MOLECULAR SIMULATIONS OF TRANSPORT AND SEPARATION IN PROTEIN CRYSTALS

HU ZHONGQIAO

NATIONAL UNIVERSITY OF SINGAPORE

2009

MOLECULAR SIMULATIONS OF TRANSPORT AND SEPARATION IN PROTEIN CRYSTALS

HU ZHONGQIAO

(***B. Eng. & M. Eng., Tsinghua University***)

A THESIS SUBMITTED

FOR THE DEGREE OF PhD

DEPARTMENT OF CHEMICAL AND BIOMOLECULAR ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2009

Acknowledgments

First and foremost, I would like to extend my greatest appreciation to my supervisor, Prof. Jiang Jianwen, for his guidance and encouragement throughout the course of my PhD program. His critical suggestions and ideas have helped me substantially in completing my PhD research. I sincerely treasure this wonderful experience and I strongly believe that the advice and lessons would be very valuable for my future undertaking to a significant extent.

I would like to express my deep gratitude to all the group members: Mr. Babarao Ravichandar, Dr. Li Jianguo, Dr. Yin Jian, Mr. Anjaiah Nalaparaju, Ms. Gnanasambandam Sivashangari, Mr. Ramakrishnan Vigneshwar, Ms. Liang Jianchao, Dr. Fan Yanping, Ms. Chen Yifei, Dr. Zhang Liling, Dr. Luo Zhonglin, Mr. Fang Weijie, Mr. Zhuo Shengchi, Mr. Liu Yu for valuable discussions and comments. I have enjoyed much pleasure shared with them.

I wish to express special thanks to Mr. Ifan for his technical support on the installation, use and upgrade of Gromacs package and to Mr. Zhang Xinhuai for his help for using clusters at SVU in the early stage of my research.

Finally, I want to thank my wife Ms. Huang Haiying for her patient love and understanding. Without her encouragement and support, this work could not have been completed successfully.

Table of Contents

Summary

As novel bionanoporous materials, protein crystals have demonstrated increasing potentials in a wide variety of applications such as bioseparation, biocatalysis and biosensing. Deep insight into the transport properties and separation mechanisms in protein crystals is crucial to better exploring their emerging applications. Toward this end, molecular dynamics (MD) simulations are employed in this thesis to investigate transport and separation in different protein crystals.

The structural and dynamic properties of water and ions are studied systematically in protein crystals with various topologies and morphologies. The solvent-accessible surface area per residue is found to be nearly identical in different protein crystals. Water and ions exhibit layered structures on protein surface. Diffusivities in protein crystals are reduced by one - two orders of magnitude than in bulk phase. The mobility in the crystals is enhanced with increasing porosity. Anisotropic diffusion is found preferentially along the pore axis, as experimentally observed.

Electrophoresis of ion mixture in a lysozyme crystal is investigated. Upon exposure to electric field, the stability of protein is found to reduce slightly. Water molecules tend to align preferentially parallel to the electric field, and the dipole moment along the pore axis rises linearly with increasing field strength. Electric field has a marginal effect on the structures of water and ions. Electrical current exhibits a linear relationship with the field strength. Equilibrium and non-equilibrium MD simulations give consistent electrical conductivity in the crystal.

Separation of amino acids (Arg, Phe and Trp) in a liquid chromatography is investigated using glucose isomerase crystal as the stationary phase. The elution order is Arg > Phe > Trp and consistent with experiment. Arg is highly hydrophilic and charged, interacts with water the most strongly, and thus moves with flowing water the fastest. Trp has the largest van der Waals volume and encounters the largest steric hindrance, leading to the slowest velocity. The solvent-accessible surface areas of amino acids and the numbers of hydrogen bonds further elucidate the observed velocity difference.

Chiral separation of racemic *D/L*-phenylglycines in thermolysin crystal is examined. *D*-phenylglycine is observed to transport slower than *L*-phenylglycine, in accord with experimental elution order. From energetic and structural analysis, it is found that *D*-phenylglycine interacts more strongly with thermolysin than *L*-phenylglycine; consequently, it stays more proximally to thermolysin for a longer time. The chiral discrimination of *D/L*-phenylglycines is attributed to the collective contribution from the chiral centers of thermolysin residues.

Three biomolecular force fields (OPLS-AA, AMBER03 and GROMOS96) in conjunction with three water models (SPC, SPC/E and TIP3P) are assessed for the transport of water and ions in a lysozyme crystal. All the three force fields predict similar pattern in *B*-factors, whereas OPLS-AA and AMBER03 accurately reproduce experimental measurements. Water diffusivities from OPLS-AA and AMBER03 along with SPC/E model match fairly well with experimental data. A combination of OPLS-AA for lysozyme and Kirkwood-Buff model for NaCl is superior to others in predicting ion mobility.

List of Tables

List of Figures

List of Abbreviations

AMBER	Assisted Model Building with Energy Refinement
BD	Brownian Dynamics
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CC	Carbonyl Carbon
CLECs	Cross-Linked Enzyme Crystals
CLPCs	Cross-Linked Protein Crystals
DSSP	Database of Secondary Structure Assignments
EMD	Equilibrium Molecular Dynamics
GI	Glucose Isomerase
GROMOS	GROningen Molecular Simulation
HW	Hydrogen atom of Water molecule
KB	Kirkwood-Buff
LJ	Lennard-Jones
MC	Monte Carlo
MD	Molecular Dynamics
MOFs	Metal-Organic Frameworks
MSD	Mean-Squared Displacement
NEMD	Non-Equilibrium Molecular Dynamics
NPR	Non-Position Restraint
OPLS	Optimized Potentials for Liquid Simulations
OPLS-AA	OPLS All-Atom
OW	Oxygen of Water

PDB	Protein Data Bank
PME	Particle-Mesh-Ewald
PNP	Poisson-Nernst-Planck
PR	Position Restraint
RDF	Radial Distribution Function
RMSF	Root-Mean-Squared Fluctuation
RMSD	Root-Mean-Squared Deviation
SASA	Solvent-Accessible Surface Area
SPC	Simple Point Charge
SPC/E	Simple Point Charge/Extended
TIP3P	Transferable Intermolecular Potential 3 Points

Chapter 1 Introduction

1.1 Protein Crystals

1.1.1 Features

Protein crystals have demonstrated increasing potentials as novel separation materials and biocatalysts.[1,2] The pore sizes, surface areas, and porosities of protein crystals can vary with the nature of proteins and crystallization conditions, and thus offer a wide variety of porous structures. **Figure 1.1** shows six protein crystals with different pore sizes and morphologies. It is also known that even one protein can form various polymorphic crystals depending upon additive, pH, and temperature. For instance, lysozyme exists at least in four crystal forms: tetragonal, orthorhombic, monoclinic, and triclinic,[3] three of which are shown in **Figure 1.2**. Therefore, protein crystals have an immense diversity compared to other porous materials. **Table 1.1** compares the pore features between protein crystals and zeolites.[1]

Table 1.1. Comparison between protein crystals and zeolites.

	Pore size (nm)	**Porosity**	**Pore volume (ml/g)**	**Pore surface areas (m^2/g)**
Protein crystals	1.5–10	0.5–0.8	0.9–3.6	800–2000
Zeolites	0.2–1.0	0.3–0.5	0.2–0.4	200–500

Another salient advantage of protein crystals over other porous materials is the inherently chiral nature of protein molecules. The *L*-amino acids as building blocks of proteins create an asymmetric environment, which could lead to selective separation of enantiomers using protein crystals. In addition, as proteins are weak ion exchangers

at isoelectric points from 2 to 12, one can easily manipulate the binding of small molecules by changing the pH or buffer content in eluent.[1]

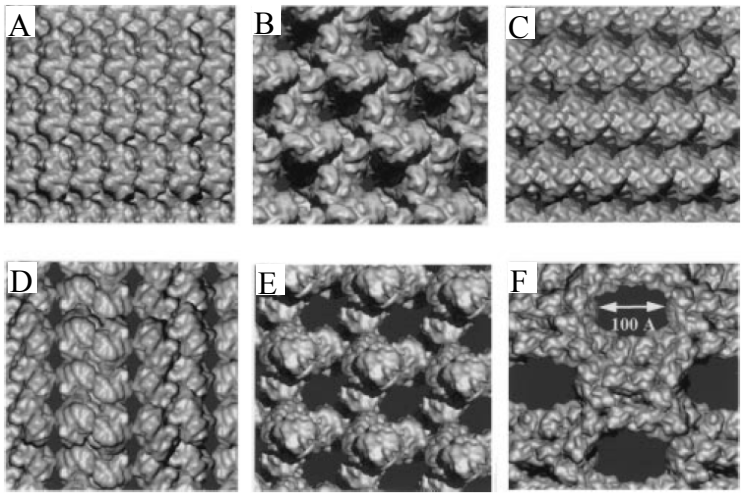

Figure 1.1. Six protein crystal lattices with different pore sizes.[1] The pore sizes have been scaled. (A) Superoxide dismutase (1XSO); (B) Themolysin (8TLN); (C) Penicillin acylase (1PNL); (D) Candida rugosa Lipase (1CRL); (E) Carboxypeptidase (1WHS); (F) Superoxide dismutase (1SOS).

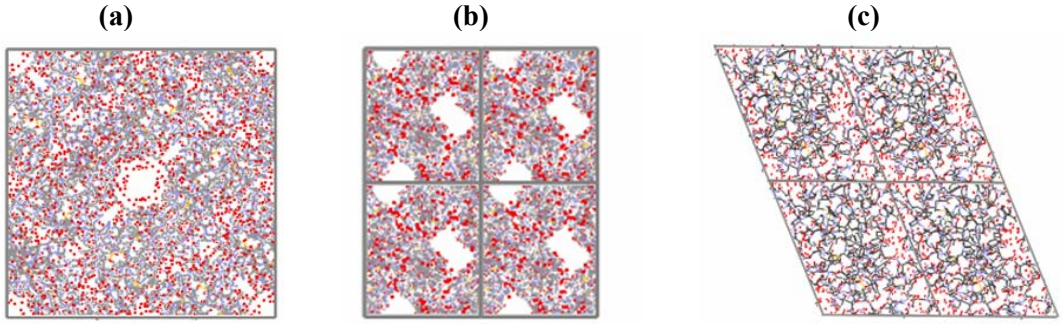

Figure 1.2. Three lysozyme crystals: (a) tetragonal, (b) orthorhombic, and (c) triclinic. Lysozyme is in gray and crystallographic water is in red.

1.1.2 Stability and Production

In the past, the applications of porous protein crystals were limited by their fragility and instability under unfavorable conditions.[1] Compared to the strong covalent bonds within zeolites, protein molecules are virtually located at lattice sites via weaker noncovalent van der Waals and electrostatic interactions. As a result, protein crystals are mechanically soft and can easily disintegrate or dissolve in unfavorable

environments.[1,4] Nowadays this limitation has been largely removed by cross-linking technology. Cross-linked protein crystals (CLPCs) can be created by protein crystallization and followed by chemical cross-linking. The strong intermolecular chemical bonds by cross-linking constrain crystalline protein molecules, leading to the formation of an insoluble, mechanically robust, and microporous protein matrix.[1] A cross-linked glucose isomerase (GI) crystal was packed in chromatographic column at 30 MPa to separate *n*-alcohols or amino acids,[5] and thermolysin and human serum albumin (HSA) crystals at 10 MPa to separate the *D*/*L*-phenylglycine.[1] Both studies demonstrated the mechanical rigidity of CLPCs. Cross-linking can also enhance the chemical stability of proteins in crystalline form. A single column packed with thermolysin or HSA crystal was used for more than 500 injection cycles without any loss of separation efficiency.[1] Cross-linked crystals are resistant to digestion by proteolytic enzymes.[6,7] The cross-linked GI crystal is about five times more stable than the native form in high substrate solution.[8] In addition, cross-linked enzyme crystals (CLECs) are more convenient than aqueous enzymes to be separated from reaction broth for reuse. The most commonly used cross-linker is glutaraldehyde, a bifunctional aldehyde. In most cases, only lysine residues react with glutaraldehyde during cross-linking.[9,10] It is worthwhile to note that chemical cross-linking, while inducing slight change in pore structures as demonstrated by X-ray diffraction patterns, does not substantially influence the properties of protein crystals as microporous materials and catalysts.[1,11,12] For instance, crystallographic studies showed that glutaraldehyde cross-linking reaction had only minor effect on lysozyme structure.[13] Chemical stable CLECs were first developed in the 1960's.[14] Several techniques have been recently reported to effectively produce CLPCs or CLECs with high yield and good quality or on a large scale.[15-21] Falkner et al. proposed a method

to economically produce size-tunable submicrometer CLPCs (hen egg white lysozyme) with a good reproducibility and good quality on a large scale.[20] However, the development of cross-linking technology is still ongoing and the number of proteins commercially available in cross-linked form is currently limited.

1.1.3 Applications

The collective features mentioned above have allowed protein crystals to be used as a novel class of molecular sieves[1] for analytical or preparative separation and as catalysts[2] for the synthesis of fine chemicals, chiral intermediates, and peptides in laboratory or on a commercial scale. Major applications of protein crystals are briefly summarized below.

CLPCs can be used as stationary phase in liquid chromatography for chiral and achiral separation. Vilenchik et al. showed that cross-linked thermolysin crystals separated several mixtures.[1] Cross-linked GI crystals were used to separate racemic mixture of *D*/*L*-arabitol, or *D*/*L* pairs of some amino acids.[5] Pastinen et al. showed that cross-linked GI crystals could separate mixtures of amino acids, *n*-alcohols from C_1 to C_8 or nucleosides (e.g., uridine, cytidine, adenosine, and guanosine).[5,22]

Protein crystals can also be used as biocatalysts, biosensors, drug delivery carriers and biotemplates etc. Compared to aqueous or immobilized enzymes traditionally used in biocatalysis, CLECs have higher stability and better operational performance in biocatalysis. Many enzymes including subtilisin protease, candida rugosa lipase, alcohol dehydrogenase, glucose isomerase and penicillin acylase, have been successfully crystallized and used to catalyze important bioreactions. Enzymes are immobilized on solid supports and used as biosensors. Alternatively, CLECs could be directly used as biosensors. The biosensors of enzymes in crystal form exhibited greater sensitivity and higher operational stability with a lower limit of detection

compared to those in non-crystal form because crystalline enzymes have higher specific volumetric activity and higher stability against the changes of external environments (e.g., pH, temperature and solvent).[23-25] Protein crystals can also be used as biotemplates for the fabrication of nano-structural composite materials.[26-29] These important applications are summarized in **Table 1.2**.

Table 1.2. Applications (excluding separation) of proteins crystals.

Proteins	Applications	Refs
Biocatalysts		
Subtilisin protease	α-methyltryptamine was resolved.	30
Candida rugosa lipase	Ketoprophen was resolved.	31
Alcohol dehydrogenase	Cinnamaldehyde was reducted.	32
Glucose isomerase	High-fructose corn syrup was produced.	8
Hydroxynitrile lyase	Cyanohydrins were synthesized.	33
Penicillin acylase	hydrolysis of penicillin G.	2
Biosensors		
Glucose oxidase	Biosensor towards hydrogen peroxide and glucose.	25
Laccase	Biosensor towards phenols like 2-amino phenol, guaiacol, catechol and pyrogallol etc.	23
Organophosphate hydrolase	Electrochemical biosensors for the detection of organophosphate pesticides.	24
Lysozyme	Scanned probes in recognition atomic force microscopy for detection.	34
Drug delivery carrier		
Human serum albumin	Vaccine delivery.	35
Biotemplates		
Lysozyme	Template for the fabrication of a nano-structural protein-synthetic-hydrogel hybrid.	28

1.2 Molecular Simulations

With the continually growing computational power, molecular simulations have been playing an increasingly important role in material and life sciences. Simulations at the molecular scale can provide microscopic pictures that are experimentally intractable or difficult to obtain, if not impossible. Consequently, fundamental insight gained from molecular simulations can assist in the rational design of new materials and the optimization of engineering processes.

Nowadays molecular simulations are more tightly coupled with experiments than ever and have been used to interpret experimental data, trigger new experiments, or even substitute experiments.[36] For instance, simulations can resolve the contradictions between experimental NMR and X-ray data, and provide quantitative interpretation for experimentally observed stability difference between protein mutants. In the study of water and ion channels across biomembranes, simulations are also very useful to complement experiments. On the basis of simple polypeptide models, simulation studies found that denatured or unfolded polypeptides are composed of less relevant conformations (about 10^2–10^3) than expected (10^8), which triggered experimentalists to find a new methodology to characterize the small unfolded conformations in terms of residual structure.[36] Interestingly, simulation first discovered that water transports through carbon nanotube with widths of a few nanometers at a much greater rate than expected,[37] and this finding was subsequently confirmed by experiment.[38] Simulations can also substitute labor-intensive and/or high-capital-cost experiments. After a set of simulation methods for certain properties are established with acceptable accuracy, simulations can be carried out to predict the properties that would be formidable or expensive to perform experimentally or raise environmental or safety issues.

A few widely used molecular simulation methods including molecular dynamics (MD), Monte Carlo (MC), Brownian dynamics (BD) are briefly introduced.

1.2.1 Molecular Dynamics Simulation

MD simulation was first used in the 1950's to simulate simple fluids. The main principle of MD simulation is as follows. Given the state $S(t_0)$ of a system with N particles, that is, the position $\mathbf{r}_i$ and velocity $\mathbf{v}_i$ (i = 1, …, N) at time t_0, the subsequent states $S(t_0 + \Delta t)$, $S(t_0 + 2\Delta t)$, … are calculated using Newton's equation of motion

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, \quad i = 1, 2, \cdots, N \tag{1.1}$$

where m_i is the mass of particle i, and $\mathbf{F}_i$ is the force exerted on particle i. $\mathbf{F}_i$ is the negative derivative of a potential function $\mathbf{V}\left(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N\right)$

$$\mathbf{F}_i = -\frac{\partial \mathbf{V}}{\partial \mathbf{r}_i} \tag{1.2}$$

These equations are solved numerically using a small time step t_0, usually in femtosecond scale, for energy conservation.

It should be noted that MD and other simulation methodologies are largely dependant on the availability of a suitable potential function in Eq. (1.2) to describe the energy landscape of the system.

1.2.2 Monte Carlo Simulation

MC simulation is a stochastic method to generate a set of representative configurations at given conditions (statistical ensembles) such as temperature, volume, pressure, or chemical potential. One attractive aspect of the conventional MC simulation is that only potential energy rather than force is evaluated in sampling configurations, leading to a very efficient calculation. Nevertheless, some biased MC

methods can evaluate force. MC can perform physically unnatural motions, e.g. a jump from one position to the other or insertion/deletion of a new molecule, and thus significantly increase efficiency. Depending on the system of interest, various types of trial moves can be attempted to lead the system to equilibration. Thereafter, ensemble properties are statistically averaged.

1.2.3 Brownian Dynamics Simulation

For systems with a large amount of solvent molecules, e.g. dilute colloidal and protein solutions, the interest is usually in the solute rather than the solvent; consequently, the effect of solvent can be considered implicitly. BD is such a simplified method to smear out the solvent molecules. In BD, the motion of particle *i* is governed by the Langevin equation

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -m_i\gamma_i \frac{d\mathbf{r}_i}{dt} + \mathbf{F}_i^{\mathrm{R}}(\mathbf{r}) + q_i\mathbf{E}_i + \mathbf{F}_i^{\mathrm{S}},\ i = 1, 2, \cdots, N \qquad (1.3)$$

where m_i, γ_i and q_i are the mass, friction coefficient and charge of particle *i*, while $\mathbf{F}_i^{\mathrm{R}}$, $\mathbf{E}_i$ and $\mathbf{F}_i^{\mathrm{S}}$ are the random stochastic force, electric field, and short range force, respectively, experienced by particle *i*. The use of BD simulation allows one to gain insight at larger time and length scales.

1.2.4 Technical Issues

The simulation methods mentioned above are in principle very simple, but many crucial technical issues as discussed below have to be implemented during simulations.[39]

Periodic boundary conditions

Because of the limitation of CPU power, in most molecular simulations 10^3–10^6 particles are typically involved. The simulation system with such a small size suffers

from finite-size effect. For example, because of the surface tension, the pressure in a spherical droplet of water consisting of 2×10^4 molecules is approximately 275 bar.[39] Therefore, most simulations are performed with periodic boundary conditions, in which a simulation box is surrounded by an infinite number of identical replica boxes. The particles in the central box and their images in replica boxes behave in the same manner, and can freely cross box boundaries. When a particle leaves the box, its image from the adjacent box will enter from the opposite side.

Constraint dynamics

In MD simulations, the allowed maximal time step is dictated by the highest frequency in the system. In a biomolecular system, the vibrations of bonds involving H atoms have the highest frequency. For example, the bond stretching frequency of O-H bond is typically 10^{14} Hz with the average period of 10 fs. This implies that the time step in simulation should be about 0.5 fs based on a rule of thumb, which states that samples are needed to be taken at least 20 times per period for a reasonable sampling of a periodic function. In many situations, however, the bond with a high vibrational frequency is very rigid and the bond length is almost unvaried. Furthermore, the bonded interaction is generally not part of the physics of interest in MD simulation because the bond vibration is practically uncoupled to the other vibrations, indicating that it does not play a significant role in the dynamics of the system. Thus, in most MD simulations, chemical bonds (at least H-involving bonds) are handled using constraint dynamics to keep constant bond lengths. Then the time step may be increased to 2 fs, that is, the speedup factor is about 4. However, constraint algorithms should be implemented effectively to prevent time-consuming computations. Currently there exists a fast iterative method, called LINCS, to solve this problem.[40]

Particle-Mesh Ewald

In molecular simulations, the van der Waals interactions are short-ranged and thus can be cut off directly beyond a short distance. However, electrostatic terms are inversely proportional to the distances of charged particles and thus are long-ranged. The evaluation of electrostatic interactions based on the Coulomb's law is very time consuming, because the required time is proportional to N^2 (N is the number of particles). The most straightforward simplification is cut-off method with an order of N, but significant artifacts are introduced. Several alternative methods have been developed in the literature, including Ewald methods with an order of $N^{3/2}$ which is not well suited for large biomolecular systems, particle-particle particle-mesh (PPPM) and particle-mesh Ewald (PME)[41,42] that scale with $N\log(N)$. PME is used for all simulations in this thesis.

1.2.5 Force Fields

In molecular simulations, the selection of potential function $\mathbf{V}\left(\mathbf{r}_1,\mathbf{r}_2,\cdots,\mathbf{r}_N\right)$ in Eq. (1.2) and its parameterization are one of major concerns. A set of suitable potential functions and precise parameters, which are referred to as force field, is crucial to the accuracy of simulations.

For an additive force field, the potential function is generally decomposed into the bonded term $\mathbf{V}_{\text{bonded}}$ for covalently bonded atoms and the nonbonded term $\mathbf{V}_{\text{nonbonded}}$ for electrostatic and van der Waals interactions

$$\mathbf{V} = \mathbf{V}_{\text{bonded}} + \mathbf{V}_{\text{nonbonded}} \tag{1.4}$$

where $\mathbf{V}_{\text{bonded}}$ and $\mathbf{V}_{\text{nonbonded}}$ are further decomposed into several terms as follows

$$\mathbf{V}_{\text{bonded}} = \mathbf{V}_{\text{bond}} + \mathbf{V}_{\text{angle}} + \mathbf{V}_{\text{dihedral}} \tag{1.5}$$

$$\mathbf{V}_{\text{nonbonded}} = \mathbf{V}_{\text{electrostatic}} + \mathbf{V}_{\text{vanderWaals}} \tag{1.6}$$

For common force fields, the nonbonded terms only include two-body terms and multi-body interactions are excluded for computational efficiency. The van der Waals potential $\mathbf{V}_{\text{vanderWaals}}$ is generally described by 12-6 Lennard-Jones (LJ) potential

$$\mathbf{V}_{\text{LJ}} = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \tag{1.7}$$

Force field parameters are usually derived from both experimental data and high-level quantum mechanical calculations. For biomolecules the most commonly used force fields include CHARMM (Chemistry at HARvard Macromolecular Mechanics),[43] AMBER (Assisted Model Building with Energy Refinement),[44,45] OPLS (Optimized Potentials for Liquid Simulations),[46,47] and GROMOS (GROningen MOlecular Simulation).[48] Among them the first three belong to high-resolution fully atomistic force fields that provide parameters for all atoms (including hydrogens), while GROMOS is a united atom force field that incorporates a carbon atom and all its neighbored hydrogen atoms into a single united particle. Recently, several coarse-grained (CG) force fields with low resolution have been developed and are increasingly used in protein simulations to enhance computational efficiency.[49] Along with the development of biomolecular force fields, commercial and free software packages are also released, including AMBER, CHARMM, GROMOS, and GROMACS (GROningen Machine for Chemical Simulations).[50,51]

1.3 Literature Review

Due to the unique characteristics of protein crystals, a large number of studies have been conducted, particularly by experiments, to investigate the properties of guest molecules confined in protein crystals. Here we review recent advances in transport and separation in protein crystals, which are the central topics of this thesis.

1.3.1 Experimental Studies

Experiments have been extensively conducted to investigate the dynamic properties of guest molecules and separation processes in protein crystals, as listed in **Table 1.3**. A few earlier studies reported the diffusion of solvent and ions in protein crystals and found that the diffusion coefficients of water and ions were reduced by 1–2 orders of magnitudes than in bulk phase.[52-54] Morozov et al. found that the diffusion coefficient of intracrystalline water in a tetragonal lysozyme crystal was reduced by about 30–40% compared with that in bulk phase.[52] Morozova et al. measured the conductivity and transference number of ions in a tetragonal lysozyme crystal and further calculated the mobility of ions. The mobility of cations was 4–50 fold lower and the mobility of anions 100–300 fold lower than in bulk.[53] Bon et al. found that diffusion coefficients of water in the triclinic lysozyme crystal were reduced by 5–50 times than in bulk phase and that water molecules exhibited ordered structure in the hydration shell near to protein surface.[54] Subsequently, several studies were reported on the diffusion of intermediate-sized inorganic and organic molecules such as surfactants and dyes in protein crystals. The adsorption and diffusion of solutes within different lysozyme crystals were experimentally examined in detail.[55-59] Transport of dyes in crystals and adsorption capacities of the crystals were found to depend on solute type, crystal morphology, and solution characteristics (e.g. pH). The results indicated the potentially interesting ability of protein crystals to concentrate, collect, and store solutes from a surrounding solution. From these studies, it was concluded that the dominant factors influencing transport of guest molecules confined in protein crystals are steric repulsion, cross-linker, and electrostatic interaction.[52,53,55-57]

Table 1.3. Experimental studies on transport in protein crystals.

Guests	Proteins	Experiments	Refs
Water	**Lysozyme**	Water mobility was measured and several factors were proposed to explain the discrepancy between experimental and theoretical water diffusion coefficients.	52
	Lysozyme	The diffusion coefficient of water was reduced by 5−50 times compared with in bulk phase.	54
	Lysozyme	Ion mobility in crystal was obviously lower than in solution. The steric hindrance and charges in both ion and protein were responsible for this reduction.	53
Salts	***β*-lactoglobulin**	The diffusion coefficients of several salts like KBr were measured.	60
	Glycogen phosphorylase b	Diffusion of salts lilke LiBr-uridylate was studied.	61
Organic molecules	**Lysozyme**	The diffusion coefficients of lysozyme adjacent to the lysozyme crystal surface were measured.	62
	Lysozyme	The diffusion coefficients of surfactants were measured. A strong adsorption of surfactants to crystal lattice lowered the infusion into crystal.	63
	Lysozyme	The adsorption and transport of dyes in four lysozyme crystals (e.g., tetragonal, orthorhombic, monoclinic and triclinic) were studied. Anisotropic diffusion was found and modeled.	55-59

Separation of mixtures in protein crystals has also been investigated including racemic separation.[1,5,64] In chiral and affinity separation, for instance, proteins such as lysozyme, bovine/human serum albumin and glycoproteins are immobilized on solid supports.[65] However, the utilization of support matrix results in a low volumetric specific activity of proteins and thus decreases the separation efficacy. Separation efficacy can be largely improved if crystalline proteins are directly packed in chromatographic column.[5] Furthermore, the compact arrangement of protein molecules in crystalline phase inhibits protein unfolding and thus maintains their

native conformations more effectively, especially at elevated temperatures or in organic solvents as compared to amorphous proteins.[1,2] **Table 1.4** summarizes recent experimental studies on separation in protein crystals.

Table 1.4. Experimental studies on separation in protein crystals.

Proteins	Mixtures	Refs
Glucose isomerase	Four nucleosides	22
	Amino acids	5
	n-alcohols from C_1 to C_8	
	D/*L*-arabitol	
Thermolysin	PEG molecules with different sizes	1
	R/*S*-phenylglycines	
	R/*S*-phenyllactic acids	
	S-ibuprofen and *R*-phenyllactic acid	
Human serum albumin	*R*/*S*-folinic acid	1

Glucose isomerase (GI) is the first enzyme crystallized on an industrial scale. Cross-linked GI crystal successfully separated a mixture of four nucleosides (e.g., uridine, cytidine, adenosine, and guanosine).[22] GI crystal has capability to separate *n*-alcohols from C_1 to C_8 and the mixture of different amino acids, based on differential hydrophobic interactions of solutes with protein.[5] GI crystal also showed a strong chiral separation ability for racemic *D*/*L*-arabitol and a weak chiral discrimination ability for *D*/*L* pairs of some amino acids.[5] Besides GI crystals, cross-linked thermlysin crystal is another important separation material. Thermolysin crystal effectively separated a mixture of ibuprofen and phenyllactic acid, racemic *R*/*S*-phenylglycines, or *R*/*S*-phenyllactic acids, indicating thermlysin is a good chiral selector.[1] Interestingly, Vilenchik et al. showed that cross-linked HSA crystals gave good chiral separation, in contrast to the cross-linked precipitate of HSA. The result suggests that crystallinity is needed in this separation process.[1] Different separation mechanisms were proposed, including size exclusion, adsorption, charge,

hydrophobicity and chirality.[1,5] Various factors such as eluent pH, type and concentration of organic and charged modifiers, ionic strength and temperature were identified to affect the retentivity and enantioselectivity of solutes and required to be optimized.

Experimental studies on protein crystals as separation materials are still much fewer compared to those on inorganic or organic materials such as zeolites or metal-organic frameworks (MOFs), largely because of the long-held prejudice that protein crystals are not stable. Another reason is that the production processes of protein crystals are more difficult compared to those of zeolites or MOFs. Due to the diversity and complexity of protein crystal structures, systematic experimental studies are desired in order to establish semi-quantitative or even quantitative relationships that can effectively describe the transport and separation in protein crystals.

1.3.2 Simulation Studies

Most simulation studies on proteins are focused on the conformational change (folding and unfolding) of a single protein molecule in solution. Simulation studies on protein crystals are relatively rare. A few earlier simulations primarily examined the conformations of protein in crystal form.[66-76] It was demonstrated that lysozyme structures in an orthorhombic crystal and aqueous solution were very similar with respect to crystallographic *B*-factors, NOE atom-atom distance bounds, $^{3}J_{HN\alpha}$ coupling constants and ^{1}H-^{15}N bond vector order parameters, but crystalline structure reproduced X-ray NMR data slightly better than in solution.[74] The characteristics of two Aib-rich peptides in crystal and solution states were examined; one peptide exhibited very similar conformations in the two states, while the other exhibited much narrower conformational distribution in crystal.[76] Recently a streptavidin-biotin complex was simulated in both crystal and solution.[77,78] Although the mobility of

protein molecules was comparable in both states, the initial X-ray structure was better maintained in crystal. From these simulations, it was concluded that protein conformations in crystalline form could be slightly or distinctly different from those in aqueous state. Therefore, simulations in crystallization conditions can better validate potential function parameters if experimental crystallographic structural data are used.

Compared to the tremendous simulations reported for adsorption and transport in inorganic and organic zeolites[79-84] and carbonaceous materials,[85-90] only recently have a few simulation studies been conducted on transport in protein crystals.[52,91-95] A dynamic MC simulation revealed that the steric restriction was a predominant factor for the reduction of water diffusivity within a lysozyme crystal and that water diffusion in the lysozyme crystal was nearly ten times slower than in bulk phase.[52] A combined dynamic MC and BD simulations were carried out to study the diffusion of small and large molecules with or without net charges within orthorhombic and tetragonal lysozyme crystals.[91] The results demonstrated how the electrostatic interaction and steric confinement restricted the mobility of spherical probes in lysozyme crystals, and it was found that there existed a transition between the dominance of electrostatic effect for small probes and the steric confinement for larger molecules. However, the structure of guest molecules was not considered, which is a key factor to separate different molecules with similar size (e.g. chiral enantiomers). The dynamics of water and Na^+ counterions in an orthorhombic β-lactoglobulin crystal was investigated by a 5-ns MD simulation.[92] Within the pore with a radius of ca. 0.6–1.0 nm, water undergoes an anomalous diffusion in the proximity of the protein surface. Compared to water, the dynamics of Na^+ ions is disordered. However, the simulation time was not sufficiently long to provide a

conclusive description for the diffusion of Na^+ ions. A simple model was developed for evaluation of the diffusion times of small molecules into protein crystals, which accounted for the physical and chemical properties of both the protein crystal and diffusing molecules.[93] The transport of *L*-arabinose in an orthorhombic lysozyme crystal was investigated and the computed diffusion coefficients within the crystal were several orders of magnitude lower than in water.[94,95]

Overall, there have been few attempts in the theoretical or computational studies on the microscopic transport in protein crystals. It is generally concluded from these studies that the diffusion coefficients of guest molecules in protein crystals are greatly reduced than in bulk phase, consistent with experimental observation. However, the effects of crystal morphology and pore size on the transport of guest molecules are scarcely addressed. To the best of our knowledge, no computational study has been carried out on the effects of operating conditions such as electric field on the transport of guest molecules in protein crystals; there is yet no simulation study on the separation of mixtures in protein crystals used as stationary phase in liquid chromatography.

1.4 Objectives

The study on transport and separation in protein crystals is scarce; therefore, a number of important issues associated with the utilization of protein crystals as separation media have yet to be addressed. In order to facilitate the development of technically feasible and economically competitive separation technologies using protein crystals, a deeper understanding of transport and separation in protein crystals is required. Molecular simulations have unique advantages to shed light on this field as they can provide atomistic/molecular pictures that would otherwise be

experimentally intractable or impossible to obtain. In addition, molecular simulations can also complement experimental measurements.

The objectives of this thesis are to study the transport and separation of guest molecules in protein crystals using MD simulations, and subsequently provide molecular insights and guidelines for the development of high-performance protein crystals in separation technology. First, protein conformations and biological nanopores are characterized in protein crystals of various morphologies and topologies. Then, the dynamic and spatial properties of water and ions are examined in detail. Water and ions play a crucial role in determining the structure, dynamics, and functionality of proteins; and they are ubiquitously involved in separation processes. A clear understanding of their properties in protein crystals is very important. On the other hand, external environment (e.g. electric field) has a crucial effect on the properties of protein, water and ions; and the study will help optimize the separation technology using protein crystals. Therefore, the effects of electric field on the transport of electrolytes (electrophoresis) are also investigated. Following these, the chiral and achiral separation mechanisms in protein crystals are explored from the microscopic scale. In addition, the capability of different biomolecular force fields to predict the transport of water and ions in protein crystals is assessed. An appropriate biomolecular force field plays a deterministic role in the accuracy and reliability of simulations for protein crystals.

1.5 Thesis Outline

This thesis consists of seven chapters including the current one. Chapter 2 presents the diffusion of water and ions in three different protein crystals. In Chapter 3, the electrophoresis is investigated in a lysozyme crystal with the emphasis on the change

of protein structures and the electrical conductivity. Achiral and chiral separation processes in liquid chromatography with protein crystals as stationary phase are presented in Chapters 4-5. The elution orders are compared with experimental results and the separation mechanisms involved are discussed in detail. In Chapter 6, three biomoleculasr force fields are evaluated for their capability to predict the diffusion of water and electrolyte in a lysozyme crystal. Finally, general conclusions and outlook are summarized in Chapter 7.

Chapter 2 Water and Ions in Protein Crystals

2.1 Introduction

Protein crystals have emerged as promising bionanoporous materials for a wide range of applications such as separation, biocatalysis and biosensing. Known as bioorganic zeolites, protein crystals possess high porosities (0.5–0.8), large surface areas (800–2000 m^2/g), and a wide range of pore sizes (1.5–10 nm).[1] The pore size and porosity in protein crystals vary with the nature of the protein and crystallization conditions. Intriguingly, a protein can form various morphologies depending on the additive, pH, and temperature. For example, lysozyme exists in at least four crystalline forms: tetragonal, orthorhombic, monoclinic, and triclinic.[3] Compared with inorganic and organic zeolites, the inherently chiral nature is a salient feature of protein crystals. *L*-amino acids that constitute protein molecules create a chiral environment, which could separate pharmaceutically important enantiomers.[1]

In the past, the applications of protein crystals were limited by their fragility.[1] Crystalline protein molecules are virtually located at lattice sites via noncovalent van der Waals and electrostatic interactions, in contrast to the covalent bonds in zeolites. As such, they are mechanically soft and can easily disintegrate under unfavorable conditions.[1,96] This limitation has been largely reduced by cross-linking technology. Cross-linked protein crystals (CLPCs) are stable against mechanical disruption and shearing under mixing, filtration and pumping.[1] Cross-linked enzyme crystals (CLECs) are more convenient than solution enzymes to be separated from the reaction broth. Several techniques have been reported to effectively produce CLPCs or CLECs.[18-20] These collective features could allow CLPCs to be used as a novel class of molecular sieves[1] in biotechnological separation and as biocatalysts[2] in the

synthesis of fine chemicals, chiral intermediates, and peptides in the laboratory or on a commercial scale.

A number of studies, primarily experiments, have been carried out to investigate the properties of protein crystals and the mechanistic behavior of guest molecules therein. Adsorption in different lysozyme crystals revealed that the uptake capacity depends on solute type, crystal morphology, and solution characteristics (e.g., pH).[58,59] Steric repulsion, surface binding, cross-linker, and electrostatic interaction (especially for ions) were found to influence diffusion in protein crystals.[53,55,56,58] Furthermore, the separation of mixtures including racemic enatiomers in protein crystals was explored.[1,5,55,64] Different mechanisms have been proposed to resolve the separation on the basis of size exclusion or the difference in adsorption, charge, hydrophobicity and chirality.[1] Various factors such as eluent pH, type and concentration of organic and charged modifiers, ionic strength and temperature were identified to be important in the retentivity and enantioselectivity of solutes and, consequently, should be optimized in practice.

With the continually growing computational power and resource, molecular simulations have been playing an increasingly important role in life sciences. Simulations at the molecular scale can provide microscopic pictures that are experimentally inaccessible or difficult to obtain, if not impossible. Fundamental insight gained from molecular simulations can assist in the rational design of new materials and optimization of engineering processes. Numerous MC and MD simulations have been reported for adsorption and diffusion of fluids in inorganic and organic zeolites,[79-84] and in carbonaceous materials.[85-90] Nevertheless, fluid behavior in protein crystals has been scarcely investigated at the molecular level. Several earlier studies primarily focused on the difference of protein conformations in

solution and crystalline environments.[69,70,75,76] The water content is usually rather high in protein crystals ranging from 30% to 65%.[97] A handful of water molecules occupy the well-defined crystallographic sites and most water molecules are dispersed in the pores. From this aspect, crystalline protein is comparable to protein in solution. Recently a few simulation studies examined microscopic diffusion in protein crystals.[52,91,93] A random-walk algorithm was applied to estimate the effective diffusion coefficient of water in a tetragonal lysozyme crystal, and the reduction of water diffusion in the crystal was attributed primarily to steric limitations.[52] Dynamic MC and BD simulations were carried out to simulate the diffusion of spherical probes in lysozyme crystals; the electrostatic interaction and steric confinement were found to restrict the mobility. However, the structure of the probes was not taken into account, which is a key factor in separating different molecules with similar sizes, especially for chiral molecules.[91] A simple model was developed for evaluation of the diffusion times of small molecules into protein crystals, which accounts for the physical and chemical properties of both the protein crystal and diffusing molecules.[93]

Currently, our understanding of fluids in protein crystals remains largely obscure. The characteristics of protein crystals and subsequently the behavior of confined fluids would vary with medium (e.g., pH) and external environment (e.g., electric field), and little is known about the influence of crystal morphology. A set of guidelines on how to select a specific protein crystal and to optimize operation conditions are crucial to the new development of technically feasible and economically competitive separation technology using protein crystals.

Here we employ MD simulations to investigate the spatial and temporal behavior of water and ions in three protein crystals, particularly with different morphologies and topologies. Water and ions play a crucial role in determining the structure,

dynamics, and functionality of proteins.[98] As Szent-Györgyi (Nobel Laureate for the discovery of vitamin C) profoundly pointed out that the dominant feature of living state is macromolecule-water interaction.[99] Therefore, a clear understanding of a confined solvent or solute in different protein crystals under a variety of conditions is of central importance. This is also of fundamental significance for biomembrane channels, a topic for which MacKinnon won the Nobel Chemistry Prize in 2003.[100] Water or a specific ion can selectively permeate these channels, but the mechanism is far from complete. Due to the similarity of the pores/channels in protein crystals and biomembranes, and the more readily available atomic structures of protein crystals from experimental techniques such as X-ray diffraction, protein crystals could serve as a remarkable benchmark to examine biomembrane channels *in vivo*.[52] Consequently, fluid behavior in protein crystals can provide a direct insight into the less clear behavior in biomembrane channels.

2.2 Models and Methods

Three protein crystals, tetragonal lysozyme, orthorhombic lysozyme, and tetragonal thermolysin were studied. These proteins were crystallized at room temperature. The PDB IDs are 1HEL,[101] 1AKI,[102] and 1L3F[103] from the RCSB Protein Data Bank (PDB). The IDs are simply used below to denote the specific protein crystals. Lysozyme, an enzyme with the function to kill bacteria, is commonly referred to as the body's own antibiotic. The structure and function of the readily-available lysozyme have been widely studied. Lysozyme shows polymorphism in the crystal structure, thus providing a platform to study the effect of crystalline packing fashion on the behavior confined fluids. Thermolysin is a thermally stable metalloproteinase and hydrolyzes peptide bonds specifically on the amino side of bulky hydrophobic

residues such as Leu, Ile, Val, and Phe. Thermolysin crystal has been successfully utilized to separate several chiral mixtures.[1]

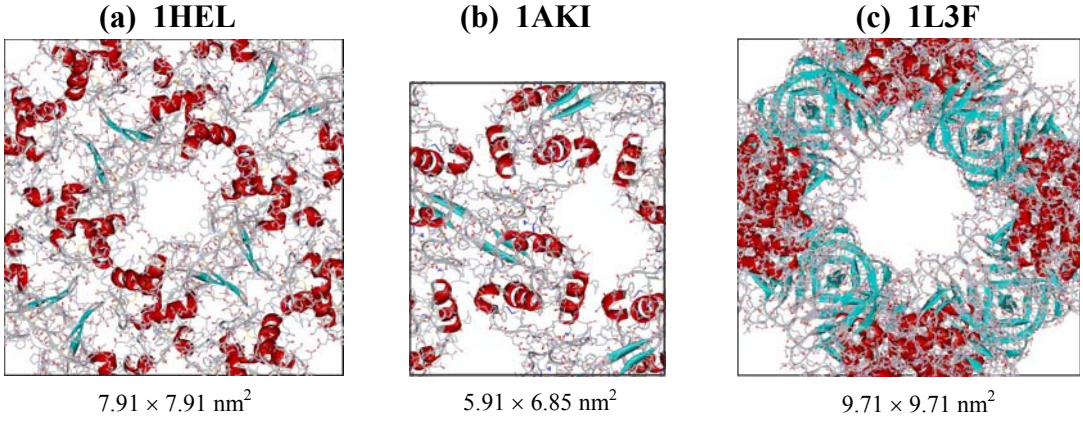

Figure 2.1. Three protein crystals: (a) tetragonal lysozyme, (b) orthorhombic lysozyme, and (c) tetragonal thermolysin. The views are on the *xy* plane, and the unit cell lengths in the *x* and *y* directions are indicated.

Crystal structures of the three protein crystals were constructed according to their space groups. **Figure 2.1** shows the unit cells of the three crystals, in which α-helices, β-sheets and random coils are illustrated in red, cyan and gray, respectively. Formation of the secondary structures in protein molecules is attributed to the hydrogen bonding. For all the three protein crystals in our simulations, pH was assumed to be 7. Consequently, Arg and Lys residues were protonated, while Asp and Glu residues were deprotonated on the basis of experimental p*K*a values.[104] **Table 2.1** lists the system parameters used for the simulations of the three protein crystals. In each crystal, a certain number of chloride counterions were introduced randomly to neutralize the system and water molecules were added to mimic a fully hydrated protein crystal sample. Two unit cells (1 × 1 × 2) were used for 1HEL and 1AKI; therefore, the simulation box was approximately cubic. In the latter, the system was rather large with 90 783 atoms. The overall mass density in the three crystals is approximately equal (~ 1.1×10^3 kg/m^3). The porosity changes from 38.6% to 42.9%

to 66.5% in 1HEL, 1AKI and 1L3F; and water content increases from 32.1 % to 35.1% to 58.8%.

Table 2.1. System parameters for three protein crystals.

	Tetragonal lysozyme (1HEL)	Orthorhombic lysozyme (1AKI)	Tetragonal thermolysin (1L3F)
Space group	$P4_32_12$	$P2_12_12_1$	$P4_12_12$
No. of unit cells	1 × 1 × 2	1 × 1 × 2	1 × 1 × 1
No. of protein chains	16	8	8
No. of protein atoms	21136	10568	25248
No. of Cl^- ions	128	64	16
No. of water molecules	5985	3465	21821
No. of total atoms	39219	21027	90783[b]
Box size (nm^3)	7.91 × 7.91 × 7.58	5.91 × 6.85 × 6.10	9.71 × 9.71 × 10.66
Mass density (kg/m^3)	1174.1	1197.3	1104.7
Water content (%, w/w)	32.1	35.1	58.8
Porosity[a] (%)	38.6	42.9	66.5

[a] Porosity is estimated as a ratio of the number of water molecules in the protein crystal to the total number of bulk water molecules that can fill in the simulation box without protein. The bulk water density of the SPC model is 977 kg/m^3 at 300 K.

[b] Each thermolysin consists of 4 Ca^{2+} and 3 Zn^{2+} ligand ions necessary for activity and stability.

GROMOS96 united-atom force field was used to model the protein molecules with each CH_x group as a single interaction site.[105] The pairwise additive interactions have two contributions; one is bonded and the other is nonbonded. The bonded interaction V_{bonded} includes stretching, bending, and proper and improper torsional potentials and is written as

$$V_{\text{bonded}} = \sum_{\text{bond}} \frac{1}{4} k_{ij}^{\text{b}} \left(r_{ij}^2 - b_{ij}^2 \right)^2 + \sum_{\text{angle}} \frac{1}{2} k_{ijk}^0 \left(\cos(\theta_{ijk}) - \cos(\theta_{ijk}^0) \right)^2 + \sum_{\text{dihedral}} k^{\phi} \left\{ 1 + \cos\left[\left(n\phi - \phi^0 \right) \right] \right\} + \sum_{\text{improper}} \frac{1}{2} k_{\xi} \left(\xi_{ijkl} - \xi_0 \right)^2 \tag{2.1}$$

where k_{ij}^{b}, k_{ijk}^{0}, k^{ϕ} and k_{ξ} are force constants of stretching, bending, and proper and improper torsional potentials respectively, r_{ij}, θ_{ijk}, ϕ and ξ_{ijkl} are the distance, angle, dihedral angle and improper dihedral angle respectively, b_{ij}, θ_{ijk}^{0}, ϕ^{0} and ξ_{0} are corresponding equilibrium values, and n is multiplicity. The nonbonded interaction $V_{\text{nonbonded}}$ includes LJ and Coulombic potentials and is written as

$$V_{\text{nonbonded}} = \sum_{i<j}\left[\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} + \frac{\text{erfc}(\beta r_{ij})}{4\pi\varepsilon_0 r_{ij}} q_i q_j\right] \tag{2.2}$$

where q_i is the partial charge, $C_{ij}^{(6)}$ and $C_{ij}^{(12)}$ are LJ paramters, β is a weight parameter, and erfc is the complementary error function.

GROMACS v3.3.1 simulation package was employed to perform our simulations because it is fast and particularly well-suited for biomolecular systems.[51] The periodic boundary conditions were applied in three dimensions to mimic infinitely large crystal structures. The LJ interactions were calculated with a spherical cutoff distance of 1.4 nm. The particle-mesh-Ewald (PME) method[41,42] was applied to calculate the Coulombic interactions, with a grid spacing of 0.12 nm and a fourth-order interpolation. The bond lengths with dangling hydrogen atoms in protein molecules were constrained using the LINCS algorithm.[40] Water was represented by the SPC model,[106] and the SETTLE algorithm[107] was used to constrain the geometry of the water molecules.

For each system, we started with a 1000-step energy minimization. The steepest descent method was used with a maximum force criterion of 1.0 kJ mol^{-1} nm^{-1}. Thereafter, initial velocities were assigned to the system according to the Maxwell-Boltzmann distribution at 300 K. Equilibrium simulation was performed in a canonical (NVT) ensemble at 300 K. In experiments, the crystallization temperature

was about 300 K for all the three proteins considered. The simulation temperature was controlled by a Berendsen thermostat with a relaxation time of 0.1 ps. After 1 ns of equilibration, subsequent production simulation was continued, 10 ns for the two lysozyme crystals and 5 ns for the thermolysin crystal. For all simulations, the integration time step was 2 fs. The neighbor list and long-range forces were updated every 10 steps (0.02 ps). Atomic coordinates and velocities were saved every 500 steps (1 ps) for further analysis. To examine the effect of thermal fluctuations of protein atoms, simulations were run with or without a position restraint, referred to as PR and NPR, respectively. In PR simulation, a harmonic position restraint of 1000 kJ mol^{-1} nm^{-2} was exerted on the heavy atoms of the proteins.

2.3 Results and Discussion

2.3.1 Fluctuations and Solvent-Accessible Surface Areas

The thermal motions of protein molecules were evaluated from NPR simulation. **Figure 2.2** shows the root-mean-squared fluctuations (RMSFs) of C_α atoms for each protein chain in the three crystals. Note that each curve is subsequently shifted for clarity. The RMSF patterns of different protein chains are similar in each crystal. For instance, some of C_α atoms in 1HEL fluctuate around 0.1–0.15 nm, while others exhibit peaks with fluctuations of about 0.2–0.3 nm. The weaker fluctuations are attributed to the formation of the secondary structures in protein molecules. Shown on the top of **Figure 2.2a-c** are the locations of the secondary structures, including α-helices, 3_{10}-helices, and β-sheets. In 1HEL, α-helices are Gly4–Gly16, Leu25–Phe34, Ile88–Ser100, and Val109–Arg114, 3_{10}-helices are Pro79–Leu84, Gly104–Trp108, and Val120–Ile124, and β-sheets are Ala42–Asn46, Gly49–Gly54, and Leu56–Ser60.[101]

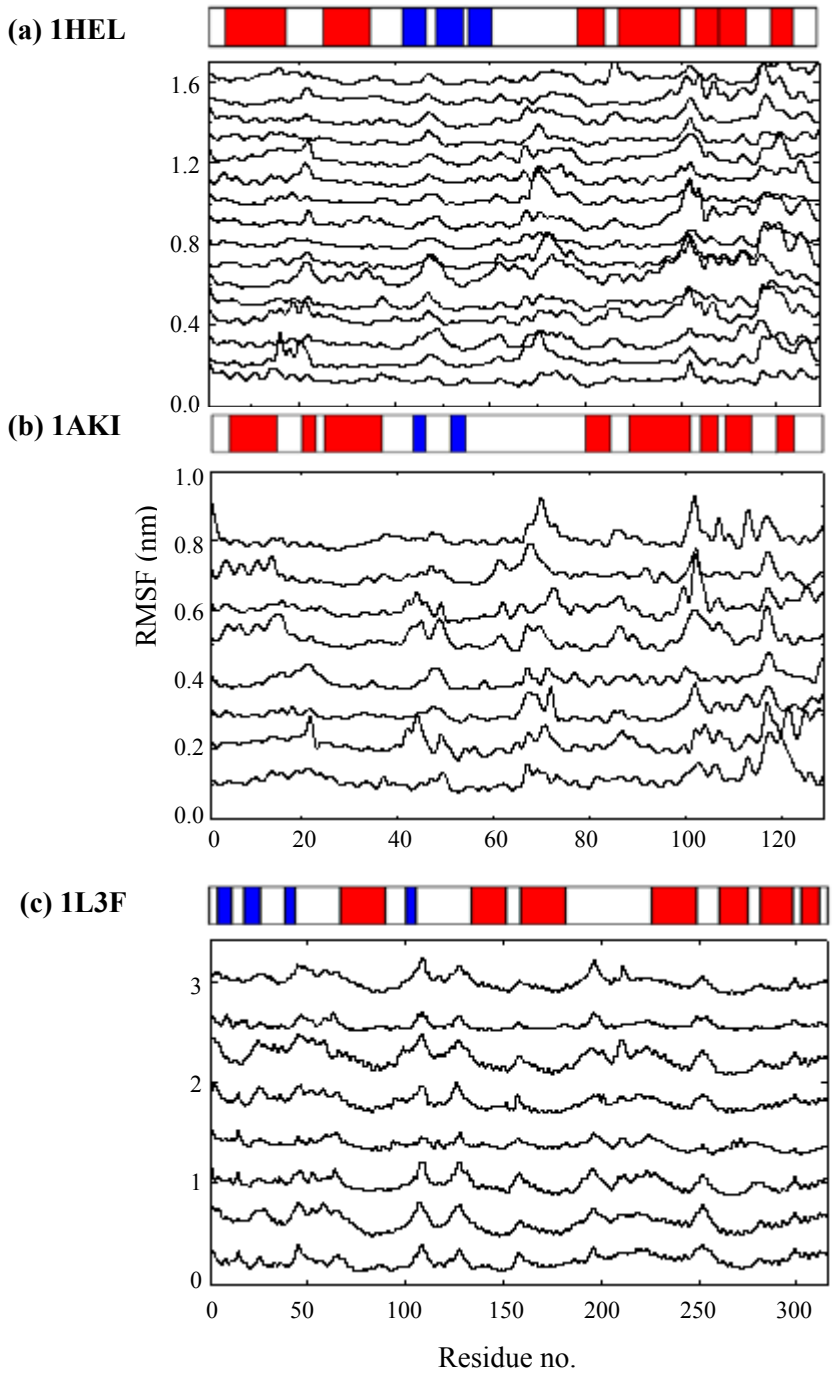

Figure 2.2. RMSFs of C_α atoms for each protein chain in (a) tetragonal lysozyme, (b) orthorhombic lysozyme, and (c) tetragonal thermolysin. For clarity, each curve is subsequently shifted by 0.1 nm for 1HEL and 1AKI and 0.4 nm for 1L3F in the vertical direction. On the top of each subfigure, the red regions denote helices and the blue regions denote sheets.

The secondary structures, especially α-helices and β-sheets, are two primary building blocks in proteins. Helix consists of turns and successive turns held by

hydrogen bonding. Usually, two helices twist around one another and form a coil. In β-sheet, polypeptide chain folds back and runs in a direction opposite to that of its neighbor by hydrogen bonding. The residues forming the secondary structures are bound more tightly and consequently exhibit relatively lower mobility. In contrast, some residues are random coils (e.g., Gly22, Thr47, Asp48, Pro70, and Gly102 in 1HEL) and thus exhibit greater RMSFs. Similar behavior is also observed in 1AKI and 1L3F. Though not shown, as expected the RMSFs from PR simulation are significantly smaller than from NPR simulation because of the exerted position restraint on protein atoms.

We also computed the solvent-accessible surface areas (SASAs) for the three protein crystals. The active sites of protein tend to locate at the surface; therefore, knowing the residue accessibility to surrounding solvent is important for the structure-property relation of protein molecule. **Table 2.2** lists the total, hydrophobic, and hydrophilic SASAs, as well as the average SASA per residue. An atom in protein is considered to be hydrophobic if its charge falls within ($-0.2e$, $0.2e$), and is hydrophilic otherwise.[50] SASAs are then determined using a probe with a diameter of 0.28 nm (approximately the size of a water molecule) to roll on van der Waals surface of protein crystal. There are two interesting observations. First, the hydrophobic SASA is close to the hydrophilic SASA for each protein crystal; second, the average SASA is roughly the same for all three proteins studied here. Our findings are consistent with a study by Lins et al. for 587 proteins in native or folded states.[108] They found that the ratio of hydrophilic to hydrophobic SASA is approximately equal to unity and the SASA of a protein increases with the number of residues. Not unexpected, however, the SASA of a protein in a crystalline environment studied here is much smaller than in solution. One of such consequences in enzyme crystals is that

the catalytic ability of enzyme is reduced as the number of active sites available for substrates decreases.

Table 2.2. SASAs (nm^2) of proteins and diffusivities (10^{-9} m^2/s) of water in three protein crystals.

	1HEL	**1AKI**	**1L3F**
Total SASA	768 ± 22	384 ± 6	950 ± 6
Hydrophobic SASA	390 ± 12	194 ± 4	503 ± 5
Hydrophilic SASA	378 ± 11	190 ± 3	447 ± 4
Average SASA per residue	0.37 ± 0.01	0.37 ± 0.01	0.38 ± 0.01
D **(PR)**	0.78 ± 0.01	0.86 ± 0.01	2.40 ± 0.08
D **(NPR)**	0.93 ± 0.03	0.91 ± 0.02	2.44 ± 0.04
D_x **(NPR)**	0.82 ± 0.03	0.68 ± 0.06	2.39 ± 0.02
D_y **(NPR)**	0.88 ± 0.01	0.78 ± 0.01	2.32 ± 0.06
D_z **(NPR)**	1.09 ± 0.04	1.28 ± 0.01	2.61 ± 0.04
D_z/D_x **(NPR)**	1.33 ± 0.07	1.89 ± 0.17	1.09 ± 0.02
D_z/D_y **(NPR)**	1.23 ± 0.05	1.65 ± 0.03	1.12 ± 0.04

2.3.2 Biological Nanopores and Water Densities

In protein crystals, biological nanopores are periodically distributed with different sizes and shapes, which is a common feature shared by many protein crystals. These pores are interconnected and partially or fully closed; consequently, it is difficult to characterize them separately. There might exist major pores, minor pores, and cavities in a protein crystal, in which the major pores have the largest sizes and fairly well-defined structures. Each protein crystal considered here possesses one type of major pore along the *z* axis, and there are two such major pores per unit cell with identical morphology and size but at different locations.

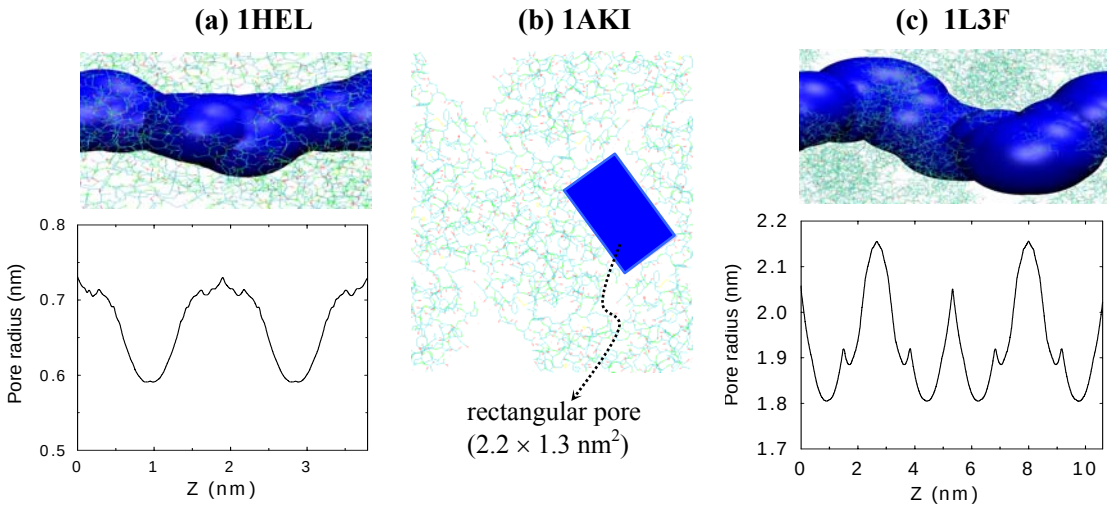

Figure 2.3. Pore structures and sizes in (a) tetragonal lysozyme, (b) orthorhombic lysozyme, and (c) tetragonal thermolysin. In orthorhombic lysozyme, the pore is approximately rectangular with the area of 2.2 × 1.3 nm^2 and assumed to be uniform along the *z* direction, and the view in (b) is on the *xy* plane.

Figure 2.3 shows that the morphologies of the major pores in 1HEL and 1L3F crystals are nearly cylindrical. Their sizes were calculated using the HOLE program.[109] The van der Waals radii of protein atoms at the pore surface were adopted from the GROMOS96 united-atom force field. In the 1HEL crystal, the pore is zigzag periodically along the *z* axis and the pore radius is 0.6–0.7 nm. A similar zigzag pore exists in the 1L3F crystal, but with a larger radius of 1.8–2.2 nm. The major pore in the 1AKI crystal (see **Figure 2.3b**) is considered to be approximately rectangular (2.2 × 1.3 nm^2) and uniform along the *z* axis. These estimated pore structures and sizes were used to calculate water densities within the pores.

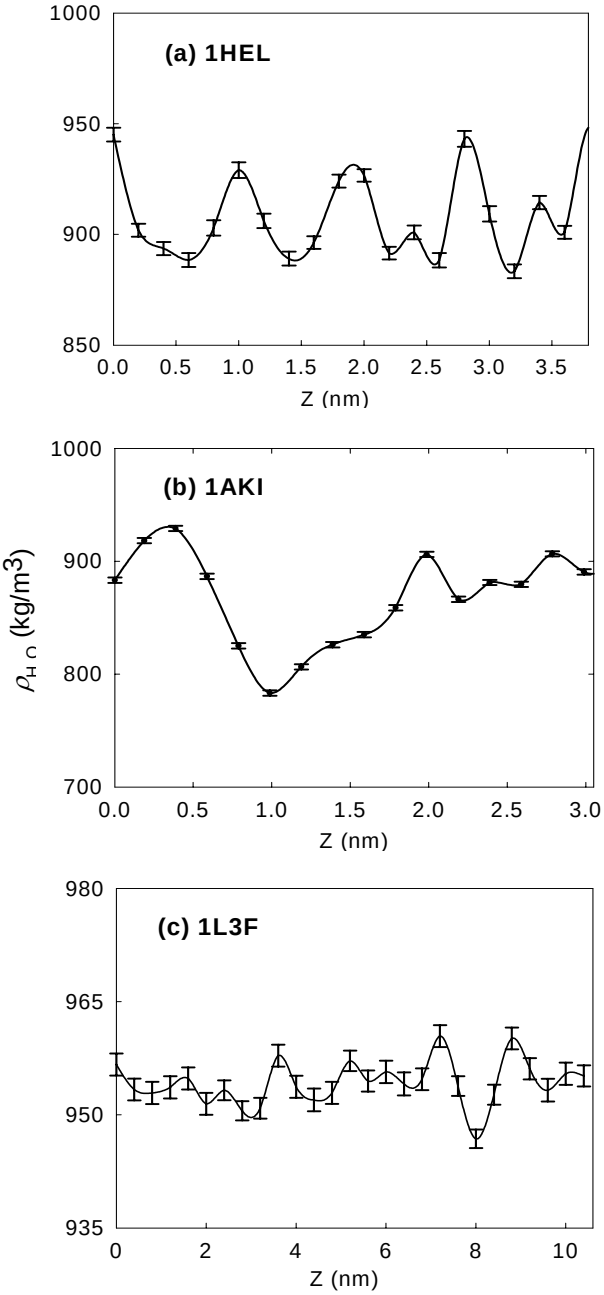

Figure 2.4. H_2O densities along the *z* direction within the major pores of the three protein crystals.

To calculate water density within a biological nanopore, the pore was divided into segments with an identical thickness of 0.01 nm along the pore axis. Each segment is nearly cylindrical, and its volume was calculated using the radius profiles in **Figure 2.3**. The number of water molecules in every segment was counted each picosecond.

After the water number and volume in every twenty sequential segments were summed up, which forms a slice of 0.2 nm (0.01 × 20 = 0.2) in thickness, the water density in each slice was calculated. **Figure 2.4** shows the water densities within the pores of the three protein crystals from PR simulations with position restraints on the protein atoms. Water distributes heterogeneously and fluctuates along all the three pores at the level of the simulated cell. These reflect the pore surface characteristics such as the hydrophobicity and connectivity with other pores/cavities. Compared to the other two protein crystals, water within the pore of 1AKI exhibits a larger fluctuation. This is due to the simple assumption of a uniformly rectangular pore of 1AKI. In contrast, the fluctuation within the pore of 1L3F is small because of its large pore, which reduces the effect of pore surface characteristics.

Protein crystals can serve as benchmark models for biomembranes, and it is interesting to compare water behavior within the pores of protein crystals with that in biomembranes. One difference between the pores/channels in protein crystals and biomembranes is the side opening. Normally, the channels of membranes are fully or almost fully closed in the side surface, which is crucial to selective permeation. However, it is a common feature that most regions of the pores in protein crystals are partially or almost fully open. In other words, the pore surface in protein crystals is not closed; water inside and outside of the pore can exchange through not only the pore mouth, but also the pore surface. As a consequence, water density within the pore of the protein crystals is comparable to the bulk density. However, water density within membrane channels distinctly deviates from the bulk density[110] and the profile is consistent with the cross-section area.[111] Water exhibits a large dynamic fluctuation,[112] even oscillates between the liquid and vapor states,[110,113] and water structure is well ordered as observed within membrane channels.[114,115] Within the pore

of the protein crystals, however, a less ordered structure is expected due to the highly open pore surface. In addition to side opening, pore/channel radius, interaction with water, and the ratio of hydrophobic/hydrophilic SASAs also play a role in the difference between protein crystals and membranes.

2.3.3 Radial Distributions of Water and Ions

In a protein crystal, the structure of fluid is expected to be affected by the nature of the protein molecule and other factors such as the packing pattern. A straightforward way to characterize the structural information is radial distribution function $g_{ij}(r)$ that is defined as

$$g_{ij}(r) = \frac{\langle \Delta N(r, r+\Delta r) \rangle V}{4\pi r^2 \Delta r \, N_i N_j} \tag{2.3}$$

where r is the distance from species j to i, $\Delta N(r, r+\Delta r)$ is the ensemble averaged number of species j around i within a shell from r to $r + \Delta r$, and N_i and N_j are the numbers of species i and j. **Figure 2.5a** shows $g_{\mathrm{Cl^-\text{-}OW}}(r)$ between the counterion Cl^- and the oxygen atom (OW) of the water molecules in the three protein crystals from NPR simulation. The $g_{\mathrm{Cl^-\text{-}OW}}(r)$ profiles exhibit similar behavior in all the three protein crystals. The first peak of $g_{\mathrm{Cl^-\text{-}OW}}(r)$ is located at 0.321 nm, the second is at 0.501 nm, and between them there is a minimum at 0.393 nm. Despite the different morphologies of the two lysozyme 1HEL and 1AKI crystals, the peak values of $g_{\mathrm{Cl^-\text{-}OW}}(r)$ are nearly identical with the same nature of the protein molecule. Both of them are larger than those in the thermolysin 1L3F crystal because of the higher porosity in the latter. We also estimated the hydration number of Cl^- from

$$N_{\mathrm{hy}} = \rho_{\mathrm{OW}} \int_0^{r_{\min}} [\mathrm{g_{Cl\text{-}OW}}(r)] 4\pi r^2 \mathrm{d}r \tag{2.4}$$

where ρ_{OW} is the average density of OW and r_{min} is the position of the first minimum. N_{hy} values are found to be 6.0, 6.3, and 7.3 in 1HEL, 1AKI, and 1L3F, respectively.

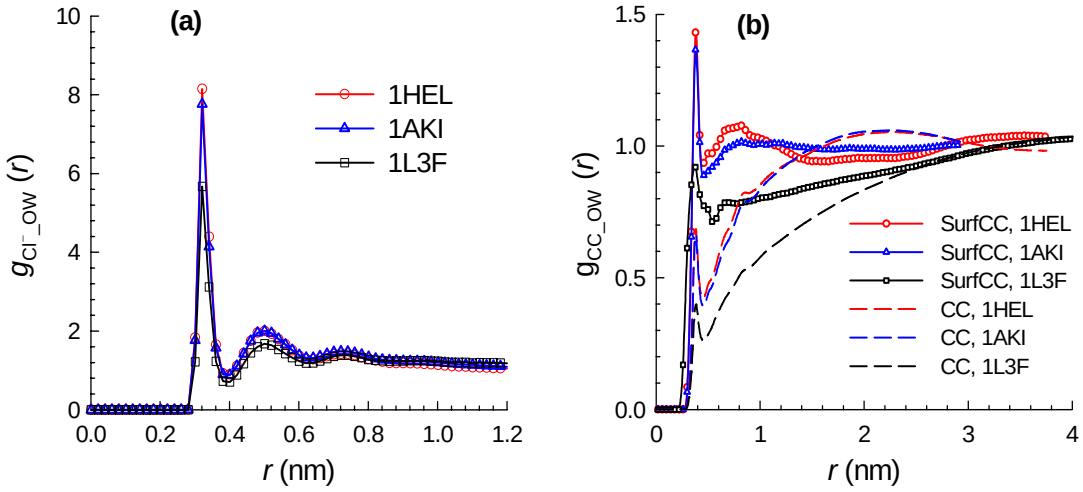

Figure 2.5. Radial distribution functions between (a) Cl^- and OW and (b) CC and OW in the three protein crystals. SurfCC denotes the carbonyl carbon atoms near the protein surface with SASA ≥ 0.03 nm^2, while CC denotes all the carbonyl carbon atoms.

The local structures of water and Cl^- are expected to vary at different regions of the protein surface as attributed to the heterogeneity of the protein surface.[116] **Figure 2.5b** shows $g_{CC-OW}(r)$ between the carbonyl carbon (CC) atom and OW atom for the three protein crystals (dashed curves). Although at large distance $g_{CC-OW}(r)$ approaches to unity, the peak values of $g_{CC-OW}(r)$ are rather small, about 0.70 in 1HEL and 1AKI crystals and 0.40 in 1L3F crystal. However, this way of calculating $g_{CC-OW}(r)$ is not physically sound because a large fraction of CC atoms are actually buried within protein molecules and hence inaccessible to water molecules. To obtain physically more meaningful $g_{CC-OW}(r)$, CC atoms with SASA larger than 0.03 nm^2 representing the surface CC atoms were used instead to calculate $g_{CC-OW}(r)$, as indicated by solid curves in **Figure 2.5b**. In this case, the peak values of $g_{CC-OW}(r)$ are observed to

increase. In 1HEL and 1AKI crystals, again water shows similar structures with a peak of 1.45. In the 1L3F crystal, the peak is 0.93 and less than that in 1HEL and 1AKI crystals, which is attributed to the larger pore and higher porosity in 1L3F as mentioned earlier.

2.3.4 Number Distributions of Water and Ions

Because of the favorable interaction with the protein molecules, water exhibits an ordered structure in the vicinity of the protein surface in the crystal. To quantify, we calculated the number distribution $N_{\mathrm{dis}}(r)$ of water around the protein surface as shown in **Figure 2.6a** from NPR simulation. $N_{\mathrm{dis}}(r)$ is defined as

$$N_{\mathrm{dis}}(r) = \frac{\langle \delta N(r) \rangle}{N_{\mathrm{t}}\, \delta r} \tag{2.5}$$

where r is the distance between target species (e.g., water) and the nearest protein atom, and the van der Waals radii of protein atoms were taken into account. $\langle \delta N(r) \rangle$ is the ensemble averaged number of target species within a layer of thickness δr (δr = 0.01 nm in our calculation). N_{t} is the total number of target species in a protein crystal. Note that we did not calculate the density distribution because the volume profile away from protein surface is hard to quantify with highly irregular pores or cavities in the crystal structure. From **Figure 2.6a**, a distinct peak of $N_{\mathrm{dis}}(r)$ is observed at 0.15 nm away from the protein surface in each crystal, which is roughly equal to the radius of a water molecule. This clearly demonstrates the existence of a hydration shell surrounding the protein surface. If the shell thickness is set to 0.3 nm, about the size of a water molecule, the number of water molecules within the shell is estimated to be 4800, 2610, and 7370 separately in 1HEL, 1AKI and 1L3F crystals. Correspondingly, water within the hydration shell overwhelmingly accounts for 80%

and 75% of the total water content in 1HEL and 1AKI crystals. The percentage is low (34%) in the 1L3F crystal because of its large pore and high porosity. However, water exhibits a wider distribution in 1L3F than in the two lysozyme crystals.

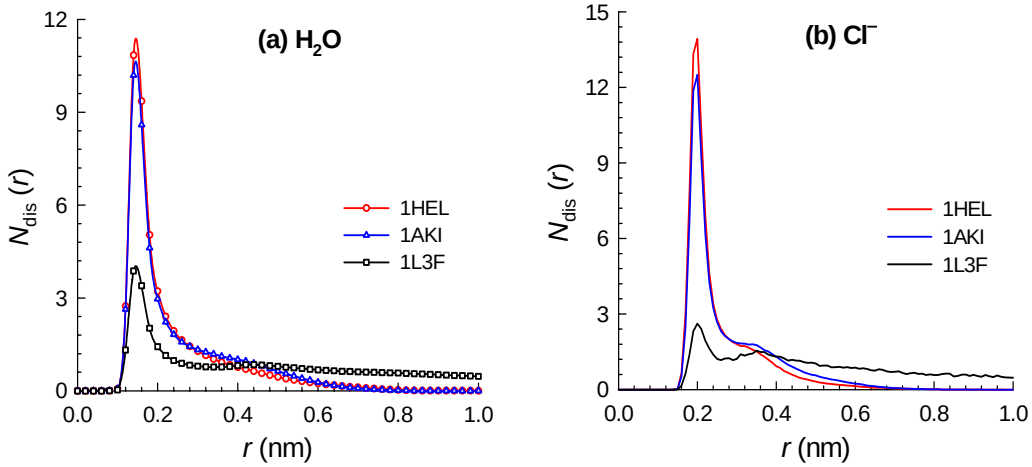

Figure 2.6. Number distributions of (a) H_2O and (b) Cl^- as a function of the distance from the protein surface in the three protein crystals.

Water molecules are found in abundance near the protein surface and govern the structure, functionality, and reactivity of natural and biological systems.[117,118] To characterize the water hydration shell, one should estimate the time scale of water motion within the shell. Due to the strong interactions with protein surface atoms, water molecules have a much longer residence time within the hydration shell than in other regions away from the protein surface (data not shown here). Furthermore, water dynamics within the hydration shell is believed to be partially determined by the lifetime kinetics of hydrogen bonding, which in turn depends on the structure of the hydration shell induced by the protein surface.[119]

The affinity of counterions for binding sites leads to the adsorption of Cl^- on the accessible protein surface. It is well recognized that the binding of specific ions, also called cofactors, is essential for the performance of catalytic enzymes. **Figure 2.6b** shows the normalized number distribution of Cl^- from the protein surface in the three

crystals. A preferential layer and a diffusion layer are observed for Cl^- around the protein surface, which reveals the structures of the electric double layers. The preferential layer is at 0.20 nm from the protein surface, roughly equal to the radius of Cl^- (about 0.22 nm). Similar to **Figure 2.6a** for water, the peak here is considerably lower in 1L3F than in 1HEL and 1AKI; however, the distribution is broader.

2.3.5 Diffusions of Water and Ions

Diffusions of water and counterions were examined in the three protein crystals. **Figure 2.7** shows the mean-squared displacements (MSDs) of water and Cl^- from both NPR and PR simulations. The MSD was calculated by

$$\mathrm{MSD}(t) = \left\langle \frac{1}{N} \sum_{i=1}^{N} \left| \mathbf{r}_i(t) - \mathbf{r}_i(0) \right|^2 \right\rangle \tag{2.6}$$

where N is the number of target molecules, and $\mathbf{r}_i(t)$ is the position of molecule i at time t. The multiple-origin method was employed to improve the statistical accuracy. **Table 2.2** gives the corresponding diffusivities from the Einstein equation[120]

$$D = \frac{1}{2d} \lim_{t \to \infty} \frac{\mathrm{MSD}(t)}{t} \tag{2.7}$$

where d is dimensionality. Our simulated diffusivities are in close agreement with the calculated values of small molecules by Geremia et al.[93]

The mobility of water is generally enhanced with increasing porosity of the crystal, simply as a consequence of more free space available in the crystalline environment. As expected, water diffusivities in all three protein crystals are reduced compared to that in bulk. On the basis of the SPC model, the simulated bulk water diffusivity is 4.3 $\times$ 10^{-9} m^2/s at room temperature.[121,122] Without position restraints on protein atoms, water diffusion is enhanced in both lysozyme crystals. The thermal fluctuations of protein atoms in NPR simulation can open pathways for diffusion, which are

inaccessible in PR simulation, and thus promote mobility of confined fluid. Nevertheless, there is no enhancement of water mobility in the 1L3F crystal within the statistical uncertainty. This is attributed to its high porosity and large pore, which reduces the influence of protein atoms on water dynamics. It is instructive to compare the mobility of water in the two lysozyme 1HEL and 1AKI crystals from PR and NPR simulations. With position restraints on the protein atoms, water mobility is larger in 1AKI with higher porosity than in 1HEL. Without position restraints, however, the mobility is nearly the same in 1AKI and 1HEL. Therefore, both the porosity and flexibility of the protein framework play a role in water diffusion in protein crystals, and there is a counterbalance between them.

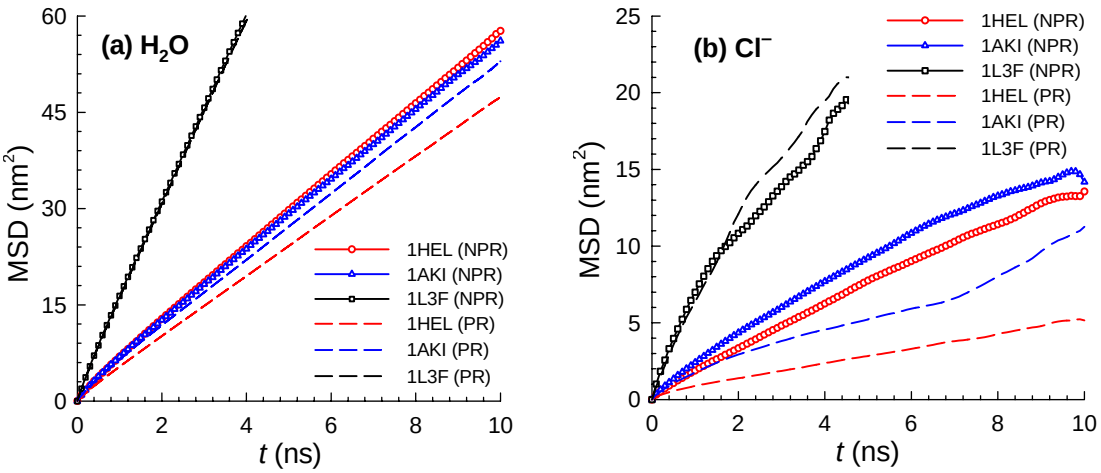

Figure 2.7. Mean-squared displacements of (a) H_2O and (b) Cl^- in the three protein crystals.

Figure 2.7b shows the MSDs of Cl^- in the three crystals. While the general behavior here resembles that in **Figure 2.7a** of water, the MSDs exhibit a bit of a nonlinear fashion as a function of time, particularly in 1L3F. This is because the number of Cl^- ions is significantly fewer compared to that of water molecules in the

crystals. To improve the accuracy, one may use a larger simulation system which consists of more counterions.

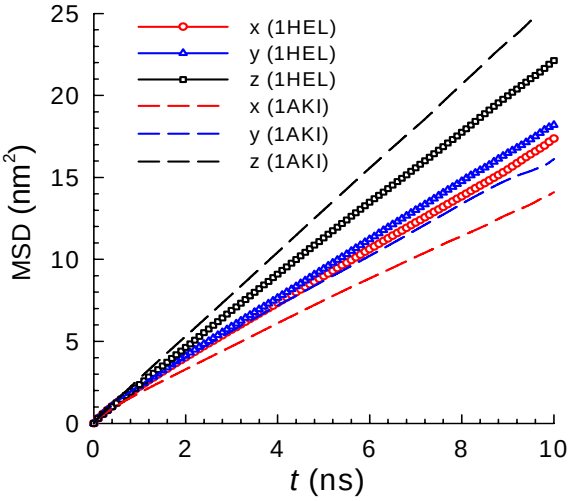

Figure 2.8. Mean-squared displacements of H_2O along the x, y, and z directions in the tetragonal (1HEL) and orthorhombic (1AKI) lysozyme crystals.

Diffusion in protein crystals has been demonstrated experimentally to be anisotropic, and the level of anisotropy depends on crystal morphology.[57] **Figure 2.8** shows the calculated MSDs of water separately along the x, y, and z directions in 1AKI and 1HEL crystals from NPR simulation, and the estimated diffusivities are listed in **Table 2.2**. In both 1HEL and 1AKI, D_y is slightly larger than D_x, whereas D_z is considerably larger than D_y or D_x. The reason is that the major pores in both crystals are preferentially along the z axis; as a consequence, the steric hindrance is less for water to travel in the z direction. This feature simply implies that the pore distribution is anisotropic. Furthermore, D_z / D_x and D_z / D_y in 1AKI are larger than in 1HEL, indicating a more pronounced anisotropy along the z direction in 1AKI because there are no other substantial pores in the x or y direction.[91] In 1L3F crystal, D_z / D_x and D_z / D_y are less than in the two lysozyme crystals; therefore, the diffusion is more isotropic in 1L3F crystal. Similar to 1HEL and 1AKI crystals, the major pore

in 1L3F is along the *z* direction, but the pore is much larger in size and the degree of anisotropy is hence reduced. The anisotropy of pore distribution in the three crystals increases in the order of 1L3F < 1HEL < 1AKI.

Recently, Cvetkovic et al.[57] experimentally measured the anisotropic diffusion of fluorescein in four lysozyme crystals with various morphologies (tetragonal, orthorhombic, monoclinic, and triclinic). In the tetragonal lysozyme, diffusions in three directions change in the order of $D_z > D_y > D_x$. What we have found for water diffusion in the same crystal is consistent with the experimental observation. Cvetkovic et al. further proposed an anisotropic diffusion model to correlate the relation between pore size and solute diffusivity. The solution characteristics, crystal porosity, and density were identified to be irrelevant to the observed anisotropies. It was thus hypothesized that diffusion anisotropy is caused by anisotropy of the pore size and connectivity (network) in the protein crystals. Nevertheless, how to fully explain the anisotropic diffusion is still an open question and requires more detailed investigation to improve our understanding. Further studies in this aspect are instructive for the emerging application of protein crystals. For instance, protein crystals used as a biosensor could respond rapidly if the crystal particles are oriented with major pores directly toward the specimen.[57]

2.4 Conclusions

From MD simulations, water and ions have been studied in three protein crystals with various morphologies and topologies. A distinct hydration shell surrounding the protein surface is identified from water number distribution, and 75–80% of water molecules are found to locate within the hydration shell in the two lysozyme 1HEL and 1AKI crystals. Water density within the major biological nanopores is inhomogeneous and depends on the characteristics of the pore surface. The diffusions

of water and ions in the protein crystals are enhanced as the porosity increases. Relatively smaller thermal fluctuations are observed in the secondary structures of protein molecules as a result of hydrogen bonding. The flexibility of the protein framework promotes the mobility of water and ions, particularly in the two lysozyme crystals. In agreement with experimental measurement, diffusions in the three protein crystals are anisotropic with a larger diffusivity along the pore axis than in the other two directions. The simulation results provide a clearer understanding of the dynamical coupling between guest fluids and protein crystals, and could inspire the emerging applications of protein crystals as bioorganic nanoporous materials. Further improved mechanistic insights into the behavior of confined fluids in protein crystals are crucial to better elucidate protein functionality and innovative biotechnological applications.

Chapter 3 Electrophoresis in a Lysozyme Crystal

3.1 Introduction

Electrophoresis refers to the transport or migration of ions under the influence of electric field and plays a crucial role in separation, sensing, biology, etc. For instance, capillary electrophoresis is a commonly used method for the purification of ionic species in a buffer-filled narrow capillaries with 25–100 μm in diameter.[123] A recent work showed that DNA adsorbed on a nanopatterned surface had a length-dependent mobility under an electric field and thus was separated. This opened up a new way for the rational design of nanodevices using surface-directed separation.[124] In cell membranes, there is instantaneous electric field between intra- and extra-cellular environments, which drives ions and pharmaceutical molecules across membranes in many bioprocesses.[125] As a consequence, a better understanding of electrophoresis in confined space is of central importance to the new development of engineering controllable nanofluidic channels, function-tuning lap-on-a-chip devices, and biologically significant electrical signaling in nervous systems.

Over decades, Poisson-Nernst-Planck (PNP) model has been prevailed to describe electrophoresis.[126] The mean-field based PNP model provides a classical continuum framework; however, the molecular-level details and correlations are not incorporated. Therefore, PNP model is unable to examine the complexity of electrophoresis at the molecular scale, particularly if the confined channel is in nano-domain, the surface-to-volume ratio is large, and the interfacial effect dominates. Alternatively, with ever-growing computational power, molecular simulations have been increasingly used with their ability to give microscopic insights that are otherwise experimentally inaccessible or difficult to obtain. Tang et al. conducted MD simulations to investigate

the effects of confinement on the structural and transport properties of 0.5 M KCl in cylindrical nanopores of radii 4.75–15.8 Å.[127,128] The results revealed a decrease in the hydration of ions and hydrogen bonding of water as pore radius reduces, and the external electric field has a strong influence on the orientation of water molecules. Dzubiella et al. reported an MD simulation study for the electric field-controlled water permeation coupled to ion transport through a hydrophobic nanopore. Ionic charge imbalance across the nanopore was suggested to induce water permeation and thus permeable to ions.[129,130] Murad et al. studied the separation of supercritical aqueous electrolyte solutions in thin zeolite membranes from MD simulations. They showed the important role of electric field in enhancing the separation rate and found the serious risk in applying macroscopic hydrodynamics to nanoscale systems.[131,132] Hwang et al. developed a kinetic lattice grand canonical MC simulation method for a model ion channel system. The simulated ion currents, electrostatic potentials and ion concentrations agree well with the PNP predictions if the channel has the same dielectric constant as water; however, there is considerable difference if the channel has a lower dielectric constant and thus the reaction field effect is missing in PNP theory. [133,134]

Electric field may enhance conductivity and permeability of molecules in biomembranes. This can trigger drug delivery across cell membranes (electroporation) and be used for cancer treatment and gene therapy.[135] However, understanding the fundamental mechanism of transport in biomembranes is difficult from the microscopic scale due to the unknown atomic structures of many membrane proteins (not easy to be crystallized). Recently, it has been recognized that protein crystals can serve as a remarkable benchmark for biomembranes because of the biological similarity between the pores in protein crystals and the channels in biomembranes;

also importantly, structures of a large number of proteins are readily available.[2] Therefore, the study of fluid behavior in protein crystals under electric field is of significance to provide insight into the less clear behavior in biomembranes. With high porosities, large surface areas, wide range of pore sizes and unique chiral environments, protein crystals have also emerged as new separation media for chemically or optically different molecules through size exclusion or chiral discrimination.[1] Water and ions are ubiquitously involved in separation processes and a better understanding of their behavior from a molecular level is of central importance for the rational design of high-performance bioseparation technologies.

In Chapter 2, we reported an equilibrium MD (EMD) simulation study to investigate the spatial and temporal properties of ions and water in three protein crystals with different morphologies and topologies. The diffusion in the crystals was found to be anisotropic preferentially along the pore axis and to be enhanced with increasing porosity. In the current work, non-equilibrium MD (NEMD) simulations are performed to explore electrophoresis of NaCl and $CaCl_2$ mixture in a lysozyme crystal under the electric field. NEMD method was first proposed by Evans and Morris[136] and later applied to bulk electrolytes.[137] We choose lysozyme as a model protein because it is readily available with well-known structure. Lysozyme functions to kill bacteria and is commonly regarded as the body's own antibiotic. An electrolyte mixture of NaCl and $CaCl_2$ is considered representing a general buffer solution. In addition to the static and transport properties of Na^+, Ca^{2+} and Cl^- ions, we examine the orientation of water molecules and the stability of lysozyme molecules under the electric field. Water plays a critical role in the structure, dynamics, and functionality of proteins; a clear understanding of confined water in protein crystal is of fundamental interest. It is also intriguing to study the stability and structural change of

protein under the electric field, which have been rarely examined. Nevertheless, this is an important topic based on the fact that there has been increasingly large number of electrical devices used daily around us.

3.2 Models and Methods

Lysozyme exists in different crystalline forms, namely, tetragonal, orthorhombic, monoclinic and triclinic.[3] In this work, tetragonal lysozyme (PDB ID: 1HEL) is considered with a $P4_32_12$ space group.[101] The lattice constants are $a = b = 7.91$ nm, and $c = 3.79$ nm. **Figure 3.1** shows the surface representations of the hydrophobic and hydrophilic parts of tetragonal lysozyme crystal on the *xy* plane. In the crystal, biological nanopores are interconnected and periodically distributed with different sizes and shapes. Among them, there are two major pores in a unit cell; one is at the center and the other at the corner, as clearly seen in **Figure 3.1**. The major pore is approximately cylindrical along the *z* axis with a slight zigzag and the radius is in the range of 0.6–0.7 nm, as discussed in Chapter 2.

The simulation box contained two unit cells with a size of $7.91 \times 7.91 \times 7.58$ nm^3. At the physiological condition (pH ≈ 7), Arg and Lys residues were protonated, while Asp and Glu residues were deprotonated based on the experimental pK_a.[104] As a consequence, each lysozyme molecule carried eight positive unit charges. Water molecules were added to mimic a fully hydrated crystal and three types of ions (Cl^-, Na^+ and Ca^{2+}) were introduced randomly while keeping the system neutral. Overall, the simulation box consisted of 16 lysozyme molecules, 5270 water molecules, 557 Cl^- ions (1.95 M), 143 Na^+ ions (0.50 M) and 143 Ca^{2+} ions (0.50 M). GROMOS96 united-atom force field was adopted for lysozyme and ions.[105] Water was mimicked by the SPC model.[106] Some MD simulation settings were same as in Chapter 2, and

thus not described here. **Table 3.1** gives the LJ parameters and charges of the three ions and oxygen (OW) and hydrogen (HW) in water.

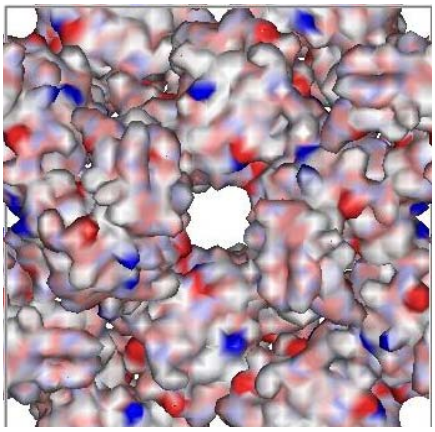

Figure 3.1. Surface representations of a unit cell of tetragonal lysozyme crystal on the *xy* plane. The hydrophobic and hydrophilic (blue) parts are in red and blue respectively.

Table 3.1. Lennard-Jones potential parameters and charges.

	Ca^{2+}	Na^{+}	Cl^{-}	OW	HW
ε **(kJ/mol)**	0.507	0.062	0.446	0.650	0
σ **(nm)**	0.281	0.258	0.445	0.317	0
q **(*e*)**	+2	+1	–1	–0.82	+0.41

Simulations were performed with GROMACS v3.3.1 package because it is extremely fast and particularly well-suited for biomolecular systems.[51] The system was initially subject to energy minimization using the steepest descent method with a maximum step size 0.01 nm and a force tolerance 10 kJ mol^{-1} nm^{-1}. Then velocities were assigned according to the Maxwell-Boltzmann distribution at 300 K, followed by 2 ns MD simulation. Drift velocities were tracked in NEMD simulation to assure the system reaching a steady state. In the absence of electric field, EMD was run with energy monitored for the system to reach equilibrium. Finally, 10 ns MD simulations

were conducted for production. MD simulations were performed in a canonical ensemble (NVT) at 300 K using the Berendsen thermostat with a relaxation time of 0.1 ps. An integration time step of 2 fs was used and trajectories were saved every 1 ps.

For NEMD simulation in this study, a uniform electric field was exerted along the nanopore axis (the z direction) ranging from E_z = 0.1, 0.2, 0.3 to 0.4 V/nm. It is worthy to note that the electric field strength adopted here is typically 1–2 orders of magnitude stronger than that across cell membranes or used experimentally.[138,139] Such a strong electric field is indeed commonly used in NEMD to reduce the impact of thermal noise and therefore to enhance the "signal-to-noise" ratio within a nanosecond time scale.[129] Furthermore, it is subtle to maintain the temperature in NEMD simulation since the ohmic heat generated by ion flux under the electric field must be removed.[140] To calculate the temperature in NEMD run, ion velocities along the direction of the electric field were subtracted. In our NEMD simulation, it was observed that the system temperature was kept fairly well around 300 K with an acceptable small fluctuation of approximately 1.1 K.

3.3 Results and Discussion

3.3.1 Protein Stability and Structural Change

The stability of protein molecules exposed to an electric field is of importance from a practical point of view. It has been reported that cell membranes exposed to a strong electric field would deform and lose part of surface or even disintegrate.[135] In the crystal, lysozyme molecules are constrained at the lattice sites via noncovalent van der Waals and electrostatic interactions, and intermolecular hydrogen bonds. Under the electric field E_z, the positions of lysozyme molecules are shifted to some extent.

Figure 3.2a shows the averaged RMSDs of lysozyme heavy atoms from the initial crystallographic structure as a function of E_z. The values of RMSD increase monotonically with E_z and rise from 0.35 to 0.57 nm upon increasing E_z from 0 to 0.4 V/nm. This indicates that the stability of protein crystal is reduced under the electric field, and could cause protein denaturation or damage in the biological functions of protein active sites. Nevertheless, as mentioned above, for the sake of reducing the impact of thermal noise, the electric field used is about several orders of magnitude stronger than normally used in experiments. In reality, the effect of electric field is substantially smaller and protein molecules can maintain their stability fairly well.

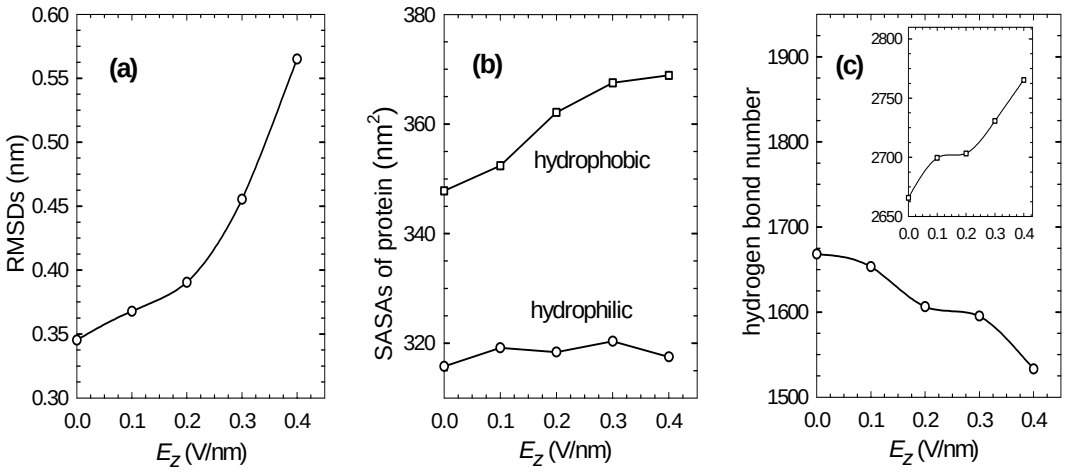

Figure 3.2. (a) Averaged RMSDs of lysozyme heavy atoms from the initial crystallographic structure. (b) Hydrophobic and hydrophilic solvent-accessible surface areas (SASAs). (c) Number of hydrogen bonds between lysozyme molecules, the inset is number of hydrogen bonds between lysozyme and water molecules.

As shown in **Figure 3.2b**, the hydrophobic SASA increases with E_z and appears to approach a constant at high E_z, while the hydrophilic SASA is essentially independent of E_z. It is recognized that the hydrophobic interaction is one of major factors in protein stabilization.[141] Without the electric field applied, the side chains of hydrophobic amino acids are primarily buried inside protein molecules and thus largely inaccessible to solvent. Upon exposure to an external electric field, however,

protein-solvent interface is distorted and some buried hydrophobic residues become accessible to solvent resulting in a larger SASA. Our results are consistent with an earlier report, which suggests that protein destabilization in solution is associated with an increase in SASA.[142] **Figure 3.2c** presents the effect of E_z on the numbers of hydrogen bonds between lysozyme molecules, and between lysozyme and water molecules, respectively. The intra-/inter-molecular hydrogen bonding is a key factor to stabilize protein crystal. As seen, the external E_z leads to a reduction in the number of hydrogen bonds between lysozyme molecules, in contrast, an increase between lysozyme and water molecules.

Lysozyme consists of several stable structural domains (e.g. α-helices and β-sheets) and its stability is well maintained by the intramolecular hydrogen bonds and four disulfide bonds.[101,143] The impact of electric field is further demonstrated by the evolution of lysozyme secondary structures as a function of time, which was evaluated using the DSSP (database of secondary structure assignments) algorithm.[144] As shown in **Figure 3.3**, the secondary structures including α-helices and β-sheets are generally well maintained up to E_z = 0.4 V/nm, despite some variations. For example, the number of residues (5^{th}–17^{th} and 25^{th}–35^{th} residues) forming two α-helices becomes smaller as the electric field increases, implying the partial loss of the secondary structures. Under a strong electric field, a handful of the secondary structures are shifted and protein is re-arranged to some extent in order to adopt a favorable conformation.

We point out that two factors, cross-linking and polarizable charges, were not included in our simulations. Protein crystals are usually cross-linked among each other to enhance the thermal and mechanical stability. The inclusion of cross-linking could practically improve protein stability even at a high electric field. The charges in

GROMOS96 force field for protein are fixed and not allowed to change. However, the charges could be polarized and vary with the electric field. In future study, these factors need to be taken into account.

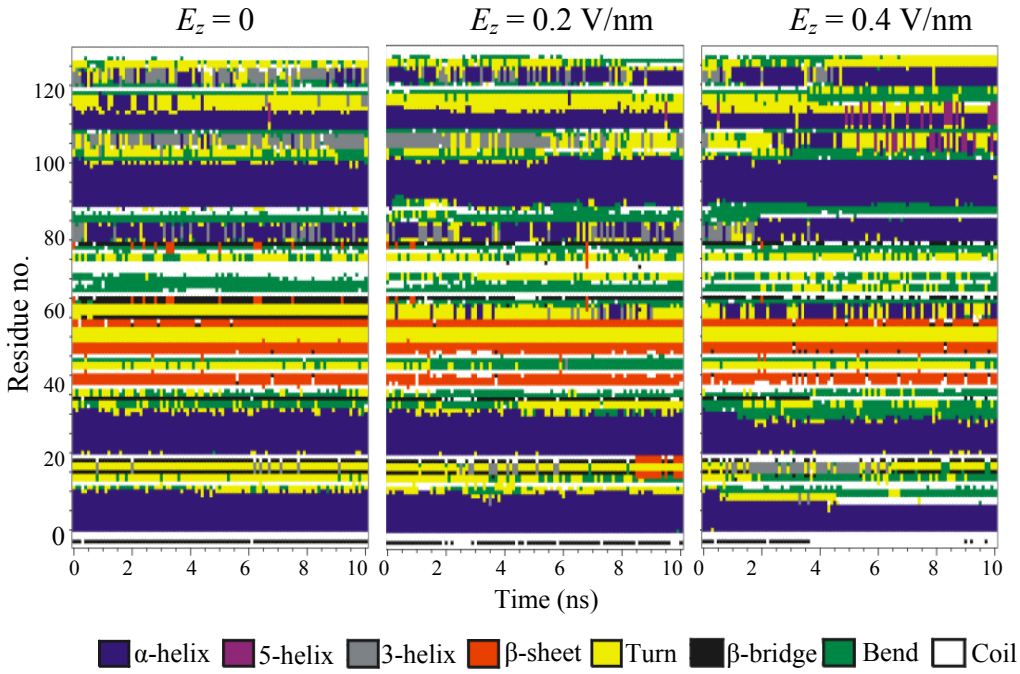

Figure 3.3. Evolution of lysozyme structures as a function of time at three electric fields E_z = 0, 0.2 and 0.4 V/nm, respectively.

3.3.2 Structures of Water and Ions

Water is a highly polar molecule and its behavior could be significantly altered upon exposure to an electric field. Recently, Cramer et al. performed MD simulation for water adsorbed on a polar surface under an external electric field and found the sudden formation of a water pillar above a threshold field of 1.2 V/nm as a result of the competition between orientational confinement and electric field.[145,146] In our study, we examine the orientation of the dipole moment of water in the lysozyme crystal. The dipole moment of water is defined by the vector directing from OW atom to the middle point of two HW atoms, as illustrated in the inset. The probability distribution function $P(\theta)$ was calculated by[128]

$$P(\theta) = \frac{\langle \delta N(\theta) \rangle}{\sin\theta} \tag{3.1}$$

and then normalized by $\sum P(\theta)$, in which $\langle \delta N(\theta) \rangle$ is the ensemble averaged number of water molecules within an angle ranging from $\theta - \delta\theta$ to $\theta + \delta\theta$ ($\delta\theta$ = 0.5° in our calculation). **Figure 3.4a** shows the probability distribution function $P(\theta)$ as a function of angle θ between the dipole moment of water and the z axis.

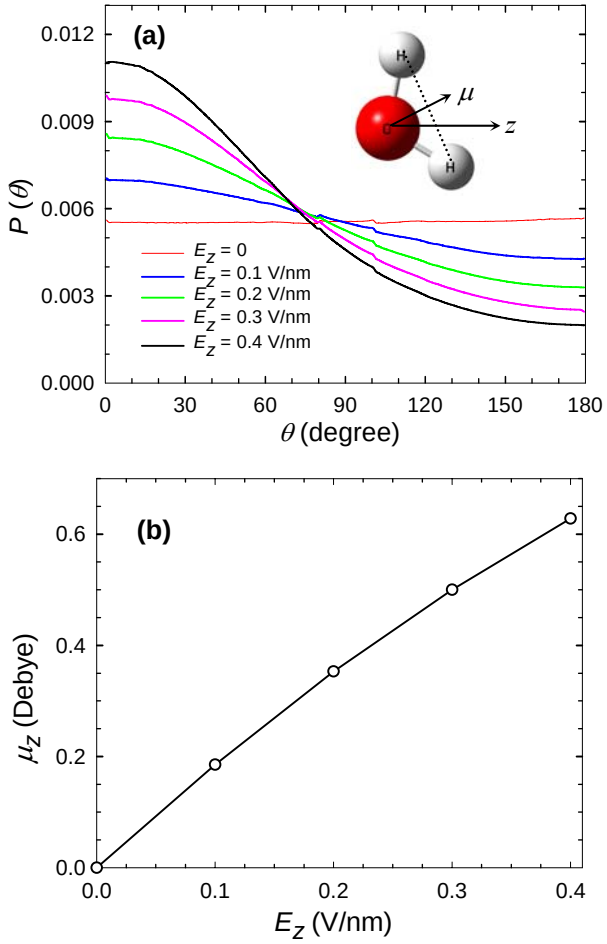

Figure 3.4. (a) Probability distribution function of angle θ between the dipole moment of water and the z axis. (b) Dipole moment of water along the z axis as a function of the electric field strength.

At E_z = 0 (from EMD simulation), the dipole moment shows an equally probabilistic distribution at any θ over the range 0–180°. That is, water molecules

exhibit a disordered orientation in the complex environment of lysozyme crystal, in which the distribution of nanopores is highly heterogeneous. Upon increasing E_z, $P(\theta)$ starts to exhibit a sigmoid shape with a larger probability at θ between 0 and 90°. This implies that the dipole moment tends to orient preferentially along the z axis under the electric field, and the preference becomes more distinct with increasing field strength. A quantitative estimation of the orientation was characterized by calculating the dipole moment along the z axis (μ_z). As shown in **Figure 3.4b**, μ_z increases with the electric field strength approximately in a linear fashion. The largest μ_z (at E_z = 0.4 V/nm) is about 25% of the dipole moment of SPC water molecule (2.27 Debye).[106] The preferred orientation of polar molecules under an electric field may influence the permeation process through protein crystals or biomembranes;[145] therefore, the electric field could be used for tuning transport rates.

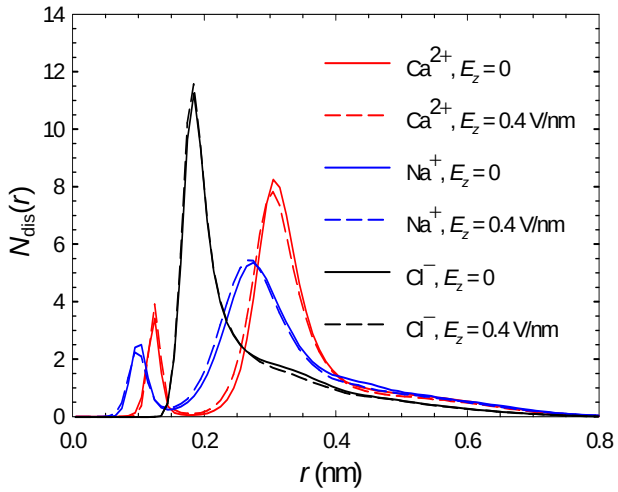

Figure 3.5. Number distributions of ions as a function of the distance from protein surface at E_z = 0 (solid lines) and E_z = 0.4 V/nm (dashed lines), respectively.

The binding of an ion, also called cofactor, is essential for the functionality of highly selective catalytic enzymes. It is instructive to examine ion distributions

around protein surface. Using Eq. (2.3), we calculated the number distributions $N_{\mathrm{dis}}(r)$ of Cl^-, Na^+ and Ca^{2+} around the protein surface. **Figure 3.5** shows the number distributions at two cases, E_z = 0 and 0.4 V/nm. The distributions of all ions are apparently independent of the electric field. While Cl^- exhibits only one pronounced peak at r = 0.19 nm, Ca^{2+} has two peaks at r = 0.12 and 0.31 nm; Na^+ also has two peaks at r = 0.10 and 0.27 nm. The atomic radii of Ca^{2+}, Na^+ and Cl^- are 0.14, 0.13 and 0.23 nm, respectively (see **Table 3.1**), slightly larger than the positions of the first peaks for the three ions. This indicates that some ions of each type are closely bound to lysozyme as a result of the electrostatic interactions with the oppositely charged residues in lysozyme. In particular, the first peak of Cl^- (11.4) is much higher than those of Ca^{2+} (3.7) and Na^+ (2.3) because lysozyme is overwhelmingly positively charged and has a significantly stronger affinity to Cl^-. From a rough estimation based on the number distribution profile, about 10% Ca^{2+} or Na^+ cations are located within the first layer, much less than the percentage of Cl^-. Both Ca^{2+} and Na^+ exhibit the second peaks, which are much higher than the first peaks and attributed primarily to their interactions with water molecules. Water forms a hydration layer with a thickness of about 0.3 nm around the protein surface and there is a minimum in its local density profile at $r \approx 0.3$ nm. Therefore, Ca^{2+} and Na^+ ions can readily intercalate into this less packed region and show the second peaks at $r \approx 0.3$ nm. Nevertheless, Cl^- has a larger radius compared to Ca^{2+} and Na^+ and thus is largely prohibited to access this region. Similar behavior was observed in previous studies of electrolytes in nanopores.[147-149]

Figure 3.6 shows the water coordination numbers (also called hydration numbers) of three ions and the Cl^- coordination numbers of Ca^{2+} and Na^+ cations as a function

of the electric field strength. The coordination numbers were estimated by integrating the radial distribution function up to the first minimum position, as indicated in **Figure 3.6**. Both water and Cl^- coordination numbers of ions change marginally with the electric field. This is also true for the radial distribution functions between any two species (not shown), which are nearly independent on the electric field. These structural properties are essentially the *static* not *dynamic* characteristics of the system, thus they are highly conserved even at a strong electric field. The hydration number of ion is relevant to ionic charge and interaction with water. The bivalent Ca^{2+} strongly interacts with water molecules via the electrostatic and van der Waals interactions, and has the largest hydration number, while Na^+ has the smallest. These are also observed for ion mixture (0.50 M Ca^{2+}, 0.50 M Na^+ and 1.50 M Cl^-) in an aqueous bulk solution as in **Table 3.2**, but with larger hydration numbers for all ions. Although Cl^- can interact with Ca^{2+} more strongly than with Na^+, counterintuitively, Cl^- coordination number of Na^+ (2.24) is almost twice as that of Ca^{2+} (1.13). Comparing to the case in aqueous bulk solution, Cl^- coordination number is increased in the crystal, particularly for Ca^{2+}. Water and Cl^- interact with Ca^{2+} and Na^+ in a competitive way and thus there is a counterbalance between water and Cl^- coordination numbers. The number of water molecules in the system is overwhelmingly larger than that of Cl^-; consequently, Ca^{2+} are strongly hydrated by water rather than by Cl^- and has a weaker interaction with Cl^- compared to the less hydrated Na^+. Overall, the hydrated Na^+ has a larger Cl^- coordination number than the hydrated Ca^{2+}.

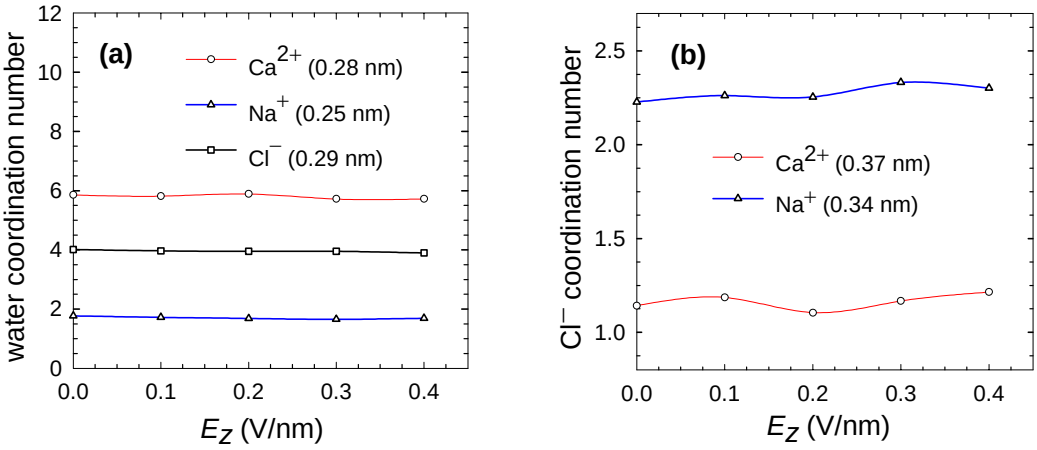

Figure 3.6. (a) Water coordination numbers of Ca^{2+}, Na^{+} and Cl^{-}. (b) Cl^{-} coordination numbers of Ca^{2+} and Na^{+}. The first minimum positions in the radial distribution functions are indicated in the parenthesis.

Table 3.2. Water and Cl^{-} coordination numbers and self-diffusivities D_z in lysozyme crystal ($E_z = 0$) and in aqueous bulk solution, respectively.

	H_2O coordination no.		**Cl^- coordination no.**		**D_z (10^{-9} m^2/s)**	
	In crystal	In solution	In crystal	In solution	In crystal	In solution
Ca^{2+}	5.95	7.67	1.13	0.16	0.018	1.09
Na^{+}	1.80	3.69	2.24	1.06	0.037	1.13
Cl^{-}	4.01	6.17	–	–	0.056	1.58

3.3.3 Ion Mobility

In the absence of electric field, ion mobility along the pore axis (e.g., the z direction) was examined by calculating the ensemble averaged mean-squared displacement (MSD_z) using Eq. (2.4). The multiple-origin method was used to improve the statistical accuracy. Subsequently the self-diffusivity in the z axis was calculated using Eq. (2.5).

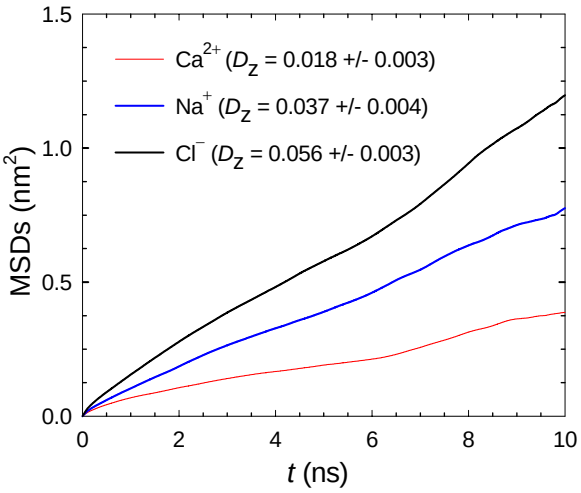

Figure 3.7. MSDs of ions along the z axis from EMD simulation. The estimated self-diffusivities D_z are given in parenthesis with a unit of 10^{-9} m^2/s.

Figure 3.7 shows the MSDs of the three ions from EMD simulation. The estimated diffusivities indicated in the parenthesis are 0.018×10^{-9} m^2/s for Ca^{2+}, 0.037×10^{-9} m^2/s for Na^+ and 0.056×10^{-9} m^2/s for Cl^-, which are two orders of magnitude smaller comparing to the diffusivities in aqueous bulk solution (in **Table 3.2**). Such a dramatic reduction in the mobility of ions in the lysozyme crystal is attributed to the interaction and steric obstacle of the lysozyme molecules.

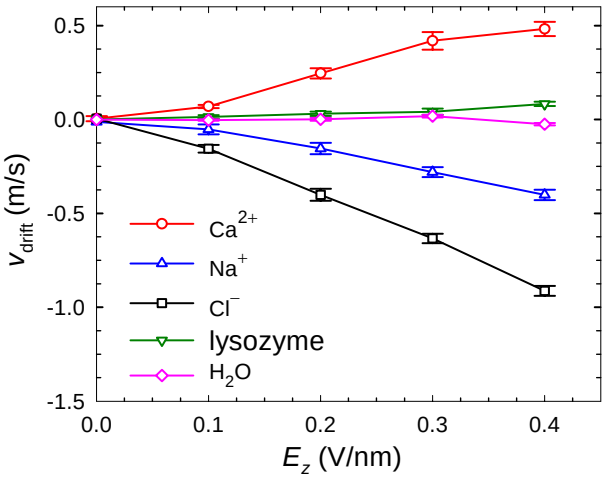

Figure 3.8. Drift velocities of ions, lysozyme and water along the z axis as a function of the electric field.

Under the electric field, ions are driven to migrate in the nanopore of the lysozyme crystal. At the steady state, each type of ion moves at a constant velocity, referred to as drift or streaming velocity. **Figure 3.8** shows the effect of electric field on the drift velocities v_{drift} of the three ions, water and lysozyme, respectively. The drift velocities were computed by tracking the centers-of-mass over time. It is known that solvent can migrate in terms of the electro-osmotic flow dragged by ionic stream.[150] Nevertheless, in this study water was found to have a negligible drift velocity due to the electroneutrality of water molecule. On average, the dragging forces from positive and negative ions on water are cancelled out. Lysozyme carrying positive charges exhibits a slight positive displacement, albeit the magnitude is negligible due to the high molecular weight. As expected, Ca^{2+} moves at a positive drift velocity, Cl^- migrates at a negative velocity. Interestingly, Na^+ also moves at a negative velocity like Cl^-. The electrophoretic mobility of ion is not only driven by the external electric field, but affected by the interactions with protein atoms, the hydration states, the dragging and resisting forces of neighboring ions, etc. As a counterbalance of all these complex factors, Cl^- moves at a faster negative velocity under the electric field. The driving force by the electric field and the dragging force by the Cl^- ionic stream have the opposite directions for Na^+ and Ca^{2+}. As indicated in **Figure 3.6b**, Na^+ interacts with Cl^- more strongly than Ca^{2+} and has a larger Cl^- coordination number. As a consequence, the Cl^- ionic stream drags Na^+ to flow oppositely to the direction of the electric field. However, Ca^{2+} is imposed by a stronger electrical force than Na^+ and the Cl^- coordination number of Ca^{2+} is smaller, therefore, Ca^{2+} exhibits a positive drift velocity. In other words, the dragging force by the Cl^- ionic stream dominates the electrical force for Na^+, but the reverse is true for Ca^{2+}. Similar phenomenon was

observed for an electrolyte mixture in a silicon nano-channel under an electric field, in which K^+ and Cl^- migrated in the same direction.[148,149] In that work, ionic flux was decomposed into two items, namely, an electrical migration component driven directly by the electric field, and a convection component dragged by the solvent stream. The strong convection component dominated and resulted in the migration of Cl^- in the reversed direction. Slightly differently, our results here show that the direction of ion migration can also be reversed by ions. In our study, the migration of Na^+ is affected by the stream of Cl^-, and thus the former migrates along the direction opposite to the electric field.

3.3.4 Electrical Conductivity

Measurement of electrical conductivity in protein crystal is of significance for biosensing and ion exchanging. An earlier work studied the conductivity in a dry lysozyme crystal with orthorhombic symmetry and observed an Arrhenius relationship between the conductivity and temperature.[151] The movement of trace proton or electron was regarded to dominate the conduction process in dry crystal, and the conductivity was found to be very low (10^{-7}–10^{-9} S/m). Conductivity was also examined for tetragonal lysozyme crystal immersed in NaCl solution.[53] In this case, the conductivity was greatly enhanced to 10^{-2}–10^{-3} S/m.

Electrical conductivity κ can be estimated in two different ways, separately, from EMD and NEMD simulations. In EMD, κ is obtained from the classical Nernst-Einstein equation[152]

$$\kappa = \frac{e^2}{k_B T} \sum_i D_i \rho_i |q_i|^2 (1-\Delta) = \frac{F_a^2}{RT} \sum_i D_i c_i |q_i|^2 (1-\Delta) \tag{3.2}$$

where e is the electron charge, T is the absolute temperature, k_B the Boltzmann constant, ρ_i is the number density, c_i is the molar concentration, D_i is the self-diffusivity, q_i is the charge of ith species, and Δ is the cross correlation term which reflects the correlations between different species. F_a is the Faraday constant equal to 9.6485×10^4 C/mol and R is the gas constant. In the Nernst-Einstein equation, conductivity is assumed to be directly proportional to self-diffusivity. The cross term Δ is usually small and can be reasonably neglected. The conductivity in the lysozyme crystal studied here is estimated to be 0.61 ± 0.05 S/m.

In NEMD simulation, the conductivity is obtained from the relationship between the electrical current and the electric field strength. In our study, the current density J_z along the z axis can be expressed as[152]

$$J_z = F_a \sum_i c_i q_i v_{i,z} \tag{3.3}$$

where $v_{i,z}$ is the velocity along the z axis. Within the linear response range, the zero-field conductivity κ can be extrapolated from

$$\kappa = \lim_{E_z \to 0} \frac{J_z}{E_z} \tag{3.4}$$

Figure 3.9 shows the current density *Jz* as the function of the electric field strength. A very good linear relationship is observed in the range of the electric field from our NEMD simulation. The zero-field conductivity κ of the system is estimated as 0.56 ± 0.03 S/m, which is close to 0.61 ± 0.05 S/m, within the statistical uncertainty, determined from EMD simulation based on the Nernst-Einstein equation.

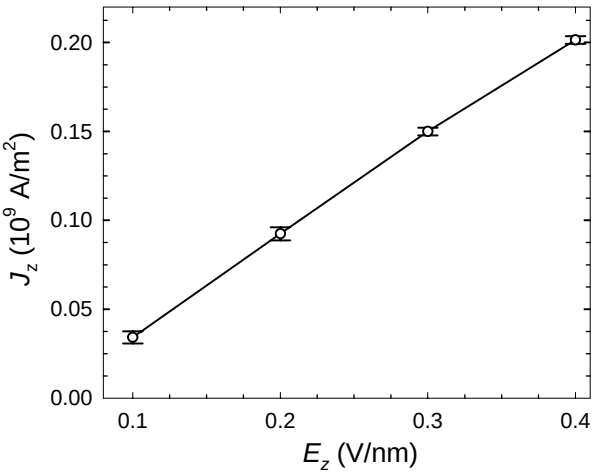

Figure 3.9. Electrical current density along the *z* axis as a function of the electric field.

3.4 Conclusions

From MD simulations, we have investigated the electrophoretic flow in a tetragonal lysozyme crystal. With increasing electric field strength, the root-mean-squared deviations of lysozyme atoms are enhanced, more hydrophobic residues are exposed to water, and the secondary structures are slightly destroyed. All these indicate that the stability of lysozyme is reduced to some extent by the external electric field. The dipole moment of water in the pore axis increases linearly with the field strength due to the reorientation of water molecules. As a counterbalance between the favorable interaction with protein and the size effect, cations exhibit two pronounced layers around lysozyme surface, while Cl^- has only one enriched layer. The electric field has a negligible effect on the static and structural properties of ions, such as the number distributions and coordination numbers. Compared to aqueous bulk solution, the confinement in the crystal remarkably reduces ion mobility. The drift velocities of ions exhibit approximately a linear relationship with the field strength. The movement of Ca^{2+} and Cl^- is primarily dominated by the electric field; nevertheless, Cl^- ion stream also affects the movement of Na^+ as a consequence of

coordination. The electrical conductivity of the system predicted from NEMD and EMD is in good agreement. The simulation results provide a deeper understanding of the microscopic behavior of protein, water and ions under electric field. This is important to elucidate the fundamental mechanism of transport in biomembranes and useful for the emerging applications of protein crystals as new biosensors and bioseparation media.

Chapter 4 Separation of Amino Acids in a Glucose Isomerase Crystal

4.1 Introduction

Liquid chromatography is one of the major techniques for the analysis and separation of organic and biological molecules.[153] Despite the wide applications in both academic and industrial communities, many fundamental issues on chromatographic retention processes remain obscure. The operation conditions are often adjusted by empirical "trial and error". This is largely due to the complexity of chromatographic systems and the lack of detailed microscopic understanding. Conventional experimental methods are unable to examine this level of complexity from the microscopic scale. To facilitate the rational design of new column materials and the optimization of separation processes, it is indispensable to obtain insight into chromatographic separation at the molecular level. In this regard, molecular simulations have played an increasingly important role as they can provide atomistic/molecular pictures that would otherwise be experimentally intractable or impossible to obtain. Compared to numerous experimental investigations in liquid chromatography, few simulation studies have been reported. MC simulations were carried out to explore the retention mechanism and the influence of various stationary phases in reversed-phase liquid chromatography.[154] MD simulations were used to examine solute conformations at chiral surfaces embedded with selectors in different solvents,[154-158] and solvent structures and solute diffusion at chromatographic interfaces.[159-161] A clearer and deeper understanding is desirable from the molecular

level as novel stationary phases are being developed to separate unusual mixtures that are difficult to separate by common systems.

Conventional materials utilized as stationary phase in a liquid chromatographic column are inertial supports coated with polymers or biopolymers. In chiral and affinity separation, for instance, proteins such as lysozyme, bovine/human serum albumin and glycoproteins are usually immobilized on solid supports.[65] However, the utilization of support matrix results in a low volumetric specific activity of proteins and thus decreases the separation efficacy. Recently, it was observed that separation efficacy can be largely improved if crystalline proteins are directly packed in chromatographic column.[5] Furthermore, the compact arrangement of protein molecules in crystalline phase inhibits protein unfolding and thus maintains their native conformations more effectively even at elevated temperatures or in organic solvents as compared to amorphous proteins.[1,2] Therefore, protein crystals have been proposed as a class of novel stationary phases in liquid chromatography to separate mixtures of amino acids, nucleoside, or racemic drug enantiomers. Known as bioorganic zeolites, protein crystals possess many fascinating features, including high porosity (0.5–0.8), large surface area (800–2000 m^2/g), and a wide range of pore size (1.5–10 nm).[1] The inherently chiral nature of *L*-amino acids as building blocks of proteins creates an asymmetrical environment, which could lead to the selective separation of pharmaceutically important enantiomers. Decades ago, the applications of protein crystals were severely restricted by their fragility and unpredictable growth patterns. These problems have to some degree been solved and mechanically stable protein crystals can be made by cross-linking techniques.[1,2] Cross-linking can also enhance the chemical stability of proteins in crystalline form. For instance, the cross-linked glucose isomerase crystal is about five times more stable than the native form

in high substrate solution.[8] Several techniques have been reported to effectively produce cross-linked protein crystals with high yield and good quality or on a large scale.[18-20] As a consequence, protein crystals have been increasingly utilized as new separation materials.[1,2,22,64,162,163]

A handful of experimental and computational studies have been reported in order to elucidate the underlying physics of fluids confined in protein crystals. Separation of mixtures differing in size, charge, hydrophobicity and chirality was examined in protein crystals.[1] Diffusion in protein crystals was found to depend on steric repulsion, surface affinity, electrostatic interaction and cross-linker.[53,55-57] In Chapters 2 and 3, the spatial and temporal properties of water and ions were studied in protein crystals with different morphologies and under electric field. Although endeavors have been taken to explore the properties and applications of protein crystals as stationary phase in liquid chromatography, the underlying mechanisms are not fully understood. Many complex factors such as temperature, pH, ionic strength, modifier type, all come into play. A set of guidelines on how to select a protein crystal and to optimize operation conditions are crucial to the new development of technically feasible and economically competitive separation technologies using protein crystals.

The objective of this work is to examine the capability of a glucose isomerase (GI) crystal to separate the mixtures of amino acids using MD simulations. The experimental liquid chromatographic separation process is mimicked, in which GI crystal acts as a stationary phase and water as a mobile phase.[5] GI, also called *D*-xylose isomerase, is an extremely stable enzyme against pH (6–9) and temperature. It is used in industrial processes at temperatures around 60 °C to catalyze the reversible isomerization of *D*-glucose and *D*-xylose to *D*-fructose and *D*-xylulose, respectively.[164] GI is the first enzyme crystallized on an industrial scale, Visuri

developed an industrial-scale process for generating crystalline GI crystals over one decade ago.[165,166] Cross-linked GI crystals have been utilized in chemical synthesis of polar solvents and chiral separation.[162,163] Experimental studies have evidenced that cross-linked GI crystals could be used as stationary phase in liquid chromatography to separate mixtures of amino acids or nucleosides, or *D/L* pairs of amino acids or arabitol.[5,167] However, there is yet no computational study to explore the separation mechanisms involved. Therefore, insightful information can be obtained from our simulation study. In the next section, we begin with a description of models and methods used to simulate mixtures of three amino acids. The amino acids considered are arginine (Arg), phenylalanine (Phe), and tryptophan (Trp) differing in size, charge and hydrophobicity. Then, the transport velocities of three amino acids in GI crystal are compared with experimental observation and the underlying mechanisms for separation are elucidated. Finally, the concluding remarks are summarized.

4.2 Models and Methods

Figure 4.1 illustrates the model used in this work for the separation of Arg, Phe and Trp in GI crystal. The GI crystal has an I_{222} space group and lattice constants a = 9.388 nm, b = 9.968 nm and c = 10.290 nm. The crystal structure was constructed based on the experimental crystallographic data (PDB ID: 1XIB).[168] The largest pore extending along the z direction is nearly cylindrical with a radius of 1.2–1.8 nm and the pores in x and y directions are relatively small and irregular. Simulations were carried out at the physiological condition (pH ≈ 7), which was very close to the pH in the experimental crystallization of GI (7.4)[168] and in the chromatographic separation process (7.0).[5] Consequently, Arg and Lys residues were protonated, while Asp and Glu residues were deprotonated, leading to twenty negative unit charges in a GI

molecule. The simulation box contained two unit cells (1 × 1 × 2) and the periodic boundary conditions were exerted in all three directions.

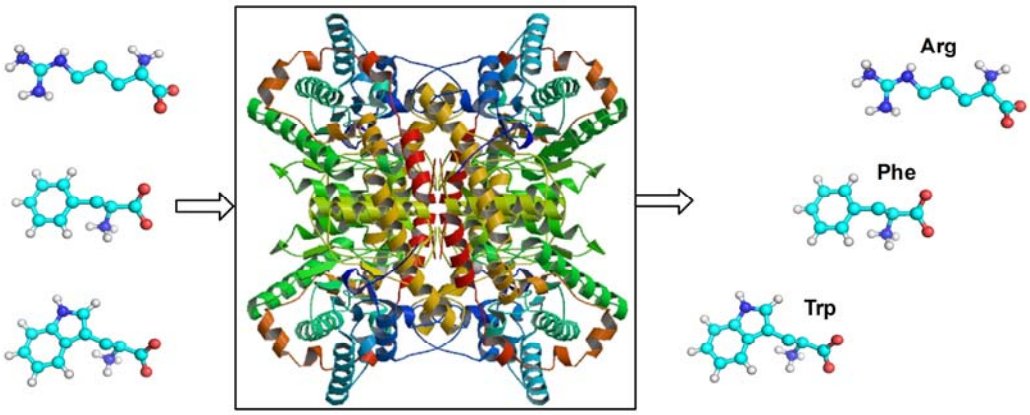

Figure 4.1. Schematic illustration for the separation of Arg, Phe and Trp in glucose isomerase crystal. Water and ions are not shown for clarity.

The structures and characteristic parameters of the three amino acids Arg, Phe and Trp are shown in **Figure 4.1** and **Table 4.1**, respectively. Arg carried one positive unit charge, while Phe and Trp were neutral. In order to examine the effect of solute concentration, simulations were conducted with different numbers (N_{AA} = 20, 40, and 80 respectively) of each amino acid. In addition, a certain number of Na^+ ions were introduced randomly for electroneutrality and H_2O molecules were then added to mimic a fully hydrated GI crystal. All simulations were performed in a canonical (NVT) ensemble and the system density was approximately 1100 kg/m^3.

Table 4.1. Characteristic parameters of Arg, Phe and Trp.

	Formula	Molecular Weight	van der Waals Volume (nm^3)	Charge (*e*)
Arg	$C_6H_{14}N_4O_2$	174.2	0.139	+1
Phe	$C_9H_{11}NO_2$	165.2	0.140	0
Trp	$C_{11}H_{11}N_2O_2$	204.2	0.170	0

[a] pH value was set at 7 and the three amino acids were in zwitterionic state.
[b] The van der Waals volume was defined as the volume one molecule occupied based on the atomic collision diameters, which were adopted from GROMOS96 43a1 force field. The grid method was used with a cubic grid size being 0.002 nm.

The GI and amino acids were modeled by GROMOS96 force field,[105] and water by SPC/E model.[122] GROMACS v3.3.1 package was used to perform MD simulations because it is extremely fast and particularly well-suited for biomolecular systems.[51] Some MD simulation settings were same as in Chapter 2 and thus not described here. The system was initially subject to energy minimization using the steepest descent method with a maximum step size 0.01 nm and a force tolerance 10 kJ mol^{-1} nm^{-1}. Then velocities were assigned according to the Maxwell-Boltzmann distribution at 323 K. Non-equilibrium MD simulations were performed at 323 K as in experimental separation process.[5] The temperature was controlled by the Berendsen thermostat with a relaxation time of 0.1 ps. Integration time step was 2 fs and trajectory was saved for 30 ns at 2 ps interval. The C_α atoms of GI molecules were fixed and GI molecules could move only slightly thus acting as a stationary phase. A constant external acceleration a_{ext} was exerted on water molecules so that water flowed as a mobile phase. In order to examine the effect of the external force, three accelerations (0.01, 0.02, and 0.04 nm/ps^2) were considered. In total, five runs were performed (1) N_{AA} = 80, a_{ext} = 0.04 nm/ps^2; (2) N_{AA} = 80, a_{ext} = 0.02 nm/ps^2; (3) N_{AA} = 80, a_{ext} = 0.01 nm/ps^2; (4) N_{AA} = 40, a_{ext} = 0.02 nm/ps^2; (5) N_{AA} = 20, a_{ext} = 0.02 nm/ps^2. It is noteworthy that the external forces used were typically strong. For instance, a constant external acceleration 0.02 nm/ps^2 corresponds to a force 3.2×10^{11} N applied on one mole of water. Such a strong driving force is commonly used in non-equilibrium MD simulations to reduce thermal noise and therefore to enhance the signal/noise ratio within a nanosecond timescale.[169]

Amino acid molecules were dragged by flowing water to transport through the pores in GI crystal. The drift velocities of water and amino acids were tracked to assure the system reaching a steady state. Note that it is subtle to maintain

temperature in non-equilibrium MD simulations since the friction heat generated by flux must be removed.[139] To calculate temperature, the mean velocity of each group (e.g., water and amino acids) was subtracted from the instantaneous velocity. In all simulation, temperature was kept fairly well around 323 K with a small fluctuation of approximately 0.5 K. The steady state was found to reach in less than 10 ns, and the subsequent 20 ns trajectory was used for analysis. The displacements of amino acids, interaction energies, hydrogen bond numbers, and solvent-accessible surface areas were estimated using built-in tools in GROMACS, and other properties including van der Waals volumes, number distributions, hydration numbers, and residence times were obtained using in-house developed codes.

4.3 Results and Discussion

4.3.1 Effects of Solute Concentration and Solvent Flowing Rate

Figure 4.2 shows the displacements of amino acids in *x*, *y* and *z* directions as a function of time from five runs. The displacements were calculated based on the centers-of-mass of amino acid molecules. We can see that Arg transports the fastest and followed by Phe and Trp. The solvent flowing rate and solute concentration do not qualitatively alter the elution order of three amino acids, though they influence the overall velocities ($v = \sqrt{v_x^2 + v_y^2 + v_z^2}$) of solute and solvent as listed in **Table 4.2**. With deceasing flowing rate from a_{ext} = 0.04 to 0.01 nm/ps^2 (runs 1–3), the velocities of both solute and solvent become smaller; the ratio v_{Arg}:v_{Phe}:v_{Trp} changes from 1:0.37:0.23 to 1:0.35:0.06. Specifically at a_{ext} = 0.01 nm/ps^2, the velocity difference between three amino acids is the largest. This implies that the separation efficiency is improved by decreasing flowing rate.

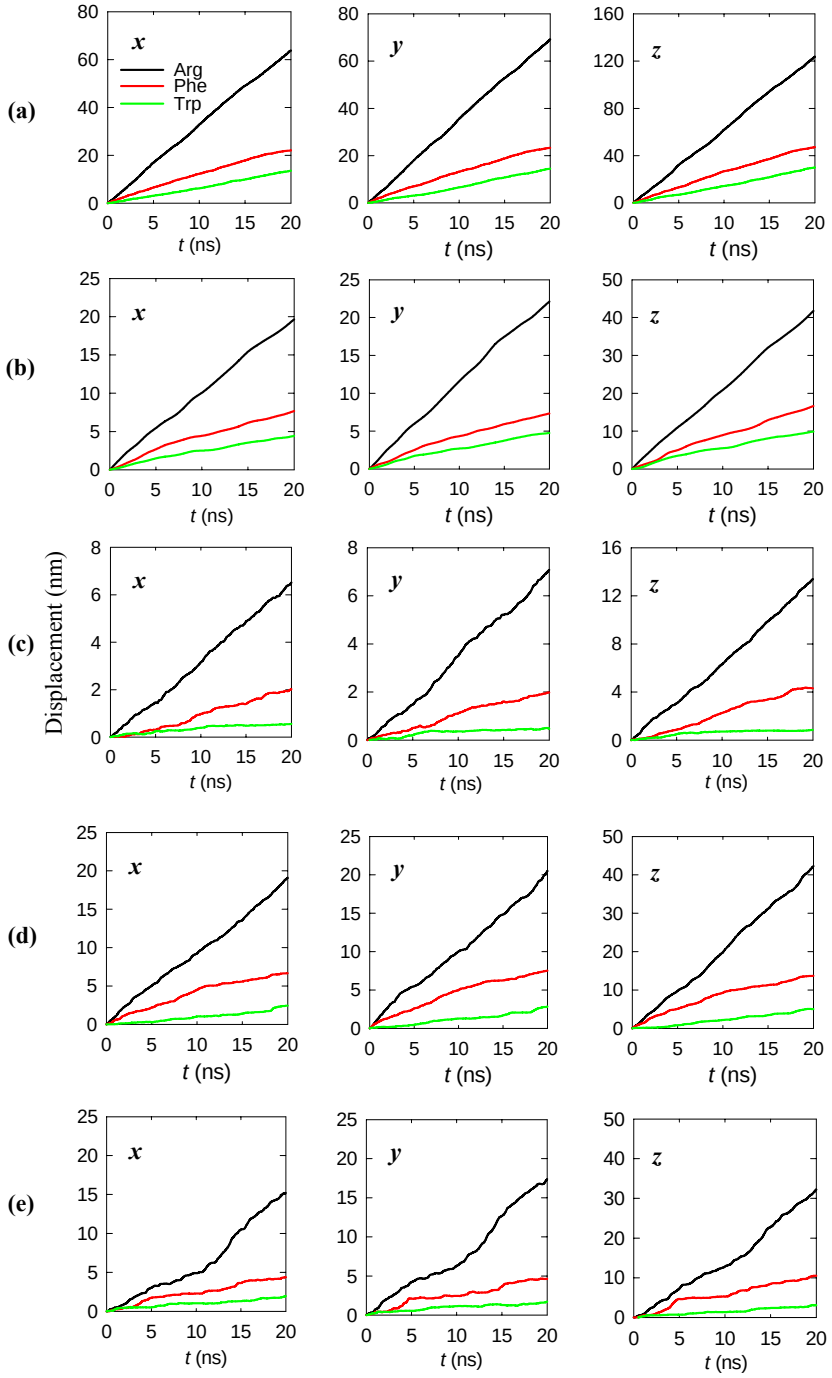

Figure 4.2. Displacements of amino acids in x, y and z directions as a function of time from (a) run 1 (N_{AA} = 80, a_{ext} = 0.04 nm/ps^2), (b) run 2 (N_{AA} = 80, a_{ext} = 0.02 nm/ps^2), (c) run 3 (N_{AA} = 80, a_{ext} = 0.01 nm/ps^2), (d) run 4 (N_{AA} = 40, a_{ext} = 0.02 nm/ps^2), and (e) run 5 (N_{AA} = 20, a_{ext} = 0.02 nm/ps^2).

With decreasing N_{AA} = 80 to 20 (runs 2, 4 and 5), the velocity decreases for amino acid but increases for water; the ratio v_{Arg}:v_{Phe}:v_{Trp} changes from 1:0.38:0.23 to 1:0.30:0.09. Such a trend suggests that a lower concentration is more efficient for separation. Consequently, both flowing rate and solute concentration could be tuned to optimize the separation process. Also observed in **Figure 4.2** is that the displacement curve is not smooth at a low solute concentration. This is attributed to the small number of solute molecules which leads to large fluctuations. Therefore, it is necessary to have a sufficiently large number of solute molecules in simulations.

To provide insight into the retention mechanisms, the dynamic and static properties presented below are based on run 2 unless otherwise stated.

Table 4.2. Overall velocities from five runs.

	Runs				
	1	**2**	**3**	**4**	**5**
N_{AA}	80	80	80	40	20
a_{ext} (nm/ps^2)	0.04	0.02	0.01	0.02	0.02
v_{Arg} (m/s)	7.826	2.545	0.823	2.490	1.966
v_{Phe} (m/s)	2.900	0.956	0.285	0.817	0.598
v_{Trp} (m/s)	1.799	0.576	0.051	0.312	0.177
v_{water} (m/s)	13.223	6.006	2.876	6.910	7.229
v_{Arg} : v_{Phe} : v_{Trp}	1:0.37:0.23	1:0.38:0.23	1:0.35:0.06	1:0.33:0.13	1:0.30:0.09

4.3.2 Directional Velocities

Table 4.3 lists the directional velocities of amino acids and water. For each amino acid, the velocity in the *x* direction is approximately equal to that in the *y* direction; both are about half of the velocity in the *z* direction. This is because the largest pores of GI crystal in the *z* direction are wider than in the other two directions. The three amino acids differ in molecular weight, van der Waals volume and charge (see **Table**

4.1), but they exhibit similar anisotropic transport reflected in v_x:v_y:v_z. This implies that anisotropic degree in the transport of amino acid in GI crystal does not solely depend on solute size, steric effect or hydrophobicity. From **Table 4.2**, the anisotropic transport is also observed for water. As mentioned earlier, the velocity of water in our simulation system was greater than the flowing velocity in a practical liquid chromatography. The driving force for water in simulation is sufficiently large in order to reduce thermal noise and high velocity improves the sampling efficiency of simulation in an affordable time scale. Therefore, solute molecules dragged by high-speed flowing water could translocate through a long distance and experience various heterogeneous locations in the crystal.

Table 4.3. Directional velocities, nonbonded interaction energies, entry numbers and residence times from run 2.

	Arg	**Phe**	**Trp**	**Water**
v_x (m/s)	0.976	0.365	0.214	2.509
v_y (m/s)	1.116	0.353	0.230	2.527
v_z (m/s)	2.069	0.810	0.483	4.837
v_x: v_y: v_z	0.47:0.54:1	0.45:0.44:1	0.44:0.48:1	0.52:0.52:1
ΔE_1 (kJ/mol)	-590.3 ± 6.4	-326.8 ± 8.0	-324.7 ± 8.4	
ΔE_2 (kJ/mol)	-222.9 ± 4.1	-155.7 ± 4.6	-177.4 ± 3.6	
$\Delta\Delta E$ (kJ/mol)	-367.4 ± 5.3	-171.1 ± 6.3	-147.3 ± 6.0	
Entry number	442	428	377	
Residence time (ps)	26.6	34.2	41.4	
Hydration number	25.9	19.7	20.0	

Although the three amino acids exhibit similar degree of anisotropic transport, the velocity in each direction decreases in the order of $v_{\mathrm{Arg}} > v_{\mathrm{Phe}} > v_{\mathrm{Trp}}$. That is, Arg moves the fastest, while Trp moves the slowest. The overall velocity of Arg is about 1.7 and 3.4 times larger than those of Phe and Trp, respectively. The solute size and

hydrophobicity appear to play a key role in the observed difference in velocity. Compared with Phe and Trp, Arg is the most hydrophilic with five polar groups; consequently, it interacts with water the most strongly and is dragged by flowing water to move at the largest velocity. Furthermore, Arg has a small size and thus encounters a small steric retarding force. In contrast, Trp moves at the slowest velocity because it is highly hydrophobic and has the largest size. Phe has the close size as Arg and the similar hydrophobicity to Trp; consequently, Phe has a velocity between Arg and Trp. Quantitative analysis will be discussed in more details below based on interaction energies, structural and geometrical perspectives.

The decreasing order in velocity $v_{Arg} > v_{Phe} > v_{Trp}$ from our simulation is consistent with the experimental observation by Pastinen et al.[5] They measured the retention times of twenty amino acids separately in a chromatographic column packed with GI crystal. Although our simulation system contained the mixture of three amino acids and was not exact the same as in experiment, the same elution order $v_{Arg} > v_{Phe} > v_{Trp}$ was observed. The experiment suggested that the elution properties of amino acids are determined by hydrophobicity and/or other factors.[5] The distinct difference in the velocities of three amino acids reveals that they can be separated in GI crystal.

4.3.3 Interaction Energies

Figure 4.3 shows the interaction energies of amino acids with the mobile phase (water) and the stationary phase (protein). The contributions from LJ and Coulombic potentials were calculated separately. Na^+ ions were not included in the calculations for two reasons. First, the number of Na^+ ions was substantially smaller than those of water and protein atoms and its effect on solute transport is small. Second, it is a bit ambiguous to classify Na^+ ions as mobile phase or stationary phase. Some Na^+ ions

were strongly bound to protein, had negligible mobility, and could be regarded as the stationary phase along with GI crystal. However, other Na^+ ions were highly hydrated, moved with flowing water, and hence acted as mobile phase. Consequently, in the calculations of interaction energies the effect of Na^+ was not considered. For the interactions between amino acids and water (ΔE_1), the Coulombic potential dominates and the LJ potential is vanishingly small. Because of the strong Coulombic attraction, amino acid molecules are very adjacent to water leading to a slight repulsion in the LJ potential. For the interactions between amino acids and protein (ΔE_2), the Coulombic potential is also the major contribution. Among the three amino acids, Arg interacts with both water and protein most strongly because Arg is most hydrophilic and positively charged. Interestingly, all three amino acids interact with water more strongly than with protein. The differences $\Delta\Delta E$ (= ΔE_1 – ΔE_2) in the interaction energies with water and protein are reported in **Table 4.3**. The $\Delta\Delta E$ values for Arg, Phe and Trp are –367.4, –171.1 and –147.3 kJ/mol, following the order of velocity $v_{Arg} > v_{Phe} > v_{Trp}$. The strongest attraction between Arg and water determines its transport as Arg is dragged by the mobile phase to move at the highest velocity.

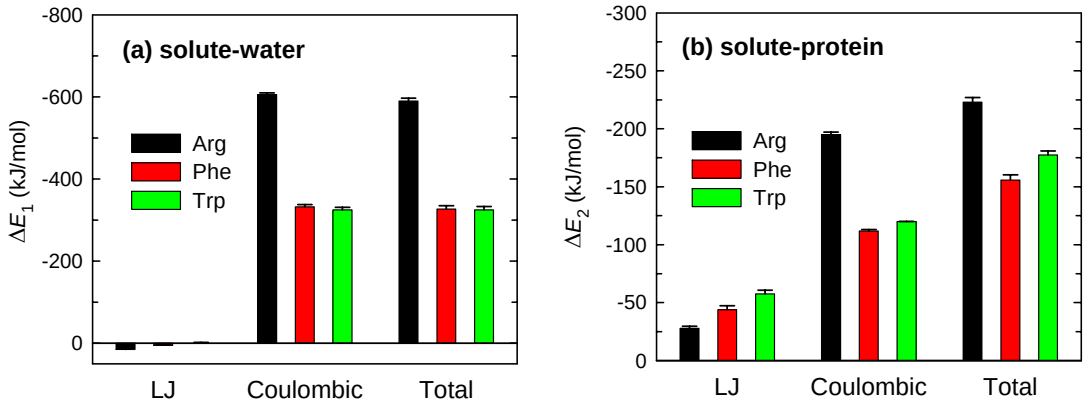

Figure 4.3. Nonbonded interaction energies of amino acids with (a) water and (b) protein.

Our simulation results reveal that the separation of amino acids in GI crystal is the consequence of counterbalance between solvophilic interaction with the aqueous mobile phase and protein-philic interaction with the stationary phase (especially the hydrophobic interactions with hydrophobic residues). Both interactions play an important role in the retention of amino acids in chromatography. Interestingly, this was also observed from an MC simulation in which the solvophilic interactions with mobile phase and the hydrophobic interactions with stationary phase were found to be important in reversed-phase liquid chromatography.[170] The interpretation here for the velocity difference of the three amino acids is based on energetic point of view, and we will further analyze the mechanisms from structural and geometrical perspectives, including number distributions, contact numbers, hydrogen bonds and surface areas.

4.3.4 Number Distributions and Contact Numbers

To examine the locations of amino acids around stationary phase (protein), we calculated the normalized accumulative number distribution $N_{\mathrm{acc}}(r)$ of amino acids as a function of distance *r* from the nearest protein atom. $N_{\mathrm{acc}}(r)$ is defined as

$$N_{\mathrm{acc}}(r) = \int_0^r N_{\mathrm{dis}}(r)\mathrm{d}r \tag{4.1}$$

where $N_{\mathrm{dis}}(r)$ is the number distribution defined by Eq. (2.3). In the calculations, the van der Waals radii of protein atoms were taken into account. The number distribution is a measurement to estimate how amino acid molecules are proximal to protein surface. As shown in **Figure 4.4a**, $N_{\mathrm{acc}}(r)$ are quantitatively different for Arg, Phe and Trp. We note that the interaction energy differences $\Delta\Delta E$ of amino acids are Arg > Phe > Trp. Consequently, Arg moves the fastest with flowing water and Trp moves the slowest. In other words, Trp stays the most closely to the stationary phase (protein). We also calculated the number distributions of water around amino acids

and found that water is the most adjacent to Arg. As a result, $N_{acc}(r)$ around protein follows the order Arg < Phe < Trp at a given r.

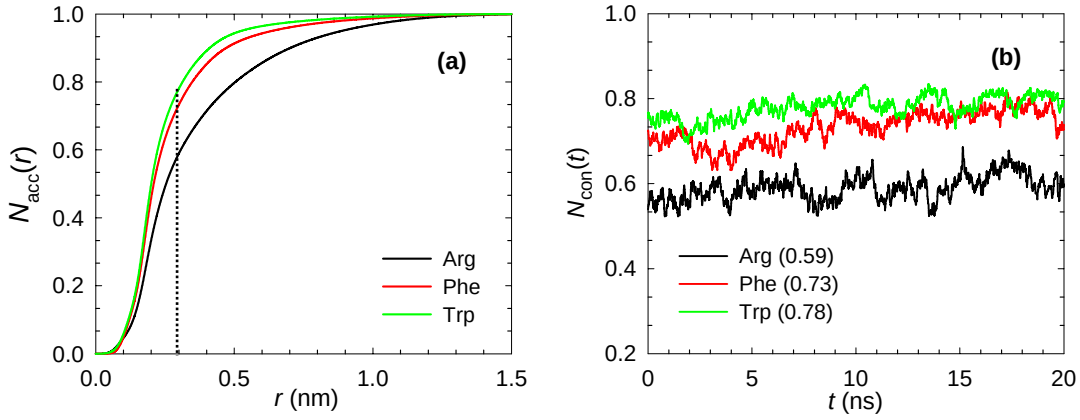

Figure 4.4. (a) Accumulative number distributions of amino acids as a function of distance from protein surface. The dotted line indicates r = 0.3 nm. (b) Contact numbers of amino acids as a function of time. Contact number is defined as the accumulative number at r = 0.3 nm from protein surface. The contact numbers averaged over time are shown in the parentheses.

If a contact layer is defined within the distance from protein surface to be less than 0.3 nm, $N_{acc}(r=0.3)$ gives the contact number. The thickness of the contact layer defined does not qualitatively affect the result. Here, the thickness (0.3 nm) is based on the fact that if the distance between protein and amino acid is less than 0.3 nm, a water molecule can not penetrate the space in between. **Figure 4.4b** shows the temporal evolution of contact numbers as a function of time. The contact numbers fluctuate with time; nevertheless, Arg exhibits the smallest value while Trp has the largest one over the whole simulation duration. The average contact numbers are 0.59, 0.73 and 0.78, respectively, for Arg, Phe and Trp. This suggests that 59% Arg molecules stay within the contact layer during the simulation time span. The number distribution and contact number are structural indicators and consistent with the energetic analysis discussed earlier. If a species interacts with protein more strongly

than other species, it is in closer contact with protein and encounters a larger retarding force and thus transports slower, and vice versa.

We also calculated the entry numbers and residence times of amino acids in the contact layer of protein surface. The entry number is the times amino acid enters into the contact layer during the 20 ns simulation duration. The residence time is the time span of amino acid continually staying in the contact layer without escaping. Note that both quantities were ensemble averaged at the atomistic scale. As reported in **Table 4.3**, Arg enters the contact layer the most frequently (442 times), while Trp enters the least frequently (377 times). On average, Arg stays in the contact layer for 26.6 ps and Trp stays for 41.4 ps. These results further support our earlier discussion that Trp is the most favorable being adsorbed onto protein surface, stays there for the longest time, and moves the slowest.

Though not shown, it was also found that the contact numbers of amino acids become greater with decreasing solute concentration or solvent flowing rate. This suggests that a higher percentage of amino acids are bound closely to protein surface at a lower concentration or a slower flowing rate of mobile phase.

4.3.5 Hydrogen Bonds and Solvent-accessible Surface Areas

The hydrogen bonds and solvent-accessible surface areas were calculated from geometrical perspective to further characterize the observed velocity difference of the three amino acids. A hydrogen bond between a donor and an acceptor forms if two geometrical criteria are satisfied: (a) the distance between the donor and the acceptor is less than 0.35 nm and (b) the hydrogen-donor-acceptor angle is less than 30°.[50] **Figure 4.5a** shows the number of hydrogen bonds of amino acids with protein and water, respectively. The numbers are based on one molecule of amino acid with all protein molecules or all water molecules. Approximately, one hydrogen bond forms

with protein for each amino acid. Nevertheless, the number of hydrogen bonds formed with water is significantly different and Arg has the largest number. Arg has five polar groups that provide numerous donors and acceptors for hydrogen bonding, leading to nine hydrogen bonds per amino acid molecule. Further analysis shows that the side chain of Arg contributes one third (i.e. three) of the hydrogen bonds formed with water. The side chain of Phe has neither donors nor acceptors, and the side chain of Trp has one NH group; consequently, both Phe and Trp form fewer hydrogen bonds with water.

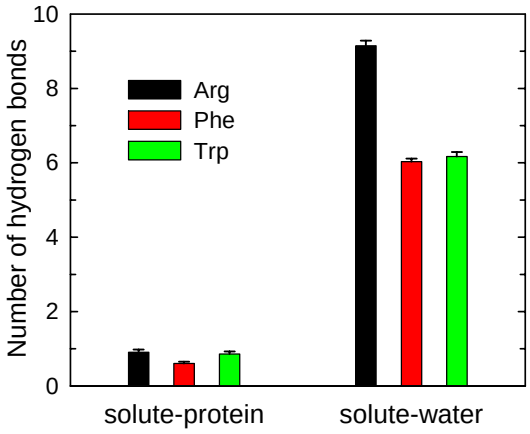

Figure 4.5. Numbers of hydrogen bonds between amino acids and protein or water. The values are based on one amino acid molecule.

Figure 4.6 shows the solvent-accessible surface areas (SASAs) of amino acids. The three amino acids have different hydrophobic SASAs in increasing order Arg < Phe < Trp. Trp has two hydrophobic rings (one six-member and the other five-member) and Phe has one hydrophobic ring; consequently, Trp has the largest hydrophobic SASA and followed by Phe and Arg. Conversely, Arg exhibits the largest hydrophilic SASA, while Phe and Trp have approximately the same hydrophilic SASA. This is because Arg is a linear molecule with a very hydrophilic side chain (three NH_2 groups) and more favorable to interact with water.

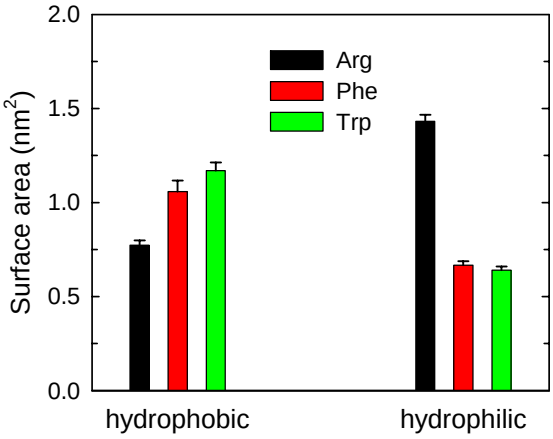

Figure 4.6. Solvent-accessible surface areas of amino acids. The values are based on one amino acid molecule.

The differences in the interactions of amino acids with water are further evidenced by the hydration numbers reported in **Table 4.3**. Here the hydration numbers are the averaged numbers of water molecules within 0.3 nm from the surface of one amino acid molecule. Arg has the largest hydration number (25.9), while Phe and Trp have smaller values (~ 20). This further confirms the strongest interaction of Arg with water, as discussed earlier.

4.4 Conclusions

The separation of Arg, Phe and Trp was examined using MD simulations. The study was to mimic a liquid chromatographic separation process with GI crystal as the stationary phase and water as the mobile phase. The simulation results reveal that the retention mechanisms of amino acids in GI crystal are the counterbalance between solute-water interaction and solute-protein interactions. The Coulombic potential dominates the interactions between amino acids and water or protein. Among the three amino acids, Arg interacts with both water and protein the most strongly because Arg is highly hydrophilic and positively charged. The differences in the interaction energies with water and protein follow the order Arg > Phe > Trp. The

strongest attraction between Arg and water results in its fastest transport velocity dragged by water. The velocities decrease in the order Arg > Phe > Trp in all directions. These are also supported by the structural and geometrical properties. The number distributions of amino acids around protein follows the order Arg < Phe < Trp. The hydrophobic SASAs increase in the order Arg < Phe < Trp as a consequence of hydrophobic rings. Arg is highly hydrophilic and exhibits the largest number of hydrogen bonds with water, hydrophilic SASA and hydration number.

The predicted transport velocity decreases in the order Arg > Phe > Trp, which agrees qualitatively with the available experimental observation. The elution order of amino acids is independent of different solute concentration or solvent flowing rate, but these two factors affect separation efficiency. Nevertheless, we would like to point out the differences between our simulation and experiment. First, the mobile phase (water) in experiment was driven to flow by a pressure gradient and water molecules at different locations in chromatographic column may experience different driving force. However, a uniform external force was exerted on each water molecule in simulation. This would not lead to qualitative deviation in the elution order of the test mixture. Second, the stationary phase in simulation was an ideal crystal with infinitely large periodical network and hence only intra-crystalline pores existed. In contrast, the stationary phase in experiment consisted of numerous micrometer-scale crystalline particles and there were inter-crystalline pores in addition to the intra-crystalline pores. Consequently, solute and solvent could enter both inter- and intra-crystalline pores. This effect was not fully captured by our simulation. Finally the GI crystal in experiment was cross-linked by glutaraldehyde, while the crystal in simulation was not.

Despite the ineluctable deviations between simulation and experiment, the simulation methodology adopted here is general and could be extended to separation processes in other nanostructures such as zeolites, metal-organic frameworks and others. The salient feature of molecular simulations is that it can treat idealized model system and thus provide microscopic insight that would otherwise be experimentally intractable. In principle, simulations with very detailed atomic-level description for system of interest are possible, but the computational effort required may be beyond current resources. As a compromise of accuracy and computational time, multiscale modeling approach has been proposed ranging from the low-resolution coarse-grained (CG) models to the high-resolution fully atomistic counterparts.[171,172] The length scale for system consisting of protein crystal is usually large at the atomistic level. For instance, there are over 150,000 atoms for the simulation system under this study and atomistic simulations are very computationally extensive. Therefore, the CG model is a feasible option to achieve longer time scale for such a large system without the loss of fundamental features.

Chapter 5 Chiral Separation of Racemic Phenylglycines in a Thermolysin Crystal

5.1 Introduction

Separation of chiral enantiomers has attracted extensive interest in chemical and pharmaceutical industries. Usually one enantiomer possesses therapeutic activity, while the other is either of no therapeutic value or even causes adverse effect. As a consequence, enantiomerically pure drugs are being increasingly used in clinic. In 2004, five of the top ten selling drugs were single-enantiomer products, which accounted for about 61% of the global sale of pharmaceuticals.[173] However, the separation of chiral compounds is very challenging because two enantiomers have identical chemical and physical properties under symmetric environment.

One of commonly used techniques for chiral separation is chiral chromatography, which utilizes chiral stationary phases to discriminate enantiomers. Although numerous experiments successfully reported enantioseparation in chiral chromatography, many fundamental issues on chiral recognition mechanisms remain elusive. This is largely due to the complexity of chromatographic systems and the lack of detailed microscopic understanding. Conventional experimental methods are unable to examine this level of complexity from the microscopic scale. To facilitate the rational design of new chiral stationary phase and optimization of separation processes, it is indispensable to obtain insight into chiral chromatographic separation at the molecular level. In this regard, molecular simulations have played an increasingly important role as they can provide atomistic/molecular pictures that would otherwise be experimentally intractable or impossible to obtain. A number of

simulation studies have been reported on chiral discrimination. For instance, Lipkowitz et al. located the chiral recognition regions by determining the Boltzmann-weighted intermolecular energies between analytes and *β*-cyclodextrin from MD simulations.[174,175] Cann and coworkers used MD simulations to investigate the solute conformations at chiral surfaces embedded with selectors in different solvents.[155-157,176] Szabelski and Sholl performed MC simulations to differentiate the adsorption of enantiomers on a chiral stationary phase.[177,178] Kasat et al. examined the effects of solute structure on its enantioresolution in cellulose using MD simulations and the elution orders predicted were consistent with HPLC results.[179,180]

The inherently chiral nature of *L*-amino acids as building blocks of proteins creates a chiral environment, which could lead to the selective separation of enantiomers using protein-based stationary phases. Many proteins such as lysozyme, albumins, and glycoproteins are commonly immobilized on achiral supports as chiral stationary phases.[65] However, the utilization of support matrix results in a low volumetric specific activity of proteins and thus decreases the separation efficacy. Vilenhick et al. showed that cross-linked protein crystals can be directly packed in a chromatographic column without support and act as chiral stationary phase to improve separation efficacy.[1] In addition, the compact arrangement of protein molecules in crystalline phase inhibits protein unfolding and thus more effectively maintains their native conformations even at elevated temperatures or in organic solvents as compared to amorphous proteins.[1,2] Also known as bioorganic zeolites, protein crystals possess many fascinating features, including high porosity (0.5–0.8), large surface area (800–2000 m^2/g), and wide range of pore size (1.5–10 nm).[1] Therefore, protein crystals have been proposed as a novel class of stationary phases in chiral chromatography to separate mixtures of racemic drug enantiomers. Decades ago, the

applications of protein crystals were severely restricted by their fragility and unpredictable growth patterns. These problems have to some degree been solved and mechanically stable protein crystals can be made by cross-linking techniques.[1,2] For instance, cross-linked thermolysin crystal was packed in a column to separate several racemic mixtures, which demonstrated the mechanical rigidity of cross-linked crystal.[1] Cross-linking also enhances the chemical stability of proteins in crystalline form. A single column of thermolysin or human serum albumin crystal was used for more than 500 injection cycles without the loss of separation efficiency.[1] Several techniques have been reported to effectively produce cross-linked protein crystals with high yield and good quality or on a large scale.[18-20] As a consequence, protein crystals have been increasingly utilized as new class of chiral separation materials.[1,2,5,162,163]

A handful of experimental and computational studies have been reported in order to elucidate the underlying physics of fluids confined in protein crystals.[55-57,181] In Chapters 2 and 3, we studied the properties of water and ions in protein crystals with varying morphology and under electric field. In Chapter 4, we reported the separation of three amino acids (arginine, phenylalanine and tryptophan) in glucose isomerase crystal and the elution order was found to be in accord with experimental measurement. To our knowledge, chiral separation process in protein crystals has not been computationally explored. A clear understanding on the separation mechanisms is crucial to the new development of chiral separation technologies using protein crystals.

The objective of this work is to use MD simulations to examine the capability of thermolysin crystal for chiral separation. The simulations are to mimic the experimental liquid chromatographic separation process, in which thermolysin crystal

acts as the stationary phase and water as the mobile phase. Thermolysin is a thermostable metalloproteinase enzyme and cross-linked thermolysin crystals have been utilized as stationary phase in liquid chromatography for chiral separation.[1] However, there is yet no computational effort to explore the separation mechanisms involved; therefore, insightful information can be obtained from the present simulation study. In the next section, we begin with a description of the model and method used to simulate a mixture of *D*/*L*-Phenylglycine (Phg) in thermolysin crystal. The elution order of two enantiomers is compared with experiment and the underlying mechanisms for separation are elucidated. Finally, the concluding remarks are summarized.

5.2 Models and Methods

Thermolysin crystal has a $P4_12_12$ space group and lattice constants $a = b = 9.705$ nm, and $c = 10.658$ nm (PDB ID: 1L3F).[103] The largest pore along the z direction is nearly cylindrical with a radius of 1.8–2.2 nm; the pores in the x and y directions are relatively small and irregular. The pores in different directions are interconnected and form a complicated nanoporous network, through which solvent and solute molecules can transport.

Simulations were carried out at the physiological condition (pH ≈ 7), which is close to the pH in the experimental chromatographic separation (7.5).[1] Consequently, Arg and Lys residues were protonated, while Asp and Glu residues were deprotonated, leading to two positive unit charges for each thermolysin molecule. Each thermolysin consists of 4 Ca^{2+} and 3 Zn^{2+} ligand ions necessary for activity and stability. The simulation box contained two unit cells (1 × 1 × 2) and the periodic boundary conditions were exerted in all three directions. Eighty *D*- and *L*-Phg molecules were

added into the box; in addition, 32 Cl^- ions were introduced randomly for electroneutrality and H_2O molecules were added to mimic a fully hydrated thermolysin crystal. The final simulation system consisted of 16 thermolysin molecules, 80 *D*-Phg, 80 *L*-Phg, 64 Ca^{2+}, 48 Zn^{2+}, 32 Cl^- ions, and 42214 H_2O molecules. **Figure 5.1** illustrates the model used in this work for the separation of *D*/*L*-Phg in thermolysin crystal.

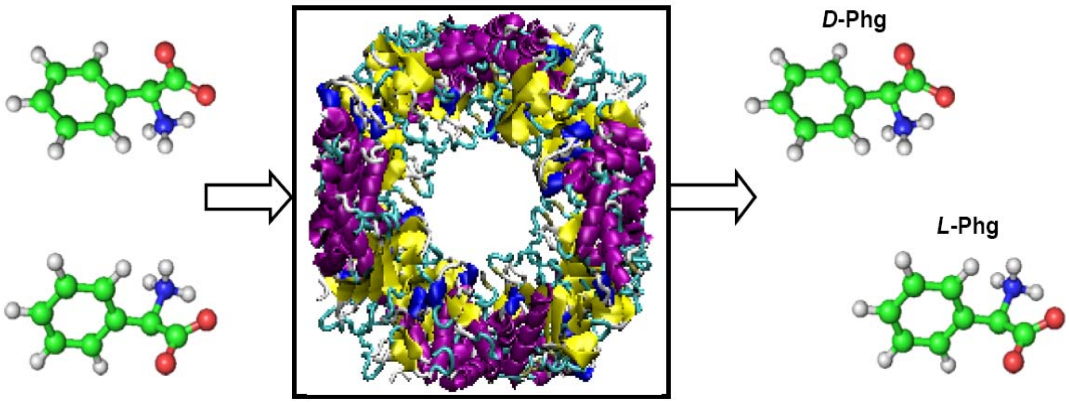

Figure 5.1. Schematic illustration for the separation of *D*/*L*-Phg through thermolysin crystal. Thermolysin is shown as cartoons on the *xy* plane; α-helices, β-sheets and random coils are in purple, yellow and cyan respectively. Water and ions are not shown for clarity.

Thermolysin, *D*/*L*-Phg and ions were modeled by GROMOS96 force field,[105] and water by SPC/E model.[122] GROMACS v3.3.1 package was used to perform MD simulations because it is extremely fast and particularly well-suited for biomolecular systems.[51] Some simulation settings were same as in Chapter 2 and thus not presented here. The system was initially subject to energy minimization using the steepest descent method with a maximum step size 0.01 nm and a force tolerance 10 kJ mol^{-1} nm^{-1}. Then velocities were assigned according to the Maxwell-Boltzmann distribution at 300 K. Non-equilibrium MD simulations were performed at 300 K as in the

experimental study.[1] The temperature was controlled by the Berendsen thermostat with a relaxation time of 0.1 ps. Integration time step was 2 fs and trajectory was saved for 30 ns at 2 ps intervals. The C_α atoms of thermolysin molecules were fixed and thermolysin molecules could not move thus acting as the stationary phase. A constant external acceleration a_{ext} was exerted on water molecules so that water flowed as the mobile phase. In order to examine the effect of the external force, three runs were performed with different accelerations (1) a_{ext} = 0.03 nm/ps^2, (2) a_{ext} = 0.05 nm/ps^2 and (3) a_{ext} = 0.07 nm/ps^2. It is noteworthy that the external forces applied were typically strong and commonly used in non-equilibrium MD simulations in order to reduce the thermal noise and enhance the signal/noise ratio within a nanosecond timescale.[169] Phg molecules were dragged by flowing water to transport through the pores within thermolysin crystal. The drift velocities of water and *D*/*L*-Phg were tracked to assure the system reaching a steady state. It is subtle to maintain temperature in non-equilibrium MD simulation since the friction heat generated by flux must be removed.[139] To calculate temperature, the mean velocity of each group (e.g., water and *D*/*L*-Phg) was subtracted from the instantaneous velocity. In our simulation, temperature was kept fairly well around 300 K with a small fluctuation of approximately 0.6 K. The system potential energy and drift velocities of water and amino acids were tracked and a steady state was reached in less than 10 ns, and the final 20 ns trajectories were used for analysis. The displacements, interaction energies, hydrogen bond numbers, and pair correlation functions were estimated using the built-in tools in GROMACS, and other properties including number distributions and residence times were calculated using in-house developed codes.

5.3 Results and Discussion

5.3.1 Effect of Solvent Flowing Rate

We first examine the effect of solvent flowing rate on the displacements of *D*- and *L*-Phg as shown in **Figure 5.2**. The displacements were calculated for the centers-of-mass of Phg molecules. The directional and overall ($v = \sqrt{v_x^2 + v_y^2 + v_z^2}$) velocities of of *D*- and *L*-Phg are reported in **Table 5.1**. It is found that *L*-Phg transports faster than *D*-Phg in all cases, demonstrating that thermolysin crystal can be used as stationary phase in liquid chromatography to separate *D*/*L*-Phg. The solvent flowing rate does not qualitatively affect the elution order of *D*/*L*-Phg, but it quantitatively influences the relative velocities between *D*/*L*-Phg. When the flowing rate increases from a_{ext} = 0.03 to 0.07 nm/ps^2, the ratio $v_{D\text{-Phg}}$:$v_{L\text{-Phg}}$ changes from 0.75:1 to 0.91:1. At a_{ext} = 0.03 nm/ps^2, the velocity difference between *D*/*L*-Phg is the largest. This suggests that the separation efficacy may be improved by decreasing flowing rate. As a consequence, solvent flowing rate could be tuned to optimize separation processes. To provide insightful information for chiral recognition mechanisms, the detailed analysis below is based on run 2 (a_{ext} = 0.05 nm/ps^2) unless otherwise stated.

Table 5.1. Velocities of *D*/*L*-Phg and water from three runs.

Runs	a_{ext} (nm/ps^2)	$v_{D\text{-Phg}}$ (m/s)	$v_{L\text{-Phg}}$ (m/s)	v_{H2O} (m/s)	$v_{D\text{-Phg}}$: $v_{L\text{-Phg}}$
1	0.03	2.74	3.64	23.60	0.75:1
2	0.05	22.84	28.68	45.83	0.80:1
3	0.07	53.12	58.14	80.08	0.91:1

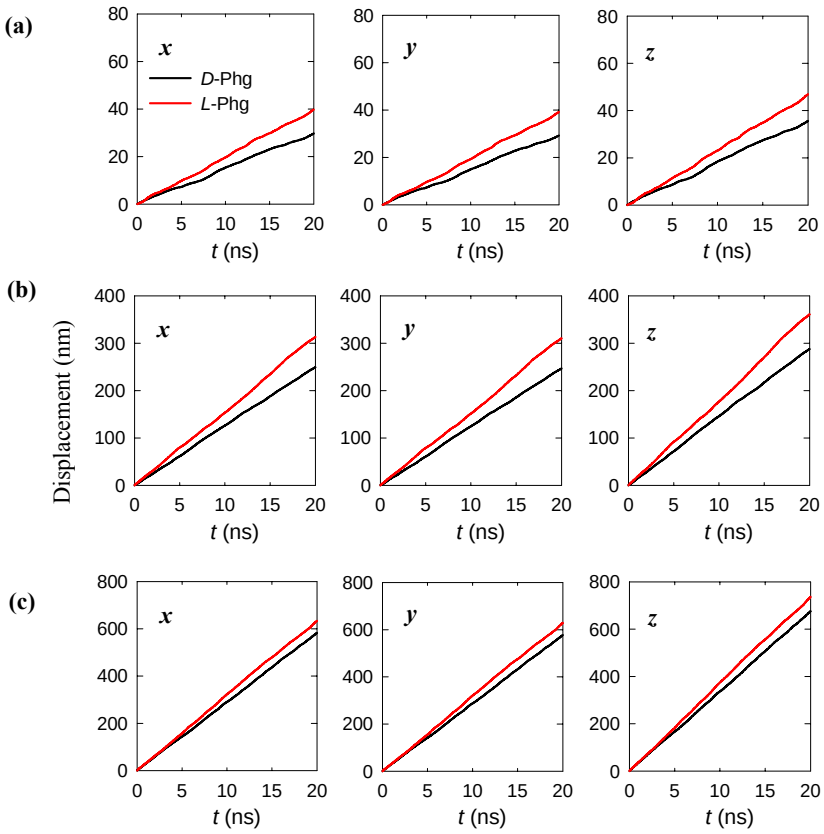

Figure 5.2. Displacements of *D*/*L*-Phg in *x*, *y* and *z* directions as a function of time in thermolysin crystal from three runs (a) a_{ext} = 0.03 nm/ps^2, (b) a_{ext} = 0.05 nm/ps^2, and (c) a_{ext} = 0.07 nm/ps^2.

5.3.2 Transport of Enantiomers

Table 5.2 lists the directional velocities of solutes and water. For *D*/*L*-Phg, the velocity in *x* direction is approximately equal to that in *y* direction; both are smaller than that in *z* direction. This is because the largest pores of the thermolysin crystal are in *z* direction and wider than in the other two directions. *D*/*L*-Phg molecules exhibit similar anisotropic transport reflected in v_x:v_y:v_z. It indicates that chiral discrimination occurs almost evenly in all directions, primarily because thermolysin crystal has multiple chiral discrimination sites dispersed on the pore surface extending along different directions. The elution order of *D*/*L*-Phg from our simulation is consistent with the experimental observation by Vilenchik et al.[1] They measured the retention

times of *D*/*L*-Phg in a chromatographic column packed with thermolysin crystal and observed that *L*-Phg moved faster than *D*-Phg. The separation factor α between *L*- and *D*-Phg was calculated by

$$\alpha = \frac{1/v_{D\text{-Phg}} - 1/v_{H_2O}}{1/v_{L\text{-Phg}} - 1/v_{H_2O}} \tag{5.1}$$

The experimental α was about 1.2, slight smaller than 1.45 estimated from our simulation. Note that the simulation system is simplified and cannot compare exactly with experiment. Nevertheless, the distinct difference in the velocities of *D*/*L*-Phg from both experiment and simulation reveals that *D*/*L*-Phg can be separated in thermolysin crystal.

Table 5.2. Velocities, residence times and numbers of H-bonds from run 2.

	***D*-Phg**	***L*-Phg**	**Water**
v_x (m/s)	12.56	15.76	24.77
v_y (m/s)	12.42	15.61	24.75
v_z (m/s)	14.49	18.18	29.56
v_x : v_y : v_z	0.87:0.86:1	0.87:0.86:1	0.84:0.84:1
Residence time (ps)	46.7	41.3	
Number of H-bonds with protein	0.67 ± 0.08	0.59 ± 0.07	

It is intriguing to observe that as a bioorganic microporous material, thermolysin crystal possesses the capability of chiral discrimination. In order to examine the capability of chiral separation in an inorganic counterpart, we also carried out simulations for *D*/*L*-Phg molecules transporting in a single-walled carbon nanotube (SWNT). Carbon nanotubes can exist in chiral form by helical winding of graphene sheets along the tube axis. The SWNT considered is chiral with index (22, 6) and has a diameter of approximately 2.0 nm. **Figure 5.3** shows the displacements of *D*/*L*-Phg in the SWNT. Two enantiomers exhibit approximately equal velocity in the SWNT,

implying the SWNT has no chiral discrimination toward *D*/*L*-Phg. A recent study by Power et al. also demonstrated that chiral SWNTs have no enentiospecific adsorption for enantiomers.[182]

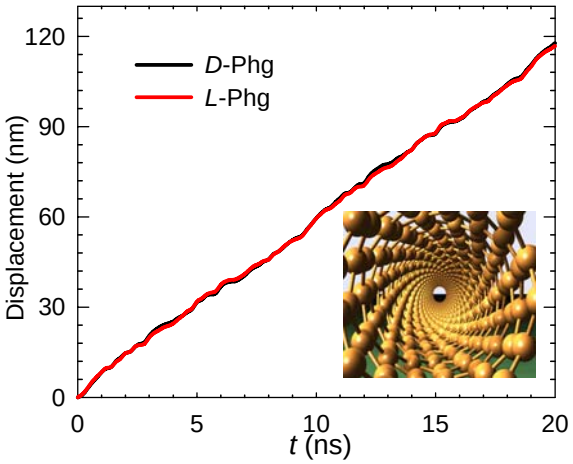

Figure 5.3. Displacements of *D*/*L*-Phg as a function of time along the axis in a chiral (22, 6) single-walled carbon nanotube.

The MD simulations successfully reproduced experimental chiral separation of *D*/*L*-Phg enantiomers in thermolysin crystal. To further understand enantioselectivity mechanisms involved, quantitative analysis will be discussed in more details below from both energetic and structural perspectives.

5.3.3 Energetic Analysis

Chiral separation in chromatography is a signature for the energy difference of enantiomers interacting with stationary phase. **Figure 5.4a** shows the nonbonded interaction energies of *D*/*L*-Phg with protein (ΔE_1). Compared to *L*-Phg, *D*-Phg interacts with protein more strongly; consequently, *D*-Phg is preferentially adsorbed onto the stationary phase protein surface and moves at a lower velocity with the mobile phase. In a previous study, the LJ force was found to dominate chiral discrimination in some cases, while the Coulombic force does in other cases.[174] Here

both LJ and Coulombic forces have a significant contribution to the energy difference of *D*/*L*-Phg with thermolysin crystal. For one Phg molecule with all sixteen thermolysin molecules in the system, the total energy difference Δ ΔE_1 (= $\Delta E_{1,\ D\text{-Phg}}$ − $\Delta E_{1,\ L\text{-Phg}}$) is approximately 9.9 kJ/mol. In terms of one Phg molecule with one thermolysin molecule, the energy difference is 0.62 (= 9.9/16) kJ/mol. Such a small energy difference between two enantiomers reveals that chiral separation is subtle and the selection of stationary phase is crucial. Directly packed in a chromatographic column, protein crystals have a higher volumetric specific activity compared to the case they are supported on a surface; therefore, energy difference and enantioseparation efficacy could be enhanced using protein crystals as stationary phase.

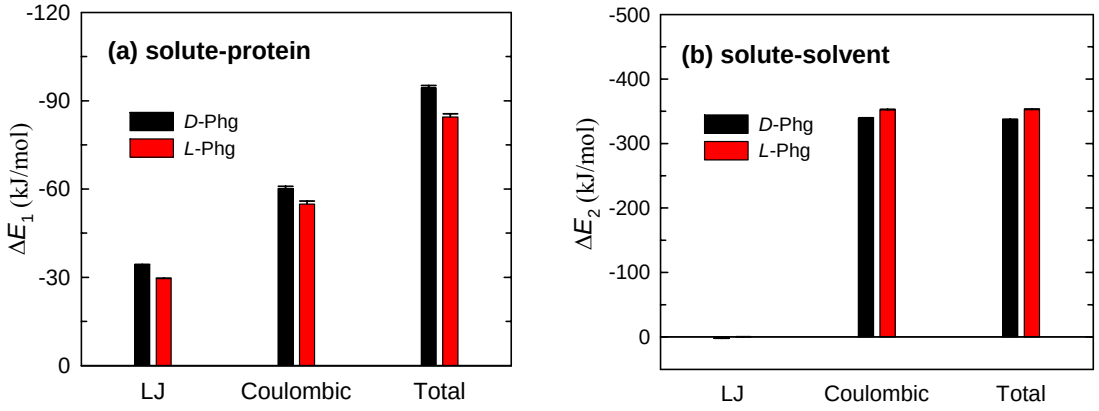

Figure 5.4. Nonbonded interaction energies of *D*/*L*-Phg with (a) protein and (b) water.

Figure 5.4b shows the interaction energies of *D*/*L*-Phg with water (ΔE_2). *D*-Phg interacts with water more weakly than *L*-Phg; the Coulombic potential dominates and the LJ potential is vanishingly small. The weaker interaction of *D*-Phg with water is inherently attributed to the existence of chiral selector. Due to the stronger interaction of *D*-Phg with protein, *D*-Phg is more adjacent to the protein surface and has smaller opportunity to contact water molecules, and hence it interacts with water more weakly.

At the same time, the water molecules on the surface of chiral selector can be induced to become chiral, as reported by Wang and Cann;[183] this factor may also lead to differential interaction of *D/L*-Phg with water. In other words, there should be no discernible energy difference of *D*/*L*-Phg with water unless chiral stationary phase exists. As demonstrated by the simulation of *D*/*L*-Phg in chiral (22, 6) SWNT, *D*/*L*-Phg exhibit nearly identical interaction with SWNT and there is no energy difference of *D*/*L*-Phg with water (data not shown). Therefore, the mechanisms for chiral separation explored here are different from achiral separation. In Chapter 4, our study showed that the (achiral) separation of amino acids in glucose isomerase crystal was due to the counterbalance between solute-water and solute-protein interactions, and both play an important role in the retention of amino acids.

5.3.4 Structural Analysis

The analysis above is based on energetic point of view. We will further provide structural analysis including number distributions, contact numbers, residence times, hydrogen bonds and pair correlation functions.

D-Phg interacts with protein more strongly than *L*-Phg, it is thus expected that *D*-Phg is closer to protein surface. To quantify, the normalized accumulative number distributions $N_{\mathrm{acc}}(r)$ of *D*/*L*-Phg were calculated as a function of distance *r* from the nearest protein atom using Eq. (4.1). The van der Waals radii of protein atoms were taken into account in the calculations. The number distribution is a measurement to estimate how solute molecules distribute from the protein surface. As shown in **Figure 5.5a**, $N_{\mathrm{acc}}(r)$ at a given *r* follows the order *D*-Phg > *L*-Phg. Therefore, *D*-Phg stays more closely to the surface of stationary phase (protein) and moves slower than *L*-Phg. This is attributed to the stronger interaction of *D*-Phg with protein. The

number distributions of water around *D*/*L*-Phg were also evaluated and it was found that water is adjacent more to *L*-Phg. As a consequence, *L*-Phg transports faster than *D*-Phg dragged by the mobile phase (water).

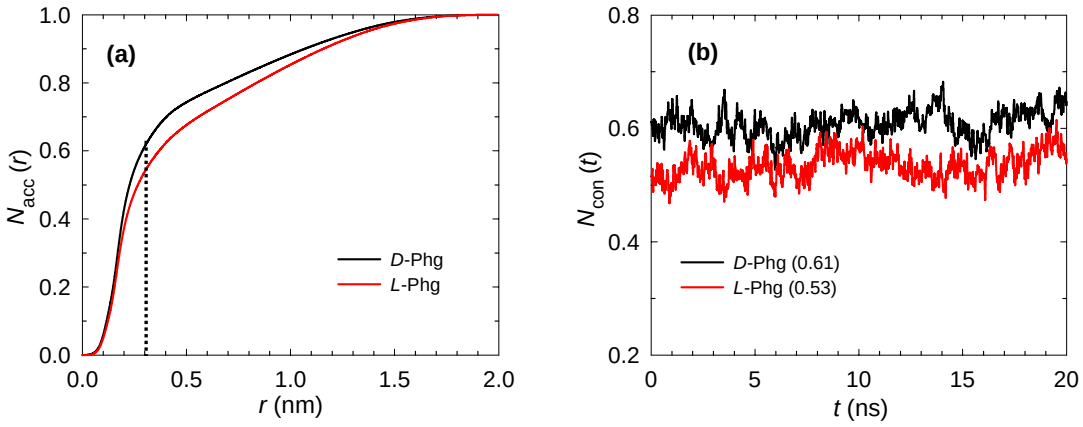

Figure 5.5. (a) Accumulative number distributions of *D*/*L*-Phg as a function of distance from protein surface. The dotted line indicates r = 0.3 nm. (b) Contact numbers of *D*/*L*-Phg as a function of time. The contact numbers averaged over time are shown in the parentheses.

If we define a contact layer to be within 0.3 nm from the protein surface, the number of solute molecules in the contact layer (i.e., the contact number) is then given by $N_{acc}(r = 0.3)$. The thickness of the contact layer defined does not qualitatively affect the result. The thickness of 0.3 nm is based on the fact that if the distance between protein and solutes is less than 0.3 nm, a water molecule can not penetrate the space in between and thus solutes molecules directly "contact" protein surface. **Figure 5.5b** shows the temporal evolution of contact numbers of *D*/*L*-Phg as a function of time. The contact numbers fluctuate with time; nevertheless, *D*-Phg exhibits greater value than *L*-Phg over the whole simulation duration. The average contact numbers are 0.61 and 0.53 for *D*-Phg and *L*-Phg, respectively. This implies that 61% *D*-Phg and 53% *L*-Phg stay within the contact layer during 20 ns. The number distributions and contact numbers are structural indicators and consistent with

the energetic analysis discussed earlier. *D*-Phg interacts with protein more strongly than *L*-Phg, and thus it is in closer contact with protein, encounters a larger retarding force and transports slower.

The residence times of *D*/*L*-Phg were calculated to estimate the time span of solute continually staying in the contact layer of protein surface without escaping. As reported in **Table 5.2**, *D*-Phg is found to stay in the contact layer for a longer time (46.7 ps) than *L*-Phg (41.3 ps). It further supports our earlier energetic analysis that *D*-Phg is more favorably adsorbed onto protein surface, stays there for a longer time, and moves slower.

Hydrogen bonding was considered to be primarily responsible for chiral discrimination.[176] To examine the role of hydrogen bonding for our system here, the numbers of hydrogen bonds between *D*/*L*-Phg and protein were calculated and listed in **Table 5.2**. A hydrogen bond between a donor and an acceptor forms if two geometrical criteria are satisfied: (a) the distance between the donor and the acceptor is less than 0.35 nm and (b) the hydrogen-donor-acceptor angle is less than 30°.[50] The side chain of Phg molecule is a hydrophobic phenyl ring and has no contribution to hydrogen bonding. On average, there is 0.67 ± 0.08 hydrogen bond formed between one *D*-Phg molecule with protein; and 0.59 ± 0.07 for one *L*-Phg molecule. The difference in the number of hydrogen bonds between *D*/*L*-Phg and protein is marginal within statistical uncertainty, indicating that hydrogen bonding might not play a dominating role in the chiral discrimination of *D*/*L*-Phg.

The capability of thermolysin to discriminate *D*/*L*-Phg is due to the fact that thermolysin contains chiral amino acids, which create many chiral centers and interact with two enantiomers differently. The primary chiral centers in thermolysin are Cα atoms; a portion of them are located on the pore surfaces and contribute to chiral

recognition. To characterize the interactions of chiral centers in thermolysin with *D*/*L*-Phg, the pair correlation functions were calculated using Eq. (2.1).

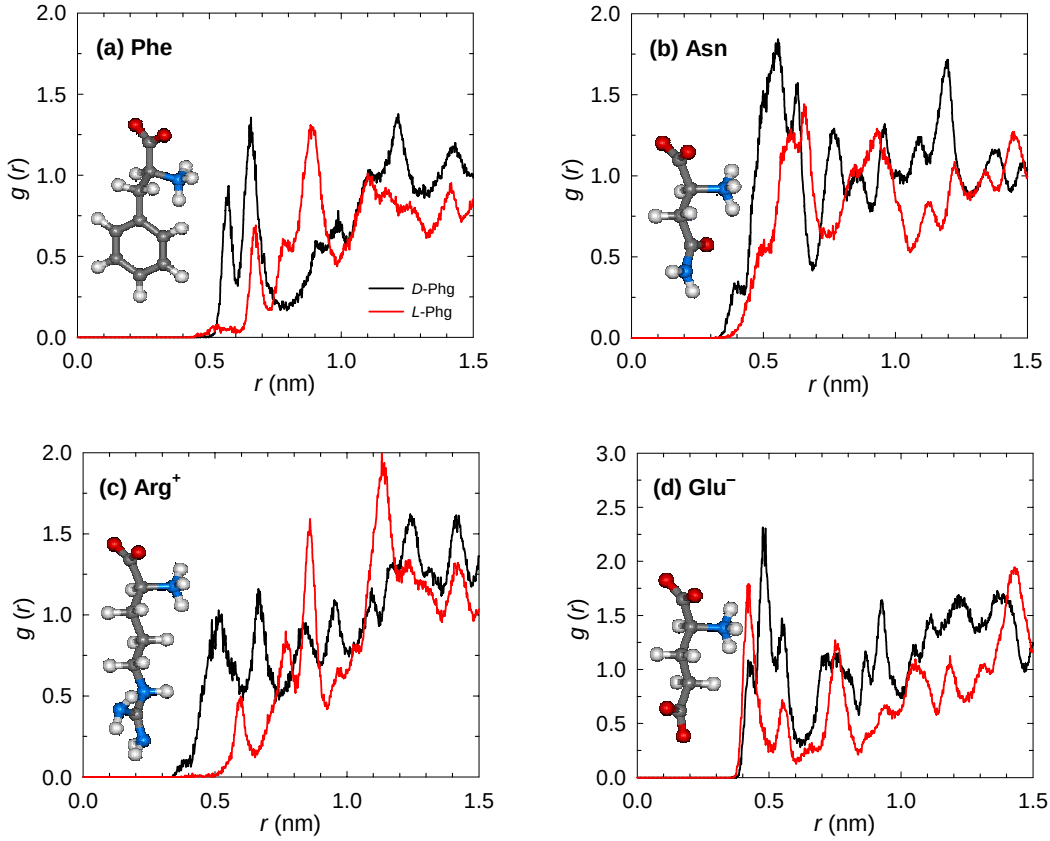

Figure 5.6. Pair correlation functions g(*r*) between C_α atoms of *D*/*L*-Phg and Cα atoms of Phe, Asn, Arg and Glu residues in thermolysin. Phe: nonpolar, Asn: polar, Arg: basic, Glu: acidic. The molecular structures of residues are shown in the inset. Color code: C, grey; O, red; N, blue; H, white.

Figure 5.6 shows *g*(*r*) between C_α atoms of *D*/*L*-Phg and C_α atoms of Phe, Asn, Arg and Glu residues in thermolysin. The four residues are nonpolar, polar, basis and acidic, respectively; each represents a typical type of residue. The complex profiles are due to the irregular distribution of residues in thermolysin. For Phe, Asn and Arg residues, the first and second peaks in *g*(*r*) of *D*-Phg are closer to protein surface, higher and broader than those of *L*-Phg. For Glu residue, although the first peak of *L*-Phg is higher than that of *D*-Phg, the second peak of *D*-Phg latter is much higher and broader. Consequently, the four representative residues or chiral centers have stronger

interactions with *D*-Phg rather than *L*-Phg. These chiral centers with others collectively make a contribution to the chiral discrimination.

5.4 Conclusions

The separation of *D*/*L*-Phg was examined using non-equilibrium MD simulations. The study was to mimic a liquid chromatographic separation process with thermolysin crystal as the stationary phase and water as the mobile phase. The simulation results show that *D*-Phg has a smaller transport velocity compared with *L*-Phg. The chiral discrimination is the consequence of stronger interaction of *D*-Phg with thermolysin crystal, as reflected in interaction energy and structural properties. The number distribution of *D*-Phg around protein is larger and *D*-Phg stays near the protein surface for a longer time. The chiral centers of different residues in thermolysin are responsible for chiral recognition. There is a marginal difference in the number of hydrogen bonds between *D*/*L*-Phg with protein, thus hydrogen bonding is not a key factor to chiral discrimination. The predicted elution order of *D*/*L*-Phg in thermolysin crystal is consistent with experimental observation, and the calculated separation factor is close to measured value.

It is worthwhile to point out that the model adopted here is not completely the same as real experimental sample. For instance, thermolysin crystal in experiment was cross-linked by glutaraldehyde, which was not accounted for in simulation. Nevertheless, major features of the sample have been captured in the model and good agreement with experiment is obtained. Our simulation represents a realistic liquid chromatographic separation process under a non-equilibrium condition, in which the mobile phase flows through the stationary phase. Most previous simulation studies employed MC for the adsorption of *D*/*L*-isomers onto chiral selectors, or equilibrium MD for the structural and thermodynamic properties and docking processes of

enantiomers near chiral selectors. In principle, the simulation methodology adopted here could be readily extended to probe other enantioseparation processes in chromatography and examine the enantioseparation capability of various chiral stationary phases directly from non-equilibrium MD. However, caution should be drawn and simulations should be conducted carefully. Because chiral discrimination is very subtle and energy difference between enantiomers is usually very small, it is formidable to observe distinct separation in a limited simulation time scale. The chiral separation is pronounced in current study as attributed to two factors. First, the stationary phase (thermolysin crystal) in our simulation possesses a significant number of chiral centers. Second, a high flowing rate of the mobile phase was used to enhance the signal/noise ratio within a nanosecond timescale. From both energetic and structural analysis, this work provides a deeper understanding for chiral discrimination in thermolysin crystal, and suggests that protein crystal might be useful for chiral separation.

Chapter 6 Assessment of Biomolecular Force Fields

6.1 Introduction

Force fields govern interaction energies and play a central role in molecular simulations.[184,185] For biomolecules the most commonly used force fields include OPLS,[46,47,186] CHARMM,[43] AMBER,[44,45,187] and GROMOS.[48] In OPLS and CHARMM, the nonbonded parameters were developed by computing thermodynamic and structural properties, whereas the torsional parameters and partial charges were determined by fitting separately to rotational energies and electrostatic potentials from ab initio calculations. Earlier versions of AMBER force fields (e.g. AMBER94 and AMBER99) estimated the partial charges using ab initio calculations in gas phase.[44] In the latest AMBER03, however, the calculations were carried out in a condensed phase with continuum-solvent of an effective dielectric constant.[45] GROMOS force field (e.g. GROMOS96) is based on united-atom model and was parameterized by the free enthalpies of solvation.[48]

As the matrix of life, water is crucial in determining the structure, dynamics and functionality of biomolecules. Consequently, water is regarded as an integral part of biomolecular organization and a precise water model is tremendously important in biomolecular simulations. The widely used water models are SPC,[106] SPC/E,[122] TIP3P,[188] TIP4P[188] and TIP5P.[189] A specific water model is usually used in the development of biomolecular force fields. For example, OPLS, CHARMM and AMBER force fields were parameterized with TIP3P water model, while GROMOS with SPC water model. In biomolecular simulations, one should carefully select the possible combination of force field with water model.[190] In addition to water, ions also exist ubiquitously in biomolecular systems, e.g., acting as cofactors. All the four

popular force fields, OPLS, CHARMM, AMBER and GROMOS provide the parameters of common ions such as Na^+ and Cl^-. However, these parameters were derived from individual ions, not based on a neutral pair of ions.[191] It is interesting to examine whether they can be incorporated with biomolecular force fields to describe the behavior of ions in biomolecular systems. For instance, a few simulation studies have been conducted for biomolecules using different force fields. The dynamic properties of 30 proteins were examined using the above-mentioned force fields and it was found that these force fields give comparable results.[192] The water model was found to largely affect the accuracy of thermodynamic properties of 13 amino acids and SPC/E performs best along with OPLS-AA (all atom), AMBER99 and GROMOS 53A6 force fields.[193]

Recently, protein crystals have emerged as a new class of bio- and nano-materials. With a wide range of porosities, surface areas and pore sizes and unique chiral environments, protein crystals can be used as high-performance separation media for chemically or optically different molecules through size exclusion or chiral discrimination.[1] Due to the chemical and structural similarity, protein crystals also serve as an ideal prototype to understand the functioning of biomembrane channels or as benchmarks to examine fluids within highly crowded media (e.g. mitochondria with confined water). Nevertheless, current understanding of fluids in protein crystals remains largely obscure. The characteristics of protein crystals and subsequently the behavior of confined fluids would vary with media and external environment. A set of guidelines on how to select a specific protein crystal and to optimize operation condition are crucial to the new development of technically feasible and economically competitive separation technology using protein crystals. A number of experimental studies have been carried out to investigate the permeation of solvent or solutes

through protein crystals.[52-57,60] Several earlier computational studies focused on the difference of protein conformations in solution and crystalline environment.[69,70,75,76] A simple model was developed to evaluate the diffusion times of small molecules into protein crystals, which accounts for the physical and chemical properties of both protein crystal and diffusing molecules.[93] In Chapters 2 and 3, we studied the properties of water and ions in protein crystals with varying morphology and under electric field. In Chapters 4 and 5, we examined the liquid chromatographic separation processes using protein crystals as stationary phase. To save the computational time, we have used united-atom based Gromos96 force field for proteins in these Chapters.

The objective of this Chapter is to examine the capabilities of various biomolecular force fields in conjunction with different models for water and ions to predict the static and dynamic properties in a lysozyme crystal with counterions Cl^- or the addition of NaCl. The simulated results are compared with available experimental data. Lysozyme is chosen as the host crystal because it is one of the most commonly studied proteins with readily available structure. Three biomolecular force fields, OPLS-AA, AMBER03 and GROMOS96 are tested along with three water models, SPC, SPC/E and TIP3P. Additionally, different models for NaCl are considered including the Kirkwood-Buff (KB) model, which was developed by fitting the KB integrals to experimental data in NaCl aqueous solution.[194] In the next section, the simulated models and methods are briefly described. Thereafter, the properties of lysozyme, water and ions in the crystal are presented using different force fields and compared with measured data. Concluding remarks are summarized finally.

6.2 Models and Methods

As a readily available protein, the structure of lysozyme has been well resolved. Interestingly, lysozyme exhibits various crystalline forms, namely, tetragonal,

orthorhombic, monoclinic and triclinic.[3] In this work, the tetragonal structure was considered because experimental data are available in this morphology for comparison. Morozov et al. measured water diffusion in the tetragonal lysozyme crystal at room temperature and the diffusivities in x, y and z directions were about (0.18, 0.18 and 0.46) × 10^{-9} m^2/s, respectively.[52] They also measured the mobility of NaCl and the induced electrical conductivity in the crystal.[53] In our simulations, the tetragonal lysozyme crystal was constructed from the experimental crystallographic data (PDB ID: 1HEL[101]). It has a $P4_32_12$ space group and lattice constants $a = b$ = 7.91 nm and c = 3.79 nm. Two main nanopores exist in a unit cell; one is at the center and the other at the corner. In order to simulate the properties of water and ions under the corresponding experimental conditions,[52,53] two model systems were examined, one for water diffusivity (system I) and the other for ion mobility and electrical conductivity of NaCl (system II). System I consisted of two unit cells (1 × 1 × 2) with a size of 7.91 × 7.91 × 7.58 nm^3, as shown in **Figure 6.1**. Experiment revealed that water diffusivity in the lysozyme crystal was almost irrelevant to pH.[52] Therefore, system I was set at the physiological condition (pH ≈ 7). Arg and Lys residues were protonated, while Asp and Glu residues were deprotonated on the basis of experimental pK_a values,[53] thus leading to eight positive unit charges for each lysozyme molecule. Water molecules were added to fully solvate lysozyme molecules and Cl^- counterions were introduced randomly into the box for electroneutrality. The ratio of water number to lysozyme number was 374, close to 358 experimentally determined.[52] There were 16 lysozyme molecules, 5985 water molecules and 128 Cl^- ions in system I.

To mimic the experimental study for ion mobility and conductivity,[53] system II contained the addition of NaCl. An initial tentative simulation using two (1 × 1 × 2)

unit cells revealed that the statistical uncertainty was a bit large for ion mobility. Thus, a larger system with four unit cells (1 × 1 × 4) was used. According to the experimental condition (pH ≈ 9),[53] Arg residues were protonated, Asp and Glu residues were deprotonated, and the protonation states of Lys residues were determined based on their pK_a values. A macroscopic electrostatic model was used to evaluate pK_a by taking into account the interactions of a titratable site with solvent and other sites.[195] Thereafter, different protonation states were assigned to six Lys residues in each lysozyme. Specifically, Lys1, 13, and 116 (with pK_a > 9) were protonated, while Lys33, 96 and 97 (with pK_a < 9) were neutralized. As a consequence, one lysozyme carried five positive unit charges, close to experimentally determined 4.5.[53] After lysozymes were fully solvated by water, 130 Na^+ and 290 Cl^- ions were introduced randomly into the box, leading to [Na^+] = 0.228 M and [Cl^-] = 0.508 M. The Na^+ concentration was nearly the same as in experiment (0.229 M), while the Cl^- concentration was slightly larger than in experiment (0.464 M).[53]

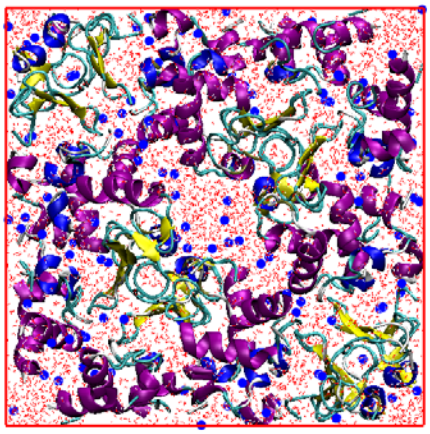

Figure 6.1. System I on the *xy* plane (7.91 × 7.91 nm^2). Lysozymes are shown as cartoons, in which α-helices, β-sheets and random coils are illustrated in purple, yellow and cyan respectively. Counterions and waters are represented by blue spheres and red sticks, respectively.

GROMACS v3.3.1 simulation package was used to perform the simulations because it is extremely fast and particularly well-suited for biomolecular systems.[51] It is also flexible, allowing for selecting suitable biomolecular force field and water model. As mentioned, we chose three biomolecular force fields, namely, OPLS-AA,[46] AMBER03[45] and GROMOS96,[48] and three water models, SPC,[106] SPC/E[122] and TIP3P.[188] Furthermore, the KB model for Na^+ and Cl^- was considered.[194] The parameters of LJ and Coulombic potentials for water and ions in different force fields are listed in **Table 6.1**.

Table 6.1. Lennard-Jones potential parameters and atomic charges of Na^+, Cl^- and water.

Force fields	Atoms	ε (kJ/mol)	σ (nm)	q (e)	Refs
GROMOS96	Na^+	0.0617	0.2575	+1	[48]
	Cl^-	0.4457	0.4448	–1	
AMBER03	Na^+	0.0116	0.3328	+1	[44]
	Cl^-	0.4184	0.4401	–1	
OPLS̃AA	Na^+	0.0116	0.3330	+1	[51]
	Cl^-	0.4928	0.4417	–1	
SPC	OW	0.6506	0.3166	+0.82	[106]
	HW	0.0	0.0	–0.41	
SPC/E	OW	0.6506	0.3166	+0.8476	[122]
	HW	0.0	0.0	–0.4238	
TIP3P	OW	0.6354	0.3151	+0.834	[188]
	HW	0.0	0.0	–0.417	
KB	Na^+	0.3200	0.2450	+1	[194]
	Cl^-	0.4700	0.4400	–1	

The geometric- or arithmetic-mean combining rules were used for ε_{ij} and σ_{ij}. A scaling factor of 0.75 was used for ε_{ij} between oxygen atom in water and Na^+ in KB model.[194] There were no parameters for Na^+ and Cl^- in the original OPLS-AA force field and these parameters were adopted from GROMACS v3.3.1.[51]

A cutoff of 1.4 nm was used to evaluate the LJ interactions. The long-ranged Coulombic interactions were evaluated using the particle-mesh Ewald (PME) method[41,42] with a grid spacing of 0.12 nm and a fourth-order interpolation. The bond lengths with dangling hydrogen atoms in lysozymes were constrained using the LINCS algorithm,[40] and water geometry was constrained using the SETTLE algorithm.[107] All simulations were performed in a canonical (NVT) ensemble with 298 K in system I and 291 K in system II. Temperature was controlled by the Berendsen thermostat with a relaxation time of 0.1 ps. The systems were initially subject to energy minimization, followed by 10 ns MD simulation for equilibration and subsequent 10 ns for production. The time step was 2 fs and trajectories were saved every 1 ps. The system energy and protein structure were monitored to ensure the system reached equilibration.

6.3 Results and Discussion

6.3.1 Lysozyme Structure and Water Diffusion in System I

Figure 6.2 shows the positional root-mean-squared deviations (RMSDs) for the C_α atoms of lysozymes with the initial X-ray crystallographic structure as a reference. The RMSDs from both AMBER03 and OPLS-AA reach plateaus rapidly in less than 2 ns and thereafter undergo small fluctuations. On average, the RMSD is about 0.10 nm from AMBER03 and 0.17 nm from OPLS-AA. Nevertheless, the RMSD from GROMOS96 slowly approaches a constant of 0.40 nm after approximately 10 ns. Compared to AMBER03 and OPLS-AA, GROMOS96 exhibits larger RMSD. A plausible interpretation is because GROMOS96 is a unit-atom model. For a given biomolecular force field (GROMOS96, AMBER03 or OPLS-AA), the RMSD is not significantly influenced by water model.

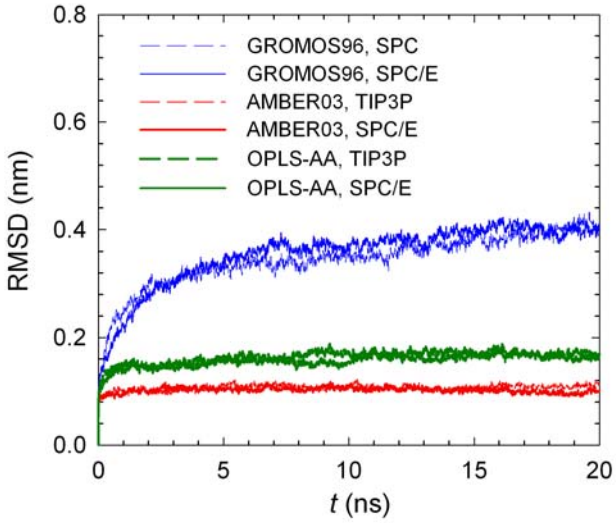

Figure 6.2. RMSDs for the Cα atoms of lysozymes in system I.

Temperature factor (also called *B*-factor) is usually used to evaluate the fluctuations of atoms around their averaged positions. *B*-factor is defined by[196,197]

$$B = \frac{8\pi^2}{3}\left\langle \mathbf{u}^2 \right\rangle \tag{6.1}$$

where **u** is the RMSF. **Figure 6.3** shows the *B*-factors for 129 C_α atoms of lysozymes. In Chapter 2, we showed that various lysozyme chains in the tetragonal crystal have approximately the same RMSF; consequently, the *B*-factors were averaged over all lysozymes here. Depending on the residue locations, *B*-factors exhibit different extents of fluctuations. The small *B*-factors are attributed to the secondary structures in peptide chains. In the tetragonal lysozyme, there are various secondary structures including α-helices, 3_{10}-helices and β-sheets as shown on the top of **Figure 6.3**. Specifically, α-helices are Gly4–Gly16, Leu25–Phe34, Ile88–Ser100, and Val109–Arg114; 3_{10}-helices are Pro79–Leu84, Gly104–Trp108, and Val120–Ile124; β-sheets are Ala42–Asn46, Gly49–Gly54, and Leu56–Ser60.[101] The secondary structures, especially, α-helices and β-sheets, are two primary building blocks in

proteins formed by hydrogen bonding. The residues forming secondary structures are bound more tightly and consequently exhibit relatively smaller fluctuations. In contrast, a number of residues are random coils (e.g. Gly22, Thr47, Asp48, Pro70 and Gly102) and thus exhibit large fluctuations.

Similar pattern of *B*-factors is observed in **Figure 6.3** from different force fields, though GROMOS96 gives larger *B*-factors than those from AMBER03 and OPLS-AA. For each force field, different water models give very close *B*-factors as in the case of RMSD in **Figure 6.2**. This indicates that the water model has an insignificant effect on the thermal motion of lysozyme, which is determined largely by protein force field. The *B*-factors predicted from OPLS-AA is slightly larger than from AMBER03 and both are in good accord with experiments.[101]

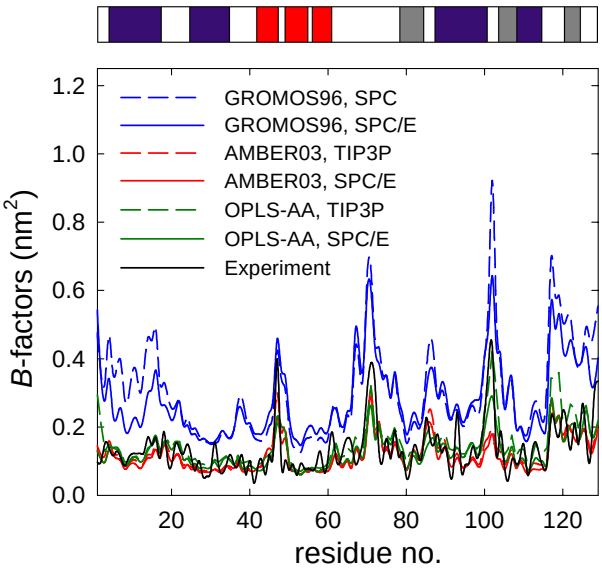

Figure 6.3. *B*-factors for the Cα atoms of lysozymes in system I. On the top, the dark blue, grey and red regions denote α-helices, 3_{10}-helices and β-sheets, respectively.

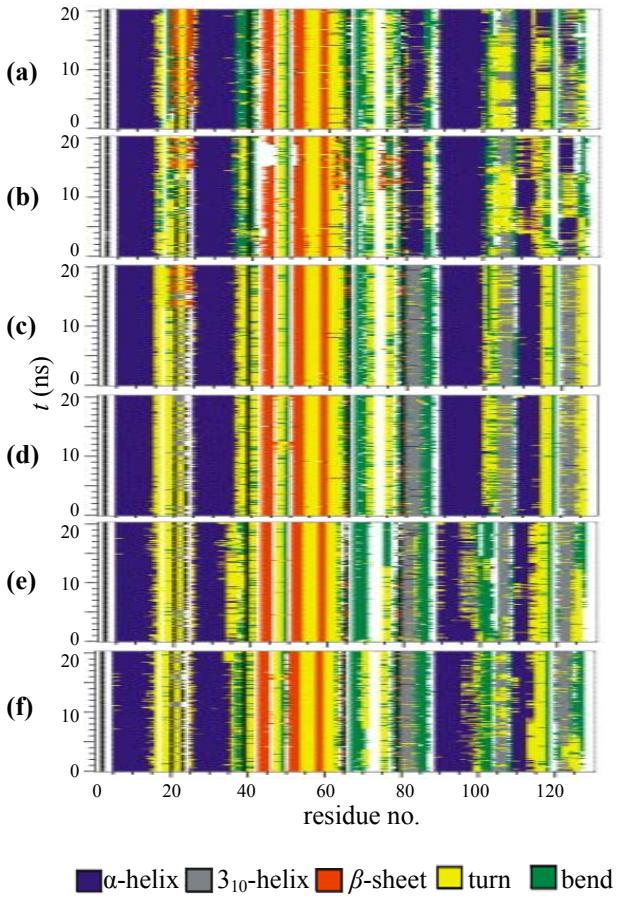

Figure 6.4. Evolution for the secondary structures of one lysozyme chain in system I. (a) Gromos96, SPC; (b) Gromos96, SPC/E; (c) AMBER03, TIP3P; (d) AMBER03, SPC/E; (e) OPLS-AA, TIP3P; (f) OPLS-AA, SPC/E.

The evolution of secondary structures in the lysozyme crystal was evaluated as a function of simulation time using the DSSP (database of secondary structure assignments) algorithm.[144] It was found that all the sixteen lysozyme chains in the simulation box share similar pattern of secondary structures using different force fields; therefore, **Figure 6.4** shows the evolution for only one chain. Again, the water model has a negligible effect on the evolution of secondary structures. The primary secondary structure segments, two α-helices (Gly4–Gly16 and Leu25–Phe34) and three β-sheets (Ala42–Asn46, Gly49–Gly54, and Leu56–Ser60) are well conserved

over the entire simulation duration using all force fields. However, several secondary structures undergo variations and the degree of variation depends on the force field. For instance, α-helix (Ile88–Ser100) is maintained using GROMOS96 and AMBER03, but diminishes to some extent using OPLS-AA. While three 3_{10}-helices (Pro79–Leu84, Gly104–Trp108, and Val120–Ile124) are essentially lost or converted to α-helices using GROMOS96, their structures are largely unchanged using AMBER03 and OPLS-AA.

Water diffusion was examined by calculating MSD using Eq. (2.4). The multiple-origin method was used to improve the statistical accuracy. Thereafter, diffusivity was calculated using Eq. (2.5). At ambient conditions, SPC/E water model gives (2.49 ± 0.05) × 10^{-9} m^2/s for bulk phase and agrees well with the experimental diffusivity 2.30 × 10^{-9} m^2/s. However, SPC and TIP3P models yield (3.85 ± 0.09) × 10^{-9} m^2/s and (5.19 ± 0.08) × 10^{-9} m^2/s, respectively, much larger than the experimental value.[198] It is intriguing to examine whether SPC/E model also gives a good description of water diffusion in a confined space such as the lysozyme crystal under this study, in which the host structure may play a significant role.

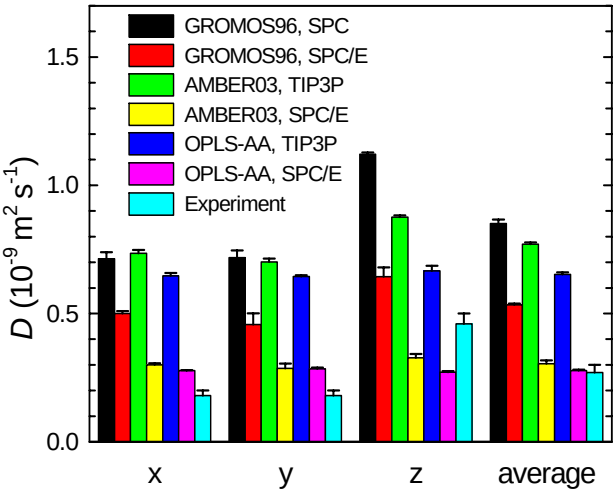

Figure 6.5. Directional and average water diffusivities in system I.

Figure 6.5 shows water diffusivities in the lysozyme crystal using different force fields and water models. GROMOS96 predicts larger water diffusivities than OPLS-AA and AMBER03. In conjunction with SPC/E model, AMBER03 and OPLS-AA give nearly identical diffusivities, which are the closest to experimental data. All other combinations considerably overestimates the directional (in the x, y and z directions) and average diffusivities. It can thus be concluded that compared to SPC and TIP3P models, SPC/E model is better in predicting water diffusivities not only in bulk phase but in protein crystal. As a consequence of the geometric restraints and surface interactions of crystal framework, water diffusion in the lysozyme crystal is nearly ten times slower than in bulk phase. All the three biomolecular force fields adopt similar functional forms; therefore, the difference in their predictions is apparently caused by the potential parameters, which will be discussed in detail below. AMBER03 and OPLS-AA along with SPC/E model give D_x and D_y slightly larger than the experimental values, which could be attributed primarily to deviation between simulation model and experimental sample. The model is a perfect crystal and hence does not mimic exactly the experimental sample. The lysozyme crystal in experiment was cross-linked by glutaraldehyde, which retarded water motion in the crystal. Morozov et al. estimated that the volume occupied by glutaraldehyde was approximately 10% of the whole volume and the existence of cross-linkers might cause 17% drop in water diffusivity along the x or y direction, but a smaller drop in the z direction.[52] Taking the effect of cross-linkers into account, therefore, it is reasonable that D_x and D_y predicted in the model system are larger than the experimental data. Nevertheless, the predicted D_z in the z direction along the pore axis is smaller. The values of diffusion anisotropy (defined as D_z /D) are 1.2, 1.1 and 1.0, respectively, from GROMOS96, AMBER03 and OPLS-AA; all these values are

smaller than the experimental value 1.7.[52] The fact not including cross-linkers in simulation leads to an overestimated average D, which is attributed to the underestimation of anisotropy. A better representation of the experimental sample and a more accurate force field may improve the predictions.

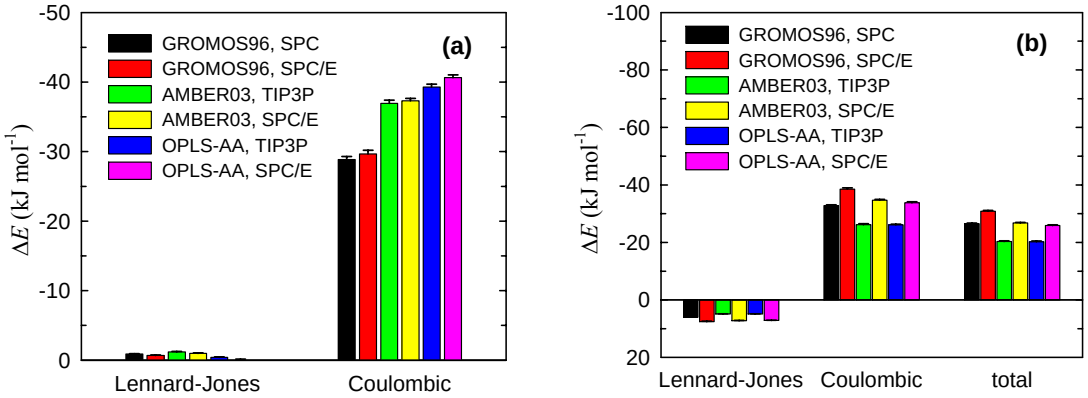

Figure 6.6. Interaction energies (a) between water and lysozyme and (b) between water molecules in system I.

The effect of host framework on water diffusion in the lysozyme crystal is further examined by analyzing interactions in the system. **Figure 6.6a** shows the interaction energies between water and protein predicted from GROMOS96, AMBER03 and OPLS-AA along with SPC, SPC/E and TIP3P models. For all the force fields, the LJ interaction energy is approximately –1 kJ/mol, while the Coulombic interaction energy is within (–30, –40) kJ/mol. Therefore, the interaction between water and lysozyme is dominated by the Coulombic contribution. OPLS-AA predicts the largest interaction energy, followed by AMBER03, and then by GROMOS96. GROMOS96 is a united-atom force field, in which some H atoms are united to the neighboring C atoms, thus charge density is reduced in the united groups. This leads to a weaker Coulombic interaction between water and lysozyme. From the energetic perspective, lysozyme exhibits less affinity for water using GROMOS96; as a consequence, water diffusivity predicted from GROMOS96 is larger as observed in **Figure 6.5**. Our

analysis also demonstrated that the interaction between protein atoms is stronger using OPLS-AA and AMBER03 than using GROMOS96 (data not shown). The stronger interactions of protein with protein and with solvent using OPLS-AA and AMBER03 result in a more stable structure of protein, which in turn decreases water mobility therein. This is also consistent with our earlier results in Chapter 2, in which position restraints on protein atoms were found to slightly reduce water diffusivity.

Although water diffusion is significantly affected by water model, the interaction energies between water and protein are nearly the same for a force field with different water models. We further examine the interaction energies between water molecules. As shown in **Figure 6.6b**, again the Coulombic interaction is dominant for all force fields. The LJ interaction is positive, indicating that water molecules are closely adjacent to each other attributed to the strong Coulombic attraction. For each force field, the Coulombic interaction and the overall interaction predicted using SPC/E are stronger (more negative) than SPC and TIP3P; consequently, smaller water diffusivity are observed with SPC/E model.

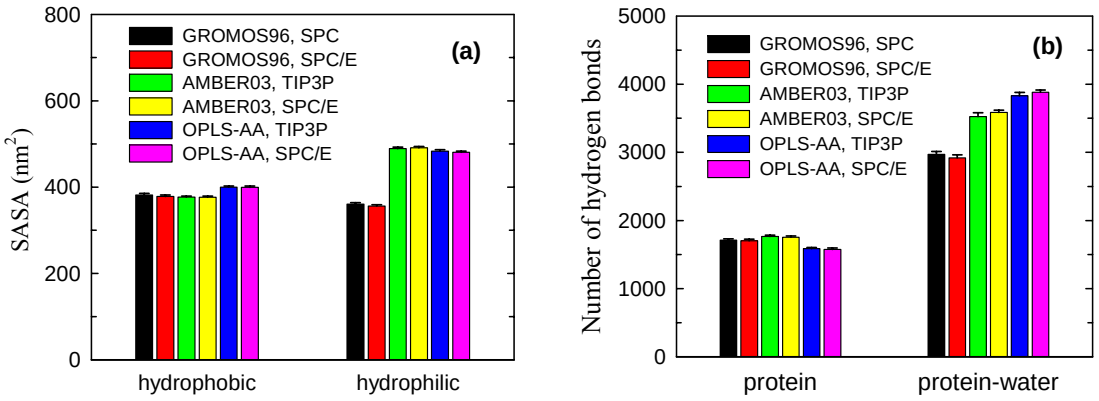

Figure 6.7. (a) Hydrophobic and hydrophilic solvent-accessible surface areas of lysozymes and (b) numbers of hydrogen bonds between lysozymes (including intra- and inter-) and between lysozyme and water in system I.

Figure 6.7a shows the hydrophobic and hydrophilic solvent-accessible surface areas (SASAs) of lysozymes. An atom in protein is considered to be hydrophobic if its charge falls within (–0.2*e*, 0.2*e*), and is hydrophilic otherwise.[50] SASAs are then determined using a probe with a diameter of 0.28 nm (approximately the size of a water molecule) to roll on van der Waals surface of protein crystal. The three force fields give similar hydrophobic SASAs and water model has a negligible influence. Nevertheless, AMBER03 and OPLS-AA predict greater hydrophilic SASAs than GROMOS96. As mentioned, the charge density of united groups is reduced in GROMOS96. In addition, the diameters for the major atoms (C, O, N, H) of lysozyme in GROMOS96 are generally smaller than in AMBER03 and OPLS-AA, as shown in **Table 6.2**. These two factors lead to a smaller surface area in the hydrophilic region predicted by GROMOS96.

Table 6.2. Lennard-Jones collision diameters of the major atoms (C, O, N and H) in lysozyme.

	GROMOS96	**AMBER03**	**OPLS-AA**
O (carbonyl)	0.263	0.296	0.296
N (amide)	0.298	0.325	0.325
C (aliphatic)	0.336 (C) 0.380 (CH_1)[a] 0.392 (CH_2)[a] 0.387 (CH_3)[a]	0.340	0.350
C (carbonyl)	0.336	0.340	0.375
C (aromatic)	0.336	0.340	0.355
H (aliphatic)	united to carbon	0.265	0.250
H (aromatic)	0.237	0.260	0.242

[a] CH_1, CH_2 and CH_3 stand for the united atoms wherein one, two and three hydrogen atoms are united to the neighboring carbon atoms, respectively.

Figure 6.7b shows the numbers of hydrogen bonds between lysozyme-lysozyme and of lysozyme-water. A hydrogen bond between a donor and an acceptor forms if

two geometric criteria are satisfied: (a) the distance between the donor and the acceptor is smaller than 0.35 nm and (b) the hydrogen-donor-acceptor angle is less than 30°. The three force fields give approximately the same number of hydrogen bonds between lysozymes (including intra- and inter-), but AMBER03 and OPLS-AA predict a greater number between lysozyme and water. This is consistent with the larger hydrophilic SASAs predicted by these two force fields. With AMBER03 and OPLS-AA, the surface area accessible to water is larger and more water molecules can approach lysozyme, thus the probability to form hydrogen bonds between water and lysozyme is enhanced. As in SASA, the number of hydrogen bonds weakly depends on water model.

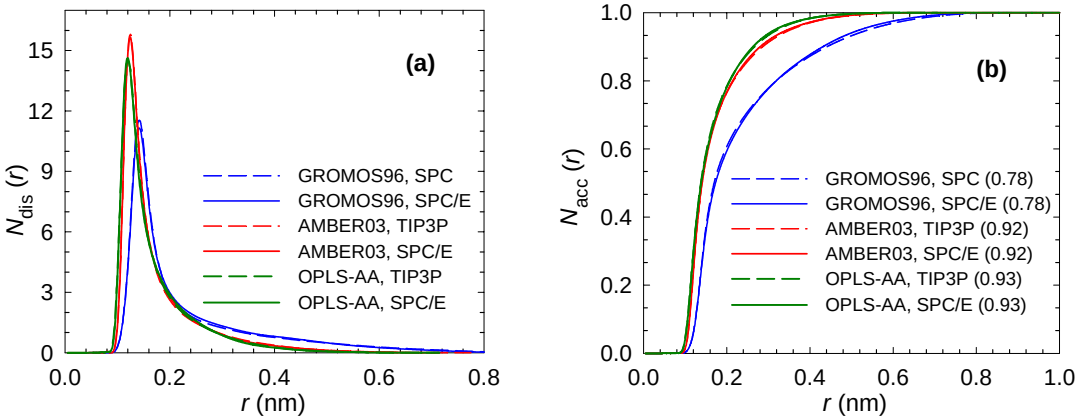

Figure 6.8. (a) Number distributions and (b) normalized accumulative number distributions of water molecules versus the distance from lysozyme surface in system I. Indicated in the parenthesis are the percentage of water within 0.3 nm thick hydration shell around lysozyme surface.

Due to the strong affinity for protein, water exhibits a layered structure proximal to protein surface. To quantify, the number distribution $N_{dis}(r)$ of water versus the distance r from lysozyme surface was calculated using Eq. (2.3). The van der Waals radii of protein atoms were taken into account. In **Figure 6.8a**, a distinct peak of $N_{dis}(r)$ is observed at 0.14 nm away from the surface using GROMOS96 with SPC or

SPC/E, which is equal to the radius of a water molecule. This clearly demonstrates the existence of a hydration shell surrounding the lysozyme surface. Using AMBER03 and OPLS-AA force fields, the location of peak moves to 0.12 nm. As clarified in **Figure 6.6**, AMBER03 and OPLS-AA predict stronger interaction energy between water and lysozyme than GROMOS96. Therefore, the location of peak is closer to lysozyme. **Figure 6.8b** shows the normalized accumulative number distributions of waters that are defined by Eq. (4.1).

As seen, more water molecules are located near the surface predicted by AMBER03 and OPLS-AA. If the hydration shell thickness is set as 0.3 nm, about the size of a water molecule, the fractions of water molecules within the shell are approximately 0.78, 0.92 and 0.93 predicted from GROMOS96, AMBER03 and OPLS-AA, respectively.

We can conclude that AMBER03 and OPLS-AA give very similar results for the thermal fluctuations of lysozyme atoms, the directional and average diffusivities of water, the SASAs of hydrophobic and hydrophilic residues, the numbers of hydrogen bonds. The predictions from AMBER03 and OPLS-AA reproduce the available experimental data fairly well and outperform over GROMOS96. While SPC, SPC/E and TIP3P models give close results for most energetic and structural properties, SPC/E is better in predicting water diffusivities.

6.3.2 Ion Mobility and Electrical Conductivity in System II

In this section, we examine ion mobility and electric conductivity of NaCl in the lysozyme crystal. A clear understanding toward the behavior of ions in a protein crystal is of significance for separation, ion exchange and biosensing. As discussed in the previous section, SPC/E model describes water diffusion in the crystal more precisely than SPC and TIP3P. Consequently, simulations in system II were carried

out only with SPC/E model and three biomolecular force fields. Indeed, a test revealed that SPC/E is superior to other water models (data not shown). In addition, the KB model for NaCl was considered by incorporating with OPLS-AA.

The mobility Mi of ion species i was estimated by the Einstein equation[152]

$$M_i = \frac{D_i |q_i| F_a}{RT} \tag{6.2}$$

where D_i is diffusivity, q_i is the charge of ith species, F_a is the Faraday constant (9.6485×10^4 C/mol), R is the gas constant, and T is temperature. The mobility M_z of Na^+ and Cl^- was calculated along the z direction, which is the axis of main pore in the tetragonal lysozyme crystal. As shown in **Figure 6.9a**, all the three force fields overestimate the mobility of Na^+ and Cl^- upon comparison with the experimental values. Particularly, GROMOS96 gives much higher mobility than AMBER03 and OPLS-AA. Incorporated with KB model for NaCl, the predictions of OPLS-AA are improved.

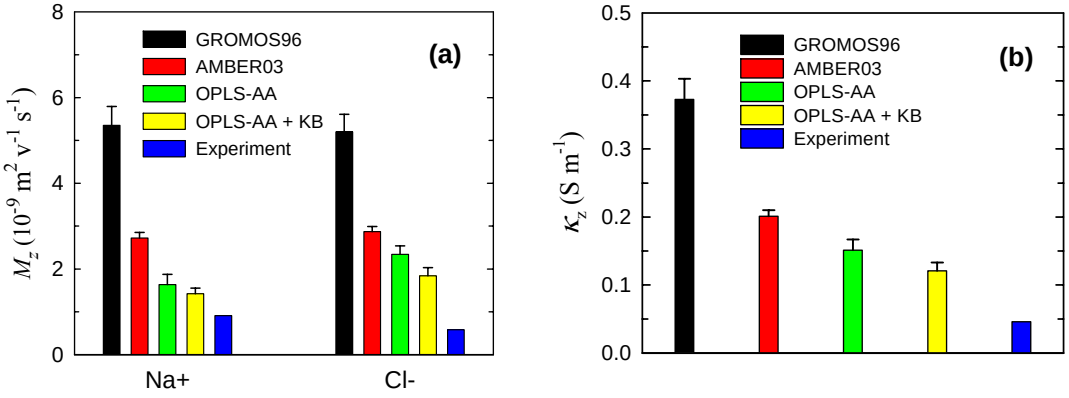

Figure 6.9. (a) Mobility of Na^+ and Cl^- and (b) electrical conductivity along the z direction in system II. OPLS-AA + KB denotes that OPLS-AA was used for protein and KB model was used for Na^+ and Cl^-.

The electrical conductivity κ was evaluated from the variation of Nernst-Einstein relation[152]

$$\kappa = \frac{F_a^2}{RT}\sum_i D_i c_i |q_i|^2 (1-\Delta) = F_a \sum_i M_i c_i |q_i| (1-\Delta) \tag{6.3}$$

where c_i is molar concentration, and Δ is the cross correlation term that reflects the correlations between ions.[199] The original Nernst-Einstein relation is described in Eq. (3.2). The cross term Δ is usually marginal and was neglected in our calculation. **Figure 6.9b** shows the electrical conductivity κ_z along the z direction. Similar to **Figure 6.9a** for the ion mobility, all force fields overestimate the electrical conductivity. Nevertheless, OPLS-AA with KB model yields the best prediction. A separation simulation was carried out in 0.5 M NaCl bulk solution with SPC/E model for water and KB model for NaCl. The mobility of Na^+ and Cl^- is (3.79 ± 0.23) and (5.37 ± 0.28) × 10^{-8} m^2 v^{-1} s^{-1} respectively, and the conductivity is (4.42 ± 0.25) S m^{-1}, close to experimental 4.05 S m^{-1} at 18 °C.[200] Compared to these bulk values, the measured ion mobility and conductivity in the crystal are reduced by one ~ two orders of magnitude due to the steric restraints and surface interactions with the crystal framework.

The large deviations in the mobility and conductivity of NaCl between predictions and experiments are plausibly due to several reasons. First, the force fields used in our simulations were developed for protein, water and individual ions. These force fields were fit to some experimental data over a (limited) range of conditions. From them, useful information can be obtained for the properties that have been fit; however their semi-empirical nature may lead to inaccurate predictions for other properties, other systems or at other conditions. Though a force field can predict the properties of lysozyme, water, and counterions fairly well, it is not necessarily accurate for a pair of ions (NaCl) in the lysozyme crystal. As mentioned, most force fields for Na^+ and Cl^- were derived from the properties of individual ions, thus they may not be able to

precisely describe the combination of individual ions (NaCl). For instance, the abnormal clustering of excess K^+ or Na^+ with Cl^- observed using AMBER03 suggests that ions are not well parameterized in the current biomolecular force fields.[201] Furthermore, the use of pair-wise potentials could also be the cause of inaccurate predictions. Second, as discussed earlier, the simulation model is not exactly the same as the experimental sample. There were cross-linkers in the real sample, which exerted additional steric obstacles for the motion of ions. However, we have no exact knowledge about the degree to which cross-linkers decrease the mobility. Morozov et al. estimated that the existence of cross-linkers in the tetragonal lysozyme crystal might cause 17% decrease in water diffusivity along the *x* or *y* direction and a smaller drop in the *z* direction.[52] Although their estimation was for water, similar effect also came into play for ions. Besides cross-linkers, Cl^- concentration (0.508 M) in our simulation was a bit larger than that in experiment (0.464 M), and this may lead to 7% increase in electrical conductivity by assuming that ion mobility is independent of concentration within the range under study. In addition, the simulation system was an infinite crystal, as opposed to the real sample with a finite size and interface between sample and buffer solution. However, the influence of interface on ion mobility and electrical conductivity is not clear. Third, the cross correlation term Δ in eq. (6.7) was assumed to be zero. This may be inaccurate because ion clusters were formed between Na^+ and Cl^-, which contributes a non-zero cross correlation term and hence reduces the predicted conductivity. Another reason is the lack of sufficient accuracy of experimental data. The concentrations of Na^+ and Cl^- in the protein crystal were not directly measured in the experiment, but estimated using theoretical equations and empirical parameters.[53]

6.4 Conclusions

We have assessed different biomolecular force fields (OPLS-AA, AMBER03 and GROMOS96) and water models (SPC, SPC/E and TIP3P) for MD simulations in the tetragonal lysozyme crystal. Using OPLS-AA and AMBER03, the system approaches equilibration faster and exhibits smaller RMSDs. Similar pattern is observed using different force fields in temperature factors, in which small values are attributed to the formation of relatively stable secondary structures. OPLS-AA and AMBER03 reproduce the experimental temperature factors fairly well and perform better than GROMOS96. Compared to the bulk phase, water diffusion in the crystal is reduced by approximately ten fold due to the geometric restraints and surface interactions. Experimentally measured directional and average diffusivities of water are well predicted by AMBER03 and OPLS-AA with SPC/E model, but overestimated by other combinations. The interaction between water and lysozyme is dominated by the Coulombic contribution and predicted to be the largest by OPLS-AA followed by AMBER03 and GROMOS96. Larger hydrophilic solvent accessible surface area of lysozyme, more hydrogen bonds between lysozyme and water, and higher percentage of water molecules in hydration shell are predicted by AMBER03 and OPLS-AA than GROMOS96. However, the hydrophobic surface area of lysozyme and the number of hydrogen bonds between lysozymes are found being close from different force fields. The water model has a significant effect on the dynamic diffusion of water, but not on the energetic and structural properties of the system. Compared to aqueous solution, the mobility and conductivity are reduced by approximately one ~ two orders of magnitude. All the force fields overestimate the experimental mobility and conductivity of NaCl in the lysozyme crystal, while a combination of OPLS-AA for protein and KB model for ions exhibits the best performance. Our results provide

useful guidance on the selection of force fields for molecular simulations in protein crystal and on the aspect for improving the current force fields to more accurately describe confined fluids in biological nanospace.

Chapter 7 Conclusion and Outlook

7.1 Conclusion

In this thesis, transport and separation in protein crystals have been investigated using MD simulations. The first part was focused on the transport of water and ions in different protein crystals, as well as the effects of electric field on the transport of ions. In the second part, achiral and chiral separation processes were examined for amino acids, and the separation mechanisms were subsequently elucidated. Finally, biomolecular force fields were evaluated for simulation in a protein crystal. Major findings are summarized below.

Structural and dynamic features of proteins

Protein structures were examined in three crystals including tetragonal lysozyme, orthorhombic lysozyme and tetragonal thermolysin; and the structural stability of tetragonal lysozyme was further studied under electric field.

(1) The residues in the secondary structures of proteins exhibit relatively weaker fluctuations than other residues due to hydrogen bonding.

(2) The SASA per residue is nearly identical in the three protein crystals; the hydrophobic and hydrophilic SASAs are approximately the same in crystal.

(3) The stability of lysozyme is reduced slightly upon exposure to electric field as evidenced from the increases in the RMSDs of heavy atoms and in hydrophobic SASA.

Structural and dynamic properties of water and ions

The structural and dynamic properties of water and ions are significantly affected by the confinement in protein crystals.

(1) Water and ions exhibit layered structures on the protein surface; cations (e.g., Na^+ and Ca^{2+}) exhibit two layers structures in the lysozyme surface, but anions (e.g., Cl^-) only have one. The local structures of water and ions on the protein surface depend on the type of protein.

(2) Water coordination numbers of ions in the crystal are smaller than in aqueous bulk solution; however, the reverse is true for anionic (e.g., Cl^-) coordination numbers of cations (e.g., Na^+ and Ca^{2+}).

(3) Diffusivities of water and ions in protein crystals are reduced by one or two orders of magnitude than in bulk solution. The higher the porosity of protein crytals, the smaller is the reduction.

(4) Water and ions exhibit anisotropic diffusion in protein crystals with anisotropic nanoporous structures.

Effects of external electric field

(1) Electric field has a marginal effect on the structural properties of water and ions, including the local layered structures on protein surface and the coordination numbers of ions or water.

(2) Water molecules tend to align preferentially parallel to the electric field and the dipole moment along the direction of electric field rises linearly with increasing electric field strength.

(3) The drift velocities of ions and electrical current increase proportionally to the electric field strength.

(4) Equilibrium and non-equilibrium MD simulations give consistent electrical conductivity.

Separation of amino acids

Non-equilibrium MD simulations were used to explore the separation of amino acid mixtures in a glucose isomerase crystal and the separation of *D*/*L*-phenylglycines in a thermolysin crystal. The simulated elution orders in the two cases were in accord with experimental observations.

(1) The flowing rate of mobile phase and the concentration of solutes can be used as operating variables to optimize separation process.

(2) Amino acids differing in size, charge and hydrophobicity are separated as a consequence of the counterbalance between solute-water and solute-protein interactions; however, the chiral separation of racemic *D*/*L*-phenylglycines is primarily attributed to different solute-protein interactions.

(3) Geometrical analysis (e.g., number distribution) is consistent with energetic analysis, and both provide microscopic insights into the separation mechanisms.

Simulations are a useful tool to study the separation capability of protein crystals and probe the separation mechanisms involved from a microscopic perspective.

Assessment of force fields

Three biomolecular force fields (OPLS-AA, AMBER03 and GROMOS96) in conjunction with SPC, SPC/E and TIP3P water models were assessed for simulation in a tetragonal lysozyme crystal.

(1) OPLS-AA and AMBER03 accurately reproduce experimental *B*-factors, while GROMOS96 overestimate them. Water model plays a marginal role.

(2) OPLS-AA and AMBER03 predict similar properties including hydrophilic SASA of lysozyme, hydrogen bond number between lysozyme and water, and percentage of water in hydration shell; however, GROMOS96 predicts lower values for all these properties.

(3) OPLS-AA and AMBER03 along with SPC/E model fairly well reproduce the experimental diffusivities of water, while GROMOS96 and other two water models (e.g., SPC and TIP3P) overestimate.

(4) SPC, SPC/E and TIP3P water models have similar performance in most energetic and structural properties, but SPC/E outperforms in water diffusion.

(5) A combination of OPLS-AA for lysozyme and the Kirkwood-Buff model for ions is superior in predicting the mobility of ions and electrical conductivity of NaCl.

7.2 Outlook

To facilitate the emerging applications of protein crystals as bionanoporous materials, further studies on protein crystals in both experimental and computational aspects are required. Molecular simulations can quantitatively interpret experimental results and complement experimental measurements. In the context of molecular simulations, several recommendations are presented below.

First, advanced simulation techniques need to be developed and adopted. Currently most MD simulations are run in the scale of nanoseconds. For the transport or separation of intermediate-sized molecules such as surfactants and peptides in protein crystals, a longer time scale is indispensable for sufficient sampling. It has to resort to more effective simulation techniques. As a compromise of accuracy and computational time, multiscale modeling approaches have been proposed ranging from the low-resolution coarse-grained (CG) models to the high-resolution fully atomistic counterparts.[171,172] The CG model is a feasible option to achieve longer time

scale for large systems without the loss of fundamental features. Fore instance, the MARTINI CG force field could achieve a speedup factor of 3–4 orders of magnitude over atomistic simulations.[202] The flowing rate of the mobile phase in the atomistic simulations in Chapters 4–5 are several orders of magnitude greater than in practice. If the CG models are employed, the mobile phase can be simulated with a flowing rate comparable to the practical rate in a liquid chromatography.

Protein crystals are used not only as separation materials, but also as biocatalysts and biosensors. However, biocatalysis and biosensing involve reactions that are beyond the capability of classical simulations. Quantum-level simulation techniques are necessary to describe such systems involving the movements of electrons; however, they are very computationally extensive. One feasible option is to use hybrid quantum-classical simulation techniques in which the reacting core of the system is treated quantum-mechanically with the reminder being simulated by classical method.

Another issue is to develop more realistic models for protein crystals. Our simulations used ideal models for protein crystals, in which cross-linkers were neglected. Nevertheless, the stability of proteins and the transport of guest molecules would be varied in the presence of cross-linkers. The structure of tetragonal lysozyme crystal cross-linked by glutaraldehyde was resolved recently,[203] and thus it is possible now to incorporate the effect of cross-linker without a serious effect on computation efficiency, if a suitable force field of glutaraldehyde is developed.

Bibliography

(1) Vilenchik, L. Z.; Griffith, J. P.; St Clair, N.; Navia, M. A.; Margolin, A. L. *J. Am. Chem. Soc.* **1998**, 120, 4290-4294.

(2) Margolin, A. L.; Navia, M. A. *Angew. Chem. Int. Ed.* **2001**, 40, 2205-2222.

(3) Oki, H.; Matsuura, Y.; Komatsu, H.; Chernov, A. A. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **1999**, 55, 114-121.

(4) McPherson, A. *Preparation and analysis of protein crystals.*; John Wiley & Sons, Inc.: New York, **1982**.

(5) Pastinen, O.; Jokela, J.; Eerikainen, T.; Schwabe, T.; Leisola, M. *Enzyme Microb. Technol.* **2000**, 26, 550-558.

(6) Stclair, N. L.; Navia, M. A. *J. Am. Chem. Soc.* **1992**, 114, 7314-7316.

(7) Wang, Y. F.; Yakovlevsky, K.; Zhang, B. L.; Margolin, A. L. *J. Org. Chem.* **1997**, 62, 3488-3495.

(8) Visuri, K.; Pastinen, O.; Wu, X. Y.; Makinen, K.; Leisola, M. *Biotechnol. Bioeng.* **1999**, 64, 377-380.

(9) Persichetti, R. A.; Stclair, N. L.; Griffith, J. P.; Navia, M. A.; Margolin, A. L. *J. Am. Chem. Soc.* **1995**, 117, 2732-2737.

(10) Quiocho, F. A.; Richards, F. M. *Biochemistry* **1966**, 5, 4062-4076.

(11) Fitzpatrick, P. A.; Steinmetz, A. C. U.; Ringe, D.; Klibanov, A. M. *Proc. Natl. Acad. Sci. USA.* **1993**, 90, 8653-8657.

(12) Lusty, C. J. *J. Appl. Crystallogr.* **1999**, 32, 106-112.

(13) Yonath, A.; Sielecki, J. M.; Moult, J.; Podjarny, A.; Traub, W. *Biochemistry* **1977**, 16, 1413-1417.

(14) Quiocho, F. A.; Richards, F. M. *Proc. Natl. Acad. Sci. USA.* **1964**, 52, 833-839.

(15) Cabirol, F. L.; Hanefeld, U.; Sheldon, R. A. *Adv. Synth. Catal.* **2006**, 348, 1645-1654.

(16) Wang, Y. J.; Caruso, F. *Adv. Mater.* **2006**, 18, 795-800.

(17) Di Profio, G.; Perrone, G.; Curcio, E.; Cassetta, A.; Lamba, D.; Drioli, E. *Ind. Eng. Chem. Res.* **2005**, 44, 10005-10012.

(18) Lee, T. S.; Vaghjiani, J. D.; Lye, G. J.; Turner, M. K. *Enzyme Microb. Technol.* **2000**, 26, 582-592.

(19) Roy, J. J.; Abraham, T. E. *Chem. Rev.* **2004**, 104, 3705-3721.

(20) Falkner, J. C.; Al-Somali, A. M.; Jamison, J. A.; Zhang, J. Y.; Adrianse, S. L.; Simpson, R. L.; Calabretta, M. K.; Radding, W.; Phillips, G. N.; Colvin, V. L. *Chem. Mat.* **2005**, 17, 2679-2686.

(21) Klyushnichenko, V. *Curr. Opin. Drug Discov. Dev.* **2003**, 6, 848-854.

(22) Jokela, J.; Leisola, M. *J. Sep. Sci.* **2004**, 27, 1491-1497.

(23) Roy, J. J.; Abraham, T. E.; Abhijith, K. S.; Kumar, P. V. S.; Thakur, M. S. *Biosens. Bioelectron.* **2005**, 21, 206-211.

(24) Laothanachareon, T.; Champreda, V.; Sritongkham, P.; Somasundrum, M.; Surareungchai, W. *World J. Microbiol. Biotechnol.* **2008**, 24, 3049-3055.

(25) de Mattos, I. L.; Lukachova, L. V.; Gorton, L.; Laurell, T.; Karyakin, A. A. *Talanta* **2001**, 54, 963-974.

(26) Cohen-Hadar, N.; Wine, Y.; Lagziel-Simis, S.; Moscovich-Dagan, H.; Dror, Y.; Frolow, F.; Freeman, A. *J. Porous Media* **2009**, 12, 213-220.

(27) Lagziel-Simis, S.; Cohen-Hadar, N.; Moscovich-Dagan, H.; Wine, Y.; Freeman, A. *Curr. Opin. Biotech.* **2006**, 17, 569-573.

(28) Cohen-Hadar, N.; Wine, Y.; Nachliel, E.; Huppert, D.; Gutman, M.; Frolow, F.; Freeman, A. *Biotechnol. Bioeng.* **2006**, 94, 1005-1011.

(29) Falkner, J. C.; Turner, M. E.; Bosworth, J. K.; Trentler, T. J.; Johnson, J. E.; Lin, T. W.; Colvin, V. L. *J. Am. Chem. Soc.* **2005**, 127, 5274-5275.

(30) Wang, Y. F.; Yakovlevsky, K.; Khalaf, N.; Zhang, B. L.; Margolin, A. L. *Enzyme Eng. Xiii* **1996**, 799, 777-783.

(31) Lalonde, J. J.; Govardhan, C.; Khalaf, N.; Martinez, A. G.; Visuri, K.; Margolin, A. L. *J. Am. Chem. Soc.* **1995**, 117, 6845-6852.

(32) St Clair, N.; Wang, Y. F.; Margolin, A. L. *Angew. Chem. Int. Ed.* **2000**, 39, 380-383.

(33) Costes, D.; Wehtje, E.; Adlercreutz, P. *J. Mol. Catal. B-Enzym.* **2001**, 11, 607-612.

(34) Wickremasinghe, N. S.; Hafner, J. H. *Nano Lett.* **2005**, 5, 2418-2421.

(35) Clair, N. S.; Shenoy, B.; Jacob, L. D.; Margolin, A. L. *Proc. Natl. Acad. Sci. USA.* **1999**, 96, 9469-9474.

(36) van Gunsteren, W. F.; Dolenc, J. *Biochem. Soc. Trans.* **2008**, 36, 11-15.

(37) Kalra, A.; Garde, S.; Hummer, G. *Proc. Natl. Acad. Sci. USA.* **2003**, 100, 10175-10180.

(38) Holt, J. K.; Park, H. G.; Wang, Y. M.; Stadermann, M.; Artyukhin, A. B.; Grigoropoulos, C. P.; Noy, A.; Bakajin, O. *Science* **2006**, 312, 1034-1037.

(39) Bekker, H. *PhD dissertation in University of Groningen* **1996**.

(40) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, 18, 1463-1472.

(41) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, 98, 10089-10092.

(42) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, 103, 8577-8593.

(43) 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. *J. Phys. Chem. B* **1998**, 102, 3586-3616.

(44) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, 117, 5179-5197.

(45) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J. Comput. Chem.* **2003**, 24, 1999-2012.

(46) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, 118, 11225-11236.

(47) Jorgensen, W. L.; Tiradorives, J. *J. Am. Chem. Soc.* **1988**, 110, 1657-1666.

(48) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular simulations: GROMOS96 manual and user guide*; Hochschulverlag AG an der ETH Zurich: Zurich, Switzerland, **1996**.

(49) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, 111, 7812-7824.

(50) Van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C. *Gromacs user mannal v3.3*, **2006**.

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

(52) Morozov, V. N.; Kachalova, G. S.; Evtodienko, V. U.; Lanina, N. F.; Morozova, T. Y. *Eur. Biophys. J.* **1995**, 24, 93-98.

(53) Morozova, T. Y.; Kachalova, G. S.; Lanina, N. F.; Evtodienko, V. U.; Botin, A. S.; Shlyapnikova, E. A.; Morozov, V. N. *Biophys. Chem.* **1996**, 60, 1-16.

(54) Bon, C.; Dianoux, A. J.; Ferrand, M.; Lehmann, M. S. *Biophys. J.* **2002**, 83, 1578-1588.

(55) Cvetkovic, A.; Picioreanu, C.; Straathof, A. J. J.; Krishna, R.; van der Wielen, L. A. M. *J. Phys. Chem. B* **2005**, 109, 10561-10566.

(56) Cvetkovic, A.; Picioreanu, C.; Straathof, A. J. J.; Krishna, R.; van der Wielent, L. A. M. *J. Am. Chem. Soc.* **2005**, 127, 875-879.

(57) Cvetkovic, A.; Straathof, A. J. J.; Hanlon, D. N.; Van der Zwaag, S.; Krishna, R.; Van der Wielen, L. A. M. *Biotechnol. Bioeng.* **2004**, 86, 389-398.

(58) Cvetkovic, A.; Straathof, A. J. J.; Krishna, R.; van der Wielen, L. A. M. *Langmuir* **2005**, 21, 1475-1480.

(59) Cvetkovic, A.; Zomerdijk, M.; Straathof, A. J. J.; Krishna, R.; van der Wielen, L. A. M. *Biotechnol. Bioeng.* **2004**, 87, 658-668.

(60) Bishop, W. H.; Richards, F. M. *J. Mol. Biol.* **1968**, 38, 315-328.

(61) Johnson, L. N.; Hajdu, J. *Synchrotron studies on enzyme catalysis in crystals. In Synchrotron Radiation in biophysics*; Ellis Horwood Ltd.: Chichester, UK, **1990**.

(62) Gorti, S.; Zuk, W. M.; Konnert, J.; Ward, K.; Tanaka, T.; Yang, H. *J. Cryst. Growth* **2001**, 232, 256-261.

(63) Velev, O. D.; Kaler, E. W.; Lenhoff, A. M. *J. Phys. Chem. B* **2000**, 104, 9267-9275.

(64) Pastinen, O.; Visuri, K.; Leisola, M. *Biotechnol. Tech.* **1998**, 12, 557-560.

(65) Haginaka, J. *J. Chromatogr. A* **2001**, 906, 253-273.

(66) Hermans, J.; Vacatello, M. *Biophys. J.* **1980**, 32, 87-88.

(67) Vangunsteren, W. F.; Karplus, M. *Nature* **1981**, 293, 677-678.

(68) Vangunsteren, W. F.; Karplus, M. *Biochemistry* **1982**, 21, 2259-2274.

(69) Vangunsteren, W. F.; Berendsen, H. J. C.; Hermans, J.; Hol, W. G. J.; Postma, J. P. M. *Proc. Natl. Acad. Sci. USA.-Biol.* **1983**, 80, 4315-4319.

(70) Heiner, A. P.; Berendsen, H. J. C.; Vangunsteren, W. F. *Proteins* **1992**, 14, 451-464.

(71) York, D. M.; Darden, T. A.; Pedersen, L. G.; Anderson, M. W. *Biochemistry* **1993**, 32, 1443-1453.

(72) vanAalten, D. M. F.; Conn, D. A.; deGroot, B. L.; Berendsen, H. J. C.; Findlay, J. B. C.; Amadei, A. *Biophys. J.* **1997**, 73, 2891-2896.

(73) Hery, S.; Genest, D.; Smith, J. C. *J. Mol. Biol.* **1998**, 279, 303-319.

(74) Stocker, U.; Spiegel, K.; van Gunsteren, W. F. *J. Biomol. NMR* **2000**, 18, 1-12.

(75) Walser, R.; Hunenberger, P. H.; van Gunsteren, W. F. *Proteins* **2002**, 48, 327-340.

(76) Yu, H. B.; Ramseier, M.; Burgi, R.; van Gunsteren, W. F. *ChemPhysChem* **2004**, 5, 633-641.

(77) Cerutti, D. S.; Le Trong, I.; Stenkamp, R. E.; Lybrand, T. P. *Biochemistry* **2008**, 47, 12065-12077.

(78) Cerutti, D. S.; Le Trong, I.; Stenkamp, R. E.; Lybrand, T. P. *J. Phys. Chem. B* **2009**, 113, 6971-6985.

(79) Jiang, J. W.; Sandler, S. I. *Langmuir* **2006**, 22, 5702-5707.

(80) Babarao, R.; Hu, Z. Q.; Jiang, J. W.; Chempath, S.; Sandler, S. I. *Langmuir* **2007**, 23, 659-666.

(81) Duren, T.; Sarkisov, L.; Yaghi, O. M.; Snurr, R. Q. *Langmuir* **2004**, 20, 2683-2689.

(82) Smit, B.; Krishna, R. *Curr. Opin. Solid. St. M.* **2001**, 5, 455-461.

(83) Sanborn, M. J.; Snurr, R. Q. *Sep. Purif. Technol.* **2000**, 20, 1-13.

(84) Skoulidas, A. I.; Sholl, D. S. *J. Phys. Chem. B* **2005**, 109, 15760-15768.

(85) Jiang, J. W.; Sandler, S. I.; Smit, B. *Nano Lett.* **2004**, 4, 241-244.

(86) Jiang, J. W.; Sandler, S. I. *J. Am. Chem. Soc.* **2005**, 127, 11989-11997.

(87) Jiang, J. W.; Klauda, J. B.; Sandler, S. I. *J. Phys. Chem. B* **2005**, 109, 4731-4737.

(88) Jiang, J. W.; Klauda, J. B.; Sandler, S. I. *Langmuir* **2003**, 19, 3512-3518.

(89) Shi, W.; Johnson, J. K. *Phys. Rev. Lett.* **2003**, 91, 015504.

(90) Skoulidas, A. I.; Ackerman, D. M.; Johnson, J. K.; Sholl, D. S. *Phys. Rev. Lett.* **2002**, 89, 185901.

(91) Malek, K.; Odijk, T.; Coppens, M. O. *Chemphyschem* **2004**, 5, 1596-1599.

(92) Malek, K.; Odijk, T.; Coppens, M. O. *Nanotechnology* **2005**, 16, S522-S530.

(93) Geremia, S.; Campagnolo, M.; Demitri, N.; Johnson, L. N. *Structures* **2006**, 14, 393-400.

(94) Malek, K. *Comput. Sci. Eng.* **2007**, 9, 70-75.

(95) Malek, K. *Biotechnol. Lett.* **2007**, 29, 1865-1873.

(96) McPherson, A. In Preparation and Analysis of Protein Crystals; McPherson, A., Ed.; John Wiley & Sons: New York, **1982**.

(97) Matthews, B. W. *J. Mol. Biol.* **1968**, 33, 491-497.

(98) Bizzarri, A. R.; Cannistraro, S. *J. Phys. Chem. B* **2002**, 106, 6617-6633.

(99) Szent-Gyorgyi, A. *The living state*; Academic Press: NY, **1972**.

(100) MacKinnon, R. *FEBS Lett.* **2003**, 555, 62-65.

(101) Wilson, K. P.; Malcolm, B. A.; Matthews, B. W. *J. Biol. Chem.* **1992**, 267, 10842-10849.

(102) Artymiuk, P. J.; Blake, C. C. F.; Rice, D. W.; Wilson, K. S. *Acta Crystallogr. Sect. B-Struct. Commun.* **1982**, 38, 778-783.

(103) Hausrath, A. C.; Matthews, B. W. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2002**, 58, 1002-1007.

(104) Kuramitsu, S.; Hamaguchi, K. *J. Biochem.* **1980**, 87, 1215-1219.

(105) Schuler, L. D.; Daura, X.; Van Gunsteren, W. F. *J. Comput. Chem.* **2001**, 22, 1205-1218.

(106) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Interaction models for water in relation to protein hydration. In: Intermolecular Forces. Pullman, B. ed. . D. Reidel*; Publishing Company Dordrecht, **1981**.

(107) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, 13, 952-962.

(108) Lins, L.; Thomas, A.; Brasseur, R. *Protein Sci.* **2003**, 12, 1406-1417.

(109) Smart, O. S.; Neduvelil, J. G.; Wang, X.; Wallace, B. A.; Sansom, M. S. P. *J. Mol. Graph.* **1996**, 14, 354-360.

(110) Beckstein, O.; Sansom, M. S. P. *Proc. Natl. Acad. Sci. USA.* **2003**, 100, 7063-7068.

(111) Im, W.; Roux, B. *J. Mol. Biol.* **2002**, 319, 1177-1197.

(112) Beckstein, O.; Biggin, P. C.; Sansom, M. S. P. *J. Phys. Chem. B* **2001**, 105, 12902-12905.

(113) Beckstein, O.; Sansom, M. S. P. *Phys. Biol.* **2004**, 1, 42-52.

(114) Tieleman, D. P.; Biggin, P. C.; Smith, G. R.; Sansom, M. S. P. *Q. Rev. Biophys.* **2001**, 34, 473-561.

(115) Law, R. J.; Tieleman, D. P.; Sansom, M. S. P. *Biophys. J.* **2003**, 84, 14-27.

(116) Hua, L.; Huang, X. H.; Zhou, R. H.; Berne, B. J. *J. Phys. Chem. B* **2006**, 110, 3704-3711.

(117) Merzel, F.; Smith, J. C. *Proc. Natl. Acad. Sci. USA.* **2002**, 99, 5378-5383.

(118) Svergun, D. I.; Richard, S.; Koch, M. H. J.; Sayers, Z.; Kuprin, S.; Zaccai, G. *Proc. Natl. Acad. Sci. USA.* **1998**, 95, 2267-2272.

(119) Bagchi, B. *Chem. Rev.* **2005**, 105, 3197-3219.

(120) Allen, M. P.; Tildesley, D. J. *Computer simulations of liquids*; Oxford Science: Oxford, **1987**.

(121) Wu, Y. J.; Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2006**, 124, 024503.

(122) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, 91, 6269-6271.

(123) Yeung, K. K.-C.; Lucy, C. A. *Electrophoresis* **1999**, 20, 2554-2559.

(124) Seo, Y. S.; Luo, H.; Samuilov, V. A.; Rafailovich, M. H.; Sokolov, J.; Gersappe, D.; Chu, B. *Nano Lett.* **2004**, 4, 659-664.

(125) Voet, D.; Voet, J. G. *Biochemistry*; Wiley: New York, **2004**.

(126) Jackson, J. D. *Classical Electrodynamics*; Wiley: New York, **1975**.

(127) Tang, Y. W.; Szalai, I.; Chan, K. Y. *J. Phys. Chem. A* **2001**, 105, 9616-9623.

(128) Tang, Y. W.; Chan, K. Y.; Szalai, I. *J. Phys. Chem. B* **2004**, 108, 18204-18213.

(129) Dzubiella, J.; Allen, R. J.; Hansen, J. P. *J. Chem. Phys.* **2004**, 120, 5001.

(130) Dzubiella, J.; Hansen, J. P. *J. Chem. Phys.* **2005**, 122, 234706.

(131) Murad, S.; Lin, J. *Ind. Eng. Chem. Res.* **2002**, 41, 1076-1083.

(132) Murad, S.; Jia, W.; Krishnamurthy, M. *Mol. Phys.* **2004**, 102, 2103-2112.

(133) Hwang, H.; Schatz, G. C.; Ratner, M. A. *J. Phys. Chem. B* **2006**, 110, 6999-7008.

(134) Hwang, H.; Schatz, G. C.; Ratner, M. A. *J. Chem. Phys.* **2007**, 127.

(135) Dimova, R.; Riske, K. A.; Aranda, S.; Bezlyepkina, N.; Knorr, R. L.; Lipowsky, R. *Soft Matter* **2007**, 3, 817-827.

(136) Evans, D. J.; Morriss, G. P. *Comput. Phys. Rep.* **1984**, 1, 297-343.

(137) Svishchev, I. M.; Kusalik, P. G. *Physica A* **1993**, 192, 628-646.

(138) Pethica, B. A.; Hall, D. G. *J. Colloid Interface Sci.* **1982**, 85, 41-51.

(139) Siu, S. W. I.; Bockmann, R. A. *J. Struct. Biol.* **2007**, 157, 545-556.

(140) Tang, Y. W.; Zhang, Q. Y.; Chan, K. Y. *Chem. Phys. Lett.* **2004**, 385, 202-207.

(141) Dill, K. A. *Biochemistry* **1990**, 29, 7133-7155.

(142) Collins, K. D. *Methods* **2004**, 34, 300-311.

(143) Acharya, A. S.; Taniuchi, H. *Proc. Natl. Acad. Sci. USA.* **1977**, 74, 2362-2366.

(144) Kabsch, W.; Sander, C. *Biopolymers* **1983**, 22, 2577-2637.

(145) Robertson, K. M.; Tieleman, D. P. *FEBS Lett.* **2002**, 528, 53-57.

(146) Cramer, T.; Zerbetto, F.; Garcia, R. *Langmuir* **2008**, 24, 6116-6120.

(147) Freund, J. B. *J. Chem. Phys.* **2002**, 116, 2194-2200.

(148) Qiao, R.; Aluru, N. R. *J. Chem. Phys.* **2003**, 118, 4692-4701.

(149) Qiao, R.; Aluru, N. R. *Colloid Surf. A-Physicochem. Eng. Asp.* **2005**, 267, 103-109.

(150) Lyklema, J. *Fundamentals of Interface and Colloid Science*; Academic Press: San Diego, **1995**.

(151) Ataka, M.; Tanaka, S. *Biopolymers* **1980**, 19, 669-679.

(152) Plawsky, J. L. *Transport Phenomena Fundamentals*; CRC Press, **2001**.

(153) Stevenson, R.; Johnson, E. L. *Basic liquid chromatography*; Varian: New York, **1978**.

(154) Rafferty, J. L.; Siepmann, J. I.; Schure, M. R. *J. Chromatogr. A* **2008**, 1204, 20-27.

(155) Zhao, C. F.; Cann, N. M. *J. Chromatogr. A* **2007**, 1149, 197-218.

(156) Zhao, C.; Cann, N. M. *J. Chromatogr. A* **2006**, 1131, 110-129.

(157) Nita, S.; Cann, N. M. *J. Phys. Chem. B* **2008**, 112, 13022-13037.

(158) Rafferty, J. L.; Siepmann, J. I.; Schure, M. R. *J. Chromatogr. A* **2008**, 1204, 11-19.

(159) Braun, J.; Fouqueau, A.; Bemish, R. J.; Meuwly, M. *Phys. Chem. Chem. Phys.* **2008**, 10, 4765-4777.

(160) Fouqueau, A.; Meuwly, M.; Bemish, R. J. *J. Phys. Chem. B* **2007**, 111, 10208-10216.

(161) Beck, T. L.; Klatte, S. J. *Unif. Chromatogr.* **2000**, 748, 67-81.

(162) Leisola, M.; Jokela, J.; Finell, J.; Pastinen, O. *Biotechnol. Bioeng.* **2001**, 72, 501-505.

(163) Jokela, J.; Pastinen, O.; Leisola, M. *Enzyme Microb. Technol.* **2002**, 31, 67-76.

(164) Asboth, B.; Naray-Szabo, G. *Curr. Protein. Pept. Sci.* **2000**, 1, 237-254.

(165) Visuri, K. J. *US Patent 4699882* **1987**.

(166) Visuri, K. J. *US Patent 5437993* **1995**.

(167) Hanson, B. L.; Langan, P.; Katz, A. K.; Li, X. M.; Harp, J. M.; Glusker, J. P.; Schoenborn, B. P.; Bunick, G. J. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2004**, 60, 241-249.

(168) Carrell, H. L.; Hoier, H.; Glusker, J. P. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **1994**, 50, 113-123.

(169) Koplik, J.; Banavar, J. R.; Willemsen, J. F. *Phys. Rev. Lett.* **1988**, 60, 1282-1285.

(170) Rafferty, J. L.; Zhang, L.; Siepmann, J. I.; Schure, M. R. *Anal. Chem.* **2007**, 79, 6551-6558.

(171) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, 17, 192-198.

(172) Vvedensky, D. D. *J. Phys.-Condens. Matter* **2004**, 16, R1537-R1576.

(173) McConnell, O.; Bach, A.; Balibar, C.; Byrne, N.; Cai, Y. X.; Carter, G.; Chlenov, M.; Di, L.; Fan, K.; Goljer, I.; He, Y. N.; Herold, D.; Kagan, M.; Kerns, E.; Koehn, F.; Kraml, C.; Marathias, V.; Marquez, B.; McDonald, L.; Nogle, L.; Petucci, C.; Schlingmann, G.; Tawa, G.; Tischler, M.; Williamson, R. T.; Sutherland, A.; Watts, W.; Young, M.; Zhang, M. Y.; Zhang, Y. R.; Zhou, D. H.; Ho, D. *Chirality* **2007**, 19, 658-682.

(174) Lipkowitz, K. B.; Coner, R.; Peterson, M. A. *J. Am. Chem. Soc.* **1997**, 119, 11269-11276.

(175) Lipkowitz, K. B. *Acc. Chem. Res.* **2000**, 33, 555-562.

(176) Zhao, C. F.; Cann, N. M. *Anal. Chem.* **2008**, 80, 2426-2438.

(177) Szabelski, P.; Sholl, D. S. *J. Chem. Phys.* **2007**, 126, 144709.

(178) Szabelski, P. *App. Surf. Sci.* **2004**, 227, 94-103.

(179) Kasat, R. B.; Wang, N. H. L.; Franses, E. I. *J. Chromatogr. A* **2008**, 1190, 110-119.

(180) Kasat, R. B.; Wee, S. Y.; Loh, J. X.; Wang, N. H. L.; Franses, E. I. *J. Chromatogr. B* **2008**, 875, 81-92.

(181) Malek, K.; Coppens, M. O. *J. Phys. Chem. B* **2008**, 112, 1549-1554.

(182) Power, T. D.; Skoulidas, A. I.; Sholl, D. S. *J. Am. Chem. Soc.* **2002**, 124, 1858-1859.

(183) Wang, S.; Canna, N. M. *J. Chem. Phys.* **2009**, 130.

(184) Yang, W.; Nymeyer, H.; Zhou, H. X.; Berg, B.; Bruschweiler, R. *J. Comput. Chem.* **2007**, 29, 668-672.

(185) MacKerell, A. D. *J. Comput. Chem.* **2004**, 25, 1584-1604.

(186) Kaminski, G. A.; Friensner, R. A.; TiradoRives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, 105, 6474-6487.

(187) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, 21, 1049-1074.

(188) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, 79, 926-935.

(189) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, 112, 8910-8922.

(190) Nutt, D. R.; Smith, J. C. *J. Chem. Theo. Comput.* **2007**, 3, 1550-1560.

(191) Hess, B.; Holm, C.; van der Vegt, N. *J. Chem. Phys.* **2006**, 124, 164509.

(192) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. USA* **2007**, 104, 796-801.

(193) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, 110, 17616-17626.

(194) Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, 119, 11342-11349.

(195) Bashford, D.; Karplus, M. *Biochemistry* **1990**, 29, 10219-10225.

(196) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. *Biophys. J.* **2006**, 90, L36-L38.

(197) Meinhold, L.; Smith, J. C. *Biophys. J.* **2005**, 88, 2554-2563.

(198) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2001**, 114, 363-366.

(199) Rey-Castro, C.; Tormo, A. L.; Vega, L. F. *Fluid Phase Equilib.* **2007**, 256, 62-69.

(200) Speight, J. G., Ed. Lange's Handbook of Chemistry; McGraw-Hill: New York, **2004**.

(201) Auffinger, P.; Hashem, Y. *Curr. Opinion. Struct. Biol.* **2007**, 17, 325-333.

(202) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theo. Comput.* **2008**, 4, 819-834.

(203) Wine, Y.; Cohen-Hadar, N.; Freeman, A.; Frolow, F. *Biotechnol. Bioeng.* **2007**, 98, 711-718.

Publications

1. Hu Z. Q., Jiang J. W. Assessment of Biomolecular Force Fields for Molecular Dynamics Simulations in a Protein Crystal. *J. Comput. Chem.* 2010, 31, 371-380.

2. Hu Z. Q., Jiang J. W. Chiral Separation of Racemic Phenylglycines in Thermolysin Crystal: A Molecular Simulation Study, *J. Phys. Chem. B* 2009, 113, 15851-15857.

3. Hu Z. Q., Jiang J. W. Separation of Amino Acids in Glucose Isomerase Crystal: Insight from Molecular Dynamics Simulations. *J. Chromatogr. A* 2009, 1216, 5122-5129.

4. Hu Z. Q., Jiang J. W. Electrophoresis in Protein Crystal: Non-Equilibrium Molecular Dynamics Simulations. *Biophys. J.* 2008, 95, 4148-4156.

5. Hu Z. Q., Jiang J. W. Molecular Dynamics Simulations for Water and Ions in Protein Crystals. *Langmuir* 2008, 24, 4215-4223.

6. Hu Z. Q., Jiang J. W. Mechanistic Insight into the Biological Nanopore in Tetragonal Lysozyme Crystal. *J. Membr. Sci*. 2008, 324, 192-197.

7. Hu Z. Q., Jiang J. W. Comment on 'Diffusion of Water and Sodium Counter-ions in Nanopores of a β-lactoglobulin Crystal: A Molecular Dynamics Study.' *Nanotechnology* 2008, 19, 438001.

8. Hu Z. Q., Jiang J. W., Sandler S. I. Water in Hydrated Orthorhombic Lysozyme Crystal: Insight from Atomistic Simulations. *J. Chem. Phys*. 2008, 129, 075105.

9. Hu Z. Q., Jiang J. W., Rajagopalan R. Effects of Macromolecular Crowding on Biochemical Reaction Equilibria: A Molecular Thermodynamic Perspective. *Biophys. J*. 2007, 93, 1464-1473.

10. Gnanasambandam S., Hu Z. Q., Jiang J. W., Rajagopalan R. Force Field for Molecular Dynamics Studies of Glycine/Water Mixtures in Crystal/Solution Environments. *J. Phys. Chem. B* 2009, 113, 752-758.

11. Nalaparaju A., Hu Z. Q., Zhao X. S., Jiang J. W. Exchange of Heavy Metal Ions in Titanosilicate Na-ETS-10 Membrane from Molecular Dynamics Simulations. *J. Membr. Sci*. 2009, 335, 89-95.

12. Babarao R., Hu Z. Q., Jiang J. W., Chempathy S., Sandler S. I. Storage and Separation of CO_2 and CH_4 in Silicalite, C_{168} Schwarzite and IRMOF-1: A Comparative Study from Monte Carlo Simulation. *Langmuir* 2007, 23, 659-666.

Presentations

1. Hu Z. Q., Jiang J. W. "Molecular dynamics simulations in protein crystals", *4th Asian Pacific Conference of Theoretical and Computational Chemistry*, Port Dickson, Malaysia (Dec. 2009).

2. Hu Z. Q., Jiang J. W. "Biomolecular force fields for lysozyme crystal: Assessed from molecular dynamics simulations", *AIChE Annual Conference*, Nashville, Tennessee, USA (Nov. 2009).

3. Hu Z. Q., Jiang J. W. Separation of amino acid mixture in glucose isomerase crystal: A non-equilibrium moleccular dynamics simulation. *Collaborative Computational Project for Biomolecular Simulation Conference*, York, UK (Jan. 2009).

4. Hu Z. Q., Jiang J. W. "Separation of amino acid mixtures in glucose isomerase crystal: A computational study", *1st Biological Physics International Conference*, National University of Singapore, Singapore (Dec. 2008).

5. Hu Z. Q., Jiang J. W. A molecular thermodynamic model for crowding effects on reactions. *AIChE Annual Meeting*, Philadelphia, USA (Nov. 2008).

6. Hu Z. Q., Jiang J. W. Diffusion of water, ions and amino acids in protein crystals: A molecular simulation study. *AIChE Annual Meeting*, Philadelphia, USA (Nov. 2008).

7. Hu Z. Q., Jiang J. W. "Electrophoresis in protein crystal: molecular dynamics simulation", *International Congress on Membranes and Membrane Processes*, Hawaii, USA (Jul. 2008).

8. Hu Z. Q., Jiang J. W. Nanoconfined Fluids in Protein Crystals: A Computational Perspective. *International Conference on Materials for Advanced Technologies*, Singapore (Jul. 2007).

9. Hu Z. Q., Jiang J. W. "Mechanistic insight into protein crystals from atomistic simulations", *11th International Conference on Properties and Phase Equilibria for Product and Process Design*, Greece (May 2007).

Appendix A

Copy right permission of Figure 1 in chapter 1 from JACS

AMERICAN CHEMICAL SOCIETY LICENSE TERMS AND CONDITIONS

May 27, 2010

This is a License Agreement between zhongqiao hu ("You") and American Chemical Society ("American Chemical Society") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Chemical Society, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2437411352136
License Date	May 27, 2010
Licensed content publisher	American Chemical Society
Licensed content publication	Journal of the American Chemical Society
Licensed content title	Protein Crystals as Novel Microporous Materials
Licensed content author	Lev Z. Vilenchik et al.
Licensed content date	May 1, 1998
Volume number	120
Issue number	18
Type of Use	Thesis/Dissertation
Requestor type11	Not specified
Format	Print and Electronic
Portion	Table/Figure/Micrograph
Number of Table/Figure/Micrographs	1
Author of this ACS article	No
Order reference number	
Title of the thesis / dissertation	Molecular Simulations of Transport and Separation in Protein Crystals
Expected completion date	Jun 2010
Estimated size(pages)	140
Billing Type	Invoice

Billing Address	Dept of Chem. Biomol. Engi. 4 Engineering Drive 4 Singapore, other 117576 Singapore
Customer reference info	
Total	0.00 USD
Terms and Conditions	

Thesis/Dissertation

ACS / RIGHTSLINK TERMS & CONDITIONS
THESIS/DISSERTATION

INTRODUCTION
The publisher for this copyrighted material is the American Chemical Society. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

LIMITED LICENSE
Publisher hereby grants to you a non-exclusive license to use this material. Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process; any form of republication must be completed within 60 days from the date hereof (although copies prepared before then may be distributed thereafter).

GEOGRAPHIC RIGHTS: SCOPE
Licenses may be exercised anywhere in the world.

RESERVATION OF RIGHTS
Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

PORTION RIGHTS STATEMENT: DISCLAIMER
If you seek to reuse a portion from an ACS publication, it is your responsibility to examine each portion as published to determine whether a credit to, or copyright notice of, a third party owner was published adjacent to the item. You may only obtain permission via Rightslink to use material owned by ACS. Permission to use any material published in an ACS publication, journal, or article which is reprinted with permission of a third party must be obtained from the third party owner. ACS disclaims any responsibility for any use you make of items owned by third parties without their permission.

REVOCATION
The American Chemical Society reserves the right to revoke a license for any reason, including but not limited to advertising and promotional uses of ACS content, third party usage, and incorrect figure source attribution.

LICENSE CONTINGENT ON PAYMENT
While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

COPYRIGHT NOTICE: DISCLAIMER
You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Reprinted (“Adapted” or “in part”) with permission from REFERENCE CITATION. Copyright YEAR American Chemical Society."

WARRANTIES: NONE
Publisher makes no representations or warranties with respect to the licensed material.

INDEMNITY
You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

NO TRANSFER OF LICENSE
This license is personal to you or your publisher and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

NO AMENDMENT EXCEPT IN WRITING
This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

OBJECTION TO CONTRARY TERMS
Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions

(which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

JURISDICTION
This license transaction shall be governed by and construed in accordance with the laws of the District of Columbia. You hereby agree to submit to the jurisdiction of the courts located in the District of Columbia for purposes of resolving any disputes that may arise in connection with this licensing transaction.

THESES/DISSERTATION TERMS
Publishing implications of electronic publication of theses and dissertation material
Students and their mentors should be aware that posting of theses and dissertation material on the Web prior to submission of material from that thesis or dissertation to an ACS journal may affect publication in that journal. Whether Web posting is considered prior publication may be evaluated on a case-by-case basis by the journal's editor. If an ACS journal editor considers Web posting to be "prior publication", the paper will not be accepted for publication in that journal. If you intend to submit your unpublished paper to ACS for publication, check with the appropriate editor prior to posting your manuscript electronically.

If your paper has already been published by ACS and you want to include the text or portions of the text in your thesis/dissertation in **print or microfilm formats**, please print the ACS copyright credit line on the first page of your article: "Reproduced (or 'Reproduced in part') with permission from [FULL REFERENCE CITATION.] Copyright [YEAR] American Chemical Society." Include appropriate information.

Submission to a Dissertation Distributor: If you plan to submit your thesis to UMI or to another dissertation distributor, you should not include the unpublished ACS paper in your thesis if the thesis will be disseminated electronically, until ACS has published your paper. After publication of the paper by ACS, you may release the entire thesis (**not the individual ACS article by itself**) for electronic dissemination through the distributor; ACS's copyright credit line should be printed on the first page of the ACS paper.

Use on an Intranet: The inclusion of your ACS unpublished or published manuscript is permitted in your thesis in print and microfilm formats. If ACS has published your paper you may include the manuscript in your thesis on an intranet that is not publicly available. Your ACS article cannot be posted electronically on a publicly available medium (i.e. one that is not password protected), such as but not limited to, electronic archives, Internet, library server, etc. The only material from your paper that can be posted on a public electronic medium is the article abstract, figures, and tables, and you may link to the article's DOI or post the article's author-directed URL link provided by ACS. This paragraph does not pertain to the dissertation distributor paragraph above.

Other conditions:

v1.1

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK10791571.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2437411352136 and noting the reason for cancellation.

Questions? customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.