Simulation study of antimicrobial agents : theories, properties, and interactions with biomembrane

Liu, Yang

2018

Liu, Y. (2018). Simulation study of antimicrobial agents : theories, properties, and interactions with biomembrane. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/87817>

<https://doi.org/10.32657/10220/46935>

**SIMULATION STUDY OF ANTIMICROBIAL AGENTS:
THEORIES, PROPERTIES, AND INTERACTIONS WITH
BIOMEMBRANE**

LIU YANG

SCHOOL OF BIOLOGICAL SCIENCES

2018

**SIMULATION STUDY OF ANTIMICROBIAL AGENTS:
THEORIES, PROPERTIES, AND INTERACTIONS WITH
BIOMEMBRANE**

LIU YANG

LIU YANG

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological
University in partial fulfillment of the requirement for the
degree of Doctor of Philosophy

2018

Acknowledgments

First and foremost, I would like to give my thanks to my supervisor Assoc. Prof. Mu Yuguang, an affable and erudite person. This thesis cannot be finished without his informative instructions and suggestions.

Second, I want to thank my co-supervisor Prof. Mary B. Chan-Park for her guidance and support during the four-year study.

In addition, my thanks also go to my colleagues in the computational biology lab, for their helpful discussions and suggestions in the group meeting.

I also want to thank the School of Biological Sciences, Nanyang Technological University for the financial support of my PhD project.

Last but not least, I would like to express my special thanks to my family for supporting me spiritually.

Contents

List of Figures	I
List of Tables.....	III
Abbreviations	V
Abstract.....	VII
1 Introduction	1
1.1 Traditional antibiotics: targets and resistance	2
1.1.1 Main targets of traditional antibiotics	2
1.1.2 Mechanisms of drug-resistance	3
1.2 Bacterial envelope targeting antibiotics.....	5
1.2.1 Mammalian cell membrane	5
1.2.2 Bacterial cell envelope	6
1.2.3 Bacterial envelope targeting antibiotics and their modes of action.....	9
1.3 From mechanism study to rational design	14
2 Methodology.....	17
2.1 Conventional MD simulation	17
2.1.1 Force field.....	18
2.1.2 Simulation ensembles	19
2.2 Enhanced sampling methods.....	20
2.2.1 MetaD and well-tempered MetaD.....	21
2.2.2 PT and PT-WTE	26
2.3 MM-PBSA calculations.....	29
3 Simulation study of a membrane targeting antibiotic - CSM5-K5	31
3.1 Background	31
3.2 Parameters.....	33
3.2.1 General parameters	33
3.2.2 Well-tempered MetaD simulation of CSM5-K5's dimerization	34
3.2.3 Conventional MD simulation of the CSM5-K5 and membrane system	35
3.3 Results and discussion	35
3.3.1 Dimerization of the CSM5-K5 monomers	35
3.3.2 Simulation of the CSM5-K5 with membrane systems.....	39
3.4 Conclusion.....	44
4 Simulation study of a L_{II} targeting antibiotic - TXB.....	47

4.1	Background	47
4.2	Parameters	49
4.2.1	Unbiased MD simulation details.....	49
4.2.2	PTMetaD-WTE	51
4.3	Results and discussion.....	53
4.3.1	Unbiased MD simulations of TXB and L _{II} binding with a ratio of 1:1.....	53
4.3.2	TXB-L _{II} binding modes explored by PTMetaD-WTE simulation	54
4.3.3	Prediction of TXB-L _{II} complex with a ratio of 2:1.....	66
4.4	Conclusion	72
5	The necessity of D-threonine in TXB	75
5.1	Background	75
5.2	Parameters	75
5.2.1	General parameters.....	75
5.2.2	Well-tempered MetaD simulations of the open ring motifs.....	77
5.2.3	Unbiased MD simulations of the TXB-L _{II} complex	78
5.2.4	PT-WTE simulation of the closed ring motifs	78
5.3	Results and discussion.....	79
5.3.1	Simulation of the open ring motifs.....	79
5.3.2	Simulation of the TXB-L _{II} complexes.....	82
5.3.3	Simulation of the closed ring motifs.....	85
5.4	Conclusion	89
6	Optimization of the temperature ladder in PT-WTE simulations.....	91
6.1	Background	91
6.2	Parameters	94
6.3	Results and discussion.....	95
6.3.1	Generation of temperature sequences and bias factors.....	95
6.3.2	Sampling efficiency of PT-WTE simulations with different bias factors.....	98
6.3.3	A smaller bias factor works better with our algorithm	100
6.3.4	Improvement of the performance of large bias factor	102
6.4	Conclusion	104
7	Summaries and recommendations	105
	References.....	109
	Author's publications	115

List of Figures

Figure 1.1 Head groups of different phospholipids in the mammalian cell membrane.....	5
Figure 1.2 Bacterial envelop of Gram-positive and Gram-negative bacteria.	7
Figure 1.3 Head groups of different phospholipids in the bacterial cell membrane.....	8
Figure 1.4 Schematic illustrations of bacterial killing mechanisms by AMPs.	10
Figure 1.5 Proposed bactericidal mechanism of daptomycin.....	14
Figure 1.6 Chemical structure of polynorbornenes (a) and polymethacrylates (b).....	15
Figure 2.1 Potential distributions of canonical ensembles (a) and the effects from WTE (b).	27
Figure 3.1 Two composition units of chitosan: N-acetyl-D-glucosamine (left) and D-glucosamine (right).	31
Figure 3.2 Molecular model used in this study.....	33
Figure 3.3 FESs of CSM5-K5's dimerization at different simulation time.	36
Figure 3.4 The representative structures and H-bonding architectures of minimum A and B.	37
Figure 3.5 Components of the CC H-bonds and stability of two minima.	38
Figure 3.6 Simulation of dimerized CSM5-K5 in the MM and IM membrane environments.	40
Figure 3.7 Area per lipid (a) and membrane thickness (b) distribution in different systems.....	41
Figure 3.8 Deuterium order parameters (S_{cd}) for two acyl chains of POPC and POPG in different systems.....	42
Figure 3.9 Lateral radial distribution function (RDF) of POPG lipids with respect to themselves in membrane systems with or without the presence of polycations.	43
Figure 3.10 Accumulation of POPG in the different IM-polycation system versus the control.....	44
Figure 4.1 Structure of simulated TXB (a) and L_{II} (b).....	48
Figure 4.2 Number of contacts between TXB and L_{II} in three unbiased MD simulation.	54
Figure 4.3 Diffusion of the number of contacts (a) and decay of the Gaussian height (b) as a function of time in the PTMetaD-WTE simulation.....	55
Figure 4.4 Diffusive movement of replicas in temperature space.....	55
Figure 4.5 Atom-based contact map (a) and the histogram of different clusters (b).....	56
Figure 4.6 Four binding modes and their featured contacts.	57
Figure 4.7 Residue based contact maps of the first 10 ns simulation and last 10 ns for the four binding modes.....	58
Figure 4.8 RMSD in the three repeats of stability-checking simulations for the four different binding modes.....	59
Figure 4.9 Distance between ring motif of TXB and phosphate group (C15P, left panel) or pyrophosphate group (C15PP and L_{II} , middle and right panel, respectively).	62
Figure 4.10 Binding energy of different binding modes and their comparison with experimental structures.	63
Figure 4.11 Stability of the nisin- L_{II} complex in 100 ns of unbiased MD simulations.	65
Figure 4.12 Contact map of the different TXB- L_{II} complex with a ratio of 2:1.....	67
Figure 4.13 Study of TXB dimerization based on 400 ns PTMetaD-WTE simulation of two TXB molecules.	68
Figure 4.14 Stability of the two TXB dimer models.....	69
Figure 4.15 Stability of the 2:1 TXB- L_{II} complex where the two TXB molecules dimerize first.	70
Figure 4.16 The 2:1 TXB- L_{II} complex in the membrane environment.	71

Figure 4.17 Contact map between two TXB molecules in the BM1&BM2 system.	72
Figure 5.1 Structures of the open and closed ring motif of TXB.	77
Figure 5.2 FES projections of O-DLLL and O-LLLL on two defined CVs.	80
Figure 5.3 Ramachandran plots of each residue in the O-DLLL and O-LLLL.	80
Figure 5.4 The most popular groups of O-DLLL (a) and O-LLLL (b).	81
Figure 5.5 Stability of the ring-pyrophosphate binding.	83
Figure 5.6 Ramachandran plot of each residue in the C-DLLL and C-LLLL in the 300K ensemble...	86
Figure 5.7 The most popular groups of C-DLLL (a) and C-LLLL (b).	87
Figure 5.8 Orientations of NH groups on the ring motif.	88
Figure 6.1 Thermodynamic properties of the system.	96
Figure 6.2 Conformational sampling of the two simulation series.	99
Figure 6.3 Distribution of the biased potential energy in S_{16} (a) and S_8 (b).	101
Figure 6.4 Improvement of S_8 simulation by extending the WTE-establishing time to 100 ns.....	103

List of Tables

Table 4.1 Details of the H-bonds maintaining the featured contacts in each binding mode.	61
Table 5.1 Occupancies of the H-bonds between the ring motif and pyrophosphate group in the D-Thr TXB-L _{II} complex.	84
Table 6.1 Calculated bias factors and temperature sequences for simulation series S ₁₆ and S ₈	97
Table 6.2 The RTTs for replicas in different simulation series (continuous trajectories).	98
Table 6.3 The AAPs for neighboring replica-pairs.	100

Abbreviations

MD	Molecular Dynamics
PT	Parallel Tempering
WTE	Well-Tempered Ensemble
MetaD	Meta-Dynamics
PT-WTE	Parallel Tempering simulation in the Well-Tempered Ensemble
PTMetaD-WTE	Parallel Tempering Meta-dynamics simulation in the Well-Tempered Ensemble
TXB	Teixobactin
L _{II}	Lipid II
NMR	Nuclear Magnetic Resonance
FES	Free Energy Surface
CV	Collective Variable
MM-PBSA	Molecular Mechanics Poisson–Boltzmann Surface Area
PME	Particle mesh Ewald
MM	Mammalian cell Membrane
IM	Inner Membrane
RMSD	Root Mean Square Deviation
APL	Area per Lipid
RDF	Radial Distribution Function
MIC	Minimum Inhibitory Concentration
SAR	Structure-Activity Relationships

RTT	Round Trip Time
O-DLLL	Open ring motif D-Thr-L-Ala-L-End-L-Ile
O-LLLL	Open ring motif L-Thr-L-Ala-L-End-L-Ile
C-DLLL	Closed ring motif D-Thr-L-Ala-L-End-L-Ile
C-LLLL	Closed ring motif L-Thr-L-Ala-L-End-L-Ile
AAP	Average Acceptance Probability

Abstract

The competition between the evolution of drug-resistant bacteria and the development of new antibiotics has never stopped since the first application of antimicrobial agent. Considering the long development cycle for a new agent and the invalidation of more and more old agents due to resistance, sustained investments are necessary to keep drug design leading in this competition. MD simulation as a useful tool for rational drug design reveals the drug-target interactions at an atom level, offering great assistance in understanding the killing mechanisms of antibiotics. Using MD simulations in combination with enhanced sampling methods, two novel antimicrobial agents and their self-assembly, binding modes with the targets, and conformational changes in solution were studied in this thesis. Furthermore, for a better application of one of the enhanced sampling methods, PT-WTE, we also came up with an algorithm to generate the temperature sequences and corresponding bias factors.

The first study case, CSM5-K5, which is synthesized by grafting polylysine chains onto a chitosan backbone, was experimentally proved to self-assemble in aqueous medium and function as a membrane active antibiotic. According to our MetaD simulation results, the $\text{--NHC(=O)}\cdots\text{--NHC(=O)}$ H-bonds between chitosan chains are strong enough to counteract the electrostatic repulsion introduced by the polylysine, and play a critical role in maintaining the dimerization of CSM5-K5. And as proved by the later conventional MD simulation, this dimerization also accelerates the lipid rearrangement process upon binding to the bacterial membrane, which will eventually cause membrane defects and lead to the cell death. CSM5-K5 presents a successful example of drug design by employing the polylysine chains to ensure the selectivity of bacteria over the mammalian cell and the

antimicrobial activity, while the chitosan chains to supply the aggregation mechanism to further enhance the effectiveness.

In the second case, the binding phase space of a recently discovered antibiotic, TXB, and its target, L_{II} , is investigated. Totally four binding modes were identified from this highly flexible binary complex ensemble generated by PTMetaD-WTE. Three of them involve the ring motif of TXB and have relatively higher binding affinities, indicating the importance of the ring motif of TXB in L_{II} recognition. These three binding modes where all the four NH groups on the ring form H-bonds with the negatively charged residues of L_{II} , were also demonstrated indirectly by a crystallographic structure of a TXB analogue binding to a chloride anion. Among the four amino acids on the ring motif, the chiral D-Thr residue is critical for the antimicrobial activity. We then performed MetaD and PT-WTE simulations to study the necessity of this chiral D-Thr in the binding with L_{II} . Our results suggest that different chirality leads to different NH orientation of Thr with respect to the ring plane. Only in the ring motif with D-Thr, a favored binding cavity is achievable with all the four NH groups pointing to the same side of the ring plane.

To improve the application of PT-WTE, we came up with a scheme to generate the temperature sequence and the corresponding bias factor optimally. The number of replicas for a complete coverage of a specific temperature range is adjustable in this scheme while keeping the AAP between neighboring replica-pairs unchanged. Two series of PT-WTE simulations were tested with the number of replicas as 16 and 8, and the corresponding bias factors as 37.9 and 174.2, respectively. However, the bias factor of 174.2 is oversized for a better application of PT-WTE.

1 Introduction

One of the most significant medical advances of the twentieth century is the development of antibiotics [1]. Initial use of antibiotics could be traced back centuries when people treated wounds with moldy foods without even knowing their antibacterial mechanisms [2], however, not until the early 1940s with the popularization of effective antimicrobials like penicillin [3], were the antibiotics applied to deal with bacterial infections widely and effectively. The race between the evolution of drug-resistant bacteria and the development of new antibiotics has never stopped since then. Unfortunately, last decades had not seen any discovery of new structurally or functionally distinct antibiotics from nature until Ling et al. made a breakthrough in the field with TXB [4]. Most of the antibiotics in the clinical application today were discovered in 1940-1970s, the so-called “golden period” of antibiotic discovery [5], however, some of them were resisted only within a few years because of the misuse [6]. The emergence of drug-resistant bacteria, especially the multi-drug resistant ones [7], poses a threat to global public health. There is no doubt that continued efforts are necessary for the development of new antibiotics.

One of the traditionally used strategies for antibiotic discovery by exploiting the cultivable soil microbes is largely limited due to their small proportion in the whole soil microbes (approximately %1) [8]. Luckily, the modern technological progress like the iChip [9], which offers a higher sensitivity than the conventional approaches in isolating the growing uncultured microorganisms [10], brings new hopes to the field. Another option is to turn to synthetic approaches for help. Antibiotics kill bacteria rather than mammalian cells by using the biochemical differences in, such as, metabolic processes or components

of the envelope between the prokaryotic and eukaryotic cells. By rational designs, the underlying interaction-patterns are able to be grafted to the synthesized molecules for antibacterial activity and selectivity, accelerating the paces of antibiotic discovery. This cannot be accomplished without a better understanding of the modes of action of the revealed antibiotics.

1.1 Traditional antibiotics: targets and resistance

1.1.1 Main targets of traditional antibiotics

Most of the traditional antibiotics kill bacteria by disturbing their essential biological processes, such as (1) replication of DNA, (2) synthesis of bacterial cell-wall, and (3) synthesis of protein [11]. Despite these diverse mechanisms of antibiotics, their targets are mainly proteins. For example, a category of type 1 antibiotics, quinolones, target the enzyme topoisomerase, which controls the chromosomal topology change in DNA replication [12]. Normally, topoisomerase introduces a pair of single-strand breaks into DNA for the subsequently bonding with DNA cleavages [13]. However, quinolones bind to form the DNA-topoisomerase-quinolone complex, trapping the enzyme on DNA strand and inhibiting the rejoining of DNA cleavages [14]. It is reported that the targeting topoisomerase varies across bacterial species: topoisomerase II in Gram-negative bacteria [15] and topoisomerase IV in Gram-positive bacteria [16], although both topoisomerases have similar functions. The bacterial cell wall plays critical roles in providing structural integrity of the cell. Its rigidity is mainly maintained by the peptidoglycan layers, where there are covalent cross-links through peptide chains of the peptidoglycan units [17]. One kind of the type 2 antibiotics, β -lactams, kills bacteria by attacking the enzyme responsible for the cross-linking stage, namely transpeptidase. Using its analogous structure with D-

analyl-D-alanine dipeptide of peptidoglycan unit, β -lactams bind preferentially to the transpeptidases [18,19], along with a covalent modification of the enzyme's active site, thereby disabling the enzyme [20]. The ribosome is involved in the process of protein synthesis. The type 3 antibiotics interfere with this process by targeting the ribonucleoprotein subunits, the 50S, and 30S. The 50S inhibitors either block the binding of initiator tRNA [21] or the translocation of peptidyl tRNAs [11]. By contrast, the 30S inhibitors either block the binding of aminoacyl tRNA to the ribosomes [22] or destabilize the binding of peptidyl tRNA [23,24]. There are also antibiotics targeting other components instead of proteins, which will be introduced in detail in the later section. These drug-target interactions provide cell recognition with a good starting point, and their primary effects intrigue the subsequent antibiotic-mediated cell death process.

1.1.2 Mechanisms of drug-resistance

It is almost inevitable that a small percentage of the microbial pathogens in a population develop their resistance sooner or later with the constant application of antibiotics, to confront the inhospitable environment and survive. Once such mutation happens, antibiotics will wipe out the major part of the bacterial population – the nonresistant ones, providing an ideal environment for the proliferation of the resistant ones, which would be the main components of the new and drug-resistant population [25]. Several countermeasures against the antibiotics have been discovered.

One of the common mechanisms is the synthesis of enzymes to destroy antibiotics before the binding with their targets. This is the strategy of developing resistance to the β -lactam: the penicillinase was identified even several years before its widespread clinical application [26]. Although β -lactams contain a variety of derivatives, the random gene

mutation in bacterial proliferation extends the spectra of resistance by enabling the modification of β -lactamase [27]. As another example, the surface protease is demonstrated to facilitate the resistance to antimicrobial peptides (AMPs). For example, the protease PgtE and OmpT cleave AMP with specificity for paired basic residues [28,29].

Besides the translation of diversiform protease enzymes, gene mutations also introduce resistance through the modification of antibiotics' targets. In such a manner, the antibiotics targeting the ribosome subunit S50 dysfunction with the modification of their binding site [30].

For the antibiotics targeting the bacterial envelope, pathogens develop mechanisms to vary their extracellular structures [31]. These inducible resistances include the modifications of lipid A [32,33], lipopolysaccharide (LPS) [34,35], cell wall teichoic acid [36,37], and expressions of mammalian-lipids-mimicking molecules [38] to circumvent the antibiotics binding. There are also reports of repelling the cationic AMPs through modifications to the cytoplasmic membrane. For instance, *Pseudomonas fluorescens* obtained resistance to polymyxin B in the phosphate-limiting medium by synthesizing positively charged ornithine amide lipid instead of the phosphate-containing lipids [39]. In another case, pathogens reduce the negative charge density on the membrane with a lysine modification to their anionic membrane lipids [40].

As mentioned above, there are no antibiotics could deal with bacteria once and for all, thus, continuous efforts are necessary to prevail the battle. In spite of the developed resistance to the envelope-targeting antibiotics, many components in the envelope still provide favorable binding motivation. Most of the agents bind to bacterial envelope via not a specific receptor but the charges dispersed on the surface based on the biophysical

nature, which is believed more difficult for resistance developing when compared to the protein-targeting antibiotics [25]. Antibiotic-discovery on the basis of differences between mammalian cells membrane and bacterial envelop remains to be a promising direction.

1.2 Bacterial envelope targeting antibiotics

1.2.1 Mammalian cell membrane

If we take no account of the membrane proteins and carbohydrates, the mammalian cell membrane is mainly composed of four major phospholipids: phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), and sphingomyelin (SM). Their head groups are shown in Fig. 1.1. They are asymmetrically distributed in the two leaflets of the membrane bilayer [41]: PC and SM predominate in the outer leaflet, whereas PS and PE prevail in the inner leaflet. Besides the phospholipids, cholesterol take a large mole proportion of the total lipids and are important for the membrane properties [42]. They

Figure 1.1 Head groups of different phospholipids in the mammalian cell membrane. R1 and R2 represent for the hydrocarbon chains.

order the PC chains and reduce area per lipid by forming small subunits with PC [43,44]. Among all the phospholipids in the mammalian cell membrane, PC accounts for more than 50% [45], thus the pure PC bilayers are frequently employed to model the mammalian cell membrane [46-48].

Different from the other three phospholipids in Fig. 1.1, who have net charges of zero, PS is negatively charged because of the covalently bonded serine group to the phosphate group. PS constitute about 20% of the phospholipids in inner leaflet [49,50], and function as the target of an assortment of proteins due to the negatively charged head group and unique structural features [51]. Although PS could flip to the outer leaflet with assistance from enzymes when the apoptotic pathway is activated [52,53], under most of the circumstances, the mammalian cell remains a neutrally charged surface, offering a recognizable marker for antibiotics to behave their selectivity.

1.2.2 Bacterial cell envelope

Bacteria have much more complexed envelopes than the mammalian cell (Fig. 1.2). Both Gram-positive and Gram-negative bacteria possess a cell wall layer covering the plasma membrane. The cell wall is mainly peptidoglycan made up of repeating units of the N-acetyl-glucosamine-N-acetyl-muramic acid, cross-linked by penta-peptide side chains [54]. This scaffold of peptidoglycan layers supplies enough physical strength for the bacteria to adapt to larger osmotic pressure variation [55]. Compared to Gram-negative bacteria, Gram-positive bacterial cell wall is much thicker and coupled with teichoic acids, one kind of anionic cell surface polymer. Since teichoic acids are anionic, they could attract metal cations and form networks, which could influence the rigidity and porosity of the cell wall [56]. Gram-negative bacteria have a thinner cell wall, but an extra outer membrane bilayer

Figure 1.2 Bacterial envelop of Gram-positive and Gram-negative bacteria. From cytosol to extracellular space, there are cytoplasmic membrane (mainly phospholipid) and cell wall (mainly peptidoglycan) in Gram-positive envelop, and one more outer membrane (mainly phospholipid and lipopolysaccharide) in Gram-negative envelop.

with mainly lipopolysaccharide (LPS) in the outer leaflet and phospholipid in the inner leaflet. Structurally, an LPS molecule could be divided into lipid A, the inner core, the outer core, and the O-polysaccharide chains [57]. Well known as the inducer of septicemia and endotoxic shock syndromes associated with Gram-negative infections [58,59], LPS molecules have some properties contributing to the permeability barrier function of outer membrane: they have highly saturated acyl chains facilitating the tight packing; they could bind to each other tightly to form non-fluid continuum which could effectively prevent hydrophobic molecules [56]. Teichoic acids and LPS both are negatively charged molecules and both locate at the extracellular face of the bacteria, leading to a negatively charged cell surface.

The plasma membrane in Gram-positive bacteria is different from that of Gram-negative

bacteria too. The principal components of both kinds of membrane models are PE, phosphatidylethanolamine (PG, negatively charged), and cardiolipin (CL, negatively charged) (Fig. 1.3), but their ratio varies greatly over different bacteria strains [60]. Gram-negative strains usually contain more PE than anionic lipids, while Gram-positive is the opposite, or in some cases, no PE at all [60], leading to the fact that Gram-positive membrane is more negatively charged than the Gram-negative membrane. And as a result, a mixture of PE and PG lipids is commonly used to mimic the bacterial membrane, more specifically, with a ratio of 3:1 for Gram-negative bacteria [61] and a ratio of 1:3 for Gram-positive bacteria [62].

Taken together, the charged surface and the unique components make the bacterial envelope differentiating from the mammalian membrane and have the potential to be the

Figure 1.3 Head groups of different phospholipids in the bacterial cell membrane. Rs represent for the hydrocarbon chains.

targets for envelope-targeting antibiotics design. Although some resistances are developed towards some of these antibiotics, such as the secreted protease against the AMPs, there are also studies showed that AMPs with specific structures or D-amino acids withstand the proteolytic degradation from protease [63,64], this direction of antibiotics design is still full of hopes and possibilities.

1.2.3 Bacterial envelope targeting antibiotics and their modes of action

a. Membrane-active AMPs

One kind of intensively studied antibiotics during the last decades is the membrane-active AMP. AMPs are produced by most of the living organisms, like animals, plants, and bacterial species, and play an important role in the innate immune system [65]. An AMP is usually a positive charged and amphipathic peptide composed of 5 to 50 amino acids [66]. The positive charges promote the binding with the negatively charged cell wall or outer membrane. And then the permeation process follows after the accumulation of AMPs to a threshold concentration. On one hand, AMPs across the cell wall easily due to the mesh-like and unattractive structure of peptidoglycan layers [36]; on the other, the outer membrane is also passable to AMPs by charge-competing with the divalent cations distributed in the LPS layers [36,67]. Once they access the cytoplasmic membrane, AMPs collapse the membranes through one of the several proposed mechanisms [25,31,60,67,68]:

(1) In the barrel-stave model, AMPs undergo a conformational phase transition to form an amphipathic secondary structure after binding to the membrane. The pose of AMPs at this point, in which the charged residues are attracted to the phospholipid head groups and the hydrophobic domains are inserted into the membrane, promotes the further

penetration process. Several AMPs then aggregate to minimize the unfavorable contacts between their hydrophilic domains and the hydrophobic area of the membrane. The barrel-stave model is finally established with all the AMPs aligning around an aqueous pore in a transmembrane configuration, and facing their hydrophobic domains to the acyl chains and their hydrophilic domains to the center of the pore (Fig. 1.4 barrel-stave model). This model requires sufficient length of AMPs to traverse the hydrophobic core of the membrane.

(2) In the toroidal-pore model, AMPs are usually not long enough to span the two leaflets of the membrane. Most of the process is similar with the barrel-stave model except that the penetration of AMPs induces a curvature strain to force the phospholipid head groups to bend to the opposite leaflet for the formation of a toroidal pore (Fig. 1.4 toroidal-pore model). In this model, the wall of the pore is composed by both AMPs and the head groups of the lipids, and this is the major difference to the barrel-stave model.

Figure 1.4 Schematic illustrations of bacterial killing mechanisms by AMPs. [67]

(3) In the carpet model, no formation of the regular transmembrane channel is involved. AMPs function in a detergent-like manner by forming micelles with lipids (Fig. 1.4 carpet model), and the specific secondary structures are unnecessary either. The initiation of the process is triggered by the changes in membrane fluidity and decreases in membrane barrier properties upon the accumulation of AMPs on the membrane surface.

The membrane disruption induced by the above three mechanisms first causes leakage of ions and metabolites, and ultimately membrane rupture and cell death. Interestingly, associated with the membrane disruption, lots of AMPs are able to be translocated into the cytoplasm to interact with their intracellular targets [69-71], suggesting a correlation between different antimicrobial mechanisms.

Another killing mechanism of membrane-active AMPs is through the clustering of anionic lipids in the membrane. Like those raft domains existed in the mammalian membrane [72], lipid components in the bacterial membrane are not uniformly distributed as well. The formation of bacterial lipid domain is promoted by membrane curvature which is related to the shape of the cell [73,74], or by interactions with membrane components, such as membrane proteins [75]. Lipid domains are also inducible by AMPs, however, unlike those forming spontaneously, which are believed to have specific functions [76], these induced ones could lead to the cell death due to the caused phase boundary defects between the domain interfaces. Compared to the defects between natural lipid domains, those between the induced domains arise quickly and leave no compensating time for the bacteria [77].

Apart from their broad-spectrum antimicrobial activity, AMPs have significant disadvantages for therapeutic use. Many AMPs might cause lysis of human red blood cells

(cytotoxicity) [47], and some of them active sensitively to the salt concentration or pH value, thus only a few out of thousands of AMPs proceeded into clinical trials. Improvements are required towards the sequence design of AMPs.

b. L_{II} targeting antibiotics

L_{II}, one of the most essential intermediate products for the biosynthesis of bacterial cell walls, has been recognized as an antibiotic target for decades [78]. Structurally, one L_{II} molecule is composed of one bactoprenol hydrocarbon chain (C55), which is embedded in the cell membrane, a disaccharide of N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc), a penta-peptide attached to the MurNAc (typically with the sequence of L-alanyl-γ-D-glutamyl-L-lysyl-D-alanyl-D-alanine), and a pyrophosphate group (PP) linking the bactoprenol anchor and MurNAc [79,80]. The disaccharide and penta-peptide form the peptidoglycan subunit is further cross-linked through the penta-peptide to provide mechanical strength to the cell wall. L_{II} is critical to the cell wall synthetic pathway [81]: the molecule is first assembled to the C55 anchor in the cytoplasmic side of the plasma membrane, and then translocated to the periplasmic side where the pentapeptide-linked disaccharide is released from the anchor to build the cell wall. The remaining C55-PP is translocated back to the cytoplasmic side of the membrane by a not well characterized recycling pathway [82]. Thus, preventing the translocation of L_{II} is a bona fide strategy to kill bacteria by interfering with a critical step in the cell wall synthetic pathway.

Some antibiotics, such as vancomycin, nisin, ramoplanin, and mannopeptimycin [79] exert their antimicrobial activity by interacting with L_{II} synthesis. Despite the identity of their targets, these antibiotics bind to L_{II} through the different patterns. For example, the interaction between vancomycin and L_{II} was well investigated experimentally and

computationally: vancomycin binds to L_{II} through five hydrogen bonds (H-bonds) with the D-Ala-D-Ala segment of the penta-peptide of L_{II} [83,84]. Several of these complexes can even form a super-complex through the D-Ala-D-Ala segment [85,86]. On the other hand, nisin, the best-known antimicrobial agent from the lantibiotic family, binds to the pyrophosphate moiety of L_{II} through the H-bonds with its lanthionine ring [87]. However, unlike its cell-wall-synthesis-blocking variant [88], nisin itself works as a membrane disruptor: the high binding affinity with L_{II} is utilized to dock onto the membrane, where nisin formed transmembrane pore together with L_{II}.

c. Phospholipid targeting antibiotics

Cyclic peptides antibiotics are widely applied in therapeutics for their good binding affinity, target selectivity, and low toxicity [89]. The above-mentioned vancomycin and nisin are both cyclic peptides. Some of the cyclic peptides bind to phospholipid components through specific binding sites. An NMR study showed that cinnamycin has a strong binding affinity to PE component [90]. In the complex, the hydrophobic domain of cinnamycin is clearly distinguished from the hydrophilic domain. Together with one hydrophilic residue in the boundary area, several hydrophobic residues form a binding pocket for the head group of PE. The selectivity of cinnamycin is assured by the size of the binding pocket. As another case, cyclic peptide daptomycin shows its selectivity to PG components [91,92]. This binding is mediated by calcium ions, whose association induces the conformational change of the daptomycin to a higher amphipathicity. Bactericidal mechanism of the daptomycin is shown in Fig. 1.5. Modes of action of the phospholipid targeting antibiotics principally involve the alteration of membrane properties, such as the fluidity or curvature of the membrane, as well as the lipid flip-flop [93].

Figure 1.5 Proposed bactericidal mechanism of daptomycin. (A) Daptomycin first undergoes a conformational transition after association with calcium ions. This transition increases the molecular amphipathicity, facilitating the interaction with the PG lipids and the insertion into the membrane. (B) Daptomycin oligomerizes in the membrane, disrupting the membrane integrity, and (C) induces leakage of the membrane and eventually the cell death [94]. Dap: daptomycin.

The negatively charged cell surface and the unique lipid components make bacterial envelop a versatile target for antibiotics. Both the anionic and zwitterionic components are recognizable for antibiotics, and the modes of action of antibiotics include but are not limited to the collapse of the membrane, rearrangement of the lipid components, variation of the bulk physical properties of the membrane, and the disruption of cell wall synthesis. These studies provide us with hints in the search for new therapeutic options.

1.3 From mechanism study to rational design

The studies of drug-target interactions aid drug design from two aspects: Firstly, it is possible to extract the characteristics of drugs that trigger the antimicrobial activities and graft them to a purely new synthesis. In the case of AMP, the positive charges provide the driven force in the initial binding with bacteria, and the amphiphilicity is necessary for the later secondary structure formation and the further permeation process. By combining

these features, many new chemical structures mimicking the AMPs were synthesized. For example, the synthetic cationic polymers such as polynorbornenes (Fig. 1.6a), reported by Lienkamp et al. [95], and polymethacrylates (Fig. 1.6b), reported by Sovadinova et al.[96], showed excellent broad-spectrum antimicrobial properties and selectivity. Same as the AMPs, both polymers are composed of hydrophilic units (positively charged) and hydrophobic units. A polymerizable interface lies in each repeating-unit pair for the control of molecular weight, which is another important parameter for the antibacterial activity [95].

Much more need to be considered for the optimal activity as well as avoiding the cytotoxicity [97] in the de novo drug design, such as the balance between the hydrophobicity and the hydrophilicity, the cationic group structure, the molecular size and so on [98]. By contrast, acquisition of new drugs through a minor modification of existing drugs is relatively easier. This is the second strategy that the mechanism studies are valuable in. Such an operation is usually with the intention of improving drug activity [99], but sometimes it brings back activities for the resisted drugs. As we mentioned earlier,

Figure 1.6 Chemical structure of polynorbornenes (a) and polymethacrylates (b). R is the alkyl group.

vancomycin interferes with cell wall synthesis through its binding with the D-Ala-D-Ala segments of L_{II}, however, resistance to this drug have been detected in bacterial strains with mutation of D-Ala-D-Ala into D-Ala-D-Lac (lactate). Several vancomycin analogues were proved to regain activity against vancomycin-resistant pathogen [100]. Xie et al. reported that the binding affinity of vancomycin to the mutated L_{II} is 1000-fold lower due to the introduction of lone-pair repulsion and the loss of central H-bonds [101]. By changing one critical group of vancomycin, they came up with a derivative having a similar binding affinity to both the normal and mutated L_{II}.

A deep understanding of the drug-target interaction offers opportunities for perfecting the existing drugs and developing the unrevealed ones. In this thesis, we focused on two new antibiotics, a synthesized cationic polymer CSM5-K5, and a recently discovered natural product TXB. For CSM5-K5 molecule, the mechanism of its self-assembly and interaction with membrane models were investigated in this thesis. TXB was reported in 2015 as a L_{II} recognizer, however, the conformation of the binding complex is not available. With the help of MD simulation, the binding modes of TXB and L_{II} were studied. The pharmacophore of TXB was confirmed to be the ring motif, and the importance of a D-amino acid on the ring motif was discussed. Hopefully, these findings of CSM5-K5 and TXB would be instructively meaningful for the future drug discovery.

2 Methodology

Experimental techniques, such as X-ray diffraction or NMR, play central roles in exploring structure and function of the large molecules of biological interest during the last century. Despite the increasingly sophisticated technologies nowadays, biological researches on an atomic-level, which greatly help to understand biological processes, is still experimentally challenging. MD simulations provide the ultimate detail concerning individual particle movement as a function of time, pointing out an effective approach to solve this problem [102].

2.1 Conventional MD simulation

Since its first application about macromolecule was published by McCammon et al. in 1977 [103], MD simulation has been developed to be a useful tool for investigating the biomolecule systems, including proteins, nucleic acids, lipids and so on. The essence of MD is to model a physical system by treating all its atoms as point charges associated with masses. The movements of the atoms obey the Newton equation of motion. For a system composed of n particles, potential U is a function of particle positions,

$$U = U(r_1, r_2, \dots r_n) \quad (2.1)$$

thus the force on particle i is:

$$F_i = -\frac{\partial}{\partial r_i} U(r_1, r_2, \dots r_n) \quad (2.2)$$

By integrating equation 2.2, the position and velocity changes of a particle in a time step are accessible with the awareness of its initial position and velocity. The trajectory of movement for the system is then predictable with the iterative solution of the equation.

2.1.1 Force field

To get an accurate result, the most important thing to be addressed is to describe the interactions between atoms correctly. This is usually handled by force field, which is essentially a combination of potential formulas like the following:

$$\begin{aligned} U_{total} &= U_{bonds} + U_{angles} + U_{torsions} + U_{vdw} + U_{columb} \\ &= \sum_{bonds} k_b(l - l_0)^2 + \sum_{angles} k_a(\theta - \theta_0)^2 + \sum_{torsions} \sum_n \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \\ &\quad + \sum_{j=1}^{N-1} \sum_{i=j+1}^N \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^6 \right] + \sum_{j=1}^{N-1} \sum_{i=j+1}^N k \frac{q_i q_j}{r_{ij}} \end{aligned} \quad (2.3)$$

This is a typical potential formula used in AMBER force field [104]. The bonded interactions include the first three terms: The first two terms represent the deformational energies of bond lengths and bond angles from their equilibrium values, and a harmonic force is used in these two terms to keep the correct chemical structure from bond breaking. l_0 , θ_0 in equation 2.3 are the equilibrium values for bond length and angle respectively, and k_b , k_a are their force constants. The third term describes rotations around the chemical bond with a Fourier series, in which periodicity is determined by n and heights of rotational barriers are defined by V_n . The non-bonded interactions, i.e. the van der Waals and electrostatic interactions between all atom pairs are described by a Lennard-Jones 12-6 potential and a Coulomb electrostatic potential respectively. In equation 2.3, r_{0ij} is the equilibrium distance, ϵ_{ij} is the potential well depth, and k is the electrostatic constant.

Different force fields, such as CHARMM force field [105] or GROMOS force field [106], usually have different functional forms, thus, their performances on the same system

could be variable [107]. In a comparison study of AMBER ff99SB and CHARMM22/CMAP, the former generated experimentally matchable results by predicting the unfolding of cecropin P1 in bulk water and folding on membrane [108]. But another study indicated comparable behaviors of protein under the description of different force fields [109]. The variation of performances depends sometimes on the studying systems, thus it is always recommended to select a force field based on the literature search of a similar but studied system.

2.1.2 Simulation ensembles

The Hamiltonian of the simulation system is conserved by the equations of motion [110], thus providing an important link between MD and statistical mechanics, which connects the microscopic details of the system. According to the Gibbs' ensemble concept, a large system exhibits the same macroscopic properties as a set of its inclusive individual microscopic configurations, it is possible to calculate the macroscopic observables of a system by simply averaging over a large number of different microscopic configurations. Statistical ensemble are usually characterized by fixed values of thermodynamics variables. The generally used ensembles in MD simulations are the Microcanonical Ensemble (NVE) with constant volume (V), number of particles (N) and energy (E); the Canonical Ensemble (NVT) with constant volume (V), number of particles (N), and temperature (T); the Isothermal-isobaric Ensemble (NPT) with constant pressure (P), number of particles (N) and temperature (T). The NVE ensemble is isolated, thus no energy exchange with outside, however there are exothermic and endothermic processes in the NVT and NPT ensemble, and their energy can be exchanged with an outside thermostat. Besides, the NPT ensemble

is also coupled with an outside barostat to maintain the desired pressure. In this thesis, we performed our simulations mainly using the NVT ensemble and NPT ensemble.

2.2 Enhanced sampling methods

Conventional MD simulation has been regarded as a general sampling method for the conformational study of biomolecules, however, it is useful in exploring interesting events only when the system is ergodic in the simulated timescale. Most of the simulated systems exhibit metastability, i.e., having a rugged FES. The sampling could be trapped by the separated local minima on the FES for a very long time, making sufficient sampling unaffordable. Therefore, it is always a challenge for MD simulation to efficiently explore the large regions of phase space due to the limited time scale that the computational resources could reach. To handle this issue, several enhanced sampling methods, such as MetaD and PT, have been developed in the past few decades. In MetaD, the selected CVs evolve in response to a biased potential composed by the real potential and a history-dependent bias. The history-dependent bias is a sum of Gaussians which center on the visited CV values, thus pushing the system to visit the less favored region. In PT, several replicas at different temperatures are simulated simultaneously. The neighboring replicas can swap their configurations with an acceptance probability determined by their temperature and instantaneous potential. The root of PT relies on an annealing procedure, where the system is first heated to obtain enough energy to escape the local minima, and then cooled to produce uncorrelated configurations quickly [111]. By annealing the system and manipulating the Hamiltonian of the system with extra biases respectively, configurations in the MetaD and PT simulations could escape the local minima more easily when compared with the conventional MD. These enhanced sampling methods have been

widely applied in studying the rare events including protein folding [112,113], peptide adsorption onto surfaces [114], protein oligomerization [115,116], and so on.

2.2.1 MetaD and well-tempered MetaD

MetaD [117,118] is one of the most popular sampling-enhancing methods, especially to validate and efficiently probe rare events, and to reconstruct the FES [119-122]. In MetaD, an external history-dependent bias potential, built as a sum of Gaussians deposited along the selected CVs, is used to stimulate the system to visit its unfavorable region of the FES to enhance the dynamics. When the system becomes diffusive in the CV space, a negative image of the FES is also reconstructed by the sum of the Gaussian potential [117].

For a system with a set of coordinates R and a potential $U(R)$ at temperature T , we define a finite number of CVs $s(R) = (s_1(R), s_2(R), \dots, s_d(R))$ as any explicit function of R , then the equilibrium behavior of these CVs is completely defined by the probability distribution:

$$P(s) = \frac{\exp(-\beta F(s))}{\int ds \exp(-\beta F(s))} \quad (2.4)$$

and free energy $F(s)$ is given by

$$F(s) = -\frac{1}{\beta} \ln \left(\int dR \exp(-\beta U(R)) \delta(s - s(R)) \right) \quad (2.5)$$

where s is the value of CVs, and $\beta = 1/(k_B T)$, with k_B the Boltzmann constant. Apparently, from equation 2.4, the system is fond of visiting regions with lower free energy. In MetaD, a small repulsive Gaussian potential is added to the system every τ_G MD time steps, at time t , the accumulated Gaussian potential $V_G(s, t)$ is given by:

$$V_G(s, t) = \int_0^t dt' \omega \exp \left(- \sum_{i=1}^d \frac{(s_i(R) - s_i(R(t')))^2}{2\sigma_i^2} \right) \quad (2.6)$$

where σ_i is the Gaussian width and ω is an energy rate, which could be expressed as the ratio of Gaussian height W and deposition stride τ_G : $\omega = W/\tau_G$. Impact of this external bias potential is to disfavor the more frequently visited configurations. In this way, sufficient statistics for every CV value could be collected and the FES could be reconstructed along the CV space. After a sufficiently long time deposition, the Gaussian potential is related to the underlying free energy as following:

$$V_G(s, t \rightarrow \infty) = -F(s) + C \quad (2.7)$$

where C is an irrelevant constant. However, MetaD has a convergence issue that V_G would oscillate around the underlying free energy rather than converge to it [118], remaining to be the major drawback of this method.

Barducci et al. published a well-tempered MetaD in 2008 [123] to improve this convergence problem by utilizing a modified external bias:

$$V_G(s, t) = k_B \Delta T \ln \left(1 + \frac{\omega N(s, t)}{k_B \Delta T} \right) \quad (2.8)$$

where $N(s, t) = \int_0^t \delta(s - s(t')) dt'$, is the histogram of CV and its time derivative $\dot{N}(s, t) = \delta(s - s(t))$. ΔT has a dimension of temperature. We can get the variation rate of this bias potential from the time derivative of equation 2.8:

$$\dot{V}_G(s, t) = \frac{\omega \delta_{s,s(t)}}{1 + \frac{\omega N(s, t)}{k_B \Delta T}} = \omega \exp \left(- \frac{V_G(s, t)}{k_B \Delta T} \right) \delta(s - s(t)) \quad (2.9)$$

If the $\delta(s - s(t))$ term is replaced by Gaussian, the difference between equation 2.9 and time derivative of equation 2.6 would only lie in the factor $e^{-\frac{V_G(s,t)}{k_B\Delta T}}$, which is usually integrated into Gaussian height in practice: $\dot{W} = \omega e^{-\frac{V_G(s,t)}{k_B\Delta T}} \tau_G$. By using the new bias, Gaussian deposition rate decreases over simulation time, leading to the convergence of $V_G(s, t)$ to the underlying free energy and greatly increasing the availability of MetaD.

According to equation 2.4, the probability distribution in well-tempered MetaD after long enough Gaussian deposition time becomes $P(s, t)ds \propto \exp(-\beta(F(s) + V_G(s, t))) ds$, and we will get:

$$\begin{aligned} \dot{V}_G(s, t) &= \omega \exp\left(-\frac{V_G(s, t)}{k_B\Delta T}\right) P(s, t) \\ &= \omega \exp\left(-\frac{V_G(s, t)}{k_B\Delta T}\right) \frac{\exp(-\beta(F(s) + V_G(s, t)))}{\int ds \exp(-\beta(F(s) + V_G(s, t)))} \quad (2.10) \end{aligned}$$

When $t \rightarrow \infty$, we have

$$V_G(s, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(s) + C \quad (2.11)$$

where C is an immaterial constant. Thus the bias potential is not fully compensated the free energy. And combined with equation 2.8, the free energy could be estimated as:

$$\tilde{F}(s, t) = -\frac{T + \Delta T}{\Delta T} V_G(s, t) = -k_B(T + \Delta T) \ln\left(1 + \frac{\omega N(s, t)}{k_B\Delta T}\right) \quad (2.12)$$

The parameter ΔT regulates the extent of free-energy exploration: when $\Delta T = 0$, we obtain a conventional MD simulation with a zero bias potential; and if $\Delta T \rightarrow \infty$, we have a standard MetaD simulation with a full compensation of the underlying free energy. With

a proper ΔT the relevant free energy barrier could be crossed efficiently, and practically, a bias factor $\gamma = \frac{T+\Delta T}{T}$, is usually employed to set the value of ΔT .

Under convergence, the partition function of the selected CV is deducible. Transformed from equation 2.11, the bias potential is expressed as:

$$V_G(s, t \rightarrow \infty) = -(1 - \gamma^{-1})F(s) \quad (2.13)$$

And within an irrelevant constant,

$$\begin{aligned} F(s) &= -\beta^{-1} \ln P(s) = -\beta^{-1} \ln \int dR \delta(s - s(R)) e^{-\beta U(s(R))} \\ &= U(s) - \beta^{-1} \ln N(s) \end{aligned} \quad (2.14)$$

where $P(s) = \int dR \delta(s - s(R)) e^{-\beta U(s(R))} = \ln N(s) e^{-\beta U(s(R))}$ is the distributional probability in canonical ensemble, and $N(s) = \int dR \delta(s - s(R))$ is the number of states of CV with a value s . By combining equation 2.13 and 2.14, the biased potential of the system $U_\gamma(s(R))$ could be written as:

$$U_\gamma(s(R)) = U(s(R)) + V_G(s, t \rightarrow \infty) = U(s(R)) - (1 - \gamma^{-1})F(s(R)) \quad (2.15)$$

Thus the partition function of the biased system is:

$$\begin{aligned} Z_\gamma &= \int dR e^{-\beta[U(s(R)) - (1 - \gamma^{-1})F(s(R))]} \\ &= \iint dR ds \delta(s - s(R)) e^{-\beta[U(s(R)) - (1 - \gamma^{-1})F(s(R))]} \end{aligned} \quad (2.16)$$

to have a further simplification

$$\begin{aligned}
Z_\gamma &= \int ds e^{-\beta F(s)} * e^{\beta(1-\gamma^{-1})F(s)} = \int ds e^{-\beta \gamma^{-1} F(s)} = \int ds e^{-\beta \gamma^{-1} (U(s) - \beta^{-1} \ln N(s))} \\
&= \int ds e^{-\beta U(s) \gamma^{-1}} * N(s)^{\gamma^{-1}} = \int ds P(s)^{1/\gamma}
\end{aligned} \tag{2.17}$$

Based on equation 2.17, the distributional probability of the selected CV is tuned by bias factor in well-tempered MetaD. The bias potential makes the system more dispersive in the CV space by smoothing the distributional probability.

Another thing needs to be considered in dealing with the frames generated in the biased simulation is the reweighting. Different from the conventional MD simulation, where the frames at different time steps have the same weight value, the biased MD simulation has uneven weight value for its frames because of the alteration of the instantaneous probability induced by the bias potential. As discussed above, the instantaneous probability at time t $P(R, t)$ in well-tempered MetaD is under the action of both the internal potential $U(s(R))$ and the bias potential $V_G(s(R), t)$.

$$P(R, t) = \frac{\exp\left(-\beta \left(U(s(R)) + V(s(R), t)\right)\right)}{\int dR' \exp\left(-\beta \left(U(s(R')) + V(s(R'), t)\right)\right)} \tag{2.18}$$

based on equation 2.14, the above equation is rewritten as [124,125]:

$$P(R, t) = \frac{\exp\left(-\beta \left(U(s(R)) + V(s(R), t)\right)\right) \int ds \exp(-\beta F(s))}{\int ds \exp\left(-\beta \left(F(s) + V(s, t)\right)\right)} \tag{2.19}$$

and further recast as:

$$P(R, t) = e^{-\beta(V(s(R), t) - c(t))} \cdot P_0(R) \tag{2.20}$$

where $P_0(R)$ is the unbiased Boltzmann probability density and the function $c(t)$ is defined as

$$c(t) = \frac{1}{\beta} \ln \frac{\int ds \exp(-\beta F(s))}{\int ds \exp(-\beta (F(s) + V(s, t)))} \quad (2.21)$$

So far, the biased instantaneous probability is connected to the unbiased probability density. All the data analysis of the biased MD simulation in this thesis is based on this reweighting method.

2.2.2 PT and PT-WTE

Another commonly used method to enhance sampling is Replica Exchange MD (REMD), such as PT. MD simulations at room-temperature are usually of much interest but easily trapped in the metastable states. Simulations at higher-temperatures with correspondingly higher kinetic energies are able to overcome the energy barrier between different metastable states, however, the FESs obtained are different from the one at room temperature. In PT, a series of replicas, whose temperatures are arranged in an incremental ladder, are simulated simultaneously. The enhanced sampling is realized by allowing the exchange of configurations of lower- and higher-temperature replicas. In practice, the exchange is attempted every specified number of simulation time-steps, with an acceptance probability described by [126]:

$$P(T_1 \leftrightarrow T_2) = \min(1, e^{(\beta_1 - \beta_2) \cdot (U_1 - U_2)}) \quad (2.22)$$

where $\beta_1 = 1/(k_B T_1)$, $\beta_2 = 1/(k_B T_2)$, and U_1, U_2 are the instantaneous potential energies of the two exchanging configurations. The efficiency of PT simulation is largely

impacted by the AAP, which is related to the potential distribution of the two neighboring replicas. As shown in Fig. 2.1a, instantaneous potential energies of two replicas fluctuate around their individual average values during the simulation. To have a higher AAP for these two replicas, their fluctuations must have enough intersection, which is displayed in their distribution as sufficient overlap. The position of the potential distribution is decided by the ensemble temperature, thus, a small temperature gap between neighboring replicas is necessary for an efficient PT simulation.

However, there is still a big obstacle that limits the wider applications of PT. The requirements that the highest temperature must be high enough to keep all the replicas diffusive in phase space and the temperature gap between neighboring replicas must be

Figure 2.1 Potential distributions of canonical ensembles (a) and the effects from WTE (b). In WTE, the position of the potential distribution is not affected, while the width is enlarged. These properties of WTE guarantee the coverage of same temperature range with fewer replicas in WTE than in canonical ensemble.

small enough to keep sufficient exchange, determine the lower bound of the number of replicas, which is usually from dozens to more than one hundred. That is a disaster for researchers with limited computational resources, especially when simulating systems being large. The WTE framework is able to alleviate this issue [127]. Intrinsically, WTE is well-tempered MetaD, but using the system potential as a CV, therefore, distribution of the system potential in WTE can also be described by equation 2.17. Its partition function is as following:

$$Z_\gamma = \int dU P(U) \gamma^{-1} \quad (2.23)$$

Unlike other CVs, the system potential in a canonical ensemble follows a Gaussian distribution [128], which means $P(U) \propto e^{-\frac{(U-\mu)^2}{2\sigma^2}}$, where μ is the average potential and σ is the standard deviation of the distribution. In WTE, potential distribution function $P_\gamma(U)$ is tuned by the bias factor but still has a Gaussian shape:

$$P_\gamma(U) = P(U) \gamma^{-1} \propto e^{-\frac{(U-\mu)^2}{2\gamma\sigma^2}} \quad (2.24)$$

As shown in Fig. 2.1b, this change leads to a same average potential, μ , but $\sqrt{\gamma}$ times larger distributional width in the WTE when compared with the canonical ensemble, which means the PT simulation could still have intact acceptance probability of configurational exchange between replicas with larger temperature gaps when it is performed in WTE (PT-WTE). PT-WTE simulation offers a huge economization on computational resources by decreasing the number of replicas that are required to cover a certain temperature range. Several systems in this thesis were simulated with PT-WTE, and we also developed a strategy to generate an optimal temperature sequence for PT-WTE.

2.3 MM-PBSA calculations

The MM-PBSA approach is widely used to calculate the binding free energy of a biomolecular complex for an estimation of their binding strength [129]. By definition, the binding free energy of a receptor-ligand complex $\Delta G_{binding}$ is defined as:

$$\Delta G_{binding} = G_{complex} - (G_{receptor} + G_{ligand}) \quad (2.25)$$

where $G_{complex}$, $G_{receptor}$, and G_{ligand} are the free energy of the complex, isolated receptor and ligand in solvent, respectively, and can be calculated by

$$G_x = \langle E_{MM} \rangle - TS + \langle G_{solvation} \rangle \quad (2.26)$$

where $\langle E_{MM} \rangle$ is the average molecular mechanics potential energy in a vacuum, T and S are the temperature and entropy, respectively, and $\langle G_{solvation} \rangle$ is the free energy of solvation.

The potential energy E_{MM} is the sum of both bonded (E_{bonded}) and non-bonded interactions ($E_{non-bonded}$), within which the former is composed of bond, angle, dihedral and improper interactions, while the latter consists of electrostatic (E_{el}) and van der Waals (E_{vdW}) interactions. When calculating the free energy of the bound and isolated state of the receptor and ligand, their conformations are considered as identical, thus the variation of the E_{bonded} term after binding is zero [130]. The E_{el} and E_{vdW} terms are calculated by Coulomb's law and Lennard-Jones function, respectively. However, choice of the dielectric constant in the E_{el} calculation is usually tricky. In original applications of MM-PBSA, the dielectric constant was set as 1, nonetheless, people found that the results might be improved with a larger dielectric constant [131]. It is suggested that the selection of a

proper dielectric constant depends on the characteristics of the binding site, but often the results are best for the value of 2-4 [131].

The free energy of solvation, $G_{solvation}$, denotes the energy cost of transferring a solute from vacuum into the solvent. It is the sum of the polar solvation term (G_{polar}) and the non-polar solvation term ($G_{non-polar}$), which represent the electrostatic and non-electrostatic contributions to the solvation free energy, respectively. In calculation of the $G_{solvation}$, the G_{polar} term is estimated by solving Poisson–Boltzmann (PB) equation, and the $G_{non-polar}$ term, which includes the energy cost of making a cavity in the solvent, as well as the van der Waals interaction between the solvent and solute, is usually calculated using the solvent accessible surface area (SASA) model, based on the assumption that the SASA is linearly related to the $G_{non-polar}$ term [132].

The entropic term is calculated by quasi harmonic analysis or normal-mode analysis of the trajectory [133]. However, the entropy calculations is expensive, and there are studies reporting that the entropic contribution has only small improvements to the result [134,135]. Therefore, the entropy term is usually ignored in MM-PBSA studies [131]. In the MM-PBSA calculation of this thesis, we also ignored the entropy contribution.

3 Simulation study of a membrane targeting antibiotic - CSM5-K5

3.1 Background

Compared to the synthetic polymers, natural ones are usually non-toxic, derived from renewable resources, biocompatible, biodegradable and possess low immunogenicity [136]. Chitosan is one type of widely studied cationic polymer obtained by treating chitin with alkali. Its long-chain molecule is composed of randomly distributed β -(1-4)-linked D-glucosamine, the deacetylated units, and N-acetyl-D-glucosamine, the acetylated units (Figure 3.1). Amino group ($-NH_2$) on D-glucosamine units could be protonated to form $-NH_3^+$ when the pH value of the environment is lower than 6.5, making chitosan water soluble in acidic solution [137].

Chitosan shows antimicrobial activity only in the acidic medium due to the poor solubility at high pH [138,139]. The molecular weight and degree of deacetylation are another two important factors of its antimicrobial activity [140,141]. A generally accepted killing mechanism is that chitosan molecules precipitate on the microbial cell surface based

Figure 3.1 Two composition units of chitosan: N-acetyl-D-glucosamine (left) and D-glucosamine (right).

on the electrostatic interaction, and then this long-chain molecule could form an impervious layer around the cell surface, preventing the exchange of crucial material by blocking the channels [142]. During this process, chitosan with too large molecular weight would lead to an incompact layer, whereas a too small molecule would be difficult to form the layer [143]. The blocking layer could finally destabilize the cell envelop and lead to the death of the cell [138].

The limitation of chitosan for therapeutic application caused by its low solubility at physiological pH is conquerable by modifying the reactive sites on the glucosidic units, including one primary amine and two primary or secondary hydroxyl groups. Mary's group synthesized one kind of cationic peptido-polysaccharide named chitosan-graft-polylysine (CS-g-Kn) by grafting polylysine chain onto the amide group of chitosan [144]. Later in 2017, their group reported another polymer of the same category but possessing a smaller molecular weight, CSM5-K5 [145]. This kind of chitosan derivative shows broad-spectrum antimicrobial activities with high solubility and a large number of positive charges at neutral pH due to the integration with polylysine.

Like most of the antimicrobial cationic polymers in the aqueous medium, which self-assemble easily thus improving their antibacterial properties [146], aggregated nanoparticles of CSM5-K5 were also detected by a dynamic light-scattering setup, and their membrane activity was observed by confocal microscopy and cryo-TEM [145]. However, the mechanism of this aggregation and its interaction with the membrane is hard to reveal by experimental techniques. In this chapter, two CSM5-K5 monomers are simulated with the help of MetaD for their FES of dimerization, and their interactions as

monomers or dimer with the inner membrane of Gram-negative bacteria are studied by conventional MD simulations.

3.2 Parameters

3.2.1 General parameters

Experimentally, the molecular weight of CSM5-K5 is controlled to be around 1450 Da [145]. The backbone of CSM5-K5 is a chitosan chain composed of 5 sugar units, four of which is covalently bonded to four polylysine chains since the deacetylation degree of chitosan is 80%. Based on the molecular weight, the total number of lysine residues in the polylysine chains is approximately equal to 5, thus the molecular model for CSM5-K5 is built as shown in Fig. 3.2. We assume all the lysine residues are protonated in neutral pH due to their PKa value (around 10.5), therefore, each CSM5-K5 molecule contains 9 positive charges.

Parameters in this section are applied for both enhanced sampling and conventional MD simulations of the membrane system. All MD simulations were performed using the GROMACS package, version 4.6.3 [147]. The CSM5-K5 molecule is described by a hybrid force field: Glycam_06j force field [148] for the chitosan part and AMBER99SB force field

Figure 3.2 Molecular model used in this study. Four polylysine chains are grafted to the 80% deacetylated chitosan backbone. The polylysine chains are 100% protonated.

[149] for the polylysine part. A leap-frog algorithm was used to integrate Newton's equation of motion, and the time step was set to 2 fs along with the application of LINCS algorithm [150], by which all the covalent bonds between hydrogen atoms and heavy atoms were constrained. Simulation systems were solvated in TIP3P water model [151]. Sodium and chlorine ions, modified by Joung et al.[152], were used as the counter ions and added to reach a concentration of 0.17 M physiological salt solution. A cutoff of 1.2 nm was used for both van der Waals interaction calculation and short-range electrostatic interaction calculation, and correspondingly PME [153] method was employed to deal with the long-range electrostatic interaction. All simulations were carried out at 300 K temperature and 1 bar pressure and the V-rescale method [154] and Parrinello-Rahman method [155] (semi-isotropic Parrinello-Rahman for the membrane systems) were applied for temperature coupling and pressure coupling. All the visualization work was done by PyMol, version 1.5.0.3.

3.2.2 Well-tempered MetaD simulation of CSM5-K5's dimerization

The well-tempered MetaD simulation was carried out by the plumed plug-in [156]. Two CSM5-K5 molecules were solved in a 6*6*6 nm³ water box to study their dimerization behavior. The system was equilibrated with position restraint to CSM5-K5 for 100 ps in NVT ensemble, and another 100 ps in NPT ensemble, and then followed by 400 ns production run in NPT ensemble. We defined two CVs as the biased CVs: Distance refers to the center of the mass distance between the two CSM5-K5 molecules and angle is the included angle between two chitosan chains. The bias potential was deposited every 500 calculation steps (1 ps) with the initial Gaussian height equal to 2 kJ/mol and the Gaussian widths for distance and angle as 0.1 nm and 0.05 rad respectively. A constrained potential

was added on distance CV to limit the size of CV space and enhance the simulation efficiency. Bias factor used in the simulation was 8.

3.2.3 Conventional MD simulation of the CSM5-K5 and membrane system

Two kinds of membranes were involved in this study. The mammalian cell membrane (MM) was mimicked by a zwitterionic POPC bilayer with 81 POPC molecules in each leaflet. The bacterial inner membrane (IM) was modeled by a mixture of zwitterionic POPC and negatively charged POPG, and each leaflet contains 65 POPC and 16 POPG. The membrane models were built by a web server MemBuilder II and parameterized with a Slipids force field [157-159], which is compatible with the AMBER force field. To study the selectivity of CSM5-K5 to different membrane models, a CSM5-K5 dimer was simulated together with the mammalian cell membrane (MM-dimer) or bacterial inner membrane (IM-dimer) for 200 ns, respectively. To compare the influence of the dimerization to the morphology of the IM, we also performed simulations of the IM with two CSM5-K5 monomers (IM-2mon) or two 9-residue polylysine chains (IM-2PK) for 200 ns. The MM or IM models were equilibrated for 200 ns before its interaction with the cationic polymers. After equilibration, the area per lipid (APL) in MM and IM reached values of 0.64 nm² and 0.63 nm², respectively, which were comparable with the experimental values of pure POPC (0.64 nm² at 303 K [160]) and pure POPG (0.64 nm² at 293 K [161]). All the membrane systems were simulated for 3 repeats.

3.3 Results and discussion

3.3.1 Dimerization of the CSM5-K5 monomers

FES of the dimerization was obtained by plumed sum_hills. Our simulation was converged in 400 ns, since the FES was almost unchanged in the last 40 ns sampling,

especially in the region with small distance value (Fig. 3.3). Two local minima (labelled as A and B) are discovered in the region with a distance smaller than 0.7 nm and angle close to 0 or π , indicating that CSM5-K5 dimerize with the posture in which the chitosan chains are parallel or antiparallel to each other.

We clustered the structures in each minimum using gromos method [162] based on the RMSD with respect to the two chitosan chains with a cutoff of 0.1 nm. The representative structures of minimum A and B are shown in Fig. 3.4 A1 and B1, respectively. Along with the grafting of polylysine chains, the electrostatic repulsion between CSM5-K5 is highly enhanced as well. CSM5-K5 molecules self-assemble only when the attractive interactions, which are dominated by the inter-molecular H-bonds, overcome the repulsive interactions. We roughly divided the H-bonds into three categories: the chitosan-chitosan H-bonds (CC), the chitosan-lysine H-bonds (CK), and the lysine-lysine H-bonds (KK). All structures in the two minimum were picked out for an H-bonding statistics. As shown in Fig. 3.4 A2 and B2,

Figure 3.3 FESs of CSM5-K5's dimerization at different simulation time. Distance and Angle are used as biased CVs: the distance refers to the center of the mass distance between the two CSM5-K5 molecules and the angle is the included angle between two chitosan chains. Two local minima are discovered (labelled as A and B) in the dimerization region (distance < 1 nm), and they have almost no variation during the last 40 ns sampling, indicating the convergence of our simulation.

Figure 3.4 The representative structures and H-bonding architectures of minimum A and B. The representative structures in figure A1 and B1 are shown as stick model with carbon atoms colored as green and cyan to differentiate different CSM5-K5 monomers. Oxygen and nitrogen atoms are colored as red and blue respectively. All hydrogen atoms are hidden. The H-bonding architectures were generated based on all structures in the minima (figure A2 for minimum A and B2 for minimum B). From the average number of H-bonds (shown in the parenthesis), CC H-bonds are more important than KK or CK H-bonds in both cases.

CC H-bonds (average number of 5.0 and 3.1 for minimum A and B, respectively) occupy 84.7% and 79.5% of all the H-bonds in minimum A and B respectively, indicating the more important role they play in the aggregation as compared with the KK and CK H-bonds. On the other hand, lysine residues can hardly form H-bonds with each other because of the strong repulsion, in fact, polylysine exists as a random coil rather than α -helix or β -sheet in neutral pH solution due to the same reason [163]. These results also explain the posture of the parallel chitosan chains in the aggregation, since such a posture is favorable for the CC H-bonds formation.

Two H-bonding patterns are observed from the CC H-bonds: the one formed between two amide bonds, labelled as $\text{--NHC(=O)}\cdots\text{--NHC(=O)}$ (Fig. 3.5 a1 for grafted units and a2 for free units), and the one formed between two hydroxyl groups, labelled as $\text{--OH}\cdots\text{--OH}$ (Fig. 3.5 a3). Minimum A has a much higher occupancy of $\text{--NHC(=O)}\cdots\text{--NHC(=O)}$ H-bonds than minimum B (Fig 3.5 b), while their occupancies of $\text{--OH}\cdots\text{--OH}$ H-bonds are

Figure 3.5 Components of the CC H-bonds and stability of two minima. Two H-bonding patterns are observable from the CC H-bonds of the minima: the H-bond between two amide bonds (a1 and a2, labelled as $\text{--NHC(=O)}\cdots\text{--NHC(=O)}$), and the H-bond between two hydroxyl groups (a3, labelled as $\text{--OH}\cdots\text{--OH}$). Minimum A has a much higher occupancy of $\text{--NHC(=O)}\cdots\text{--NHC(=O)}$ H-bonds (b), and a similar occupancy of $\text{--OH}\cdots\text{--OH}$ H-bonds (c) as compared to minimum B. Starting from the representative structures presented in Fig. 3.4, 200 ns conventional MD simulations were performed. The distance between two CSM5-K5 monomers (d) suggests that only minimum A is stable.

comparable with each other (Fig. 3.5 c). Using the representative structures presented in Fig. 3.4 as the initial structures, we performed 200 ns conventional MD simulations to check the stability of each minimum. Distances between the two CSM5-K5 monomers (Fig. 3.5 d) suggest that minimum B, *i.e.*, aggregation with antiparallel chitosan chains, is not stable, which means the $\text{--NHC(=O)}\cdots\text{--NHC(=O)}$ H-bonds is more critical in self-assembly of CSM5-K5 than the $\text{--OH}\cdots\text{--OH}$ H-bonds. The strength of the $\text{--NHC(=O)}\cdots\text{--NHC(=O)}$ H-bonds is well known in chitin (chitosan with less deacetylation) [164,165], which stays as fibrils due to the great insolubility caused by such H-bonds between chains. Therefore, we speculate that the chitosan chains afford the aggregation mechanism in CSM5-K5 molecule by providing the $\text{--NHC(=O)}\cdots\text{--NHC(=O)}$ H-bonds, improving the antimicrobial activity.

3.3.2 Simulation of the CSM5-K5 with membrane systems

We first studied the selectivity of CSM5-K5. A CSM5-K5 dimer was placed 5 nm above the center of the mammalian cell membrane (MM-dimer) or bacterial inner membrane (IM-dimer) as initiation. The stability of the dimer is not affected by both kinds of membrane environments in 200 ns (Fig. 3.6 a). The dimer exhibits its selectivity by attaching to the IM model very quickly and strongly (Fig. 3.6 b), and this selectivity is also proved by isothermal titration calorimetry (ITC) measurement [145]. The binding is firstly driven by the long-range electrostatic attraction between the positively charged lysine residues and negatively charged IM model, which includes the negatively charged POPG lipids beyond the neutrally charged POPC lipids (Fig. 3.6 b). Extra H-bonds could be established between the lysine residues and the phosphate groups or the ester groups of the IM model (Fig. 3.6 c and d), to further stabilize the binding, when the dimer and IM model is close enough. The polylysine chains here function as the recognizer of the

Figure 3.6 Simulation of dimerized CSM5-K5 in the MM and IM membrane environments. Distances between the two CSM5-K5 monomers (a) indicate the stability of the dimer in both environments. The z-distances between CSM5-K5 and the center of the membrane suggest that CSM5-K5 binds to IM quickly and strongly (b). The z-distances between lysine residues and the center of the membrane is also shown in Figure b. The binding is driven by the long-range electrostatic attraction between lysine residues and the POPG lipids in the beginning. H-bonds between lysine residues and IM membrane are established to further stabilize the binding when the CSM5-K5 molecule attaches to the membrane. Figure c shows that the lysine residues form H-bonds with the phosphate groups or the glycerol moieties of the IM model. Figure d shows the time evolution of the number of such H-bonds.

bacterial membrane through its charge interaction, which is also proved responsible for membrane permeability by previous studies [166,167]. Taken together, our simulation results suggest that a combination of the specialties of chitosan and polylysine makes CSM5-K5 an effective antimicrobial polymer.

To further address the influences brought by aggregation, we simulated the IM model with two dissociated CSM5-K5 monomers (IM-2mon) as well as two 9-residue polylysine chains (IM-2PK). As expected, they both can bind to the IM quickly like the CSM5-K5 dimer because of the electrostatic attraction. However, only tiny modifications are inflicted to the IM in the aspects of area per lipid (APL), membrane thickness (Fig. 3.7) and deuterium order parameter (Fig. 3.8). Despite all these subtle influences to the membrane morphology, the lateral radial distribution function (RDF) curves of POPG lipids with respect to themselves (Fig. 3.9) indicate that the dispersion of the POPG is greatly changed in the IM-dimer system due to the accumulation of POPG around the CSM5-K5 dimer (Fig. 3.10). From Fig. 3.10, we also see slightly accumulated POPG in the IM-2mon system and a hardly influenced POPG enrichment in the IM-2PK system versus the control. We believe these results are related to the distribution of the positive charges in the polycations. The

Figure 3.7 Area per lipid (a) and membrane thickness (b) distribution in different systems. The pure membrane system (IM) is used as a control. The APL is slightly decreased (0~0.8%) with the binding of polycations, and the membrane thickness has a slight increase (0~1%) consequently. Figures are generated based on the last 50 ns simulation. The distance between phosphorus atoms in different leaflets is used to calculate the membrane thickness.

Figure 3.8 Deuterium order parameters (S_{cd}) for two acyl chains of POPC and POPG in different systems.

S_{cd} are calculated based on $S_{cd}(z) = \frac{3}{2} \langle \cos^2 \theta(z) \rangle - \frac{1}{2}$, where $\theta(z)$ was the angle between the normal to the bilayer and the C-C vector located at carbon number z . The polycations influence POPG more than POPC.

IM-2mon system contains two positive charge centers, each belonging to one monomer, in contrast, the stable dimerization provides the IM-dimer system with only one charge center. Such configuration in the IM-dimer system enhances the electrostatic attraction to POPG while avoiding the competition between different charge centers in the IM-2mon system, resulting in the most significant accumulation of POPG among all the systems. Although the IM-2PK system has a comparable number of charge centers as in the IM-2mon system, the CSM5-K5 monomer has a higher charge density than the PK monomer, since all the positive charges are grafted around the chitosan backbone in the CSM5-K5 case, while those involved in the PK monomer has a linear distribution.

Figure 3.9 Lateral radial distribution function (RDF) of POPG lipids with respect to themselves in membrane systems with or without the presence of polycations. The peak around 0.6 nm illustrates that the dispersion of POPG is greatly changed in IM-dimer system.

The CSM5-K5 dimer is capable of rearranging the anionic and zwitterionic lipids components in the bacterial membrane model, which in a larger timescale might cause the lipid domain formation in the membrane model. This induced lipid domain formation has been discovered in many antimicrobial agents [168-170], which all contain many positive charges as well. Their possible killing mechanism is that the line tension between the domain-boundaries is enlarged with the domain separation thus cause defects to the boundaries and lower the resistance of the membrane [76,171]. We suggest that the lipid domain formation contributes to the antimicrobial activity of CSM5-K5, and the concentration of POPG in different systems is consistent with the minimum inhibitory

Figure 3.10 Accumulation of POPG in the different IM-polycation system versus the control. The last 1 ns trajectories are used to compute the average number of POPG molecules as a function of the X and Y coordinates. Each figure is a top view of the simulation box with a size of approximately 7.1*7.1 nm². Polycations are labeled with the magenta color and may cross the box boundary due to the periodic boundary conditions (PBC). POPG lipids in the IM-dimer system are more concentrated than the others.

concentration (MIC) determination of CSM5-K5 and polylysine [145].

3.4 Conclusion

Using well-tempered MetaD simulation and conventional MD simulation, we studied the self-assembly of CSM5-K5 molecules and their interaction with neutrally or negatively charged membrane models. CSM5-K5 dimerizes with the posture of parallel chitosan chains so that maximizing the formation of $\text{—NHC(=O)•••—NHC(=O)}$ H-bonds, which are strong enough to overcome the electrostatic repulsion from the lysine residues. CSM5-K5

recognizes the negatively charged membrane model based on the electrostatic attraction. Binding of CSM5-K5 rearranges the lipid components for domain formation in the membrane, and this might cause defects to the membrane thus kill the bacteria.

Grafting of the polylysine largely ensures the solubility and antimicrobial activity of CSM5-K5, while using chitosan as a backbone introduces aggregation mechanism into CSM5-K5 to further increase the antimicrobial activity. This combination conduces to the effectiveness of CSM5-K5 as an antibiotic.

4 Simulation study of a L_{II} targeting antibiotic - TXB

4.1 Background

L_{II} is one of the most essential intermediate products for the biosynthesis of bacterial cell wall. Recently, Ling *et al.* reported a new L_{II}-targeted antibiotic named TXB by screening the uncultured bacteria with the help of iChip technology [4]. TXB is produced by the Gram-negative bacterium *Eleftheria terrae* and can effectively kill Gram-positive pathogens without any drug-resistance developed. TXB is an undecapeptide with a sequence of N-Me-D-Phe-L-Ile-L-Ser-D-Gln-D-allo-Ile-L-Ile-L-Ser-D-Thr-L-Ala-L-End-L-Ile, including a rare amino acid enduracididine as the 10th residue (End-10). The structures of TXB and L_{II} are shown in Fig. 4.1. There is a methylation on its N-terminal Phe residue (NmPhe-1) as well as a cyclization between the C-terminal Ile-11 and the side chain hydroxyl group of Thr-8 through an ester linkage. TXB forms a complex with L_{II} with a ratio of 2:1 (TXB: L_{II}) and the interaction of TXB with pyrophosphate and N-acetylmuramic acid (MurNAc) group of L_{II} is critical for binding.

The finding of such novel antibiotics stimulated follow-up studies to establish the TXB structure-activity relationships (SAR). Although a few groups have reported the total synthesis of TXB [172,173], or the synthesis of the TXB ring motif [174], the great difficulties in the synthesis of the End-10 residue are still the main obstacles in the synthesis of TXB. Jad *et al.* synthesized a TXB analogue by replacing the End-10 residue with an arginine residue [175] and found this analogue only had a slight

Figure 4.1 Structure of simulated TXB (a) and L_{II} (b). TXB contains 11 residues: N-Me-D-Phe-L-Ile-L-Ser-D-Gln-D-allo-Ile-L-Ile-L-Ser-D-Thr-L-Ala-L-End-L-Ile, and the last four residues form a closed ring. (NmPhe: N-methylated phenylalanine, End: enduracididine). L_{II} model used in our simulation was published by Hst *et al.* [87], which has a shortened hydrocarbon chain (15 carbon atoms) than the wild-type (55 carbon atoms).

increase of the MIC against Gram-positive bacteria when compared to TXB. SARs of TXB and several new analogues indicate that the D-amino acids are important for activity, as almost a total loss of activity, was observed when they were replaced by L-amino acids [176-178]. Further, modification of NmPhe-1 by the substitution of a methyl group with an acetyl group [178], or substitution of the NH proton with one more methyl group, or introduction of a hydroxyl group to aromatic ring [179], reduced the antimicrobial activity. Nowick's group reported the importance of the ring motif and suggested the TXB-L_{II} binding was through H-bond interaction between amide NH groups on the ring motif of TXB and pyrophosphate group of L_{II} [180], which was similar to nisin-L_{II} binding [87]. The importance of hydrophobicity from the N-terminal tail was also discussed in their work: the TXB analogue still remained bactericidal after replacing the residues NmPhe-1 to Ile-5 with a dodecanoyl group.

These previous studies provided clues on the important functional groups of TXB interacting with L_{II}. However, a detailed structural model of the TXB-L_{II} complex is not available. In this chapter, MD simulations with an enhanced sampling technique were performed to explore the binding phase space of TXB-L_{II} complex with a ratio of 1:1, to figure out the specific binding modes of TXB and L_{II}. Four different binding modes are discovered in our simulations, and the results indicate the ring motif of TXB plays an important role in recognizing L_{II}. Structural models of the TXB-L_{II} complex with a ratio of 2:1 are suggested, which are also found stable in the lipid bilayer environment. These discoveries provide new insights to understand the binding mechanism of TXB to L_{II}.

4.2 Parameters

4.2.1 Unbiased MD simulation details

All the amino acid residues in TXB and L_{II} molecules are described by AMBER99SB force field [149], except End-10 in TXB, which is not defined in AMBER99SB. We chose GLYCAM_06j-1 force field [148] to describe the sugar residues GlcNAc-2 in L_{II}. However, the other sugar residue MurNAc was not available in the force field. MurNAc could be considered as the ether of lactic acid and GlcNAc, so it was divided into two residues GlcNAc-1 and Lac-5. Same as End-10 in TXB, Lac-5 is not a standard residue either, therefore, we generated their topologies, as well as the topologies of other undefined residues like C15-3 and PP-4 in L_{II}, based on generalized amber force field (GAFF) [181]. Partial charges for the undefined residues (End-10 of TXB, C15-3, PP-4 and Lac-5 of L_{II}) were fitted based on Restrained Electrostatic Potential (RESP) method in which the electric potential was calculated with Gaussian 09 at the level of HF/6-31 G*. The partial charges

of some residues on the ring motif of TXB and Glu-7 of L_{II} bonded with the neighboring residues in a non-standard way were recalculated using the same strategy.

MD simulations were performed with GROMACS package, version 4.6.7 [147]. We mainly focused on the interaction between TXB and L_{II}; therefore, the membrane was not included in the exploration of the binding phase space of TXB-L_{II} complex to save the computational resources. This is reasonable according to the original study [4], which claimed that purified L_{II} could form complex with TXB in vitro, indicating the membrane is not involved in TXB-L_{II} binding. The simulation systems were composed of TXB, L_{II} (ratio of them would be mentioned in the results and discussion section), TIP3P water [151], and counter ions NaCl. Counter ions were modified by Joung *et al.* [152] and added to a concentration of 0.17 M while neutralizing charges of the simulation systems. A leap-frog algorithm was used to integrate Newton's equation of motion, and the time step was set to 2 fs along with the application of LINCS algorithm [150], by which all the covalent bonds between hydrogen atoms and heavy atoms were constrained. The cutoff for both van der Waals interaction and short-range electrostatic interaction were set to 1.0 nm, and a PME method [153] was employed to deal with the long-range electrostatic interaction. After the energy minimization, two short simulations with position restrains of all atoms of TXB and L_{II} were performed to equilibrate the simulation system: 100 ps simulation in NVT ensemble to heat up the system temperature to 300 K and another 100 ps simulation in NPT ensemble to adjust the system pressure to 1 bar. After that, the product simulations were run 200 ns in NPT ensemble with all position restrains removed. The temperature and pressure of the system were maintained by V-rescale temperature coupling method

[154] and Parrinello-Rahman pressure coupling method [155], respectively. Each kind of simulation was repeated three times by random assigning initial velocities.

After the TXB-L_{II} complex was stabilized in the aqueous simulations, we checked the stability of the complex in a membrane environment. The membrane model was built utilizing a web server MemBuilder II and parameterized with a Slipids force field [157-159], which is compatible with AMBER force field. In the prepared lipid bilayer, each leaflet of the membrane contained 48 negatively charged POPG and 16 zwitterionic POPE. Then one POPG was removed from the upper leaflet and our L_{II} model was embedded. The membrane model was equilibrated for 50 ns at 300 K and 1 bar, and its APL reached a value of 0.60 nm², which was comparable with experimental results (0.59 nm² for pure POPE at 303K [182], and 0.64 nm² for pure POPG at 293K [161]). After that, two TXB molecules were added onto the L_{II} molecule in the predicted 2:1 complex configuration in the simulation box and 100 ns production MD simulation was performed. The distances between the ring motifs of the two TXBs and their respective targeting residues of L_{II} were measured to illustrate the stability of the TXB-L_{II} complex.

4.2.2 PTMetaD-WTE

Our PTMetaD-WTE simulation was performed using the Gromacs 4.6.7 package with the help of plumed 2.1.3 plug-in [156]. 16 replicas were simulated in NVT ensemble to avoid the instability of the simulation box in NPT ensemble at high temperature. Temperatures varied from 300 K to 546.6 K and followed an exponential distribution: $T_i = T_0 * e^{k*i}$, where T_i was the temperature of replica i , and $k = 0.04$. Each replica contained 1 TXB, 1 L_{II}, 6506 TIP3P water molecules as well as 24 sodium ions and 21 chlorine ions. 16 starting conformations of the TXB-L_{II} complex with different contacts were generated from a

conventional MD simulation at high temperature (500 K). Before the PTMetaD-WTE simulations, 500 ps simulation was performed to equilibrate each replica.

PTMetaD-WTE simulations were separated into two steps. First, 20 ns PTMetaD simulation was run with biases only on the potential energy. During this step, Gaussian potential was added every 500 calculation steps (1 ps), and a bias factor equal to 60 was adopted. Gaussian height was set to 2 kJ/mol and the Gaussian width was set to 500 kJ/mol. Probabilities of neighboring-replica-exchange could reach 45%. In the second step, the accumulated biases on the potential energy CV in the first step were kept as a static additional bias potential, but no more new biases were added to the potential energy. The number of contacts between TXB and L_{II} heavy atoms was used as the new biased CV. In the MetaD simulation, CV should be wisely chosen to describe the degrees of freedom of system to achieve convergence. However, with the benefit from WTE, such limitation is largely reduced. The reason we chose the number of contacts as CV, instead of other generic CV such as inter-molecular distance, is that the contacts naturally include the information of interacting pairs between the two molecules, which is closely related to the binding modes. The number of contacts S was defined as the number of heavy atom pairs with a distance smaller than 0.6 nm, within which each heavy atom was from different molecules. Its value became differentiable by adopting a switching function:

$$S = \sum_{i \in A} \sum_{j \in B} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^6}{1 - \left(\frac{r_{ij}}{r_0}\right)^{12}} \quad (4.1)$$

where i and j were labels of heavy atoms from molecule A and B respectively, r_{ij} was the pair distance and r_0 (0.6 nm) was the distance cutoff. During this step, the values of bias factor and Gaussian width were set as 8 and 5 respectively, Gaussian height and time

interval of Gaussian deposition were still 2 kJ/mol and 1 ps. 400 ns PTMetaD-WTE simulation was finished for each replica.

In prediction of the 2:1 TXB-L_{II} complex, two possibilities were discussed in this chapter: the one in which one L_{II} binds to two TXB separately, and the one in which one L_{II} binds to a TXB dimer. In the former situation, the complex was constructed by combining two of the discovered binding modes in the study of the 1:1 TXB-L_{II} complex. However, in the latter case, the TXB dimer should be carefully designed. Thus we went through the above procedures for another 400 ns PTMetaD-WTE simulation to study the dimerization of two TXB molecules. The same parameters were utilized except each replica of this simulation contained 2 TXB, 6510 water molecules, 21 sodium ions, and 21 chlorine ions.

4.3 Results and discussion

4.3.1 Unbiased MD simulations of TXB and L_{II} binding with a ratio of 1:1

The high flexibility for both TXB and L_{II} increases the diversity of binding gestures. With so many polar residues, such as charged residues PP-4, Glu-7, Lys-8, Ala-10 (charged C-terminal residue) in L_{II}, and the polar ring motif in TXB, a lot of local minima could be foreseen on the FES of TXB-L_{II} binding. Three repeats of the simulation were run for 200 ns to investigate the binding of one TXB and one L_{II}. The number of contacts between the two molecules (Fig. 4.2) showed different binding status for each different repeat. However, no stable binding mode could be found from these three repeats and in some cases (Fig. 4.2 repeat2 & 3) the configuration was trapped in the local minima for about 100 ns. It is

Figure 4.2 Number of contacts between TXB and L_{II} in three unbiased MD simulation. Only heavy atoms were counted in both molecules. The cutoff was set as 0.6 nm.

very difficult for an exhaustive exploration by unbiased MD simulations because they are often trapped in the local minima; therefore, an enhanced sampling method is required.

4.3.2 TXB- L_{II} binding modes explored by PTMetaD-WTE simulation

The number of contacts between 1 TXB and 1 L_{II} in our PTMetaD-WTE simulation was selected as a biased CV. Based on the diffusion of the CV (Fig. 4.3 a) and decay of the Gaussian height at 300 K (Fig. 4.3 b), as well as the diffusive movement of replicas in the temperature space (Fig. 4.4), the PTMetaD-WTE simulation was converged. Searching for specific binding modes is difficult for the highly flexible molecules with the structure-based clustering methods, instead, we utilized the contact map (Fig. 4.5a) established based on the heavy-atom pairs of all frames in the 300 K ensemble. The unbiased frequency of each

Figure 4.3 Diffusion of the number of contacts (a) and decay of the Gaussian height (b) as a function of time in the PTMetaD-WTE simulation.

Figure 4.4 Diffusive movement of replicas in temperature space.

Figure 4.5 Atom-based contact map (a) and the histogram of different clusters (b). In (a) axis labels in red color with smaller size are the heavy atom ID, axis labels in black color with larger size are the residue ID. Based on the contact map, 25 contacted residue-pairs were labelled with the letters from a to y. Each simulated frame was relabeled with these 25 letters by identifying the existing contacts and clustered into groups. (b) Presents the 10 most popular clusters.

frame was estimated as $e^{\beta(V(s(R),t)-c(t))}/Z$ based on Tiwary and Parrinello's method [125], hence the contact-frequency of heavy atoms i and j was calculated as $freq_{i,j} = \frac{\sum P(t)_{i,j} \cdot e^{\beta(V(s(R),t)-c(t))}}{Z}$, where $P(t)_{i,j} = 1$, when the distance between atoms i and j is no larger than 0.4 nm, and $P(t)_{i,j} = 0$, when their distance is larger than 0.4 nm. Compared with the residue-based contact map, this atom-based contact map avoids the accumulative effects, i.e. several atom pairs with low frequency ends up to one residue pair with high frequency, thus provides contact information with higher accuracy.

According to the contact map, 25 residue-pairs with high occupancy were selected and labelled with the letters from a to y, respectively. The 25 contacts are statistically meaningful, but for a specific binding mode, usually not all the 25 contacts are involved. The next step was to re-label each frame obtained from the simulation by identifying which

contacts existed. For example, if a frame is labeled as “opqrst”, it means the contacts of o, p, q, r, s, and t exist in this frame. In this way, all the frames could be re-grouped by their labels. The label, which is a combination of the contact letter a to y, functions as a natural metrics for binding mode identification. The 10 most popular binding modes are shown in Fig. 4.5b. Among the ten modes, four predominant binding modes were identified, while the other six binding modes in the top 10 list are derivatives of the first 4 modes from adding or missing 1 or 2 contacts.

The gromos method [162] was used to classify the frames of each binding mode, based

Figure 4.6 Four binding modes and their featured contacts. In BM1, featured contacts are maintained by four amide groups of the ring motif of TXB form H-bonds with Glu-7 of L_{II}. In BM2, besides the H-bonds between four amide groups of the ring motif of TXB and PP-4 of L_{II}, there are also one H-bond between Ser-7 of TXB and PP-4 of L_{II}, as well as two H-bonds between the side chain of End-10 of TXB and GlcNAc-1 of L_{II}. In BM3, H-bonds are formed between Ser-3 of TXB and Glu-7 of L_{II}, Gln-4 of TXB and Lys-8 of L_{II}, and Gln-4 of TXB and Ala-10 of L_{II}. In BM4, H-bonds are between ring motif of TXB and Ala-10 of L_{II}. Polar hydrogen, nitrogen, oxygen and phosphorus atoms are colored as white, blue, red, and orange respectively, non-polar hydrogen atoms are hidden. Carbon atoms are colored as cyan (TXB) and green (L_{II}) to differentiate different molecules. H-bonds donor and acceptors are labelled with the red letters.

on RMSD with respect only to the contacted residues with a cutoff of 0.25 nm. Concretely, that is RMSD of residues Ala-6, Glu-7 (L_{II}) and Ser-7, Thr-8, Ala-9, End-10, Ile-11 (TXB) for binding mode 1 (BM1); RMSD of residues GlcNAc-1, PP-4 (L_{II}) and Ser-7, Thr-8, Ala-9, End-10 (TXB) for binding mode 2 (BM2); RMSD of residues Glu-7, Lys-8, Ala-9, Ala-10 (L_{II}) and Ile-2, Ser-3, Gln-4 (TXB) for binding mode 3 (BM3); and RMSD of residues Ala-10 (L_{II}) and Ser-7, Thr-8, Ala-9, End-10, Ile-11 (TXB) for binding mode 4 (BM4). The representative conformations of each binding mode are presented in Fig. 4.6 a1, b1, c1 and d1, and their featured contacts are in Fig. 4.6 a2, b2, c2 and d2, respectively. Starting from these representative structures, three repeats of standard MD simulation, each of which had a

Figure 4.7 Residue based contact maps of the first 10 ns simulation and last 10 ns for the four binding modes. Representative structures of each binding mode were simulated for 3 repeats with 200 ns each to check their stability. The contact cutoff was set as 0.4 nm and only heavy atoms were considered. The region of featured contacts for all the four binding modes have no changes when comparing the contact maps of the first 10 ns simulation and the last 10 ns simulation, indicating the strong stability of the four binding modes.

length of 200 ns, were performed for each binding mode. Stabilities of these binding modes were confirmed by comparing the contact maps for the first 10 ns simulation and last 10 ns simulation of each MD repeat (Fig. 4.7). In all four cases, both molecules in the complex remain very flexible (at the end of the simulations, the whole molecule RMSDs could reach 1.0 nm, 0.9 nm, 0.4 nm, and 0.8 nm for BM1, BM2, BM3 and BM4 systems, respectively. The RMSDs are shown in Fig. 4.8) and more importantly, the featured contacts between them are maintained throughout the whole simulations (the featured contacting residues RMSDs could only reach 0.15 nm, 0.15 nm, 0.2 nm, and 0.15 nm for BM1, BM2, BM3 and BM4 systems, respectively Fig. 4.8).

Figure 4.8 RMSD in the three repeats of stability-checking simulations for the four different binding modes.

Black, red and blue colors represent repeat1, repeat2, and repeat3, respectively, and solid lines and dot lines represent the RMSD of whole complex (TXB and L_{II}) and RMSD of featured-contacting residues, respectively.

The featured contacts of BM1 are mainly maintained by the H-bonds interaction between the four amide groups on the TXB ring motif (from residue Thr-8 to residue Ile-11) and the carboxyl group of Glu-7 on L_{II}. BM2 is characterized by the H-bonds between four amide groups on TXB ring motif and PP-4 of L_{II}, additionally, one H-bond is also found between the hydroxyl group of Ser-7 of TXB and PP-4 of L_{II}, as well as two H-bonds between the sidechain-ring of End-10 of TXB and GlcNAc-1 of L_{II}. BM3 is stabilized by three H-bonds, one of which is formed between the backbone carbonyl group of Gln-4 and the amide group of Ala-10, one is between the backbone amide group of Gln-4 and the carbonyl group of Lys-8, and another one is formed between the hydroxyl group of Ser-3 and the carboxyl group of Glu-7. And BM4 is similar to BM1, the difference is the carboxyl group is from Ala-10 instead of Glu-7. Occupancy for each of these H-bonds during 600 ns simulation (3 repeats with 200 ns each) is listed in Table 4.1.

Further analysis of the featured contacts of BM1 and BM4 finds that the carboxyl groups of Glu-7 and Ala-10 come from the backbone: Glu-7 covalently bonds to Lys-8 via the side chain carboxylate instead of the backbone carboxyl and Ala-10 is located at the C-terminus. These two carboxyl groups cannot be changed by side-chain mutations. Both carboxyl groups could rotate during the binding. Therefore, both oxygen atoms in each carboxyl have the opportunity to form H-bonds with the four amide groups in the ring motif of TXB (Table 4.1). On average, BM1 has 4.27 H-bonds (obtained by simply adding the occupancy) between the carboxyl and the ring motif, while BM4 has 3.92. Thus, we suggest that BM1 is slightly more stable than BM4.

In BM2, the phosphate groups of L_{II} interact with the ring motif and Ser-7 of TXB individually (Fig. 4.6b2). They both have H-bonds with occupancy higher than 90%

Table 4.1 Details of the H-bonds maintaining the featured contacts in each binding mode. The occupancy was calculated over 3*200 ns of MD simulations. Atom names could be found in Fig. 4.6.

Binding Mode	Donor	Acceptor	Occupancy (%)
BM1	N1H (Thr-8)	O1 (Glu-7)	63.0
	N2H (Ala-9)	O1 (Glu-7)	52.8
	N3H (End-10)	O1 (Glu-7)	50.1
	N4H (Ile-11)	O1 (Glu-7)	45.3
	N1H (Thr-8)	O2 (Glu-7)	69.9
	N2H (Ala-9)	O2 (Glu-7)	47.6
	N3H (End-10)	O2 (Glu-7)	47.3
	N4H (Ile-11)	O2 (Glu-7)	51.0
BM2	N1H (Thr-8)	O3 (PP-4)	99.4
	N2H (Ala-9)	O3 (PP-4)	81.5
	N3H (End-10)	O3 (PP-4)	70.7
	N4H (Ile-11)	O3 (PP-4)	83.8
	O5H (Ser-7)	O4 (PP-4)	90.5
	N5H (End-10)	O6 (GlcNAc-1)	57.5
	N6H (End-10)	O7 (GlcNAc-1)	51.5
BM3	O10H (Ser-3)	O1 (Glu-7)	41.2
	O10H (Ser-3)	O2 (Glu-7)	34.5
	N7H (Gln-4)	O8 (Lys-8)	98.2
	N8H (Ala-10)	O9 (Gln-4)	81.3
BM4	N1H (Thr-8)	O11 (Ala-10)	75.4
	N2H (Ala-9)	O11 (Ala-10)	66.4
	N3H (End-10)	O11 (Ala-10)	49.1
	N4H (Ile-11)	O11 (Ala-10)	58.8
	N1H (Thr-8)	O12 (Ala-10)	40.9
	N2H (Ala-9)	O12 (Ala-10)	34.6
	N3H (End-10)	O12 (Ala-10)	42.2
	N4H (Ile-11)	O12 (Ala-10)	24.7

(Table 4.1), indicating the interaction is very strong. The sidechain of End-10 of TXB also forms two H-bonds with GlcNAc-1 of L_{II} to further strengthen this binding. On average, the total number of H-bonds in BM2 could reach 5.35, which is higher than BM1 and BM4.

Since the ring motif of TXB makes polar-contacts only with one phosphate group of the pyrophosphate residue, we examined whether one phosphate group is enough for a stable binding. Two new simulation systems were designed to verify the importance of the PP-4 motif. One system contained one TXB molecule and one C15P molecule, which was

Figure 4.9 Distance between ring motif of TXB and phosphate group (C15P, left panel) or pyrophosphate group (C15PP and L_{II}, middle and right panel, respectively). Binding in TXB & L_{II} system is the most stable out of three, indicating that both pyrophosphate moiety and the first sugar moiety of L_{II} are required for a stable binding with ring motif of TXB.

composed of a bactoprenol carbon chain covalently bonded to a phosphate group. The other system contained one TXB and one C15PP molecule, which was made of a carbon chain and a pyrophosphate group. Force field parameters for both C15P and C15PP molecules were taken from GAFF. Each system was run by triplicate and simulations started from the bound conformation. Distances between the ring motif and the phosphate (TXB & C15P system) or the pyrophosphate (TXB & C15PP system and TXB & L_{II} system) are shown in Fig. 4.9. Data show that the binding of the ring motif and the phosphate (TXB & C15P system) is unstable. Adding one more phosphate to the system (TXB & C15PP system) rendered binding much more stable because of the higher polarity of the pyrophosphate group than one phosphate group. However, binding in TXB & C15PP system was not as stable as that in TXB & L_{II} system. The inter-group distance in TXB-C15PP system showed larger fluctuation than that in TXB-L_{II} system at the end of simulations and unbinding happened in repeat1 of TXB-C15PP system, indicating that the introduction of GlcNAc-1 in L_{II} further stabilized the binding. Our results agree with the experimental finding [4] that the minimal motif of L_{II} required for stable binding with TXB is the

Figure 4.10 Binding energy of different binding modes and their comparison with experimental structures.

Binding energies (a) were calculated based on 100 conformations from last 5 ns of the stability-check simulation of each binding mode. Binding affinity in a descending order is BM2 > BM1 > BM4 > BM3, indicating the importance of ring motif. We superimposed the ring motifs (BM1, BM2, and BM4) onto the ring motif of a published x-ray crystallographic structure (b embedded figure), and calculated RMSD of our ring motif with respect to the published ring motif (b). Side chains were not included in the RMSD calculation since the published structure was a TXB analogue, which has an Arg-10 residue rather than the End-10 residue. The fundamental interactions in these ring-involved binding modes share features with those in nisin- L_{II} binding: in nisin- L_{II} binding, amide groups on the ring backbone H-bond with negatively charged pyrophosphate (c); similarly, in TXB- L_{II} binding, amide groups on the ring backbone also form H-bonds with negatively charged pyrophosphate (BM2, d), or with negatively charged carboxyl (BM1/BM4, e). Negatively charged groups are shown in sphere models: pyrophosphate groups in (c,d), and the carboxyl group in (e).

pyrophosphate (PP-4) and the first sugar (GlcNAc-1) moieties, both of which are highly conserved in bacteria [79].

The binding energies of the four binding modes were estimated (Fig. 4.10a) using the MM-PBSA method with the tool of g_mmpbsa published by Kumari et al. [129]. However, in MM-PBSA method [132], the selection of the dielectric constant of the solute would be tricky [131,183]. The charge-charge interaction related terms in binding energy, such as the polar solvation term, and especially the electrostatics term, can be largely affected by the solute dielectric constant. We adopted three different dielectric constants, and although the binding energy of each binding mode varied at different dielectric constants, the overall trend at the same dielectric constant is that $E_{BM2} < E_{BM1} < E_{BM4} < E_{BM3}$, suggesting that BM2 has the highest binding affinity among the four.

Three out of the four binding modes (BM1, BM2, and BM4) utilize the ring motif of TXB to attract the charged residues of L_{II}, indicating that TXB ring motif plays a critical role in recognizing L_{II}. Recently, a series of TXB analogues with varying structure and stereochemistry were synthesized to elucidate the TXB pharmacophore. The ring structure was found critical to bactericide function [180]. Further, a crystallographic structure of a TXB analogue binding with a chloride anion [184] shows the H-bonding between the ring motif and a negatively charged ion (Fig. 4.10b embedded). To validate our predicted ring structure, the root-mean-squared-deviations (RMSD) of the backbone atoms of the ring motif with respect to the crystallographic structure were calculated (Fig. 4.10b). The RMSD values are only around 0.4 Å for BM1, BM2 and BM4, indicating our predicted ring structure agrees with the crystallized structure. The published structure also supported the H-bonding pattern predicted by our model. However, our model has some differences with the crystallographic structure where the amide group of Ala-9 (TXB) is H-bonded with the hydroxyl group of Ser-7 (TXB), and the other three amide groups interact with the

chloride anion. In the ring motif of our model, all four amide groups on the backbone form H bonds with the negatively charged residues.

Comparisons of the binding modes of TXB with other L_{II} binders show some differences and similarities. For instance, vancomycin binds to the negatively charged residue (the D-Ala-D-Ala segment of L_{II}) [83,84] through the formation of a “carboxylate anion binding pocket” [185]. Tridecaptin A1, a Gram-negative L_{II} binder (Gram-negative L_{II} has a meso-diaminopimelic acid instead of a lysine on the penta-peptide motif), forms H-bonds with the carboxylate of meso-diaminopimelic acid, and the pyrophosphate moiety is not

Figure 4.11 Stability of the nisin-L_{II} complex in 100 ns of unbiased MD simulations. This figure shows the distance between the featured-contacting residues, i.e. the ring motif of nisin and the pyrophosphate of L_{II}. According to our simulation, the binding for nisin and L_{II} shows high stability, indicating our force field parameters are reliable for the binding study.

involved in the binding [186]. Mersacidin has an electrostatic-interaction dominated binding to L_{II}, and the terminal GlcNAc sugar is included in the binding [187]. In contrast, Nisin binds to L_{II} through the H-bonds between the backbone amides on the ring structure and the pyrophosphate moiety [87], and shares features with TXB (Fig. 4.10c–e). We performed 100 ns unbiased MD simulation of nisin-L_{II} complex starting from the published structure (pdb: 1WCO) to test the reliability of our force field parameters. The topology file of nisin was developed with the same strategy as we described in the Parameters section of this chapter. The nisin-L_{II} binding mode was well maintained during the 100 ns (Fig. 4.11), indicating that our force field parameters are reliable for the binding study.

4.3.3 Prediction of TXB-L_{II} complex with a ratio of 2:1

Experiments found that one L_{II} could form a stable complex with two TXB molecules [4]. Based on the four 1:1 binding modes found in this study, the structural models of six different 2:1 complexes, labelled as BM1&BM2, BM1&BM3, BM1&BM4, BM2&BM3, BM2&BM4, and BM3&BM4 were constructed and validated through MD simulations with three repeats each for a length of 200 ns.

The contact map of the last 50 ns is presented in Fig. 4.12, and all the featured contacts region were highlighted with the ellipses. In both BM1&BM3 and BM2&BM3, the featured contacts for BM3 were not found in one repeat, and in BM3&BM4, the featured contacts of BM4 partially disappeared, indicating that all these three 2:1 models were not as stable as the BM1&BM2, BM1&BM4 or BM2&BM4. Our results suggest that in TXB-L_{II} binding, two ring motifs of TXB could bind to any two out of the three binding sites supplied by L_{II} (PP-4 & GlcNAc-1 moiety, Glu-7, and Ala-10). However, considering the relatively higher

Figure 4.12 Contact map of the different TXB- L_{II} complex with a ratio of 2:1. Each complex was run for 3 repeats with 200 ns each, and the last 50 ns trajectory was used for the calculation of contact map. Only heavy atoms were considered, and the contact cutoff was set as 0.4 nm. Residue number from 1 to 11 on x-axis belongs to one TXB, and from 12 to 22 belongs to another TXB. Featured contacts areas are highlighted with ellipses.

binding affinity of BM1 and BM2, we believe that the model BM1&BM2 is the most probable model.

Another possibility for the 2:1 binding mode is that TXB molecules first form a dimer, and then bind to one L_{II} molecule. Such binding mode was found for ramoplanin, which dimerizes under membrane environment before binding to L_{II} [188]. Therefore, another 400 ns PTMetaD-WTE simulation was carried out to simulate the interaction between two

Figure 4.13 Study of TXB dimerization based on 400 ns PTMetaD-WTE simulation of two TXB molecules.

An atom-based contact map is shown in the left panel. Axis labels in red color with smaller size are the heavy atom ID, axis labels in black color with larger size are the residue ID. The residue-pairs with high contacting frequencies are labelled with different single letters. Using the same grouping strategy as in Fig. 4.5, all the frames in the simulation are clustered into groups. The right panel presents the 10 most popular clusters. We chose the first two groups (group yzA and group ghklmnpI) to do further analysis since they have common features with the following groups. Group yzA has featured-contacts between the ring motif of TXB1 and Ile-6 of TXB2, and group ghklmnpI has featured-contacts between the Ser-3-Gln-4 segment of TXB1 and ring motif of TXB2.

TXB molecules in solution. Based on the contact map and the subsequent unbiased MD simulation (Figs 4.13, 4.14), a TXB dimer with a featured-contact between the ring motif of one TXB molecule and the residue Ile-6 of another TXB molecule was found stable. A possible 2:1 TXB- L_{II} model was created by positioning an L_{II} molecule close to the other ring motif of the TXB dimer with the pyrophosphate group based on the binding mode of BM2. Nevertheless, this complex was not stable: the ring motif which bound to the Ile-6 promptly lost contacts while the other ring motif strongly attached to the pyrophosphate group of L_{II} (Fig. 4.15). These results suggest that TXB molecules could dimerize in solution, but this dimer is not stable under the competition of L_{II} .

Figure 4.14 Stability of the two TXB dimer models. Starting from the representative structures of group yzA and group ghklmnpI, three repeats of unbiased MD simulations, 200 ns each, were performed to check the stabilities of these two groups. Contact maps between two TXB molecules of the first 10 ns and last 10 ns trajectory for each simulation were compared, and only featured-contacts of group yzA could be maintained during the simulation. Group ghklmnpI is not stable.

Figure 4.15 Stability of the 2:1 TXB-L_{II} complex where the two TXB molecules dimerize first. An L_{II} molecule was inserted into the conformation of group yzA to form a 2:1 TXB-L_{II} complex. This new complex, which had the pyrophosphate group of L_{II} H-bonding with the free ring-motif of TXB2, was used as an initial structure to produce another series of unbiased MD simulation (three repeats with 200 ns each). Above figures present the contacts inside the complex based on the last 50 ns trajectories, and the featured contacts between two TXB molecules are disappeared. These results reveal that TXB molecules form a stable dimer in solution, however, the dimer could not be maintained with the competition of L_{II} molecule.

Therefore, we favor the structural model of 2:1 TXB-L_{II} complex (Fig. 4.16a), in which the ring motif of two TXB molecules bound to the PP-4 & GlcNAc-1 moiety and Glu-7 residue of the same L_{II} respectively. These two contacts function as anchors, while the rest of TXB can move freely since no specific binding patterns were found between the rest residues of both TXB and L_{II} (Fig. 4.12 BM1&BM2) or between the two TXB molecules (Fig. 4.17). Further, this complex was simulated under a membrane environment to verify its stability (Fig. 4.16b). The distances between the ring motif of the first TXB and Glu-7 and the distance between the ring motif of the second TXB and PP-4 were monitored (Fig. 4.16c). The complex was stable in 100 ns MD simulation.

Figure 4.16 The 2:1 TXB-L_{II} complex in the membrane environment. Surface model of the complex is shown in (a). This complex was built based on featured contacts of BM1 and BM2, e.g., the ring motifs of two TXB molecules bound to the PP-4 & GlcNAc-1 moiety and Glu-7 residue of the same L_{II}. The complex in the membrane environment is shown in (b). The distance between the first ring motif and Glu-7 (BM1) and the distance between the second ring motif and PP-4 (BM2) during the simulation in membrane environment are shown in (c). Two TXB molecules are colored as pink and yellow, respectively, and the L_{II} molecule is colored as gray. Lipid molecules in membrane model are colored as gray lines, while their phosphorus atoms are shown as orange spheres to state the position of their head groups.

Figure 4.17 Contact map between two TXB molecules in the BM1&BM2 system. Last 50 ns simulation was used to generate the contact map. No specific binding patterns between the two TXB molecules are found over the three repeats.

4.4 Conclusion

Using PTMetaD-WTE simulations, we explored the binding phase space between the two molecules of TXB and L_{II}. We also developed a strategy to accurately cluster structures for a system with very high flexibility, which was realized by labeling each structure with the featured residue-contacts and grouping the structures with these labels. Four binding modes (BM1, BM2, BM3, and BM4) were discovered. These structural models explain several experimental observations by confirming that: 1) the pyrophosphate-MurNAc moiety of L_{II} is the minimal motif for a stable binding in BM2; 2) The ring motif of TXB is critical for its bactericide function; and 3) Binding sites in L_{II}, the pyrophosphate-MurNAc moiety, the carboxyl group on the backbone of glutamic acid residue and the carboxyl group on the backbone of C-terminus alanine residue play essential roles in the synthesis of the bacterial cell wall. Thus, TXB resistance mutants of L_{II} are unlikely to develop. To further validate such binding modes, X-ray and NMR approaches may be needed. Based on the flexible nature of such binding it is anticipated that direct structure characterization with the crystallography or NMR may be difficult. With our predicted complex models as

a working hypothesis, experiments can be designed to explore the recognition mechanism between TXB and L_{II}.

5 The necessity of D-threonine in TXB

5.1 Background

Among all the SARs of TXB, one of them intrigued our study interests: TXB loses all activity when the D-Thr residue on its ring motif is replaced by the L-isomer [180], although L- and D-isomers share almost all the physicochemical properties except the stereochemistry. And unlike the chiral change of the other D-amino acids on the tail motif [176-178], which are proved important for maintaining the unstructured molecule for the antimicrobial activity [176], chiral change of the D-Thr doesn't alter too much of the molecular flexibility, indicating that the major function of D-Thr is related to the L_{II}-recognition. In the previous chapter, we discussed the binding of TXB ring motif and L_{II} molecule and learning from the SAR study, we realize that the chiral change of D-Thr residue possibly interferes with the existing binding pattern to lose activity. Besides, the bacteria select the D-Thr for the production of TXB by nature, thus, we would speculate that the chiral change interferes with the cyclization process of TXB as well. In this chapter, we try to find out the deterministic reasons why D-Thr instead of L-Thr is needed in this ring motif through computational structural modeling studies.

5.2 Parameters

5.2.1 General parameters

Two possible factors affecting the stability and the structure of the ring motif are considered here: 1) whether D/L-Thr has big effect on the ring structure; 2) whether D/L-Thr affects the configuration of the ring structure and then the binding affinity with the L_{II}. Thus three different kinds of simulations are employed in this study: well-tempered MetaD

simulations of an open ring motif of TXB with sequences of D-Thr-L-Ala-L-End-L-Ile (referred as O-DLLL) and L-Thr-L-Ala-L-End-L-Ile (referred as O-LLLL) to study the FES of the cyclization; unbiased MD simulations of TXB and L_{II} complexes within which the TXB is a normal one or an L-Thr-mutated one to study the influence of the Thr stereochemistry to the binding stability; PT-WTE of closed ring motifs with a D- or L-Thr residue (referred as C-DLLL or C-LLLL respectively) to study their behaviors in solution. Structures of the open or closed ring motifs are shown in Fig. 5.1.

All the MD simulations were performed using GROMACS 4.6.7 [147], with the addition of plumed 2.1.3 plug-in [156] only for the biased MD simulations. The open and closed ring motifs were characterized with AMBER99SB force field [149] (for the Thr, Ala, and Ile residues) and GAFF [181] (for the End residue). The method for partial charges calculation was the same as the one described in the last chapter. During the simulation, the TIP3P water model [151] was used as the solvent. Na⁺ and Cl⁻ were used as the counter ions and added to a concentration of 0.15M. All the covalent bonds in peptides and solvent were constrained by the LINCS algorithm [150] and the SETTLE algorithm [189], respectively. The time step for the integration of the Newton equation of motion was set to 2 fs. The cutoff for both van der Waals interactions and short-range electrostatic interactions were set to 1.0 nm. The PME algorithm [153] was employed to deal with the long-range electrostatic interactions. The V-rescale method [154] and Parrinello-Rahman method [155] were applied for temperature coupling and pressure coupling.

Figure 5.1 Structures of the open and closed ring motif of TXB. A cap residue is added to the N-terminal of the molecules.

5.2.2 Well-tempered MetaD simulations of the open ring motifs

The open ring motif (O-DLLL or O-LLLL) was dissolved into a water box with a size of $4 \times 4 \times 4 \text{ nm}^3$, which contained 2012 water molecules, 6 Na^+ ions, and 6 Cl^- ions. Before the production run in NPT ensemble, the system was equilibrated with position restraint to the peptide in NVT ensemble for 100 ps and NPT ensemble for another 100 ps to reach the desired 300 K temperature and 1 bar pressure. We applied two CVs in the production simulation: one is the distance between the hydroxyl group of the Thr residue (the oxygen atom) and the carboxyl group of the Ile residue (the carbon atom), and the other one is the dihedral angle defined by the four C_α atoms of the motif with a residue order of Thr-

Ala-End-Ile. We believe these two CVs provide good descriptions of the cyclization process. The Gaussian widths of the two CVs were set to 0.04 nm and 0.06 rad respectively. The additional biases were added every 1 ps with an initial Gaussian height 2 kJ/mol. The bias factor was set as 8. Totally 600 ns simulation was performed for each production run.

5.2.3 Unbiased MD simulations of the TXB-L_{II} complex

The initial complex with D-Thr-TXB was obtained from our previously discussed binding mode 2 (BM2), which is mainly maintained by the H-bond interactions between the four amide groups on the ring motif and the pyrophosphate group of L_{II}. The complex with L-Thr-TXB was then generated by mutating only the D-Thr to L-Thr. Both complexes were dissolved into a 5*5*5 nm³ box containing 3906 water molecules, 12 Na⁺ ions, and 11 Cl⁻ ions, respectively. After short equilibration (same as described above), both systems were simulated for 200 ns * 3 repeats in NPT ensemble at 300 K and 1 bar with randomly assigned initial velocities.

5.2.4 PT-WTE simulation of the closed ring motifs

The closed ring motif (C-DLLL or C-LLLL) was placed into a 4*4*4 nm³ box together with another 2015 water molecules, 6 Na⁺ ions, and 7 Cl⁻ ions. A 16-replica-simulation was performed for each ring motif. The temperature sequence in our simulation, 300.00, 310.48, 321.31, 332.50, 344.08, 356.04, 368.40, 381.18, 394.38, 408.03, 422.13, 436.70, 451.76, 467.31, 483.39, 500.00, was generated based on the algorithm of next chapter, with the corresponding bias factor of 11.2. These series of simulations were undertaken in the NVT ensemble. The replicas were equilibrated for 500 ps at their own temperature. Then the WTE frameworks were built within 30 ns simulation: potential energy of each replica was used as CV, and the Gaussian width, Gaussian height, and Gaussian deposition

time interval were set as 500 kJ/mol, 2 kJ/mol, 1 ps, respectively. After that, the added biases were maintained to perform another 200 ns production run. An exchange was attempted every 1 ps.

5.3 Results and discussion

5.3.1 Simulation of the open ring motifs

We performed 600 ns well-tempered MetaD simulations for both O-DLLL and O-LLLL with the biasing distance CV and dihedral CV. The simulations are converged since the FESs in the last 100 ns sampling have only minor changes (Fig. 5.2). The free energy landscapes (Fig. 5.2) are similar for both L- and D-Thr loops. Both the local minima at the distance CV (the distance between the oxygen atom in the hydroxyl group of the Thr residue and the carbon atom in the carboxyl group of the Ile residue) of 0.3~0.4 nm and dihedral CV (the dihedral angle defined by the four C α atoms of Thr-Ala-End-Ile) of -1~1 rad with the depth of -105.1~-101.5 kJ/mol (Fig. 5.2, marked with arrows) indicate that the chiral change of the Thr residue makes no big difference in terms of the propensity to form a closed-loop configuration. All the conformations in these two local minima were analyzed for the detailed distributions of the backbone dihedrals. The Ramachandran plot for each residue in O-DLLL and O-LLLL (excluding the terminal Ile residue) was shown in Fig. 5.3 with the highly visited regions labeled. The frequencies were reweighted based on Tiwary and Parrinello's publication [125]. In Ramachandran plots, the favored regions of residues with different chirality are usually central symmetrical [190]. However, no such correspondences are found for D-Thr and L-Thr in this study. In fact, due to the immobilization of relative

Figure 5.2 FES projections of O-DLLL and O-LLLL on two defined CVs. Definitions of the two CVs are available in section 5.2.2. FESs of the O-DLLL in 500 ns, 550 ns, and 600 ns sampling are shown in the upper three panels, and FESs of the O-LLLL in different sampling time are shown in the lower three panels. No obvious variations are discovered in the FESs of the cyclized region (distance < 0.4 nm) since 500 ns sampling, indicating the convergence of the simulation. The local minima in the cyclized region are marked with black arrows in the 600-ns-sampling FESs.

Figure 5.3 Ramachandran plots of each residue in the O-DLLL and O-LLLL. The terminal Ile residue is not included. The distribution is generated only based on the structures of the cyclization area of the FESs.

position between the hydroxyl group of Thr residue and the carboxyl group of the Ile residue, the dihedral regions of D- and L-Thr in our simulation only preserve parts of the features of the originally symmetrical ones [190]. The Ala and End residues exhibit higher flexibility than the Thr residues by visiting different regions in the phi/psi space and behave quite similar in the two motif.

Each structure was re-labelled based on the favorable backbone dihedral angles in Fig. 5.3 and was subsequently separated into groups according to their labelling. Ranked by their occupancies, the first six groups for O-DLLL and O-LLLL are shown in Fig. 5.4a, and 5.4b respectively, together with the representative structures of the groups with occupancies larger than 10%. There are two popular conformations for O-DLLL ($a_1d_1e_1$ and $a_1c_1f_1$, with a total occupancy of 83.7%), versus three for O-LLLL ($b_2c_2f_2$, $b_2d_2e_2$, and $b_2c_2g_2$, with a total occupancy of 86.8%), indicating the lower flexibility of O-DLLL as compared to O-LLLL when the motifs are close to cyclization. These results allow us to conclude that the

Figure 5.4 The most popular groups of O-DLLL (a) and O-LLLL (b). The representative structures of the high occupancy groups are embedded. In the embedded structures, carbon, hydrogen, nitrogen, and oxygen atoms are colored as green, white, blue and red respectively. C_α atoms are shown as sphere model, and backbones are shown as stick model, while the rest parts of the molecules are shown as line model. The non-polar hydrogen atoms are hidden.

flexibilities during the cyclization are mainly induced by the flexible Ala and End residues, and are reduced with the replacement of L-Thr with D-Thr. In addition, we found the NH groups of the D-/L-Thr residues point to different direction with respect to the ring plane in spite of the flexibilities of all the rest residues (embedded figures in Fig. 5.4), and this might change the interaction between this motif and L_{II} .

5.3.2 Simulation of the TXB- L_{II} complexes

To investigate the chiral influence of Thr residue to the binding, we performed unbiased MD simulations of TXB- L_{II} complexes with D-Thr, as well as L-Thr. We chose the ring-pyrophosphate binding as the initial structure. The complex with L-Thr-TXB was built by only changing the chirality of the D-Thr residue. Despite the high similarities between the initial structures of the two complexes, the complex of L-Thr-TXB turns out to be a weaker binding. The distance between the center of the mass of the ring motif and the pyrophosphate group is shown as a function of simulation time in Fig. 5.5a and 5.5b. The ring-pyrophosphate-binding lasts no longer than 20 ns in the complex with L-Thr-TXB. The initial structure for each complex contains four H-bonds between the amide groups on the ring motif and the pyrophosphate group of L_{II} , however, such configuration is not comfortable for the ring motif with L-Thr. A backbone dihedral transformation was observed only for the L-Thr (Fig. 5.5c and 5.5d) during the first 10 ns simulation. Both backbone dihedral-pairs of Thr end up in their favorable region as shown in Fig. 5.3, and the number of H-bonds between the amide groups and the pyrophosphate group is lessened in the L-Thr complex (2.4 H-bonds on average) than in the D-Thr complex (3.1 H-bonds on average, Fig. 5.5e). Similar with the conformations found in the simulations of the open ring motifs, the amide groups of the Thr residues with different

Figure 5.5 Stability of the ring-pyrophosphate binding. Center of mass distance between the ring motifs and the pyrophosphate groups in the D-Thr-TXB complex (a) keeps a small value in 200 ns simulation, while that of the L-Thr-TXB complex (b) varies largely. The initial backbone dihedral is comfortable for the D-Thr residue (c) rather than the L-Thr residue (d), leading to a big conformational transformation of L-Thr, and as a result, the number of H-bonds between the amide groups and pyrophosphate group is lessened in the L-Thr-TXB complex (e). Binding model is presented in figure (f) for D-Thr-TXB complex and figure (g) for L-Thr-TXB complex. Only the section from Ser₇ to Ile₁₁ of TXB (in stick model), as well as the pyrophosphate group of L₁₁ (in sphere model), is illustrated. In figure f and g, carbon, hydrogen, nitrogen, and oxygen atoms are colored as green, white, blue and red respectively. The non-polar hydrogen atoms are hidden.

chirality have different preferences of pointing direction (Fig. 5.5f and 5.5g).

In the published X-ray crystallographic structure [184], the TXB analogue captures a chloride anion through H-bonds “cage” formed by the NH groups of residue L-Ser-7, D-Thr-8, L-Arg-10, and L-Ile-11, as well as the side chain of L-Arg-10. The NH group of L-Ala-9 is H-bonding to the hydroxyl group of L-Ser-7. This structure reveals a possible binding pose of the ring motif for the first time. However, the pose could have alteration when the anion changes to the size of a pyrophosphate. Our prediction of the binding mode is quite similar to the crystal structure model with only a few differences. In our simulation, the NH groups of D-Thr-8, L-Ala-9, L-End-10, and L-Ile-11 H-bonds to half of the pyrophosphate, while the hydroxyl group H-bonds to the other half. The contribution of each H-bond to the binding stability is different due to the different donor location: NH groups of L-Ala-9, L-End-10, and L-Ile-11 situate right on the ring plane, whereas the NH group of D-Thr-8 branches out by a C-N covalent bond. The protruding NH group deepens the binding cavity for a stable contact, thus the correlated H-bonds seems more important than the others, which is indicated by the highest occupancy among the four (Table 5.1).

Unlike the other three membrane-anchoring-related D-amino acids on the tail motif of TXB [180], the D-Thr directly participates in the formation of the binding cavity, thus its chirality change is sensitive to the recognition of L_{II}. Our simulation indicates that

Table 5.1 Occupancies of the H-bonds between the ring motif and pyrophosphate group in the D-Thr TXB-

L_{II} complex. Data were averaged over 200 ns * 3 repeats.

Donor	Acceptor	Occupancy (%)
D-Thr	pyrophosphate	92.2
L-Ala		63.2
L-End		87.2
L-Ile		65.6

the conformational change from D-Thr to L-Thr cracks one of the H-bonds between the ring motif and the pyrophosphate group, and destabilizes the binding eventually. Parmar et al. reported the stable binding of the L_{II} and an inactive TXB analogue with all four D-amino acids substituted by their L-counterparts [176]. Nevertheless, this does not sufficiently prove that the ring motif with L-Thr binds to the pyrophosphate with the same binding pose. The orientation change of the NH group is based on the nature of the chiral Thr, thus the loss of the related H-bond is inevitable. Although the presented binding pattern is not stable for the ring motif with L-Thr, it is still possible that the pyrophosphate slip to a new binding site, which probably impacts the function of the tail motif. However, such a process is not accessible in the present 200 ns unbiased MD simulation.

5.3.3 Simulation of the closed ring motifs

Based on the above simulations, the different pointing orientation of the NH groups in the L-Thr disturbs H-bonding of the ring motif with the charged residues, thereby interfering with the binding stability. To study the NH-orientation of the closed ring motif and get rid of the influences from other residues, we performed 200 ns PT-WTE simulations. The sampling is efficient enough with the round trip time (RTT) of 1.87 ± 0.27 ns and 1.76 ± 0.29 ns for C-DLLL and C-LLLL, respectively. The structures in the 300 K ensemble were collected for analysis. With the constraints of closed ring, the Ramachandran plot of residues in the first 150 ns simulation is similar with that in the whole 200 ns simulation, indicating that our simulation is converged (Fig. 5.6). In addition, the plots for C-LLLL varies from that of O-LLLL: the number of favored regions for Ala and End reduces from 2 and 3 in O-DLLL to 1 and 1 in C-DLLL, respectively (Fig. 5.6). With the same strategy, structures in the 300K ensemble were grouped and ranked by

Figure 5.6 Ramachandran plot of each residue in the C-DLLL and C-LLLL in the 300K ensemble. The Ile residue is not included. The plot is build based on 150 ns sampling (upper panel) and 200 ns sampling (lower panel). The distribution doesn't change during the last 50 ns simulation, indicating the convergence of our simulation.

occupancies in Fig. 5.7. Only one and two popular conformations are found for C-DLLL ($a_1d_1e_1$, with occupancy of 92.8%) and C-LLLL ($b_2c_2f_2$ and $b_2c_2g_2$, with a total occupancy of 72.0%), respectively. The molecular flexibility is further reduced in the closed ring motifs than in the open ring motifs, especially for the ring motif with D-Thr. The NH groups of D-Thr or L-Thr in the closed ring motifs show no differences on their pointing orientations

Figure 5.7 The most popular groups of C-DLLL (a) and C-LLLL (b). The representative structures of the high occupancy groups are embedded. The coloring schemes are the same with Fig. 5.4.

versus the open ring motifs. It is also noticeable from the representative structures that the four NH groups point to the same side of the ring plane only in C-DLLL. This specialty of C-DLLL offers a better recognition platform in forming H-bonds with the charged residues.

To do a quantitative investigation on the orientation of all the NH groups, we defined a plane based on the C_α atoms of the residue Ala, End, and Ile, and the positive direction of the plane normal vector is set by the right hand rule. The cosine values of the angle between each NH vector and the normal vector of the defined plane were monitored for 20001 frames (Fig. 5.8a), which were generated from the trajectories of the 300K ensemble with a time interval of 10 ps. The width of each frame in the color map is adjusted based on their weight factor. Having a positive value or a negative value indicates that the corresponding NH group pointing to the same direction or the opposite direction as the normal vector of the ring plane. NH group of the Thr residue attaches to the ring plane through one more C-N covalent bond, thus the absolute value of its cosine is closer to zero than others. We distributed the sum of the four cosine values for C-DLLL and

Figure 5.8 Orientations of NH groups on the ring motif. Orientations of different NH groups (the 300 K ensemble) of C-DLLL and C-LLLL are shown in figure a upper panel and lower panel, respectively. Cosine value of the included angle between the NH vector and the normal vector of the ring plane is indicated by the color from red to blue. The width of each frame in the x axis is adjusted based on its weight factor. The four cosine values of each ring motif are summed up and the distribution of the sum value is shown in figure b.

C-LLLL in Fig. 5.8b. In the case of C-DLLL, the four NH groups face the opposite direction of the normal vector of the ring plane in most of the time, except the NH group of L-End residue, which reverses only with a small probability. As a reflection, there is only one major peak for C-DLLL in Fig. 5.8b. For the C-LLLL, on the other hand, the directions of NH groups are more flexible: the NH groups of the L-Thr and the L-Ala point to the same and opposite direction as the normal vector respectively, and the NH groups of L-End and L-Ile were observed to flip over with considerable probabilities. Accordingly, two major peaks are observable for C-LLLL in Fig. 5.8b. Peaks in Fig. 5.8b correspond to the three popular groups in Fig. 5.7.

Clearly only with D-Thr is the ring motif capable of adopting a conformation with all four NH groups facing the same side of the ring plane. Such a structural configuration of the ring motif guarantees an ideal binding cavity for the pyrophosphate group. It is easy to deduce that a ring motif with a sequence of L-Thr-D-Ala-D-End-D-Ile is also functional to

the binding. Since all the four NH groups flip over along with the chiral change, an opposite binding cavity is achievable for the achiral pyrophosphate group. This deduction has already been proved by experiments [180]. Any modification to the relative chirality of the ring motif causes a loss of activity to varying degrees, such as L-Thr₈-L-Ala₉-L-Arg₁₀-L-Ile₁₁-TXB [180], D-Thr₈-L-Ala₉-L-Arg₁₀-D-Ile₁₁-TXB [180], and D-Thr₈-L-Ala₉-D-Ala₁₀-L-Ile₁₁-TXB [191], because of the loss of corresponding H-bonds.

5.4 Conclusion

With different chiral Thr residue, we simulated the closed and open ring motifs of TXB as well as the TXB-L_{II} complexes in this chapter. Chirality of Thr residues causes no energetic difference during the cyclization process but leads to different NH orientations with respect to the ring plane. This distinction provides the D-Thr-ring with a deeper binding cavity for the pyrophosphate group, and the transformation from D-Thr to L-Thr destabilizes the binding mode by cracking the most important H-bond between the ring motif and the pyrophosphate group. The other NH groups on the ring motif are relatively flexible to be adjusted. Only in ring motifs with D-Thr, a favored binding cavity for pyrophosphate is achievable with all the four NH groups facing the same side of the ring plane. This study develops deep understandings of the binding mechanism of TXB and L_{II}, and is expected to be beneficial to the TXB based drug design.

6 Optimization of the temperature ladder in PT-WTE simulations

6.1 Background

PT achieves the highest sampling efficiency when the AAP between neighboring replica pairs is uniform [192]. In such a case, a well-designed temperature sequence would be required for the optimized performance of PT. Researchers have developed several methods to generate the temperature sequence. For instance, Kofke showed that for systems with a constant heat capacity, a geometric progression of temperature sequence assured the same AAP among all the neighboring replica pairs [193-195]. Rathore et al. found that AAP was relevant to the value of $\frac{U(T_i) - U(T_{i-1})}{\frac{1}{2}(\sigma(T_i) + \sigma(T_{i-1}))}$, where $U(T_i)$ is the average potential of replica at temperature T_i , and $\sigma(T_i)$ is the distribution width of the Gaussian-like potential at that temperature [192]. Before using this algorithm, short simulations for the system at several different temperatures are necessary to estimate the average potential and the potential distribution width. Then the temperatures for PT could be obtained by iteratively solving the equation $\frac{U(T_i) - U(T_{i-1})}{\frac{1}{2}(\sigma(T_i) + \sigma(T_{i-1}))} = \frac{U(T_{i+1}) - U(T_i)}{\frac{1}{2}(\sigma(T_{i+1}) + \sigma(T_i))}$. Patriksson et al. came up with a much friendlier using method [196], on the basis of the assumption of constant heat capacity: description of $U(T_i)$ and $\sigma(T_i)$ in the iterative equation were detailed down to the level of degree of freedom, thus users only needed to provide the information of system size and number of constraints to get the optimal allocation of temperatures with a specified AAP.

In PT, the AAP between two neighboring replicas $\langle P_{acc}(T_1 \leftrightarrow T_2) \rangle$ could be estimated as:

$$\langle P_{acc}(T_1 \leftrightarrow T_2) \rangle = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{U_1} P_{(2)}(U_2) \cdot P_{(1)}(U_1) dU_2 + \int_{U_1}^{+\infty} e^{(\beta_1 - \beta_2) \cdot (U_1 - U_2)} \cdot P_{(2)}(U_2) \cdot P_{(1)}(U_1) dU_2 \right) dU_1 \quad (6.1)$$

where $P_{(1)}(U_1)$ and $P_{(2)}(U_2)$ are the probability of having instantaneous potential energy, U_1 and U_2 , for replicas at temperature T_1 and T_2 , respectively. $P_{(1)}(U_1)$ and $P_{(2)}(U_2)$ should have Gaussian shapes with their normalized expression as: $P_{(1)}(U_1) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(U_1 - \mu_1)^2}{2\sigma_1^2}}$ and $P_{(2)}(U_2) = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(U_2 - \mu_2)^2}{2\sigma_2^2}}$ respectively, where μ_1 , σ_1 , μ_2 and σ_2 are the averaged value and standard deviation of each individual replica. Equivalently, equation 6.1 could be described as [196]:

$$\langle P_{acc}(T_1 \leftrightarrow T_2) \rangle = \int_{-\infty}^0 P_{(1,2)}(U_{1,2}) dU_{1,2} + \int_0^{+\infty} e^{(\beta_1 - \beta_2) \cdot U_{1,2}} \cdot P_{(1,2)}(U_{1,2}) dU_{1,2} \quad (6.2)$$

where $P_{(1,2)}(U_{1,2})$ is the probability of having energy different $U_{1,2} = U_1 - U_2$, with its normalized form: $P_{(1,2)}(U_{1,2}) = \frac{1}{\sigma_{1,2} \sqrt{2\pi}} e^{-\frac{(U_{1,2} - \mu_{1,2})^2}{2\sigma_{1,2}^2}}$, where $\mu_{1,2} = \mu_1 - \mu_2$ and $\sigma_{1,2} = \sqrt{\sigma_1^2 + \sigma_2^2}$ are the new average and standard deviation respectively.

Based on equation 2.24, the biased potential distribution in WTE still has a Gaussian-shape, thus, AAP between replica pairs in PT-WTE is also predictable:

$$\langle P_{\gamma, acc}(T_1 \leftrightarrow T_2) \rangle = \int_{-\infty}^0 P_{\gamma, (1,2)}(U_{1,2}) dU_{1,2} + \int_0^{+\infty} e^{\frac{1}{\gamma}(\beta_1 - \beta_2) \cdot U_{1,2}} \cdot P_{\gamma, (1,2)}(U_{1,2}) dU_{1,2} \quad (6.3)$$

where

$$P_{\gamma, (1,2)}(U_{1,2}) = \frac{1}{\sigma_{1,2} \sqrt{2\pi\gamma}} e^{-\frac{(U_{1,2} - \mu_{1,2})^2}{2\gamma\sigma_{1,2}^2}} \quad (6.4)$$

It should be noticed that the instantaneous acceptance probability expressed in equation 2.22, is tuned by γ when compared to the canonical ensemble, since the Boltzmann factor is tuned by γ , based on equation 2.23. In equation 6.3, the AAP is a function of the potential average and potential distributional width of the two replicas, as well as the bias factor. This relationship is able to be further simplified by replacing the potential average and distributional width with the expression of their individual temperatures, such as the linear relationships $\mu(T) = A_1 + B_1T$ and $\sigma(T) = A_2 + B_2T$ (A_1 , B_1 , A_2 , and B_2 are four constants) for systems with constant heat capacity [196]. Consequently, the AAP is only related to the bias factor and the two temperatures of the involved replicas. To cover a temperature gap from T_1 to T_n with n replicas in PT-WTE, and have a desired AAP $\langle P_{\gamma,acc} \rangle$, we will obtain a solvable equation set composed of $n - 1$ equation 6.3 which correspondingly contains $n - 1$ unknown variables including $n - 2$ temperatures from T_2 to T_{n-1} and the bias factor γ . The temperature ladder and the bias factor are generated by solving the equation set.

WTE framework demonstrates its advantages when combined with PT: the number of replicas is adjustable by changing the value of bias factor, in which way considerable amounts of computational resources are saved. Inspired by the works of Rathore et al.[192] and Patriksson et al. [196], we provided an operable scheme to generate reasonable temperature intermediates and a corresponding bias factor for PT-WTE within a given temperature range. By taking different bias factors, the number of replicas could be adjusted in this scheme to keep the desired AAP unchanged. This method was demonstrated on an explicit solvated mini-protein system. Two series of PT-WTE

simulations with different bias factors and number of replicas were carried out. Discussions on the choice of proper bias factors were given.

6.2 Parameters

All simulations were performed with the help of Gromacs package [147] (version 4.6.7) and plumed plug-in [156] (version 2.1.2). An explicitly solvated tryptophan-cage protein model (PDB code 1L2Y [197]) was simulated to compare the performance of PT-WTE under different bias factors and number of replicas. All interactions were described by the AMBER99SB force field [149]. Two series of PT-WTE simulations, including S_{16} , a 16-replica PT-WTE simulation with bias factor $\gamma = 37.9$, and S_8 , an 8-replica PT-WTE simulation with $\gamma = 174.2$ were simulated for 200 ns per replica, so totally we have 4.8 μ s length of trajectories in this study.

The desired average acceptance probabilities and the temperature boundary for both series were set to 0.5 and 300K ~ 500K, respectively. Their detailed temperature sequences are listed in the Results and Discussion section. The simulations were prepared as following before the production sampling: different unfolded structures were generated from an unbiased MD simulation at 500K and were used as the initial configurations for different replicas. Each structure was then solved in a 6*6*6 nm³ water box which contained 6769 TIP3P water molecules [151] and one chlorine ion as the counter ion.

After energy minimization of the simulation box, 500 ps of equilibrium simulation were carried out in the NVT ensemble with position restrains of the protein. We then removed the position restrains and constructed the WTE framework by a 30 ns simulation. During this process, the potential energy was used as CV with biasing potentials of the Gaussian width of 500 kJ/mol and Gaussian height of 2 kJ/mol, and the Gaussian potential was

deposited every 500 calculation steps. It was then followed the 200 ns production run keeping the added biasing potential with no more new biases added to the system. The NVT ensemble was adopted during the whole simulation with the V-rescale temperature coupling method to maintain the desired temperature [154]. A leap-frog algorithm was applied for the integration of Newton equation of motion, and the LINCS algorithm [150] and the SETTLE algorithm [189] were used to constrain all the covalent bonds in protein and water with a time step of 2 fs. The cutoff for both van der Waals interactions and short-range electrostatic interactions were set to 1.0 nm. The long-range electrostatic interactions were handled by the PME [153] algorithm. An exchange was attempted between neighboring replicas every 500 calculation steps.

6.3 Results and discussion

6.3.1 Generation of temperature sequences and bias factors

Short unbiased MD simulations (10 ns each) were performed at different temperatures for the fitting of the potential energy average $\mu(T)$ and the potential energy distributional width $\sigma(T)$. Potential energies at different temperatures follow Gaussian distributions (Fig. 6.1a). Raw data of the potential averages are shown in Fig. 6.1b as black dots. A linear fitting (Fig. 6.1b blue line) was first performed as suggested by Patriksson et al. [196], however, instead of a proportional increase of the potential average to the temperature, we did observe a decay of the slope. This slope has a dimension of heat capacity, indeed it can be considered, by definition, as a reduced heat capacity of the system with the exclusion of the kinetic energy part. In our simulation, this reduced heat capacity decreased by $135.2 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ as the temperature increases from 300K to 500K (Fig. 6.1c).

Figure 6.1 Thermodynamic properties of the system. Potential energy distributions in the canonical ensemble are Gaussian-shaped (figure a). Quadratic fitting (figure b, red curve) is more precise than linear fitting (figure b, blue line) for the potential average (AVE) and temperature relationship. The heat capacity decays almost linearly with the temperature (figure c). The heat capacity is calculated according to $C_V(T) = \sigma(T)^2 / (k_B T^2)$.

It is well known that the heat capacity for proteins is usually not a unique value at different temperatures [198], especially at the phase transition temperature, as shown by Rathore et al. [192]. In our simulation system, we have almost 7000 water molecules. As a result, the water molecules contribute much more to the heat capacity than the protein. Experiments proved the decrease of the heat capacity of water at a constant volume at high temperature [199,200]. According to the calculation of the online water heat capacity calculator [201], the heat capacity decrease of a water system with comparable size at the same temperature range was equal to $109 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$. This decrease, which should come from the potential energy part since the kinetic energy part is supposed to be a constant, was in the same order of magnitude as the decrease of our system, indicating that the major change of heat capacity comes from water molecules.

The expectation of this study is to provide the PT-WTE in a specific temperature range with an optimal temperature ladder, which is supposed to be sparser than that in conventional PT. However, along with the heat capacity decay, a linear fitting of the

potential average would bring errors into the temperature ladder, and unfortunately, the errors eventually overestimate the AAPs in higher temperature region and underestimate the AAPs in lower temperature. Thus the linear fitting doesn't apply to our case and a more precise fitting method is required. Since the isochoric heat capacity of water decreases with temperature in a linear manner [202], so does the heat capacity in our system (Fig. 6.1c). This means the potential average is quadratically related to the temperature. We assume the expression of the potential average as $\mu(T) = A_1 + B_1T + C_1T^2$ and the heat capacity as $C_V(T) = B_1 + 2C_1T$, correspondingly, the potential distributional width is expressed as $\sigma(T) = \sqrt{k_B B_1 T^2 + 2k_B C_1 T^3}$, based on $C_V(T) = \sigma(T)^2 / (k_B T^2)$. The $\mu_{1,2}$ and $\sigma_{1,2}$ in equation 6.3 are then written as:

$$\mu_{12} = B_1(T_1 - T_2) + C_1(T_1^2 - T_2^2) \quad (6.5)$$

and

$$\sigma_{12} = \sqrt{k_B B_1(T_1^2 + T_2^2) + 2k_B C_1(T_1^3 + T_2^3)} \quad (6.6)$$

respectively. The quadratic function provided a good prediction of potential average (Fig. 6.1b, red curve) with an R-squared value of 0.99994, and the three constants A_1 , B_1 , and C_1 , were fitted with values of $-391990.46 \text{ kJ/mol}$, $499.38 \text{ kJ/(mol} \cdot \text{K)}$, and $-0.32 \text{ kJ/(mol} \cdot \text{K}^2)$ respectively. The resultant temperature sequences and the corresponding

Table 6.1 Calculated bias factors and temperature sequences for simulation series S₁₆ and S₈.

Series	Bias factors	Temperature sequences (K)							
S ₁₆	37.9	300.00	309.35	319.10	329.26	339.87	350.96	362.57	374.73
		387.50	400.93	415.09	430.05	445.90	462.74	480.72	500.00
S ₈	174.2	300.00	320.52	342.99	367.71	395.09	425.69	460.27	500.00

bias factors (Table 6.1) could be obtained by solving the equation set composed of equation 6.3, with the substitution of equation 6.5 and equation 6.6, as well as the specified temperature range (300 K ~ 500 K), the number of replicas (16 or 8) and the AAPs (0.5). Theoretically, this algorithm is capable of obtaining temperature intermediates and bias factors for simulations with any number of replicas. In this work, we only use a 16-replica PT-WTE and an 8-replica PT-WTE as demonstrations and compare their performances. A large amount of computational resources is saved as compared to the PT simulation in canonical ensemble, where more than 120 replicas are required for the same temperature range and the same AAP, based on the Temperature Generator [196].

6.3.2 Sampling efficiency of PT-WTE simulations with different bias factors

One of the commonly used indicators to evaluate the efficiency of PT is the RTT, which refers to the time that a replica spends to traverse all the temperatures and return to the start. RTTs in this study are shown in Table 6.2 and their average values are 18.7 ± 21.2 ns and 5.1 ± 1.3 ns for S_{16} and S_8 respectively. Interestingly, the RTT values directly correlate with the number of replicas and decline with the increase of bias factors, indicating a higher PT efficiency happens with a larger bias factor (or fewer replicas). In a recent study, the RTT was found to be around 5 ns in a 10-replica simulation [203]. Our simulation is more efficient than their work even though our frequency of exchange attempt (1 ps^{-1}) is lower than theirs (2.5 ps^{-1}), this is probably because the AAPs in this

Table 6.2 The RTTs for replicas in different simulation series (continuous trajectories).

Series	RTTs (ns)							
S_{16}	11.8	15.4	12.5	11.1	10.5	8.7	15.4	16.7
	20	15.4	9.5	15.4	11.8	13.3	100	11.8
S_8	7.7	4.4	5.7	4.8	3.2	4.2	5.3	5.9

Figure 6.2 Conformational sampling of the two simulation series. FESs of the 300K ensemble are shown in figure (a), and S_8 is not able to reproduce the folded local minimum precisely. The time evolution of the C_α RMSD of each continuous trajectory is shown in figure (b): the S_{16} in the upper panel and S_8 in the lower panel. Each RMSD value is averaged over 100 ps time block with reweighting and is labelled by the colors from blue to red.

study is higher.

The RMSD of C_α atoms with respect to those of the NMR structure is calculated and one conformation is considered fully folded when its RMSD value is smaller than 0.2 nm [203]. The biases were removed based on the work of Tiwary et al. [125] during data processing. FES of the 300K ensemble in S_{16} (Fig. 6.2a) is able to sample the folded and unfolded states of trp-cage with a free energy difference of -1.4 kJ/mol, which is comparable to the experimental value of -2.1 kJ/mol [204]. However, FES in S_8 (Fig. 6.2a) shows quite a poor sampling of the folded states. The reason is shown by the RMSD change of each continuous trajectory in Fig. 6.2b, where we only monitored one of the replicas spent considerable time in the folded state in the S_8 simulation while there were more in the S_{16} simulation. Compared to the S_{16} simulation, the S_8 simulation requires much longer simulation time to converge.

If a folding event is defined by a transition of C_α RMSD from a value larger than 0.8 nm to that smaller than 0.2 nm, we only confirmed 0.94 folding events per replica in the S_{16} simulation, and even less (0.88 per replica) in the S_8 simulation. This also restricted the conformational sampling in our study. To further enhance the conformational sampling from this perspective, a recommended solution is to apply bias on other structure-related CVs, as proved by the above mentioned work [203], where the backbone hydrogen bonds and hydrophobic core contacts were used as CVs in a PT-WTE simulation of the trp-cage system with 10 replicas, the folding events per replica increased to almost 30 times in 250 ns simulation.

6.3.3 A smaller bias factor works better with our algorithm

The AAPs in each simulation series are nearly uniform (Table 6.3), and their averages are 0.439 ± 0.031 (S_{16}) and 0.424 ± 0.072 (S_8) respectively, which are both close to the predicted value of 0.5. These results indicate that our algorithm generates satisfactory temperature ladders for the two cases. Nonetheless, it is still notable that the AAP sequence in the S_8 simulation (bias factor = 174.2) is more dispersive than that in the S_{16} simulation (bias factor = 37.9). Such a behavior for the series under a larger bias factor is a consequence of the abnormal potential distributions. The biased potential distributions of the two simulation series (Fig. 6.3a and 6.3b) exhibit the Gaussian-shaped curves when using the bias factor of 37.9 and the misshaped distribution when the bias factor equals to 174.2.

Table 6.3 The AAPs for neighboring replica-pairs. Lower temperature replica-pairs are on the left.

Series	AAPs of neighboring replica-pairs														
S_{16}	0.43	0.42	0.48	0.40	0.48	0.42	0.46	0.36	0.46	0.45	0.44	0.43	0.44	0.44	0.48
S_8	0.48	0.36	0.56	0.39	0.41	0.33	0.44

Figure 6.3 Distribution of the biased potential energy in S_{16} (a) and S_8 (b). The Gaussian shapes become poorer in S_8 . The Gaussian height decay of 300 K ensembles during the 30 ns WTE-establishing period (c). Only the regions around the potential average are counted. The x-axis is the times that the samples visit the considered region.

In WTE, a history dependent bias [123]

$$V(U, t) = k_B \Delta T \ln \left(1 + \frac{\omega N(U, t)}{k_B \Delta T} \right) \quad (6.7)$$

is added to the Hamiltonian with a rate of

$$\dot{V}(U, t) = \omega e^{-\left(\frac{V(U, t)}{k_B \Delta T}\right)} \delta_{U, U(t)} \quad (6.8)$$

where ω is the initial bias deposition rate, $N(U, t) = \int_0^t \delta_{U, U(t')} dt'$ is the histogram of potential U in the biased simulation, which rises linearly with simulation time. The Gaussian deposition rate at a given point U , $\omega e^{-\left(\frac{V(U, t)}{k_B \Delta T}\right)} \delta_{U, U(t)}$ decreases to zero if the time spent there is long enough, and that's the sign to reach a converged WTE. This is the foundation of the whole WTE framework and ensures the effectiveness of equation 2.23 and equation 6.3. To verify if the WTE is well established, we checked the time evolution of Gaussian height. Practically we used a finite width Gaussian to replace the $\delta_{U, U(t)}$. The Gaussian height decay in a CV range ($\mu - \sigma_G \sim \mu + \sigma_G$, where μ is the potential average and

σ_G is the Gaussian width in the 300K ensemble) is plotted as a function of visiting times (Fig. 6.3c). In 30 ns WTE-establishing time, Gaussian height with bias factor of 37.9 decreased to a negligible value (0.019 kJ/mol), while that with bias factor of 174.2 is still sizeable (0.159 kJ/mol). As a result, the distributions deviate from the ideal Gaussian shapes in S_8 , thus misleading the solution of equation 6.3 and leading to a more disperse AAP sequence. Based on these results, we speculate that the larger bias factor retards the convergence and offers a worse performance of PT-WTE.

6.3.4 Improvement of the performance of large bias factor

The performance of a large bias factor is improvable by extending the WTE-establishing time to achieve the convergence. To obtain a small value of $\omega e^{-\left(\frac{V(U,t)}{k_B\Delta T}\right)} \delta_{U,U(t)}$, the $\frac{V(U,t)}{k_B\Delta T}$ should be large enough, which, based on equation 6.7, means the value of $\frac{\omega N(U,t)}{k_B\Delta T}$ needs to be large enough. Noting that $\Delta T = T(\gamma - 1)$ is increasing with bias factor, however, $N(U, t)$ grows linearly with the simulation time, indicating that for a specific bias factor, the convergence is always achievable by elongating the sampling time. For example, we extended the WTE-establishing time of S_8 simulation from 30 ns to 100 ns and obtained a further reduced Gaussian height (Fig. 6.4a, 0.054 kJ/mol), and correspondingly, the biased potential distribution in the following 20 ns PT-WTE simulation is more Gaussian-shaped (Fig. 6.4b). The AAPs were also improved to a value of 0.454 ± 0.058 with a smaller standard deviation.

The performance of the improved S_8 simulation was still not as good as S_{16} . Here we make a rough evaluation based on the $\frac{\omega N(U,t)}{k_B\Delta T}$ values of two simulation cases. For a fully converged simulation, the biased potential energy is supposed to distribute in a Gaussian shape. We represent the histogram of $N(U, t)$ from equation 6.7 in the region of $\mu -$

Figure 6.4 Improvement of S_8 simulation by extending the WTE-establishing time to 100 ns. The Gaussian height decay of 300K ensemble is further reduced (a), and correspondingly the potential distribution is more Gaussian-shaped (b).

$\sigma_G \sim \mu + \sigma_G$ in a period of time as $N(\mu - \sigma_G \sim \mu + \sigma_G)$, then $N(\mu - \sigma_G \sim \mu + \sigma_G)$ is countable based on:

$$N(\mu - \sigma_G \sim \mu + \sigma_G) = N(-\infty \sim +\infty) \cdot \int_{\mu - \sigma_G}^{\mu + \sigma_G} P(U) dU \quad (6.9)$$

where $N(-\infty \sim +\infty)$ is equal to the total number of Gaussian deposition steps, and

$P(U) = \frac{1}{\sigma \sqrt{2\pi\gamma}} e^{-\frac{(U-\mu)^2}{2\gamma\sigma^2}}$ is the biased potential distribution. The ratio of $\frac{\omega N_{16}(\mu - \sigma_G \sim \mu + \sigma_G)}{k_B \Delta T_{16}}$:

$\frac{\omega N_8(\mu - \sigma_G \sim \mu + \sigma_G)}{k_B \Delta T_8}$ (subscript 16 for S_{16} and 8 for S_8) is simplified to:

$$\frac{N_{16}(-\infty \sim +\infty) \cdot \int_{\mu - \sigma_G}^{\mu + \sigma_G} P_{16}(U) dU}{\gamma_{16} - 1} : \frac{N_8(-\infty \sim +\infty) \cdot \int_{\mu - \sigma_G}^{\mu + \sigma_G} P_8(U) dU}{\gamma_8 - 1} \quad (6.10)$$

In equation 6.10, $P_8(U)$ has a larger width than $P_{16}(U)$, thus fewer samples visit a specific CV region in equal simulation time in S_8 than in S_{16} . This is also illustrated in Fig. 6.3c, where the visiting times for the selected region in S_8 is only half as many as in S_{16} . The expansion of simulation time only enlarges the $N(-\infty \sim +\infty)$ term. If equation 6.10 has a value of 1, we will have the ratio of $N_{16}(-\infty \sim +\infty) : N_8(-\infty \sim +\infty) \approx 1 : 10$, which means that in

our simulation, the S_8 reaches the same convergence level as S_{16} in 300 ns of WTE-establishing time. Clearly, the cost is huge to build up the WTE framework with a large bias factor. Thus, an oversized bias factor in our algorithm is not recommended, in spite of the reduction of the total number of replicas.

6.4 Conclusion

In this chapter, we come up with an algorithm to generate the temperature sequence and the corresponding bias factor for PT-WTE simulations to achieve desired exchange acceptance rates. The number of replicas that are required to cover a specific temperature range is adjustable by taking bias factor into consideration keeping all AAPs between neighboring replica-pairs to a fixed value. We have tested two simulation series with the number of replicas as 16 and 8 and bias factors as 37.9 and 174.2, respectively. Bias factor of 174.2 is proved to be oversized for the application of the algorithm. This algorithm is beneficial for the preparation of PT-WTE or PTMetaD-WTE simulations with greatly reduced computational resource requirement.

7 Summaries and recommendations

Interactions between antibiotics and their targets usually behave multistability, which impede the thorough exploration of the binding phase space for the conventional MD simulation in accessible timescale. The enhanced sampling methods, by annealing the system (PT), or manipulating the Hamiltonian of the system with extra biases (MetaD), accelerate the sampling process in some degree. With the help of the enhanced sampling methods, the binding of two new antimicrobial agents, namely, CSM5-K5 and TXB, to their targets are studied in this thesis. The presented results may contribute to the future drug design.

As a product of the synthetic approaches, CSM5-K5 demonstrates a successful drug design by combining two mechanisms which are positive to the antimicrobial activity. On one hand, the introduction of polylysine side chains provides CSM5-K5 with the motivation of recognizing bacteria over mammalian cell; on the other hand, the choice of chitosan as the molecular backbone counteracts the electrostatic repulsion brought by polylysine through the strong $\text{-NHC(=O)}\cdots\text{-NHC(=O)}$ H-bonds to aggregate. In the other study case, the natural product TXB points out a direction for the future drug-discovery. Our simulations reveal that TXB binds to the negatively charged groups of L_{II} by forming H-bonds with its four NH groups on the ring motif. With a deeper study, we suggest that the chirality of the D-Thr residue on the ring motif plays a critical role in maintaining the binding cavity to L_{II} . Our simulations confirm a working drug design philosophy and a pharmacophore based on the simulation of CSM5-K5 and TXB, providing theory basis for experimentally developing the derivatives of these antibiotics and even new designs. More than that, our studies put forward a feasible solution for investigating the interactions

between biomolecules with high flexibilities: a thoroughly sampling of the binding phase space is always necessary, and this could be realized by the enhanced sampling methods; after the sampling, structures should be clustered based on the contacting information rather than the configurational criterion. MD simulation in combination with enhanced sampling methods exhibits its power in searching configurations of binding complexes, and it has already grown to be a good supplementary of the experimental techniques in drug design.

The enhanced sampling method PT-WTE rendered itself well in the study of TXB. However, no mature method is available in selecting the temperature sequence and bias factor for PT-WTE. We provide a scheme in this thesis to generate the temperature sequence and the corresponding bias factor optimally. In this scheme, the AAP of replica-pairs is customizable, as well as the number of replicas for a complete coverage of a specific temperature range. By comparing the performances of different bias factors, the oversized one is not recommended for the PT-WTE simulation. This scheme, which assists the promotion of PT-WTE simulations from the perspective of optimizing performances and saving computational resources, leads to a better understanding and application of PT-WTE simulations.

The studies in this thesis are sustainable from the following aspects:

- a. Although the pharmacophore of TXB is confirmed as the ring motif, agents that based on this binding site are still synthetically difficult because of the inaccessibility of the unusual enduracididine (End) residue. Several attempts had been made experimentally to replace the End residue with other usual amino acids, however, not all of them were able to maintain the activity. Thus, the principles of this substitution could accelerate the

development of the TXB-based drug design. Such principles are also expandable to other residues on the ring motif to further lower the limitation in selecting residues for the design of the ring motif.

b. In PT-WTE simulation, the instantaneous acceptance probability is described by $e^{\frac{1}{\gamma}(\beta_1 - \beta_2) \cdot (U_1 - U_2)}$, because when WTE is well-established, the biased instantaneous potential $U_\gamma = \frac{1}{\gamma}U$. However, the situation is different with more CVs being biased in PTMetaD-WTE simulations, since there would be new biases kept adding to the Hamiltonian of the system. Usually, the deposited biases of different replicas vary from each other and lead to the differences of the Hamiltonians, so the PTMetaD-WTE simulation is more like a Hamiltonian- REMD. Description of the instantaneous acceptance probability in PTMetaD-WTE simulation could help to deeply understand the dynamic properties of the system.

References

- [1] C.J. Thomson, E. Power, H. Ruebsamen-Waigmann, H. Labischinski, *Curr. Opin. Microbiol.* 7 (2004) 445.
- [2] M. Wainwright, *Top. Catal.* 3 (1989) 21.
- [3] A. Fleming, *Br. J. Exp. Pathol.* 10 (1929) 226.
- [4] L.L. Ling, T. Schneider, A.J. Peoples, A.L. Spoering, I. Engels, B.P. Conlon, A. Mueller, T.F. Schaberle, D.E. Hughes, S. Epstein, M. Jones, L. Lazarides, V.A. Steadman, D.R. Cohen, C.R. Felix, K.A. Fetterman, W.P. Millett, A.G. Nitti, A.M. Zullo, C. Chen, K. Lewis, *Nature* 517 (2015) 455.
- [5] K.J. Simmons, I. Chopra, C.W. Fishwick, *Nat. Rev. Microbiol.* 8 (2010) 501.
- [6] C.M. Kunin, *Ann. Intern. Med.* 118 (1993) 557.
- [7] I.A. Edwards, A.G. Elliott, A.M. Kavanagh, J. Zuegg, M.A.T. Blaskovich, M.A. Cooper, *ACS Infect. Dis.* 2 (2016) 442.
- [8] C. Arias, *N. Engl. J. Med.* 372 (2015) 1168.
- [9] D. Nichols, N. Cahoon, E.M. Trakhtenberg, L. Pham, A. Mehta, A. Belanger, T. Kanigan, K. Lewis, S.S. Epstein, *Appl. Environ. Microbiol.* 76 (2010) 2445.
- [10] L.J. Piddock, *J. Antimicrob. Chemother.* 70 (2015) 2679.
- [11] M.A. Kohanski, D.J. Dwyer, J.J. Collins, *Nat. Rev. Microbiol.* 8 (2010) 423.
- [12] O. Espeli, K.J. Mariani, *Mol. Microbiol.* 52 (2004) 925.
- [13] K. Drlica, M. Malik, R.J. Kerns, X. Zhao, *Antimicrob. Agents Chemother.* 52 (2008) 385.
- [14] C. Chen, M. Malik, M. Snyder, K. Drlica, *J. Mol. Biol.* 258 (1996) 627.
- [15] R.J. Belland, S.G. Morrison, C. Ison, W.M. Huang, *Mol. Microbiol.* 14 (1994) 371.
- [16] R. Munoz, A.G. De La Campa, *Antimicrob. Agents Chemother.* 40 (1996) 2252.
- [17] S.J. Kim, J. Chang, M. Singh, *Biochim. Biophys. Acta (BBA) - Biomembranes* 1848 (2015) 350.
- [18] E.M. Wise, J.T. Park, *Proc. Natl. Acad. Sci. U S A* 54 (1965) 75.
- [19] D.J. Tipper, J.L. Strominger, *Proc. Natl. Acad. Sci. U S A* 54 (1965) 1133.
- [20] D.J. Waxman, R.R. Yocum, J.L. Strominger, *Phil. Trans. R. Soc. Lond. B* 289 (1980) 257.
- [21] U. Patel, Y.P. Yan, F.W. Hobbs, Jr., J. Kaczmarczyk, A.M. Slee, D.L. Pompliano, M.G. Kurilla, E.V. Bobkova, *J. Biol. Chem.* 276 (2001) 37199.
- [22] I. Chopra, M. Roberts, *Microbiol. Mol. Biol. Rev.* 65 (2001) 232.
- [23] B.D. Davis, *Microbiol. Rev.* 51 (1987) 341.
- [24] J. Davies, B.D. Davis, *J. Biol. Chem.* 243 (1968) 3312.
- [25] E. Matyus, C. Kandt, D.P. Tieleman, *Curr. Med. Chem.* 14 (2007) 2789.
- [26] E.P. Abraham, E. Chain, *Rev. Infect. Dis.* 10 (1990) 677.
- [27] M. Gniadkowski, *Clin. Microbiol. Infect.* 14 (2008) 11.
- [28] T. Guina, E.C. Yi, H. Wang, M. Hackett, S.I. Miller, *J. Bacteriol.* 182 (2000) 4077.
- [29] K. Sugimura, T. Nishihara, *J. Bacteriol.* 170 (1988) 5625.
- [30] M.C. Roberts, *FEMS Microbiol. Lett.* 282 (2008) 147.
- [31] M.R. Yeaman, N.Y. Yount, *Pharmacol. Rev.* 55 (2003) 27.
- [32] L. Guo, K.B. Klim, C.M. Poduje, M. Daniel, J.S. Gunn, M. Hackett, S.I. Miller, *Cell* 95 (1998).
- [33] Z. Zhou, S. Lin, R.J. Cotter, C.R.H. Raetz, *J. Biol. Chem.* 274 (1999) 18503.
- [34] R.K. Ernst, T. Guina, S.I. Miller, *J. Infect. Dis.* 179 (1999) S326.
- [35] R.K. Ernst, E.C. Yi, L. Guo, K.B. Lim, J.L. Burns, M. Hackett, S.I. Miller, *Science* 286 (1999) 1561.
- [36] N. Malanovic, K. Lohner, *Biochim. Biophys. Acta* 1858 (2016) 936.
- [37] M. Simanski, R. Glaser, B. Koten, U. Meyer-Hoffert, S. Wanner, C. Weidenmaier, A. Peschel, J. Harder, *Exp. Dermatol.* 22 (2013) 294.
- [38] E.S. Lysenko, J. Gould, R. Bals, J.M. Wilson, J.N. Weiser, *Infect. Immun.* 69 (2000) 4916.

- [39] E. Dorrer, M. Teuber, *Arch. Microbiol.* 114 (1977) 87.
- [40] A. Peschel, R.W. Jack, M. Otto, L.V. Collins, P. Staubitz, G. Nicholson, H. Kalbacher, W.F. Nieuwenhuizen, G. Jung, A. Tarkowski, K.P. van Kessel, J.A. Van Strijp, *J. Exp. Med.* 193 (2001) 1067.
- [41] J. Delaunay, *Blood Rev.* 21 (2007) 1.
- [42] Y. Andoh, S. Okazaki, R. Ueoka, *Biochim. Biophys. Acta* 1828 (2013) 1259.
- [43] C. Hofstab, E. Lindahl, O. Edholm, *Biophys. J.* 84 (2003) 2192.
- [44] S.W. Chiu, E. Jakobsson, R.J. Mashl, H.I. Scott, *Biophys. J.* 83 (2002) 1842.
- [45] G. van Meer, D.R. Voelker, G.W. Feigenson, *Nat. Rev. Mol. Cell Biol.* 9 (2008) 112.
- [46] T. Yuan, X. Zhang, Z. Hu, F. Wang, M. Lei, *Biopolymers* 97 (2012) 998.
- [47] Y. Wang, D.E. Schlamadinger, J.E. Kim, J.A. McCammon, *Biochim. Biophys. Acta* 1818 (2012) 1402.
- [48] F. Mehrnejad, M. Zarei, *J. Biomol. Struct. Dyn.* 27 (2010) 551.
- [49] M. Xiong, M.W. Lee, R.A. Mansbach, Z. Song, Y. Bao, R.M. Peek, Jr., C. Yao, L.F. Chen, A.L. Ferguson, G.C. Wong, J. Cheng, *Proc. Natl. Acad. Sci. U S A* 112 (2015) 13155.
- [50] P. Jurkiewicz, L. Cwiklik, A. Vojtkova, P. Jungwirth, M. Hof, *Biochim. Biophys. Acta* 1818 (2012) 609.
- [51] L.M. A., *Nat. Rev. Mol. Cell Biol.* 9 (2008) 99.
- [52] J. Razzokov, M. Yusupov, S. Vanuytsel, E.C. Neyts, A. Bogaerts, *Plasma Process Polym.* 14 (2017) 1700013.
- [53] P.A. Leventis, S. Grinstein, *Annu. Rev. Biophys.* 39 (2010) 407.
- [54] W. Vollmer, D. Blanot, M.A. de Pedro, *FEMS Microbiol. Rev.* 32 (2008) 149.
- [55] P. Giesbrecht, T. Kersten, H. Maidhoe, J. Wecke, *Microbiol. Mol. Biol. Rev.* 62 (1998) 1371.
- [56] T.J. Silhavy, D. Kahne, S. Walker, *Cold Spring Harb. Perspect. Biol.* 2 (2010) a000414.
- [57] C. Erridge, E. Bennett-Guerrero, I.R. Poxton, *Microb. Infect.* 4 (2002) 837.
- [58] C.R. Raetz, C. Whitfield, *Annu. Rev. Biochem.* 71 (2002) 635.
- [59] A. Bhunia, H. Mohanram, P.N. Domadia, J. Torres, S. Bhattacharjya, *J. Biol. Chem.* 284 (2009) 21991.
- [60] R.M. Epand, R.F. Epand, *J. Pept. Sci.* 17 (2011) 298.
- [61] V. Vivcharuk, Y.N. Kaznessis, *J. Phys. Chem. B* 115 (2011) 14704.
- [62] J. Lee, S.W. Jung, A.E. Cho, *Langmuir* 32 (2016) 1782.
- [63] K. Osapay, D. Tran, A.S. Ladokhin, S.H. White, A.H. Henschen, M.E. Selsted, *J. Biol. Chem.* 275 (2000) 12017.
- [64] C. de la Fuente-Nunez, F. Reffuveille, S.C. Mansour, S.L. Reckseidler-Zenteno, D. Hernandez, G. Brackman, T. Coenye, R.E. Hancock, *Chem. Biol.* 22 (2015) 196.
- [65] E.H. Mattar, H.A. Almeshdar, H.A. Yacoub, V.N. Uversky, E.M. Redwan, *Cytokine Growth Factor Rev.* 28 (2016) 95.
- [66] N. Shagaghi, E.A. Palombo, A.H.A. Clayton, M. Bhave, *World J. Microbiol. Biotechnol.* 34 (2018) 62.
- [67] M. Mahlapuu, J. Hakansson, L. Ringstad, C. Bjorn, *Front. Cell. Infect. Microbiol.* 6 (2016) 194.
- [68] B. Gomes, M.T. Augusto, M.R. Felicio, A. Hollmann, O.L. Franco, S. Goncalves, N.C. Santos, *Biotechnol. Adv.* 36 (2018) 415.
- [69] K.A. Brogden, *Nat. Rev. Microbiol.* 3 (2005) 238.
- [70] N.Y. Yount, A.S. Bayer, Y.Q. Xiong, M.R. Yeaman, *Biopolymers* 84 (2006) 435.
- [71] L.T. Nguyen, E.F. Haney, H.J. Vogel, *Trends Biotechnol.* 29 (2011) 464.
- [72] R.F. de Almeida, L.M. Loura, M. Prieto, *Chem. Phys. Lipids* 157 (2009) 61.
- [73] F. Kawai, M. Shoda, R. Harashima, Y. Sadaie, H. Hara, K. Matsumoto, *J. Bacteriol.* 186 (2004) 1475.
- [74] E. Mileykovskaya, W. Dowhan, *J. Bacteriol.* 182 (2000) 1172.

- [75] R.F. Epand, M. Tokarska-Schlattner, U. Schlattner, T. Wallimann, R.M. Epand, *J. Mol. Biol.* 365 (2007) 968.
- [76] R.M. Epand, R.F. Epand, *Mol. Biosyst.* 5 (2009) 580.
- [77] J. Oreopoulos, R.F. Epand, R.M. Epand, C.M. Yip, *Biophys. J.* 98 (2010) 815.
- [78] M.H. McCormick, J.M. McGuire, G.E. Pittenger, R.C. Pittenger, W.M. Stark, *Antibiot. Annu.* 3 (1956) 606.
- [79] E. Breukink, B. de Kruijff, *Nat. Rev. Drug Discov.* 5 (2006) 321.
- [80] B. de Kruijff, V. van Dam, E. Breukink, *Prostaglandins Leukot. Essent. Fatty Acids* 79 (2008) 117.
- [81] M.A. McCloskey, F.A. Troy, *Biochemistry* 19 (1980) 2061.
- [82] K. Schirner, Y.J. Eun, M. Dion, Y. Luo, J.D. Helmann, E.C. Garner, S. Walker, *Nat. Chem. Biol.* 11 (2015) 38.
- [83] H. Molinari, A. Pastore, L. Lian, G.E. Hawkes, K. Sales, *Biochemistry* 29 (1990) 2271.
- [84] B. Xing, T. Jiang, X. Wu, R. Liew, J. Zhou, D. Zhang, E.K. Yeow, *Chemistry* 17 (2011) 14170.
- [85] P.J. Loll, A. Derhovanessian, M.V. Shapovalov, J. Kaplan, L. Yang, P.H. Axelsen, *J. Mol. Biol.* 385 (2009) 200.
- [86] Z. Jia, M.L. O'Mara, J. Zuegg, M.A. Cooper, A.E. Mark, *FEBS J.* 280 (2013) 1294.
- [87] S.T. Hsu, E. Breukink, E. Tischenko, M.A. Lutters, B. de Kruijff, R. Kaptein, A.M. Bonvin, N.A. van Nuland, *Nat. Struct. Mol. Biol.* 11 (2004) 963.
- [88] H.E. Hasper, N.E. Kramer, J.L. Smith, J.D. Hillman, C. Zachariah, O.P. Kuipers, B. de Kruijff, E. Breukink, *Science* 313 (2006) 1636.
- [89] A. Zorzi, K. Deyle, C. Heinis, *Curr. Opin. Chem. Biol.* 38 (2017) 24.
- [90] K. Hosoda, M. Ohya, T. Kohno, T. Maeda, S. Endo, K. Wakamatsu, *J. Biochem.* 119 (1996) 226.
- [91] W.R. Scott, S.B. Baek, D. Jung, R.E. Hancock, S.K. Straus, *Biochim. Biophys. Acta* 1768 (2007) 3116.
- [92] D. Jung, A. Rozek, M. Okon, R.E. Hancock, *Chem. Biol.* 11 (2004) 949.
- [93] R.M. Epand, C. Walker, R.F. Epand, N.A. Magarvey, *Biochim. Biophys. Acta* 1858 (2016) 980.
- [94] N. Bionda, J.P. Pitteloud, P. Cudic, *Future Med. Chem.* 5 (2013) 1311.
- [95] K. Lienkamp, A.E. Madkour, A. Musante, C.F. Nelson, K. Nusslein, G.N. Tew, *J. Am. Chem. Soc.* 130 (2008) 9836.
- [96] I. Sovadinova, E.F. Palermo, M. Urban, P. Mpiga, G.A. Caputo, K. Kuroda, *Polymers* 3 (2011) 1512.
- [97] P. Maturana, M. Martinez, M.E. Noguera, N.C. Santos, E.A. Disalvo, L. Semorile, P.C. Maffia, A. Hollmann, *Colloids Surf. B Biointerfaces* 153 (2017) 152.
- [98] C. Ergene, K. Yasuhara, E.F. Palermo, *Polym. Chem.* 9 (2018) 2407.
- [99] L. Migliolo, M.R. Felicio, M.H. Cardoso, O.N. Silva, M.A. Xavier, D.O. Nolasco, A.S. de Oliveira, I. Roca-Subira, J. Vila Estape, L.D. Teixeira, S.M. Freitas, A.J. Otero-Gonzalez, S. Goncalves, N.C. Santos, O.L. Franco, *Biochim. Biophys. Acta* 1858 (2016) 1488.
- [100] L. Chen, D. Walker, B. Sun, Y. Hu, S. Walker, D. Kahne, *Proc. Natl. Acad. Sci. U S A* 100 (2003) 5658.
- [101] J. Xie, J.G. Pierce, R.C. James, A. Okano, D.L. Boger, *J. Am. Chem. Soc.* 133 (2011) 13946.
- [102] M. Karplus, J.A. McCammon, *Nat. Struct. Biol.* 9 (2002) 646.
- [103] J.A. McCammon, B.R. Gelin, M. Karplus, *Nature* 267 (1977) 585.
- [104] J.W. Ponder, D.A. Case, *Adv. Prot. Chem.* 66 (2003) 27.
- [105] A.D. MacKerell Jr., D. Bashford, M. Bellott, J.R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *J. Phys. Chem. B* 102 (1998) 3586.
- [106] C. Oostenbrink, A. Villa, A.E. Mark, W.F. vanGunsteren, *J. Comput. Chem.* 25 (2004) 1656.
- [107] Y. Wang, T. Zhao, D. Wei, E. Strandberg, A.S. Ulrich, J.P. Ulmschneider, *Biochim. Biophys. Acta* 1838 (2014) 2280.

- [108] A. Kia, E. Darve, *Mol. Simul.* 39 (2013) 922.
- [109] P.D. J., B.C. L., *J. Comput. Chem.* 23 (2002) 1045.
- [110] M.E. Tuckerman, G.J. Martyna, *J. Phys. Chem. B* 104 (2000) 159.
- [111] C. Abrams, G. Bussi, *Entropy* 16 (2013) 163.
- [112] G. Bussi, F.L. Gervasio, A. Laio, M. Parrinello, *J. Am. Chem. Soc.* 128 (2006) 13435.
- [113] R. Zhou, *Proc. Natl. Acad. Sci. U S A* 100 (2003) 13280.
- [114] M. Deighan, J. Pfaendtner, *Langmuir* 29 (2013) 7999.
- [115] A. Barducci, M. Bonomi, M.K. Prakash, M. Parrinello, *Proc. Natl. Acad. Sci. U S A* 110 (2013) E4708.
- [116] H. Kokubo, Y. Okamoto, *Chem. Phys. Lett.* 392 (2004) 168.
- [117] A. Laio, F.L. Gervasio, *Rep. Prog. Phys.* 71 (2008) 126601.
- [118] A. Barducci, M. Bonomi, M. Parrinello, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1 (2011) 826.
- [119] D. Bochicchio, E. Panizon, R. Ferrando, L. Monticelli, G. Rossi, *J. Chem. Phys.* 143 (2015) 144108.
- [120] D. Luo, Y. Mu, *J. Phys. Chem. B* 119 (2015) 4955.
- [121] V. Babin, C. Roland, *J. Chem. Phys.* 125 (2006) 204909.
- [122] V. Leone, G. Lattanzi, C. Molteni, P. Carloni, *PLoS Comput. Biol.* 5 (2009) e1000309.
- [123] A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* 100 (2008) 020603.
- [124] M. Bonomi, A. Barducci, M. Parrinello, *J. Comput. Chem.* 30 (2009) 1615.
- [125] P. Tiwary, M. Parrinello, *J. Phys. Chem. B* 119 (2015) 736.
- [126] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* 314 (1999) 141.
- [127] M. Bonomi, M. Parrinello, *Phys. Rev. Lett.* 104 (2010) 190601.
- [128] A. Amadei, M.E.F. Apol, A. Di Nola, H.J.C. Berendsen, *J. Chem. Phys.* 104 (1996) 1560.
- [129] R. Kumari, R. Kumar, A. Lynn, *J. Chem. Inf. Model.* 54 (2014) 1951.
- [130] N. Homeyer, H. Gohlke, *Mol. Inform.* 31 (2012) 114.
- [131] S. Genheden, U. Ryde, *Expert. Opin. Drug Discov.* 10 (2015) 449.
- [132] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, *Acc. Chem. Res.* 33 (2000) 889.
- [133] S. Zhang, S. Guo, Z. Chen, Y. Wang, H. Gao, J. Gomez-Herrero, P. Ares, F. Zamora, Z. Zhu, H. Zeng, *Chem. Soc. Rev.* 47 (2018) 982.
- [134] N. Spackova, T.E. Cheatham, F. Ryjacek, F. Lankas, L. Meervelt, P. Hobza, J. Sponer, *J. Am. Chem. Soc.* 125 (2003) 1759.
- [135] T. Yang, J.C. Wu, C. Yan, Y. Wang, R. Luo, M.B. Gonzales, K.N. Dalby, P. Ren, *Proteins* 79 (2011) 1940.
- [136] S.K. Samal, M. Dash, S. Van Vlierberghe, D.L. Kaplan, E. Chiellini, C. van Blitterswijk, L. Moroni, P. Dubruel, *Chem. Soc. Rev.* 41 (2012) 7147.
- [137] M. Rinaudo, G. Pavlov, J. Desbrieres, *Polymer* 40 (1999) 7029.
- [138] H. Liu, Y. Du, X. Wang, L. Sun, *Int. J. Food Microbiol.* 95 (2004) 147.
- [139] E.I. Rabea, M.E.T. Badawy, C.C. Stevens, G. Smagghe, W. Steurbaut, *Biomacromolecules* 4 (2003) 1457.
- [140] Y.J. Jeon, P.J. Park, S.K. Kim, *Carbohydr. polym.* 44 (2001) 71.
- [141] Y. Omura, M. Shigemoto, T. Akiyama, H. Saimoto, Y. Shigemasa, I. Nakamura, T. Tsuchido, *Biocontrol Sci.* 8 (2003) 25.
- [142] G.B. Ralston, M.V. Tracey, P.M. Wrench, *Biochim. Biophys. Acta* 93 (1964) 652.
- [143] C. Qin, H. Li, Q. Xiao, Y. Liu, J. Zhu, Y. Du, *Carbohydr. Polym.* 63 (2006) 367.
- [144] P. Li, C. Zhou, S. Rayatpisheh, K. Ye, Y.F. Poon, P.T. Hammond, H. Duan, M.B. Chan-Park, *Adv. Mater.* 24 (2012) 4130.

- [145] Z. Hou, Y.V. Shankar, Y. Liu, F. Ding, J.L. Subramanion, V. Ravikumar, R. Zamudio-Vazquez, D. Keogh, H. Lim, M.Y.F. Tay, S. Bhattacharjya, S.A. Rice, J. Shi, H. Duan, X.W. Liu, Y. Mu, N.S. Tan, K.C. Tam, K. Pethe, M.B. Chan-Park, *ACS Appl. Mater. Interfaces* 9 (2017) 38288.
- [146] S.R. Deka, A.K. Sharma, P. Kumar, *Curr. Top. Med. Chem.* 15 (2015) 1179.
- [147] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, *Comp. Phys. Comm.* 95 (1995) 43.
- [148] K.N. Kirschner, A.B. Yongye, S.M. Tschampel, J. Gonzalez-Outeirino, C.R. Daniels, B.L. Foley, R.J. Woods, *J. Comput. Chem.* 29 (2008) 622.
- [149] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* 65 (2006) 712.
- [150] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, *J. Comput. Chem.* 18 (1997) 1463.
- [151] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, *J. Chem. Phys.* 79 (1983) 926.
- [152] I.S. Joung, T.E. Cheatham III, *J. Phys. Chem. B* 112 (2008) 9020.
- [153] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, *J. Chem. Phys.* 103 (1995) 8577.
- [154] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* 126 (2007) 014101.
- [155] M. Parrinello, A. Rahman, *Phys. Rev. Lett.* 45 (1980) 1196.
- [156] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia, *Comp. Phys. Comm.* 180 (2009) 1961.
- [157] J.P.M. Jambeck, A.P. Lyubartsev, *J. Chem. Theory Comput.* 9 (2012) 774.
- [158] J.P.M. Jambeck, A.P. Lyubartsev, *J. Chem. Theory Comput.* 8 (2012) 2938.
- [159] J.P.M. Jambeck, A.P. Lyubartsev, *J. Phys. Chem. B* 116 (2012) 3164.
- [160] N. Kucerka, M.P. Nieh, J. Katsaras, *Biochim. Biophys. Acta* 1808 (2011) 2761.
- [161] J. Pan, F.A. Heberle, S. Tristram-Nagle, M. Szymanski, M. Koepfinger, J. Katsaras, N. Kucerka, *Biochim. Biophys. Acta* 1818 (2012) 2135.
- [162] X. Daura, K. Gademann, B. Jaun, D. Seebach, W.F. Gunsteren, A.E. Mark, *Angew. Chem. Int. Ed.* 38 (1999) 236.
- [163] J.J. Grigsby, H.W. Blanch, J.M. Prausnitz, *Biophys. Chem.* 99 (2002) 107.
- [164] M.H. Struszczyk, *Polimery* 47 (2002) 316.
- [165] V.L. Deringer, U. Englert, R. Dronskowski, *Biomacromolecules* 17 (2016) 996.
- [166] A.A. Yaroslavov, O.Y. Kuchenkova, I.B. Okuneva, N.S. Melik-Nubarov, N.O. Kozlova, V.I. Lobyshev, F.M. Menger, V.A. Kabanov, *Biochim. Biophys. Acta Biomembr.* 1611 (2003) 44.
- [167] T.A. Spurlin, A.A. Gewirth, *Biophys. J.* 91 (2006) 2919.
- [168] R.F. Epand, M.A. Schmitt, S.H. Gellman, R.M. Epand, *Biochim. Biophys. Acta* 1758 (2006) 1343.
- [169] R.F. Epand, B.P. Mowery, S.E. Lee, S.S. Stahl, R.I. Lehrer, S.H. Gellman, R.M. Epand, *J. Mol. Biol.* 379 (2008) 38.
- [170] F. Jean-Francois, S. Castano, B. Desbat, B. Odaert, M. Roux, M.H. Metz-Boutigue, E.J. Dufourc, *Biochemistry* 47 (2008) 6394.
- [171] R.M. Epand, R.F. Epand, *Biochim. Biophys. Acta* 1788 (2009) 289.
- [172] K. Jin, I.H. Sam, K.H. Po, D. Lin, E.H. Ghazvini Zadeh, S. Chen, Y. Yuan, X. Li, *Nat. Commun.* 7 (2016) 12394.
- [173] A.M. Giltrap, L.J. Dowman, G. Nagalingam, J.L. Ochoa, R.G. Linington, W.J. Britton, R.J. Payne, *Org. Lett.* 18 (2016) 2788.
- [174] S. Dhara, V.B. Gunjal, K.L. Handore, D. Srinivasa Reddy, *Eur. J. Org. Chem.* 2016 (2016) 4289.
- [175] Y.E. Jad, G.A. Acosta, T. Naicker, M. Ramtahal, A. El-Faham, T. Govender, H.G. Kruger, B.G. de la Torre, F. Albericio, *Org. Lett.* 17 (2015) 6182.
- [176] A. Parmar, S.H. Prior, A. Iyer, C.S. Vincent, D. Van Lysebetten, E. Breukink, A. Madder, E.J. Taylor, I. Singh, *Chem. Commun.* 53 (2017) 2016.

- [177] A. Parmar, A. Iyer, C.S. Vincent, D. Van Lysebetten, S.H. Prior, A. Maddar, E.J. Taylor, I. Singh, *Chem. Commun.* 52 (2016) 6060.
- [178] S.A.H. Abdel Monaim, Y.E. Jad, G.A. Acosta, T. Naicker, E.J. Ramchuran, A. El-Faham, T. Govender, H.G. Kruger, B.G. de la Torre, F. Albericio, *RSC Adv.* 6 (2016) 73827.
- [179] C. Wu, Z. Pan, G. Yao, W. Wang, L. Fang, W. Su, *RSC Adv.* 7 (2017) 1923.
- [180] H. Yang, K.H. Chen, J.S. Nowick, *ACS Chem. Biol.* 11 (2016) 1823.
- [181] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. CKollman, *J. Comput. Chem.* 25 (2004) 1157.
- [182] M. Rappolt, A. Hickel, F. Bringezu, K. Lohner, *Biophys. J.* 84 (2003) 3111.
- [183] S. Genheden, U. Ryde, *Proteins* 80 (2012) 1326.
- [184] H. Yang, D.R. Du Bois, J.W. Ziller, J.S. Nowick, *Chem. Commun.* 53 (2017) 2772.
- [185] D.H. Williams, M.P. Williamson, D.W. Butcher, S.J. Hammond, *J. Am. Chem. Soc.* 105 (1983) 1332.
- [186] S.A. Cochrane, B. Findlay, A. Bakhtiary, J.Z. Acedo, E.M. Rodriguez-Lopez, P. Mercier, J.C. Vederas, *Proc. Natl. Acad. Sci. U S A* 113 (2016) 11561.
- [187] S.T. Hsu, E. Breukink, G. Bierbaum, H.G. Sahl, B. de Kruijff, R. Kaptein, N.A. van Nuland, A.M. Bonvin, *J. Biol. Chem.* 278 (2003) 13110.
- [188] Y. Hu, J.S. Helm, L. Chen, X.Y. Ye, S. Walker, *J. Am. Chem. Soc.* 125 (2003) 8736.
- [189] S. Miyamoto, P.A. Kollman, *J. Comput. Chem.* 13 (1992) 952.
- [190] C.L. Towse, G. Hopping, I. Vulovic, V. Daggett, *Protein Eng. Des. Sel.* 27 (2014) 447.
- [191] A. Parmar, A. Iyer, S.H. Prior, D.G. Lloyd, E.T. Leng Goh, C.S. Vincent, T. Palmai-Pallag, C.Z. Bachrati, E. Breukink, A. Maddar, R. Lakshminarayanan, E.J. Taylor, I. Singh, *Chem. Sci.* 8 (2017) 8183.
- [192] N. Rathore, M. Chopra, J.J. de Pablo, *J. Chem. Phys.* 122 (2005) 024111.
- [193] D.A. Kofke, *J. Chem. Phys.* 120 (2004) 10852.
- [194] D.A. Kofke, *J. Chem. Phys.* 121 (2004) 1167.
- [195] D.A. Kofke, *J. Chem. Phys.* 117 (2002) 6911.
- [196] A. Patriksson, D. van der Spoel, *Phys. Chem. Chem. Phys.* 10 (2008) 2073.
- [197] J.W. Neidigh, R.M. Fesinmeyer, N.H. Andersen, *Nat. Struct. Biol.* 9 (2002) 425.
- [198] M. Hackel, H.J. Hinz, G.R. Hedwig, *J. Mol. Biol.* 291 (1999) 197.
- [199] T. Kuroki, N. Kagawa, H. Endo, S. Tsuruno, J.W. Magee, *J. Chem. Eng. Data* 46 (2001) 1101.
- [200] I.M. Abdulagatov, V.I. Dvoryanchikov, A.N. Kamalov, *J. Chem. Eng. Data* 43 (1998) 830.
- [201] E. ToolBox, Water-Heat Capacity (Specific Heat), https://www.engineeringtoolbox.com/specific-heat-capacity-water-d_660.html, 2004.
- [202] J. Pátek, J. Hrubý, J. Klomfar, M. Součková, A.H. Harvey, *J. Phys. Chem. Ref. Data* 38 (2009) 21.
- [203] M. Deighan, M. Bonomi, J. Pfendtner, *J. Chem. Theory. Comput.* 8 (2012) 2189.
- [204] L. Qin, S.A. Pabit, A.E. Roitberg, S.J. Hagen, *J. Am. Chem. Soc.* 124 (2002) 12952.

Author's publications

1. **Yang Liu**, Weifeng Li, and Yuguang Mu, Optimization of replica exchange temperature ladder under the well-tempered ensemble, *Chem. Phys. Lett.*, 2018. 711: p.66-72
2. **Yang Liu**, Yaxin Liu, Mary B. Chan-Park, and Yuguang Mu, Binding Modes of Teixobactin to Lipid II: Molecular Dynamics Study, *Sci. Rep.*, 2017. **7**: p. 17197
3. Zheng Hou*, Yogesh Vikhe Shankar*, **Yang Liu***, Feiqing Ding, Jothy Lachumy Subramanion, et. al., Nanoparticles of Short Cationic Peptidopolysaccharide Self-Assembled by Hydrogen Bonding with Antibacterial Effect against Multidrug-Resistant Bacteria, *ACS Appl. Mater. Interfaces*, 2017. 9: p. 38288-38303