Computational Study of Bovine β-Lactoglobulin Complexes with Fatty Acids

by

Nobar Jalili

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Chemistry

University of Alberta

Abstract

This thesis presents studies aimed at delivering a deeper understanding of protein-fatty acids recognition and dissociation processes using molecular dynamics simulations. The focus of this thesis is on theoretical modeling of β-lactoglobulin protein in complex with fatty acid ligands (fluorinated and non-fluorinated). The dynamics of ligand exit from protein binding site is unclear and it is desired to understand whether ligands dissociate from the protein binding site along a well defined dissociation pathway or through a collection of exit pathways. This computational study of β-lactoglobulin and fatty acid complexes was inspired by recent mass spectrometry experiments using blackbody infrared radiative dissociation technique where the dissociation kinetics of these complexes was measured. Potential of mean force calculations and transition state theory were utilized to compute the dissociation rate constant of β-lactoglobulin-fatty acids complexes. Analysis of the calculated free energy profiles provided a more complete picture of the probable intermolecular interactions. The carboxyl group of the fatty acids interacts with variety of the residues on the flexible loops via hydrogen bonds but it is not involved in the interactions with the charged amino acids. There is a late transition state for the dissociation of β-lactoglobulin-fatty acid complexes and most probably the cleavage of the nonpolar interactions of the fatty acid aliphatic chain with protein residues lined in binding cavity is the last step of the activation process.

It is not clear how fluorination influences the stability of protein-ligand complexes. Recently, quantitative investigation of the energetics of β-lactoglobulin complex with fluorinated fatty acids proved that fluorocarbon binding within the binding cavity of β-lactoglobulin is stronger than hydrocarbon binding. MD simulations were performed on

β-lactoglobulin-fluorinated fatty acids complexes to probe the nature of stabilizing intermolecular interactions in further details. Analysis of the trajectory files revealed fluorine bonding to the polar hydrogen atoms is primarily responsible for the stabilizing effects of fluorination.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

$(Lg+FA)_f$	Fast component of β-lactoglobulin complex with fatty acid
$(Lg+FA)_s$	Slow component of β-lactoglobulin complex with fatty acid
$(Lg+XF\text{-}SA)_f$	Fast component of β-lactoglobulin complex with fluorinated analogs of stearic acid
$(Lg+XF\text{-}SA)_s$	Slow component of β-lactoglobulin complex with fluorinated analogs of stearic acid
β-AR	β-adrenergic receptors
Ab	Abundance
Amber	Assisted Model Building and Energy Refinement
Ala	Alanine
Asn	Asparagine
Asp	Aspartic acid
BIRD	Blackbody infrared radiative dissociation
BK	Brünger-Brooks-Karplus
COM	Center of mass
Cys	Cysteine

DCM	Dichloromethane
DMF	Dimethylformamide
ESI	Electrospray Ionization
ESP	Electrostatic Potential
FA	Fatty acid
FK506	Tacrolimus
FKBP	FK506 Binding Protein
FTICR	Fourier Transform Ion Cyclotron Resonance
GAFF	Generalized Amber Force Field
Glu	Glutamic acid
Gln	Glutamine
GW0072	Thiazolidinone
HCA II	Human carbonic anhydrase II
iGluRs	Ionotropic glutamate receptors
Ile	Isoleucine
LA	Lauric acid
Leu	Leucine

Lg	β-lactoglobulin
Lg+FA	β-lactoglobulin complex with fatty acid
Lg+LA	β-lactoglobulin complex with lauric acid
Lg+MA	β-lactoglobulin complex with myristic acid
Lg+PA	β-lactoglobulin complex with palmitic acid
Lg+SA	β-lactoglobulin complex with stearic acid
Lg+XF-SA	β-lactoglobulin complex with fluorinated analogs of stearic acid
Lys	Lysine
MA	Myristic acid
MC	Monte Carlo
MD	Molecular Dynamics
MeOH	Methanol
Met	Methionine
MM/PB-SA	Molecular Mechanics/Poisson-Boltzmann and Surface Area
NMR	Nuclear Magnetic Resonance
PA	Palmitic acid
PES	Potential energy surface

Phe	Phenylalanine
PMF	Potential of mean force
PPARγ	Peroxisome proliferator activated receptor γ
SA	Stearic acid
Src	Proto-oncogene tyrosine-protein kinase Src
RESP	Restrained Electrostatic Potential
RMSD	Root Mean Squared Deviation
THF	Tetrahydrofuran
TST	Transition state theory
Tyr	Tyrosine
Val	Valine
VMD	Visual Molecular Dynamic
WHAM	Weighted histogram analysis method
XF-SA	Fluorinated analogs of stearic acid

Chapter 1

Introduction

1. 1. Motivation

Molecular recognition phenomena play significant roles in many biomolecular processes. Non-covalent, specific association of biomolecules is central in many biological processes ranging from signal transduction to immune response and bacterial and viral infections. [1-3] Therefore, an accurate description of the nature and strength of bio molecules interactions (such as protein-ligand interactions) leads us to a comprehensive understanding of the function of biomolecules. [4] Free energy calculations attempt to provide microscopic insight into the experimental thermodynamic measurements. They elucidate the underlying chemical and physical principles governing the association and dissociation processes in complex and large biomolecular systems and shed light on molecular recognition phenomena. [5]

Addressing protein-ligand chemistry questions such as affinity of a ligand for a given protein with the aid of free energy calculations is important in the realm of pharmaceutical sciences. [6,7] Protein-ligand binding affinity K identifies the potential lead drug that binds to the target protein with high specificity and affinity. [8] Consider a solution comprising of receptor proteins and ligands, which are able to associate as follows:

$$P + L \rightleftharpoons PL$$

The corresponding binding affinity is defined as in terms of the concentrations of each species:

$$K = \frac{[PL]}{[P][L]} \qquad (1.1)$$

where $[PL]$, $[P]$ and $[L]$ are the equilibrium concentrations of the complex, free protein and ligand, respectively. Free energy simulations are able to draw a distinction between the regions of configuration space corresponding to the protein receptor with a ligand bound and unbound. Let ρ_0 and ρ_1 be the fraction of protein receptor with no ligand bound or one ligand bound, respectively. Then, free energy calculations give the binding affinity of the protein-ligand as:

$$K = \frac{\rho_1[P]_{tot}}{[L]\rho_0[P]_{tot}} = \frac{1}{[L]} \times \frac{\rho_1}{\rho_0} \qquad (1.2)$$

$\frac{\rho_1}{\rho_0}$ is calculated during free energy simulation and it is related to the reversible work required to bring a ligand molecule from the bulk to the binding site.[9] In fact, computer simulations play a central role in guiding molecular design via prediction of the binding free energies or ranking the relative affinities for a series of structurally similar molecules.

The ability of free energy calculations to predict the binding free energies of small rigid molecules to the proteins is promising.[10-14] However, application of free energy simulations in studying the kinetics of association/dissociation processes is not well established. Recently, dissociation rate constants of gaseous β-lactoglobulin and fatty acid (stearic acid, palmitic acid, myristic acid and lauric acid) complexes were reported using black body infrared radiative dissociation technique.[15] It is well known that desolvated β-lactoglobulin and fatty acid complexes resemble the solvated complexes.[15,16] Such a study provides the opportunity to explore the practical feasibility of free energy calculation to predict the dissociation kinetics of protein-fatty acid complexes in the absence of ligand solvation energy. Evaluation of solvation energy is one of major difficulties in free energy computations.[17] Furthermore, theoretical study of this system

sheds light on the challenges of free energy calculations to reproduce the thermodynamics and specifically kinetics of flexible ligand dissociation from the protein binding site through potential of mean force calculations. Investigation of the wide range of systems illustrates the strengths and shortcomings of computation and ultimately algorithmic and methodological improvement and development can lead us to the routine and generalized use of computation as a predictive tool to calculate free energies. [18]

Potential of mean force calculations along the reaction coordinate provide a useful context for understanding how proteins overcome activation barriers, which ultimately gives a molecular picture of dissociation pathways. A well-characterized free energy landscape assists in understanding whether the ligand dissociation occurs via a well-ordered sequence of bond breakings and protein conformational rearrangements or through a broad collection of dissociation pathways. [19] The aim of our work is to examine whether molecular simulation of the model system is able to reproduce experimental dissociation rate constants. Agreement between theory and experiment allows us to obtain a molecular view of the fatty acid dissociation from β-lactoglobulin binding site, understand the nature of the interactions in protein-ligand complexes and to propose a transition state for the dissociation. Details of the model systems are presented in the following section and results of free energy simulations will be described in Chapter 2.

In general, presence of fluorine atoms influences physicochemical properties of the compounds, which results in the change in their non-covalent interactions with other molecules and more interestingly with proteins. Normally, fluorine atom impacts the electronic structure of the molecule or its hydrophobicity. [20-23] Numerous experimental

and theoretical studies carried out on fluorine containing molecules to understand the origin of fluorination effects on intermolecular interactions. The results of these studies indicate that fluorine can affect non-covalent interactions directly or via modulation of the polarity of the other groups involved in intermolecular interactions. [24-31] In despite of our knowledge of the inductive effects of fluorine atoms, the influence of fluorine on docking interactions, which alters the affinity and selectivity of protein-ligand interactions, is not understood comprehensively. [32]

Recently, binding thermodynamics of alkyl- and fluoroalkyl-substituted benzene sulfonamide ligands to human carbonic anhydrase II (HCA II) was measured using isothermal titration calorimetry. The author concluded that differences in the binding thermodynamics are arisen from differences in hydrophobic surface area and not differences in the strength of the intermolecular interactions. [33] The aim of our work is to use desolvated complexes of β-lactoglobulin with fluorinated analogs of stearic acid containing X = 13, 15, 17 or 21 fluorine atoms as a model system and investigate the intrinsic energetics of fluorine bonding in the absence of solvent. Arrhenius parameters for the loss of neutral ligand from the desolvated complexes were measured using the blackbody infrared radiative dissociation technique. Molecular dynamics simulations were utilized to probe the nature of the intermolecular interactions in the desolvated complexes of β-lactoglobulin with fluorinated ligands. [34] The model systems are described in the following section and results will be presented in Chapter 3.

1. 2. Studied systems

Non-covalent complexes of β-lactoglobulin and fluorinated and non-fluorinated fatty acids

Bovine β-lactoglobulin is an ~18 kDa protein with 162 amino acid residues and belongs to the lipocalin protein family. Its ability to bind to a variety of hydrophobic molecules including fatty acids is well characterized. It is suggested that β-lactoglobulin is involved in the hydrophobic ligands transport, passive immune transfer from mother to offspring and enzyme regulation. However, its functionality is not consistent among different species. [35,36]

Lipocalin proteins share a common structure: an eight-stranded anti-parallel β-barrel (A-H) to accommodate the hydrophobic ligands and one strand by β-sheet hydrogen bonding template (strand I). Strand I is linked to C-terminal in monomer or to the next subunit in dimer. The strands are connected via some β-hairpin loops that cover the cavity. Also, there is a short helix before the first strand that closes the other end of cavity and a α-helix beyond strand H that folds back to pack against the cavity. [37] There are numerous experimental and theoretical studies on the structure and energetics of these systems. Furthermore, there is an interest in understanding the mechanism of ligand release and uptake. [38-48] The central binding cavity of the protein is large and dry (in absence of ligand), lined with the following residues: Leu^{58}, Val^{41}, Val^{43}, Leu^{46}, Leu^{54}, Ile^{56}, Leu^{58}, Ile^{71}, Leu^{87}, Val^{92} and Leu^{103} and Phe^{105}. [35,36] It accommodates various hydrophobic ligands with low specificity but high affinity. NMR studies on β-lactoglobulin and palmitic acid indicates that palmitic acid aliphatic chain is buried in the cavity and it makes rigid connections with the residues located in the lower region of the binding cavity. The residues located in the entrance are more flexible and variable conformations are present for the carboxyl end of palmitic acid. [41,49]

The Klassen group investigated the dissociation kinetics and energetics of gas phase complexes of β-lactoglobulin and hydrophobic ligands for the first time.[15] They employed blackbody infrared radiative dissociation (BIRD) technique[50,51] to measure time-resolved thermal dissociation kinetics and energetics of a series of structurally similar fatty acid ligands (FA) and β-lactoglobulin (Lg).[15] The ligands are the fatty acids $CH_3(CH_2)_XCOOH$ with increasing chain length as following: X = 10 (lauric acid ≡ LA), X = 12 (myristic acid ≡ MA), X= 14 (palmitic acid ≡ PA) and X = 16 (stearic acid ≡ SA). The results proved that nonpolar intermolecular interactions are preserved in the gas phase. Temperature dependent dissociation rate constants were determined for the gaseous deprotonated ions $(Lg+FA)^{-7}$ and importantly the dissociation proceeds by the loss of a neutral ligand. The kinetic plots indicate that there are two kinetically distinct structures for $(Lg+FA)^{-7}$ ions referred to as slow $(Lg+FA)_s^{-7}$ and fast $(Lg+FA)_f^{-7}$ components. The dissociation activation energies were obtained using Arrhenius plots and the results indicate that the dissociation activation energies increase almost linearly with the aliphatic chain length ($E_a^{SA} > E_a^{PA} > E_a^{MA} > E_a^{LA}$) for the fast component and the contribution of each methylene group is 0.82 ± 0.04 kcal mol^{-1}. Therefore, in the fast component the aliphatic chain is buried in the binding cavity and dissociation activation energies reflect the required energy to cleave the nonpolar interactions within cavity. Nevertheless, dissociation activation energies for the slow component are higher than the corresponding values for the fast component and no simple trend is observed for the slow component. They concluded hydrogen bond interactions of carboxyl group contribute predominantly in stabilizing the slow component of ion complex.[15]

In another recent study by the Klassen group, a non-covalent complex of β-lactoglobulin and fluorinated analogs of stearic acid (Lg+XF-SA) were chosen as model systems to clarify whether fluorine bonding in protein-ligand complexes alters the intrinsic strength of the intermolecular interactions. XF-SA refers to the fluorinated analogs of stearic acid where X = 13, 15, 17, 21 represents the number of fluorine atoms. [34] They measured temperature dependent dissociation rate constants for the loss of neutral ligand from the most abundant gaseous complexes $(Lg+XF\text{-}SA)^{-7}$ using the BIRD technique. Similar to the $(Lg+FA)^{-7}$ complexes, there are two kinetically distinct structures for deprotonated $(Lg+XF\text{-}SA)^{-7}$ ions known as the fast $(Lg+XF\text{-}SA)_f^{7-}$ and slow $(Lg+XF\text{-}SA)_s^{7-}$ components. [15] Arrhenius parameters obtained from measured rate constants indicate that fluorination increases the dissociation activation energies E_a for the both slow and fast components. In contrast to the slow component, the dissociation activation energies have a linear relation with the number of fluorine atoms for the fast component. The results demonstrate each fluorine atom enhances the activation energy by ~ 0.15 kcal mol^{-1} for the fast component. Thus, the average contribution of $>CF_2$ is 1.12 ± 0.01 kcal mol^{-1}. Moreover, the energetic contribution of $–CF_3$ was estimated to be 1.85 ± 0.15 kcal mol^{-1}. There was no correlation between E_a and the number of fluorine atoms for the slow component. The studies on Lg complex with non-fluorinated and fluorinated fatty acids suggest that the nature of carboxyl group interactions with the protein is different in the fast and slow components. [15,34]

1. 3. Literature review of free energy simulations

About 35 years ago, McCammon *et al*. extended molecular dynamics simulations to the realm of biological molecules by studying bovine pancreatic trypsin inhibitor and it was a

turning point in computational structural biology. [52] Their work provided a new perspective for the exploration of biological processes using *numerical experiments*. A few years later, Jorgensen and Ravimohan computed the relative free energies of hydration of methanol and ethane in dilute solutions. [53] Later on, free energy calculations were extended to tackle the computation of absolute or relative binding free energies and hydration free energies in biological systems. [54-56] In fact, calculation of free energies of biological macromolecules is one of the most important applications of biomolecular simulations.

Protein-ligand association is a particularly attractive candidate for free energy computations because the knowledge of binding affinity of a ligand to a given protein is highly demanded in pharmaceutical sciences. [7] Quantitatively accurate prediction of binding affinities of diverse ligands to biological macromolecules results in identifying novel molecules that can bind to the target receptors and act as therapeutic drugs. [9] Furthermore, computational studies enhance the ability to screen large databases of compounds *in silico* to determine potential lead drug molecules and their binding affinities, which can improve the structure based drug design.

Many biological processes rely on the specific interactions between the molecules. In other words, molecular recognition phenomena lie at the heart of biological processes. The interest in elucidating physical and chemical principles governing molecular recognition phenomena motivated numerous experimental and theoretical studies. [57-59] Free energy simulations try to reconcile the experimental thermodynamic measurements with a microscopic insight into the specific interactions involved in chemical phenomena such as protein-ligand association/dissociation processes. The role

of particular interactions in the protein-ligand association/dissociation processes may be elucidated by carrying out the mutant studies.[19] Free energy simulations are able to determine the individual contribution of solvent, protein or even particular chemical groups to the total free energy change associated with a reaction. Moreover, they can clarify the intermolecular forces underlying the association/dissociation processes by further breaking down the free energy changes arising from different components to the individual contribution of repulsive, van der Waals dispersion and electrostatic interactions.

Free energy simulations have become a useful tool to validate and refine biomolecular force fields.[60,61] Calculation of absolute hydration free energies for amino acids side chains and a variety of small molecules are examples of benchmark free energy simulations for force field refinement and validation.[62-65] Moreover, decomposition of free energies into the individual contribution of repulsion, van der Waals dispersion and electrostatic interactions provides insight into the governing intermolecular forces during solvation processes.[63]

Improvements in computational resources and theoretical methods and algorithms for carrying out free energy computations has made free energy simulations a promising tool in drug discovery and optimization. Investigations of biomolecular systems has allowed identification of the strengths and shortcomings of free energy simulations and has opened vistas to further improvement of methodologies, sampling strategies and ultimately the reliability of free energy estimations.

In recent years, there has been promising progress in predicting binding free energies.[10-12,14,66] However, in most of the studies the binding of small, rigid ligands to

small proteins has been investigated. Extensively studied systems include the T4-lysozyme mutants with small non-polar aromatic ligands, [5,59,67-69] FK506 binding protein (FKBP) and related ligands [17,70,71] and the SH2 domain of the Src protein with phosphotyrosine peptide pYEEI and its non-peptide inhibitors. [18,72] More recently, the standard binding free energy of proline-rich peptides to Src homology domain of tyrosine kinase Abl (SH3) was reported. [73] The free energy landscapes of GluA2 (belongs to ionotropic glutamate receptors (iGluRs) family) and nine ligands including glutamate were studied. [74] All of these studies have been limited to fairly small, rigid ligands and small proteins. In general, free energy calculations for the binding of large flexible ligands to large proteins is challenging and computationally prohibitive due to the intrinsic difficulty of calculating the entropy of the system. [75]

There are three main obstacles in free energy calculations: accuracy of force fields, evaluation of solvation energy and difficulties arisen from a deficiency in sampling of the configuration space, which hinders an accurate free energy calculation. [17] In terms of solvation energy evaluation, the binding free energy may be estimated based on a continuum solvent approximation or using the molecular mechanics/Poisson-Boltzmann and surface area (MM/PB-SA) method. [76,77] The latter is a mixed scheme combining configurations sampled from MD simulations with explicit solvent, together with free energy estimators based on an implicit continuum solvent model. Generally, the free energy of the system is calculated using the *ensemble average* expression. An *ensemble* is the collection of all possible microstates that are consistent with the thermodynamic or macroscopic state of the system. In general, from statistical mechanics, a conventional simulation of the practical length cannot adequately sample

the high-energy states that contribute to the free energy profile.[75] Thus, a deficiency in sampling of the configuration space leads us to inaccurate free energy prediction.

Coupling parameter approaches circumvent the sampling problem by casting the potential energy in terms of a continuous parameter referred to as the coupling parameter. The coupling parameter λ may be considered as a general extent parameter, defining the progress of the system along a path between two well-defined states. The free energy difference between two states can be calculated by integrating the derivative of the free energy along λ (thermodynamic integration);[78] designating closely spaced intermediate states and stepwise calculation of the free energies (perturbation method);[79] or actually developing the free energy on a variable λ and calculate $A\ (\lambda)$. The latest method is called potential of mean force where free energy is defined as a function of λ. On a physical path, $A\ (\lambda)$ can be used to determine the free energy of activation, as well as free energy difference between two initially defined states and thus the rate constants can be estimated via transition state theory.

The focus of this thesis is on calculation of the potential of mean force (PMF) via geometrical transformation. In this approach the alchemical decoupling of the ligand with its surrounding (widely used in thermodynamic integration and perturbation method) is prevented. It is advantageous in the sense that PMF method mimics roughly dissociation process whereby the ligand moves along a reaction path from the binding site to the point far from the binding site. In this way, the reaction path will reflect a meaningful physical pathway and dissociation activation free energy can be calculated.[80] The PMFs along the reaction coordinate are computed using biased molecular dynamics simulations. Therefore, we offer a brief description of Born-Oppenheimer approximation in the next

section, which is exploited in classical molecular dynamics (MD) simulation of large systems.

1. 4. Born-Oppenheimer approximation

Schrödinger equation lies at the heart of quantum mechanics and it describes the behavior of the wave-like particles. This eigenvalue equation is as following:

$$\widehat{H}\,\Psi\left(\vec{r},\vec{R}\right) = E_{tot}\,\Psi\left(\vec{r},\vec{R}\right) \quad (1.3)$$

where $\widehat{H}$ is Hamiltonian and E_{tot} is the total energy of the system. $\Psi\left(\vec{r},\vec{R}\right)$ represents the wave function of the system as a function of $\vec{r}$ and $\vec{R}$, position vector of the electrons and nuclei, respectively. The Hamiltonian of a molecular system composed of electrons and nuclei associated with m_e and m_A masses is given by:

$$\widehat{H} = -\sum_A \frac{\hbar^2}{2m_A}\nabla_A^2 - \frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \frac{e^2}{4\pi\varepsilon_0}\sum_{i,A}\frac{Z_A}{|r_i - R_A|} + \frac{e^2}{4\pi\varepsilon_0}\sum_{i<j}\frac{1}{|r_j - r_i|} + \frac{e^2}{4\pi\varepsilon_0}\sum_{A<B}\frac{Z_A Z_B}{|R_B - R_A|} \quad (1.4)$$

In the above equation, A and B refer to nuclei and R represents their positions. i and j refer to electrons and r denotes the position of electrons. The first two terms originate from the kinetic energy of nuclei and electrons, respectively. The third term results from the attractive Columbic interaction of electrons and nuclei. The last two terms correspond to the repulsive interaction of like-charge particles. [81]

It is not feasible practically to solve Schrödinger equation for a complex system associated with the wave function $\Psi\left(\vec{r},\vec{R}\right)$, which includes all of the position variables. Therefore, several approximations need to be considered for a system with the large number of degrees of freedom. Nuclei are much heavier than electrons and move much more slowly. The time scale separation of electronic and nuclear motions forms the basis

of Born-Oppenheimer approximation that was proposed in 1927. [82] Therefore, it is assumed that the nuclei are fixed with respect to the electrons. The wave function is rewritten as the product of electronic and nuclear wave functions but the electronic wave function depends on the position of the nuclei parametrically:

$$\psi\left(\vec{r},\vec{R}\right) = \psi\left(\vec{r};\ \vec{R}\right)\varphi\left(\vec{R}\right) \quad (1.\ 5)$$

Then, Schrödinger equation is reformulated in two separate equations:

$$\widehat{H}_{el}\,\psi\left(\vec{r};\ \vec{R}\right) = E_{el}\ \psi\left(\vec{r};\ \vec{R}\right) \quad (1.\ 6)$$

$$\widehat{H}_{nuc}\,\varphi\left(\vec{R}\right) = E_{tot}\,\varphi\left(\vec{R}\right) \quad (1.\ 7)$$

$$\widehat{H}_{el} = \widehat{H} - \sum_{A}\frac{\hbar^2}{2m_A}\nabla_A^2 \quad (1.\ 8)$$

Equation (1. 6) is used to obtain the electronic energy E_{el} for a given nuclei positions. E_{el} includes the kinetic energy of the electrons and potential energy terms corresponding to electron-nuclei attraction and the repulsion between like-charge particles. Equation (1. 6) is solved at slightly varying nuclei positions and results in electronic energy as a function of $\vec{R}$ that is called potential energy surface (PES). The PES is exploited to construct $\widehat{H}_{nuc}$ by adding nuclear kinetic energy operator. All of the electronic effects are buried in PES implicitly and one can describe the nuclear motions on the potential energy surface by solving equation (1. 7). Ab initio quantum methods are utilized to calculate a series of electronic structures at various nuclei positions and provide PES. However, the quantum computations are so expensive and demanding for the larger systems. Thus, a model PES called force field offers an alternative approach. In addition, the nuclear motions can be treated classically instead of solving nuclear Schrödinger equation using molecular dynamics. This approximation is safe due to negligible quantum mechanical effects for the heavier nuclei and it would be more efficient computationally.

A brief introduction to related concepts of statistical mechanics is presented in the following section. It is explained how statistical mechanics principles are exploited in MD simulations to calculate physical observables.

1. 5. Statistical mechanics and a canonical ensemble

Statistical mechanics is utilized to calculate physical observables or the macroscopic properties of a system by averaging over microscopic states of a given system. The key postulate of statistical mechanics is that if one waits long enough, eventually, all of the microscopic states of the system are explored. [83,84]

A conventional experimental measurement consists of a series of independent measurements and the observed value of the desired property A is determined by:

$$A_{obs} = \frac{1}{N} \sum_{i=1}^{N} A_i \qquad (1.\ 9)$$

where N measurements are carried out and A_i is the value obtained from the i^{th} measurement assuming the system is in a single microstate. Equivalently, statistical mechanics considers all microstates consistent with a macro state such that equation (1. 9) can be rewritten as:

$$A_{obs} = \sum_{\nu} P_{\nu} A_{\nu} \equiv \langle A \rangle \qquad (1.\ 10)$$

In the above equation, P_{ν} denotes the probability of finding the system in macrostate ν, which is equal to the ratio of the number of times that state ν is observed to the total number of the measurements. A_{ν} represents the expected value corresponding to state ν and the angular brackets refer to an "ensemble" average. [83,85]

Typically, one is often interested in systems at equilibrium, with a constant number of particles N, constant temperature T and constant volume V (canonical

ensemble). Under these conditions, the probability P_i of finding the system in state i with energy E_i is given by the Boltzmann distribution:

$$P_i = \frac{e^{-E_i/k_BT}}{Q} \tag{1. 11}$$

where k_B is the Boltzmann constant and Q is the partition function given by:

$$Q = \sum_i e^{-E_i/k_BT} \tag{1. 12}$$

Therefore, the ensemble average of an observable in quantum mechanics can be calculated by:

$$\langle A \rangle = \frac{\sum_i \left\langle i \left| e^{-\hat{H}/k_BT} \hat{A} \right| i \right\rangle}{\sum_i \left\langle i \left| e^{-\hat{H}/k_BT} \right| i \right\rangle} \tag{1. 13}$$

where $\hat{H}$ is Hamiltonian of the system and $\hat{A}$ is the corresponding operator of observable A. Equation (1. 13) can be rewritten in the more familiar form:

$$\langle A \rangle = \frac{\sum_i e^{-E_i/k_BT} \langle i|\hat{A}|i\rangle}{Q} \tag{1. 14}$$

where $\langle i \left| \hat{A} \right| i \rangle$ is the expected value of observable A in quantum state i and E_i is the energy of that state.

The classical mechanical analogy of the expectation value of A is given by:

$$A_{obs} = \frac{1}{Q_{cl}} \int \dots \int d\vec{p}\, d\vec{q}\, e^{-\beta H(p,q)}\, A(\vec{p}, \vec{q}\,) \tag{1. 15}$$

$$Q_{cl} = \frac{1}{N!\, h^{3N}} \int \dots \int d\vec{p}\, d\vec{q}\, e^{-\beta H(\vec{p},\vec{q})} \tag{1. 16}$$

where Q is the classical canonical partition function, h is Plank's constant and H represents Hamiltonian of the system. $1/N!$ factor accounts for the N indistinguishable particles. Just as the wave function allows us to determine all the properties of a system in quantum mechanics, the partition function in statistical mechanics is the key to

calculating many properties of a system such as heat capacity, internal energy, free energy and so on. [86]

One can calculate the value of an observable using computer simulations by two strategies: an MD simulation via the time average of the observable or a Monte Carlo (MC) simulation via the ensemble average. In this thesis we rely on MD simulations.

1. 6. Molecular dynamics

A classical MD simulation may be used to compute equilibrium and transport properties of a many-particle system by calculating the time average. A classical simulation may be justified if the motion of the nuclei can be well approximated by classical dynamics. In the spirit of an experimental measurement, which occurs over a finite length of time, a molecular dynamics simulation of a many-particle system can be used to calculate the ensemble average of an observable. For an ergodic system, the time average of an observable and its ensemble average are equal if the time evolution of the particles' positions and momenta, $\Gamma(t)$, is calculated by a sufficiently long MD simulation.

$$A_{obs} = \langle A \rangle = \lim_{t_{obs}\to\infty} \frac{1}{t_{obs}} \int_0^{t_{obs}} A\left(\Gamma(t)\right) dt \quad (1.\ 17)$$

where $\Gamma(t) = \{p_1(t), \dots, p_{3N}(t), q_1(t), \dots, q_{3N}(t)\}$ (1. 18)

Note that for a system consisting of N particles, there are 3N positions and 3N momenta. Each point in the phase space $\Gamma(t)$ represents a state of the system at time t. This 6N dimensional phase space is explored during an MD simulation.

To start the simulation, one specifies the initial positions $\vec{p}\ (0)$ and momenta $\vec{q}(0)$ of all the particles. The initial positions of the particles are often obtained by the experimental structures provided by techniques such as X-ray crystallography or NMR spectroscopy. Maxwell-Boltzmann distribution is used to assign the velocities to each

atom at the specified temperature.[87] The interaction between the constituting particles is determined by the selected force field. Then, Newton's equations of motion are solved for all of the particles to obtain the acceleration on each atom and consequently, the new $\vec{p}\,(t)$ and $\vec{q}(t)$. Repetition of this procedure at regular time steps until the end of simulation time gives the dynamical trajectory of the system in phase space. Therefore, the time average of A may be obtained from a finite number of time steps τ_{obs}via:

$$A_{obs} = \frac{1}{\tau_{obs}} \sum_{\tau=1}^{\tau_{obs}} A\big(\Gamma(\tau)\big) \qquad (1.\ 19)$$

The length of the time step $\Delta t = {}^{t_{obs}}/_{\tau_{obs}}$ is required to be shorter than the fastest motion in the system. Similar to experiment, the property of the interest is measured over a certain time interval, which is called production. However, the system must reach thermal equilibration prior to measurement. Thermal equilibration is reached if the average temperature of the system stays steady. After equilibration, the previous history of the calculation is discarded and the actual data is collected during the production time.[88]

The gradient of the potential energy with respect to the position of each nucleus gives the force acting on the nucleus. The most time consuming step of the MD procedure is the calculation of the forces on the nuclei. An integration method is needed to solve the equations of motion. Almost all of the methods are designed based on finite difference schemes. The most widely used integrator is called the velocity-Verlet algorithm. First, the positions around time t are approximated by a Taylor series expansion (the higher order sentences in Taylor expansion are neglected):[89]

$$\vec{q}_i\,(t + \Delta t) = \vec{q}_i(t) + \vec{v}_i \Delta t + \frac{1}{2}\,\vec{a}_i \Delta t^2 \qquad (1.\ 20)$$

$$\vec{q}_i\,(t-\Delta t) = \vec{q}_i(t) - \vec{v}_i\Delta t + \frac{1}{2}\,\vec{a}_i\Delta t^2 \quad (1.\,21)$$

$$\vec{a}_i = \frac{\vec{F}_i(t)}{m_i} \quad (1.\,22)$$

Then, equation (1. 13) and (1. 14) are added to find the position of the next time step (t + Δt):

$$\vec{q}_i\,(t+\Delta t) = 2\vec{q}_i(t) - \vec{q}_i\,(t-\Delta t) + \vec{a}_i\Delta t^2 \quad (1.\,23)$$

The above equation is referred to position-Verlet algorithm since there is no explicit term involving velocity. However, the velocity can be calculated by one the equivalent equations shown in below:

$$\vec{v}_i(t+\Delta t) = \frac{\vec{q}_i(t+\Delta t) - \vec{q}_i(t-\Delta t)}{2\Delta t} \quad (1.\,24)$$

$$\vec{v}_i(t+\Delta t) = \vec{v}_i(t) + \frac{\vec{F}_i\,(t)+\vec{F}_i\,(t+\Delta t)}{2m_i}\,\Delta t \quad (1.\,25)$$

In both equations, the new positions should be calculated prior to velocity calculation.

In order to select an appropriate integration algorithm a few considerations must be taken into account. (i) Conservation of energy is one of the most significant considerations i.e the sum of the kinetic and potential energies remains constant. The root-mean squared fluctuation of energy is proportional to Δt^2 for the Verlet algorithm so short time step fulfills the energy conservation criterion. (ii) Calculation of the forces is the most demanding task in an MD simulation so a longer time step is desirable. However, there is a trade off between the cost of the computation and its accuracy.[90] (iii) A good algorithm conserves momentum and is time-reversible. [91]

Special attention is required for simulations performed at higher temperatures or involving lighter atoms. It is mentioned earlier that the time step in an MD simulation must be shorter than the fastest motion in the system. Simulation of fast particles at high

temperatures and with higher vibrational frequencies requires shorter time steps. One can apply constraints on an MD simulation in order to use longer time steps. In the works presented in this thesis, the SHAKE algorithm was utilized to constrain the bonds involving hydrogen atoms.[92] Two types of forces are present in a constrained system: the intra-molecular and intermolecular forces described by force field terms and the forces due to the constraints. To fix the bond lengths, the forces are imposed along the bonds. Since there is an equal but opposite force on each atom, there is no net force. The forces associated with the constraints are calculated by differentiating the constraints with respect to the coordinates of the atoms and multiplying by a Lagrange multiplier. Then, these forces are incorporated into the equations of motion. The new positions are obtained using the unconstrained positions in the presence of the constraint forces. The next step is determination of the multipliers in a way that all of the constraints are satisfied simultaneously. SHAKE algorithm uses an iterative approach to achieve this.

In order to perform an MD simulation, one needs to choose a system and describe the underlying chemistry and physics principles among N particles using a model, which is described in further details in the following section.

1. 6. 1. Molecular modeling and force fields

The goal of the computational study of a molecular system is predicting some physical properties based on the numerical solution of the mathematical equations that entail the physical laws governing the behavior of the system. In an MD simulation, the electronic effects are taken into account in the model and the motion of the nuclei is treated classically. The molecular model includes the coordinates of the constituting atoms and description of their interactions using a functional form and parameter sets called force

filed. The kinetic energy of the electrons and their interactions are buried in the model potential energy, which is provided using three approaches: (i) Quantum mechanics (ii) Analytical functions to describe the interactions such as Columbic forces, van der Waals interactions and etc. (iii) Empirical fitting where the experimental data are used and the analytical functional forms are fitted to the potential.

The force fields are often represented by the sum of the bonded and non-bonded interactions. The former includes the interactions due to chemical bonds, bond angles and dihedral angles. In an additive force field, which was employed in this study, non-bonded energies consist of the terms due to long range electrostatic and van der Waals forces. The mostly applied force fields are two-body additive force fields and one example is as following: [93]

$$E_{tot} = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta (r - r_\theta)^2 + \sum_{dihedrals} \frac{V}{2} [1 + cos\,(n\varphi - \gamma)] + \sum_{i<j} 4\varepsilon_{ij}\left(\frac{\sigma_{ij}}{R_{ij}^{12}} - \frac{\sigma_{ij}}{R_{ij}^{6}}\right) + \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon_0 R_{ij}} \quad (1.\ 26)$$

where the first three terms describe the stretching, bending and torsional bonded interactions, respectively. Harmonic oscillator approximation is considered to describe the associated energies with the bond stretching and bond angle bending where K_r and K_θ are the harmonic force constants for the bond stretching and bond angle bending. r and θ denote the bond length and bond angle and r_{eq} and θ_{eq} represent their values at equilibrium. This approximation does not take into account anharmonicity, bond breaking and cross terms. The third term of equation (1. 26) represents the torsional energy where V is the energy barrier to rotation, φ is the torsional angle, n represents the periodicity in one full rotation and γ is the phase angle. The forth term represents van der Waals

interactions approximated by a Lennard-Jones 6-12 potential. ε_{ij} is van der Waals well depth, σ_{ij} is the distance at which $E_{vdW}^{LJ} = 0$ and R_{ij} is the distance between atom i and j. Van der Waals parameters are chosen to fit the properties of the interacting atoms. The last term corresponds to the electrostatic interactions between point charges due to uneven distribution of charges in the system. ε_0 is vacuum permittivity, q_i and q_j are charges on atom i and j and R_{ij} is their distance. The atomic multipoles and induced dipoles are neglected in an additive force field.

Equation (1. 26) presents the simplest description of the interactions. Some force fields separate hydrogen bond interaction and replace Lennard-Jones 6-12 potential term by Lennard-Jones 10-12 function $E = \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{10}}$ for the atoms involved in hydrogen bond interactions. In addition, one can use more sophisticated force fields such as additive (polarizable) force fields.[90]

Each particular force field includes a set of parameters for each atom type such as atomic mass, van der Waals parameters, force constants, bond lengths, bond angles and dihedral angles. In addition, partial charges must be provided for the chosen force field to construct the potential energy of the system. Amber (Assisted Model Building and Energy Refinement) force fields are widely used for proteins and other biological molecules.[94]

1. 6. 2. Constant temperature molecular dynamics

Many experimental measurements are carried out at constant temperature T and constant pressure P or constant volume V. In order to generate the correct distributions of positions and momenta for the canonical (NVT) and isothermal-isobaric (NPT) ensembles, modifications must be made to the Newtonian equations of motion. [91,95]

In order to perform constant temperature simulations, "thermostats" are used by either deterministic or stochastic coupling of the system to a reservoir. When a system is coupled to a reservoir, the particles of the system and reservoir can exchange energy to achieve thermal equilibration. [95-99] In the deterministic approach, extra degrees of freedom corresponding to the reservoir are added to the equations of motion. The Hamiltonian of this extended system is conserved. [99] The Nosé [100] and Nosé-Hoover [101] thermostats are examples of deterministic approaches to perform constant temperature MD simulations.

Several stochastic coupling methods are available, for instance, the Andersen thermostat, Berendsen thermostat and Langevin method. [97,98,102] In the studies reported in this thesis, the Langevin method is utilized and thus it is explained in further details here. The Langevin equation is a non-Newtonian equation of motion with the following general form: [102,103]

$$M\dot{\vec{v}} = F\,(\vec{r}) - \gamma\vec{v} + \left(\frac{2\gamma k_B T}{M}\right)^{1/2} R\,(t) \qquad (1.27)$$

where M is the mass, $\dot{\vec{v}}$ is acceleration and $\vec{v}$ is velocity, $F\,(\vec{r})$ is the force as a function of position $\vec{r}$, γ is the friction coefficient, k_B is the Boltzmann constant, T is the temperature and R (t) is a univariate Gaussian random process. This equation indicates the forces on a particle result from three sources: (i) The forces due to the interactions with other particles represented by $F\,(\vec{r})$. (ii) A dissipative force proportional to the velocity (the proportionality constant is the friction coefficient). The friction coefficient and the collision frequency ξ are related by the mass of the particle $\xi = \gamma/m$. ξ^{-1} is the velocity relaxation time, the elapsed time for a particle to lose the memory of its initial velocity. It is required to assign a value to the collision frequency in a stochastic

dynamics. (iii) A random fluctuating force that arises from the interactions with solvent molecules. It is assumed that the random forces obey a Gaussian distribution with zero mean. [90]

An extension of the Verlet algorithm called the Brünger-Brooks-Karplus (BBK) method is used in the NAMD simulation package to integrate the Langevin equation: [91,104]

$$\vec{r}_{(t+\Delta t)} = \vec{r}_t + \frac{1-\gamma\Delta t/2}{1+\gamma\Delta t/2}\left(\vec{r}_t - \vec{r}_{(t-\Delta t)}\right) + \frac{1}{1+\gamma\Delta t/2}\Delta t^2\left[M^{-1}F(\vec{r})_t + \sqrt{\frac{2\gamma k_B T}{\Delta M}}\ Z_t\right] \quad (1.28)$$

where Z_t is a set of Gaussian random variables with zero mean and variance 1 and each degree of freedom necessitates only one random number in the BBK method.

1. 7. Free energy via molecular simulation

The free energy is a state function given by the subtraction between the internal energy of a system and the product of its temperature and entropy. The Helmholtz free energy applies to systems with a fixed number of particles N, T and V. In classical statistical mechanics all of the observables are calculated using the partition function of the corresponding ensemble. [90] The following equation give the Helmholtz free energy of the canonical (NVT):

$$A = -k_B T \ln Z_N(N,V,T) \quad (1.29)$$

where Z_N is the configurational partition function of the ensemble represented by:

$$Z_N = \int \ldots \int e^{-E(\vec{q}^N)/k_B T}\ d\vec{q}^N \quad (1.30)$$

and $E(\vec{q}^N)$ is the configurational energy and the integration expands over the position element of all particles $d\vec{q}^N$. The formalism generalized to phase space is not the focus of this work. Computation of the absolute value of the partition function is a challenging

task because an extensive sampling of the configuration space is required, including the high-energy regions. In consequence, it is difficult to determine the absolute value of the free energies. Nevertheless, it is easier to calculate relative free energies because the sampling of a smaller region is required. One can define the free energy of the system relative to an ideal gas reference state, $E(\vec{q}^N) = 0$, and multiply the configurational partition function by a normalization factor of $1/(8\pi^2 V)^N$. The orientational factor $8\pi^2$ is replaced by unity for a macromolecular system. [75,83] The relative free energy ΔA is the energy difference between two well-defined n and m. The relative free energy is expressed in terms of the ratio of the partition functions corresponding to two states:

$$\Delta A = A_m - A_n = -k_B T \ln\left(\frac{Z_m}{Z_n}\right) \quad (1.31)$$

where A_m and A_n denote the free energy associated with state m and n, respectively, and Z_m and Z_n are the corresponding partition functions. Individual determination of Z_m and Z_n suffers from the same difficulties addressed previously. However, the potential energy can be defined as a function of a continuous coupling parameter λ, i.e. $E(\lambda)$. Then, λ is varied from n to m such that $E(\lambda)$ passes smoothly from E_n to E_m. The original idea of using coupling parameter was first introduced in Kirkwood's work. [105] One may consider the coupling parameter λ as a generic extent parameter, which describes the progress of the system along a path connecting the initial (n) and final (m) states. In general, λ involves the molecular topography changes and it is selected by considering the conceptual and numerical requirements associated with the studied problem. Thus, equation (1. 31) may be rewritten as:

$$A(\lambda) = -k_B T \ln Z(\lambda) \quad (1.32)$$

Various methods for calculating relative free energies exist. In the thermodynamic integration approach, an integration over the derivative of $A(\lambda)$ along λ is carried out. [105] The perturbation method designs closely related intermediate states associated with $A(\lambda_i)$ and computes ΔA in a stepwise manner using equation (1. 31). [106] In an alternative approach, $A(\lambda)$ is actually developed on the [n, m] interval over the course of the simulation where λ is variable and finally ΔA is calculated via $A(\lambda = m) - A(\lambda = n)$. The free energy along λ is known as potential of mean force.

λ can represent a physically meaningful pathway and result in the determination of the free energy of activation $\Delta A^{\ddagger}$ as well as ΔA. Having the free energy of activation in hand allows one to estimate the rate constant using transition state theory. [107] However, since the free energy is a state function, one may construct a non-physical or fictitious path between the initial and final states. It is often convenient computationally to consider non-physical pathways. This strategy is exploited in "mutational processes" (employed frequently in energy perturbation calculations) to define a chemical process by changing a functional group, a subunit and even the whole molecule into another. [53]

In general, theoretical methods designed to calculate free energy are distinguished in terms of geometrical and alchemical transformations. Alchemical free energy calculations are often achieved using perturbation theory and thermodynamic integration methods. There are two classes of algorithms to carry out alchemical transformations known as single-topology and dual-topology models. In the former, the transformation is performed using a general extent parameter to modify selective non-bonded parameters of the force field and express them as a linear combination of their end-point values. [53] In the latter, the initial and final states of the alchemical transformation are defined by

closely related but non-interacting states. The intermolecular interaction of the perturbed moiety with its surrounding is obtained by a linear combination of the end-state potential energy functions through the same general extent parameter.[108] On the other hand, geometrical free energy calculations do not require any chemical alteration of the studied system. In this approach a reaction coordinate is defined and the free energy change is calculated along the selected reaction coordinate frequently referred to as collective-variable. The reaction coordinate describes the minimum free energy pathway connecting the initial and final states of the transformation. In the simplest scenario, λ represents the separation of two tagged particles, which are initially at the distance $\vec{r}$ and then their distance increase to infinity.

There is a separate class of free energy calculation methods that refers to non-equilibrium work. Despite equilibrium geometrical free energy calculations, a constant or time-dependent force is applied to explore rare events, which leads to considerable deviation from the equilibrium conditions. For example, steered and force-probe molecular dynamics involve non-equilibrium work. [109,110]

There are a variety of collective-variable-based methods to achieve a geometrical transformation. One of the most popular approaches called umbrella sampling will be described in detail in the following section.

1. 7. 1. Potential of mean force and umbrella sampling

Potential of mean force (PMF), known as the free energy along a chosen coordinate, is a key quantity in computational studies of the macromolecules because it expresses conformational equilibrium properties or transition rate of dynamical activated processes. The coordinate is a function depending on a few or several degrees of freedom (such as

an angle, a distance and etc.) in a dynamical system. PMF calculation for a process along a physical pathway results in determination of transition state as the highest energy point on the free energy profile. Therefore, it is feasible to compute kinetic quantities such as rate constant. The generalized formalism of PMF is as following: [111]

$$W(\xi) = W(\xi^*) - k_B T \ln\left[\frac{\langle\rho(\xi)\rangle}{\langle\rho(\xi^*)\rangle}\right] \quad (1.33)$$

where ξ^* and $W(\xi^*)$ are the arbitrary reference values. The Boltzmann weighted average gives the average distribution function by:

$$\langle\rho(\xi)\rangle = \frac{\int \delta(\xi'[\vec{r}] - \xi)\, e^{-U(\vec{r})/k_B T}\, d\vec{r}}{\int e^{-U(\vec{r})/k_B T}\, d\vec{r}} \quad (1.34)$$

where $U(\vec{r})$ is the total energy of the system and $\delta(\xi'[\vec{r}] - \xi)$ is the Dirac function for the geometrical coordinate ξ and $\xi'[\vec{r}]$. $\xi'[\vec{r}]$ denotes a function that depends on one or several degrees of freedom in studied system.

A conventional molecular dynamics is used to explore the time evolution of the many-body systems on the time scale of 10^{-14} s to 10^{-8} s. However, the upper limit can change based on available computing power. [86] In a conventional MD, the presence of high-energy barrier along a selected coordinate ξ hinders a sufficient sampling of the phase space during the affordable simulation time. Traditional technique utilized to prevent this problem is known as umbrella sampling. This technique was first introduced in Torrie and Valleau work to estimate the free energy using Monte Carlo method. [112] In this method potential energy function of a system is modified by adding an artificial biasing window potential $V(\xi)$. The window potential makes neighboring region of ξ more favorable in terms of potential energy and enhances the sampling around the chosen ξ. The use of word "window" refers to the fact that the coordinate changes within an interval around some predetermined ξ value as the result of adding biasing potential. One

of the widely used functions for biasing potential is harmonic function. Thus, the potential energy function is described by:

$$U'(\vec{r}) = U(\vec{r}) + V(\xi) \quad (1.35)$$

$$V_i(\xi) = \frac{1}{2} K (\xi - \xi_i)^2 \quad (1.36)$$

The harmonic function is centered on the adjacent values of ξ_i and K is the force constant. Due to the harmonic function, the sampling is confined to a small region around ξ_i and thus the adequate sampling of this region gives an accurate small piece of PMF. Therefore, several numbers of the biasing window potentials are used along ξ to sample the entire region and obtain the complete PMF. Nevertheless, the distributions that are resulted from various windows must be unbiased first and then combined together to estimate the final $W(\xi)$. Biased distribution corresponding to *i*th window is obtained by:

$$\langle \rho(\xi) \rangle_i = e^{-V_i(\xi)/k_B T} \langle \rho(\xi) \rangle \langle e^{-V_i(\xi)/k_B T} \rangle^{-1} \quad (1.37)$$

Unbiased PMF resulting from *i*th window is calculated by:

$$W_i(\xi) = W(\xi^*) - k_B T \ln \left[\frac{\langle \rho(\xi) \rangle_i}{\langle \rho(\xi^*) \rangle} \right] - V_i(\xi) + F_i \quad (1.38)$$

The potential of each window $V_i(\xi)$ is known and F_i is an undetermined constant that represents the free energy due to employed window potential. It is obtained from:

$$e^{-F_i/k_B T} = \langle e^{-V_i(\xi)/k_B T} \rangle \quad (1.39)$$

In traditional methods the unknown energy constants F_i are determined while $W_i(\xi)$ of the neighboring windows are adjusted in the overlapped region until they match. [113] Least-squares procedure is often used to perform the matching. [114] Next, multiple $W_i(\xi)$ are connected to generate PMF and superfluous data in the overlapped regions are discarded. Because of individual determination of F_i in each window, a considerable overlap between the neighboring windows is required to circumvent the statistical errors.

Therefore, a large number of data points are not utilized during the process. Moreover, the matching process is somehow arbitrary and the cumulative errors resulting from the uncertainty involved in the process increases with the number of windows. Kumar et al proposed an alternative approach called weighted histogram analysis method (WHAM) in order to use all of the information obtained by umbrella sampling.[115] WHAM is the most reliable method to unbias and recombine all of the data from various windows. Furthermore, it prevents the problems mentioned earlier. It is also feasible to extend it to multidimensional free energy problem. In this approach, optimal unbiased distribution function is estimated as a weighted sum over the data obtained from N_w biased windows:

$$\langle \rho(\xi) \rangle = \sum_{i=1}^{N_w} [\langle \rho(\xi) \rangle]_i^{unbiased} \times \left[\frac{n_i \, e^{-[V_i(\xi) - F_i]/k_B T}}{\sum_{j=1}^{N_w} n_j \, e^{-[V_j(\xi) - F_j]/k_B T}} \right] \quad (1.40)$$

$$[\langle \rho(\xi) \rangle]_i^{unbiased} = e^{+V_i(\xi)/k_B T} \, \langle \rho(\xi) \rangle_i \, e^{-F_i/k_B T} \quad (1.41)$$

where n_i denotes the number of independent data points used to generate the biased distribution function. Then, the optimal estimate for the distribution function is used to give the free energy constants F_i:

$$e^{-F_i/k_B T} = \int d\xi \, e^{-V_i(\xi)/k_B T} \, \langle \rho(\xi) \rangle \quad (1.42)$$

The distribution function itself depends on the free energy constants. Thus, one must solve equations (1. 40) and (1. 42) self-consistently via an iterative procedure. First, an initial guess of N_w free energy constants is used to estimate the unbiased distribution function. Next, this estimate is exploited to construct a new set of free energy constants. The iteration cycle is repeated until both equations are satisfied within the predetermined threshold for the convergence. Finally, the functional form of the weighted factors is

provided that minimizes the statistical error. A review on different available techniques to unbias and combine the umbrella sampling data is available in reference [111].

PMF calculation provides the free energy landscape of the process of the interest. Transition state theory can be used to extract the rate constants from the calculated activation free energy. Following section presents on overview of the transition state theory.

1. 8. Transition state theory

In a theoretical study, a reaction's rate constant is determined accurately by solving the time independent Schrödinger equations for a variety of the nuclei configurations to obtain complete potential energy surface. Then, this PES is used to carry out classical trajectory calculations for the numerous initial reactant states and average over the results to extract the rate constant. This method is applicable to heavy atoms and also tunneling effects must be negligible. Force fields are practical alternatives for the PES to study larger systems. Furthermore, rate constants can be approximated by a simpler approach such as transition state theory (TST) to overcome the enormous difficulties involved in the described procedure.

TST has been used over decades, since its development in the 1930s to present, to approximate a reaction's rate constant by choosing a boundary surface located between reactant and product regions.[116-118] The boundary surface is known as a dividing surface, which reactants pass through its saddle point to produce products. Nuclear configuration corresponding to the saddle point of the dividing surface or any point within a short distance beyond this surface (δ) forms the transition state of the reaction. A basic

assumption of TST indicates that all of the transition states cross the saddle point of the dividing surface and form the products.

Free energy difference between two molecular states is one of the central experimental observables. It can be used to determine binding affinities in equilibrium thermodynamics or to calculate rate constants in kinetics. In the present study, dissociation rate constants were computed from the free energy difference between the bound protein-ligand complex and newly created unbound protein-ligand, which is estimated from the PMF along the reaction coordinate. Thus, the thermodynamic formulation of TST for a gas phase reactions is presented here. The reaction rate for an elementary reaction $B + C + \cdots \rightarrow products$ is given by:

$$r = k\,[B]\,[C]\,\ldots \tag{1.43}$$

$$k = \frac{k_B T}{h}\,\frac{Z'_{\ddagger}/N_A V}{(Z_B/N_A V)\,(Z_C/N_A V)\ldots}\,e^{-\Delta\varepsilon_0^{\ddagger}/k_B T} \tag{1.44}$$

where k denotes the reaction rate constant and $Z'_{\ddagger}$ represents the transition state partition function. However, partition function for the motion along the reaction coordinate is omitted in $Z'_{\ddagger}$. Z_B and Z_C are the partition function of the reactants, N_A is Avogadro's number and V is the volume. N_A and V constants are used to cast the equation in terms of the concentration. $\Delta\varepsilon_0^{\ddagger}$ is the energy difference between the transition state in its lowest energy state and the reactants B, C, … in their lowest energy states. The formulation of rate constant can be rewritten based on thermodynamics quantities as following:

$$k = \frac{k_B T}{h}\,K_c^{\ddagger} \tag{1.45}$$

$$K_c^{\ddagger} = K_f\,\left(\frac{2}{\pi m_{rc} k_B T}\right)^{1/2}\,\frac{h}{\delta} \tag{1.46}$$

where $K_f = \left[X_f^{\ddagger}\right]/[B][C]$ is the formation constant of the transition state. $\left(\frac{2}{\pi m_{rc} k_B T}\right)^{1/2} \frac{h}{\delta}$ takes into account one-dimensional translational motion of the transition state as a free particle (there is no force at the saddle point), which was neglected in $Z_{\ddagger}'$ function. $\frac{1}{\delta}$ denotes that the transition state passes the saddle point of the dividing surface or a surface located in δ distance beyond the dividing surface to form the products. m_{rc} is the effective mass for the motion along the reaction coordinate. Finally, the rate constant equation can be rewritten in terms of free energy difference between the reactants and the transition state as following:

$$k = \frac{k_B T}{h} (c^0)^{1-n} e^{-\Delta G_c^{0\ddagger}/RT} \quad (1.47)$$

where $\Delta G_c^{0\ddagger} = -RT \ln\left[K_c^{\ddagger} (c^0)^{n-1}\right]$ (1. 48)

The ^{0} symbol represents the standard condition where c^0 denotes the concentration equal to 1 mol dm^{-3} and the molecularity of the reaction is shown by n. Equation (1. 47) indicates that higher value of $\Delta G_c^{0\ddagger}$ results in slower reaction. [119]

Based on the definition of PMF, equation (1. 47) can be rewritten as following for an unimolecular reaction:

$$k = \frac{k_B T}{h} e^{-[-W(s) - A_{\xi}^{R}]/k_B T} \quad (1.49)$$

where $W(s)$ is the PMF value at $\xi = s$ (transition state) and A_{ξ}^{R} is the reversible work associated with setting ξ to the reactant value R for a non-separable reaction coordinate.

1. 9. Goals and overview of thesis

The goals of the studies reported in this thesis are: (i) To elucidate the microscopic details of the intrinsic interactions between fluorinated and non-fluorinated fatty acid ligands and Lg protein. (ii) To explore whether free energy calculations allow us to predict the

dissociation kinetics of Lg-fatty acid complexes in the absence of solvent. (iii) We are interested in investigation of the sequence of bond breaking events during the ligand exit from the protein binding site. (iv) To investigate how fluorination influences the protein-ligand intermolecular interactions. The theoretical modeling presented in this thesis is complementary to the experimental studies. In general, the works presented in this thesis rely on the interplay between theory and experiment, where the experimental dissociation rate constants were obtained using the BIRD technique in mass spectrometry. [15,34]

Free energy calculations were performed on desolvated protein-fatty acid (β-lactoglobulin and stearic acid, palmitic acid, myristic acid and lauric acid ligands) complexes. The details of the computational procedure and results will be described in Chapter 2. In this work, umbrella sampling is used to compute the potential of mean force along the dissociation coordinate for these complexes in two different charge configurations. The calculated PMFs were used to provide an estimation of activation free energies for dissociation of the fatty acid ligands from Lg. Ultimately; the calculated activation free energies were utilized to extract the dissociation rate constants via transition state theory. Consistent with the experiment, dissociation of the stearic acid was predicted as the slowest dissociation process in the first studied charge configuration. The trend of the experimental dissociation rate constants was reproduced for the second charge configuration, except for one of the fatty acid ligands. The results provide an opportunity to identify the strengths and shortcomings of the theoretical approach in the prediction of the kinetics of the protein-ligand dissociation process. The theoretical results suggest that the ligand dissociation from the protein binding site occurs upon the

cleavage of the carboxyl group interactions followed by the dissociation of the aliphatic chain from the binding cavity.

In Chapter 3, the description of MD simulations performed on protein-fluorinated fatty acid complexes (β-lactoglobulin and fluorinated analogs of stearic acid containing X fluorine atoms (X = 13 and 21)) is presented. The role of fluorine atoms in enhancing the intrinsic interactions between protein and ligands was clarified. The histograms of the distances and angles between fluorine atoms and neighboring protein atoms demonstrate that the fluorine atoms are involved in the interactions with polar hydrogen atoms in the binding cavity. Also, it was shown that the degree of fluorination influences the fatty acid carboxyl group interactions with protein residues.

References

1. Huang, J. F.; Yang, Y.; Sepulveda, H.; Shi, W. X.; Hwang, I.; Peterson, P. A.; Jackson, M. R.; Sprent, J.; Cai, Z. L. TCR-mediated internalization of peptide-MHC complexes acquired by T cells. *Science* **1999***, 286*, 952-954.

2. Arikawa, T.; Simamura, E.; Shimada, H.; Nakamura, T.; Hatta, T.; Shoji, H. Significance of sugar chain recognition by galectins and its involvement in disease-associated glycosylation. *Congenital anomalies* **2014***, 54*, 77-81.

3. Igakura, T.; Stinchcombe, J. C.; Goon, P. K. C.; Taylor, G. P.; Weber, J. N.; Griffiths, G. M.; Tanaka, Y.; Osame, M.; Bangham, C. R. M. Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton. *Science* **2003***, 299*, 1713-1716.

4. Patra, S. M.; Bastug, T.; Kuyucak, S. Binding of organic cations to gramicidin a channel studied with AutoDock and molecular dynamics simulations. *J. Phys. Chem. B* **2007***, 111*, 11303-11311.

5. Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009***, 113*, 2234-2246.

6. Shirts, M. R.; Mobley, D. L.; Chodera, J. D. Alchemical Free Energy Calculations: Ready for Prime Time? *Ann. Rep. Comput. Chem.* **2007***, 3*, 41-59.

7. Chipot, C. Frontiers in free-energy calculations of biological systems. *Rev. Comput. Mol. Sci.* **2014**, *4*, 71-89.

8. Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395-2417.

9. Woo, H.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825-6830.

10. Lee, M.; Olson, M. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys. J.* **2006**, *90*, 864-877.

11. Jorgensen, W. L. Interactions between Amides in Solution and the Thermodynamics of Weak Binding. *J. Am. Chem. Soc.* **1989**, *111*, 3770-3771.

12. Hermans, J.; Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.* **1997**, *119*, 2707-2714.

13. Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006**, *125*, 084902.

14. Woo, H.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825-6830.

15. Liu, L.; Bagal, D.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Hydrophobic Protein-Ligand Interactions Preserved in the Gas Phase. *J. Am. Chem. Soc.* **2009***, 131*, 15980-15981.

16. Liu, L.; Michelsen, K.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Energetics of Lipid Binding in a Hydrophobic Protein Cavity. *J. Am. Chem. Soc.* **2012***, 134*, 3054-3060.

17. Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J. Chem. Phys.* **2006***, 125*, 84901-84905.

18. Chipot, C.; Rozanska, X.; Dixit, S. B. Can free energy calculations be fast and accurate at the same time? Binding of low-affinity, non-peptide inhibitors to the SH2 domain of the src protein. *J. Comput. Aided Mol. Des.* **2005***, 19*, 765-770.

19. Freiteg, S.; Chu, V.; Penzotti, J.; Klumb, L.; To, R.; Hyre, D.; Trong, I.; Lybrand, T.; Stenkamp, R.; Stayton, P. A structural snapshot of an intermediate on the streptavidin–biotin dissociation pathway. *Proc. Natl. Acad. Sci.* **1999***, 96*, 8384-8389.

20. Benitex, Y.; Baranger, A. M. Control of the Stability of a Protein-RNA Complex by the Position of Fluorine in a Base Analogue. *J. Am. Chem. Soc.* **2011***, 133*, 3687-3689.

21. Kim, C.; Chang, J.; Doyon, J.; Baird, T.; Fierke, C.; Jain, A.; Christianson, D. Contribution of fluorine to protein-ligand affinity in the binding of fluoroaromatic inhibitors to carbonic anhydrase II. *J. Am. Chem. Soc.* **2000***, 122*, 12125-12134.

22. Lee, Y.; Zeng, H.; Ruedisser, S.; Gosser, A. D.; Hilty, C. Nuclear Magnetic Resonance of Hyperpolarized Fluorine for Characterization of Protein-Ligand Interactions. *J. Am. Chem. Soc.* **2012***, 134*, 17448-17451.

23. Bohm, H.; Banner, D.; Bendels, S.; Kansy, M.; Kuhn, B.; Muller, K.; Obst-Sander, U.; Stahl, M. Fluorine in medicinal chemistry. *Chembiochem* **2004***, 5*, 637-643.

24. Kawahara, S.; Tsuzuki, S.; Uchimaru, T. Theoretical study of the C-F-π interaction: Attractive interaction between fluorinated alkane and an electron-dericient π - system. *J. Phys. Chem. A* **2004***, 108*, 6744-6749.

25. Riley, K.; Merz, K. Effects of fluorine substitution on the edge-to-face interaction of the benzene dimer. *J. Phys. Chem. B* **2005***, 109*, 17752-17756.

26. Iwaoka, M.; Komatsu, H.; Katsuda, T.; Tomoda, S. Quantitative evaluation of weak nonbonded Se...F interactions and their remarkable nature as orbital interactions. *J. Am. Chem. Soc.* **2002***, 124*, 1902-1909.

27. Chopra, D.; Nagarajan, K.; Row, T. N. G. Analysis of weak interactions involving organic fluorine: Insights from packing features in substituted 4-keto-tetrahydroindoles. *J. Mol. Struct.* **2008***, 888*, 70-83.

28. Lee, S.; Mallik, A.; Fredrickson, D. Dipolar-dipolar interactions and the crystal packing of nitriles, ketones, aldehdyes, and C(sp(2))-F groups. *Cryst. Growth Des.* **2004***, 4*, 279-290.

29. Matta, C.; Castillo, N.; Boyd, R. Characterization of a closed-shell fluorine-fluorine bonding interaction in aromatic compounds on the basis of the electron density. *J Phys Chem A* **2005***, 109*, 3669-3681.

30. Olsen, J.; Banner, D.; Seiler, P.; Sander, U.; D'Arcy, A.; Stihle, M.; Muller, K.; Diederich, F. A fluorine scan of thrombin inhibitors to map the fluorophilicity/fluorophobicity of an enzyme active site: Evidence for C-F center dot center dot center dot C=O interactions. *Angew. Chem. Int. Edit.* **2003***, 42*, 2507-2511.

31. Bettinger, H. How good is fluorine as a hydrogen-bond acceptor in fluorinated single-walled carbon nanotubes? *ChemPhysChem* **2005***, 6*, 1169-1174.

32. Mueller, K.; Faeh, C.; Diederich, F. Fluorine in pharmaceuticals: Looking beyond intuition. *Science* **2007***, 317*, 1881-1886.

33. Mecinović, J.; Snyder, P. W.; Mirica, K. A.; Bai, S.; Mack, E. T.; Kwant, R. L.; Moustakas, D. T.; Héroux, A.; Whitesides, G. M. Fluoroalkyl and Alkyl Chains Have Similar Hydrophobicities in Binding to the "Hydrophobic Wall" of Carbonic Anhydrase. *J. Am. Chem. Soc.* **2011***, 133*, 14017-14026.

34. Liu, L.; Jalili, N.; Baergen, A.; Ng, S.; Bailey, J.; Derda, R.; Klassen, J. S. Fluorine Bonding Enhances the Energetics of Protein-Lipid Binding in the Gas Phase. *J. Am. Soc. Mass Spectrom.* **2014***, 25*, 751-757.

35. Kontopidis, G.; Holt, C.; Sawyer, L. Invited Review: beta-lactoglobulin: Binding properties, structure, and function. *J. Dairy Sci.* **2004***, 87*, 785-796.

36. Qin, B.; Bewley, M.; Creamer, L.; Baker, H.; Baker, E.; Jameson, G. Structural basis of the tanford transition of bovine beta-lactoglobulin. *Biochemistry (N. Y.)* **1998***, 37*, 14014-14023.

37. Flower, D. R.; North, A. C. T.; Attwood, T. K. Structure and Sequence Relationships in the Lipocalins and Related Proteins. *Protein Science* **1993***, 2*, 753-761.

38. Qin, B.; Bewley, M.; Creamer, L.; Baker, H.; Baker, E.; Jameson, G. Structural basis of the tanford transition of bovine beta-lactoglobulin. *Biochemistry (N. Y.)* **1998***, 37*, 14014-14023.

39. Kontopidis, G.; Holt, C.; Sawyer, L. Invited Review: beta-lactoglobulin: Binding properties, structure, and function. *J. Dairy Sci.* **2004***, 87*, 785-796.

40. Collini, M.; D'Alfonso, L.; Molinari, H.; Ragona, L.; Catalano, M.; Baldini, G. Competitive binding of fatty acids and the fluorescent probe 1-8-anilinonaphthalene sulfonate to bovine beta-lactoglobulin. *Protein Sci.* **2003***, 12*, 1596-1603.

41. Konuma, T.; Sakurai, K.; Goto, Y. Promiscuous binding of ligands by beta-lactoglobulin involves hydrophobic interactions and plasticity. *J. Mol. Biol.* **2007***, 368*, 209-218.

42. Ohtomo, H.; Konuma, T.; Utsunoiya, H.; Tsuge, H.; Ikeguchi, M. Structure and stability of Gyuba, a beta-lactoglobulin chimera. *Protein Sci.* **2011***, 20*, 1867-1875.

43. Sawyer, L.; Kontopidis, G. The core lipocalin, bovine beta-lactoglobulin. *Biochim. Biophys. Acta-Protein Struct. Molec. Enzym.* **2000***, 1482*, 136-148.

44. Levin, L. B.; Ganoth, A.; Amram, S.; Nachliel, E.; Gutman, M.; Tsfadia, Y. Insight into the interaction sites between fatty acid binding proteins and their ligands. *J. Mol. Model.* **2010***, 16*, 929-938.

45. Long, D.; Mu, Y.; Yang, D. Molecular Dynamics Simulation of Ligand Dissociation from Liver Fatty Acid Binding Protein. *PLoS One* **2009***, 4*, e6081.

46. Tsfadia, Y.; Friedman, R.; Kadmon, J.; Selzer, A.; Nachliel, E.; Gutman, M. Molecular dynamics simulations of palmitate entry into the hydrophobic pocket of the fatty acid binding protein. *FEBS Lett.* **2007***, 581*, 1243-1247.

47. Friedman, R.; Nachliel, E.; Gutman, M. Fatty acid binding proteins: Same structure but different binding mechanisms? Molecular dynamics simulations of intestinal fatty acid binding protein. *Biophys. J.* **2006***, 90*, 1535-1545.

48. Friedman, R.; Nachliel, E.; Gutman, M. Molecular dynamics simulations of the adipocyte lipid binding protein reveal a novel entry site for the ligand. *Biochemistry (N. Y.)* **2005***, 44*, 4275-4283.

49. Ragona, L.; Fogolari, F.; Zetta, L.; Perez, D.; Puyol, P.; De Kruif, K.; Lohr, F.; Ruterjans, H.; Molinari, H. Bovine beta-lactoglobulin: Interaction studies with palmitic acid. *Protein Sci.* **2000***, 9*, 1347-1356.

50. Dunbar, R.; McMahon, T. Activation of unimolecular reactions by ambient blackbody radiation. *Science* **1998***, 279*, 194-197.

51. Price, W.; Schnier, P.; Jockusch, R.; Strittmatter, E.; Williams, E. Unimolecular reaction kinetics in the high-pressure limit without collisions. *J. Am. Chem. Soc.* **1996***, 118*, 10640-10644.

52. McCammon, J.; Gelin, B.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977***, 267*, 585-590.

53. Jorgensen, W.; Ravimohan, C. Monte-Carlo Simulation of Differences in Free-Energies of Hydration. *J. Chem. Phys.* **1985***, 83*, 3050-3054.

54. Lybrand, T.; Mccammon, J.; Wipff, G. Theoretical Calculation of Relative Binding-Affinity in Host Guest Systems. *Proc. Natl. Acad. Sci. U. S. A.* **1986***, 83*, 833-835.

55. Postma, J.; Berendsen, H.; Haak, J. Thermodynamics of Cavity Formation in Water - a Molecular-Dynamics Study. *Faraday Symposia Chem. Soc.* **1982**, 55-67.

56. Warshel, A. Dynamics of Reactions in Polar-Solvents - Semi-Classical Trajectory Studies of Electron-Transfer and Proton-Transfer Reactions. *J. Phys. Chem.* **1982**, *86*, 2218-2224.

57. Kroon, G.; Mo, H.; Martinez-Yamout, M.; Dyson, H.; Wright, P. Changes in structure and dynamics of the Fv fragment of a catalytic antibody upon binding of inhibitor. *Protein Sci.* **2003**, *12*, 1386-1394.

58. Thorpe, I. F.; Brooks,Charles L.,,III Molecular evolution of affinity and flexibility in the immune system. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 8821-8826.

59. Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255-1273.

60. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field (vol 25, pg 1157, 2004). *J. Chem. Theory Comput.* **2005**, *26*, 114-114.

61. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

62. Shirts, M.; Pitera, J.; Swope, W.; Pande, V. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003***, 119*, 5740-5761.

63. Deng, Y.; Roux, B. Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B* **2004***, 108*, 16567-16576.

64. Wan, S.; Stote, R.; Karplus, M. Calculation of the aqueous solvation energy and entropy, as well as free energy, of simple polar solutes. *J. Chem. Phys.* **2004***, 121*, 9539-9548.

65. Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *J. Med. Chem.* **2008***, 51*, 769-779.

66. Mobley, D. L.; Chodera, J. D.; Dill, K. A. Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput.* **2007***, 3*, 1231-1235.

67. Morton, A.; Baase, W. A.; Matthews, B. W. Energetic Origins of Specificity of Ligand-Binding in an Interior Nonpolar Cavity of T4 Lysozyme. *Biochemistry (N. Y.)* **1995***, 34*, 8564-8575.

68. Rodinger, T.; Howell, P. L.; Pomes, R. Calculation of absolute protein-ligand binding free energy using distributed replica sampling. *J. Chem. Phys.* **2008***, 129*, 155102.

69. Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. Predicting Ligand Binding Affinity with Alchemical Free Energy Methods in a Polar Model Binding Site. *J. Mol. Biol.* **2009***, 394*, 747-763.

70. Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006***, 91*, 2798-2814.

71. Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C.; Shirts, M.; Sorin, E.; Pande, V. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **2005***, 123*, 084108.

72. Gan, W.; Roux, B. Binding specificity of SH2 domains: Insight from free energy simulations. *Proteins: Struct. Funct. Bioinf.* **2009***, 74*, 996-1007.

73. Gumbart, J. C.; Roux, B.; Chipot, C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.* **2013***, 9*, 794-802.

74. Lau, A. Y.; Roux, B. The hidden energetics of ligand binding and activation in a glutamate receptor. *Nat. Struct. Mol. Biol.* **2011***, 18*, 283-288.

75. Beveridge, D.; Dicapua, F. Free-Energy Via Molecular Simulation - Applications to Chemical and Biomolecular Systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989***, 18*, 431-492.

76. Luo, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc. Natl. Acad. Sci.* **2002***, 99*, 10399-10404.

77. Massova, I.; Kollman, P. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery* **2000***, 18*, 113-135.

78. Kirkwood, J. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935***, 3*, 300-313.

79. Zwanzig, R. High-Temperature Equation of State by a Perturbation Method .1. Nonpolar Gases. *J. Chem. Phys.* **1954***, 22*, 1420-1426.

80. Allen, T.; Andersen, O.; Roux, B. Energetics of ion conduction through the gramicidin channel. *Proc. Natl. Acad. Sci. U. S. A.* **2004***, 101*, 117-122.

81. Levine, I. N. *Physical Chemistry;* McGraw-Hill: New York, 2009; .

82. Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys. (Berlin)* **1927***, 389*, 457-484.

83. McQuarrie, D. A. *Statistical Mechanics;* University Science Books: Sausalito, CA, 2000.

84. Pathria, R. K. *Statistical Mechanics;* Reed Educational and Professional Publishing Ltd: Oxford, 1996.

85. Chandler, D. *Introduction to modern statistical mechanics;* Oxford University Press: New York, 1987.

86. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: from Algorithm to Application;* Academic Press: San Diego, CA, 1996.

87. Maxwell, J. C. Illustrations of the dynamical theory of gases. *Philos. Mag.* **1860***, 19*, 19-32.

88. Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993***, 93*, 2395-2417.

89. Verlet, L. Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967***, 159*, 98-103.

90. Leach, A. R. *Molecular Modelling: Principles and Applications;* Addison Wesley Longman: England, 1996.

91. Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005***, 26*, 1781-1802.

92. Rychaert, J.; Ciccotti, G.; Berendsen, H. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977***, 23*, 327-341.

93. Wang, W.; Donini, O.; Reyes, C.; Kollman, P. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001***, 30*, 211-243.

94. Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995***, 117*, 5179-5197.

95. Nose, S. Constant-Temperature Molecular-Dynamics. *J. Phys. Condes. Matter* **1990***, 2*, SA115-SA119.

96. Hoover, W.; Ladd, A.; Moran, B. High-Strain-Rate Plastic-Flow Studied Via Non-Equilibrium Molecular-Dynamics. *Phys. Rev. Lett.* **1982***, 48*, 1818-1820.

97. Berendsen, H.; Postma, J.; Vangunsteren, W.; Dinola, A.; Haak, J. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984***, 81*, 3684-3690.

98. Andersen, H. C. Molecular-Dynamics Simulations at Constant Pressure And-Or Temperature. *J. Chem. Phys.* **1980***, 72*, 2384-2393.

99. Nose, S. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984***, 52*, 255-268.

100. Nosé, S. A Molecular-dynamics Method For Simulation in the Canonical Ensemble. *J. Mol. phys.* **1984***, 52*, 255-268.

101. Hoover, W. G. Canonical Dynamics - Equilibrium phase-space Distributions. *Phys. Rev. A* **1985***, 31*, 1695-1697.

102. Kubo, R.; Toda, M.; Hashitsume, N. *Statistical Physics II: Nonequilibrium Statistical Mechanics;* Springer: New York, 1991.

103. Adelman, S.; Doll, J. Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - General Formulation for Classical Scattering Off Harmonic Solids. *J. Chem. Phys.* **1976***, 64*, 2375-2388.

104. Brunger, A.; Brooks, C.; Karplus, M. Stochastic Boundary-Conditions for Molecular-Dynamics Simulations of St2 Water. *Chem. Phys. Lett.* **1984***, 105*, 495-500.

105. Kirkwood, J. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935***, 3*, 300-313.

106. Zwanzig, R. High-Temperature Equation of State by a Perturbation Method .1. Nonpolar Gases. *J. Chem. Phys.* **1954***, 22*, 1420-1426.

107. Glasstone, S.; Laidler, K.; Eyring, H. *The Thoery of Rate Processes;* McGraw-Hill: New York, 1941.

108. Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden Thermodynamics of Mutant Proteins - a Molecular-Dynamics Analysis. *Science* **1989***, 244*, 1069-1072.

109. Izrailev, S.; Stepaniants, S.; Schulten, K. Applications of steered molecular dynamics to protein-ligand/membrane binding. *Biophys. J.* **1998***, 74*, A177-A177.

110. Grubmuller, H.; Heymann, B.; Tavan, P. Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **1996***, 271*, 997-999.

111. ROUX, B. The Calculation of the Potential of Mean Force using Computer-Simulations. *Comput. Phys. Commun.* **1995***, 91*, 275-282.

112. Torrie, G.; Valleau, J. Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *J. Comput. Phys.* **1977***, 23*, 187-199.

113. Torrie, G.; Valleau, J. Monte-Carlo Free-Energy Estimates using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974***, 28*, 578-581.

114. Shen, J.; McCammon, J. Molecular-Dynamics Simulation of Superoxide Interacting with Superoxide-Dismutase. *Chem. Phys.* **1991***, 158*, 191-198.

115. Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. the Method. *J. Comput. Chem.* **1992***, 13*, 1011-1021.

116. Laidler, K.; King, M. The Development of Transition-State Theory. *J. Phys. Chem.* **1983***, 87*, 2657-2664.

117. Evans, A.; Polanyi, M. Some Aspects of the Chemistry of Macromolecules : Polymerisation of Isobutene by Friedel-Crafts Catalysts. *J. Chem. Soc.* **1947**, 252-257.

118. Pelzer, H.; Wigner, E. The speed constansts of the exchange reactions. *Z. Phys. Chem. B-Chem. Elem. Aufbau. Mater.* **1932***, 15*, 445-471.

119. Levine, I. N. *Physical Chemistry;* McGraw-Hill: New York, 2009.

Chapter 2

Dissociation kinetics of protein-fatty acid complexes

2. 1. Introduction

A comprehensive understanding of molecular recognition is achieved through the knowledge of free energy associated with the process. Free energy calculations provide insight into the mechanisms involved in recognition phenomena on an atomic level. For instance, they are able to identify the process of bond formation or bond breaking within involved moieties and can predict the probability of finding a system in a given state. [1] Generally, the dynamics of ligand entry and exit from protein binding sites, which is an important aspect of small-molecule recognition, can be understood by characterizing the energy landscape of the process. [2] Experimental and theoretical studies suggest that a relatively well-defined and ordered sequence of conformational change and bond formation events mediate the association of small molecules to the target proteins. Therefore, even though the association rates of small molecules normally fall near the diffusion limit, the dehydration and/or conformational rearrangement steps are involved as well. Moreover, some proteins interact with the ligand via some specific side chains to steer the molecule electrostatically and enhance the association rates. [3-5] It raises the question whether there is a well-defined ligand exit pathway as well. Complementary computational studies assist to answer whether the small molecules dissociate in a concerted sequence of bond breakings and protein conformational rearrangements or it is a stochastic process with a broad collection of pathways and barrier topologies. [2]

There are examples of well-characterized dissociation pathways such as dissociation of small gaseous ligands of heme-proteins (O_2 and CO) from myoglobin [3,6]

and biotin from streptavidin.[2,7,8] For these examples, potential of mean force calculations were used to obtain the free energies associated with the ligand dissociation and agreement between experimental and predicted values was satisfactory. More recently, the dissociation pathway of GW0072 drug from peroxisome proliferator activated receptor (PPARγ) was investigated using MD simulations and conformational rearrangement of protein was found to accompany the dissociation of the ligand.[9] Also, steered molecular dynamics was used to study the molecular basis of dissociation of the ligand from human β-adrenergic receptors (β_1AR and β_2AR). The authors suggested that there are some channels in extracellular side that are used as the exit pathway.[10] At present, there are a few examples of quantitative studies on dissociation of ligands from protein binding sites. Also, it is not clearly understood whether there is a well-defined ligand exit pathway. Experimental dissociation rate constants of desolvated β-lactoglobulin-fatty acids complexes are available and resemblance of desolvated and solvated complexes is well known.[11] Desolvated β-lactoglobulin-fatty acids complexes can be used as a model system to explore the feasibility of dissociation kinetics prediction using free energy calculations in the absence of solvent. Molecular simulation of fatty acids from β-lactoglobulin binding site may provide more insight into dissociation pathway of fatty acid ligands from protein binding site.

In the present work, geometrical transformation of the ligand along the reaction coordinate and the corresponding free energy change were studied using the umbrella sampling and WHAM methods.[12,13] The dissociation rate constants are calculated based on the free energy difference between bound and unbound states via transition state theory.[14-16] The free energy change associated with different fatty acid dissociation

processes are reflected in the corresponding dissociation rate constants. Details of the studied systems, umbrella sampling and WHAM were explained in Chapter 1 (Section 1. 2 and Section 1. 7. 1, respectively). Following section presents description of simulation details.

2. 2. Simulation details

2. 2. 1. Potential of mean force

Free energy profile along some chosen dissociation coordinate *r* refers to as PMF, which can be constructed from the average distribution function by: [17]

$$W(\xi) = W(\xi^*) - k_B T \ln\left[\frac{\langle\rho(\xi)\rangle}{\langle\rho(\xi^*)\rangle}\right] \quad (2.\ 1)$$

where ξ^* and $W(\xi^*)$ are arbitrary reference values, k_B is the Boltzmann constant and T is temperature. The PMF difference between bound and unbound states is required to calculate dissociation rate constants and thus choice of reference values is arbitrary. In order to obtain an accurate free energy profile, a sufficient sampling of configuration space is necessary, even high-energy regions. Umbrella sampling was used to enhance the sampling efficiency. [12]

2. 2. 1. 1. Reaction coordinate

Equation (2. 1) indicates that the PMF depends on the choice of reaction coordinate. Reaction coordinate is a mathematical object, which defines minimum free energy pathway connecting the initial state to final state of the transformation. In most of the studies geometrical parameters such as a distance, a dihedral angle and a torsional angle are chosen as reaction coordinate. [18-20] Improper choice of reaction coordinate can result in a simulation bias and slow convergence. Chandler *et al.* concluded that sometimes

coupling of additional degrees of freedom to the principle reaction coordinate is required to reproduce the accurate dynamical behavior of the system. [21]

In the current study, a Jacobi distance, which is often used in unimolecular decomposition reactions, was used as the reaction coordinate. The reaction coordinate $\xi = r_{AB}$, where $r_{AB} = |r_B - r_A|$, is defined as the distance between center of mass (COM) of two molecules or two molecular fragments. [22] In the present work, the distance between the COM of the fatty acid head group and the COM of protein binding cavity was chosen to represent the dissociation coordinate. The atoms constituting the carboxyl group of fatty acid (one carbon and two oxygen atoms) and the closest aliphatic carbon atom were selected to define the COM of the fatty acid head group. The backbone carbon atoms (C_α) of the amino acid residues located on eight β-strands enclosing the binding cavity were chosen to determine the COM of protein binding cavity. The residues involved in the definition of the COM of protein binding cavity are as follows: Tyr^{42}, Val^{43}, Glu^{44}, Glu^{45}, Leu^{46}, Lys^{47}, Leu^{54}, Glu^{55}, Ile^{56}, Leu^{57}, Leu^{58}, Gln^{59}, Gln^{68}, Lys^{69}, Lys^{70}, Ile^{71}, Ile^{72}, Ala^{73}, Lys^{91}, Val^{92}, Leu^{93}, Val^{94}, Leu^{95}, Asp^{96}, Leu^{103}, Leu^{104}, Phe^{105}, Cys^{106}, Met^{107}, Glu^{108}, Leu^{117}, Ala^{118}, Cys^{119}, Gln^{120}, Cys^{121}, Leu^{122} (See Figure 2. 1).

The choice of reaction coordinate relies on the fact that the aliphatic chain of the fatty acids is buried in the binding cavity and interacts with the residues in the lower region of the cavity with low specificity. This can be inferred from the ability of Lg binding cavity to accommodate a variety of lipid ligands such as saturated, unsaturated and branched fatty acids, retinoic acid and retinol. [23] Moreover, the linear relationship between the dissociation activation energy and the length of the fatty acid aliphatic chain in the fast component proves that $–CH_3$ and all of $–CH_2$ groups contribute in the

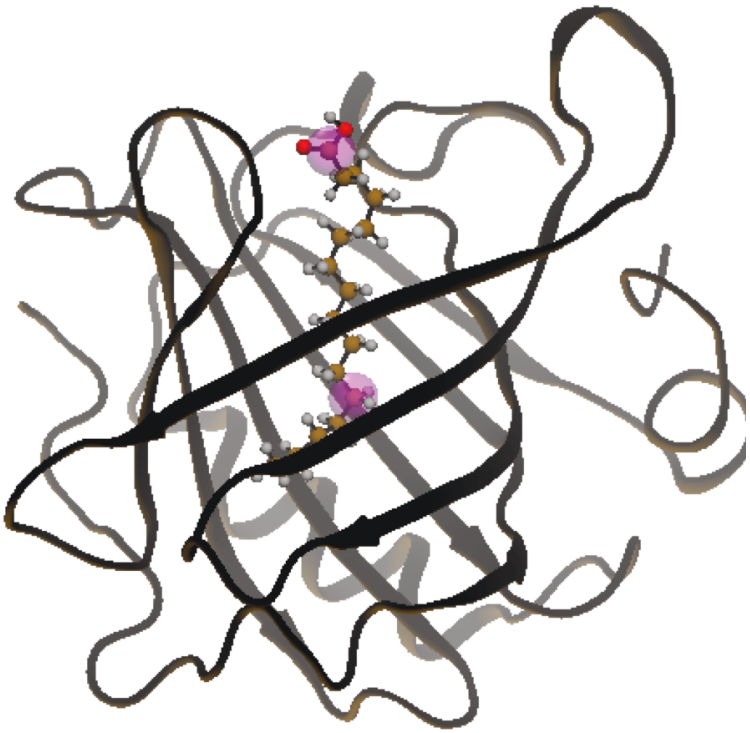


Figure 2. 1. The reaction coordinate used in free energy calculation: a representative example for (Lg+PA) where COM of fatty acid head group and protein binding cavity are shown by purple circles.

nonspecific nonpolar interactions with the residues in the homogeneous binding cavity. However, the head group of the fatty acid can be involved in directed hydrogen bond interactions. [24]

The simulation is converged if all of the relevant regions of configuration space are sampled sufficiently. To prevent the convergence problem due to short simulation time or being trapped in a metastable state, some additional geometrical restraints can be applied during the simulation. [25-27] In the present study, a single distance restraint between protein and ligand was used. Thus, all degrees of freedom of the ligand and protein must be sampled during the simulation. This could provide a more realistic picture of the ligand dissociation mechanism in the absence of knowledge of exact degrees of freedom accompanying the ligand dissociation from the binding cavity.

2. 2. 2. Umbrella sampling

Sampling enhancement along a coordinate of interest was achieved using the umbrella sampling method. [12] A detailed description of umbrella sampling method was presented in Chapter 1 (Section 1. 7. 1). A harmonic restraining potential given by equation (2. 2), was employed to bias the ligand around the specific distance at each window:

$$V = k_r (r - r_0)^2 \quad (2.\ 2)$$

where k_r is force constant, r is distance and r_0 is the center of applied force at each window. The force constants can be determined by the average values based on unbiased simulation. Therefore, the following equation is normally used to obtain the magnitude of the force constant: [27]

$$k_r = \frac{k_B T}{\langle \Delta r^2 \rangle} \quad (2.\ 3)$$

where k_B is the Boltzmann constant, T is temperature and Δr is the fluctuation of associated coordinate. In this work, however, the values determined by the above equation were not efficient in sampling the relevant regions. On the other hand, multiple restraining potentials are useful to sample long coordination path. Thus, the magnitudes of the force constants were determined by a few trials as follows: an 8.0 kcal $mol^{-1}Å^{-2}$ force constant was used to sample the region confined within r = 0 – 30 Å. The force constant magnitude was decreased to 5.0 kcal $mol^{-1}Å^{-2}$ for the longer distances where r = 30 – 50 Å. Finally, the force constant value was further reduced to 1.0 kcal $mol^{-1}Å^{-2}$ for r = 50 – 60 Å. Applying lower force constants at longer distances results in sampling a wider range of r and thus, less number of windows are needed in order to sample those regions efficiently. The total numbers of 180 windows were utilized to sample the entire reaction coordinate (r = 0 – 60 Å). The distance r = 0 – 30 Å that most likely

corresponds to the bound protein-ligand was sampled by windows centered at 0.25 Å intervals. The interval between the centers of windows at longer distances, r = 30 – 60 Å, was incremented to 0.5 Å.

2. 2. 3. Computational details

Structures of the fatty acid ligands were constructed using Avogadro 1.1.0 [28] in order to calculate partial charges on their atoms. The partial charge on each atom was computed using Gaussian 09 (C.01) on RESP ESP charge Drive Server. [29-31] Geometry optimization and charge derivation was performed using HF/6-31G* and RESP (Restrained Electrostatic Potential) methods. [32] In RESP method, atom-center charges are computed by fitting the charges to reproduce electrostatic potential (ESP) calculated at large number of grid points around the molecule. This approach introduces some restraints in the form of penalty functions in the fitting process to make the charges independent of the molecular conformations. To study deprotonated $(Lg+FA)^{-7}$ ions and all of their possible charge configurations, the partial charges for neutral arginine, neutral N-terminal leucine and neutral C-terminal isoleucine were calculated using the same method as mentioned for fatty acids. The crystal structure of Lg and palmitic acid (PDB 1B0O) was used to construct coordinate files for $(Lg+FA)^{-7}$ ions by removing or adding $–CH_2$ groups. [33] The atom parameter files of $(Lg+FA)^{-7}$ ions were generated using Amber 12SB and GAFF (Generalized Amber Force Field) force fields for the protein and fatty acids, respectively. [34,35]

The NAMD 2.9 simulation package [36] was used to carry out all of MD simulations reported in this chapter. Due to the enormous number of possible charge configurations for (Lg+FA) in charge state -7, the electrostatic potential was first calculated for all the

charge configurations of $(Lg+PA)^{-7}$. Then, 1% of the most stable charge configurations (~8880) were selected to perform 5000 steps conjugate gradient energy minimization. Finally, they were ranked based on the sum of electrostatic and van der Waals energies and the most stable charge configuration was chosen when the following residues are deprotonated: Asp^{11}, Asp^{28}, Asp^{85}, Glu^{112}, Glu^{127}, Glu^{131} and C-terminus-Ile^{162} (See Figure 2. 2).

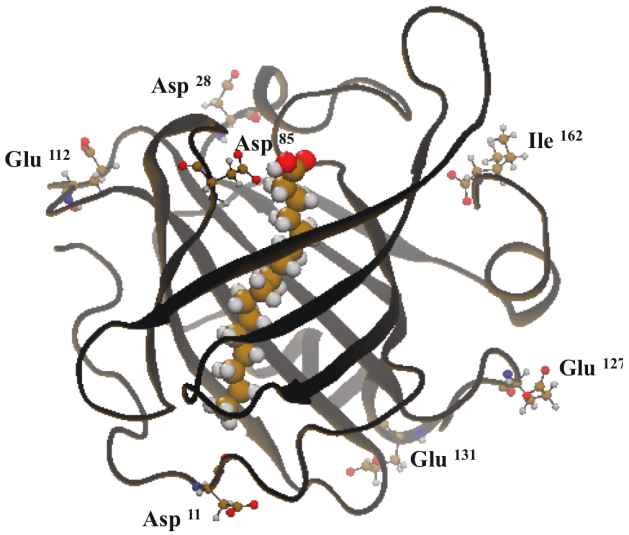


Figure 2. 2. The most stable charge configuration represented for $(Lg+PA)^{-7}$: the charged residues Asp^{11}, Asp^{28}, Asp^{85}, Glu^{112}, Glu^{127}, Glu^{131} and C-terminus-Ile^{162} are shown in ball-stick representation.

Simulations were carried out under vacuum (no pressure) and at fixed temperatures of 299 K to mimic the gas phase and isothermal conditions as in the experiments. First, 20 ns of MD simulation was performed on the most stable charge configuration to obtain the initial structure for PMF calculation. The time step was set to 2 fs and the system was equilibrated for 1 ns. Production dynamics was carried out after this period and data was collected every 2 ps. All of the bonds to hydrogen atoms were constrained using the SHAKE algorithm. [37] No cut-off was used for non-bonded

interactions. Langevin dynamics was applied to fix the temperature at 299 K and damping coefficient (collisional frequency) was set to 1 ps^{-1}. [38]

2. 2. 4. Equilibration and overlap between windows

Equilibration is one of the important conditions that are required to be satisfied to obtain an accurate PMF. In order to ensure the data points used in PMF construction are resulted from equilibrated ensemble and converged simulation, for each window, the r value along the simulation time was examined to be equilibrated by the following strategy: The biased simulations were continued as long as they generate 10 ns equilibrated data points. To fulfill this criterion, first, the initial 5% data points of each window were discarded. Next, the data points were trimmed up to the points that they are within 10% standard deviation from the mean value of the window for at least 10 ns. Since the data were collected every 2 ps, there were at least 5000 equilibrated data points at each window that were used to generate PMF.

Adjacent windows are required to overlap along the reaction coordinate to obtain an accurate PMF using umbrella sampling technique. The windows overlap each other adequately along r for all of the PMFs presented in this chapter. To illustrate, one example for $(Lg+MA)^{-7}$ complex is presented in Figure 2. 3. Distribution of r values along r = 10 – 13.5 Å, clearly indicates that the windows separated by 0.25 Å overlap each other sufficiently.

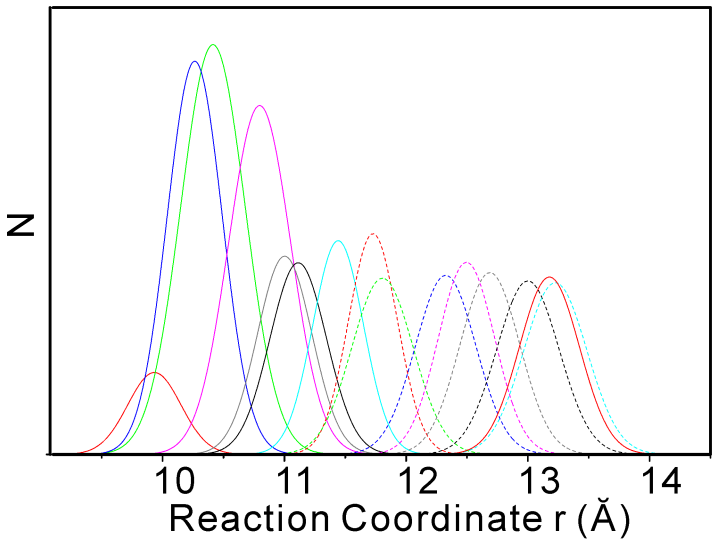


Figure 2. 3. The distribution of r along $\boldsymbol{r}$ = 10 – 13.5 Å for $(Lg+MA)^{-7}$: 15 windows separated by 0.25 Å overlap each other; N refers to the number of occurrence.

The results of all of the simulations (multiple windows) were unbiased and combined using the WHAM. [13] The PMF must be corrected to eliminate the influence of entropy after dissociation, which arises from the escape of the free ligand into an infinite volume. The correction is done using the following equation: [22,39]

$$W(r)^{corrected} = W(r) - 2k_B T \ln\left(\frac{r}{r^*}\right) \quad (2.4)$$

where r^* corresponds to the infinite distance between the ligand and protein. In the absence of the correction, the PMF decreases along the reaction coordinate after dissociation and separated ligand and protein appear more favorable at very long distances. Moreover, the accurate free energy difference between reactant and transition state was determined by including the effect of the PMF well width and shape. The Boltzmann factor of potential of mean force was integrated as a function of reaction coordinate. [25,40,41]

2. 3. Results and discussion

2. 3. 1. Results of the PMF calculations for the most stable charge configuration of $(Lg+FA)^{-7}$ ions

Calculated PMFs for all of the gaseous protein-fatty acid complexes are plotted in Figure 2. 4. The dissociation E_a reflects the required energy to cleave the nonpolar interactions of the fatty acid aliphatic chain within binding cavity and its carboxyl group interactions. In other words, the results of the experiments reveal that there is a late transition state for the dissociation of $(Lg+FA)^{-7}$. As a consequence, the free energy difference between the unbound protein–ligand and bound protein–ligand was considered as the activation free energy. Unbound state corresponds to the plateau of the PMF plot and the minimum of the PMF plot represents the bound state (See Figure 2. 4).

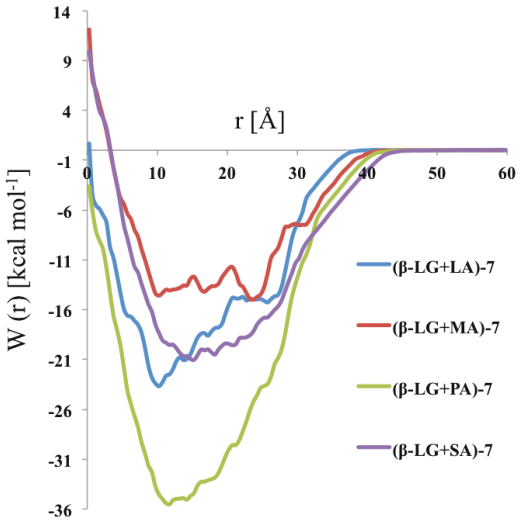


Figure 2. 4. PMF [W(r)] plots of the protein-fatty acid complexes along the reaction coordinate r calculated using single initial structure; All of the complexes are labeled on the plots. Deprotonated residues are Asp^{11}, Asp^{28}, Asp^{85}, Glu^{112}, Glu^{127}, Glu^{131}, C-terminus-Ile^{162}.

The PMF plots give activation free energies ranging from ~16 kcal mol^{-1} to ~36 kcal mol^{-1}, which deviate significantly from the experimental values (ranging from 19.2 to 20.1 kcal mol^{-1}). They also fail to predict the experimental trend of the dissociation activation energies for the fast component. The calculated PMFs do not reflect the trend of dissociation activation free energies of the slow component, either.

To explore whether equilibration time influences the calculated PMFs, the biased simulations of $(Lg+PA)^{-7}$ complex were continued for longer time to collect 20 ns equilibrated data points. Figure 2. 5 shows the change in PMF plot was negligible.

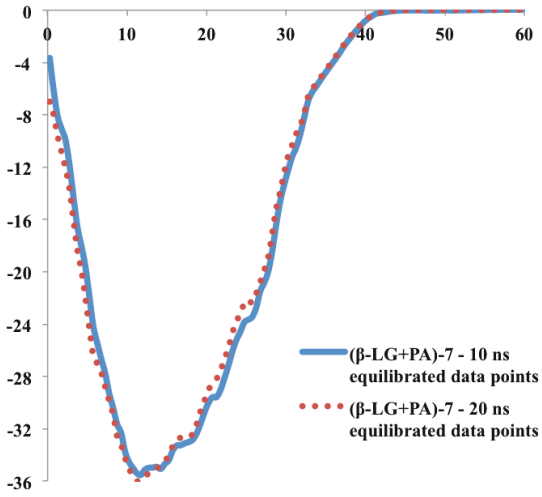


Figure 2. 5. Comparison between PMF plots of $(Lg+PA)^{-7}$ complex: solid blue line) 10 ns equilibrated data points, dotted red line) 20 ns equilibrated data points were used to generate the PMF plots.

One can calculate the classical trajectories for a number of initial states of reactant and average over them to extract the rate constant. Therefore, to modify the results obtained using single structure calculations, two initial structures were used to calculate the PMF along the dissociation coordinate. The second structure was obtained by performing 40 ns MD on the most stable charge configuration. These two initial

structures are slightly different at the beginning of the simulation but they are required to be statistically identical after converged simulations. An illustration of two initial structures for $(Lg+MA)^{-7}$ complex is depicted in Figure 2. 6 as an example. The obtained structures after 20 ns (first) and 40 ns MD (second) are slightly different in terms of the interacting residues and conformation of the ligand and protein. In the first structure, the carboxyl group of myristic acid interacts with Pro^{38} via hydrogen bond, whereas in the second structure Asn^{90} forms two hydrogen bonds with the myristic acid head group (heavy atoms distance $\leq$ 3 Å and donor-hydrogen-acceptor angle $\geq$ 160°). Moreover, RMSD of C_α of the protein residues in the second structure relative to the first structure is 1.31 Å. It indicates that there is no significant difference between two initial configurations in terms of the protein structure. [42]

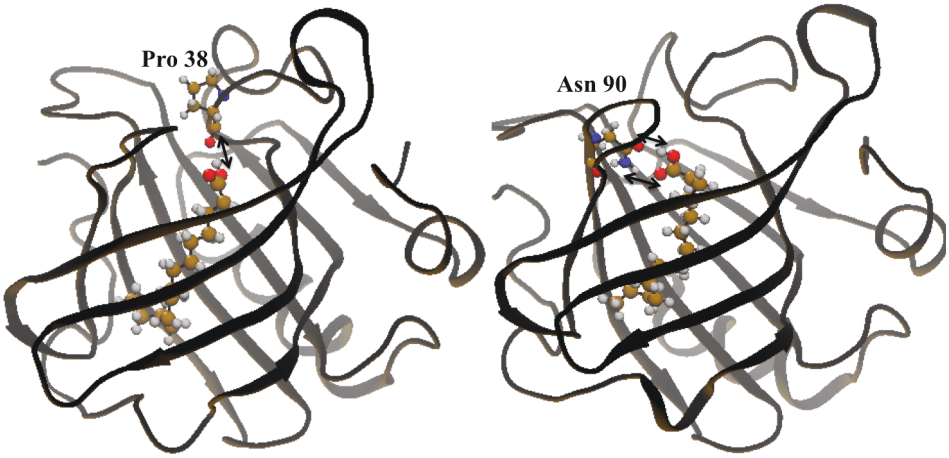


Figure 2. 6. The structure of $(Lg+MA)^{-7}$ obtained after 20ns MD (left) and 40ns MD (right): In the first structure the head group of the MA interacts with Pro^{38} via H-bond MA-OH•••O=C(Pro^{38})) (shown by an arrow) whereas in the second structure the head group of MA forms H-bonds with Asn^{90} (MA-OH•••O=C(Asn^{90}-side chain), MA-C=O•••H_2N(Asn^{90}-side chain)).

In order to modify the results, two parallel simulations on two initial structures must be equilibrated and converged i.e. each simulation must find the most important region of configuration space (equilibrate) and also make a remarkable number of visits to the relevant region (converge). [43] Firstly, the mean value of each window was calculated for each structure. Secondly, the mean values were swapped and the simulations were considered converged if there were 10 ns of data points (at least 5000 points) within 10% standard deviation from the swapped mean value. There were a few windows at shorter distances r = 0 – 30 Å that did not reach equilibration even after continuing them for 10 μs. Higher force constants ranging from 12.0 to 22.0 kcal $mol^{-1}Å^{-2}$ were used to bias these windows around their center and obtain equilibrated data points. WHAM was utilized to unbias and combine the data points of both structures and the calculated PMFs are plotted in Figure 2. 7.

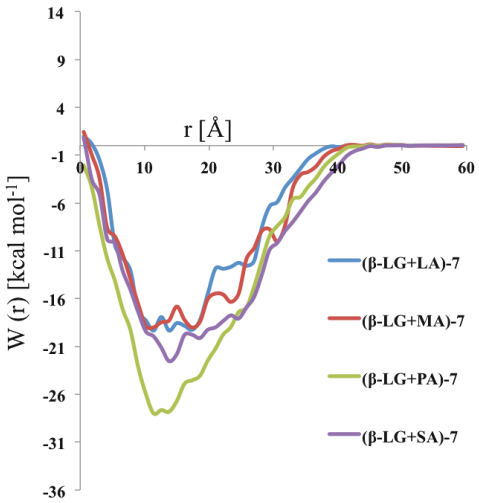


Figure 2. 7. PMF [W(r)] plots of the protein-fatty acid complexes along the reaction coordinate r calculated using two initial structure; Deprotonated residues are Asp^{11}, Asp^{28}, Asp^{85}, Glu^{112}, Glu^{127}, Glu^{131}, C-terminus-Ile^{162}.

Generally, averaging over the trajectories of two initial structures improves the agreement between calculated and experimental activation free energies. The PMF plot of $(Lg+PA)^{-7}$ complex gives the highest activation free energy (28.7 kcal mol^{-1}) and it does not fall into the observed trend for the dissociation activation free energies. Investigation of initial structures for PMF calculations reveals palmitic acid is the only ligand, which interacts with the charged residue (Asp^{85}) of the protein (See Figure 2. 8).

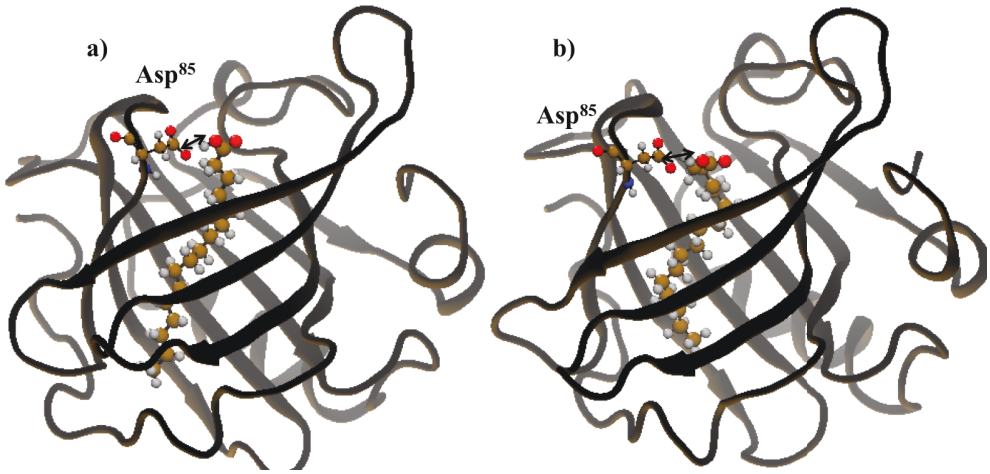


Figure 2. 8. The structure of $(Lg+PA)^{-7}$ obtained after 20 ns MD (left) and 40 ns MD (right): The head group of PA forms hydrogen bond (shown by arrows) with the charged Asp^{85}. Deprotonated residues are Asp^{11}, Asp^{28}, Asp^{85}, Glu^{112}, Glu^{127}, Glu^{131}, C-terminus-Ile^{162}.

The plot of experimental dissociation activation energies of the fast component versus the size of the aliphatic chain (number of $–CH_2$ groups) is provided in Figure 2. 9. The interception of the line gives an estimation of the hydrogen bond contribution to the dissociation activation energy, which results from the head group interactions with the protein residues. Deduction of the energetic contribution of $–CH_3$ group (1.29 kcal mol^{-1}) from intercept yields the contribution of the head group hydrogen bond interactions, which is 3.6 $kcal.mol^{-1}$. The classical hydrogen bonds in the biological molecules are

presumed to be of intermediate strength and in the range of 5-6 kcal mol^{-1}. The hydrogen bonds involving the charged moiety (e.g. OH•••O^-) are stronger and they are associated with the higher energies > 10 kcal mol^{-1}. [44] Thus, the head group of the fatty acids do not form hydrogen bonds with the protein charged residues in the fast component. In conclusion, the theoretical results are consistent with the experimental data in terms of the absence of hydrogen bonds involving the charged residues in the fast component. Furthermore, it clarifies that the dissociation activation free energy of $(Lg+PA)^{-7}$ complex does not fall into the experimental trend due to presence of the interactions with the charged protein residue.

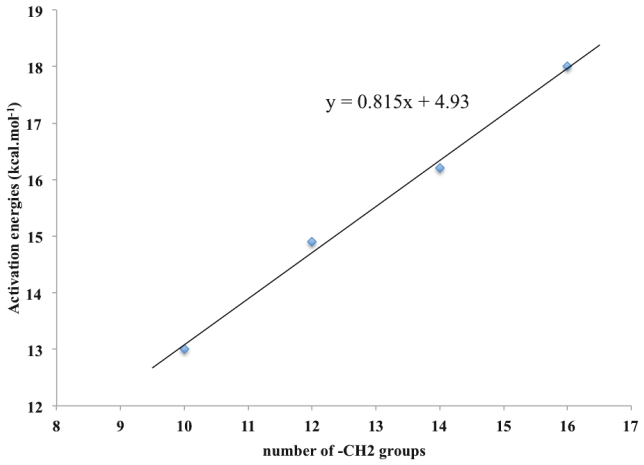


Figure 2. 9. The plot of dissociation activation energies versus the size of aliphatic chain length for the fast component of $(Lg+FA)^{-7}$: deduction of energetic contribution of $–CH_3$ group from intercept ~4.9 represents the contribution of fatty acid head group interactions in the dissociation activation energies.

To make the comparison between calculated quantity $\Delta A^{\ddagger}_{calc}$ and the experimental measurements easier, the TST was used to compute the activation free energies from the experimental dissociation rate constants. The results are presented in Table 2. 1. $A^{\ddagger exp}_{fast}$

denotes the experimental activation free energies for the fast component. $\Delta A_{calc}^{\ddagger single}$ and $\Delta A_{calc}^{\ddagger two}$ refer to the calculated activation free energies using single initial structure and two initial structures, respectively.

Table 2. 1. The experimental and calculated activation free energies. Computation was performed at T = 299 K. The experimental data obtained at T = 302, 298, 298 and 303 K for $(Lg+FA)^{-7}$; FA = LA, MA, PA and SA, respectively. The experimental activation free energies were calculated from the measured rate constants using TST. Activation free energies are reported in kcal mol^{-1}.

Complex	$A_{fast}^{\ddagger exp}$	$\Delta A_{calc}^{\ddagger single}$	$\Delta A_{calc}^{\ddagger two}$
$(Lg+LA)^{-7}$	19.2	23.8	20.5
$(Lg+MA)^{-7}$	19.5	16.3	20.2
$(Lg+PA)^{-7}$	19.7	36.2	28.7
$(Lg+SA)^{-7}$	20.1	22.1	23.1

The data presented in Table 2. 1 indicates averaging over trajectories of two initial structures improves the results. Predicted activation free energies are within ~1.0 kcal mol^{-1} of the actual values for the $(Lg+LA)^{-7}$ and $(Lg+MA)^{-7}$ ions, which is normally considered as a good quantitative agreement. [45,46] For instance, Roux and coworkers calculated binding free energies of nonpolar aromatic ligands to L99A mutant of T4 lysozyme using free energy perturbation method. They claimed that there is an excellent agreement with the experimental values for small ligands such as benzene, toluene, and ethylbenzene whereby the calculated free energies were within 1.1 kcal mol^{-1} of the experimental values. [46] The activation free energy of $(Lg+SA)^{-7}$ complex was predicted as the highest activation free energy as measured by experiments (excluding $(Lg+PA)^{-7}$

complex from the colclusion). Also, it is worthwhile to mention that all of the predicted activation free energies are larger than the experimental values.

Figure 2. 7 shows that PMF plots of $(Lg+FA)^{-7}$ complexes; FA = LA, MA and SA are more similar in terms of the potential well depth and width compared with $(Lg+PA)^{-7}$ complex. The wide potential well indicates that the protein and ligand are bound over a range of r values. r denotes the distance between the COM of the protein and the COM of the fatty acid head group. The experimental and theoretical data indicates that the minimum of the potential well corresponds to states where the aliphatic chain is buried in the binding cavity and the head group forms hydrogen bonds. The result of MD simulations showed that the carboxyl group of the fatty acid interacts with variety of the residues around the opening of the cavity (See Figure 2. 4). All together, the head group of the fatty acid is able to form the hydrogen bonds over a range of r values due to the presence of the flexible residues at the opening of the cavity. NMR studies on the solvated (Lg+PA) complex demonstrated that there are various conformations for the carboxyl end of the fatty acid. [47,48] These findings are indicative of the similarity of solvated and desolvated protein-fatty acid complexes.

In general, the wide potential well and PMF fluctuations may suggest that there is a conformational change accompanying the ligand dissociation process. In consequence, incomplete sampling of the conformational degrees of freedom in $(Lg+MA)^{-7}$ and $(Lg+LA)^{-7}$ complexes (two smallest fatty acids) causes the discrepancies in the predicted dissociation activation free energies. It is worthwhile to mention that the PMF roughness affects the evaluation of activation free energies as one must integrate the Boltzmann factor of potential of mean force as a function of reaction coordinate in order to obtain a

good approximation of activation free energies. The deficiency in sampling becomes more important in kinetic study of a system because the rate constant has an exponential relation with the activation free energy.

Calculated dissociation activation free energies were used to compute the dissociation rate constants via transition state theory. The experimental dissociation rate constants for the fast k_{fast}^{exp} and slow k_{slow}^{exp} components and calculated dissociation rate constants extracted from the PMFs involving single structure k_{single}^{TST} and two structures k_{two}^{TST} are reported in Table 2. 2. The results indicate that none of the obtained trends from the single structure or two-structure PMF calculations are consistent with the slow component. Calculated dissociation rate constant of $(Lg+PA)^{-7}$ complex is extremely low and the reason was discussed. Generally, there is a better agreement between calculated dissociation rate constants and relative dissociation rate constants for the fast components. For instance, theoretical data gives the lowest dissociation rate constant for stearic acid.

Table 2. 2. The experimental and calculated rate constants for $(Lg+FA)^{-7}$ complexes at T = 299 K. The experimental data obtained at T = 302, 298, 298 and 303 K for $(Lg+FA)^{-7}$; FA = LA, MA, PA and SA, respectively.

Complex	k_{slow}^{exp} **(s^{-1})**	k_{fast}^{exp} **(s^{-1})**	k_{single}^{TST} **(s^{-1})**	k_{two}^{TST} **(s^{-1})**
$(Lg+LA)^{-7}$	9.99×10^{-4}	8.90×10^{-2}	2.82×10^{-5}	6.25×10^{-3}
$(Lg+MA)^{-7}$	1.36×10^{-3}	3.57×10^{-2}	7.17	1.07×10^{-2}
$(Lg+PA)^{-7}$	8.13×10^{-4}	2.55×10^{-2}	2.39×10^{-14}	6.49×10^{-9}
$(Lg+SA)^{-7}$	1.60×10^{-3}	2.00×10^{-2}	4.87×10^{-4}	8.59×10^{-5}

A small difference in calculated dissociation activation free energies of the $(Lg+LA)^{-7}$ and $(Lg+MA)^{-7}$ complexes (~ 0.3 kcal mol^{-1}) results in approximately 1.7 times difference in their dissociation rate constants because there is an exponential relation between activation free energy and reaction rate constant. Therefore, even though the calculated activation free energies are quantitatively so close to the experimental values, the present level of accuracy is not able to capture 2.5 times difference in their dissociation rate constants. The experimental dissociation rate constant of lauric acid is almost 2.5 times higher than dissociation rate constant of myristic acid. In conclusion, the evaluation of the accuracy of the computations could be done more reliably if the activation free energies were calculated at multiple temperatures. Then, dissociation activation energies could be compared for different ligands, as it was preliminary goal of the present study. Nevertheless, multiple temperature calculations were proved to be computationally expensive and the calculations were carried out at single temperature.

Improvement of the computed rate constants using two initial structures indicates that there could be a probable problem arising from the deficiency in sampling. Including the second structure in PMF calculations enhances the sampling efficiency and modifies the results. If the reaction coordinate is simply considered as the distance between the protein binding cavity and the fatty acid, two global minima on the potential energy surface of $(Lg+FA)^{-7}$ along the reaction coordinate are expected, corresponding to the slow and fast components. It is evident from the PMF plots (See Figure 2. 4 and Figure 2. 7) that the potential energy surfaces of $(Lg+FA)^{-7}$ complexes are not smooth along the reaction coordinate and apparently several local minima are present along the reaction coordinate. The presence of large number of degrees of freedom in the system justifies

the PMF roughness. The sampling of the entire configurational space becomes more challenging in the presence of various local minima if the states are kinetically distinct. It also rationalizes the improvement of the results by including the second structure, which is slightly different from the first structure but belongs to the same ensemble based on the RMSD values of C_α atoms.

It is inferred from the results that the computed free energies depend on the initial structures of the protein-ligand complexes. This becomes more evident when there is a significant conformational change in protein or ligand upon dissociation reaction, which is clear from the experimental activation entropies. The experimental activation entropies are negative, ranging from -22 to -9 cal mol^{-1} K^{-1} (LA = -22 ± 2, MA = -17 ± 1, PA = -14 ± 1, SA = -9 ± 2). The ligand dissociation is expected to be entropically favorable process. Moreover, the activation entropy becomes less negative with increasing the aliphatic chain length. Thus, it was concluded that the conformational entropy loss of protein in the transition state results in negative activation entropies. An extensive umbrella sampling with overlapped windows along the reaction coordinate was used in this work to calculate the PMFs. Moreover, the biased simulations were continued over 10 μs to collect the equilibrated and converged data points. However, the association of the conformational change with ligand dissociation process creates some meta-stable states, which are not accessible during the affordable simulation time. Thus, the kinetic study of the system becomes difficult due to insufficient sampling of the configurational space.

The issue of sampling has been reported in previous studies, which are mostly performed on the solvated systems and using alchemical transformation methods. [43,49,50]

To avoid the sampling problem in alchemical transformation methods, first, the free energy of restrained ligand in protein binding site is calculated. Next, the ligand is annihilated (its interactions are switched off) and the associated free energy due to its interactions is computed in the presence of some translational restraints. Finally, the ligand is released into standard volume and the free energy corresponding to its interactions with solvent is calculated while the protein and released ligand are restrained in their bound orientation. It is not possible to follow this strategy in the present study by applying restraining potentials because one must know or be able to predict all of relevant slow degrees of freedom involved in dissociation process. On top of that, there must be a few of these degrees of freedom to trace their sampling purposefully. [50] All of the studied ligands consist of a carboxyl group and a saturated aliphatic chain and in consequence, there are a large number of degrees of freedom. On the other hand, the number of potential ligand orientations grows rapidly when the ligand internal degrees of freedom are also relevant for the flexible ligands. Furthermore, it is known that the aliphatic chain of the fatty acid interacts with considerable number of residues lined in the binding cavity (Val^{41}, Val^{43}, Leu^{46}, Leu^{54}, Ile^{56}, Leu^{58}, Ile^{71}, Ile^{84}, Val^{92}, Val^{94}, Leu^{103}, Phe^{105}, Leu^{122}). Moreover, similar to the experimental observations on solvated (Lg+PA), the result of MD simulations for the first and second structures demonstrate that carboxyl end of the fatty acid interacts with variety of residues located on the flexible loops. [47,48,51,52] In conclusion, it is computationally prohibitive to improve the sampling efficiency using extra restraining potentials due to large number of involved moieties and umbrella sampling may prove impractical.

As it was mentioned earlier, palmitic acid interacts with the charged residue and gives the highest dissociation activation free energy. Another charge configuration was studied to circumvent this problem and explore the influence of charge configuration on calculated activation free energies.

2. 3. 2. Results of the PMF calculations for the second studied charge configuration of $(Lg+FA)^{-7}$ ions

The new charge configuration was selected such that there is no charged residue on the opening of the binding cavity. It eliminates the possibility of the interaction of the ligand with the charged amino acids. Deprotonated residues of $(Lg+FA)^{-7}$ in the new charge configuration are as following: Asp^{11}, Glu^{45}, Glu^{51}, Glu^{74}, Glu^{108}, Glu^{114} and C-terminus-Ile^{162} (See Figure 2. 10).

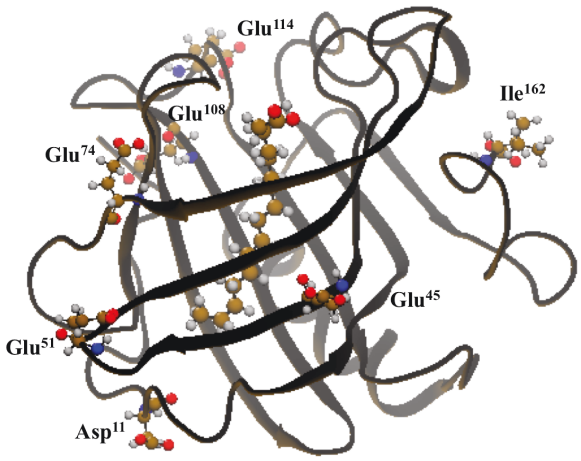


Figure 2. 10. The second studied charge configuration represented for $(Lg+PA)^{-7}$, the charged residues are shown in ball-stick representation.

The PMF plots of the $(Lg+FA)^{-7}$ complexes are presented in Figure 2. 11. The results indicate that the relative activation free energies obtained by PMF calculations

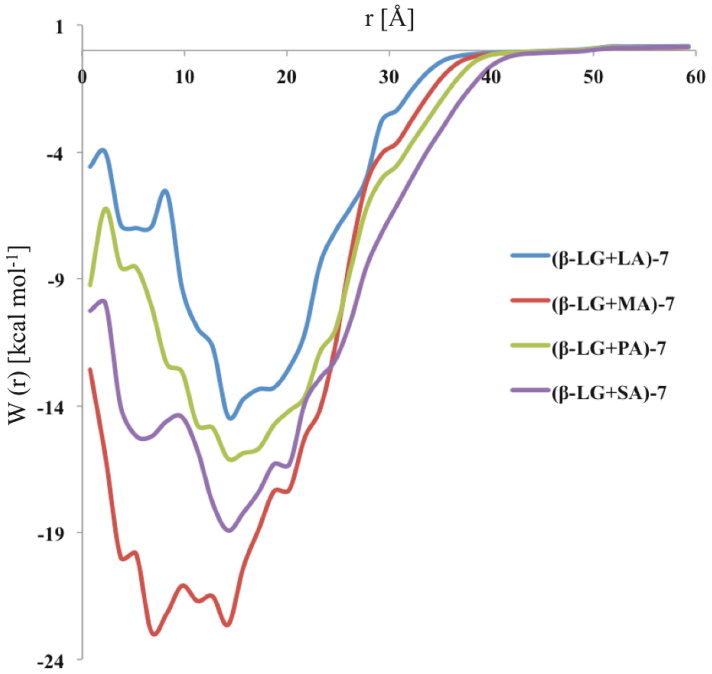


Figure 2. 11. PMF [W(r)] plots of the protein-fatty acid complexes along the reaction coordinate r; All of the complexes are labeled on the plot. Deprotonated residues are Asp^{11}, Glu^{45}, Glu^{51}, Glu^{74}, Glu^{108}, Glu^{114} and C-terminus-Ile^{162}.

reproduce the experimental trend observed for the fast components of $(Lg+FA)^{-7}$ complexes, except for $(Lg+MA)^{-7}$. Consistent with the experimental data, $(Lg+LA)^{-7}$ complex possesses the lowest activation free energy followed by $(Lg+PA)^{-7}$ and $(Lg+SA)^{-7}$ complexes. This is expected from the correlation between dissociation rate constants of the fast components and the size of fatty acid aliphatic chain. Moreover, the PMF plots of $(Lg+LA)^{-7}$, $(Lg+PA)^{-7}$ and $(Lg+SA)^{-7}$ complexes are similar (Figure 2. 11) whereby the minima of the PMF plots correspond to $r \approx 14.25$ Å.

The experimental dissociation rate constants for the fast k_{fast}^{exp} and slow k_{slow}^{exp} components and calculated dissociation rate constants k_{two}^{TST} are reported in Table 2. 3. The experimental activation free energies for the fast component $\Delta A_{fast}^{\ddagger}$ and calculated $\Delta A_{calc}^{\ddagger two}$ activation energies are also included in Table 2. 3.

Table 2. 3. The experimental (k_{slow}^{exp} for the slow and k_{fast}^{exp} for the fast component) and calculated k_{two}^{TST} rate constants for $(Lg+FA)^{-7}$ complexes are presented in three first columns. The theoretical rate constants were computed at T = 299 K. The experimental data were obtained at T = 302, 298, 298 and 303 K for $(Lg+FA)^{-7}$; FA = LA, MA, PA and SA, respectively. The last two columns present the experimental $\Delta A_{fast}^{\ddagger}$ and calculated $\Delta A_{calc}^{\ddagger}$ activation free energies, respectively. TST was used to calculate the experimental activation free energies from the dissociation rate constants of the fast component.

Complex	k_{slow}^{exp} **(s^{-1})**	k_{fast}^{exp} **(s^{-1})**	k_{two}^{TST} **(s^{-1})**	$\Delta A_{fast}^{\ddagger exp}$ **(kcal mol^{-1})**	$\Delta A_{two}^{\ddagger calc}$ **(kcal mol^{-1})**
$(Lg+LA)^{-7}$	9.99×10^{-4}	8.90×10^{-2}	4.03×10^{1}	19.2	15.3
$(Lg+MA)^{-7}$	1.36×10^{-3}	3.57×10^{-2}	4.42×10^{-5}	19.5	23.5
$(Lg+PA)^{-7}$	8.13×10^{-4}	2.55×10^{-2}	1.60	19.7	17.3
$(Lg+SA)^{-7}$	1.60×10^{-3}	2.00×10^{-2}	2.50×10^{-2}	20.1	19.7

There is an agreement between the trend of calculated dissociation rate constants and relative dissociation rate constants of the fast components, except for $(Lg+MA)^{-7}$. Strikingly, the calculated rate constant for dissociation of stearic acid from Lg protein (2.50×10^{-2} s^{-1}) is in close agreement with the experimental value (2.00×10^{-2} s^{-1}). Dissociation rate constant of $(Lg+MA)^{-7}$ complex is predicted as the lowest rate constant and does not fall into the observed trend for dissociation rate constants. Moreover, the minimum of PMF plot of $(Lg+MA)^{-7}$ complex (Figure 2. 11) is located at $r \approx 7$ Å. However, the second minimum energy point corresponds to $r \approx 14.25$ Å, as was observed for the other protein-fatty acid complexes. In contrast to what was observed for $(Lg+PA)^{-7}$ complex in the most stable charge configuration (head group interaction with the charged residue), no significant difference was identified between $(Lg+MA)^{-7}$ complex and other protein-fatty acid complexes from the analysis of the trajectories computed

along the reaction coordinate. The only difference, which may explain deviation of $(Lg+MA)^{-7}$ complex from the expected trend, is related to the interactions involving myristic acid head group. The head group of myristic acid forms a hydrogen bond with Glu^{120} residue (–OH/Glu^{120} (side chain)) when it is deeply buried in the binding cavity ($r \leq 7$ Å). Meanwhile, the hydrogen bond interactions are absent in the other $(Lg+FA)^{-7}$ complexes at shorter distances. To illustrate the difference between $(Lg+MA)^{-7}$ and other $(Lg+FA)^{-7}$ complexes, the structures obtained after performing 60 ns biased MD in the presence of restraining potentials centered at $r = 0.25$ Å are presented in Figure 2. 12. The hydrogen bond between the head group of myristic acid and side chain of Glu^{120} residue exist for more than 40% of the calculated trajectory. The criteria used to identify hydrogen bonds were: heavy atoms distance ≤ 3 Å and donor-hydrogen-acceptor angle ≥ 160°.

Furthermore, the structures of $(Lg+MA)^{-7}$ and $(Lg+PA)^{-7}$ complexes are compared in Figure 2. 13 in the presence of the restraining potential centered at $r = 7$ Å. It is one example to demonstrate the evolution of fatty acids head group interactions in $(Lg+FA)^{-7}$ complexes when the center of restraining potential changes along the reaction coordinate to sample the longer distances. There is a hydrogen bond between the head group of myristic acid and Glu^{120} (Glu^{120} (side chain)/O=C–) for approximately 20% of the trajectory. Thus, the myristic acid interacts with the same amino acid residue even when the center of restraining potential has incremented from 0.25 to 7 Å. In the presence of the same restraining potential, palmitic acid interacts with Asn^{90} (Asn^{90} (side chain)/ O=C–) and Lys^{69} (–OH/Lys^{69}) residues via hydrogen bonds and these bonds exist for approximately 23% and 12% of the trajectory, respectively.

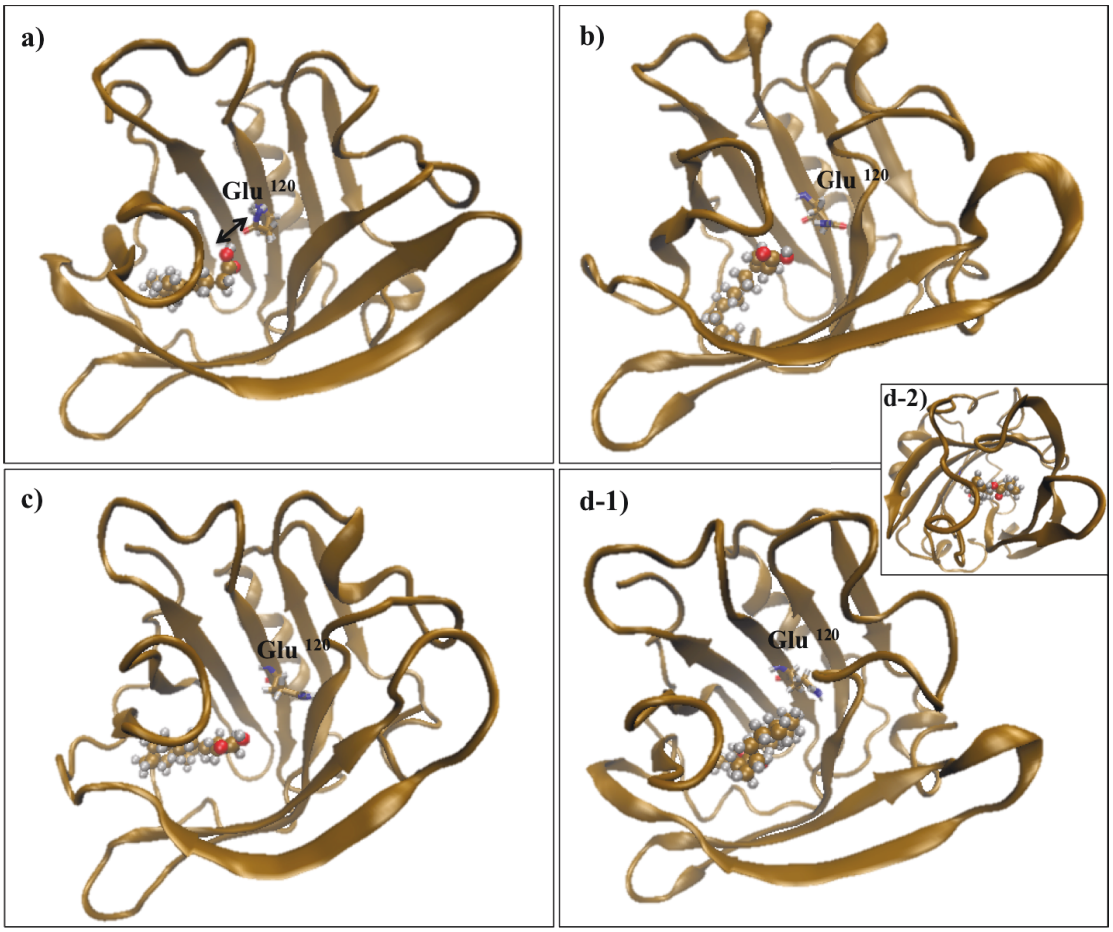


Figure 2. 12. The top view of a) $(Lg +MA)^{-7}$, b) $(Lg +LA)^{-7}$, c) $(Lg +PA)^{-7}$, d-1) $(Lg +SA)^{-7}$ complexes obtained after 60 ns biased MD in the presence of restraining potential centered at r = 0.25 Å. d-2) shows the bottom view of $(Lg+SA)^{-7}$ complex. The fatty acids and Glu^{120} are shown in ball-stick and bond representation, respectively. Only myristic acid head group (a) forms the hydrogen bond (shown by arrow).

It is worth mentioning that all of the structures presented here obtained from the first initial structure of the $(Lg+FA)^{-7}$ complexes, which obtained after performing 20 ns MD simulation.

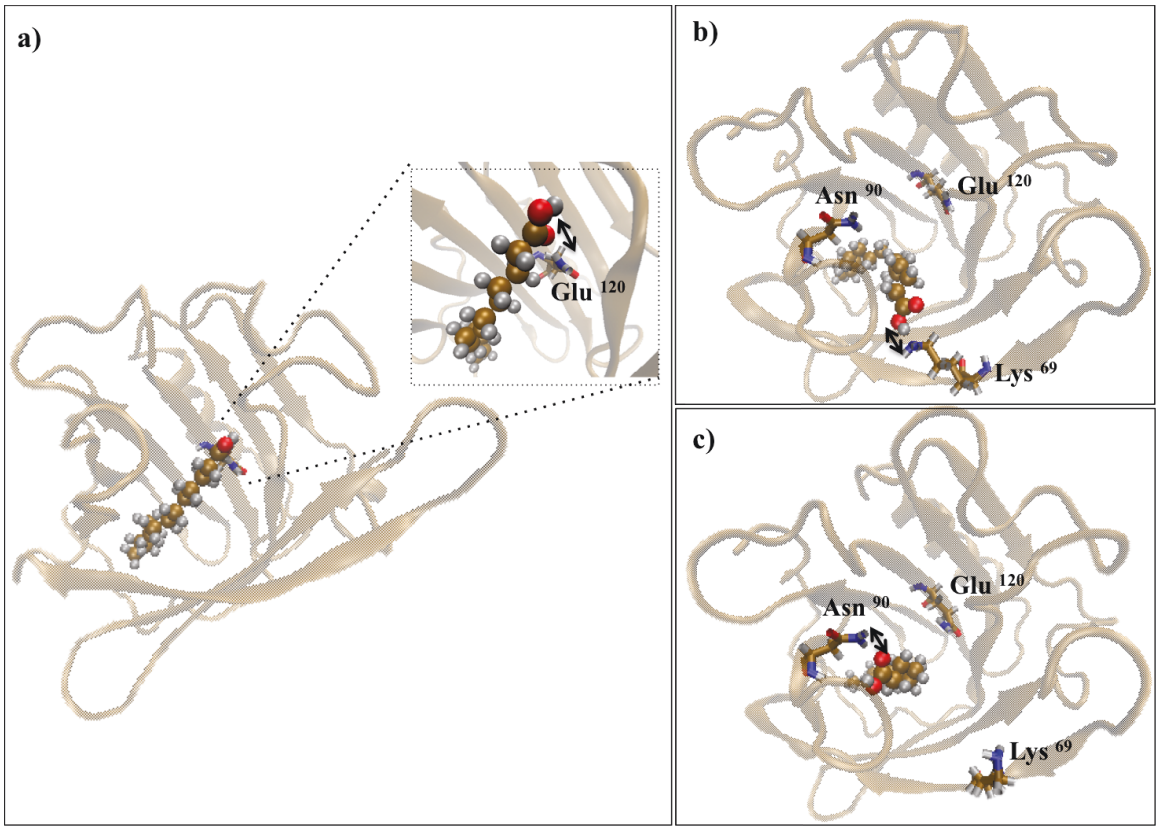


Figure 2. 13. The top view of a) $(Lg+MA)^{-7}$, b and c) $(Lg+PA)^{-7}$ complexes obtained after 60 ns biased MD in the presence of restraining potential centered at r = 7.0 Å. The head group of myristic acid forms H-bond with Glu^{120} whereas the head group of palmitic acid interacts with Asn^{90} and Lys^{69} via hydrogen bonds. The fatty acids and amino acid residues are shown in ball-stick and bond representation, respectively.

The results suggest that $(Lg+MA)^{-7}$ complex is trapped in a kinetically distinct region of the configurational space normally referred to metastable state and thus the other states are not easily accessible regarding to the simulation time. The relative dissociation rate constants of the other protein-fatty acid complexes are consistent with the experimental trend and also their PMF plots are similar. Together, it can be concluded that the experimentally determined trend of dissociation rate constants for $(Lg+FA)^{-7}$ complexes are reproducible by computations if all of the complexes are sampled from

kinetically similar states. However, it also implies that there is a deficiency in sampling of the configurational space since the entire configurational space is not accessible during the simulation.

2. 3. 3. Comparison of two studied charge configurations

The results of the PMF calculations for two different charge configurations of $(Lg+FA)^{-7}$ complexes demonstrate that the deficiency in sampling leads to the prediction of the dissociation rate constants that deviate from the experimental values. However, computations are able to reconcile the experimental trend of dissociation rate constant if the initial structures of the complexes are sampled from similar states. A short simulation time in the range of μs cannot access the entire region of configurational space and thus if the system is trapped in a metastable state, the predicted activation energy differs significantly from the values obtained for the complexes that are not trapped.

Free energies of the protein-fatty acid complexes when the fatty acid is deeply buried in the binding cavity ($r \approx 0$ Å) are consistent for each charge configuration: The free energies of all protein-fatty acid complexes are around zero for the most stable charge configuration (See Figure 2. 7). All of the protein-fatty acid complexes possess favorable free energies for the second studied charge configuration (See Figure 2. 11). In addition, all the activation free energies obtained using the most stable charge configuration were greater than the experimental values (Table 2. 1). However, the activation free energies achieved by the second studied charge configuration are lower than the experimental values (Table 2. 3). In other words, the values of dissociation rate constants depend on the charge configuration of the protein. These results may suggest

that the experimentally determined values are the averaged dissociation rate constants for the dissociation of the ligand from the protein with variety of charge configuration.

The results of free energy simulations were more consistent with the trend of dissociation rate constants for the fast components. In the fast components, the aliphatic chain of the fatty acid is buried in the binding cavity and contribution of the carboxyl group interactions is ~3.6 kcal mol^{-1}. As stated earlier, it was suggested that hydrogen bond interactions of carboxyl group contribute predominantly in stabilizing the slow component of ion complex. It was shown for the $(Lg+MA)^{-7}$ complex in the second studied charge configuration, the differences in the fatty acid head group interactions results in being trapped in a kinetically distinct region of the configurational space. It proposes that the slow components of $(Lg+FA)^{-7}$ complexes are possibly the extreme examples of stabilization by hydrogen bond interactions of the fatty acid head group.

The PMF plots obtained for $(Lg+FA)^{-7}$ complexes in two different charge configurations (Figure 2. 7 and Figure 2. 11) indicate the r value corresponding to the unbound protein-ligand is smaller for the fatty acid with shorter aliphatic chain and it becomes greater with increasing the size of fatty acid. Figure 2. 14 gives representative structures of the transition states of the dissociation process for $(Lg+LA)^{-7}$ and $(Lg+SA)^{-7}$ complexes. One must note that there is no single structure for the dissociation of fatty acid from the binding cavity of Lg. Indeed, there is a distribution of the structures corresponding to the probable transitions states of the dissociation process and some representative structures are shown in Figure 2. 14. Transition states of the dissociation process are suggested based on two assumptions: (i) Interactions cut-off where protein

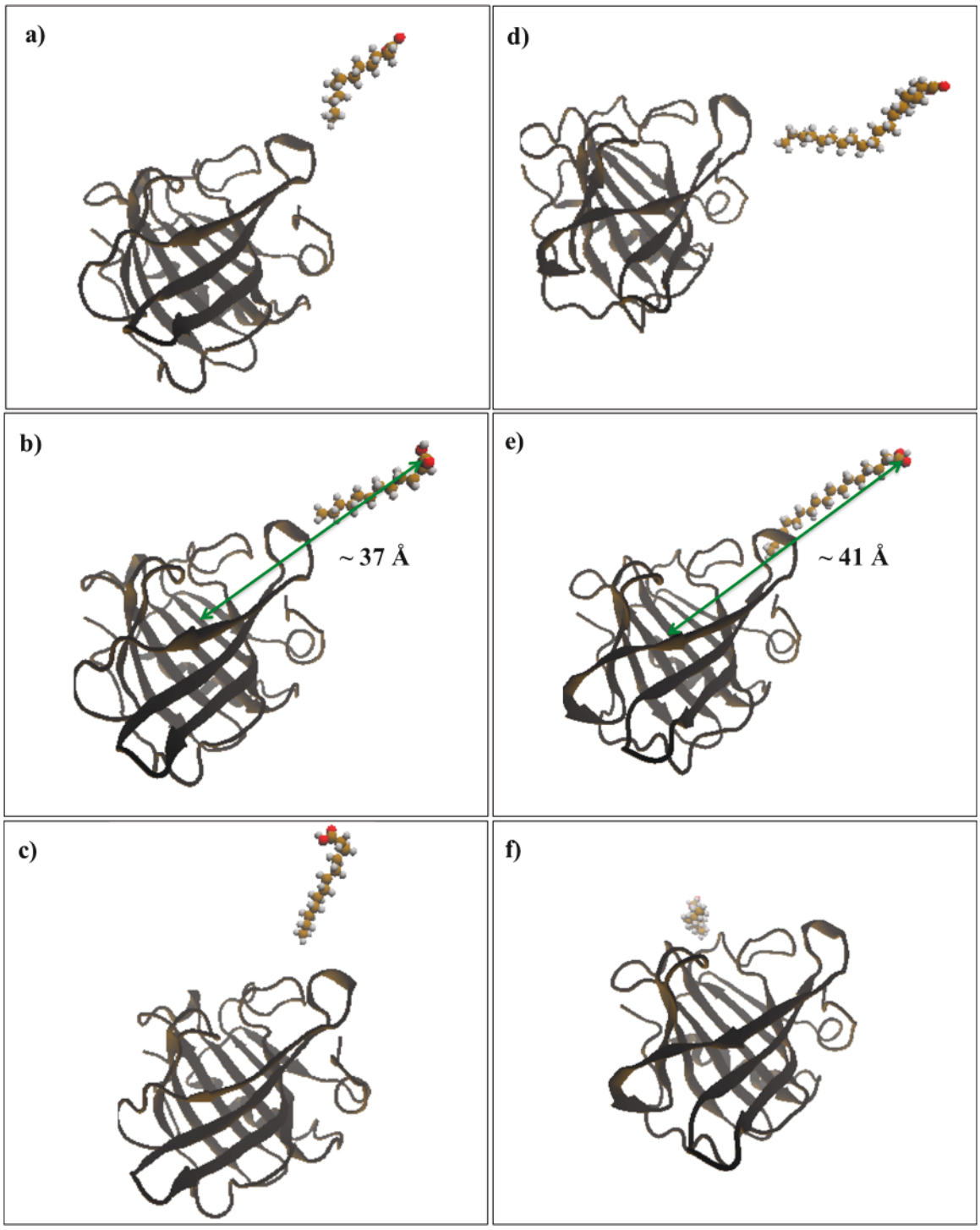


Figure 2. 14. Some representative structures of suggested transition state for dissociation of $(Lg+FA)^{-7}$ complexes; a-c) transition states of $(Lg+LA)^{-7}$ and d-f) transition states of $(Lg+SA)^{-7}$ complexes. The structures are generated from the trajectory files of the second studied charge configuration.

and ligand are not interacting any longer is the same for all of protein-fatty acid complexes. (ii) There is a correlation between the length of the aliphatic chain of the fatty acid and dissociation distance. Dissociation distance is defined as the distance between the COM of the fatty acid head group and the COM of the protein binding cavity. Thus, assuming that protein and ligand are not able to interact with each other beyond some distance cut-off, the bonds of carboxyl end of the fatty acid must break first and aliphatic chain of the fatty acid leaves the binding cavity in the next step.

2. 4. Conclusions

In summary, calculated PMFs for the most stable charge configuration were able to predict the activation free energies of the $(Lg+LA)^{-7}$ and $(Lg+MA)^{-7}$ complexes within less than 1.0 kcal mol^{-1} deviation from the experimental values, which is an excellent agreement with the experimental measurement. Moreover, consistent with the experiment, the dissociation rate constant of the $(Lg+SA)^{-7}$ was calculated as the lowest dissociation rate constant. Furthermore, the results of the free energy calculations illustrated the absence of interaction between the carboxyl groups of the fatty acids with the charged amino acid residues in the fast components.

Calculated PMFs for the second studied charge configuration reproduced the trend of dissociation rate constants for the $(Lg+FA)^{-7}$ complexes, except for $(Lg+MA)^{-7}$ complex. The trajectory analysis indicated that $(Lg+MA)^{-7}$ complex is trapped in a metastable state due to the interaction of the fatty acid head group with Glu^{120} residue.

The PMF plots reported in the present work were obtained using an extensive umbrella sampling and rigorous criteria for the equilibration, convergence and overlap between the windows. The agreement between computed and experimental

thermodynamic quantities was satisfactory for the most stable charge configuration. Calculated activation free energies for the second studied charge configuration reproduced the trend of activation free energies for the fast components. Insufficient sampling of conformational degrees of freedom and being trapped in metastable states are the issues must overcome in a relatively short simulation time. The results of the PMF calculations can improve further using some restraining potentials but prior knowledge of slow degrees of freedom is required in order to apply suitable restraining potentials. The results of the present study suggest that the carboxyl group of the fatty acids or flexible residues on the opening of the protein cavity are the candidates for applying restraining potentials.

The proposed sequence of the ligand exit is consistent with the measured dissociation activation energies. Dissociation activation energies reflect the required energy to cleave the nonpolar interactions in addition to the interactions of carboxyl group of the fatty acid with the protein residues. Therefore, there is a late transition state for the dissociation of $(Lg+FA)^{-7}$ complexes where the ligand exit occurs upon cleaving the carboxyl group interactions followed by the exit of the aliphatic chain from the binding cavity.

References

1. Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Feature Article Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2010**, *31*, 1569-1582.

2. Freiteg, S.; Chu, V.; Penzotti, J.; Klumb, L.; To, R.; Hyre, D.; Trong, I.; Lybrand, T.; Stenkamp, R.; Stayton, P. A structural snapshot of an intermediate on the streptavidin–biotin dissociation pathway. *Proc. Natl. Acad. Sci.* **1999**, *96*, 8384-8389.

3. Buckle, A.; Schreiber, G.; Fersht, A. Protein-protein recognition: Crystal structural analysis of barnase barstar complex at 2.0-angstrom resolution. *Biochemistry* **1994**, *33*, 8878-8889.

4. Radic, Z.; Kirchhoff, P.; Quinn, D.; McCammon, J.; Taylor, P. Electrostatic influence on the kinetics of ligand binding to acetylcholinesterase - Distinctions between active center ligands and fasciculin. *J. Biol. Chem.* **1997**, *272*, 23265-23277.

5. Sines, J.; Allison, S.; McCammon, J. Point-charge distribution and electrostatic steering in enzyme substrate encounter- Brownian dynamics of modified copper-zinc superoxide dismutases. *Biochemistry* **1990**, *29*, 9403-9412.

6. Case, D.; Karplus, M. Dynamics of ligand-binding to heme-proteins. *J. Mol. Biol.* **1979**, *132*, 343-368.

7. Hyre, D.; Amon, L.; Penzotti, J.; Le Trong, I.; Stenkamp, R.; Lybrand, T.; Stayton, P. Early mechanistic events in biotin dissociation from streptavidin. *Nat. Struct. Mol. Biol.* **2002**, *9*, 582-585.

8. Grubmuller, H.; Heymann, B.; Tavan, P. Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **1996**, *271*, 997-999.

9. Genest, D.; Garnier, N.; Arrault, A.; Marot, C.; Morin-Allory, L.; Genest, M. Ligand-escape pathways from the ligand-binding domain of PPARγ receptor as probed by molecular dynamics simulations. *Eur. Biophys. J.* **2008**, *37*, 369-379.

10. Gonzalez, A.; Perez-Acle, T.; Pardo, L.; Deupi, X. Molecular Basis of Ligand Dissociation in b-Adrenergic Receptors. *Plos One* **2011**, *6*, e23815.

11. Liu, L.; Bagal, D.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Hydrophobic Protein-Ligand Interactions Preserved in the Gas Phase. *J. Am. Chem. Soc.* **2009**, *131*, 15980-+.

12. Torrie, G.; Valleau, J. Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187-199.

13. Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. the Method. *J. Comput. Chem.* **1992**, *13*, 1011-1021.

14. Pelzer, H.; Wigner, E. The speed constansts of the exchange reactions. *Z. Phys. Chem. B-Chem. Elem. Aufbau. Mater.* **1932**, *15*, 445-471.

15. Laidler, K.; King, M. The Development of Transition-State Theory. *J. Phys. Chem.* **1983**, *87*, 2657-2664.

16. Evans, A.; Polanyi, M. Some Aspects of the Chemistry of Macromolecules - Held at the Institution-Of-Mechanical-Engineers, London, on April 4th, 1946 .1. Polymerisation of Isobutene by Friedel-Crafts Catalysts - Discussion. *J. Chem. Soc.* **1947**, 252-257.

17. Roux, B. The Calculation of the Potential of Mean Force using Computer-Simulations. *Comput. Phys. Commun.* **1995**, *91*, 275-282.

18. Truhlar, D.; Gao, J.; Alhambra, C.; Garcia-Viloca, M.; Corchado, J.; Sanchez, M.; Villa, J. The incorporation of quantum effects in enzyme kinetics modeling. *Acc. Chem. Res.* **2002**, *35*, 341-349.

19. Zhang, H.; Tan, T.; Feng, W.; van der Spoel, D. Molecular Recognition in Different Environments: beta-Cyclodextrin Dimer Formation in Organic Solvents. *J. Phys. Chem. B* **2012**, *116*, 12684-12693.

20. Bastug, T.; Chen, P.; Patra, S. M.; Kuyucak, S. Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. *J. Chem. Phys.* **2008**, *128*, 155104.

21. Bolhuis, P.; Dellago, C.; Chandler, D. Reaction coordinates of biomolecular isomerization. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5877-5882.

22. Schenter, G.; Garrett, B.; Truhlar, D. Generalized transition state theory in terms of the potential of mean force. *J. Chem. Phys.* **2003**, *119*, 5828-5833.

23. Liu, L.; Michelsen, K.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Energetics of Lipid Binding in a Hydrophobic Protein Cavity. *J. Am. Chem. Soc.* **2012**, *134*, 3054-3060.

24. Liu, L.; Bagal, D.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Hydrophobic Protein-Ligand Interactions Preserved in the Gas Phase. *J. Am. Chem. Soc.* **2009**, *131*, 15980-15981.

25. Woo, H.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825-6830.

26. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535-9551.

27. Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91*, 2798-2814.

28. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Chem. informatics* **2012**, *4*, 17.

29. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. RED Server: a web service for deriving RESP and ESP

charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res*. **2011**, *39*, W511-W517.

30. Dupradeau, F.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.ED. tools: advances in RESP and ESP charge derivation and force field library building. *Phys*. *Chem*. *Chem*. *Phys*. **2010**, *12*, 7821-7839.

31. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A. J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09. *Gaussian Inc*. **2009**.

32. Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. A Well-Behaved Electrostatic Potential Based Method using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J*. *Phys*. *Chem*. **1993**, *97*, 10269-10280.

33. Wu, S.; Perez, M.; Puyol, P.; Sawyer, L. Beta-Lactoglobulin Binds Palmitate within its Central Cavity. *J. Biol. Chem.* **1999**, *274*, 170-174.

34. Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

35. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Case, D. A.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. AMBER 12. **2012**.

36. Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781-1802.

37. Ryckaert, J.; Ciccotti, G.; Berendsen, H. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

38. Loncharich, R.; Brooks, B.; Pastor, R. Langevin Dynamics of Peptides - the Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers* **1992**, *32*, 523-535.

39. Truhlar, D.; Garrett, B. Multidimensional transition state theory and the validity of Grote-Hynes theory. *J Phys Chem B* **2000**, *104*, 1069-1072.

40. Luo, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 10399-10404.

41. Jorgensen, W. Interactions between Amides in Solution and the Thermodynamics of Weak Binding. *J. Am. Chem. Soc.* **1989**, *111*, 3770-3771.

42. Maiorov, V.; Crippen, G. Size-Independent Comparison of Protein 3-Dimensional Structures. *Proteins* **1995**, *22*, 273-283.

43. Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006**, *125*, 084902.

44. Weiss, M.; Brandl, M.; Suhnel, J.; Pal, D.; Hilgenfeld, R. More hydrogen bonds for the (structural) biologist. *Trends Biochem. Sci.* **2001**, *26*, 521-523.

45. Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395-2417.

46. Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255-1273.

47. Konuma, T.; Sakurai, K.; Goto, Y. Promiscuous binding of ligands by beta-lactoglobulin involves hydrophobic interactions and plasticity. *J. Mol. Biol.* **2007**, *368*, 209-218.

48. Ragona, L.; Fogolari, F.; Zetta, L.; Perez, D.; Puyol, P.; De Kruif, K.; Lohr, F.; Ruterjans, H.; Molinari, H. Bovine beta-lactoglobulin: Interaction studies with palmitic acid. *Protein Sci.* **2000**, *9*, 1347-1356.

49. Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.***2006**, *2*, 1255-1273.

50. Mobley, D. L.; Chodera, J. D.; Dill, K. A. Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput.* **2007**, *3*, 1231-1235.

51. Kontopidis, G.; Holt, C.; Sawyer, L. Invited Review: beta-lactoglobulin: Binding properties, structure, and function. *J. Dairy Sci.* **2004**, *87*, 785-796.

52. Qin, B.; Bewley, M.; Creamer, L.; Baker, H.; Baker, E.; Jameson, G. Structural basis of the tanford transition of bovine beta-lactoglobulin. *Biochemistry (N. Y.)* **1998**, *37*, 14014-14023.

Chapter 3

Fluorine bonding enhances the energetics of protein-lipid binding in the gas phase

Reproduced with permission from Lan Liu, Nobar Jalili, Alyson Baergen, Simon Ng, Justin Bailey, Ratmir Derda and John S. Klassen, *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 751-757: "Fluorine Bonding Enhances the Energetics of Protein-Lipid Binding in the Gas Phase". The experimental measurements were performed by Lan Liu and Alyson Baergen.

3. 1. Introduction

The incorporation of fluorine substituents into pharmacologically active compounds is a common strategy used to improve their bioavailability, metabolic stability, and distribution.[1-5] Fluorination of drug molecules can also affect their non-covalent interactions with other molecules, notably proteins.[6-10] Understanding the molecular basis of how fluorination influences the affinity and selectivity of protein-ligand interactions is attracting increased attention.[8,11,12]

Theoretical and experimental studies carried out on non-covalent complexes involving small fluorine-containing molecules indicate that fluorine can alter non-covalent binding either directly, through intermolecular interactions, or indirectly, through the modulation of the polarity of other groups that engage in intermolecular interactions.[13-21] At present, the influence of fluorine bonding on the stability of protein-ligand complexes is unclear.

A survey of reported crystal structures for proteins bound to fluorinated ligands revealed a large number of interactions between the carbon-fluorine (C-F) unit and polar and non-polar hydrogens.[22] According to theoretical calculations performed on model systems, such interactions, although relatively weak, are expected to be energetically more significant than similar interactions involving hydrocarbons.[22] However, from a detailed investigation into the binding of alkyl- and fluoroalkyl-substituted benzene sulfonamide ligands to human carbonic anhydrase II (HCA II), it was recently concluded that differences in the binding thermodynamics reflect differences in hydrophobic surface area and not differences in the strength of the intermolecular interactions.[23]

Here, we report on the first quantitative investigation of the intrinsic energetics (free from solvent effects) of fluorine bonding in a protein-ligand complex. Bovine β-lactoglobulin (Lg), which possesses a large and dry (in absence of ligand) hydrophobic cavity, lined with 12 aliphatic residues (Leu^{58}, Val^{41}, Val^{43}, Leu^{46}, Leu^{54}, Ile^{56}, Leu^{58}, Ile^{71}, Leu^{87}, Val^{92} and Leu^{103}) and one aromatic residue (Phe^{105}), [24,25] and its interactions with four fluorinated analogs of stearic acid (XF-SA) containing X = 13, 15, 17 or 21 fluorine atoms served as model systems for this study (See Figure 3. 1). The Arrhenius parameters for the loss of neutral ligand from the gaseous $(Lg+XF\text{-}SA)^{7-}$ ions were measured using Fourier-transform ion cyclotron resonance mass spectrometry and the blackbody infrared radiative dissociation (BIRD) technique.[26,27] The nature of the intermolecular interactions in the gaseous $(Lg+XF\text{-}SA)^{7-}$ ions were investigated using molecular dynamics (MD) simulations. Notably, the results of this study provide unambiguous evidence that fluorination of ligand alkyl chains enhances the energetics of their intermolecular interactions with proteins.

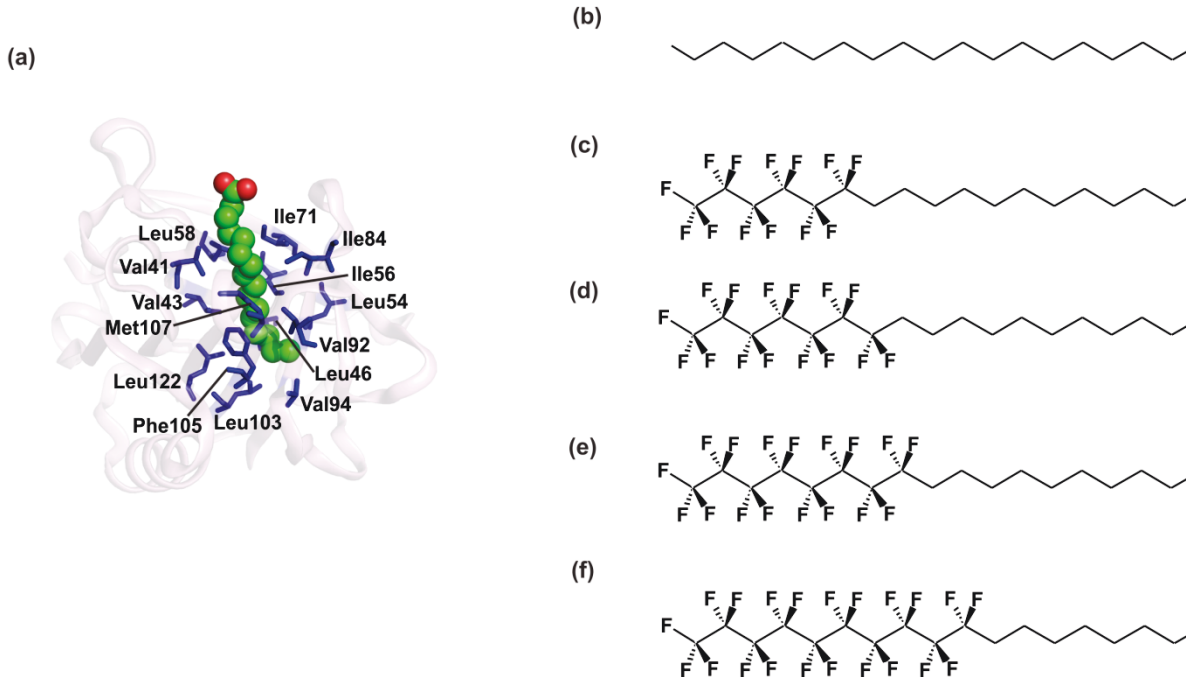


Figure 3. 1. (a) Crystal structure of Lg complex with palmitic acid (PDB 1B0O) and residues within the hydrophobic cavity of β-LG. Structures of (b) SA, (c) 13F-SA, (d) 15F-SA, (e) 17F-SA and (f) 21F-SA.

3. 2. Experimental Methods

3. 2. 1. Synthesis of XF-SA ligands

13,13,14,14,15,15,16,16,17,17,18,18,18-Tridecaflurooctadecanoic acid (13F-SA)

Perfluorohexyl iodide (0.120 mL, 0.55 mmol), 11-dodecenoic acid (0.100 g, 0.50 mmol) and AIBN (0.003 g, 0.02 mmol) were added to a round-bottom flask. The flask was installed with a condenser and cooled to -78 °C. The flask was evacuated and refilled with nitrogen a total of 3 times. The reaction mixture was heated to 80 °C and stirred under neat condition with nitrogen flow for 10 hours. After cooling and evaporation under reduced pressure, the solid was dissolved in anhydrous THF (10 mL). The solution was cooled to 4 °C, and into it lithium triethylborohydride solution (2 mL, 2.00 mmol)

was added. After stirring for 5 hours at room temperature, the mixture was quenched with acetic acid/methanol solution (1:4) until no more bubbles evolved. The solvent was removed under reduced pressure and the crude mixture was dissolved in ethyl acetate (40 mL). The solution was washed with water (40 mL) and brine (20 mL), and dried with sodium sulfate. After filtration and evaporation, the residue was purified on silica gel (40 g) with a gradient of 0–10% ethyl acetate in hexane using CombiFlash Rf. Removal of solvent and drying in vacuo yielded the desired product as a white solid (0.115 g, 44%).

12,12,13,13,14,14,15,15,16,16,17,17,18,18,18-Pentadecaflurooctadecanoic acid (15F-SA)

The synthetic procedure was similar to that described above, except that perfluoroheptyl iodide and 10-undecenoic acid were used as substrates. The product was obtained as a white solid (83 mg, 30%).

11,11,12,12,13,13,14,14,15,15,16,16,17,17,18,18,18 Heptyldecaflurooctadecanoic acid (17F-SA)

The synthetic procedure was similar to that described above, except that perfluorooctyl iodide and 9-decenoic acid were used as the substrates. The product was obtained as a white solid (142 mg, 48%).

9,9,10,10,11,11,12,12,13,13,14,14,15,15,16,16,17,17,18,18,18-Heptyldecaflurooctadecanoic acid (21F-SA)

Perfluorodecyl iodide (0.500 g, 0.77 mmol) was dissolved in a solution of 7 octenoic acid (0.1 mL, 0.65 mmol) and 0.5 mL toluene under argon. The reaction mixture was heated to 80°C and AIBN (15.6 mg, 0.095 mmol) was added in six portions at 30 minute

intervals. After 20 hours, tributyltin hydride (0.21 mL, 0.77 mmol) and AIBN (5.2 mg, 0.032 mmol) were added to the reaction. The solution was stirred for 2 hours at 80°C under argon. After cooling and evaporation under reduced pressure, the solid was dissolved in ether and treated with potassium fluoride to remove tributyltin iodide. The ether solution was filtered through celite enriched with potassium fluoride and evaporated under reduced pressure. The solid was first purified via fluorous solid phase extraction using a 9:1 DMF/H_2O eluent to elute non-fluorous contaminants, followed by elution of the product and excess perfluorodecyl contaminants using a 1:1 DMF/ether eluent. The residue was then purified by flash chromatography using 9:1 DCM/ether to produce a white solid (0.9 g, 67%).

3. 2. 2. Sample preparation

Bovine β-lactoglobulin variant B (Lg, monomer MW 18281 Da) was purchased from Sigma-Aldrich Canada (Oakville, Canada). The Lg was dissolved and exchanged directly into Milli-Q water, using an Amicon microconcentrator with a 10 kDa molecular weight cutoff, followed by filtration using an Amicon ultra centrifugal filter with a molecular weight cutoff of 100 kDa to remove Lg aggregates from solution. The concentration of the Lg solution was determined by lyophilizing a known volume of the filtrate and measuring the mass of the protein. The protein stock solution was stored at -20 °C until needed. Ligand stock solutions were prepared by dissolving a known amount of XF-SA into a known volume of MeOH. For ESI, solutions were prepared from the stock solutions of Lg and XF-SA. The percent (volume) MeOH in the ESI solutions was 2%. Aqueous ammonium hydroxide was added to adjust the pH to 8.5; imidazole (10 mM) was also added in order to minimize the occurrence of in-source dissociation.

3. 2. 3. Mass spectrometry

All experiments were performed on a 9.4 T ApexII FTICR mass spectrometer (Bruker, Billerica, MA) equipped with a nanoflow ESI ion source. Complete details of the instrumental and experimental conditions used for the BIRD measurements, as well as a description of how the kinetic data were analyzed, can be found elsewhere.[28]

3. 2. 4. Computational modeling

The initial structures of 13F-SA and 21F-SA were constructed using Avogadro 1.1.0.[29] The partial charges of the ligands, neutral arginine, neutral N-terminal leucine and neutral C-terminal isoleucine were calculated using the RESP ESP charge Driver Server[30,31] and Gaussian 09 (C.01).[32] Since there are no crystal structures available for the (Lg+XF-SA) complexes, the crystal structure of Lg and palmitic acid complex (PDB 1B0O)[33] was used to produce the initial coordinate file for each of the $(Lg+XF\text{-}SA)^{7-}$ ions. All the coordinate and atom parameter files for $(Lg+XF\text{-}SA)^{7-}$ were constructed using AmberTool 12.[34] MD simulations were performed using the NAMD 2.9 package, the Amber 12SB force field for Lg and the general Amber force filed (GAFF) for the XF-SA ligands.[34-36] The electrostatic potential energy was calculated for all the possible charge configurations. Energy minimization (5000 steps conjugate gradient) was performed on the lowest energy configurations (8880 charge configurations were considered). All the covalent bonds to hydrogen atoms were constrained using the SHAKE algorithm[37] and no cut-off was considered for non-bonded interactions. The sum of the electrostatic potential and Van der Waals energies was considered to determine the most stable charge configurations. Langevin dynamics was implemented using a 1 ps^{-1} damping coefficient at a temperature of 299 K. For all simulations, the time step was set to 2 fs, a 50 ns

equilibration period was followed by 150 ns of production dynamics. The root mean-squared deviation (RMSD) of Cα atoms of the protein was examined to ensure that equilibration was reached prior to production. The distances and angles between possible hydrogen donors and acceptors (fluorine atoms) were measured using the VMD software[38] and the distance and angle distributions were calculated using Origin 9.0.[39] The criteria used to identify hydrogen bonds were: heavy atom distance ≤ 4 Å and donor-hydrogen-acceptor angle ≥110°.

3. 3. Results and discussion

Deprotonated gas-phase ions of the (Lg+XF-SA) complexes were readily detected by ESI-MS performed in negative ion mode on aqueous ammonium acetate (10 mM) solutions containing Lg and one of the XF-SA ligands, at 25 °C and pH 8.5 (See Figure 3. 2). Imidazole (10 mM), which is known to minimize the extent of in-source (gas-phase) dissociation, was also added to each solution.[40,41] At pH > 8, Lg is known to exist predominantly as a monomer and to adopt an "open" structure that allows ligand access to the hydrophobic cavity.[24,25] Time-resolved BIRD measurements were performed on the $(Lg+XF\text{-}SA)^{7-}$ ions at temperatures ranging from 37 to 77°C. As illustrated in Figure 3. 3, BIRD of the $(Lg+X\text{-}FSA)^{7-}$ ions proceeds exclusively by the loss of neutral XF-SA, (equation 3. 1):

$$(Lg+XF\text{-}SA)^{7-} \rightarrow Lg^{7-} + XF\text{-}SA \qquad (3.1)$$

Natural log plots of the normalized abundance (Ab/Ab_{total}) of the $(Lg+XF\text{-}SA)^{7-}$ ions versus reaction time are shown in Figure 3. 4, for each ligand investigated. In all cases,

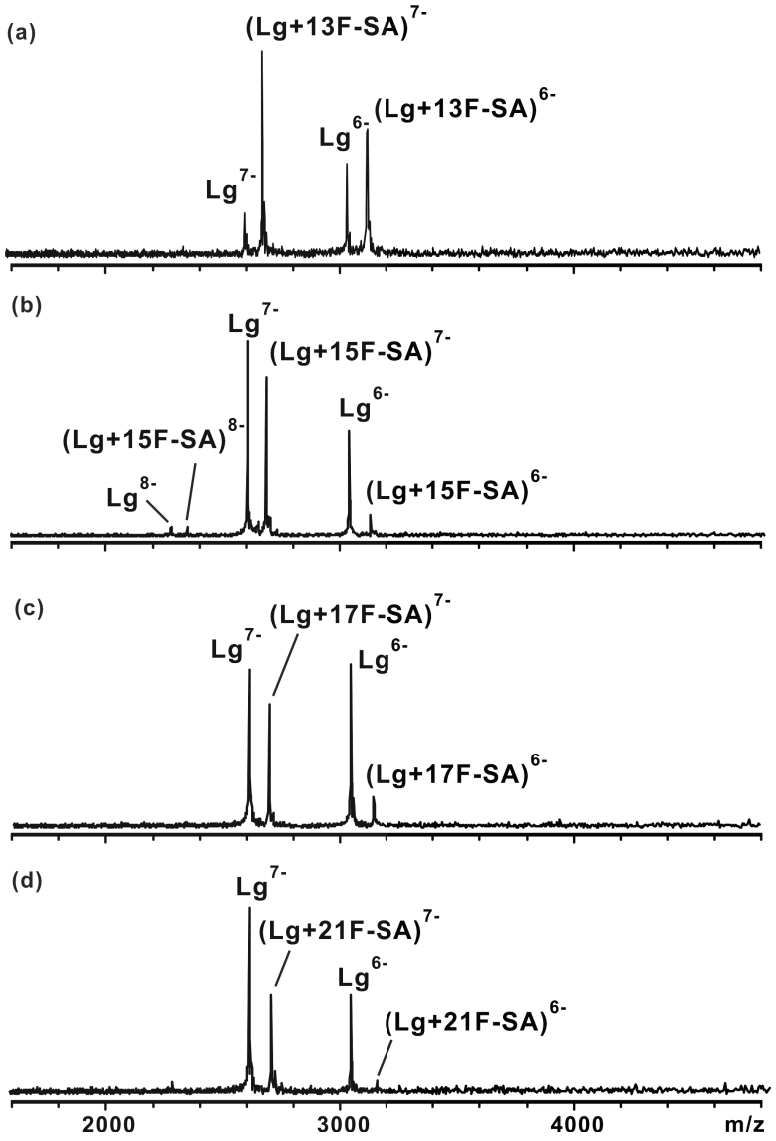


Figure 3. 2. ESI mass spectra acquired for aqueous solutions (pH 8.5, 25 °C) of Lg (15 μM) and (a) 13F-SA, (b) 15F-SA, (c) 17F-SA, and (d) 21F-SA. Each solution contained 10 mM ammonium acetate and 10 mM imidazole.

the kinetic plots exhibit non-linear behavior that can be described by a double exponential function, indicating the presence of two kinetically distinct (non-interconverting) structures. These findings are consistent with those reported recently for the dissociation of the deprotonated ions of Lg complexes with saturated, unsaturated and branched fatty

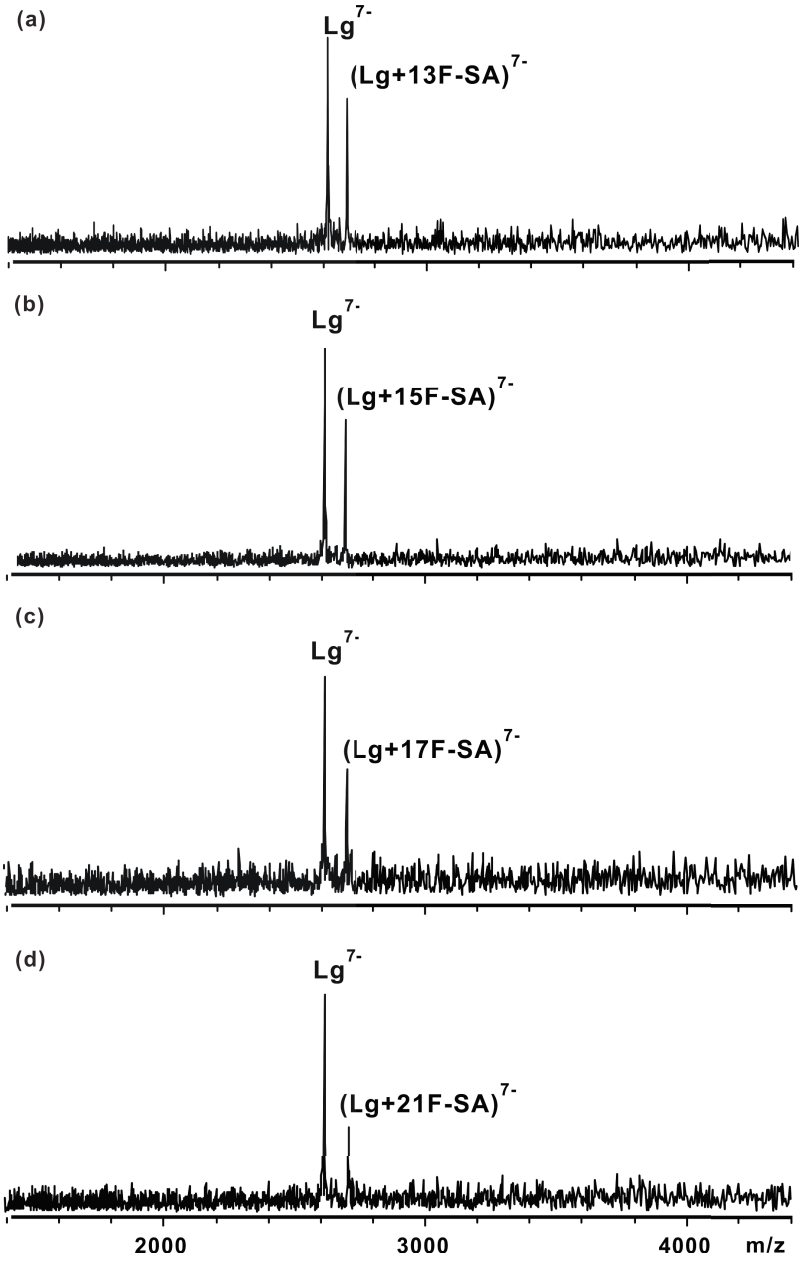


Figure 3. 3. Illustrative BIRD mass spectra measured for (a) (Lg + 13F-SA)$^{7-}$ at a reaction temperature of 61 °C and a reaction time of 30 s; (b) (Lg + 15F-SA)$^{7-}$ at 52 °C and 58 s; (c) (Lg + 17F-SA)$^{7-}$ at 60 °C and 52 s; and (d) (Lg + 21F-SA)$^{7-}$ at 69 °C and 10 s.

acids (FA).[28,42] In these earlier studies, the two non-interconverting structures were designated as the fast and slow (dissociating) components, i.e., $(Lg+FA)_f^{n-}$ and $(Lg+FA)_s^{n-}$

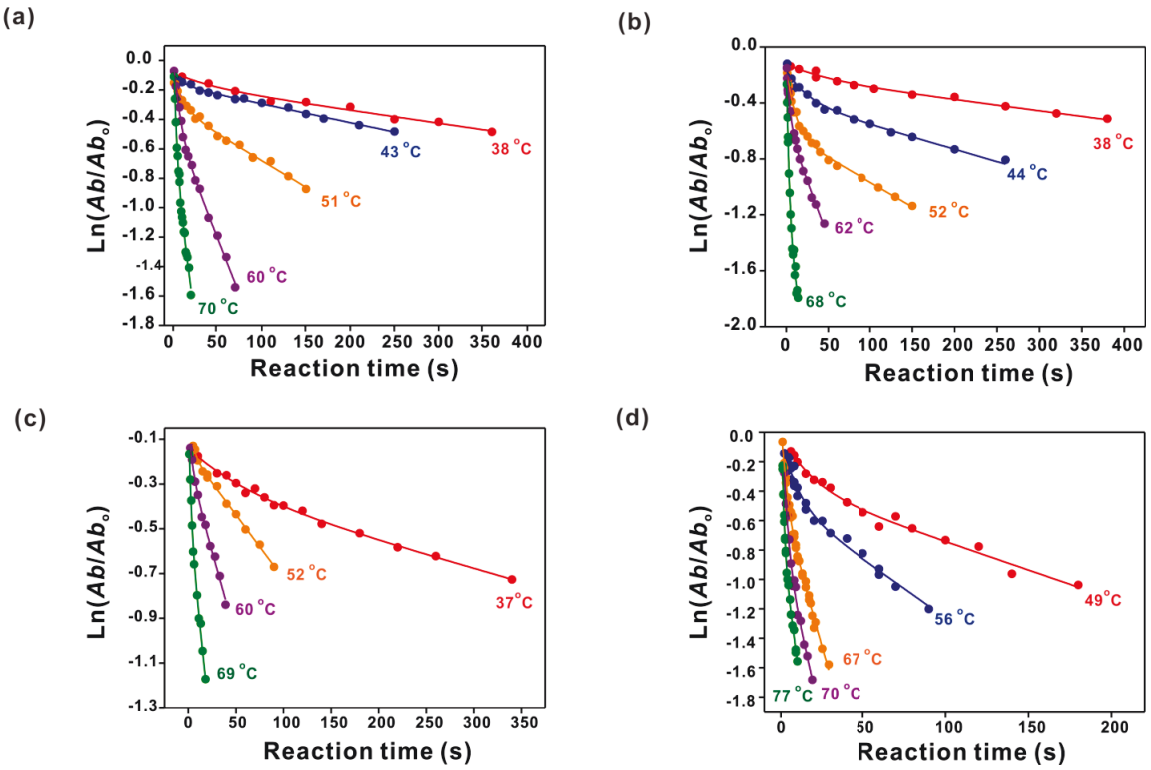


Figure 3. 4. Plots of the natural logarithm of the normalized abundance (Ab/Ab_o) of the (a) $(Lg+13F\text{-}SA)^{7-}$, (b) $(Lg+15F\text{-}SA)^{7-}$, (c) $(Lg+17F\text{-}SA)^{7-}$, and (d) $(Lg+21F\text{-}SA)^{7-}$ ions versus reaction time, at the temperatures indicated.

ions, respectively. According to the results of MD simulations, the acyl chain of the FA remains buried in the hydrophobic cavity in both the fast and slow structures and the main structural difference between the two structures is the position of the flexible EF loop of Lg.[28] In the $(Lg + FA)_f^{7-}$ ions the loop is in an "open" position, such that the $(Lg+FA)_f^{7-}$ is stabilized predominantly by protein-lipid interactions involving the acyl chain of the FA and the hydrophobic residues that line the cavity of Lg.[28] Available experimental data suggest that the fast component resembles the native structure in solution.[28,42] In the $(Lg+FA)_s^{7-}$ ions, the loop is in a "closed" position and H-bonds between the FA carboxyl group and the residues that make up the entrance of Lg cavity also contribute to the stability of the complex.[28] Given the structural similarities of the

FAs considered here and those investigated previously,[28,42] it is reasonable to expect that similar structural differences are responsible for the double exponential kinetics observed for the $(Lg+XF\text{-}SA)^{7-}$ ions.

Arrhenius plots (See Figure 3. 5) were constructed from the rate constants measured for both the $(Lg+XF\text{-}SA)_f^{7-}$ and $(Lg+XF\text{-}SA)_s^{7-}$ ions, i.e., k_f and k_s, respectively, and the corresponding Arrhenius parameters (E_a and A) are listed in Table 3. 1. For comparison purposes, the Arrhenius plots and parameters for the dissociation of the $(Lg+SA)_f^{7-}$ and $(Lg+SA)_s^{7-}$ ions are also included.[28]

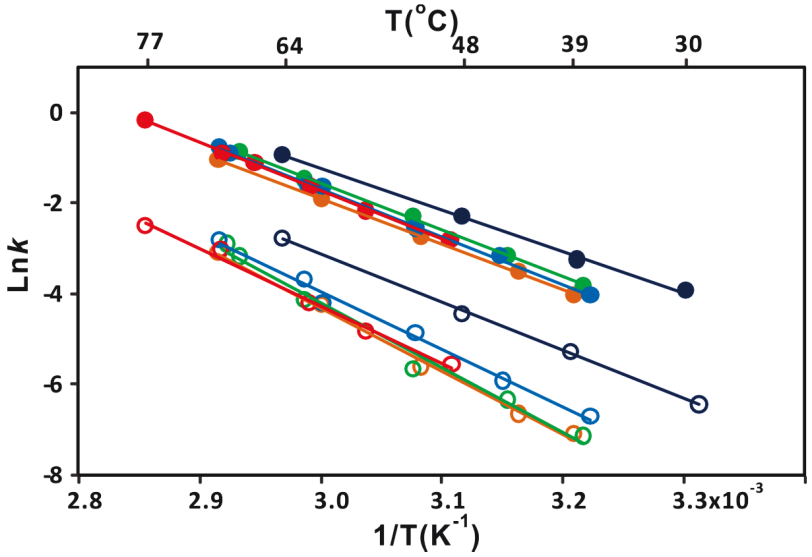


Figure 3. 5. Arrhenius plots for the loss of neutral ligand from the $(Lg+XF\text{-}SA)_f^{7-}$ (solid circles) and $(Lg+XF\text{-}SA)_s^{7-}$ ions (open circles) where X = 0 (●), 13 (●), 15 (●), 17 (●) and 21 (●).

Inspection of the Arrhenius parameters reveals that fluorination of the acyl chain of SA results in a significant increase in the E_a values for both the fast and the slow components. A plot of E_a versus number of fluorine substitutions is shown in Figure 3. 6. In the case of the fast component, E_a increases linearly with the number of fluorines, with each fluorine contributing an additional ~0.15 kcal mol^{-1}. The average contribution of

Table 3. 1. Arrhenius parameters (E_a, A) determined for the loss of neutral ligand from the gaseous, deprotonated $(Lg+XF\text{-}SA)_f^{7-}$ and $(Lg+XF\text{-}SA)_s^{7-}$ ions. The reported errors are one standard deviation. a. Values taken from reference [28].

Ligand	E_a (kcal mol^{-1})	A (s^{-1})
	Fast	
SA	18.0 ± 0.6 [a]	$10^{11.3 \pm 0.4}$ [a]
13F-SA	20.1 ± 0.2	$10^{12.4 \pm 0.1}$
15F-SA	20.5 ± 0.3	$10^{12.8 \pm 0.4}$
17F-SA	20.8 ± 0.4	$10^{13.0 \pm 0.2}$
21F-SA	21.1 ± 0.5	$10^{13.1 \pm 0.3}$
	Slow	
SA	21.5 ± 0.5 [a]	$10^{12.7 \pm 0.3}$ [a]
13F-SA	27.9 ± 1.1	$10^{16.4 \pm 0.7}$
15F-SA	26.3 ± 1.7	$10^{15.4 \pm 2.2}$
17F-SA	24.7 ± 1.1	$10^{14.4 \pm 0.8}$
21F-SA	25.6 ± 1.5	$10^{14.9 \pm 2.1}$

each $>CF_2$ group to E_a is ~1.12 ± 0.01 kcal mol^{-1} for the fast component, which is 0.30 kcal mol^{-1} higher than the value (0.82 ± 0.04 kcal mol^{-1}) reported for $>CH_2$ groups.[28] The linear increase in E_a with number of fluorine substitutions suggests that the Lg cavity presents a relatively homogeneous environment to the $>CF_2$ groups. The average energetic contribution of $-CF_3$ to the E_a of the fast component (estimated as the difference between the measured E_a value and the combined energetic contributions of the $>CF_2$ and $>CH_2$ groups) is 1.85 ± 0.15 kcal mol^{-1}, compared to 1.29 kcal mol^{-1} for the $-CH_3$ group.[42] Taken together, these results provide compelling evidence that Lg bonding to fluoroalkyl chains is energetically more favourable than to the corresponding alkyl chains in the gas phase. No correlation between E_a and the number of fluorines in the XF-SA ligands is evident for the slow component. This finding is consistent with results obtained

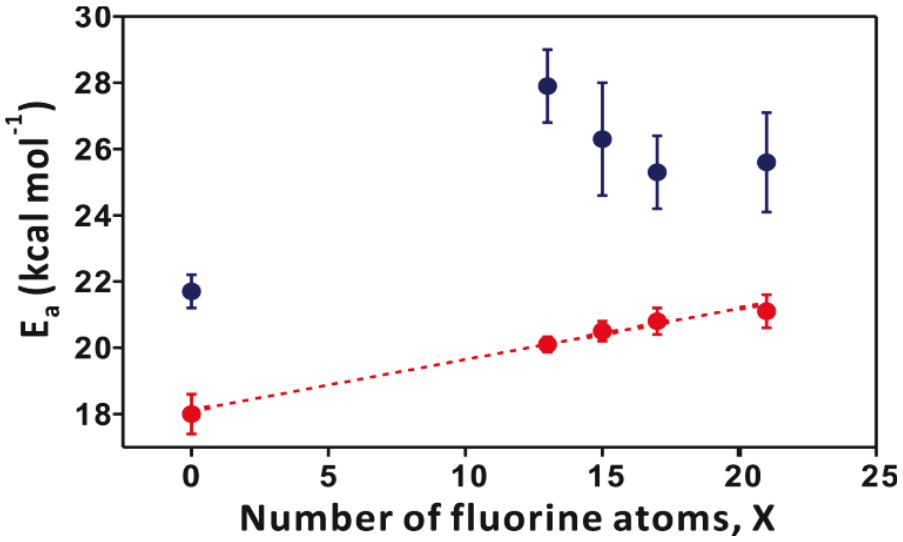


Figure 3. 6. Plot of E_a for dissociation of the fast (●) and slow (●) components of the (Lg + XF-SA)$^{7-}$ions versus X, the number of fluorine substitutions.

previously for the slow components of (Lg+FA)$^{n-}$ ions composed of saturated and unsaturated FAs.[28,42] In these earlier studies it was found that the E_a values (for the slow component) did not correlate with the structure (length of the acyl chain and degree of unsaturation) of the FAs. It was also found that the dissociation E_a values for the slow components were 2 to 12 kcal mol^{-1} larger than those of the fast component.[28,42] As described above and elsewhere,[28,42] the slow component is believed to be stabilized by both protein-lipid interactions (similar to those formed in the fast component) and hydrogen bonding involving the FA carboxyl group. According to the results of MD simulations, the nature of the intermolecular hydrogen bonds is sensitive to the structure of the FA (length of the acyl chain).[28] Consequently, the absence of correlation between E_a and the number of fluorines atoms in the XF-SA ligands suggests that the degree of fluorination may also influence the nature of the hydrogen bonds involved in stabilizing the (Lg+XF-SA)$_s^{7-}$ ions, *vide infra*.

Shang and coworkers have calculated energies for C-F interactions representative of those found in protein-ligand complexes.[11] Comparison of the average energetic contribution that C-F makes to the E_a for the $(Lg+XF\text{-}SA)_f^{7-}$ ions with the calculated energies reveals that the experimental value most closely resembles those found for C-F interactions with polar hydrogens. For example, at the MP2 level of theory and using the 6-311++G(d,p) basis set, an interaction energy of 1.60 kcal mol^{-1} was found.[11] Other neutral interactions suggested to be important in protein-ligand binding, such those involving nonpolar hydrogens or amide carbonyls (so-called orthogonal multipolar interactions) are too weak to account for the experimental results.[11]

To probe the nature of the stabilizing intermolecular interactions in more detail, MD simulations were performed on the $(Lg+13F\text{-}SA)^{7-}$ and $(Lg+21F\text{-}SA)^{7-}$ ions using the Amber 12SB force field for Lg and the general Amber force field for the ligands.[34,36] Analysis of the trajectories for the $(Lg+21F\text{-}SA)^{7-}$ ion reveals that four amino acid residues (Lys^{69}, Asn^{88}, Asn^{90} and Glu^{120}) are involved in fluorine bonding and that approximately half the fluorine atoms (11 out of 21) form interactions (of the F•••H_2N- type) with polar hydrogens associated with the side chains of these four amino acids (See Figures 3. 7a and 3. 7c). For the $(Lg+13F\text{-}SA)^{7-}$ ion, three amino acid residues (Leu^{93}, Val^{94}, and Phe^{82}) are involved in fluorine bonding and 5 out of 13 fluorine atoms form interactions (of the F•••HN- type) with polar hydrogens associated with backbone amide groups of these three amino acids (See Figures 3. 7b and 3. 7d). However, these interactions are transient (a given fluorine bond exists for only 10-50% of the trajectory calculated for $(Lg+21F\text{-}SA)^{7-}$ and 5-45% of the trajectory calculated for $(Lg+13F\text{-}SA)^{7-}$)

and individual residues interact, in an alternating fashion, with multiple fluorine atoms (See Figures 3. 8 and 3. 9)

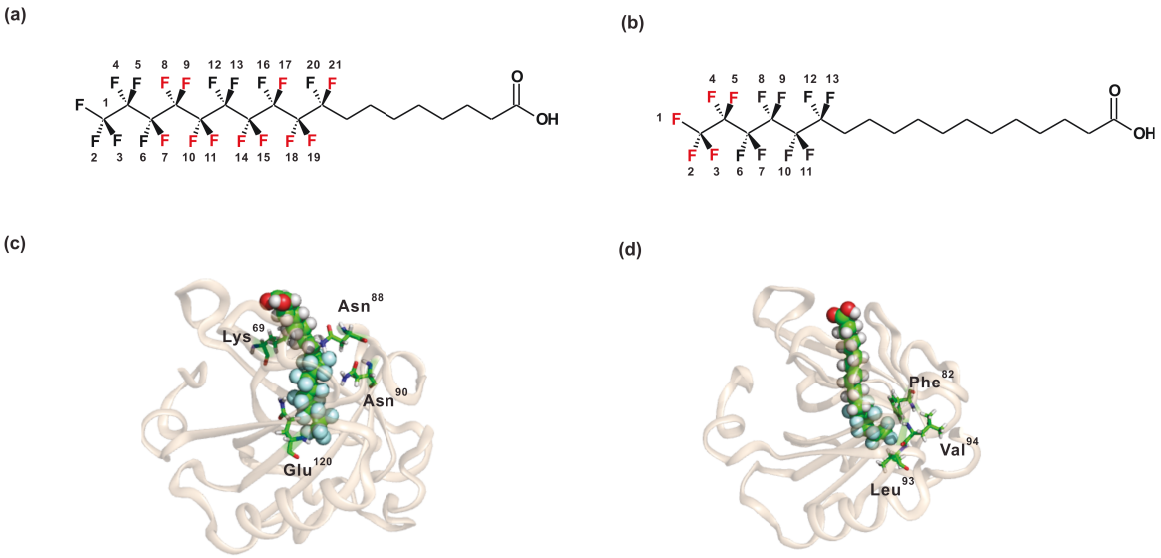


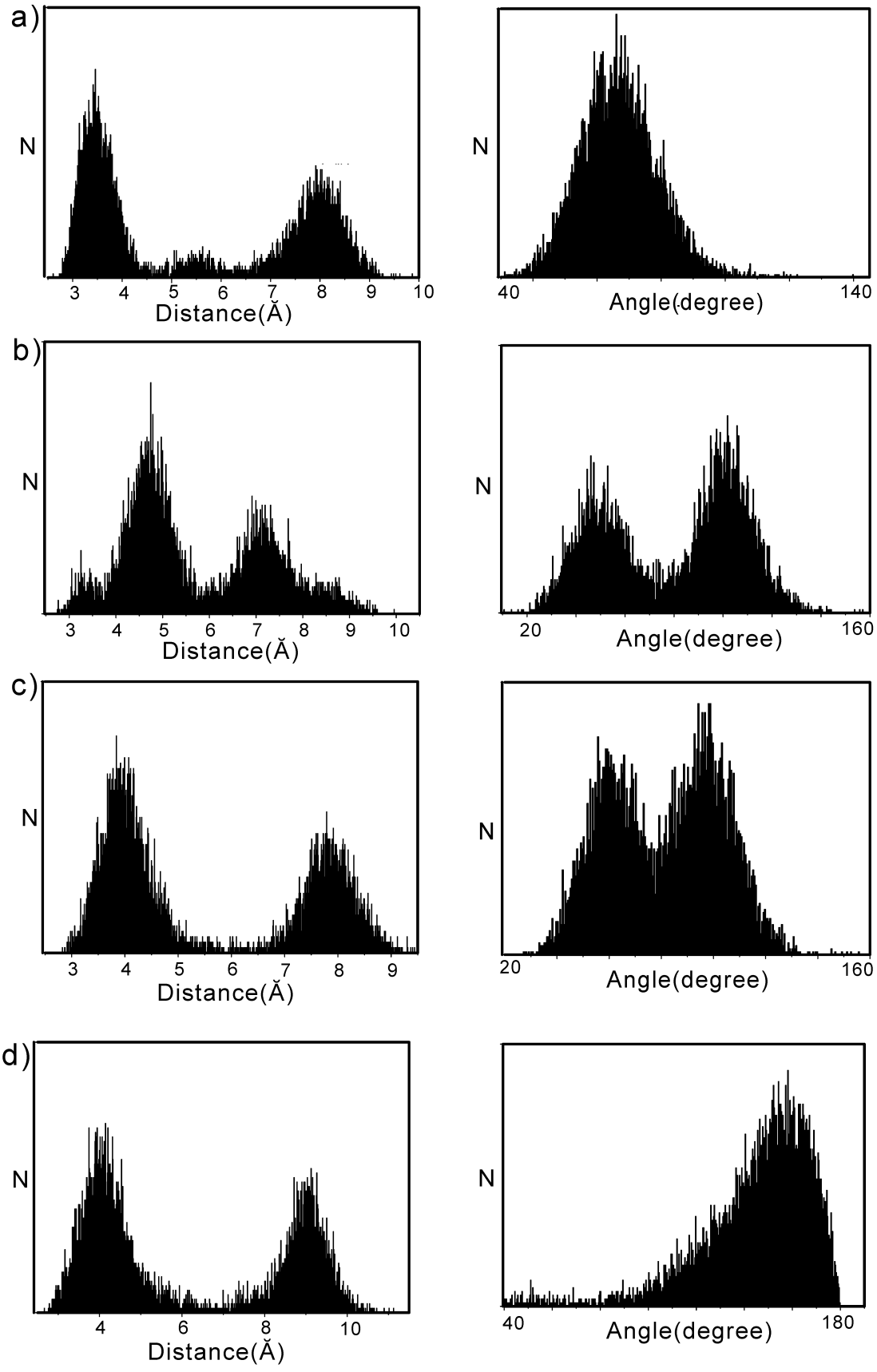
Figure 3. 7. Labelled in red are the fluorine atoms in (a) 21F-SA and (b) 13F-SA that are involved in intermolecular interactions with polar hydrogen in Lg, as determined from MD simulations performed on the $(Lg+21F\text{-}SA)^{7-}$ ion and $(Lg+13F\text{-}SA)^{7-}$, respectively. The Lg residues that were deprotonated for this simulation were Asp^{11}, Asp^{28}, Asp^{85}, Asp^{129}, Glu^{112}, Glu^{127}, Glu^{131} and C-terminus Ile^{162}. (c) Representative structure of the $(Lg+ 21F\text{-}SA)^{7-}$ ion, obtained from the MD simulations, showing the amino acid residues involved in fluorine bonding. The corresponding distance and angle distributions for these interactions are shown in Figure 3. 8. (d) Representative structure of the $(Lg+13F\text{-}SA)^{7-}$ ion, obtained from the MD simulations, showing the amino acid residues involved in fluorine bonding. The corresponding distance and angle distributions for these interactions are shown in Figure 3. 9.

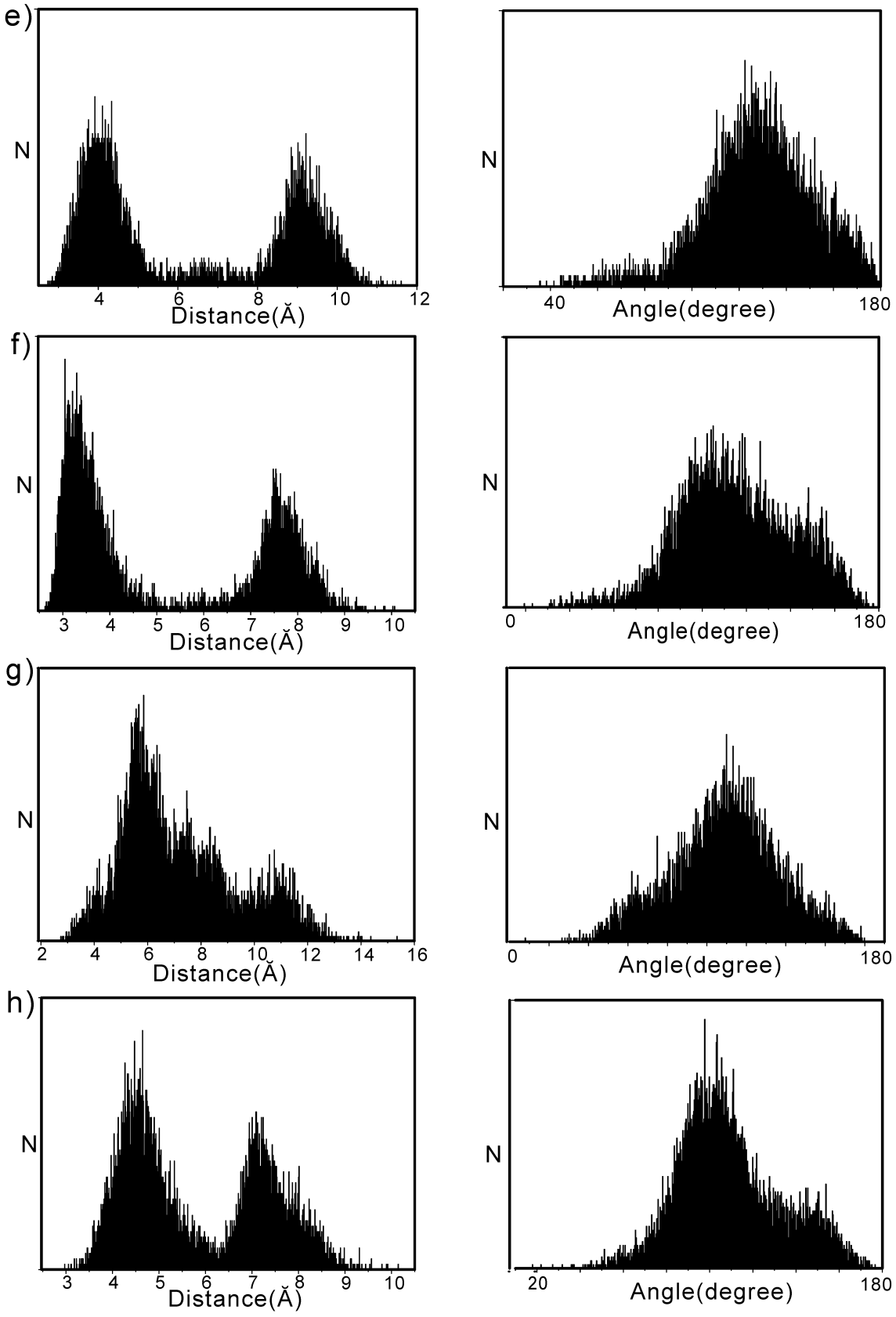
Although the individual fluorine bonds are apparently weak (given their transient nature), their combined effects would, nevertheless, be expected to enhance the kinetic stabilities of the $(Lg+XF\text{-}SA)^{7-}$ ions over those of the corresponding $(Lg+SA)^{7-}$ ions.

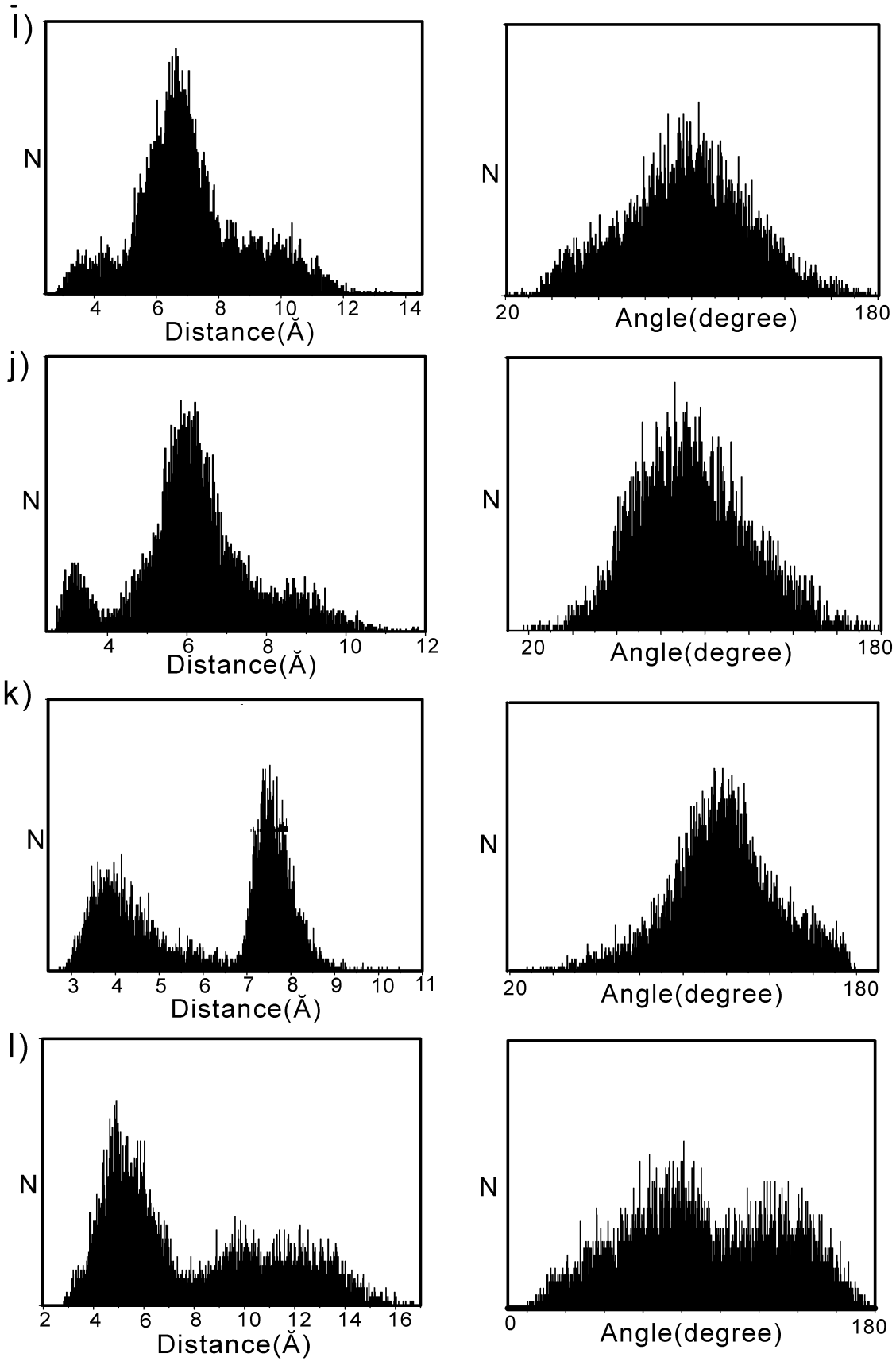
The results of the MD simulations also suggest that the degree of fluorination of SA influences the formation of hydrogen bonds between the XF-SA carboxyl group and Lg. For the $(Lg+21F\text{-}SA)^{7-}$ ion no stable intermolecular hydrogen bonds were identified, while for the $(Lg+13F\text{-}SA)^{7-}$ ion, the carboxyl group can form hydrogen bonds with

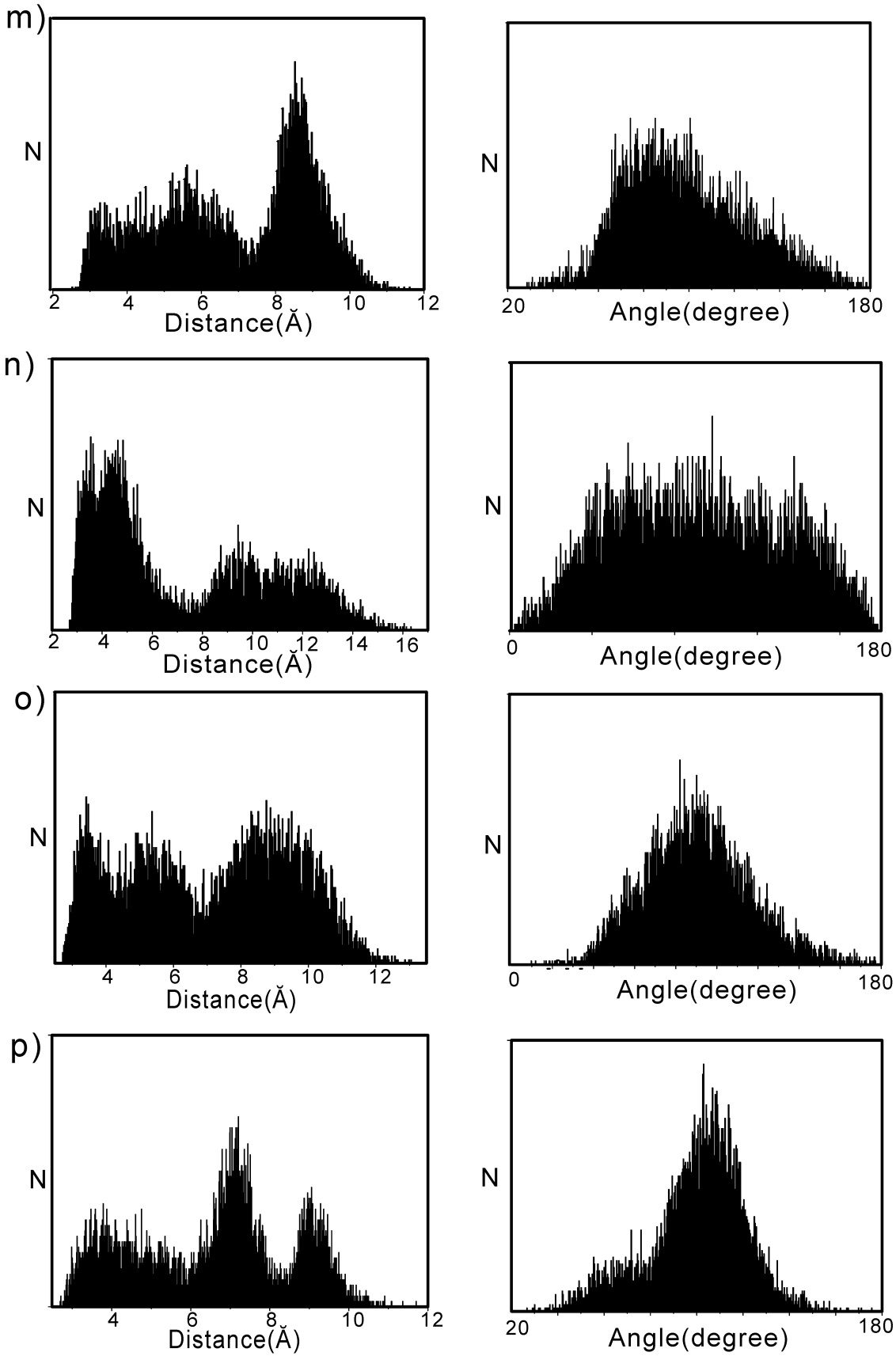
Lys[60], Glu[62], and Asn[88] (See Figure 3. 10). Differences in hydrogen bonding for the different XF-SA ligands could explain the absence of correlation between the degree of fluorination and the E_a values for the $(Lg + XF\text{-}SA)_s^{7-}$ ions and the larger E_a measured for $(Lg + 13F\text{-}SA)_s^{7-}$ (27.9 kcal mol^{-1}), compared to $(Lg + 21F\text{-}SA)_s^{7-}$ (25.6 kcal mol^{-1}).

The present results clearly demonstrate that fluorocarbon binding within the hydrophobic cavity of Lg is energetically preferred to hydrocarbon binding. Given that there is nothing remarkable about the residues that make up the Lg cavity, it is reasonable to conclude that, generally, fluorocarbon binding inside hydrophobic protein cavities will be energetically more favourable than hydrocarbon binding and, in the absence of differential solvent effects, will lead to enhanced binding in aqueous solution. If that is indeed the case, then the present findings would seem to argue for a refinement of the conclusions drawn recently by Whitesides and coworkers regarding the origin of the thermodynamic differences in binding of alkyl and fluoroalkyl groups to HCA II.[23] Fluorination was found to enhance both the enthalpy and entropy of ligand binding to the protein. The authors concluded that these changes arose primarily from differences in the solvent accessible surface area of hydrocarbon and fluorocarbon moieties and not from differences in intermolecular interaction energies. While it is possible that in the case of HCA II the interactions with alkyl and fluoroalkyl chains are energetically similar, it is more likely that the thermodynamic differences do reflect, at least in part, a greater intrinsic stability of the fluorinated ligands.


a)
N
3 4 5 6 7 8 9 10
Distance(Ă)
N
40
Angle(degree)
140
b)
N
3 4 5 6 7 8 9 10
Distance(Ă)
N
20
Angle(degree)
160
c)
N
3 4 5 6 7 8 9
Distance(Ă)
N
20
Angle(degree)
160
d)
N
4 6 8 10
Distance(Ă)
N
40
Angle(degree)
180


e)
N
4
6
8
10
12
Distance(Å)
N
40
Angle(degree)
180
f)
N
3
4
5
6
7
8
9
10
Distance(Å)
N
0
Angle(degree)
180
g)
N
2
4
6
8
10
12
14
16
Distance(Å)
N
0
Angle(degree)
180
h)
N
3
4
5
6
7
8
9
10
Distance(Å)
N
20
Angle(degree)
180


i)
N
4 6 8 10 12 14
Distance(Ă)
N
20
Angle(degree)
180
j)
N
4 6 8 10 12
Distance(Ă)
N
20
Angle(degree)
180
k)
N
3 4 5 6 7 8 9 10 11
Distance(Ă)
N
20
Angle(degree)
180
l)
N
2 4 6 8 10 12 14 16
Distance(Ă)
N
0
Angle(degree)
180


m)
N
2
4
6
8
10
12
Distance(Ă)
N
20
Angle(degree)
180
n)
N
2
4
6
8
10
12
14
16
Distance(Ă)
N
0
Angle(degree)
180
o)
N
4
6
8
10
12
Distance(Ă)
N
0
Angle(degree)
180
p)
N
4
6
8
10
12
Distance(Ă)
N
20
Angle(degree)
180

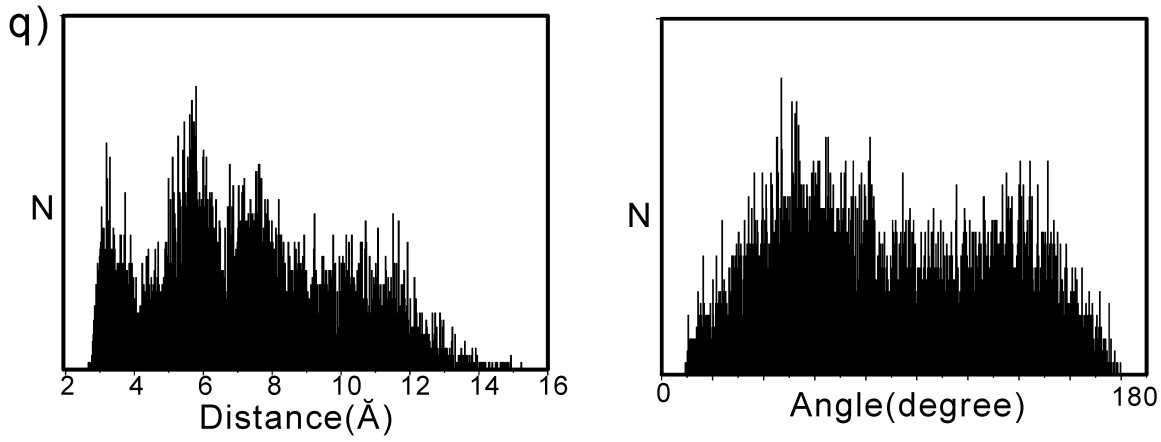
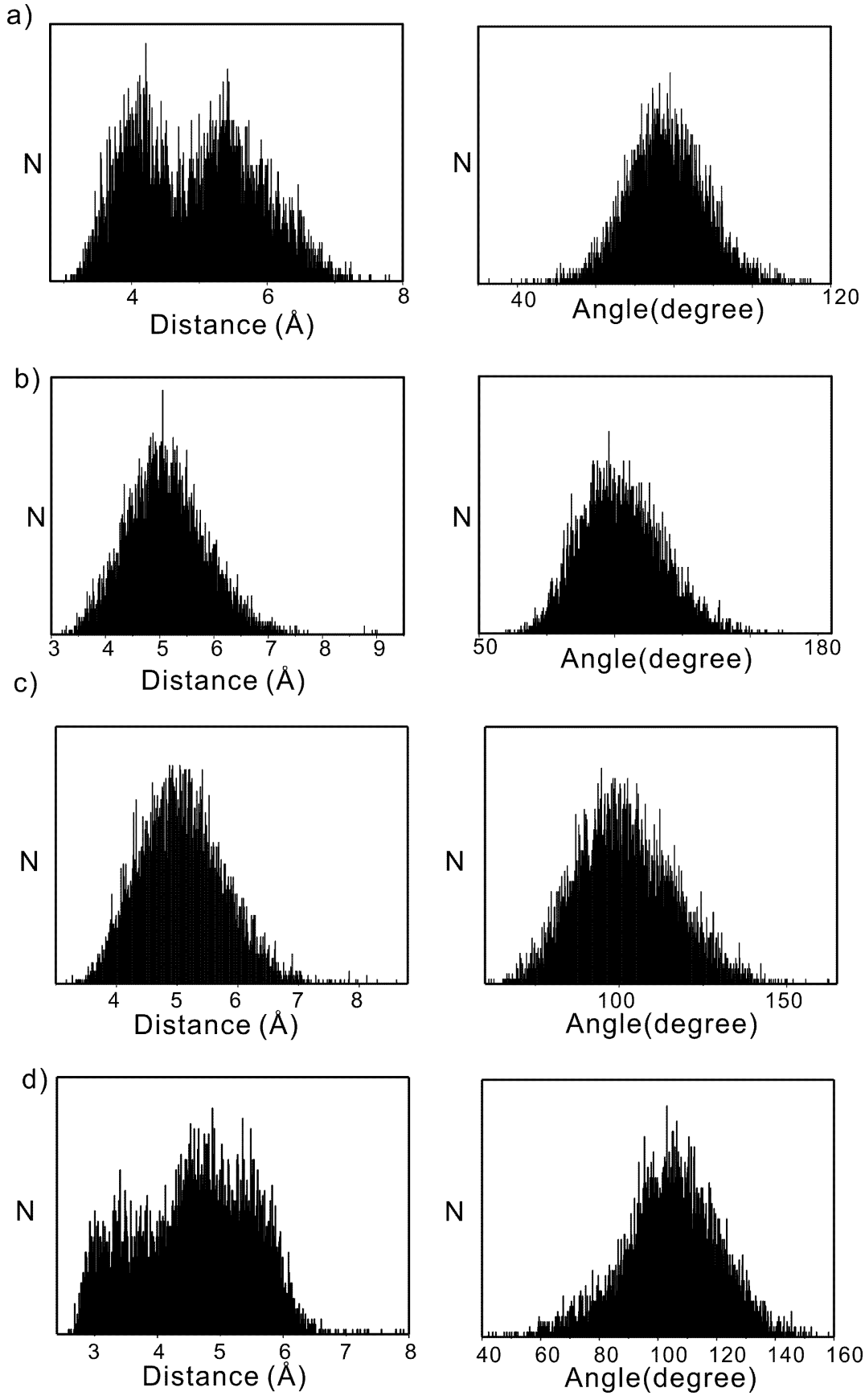


Figure 3. 8. Distance (left) and angle (right) distributions obtained from MD simulations performed on the $(Lg+21F\text{-}SA)^{7-}$ ion. The deprotonated residues are: Asp^{11}, Asp^{28}, Asp^{85}, Asp^{129}, Glu^{112}, Glu^{127}, Glu^{131} and C-terminus Ile^{162}. (a) F^{7}/Gln^{120} H_2N (side chain), (b) F^{8}/Gln^{120} H_2N (side chain), (c) F^{9}/Gln^{120} H_2N (side chain), (d) F^{10}/Asn^{90} H_2N (side chain), (e) F^{11}/Asn^{90} H_2N (side chain), (f) F^{14}/Asn^{90} H_2N (side chain), (g) F^{14}/Asn^{88} H_2N (side chain), (h) F^{15}/Asn^{90} H_2N (side chain), (i) F^{15}/Asn^{88} H_2N (side chain), (j) F^{17}/Asn^{88} H_2N (side chain), (k) F^{18}/Asn^{90} H_2N (side chain), (l) F^{18}/Lys^{69} H_2N (side chain), (m) F^{19}/Asn^{88} H_2N (side chain), (n) F^{19}/Lys^{69} H_2N (side chain), (o) F^{21}/Asn^{88} H_2N (side chain), (p) F^{21}/Asn^{90} H_2N (side chain), (q) F^{21}/Lys^{69} H_2N (side chain); The fluorine numbering scheme is the same shown in Figure 3. 7a and N is the number of occurrence.


a)
N
4
6
8
Distance (Å)
N
40
120
Angle(degree)
b)
N
3
4
5
6
7
8
9
Distance (Å)
N
50
180
Angle(degree)
c)
N
4
5
6
7
8
Distance (Å)
N
100
150
Angle(degree)
d)
N
3
4
5
6
7
8
Distance (Å)
N
40
60
80
100
120
140
160
Angle(degree)

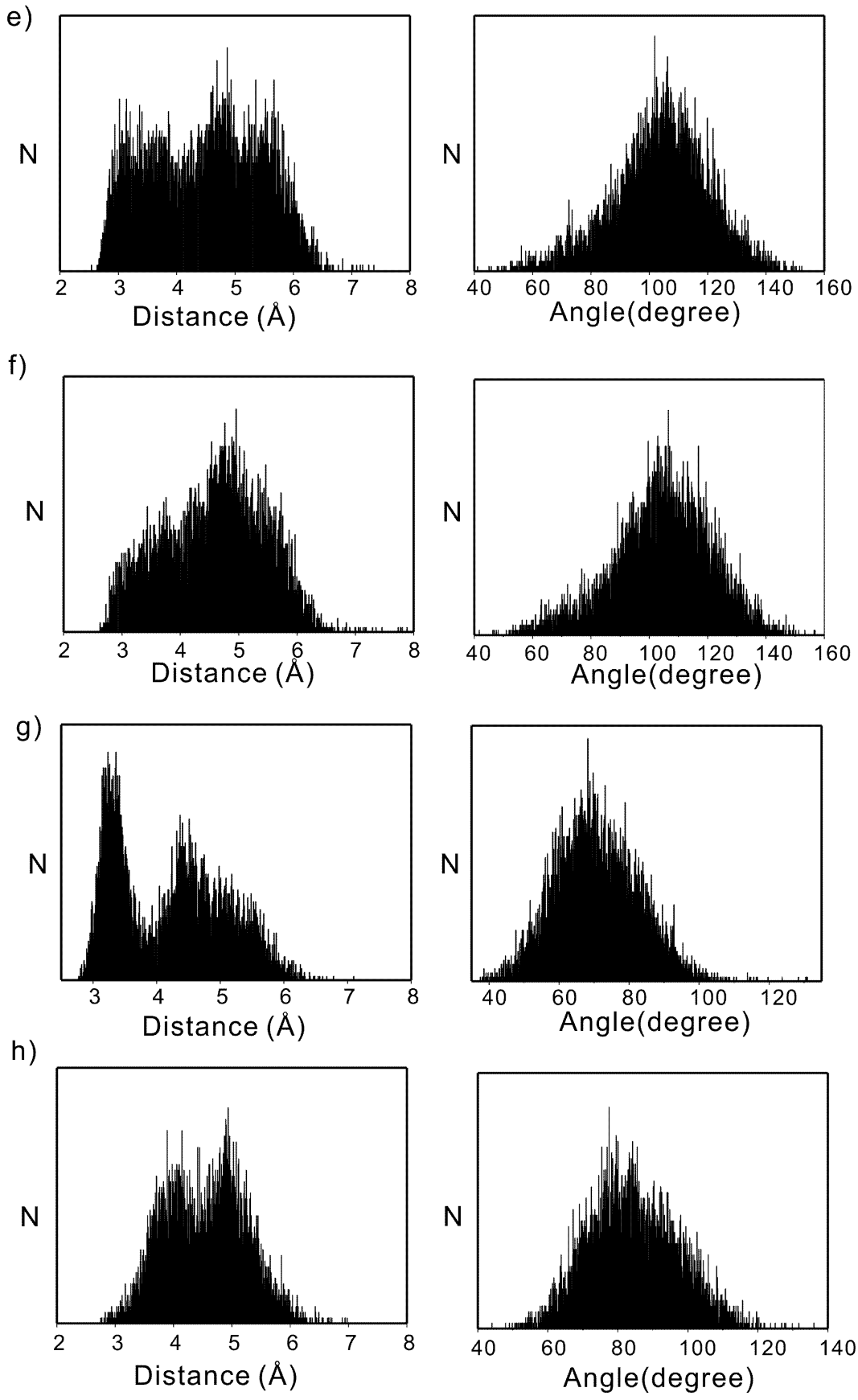
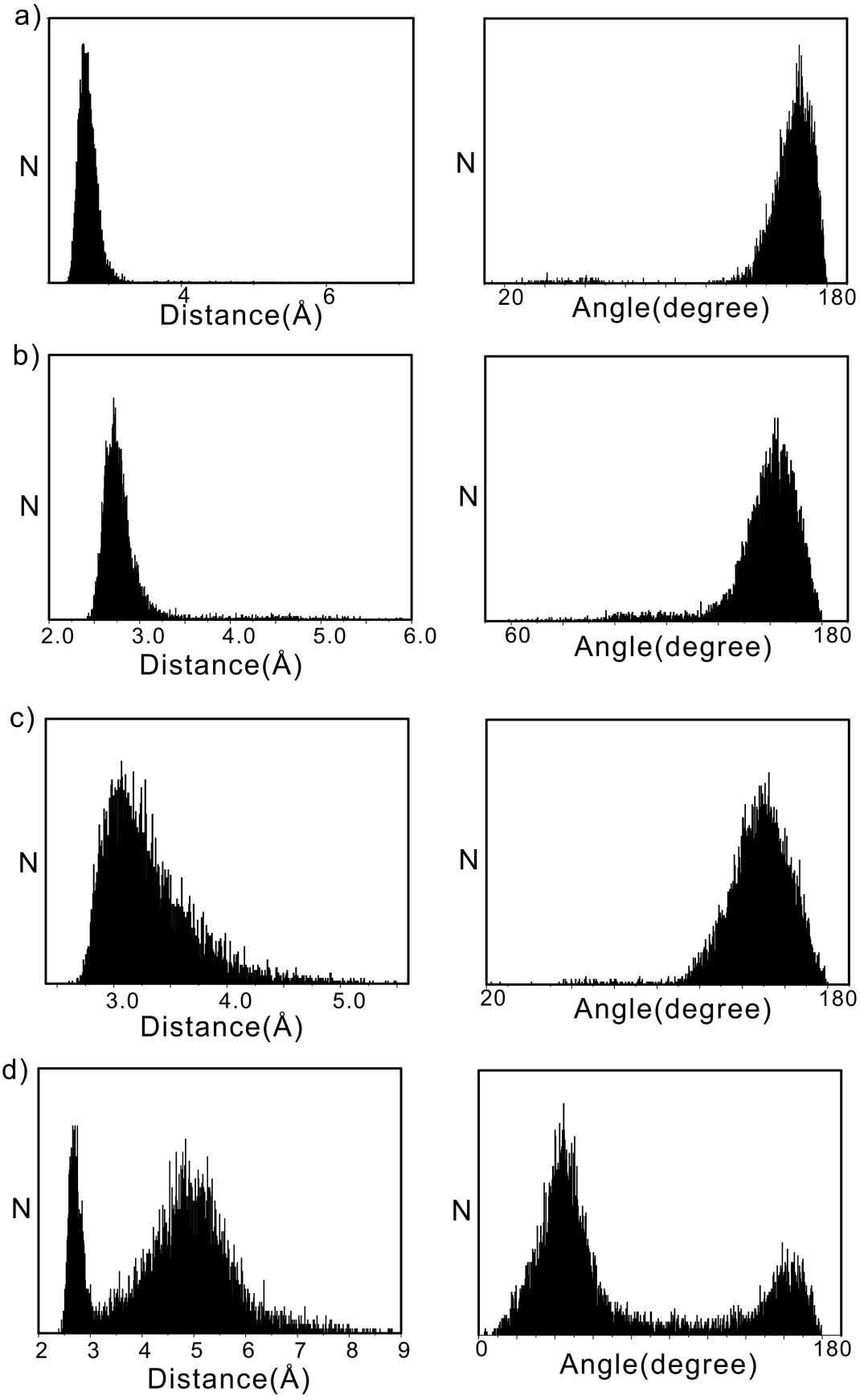

e)
N
2 3 4 5 6 7 8
Distance (Å)
N
40 60 80 100 120 140 160
Angle(degree)
f)
N
2 3 4 5 6 7 8
Distance (Å)
N
40 60 80 100 120 140 160
Angle(degree)
g)
N
3 4 5 6 7 8
Distance (Å)
N
40 60 80 100 120
Angle(degree)
h)
N
2 3 4 5 6 7 8
Distance (Å)
N
40 60 80 100 120 140
Angle(degree)

Figure 3. 9. Distance (left) and angle (right) distributions obtained from MD simulations performed on the $(Lg+13F\text{-}SA)^{7-}$ ion. The deprotonated residues are: Asp^{11}, Asp^{28}, Asp^{85}, Asp^{129}, Glu^{112}, Glu^{127}, Glu^{131} and C-terminus Ile^{162}. (a) F^4/Leu^{93} HN (amide N), (b) F^2/Phe^{82} HN (amide N), (c) F^3/Phe^{82} HN (amide N), (d) F^1/Val^{94} HN (amide N), (e) F^2/Val^{94} HN (amide N), (f) F^3/Val^{94} HN (amide N), (g) F^4/Val^{94} HN (amide N) and (h) F^5/Val^{94} HN (amide N); The fluorine numbering scheme is the same as shown in Figure 3. 7b and N is the number of occurrence.


a)
N
4
6
Distance(Å)
N
20
Angle(degree)
180
b)
N
2.0
3.0
4.0
5.0
6.0
Distance(Å)
N
60
Angle(degree)
180
c)
N
3.0
4.0
5.0
Distance(Å)
N
20
Angle(degree)
180
d)
N
2
3
4
5
6
7
8
9
Distance(Å)
N
0
Angle(degree)
180

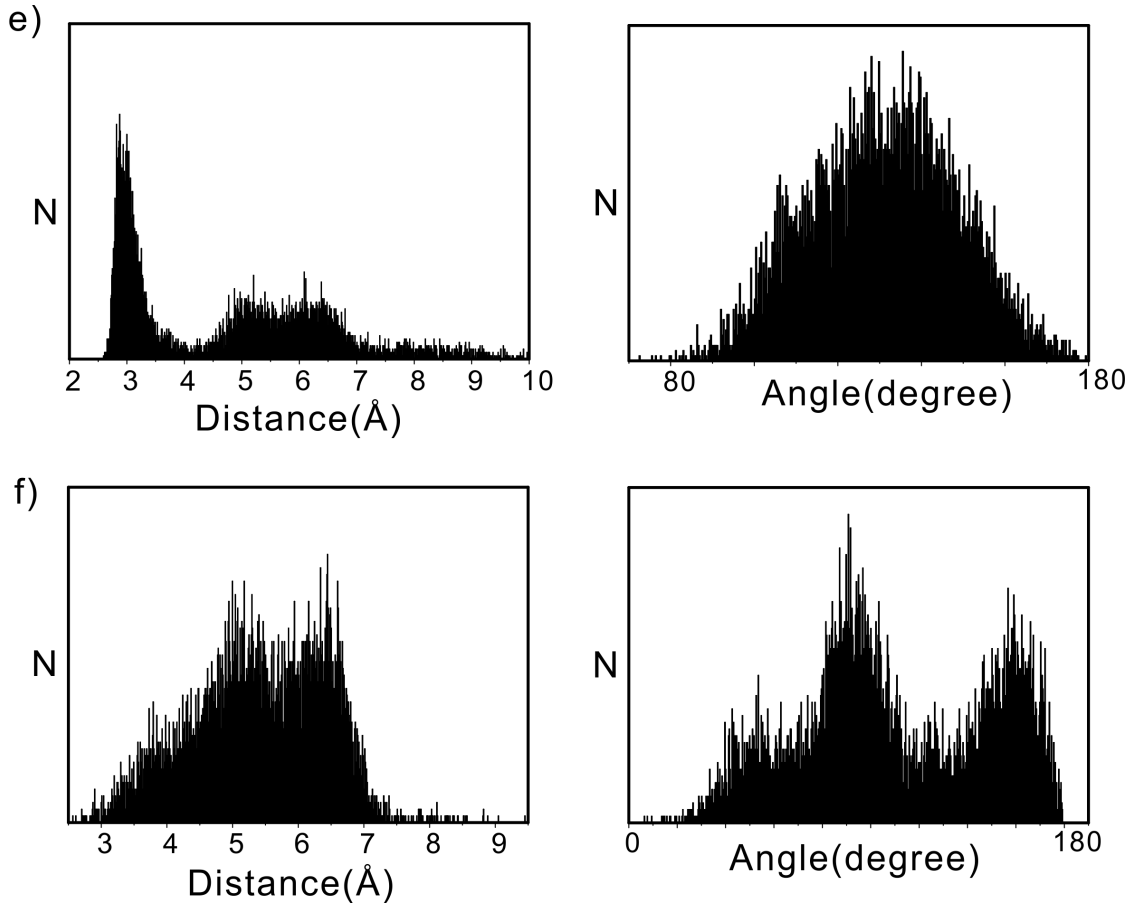


Figure 3. 10. Distance (left) and angle (right) distributions for H-bonds obtained from MD simulations performed on the (a-c) $(Lg+13F\text{-}SA)^{7-}$ ion and (d-f) $(Lg+21F\text{-}SA)^{7-}$ ion. The deprotonated residues are: Asp^{11}, Asp^{28}, Asp^{85}, Asp^{129}, Glu^{112}, Glu^{127}, Glu^{131} and C-terminus Ile^{162}. (a) 13F-SA C=O/Glu^{62} OH side chain (hydrogen donor), (b) 13F-SA –OH/Asn^{88} O side chain (hydrogen acceptor), (c) 13F-SA C=O/Lys^{60} H_2N, side chain (hydrogen donor), (d) 21F-SA C=O/Glu^{62} OH side chain (hydrogen donor), (e) 21F-SA C=O/Asn^{63} amide NH (hydrogen donor), and (f) 21F-SA -OH/Ser^{36} amide O (hydrogen acceptor); N is the number of occurrence.

3. 4. Conclusions

In summary, the intrinsic energetics of non-covalent fluorine bonding in a protein-ligand complex were measured for the first time. Fluorination of SA was found to increase the dissociation E_a for ligand loss from complexes with Lg in the gas phase. The average

energetic contribution of $>CF_2$ groups to E_a is measurably larger than the value reported previously for $>CH_2$ groups and relatively insensitive to position on the alkyl chain, suggesting that the Lg cavity presents a relatively homogeneous solvation environment to the fluoroalkyl chains. According to the results of MD simulations, fluorine bonding to polar hydrogens is primarily responsible for the stabilizing effect of fluorination. Future efforts will investigate the stabilities of other protein-ligand complexes in the gas phase, including those of HCA II, with the goal of more generally establishing the influence fluorine bonding on the thermodynamics of protein-ligand interactions.

References

1. Banks, R. E.; Smart, B. E.; Tatlow, J. C. *Organofluorine Chemistry: Principles and Commercial Applications*; Plenum Press: New York, **1994**.

2. Clader, J. W.: The Discovery of Ezetimibe a View from Outside the Receptor. *J. Med. Chem.* **2004**, *47*, 1-9.

3. Hugel, H. M.; Jackson, N. Special Feature Organo-Fluorine Chemical Science *Appl. Sci.* **2012**, *2*, 558-565.

4. Gunduz, M.; Argikar, U. A.; Kamel, A.; Colizza, K.; Bushee, J.; Cirello, A.; Lombardo, F.; Harriman, S. Oxidative *ipso* Substitution of 2,4-Difluoro-benzylphthalazines: Identification of a Rare Stable Quinone Methide and Subsequent GSH Conjugate. *Drug Metab. Dispos.* **2012**, *40*, 2074-2080.

5. Vasdev, N.; Dorff, P. N.; O'Neil, J. P.; Chin, F. T.; Hanrahan, S.; Van Brocklin, H. F. Metabolic Stability of 6,7-dialkoxy-4-(2-,3-and 4-[^{18}F] fluoroanilino) quinazolines, Potential EGFR Imaging Probes. *Bioorg. Med Chem.* **2011**, *19*, 2959-2965.

6. Böhm, H.; Banner, D.; Bendels, S.; Kansy, M.; Kuhn, B.; Müller, K.; Obst-Sander, U.; Stahl, M. Fluorine in Medicinal Chemistry. *ChemBioChem.* **2004**, *5*, 637-643.

7. Kim, C.; Chang, J. S.; Doyon, J. B.; Baird, T. T. Jr.; Fierke, C. A.; Jain, A.; Christianson, D. W. Contribution of Fluorine to Protein-Ligand Affinity in the Binding of Fluoroaromatic Inhibitors to Carbonic Anhydrase II. *J. Am. Chem. Soc.* **2000**, *122*, 12125-12134.

8. Rendine, S.; Pieraccini, S.; Forni, A.; Sironi, M.: Halogen Bonding in Ligand–Receptor Systems in the Framework of Classical force fields. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19508-19516.

9. Benitex, Y.; Baranger, A. M. Control of the Stability of a Protein-RNA Complex by the Position of Fluorine in a Base Analogue. *J. Am. Chem. Soc.* **2011**, *133*, 3687-3689.

10. Lee, Y.; Zeng, H.; Ruedisser, S.; Gossert, A. D.; Hilty, C. Nuclear Magnetic Resonance of Hyperpolarized Fluorine for Characterization of Protein-Ligand Interactions. *J. Am. Chem. Soc.* **2012**, *134*, 17448-17451.

11. Zhou, P.; Zou, J.; Tian, F.; Shang, Z. Fluorine Bonding: How Does It Work In Protein-Ligand Interactions. *J. Chem. Inf. Model.* **2009**, *49*, 2344-2355.

12. Kuhn, B.; Kollman, P. A. A Ligand That Is Predicted to Bind Better to Avidin than Biotin: Insights from Computational Fluorine Scanning. *J. Am. Chem. Soc.* **2000**, *122*, 3909-3916.

13. Kawahara, S.; Tsuzuki, S.; Uchimaru, T. Theoretical Study of the C-F/π Interaction: Attractive Interaction between Fluorinated Alkane and an Electron-Deficient π–System. *J. Phys. Chem. A.* **2004**, *108*, 6744-6749.

14. Riley, K. E.; Merz, K. M. Effects of Fluorine Substitution on the Edge-to-Face Interaction of the Benzene Dimer. *J. Phys. Chem. B* . **2005**, *109,* 17752-17756.

15. Chopra, D.; Nagarajan, K.; Row, T. N. G. Analysis of Weak Interactions Involving Organic Fluorine: Insights from Packing Features in Substituted 4-keto-tetrahydroindoles, *J. Mol. Struct*. **2008**, *888*, 70-83.

16. Iwaoka, M.; Komatsu, H.; Katsuda, T.; Tomoda, S. Quantitative Evaluation of Weak Nonbonded Se···F Interactions and Their Remarkable Nature as Orbital Interactions. *J. Am. Chem. Soc.* **2002**, *124*, 1902-1909.

17. Matta, C. F.; Castillo, N.; Boyd, R. J. Characterization of a Closed-Shell Fluorine-Fluorine Bonding Interaction in Aromatic Compounds on the Basis of the Electron Density. *J. Phys. Chem. A.* **2005**, *109*, 3669-3681.

18. Lee, S.; Mallik, A. B.; Fredrickson, D. C. Dipolar-Dipolar Interactions and the Crystal Packing of Nitriles, Ketones, Aldehydes, and C(sp^2)-F Groups. *Cryst. Growth Des*. **2004**, *4,* 279-290.

19. Olsen, J. A.; Banner, D. W.; Seiler, P.; Sander, U. O.; D'Arcy, A.; Stihle, M.; Müller, K.; Diederich, F. A Fluorine Scan of Thrombin Inhibitors to Map the Fluorophilicity/Fluorophobicity of an EnzymeActive Site: Evidence for C-F···C=O Interactions. *Angew. Chem. Int.Ed.* **2003**, *42,* 2507-2511.

20. Plenio, H.: The Coordination Chemistry of Fluorine in Fluorocarbons. *ChemBioChem.* **2004***, 5*, 650-655.

21. Bettinger, H. F. How Good is Fluorine as a Hydrogen-Bond Acceptor in Fluorinated Single-Walled Carbon Nanotubes? *ChemPhysChem*. **2005**, *6,* 1169-1174.

22. Müller, K.;Faeh, C.; Diederich, F. Fluorine in Pharmaceuticals: Looking Beyond Intuition. *Science* **2007**, *317*, 1881-1886.

23. Mecinović, J.; Snyder, P. W.; MIRICA, K. A; Bai, S.; Mack, E. T.; Kwant, R. L.; Moustakas, D. T.; Héroux, A.; Whitesides, G. M. Fluoroalkyl and Alkyl Chains Have Similar Hydrophobicities in Binding to the "Hydrophobic Wall" of Carbonic Anhydrase. *J. Am. Chem. Soc.* **2011***, 133*, 14017-14026.

24. Kontopidis, G.; Holt, C.; Sawyer, L. Beta-Lactoglobulin: Binding Properties, Structure, and Function. *J. Dairy Sci*. **2004**, *87*, 785-796.

25. Qin, B.Y.; Bewley, M.C.; Creamer, L.K.; Baker, H.M.; Baker, E.N.; Jameson, G.B. Structural Basis of the Tanford Transition of Bovine Beta-Lactoglobulin. *Biochem.* **1998**, *37*, 14014-14023.

26. Dunbar, R.C., McMahon, T.B. Activation of Unimolecular Reactions by Ambient Blackbody Radiation. *Science* **1998**, *279*, 194-197.

27. Price, W.D.; Schnier, P.D.; Jockusch, R.A.; Strittmatter, E.R.; Williams, E.R. Unimolecular Reactions Kinetics in the High-Pressure Limit Without Collisions. *J. Am. Chem. Soc.* **1996**, *118*, 10640-10644.

28. Liu, L.; Bagal, D.;Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Hydrophobic Protein-Ligand Interactions Preserved in the Gas Phase. *J. Am. Chem. Soc.* **2009**, *131*, 15980-15981.

29. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Chem. Inform.* **2012**, *4*, 17-22.

30. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. Y. R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.*, **2011**, *39*, W511−W517.

31. Dupradeau, F.-Y.;Pigache, A.; Zaffran, T.; Savineau, C.;Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak P. The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Feld Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821−7839.

32. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.;Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima,T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09. *Gaussian Inc*. **2009**.

33. Wu, S.Y.; Perez, M. D.; Puyol, P.; Sawyer, L. β-Lactoglobulin Binds Palmitate within its Central Cavity. *J. Biol. Chem.* **1999**, *274*, 170-174.

34. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. -J.; Cui, G.; Roe D. R.; Mathews D. H.; Seetin M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin V.; Luchko, T.; Gusarov, S.; Kovalenko A.; Kollman, P. A. AMBER 12. **2012**.

35. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD, *J. Comput. Chem.* **2005**, *26*, 1781-1802.

36. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development of Testing of a General Amber Force Field, *J. Comput. Chem.* **2004***, 25*, 1157-1174.

37. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes, *J. Comput. Phys.* **1977**, *23*, 327-341.

38. Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics, *J. Molec. Graphics* **1996***, 14.1*, 33-38.

39. *Origin;* OriginLab, Northampton, MA 01060 USA.

40. Sun, J.; Kitova, E.N.; Klassen, J.S. Method for Stabilizing Protein-Ligand Complexes in Nnanoelectrospray Ionization Mass Spectrometry. *Anal. Chem*. **2007**, *79*, 416-425.

41. Bagal, D.; Kitova, E.N.; Liu, L.; El-Haweit, A.; Schnier, P. D.; Klassen, J.S. Gas Phase Stabilization of Noncovalent Protein Complexes Formed by Electrospray Ionization. *Anal. Chem*. **2009**, *81*, 7801-7806.

42. Liu, L.; Michelsen, K.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Energetics of Lipid Binding in a Hydrophobic Protein Cavity. *J. Am. Chem. Soc.* **2012**, *134*, 3054-3060.

Chapter 4

Conclusions and future works

In Chapter 2, PMF calculations were used to calculate the activation free energy for the dissociation of fatty acid ligands from Lg binding cavity. Consequently, the dissociation rate constants were extracted from the calculated PMF via transition state theory. The ultimate goals were to reproduce the dissociation rate constants of the fatty acids Lg binding cavity and also gain more insight into the exit pathway of the fatty acids. The PMFs of desolvated protein-fatty acid complexes were computed along the reaction coordinate for two different charge configurations of protein. The reaction coordinate was defined as the distance between the COM of carboxyl group of the fatty acid and the COM of the protein binding cavity. An extensive umbrella sampling was employed along the dissociation coordinate in order to enhance the sampling. The PMF plots of the protein-fatty acid complexes were used to estimate the dissociation activation free energies. In general, the results of PMF calculations demonstrate that the dissociation activation free energies cannot be determined accurately using only single protein and ligand conformation. However, the PMF calculations presented in this thesis provided the dissociation activation free energies of two ligands within 1.0 kcal mol^{-1} of the experimental values for one of the charge configurations. Moreover, computations were relatively successful in predicting the trend of dissociation activation free energies for the other charge configuration. Conformational rearrangements and being trapped in metastable states lead to a deficiency in sampling, which was the main obstacle in free energy calculations. Free energy profiles along the dissociation coordinate and analysis of the trajectory files demonstrated that the carboxyl group of the fatty acid interacts with a

variety of the residues on the flexible loops at opening of the cavity. Nevertheless, there is no hydrogen bond between the fatty acid head group and charged amino acids in the fast components of the protein-fatty acid complexes. The calculated activation free energies were used to compute the dissociation rate constants for protein-fatty acid complexes. The results indicated that prediction of the kinetics of the dissociation process using computations is even more challenging but the predicted dissociation rate constant of stearic acid in one of the charge configurations has an excellent agreement with the experimental value. The PMF plots revealed that the dissociation distance is longer for the larger fatty acid. The dissociation coordinate is the distance between fatty acid head group and the COM of the protein binding site. Thus, the cleavage of the nonpolar interactions is the last step of the activation process. In summary, the dissociation rate constants of the ligands from the protein binding site were not predicted accurately using computations where the ligands were flexible and there was a conformational change upon dissociation.

Investigation of the trajectory files along the reaction coordinate and comparison between the protein conformations in bound and unbound states may reveal which degrees of freedom facilitate the ligand exit. However, caution must be taken in interpretation of the accompanying degrees of freedom due to the fact that the entire configurational space has not sampled. The other solution for identifying the slow degrees of freedom is calculation of the correlation time. Calculation of the correlation time for the dissociation coordinate identifies the most probable residues involved in ligand dissociation by determining the distance, which shows the highest decorrelation time. These studies may provide a more detailed picture of the ligand dissociation

pathway. Estimation of the dissociation kinetics can also improve using free energy calculations at lower temperatures where there is a larger difference between the dissociation activation energies of the ligands.

In Chapter 3, the nature of the stabilizing intermolecular interactions in protein-fluorinated fatty acid complexes was probed using MD simulations. A more complete molecular picture of the fluorocarbon binding in the protein binding cavity was provided. Distributions of distance and angle of the possible hydrogen donors and fluorine atoms revealed that fluorine atoms are involved in interactions with polar hydrogen atoms within binding cavity. These interactions were transient but individual residues in the binding cavity interact with multiple fluorine atoms. Therefore, the combined effects of the interactions are considerable. Moreover, the results of the MD simulation suggest that the degree of fluorination influences the hydrogen bond interactions of the carboxyl end of the fatty acid with protein residues.

Currently, is not clear how fluorine bonding influences the interactions and conformations of the protein and ligand in protein-ligand complexes. Studying various systems with fluorinated ligands provides a better understanding of the molecular basis of the fluorination influences on the affinity and selectivity of protein-ligand interactions.

Bibliography

Chapter 1

1. Huang, J. F.; Yang, Y.; Sepulveda, H.; Shi, W. X.; Hwang, I.; Peterson, P. A.; Jackson, M. R.; Sprent, J.; Cai, Z. L. TCR-mediated internalization of peptide-MHC complexes acquired by T cells. *Science* **1999**, *286*, 952-954.
2. Arikawa, T.; Simamura, E.; Shimada, H.; Nakamura, T.; Hatta, T.; Shoji, H. Significance of sugar chain recognition by galectins and its involvement in disease-associated glycosylation. *Congenital anomalies* **2014**, *54*, 77-81.
3. Igakura, T.; Stinchcombe, J. C.; Goon, P. K. C.; Taylor, G. P.; Weber, J. N.; Griffiths, G. M.; Tanaka, Y.; Osame, M.; Bangham, C. R. M. Spread of HTLV-I between lymphocytes by virus-induced polarization of the cytoskeleton. *Science* **2003**, *299*, 1713-1716.
4. Patra, S. M.; Bastug, T.; Kuyucak, S. Binding of organic cations to gramicidin a channel studied with AutoDock and molecular dynamics simulations. *J. Phys. Chem. B* **2007**, *111*, 11303-11311.
5. Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234-2246.
6. Shirts, M. R.; Mobley, D. L.; Chodera, J. D. Alchemical Free Energy Calculations: Ready for Prime Time? *Ann. Rep. Comput. Chem.* **2007**, *3*, 41-59.
7. Chipot, C. Frontiers in free-energy calculations of biological systems. *Rev. Comput. Mol. Sci.* **2014**, *4*, 71-89.
8. Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395-2417.

9. Woo, H.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005***, 102*, 6825-6830.

10. Lee, M.; Olson, M. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys. J.* **2006***, 90*, 864-877.

11. Jorgensen, W. L. Interactions between Amides in Solution and the Thermodynamics of Weak Binding. *J. Am. Chem. Soc.* **1989***, 111*, 3770-3771.

12. Hermans, J.; Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.* **1997***, 119*, 2707-2714.

13. Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006***, 125*, 084902.

14. Woo, H.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005***, 102*, 6825-6830.

15. Liu, L.; Bagal, D.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Hydrophobic Protein-Ligand Interactions Preserved in the Gas Phase. *J. Am. Chem. Soc.* **2009***, 131*, 15980-15981.

16. Liu, L.; Michelsen, K.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Energetics of Lipid Binding in a Hydrophobic Protein Cavity. *J. Am. Chem. Soc.* **2012***, 134*, 3054-3060.

17. Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J. Chem. Phys.* **2006***, 125*, 84901-84905.

18. Chipot, C.; Rozanska, X.; Dixit, S. B. Can free energy calculations be fast and accurate at the same time? Binding of low-affinity, non-peptide inhibitors to the SH2 domain of the src protein. *J. Comput. Aided Mol. Des.* **2005***, 19*, 765-770.

19. Freiteg, S.; Chu, V.; Penzotti, J.; Klumb, L.; To, R.; Hyre, D.; Trong, I.; Lybrand, T.; Stenkamp, R.; Stayton, P. A structural snapshot of an intermediate on the streptavidin–biotin dissociation pathway. *Proc. Natl. Acad. Sci.* **1999***, 96*, 8384-8389.

20. Benitex, Y.; Baranger, A. M. Control of the Stability of a Protein-RNA Complex by the Position of Fluorine in a Base Analogue. *J. Am. Chem. Soc.* **2011***, 133*, 3687-3689.

21. Kim, C.; Chang, J.; Doyon, J.; Baird, T.; Fierke, C.; Jain, A.; Christianson, D. Contribution of fluorine to protein-ligand affinity in the binding of fluoroaromatic inhibitors to carbonic anhydrase II. *J. Am. Chem. Soc.* **2000***, 122*, 12125-12134.

22. Lee, Y.; Zeng, H.; Ruedisser, S.; Gosser, A. D.; Hilty, C. Nuclear Magnetic Resonance of Hyperpolarized Fluorine for Characterization of Protein-Ligand Interactions. *J. Am. Chem. Soc.* **2012***, 134*, 17448-17451.

23. Bohm, H.; Banner, D.; Bendels, S.; Kansy, M.; Kuhn, B.; Muller, K.; Obst-Sander, U.; Stahl, M. Fluorine in medicinal chemistry. *Chembiochem* **2004***, 5*, 637-643.

24. Kawahara, S.; Tsuzuki, S.; Uchimaru, T. Theoretical study of the C-F-π interaction: Attractive interaction between fluorinated alkane and an electron-dericient π -system. *J. Phys. Chem. A* **2004***, 108*, 6744-6749.

25. Riley, K.; Merz, K. Effects of fluorine substitution on the edge-to-face interaction of the benzene dimer. *J. Phys. Chem. B* **2005***, 109*, 17752-17756.

26. Iwaoka, M.; Komatsu, H.; Katsuda, T.; Tomoda, S. Quantitative evaluation of weak nonbonded Se...F interactions and their remarkable nature as orbital interactions. *J. Am. Chem. Soc.* **2002***, 124*, 1902-1909.

27. Chopra, D.; Nagarajan, K.; Row, T. N. G. Analysis of weak interactions involving organic fluorine: Insights from packing features in substituted 4-keto-tetrahydroindoles. *J. Mol. Struct.* **2008***, 888*, 70-83.

28. Lee, S.; Mallik, A.; Fredrickson, D. Dipolar-dipolar interactions and the crystal packing of nitriles, ketones, aldehdyes, and C(sp(2))-F groups. *Cryst. Growth Des.* **2004***, 4*, 279-290.

29. Matta, C.; Castillo, N.; Boyd, R. Characterization of a closed-shell fluorine-fluorine bonding interaction in aromatic compounds on the basis of the electron density. *J Phys Chem A* **2005***, 109*, 3669-3681.

30. Olsen, J.; Banner, D.; Seiler, P.; Sander, U.; D'Arcy, A.; Stihle, M.; Muller, K.; Diederich, F. A fluorine scan of thrombin inhibitors to map the fluorophilicity/fluorophobicity of an enzyme active site: Evidence for C-F center dot center dot center dot C=O interactions. *Angew. Chem. Int. Edit.* **2003***, 42*, 2507-2511.

31. Bettinger, H. How good is fluorine as a hydrogen-bond acceptor in fluorinated single-walled carbon nanotubes? *ChemPhysChem* **2005***, 6*, 1169-1174.

32. Mueller, K.; Faeh, C.; Diederich, F. Fluorine in pharmaceuticals: Looking beyond intuition. *Science* **2007***, 317*, 1881-1886.

33. Mecinović, J.; Snyder, P. W.; Mirica, K. A.; Bai, S.; Mack, E. T.; Kwant, R. L.; Moustakas, D. T.; Héroux, A.; Whitesides, G. M. Fluoroalkyl and Alkyl Chains

Have Similar Hydrophobicities in Binding to the "Hydrophobic Wall" of Carbonic Anhydrase. *J. Am. Chem. Soc.* **2011***, 133*, 14017-14026.

34. Liu, L.; Jalili, N.; Baergen, A.; Ng, S.; Bailey, J.; Derda, R.; Klassen, J. S. Fluorine Bonding Enhances the Energetics of Protein-Lipid Binding in the Gas Phase. *J. Am. Soc. Mass Spectrom.* **2014***, 25*, 751-757.

35. Kontopidis, G.; Holt, C.; Sawyer, L. Invited Review: beta-lactoglobulin: Binding properties, structure, and function. *J. Dairy Sci.* **2004***, 87*, 785-796.

36. Qin, B.; Bewley, M.; Creamer, L.; Baker, H.; Baker, E.; Jameson, G. Structural basis of the tanford transition of bovine beta-lactoglobulin. *Biochemistry (N. Y.)* **1998***, 37*, 14014-14023.

37. Flower, D. R.; North, A. C. T.; Attwood, T. K. Structure and Sequence Relationships in the Lipocalins and Related Proteins. *Protein Science* **1993***, 2*, 753-761.

38. Qin, B.; Bewley, M.; Creamer, L.; Baker, H.; Baker, E.; Jameson, G. Structural basis of the tanford transition of bovine beta-lactoglobulin. *Biochemistry (N. Y.)* **1998***, 37*, 14014-14023.

39. Kontopidis, G.; Holt, C.; Sawyer, L. Invited Review: beta-lactoglobulin: Binding properties, structure, and function. *J. Dairy Sci.* **2004***, 87*, 785-796.

40. Collini, M.; D'Alfonso, L.; Molinari, H.; Ragona, L.; Catalano, M.; Baldini, G. Competitive binding of fatty acids and the fluorescent probe 1-8-anilinonaphthalene sulfonate to bovine beta-lactoglobulin. *Protein Sci.* **2003***, 12*, 1596-1603.

41. Konuma, T.; Sakurai, K.; Goto, Y. Promiscuous binding of ligands by beta-lactoglobulin involves hydrophobic interactions and plasticity. *J. Mol. Biol.* **2007***, 368*, 209-218.

42. Ohtomo, H.; Konuma, T.; Utsunoiya, H.; Tsuge, H.; Ikeguchi, M. Structure and stability of Gyuba, a beta-lactoglobulin chimera. *Protein Sci.* **2011***, 20*, 1867-1875.

43. Sawyer, L.; Kontopidis, G. The core lipocalin, bovine beta-lactoglobulin. *Biochim. Biophys. Acta-Protein Struct. Molec. Enzym.* **2000***, 1482*, 136-148.

44. Levin, L. B.; Ganoth, A.; Amram, S.; Nachliel, E.; Gutman, M.; Tsfadia, Y. Insight into the interaction sites between fatty acid binding proteins and their ligands. *J. Mol. Model.* **2010***, 16*, 929-938.

45. Long, D.; Mu, Y.; Yang, D. Molecular Dynamics Simulation of Ligand Dissociation from Liver Fatty Acid Binding Protein. *PLoS One* **2009***, 4*, e6081.

46. Tsfadia, Y.; Friedman, R.; Kadmon, J.; Selzer, A.; Nachliel, E.; Gutman, M. Molecular dynamics simulations of palmitate entry into the hydrophobic pocket of the fatty acid binding protein. *FEBS Lett.* **2007***, 581*, 1243-1247.

47. Friedman, R.; Nachliel, E.; Gutman, M. Fatty acid binding proteins: Same structure but different binding mechanisms? Molecular dynamics simulations of intestinal fatty acid binding protein. *Biophys. J.* **2006***, 90*, 1535-1545.

48. Friedman, R.; Nachliel, E.; Gutman, M. Molecular dynamics simulations of the adipocyte lipid binding protein reveal a novel entry site for the ligand. *Biochemistry (N. Y.)* **2005***, 44*, 4275-4283.

49. Ragona, L.; Fogolari, F.; Zetta, L.; Perez, D.; Puyol, P.; De Kruif, K.; Lohr, F.; Ruterjans, H.; Molinari, H. Bovine beta-lactoglobulin: Interaction studies with palmitic acid. *Protein Sci.* **2000***, 9*, 1347-1356.

50. Dunbar, R.; McMahon, T. Activation of unimolecular reactions by ambient blackbody radiation. *Science* **1998***, 279*, 194-197.

51. Price, W.; Schnier, P.; Jockusch, R.; Strittmatter, E.; Williams, E. Unimolecular reaction kinetics in the high-pressure limit without collisions. *J. Am. Chem. Soc.* **1996***, 118*, 10640-10644.

52. McCammon, J.; Gelin, B.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977***, 267*, 585-590.

53. Jorgensen, W.; Ravimohan, C. Monte-Carlo Simulation of Differences in Free-Energies of Hydration. *J. Chem. Phys.* **1985***, 83*, 3050-3054.

54. Lybrand, T.; Mccammon, J.; Wipff, G. Theoretical Calculation of Relative Binding-Affinity in Host Guest Systems. *Proc. Natl. Acad. Sci. U. S. A.* **1986***, 83*, 833-835.

55. Postma, J.; Berendsen, H.; Haak, J. Thermodynamics of Cavity Formation in Water - a Molecular-Dynamics Study. *Faraday Symposia Chem. Soc.* **1982**, 55-67.

56. Warshel, A. Dynamics of Reactions in Polar-Solvents - Semi-Classical Trajectory Studies of Electron-Transfer and Proton-Transfer Reactions. *J. Phys. Chem.* **1982***, 86*, 2218-2224.

57. Kroon, G.; Mo, H.; Martinez-Yamout, M.; Dyson, H.; Wright, P. Changes in structure and dynamics of the Fv fragment of a catalytic antibody upon binding of inhibitor. *Protein Sci.* **2003***, 12*, 1386-1394.

58. Thorpe, I. F.; Brooks,Charles L.,,III Molecular evolution of affinity and flexibility in the immune system. *Proc. Natl. Acad. Sci. U. S. A.* **2007***, 104*, 8821-8826.

59. Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006***, 2*, 1255-1273.

60. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field (vol 25, pg 1157, 2004). *J. Chem. Theory Comput.* **2005***, 26*, 114-114.

61. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998***, 102*, 3586-3616.

62. Shirts, M.; Pitera, J.; Swope, W.; Pande, V. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* **2003***, 119*, 5740-5761.

63. Deng, Y.; Roux, B. Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B* **2004***, 108*, 16567-16576.

64. Wan, S.; Stote, R.; Karplus, M. Calculation of the aqueous solvation energy and entropy, as well as free energy, of simple polar solutes. *J. Chem. Phys.* **2004***, 121*, 9539-9548.

65. Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *J. Med. Chem.* **2008***, 51*, 769-779.

66. Mobley, D. L.; Chodera, J. D.; Dill, K. A. Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput.* **2007***, 3*, 1231-1235.

67. Morton, A.; Baase, W. A.; Matthews, B. W. Energetic Origins of Specificity of Ligand-Binding in an Interior Nonpolar Cavity of T4 Lysozyme. *Biochemistry (N. Y.)* **1995***, 34*, 8564-8575.

68. Rodinger, T.; Howell, P. L.; Pomes, R. Calculation of absolute protein-ligand binding free energy using distributed replica sampling. *J. Chem. Phys.* **2008***, 129*, 155102.

69. Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. Predicting Ligand Binding Affinity with Alchemical Free Energy Methods in a Polar Model Binding Site. *J. Mol. Biol.* **2009***, 394*, 747-763.

70. Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006***, 91*, 2798-2814.

71. Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C.; Shirts, M.; Sorin, E.; Pande, V. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **2005***, 123*, 084108.

72. Gan, W.; Roux, B. Binding specificity of SH2 domains: Insight from free energy simulations. *Proteins: Struct. Funct. Bioinf.* **2009***, 74*, 996-1007.

73. Gumbart, J. C.; Roux, B.; Chipot, C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.* **2013***, 9*, 794-802.

74. Lau, A. Y.; Roux, B. The hidden energetics of ligand binding and activation in a glutamate receptor. *Nat. Struct. Mol. Biol.* **2011***, 18*, 283-288.

75. Beveridge, D.; Dicapua, F. Free-Energy Via Molecular Simulation - Applications to Chemical and Biomolecular Systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989***, 18*, 431-492.

76. Luo, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc. Natl. Acad. Sci.* **2002***, 99*, 10399-10404.

77. Massova, I.; Kollman, P. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery* **2000***, 18*, 113-135.

78. Kirkwood, J. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935***, 3*, 300-313.

79. Zwanzig, R. High-Temperature Equation of State by a Perturbation Method .1. Nonpolar Gases. *J. Chem. Phys.* **1954***, 22*, 1420-1426.

80. Allen, T.; Andersen, O.; Roux, B. Energetics of ion conduction through the gramicidin channel. *Proc. Natl. Acad. Sci. U. S. A.* **2004***, 101*, 117-122.

81. Levine, I. N. *Physical Chemistry;* McGraw-Hill: New York, 2009; .

82. Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys. (Berlin)* **1927***, 389*, 457-484.

83. McQuarrie, D. A. *Statistical Mechanics;* University Science Books: Sausalito, CA, 2000.

84. Pathria, R. K. *Statistical Mechanics;* Reed Educational and Professional Publishing Ltd: Oxford, 1996.

85. Chandler, D. *Introduction to modern statistical mechanics;* Oxford University Press: New York, 1987.

86. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: from Algorithm to Application;* Academic Press: San Diego, CA, 1996.

87. Maxwell, J. C. Illustrations of the dynamical theory of gases. *Philos. Mag.* **1860***, 19*, 19-32.

88. Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993***, 93*, 2395-2417.

89. Verlet, L. Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967***, 159*, 98-103.

90. Leach, A. R. *Molecular Modelling: Principles and Applications;* Addison Wesley Longman: England, 1996.

91. Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005***, 26*, 1781-1802.

92. Rychaert, J.; Ciccotti, G.; Berendsen, H. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977***, 23*, 327-341.

93. Wang, W.; Donini, O.; Reyes, C.; Kollman, P. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001***, 30*, 211-243.

94. Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. A 2nd Generation Force-Field for the

Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995***, 117*, 5179-5197.

95. Nose, S. Constant-Temperature Molecular-Dynamics. *J. Phys. Condes. Matter* **1990***, 2*, SA115-SA119.

96. Hoover, W.; Ladd, A.; Moran, B. High-Strain-Rate Plastic-Flow Studied Via Non-Equilibrium Molecular-Dynamics. *Phys. Rev. Lett.* **1982***, 48*, 1818-1820.

97. Berendsen, H.; Postma, J.; Vangunsteren, W.; Dinola, A.; Haak, J. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984***, 81*, 3684-3690.

98. Andersen, H. C. Molecular-Dynamics Simulations at Constant Pressure And-Or Temperature. *J. Chem. Phys.* **1980***, 72*, 2384-2393.

99. Nose, S. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984***, 52*, 255-268.

100. Nosé, S. A Molecular-dynamics Method For Simulation in the Canonical Ensemble. *J. Mol. phys.* **1984***, 52*, 255-268.

101. Hoover, W. G. Canonical Dynamics - Equilibrium phase-space Distributions. *Phys. Rev. A* **1985***, 31*, 1695-1697.

102. Kubo, R.; Toda, M.; Hashitsume, N. *Statistical Physics II: Nonequilibrium Statistical Mechanics;* Springer: New York, 1991.

103. Adelman, S.; Doll, J. Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - General Formulation for Classical Scattering Off Harmonic Solids. *J. Chem. Phys.* **1976***, 64*, 2375-2388.

104. Brunger, A.; Brooks, C.; Karplus, M. Stochastic Boundary-Conditions for Molecular-Dynamics Simulations of St2 Water. *Chem. Phys. Lett.* **1984***, 105*, 495-500.

105. Kirkwood, J. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935***, 3*, 300-313.

106. Zwanzig, R. High-Temperature Equation of State by a Perturbation Method .1. Nonpolar Gases. *J. Chem. Phys.* **1954***, 22*, 1420-1426.

107. Glasstone, S.; Laidler, K.; Eyring, H. *The Thoery of Rate Processes;* McGraw-Hill: New York, 1941.

108. Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden Thermodynamics of Mutant Proteins - a Molecular-Dynamics Analysis. *Science* **1989***, 244*, 1069-1072.

109. Izrailev, S.; Stepaniants, S.; Schulten, K. Applications of steered molecular dynamics to protein-ligand/membrane binding. *Biophys. J.* **1998***, 74*, A177-A177.

110. Grubmuller, H.; Heymann, B.; Tavan, P. Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **1996***, 271*, 997-999.

111. ROUX, B. The Calculation of the Potential of Mean Force using Computer-Simulations. *Comput. Phys. Commun.* **1995***, 91*, 275-282.

112. Torrie, G.; Valleau, J. Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *J. Comput. Phys.* **1977***, 23*, 187-199.

113. Torrie, G.; Valleau, J. Monte-Carlo Free-Energy Estimates using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974***, 28*, 578-581.

114. Shen, J.; McCammon, J. Molecular-Dynamics Simulation of Superoxide Interacting with Superoxide-Dismutase. *Chem. Phys.* **1991***, 158*, 191-198.

115. Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. the Method. *J. Comput. Chem.* **1992***, 13*, 1011-1021.

116. Laidler, K.; King, M. The Development of Transition-State Theory. *J. Phys. Chem.* **1983***, 87*, 2657-2664.

117. Evans, A.; Polanyi, M. Some Aspects of the Chemistry of Macromolecules : Polymerisation of Isobutene by Friedel-Crafts Catalysts. *J. Chem. Soc.* **1947**, 252-257.

118. Pelzer, H.; Wigner, E. The speed constansts of the exchange reactions. *Z. Phys. Chem. B-Chem. Elem. Aufbau. Mater.* **1932***, 15*, 445-471.

119. Levine, I. N. *Physical Chemistry;* McGraw-Hill: New York, 2009.

Chapter 2

1. Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Feature Article Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2010***, 31*, 1569-1582.

2. Freiteg, S.; Chu, V.; Penzotti, J.; Klumb, L.; To, R.; Hyre, D.; Trong, I.; Lybrand, T.; Stenkamp, R.; Stayton, P. A structural snapshot of an intermediate on the streptavidin–biotin dissociation pathway. *Proc. Natl. Acad. Sci.* **1999***, 96*, 8384-8389.

3. Buckle, A.; Schreiber, G.; Fersht, A. Protein-protein recognition: Crystal structural analysis of barnase barstar complex at 2.0-angstrom resolution. *Biochemistry* **1994**, *33*, 8878-8889.

4. Radic, Z.; Kirchhoff, P.; Quinn, D.; McCammon, J.; Taylor, P. Electrostatic influence on the kinetics of ligand binding to acetylcholinesterase - Distinctions between active center ligands and fasciculin. *J. Biol. Chem.* **1997**, *272*, 23265-23277.

5. Sines, J.; Allison, S.; McCammon, J. Point-charge distribution and electrostatic steering in enzyme substrate encounter- Brownian dynamics of modified copper-zinc superoxide dismutases. *Biochemistry* **1990**, *29*, 9403-9412.

6. Case, D.; Karplus, M. Dynamics of ligand-binding to heme-proteins. *J. Mol. Biol.* **1979**, *132*, 343-368.

7. Hyre, D.; Amon, L.; Penzotti, J.; Le Trong, I.; Stenkamp, R.; Lybrand, T.; Stayton, P. Early mechanistic events in biotin dissociation from streptavidin. *Nat. Struct. Mol. Biol.* **2002**, *9*, 582-585.

8. Grubmuller, H.; Heymann, B.; Tavan, P. Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **1996**, *271*, 997-999.

9. Genest, D.; Garnier, N.; Arrault, A.; Marot, C.; Morin-Allory, L.; Genest, M. Ligand-escape pathways from the ligand-binding domain of PPARγ receptor as probed by molecular dynamics simulations. *Eur. Biophys. J.* **2008**, *37*, 369-379.

10. Gonzalez, A.; Perez-Acle, T.; Pardo, L.; Deupi, X. Molecular Basis of Ligand Dissociation in b-Adrenergic Receptors. *Plos One* **2011**, *6*, e23815.

11. Liu, L.; Bagal, D.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Hydrophobic Protein-Ligand Interactions Preserved in the Gas Phase. *J. Am. Chem. Soc.* **2009**, *131*, 15980-+.

12. Torrie, G.; Valleau, J. Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187-199.

13. Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. the Method. *J. Comput. Chem.* **1992**, *13*, 1011-1021.

14. Pelzer, H.; Wigner, E. The speed constansts of the exchange reactions. *Z. Phys. Chem. B-Chem. Elem. Aufbau. Mater.* **1932**, *15*, 445-471.

15. Laidler, K.; King, M. The Development of Transition-State Theory. *J. Phys. Chem.* **1983**, *87*, 2657-2664.

16. Evans, A.; Polanyi, M. Some Aspects of the Chemistry of Macromolecules - Held at the Institution-Of-Mechanical-Engineers, London, on April 4th, 1946 .1. Polymerisation of Isobutene by Friedel-Crafts Catalysts - Discussion. *J. Chem. Soc.* **1947**, 252-257.

17. Roux, B. The Calculation of the Potential of Mean Force using Computer-Simulations. *Comput. Phys. Commun.* **1995**, *91*, 275-282.

18. Truhlar, D.; Gao, J.; Alhambra, C.; Garcia-Viloca, M.; Corchado, J.; Sanchez, M.; Villa, J. The incorporation of quantum effects in enzyme kinetics modeling. *Acc. Chem. Res.* **2002**, *35*, 341-349.

19. Zhang, H.; Tan, T.; Feng, W.; van der Spoel, D. Molecular Recognition in Different Environments: beta-Cyclodextrin Dimer Formation in Organic Solvents. *J. Phys. Chem. B* **2012**, *116*, 12684-12693.

20. Bastug, T.; Chen, P.; Patra, S. M.; Kuyucak, S. Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. *J. Chem. Phys.* **2008**, *128*, 155104.

21. Bolhuis, P.; Dellago, C.; Chandler, D. Reaction coordinates of biomolecular isomerization. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5877-5882.

22. Schenter, G.; Garrett, B.; Truhlar, D. Generalized transition state theory in terms of the potential of mean force. *J. Chem. Phys.* **2003**, *119*, 5828-5833.

23. Liu, L.; Michelsen, K.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Energetics of Lipid Binding in a Hydrophobic Protein Cavity. *J. Am. Chem. Soc.* **2012**, *134*, 3054-3060.

24. Liu, L.; Bagal, D.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Hydrophobic Protein-Ligand Interactions Preserved in the Gas Phase. *J. Am. Chem. Soc.* **2009**, *131*, 15980-15981.

25. Woo, H.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825-6830.

26. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535-9551.

27. Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91*, 2798-2814.

28. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Chem. informatics* **2012**, *4*, 17.

29. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **2011**, *39*, W511-W517.

30. Dupradeau, F.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.ED. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821-7839.

31. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.;

Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A. J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09. *Gaussian Inc*. **2009**.

32. Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. A Well-Behaved Electrostatic Potential Based Method using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem*. **1993***, 97*, 10269-10280.

33. Wu, S.; Perez, M.; Puyol, P.; Sawyer, L. Beta-Lactoglobulin Binds Palmitate within its Central Cavity. *J. Biol. Chem*. **1999***, 274*, 170-174.

34. Wang, J.; Wolf, R.; Caldwell, J.; Kollman, P.; Case, D. Development and testing of a general amber force field. *J. Comput. Chem*. **2004***, 25*, 1157-1174.

35. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Case, D. A.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.;

Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. AMBER 12. **2012**.

36. Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781-1802.

37. Ryckaert, J.; Ciccotti, G.; Berendsen, H. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

38. Loncharich, R.; Brooks, B.; Pastor, R. Langevin Dynamics of Peptides - the Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers* **1992**, *32*, 523-535.

39. Truhlar, D.; Garrett, B. Multidimensional transition state theory and the validity of Grote-Hynes theory. *J Phys Chem B* **2000**, *104*, 1069-1072.

40. Luo, H.; Sharp, K. On the calculation of absolute macromolecular binding free energies. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 10399-10404.

41. Jorgensen, W. Interactions between Amides in Solution and the Thermodynamics of Weak Binding. *J. Am. Chem. Soc.* **1989**, *111*, 3770-3771.

42. Maiorov, V.; Crippen, G. Size-Independent Comparison of Protein 3-Dimensional Structures. *Proteins* **1995**, *22*, 273-283.

43. Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys.* **2006**, *125*, 084902.

44. Weiss, M.; Brandl, M.; Suhnel, J.; Pal, D.; Hilgenfeld, R. More hydrogen bonds for the (structural) biologist. *Trends Biochem. Sci.* **2001**, *26*, 521-523.

45. Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395-2417.

46. Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255-1273.

47. Konuma, T.; Sakurai, K.; Goto, Y. Promiscuous binding of ligands by beta-lactoglobulin involves hydrophobic interactions and plasticity. *J. Mol. Biol.* **2007**, *368*, 209-218.

48. Ragona, L.; Fogolari, F.; Zetta, L.; Perez, D.; Puyol, P.; De Kruif, K.; Lohr, F.; Ruterjans, H.; Molinari, H. Bovine beta-lactoglobulin: Interaction studies with palmitic acid. *Protein Sci.* **2000**, *9*, 1347-1356.

49. Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.***2006**, *2*, 1255-1273.

50. Mobley, D. L.; Chodera, J. D.; Dill, K. A. Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput.* **2007**, *3*, 1231-1235.

51. Kontopidis, G.; Holt, C.; Sawyer, L. Invited Review: beta-lactoglobulin: Binding properties, structure, and function. *J. Dairy Sci.* **2004**, *87*, 785-796.

52. Qin, B.; Bewley, M.; Creamer, L.; Baker, H.; Baker, E.; Jameson, G. Structural basis of the tanford transition of bovine beta-lactoglobulin. *Biochemistry (N. Y.)* **1998**, *37*, 14014-14023.

Chapter 3

1. Banks, R. E.; Smart, B. E.; Tatlow, J. C. *Organofluorine Chemistry: Principles and Commercial Applications*; Plenum Press: New York, **1994**.

2. Clader, J. W.: The Discovery of Ezetimibe a View from Outside the Receptor. *J. Med. Chem.* **2004**, *47*, 1-9.

3. Hugel, H. M.; Jackson, N. Special Feature Organo-Fluorine Chemical Science *Appl. Sci.* **2012**, *2*, 558-565.

4. Gunduz, M.; Argikar, U. A.; Kamel, A.; Colizza, K.; Bushee, J.; Cirello, A.; Lombardo, F.; Harriman, S. Oxidative *ipso* Substitution of 2,4-Difluoro-

benzylphthalazines: Identification of a Rare Stable Quinone Methide and Subsequent GSH Conjugate. *Drug Metab. Dispos.* **2012**, *40*, 2074-2080.

5. Vasdev, N.; Dorff, P. N.; O'Neil, J. P.; Chin, F. T.; Hanrahan, S.; Van Brocklin, H. F. Metabolic Stability of 6,7-dialkoxy-4-(2-,3-and 4-[^{18}F] fluoroanilino) quinazolines, Potential EGFR Imaging Probes. *Bioorg. Med Chem.* **2011**, *19*, 2959-2965.

6. Böhm, H.; Banner, D.; Bendels, S.; Kansy, M.; Kuhn, B.; Müller, K.; Obst-Sander, U.; Stahl, M. Fluorine in Medicinal Chemistry. *ChemBioChem.* **2004**, *5*, 637-643.

7. Kim, C.; Chang, J. S.; Doyon, J. B.; Baird, T. T. Jr.; Fierke, C. A.; Jain, A.; Christianson, D. W. Contribution of Fluorine to Protein-Ligand Affinity in the Binding of Fluoroaromatic Inhibitors to Carbonic Anhydrase II. *J. Am. Chem. Soc.* **2000**, *122*, 12125-12134.

8. Rendine, S.; Pieraccini, S.; Forni, A.; Sironi, M.: Halogen Bonding in Ligand–Receptor Systems in the Framework of Classical force fields. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19508-19516.

9. Benitex, Y.; Baranger, A. M. Control of the Stability of a Protein-RNA Complex by the Position of Fluorine in a Base Analogue. *J. Am. Chem. Soc.* **2011**, *133*, 3687-3689.

10. Lee, Y.; Zeng, H.; Ruedisser, S.; Gossert, A. D.; Hilty, C. Nuclear Magnetic Resonance of Hyperpolarized Fluorine for Characterization of Protein-Ligand Interactions. *J. Am. Chem. Soc.* **2012**, *134*, 17448-17451.

11. Zhou, P.; Zou, J.; Tian, F.; Shang, Z. Fluorine Bonding: How Does It Work In Protein-Ligand Interactions. *J. Chem. Inf. Model*. **2009**, *49*, 2344-2355.

12. Kuhn, B.; Kollman, P. A. A Ligand That Is Predicted to Bind Better to Avidin than Biotin: Insights from Computational Fluorine Scanning. *J. Am. Chem. Soc*. **2000**, *122*, 3909-3916.

13. Kawahara, S.; Tsuzuki, S.; Uchimaru, T. Theoretical Study of the C-F/π Interaction: Attractive Interaction between Fluorinated Alkane and an Electron-Deficient π–System. *J. Phys. Chem. A*. **2004***, 108*, 6744-6749.

14. Riley, K. E.; Merz, K. M. Effects of Fluorine Substitution on the Edge-to-Face Interaction of the Benzene Dimer. *J. Phys. Chem. B* . **2005***, 109,* 17752-17756.

15. Chopra, D.; Nagarajan, K.; Row, T. N. G. Analysis of Weak Interactions Involving Organic Fluorine: Insights from Packing Features in Substituted 4-keto-tetrahydroindoles, *J. Mol. Struct*. **2008**, *888*, 70-83.

16. Iwaoka, M.; Komatsu, H.; Katsuda, T.; Tomoda, S. Quantitative Evaluation of Weak Nonbonded Se···F Interactions and Their Remarkable Nature as Orbital Interactions. *J. Am. Chem. Soc.* **2002**, *124*, 1902-1909.

17. Matta, C. F.; Castillo, N.; Boyd, R. J. Characterization of a Closed-Shell Fluorine-Fluorine Bonding Interaction in Aromatic Compounds on the Basis of the Electron Density. *J. Phys. Chem. A.* **2005**, *109*, 3669-3681.

18. Lee, S.; Mallik, A. B.; Fredrickson, D. C. Dipolar-Dipolar Interactions and the Crystal Packing of Nitriles, Ketones, Aldehydes, and C(sp^2)-F Groups. *Cryst. Growth Des*. **2004**, *4,* 279-290.

19. Olsen, J. A.; Banner, D. W.; Seiler, P.; Sander, U. O.; D'Arcy, A.; Stihle, M.; Müller, K.; Diederich, F. A Fluorine Scan of Thrombin Inhibitors to Map the Fluorophilicity/Fluorophobicity of an EnzymeActive Site: Evidence for C-F···C=O Interactions. *Angew. Chem. Int.Ed.* **2003**, *42,* 2507-2511.

20. Plenio, H.: The Coordination Chemistry of Fluorine in Fluorocarbons. *ChemBioChem.* **2004***, 5*, 650-655.

21. Bettinger, H. F. How Good is Fluorine as a Hydrogen-Bond Acceptor in Fluorinated Single-Walled Carbon Nanotubes? *ChemPhysChem*. **2005**, *6,* 1169-1174.

22. Müller, K.;Faeh, C.; Diederich, F. Fluorine in Pharmaceuticals: Looking Beyond Intuition. *Science* **2007**, *317*, 1881-1886.

23. Mecinović, J.; Snyder, P. W.; MIRICA, K. A; Bai, S.; Mack, E. T.; Kwant, R. L.; Moustakas, D. T.; Héroux, A.; Whitesides, G. M. Fluoroalkyl and Alkyl Chains Have Similar Hydrophobicities in Binding to the "Hydrophobic Wall" of Carbonic Anhydrase. *J. Am. Chem. Soc.* **2011***, 133*, 14017-14026.

24. Kontopidis, G.; Holt, C.; Sawyer, L. Beta-Lactoglobulin: Binding Properties, Structure, and Function. *J. Dairy Sci*. **2004**, *87*, 785-796.

25. Qin, B.Y.; Bewley, M.C.; Creamer, L.K.; Baker, H.M.; Baker, E.N.; Jameson, G.B. Structural Basis of the Tanford Transition of Bovine Beta-Lactoglobulin. *Biochem.* **1998**, *37*, 14014-14023.

26. Dunbar, R.C., McMahon, T.B. Activation of Unimolecular Reactions by Ambient Blackbody Radiation. *Science* **1998**, *279*, 194-197.

27. Price, W.D.; Schnier, P.D.; Jockusch, R.A.; Strittmatter, E.R.; Williams, E.R. Unimolecular Reactions Kinetics in the High-Pressure Limit Without Collisions. *J. Am. Chem. Soc.* **1996**, *118*, 10640-10644.

28. Liu, L.; Bagal, D.;Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Hydrophobic Protein-Ligand Interactions Preserved in the Gas Phase. *J. Am. Chem. Soc.* **2009**, *131*, 15980-15981.

29. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Chem. Inform.* **2012**, *4*, 17-22.

30. Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F. Y. R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.*, **2011**, *39*, W511−W517.

31. Dupradeau, F.-Y.;Pigache, A.; Zaffran, T.; Savineau, C.;Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak P. The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Feld Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821−7839.

32. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.;Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima,T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.;

Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09. *Gaussian Inc*. **2009**.

33. Wu, S.Y.; Perez, M. D.; Puyol, P.; Sawyer, L. β-Lactoglobulin Binds Palmitate within its Central Cavity. *J. Biol. Chem.* **1999**, *274*, 170-174.

34. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. -J.; Cui, G.; Roe D. R.; Mathews D. H.; Seetin M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin V.; Luchko, T.; Gusarov, S.; Kovalenko A.; Kollman, P. A. AMBER 12. **2012**.

35. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD, *J. Comput. Chem.* **2005**, *26*, 1781-1802.

36. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development of Testing of a General Amber Force Field, *J. Comput. Chem.* **2004***, 25*, 1157-1174.

37. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes, *J. Comput. Phys.* **1977**, *23*, 327-341.

38. Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics, *J. Molec. Graphics* **1996***, 14.1*, 33-38.

39. *Origin;* OriginLab, Northampton, MA 01060 USA.

40. Sun, J.; Kitova, E.N.; Klassen, J.S. Method for Stabilizing Protein-Ligand Complexes in Nnanoelectrospray Ionization Mass Spectrometry. *Anal. Chem*. **2007**, *79*, 416-425.

41. Bagal, D.; Kitova, E.N.; Liu, L.; El-Haweit, A.; Schnier, P. D.; Klassen, J.S. Gas Phase Stabilization of Noncovalent Protein Complexes Formed by Electrospray Ionization. *Anal. Chem*. **2009**, *81*, 7801-7806.

42. Liu, L.; Michelsen, K.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Energetics of Lipid Binding in a Hydrophobic Protein Cavity. *J. Am. Chem. Soc.* **2012**, *134*, 3054-3060.