Study of integral equation theory and human beta defensin 28

Chen, Wenwen

2012

Chen, W. (2012). Study of integral equation theory and human beta defensin 28. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/47998>

<https://doi.org/10.32657/10356/47998>

**STUDY OF INTEGRAL EQUATION
THEORY AND HUMAN BETA
DEFENSIN 28**

CHEN WENWEN

School of Chemical and Biomedical Engineering

A thesis submitted to Nanyang Technological University

in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

2012

ACKNOWLEDGEMENT

It is a great pleasure to extend my most sincere acknowledgement and gratitude to a number of people whose help, support and suggestions were immensely valuable in my research.

First and foremost, I gratefully acknowledge my supervisor Dr. Kwak Sang Kyu for the opportunity to work in his group and for his guidance, encouragement and fully support throughout my study. I could not have completed the project without absorbing his enthusiasm in molecular simulations. He has opened me the door of a world with so much joy and imagination.

I appreciate my colleagues in Molecular Simulation Group of SCBE for all their help during my research.

I am also grateful to the Nanyang Technological University for certain research scholarship and the computational resources provided by the School of Chemical and Biomedical Engineering.

Equally important, I would like to thank my family. They always encourage me my pursuits and have been providing love and support all through my life.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	I
TABLE OF CONTENTS	II
LIST OF ACRONYMS	VIII
LIST OF FIGURES	X
LIST OF TABLES	XVI
ABSTRACT	XVIII
PART I: INTEGRAL EQUATION THEORY	1
CHAPTER 1 INTRODUCTION	2
1.1 RADIAL DISTRIBUTION FUNCTION	6
1.2 INTEGRAL EQUATION THEORY	8
1.2.1 Classical Closures	9
1.2.2 Truncated series of the bridge diagrams.....	10
1.2.3 Extension to Mixtures.....	11
1.3 NUMERICAL SOLUTIONS OF INTEGRAL EQUATIONS.....	12
1.3.1 Picard's Method	12
1.3.2 Newton-Raphson Method.....	13

1.3.3 Combination Method	14
1.4 DIAGRAMMATIC EXPRESSION	15
1.4.1 Diagrammatic Technique	15
1.4.2 Cluster Expansion	16
1.4.3 Bridge Diagrams.....	18
CHAPTER 2 GENERAL METHODOLOGY	21
2.1 MONTE CARLO SIMULATION	21
2.1.1 Monte Carlo	22
2.1.2 Metropolis Monte Carlo	24
2.1.3 Transition Matrix Monte Carlo.....	25
2.1.4 Umbrella Sampling.....	27
2.1.5 Grand Canonical Ensemble Monte Carlo.....	28
2.2 BRIDGE DIAGRAM CALCULATION	29
2.2.1 Legendre Polynomial Expansion.....	30
2.2.2 Biased Monte Carlo Method	31
2.2.3 Free-Energy Perturbation Method	31
2.3 SIMULATION MODELS	34

2.3.1 <i>Hard Sphere Model</i>	34
2.3.2 <i>Hard Sphere-Wall Model</i>	35
2.4 SIMULATION ENVIRONMENT	37
CHAPTER 3 NEW METHODS OF BRIDGE FUNCTION CALCULATION	39
3.1 INTRODUCTION	39
3.2 SIMULATION METHODOLOGY	41
3.2.1 <i>TMMC Mayer-Sampling</i>	41
3.2.2 <i>Overlap Sampling</i>	43
3.3 SIMULATION MODELS AND DETAILS	45
3.4 RESULTS AND DISCUSSION	46
3.5 CONCLUSION	58
CHAPTER 4 SINGLET IE THEORY OF THE HARD SPHERE FLUID CONFINED IN A SLIT-PORE .	60
4.1 INTRODUCTION	60
4.2 THEORETICAL APPROACH	62
4.2.1 <i>Singlet Integral Equation Theory</i>	62
4.2.2 <i>Slit-Fluid Bridge Function</i>	63
4.2.3 <i>TMMC Mayer-Sampling</i>	65

4.3 SIMULATION MODELS AND DETAILS	66
4.4 RESULTS AND DISCUSSION.....	68
4.5 CONCLUSION.....	73
PART II: HUMAN BETA DEFENSIN 28	76
CHAPTER 5 BACKGROUND	77
5.1 INTRODUCTION.....	77
5.2 HUMAN β -DEFENSINS	80
5.2.1 Amino Acids and Peptides	80
5.2.2 Characterization.....	82
5.2.3 Antimicrobial activities.....	83
5.2.4 Structures	84
5.2.5 Structure-Function Relationship.....	87
5.3 HOMOLOGY MODELING	87
5.3.1 Template Recognition and Sequence Alignment	89
5.3.2 Model Building, Refining and Evaluation	90
5.4 MOLECULAR MECHANICS	91
5.4.1 Force Fields.....	92

5.4.2 Energy Minimization	93
5.4.3 Molecular Dynamic Simulation	93
CHAPTER 6 STRUCTURE OF HUMAN β-DEFENSIN 28	95
6.1 INTRODUCTION	95
6.2 SIMULATION AND EXPERIMENTAL DETAILS	96
6.2.1 Homology Modeling	96
6.2.2 Molecular Dynamics Simulation	97
6.2.3 Simulation Data Analysis	99
6.2.4 Circular Dichrorism Spectra	100
6.3 RESULTS AND DISCUSSION	101
6.3.1 Sequence Alignment	101
6.3.2 Initial HBD-28 Models	103
6.3.3 Stability of the Initial HBD-28 Models	106
6.3.4 Final HBD-28 Models	111
6.3.5 Comparison with Circular Dichroism Spectra	114
6.4 CONCLUSION	117
CHAPTER 7 CONCLUSIONS	119

APENDIX A	122
APPENDIX B.....	125
APPENDIX C.....	128
REFERENCES	130
PUBLICATION LIST	148

LIST OF ACRONYMS

IE: Integral Equation

TMMC: Transition Matrix Monte Carlo

RDF: Radial Distribution Function

OZ: Ornstein-Zernike

HNC: Hypernetted Chain

PY: Percus–Yevick

NR: Newton-Raphson

MC: Monte Carlo

GCMC: Grand Canonical Monte Carlo

API: Application Program Interface

IDE: Integrated Development Environment

AMP: Antimicrobial Peptide

HBD: Human Beta Defensin

3D: Three-Dimensional

CD: Circular Dichroism

MD: Molecular Dynamics

EM: Energy Minimization

TFE: Trifluoroethanol

PDB: Protein Data Bank

PSVS: Protein Structure Validation Software Suite

Gromacs: Groningen Machine for Chemical Simulation

SD: Steepest Descent

DS: Discovery Studio

DSSP: The Dictionary of Secondary Structure of Proteins

RMSD: Root Mean Square Deviation

RMSF: Root Mean Square Fluctuation

Tris-HCl: Tris(hydroxymethyl)aminomethane-HCl

LIST OF FIGURES

- Figure 1.1. Virial coefficients B_2 and B_3 in the f -bond expansion. The open circle represents one root particle (labeled by 1) and the solid circles (labeled by 2 and 3) denote field points, which are subjected to be integrated over the connecting f -bond(s). 16
- Figure 1.2. Cavity distribution function coefficients y_1 and y_2 in the f -bond density expansion. The open circles represent two root particles and the solid circles denote field points. The first row shows the single diagram in y_1 and the second row shows 5 diagrams in y_2 , respectively. The last row represents the diagrammatic expansion y . The number of solid circles is related to the number of powers of density and thus the order of each term. 17
- Figure 1.3. Diagrammatic expansions of direct correlation function $c(r)$ and correlation function $h(r)$. Circles and bonds are the same as in 1.2. 18
- Figure 1.4. Bridge diagrams b_2 and b_3 in the f -bond density expansion of bridge function. The open circles represent two root particles (labeled by 1 and 2, for clarity, these labels are omitted in the other diagrams) and the solid circles denote field points to be integrated over the connecting f -bonds. The first row represents b_2 and its integral expression. The second row shows 13 diagrams in b_3 and the value of each coefficient is the sum of all diagrams with their weights, respectively. Number below diagrams denotes their symmetry

numbers which is the reciprocal value of the weight.....	19
Figure 1.5. Bridge diagrams b_2 and b_3 in the h -bond density expansion of bridge function. Circles are the same as in Figure 1.4, but the connecting lines signify the total correlation function $h(r)$. The first row represents b_2 and its integral expression. The second row shows 7 diagrams in b_3 and the value of each coefficient are calculated in same way as that of f -bond.	20
Figure 2.1. Excess free energy in the f -bond density expansion. The solid circles denote field points and integrations. The number of circles is to the number of powers of density. The coefficients before the diagrams are omitted for clarity, which is $1/2$ for 1 st , $1/6$ for 2 nd , $1/8$ for 3 rd , $1/4$ for 4 th and $1/24$ for 5 th	32
Figure 2.2. The schematics of the potentials for hard sphere model (left) and sphere-wall model (right).....	36
Figure 2.3 Screenshot of the simulation environment based on eclipse java IDE	37
Figure 3.1. The phase-space schematic of appropriate intermediate system (labeled C) for the bridge coefficient calculation constructed between target system A and reference system B for (a) umbrella sampling; (b) overlap sampling. Arrows from one system to indicate the sampling direction.	41
Figure 3.2. Development of the weight function for $b_3(r)$ at density 0.8 in the TMMC Mayer-sampling. The different symbols represent estimates	

of the weight function after different numbers of MC steps; circle, triangle and square represent the weight function after 20, 200, 2000 and 200000×10^4 MC steps, respectively. The upper-left plot indicates the typical configuration distributions for the MC Mayer-sampling and the TMMC Mayer-sampling, labeled by square and circle symbols, respectively.50

Figure 3.3. Configuration distributions of reference and target systems in sampling for $b_3(r)$. The curves are normalized from 2×10^7 configurations over 8σ with the bin width as 0.1σ52

Figure 3.4. Plot of the bridge coefficient $b_3(r)$ vs. reduced moderate distance. Solid line is the reference data from the long MC Mayer-sampling simulation, and dash, dot, and dot-dash lines represent the results calculated from the short MC Mayer-sampling, the TMMC Mayer-sampling, and the overlap sampling, respectively. The reference data was evaluated from 5×10^{10} configurations at density 0.2 and 10^{11} configurations at densities 0.5 and 0.8.53

Figure 3.5. Plot of bridge coefficient $b_3(r)$ vs. reduced short distance. The same notations as in Figure 3.4 are applied. The lower-right plot shows $b_3(r)$'s within range 0.01σ , note that values at density 0.2 are added by 2.0 for clarity.54

Figure 3.6. Plot of bridge coefficient $b_3(r)$ vs. reduced long distance. The same notations as in Figure 3.4 are applied.55

Figure 3.7. Plot of pressure versus reduced density. Solid line is the equation of

state of Carnahan and Starling. Subscript v stands for ‘virial’ pressure. Dashed line are from pure HNC theory, and square, triangle and cycle symbols are from solution of the OZ equation with a bridge function closure obtained from the h -bond cluster expansion to the second order in density by the MC Mayer-the TMMC Mayer-sampling and the overlap sampling, Note here, the lines from reference [29] are guides to an eye and triangle and cycle symbols at densities 0.2, 0.5 and 0.8 are from $h(r)$ via the TMMC Mayer-sampling and the overlap sampling, respectively by collecting $5\sim 10\times 10^9$ configurations.....57

Figure 4.1. Schematic of the wall-particle bridge diagram in the second order slit-fluid bridge coefficient. The open circles represent two root particles (labeled by 1 and 2) and the solid circles denote the field points (labeled by 3 and 4), which are subjected to be integrated the connecting h -bonds.64

Figure 4.2. Plots of two slit-fluid bridge coefficients $b_2(z)$, $b_3(z)$ and a bridge function $b_s(z)$ versus the reduced distance from the wall surface for $L=3.0\sigma$ and $L=4.0\sigma$, respectively. The solid, dash and dash-dot represent the values at density 0.3, 0.5 and 0.7, respectively.69

Figure 4.3. Density profiles of hard sphere fluids within $L=3.0\sigma$. The slit pore fluid is in equilibrium with a bulk fluid of the reduced density at 0.5 and 0.7, respectively. The dot, dash-dot and solid lines results of the GCMC simulation, the HNC closure and the HNC closure with $b_s(z)= b_2(z)\rho_b^2+b_3(z)\rho_b^3$, respectively.71

Figure 4.4. Density profiles of hard sphere fluids within $L=4.0\sigma$. The symbols the same as shown in Figure 4.3.	72
Figure 5.1. A generic structure of a peptide, consisting of three amino acids and two peptide bonds.	80
Figure 5.2. Structural comparison of human β -defensins. (a) The amino acid sequence alignment of HBD-1, HBD-2 and HBD-3. The extent of the secondary structure elements is shown above each sequence and the numbers in the brackets represent charges. Six conserved cysteine residues are colored in yellow. (b) The ribbon representations of monomeric β -defensins (left) are accompanied by the electrostatic potential maps projected on the solvent-solvent-accessible surfaces shown in the equivalent orientations (center) and after 180° rotation (right). The positively and negatively charged areas of the solvent-accessible surfaces are colored in blue and red, respectively. [107].....	85
Figure 5.3. The limiting steps in homology modeling as function of percentage sequence identity between the structure and the model [127].	89
Figure 6.1. Average RMSD of all $C\alpha$ atoms of the HBD-28 models relative to the averaged structure as a function of residue number.....	104
Figure 6.2. Verify-3D score of the HBD-28 models as a function of residue number. The solid, dot, dash-dot-dot and dash lines represent the values for HBD-28_A, HBD-28_B, HBD-28_C and HBD-28_D, respectively.	105

Figure 6.3. Secondary structure of each residue (on y-axis) as a function of time (x-axis) of four models in pure water simulation under NPT condition. Here, different color represents different secondary structure types, red color denotes β -strand, blue and violet colors denote helical conformation. Secondary structures are obtained the DSSP program.....	108
Figure 6.4. The RMSD value of the HBD-28 models as a function of time. The black and red lines represent the values for HBD-28_A and HBD-HBD-28_D, respectively.....	110
Figure 6.5. The RMSF value of the HBD-28 models as a function of residue number. Other symbols are the same as in Figure 6.4.	110
Figure 6.6. Secondary structure of each residue (on y-axis) as a function of time (x-axis) of four models in pure water simulation under NVT condition. Other symbols are the same as in Figure 6.3.	111
Figure 6.7. Comparison of the 3D structures between HBD-28_A and HBD-HBD-28_D after NVT MD simulation. The α -helix is colored in β -strands in yellow and others in blue.	112
Figure 6.8. Structural comparison of HBD-28_A (top) and HBD-28_D (bottom) by surfaces shown in the front view (left) and back view (right). α -helix is colored in red, β -strands in yellow and others in white.	113

LIST OF TABLES

Table 3.1. Standard deviations of short MC Mayer-sampling, TMMC Mayer-sampling and overlap sampling simulations against long Mayer-sampling simulation for $b_2(r)$ at densities 0.2, 0.5, and 0.8, respectively. Note that the reference values (long MC Mayer-sampling) were evaluated from 5×10^{10} configurations for $b_2(r)$ at all densities.	47
---	----

Table 3.2. Standard deviations of short MC Mayer-sampling, TMMC Mayer-sampling and overlap sampling simulations against long Mayer-sampling simulation for $b_3(r)$ at densities 0.2, 0.5, and 0.8, respectively. Note that the reference values (long MC Mayer-sampling) were evaluated from 5×10^{10} configurations for $b_3(r)$ at density 0.2 and 10^{11} configurations at densities 0.5 and 0.8. The notations are the same as in Table 3.1.	48
---	----

Table 4.1. The reduced pressures normal to the wall (p_w^*) at slit widths 3.0σ and 4.0σ for reduced bulk densities $\rho_b\sigma^3=0.3, 0.5$ and 0.7	73
--	----

Table 5.1. Amino acid full names, three-letter abbreviations (in bold type), and one-letter abbreviations (between brackets).	81
--	----

Table 5.2. Antimicrobial spectrum of HBDs 1-4. [106, 119-121]	83
---	----

Table 6.1. The sequence alignments of HBD-28 with different templates. Here HBD-28 is aligned with HBD-2, HBD-3 and their combinations. Conserved residues are highlighted by asterisks at the bottom of	
--	--

each alignment and <i>A</i> , <i>B</i> , and <i>C</i> are determined using CLUSTALW tool.	102
Table 6.2. Comparison of the secondary structure elements in different HBD-28 models predicted by DS, EM and MD. Here numbers represent the residue numbers in each HBD.	103
Table 6.3. The molar percentage of secondary structure configurations of HBD-HBD-28 by experiment and MD simulation in pure water. Here antiparallel β -strand stands for β_2 and parallel β -strand contains β_1 and β_2	115
Table 6.4. The molar percentage of secondary structure configurations of HBD-HBD-28 by experiment and by MD simulation in the present of TFE. Same notations as shown in Table 6.3.	116

ABSTRACT

This thesis addresses two different topics. The major part of this thesis deals with the precise evaluation of bridge functions required to yield the integral equation theories up to the second power in density for both bulk and confined hard sphere fluids. We propose two modified Mayer-sampling methods based on Transition Matrix Monte Carlo (TMMC) and overlap sampling for calculating the integrated diagrams appearing in the coefficients of the bridge function. The results from these methods compared with the generic Monte Carlo Mayer-sampling are analyzed in detail for the bulk hard sphere system. Next TMMC Mayer-sampling method is applied to evaluate the bridge function for the hard sphere fluid in a slit-pore. We first construct the slit-fluid bridge function by wall-particle bridge diagrams with h_b -bond (i.e. bulk-fluid total correlation function h_b). The quantity of the slit-fluid bridge function is assessed by comparing density profiles and reduced normal pressures obtained from the singlet integral equation theory with those from the grand canonical Monte Carlo simulation.

The second part of this thesis concerns the 3D structure of human beta defensin (HBD) 28 via homology modeling and molecular dynamics approach. Implementations of these methods are described, as well as the comparison among the resulting models that based on HBD-2 and HBD-3 as templates. The highly probable candidates for HBD-28 are refined through molecular simulations in pure water and 50% trifluoroethanol/water mixture. In comparison with circular dichroism experimental data as well as stabilities of β -sheet and α -helix, HBD-28 with a reliable 3D structure was found.

Part I: Integral Equation Theory

Chapter 1

Introduction

The determination of structural and thermodynamic properties of equilibrium classical fluids is of central importance to the field of statistical mechanics [1]. One of the most important properties in the structural description of equilibrium fluids is the pair correlation function [1, 2]. It gives the probability of the spatial arrangements of pair particles and has been used to calculate various thermodynamic properties, such as pressure, compressibility, and internal energy. Pair correlation function is usually calculated by computer simulations [3-5] or integral equation (IE) theory [6-8]. The latter has been more widely used since it requires much less computational cost than direct simulation methods like Monte Carlo or molecular dynamics, and in some cases it can yield completely to analytic treatment [9, 10]. The IE approach that yields the pair correlation function of classical fluids from the pair potentials is based on the Ornstein-Zernike equation supplemented by a closure that establishes the bridge function [11]. Usually, the bridge function in the IE theory is either ignored completely or approximated by other expressions, which leads to deviations with respect to the ‘exact’ simulation results [12-16]. In addition, it also leads to certain thermodynamic inconsistent, for instance, the pressure calculated by the virial expansion differs from the pressure via the compressibility integration [17]. All these deficiencies in the implementations of the IE theory arise from the inaccuracy of bridge function approximations. It is therefore of fundamental importance to construct accurate bridge function in the improvement of the IE theory.

Typically, there are two main routes to obtain the bridge function. The first route is related to the IE theory framework, which evaluates the bridge function through empirical approximations or measurements. The former method mostly approximates the bridge function of interesting with that of a simpler reference system [18, 19]. Whereas in the latter approach, the bridge function is derived from the accurate pair correlation function that is obtained from molecular simulations [20]. The resulting shape of bridge function undergoes fitting procedure to yield an empirical expression with adjustable parameters [21-23]. The parameterization provides fast and reliable ways to construct bridge function for some model systems of interest. This approach, however, shows some inaccuracy with respect to the signs of bridge function [24] and zero-zero-separation theorems [25]. Furthermore, the true characteristic of the bridge function is still on debate. In contrast to these empirical approaches, the second route is the direct evaluation of the bridge function by its definition that is expressed as an infinite sum of highly connected bridge diagrams [24, 26-29]. This theoretical approach should lead to a more accurate bridge function, correspondingly, a more accurate IE theory. However, the evaluation of series of irreducible diagrams forming the bridge function is a computationally demanding task, because the diagram number increases rapidly with increasing order of the diagrams and each diagram represents a high-dimensional integral over products of pair functions. Since the high computational requirement restricts the direction evaluation of bridge function with certain accuracy, the goal of this work is to develop efficient and stable methods for precise evaluation of bridge function up to higher order of the diagrams and as a consequence to improve the IE theory.

To date, one of the successful methods to compute the high-dimensional integrals of bridge function is the biased Monte Carlo sampling, which is proposed by Rast *et al.* [24] and improved by Labik *et al.* [27]. Another successful approach is a free energy perturbation method, so-called Mayer-sampling, which is first introduced to evaluate accurate virial coefficients represented as cluster diagrams by Singh and Kofke [28] and later applied to calculate the bridge diagrams by Kwak and Kofke [29]. However, the bridge function calculated from these approaches converges very slowly. In addition, the accurate prediction has been hindered where the relative distance of two particles are short and long, respectively, as the density and the order of the diagrams increase. Moreover, it is noticed that the IE theory generically suffers from prediction of accurate pair correlation function at high density close to freezing, but it is possible to obtain better pair correlation function by improving the sampling technique for bridge diagram configurations. In our work, we introduce two sampling techniques based on the original Mayer-sampling. One is the Mayer-sampling incorporated with Transition Matrix Monte Carlo [30] (i.e. TMMC Mayer-sampling), and another is the overlap sampling [31, 32]. The feasibility and efficiency of these two approaches for generating approximation of bridge function are both tested for the hard sphere fluid, and the results calculated by these approaches are compared with those deduced from Mayer-sampling and empirical expressions. We also show how these two approaches influence the properties of the hard sphere fluid in the closure of Ornstein-Zernike equation.

So far, most studies of the IE theory for liquids are related to the bulk fluid, limited effort has been devoted to the confined fluid. It has been poorly

understood even for the simplest model of a hard sphere fluid confined in a slit-slit-pore. Since the confined fluid plays an important role in scientific phenomena and technological applications, it is of essential importance to study the confined fluid. Until now, two types of IE theories have been developed for the confined fluid: the singlet level [33-36] and pair level theories [37-40]. The latter is rather complex and has no advantage over the singlet one in most cases [35, 39], thus we will focus on the singlet IE theory in this thesis. However, like most of the IE theories for the bulk fluid, the IE theories for the confined fluid also suffer from the uncertainty of the bridge function. The bridge function involving the fluid confinement is trivial to be formulated or approximated empirically. Therefore an alternative route to evaluate the confined-fluid bridge function is through the application of the bridge diagrams. The question arises the construction theorem of the bridge diagrams for the confined fluid is the same as that for the bulk fluid and how to apply them correctly in the confined fluid. Therefore, another goal of this work is to study the accuracy and applicability of the confined-fluid bridge function for the singlet IE theory to predict the structural and thermodynamic properties of the hard sphere fluid in a slit-pore in equilibrium with its corresponding bulk fluid. In this work, the recently developed TMMC Mayer-sampling method is used to evaluate the confined-fluid bridge function, and with the help of this bridge function the confinement effects on the local structure of the confined fluid is studied by the singlet IE approach. The quality of our approach is assessed by comparing the results of this method with those from the grand canonical Monte Carlo simulation.

The first part of this thesis is arranged as follows. In the next section of this

chapter, we give a brief review of the radial distribution function and its related thermodynamic properties, the IE theories and their numerical solutions, as well as the diagrammatic expansions of pair functions, especially, the bridge function. Chapter 2 describes the general simulation methods: Monte Carlo methods and bridge diagram calculation methods, as well as the models we in this work. In Chapter 3, two new bridge function calculation methods are introduced, i.e. TMMC Mayer-sampling and overlap sampling methods. The applications of these two methods are described in Chapter 3 and Chapter 4. Chapter 3 investigates the feasibility of TMMC Mayer-sampling, overlap samplings and their implementations on the evaluation of the bridge function the bulk hard sphere fluid. The precision and convergence of these two newly developed methods were compared with those of the original sampling method small distance range, moderate distance range and tails region of interaction range of the pair particles, respectively. The pressures determined by the IE theory with our new approximations of bridge function are compared with those from original sampling method with reduced densities at 0.2, 0.5 and 0.8. In Chapter 4 we use the TMMC Mayer-sampling to determine the slit-fluid bridge functions of a hard sphere fluid in slit-pores with slit widths 3.0σ and 4.0σ , where the structural and thermodynamic properties such as density profiles and contact pressures perpendicularly to the wall are investigated.

1.1 Radial Distribution Function

A liquid is characterized by its structural and thermodynamic properties. The *pair correlation function* $g(r_1, r_2)$ is a fundamental quality in the structural descriptions of liquids [2]. It represents normalized probability of finding a

particle at r_1 and another particle at r_2 . For a spatially homogeneous fluid and isotropic system, it is also called *radial distribution function* (RDF) $g(r)$, where r is the relative distance between pair particles.

The RDF can be generalized to other structural functions, such as the *total correlation function* $h(r)$,

$$h(r) = g(r) - 1 \quad (1.1)$$

and *cavity distribution function* (or *background correlation function*) $y(r)$,

$$y(r) = g(r) \exp[\beta u(r)] \quad (1.2)$$

where $\beta = 1/k_B T$ is the inverse temperature, k_B is the Boltzmann constant, T is the absolute temperature and $u(r)$ is the pair potential between two particles.

In addition, all thermodynamic quantities can be expressed in terms of RDF. The main thermodynamic quantity *pressure* P can be computed directly from the pressure (virial) equation,

$$\frac{\beta P}{\rho} = 1 - \frac{2\pi\beta\rho}{3} \int_0^\infty \frac{du(r)}{dr} g(r) r^3 dr \quad (1.3)$$

or via the integration from the compressibility equation,

$$\frac{1}{\beta} \left(\frac{\partial \rho}{\partial P} \right)_T = 1 + 2\pi\rho \int_0^\infty (g(r) - 1) r^2 dr \quad (1.4)$$

where ρ is the density. The equations (1.3) and (1.4) provide two different routes to compute the pressure, and they should yield the same result. However,

for approximate theories two routes will give different values and this difference is called the *thermodynamic inconsistency* of the theory [2].

Since the RDF connects the structural and thermodynamic properties of liquids, it is of fundamental importance to characterize the RDF in liquid theory. The most powerful theoretical approach for calculating the RDF is the IE theory that based on the Ornstein-Zernike equation supplemented with a closure equation.

1.2 Integral Equation Theory

Considered a system of homogeneous and isotropic fluids, the Ornstein-Zernike (OZ) equation [2, 11] defines the *direct correlation function* $c(r)$ in terms of the *total correlation function* (i.e. the deviation of $g(r)$) through

$$h(r) = c(r) + \rho \int c(r') h(|r - r'|) dr' \quad (1.5)$$

where r' is the relative distance of the third particle from one of the pair particles. Equation (1.5) shows that the total influence of pair particles at a separation distance r is given by the sum of the direct influence of pair particles and an indirect part (i.e. appeared as a convolution integral, which is all indirect influences from the third particle exerted on the pair particles) [2]. By taking the Fourier transform of the OZ equation, we obtain a simpler form in k space,

$$\tilde{h}(k) = \tilde{c}(k) + \rho \tilde{h}(k) \tilde{c}(k) \quad (1.6)$$

where the symbol ‘tilde’ denotes the Fourier transform of a function. This transform enable us to find the solutions of the OZ equation easier.

As we have seen, both functions $h(r)$ and $c(r)$ are unknown and there is only one relation in the OZ equation. To solve these quantities, another relation, so-called closure equation, is required to complement the OZ equation [2]. Generally, the cavity distribution function, generalized from RDF by equation (1.2), is used as below,

$$y(r) = \exp[h(r) - c(r) + b(r)] \quad (1.7)$$

Combining equations (1.2) and (1.7), we obtain

$$h(r) = \exp[-\beta u(r) + h(r) - c(r) + b(r)] - 1 \quad (1.8)$$

which is an exact relation between $h(r)$ and $c(r)$, but includes $u(r)$ and unknown *bridge function* $b(r)$. For a given potential $u(r)$, we are able to find $h(r)$, $c(r)$ and corresponding $g(r)$ if we know the exact $b(r)$. Unfortunately, although the $b(r)$ can be represented by an infinite sum of bridge diagrams, the exact $b(r)$ is not clear for any system. Therefore, the bridge function plays a key role in the IE approach and in this sense a closure equation is an approximation for the bridge function. Some of the approximations are described in the following subsections.

1.2.1 Classical Closures

The simplest approximation for the bridge function is the hypernetted chain (HNC) equation which takes the bridge function as,

$$b_{HNC}(r) = 0 \quad (1.9)$$

and the $g(r)$ becomes

$$g(r) = \exp[-\beta u(r) + h(r) - c(r)] \quad (1.10)$$

This closure works well for long-ranged attractive potentials [2]. Another classical closure is the Percus–Yevick (PY) equation given by

$$b_{PY}(r) = t(r) - \ln[t(r) + 1] \quad (1.11)$$

where $t(r)=h(r)-c(r)$ and $g(r)$ is written as,

$$g(r) = (1 - e^{-\beta u(r)})c(r) \quad (1.12)$$

This closure gives very good results for fluids with short-ranged and sharply repulsive potentials, such as hard and soft spheres [2]. For hard spheres, the PY equation along with the OZ equation has been solved analytically [9, 10].

Since the HNC and PY equations are approximate, the pressure obtained from the virial route (i.e. equation (1.3)) differs from that of the compressibility route (i.e. equation (1.4)) [17]. Many attempts have been devoted to improve the performance of the HNC and PY theories. However, most of them are only empirical modifications of closures with the goal of improving the accuracy of the resulting $g(r)$ [41-43]. Another sort of approaches use a self-consistent procedure in which certain constraints for thermodynamic consistency are imposed to determine the closure with unspecified parameters [44-48].

1.2.2 Truncated series of the bridge diagrams

A third way to obtain a nonempirical or nonanalytical representation of is to evaluate diagrammatic expansion of $b(r)$ directly and truncate the infinite series at some point [24, 26-29, 49]. The bridge function can be expressed as a

series in powers of density,

$$b(r) = \sum_{n=2}^{\infty} b_n(r) \rho^n \quad (1.13)$$

where the coefficient $b_n(r)$ are multi-dimensional integrals, which can be written in terms of bridge diagrams. The bridge diagrams can be represented in terms of Mayer function $f(r)$ or the total correlation function $h(r)$. So far, the 2nd to 5th coefficients in terms of $f(r)$ have been found by Labik *et al.* [27] and up to 6th coefficient by Kolafa and Labik [50] for hard sphere systems. Recently, Kwak and Kofke calculated bridge diagrams of a hard sphere fluid up to order ρ^4 in terms of $h(r)$ [29]. Their results at high density are better than those using the $f(r)$ representation and classical closures (i.e. HNC and PY). Unfortunately, the convergences of the series using both $f(r)$ and $h(r)$ representations are slow at high density and the calculation of higher order terms is necessary to improve the accuracy of the IE theory to the next level.

1.2.3 Extension to Mixtures

Up to this point, we have concerned the one-component system. The multi-component system will be considered in this subsection. All aforementioned equations have their analogous expressions for mixtures and can be generalized in a similar way. The OZ equation for multi-components fluid takes the form

$$h_{\alpha\beta}(r) = c_{\alpha\beta}(r) + \rho \sum_{\nu=1}^n x_{\nu} \int c_{\alpha\nu}(r') h_{\nu\beta}(|r-r'|) dr' \quad (1.14)$$

and the analogue of exact closure is given by

$$h_{\alpha\beta}(r) = \exp[-\beta u_{\alpha\beta}(r) + h_{\alpha\beta}(r) - c_{\alpha\beta}(r) + b_{\alpha\beta}(r)] - 1 \quad (1.15)$$

where n is the number of species, subscripts α , β and ν represent the quantities for species α , β and ν , respectively, x_ν is the mole fraction of species ν .

The IE theory for mixtures can be extended to the wall-fluid system with certain assumptions and simple modifications [33, 51-54]. This extension is called singlet IE theory, in which the wall is treated as a giant molecule in an infinitely dilute solution (i.e. including the wall we have a binary mixture). We will not go into the details of the singlet IE theory in this chapter, as it will be explained in Chapter 4.

1.3 Numerical Solutions of Integral Equations

The OZ equation coupled with a chosen closure must be solved numerically, except the PY theory for the hard sphere system and a few other potentials [9, 10]. A number of numerical methods for solving OZ equation have been proposed [55-64]. In general, iterative algorithms are applied with some suitable convergence criteria. We introduce three most frequently used methods as follows.

1.3.1 Picard's Method

Picard's method refers to the solution by direct substitutions [55]. For a given bridge function, it starts from the closure equation (i.e. equation (1.8)),

which is usually written as

$$c(r) = \exp[-\beta u(r) + t(r) + b(r)] - 1 - t(r) \quad (1.16)$$

An initial guess of solution $t(r)=h(r)-c(r)$ is substituted into the RHS of equation (1.16) to get an output function $c(r)$. The output function is then substituted into the OZ equation (i.e. equation (1.6)) to get a new solution value in Fourier space. Next this new solution is again substituted into the RHS of equation (1.16) as a new input to yield a second output. The process is repeated until certain convergence criterion is satisfied. In general, the iterative procedure is given as following [56, 57]:

$$t(r) \xrightarrow{\text{Closure}} c(r) \xrightarrow[\text{Transfer}]{\text{Fourier}} \tilde{c}(k) \xrightarrow{\text{OZ}} \tilde{t}(k) \xrightarrow[\text{Transfer}]{\text{Inverse}} t(r)^{\text{new}} \quad (1.17)$$

An advantage of this numerical method is its simplicity. Its disadvantage is the poor convergence or even diverged at medium and high densities, which can be efficiently improved by mixing the old and new solutions with a dumping parameter ω : $t(r) = \omega t(r)^{\text{old}} + (1 - \omega)t(r)^{\text{new}}$. This mixing iteration algorithm was used by Duh and Haymet [57] to estimate the $h(r)$ and $c(r)$ at low and medium densities efficiently. However, at high densities the dumping parameter ω is close to 1 and the convergence is very slow.

1.3.2 Newton-Raphson Method

Another approach is to use the integral equations in a discrete form and to solve a set of non-linear equations by Newton-Raphson (NR) method [58]. Rapid convergence is obtained if a good initial estimate is used. This method

has been used to find solutions of the HNC, PY, and self-consistent approximations, and proved to be surprisingly valuable at high density, even in the region of the critical points [59]. However, it is time consuming because the number of equations is usually with the order of 100~1000, which means one might need to invert a large (typically 1000×1000) matrix.

1.3.3 Combination Method

It could be concluded that the direct iterative procedure is slow and the NR method is not convenient when applied to large set of equations. As a result, a combination of the two methods was proposed by Gillan [60]. Gillan's method reduces the size of the matrix in inversion and the procedure is very efficient. However, it is rather complex to implement the algorithm due to the complicated basis function (i.e. roof function). Later, Labik *et al.* [61] use the Fourier series as the basis function instead of the roof function to overcome this problem. A highlighted feature of the combination method is its insensitivity to the initial guesses for the numerical solution compared to earlier methods. Moreover, for most cases, the convergence is prompt. These advantages have made this method the primary choice to obtain solutions of integral equations. Thus we also use this method to solve the integral equations for the bulk fluid and the detailed procedure, which can be found in appendix A.

So far, the solutions introduced above are related to the bulk-fluid OZ equation. Since the OZ equation for the confined fluids is nonuniform due to the different confinements and boundary conditions, the solution is not unique. However, Pospisil *et al.* [52] proposed an efficient method to solve the

confined-fluid OZ equation for the fluid confined in a slit-pore (i.e. singlet OZ equation), which is similar to the Labik's algorithm proposed for the bulk fluid. This method is suitable for all densities and its convergence is fast. In our study, we will choose this algorithm for solving the singlet OZ equation, and the bulk properties involved in the singlet OZ equation can be obtained from the bulk-bulk-fluid OZ equation by using Labik's algorithm. The numerical procedure is shown in appendix B.

1.4 Diagrammatic Expression

1.4.1 Diagrammatic Technique

The functions in liquid theory are commonly defined as an infinite series of multi-dimensional integrals of one or more other functions [2]. It is very convenient to represent such series of integrals by *diagrams* or *graphs*. In this subsection, we will introduce some basic concepts in diagrammatic techniques. For example, the virial expansion of the equation of state is given by

$$\frac{\beta P}{\rho} = 1 + \sum_{i=2}^{\infty} B_i(T) \rho^{i-1} \quad (1.18)$$

The integrals appeared in the second and third virial coefficients B_2 , B_3 are given by,

$$B_2(T) = -\frac{1}{2} \int f_{12} d\mathbf{r}_2 \quad (1.19)$$

$$B_3(T) = -\frac{1}{3} \iint f_{12} f_{13} f_{23} d\mathbf{r}_2 d\mathbf{r}_3 \quad (1.20)$$

In the integrals $f_{ij} = [\exp(-\beta u_{ij}) - 1]$ is the *Mayer function*. For each integral there is a corresponding *labeled diagram* which contains a number of *circles* linked by *bonds*. The circles represent particle coordinates in a space and have two types: *white circles* (or *root points*), which represent coordinates held constant in the integration, and *black circles* (or *field points*), which stand for the variables of integration. Generally, the circle is often a function of coordinates and integration. The bond drawn as a line between two circles i and j represents certain function, say f_{ij} , which we name it as an f -bond. The value of a diagram is the value of the integral that the diagram represents, which is a function of the coordinates attached to the white circles and a functional of the functions associated with the black circles and bonds. Therefore, the integrals in equations (1.19) and (1.20) can be represented by simple labeled diagrams consisting of white and black cycles and f -bonds as following:

$$B_2 = -\frac{1}{2} \begin{array}{c} \text{---} \circ \text{---} \bullet \\ 1 \quad 2 \end{array}$$

$$B_3 = -\frac{1}{3} \begin{array}{c} \quad \bullet \quad 3 \\ \quad \diagup \quad \diagdown \\ \circ \quad \bullet \\ 1 \quad 2 \end{array}$$

Figure 1.1. Virial coefficients B_2 and B_3 in the f -bond expansion. The open circle represents one root particle (labeled by 1) and the solid circles (labeled by 2 and 3) denote field points, which are subjected to be integrated over the connecting f -bond(s).

1.4.2 Cluster Expansion

Similar to the equation of state, all aforementioned pair functions can be

represented as an infinite series in powers of density, where the coefficients are sums of integrals over products of other simpler pair functions which can be depicted by diagrams [2]. For example, the cavity distribution function $y(r)$,

$$y(r) = 1 + y_1(r)\rho + y_2(r)\rho^2 + \dots \quad (1.21)$$

in which the coefficients $y_i(r)$ are series of f -bonds diagrams. The first two of diagrams are shown in Figure 1.2 and we use the last row in Figure 1.2 to represent this diagrammatic expansion. Then the direct correlation function $c(r)$ and total correlation function $h(r)$ in the OZ equation can be represented as Figure 1.3.

Figure 1.2. Cavity distribution function coefficients y_1 and y_2 in the f -bond density expansion. The open circles represent two root particles and the solid circles denote field points. The first row shows the single diagram in y_1 and the second row shows 5 diagrams in y_2 , respectively. The last row represents the diagrammatic expansion of y . The number of solid circles is related to the number of powers of density and thus the order of each term.

$$\begin{aligned}
c &= \text{---} + \text{---} + \text{---} + \text{---} + \text{---} + \text{---} + \\
&\quad \text{---} + \text{---} + \text{---} + \dots \\
h &= \text{---} + \text{---} + \dots
\end{aligned}$$

Figure 1.3. Diagrammatic expansions of direct correlation function $c(r)$ and total correlation function $h(r)$. Circles and bonds are the same as in Figure 1.2.

Note that the diagrammatic expansion is also called cluster expansion; as a consequence the integrals arising in cluster expansion is called cluster integrals and the diagram, referred as cluster diagram (or simply cluster).

1.4.3 Bridge Diagrams

To this point, the pair functions in the IE theory are all expressed in diagrammatic form, except the bridge function. Recalling equation (1.13), it is shown that the bridge function also can be represented by cluster expansion. coefficients $b_n(r)$ in equation (1.13) are the sum of highly connected cluster diagrams that representing multi-dimensional integrals, which consists of n points, two root points in a relative distance r and bonds. There are two famous routes to correlate the bonds appearing in the bridge diagrams: the Mayer function (i.e. f -bond) and the total correlation function (i.e. h -bond). With f - f -bonds, the first two orders of bridge diagrams are given in Figure 1.4.

Figure 1.4. Bridge diagrams b_2 and b_3 in the f -bond density expansion of bridge function. The open circles represent two root particles (labeled by 1 and 2, for clarity, these labels are omitted in the other diagrams) and the solid circles denote field points to be integrated over the connecting f -bonds. The first row represents b_2 and its integral expression. The second row shows 13 diagrams in b_3 and the value of each coefficient is the sum of all diagrams with their weights, respectively. Number below diagrams denotes their symmetry numbers which is the reciprocal value of the weight.

As shown in Figure 1.3, the $h(r)$ can be represented by a series in powers of density and the coefficients are integrals over products of $f(r)$, therefore the h -bond implicitly incorporates the f -bonds. Due to this property of the h -bond, one might expect a better representation of the bridge function with fewer terms by the h -bond expansion that is derived by Stell [65]. The first two orders of bridge diagrams in terms of h -bonds are shown as following:

Figure 1.5. Bridge diagrams b_2 and b_3 in the h -bond density expansion of bridge function. Circles are the same as in Figure 1.4, but the connecting lines signify the total correlation function $h(r)$. The first row represents b_2 and its integral expression. The second row shows 7 diagrams in b_3 and the value of each coefficient are calculated in the same way as that of f -bond.

The Figure 1.4 and Figure 1.5 show that there is only one diagram of four particles in f -bond b_2 , and 13 diagrams of five particles in f -bond b_3 ; for h -bond expression, 1 for b_2 and 7 for b_3 . Moreover, the bridge diagrams in h -bond expression are a subset of those with f -bonds, which indicates that the h -bond expansion indeed reduces the number of diagrams in the bridge function, as a consequence h -bond expression may improve the convergence of the series and save the computational time. This is the reason we choose h -bond expression to study bridge function through this research. It is worth nothing that the number diagrams in the second tern b_3 with f -bond could be further reduced to 9 and those with h -bond to 5, as some of diagrams are symmetric and belong to the same set of topology and thus have identical values. Another remarkable feature of h -bond expression is that the b_n using h -bond is dependent on density, to that in terms of f -bonds is density independent.

Chapter 2

General Methodology

2.1 Monte Carlo Simulation

Monte Carlo (MC) simulation [66] is very impotent to the development of liquid state theory [2]. It provides essentially exact results of well-defined models compared to experimental data and sometimes it is the only way to data for models that do not exist in nature. The MC method is designed to generate static configurations of the system of interest. The quantity of interest obtained as an *ensemble* average, which only depends on the configurational variables. Therefore, the MC method is a very efficient approach to calculate static properties of the system. Nowadays, using MC simulation to obtain the mechanical properties such as pressure and internal energy that can be as ensemble averages is a relatively routine matter. Moreover, MC method has a wider application and can be used to evaluate the multi-dimensional integrals, such as bridge diagrams.

The MC simulation can be carried out in various ensembles, such as the canonical ensemble and grand canonical ensemble. The conventional MC probes the canonical ensemble and the Metropolis scheme, usually referred as Metropolis MC [67], which involves the importance sampling technique. There are other sampling techniques that beyond Metropolis, so-called *biased techniques*, such as multicanonical sampling [66] in Transition Matrix Monte Carlo [30, 68] in this work and umbrella sampling [66]. A brief review of these

simulations and theoretical approaches is given below.

2.1.1 Monte Carlo

We introduce the concepts of MC simulation for *canonical ensemble* which stands for a system with a fixed particles number N in a given volume V at a temperature T . The macroscopic quantity A of the system can be calculated as an ensemble average

$$\langle A \rangle = \frac{\int d\mathbf{r}^N A(\mathbf{r}^N) \exp[-\beta U(\mathbf{r}^N)]}{\int d\mathbf{r}^N \exp[-\beta U(\mathbf{r}^N)]} \quad (2.1)$$

where $U(\mathbf{r}^N)$ is the potential energy, $\mathbf{r}^N$ stands for the coordinates of all N particles and $\exp[-\beta U(\mathbf{r}^N)]$ is the *Boltzmann factor*. The *configurational part of the partition function* Z is defined by

$$Z \equiv \int d\mathbf{r}^N \exp[-\beta U(\mathbf{r}^N)] \quad (2.2)$$

The sample configurations that contribute to the average are generated according to the probability distribution $p(\mathbf{r}^N)$, which is the probability density of observing the system in the configuration around $\mathbf{r}^N$. A series of successive configurations (i.e. *Markov chain*) corresponding to the probability distribution is produced by *Markov process* which consists of Markov transitions generating a new microstate (i.e. configuration) t out of a given microstate s by a random walk [69]. The transitions often obey the *detailed balance* given by

$$p(s)Q(s \rightarrow t) = p(t)Q(t \rightarrow s) \quad (2.3)$$

where $Q(s \rightarrow t)$ is the transition probability of generating a new microstate t from the current microstate s , which is the product of a selection probability $\phi(s \rightarrow t)$ and a acceptance ratio $acc(s \rightarrow t)$

$$Q(s \rightarrow t) = \phi(s \rightarrow t) \times acc(s \rightarrow t) \quad (2.4)$$

where $\phi(s \rightarrow t)$ is the probability of the system moving to a new microstate t provided the current microstate s and $acc(s \rightarrow t)$ is the probability of the generated microstate t is accepted as the new microstate of the system. For a specific probability distribution, $acc(s \rightarrow t)$ can be determined through the condition of detailed balance (i.e. equations (2.3) and (2.4)). To summarize, the standard MC procedure has two stages.

1. *Trial moves*: A new microstate t is generated from a given current microstate s of the system by a trial move. In canonical ensemble, the trial move is given by a random displacement of a particle i

$$\mathbf{r}_i \rightarrow \mathbf{r}_i + \Delta \boldsymbol{\xi}_i \quad (2.5)$$

where $\boldsymbol{\xi}_i$ is a vector of real random numbers uniformly distributed in the interval $[-1,1]$ and Δ is the maximum allowed displacement with upper limit (i.e. $\Delta \leq H/2$, H is the box length).

2. *Accept or reject*: The trial move is conditionally accepted with an acceptance probability $acc(s \rightarrow t)$. If rejected, the old microstate s is kept.

Many schemes are proposed to fulfill above two stages and a brief review given below.

2.1.2 Metropolis Monte Carlo

Metropolis *et al.* introduced a particular and very efficient method, so-called *importance sampling*, to select microstates according to the *Boltzmann distribution*. Apart from the *simple sampling* in ‘crude’ MC which all microstates are chosen with equal probability, the importance sampling is not to sample all possible microstates, but relatively few and representative microstates that contribute much to the average. This is achieved by setting the selection probabilities all equal and the condition of detailed balance (i.e. equations (2.3) and (2.4)) becomes

$$\frac{Q(s \rightarrow t)}{Q(t \rightarrow s)} = \frac{acc(s \rightarrow t)}{acc(t \rightarrow s)} = \frac{p(t)}{p(s)} \quad (2.6)$$

The probability distribution $p(s)$ that used in the Metropolis MC is the Boltzmann distribution given by

$$p(s) = \frac{\exp[-\beta U(s)]}{Z} \quad (2.7)$$

Therefore the Metropolis algorithm of acceptance ratio follows

$$acc(s \rightarrow t) = \min \left[1, \frac{p(t)}{p(s)} \right] = \min \left[1, e^{-\beta[U(t)-U(s)]} \right] \quad (2.8)$$

Although the Metropolis scheme that concentrates sampling in the most important region of the configuration space can be used to sample any system, has its limitations. One drawback of the Metropolis scheme is that the times using the importance sampling can be prohibitively long in systems that

require traversing a path where the relative probabilities differ by many orders magnitude. To overcome such probability barriers, the biased samplings are proposed.

2.1.3 Transition Matrix Monte Carlo

One possible choice is the Transition Matrix Monte Carlo (TMMC) scheme [30, 70-74] which provides a highly efficient mechanism to speed up the simulation by utilizing the statistics about attempted transitions between microstates in the Metropolis scheme. The TMMC scheme employed here is an adaptation of the algorithm proposed by Fitzgerald *et al.* [68] for studying lattice system and was first used by Errington *et al.* [30] to study the liquid-vapor phase equilibrium. Implementation of the transition matrix algorithm is very simple that one only needs to add a third stage to the standard MC procedure.

To connect with macroscopic quantities, the macrostate S that encompass a set of microstates ($s \in S$) is defined by one or more prespecified macrovariables. The probability $\Pi(S)$ of finding the system in macrostate S is given by the summation of probabilities over all microstates that belong to macrostate S ,

$$\Pi(S) = \sum_{s \in S} p(s) \quad (2.9)$$

In order to estimate the macrostate probability $\Pi(S)$, a third stage is added to the MC cycle. In this stage, the statistics regarding attempted transitions between macrostates is collected in a collection matrix C . The collection matrix is after each trial move as follows:

$$C(S \rightarrow T) = C(S \rightarrow T) + acc(s \rightarrow t) \quad (2.10)$$

$$C(S \rightarrow S) = C(S \rightarrow S) + 1 - acc(s \rightarrow t) \quad (2.11)$$

It is important to emphasize that the matrix is updated whether the MC move is accepted or rejected. The collection matrix is used to calculate the macrostate transition probability $P(S \rightarrow T)$ by

$$P(S \rightarrow T) = \frac{C(S \rightarrow T)}{\sum_{\Delta S} C(S \rightarrow S + \Delta S)} \quad (2.12)$$

Once $P(S \rightarrow T)$ which indicates the likelihood of the system moving from current macrostate S to a new macrostate T is obtained, the detailed balance expression is employed to determine the $\Pi(S)$ through

$$\Pi(S)P(S \rightarrow T) = \Pi(T)P(T \rightarrow S) \quad (2.13)$$

To bias the simulation, one may introduce a weight function $\eta(S)$ such that all macrostates of the system are sampled with uniform probability. This is achieved by using *multicanonical sampling* technique, in which the weight function is set inversely proportional to the macrostate probability

$$\eta(S) = -\ln \Pi(S) \quad (2.14)$$

Correspondingly, the acceptance criterion for a trial move becomes

$$acc(s \rightarrow t) = \min \left[1, \frac{p(t)e^{\eta(T)}}{p(s)e^{\eta(S)}} \right] = \min \left[1, e^{-\beta[U(t)-U(s)]+\eta(T)-\eta(S)} \right] \quad (2.15)$$

Note that we still use the unadjusted acceptance criterion to update the matrix

2.1.4 Umbrella Sampling

Another possible choice to bias a simulation is the *umbrella sampling* [66] scheme, which is usually used to calculate the free energies. It is well known that the conventional MC is unable to calculate properties such as entropy and free energy directly, as the free energy cannot be expressed as ensemble averages. In general, the free energy of a system can be determined by evaluation the difference between the system of interest and some reference systems of known free energies. This approach is called *free energy perturbation* [66, 75], which permits the free energy difference between two systems to be written as an ensemble average easily, but the conventional MC methods of estimating such average are usually inadequate due to that the Boltzmann-weighted sampling distribution are inefficient in some cases. To overcome this difficulty, the umbrella sampling method is developed.

The free energy is directly related to the partition function Z (i.e. equation (2.2)) [66] and the free energy difference between two systems A and B (i.e. $\Delta F = F_B - F_A$) in canonical ensemble is given by

$$\begin{aligned} -\beta\Delta F &= \ln \left(\frac{\int d\mathbf{r}^N \exp[-\beta U_B(\mathbf{r}^N)]}{\int d\mathbf{r}^N \exp[-\beta U_A(\mathbf{r}^N)]} \right) \\ &= \ln \left(\frac{\int d\mathbf{r}^N \exp[-\beta(U_A(\mathbf{r}^N) + \Delta U)]}{\int d\mathbf{r}^N \exp[-\beta U_A(\mathbf{r}^N)]} \right) \end{aligned} \quad (2.16)$$

We introduce the notation $\langle \rangle_A$ to denote an average over a canonical ensemble of the reference system A and the equation (2.16) becomes

$$\exp(-\beta\Delta F) = \langle \exp(-\beta\Delta U) \rangle_A \quad (2.17)$$

This is the basic formula for free energy perturbation. To improve this approach, a bias is introduced by replacing the Boltzmann factor with a non-negative weight function $w(\mathbf{r}^N)$. The weight function $w(\mathbf{r}^N)$ is chosen to favor those configurations that contribute much to both systems A and B and should have an appreciable overlap with both regions of configuration space relevant to systems A and B . The term ‘umbrella sampling’ is given to describe this connecting property. To this end, the probability of visiting a configuration around $\mathbf{r}^N$ is now proportional to $w(\mathbf{r}^N)$ and the expression for $\langle \exp(-\beta\Delta U) \rangle_A$ is given by

$$\langle \exp(-\beta\Delta U) \rangle_A = \frac{\int d\mathbf{r}^N w(\mathbf{r}^N) \exp[-\beta U_B(\mathbf{r}^N)] / w(\mathbf{r}^N)}{\int d\mathbf{r}^N w(\mathbf{r}^N) \exp[-\beta U_A(\mathbf{r}^N)] / w(\mathbf{r}^N)} \quad (2.18)$$

Representing an ensemble average over a probability distribution proportional to $w(\mathbf{r}^N)$ by $\langle \rangle_w$, one obtains

$$\langle \exp(-\beta\Delta U) \rangle_A = \frac{\langle \exp(-\beta U_B) / w \rangle_w}{\langle \exp(-\beta U_A) / w \rangle_w} \quad (2.19)$$

To account for the bias, the acceptance criterion for a MC trial move becomes

$$acc(s \rightarrow t) = \min \left[1, \frac{w(t)}{w(s)} \right] \quad (2.20)$$

2.1.5 Grand Canonical Ensemble Monte Carlo

So far, we have considered the simulation in canonical ensemble. In to canonical ensemble, the simulation in *grand canonical ensemble* (GCMC)

is more convenient for the confined fluids and adsorptions as it is a thermodynamic open which means that the particles and energy between the system and a reservoir are allowed to transfer. In grand canonical ensemble, the volume V , inverse temperature β and chemical potential μ are fixed, while the particle number N and the energy E are allowed to fluctuate. The algorithm presented in the Metropolis MC can be easily generalized to grand canonical ensemble with simple modifications.

Apart from using the displacements of particles as trial moves in the canonical ensemble, we use trial insertions and deletions of particles to incorporate fluctuations in the number of particles. Then Metropolis rule for acceptance of these attempted particle number changes are given by

$$acc(N \rightarrow N-1) = \min \left[1, \frac{N}{V} e^{-\beta[U(r^{N-1}) - U(r^N) - \mu]} \right] \quad (2.21)$$

$$acc(N \rightarrow N+1) = \min \left[1, \frac{V}{N+1} e^{-\beta[U(r^{N+1}) - U(r^N) + \mu]} \right] \quad (2.22)$$

As a result, both the particle number and energy fluctuate in the simulation. The GCMC simulation can be easily applied to confined fluids by fixing the chemical potential of fluids inside the confinement equal to that of bulk fluid and letting the particle number fluctuates.

2.2 Bridge Diagram Calculation

Evaluating the bridge diagrams with either f -bonds or h -bonds is a painstaking task, because their diagram numbers rapidly expand as the order

increases and each diagram represents a high-dimensional integral over the coordinates of pair particles. Although the h -bond expansion reduces the number of diagrams (see Figure 1.4 and Figure 1.5), there are still 56 distinct diagrams (i.e. 88 diagrams) that contribute to order ρ^4 in h -bond expansion and 956 distinct diagrams for the next order in density. Therefore it is time consuming to evaluate the higher order bridge diagrams. Except that the first order term in f - f -bond expansion of the bridge function of hard sphere fluid is known analytically [76], the higher order terms must be calculated numerically. There have been several recent attempts to obtain higher order terms in the bridge function. In this section, we review several frequently used methods as shown below.

2.2.1 Legendre Polynomial Expansion

One possible route to evaluate the bridge diagram is using the Legendre polynomial expansions. The Legendre expansion expressions were first proposed for f -bond b_1 and b_2 by Attard and Patey [77] for pure hard sphere and mixture systems and later simply adjusted for h -bond b_1 and b_2 by Perkyns and Pettitt [26] for pure Lennard-Jones system. These Legendre formulas were also applied to other systems by others [78, 79]. Although the Legendre expansion reduces the computational requirements by reducing the high-dimensional integration to two- or three-dimensions, it is not straightforward and rather complex since every diagram needs a special treatment. Therefore, it is not suitable for calculation of the higher order bridge diagrams beyond b_3 . The details of the Legendre expansion expressions are not presented here, but it can be found elsewhere [26, 77].

2.2.2 Biased Monte Carlo Method

A more general route to evaluate the higher order bridge diagrams is using the biased MC Methods. There are several MC methods developed for bridge diagrams calculation. Rast *et al.* [24] first proposed a general biased MC method with a suitable nonuniform distribution of integration points to integrate cluster integrals, and it is applied to compute b_1 and b_2 in term of h -bond for hard sphere and Lennard-Jones systems. This approach is a direct MC numerical integration of cluster integrals appearing in the bridge diagrams. Labik *et al.* [27] proposed a new biased MC Method by replacing the Boltzmann distribution in Metropolis MC with a probability density provided by the *framework* cluster to evaluate the 2nd to 5th coefficients of bridge function with f -bonds of a hard sphere fluid. Unlike the approach of Rast *et al.*, Labik *et al.* did not attempt to evaluate the desired clusters directly, but to calculate the ratio of the desired cluster to the framework cluster with known integral. The framework cluster is chosen to have a simpler form and can be calculated analytically. Several framework clusters were needed for evaluating b_n higher than b_3 and the clusters in b_n term were calculated separately in different MC simulations. It is not convenient for higher orders beyond ρ^3 since each b_n term has to be divided into small groups and each group has to be treated particularly.

2.2.3 Free-Energy Perturbation Method

Recently, a free energy perturbation formula, called Mayer-sampling, was proposed to calculate cluster integrals by Singh and Kofke [28] and applied to h -bond bridge diagrams up to order ρ^4 by Kwak and Kofke [29]. Their main

argument lies in the viewpoint of the relation between the thermodynamic free energy and the configurational integral, which can be explained as cluster integrals. The excess free energy of classical fluids that directly connected to partition function Z is definitely a configurational integral which can be by diagrams as shown in Figure 2.1 [80], therefore the methods to calculate free energy are methods for calculation of cluster integrals, if casted in the proper form, can be used to evaluate the cluster integrals that appearing in the bridge function. In this subsection, we give a brief description of two adaptations of free-energy methods used to calculate the bridge diagrams.

$$-\beta F = \bullet\text{---}\bullet + \begin{array}{c} \bullet\text{---}\bullet \\ | \quad \diagup \\ \bullet \end{array} + \begin{array}{c} \bullet\text{---}\bullet \\ | \quad | \\ \bullet\text{---}\bullet \end{array} + \begin{array}{c} \bullet\text{---}\bullet \\ | \quad | \quad \diagup \\ \bullet \end{array} + \begin{array}{c} \bullet\text{---}\bullet \\ | \quad | \quad \diagup \quad \diagdown \\ \bullet \end{array} + \dots$$

Figure 2.1. Excess free energy in the f -bond density expansion. The solid circles denote field points and integrations. The number of circles is related to the number of powers of density. The coefficients before the diagrams are omitted for clarity, which is $1/2$ for 1st, $1/6$ for 2nd, $1/8$ for 3rd, $1/4$ for 4th and $1/24$ for 5th.

2.2.3.1 Free Energy Perturbation

The free energy perturbation is a straightforward free energy method as show in subsection 2.1.4, a key ideal of which is to work with free energy difference between the target and reference systems to avoid direct evaluation of free energy in the target system, which is also adaptable for cluster integrals. The basic working equation (i.e. equation (2.17)) for this method can be recovered as

$$\Gamma(r) = \Gamma_o \langle \gamma(r) / \gamma_o \rangle_{\gamma_o} \quad (2.23)$$

where $\Gamma(r)$ represents a general cluster integral or sum of integrals, with integrand (or sum of integrands) $\gamma(r)$; for example, if Γ is the second bridge coefficient b_2 , then $\gamma = h_{12}h_{23}h_{34}h_{24}$. The angle bracket indicates the ensemble average over all configurations and the subscript “o” indicates the reference system, for which Γ_o is known. This approach resembles the Metropolis MC with the sampling distribution given by γ_o which is positive. In addition, this formula is compatible to the approach of Labik *et al.* who use the framework cluster as a reference γ_o .

2.2.3.2 Mayer-Sampling

A more flexible method is the umbrella sampling, referred as Mayer-sampling in cluster integrals calculation, for which the working formula (i.e. equation (2.19)) now becomes

$$\Gamma(r) = \Gamma_o \frac{\langle \gamma(r) / \pi \rangle_{\pi}}{\langle \gamma_o / \pi \rangle_{\pi}} \quad (2.24)$$

where the subscript π is the normalized probability distribution that governs the sampling. In this approach, the MC sampling is performed on a number of particles equal to the order of the integral and cluster configurations are using the Metropolis importance sampling with weights based on the magnitude of the interactions that are represented in the given cluster. An umbrella average yields the value of a desired cluster integral to a reference cluster. It is important to emphasize that π is non-negative and should be chosen to

sample important regions for both target cluster $\gamma(r)$ and reference cluster γ_0 . Since the reference cluster is separated from its role as sampling function, the choice of π is arbitrary; any cluster based on any interaction potential may be used.

To summarize, there are two key ideas in the Mayer-sampling when applied to cluster integrals. First, we generate the molecule configurations using the Metropolis importance sampling. Second, we aim to evaluate the ratio of the desired cluster integral to a known reference integral. These two ideas open two doors to improve the Mayer-sampling. One is replacing the sampling function via application of other sampling techniques, such as TMMC; another is introducing one or more intermediate systems to maximize the overlap regions between the target and reference systems. According to the above two routes, we introduce two new methods to improve the performance of Mayer-sampling and give the name TMMC Mayer-sampling and overlap sampling to represent them. We will not present the details of these two methods here, but in the next chapter.

2.3 Simulation Models

In this section, I give a brief review of two models used in this work. A summary of properties is given below.

2.3.1 Hard Sphere Model

Hard spheres are defined as spherical impenetrable particles that can not overlap in space. The hard sphere potential is widely used in the simplest nontrivial models of fluids and solids due to theoretical tractability. It mimics

strong repulsion that atoms and molecules of spherical shape experienced at contact distances. The particles of diameter σ have the following pair-wise interaction potential:

$$U(r) = \begin{cases} 0 & r > \sigma \\ \infty & r \leq \sigma \end{cases} \quad (2.25)$$

where $U(r)$ is the potential energy between two particles at a relative distance r . The potential is infinite if two particles overlaps, otherwise zero. For hard spheres, the virial equation of state (i.e. equations (1.3) and (1.18)) becomes,

$$\frac{\beta P}{\rho} = 1 + \frac{2\pi\rho}{3} \sigma^3 g(\sigma^+) \quad (2.26)$$

in which the pressure only depends on the value of the RDF at contact $g(\sigma^+)$.

The hard sphere potential is a very good checking ground for many IE theories and we use it to test our new methods as shown in next chapter. Moreover, the extension of the hard sphere potential to binary mixtures has following from:

$$U_{ij}(r) = \begin{cases} 0 & r > \sigma_{ij} \\ \infty & r \leq \sigma_{ij} \end{cases} \quad \sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \quad (2.27)$$

where two particles i and j have strictly additive diameters.

2.3.2 Hard Sphere-Wall Model

The hard sphere-wall models have been considered as the simplest models of confined fluids. A commonly investigated hard sphere-wall model is a hard

sphere fluid confined in a slit-pore. In this model, the system consists of N hard spheres of diameter σ confined between two parallel planar hard walls. The pair interaction potential $U(r)$ of hard spheres that inside the slit-pore at a relative distance r is given by equation (2.25), and the interaction potential $U(z)$ a wall and a hard sphere at a distance z is given by

$$U(z) = \begin{cases} \infty & |z| \geq (L - \sigma)/2 \\ 0 & |z| < (L - \sigma)/2 \end{cases} \quad (2.28)$$

where L is the slit width, the origin is set on the centre of the walls and z -axis is selected perpendicular to the walls. The potential energy of a particle is infinite when it collides with the wall, otherwise zero. Compared to equation (2.27), we can see that the hard sphere-wall model can be considered as an extension to hard sphere binary mixtures by assuming one of the species has an infinite dilute concentration and diameter. This is responsible for the theoretical tractability of the singlet IE theory.

So far, we have shown the working equations of two model potentials. The schematics of the two potentials are shown as following:

Figure 2.2. The schematics of the potentials for hard sphere model (left) and sphere-wall model (right).

2.4 Simulation Environment

We apply the programming package *Etomica* in our studies [81, 82]. *Etomica* is an application program interface (API) for the construction of Java-based molecular simulation. The *Etomica* molecular simulation API is a collection of Java classes that are written to represent all the components of a simulation: species of molecules, the space, the phase, and the potentials or molecular interactions as well as that carry out MD or MC simulation, control the simulation process, and record the properties of the system. The 3D visualization of *Etomica* is supported by the OpenGL technology.

Figure 2.3 Screenshot of the simulation environment based on eclipse java IDE

Our simulations can be performed graphically or non-graphically as well on the PC (*Windows* system) level or in a computer Cluster (*Linux* system).

Usually, we run small systems on PC graphically to code and test the coding the actual simulations are performed in Cluster non-graphically. We use *eclipse* to manipulate the *Etomica* package. Eclipse is a Java Integrated Development Environment (IDE) developed by IBM. A screenshot of the simulation system shown in Figure 2.3.

Chapter 3

New Methods of Bridge Function Calculation

3.1 Introduction

The primary goal of the IE theory [2, 11] is to predict structural properties of classical fluids from the knowledge of intermolecular interactions. In the past, the IE theory has been recognized [38] as a relatively weak field among its sister methods such as analytical or semi-empirical equation of states, and molecular simulation because it relies on accurate prediction of the bridge function, which can be expressed by the infinite sum of bridge coefficients (i.e. equation (1.13)). During recent years, much effort has been put into getting accurate approximations [12-16, 20, 44, 45, 83, 84] of the bridge function. A route is to evaluate $b(r)$ by direct calculations of the bridge coefficients $b_n(r)$ in equation (1.13); $b_n(r)$ might be evaluated from treating pair interactions in bridge cluster diagrams as $h(r)$ (i.e. Figure 1.5). Also, other functions such as $f(r)$ (i.e. Figure 1.4), and etc., have also been used [8, 24, 26, 27, 49, 85]. The advantage of using $h(r)$ is to allow one to handle fewer diagrams in the bridge coefficients, but the difficulty still remains due to complex interaction schemes. Recently, Singh and Kofke [28] have developed a particular type of the free energy perturbation method called the Mayer-sampling technique, which was used to evaluate configurational integrals arising in the cluster expansion of virial coefficients. Benjamin *et al.* [28, 31, 32] calculated virial coefficients of several models with the help of Mayer-sampling and overlap sampling

methods. The first use of the Mayer-sampling with the conventional Monte Carlo (MC) simulation to obtain the bridge function in terms of h -bond expansion was done by Kwak and Kofke [29], but the accurate prediction has been hindered where the relative distance of two root points is short or long as the density and the bridge coefficient order increase. We notice that the IE theory generically suffers from prediction of accurate $g(r)$ at high density close to freezing, but there exists a possibility to obtain better $g(r)$ by improving the sampling technique for bridge cluster configurations, and that is the main focus of this chapter.

The generic MC method produces the uneven configuration distribution over a fixed distance (i.e. Figure 3.2). However, the TMMC [30, 70-74] provides an efficient reweighting technique thus paths for particles to move to low-probability regions. Therefore, we adopt the TMMC scheme with the Mayer-sampling method to improve the accuracy of the bridge coefficients over core and tail regions of the relative distance of two root points. Another issue to consider in the Mayer-sampling technique is the direction of perturbation (i.e. Figure 3.1). Kwak and Kofke's work chose a ring structure with f -bond of hard sphere as a reference system, which follows the umbrella sampling. In this chapter, we present a different approach, the overlap sampling method [31, 32, 86], which has been applied to improve the precision of virial coefficients. The phase spaces of bridge coefficient calculations corresponding to umbrella (i.e. Mayer) and overlap samplings are depicted in Figure 3.1. There are two options to define the intermediate system C between the target system A and the reference system B . In the umbrella sampling, C is formulated to contain both A and B , and the sampling is performed from C into each system A and B . In the

overlap sampling, C is formulated to be a subset of both A and B . Hence, the sampling is performed from each system A and B into C [75, 87]. In order to give a better precision of $b_n(r)$, C should be appropriately selected. In the umbrella sampling, many configurations important to the f -bond ring structure (i.e. reference system) are not efficiently sampled, but this deficit can be overcome if the overlap sampling is applied.

Figure 3.1. The phase-space schematic of appropriate intermediate system (labeled C) for the bridge coefficient calculation constructed between target system A and reference system B for (a) umbrella sampling; (b) overlap sampling. Arrows from one system to another indicate the sampling direction.

3.2 Simulation Methodology

3.2.1 TMMC Mayer-Sampling

The TMMC Mayer-sampling method is similar in nature to the Mayer-

Mayer-sampling (i.e. equation (2.24)) for evaluating cluster integrals. Apart from the Mayer-sampling that incorporated the Metropolis MC approach, it is based on the TMMC scheme and the governing sampling involved in this method is the magnitude of the interactions that are represented in the given cluster with the TMMC weight function η . The weight function will be reevaluated at regular intervals during simulation. Mathematically, the general cluster integral can be expressed as,

$$\Gamma(r) = \Gamma_o \frac{\left\langle \gamma(r) / e^{\eta(N)} \pi \right\rangle_{e^{\eta(N)} \pi}}{\left\langle \gamma_o / e^{\eta(N)} \pi \right\rangle_{e^{\eta(N)} \pi}} \quad (3.1)$$

where $\Gamma(r)$ represents a general cluster integral or sum of integrals, with integrand (or sum of integrands) $\gamma(r)$. The subscript “o” indicates the reference system, for which Γ_o is known. The angle bracket indicates the ensemble average over all configurations and the subscript $e^{\eta(N)} \pi$ is the probability distribution that governs the sampling. The N in the weight function represents the quantity for macrostate that are needed in TMMC scheme. It is defined by regularly partitioned distances in the relative distance r of two root particles and is equal to the total number of bins relevant to r .

To obtain the weight function η , we employ the TMMC scheme described Chapter 2, in which the weight function is set to be inversely proportional to the current macrostate probability $\ln \Pi(N): \eta(N) = -\ln \Pi(N)$. Since a way to discretize r is arbitrary, the bin width δr in the definition of macrostate is set to $\delta r = 0.1 \sigma$ to capture the probability distribution by TMMC moves that restrict the new separation distance of the random range within $r - \delta r$ and $r + \delta r$. This

sequential approach provides a suitable means for obtaining the macrostate probabilities as follows,

$$\ln(N+1) = \ln(N) + \ln \left[\frac{P(N \rightarrow N+1)}{P(N+1 \rightarrow N)} \right] \quad (3.2)$$

$$P(N \rightarrow N') = \frac{C(N \rightarrow N')}{\sum_{\Delta N} C(N \rightarrow N + \Delta N)} \quad (3.3)$$

where $P(N \rightarrow N')$ is the transition probability and C is a collection matrix which contains statistics about the attempted transitions between macrostates. The collection matrix is updated after each MC step as follows,

$$C(N \rightarrow N') = C(N \rightarrow N') + acc(r \rightarrow r') \quad (3.4)$$

$$C(N \rightarrow N) = C(N \rightarrow N) + 1 - acc(r \rightarrow r') \quad (3.5)$$

where acc is the importance sampling acceptance criterion based on the Metropolis MC algorithm; $acc(r \rightarrow r') = \min[1, \pi(r)/\pi(r')]$. To account for the weight function, the acceptance probability for a MC trial move now becomes

$$acc(r \rightarrow r') = \min \left[1, \frac{e^{\eta(N)} \pi(r)}{e^{\eta(N')} \pi(r')} \right] \quad (3.6)$$

Note that we still use the unadjusted acceptance criterion to update the matrix C .

3.2.2 Overlap Sampling

Both Mayer-sampling and TMMC Mayer-sampling correspond to *direct-direct*-sampling methods as they involve direct perturbations between the target

system, of which the absolute value of the target cluster(s) governs the sampling, and the reference system. The overlap sampling [31, 32, 86] is a desirable alternative, which takes the target and reference systems perturbed into their overlapping regions in the phase-space. One can define an overlap function to represent the region important only to both target and reference systems as follows:

$$\gamma_{os} = \frac{|\gamma_o||\gamma|}{\lambda|\gamma_o| + |\gamma|} \quad (3.7)$$

where γ_{os} is the overlap function and λ is an optimization parameter. In this formula, we omit r from $|\gamma|$ since it particularly represents the sum of the absolute value of cluster(s) over some MC steps. Thus a cluster integral can be expresses as follows,

$$\Gamma(r) = \Gamma_o \frac{\langle \gamma(r) / \pi \rangle_\pi / \langle \gamma_{os} / \pi \rangle_\pi}{\langle \gamma_o / \pi_o \rangle_{\pi_o} / \langle \gamma_{os} / \pi_o \rangle_{\pi_o}} \quad (3.8)$$

To find a good value of λ , we employ an optimization method [31, 32], which uses the following criterion,

$$\langle \gamma_{os} / \pi_o \rangle_{\pi_o} = \lambda \langle \gamma_{os} / \pi \rangle_\pi \quad (3.9)$$

The overlap sampling calculation is conducted in the same way as the Mayer-sampling except that we used two systems. One with the sampling governed by π_o yields the average for the denominator in equation (3.8), and second with the sampling governed by π yields the average for the numerator of the same equation. Since the criterion in equation (3.9) assumes that the

statistical uncertainties for both systems of overlapping region are equal [32], we monitor the statistical uncertainty of the results from each simulation to help us choose the simulation steps of each system.

3.3 Simulation Models and Details

We define $\gamma(r)$ as the sum of the cluster(s) defining a term $b_n(r)$ in the h - h -bond expansion of the bridge function and the reference γ_o as a single ring-ring-cluster of field particles (i.e. no root particles) with the same number of particles as appropriate to $\gamma(r)$. The bonds in γ_o are based on f -bonds for a hard sphere potential of unit diameter, which is the same diameter as the target of interest. Values of $\Gamma(r)$ in equation (3.1) can be collected at different values r in one simulation and the contributions to the averages for $\gamma(r)$ are binned according to the value of r (more information can be found elsewhere [28, 29]). Regarding the probability distribution π and π_o , it is necessary that they are non-negative. In the Mayer-sampling, the choice of π should have the ability to sample configurations belonging to both target and reference systems, as shown in Figure 3.1(a) so that we select the absolute value of $\gamma(r)$. It is clear that the configurations important to γ_o and $\gamma(r)$ are subsets of these sampled by π . Similarly, in the overlap sampling we choose the absolute values of $\gamma(r)$ and γ_o for π and π_o , respectively, and this choice satisfies the phase-space relationship for the overlap sampling as shown in Figure 3.1(b). To obtain $b_2(r)$ and $b_3(r)$, we use the simulation values of $g(r)$ thus $h(r)$, which are from Kolafa *al.*'s work [5, 21] over a wide range of reduced densities. We use a bin size $\Delta r = 0.1$ to obtain $b_n(r)$ and apply Neville's algorithm [88] to interpolate between

points of $b_n(r)$ to produce the smooth results with the bin size $\Delta r = 0.01$.

The simulation is conducted in the free space and its box contains 4 or 5 particles including root and field particles for $b_2(r)$ or $b_3(r)$ calculations, respectively. For efficient simulation, we fixed one root particle in the center of the box to bypass the periodic boundary condition. This is acceptable since the other 3 or 4 particles do not move out of the box due to interconnected h -bonds the bridge cluster within the density range of interest (i.e. 0.2~0.8). In the Mayer-sampling, we found it is helpful to first select the other root particle to undergo one TMMC trial, and next field particles to undergo MC trial(s). In the overlap sampling, we select a random number of particles including the other root particle to perform MC trials. The results presented in this work are based two independent simulations with each collection of $5 \times 10^9 \sim 10^{10}$ configurations excluding initialization steps. In general, the value of the optimization λ is in order 1 at density 0.2 and less than 10 for the rest except $b_3(r)$ (i.e. ~ 66) density 0.8. To fix the number of macrostates, the maximum separation distance of two root particles is determined prior to the TMMC Mayer-sampling simulation; a reasonable distance can be obtained through a preliminary simulation by using the MC Mayer-sampling. The root particles are found to move less than 4σ at density 0.2, 5σ for 0.5 and 8σ for 0.8, which resulted in to 80 macrostates, respectively. Since the moves of two root particles are from a zero separation distance, the weight function is developed from the distance to the entire region.

3.4 Results and Discussion

Table 3.1. Standard deviations of short MC Mayer-sampling, TMMC Mayer-sampling and overlap sampling simulations against long MC Mayer-sampling simulation for $b_2(r)$ at densities 0.2, 0.5, and 0.8, respectively. Note that the reference values (long MC Mayer-sampling) were evaluated from 5×10^{10} configurations for $b_2(r)$ at all densities.

		$b_2(r)$		
		0.2	0.5	0.8
Mayer-sampling	$\sigma_{\text{total}}^{\text{a}}$	0.1390	0.2688	0.9393
	$\sigma_{\text{s}}^{\text{b}}$	0.1319	0.2505	0.8571
	$\sigma_{\text{m}}^{\text{c}}$	0.0071	0.0182	0.0785
	$\sigma_{\text{l}}^{\text{d}}$	8.15×10^{-7}	5.29×10^{-5}	0.0027
TMMC Mayer-sampling	σ_{total}	0.1696	0.2642	0.7534
	σ_{s}	0.1588	0.2420	0.5297
	σ_{m}	0.0108	0.0222	0.2222
	σ_{l}	1.92×10^{-7}	1.81×10^{-5}	0.0015
Overlap sampling	σ_{total}	0.0869	0.1385	0.9361
	σ_{s}	0.0825	0.1210	0.8301
	σ_{m}	0.0044	0.0174	0.1035
	σ_{l}	8.77×10^{-7}	6.97×10^{-5}	0.0024

a. total represents the total standard deviation over entire range of r .

b. s represents separated deviation for short distance ($0 \leq r < 1$).

c. m represents separated deviation for moderate distance ($1 \leq r < 3.5$).

d. 1 represents separated deviation for long distance($r \geq 3.5$).

Table 3.2. Standard deviations of short MC Mayer-sampling, TMMC Mayer-sampling and overlap sampling simulations against long MC Mayer-sampling simulation for $b_3(r)$ at densities 0.2, 0.5, and 0.8, respectively. Note that the reference values (long MC Mayer-sampling) were evaluated from 5×10^{10} configurations for $b_3(r)$ at density 0.2 and 10^{11} configurations at densities 0.5 and 0.8. The notations are the same as in Table 3.1.

		$b_3(r)$		
		0.2	0.5	0.8
Mayer-sampling	$\sigma_{\text{total}}^{\text{a}}$	1.0842	1.7054	17.277
	$\sigma_{\text{s}}^{\text{b}}$	1.0368	1.5710	16.170
	$\sigma_{\text{m}}^{\text{c}}$	0.0474	0.1340	1.0838
	$\sigma_{\text{l}}^{\text{d}}$	7.18×10^{-6}	5.11×10^{-4}	0.0237
TMMC Mayer-sampling	σ_{total}	1.5026	1.4777	9.2456
	σ_{s}	1.4408	1.3332	7.4326
	σ_{m}	0.0618	0.1444	1.7998
	σ_{l}	1.23×10^{-6}	1.72×10^{-4}	0.0132
Overlap sampling	σ_{total}	0.5749	1.6326	14.557
	σ_{s}	0.5392	1.5177	12.900
	σ_{m}	0.0358	0.1145	1.6314
	σ_{l}	8.48×10^{-6}	4.40×10^{-4}	0.0254

Since there exist little data for bridge coefficients calculated with $h(r)$, we

collected a great many configurations through long MC Mayer-sampling simulations (i.e. 5×10^{10} configurations for $b_2(r)$ at all densities and $b_3(r)$ at density 0.2, and 10^{11} configurations for $b_3(r)$ at densities 0.5 and 0.8) to check precision of the results from the TMMC Mayer-sampling and the overlap sampling compared with the short MC Mayer-sampling. From computed results of $b_2(r)$'s from the short MC Mayer-sampling, the TMMC Mayer-sampling, and the overlap sampling simulations, we obtained the very small differences since there is only one cluster diagram to be evaluated with relatively fewer configurations. Thus, this discussion is focused more on $b_3(r)$. Nevertheless, the detailed comparison for $b_2(r)$'s is presented in Table 3.1.

Typical development of the weight function in the TMMC Mayer-sampling is shown in Figure 3.2, where the estimates of the weight functions at different MC steps are presented, along with the comparison of configuration collected by using TMMC and MC Mayer-samplings. While most fall in the moderate range by the MC Mayer-sampling simulation, the TMMC Mayer-sampling technique enables the two root particles to move over the range so that one can produce equally well-distributed number of configurations. Thus, the TMMC Mayer-sampling can accelerate the convergence of bridge coefficients at core and tail regions; from Table 3.1 and Table 3.2, one can see that the standard deviation is generally large in the core region and the TMMC Mayer-sampling produces more precise values than other methods. The procedure to incorporate the TMMC Mayer-sampling method into a simulation initiated by setting $\ln \Pi(0) = 0$, which is subsequently updated with using equations (3.2) to (3.6) during simulation. At density 0.8 for $b_3(r)$, the weight

function is developed at shorter range of r , rapidly spread over the whole range after 2×10^6 MC steps, and is converged at 2×10^7 MC steps. After a reasonable estimate of the weight function is obtained, all macrostates are sampled with roughly equal probability as shown by the upper-left plot in Figure 3.2. The discrepancy of the weight function between 2×10^7 and 2×10^9 MC steps is very small. Since the analysis shown here is done for relatively high density in fluid regime, it is anticipated that the convergence of the weight function becomes more rapid as the order of bridge coefficient and the density decrease.

Figure 3.2. Development of the weight function for $b_3(r)$ at density 0.8 in the TMMC Mayer-sampling. The different symbols represent estimates of the weight function after different numbers of MC steps; diamond, circle, triangle and square represent the weight function after 20, 200, 2000 and 200000×10^4

MC steps, respectively. The upper-left plot indicates the typical configuration distributions for the MC Mayer-sampling and the TMMC Mayer-sampling, labeled by square and circle symbols, respectively.

In the overlap sampling, it was found that the numerator expressed in equation (3.8) is the same form as equation (3.1) by replacing γ_o by γ_{os} ; we evaluate $\gamma(r)$ from γ_{os} instead of γ_o . As mentioned above, the ring structure γ_o does not contain root particles and is built with f -bond, of which the relative distance of two particles is restricted within 1σ , while the target system has two root particles without bond and it can generally move much further than 1σ . the difference between $\gamma(r)$ and γ_o is observable in the Mayer-sampling simulation. We reduce this difference by introducing the overlap function γ_{os} . The overlap sampling is performed in two systems by setting two different probability distribution functions π and π_o . From Figure 3.3, we can see that π_o is independent on the density and restricted within 1σ if we choose two particles in the ring structure same as root particles in $\gamma(r)$, while π depends on the density and it spreads out as the density increases. In the Mayer-sampling, π is applied to evaluate both $\gamma(r)$ and γ_o , in which most configurations are beyond 1σ and γ_o is zero (i.e. not existed) if the distance of two particles is apart larger than 1σ , therefore the value of $\langle \gamma_o / \pi \rangle_\pi$ in the Mayer-sampling may not be precise to get $\Gamma(r)$. On the other hand, in the overlap sampling, we applied π to evaluate $\gamma(r)$ and π_o for γ_o separately, and connect them by the overlap function γ_{os} . While considering the configuration space, the covered range of the reference system is show by vertical lines in Figure 3.3. They overlap with the

configuration space of the target system over the range below 1σ . The overlapped area is large at low density and becomes smaller as the density increases. In the systems of interest, it was found that the overlap sampling bears higher precision than the umbrella sampling if the overlapping region is large. Thus, it produced more precise results compared to the short MC Mayer-sampling at densities 0.2 and 0.5 but comparable at density 0.8 as shown in Table 3.1 and Table 3.2.

Figure 3.3. Configuration distributions of reference and target systems in overlap sampling for $b_3(r)$. The curves are normalized from 2×10^7 configurations over 8σ with the bin width as 0.1σ .

Figure 3.4 to Figure 3.6 show the comparing results for $b_3(r)$'s of aforementioned three methods for moderate, short, and long ranges of r ,

respectively. The results are generally consistent with each other at low density 0.2 and moderate density 0.5, however noticeable discrepancy is observed at high density 0.8.

Figure 3.4. Plot of the bridge coefficient $b_3(r)$ vs. reduced moderate distance. Solid line is the reference data from the long MC Mayer-sampling simulation, and dash, dot, and dot-dash lines represent the results calculated from the short MC Mayer-sampling, the TMMC Mayer-sampling, and the overlap sampling, respectively. The reference data was evaluated from 5×10^{10} configurations at density 0.2 and 10^{11} configurations at densities 0.5 and 0.8.

As shown in Figure 3.4, the calculated bridge coefficients at moderate distance considerably differ at the first peak and the saddle area between peaks. Over this range, the short MC Mayer-sampling produces closer results to the

MC Mayer-sampling results since the MC Mayer-sampling produces more configurations than other methods in this range. However, as shown in Figure 3.5, $b_3(r)$'s are almost the same except for values near zero separation distance. We notice that the TMMC Mayer-sampling produces the best result and the overlap sampling is better than the short MC Mayer-sampling at close to zero separation as shown in inserted plot in Figure 3.5. Lee [25] has shown that the excess chemical potential for hard spheres system is essentially made up (i.e. up to 98%) by contributions from the correlation functions inside the core (i.e. especially $b(r)$ when $r < \sigma$). Therefore, we expect that the TMMC Mayer-sampling can provide more accurate value of the correlation function via fast evaluation of $b_n(r)$ in the core.

Figure 3.5. Plot of bridge coefficient $b_3(r)$ vs. reduced short distance. The same

notations as in Figure 3.4 are applied. The lower-right plot shows $b_3(r)$'s within range 0.01σ , note that values at density 0.2 are added by 2.0 for clarity.

Figure 3.6. Plot of bridge coefficient $b_3(r)$ vs. reduced long distance. The same notations as in Figure 3.4 are applied.

Figure 3.6 shows $b_3(r)$ over the tail region. All three methods did not produce very good results at high density 0.8. Even the reference data, which obtained from collecting 10^{11} configurations, shows severe oscillations since number of probable bridge configurations become very high at long distance range at high density. Nevertheless, the results from collecting 2×10^{10} configurations show that the TMMC Mayer-sampling is the best and the sampling is comparable with the short MC Mayer-sampling. To clarify our claims, we provide the results of the overall convergence in terms of standard

deviation in Table 3.1 and Table 3.2. The standard deviation $\sigma(r)$ is calculated

by using the formula $\sigma(r) = \sqrt{\sum_{i=1}^2 (x_i(r) - x(r))^2}$ for each distance r , where $x(r)$

represents reference data and $x_i(r)$ is the value from two simulations with each run for 5×10^9 MC steps for $b_2(r)$ at all densities and $b_3(r)$ at density 0.2, while using 10^{10} MC steps for $b_3(r)$ at densities 0.5 and 0.8, respectively. At low to moderate density, the overlap sampling shows the best precision, and at high density, the TMMC Mayer-sampling shows the best. We found that the overlap sampling generally yields better convergence than the short MC Mayer-over low density range, but as the density increases, the TMMC Mayer-shows the better precision.

Finally, we examine the pressure as a function of the density obtained from the total correlation function $h(r)$ via the virial equation. The calculation of $h(r)$ performed as the same way as Kwak and Kofke [29] performed, starting from pure HNC solution for $h(r)$ using Picard iteration algorithm presented by Duh and Haymet [57] to estimate the $b(r)$ including only $b_2(r)$ term, which is evaluated by the TMMC Mayer-sampling and the overlap sampling, This $b(r)$ is then used in another Duh-Haymet calculation of $h(r)$ and this is in a new calculation of $b_2(r)$. This $b_2(r)$ is considered as the converged value is used in the next iteration process to calculate $b_3(r)$. Another two calculations $b_3(r)$ are performed and used in the Dub-Haymet algorithm to get the final $h(r)$, which is calculated from HNC with $b(r) = b_2(r)\rho^2 + b_3(r)\rho^3$. Figure 3.7 shows our results compared to the accurate Carnahan-Starling equation of state [89] Kwak and Kofke's results. It is well known that the pure HNC virial pressure

overestimates the actual pressure and this discrepancy can be eliminated by adding higher order terms in the bridge function expansion [29] as shown in Figure 3.7. At low and moderate densities, there is little difference among pressures obtained from $h(r)$ as a solution of the OZ equation with the bridge function closure calculated from the MC Mayer-sampling, the TMMC Mayer-Mayer-sampling, and the overlap sampling. However, at high density, slight difference can be seen for the virial pressures obtained from the three methods; the overlap sampling yields the best result and the TMMC Mayer-sampling yields a better one than the MC Mayer-sampling.

Figure 3.7. Plot of pressure versus reduced density. Solid line is the equation of state of Carnahan and Starling. Subscript v stands for ‘virial’ pressure. Dashed line are from pure HNC theory, and square, triangle and cycle symbols are from solution of the OZ equation with a bridge function closure obtained from the h -

h -bond cluster expansion to the second order in density by the MC Mayer-sampling, the TMMC Mayer-sampling and the overlap sampling, respectively. Note here, the lines from reference [29] are guides to an eye and triangle and cycle symbols at densities 0.2, 0.5 and 0.8 are calculated from $h(r)$ via the TMMC Mayer-sampling and the overlap sampling, respectively by collecting $5\sim 10\times 10^9$ configurations

3.5 Conclusion

In this chapter, we have presented modified methods, the TMMC Mayer-sampling and the overlap sampling, to determine the bridge coefficients (i.e. 2nd and 3rd) that can be represented as bridge cluster diagrams. The TMMC scheme bears the ability for a pair of particles to pass through low-probability regions to reach extreme separation distances and evenly sample an entire It has been found that the Mayer-sampling based on the TMMC method shows fast convergence of bridge coefficients over a wide range of distance except moderate distance region and better precision in core and tail regions of the bridge coefficients. We also demonstrated the overlap sampling method, which has been applied previously only to cluster(s) relevant to virial coefficients (one root point), can be applied to calculate bridge coefficients (two root points). Its reliability and precision have been tested and the method improved the of the bridge coefficients over the densities of interest compared to the generic MC Mayer-sampling. The efficiency of this method is related to the degree of overlapping region between the target and reference systems, therefore we expect that a better reference system rather than the f -bond ring cluster can be found to improve the efficiency. We performed the calculation, in which the h -

h -bond ring cluster is used instead of the f -bond ring cluster as the reference system to increase the overlapping region. In this case, the result was more precise than that of the Mayer-sampling using the h -bond ring cluster and those are not reported here. It is anticipated that as the density and the order of the bridge coefficient increase, the modified methods presented in this work would be more suitable to produce reliable data than the generic MC Mayer-sampling method. Lastly, our approach can be extended to any model systems with relatively simple modifications.

Chapter 4

Singlet IE Theory of the Hard Sphere Fluid Confined in a Slit-Pore

4.1 Introduction

Theoretical study of confined fluids by statistical mechanical methods is known to be practical for evaluating thermophysical properties of bulk fluids also important for understanding various problems related to surface such as adsorption and capillary condensation [80]. Recently progress has been made studying confined fluids using two approaches. One is the IE theory [51, 90-92] and the other is based on molecular modeling and simulations [93-99]. In the IE approach both structural and thermodynamic properties can be determined using a singlet level [33-36] or a pair level theory [37-40]. The shortcomings of the latter are considerable, demanding more computational resources than the singlet level one. Additionally, the pair level theory does not significantly improve fluid properties against the singlet level one in many cases [35, 39], such as for fluids inside a micropore, which is one of the important models of confined fluids [90]. Many singlet IE theories have been investigated including the PY/PY and the HNC/HNC equations for hard sphere fluids inside spherical and slit pores [51, 52], the HNC/MSA (mean spherical approximation) equation for hard sphere [53] and ionic systems [100, 101] a slit-pore, and the HNC/HNC equation for Yukawa fluids [54] inside a slit-pore. Although theoretical results of the singlet IE theory are generally in agreement with those from simulations, it shows some limitations. When applied to a

confined fluid, the singlet PY/PY and HNC/HNC theories predict incorrect contact values of the density profiles, consequently, those of the normal [51, 52].

It is well known that the primary source of inaccuracy in the implementations of the IE theory for a bulk fluid is due to the uncertainty of the bridge function, which can be expressed as an infinite sum of highly connected bridge diagrams [38]. This is also true for the singlet IE theory of the confined fluid. The bridge functions involving the fluid confinement cannot be easily evaluated or even formulated. However, direct evaluation of the confined-fluid bridge function is possible if one can find an easy way to evaluate bridge coefficients (i.e. diagrams, see Figure 1.5). Note that the construction theorem of the bridge diagrams would be the same for bulk and slit-pore since the interaction between fluid and wall can be decoupled from the interaction between fluid and fluid. A system of hard spheres inside a slit-pore in equilibrium with a bulk hard sphere fluid is an efficient checking ground due to theoretical tractability. Therefore, this work deals with accuracy and applicability of the confined-fluid bridge function for the singlet IE theory to predict the local structure of the hard sphere fluid in a slit-pore. In this chapter, the bridge function is formulated by two bridge coefficients that are determined by the TMMC [102] Mayer-sampling [28, 29] method, which was recently developed in Chapter 3. The quality of the performance of our approach has been assessed by comparing the results from GCMC [94, 103] simulation with those from this method.

4.2 Theoretical Approach

4.2.1 Singlet Integral Equation Theory

Originally the singlet IE theory was derived by Henderson *et al.* [33] and extended by Zhou and Stell [51] to study the hard sphere fluid inside a pore and later by others [52-54] to study the fluid confined in a slit-pore. Starting from the OZ equation for a multi-components fluid (i.e. equation (1.14)) introduced in Chapter 1, we obtain the OZ equation for a binary homogeneous mixture shown as follows,

$$\begin{aligned} h_{11}(r) &= c_{11}(r) + \rho x_1 \int c_{11}(|r-r'|) h_{11}(r') dr' + \rho x_2 \int c_{12}(|r-r'|) h_{21}(r') dr' \\ h_{12}(r) &= c_{12}(r) + \rho x_1 \int c_{11}(|r-r'|) h_{12}(r') dr' + \rho x_2 \int c_{12}(|r-r'|) h_{22}(r') dr' \\ h_{22}(r) &= c_{22}(r) + \rho x_1 \int c_{21}(|r-r'|) h_{12}(r') dr' + \rho x_2 \int c_{22}(|r-r'|) h_{22}(r') dr' \end{aligned} \quad (4.1)$$

where subscripts 1, 2 represent the quantities for species 1 and 2, ρ is the homogeneous density of bulk fluid, x_i is the concentration of species i and h and c are the total correlation functions, respectively. Under the assumption that the size of species 2 grows to infinite $\sigma_2 \rightarrow \infty$ and the concentration approaches zero $x_2 \rightarrow 0$, the set of above equation decouples to the OZ equation for the bulk-bulk interaction (i.e. equation (1.5)),

$$h_b(r) = c_b(r) + \rho_b \int c_b(|r-r'|) h_b(r') dr' \quad (4.2)$$

and a singlet OZ equation for the bulk-wall interaction,

$$h(z) = c(z) + \rho_b \int c_b(|r-r'|) h(z') dr' \quad (4.3)$$

where ρ_b is the reduced density of the bulk fluid, c_b is direct correlation function of the bulk fluid, which is independently given by an appropriate theory for the bulk fluid, $h(z)$ and $c(z)$ are wall-particle total and direct correlation functions, respectively. Here z represents the distance of the particle from the wall. The wall-particle distribution function $g(z)$ is given by $g(z) = h(z) + 1$ and the prime quantity in the confined fluids of interest is the density profile $\rho(z)$, which can be calculated by

$$\rho(z) = \rho_b g(z) \quad (4.4)$$

To solve the singlet OZ equation above, another relationship between $h(z)$ and $c(z)$ is needed. The suitable choice is in the form of closure,

$$h(z) = \exp[-\beta u_{wp}(z) + h(z) - c(z) + b_s(z)] - 1 \quad (4.5)$$

where the $u_{wp}(z)$ is the wall-particle potential interaction and $b_s(z)$ is the slit-fluid bridge function.

4.2.2 Slit-Fluid Bridge Function

The integral equations (4.3) and (4.5) are exact, yet the slit-fluid bridge function is unknown, thus all present implementations are approximate by necessity. Recalling that the bulk-fluid bridge function can be expressed as a series in the density expansion (i.e. equation (1.13)) and the first two orders of bridge diagrams constructed with h -bonds are shown in Figure 1.5. In this subsection, we will obtain the slit-fluid bridge function from these diagrams. In the singlet OZ equation, the total correlation function $h(z)$ between the wall and fluid particle at a distance z is the sum of the direct correlation $c(z)$ between

and the indirect interactions (i.e. the second term in the RHS of equation (4.3)) from the other fluid particle exerted on them; the slit-fluid bridge function include the interactions between the wall and fluid particle at a distance z . we can construct the wall-particle bridge diagrams by taking the second order bridge coefficient $b_{2(wp)}$ as an example, as shown in Figure 4.1, one of the root particles (i.e. white circles) labeled by 1 is fixed at one of the wall surfaces and the other root particle labeled by 2 moves perpendicularly to the wall in z direction, the field particles (i.e. black circles) inside the slit-pore can move with periodic boundaries in x and y directions and the bonds represent the total correlation functions of the bulk fluid, which are $h_b(r)$. Note here L is the slit width and the walls are located at $z = \pm L/2$.

Figure 4.1. Schematic of the wall-particle bridge diagram in the second order slit-fluid bridge coefficient. The open circles represent two root particles (labeled by 1 and 2) and the solid circles denote the field points (labeled by 3 and 4), which are subjected to be integrated over the connecting h -bonds.

Since the slit pore is symmetric, the slit-fluid bridge function can be represented by

$$b_s(|z|) = b_{wp}[L/2 + |z|] + b_{wp}[L/2 - |z|] \quad (4.6)$$

Note that the first term in the RHS represents the effect of the upper wall and the second term for the lower wall as shown in Figure 4.1.

4.2.3 TMMC Mayer-Sampling

The wall-particle bridge diagrams can be evaluated by the TMMC Mayer-sampling [102] method proposed in previous chapter. This method is developed from the general MC Mayer-sampling method [28, 29] and the governing sampling involved is based on the magnitude of the interactions with the TMMC weight function η , for which a working equation (i.e. equation (3.1)) can be expressed as,

$$\Gamma(z) = \Gamma_o \frac{\langle \gamma(z) / e^{\eta(N)} \pi \rangle_{e^\eta \pi}}{\langle \gamma_o / e^{\eta(N)} \pi \rangle_{e^\eta \pi}} \quad (4.7)$$

where $\Gamma(z)$ is the value of a desired cluster integral or sum of integrals in a relative distance z of two root points and $\gamma(z)$ is the value of integrand (or sum integrands). The subscript $e^\eta \pi$ is the probability distribution that governs the sampling of all configurations and the ensemble average over all configurations is indicated by the angle brackets. The subscript “o” refers to the reference system, for which Γ_o is known. The N stands for the quantity for macrostate and is defined by regularly partitioned distances in the relative distance z of two

particles, which is equal to the total number of bins relevant to z . The TMMC weight function η is introduced to ensure that all ranges are sampled with sufficient frequency and is given by $\eta(z) = -\ln \Pi(z)$. To obtain η , a TMMC scheme is employed and more details can be found in Chapter 3.

4.3 Simulation Models and Details

For the hard sphere fluid inside a slit-pore, which is formed by two planar walls (i.e. the hard sphere-wall model introduced in Chapter 2), the pair interaction potential of hard spheres at a distance r is calculated as same as equation (2.25) and the interaction potential between a planar wall and a hard sphere at a distance z is calculated as same as equation (2.28).

We introduced $\gamma(z)$ as the sum of the diagram(s) defining a term $b_n(z)$ of slit-fluid bridge function $b_s(z)$. Further, we chose the reference γ_o as a single ring-cluster of field particles (i.e. no root particles) with the same number of particles as appropriate to $\gamma(z)$. The interaction bonds in γ_o are based on the Mayer function of a hard sphere potential of unit diameter, which is the same diameter as the target system of interest. The value of h_b -bond in $\gamma(r)$ and $c_b(r)$ in the singlet OZ equation is obtained from the bulk-fluid OZ equation with the HNC closure with the cut-off length of the bond r as 10.24σ and $\Delta r=0.005$, indicate the bin width and the total number of bins is 2048. Regarding the probability distribution π and π_o , they are necessarily non-negative and we them as the absolute values of $\gamma(z)$ and γ_o , respectively, which consequently simulate the umbrella sampling. To obtain η , the distance z between two root points are regularly partitioned and the index follows the number of bins. Since

natural discretization of z does not exist, a bin width is set to $\delta z = 0.025\sigma$ to capture the probability distribution by TMMC moves that restrict the new separation distance to the random range within $z-\delta z$ and $z+\delta z$ given the current distance z . Values of $\Gamma(z)$ in equation (4.7) can be collected at different values z in one simulation and the contributions to the averages for $\gamma(z)$ are binned according to the value of z . Reported values of $\gamma(z)$ are

$$\gamma(z_i) = \frac{1}{LM} \sum_{k=1}^{m_i} \gamma_{i,k} \quad (4.8)$$

where m_i is the number of times where the movable root particle was observed in a distance within z_i to $z_i+\delta z$, $\gamma_{i,k}$ is the k^{th} contribution in the bin i to the average, $M = \sum m_i$ is the total number of contributions made to all bins, and $L = \frac{4}{3}\pi(\Delta z)$ is the length of the bin i with weight. Here the weight is due to the way we generate MC steps, if the root particles can move freely in 3D, then the weight is $\frac{4}{3}\pi[(r+\Delta r)^3 - r^3]$, if the root particles move in a 2D planar, then the weight is $\frac{4}{3}\pi[(r+\Delta r)^2 - r^2]$, else if the root particles move in a 1D line, then the weight is $\frac{4}{3}\pi(\Delta r) = \frac{4}{3}\pi(\Delta z)$. These weights have been tested by calculating the bulk-fluid bridge function using different root particle movements and the results are the same with employing the weights.

We performed simulations for two slit sizes 3.0σ and 4.0σ , for three densities $\rho_b\sigma^3 = 0.3, 0.5$ and 0.7 . The number of configurations generated is

for $b_2(z)$ and 2×10^{10} for $b_3(z)$ and update the weight functions at every 10^5 MC steps. We applied Neville's algorithm [88] to interpolate between points of $b_n(z)$ to produce the smooth results with the bin size $\Delta z = 0.005$. The calculated wall-wall-particle correlation functions are then used to solve the slit-fluid OZ equation using Labik's algorithm [52] to determine the structure of the fluid in the pore. To test the theory, we will compare our results with GCMC simulation data. It is convenient to perform the GCMC for confined fluids by fixing the fluid chemical potential at certain value and inserting and deleting hard sphere particles until the chemical potential of the slit pore fluid is equal to that of bulk fluid at the same density. The value of bulk fluid chemical potential is obtained from Labik and Smith's work [94]. During the GCMC simulation, the displacement moves for the hard spheres were about 70% and the additions or deletions of hard spheres were 15% each. After an equilibration cycles of 10^9 to 2×10^9 , the final results were collected by 10^{10} to 2×10^{10} cycles depending on density. The density profiles of the slit pore fluid were determined by counting the particles located within slabs parallel to the walls.

4.4 Results and Discussion

The top and middle plots in Figure 4.2 show the behaviors for $b_2(z)$ and $b_3(z)$ at slit widths $L=3.0\sigma$ and 4.0σ respectively at densities 0.3, 0.5 and 0.7, respectively. For the slit width 3.0σ , $b_2(z)$ shows a single maximum at the center for all densities and minimum at the edges and this behaviour is strengthened as the density increases. While $b_3(z)$ shows a minimum at the center and maximum at the edges for low density 0.3. As the density increases to 0.5 and 0.7, single peaks at the centre appear, but the maximum still appears at the edges.

for the slit width 4.0σ , $b_2(z)$ shows a single peak at the center at low densities and 0.5, then become lower and splits into peaks, if the density is further increased as shown in Figure 4.2. In the same way, we can see $b_3(z)$ shows two minima at vicinities of the edges with a maximum at the center at low density followed by the minimum plumps up and the maximum flattens as the density increases.

Figure 4.2. Plots of two slit-fluid bridge coefficients $b_2(z)$, $b_3(z)$ and a bridge

function $b_s(z)$ versus the reduced distance from the wall surface for $L=3.0\sigma$ and $L=4.0\sigma$, respectively. The solid, dash and dash-dot lines represent the values at density 0.3, 0.5 and 0.7, respectively.

Comparing $b_2(z)$ and $b_3(z)$ at the same slit width, the cancellation of the odd and even coefficients is observed especially around the edges, and we speculate that this behaviour may remain the same, which results in the odd and even functions affecting the correlation function such that $b_s(z)$ oscillates around 0; the interactions of the bridge function are all negative, yet as pore increases, oscillation around 0 is observed in the bulk fluid. We know that the value of bridge coefficients $b_2(r)$ and $b_3(r)$ for the bulk fluids are always negative at the small distance region and values increase as the relative distance r increases, then oscillates around 0 at the middle distance region. Therefore, the cancellation of the odd and even coefficients is not observed for the bulk fluid at the small regions, this differs from the confined fluid, and we speculate that the cancellation of the odd and even coefficients around the edges for the confined fluid is due to the influence of the walls [29].

We also obtain the slit-fluid bridge function $b_s(z)$ containing $b_2(z)$ and $b_3(z)$ terms as shown in bottom plots in Figure 4.2. The bridge function containing $b_2(z)$ and $b_3(z)$ shows a similar behaviour as $b_2(z)$ when the density increases since $b_2(z)$ is a dominant term in the bridge function under low densities. The value of the bridge function is negative near the wall surface and closer to zero at the central region, and the behaviour is changed as pore size varies. When the pore size is 3.0σ , all the values are negative at all densities of interest. The value of the function oscillates around 0 as the pore size increase to 4.0σ at density 0.7, but

negative for other densities.

Figure 4.3. Density profiles of hard sphere fluids within $L=3.0\sigma$. The slit pore fluid is in equilibrium with a bulk fluid of the reduced density at 0.3, 0.5 and respectively. The dot, dash-dot and solid lines represent results of the GCMC simulation, the HNC closure and the HNC closure with $b_s(z)=$ respectively.

Figure 4.3 and Figure 4.4 show the density profiles as the functions of the particle-wall separation distance at densities 0.3, 0.5 and 0.7, of slit widths 3.0σ and 4.0σ , respectively. Generally, the good agreements between the theoretical results (i.e. HNC and HNC with $b_s(z)$) and the GCMC simulation data are observed except at the contact values of the pure HNC density profiles, which overestimate their corresponding GCMC values for all densities. These

discrepancies are reduced with the inclusion of the bridge function proposed in this chapter as shown in the figures. Additionally, the qualitative behavior is predicted correctly by the HNC closure with $b_s(z) = b_2(z)\rho_b^2 + b_3(z)\rho_b^3$ and the quantitative agreement with the GCMC simulation data is superior to the pure HNC closure especially at the center and the edges. All above results implicitly represent that the slit-fluid bridge function provides a way to show a correct behavior of the density profile and as more coefficients are included in the function, a better prediction of the GCMC data is expected at moderate range.

Figure 4.4. Density profiles of hard sphere fluids within $L=4.0\sigma$. The symbols are the same as shown in Figure 4.3.

We also obtain the reduced pressures p_w^* normal to the wall, which is equal to the values of the wall-particle density profiles at contact [51].

$$p_w^* = \beta p_w = \rho[(L - \sigma)/2] = \rho_b g[(L - \sigma)/2] \quad (4.9)$$

The results are compared with the GCMC values from literature data [52] in Table 4.1 at slit widths 3.0σ and 4.0σ with densities at 0.3, 0.5 and 0.7, respectively. The pressures from the pure HNC are higher than the GCMC results at low density 0.3 and medium one 0.5, but a bit higher at density 0.7. These discrepancies may occur due to the truncation of the bridge coefficients in the density expansion. We expected that if one or two more bridge coefficients are included in equation (1.13), more accurate density profiles and normal pressures can be obtained with these pore ranges at densities of fluid phase.

Table 4.1. The reduced pressures normal to the wall (p_w^*) at slit widths 3.0σ and 4.0σ for reduced bulk densities $\rho_b\sigma^3=0.3, 0.5$ and 0.7 .

$\rho_b\sigma^3$	GCMC*	HNC	HNC with $b_s(z) = b_2\rho_b^2 + b_3\rho_b^3$
$L/\sigma=3.00$			
0.3	0.58 ± 0.01	0.63	0.59
0.5	1.64 ± 0.02	1.93	1.66
0.7	3.92 ± 0.05	5.33	4.56
$L/\sigma=4.00$			
0.3	0.59 ± 0.01	0.64	0.60
0.5	1.64 ± 0.02	1.90	1.64
0.7	4.15 ± 0.04	5.04	4.31

4.5 Conclusion

In this chapter, we constructed the bridge function of the hard sphere fluid

inside a slit-pore from the wall-particle bridge diagrams, which were evaluated by using previously proposed TMMC Mayer-sampling. The slit-fluid bridge function's second order coefficient $b_2(z)$ shows a cancellation with the third coefficient $b_3(z)$ especially around the edges of the walls. The slit-fluid bridge function's even coefficients were cancelled with the odd ones to some extent especially around edges, compared with the bulk-fluid bridge coefficients, the cancellation does not happen at the small distance region, thus we speculate that the existence of the confining walls in fact shows a strong influence on that particular cancellation of the bridge coefficients. A comparison shows a very good agreement between the HNC closure with the slit-fluid bridge function the GCMC result; the slit-fluid bridge function studied in this chapter can be the pure HNC under the slit confinement so that more accurate density profiles can be obtained. We also calculated the normal pressures and the HNC closure with the slit-fluid bridge function correctly predicts those at low and medium densities, while a little larger at high density was obtained compared to the GCMC simulation result. We expect to reduce the discrepancy by adding higher order bridge coefficients. These results together with the density profile comparison show that the direct bridge function evaluation can in fact improve the quality of the singlet IE theory for the fluid confined in a narrow slit-pore to predict its structural and thermodynamic properties. The method presented in chapter is not restricted to the study of hard sphere systems but can be easily extended to study any single component systems; one can expect to use the method for studying unified-atom or coarse-grained systems of complex molecules (i.e. drug particulates, monomers, caged molecules, etc.) confined in planar walls by relatively simple modification on the total correlation function.

Thus, this approach can reduce tremendously the computational recourses needed to run such large-scale simulations. Nevertheless, more insightful investigation might be useful to apply the present method to reveal the first-phase transition, which is an inherent problem continuously challenged by the theory.

PART II: Human Beta Defensin 28

Chapter 5

Background

5.1 Introduction

Antimicrobial peptides (AMPs) contain sets of small molecules that contribute to host defense in multicellular organisms [104]. Among these peptides, defensins that contain six invariant cysteine residues are of particular interest as they exhibit a broad spectrum of antimicrobial activity against Gram-positive and Gram-negative bacteria, yeast, fungi and some enveloped viruses [105, 106]. The human β -defensins (HBDs) are members of the defensins family and deserve a special attention due to their prominence in potent antimicrobial activity and as a chemo-attractant, thereby playing a key role in innate and adaptive immunity [107]. Their great potentials are found in developing ideal therapeutic drugs as antibiotics and modulators of inflammation; however, the less knowledge has been developed for this HBDs subfamily to further study the structural effect on their antimicrobial property; the interaction mechanisms involving membrane depolarization and permeabilization are still a matter of dispute. Therefore, much effort are to investigate the membrane interactions of HBDs. In order to investigate these interactions, insight into three-dimensional (3D) structural information about HBDs is of critical importance.

To date, the predicted peptide sequence is available for over 40 HBDs, and only six HBDs (HBD-1 to HBD-6) have been characterized and only three

structures of them (1 through 3) have been determined [108]. Despite low acid sequence conservation among HBD-1 to HBD-3, they share a similar tertiary structure containing a triple-stranded, antiparallel β -sheet with a short *N*-terminal α -helical region. The β -sheet constrained by three disulfide bridges constitutes the core of defensin and is suggested as the central to the defensin's structure, and presumably, its function. The high cationicity and ability to dimerize have been considered to contribute to the broad spectrum and salt-salt-insensitive behavior of HBD-3 [109]. However, it is also postulated that the hydrophobicity and cationicity are more important for good antimicrobial behavior than the ability to oligomerize [110]. There is no universal rule to suggest a fixed structure-function relationship for HBDs and remains an active area of research. To better understand the relationships between the structure of HBDs and their multiple biological properties, functions, and their potential for pharmaceutical applications, it should be necessary antecedent to determine 3D structures of HBDs.

NMR data indicates that the HBDs can form dimers or higher-order oligomers in solution by the interactions between residues on the first or second β -strand. Recent studies [111, 112] also discover that circular dichroism (CD) spectroscopy shows variations in composition with respect to HBDs, indicating the presence of several polymorphic variants. Therefore, it might be difficult to correlate with the behavior of a pure HBD molecule in the physiological conditions due to the lack of precise structure and its possible conformational heterogeneity in solution. To obtain the 3D structure of protein, NMR or X-ray can be used after a series of purifications or crystallization, but we have not yet

found any related experimental observations in newly discovered HBDs. Therefore, apart from the generic laboratory investigations of HBDs, the research specifically aims to employ protein modeling and simulations as following:

1. Design possible 3D models of the HBDs using homology modeling and validate the homology models using molecular dynamics (MD).
2. Study the conformations changes of HBDs models in different solution environments by molecular simulations.

In order to address these two objectives, we choose the HBD-28 which is a newly identified defensin as our target protein [113]. We describe the modeling methods implemented as well as the comparison among the resulting models based on the HBD-2 and HBD-3 as templates. We use the stabilities of the models after energy minimizations (EMs) and MD simulations, as well as the secondary structure information with respect to consistence with CD data, to probe the validity of the models. Additionally, the configurational changes of HBD-28 models in trifluoroethanol (TFE) /water mixture are investigated.

The second part of this thesis is arranged as following: in the rest of this chapter, we give a brief overview of amino acids, peptides and HBDs regard their characterizations, antimicrobial activities, structures and structure-function relationship, as well as simulation methodologies including homology energy minimization and molecular dynamics. In the next Chapter, we construct initial and refined homology models for the HBD-28 based on HBD-2 and HBD-3 of known structures as templates. Since the 3D structures of homology

models for a protein with low sequence identity to template protein(s) are unavoidably affected with some uncertainties, we investigate extra information during the creation of the 3D models. Therefore the modeling and validation processes are guided by continually testing the model stability and Once possible HBD-28 models are obtained, the configurational changes of HBD-28 models induced by TFE are studied and compared with the CD experimental results.

5.2 Human β -Defensins

5.2.1 Amino Acids and Peptides

An amino acid is a small molecule consisting of an amino group ($-\text{NH}_2$), a carboxyl group ($-\text{COOH}$) and a side-chain ($-\text{R}$) that characterizes an amino acid as shown in Figure 5.1.

Figure 5.1. A generic structure of a peptide, consisting of three amino acids and two peptide bonds.

From the figure, we can see that the general formula of an amino acid is $R-CH(NH_2)COOH$. There are 20 kinds of amino acids residues in living cell which are usually represented by three (or one) letters, as shown in Table 5.1.

Table 5.1. Amino acid full names, three-letter abbreviations (in bold type), and one-letter abbreviations (between brackets).

Alanine	(A)	Cysteine	(C)
Aspartic Acid	(D)	Glutamic	(E)
Phenylalanine	(F)	Glycine	(G)
Histidine	(H)	Isoleucine	(I)
Lysine	(K)	Leucine	(L)
Methionine	(M)	Asparagine	(N)
Proline	(P)	Glutamine	(Q)
Arginine	(R)	Serine	(S)
Threonine	(T)	Valine	(V)
Tryptophan	(W)	Tyrosine	(Y)

A peptide is a short biopolymer consists of amino acids joined together by peptide bonds between the carboxyl and amino groups of adjacent amino acid residues which are the units of a peptide (see Figure 5.1). Usually, a peptide has an amino end (i.e. N-terminus) and a carboxyl end (i.e. C-terminus).

The primary structure of a peptide is its amino acid sequence, which is the order of the amino acid residues as they appeared in a peptide. It is of essential importance in the peptide structure prediction. The secondary structure of a peptide is the characteristic conformation of amino acids in local regions of a

peptide, which is typically maintained by hydrogen bonds. There are two main secondary elements, i.e. α -helix and β -sheet structures. The tertiary structure of peptide is the three-dimensional conformation of all atoms within a peptide, which is largely stabilized by disulfide bonds.

5.2.2 Characterization

AMPs are short, cationic and amphiphilic peptides, consisting 12 to 60 amino acids, which are very important in the innate immune system [114]. An important class of AMPs is the defensin that is rich in sheet structures and cysteine residues. The defensins family can be further classified into three different subfamilies (α , β , θ -defensins) based on the pattern of six conserved cysteine residues forming three disulfide bridges [115]. However, it has been found that only α - and β -defensins are endogenously present in humans.

HBDs are members of the β -defensins subfamily and have attracted the attention of researchers due to their potent antimicrobial activities and important roles in the human immunity system. The first HBD (i.e. HBD-1) was found and isolated from human plasma [116], later HBD-2 and HBD-3 were discovered in the extract of lesional scales from patients suffering from psoriasis and their coding nucleotide sequences also have been identified [105, 106, 110]. With the development of bioinformatics (e.g. biological sequence analysis using 'Profile Hidden Markov Models' and 'Basic Local Alignment Search Tool'), new HBDs are continuously being identified, of which the recent ones are HBD-25 to HBD-29.

HBDs are also small (3-5 kDa), cationic and amphipathic cysteine-rich peptides. The spacing between the cysteine residues and the connectivity of the disulfide bridges in HBDs are different from that of α -defensins. The arrangement found in α -defensins are Cys¹-Cys⁶, Cys²-Cys⁴ and Cys³-Cys⁵ (according to the relative Cys numbering), whereas in HBDs are Cys¹-Cys⁵, Cys²-Cys⁴ and Cys³-Cys⁶ meaning that the first cysteine in the amino acid sequence form a disulfide bond with the fifth, and so on [115]. Although disulfide bonds are important for tertiary structure maintenance of β -defensins, some studies have indicated that disulfide bond is not absolutely necessary for the antimicrobial ability [109, 117, 118].

5.2.3 Antimicrobial activities

Table 5.2. Antimicrobial spectrum of HBDs 1-4. [106, 119-121]

Defensin	Pathogen
HBD-1	E. coli, P. aeruginosa, K. pneumonia, S.aureus
HBD-2	E.coli, P.aeruginosa, K.pneumonia, S.aureus, S.pneumoniae, c.albicans, C.parapsilosis, C.krusei, E.faecalis, HIV-1
HBD-3	E.coli, P.aeruginosa, K.pneumonia, S.aureus, S.pyogenes, E.Faecium, S. pneumonia, S.carnosus, B.cepacia, S.cerevisiae, C.albicans, C.parapsilosis, C.krusei, HIV-1
HBD-4	E.coli, S.carnosus, P.aeruginosa, B.cepacia, S.Pneumoniae, S.aureus, S.cerevisiae
HBD-28	E.coli, P.aeruginosa, K.pneumoniae, S.Pneumoniae, S.aureus
HBD-18	E.coli
HBD-6	E.coli

HBDs have been recognized to exhibit a broad spectrum of antimicrobial activity. Among all HBDs, only the antimicrobial activities of HBD-1 to HBD-4 have been characterized in detail [106, 119-121]. The antimicrobial properties of HBDs are shown in Table 5.2.

5.2.4 Structures

5.2.4.1 Primary Structures

To present, the predicted peptide sequence is available for over 40 HBDs, which typically consists 41 to 50 amino acid residues and is rich in basic and hydrophobic amino acids. It has an identifiable consensus sequence as $X_{2-10}CX_{5-6}(G/A)XCX_{3-4}CX_{9-13}CX_{4-7}CCX_n$ (where X is any amino acid). The sequence alignment of HBD-1, HBD-2 and HBD-3 is shown in Figure 5.2 (a). By comparing the sequences, it is shown that HBD-1 to HBD-3 contain six conserved cysteine residues that form three common disulfide bridges (Cys¹-Cys⁵, Cys²-Cys⁴ and Cys³-Cys⁶), as well as a conserved motif Gly-X-Cys that forms a β -bulge. In addition, the first conserved cysteine residue is situated close to N-terminus and most cationic residues (Lys, Arg) clustered close to the C-terminus, as shown in HBD-1 to HBD-3. It is also reported that high cationicity and its ability to aggregate are important for antimicrobial activity.

Figure 5.2. Structural comparison of human β -defensins. (a) The amino acid sequence alignment of HBD-1, HBD-2 and HBD-3. The extent of the secondary structure elements is shown above each sequence and the numbers in the brackets represent charges. Six conserved cysteine residues are colored in yellow. (b) The ribbon representations of monomeric β -defensins (left) are accompanied by the electrostatic potential maps projected on the solvent-accessible surfaces shown in the equivalent orientations (center) and after 180° rotation (right). The positively and negatively charged areas of the solvent-accessible surfaces are colored in blue and red, respectively. [107]

5.2.4.2 Secondary Structures

The secondary structure of HBDs typically consists of a α -helix and three β -strands (see Figure 5.2 (a)). The α -helix is formed by the N-terminal fragment of the peptide and its length is usually very short and variable. However, it is suggested that the α -helix is not necessary needed in the structure of HBDs due to its small size and adjustable length. The three β -strands are arranged in an antiparallel sheet and stabilized by three intramolecular disulfide bonds (see Figure 5.2 (b)), which have been considered to contribute to the oligomerization of the peptides. For some HBDs, a β -bulge structure is found in the second β -strand, which consists of the conserved motif Gly-X-Cys and contributes to the twist of the β -strands.

5.2.4.3 Tertiary Structures

The tertiary structures have been determined only for three HBDs (i.e. HBD-1 to HBD-3) and their representative 3D structures can be found in the Protein Data Bank (PDB) [122] (i.e. with accession ID: 1E4S for HBD-1 [123], 1FD3 for HBD-2 [124] and 1KJ6 for HBD-3 [125]). Despite the low amino acid sequence conservation, the tertiary structures of HBDs are very similar as shown in Figure 5.2 (b). The core of the HBDs' structure contains an antiparallel β -sheet that stabilized by three disulfide bonds. The β -sheet is flanked by a α -helix whose orientation in relation to the β -sheet is also stabilized by the disulfide bond (Cys¹-Cys⁵) and is thought to be a contribution to the antimicrobial activity and chemo-attractant of the HBDs [107].

5.2.5 Structure-Function Relationship

The relationship between structure and function of defensins is still poorly understood. The β -sheet constitutes the core of defensin and is suggested as the central to the defensin's structure, and presumably, its function [107]. It is found that disulfide bridges are not necessary for maintain the antimicrobial activity of some HBDs [109, 117, 118]. It is also postulated that the cationicity, amphipathicity and ability to oligomerize are considered as key factors to the mechanisms of antimicrobial activity [109, 125]. However, there is no universal rule to suggest a fixed structure-function relationship for HBDs and remains an active area of research.

5.3 Homology Modeling

The 3D structure of a protein is a prerequisite for a better understanding of its functions at the molecular level. There are several experimental techniques determine 3D structure of proteins. Among them, X-ray crystallography and NMR spectroscopy are the most widely used methods. However, it may be a challenge to determine structures by using these techniques for some reasons. Generally, many proteins cannot crystallized for X-ray crystallography and are simply too large for NMR spectroscopy. In addition, the acquisition of information by these techniques is a slow and expensive process. Thus protein modeling provides us an alternative route to obtain the structure and it is the way when experimental fails. There are three major methods for protein prediction, which are *homology modeling*, *threading* and *ab-initio* methods. Ab-initio folding is a simulation-based method. The basic idea is to build

empirical function that simulates real physical force and potentials of chemical contacts. Since there is no use of sequence alignment and no direct use of structure, this method is not very accurate and only small proteins can be properly simulated. In contrast, the template-based methods have achieved success in recent years, due to the enlargement of the PDB database. The idea of the template-based method is to find a similar structure from PDB for new protein. Both homology (comparative) modeling and threading belong to this kind of method. Homology modeling has a good prediction accuracy, but it works only if the new protein and the template share more than 25% identical residues. The threading method is more sensitive than the homology modeling method. However, the prediction accuracy of the threading method is not very good. In addition, homology modeling is by far the easiest prediction method in this field [127].

Homology modeling is a technique based on the observation that both the structure of proteins and their sequence are tend to conserved, at least for some key amino acids, during evolution in the folding process [127]. Therefore, similar aminoacid sequences usually fold into similar structures. It is now widely used to produce reasonable protein structures [128-133]. This trend implies the possibility of generating sequence alignments of the target protein one or several templates having 3D structures already solved. For example, homology modeling is used to build structures of toxins [128, 129] and enzyme-substrate complexes [133] based on single-template modeling, and structures of glycine receptor [130] and its ligand binding domain [131] based multiple-templates modeling. In general, homology modeling is a process involving two main steps: (i) template recognition and sequence alignment, and

(ii) model building, refining and finally evaluation of the model. We give a brief overview of these two steps in the subsection below.

5.3.1 Template Recognition and Sequence Alignment

An important step in homology modeling is template recognition which is identification of well suited protein structures related to target sequence and choosing those with higher homology score as template for the desired model [127]. This is done by sequence alignment between target and template. Homology modeling produces reasonable protein structures, as long as an alignment with at least 25% sequence identity has been found as shown in Figure 5.3. From the figure, it is shown that the homology modeling is fast and simple when the sequence identity of target and template proteins is higher than 75%. Moreover, the accuracy of the homology modeling in this range is comparable to that of NMR and X-ray. When the sequence identity lowers to 50% to 75%, more time is needed to find the details of the model and to correct the alignment. For sequence identity is 50% to 25%, the limiting step is the sequence alignment. More effort has been made to obtain most possible alignment in this range.

Figure 5.3. The limiting steps in homology modeling as function of percentage sequence identity between the structure and the model [127].

There are two common techniques for sequence alignments. One is single-template alignment, and another is multi-templates alignment that aligns several homologous protein sequences together. In some cases, the methods rely on multiple templates that combine the good parts of several templates in one model have been proven successful in identifying more distantly related homologues [134, 135]. It would be anticipated that using multiple templates rather than a single one is better in the case the target-template similarity is low.

5.3.2 Model Building, Refining and Evaluation

5.3.2.1 Model Building and Refining

Once a reasonable alignment is found, the actual model building can be carried out. Generally, the coordinates of backbone atoms of the target protein simply copy the coordinates of those template atoms according to the sequence alignment. It is simple to arrange all the atoms according to the template when they are conserved. However, there are loop regions that often show little sequence conservation and may diverge in length from the template sequence. For these regions, it is very trivial to generate the coordinates. There are three major methods for loop modelling. A widely used method is employing a loop search, which searches PDB for a reference peptide with identical length that can match the gap in the protein model without introducing large distortions. This approach is supported for major molecular modelling programs and servers, such as Modeller [136], Swiss-Model [137] and WHAT IF [138].

So far, the backbone of the target protein is created and we need to add the sidechains. It is reported that aminoacid sidechains have certain energetically

favoured conformations (referred as rotamers). Thus, one of commonly used methods is adding missing sidechains according to backbone conformations from backbone-dependent rotamer library, which is based on that certain backbone conformations strongly favour certain rotamers. To predict the sidechain rotamers accurately, we need the correct backbone, which in turn depends on the rotamers and their packing. Therefore, iteration is introduced to refine the model until certain convergence is satisfied.

5.3.2.2 Model Evaluation

A 3D protein structure created from homology modelling will be subject to many source of error. Therefore, it is essential to have an assessment of a structure's overall quality and be able to identify poorly folded regions. This is achieved by carry out the model evaluation, which aims to determine whether the model is acceptable or not. There are serveral softwares developed for this purpose, and in this research we use the protein structure validation software suite (PSVS) [139] which is a package including Procheck, MolProbity, Verify-3D and various structure-validation tools.

5.4 Molecular Mechanics

Molecular mechanics is a widely used technique in molecular modeling using the classical mechanics/Newtonian mechanics to search low energy states of molecules based on force field (i.e. empirical potential energy function) It provides the physical basis behind the model, and is also commonly used in several biochemical and biophysical systems, such as conformational analysis of proteins, protein-membrane interactions and drug designs.

The molecular mechanics simulation program used in this work is Gromacs (Groningen Machine for Chemical Simulation), which is a free software for performing energy minimization (EM) and molecular dynamics (MD) based on molecular mechanics force fields [140]. A brief overview of force field, EM and MD methods are given below.

5.4.1 Force Fields

A force field is a set of equations with some parameters, which defines the potential energy of the system. Usually, the parameters in the equation describe the properties of atoms and molecules, such as atomic masses, charges, bonds and angles, as well as force constants and reference which have been determined experimentally. The potential functions of the system can be divided into the bonded interactions part, non-bonded interactions part and restraints part, as given in equation (5.1),

$$E_{total} = E_{bonded} + E_{non-bonded} + E_{restraints} \quad (5.1)$$

The bonded interactions can be further subdivided into the *bond stretching* (2-body), *bond angle* (3-body) and *dihedral angle* (4-body) interactions based on the number of atoms involved, as shown in equation (5.2),

$$E_{bonded} = E_{stretch} + E_{angle} + E_{dihedral} \quad (5.2)$$

There are several functions to represent the bonded interactions: typically, the Harmonic potential is used for bond stretching and angle bending interactions, and periodic cosine function for the torsion potential. The non-bonded interactions are computed between all pairs of atoms within a

certain separation distance. Generally, it contains der Waals interactions (short range) and electrostatic interactions (long range) as show in equation (5.3),

$$E_{non-bonded} = E_{electrostatic} + E_{vdw} \quad (5.3)$$

Usually, van der Waals interactions are calculated based on Lennard-Jones potential and the Coulomb potential is used for electrostatic interactions.

5.4.2 Energy Minimization

EM method is a technique to minimize the potential energy of a system to local minima [140]. Since the protein structures obtained from both experiments and computer modelings may have steric clashes and distorted bonds, it is necessary to perform EM to relax the worst conflicts and to find a more energetically favorable conformation. There are mainly two EM methods involved in Gromacs: the steepest descent (SD) method and the conjugated gradient method. The SD method is a first order EM which simply moves along negative gradient and hence the convergence is quite slow. This drawback could be improved by using conjugated gradient method. However, the latter method may bring you away from the minimum. Thus it is recommended to run a SD EM followed by a conjugated gradient EM.

5.4.3 Molecular Dynamic Simulation

Proteins are non-rigid molecules and can twist along the bonds of the backbone. Therefore, there are more than one local minimum in the potential energy surface of a protein, and consequently the protein has a large number of

possible conformations. Based on the Anfinsen's dogma thermodynamic hypothesis, the protein's native (or stable) structure is the one that minimizes the global free energy of the protein [126]. However, it is impossible to escape from one local minimum to another and find a global energy minimum by using EM, but this problem can be overcome by using MD which is a more sophisticated method for searching the all possible conformations of the molecules.

MD provides information about the dynamical behavior of molecular systems such as molecular conformations by solving the classical Newtonian equations of motion for the molecules. It is a time-dependent process which provides an ensemble of molecular conformations (usually converged to a local energy minimum). These generated molecular conformations are averaged to obtain the global minimized conformation of molecules. This time-averaged structure is considered as a stable structure that describes the geometric average shape of molecules. Therefore, MD simulation is a powerful method for molecular structure validation and refinement. In our work, we used MD to validate the homology models and to obtain a more energetically favorable protein model with lower potential energy from the initial models.

Chapter 6

Structure of Human β -Defensin 28

6.1 Introduction

HBD-28 is a member of the human β -defensin subfamily and was first found in the male genital tract in 2003 [113]. It is a newly identified defensin which shows a similar broad range *in vitro* antimicrobial activity against all microbes tested and is only slightly less active than HBD-3, the most potent HBD [141]. Although its explicit functions are yet to be studied in detail, HBD-28 makes a promising candidate for antibiotic therapy due to its low susceptibility to develop resistance. However, the extraction cost of HBD-28 is very expensive due to the concentration of HBD-28 is low in its natural resource. Therefore there are insufficient amount of pure HBD-28 for detailed characterization and structural studies. Thus the exact tertiary structure and antimicrobial mechanism of HBD-28 were poorly understood. Therefore finding an efficient simulation method to obtain the structure of HBD-28 is critical necessary for further study of HBD-28. By employing molecular modeling (homology) and simulation (MD) methodologies in a systematic way, we bypass the difficulties in the structural determination in this work and study the antimicrobial behaviors of HBDs later in future work.

In the rest of this chapter, we first build initial models of HBD-28 based on HBD-2 and HBD-3 of known structures by homology modeling. This approach is justified by observing that similar amino acid sequences usually fold into

similar conformations [127]. However, the target-template sequences identified are low (mostly less than 25%) in this case. To resolve this issue, both single-multiple-template sequence alignments, which may provide more comprehensive coverage across the target sequence, were attempted to improve overall alignment accuracy. In order to determine which of the possibly alignment produces the most physically reasonable model, we generated quality scores for models based on different alignments using the PSVS [139] . The models of good qualities are chosen for the subsequent MD simulations to their structural stability in pure water and TFE/water mixture. Previous study [112] has shown that the HBD-3 was flexible and random in aqueous solution, and its conformation was dictated by solution environment. Therefore, it is interesting to examine the conformations of HBD-28 in different solution environments. The TFE has been widely used in protein structure studies as it can stabilize the secondary structure of protein in aqueous solution, especially α -helix. It is found that the HBD-28 underwent a marked conformational transition in the presence of TFE as observed in experimental CD spectra, which shows the increased helical conformation [142]. By molecular modeling and simulation, we display the similar experimental phenomena.

6.2 Simulation and Experimental Details

6.2.1 Homology Modeling

The structure of templates HBD-2 and HBD-3 used in this work were in PDB with accession ID: 1FD3 for HBD-2 and 1KJ6 for HBD-3. Sequence alignments of single and multiple templates were generated by the CLUSTALW

[143] program. For the multiple-template sequence alignment, the alignment was manually adjusted to conserve the Cys residues. Model building was performed using the MODELLER [136] protocol in Discovery Studio (DS) 2.5 software [144], which is an automated homology modeling software. Single-multiple-template modelings (see Table 6.1) were performed using the alignments generated in the previous step to obtain the initial models of HBD-. The positions of Cys residues that form the disulfide bridge were constrained to the aligned Cys residues and corresponding disulfide bridges in templates using “Disulfide Bridges” parameter in DS. We apply CHARMM 27 force field [145] to all models of our interest.

6.2.2 Molecular Dynamics Simulation

The preliminary 3D models obtained from the homology modeling using DS were first refined by EM in vacuum, subsequently the energy-minimized systems were subjected to MD simulation in water. In these refinement steps, the structural stabilities of the models are investigated. We consider a model to be acceptable if its overall structure is reserved after EM in vacuum (i.e. intrinsic stability) and MD in water. The models generated in this process were put together with 50% TFE/water mixtures to adopt similar experimental environment in MD simulation.

The intrinsic stabilities of initial HBD-28 models were tested by a combination of two EM methods in vacuum: the SD method and the conjugate gradient method. The SD was first run until a minimization criteria (i.e. the of maximum force being less than $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$.) was satisfied, followed

by conjugate gradient until the maximum force was smaller than $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$.

The intrinsically stable structures obtained from EM steps were then placed with water molecules as a starting model for MD treatment. After SD minimization of solvated protein for 500 iteration steps, the temperature of the system was gradually increased to 300K over 20 ps, followed by 80 ps of MD steps with position restraints on non-hydrogen atoms of the protein at constant pressure 1 bar to relax the system. The restraints on non-hydrogen atoms were removed, after that the simulation was then continued for 10 ns to assess the stabilities of the structures, where the coordinates were saved every 1 ps for a detailed analysis. After NPT MD simulation (i.e. constant number of atoms N , pressure P and temperature T) is completed, another 6 ns independent simulation was performed with different simulation ensemble scheme (i.e. NVT MD simulation - constant number of atoms, volume V and temperature) in order to check the stability of the model reliably. Note that the volume predicted from the NPT MD simulation was applied in the NVT MD simulation.

After the completion of NVT simulation, the highly plausible HBD-28 models were obtained; in order to evaluate the influence of the realistic solution to the structure of HBD-28, the MD simulations of the highly plausible HBD-models were performed in 50% water/TFE mixtures for 20 ns under NPT condition. The simulation results of the secondary structure of HBD-28 models in the presence of TFE were compared with the CD experiments (see Section 6.3.5). These simulations further prove the reliability of the NVT-simulation-NVT-simulation-refined HBD-28 models.

All EM and MD simulations were performed with the GROMACS 4.07 [140] program package. We also employed the CHARMM force field for the TIP3P [146] water model. The TFE model was obtained from the work of Chitra *et al.* [147, 148]. To neutralize the system, sodium counterions were added and water molecules were removed if they overlapped with the sodium ions. Covalent bonds in the protein were constrained using the SHAKE algorithm [149], whereas the water molecules were constrained using the SETTLE algorithm [150]. The simulation box was $54 \times 54 \times 40$ Å in dimensions and periodic boundary conditions were applied. The amount of water molecules is closer to 3000 for all models, which provides a layer of pure water with a width of approximately 10 Å to avoid interface with the boundaries. The particle mesh Ewald [84] method was used to treat electrostatic interactions with a cut-off value of 10 Å and space grid size of 1.2 Å. The cut-off distance of van der Waals interactions was 12 Å. The time step for integration was 2 fs and the neighbor list was updated every five steps. The temperatures were maintained at 300 K by velocity-rescaling thermostat with a coupling time of 0.1 ps in both NPT and NVT simulations [151]. The constant pressure in NPT simulation was 1 bar using the Parrinello-Rahman barostat with a coupling constant of 2 ps [152, 153]. All data for analysis were collected after equilibrium, which was achieved until the averaged temperature, pressure and structure remain stable.

6.2.3 Simulation Data Analysis

The qualities of the initial models generated by homology modeling using DS were assessed by the PSVS scores [139]. The secondary structure of was analyzed by the dictionary of secondary structure of proteins (DSSP) [154]

program every 10 ps and the 3D structure of peptide was visualized using the visual MD [155] software. All analyses of resulting trajectories were performed by using GROMACS program package. The value of root mean square (RMSD) of the backbone atoms related to the structure at the beginning of equilibrium was calculated by

$$RMSD(t) = \sqrt{\frac{\sum_{i=1}^N (x_i^t - x_i^0)^2 + (y_i^t - y_i^0)^2 + (z_i^t - z_i^0)^2}{N}} \quad (6.1)$$

where (x_i^t, y_i^t, z_i^t) and (x_i^0, y_i^0, z_i^0) are the positions of the atom i at time t and beginning time 0, respectively, and N is the total number of atoms in the molecule. The root mean square fluctuation (RMSF) measures the molecular fluctuations about the time-averaged molecule structure. The RMSF of particular atom j can be calculated by

$$RMSF(j) = \sqrt{\frac{1}{n_j M} \sum_{i=1}^N \sum_{k=1}^M [(x_{i,t_k} - \langle x_i \rangle)^2 + (y_{i,t_k} - \langle y_i \rangle)^2 + (z_{i,t_k} - \langle z_i \rangle)^2]} \quad (6.2)$$

where M is the total number of recording configurations. $(x_{i,t_k}, y_{i,t_k}, z_{i,t_k})$ and $(\langle x_i \rangle, \langle y_i \rangle, \langle z_i \rangle)$ denote the position of atom i at time step t_k and the averaged position of atom i over M time steps, respectively.

6.2.4 Circular Dichroism Spectra

The CD spectra were recorded on the ChirascanTM CD Spectrometer (Applied Photophysics, Surrey, UK) in the range of 190-260 nm at room temperature. HBD-28 solutions were prepared at concentrations of 100 and 50

$\mu\text{g/ml}$ in 20 mM tris(hydroxymethyl)aminomethane-HCl (Tris-HCl) buffer at 7.4 and 20 mM Tris-HCl with 50% v/v TFE at pH 7.4 (Fluka, Spain), An average of 5 repeat scans for each sample was calculated as the final CD The data were also analyzed using the CDNN/PFPPFIT program [156], which predict the percentage of α -helix, parallel and antiparallel β -strands, turns, and others.

6.3 Results and Discussion

6.3.1 Sequence Alignment

The most important step in homology modeling method is the sequence alignment. A number of modeling examples have shown that the differences in alignments affect the final models, in turn the quality of homology modeling [134, 135, 157, 158]. There are two important factors in the accuracy of alignment: (i) the template identification and (ii) the target-template(s) [135]. Usually, the methods developed for template identification relies on the single template [159, 160]. However, in some cases, the methods rely on templates have proven successful in identifying more distantly related homologue [134, 135]. It would be anticipated that using multiple templates rather than a single one is better in the case the target-template similarity is low. In our study, the quality of the single-template sequence alignment was by sequence identity, which is 21.6% for HBD-1, 30.2% for HBD-2 and 28.9% for HBD-3. Since the sequence identity of HBD-28 to HBD-1 is lower than which is the threshold for safe homology modeling [127], it is considered as an improper template. The sequence identities of the other two templates are a

higher than 25%, thus we use multiple-template modeling to increase the accuracy of the modeling. Table 6.1 shows the sequence alignments of HBD-28 with different templates, and *A*, *B*, and *C* alignments were determined using CLUSTALW tool. It is reported that the HBDs have an identifiable consensus sequence: $X_{2-10}CX_{5-6}(G/A)XCX_{3-4}CX_{9-13}CX_{4-7}CCX_n$ (where *X* is any amino acid) and three common disulfide bridges (Cys¹-Cys⁵, Cys²-Cys⁴ and Cys³-Cys⁶). From the Table, we found that all defensins were aligned according to the conserved pattern of six cysteine residues except alignment *B*. As cysteine residues that sit in disulfide bridges do not mutate easily, the cysteine residues should be preserved in the alignment. In this sense, we manually adjusted the alignment *B* to conserve the cysteine residues and to obtain the alignment *D*. Based on these four alignments, we constructed four models and labeled as HBD-28_A, HBD-28_B, HBD-28_C and HBD-28_D, respectively.

Table 6.1. The sequence alignments of HBD-28 with different templates. Here HBD-28 is aligned with HBD-2, HBD-3 and their combinations. Conserved residues are highlighted by asterisks at the bottom of each alignment and *A*, *B*, and *C* are determined using CLUSTALW tool.

Alignments		Single-Template Alignment (CLUSTALW)	
A	Template	HBD-3	GIGDPVTCL-KSGAICHVPVFCP--RRYKQIGTCGLPGTKCKKKP
	Target	HBD-28	--ARLKKCFNKVTGYCRKK-CKVGERY-FIG-C-LSGKLCCAN--
		* * * * *	
B	Template	HBD-2	GIINTLQKYYCRVRGGRCVLSCLPKKEQIGKCSTRGRKCCRKK
	Target	HBD-28	---ARLKKCFNKVTG-YCRKKCKVGERYEIG-CLS-GKLCCAN---
		* * * * *	
C			Multiple-Template Alignment (CLUSTALW)
	Template_1	HBD-2	GIG---DPVTCCLKS-GAICHVPVFCPPRYK-QIGTCGLPGTKCKKKP-
	Template_2	HBD-3	GIINTLQKYYCRVR-GGRCVLSCLPKKE-QIGKCSTRGRKCCRKK
	Target	HBD-28	-----ARLKKCFNKVTGYCR-KKCKVGERYEIG-C-LSGKLCCAN--
		* * * * *	
D			Adjusted Single-Template Alignment (Manual Correction)
	Template	HBD-2	GIINTLQKYYCRVRG-GRCVLSCLPKKEQIGKCSTRGRKCCRKK
	Target	HBD-28	-----ARLKKCFNKVTGYCR-KKCKVGERYEIG-CLS-GKLCCAN---
		* * * * *	

6.3.2 Initial HBD-28 Models

Table 6.2 shows the detailed residue numbers of the predicted secondary structure elements for four initial models generated by DS in α -helix and β - β -strands. Despite the fact that the secondary structures of four models are similar and generally agreed with each other, minor differences are observed in β_2 and large differences appeared in the location of α -helix.

Table 6.2. Comparison of the secondary structure elements in different HBD-28 models predicted by DS, EM and MD. Here numbers represent the residue numbers in each HBD.

Model		<i>Secondary Structure Elements</i>			
		α -helix	β_1	β_2	β_3
HBD-28_A	DS	3-7	13-15	23-24	33-36
	EM	3-5	13-15	23-24	33-36
	MD	3-8	13-15	24-27	33-35
HBD-28_B	DS	7-10	13-15	24-25	33-36
	EM	2-9	0	0	0
	MD	0	0	25-27	33-36
HBD-28_C	DS	2-8	13-15	23-24	33-36
	EM	3-6	13-15	0	33-35
	MD	0	12-15	23-25	33-36
HBD-28_D	DS	2-9	13-15	24-25	33-36
	EM	3-5	13-15	24-25	33-36
	MD	3-7	14-15	24-27	33-36

To illustrate the magnitude of the difference observed in models, we

superposed four models to obtain the averaged structure, and calculated the per-residue RMSD of superimposed structures from the averaged structure by using SuperPose [161] software. The results are shown in Figure 6.1. The largest deviations are found in initial residues from 1 to 10 where α -helix located with RMSD values higher than 2.0 Å. Most RMSD values of residues from 11 to 37 are less than 2.0 Å, and those become less than 1.5 Å for β_1 and β_3 regions. From this observation, it is proposed that this is a structural distinctness and all four models are considered as reasonable candidates for the 3D structure of HBD-28 by homology modeling.

Figure 6.1. Average RMSD of all $C\alpha$ atoms of the HBD-28 models relative to the averaged structure as a function of residue number.

The quality of the four initial models obtained from homology modeling step is first checked using the PSVS software, which includes Procheck,

MolProbity, Verify-3D and various structure validation tools. The results of Verify-3D [162] score are plotted with respect to the residue numbers as shown in Figure 6.2. The Verify-3D indicates how well each aminoacids fits in its current environment and it is very sensitive to distinguish misfolded models. Generally, the value approaching to or falling below 0 indicates incorrect and a higher positive value indicates a better one. From the figure, we can see that models HBD-28_A and HBD-28_D generally yield higher scores than models HBD-28_B and HBD-28_C, and the model HBD-28_D has the highest Verify-3D score.

Figure 6.2. Verify-3D score of the HBD-28 models as a function of residue number. The solid, dot, dash-dot-dot and dash lines represent the values for HBD-28_A, HBD-28_B, HBD-28_C and HBD-28_D, respectively.

We also calculated other structure-quality assessment scales of four

with Ramachandran plot (phi/psi dihedral angle distribution), ProsaII (correct and incorrect folds), Procheck G-factor (all dihedral angle distribution) and MolProbity clash (accuracy of predicted models), and the results yielded the same outcome (detailed values are provided in appendix C). All measures indicate that the quality of models HBD-28_A and HBD-28_D is better than of models HBD-28_B and HBD-28_C. Considering that the reason of the difference between models HBD-28_B and HBD-28_D was inherited from the different alignments where the position of the cysteine residues were different, we can deduce that the cysteine residues should be conserved when aligning the sequence to template in the case of HBDs. By comparing quality of models HBD-28_A (single template: HBD-2) and HBD-28_D (single template: HBD-2) with HBD-28_C (multiple templates: HBD-2 and -3), we also observe that multiple-template alignment may not improve the quality.

6.3.3 Stability of the Initial HBD-28 Models

A comparison of the initial secondary structure elements of HBD-28 predicted by DS with that refined by EM in vacuum is given in Table 6.2. Table 6.2 shows that the application of EM has caused significant changes in the predicted structures. The secondary structures of initial models HBD-28_A and HBD-28_D do not vary much, but those of models HBD-28_B and HBD-28_C change significantly. The β_2 of the model HBD-28_C disappeared and all the β - β -strands lost in the model HBD-28_B, which means that the models built on alignment C and B are very intrinsically unstable. Furthermore, to observe this phenomenon in water, we run models HBD-28_A to D, which were treated with EM in vacuum, to explore the conformational changes in water solution. Table

6.2 also shows the results refined by MD simulation. We found that the β_2 was extended and α -helix was shortened after the EM and MD refinement for the model HBD-28_A. For the model HBD-28_D, it shows small variations, which are shortened α -helix and β_1 with longer β_2 . However, the models HBD-28_B and HBD-28_C are either free of β -strand(s) or α -helix, which indicates their structural instability in water due to incorrectly folded states. From two modeling analyses, the stabilities of models HBD-28_A and HBD-28_D are better than the others, and this finding consequently supports that cysteine residues should be conserved in the alignment and the multiple-template alignment does not necessarily improve the quality of the model.

Now we turn our focus to two stable models for further study. Figure 6.3 shows the secondary structure of each residue as a function of time for models HBD-28_A and HBD-28_D in water under NPT simulation. Different colors represent different secondary structures: red color represents β -strand and blue (and violet) color represents helical conformation. From the figure, there are continuous fluctuations in secondary structure conformations of models HBD-28_A and HBD-28_D over the time. The β_2 and α -helix in the model HBD-28_A were disappeared at the beginning of the simulation and reappeared at later time. Compared to the model HBD-28_A, which is unfolded at the beginning and refolded later, the secondary structure of the model HBD-28_D is well preserved during the simulation as shown in Figure 6.3, in which several clear color-bands without wide gap were observed. This result shows that the most stable structure in water environment was the model HBD-28_D.

Figure 6.3. Secondary structure of each residue (on y-axis) as a function of time (x-axis) of four models in pure water simulation under NPT condition. Here, different color represents different secondary structure types, red color denotes β -strand, blue and violet colors denote helical conformation. Secondary structures are obtained from the DSSP program.

In order to evaluate the relative structural change or magnitude of the displacements of each structure in MD simulation, we calculate the RMSD (based on the backbone atoms) from the relevant initial structure as a function time as shown in the Figure 6.4. The values of RMSD for models HBD-28_A and HBD-28_D were measured every 1 ps. We found that the RMSD values

rapidly increase at the beginning of the simulation, illustrating that the structures of the models rapidly alter from the initial structures. For the model HBD-28_A, a slow increase of RMSD after 1 ns is observed, which means the structure gradually stabilized. For the model HBD-28_D, the structure significantly deviate around 2 ns, and then become relatively stable after 4 ns. The RMSD values in both models HBD-28_A and HBD-28_D are less 2.5 Å during the entire MD simulation, which indicates the total structure change is small and overall stability is reserved.

The relative mobility of each residue in HBD-28 models can be measured by plotting the RMSF with respect to the average structure. Figure 6.5 shows the RMSF as a function of the residue number, and we can see that residues in the β -strands were least mobile with RMSF values below 1 Å. The most flexible parts of HBD-28 models are *N* and *C* termini, but within the aminoacid sequence, there are also regions with large RMSF values, which represent the region connecting β_1 and β_2 . This is partly due to the Gly-21, which is the smallest aminoacid and therefore it is very flexible. Comparing these two models, the model HBD-28_D has smaller RMSF values than the model HBD-28_A, thus the model HBD-28_D is mechanically more stable structure.

Figure 6.4. The RMSD value of the HBD-28 models as a function of time. The black and red lines represent the values for HBD-28_A and HBD-28_D, respectively.

Figure 6.5. The RMSF value of the HBD-28 models as a function of residue number. Other symbols are the same as in Figure 6.4.

6.3.4 Final HBD-28 Models

In order to confirm the stabilities of the models HBD-28_A and HBD-28_D, those weakly fluctuated structures seen after NPT refinement were submitted to another round of MD simulation with NVT ensemble, which is run for 6 ns, to check any abnormality occurred. If statistical behaviors of the model in these two simulations were similar, one might consider a model to be a highly plausible model for HBD-28.

Figure 6.6. Secondary structure of each residue (on y-axis) as a function of time (x-axis) of four models in pure water simulation under NVT condition. Other symbols are the same as in Figure 6.3.

Figure 6.6 shows the secondary structures of each residue as a function of time for models HBD-28_A and HBD-28_D under the NVT condition. We observe that short transitions occur in the middle of the simulation but the structure elements quickly return to their previous states for both models HBD-28_A and HBD-28_D. The analysis of the RMSD value obtained between the starting conformation and the final structure shows 0.96 Å for the model HBD-28_A and 0.51 Å for the model HBD-28_D, respectively. Both structures were stable during NVT MD, and mechanical stability favors to the model HBD-28_D. Finally, the RMSD between the models HBD-28_A and HBD-28_D was calculated to see their structural similarity, and the value was 2.97 Å before the NVT MD refinement and 1.95 Å after the NVT MD refinement.

Figure 6.7. Comparison of the 3D structures between HBD-28_A and HBD-

HBD-28_D after NVT MD simulation. The α -helix is colored in red, β -strands yellow and others in blue.

Figure 6.8. Structural comparison of HBD-28_A (top) and HBD-28_D (bottom) by surfaces shown in the front view (left) and back view (right). The α -helix is colored in red, β -strands in yellow and others in white.

Recalling Figure 6.4, refined models HBD-28_A and HBD-28_D are very similar in shape. The final conformations obtained during this step were presented in Figs.7 and 8. It is shown that the secondary structures of models HBD-28_A and HBD-28_D were well conserved and their 3D structures are very similar. With surface displayed (Figure 6.8), positions of α -helix and β -strands are located in similar space but with different surface areas.

6.3.5 Comparison with Circular Dichroism Spectra

These final structures obtained from the NVT MD refinement were compared with CD experimental data in Tris-HCl buffer. In these experiments, the Tris-HCl was added to control the pH at 7.4 and the amount of Tris-HCl is very small (50 μ M in 20 mM Tris-HCl buffer). For the simulation, we construct a similar pH environment with pure water to mimic the Tris-HCl buffer at pH 7.4, as there is no good forcefield available for Tris-HCl buffer in GROMACS and the major component in the solution is water (i.e. two Tris-HCl molecules per 5546 water molecules). The secondary structural analysis of the HBD-28 by the CD spectra in Tris-HCl buffer indicated that HBD-28 has three β -strands with a short α -helical segment. Table 6.3 shows the presence of the molar populations of both α -helical and β -strand conformations in NVT-simulation -refined models HBD-28_A and HBD-28_D with that of CD experiment. From the experiment, the helical segment is estimated to be 24.3%, and β - β -strands to be 22.9%. From the Table, we can see that the populations of helical conformations of models HBD-28_A and HBD-28_D are slightly lower than experimental values and β -strands are comparable to those. Note that solution environments from experiment (Tris-HCl) and simulation (H_2O) are similar in pH, but different due to unavailability of force-field parameters for Tris-HCl.

Table 6.3. The molar percentage of secondary structure configurations of HBD-HBD-28 by experiment and MD simulation in pure water. Here antiparallel β - β -strand stands for β_2 and parallel β -strand contains β_1 and β_2 .

	Pure Water		20mM Tris-HCl
	HBD-28_A	HBD-28_D	buffer (pH 7.4)
Helix	16.2%	13.5%	24.3%
Antiparallel β -strand	10.8 %	10.8%	10.5%
Parallel β -strand	16.2%	16.2%	12.4%

To further elucidate the structure and folding of HBD-28 in different environment, TFE/water mixture was chosen. The TFE has been extensively used in CD and NMR studies of peptides in solution since it plays a role as a α - α -helix promoting solvent. In the Tris-HCl with 50% v/v TFE buffer CD experiment, the α -helical conformation of HBD-28 is largely increased in the presence of 50% TFE in water, which is known to stabilize this type of conformation. The stability of β -strands is sensitive to solvation force so that it decreases in the TFE buffer. The comparison of experimental results and simulation data is shown in Table 6.4. We observe that the α -helix of the model HBD-28_D increases a little from 13.5% to 21.6% and that of the model HBD-HBD-28_A is invariant. The antiparallel β -strand (β_2) of the model HBD-28_A decreases but remains same in the model HBD-28_D, but the opposite is observed for the parallel β -strand (β_1 and β_3). This may be due to the different surface characteristics (i.e. different contributions of α -helix and β -strands, cationicity and amphipathicity) of models HBD-28_A and HBD-28_D as

in Figure 6.8. Although the MD results were different from the experiment data, the tendency to form helix and decreased β -strands was still observed in the model HBD-28_D. There are three possible reasons for the differences in α - and β -strands percentages between refined models and experiment. First, the solution environments in simulation are not exactly same as in experiment. In the simulation, the HBD-28 and Tris-HCl concentrations are so low that we can treat them as being infinitely dilutes. These approximations assume that protein-protein, protein-Tris, protein-HCl, TFE-Tris and TFE-HCl interactions are negligible. Second, it is assumed that there is only one state of HBD-28 (monomer) and no oligomers are involved. However, it is known that the HBDs have the ability to oligomerize and this possible conformational heterogeneity may be contributing to the helix formation in the present of TFE. The third possible reason is the less accurate forcefield. It is very difficult to evaluate a forcefield that can accurately describe the properties of interest. As a result, the differences between experimental observation and MD simulations are unavoidable as the existing errors in the forcefield.

Table 6.4. The molar percentage of secondary structure configurations of HBD-HBD-28 by experiment and by MD simulation in the present of 50% TFE. notations as shown in Table 6.3.

	50% TFE/Water		20mM Tris-HCl
	HBD-28_A	HBD-28_D	50% TFE (pH 7.4)
Helix	16.2%	21.6%	69.0%
Antiparallel β -strand	5.4 %	10.8%	3.2%
Parallel β -strand	16.2%	10.8%	3.0%

6.4 Conclusion

In this chapter, we have obtained a stable 3D structure of HBD-28 by homology modeling and MD simulation in the absence of experimental determined by X-ray or NMR. Both single- and multiple-template alignments were used to improve the alignment quality due to the low sequence identity of HBD-28 matched to HBD-2 and -3. However, the results show that multiple-multiple-template alignment do not improve a model generated from the single-single-template alignment. Modeling quality is also proportionally related to the sequence alignment accuracy. In case of HBDs, we have shown that the six cysteine residues are significant to the sequence alignment accuracy and should be conserved in the alignment. We carried out MD simulations of four HBD-28 models in water solution using CHARMM force field by the GROMACS simulator. Structural analysis of the MD results showed that HBD-28_A and HBD-28_D are initially deviated slightly from the initial model and remain stable during simulation, whereas models HBD-28_B and HBD-28_C lose some characteristic secondary structures, e.g. α -helix or β - β -strands. Additionally, the model HBD-28_D showed a more satisfactory stability of the core in the MD simulation and more stable secondary structure as well than any other models of our interest. The overall structure drift of models HBD-28_A and HBD-28_D did not exceed 2.5 Å, which means that are very similar. By comparing the simulation result with CD experiment data, we found that models HBD-28_A and HBD-28_D are possible models for HBD-28. For NVT-simulation-refined models HBD-28_A and HBD-28_D, we studied the structure and folding of the HBD-28 in 50% TFE/water mixture. By

comparing the CD spectra results with the data obtained from MD simulations refined models HBD-28_A and HBD-28_D in TFE buffer, we found that the refined model HBD-28_D exhibited a higher variation in the presence of TFE and showed elongated α -helix and shortened β -strands, which implicitly illustrate the experimental behavior of HBD-28. To this end, the model HBD-HBD-28_D is currently the closest tertiary structure to the real one. This is an encouraging result as the HBDs template has low sequence identity (one of the main areas of difficulty in homology modeling) with our target sequence. Our approach including homology modeling followed by MD treatment is not likely restricted only to HBDs but applicable to aminoacid mutants or to other proteins.

Chapter 7

Conclusions

Molecular simulation techniques were employed to improve the performance of the IE theories via application of bridge functions and also to predict the structure of HBD-28 via homology modelling and MD approaches.

For the IE theory, we proposed two modified Mayer-sampling methods to calculate the integrated diagrams appearing in the coefficients of bridge namely TMMC Mayer-sampling and overlap sampling. The bridge coefficient $b_n(r)$ was constructed in terms of $h(r)$ for a hard sphere system and calculations were performed using prescribed $h(r)$ for reduced densities at 0.2, 0.5 and 0.8 to the third order expansion in density. The reliability and precision of these two methods were tested in detail. The comparison of the results from the proposed methods and the original Mayer-sampling method revealed that the TMMC Mayer-sampling approach shows better precision over the core and tail regions of $b_2(r)$ and $b_3(r)$, except the moderate distance region. This is due to the scheme enables the root particles to pass through low-probability regions to reach extreme separation distances and evenly sample an entire range. It was concluded that overlap sampling method shows overall improvement in the precision of the bridge coefficients and the efficiency of this method is related the degree of the overlapping region between the target and reference systems. Moreover, both methods can be straightforwardly applied to calculate the order bridge function coefficients and to any model systems with relatively simple modifications.

We have also investigated the bridge function required to yield a singlet IE theory up to the second power in density for the hard sphere fluid confined in a slit-pore. The slit-fluid bridge function can be divided into wall-particle bridge diagrams with h_b -bond which were evaluated by recently proposed TMMC Mayer-sampling method. The $h_b(r)$ used in cluster integrals was determined by the solution of the bulk-fluid OZ equation with a HNC closure. The calculation was performed for the reduced density of bulk fluid in equilibrium with the fluid in slit-pores from 0.3 to 0.7 with narrow slit width of 3.0σ and 4.0σ . The quantity of the slit-fluid bridge function was assessed by a comparison of the density profiles obtained from the singlet IE theory and GCMC simulation. Good agreement between the proposed approach and the GCMC data was observed. The reduced normal pressure was also calculated. The calculated result agrees with the simulation data at low to medium densities but becomes a little larger at high density. It was expected that the data can be improved by adding higher order bridge coefficients. The direct evaluation of the slit-fluid bridge function seemed to be practical since a great improvement of the quality of the singlet IE theory had been achieved for prediction of the structural and thermodynamic properties of fluids confined in narrow slit-pores.

For the HBD-28, we created a stable 3D structure of HBD-28 by modeling and MD simulations in the absence of experimental structure. With known high-resolution structures of HBD-2 and HBD-3, single- and multiple-template sequence alignments were applied to find highly probable candidates for HBD-28, which were in turn refined through MD simulations in pure water and 50% TFE/water mixture. The models obtained from homology modeling were assessed by PSVS software, which includes Verify-3D,

MolProbity and various structure validation tools. The results show that multiple-template alignment has no advantage over the single-template alignment and even worse in our case. We also found that modeling quality is proportionally related to the sequence alignment accuracy, and the six invariant cysteine residues play a key role in the alignment accuracy which should be conserved in the alignment. In Comparison with CD experimental data as well as stabilities of β -sheet and α -helix, one of HBD-28 candidates exhibited a higher variation in the presence of TFE and implicitly illustrated the experimental behavior of HBD-28, which indicates a reliable 3D structure of HBD-28 was found. Last, our approach involving homology modeling followed by MD treatments is not likely restricted only to HBDs but adaptable to aminoacid mutants or to other proteins.

Appendix A

Numerical Solution for Bulk Fluids

The numerical solution of the OZ equation for hard sphere bulk fluid employed in this work is Labik's algorithm and is summarized as below.

First all functions in r space are discretized on a linear mesh of N points $r_i = i\Delta r$. Functions in k spaces are treated similarly. Secondly, the relationship between Δr and Δk is $\Delta r \Delta k = \frac{\pi}{N}$.

To improve numerical efficiency, the following two new functions are introduced,

$$\Gamma_i = \Gamma(r_i) = r_i \gamma(r_i) \quad (\text{A1})$$

$$C_i = C(r_i) = r_i c(r_i) \quad (\text{A2})$$

such that the Fourier transform of equation (A1) becomes a one-dimensional sine transform, i.e.,

$$\tilde{\Gamma}_j = 4\pi\Delta r \sum_{i=1}^{N-1} \Gamma_i \sin\left(\frac{\pi}{N} ij\right), j = 1, 2, \dots, N-1 \quad (\text{A3})$$

$$\Gamma_i = \frac{\Delta r}{2\pi^2} \sum_{j=1}^{N-1} \tilde{\Gamma}_j \sin\left(\frac{\pi}{N} ij\right), i = 1, 2, \dots, N-1 \quad (\text{A4})$$

The following relations also hold in k space,

$$\tilde{\Gamma}_i = \tilde{\Gamma}(k_i) = k_i \tilde{\gamma}(k_i) \quad (\text{A5})$$

$$\tilde{C}_i = \tilde{C}(k_i) = k_i \tilde{c}(k_i) \quad (\text{A6})$$

Then the OZ equation (i.e. equation (1.5)) may be rewritten as

$$\tilde{\Gamma}_j = \frac{\rho \tilde{C}_j^2}{t_j - \rho \tilde{C}_j} \quad (\text{A7})$$

In a similar way, the closure equation (i.e. equation (1.8)) can be recovered such that C_i is expressed as a function of Γ_i , i.e.,

$$C_i = F(\Gamma_i) = (e^{-u(r_i)} + \frac{\Gamma_i}{r_i} + b(r_i)) - r_i - \Gamma_i \quad (\text{A8})$$

If take Fourier transform of the first-order expansion of function C_i in equation (A8) and substituted in equation (A7), We got the linearized set of equation (A7) for the unknown Γ_i , i.e.,

$$\Delta \tilde{\Gamma}_j - \frac{\rho \tilde{C}_j}{t_j - \rho \tilde{C}_j} (2 + \frac{\rho \tilde{C}_j}{t_j - \rho \tilde{C}_j}) \sum_{k=1}^N \tilde{C}_{j,k} \Delta \tilde{\Gamma}_k = \frac{\rho \tilde{C}_j^2}{t_j - \rho \tilde{C}_j} - \tilde{\Gamma}_j \quad (\text{A9})$$

with

$$\tilde{C}_{j,k} = \frac{2}{N} \sum_{i=1}^N \phi_i^0 \sin(\frac{\pi}{N} ik) \sin(\frac{\pi}{N} ij) \quad j, k = 1, 2, \dots, N \quad (\text{A10})$$

$$\phi_i^0 = \left(\frac{dF}{d\Gamma_i} \right)_{\Gamma_i = \Gamma_i^0} \quad (\text{A11})$$

The algorithm used to solve the OZ equation is as follows: (1) Choose N ,

and initial guess of $\Gamma_i^0, i = 1, 2, \dots, N-1$. Calculate $\tilde{\Gamma}_j^0$. (2) From equation (A8) to calculate C_i^0, ϕ_i^0 and $\tilde{C}_j^0$. (3) Choose M , calculate $\tilde{C}_{j,k}, j, k = 1, 2, \dots, M$, from equation (A10). (4) The initial estimate for NR iterations is $\tilde{\Gamma}_j = \tilde{\Gamma}_j^0$. (5) Solve linear equation (A9). If

$$\left[\sum_{j=1}^M (\Delta \tilde{\Gamma}_j)^2 \right]^{1/2} > 10^{-5} \quad (\text{A12})$$

Correct $\tilde{\Gamma}_j$: $\tilde{\Gamma}_j = \tilde{\Gamma}_j + \Delta \tilde{\Gamma}_j, j = 1, 2, \dots, M$ and repeat (5). If inverse is true, pass on to the next operation. (6) For $j > M$, correct the values of $\tilde{\Gamma}_j$ by equation (A7). Inverse transform of $\tilde{\Gamma}_j$ by equation (A4) to obtain $\Gamma_i, i = 1, 2, \dots, N-1$. If

$$\left[\Delta r \sum_{i=1}^{N-1} (\Gamma_i - \Gamma_i^0)^2 \right]^{1/2} > 10^{-5} \quad (\text{A13})$$

Set $\Gamma_i^0 = \Gamma_i$ and $\tilde{\Gamma}_i^0 = \tilde{\Gamma}_i, i, j = 1, 2, \dots, N-1$ and return to operation (2). If the reverse is true, the computation is finished.

Appendix B

Numerical Solution for Fluids in a Slit-Pore

The numerical solution of the singlet OZ equation for hard sphere fluid confined in a slit-pore employed in this work is also Labik's algorithm and is summarized as following.

First all functions in z space are discretized on a linear mesh of N points $z_i = i\Delta z$. Functions in k spaces are treated similarly. Secondly, the relationship between Δz and Δk is $\Delta z \Delta k = \frac{\pi}{N}$.

For computational reasons it is convenient to introduce the auxiliary function

$$\delta(z) = \gamma(z) - \beta \left(\frac{\partial p}{\partial \rho} \right)_\beta + 1 \quad (\text{B1})$$

where $\beta(\partial p / \partial \rho)_\beta$ is the bulk-fluid inverse isothermal compressibility and p is the pressure. The singlet OZ equation (i.e. equation (4.3)) and the closure equation (i.e. equation (4.5)) may be written in terms of $\delta(z)$ as

$$\delta(z) = \rho_b \int c_b(|r - r'|) g(z') dr' \quad (\text{B2})$$

and

$$g(z) = \exp \left[-\beta u(z) + \delta(z) + \beta \left(\frac{\partial p}{\partial \rho} \right)_\beta - 1 + b(z) \right] \quad (\text{B3})$$

The required bulk-fluid quantities can be obtained by solving the bulk-fluid OZ equation with an appropriate closure as shown in appendix A above. Taking the one-dimensional Fourier transform of equation (B2) gives

$$\tilde{\delta}(k_j) = \rho_b \tilde{c}_b(k_j) \tilde{g}(k_j) \quad (\text{B4})$$

where

$$\tilde{c}_b(k_j) = \frac{4\pi}{j\Delta k} \Delta r \sum_{i=1}^{N-1} r_i c_b(r_i) \sin\left(\frac{\pi}{N} ij\right), j = 1, 2, \dots, N-1 \quad (\text{B5})$$

$$\tilde{g}(k_j) = 2\Delta z \sum' g(z_i) \cos(k_j z_i) \quad (\text{B6})$$

Defining $\Delta\tilde{\delta}(k_j) = \tilde{\delta}(k_j) - \tilde{\delta}^0(k_j)$ and expanding equation (B3) to the first order, then taking Fourier transform and combining with equation (B4) we have

$$\Delta\tilde{\delta}(k_j) = -\tilde{\delta}^0(k_j) + \rho_b c_b(k_j) \left\{ \tilde{g}^0(k_j) + \frac{\Delta k}{\pi} \sum' [\tilde{D}^0(k_j - l_n) + \tilde{D}^0(k_j + l_n)] \Delta\tilde{\delta}(l_n) \right\} \quad (\text{B7})$$

where $D^0(z)$ is a derivative depending on particular closure used for $b(z)$

$$\tilde{D}^0(k_j) = \Delta z \sum D^0(z_i) \cos(k_j z_i) \quad (\text{B8})$$

$$D^0(z_i) \equiv \left(\frac{\partial g(z_i)}{\partial \delta(z_i)} \right)^0 = g^0(z_i) \left[1 + \left(\frac{\partial b(z_i)}{\partial \delta(z_i)} \right)^0 \right] \quad (\text{B9})$$

and the symbol $\sum'$ denotes

$$\sum' f(x_i) = \frac{1}{2} f(x_0) + f(x_1) + \cdots + f(x_{N-1}) + \frac{1}{2} f(x_N) \quad (\text{B10})$$

The numerical solution procedure can be summarized as following: (1)

Starting from an initial estimate $\tilde{\delta}^0(k_j), j=0,1,\dots,N$, the initial $\delta(z_i)$ is calculated by

$$\delta(z_i) = \frac{\Delta k}{\pi} \sum' \tilde{\delta}(k_j) \cos(k_j z_i) \quad (\text{B11})$$

(2) Then $g^0(z_i)$ and $D^0(z_i)$ are obtained from equations (B3) and (B9), respectively, and their Fourier transforms $\tilde{g}^0(k_j)$ and $\tilde{D}^0(k_j)$ from equations (B6) and (B8). (3) For the next NR iteration, choose M , the first M Fourier components $\tilde{\delta}(k_j)$ are obtained by setting $\Delta\tilde{\delta}(k_j)=0, j \geq M$ and solving the first M components of equation(B7), and remaining components are obtained from equation (B4). (4) This procedure is repeated until convergence is considered to be obtained using the criterion

$$\left\{ \sum_j \Delta\tilde{\delta}(k_j)^2 / N \right\}^{1/2} \leq 10^{-5}, j=0,1,\dots,N \quad (\text{B12})$$

Appendix C

PSVS Results and Discussions

Ramachandran plot is a way to visualize dihedral angles ψ against ϕ of amino acid residues in protein structure. It shows the possible conformations of ψ and ϕ angles for a peptide. From the Table C. I, we can see that the models HBD-28_B and HBD-28_C have some unfavorable conformations for some residues, while all the ψ - ϕ conformational angles in models HBD-28_A and HBD-28_D are reasonable.

Table C. I. The Ramachandran statistics of initial HBD-28 models

Model	<i>Ramachandran statistics from Richardson</i>			
	<i>Lab's Molprobit</i>			
	Favored regions	Allowed regions	Disallowed regions	Disallowed regions
<i>HBD-28_A</i>	85.7%	14.3%	0%	
<i>HBD-28_B</i>	88.6%	8.6%	2.9%	12:GLY
<i>HBD-28_C</i>	85.7%	11.4%	2.9%	30:SER
<i>HBD-28_D</i>	88.6%	11.4%	0%	

Z-scores of Verify-3D and ProsaII are particularly useful in discrimination between homology models with wrong folds from those that are correctly folded. Homology models with higher Verify-3D Z-score and ProsaII Z-score indicate better folded ones. From the Table C. II, we can see that models HBD-28_A and HBD-28_D generally yield higher scores than models HBD-28_B and HBD-

HBD-28_C and the model HBD-28_D has the highest Z-scores. It is difficult to elucidate the accuracy of models by MolProbity clash and Procheck G-factor as Z-scores from both are comparable in quantity. Nevertheless, overall results Ramachandran statistics, Verify-3D and ProsaII indicate that the quality of models HBD-28_A and HBD-28_D is better than that of models HBD-28_B and HBD-28_C.

Table C. II. The PROCHECK G-factor, MolProbity clashscore and ProsaII Z-scores of initial HBD-28 models. Note that all data are Z-score values calculated based on a set of 252 X-ray structures < 500 residues, of resolution 1.80 Å, R-factor <= 0.25 and R-free <= 0.28.

Model	<i>Verify-3D*</i>	<i>ProsaII*</i>	<i>Procheck G-factor*</i>	<i>MolProbity clash*</i>
<i>HBD-28_A</i>	-0.96	-0.17	-3.49	-4.0
<i>HBD-28_B</i>	-4.98	-2.44	-1.48	-8.54
<i>HBD-28_C</i>	-3.69	-1.45	-2.90	-3.71
<i>HBD-28_D</i>	-1.61	-0.95	-3.25	-5.05

REFERENCES

1. Balescu, R., *Equilibrium and Nonequilibrium Statistical Mechanics*, 1975, New York: John Wiley and Sons.
2. Hansen, J.P. and I.R. McDald, *Theory of Simple Liquids*, 2nd ed., 1986, London.
3. Alder, B.J., S.P. Frankel, and V.A. Lewinson, *Radial distribution function calculated by the Monte Carlo method for a hard sphere fluid*. J. Chem. Phys., 1955. **23**(3): p. 417-419.
4. Barker, J.A. and Henderso.D, *Monte Carlo values for the radial distribution function of a system of fluid hard spheres*. Mol. Phys., 1971. **21**(1): p. 187-191.
5. Kolafa, J., S. Labik, and A. Malihevsky. *Radial distribution function of the hard sphere fluid*. in *IUPAC International Conference on Chemical Thermodynamics*. 2004. Beijing, China.
6. Waisman, E., *The radial distribution function for a fluid of hard spheres at high densities. Mean spherical integral equation approach*. Mol. Phys., 1973. **25**(1): p. 45-48.
7. Wehner, M.F. and W.G. Wolfer, *A new integral equation for the radial distribution function of a hard sphere fluid*. J. Stat. Phys., 1986. **42**(3-4): p. 493-508.
8. M.-E. and Boudh-Hir, *Retention of the bridge diagrams in the pair-distribution function's integral equation*. Physica A: Statistical Mechanics and its Applications, 1992. **181**(3-4): p. 297-308.
9. Wertheim, M.S., *Exact solution of the percus-yevick integral equation*

- hard spheres*. *Phy. Rev. Lett.*, 1963. **10**(8): p. 321.
10. Wertheim, M.S., *Analytic solution of the Percus-Yevick equation*. *J. Math. Phys.*, 1964. **5**(5): p. 643.
 11. Caccamo, C., *Integral equation theory description of phase equilibria in classical fluids*. *Phys. Rep.*, 1996. **274**(1-2): p. 1-105.
 12. Henderson, D., K.-Y. chan, and L. Degreve, *The direct correlation functions and bridge functions for hard spheres near a large hard sphere*. *J. Chem. Phys.*, 1994. **101**(8): p. 6975-6978.
 13. Henderson, D.J. and S. Sokolowski, *Hard-sphere bridge function calculated from a second-order Percus–Yevick approximation*. *J. Chem. Phys.*, 1995. **102**(17): p. 7541-7544.
 14. Bomont, J.M. and J.L. Bretonnet, *A new approximate bridge function for pure fluids*. *Mol. Phys.*, 2003. **101**(22): p. 3249-3261.
 15. Kunor, T.R. and S. Taraphder, *Bridge functions near the liquid-vapor coexistence curve in binary Lennard-Jones mixtures*. *Phy. Rev. E*, 2006. **74**(1): p. 011201.
 16. Cheung, D.L., et al., *Structure of molecular liquids: Cavity and bridge functions of the hard spheroid fluid*. *Phy. Rev. E*, 2006. **73**(6): p. 061204.
 17. Malijevsky, A. and S. Labik, *Test of simple fluid theories for the Lennard-Jones system*. *Czech. J. Phys.*, 1988. **38**(3): p. 250-256.
 18. Lado, F., *A local thermodynamic criterion for the reference-hypernetted chain equation*. *Phys. Lett. A*, 1982. **89**(4): p. 196-198.
 19. Rosenfeld, Y. and N.W. Ashcroft, *Mean-spherical model for soft potentials: The hard core revealed as a perturbation*. *Phys. Rev. A*, 1979. **20**(5): p. 2162-2169.

20. Kolafa, J., S. Labik, and A. Malijevsky, *The bridge function of hard spheres by direct inversion of computer simulation data*. Mol. Phys., 2002. **100**(16): p. 2629-2640.
21. Verlet, L., *Integral equations for classical fluids. I. The hard sphere case*. Mol. Phys., 1980. **41**(1): p. 183-190.
22. Grundke, E.W. and Henderso.D, *Distribution functions of multi-component fluid mixtures of hard spheres*. Mol. Phys., 1972. **24**(2): p. 269-281.
23. Malijevsky, A. and S. Labik, *The bridge function for hard-spheres*. Mol. Phys., 1987. **60**(3): p. 663-669.
24. Rast, S., P.H. Fries, and H. Krienke, *A new biased Monte-Carlo method for computing coefficients of the bridge functions of liquids*. Mol. Phys., 1999. **96**(10): p. 1543-1557.
25. Lee, L.L., *An accurate integral equation theory for hard spheres: Role of the zero-separation theorems in the closure relation*. J. Chem. Phys., 1995. **103**(21): p. 9388.
26. Perkyns, J.S., K.M. Dyer, and B.M. Pettitt, *Computationally useful bridge diagram series. II. Diagrams in h-bonds*. J. Chem. Phys., 2002. **116**(21): p. 9404.
27. Labik, S., et al., *Calculation of elementary diagrams using a Metropolis-like simulation method*. Mol.Phys., 2003. **101**(8): p. 1139-1146.
28. Singh, J.K. and D.A. Kofke, *Mayer sampling: calculation of cluster integrals using free-energy perturbation methods*. Phy. Rev. Lett., 2004. **92**(22): p. 220601.

29. Kwak, S.K. and D.A. Kofke, *Evaluation of bridge-function diagrams via Mayer-sampling Monte Carlo simulation*. J. Chem. Phys., 2005. **122**(10): p. 104508.
30. Errington, J.R., *Direct calculation of liquid-vapor phase equilibria from transition matrix Monte Carlo simulation*. J. Chem. Phys., 2003. **118**(22): p. 9915.
31. Benjamin, K.M., A.J. Schultz, and D.A. Kofke, *Gas-phase molecular clustering of TIP4P and SPC/E water models from higher-order virial coefficients*. Ind. Eng. Chem. Res, 2006. **45**(16): p. 5566-5573.
32. Benjamin, K.M., et al., *Higher-order virial coefficients of water models*. J. Phys. Chem. B, 2007. **111**(39): p. 11463-11473.
33. Henderson, D., F.F. Abraham, and J.A. Barker, *The Ornstein-Zernike equation for a fluid in contact with a surface*. Mol. Phys., 1976. **31**(4): p. 1291-1295.
34. Giaquinta, P.V. and M. Parrinello, *Pair correlations at an electrified Interface* J. Chem. Phys., 1984. **81**(9): p. 4074-4079.
35. Plischke, M. and D. Henderson, *Pair correlation of hard spheres near a hard wall*. J. Phys. Chem., 1984. **88**(26): p. 6544-6547.
36. Rickayzen, G. and A. Augousti, *Integral equations and the pressure at the liquid-solid interface*. Mol. Phys., 1984. **52**(6): p. 1355-1366.
37. Sokolowski, S., *Adsorption of hard spheres via the non-uniform Percus-Yevick equation*. Mol. Phys., 1983. **49**(6): p. 1481-1488.
38. Zhou, Y. and G. Stell, *The hard-sphere fluid: New exact results with applications*. J. Stat. Phys., 1988. **52**(5-6): p. 1389-1412.
39. Attard, P., *Spherically inhomogeneous fluids. I. Percus-Yevick hard*

- spheres: Osmotic coefficients and triplet correlations.* J. Chem. Phys., 1989. **91**(5): p. 3072-3082.
40. Henderson, D. and M. Plischke, *Pair correlation function in a fluid with density inhomogeneities: parametrization for hard spheres near a hard wall.* Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences, 1985. **400**(1818): p. 631-674.
 41. Ballone, P., et al., *Additive and non-additive hard sphere mixtures.* Mol. Phys., 1986. **59**(2): p. 275-290.
 42. Labik, S., A. Malijevsky, and W.R. Smith, *An accurate integral equation for molecular fluids. I. Hard homonuclear diatomics.* Mol. Phys., 1991. **73**(1): p. 87-98.
 43. Foiles, S.M., N.W. Ashcroft, and L. Reatto, *A crossover integral equation for the structure of simple liquids.* J. Chem. Phys., 1984. **80**(9): p. 4441-4447.
 44. Bomont, J.M. and J.L. Bretonnet, *A self-consistent integral equation: Bridge function and thermodynamic properties for the Lennard-Jones fluid.* J. Chem. Phys., 2003. **119**(4): p. 2188-2191.
 45. Bomont, J.M., N. Jakse, and J.L. Bretonnet, *Self-consistent determination of the bridge function for simple fluids with the Lennard-Jones potential.* J. Chem. Phys., 1997. **107**(19): p. 8030-8035.
 46. Vompe, A.G. and G.A. Martynov, *The bridge function expansion and the self-consistency problem of the Ornstein-Zernike equation solution.* J. Chem. Phys., 1994. **100**(7): p. 5249-5258.
 47. Rogers, F.J. and D.A. Young, *New, thermodynamically consistent, integral equation for simple fluids.* Phys. Rev. A, 1984. **30**(2): p. 999-

- 999-1007.
48. Martynov, G.A. and G.N. Sarkisov, *Exact equations and the theory of liquids*. Mol. Phys., 1983. **49**(6): p. 1495-1504.
 49. Samec, A., *Integral equations for classical fluids using bridge diagrams: Hard-sphere solutions*. J. Chem. Phys., 1971. **54**(4): p. 2178-2184.
 50. Kellam, J. and S. Labik, *Density expansion of the radial distribution and bridge functions of the hard sphere fluid*. Mol. Phys., 2006. **104**(12): p. 1915-1924.
 51. Zhou, Y.Q. and G. Stell, *Fluids inside a pore—an integral-equation approach I. General formalism and hard spheres inside spherical and slit pores* Mol. Phys., 1989. **66**(4): p. 767-789.
 52. Pospisil, R., et al., *Integral equation and computer simulation studies of hard spheres in a slit pore* Mol. Phys., 1993. **78**(6): p. 1461-1474.
 53. Alejandre, J., M. Lozada-Cassou, and L. Degreve, *Effect of pore geometry on a confined hard sphere fluid*. Mol. Phys., 1996. **88**(5): p. 1317-1336.
 54. Karanikas, S., et al., *Density profiles and solvation forces for a Yukawa fluid in a slit pore*. J. Chem. Phys., 2008. **128**(20): p. 204704.
 55. Broyles, A.A., *Radial distribution functions from the born-green integral equation*. J. Chem. Phys., 1960. **33**(2): p. 456.
 56. Lado, F., *Pressure-consistent integral equation for classical fluids: Hard-sphere solutions*. J. Chem. Phys., 1967. **47**(11): p. 4828.
 57. Duh, D.-M. and A.D.J. Haymet, *Integral equation theory for uncharged liquids: The Lennard-Jones fluid and the bridge function*. J. Chem. Phys., 1995. **103**(7): p. 2625.

58. Watts, R.O., *Specialist Periodical Report of Chemical Society. Statistival Mechanics*, 1973, London: British Chemical Society.
59. Watts, R.O., *Percus-yevick equation applied to a lennard-jones fluid*. J. Chem. Phys., 1968. **48**(1): p. 50.
60. Gillan, M.J., *A new method of solving the liquid structure integral equation*. Mol. Phys., 1979. **38**(6): p. 1781-1794.
61. Labik, S., A. Malijevsky, and P. Vonka, *A rapidly convergent method of solving the OZ equation*. Mol. Phys., 1985. **56**(3): p. 709-715.
62. Zerah, G., *An efficient Newton's method for the numerical solution of fluid integral equations*. J. Comput. Phys., 1985. **61**(2): p. 280-285.
63. Kinoshita, M. and M. Harada, *Numerical solution of the HNC equation for ionic systems*. Mol. Phys., 1988. **65**(3): p. 599-618.
64. Lomba, E., *An efficient procedure for solving the reference hypernetted chain equation (RHNC) for simple fluids. Illustrative results with application to phase coexistence for a Lennard-Jones fluid*. Mol. Phys., 1989. **68**(1): p. 87-95.
65. Stell, G., in *The Equilibrium Theory of Classical Fluids*, edited by Frisch, H.L. and Lebowitz, J.L., 1964, New York: Benjamin.
66. Frenkel, D. and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2002, San Diego: Academic Press.
67. Metropolis, N., et al., *Equation of state calculations by fast computing machines*. J. Chem. Phys., 1953. **21**(6): p. 1087.
68. Fitzgerald, M., R.R.Picard, and R.N.Silver, *Canonical transition probabilities for adaptive Metroplis simulation*. Europhys. Lett, 1999. **46**(3): p. 282.

69. Norris, J.R., *Markov Chain*, 1997, Cambridge: Cambridge University Press.
70. Shen, V.K. and J.R. Errington, *Metastability and Instability in the Lennard-Jones Fluid Investigated by Transition-Matrix Monte Carlo*. J. Chem. Phys., 2004. **108**(51): p. 19595-19606.
71. Shen, V.K. and J.R. Errington, *Determination of fluid-phase behaviour using transition-matrix Monte Carlo: Binary Lennard-Jones mixtures*. J. Chem. Phys., 2005. **122**(6): p. 064508.
72. Shen, V.K. and J.R. Errington, *Determination of surface tension in binary mixtures using transition-matrix Monte Carlo*. J. Chem. Phys., 2006. **124**(2): p. 024721.
73. Cichowski, E.C., T.R. Schmidt, and J.R. Errington, *Determination of Hery's law constants through transition matrix monte carlo simulation*. Fluid Phase Equilib., 2005. **236**(1-2): p. 58-65.
74. Singh, J.K. and J.R. Errington, *Calculation of phase coexistence properties and surface tensions of n-Alkanes with grand-canonical transition-matrix Monte Carlo simulation and finite-size scaling*. J. of Chem. Phys., 2006. **110**(3): p. 1369-1376.
75. Kofke, D.A. and D. Frenkel, *Handbook of Materials Modeling, Perspective: Free Energies and Phase Equilibria*, 2005, Netherlands: Springer Netherlands.
76. Nijboer, B.R.A. and L. Van Hove, *Radial distribution function of a gas of hard spheres and the superposition approximation*. Phys. Rev., 1952. **85**(5): p. 777-783.
77. Attard, P. and G.N. Patey, *Hypernetted-chain closure with bridge*

- diagrams. Asymmetric hard sphere mixtures.* J. Chem. Phys., 1990. p. 4970-4982.
78. Dyer, K.M., J.S. Perkyns, and B.M. Pettitt, *Effective density terms in proper integral equations.* J. Chem. Phys., 2005. **123**(20): p. 204512.
 79. Dyer, K.M., et al., *A molecular site-site integral equation that yields the dielectric constant.* J. Chem. Phys., 2008. **129**(10): p. 104512.
 80. Henderson, D., *Fundamentals of Inhomogeneous Fluids*, 1992, New York: CRC Press.
 81. Kofke, D.A., et al. *A java-based development environment for molecular simulations.* in *PPEPPD*. 2004. Snowbird, UT, USA.
 82. Kofke, D.A. and B.C. Mihalick, *Web-based technologies for teaching and using molecular simulation.* Fluid Phase Equilib., 2002. **194**: p. 327-335.
 83. Llano-Restrepo, M. and W.G. Chapman, *Bridge function and cavity correlation function for the Lennard-Jones fluid from simulation.* J. Chem. Phys., 1992. **97**(3): p. 2046-2054.
 84. Essmann, U., et al., *A smooth particle mesh Ewald method.* J. Chem. Phys., 1995. **103**(19): p. 8577-8593.
 85. Taylor, M.P. and J.E.G. Lipson, *On the hard sphere bridge function* J. Chem. Phys., 1992. **97**(10): p. 7851-7852.
 86. Bennett, C.H., *Efficient estimation of free energy differences from Monte Carlo data.* J. Comput. Phys., 1976. **22**(2): p. 245-268.
 87. Kofke, D.A., *Free energy methods in molecular simulation.* Fluid Phase Equilib., 2005. **228-229**: p. 41-48.
 88. Press, W.P., et al., *Numerical Recipes: The Art of Scientific Computing*,

2nd ed., 1992, Cambridge: Cambridge University Press.

89. Carnahan, N.F. and K.E. Starling, *Equation of state for non-attracting rigid spheres*. J. Chem. Phys., 1969. **51**(2): p. 635.
90. Alejandre, J., et al., *Molecular dynamics of a hard-sphere fluid between two walls: a comparison with the three-point extension hypernetted chain approximation*. Chem. Phys. Lett., 1990. **175**(1-2): p. 111-116.
91. Kjellander, R. and S. Sarman, *On the statistical mechanics of inhomogeneous fluids in narrow slits. An application to a hard-sphere fluid between hard walls*. Chem. Phys. Lett., 1988. **149**(1): p. 102-108.
92. Kjellander, R. and S. Sarman, *Pair correlations of non-uniform hard-sphere fluids in narrow slits and the mechanism of oscillatory solvation forces*. J. Chem. Soc., Faraday Trans., 1991. **87**(12): p. 1869-1881.
93. Karlstrom, G., *Solvation forces studied by Grand Canonical Monte Carlo simulations of hard spheres between hard walls*. Chemica Scripta, 1985. **25**(1): p. 89-91.
94. Labik, S. and W.R. Smith, *Computer simulation of the chemical potential and adsorption isotherm of hard spheres in a hard slit-like pore*. Mol. Phys., 1996. **88**(5): p. 1411-1418.
95. Lupkowski, M. and F. Vanswol, *Computer simulation of fluids interacting with fluctuating walls*. J. Chem. Phys., 1990. **93**(1): p. 737-745.
96. Macelroy, J.M.D. and S.H. Suh, *Computer simulation of moderately dense hard-sphere fluids and mixtures in microcapillaries*. Mol. Phys., 1987. **60**(2): p. 475-501.

97. Macpherson, A.K., Y.P. Carignan, and T. Vladimiroff, *Molecular dynamics of hard spheres. I. Hard spheres in a channel*. J. Chem. Phys., 1987. **86**(7): p. 4228-4232.
98. Peterson, B.K. and K.E. Gubbins, *Phase transitions in a cylindrical pore grand canonical Monte Carlo, mean field theory and the Kelvin equation* Mol. Phys., 1987. **62**(1): p. 215-226.
99. Schoen, M., D.J. Diestler, and J.H. Cushman, *Fluids in micropores. I. Structure of a simple classical fluid in a slit-pore*. J. Chem. Phys., 1987. **87**(9): p. 5464-5476.
100. Lozada-Cassou, M., *The force between two planar electrical double layers*. J. Chem. Phys., 1984. **80**(7): p. 3344-3349.
101. Lozada-Cassou, M. and D. Henderson, *The force between two planar electrical double layers. Some numerical results* Chem. Phys. Lett., 1986. **127**(4): p. 392-397.
102. Chen, W.W. and S.K. Kwak, *Calculation of bridge function coefficients via modified Mayer-samplings*. Mol. Phys., 2009. **107**(20): p. 2213-2220.
103. ~~3120~~ 3120, P., *Hard-sphere fluids inside spherical, hard pores. Grand canonical ensemble Monte Carlo calculations and integral equation approximations*. J. Chem. Phys., 1990. **93**(2): p. 1292-1298.
104. Izadpanah, A. and R.L. Gallo, *Antimicrobial peptides*. J. Amer. Acad. Dermat., 2005. **52**(3): p. 381-392.
105. Harder, J., et al., *Identification and gene expression of antimicrobial proteins in human keratinocytes*. J. Invest. Dermat., 1997. **109**(3): p. 458-458.
106. Harder, J., et al., *Isolation and characterization of human beta-defensin-*

- a novel human inducible peptide antibiotic*. J. Biol. Chem., 2001. p. 5707-5713.
107. Pazgier, M., et al., *Human beta-defensins*. Cell. Mol. Life Sci., 2006. **63**(11): p. 1294-1313.
 108. Taylor, K., P.E. Barran, and J.R. Dorin, *Review: Structure-activity relationships in beta-defensin peptides*. Biopolymers, 2008. **90**(1): p. 1-
 109. Campopiano, D.J., et al., *Structure-activity relationships in defensin dimers - A novel link between beta-defensin tertiary structure and antimicrobial activity*. J. Biol. Chem., 2004. **279**(47): p. 48671-48679.
 110. Garcia, J.R.C., et al., *Identification of a novel, multifunctional beta-defensin (human beta-defensin 3) with specific antimicrobial activity - Its interaction with plasma membranes of Xenopus oocytes and the induction of macrophage chemoattraction*. Cell Tissue Res., 2001. **306**(2): p. 257-264.
 111. Chandrababu, K.B., B. Ho, and D.W. Yang, *Structure, dynamics, and activity of an all-cysteine mutated human beta defensin-3 peptide analogue*. Biochemistry, 2009. **48**(26): p. 6052-6061.
 112. Liu, S.P., et al., *Linear analogues of human beta-defensin 3: Concepts for design of antimicrobial peptides with reduced cytotoxicity to mammalian cells*. Chembiochem, 2008. **9**(6): p. 964-973.
 113. Rodriguez-Jimenez, F.J., et al., *Distribution of new human beta-defensin genes clustered on chromosome 20 in functionally different segments of epididymis*. Genomics, 2003. **81**(2): p. 175-183.
 114. Yount, N.Y., et al., *Advances in antimicrobial peptide immunobiology*. Biopolymers, 2006. **84**(5): p. 435-458.

115. Selsted, M.E. and A.J. Ouellette, *Mammalian defensins in the antimicrobial immune response*. Nat. Immunol., 2005. **6**(6): p. 551-557.
116. Bensch, K.W., et al., *Hbd-1 - a Novel Beta-Defensin from Human Plasma*. FEBS Lett., 1995. **368**(2): p. 331-335.
117. Wu, Z.B., et al., *Engineering disulfide bridges to dissect antimicrobial and chemotactic activities of human beta-defensin 3*. Proc. Natl. Acad. Sci. U. S. A., 2003. **100**(15): p. 8880-8885.
118. Mandal, M., M.V. Jagannadham, and R. Nagaraj, *Antibacterial activities and conformations of bovine beta-defensin BNBD-12 and analogs: structural and disulfide bridge requirements for activity*. Peptides, 2002. **23**(3): p. 413-418.
119. Garcia, J.R.C., et al., *Human beta-defensin 4: a novel inducible peptide with a specific salt-sensitive spectrum of antimicrobial activity*. FASEB J., 2001. **15**(8): p. 1819-21.
120. Jia, H.P., et al., *Discovery of new human beta-defensins using a genomics-based approach*. Gene, 2001. **263**(1-2): p. 211-8.
121. Schroder, J.M. and J. Harder, *Human beta-defensin-2*. Int. J. Biochem. Cell Biol., 1999. **31**(6): p. 645-51.
122. Berman, H.M., et al., *The protein data bank*. Nuc. Acid. Res., 2000. **28**(1): p. 235-242.
123. Hoover, D.M., O. Chertov, and J. Lubkowski, *The structure of human beta-defensin-1 - New insights into structural properties of beta-defensins*. J. Biol. Chem., 2001. **276**(42): p. 39021-39026.
124. Hoover, D.M., et al., *The structure of human beta-defensin-2 shows evidence of higher order oligomerization*. J. Biol. Chem., 2000.

- p. 32911-32918.
125. Schibli, D.J., et al., *The solution structures of the human beta-defensins lead to a better understanding of the potent bactericidal activity of HBD3 against staphylococcus aureus*. J. Biol. Chem., 2002. **277**(10): p. 8279-8289.
 126. Anfinsen, C.B., *Principles that govern the folding of protein chains*. Science, 1973. **181**(4096): p. 223-230.
 127. Jenny, G. and E.B. Philip, *Homology modeling in structural bioinformatics*, 2nd ed. 2009, Canada: John Wiley & Sons, Inc., Hoboken, N. J.
 128. Siqueira, A.M., et al., *A proposed 3D structure for crotonine based on homology building, molecular simulations and circular dichroism*. J. Mol. Graphics Modell., 2002. **20**(5): p. 389-398.
 129. Xie, J., P. Butko, and D.X. Xie, *Molecular dynamics simulations of the cytolytic toxin Cyt1A in solution*. IEEE Trans. Nanobiosci., 2005. **4**(3): p. 235-240.
 130. Cheng, M.H.Y., M. Cascio, and R.D. Coalson, *Homology modeling and molecular dynamics simulations of the alpha 1 glycine receptor reveals different states of the channel*. Prot. Struct. Func. Bioinf., 2007. **68**(2): p. 581-593.
 131. Speranskiy, K., M. Cascio, and M. Kurnikova, *Homology modeling and molecular dynamics simulations of the glycine receptor ligand binding domain*. Prot. Struct. Func. Bioinf., 2007. **67**(4): p. 950-960.
 132. Zhao, Y.S., et al., *Homology modeling of human serine racemase and its molecular docking with peptide inhibitors*. Acta Phys. Chim. Sin., 2009.

- 25**(3): p. 417-422.
133. Manzetti, S., et al., *Modeling of enzyme-substrate complexes for the metalloproteases MMP-3, ADAM-9 and ADAM-10*. J. Comput. Aided Mol. Des., 2003. **17**(9): p. 551-565.
 134. Venclovas, C. and M. Margelevicius, *Comparative modeling in CASP6 using consensus approach to template selection, sequence-structure alignment, and structure assessment*. Prot. Struct. Func. Bioinf., 2005. **61**(7): p. 99-105.
 135. Chakravarty, S., et al., *Systematic analysis of the effect of multiple templates on the accuracy of comparative models of protein structure*. BMC Struct. Biol., 2008. **8**: p. 31.
 136. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints*. J. Mol. Biol., 1993. **234**(3): p. 779-815.
 137. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling*. Bioinformatics, 2006. **22**(2): p. 195-201.
 138. Vriend, G., *WHAT IF: A molecular modeling and drug design program*. J. Mol. Graphics, 1990. **8**(1): p. 52-&.
 139. Bhattacharya, A., R. Tejero, and G.T. Montelione, *Evaluating protein structures determined by structural genomics consortia*. Proteins, 2007. **66**(4): p. 778-795.
 140. Hess, B., et al., *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation*. J. Chem. Theo. Comp., 2008. **4**(3): p. 435-447.
 141. Schulz, A., et al., *Engineering disulfide bonds of the novel human beta-*

- beta-defensins hBD-27 and hBD-28: Differences in disulfide formation and biological activity among human beta-defensins.* Biopolymers, **80**(1): p. 34-49.
142. Tay, D.K.S., et al., *A new bioproduction route for a novel antimicrobial peptide.* Biotechnol. Bioeng., 2011. **108**(3): p. 572-581.
 143. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nuc. Acid. Res., 1994. **22**(22): p. 4673-4680.
 144. <http://accelrys.com/products/discovery-studio/>.
 145. Bjelkmar, P., et al., *Implementation of the CHARMM force field in GROMACS: analysis of protein stability effects from correction maps, virtual interaction sites, and water models.* J. Chem. Theo. Comp., 2010. **6**(2): p. 459-466.
 146. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water.* J. Chem. Phys., 1983. **79**(2): p. 926-935.
 147. Chitra, R. and P.E. Smith, *A comparison of the properties of 2,2,2-trifluoroethanol and 2,2,2-trifluoroethanol/water mixtures using different force fields.* J. Chem. Phys., 2001. **115**(12): p. 5521-5530.
 148. Chitra, R. and P.E. Smith, *Properties of 2,2,2-trifluoroethanol and water mixtures.* J. Chem. Phys., 2001. **114**(1): p. 426-435.
 149. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes.* J. Comp. Phys., 1977. **23**(3): p. 327-341.

150. Miyamoto, S. and P.A. Kollman, *Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models*. J. Comp. Chem., 1992. **13**(8): p. 952-962.
151. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. J. Chem. Phys., 2007. **126**(1): p. 014101.
152. Nose, S. and M.L. Klein, *Constant pressure molecular dynamics for molecular systems*. Mol. Phys., 1983. **50**(5): p. 1055-1076.
153. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*. J. Appl. Phys., 1981. **52**(12): p. 7182-7190.
154. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features*. Biopolymers, 1983. **22**(12): p. 2577-2637.
155. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics*. J. Mol. Graphics, 1996. **14**(1): p. 33-38.
156. Poschner, B.C., et al., *An automated application for deconvolution of circular dichroism spectra of small peptides*. Anal. Biochem., 2007. **363**(2): p. 306-308.
157. Dalton, J.A.R. and R.M. Jackson, *An evaluation of automated homology modelling methods at low target-template sequence similarity*. Bioinformatics, 2007. **23**(15): p. 1901-1908.
158. Nayeem, A., D. Sitkoff, and S. Krystek, *A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models*. Protein Sci., 2006. **15**(4): p. 808-824.
159. Bowie, J.U., R. Luthy, and D. Eisenberg, *A method to identify protein*

- sequences that fold into a known three-dimensional structure. Science, 1991. 253(5016): p. 164-170.*
160. Jones, D.T., W.R. Taylor, and J.M. Thornton, *A new approach to protein fold recognition. Nature, 1992. 358(6381): p. 86-89.*
161. Maiti, R., et al., *SuperPose: A simple server for sophisticated structural superposition. Nuc. Acid. Res., 2004. 32(Web Server issue): p. W590-W594.*
162. Luthy, R., J.U. Bowie, and D. Eisenberg, *Assessment of protein models with three-dimensional profiles. Nature, 1992. 356(6364): p. 83-85.*

PUBLICATION LIST

Journal Papers

1. **W.W. Chen** and S.K. Kwak, "Calculation of Bridge Function Coefficients via Modified Mayer-Samplings", *Molecular Physics*, **107**, 2213-2220, (2009)
2. **W.W. Chen**, H.C. Huang and S.K. Kwak, "Evaluation of Bridge Function for Hard Sphere Fluids Confined in a Narrow Slit-Pore via TMMC Mayer-Sampling", *Molecular Physics*, **108**, 1531-1537, (2010)
3. H.C. Huang, **W.W. Chen**, J.K. Singh and S.K. Kwak, "Direct Determination of Fluid-Solid Coexistence of Square-Well Fluids Confined in Narrow Cylindrical Hard Pore", *Journal of Chemical Physics*, **132**, 224504, (2010)
4. **W.W. Chen**, Derrick Kok Sing Tay, Susanna Su Jan Leong and S.K. Kwak, "Three-Dimensional Structure of Human β -defensin 28 via Homology Modeling and Molecular Dynamics", *Molecular Simulation*, **38**, 90-101, (2012)

Conference Presentations

1. **W.W. Chen** and S.K. Kwak, "Bridge function for hard sphere fluid confined in slit ", AICHE Annual Meeting, 2009, Nashville, USA
2. **W.W. Chen**, Derrick Kok Sing Tay, Susanna Su Jan Leong and S.K. Kwak, "Homology Modeling and Molecular Dynamics Simulations of Human β -defensin 28", International Conference on Cellular and Molecular Bioengineering, 2010, Singapore
3. **W.W. Chen**, Derrick Kok Sing Tay, Susanna Su Jan Leong and S.K. Kwak, "Homology Modeling and Molecular Dynamics Simulations of Human β -

β -defensin 28", The International Symposium on Exploring the Frontiers of Chemical and Biomedical Engineering, 2010, Singapore

4. H.C. Huang, **W.W. Chen** and S.K. Kwak, "Characterization of Fluid-Solid Phase Transition of Cylindrically Confined Atomic Fluids", Pacific Basin Conference on Absorption Science and Technology (PBAST 5), 2009, Singapore
5. **W.W. Chen** and S.K. Kwak, "Bridge function for hard sphere fluid confined between planar hard walls", The International Symposium on Advanced Bio, Nano, and Pharmaceutical Science and Technology, 2009, Beijing, China
6. **W.W. Chen** and S.K. Kwak, "Evaluation of Bridge Diagrams via Overlap-Mayer Sampling: Hard Sphere Fluids", 15th Regional Symposium on Chemical Engineering in conjunction with 22nd Symposium of Malaysian Chemical Engineers , 2008, Kuala Lumpur, Malaysia