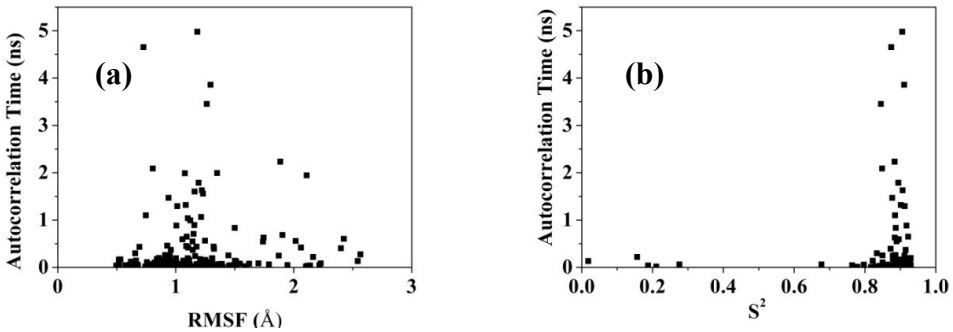


Figure 2.3 Autocorrelation time of MM/GBSA interaction energy between protein A and individual residues of protein B that has varies RMSF (a) or S^2 (b) extracted from 20ns simulation using a window size of 5ns.

To further investigate the factors affect the autocorrelation time, we measured the autocorrelation time of binding free energy contributed by individual residues on TEM8, and compared with RMSF and an NMR property, S^2. Because none of the TEM8 residues are under any restraint and these residues are more than 14 Å from the binding site, the dynamics of these residues are considered unaffected by the restraint applied on the residues near the truncation site. Figure 2.3 shows that there is not an obvious correlation between energy autocorrelation time and RMSF or S^2. NMR experiments showed that the relaxation times of the residues are also independent from RMSF or S^2 [96-98]. Although relaxation time is not a direct indicator of energy autocorrelation time, the dynamics information given by relaxation time is usually correlated with the autocorrelation time of thermodynamics properties[96,99]. Weak correlation between energy autocorrelation time and dynamics properties suggests that the sampling interval of property A cannot be

determined by measuring the dynamics properties of the system.

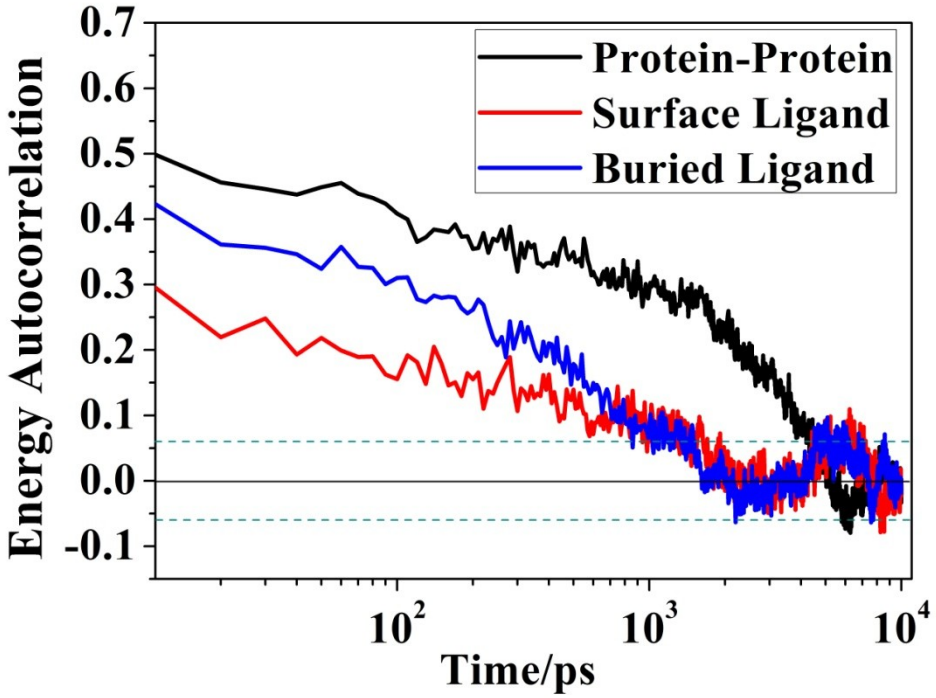


Figure 2.4 Autocorrelation time for MM/GBSA interaction energy in Protein-Protein interaction and protein-ligand interactions when the ligand is on protein surface or buried, using a window size of 5ns. Green dot lines are results from a Runs test at confidential interval of 95%. Samples have autocorrelations between the green dot-lines can be statistically considered as independent.

Autocorrelation time in binding free energy is also related to the size and location of ligand. We compare the MM/GBSA interaction energy autocorrelation time of PA-TEM8 complex (protein-protein), PDB: 4MRS with a buried small molecule ligand[100], and PDB:

4MQT with a small molecule ligand bind to the surface[101] (Figure 2.5). The autocorrelation protein-protein interaction energy starts higher (0.5) than that of the other two protein-ligand interactions (0.42 and 0.29). Its autocorrelation time (the time required to reach the green dot lines) is ~5 times longer than the other two. One possible reason is that the protein-protein binding has larger contact surface area and involves more residues, and more and larger memory kernels can be possibly formed, resulting in longer autocorrelation time. Second, small molecules are not usually believed to have the memory kernels or fractal behavior that were found in large protein molecules. The protein-ligand interaction can be considered as a fractal object binding to a rigid object, of which the autocorrelation is only contributed by the fractal object. But the protein-protein interaction, theoretically, is the interaction between two fractal objects, of which the autocorrelation is contributed by both fractal objects. The increased source of autocorrelation leads to longer autocorrelation time and higher starting value. Third, there is no clear evidence show that the location of the ligand affects autocorrelation time, but it affects the starting value of the autocorrelation. The relationship between ligand location and the autocorrelation time will need further investigation.

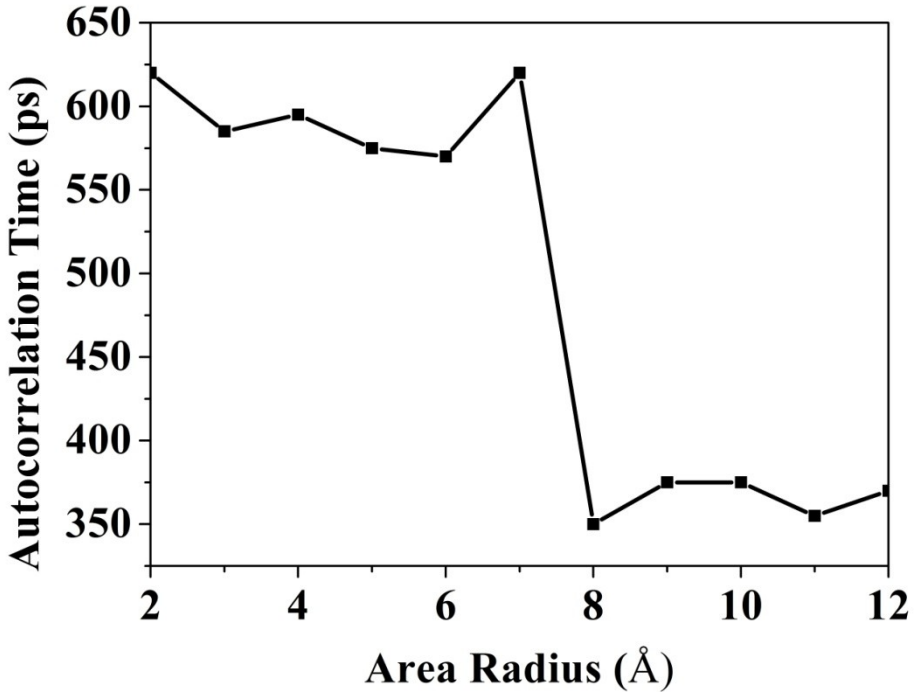


Figure 2.5 Autocorrelation time of MM/GBSA interaction energy between selected residues on TEM8 and PA, versus the radius of contact area centered to L113 residue on TEM8. Autocorrelation time is measured by the first time the autocorrelation reached zero. A window size of 1ns is used.

The correlation between autocorrelation time and the size of the contact surface, however, is not always positive. The auto correlation time in MM/GBSA interaction energy becomes shorter when the radius of selected area became larger and number of selected residue becomes larger (Figure 2.4). L113 residue locates at about the center of the contact surface between TEM8 and PA. When the selected area becomes larger, the interaction energies of more residues are included. The scale of fluctuation in energy will

be changed from a smaller binding site to a larger binding site. Thus, the fractal factor is also changed from a smaller one to a bigger one. When using certain observation windows to calculate autocorrelation time, a change of autocorrelation time should be expected if the fractal factor is changed. The change can be either increment or decrement[71]. Although L113 was randomly selected in this case, the sharp change in energy autocorrelation time may indicate the edge of a fractal group that share similar dynamics property. Thus we may measure the autocorrelation time of residue groups to determine the function group in protein complexes.

2.4.4 Langevin damping coefficient and autocorrelation

Motions in proteins can be approximately simplified to a damped harmonic oscillator system[102-104]. The complex can be simplified to a two rigid balls connected by a spring and merged in a viscous solvent. Each ball represents one protein in the complex, and the spring mimics the interaction between the proteins (Figure 2.6). Once an excitation is given by either the collision of Brownian motion or external excitation, the ball-spring model starts to vibrate. When the viscosity of the solvent is very high, the damping strength is also very high; the system needs a long time to recover the equilibration state. When the viscosity is very low, the vibration takes a long time to decay. Both case lead to long autocorrelation time. At some critical damping strength, the system needs the

shortest time to fully absorb the excitation, and yields the shortest autocorrelation time. In Langevin dynamics, the viscosity of the solvent can be modified by tuning τ in Equation 2.13. Good sampling methods obtains more independent sample with given resource. Sampling at the critical damping strength can decrease the autocorrelation time, thus obtain more the samples per unit time.

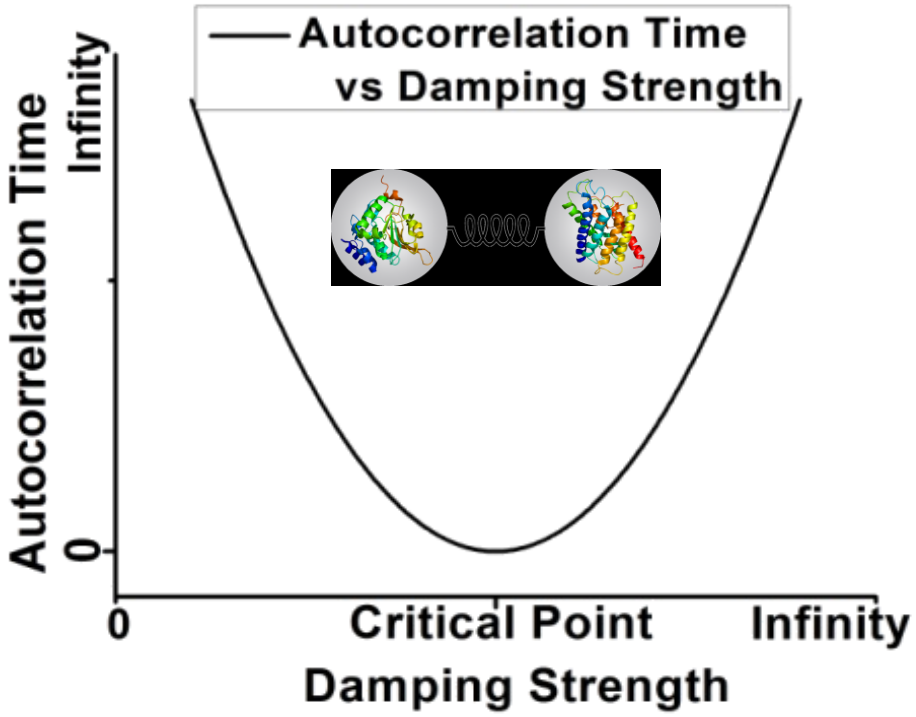


Figure 2.6 A scheme of relationship between autocorrelation time and damping strength in a damped harmonic oscillator system.

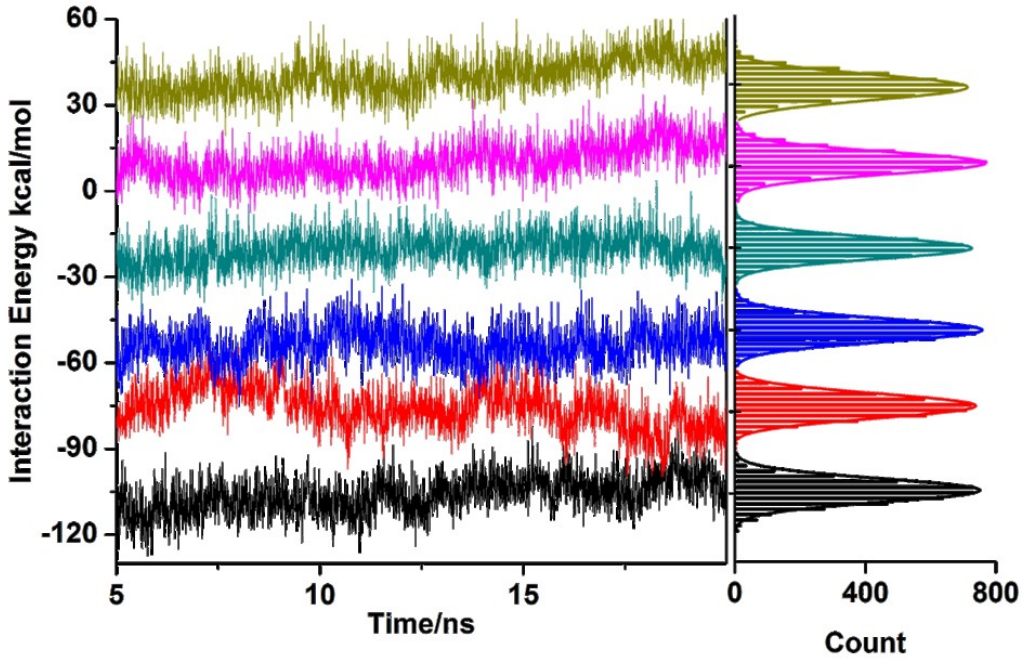


Figure 2.7 MM/GBSA interaction energy of PA-TEM8(WT) complex under Langevin damping coefficient 1 ps^{-1} (black), 5 ps^{-1} (red, offset by 30 kcal/mol), 10ps^{-1} (blue, offset by 60 kcal/mol), 50 ps^{-1} (dark cyan, offset by 90 kcal/mol),100 ps^{-1} (pink, offset by 120 kcal/mol), 200 ps^{-1} (dark yellow, offset by 150 kcal/mol). On the right, histograms show the distribution of the energy data with Gaussian fitting overlaid.

Although the motions are known to have damped harmonic oscillator like behavior, there is no direct evidence that the energies also have similar behaviors. To investigate the existence of these behaviors, we ran MD simulation for our protein-protein complex using varies damping Langevin damping coefficient ranging from 1 ps^{-1} to 200 ps^{-1}. MM/GBSA interaction free energy between PA and wild type TEM8 obtained from all

these simulations follow Gaussian distribution, which indicates that there is not significant bias in the MD sampling processes (Figure 2.6). When damping coefficient 10 ps^{-1} and 50 ps^{-1} were used, the autocorrelation in MM/GBSA interaction energy decays significantly faster than those using other damping coefficients. The autocorrelation times are also ~10 times shorter than the other cases. Energies calculated using damping coefficient 1 ps^{-1}, 5 ps^{-1}, 100 ps^{-1}, and 200 ps^{-1} have similar autocorrelation time. If we compare the results with the theory shown in Figure 2.6, the critical damping coefficient should be some value between 10 ps^{-1} and 50 ps^{-1}.

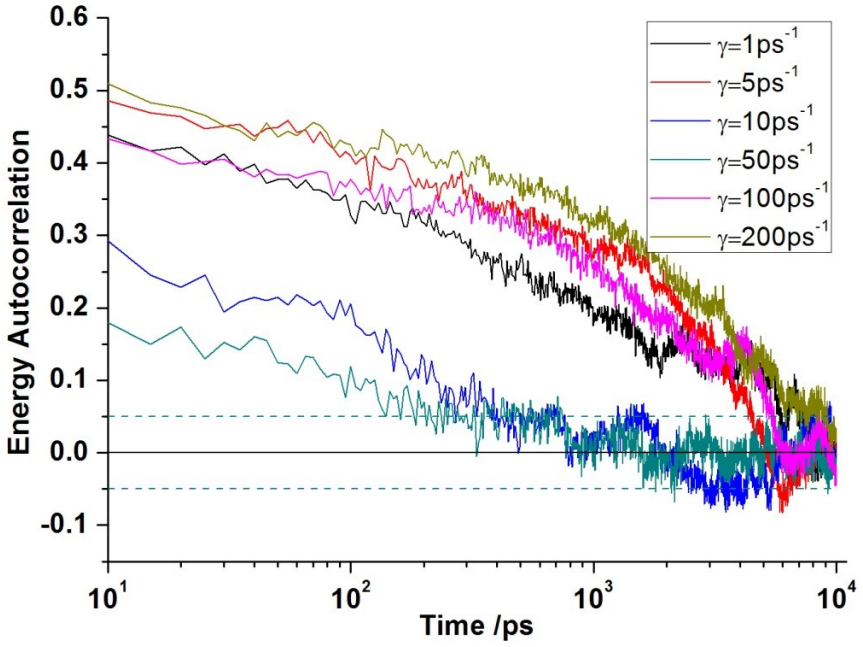


Figure 2.8 Autocorrelation of MM/GBSA interaction free energy PA-TEM8(WT) complex exacted from 15 ns MD simulations under different Langevin damping coefficient γ. Green dot lines are results from a Runs

test at confidential interval of 95%. Samples have autocorrelation within the green dot lines is calculated can be statistically considered as independent.

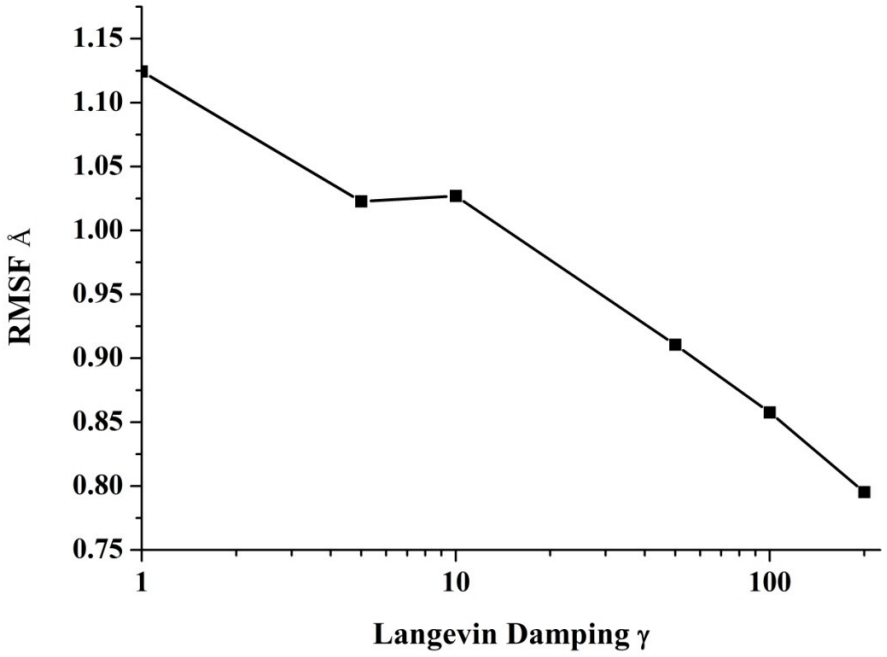


Figure 2.9 Average RMSF of all Cα atoms on TEM8 during 20ns MD simulation of PA-TEM8 complex using Langevin Damping coefficient from 1 ps^{-1} to 200 ps^{-1}.

Changing Langevin damping coefficients also brings side effects. Damping coefficients of 1 ps^{-1} to 5 ps^{-1} are commonly used to recover the dynamics of water molecules[86,105]. The change in viscosity in solvent have been shown to inhibit the local dynamics of protein, and enhanced the sampling of local conformation[106]. However, in our simulation,

the damping coefficients were modified for both protein and solvent. We found the increased Langevin Damping coefficient lowered the RMSF of residues in MD simulation, which may limit the conformation space visited in given simulation time, thus decrease the sampling efficiency.

In all, the MM/GBSA interaction free energy in protein-protein complex has similar behaviors as damped harmonic oscillator. A critical damping coefficient that can yield the shortest autocorrelation time exists. However, the side effected comes with the advantages hasn't been compared with experimental data yet. The use of damping Langevin damping coefficient needs to be further investigated. Considering the benefits and risks, the simulations in the following chapters were conducted using $5ps^{-1}$ Langevin damping coefficient.

2.4.5 Equilibration and production time

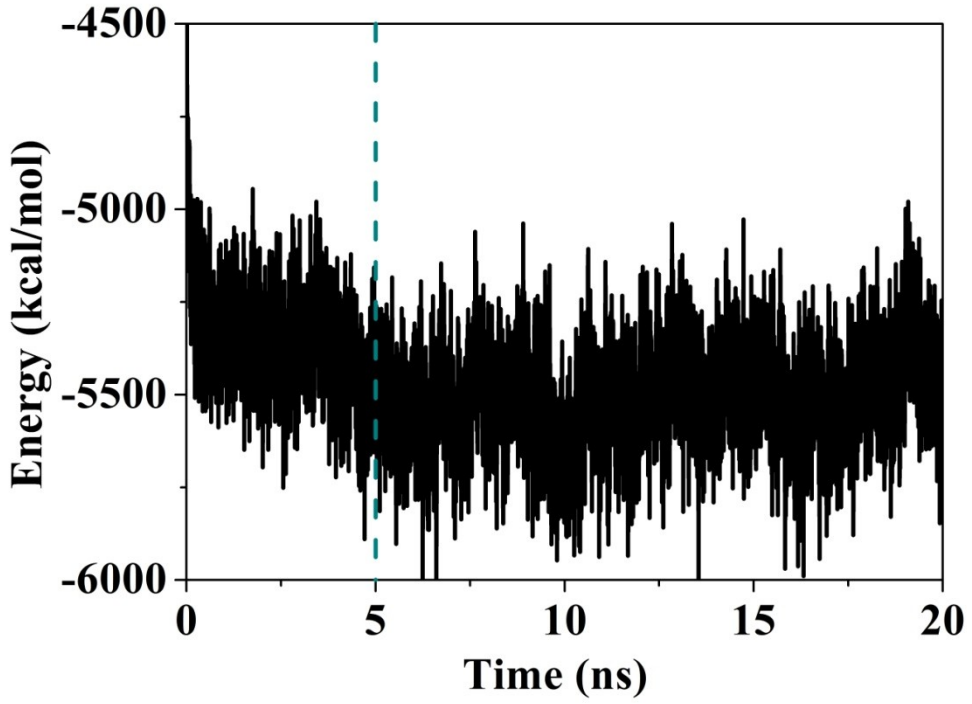


Figure 2.10. Solvation energy (van der Waals and Coulombic interaction energy between proteins and solvent) in 20ns MD simulation. The green dot line is marked at 5ns as visual aid.

The initial part of a Molecular Dynamics trajectory is usually considered as equilibration phase, and not used for further analysis[40,107,108]. There exists numerous of methods to set the length of equilibration phase, such as by measuring the RMSD[41], total energy[109], or imperial[110,111], etc . Because solvation energy is one major component of MM/GBSA binding free energy, this dissertation uses the interaction energy between the protein complex and water as an indicator of "equilibration", at which the layers of solvent

shells[84] have been formed. Figure 2.10 shows that the solvation energy of the PA-TEM8 complex decreased in the first 5ns and then fluctuated around -5500kcal/mol from 5ns to 20ns. Thus, simulation after 5ns is assumed to be "equilibrated".

If we consider the assignment of initial velocity as an external excitation to the protein complex, the oscillators started to vibrate following the external energy source. It is similar to observations in protein far infrared excitation experiments[112,113], where the vibration resulted from excitation lasted more than 30 periods. To reveal the vibration modes of the protein complex, the influence from artificial excitation needs to be minimized. The memory-kernel-like behaviors are usually more prevalent in large biomolecules rather than small inorganic molecules. Thus, a longer equilibration phase may be required for protein-protein complexes than for small molecules or protein-small-molecule complexes.

To test our hypodissertation, we ran 10 repeats of MD simulations to measure the standard deviation in energies calculated from these 10 repeats using different equilibration lengths and production length. Longer equilibration times and longer production times indeed yielded smaller standard deviations in the results, indicating higher precision. Unlike the commonly used 10-100ps equilibration time in protein-small

molecule binding[40,107,114], equilibration times shorter than 1ns in this protein-protein binding may still give low precision even the production time is longer than 8 n. If we consider running simulations at a predetermined computing resource, the equilibration time and production time add up to a given length. Imagine there exists a line $x + y = t$ in Figure 2.11, where x is equilibration time, y is production time, and t is the given simulation length. When simulation time is longer than 10ns ($x + y > 10$), the simulation precision will benefit from the extension of production time, especially when more than 5ns equilibration time is used.

For these experiments, we chose to use 5ns equilibration time out of the 20ns MD simulation to obtain MM/GBSA interaction free energy at the best possible precision. Only the other 15ns simulation was used for free energy calculation.

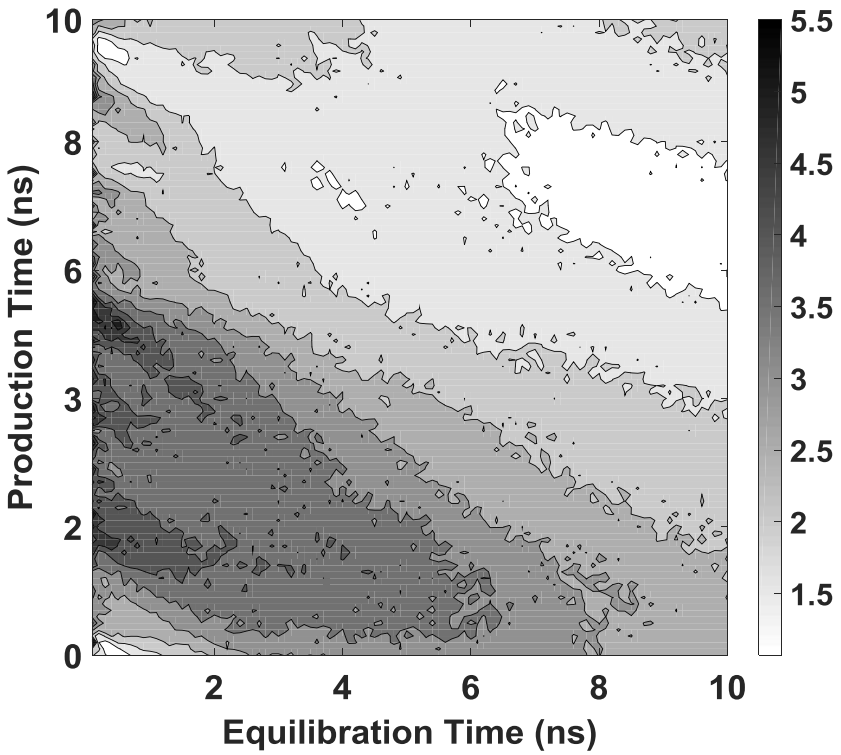


Figure 2.11 Standard deviation of the average interaction energy between two proteins extracted from 10 repeats of 20ns simulation vs length of equilibration time and production time.

2.4.6 The accuracy of MM/GBSA

After all simulation parameters were selected, we calculated the MM/GBSA interaction free energy between PA and 13 TEM8 mutants from 10 repeats of 20ns MD simulation, and compared the results with experimentally measured binding affinity. The experimental data were dissociation constants collected by our collaborators using Surface Plasmon Resonance experiments. The calculated binding free energy has a

correlation of 0.57 with experimental values. With 10 repeats of independent simulation, the standard error in calculated free energy is ~1kcal/mol. 20ns simulation yielded satisfactory results in estimating the binding free energy of protein-protein interactions. However, the absolute calculated energies are ~10 times exaggerated compared to the experimental values, suggesting that large conformation changse and unfolding caused by mutations may not be recovered in the simulation.

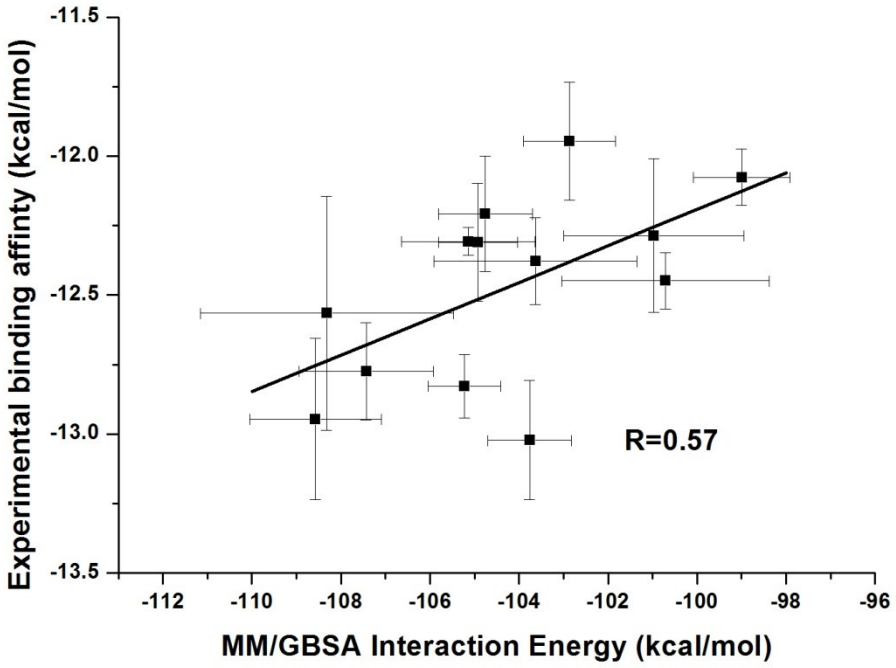


Figure 2.12. Comparison of calculated MM/GBSA interaction free energy and experimental binding affinity. Error bars show the standard error of values.

2.5 Conclusion

Protein-protein interactions exhibit memory-kernel-like fractal behavior, which causes anomalous autocorrelation times. To obtain statistically meaningful value of simulation properties, independent samples need to be obtained from simulation. The slowly decaying autocorrelation in protein-protein interaction makes the sampling process very challenging. A carefully tuned Langevin damp coefficient can serve as the critical damping strength in damped harmonic oscillator model, yielding minimal autocorrelation time. The tuning of damping coefficient may come with unwanted side effects, and requires further investigation.

Crystallographic water molecules in the binding site are necessary to model the binding interactions. Failure to correctly locate the crystallographic water molecules may lead to artificial conformational change or unfolding that lead to simulation errors.

Compared to the simulation of protein and small ligand binding, longer equilibration phases are necessary before the production MD simulation of protein-protein interactions, e.g. at least 1ns. Independent MD simulation using different initial velocities are chosen to be the sampling method. For PA-TEM8 and similar protein-protein complexes, we used 10 repeats of a 20ns simulation, including 5ns as the equilibration phase, and

obtained a correlation of 0.57 with experimental binding affinity. Similar sampling methods were used in the following chapters to estimate the MM/GBSA binding free energy.

CHAPTER 3

EFFECT FROM METAL ION IN TUMOR ENDOTHELIAL MARKER 8 AND ANTHRAX PROTECTIVE ANTIGEN

ABSTRACT

Anthrax toxin, the causative agent of anthrax, infects the host cells after the protective antigen binds to a cellular receptor. One of the receptors, tumor endothelial marker 8 (TEM8), is reported to be a potential anticancer target due to its over-expression during tumor angiogenesis. To extend our surface plasmon resonance (SPR) study in PA/TEM8 binding, we present a systematic computational approach to reveal the role of an integral metal ion on receptor structure and binding thermodynamics. We estimated the interaction energy between PA and TEM8 using computer simulation methods. In addition, the calculated relative dissociation constant between TEM8 and PA in the presence of different divalent metal ions was verified via SPR and compared with previous publications. Consistent with our experimental study, computational results indicate the metal ion in TEM8 contributes significantly to the binding affinity, and TEM8/PA binding is more favorable in the presence of Mg^{2+} than Ca^{2+}. In addition, some of the residues coordinated to the metal ion partially compensate the loss in interaction

energy resulting from metal ion replacement. Further, computational analyses suggest that the differences in TEM8/PA binding affinity are comparable to behaviors observed in closely related integrin proteins, which are known to adopt two conformations linked to changes in activity. Specifically, we found TEM8 remains in a conformation analogous to an integrin open (high-affinity) conformation when Mg^{2+} is bound to the TEM8 metal ion-dependent adhesion site (MIDAS). Nonetheless, molecular dynamics simulations suggest that when Ca^{2+} occupies the MIDAS, TEM8 may favor a locally unfolded conformation analogous to integrin α_L.

3.1 Introduction

Since the bioterrorism attack in 2001, the need to understand and develop countermeasures for anthrax infection has demanded greater attention[7]. The infection is typically transmitted through ingestion, inhalation or cutaneous, followed by cellular infection at the site of contact, causing distinct clinical symptoms[115]. The mechanism of cellular anthrax infection begins when protective antigen (PA), a component of anthrax toxin, binds to anthrax receptors on the cell surface[17].

Tumor endothelial marker 8 (TEM8) is one of the anthrax cell surface receptors in the human body[17]. Anthrax protective antigen can transport anthrax lethal factor and edema factor into human cells once it forms a pore after binding to its receptors[9]. TEM8 is also known to be over expressed in tumor cells[17], and was originally identified as the product of gene upregulation in tumor endothelium[116]. Since TEM8 functions in angiogenic processes that are required for tumor growth[117,118], the receptor has generated much interest as either a cancer marker[119,120] or a target for tumor-specific therapies[120]. Anti-TEM8 antibodies were developed by modifying PA to inhibit function of TEM8[121,122].

X-ray crystallographic studies[9] demonstrated that TEM8 is a von Willebrand factor type

A protein that contains a divalent cation in a metal ion dependent adhesion site (i.e. the MIDAS domain). Most other von Willebrand factor A proteins, including integrins, bind their relevant physiological ligand(s) via the metal cation. TEM8 mutants that disrupt metal binding, or wild-type TEM8 that lacks the divalent metal, yield a significant decrease in TEM8-PA binding affinity[29]. PA is believed to occupy the same binding site as physiological extracellular membrane ligand(s)[123]. Consequently, TEM8-PA interactions can be used as a model for studying TEM8 binding behavior both to its physiological ligand and anthrax protective antigen, to modulate the angiogenic effect(s) of TEM8 and the anthrax infection pathway. However, the exact structure of TEM8-PA complex and the molecular mechanism by which TEM8 exerts its angiogenic effect remain unclear.

Another anthrax receptor, capillary morphogenesis gene 2 (CMG2), shares 40% amino acid identity with TEM8, with close to 60% identity in the PA binding domain. The affinity between another anthrax receptor CMG2 and PA is significantly affected by the choice of metal cation[24]. Three possible metal ions facilitate the binding by coordinating to the MIDAS in CMG2: Mg^{2+} has the strongest binding affinity, followed by Zn^{2+}, and then Ca^{2+}. Similar results were found for TEM8[124]. However, a systematic examination of the role of the divalent cation on the TEM8-PA interaction has not been carried out.

In addition, several groups have taken advantage of the structural similarity between integrin α and TEM8 MIDAS domains. They speculate that the binding of TEM8 cytosolic domains to actin can result in conformational changes that switch the TEM8 MIDAS domain from high to low affinity states, and that these two conformational states could resemble the "open" and "closed" conformations of integrin MIDAS domains, respectively[125,126]. Presumably, such conformational changes could be responsible for TEM8 signaling. TEM8 X-ray structures show greater similarity to the integrin α I domain "open" conformation than to the "closed" conformation. Many integrin α I domain crystal structures are always bound to a ligand in "open" conformation, and not bound to a ligand in "closed" conformation. Although the TEM8 crystal structure doesn't contain a ligand, it is believed to be an "open" conformation. The integrin α conformation change is believed to be strongly coupled to the conformation of a phenylalanine on the C-terminal and a tyrosine near the MIDAS domain[127]. F205 and T118 are the corresponding residues conserved in TEM8, comparing to the phenylalanine and tyrosine in integrin α. Mutation of the TEM8 phenylalanine 205, highly conserved among related integrins, to tryptophan (F205W) has been shown to lock TEM8 into a high affinity state. Conversely, the T118A mutation lowers the binding affinity to PA by ~103 fold[128]. However, no structural data directly supporting different conformational states has been

generated for TEM8.

To investigate the role of the metal ion in the TEM8 and PA interaction, we generated a TEM8-PA complex structure model based on the highly homologous CMG2-PA crystal structure and evaluated the molecular model in the presence of different divalent cations. We also used our model to examine the possibility of "open" to "closed" conformational changes previously suggested for TEM8[125,128]. To validate the computational model, our experimental collaborators measured binding affinity through surface plasmon resonance (SPR) experiments that focused on the PA-TEM8 interaction in the presence of Mg^{2+} and Ca^{2+}. We found the molecular reason for the difference binding affinity between PA and TEM8, the "hot-spot" residues contribute most binding free, and examined the possibility of TEM8 "open" and "closed" conformation change.

3.2 Method

3.2.1 Preparation of Complex

To model TEM8-PA complex, we started from the crystal structure of CMG2-PA complex (1T6B[25]) and the crystal structure of TEM8 (3N2N[29]), a structure of PA binding TEM8

was obtained by using the TOPMATCH[80] protein structural precise alignment algorithm. A rigid body structural alignment was applied to replace CMG2 in 1T6B using TEM8 from 3N2N chain A at RMSD=0.66 Å. Missing loops in PA far (more than 14Å) from the binding surface were patched using the optimized conformation generated by Modeller9.11[81].

To reduce the computing cost, PA domain I and III were truncated in the simulation to improve efficiency. We consider the truncation will not significantly affect the binding interaction energy, because the truncated domains are more than 20Å away from the binding surface and are not known to contribute binding interaction energy[82]. To keep the truncated PA folded, a harmonic restraint were applied on the residues more than 14 Å away from the binding surface in MD simulation. A harmonic restraint of 1 $kcal \cdot mol^{-1}$ was chosen based on our tests of multiple restraint strengths as shown in section 2.4.1, to minimized the potential artificial impact from harmonic restraint. The harmonic restraint didn't significantly affect the dynamics properties of the protein complex. Hence, we assume the interaction energy was not significantly affected either. Mg^{2+} ion was used in the original crystal structure of TEM8. To study the effects of different metal ions, we replaced the Mg^{2+} Ca^{2+}, and set their initial coordinates as the same as Mg^{2+}.

To model free TEM8, the crystal structure of the TEM8 monomer chain "A" was also used. Ion replacement, energy minimization and molecular dynamics were done in the same way as for TEM8-PA complex, but without restraints.

3.2.2 Molecular Dynamics simulation

To minimize the possible errors resulting from homology modeling, we carefully designed the modeling process. An RMSD of 0.14Å (comparing to initial structure, and not including missing loops) on all heavy atoms was achieved after all energy minimization (before MD simulation in NAMD), which can yield satisfactory accuracy in binding affinity calculation[129].

Before dissolving the protein in explicit solvent, the initial atomic coordinates of PA-TEM8 complexes were first subjected to an energy minimization in vacuum to remove energy clashes. The energy minimization has 10 cycles of 1,000 steps, reducing the harmonic restraint each cycle on all protein atoms from 10 $kcal \cdot mol^{-1} \cdot Å^{-2}$ to 1 $kcal \cdot mol^{-1} \cdot Å^{-2}$ in decrements of 1 $kcal \cdot mol^{-1} \cdot Å^{-2}$ under Steepest Descent method in CHARMM 35b6 software package[83] and charmm27 force field parameters[46]. The protein was then dissolved in a TIP3[84] water and 0.15$mol \cdot L^{-1}$ NaCl box of 126Å×86Å ×86Å at a mixed density of 0.947g/cm^3. Extra chloride anions were used to neutralize the positive

charge of the protein complex.

After the protein solvent was generated, further energy minimization and molecular dynamics simulations of protein solution were run using the NAMD 2.10-GPU software package[85] with charmm27 force field parameters. Two cycles of 10000 steps of Conjugated Gradient minimization were run NAMD with fixed protein atoms and without constraints, respectively. Finally, 10 repeats of 20 ns molecular dynamics using different random seeds for starting velocities were produced[114]. Temperature was set at 300 K with a damping coefficient of $5ps^{-1}$ using Langevin dynamics[86]. Pressure was set at 1 bar using a Langevin-Noose Hoover piston[87] with a damping time of $50ps^{-1}$. Particle Mesh Ewald[88] was used to calculate long-range electrostatic interactions. The time step was set at 2fs with the use of Rigid Bond algorithm[89] between hydrogen atoms and heavy atoms.

The PA-TEM8 system has ~90000 atoms, the TEM8 system has ~60000 atoms, and the total simulation time is 800 ns. The MD simulations were run on 16 cores Intel Xeon E5-2665 and 2 NVIDIA K20 GPU.

3.2.3 Simulation Stability

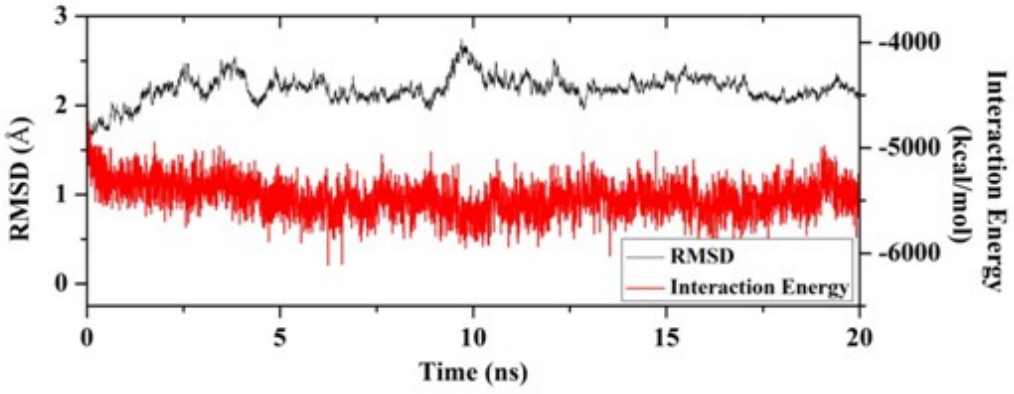


Figure 3.1 Molecular Dynamics simulation stability and equilibration. Black line shows RMSD of PA-TEM8 complex in 20ns MD simulation. Red line shows van de Waals and Columbic interaction energy between the protein complex and surrounding water molecules.

An optimized harmonic restraint parameter of 1kcal·mol^{-1}·Å^{-2} was applied on PA residues that are more than 14 Å from any atom in TEM8 (Section 2.4.1). Comparing to the superposed crystal structure, the RMSD of the TEM8-PA complex simulation fluctuated around 2.3 Å after the first 5ns of simulation (black line in Figure 3.1b).

3.2.4 MM/GBSA energy calculation

MM/GBSA is an end point energy calculation method that has been widely used to estimate the relative binding interaction energy[35,40,107,130]. In addition to the molecular

mechanics energies, the polar and non-polar solvation free energies are simplified with generalized Born implicit solvent approximation and solvent-accessible surface area term, respectively[131]. In this study, because the only difference in the two protein compounds is the metal ion, we assume the difference in translational and rotational entropies upon binding can be omitted. Thus, the binding interaction energies can be estimated according to the equation $\Delta G_{bind} = \Delta E_{MM} + \Delta G_{GB} + \Delta G_{nonpolar} - T\Delta S$, where ΔE_{MM} is the difference of gas-phase interaction energy between proteins, including the Columbic, van der Waals, bond, angle and dihedral energies; ΔG_{GB} and $\Delta G_{nonpolar}$ are the polar and nonpolar components of the desolvation free energy, respectively; $-T\Delta S$ is the change in conformational entropy during the binding process.

The change in configurational entropy upon binding was calculated using the Quasi-Harmonic estimation. The variance in entropic contribution ($-T\Delta S$) of PA-TEM8 in the presence of different metal ion is relatively small (0.03kcal/mol) compared to the variance in MM/GBSA energies (2.23 kcal/mol). The Quasi-Harmonic entropy did not converge in 20ns simulation trajectory for a large protein compound. Moreover, the difference is much smaller than the precision of the interaction energy calculation. Therefore, to save computing time and exclude the possible errors in not converged Quasi-Harmonic entropy calculation, we assume that the volumes of configuration space

occupied by the ligand and protein change negligibly upon association, and the $-T\Delta S$ term was not included in binding free energy calculation.

MM/GBSA free energies were calculated from snapshots taken from MD trajectory at commonly used[40] 5 ps intervals from 20 ns MD simulation. The first 5 ns simulation was considered as pre-equilibrium stage based on the interaction between the protein complex and water (red line in Figure 3.1b), thus any data from the first 5ns simulation were not used in further analysis. The single trajectory interaction energy calculation technique was used to cancel the errors resulted from internal strength energy, thermal noise and potentially inadequate configuration sampling when the energies were calculated from multiple simulations[43]. The conformation ensemble of the PA, TEM8 and PA-TEM8 complex were obtained from the same MD trajectory.

For each PA-TEM8 complex, the binding free energy of MM/GBSA was estimated as follows: $\Delta G_{bind} = G_{complex} - G_{PA} - G_{TEM8}$ where ΔG_{bind} is the binding free energy and $G_{complex}$, G_{PA}, and G_{TEM8} are the free energies of complex, PA, and TEM8, respectively. When calculating the MM/GBSA interaction energy between PA and TEM8, the unwanted protein and solvent molecules were removed, e.g. PA and solvent molecules were removed when calculating G_{TEM8}. When calculating interaction energy contributed

by individual residue, all solvent molecules were also removed; but for the unwanted protein, only the charges were removed, the structure was kept to maintain the proper Born radii for the residues. Graphs and visual inspections were made using VMD[132].

The free energies of each system were calculated using an MM/GBSA method implemented in CHARMM. E_{MM} was determined by CHARMM with 999Å cutoff distance (essentially no cutoff), and $G_{nonpolar} = \gamma A + b$ was estimated from the solvent accessible surface area (A in the equation) using SASA package in CHARMM[90], and $\gamma = 0.00542 kcal \bullet mol^{-1} \bullet Å^{-2}$, $\beta = 0.92 kcal/mol$. G_{GB} was calculated with the GBSW approach[91,92] implemented in CHARMM. Dielectric constants of 4 and 80 were used for solute and solvent, respectively. CHARMM default optimized parameters for GB-calculations[93,94] were used. For each PA-TEM8 complex, the binding free energy of MM/GBSA was estimated as follows: $\Delta G_{bind} = G_{complex} - G_{PA} - G_{TEM8}$ where ΔG_{bind} is the binding free energy and $G_{complex}$, G_{PA}, and G_{TEM8} are the free energies of complex, PA, and TEM8, respectively.

3.2.5 Energy Decomposition

Biding free energies resulting from non-bond interactions were decomposed at atomic level to evaluate the contribution of individual residues to the binding free energy. Van

der Waals and Columbic interaction energies were calculated using a traditional Molecular Mechanics method (INTE in CHARMM). For residue i in TEM8, j in PA, the MM interaction energy contribution of atom i between the two proteins was calculated by

$$E_i^{MM} = \sum_j E_{ij}^{MM} \tag{3.1}$$

The polar solvation energy was calculated using generalized Born (GBSW in CHARMM) approach. For residue i in TEM8, and PA, the polar solvation free energy contribution of atom i between the two proteins was calculated by

$$\Delta G_i^{p-sol} = \Delta G_{i+PA}^{p-sol} - \Delta G_{PA}^{p-sol} - \Delta G_i^{p-sol} \tag{3.2}$$

The non-polar solvation energy was calculated using solvent accessible surface area (SASA in charm) approach. For residue i in TEM8, and PA, the non-polar solvation free energy contribution of atom i between the two proteins was calculated by

$$\Delta G_i^{np-sol} = \gamma \times \left(SASA_i^{TEM8+PA} - SASA_i^{TEM8} - SASA_i^{PA}\right) \tag{3.3}$$

The summation of energies calculating from equation (3.1-3.3) gives the free energy contribution of residue i.

3.3 Results

3.3.1 Mg^{2+} results in higher TEM8/PA binding affinity than Ca^{2+}

MM/GBSA Interaction Energy								
Metal ion in PA/TEM8	ΔE_{ele}	ΔE_{vdw}	ΔG_{SA}	ΔG_{GB}	$\Delta E_{vdw}+\Delta G_{SA}$	$\Delta E_{ele}+\Delta G_{GB}$	$\Delta G_{cal.}$	ΔG_{exp}
Mg^{2+}	-28.65±3.78	-68.75±0.63	-13.49±0.10	7.13±3.01	-82.24±0.71	-21.52±0.45	-103.76±0.94	-12.82±0.10
Ca^{2+}	-22.75±4.12	-72.58±1.24	-13.48±0.14	7.28±3.54	-86.06±1.24	-15.47±0.85	-101.53±0.87	-10.01±0.10

Table 3.1 Experimental and calculated interaction energies (kcal/mol) for PA-TEM8 binding system. The standard error was estimated over the mean of 10 repeats, each repeat has 3000 data points. E_{ele} is the Columbic energy, E_{vdw} is the van der Waals energy, G_{SA} is the non-polar solvation free energy, G_{GB} is the polar solvation free energy, G_{cal} is the calculated free energy, G_{exp} is the experimental free energy.

To measure the binding interaction energy between TEM8 and PA in the presence of different metal ions, the dissociation constant was measured experimentally by SPR in the presence of Mg^{2+} and Ca^{2+}. Corresponding cations were placed in MIDAS domain in simulation models. As shown in Table 3.1, experiments showed TEM8 bound to PA in the presence of Mg^{2+} 2.81kcal·mol^{-1} stronger than in the presence of Ca^{2+}. Simulation data shows a difference of 2.23 kcal·mol^{-1}.

Stronger electrostatic interaction ($\Delta E_{ele}+\Delta G_{GB}$) was observed in the presence of Mg^{2+} comparing to Ca^{2+}. Weaker van der Waals interactions were observed in the presence of Mg^{2+}. The major difference comes from the Coulombic interaction. On the contrary, the GB energies do not show much difference. The GB energy mimics the solvation screening effect to partially cancel out about 25% of the Coulombic term. Surprisingly,

the summation of Coulombic term and GB term only has less than 1kcal/mol in standard error, which is, more than 60% less than the errors in either ΔE_{ele} or ΔG_{GB}. We ascribe it to the exaggerated error in Coulombic energy calculated in vacuum, which results from the lack of screening terms for the interaction between exposed charged residues. Adding in GB terms largely cancels the exaggerated Coulombic energy between these exposed residues.

The fluctuation (standard deviation) in vdW energy is comparable to the electrostatic term, but much smaller than the Coulombic term. It indicates the fluctuation in short range non-bond energy is much less than the long range electrostatic energies. And the short range energy term is less affected by the lack of solvation free energy, which can be further supported by the small variance and error in the non-polar solvation energy term G_{SA}.

In SPR experiments, TEM8 shows more than 150 times higher binding affinity towards PA in 1mmol Mg^{2+} than in 1mmol Ca^{2+} solution. The trend and magnitude are consistent with our theoretical results. When replacing the metal ion from Mg^{2+} to Ca^{2+}, the on rate is about 5 fold lower, however, the dissociation rate becomes 100 times faster. k_{on} and k_{diss} in the Mg^{2+} and Ca^{2+} mixture are similar to the results from the Mg^{2+} solution. First,

it suggests the TEM8 is less active towards PA in the aqueous solution containing Ca^{2+}. Secondly, even if it binds to PA in Ca^{2+} solution, the complex it still less stable than in the presence of Mg^{2+}. Above all, while the Mg^{2+} ion stabilizes the protein complex mainly by decelerating the dissociation, it also slightly accelerates the binding process.

Experimental Binding Affinities between PA and TEM8			
Metal ion in PA/TEM8	K_d (nM)	$k_{on}(10^4/(M\bullet sec))$	$k_{diss}(10^{-5}/sec)$
Mg^{2+}	3.33±1.12	1.57±0.19	5.17±1.65
Ca^{2+}	571±170	8.67±1.86	487±146
Mg^{2+}&Ca^{2+}	3.98±0.20	1.07±0.13	4.22±0.35

Table 3.2 SPR Experimental data of PA-TEM8 binding in the present of Mg^{2+} and Ca^{2+} ion. Data contributed by experimental collaborator.

3.3.2 Mg^{2+} interacts stronger to the surroundings than Ca^{2+}

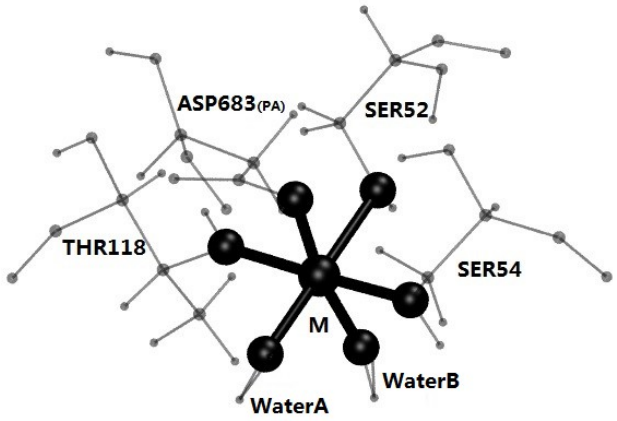


Figure 3.2A Diagram of MIDAS coordination structure in TEM8-PA complex. M is the metal ion on MIDAS domain. THR118, SER52 and SER54 are residues on TEM8. ASP683 is a residue on PA. The metal ion and oxygen atoms involve the coordinating are in bigger size. The other atoms are in smaller size.

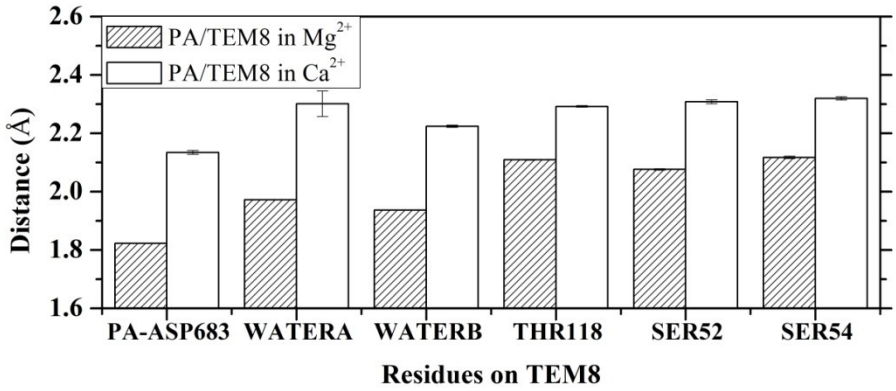


Figure 3.2B Distance between metal ion and the six coordination oxygen atoms in MIDAS. The residue names are the same as shown in Figure 3.2a. Error bars show standard error of the mean from 10 individual 20ns MD simulations.

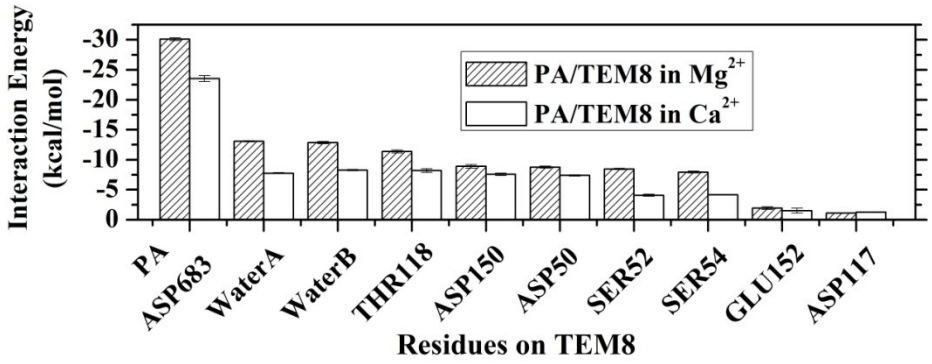


Figure 3.2C MM/GBSA interaction energy between metal ion and residues on PA/TEM8 complex. Residues are on TEM8 if not marked on PA. WaterA and WaterB are the same molecules shown in Figure 3.2a.

The distances between Mg^{2+} and all the coordinating oxygen atoms are shorter than those between Ca^{2+} and the oxygen atoms (Figure 3.2B). This results in Mg^{2+} interacting more strongly with the residues and water molecules around it (within 4.0 Å) relative to Ca^{2+} (Fig 2C). Among the three surrounding residues and the three water molecules directly interacts with the metal ion in MIDAS (as shown in Figure 3.2A), PA-ASP683 and TEM8-SER52/54 have larger RMSF in the presence of Ca^{2+} than Mg^{2+} (Table S3.1). The results can be interpreted as Ca^{2+} expands the size of MIDAS domain and leads to weaker interactions with residues.

Metal ion, Domain	$<\Delta E_{vdw}>$	$<\Delta E_{elec}>$	$<\Delta G_{GB}>$	$<\Delta E_{elec}+\Delta G_{GB}>$	$<\Delta G_{bind}>$
Mg^{2+}					
PA	9.56±0.14	-111.08±0.28	71.72±0.25	-27.09±0.08	-29.80±0.14
TEM8	7.20±0.09	-173.43±1.52	119.39±1.44	-54.03±0.36	-46.83±0.29
Water	12.16±0.06	-36.66±0.04	-1.52±0.08	-33.21±0.08	-26.03±0.11
Total	28.92±0.10	-321.17±0.89	189.59±0.85	-114.33±0.22	-102.67±0.20
Ca^{2+}					
PA	9.16±0.18	-106.58±0.97	73.74±0.97	-20.16±0.90	-23.67±0.84
TEM8	12.36±0.34	-164.35±1.43	118.41±1.36	-25.78±0.98	-33.58±0.86
Water	11.53±0.63	-26.29±1.06	0.47±0.23	-22.80±1.24	-14.30±0.73
Total	33.05±0.43	-297.22±1.17	192.62±0.98	-68.74±1.05	-71.55±0.81
Note: Energies are the mean of 10 repeats ± standard error, in kcal/mol.					

Table 3.3 van der Waals, Columbic and generalized Born interaction energy between PA, TEM8, coordinate water and metal ions.

Mg^{2+} interacts more strongly to PA, TEM8 and the coordinated water molecules than Ca^{2+} in terms of both vdW and electrostatic energy (Table 3.3). As the distances between the metal ion and the surrounding residues increase, all interaction energies between the metal ion and PA/TEM8/Water are weakened. This is one possible reason for the lower binding affinity in Ca^{2+} solution we observed in experiment.

The interaction energy with PA contributed by individual residues on TEM8 are also affected by the replacement of metal ion. Figure 3.3 shows the changes on the non-MIDAS residues (LYS111, ASP117, TYR119) are from the errors in calculation. The residues on MIDAS domain interact more weakly to PA when Mg^{2+} is in position, compared to Ca^{2+}. This is opposite to the results of interaction energy between the

residues and metal ions. When Ca^{2+} is in position, the expanded MIDAS coordination and weakened local interaction (Figure 3.2) makes the residues on TEM8 MIDAS domain interact stronger to PA in compensate.

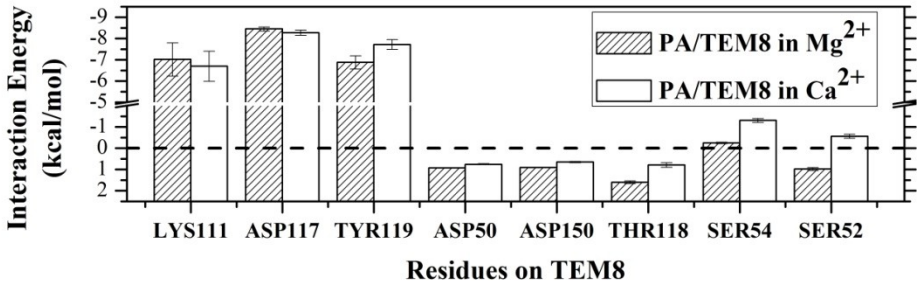


Figure 3.3 Biggest changes in interaction energy with PA (summation of vdW, Columbic and Generalized Born energies) contributed by individual residues on TEM8 resulting from metal ion replacement.

3.3.3 Metal ion in unbound TEM8

To explore further the reason for the different binding affinity contributed from metal ion, we simulated the unbound TEM8 containing Mg^{2+}, Ca^{2+} or no metal ion in its MIDAS domain. Each simulation was 20ns long. The crystal structure 3N2N chain A has Mg^{2+} as the MIDAS metal ion, and its MIDAS domain retained the original conformation by holding an RMSD of 0.28Å after 20ns(Table 3.4). As a substitute ion, Ca^{2+} does not fit into the MIDAS binding site as well as Mg^{2+} because of its larger ion radius. Whereas, it

expanded the MIDAS domain more severely than the distinction in ion size, so much so that the coordination residues cannot hold their original conformation and started to unfold. If no metal ion was present, the MIDAS domain unfolded faster than in the presence of Ca^{2+}. This agrees with the results of integrin α_L crystal structure in different metal ion conditions[133].

RMSD of MIDAS Residues				
Metal ion in PA/TEM8	RMSD (Å)			
	0.1ns	0.5ns	5ns	20ns
Mg^{2+}	0.27±0.03	0.29±0.04	0.29±0.04	0.28±0.04
Ca^{2+}	0.67±0.22	1.07±0.24	1.25±0.29	1.23±0.26
None	0.82±0.32	1.32±0.25	2.24±0.22	2.65±0.27
All data are mean of 10 repeats of 20ns simulation ± SE.				

Table 3.4. RMSD of residues (SER52, SER54, THR118) coordinate to the metal ion in 20 ns simulation, comparing to crystal structure 3N2N chain A.

Unfold was observed in all of the 10 repeats of 20ns simulation for TEM8 includes either Ca^{2+} or no metal ion in MIDAS domain. This could be a result of unstable conformation generated by homology modeling method. Comparing to the structural study of integrin α_L, it may imply that TEM8 possesses a slightly different conformation when Ca^{2+} is present and another obviously different conformation when there is no metal ion in MIDAS domain.

3.3.4 PA-TEM8 interaction mechanism

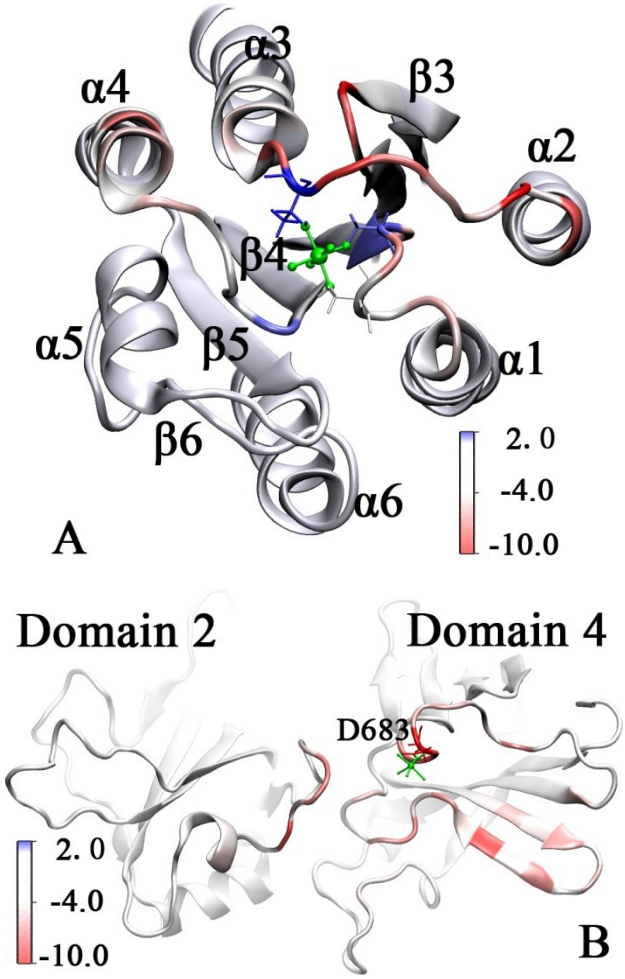


Figure 3.4 MM/GBSA interaction energy contributed by individual residues on TEM8 (A) and PA(B), in kcal/mol. Residues are colored by the strength of interaction energy.

TEM8 binds to PA via a buried surface area of more than 2200 $Å^2$ (Figure 3.4A). Among all the residues in the two proteins, residues on the contact surface contribute most of the

interaction energy. Among the contact surface on TEM8, α2-α3 loop and β2-β3 loop contribute more interaction energy than the other area. From β2-β3 loop to β5-β6 loop, the interaction energy donated by individual residues becomes less and less. Residues near the metal ion, D50, S52, T118 and D150, show repulsive interaction to PA. As expected, residues near the contact surface contribute more interaction energy than the further ones.

PA interacts with TEM8 mainly through its domain 4. PA domain 2 barely shows any interaction to TEM8, except a few residues adjacent to domain 4. The loop contains D683 and the two beta strands interact with TEM8 most strongly. D683 is the only residue on PA that involves in the formation of MIDAS. It exhibits the strongest interaction with TEM8 among all the residues on PA. Strong Columbic interaction combined with favorable cross polarization energy (electrostatic screening term)[134] composite the extremely attractive interaction between D683 and TEM8. The favorable electrostatic screening term indicates the stronger cross polarization effect between D683 and TEM8 in the bound complex than in the separated proteins, which accelerates the binding process. Slightly repulsive vdW interaction was discovered between D683 and TEM8, to partially compensate the extremely strong electrostatic interaction.

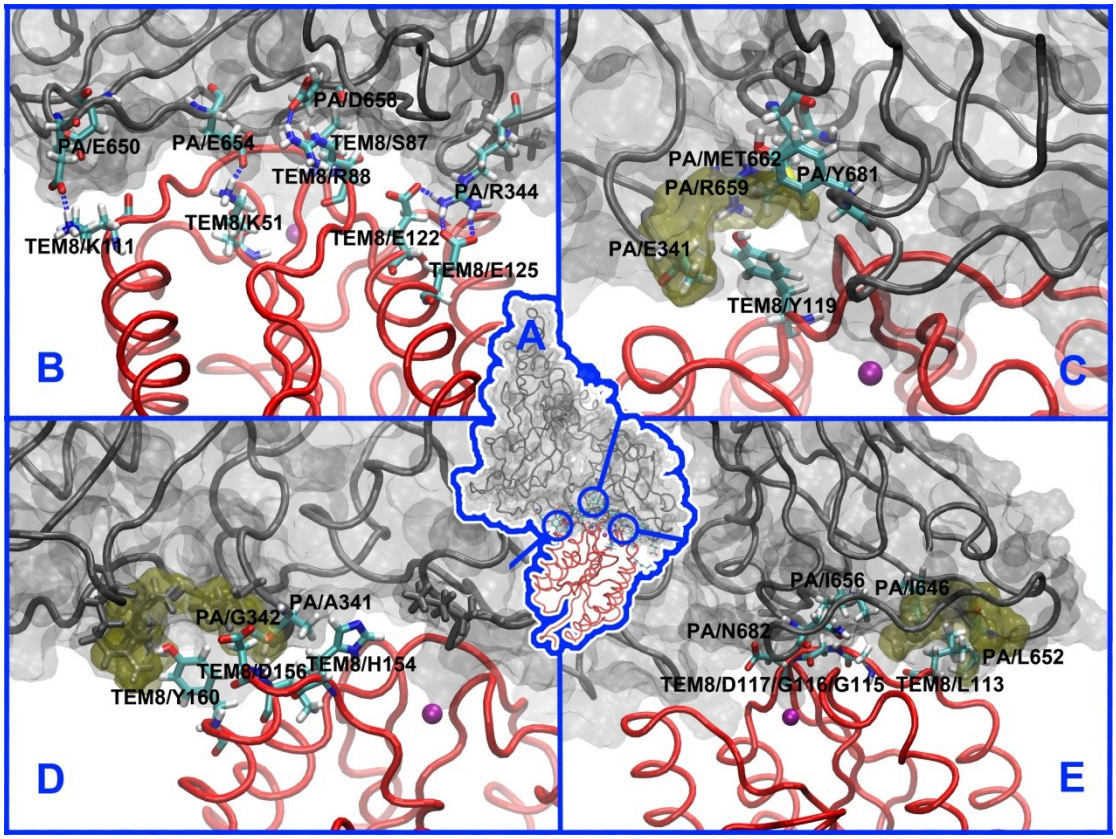


Figure 3.5 PA-TEM8 contact surface model and details of their interaction. (A) A model of PA-TEM8 contact surface (PA, gray; TEM8, red; Mg^{2+}, purple). (B) The hydrogen bonds formed between PA and TEM8. Hydrogen bonds are shown in dash lines. (C) Y119 residue on TEM8 inserts to the gap between PA domain 2 and domain 4. (D) Y160 residue on TEM8 inserts to the pocket on the side wall of PA domain 2. (E) L113 residue on TEM8 inserts to the hydrophobic pocket on PA domain 4. And there is a large contact surface between TEM8 residue 115 to 117 and N682 and I656 on PA. Yellow clouds in (C,D,E) are the pockets in PA. The pockets are made using a surface probe for PA atoms within 4.0Å of the inserting residue on TEM8.

In addition to D683, the salt bridges and hydrophobic insertion on large contact area between PA and TEM8 are also responsible for the strong binding. 4 salt bridges (Figure 3.5B), 3 hydrophobic pocket insertion (Figure 3.5CDE) and the buried surface area near MIDAS domain (Figure 3.5E) constitute the majority of PA-TEM8 interaction. TEM8-K111&PA-E650 (Figure 3.5B) salt bridge and TEM8-L113 insertion (Figure 3.5E) contribute the majority of interaction energy on theα2-α3 loop (based on Figure 3.3A) of contact surface. On the left side, salt bridges formed by TEM8-E122/E125&PA-R344 and the hydrophobic pocket formed by TEM8-Y160&PA-G342/A341 locked the binding surface. TEM8-K51&PA-E654 and TEM8-R88&PA-D658 form salt bridges near the MIDAS metal ion to secure the shape of MIDAS pocket. TEM8-Y119 inserts into a big hydrophobic pocked in PA formed by PA-E342/R659/M662/Y681. Y119 is directly connected to T118, its insertion also stabilized T118 in its high affinity conformation. Comparing to the residues on the inner contact surface, the residues on the edges of the contact surface have larger standard errors in interaction energies. This is understandable because the less buried residues are usually more flexible and, hence, suffer more energetic and conformational fluctuation originated from the solvent.

3.3.5 Ratchet like Hydrophobic Lock

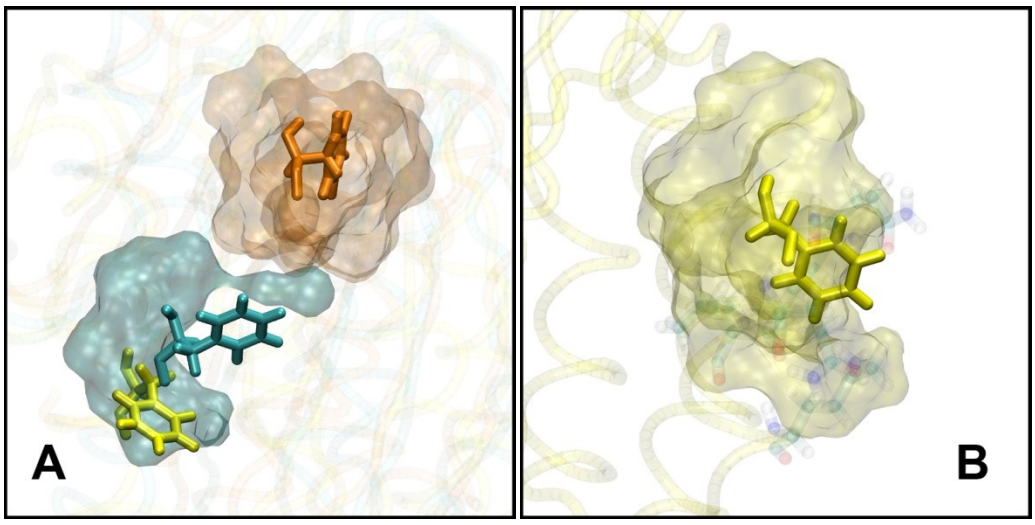


Figure 3.6 Comparison of the key phenylalanine on integrins and TEM8 controlling conformation change. Proteins in red and cyan are integrin M in closed and open conformations, respectively. Protein in yellow is TEM8. The key phenylalanine residues (F302 on integrin M and F205 on TEM8) are highlighted. The clouds show the residues within 4.0Å of the key phenylalanine made using 1.4Å surface probe. (A) Comparing integrin M close and open conformation with TEM8. (B) F205 on TEM8 and the residues within 4.0 Å.

A hydrophobic lock regulates the conformation change in integrin α proteins, which are 40% structurally similar to TEM8. In integrin α_M, F302 is well inserted in the hydrophobic lock when it is in the closed conformation, and it becomes exposed in open conformation (Orange and cyan parts in Figure 3.6A). The conversion from closed to

open conformation comes with more than 10Å displacement on F302 in the integrin α_M. In TEM8, F205 plays a similar role as F302 in integrinα_M. Based on the location of F205, TEM8 is in an open conformation (Yellow residue in Figure 3.6A). Different to integrins, although TEM8 is considered in an open conformation, F205 in TEM8 is more buried than F302 in integrin open conformation.

	SASA (Å)	E_{F205}
TEM8	28.33±0.16	9.63±0.07
α_M *open*	161.95±0.27	20.75±0.07
α_M *close*	13.68±0.07	8.23±0.07
α_L *open*	53.28±0.25	18.55±0.07
α_L *close*	3.99±0.08	9.85±0.07
All data are the mean of 3000 frames exact from 15ns simulation ± SE. Energies are in kcal/mol.		

Table 3.5 Solvent accessible surface area (SASA) of TEM8 F205, integrin αL F292, integrin αM F302, and their interaction energy (vdW, Columbic and Generalized Born energies) between the rest part of the protein.

In TEM8, F205 is more exposed than in close conformation by possessing larger solvent accessible surface area, but not as exposed as the phenylalanine in integrin open conformation (Table 3.5). Comparing to TEM8, in integrins, the phenylalanine (F292 in integrin α_L and F302 in integrin α_M) has 10 times larger solvent accessible surface area in open conformation than in closed conformation. It also suffers a penalty in energy of

9~12 kcal/mol (Table 3.5) when integrins shifted conformation from closed to open. The energy terms of TEM8 F205 appears to be on the same magnitude as phenylalanine in integrin closed conformation. Referring to the integrins, we suggest that TEM8 is in an open conformation, but it does not suffer the energy penalty as integrins. It can be considered as a stabilized high affinity conformation.

3.4 Discussion

Our simulations results show that TEM8 yields higher binding affinity to PA in the presence of Mg^{2+} than Ca^{2+}. The metal ion in MIDAS domain of TEM8 contributes a large part (26% to 28%) of the binding affinity. The size of metal ion is the major factor affecting the binding affinity. Smaller metal ions interact more strongly to both protein molecules in the complex, and result in higher binding affinity. The change in MM/GBSA interaction energy between metal ion and PA is larger than the change in total binding free energy. Residues around the metal ion compensate part of the change in binding affinity resulted from metal ion. Four salt bridges and three hydrophobic insertions also contribute a large part of the binding affinity. The interactions are different from the structural alignment study in 2010 [29]. The conformation of TEM8 is confirmed as “open”, and is stabilized in “open”.

In experiments, we note that TEM8-PA affinity in the presence of either Mg or Ca, obtained by SPR, has been previously reported for alternate truncations of soluble TEM8 [124,128]. Our K_d values agree with previous studies qualitatively, but are not identical quantitatively. For example, while our Kd in Mg (3.3 ± 1.1 nM) obtained via SPR is much lower than previous SPR data obtained by Scobie et al (130 ± 46 nM) [124] and Fu et al (30 ± 9 nM) [29]. Our measured Kd in Ca (570 ± 170 nM) is two-fold lower than the Scobie value (1100±41 nM)). Observed differences may simply reflect different MIDAS domain truncations. However, TEM8-PA off-rates are slow (10^{-4}/s, as measured by Fu et al by SPR and confirmed our own SPR data. Previously reported SPR values result from short (3-5 min) observation of complex dissociation (when reported), which could contribute to inaccuracies in the resulting reported K_d values.

3.4.1 Simulation Precision

The calculated difference in TEM8-PA interaction energy in the presence of Mg^{2+} and Ca^{2+} is 2.23±1.29kcal/mol, quantitatively agree with the experimental 2.82±0.14kcal/mol. The calculated relative interaction energy was averaged over 10 repeats of independent 20ns simulations. Each simulation generates 15ns trajectory for analysis, after discarding the first 5ns as pre-equilibration phase. We calculated the MM/GBSA energy over 15ns trajectory to average the fast motion. The fast motions are considered to be dominated by

the motion of small domains, and typically have vibration period between 5-8ns in NMR experiments[97].

The Generalized Born term moderated the overestimated fluctuation in Columbic energy. An anomalous standard error of more than 3kcal/mol was observed in the Columbic interaction energy, much larger than the ~1kcal/mol in total electrostatic interaction energy (Columbic plus GB). The standard error comes from the strong fluctuation in Columbic interaction energy of the residues formed salt bridges between the two proteins (e.g. K111, E122, E125 on TEM8). The fluctuation is a result of lack of screening term for the Columbic interaction energy, because the energy was calculated in vacuum, but the simulation was conducted in TIP3 explicit solvent. GB term mimics the screening effect of solvent and, as a result, partially cancels the exaggerated error in Columbic energy.

3.4.2 The Size of Metal Ion Matters

The size of metal ion leads to the major difference in binding affinity in the presence of Mg^{2+} and Ca^{2+}. Ca^{2+} has a larger ion radius than Mg^{2+}. Larger ion radius leads to larger distance between metal ion and the other residues in MIDAS. When the distance increased, vdW, Columbic and Generalized Born interaction energies become weaker,

according to their definitions. Similar results were found in quantum mechanical study[135], interaction between metal ions have 2 positive charges and residues on MIDAS are ranked as Mg^{2+}>Zn^{2+}>Ca^{2+}, the ion radius of Zn^{2+} is between $Ca^{2+}$and Mg^{2+}. When expanded to PA-TEM8 complex, smaller metal ion interacts stronger to both proteins in the complex, and results in higher binding affinity.

Similar to TEM8, many integrin α I domains, remain in a low affinity "closed" state [136] the presence of Ca^{2+}. In integrin α_L, the size of the metal ion was shown to affects the conformation of MIDAS domain[137]. Based on the relatively lower experimental and theoretical binding affinity of TEM8 containing Ca^{2+} in MIDAS, plus the locally unfolding results from the substitution of metal ion, we suggest that TEM8, analogous to the closely related integrin α_L, can possibly retain a low affinity conformation when Ca^{2+} is the ion in MIDAS pocket, which is slightly different from the high affinity conformation.

When the metal ion was stripped out, integrin α_L suffers a moderate conformation change around the MIDAS domain, and lose most of the binding affinity towards ligand [138]. Experiments show that TEM8 lost binding affinity to PA in EDTA solution [24]. An unfolding in MIDAS domain was observed after the metal ion from MIDAS was

removed in free TEM8 simulation.

We suspect that TEM8 may have a different conformation in the presence of Ca^{2+} or no metal ion in MIDAS. The possible conformation change can cause larger difference in binding free energy in the presence of difference metal ions, in addition to the MM/GBSA interaction energy. To confirm the hypodissertation, results from future NMR or crystallization structural study on TEM8 in Ca^{2+} or EDTA/EGTA can serve as powerful evidences.

Replacing metal ion itself would have changed the interaction energy between TEM8 and PA by 6.13kcal/mol (Table 3.2), regardless the energy change on the other residues. It is 3.90kcal/mol more than the difference in total binding affinity. As shown in Figure 3.3, residues on the MIDAS interact more strongly to PA when Ca^{2+} is in the pocket. They partially compensated the lost in interaction energy resulted from replacing metal ion. Hence, when calculating the relative binding affinity, it is necessary to calculate the energy of all the residues on the same binding pocket, not only the mutated residue or atoms.

3.4.3 Large Contact Area Leads to Tight Binding

The 4 salt bridges and 3 hydrophobic insertions in the 2200Å^2 buried surface area are responsible for the strong binding affinity. TEM8 K51, R88, K111 and E122/E125 form salt bridges with PA E654, D658, E650 and R344, respectively. Different than the strong interaction between TEM8 K51, R88, K111 and PA (Table S3.1), TEM8 E122/E125 shows much weaker interaction to PA. In PA-CMG2 crystal structure [25], CMG2 E122 forms a salt bridge with PA R344. When pH is decreased, the CMG2 E122 drift away from R344 after 2.6ns of MD simulation [82]. Similar phenomena were observed in our PA-TEM8 simulation at pH=7. Further, TEM8 E125 can also form salt bridge with PA R344, competing with PA E122. E122 is considered as the base of ligand binding for CMG2. Hence, pH may affect the binding preference of PA towards CMG2 and TEM8. The stability of this salt bridge and the possibility for PA R344 to form salt bridge with other residues still need further investigation.

TEM8 loop α2-α3 and loop β2-β3 contributes a large portion of the interaction energy (Figure 3.4). Mutating residue in this domain to another similar residue is not likely to affect the binding affinity [29]. In addition to the buried surface area, the 4 salt bridges and 3 hydrophobic insertions are responsible for the binding. Mutations on residues that formed salt bridges or hydrophobic insertions e.g. K51A, R88Q, E122A, Y119H, show

big impact to the interaction energy, in terms of IC50 [29]. L56 on α1-β1 has attractive interaction with PA (Table S3.1). V55 and H57 on the same loop also show attractive interaction, but slightly weaker. L56A mutation exhibits 10 times better IC 50 without affecting the pH threshold for pore formation. It suggests that L56A mutation decreases the distance betweenα1-β1 loop and PA, thus allows local residues to contribute stronger interaction energy.

3.4.4 TEM8 Adopts a Stabilized Open Conformation

TEM8 extracellular vWA domain (3N2N,A[29]) shows high similarity with integrin alpha I domains in open conformations. It has C_α atoms RMSD of 2.81Å to integrin α_L (PDB ID 1MQ9 [137]), 3.18 Å to integrin α_M (PDB ID 1IDO [139]), 2.71 Å to integrin α_X (PDB ID 1N3Y [140]), 2.75 Å to integrin α_1 (PDB ID 1QCY [141]), 2.81 Å to integrin α_2 (PDB ID 1DZI [142]), although they have low sequence identity (13%-20%). The integrin I domains have two possible conformations. They are open and closed conformations (Figure 3.6A), representing the active and inactive states, respectively[136,12]. In the closed conformation, integrins are more stable than in the open conformation. The key-like residue (phenylalanine or glutamate acid, orange residue in Figure 3.6A) inserts into the hydrophobic lock near C-terminal (orange cluster in Figure 3.6A). In the open conformation, the key like residue is pulled out of the hydrophobic lock, and the structure

becomes less stable. The hydrophobic lock is formed by a valine (or leucine for integrin α_L) and a leucine in integrins. In TEM8 and CMG2, the sequence distance between the valine and the leucine is one amino acid further. As shown in the crystal structure we have for TEM8, the hydrophobic pocket in not well formed. It makes TEM8 less possible to be stabilized in a closed conformation.

F205 on TEM8 interacts to the surrounding protein and solvent molecules similarly to "closed" conformation integrins. Although the activity and conformation of TEM8 can be considered as "open" conformation [29], the hydrophobic ratchet pocket controlling conformation change has 28.33 $Å^2$ solvent accessible surface area, much smaller than 161.95 $Å^2$ and 53.28 $Å^2$ for "open" conformation integrins. It is also more stable in terms of energies, 9.63 kcal/mol comparing to 20.75 kcal/mol and 18.55 kcal/mol for "open" conformation integrins. Unlike the highly buried in "closed" integrins or largely exposed in "open" integrins, F205 in TEM8 is partially buried on the protein surface. Thus, the F205W mutation did not make significant change in TEM8 conformation [128]. It is still possible that TEM8 has a "closed" conformation. However, because the TEM8 "open" conformation is more stabilized comparing to integrin "open" conformation, TEM8 is not likely to change to its "closed" conformation spontaneously.

Integrin	Starting Res	C Terminal residues
aL	288	I **L** D T **F** E K **L** K D L F T E L Q K K I Y V I E
aM	298	Q **V** N N **F** E A **L** K T I Q N Q L R E K I F A I E
aX	296	K **V** E D **F** D A **L** K D I Q N Q L K E K I F A I E
a1	313	N **V** S D **E** L A **L** V T I V K T L G E R I F A L E
a2	314	N **V** S D **E** A A **L** L E K A G T L G E Q I F S I E
TEM8	201	**V** N D G **F** Q A **L** Q G I I H
CMG2	199	**V** K G G **F** Q A **L** K G I I

Table 3.6 Residues on integrins and ANTRX coordinate conformational change.

An RMSD of 2.1Å was observed in 10 repeats of 20ns free TEM8 simulation, comparing to the initial structure. According to NMR experimental study [97], 20ns is sufficient for protein fast motions, such as the fluctuation of side chains and loops. It is not as powerful to predict the movement of large domains. Although structural change on the dimension of domains was not observed, it may still happen in longer simulation.

3.5 Conclusion

Both experimental and simulation data point out TEM8 binds to PA better in the presence of Mg^{2+} than Ca^{2+}. Because Mg^{2+} is smaller in size, it interacts with PA, TEM8 and the coordinate water molecules more strongly than Ca^{2+}, and leads to higher binding affinity. Introducing Ca^{2+} to Mg^{2+} solution does not affect the binding affinity. It indicates Mg^{2+} interacts more strongly to TEM8, in either free TEM8 or TEM8-PA complex, than Ca^{2+}. On the other hand, the residues around the metal ion partially compensate the change in

interaction energy by interacting more strongly to PA when Ca^{2+} is in MIDAS. Other than the metal ion, residues on the buried surface area contribute almost all the interaction energy.

We suggest the existence of a low affinity conformation of TEM8 in the presence of Ca^{2+} in MIDAS and a locally unfolded conformation in the absence of metal ion. Limited by the simulation time scale, large domain motion or conformation change was not observed in the simulation. Future structural study can provide more evidence to confirm the hypodissertation.

TEM8 stays in a stabilized open conformation. Although the overall backbone structure, binding affinity and the conformation of F205 all indicate TEM8 is in an open conformation, the SASA and energy penalty of F205 suggest the conformation is stabilized. TEM8 may still have a closed conformation, but, unlike integrins, it may not change to closed conformation spontaneously.

Loud thermo noise comes with the large contact surface area and strong interaction. Adding GBSA terms into traditional MM (vdw and Coulombic energies) reduced the standard deviation in interaction energy by 70%. With the help of MM/GBSA method, it

is less expensive to obtain 1kcal/mol precision for protein-protein interaction energy calculation.

3.6 Appendix

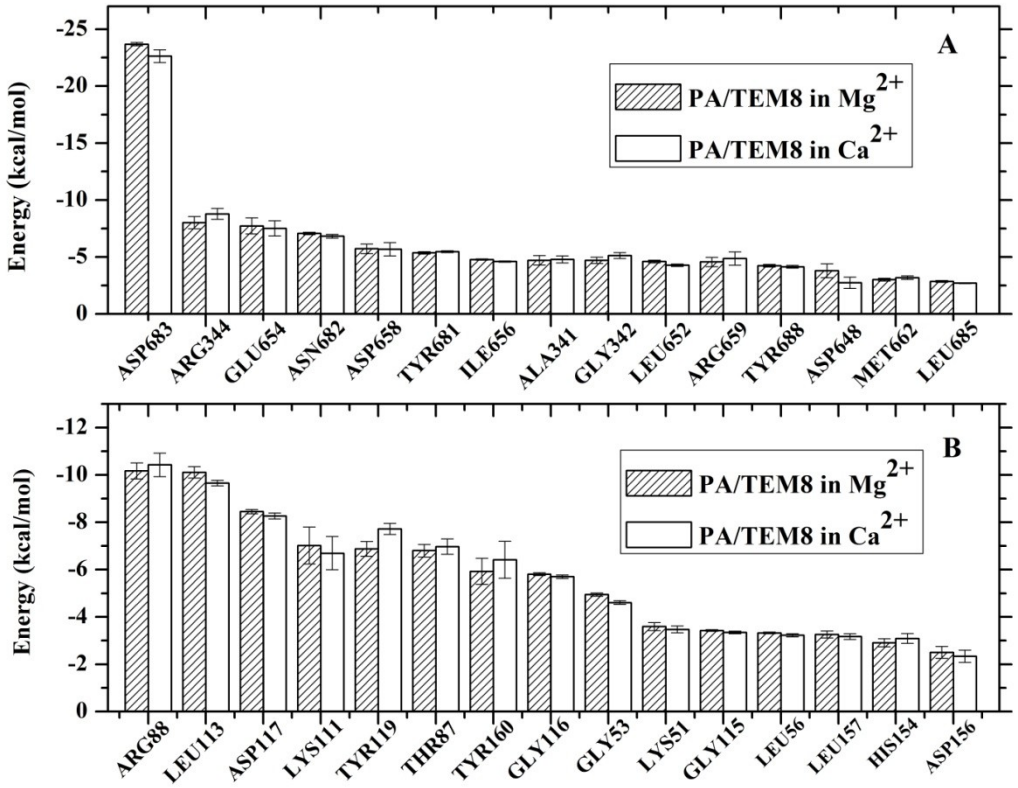


Figure S3.1 Interaction energy contributed by individual residues. (a) Residues on PA contribute most (top 15) to the interaction energy. (b) Residues on TEM8 contribute most (top 10) to the interaction energy. Error bars show standard error of the mean from 10 individual 20ns MD simulations.

CHAPTER 4

FIND THE DIFFERENCES FROM SIMILARITY: THE DISTINCT INTERACTION MECHANISM OF ANTHRAX RECEPTORS CMG2/TEM8

ABSTRACT

Other than the anthrax toxin receptor TEM8, which is over-expression during tumor angiogenesis, of the receptors, capillary morphogenesis gene 2 (CMG2), is reported to be a angiogenesis regulation protein widely distributed in human tissues. The two receptors are structurally highly similar to each other, but are responsible for distinct human diseases. To facilitate the development of drugs specific to either CMG2 or TEM8, we examine the differences in mechanism between CMG2 and TEM8 bind to PA. We estimated the interaction energy between PA and CMG2 using computer simulation methods, and compared with our previous study of PA-TEM8 binding. In addition, the calculated relative dissociation constant between CMG2 and PA in the presence of different divalent metal ions was verified via SPR. Consistent with our experimental study, computational results indicate CMG2 binds to PA on a significantly higher affinity,

and the binding is tighter in the present of Mg^{2+} than Ca^{2+}. Further, computational analyses suggest that the differences in CMG2-PA and TEM8-PA binding affinity are mainly due to the differences in α4 helices in CMG2 and TEM8, plus contribution from the difference in other residues on the contact surface. Specifically, we found PA domain 2 has a strong preference to CMG2 than to TEM8. The correlated motion between CMG2 and PA domain 2 was not found in TEM8-PA interaction. Thus, there may exist distinct signaling pathways in CMG2-PA binding and TEM8-PA binding.

4.1 Introduction

Anthrax infection begins from the contact of anthrax protective antigen (PA) and its cell surface receptors. The two known receptors tumor endothelial marker 8 (TEM8, ANTRX1) and capillary morphogenesis gene 2 (CMG2, ANTRX2) form a toxin pore upon binding with PA that transmits the anthrax lethal factor and edema factor into cells [9]. The extracellular von Willebrand Factor A domains of two receptors (TEM8 splice variant 1 and CMG2-489) are structurally high similar to each other (58% identities, 78% positives, BLAST2[143], 1.2Å RMSD)[11] with a fully conserved metal ion dependent adhesive site[29]. Previous study [cite PA-TEM8 ion study] shown that they possibly bind to PA through similar binding surface. The loops on the binding contact surface of TEM8 and CMG2 share 48% identities and 76% of similarities. Although the exact physiological functions of TEM8 and CMG2 are not known yet, they are known to regulate the angiogenesis process or the growth of new blood vessels[144]. Malfunctioning angiogenesis causes many pathological processes including metastasis, tumor growth and proliferative retinal diseases, as a consequence of defective TEM8[145] or CMG2[20].

TEM8, according to its name "tumor endothelial marker 8", is overexpressed in malignant tumor endothelium rather than normal endothelium[16,116]. It has been confirmed that clinical results in cancer vaccines are correlated to TEM-8 expression levels[146]. For

tumor cells[147,148] and other cell types[149,150] outside the vasculature, the distribution of TEM8 is well established. In tumor vasculature, the expression of TEM8 was also tested under the impaction of a good number of potential anti-tumorigenic therapeutics, not including anthrax toxin. Most TEM88 targeted anti-angiogenic therapies combine an anti-TEM8 antibody with truncated tissue factor to create a fusion protein. The fusion protein can localize at tumor blood vessels and promote local thrombosis, thus disrupt tumor vasculature and reduce tumor volume. On the other hand, CMG2 is not overexpressed in tumor endothelium comparing to TEM8. The mutation and defection of CMG2 are shown to be linked to infantile systemic hyalinosis (ISH)[20], juvenile hyaline fibromatosis (JHF) and breast cancer[151].

Protective antigen based anti-angiogenesis therapies targeting TEM8 is being developed as anti-cancer agents[152,153]. Naturally, PA binds to CMG2 at 100~1000 folds higher affinity than to TEM8[124] with the presence of Mg^{2+} cation in MIDAS. Experiments showed that mutations in PA domain 4 are able to alter the binding affinity to anthrax receptors[154,155]. Furthermore, PAD7 variant (R659S/M662R) binds preferentially to TEM8 rather than CMG2 resulting 10 times better EC50[152]. However, the selectivity still needs to be improved to be considered for therapy purpose.

Crystal structure of PA-CMG2 bound complex showed that PA binds to CMG2 through its domain 2 and 4[25]. Benefiting from the high structural similarity between TEM8[29] and CMG2[11], plus the previous simulation results (Chapter 3), we believe TEM8 binds to PA via similar binding site. Alanine scanning[154] and simulation study identified the "hot spots"[134] on PA for the binding were loop 680-688, loop 652-658 and loop 340-344. However, the preference between TEM8 and CMG2 of these loops and the individual residues on the loops has not yet been investigated.

To further identify the "hot spots" on PA for PA-TEM8/CMG2 binding, we employed homology modeling method to build conformation of PA-TEM8 bound complex, and did simulation to identify the contribution to binding free energy by each individual residue on PA in PA-TEM8 and PA-CMG2 binding. We found loop 652-658 on domain 4 prefers CMG2, whereas, loop 340-344 prefers TEM8. We also simulated the binding mechanism of PA binding to TEM8 in the presence of Ca^{2+} or Mg^{2+} cation in MIDAS. Ultimately, our simulation binding free energy agrees with surface plasmon resonance experiment results.

4.2 Method

We used the crystal structure of PA-CMG2 complex (PDB: 1T6B[11]) as the initial

structure for MD simulation. Missing loops in PA far (more than 14Å) from the binding surface were patched using the optimized conformation generated by Modeller9.11[81].

PA domain I and III were truncated in the simulation to improve efficiency using a similar to the simulation method used in Chapter 3. We consider the truncation will also not significantly affect the binding interaction energy in PA-CMG2. 1kcal•mol^{-1} harmonic restraints were applied on the residues more than 14 Å away from the binding surface in MD simulation to prevent unwanted conformation change. Detailed reasons were explained in section 2.2.3. To study the effect of different metal ions, we replaced the Mg^{2+} with Ca^{2+}, and set the initial coordinate as the same as Mg^{2+}.

Energy minimizations and Molecular Dynamics simulations were conducted under the condition as mentioned in Chapter 3. A total of 10 repeats of 20ns MD simulation trajectory were generated using different initial velocities. The first 5ns of MD simulation trajectories were considered as equilibration phase, the last 15ns were used in further analysis.

MM/GBSA interaction free energy and per residue interaction energy decomposition were also conducted under the same condition as mentioned in Chapter 3. For the same

reason as in Chapter 3, the configurational entropies were not included in the interaction free energy.

Similarity and identity of residues on CMG2 and TEM8 were obtained using blast2seq[143] and TOPMATCH[80]. Because CMG2 (chain Y of PDB: 1T6B) has fewer residue than TEM8 (chain A of PDB: 3N2N), CMG2 residues are aligned according to the residue numbers of TEM8.

4.3 Result and Discussion

To find the differences in the binding mechanism between PA-CMG2 binding and PA-TEM8 binding, we calculated the interaction energy of PA-CMG2 binding and decomposed the interaction free energy into per-residue contribution. The per-residue contribution of PA and CMG2/TEM8 are investigated in details. Furthermore, we also investigated how the replacement of MIDAS metal ion influents the binding.

4.3.1 CMG2 binds to PA stronger than TEM8

MM/GBSA Interaction Energy						
Protein	ΔE_{ele}	ΔE_{vdw}	ΔG_{SA}	ΔG_{GB}	$\Delta G_{cal.}$	ΔG_{exp}
CMG2	-185.31±3.42	-74.04±0.92	-13.94±0.10	163.17±2.72	-110.12±0.87	-14.05±0.08
TEM8	-28.65±3.78	-68.75±0.63	-13.49±0.10	7.13±3.01	-103.76±0.94	-12.82±0.10

Table 4.1 Experimental and calculated MM/GBSA interaction energy between CMG2/TEM8 and PA in the presence of Mg^{2+} ion in MIDAS. E_{tot} is total MM/GBSA interaction energy; E_{vDW} is the van der Waals interaction energy; E_{elec} is the Coulombic interaction energy; E_{GB} is the generalized Born interaction energy; E_{SASA} is the solvent accessible surface area interaction energy. Values are differences ± standard error. Standard errors are calculated from 10 repeats of 20ns simulation.

The MM/GBSA interaction free energy between CMG2 and PA is compared with the binding free energy converted from the dissociation constant measured experimentally by SPR in the presence of Mg^{2+}. Corresponding catiosn were placed in MIDAS domain in simulation models. As shown in Table 4.1, experiments showed CMG2 bound to PA in the presence of Mg^{2+} 6.36kcal·mol^{-1} stronger than in the presence of Ca^{2+}. Simulation data shows a difference of 1.23 kcal·mol^{-1}, matches with experimental results quantitatively.

Stronger Coulombic interaction (ΔE_{ele}) was observed in PA-CMG2 binding comparing PA-TEM8. If we add it up with the screening term (ΔE_{GB}), the total electrostatic interaction energy of PA-CMG2 binding equals to -22.14 kcal/mol, is still stronger than

the -21.52 kcal/mol for PA-TEM8 binding. Stronger van der Waals interactions were observed in the PA-CMG2 binding, which leads to the major difference. This may indicate that most of the salt bridges were conserved in PA-CMG2 interaction, with an addition of some hydrophobic binding pockets comparing to PA-TEM8 binding.

4.3.2 α4 helix and S113/L113 differ the CMG2/TEM8-PA binding affinity

To further investigate the reason of the difference in binding interaction free energy between CMG2 and TEM8, we calculated the binding interaction free energy contributed by individual residues on CMG2 and TEM8 (Figure 4.1, Table 4.2). The residues contributed to the difference in binding free energy to PA are mainly located close to the contact surface with PA.

α4 helix (residue 150-158 on CMG2, 152-160 on TEM8) that interacts mainly with PA domain 4 makes CMG2 interact with PA more strongly. The decomposed energy terms in Table 4.2 shows an extremely more unfavorable Coulombic interaction between TEM8 ASP156, GLU 155 and PA, though the electrostatic screening term cancels most of the difference. On the same spot, CMG2 LEU154, GLY153 have much stronger van der Waals interaction with PA. TYR158 on CMG2 also have stronger van der Waals interaction with PA than TYR160 on TEM8. To further examine the reason, the residues

on CMG2/TEM8 α4 helices are listed in Table 4.3, and the residues on the PA domain 2 loops close to the α4 helices are listed in Table 4.4. The α4 helix in CMG2 and the nearby loops in PA domain 2 both have neutral charge, whereas, the α4 helix in TEM8 has a total charge of negative 3. The negatively charged α4 helix in TEM8 is less favored by PA domain 2 comparing to the charge neutral helix in CMG2.

SER87, GLN88 in CMG2 show weaker interaction free energy towards PA comparing to THR87, ARG88 in TEM8, which form salt bridges with residues on PA (Chapter 3). The distinct Coulombic interactions between GLN88 in CMG2 and ARG88 in TEM8, plus the less difference in generalized Born screening term together indicate the loss of a salt bridge in CMG2-PA interaction. The loss in Coulombic interactions is partially compensated by the gain in vdW interactions. The difference in SER87/THR87 is mainly from vdW interactions, which roughly cancels the gain in vdW interaction of GLN88. In all, the difference in SER87/THR87 and GLN88/ARG88 is mainly electrostatics interaction energy.

Another spot that raises major difference locates in α2-α3 loop in CMG2/TEM8, SER113 and VAL 115 were found in TEM8, LEU113 and GLY115 in CMG2. If we consider the

sequence of the entire loop 112-118, TEM8 has VSPVGET, CMG2 has VLPGGDT. The change from SER113 in TEM8 to LEU113 in CMG2 decreased the contribution to binding affinity by 6.78kcal/mol, mainly vdW interaction energy. Interestingly, the change from VAL115 in TEM8 to GLY115 in CMG2 compensates the loss by 5.01kcal/mol. These two changes are the largest difference in per residue interaction free energy contribution when compared CMG2 with TEM8. Considering the flexibility of glycine and the larger size of leucine versus serine, LEU113 leads to a further distance between the loop in CMG2 and PA, hence weaker vdW interaction energy; whereas, the combination of SER113 and VAL115 in TEM8 loop allows a smaller distance between the contact surfaces and stronger interaction energy.

TYR119 on both CMG2 and TEM8 inserts into the hydrophobic pocket between PA domain 2 and domain 4. Although the interaction free energy of this insertion makes CMG2 binds more strongly to PA, it is not considered as one of the residues actually contribute the difference because of the large standard error of the interaction free energy of this insertion (1.07±0.90 kcal/mol in Table 4.2). The standard error in the per residue free energy contribution of ARG111/LYS111 is also comparable to the difference in energy. Thus, TYR119/TYR119 and ARG111/LYS111 were not considered as factors that have changed the binding affinity.

Thus, the difference in CMG2/TEM8 binding affinity towards PA is mainly caused by the difference in α4 helix, S113/L113, G115/V115, GLN88/ARG88, SER87/THR87. Most of the difference in affinity is caused by the different charges on α4 helix. Experiments have shown that mutating LEU56 in TEM8 using the corresponding ALA56 in CMG2 can improve TEM8-PA binding[29]. We haven't found evidence to explain such finding. We predict that mutations on TEM8 that neutralized the negative total charge may enhance the binding between TEM8 and PA; S113L mutation on CMG2 may enhance CMG2-PA binding.

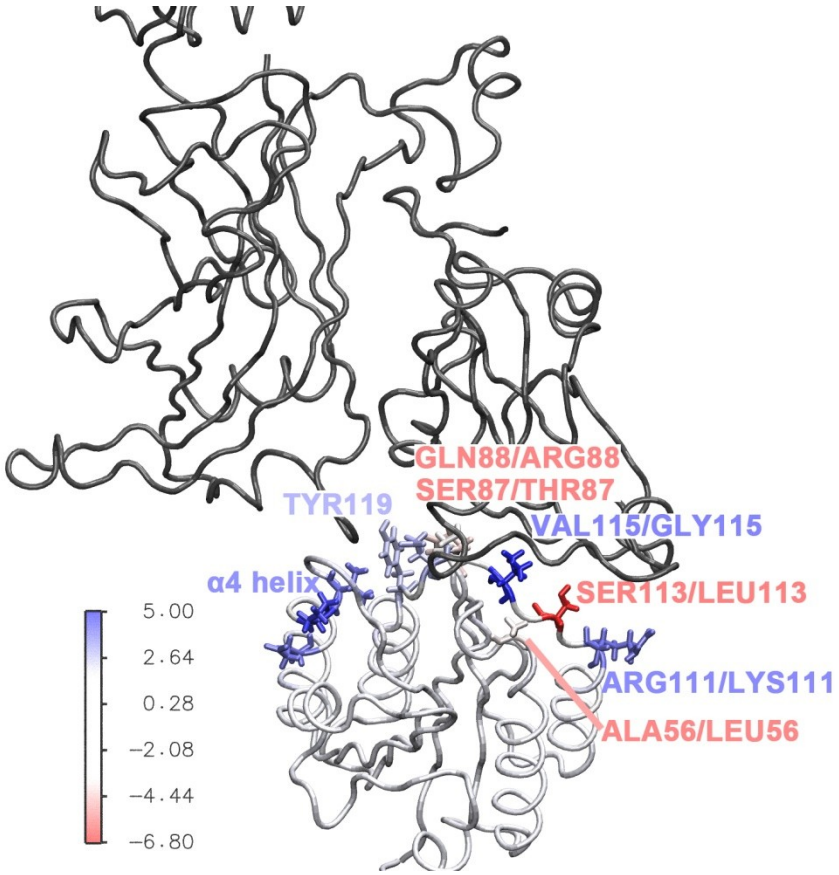


Figure 4.1 The colors for residues show the difference in MM/GBSA interaction energy between PA and individual residues on CMG2/TEM8. The unmarked grey protein in the upper part represents PA. The white protein on the bottom part represents CMG2/TEM8. Residue name and numbers are marked by name and number in CMG2/TEM8. Positive energies mean CMG2 residue interacts more strongly to PA; negative energy values mean the corresponding residues on TEM8 interact more strongly.

Difference in MM/GBSA interaction energy between PA and individual residues on TEM8/CMG2(E_{TEM8}-E_{CMG2})					
CMG2/TEM8 Residue	ΔE_{tot}	ΔE_{vDW}	ΔE_{elec}	ΔE_{GB}	ΔE_{SASA}
Stronger in TEM8					
SER113/LEU113	-6.78±0.62	-6.66±0.61	0.97±0.19	-0.68±0.17	-0.41±0.03
GLN88/ARG88	-2.38±1.14	1.52±0.96	-54.55±2.82	50.68±1.34	-0.03±0.05
ALA56/LEU56	-1.74±0.41	-1.60±0.39	-0.04±0.09	0.06±0.07	-0.17±0.03
SER87/THR87	-1.51±0.60	-1.17±0.55	-0.14±0.51	-0.06±0.50	-0.13±0.01
Stronger in CMG2					
VAL115/GLY115	5.01±0.26	4.35±0.26	0.27±0.16	-0.05±0.13	0.45±0.01
LEU154/ASP156	4.35±0.73	3.08±0.51	23.63±1.43	-22.58±1.05	0.22±0.04
ARG111/LYS111	2.88±2.62	0.09±1.50	-1.30±10.31	3.96±6.71	0.13±0.12
TYR158/TYR160	2.88±1.33	2.97±1.25	0.23±0.35	-0.41±0.15	0.09±0.06
GLY153/GLU155	2.87±0.21	2.29±0.16	19.58±0.45	-19.19±0.42	0.19±0.02
GLU117/ASP117	1.50±1.12	0.73±0.48	4.16±1.62	-3.46±1.54	0.08±0.02
TYR119/TYR119	1.07±0.90	1.22±0.59	-0.10±0.50	-0.08±0.28	0.02±0.04

Table 4.2 Residues on CMG2 and corresponding residues in TEM8 with difference in MM/GBSA interaction energy between the residue and PA greater than 1kcal/mol. Negative energy values indicate the residues on CMG2 interact with PA more strongly; positive energies indicate the corresponding residues on TEM8 interact with PA more strongly. E_{tot} is total MM/GBSA interaction energy; E_{vdW} is the van der Waals interaction energy; E_{elec} is the Coulombic interaction energy; E_{GB} is the generalized Born interaction energy; E_{SASA} is the solvent accessible surface area interaction energy. Values are differences ± standard error. Standard errors are calculated from 10 repeats of 20ns simulation.

Residues on α4 helices of CMG2 and TEM8			
CMG2	*Charge*	*TEM8*	*Charge*
LYS150	+1	GLU152	-1
LEU151	0	LEU153	0
ASP152	-1	HIS154	0
GLY153	0	GLU155	-1
LEU154	0	ASP156	-1
VAL155	0	LEU157	0
PRO156	0	PHE158	0
SER157	0	PHE159	0
TYR158	0	TYR160	0
Total	0	*Total*	-3

Table 4.3 Residue names and their charges on α4 loops of CMG2 and TEM8.

Residues on PA domain 2 loops on	
Residue	*Charge*
LEU338	0
SER339	0
LEU340	0
ALA341	0
GLY342	0
GLU343	-1
ARG344	+1
THR345	0
TRP346	0
ALA347	0

Table 4.4 Residue names and their charges on PA domain 2 loops on contact surface with CMG2 and TEM8.

4.3.3 CMG2 binds tighter to PA domain2 than TEM8

Residues on protective antigen show distinct preferences between CMG2 and TEM8. PA binds to CMG2 and TEM8 though its domain 2 and 4. Table 4.5 shows that PA binds to CMG2 and TEM8 mainly through domain 4. PA domain 4 interacts with CMG2 at 4.08kcal/mol more than towards TEM8, considering the 85-90kcal/mol total, it's about 5% differences. PA domain 2 interacts with CMG2 and TEM8 less strongly than its domain 4, but the difference in binding free energy is 13.36kcal/mol, it is more than 50% of the 24.97 kcal/mol energy between TEM8 and PA domain 2. Hence, PA domain 2 binds to CMG2 more strongly, and PA domain 4 binds to TEM8 more strongly.

If we decompose the energy into different terms, van der Waals interaction contributes 10.20kcal/mol of the 13.66kcal/mol difference in domain 2 and 5.39kcal/mol of the 4.08kcal/mol difference in domain 4. The summation of Coulombic and generalized Born interaction energy can be considered as the electrostatic energy. It raises 2.68kcal/mol difference for domain 2 and partially cancels the difference in vdW energy by 1.65kcal/mol for domain 4.The rest of the difference is from SASA energy. Thus, the majority of the difference in binding free energy results from vdW interaction energy.

In PA domain 2, loop 340-344 is responsible for 12.12kcal/mol in the 13.66kcal/mol

difference in interaction free energy between PA domain 2 and CMG2/TEM8. The two charged residues, GLU343 and ARG 344 composed the neutral charge on this loop (Table 4.6). When bound with the neutrally charged CMG2 loops and negatively charged (-3) TEM8 loops, the electrostatic (Coulombic and generalized Born) energies on these two residues are more favorable towards CMG2. For interaction energy between each residue in this loop and PA, vdW interaction energies are stronger in PA-CMG2 complex. This observation confirms that the negatively charged loop 152-160 in TEM8 in less favored by PA domain 2.

In PA domain 4, the difference in binding free is mainly contributed by the residues on loop 652-658 (Table 4.6). Because of the large standard error in the energies of GLU654 and ASP 658, the differences caused by these two residues will not be further analyzed. The difference in LEU652 is mainly from the vdW interaction energy. Considering the differences in binding free energy cause by residues on TEM8/CMG2 SER113/LEU113 and VAL115/GLY115, this result confirms that the change in TEM8/CMG2 may have changed the distance between CMG2/TEM8 and PA, furthermore, caused the difference in interaction energy between PA domain 4 and CMG2/TEM8. This is consistent with the hydrophobic binding pocket observed in Chapter 3.

Thus, the different in CMG2/TEM8-PA binding affinity mainly results from the discrimination from PA domain 2. If PA domain2 can bind to CMG2 or TEM8 without the other domains, the difference in binding affinity would be larger than using all PA domains. To enhance the specificity towards CMG2, antibodies can be developed based on PA domain 2 or the loops on PA domain 2 that bind to CMG2 (340-344).

MM/GBSA interaction energy between PA domains and TEM8/CMG2					
Protein Domains	ΔE_{tot}	ΔE_{vDW}	ΔE_{elec}	ΔE_{GB}	ΔE_{SASA}
TEM8					
Domain 2	-24.97±0.74	-20.38±0.75	43.68±6.75	-46.62±2.80	-1.66±0.00
Domain 4	-89.31±1.47	-48.96±0.62	-71.38±15.03	35.56±6.36	-4.54±0.00
CMG2					
Domain 2	-38.33±0.51	-30.58±0.39	-38.67±4.18	33.05±1.84	-2.12±0.00
Domain 4	-85.23±1.36	-43.47±0.73	-147.25±12.27	109.78±5.23	-4.29±0.00

Table 4.5. MM/GBSA interaction energy between CMG2/TEM8 and PA domain 2/4. E_{tot} is total MM/GBSA interaction energy; E_{vdW} is the van der Waals interaction energy; E_{elec} is the Coulombic interaction energy; E_{GB} is the generalized Born interaction energy; E_{SASA} is the solvent accessible surface area interaction energy. Values are differences ± standard error. Standard errors are calculated from 10 repeats of 20ns simulation.

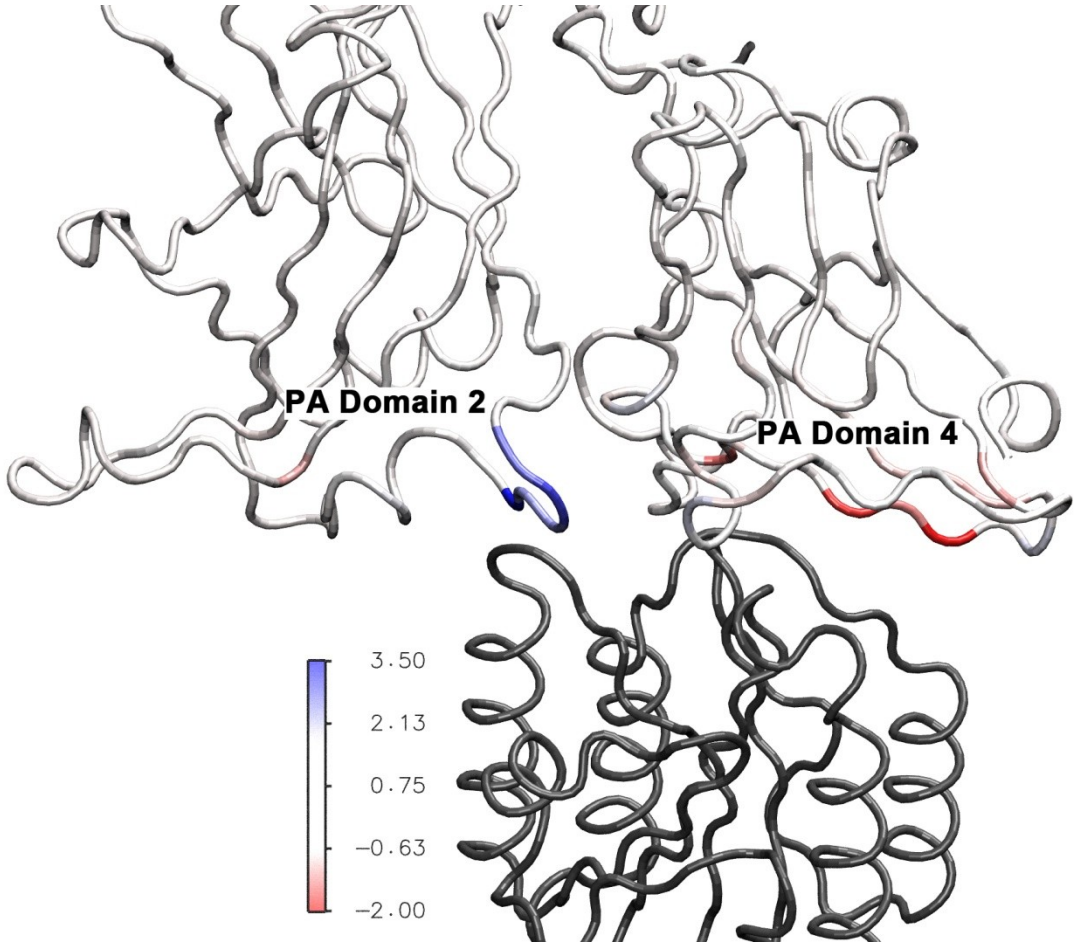


Figure 4.2 The colors for residues show the difference in MM/GBSA interaction energy between CMG2 /TEM8 and individual residues on PA. The unmarked grey protein in the lower part represents TEM8 or CMG2/TEM8. The white proteins on the top part represents protective antigen. Positive energies mean the PA residue interacts more strongly to CMG2; negative energy values mean the residues interact with TEM8 more strongly.

Difference in MM/GBSA interaction energy between TEM8/CMG2(E_{TEM8}-E_{CMG2}) and individual residues on PA					
PA Residue	ΔE_{tot}	ΔE_{vDW}	ΔE_{elec}	ΔE_{GB}	ΔE_{SASA}
Discriminate TEM8					
LEU652	-2.10±0.35	-2.07±0.35	-0.24±0.20	0.40±0.16	-0.19±0.03
GLU654	-1.89±1.62	-0.01±0.94	16.35±5.53	-18.21±3.30	-0.02±0.06
ASP658	-1.27±1.21	0.88±0.66	8.64±3.38	-10.71±1.76	-0.08±0.01
Discriminate CMG2					
GLU343	3.51±1.76	1.77±2.09	-26.91±6.13	28.49±3.62	0.16±0.05
GLY342	2.88±0.80	1.42±0.80	3.80±0.18	-2.38±0.17	0.04±0.06
LEU340	2.27±0.64	2.26±0.57	-0.07±0.20	-0.07±0.18	0.15±0.06
ALA341	1.87±0.63	1.69±0.55	0.60±0.20	-0.40±0.13	-0.02±0.02
ARG344	1.73±0.28	1.59±0.31	23.03±1.26	-22.94±1.00	0.05±0.02

Table 4.6 Residues on PA discriminate CMG2 or TEM8 with difference in MM/GBSA interaction energy between the residue and CMG2/TEM8 greater than 1kcal/mol. Negative energy values indicate the PA residues interact more strongly to CMG2; positive energies indicate the residues interact more strongly to TEM8. E_{tot} is total MM/GBSA interaction energy; E_{vdW} is the van der Waals interaction energy; E_{elec} is the Coulombic interaction energy; E_{GB} is the generalized Born interaction energy; E_{SASA} is the solvent accessible surface area interaction energy Values are differences ± standard error. Standard errors are calculated from 10 repeats of 20ns simulation.

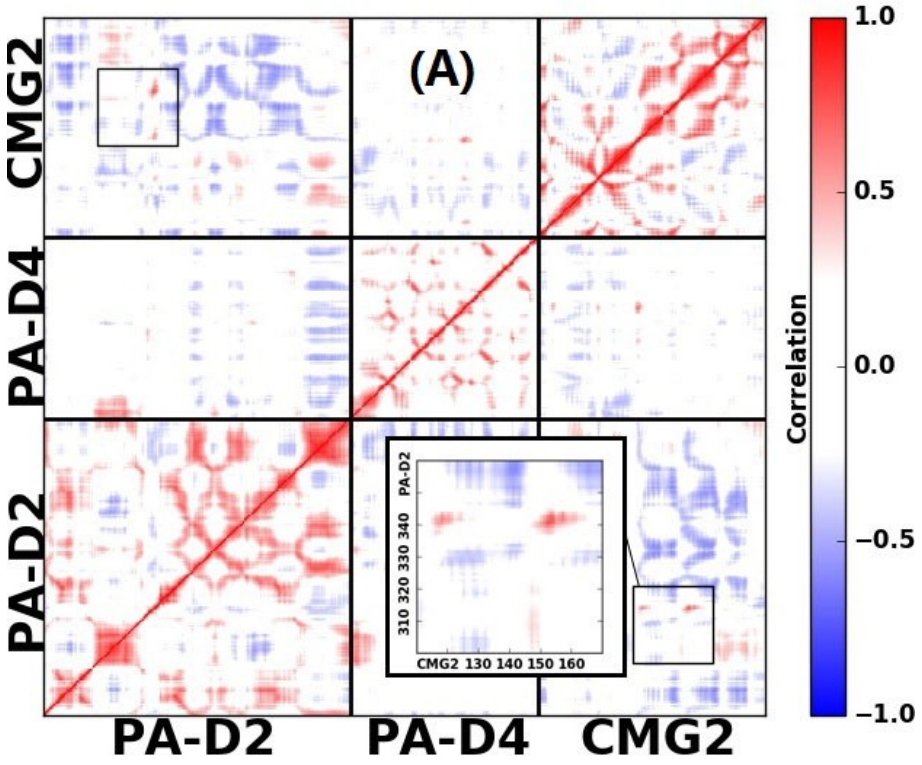
4.3.4 Correlated motion between CMG2 and PA domain 2

Covariance analysis can identify the correlated motions in protein dynamics in PA-CMG2 and PA-TEM8 complex. The highly correlated motions in groups of residues usually indicate strong binding or signaling pathway. To further investigate the differences in PA-CMG2 and PA-TEM8 binding, I analyzed the covariance in the motion of Cα atoms for each residue in the complex and shown in Figure 4.3.

In both PA-CMG2 and PA-TEM8 complexes, strong correlations were observed in each local domain of PA domain 2, 4 and CMG2/TEM8. Surprisingly, despite of the strong interaction free energy between PA domain 4 and CMG2/TEM8, almost no strong correlation was discovered between them; whereas some strong anti-correlations were discovered between CMG2/TEM8 and PA domain 2. This indicates the existence of motions conduction between CMG2/TEM8 and PA domain 2; and PA domain 4 may be mainly responsible for the binding and recognition.

The difference in covariance between PA-CMG2 and PA-TEM8 lies in the PA domain 2 (sub-figures in Figure 3 A, B). After applying the filter to screen weak correlations below 0.25, strong positive correlations were observed between the PA domain 2 loops near the contact surface (340-344) and CMG2 residue 118-120, 150-168. However, such

correlations were not observed for the corresponding residues in PA-TEM8 complex. Considering the difference in binding interaction energy contributed by PA domain 2 in PA-CMG2 and PA-TEM8 interactions, the difference in correlated motion could be a result of the lack of vdW interaction between PA domain 2 and TEM8. This may suggest TEM8 and CMG2 have distinct signaling pathway when bound with PA.



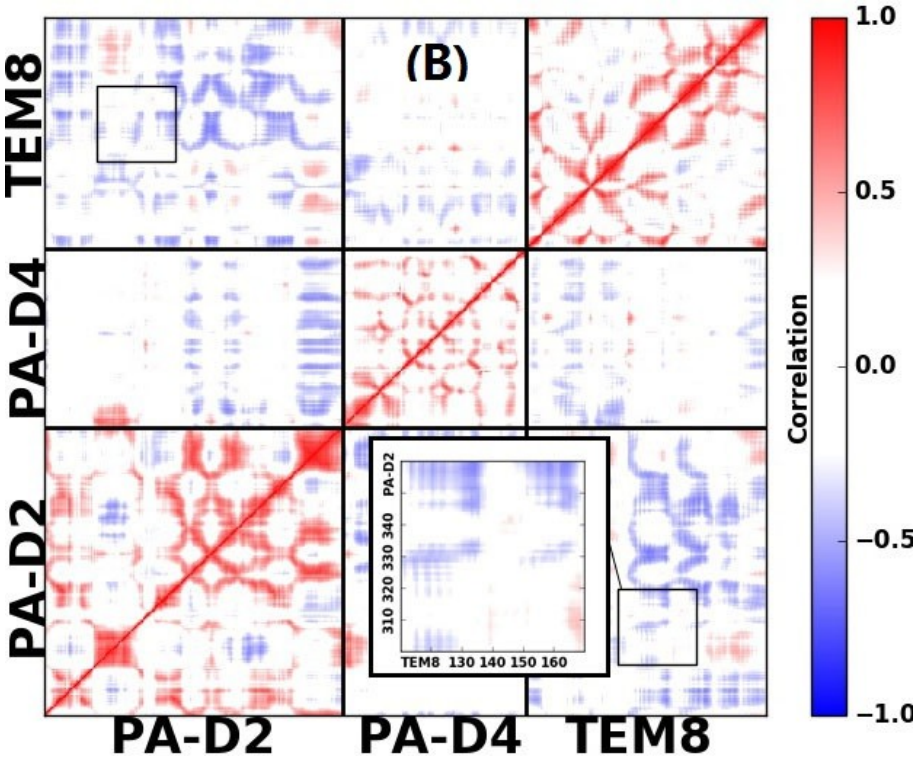


Figure 4.3 Correlation in the motion of Cα atoms in each residue of the PA-CMG2/TEM8 complex. PA-D2 is PA domain 2, PA-D4 is PA domain 4. Residue numbers are labeled in the small windows. A cut off of 0.25 is used to filter out the weakly correlated motions, correlation between -0.25 to 0.25 is not shown in the figures.

4.3.4 Metal ion in MIDAS affects PA-CMG2 binding affinity

To measure the binding interaction energy between CMG2 and PA in the presence of different metal ions, the dissociation constant was calculated in the presence of Mg^{2+} and Ca^{2+} and compared with results from SPR experiments. As shown in Table 4.7, CMG2 binds to PA more strongly in the presence of Mg^{2+} than in the presence of Ca^{2+}. Both simulation results and experimental results give the same qualitative results. The structure of MIDAS is shown in Figure 3.2A.

Stronger electrostatic interaction ($\Delta E_{ele}+\Delta G_{GB}$) was observed in the presence of Mg^{2+} comparing to Ca^{2+}. Weaker van der Waals interactions were observed in the presence of Mg^{2+}. The major difference comes from the Coulombic interaction. On the contrary, the GB energies do not show much difference. The GB energy mimics the solvation screening effect to partially cancel out more than 80% of the Coulombic term. Surprisingly, the GB energy only cancels more 0.6kcal.mol electrostatic energy, comparing to the 10.4kcal/mol difference in Coulombic energy. It indicates that the replacement of metal ion affects the interaction between CMG2 and PA more, and less to the dissolvation process.

The distances between Mg^{2+} and all the coordinating oxygen atoms are shorter than those

between Ca^{2+} and the oxygen atoms (Table 4.8), very similar to the results for TEM8-PA binding (Figure 3.2C). The results can be also interpreted as Ca^{2+} expands the size of MIDAS domain and leads to weaker interactions with residues.

Metal ion	ΔE_{ele}	ΔE_{vdw}	ΔG_{SA}	ΔG_{GB}	$\Delta G_{cal.}$	ΔG_{exp}
Mg^{2+}	-185.31±3.42	-74.04±0.92	-13.94±0.10	163.17±2.72	-110.12±0.87	-14.05±0.08
Ca^{2+}	-175.73±5.20	-75.99±1.01	-13.76±0.11	162.53±4.37	-102.95±0.95	-13.43±0.03

Table 4.7 Experimental and calculated interaction energies (kcal/mol) for PA-CMG2 binding system. The standard error was estimated over the mean of 10 repeats, each repeat has 3000 data points. E_{ele} is the Columbic energy, E_{vdw} is the van der Waals energy, G_{SA} is the non-polar solvation free energy, G_{GB} is the polar solvation free energy, G_{cal} is the calculated free energy, G_{exp} is the experimental free energy.

Metal Ion/ Distance(Å)	ASP344	SER52	SER54	THR118	WATERA	WATERB
Mg^{2+}	1.82±0.01	2.07±0.01	2.09±0.01	2.09±0.01	1.98±0.01	1.93±0.01
Ca^{2+}	2.17±0.01	2.33±0.01	2.33±0.01	2.29±0.01	2.28±0.01	2.25±0.01

Table. 4.8 Distance between the metal ion in MIDAS and the coordination oxygen atoms in MIDAS residues.

CONCLUSION

CMG2 binds to PA better than TEM8 in the present of either Mg^{2+} or Ca^{2+} cation. The stronger binding affinity mainly results from the difference in α4 helices in CMG2 and TEM8, and also a few differences in residues on the binding surface such as residue 87, 88, 113, 115 on both CMG2 and TEM8. Mutations on these residues could change the binding affinity, and possibly reverse the specificity of PA to CMG2 and TEM8.

PA domain 2 interacts more strongly towards CMG2 than to TEM8, because of the neutrally charged α4 helix in CMG2 is more favored by the loops on PA domain 2 binding surface. Based on the difference in affinity, mutations that neutralize the negative overall charges on TEM8 α4 helix can possibly increase the binding affinity between TEM8 and PA. On the other hand, PA domain 4 interacts more strongly towards TEM8, especially that L113/S113-V115/G115 in TEM8/CMG2 contributes the most difference.

The correlated motion between PA domain 2 and CMG2 was not found in TEM8-PA complex. This may results from the absence of strong interaction between TEM8 and PA domain 2. If the interaction between CMG2 and PA domain 2 is responsible for signaling process, such process is possibly missing or differently organized in TEM8 and PA interaction.

Hypothesizes of binding affinity in PA and CMG2/TEM8 mutant interaction can be further confirmed by experiments. The specificity of antibodies targeting CMG2 may be improved if α4 helix on CMG2 is selected as the binding site. Antibodies targeting the L113/S113-V115/G115 loop may have the preference towards TEM8.

APPENDIX

DATASET: TARGETED MD STUDY OF INTEGRIN α_M SIGNALING PATHWAY

5.1 Background

Integrins are another group of proteins share the same family with anthrax receptors. Integrins get their name from their function, which is integrating the extracellular and intracellular environments. They also bind to signal molecules inside the cells and ligands outside the cells[136]. The signaling processes they involve are heavily relied by the immune system and many other physiology systems[156]. Integrins are known to have α and β subunits, which are coupled to in the upstream and downstream signaling pathways[157].

More than half of integrin α subunits contain a von Willebrand factor A domain, which have about 200 amino acid, and are similar to CMG2 and TEM8 mentioned in the previous chapters. These domains are known as an inserted domain, or I domain. Integrin α I domain is the major ligand binding site in integrins. Several crystal structures of independent integrin I domain have been obtains with or without ligand bound to them. Integrin α I domain binds to their ligands through a highly conserved metal ion dependent adhesion site (MIDAS). The MIDAS in integrin α I domain is identical to the one in anthrax receptors. In MIDAS, a divalent metal ion is ligated by five side chains (Figure 5.1).

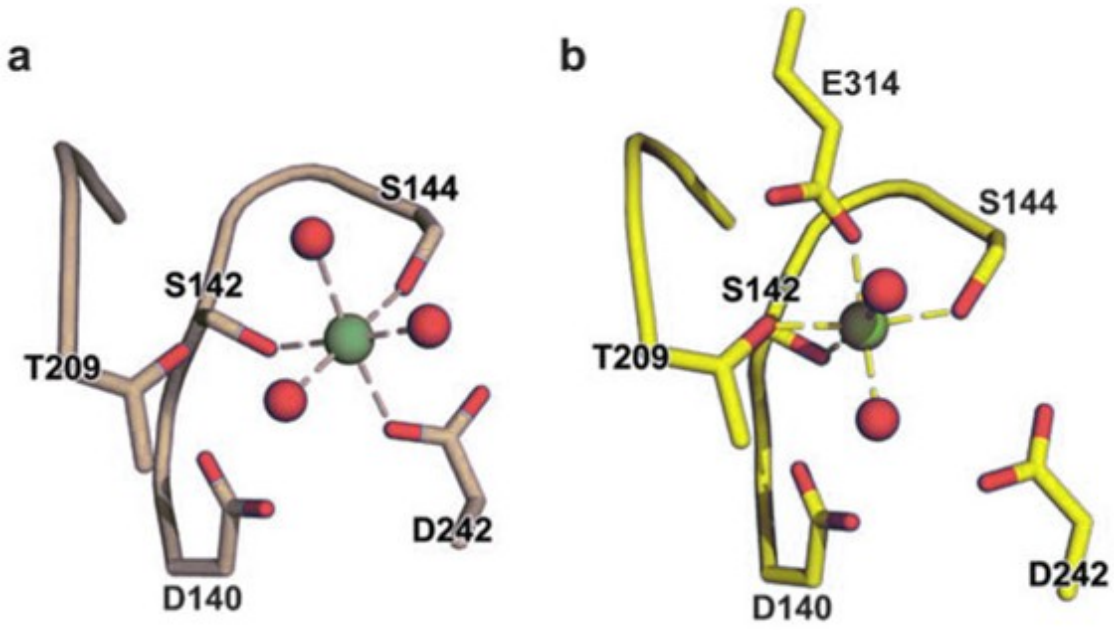


Figure 5.1 Structures of the αM I domain MIDAS in closed conformation (a) and open conformation (b). E314 is from a ligand coordinates with the MIDAS magnesium. Green particles are the divalent metal ions in MIDAS, and red spheres are oxygen atoms in coordinating water-molecules. [From PDB 1JLM[158] and 1IDO[159]]

Integrin α I domains have been crystallized in three distinct conformations: open, intermediate, and closed[137,160]. These conformations demonstrate high, intermediate, and low binding affinity to their ligands, respectively. The arrangement of MIDAS, β6-α7 loop, and C-terminal loops are believed to be correlated to the change in conformation and affinity[161]. The rearrangements of MIDAS in integrin α I domain are accompanied by a 2.3Å shift away from the metal ion THR209 in open to closed conformation change. The closed to open conformation change is believed to initiated by the pulling of C-terminal residues and therefore resulted the rearrangement in MIDAS residues[161].

However, the detailed pathway conformation change has not been found yet.

The displacement of residues on α7 helix, especially PHE302, regulates the critical[160] hydrophobic pocket that stabilize the conformation in open or closed. Engineered disulfide bond introduced to lock the hydrophobic pocket can stabilized the open conformation[137]. An engineered intermediate state has been crystallized for integrin α_L, which has the C-terminal in open conformation but MIDAS in closed conformation[162]. Steered molecular dynamics studies also showed the existence of intermediate state of C-terminal in the pathway from close to open conformation[127]. However, the mechanism of the hydrophobic pocket has not been investigated in details.

In this data set, we preformed a targeted molecular dynamics to simulate the integrin conformation process. The sequence of displacement on key residues was measured to demonstrate the mechanism of conformation regulation. The popularities of open, intermediate, and closed conformation were calculated to investigate the energy barriers in the conformation transition process. We also studied the correlated motions in conformation change to find out the transmission of conformational signals. In addition, we made mutations on the α7 helix and C-terminals to further confirm the existence of conformation change in anthrax receptors.

5.2 Method

5.2.1 Preparation of Complex

The crystal structures of integrin α_M in open[159] and closed[158] were used as initial structure for molecular dynamics simulation. Different numbers of residues were included in the crystal structures. For consistency, we truncated the residues that were not shared in these two crystal structures. The truncated residues are on the C and N terminals, and are not considered as the residues highly correlated to the binding of conformation changes[161].

Mutations on α7 helix and C-terminal were made using the mutation module in Modeller9.11[81]. 1000 cycles of energy iteration were conducted to obtain the optimized structure.

5.2.2 Molecular Dynamics simulation

To minimize the possible errors resulting from truncation, the initial atomic coordinates of integrin α_M I domains were first subjected to an energy minimization in vacuum to remove energy clashes. The energy minimization has 10 cycles of 1,000 steps, reducing the harmonic restraint each cycle on all protein atoms from 10 $kcal \cdot mol^{-1} \cdot Å^{-2}$ to 1

kcal·mol^{-1}·Å^{-2} in decrements of 1 kcal·mol^{-1}·Å^{-2} under Steepest Descent method in CHARMM 35b6 software package[83] and charmm27 force field parameters[46]. The protein was then dissolved in a TIP3[84] water and 0.15mol·L^{-1} NaCl box of 86Å×86Å ×86Å at a mixed density of 0.951g/cm^3. Extra chloride anions were used to neutralize the positive charge of the protein complex.

After the protein solvent was generated, further energy minimization and molecular dynamics simulations of protein solution were run using NAMD 2.10-GPU software package[85] with charmm27 force field parameters. Two cycles of 10000 steps of Conjugated Gradient minimization were run NAMD with fixed protein atoms and without constraints, respectively. Finally, 50 ns of targeted molecular dynamics was conducted using a spring constant k=200 in an artificial potential

– –

where RMS(t) is the instantaneous best-fit RMS distance of the current coordinates from the target coordinates, RMS*(t) evolves linearly from the initial RMSD at the first TMD step to the final RMSD at the last TMD step, and N is the number of targeted atoms. The structure of closed conformation was used as a target when using open conformation as initial structure, or vise versa. Temperature was set at 300k with a damping coefficient of 5ps^{-1} using Langevin dynamics[86]. Pressure was set at 1 bar using a Langevin-Noose

Hoover piston[87] with a damping time of $50ps^{-1}$. Particle Mesh Ewald[88] was used to calculate long-range electrostatic interactions. The time step was set at 2fs with the use of Rigid Bond algorithm[89] between hydrogen atoms and heavy atoms.

5.3 Result and discussion

5.3.1 Movement of residues regulating conformation change

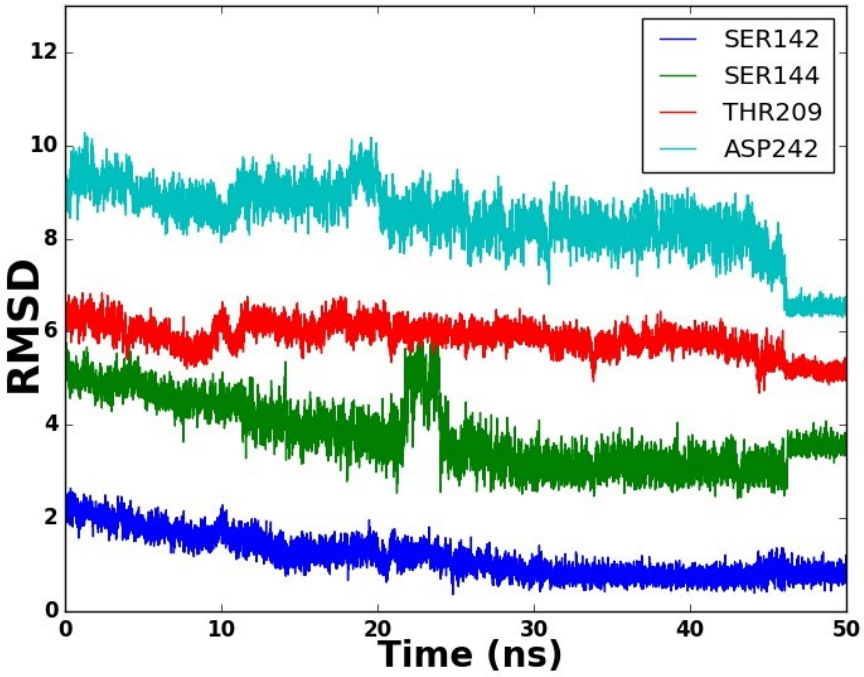


Figure 5.2 RMSD of MIDAS residue in wild type integrin αM I domain during open to closed conformation change.

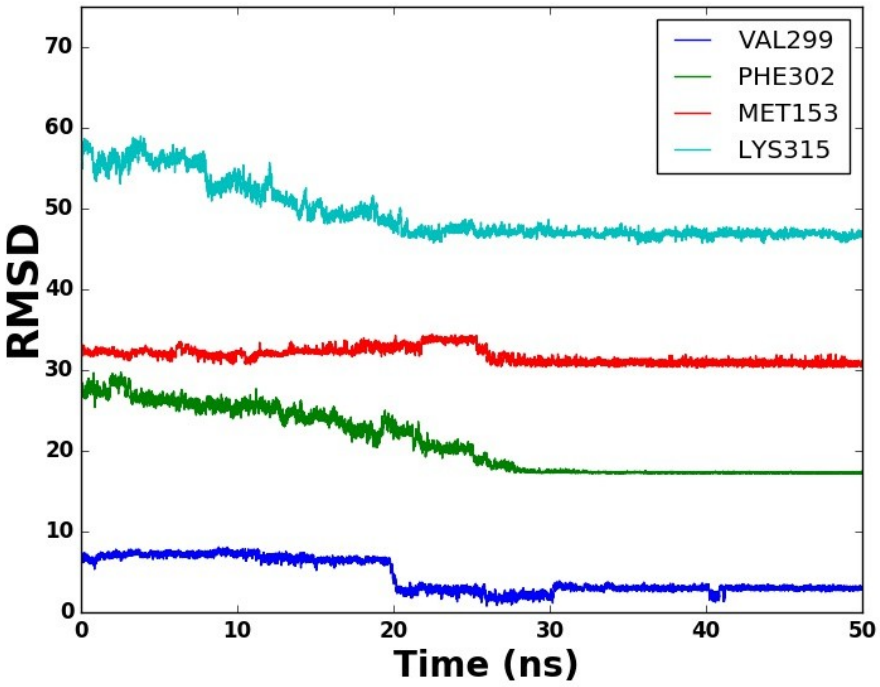


Figure 5.3 RMSD of C-ternimal residue in wild type integrin αM I domain during open to closed conformation change.

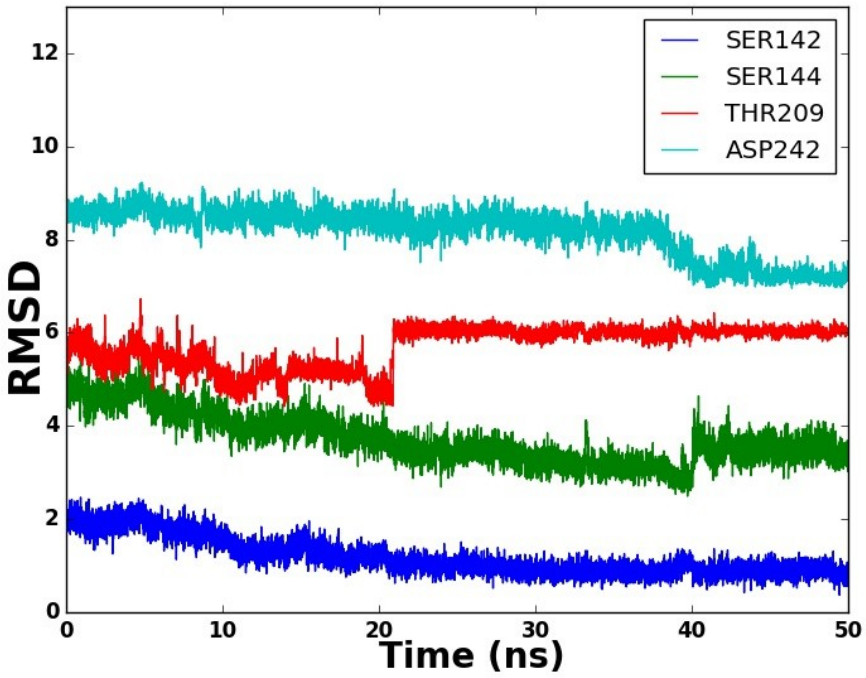


Figure 5.4 RMSD of MIDAS residue in wild type integrin αM I domain during closed to open conformation change.

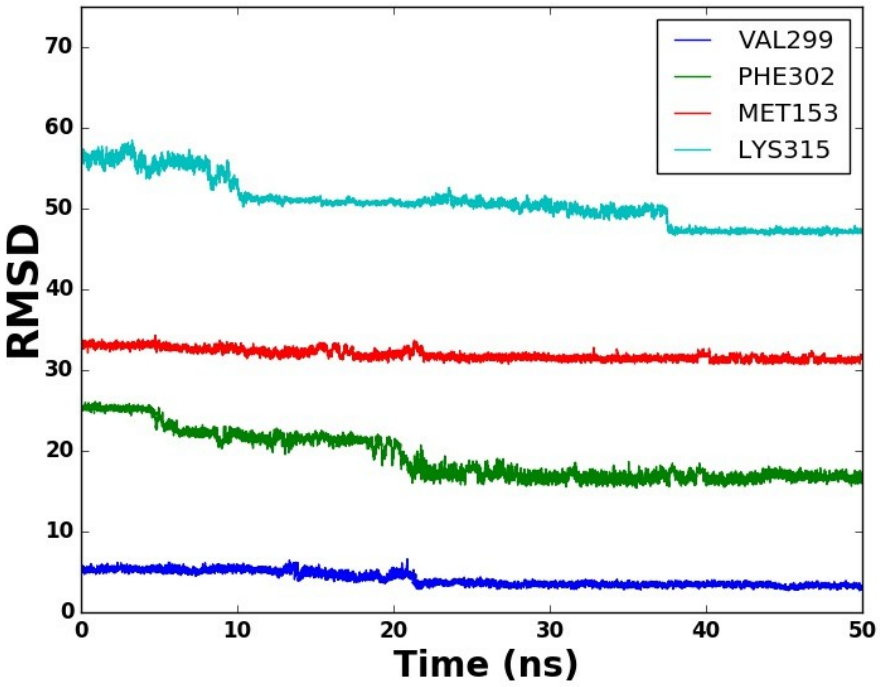


Figure 5.5 RMSD of C-terminal residue in wild type integrin αM I domain during closed to open conformation change.

5.3.2 The existence of intermediate state

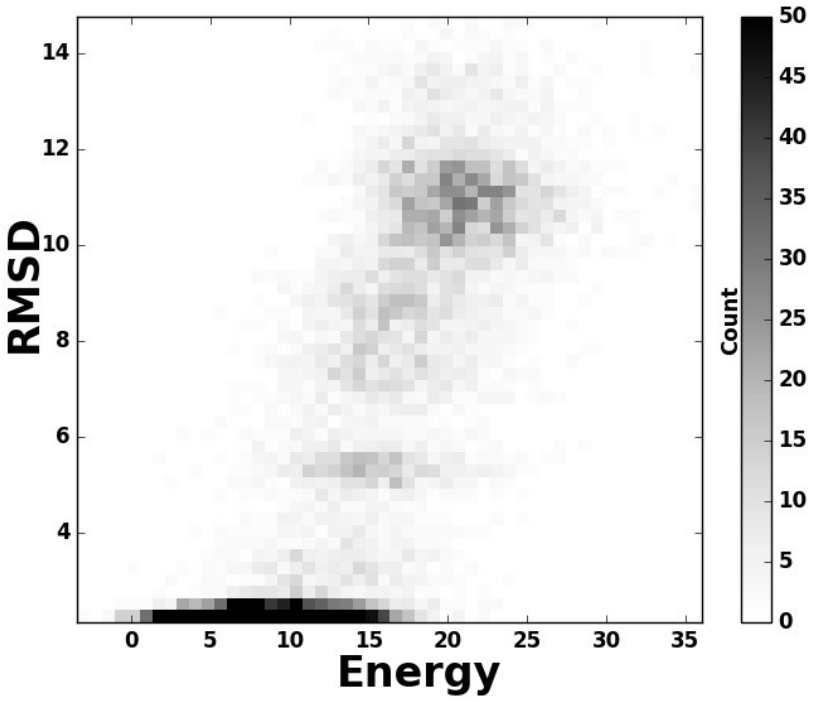


Figure 5.6 Population of conformations during open to closed conformation change of wild type integrin αM I domain.

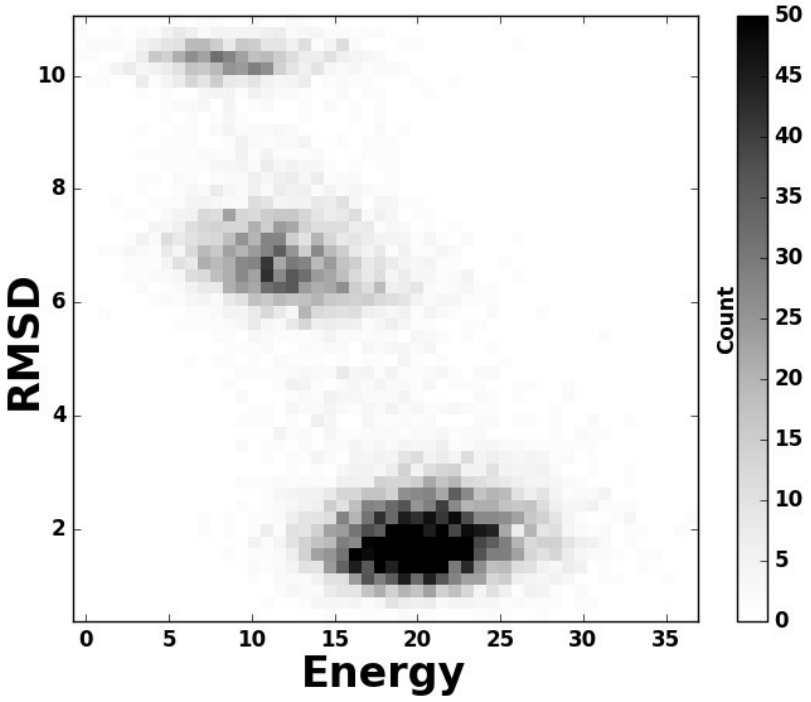


Figure 5.7 Population of conformations during closed to open conformation changeof wild type integrin αM I domain.

5.3.3 Correlated motion in conformation change

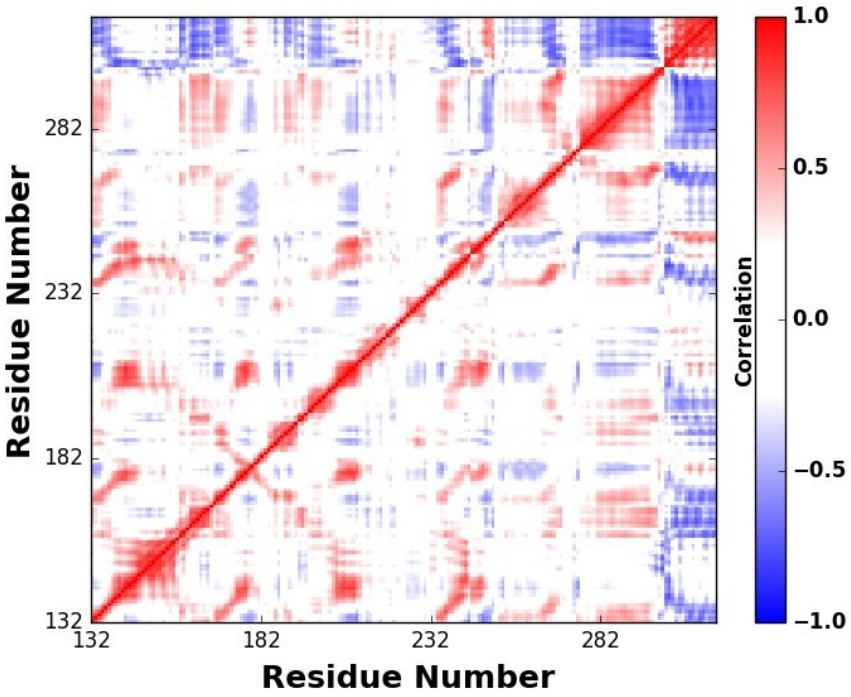


Figure 5.8 Correlated motion in residues during open to closed conformation change of wild type integrin αM I domain.

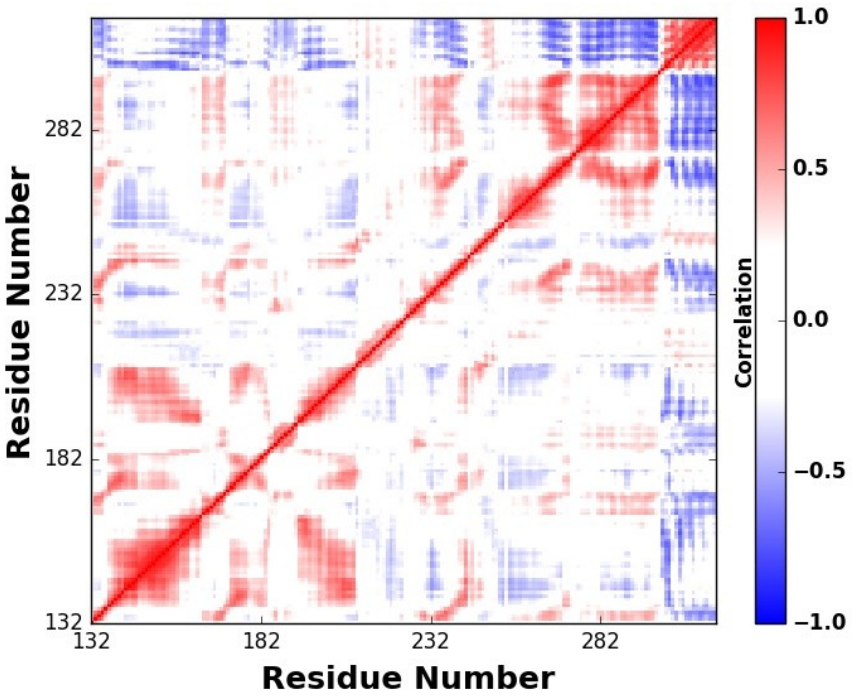


Figure 5.9 Correlated motion in residues during closed to open conformation change of wild type integrin αM I domain.

5.3.4 F302W mutation study

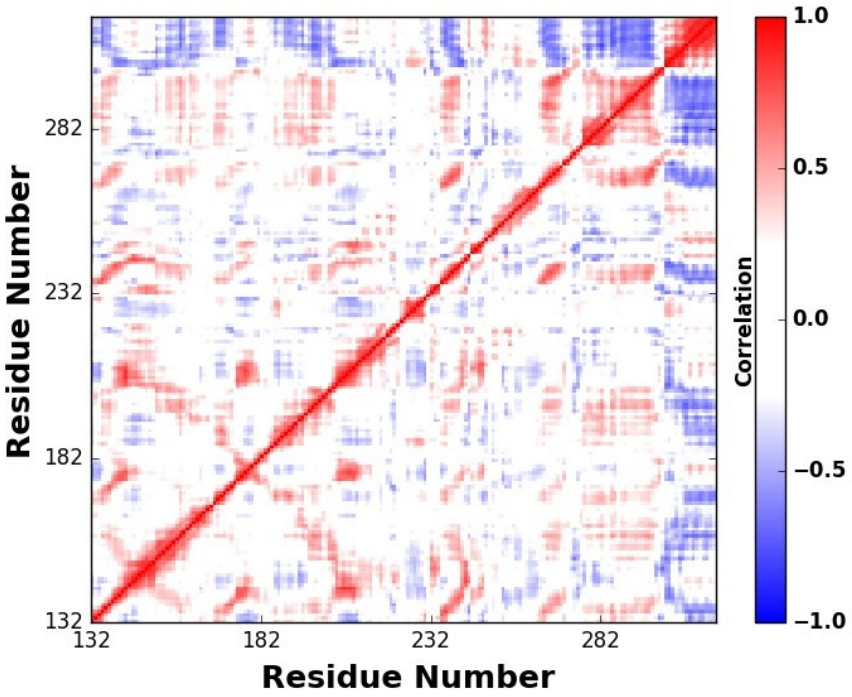


Figure 5.10 Correlated motion in residues during open to closed conformation change of integrin αM I domain F302W mutant.

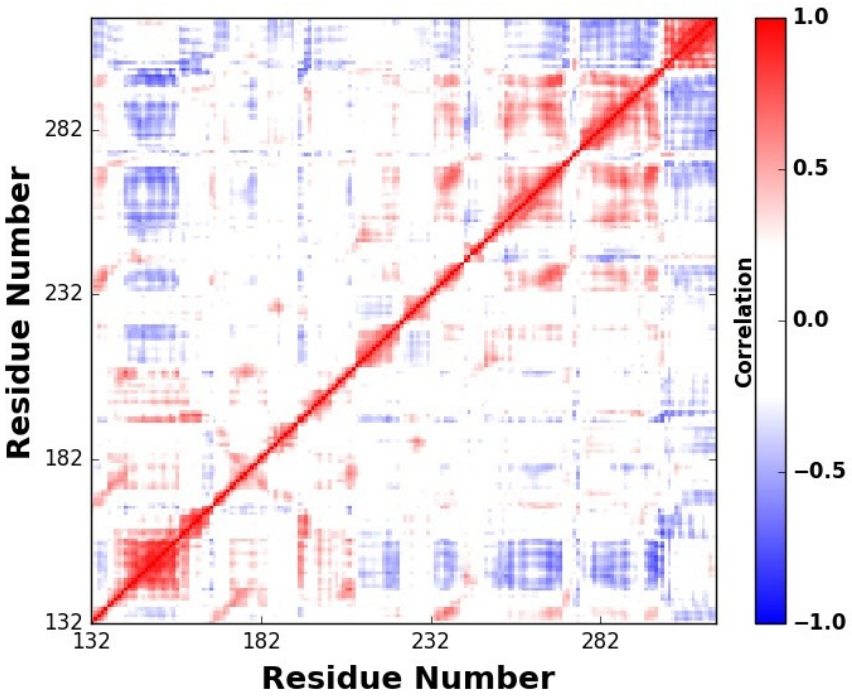


Figure 5.11 Correlated motion in residues during closed to open conformation change of integrin αM I domain F302W mutant.

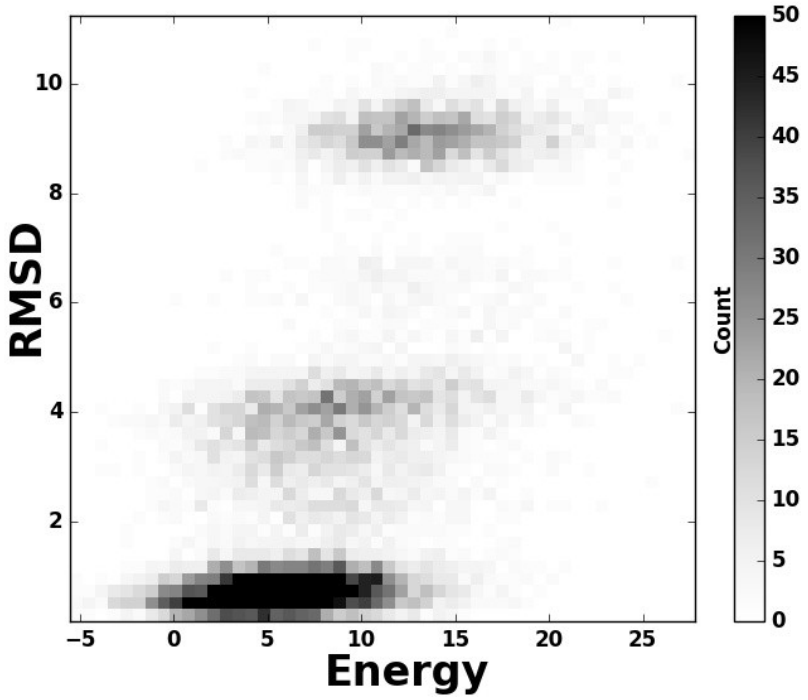


Figure 5.12 Population of conformations during open to closed conformation change of integrin αM I domain F302W mutant.

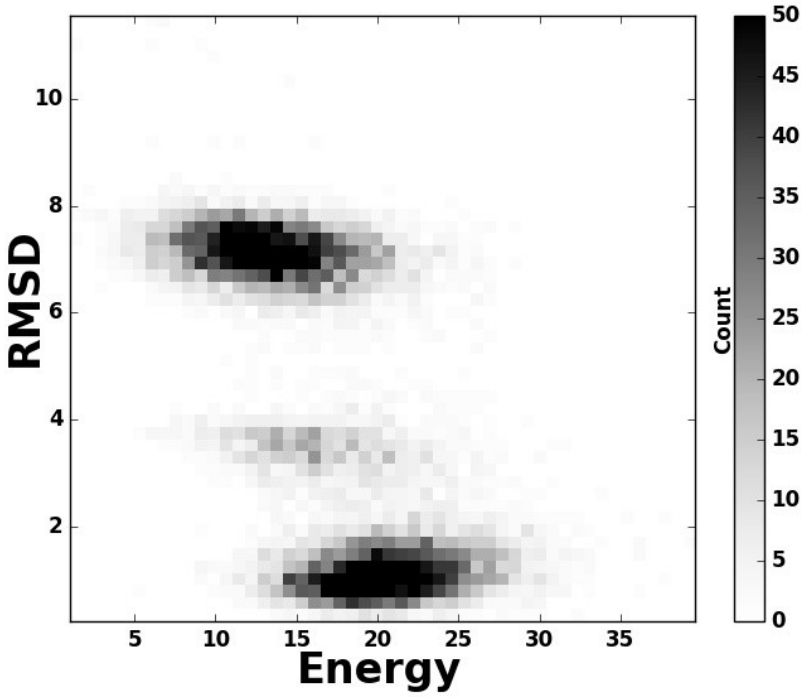


Figure 5.13 Population of conformations during closed to open conformation change of integrin αM I domain F302W mutant.

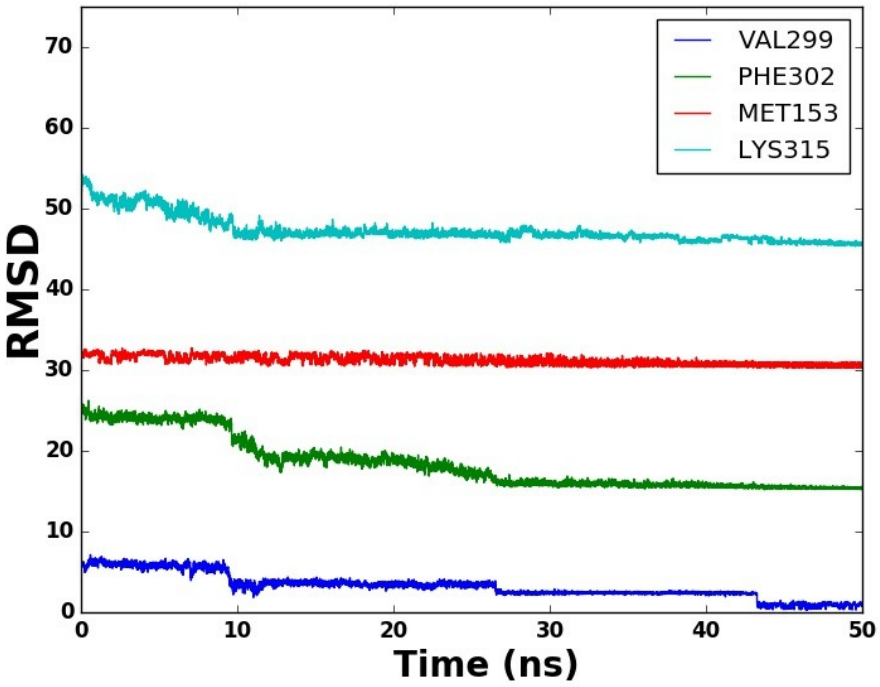


Figure 5.14 RMSD of C-ternimal residue in integrin αM I domain F302W mutant during open to closed conformation change.

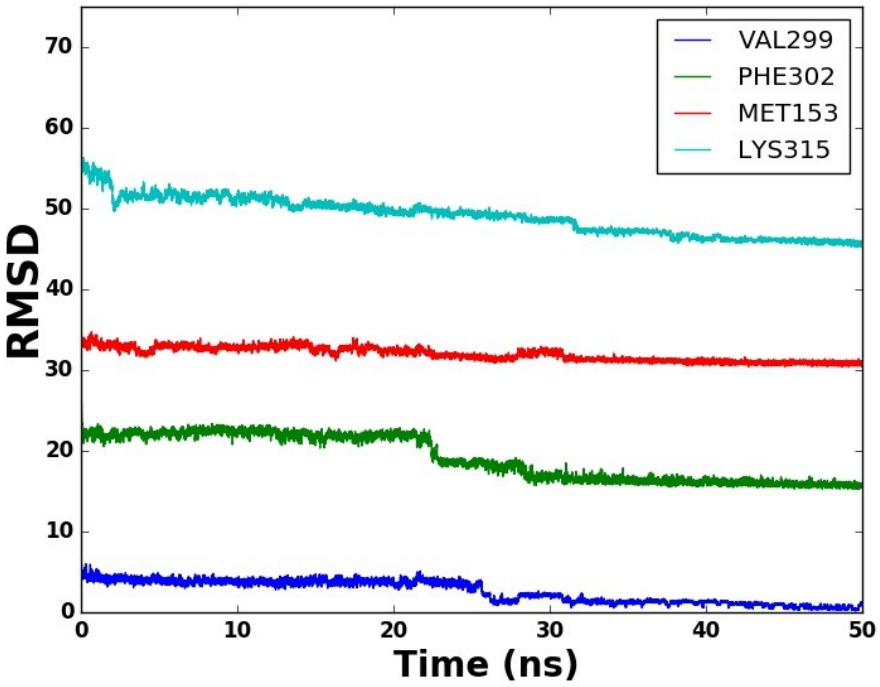


Figure 5.15 RMSD of C-ternimal residue in integrin αM I domain F302W mutant during closed to open conformation change.

5.3.5 TEM8-like mutation study

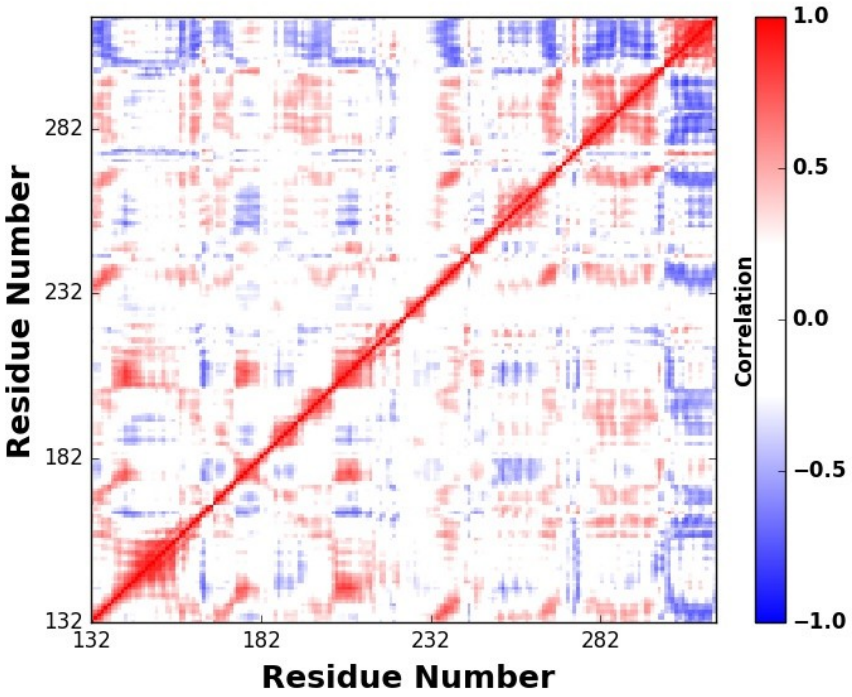


Figure 5.16 Correlated motion in residues during open to closed conformation change of integrin αM I domain TEM8-like mutant.

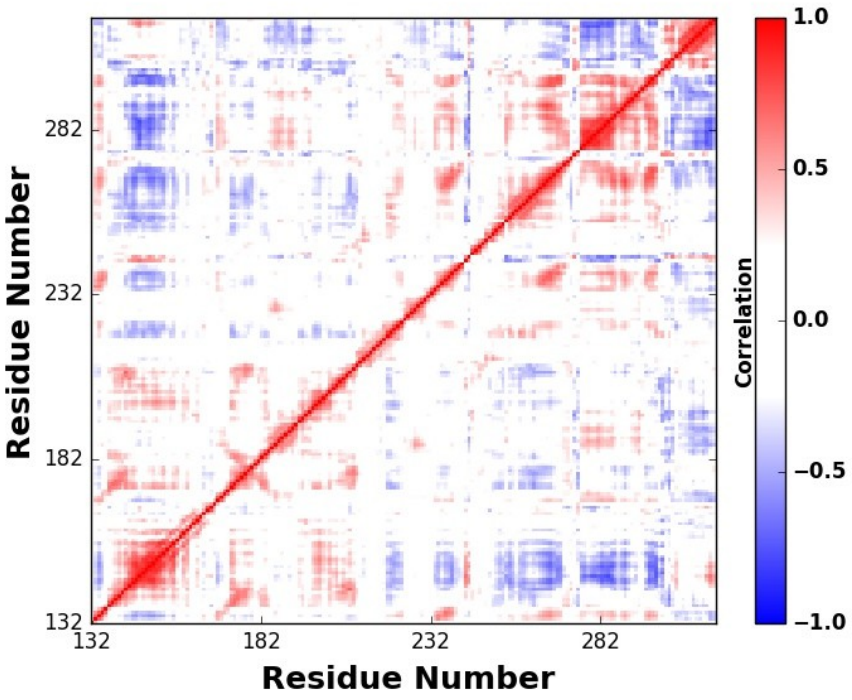


Figure 5.17 Correlated motion in residues during closed to open conformation change of integrin αM I domain TEM8-like mutant.

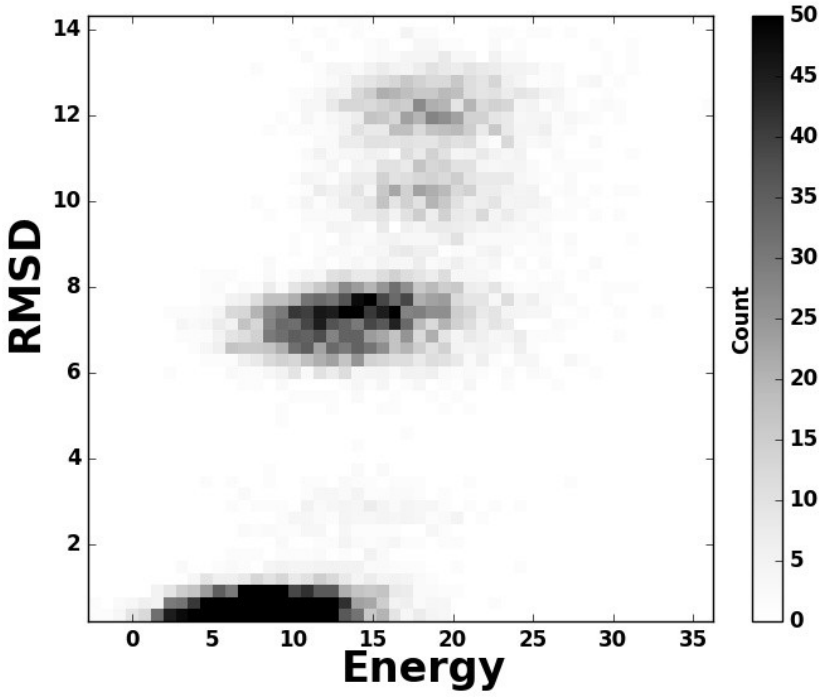


Figure 5.18 Population of conformations during open to closed conformation change of integrin αM I domain TEM8-like mutant.

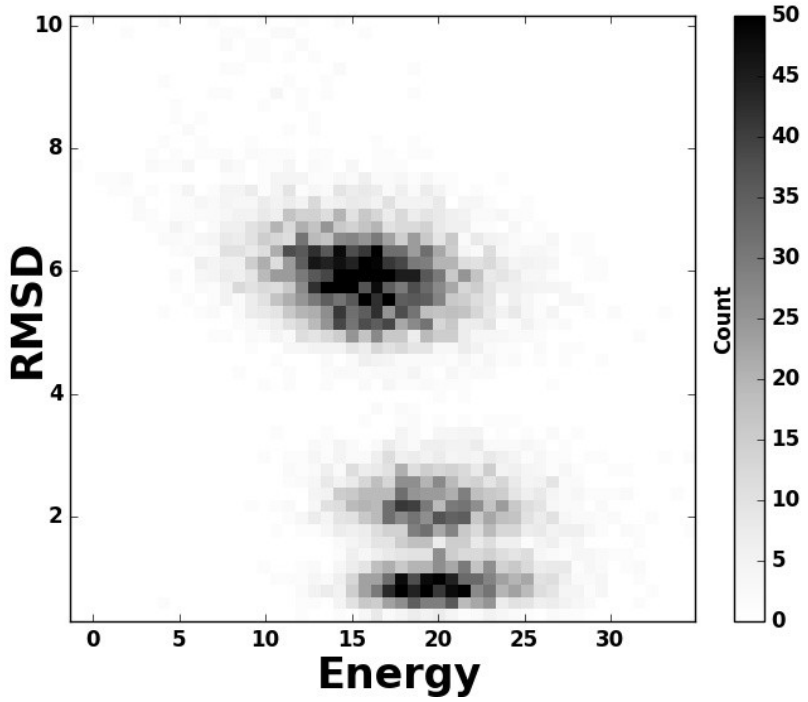


Figure 5.19 Population of conformations during closed to open conformation change of integrin αM I domain TEM8-like mutant.

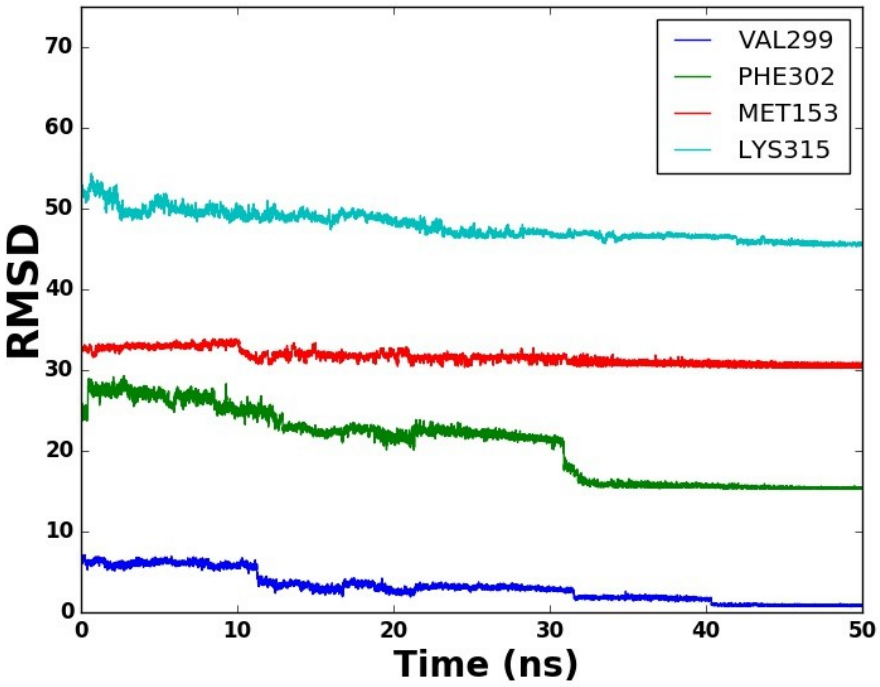


Figure 5.20 RMSD of C-ternimal residue in integrin αM I domain TEM8-like mutant during open to closed conformation change.

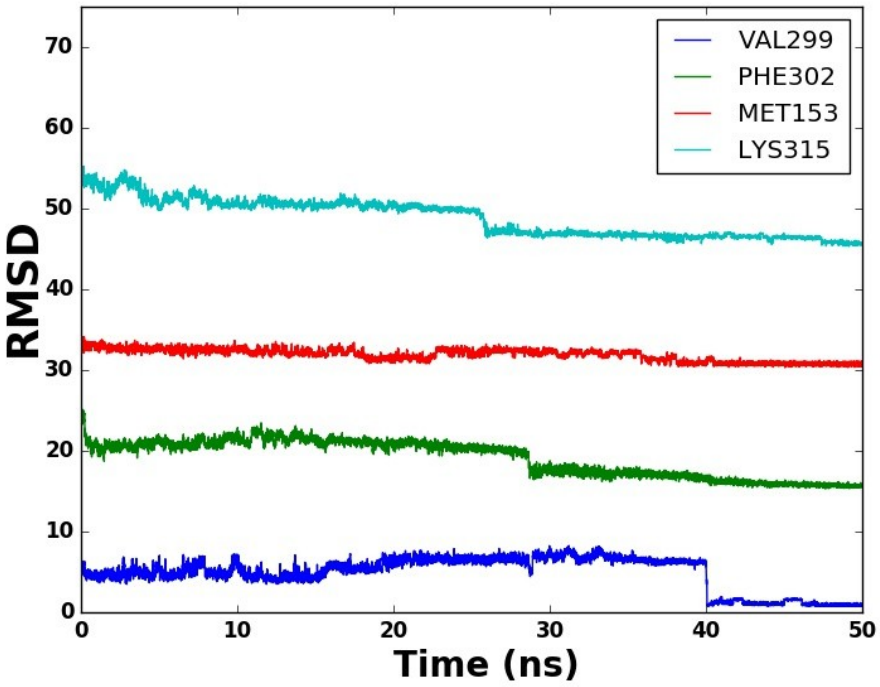


Figure 5.21 RMSD of C-ternimal residue in integrin αM I domain TEM8-like mutant during closed to open conformation change.

BIBLIOGRAPHY

(1) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys J* **1997**, *72*, 1047.

(2) Luo, H.; Sharp, K. *Proc Natl Acad Sci U S A* **2002**, *99*, 10399.

(3) Woo, H. J.; Roux, B. *Proc Natl Acad Sci U S A* **2005**, *102*, 6825.

(4) Ascenzi, P.; Visca, P.; Ippolito, G.; Spallarossa, A.; Bolognesi, M.; Montecucco, C. *FEBS Lett* **2002**, *531*, 384.

(5) Collier, R. J.; Young, J. A. *Annu Rev Cell Dev Biol* **2003**, *19*, 45.

(6) Chaudry, G. J.; Moayeri, M.; Liu, S.; Leppla, S. H. *Trends Microbiol* **2002**, *10*, 58.

(7) Young, J. A.; Collier, R. J. *Annu Rev Biochem* **2007**, *76*, 243.

(8) Mock, M.; Fouet, A. *Annu Rev Microbiol* **2001**, *55*, 647.

(9) Bradley, K. A.; Young, J. A. *Biochem Pharmacol* **2003**, *65*, 309.

(10) Scobie, H. M.; Rainey, G. J.; Bradley, K. A.; Young, J. A. *Proc Natl Acad Sci U S A* **2003**, *100*, 5170.

(11) Lacy, D. B.; Wigelsworth, D. J.; Scobie, H. M.; Young, J. A.; Collier, R. J. *Proc Natl Acad Sci U S A* **2004**, *101*, 6367.

(12) Blaustein, R. O.; Koehler, T. M.; Collier, R. J.; Finkelstein, A. *Proc Natl Acad Sci U S A* **1989**, *86*, 2209.

(13) Petosa, C.; Collier, R. J.; Klimpel, K. R.; Leppla, S. H.; Liddington, R. C. *Nature* **1997**, *385*, 833.

(14) Liddington, R.; Pannifer, A.; Hanna, P.; Leppla, S.; Collier, R. J. *J Appl Microbiol* **1999**, *87*, 282.

(15) Brossier, F.; Mock, M.; Sirard, J. C. *J Appl Microbiol* **1999**, *87*, 298.

(16) Carson-Walter, E. B.; Watkins, D. N.; Nanda, A.; Vogelstein, B.; Kinzler, K. W.; St Croix, B. *Cancer Res* **2001**, *61*, 6649.

(17) Bradley, K. A.; Mogridge, J.; Mourez, M.; Collier, R. J.; Young, J. A. *Nature* **2001**, *414*, 225.

(18) Bell, S. E.; Mavila, A.; Salazar, R.; Bayless, K. J.; Kanagala, S.; Maxwell, S. A.; Davis, G. E. *J Cell Sci* **2001**, *114*, 2755.

(19) Hanks, S.; Adams, S.; Douglas, J.; Arbour, L.; Atherton, D. J.; Balci, S.; Bode, H.; Campbell, M. E.; Feingold, M.; Keser, G.; Kleijer, W.; Mancini, G.; McGrath, J. A.; Muntoni, F.; Nanda, A.; Teare, M. D.; Warman, M.; Pope, F. M.; Superti-Furga, A.; Futreal, P. A.; Rahman, N. *Am J Hum Genet* **2003**, *73*, 791.

(20) Dowling, O.; Difeo, A.; Ramirez, M. C.; Tukel, T.; Narla, G.; Bonafe, L.; Kayserili, H.; Yuksel-Apak, M.; Paller, A. S.; Norton, K.; Teebi, A. S.; Grum-Tokars, V.; Martin, G. S.; Davis, G. E.; Glucksman, M. J.; Martignetti, J. A. *Am J Hum Genet* **2003**, *73*, 957.

(21) Scobie, H. M.; Thomas, D.; Marlett, J. M.; Destito, G.; Wigelsworth, D. J.; Collier, R. J.; Young, J. A.; Manchester, M. *J Infect Dis* **2005**, *192*, 1047.

(22) Nanda, A.; Carson-Walter, E. B.; Seaman, S.; Barber, T. D.; Stampfl, J.; Singh, S.; Vogelstein, B.; Kinzler, K. W.; St Croix, B. *Cancer Res* **2004**, *64*, 817.

(23) Bradley, K. A.; Mogridge, J.; Rainey, G. J. A.; Batty, S.; Young, J. A. T. *Journal of Biological Chemistry* **2003**, *278*, 49342.

(24) Wigelsworth, D. J.; Krantz, B. A.; Christensen, K. A.; Lacy, D. B.; Juris, S. J.; Collier, R. J. *J Biol Chem* **2004**, *279*, 23349.

(25) Santelli, E.; Bankston, L. A.; Leppla, S. H.; Liddington, R. C. *Nature* **2004**, *430*, 905.

(26) Hotchkiss, K. A.; Basile, C. M.; Spring, S. C.; Bonuccelli, G.; Lisanti, M. P.; Terman, B. I. *Exp Cell Res* **2005**, *305*, 133.

(27) Liu, S.; Crown, D.; Miller-Randolph, S.; Moayeri, M.; Wang, H.; Hu, H.; Morley, T.; Leppla, S. H. *Proc Natl Acad Sci U S A* **2009**, *106*, 12424.

(28) Lee, J. O.; Bankston, L. A.; Arnaout, M. A.; Liddington, R. C. *Structure* **1995**, *3*, 1333.

(29) Fu, S.; Tong, X.; Cai, C.; Zhao, Y.; Wu, Y.; Li, Y.; Xu, J.; Zhang, X. C.; Xu, L.; Chen, W.; Rao, Z. *PLoS One* **2010**, *5*, e11203.

(30) Burgoyne, N. J.; Jackson, R. M. *Bioinformatics* **2006**, *22*, 1335.

(31) Liu, S.; Redeye, V.; Kuremsky, J. G.; Kuhnen, M.; Molinolo, A.; Bugge, T. H.; Leppla, S. H. *Nat Biotechnol* **2005**, *23*, 725.

(32) Rogers, M. S.; Christensen, K. A.; Birsner, A. E.; Short, S. M.; Wigelsworth, D. J.; Collier, R. J.; D'Amato, R. J. *Cancer Res* **2007**, *67*, 9980.

(33) Cai, C.; Che, J.; Xu, L.; Guo, Q.; Kong, Y.; Fu, L.; Xu, J.; Cheng, Y.; Chen, W. *PLoS One* **2011**, *6*, e20646.

(34) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd *Accounts of chemical research* **2000**, *33*, 889.

(35) Greenidge, P. A.; Kramer, C.; Mozziconacci, J. C.; Wolf, R. M. *Journal of Chemical Information and Modeling* **2013**, *53*, 201.

(36) Zwanzig, R. W. *Journal of Chemical Physics* **1954**, *22*, 1420.

(37) Kirkwood, J. G. *Journal of Chemical Physics* **1935**, *3*, 300.

(38) Hansson, T.; Marelius, J.; Aqvist, J. *J Comput Aided Mol Des* **1998**, *12*, 27.

(39) Foloppe, N.; Hubbard, R. *Current medicinal chemistry* **2006**, *13*, 3583.

(40) Hou, T.; Wang, J.; Li, Y.; Wang, W. *J Chem Inf Model* **2011**, *51*, 69.

(41) Li, M. H.; Petukh, M.; Alexov, E.; Panchenko, A. R. *Journal of chemical theory and computation* **2014**, *10*, 1770.

(42) Kodchakorn, K.; Dokmaisrijan, S.; Chong, W. L.; Payaka, A.; Wisitponchai, T.; Nimmanpipug, P.; Zain, S. M.; Abd Rahman, N.; Tayapiwatana, C.; Lee, V. S. *Integr Ferroelectr* **2014**, *156*, 137.

(43) Dominy, B. N.; Brooks, C. L. *Proteins-Structure Function and Genetics* **1999**, *36*, 318.

(44) Ponder, J. W.; Case, D. A. *Adv Protein Chem* **2003**, *66*, 27.

(45) Mackerell, A. D. *Journal of Computational Chemistry* **2004**, *25*, 1584.

(46) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *Journal of Physical Chemistry B* **1998**, *102*, 3586.

(47) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1996**, *118*, 2309.

(48) Jorgensen, W. L.; Tiradorives, J. *Journal of the American Chemical Society* **1988**, *110*, 1657.

(49) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *Journal of Computational Chemistry* **2004**, *25*, 1656.

(50) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd *J Comput Chem* **2004**, *25*, 1400.

(51) Israelachvili, J. N. *Intermolecular and Surface Forces, 3rd Edition* **2011**, 1.

(52) Leach, A. R. *Molecular modelling : principles and applications*; 2nd ed.; Prentice Hall: Harlow, 2001.

(53) Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* **1993**, *98*, 10089.

(54) Born, M.; Infeld, L. *P R Soc Lond a-Conta* **1934**, *144*, 0425.

(55) Westergren, J.; Lindfors, L.; Hoglund, T.; Luder, K.; Nordholm, S.; Kjellander, R. *J Phys Chem B* **2007**, *111*, 1872.

(56) Hermann, R. B. *J Phys Chem-Us* **1972**, *76*, 2754.

(57) Sitkoff, D.; Sharp, K. A.; Honig, B. *J Phys Chem-Us* **1994**, *98*, 1978.

(58) Sharp, K. A.; Honig, B. *Annu Rev Biophys Bio* **1990**, *19*, 301.

(59) Klamt, A. *COSMO-RS : from quantum chemistry to fluid phase thermodynamics and drug design*; Elsevier: Amsterdam ; Oxford, 2005.

(60) Fogolari, F.; Brigo, A.; Molinari, H. *J Mol Recognit* **2002**, *15*, 377.

(61) Bashford, D.; Case, D. A. *Annu Rev Phys Chem* **2000**, *51*, 129.

(62) Newton, I. *Filozofia* **2001**, *56*, 341.

(63) Verlet, L. *Phys Rev* **1967**, *159*, 98.

(64) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *Journal of Chemical Physics* **1982**, *76*, 637.

(65) Hockney, R. W. *Methods Comput. Phys.* **1970**, 136.

(66) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J Comput Phys* **1977**, *23*, 327.

(67) Hunenberger, P. *Adv Polym Sci* **2005**, *173*, 105.

(68) Wu, X. W.; Brooks, B. R. *Chemical physics letters* **2003**, *381*, 512.

(69) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *Journal of Chemical Physics* **1995**, *103*, 4613.

(70) Grossfield, A.; Zuckerman, D. M. *Ann Rep Comp Chem* **2009**, *5*, 23.

(71) Madsen, H. *Time series analysis*; Chapman & Hall/CRC: Boca Raton, Fla. ; London, 2008.

(72) Bishop, M.; Frinks, S. *Journal of Chemical Physics* **1987**, *87*, 3675.

(73) Yang, W.; Bitetti-Putzer, R.; Karplus, M. *Journal of Chemical Physics* **2004**, *120*, 2618.

(74) Kampf, K.; Klameth, F.; Vogel, M. *Journal of Chemical Physics* **2012**, *137*.

(75) Berkowitz, M.; Morgan, J. D.; Kouri, D. J.; Mccammon, J. A. *Journal of Chemical Physics* **1981**, *75*, 2462.

(76) Lawrenz, M.; Baron, R.; McCammon, J. A. *Journal of chemical theory and computation* **2009**, *5*, 1106.

(77) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. *Journal of Chemical Physics* **2005**, *123*.

(78) Sadiq, S. K.; Wright, D. W.; Kenway, O. A.; Coveney, P. V. *Journal of Chemical Information and Modeling* **2010**, *50*, 890.

(79) Chang, C. E.; Chen, W.; Gilson, M. K. *Proc Natl Acad Sci U S A* **2007**, *104*, 1534.

(80) Sippl, M. J.; Wiederstein, M. *Bioinformatics* **2008**, *24*, 426.

(81) Fiser, A.; Do, R. K. G.; Sali, A. *Protein Science* **2000**, *9*, 1753.

(82) Gao, M.; Schulten, K. *Biophys J* **2006**, *90*, 3267.

(83) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.

(84) Mikulskis, P.; Genheden, S.; Ryde, U. *Journal of Molecular Modeling* **2014**, *20*.

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

(86) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. *Journal of Chemical Physics* **2001**, *114*, 2090.

(87) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *Journal of Chemical Physics* **1994**, *101*, 4177.

(88) Harvey, M. J.; De Fabritiis, G. *Journal of chemical theory and computation* **2009**, *5*, 2371.

(89) Ruymgaart, A. P.; Elber, R. *Journal of chemical theory and computation* **2012**, *8*, 4624.

(90) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins-Structure Function and Genetics* **2002**, *46*, 24.

(91) Im, W.; Lee, M. S.; Brooks, C. L., 3rd *J Comput Chem* **2003**, *24*, 1691.

(92) Dominy, B. N.; Brooks, C. L. *Journal of Physical Chemistry B* **1999**, *103*, 3765.

(93) Chen, J.; Im, W.; Brooks, C. L., 3rd *J Am Chem Soc* **2006**, *128*, 3728.

(94) Im, W.; Feig, M.; Brooks, C. L. *Biophysical Journal* **2003**, *85*, 2900.

(95) Dahanayake, J. N.; Gautam, D. N.; Verma, R.; Mitchell-Koch, K. R. *Mol Simulat* **2016**, 1.

(96) Dyson, H. J.; Wright, P. E. *Chem Rev* **2004**, *104*, 3607.

(97) Xia, J. C.; Deng, N. J.; Levy, R. M. *Journal of Physical Chemistry B* **2013**, *117*, 6625.

(98) Boehr, D. D.; Dyson, H. J.; Wright, P. E. *Chem Rev* **2006**, *106*, 3055.

(99) Min, W.; Luo, G. B.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. *Physical Review Letters* **2005**, *94*.

(100) Xia, B.; Liu, Y. L.; Li, W.; Brice, A. R.; Dominy, B. N.; Cao, W. G. *Journal of Biological Chemistry* **2014**, *289*, 18413.

(101) Agrawal, V. *Clemson University PhD Dissertation* **2016**.

(102) Hayward, S.; Kitao, A.; Hirata, F.; Go, N. *Journal of Molecular Biology* **1993**, *234*, 1207.

(103) Nadler, W.; Brunger, A. T.; Schulten, K.; Karplus, M. *P Natl Acad Sci USA* **1987**, *84*, 7933.

(104) Cartling, B. *Journal of Chemical Physics* **1989**, *91*, 427.

(105) Godschalk, F.; Genheden, S.; Soderhjelm, P.; Ryde, U. *Physical Chemistry Chemical Physics* **2013**, *15*, 7731.

(106) Feig, M. *Journal of chemical theory and computation* **2007**, *3*, 1734.

(107) Genheden, S.; Ryde, U. *Journal of Computational Chemistry* **2010**, *31*, 837.

(108) Homeyer, N.; Gohlke, H. *Mol Inform* **2012**, *31*, 114.

(109) Kong, X. T.; Pan, P. C.; Li, D.; Tian, S.; Li, Y. Y.; Hou, T. J. *Physical Chemistry Chemical Physics* **2015**, *17*, 6098.

(110) Adasme-Carreno, F.; Munoz-Gutierrez, C.; Caballero, J.; Alzate-Morales, J. H. *Physical Chemistry Chemical Physics* **2014**, *16*, 14047.

(111)Sun, H. Y.; Li, Y. Y.; Shen, M. Y.; Tian, S.; Xu, L.; Pan, P. C.; Guan, Y.; Hou, T. J. *Physical Chemistry Chemical Physics* **2014**, *16*, 22035.

(112) Xie, A.; van der Meer, L.; Austin, R. H. *J Biol Phys* **2002**, *28*, 147.

(113) Xie, A. H.; van der Meer, A. F. G.; Austin, R. H. *Physical Review Letters* **2002**, *88*.

(114) Genheden, S.; Ryde, U. *Journal of Computational Chemistry* **2011**, *32*, 187.

(115) Inglesby, T. V. *Jama-J Am Med Assoc* **2000**, *283*, 1963.

(116) St Croix, B.; Rago, C.; Velculescu, V.; Traverso, G.; Romans, K. E.; Montgomery, E.; Lal, A.; Riggins, G. J.; Lengauer, C.; Vogelstein, B.; Kinzler, K. W. *Science* **2000**, *289*, 1197.

(117) Cryan, L. M.; Rogers, M. S. *Frontiers in bioscience* **2011**, *16*, 1574.

(118) Nanda, A.; St Croix, B. *Curr Opin Oncol* **2004**, *16*, 44.

(119) Raeisossadati, R.; Farshchian, M.; Ganji, A.; Tavassoli, A.; Velayati, A.; Dadkhah, E.; Chavoshi, S.; Mehrabi Bahar, M.; Memar, B.; Rajabi Mashhadi, M. T.; Naseh, H.; Forghanifard, M. M.; Moghbeli, M.; Moaven, O.; Abbaszadegan, M. R. *International journal of colorectal disease* **2011**, *26*, 1265.

(120) Frankel, A. E.; Carter, C.; Kuo, S. R.; Woo, J. H.; Mauldin, J.; Liu, J. S. *Anticancer Agents Med Chem* **2011**, *11*, 983.

(121) Chaudhary, A.; St Croix, B. *Cell Cycle* **2012**, *11*, 2253.

(122) Chaudhary, A.; Hilton, M. B.; Seaman, S.; Haines, D. C.; Stevenson, S.; Lemotte, P. K.; Tschantz, W. R.; Zhang, X. M.; Saha, S.; Fleming, T.; St Croix, B. *Cancer Cell* **2012**, *21*, 212.

(123) Cryan, L. M.; Habeshian, K. A.; Caldwell, T. P.; Morris, M. T.; Ackroyd, P. C.; Christensen, K. A.; Rogers, M. S. *Journal of biomolecular screening* **2013**, *18*, 714.

(124) Scobie, H. M.; Young, J. A. *Curr Opin Microbiol* **2005**, *8*, 106.

(125) Go, M. Y.; Chow, E. M.; Mogridge, J. *Infect Immun* **2009**, *77*, 52.

(126) Yang, M. Y.; Chaudhary, A.; Seaman, S.; Dunty, J.; Stevens, J.; Elzarrad, M. K.; Frankel, A. E.; St Croix, B. *Biochim Biophys Acta* **2011**, *1813*, 39.

(127) Jin, M.; Andricioaei, L.; Springer, T. A. *Structure* **2004**, *12*, 2137.

(128) Ramey, J. D.; Villareal, V. A.; Ng, C.; Ward, S. C.; Xiong, J. P.; Clubb, R. T.; Bradley, K. A. *Biochemistry* **2010**, *49*, 7403.

(129) Singh, M. K.; Dominy, B. N. *Proteins-Structure Function and Bioinformatics* **2010**, *78*, 1613.

(130) Brice, A. R.; Dominy, B. N. *J Comput Chem* **2011**, *32*, 1431.

(131) Swanson, J. M.; Henchman, R. H.; McCammon, J. A. *Biophys J* **2004**, *86*, 67.

(132) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics & Modelling* **1996**, *14*, 33.

(133) Qu, A. D.; Leahy, D. J. *Structure* **1996**, *4*, 931.

(134) Zoete, V.; Meuwly, M.; Karplus, M. *Proteins-Structure Function and Bioinformatics* **2005**, *61*, 79.

(135) San Sebastian, E.; Mercero, J. M.; Stote, R. H.; Dejaegere, A.; Cossio, F. P.; Lopez, X. *J Am Chem Soc* **2006**, *128*, 3554.

(136) Plow, E. F.; Haas, T. A.; Zhang, L.; Loftus, J.; Smith, J. W. *J Biol Chem* **2000**, *275*, 21785.

(137) Shimaoka, M.; Xiao, T.; Liu, J. H.; Yang, Y. T.; Dong, Y. C.; Jun, C. D.; McCormack, A.; Zhang, R. G.; Joachimiak, A.; Takagi, J.; Wang, J. H.; Springer, T. A. *Cell* **2003**, *112*, 99.

(138) Lee, J. Y.; Tsai, Y. M.; Chao, S. C.; Tu, Y. F. *Clin Exp Dermatol* **2005**, *30*, 176.

(139) Lee, J. O.; Rieu, P.; Arnaout, M. A.; Liddington, R. *Cell* **1995**, *80*, 631.

(140) Vorup-Jensen, T.; Chi, L.; Gjelstrup, L. C.; Jensen, U. B.; Jewett, C. A.; Xie, C.; Shimaoka, M.; Linhardt, R. J.; Springer, T. A. *J Biol Chem* **2007**, *282*, 30869.

(141) Nymalm, Y.; Puranen, J. S.; Nyholm, T. K. M.; Kapyla, J.; Kidron, H.; Pentikainen, O. T.; Airenne, T. T.; Heino, J.; Slotte, J. P.; Johnson, M. S.; Salminen, T. A. *Journal of Biological Chemistry* **2004**, *279*, 7962.

(142) Emsley, J.; Knight, C. G.; Farndale, R. W.; Barnes, M. J.; Liddington, R. C. *Cell* **2000**, *101*, 47.

(143) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *Journal of Molecular Biology* **1990**, *215*, 403.

(144) Reeves, C. V.; Dufraine, J.; Young, J. A. T.; Kitajewski, J. *Oncogene* **2010**, *29*, 789.

(145) Jinnin, M.; Medici, D.; Park, L.; Limaye, N.; Liu, Y.; Boscolo, E.; Bischoff, J.; Vikkula, M.; Boye, E.; Olsen, B. R. *Nat Med* **2008**, *14*, 1236.

(146) Venanzi, F. M.; Petrini, M.; Fiammenghi, L.; Bolli, E.; Granato, A. M.; Ridolfi, L.; Gabrielli, F.; Barucca, A.; Concetti, A.; Ridolfi, R.; Riccobon, A. *Cancer Immunol Immun* **2010**, *59*, 27.

(147) Rouleau, C.; Menon, K.; Boutin, P.; Guyre, C.; Yoshida, H.; Kataoka, S.; Perricone, M.; Shankara, S.; Frankel, A. E.; Duesbery, N. S.; Woude, G. V.; Biemann, H. P.; Teicher, B. A. *International Journal of Oncology* **2008**, *32*, 739.

(148) Felicetti, P.; Mennecozzi, M.; Barucca, A.; Montgomery, S.; Orlandi, F.; Manova, K.; Houghton, A. N.; Gregor, P. D.; Concetti, A.; Venanzi, F. M. *Cytotherapy* **2007**, *9*, 23.

(149) Bonuccelli, G.; Sotgia, F.; Frank, P. G.; Williams, T. M.; de Almeida, C. J.; Tanowitz, H. B.; Scherer, P. E.; Hotchkiss, K. A.; Terman, B. I.; Rollman, B.; Alileche, A.; Brojatsch, J.; Lisanti, M. P. *Am J Physiol-Cell Ph* **2005**, *288*, C1402.

(150) Fernando, S.; Fletcher, B. S. *Cancer Research* **2009**, *69*, 5126.

(151) Ye, L.; Sun, P. H.; Malik, M. F. A.; Mason, M. D.; Jiang, W. G. *Journal of cancer research and clinical oncology* **2014**, *140*, 957.

(152) Chen, K. H.; Liu, S.; Bankston, L. A.; Liddington, R. C.; Leppla, S. H. *Journal of Biological Chemistry* **2007**, *282*, 9834.

(153) Liu, S.; Schubert, R. L.; Bugge, T. H.; Leppla, S. H. *Expert Opin Biol Ther* **2003**, *3*, 843.

(154) Rosovitz, M. J.; Schuck, P.; Varughese, M.; Chopra, A. P.; Mehra, V.; Singh, Y.; McGinnis, L. M.; Leppla, S. H. *J Biol Chem* **2003**, *278*, 30936.

(155) Mechaly, A.; McCluskey, A. J.; Collier, R. J. *Mbio* **2012**, *3*.

(156) Pribila, J. T.; Quale, A. C.; Mueller, K. L.; Shimizu, Y. *Annual review of immunology* **2004**, *22*, 157.

(157) Kinashi, T. *Nat Rev Immunol* **2005**, *5*, 546.

(158) Lee, J. O.; Rieu, P.; Arnaout, M. A.; Liddington, R. *Cell* **1995**, *80*, 631.

(159) Lee, J. O.; Bankston, L. A.; Arnaout, M. A.; Liddington, R. C. *Structure* **1995**, *3*, 1333.

(160) Emsley, J.; Knight, C. G.; Farndale, R. W.; Barnes, M. J.; Liddington, R. C. *Cell* **2000**, *101*, 47.

(161) Luo, B. H.; Carman, C. V.; Springer, T. A. *Annual review of immunology* **2007**, *25*, 619.

(162) Shimaoka, M.; Lu, C. F.; Palframan, R. T.; von Andrian, U. H.; McCormack, A.; Takagi, J.; Springer, T. A. *P Natl Acad Sci USA* **2001**, *98*, 6009.