UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Molecular Recognition Modeling: Free Energy, Protein Dynamics and Unbinding Pathways

Permalink

https://escholarship.org/uc/item/82z6s4c9

Author

You, Wanli

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Molecular Recognition Modeling: Free Energy, Protein Dynamics and Unbinding
Pathways

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Wanli You

September 2017

Dissertation Committee:
      Dr. Chia-en A. Chang, Chairperson
      Dr. Gregory Beran
      Dr. Ryan Julian

The Dissertation of Wanli You is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

The Text of this thesis, in part or in full, is a reprint of the material as it appears in the following publications:

Chapter 2: You W, Huang YMM, Kizhake S, Natarajan A, Chang, CEA (2016) Characterization of Promiscuous Binding of Phosphor Ligands to Breast-Cancer-Gene 1 (BRCA1) C-Terminal (BRCT): Molecular Dynamics, Free Energy, Entropy and Inhibitor Design. PLOS Computational Biology 12(8): e1005057.

Chapter 4: Huang YMM., You W, Caulkins BG, Dunn MF, Mueller L, Chang CEA (2016) Protonation states and catalysis: Molecular dynamics studies of intermediates in tryptophan synthase. Protein Sciences 25 (1), 166-183.

Among them Chapter 2 includes the entire article without being rewritten. Chapter 4 only includes rewritten article about the portion of the work that I contributed to.

ABSTRACT OF THE DISSERTATION

Molecular Recognition Modeling: Free Energy, Protein Dynamics and Unbinding
Pathways

by

Wanli You

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, September 2017
Dr. Chia-en A. Chang, Chairperson

Molecular recognition is fundamentally important in biological chemistry. Nowadays, with the rapid development of computational technology and algorithm, molecular modeling has become a powerful tool in studying molecular recognition, such as exploring molecular interactions and understanding biological dynamic processes, making significant contributions to modern biology and drug discovery. The state-of-the-art techniques of computational chemistry and molecular modeling can be applied to study a wide range of chemical and biological systems of interest. This enables us to study structural details at the atomic level and obtain chemical/biological information which is not available by experimental measurements. This dissertation project focused on modeling the recognition mechanisms of biomolecules and their conjugated ligands. Multiple computational techniques, such as molecular dynamics simulation, enhanced sampling methods and free energy calculation were applied. The model systems included

signaling domains (BRCT domain), kinase (p38 kinase), enzyme system (TRPS) and small biomolecular system (cyclodextrin). The details of protein-ligand interactions, including both enthalpic and entropic contribution within protein domain-phosphopeptide systems were investigated, based on which new inhibitors were proposed. Several enhanced sampling methods like accelerated molecular dynamics simulation, pathway search guided by internal motions (PSIM) and umbrella sampling, were applied to explore the dissociation pathway of kinase-ligand systems and the motions of kinase during dissociation process were studied both thermodynamically and kinetically, protein conformational rearrangement was found to differentiate slow and fast unbinding inhibitors, casting light on high efficacy inhibitor design. Furthermore, using full structural molecular modeling, we explored how the position of a single proton can change the overall protein dynamics and further activate or inactivate enzyme catalysis, elucidating the catalytic mechanism of TRPS. In addition, we performed systematically evaluation to the performance of umbrella sampling, investigated the influence of subtle changes in the dissociation pathways and conformational sampling methods that provide the initial conformations, paving way for future improvement of umbrella sampling. This project studies the details of receptor-ligand interaction and provides a more complete picture of molecular recognition.

Table of Contents

List of Tables

List of Figures

xix

Chapter 1 Introduction

1.1 Overview

Molecular recognition, including both static molecular interaction and dynamic behaviors, is of profound importance in biology and therapeutics. Structure-based ligand design in medicinal chemistry relies on the identification and quantification of noncovalent interactions. Study of dynamic behaviors in ligand-receptor systems reveal key factors behind drug efficacy such as the change of association or dissociation rates. This project studies the molecular recognition in biomolecular systems, such as protein-ligand and small molecular host-guest systems. To approach this goal, multiple tools of molecular modeling were applied. Originated from the early models of ball and stick model and evolved into modern technique with the appearance and development of computer science, molecular modeling has become an important component in fundamental studies in chemistry and biology, playing both explanatory and predictive roles. Subsequent sections in this chapter describe the theories and techniques of molecular modeling, such as molecular dynamics simulation and free energy calculation. The chapters that follow present several studies about molecular interactions and unbinding process of bimolecular systems, as well as an evaluation of umbrella sampling method in building free energy profile along ligand dissociation pathway.

Targeting protein–protein interactions (PPIs) has emerged as a viable approach in modern drug discovery. The use of phosphopeptides and phosphor compounds as inhibitors to protein–protein interactions have attracted increasing interest for years. Chapter 2 focuses

on study of interactions between phosphor ligands and a protein domain. Free energy calculation were performed to provide insights into the mechanism of binding, and were evaluated from enthalpical and entropical aspects, based on which new inhibitors were proposed to improve binding affinity.

A sole strong binding affinity doesn't always guarantee good drug efficacy, another crucial factor is the mean residence time (RT) in the binding pocket, which is usually quantified by dissociation rate constant through the equation $RT = 1/k_{off}$. A thorough understanding of ligand dissociation process is important in drug development. In Chapter 3, we applied multiple enhanced sampling methods to investigate the dissociation pathway of inhibitors of a kinase system p38α. Umbrella sampling method was applied to build free energy profile of ligand dissociation process.

The mechanism of catalysis in many enzymes depends on simple acid-base reactions involving a series of proton transfers. The protonation state of a specific atom is crucial to initiate chemical reactions in enzyme active sites. Chapter 4 focuses on one detail of ligand-enzyme interaction. Using full structural models of the protein, we explored how the position of a single proton can change the overall protein dynamics and further activate or inactivate enzyme catalysis of TRPS.

Umbrella sampling is a widely used method to build free energy profile, but how accurate and sensitive it is on different scales of molecular systems has never been explored. In Chapter 5 we applied umbrella sampling to study the dissociation process of both a small host-guest system and a large protein-ligand system. The influence of conformation of

receptor and selection of dissociation path on free energy profile were carefully assessed in this chapter.

1.2 Molecular dynamics

Molecular dynamics (MD) simulation, which is first developed over 30 years ago [1], has advanced from a simple method to simulate movements of several hundreds of atoms to a widely used way to model large biomolecular systems, including peptides, proteins, nucleotides, lipids and viruses [2-6]. Simulation of systems with ~50,000–100,000 atoms are now routine. Nowadays, with the rapid development of computational technology and algorithms, MD simulation has become a powerful tool to study the structure and dynamics of macromolecules.

Molecular dynamics is a method to model the motion of some group of particles (e.g., atoms) governed by Newton's law. Forces acting on each atom are obtained by deriving equations of potential energy. Then forces acting on individual atoms are used for the calculation of accelerations and velocities with classical Newton's law of motion. Therefore atom positions are updated after each time step. The result of MD is a trajectory of the system over a certain period of time, usually in nanosecond (ns) scale. Various structural and dynamic properties of the system then are analyzed based on the trajectories, and some of them are used to directly compare with experimental data.

1.2.1 Force fields

The potential energy function of MD simulations can be described as equations with parameters such as equilibrium bond length or angle, partial atomic charge, and van der Waals atomic radii (called "force-field" [7-9]).). The commonly used force fields include Amber [10, 11], Charmm [12], GROMOS [13], and OPLS [14]. The potential energy U can be divided into two main terms, bonded and non-bonded terms. Chemical bonds, atomic angles and improper angles are modeled by harmonic motions, and dihedral angles are modeled by using a sinusoidal function that approximates the energy differences between eclipsed and staggered conformations. Non-bonded terms include van der Waals interactions, modeled by using the Lennard-Jones potential, and electrostatic interactions, modeled by using Coulomb's law. The equations can be expressed as

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{improper} + U_{vdw} + U_{elec} \qquad \text{eq. 1-1}$$

Each term of this equation can be further represented as in Figure 1.1.

$$U(R) = \sum_{bonds} k_r \left(r - r_{eq}\right)^2 \qquad \qquad \text{\textit{bond}}$$

$$+ \sum_{angles} k_\theta \left(\theta - \theta_{eq}\right)^2 \qquad \qquad \text{\textit{angle}}$$

$$+ \sum_{dihedrals} k_\phi \left(1 + \cos[n\phi - \gamma]\right) \qquad \text{\textit{dihedral}}$$

$$+ \sum_{impropers} k_\omega \left(\omega - \omega_{eq}\right)^2 \qquad \qquad \text{\textit{improper}}$$

$$+ \sum_{i<j}^{atoms} \varepsilon_{ij} \left[\left(\frac{r_m}{r_{ij}}\right)^{12} - 2 \left(\frac{r_m}{r_{ij}}\right)^6 \right] \qquad \text{\textit{van der Waals}}$$

$$+ \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \qquad \text{\textit{electrostatic}}$$
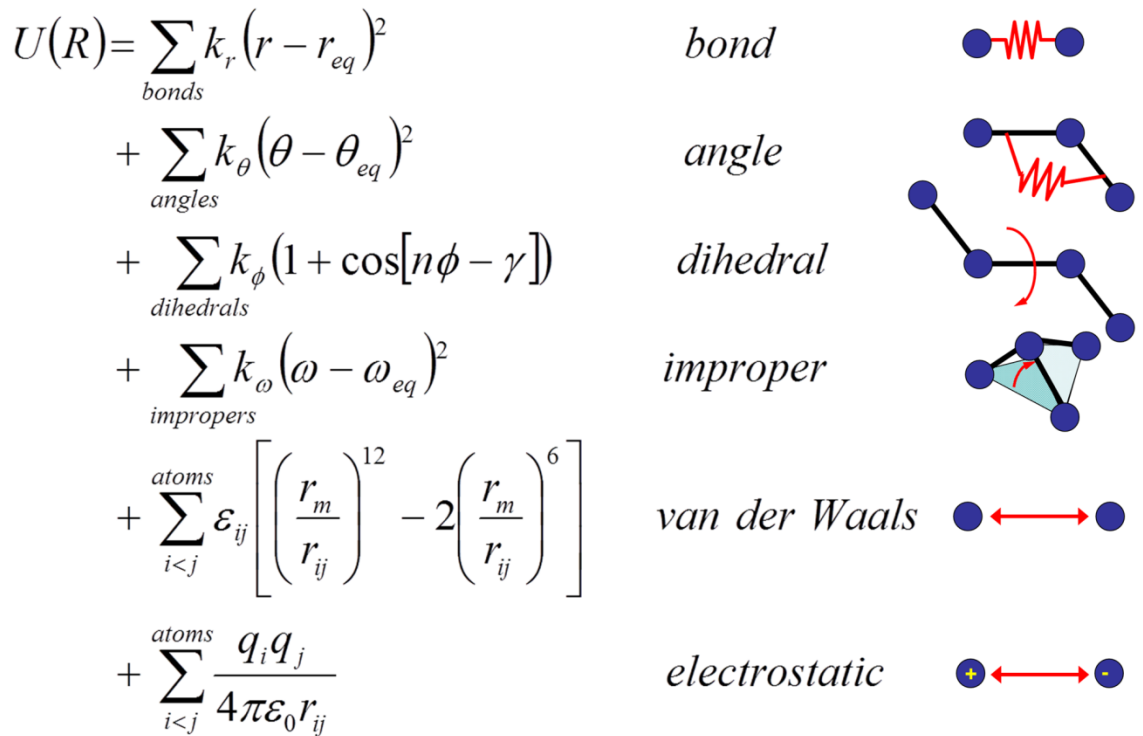
Figure 1.1. Classical force fields used for MD simulations. Right: potential energy terms in a force field. Left: energy function used to derive atomic forces for molecular movement.

The chemical bonds are modeled as harmonic springs with a reference value equal to the equilibrium length of the bond, and a force constant describing the strength of the bond, In the bond term of the functional form, k_r is the force constant of chemical bonds, r is the bond distance and r_{eq} is the equilibrium distance.

The angle term is similar to the bond term, a harmonic potential is used to describe the angle between the two bonds sequentially connecting the three atoms. In the angle term of the functional form, k_θ is the force constant of the angle form by three atoms, θ is the angle in radians and θ_{eq} is the equilibrium angle.

The dihedral term is the pivotal term which further defines the structure and differentiates the conformations of the molecule. In the dihedral term of the functional form, k_φ is the force constant of the dihedral angle form by four atoms, φ is the dihedral in radians determining the minima and maxima of the potential, and multiplicity n is the nonnegative integer that indicates the periodicity. For $n > 0$, γ is the phase shift angle, while for $n = 0$, γ acts as an equilibrium dihedral.

The improper term defines the angle at which the fourth atom deviates from the plane defined by the first three atoms. The improper torsion is the least frequent term in the potential energy function and is only used to enforce the planar geometry, for example, an $sp2$ hybridized atom. In the improper term of the functional form, K_φ is the force constant of the improper angle, φ is the improper angle in radians and φ_0 is the equilibrium improper angle.

The non-bonded term are to model van der Waals and electrostatic interactions. In molecular dynamics, the electrostatic interaction is modeled in a partial charge based monopole approach. The partial charge of the atom is usually assigned by empirical, semi-empirical, or quantum calculations. Both electrostatic and van der Waals interaction are truncated at two body interactions and higher order of interactions are captured by empirically tuning the two body interaction parameters. The behavior of any pair of atoms in a system is under the government of non-bonded forces unless the two atoms are involved in a bond or angle term.

In the van der Waals interaction term of the functional form, r_{ij} gives the distance between the pair of atoms, ε_{ij} is related to the Lennard-Jones well depth, and r_m is the distance at which the potential reaches its minimum.

In the electrostatic interaction term of the functional form, r_{ij} is the distance between the pair of atoms, q_i and q_j are the charges on the respective atoms and the dielectric constant of free space ε_0 is fixed for all electrostatic interactions.

1.2.2 Molecular dynamics algorithm

Once forces acting on atoms obtained from deriving equations of force filed, we can determine the acceleration of each atom in the system using Newton's second law, $F = ma$, where F is the force exerted on the particle, m is its mass and a is its acceleration. Next, integration of the equations of motion will yield a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. Given the inital positions and velocities of each atom, the state of the system can be predicted at any time in the future. A simplicity of the basic MD algorithm is shown in Figure 1.2.
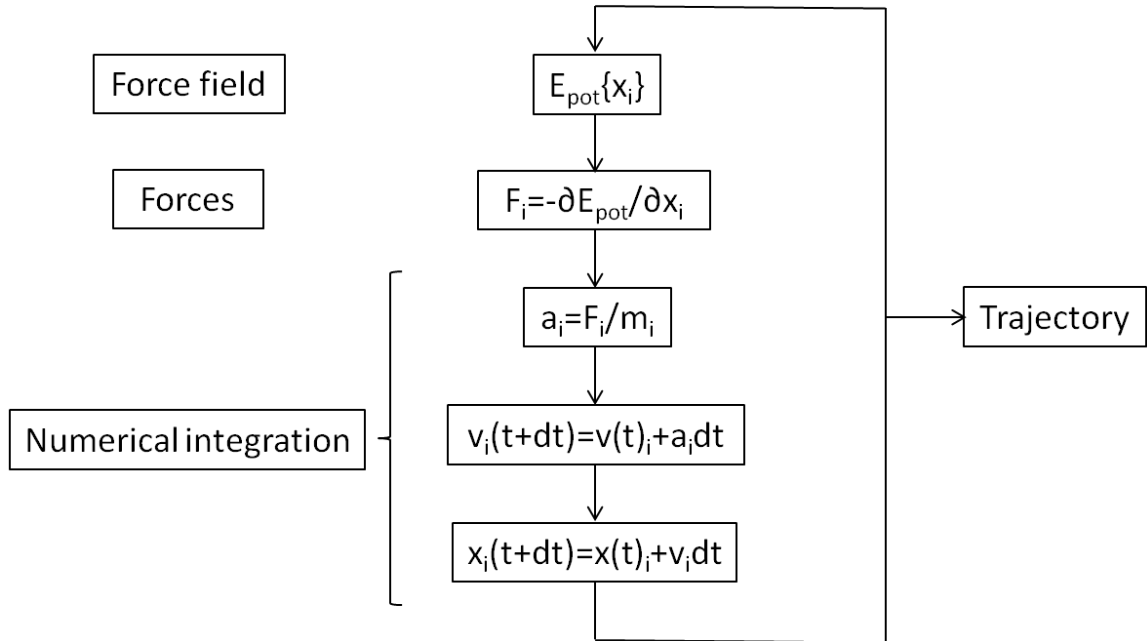
Figure 1.2. Molecular dynamics basic algorithm. Epot, potential energy; t, simulation time; dt, iteration time; For each spatial coordinate of the N simulated atoms (i): x, atom coordinate; F, forces component; a, acceleration; m, atom mass; v, velocity.

Numerical integration is applied to update position of atoms because there is no analytical solution to the equation with large degrees of freedom. There are many numerical algorithms developed for integrating the equations. All the integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion:

$$r(t + dt) = r(t) + v(t)dt + \frac{1}{2}a(t)dt^2 + \cdots \qquad \text{eq. 1-2}$$

$$v(t + dt) = v(t) + a(t)dt + \frac{1}{2}b(t)dt^2 + \cdots \qquad \text{eq. 1-3}$$

$$a(t + dt) = a(t) + b(t)dt + \frac{1}{2}c(t)dt^2 + \cdots, \qquad \text{eq. 1-4}$$

where r is the position, v is the velocity (the first derivative of position with respect to time), a is the acceleration (the second derivative of position with respect to time), etc.

Here, we introduce the two most widely used algorithms.

(1) Verlet algorithm

To derive the Verlet algorithm, we can first write

$$r(t + dt) = r(t) + v(t)dt + \frac{1}{2}a(t)dt^2 \qquad \text{eq. 1-5}$$

$$r(t - dt) = r(t) - v(t)dt + \frac{1}{2}a(t)dt^2. \qquad \text{eq. 1-6}$$

Then we will add up these two equations, we can obtain

$$r(t + dt) = 2r(t) - r(t - dt) + a(t)dt^2. \qquad \text{eq. 1-7}$$

The new positions at time ($t+dt$) can be calculated by the positions from time ($t-dt$) and the positions and accelerations at time t.

(2) Leap-frog algorithm

In this algorithm, the velocities are the first calculated at time ($t+1/2dt$). These are used to calculate the positions, r, at time ($t+dt$).

$$v\left(t + \frac{1}{2}dt\right) = v\left(t - \frac{1}{2}dt\right) + a(t)dt \qquad \text{eq. 1-8}$$

$$r(t + dt) = r(t) + v\left(t + \frac{1}{2}dt\right)dt \qquad \text{eq. 1-9}$$

9

In this way, the velocities leap over the positions, and then the positions leap over the velocities. The advantage of this algorithm is that the velocities are explicitly calculated, however, the disadvantage is that they are not calculated at the same time as the positions. The velocities at time t can be approximated by the relationship:

$$v(t) = \frac{1}{2}\left[v\left(t - \frac{1}{2}dt\right) + v\left(t + \frac{1}{2}dt\right)\right] \qquad \text{eq. 1-10}$$

1.2.3 Setting up a simulation

An all atom molecular dynamics simulation requires careful preparation of initial coordinates, protonation state, solvent model, proper minimization and equilibrium because the potential energy function is sensitive to the details of the model. The common protocol for performing MD simulations consists of a number of steps and is shown in figure 1.3.
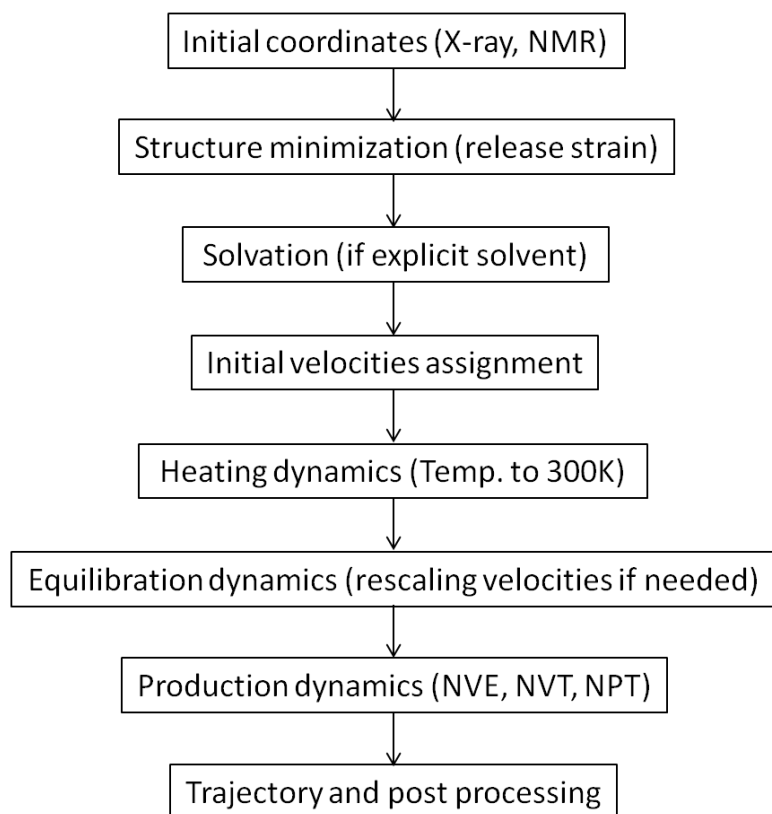
Figure 1.3. General steps used in molecular dynamics simulations.

Details for each step of MD simulations are summarized as following:

1.2.3.1 Preparing initial structures

In MD simulations of biomolecules, an X-ray crystalstructure or an nuclear magnetic resonance (NMR) structure obtained from the online databank is commonly used as the initial structure. For apo protein and protein-ligand crystal structures, the Protein Data Bank is a publicly available database which stores the experimentally determined crystal structures of various protein systems. For generic molecules, the Cambridge Crystallographic Data Centre is commonly used database. It is also possible to use a

theoretical structure developed by homology modeling (also known as comparative modeling) [15] if experimental data is not available. Homology modeling is a tool which builds the undetermined protein structure based on its sequence and a chosen template that is structurally similar to the target protein. In the case of small molecules, there are various visualization and editing programs that can build the structure manually, such as Avogadro and VEGA ZZ. If the structures of protein and ligand are obtained by either means while the bound state structure is not been determined, docking can be a good option to find the best reasonable bound conformations.

Before moving on to structure minimization, we still need to determine the protonation state of each titratable residue. We apply pKa value to determine the protonation state of a residue. Because the pKa value of each residue is sensitive to hydrogen bonding, desolvation effect and electrostatic interactions in its local environment, it is different from the pKa of standard amino acid. To calulate the true pKa values of residues in different surroundings, multiples programs are available, such as MCCE, PROPKA, DEPTH, H++ and UHBD [16-18].

1.2.3.2 Minimization

Because a protein has to be crystallised to apply X-ray crystallography, or due to small deviations in NMR, the position of its constituent atoms may be distorted from their natural positions. Consequently, bond lengths and bond angles may be distorted and steric clashes in between atoms may occur. Therefore, it is recommended to minimize the potential energy of the structure to create a more realistic structure. The restraints on a

part of protein atoms are designed to gradually bring the system to the target condition. The hydrogen atoms of the system are usually minimized first by restraining the all atoms except hydrogen. The next is to minimize the sidechain and whole biomolecule. The entire system including protein and water molecules is minimized finally. Several minimum search algorithms are commonly used to bring the energy to local energy minima, such as steepest descent, conjugate gradient, Newton-Raphson method and quasi-Newton method [19-21]. Among these methods, steepest descent and conjugate gradient are first order methods which only utilize the gradient of the potential energy function, Newton-Raphson method and quasi-Newton method are second order methods which calculate both gradient and hessian matrix for minimization. The first order minimization methods are not as efficient as the second order methods, but the advantage is that the starting point does not have to be near the minimum. By combining the first order and second order methods, we can minimize structures with high efficiency. First we can pick a random or coarse initial conformation of the molecular system, and use the first order methods to bring the conformation closer to the minimum, then use the second order methods to fully minimize the system.

1.2.3.3 Solvation

Biomolecular systems do not exist in vacumm. There are water molecules and ions surrounding them. To obtain a realistic simulation, the structure of interest must be solvated. The solvation can be either implicit or explicit. In the case of an implicit solvation, the water molecules are replaced by a potential, which describe their average

13

action while, in the case of an explicit solvation, the molecule is surrounded by a solvation box constituted of water molecules. Various explicit solvent models have been used, such as TIPS, TIP3P, TIP4P and TIP5P [22-24]. which are differentiated by the number of points used, for example, whether the structure is rigid or flexible, or whether the model considers polarization effects or not. These models include a molecular mechanical description of the water molecule structure and intermolecular force field parameters that can accurately describe water both structurally and energetically. The explicit water models represent the most accurate molecular properties of water. However, the large number of water molecules used to solvate a molecular system can significantly slows the calculation of the potential energy function. In some cases it may make a certain particular model impractical for a desired application, then we can choose implicit water model instead.

Implicit solvent model, also known as a continuum model, is a method to represent solvent as a continuous medium instead of individual water molecules. Many different methods can be applied to approximate the averaged behavior of water molecules. While two methods, Poisson-Boltzmann surface area (PBSA) and Generalized Born surface area (GBSA), are the most commonly used implicit models in MD simulations. The Poisson-Boltzmann (PB) equation is an accurate function that models the electrostatic properties of a charged solute in an ionic solution [25, 26]. The equation can be written as

$$\vec{\nabla} \cdot \left[\epsilon(\vec{r}) \vec{\nabla} \Psi(\vec{r}) \right] = -4\pi \rho^f(\vec{r}) - 4\pi \sum_i c_i^\infty z_i q \lambda(\vec{r}) \exp \left[-\frac{z_i q \Psi(\vec{r})}{k_b T} \right], \qquad \text{eq. 1-11}$$

14

where $\epsilon(\vec{r})$ is the position-dependent dielectric, $\Psi(\vec{r})$ is the electrostatic potential, $\rho^f(\vec{r})$ is the charge density of the solute, c_i^∞ is the concentration of the ion i at a distance of infinity from the solute, z_i is the valence of the ion, q is the charge of a proton, k_b is the Boltzmann constant, T is the temperature, and $\lambda(\vec{r})$ is a factor for the position-dependent accessibility of position r to the ions in solution. Because of the computational expense of solving the PB equation, it is usually used to investigate energies of static structures of a molecular model rather than a dynamics simulation. The relatively efficient generalized Born (GB) equation is then used to approximate the solvation free energy.

The GB model is an approximation of the linearized PB equation. It is based on modeling the solute using a set of spheres whose internal dielectric constant differs from the external solvent. The model is written as:

$$G = \frac{1}{8\pi}(\frac{1}{\epsilon_0} - \frac{1}{\epsilon}) \sum_{i,j}^{N} \frac{q_i q_j}{f_{GB}}, \qquad \text{eq. 1-12}$$

where $f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}}$, $D = (\frac{r_{ij}}{2a_{ij}})^2$ and $a_{ij} = \sqrt{a_i a_j}$,

where ϵ_0 is the permittivity of free space, ϵ is the dielectric constant of the solvent, q_i is the electrostatic charge on particle i, r_{ij} is the distance between particles i and j, and a_{ij} is the effective Born radius. Accurate estimation of the effective Born radius is important for the GB model [27, 28].

Besides water, the buffer solution contains different types of ions in most biological systems. We can add ions, such as Na^+ and Cl^-, in a system to provide a similar condition

as experiments. The added ions should neutralize the net charge of the biomolecule, so the total charge of the simulation system is zero.

If explicit water molecules are used to solvate the system, because the orientation is fixed for all added water molecules, minimization is again necessary to move and rotate water molecules to create a realistic structure before moving on to equilibration step. First, the protein position is fixed to allow minimization of water molecules. Once that is done, the constraints on the protein can be removed and the whole system will be minimized together.

1.2.3.4 Equilibrium

After minimizations, the system should be equalized by gradually increasing temperature. At the beginning of simulations velocities are assigned at low temperature and simulation proceeds via heating phase. Periodically, new velocities are assigned at a slightly higher temperature and the simulation is allowed to continue. This is repeated until the desired temperature is reached. The temperature is usually simulated from 50 K, 100 K, 150 K, 200 K, 250 K to 300 K. The Andersén thermostat, Langevin thermostat and Nosé-Hoover thermostat can help to control the temperature [29]. The ensembles, e.g., NPT, NVT or mVT, used in an MD simulation should be chosen according to the nature of the biomolecular system. The purpose of equilibration dynamics is to run the simulation until these properties become stable with respect to time.

1.2.3.5 Production run

The final step of the simulation is to run the simulation in "production" phase for the desired time length. The MD trajectories from the production dynamics can be analyzed and post-processed to obtain conformational dynamics, thermodynamic parameters and free energy values for the system of interest. Multiple packages can help to perform MD simulations, such as Amber, NAMD, Charmm and GROMOS [11, 13, 30, 31]. To speed up the calculations, MD simulations can use hundreds of CPUs at one time, which is called the Message Passing Interface (MPI). By using multiple processors for one calculation task simultaneously, we can largely reduce computation time. In recent years, the use of graphical processing unit (GPU) cards to accelerate calculations has been a major breakthrough in computational simulation field. With the use of GPU cards, which include many arithmetic units working in parallel, MD simulations can be accelerated by tens of times, so a single PC with such a card has the power similar to that of a cluster of workstations with multiple processors. Many major MD codes have already been rewritten to incorporate GPUs like ACEMD and OpenMM.

1.2.4 Post-MD analysis

MD simulations generate a wealth of data. Deducing meaningful conclusions from simulations requires analysis of MD trajectories in terms of the individual positions of all atoms or a selected subset of atoms for each time frame of a trajectory. Several common quantitative properties that can be analyzed from MD simulations are listed below.

1.2.4.1 Root-mean-square deviation

Root-mean-square deviation (RMSD) is a measure of the deviation of the atom positions at specific time in respect to refenrence positions. RMSD is often used on superimposed structures, either the whole structure or part of structure, such as a certain loop or only backbone atoms, for structure check, for example, to check if system reaches equlibrium, to find timepoints when conformation changes, and to define folding procedures. The definition of RMSD can be shown as the following:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_0)^2},$$
eq. 1-13

where, x is the coordinate or measured distance, 0 denotes the reference and N is the number of atoms.

1.2.4.2 Root-mean-square fluctuation

Root-mean-square fluctuation (RMSF) is a measure of the average distance between the position of atom i and one reference position over the time zone T. RMSF is often used to find high fluctuating areas in the system, and compare the results with experiments.

$$RMSF = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T}(x_i(t_j) - \bar{x_i})^2},$$
eq. 1-14

where x_i is the coordinate or measured distance at time t_j and $\bar{x_i}$ is the reference position of atom i.

The difference between RMSD and RMSF is, RMSD is the average taken over the particles given the specific time, while RMSF is the average taken over the time given a value of each atom i.

1.2.4.3 Configurational entropy calculation

The binding affinity of a ligand for its protein partner depends on the balance between intermolecular interactions and configurational entropy changes. The restriction of motion of a small ligand upon binding to a protein causes a loss of configurational entropy, and thus a penalty in binding affinity. Configurational entropy includes both conformational and vibrational terms, which are reflected by the number of energy wells and the average width of the occupied wells, respectively [32]. Using bond-angle-torsion (BAT) coordinate instead of cartesian coordinates, we can decompose the total entropy into individual terms based on the torsion coordinates, including backbone dihedrals (phi, psi and omega) and sidechain dihedrals [33]. The Gibbs entropy formula is used to calculate torsional entropy:

$$S_{conf(i)} = -R \sum P_i \ln (P_i),$$ eq. 1-15

where P_i is the probability distribution of each torsional angle i and R is the gas constant.

1.2.4.4 Principal Component Analysis

Principal component analysis (PCA) is a commonly used tool to extract useful information from the random motions of a long MD trajectory of our molecular systems.

Similar to quasi-Harmonic approximation, PCA calculates the eigenvectors of the covariance matrix for intrinsic motion creation. First, covariance matrix is calculated from the atom coordinates in the trajectory aligned to the reference structure, then we diagonalize it for eigenvectors, and apply the eigenvectors on the average atom positions to create the motion of each principal component mode. Due to the minimal data required for convergence, we usually only use the α-carbon atoms in each residue of the protein to construct the covariance matrix. Therefore we refer this commonly used PCA as α-carbon PCA. The principal component of each conformation along the PC mode can be calculated by

$$q_i = R^T(x_i - x_a),$$ eq. 1-16

where the x is the conformation along the trajectory, i is the conformation number, a is the average conformation of the trajectory, q is the principal component on one PCA mode, and RT is the transpose of the eigenvector matrix of the covariance matrix.

By ordering the eigenvalues of the transformation decreasingly, it has been shown that a large part of the system's fluctuations can be described by the first few principal components. The resulting low-dimensional representation of the dynamics can then be used to construct the free energy landscape by projecting the trajectory onto the space of the first few principal component modes and calculating the free energy from population on the these dimension space, which reveals the metastable conformational states and the transition states of the system.

1.2.4.5 MM-PBSA and MM-GBSA energy calculations

MMPBSA (Molecular Mechanics Poisson Boltzmann/Surface Area) and MMGBSA (Molecular Mechanics Generalized Born/Surface Area) are methods for assessing specific intermolecular interactions and estimating the binding energy of a ligand receptor system from the trajectory of a dynamics simulation [34-36]. The trajectory of a simulation is stripped of explicit solvent if it is present, then these methods compute binding energies using enthalpy/entropy decomposition approach with implicit solvation model, PB or GB. The binding interaction energy ΔE_{bind}, associated with the binding of a protein and ligand to form protein-ligand complex, can be calculated as the following:

$$\Delta E_{bind} = <E_{complex}> - <E_{protein}> - <E_{ligand}>, \qquad \text{eq. 1-17}$$

where the bracket $<E>$ denotes the average energy computed from a given MD trajectory. The change of interaction energy of each molecule can be further decomposed as

$$\Delta E_{total} = \Delta E_v + \Delta E_{vdw} + \Delta E_{Coul} + \Delta W_{PB}/\Delta W_{GB} + \Delta W_{np}, \qquad \text{eq. 1-18}$$

representing the changes in valance energy (E_v) (bond, angle, dihedral and improper energy), van der Waal interaction energy (E_{vdw}), Coulombic interactions energy (E_{Coul}) and the polar contribution of solvation free energy (W_{PB}/W_{GB}) and the nonpolar tern with cavity/surface area energy (W_{np}). We note that the solvation free energy here includes water enthalpy and entropy.

Although MM-PBSA and MM-GBSA serve as robust methods of energy calculation in drug design and biomolecular analyses, the studies also revealed some weaknesses of this method [37, 38]. Source errors include the prediction of solute entropies, the estimation of solvation free energies and parameter selection. Moreover, contributions of water molecules to the binding free energy are not accurate by applying implicit solvation models. Despite several limitations, this method is still with great potential to quantitatively compute binding energy in various molecular systems at low computational cost.

1.3 Applications

MD simulations have more than 40 years of history [39, 40]. Along the path MD simulations have evolved rapidly and now achieved time scales up to microseconds (μs) / milliseconds (ms), which is compatible with biological processes. At present, processes like conformational changes and ligand binding/unbinding can be effectively simulated. The improvement of the computational power, especially the use of GPUs, and the improvements made in the optimization of MD algorithms, including enhanced sampling methods like accelerated MD, allow us to move from the analysis of single structures, the basis of the molecular modeling as we know it, to the analysis of biological processes that requires large conformational changes.

MD has been widely applied in many fields of biomolecular studies and there are many advanced methods derived from MD simulations. For example, accurate calculations of absolute free energy has always been a great chanllenge due to insufficient sampling in a

finite length and time scale simulation. Different methods have been developed to estimate free energy, such as thermodynamical integration, free energy perturbation, mining minima and umbrella sampling. Ligand association and dissociation processes have significant meaning in understand the binding/unbinding kinetics and development of new drugs. With the help of various enhanced sampling methods, we can now fully simulate association or dissociation processes of ligands.

1.3.1 Free energy calculations

1.3.1.1 The concepts

The difference in free energy between two states will determine the probability of finding a molecular system in one state or the other. As a consequence, free-energy differences are directly related to a wide range of important chemical quantities such as binding constants, solubilities and partition coefficients. By means of statistical mechanics, thermodynamic values, such as total energy, entropy, free energy, and pressure, can be be expressed in terms of averages over ensembles of atomic configurations for the molecular system of interest, in other words, derived from the partition function of a system. In modeling of molecular recognition processes, the calculation of these thermodynamic properties can allow us to investigate into the stability of a drug bound to protein receptor, and more importantly, to rank the relative affinity of a series of drug candidates to a target protein.

One commonly used ensemble in MD simulations is the canonical ensemble, which is used for investigation of a molecular system with a fixed number of particles, a specific volume and temperature. The classical canonical ensemble partition function Z is given by,

$$Z = N \int d\mathbf{x} \exp[-\beta E(\mathbf{x})] \qquad \text{eq. 1-19}$$

where $\mathbf{x}$ is condition of the system that defines the microstate (momentum and position), β is the Boltzmann factor $(k_b T)^{-1}$, k_b is the Boltzmann's constant, T is temperature, E denotes the total energy function of the system, and N is a constant that renders Z unitless. The probability of the system in a microstate defined by x is given by,

$$P(\mathbf{x}) = \frac{N \exp[-\beta E(\mathbf{x})]}{Z} = \frac{\exp[-\beta E(\mathbf{x})]}{\int d\mathbf{x} \exp[-\beta E(\mathbf{x})]}. \qquad \text{eq. 1-20}$$

The probability function is then used to calculate ensemble averages, or expectation values, of an observable, such as the total energy, E, of a system:

$$<E> = \int d\mathbf{x}\, E(\mathbf{x}) P(\mathbf{x}) = \frac{N \int d\mathbf{x}\, E(\mathbf{x}) \exp[-\beta E(\mathbf{x})]}{Z}, \qquad \text{eq. 1-21}$$

The Helmholtz free energy A, which is applied in the constant volume condition, is given in terms of a temperature T and the partition function:

$$A = -k_b T \ln Z, \qquad \text{eq. 1-22}$$

The entropy can also be defined according to the probability density function:

$$S = -k_b \int d\mathbf{x}\, P(\mathbf{x}) ln P(\mathbf{x}),$$
<div align="right">eq. 1-23</div>

which is used to calculate configuration entropy before. The entropy can also be derived as following function:

$$S = -\left(\frac{\partial A}{\partial T}\right)_v$$

$$= k_b \ln Z + \frac{k_B T}{Z}\left(\frac{\partial Z}{\partial T}\right)_v$$

$$= k_b \ln Z + \frac{N}{TZ}\int dx\, E(x) \exp[-\beta E(x)]$$

$$= \frac{-A + <E>}{T}.$$
<div align="right">eq. 1-24</div>

Traditional free energy calculation methods include thermodynamical integration, free energy perturbation. However, with the big conformational changes of receptor upon binding of different ligands, they are not suitable. Here, we are going to introduce a relatively new method called mining minima (M2).

1.3.1.2 Mining minima method

Mining minima (M2) method is one of the end point methods. It is solidly grounded in a statistical thermodynamics approach to the prediction of binding affinities. It aims to predict standard free energies of binding. For a standard molecular recognition of a protein-ligand system, the reaction can be viewed as the formation of a complex *RL* by a receptor *R* and a ligand molecule *L* in equilibrium [41], which is expressed as

$$R + L \rightleftharpoons RL .$$ eq. 1-25

Experiments can determine standard binding free energies through the equilibrium constant Kb, where the equilibrium constant itself is determined via the receptor, ligand, receptor concentrations along with the standard concentration.

$$\Delta G_b^0 = -k_B T \ln K_b,$$ eq. 1-26

$$K_b = [\frac{[RL]/C^0}{([R]/C^0)([L]/C^0)}]_{eq},$$ eq. 1-27

The standard binding free energy of a non-covalent complex can be calculated via the standard chemical potentials of the receptor, ligand, and receptor-ligand complex.

$$\Delta G^0 = \mu_{RL}^0 - \mu_R^0 - \mu_L^0,$$ eq. 1-28

where μ_x^0 is the standard chemical potential of molecule $x = RL$, R and L, which can be expressed in terms of the configurational integral Z.

$$\mu^0 = -RTln(\frac{8\pi^2}{C^0} Z)$$ eq. 1-29

$$Z = \int d\boldsymbol{r} \exp [-\frac{U(\mathbf{r})+W(\mathbf{r})}{RT}],$$ eq. 1-30

where R is the gas constant, C^0 provides a correction to the standard state. U(r) is the potential energy and W(r) is the solvation energy.

Formally, the configuration integral must consider all space along the remaining internal degrees of freedom, however, M2 method approximates this using the concept of mining

minima. The central idea of the M2 method is to replace the configurational integral over all space with a sum over separate local configurational integrals associated with low energy minima of the system. The algorithm of the M2 method can be divided into several parts shown in the following [42].

First, M2 method uses the Tork algorithm to search stable conformations of molecules, including protein, ligand and complex [43]. One minimized conformation will be provided as an initial search structure. Natural motions of the molecules can be calculated by computing normal modes using bond-angle-torsion (BAT) coordinates. Next, the molecules are distorted along each torsional angle of the modes to generate new conformations with low energy. These iterations will keep processing until the free energy of the system is converged to the criteria, i.e. 10^{-4} kcal/mol. The conformations generated by Tork are then filtered to delete the repeat conformations for further calculations.

Numerical estimation of configurational integrals is used to compute the binding free energy. An implicit solvent model is applied to the molecular system. The configuration integrals Z can be approximated by a sum of local configurational integral z_i over low energy conformations.

$$z_i = \int_i d\mathbf{r} \exp\left[-\frac{U(\mathbf{r})+W(\mathbf{r})}{RT}\right],$$
eq. 1-31

$$G^0 = -RTln\left(\frac{8\pi^2}{C^0}\sum_i z_i\right)$$
eq. 1-32

where $U(r)$ and $W(r)$ are the vacuum potential energy and the solvation energy along the coordinates r, respectively. In BAT coordinates, the configurational integral of an energy well can be written as

$$z_i = b_i^2 \int_i d\vec{b} d\vec{\theta} d\vec{\phi} \prod_{i=3}^{N_{atom}} (b_i^2 \sin \theta_i) \exp[\frac{E(\vec{b},\vec{\theta},\vec{\phi})}{RT}], \qquad \text{eq. 1-33}$$

where $E = U + W$ and $b_i, \theta_i \ and \ \phi_i$ are bond length, angle and dihedral angle of atom i, respectively. In the standard harmonic approximation, the energy E is related to the second derivative matrix of E. Through diagonalizing this matrix and calculating each eigenvector, the configurational integral can be further modified and approximated numerically. This is the harmonic approximation/mode-scanning method.

Calculations of the local configurational integrals Zi allows a probability to be associated with each energy well, which allows a Boltzmann averaged energy <U+W> to be determined. Configurational entropy S_{config} can be calculated by subtracting $\langle U + W$ from the total free energy

$$-T\Delta S^0_{config} = G^0 - \langle U + W \rangle \qquad \text{eq. 1-34}$$

To speed up the calculation for large biomolecular systems, M2 program treats part of protein rigid, called real set, and part of protein flexible, called live set. The rigid set of the molecule acts as a framework that can hold the structure. The acceleration of M2 method can be reached by neglecting the rigid part during conformational search or free energy calculations. The GB solvation model is applied during the conformational search

procedure, and more accurate PB model is used for the free energy calculations of each energy minima.

The mining minima algorithm serves as a method which can perform aggressive conformational searches and more accurate calculations of binding free energy. Broad applications include the study of molecular association, such as host-guest systems [42, 44, 45]. The applications of large system to study protein-ligand binding affinities are also reported [46].

1.3.2 Protein-ligand association/dissociation

Non-covalent molecular recognition plays a crucial role in biology, chemistry and pharmaceutics. Exploring binding/unbinding pathways will help elucidate mechanisms that allostery, induced fit, and the free energetics of association/dissociation, which will later guide molecular designs. Kinetic properties are also important in drug activity [47-49]. For example, the mean residence time (RT) in the binding pocket of a drug can be quantified by dissociation rate constant through the equation $RT = 1/k_{off.}$ Because longer residence time means the drug stays longer in target site, leading to better drug efficacy, exploring binding kinetics has extreme practical meaning in drug design. However, association and dissociation processes are mostly at microsecond or even second or minute scale, due to limitations of computer power, it has been challenging to bring a full picture of molecular recognition. Besides the long time scale of the process, in order to fully sample an association or dissociation process, the system has to cross multiple energy barriers and there are enormous thermo-states need to sample, not to mention

complicated details, such as protein rearrangements, ligand rearrangements, solvent effects, and detaild molecular interactions.

Coarse-grained simulation is useful with large systems or long simulations [50], but it's in compense of neglecting molecular details. Another way to improve of efficiency of MD is through enhanced sampling simulations. Various enhanced sampling simulation methods have been developed over the past decade. Approaches like hyperdynamics [51], accelerated MD [52] and Gaussian accelerated MD (GaMD) [53] accelerate the simulation by raising the potential energy well to lower the energy barrier. Another group of enhanced sampling methods improve sampling by employing additional forces to the region of interest, including steered MD [54, 55], target MD [56] and self-guided Langevin dynamics (SGLD) [57, 58]. There are also other important MD based simulation methods that can enhance sampling. LowModeMD [59] uses low frequency modes from normal mode analysis to guide MD simulation. Pathway search guided by internal motions (PSIM) search for dissociation along principle component motions. Replica exchange molecular dynamics (REMD) [60, 61] enables configurations at high temperatures to the simulations at low temperatures and vice versa, thus sampling both low and high energy configurations efficiently. Umbrella sampling apply multiple overlapping biasing potentials along the association/dissociation pathway in order to sample all points suffciently. Here, we will mainly introduce three methods used in the following sections, accelerated MD, PSIM and umbrella sampling.

1.3.2.1 Accelerated molecular dynamics simulation

Accelerated MD (aMD) [52] is a widely used enhanced sampling method. The central idea of aMD is to improve the conformational space sampling by reducing energy barriers between different states of a system. The method modifies the potential energy landscape by raising energy wells that are below a reference energy, while leaving those above this energy unaffected. As a result, barriers between adjacent energy basins are reduced, allowing the system to cross the energy barriers much easier than that in a conventional MD simulation. When potential energy of the system falls below a reference energy E, a boost potential energy $\Delta V(r)$ will be added to make the modified potential energy $V^*(r)$ as a sum of the original potential energy $V(r)$ and a boost potential energy.

The aMD modification of the potential is defined by the following equation:

$$V^*(r) = V(r) + \Delta V(r), \qquad\qquad \text{eq. 1-35}$$

$$\Delta V(r) = \begin{cases} 0 & if\ V(r) \geq E \\ \frac{(E-V(r))^2}{\alpha+E-V(r)} & if\ V(r) < E \end{cases}, \qquad\qquad \text{eq. 1-36}$$

where V(r) is the original potential, E is the reference energy, and V*(r) is the modified potential. $\Delta V(r)$ is the boost potential, α is the acceleration factor that determines the shape of the modified potential.

1.3.2.2 Pathway search guided by internal motions

Pathway search guided by internal motions (PSIM) applies internal principle component (PC) modes computed from a MD trajectory to guide the searches. Internal coordinates and selected dihedrals instead of Cartesian coordinate are used to compute PC modes. Due to the fact that biomolecules are significantly larger than chemical compounds, multi-layer internal coordinates are developed. First of all, the internal PC modes are sorted by the extend the ligand is driven to move. The conformational search is conducted in a systematic manner. First, the initial conformation is distorted stepwise along both positive and negative directions of each PC modes until new conformation is rejected by the acceptance test. Next, starting from the new conformations, the conformational search continues along two directions of each mode as the starting point after the PC modes screened before being used in the sub branches. During the conformational search, PSIM will track the used modes and the number of steps of distortion to generate multiple trajectories. When the number of steps reaches a predefined parameter, i.e. 100 steps, a motion test is performed to make sure that the ligand moves away from its last position and the protein maintains a reasonable conformation while being distorted. A failed motion test will block the search along the current search path. If the motion test is passed, a short minimization will be performed to reduce the artifacts induced by the distortion, and the search will be continued from the corrected conformation. The search algorithm will stop the distortion path after the total step of distortion has exceeded a predefined path length, i.e. 2000 steps, and the algorithm will be exited after the systematic search exhausts the search tree.

1.3.2.3 Umbrella sampling

Umbrella sampling searves as both conformational sampling method and free energy calculation method for association and dissociation pathways. By adding multiple overlapping biasing potentials along the dissociation pathway as the reaction coordinate (RC), umbrella sampling can sample all points on the RC suffciently. First, the whole RC is divided into a series of continuous windows, then a biased potential, mostly harmonic potential, is applied to add on the original potential in each window. The equation of harmonic potential is shown as below:

$$u_i = k_i (r - r_i)^2 \qquad\qquad \text{eq. 1-37}$$

where u_i is the biased potential in window i, r is the current position of RC, r_i is the reference position in window i, and k_i is the force constant used to restrain the biased molecule in the biased potential. Many programs are available for construction of the potential of the mean force (PMF) along the RC using data from umbrella sampling, such as WHAM and MBAR [62, 63].

1.4 References

1.   McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977;267(5612):585-90.   doi: 10.1038/267585a0.   PubMed   PMID: WOS:A1977DJ69100017.

2.   Amrhein S, Oelmeier SA, Dismer F, Hubbuch J. Molecular dynamics simulations approach for the characterization of peptides with respect to hydrophobicity. J Phys Chem   B.   2014;118(7):1707-14.   doi: 10.1021/jp407390f.   PubMed   PMID: WOS:000331861500001.

3.   Zhang J, Zhang Y. Molecular dynamics studies on the NMR and X-ray structures of rabbit prion proteins. J Theor Biol. 2014;342:70-82. doi: 10.1016/j.jtbi.2013.10.005. PubMed PMID: WOS:000330257900008.

4.   Nguyen THT, Liu Z, Moore PB. Molecular dynamics simulations of homo-oligomeric bundles embedded within a lipid bilayer. Biophys J. 2013;105(7):1569-80. doi: 10.1016/j.bpj.2013.07.053. PubMed PMID: WOS:000325383000006.

5.   Tumbi KM, Nandekar PP, Shaikh N, Kesharwani SS, Sangamwar AT. Molecular dynamics simulation studies for DNA sequence recognition by reactive metabolites of anticancer compounds. J Mol Recognit. 2014;27(3):138-50. doi: 10.1002/jmr.2342. PubMed PMID: WOS:000331064000004.

6.   de Almeida H, Bastos IMD, Ribeiro BM, Maigret B, Santana JM. New binding site conformations of the dengue virus NS3 protease accessed by molecular dynamics simulation. PLoS ONE. 2013;8(8):e72402. doi: 10.1371/journal.pone.0072402. PubMed PMID: WOS:000324470100102.

7.   Alder BJ, Wainwright TE. Studies in molecular dynamics .1. General method. J Chem Phys. 1959;31(2):459-66. doi: 10.1063/1.1730376. PubMed PMID: WOS:A1959WA66500032.

8.   Frenkel D, Smit B. Understanding molecular simulation from algorithms to applications. Academic Press, California. 2002.

9.   Rahman A. Correlations in motion of atoms in liquid argon. Phys Rev A. 1964;136(2A):A405. doi: 10.1103/PhysRev.136.A405. PubMed PMID: WOS:A19641532C00031.

10.   Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins: Struct, Funct, Bioinf. 2006;65(3):712-25. doi: 10.1002/prot.21123. PubMed PMID: WOS:000241247100017.

11.   Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005;26(16):1668-88. doi: 10.1002/jcc.20290. PubMed PMID: WOS:000233021400002.

12.   Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-Atom additive biological force fields. J Comput Chem. 2010;31(4):671-90. doi: 10.1002/jcc.21367. PubMed PMID: WOS:000274922000002.

13.   Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, et al. The GROMOS biomolecular simulation program package. J Phys Chem A.

1999;103(19):3596-607. doi: 10.1021/jp984217f. PubMed PMID: WOS:000080552200002.

14. Jorgensen WL, Maxwell DS, TiradoRives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc. 1996;118(45):11225-36. doi: 10.1021/ja9621760. PubMed PMID: WOS:A1996VT46600030.

15. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct. 2000;29:291-325. doi: 10.1146/annurev.biophys.29.1.291. PubMed PMID: WOS:000088492300011.

16. Song Y, Mao J, Gunner MR. MCCE2: Improving protein pK(a) calculations with extensive side chain rotamer sampling. J Comput Chem. 2009;30(14):2231-47. doi: 10.1002/jcc.21222. PubMed PMID: WOS:000269939600008.

17. Bas DC, Rogers DM, Jensen JH. Very fast prediction and rationalization of pK(a) values for protein-ligand complexes. Proteins: Struct, Funct, Bioinf. 2008;73(3):765-83. doi: 10.1002/prot.22102. PubMed PMID: WOS:000260515500020.

18. Davis ME, Madura JD, Luty BA, McCammon JA. Electrostatics and diffusion of molecules in solutions - simulations with the university-of-houston-brownian dynamics program. Comput Phys Commun. 1991;62(2-3):187-97. doi: 10.1016/0010-4655(91)90094-2. PubMed PMID: WOS:A1991FF81900004.

19.   Curry HB. The method of steepest descent for non-linear minimization problems. Quarterly of Applied Mathematics. 1944;2:258-61. PubMed PMID: INSPEC:1946A00364.

20.   Stich I, Car R, Parrinello M, Baroni S. Conjugate-gradient minimization of the energy functional - A new method for electronic-structure calculation. Physical Review B. 1989;39(8):4997-5004. doi: 10.1103/PhysRevB.39.4997. PubMed PMID: WOS:A1989T691200018.

21.   Jensen F. Introduction to computational chemistry: John Wiley and Sons Ltd.; 1999.

22.   Jorgensen WL. Quantum and statistical mechanical studies of liquids .10. transferable intermolecular potential functions for water, alcohols, and ethers - application to liquid water. J Am Chem Soc. 1981;103(2):335-40. doi: 10.1021/ja00392a016. PubMed PMID: WOS:A1981KZ44800016.

23.   Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79(2):926-35. PubMed PMID: WOS:A1983QZ31500046.

24.   Mahoney MW, Jorgensen WL. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J Chem Phys. 2000;112(20):8910-22. doi: 10.1063/1.481505. PubMed PMID: WOS:000086851200022.

25.   Shestakov AI, Milovich JL, Noy A. Solution of the nonlinear Poisson-Boltzmann equation using pseudo-transient continuation and the finite element method. J Colloid

Interface Sci. 2002;247(1):62-79. doi: 10.1006/jcis.2001.8033. PubMed PMID: WOS:000173998300008.

26. Lu BZ, Zhang DQ, McCammon JA. Computation of electrostatic forces between solvated molecules determined by the Poisson-Boltzmann equation using a boundary element method. J Chem Phys. 2005;122(21):1297-304. doi: 10.1063/1.1924448. PubMed PMID: WOS:000229700600002.

27. Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular Simulations. Biopolymers. 2000;56(4):275-91. PubMed PMID: WOS:000173077900005.

28. Onufriev A, Case DA, Bashford D. Effective Born radii in the generalized Born approximation: The importance of being perfect. J Comput Chem. 2002;23(14):1297-304. doi: 10.1002/jcc.10126. PubMed PMID: WOS:000178349800001.

29. Andersen HC. Molecular dynamics simulations at constant pressure and-or temperature. J Chem Phys. 1980;72(4):2384-93. doi: 10.1063/1.439486. PubMed PMID: WOS:A1980JK06800026.

30. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26(16):1781-802. doi: 10.1002/jcc.20289. PubMed PMID: WOS:000233021400007.

31. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM - A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem. 1983;4(2):187-217. doi: 10.1002/jcc.540040211. PubMed PMID: WOS:A1983QP42300010.

32.   Ai R, Fatmi MQ, Chang C-eA. T-Analyst: A program for efficient analysis of protein conformational changes by torsion angles. J Comput-Aided Mol Des. 2010; 24(10):819-27.   doi: :10.1007/s10822-010-9376-y.   PubMed   PMID: BIOSIS:PREV201000574011.

33.   Pitzer KS. Energy levels and thermodynamic functions for molecules with internal rotation .2. unsymmetrical tops attached to a a rigid frame. J Chem Phys. 1946;14(4):239-43. PubMed PMID: WOS:A1946UB22900003.

34.   Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. J Am Chem   Soc.   1998;120(37):9401-9.   doi: 10.1021/ja981844+.   PubMed   PMID: WOS:000076117200001.

35.   Massova I, Kollman PA. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. Perspect Drug Discov. 2000;18:113-35. PubMed PMID: WOS:000086013500006.

36.   Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model. 2011;51:14.

37.   Huo SH, Wang JM, Cieplak P, Kollman PA, Kuntz ID. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: Insight into structure-based ligand design. J Med Chem. 2002;45(7):1412-9. doi: 10.1021/jm010338j. PubMed PMID: WOS:000174588100004.

38.   Homeyer N, Gohlke H. Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. Molecular Informatics. 2012;31(2):114-22. doi: 10.1002/minf.201100135. PubMed PMID: WOS:000300675200001.

39.   Levitt M, Warshel A. Computer simulation protein folding. Nature. 1975; 253(5494):694-8. doi: 10.1038/253694a0. PubMed PMID: WOS:A1975V689000024.

40.   Warshel A. Bicycle-pedal for 1st step in vision process. Nature. 1976; 260(5553):679-83. doi: 10.1038/260679a0. PubMed PMID: WOS:A1976BN79200020.

41.   Gallicchio E, Levy RM. Recent theoretical and computational advances for modeling protein-ligand binding affinities. Adv Protein Chem Struct Biol. 2011;85:27-80. doi: 10.1016/b978-0-12-386485-7.00002-8. PubMed PMID: WOS:000295817200002.

42.   Chang CE, Gilson MK. Free energy, entropy, and induced fit in host-guest recognition: Calculations with the second-generation mining minima algorithm. J Am Chem Soc. 2004;126(40):13156-64. doi: 10.1021/ja047115d. PubMed PMID: WOS:000224357700085.

43.   Chang CE, Gilson MK. Tork: Conformational analysis method for molecules and complexes. J Comput Chem. 2003;24(16):1987-98. doi: 10.1002/jcc.10325. PubMed PMID: WOS:000186375200004.

44.   Rekharsky MV, Mori T, Yang C, Ko YH, Selvapalam N, Kim H, et al. A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy-entropy compensation. Proc Natl Acad Sci U S A. 2007;104(52):20737-42. doi: 10.1073/pnas.0706407105. PubMed PMID: WOS:000252077400023.

45.   Moghaddam S, Inoue Y, Gilson MK. Host-guest complexes with protein-ligand-like affinities: Computational analysis and design. J Am Chem Soc. 2009;131(11):4012-21. doi: 10.1021/ja808175m. PubMed PMID: WOS:000264792700059.

46.   Chen W, Gilson MK, Webb SP, Potter MJ. Modeling protein-ligand binding by mining minima. J Chem Theory Comput. 2010;6(11):3540-57. doi: 10.1021/ct100245n. PubMed PMID: WOS:000283884300024.

47.   Janin J. The kinetics of protein-protein recognition. Proteins: Struct, Funct, Bioinf. 1997;28(2):153-61. doi: 10.1002/(sici)1097-0134(199706)28:2<153::aid-prot4>3.0.co;2-g. PubMed PMID: WOS:A1997XD03700004.

48.   Schreiber G, Haran G, Zhou HX. Fundamental aspects of protein-protein association kinetics. Chem Rev. 2009;109(3):839-60. doi: 10.1021/cr800373w. PubMed PMID: WOS:000264199700001.

49.   Zhou H-X. Rate theories for biologists. Q Rev Biophys. 2010;43(2):219-93. doi: 10.1017/s0033583510000120. PubMed PMID: WOS:000282243900003.

50.   Orozco M, Orellana L, Hospital A, Naganathan AN, Emperador A, Carrillo O, et al. Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. Advances in protein chemistry and structural biology. 2011;85:183-215. Epub 2011/09/17. doi: 10.1016/b978-0-12-386485-7.00005-3. PubMed PMID: 21920324.

51.   Voter AF. Hyperdynamics: Accelerated molecular dynamics of infrequent events. Physical Review Letters. 1997;78(20):3908-11. doi: 10.1103/PhysRevLett.78.3908. PubMed PMID: WOS:A1997WZ24400027.

52.   Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. Journal of Chemical Physics. 2004;120(24):11919-29.   doi: 10.1063/1.1755656.   PubMed   PMID: WOS:000221907800065.

53.   Miao Y, Feher VA, McCammon JA. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. Journal of Chemical Theory and Computation. 2015;11(8):3584-95. doi: 10.1021/acs.jctc.5b00436. PubMed PMID: WOS:000359500000007.

54.   Leech J, Prins JF, Hermans J. SMD: Visual steering of molecular dynamics for protein design. Ieee Computational Science & Engineering. 1996;3(4):38-45. doi: 10.1109/99.556511. PubMed PMID: WOS:A1996VZ99800007.

55.   Nicolini P, Frezzato D, Gellini C, Bizzarri M, Chelli R. Toward quantitative estimates of binding affinities for protein-ligand systems involving large inhibitor compounds: A steered molecular dynamics simulation route. Journal of Computational Chemistry.   2013;34(18):1561-76.   doi: 10.1002/jcc.23286.   PubMed   PMID: WOS:000320177700005.

56.   Schlitter J, Engels M, Kruger P. TARGETED MOLECULAR-DYNAMICS - A NEW APPROACH FOR SEARCHING PATHWAYS OF CONFORMATIONAL TRANSITIONS. Journal of Molecular Graphics. 1994;12(2):84-9. doi: 10.1016/0263-7855(94)80072-3. PubMed PMID: WOS:A1994NT17800002.

57.   Wu XW, Brooks BR. Self-guided Langevin dynamics simulation method. Chemical Physics Letters. 2003;381(3-4):512-8. doi: 10.1016/j.cplett.2003.10.013. PubMed PMID: WOS:000186598200038.

58.   Wu X, Brooks BR, Vanden-Eijnden E. Self-Guided Langevin Dynamics via Generalized Langevin Equation. Journal of Computational Chemistry. 2016;37(6):595-601. PubMed PMID: CCC:000370254600007.

59.   Labute P. LowModeMD-Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops. Journal of Chemical Information and Modeling. 2010;50(5):792-800. doi: 10.1021/ci900508k. PubMed PMID: WOS:000277911600008.

60.   Voter AF. Parallel replica method for dynamics of infrequent events. Physical Review B. 1998;57(22):13985-8. PubMed PMID: WOS:000074252500007.

61.   Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chemical Physics Letters. 1999;314(1-2):141-51. doi: 10.1016/s0009-2614(99)01123-9. PubMed PMID: WOS:000083955300022.

62.   Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. Journal of computational chemistry. 1992;13(8):1011-21. doi: 10.1002/jcc.540130812.

63.   Souaille M, Roux Bt. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Computer physics communications. 2001;135(1):40-57. doi: http://dx.doi.org/10.1016/S0010-4655(00)00215-0.

Chapter 2 Characterization of Promiscuous Binding of Phosphor Ligands to Breast-Cancer-Gene 1 (BRCA1) C-Terminal (BRCT): Molecular Dynamics, Free Energy, Entropy and Inhibitor Design

2.1 Introduction

The tandem ~100-amino acid repeats of breast-cancer-gene 1 (BRCA1) C-terminal (BRCT) are known to bind to phosphorylated proteins which are important for a number of tumor suppressor functions, which include, DNA repair, cell-cycle checkpoint, and transcription regulation [1-4]. The BRCT repeats recognize and bind phosphorylated protein partners such as CCDC98/Abraxas, BACH1 and CtIP in response to DNA damage [5-10]. Mutations in the BRCT domain of BRCA1 predispose women to breast and ovarian cancers [11]. A recent study showed that inhibitors of BRCT(BRCA1)–phosphoprotein interface can be combined with DNA damaging agents as a viable therapeutic strategy for non-BRCA mutation carriers [12]. The same binding interface on BRCT(BRCA1) promiscuously interacts with various phosphoproteins and short phosphopeptides containing the pSer-X-X-Phe sequence, where X denotes any residue [5-10]. Several modular domains, such as SH3, SH2, FHA, WW, Polo-box and PDZ, are also known to interact with multiple proteins through a consensus recognition sequence [13-18]. Here, we investigated the promiscuous recognition of the BRCT(BRCA1) domain to better understand the mechanism that drives diverse ligands to bind to the same binding site. Our studies will provide insights into molecular detection, inhibitor discovery, and the search for binding partners.

44

The BRCT (BRCA1) domain is a tandem repeat; each N-terminal BRCT and C-terminal BRCT contain 90-100 residues with a central four-stranded β sheet (β1-β4 and β1′-β4′) and three α-helices (α1-α3 and α1′-α3′). The BRCT–pSXXF interaction is anchored via a two-point binding mode: a hydrophilic contact made by the phosphoserine (pS) residue formed by N-terminal BRCT and a hydrophobic binding pocket from C-terminal BRCT for the phenylalanine (F) residue (Figure 2.1). The two-point binding scheme is also conserved for compounds with phosphate groups via a phosphate group and a hydrophobic ring group. Unlike most classical pharmaceutical targets such as enzymes with very defined binding cavity, the mostly solvent-exposed and plastic binding pockets such as the phosphoprotein binding interface of BRCT (BRCA1) were considered un-druggable years ago [19-21].
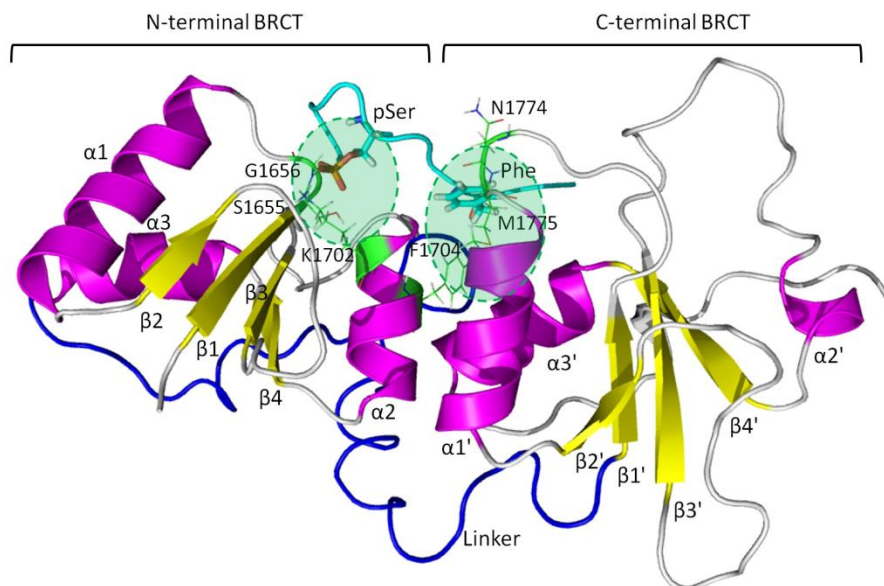
Figure 2.1. Breast-cancer-gene 1 (BRCA1) C-terminal (BRCT) binding with a phosphoserine (pSer) peptide. pSer forms hydrogen bonds with S1655, G1656 and K1702, and the P+3 Phe locates in the hydrophobic packet formed by M1775, N1774 and F1704. The two points of contact (pSer and P+3 Phe) shown in all of our calculations are highlighted by green circles.

The phosphopeptides are successful inhibitors of protein–protein interactions (PPI) [22-24]. Recently, many new PPI inhibitors have been developed for the BRCT domain, which include a number of short pSXXF tetra-phosphopeptides [12, 24-26] and new phosphopeptide analogs with phosphate groups [27, 28]. Although challenging to design, the demand for inhibitors of PPI has steadily increased [29, 30]. Significant progress has been made in developing inhibitors targeting PPIs, and the development of effective therapeutics from PPI inhibitors will be improved by both experimental and computational approaches.

Recent advances in computer modeling have provided powerful tools to study peptide-domains binding and protein dynamics. Molecular dynamics (MD), Brownian dynamics simulations, and molecular docking have been used to investigate BRCT dynamics, interactions between inhibitors and BRCT, and the ligand association processes [25, 27, 31, 32]. Bioinformatics tools were used to assess the functional impact and likelihood of pathogenicity of variants in the BRCT domain [33, 34]. The promiscuous recognition of BRCT also makes it convenient to investigate the relationship between binding entropy and enthalpy changes. In addition to BRCT, other modular domains serve as good model systems for inspecting promiscuous recognition and the paradox associated with changes in entropy and enthalpy upon ligand binding that targets PPIs by computational methods [35-43].

This study aimed to further understand ligand–BRCT binding and provide strategies for designing inhibitors of PPIs. We selected several tetrapeptides and compounds with phosphate groups to computationally evaluate their driving forces to bind to BRCT(BRCA1). We performed MD simulations and detailed analysis of MD trajectories to examine the approaches BRCT uses to achieve promiscuous binding and the interaction energy of the ligand-BRCT. The MD simulations illustrated the molecular flexibility in the free and bound states for BRCT(BRCA1) and ligands. We analyzed loop movements and the population of dihedral rotations of backbone and side-chains. Conformations from MD simulations were used as initial structures for thorough conformational search and free energy calculations with the M2 method, to reveal the contribution of configuration entropy and enthalpy to ligand binding affinities. We

focused on how to optimize the balance between enthalpy gain and entropy loss. Using an accepted practice in ligand design, we synthesized a ligand that incorporates a benzene ring to possibly constrain its conformation. Upon ligand binding, changes of each energy term, conformations, rotameric state, and configurational entropy were evaluated by both MD and M2 tools; and the findings were used to suggest new inhibitors.

2.2 Materials and Methods

2.2.1 Molecular systems

Table 2.1 lists 14 short peptides (P1-P14, among which P11 contains a phosphate mimic and others contain phosphorylated amino acids) [24], one compound (C1) [27], one new compound (N1), one designed compound (D1) and 4 long phosphopeptides (L1-L4) that bind to the BRCT domain: pS and pT is phosphorylated amino acid serine and threonine, respectively, and γcE is γ-carboxyglutamate, which is chosen to mimic pS interaction as a non-phosphorylated peptide binder.

Table 2.1. Ligand library of BRCT used for binding affinity exploration and study of flexibility of binding site. The major binding residues, pSer and Phe (P+3), are in bold. The relative binding free energy for ligand X (X=P2-P14, C1, N1) to ligand P1 is approximated using the half maximal inhibitory concentration IC50 as $\Delta\Delta G_{exp} = RT \ln$ IC50(X)/IC50(P1) based on equation $\Delta G = RT \ln K_d = RT \ln(\text{IC50} + 0.5 C_{enzyme}) \approx RT \ln$ IC50 [44, 45]. Binding free energies for L1-L4 are calculated through equation $\Delta G_{exp} = RT \ln (K_d)$.

Tetrapeptides			
No.	**Sequence**	**IC$_{50}$ (µM)** [a]	**$\Delta\Delta G_{exp}$ (kcal/mol)**
P1	Ac-**pS**PT**F**-COOH	1.0±0.2	0
P2	Ac-**pS**PV**F**-COOH	1.6±0.3	0.28
P3	Ac-**pS**PV**F**-CONH$_2$	3.2±0.8	0.69
P4	Ac-**pS**PT**F**-CONH$_2$	4.6±0.9	0.91
P5	Ac-**pS**PI**F**-CONH$_2$	7.1±1.4	1.17
P6	Ac-**pS**PT**Y**-CONH$_2$	14.9±2.8	1.61
P7	Ac-**pS**AT**F**-CONH$_2$	15.0±1.7	1.61
P8	Ac-**pS**PL**F**-CONH$_2$	18.4±1.8	1.74
P9	Ac-**pS**PS**F**-CONH$_2$	30.1±7.2	2.03
P10	Ac-**pS**PA**F**-CONH$_2$	35.0±7.9	2.12
P11	Ac-**γcE**PT**F**-CONH$_2$	52.8±1.6	2.36
P12	Ac-**pS**AA**F**-CONH$_2$	98.4±23.1	2.74
P13	Ac-**pS**PP**F**-CONH$_2$	>250	>3.29
P14	Ac-**pT**PT**F**-CONH$_2$	>250	>3.29
Compounds			
No.	**Structure**	**IC$_{50}$ (µM)**	**$\Delta\Delta G_{exp}$ (kcal/mol)**
C1	See Figure 2.2	0.31	-0.70
N1	See Figure 2.2	>250	>3.29
D1	See Figure 2.2	N/A	N/A
Long peptides			
No.	**Sequence**	**K$_d$ (µM)** [b]	**ΔG_{exp} (kcal/mol)**
L1	ISRST**pS**PT**F**NKQ	0.9	-8.30
L2	PTRVS**pS**PV**F**GA	3.7	-7.46
L3	AAYDI**pS**QV**F**PFA	0.4	-8.78
L4	PQ**pS**PT**F**PEAG	5.2	-7.25

[a] IC50 values of P1-P14 were taken from ref [24]. IC50 values of C1 was taken from ref [27].
[b] K$_d$ values of L1-L4 were taken from ref [46], ref [47], ref [48] and ref [49], respectively.
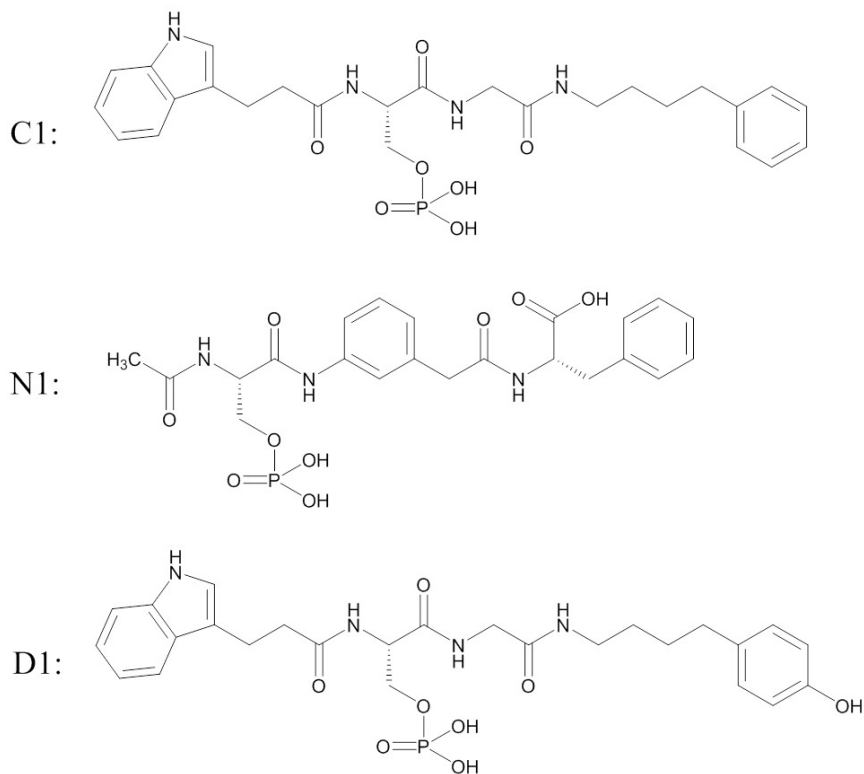
Figure 2.2. Structures of C1, N1 and D1 that bind to the BRCT domain.

2.2.2 Molecular dynamics simulations

We ran MD simulations on BRCT-ligand complexes, free ligands and free protein, and the PDB IDs used as initial structures to perform MD simulations were listed in Table 2.2. The initial bound conformation of all tetrapeptides was generated by superimposing the backbone atoms of -pSXXF- within phosphorylated BACH1 peptide ISRSTpSPTFNKQ in the C-terminal domain of the BRCA1 protein (PDB code 1T29) [46]. Besides 1T29, we included the other three BRCT domain structures in complex with long phosphopeptides from CtIP, ACC1 proteins and library screening, with PDB IDs 1Y98 (PTRVSpSPVFGA), 3COJ (PQpSPTFPEAG) and 1T2V (AAYDIpSQVFPFA),

respectively, for promiscuous molecular recognition study [47-49]. The initial structure of the bound conformation of C1, N1 and D1, where no available crystal structures, were from docking with Autodock tools 1.5.6 [50, 51] and then further checked manually by ensuring important interactions hold. Notably, Autodock was used for only the three ligands that did not have co-crystal structures with BRCT. The docking method used the Lamarckian genetic algorithm, which fixed the protein and allowed the ligand to move around in the docking box. The partial charges of ligands were calculated by using the Vcharge program [52]. The Autodock scoring function is a subset of the AMBER force field that treats molecules using the united atom model. Autogrid version 4.0 was used to create affinity grids with 0.375 Å spacing in 19.5 x 11.25 x 11.25 Å^3 space at binding site. The final docking result was obtained by 10 runs of simulation with 2.5 million rounds of energy evaluation in each run. Ligand conformations with the lowest docked energies and reasonable conformation (pSer forms hydrogen bonds with S1655, G1656 and K1702, and the P+3 Phe locates in the hydrophobic packet formed by M1775, N1774 and F1704) were further analyzed. We selected two initial conformations with similar low energy computed by Autodock for ligands C1 and D1, and N1 has one initial conformation (Figure 2.3).

Table 2.2. Sources of initial bound conformations of ligands for MD simulation.

L1	Crystal structure 1T29
L2	Crystal structure 1Y98
L3	Crystal structure 1T2V
L4	Crystal structure 3COJ
P1-P14	pSXXF sequence superimposed to crystal structure 1T29
C1, N1, D1	Docked to crystal structure 1T29
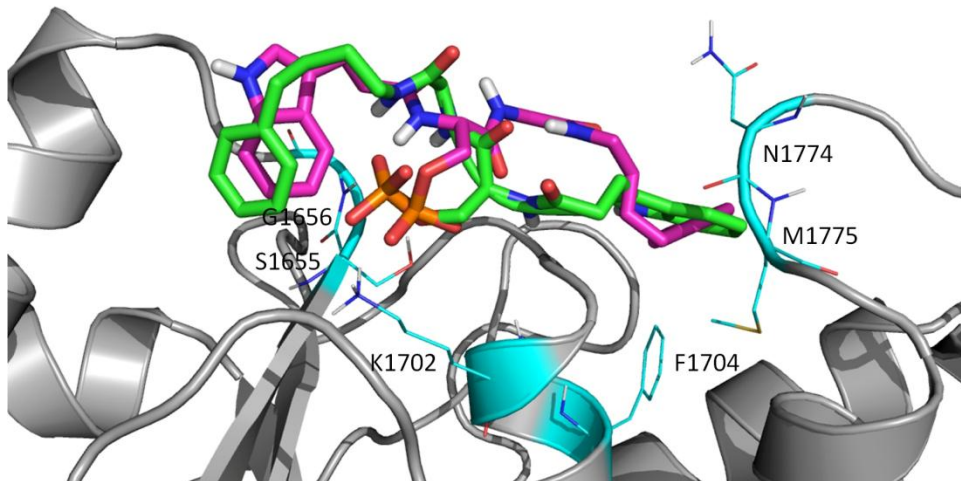
Figure 2.3. Two initial bound structures of C1 from docking. The trajectory that covers the conformations close to the three bound structures in M2 search (Figure 2.11) was further used for MM/PBSA calculation.

We performed MD simulations on an apo BRCT domain, 21 complexes, and 21 free ligands to study the dynamic nature of a given system. The standard simulation package, Amber14 [53] with the Amber 99SB force field [54-57], was used. For pSer and pThr, we used the force field reported by Homeyer *et al* [58]. Amber atom types were manually assigned to non-standard amino acid and functional groups of the ligands C1, N1 and D1. Each system was set up as follows. First, we minimized the hydrogen, side-chain and whole system for 500, 5 000 and 5 000 steps, respectively; then the systems were solvated in a rectangular box of a 12-Å explicit TIP3P water model by the tleap program in Amber14. Each system contains about 50 000 atoms. Counter ions Na^+ were added to keep the whole system neutral, and particle mesh Ewald was used to consider long-range electrostatic interactions [59]. Before equilibration, we ran energy minimization of 10

000 and 20 000 steps for the waters and system, respectively; next, we ran equilibrium of solvent molecules for 40 ps. Then the systems were gradually heated from 250 K for 20 ps, 275 K for 20 ps, and 300 K for 160 ps. We saved a frame every 1 ps with a time step of 2 fs in the isothermic−isobaric (NPT) ensemble. The Langevin thermostat with a damping constant of 2 ps^{-1} was used to maintain a temperature of 300 K, and the hybrid Nosé−Hoover Langevin piston method was used to control the pressure at 1 atm. We also used the SHAKE procedure to constrain hydrogen atoms during MD simulations [60]. Finally, all production runs were performed for 100 ns at 300 K. To ensure that all simulations reached stable energy fluctuations, we considered only trajectories during 20−100 ns for post-analysis.

2.2.3 M2 method

The second-generation mining minima method, M2, calculates the standard free energy of binding by computing the free energy of the free BRCT ($G°_{BRCT}$), ligand ($G°_{ligand}$), and ligand-BRCT complex ($G°_{comp}$).

$$\Delta G^o = G^o_{Comp} - G^o_{BRCT} - G^o_{Ligand} \qquad (1)$$

M2 uses the classical formulation of the partition function for calculating free energy $G°$.

$$G^o \approx -RT \ln\left(\frac{8\pi^2}{c^o}\sum_i Z_i\right) \qquad (2)$$

$$Z_i = \int_{well\,i} e^{-\beta(U(r)+W(r))} dr\_int \qquad (3)$$

where U is potential energy, W is the solvation free energy and Z_i is the local configuration integral from distinct energy wells. The external degrees of freedom were integrated out and C^0 provides a correction to the standard state, and *r_int* indicates the variables of the internal bond-angle-torsion coordinates. Formally, the configuration integral must be determined over all spaces along the remaining internal degrees of freedom. M2 approximates this configuration integral by using the concept of considering local energy minima only [61, 62]. Therefore, the M2 approach replaces the configurational integral over all spaces with a sum over separate local configurational integrals (Z_i) associated with the low energy minima of the system. Determining Z_i allows for the probability to be associated with each energy well, which in turn, allows for determining a Boltzmann averaged energy <U+W>, which is then subtracted from the total free energy to give the system configurational entropy, useful when analyzing and interpreting predicted binding affinities.

$$-TS^o_{config} = G^o - <U+W>$$ (4)

Note that the configurational entropy S°_{config} includes both a conformational part, which reflects the number of energy wells (conformations), and a vibrational part, which reflects the average width of the energy wells. The solvent entropy is included in the solvation free energy, W. Therefore, the computed configurational entropy changes cannot be directly compared with experimentally measured entropy changes, which contain both configurational and solvent entropy.

54

In brief, M2 contains two parts: 1) an aggressive conformational search for distinct low-energy wells, with repeats detected and removed; and 2) an enhanced harmonic approximation for computing the configuration integral Z_i of each well i. Each distinct conformation is energy minimized, first by conjugate gradient method and then by Newton-Raphson method. Both parts involve the Hessian matrix with respect to bond-angle-torsion coordinates, and our harmonic approximation accounts for anharmonicity of eigenvectors of the Hessian matrix with eigenvalues < 2 kcal/mol/Å or 2 kcal/mol/rad. The correlation between different degrees of freedom (e.g., multiple dihedrals may rotate in concert or move with ligand translation/rotation) is captured in the Hessian matrix. We used the VM2 package for the calculation [63-65] and performed three iterations for each ligand and 3 to 10 iterations for the free BRCT and the complexes until the cumulated free energy converged (Figure 2.4). To reduce the computational cost, only parts of BRCT were flexible, called the "live set" (Figure 2.5), which are residues within 7 Å of a long peptide ISRSTpSPTFNKQ in complex with BRCT (PDB code 1T29). The rigid set, called the "real set", contained the residues within 5 Å of the live set. Other atoms not included in these two sets were not considered during the M2 calculations. All ligands were completely flexible and can freely translate and rotate within the binding site, and the same rigid and flexible parts of BRCT were applied to all systems.
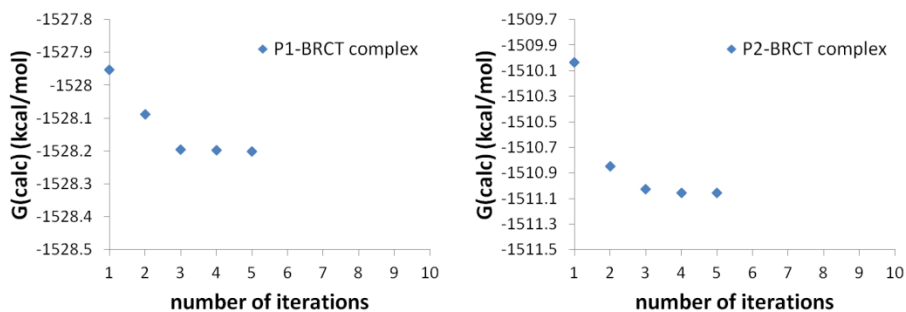
Figure 2.4. Convergence plots for cumulated free energy of complex BRCT and P1/P2.
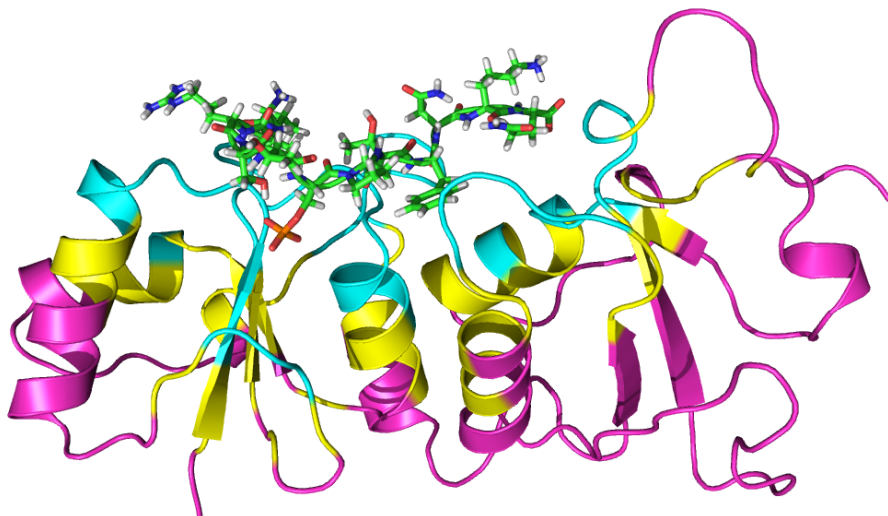
Figure 2.5. BRCT domain with ligand L1. Ligand L1 is shown in green licorice structure. Only residues within 7 Å of the ligand (live set, labeled in blue) is set flexible in M2 calculations. The rigid set (real set, labeled in yellow) contains the residues within 5 Å of the flexible set; other atoms outside the real set (labeled in pink) were not considered during M2 calculations. Notably, the computed entropy and enthalpy terms from M2 consider the contribution of BRCT (live set) and the ligand.

2.2.4 Post-MD analysis: Identifying rotamer states and MM/PBSA calculations

To compare the conformational changes of a molecular system between its free and bound states, we analyzed the selected ligand and BRCT dihedral angles during MD simulations and M2 calculations. Dihedral angles were measured by using T-analyst [66], which can detect the angle population to find discontinuity in a dihedral distribution such as one energy well splitting into two wells near -180 ^{o} and +180 ^{o}. A shifted angle by adding or subtracting 360° is then applied to illustrate proper rotamer states. The population of each dihedral was then plotted by using Matlab with a histogram of 114 bins ranging from -360 ^{o} to +360 ^{o} to ensure coverage of all rotamer states after angle shifting. When analyzing the rotameric states, because the analysis does not need more than 1000 data points [66], we used trajectories with a smaller file size that a frame was saved every 100 ps (1000 frames) for each 100 ns MD run.

We used the molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA)-type post-processing method to compute ligand-BRCT intermolecular interactions during MD simulations [67-76]. The interaction energy, $\Delta(U+W)$ associated with BRCT and a ligand is computed by $\Delta(U+W) = <E_{complex}> - <E_{bound\ BRCT}> - <E_{bound\ ligand}>$. The bracket $<E>$ denotes the average energy computed from a given MD trajectory and the energy terms include a valence term (bond, angle and dihedral), van der Waals (U_{VDW}), Coulombic (U_{Coul}), solvation free energy computed by the Possion-Boltzmann equation (W_{PB}) and by cavity/surface area (W_{NP}). The dielectric constants of the interior and exterior protein

were set to 1 and 80, respectively. The valence term was canceled because of the single

trajectory approach.

2.2.5 Ligand N1 synthesis, purification, and determination of IC50

Peptide synthesis. The peptide with the modified amino acid was synthesized by using

standard Fmoc

chemistry following previously reported methods [24, 77, 78]. The peptide was purified

by preparative LC to >95% as assessed by HPLC and characterized by mass spectrometry

(Figure 2.6).
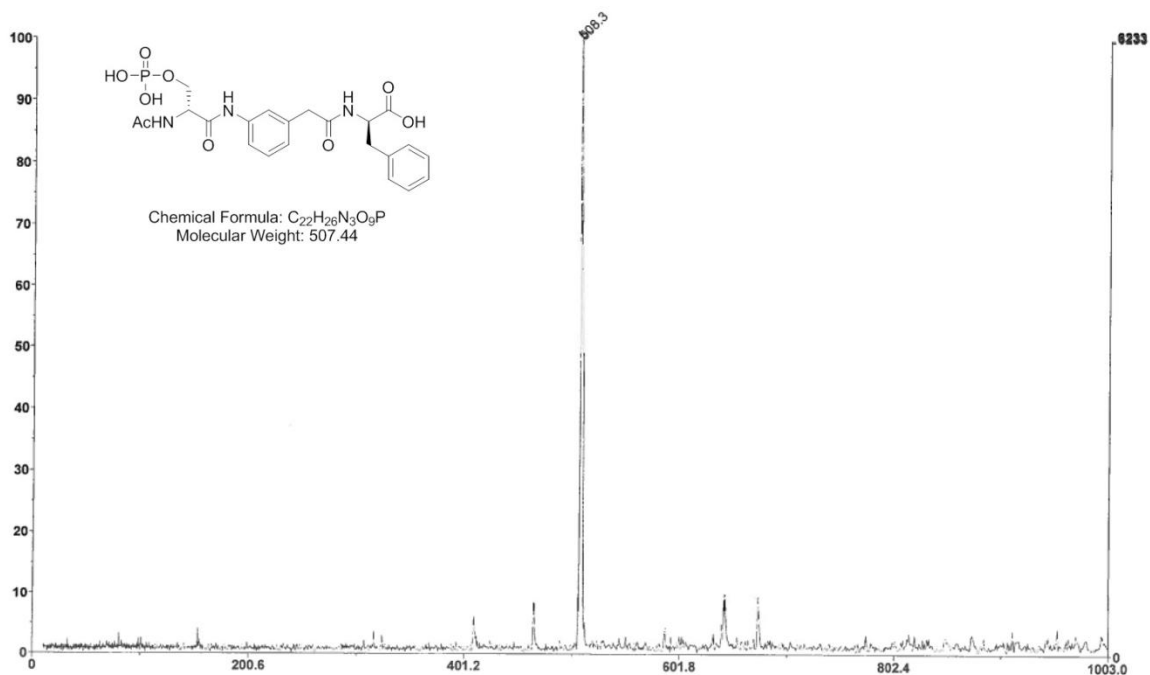
Figure 2.6. Mass spectrum of N1.

Protein expression and purification. The plasmid construct (pAM15, gift from Luc Gadraeu, University De Sherbrook) encoding six his-tagged BRCT domains of BRCA-1 (amino acids 1646-1859) was used to transform BL21(DE3) RIL (Stratagene). Protein expression was induced by 1mM IPTG and the recombinant protein was purified by nickel affinity chromatography (Qiagen). Homogeneity of the purified protein preparation was assessed by SDS-PAGE and concentration estimated by BCA method (Pierce).

Fluorescence polarization assay. The peptide was evaluated in a BRCT assay following previously reported methods [79-81], representative dose-response curves from our previous Fluorescence polarization assay study was shown in Figure 2.7. It was carried out in a 384-well low volume corning plate. The polarization and fluorescence were measured on a Spectramax M5 (molecular devices) plate reader. The peptide was titrated into a mixture of BRCT(BRCA1) (1000 nM) and Fluorescently labeled peptide Flu-βA-pSPTF-CONH$_2$ (100 nM) where βA is beta-alanine. The IC50 value was calculated by using SigmaPlot. Unfortunately, N1 peptide was inactive even at 1000 μM (1 mM) concentration and all we got was a flat line.
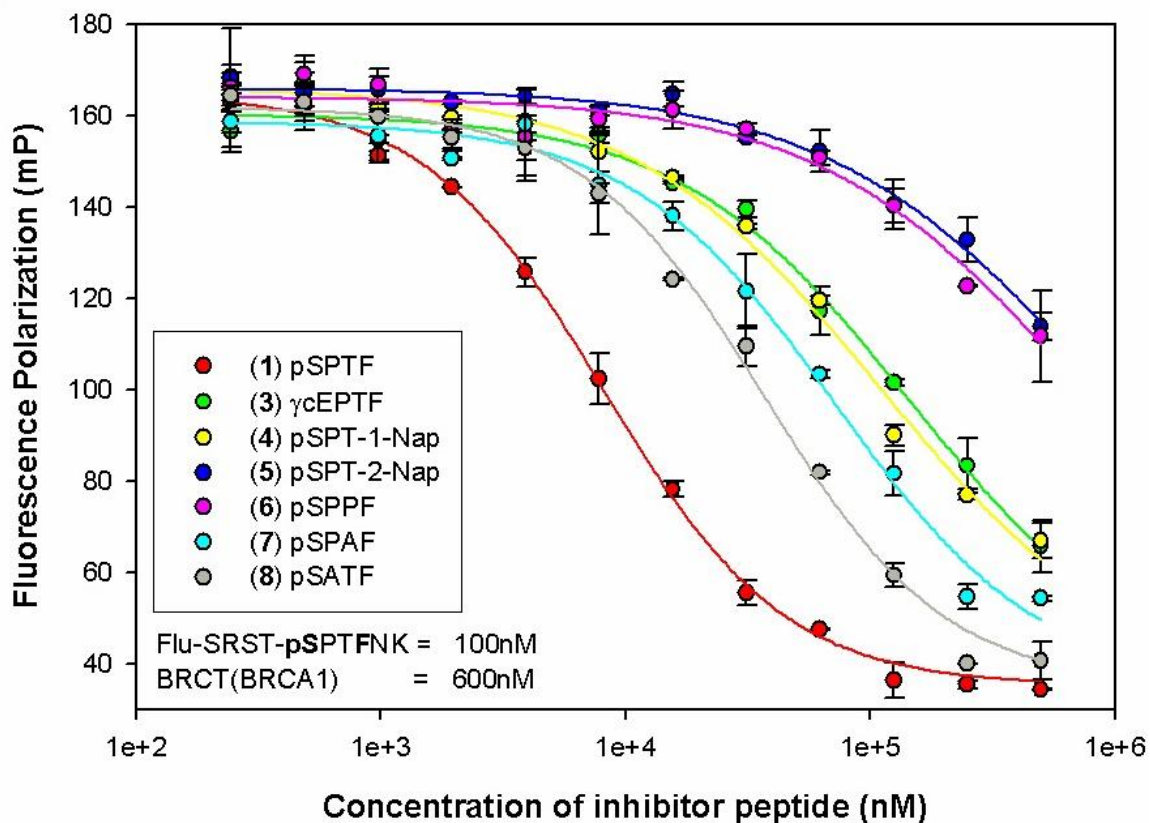
Figure 2.7. Representative dose-response curves from an fluorescence polarization assay study that were used to determine the IC50 values shown in Table 2.1 (1 = P4; 3 = P11; 6 = P13; 7 = P10; 8 = P7).

2.3 Results and Discussions

We first applied MD simulations and post-MD analysis for the peptides (P1-P14, L1-L4) and compound C1 to study the fluctuations in various complexes, followed by more rigorous free energy calculations with the M2 method for short peptides (P1-P14) and compound C1 to illustrate detailed energetic and entropic changes upon ligand binding. The new ligand N1 based on consensus ideas that impose structure constraints, was examined experimentally and computationally. Based on our results, compound D1 was derived from the tight binder C1.

2.3.1 Conformational flexibility of the molecular systems

One unique feature of promiscuous protein systems such as BRCT is to bind to various ligands with significantly different size and shape by using the same binding interface. BRCT needs to provide adequate conformational isomers to recognize these ligands, which involves both side-chain rotation and additional plasticity provided by the backbone. As what shown in crystal structures of BRCT, the relatively rigid alpha helix and beta sheets hold the overall geometry. The variety of side-chains of residues in loops (β3-α2 connection loop, β1'-α1' connection loop and linker between N-terminal and C-terminal) creates a binding surface for ligand recognition except for the reserved binding region for the phosphate group [46]. The backbone nitrogen of G1656 and side-chain of S1655 of the β1 sheet and K1702 of the α2 helix form at least three stable hydrogen bonds with the phosphate group and also orient a ligand in the binding site (Figure 2.1). Notably, the pocket reserved to bind the phosphate group is located between a structurally rigid region constructed by a helix and a sheet. In contrast, the hydrophobic pocket for the P+3 phenylalanine is built by M1775 and N1774 of the β1'-α1' connection loop and F1704 of the α2 helix, with the β1'-α1' connection loop providing a certain flexibility for peptide binding [48].

To study the flexible regions in the binding pocket of BRCT, we measured the root mean square fluctuation (RMSF) of C_α and the standard deviation of phi and psi angles of residues in the BRCT backbone within 7 Å of 18 peptides (P1-P14, L1-L4) and compound C1. The RMSF in Figure 2.8 shows that residues contacting with a ligand

generally have smaller fluctuations and residues without contact with a ligand generally have larger fluctuations, Except for P13, where the middle two proline residues of tetrapeptides do not form optimized contacts with BRCT. Although RMSF plot suggested that residues contacting with a ligand have small fluctuations in the Cartesian space, the standard deviation of phi and psi angles in Figure 2.9(A) shows that the backbone dihedral angle can still rotate considerably. As illustrated in Figure 2.9(B), the most flexible region in the center part of the binding pocket, which directly contacts with the middle two residues of a pSer-X-X-Phe peptide and middle atoms of compound C1. Utilizing the flexible loop region allows for the polar residues E1698 and R1699 of the β3-α2 connection loop to form a hydrogen bond with backbone atoms of the phophopeptides and also accommodate ligands with different shapes. For example, the standard deviations for E1698, R1699 and T1700 were especially large when BRCT bound to P13 and C1, followed by concerted motions of N1742 and G1743 in the linker region. Although P13 still can fit into the binding cavity, the two proline residues limit the arrangement of both molecules to optimize the intermolecular interactions. In contrast, C1 was flexible and adopted multiple bound conformations to strengthen its binding affinity, as discussed in the following sections. For the long peptides, F1772, T1773 from the β1'-α1' connection loop and D1692, A1693 from the β3-α2 connection loop fluctuate to adjust the size of the binding cavity. The size change of binding site agrees with our previous molecular dynamics study, where the size of cavity can be characterized by two angles E1698-A1752-E1836 and S1655-A1752-N1774, which can have difference of 10° upon binding of different peptides (Figure 2.10) [25]. In summary, BRCT uses the power

of loops to alter the shape and size of the binding site to fit various ligands, combined

with a rigid region designed to form stable hydrogen bonds with the phosphate group.
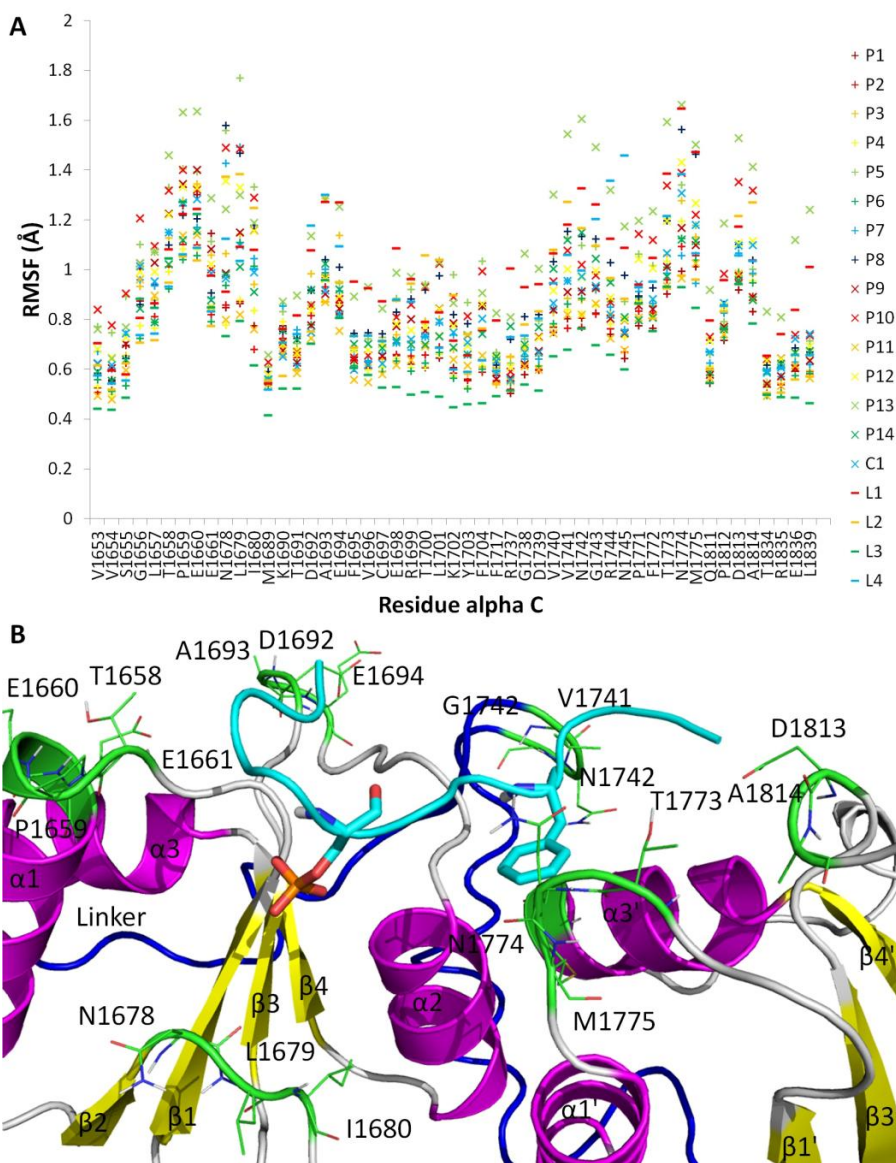
Figure 2.8. Flexibility of active site of BRCT. (A). Root mean square fluctuation (RMSF) of C_α of the residues of the receptor within 7 Å of ligands during MD simulations. (B). Flexible region of the active site. Flexible residues of the protein are shown in a green line representation. Ligand is shown as a blue tube with pSer and Phe (P+3) residues in licorice representation.
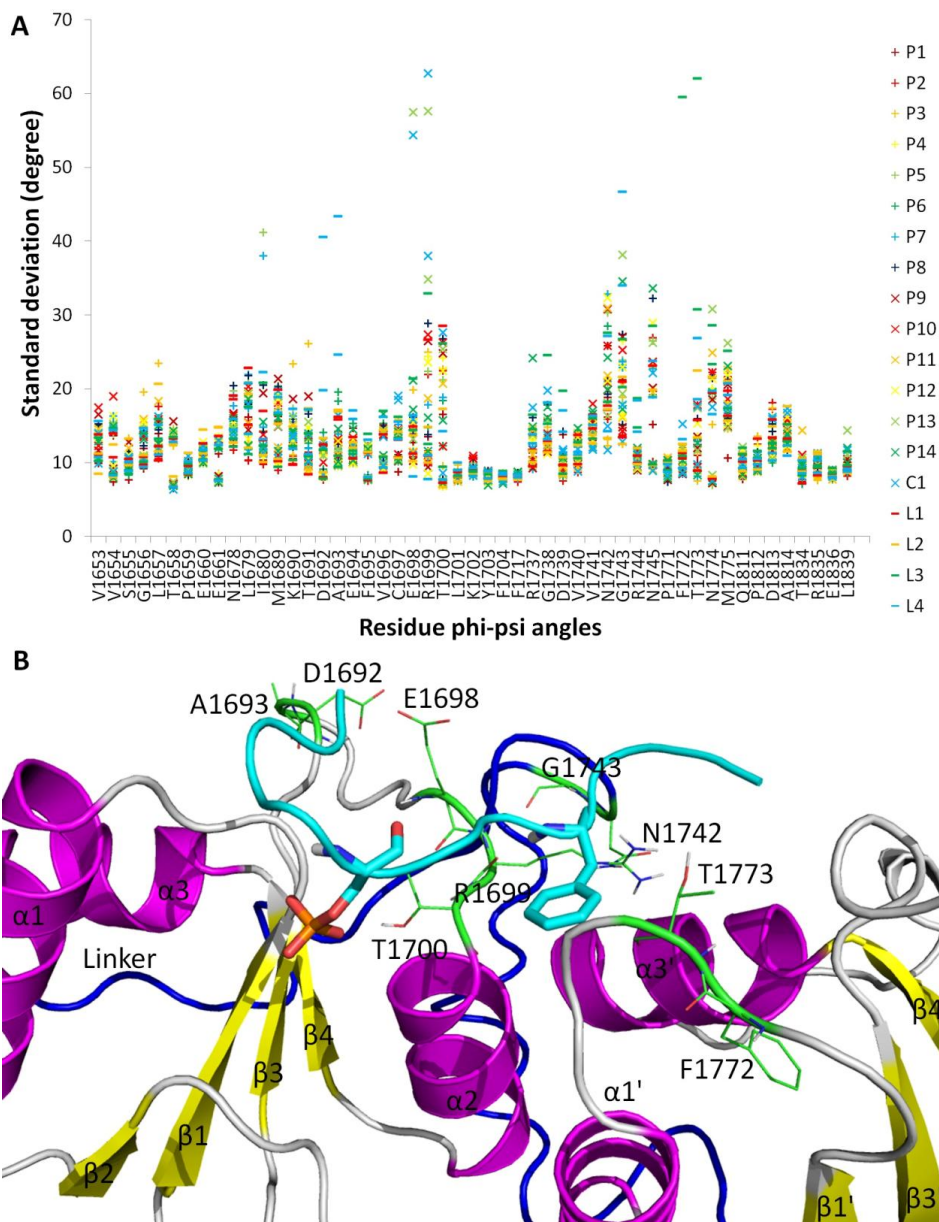
Figure 2.9. Flexibility of active site of BRCT. (A). Standard deviation of phi and psi angles of the residues of the receptor within 7 Å of ligands during MD simulations. Each residue has one column containing two standard deviation values for the phi angle and psi angle, respectively. (B). Flexible region of the active site. Flexible residues of the protein are shown in a green line representation. Ligand is shown as a blue tube with pSer and Phe (P+3) residues in licorice representation.
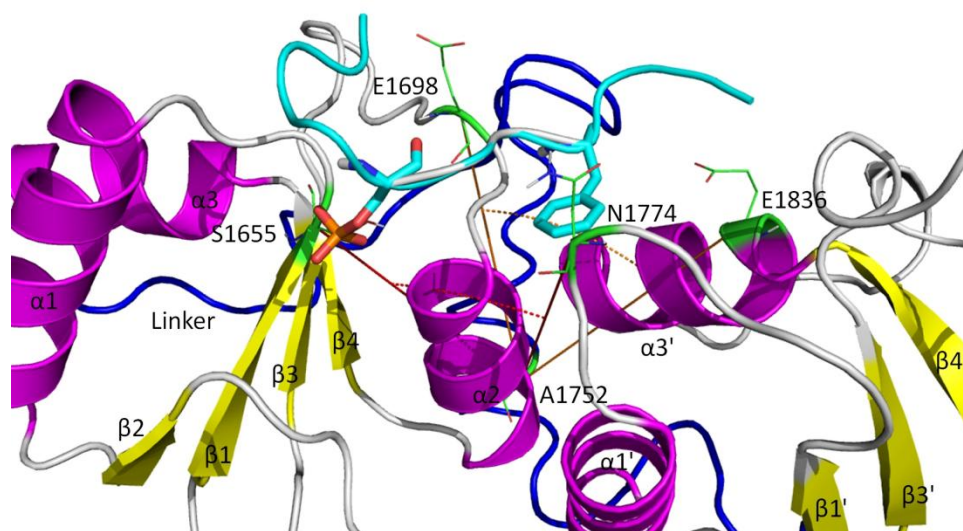
Figure 2.10. Angles E1698-A1752-E1836 and S1655-A1752-N1774 as indications of size change of BRCT binding site.

2.3.2 Ligand binding modes and intermolecular interactions computed by MM/PBSA calculations

Because the BRCT domain has a highly adaptable binding pocket, we hypothesized that some ligands may feature diverse binding modes. We therefore examined the ligand binding modes and the rotamer of each rotatable bond for every ligand to discover their differences between the free and bound states. For all peptides P1-P14, only one major bound conformation was observed: pSer forms hydrogen bonds with S1655, G1656 and K1702 and the P+3 Phe locates in the hydrophobic packet (Figure 2.1). Interestingly, compound C1 can establish multiple bound conformations in the binding site by fitting either a benzene ring into the hydrophobic pocket and an indole ring into a cluster of residues G1656, L1657, D1658 of the β1-α1 connection loop and K1690 of the β3-α2 connection loop, and vice versa (Figure 2.11(C) and 5(B)). C1 can also bind to BRCT

65

with its folded form, whereby two rings form a T-shape stacking interaction (Figure 2.11(A)).
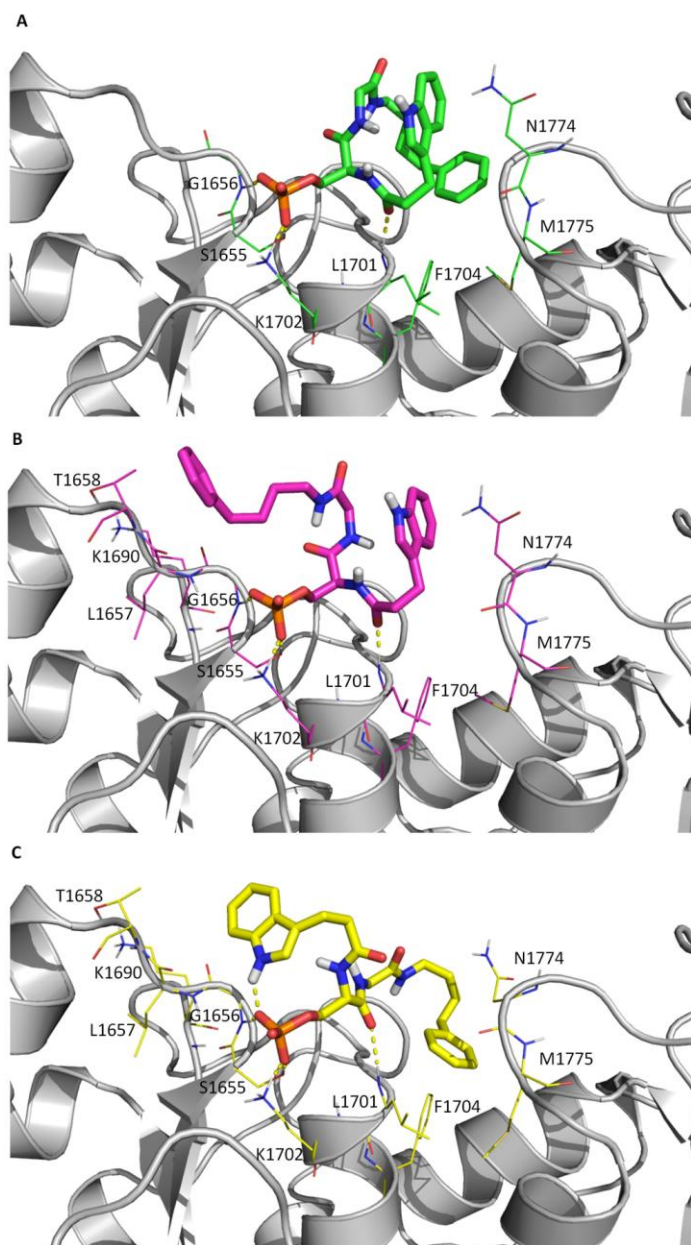
Figure 2.11. (A, B, C) Three distinct bound conformations of C1 from M2 calculation. Residues of BRCT are shown in line representation and ligand is shown in licorice representation, hydrogen bonds are drawn in dash lines (free energies of A, B and C bound conformations are -1461.16, -1457.04 and -1453.72 kcal/mol, respectively).

Figure 2.12 illustrates the rotameric states of selected rotatable bonds of P4 and C1 in their free and bound states. All peptides show the same trend as in the histogram plots of P4, with most rotatable bonds becoming more rigid and losing rotameric states in their bound state. However, compound C1 does not lose rotamers in the bound state, and a few dihedrals are even more flexible in the bound form. BRCT does not reduce the number of rotamers after binding to C1 either, which differs from the bound states with other peptides (Figure 2.13). MM/PBSA calculations suggested that the intermolecular interactions between all the peptides/ligands and BRCT are about the same, which agrees with experiments finding that $\Delta\Delta G_{exp}$ is within 3 kcal/mol (Table 2.3).
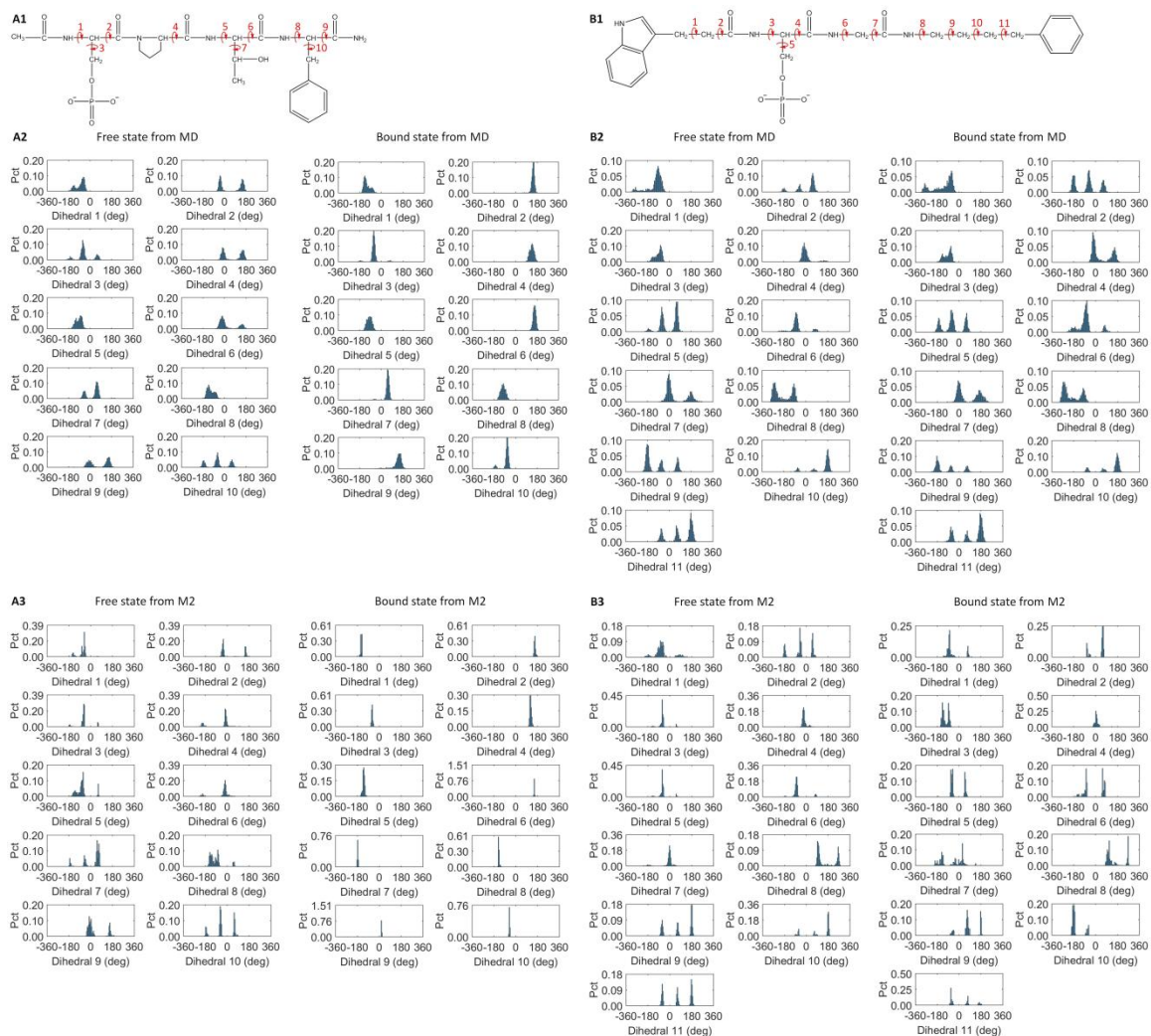
Figure 2.12. The rotameric states of selected rotatable bonds of P4 and C1 in both free and bound states. (A1), (B1). Selected rotatable bonds of ligand P4 and C1 structures, respectively. (A2), (B2). The dihedral angle distribution from 1000 frames collected during 100-ns MD simulations of P4 and C1, respectively. (A3), (B3). The dihedral angle distribution for distinct energy minima found by M2 calculations of P4 and C1, respectively.

(A1) The first side-chain dihedral angles of part of live set residues of C1-BRCT complex from MD
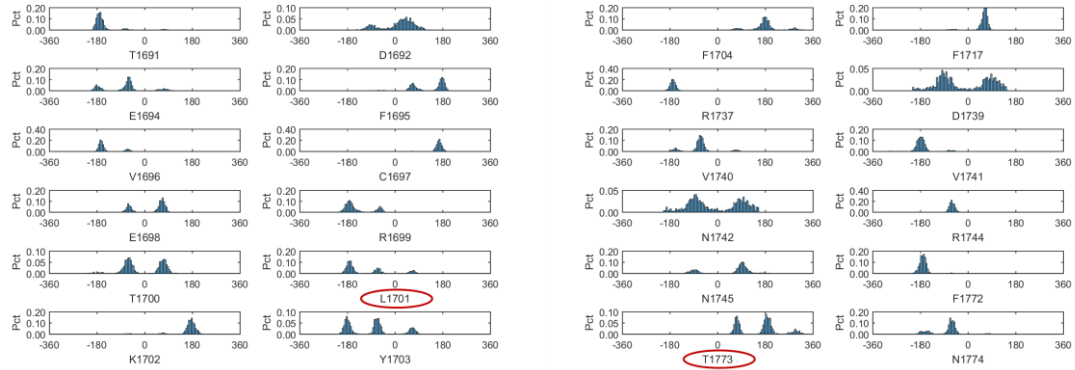
(A2) The first side-chain dihedral angles of part of live set residues of P4-BRCT complex from MD
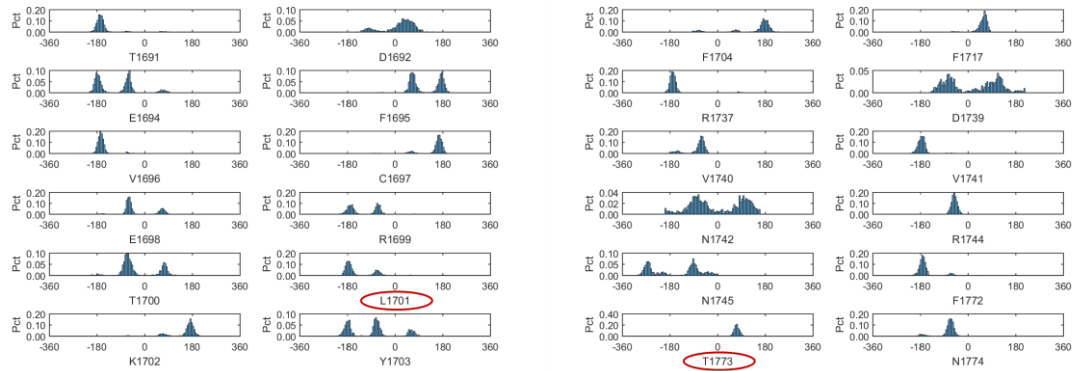
(B1) The first side-chain dihedral angles of part of live set residues of C1-BRCT complex from M2
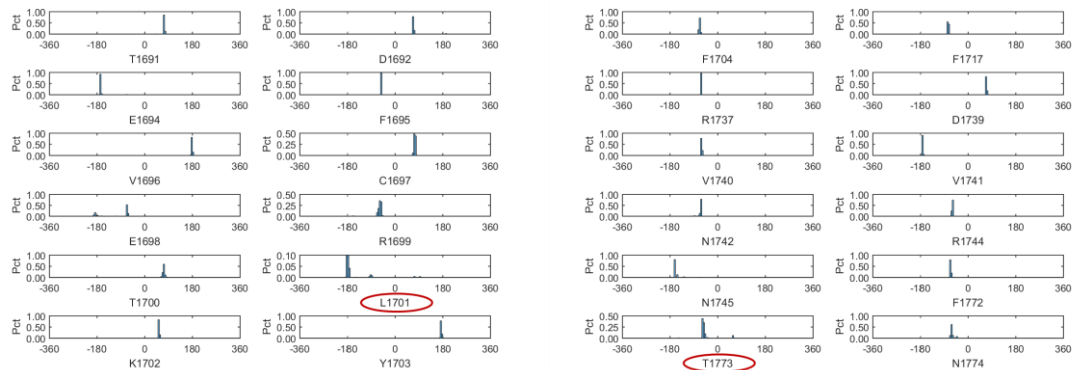
(B2) The first side-chain dihedral angles of part of live set residues of P4-BRCT complex from M2
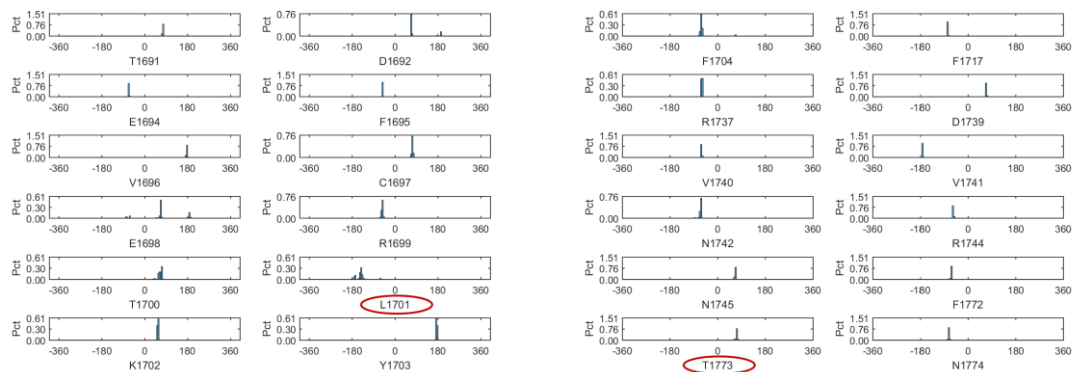
Figure 2.13. Comparison of the first side-chain dihedral angles of part of live set residues of C1 and P4-BRCT complex from MD and M2, respectively. (A1), (A2). The first side-chain dihedral angles of part of live set residues of C1 and P4-BRCT complexes from MD, respectively. (B1), (B2). The first side-chain dihedral angles of part of live set residues of C1 and P4-BRCT complexes from M2, respectively. The difference is highlighted by red circle.

Table 2.3. BRCT domain−ligand Interaction Energy (kcal/mol) of P1-14 and C1, N1 and D1 calculated by molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA). The binding interaction energy was computed by $\Delta E_{cal}=E_{complex}-E_{free\ protein}-E_{free\ ligand}$. Decomposed interaction energy, E_{cal}, from our calculations includes Lennard-Jones energy $\langle U_{VDW}\rangle$, nonpolar solvation free energy $\langle W_{NP}\rangle$, Coulombic energy $\langle U_{Coul}\rangle$, and PB solvation free energy $\langle W_{PB}\rangle$. $\langle E_{np}\rangle$ represents the sum of $\langle U_{VDW}\rangle$ and $\langle W_{NP}\rangle$; $\langle E_{ploar}\rangle$ represents the sum of $\langle U_{Coul}\rangle$ and $\langle W_{PB}\rangle$, bonded terms $\langle U_{val}\rangle$ are zero due to energy cancelling out and therefore not listed here.

No.	$\Delta\Delta G_{exp}$	$\Delta(U+W)$	ΔU_{VDW}	ΔW_{NP}	ΔE_{NP}	ΔU_{Coul}	ΔW_{PB}	ΔE_{polar}
P1	0.00	-3.19±0.72[a]	-32.7±0.5	20.8±0.1	-11.9±0.5	-84.6±3.7	93.3±3.67	8.73±0.86
P2	0.28	-2.24±0.66	-31.9±0.5	20.6±0.1	-11.3±0.5	-76.4±3.3	85.4±3.28	9.06±0.56
P3	0.69	-3.48±0.34	-31.0±0.3	20.6±0.1	-10.4±0.4	-160±2	167±2	6.91±0.42
P4	0.91	-4.68±0.50	-32.8±0.2	21.3±0.1	-11.5±0.2	-160±2	167±1	6.79±0.42
P5	1.17	-2.45±0.58	-35.0±0.8	21.5±0.3	-13.4±0.6	-151±3	162±2	11.0±0.5
P6	1.61	-1.53±0.56	-31.3±0.4	21.4±0.2	-9.83±0.25	-163±2	171±2	8.30±0.75
P7	1.61	-2.36±0.72	-28.8±0.8	19.6±0.2	-9.17±0.68	-152±3	159±2	6.81±0.58
P8	1.74	-4.86±0.56	-33.4±0.5	19.9±0.3	-13.5±0.5	-139±2	148±2	8.64±0.38
P9	2.03	-0.53±0.66	-31.5±0.7	20.4±0.1	-11.1±0.6	-150±4	161±3	10.6±1.0
P10	2.12	-1.06±1.35	-30.5±0.5	20.0±0.2	-10.4±0.4	-142±12	151±11	9.38±1.37
P11	2.36	-0.50±0.85	-32.3±0.4	21.6±0.1	-10.6±0.3	-122±3	132±3	10.1±0.7
P12	2.74	-3.15±0.46	-30.9±0.4	19.8±0.2	-11.0±0.3	-152±2	160±2	7.89±0.48
P13	3.29	-3.05±0.41	-29.3±0.2	18.5±0.1	-10.8±0.2	-123±2	130±2	7.74±0.50
P14	3.29	0.39±0.99	-29.4±0.6	20.8±0.2	-8.61±0.43	-149±2	158±2	9.00±1.14
C1	-0.70	-1.59±0.66	-18.9±1.7	15.3±0.7	-3.55±0.99	-134±3	136±4	1.96±1.02
N1	3.29	-3.05±0.70	-29.8±0.5	18.6±0.2	-11.2±0.5	-59.4±5.4	67.6±4.9	8.12±0.68
D1	N/A	0.21±1.82	-33.0±1.8	21.0±0.7	-12.0±1.2	-121±9	133±7	12.2±2.7

[a] The statistical error was estimated on the basis of the deviation between block averages [82].

To understand why or why not a ligand loses the rotamers after binding, we clustered conformations of the free peptides and ligands and compared them with those in the bound complexes. For the peptides and C1, they generally have two distinct conformations in the free state, folded and extended, which except for P13 (Ac-pSPPF-NH$_2$), can switch back and forth in MD simulations for most free ligands (Figure 2.14). However, the bound peptides are locked to only the extended form, which results in reduced rotamers in side-chains and also backbone φ and ψ angles (Figure 2.12). To test

the robustness of MD simulation on rotameric states analysis, we ran and analyzed another MD run with different initial conformations for several ligands. The simulated rotameric states are nearly identical to the other MD, showing that multiple rotameric states in free states reduce to single rotameric state in bound state (Figure 2.15). For C1, both folded and extended forms are observable in the bound states; free energy calculations with M2 further revealed that all these distinct ligand conformations are stable energy minima (Figure 2.11).
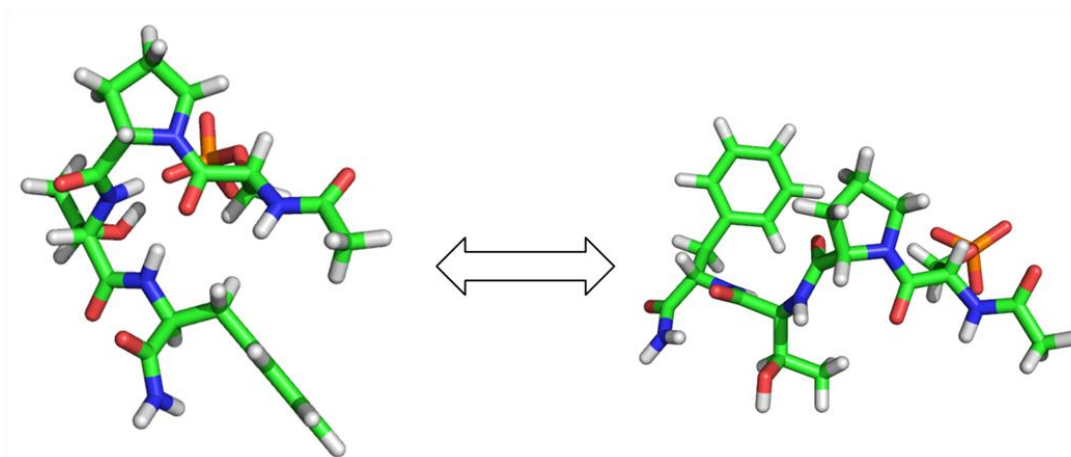
Figure 2.14. Conformational change of P4 between bent and stretched in free ligand state.
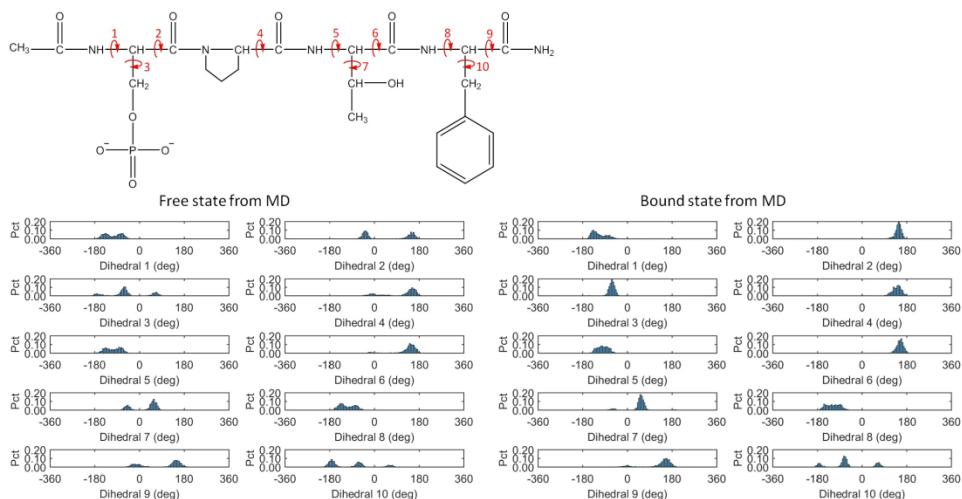
Figure 2.15. Representative robustness test of MD simulations on rotameric states analysis with P4.

2.3.3 Binding free energies with M2 method

To gain insights into the mechanism of binding, we needed thorough sampling and accurate ligand binding free energy calculations that included both enthalpic and configurational entropic contributions for molecular recognition. Although MM/PBSA calculations provide valuable information for intermolecular interactions, our calculations based on 100-ns MD simulations may have missed some important conformations, and contributions from changing configurational entropy and molecular conformations are neglected in Table 2.3. In addition, because of different non-polar solvation models and use of a real set in M2 for energy calculations (Figure 2.5), the values of non-polar and polar interaction energies, ΔE_{NP} and ΔE_{Polar}, from MM/PBSA and M2 cannot be compared directly. We therefore computed ligand-binding free energy with the M2 method, which involved an aggressive conformational search engine to locate local

energy minima and a rigorous modified harmonic approximation approach to compute free energy for each minimum found.

Table 2.4 and Figure 2.16 show that the computed related binding free energy, $\Delta\Delta G_{calc}$, was in good agreement with experimental values, which validated the method as well. Because M2 uses accumulated energy which is different from dynamics-based method, it does not have fluctuated energy. Equation 2 shows that when a low energy minimum is found by M2 conformational search and added to the accumulated energy, the computed free energy G^o drops. Search and computation continue until the accumulated free energy is converged. Here we calculated error interval for y-intercept of linear regression line in Figure 2.16 [83-85]. Part of the variance comes from experimental noise, which is typically about 0.3-0.5 kcal/mol for accurate binding free energy measurements [86]. If the binding free energies of two ligands are measured independently in experiments, then experimental relative binding free energies between the two ligands would have error around 0.4-0.7 kcal/mol. Therefore, the errors for free energy calculation method versus experimental data can only be larger than experimental noise of 0.4-0.7 kcal/mol, indicated by the range of y-intercept of linear regression line (~3 kcal/mol), and experimental noise is expected to be a at least 13% of the total observed error. With agreement of early studies on ligand–protein binding, the strong Coulombic attraction is largely compensated by the solvation free energy, and the vdW attraction is the major driving force for ligand binding [32, 64]. Moreover, peptides with large non-polar residues at the P+2 position, such as P2, P3, P5, P8, P10 and P12, generally have stronger vdW interaction (Table 2.4). Although M2 revealed more bound conformations for the

complex from various combinations of side-chain rotations, the major binding mode of

BRCT-pSXXF is the same as that obtained by MD sampling, whereby the phosphate

group forms hydrogen bonds with S1655, G1656 and K1702, and P+3 Phe or Tyr locates

in the hydrophobic pocket (Figure 2.17). M2 also revealed more conformations for free

ligands, including the folded and extended forms, and their computed conformational free

energies are similar. Therefore, the folded and extended conformations may have similar

population in the free ligands.

Table 2.4. Binding free energy, average binding potential energy, and solvation free energy (kcal/mol) of P1-14, C1, N1 and D1 calculated by M2. The binding free energy was computed by $\Delta G_{cal}=G_{complex}-G_{free\ protein}-G_{free\ ligand}$. Each decomposed energy is obtained by $\Delta E_{cal}=E_{complex}-E_{free\ protein}-E_{free\ ligand}$. Decomposed free energy, G_{cal}, from our calculations includes the average potential energy $<U+W>$, configurational entropy $-TS$, bonded terms $<U_{val}>$, Lennard-Jones energy $<U_{VDW}>$, nonpolar solvation free energy $<W_{NP}>$, Coulombic energy $<U_{Coul}>$, and PB solvation free energy $<W_{PB}>$. $<E_{np}>$ represents the sum of $<U_{VDW}>$ and $<W_{NP}>$; $<E_{ploar}>$ represents the sum of $<U_{Coul}>$ and $<W_{PB}>$.

No.	$\Delta\Delta G_{exp}$	ΔG_{cal}	$\Delta\Delta G_{cal}$	$\Delta(U+W)$	$-T\Delta S$	ΔU_{Val}	ΔU_{VDW}	ΔW_{NP}	ΔE_{NP}	ΔU_{Coul}	ΔW_{PB}	ΔE_{polar}
P1	0.00	-10.5	0.00	-42.4	31.9	5.08	-33.2	-4.40	-37.6	-291	281	-9.90
P2	0.28	-11.9	-1.33	-42.3	30.4	-0.80	-37.6	-4.35	-42.0	-274	275	0.49
P3	0.69	-9.76	0.76	-39.1	29.3	0.98	-39.7	-4.28	-44.0	-245	249	3.93
P4	0.91	-10.5	0.02	-37.9	27.4	0.20	-34.9	-4.03	-38.9	-242	243	0.84
P5	1.17	-8.94	1.58	-37.8	28.9	-1.28	-38.6	-4.22	-42.8	-238	244	6.32
P6	1.61	-8.97	1.55	-38.7	29.7	-1.10	-31.8	-4.02	-35.8	-243	241	-1.82
P7	1.61	-10.8	-0.29	-37.6	26.8	-3.61	-29.7	-3.80	-33.5	-249	249	-0.53
P8	1.74	-10.4	0.09	-37.5	27.1	0.48	-37.4	-4.11	-41.5	-239	242	3.61
P9	2.03	-10.2	0.34	-37.7	27.5	0.80	-38.3	-3.98	-42.2	-234	238	3.73
P10	2.12	-9.91	0.61	-39.5	29.6	-2.30	-37.5	-4.08	-41.6	-227	231	4.44
P11	2.36	-7.95	2.57	-35.7	27.8	-1.62	-29.0	-4.03	-33.0	-248	247	-1.05
P12	2.74	-7.26	3.26	-35.3	28.0	-2.09	-36.3	-3.69	-40.0	-217	224	6.84
P13	3.29	-5.82	4.70	-34.6	28.8	-2.62	-35.7	-4.22	-40.0	-231	239	8.01
P14	3.29	-6.02	4.50	-36.3	30.3	-5.33	-38.6	-4.21	-42.8	-227	239	11.8
C1	-0.70	-12.4	-1.84	-38.0	25.6	-4.17	-30.9	-3.91	-34.9	-212	213	1.04
N1	3.29	-7.60	2.90	-37.7	30.1	-0.74	-33.4	-4.41	-37.9	-270	271	0.886
D1	N/A	-14.4	-3.84	-40.9	26.5	-2.70	-30.6	-3.92	-34.5	-219	215	-3.69
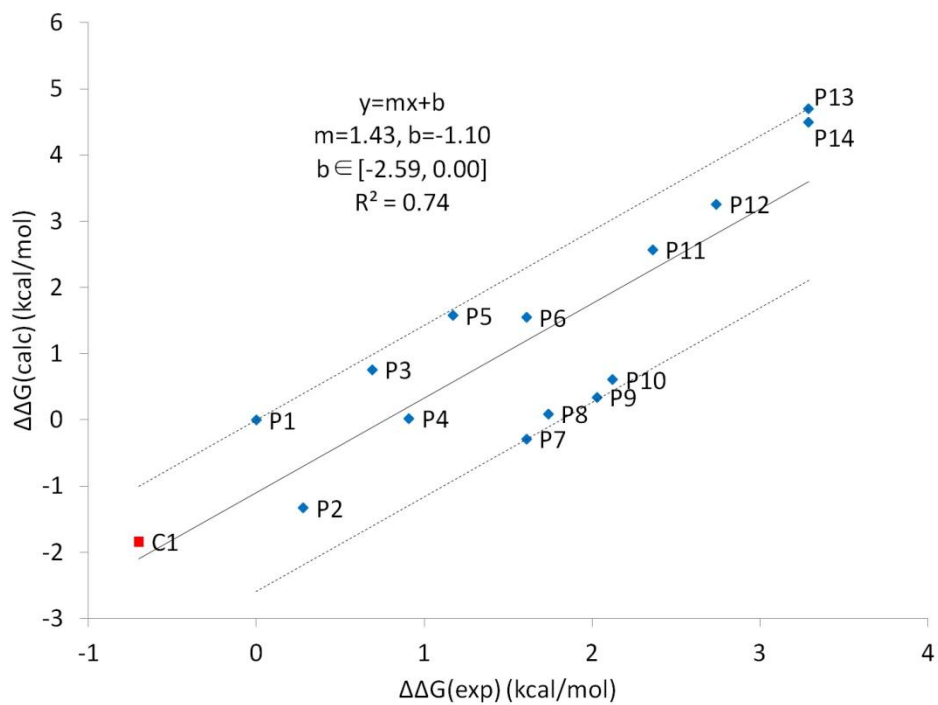
Figure 2.16. Calculated versus experimental relative binding free energies ΔΔG (kcal/mol) for P1-P14 and C1.
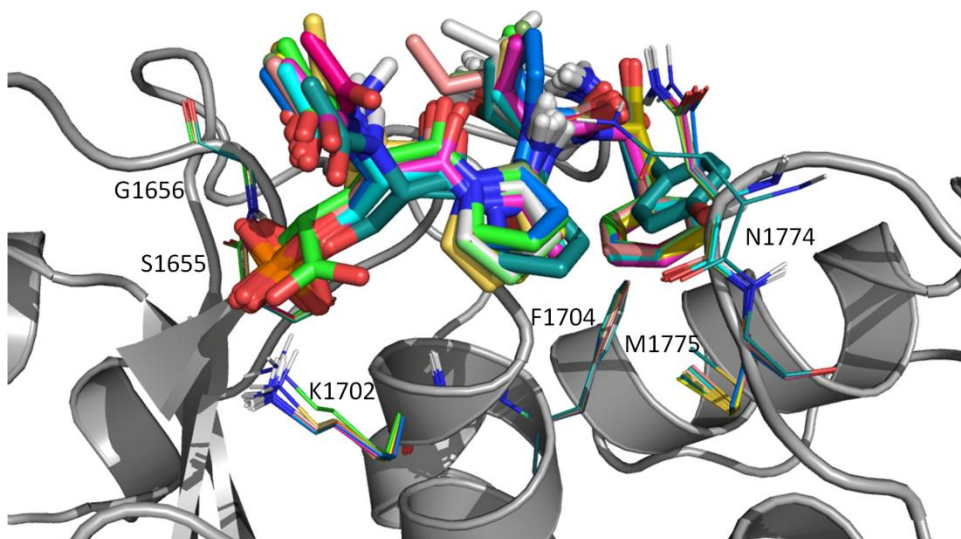
Figure 2.17. Superimposed most stable bound conformaitons of peptide (P1 to P14). The phosphate groups and phosphate mimic are anchored by S1655, G1656, K1702 and phenylalanine/tyrosine surrounded by F1704, N1774, M1775. Ligands are shown in licorice representation, residues of BRCT are shown in line representation.

It is not surprising to observe the enthalpy $\Delta<U+W>$ and configuration entropy $-T\Delta<S>$ compensation for tight binders; however, the outlier C1 is particularly of interest (Figure 2.18). Although C1 forms a moderate enthalpy attraction with BRCT, ~ -38 kcal/mol, which is similar to that for peptides P4-P9, with the remarkable ~2-4 kcal/mol small configuration entropy loss, C1 outperforms other peptides (Table 2.4). Compared with peptides, some rotamers of BRCT and C1 can gain new rotameric states rather than losing them, and the vibrational entropy loss is smaller than that for P4-P9, as seen from the change in width of M2 histogram peaks that correspond to the width of energy wells (Figure 2.12). Upon ligand binding, M2 histogram peaks for P1-P14 become narrower, whereas C1 has the same or even wider peaks. In Table 2.5, we list the number of complex, ligand and protein conformations from M2 calculations. For example, M2

calculations generated 482 distinct conformations of free P1 within 10 RT of the most stable free conformation. Even if free P1 were equally stable in all 482 energy wells with only one bound conformation, the maximum change in conformational entropy would only be reduced by RTln 482 = ~3.7 kcal/mol, which is significantly smaller than the -TΔS values in Table 2.4. We may approximate vibrational entropy through -TΔS_{vib}=-TΔS_{config} + TΔS_{conf}. Table 2.6 shows that C1 has much smaller vibrational entropy loss than peptides P1-P14. In sum, both conformational entropy and vibrational entropy are attributed to the smaller configuration entropy loss of C1. Interestingly, P7, with a small residue alanine in the P+2 position, has the second smallest entropic penalty in M2 results but not P13, which has two proline residues in the middle of the peptide. P13 managed to partially eliminate the folded conformations because of the geometric constraint proline residues; however, the entropy cost does not decrease substantially due to the big vibrational entropy loss (Table 2.6). Moreover, the restraint by the two prolines resulted in the incorrect orientation of ligand-bound conformations, which significantly weakens the polar attractions (Figure 2.19).
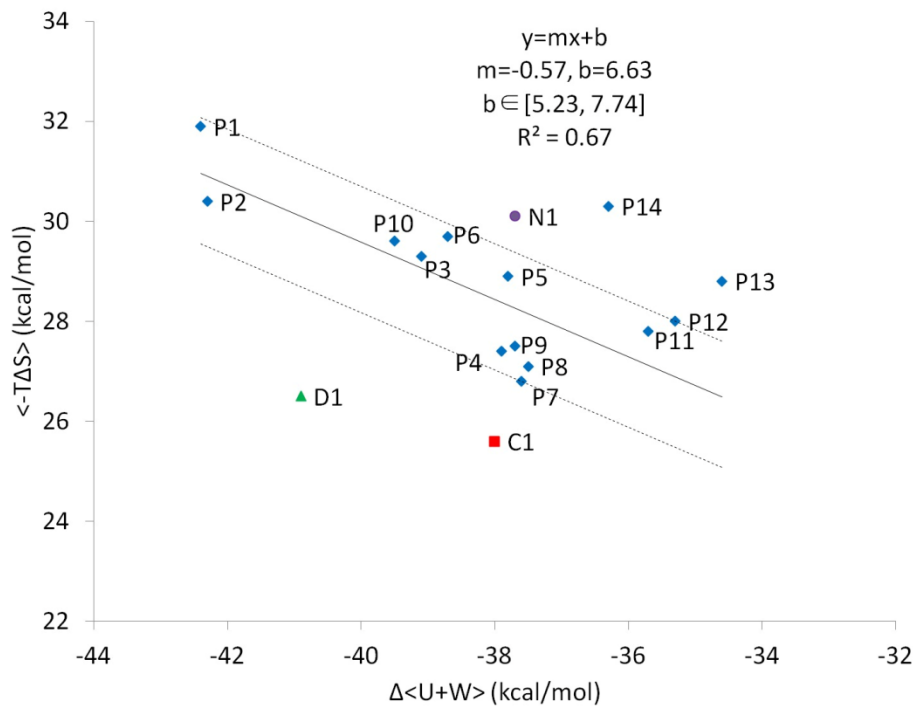
Figure 2.18. Computed configurational entropy contribution, $<-T\Delta S>$ and energy contribution, $\Delta<U+W>$, for P1-14 and C1, N1 and D1. $<-T\Delta S>$ vs $\Delta<U+W>$ is plotted using tight peptide binders P1-P12.

Table 2.5. Numbers of complex, free ligand and protein conformations from M2 calculation. M2 uses a rigorous conformational search through dihedral distortion for new conformations. Molecular torsional modes are calculated via diagonalization of matrix of energy 2nd-derivatives transformed into internal coordinates with all bond and angle rows and columns removed. After a complete distortion along these modes, the whole system is energy minimized via a quasi-Newton geometry optimization to get new conformations. 1RT means numbers of conformations within 1RT above the global energy minimum.

No.	Numbers of complex conformations in M2					Numbers of free ligand conformations in M2				
	Total	1RT	3RT	5RT	10RT	Total	1RT	3RT	5RT	10RT
P1	864	3	28	54	181	824	3	61	171	482
P2	1086	9	31	64	216	373	3	20	64	200
P3	581	3	25	58	165	589	26	99	163	370
P4	661	4	35	65	171	507	6	37	99	323
P5	755	3	20	36	133	990	29	142	279	663
P6	727	4	25	55	153	388	2	37	96	286
P7	581	2	24	58	165	481	12	54	126	355
P8	538	2	21	41	87	957	10	70	199	627
P9	535	4	29	50	140	470	8	46	126	316
P10	501	5	23	54	148	563	1	49	130	338
P11	764	3	17	51	176	747	6	62	171	488
P12	594	3	16	53	216	408	3	30	98	263
P13	362	1	12	28	125	448	28	92	148	284
P14	363	3	19	34	145	328	9	47	94	254
C1	1992	5	41	80	301	1119	5	41	148	673
N1	367	4	6	17	60	393	9	49	91	233
D1	1472	2	18	51	229	660	2	23	90	386
protein	802	8	47	128	409					

Table 2.6. Approximated conformational and vibrational entropy (kcal/mol) for P1-P14, C1, N1 and D1. The conformational entropy penalty is approximated through RTln M (M is the number of conformations within 10RT of most stable free ligand conformation from Table 2.5). The vibrational entropy penalty was computed by $-T\Delta S_{vib}=-T\Delta S_{config} + T\Delta S_{conf}$.

No.	$-T\Delta S_{config}$	$-T\Delta S_{conf}$	$-T\Delta S_{vib}$
P1	30.1	3.68	26.5
P2	28.6	3.16	25.5
P3	27.5	3.53	24.0
P4	25.6	3.44	22.2
P5	27.0	3.87	23.2
P6	28.0	3.37	24.6
P7	25.0	3.50	21.5
P8	25.2	3.84	21.4
P9	25.7	3.43	22.3
P10	27.8	3.47	24.3
P11	25.9	3.69	22.2
P12	26.2	3.32	22.9
P13	27.0	3.37	23.6
P14	28.5	3.30	25.2
C1	23.8	3.88[a]	20.0
N1	28.3	3.25	25.1
D1	24.8	3.55[a]	21.2

[a] For C1 and D1, they have at least three distinct bound conformations (Fig 5 and S13), so the conformational entropy penalty of C1 and D1 is approximated through RTln (M/3).

81
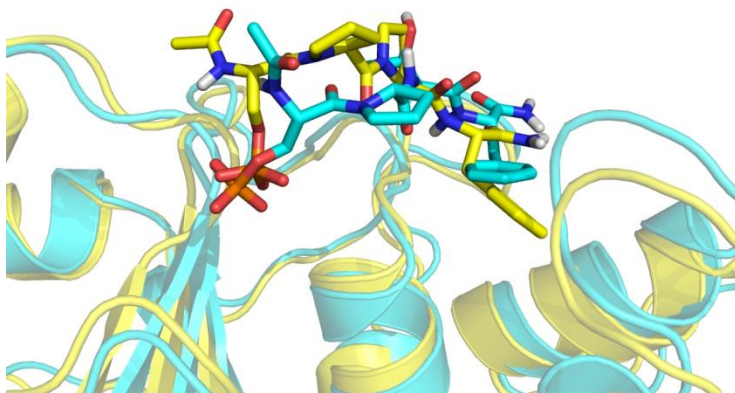
Figure 2.19. Superimposed average bound conformation of P4 (blue) and P13 (yellow) during MD simulations. The bound conformation of P4 is represents the standard bound conformation of most phosphopeptides. Changes in the bound conformation start showing up right after the mutation at the P+2 position from threonine or valine to proline. In P13, in order to align phosphate group and benzene ring of phenylalanine, the whole backbone frame of the ligand has to move towards solvent to moderate the restrain from two rigid proline residues in the middle, which causes the improper fit of P13 in the cavity, resulting in increased enthalpy change and high entropy cost.

2.3.4 Inhibitor design: New strategy for promiscuous modular domains?

Two strategies are commonly used in ligand design for enhancing binding affinities: increasing intermolecular attractions and decreasing entropy loss upon binding. For example, new interactions between ligands and receptors, such as adding hydrogen bonds, can be introduced to increase enthalpic attractions [87-93]. The other way is via reducing the entropy cost by pre-rigidifying the ligand to its bound conformation [94, 95]. This pre-organization of the ligand to its bound conformation lessens the decrease in number of rotameric states, and thus affinity is increased primarily because of optimizing the entropic term.

Because the number of potential hydrogen bonds may already be maximized by the presence of the phosphate group, we used the latter strategy to pre-organize a ligand by introducing a benzene ring in the ligand backbone to limit its conformational flexibility. Having a benzene ring in the middle at a certain level prevents the ligand from bending and forming intra-molecular hydrogen bonds like other tetrapeptides do. A new ligand, N1, was synthesized (Figure 2.2) and its binding to BRCT was tested experimentally. Although the conformations were constrained to some degree to reduce conformational entropy penalty, the loss from the vibrational part was not reduced enough. The conformational constraints by the benzene ring restricted the ligand rearrangement to optimize the polar and non-polar contacts to the protein, thereby resulting in weak binding (Table 2.4 and Figure 2.20). N1 performed similar to P13, so over-rigidifying a ligand is not advantageous, which suggests the challenge in retaining optimized intermolecular interactions in pre-rigidifying a peptidomimic compound. Previous work in design of potent Cbl(TKB)-binding peptides drew the same conclusion [96]. Therefore, because of conformational flexibility at the binding interface of a modular domain, flexible ligands may be favorable.
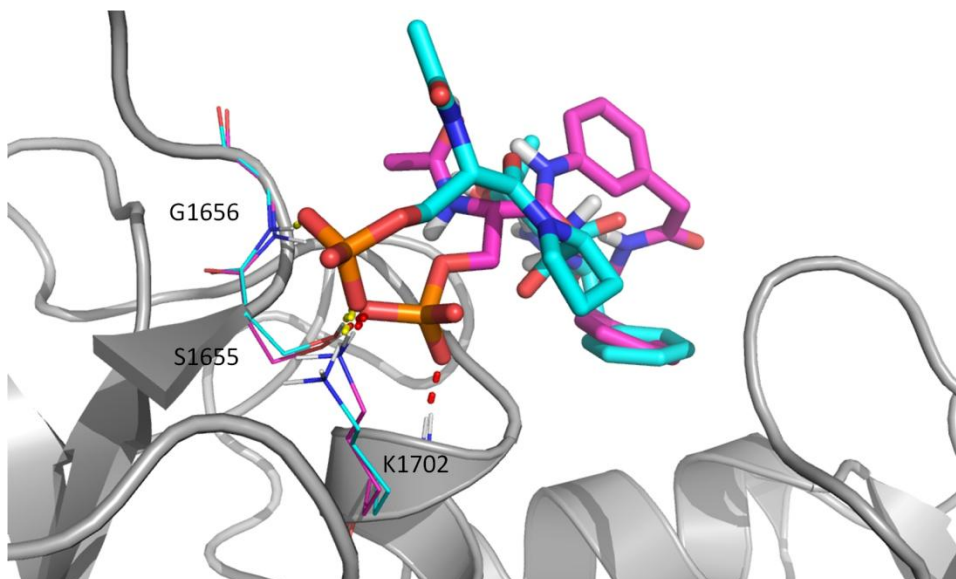
Figure 2.20. Superimposed most stable bound conformations of of P4 (blue) and N1 (pink) from M2 calculation. Ligands are shown in licorice representation, residues of BRCT are shown in line representation. Hydrogen bonds are drawn in dash lines.

Another strategy to lower entropy penalty, although less common, is by introducing a less rigid complex while the molecules bind. Because the strategies to further modify the short peptides to increase their bound conformations may be exhausted, compounds with phosphate groups are a better alternative. On the basis of our calculations and the structure of compound C1, we further modified it to enhance intermolecular attractions by the formation of additional hydrogen bonds between the ligand and BRCT. In the meantime, we kept the template structure intact to maintain its flexibility. We added one hydroxyl group to the para site of the benzene ring of C1, which can form hydrogen bonds with K1690, D1773 or N1774 with different bound conformations (Figure 2.21). Therefore, designed compound D1 shows improved binding affinity, by 2 kcal/mol, with more negative $\Delta(U+W)$ as compared with C1 (Figure 2.18 and Table 2.4). As compared with C1, D1 has a stronger Coulombic interaction with BRCT because of the additional

hydogen bonds (Table 2.4). Moreover, because D1 can also adopt mutiple distinct bound conformations, the entropy cost is minimal, as is found in C1. The enthalpy-entropy compensation plot shown in Figure 2.18 clearly indicates that D1 outperforms other peptides by both increasing intermolecular attraction and reducing entropic penalty.
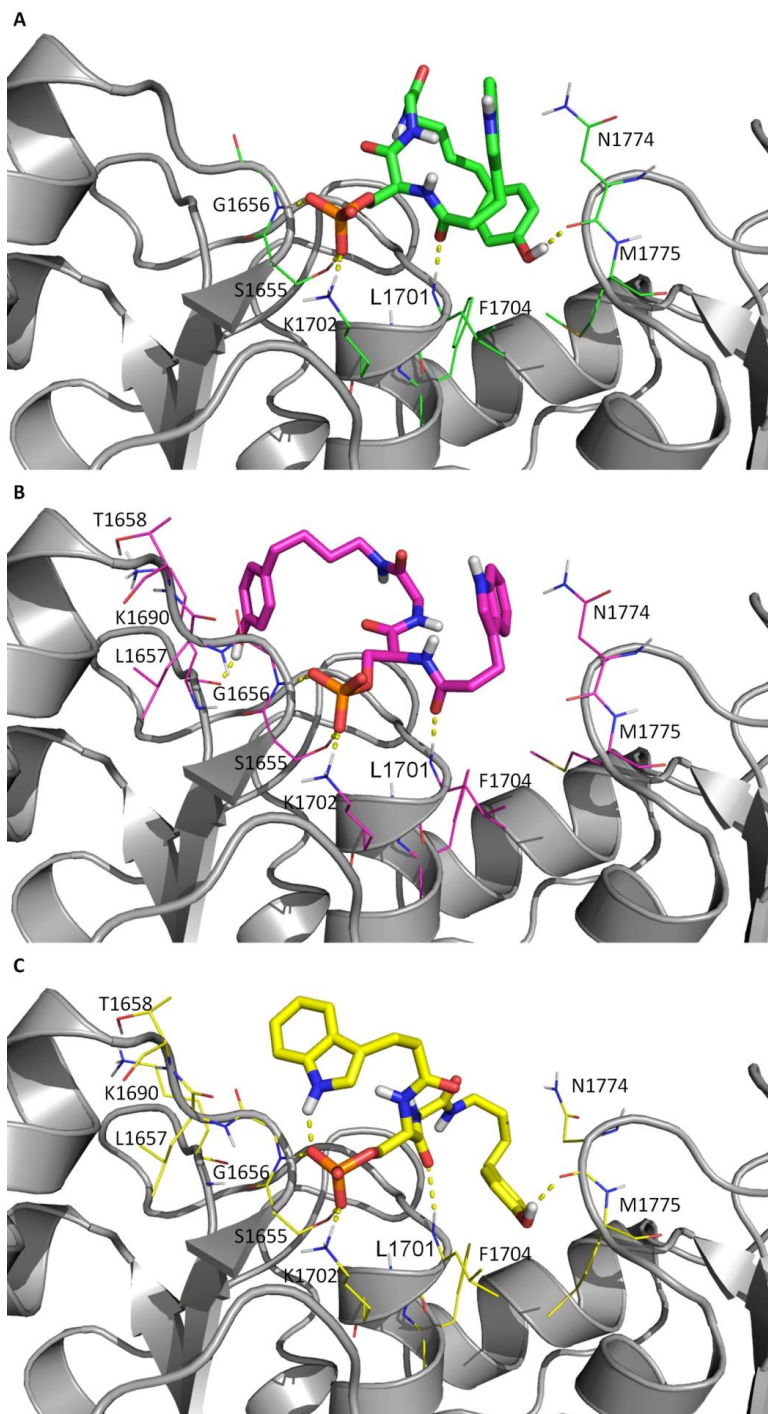
Figure 2.21. (A, B, C) Three distinct bound conformations of D1 from M2 calculations. Residues of BRCT are shown in line mode and the ligand is shown in licorice mode, hydrogen bonds are drawn in dash lines (free energies of A, B and C bound conformations are -1476.72, -1476.28 and -1465.36 kcal/mol, respectively).

In summary, designing a pre-rigidified ligand to reduce entropy cost can be tricky considering the potential loss of intermolecular attraction due to lack of proper rearrangement in the bound state. Fortunately, making a ligand more flexible and able to retain its plasticity in the bound conformation provides an effective strategy to reduce entropy cost, while the optimization of interactions between such a flexible ligand and a target protein can further improve binding affinity. Although for designing tight binders such as many drug-protein binding systems, pre-rigidified may still be the best strategy, our study points out a new direction for designing inhibitors targeting promiscuous modular domains and PPIs.

2.4 References

1.   Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF, Koonin EV. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 1997;11(1):68-76. Epub 1997/01/01. PubMed PMID: 9034168.

2.   Cantor S, Drapkin R, Zhang F, Lin Y, Han J, Pamidi S, et al. The BRCA1-associated protein BACH1 is a DNA helicase targeted by clinically relevant inactivating mutations. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(8):2357-62. Epub 2004/02/26. PubMed PMID: 14983014; PubMed Central PMCID: PMCPmc356955.

3.   Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annual review of biochemistry.   2004;73:39-85.   Epub   2004/06/11.   doi: 10.1146/annurev.biochem.73.011303.073723. PubMed PMID: 15189136.

4.   Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. Cell. 2002;108(2):171-82. Epub 2002/02/08. PubMed PMID: 11832208.

5.   Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, et al. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. Cell. 2001;105(1):149-60. Epub 2001/04/13. PubMed PMID: 11301010.

6.   Kim H, Huang J, Chen J. CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response. Nature structural & molecular biology. 2007;14(8):710-5. Epub 2007/07/24. doi: 10.1038/nsmb1277. PubMed PMID: 17643122.

7.   Manke IA, Lowery DM, Nguyen A, Yaffe MB. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. Science (New York, NY). 2003;302(5645):636-9. Epub 2003/10/25. doi: 10.1126/science.1088877. PubMed PMID: 14576432.

8.   Wang B, Matsuoka S, Ballif BA, Zhang D, Smogorzewska A, Gygi SP, et al. Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. Science (New York, NY). 2007;316(5828):1194-8. Epub 2007/05/26. doi: 10.1126/science.1139476. PubMed PMID: 17525340; PubMed Central PMCID: PMCPmc3573690.

9.   Yu X, Chen J. DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains. Molecular and cellular biology. 2004;24(21):9478-86. Epub 2004/10/16. doi: 10.1128/mcb.24.21.9478-9486.2004. PubMed PMID: 15485915; PubMed Central PMCID: PMCPmc522253.

10.   Yu X, Chini CC, He M, Mer G, Chen J. The BRCT domain is a phospho-protein binding domain. Science (New York, NY). 2003;302(5645):639-42. Epub 2003/10/25. doi: 10.1126/science.1088753. PubMed PMID: 14576433.

11.   Thompson ME. BRCA1 16 years later: nuclear import and export processes. The FEBS journal. 2010;277(15):3072-8. Epub 2010/07/09. doi: 10.1111/j.1742-4658.2010.07733.x. PubMed PMID: 20608972.

12.   Pessetto Z, Yan Y, Bessho T, Natarajan A. Inhibition of BRCT(BRCA1)-phosphoprotein interaction enhances the cytotoxic effect of olaparib in breast cancer cells: a proof of concept study for synthetic lethal therapeutic option. Breast Cancer Res Treat. 2012;134(2):511-7. doi: 10.1007/s10549-012-2079-4.

13.   Cheng KY, Lowe ED, Sinclair J, Nigg EA, Johnson LN. The crystal structure of the human polo‑like kinase‑1 polo box domain and its phospho‑peptide complex. The EMBO Journal. 2003;22(21):5757-68. doi: 10.1093/emboj/cdg558.

14.   Ilsley JL, Sudol M, Winder SJ. The interaction of dystrophin with beta-dystroglycan is regulated by tyrosine phosphorylation. Cellular signalling. 2001;13(9):625-32. Epub 2001/08/10. PubMed PMID: 11495720.

15.   Liu BA, Jablonowski K, Shah EE, Engelmann BW, Jones RB, Nash PD. SH2 domains recognize contextual peptide sequence information to determine selectivity. Molecular & cellular proteomics : MCP. 2010;9(11):2391-404. Epub 2010/07/16. doi: 10.1074/mcp.M110.001586. PubMed PMID: 20627867; PubMed Central PMCID: PMCPmc2984226.

16.   Mahajan A, Yuan C, Lee H, Chen ES, Wu PY, Tsai MD. Structure and function of the phosphothreonine-specific FHA domain. Science signaling. 2008;1(51):re12. Epub 2008/12/26. doi: 10.1126/scisignal.151re12. PubMed PMID: 19109241.

17.   Sudol M, Bedford M. Competitive Binding of Proline-Rich Sequences by SH3, WW, and Other Functionally Related Protein Domains. In: Waksman G, editor. Proteomics and Protein-Protein Interactions. Protein Reviews. 3: Springer US; 2005. p. 185-201.

18.   Yip KY, Utz L, Sitwell S, Hu X, Sidhu SS, Turk BE, et al. Identification of specificity determining residues in peptide recognition domains using an information theoretic approach applied to large-scale binding maps. BMC biology. 2011;9:53. Epub 2011/08/13. doi: 10.1186/1741-7007-9-53. PubMed PMID: 21835011; PubMed Central PMCID: PMCPmc3224579.

19.   Laraia L, McKenzie G, Spring David R, Venkitaraman Ashok R, Huggins David J. Overcoming Chemical, Biological, and Computational Challenges in the Development of Inhibitors Targeting Protein-Protein Interactions. Chemistry & Biology. 2015;22(6):689-703. doi: http://dx.doi.org/10.1016/j.chembiol.2015.04.019.

20.   Watanabe N, Osada H. Phosphorylation-Dependent Protein-Protein Interaction Modules As Potential Molecular Targets for Cancer Therapy. Current Drug Targets. 2012;13(13):1654-8. doi: 10.2174/138945012803530035.

21.   Tan YS, Spring DR, Abell C, Verma CS. The Application of Ligand-Mapping Molecular Dynamics Simulations to the Rational Design of Peptidic Modulators of Protein–Protein Interactions. Journal of chemical theory and computation. 2015;11(7):3199-210. doi: 10.1021/ct5010577.

22.   Lokesh GL, Muralidhara BK, Negi SS, Natarajan A. Thermodynamics of Phosphopeptide Tethering to BRCT: The Structural Minima for Inhibitor Design.

Journal of the American Chemical Society. 2007;129(35):10658-9. doi: 10.1021/ja0739178.

23.   Ren Z, Cabell LA, Schaefer TS, McMurray JS. Identification of a high-affinity phosphopeptide inhibitor of Stat3. Bioorganic & medicinal chemistry letters. 2003;13(4):633-6. Epub 2003/03/18. PubMed PMID: 12639546.

24.   Yuan Z, Kumar EA, Kizhake S, Natarajan A. Structure-activity relationship studies to probe the phosphoprotein binding site on the carboxy terminal domains of the breast cancer susceptibility gene 1. Journal of medicinal chemistry. 2011;54(12):4264-8. Epub 2011/05/18. doi: 10.1021/jm1016413. PubMed PMID: 21574625; PubMed Central PMCID: PMCPmc3117075.

25.   Anisimov V, Ziemys A, Kizhake S, Yuan Z, Natarajan A, Cavasotto C. Computational and experimental studies of the interaction between phospho-peptides and the C-terminal domain of BRCA1. Journal of computer-aided molecular design. 2011;25(11):1071-84. doi: 10.1007/s10822-011-9484-3.

26.   Yuan Z, Kumar EA, Campbell SJ, Palermo NY, Kizhake S, Glover JNM, et al. Exploiting the P-1 Pocket of BRCT Domains Toward a Structure Guided Inhibitor Design. ACS medicinal chemistry letters. 2011;2(10):764-7. doi: 10.1021/ml200147a.

27.   Na Z, Pan S, Uttamchandani M, Yao SQ. Discovery of cell-permeable inhibitors that target the BRCT domain of BRCA1 protein by using a small-molecule microarray. Angewandte Chemie (International ed in English). 2014;53(32):8421-6. Epub 2014/06/26. doi: 10.1002/anie.201405169. PubMed PMID: 24961672.

28.   White ER, Sun L, Ma Z, Beckta JM, Danzig BA, Hacker DE, et al. Peptide library approach to uncover phosphomimetic inhibitors of the BRCA1 C-terminal domain. ACS chemical biology. 2015;10(5):1198-208. Epub 2015/02/06. doi: 10.1021/cb500757u. PubMed PMID: 25654734; PubMed Central PMCID: PMCPmc4433557.

29.   Ivanov AA, Khuri FR, Fu H. Targeting protein-protein interactions as an anticancer strategy. Trends in pharmacological sciences. 2013;34(7):393-400. Epub 2013/06/04. doi: 10.1016/j.tips.2013.04.007. PubMed PMID: 23725674; PubMed Central PMCID: PMCPmc3773978.

30.   Mullard A. Protein–protein interaction inhibitors get into the groove. Nature reviews Drug discovery. 2012;11(3):173-5.

31.   Gough CA, Gojobori T, Imanishi T. Cancer-related mutations in BRCA1-BRCT cause long-range structural changes in protein-protein binding sites: a molecular dynamics study. Proteins. 2007;66(1):69-86. Epub 2006/10/26. doi: 10.1002/prot.21188. PubMed PMID: 17063491.

32.   Huang YM, Kang M, Chang CE. Mechanistic insights into phosphopeptide--BRCT domain association: preorganization, flexibility, and phosphate recognition. The journal of physical chemistry B. 2012;116(34):10247-58. Epub 2012/08/04. doi: 10.1021/jp305028d. PubMed PMID: 22857521.

33.   Carvalho RS, Abreu RB, Velkova A, Marsillac S, Rodarte RS, Suarez-Kurtz G, et al. Probing structure-function relationships in missense variants in the carboxy-terminal region of BRCA1. PloS one. 2014;9(5):e97766. Epub 2014/05/23. doi:

10.1371/journal.pone.0097766. PubMed PMID: 24845084; PubMed Central PMCID: PMCPmc4028255.

34.	Iversen ES, Jr., Couch FJ, Goldgar DE, Tavtigian SV, Monteiro AN. A computational method to classify variants of uncertain significance using functional assay data with application to BRCA1. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2011;20(6):1078-88. Epub 2011/03/31. doi: 10.1158/1055-9965.epi-10-1214. PubMed PMID: 21447777; PubMed Central PMCID: PMCPmc3111818.

35.	Basdevant N, Weinstein H, Ceruso M. Thermodynamic Basis for Promiscuity and Selectivity in Protein−Protein Interactions: PDZ Domains, a Case Study. Journal of the American Chemical Society. 2006;128(39):12766-77. doi: 10.1021/ja060830y.

36.	DeLorbe JE, Clements JH, Teresk MG, Benfield AP, Plake HR, Millspaugh LE, et al. Thermodynamic and Structural Effects of Conformational Constraints in Protein−Ligand Interactions. Entropic Paradoxy Associated with Ligand Preorganization. Journal of the American Chemical Society. 2009;131(46):16758-70. doi: 10.1021/ja904698q.

37.	Gerek ZN, Keskin O, Ozkan SB. Identification of specificity and promiscuity of PDZ domain interactions through their dynamic behavior. Proteins. 2009;77(4):796-811. Epub 2009/07/09. doi: 10.1002/prot.22492. PubMed PMID: 19585657.

38.	Huang YM, Chang CE. Achieving peptide binding specificity and promiscuity by loops: case of the forkhead-associated domain. PloS one. 2014;9(5):e98291. Epub

2014/05/30. doi: 10.1371/journal.pone.0098291. PubMed PMID: 24870410; PubMed Central PMCID: PMCPmc4037201.

39.   Roy A, Hua DP, Ward JM, Post CB. Relative Binding Enthalpies from Molecular Dynamics Simulations Using a Direct Method. Journal of chemical theory and computation. 2014;10(7):2759-68. doi: 10.1021/ct500200n.

40.   Shi Y, Zhu CZ, Martin SF, Ren P. Probing the Effect of Conformational Constraint on Phosphorylated Ligand Binding to an SH2 Domain Using Polarizable Force Field Simulations. The Journal of Physical Chemistry B. 2012;116(5):1716-27. doi: 10.1021/jp210265d.

41.   Udugamasooriya DG, Spaller MR. Conformational constraint in protein ligand design and the inconsistency of binding entropy. Biopolymers. 2008;89(8):653-67. Epub 2008/03/13. doi: 10.1002/bip.20983. PubMed PMID: 18335423.

42.   Wang W, Weng J, Zhang X, Liu M, Zhang M. Creating Conformational Entropy by Increasing Interdomain Mobility in Ligand Binding Regulation: A Revisit to N-Terminal Tandem PDZ Domains of PSD-95. Journal of the American Chemical Society. 2009;131(2):787-96. doi: 10.1021/ja8076022.

43.   Yang C-Y, Wang S. Hydrophobic Binding Hot Spots of Bcl-xL Protein−Protein Interfaces by Cosolvent Molecular Dynamics Simulation. ACS medicinal chemistry letters. 2011;2(4):280-4. doi: 10.1021/ml100276b.

44.   Wang J, Morin P, Wang W, Kollman PA. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode

to HIV-1 RT of efavirenz by docking and MM-PBSA. Journal of the American Chemical Society. 2001;123(22):5221-30. Epub 2001/07/18. PubMed PMID: 11457384.

45.   Itoh Y, Shichinohe S, Nakayama M, Igarashi M, Ishii A, Ishigaki H, et al. Emergence of H7N9 Influenza A Virus Resistant to Neuraminidase Inhibitors in Nonhuman Primates. Antimicrobial agents and chemotherapy. 2015;59(8):4962-73. Epub 2015/06/10. doi: 10.1128/aac.00793-15. PubMed PMID: 26055368; PubMed Central PMCID: PMCPmc4505273.

46.   Shiozaki EN, Gu L, Yan N, Shi Y. Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling. Molecular cell. 2004;14(3):405-12. Epub 2004/05/06. PubMed PMID: 15125843.

47.   Varma AK, Brown RS, Birrane G, Ladias JA. Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex. Biochemistry. 2005;44(33):10941-6. Epub 2005/08/17. doi: 10.1021/bi0509651. PubMed PMID: 16101277.

48.   Williams RS, Lee MS, Hau DD, Glover JN. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. Nature structural & molecular biology. 2004;11(6):519-25. Epub 2004/05/11. doi: 10.1038/nsmb776. PubMed PMID: 15133503.

49.   Shen Y, Tong L. Structural evidence for direct interactions between the BRCT domains of human BRCA1 and a phospho-peptide from human ACC1. Biochemistry. 2008;47(21):5767-73. Epub 2008/05/03. doi: 10.1021/bi800314m. PubMed PMID: 18452305; PubMed Central PMCID: PMCPmc2392887.

50.   Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of computational chemistry. 1998;19(14):1639-62.

51.   Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry. 2009;30(16):2785-91. Epub 2009/04/29. doi: 10.1002/jcc.21256. PubMed PMID: 19399780; PubMed Central PMCID: PMCPmc2760638.

52.   Gilson MK, Gilson HS, Potter MJ. Fast assignment of accurate partial atomic charges: an electronegativity equalization method that accounts for alternate resonance forms. Journal of chemical information and computer sciences. 2003;43(6):1982-97. Epub 2003/11/25. doi: 10.1021/ci034148o. PubMed PMID: 14632449.

53.   D.A. Case VB, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E., Duke HG, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko,, T.S. Lee SL, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui,, R. Salomon-Ferrer GS, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X., Wu PAK. AMBER 14. 2014;University of California, San Francisco.

54.   Case DA, Cheatham TE, 3rd, Darden T, Gohlke H, Luo R, Merz KM, Jr., et al. The Amber biomolecular simulation programs. Journal of computational chemistry. 2005;26(16):1668-88. Epub 2005/10/04. doi: 10.1002/jcc.20290. PubMed PMID: 16200636; PubMed Central PMCID: PMCPmc1989667.

55.   Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 2006;65(3):712-25. Epub 2006/09/19. doi: 10.1002/prot.21123. PubMed PMID: 16981200.

56.   Okur A, Strockbine B, Hornak V, Simmerling C. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. Journal of computational chemistry. 2003;24(1):21-31. Epub 2002/12/17. doi: 10.1002/jcc.10184. PubMed PMID: 12483672.

57.   Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. Journal of computational chemistry. 2005;26(16):1781-802. Epub 2005/10/14. doi: 10.1002/jcc.20289. PubMed PMID: 16222654; PubMed Central PMCID: PMCPmc2486339.

58.   Homeyer N, Horn AH, Lanig H, Sticht H. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. Journal of molecular modeling. 2006;12(3):281-9. Epub 2005/10/22. doi: 10.1007/s00894-005-0028-4. PubMed PMID: 16240095.

59.   Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. The Journal of Chemical Physics. 1995;103(19):8577-93. doi: doi:http://dx.doi.org/10.1063/1.470117.

60.   Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.

Journal of Computational Physics. 1977;23(3):327-41. doi: http://dx.doi.org/10.1016/0021-9991(77)90098-5.

61.   Chang CE, Gilson MK. Tork: Conformational analysis method for molecules and complexes. Journal of computational chemistry. 2003;24(16):1987-98. Epub 2003/10/08. doi: 10.1002/jcc.10325. PubMed PMID: 14531053.

62.   Chang C-E, Potter MJ, Gilson MK. Calculation of Molecular Configuration Integrals. The Journal of Physical Chemistry B. 2003;107(4):1048-55. doi: 10.1021/jp027149c.

63.   Chang CE, Gilson MK. Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. Journal of the American Chemical Society. 2004;126(40):13156-64. Epub 2004/10/08. doi: 10.1021/ja047115d. PubMed PMID: 15469315.

64.   Chen W, Gilson MK, Webb SP, Potter MJ. Modeling Protein-Ligand Binding by Mining Minima. Journal of chemical theory and computation. 2010;6(11):3540-57. Epub 2010/11/09. doi: 10.1021/ct100245n. PubMed PMID: 22639555; PubMed Central PMCID: PMCPmc3359898.

65.   Huang Y-ming M, Chen W, Potter Michael J, Chang C-en A. Insights from Free-Energy Calculations: Protein Conformational Equilibrium, Driving Forces, and Ligand-Binding Modes. Biophysical Journal. 2012;103(2):342-51. doi: http://dx.doi.org/10.1016/j.bpj.2012.05.046.

66.   Ai R, Qaiser Fatmi M, Chang CE. T-Analyst: a program for efficient analysis of protein conformational changes by torsion angles. Journal of computer-aided molecular

design. 2010;24(10):819-27. Epub 2010/08/07. doi: 10.1007/s10822-010-9376-y. PubMed PMID: 20689979; PubMed Central PMCID: PMCPmc2940022.

67.   Amaro RE, Cheng X, Ivanov I, Xu D, McCammon JA. Characterizing loop dynamics and ligand recognition in human- and avian-type influenza neuraminidases via generalized born molecular dynamics and end-point free energy calculations. Journal of the American Chemical Society. 2009;131(13):4702-9. Epub 2009/03/20. doi: 10.1021/ja8085643. PubMed PMID: 19296611; PubMed Central PMCID: PMCPmc2665887.

68.   Bashford D, Case DA. Generalized born models of macromolecular solvation effects. Annual review of physical chemistry. 2000;51:129-52. Epub 2000/10/14. doi: 10.1146/annurev.physchem.51.1.129. PubMed PMID: 11031278.

69.   Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert opinion on drug discovery. 2015;10(5):449-61. Epub 2015/04/04. doi: 10.1517/17460441.2015.1032936. PubMed PMID: 25835573; PubMed Central PMCID: PMCPmc4487606.

70.   Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. Journal of chemical information and modeling. 2011;51(1):69-82. Epub 2010/12/02. doi: 10.1021/ci100275a. PubMed PMID: 21117705; PubMed Central PMCID: PMCPmc3029230.

71.   Sitkoff D, Sharp KA, Honig B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. The Journal of Physical Chemistry. 1994;98(7):1978-88. doi: 10.1021/j100058a043.

72.   Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers. 2000;56(4):275-91. Epub 2002/01/05. doi: 10.1002/1097-0282(2000)56:4<275::aid-bip10024>3.0.co;2-e. PubMed PMID: 11754341.

73.   Wang J, Cai Q, Li ZL, Zhao HK, Luo R. Achieving Energy Conservation in Poisson-Boltzmann Molecular Dynamics: Accuracy and Precision with Finite-Difference Algorithms. Chemical physics letters. 2009;468(4-6):112-8. Epub 2010/01/26. doi: 10.1016/j.cplett.2008.12.049. PubMed PMID: 20098487; PubMed Central PMCID: PMCPmc2663913.

74.   Ye X, Wang J, Luo R. A Revised Density Function for Molecular Surface Calculation in Continuum Solvent Models. Journal of chemical theory and computation. 2010;6(4):1157-69. doi: 10.1021/ct900318u.

75.   Petukh M, Li M, Alexov E. Predicting Binding Free Energy Change Caused by Point Mutations with Knowledge-Modified MM/PBSA Method. PLoS computational biology. 2015;11(7):e1004276. Epub 2015/07/07. doi: 10.1371/journal.pcbi.1004276. PubMed PMID: 26146996; PubMed Central PMCID: PMCPmc4492929.

76.   Suri C, Naik PK. Combined molecular dynamics and continuum solvent approaches (MM-PBSA/GBSA) to predict noscapinoid binding to gamma-tubulin dimer.

SAR and QSAR in environmental research. 2015;26(6):507-19. Epub 2015/08/15. doi: 10.1080/1062936x.2015.1070200. PubMed PMID: 26274780.

77.   Lokesh GL, Muralidhara BK, Negi SS, Natarajan A. Thermodynamics of phosphopeptide tethering to BRCT: the structural minima for inhibitor design. Journal of the American Chemical Society. 2007;129(35):10658-9. Epub 2007/08/10. doi: 10.1021/ja0739178. PubMed PMID: 17685618.

78.   Yuan Z, Kumar EA, Campbell SJ, Palermo NY, Kizhake S, Mark Glover JN, et al. Exploiting the P-1 pocket of BRCT domains toward a structure guided inhibitor design. ACS Med Chem Lett. 2011;2(10):764-7. doi: 10.1021/ml200147a. PubMed PMID: 22046493; PubMed Central PMCID: PMCPMC3201719.

79.   Lokesh GL, Rachamallu A, Kumar GD, Natarajan A. High-throughput fluorescence polarization assay to identify small molecule inhibitors of BRCT domains of breast cancer gene 1. Anal Biochem. 2006;352(1):135-41. doi: 10.1016/j.ab.2006.01.025. PubMed PMID: 16500609.

80.   Simeonov A, Yasgar A, Jadhav A, Lokesh GL, Klumpp C, Michael S, et al. Dual-fluorophore quantitative high-throughput screen for inhibitors of BRCT-phosphoprotein interaction. Anal Biochem. 2008;375(1):60-70. doi: 10.1016/j.ab.2007.11.039. PubMed PMID: 18158907; PubMed Central PMCID: PMCPMC3389998.

81.   Joseph PR, Yuan Z, Kumar EA, Lokesh GL, Kizhake S, Rajarathnam K, et al. Structural characterization of BRCT-tetrapeptide binding interactions. Biochem Biophys Res Commun. 2010;393(2):207-10. doi: 10.1016/j.bbrc.2010.01.098. PubMed PMID: 20122900; PubMed Central PMCID: PMCPMC2834807.

82.   Andrew RL. Molecular modelling: principles and applications. Prentice Hall; 2001.

83.   Nicholls A. Confidence limits, error bars and method comparison in molecular modeling. Part 1: the calculation of confidence intervals. Journal of computer-aided molecular design. 2014;28(9):887-918. Epub 2014/06/06. doi: 10.1007/s10822-014-9753-z. PubMed PMID: 24899109; PubMed Central PMCID: PMCPmc4175406.

84.   Nicholls A. Statistics in molecular modeling: a summary. Journal of computer-aided molecular design. 2016;30(4):279-80. Epub 2016/03/24. doi: 10.1007/s10822-016-9907-2. PubMed PMID: 27001050.

85.   Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. Journal of the American Chemical Society. 2015;137(7):2695-703. Epub 2015/01/28. doi: 10.1021/ja512751q. PubMed PMID: 25625324.

86.   Brown SP, Muchmore SW, Hajduk PJ. Healthy skepticism: assessing realistic model performance. Drug discovery today. 2009;14(7-8):420-7. Epub 2009/04/03. PubMed PMID: 19340931.

87.   Conlon JM, Ahmed E, Condamine E. Antimicrobial properties of brevinin-2-related peptide and its analogs: Efficacy against multidrug-resistant Acinetobacter baumannii. Chemical biology & drug design. 2009;74(5):488-93. Epub 2009/10/02. doi: 10.1111/j.1747-0285.2009.00882.x. PubMed PMID: 19793185.

88.   Ferenczy GG, Keseru GM. Thermodynamics guided lead discovery and optimization. Drug discovery today. 2010;15(21-22):919-32. Epub 2010/08/31. doi: 10.1016/j.drudis.2010.08.013. PubMed PMID: 20801227.

89.   Ferenczy GG, Keseru GM. Thermodynamics of fragment binding. Journal of chemical information and modeling. 2012;52(4):1039-45. Epub 2012/03/31. doi: 10.1021/ci200608b. PubMed PMID: 22458364.

90.   Hann MM, Keseru GM. Finding the sweet spot: the role of nature and nurture in medicinal chemistry. Nature reviews Drug discovery. 2012;11(5):355-65. Epub 2012/05/01. doi: 10.1038/nrd3701. PubMed PMID: 22543468.

91.   Ladbury JE, Klebe G, Freire E. Adding calorimetric data to decision making in lead discovery: a hot tip. Nature reviews Drug discovery. 2010;9(1):23-7. Epub 2009/12/05. doi: 10.1038/nrd3054. PubMed PMID: 19960014.

92.   Olsson TS, Williams MA, Pitt WR, Ladbury JE. The thermodynamics of protein-ligand interaction and solvation: insights for ligand design. Journal of molecular biology. 2008;384(4):1002-17. Epub 2008/10/22. doi: 10.1016/j.jmb.2008.09.073. PubMed PMID: 18930735.

93.   Reynolds CH, Holloway MK. Thermodynamics of ligand binding and efficiency. ACS medicinal chemistry letters. 2011;2(6):433-7. Epub 2011/06/09. doi: 10.1021/ml200010k. PubMed PMID: 24900326; PubMed Central PMCID: PMCPmc4018058.

94.   Glas A, Bier D, Hahne G, Rademacher C, Ottmann C, Grossmann TN. Constrained peptides with target-adapted cross-links as inhibitors of a pathogenic protein-

protein interaction. Angewandte Chemie (International ed in English). 2014;53(9):2489-93. Epub 2014/02/08. doi: 10.1002/anie.201310082. PubMed PMID: 24504455.

95.   Martin SF, Clements JH. Correlating structure and energetics in protein-ligand interactions: paradigms and paradoxes. Annual review of biochemistry. 2013;82:267-93. Epub 2013/06/12. doi: 10.1146/annurev-biochem-060410-105819. PubMed PMID: 23746256.

96.   Kumar EA, Chen Q, Kizhake S, Kolar C, Kang M, Chang CE, et al. The paradox of conformational constraint in the design of Cbl(TKB)-binding peptides. Sci Rep. 2013;3:1639. Epub 2013/04/11. doi: 10.1038/srep01639. PubMed PMID: 23572190; PubMed Central PMCID: PMCPmc3965358.

Chapter 3 Role of Molecular Interactions and Protein Conformational Rearrangement in the Kinetics of p38α-inhibitors Dissociation

3.1 Introduction

The study of small molecular kinase inhibitors has been the central focus in drug discovery in the past decade [1-8]. To develop a good inhibitor, one first needs to consider binding affinity so to compete with the natural substrate. However, a sole strong binding affinity doesn't guarantee good drug efficacy, another crucial factor is the mean residence time (RT) in the binding pocket, which is usually quantified by dissociation rate constant through the equation $RT = 1/k_{off}$ [9-12]. Since longer residence time means the drug stays longer in target site, leading to better drug efficacy, it has become important to fully understand the kinetic process of dissociation. In this study, we applied multiple enhanced sampling methods to investigate the dissociation pathway of inhibitors of an attractive kinase system.

One of the kinase systems of interest as drug target is p38, which belongs to the mitogen-activated protein kinases (MAPKs), a super-family of enzymes that is involved in regulation of cell functions including proliferation, gene expression, differentiation and apoptosis [13, 14]. The p38 kinases have four isoforms, p38α, β, γ, and δ. The major isoform p38α has been a drug target for treating various inflammatory diseases, including rheumatoid arthritis, asthma, and cardiovascular disease [15-17]. Great efforts have been made to develop various inhibitors to p38α as potential drugs against a wide range of diseases [18-21]. To speed up the discovery of new inhibitors, it is necessary to

investigate the interactions between p38α and current inhibitors, as well as essential conformational changes of system during dissociation process.

Like all protein kinases, p38α has a structurally conserved catalytic domain consisting of two lobes, the N-terminal and the C-terminal lobes, which are connected through a flexible hinge region [22]. The activation loop, which carries a DFG (Asp-Phe-Gly) motif, belongs to the C-terminal lobe but locates outside of the ATP binding pocket. It directly regulates the enzyme activation through its conformational changes, which can be characterized by different orientations of the sidechain of Phe from DFG motif, the active conformation where Phe buried in αC helix (DFG-in) and inactive conformation where Phe sterically interferes with ATP binding (DFG-out) (Figure 3.1). NMR studies have proved the equal population of DFG-in and DFG-out conformations in the free state of p38α [23]. And small molecule inhibitors have been found to inhibitor p38α with either DFG-in or DFG-out conformation. Most inhibitors bind to the ATP binding pocket in DFG-in conformation (type I inhibitors) [24]. Some inhibitors occupy both the ATP site and a nearby allosteric hydrophobic pocket which becomes accessible in DFG-out conformation (type-II inhibitors) [25], or bind exclusively within the allosteric pocket (type-III inhibitors) [26].
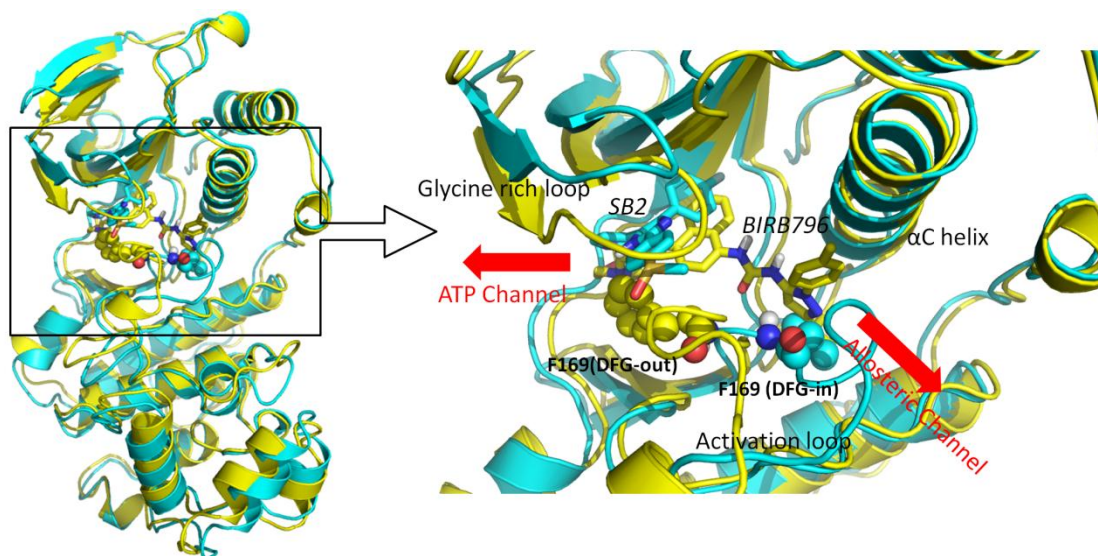
Figure 3.1. The comparison of crystal structures of p38α in DFG-in and DFG-out conformations. The left figure shows the structure alignment of DFG-in (cyan, PDB 1A9U) and DFG-out (yellow, PDB 1KV2) conformations bound with ligands SB2 and BIRB796, respectively. The right figure shows binding site structures. The Phe169 from DFG motif is shown in sphere structure, ligands SB2 and BIRB796 are shown in licorice structure.

Type I inhibitors target the DFG-in conformation of the kinases, which completely exposes the ATP-binding pocket to the solvent, thus facilitating the association and dissociation of ATP or type I inhibitors. This association/dissociation pathway can be called ATP channel. When a type II or type III inhibitor binds to the allosteric pocket of DFG-out kinase, the ATP-pocket channel narrows due to the interference of DFG motif, while another possible association/dissociation pathway known as the allosteric-pocket channel comes into sight. Many studies have been performed on the dissociations of these different types of inhibitors. For instance, Casasnovas et al. have used metadynamics simulations to study the unbinding kinetics of a p38α type-II inhibitor [27]. Yang et al. have used steered molecular dynamics (SMD) simulations to explore the possible dissociation pathways of type II inhibitor from kinases c-Kit and Abl [28]. Capelli et al.

have studied the dissociation process of type I inhibitors of VEGFR2 by using SMD [29]. Sun et al. have exploited the free energy profile change of unbinding process for different types of inhibitors of several kinase systems, ALK tyrosine kinase by using adaptive biasing force (ABF) simulations [30], and p38α kinase by using umbrella sampling (US) [31]. However, most of them used biasing force, as in SMD and ABF, which may have introduced uncertain artificial errors to the simulations of dissociation pathway.

Here we applied enhanced sampling methods, including accelerated molecular dynamics (AMD), pathway search guided by internal motions (PSIM) and umbrella sampling, to study the dissociation pathway of four p38α inhibitors, two type I (SB2, SK8), one type II (BIRB796) and one type III (LIG4) inhibitors. Interestingly, one of the type I inhibitor SB2 can bind to ATP site while DFG motif in either in or out conformation. Umbrella sampling was performed to construct PMF to get information about free energy change associated with change of interactions and protein conformations along dissociation of inhibitors and to determine the energetically favorable dissociation pathway. Correlation study was applied to find the important parts of protein that correlate with dissociation of inhibitors. The results provide useful insights about interaction and protein conformational changes during dissociation process of inhibitors, which may help design new inhibitors and guide mutation experiments in the future.

3.2 Materials and Methods

3.2.1 Molecular systems

Table 3.1 characterizes the structures and binding properties, crystal structure sources and binding modes of four inhibitors of p38α, among these four inhibitors SB2 binds to both DFG-in conformation (PDB ID: 1A9U) [32] and DFG-out conformation (PDB ID: 3GCP) [33]. Structures of free DFG-in and DFG-out conformations are from DFG-in complex (PDB ID: 1A9U) and DFG-out complex (PDB ID: 1W82) [34] with ligands removed from their crystal structures. The loop region encompassing residues 173-184 in p38α is not available in the crystal structure 3GCP, and the partial missing loop was modeled using the loop conformation from selected frame from MD simulation of free DFG-out protein. BIRB796 binds to PDB ID 1KV2 [35] whose missing loop encompassing residues 115-122, 170-184 was completed with corresponding parts in crystal structure of 1W82. After replacing the missing loop with complete loop conformation, a quick 100-step energy minimization was carried out within the substituted loop and its adjacent residues to refine the new loop conformation. The structures of p38α in complex with SK8 and LIG4 are not available. To model them, we started from the 1A9U and 1W82 whose ligands are structurally similar to SK8 and LIG4, respectively. In 1A9U, we model SK8 by replacing 4-methylsulfinylphenyl group of its original ligand with thiazole analog. In 1W82, we model LIG4 by replacing one chlorine atom of its original ligand with hydrogen atom and adding one methyl group to benzene ring. Multiple simulations were performed on these systems, as seen in the following subsections, and a summary of simulations performed on each system is listed in Table 3.2.

Table 3.1. Chemical structures, interaction modes, and protein kinase activity of p38α inhibitors used in the study. Rotatable dihedral angles of p38α inhibitors are highlighted in red. K_D, k_{on}, k_{off} were taken from [35-37].
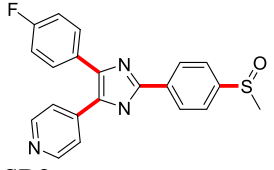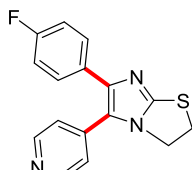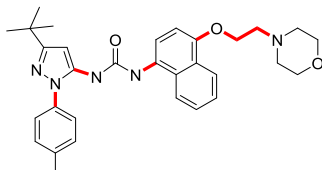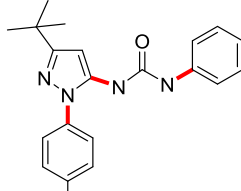
No.	Structure	K_D (nM)	k_{on} $[M^{-1}s^{-1}]$	k_{off} $[s^{-1}]$	ΔG_{exp} (kcal/mol)	PDB ID	Mode
1	SB2	11.5	1.5×10^7	1.8×10^{-1}	-10.9	1A9U (3GCP)	DFG-in (DFG-out)
2	SK8	180	4.3×10^7	7.7	-9.7	N/A	DFG-in
3	BIRB796	0.1	8.4×10^4	8.3×10^{-6}	-13.7	1KV2	DFG-out
4	LIG4	21	7.3×10^4	1.6×10^{-3}	-10.5	N/A	DFG-out

Table 3.2. Simulations performed for each system.

Free DFG-in protein	1. Three CMD runs with different random number seeds (100ns CMD1, CMD2 and CMD3). 2. CMD1, CMD2 and CMD3 were used for compute average RMSF and correlation values.
Free DFG-out protein	1. Three CMD runs with different random number seeds (100ns CMD1, CMD2 and CMD3). 2. CMD1, CMD2 and CMD3 were used for compute average RMSF and correlation values.
SB2 (DFG-in) complex	1. Three CMD runs with different random number seeds (650ns CMD1, 100ns CMD2 and CMD3). 2. One low boost 100ns AMD continued from the first 100ns of CMD1. 3. Two successful 100ns high boost AMD continued from the first 100ns of CMD1. 4. Two 10ns CMD trajectories starting from selected frames of high boost AMD were used to rebuilt the whole smooth dissociation path. 5. CMD2, CMD3 and the first 100ns CMD1 were used for compute average RMSF and correlation values.
SB2 (DFG-out) complex	1. Three CMD runs with different random number seeds (100ns CMD1, CMD2 and CMD3). 2. Two successful 100ns high boost AMD continued from CMD1. 3. Two 10ns CMD trajectories starting from selected frames of high boost AMD were used to rebuilt the whole smooth dissociation path. 4. CMD1, CMD2 and CMD3 were used for compute average RMSF and correlation values.
SK8 complex	1. Three CMD runs with different random number seeds (100ns CMD1, CMD2 and CMD3). 2. Two successful 100ns high boost AMD continued from CMD1. 3. Three 10ns CMD trajectories starting from selected frames of high boost AMD were used to rebuilt the whole smooth dissociation path. 4. CMD1, CMD2 and CMD3 were used for compute average RMSF and correlation values.
BIRB796 complex	1. Three CMD runs with different random number seeds (650ns CMD1, 100ns CMD2 and CMD3). 2. One PSIM search for allosteric pathway and ATP pathway each. 3. Two 10ns CMD trajectories starting from selected frames of PSIM were used to rebuilt the two whole smooth dissociation paths each. 4. CMD2, CMD3 and the first 100ns CMD1 were used for compute average RMSF and correlation values.
LIG4 complex	1. Three CMD runs with different random number seeds (100ns CMD1, CMD2 and CMD3). 2. One PSIM search for allosteric pathway and ATP pathway each. 3. Three 10ns CMD trajectories starting from selected frames of PSIM were used to rebuilt the whole smooth allosteric path, and two CMD for ATP path. 4. CMD1, CMD2 and CMD3 were used for compute average RMSF and correlation values.

3.2.2 Molecular dynamics simulations

We performed MD simulations on four p38α complexes (SB2 has both DFG-in and DFG-out bound conformations simulated), free DFG-in and DFG-out proteins with the standard simulation package, Amber14 [38]. The Amber 99SB force field [39-42] was used for protein and the general Amber force field (gaff) was used for ligands. The partial charges of ligands were calculated by using the Vcharge program. We set up each system as follows. First, hydrogen, side-chain and whole system were minimized for 500, 5 000 and 5 000 steps, respectively, followed by solvating the systems with a rectangular box of a 12-Å explicit TIP3P water model by the tleap program in Amber14. Each system contains 60 000 ~ 70 000 atoms. Na+ ions were added as counter ions to keep the whole system neutral, and particle mesh Ewald was used to consider long-range electrostatic interactions [43]. Before equilibration, we minimized waters and the whole system for 10 000 and 20 000 steps, respectively, followed by equilibrium of solvent molecules for 40 ps. Then the systems were gradually heated from 250 K for 20 ps, 275 K for 20 ps, and 300 K for 160 ps. Frame were saved every 1 ps with a time step of 2 fs in the isothermic-isobaric (NPT) ensemble (T = 300 K and P = 1 atm). We also used the SHAKE procedure to constrain the covalent bonds involving hydrogen atoms during MD simulations [44]. Finally, all production runs were performed for no less than 100 ns at 300 K.

3.2.3 Accelerated molecular dynamics simulations

Accelerated Molecular Dynamics (AMD) uses a bias potential introduced by the McCammon group [45]. It enhances the conformational sampling of biological systems by adding a continuous non-negative bias boost potential function $\Delta V(r)$ to the potential energy surface when the system potential is below a reference energy, therefore lowering the local barriers, allowing the calculation to advance faster.

The AMD modification of the potential is defined by the following equation:

$$V^*(r) = V(r) + \Delta V(r)$$

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E \\ \dfrac{(E - V(r))^2}{\alpha + (E - V(r))} & V(r) < E \end{cases}$$

where $V(r)$ is the original potential, E is the reference energy, and $V^*(r)$ is the modified potential. $\Delta V(r)$ is the boost potential, α is the acceleration factor. The potential energy surface is flattened as the acceleration factor α decreases, making it easier to cross energy barriers between local minimas.

The boost potential $\Delta V(r)$ can be further divided into potential-boost and dihedral-boost.

$$\Delta V(r) = \frac{(Ep - V(r))^2}{(\alpha P + Ep - V(r))} + \frac{(Ed - Vd(r))^2}{(\alpha D + Ed - Vd(r))}$$

It allows us to boost independently only the torsional terms of the potential with input parameters (Ed, αD), the whole potential at once(Ep, αP), or the whole potential with an extra boost to the torsions.

For simulations of p38α, we applied both potential-boost and dihedral-boost. The input parameters take the following form:

$Ed = Vd\_avg + 4\ N_{residues}$,

$αD = 0.8\ N_{residues}$

$Ep = Vp\_avg + 0.2\ N_{atoms}$,

$αP = 0.2\ N_{atoms}$

where N_{atoms} and $N_{residues}$ are the total number of atoms and total number of residues of solute, Vd_avg and Vp_avg are the average dihedral and total potential energies calculated from 100ns CMD simulations, respectively. For a higher acceleration, we added another 3*αD to Ed.

3.2.4 Pathway Search Guided by Internal Motions

PSIM is an enhanced conformational search method specifically designed for search of dissociation pathways of ligand-receptor systems. It generates atomistic 3D motions of the system along principal component (PC) axes using a unique multi-layer internal coordinate. By distorting the system using these 3D motions, PSIM performs systematical searches for the dissociation pathways and accepts and rejects new conformations by using geometric criteria rather than energy evaluation or minimization. To ensure reasonable conformations, short minimization on bond and angle terms is performed periodically in addition to the simple geometric criteria.

For the internal principal component analysis (PCA), p38α protein has 349 residues and was divided into 54 fragments. We considered only backbone dihedrals in residues 4 to

113, 137 to 179, 313 to 352, plus side-chain dihedrals in residues 19, 31 to 41, 51 to 56, 67 to 73, 75, 76, 79, 84 to 90, 104 to 111, 139, 141, 142, 147 to 152, 156, 158, 166, 168 to 170, 172, 174, and dihedrals in ligands when constructing the dihedral covariance matrix. All 739 PC modes were used for conformational search. Any randomly chosen frames from the MD simulation could serve as an initial conformation for our search, and we used the initial structure of MD simulation in this case. The output step number was set to 1000 steps, with the step size equal to 0.05. Starting from the initial structure, we performed the PSIM search and repeated for multiple iterations. We manually chose conformations from an existing search to start a new iteration until ligand dissociated from the binding site. We repeated this procedure and performed three to sixteen iterations to obtain dissociation pathways, therefore, the dissociation pathway is a collection of multiple trajectories obtained from each iteration.

3.2.5 Construction of PMF

Umbrella sampling was performed to compute the free energy along the dissociation pathway. By adding multiple overlapping biasing potentials along the dissociation pathway as the reaction coordinate (RC), umbrella sampling can sample all points on the RC suffciently. First, the whole RC is divided into a series of continuous windows, then a biased potential, mostly harmonic potential, is applied to add on the original potential in each window. The equation of harmonic potential is $u_i = k_i(r - r_i)^2$, where u_i is the biased potential in window i, r is the current position of RC, r_i is the reference position in window i, and k_i is the force constant used to restrain the biased molecule in the biased

potential. Here, a force constant of 5 kcal/mol*$Å^2$ was used in all the US simulation windows. WHAM was employed to construct the potential of the mean force (PMF) along the RC [46, 47].

Herein, the RC was separated into bins with 0.2 Å width for the WHAM calculation after each US simulation. The tolerance for iteration was set to 0.0001 to get convergent PMF curves. The temperature was set to 300 K to keep consistence with the simulation temperature. The distance between Cα of Arg73 and CC2 of SB2 (CC2 of SK8) was selected as the RC. For BIRB-796 and LIG4, two sets of RC were selected. The distance between Cα of Met109 and C3 of BIRB796 (C7 of LIG4) was selected as RC for allosteric pathway, and the distance between Cα of Arg73 and C3 of BIRB796 (C12 of LIG4) was selected as RC for ATP pathway. Each set of the simulations contains simulation windows with 0.25 Å in length for each. For each window, 10 ns US simulation was preformed for the systems. First, for each system, we ran multiple 10ns CMD simulations using initial structures with different RC distances selected from AMD or PSIM. In most of the simulations ligand would stay trapped where they were, while in some of them ligand would start to diffuse towards inside or outside the cavity due to its unstable initial position. Then two to three CMD simulations covering different segments of dissociation path were used to reconstruct the whole dissociation path, from which frames with desired RC distances and small root mean squire deviation (RMSD) values of ligand were selected as initial structures for US simulations. An example about reconstruction of dissociation path with CMD simulations for SB2 from p38α in DFG-in conformation is shown in Figure 3.2. Although part of the purpose of this US simulation

117

of DFG-out ligands is to reveal which pathway is more favorable for the ligand dissociation, we should note that the PMF we obtained from US here doesn't reflect the absolute binding free energy due to the fact that the biased MD with one dimensional restraint can't sample all conformations for both ligand and protein, and may miss essential states such as ligand moving along orthogonal direction to the predefined restraint and conformational change of protein when sampling dissociation path.
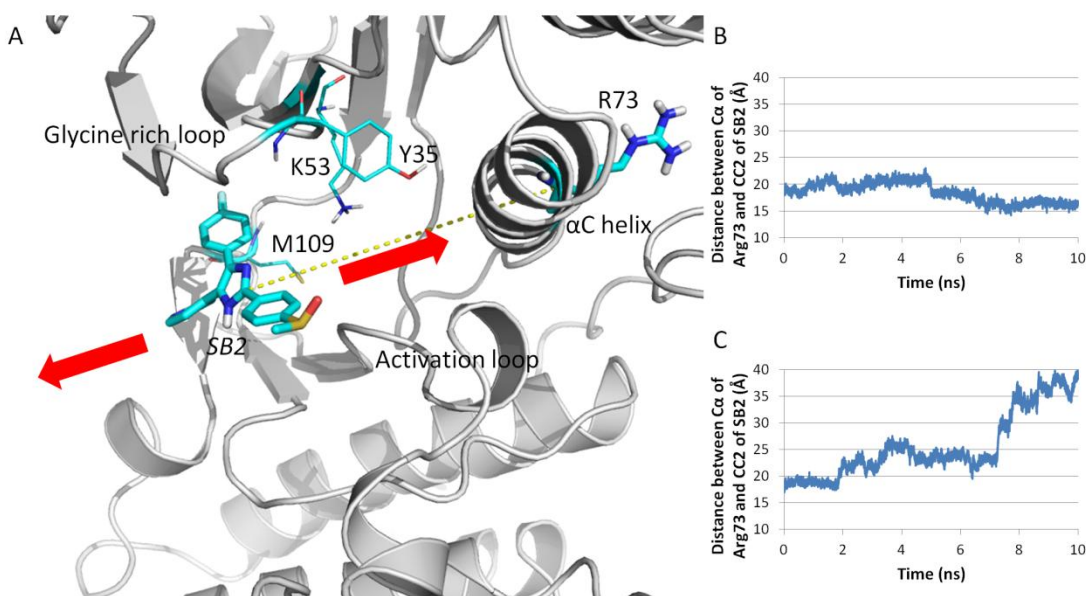
Figure 3.2. Reconstruction of dissociation path from AMD. Dissociation path of SB2 from p38α with a DFG-in conformation is rebuilt from two 10 ns CMD. (A) SB2 in one of the two 10 ns CMD moves towards inside the cavity, while SB2 in the other CMD moves towards outside. Arg73 and SB2 are shown in bold licorice structure, key interacting residues are shown in thin licorice structure. (B) SB2 moving inside the cavity indicated by the decreasing distance between SB2 and Arg73. (C) SB2 moving outside the cavity indicated by the increasing distance between SB2 and Arg73.

3.2.6 Post-MD analysis

Correlation between different parts of p38α was analyzed using the T-Analyst program [48]. First, we calculated correlations between backbone dihedral (Phi and Psi angles) of

each residue of protein, as well as rotatable dihedral angles of ligands (Table 3.1). Next, protein is divided into different parts based on the secondary structure of protein, and the absolute values of correlation of different parts were added up to get the correlation between different parts of protein, ie, the correlation between αC helix and phosphorylated loop.

3.3 Results and Discussions

Conventional molecular dynamics (CMD) simulation has been a well-established approach to study the dynamic behaviors of receptor-ligand systems. However, modeling ligand dissociation from its target protein using CMD can be impractical with current computation power. With k_{off} values ranging from 8.3×10^{-6} to 7.7 s^{-1} in our p38α systems, the dissociation time ranges from 0.1 s to $> 10^5$ s. As a result, we used AMD and a newly developed PSIM method to sample dissociation pathways.

To serve as our references, we ran 100 ns CMD for the free p38α with a DFG-in and DFG-out loop conformations, and performed correlation analysis in comparison with their ligand bound states. We also ran 650 ns CMD for the SB2 (bound with DFG-in loop) and BIRB796 complexes, and as expected, no ligand dissociation was observed.

We applied AMD simulations for five complex systems to study their dissociation pathways. However, only type-I ligands, SB2 and SK8, successfully dissociated in high boost AMD, while type-II ligand BIRB796 and type-III ligand LIG4 barely deviated from their binding position. We therefore applied PSIM, a pathway search method based

on internal PC modes, to sample dissociation pathways for BIRB796 and LIG4. For BIRB796 and LIG4, whose initial loop structures are in the DFG-out form, both allosteric and ATP pathways were sampled. After we got the complete dissociation pathways for all ligands, US were performed to illustrate the free energy profile along the dissociation processes and key interactions that contribute to the free energy changes during unbinding. We noticed that using positions of a ligand relative to p38α as a reaction coordinate in US missed important information from protein rearrangement when investigating binding kinetics. We analyzed the ligand-protein correlation during the dissociation process, and identified that protein hinge motion was a major movement as well. We therefore accessed the protein rearrangement guided by the protein hinge motion during ligand dissociation, and revealed significant differences between type-I and type-II/III ligands.

3.3.1 Dissociation pathways of type-I inhibitor, SB2 and SK8, sampled by AMD

To test the suitability and advantage of AMD, we first compared the results of CMD, low boost AMD and high boost AMD for SB2 (DFG-in). The average RMSD of heavy backbone atoms in trajectories relative to crystal structure are 2.05 for 100ns CMD, 3.05 for 100ns low boost AMD and 3.26 for high boost AMD, respectively. Figure 3.3 shows protein in low and high boost AMD didn't move too far from the crystal structure. Although in high boost AMD, the deviation increases significantly after SB2 dissociated around 64ns, indicating the role of ligand in stabilizing protein structure in AMD.

Figure 3.4 shows the projection of three trajectories (100ns CMD, 100ns low boost AMD and 100ns high boost AMD) on the first and second principal component vectors (PC1 and PC2) built from the Cα atoms of p38α in high boost AMD trajectory. In the CMD trajectory, PC1 and PC2 describe 23% and 12%, respectively, of the total variance of the motions in the simulation. It is clear from Figure 3.4A that CMD is trapped in the basin of crystal structure and never gets out. Figure 3.4B shows that the low boost AMD simulation does not explore the amount of conformational space that the high boost AMD simulation does and remains trapped around the crystal structure. The high boost AMD simulation (Figure 3.4C) exhibits a rather broad pathway from the crystallographic basin (-35, −20) to the region (15, 35) which represents SB2 outside binding cavity and is not present in the 100ns CMD or low boost AMD simulations (Figure 3.4A,B). In sum, high boost AMD can successfully simulate the dissociation of type-I ligand and maintain an overall stable structure through simulation.
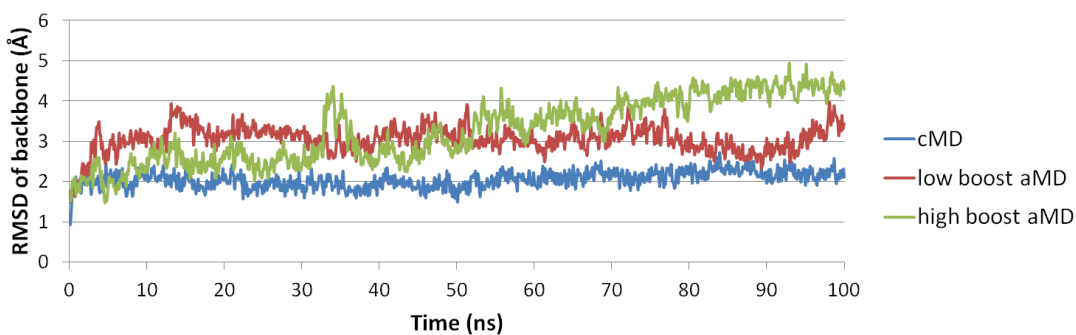
Figure 3.3. RMSD of heavy backbone atoms in trajectories of CMD, low boost AMD and high boost AMD relative to crystal structure.
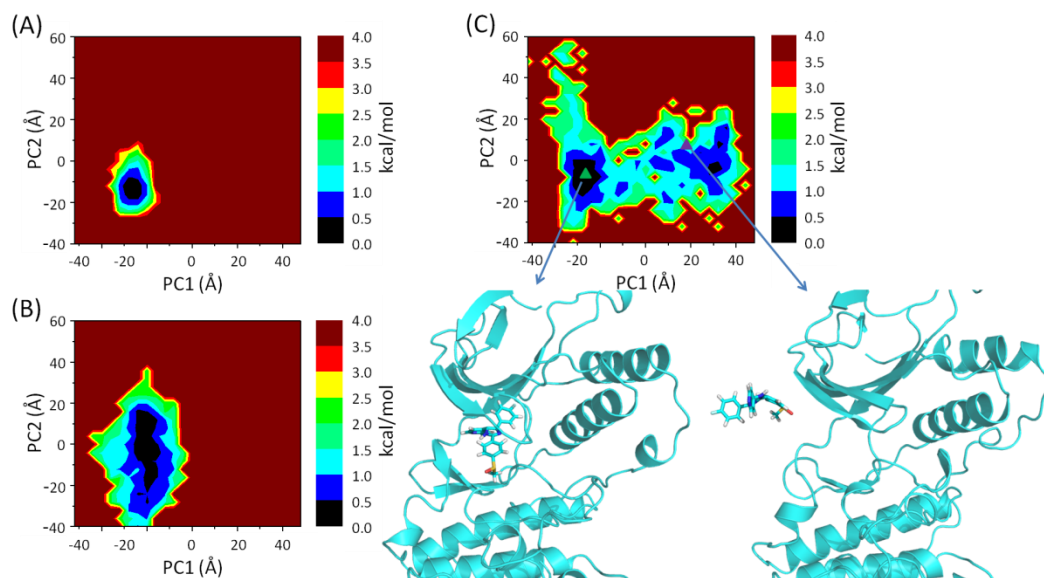
Figure 3.4. The free energy principal component projection of (A) 100ns CMD, (B) 100 ns low boost AMD, and (C) 100 ns high boost AMD onto (PC1, PC2) defined by the 100 ns high boost AMD. CMD and low boost AMD are trapped in the basin of crystal structure, while high boost AMD explores much larger conformational space.

Figure 3.5 illustrates key interactions between SB2 bound to a DFG-in loop conformation, where the attractions should be broken or loosen before or during the dissociation process. For example, there are two hydrogen bonds between the pyridine ring N and backbone nitrogen of Met109, and the N3 atom of the imidazole ring and Lys53, and the phenyl ring of 4-methylsulfinylphenyl group forms a stacking interaction with Tyr35 (Figure 3.3). In addition to these key interactions, SB2 bound to DFG-out loop conformation has an additional stacking interaction between phenyl ring and Phe169 of DFG-motif.
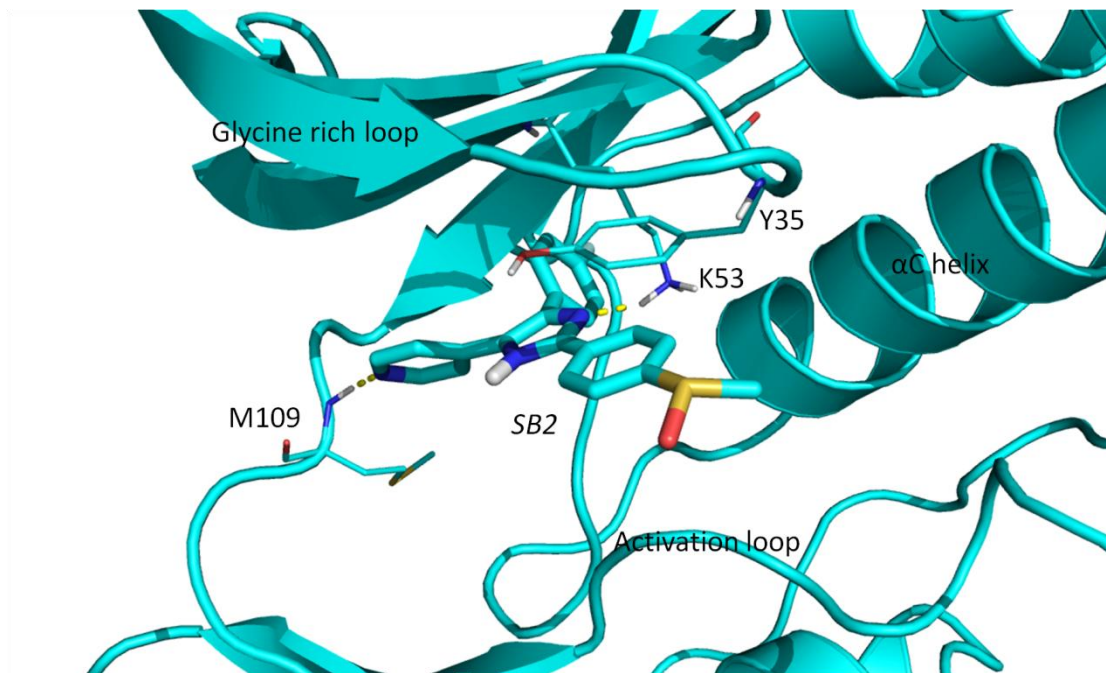
Figure 3.5. Interactions between SB2 and p38α in DFG-in conformation from crystal structure (PDB 1A9U). SB2 is shown in bold licorice structure. Key interacting residues are shown in thin licorice structure. Hydrogen bonds between SB2 and p38α are shown in dash line.

Using the dissociation pathways sampled by AMD, US was performed along the pathway to construct a PMF to further explore free energy change associated ligand unbinding. Figure 3.6A illustrates that first the hydrogen bond between SB2 and Lys52 breaks, followed by the motion of 4-methylsulfinylphenyl group (Figure 3.6B). The second hydrogen bond between pyridine nitrogen and Met109 then breaks (Figure 3.6C) and finally, ligand is outside the edge of binding cavity (Figure 3.6D) and eventually diffuses away. The same as SB2 unbinding from p38α with DFG-in conformation, SK8 unbinding and SB2 unbinding from p38 complex with DFG-out conformation were both along the ATP channel. Their PMF plots are also similar (Figure 3.7, 3.8) to Figure 3.6. The free

energy continues rising during dissociation, but no significant energy barriers were observed along the unbinding process.

Figure 3.6. PMF of dissociation process of SB2 (DFG-in) and selected snapshots from US. SB2 is shown in bold licorice structure. Key interacting residues are shown in thin licorice structure. Hydrogen bonds between SB2 and p38α are shown in dash line.
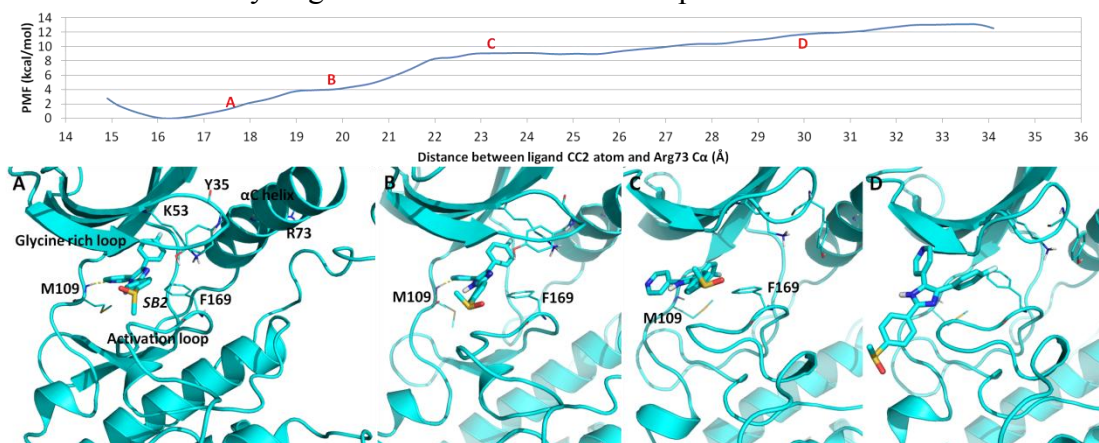
Figure 3.7. PMF of dissociation process of SB2 (DFG-out) and the selected snapshots from US. SB2 is shown in bold licorice structure, key interacting residues are shown in thin licorice structure, hydrogen bonds between SB2 and p38α are shown in dash line. (A) SB2 breaks hydrogen bond with Lys53 side-chain and stacking interaction with rotation of Tyr35. (B) 4-methylsulfinylphenyl group of SB2 diffuses towards outside the cavity, fluorophenyl ring of SB2 moves out of the hydrophobic pocket and forms stacking interaction with Phe169. (C) SB2 breaks hydrogen bond with Met109 stacking interaction with Phe169. (D) SB2 is outside the binding cavity.

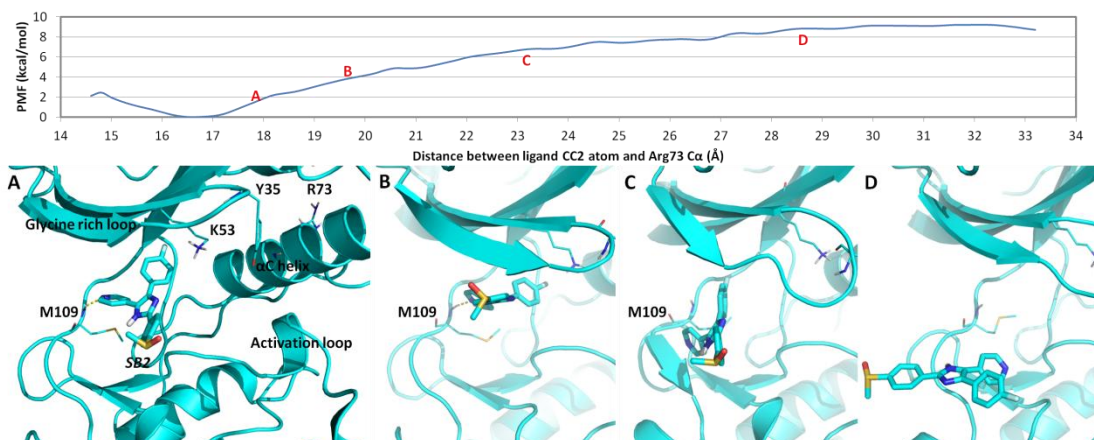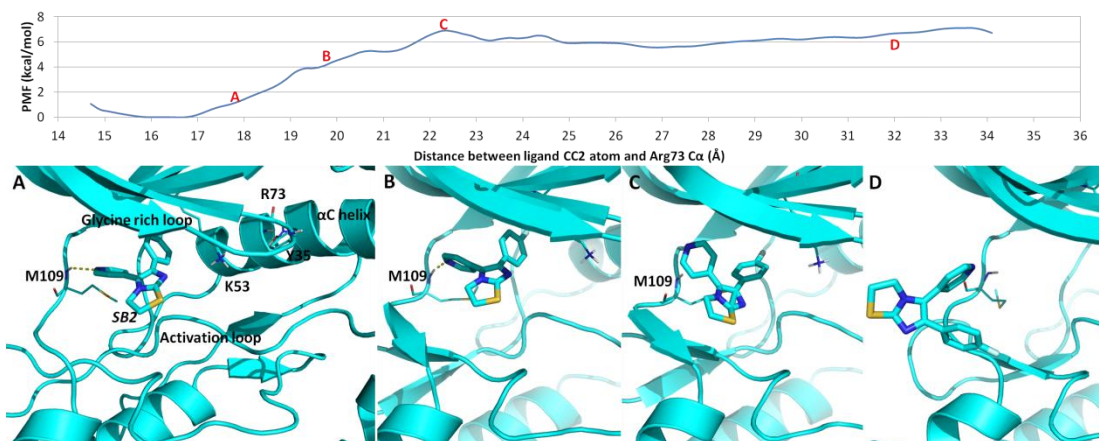Figure 3.8. PMF of dissociation process of SK8 and selected snapshots from US. SK8 is shown in bold licorice structure, key interacting residues are shown in thin licorice structure, hydrogen bonds between SK8 and p38α are shown in dash line. (A) SK8 breaks hydrogen bond with Lys53 side-chain and stacking interaction with rotation of Tyr35. (B) Fluorophenyl ring of SK8 moves out of the hydrophobic pocket. (C) SK8 breaks hydrogen bond with Met109. (D) SB2 is outside the binding cavity.

Previous researches show that water effects can be important in ligand binding kinetics [49, 50]. Therefore, we investigated if there were bridge water molecules with long residence time and also counted the number of pocket-water molecules with within 5 Å of residues in the protein pocket during SB2 unbinding from p38α with a DFG-in conformation. Figure 3.9 shows a rapid increase of number of pocket-waters when p38α moved to a position shown in Figure 3.6B, where the motion of 4-methylsulfinylphenyl group created space for re-solvating water molecules. However, the number of pocket-waters fluctuated considerably. Unlike ligands binding HIV protease that a few transient water molecules stayed very long between the ligand and protein [51], all the water molecules were replaced by each other frequently. Therefore, it is unlikely that removing bridge waters govern unbinding kinetics.
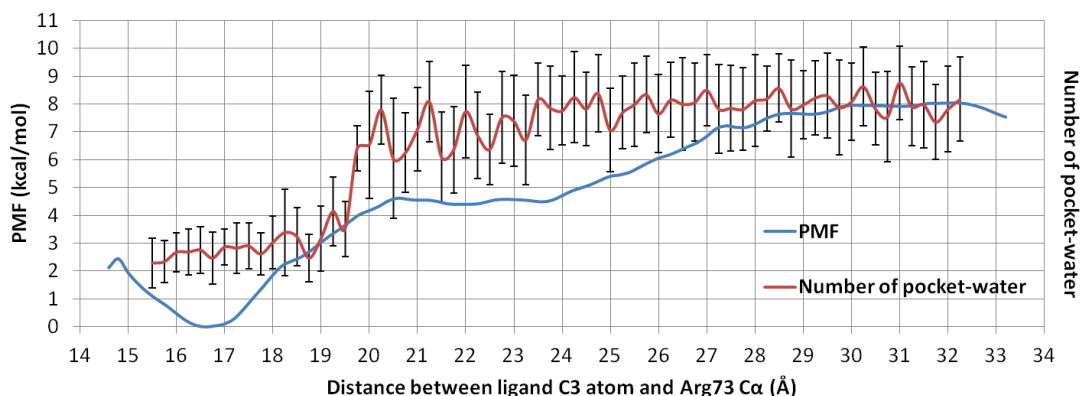
Figure 3.9. PMF of SB2 (DFG-in) dissociation (blue curve) and profile of number of pocket-water (red curve) as a function of the RC distance, standard deviation of number of pocket-water is used as error bar.

3.3.2 Dissociation pathways of type-II inhibitor, BIRB796 and type-III inhibitor, LIG4, sampled by PSIM

Although high boost AMD successfully simulated dissociation pathways for ligands SB2 and SK8 inhibitors, the method could not sample ligand dissociation for type II or type III inhibitors. Experimentally, the ligands have much longer residence time, 625 s for LIG4 and 1.2×10^5 s for BIRB796, respectively. Therefore, we applied the newly developed method PSIM to sample the dissociation pathways for BIRB796 and LIG4.

Complex conformations from CMD and crystal structures for BIRB796 and LIG4 complexes with p38α show that the urea forms two hydrogen bonds between urea NH group and Glu71 in N lobe and between urea CO group and Asp168 backbone nitrogen in C lobe (Figure 3.10). The two hydrogen bonds clamp lignads BIRB796 and LIG4 within the cleft. In addition, extended morpholino substituent forms a hydrogen bond with backbone nitrogen of Met109. To dissociate a ligand, trajectories obtained from PSIM

showed that the cleft needed to open, which also correlate with the protein hinge motion between the two lobes. The opening allows BIRB796 and LIG4 to unbind directly from the allosteric pocket, or to move further towards ATP binding site and dissociate from there. The latter pathways is similar to the unbinding pathways sampled for type-I ligands.
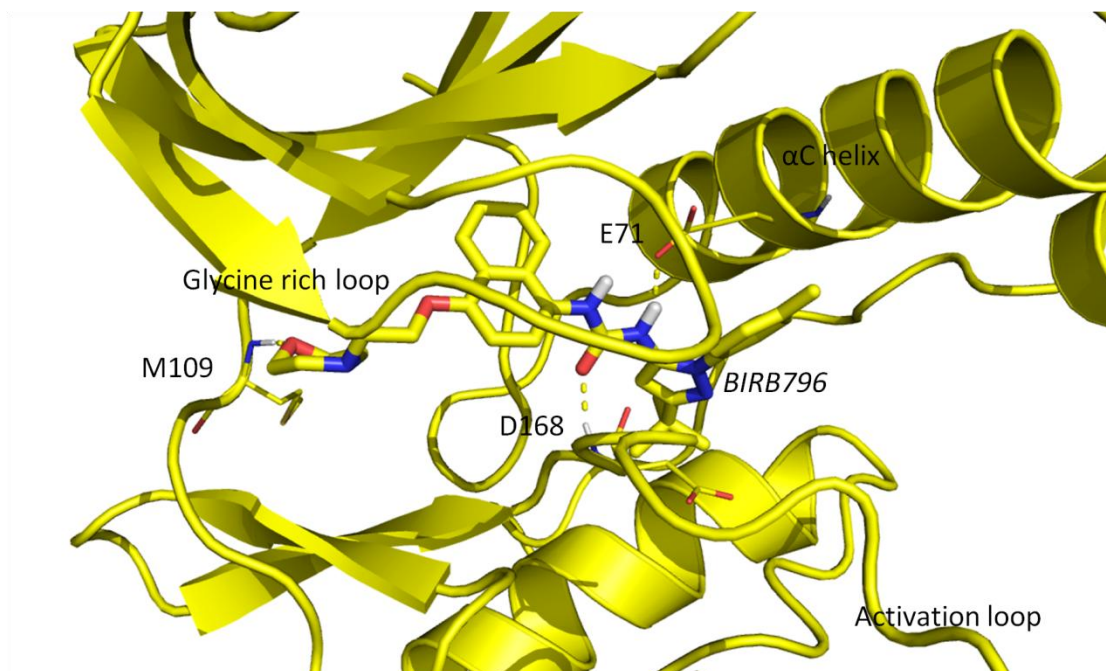
Figure 3.10. Interactions between BIRB796 and p38α in DFG-out conformation from crystal structure (PDB 1KV2). BIRB796 is shown in bold licorice structure. Key interacting residues are shown in thin licorice structure. Hydrogen bonds between BIRB796 and p38α are shown in dash line.

First we examine the PMF constructed from the BIRB796 unbinding from the allosteric pocket. Figure 3.11A shows that two hydrogen bonds between urea group of BIRB796 and Glu71 and Asp168 need to break, and the extended morpholino substituent rotates along the opened cleft, breaking another hydrogen bond with Met109. The free energy increases ~2 kcal/mol. The free energy continues to rise as BIRB796 wiggles out along the activation loop to the edge of binding cavity (Figure 3.11B), where His174 forms

stacking interaction with naphthalene group of BIRB796. The activation loop then rearranges and fluctuates in concert with the ligand to unbind the ligand, resulting in dropping free energy from 5.5 kcal/mol to 3.4 kcal/mol (Figure 3.11C). BIRB796 finally breaks the stacking interaction with His174 and moves away as shown in Figure 3.11D.

Figure 3.11. PMF of dissociation process of BIRB796 along allosteric pathway and selected snapshots from US. BIRB796 is shown in bold licorice structure. Key interacting residues are shown in thin licorice structure.

For BIRB796 dissociating from the ATP pocket, the dissociation of BIRB796 also starts with cleft opening, extended morpholino substituent rotating, accompanied with hydrogen bonds breaking. The large 5-tert-Butyl-2-p-tolyl-2H-pyrazol group rotates and moves toward the ATP pocket, while Phe169 starts to form stacking interaction with the naphthalene group of BIRB796 (Figure 3.12A). This path is not energetically favorable, and the free energy barrier rapidly increases to 10 kcal/mol. Then BIRB796 starts to dissociate from the ATP pocket similar to other type-I ligands. The free energy continues to increase to 16.8 kcal/mol until the stacking interaction between BIRB796 and Phe169 is loosen (Figure 3.12B). The glycine rich loop needs to slightly lift up to create room for

128

BIRB796 to continue unbinding from the cavity (Figure 3.12C). Once BIRB796 moves

out of cavity, the glycine rich loop returns back to its original position (Figure 3.12D).

Similar changes of interactions shown in the PMF plots for LIG4 are pointed in Figure

3.13, 3.14.
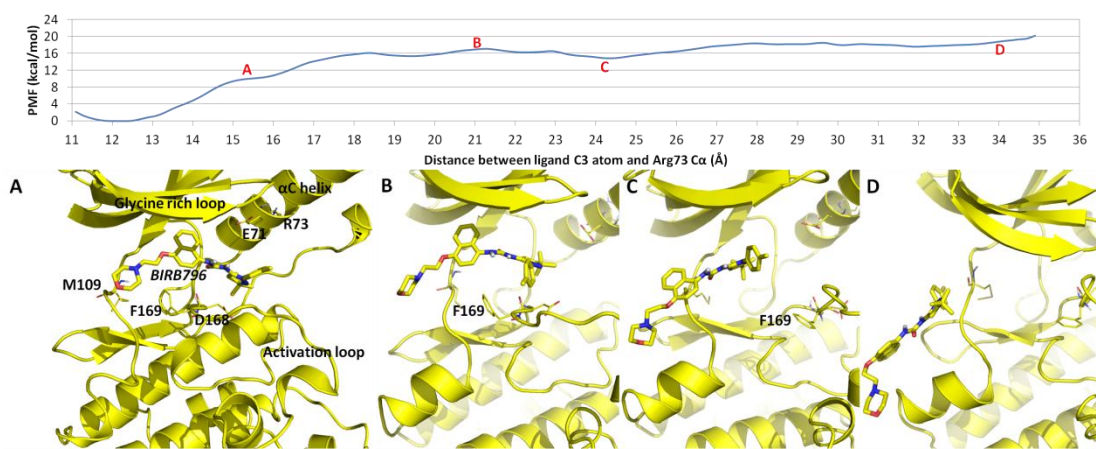
Figure 3.12. PMF of dissociation process of BIRB796 along ATP pathway and selected snapshots from US. BIRB796 is shown in bold licorice structure. Key interacting residues are shown in thin licorice structure.
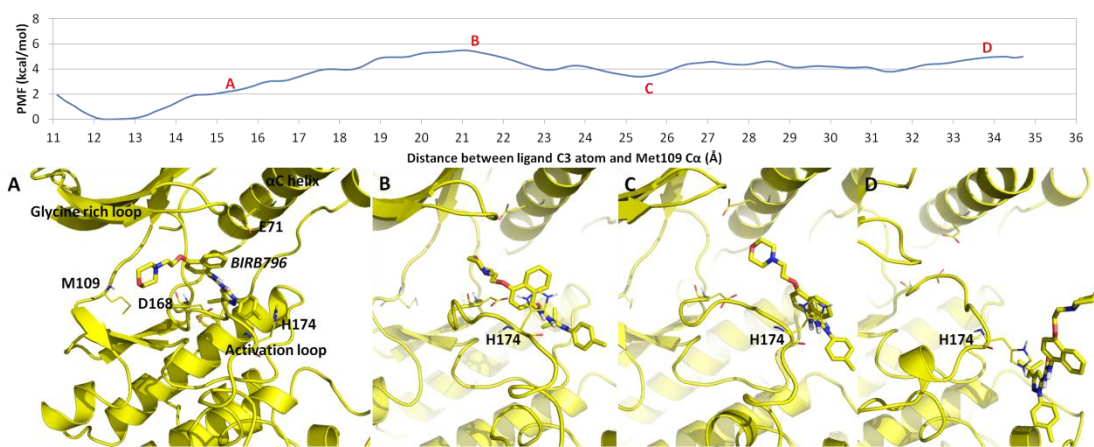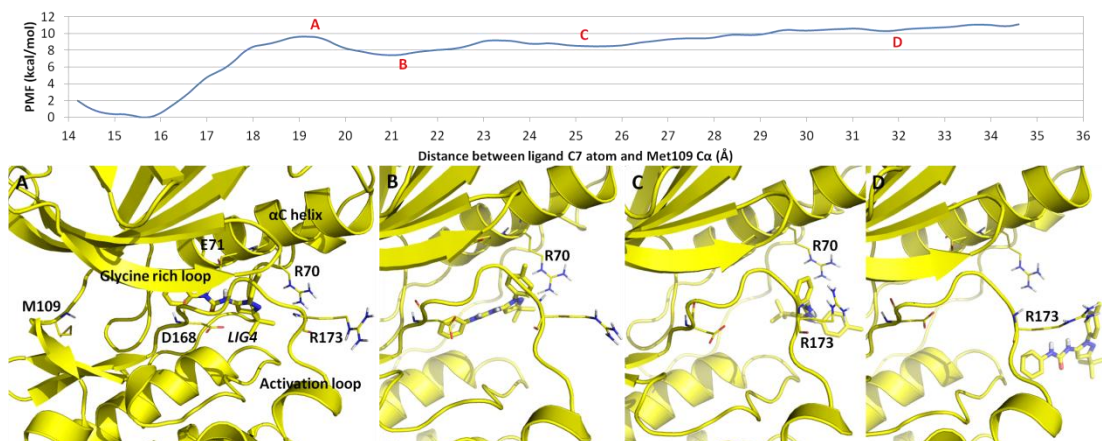
Figure 3.13. PMF of dissociation process of LIG4 along allosteric pathway and selected snapshots from US. LIG4 is shown in bold licorice structure, key interacting residues are shown in thin licorice structure, hydrogen bonds between LIG4 and p38α are shown in dash line. (A) Cleft opens up, breaking hydrogen bonds between urea group of LIG4 and Glu71 and Asp168. (B) LIG4 diffuses out, forms hydrogen bond with Arg70. (C) LIG4 breaks hydrogen bond with Arg70, forms ring-ion stacking interaction with Arg173. (D) LIG4 breaks stacking interaction with Arg173 and diffuses away.
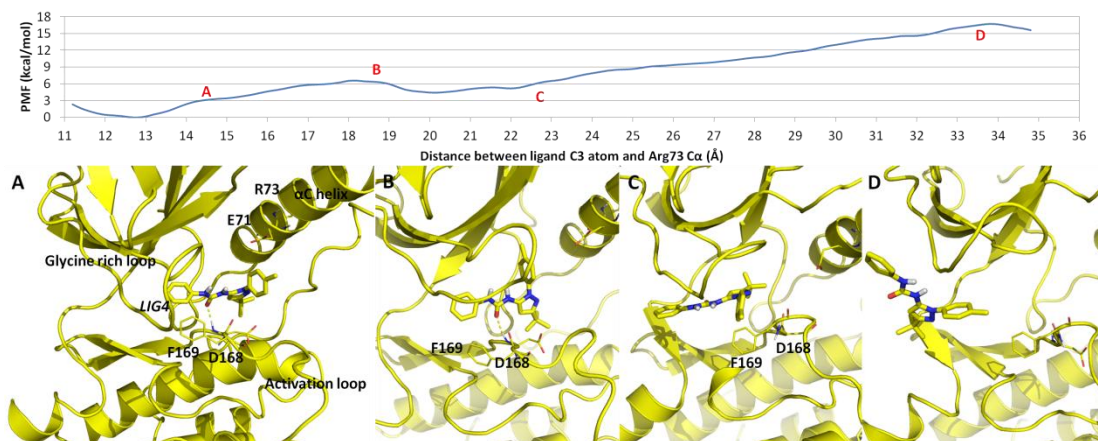
Figure 3.14. PMF of dissociation process of LIG4 along ATP pathway and selected snapshots from US. LIG4 is shown in bold licorice structure, key interacting residues are shown in thin licorice structure, hydrogen bonds between LIG4 and p38α are shown in dash line. (A) Cleft opens up, breaking hydrogen bonds between urea group of LIG4 and Glu71. LIG4 forms stacking interaction between phenyl group and Phe169. (B) LIG4 enters ATP binding site. (C) LIG4 breaks hydrogen bond with Asp168, Phe169 switches to form stacking interaction with toluene moiety. (D) LIG4 breaks stacking interaction with Phe169 and diffuses away.

130

We also examined the changes of numbers of pocket-water molecules during BIRB796 dissociation. Figure 3.15 shows a rapid increase of number of pocket-waters when the cleft between the N and C-lobes opens (Figures 5A and 6A). During ligand dissociation processes, the number of pocket-water plateaus and in average seven water molecules are in the pocket, similar to that during SB2 dissociation. The same as fast binding molecules SB2 and SK8, numbers of pocket-waters largely fluctuate and the water molecules were constantly replaced by each other. As a result, the ligand-water interactions and re-solvation processes are less likely to contribute significantly to slow BIRB796 unbinding.
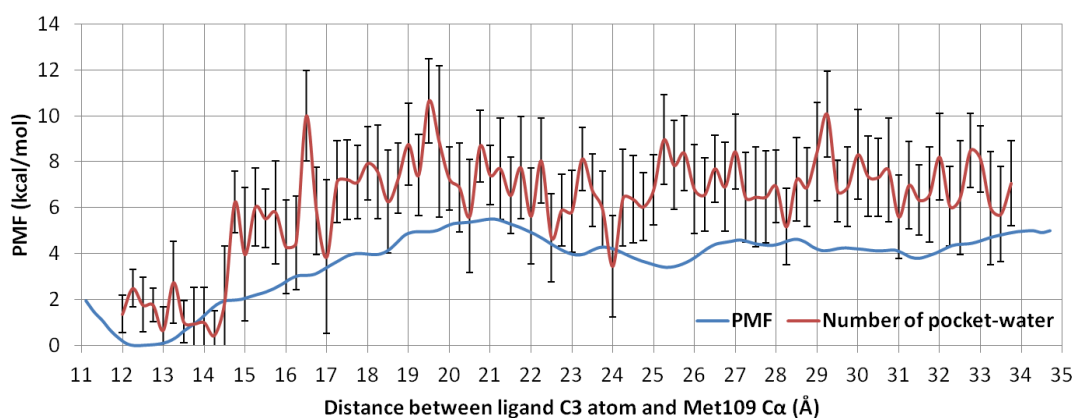
Figure 3.15. PMF of BIRB796 dissociation (blue curve) and profile of number of pocket-water (red curve) as a function of the RC distance, standard deviation of number of pocket-water is used as error bar.

Although the absolute binding free energies in one-dimensional PMF cannot accurately reproduce the ligand-p38 binding free energy (ΔG) and the barriers, PMF plots can distinguish that BIB796 and LIG4 prefer dissociating from the allosteric pocket, instead of moving to the ATP pocket and dissociating from there. Notably, the PMF plots only consider one chosen degree of freedom, which inevitably simplified and smooth out the

free energy barriers. Because we observed noticeable protein motions during ligand dissociation, we carried out further analysis to reveal the protein dynamics and used the information for another coordinate to investigate the unbinding free energy barriers.

3.3.3 PCA and Correlation Analysis

To get the major motion of protein, we first performed PCA to check the first principle component motion of both DFG-in and DFG-out proteins (Figure 3.16). It turns out that in free DFG-out protein, the most principle motion is the hinge motion between N and C lobes is mainly between activation loop and αC helix, which are on the sides of allosteric path. In free DFG-in protein, the hinge motion is mainly between glycine-rich loop (β1, L4, β2) and L9/αD helix, which are on the way of ATP path. Both PCA results suggest that hinge movement may be important in ligand dissociation.
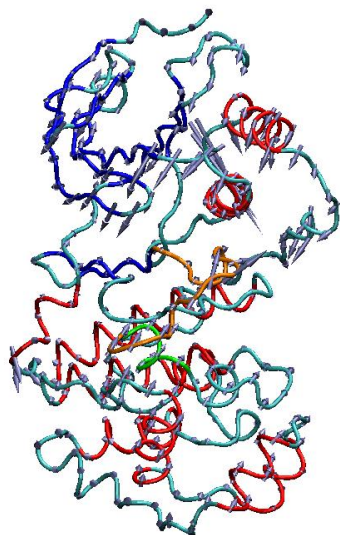
Free DFG-in   Free DFG-out
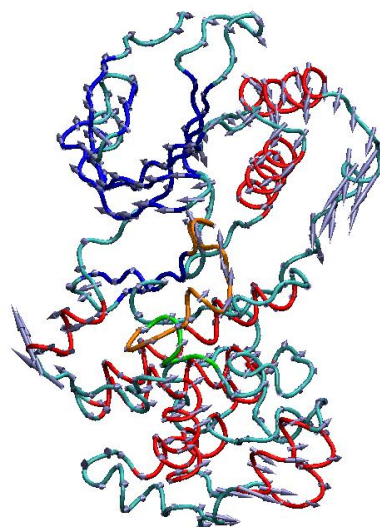
Figure 3.16. The first PC modes of free DFG-in and DFG-out proteins from CMD simulations. The gray arrows indicate the local direction and magnitude of movement. α-helix is colored in red, β-sheet is colored in blue, loop is colored in cyan, activation loop is colored in orange, P+1 substrate site is colored in green.

To better understand the protein motion during ligand dissociation, we studied the correlation between different parts of p38α complexes. Protein is divided into different fragments based on its secondary structure (Figure 3.17), while ligand is considered as one fragment, and correlations between different fragments are investigated. We first measured the correlations within CMD trajectories of both free proteins and complexes (Table 3.2).
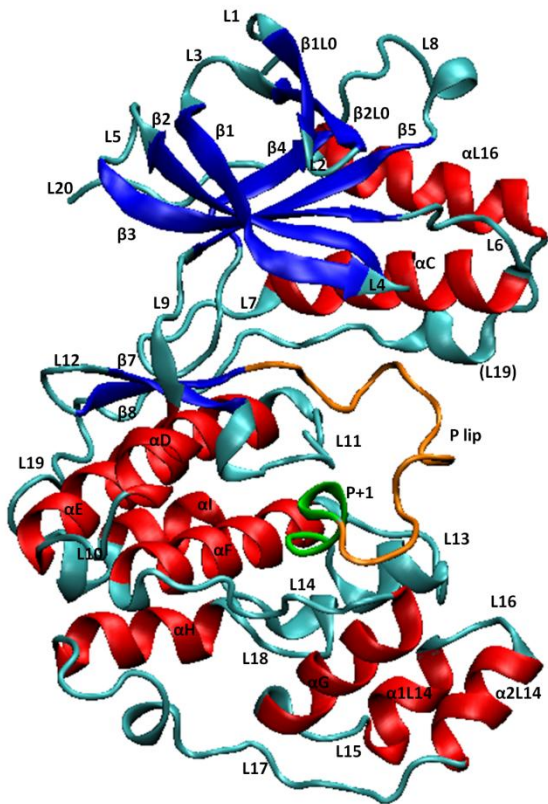
Figure 3.17. Division of p38α protein into different fragments basing on the secondary structure of protein. α-helix is colored in red, β-sheet is colored in blue, loop is colored in cyan, activation loop is colored in orange, P+1 substrate site is colored in green.

Both free DFG-in and DFG-out protein have strong correlation between hinge region (L9 loop, αD helix, L10 loop) and two lobes (N and C lobes), while free DFG-out protein has stronger correlation (Figure 3.18). The correlation results of free proteins agree with RMSF of Cα of protein in CMD trajectories, where free DFG-out protein has larger fluctuations in hinge region than free DFG-in protein (Figure 3.19).
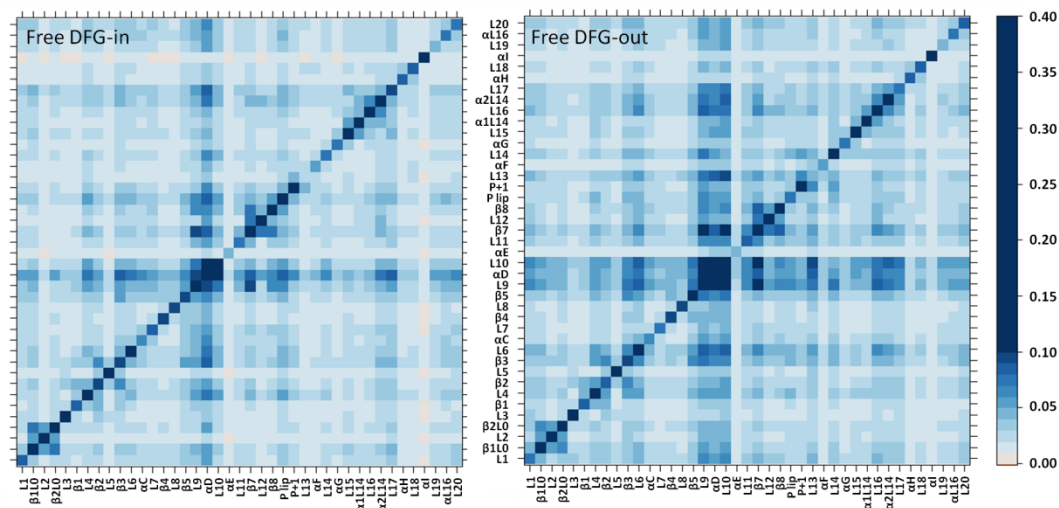
Figure 3.18. Correlation maps of free DFG-in and DFG-out proteins from CMD simulations.
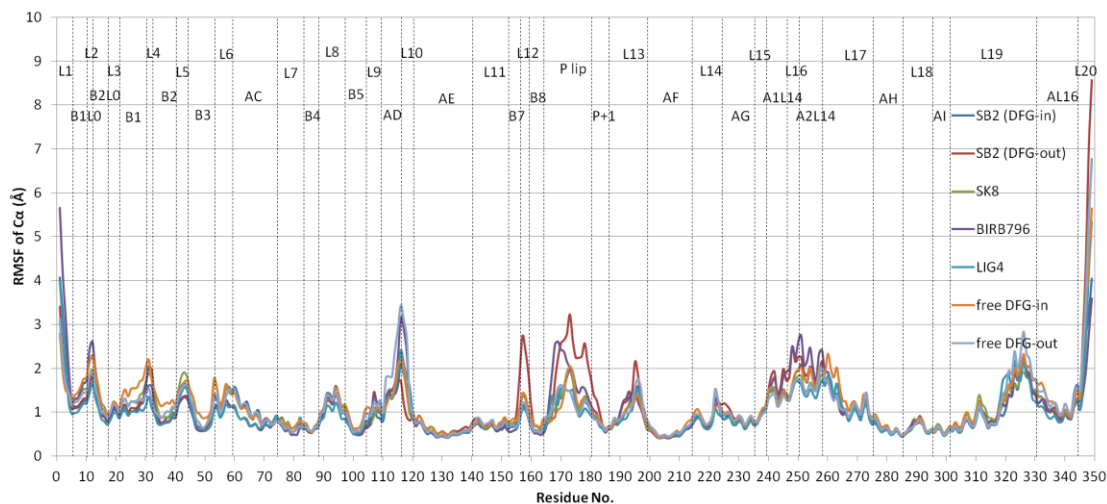
Figure 3.19. RMSF of Cα of p38α complexes and free proteins from CMD simulations. Residues are divided into different different fragments basing on the secondary structure of protein. For example, L1 indicated Loop 1, B1L0 indicates β-sheet 1L0, AC indicates α-helix C, P lip indicates activation loop, P+1 indicated P+1 substrate site.

Comparing to free protein, correlations of complexes retain strong correlation in hinge region, while the correlation between ligand and protein are mostly not obvious due to the limited power of conformational sampling using CMD (Figure 3.20). Even though

135

ligands stays steadily in energy basin of crystal bound position, there is still correlation between SB2 (both DFG-in and DFG-out) and the hinge and activation loop (P loop and P+1 loop). SB2 (DFG-out) has stronger correlation with hinge region and activation loop, especially with the activation loop because of the extra stacking interaction between the 4-methylsulfinylphenyl group of SB2 and Phe169, which is also reflected in large fluctuation in activation loop during CMD simulation. BIRB796 and LIG4 don't show any strong correlation between ligand and protein due to the grip of two hydrogen bonds from N and C lobes. Interestingly, in BIRB796 complex, there is strong correlation between L16 loop/α2L14 helix region and other part of protein, which confirms the concerted motion between L16 loop/α2L14 helix region and hinge region, as seen in PCA motion in Figure 3.16. It is possible that L16 loop/α2L14 helix region plays a role in helping the hinge region move.
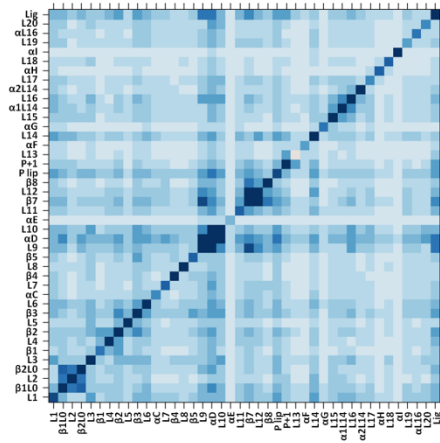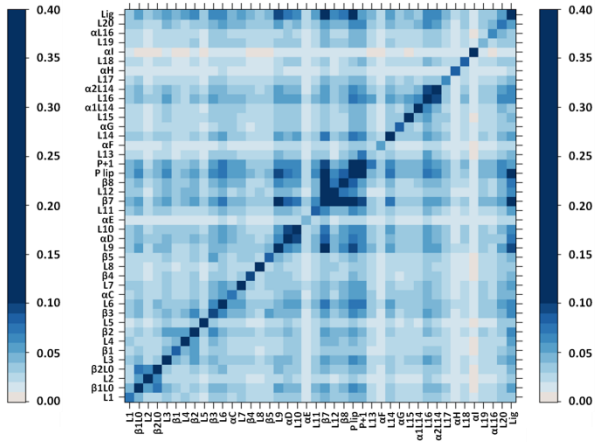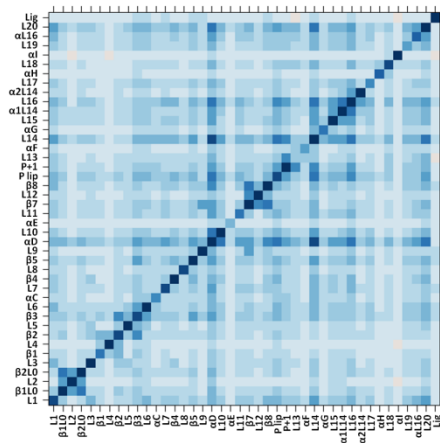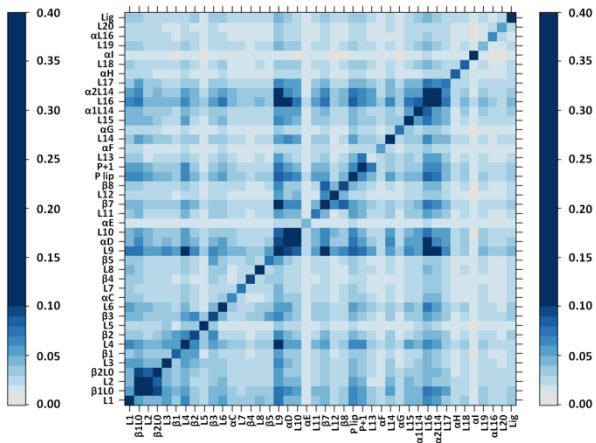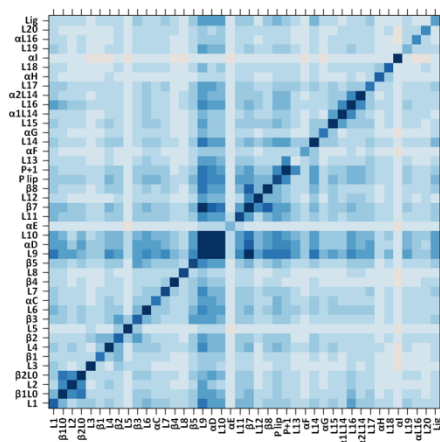
Figure 3.20. Correlation maps of p38α complexes from CMD simulations.

The differences in protein conformational correlation are more substantial while SB2 and BIRB796 are unbinding from p38 with DFG-in, and DFG-out conformations, respectively (Table 3.2, Figure 3.21). While SB2 is in the middle of dissociation pathway, p38α and SB2 mostly correlates solely within hinge region and the activation loop. In contrast, during the dissociation of BIRB796, we see strong correlation around hinge region and activation loop, and also their strong correlation with the upper arm of hinge in N lobe (L3 loop, β2 sheet and β3 sheet), suggesting p38 encounters more sizable rearrangement during BIRB796 unbinding.
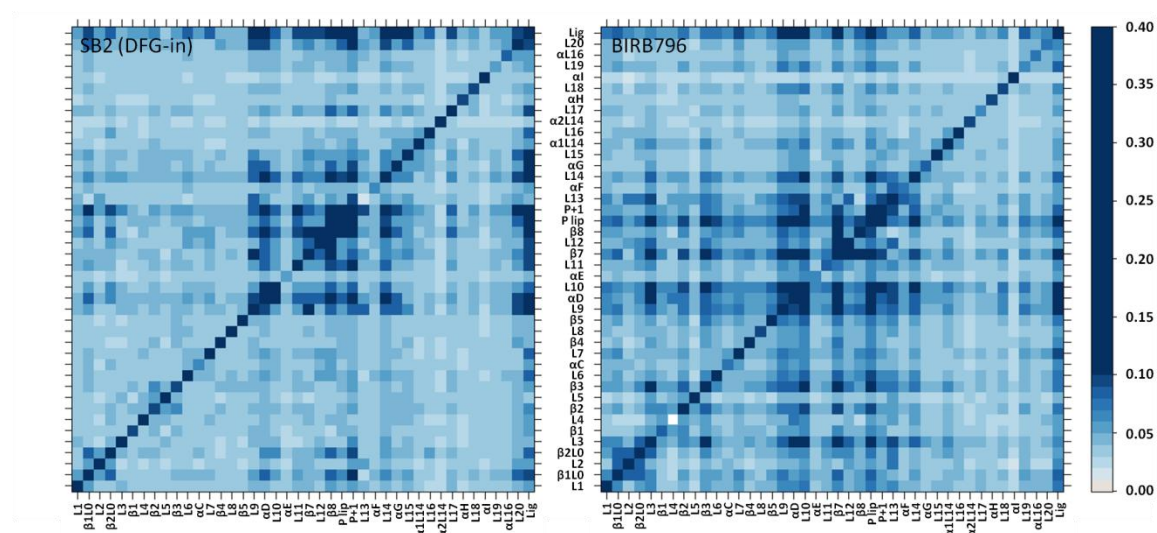
Figure 3.21. Correlation maps of SB2 (bound to DFG-in conformation) and BIRB796 p38α complexes during dissociation process.

3.3.4 The role of protein hinge motion in dissociation pathways

To quantify the contribution of protein rearrangement to free energy barrier, we select another coordinate based on PCA and correlation maps to present protein motion to construct another 1-D PMF. Notably, although multi-dimensional US can be applied to

construct PMF plots, we found that it is difficult to view and examine free energy barriers versus protein rearrangement using the 2-D PMF plots. For BIRB796 and LIG4 complex with a DFG-out conformation of p38α, the distance between Cα of Glu71 and Asp168 is used as a coordinate to represent hinge motions, and the two residues also locate on the ligand dissociation pathways. For type-I inhibitors, the distance between Cα of Val30 and Ala111 is used as a coordinate for hinge motions, and both residues are also in the unbinding pathways when type-I ligands are unbinding from the ATP binding site with a DFG-in protein conformation (Figure 3.22).
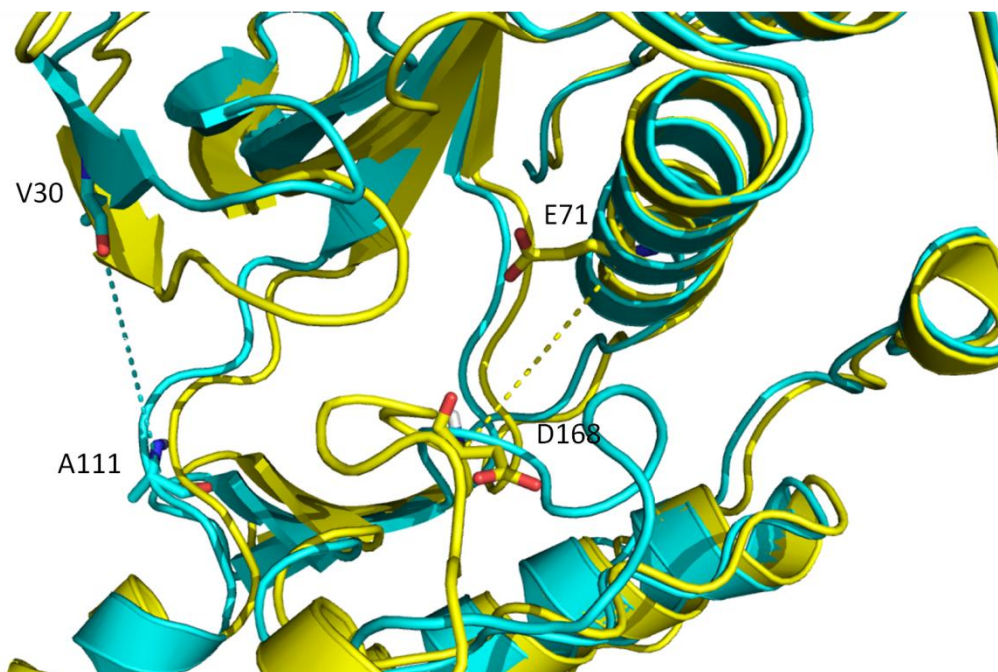
Figure 3.22. Distance used as RC for hinge motions. Hinge distances are indicated by distance between Cα of Glu71 and Asp168, which is on the allosteric path of type-II/III ligands (yellow), and distance between Cα of Val30 and Ala111, which is on the ATP path of type-I ligand (cyan).

Figure 3.23 shows the free energy change along the hinge movement while BIRB796 is located in the crystal structure bound form, middle of dissociation pathway, right outside

the binding cavity and completely dissociate from p38α with DFG-out conformation. When a ligand is far from the protein binding site, the protein motion can be presented by a shallow energy well with the equilibrium position located on 11.1 Å of the coordinate (hinge distance between Glu71 and Asp168). Interestingly, when BIRB796 is in the binding cavity, the equilibrium position shifts to 10.1 Å and the free energy well narrows. Within thermal fluctuation RT (~0.6 kcal/mol) the hinge distance can move lesser than 1 Å, resulting in much more rigid complex structure. As BIRB796 dissociates in a position shown in Figure 3.11A with the opening cleft, the equilibrium hinge distance increases to 11.9 Å. It needs 2.6 kcal/mol to widen the hinge distance from 10.1 Å to 11.9 Å. The large energy barrier from protein rearrangement explains the slow dissociation rate of BIRB796, and has been reported recently for slow ligand binding to tyrosine kinases due to the induced fit/protein rearrangement [52]. Notably, when BIRB796 locates just outside the cavity, the equilibrium position of the hinge distance is the same as that when BIRB796 is far from p38α. However, the existence of a ligand near the binding pocket perturbs p38α fluctuation of N and C-lobes, resulting in narrowing the energy well for the hinge motion. Similar to BIRB796, LIG4 requires remarkable cleft opening and rises energy barriers for ligand dissociate (Figure 3.24).

Figure 3.23. Free energy change along hinge movement of p38α in DFG-out conformation at different stages of dissociation of BIRB796. Distance between Cα of Glu71 and Asp168 is used as RC.

Figure 3.24. Free energy change along hinge movement of p38α in DFG-out conformation at different stages of dissociation of LIG4. Distance between Cα of Glu71 and Asp168 is used as RC.

For type-I ligand SB2 (bound with DFG-in loop), the equilibrium hinge distances are highly similar, regardless of where SB2 locates (Figure 3.25), suggesting that the protein hinge movement contributes insignificantly to the dissociation energy barrier. Interesting, the bound SB2 slightly rigidifies p38α movement and results in a narrower energy well. However, existing of SB2 in the pocket (Figure 3.6A) does not disturb protein fluctuation, and p38 can fluctuate as if no SB2 present, suggesting that protein rearrangement does not contribute to the unbinding free energy barriers. The shallow energy well also suggests that it is easier for SB2 (DFG-in) to move out of the cleft. Similar to SB2 (DFG-

141

in), during dissociation of SB2 (DFG-out) and SK8, the equilibrium hinge distance doesn't change much (Figure 3.26). Comparing the dissociation of SB2 (DFG-in) to SB2 (DFG-out), it's noted that the average equilibrium hinge distance shifts from 12 Å in SB2 (DFG-in) to 15 Å in SB2 (DFG-out), it's possible that cleft needs to open up as DFG motif flips from DFG-in to DFG-out conformation.

Figure 3.25. Free energy change along hinge movement of p38α in DFG-in conformation at different stages of dissociation of SB2. Distance between Cα of Val30 and Ala111 is used as RC.
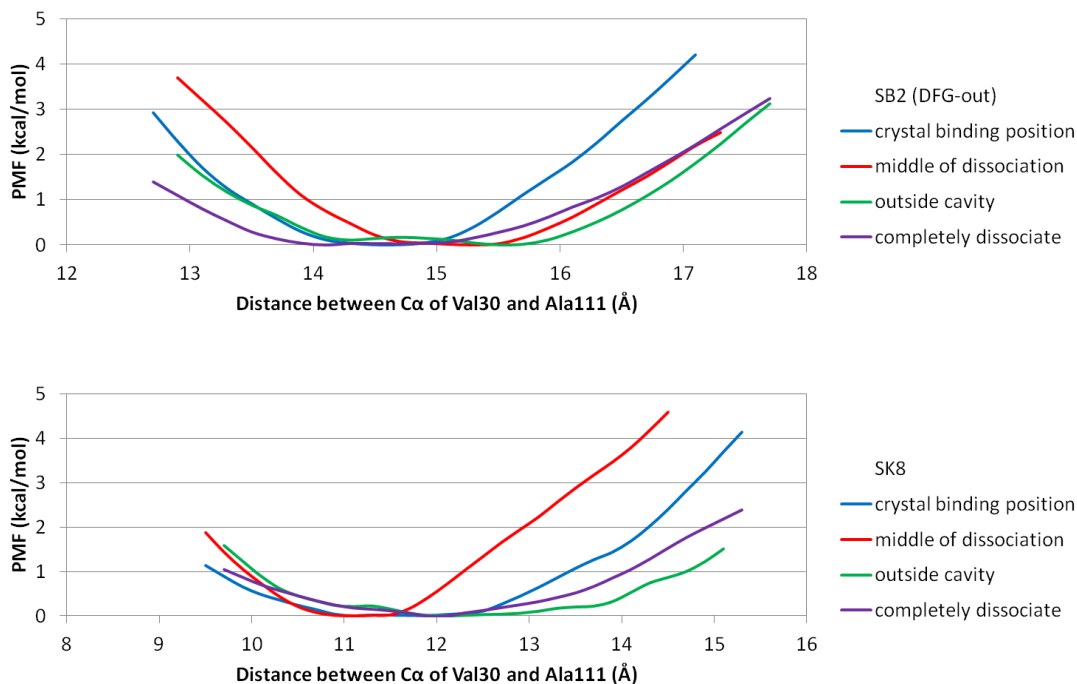
Figure 3.26. Free energy change along hinge movement of p38α at different stages of dissociation of SB2 (DFG-out) and SK8. Distance between Cα of Val30 and Ala111 is used as RC.

To make sure the results of free energy calculation along hinge movement don't depend on the predefined RC, we switched the RCs for type-I and type-II/III ligands. Free energy change along hinge movement using distance between Cα of Val30 and Ala111 as RC for BIRB796 confirmed the necessity of a hinge opening movement for its dissociation. Free energy change along hinge movement using distance between Cα of Glu71 and Asp168 as RC for SB2 (DFG-in) confirmed that its dissociation is independent from protein conformational rearrangement (Figure 3.27).

143
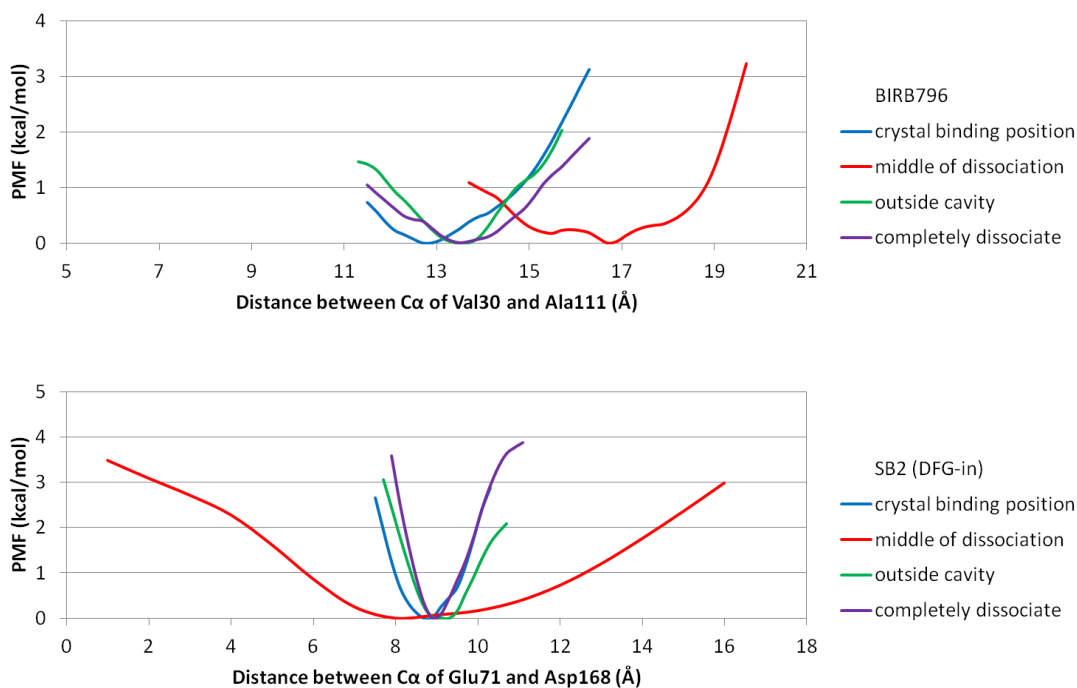
Figure 3.27. Free energy change along hinge movement of p38α at different stages of dissociation of BIRB796 and SB2 (DFG-in). Distance between Cα of Val30 and Ala111 is used as RC for BIRB796. Distance between Cα of Glu71 and Asp168 is used as RC for SB2 (DFG-in).

Since protein conformational rearrangement plays a crucial role in dissociation of type-II/III lignads, a mutation associated with the flexibility of hinge region may alter efficacy of inhibitors or activity of p38α. A more flexible hinge region can lead to faster dissociation of type-II/III ligands. To confirm that, we performed 100ns CMD simulation for BIRB796 bound with p38α in DFG-out conformation with mutated residues to make the hinge region more flexible for protein. We mutated Tyr69 in αC helix, Phe327 in L16 loop and Trp337 in L16 helix into glycine. Before mutation, Tyr69, Phe327 and Trp337 form stacking interaction with each other and can mutually stabilize L16 loop, αL16 helix and αC helix. After mutation, L16 loop is much more flexible without stacking interaction, both αL16 helix and αC helix shift up. Even though the average hinge

144

distance doesn't change much due to the concerted up shifting of activation loop (Figure 3.28). The average RMSD of heavy backbone atoms of αC helix are 1.0 for native CMD and 1.7 for mutated one, indicating the increased flexibility of the upper arm of hinge. Comparing free energy change along hinge movement of protein upon mutation, it's clear that the energy wells generally become shallower after mutation, especially when BIRB796 is in the crystal structure bound form and middle of dissociation pathway, whose equilibrium hinge distances move towards each other, ensuring a smooth transition of hinge movement during dissociation process (Figure 3.29), which may provide a new direction for future kinase study and inhibitor design.
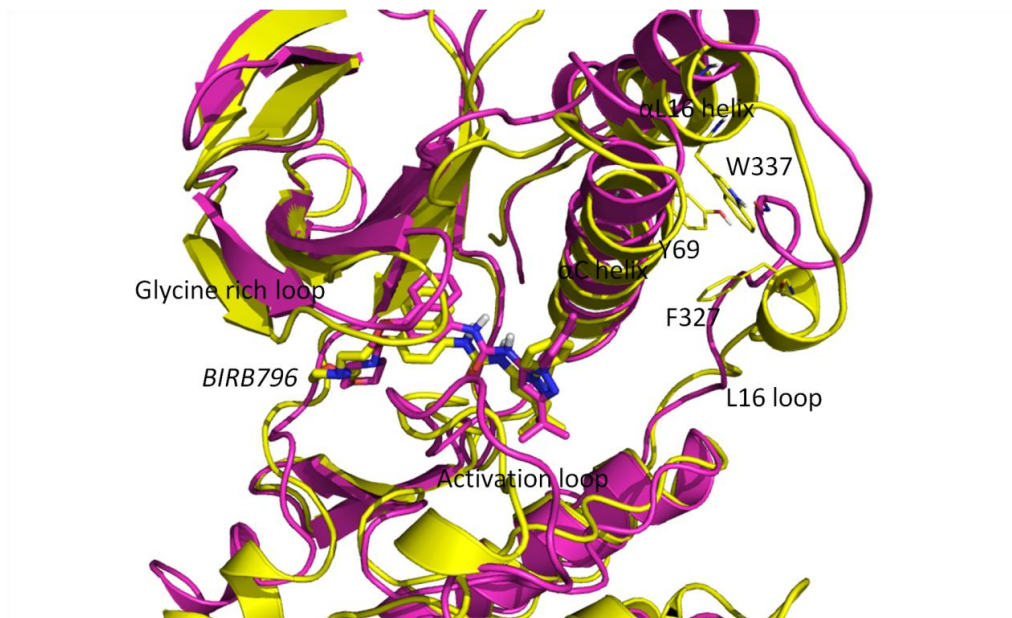
Figure 3.28. Superposition of representative structures of wild p38α-BIRB796 complex (yellow) and its mutated structure (pink) from CMD simulations. Residues Tyr69, Phe327 and Trp337 are mutated to glycine. Ligand BIRB796 is shown in bold licorice structure. Tyr69, Phe327 and Trp337 and their mutated glycine form are shown in thin licorice structure.
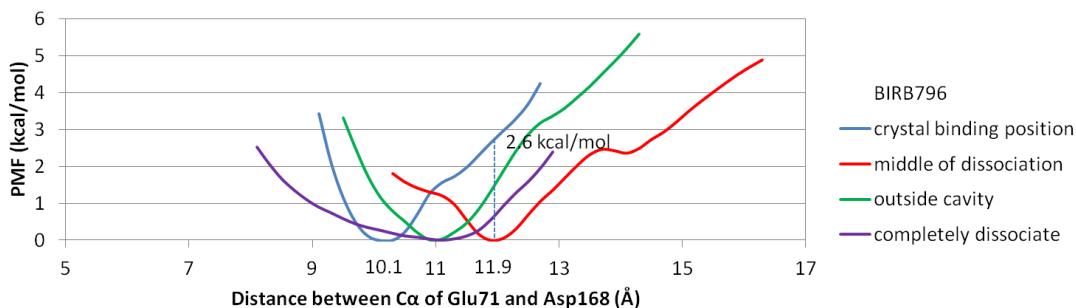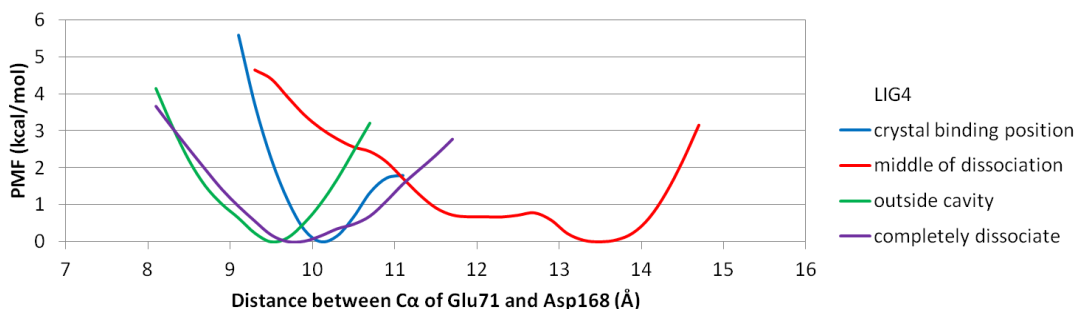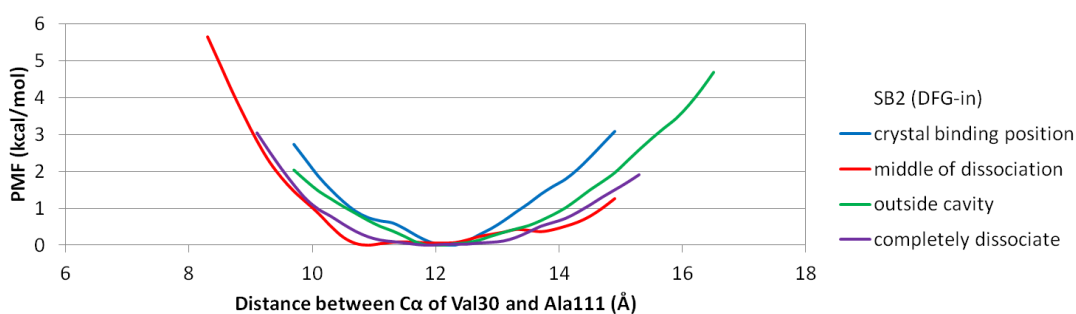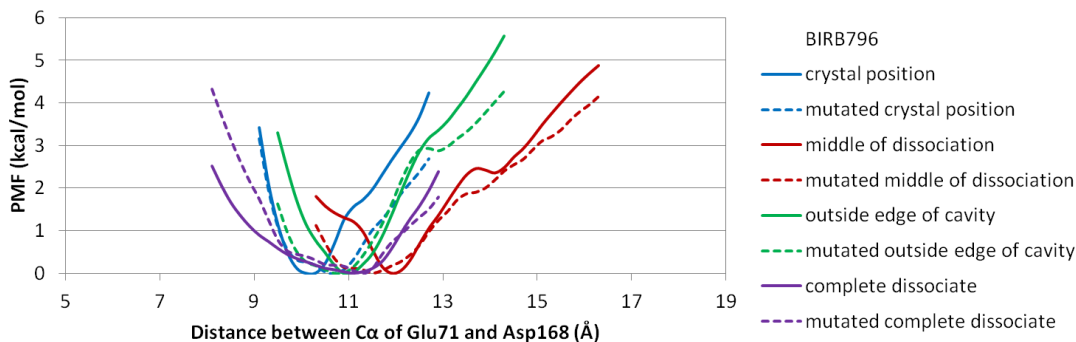
Figure 3.29. Free energy change along hinge movement of p38α in DFG-out conformation at different stages of dissociation of BIRB796 before and after mutation. Distance between Cα of Glu71 and Asp168 is used as RC.

3.4 Conclusions

In this study, we applied two enhanced sampling methods (AMD and PSIM) to investigate the dissociation processes for p38α complex systems and successfully simulated the dissociation pathways for type-I, II and III inhibitors, which provide useful guidance for PMF construction by US. We studied the detailed protein-ligand interactions that contribute to the free energy changes during dissociation. The lower PMF depths of the allosteric channel for type-II/III ligands indicate they unbind through the allosteric pathway. Correlation analysis suggested that hinge motion may play an important role in ligand dissociation process. The free energy change along hinge movement of protein proves the importance of hinge movement in the dissociation of type II/III ligands, which explains their slow dissociation rate comparing to type-I ligands, whose dissociation don't require large protein rearrangement. The mutation simulation suggests that a more flexible hinge region may facilitate the entry/exit of type-II/III ligands. The success of AMD and PSIM simulations on the dissociation of different types of p38α ligands, and

the capability of US in looking into detailed protein-ligand interactions as well as protein conformational rearrangements during ligand dissociation, provide powerful tools and useful guidance for future study of binding/unbinding mechanisms of kinase systems and inhibitor development.

3.5 References

1.   Arora A, Scholar EM. Role of tyrosine kinase inhibitors in cancer therapy. The Journal of pharmacology and experimental therapeutics. 2005;315(3):971-9. Epub 2005/07/09. doi: 10.1124/jpet.105.084145. PubMed PMID: 16002463.

2.   Biscardi JS, Tice DA, Parsons SJ. c-Src, receptor tyrosine kinases, and human cancer. Advances in cancer research. 1999;76:61-119. Epub 1999/04/28. PubMed PMID: 10218099.

3.   Dowell J, Minna JD, Kirkpatrick P. Erlotinib hydrochloride. Nature reviews Drug discovery. 2005;4(1):13-4. Epub 2005/02/05. doi: 10.1038/nrd1612. PubMed PMID: 15690599.

4.   Giamas G, Stebbing J, Vorgias CE, Knippschild U. Protein kinases as targets for cancer treatment. Pharmacogenomics. 2007;8(8):1005-16. Epub 2007/08/25. doi: 10.2217/14622416.8.8.1005. PubMed PMID: 17716234.

5.   Murphy E, Steenbergen C. Inhibition of GSK-3β as a target for cardioprotection: the importance of timing, location, duration and degree of inhibition. Expert Opinion on Therapeutic Targets. 2005;9(3):447-56. doi: 10.1517/14728222.9.3.447.

6.   Sawyers C. Targeted cancer therapy. Nature. 2004;432(7015):294-7. Epub 2004/11/19. doi: 10.1038/nature03095. PubMed PMID: 15549090.

7.   Vlahos CJ, McDowell SA, Clerk A. Kinases as therapeutic targets for heart failure. Nature reviews Drug discovery. 2003;2(2):99-113. Epub 2003/02/04. doi: 10.1038/nrd1009. PubMed PMID: 12563301.

8.   Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, et al. Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell. 2004;116(6):855-67. Epub 2004/03/24. PubMed PMID: 15035987.

9.   Copeland RA, Pompliano DL, Meek TD. Drug-target residence time and its implications for lead optimization. Nature reviews Drug discovery. 2006;5(9):730-9. Epub 2006/08/05. doi: 10.1038/nrd2082. PubMed PMID: 16888652.

10.   Nunez S, Venhorst J, Kruse CG. Target-drug interactions: first principles and their application to drug discovery. Drug discovery today. 2012;17(1-2):10-22. Epub 2011/07/23. doi: 10.1016/j.drudis.2011.06.013. PubMed PMID: 21777691.

11.   Pan AC, Borhani DW, Dror RO, Shaw DE. Molecular determinants of drug-receptor binding kinetics. Drug discovery today. 2013;18(13-14):667-73. Epub 2013/03/05. doi: 10.1016/j.drudis.2013.02.007. PubMed PMID: 23454741.

12.   Tummino PJ, Copeland RA. Residence time of receptor-ligand complexes and its effect on biological function. Biochemistry. 2008;47(20):5481-92. Epub 2008/04/17. doi: 10.1021/bi8002023. PubMed PMID: 18412369.

13.   Adams JL, Badger AM, Kumar S, Lee JC. p38 MAP kinase: molecular target for the inhibition of pro-inflammatory cytokines. Progress in medicinal chemistry. 2001;38:1-60. Epub 2002/01/05. PubMed PMID: 11774793.

14.   Cuenda A, Rousseau S. p38 MAP-kinases pathway regulation, function and role in human diseases. Biochimica et biophysica acta. 2007;1773(8):1358-75. Epub 2007/05/08. doi: 10.1016/j.bbamcr.2007.03.010. PubMed PMID: 17481747.

15.   Kumar S, Boehm J, Lee JC. p38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases. Nature reviews Drug discovery. 2003;2(9):717-26. Epub 2003/09/03. doi: 10.1038/nrd1177. PubMed PMID: 12951578.

16.   Peifer C, Wagner G, Laufer S. New approaches to the treatment of inflammatory disorders small molecule inhibitors of p38 MAP kinase. Current topics in medicinal chemistry. 2006;6(2):113-49. Epub 2006/02/04. PubMed PMID: 16454763.

17.   Schindler JF, Monahan JB, Smith WG. p38 pathway kinases as anti-inflammatory drug targets. Journal of dental research. 2007;86(9):800-11. Epub 2007/08/28. PubMed PMID: 17720847.

18.   Dominguez C, Powers DA, Tamayo N. p38 MAP kinase inhibitors: many are made, but few are chosen. Current opinion in drug discovery & development. 2005;8(4):421-30. Epub 2005/07/19. PubMed PMID: 16022178.

19.   Kong TT, Zhang CM, Liu ZP. Recent developments of p38alpha MAP kinase inhibitors as antiinflammatory agents based on the imidazole scaffolds. Current medicinal chemistry. 2013;20(15):1997-2016. Epub 2013/01/16. PubMed PMID: 23317165.

20.   Lee MR, Dominguez C. MAP kinase p38 inhibitors: clinical results and an intimate look at their interactions with p38alpha protein. Current medicinal chemistry. 2005;12(25):2979-94. Epub 2005/12/28. PubMed PMID: 16378500.

21.   Pettus LH, Wurz RP. Small molecule p38 MAP kinase inhibitors for the treatment of inflammatory diseases: novel structures and developments during 2006-2008. Current topics in medicinal chemistry. 2008;8(16):1452-67. Epub 2008/11/11. PubMed PMID: 18991731.

22.   Huse M, Kuriyan J. The conformational plasticity of protein kinases. Cell. 2002;109(3):275-82. Epub 2002/05/23. PubMed PMID: 12015977.

23.   Vogtherr M, Saxena K, Hoelder S, Grimme S, Betz M, Schieborr U, et al. NMR characterization of kinase p38 dynamics in free and ligand-bound forms. Angewandte Chemie (International ed in English). 2006;45(6):993-7. Epub 2005/12/24. doi: 10.1002/anie.200502770. PubMed PMID: 16374788.

24.   Wilson KP, McCaffrey PG, Hsiao K, Pazhanisamy S, Galullo V, Bemis GW, et al. The structural basis for the specificity of pyridinylimidazole inhibitors of p38 MAP kinase. Chemistry & biology. 1997;4(6):423-31. Epub 1997/06/01. PubMed PMID: 9224565.

25.   Kufareva I, Abagyan R. Type-II kinase inhibitor docking, screening, and profiling using modified structures of active kinase states. Journal of medicinal chemistry. 2008;51(24):7921-32. Epub 2008/12/05. doi: 10.1021/jm8010299. PubMed PMID: 19053777; PubMed Central PMCID: PMCPmc2669721.

26.   Simard JR, Getlik M, Grutter C, Schneider R, Wulfert S, Rauh D. Fluorophore labeling of the glycine-rich loop as a method of identifying inhibitors that bind to active and inactive kinase conformations. Journal of the American Chemical Society. 2010;132(12):4152-60. Epub 2010/03/06. doi: 10.1021/ja908083e. PubMed PMID: 20201574.

27.   Casasnovas R, Limongelli V, Tiwary P, Carloni P, Parrinello M. Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations.

Journal of the American Chemical Society. 2017;139(13):4780-8. Epub 2017/03/16. doi: 10.1021/jacs.6b12950. PubMed PMID: 28290199.

28.   Yang LJ, Zou J, Xie HZ, Li LL, Wei YQ, Yang SY. Steered molecular dynamics simulations reveal the likelier dissociation pathway of imatinib from its targeting kinases c-Kit and Abl. PloS one. 2009;4(12):e8470. Epub 2009/12/31. doi: 10.1371/journal.pone.0008470. PubMed PMID: 20041122; PubMed Central PMCID: PMCPmc2795779.

29.   Capelli AM, Costantino G. Unbinding pathways of VEGFR2 inhibitors revealed by steered molecular dynamics. J Chem Inf Model. 2014;54(11):3124-36. Epub 2014/10/10. doi: 10.1021/ci500527j. PubMed PMID: 25299731.

30.   Sun H, Li Y, Li D, Hou T. Insight into crizotinib resistance mechanisms caused by three mutations in ALK tyrosine kinase using free energy calculation approaches. Journal of chemical information and modeling. 2013;53(9):2376-89. Epub 2013/08/21. doi: 10.1021/ci400188q. PubMed PMID: 23952683.

31.   Sun H, Tian S, Zhou S, Li Y, Li D, Xu L, et al. Revealing the favorable dissociation pathway of type II kinase inhibitors via enhanced sampling simulations and two-end-state calculations. Scientific reports. 2015;5:8457. Epub 2015/02/14. doi: 10.1038/srep08457. PubMed PMID: 25678308; PubMed Central PMCID: PMCPmc4326958.

32.   Wang Z, Canagarajah BJ, Boehm JC, Kassisa S, Cobb MH, Young PR, et al. Structural basis of inhibitor selectivity in MAP kinases. Structure (London, England : 1993). 1998;6(9):1117-28. Epub 1998/10/01. PubMed PMID: 9753691.

33.   Simard JR, Getlik M, Grutter C, Pawar V, Wulfert S, Rabiller M, et al. Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. Journal of the American Chemical Society. 2009;131(37):13286-96. Epub 2009/07/04. doi: 10.1021/ja902010p. PubMed PMID: 19572644.

34.   Gill AL, Frederickson M, Cleasby A, Woodhead SJ, Carr MG, Woodhead AJ, et al. Identification of novel p38alpha MAP kinase inhibitors using fragment-based lead generation. Journal of medicinal chemistry. 2005;48(2):414-26. Epub 2005/01/22. doi: 10.1021/jm049575n. PubMed PMID: 15658855.

35.   Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, et al. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. Nature structural biology. 2002;9(4):268-72. Epub 2002/03/16. doi: 10.1038/nsb770. PubMed PMID: 11896401.

36.   Casper D, Bukhtiyarova M, Springman EB. A Biacore biosensor method for detailed kinetic binding analysis of small molecule inhibitors of p38alpha mitogen-activated protein kinase. Analytical biochemistry. 2004;325(1):126-36. Epub 2004/01/13. PubMed PMID: 14715293.

37.   Regan J, Pargellis CA, Cirillo PF, Gilmore T, Hickey ER, Peet GW, et al. The kinetics of binding to p38MAP kinase by analogues of BIRB 796. Bioorganic & medicinal chemistry letters. 2003;13(18):3101-4. Epub 2003/08/28. PubMed PMID: 12941343.

38.   Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, et al. {Amber 14}2014.

39.   Case DA, Cheatham TE, 3rd, Darden T, Gohlke H, Luo R, Merz KM, Jr., et al. The Amber biomolecular simulation programs. Journal of computational chemistry. 2005;26(16):1668-88. Epub 2005/10/04. doi: 10.1002/jcc.20290. PubMed PMID: 16200636; PubMed Central PMCID: PMCPmc1989667.

40.   Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 2006;65(3):712-25. Epub 2006/09/19. doi: 10.1002/prot.21123. PubMed PMID: 16981200; PubMed Central PMCID: PMCPmc4805110.

41.   Okur A, Strockbine B, Hornak V, Simmerling C. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. Journal of computational chemistry. 2003;24(1):21-31. Epub 2002/12/17. doi: 10.1002/jcc.10184. PubMed PMID: 12483672.

42.   Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. Journal of computational chemistry. 2005;26(16):1781-802. Epub 2005/10/14. doi: 10.1002/jcc.20289. PubMed PMID: 16222654; PubMed Central PMCID: PMCPmc2486339.

43.   Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. The Journal of Chemical Physics. 1995;103(19):8577-93. doi: 10.1063/1.470117.

44.   Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics. 1977;23(3):327-41. doi: http://dx.doi.org/10.1016/0021-9991(77)90098-5.

45.   Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys. 2004;120(24):11919-29. Epub 2004/07/23. doi: 10.1063/1.1755656. PubMed PMID: 15268227.

46.   Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. Journal of computational chemistry. 1992;13(8):1011-21. doi: 10.1002/jcc.540130812.

47.   Souaille M, Roux Bt. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Computer physics communications. 2001;135(1):40-57. doi: http://dx.doi.org/10.1016/S0010-4655(00)00215-0.

48.   Ai R, Qaiser Fatmi M, Chang CE. T-Analyst: a program for efficient analysis of protein conformational changes by torsion angles. Journal of computer-aided molecular design. 2010;24(10):819-27. Epub 2010/08/07. doi: 10.1007/s10822-010-9376-y. PubMed PMID: 20689979; PubMed Central PMCID: PMCPmc2940022.

49.   Mondal J, Friesner RA, Berne BJ. Role of Desolvation in Thermodynamics and Kinetics of Ligand Binding to a Kinase. Journal of chemical theory and computation.

2014;10(12):5696-705. Epub 2014/12/18. doi: 10.1021/ct500584n. PubMed PMID: 25516727; PubMed Central PMCID: PMCPmc4263462.

50. Pearlstein RA, Sherman W, Abel R. Contributions of water transfer energy to protein-ligand association and dissociation barriers: Watermap analysis of a series of p38alpha MAP kinase inhibitors. Proteins. 2013;81(9):1509-26. Epub 2013/03/08. doi: 10.1002/prot.24276. PubMed PMID: 23468227.

51. Huang Y-mM, Raymundo MAV, Chen W, Chang C-eA. Mechanism of the Association Pathways for a Pair of Fast and Slow Binding Ligands of HIV-1 Protease. Biochemistry. 2017;56(9):1311-23. doi: 10.1021/acs.biochem.6b01112.

52. Agafonov RV, Wilson C, Otten R, Buosi V, Kern D. Energetic dissection of Gleevec's selectivity toward human tyrosine kinases. Nature structural & molecular biology. 2014;21(10):848-53. Epub 2014/09/15. doi: 10.1038/nsmb.2891. PubMed PMID: 25218445; PubMed Central PMCID: PMCPmc4266587.

Chapter 4 Protonation States and Catalysis: Molecular Dynamics Studies of Intermediates in Tryptophan Synthase

4.1 Introduction

Acid-base reaction is one of the common catalysis mechanisms of many enzymes [1-4], the proton transfer in which can induce necessary interaction and conformational changes, as well as improve or reduce enzyme activity [5, 6]. Therefore, determining protonation states of substrates and key catalyticly important residues in active site is critical for understanding enzyme mechanism. However, crystal structures determined by experimental measurements are mostly have insufficient resolution of protein structures to determine the protonation states for key residues in the active site.

Isotopic labeling in solid state NMR (ssNMR) has been fast developing and used to map electrostatic and chemical environments of active site of enzymes [7, 8]. But as for the influence of proton transfer on the enzyme activity and large scale conformational change, ssNMR is not able to provide much information. Neither does the ab initio calculations of protonation state at enzyme catalytic site [9]. That's where all-atom molecular dynamics (MD) simulation comes in. With multiple simulations with different assinged protonation states, we can explore how protonation states affect overal protein motion and enzyme activity. Here, we applied all-atom MD simulation to study a well-known enzyme system, tryptophan synthase (TRPS).

TRPS is an $\alpha_2\beta_2$ tetrameric enzyme that catalyzes the last two steps in the biosynthesis of L-tryptophan. It has long been explored to understand the catalytic and allosteric regulation mechanisms of enzyme complexes [10-12]. The bacterial enzyme structure of TRPS consists of α- and β-subunits (Figure 4.1). The active sites are regulated allosterically by transitions between open, inactive, and closed, active, states in both α- and β-subunits [12-14]. The open and closed transition of the α-subunit is controlled by the motions in α-loop L6 (αL6, residues α179-193) and indicated by the H-bond between αAsp60 and αThr183. The open and close states of the β-subunit is indicated by the motion of the communication (COMM) domain (residues β102-189) and the salt bridge between βArg141 and βAsp305 [15]. The α active site is connected to a β active site by a 25 Å long hydrophobic channel contained within the enzyme. The α-site catalyzes the formation of indole and glyceraldehydes-3-phosphate (G3P) from a cleavage of indole-3-glycerol phosphate (IGP), equation 1 [16]. Through channeling, indole is transfered from α-site to β-site [17-21], where indole participates in the sythesis of L-Trp in a series of steps [22, 23]. The reactions at α- and β-sites are as follows:

α-reaction: IGP $\rightarrow$ G3P + indole $\hspace{3cm}$ (eq. 4-1)

β-reaction: Indole + L-Ser $\rightarrow$ L-Trp + H_2O $\hspace{1.5cm}$ (eq. 4-2)
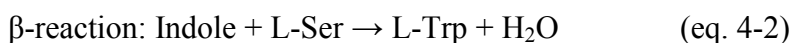
Details of each step is shown in Figure 4.2. In β-site, the first stage of reactions starts with L-Ser reacting with the cofactor in its internal aldimine resting form, E(Ain), producing gem-diamine, E(GD$_1$), then L-Ser external aldimine, E(Aex$_1$), followed by quinonoid, E(Q$_1$), end with α-aminoacrylate Schiff base (SB), E(A-A), and a water molecule. In stage II, indole makes a nucleophilic attack on E(A-A), producing the

intermediates $E(Q_2)$, $E(Q_3)$, $E(Aex_2)$, $E(GD_2)$, and, at last, L-Trp. Substrate analogues like indoline and 2-aminophenol (2AP) for indole are used during crystallizing to give quasi-stable, catalytically active species. Indoline and 2-aminophenol (2AP) can react rapidly with E(A-A) to produce the long-lived indoline quinonoid, $E(Q)_{indoline}$, and 2AP quinonoid, $E(Q)_{2AP}$, species, respectively [21, 24].
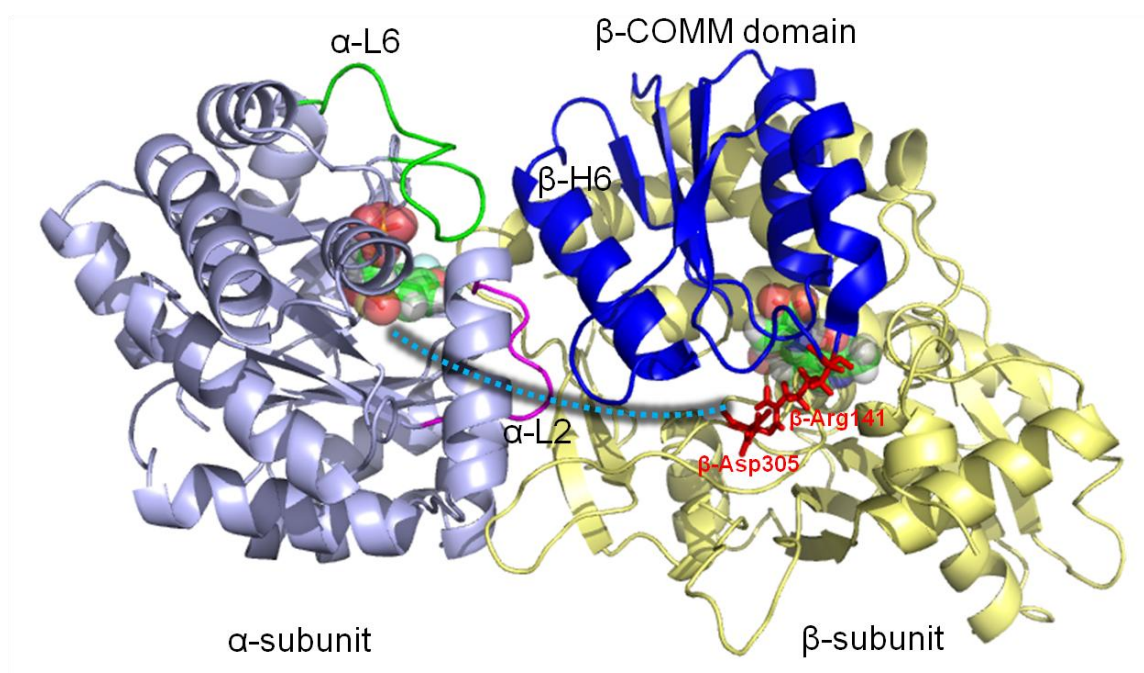
Figure 4.1: Overall structure and chemical reactions of tryptophan synthase (TRPS). TRPS is composed of an α-subunit (purple) and β-subunit (yellow). The two ligands binding to each subunit are shown in ball representation. The residues βArg141 and βAsp305, highlighted in red, are associated with the open and closed β-site conformations. The tunnel used to channel indole from the α-site to the β-site is marked as a cyan dashed line.
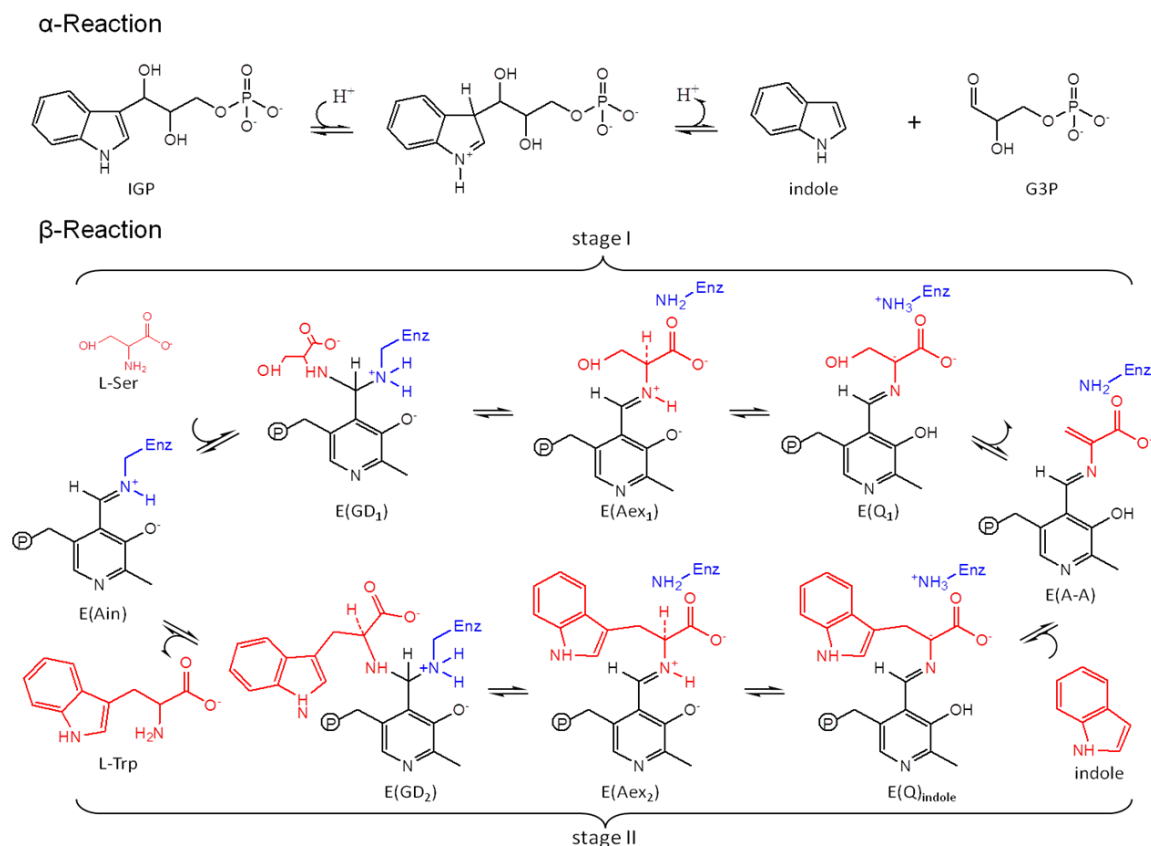
Figure 4.2. The α- and β-reactions of TRPS.

Recent NMR studies provided chemical shifts for selected atoms of the E(Ain) intermediate and suggested that the Schiff base linkage of the Lys87 ε-nitrogen is protonated and the pyridine nitrogen (PN) of Ain is deprotonated [25]. However, for the E(A-A), E(Q)$_{indoline}$, and E(Q)$_{2AP}$, previously study [24] seems to suggest that the proton transfers from the SB nitrogen to the phenolic oxygen (PO) for most of the species (Figure 4.2). In this study, we explore the dynamic behavior of TRPS for at different stages of β-reaction via MD simulations. Simulations focuses on different protonation states for the PLP ionizable groups and we discuss the roles played by protonation states in the mechanism of TRPS.

4.2. Materials and Methods

4.2.1 Molecular systems

The crystal structures of the TRPS in complex with E(Ain), E(A-A), E(Q)$_{indoline}$, and E(Q)$_{2AP}$ were obtained from the protein data bank (PDB) ID 4HT3, 4HN4, 3PR2 and 4HPJ, respectively [24, 26]. The coordinates of the TRPS:E(A-A) complex including a benzimidazole (BZI) molecule was obtained from the PDB ID 4HPX [26]. All systems studied include the ligand F9 bound in the α-active site and the MVC site in the β-subunit occupied by Cs$^+$ ion.

4.2.2 Molecular dynamics simulations

The standard Amber 14 package was used for MD simulations [27-29]. The protein and substrates were modeled with the Amber 99SB and general Amber force field, respectively [30, 31]. The atom charges of the intermediates were given by the vCharge model[32]. For MD simulation preparation, we assume that the internal coordinates in crystal structures such as bond length and angle are equilibrium positions for both protein and α- and β-substrates. In other words, the PDB structures are considered the global energy minima. We carefully minimized the systems through hydrogen atoms, sidechains, and the entire protein complex for 500, 5000, and 50000 steps, respectively. The counterion, Na$^+$ based on the Coulombic potential, was placed where necessary to maintain the overall system as neutral. After solvating the system by the TIP3P water model to create the system in a rectangular box of 12 Å [33], we minimized the water and entire system for 10000 and 20000 steps, respectively, to correct small flaws. Water

molecules were equilibrated for 40 ps. Then, all molecules, including the protein, substrates, cofactor, and solvent, were relaxed gradually by heating the system at 250, 275, and 300 K for 20, 20, and 160 ps, respectively. The particle mesh Ewald was turned on to consider long-range electrostatic interactions[34, 35]. The Langevin thermostat with a damping constant of 2 ps^{-1} was applied to maintain a temperature of 300 K, and the SHAKE algorithm was used to constrain hydrogen atoms during MD simulations [36]. We collected the resulting trajectories every 1 ps with a time step of 2 fs in the isothermic-isobaric (NPT) ensemble, and the total simulation time for each protein–substrate system was 50 ns.

4.3 Results and Discussions

Here we aimed to explore the detailed dynamics regarding a proton located at different functional groups of the β-site substrates in TRPS. Figure 4.3 shows potential sites of protonation on a PLP derivative, including the phosphoryl group (PG), pyridine nitrogen (PN), pyridoxyl phenolic oxygen (PO), Schiff base (SB) nitrogen linkage, both carboxylate oxygens (CO), and the neighboring ε-nitrogen of β-Lys87. Not all of these sites can be simultaneously protonated and we particularly focused on the proton transition within PO, SB and CO. The doubly-protonationed species including one proton on PN or PG were also considered, which gives 17 possible substrate structures. Here we refer a substrate with specific protonation states by an abbreviation; for example, a proton at PO position of E(Ain) is noted as E(Ain):PO. All potential intermediates with different protonation states in the TRPS β-reactions are listed in Figure 4.4.
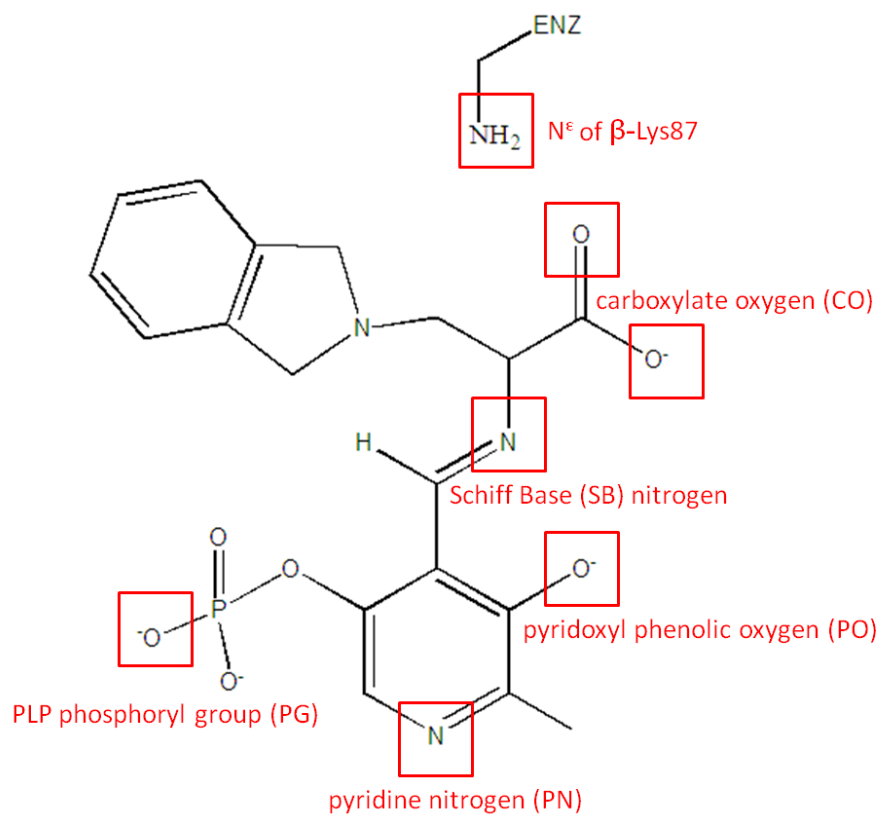
Figure 4.3. Example of the potential sites of protonation on an indoline quinonoid substrate. The ionizable groups, shown in a red box, include an atom that can be protonated or deprotonated.
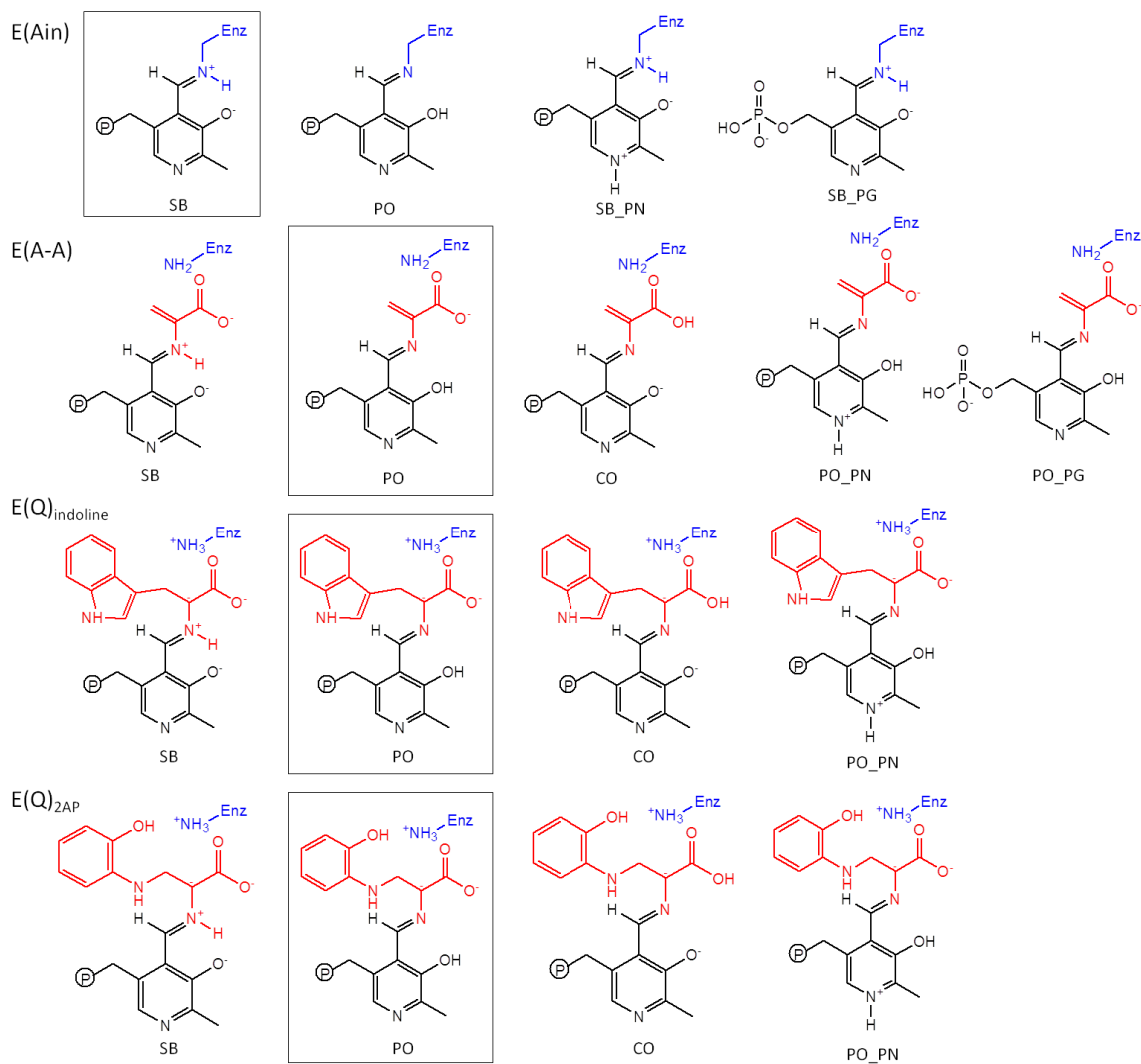
Figure 4.4. Substrates investigated in this molecular dynamics (MD) simulation study. The intermediates include the changes in proton on each ionizable group. The predominant protonation states from solid-state NMR reports are shown in boxes.

4.3.1 E(Ain)

The proton of this intermediate shows high tendency to stay with the SB nitrogen [25]. Our MD simulations present that the E(Ain):SB substrate can well stay in the binding pocket of the β-subunit. The residues that make direct contact with E(Ain):SB are shown in Figure 4.5(SB) shows the residues which can make direct contacts with E(Ain):SB.

The PG forms multiple H-bonds with neighboring residues. The amino acid, Ser377, forms interactions with PN, as the early publication suggested [25]. In addition, all E(Ain) species are surrounded by numerous water molecules.
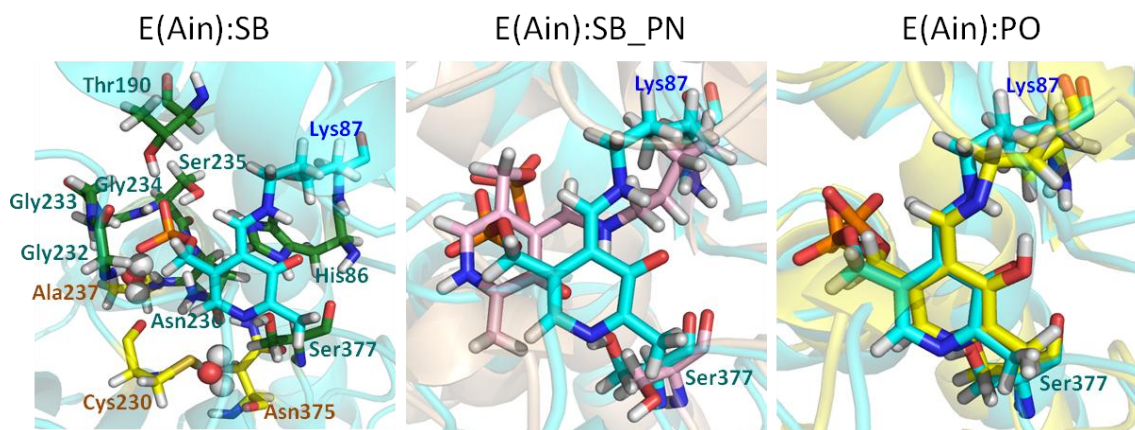
Figure 4.5. Protonation states of E(Ain). (left) The ligand (cyan) shows a proton at the Schiff base (SB) nitrogen. (middle) A proton was added at the pyridine nitrogen (PN). Motions of the ligand with two different protonation states, SB (cyan) and SB_PN (pink). (right) Comparison of a proton at the phenolic oxygen (PO) (yellow) and SB nitrogen (cyan).

To confirm that it's unlikely to have a protonated PN, we further added one more proton at PN to create another protonation state designated as E(Ain):SB_PN. As a result, it becomes too crowded in the binding pocket that E(Ain):SB_PN shifts in the position, losing its interactions between the PG of E(Ain) and Gly232, Gly233, and Gly234, in the meanwhile, the H-bond between the PN and Ser377 is lost (Figure 4.5(SB_PN)).

Although it seems that the substrate and protein dynamics are not affected significantly by the switch of a proton from the SB nitrogen to the PO, the different performances of water molecules were observed between E(Ain):SB and E(Ain):PO. Since the TRPS in complex with E(Ain) shows open conformations in the β-subunit, the substrate in the β-

site is well solvated by water molecules. MD simulation of E(Ain):PO shows that the PO is surrounded by fewer water molecules with a neutral PO group. In sum, a formal negative charge on the PO in the E(Ain):SB structure seems necessary to attract water molecules in order to fully solvate the β-site.

4.3.2 E(A-A)

E(A-A) is the first stable intermediate state after L-Ser is covalently bound to PLP that the crystal structures are available. The ssNMR spectrum of E(A-A) shows that the proton prefers to stay at the PO position most of time. Based on our MD simulations for both E(A-A) and E(Ain) intermediates and the close proximity of the two protonation sites, a proton can have smooth transition between the PO and SB position. The protein can directly make three major contacts to the substrate (see Figure 4.6(PO)). First, as the standard PLP substrates of TRPS, the PG binds to Lys87, His86, Thr190, Gly234, Ser235 and Asn236. Second, the PN form interactions with Ser377 sidechain. Third, the CO atoms interact with the hydroxyl group of Thr110 sidechain and Gly111 and His 115 backbone nitrogen.
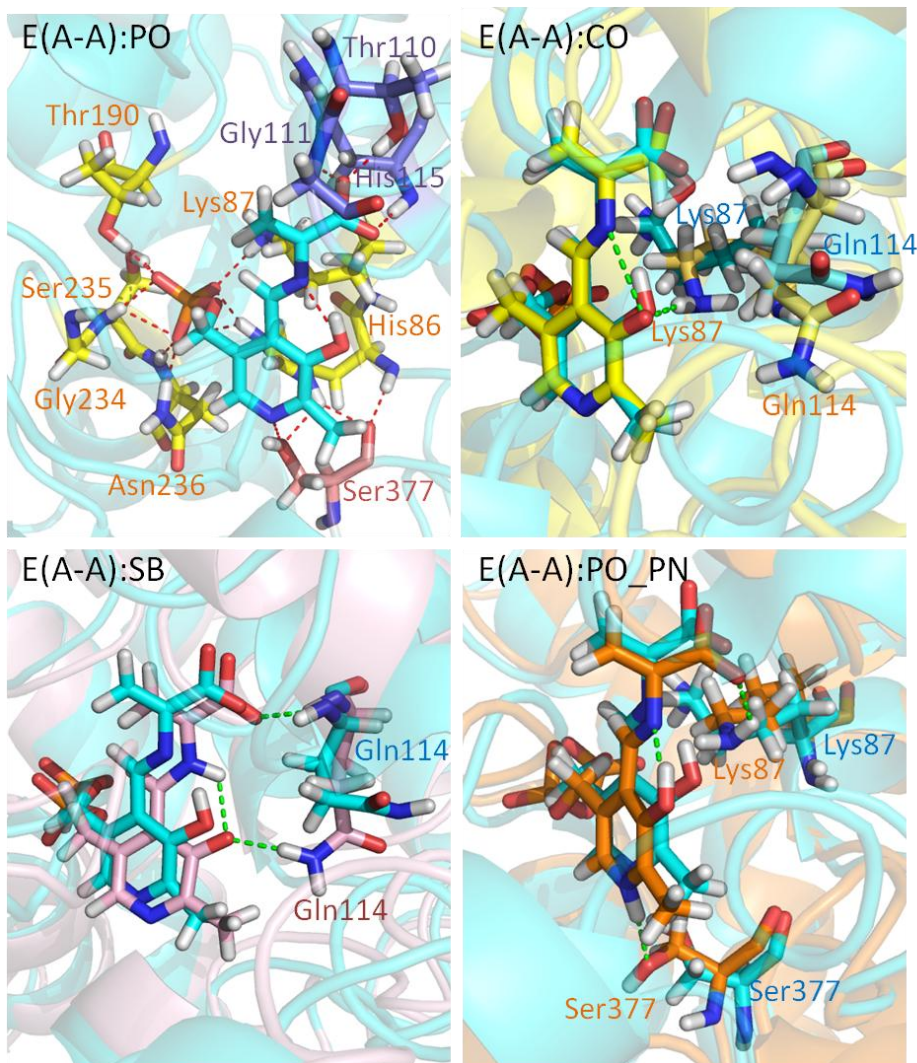
Figure 4.6. Protonation states of E(A-A). (PO) Presentation of E(A-A):PO binding to TRPS. The residues interacting with the phosphoryl group (PG), PN, and carboxylate oxygens (COs) of E(A-A):PO are shown in yellow, pink, and purple, respectively. (CO) The alignment of E(A-A):CO (yellow) and E(A-A):PO (cyan). Residues Lys87 and Gln114 are shown. The H-bond between Lys87 and the PO of E(A-A):CO is shown in green. (SB) The alignment of E(A-A):SB (pink) and E(A-A):PO (cyan). The Gln114 sidechain forming H-bonds with the PO of E(A-A):SB is shown as a green dashed line. (PO_PN) The alignment of E(A-A):PO_PN (orange) and E(A-A):PO (cyan) is depicted. The Lys87 and Ser377 residues can form H-bonds with E(A-A):PO_PN.

It is known that the chemical reactions in the α-subunit release a key compound, indole,

which would be transferred to the β-site through the channel to process a series of β-

reactions. To model the E(A-A) reacting with indole to form a E(Q)$_{indole}$ intermediate that catalyzes the process of L-Trp synthesis from the stage I to stage II in Figure 4.2, the indole analogue, benzimidazole (BZI), was synthesized to simulate the presence of the indole in a E(A-A) complex. The crystal studies have shown the bound conformation of the TRPS with E(A-A) in complex of BZI (PDB ID: 4HPX) [26]. Our MD simulations then reveal the dynamics and cooperative motions within the protein, E(A-A) and BZI. As a passenger of the channel, BZI is enclosed by most of hydrophobic residues, such as Glu109, Leu166, Cys170, Leu188, Phe280 and Phe306 (Figure 4.7(left)), which create an stable environment to for BZI in the binding pocket. Although the BZI is mobile during the simulation (Figure 4.7(right)), the interactions between E(A-A):PO and TRPS, mentioned in the above section, are still hold, which indicates that both BZI and E(A-A):PO are accessible to stay in the β-active site at the same time and the overall system is stable enough while a proton resides at the PO of E(A-A).
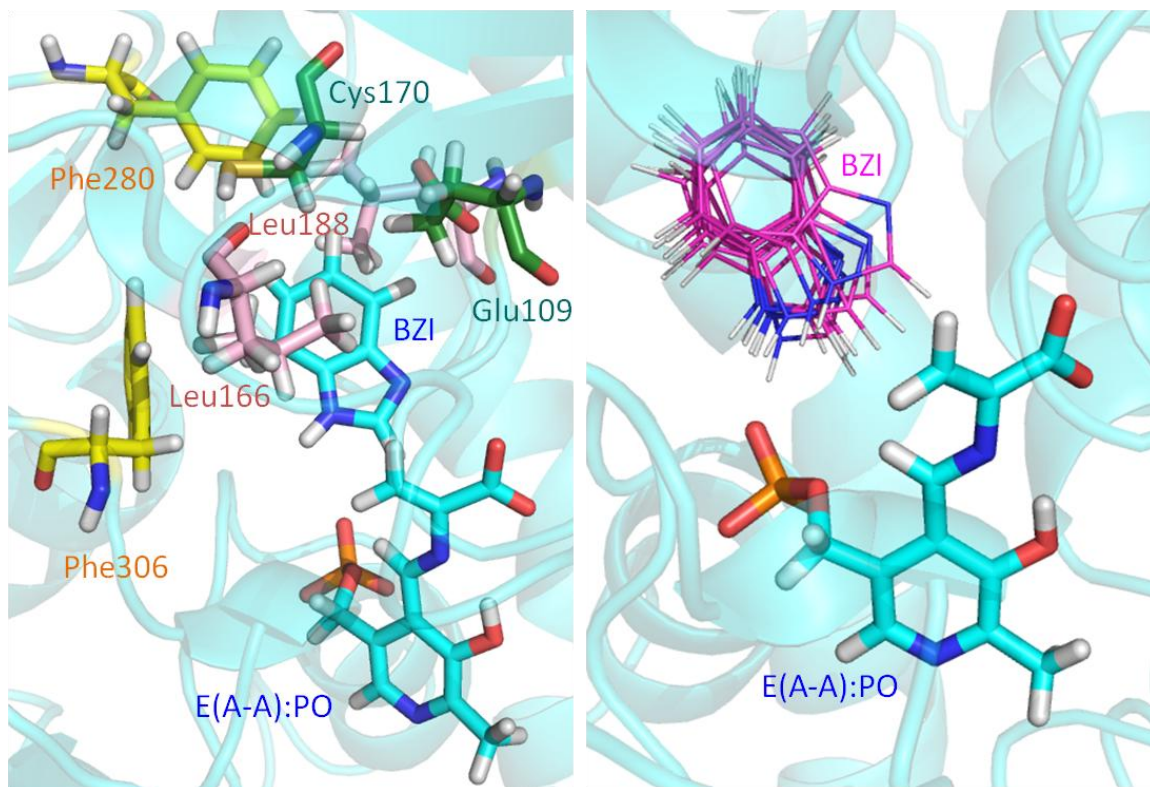
Figure 4.7. MD simulation of E(A-A):PO complex with BZI in the β-site. (left) The residues, Glu109, Leu166, Cys170, Leu188, Phe280 and Phe306, form interactions with BZI. (right) The BZI snapshot (pink) every 5 ns in the 50-ns MD simulation.

Next we checked the performance of E(A-A):CO intermediate. During 50ns MD simulation, the protein dynamics and conformations change so significantly that the β-site is able to open. The protonated carboxyl group affects the H-bond network. Compared to the E(A-A):PO, E(A-A):SB and E(A-A):PO_PN substrates which hold -1 charge at the CO atoms, fewer H-bonds between the CO of E(A-A):CO and the surrounding residues form, especially the significantly missing interactions between the CO and Thr110 sidechain. Although the alignment of the E(A-A):PO and E(A-A):CO substrate seems well (Figure 4.6(CO)), the weaker interactions around the CO of E(A-A):CO result in the destabilization of the residues around the substrate that Lys87 forms

H-bonds with the PO and Gln114 tends to move toward the PO instead of exposing to solvent, which conflicts with the findings from the X-ray structures. Furthermore, we also measured the distance between Arg141C$^{\zeta}$ and Asp305C$^{\gamma}$, which defines the open and closed conformations of the β-subunit. Figure 4.8 illustrates the salt-bridge between Arg141 and Asp305 breaks. And the TRPS in complex with E(A-A):CO shows an open β-site occasionally.
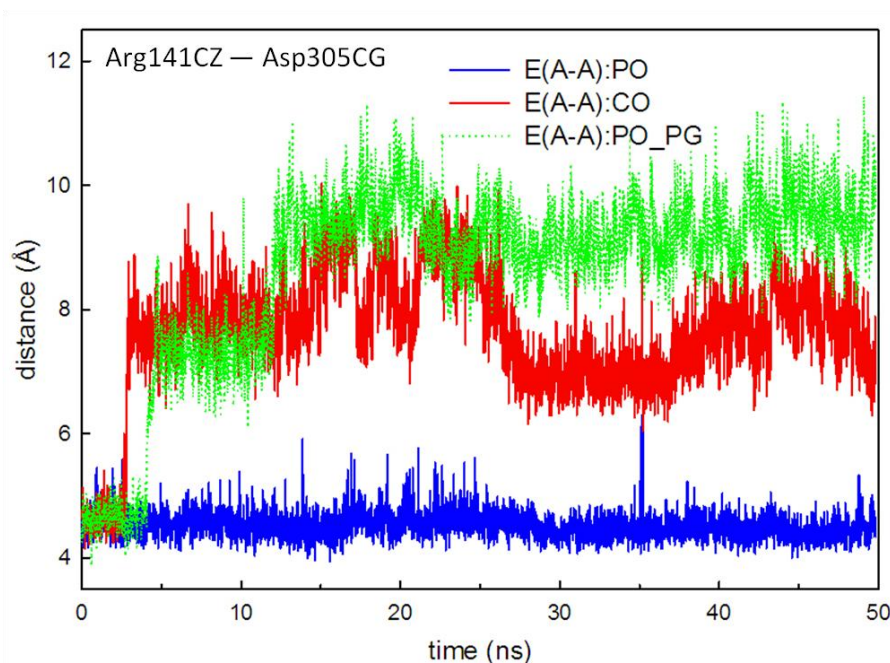
Figure 4.8. The distance between Arg141C$^{\zeta}$ and Asp305C$^{\gamma}$, which defines the open and closed conformations of the β-site.

The E(A-A) substrates are highly flexible in the β-site with another two protonation states, E(A-A):SB and E(A-A):PO_PN. RMSD of both E(A-A):SB and E(A-A):PO_PN are larger than E(A-A):PO. Similar to E(A-A):CO, the Gln114 sidechain moves toward the unprotonated PO to form H-bonds (Figure 4.6(SB)), which causes a slight shift of substrate position. Regarding to E(A-A):PO_PN, in order to keeping H-bonds between

the PN and the hydroxyl group of Ser377, the conformations of E(A-A):PO_PN twist a little bit, which allows the Lys87 sidechain rotate at the back site of the substrate. Thus, the Lys87 sidechain forms new interactions with the SB nitrogen (Figure 4.6(PO_PN)), and interactions between the PG and Lys87 are missing. However, the spectrum from X-ray diffraction does not support these large movements of the Gln114 and Lys87 sidechain. Therefore, our simulation results confirm the low population of a proton at the SB nitrogen and PN.

4.3.3 E(Q)$_{indoline}$ and E(Q)$_{2AP}$

The ssNMR results suggested that the proton of E(Q)$_{indoline}$ could shift within the PO, CO and SB nitrogen although the protonated PO is the major form. Similar to the E(A-A), three major interactions were detected in the binding of E(Q)$_{indoline}$ to TRPS that the PG interacts with Lys87, His86, Thr190, Gly234, Ser235 and Asn236, the PN interacts with Ser377 and the CO interacts with Thr110, Gly111 and His115. An intramolecular H-bond between the PO and SB nitrogen was also exposed. The binding of an indoline to the C$_\beta$ atom creates an only structural difference between the conformations of E(A-A) and E(Q)$_{indoline}$. Thus, the indoline with hydrophobic rings can well interact with the non-polar fractions from the neighboring residues, such as Glu109, Leu166, Thr190 and Phe306, which provides another contacts between the substrate and protein (Figure 4.9(PO)). Therefore, the above interactions give a highly stable intermediate of E(Q)$_{indoline}$:PO.
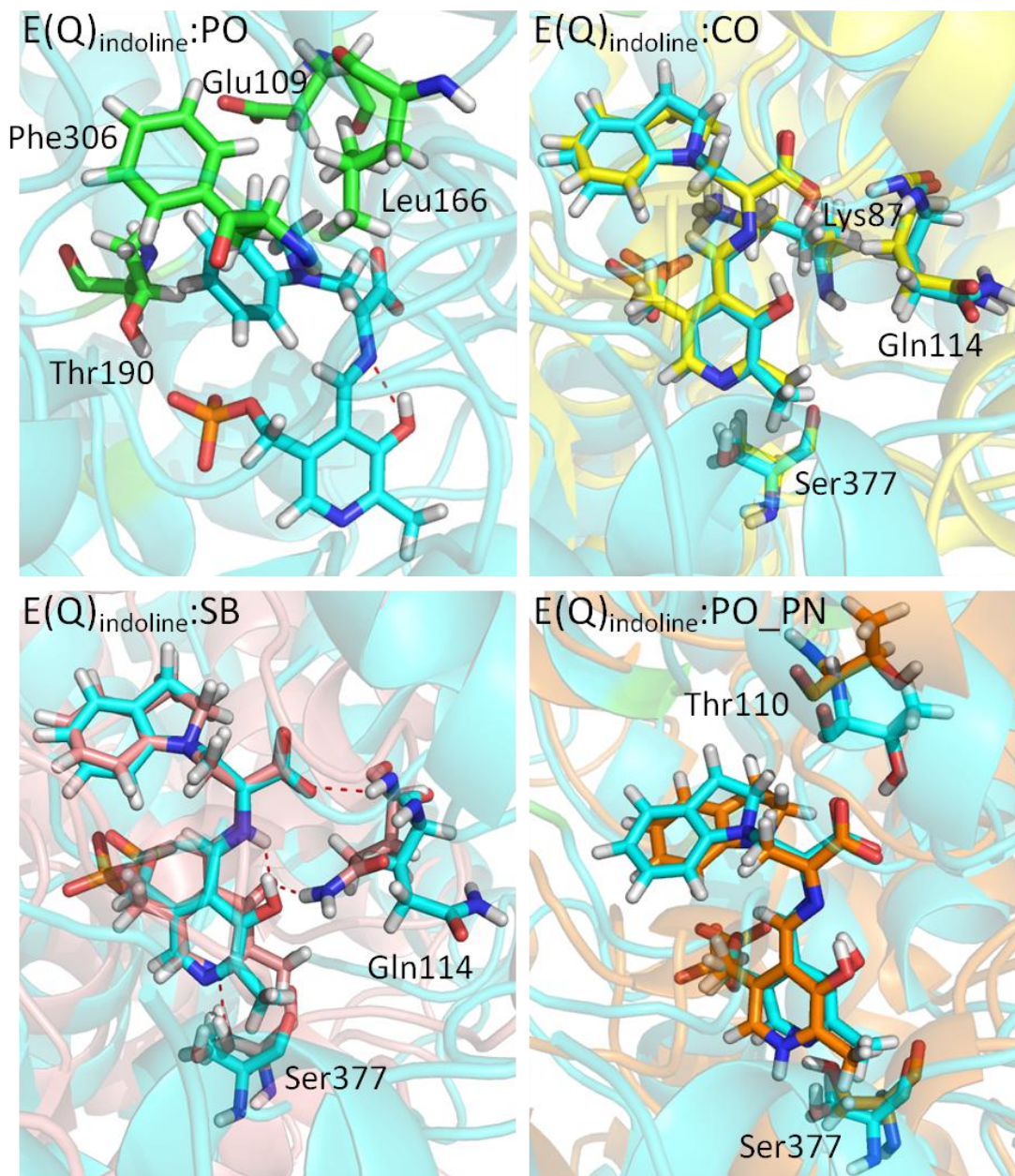
Figure 4.9. Protonation states of E(Q)$_{indoline}$. (PO) The residues interacting with the indoline ring of E(Q)$_{indoline}$:PO are shown in the green bond representation. (CO) The alignment of E(Q)$_{indoline}$:CO (yellow) and E(Q)$_{indoline}$:PO (cyan). The good alignment of the residues Lys87, Gln114, and Ser377 is shown. (SB) The alignment of E(Q)$_{indoline}$:SB (pink) and E(Q)$_{indoline}$:PO (cyan). The residues Gln114 and Ser377 form H-bonds (red dash lines) with the ligand. (PO_PN) The alignment of E(Q)$_{indoline}$:PO_PN (orange) and E(Q)$_{indoline}$:PO (cyan). The hydroxyl group of Ser377 and Thr110 rotates away from E(Q)$_{indoline}$:PO_PN.

172

Then we moved on to CO, the well alignment between E(Q)$_{indoline}$:CO and E(Q)$_{indoline}$:PO is shown in Figure 4.9(CO). The key residues, Lys87, Gln114 and Ser377, of these two protonation states also present in similar conformations, which agrees with the findings from the ssNMR study that the population of the proton covalently binding to the CO is possible. Although the transfer of the proton between the PO to CO in the E(A-A) complexes induces the movement of Gln114 toward the PO and the missing H-bonds between the CO and Thr110, it is not the case in the binding of E(Q)$_{indoline}$:CO. The H-bonds between the CO of the E(Q)$_{indoline}$:CO substrate and Thr110 is stable as the case of E(Q)$_{indoline}$:PO. The explanation of the difference between the E(Q)$_{indoline}$:CO and E(A-A):CO complex may result from the extra interactions between the protein and indoline, which can supply a better stabilization of the structure in the active site, thus the substrate and the amino acids in the binding cavity are not required to modify the original conformations for the switch of the proton. In addition, the E(Q)$_{indoline}$:SB substrate with a protonated SB nitrogen is one of the potential forms to bind to TRPS. The attractions between the protein and the key atoms of E(Q)$_{indoline}$:SB, such the CO and PN, all maintain. However, the sidechain of Gln114 forms interactions with the PO instead of exposing to solvent (Figure 4.9(SB)), which gives an extra attraction between the substrate and TRPS to promote the pyridine ring twist a little bit. These motions are similar as the binding of E(A-A):SB, which explains the lower population of E(Q)$_{indoline}$:SB.

When we added one proton to PN and created a E(Q)$_{indoline}$:PO_PN complex, the hydroxyl sidechain of Ser377 rotates away from the PN, which results in missing H-

bonds between the PN and Ser377 (Figure 4.9(PO_PN)). The proton at the PN gives a crowded environment around the PN, which promotes the PLP ring moves away from Ser377 to search a bigger space. This movement of the down ring in the substrate causes the instability of the carboxyl group. The absence of the part of H-bonds between the CO and Thr110 was found again. The salt-bridge between Arg141 and Asp305 fades out and water molecules moves into the binding site (Figure 4.10). The RMSD indicates that the order of stabilization of the bound-substrates is $E(Q)_{indoline}$:PO $>$ $E(Q)_{indoline}$:CO $>$ $E(Q)_{indoline}$:SB $>$ $E(Q)_{indoline}$:PO_PN, which agrees with what the crystal studies suggested that the protonated PN is a less likely structure.
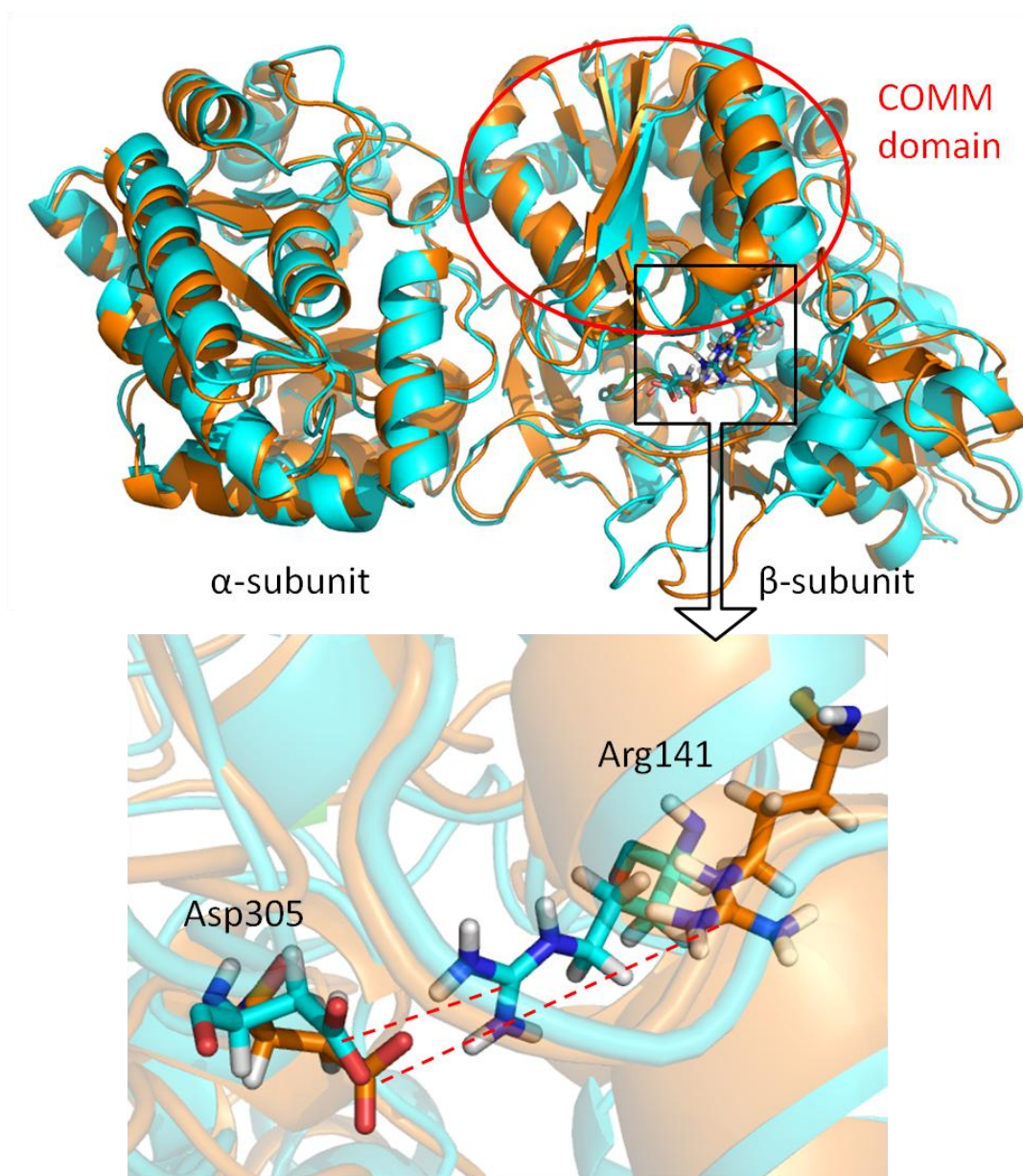
Figure 4.10. The comparison of the salt-bridge distance between $E(Q)_{indoline}$:PO and $E(Q)_{indoline}$:PO_PN complex. The up figure shows the structural alignment of the $E(Q)_{indoline}$:PO (cyan) and $E(Q)_{indoline}$:PO_PN (orange) complex. The down figure shows the salt-bridge (red dash line) forms between Arg141 and Asp305 in the simulation of the $E(Q)_{indoline}$:PO complex.

Another $E(Q)_{indole}$ analogue is $E(Q)_{2AP}$, which includes a functional group of 2-aminophenol to simulate the presence of indole. The hydroxyl group of the 2-

aminophenol forms H-bonds with Glu109, and the hydrophobic sidechains of Thr190, Leu166 and Phe306 interact with the ring conformation of the 2-aminophenol (Figure 4.11(PO)), which provides stable non-covalent interactions between TRPS and the PLP substrates. The simulations of the TRPS in complex with $E(Q)_{2AP}$:PO, $E(Q)_{2AP}$:CO and $E(Q)_{2AP}$:SB show that the substrates can well stay in the binding pocket and the attractions between the protein and the PG, CO and PN all retain during the 50-ns simulations (Figure 4.11(CO) and (SB)). The residue, Gln114, can move either toward solvent or the substrate in the $E(Q)_{2AP}$:CO and $E(Q)_{2AP}$:SB simulations. However, while a proton is added to the PN, the simulation of $E(Q)_{2AP}$:PO_PN complex displays missing H-bonds between the PN and Ser377 (Figure 4.11(PO_PN)). The binding of the $E(Q)_{2AP}$ analogues to TRPS is similar as the motions of the multiple $E(Q)_{indoline}$ complexes in our study.
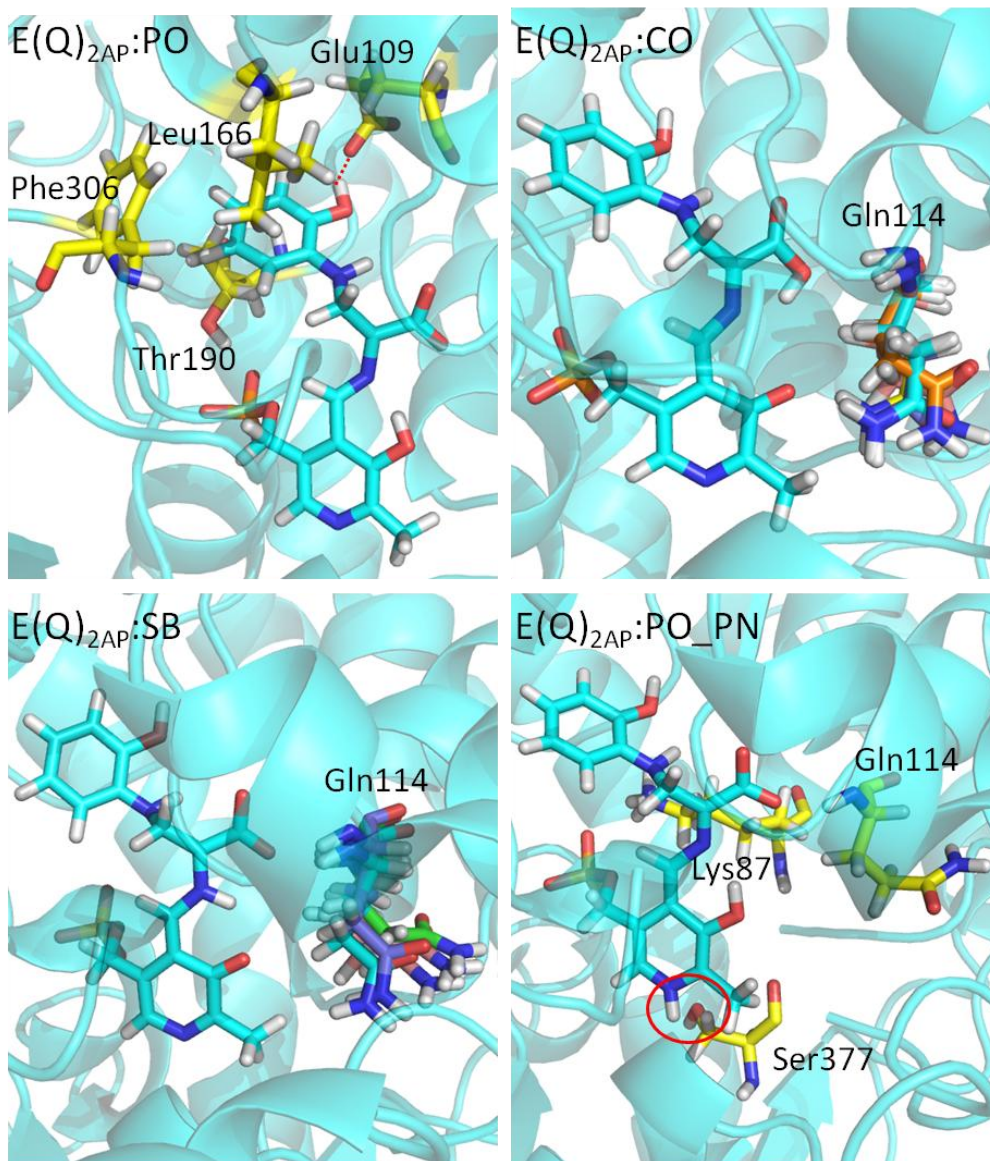
Figure 4.11. Protonation states of $E(Q)_{2AP}$. (PO) The residues interacting with the 2-aminophenol ring of $E(Q)_{2AP}$:PO are shown in a yellow bond representation. (CO) and (SB) The residue Gln114 is flexible during the 50 ns MD simulations. The sidechain of Gln114 can move toward either the PO or solvent depending on protonation state. The conformations of Gln114 at 10, 20, 30, 40, and 50 ns are shown in different colors. (PO_PN) The key residues around $E(Q)_{2AP}$:PO_PN are shown in yellow bond. The red circle indicates the missing H-bond between the PN and Ser377.

4.3.4 Roles of the pyridine nitrogen in PLP catalysis

Because TRPS precludes a proton at the PN, to support the short life-time of carbanionic intermediates to catalyze the nitrogen metabolism, remodulating the charge on the PN is necessary to provide a great delocalization system of negative charge in π-electrons [37]. The DFT computational study showed the distribution of atomic charges in PLP derivatives with a protonated and unprotonated PN [38]. However, the consideration of the substrate and only few atoms nearby is not enough to capture the full picture of protein-substrate interactions and overall conformational changes and dynamics of substrates binding to the PLP enzymes. Therefore, to study if the charge distribution on the PN plays a crucial role in controlling reaction specificity in PLP enzymes, we manually assigned the negative charge from -0.4 to -0.9 at the PN of $E(Q)_{indoline}$, but the total charges of all substrates were fixed at -4. Figure 4.12 shows the rotation of Ser sidechain with the changes of the charge at the PN. The H-bonds between the PN and the hydroxyl group of Ser377 are fully missing while the charge on the PN is -0.4. Although the -0.5, -0.6 and -0.7 charge of the PN could hold the H-bonds in the beginning of the MD simulations, the sidechain of Ser377 still rotates away at 20-40 ns. When stronger negative charges are on the PN, such as -0.75 and -0.8, the Ser sidechain could rotate to the back side in the middle of the MD simulations, while both sides of the Ser sidechain can hold the attractions with $E(Q)_{indoline}$. The MD conformations of Ser377 reveal the best match to the crystal structure while the PN charge goes down to -0.9. This indicates that the charge distribution on the pyridine ring is critical to the local stabilization of carbanionic intermediates such as $E(Q)_{indoline}$, and the protein conformations near the β-

178

active site could also be changed with the modification of a small change on the PN. Compared to E(Ain) and E(A-A) which the attractions between the substrates and Ser377 could maintain when the PN charge is from -0.3 to -0.5, the E(Q)$_{indoline}$ and E(Q)$_{2AP}$ species require more negative charges shift down to the PN.
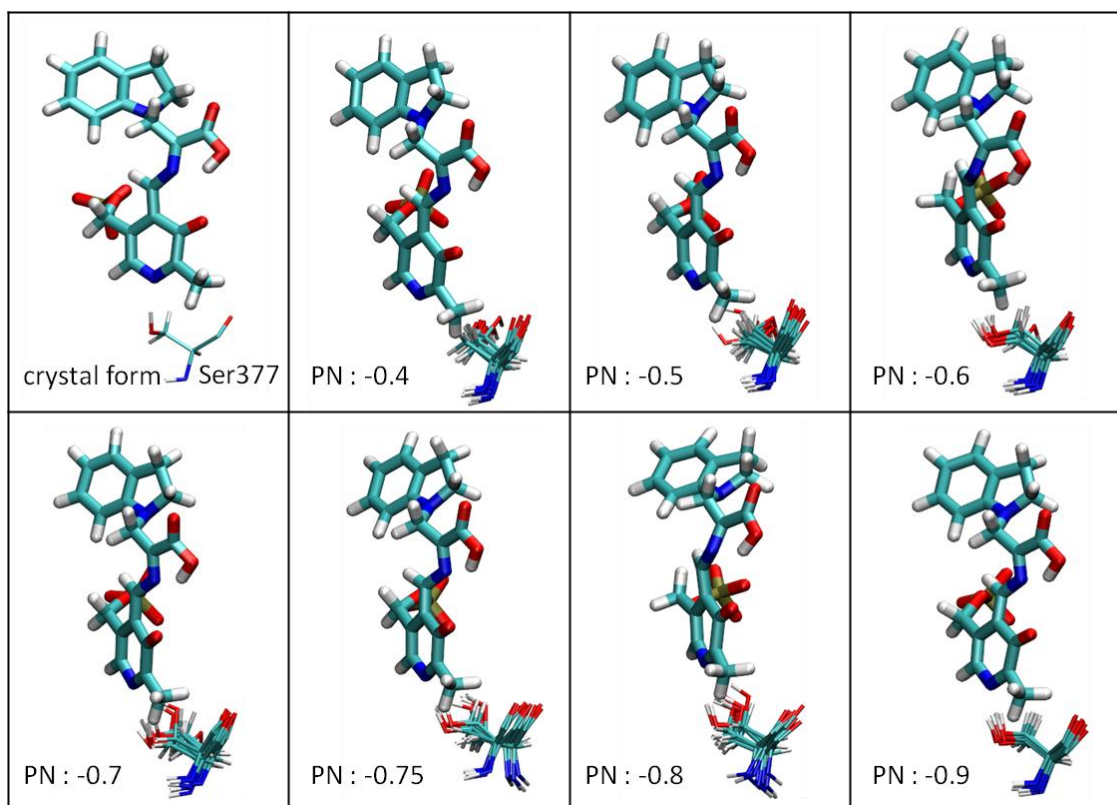
Figure 4.12. The rotation of the Ser377 sidechain in response to changes of the charge on the PN. E(Q)$_{indoline}$ is presented in the bound representation. The MD snapshots of Ser377 are shown every 5 ns.

4.3.5 Roles of the phosphate group in PLP catalysis

Phosphorylation is the most common method to modify proteins or substrates to regulate the up- or down-stream reactions [39]. Thus, the phosphate group has been a crucial inorganic modifier in a wide variety of protein-ligand systems. The PLP-dependant

enzymes play a key role in amino acid metabolism with a phosphorylated cofactor. The mutations of the 5'-phosphate group reported an effect on the PLP binding [40]. Moreover, in literature, the binding site of the PG of the most PLP enzymes are similar that the PG of the bound PLP substrates is surrounded by Gly residues and the oxygen atoms of the PG bind to the Gly mainchain by a series of H-bonds [40]. Although the deprotonated PG has been a paradigm in most systems of PLP catalysis, it still remains unknown how the protonation state of the PG controls the microenvironment in structural stability and catalysis mechanisms. Our study of the protonated and unprotonated PG in E(A-A) complexes show that, compared to E(A-A):PO, all interactions between E(A-A):PO_PG and the protein maintain except the missing H-bonds between the carboxylate group of the substrate and Thr110, His115 (Figure 4.6). It seems that the substrate with a protonated PG binds well, however, through the alignment of E(A-A):PO and E(A-A):PO_PG (Figure 4.13), the loops containing Thr190 and Ser235 shift toward the substrate in order to maintain the attractions between the PG and the residues nearby, which also induces the movement of Gln114 loop and causes the missing H-bonds around the CO of E(A-A):PO_PG, while the protein sidechain near Ser377 and Lys87 does not change significantly. The conformational changes of the loops around the PG may be attributed to the negative charge of the PG modified from -2 to -1 and fewer oxygen atoms available to interact with the protein when the PG of the substrate presents in protonated form. Although TRPS adjusts local geometries in the binding pocket to tolerate the protonation of the PG, the overall structure of the β-subunit is gradually changing from the closed to open conformations. Since the salt-bridge between Arg141

and Asp305 cannot retain (Figure 4.8). Therefore, the protonated PG not only alters the local structure of the binding site but also changes overall protein dynamics and water motions, which is not coherent with the crystal study [26].
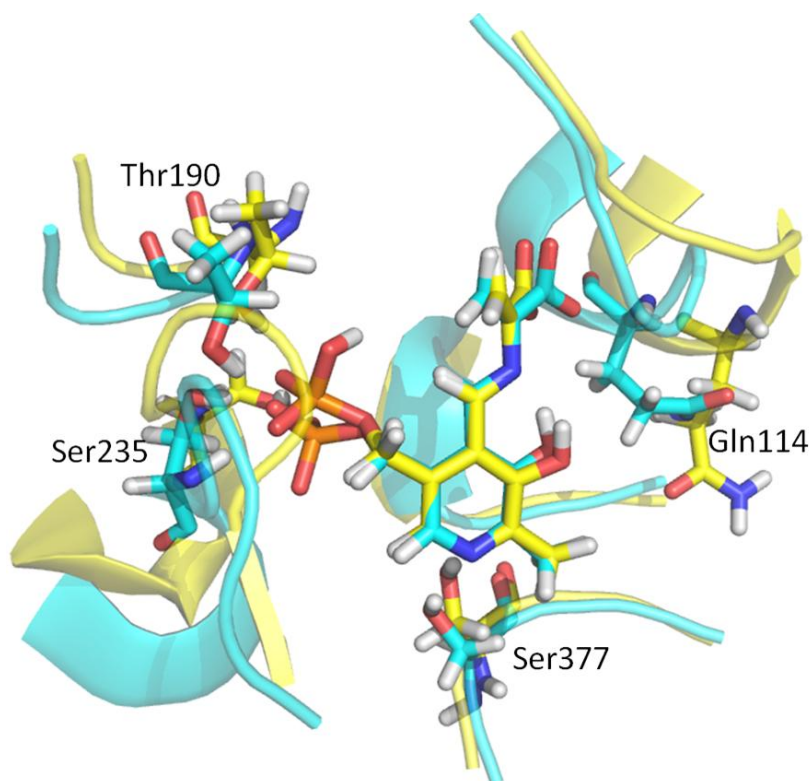
Figure 4.13. Alignment of E(A-A):PO (cyan) and E(A-A):PO_PG (yellow) complex at 50 ns. The residues within 7 Å of E(A-A) are shown in the cartoon representation. The keys residues and the E(A-A) species are shown in the stick form.

4.4 Conclusions

In this study, we applied MD simulations to investigate the protein dynamics and structural stability in the metabolic importance of the PLP-dependent enzyme, TRPS, through switching a single proton at different functional groups of the PLP derivatives. Results discussed in this work illustrate the significance of the key substructures of each PLP analogue in enzyme catalysis. The unprotonated PG with -2 charges can hold most

of the protein-substrate interactions. The PN with unprotonated form can form H-bonds with Ser377 that stabilizes the PLP ring structure. In order to create a proper shape for catalysis, the electrons could shift down to the PN in the binding of $E(Q)_{indoline}$ and $E(Q)_{2AP}$ to remodel the electrostatic and chemical environment of the substrate. The protonation of the PO could affect the motions of Gln114 that the polar sidechain of Gln114 tends to move toward the unprotonated PO, which induces the open conformations of the β-subunit in the binding of E(Ain) to TRPS. The CO typically forms stable H-bonds with Thr110, Gly111 and His115. The protonated PG, PN or CO could result in the missing of these H-bonds, which further alters the motions of the COMM domain and causes the open of the β-subunit in TRPS. Through thoroughly mapping the conformational changes of TRPS and its substrates with different protonation states, we hope this study can gain great insights into how the proteins stabilize the complex conformations with different PLP ligands.

4.5 References

1.   Kirby AJ. Efficiency of proton transfer catalysis in models and enzymes. Acc Chem Res. 1997;30(7):290-6. doi: 10.1021/ar960056r. PubMed PMID: WOS:A1997XL14200004.

2.   Hartwell E, Hodgson DRW, Kirby AJ. Exploring the limits of efficiency of proton-transfer catalysis in models and enzymes. J Am Chem Soc. 2000;122(38):9326-7. doi: 10.1021/ja002335m. PubMed PMID: WOS:000089603400045.

3.   Schowen KB, Limbach HH, Denisov GS, Schowen RL. Hydrogen bonds and proton transfer in general-catalytic transition-state stabilization in enzyme catalysis. Biochim Biophys Acta, Bioenerg. 2000;1458(1):43-62. doi: 10.1016/s0005-2728(00)00059-1. PubMed PMID: WOS:000087290100004.

4.   Hur O, Niks D, Casino P, Dunn MF. Proton transfers in the beta-reaction catalyzed by tryptophan synthase. Biochemistry. 2002;41(31):9991-10001. doi: 10.1021/bi025568u. PubMed PMID: WOS:000177228000026.

5.   Schultz BE, Chan SI. Structures and proton-pumping strategies of mitochondrial respiratory enzymes. Annu Rev Biophys Biomol Struct. 2001;30:23-65. doi: 10.1146/annurev.biophys.30.1.23. PubMed PMID: WOS:000169531600003.

6.   Alonso H, Cummins PL, Gready JE. Methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase (MeTr): Protonation state of the ligand and active-site residues. J Phys Chem B. 2009;113(44):14787-96. doi: 10.1021/jp900181g. PubMed PMID: WOS:000271105600031.

7.   deDios AC, Oldfield E. Recent progress in understanding chemical shifts. Solid State Nucl Magn Reson. 1996;6(2):101-25. doi: 10.1016/0926-2040(95)01207-9. PubMed PMID: WOS:A1996UQ60500001.

8.   McDermott A, Polenova T. Solid state NMR: New tools for insight into enzyme function. Curr Opin Struct Biol. 2007;17(5):617-22. doi: 10.1016/j.sbi.2007.10.001. PubMed PMID: WOS:000251348200017.

9.   Oldfield E. Quantum chemical studies of protein structure. Philos Trans R Soc Lond B Biol Sci. 2005;360(1458):1347-61. doi: 10.1098/rstb.2003.1421. PubMed PMID: WOS:000231317500015.

10.   Hyde CC, Ahmed SA, Padlan EA, Miles EW, Davies DR. 3-Dimensional structure of the tryptophan synthase alpha-2-beta-2 multienzyme complex from salmonella-typhimurium. J Biol Chem. 1988;263(33):17857-71. PubMed PMID: WOS:A1988R004800099.

11.   Bahar I, Jernigan RL. Cooperative fluctuations and subunit communication in tryptophan synthase. Biochemistry. 1999;38(12):3478-90. doi: 10.1021/bi982697v. PubMed PMID: WOS:000079510700004.

12.   Dunn MF, Niks D, Ngo H, Barends TRM, Schlichting I. Tryptophan synthase: The workings of a channeling nanomachine. Trends Biochem Sci. 2008;33(6):254-64. doi: 10.1016/j.tibs.2008.04.008. PubMed PMID: WOS:000257451300006.

13.   Rhee S, Parris KD, Hyde CC, Ahmed SA, Miles EW, Davies DR. Crystal structures of a mutant (beta K87T) tryptophan synthase alpha(2)beta(2) complex with ligands bound to the active sites of the alpha- and beta-subunits reveal ligand-induced

conformational changes. Biochemistry. 1997;36(25):7664-80. doi: 10.1021/bi9700429. PubMed PMID: WOS:A1997XG90100005.

14.   Schneider TR, Gerhardt E, Lee M, Liang PH, Anderson KS, Schlichting I. Loop closure and intersubunit communication in tryptophan synthase. Biochemistry. 1998;37(16):5394-406.   doi: 10.1021/bi9728957.   PubMed   PMID: WOS:000073515500008.

15.   Ferrari D, Niks D, Yang LH, Miles EW, Dunn MF. Allosteric communication in the tryptophan synthase bienzyme complex: Roles of the beta-subunit aspartate 305 arginine 141 salt bridge. Biochemistry. 2003;42(25):7807-18. doi: 10.1021/bi034291a. PubMed PMID: WOS:000183821200019.

16.   Sachpatzidis A, Dealwis C, Lubetsky JB, Liang PH, Anderson KS, Lolis E. Crystallographic studies of phosphonate-based alpha-reaction transition-state analogues complexed to tryptophan synthase. Biochemistry. 1999;38(39):12665-74. doi: 10.1021/bi9907734. PubMed PMID: WOS:000083018500016.

17.   Miles EW. Tryptophan synthase: Structure, function, and subunit interaction. Adv Enzymol Relat Areas Mol Biol. 1979;49:127-86. PubMed PMID: MEDLINE:400853.

18.   Dunn MF, Aguilar V, Brzovic P, Drewe WF, Houben KF, Leja CA, et al. The tryptophan synthase bienzyme complex transfers indole between the alpha-sites and beta-sites via a 25-30 a long tunnel. Biochemistry. 1990;29(37):8598-607. doi: 10.1021/bi00489a015. PubMed PMID: WOS:A1990DZ19500015.

19.   Pan P, Woehl E, Dunn MF. Protein architecture, dynamics and allostery in tryptophan synthase channeling. Trends Biochem Sci. 1997;22(1):22-7. doi: 10.1016/s0968-0004(96)10066-9. PubMed PMID: WOS:A1997WE01700008.

20.   Miles EW, Rhee S, Davies DR. The molecular basis of substrate channeling. J Biol Chem. 1999;274(18):12193-6. doi: 10.1074/jbc.274.18.12193. PubMed PMID: WOS:000080056800001.

21.   Ngo H, Kimmich N, Harris R, Niks D, Blumenstein L, Kulik V, et al. Allosteric regulation of substrate channeling in tryptophan synthase: Modulation of the L-Serine reaction in stage I of the ss-reaction by alpha-site ligands. Biochemistry. 2007;46(26):7740-53. doi: 10.1021/bi7003872. PubMed PMID: WOS:000247486800007.

22.   Drewe WF, Dunn MF. Detection and identification of intermediates in the reaction of L-serine with escherichia-coli tryptophan synthase via rapid-scanning ultraviolet visible spectroscopy. Biochemistry. 1985;24(15):3977-87. doi: 10.1021/bi00336a027. PubMed PMID: WOS:A1985ANQ9100027.

23.   Drewe WF, Dunn MF. Characterization of the reaction of L-serine and indole with escherichia-coli tryptophan synthase via rapid-scanning ultraviolet visible spectroscopy. Biochemistry. 1986;25(9):2494-501. doi: 10.1021/bi00357a032. PubMed PMID: WOS:A1986C353900032.

24.   Lai J, Niks D, Wang Y, Domratcheva T, Barends TRM, Schwarz F, et al. X-ray and NMR crystallography in an enzyme active site: The indoline quinonoid intermediate in tryptophan synthase. J Am Chem Soc. 2011;133(1):4-7. doi: 10.1021/a106555c. PubMed PMID: WOS:000286351100002.

25.   Caulkins BG, Bastin B, Yang C, Neubauer TJ, Young RP, Hilario E, et al. Protonation states of the tryptophan synthase internal aldimine active site from solid-state NMR spectroscopy: Direct observation of the protonated schiff base linkage to pyridoxal-5 '-phosphate. J Am Chem Soc. 2014;136(37):12824-7. PubMed PMID: CCC:000342119700006.

26.   Niks D, Hilario E, Dierkers A, Ngo H, Borchardt D, Neubauer TJ, et al. Allostery and substrate channeling in the tryptophan synthase bienzyme complex: Evidence for two subunit conformations and four quaternary states. Biochemistry. 2013;52(37):6396-411. doi: 10.1021/bi400795e. PubMed PMID: WOS:000330099600011.

27.   Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005;26(16):1668-88. doi: 10.1002/jcc.20290. PubMed PMID: WOS:000233021400002.

28.   Goetz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. J Chem Theory Comput. 2012;8(5):1542-55. doi: 10.1021/ct200909j. PubMed PMID: WOS:000303628400004.

29.   Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. WIREs Comput Mol Sci. 2013;3(2):198-210. doi: 10.1002/wcms.1121. PubMed PMID: WOS:000318044900007.

30.   Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein

backbone parameters. Proteins: Struct, Funct, Bioinf. 2006;65(3):712-25. doi: 10.1002/prot.21123. PubMed PMID: WOS:000241247100017.

31.   Ozpinar GA, Peukert W, Clark T. An improved generalized AMBER force field (GAFF) for urea. J Mol Model. 2010;16(9):1427-40. doi: 10.1007/s00894-010-0650-7. PubMed PMID: WOS:000280640700001.

32.   Gilson MK, Gilson HSR, Potter MJ. Fast assignment of accurate partial atomic charges: An electronegativity equalization method that accounts for alternate resonance forms. J Chem Inf Comput Sci. 2003;43(6):1982-97. doi: 10.1021/ci034148o. PubMed PMID: WOS:000186848700033.

33.   Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79(2):926-35. PubMed PMID: WOS:A1983QZ31500046.

34.   Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A Smooth particle mesh Ewald method. J Chem Phys. 1995;103(19):8577-93. PubMed PMID: WOS:A1995TE36400026.

35.   Salomon-Ferrer R, Goetz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. J Chem Theory Comput. 2013;9(9):3878-88. doi: 10.1021/ct400314y. PubMed PMID: WOS:000330096800004.

36.   Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-interaction of cartesian equations of motion of a system with constraints - molecular-dynamics of N-alkanes. J Comput Phys. 1977;23(3):327-41. PubMed PMID: WOS:A1977CZ25300007.

37.   Griswold WR, Toney MD. Role of the pyridine nitrogen in pyridoxal 5'-phosphate catalysis: Activity of three classes of PLP enzymes reconstituted with deazapyridoxal 5'-phosphate. J Am Chem Soc. 2011;133(37):14823-30. doi: 10.1021/ja2061006. PubMed PMID: WOS:000295148100055.

38.   Casasnovas R, Salva A, Frau J, Donoso J, Munoz F. Theoretical study on the distribution of atomic charges in the Schiff bases of 3-hydroxypyridine-4-aldehyde and alanine. The effect of the protonation state of the pyridine and imine nitrogen atoms. Chem Phys. 2009;355(2-3):149-56. doi: 10.1016/j.chemphys.2008.12.006. PubMed PMID: WOS:000263628800009.

39.   Nardozzi JD, Lott K, Cingolani G. Phosphorylation meets nuclear import: A review. Cell Commu Signal. 2010;8:32. doi: 10.1186/1478-811x-8-32. PubMed PMID: WOS:000286281200001.

40.   Schnackerz KD, Andi B, Cook PF. P-31 NMR spectroscopy senses the microenvironment of the 5'-phosphate group of enzyme-bound pyridoxal 5'-phosphate. Biochim Biophys Acta, Proteins Proteomics. 2011;1814(11):1447-58. doi: 10.1016/j.bbapap.2011.02.001. PubMed PMID: WOS:000296952800006.

Chapter 5 Insights from Umbrella Sampling: Dissociation Path, Conformational Change and Conformation Preparation

5.1 Introduction

Free energy is an important quantity that characterizes chemical and biological processes. The change of free energy governs the directionality and extent of chemical reactions. Free energy decomposes into enthalpy and entropy, where determination of entropy is challenging both experimentally and computationally. For these reasons, it is one central task for computational chemist to achieve accurate calculation of free energy, especially the free energy profile along a chemical process [1-3]. A variety of free energy calculation methods have been developed in the past decades, such as perturbation theory [4], thermodynamic integration [5], umbrella sampling [6], and partition function from density of states [7], and provided insights into various chemical and biological systems [8, 9]. Among these methods, umbrella sampling (US) is a conceptually straightforward, computationally efficient and reliable one that computes the potential of mean force (PMF) based on rigorous probability calculations [6]. It requires a well-defined reaction coordinate (RC) represented by one or a few collective variables (CVs) [10]. Intensive conformational sampling is performed by enforcing external restraints at the conformations along the RC within a series of successive overlapping windows. Finally, the PMF can be constructed by removing the external restraints.

US is particularly suitable for computing the PMF of ligand-receptor dissociation driven by non-bonding interactions. It has long been applied to calculate binding affinity of

various receptor-ligand systems, ranging from small chemical molecular system [11] to large biological systems [12]. Furthermore, by sampling local energy barriers along dissociation path, it can provide thermodynamic details for molecular recognition. However, US itself does not provide the dissociation path. Therefore, enhanced sampling methods, such as steered molecular dynamics simulation (SMD) [13-16], adaptive biasing force (ABF) [17] and metadanamics [18], are often used to provide the dissociation pathway that can be used as initial conformations for the intensive biased sampling for US. A number of methods have been proposed to improve the accuracy of US, such as the use the constrained schemes to alleviate sampling limitation [19-21], or the combination with other method like Markov model to improve the convergence [22]. Despite the natural connection between these methods and US, how accurately the conformations from these enhanced sampling methods can resemble true dissociation path, or whether they can provide reasonable initial conformations for US, have rarely been studied. Moreover, how one dissociation path compared to another in the dissociation pathway ensemble may affect the results from US remains unclear. Here we answered these two questions by studying influential factors of US using two systems, β-Cyclodextrin (β-CD)-ligand complexes and p38α-inhibitor system.

β-CD is a cyclic oligosaccharide containing seven glucopyranose units linked through 1,4 α glycosidic bonds, thus forming a truncated conical structure. With a hydrophobic inner surface and hydrophilic rims containing primary and secondary hydroxyl groups, it is able to accommodate small hydrophobic molecules, therefore enhancing the solubility and bioavailability of such molecules (Figure 5.1). The cavity of β-CD also resembles a

protein binding site, and this makes it a good host molecule to study ligand-receptor binding. Because of these properties, β-CD and its derivatives have been widely used in drug delivery, pharmaceutical, food, and chemical industries [23-29]. Therefore, a large amount of experimental reference data is available [30-37].
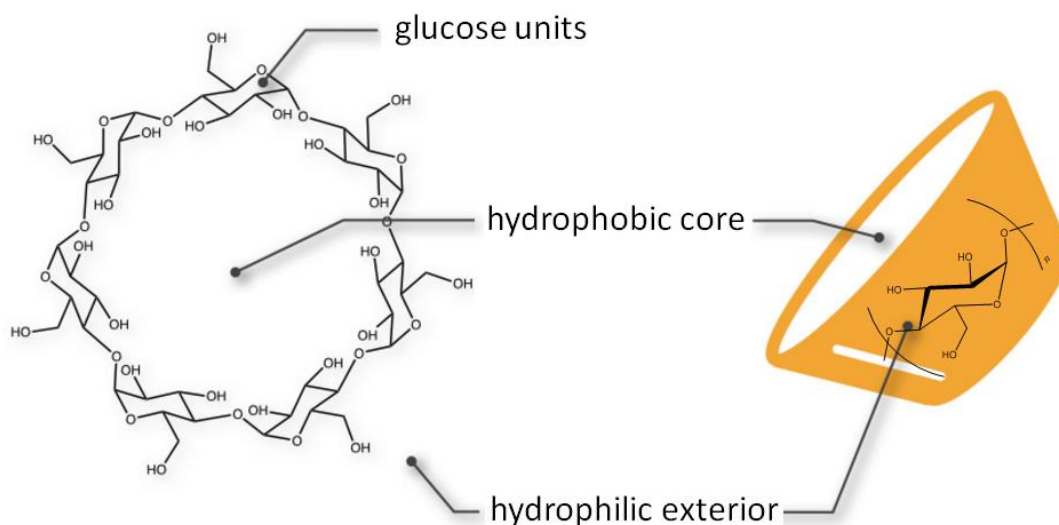
Figure 5.1. Structure of β-cyclodextrin (β-CD). β-CD is consisted of seven glucopyranose units with a hydrophobic core and hydrophilic exterior.

p38α is the major isoform of p38, which belongs to the mitogen-activated protein kinases (MAPKs), a super-family of enzymes that regulate a variety of biological processes, such as proliferation, gene expression, differentiation and apoptosis [38, 39]. p38α has been a drug target for treating various inflammatory diseases, including rheumatoid arthritis, asthma, and cardiovascular disease [40-42]. To develop new inhibitors, it is necessary to investigate the kinetics behaviors of inhibitors during dissociation process. Like all protein kinases, p38α consists of a N-terminal domain and a C-terminal domain that are connected via a hinge [43]. The activation loop that carries a DFG (Asp-Phe-Gly) motif

determines the opening or the closing of the binding cavity, where ATP binds during the activation process. The conformational change of activation loop can be characterized by different orientations of its Phe sidechain. In the active conformation, Phe is buried in αC helix (DFG-in), while in inactive conformation, Phe rotates away from the αC helix and projects into the ATP binding pocket (DFG-out). SB2 is a ligand of p38α that can bind to ATP binding site while activation loop adopts either DFG-in or DFG-out conformation [44] (Figure 5.2) and doesn't interfere with the conversion between DFG-in and DFG-out conformations of activation loop. This makes it the perfect candidate to study the influence of receptor conformational change on construction of free energy profile.

Figure 5.2. The comparison of the bound structures of SB2 in DFG-in and DFG-out conformations. The left figure shows the structure alignment of DFG-in (cyan, PDB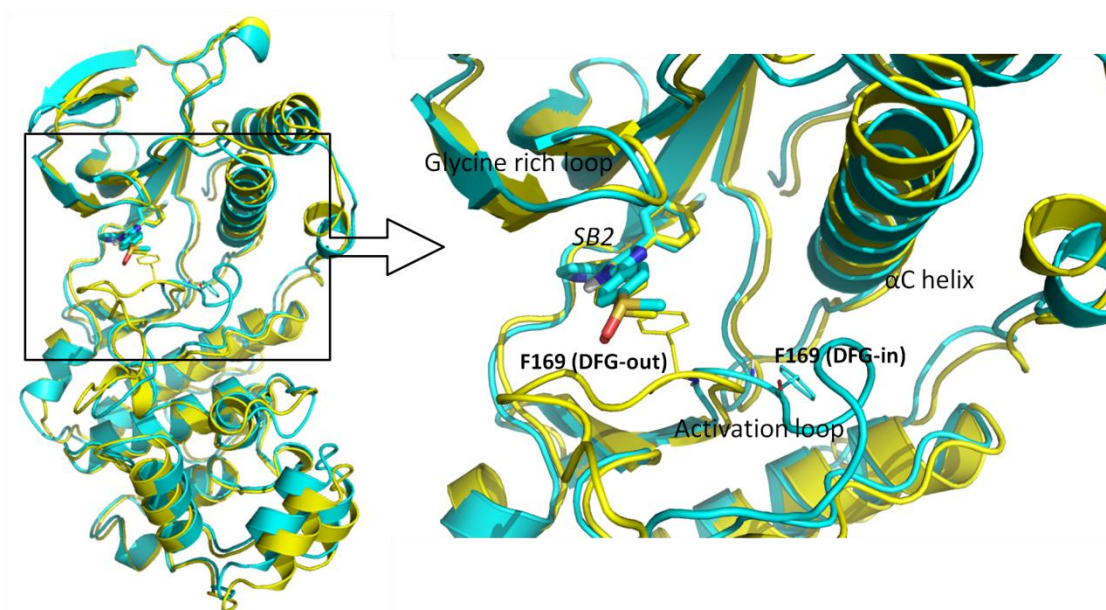 1A9U) and DFG-out (yellow, PDB 3GCP) conformations bound with ligand SB2. The right figure shows binding site structure of p38α-SB2 complex. The Phe169 from DFG motif is shown in thin licorice structure, ligand SB2 is shown in bold licorice structure.

In this work, we investigated how the PMF from US is affected by subtle changes in the dissociation pathways and conformational sampling methods that provide the initial conformations. We used dissociation of β-CD-aspirin, β-CD-1-butanol, and p38α-SB2 systems as examples. We utilized accelerated MD (AMD), SMD and manual pulling methods to provide the dissociation pathways as initial conformations for US, and investigated how these methods affect the PMF from US. By using different β-CD conformations as starting point for performing US, we found that the host conformation may fundamentally change the depth of the PMF from US, and the influence of initial conformation can hardly be removed by nanosecond level biased MD simulation, even for small systems like a β-CD complex. We also showed that AMD is a good tool to provide initial conformations for US simply by relaxing the conformations along dissociation pathways sampled by AMD using short MD simulations. However, SMD may not be a suitable method to provide the initial conformations, because the dissociation pathway sampled by SMD lacks important residue sidechain movement.

5.2 Materials and Methods

5.2.1 Structure preparation and parameters

β-cyclodextrin. We selected three β-CD conformations from previous MD simulations of β-CD-aspirin complexes and removed the ligand as initial conformation of β-CD. Two ligands, aspirin and 1-butanol, were selected to perform US along manually built dissociation pathways as detailed in later section (Figure 5.3). We used q4MD-CD force field for β-CD [45]. We manually built the ligand structures with Vega ZZ [46] and

computed the partial charges for them by using B3LYP/6-31+G(d,p) ChelpG calculations with Gaussian package [47] after optimizing the structures using the same settings. GAFF was used for the ligands.

p38α. SB2 is a ligand of p38α system that binds to both DFG-in and DFG-out conformations. We obtained p38α DFG-in conformation (PDB ID: 1A9U) [48] and DFG-out conformation (PDB ID: 3GCP) [49] from protein data bank (PDB). We built the missing activation loop (residues 173 to 184) of 3GCP by using the conformation from previous MD simulation of free DFG-out p38α (Chapter 3). Amber 99SB force field was used for proteins and GAFF was used for ligand SB2.

(A)

Conf 1

Conf 2

Conf 3

(B)

aspirin   1-butanol
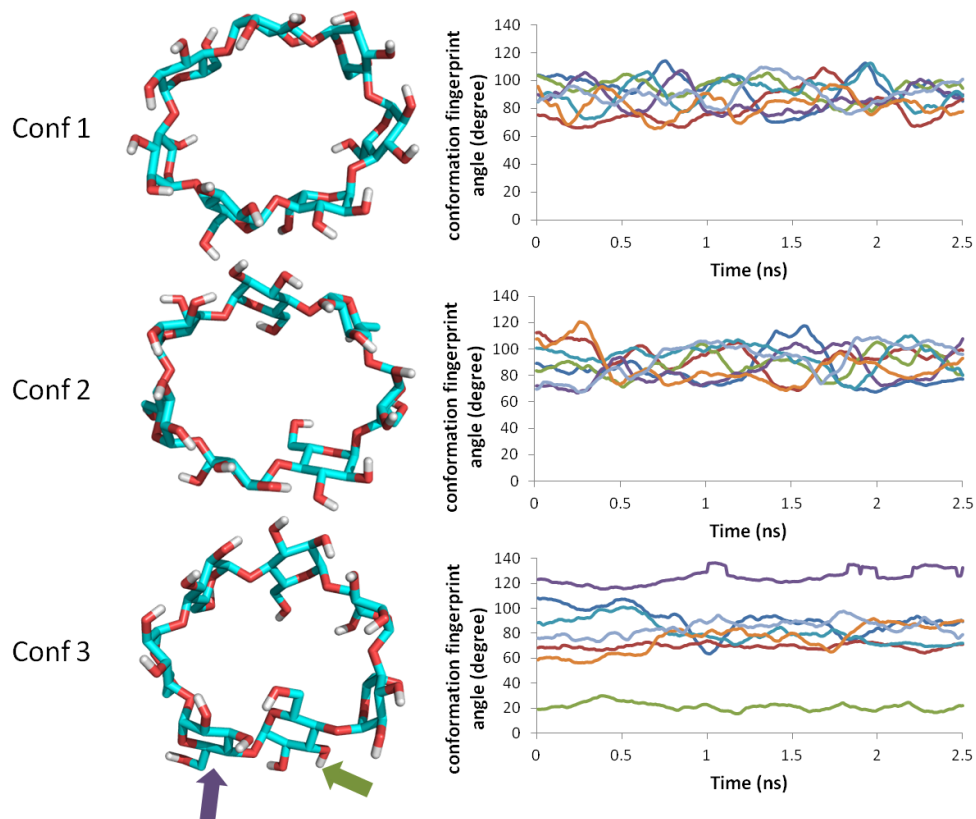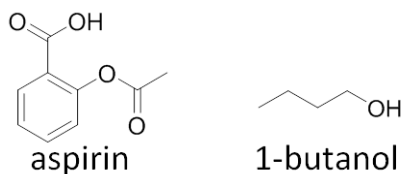
Figure 5.3. Structures of β-CD and ligands and fingerprint dihedral plot of three β-CD conformations. (A) Three different conformations of β-CD. The plots of their conformation fingerprint angles (defined in Figure 5.4) along trajectories are on the right. The fingerprint dihedral plots of β-CD are from biased MD of one US window where distance between center of mass (COM) of β-CD and aspirin is restrained to 10 Å. Before measurement of fingerprint angles, the trajectories were smoothed by averaging 100 forward and 100 backward frames on the concurrent frame throughout the whole trajectory to remove the noise. (B) Structures of aspirin and 1-butanol.

196

Table 5.1. Experimental kinetics and thermodynamics data of aspirin, 1-butanol complexed with β-CD, and SB2 complexed with p38α. K_D, k_{on} and k_{off} data for complexes were taken from [31, 34, 50], ΔG_{exp} of β-CD-1-butanol complex was taken from [35], ΔG_{exp} of β-CD-aspirin and p38α-SB2 complexes were calculated using ΔG_{exp} = RT ln K_D.

Complex system	K_D (nM)	k_{on} [M^{-1}s^{-1}]	k_{off} [s^{-1}]	ΔG_{exp} (kcal/mol)
β-CD-aspirin	1.81×10^6	7.20×10^8	1.30×10^6	-3.77
β-CD-1-butanol	1.36×10^8	2.80×10^8	3.80×10^7	-1.67
p38α-SB2	11.5	1.5×10^7	1.8×10^{-1}	-10.90
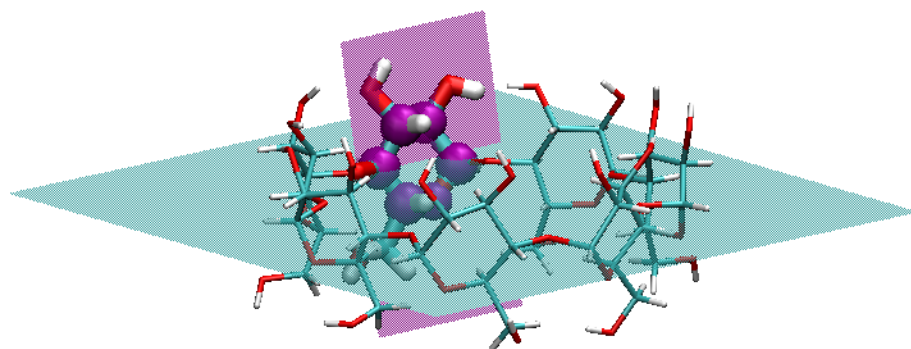
Figure 5.4. Representation of one fingerprint angle in β-CD. The regression plane of entire molecule is in shown in cyan color and the regression plane of the six atoms highlighted by purple balls in one glucose unit is shown in purple color. The fingerprint angle of one glucose unit is defined by the dihedral angle between the two regression planes.

5.2.2 Preparation of dissociation paths for US

For β-CD, we manually docked the ligands along the dissociation path. First, we put the center of mass (COM) of β-CD at the origin (0, 0, 0), and aligned its principal axes along X, Y, Z axes so that the primary cavity of β-CD faces the positive direction of X-axis. Then, we manually located the ligand so that its COM is also aligned at origin. By using this artificial bound conformation, we gradually moved the ligand along positive and negative of X-axis at a speed of 0.1 Å every step for 26 Å in both directions, until the ligand is fully dissociated. In this way, we obtained path A and B for ligand dissociation along the primary and secondary cavity of β-CD (Figure 5.5). We repeated this procedure for aspirin and 1-butanol in the three β-CD conformations respectively.
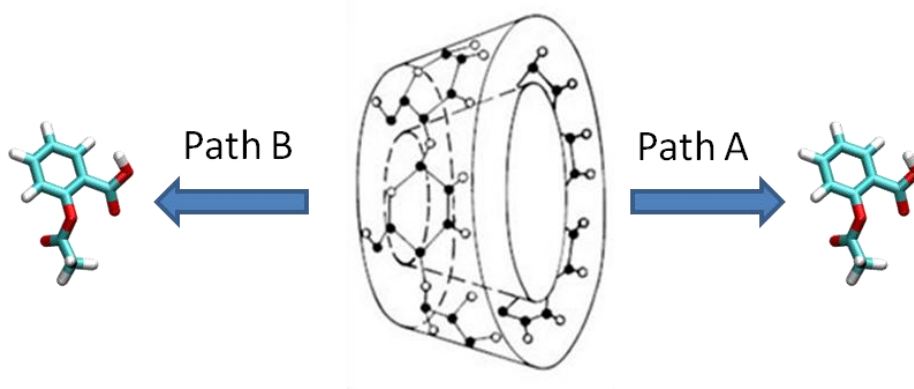
Figure 5.5. Dissociation pathways of β-CD complexes. Path A is from the primary cavity and path B is from secondary cavity of β-CD.

For p38α-SB2 system, we constructed the dissociation paths by using three ways, i.e. AMD, SMD, and manual pulling as used for β-CD. We obtained two paths from AMD

(path1 and path2), one path from SMD, one path from manual pulling for DFG-in conformation, and one path from AMD for DFG-out conformation.

5.2.3 Accelerated MD simulation

AMD, which introduces a continuous non-negative bias boost potential function $\Delta V(r)$ to the potential energy surface when the system potential is below a reference energy, to enhance the conformational sampling of biological systems, therefore lowering the local barriers to accelerate the calculation [51]. AMD uses following equations to alleviate the energy barriers,

$$V^*(r) = V(r) + \Delta V(r)$$

$$\Delta V(r) = \begin{cases} 0, & V(r) \geq E \\ \dfrac{(E - V(r))^2}{\alpha + (E - V(r))} & V(r) < E \end{cases}$$

where $V(r)$ is the original potential, E is the reference energy, and $V^*(r)$ is the modified potential. $\Delta V(r)$ is the boost potential, α is the acceleration factor.

The boost potential $\Delta V(r)$ can be applied to dihedral with input parameters (Ed, αD) and overall potential energy terms with input parameters (Ep, αP),

$$\Delta V(r) = \frac{(Ep - V(r))^2}{(\alpha P + Ep - V(r))} + \frac{(Ed - Vd(r))^2}{(\alpha D + Ed - Vd(r))}$$

For simulations of p38α, we applied both potential-boost and dihedral-boost.

5.2.4 Steered MD simulation

SMD simulation uses a time-dependent external force to drive the system to move in a predefined way [52]. The external force V(t) can be described as,

$$V(t) = k[x-x_0(t)]^2$$

where x and $x_0(t)$ are the CV in simulation and the predefined time-dependent track of the CV, k is a harmonic force constant. We selected the distances between Cα of Arg73 and CC2 of SB2 (Figure 5.6), which is also used to describe the RC in US, as the CV. $x_0(t)$ was set to move by 1.75 Å/ns and with a maximum at 33 Å. The force constant k was set to 10 kcal/mol Å^2. We equilibrated the bound state conformation for 100 ns using conventional MD, and then performed SMD for 10 ns at 300K and 1 bar in NPT ensemble. Temperature was maintained by Langevin thermostat.
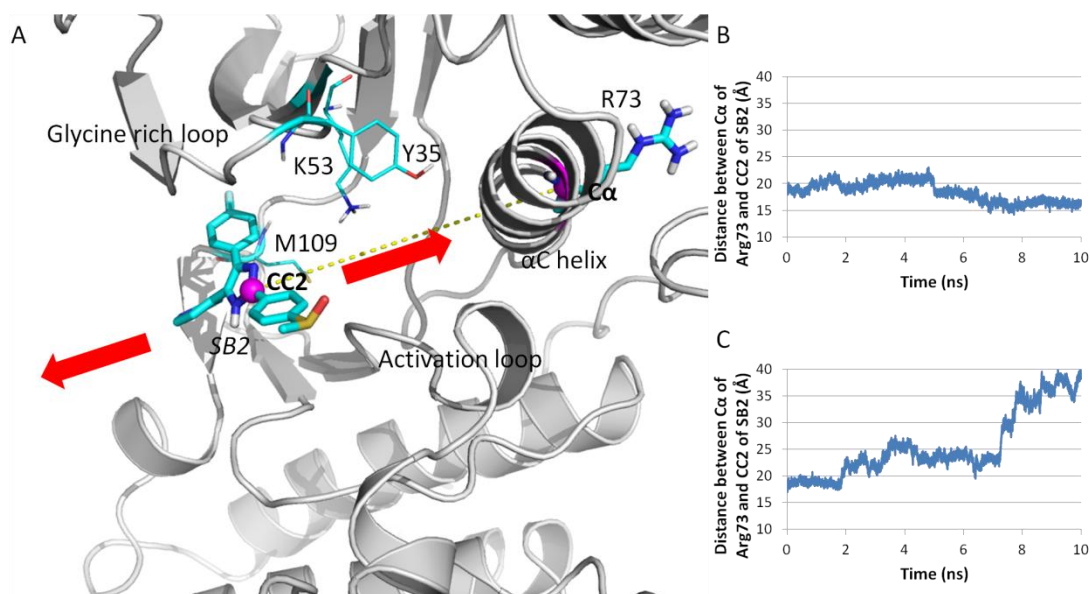
Figure 5.6. Reconstruction of dissociation path from AMD. Path 1 is built from AMD path that conformational relaxed by two 10 ns conventional MD. (A) SB2 in one of the two 10 ns conventional MD moves towards inside the cavity, while SB2 in the other conventional MD moves towards outside. Arg73 and SB2 are shown in bold licorice structure, Cα of Arg73 and CC2 of SB2 are indicated by purple ball structure, other interacting residues are shown in thin licorice structure. (B) SB2 moving inside the cavity indicated by the decreasing distance between SB2 and Arg73. (C) SB2 moving outside the cavity indicated by the increasing distance between SB2 and Arg73.

5.2.5 Manual pulling

In manual pulling, crystal structure of SB2 bound to p38α (1A9U) was used as reference conformation. Ligand SB2 was gradually moved along the vector of Cα of Arg73 and CC2 of SB2 towards outside of cavity at a speed of 0.25 Å every step for 16.75 Å, until SB2 was fully dissociated.

5.2.6 Umbrella sampling

US [53, 54] was performed to compute the free energy along the dissociation pathway. In US, a series of windows are evenly located along RC and intensive sampling in these

201

windows is achieved by enforcing an external biasing potential. The samplings in each window must overlap with adjacent windows, so that the unbiased PMF can be reproduced by removing the biasing potential. The external biasing potential u_i at window i is a harmonic function $u_i = k_i(r - r_i)^2$, where r_i is the reference position, and k_i is the harmonic force constant. All biased MD simulations were performed with Amber14 [55]. We used WHAM [56] to remove the biasing potential and reconstruct the PMF.

For β-CD, the distance between COMs of heavy atoms of β-CD and ligand was selected as the CV to represent the RC. As described in previous section, a total of 260 windows with 0.1 Å spacing along the RC were used to performed biased MD simulation. We minimized the initial conformation for each window for 1500 steps using generalized Born (GB) implicit solvation model [57] to remove clashes from manual pulling. Then we solvated β-CD-ligand complex with a 30-Å rectangular box of TIP3P water molecules using tleap module in AMBER14 [58]. After minimizing the system for 1000 steps, we equilibrated the water molecules at 298K for 1 ns with 1-fs timestep, and heated the entire system at 200K, 250K and 298K for 150 ps. In the minimization and equilibration steps, a harmonic force constant of 1000 kcal/mol $\cdot$Å^2 was used to restrain the ligand at the correct window. Finally, we performed 2.5 ns production run at 298K with restraint using harmonic force constant of 100 kcal/mol $\cdot$Å^2. In WHAM, the bin size was set to 0.05 Å. The tolerance for iteration was set to 0.0001. The temperature was set to 298 K.

For p38α, the distance between Cα of Arg73 and CC2 of SB2 was selected as the RC. We set up five sets of biased MD simulations for computing PMF using US as detailed in the

previous section. Note that we performed 10 ns conventional MD simulations on the conformations along the two DFG-in and one DFG-out paths yielded from AMD to equilibrate the conformations before we used them as initial conformations in each window for the biased simulations for US (Figure 5.6). From the SMD trajectory and the conventional MD that used to relax conformations of AMD, we selected the conformations that fall into the windows along RC and have minimal SB2 root mean square deviations (RMSD) compared to the bound state SB2 as initial conformations for the corresponding window. For the manual pulling path, since the system is not optimized or solvated, we performed these following steps on the conformations we obtained before proceeding to biased MD sampling. We optimized the conformations along the path by minimizing the hydrogen atom, sidechains, and entire complex for 500, 5000 and 5000 steps respectively for the same reason as we did for β-CD. Next, we solvated the conformation using TIP3P water model [58] so that the edge of the water box is at least 12 Å away from the solutes. We also added Na^+ ions to neutralize the system. We optimized the water molecules and the entire system for 10000 and 20000 steps respectively. After equilibrating the solvate for 40 ps at 298K in NPT, we heated the system from 250K to 300K gradually. In the minimization and equilibration process, the external force constant was 500 kcal/mol $\cdot Å^2$. In total, 62 windows from SMD, 68 windows from path 1, 2, manual pulling, and 71 windows from DFG-out path were evenly located every 0.25 Å along the RC. For all five sets of conformations, we performed a production run for 10 ns at 300K with an external restraint of 5 kcal/mol $Å^2$.

In WHAM, the bin size was set to 0.2 Å. The tolerance for iteration was set to 0.0001. The temperature was set to 300 K.

5.3 Results and Discussions

5.3.1 Unbinding process of β-CD complex system

The PMFs of path A and B are constructed by using WHAM from bound state to free state, and combined so that the free states of two paths of have the same free energy. The combined PMFs of β-CD-aspirin and β-CD-1-butanol are shown in Figure 5.7. Comparing Conf 1 and 2 with Conf 3 of the β-CD complex systems, it's clear that the host conformations have remarkable impact on the shape of the PMF. For aspirin, the binding affinities of Conf 1 and 2 are -2.8 and -2.7 kcal/mol which are similar, while the binding affinity of Conf 3 is only -1.8 kcal/mol. The binding affinities from Conf 1 and 2 are 1 kcal/mol less favorable than experimental value -3.77 kcal/mol. For 1-butanol, the binding affinities of Conf 1, 2 and 3 are -1.7, -1.7, -1.3 kcal/mol, respectively. The binding affinities of Conf 1 and 2 agree with experimental value (-1.67 kcal/mol). We only considered three β-CD conformations but in reality, β-CD can adopt much more conformations in the ligand association and dissociation. Also, in the biased MD simulations, the external harmonic potential was only applied on the direction of the RC which was represented by the COM distance CV, and the ligand is free to move on the sphere with a radius of the COM distance in that window, resulting in unrestrained deviation from the X-axis. For these two reasons, the computed binding affinities do not rigorously agree with the experimental values.

aspirin
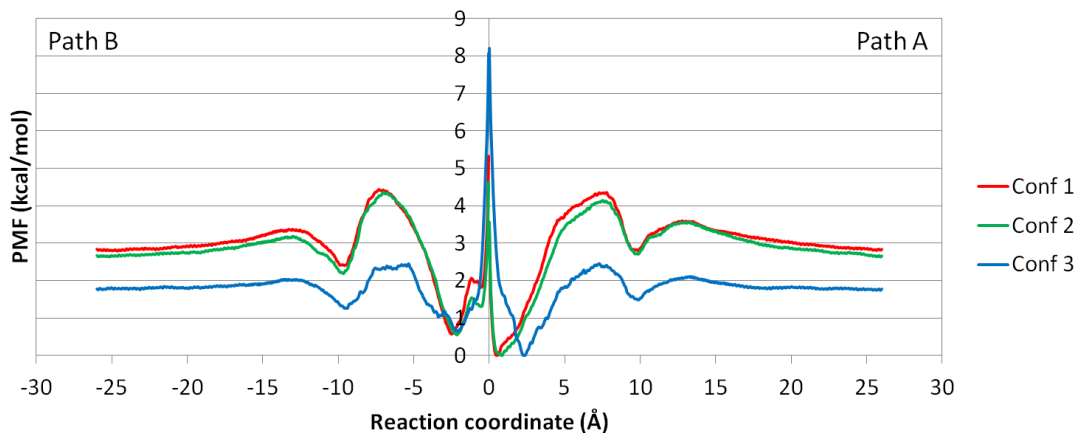
1-butanol
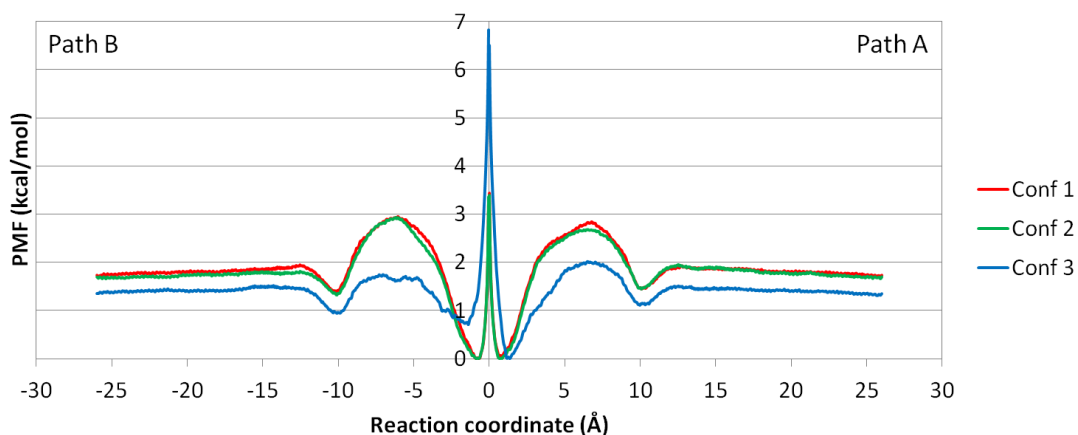
Figure 5.7. Combined PMF of dissociation of aspirin and 1-butanol from β-CD. Path A and Path B are combined for Conf 1, 2 and 3 of β-CD-aspirin and β-CD-1-butanol complexes.

The PMFs (Figure 5.7) also suggest that the association energy barriers of aspirin and 1-butanol are 1.5 and 1.1 kcal/mol in Conf 1 respectively. The similar association energy barriers of aspirin and 1-butanol agree with the fact that these two ligands have similar association rate constants (Table 5.1). Although the two-fold faster association rate

205

constant of aspirin implies a smaller association energy barrier and the computed value is actually bigger than 1-butanol, the difference in the barrier is small than 0.6 kcal/mol and can be considered as bias from the β-CD conformation and errors due to thermal fluctuation. The computed dissociation energy barriers of aspirin and 1-butanol in Conf 1 are 4.4 and 2.8 kcal/mol respectively. This agrees with the experimental dissociation rate constants perfectly.

The combined PMF unambiguously indicates that the ligands bind preferentially to the primary direction of the β-CD cavity. The ligand binds to β-CD with a range from -3 to +3 Å in depth, and there is a huge energy barrier in all combined PMFs within this range. By close investigation of the population plot (Figure 5.8), we noticed that this energy barrier near origin of the RC was caused by abnormal behavior of COM distance restraint when the two COMs were close to each other. Due to the limitation of distance restraint algorithm used in AMBER, when COM of ligand was close to 0 Å on X axis, where COM of β-CD was located, ligand could jump back and forth between path A and B. The harmonic potential added to restrain the distance between β-CD and ligand was irrelevant from directions and therefore causing an issue shown in Figure 5.8. When the distance between COMs of β-CD and ligand was larger than 1 Å, the problem disappeared because the gap had become too large for ligand to jump through. We ran multiple runs for RC smaller than 1 Å to get a rough closer PMF. The future version of restraint setting in AMBER may consider restraining vector instead distance, so this issue can be revisited and solved. It's also noted that for the 0 Å on X axis, where the COMs of β-CD and ligand were overlapped, the peak split into two located on both sides. It's an artifact of

restraint setting in AMBER, which caused the unusual high energy barrier at 0 Å position

in Figure 5.7.
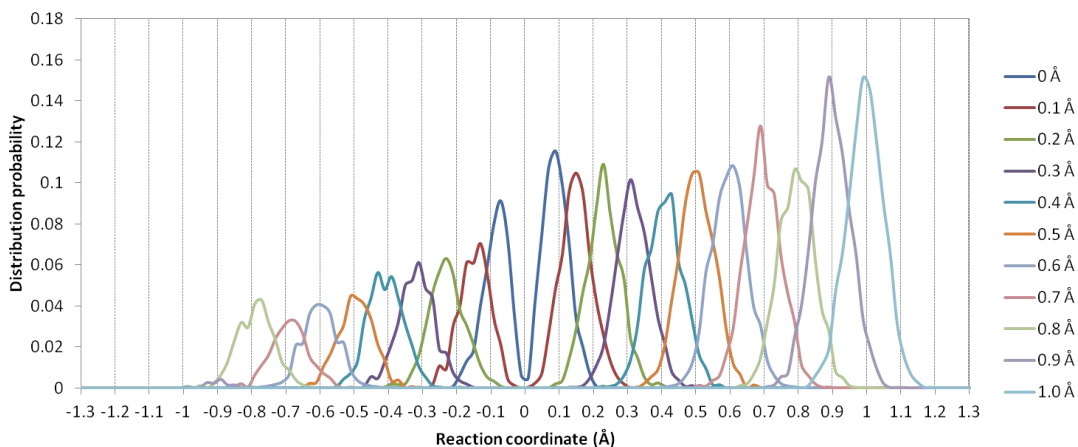
Figure 5.8. Selected distribution probability of ligand in US along path A. When distance of COMs of aspirin and β-CD is smaller than 1 Å, ligand jumps back and forth between path A and B during dissociation process of aspirin from β-CD in Conf 1 along path A, ligand stops jumping when RC distance reaches 1 Å.

Therefore, this energy barrier at the origin of the RC can be ignored. By looking at the

combined PMFs without this energy barrier, we still observe that the free energy of the

primary side is consistently 0.6 kcal/mol more favorable than the secondary side with one

exception of 1-butanol in Conf 1 and 2. This is because the primary side of the β-CD

cavity is more open and allows the aspirin to better fit into it. Because of its relatively

small size, 1-butanol can fit into both primary and secondary sides of the β-CD cavity in

Conf 1 and 2. In Conf 3, the size of cavity shrinks due to the flipping of two of the

glucopyranose units of β-CD, 1-butanol prefers to bind to the bigger primary side of β-

CD cavity (Figure 5.3).

Since we only put restraints on the 1-D RC represented by the COM distance CV, we don't have control on the position of the ligand on the sphere centered at the COM of β-CD. For example, in the window where the COM distance is 10 Å, the ligand can adopt positions anywhere on a sphere with a radius of roughly 10 Å if the interactions between the ligand and β-CD is not considered. This is not a problem in the ideal case where interactions are ignored, but in reality, the intermolecular attractions may alter the distribution of ligand on such a sphere and significantly deviates the ligand from the artificial dissociation path along the X-axis (Figure 5.9). In our simulation for aspirin, the ligand more or less follows the X-axis dissociation path way within 7 Å on the RC because of the geometrical restraints from β-CD. The ligand is free from the geometrical restraints and can diffuse on the spherical space under the government of intermolecular interactions in windows above 7 Å and within 13 Å. Note that between 10 to 13 Å, the ligand can form favorable van der Waals (vdW) interactions if the ligand is far away from the X-axis and sticks to the outer surface of β-CD (Figure 5.10). Due to this reason, the ligand deviates from the artificial path along X-axis remarkably in that region (Figure 5.11). Apparently, even at the same COM distance of 10 Å, the ligand naturally tends to stay closer to β-CD when it is on the outer surface of β-CD with stronger attractions, than in the case where it is aligned to the X-axis, where no stronger attraction can be formed. This will certainly affect the shape of the PMF. When the RC is beyond 15 Å, the two molecules do not form strong interactions any more, and the ligand is totally free to diffuse in the spherical space in the simulation for one US window. With this concern, it is interesting to investigate how PMF will be affected by the *direct dissociation* along X-

axis, and *indirect dissociation* where the ligand diffuses to the outer surface of β-CD and then dissociate.
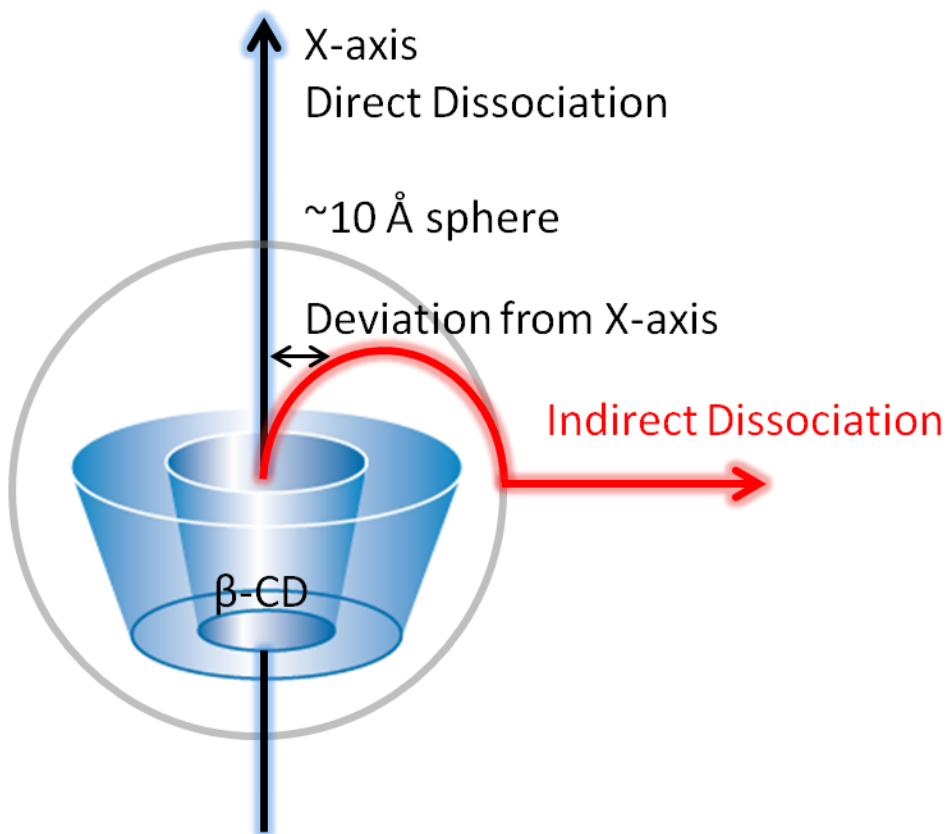
Figure 5.9. Direct and indirect ligand dissociation paths from β-CD. In direct dissociation, ligand moves out of binding site along X-axis. In indirect dissociation, ligand first diffuses to the outer surface of β-CD, which is about 10 Å from X-axis, then dissociate from there.
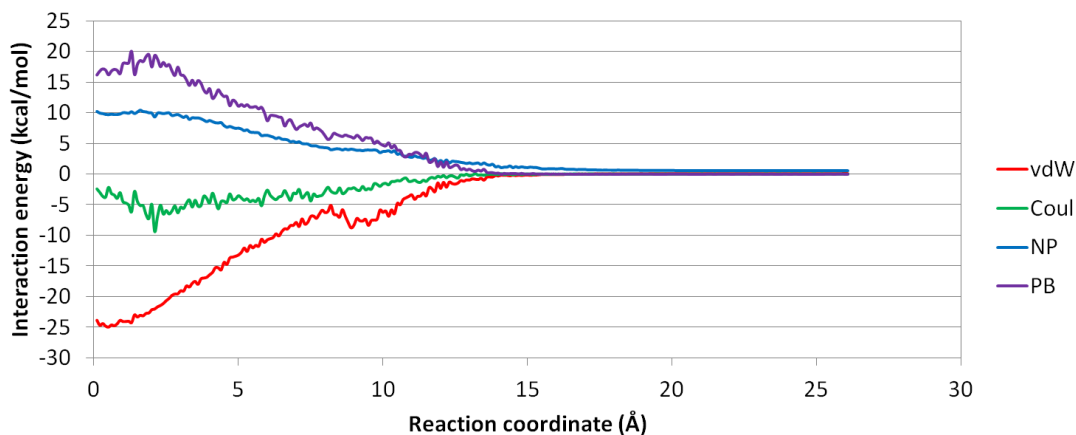
Figure 5.10. Plot of averaged MM/PBSA energy at each US window. Van der Waals energy (vdW), Coulombic energy (Coul), nonpolar solvation energy (NP) and PB solvation energy (PB) of aspirin from β-CD in Conf 1 along path A were averaged in the biased MD for each window.
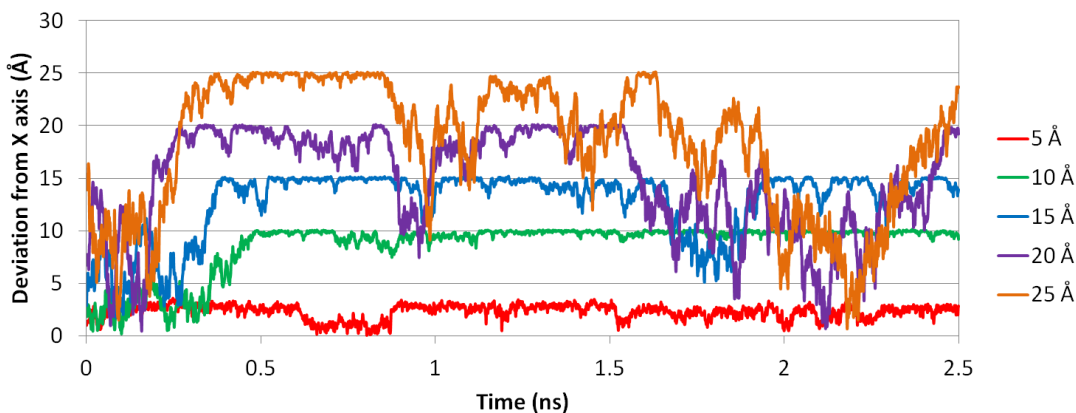
Figure 5.11. Deviation of aspirin from X axis at different RC distances. Deviations at 5, 10, 15, 20 and 25 Å during dissociation process of aspirin from β-CD in Conf 1 along path A are shown in red, green, cyan, purple and orange respectively.

According to data of aspirin deviation from X-axis (Figure 5.11), we generated a direct dissociation PMF and an indirect dissociation PMF by using the first 0.2 ns biased simulations and the last 2.0 ns biased simulations for aspirin respectively (Figure 5.12). As we anticipated, the direct PMF deviates from the indirect PMF that resembles the

210

overall PMF. The direct PMF has a higher dissociation energy barrier at roughly 7 Å, and smooths out the energy valley at 10 Å. Compared to the artificially generated direct dissociation path along X-axis, the ligand is able to find the energy minimal path by walking on the surface of β-CD and depart from there in the indirect dissociation which reproduces the reality more, resulting in a lower PMF. The strong interaction between β-CD and aspirin through the outer surface of β-CD also generates the energy valley at 10 Å, which is missing in the direct PMF. To further explore how the initial guess of the dissociation path affects the PMF from US, we randomly picked data along RC from β-CD Conf 1, 2 and 3 to construct the cross PMF. In this way, we introduced conformational exchanges between biased MD simulation starting from Conf 1, 2 and 3. Not surprisingly, the cross PMF using data from Conf 1 and 3 is located between the original PMF curves (Figure 5.13). However, the dissociation energy barrier may deviate from 2.5 kcal/mol from data of Conf 3 to 4.4 kcal/mol from data of Conf 1. Considering the normal length of biased MD simulation in each window in US application is only on the scale of nanoseconds [11, 12], it is unlikely that such short biased simulation will thoroughly explore the conformational space restrained at the CV the users use to define the RC. Therefore, an initial guess of the dissociation path that deviates from the reality too much would not be brought back to the well-equilibrated state. This stands the red flag that when using US to compute PMF for a system, the initial guess of the dissociation path plays a crucial role, and non-energy-minimal dissociation path may lead to totally wrong PMF.
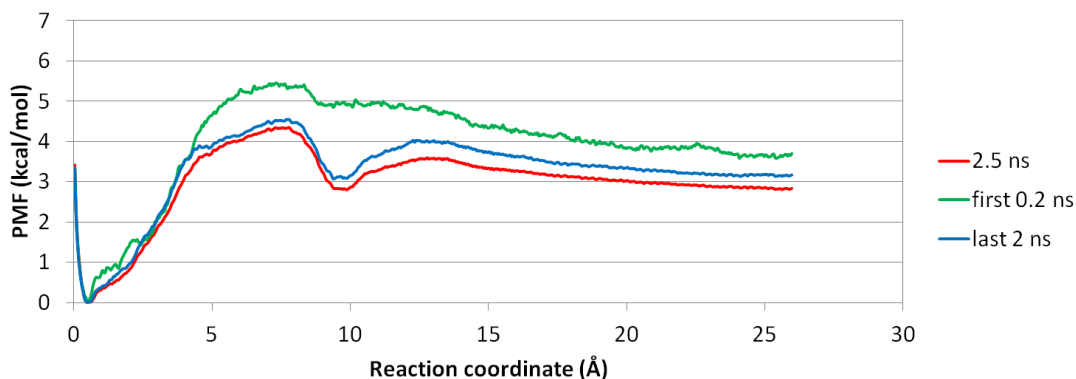
Figure 5.12. Three PMF plots of aspirin dissociation from β-CD in Conf 1 along path A. Red: using complete 2.5 ns biased MD for US. Green: using first 0.2 ns biased MD for US. Blue: using last 2.0 ns biased MD for US.
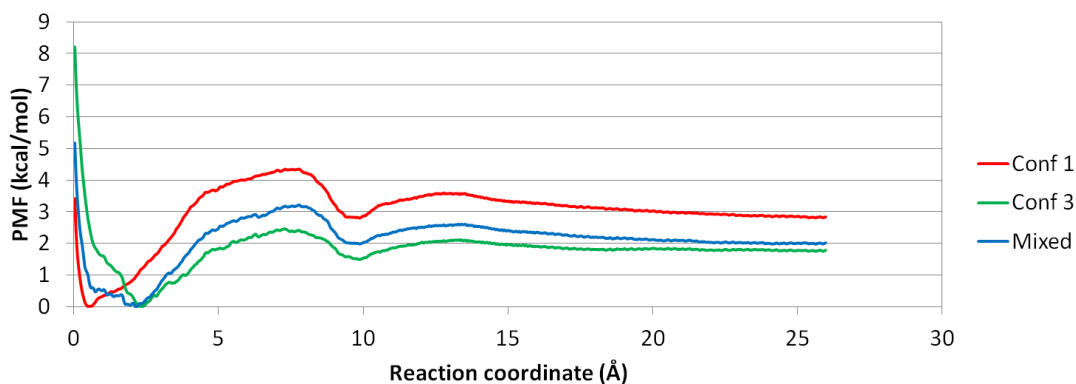
Figure 5.13. PMFs using Conf 1, Conf 3 and mixture of Conf 1 and 3 along path A. The mixed PMF in Conf 1 and 3 locates between the PMF of Conf 1 and 3.

5.3.2 Unbinding process of p38α complex system

The unbinding process of a large protein-ligand system takes usually as long as microseconds or even days, involving slow protein motions, complicated ligand rearrangement and various non-bonding interactions. This is far beyond the practical timescale of MD simulations. Therefore, it is common to use enhanced sampling methods

212

to sample the ligand dissociation pathways. Then US can be used to perform intensive sampling using such pathways as initial conformations to compute the PMF of dissociation. For p38α complex system, we prepared five paths obtained using AMD, SMD and manual pulling. Path 1 and 2 are dissociation of SB2 from DFG-in conformation using AMD equilibrated by conventional MD. One dissociation path of SB2 from DFG-out conformation was also generated in the same way. The directions of path 1 and 2 are slightly different, as indicated in Figure 5.14. Path 1 is a more straight forward dissociation where SB2 direct moves towards outside (Figure 5.15), path 2 is an indirect dissociation where SB2 adhered to the hinge region, and kept diffusing on surface of the hinge until eventually moving out. SB2 ligands in SMD and manual pulling are moving in between the two directions of path1 and 2.

Figure 5.14. Four dissociation paths for SB2-p38α complex with DFG-in conformation. Yellow: crystal bound conformation of SB2. Red: path 1 where SB2 direct moves towards outside. Green: path 2 where SB2 diffuses on the surface of the hinge region until moving out. Cyan: manual pulling path. Purple: SMD path. SB2 in manual pulling and SMD paths moves out along directions in between the two directions of path1 and 2.
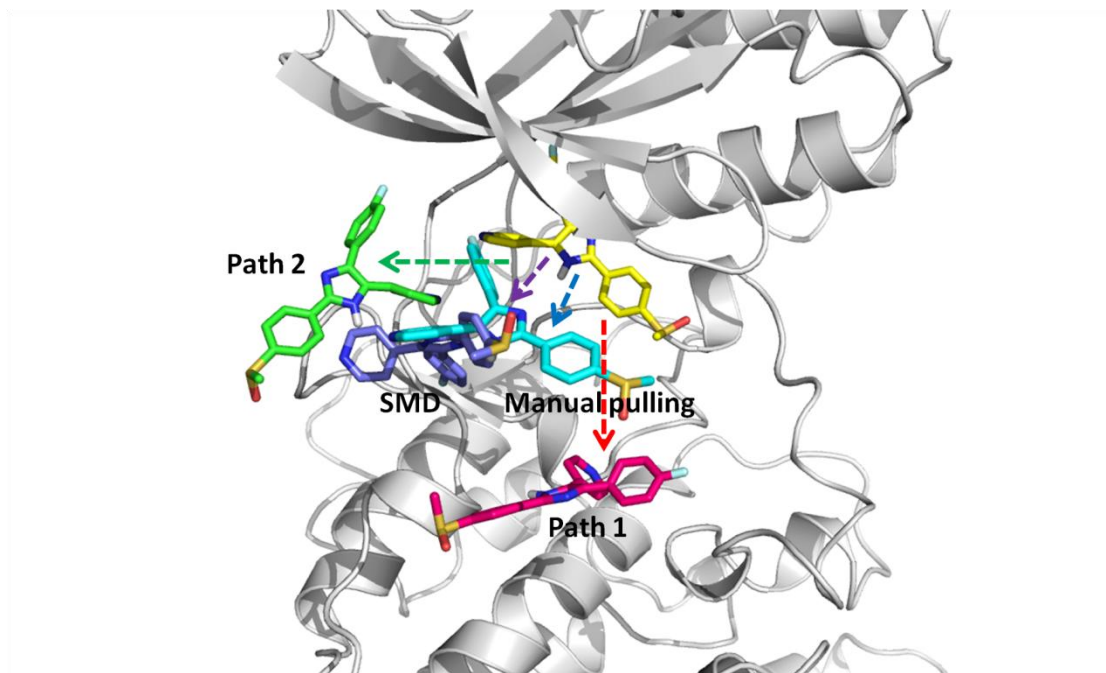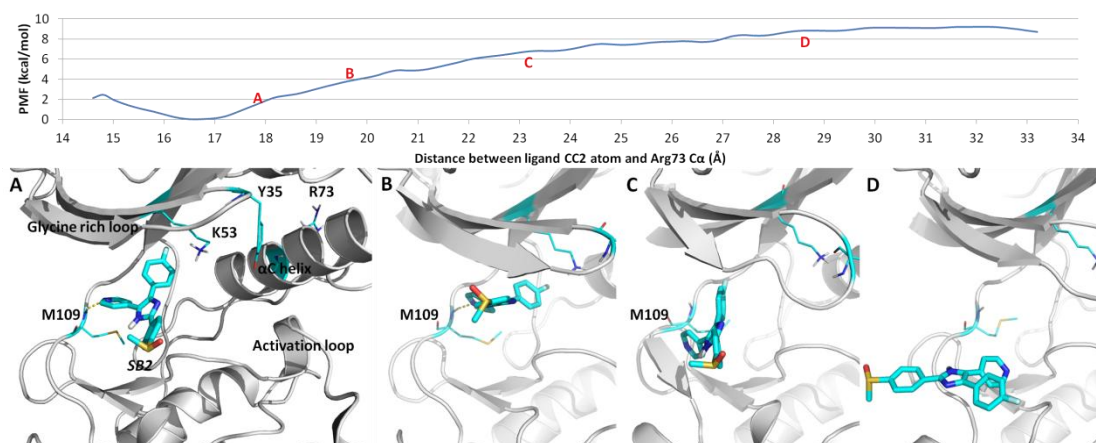
Figure 5.15. PMF of path 1 and the selected snapshots from US. Hydrogen bonds between SB2 and p38α are shown in dash line. (A) SB2 breaks hydrogen bond with Lys53 side-chain and stacking interaction with Tyr35. (B) 4-methylsulfinylphenyl group of SB2 starts diffusing towards outside the cavity (C) fluorophenyl ring of SB2 moves out of the hydrophobic pocket and hydrogen bond between pyridine nitrogen and Met109 breaks. (D) SB2 is outside the edge of binding cavity.

PMF plots of the five paths of SB2 dissociation are shown in Figure 5.16. It is very interesting that except for the PMF from SMD, the other PMF plots predict similar binding affinities which are roughly 8 to 10 kcal/mol. The binding affinities from these paths somehow fall into the common range of drug-like compound binding affinities [59] and close to experimental value (Table 5.1), regardless of the diversity of behaviors of energy barriers along the dissociation pathway and protein conformations. US using SMD path, however, predicts a binding affinity of roughly 25 kcal/mol, which is incredibly large. The PMF from US is questionable. The nanosecond timescale biased MD simulation is incapable of exploring the entire sub conformational space of protein systems restraint at the specific CV or CVs used to represent the RC. Even for small systems like β-CD-ligand complexes, nanosecond level simulations fail to smooth out the effects from initial conformations of β-CD and it takes microsecond MD simulations to

fully explore the conformational space of β-CD. Therefore, the binding affinity and energy barriers from the PMF computed using US are not undoubtedly reliable.
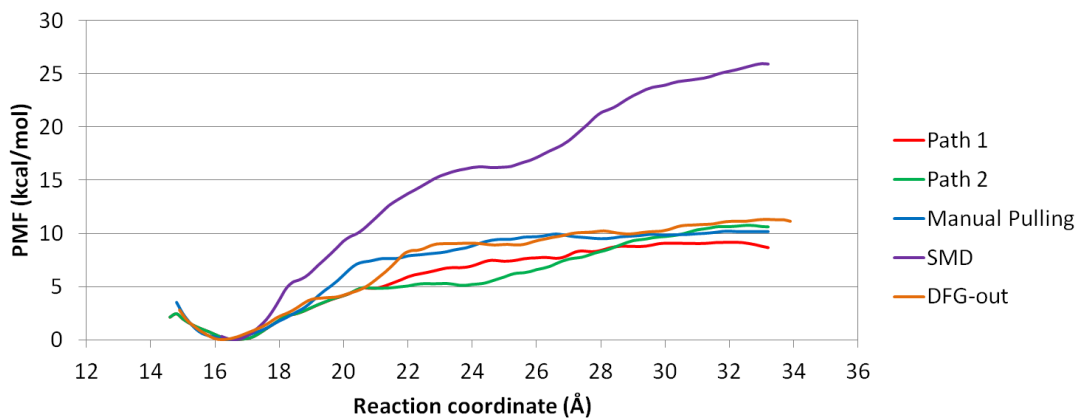
Figure 5.16. PMF plots of SB2-p38α complex. Red and green: path 1 and 2 from AMD path with conformation relaxed with short conventional MD. Cyan: manual pulling. Purple: SMD. Orange: AMD path with conformation relaxed with short conventional MD for dissociation of SB2 from DFG-out conformation.

US using paths from AMD predicts similar PMF. The binding affinities computed using path 1 and path 2 are 8.7 kcal/mol and 10.6 kcal/mol respectively, and are highly similar. In path 2, SB2 diffuses on surface of the hinge region during dissociation, which explains why its PMF reaches a low and flat region between 21 and 24 Å. In DFG-out conformation, after SB2 breaks its hydrogen bond with Met109, the 4-methylsulfinylphenyl group can rotate back inside cavity and its phenyl ring will form stacking interaction with Phe169 (Figure 5.17). However, the binding free energy of SB2 with DFG-out conformation from US is -11.2 kcal/mol, surprisingly similar to path 1 and 2 with DFG-in conformation, which agrees with previous NMR study [44] about the free conversion between DFG-in and DFG-out conformations of p38α while bound with SB2.

216

This suggests that the PMF computed using conformations from dissociation paths yielded from AMD can consistently reproduce the binding affinity if the conformations along the dissociation paths are relaxed by conventional MD simulations, even if the MD simulations are short.
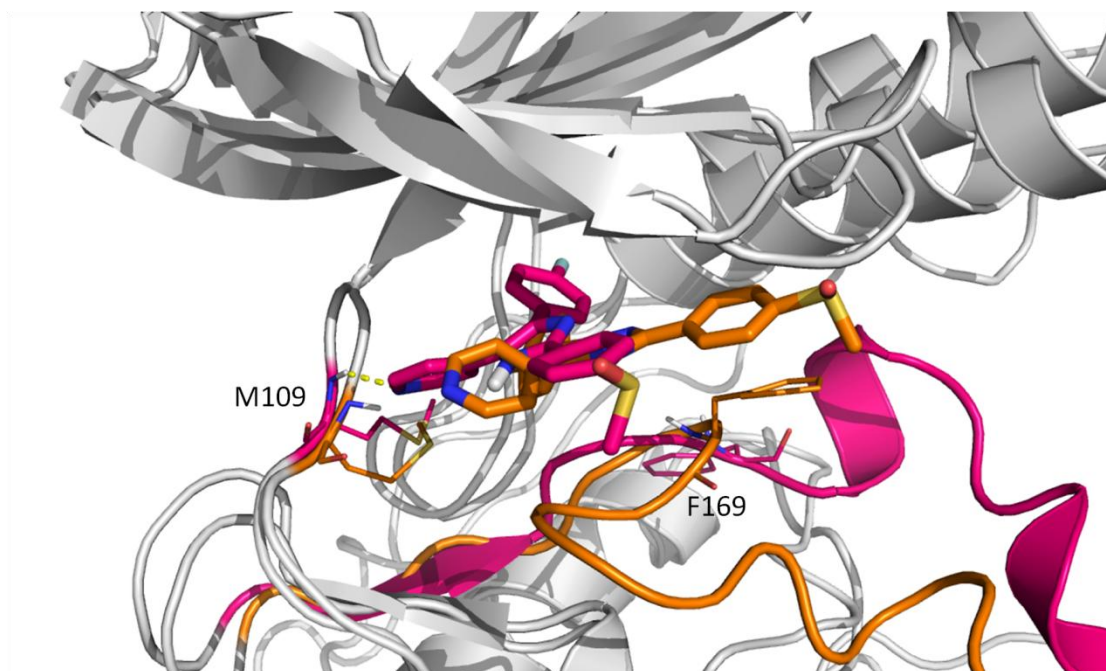
Figure 5.17. DFG-out path (orange color) VS path 1 (red color). During dissociation of SB2 from DFG-out conformation, 4-methylsulfinylphenyl group rotates back inside cavity and form stacking interaction with sidechain of Phe169. Hydrogen bonds between SB2 and Met109 are shown in dash line.

The different behaviors of protein motions and residue sidechain movement sampled from AMD, SMD and manual pulling may change the shape and height of PMF significantly. In free p38α crystal structure, Tyr35 has different orientation than SB2-p38α complex, instead of Tyr35 forming stacking interaction with the 4-methylsulfinylphenyl group of SB2, Tyr35 in free p38α rotates away and forms hydrogen

bond with the sidechain of Arg67. Both free p38α crystal structure and our AMD simulation suggest that in order for SB2 to dissociate, Tyr35 will first rotates away into its position in free p38α crystal structure, breaking its stacking interaction with SB2, thus facilitating dissociation of SB2. However, in SMD, due to the enforced pulling force acted on SB2, Tyr35 doesn't rotate away, but follows SB2 towards outside, after breaking stacking interaction with 4-methylsulfinylphenyl group at 24 Å, it immediately forms stacking interaction with fluorobenzene group of SB2, and doesn't break until 29 Å (Figure 5.18). This cause much larger energy barrier comparing to other paths. Interestingly, manual pulling US doesn't have this problem and actually performs quite similar to path 1 and 2, since all the residues of p38α are kept in the crystal structure positions, although Tyr35 doesn't rotates away from SB2 at the beginning of dissociation, it doesn't follow SB2 either, therefore avoiding unexpected interaction. Therefore, the initial guess of dissociation pathway for PMF calculation using US method must be validated and relaxed before performing biased sampling in each window for US.
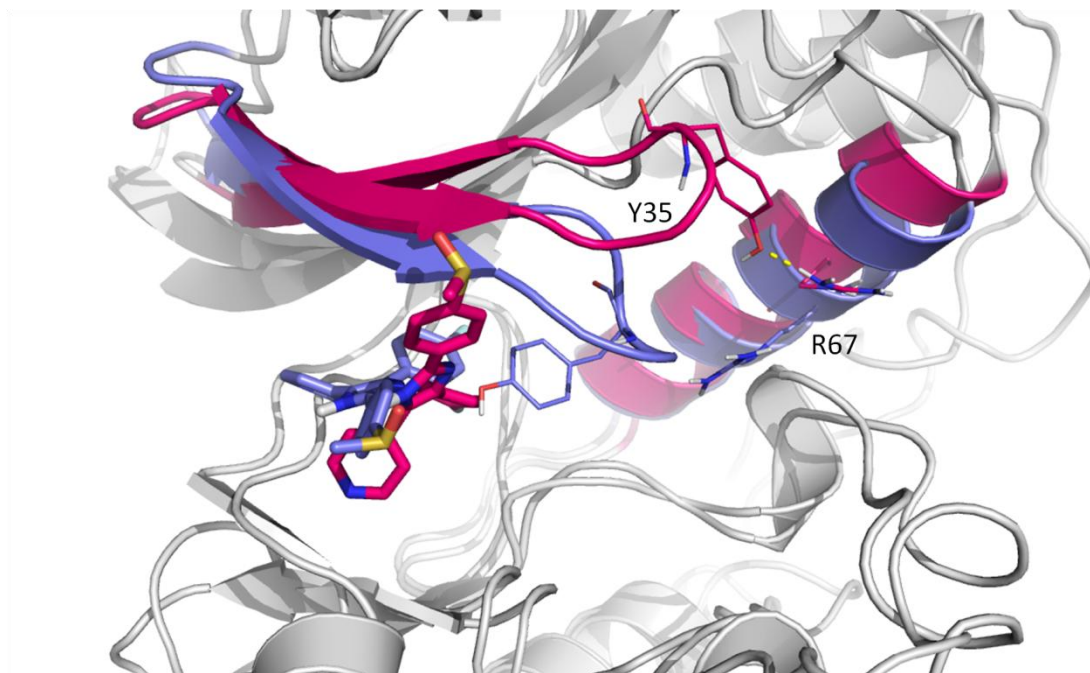
Figure 5.18. SMD path (purple color) VS path 1 (red color). Tyr35 rotates away in path 1 to form hydrogen bond (shown in dash line) with Arg67, and breaking stacking interaction with SB2, while Tyr35 in SMD follows SB2.

To confirm the lack of protein conformational change during dissociation process simulated by SMD and manual pulling simulations, we measured RMSD of backbone atoms of p38α in biased short simulations of US windows for path1, 2, manual pulling and SMD. Average RMSD were obtained in each US window with crystal structure of DFG-in p38α (1A9U) as reference (Figure 5.19). As expected, SMD and manual pulling simulations have smaller RMSD due to lack of necessary conformational change. SMD starts pulling after 100ns of conventional MD, which is enough to make the whole system reach equilibrium, but still not enough to get ready for dissociation. However, in researches using SMD to sample unbinding process of other kinase-ligand systems [60, 61], the lengths of conventional MD before SMD are mostly shorter than 100 ns, or even

10 ns, which might introduce artifacts into the simulation. Besides, previous research has shown that too large pulling speed or pulling force constant in SMD would result in instability of system or significantly higher energy barrier [62]. Therefore, future dissociation sampling with SMD should be more rigorous with system setting up and conformation selection to minimize artificial errors. Gladly, several methods have been brought up to improve free energy estimates and conformation sampling of non-equilibrium SMD simulation. Based on the forward–reverse method, Nategholeslam et al. introduced bin-passing method to better separate of the reversible and irreversible work distributions, and achieve faster convergence [63]. By combining configurational freezing and nonuniform particle-selection scheme, Riccardo Chelli developed local sampling in steered Monte Carlo simulations that can enhance the accuracy of the free energy calculation [64]. Whalen et al. applied hybrid steered MD-docking method to better rank inhibitor affinities against a flexible drug target [65].
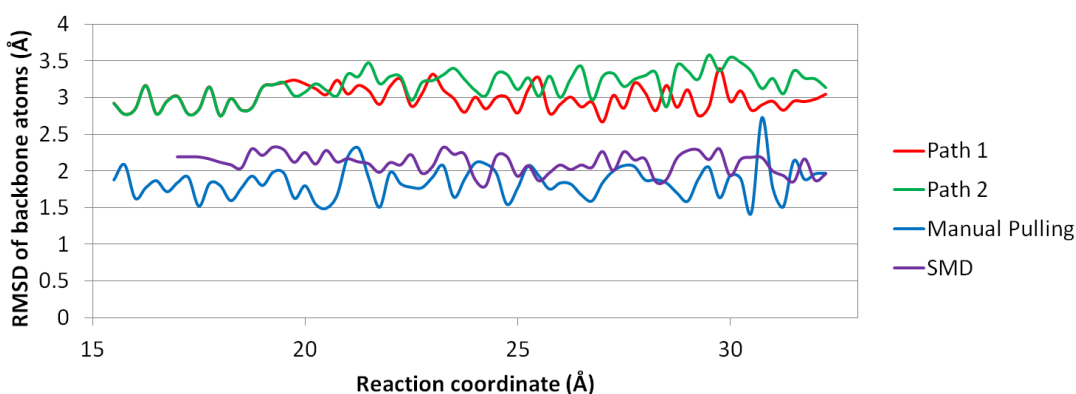

Figure 5.19. Plot of averaged RMSD of backbone atoms of p38α at each US window. RMSD values of backbone atoms of p38α from path1, 2, manual pulling and SMD paths are averaged in the biased MD for each window. Path 1 and 2 show higher RMSD due to AMD sampling more protein conformational change than SMD and manual pulling.

5.4 Conclusions

We applied US to investigate the dissociation processes of β-CD and p38α complex systems with several conformational sampling methods, including AMD, SMD and manual pulling. We also investigated the influences of the computed PMF from conformations, dissociation pathways, intensity of the biased sampling in US and dissociation pathway sampling method. Different β-CD conformations can change the depth of PMF by more than 50%, and this suggest that nanosecond timescale biased simulation is unable to remove the effects from initial conformations, even for small host like β-CD. By comparing the direct and indirect pathway of β-CD complex dissociation, we also found that different dissociation paths can result in appearance and disappearance of local minima, and non-energy-minimal dissociation path may lead to wrong PMF. Commonly used enhanced sampling methods that provide initial conformations for US were discussed. For large protein-ligand system, SMD can efficiently pull ligand out of binding cavity, however, the artificial force may introduce unexpected interactions or miss necessary conformational change in sampling of dissociation process, thus requiring careful setting up and screening of structures. Manual pulling is similar to SMD, but may be suitable for sampling dissociation path of system that not requiring much host conformational change. Compared to these two methods, universal acceleration sampling method like AMD is able to simulate adequate details along the dissociation pathway. By relaxing the conformations along dissociation pathways sampled by AMD using short MD simulations, we obtained more reliable initial conformations for US, leading to consistent and reliable PMF. Therefore, we suggest that US can be a very reliable method

for computing PMF, if a proper enhanced sampling method is used and the initial

conformations are properly relaxed to remove the bias from the enhanced technique.

5.5 References

1.   Beveridge DL, DiCapua FM. Free energy via molecular simulation: applications to chemical and biomolecular systems. Annual review of biophysics and biophysical chemistry. 1989;18:431-92. Epub 1989/01/01. doi: 10.1146/annurev.bb.18.060189.002243. PubMed PMID: 2660832.

2.   Foloppe N, Hubbard R. Towards Predictive Ligand Design With Free-Energy Based Computational Methods? Current medicinal chemistry. 2006;13(29):3583-608. doi: http://dx.doi.org/10.2174/092986706779026165.

3.   Kollman P. Free energy calculations: Applications to chemical and biochemical phenomena. Chemical Reviews. 1993;93(7):2395-417. doi: 10.1021/cr00023a004.

4.   Zwanzig RW. HIGH-TEMPERATURE EQUATION OF STATE BY A PERTURBATION METHOD .1. NONPOLAR GASES. Journal of Chemical Physics. 1954;22(8):1420-6. PubMed PMID: WOS:A1954UC16200028.

5.   Khavrutskii IV, Wallqvist A. Improved Binding Free Energy Predictions from Single-Reference Thermodynamic Integration Augmented with Hamiltonian Replica Exchange. Journal of chemical theory and computation. 2011;7(9):3001-11. doi: 10.1021/ct2003786.

6.   Torrie GM, Valleau JP. NON-PHYSICAL SAMPLING DISTRIBUTIONS IN MONTE-CARLO FREE-ENERGY ESTIMATION - UMBRELLA SAMPLING. Journal of Computational Physics. 1977;23(2):187-99. doi: 10.1016/0021-9991(77)90121-8. PubMed PMID: WOS:A1977CX19800007.

7.   Hainam D, Hirst JD, Wheatley RJ. Calculation of Partition Functions and Free Energies of a Binary Mixture Using the Energy Partitioning Method: Application to Carbon Dioxide and Methane. Journal of Physical Chemistry B. 2012;116(15):4535-42. doi: 10.1021/jp212168f. PubMed PMID: WOS:000302924800016.

8.   D L Beveridge a, DiCapua FM. Free Energy Via Molecular Simulation: Applications to Chemical and Biomolecular Systems. Annual review of biophysics and biophysical chemistry. 1989;18(1):431-92. doi: 10.1146/annurev.bb.18.060189.002243. PubMed PMID: 2660832.

9.   Chipot C. Frontiers in free-energy calculations of biological systems. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2014;4(1):71-89. doi: 10.1002/wcms.1157.

10.   Vashisth H, Skiniotis G, Brooks CL. Collective Variable Approaches for Single Molecule Flexible Fitting and Enhanced Sampling. Chemical Reviews. 2014;114(6):3353-65. doi: 10.1021/cr4005988.

11.   Velez-Vega C, Gilson MK. Force and Stress along Simulated Dissociation Pathways of Cucurbituril-Guest Systems. Journal of chemical theory and computation. 2012;8(3):966-76. Epub 2012/07/04. doi: 10.1021/ct2006902. PubMed PMID: 22754402; PubMed Central PMCID: PMCPmc3383817.

12.   Sun H, Tian S, Zhou S, Li Y, Li D, Xu L, et al. Revealing the favorable dissociation pathway of type II kinase inhibitors via enhanced sampling simulations and two-end-state calculations. Scientific reports. 2015;5:8457. Epub 2015/02/14. doi:

10.1038/srep08457. PubMed PMID: 25678308; PubMed Central PMCID: PMCPmc4326958.

13. Li W, Shen J, Liu G, Tang Y, Hoshino T. Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations. Proteins. 2011;79(1):271-81. Epub 2010/11/09. doi: 10.1002/prot.22880. PubMed PMID: 21058395.

14. Marzinek JK, Bond PJ, Lian G, Zhao Y, Han L, Noro MG, et al. Free Energy Predictions of Ligand Binding to an α-Helix Using Steered Molecular Dynamics and Umbrella Sampling Simulations. Journal of Chemical Information and Modeling. 2014;54(7):2093-104. doi: 10.1021/ci500164q.

15. Nguyen H, Le L. Steered molecular dynamics approach for promising drugs for influenza A virus targeting M2 channel proteins. European biophysics journal : EBJ. 2015;44(6):447-55. Epub 2015/06/03. doi: 10.1007/s00249-015-1047-4. PubMed PMID: 26033540.

16. Rodriguez RA, Yu L, Chen LY. Computing Protein-Protein Association Affinity with Hybrid Steered Molecular Dynamics. Journal of chemical theory and computation. 2015;11(9):4427-38. Epub 2015/09/15. doi: 10.1021/acs.jctc.5b00340. PubMed PMID: 26366131; PubMed Central PMCID: PMCPmc4565455.

17. Sun H, Li Y, Tian S, Wang J, Hou T. P-loop conformation governed crizotinib resistance in G2032R-mutated ROS1 tyrosine kinase: clues from free energy landscape. PLoS computational biology. 2014;10(7):e1003729. Epub 2014/07/18. doi:

10.1371/journal.pcbi.1003729. PubMed PMID: 25033171; PubMed Central PMCID: PMCPmc4102447.

18.   Sun H, Chen P, Li D, Li Y, Hou T. Directly Binding Rather than Induced-Fit Dominated Binding Affinity Difference in (S)- and (R)-Crizotinib Bound MTH1. Journal of chemical theory and computation. 2016;12(2):851-60. Epub 2016/01/15. doi: 10.1021/acs.jctc.5b00973. PubMed PMID: 26764587.

19.   Doudou S, Burton NA, Henchman RH. Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. Journal of chemical theory and computation. 2009;5(4):909-18. Epub 2009/04/14. doi: 10.1021/ct8002354. PubMed PMID: 26609600.

20.   Gumbart JC, Roux B, Chipot C. Standard binding free energies from computer simulations: What is the best strategy? Journal of chemical theory and computation. 2013;9(1):794-802. Epub 2013/06/25. doi: 10.1021/ct3008099. PubMed PMID: 23794960; PubMed Central PMCID: PMCPmc3685508.

21.   Woo HJ, Roux B. Calculation of absolute protein-ligand binding free energy from computer simulations. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(19):6825-30. Epub 2005/05/04. doi: 10.1073/pnas.0409005102. PubMed PMID: 15867154; PubMed Central PMCID: PMCPmc1100764.

22.   Jo S, Suh D, He Z, Chipot C, Roux B. Leveraging the Information from Markov State Models To Improve the Convergence of Umbrella Sampling Simulations. The journal of physical chemistry B. 2016;120(33):8733-42. Epub 2016/07/14. doi: 10.1021/acs.jpcb.6b05125. PubMed PMID: 27409349.

23.   Breslow R, Dong SD. Biomimetic Reactions Catalyzed by Cyclodextrins and Their Derivatives. Chem Rev. 1998;98(5):1997-2012. Epub 2002/02/19. PubMed PMID: 11848956.

24.   Challa R, Ahuja A, Ali J, Khar RK. Cyclodextrins in drug delivery: an updated review. AAPS PharmSciTech. 2005;6(2):E329-57. Epub 2005/12/16. doi: 10.1208/pt060243. PubMed PMID: 16353992; PubMed Central PMCID: PMCPmc2750546.

25.   Davis ME, Brewster ME. Cyclodextrin-based pharmaceutics: past, present and future. Nat Rev Drug Discov. 2004;3(12):1023-35. Epub 2004/12/02. doi: 10.1038/nrd1576. PubMed PMID: 15573101.

26.   Del Valle EMM. Cyclodextrins and their uses: a review. Process Biochemistry. 2004;39(9):1033-46. doi: http://dx.doi.org/10.1016/S0032-9592(03)00258-9.

27.   Lai WF. Cyclodextrins in non-viral gene delivery. Biomaterials. 2014;35(1):401-11. Epub 2013/10/10. doi: 10.1016/j.biomaterials.2013.09.061. PubMed PMID: 24103652.

28.   Marchetti L, Levine M. Biomimetic Catalysis. ACS Catalysis. 2011;1(9):1090-118. doi: 10.1021/cs200171u.

29.   Singh M, Sharma R, Banerjee UC. Biotechnological applications of cyclodextrins. Biotechnology advances. 2002;20(5-6):341-59. Epub 2003/10/11. PubMed PMID: 14550020.

30.   Barros TC, Stefaniak K, Holzwarth JF, Bohne C. Complexation of Naphthylethanols with β-Cyclodextrin. The Journal of Physical Chemistry A. 1998;102(28):5639-51. doi: 10.1021/jp9803844.

31.   Fukahori T, Kondo M, Nishikawa S. Dynamic Study of Interaction between β-Cyclodextrin and Aspirin by the Ultrasonic Relaxation Method. The Journal of Physical Chemistry B. 2006;110(9):4487-91. doi: 10.1021/jp058205n.

32.   Izatt RM, Pawlak K, Bradshaw JS, Bruening RL. Thermodynamic and Kinetic Data for Macrocycle Interaction with Cations, Anions, and Neutral Molecules. Chemical Reviews. 1995;95(7):2529-86. doi: 10.1021/cr00039a010.

33.   Nilsson M, Valente AJM, Olofsson G, Söderman O, Bonini M. Thermodynamic and Kinetic Characterization of Host−Guest Association between Bolaform Surfactants and α- and β-Cyclodextrins. The Journal of Physical Chemistry B. 2008;112(36):11310-6. doi: 10.1021/jp802963x.

34.   Nishikawa S, Ugawa T, Fukahori T. Molecular Recognition Kinetics of β-Cyclodextrin for Several Alcohols by Ultrasonic Relaxation Method. The Journal of Physical Chemistry B. 2001;105(31):7594-7. doi: 10.1021/jp010535u.

35.   Rekharsky MV, Inoue Y. Complexation Thermodynamics of Cyclodextrins. Chem Rev. 1998;98(5):1875-918. Epub 2002/02/19. PubMed PMID: 11848952.

36.   Schneider H-J, Hacket F, Rüdiger V, Ikeda H. NMR Studies of Cyclodextrins and Cyclodextrin Complexes. Chemical Reviews. 1998;98(5):1755-86. doi: 10.1021/cr970019t.

37.   Yim CT, Zhu XX, Brown GR. Kinetics of Inclusion Reactions of β-Cyclodextrin with Several Dihydroxycholate Ions Studied by NMR Spectroscopy. The Journal of Physical Chemistry B. 1999;103(3):597-602. doi: 10.1021/jp9833909.

38.   Adams JL, Badger AM, Kumar S, Lee JC. p38 MAP kinase: molecular target for the inhibition of pro-inflammatory cytokines. Progress in medicinal chemistry. 2001;38:1-60. Epub 2002/01/05. PubMed PMID: 11774793.

39.   Cuenda A, Rousseau S. p38 MAP-kinases pathway regulation, function and role in human diseases. Biochimica et biophysica acta. 2007;1773(8):1358-75. Epub 2007/05/08. doi: 10.1016/j.bbamcr.2007.03.010. PubMed PMID: 17481747.

40.   Kumar S, Boehm J, Lee JC. p38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases. Nature reviews Drug discovery. 2003;2(9):717-26. Epub 2003/09/03. doi: 10.1038/nrd1177. PubMed PMID: 12951578.

41.   Peifer C, Wagner G, Laufer S. New approaches to the treatment of inflammatory disorders small molecule inhibitors of p38 MAP kinase. Current topics in medicinal chemistry. 2006;6(2):113-49. Epub 2006/02/04. PubMed PMID: 16454763.

42.   Schindler JF, Monahan JB, Smith WG. p38 pathway kinases as anti-inflammatory drug targets. Journal of dental research. 2007;86(9):800-11. Epub 2007/08/28. PubMed PMID: 17720847.

43.   Huse M, Kuriyan J. The conformational plasticity of protein kinases. Cell. 2002;109(3):275-82. Epub 2002/05/23. PubMed PMID: 12015977.

44.   Vogtherr M, Saxena K, Hoelder S, Grimme S, Betz M, Schieborr U, et al. NMR characterization of kinase p38 dynamics in free and ligand-bound forms. Angewandte

Chemie (International ed in English). 2006;45(6):993-7. Epub 2005/12/24. doi: 10.1002/anie.200502770. PubMed PMID: 16374788.

45.   Cezard C, Trivelli X, Aubry F, Djedaini-Pilard F, Dupradeau F-Y. Molecular dynamics studies of native and substituted cyclodextrins in different media: 1. Charge derivation and force field performances. Physical Chemistry Chemical Physics. 2011;13(33):15103-21. doi: 10.1039/C1CP20854C.

46.   Pedretti A, Villa L, Vistoli G. VEGA--an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. Journal of computer-aided molecular design. 2004;18(3):167-73. Epub 2004/09/17. PubMed PMID: 15368917.

47.   Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 09. Wallingford, CT, USA: Gaussian, Inc.; 2009.

48.   Wang Z, Canagarajah BJ, Boehm JC, Kassisa S, Cobb MH, Young PR, et al. Structural basis of inhibitor selectivity in MAP kinases. Structure (London, England : 1993). 1998;6(9):1117-28. Epub 1998/10/01. PubMed PMID: 9753691.

49.   Simard JR, Getlik M, Grutter C, Pawar V, Wulfert S, Rabiller M, et al. Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. Journal of the American Chemical Society. 2009;131(37):13286-96. Epub 2009/07/04. doi: 10.1021/ja902010p. PubMed PMID: 19572644.

50.   Casper D, Bukhtiyarova M, Springman EB. A Biacore biosensor method for detailed kinetic binding analysis of small molecule inhibitors of p38alpha mitogen-

activated protein kinase. Analytical biochemistry. 2004;325(1):126-36. Epub 2004/01/13. PubMed PMID: 14715293.

51.   Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys. 2004;120(24):11919-29. Epub 2004/07/23. doi: 10.1063/1.1755656. PubMed PMID: 15268227.

52.   Jensen MO, Park S, Tajkhorshid E, Schulten K. Energetics of glycerol conduction through aquaglyceroporin GlpF. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(10):6731-6. Epub 2002/05/09. doi: 10.1073/pnas.102649299. PubMed PMID: 11997475; PubMed Central PMCID: PMCPmc124471.

53.   Torrie GM, Valleau JP. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. Chemical Physics Letters. 1974;28(4):578-81. doi: http://dx.doi.org/10.1016/0009-2614(74)80109-0.

54.   Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. Journal of Computational Physics. 1977;23(2):187-99. doi: http://dx.doi.org/10.1016/0021-9991(77)90121-8.

55.   Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, et al. {Amber 14}2014.

56.   Souaille M, Roux Bt. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Computer physics

communications. 2001;135(1):40-57. doi: http://dx.doi.org/10.1016/S0010-4655(00)00215-0.

57. Still WC, Tempczyk A, Hawley RC, Hendrickson T. SEMIANALYTICAL TREATMENT OF SOLVATION FOR MOLECULAR MECHANICS AND DYNAMICS. Journal of the American Chemical Society. 1990;112(16):6127-9. doi: 10.1021/ja00172a038. PubMed PMID: WOS:A1990DR56800038.

58. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. 1983;79(2):926-35. doi: 10.1063/1.445869.

59. Keseru GM, Makara GM. The influence of lead discovery strategies on the properties of drug candidates. Nat Rev Drug Discov. 2009;8(3):203-12. doi: http://www.nature.com/nrd/journal/v8/n3/suppinfo/nrd2796_S1.html.

60. Capelli AM, Costantino G. Unbinding pathways of VEGFR2 inhibitors revealed by steered molecular dynamics. J Chem Inf Model. 2014;54(11):3124-36. Epub 2014/10/10. doi: 10.1021/ci500527j. PubMed PMID: 25299731.

61. Yang L-J, Zou J, Xie H-Z, Li L-L, Wei Y-Q, Yang S-Y. Steered Molecular Dynamics Simulations Reveal the Likelier Dissociation Pathway of Imatinib from Its Targeting Kinases c-Kit and Abl. PLOS ONE. 2009;4(12):e8470. doi: 10.1371/journal.pone.0008470.

62. Zhang Z, Santos AP, Zhou Q, Liang L, Wang Q, Wu T, et al. Steered molecular dynamics study of inhibitor binding in the internal binding site in dehaloperoxidase-

hemoglobin. Biophysical Chemistry. 2016;211:28-38. doi: http://dx.doi.org/10.1016/j.bpc.2016.01.003.

63.   Nategholeslam M, Gray CG, Tomberli B. Implementation of the forward-reverse method for calculating the potential of mean force using a dynamic restraining protocol. The journal of physical chemistry B. 2014;118(49):14203-14. Epub 2014/11/06. doi: 10.1021/jp504942t. PubMed PMID: 25372312.

64.   Chelli R. Local Sampling in Steered Monte Carlo Simulations Decreases Dissipation and Enhances Free Energy Estimates via Nonequilibrium Work Theorems. Journal of chemical theory and computation. 2012;8(11):4040-52. Epub 2012/11/13. doi: 10.1021/ct300348w. PubMed PMID: 26605571.

65.   Whalen KL, Chang KM, Spies MA. Hybrid Steered Molecular Dynamics-Docking: An Efficient Solution to the Problem of Ranking Inhibitor Affinities Against a Flexible Drug Target. Molecular informatics. 2011;30(5):459-71. Epub 2011/07/09. doi: 10.1002/minf.201100014. PubMed PMID: 21738559; PubMed Central PMCID: PMCPmc3129543.

Chapter 6 Future Work about Compatibility of Umbrella Sampling with Other Enhanced Sampling Methods

6.1 Motivation

6.1.1 Optimization of binding free energy calculation for β-cyclodextrin system

Theoretically, umbrella sampling is able to predict binding affinity of small host-guest system more accurately than large protein-ligand system. Unlike large protein-ligand system, where ligand forms complicated interactions with surrounding residues, or large conformational changes of protein are sometimes required to facilitate ligand binding/unbinding, simple host-guest system involves simpler interactions and fewer conformational changes of host during guest binding or unbinding. We have established that different conformations of β-cyclodextrin were found to alter free energy profile and even change local energy barriers. With limited number of conformational changes of small β-cyclodextrin, we can adopts several new different conformations of β-cyclodextrin to get a more accurate binding free energy of β-cyclodextrin complex system. In turn, we can prove the accuracy of umbrella sampling when it comes to free energy calculation of small host-guest systems.

6.1.2 Improvement of performance of steered molecular dynamics simulation

In principle, one can perform a large number of repeats of the steered molecular dynamics (sMD) simulations and obtain the free energy using Jarzynski's equality [1],

which demonstrates the equivalence of the free energy change and an exponential average over the work W along nonreversible paths originating from a canonic ensemble:

$$\exp(-\beta\Delta A) = <\exp(-\beta W)>$$

This can be exploited in practical simulations by moving a constraint on the reaction coordinate relatively fast from an equilibrated system to the target system [2, 3]. However, it has been argued, based on comparing potentials of mean force obtained by umbrella sampling versus Jarzinski's equality, that Jarzinski's equality is not a practical alternative to more traditional methods to perform free energy calculations in complex biomolecular systems [4].

We have established that sMD may introduce unexpected interactions between ligand and protein when simulation the dissociation process. To further explore the limitation of sMD and ways for improvement of its performance, we can use sMD with different pulling forces or speeds.

6.1.3 Evaluation of performance of metadynamics simulation

One of the fast developing enhanced sampling methods is metadynamics simulation [5], which is based on the identification of suitable collective variables (CVs), whose fluctuations are gradually enhanced through the course of the simulation, aiding in the escape from stable configurations where the system would normally be trapped. Recently, it has been employed to explore the unbinding mechanism of an inhibitor of the pharmacologically relevant target p38α kinase. However, its accuracy of conformational

sampling hasn't been studied before. Here, we plan to evaluated the performance of metadynamics simulation by using its sampled conformations for umbrella sampling.

6.1.4 Water analysis

Previous water analysis for the binding of a inhibitor Dasatinib to src-kinase study has shown that the major free energy barrier is mainly a result of desolvation of the binding pocket on the approach of the ligand to its binding pose [6]. Even though we have established that it's not the case for p38α kinase system, we can still use water analysis to exam the performance of sMD and metadynamics simulation.

6.2 Material and Methods

6.2.1 Molecular systems

To obtain more accurate binding free energies in β-Cyclodextrin complex systems, several new different conformations of β-Cyclodextrin from previous MD simulations at 298K in NPT ensemble using q4MD-CD force field will be selected as initial bound structures. The same two ligands (aspirin and 1-butanol) will be used. Parameters of β-Cyclodextrin complex systems and settings of umbrella sampling will be the same as described in Chapter 5.

To further evaluate sMD and test metadynamics simulation, the same p38α complex system will be used. SB2 was selected as ligand for p38α system, which binds to both DFG-in conformation (PDB ID: 1A9U) and DFG-out conformation (PDB ID: 3GCP).

Parameters of p38α complex system and settings of umbrella sampling will be the same as described in Chapter 5.

6.2.2 Steered molecular dynamics simulation

As described in Chapter 5, the sMD simulation will be applied to sample dissociation of SB2 from p38α. The external potential acted on the physical system, which is in the time-dependent form as $V(t) = k[x-x_0(t)]^2$, drives a change in coordinates within a certain time, where x is the distances between Cα of Arg73 and CC2 of SB2, the same as the reaction coordinate used in umbrella sampling, $x_0(t)$ is the distances between Cα of Arg73 and the center of the restraint along the reaction coordinate. With harmonic constant k = 10 kcal mol^{-1} Å^2 throughout the run, SB2 was pulled from the bound position to the destination where the distance between SB2 and Arg73 was 33 Å. Thus the steering path was constructed. We will apply different speeds for position change of external potential, denoted as $x_0(t)/t$. The speed of previous sMD was 1.75 Å/ns, which may be too fast to allow protein conformation to reach equilibrium at different stages of ligand dissociation. We will apply another nine different speeds noted as 1.75/n Å/ns where n equals integer 2 to 10, hopefully one can capture the accurate interactions between SB2 and p38α along dissociation.

6.2.3 Metadynamics simulation

Metadynamics simulation will be performed with the NAMD. We will use the same charge and force field parameters for p38α complex system as described in Chapter 5.

Periodic boundary conditions will be used. The nonbonding interactions will be calculated with a PME scheme with a 10 Å cutoff. The integration time step will be set to 1 fs. The Langevin thermostat with a damping constant of 2 ps^{-1} will be used to maintain a temperature of 300 K, and the hybrid Nosé−Hoover Langevin piston method will be used to control the pressure at 1 atm. We will use the SHAKE procedure to constrain hydrogen atoms during MD simulations. Finally, two sets of 100 ns of metadynamics simulations will be performed.

Because metadynamics is highly dependent selection on CVs, the choice of the CVs is very important for a correct simulation. In our case, the choice of two different sets of CVs will be tested. The first set of CVs will be the same as reaction coordinate (RC) we used in our umbrella sampling, the distance between Cα of Arg73 and CC2 of SB2. The result of metadynamics simulation using this set of CVs will be compared to umbrella sampling results to test its accuracy. The second set of CVs will be a more random selection, we first selected residues with 7 Å of ligand SB2 in its crystal structure bound conformation as pocket residues, then the distance between the center of masses of pocket residues and ligand SB2 will be used as out CVs. This way the ligand will make a choice about which direction to dissociate all by itself.

After we obtain two dissociation paths from metadynamics simulation, frames with desired RC distances and small RMSD values of ligand will be selected as initial structures for US simulations.

6.3 Analysis

So far we have got the preliminary results of water analysis for the four dissociation paths of SB2 from p38α mentioned in Chapter 5. For Path 1 and 2, because their initial conformations were carefully prepared by aMD smoothed with cMD, so the conformations at each stage of dissociation were equilibriumed thoroughly. As seen in Figure 1B, water molecules entered the binding cavity as soon as SB2 was about to leave around RC distance of 19.5 Å. Manual translation can roughly simulate the resolvation process, but not quite accurately. While sMD was not able to capture the beheviour of water accurately.
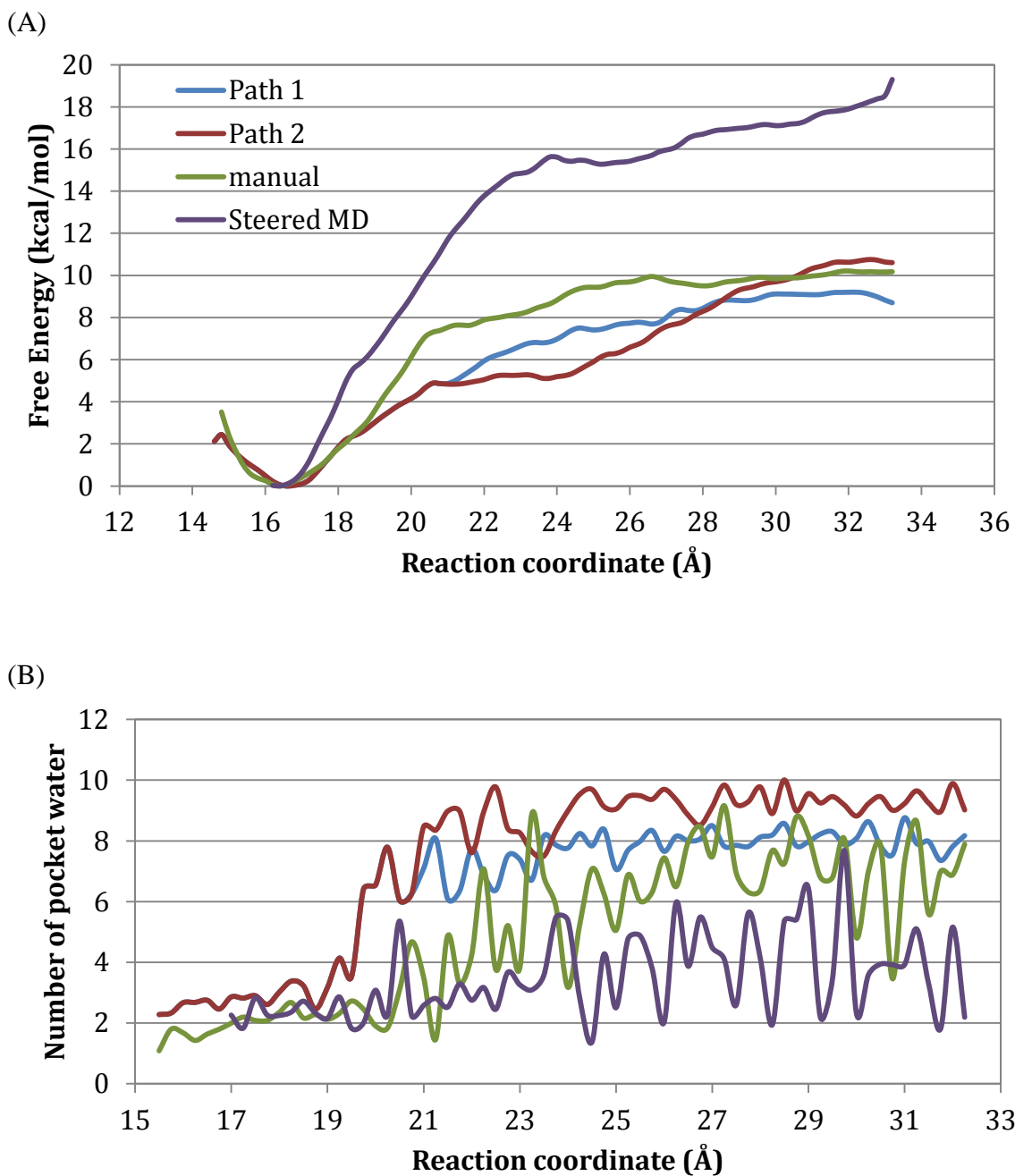
(A)

(B)

Figure 6.1 PMF and water analysis of SB2-p38α complex, including two paths (Path 1 and 2) from aMD smoothed by cMD, manual translation, sMD. (A) Calculated binding free energy from Path1, 2 and manual translation with US are close to experimental value while sMD clearly overstated the dissociation energy barrier. (B) SMD was incapable of simulating the resolvation of binding cavity as ligand leaving.

6.4 References

1.   Jarzynski C. Nonequilibrium Equality for Free Energy Differences. Physical Review Letters. 1997;78(14):2690-3.

2.   Hummer G. Fast-growth thermodynamic integration: Error and efficiency analysis. The Journal of Chemical Physics. 2001;114(17):7330-7. doi: 10.1063/1.1363668.

3.   Zuckerman DM, Woolf TB. Theory of a Systematic Computational Error in Free Energy Differences. Physical Review Letters. 2002;89(18):180602.

4.   Baştuğ T, Kuyucak S. Application of Jarzynski's equality in simple versus complex systems. Chemical Physics Letters. 2007;436(4):383-7. doi: http://dx.doi.org/10.1016/j.cplett.2007.01.078.

5.   Laio A, Parrinello M. Escaping free-energy minima. Proceedings of the National Academy of Sciences. 2002;99(20):12562-6. doi: 10.1073/pnas.202427399.

6.   Mondal J, Friesner R, Berne BJ. Role of Desolvation in Thermodynamics and Kinetics of Ligand Binding to a Protein. Biophysical Journal. 108(2):12a. doi: 10.1016/j.bpj.2014.11.093.