November 2019

Molecular Dynamics Simulations of Interaction of DNA Nucleotides and Lignin Oligomers With Small Molecules and Interfaces

Xinjie Tong

MOLECULAR DYNAMICS SIMULATIONS OF INTERACTION OF DNA NUCLEOTIDES AND LIGNIN OLIGOMERS WITH SMALL MOLECULES AND INTERFACES

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Mechanical Engineering

by
Xinjie Tong
B.S., Wuhan Institute of Technology, 2010
M.S., Illinois Institute of Technology, 2014
December 2019

To my wonderful life.

ACKNOWLEDGEMENTS

I would first like to express my greatest gratitude to my advisor Dr. Dorel Moldovan, who provided me with all the opportunities and challenges which are essential to stimulate my personal growth. His guidance lightened my way to the completion of my PhD projects. His generosity and optimism always helped me to cope with difficulties in life. I would like to thank Dr. Brian Novak, who dedicated a lot of time on supervising my projects and provided me with numerous insightful suggestions.

I would like to thank Dr. Stephen E. Rankin, Dr. Mahsa Moradipour, Dr. Pranjali, D. Muley, and Dr. Dorin Boldor for their support, patience, and invaluable advice they gave me during my work on our collaborative projects. I would like to thank Dr. Cristina M. Sabliov, who comforted me and encouraged me to be brave when I was feeling down and overwhelmed. I would also like to thank my committee members Dr. Shengming Guo, Dr. Revati Kumar, and Dr. Manas Gartia for their contribution and kind support.

I would like to thank my family and friends. My parents Zhenjiang Tong and Renzhi Chen, my brother Xinbo Tong, my uncles Dejing Chen, Defeng Chen, my aunts Jenny Wu and Lixia Chen, have always been supportive and encouraging. My sisters by heart, Lixia Huang, Rui Miller, Yiyuan Chen, and Yugui Chen, have always been there for me regardless of the distance. My friends Bin Zhang, Tairan Liu, Meng Li, and Guoling Xiang, have helped and spent exciting time with me. My cat Oreo, my fish Monster and my cactus, have been so cute around.

I would like to acknowledge the financial support from the National Science Foundation under EPSCoR RII Track-2 Program (Project OIA-1632854). The computer resources were provided by Louisiana Optical Network Infrastructure (LONI) and High Performance Computing (HPC) at LSU.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Molecular dynamics (MD) simulations of interaction of DNA nucleotides with self-assembled monolayers (SAMs) provide valuable information that is critical to the development of a new DNA sequencing technique. We investigated the interactions and transport characteristics of mononucleotides moving through nanoslits with SAMs-covered surfaces. Our simulations focused on nanoslits in which the walls were composed of three different types of SAMs: methylformyl terminated, methyl terminated, and phenoxy terminated. The results demonstrated that the phenoxy terminated surfaces have the shortest required nanoslits length for nucleotides separation.

Using MD simulations, we also investigated the interaction of mono-lignin and oligo-lignols with lipid bilayers and small molecules. The simulations showed mono and dimer lignols have different affinity for the lipid bilayers interior and as such, depending on their molecular structure, they can penetrate deep into the bilayers or absorb at bilayer-water interface. When the lignols interaction with bilayers is strong they can affect various physical characteristics of the bilayers, including bilayers gel to fluid phase transition temperature. By focusing on three lignin dimers with different structure and hydrophobicity, our MD simulations provided atomistic insight into how the lignols affect the gel to fluid phase transition temperature in DPPC. Moreover, the magnitude of the values of phase transition temperature shift from MD simulations and differential scanning calorimetry experiments are in good agreement. The MD simulations of interaction of lignin dimer and its derivative with cyclodextrins show the formation of lignin-cyclodextrin association whose stability depend on the lignin dimer structure. The potential of mean force (free energy) for cyclodextrin-lignin dimer interaction was evaluated and used to investigate various characteristics, adsorption mechanism, and binding energy between lignin dimers and cyclodextrin. The characteristic properties include: molecular positioning during association, hydrogen bonding formation, distribution of dihedral angles, adsorption/desorption frequency, lipid tails order

parameter, etc. The MD simulations of lignin tetramers in deep eutectic solvents in the presence of a microwave field show that in the externally applied time varying electric field cause stretching of certain interatomic bonds in lignin tetramers which in turn may increase the probability of bond breaking and therefore an increased rate of breakdown of lignin tetramers.

CHAPTER 1. INTRODUCTION

1.1. Background and motivations

In the long history of human beings exploring patterns of material, numerous theories were established to describe the behavior of matters. Classical mechanics, quantum mechanics, statistical mechanics, and thermodynamics, etc. have been developed into essential tools and continue to be active in specialized research topics. Dynamical study of biomolecules such as DNA/RNA, sugars, proteins and so on, is crucial to biology phenomenon including mutation, transport, and interaction with other molecules. Static and dynamic structure investigations via NMR, X-ray crystallography or homology modeling are widely used in experimental study which offers macroscopic insights, whereas atomic level information is rarely accessible. Therefore, it is challenging to interpret and provide molecular understanding of the physical processes base on experimental data alone. Under certain circumstances, it's also too expensive or not feasible to set up a real-life system to test and develop hypotheses. Atomistic level study of complex system was believed to be impossible back to decades ago. It was thought as a difficult task and even not worth the time when the first folded protein simulation was performed [1]. However, as the development of simulation algorithms, methodologies, potential functions, and computer power capabilities, molecular dynamics (MD) simulation techniques integrate experiment proven to be robust in microscopic analysis. MD also gradually received more respect and recognition in the contribution to scientific community. Nobel Prize in chemistry in 2013 was awarded to Martin Karplus, Micheal Levitt, and Arieh Warshel for "the development of multiscale models for complex chemical systems" [2].

MD simulation has attracted more and more attention, and related published papers were growing exponentially in the past twenty years [3]. It's recently widely used on the bimolecular

system in study of protein mechanism, drug design, toxicology, manifesting its significant importance to life science [4-6]. It produces high resolution of space and time since it can track each atom coordinate over time. Naturally, sequencing projects are spreading since human genome projects stated in 1990 which argues MD simulation an increased demand for structure prediction and application [7, 8]. Nanopore-based technology developed in recent year offers compelling advantages which are adopted by Oxford Nanopore Technologies Limited [9]. To investigate nucleotides with surfaces via MD simulation will provide complimentary details on how nucleotides interact with various material surfaces.

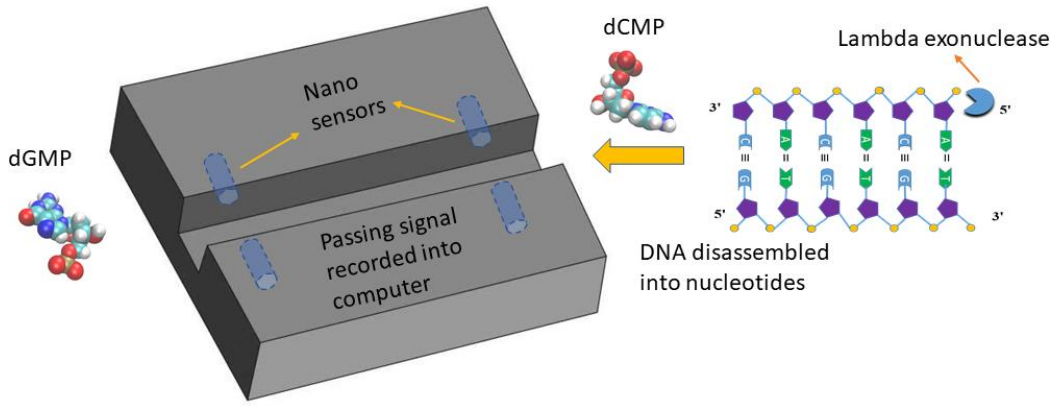


Figure 1.1. DNA sequencing based on time of flight

Another topic in this paper of lignin oligomers was motivated by the world rising concept for sustainability development, i.e., utilizing renewable resources. The trend of using lignin-rich byproduct from lignocellulosic biomass in biorefinery is expected to bring environmental and economic benefits. Lignin is one of the most abundant biopolymers in nature and is found mostly in cell walls of plants. It is a polymer composed of the aromatic monomers (so-called monolignols) p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol that bind to each other via radical coupling and form a complex, heterogeneous structure. In plants, lignin plays important roles in structuring cell walls and preventing water penetration through the xylem. It has also been

documented to have important roles in plant and tree defense against microbial attack [10]. Because it is naturally biodegradable, biocompatible, and has good stability and low toxicity, lignin derivatives are under investigation for multiple pharmacological and biomedical applications [11]. Lignin acts as additive, binders, antioxidants biological carrier, and so on has been applied in a variety field such as healthcare, cosmetic or food industry [12, 13]. To further understand the mechanism of interactions of lignin with other molecules, lignin depolymerization, MD simulation helps to boost the development for potential lignin applications.

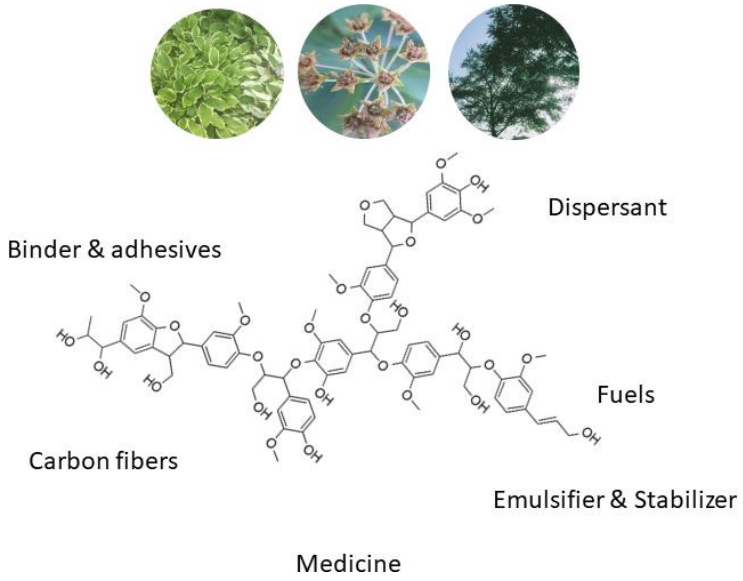


Figure 1.2. Schematic representation of lignin and its applications

The last topic focuses on MD simulation of lignin dimers and derivative binding into cyclodextrins (CDs). CDs are cyclic oligosaccharides commonly consist of 6,7 or 8 glucopyranose units which are named α-, β- and γ-cyclodextrin. Cyclodextrins have a diameter ranging from 5 to 8 Å, which allows most organic molecules to form inclusion complexes [14]. Cyclodextrins have been utilized in various fields due to their unique cyclic structure contains a hydrophobic cavity and hydrophilic exterior. The advantages of CDs are mainly as follows: to improve the solubility and stability of drugs [15]; to enhance biocompatibility [16]; to reduce undesired tastes

and odors [17]; to remove pollutants in water treatment; to use in chromatography and chemical separations [18]. In recent years, an increasing number of MD simulations have been performed to understand the process involved in the formation of CD inclusion complexes and to interpret experimental phenomenon on molecular level [19, 20].

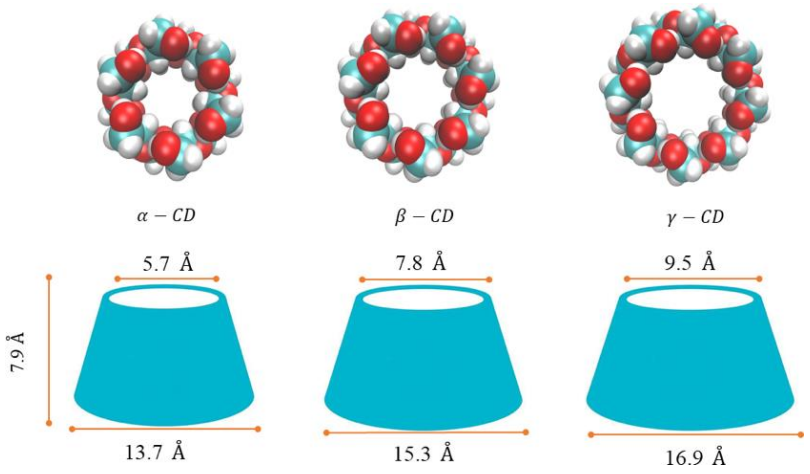


Figure 1.3. Three natural cyclodextrins and their dimensions

1.2. Goal and objectives

The study of DNA nucleotides with self-assembly monolayers (SAMs) aims to propose a more efficient method for lower cost of sequencing. MD simulation observes the DNA nucleotides physical adsorption and desorption behavior with surfaces. SAMs surfaces are friendly as its stability and high controllability. The surface properties are quite flexible as the change of its termination group. By varied alternating surfaces, we can explore shorter nucleotides separation time and optimize possible candidate as a nanopore agent. The study of lignin monomers and dimers with lipid bilayers is to understand the interaction behavior. The trajectory analysis and free energy study of lignin monomers and dimers show how likely they will affect lipid bilayers. Density profile, thickness compressibility, and other analysis from MD simulation results can quantitatively compare with experiment data. Transition temperature study of lignin dimers

supports experimental observations which help to explain what happens behind the temperature shift. Lignin tetramers under microwave observe bond stretching, which gives an idea of possible breakage thus provide insights in depolymerization in the environment of deep eutectic solvents. The study of lignin dimers with beta-cyclodextrins will include the free energy calculations, complex formation, two phases simulation and related analysis. It aims to provide molecular details for the potential application in drug discovery and the separation of small lignin oligomers from a heterogeneous solution.

1.3. Guide to the chapters

The dissertation is organized into seven sections. Chapter 1 briefly addressed the current problem, motivation, goal, and objective of this study, as well as an outline of the whole paper. Chapter 2 presents MD background and fundamental principles. It will cover details of MD knowledge which are mainly used throughout the work. Chapter 3 describes the evolution of DNA sequencing first. Then a goal is defined, and details of nucleotides with self-assembly monolayers (SAMs) are introduced. It's followed by details of MD simulation system setup, methodology, result and discussions. Chapter 4 firstly introduce the importance of studying lignin molecules and lipid bilayers. Force field preparation, simulation details and methodology are followed. Result and discussions mainly contain two parts. One is about monolignols and lignin dimers positioning and free energy with lipid bilayers. The other part talks explicitly about how different concentrations of lignin dimers affect DPPC lipid bilayer crystalline structure. Chapter 5 firstly introduce lignin depolymerization and current study from the experiment. Then the methodology and system set up information are provided. Result and discussion mainly include the total energy verse time, choline N-O vector, dipole moment of pure DES, and bond length analysis of lignin tetramer with DES under microwave field. Chapter 6 firstly provided methods and details about

force field parametrization of β -CD and lignin dimers. Then introduced system setup and information on biased and unbiased MD simulation, and MD dockings. Results and discussion are about free energy and thermodynamic property calculations involve the formation of inclusions of dimer into of β -CD, low concentrations of dimers and two phases simulation. Chapter 7 concludes and summarizes the whole work and propose the current issue which may be resolved in the future.

CHAPTER 2. MOLECULAR DYNAMICS SIMULATION METHODOLOGY

Molecular dynamics simulation has been playing an import role in the field of nanotechnology, chemical industry, and life science. Computational models are built up to mimic the real experiment where you can adjust variables and observe the outcomes easily. It saves time and money due to its efficiency and flexibility. It provides insights on a microscopic scale including atoms coordinates recording, conformational change, and velocities. Macroscopic thermodynamics properties such as internal energy, pressure, volume, and heat capacity, could be converted via statistical mechanics, which is an average of microscopic properties. General ensembles used to represents system states are canonical ensemble (NVT,constant number of atoms, constant volume and constant temperature), isobaric isothermal ensemble (NPT, constant number of atoms, constant pressure, and constant temperature), microcanonical ensemble (NVE, constant number of atoms, constant volume and constant energy) and grand canonical ensemble (μVT, constant chemical potential, constant volume and constant temperature).

2.1. Equation of motion

In MD simulations, the classical Newton's equations of motion

$$m_i \ddot{r}_i = F_i \ , \tag{2.1}$$

where, m_i and F_i are the mass and the force respectively acting on atom *i*, are integrated and the trajectories $r_i(t)$ of all atoms are determined. The force acting on a given atom *i* is given by the negative gradient of the interaction potential:

$$F_i = -\frac{\partial V}{\partial r_i} . \tag{2.2}$$

The time evolution of the system can be obtained based on the initial starting configuration, the complete description of the interatomic interaction potential function $V(\{r_i\})$, and the coupling to the external world. The total force acting on a given atom *i* is obtained by considering the interaction forces due to bonded and non-bonding interaction with the rest of the atoms present in the system and the external forces

$$F_i = \sum_j (F_{ij}) + F_{i\_ext} . \quad (2.3)$$

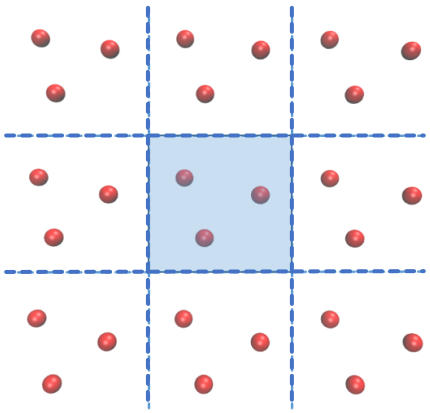


Figure 2.1. Schematic representation of periodic boundary conditions.

Very often MD simulations are carried on in systems under periodic boundary conditions, as illustrated in Figure 2.1. Often, when the interactions between atoms are of short range nature, the force acting on a given atom can be evaluated with reasonable accuracy by considering only the atom interaction with the atoms located nearby in withing some cut-off distance. When the MD simulation is acarried on in systems undr periodic boundary conditions, the maximum value for the cutoff distance for nonbonded interactions must not be greater than half of the box length used in the simulation. There are a few MD simulation package, such as Gromacs [21], that one can use to perform the MD simulations. The default integrator named leapfrog algorithm is commonly used to integrate in time the Newton's equations of motion. According to the leapfrog algorithm, first the

velocity at midpoint between time t and t + Δt is calculated, then this value is used to update the positions as follows:

$$\mathrm{v}\left(\mathrm{t}+\frac{\Delta \mathrm{t}}{2}\right)=\mathrm{v}\left(\mathrm{t}-\frac{\Delta \mathrm{t}}{2}\right)+a(t)\Delta t \tag{2.4}$$

$$\mathrm{r}(\mathrm{t}+\Delta \mathrm{t})=r(t)+\Delta t\mathrm{v}\left(\mathrm{t}+\frac{\Delta \mathrm{t}}{2}\right). \tag{2.5}$$

Another well-known integrator is called velocity verlet algorithm [22]. Based on positions r(t), the next step positions r(t+ Δt), velocity v(t) and acceleration a(t), can be derived from Tylor expansion about time t:

$$r(\mathrm{t}+\Delta \mathrm{t})=r(t)+v(t)\Delta t+\frac{a(t)}{2}\Delta t^2+\frac{\dddot{r}\Delta t^3}{3!}+\mathrm{O}(\Delta \mathrm{t})^4 \tag{2.6}$$

$$r(\mathrm{t}-\Delta \mathrm{t})=\mathrm{r}(t)-v(t)\Delta t+\frac{a(t)}{2}\Delta t^2-\frac{\dddot{r}\Delta t^3}{3!}+\mathrm{O}(\Delta \mathrm{t})^4 \tag{2.7}$$

$$\mathrm{r}(\mathrm{t}+\Delta \mathrm{t})+r(t-\Delta t)=2r(t)+a(t)\Delta t^2\ +\mathrm{O}(\Delta \mathrm{t})^4 \tag{2.8}$$

$$\mathrm{r}(\mathrm{t}+\Delta \mathrm{t})=r(t)+\Delta tv+\frac{\Delta \mathrm{t}^2}{2}a(t) \tag{2.9}$$

$$\mathrm{v}(\mathrm{t})=\frac{\boldsymbol{r}(\mathrm{t}+\Delta \mathrm{t})-\boldsymbol{r}(\mathrm{t}-\Delta \mathrm{t})}{2\Delta t}\ . \tag{2.10}$$

The Verlet integrator is a direct approach and has the advantage that requires less computational memory. The drawback is that the velocities are not obtained explicitly from the algorithm and are abtained based on the calculated trajectories.

2.2. Temperature control

To maintain temperature constant in a simulation system one need to mentain the total kinetic energy of the system at a constant value. According to the equipartition of the energy on degrees of freedom one can write:

$$\frac{3Nk_B T}{2} = \frac{\sum m_i v_i^2}{2}. \quad (2.11)$$

The distribution of the atomic vvelocities follows the Maxwell-Boltzmann distribution given by

$$f(v) = 4\pi \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2k_B T}}. \quad (2.12)$$

One of the simplest methodology that can be used to impose the constant temperature constrain is to periodicaly rescale the velocities of all the atoms in the system. An implementation of such an approach is provided by the use of the Berendsen Thermostats [23]. Other methods, such as those based on addition of stochastic forces on atoms, e.g. Andersen thermostat, Langevin thermostat, or Nosé-Hoover thermostats, have been developed and are widely used [24, 25].

Berendsen thermostat builds on the idea of weak coupling of the simulation system to a heat bath that is maintained at the desired temperature T_0 . The coupling to the heat bath will allow the system temperature, T, to relax slowly to that of the heat bath. The temperature relaxation process follows the equation:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}. \quad (2.13)$$

In practice, this process is implemented by rescaling the atoms velocities at every few time steps by using the scaling factor λ given by

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{T}{T_0} - 1\right)\right]^{\frac{1}{2}}, \quad (2.14)$$

where, Δt is the integration time step and τ_T is a relaxationtime constant characterizing the coupling to the heat bath. The larger values of τ_T leads to weaker coupling which means it will take longer for the system to reach the thermodynamic state characterized by temperature T_0 .

Velocity-rescaling thermostat was the temperature control methodology used in all of our simulations.. It is based on Berendsen thermostat except that an extra stochastic term is added. In this approach, the kinetic energy is adjusted at every few steps according to the following relation:

$$\mathrm{dK} = \frac{(\mathrm{K}_0 - K)dt}{\tau_T} + \frac{2\sqrt{KK_0}}{\sqrt{N_f}} dW/\sqrt{\tau_T} \ , \tag{2.15}$$

where N_f is the number of degrees of freedom, K the kinetic energy and dW a Wiener process.

2.3. Pressure control

The pressure in the simulation system [26] can be obtained using the virial theorem and it is given by:

$$\mathrm{p} = \frac{\mathrm{k_B} TN}{V} + \frac{1}{Vd} vir \tag{2.16}$$

$$\mathrm{vir} = \sum_{\mathrm{i<j}} \mathrm{f_{ij}} r_{ij} \ . \tag{2.17}$$

where, d is the dimensionality of the simulation system, V is the volume, f_{ij} is the force exerted on atom i by atom j. One should allow for the change of the volume of the simulation box in order to be able to maintain the systm at a constant pressure. Berendsen barostat is like a "pressure bath", in which the pressure in the system is relaxed to the target pressure via the first order differential equation:

$$\frac{\mathrm{dP}}{\mathrm{dt}} = \frac{P_0 - \mathrm{P}}{\tau_p} \ . \tag{2.18}$$

The strength of coupling is controlled by both the compressibility κ_T and the constant τ_p. In practice this is implemented by rescaling the simulation box cell length by using the matrix factors

μ_{ij} and all atoms positions and distances. That is, $x_{ij,} y_{ij}, z_{ij}$ are mapped into $\mu_{ij}x_{ij}, \mu_{ij}y_{ij}$ and $\mu_{ij}z_{ij}$ where μ_{ij} are given by

$$\mu_{ij} = \delta_{ij} - \frac{\kappa_T \Delta t}{3\tau_p}\left(P_{0ij} - P_{ij}\right). \tag{2.19}$$

2.4. Interatomic Potential Models

2.4.1. Non-bonded interactions

Periodic boundary conditions and minimum image convention are used to minimize the surface effects in simulation. The cut off radius for non-bonded interactions should not exceed half the shortest box cell length, which means, each particle should not 'see' its image under periodic boundary conditions.

The non-bonded interactions includes a repulsive term, an attractive term and a Coulomb term. The repulsive and attractive term are grouped into what is called van der Waals interaction term. The Coulomb term accounts for the electrostatic interaction. In the Gromos 54A7 force field that was used in our simulations the non-bonded interactions were described by a simpleLennard-Jones 6-12 potential [27]

$$V^{LJ}(r) = \sum_{pairs(i,j)}\left(\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}}\right). \tag{2.20}$$

The parameters $C_{ij}^{(12)}$ and $C_{ij}^{(6)}$ depend on the atom types as well as the pair interaction character. The Coulomb term is given by

$$V^{C}(r) = \sum_{pairs(i,j)} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \tag{2.21}$$

where ϵ_0 is the dielectric permittivity of vacuum and q_i and q_j are electric charge of the two atoms i and j and r_{ij} is separation distance between the atoms.

2.4.2. Bonded interactions

The bonded interactions include: bond stretching, bond angle bending, improper dihedral angle, and torsional dihedral angle interactions.

1.4.2.1. Bond stretching

The potential describing bond stretching is based on the harmonic spring model

$$\mathrm{V(r)} = \sum_{bonds} \frac{1}{2} k_{ij}^{b} \left(r_{ij} - b_{ij0}\right)^2. \quad (2.22)$$

In Gromos 54A7 this is implemented by using a more efficient formula

$$\mathrm{V(r)} = \sum_{bonds} \frac{1}{4} k_{ij}^{b} \left(r_{ij}^2 - b_{ij0}^2\right)^2, \quad (2.23)$$

where, k_{ij}^{b} is the stiffness of the spring and $\mathrm{b_{ij0}}$ is the equilibrium bond length. These parameters are assigned based on the atoms types, and are derived from experimental data.

1.4.2.2. Angle Potential

There are three atoms involved in the bond angle interaction term. In Gromos 54A7 functional form used is the cosine-based angle potential

$$\mathrm{V(\theta_{ijk})} = \frac{1}{2}\, k_{\theta}^{ijk} \left[\cos\theta_{ijk} - \cos\theta_{ijk}^{0}\right]^2 . \quad (2.24)$$

where, $k_{\theta ijk}$ is a force constant and is related to the harmonic potential force constant via the following relation $k_{\theta}^{ijk} = \frac{k_{\theta}^{harm}}{\sin^2(\theta_{ijk}^{0})}$.

1.4.2.3. Improper dihedral

Improper dihedral term is used to keep planar atomic group planar, for example, to keep aromatic rings planar.

1.4.2.4. Proper dihedral

Proper dihedral terms account for the energy change due to torsion around chemical bonds. The torsion angle ϕ is the angle between the ijk and kkl planes. A typical dihedral torsion function in Gromacs is

$$V_d(\phi_{ijkl}) = k_\phi(1 + \cos(n\phi - \phi_s)) \, . \qquad (2.26)$$

2.5. Constraint methods

To improve the limitation on the integration time step due to fast modes that characterize bond vibration, a holonomic constraint is usually used. The common algorithms through which these constraints are implemented are the SHAKE and the LINCS constraints. Both are used to correct the bond length after an unconstrained update. However, LINCS is 3 to 4 times faster than SHAKE and easier to parallelize [28].

2.6. Restraints methods

Restraints on particles are usually used to prohibit drastic rearrangements of molecular structures. For example position restraints are used to restrain certain groups of atoms in fixed configurations while allowing other groups to move freely; such an example is the restraining the motion of a nucleotide while allowing the surroung water molecules to hydrate and equilibrate in the surrounding space around nucleotide. Distance restraints are used to ensure that the distance between two atoms are not far off from the desired value. Flat bottom restraints define the bottom potential as zero while potential out of the flat region is harmonic.

2.7. Potential of mean force

Potential Mean Force (PMF) is an important feature in the analysis of molecular interactions since it provides the binding energy between the interacting molecules. In Gromacs the PMF can

be obtained by using the umbrella sampling used with the weighted histogram analysis method(WHAM) [29]. Molecular dynamics simulation performed under various ensemble conditions provides the means for evaluation of various properties via ensemble averaging. In canonical ensemble, i.e., constant number (N) of particles, constant volume(V) and constant temperature(T), the partition function, the probability of a microstate, and the entropy and the Helmholtz free energy are given by:

$$Q(N,V,T) = \sum_j e^{-Ei(N,V)/k_B T} \tag{2.27}$$

$$p_i = \frac{e^{-\frac{E_i(N,V)}{k_B T}}}{Q(N,V,T)} \tag{2.28}$$

$$S = -k_B \sum_i p_i \ln p_i \tag{2.29}$$

$$A = E - TS = -k_B T \ln Q\,. \tag{2.30}$$

where k_B is Boltzmann constant. T is temperature.

CHAPTER 3. DISTINGUISHING SINGLE NUCLEOTIDES MOVING THROUGH NANOSLITS COMPOSED OF SELF-ASSEMBLED MONOLAYERS

3.1. Introduction

3.1.1. Development of DNA sequencing methodologies

DNA sequencing has been a critical technology for molecular biology analysis. It not only provides important information for fundamental study but also has a great impact on human disease diagnostics and therapies [30]. Techniques have evolved rapidly from first-generation sequencing (Sanger's Method) to next-generation sequencing (platforms such as HiSeq®, Roche 454® are available commercially). In recent years, researchers have shifted focus from third-generation sequencing (Heliscope sequencer) to fourth-generation sequencing (Oxford Nanopore Technologies Limited). This nanopore-based technology provides a single molecule, real-time sequencing method which eliminates amplification, synthesis, or fluorescence labeling leading to higher efficiency and accuracy while reducing overall cost [30-32].

3.1.2. Fourth generation sequencing approach: Nanopores

The most prevalent fourth generation sequencing approaches are based on driving single-strand DNA (ssDNA) through nanopores. When a DNA nucleotide is in the pore, there is a blockage in the flow of other ions through the pore. Differences in the nucleotide bases cause slightly different fractions of the ion current to be blocked [31]. Materials used for nanopores can be sorted into biological and solid-state [33]. Biological nanopores mainly include the α-hemolysin membrane protein (α-HL) [34], the MspA nanopore [35, 36] and the Phi29 connector [33, 37]. Solid state nanopores such as SiN, SiO_2 [38] TiO_2, [39] Al_2O_3, boron nitride(BN), [40] graphenes [41], polymer membranes [38, 39, 42] and MoS_2 [37, 42] have been reported as promising

alternatives to biological nanopores due to their greater pore size flexibility [39] as well as better mechanical, chemical, and thermal stability [33, 37, 38].

3.1.3. Aim of the present work

DNA fragments or individual nucleotides are usually prepared before sequencing. Lambda exonuclease can selectively digest one 5'-phosphorylated chain of double-stranded DNA to generate mononucleotides(dAMP,dCMP, dGMP, dTMP) [43]. In this work, we continue our previous studies [44] based on the idea of enzymatic disassembly of DNA into its nucleotide monomers (dNMPs) followed by detection of those nucleotides at two or more locations along a nanochannel.

The flight times of the nucleotides between sensors differ due to the varied interactions of their nucleobase parts with the channel walls and potential with other species in the solution. The main advantage of this method compared to passing ssDNA through a single nanopore is that the sensors are only required to determine when a nucleotide passes. This is much easier than distinguishing between the different types of nucleotides. Information from the sensors could still be combined with the flight times to determine the nucleotide type, but the sensors alone would not have to make the determination. The disadvantage of this flight time method is the potential for nucleotides to pass each other leading to misidentification. This problem could be solved by either controlling the reaction rate of the enzyme cutting the DNA at the expense of slower analysis rates or by running multiple channels in parallel such that each segment of DNA is analyzed several times. Sampath used a mathematical analysis to show that the potential issue of dNMP disordering when DNA sequencing based on enzymatic disassembly can be overcome [45].

Although nanochannels are often constructed of amorphous polymer materials, these surfaces are rough which presents problems for atomistic molecular dynamics (MD) simulations. Using realistic roughness in simulations would require the surface tangential dimensions of the simulation box to be at least as large as the longest wavelength associated with the surface roughness which would lead to using an unrealistically large number of atoms. Even if smaller and smoother systems were used, there are still issues with time scale since with a rough surface it becomes likely that dNMPs will become trapped in holes on the surface for long periods of time relative to typical simulation times. Polymer surfaces such as poly(methyl methacrylate) (PMMA) can be chemically modified [46, 47]. Since we cannot simulate polymer surfaces realistically, in this work we study the effect of surface chemistry using self-assembled monolayers (SAMs) [48] which are smooth. SAMs usually consist of molecules with a thiol (-SH) group which spontaneously binds to the surface of a group 11 metal (usually gold is used). Surface properties can be altered by varying the terminal groups [49]. Although this is a much simpler system than an amorphous polymer with a modified surface, hopefully the results can still provide some guidance for choosing surface modifications that optimize time of flight based DNA sequencing.

The goal of this study is to estimate required channel lengths and analysis times per dNMP by statistically analyzing velocities or times of flight of dNMPs moving through nanoslits whose walls are composed of SAMs with different terminal groups in MD simulations. Although driving forces in MD simulations must necessarily be large due to limitations on the time that can be simulated, in this case, we are comparing different surfaces so we can still determine the relative performance of each surface. It is also possible to approximately extrapolate the results to realistic driving forces [44]. The simulations also provide atomistic insight into interactions between

nucleotides and SAM terminal groups and details on the dNMP adsorption and desorption dynamics.

3.2. Methodology

3.2.1. Simulation systems

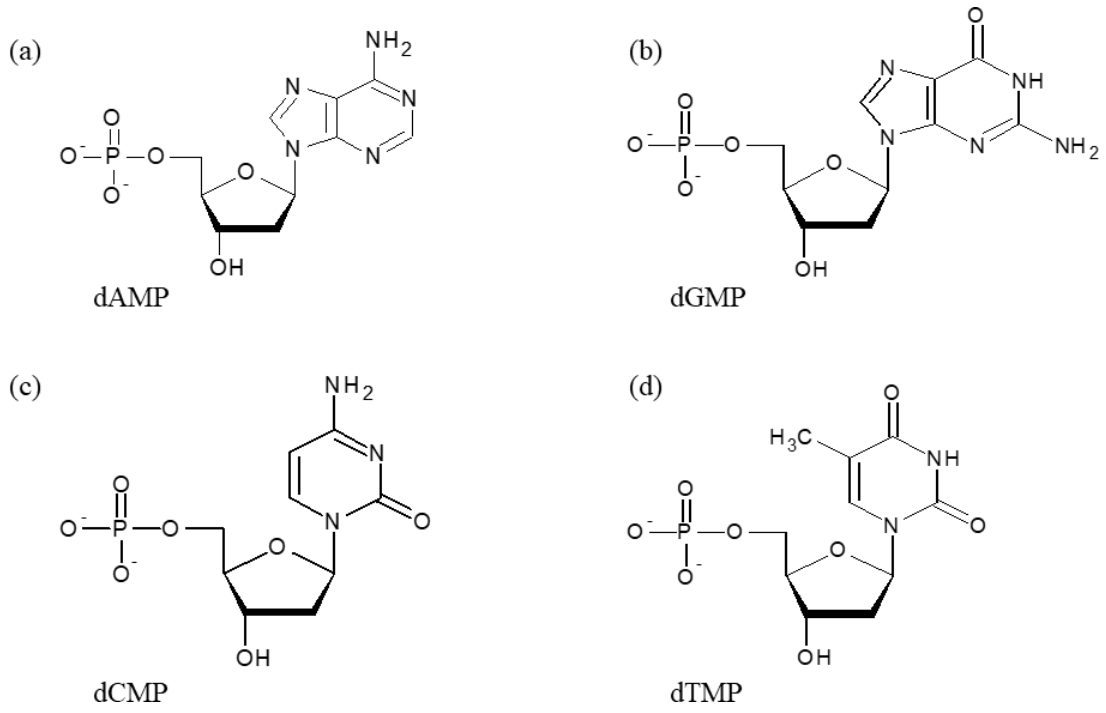


Figure 3.1. Structure of the four unmodified DNA nucleotides.

DNA is a heteropolymer whose monomer units are nucleotides (dNMPs), and each dNMP contains three components: a five-carbon sugar, a nitrogenous base, and a negatively charged phosphate group. We considered the four unmodified dNMPs (dAMP, dCMP, dGMP, and dTMP), which are shown in Figure 3.1 Since the pKa values for the phosphate groups are below 7.0 [50], we considered the -2 charged forms of the nucleotides.

The nanoslit walls were composed of three different types of SAMs: one composed of methylformyl terminated molecules (HS(CH2)11COOCH3) which will be referred to as

COOCH3-SAM, one composed of methyl terminated molecules (HS(CH2)11CH3) which will be referred to as CH3-SAM, and one composed of phenoxy terminated molecules (HS(CH2)11OC6H5) which will be referred to as OC6H5-SAM. For comparison, unmodified PMMA has methylformyl and methyl side chains. The structure of the SAM molecules is presented in Figure 3.2.

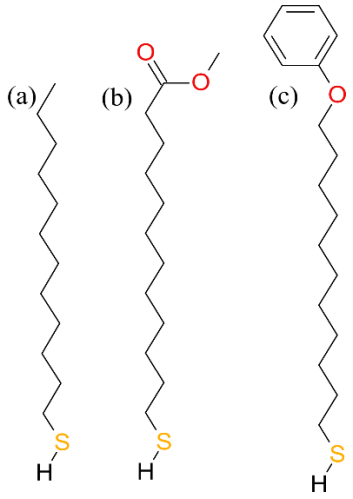


Figure 3.2. Structures of (a) methyl terminated, (b) methyl acetate terminated , and (c) phenoxy terminated thiol molecules contained in the SAMs.

A typical MD simulation system used in this study is illustrated in Figure 3.4. The SAM structures were downloaded from 'Dr. Latour's Biomolecular Modeling Lab' website [51]. Monolayers along with water were isolated from these original files which also contain peptides or other molecules. To construct a channel, a single monolayer was duplicated and reflected over a xy plane such that the SAM terminal groups point toward the center of the channel. Periodic boundary conditions were used in all directions, so a 1.5 nm gap was created between one layer of sulfur atoms and the periodic image of the other layer of sulfur atoms to ensure no short-range interaction between them. The width of the slit was defined as the average distance between sulfur atoms on both walls. Since the different types of monolayer molecules have different sizes, the slit

widths could not be chosen to be the same. To provide a reasonable comparison between systems, the channel widths were chosen so that the width of bulk water in each system was approximately 0.2 nm. This criterion leads to channel widths of about 5.6 nm for the COOCH3-SAM and CH3-SAM systems and about 6.0 nm for the OC6H5-SAM system. The number of water molecules was chosen to get approximately the correct density of bulk water of around 1.015 g/cm^3 at 300 K and 1 bar [52]. One dNMP, 3 Cl-, and 1 Na+ were then placed in the water to provide a neutral system. Water molecules with any atom within 0.18 nm of a dNMP atom were removed. An ion replaced a single water molecule. This ion concentration is about 64 mM. A snapshot from a OC6H5-SAM simulation is shown in Figure 3.4.

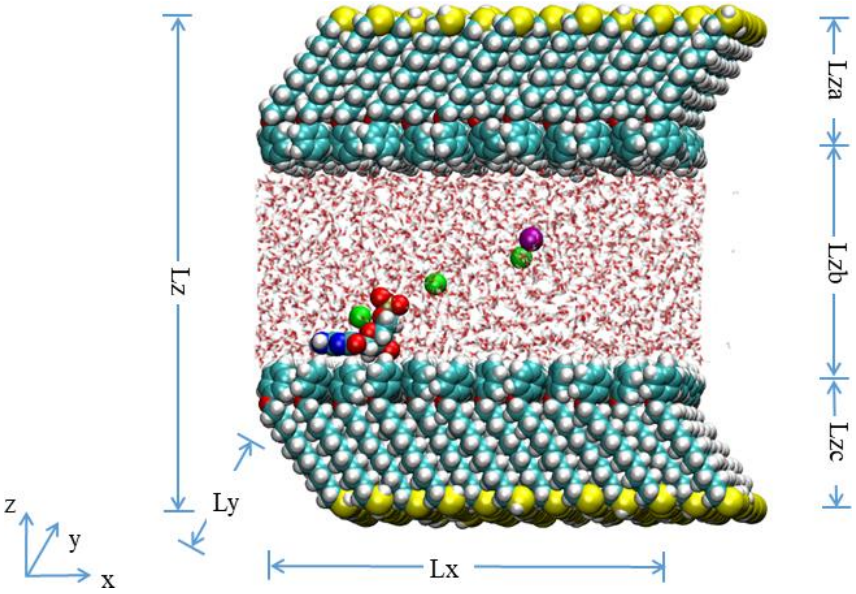


Figure 3.3. A simulation system containing a dNMP, water, and sodium chloride in a nanoslit with SAM (OC6H5-SAM shown here) walls. Tilted chains are SAMs. The dNMP is on the bottom wall. The yellow, cyan, white, red, blue, tan, green, and purple spheres represent S, C, H, O, N, P, Na^+, and Cl^- atoms or ions, respectively. For clarity, water is represented with smaller red and white transparent spheres with bonds. The labelled dimensions are Lx = 5.196 nm, Ly = 5.000 nm, Lz = 5.972 nm, Lza = Lzc

(caption cont'd)

= 1.477 nm, and Lzb = 3.020 nm. The simulation sizes (Lx×Ly×Lz) for the COOCH3-SAM and CH3-SAM systems are 5.196×5.000×5.6 nm.

3.2.3. Simulation details

Force field

The rigid CHARMM TIP3P [52] water model was used. For monolayers, ions, and dNMPs [44], the CHARMM36 force field was employed. A single molecule from the monolayer structure was extracted, converted to a stream file using Open Babel [53], and submitted to https://cgenff.paramchem.org/initguess/ to generate CHARMM General Force Field (CGenFF) [54] parameters. Sulfur atom positions, $\boldsymbol{r}_i$, were restrained near their initial positions, $\boldsymbol{r}_i^0$, using harmonic potentials, $V_{pr}(\mathbf{r}_i) = 0.5k_{pr}\left|\boldsymbol{r}_i - \boldsymbol{r}_i^0\right|^2$, with force constants k_{pr}=10000 N/m. This imitates thiol attachment onto a smooth metal substrate, which is commonly the gold (111) plane [55, 56] without requiring explicit metal atoms.

General simulation details

All simulations were performed using GROMACS 5.0.6 [57]. The short range cutoff distance for electrostatic and van der Waals interactions was 1.0 nm. The particle-mesh Ewald (PME) method was used for both long range Coulomb and van der Waals interactions. The equilibrium simulations were run in the NVT ensemble at 300 K with the thermostat of Bussi, Donadio, and Parrinello [58] with a time constant of 0.1 ps. For the non-equilibrium simulations, the thermostat was only applied on the wall atoms. An electric field with a strength of 0.1 V/nm was applied in the x direction, forcing the nucleotides to move in the negative x direction.

Potential of mean force

The potential of mean force difference with respect to some reference state, ΔF, as a function of some reaction coordinate(s) is related to the probability, *Pr*, of the dNMP having a given value(s)

of the reaction coordinate(s). We chose the distance along the surface normal direction, z, between the center of mass of the nucleobase parts of the dNMPs and the center of the sulfur atoms in one of the SAMs, d_S. The PMF profile is

$$\Delta F(d_S) = -kT\, ln\left[\frac{Pr(d_S)}{Pr(d_{S,ref})}\right] . \quad (3.1)$$

3.3. Results and discussion

3.3.1. Equilibrium PMF calculations

In Figure 3.4, the PMFs for all dNMPs in k_BT units are shown as a function of the distance from the center of mass of the nucleobase to the average position of the wall surface, d_w. This distance rather than d_S which was used to calculate the PMFs was used for comparison of different SAMs which have different depths. The average wall surface was determined by solvent accessible surface area (SASA) calculations on the SAM surfaces in water. A probe with a radius of r_{probe} = 0.388 nm was employed which is an effective size for the dNMPs (see previous supplement material [59]). To convert from d_S to d_w:

$$d_w = d_S - \left(\langle z_{SAS}\rangle - r_{probe} - \langle z_S\rangle\right) . \quad (3.2)$$

where $\langle z_{SAS}\rangle$ and $\langle z_S\rangle$ are the average positions of the solvent accessible surface and sulfur atoms in the interface normal direction, respectively.

Overall, the PMFs show a typical shape for simple adsorption. Figure 3.4 shows that when the dNMPs are close to the wall, the free energies are very high and positive because of repulsion. The PMFs drop quickly with increasing distance to negative values which is a transition from repulsion to an attractive interaction. The PMFs have global minima at around d_w = 0.2 nm, then increase out to about 0.5 nm for all systems, After 0.5 nm, there are no significant interactions between the wall and the dNMPs so the PMFs are no longer a function of distance. Since the PMFs

are shifted to zero in the bulk solvent at large d_w and all PMFs have values below zero, the dNMPs have more affinity for all SAM surfaces than for the solution.

For the COOCH3-SAM system, the PMFs have two potential wells near the wall, which could be caused by the end group, i.e. methyl ester, whose conformations are more complicated than the other two SAM terminal group. The methyl group is at the surface instead of the carbonyl group most of the time. The first small PMF well can be explained by the interaction between SAM carbonyl oxygens and the dNMPs dominates while after that the methyl group-dNMP interaction dominates. When the former situation occurs, the hydrogen bonds canceled some repulsive forces.

The global minimum values in the PMF profiles are shown in Figure 3.5 as ΔF_{min} which is directly related to the strength of the interactions between the walls and dNMPs. The magnitude of the global PMF minima is largest for OC6H5-SAM followed by CH3-SAM and COOCH3-SAM regardless of the type of dNMP, indicating that all dNMPs have the highest affinity for OC6H5-SAM and the lowest affinity for COOCH3-SAM. The location of the PMF minima are about the same for different dNMP types on the same surface. The locations are at about d_w = 0.21 nm, 0.19 nm, and 0.22 nm for COOCH3-SAM, CH3-SAM, and OC6H5-SAM systems, respectively.

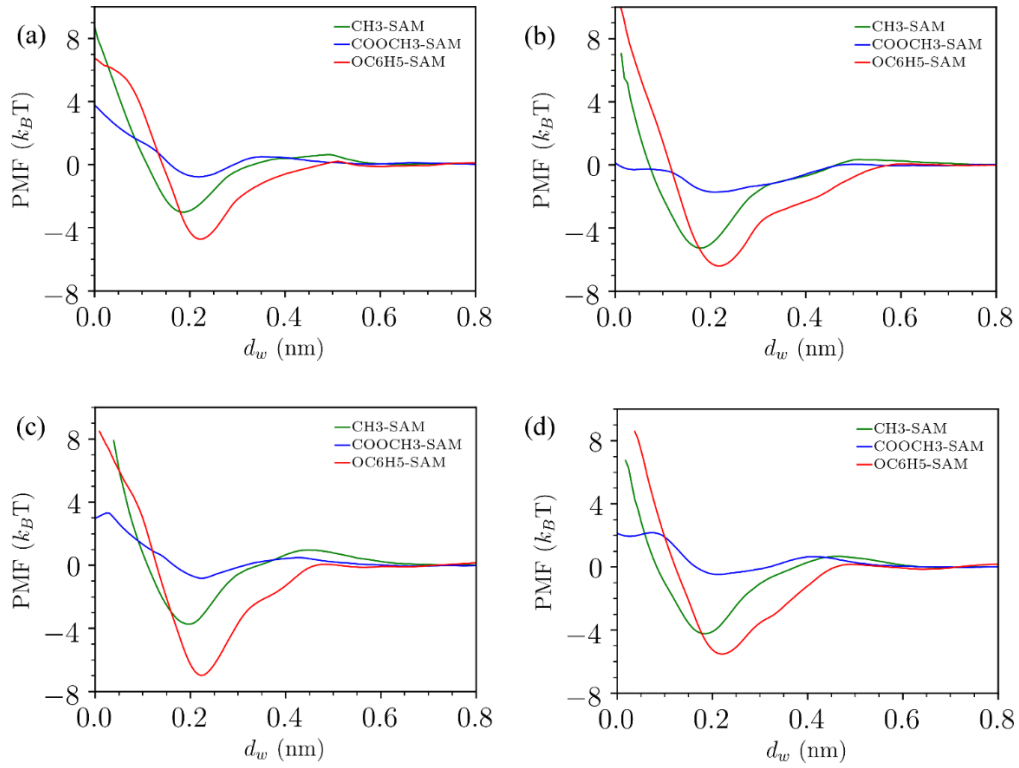


Figure 3.4. Comparison of PMFs in k_BT as a function of distance (nm) from COM of dNMPs base to the wall surfaces. The different SAMs, COOCH3-SAM, CH3-SAM, OC6H5-SAM, are represented by blue, green and red curves respectively, with nucleotides (a) dCMP (b) dTMP (c) dGMP (d) dAMP.

The PMF curve itself is not usually obtained from experiments. Instead, a mean adsorption free energy is measured. The net free energy of adsorption was obtained by integrating the PMF curves [44].

$$\Delta F_{ads} = -\ln\left(\frac{C_{ads}}{C_{bulk}}\right) = \frac{\int_0^{d_w} PMF(e^{-PMF})d(dw)}{\int_0^{d_w}(e^{-PMF})d(dw)}, \quad (3.3)$$

where C_{ads} and C_{bulk} is concentration in the adsorption region and bulk solution respectively.

Table 3.1. Global minimum values of the PMFs, ΔF_{min}, and free energies of adsorption, ΔF_{ads}, for the four dNMPs interacting with the SAM walls.

	dNMP type	ΔF_{min}/kT	ΔF_{ads}/kT
COOCH3-SAM	dCMP	-0.77	-0.02
	dTMP	-1.71	-0.93
	dGMP	-1.37	-0.06
	dAMP	-0.48	0.06
CH3-SAM	dCMP	-3.00	-1.94
	dTMP	-5.27	-4.52
	dGMP	-3.73	-2.80
	dAMP	-4.23	-3.37
OC6H5-SAM	dCMP	-4.72	-3.85
	dTMP	-6.40	-5.71
	dGMP	-6.99	-6.29
	dAMP	-5.52	-4.78

3.3.2. Nonequilibrium dNMP adsorption and desorption behavior

The interactions between dNMPs and the wall are of great importance since they are the key factor for distinguishing their identity based on time of flight. As observed previously [44, 59], when the dNMPs are adsorbed to the wall the nucleobase parts sit on the wall while the phosphate groups point away from the surface into the solvent. During translocation, nucleotides adsorbed and desorbed from the surface multiple times. This behavior can be seen in Figure 3.5 which shows the dNMP positions in the wall normal direction as a function of time for the OC6H5-SAM system where the center of the nanoslit was set at z = 0 nm. Plots for the systems with the other SAM types are shown in the APPENDIX A. The dNMPs are adsorbed to the surface when they are near ±1.0 nm. The adsorption and desorption behavior clearly depend on the dNMP type. dCMP

adsorbs and desorbs frequently, dTMP stays adsorbed for much longer times on average before desorbing, and dAMP and dGMP have similar behavior somewhere between the behavior of dCMP and dTMP.

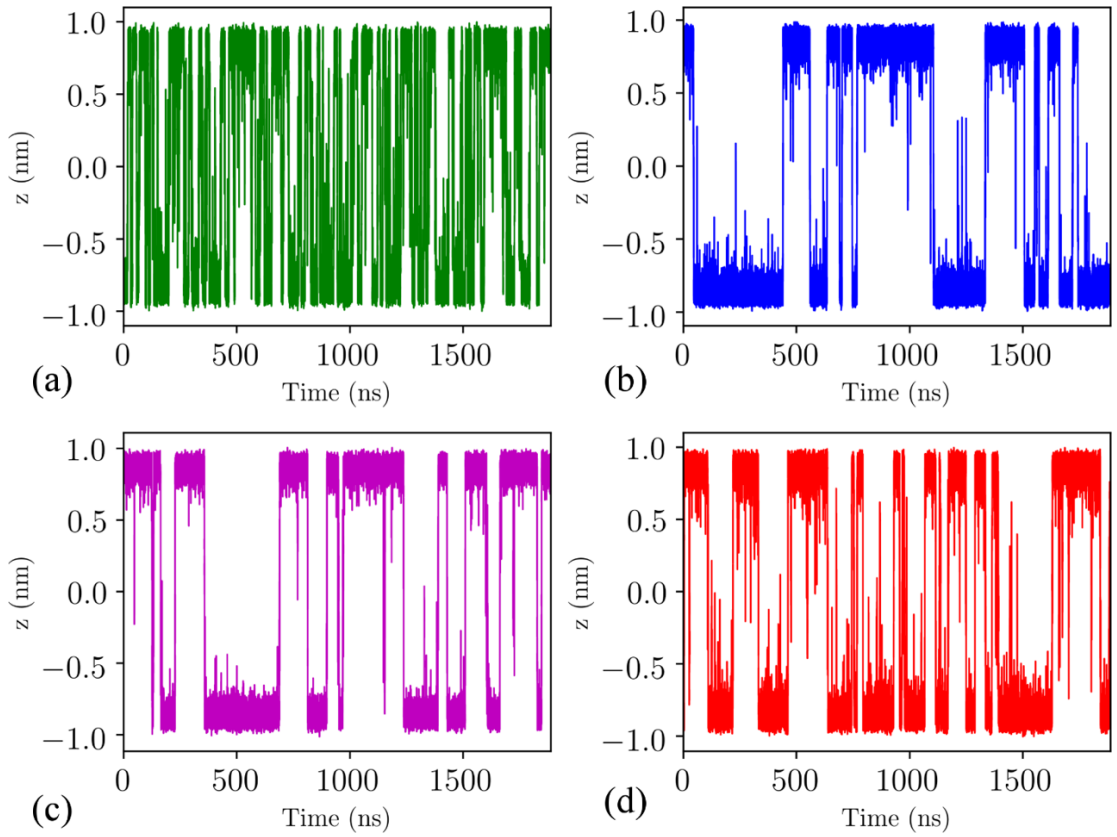


Figure 3.5. dNMP position, *z*, relative to the slit centerline as a function of time for the OC6H5-SAM walls: (a) dCMP, (b) dTMP, (c) dGMP, (d) dAMP.

Figure 3.6 shows the distance distributions from non-equilibrium simulations with electric field compared with equilibrium distance distributions obtained from PMF calculations in the OC6H5-SAM system. Plots for systems with other SAMs are in APPENDIX A. The peaks in the distributions for the equilibrium cases are higher compared to the non-equilibrium cases indicating a reduced dNMP affinity for the wall surfaces in the presence of the electric field.

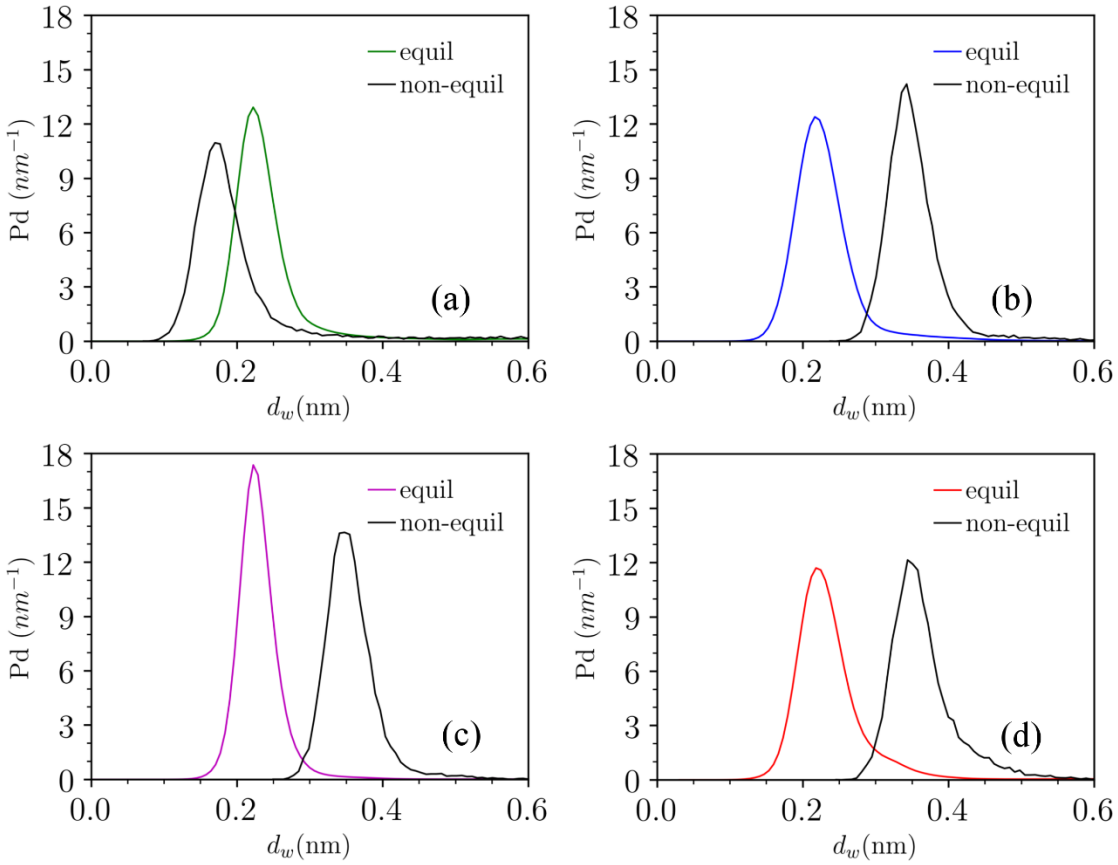


Figure 3.6. Comparison of probability densities as a function of d_w (nm) for dNMPs in the OC6H5-SAM system. Blue curves are from non-equilibrium simulation while the orange colors are obtained from equilibrium PMF simulations. (a) dCMP (b)dTMP (c) dGMP (d) dAMP.

To quantify the adsorption and desorption behavior, we analyzed the dNMP probabilities as a function of d_w to calculate fractions of time adsorbed which are shown in Table 3.2. The fraction of time adsorbed is defined as the probability that d_w is less than 0.4 nm, since the probability distributions stop changing on average for d_w greater than about 0.4 nm. The fractions of time adsorbed show that the dNMPs adsorb least to the most hydrophilic surface, COOCH3-SAM, and most to the OC6H5-SAM surface. The fractions of time adsorbed are generally higher for dNMPs whose nucleobases are more hydrophobic. The base hydrophobicity order is dTMP > dAMP > dCMP > dGMP [60] which is the same as the order of the fractions of time adsorbed for the CH3-

SAM and OC6H5-SAM systems while the order of the fractions of time adsorbed is swapped for dCMP and dGMP for the COOCH3-SAM system. Adsorption frequencies might also be calculated from the data. However, using a simple cutoff such as 0.4 nm leads to a very high frequency for some systems and especially with dCMP. Therefore, we calculated an average slit crossing time instead. This crossing time was defined as the mean time between when a desorbed dNMP adsorbed to one wall and the next time it adsorbed to the opposite wall. This means that a crossing time includes one adsorption period and one desorption period. To define adsorption, we used the same cutoff of d_w = 0.4 nm. As an example, a starting time would be when a desorbed dNMP moves across d_w = 0.4 nm near the lower wall and the ending time would be the first time after the starting time when the dNMP crosses d_w = 0.4 nm near the upper wall. This definition ignores any desorptions that end in adsorption to the same wall.

Table 3.2. Simulation time, percentage of time adsorbed , and average slit crossing time. Uncertainties are half the width of the 95% confidence intervals.

	dNMP type	SAM type		
		COOCH3-SAM	CH3-SAM	OC6H5-SAM
Simulation time (ns)	dNMP	1875	1700	1888
Percentage of time adsorbed (%)	dCMP	17.80	22.44	70.08
	dTMP	31.53	47.41	83.32
	dGMP	16.51	53.20	91.19
	dAMP	15.30	43.23	82.77
Average slit crossing time (ns)	dCMP	6.71 ± 0.35	10.00 ± 0.69	20.40 ± 1.78
	dTMP	7.73 ± 0.43	41.35 ± 5.89	91.51 ± 25.11
	dGMP	8.14 ± 0.51	62.53 ± 11.32	99.74 ± 19.17
	dAMP	7.62 ± 0.46	19.51 ± 2.57	66.15 ± 12.86

3.3.4. Mean velocities of dNMPs traveling through nanoslits

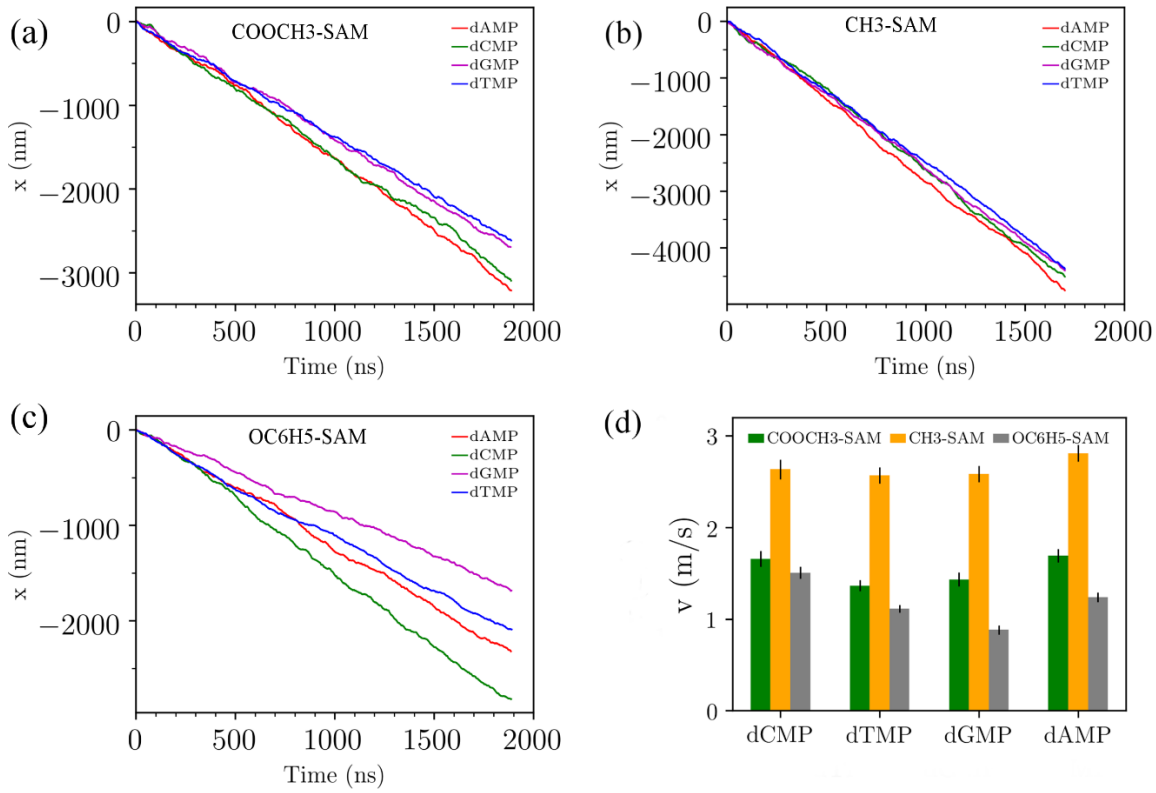


Figure 3.7. dNMP position in the x direction as a function of time : (a) COOCH3-SAM (1880 ns), (b) CH3-SAM (1700 ns), (c) OC6H5-SAM (1888 ns). Trajectories are shifted so the starting positions for the dNMPs are zero. Application of the electric field in the positive x direction moves the negatively charged dNMPs in the negative x direction. (d) Mean dNMP velocities with error bars representing 95% confidence intervals computed using a block bootstrap method [61].

Figure 3.7(a-c) shows the positions of the dNMPs in the electric field direction, x, as a function of time for each system. Driven by the electric field, the dNMPs traveled from the origin, x = 0 nm, in the negative x direction. The slope of these curves is the average dNMP velocity. dNMPs behave differently with different SAM wall surfaces. Without any further analysis, this plot already shows that the OC6H5-SAM surface provides the best separation of the dNMP velocities, the COOCH3-SAM surface separates the dTMP, dGMP pair from the dAMP, dCMP pair but not the dNMP types within those pairs, and the CH3-SAM surface has poor separation for

any of the dNMP types. Figure 3.7(d) shows the mean velocities for the dNMPs in the three systems in (a-c).

Table 3.3 shows the overall mean dNMP velocities and the mean velocities while adsorbed to or desorbed from the slit walls. Adsorption and desorption periods were defined using the same criterion that was used to compute the fraction of time adsorbed. All dNMP velocities are higher when desorbed than when adsorbed, but the ratio depends on the wall type and dNMP type.

Table 3.3. Mean dNMP velocities (m/s). Overall and while adsorbed or desorbed from the slit walls. Des/ads is the ratio of velocity while desorbed to velocity while adsorbed. Uncertainties are half the width of the 95% confidence intervals.

	COOCH3-SAM		CH3-SAM		OC6H5-SAM	
	Overall	Adsorbed	Overall	Adsorbed	Overall	Adsorbed
	Des/ads	Desorbed	Des/ads	Desorbed	Des/ads	Desorbed
dCMP	1.66±0.09	0.56±0.12	2.64±0.11	2.00±0.27	1.51±0.06	1.17±0.11
	3.41±1.67	1.91±0.20	1.45±0.89	2.90±0.24	2.03±2.09	2.37±0.23
dTMP	1.38±0.06	0.48±0.07	2.57±0.09	2.32±0.18	1.12±0.04	0.97±0.07
	3.92±2.14	1.88±0.15	1.23±1.28	2.86±0.23	2.42±5.00	2.35±0.35
dGMP	1.43±0.08	0.32±0.08	2.59±0.09	2.39±0.23	0.88±0.05	0.83±0.09
	5.22±2.13	1.67±0.17	1.16±1.30	2.77±0.3	2.29±6.11	1.90±0.55
dAMP	1.69±0.07	0.42±0.11	2.81±0.09	2.55±0.2	1.24±0.05	1.08±0.11
	4.62±1.45	1.94±0.16	1.22±1.40	3.11±0.28	1.94±2.73	2.10±0.30

The overall mean velocities in CH3-SAM system are the highest with values of for C, T, G and A approximately 2.64, 2.57, 2.59 and 2.81 m/s for dCMP, dTMP, dGMP, and dAMP, respectively (with uncertainty shown in Figure 3.7) makes it apparently the fastest group. This can be explained by the fact that the ratios of desorbed to adsorbed velocities are the smallest and the velocities while adsorbed are the largest in the CH3-SAM system. The dNMPs slide on the walls

almost as fast as they move in solution. The methyl group is the most hydrophobic group considered, so the water density is low near the wall surface which makes sliding easier. The methyl group also has no potential specific interactions with the dNMPs such as hydrogen bonds, and there is no pattern formed by SAM termini other than that imposed by the lattice.

The similar channel width in CH3-SAM and COOCH3-SAM system suggests a supplementary solution in CH3-SAM, which is attributable to its shorter chain length. It's reasonable that in CH3-SAM system, dNMPs travels fastest in desorption since the water depth is larger than the other two systems. Adsorption velocities in CH3-SAM surface are also fastest. This can be explained from the perspective of its sliding capacity near the wall. Methyl group is the most hydrophobic among all terminated groups: larger repels to water contributes to less resistance from the wall. Although the order of velocities for four dNMPs are: dAMP > dCMP > dGMP > dTMP which is the same as COOCH3-SAM system. Unlike the latter, these values are all so close to each other that an unfavorable channel length would be required in order to achieve separation. C and A, G and T are hard to be separated in COOCH3-SAM system as well as their similar transverse velocities. Due to hydrogen bonds between methyl ester group and the wall, it is reasonable that adsorption velocity is slower than other systems. Breaking hydrogen bonds so as to leave the wall requires energy although it's not that strong. OC6H5-SAM system is the most attractive one where dNMPs' overall mean velocities are the slowest. As talked before, free energy is lowest near the wall in this system. Also, as the benzyl group on the surface are able to rotate flexibly. The ring is likely to increase the contact area by lie parallel in xy plane, therefore, facilitate interaction with dNMPs.

3.3.5. Required minimum channel length for separation of dNMPs

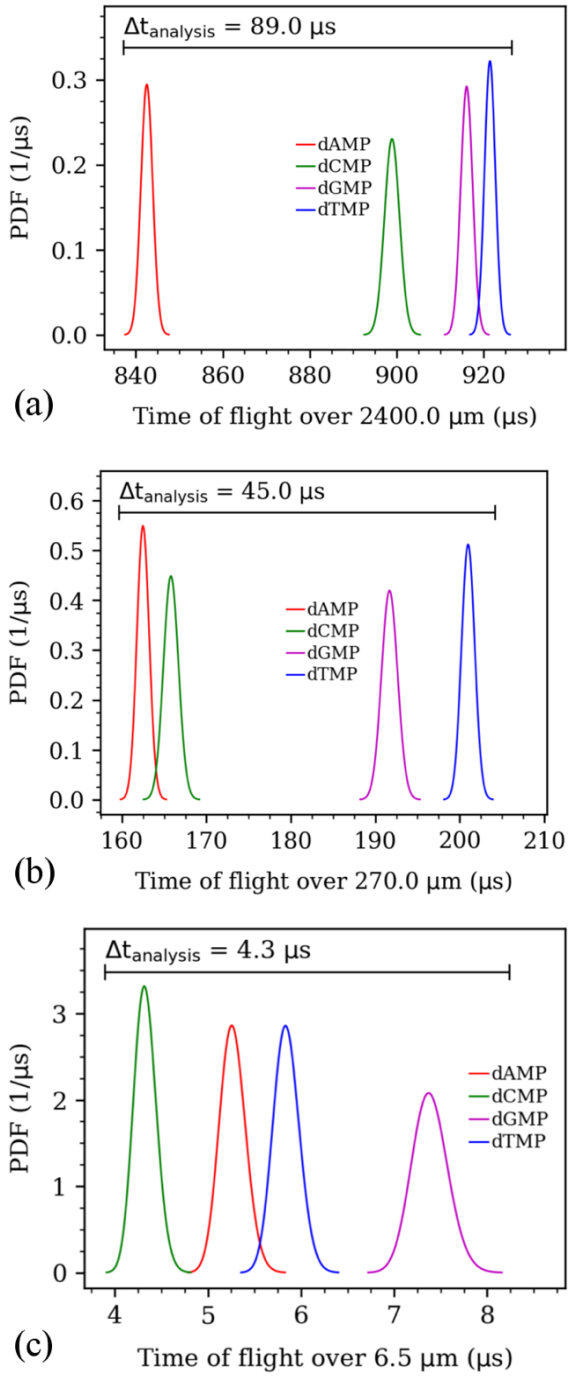


Figure 3.8. Time of flight distributions over a) 2400 um, (b) 270 um, (c) 6.5 um for CH3-SAM, COOCH3-SAM and OC6H5-SAM system respectively.

Therefore, a stronger binding force enables dNMPs to stick on the wall longer. For all systems, the desorption velocities are basically overlapped, since in the water the dNMPs are

traveling at almost the same speed without too many effects from the wall. The main differences could mainly be caused by association number of sodium ions when dNMPs desorbed from the wall.

Table 3.4. Minimum distances(μm) required to separate dNMP pairs for system (a) COOCH3, (b) CH3, (c) POXY

(a)

	dCMP	dTMP	dGMP	dAMP
dCMP	∞	2.7	5.9	**270**
dTMP	2.7	∞	40	1.8
dGMP	5.9	40	∞	3.8
dAMP	**270**	1.8	3.8	∞

(b)

	dCMP	dTMP	dGMP	dAMP
dCMP	∞	170	330	30
dTMP	170	∞	**2400**	11
dGMP	330	**2400**	∞	13
dAMP	30	11	13	∞

(c)

	dCMP	dTMP	dGMP	dAMP
dCMP	∞	0.84	0.29	2.1
dTMP	0.84	∞	1.2	**6.5**
dGMP	0.29	1.2	∞	0.67
dAMP	2.1	**6.5**	0.67	∞

When dNMPs enter to the channel, there are two detectors for tracking a dNMP's movement: one detector at the entrance and the other at the end of the channel. But it will be messed up if the dNMPs couldn't keep in order on the way of firstly passing the detector to the next detector. Thus, a minimum analysis time per dNMPs will be required which is determined by the distance between the lower edge of the distribution with the shortest flight times and the upper edge of the distribution with the longest flight times and accuracy is also taken into consideration.

Because of the large number of mean velocities which also include samplings with replacement by moving block bootstrap method, the mean velocities distribution is nearly normal distribution. The block length was chosen as 20ns which was estimated from the result of 'gmx analyze' so as it to be longer than autocorrelation time. The number(N) of mean velocities required to separate the overall mean velocity distributions to efficiency of 0.9973 was determined by approximating the simulation time is N times longer than the real ones. The minimum channel length for COOCH3-SAM system is 270 μm, for CH3-SAM system is 2400 μm and for OC6H5-SAM system is 6.5 μm. Basically, CH3-SAM and COOCH3-SAM wall didn't help too much for distinguishing dNMPs since their transverse motion weren't slowed down to an advantageous percentage.

Time of flight (TOF) PDF distribution for dNMPs are in different order from various surfaces. For CH3-SAM and COOCH3-SAM, TOF distribution in order of dAMP < dCMP < dGMP < dTMP and for OC6H5-SAM, dCMP < dAMP < dTMP < dGMP. In previous work [59], it requires at least 8 μs for the smooth carbon nanoslit and 78.9 μm of channel length and TOF distribution in order of dCMP < dTMP < dAMP < dGMP. Therefore, TOF is determined by interactions between dNMPs and surface. The higher affinity would cause dNMPs slower travel velocity inside nanoslit and may have a chance of distinguishing them. Basically, CH3-SAM and COOCH3-SAM

wall didn't help too much for distinguishing dNMPs since their transverse motion weren't slowed down to an advantageous percentage. It should be noted that the best surface for the sensors may not be the same as the best surface for the channel walls. The sensors are likely to involve some type of restriction with dimensions close to that of the dNMP molecule (nanopore). That means that strong binding of the dNMPs inside the pore may clog it as was seen with graphene, and solved by using a hydrophilic SAM on the graphene [62].

Since material had different effects on transverse of four types of nucleotides, it is promising to build a hybrid surface which leads to a high-efficiency sequencing device. Looking at the mean velocities for the CH3-SAM reveals that the pair of nucleotides with the most overlap in their velocity distributions are dGM P and dTMP. For the and COOCH3-SAM terminated system, the pair of nucleotides with the most overlap in their velocity distributions are dAMP and dCMP. This suggests that combining these two surfaces will be beneficial.

Information on complementary strand

The analysis in this paper assumes that each DNA strand is analyzed only once. If the DNA was instead analyzed multiple times where 50% of the time the forward strand was analyzed and 50% of the time the complementary strand was analyzed due to the exonuclease grabbing different ends of a double-stranded DNA segment, then the required slit length would be reduced substantially. For example, assume we analyzed the forward strand once and the complementary strand once for the CH3-SAM system where error in the forward strand comes from identifying dTMP or dGMP incorrectly. See Figure 8 (a). However, the complementary dNMPs (dAMP or dCMP) would always be identified correctly for a slit length of 2400 μm, so the overall error would be zero.

3.4. Conclusions

DNA might be sequenced by enzymatically disassembling it into individual nucleotides (dNMPs) which are passed through a nanochannel with two or more sensors used to measure the dNMP times of flight. We have used MD simulations of electrophoretic transport of dNMPs through nanoslits to study the effect of surface groups on required slit length to attain a given identification error rate. Extrapolation of the simulation data can be used to estimate the minimum slit length to obtain a given error. Self-assembled monolayers were used as a model for the slit walls since their surfaces are smooth and their terminal group can be altered. We considered methyl (CH3-SAM), methylformyl (COOCH3-SAM), and phenoxy (OC6H5-SAM) terminated SAMs. Although our systems are greatly simplified and our electric field strengths are very high by necessity, simulations can provide some insight into the type of functional groups that are most beneficial for separating the dNMP types by time of flight through a nanochannel.

Prior to the non-equilibrium simulations, we did equilibrium PMF calculations to determine the strength of interactions of the different types of SAM walls with the dNMPs. These PMFs give a good prediction of the percentages of time the dNMPs will be adsorbed to the walls, but the velocities of the dNMPs sliding on the walls also determine the required slit length. dNMPs adsorb strongest to the OC6H5-SAM walls, less strongly to the CH3-SAM walls, and weakest to the less hydrophobic COOCH3-SAM walls.

Based on the non-equilibrium simulation results, the OC6H5-SAM system is best with a required slit length of about 6.5 μm because the dNMPs adsorb to it the strongest which leads to the longest fractions of time adsorbed, and the velocities of the different dNMP types when adsorbed to the walls are significantly different. Adsorption of the dNMPs to the CH3-SAM surface is also relatively strong, but the velocities of the different dNMP types are not much

different when adsorbed to the walls. This leads to the COOCH3-SAM system being superior (required slit length of about 270 μm) to the CH3-SAM system (required slit length of about 2400 μm) even though the fractions of time adsorbed are significantly smaller than in the CH3-SAM system. The main difference between the COOCH3-SAM and CH3-SAM surfaces is that adsorbed dNMPs can form weak hydrogen bonds with the methylformyl oxygen atoms which helps to differentiate them. Although the OC6H5-SAM system is clearly the best, we note that a phenoxy modified rough polymer surface may show different behavior since the OC6H5-SAM is patterned with respect to orientations of the phenyl rings and those rings are tightly packed together which prevents the possibility of π– π stacking with the nucleobases which might occur with phenoxy groups attached to a rough surface. For the methyl and methyl ester groups, using a rough surface would probably not change the nature of the interactions of those groups with dNMPs that much.

The most difficult pair of dNMP types to separate with the CH3-SAM walls is dGMP and dTMP while the most difficult pair of dNMP types to separate with the COOCH3-SAM walls is dAMP and dCMP. Combining these two types of surfaces in the same slit reduces the slit length required to about 44% of the length required with only COOCH3-SAM walls.

CHAPTER 4. INTERACTIONS OF MONO AND DIMER LIGNOLS WITH LIPID BILAYERS

4.1. Introduction

Arvind Gupta said, “Biology is the most powerful technology ever created. DNA is software, protein is hardware, cells are factories.” The cell was first discovered by Robert Hooke in 1665 and known as a basic unit of living organisms [63]. Cell’s main function includes reproduction, metabolism and transport process. Genetic information is duplicated before cell division. Chemical reactions such as energy conversion and macromolecules synthesis happen inside the cell. Particles travel in and out cells to where it is needed [64]. Therefore, the first mission to survive is to protect integrity of cytoplasm from the surroundings. Cell membrane, which is also called plasma membrane, does the work. As a semipermeable barrier, it allows certain amounts and types of substance into the cell while blocking the others. It creates attachments to cytoskeleton to help to maintain cell shape [65]. Cell membranes are mainly composed of phospholipid bilayers embedded with proteins and cholesterol. The fluid mosaic model gives prominence to the structure of the lipid in plasma membrane [66].

Lignin is one of the most abundant biopolymers in nature and is found mostly in cell walls of plants. It is a polymer composed of the aromatic monomers (so-called monolignols) p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol that bind to each other via radical coupling and form a complex, heterogeneous structure. In plants, lignin plays important role in structuring cell walls and preventing water penetration through the xylem. Besides,it has been documented to have

---

This chapter was previously published as Tong, Xinjie, Mahsa Moradipour, Brian Novak, Poorya Kamali, Shardrack O. Asare, Barbara L. Knutson, Stephen E. Rankin, Bert C. Lynn, and Dorel Moldovan. "Experimental and Molecular Dynamics Simulation Study of the Effects of Lignin Dimers on the Gel-to-Fluid Phase Transition in DPPC Bilayers." *The Journal of Physical Chemistry B* 123, no. 39 (2019): 8247-8260. Reprinted by permission of ACS publications.

important roles in plant and tree defense against microbial attack [10]. Because it is naturally biodegradable, biocompatible, and has good stability and low toxicity, lignin derivatives are under investigation for multiple pharmacological and biomedical applications [11].

4.1.1. Antioxidant and antibacterial properties of lignins

Free radicals are known as unstable particles that accelerate oxidation process by stealing electrons from nearby molecules. Although it is regularly raised in metabolism and immune response, it can attack amino acids, lipids, DNA and proteins, and this accumulation are harmful to living cells. Antioxidants are beneficial to human health by neutralizing or removing free radicals, so they are also added to cosmetics and functional food. Besides life activities, polymers also call for radical scavengers to extend service life. Lignin extracted from wheat straw and bran boosts antioxidant property in 1% polypropylene specifically with lower molecular weight and fewer hydroxyl groups. Lignin aggregation was be observed under optical microscopy to estimate its solubility in material which is correlated to antioxidant activity [67]. Lignin materials from apple tree pruning were proven to be excellent antioxidants in ABST assay but they highly depends on obtaining process [68].

Antimicrobial mechanism has been mentioned in some research work but not specifically studied yet. Most of them attribute it to the damage on cell membrane caused by phenolic compounds. Antimicrobial properties vary concerning the lignin source as well as its processing method. Lignin extract from corn stover showed inhibitory activity only on Gram-negative bacteria (E. coli O157:H7 and S. Enteritidis) or bacteriophage MS2 but not on Gram-positive bacteria (L. monocytogenes and S. aureus) or yeast (C. lipolytica). High PH and good solubility in the culture(tryptic soy broth) were speculated as favorable terms for antimicrobial effects when compared lignin extract with commercial ligin [69]. Antimicrobial properties of intermediates of

the general phenylpropanoid and lignin specific pathways were investigated on bacteria and fungi, and hydroxycinnamaldehydes showed the highest activity [70].

It has also been proven that lignin and its derivatives (a.k.a. “lignins”) have antioxidant and antibacterial properties and they have already been used for the treatment of various diseases including diabetes, HIV, and cancer [11, 71]. The use of lignin nanoparticles (LNPs) in drug delivery has been reported in many studies [72, 73]. Besides benefits to human health, purified lignin and specific lignins have the potential to improve livestock intestine morphology and to control pathogens [74]. In many of these applications, the effects of lignins could be explained by the antioxidant capacity conferred by the phenolic character of its constituents [75]. It has been documented that low-molecular-weight oligomers derived from guaiacyl units are central to the antioxidant properties of lignins [76]. Often, in these applications, lignin absorbs into or binds to cell membranes or to membrane embedded molecular structures, such as proteins. For this reason, the investigation of the interaction of lignins with cell membranes is of critical importance.

4.1.2. Lipid bilayer transition temperature

Cell membranes are primarily comprised of lipids. In the physiological state, lipid bilayers are in their fluid state and this fluidity is pivotal to cell functions. The gel-fluid transition temperature (T_m) of lipid bilayers is the temperature where the lipids change from a gel phase where they are tightly packed in ordered domains to a more loosely packed fluid phase. Interactions between compounds such as drugs and lipid bilayers can affect their structure and permeability, and therefore T_m. Conversely, important changes in lipid and bilayer dynamics occur near the transition. For example, it was reported that the protein-mediated exchange of lipids between vesicles was greatly increased near the transition temperature [77]. Lipid transfer processes are essential to biological function.

Additive-induced changes in T_m depend on both the molecular structure and concentration of the additives as well as the identity of the lipids since these factors determine the location and concentration of the additives in the bilayer. This is supported by multiple experimental investigations. Mahendra examined the effect of a variety of small molecules including alkanols, fatty acids, detergents, and ionophores on dipalmitoyl lecithin bilayer phase behavior [78]. The plant-derived flavonoids quercetin, hesperetin, and naringenin were reported to lower T_m of 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) liposomes, while rutin (also a flavonoid) did not [79]. Prenyl-modified isoflavones had a more noticeable effect on T_m of DPPC than non-prenylated ones [80]. T_m was reduced by more than 10 K with broader transition peaks when adding a high concentration (20%) of isoflavone. Ojogun et al. investigated the partitioning and phase behavior of a series of nicotinic acid esters (nicotinates) between aqueous solution and DPPC membrane bilayers [81]. High concentrations of nicotinates increased the main phase transition temperature and decreased the membrane partition coefficient. Zhao et al. investigated the effects of cholesterol and paclitaxel on the phase behavior of DPPC liposomes, finding that both fluidize the bilayer [82]. Given that some lignin derivatives have polyphenolic drug-like structures, it is of interest to investigate their partitioning into lipid bilayers through effects on T_m.

Molecular dynamics simulation (MD) is a useful tool to gain information at the microscopic level and has contributed to understanding lipid bilayer phase behavior. However, most previous studies consider either pure lipids or lipid mixtures without additives [83-86], and most with additives were performed with coarse-grained models [87-90]. Atomistic studies of lipid bilayer phase transitions with additive molecules are still rare because of the time and length scale limitations of MD. Even if coarse-grained model improves the efficiency in system size simulation, it is still tough to parametrize for molecules that do not commonly exist in membranes. DPPC-

cholesterol phase diagram was constructed by using Martini force field. Structural properties value jump, such as area or volume per lipid, thickness, etc. were used to characterize DPPC transition temperature [88, 89]. DPPC-SP-$B_{1–25}$ (lung surfactant peptide) phase transition was also simulated with Martin model and suggested its presence fluidize the monolayer [90].

There are two approaches to estimate transition temperature using MD. The first approach is to mimic what is done experimentally by either heating a gel phase or performing several simulations at different temperatures starting from gel phase configurations. Leekumjorn and Sum [83] did heating and cooling scans for DPPC and 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine (DPPE) from 250 K to 350 K and back to 250 K where the temperatures are in the range of the gel-fluid transitions. Simulations using the Berger parameters for DPPC showed clear changes in the slope of the enthalpy during heating simulations allowing for easy identification of the phase transition [84]. Qin and Yu studied 1,2-disteroyl-sn-glycero-3-phosphocholine (DSPC) and 1,2-disteroyl-sn-glycero-3-phosphoethanolamine (DSPE) and found three stages during the DSPC transition which were from cross-tilted packing to tilted packing, tilted to “mixed” alignment, and “mixed” alignment to disordered, while DSPE didn’t have the “mixed” alignment state [85]. The second MD method involves simulating the gel and fluid phases in coexistence at various temperatures to find the temperature where neither phase grows. Coppock and Kindt used this method with DPPC and DSPC [86].

There is little existing work related to the interaction of lignin dimers and derivatives with lipid bilayers, and we are not aware of any study of their effect on the gel-fluid transition temperature. The transport of lignin monomers across the plant cell membrane has been studied and seems to involve ATP-binding cassette transporters [91-94]. Boija and Johansson studied the partitioning of monolignol and dilignol model substances in liposomes using immobilized

liposome chromatography and concluded that passive diffusion of some lignin precursors may be possible [10]. None of the compounds they considered are the same as those we consider here. Simulations of the interaction of lignin dimers with cellulose has been performed [95], but we are not aware of any MD related to their interaction with lipid bilayers.

4.1.3. Aim of the present work

Molecular dynamics simulations as a useful tool in nowadays research field also facilitate large amount of papers related to wood biopolymers. However, necessary study about lignin's compound such as monolignols, dimers and lignin's derivative has not received much attention. To investigate how lignin will interact with membrane, we can use lipid bilayer as a model in molecular simulation since it is believed to be a non-mediate process. First, from positioning and PMF study, we can understand binding energy between lignin molecules with lipid bilayer. Second, interaction of three derivatives of β-O-4' linked dimers of coniferyl alcohol (GG) with DPPC bilayers with the aim of developing a molecular-level understanding of how these molecules affect the gel-fluid phase behavior of the bilayers. We investigated the interaction of these molecules with liposomes of DPPC using high-resolution differential scanning calorimetry (DSC) and MD simulations. The DSC instrument incrementally changes the temperature of a sample and measures the heat flow. DSC allows phase transition temperatures to be identified easily by a peak in the heat flow and we applied it to identify the transition temperature of liposomes between gel and fluid phases [96]. Gel-fluid transition temperatures calculated from MD simulations showed the same trend as DSC results, and the simulations also helped to explain how the lipids were affected by the lignin dimers. All the experimental data were contributed by Moradipour Mahsa, Professor Knutson Barbara, Professor Rankin Stephen and Professor Lynn Bert from University of Kentucky.

4.2. Simulation systems

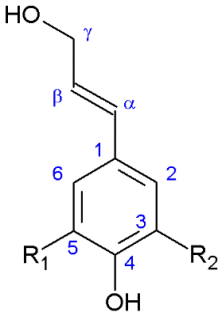


Figure 4.1. Monolignols are in three types (1) p-coumaryl alcohol (PCA) , R1=R2 = H; (2) coniferyl alcohol (CFA), R1= OCH3, R2 = H ;(3) sinapyl alcohol (SA) ,R1 = R2 = OCH3.

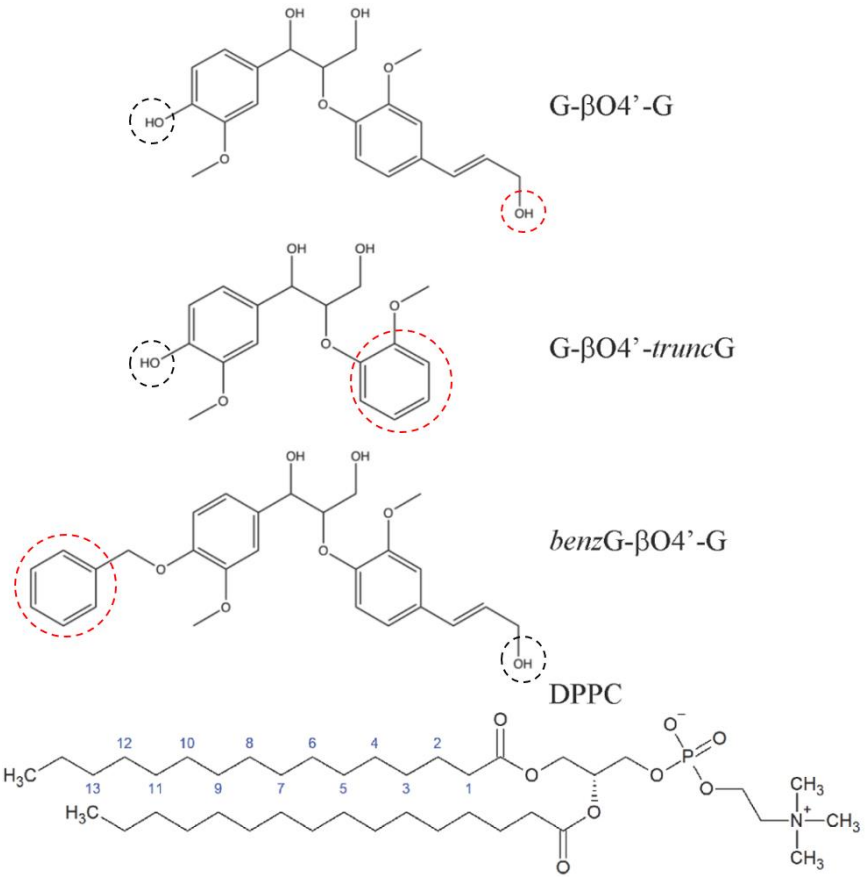


Figure 4.2. Structure of the molecules used in the experimental and simulation investigations.

4.3. Force field validation

Force field parameters for DPPC lipids and the topology file can be downloaded from the Automated Topology Builder (ATB) website [97]. To obtain parameters for the lignin molecules, several small molecules representing fragments of the lignin molecules were submitted to ATB. ATB output force constants for a few bonded interactions (bond angles in particular) which were clearly several times larger than typical values, and they were modified to more reasonable values obtained from other molecules on ATB with structures containing that type of interaction. To validate whether parameters obtained using ATB are reasonably accurate, a simulation of solid matter named erythro-2-(2,6-dimethoxy-4-methylphenoxy)-1-(4-hydroxy-3,5-dimethoxyphenyl)-propane-1,3-diol (EPD) was run as reported by Langer et al [98]. EPD has a structure that is very similar to a lignin dimer. Average unit cell dimensions and two average dihedral angles were compared with the results from experimental data. Details on the construction and validation of the lignol parameters are included in Appendix B.

4.4. Simulation details and methods

4.4.1. System setup

The MD simulation systems contained 128, 256, or 512 DPPC lipid molecules organized into flat bilayers, water, and one of the three variants of G-G lignin dimers. The bilayers were well hydrated with about 50 water molecules per lipid. All simulations were performed with the GROMACS 2016.3 package [99]. Constant number of atoms, pressure, and temperature (NPT) conditions were used for all simulations with a time step of 3 fs and fully periodic boundary conditions. The Berendsen barostat was used with a reference pressure of 1.0 bar and a time constant of 2.5 ps. The v-rescale thermostat [100] was used with a time constant of 0.1 ps. The cutoff for van der Waals interactions and short range Coulomb interactions was 1.4 nm. The

particle mesh Ewald (PME) method was used for long range Coulomb interactions. The interactions between atoms were described by the GROMOS 54a7 united atom force field [101] with the SPC water model [102].

4.4.2. Potential of mean force (PMF)

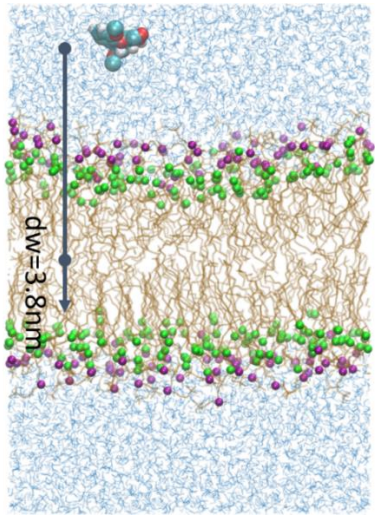


Figure 4.3. One lignin molecule was placed in water. A series of configuration will be generated along reaction coordinates for PMFs calculation.

The potential of mean force (PMF) can be used to approximate free energy differences as a function of one or more reaction coordinates, which in the case of lipid bilayers are often distances of molecules from the bilayer center. The PMF curve is related to the probabilities of observing different values of the reaction coordinate(s), and for lipid bilayers it is useful for estimating the most probable location(s) of molecules relative to the lipid molecules using the Boltzmann distribution, $\mathrm{p(i)} = \frac{e^{-\beta PMF(i)}}{\sum_{i=1}^{n} e^{-\beta PMF(i)}}$. In the case where the desired range of reaction coordinate(s) can be adequately sampled in unbiased simulations, the PMF curve can be obtained directly from those probabilities. In most cases, in order to achieve adequate sampling over the whole range of reaction coordinate(s), the system must be biased. Umbrella sampling is an approach which

involves biasing the system near different values of the reaction coordinate(s) in separate simulations (windows).

We computed PMFs in systems containing 128 DPPC molecules and a single lignin GG dimer. All three dimer variants were considered. Umbrella sampling [103] was used to calculate the PMF as a function of the distance from the center of mass of each lignin dimer to the center of lipid bilayer along the bilayer normal direction. The window spacing was 0.1 nm and 38 windows were used with equilibrium distances from 0.0 to 3.8 nm. Some starting configurations were simply extracted from initial unbiased simulations. Configurations for distances near the bilayer center that could not be obtained from the unbiased simulations were obtained by dragging the lignin molecule to the desired position at a rate of 0.00003 nm/ps. The umbrella potentials were harmonic restraints with force constants of 3000 kJ/ (mol × nm^2). Umbrella sampling simulations were run at 326 K and 1.0 bar for 80 ns for each window where 30 ns were used for equilibration time. The weighted histogram analysis method [104] is then applied to construct the unbiased PMF by combining data from all windows.

4.4.3. Partition coefficient

The partition coefficient K between two bulk phases is

$$K = e^{-\beta \Delta G} = e^{-\frac{\Delta G}{RT}} \ , \tag{4.1}$$

where R is the gas constant, T is temperature, and ΔG is the free energy difference between bulk phases. In MD simulations, the free energy difference can be obtained using the PMF as described above. However, an additional complication with lipids is that lipid bilayers are heterogeneous; there is no bulk lipid phase. This, in general, requires averaging the PMF [105] and perhaps defining a boundary between phases.

Various approximations can be made in defining the partition coefficient. When the difference between the minimum value of the PMF (PMF_{min}) and the value of PMF in bulk solution PMF_{sol} is large and the PMF curve is steep near its minimum, $\Delta G \approx PMF_{min} - PMF_{sol}$. The entire region of the system which affects the PMF of the molecule including the bilayer and interfacial water could be considered as one phase and bulk solution as the other phase. In that case, a Boltzmann weighted average of the PMF can be used:

$$log(K) = -\beta \Delta G_{avg} \approx -\beta \frac{\int_0^{z_{bulk}} PMF(z) e^{-\beta PMF(z)} dz}{\int_0^{z_{bulk}} e^{-\beta PMF(z)} dz} \quad . \tag{4.2}$$

where the PMF in bulk solution at a distance z_{bulk} from the bilayer center is set to 0. However, this definition is not appropriate for comparison to the partition coefficient, K, obtained from DSC based on eq 1 since the molecules outside the bilayer will not affect the transition temperature. We need to define a boundary between the lipid and solution phases. This is somewhat arbitrary, but we assume that the boundary is the average distance of the lipid phosphorus atoms from the bilayer center (z_{phos}) since anything outside the head groups of the lipids will have little effect on the transition temperature. We use the following definition of the partition coefficient [106]

$$log(K) \approx -\beta\left(PMF_{avg,l} - PMF_{avg,w}\right) + log\left(\frac{V_w}{V_l}\right). \tag{4.3}$$

$PMF_{avg,l}$ and $PMF_{avg,w}$ are the Boltzmann weighted average values of the PMF in the lipid phase and solution phase, respectively. They were calculated as in eq 3 except the bounds on the integrals are from 0 to z_{phos} and from z_{phos} to the furthest distance in the PMF curve, respectively. V_w and V_l are the volumes of the water and lipid phases, respectively and were defined by the same bounds as in the integrals.

4.4.4. Area and thickness compressibility modulus

Area compressibility modulus, K_A, is related to the average bilayer area, $\langle A \rangle$, and the variance of the area σ_A^2 [107]

$$\mathrm{K_A} = \frac{\langle \mathrm{A} \rangle \mathrm{k_B T}}{\sigma_\mathrm{A}^2} \quad . \tag{4.4}$$

The error estimation for σ_A^2 [108] simply assumes $A(t)$ is normally distributed. From the time autocorrelation function of $A(t)$, we can get a relaxation time, τ. The variance of σ_A^2 is related to τ by

$$\sigma^2(\sigma_A^2) = \sigma_f^2 = \frac{2\tau\sigma_A^2}{T_{run}}, \tag{4.5}$$

where T_{run} is the production simulation time. Then the variance of $\mathrm{K_A}$ can be estimated using propagation of errors as

$$\sigma_{K_A}^2 = \sigma_{<A>}^2 \left(\frac{k_B T}{\sigma_A^2}\right)^2 + \sigma_f^2 \left(\frac{\langle A \rangle k_B T}{(\sigma_A^2)^2}\right)^2 \quad , \tag{4.6}$$

where $\sigma_{<A>}^2$ is the variance of the mean area, and any covariance has been neglected. Similarly, we use the bilayer thickness D instead of area A for thickness compressibility:

$$\mathrm{K_D} = \frac{\langle D \rangle \mathrm{k_B T}}{\sigma_\mathrm{D}^2}. \tag{4.7}$$

4.4.5. Deuterium order parameter ($|\mathbf{S_{CD}}|$)

Lipid bilayers in liquid phase are very flexible and sensitive to transport process. Deuterium order parameter($|S_{CD}|$) is a measure of lipids structural orientation. In experiment, ^{2}H NMR are

generally used to get this property [109]. Nowadays, molecular simulation is also an important tool. Lipid order parameter is defined as:

$$|S_{CD}| = \langle \frac{3\cos^2 \alpha - 1}{2} \rangle \tag{4.8}$$

In this equation, α is the angle between the CD bond vector and the membrane normal. To obtain order parameter by using Gromacs function, two index files contain only carbons in the same acyl chain group are created. The number of carbons for each group should be equivalent. The results are about each atom for carbon tails.

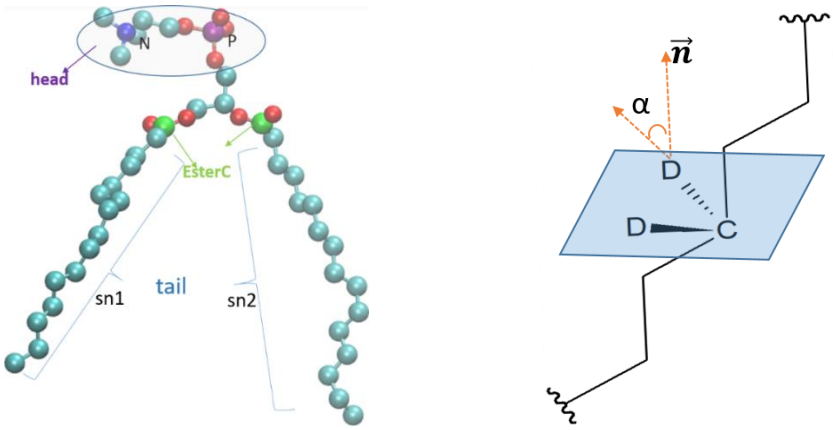


Figure 4.4. A visualized representation of a lipid. $\vec{n}$ is the normal to lipid bilayer.

4.4.6. Transition Temperature

One way to estimate a transition temperature is to simply simulate the two phases in coexistence in one simulation box. At the transition temperature, the system will remain in coexistence. At temperatures above or below the transition temperature, one phase will grow at the expense of the other and eventually only one phase will exist. This method has been used for lipids [86] and was termed the stripe growth method.

We used the coexistence method with a DPPC bilayer. First, a gel phase DPPC bilayer with 128 lipids was prepared by cooling from 326 K to 250 K at a rate of 0.253 K/ns. Note that the

freezing/melting point of SPC water model is 190 K [110]. This system was then replicated in one direction to get an elongated bilayer with 256 lipids, and the lipids were split into two regions whose boundary was perpendicular to that direction. The temperature of half of the lipids was set to 288 K so that this region remained in the gel phase. The other half was set at 353 K, so it melted into the fluid phase. The temperature of the water was set to 288 K. Pressure coupling was applied only in the direction perpendicular to the lipid phase boundaries. After melting half of the lipids, the temperature of the whole system was then changed to a temperature near the transition temperature. Lipids in the gel phase have dihedral angles in their tails close to trans configuration ($\pm$ 60^0 around trans configuration). The sign of the slope of the fraction of trans dihedrals in the lipid tails vs. time indicates whether the current temperature is above or below the transition temperature. The magnitude of this slope is an indication of how far the current temperature is from the transition temperature. The coexistence method works well for pure DPPC, but it was not used for DPPC with lignin due to the difficulty in obtaining the correct distribution of lignin molecules throughout the system. The coexistence results did serve as a useful reference for comparison with the heating method discussed next.

Another way to estimate the transition temperature is to simply heat up the system until the transition is observed as in experiments. The problem with this approach is that the heating rates in MD simulations must necessarily be very high which leads to overestimation of the transition temperature. As will be shown in the results section, by comparing to the coexistence method this overestimation was determined to be only a few degrees for pure DPPC systems. An advantage of this method is that it is simple and could be used for the systems containing DPPC and lignin molecules as easily as for pure DPPC systems.

The heating method was used with both a pure DPPC bilayer as well as systems with relatively high lignin dimer concentrations of about 0.158 or 0.289 mole fraction of dimers relative to lipids. The relativelly high concentrations were used to obtain significant shifts in the transition temperature. For bilayers with added lignin dimers, two different procedures were tried. In the first procedure, the fluid-phase systems with lignin dimers were cooled from 326 K to 250 K at a rate of 0.253 K/ns and held at 250 K for another 100 ns to create a gel phase. In the second procedure, lignin dimers were added to the solution outside the pure DPPC gel phase bilayer followed by equilibiration before heating. All gel phase sytems were heated back up from 250 K to 326 K at the same rate as in cooling, so the systems transitioned back to the fluid phase. The transition temperature was determined only on the heating step since it would be underestimated by a large amount in the cooling step due to the need to nucleate the ordered gel phase.

There are several quantities which might be used to determine the transition temperature. For a first order transition, there is a maximum in heat capacity at the transition temperature which also corresponds to a change in the slope of the enthalpy vs. temperature. However, no observable change in the slope of the enthalpy could be observed even for pure DPPC. This may indicate that the enthalpy change for the transition is small compared to the noise, or likely it is because the pure DPPC gel phase is defective or is a ripple phase. Despite trying various procedures, we could not obtain a defect free gel phase. This was also observed by Coppock and Kindt [86]. We tried using the bilayer thickness, area per lipid, and the fraction of trans dihedrals in the lipid tails as used in the coexistence method. These quantities did show a slope change during the transition, but it was very small leading to inaccuracy in determining the transition temperature. We found that the deuterium order parameters for the lipid tail atoms was a better measure and gave the clearest indication of the transition.

The deuterium order parameter, S_{CD} (i), at various positions of ethylene groups along the hydrocarbon lipid tail is given by

$$S_{CD}(i) = \langle \frac{3\cos^2\theta - 1}{2} \rangle \quad (4.9)$$

where, θ is the angle between the corresponding C-H bond vector and the bilayer normal. We used deuterium order parameters which were normalized and then averaged over carbon atoms 5-8 (see Figure 1 for carbon atom numbering) located in the middle of the lipid tails:

$$S_{CD}^{norm} = \langle \frac{S_{CD}(i) - \overline{S_{CD}^{liq}(i)}}{\overline{S_{CD}^{gel}(i)} - \overline{S_{CD}^{liq}(i)}} \rangle_i, \quad (4.10)$$

where, $\overline{S_{CD}^{gel}(i)}$ and $\overline{S_{CD}^{liq}(i)}$ indicate time averages obtained in the gel and liquid phases, and $\langle\ \rangle_i$ indicates an average over carbon atoms. S_{CD}^{norm} varies from around 1 in the gel phase to around 0 in the fluid phase and shows a change in slope during the transition. The transition temperature is taken as the temperature corresponding to the center of this transition region.

4.5. Result and discussion

4.5.1. Positioning of mono and dimer lignols in DPPC bilayers

Trajectories over time

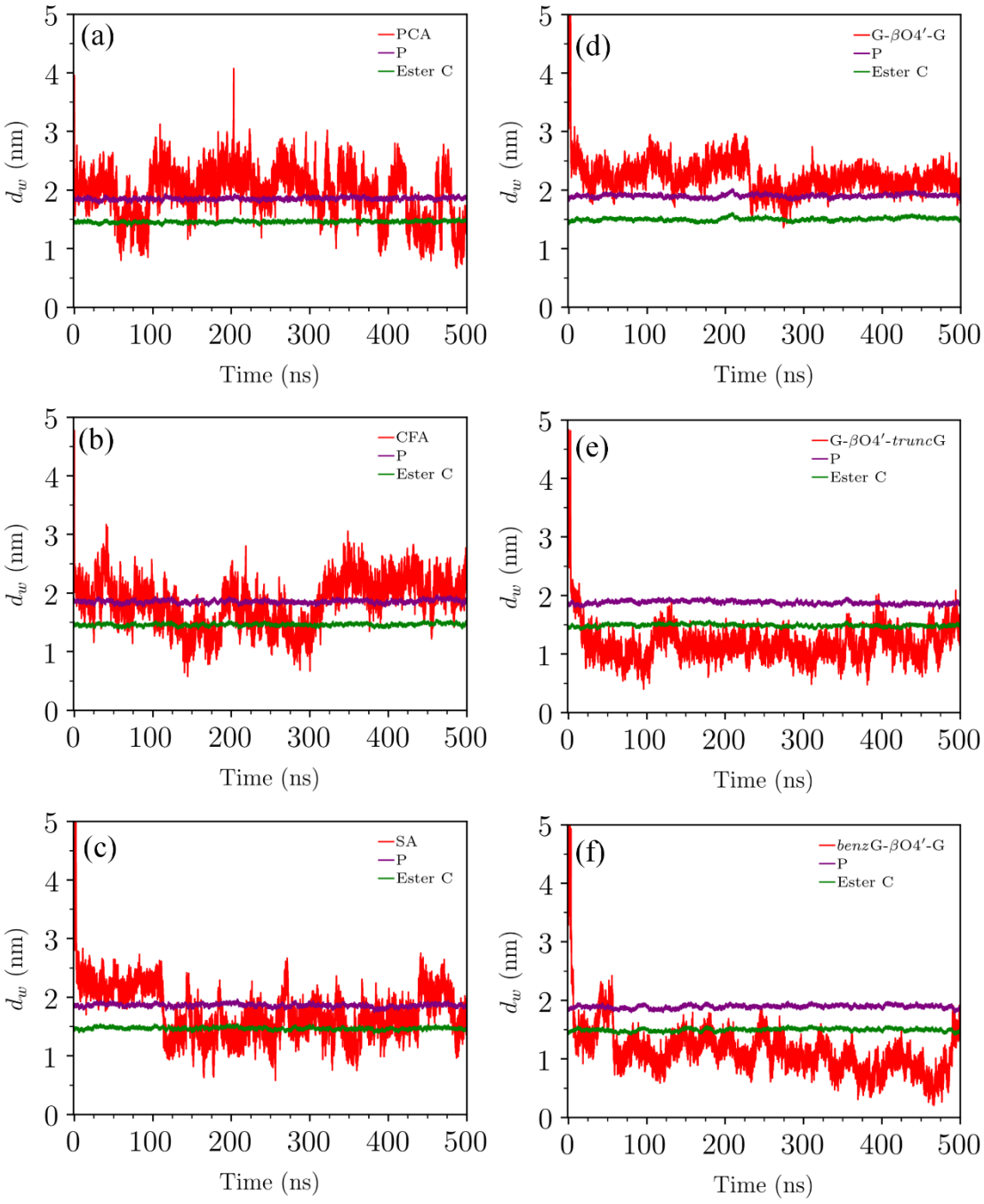


Figure 4.5. Trajectories of monolignols (a) PCA. (b)CFA. (c) SA, and dimers (d) G-βO4'-G. (e) G-βO4'-truncG. (f) benzG-βO4'-G. from the center mass to DPPC lipid bilayer center. 'P' and 'Ester C' represent lipid phosphate headgroups and ester carbons respectively.

Partial density profile of dimer groups

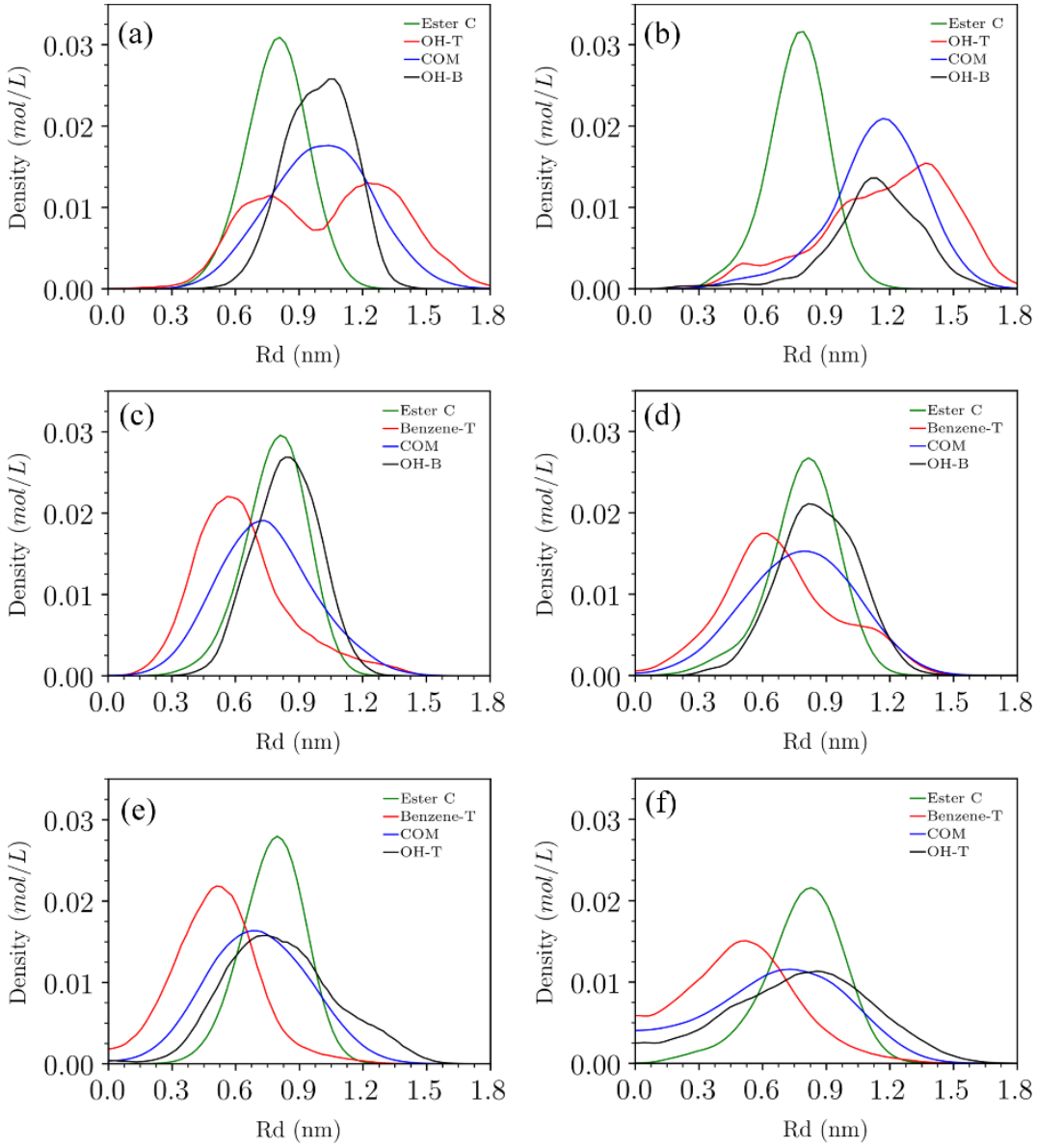


Figure 4.6. Partial density profile of certain groups as a function of relative position (Rd) from lipid bilayer center.Mole fraction (left)X=0.158, (right) X= 0.289 for all dimers: (a-b) G-βO4'-G, (c-d) G-βO4'-truncG, (e-f) benzG-βO4'-G. Ester C refer to carbon in C=O group on lipid heads. COM refer to the center mass of lignin. Benzene-T stands for one group of lignin, which was represented by one oxygen atom from the hydrophilic end of lignin. OH-T stands for hydroxyl group on the hydrophilic end which contains the C=C group. OH-B Stands for the hydroxyl group connected with (caption cont'd)

benzene ring. Benzene and hydroxyl groups are circled out in Figure 4.2 indicating in the same color as in density profile.

From positioning and PMF for we can see that lignin dimers should interact stronger than monolignols with lipid bilayers. G-βO4'-*trunc*G and *benz*G-βO4'-G are both more hydrophobic than the regular dimer and therefore would have higher affinity for bilayer interior. In order to explore how concentrations will affect the bilayer, studies involved with different numbers of dimers for the system are also presented. Partial density profile is a straightforward way of depicting how diversely the groups of lignin dimers oriented with lipid bilayer. The whole molecule of G-βO4'-G did not go deep to the lipid as the density distribution of COM, OH-T, and OH-B did not surpass ester carbon on the left-hand bound. G-βO4'-*trunc*G and *benz*G-βO4'-G, benzene terminated group is easy to access to the center of bilayer by its hydrophobicity. Hydroxyl terminated group, on the other hand, points more towards water phase and the preferable positions are close to lipid surface, i.e., where lipid ester carbon locates. *benz*G-βO4'-G are larger molecules and they spread more in lipid bilayer thus the width of density distribution is larger than the other two as well. It also showed much more occupation in the center of lipid bilayer than G-βO4'-*trunc*G.

4.5.2. Potential of mean Force

The MD calculations performed as part of this work allow us to gain greater insight not just into the general bilayer penetrating ability of dimers, but also where they prefer to reside and how their interactions alter the structure of the bilayers. The potential of mean force (PMF) curves as a function of the distance from the center of mass of single lignin dimers to the center of DPPC bilayers are shown in Figure 4.7.

Diluted system

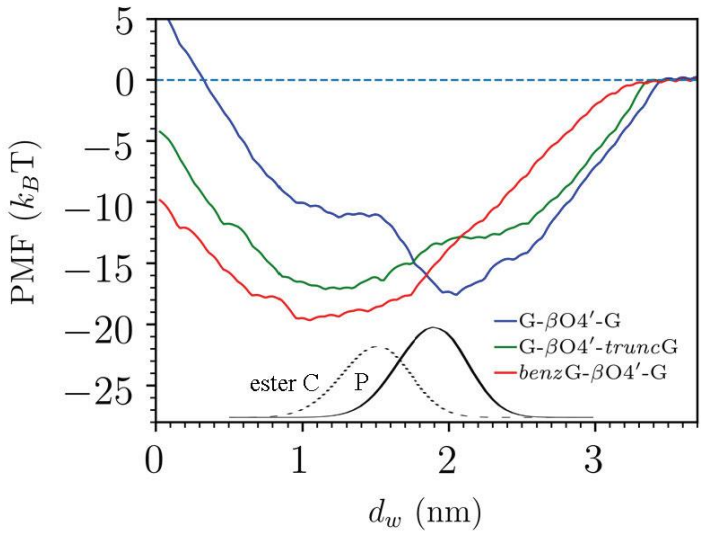


Figure 4.7. PMF curves for each of the three types of individual lignin dimer as a function of the distance from the dimer center of mass to the center of the lipid bilayer, d_w. The reference state is the dimer in bulk water, where the PMF is set to zero. For reference, the black solid and dashed lines trace the density distributions (scale not shown) for lipid phosphorus atoms (P) and lipid carbonyl carbon atoms in the ester groups (ester C).

The minimum in the G-βO4'-G curve is slightly outside the average position of the lipid phosphorus atoms indicating that most of the natural lignin dimer resides at or near the exterior bilayer surface. There is a shoulder in the G-βO4'-G curve at distances between about 1.0 and 1.5 nm indicating that there is a small probability of finding the dimer inside the bilayer. For the G-βO4'-*trunc*G curve, the minimum is inside the bilayer at around 1.2 nm. Therefore, most of the G-βO4'-*trunc*G resides inside the bilayer. There is a shoulder in the G-βO4'-*trunc*G curve at distances between about 1.9 and 2.3 nm indicating that there is a small probability of finding the truncated dimers at the exterior bilayer surfaces. The minimum in the *benz*G-βO4'-G curve is around 1.0 nm which is even deeper in the bilayer than the minimum for G-βO4'-*trunc*G. The minimum is also more negative for *benz*G-βO4'-G than for G-βO4'-*trunc*G, about -20 k_BT

compared to about -17 k_BT. There is also no shoulder in the *benz*G-βO4'-G at the bilayer surface, showing that it is expected to reside almost exclusively in the bilayer interior. These PMF results are qualitatively consistent with the experimental transition temperature shifts. Little of the natural G-βO4'-G dimer goes into the bilayer which leads to only small transition temperature shifts, while much more of the more hydrophobic G-βO4'-*trunc*G and *benz*G-βO4'-G dimers go into the bilayer leading to larger transition temperature shifts. Even though PMFs provide insight into lignin dimer-DPPC interactions, they are possibly not that accurate due to incomplete convergence issues and systematic sampling errors. Larger solutes require more sampling or additional reaction coordinates to force more adequate sampling of conformations and orientations [111].

High concentration of dimers

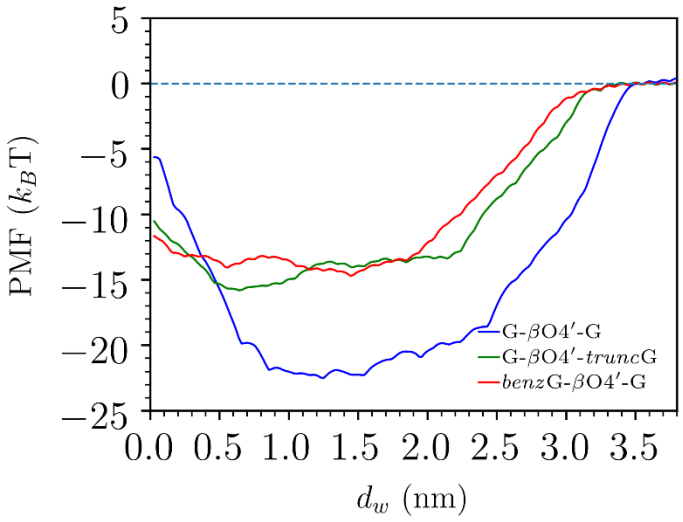


Figure 4.8. PMFs profile for a lignin dimer as a function of the distance (d_w) from the center of mass to bilayer center. The mole fraction of dimers in the system is X=0.289 in all environments.

Compare X_{dimers} = 0.289 with the diluted system, PMFs minima became shallower for G-βO4'-*trunc*G and *benz*G-βO4'-G but deeper for G-βO4'-G dimers. G-βO4'-G dimer did not go deeper to bilayers after it reached a concentration of X=0.289. Also, its energy barrier of diffusing

into bulk water is even higher. It possibly that G-βO4'-G are attracted by its own species who are already strongly attached to lipid surface. The energy cost from PMF minima to the center of bilayer dropped significantly (more than 50%) in both $X_{G\text{-}\beta O4'\text{-}truncG} = 0.289$ and $X_{benzG\text{-}\beta O4'\text{-}G} = 0.289$. This indicates that both modified dimers went deeper in bilayer at a high concentration which is consistent with density profile in Figure 4.6. From their PMFs, flipping seems unlikely to happen at high concertation due to a high energy barrier. We tested on system of $X_{G\text{-}\beta O4'\text{-}truncG} = 0.384$ and $X_{G\text{-}\beta O4'\text{-}truncG} = 0.458$ to evaluate this question. The results turned out to be dimers were observed stacking instead of going deeper into bilayer (Figure 4.9), which is a hint of saturation of the lipid bilayer. On the other hand, using center of mass to calculate energy of crossing bilayer for *benz*G-βO4'-G is not an accurate measurement. Even though the center of mass showed only $3\ k_B T$ to go through bilayer center, it doesn't mean the whole molecule can make it. Because rotation energy of this particle due to the hydroxyl group in its tail is unknown, which can be pretty large. Therefore, we did an extra PMF calculation (see APPENDIX B) of $X_{benzG\text{-}\beta O4'\text{-}G} = 0.289$ based on distance from COM of the hydroxyl group attached to benzene ring to the bilayer center. It was noticed that the PMF minima located much closer to water phase and the flipping energy is about twice as big. From here we can estimate that it will not be easy for *benz*G-βO4'-G to have a translipid movement.

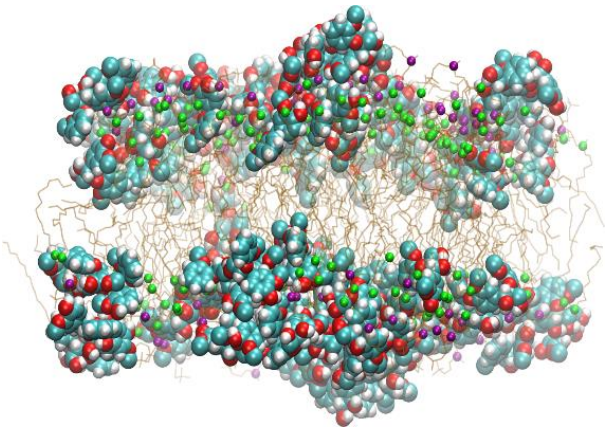


Figure 4.9. A snapshot of DPPC bilayer with $X_{G\text{-}\beta O4'\text{-}truncG}$= 0.458.

4.5.3. Bilayer thickness, area and order parameter

As an increase of benzG-βO4'-G or G-βO4'-*trunc*G dimer concentration, lipids bilayer thickness has decreased. As more dimers insert themselves into spaces between lipids head group, they expanded bilayer horizontally. Accordingly, contraction happened in vertical direction since bilayer volume was about to stay the same. As noticed, G-βO4'-G dimer did not follow the trend**.** Thickness had to be through increased and then decreased as more G-βO4'-G presented in system. This is consistent with results in PMFs, density profile and order parameter (see more in APPENDIX B). G-βO4'-G dimer had reached its saturation in lipids at lower concentrations as they bind mostly steady to bilayer surface.

Bilayer thickness or area compressibility

Table 4.1. Thickness of bilayer with different concentrations of dimer. The pure DPPC bilayer thickness is 3.733± 0.126 nm

Dimer concentration (mole fraction)	DPPC bilayer P-P distance (nm)		
	G-βO4'-G	G-βO4'-*trunc*G	*benz*G-βO4'-G
0.086	3.695 ± 0.109	3.727 ± 0.021	3.693 ± 0.036
0.158	3.589 ± 0.145	3.606 ± 0.061	3.604 ± 0.017
0.289	3.699 ± 0.007	3.463 ± 0.068	3.402 ± 0.126

Table 4.2. Area compressibility K_A (N/m) and thickness compressibility K_D(10^{-9}N)

	Pure DPPC	DPPC + G-βO4'-G		DPPC+ G-βO4'-*trunc*G		DPPC+ *benz*G-βO4'-G	
X_{dimer}	0	0.158	0.289	0.158	0.289	0.158	0.289
K_A (N/m)	1.98±0.46	0.35±0.1	1.21±0.31	0.49±0.2	0.46±0.1	0.68±0.0	0.39±0.1
K_D(10^{-9}	17.80±4.1	3.00±2.5	13.80±8.7	4.38±2.3	2.63±1.9	6.78±3.3	5.32±3.5
Mass/Area	390	41	450	39	39	38	38

Area and thickness compressibility were derived from averaging over 300ns simulation to the extent of area and thickness fluctuation. Of all these values K_A and K_D are essentially largest for pure DPPC. K_A and K_D decreasing suggested a more rigid bilayer with X_{dimer} =0.158 and 0.289 for G-βO4'-*trunc*G and *benz*G-βO4'-G dimer. Under the condition of most dimers were inside lipid bilayer, at low mole fractions, a stress relief of lipids compression were attributed to more dimers orientational freedom. But at higher concentrations, lipid compression has to subsume a compression of dimer molecules who were not capable of moving around freely [112]. Lipids interacted with G-βO4'-G dimers have a reversed behavior similar to previous measurements. At

X_{dimer} =0.158, G-βO4'-G dimers penetrate deeper inside bilayer while at X_{dimer} =0.289 it had less interactions with bilayer interior.

Area per lipid probability density distribution

Prior to performing transition temperature calculations with MD simulations, we looked at the effect of the high concentrations of dimers on the structure of the bilayer in the fluid phase at 326 K. Increasing area per lipid is correlated with decreasing transition temperature. If lipids are more widely spaced in a fluid bilayer and therefore their tails are more disordered, then it requires a lower temperature to get them to transition to a gel phase with more ordered tails.

The FATSLIM program [113] was employed to perform an area per lipid (APL) calculation. FATSLIM gives an area for each lipid, so a distribution can be obtained and not just a mean value. The areas are obtained from a Voronoi tessellation of the projection of the lipid positions onto a plane which is perpendicular to the bilayer normal. The lipid position was defined by its headgroup, and the cutoff distance for neighbor searching was 3.0 nm.

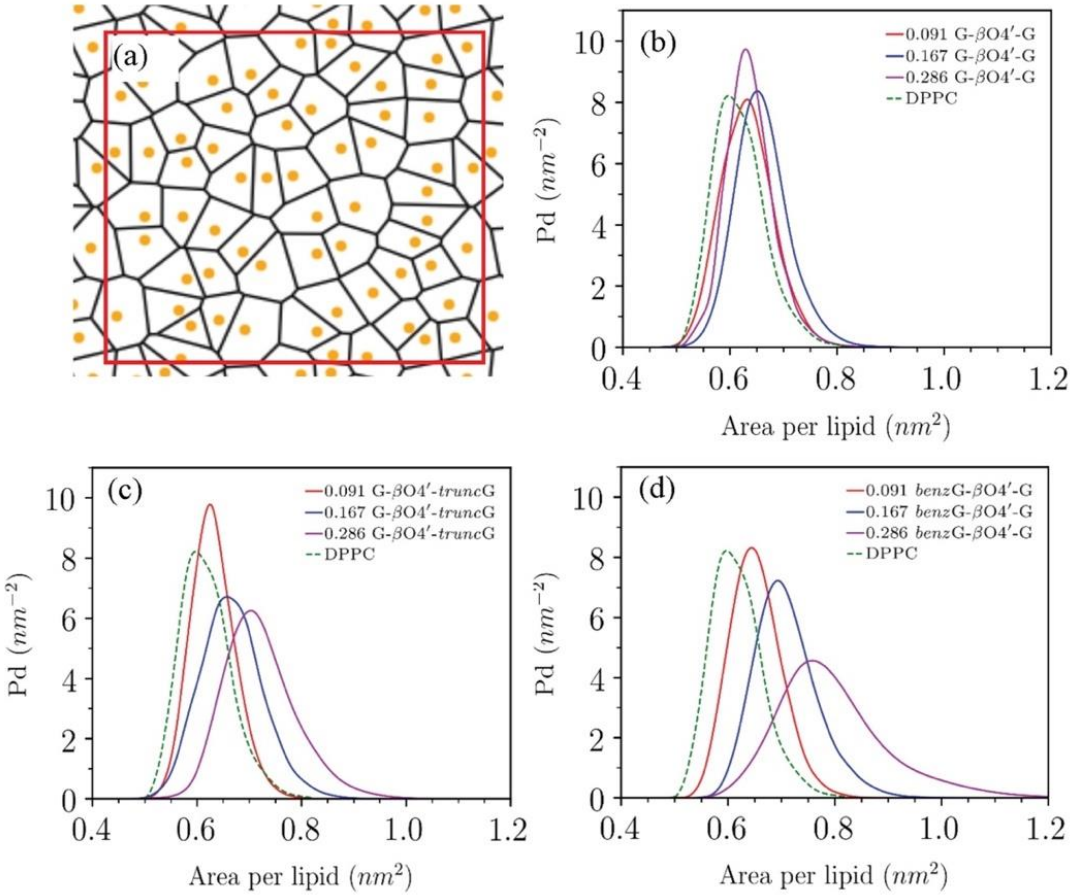


Figure 4.10. (a) Voronoi tessellation depicting the area distribution per lipid ; orange dots indicate the position of the center of mass of lipid heads. Area per lipid probability density distributions (Pd) for DPPC bilayers and DPPC bilayers with added lignin dimers: (b) G-βO4'-G dimers, (c) G-βO4'-*trunc*G dimers, (d) *benz*G-βO4'-G dimers. Numbers in the legends indicate mole fraction of lignin dimer relative to DPPC lipids.

Figure 4.10 shows the APL probability density (Pd) distributions for DPPC and for DPPC with various concentrations of lignin dimers. The distributions for DPPC with G-βO4'-G in Figure 4.10(a) are all similar to the distribution for pure DPPC, regardless of the concentration. This is consistent with experimental results showing that G-βO4'-G does not shift the gel-fluid transition temperature much, even at high concentrations. The distributions for DPPC with G-βO4'-*trunc*G or *benz*G-βO4'-G in Figure 4.10(b) and (c) show clear increases in both the median APL and variance of the APL with increasing concentration. Consistent with the log(K) results from

simulations, *benz*G-βO4'-G has the largest effect on APL at a given concentration due to its deep penetration into the lipid bilayers.

Deuterium order parameter

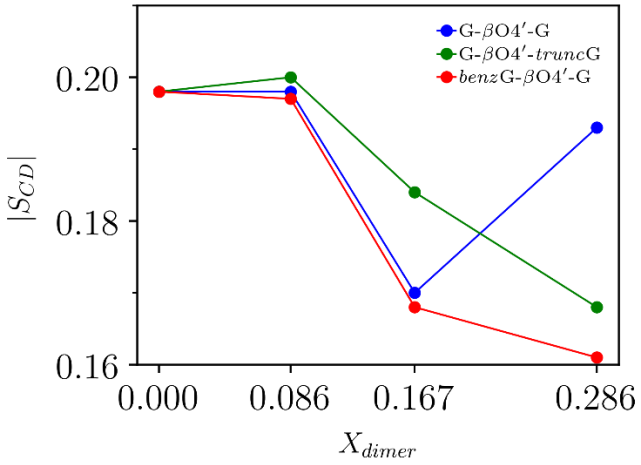


Figure 4.11. Order parameter of lipids carbon 7 with dimers at four different mole fraction(X_{dimer}).

The averaged order parameter has been calculated for three dimers at X_{dimer} = 0, 0.086,0.167 and 0.286 and a simplified summary is shown in Figure 4.11(more details in Appendix B). Pure lipids with no or a low percentage of dimers showed a high ordering of acyl Chains. As dimers started interacting with atoms in lipid tails, it was expected a diminution of order parameter. As we see, there showed a trend of reduction of order parameter as an increase of G-βO4'-*trunc*G and *benz*G-βO4'-G dimers concentration**.** Distinct from these two, $X_{G\text{-}\beta O4'\text{-}G}$ = 0.286 increased order parameter back to a larger value similar as from pure DPPC. From PMFs(Figure 4.8) and density profile(Figure 4.6), G-βO4'-*trunc*G had been sucked out some from $X_{\mathbf{G\text{-}\beta O4'\text{-}G}}$ = 0.167 to $X_{G\text{-}\beta O4'\text{-}G}$ = 0.286 and then comfortably stayed near bilayer surface, which was about 0.45nm away to lipids ester carbon. Therefore, G-βO4'-G dimers have less impact on lipids structures. This

property is also corresponding to the phenomenon of that *benz*G-βO4'-G or G-βO4'-*trunc*G dimer increased lipid bilayer fluidity while G-βO4'-G barely affected.

4.5.4. Partition coefficient

Table 4.3 compares the log of the partition coefficients (log K) calculated using the PMF curves in Figure 4.7 with eq 4 using the average lipid phosphorus atom positions as the boundary between lipid and water phases with the experimental values estimated from T_m shifts. Experimental details are shown in the previously published paper [114].The simulation results show a similar trend compared to the experimental results in that the more hydrophobic G-βO4'-*trunc*G and *benz*G-βO4'-G have larger log K values than the more hydrophilic G-βO4'-G. However, the simulations predict larger differences in log K for the different dimers compared to experiment. Simulation results for log K range from -1.42 for *benz*G-βO4'-G to 5.21 for *benz*G-βO4'-G while experimental results range from 2.72 to 3.45 for the same dimers.

Table 4.3. Comparison between log(K) from MD and experiment.

Dimer	G-βO4'-G	G-βO4'-*trunc*G	*benz*G-βO4'-G
Experiment (DSC)	2.72 ± 0.05	3.45 ± 0.20	3.38 ± 0.28
Simulation (MD)	-1.42 ± 0.13	3.83 ± 0.29	5.21 ± 0.21

The negative value for G-βO4'-G just means that more of the molecules are found outside the phosphorus atom boundary than inside. Since most of G-βO4'-G is on the bilayer surface, choosing the boundary between phases somewhat outside the phosphorus atoms of the lipids would change the calculated log K significantly. Since this approach to comparison with experiment is only qualitative, we did not try to determine a way to define the boundary between phases more

accurately. The MD simulations also predict that *benz*G-βO4'-G partitions more strongly into the bilayer than G-βO4'-*trunc*G while experimental log K values of the two dimers are about the same.

4.5.5. Effect of lignin dimers on gel to liquid-crystalline phase transition temperature in DPPC bilayers

Coexistence method

The coexistence method was used for pure DPPC as a reference for comparison to the heating method. After creating a system consisting of strips of gel and fluid phase bilayer, we first guessed a transition temperature of 308 K and performed a 200 ns NPT simulation at that temperature. We saw noticeable growth in the fraction of gel phase at 308 K, so this temperature is lower than the transition temperature. We then simulated at a series of temperatures: 312 K, 320 K, 316 K, 314 K, and 313 K; which allowed us to narrow down the transition temperature of pure DPPC to between 312 K and 313 K which is quite close to the experimental main transition of 315.5 K found here, or of 314 K reported by Biltonen and Lichtenberg. [115]

Heating method

The heating method was used for both pure DPPC and for DPPC with lignin dimers. These simulations were not performed for G-βO4'-G due to the small transition temperature shifts seen in experiments, the fact that it is found on the surface of the bilayer in simulations, and the small effect it has on area per lipid in simulations (Figure 4.10). Transition temperatures were obtained for bilayers containing G-βO4'-*trunc*G and *benz*G-βO4'-G.

Deuterium order parameter

Since the normalized deuterium order parameters averaged over several carbon atoms in the lipid tails is used as the indicator of the phase transition, we first show the deuterium order

parameters for each system in both the fluid and gel phases in Figure 4.12. As expected, the difference between the fluid and gel phases for pure DPPC is larger than for the systems with added lignin dimers. The biggest effect is in the gel phase; the lignin dimers do not allow the bilayer tails to become as ordered as in pure DPPC.

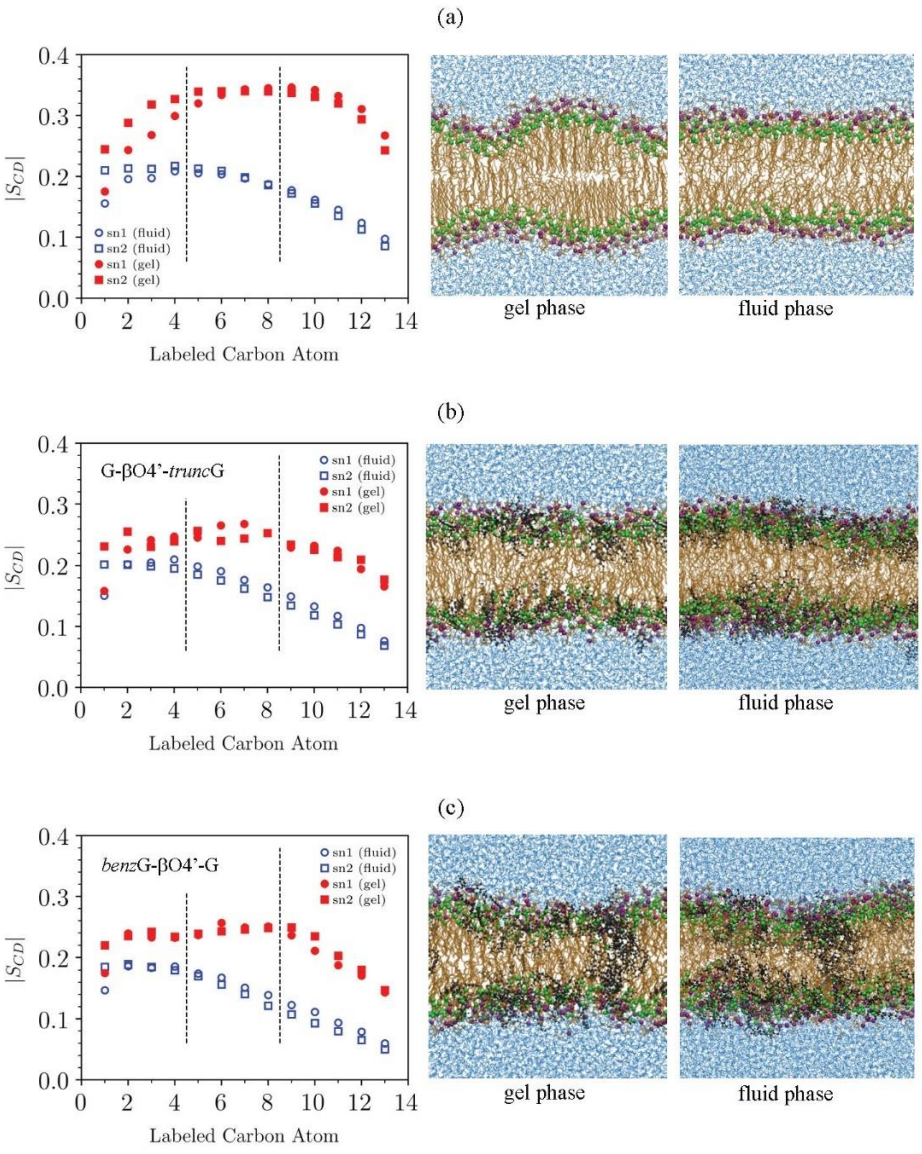


Figure 4.12. Deuterium order parameter, $|S_{CD}|$, for lipid tail sn1 and sn2 chains as function of carbon atom index in liquid (326 K) and gel phase (250 K). Lower index

(caption cont'd)

atoms are closer to the lipid heads. The vertical dashed lines indicate the range of atoms which were used to obtain a normalized and averaged order parameter used to follow the phase transition. (a) 128 DPPC lipids. (b) 512 DPPC lipids with G-βO4'-truncG dimers of concentration X_{dimer} = 0.289. (c) 512 DPPC lipids with benzG-βO4'-G dimers of concentration X_{dimer} = 0.289. Figures to the right of each plot were obtained from snapshots of the gel phase and liquid phase. In those pictures, lipids are represented in orange except for the ester carbon atoms (green) and phosphorus atoms (purple), lignin molecules are black, and water is blue (not all shown).

Metastable gel phase bilayers containing lignin dimers

Due to the high cooling rate, the gel phase structures containing lignin dimers may be in metastable states where lignin is trapped in the bilayer while in the stable state the lignin may be largely expelled. Adding lignin to the outside of a pure DPPC gel phase followed by equilibration shows that the lignin dimers stay primarily on the surface of the gel phase bilayer. Although inaccurate at the low temperature of 250 K, PMF calculations also indicate that there should be a very small concentration of dimers inside the bilayer. Therefore, for calculation of transition temperature we considered starting with the lignin trapped in the gel phase or with the lignin on the surface of the bilayer since the latter is likely a better approximation. Note that we are still ignoring the partitioning of lignin dimers to the water phase. Considering this would be expensive since instead of heating it would require doing several simulations at constant temperature with different dimer concentrations and those dimer concentrations would have to be first determined by PMF calculations which are difficult to obtain accurately, especially in the gel phase. We believe that starting with the lignin dimers outside the bilayer is a reasonable approximation since they don't affect the bilayer structure much when on the surface (see plots of deuterium order parameters in APPENDIX B) and will naturally partition into the bilayer interior as the temperature is increased.

Transition temperature

As seen in the snapshot in Figure 4.12(a), a pure DPPC lipid bilayer does not form a perfect gel phase even at 250 K. A defective DPPC gel phase was also observed in Leekumjorn's study [83]. Despite this defect, the normalized S_{CD} shows a sharp drop in Figure 4.13(a) and the center of this drop gives a transition temperature of 316.6 K. This is higher than the results from the coexistence method of 312.5 K, but only by about four degrees which is acceptable given the high heating rate.

With added lignin dimers, it is hard to visually determine from the snapshots in Figure 4.12 if the gel phase is formed but there is a clear increase in S_{CD} compared to the fluid phase seen in Figure 4.13(b) and Figure 4.13(c). The less ordered gel phase leads to a lower transition temperature and the transition occurs over a wider range of temperatures, and these effects are increased with increasing lignin dimer concentration. This can be seen in the normalized S_{CD} plots as a function of temperature Figure 4.13(b-e). For bilayers containing lignin dimers, both 128 and 512 lipid systems were considered. In some cases, determining the transition region was very difficult or impossible with only 128 lipids. Increasing the number of lipids also reduced the widths of the transition regions by about 10 K. Figure 4.13(b-e) shows only the results for the 512 lipid bilayer systems in which the lignin dimers were added to the solution outside the pure DPPC bilayers in the gel phase. The results for the 128 and 512 lipid systems with lignin added to DPPC bilayers while in the fluid phase are presented in APPENDIX B.

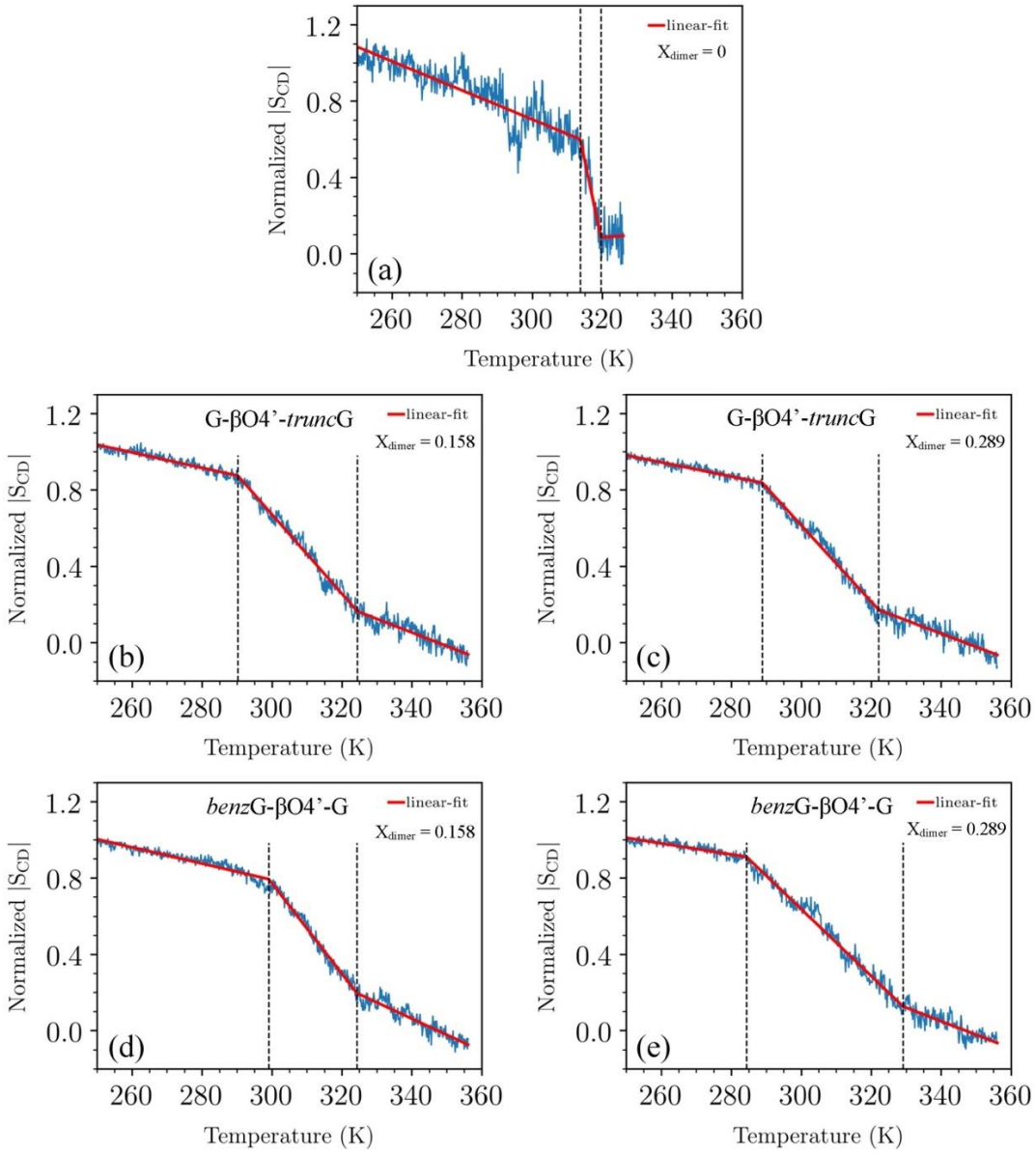


Figure 4.13. Normalized deuterium order parameter, $|S_{CD}|$, versus temperature and 3-piece linear fits. (a) 128 DPPC lipids. (b) and (c) 512 DPPC lipids with G-βO4'-truncG dimers of concentration X_{dimer} = 0.158 and 0.289 respectively, (d) and (e) 512 lipids with benzG-βO4'-G dimers of concentration X_{dimer} = 0.158 and 0.289 respectively. The vertical dashed lines indicate the turning points; the transition occurs between these lines and the transition temperature is determined to be midway between them. The results in (b) through (e) are for the 512 DPPC lipids systems in which the lignin dimers were added to the solution outside of pure DPPC bilayers in the gel phase.

Table 4.4 summarizes the transition temperatures and transition widths from MD simulations. The lignin dimers G-βO4'-*trunc*G and *benz*G-βO4'-G had about the same effect on the transition temperature. The shifts relative to pure DPPC for X_{dimer} = 0.158 of G-βO4'-*trunc*G and *benz*G-βO4'-G are about 9 K and 5 K, respectively when the lignin dimers are started outside the gel phase bilayer. This is nearly the same compared to the case of starting with the dimers inside bilayer where the shifts are both about 7 K. For low concentrations, trapping the lignin in the gel phase has little effect on the transition temperature. For X_{dimer} = 0.289 for G-βO4'-*trunc*G and *benz*G-βO4'-G, T_m shifts are about 11 K and 10 K, respectively when the lignin dimers are started outside the gel phase. This is significantly smaller than the case where the dimers start inside the bilayer where the shifts are about 24 K and 23 K, respectively. This shows that at high concentrations, the starting structure can be very important. Trapping the dimers in the bilayer lead to a much lower transition temperature since the lignin molecules create disorder in bilayer and lead to earlier onset of melting. The range of melting, ΔT_{trans} is 5 K and 8 K narrower for DPPC with X_{dimer} = 0.289 for G-βO4'-*trunc*G and *benz*G-βO4'-G, respectively when the lignin dimers are started outside the gel phase. The MD results cannot be compared directly to experimental results since experimental concentrations include all lignin in the system, while all lignin in MD systems is in the bilayer due to the small amount of water. The MD shifts for X_{dimer} = 0.158 are reasonable, but likely large compared with experiment. The MD shifts for X_{dimer} = 0.289 are comparable when started dimers inside DPPC but clearly too large from the opposite way. The MD results are qualitatively in line with experiments in that the effects of G-βO4'-*trunc*G and *benz*G-βO4'-G are similar, and the magnitude of the shifts increases with increasing concentration.

Atomistic MD simulations of lipid phase transitions are difficult due to both time and length scale limitations. It may be possible to improve results by starting with more accurate

concentrations determined from expensive PMF calculations in a series of constant temperature simulations as described above. Even with an extremely accurate calculation, it is still likely that the simulation parameters result in at least slightly different phase behavior than the real system. Although only qualitative in this case, the MD simulations still provide information about behavior at the molecular level that can be difficult to obtain from experiments.

Table 4.4. Transition temperatures (T_m) and transition widths (ΔT_{trans}) from MD simulations. N_{DPPC} is the number of DPPC lipid molecules. Transition widths are determined by temperature difference between turning points shown in Figure 4.13. The transition temperature is defined to be midway between the turning points T_0 and T_1 shown in Figure 4.13; $T_m = (T_1-T_0)/2$. Uncertainties were calculated by using two times the standard deviation from the mean value. (The details on error estimation procedure are given in Appendix B)

N_{DPPC}	256*	128	512			
X_{dimer}	0	0	0.158		0.289	
dimer type	none	none	G-βO4'-*trunc*G	*benz*G-βO4'-G	G-βO4'-*trunc*G	*benz*G-βO4'-G
Simulations starting with dimers inside DPPC gel phase						
T_m (K)	312.5*	316.6 ± 0.4	309.8 ± 1.8	309.3 ± 1.0	291.5 ± 1.6	292.2 ± 2.0
ΔT_{trans} (K)	--	6.0	30.0	35.1	38.6	52.3
Simulations starting with dimers outside DPPC gel phase						
T_m (K)	312.5*	316.6 ± 0.4	307.3 ± 0.4	311.6 ± 0.4	305.4 ± 0.4	306.7 ± 0.6
ΔT_{trans} (K)	--	6.0	34.2	25.3	33.2	44.8

* Coexistence method.

The lack of a difference in transition temperatures between bilayers with G-βO4'-*trunc*G and *benz*G-βO4'-G in the MD simulations is surprising given that *benz*G-βO4'-G is larger and partitions more into the bilayer interior. However, this is consistent with experimental observations of similar log K values from DSC for the two dimers. This behavior might be explained by the fact that *benz*G-βO4'-G clustered together more than G-βO4'-*trunc*G inside the bilayer which can be seen visually in in Figure 4.12 (b) and (c). Using all atoms in the molecules and a distance cutoff between atoms of 0.35 nm, the average cluster sizes for G-βO4'-*trunc*G and *benz*G-βO4'-G were 6 and 25, respectively. More clustering meant that there was more local disruption to the bilayer, but perhaps less overall disruption. Even though there was less G-βO4'-*trunc*G in the bilayer

interior, it spread out more which caused more disruption and lead to similar transition temperature shifts compared to *benz*G-βO4'-G.

4.5.6. Time evolution of an artificially-created pore

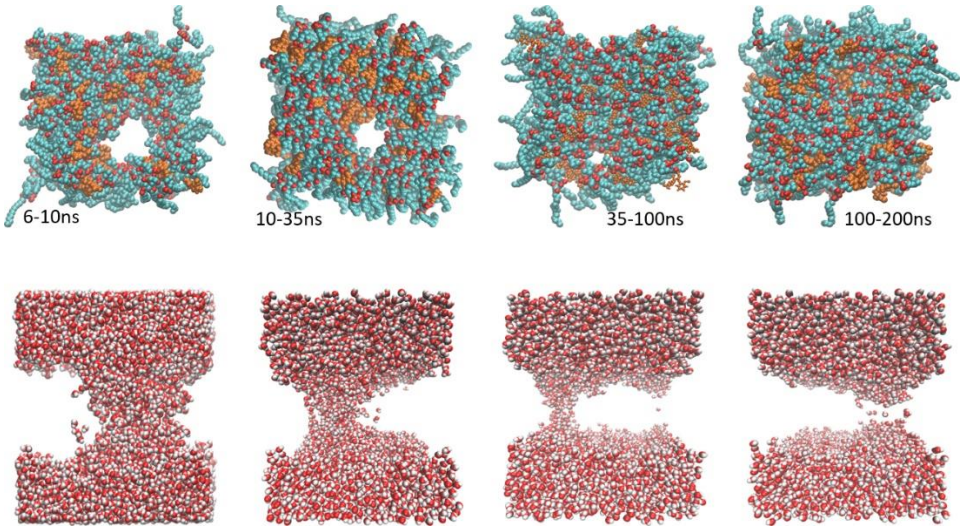


Figure 4.14. Time evolution of an artificially-created pore in a DPPC bilayer in the presence of G-βO4'-*trunc*G of concentration X_{dimer} = 0.289.

Transmembrane pore formation is vital in biological process includes cell signaling, small molecule transport, and so on. However, when a pore is caused by toxin lead to cell leakage could result in the death of a cell. Neha Awasthi [116] used three methods of umbrella sampling to calculate pore formation free energy which all turned to be efficient. Vahid Mirjalili [117] used an enhanced way which results in a more realistic pathway. However, in our simulation, it was not necessary to get an accurate result of free energy of pore formation. Based on the PMFs we speculate that it will not happen. Thus we did it in a reversed way that is we artificially created one pore and to see if it will close or not. Pore formation and closure have been simulated for G-βO4'-*trunc*G of concentration X_{dimer} = 0.289. It closes within 150ns at 326K and within 100ns at 355K. For lipid only system, it closes in 10ns. Investigation of the stability of an artificially created

pore in the bilayer indicates that, even at high concentrations, it is very unlikely that lignol dimers have the ability to mediate pore formation in bilayers.

4.8. Conclusions

Interactions between DPPC lipid bilayers and coniferyl alcohol-based lignin G-G dimers and their derivatives were studied using DSC experiments and MD simulations. Specifically, the systems studied consist of fully hydrated DPPC lipid bilayers and liposomes in the presence of different concentrations of lignin dimers for DSC or MD, respectively. Experimental liposome-water partition coefficients for the lignin dimers were obtained through a relationship between the shift in the gel-fluid transition temperature relative to pure DPPC liposomes obtained from DSC measurement. In MD simulations, partition coefficients were obtained using PMF calculations with a single dimer and a 128 DPPC bilayer. Transition temperatures in MD were estimated by first cooling the system to induce transition to the gel phase followed by heating to change back to the fluid phase. During the heating, we captured the transition from gel to fluid phase by defining and analyzing a normalized deuterium order parameter averaged over all carbon atoms located in the middle of the lipid tails. We further analyze the MD trajectories and study the location of the lignols and their effects on bilayer structure to explain the observed trends.

A commercially available truncated G-G dimer with the hydroxyproenyl tail removed, G-βO4'-*trunc*G, and G-G dimers with an added benzyl group, *benz*G-βO4'-G, show a higher affinity for the bilayer interior than natural dimers. DSC transition temperature shifts for the G-βO4'-G natural dimer were very small and consistent with that result, MD simulations showed that most of this dimer stays on the surface of the bilayer and has little effect on the lipid tail structure or area per lipid. The partition coefficient for the natural dimer was smallest from both DSC (2.72) and MD (log K = -1.42). The partition coefficients obtained using DSC measurements were similar

for G-β4'-*trunc*G (log K = 3.45) and *benz*G-βO4'-G (log K = 3.38), while MD showed a higher partition coefficient for *benz*G-βO4'-G (log K = 5.21 compared to log K = 3.83) which is the most hydrophobic dimer considered.

Both G-βO4'-*trunc*G and *benz*G-βO4'-G dimers induced larger shifts in the gel-fluid phase transition in DSC compared to the natural dimer, and MD simulations for the transition temperature with natural dimer were not attempted due to the small shifts seen in DSC. The MD results show a very clear change in the deuterium order parameter associated with the transition for a pure DPPC bilayer with T_m =316.5 K which is very close to the experimental value of 315.5 K. However, once hydrophobic lignin dimer derivatives are added to the system, the signal becomes noisier and wider consistent with disorder caused by the lignin dimers and a coexistence region in the phase diagram, respectively. It is important to note that with experimental systems, only the total amount of dimer added to the experimental system is known, so that the amount that partitions into the bilayer will be less than the nominal amount added. In MD all dimer was taken up by the bilayers, therefore the transition temperature can give just a qualitative estimate. In the MD simulations, two types of starting configurations for the heating step were considered. In the first approach, the fluid phase with lignin was cooled down to obtain a gel phase. The potential problem with this is that it may trap the lignin inside the bilayer leading to a metastable state. To alleviate this potential problem, the second type of configuration is obtained by adding lignin dimers outside a pure DPPC gel phase followed by equilibration before heating. This second type of starting configuration leads to significantly lower transition temperature shifts which are much closer to DSC results for high lignin concentrations, but gives little difference compared to the first type of starting configuration at lower concentrations. MD transition temperature shifts were similar for G-βO4'-*trunc*G and *benz*G-βO4'-G, while DSC results predict slightly larger shifts for

*benz*G-βO4’-G at high concentrations. The lack of difference in transition temperature shifts in MD simulations is surprising given that the MD calculated partition coefficient for *benz*G-βO4’-G was significantly larger. This may be due to increased clustering *benz*G-βO4’-G in the bilayer relative to G-βO4’-*trunc*G leading to a reduced impact of *benz*G-βO4’-G on the overall bilayer structure.

Overall, this study shows that chemical modification of lignin dimers can have significant impacts on their interaction with model lipid bilayers. The comparison with experiments shows that these effects can be qualitatively captured using molecular dynamics simulations of lignin dimers at equilibrium and during heating of model lipid bilayers. MD simulations also provide details about the behavior of the molecules which is difficult to obtain from experiment. These results pave the way for using MD to screen other lignin oligomers and their derivatives with respect to their interactions with lipid bilayers.

CHAPTER 5. LIGNIN OLIGOMERS IN DEEP EUTECTIC SOLVENTS IN THE PRESENCE OF A MICROWAVE FIELD

5.1. Introduction

5.1.1. Lignin classification and processing

The detection of lignin was firstly realized by Runge when spruce wood turned greenish blue after treatment with certain reagents. And now more color reactions have been proposed [118]. The determination of lignin mainly has direct and indrect methods. Direct method involves acid hydrolysis and lignin isolation, which measures acid insoluble residues. The indirect method uses spectrophometric which is suitable for determining lignin concentrations. The typical adsorbance at 281nm follows Beer's Law for cold dioxane-hydrochloric acid lignin [119]. Experimental techniques such as nuclear magnetic resonance spectroscopy(NMR), FTIR and UV−vis spectroscopy [120], infrared (IR) and X-ray crystallography [121] are helpful for understanding three dimensional lignin structures. Kinetics behavior of lignin decomposition is revealed by thermal analysis such as DGT, thermos-gravimetric analysis (TGA) and Differential Scanning Calorimetry(DSC) [122]. Despite the abundant lignin resources and promising applications in various field, this biomass byproducts are highly wasted. It’s estimated that about 50 million tones is produced in paper industry but 98% of them were burnt as low quality fuel [123]. This underused situation is mainly caused by lignin’s structural complexity [124].

For different species of plant, lignins are commonly divides into softwood, hardwood and grass lignins based on varied content of monoligonols [125]. Softwood lignin compromises mostly

---

This chapter was previously partly published as Muley, Pranjali D., Justin K. Mobley, Xinjie Tong, Brian Novak, Joseph Stevens, Dorel Moldovan, Jian Shi, and Dorin Boldor. "Rapid microwave-assisted biomass delignification and lignin depolymerization in deep eutectic solvents." *Energy Conversion and Management* 196 (2019): 1080-1088.

G type which is over 95% of total amount of units; hardwoods lignin compromises mainly G and S type with various ratios; grass lignins also have G and S type as in hardwood, but differently, it contains p-coumaric acid and ferulic acid residues. Low level H type unit is contained for all above [118, 121, 126]. However, concerns from the differences are more from industrial processing when lignin structure has an inevitable change from the original source. They can be sorted into two main group: sulfur bearing lignin and sulfur free lignin [127, 128]. Sulfur bearing lignin includes Kraft lignin, Sulphite lignin, steam explosion lignin. Kraft processing is is the most prevailing method in pulping industry. Delignification takes place in a condition of high temperature in sodium hydroxide and sodium sulfide solution. Hydrolysed lignin molecular weight ranges from 150 to 200000 dissolved in the alkaline solution will be recovered by acidification [129, 130].

5.1.2. Lignin depolymerization

Many research works related to biosynthesis or pyrolysis mechanism of lignin have been done in recent years [131, 132]. Biosynthesis hypotheses primarily include polymerization and transport mechanism, but it remains not that clear. Lignin precipitation on cellulose hinder the hydrolytic enzymes from accessing to cellulose contributes to biomass recalcitrance to hydrolysis [133]. Lignin has a random three-dimensional network. It is not crystalline ordered; severely milled lignin dissolves more in solvent; lignin macromolecules are highly compact compared with linear polymer [134]. Chemical and enzymatic conversion included processes such as pyrolysis, hydrotreating, gasification and hydrolysis [135-137]. Ionic liquids (ILs) and organosolv process have been widely used because they are efficient in delignification. But they are expensive and not environmentally friendly [137, 138]. The deep eutectic solvent (DES), a homogeneous mixture of two or more solid chemicals which melting point is lower that its independent component, have received an increasing attention. DES is composed of hydrogen-bond donors or hydrogen-bond

acceptors with low volatility, low toxicity and low cost [139]. Recent study shows that DESs has high selectivity for the separation of lignocellulose product [140]and microwave-assisted DESs will accelerate fractionation reactions [141].

5.1.3. Aim of the present work

We performed molecular dynamics (MD) simulations on three lignin G tetramers in a deep eutectic solvent (DES) composed of choline chloride and oxalic acid (1:1 ratio) with and without a microwave field. Since the lignin force used in our simulations is of non-reactive type, we could not directly investigate the potential breaking of bonds. However, the MD simulations allowed us to investigate how the transfer of the microwave energy to kinetic and internal (e.g., inter-atomic bond stretching) energies of the molecules affects the dynamical and structural characteristics, such as bond length distributions, of the system. Considering the correlation between bond stretching and the probability of bond breaking at high temperature we infer that the bonds which are stretched more in the presence of the electric field are more likely to break.

5.2. Methodology

5.2.1. Pure DES under microwave

We simulated pure DES (1:1 choline chloride: oxalic acid) using the OPLS force field [142, 143] under alternating electric fields with strengths of E = 0.0, 0.05, 0.075 and 0.1 V/nm and a frequency of 2.45 GHz. The electric field was applied in the x direction. The short-range Coulomb and van der Waals cutoff was 1.6 nm, and particle mesh Ewald (PME) was used for long range electrostatics. NVE simulations were run for the cases with electric field for 10 ns using double precision and a time step of 0.5 fs with GROMACS 2018.1 [21], allowing us to measure the rate of energy addition for simulations equilibrated at 423.15 K. A 100 ns long NVT simulation at

423.15 K was run for the case of no electric field using a Berendsen thermostat [26] with a time constant of 0.1 ps.

5.2.2. Lignin tetramers with DES under microwave

All simulations were performed with the GROMACS 2018.1 package [21]. The interactions of particles were described by the all atom CHARMM36 force field [54]. Force field parameters and topology files for choline and oxalic acid were generated from SwissParam [144]. Lignin tetramer models and topologies were created via LigninBuilder [145]. The systems were composed of 500 choline ions, 500 chloride ions, 500 oxalic acid molecules, and one lignin G tetramer. Initial configurations were built using PACKMOL [146]. A simulation system is shown inFigure 5.1. Bond constraints were applied only to bonds involving hydrogens using LINCS [28] and a 1.8 fs time step was used so that we could obtain carbon-oxygen and carbon-carbon bond length distributions. The cutoff for van der Waals and short range Coulomb interactions was 1.2 nm. A switching function was applied to the van der Waals potential to change it smoothly zero between 1.0 and 1.2 nm. Particle mesh Ewald (PME) [147] was used for long range Coulomb interactions with a Fourier spacing of 0.12 and an order of four. The thermostat of Bussi, Donadio, and Parrinello [100] was used only on solvent molecules to maintain the temperature at 423.15 K. Periodic boundary conditions were applied in all directions. The microwave electric field strength was E = 0.1 V/nm and the frequency 2.45 GHz. For reference and comparison we also investigated the behaviour of pure DES sytems in the presence of microwave field of different strengths (E= 0.0 , 0.05, 0.075, and 0.1 V/nm).

Due to the high viscocity of the DES, an annealing procesure was applied to speed up equilibration. The temperature was first set at 600 K and linearly cooled down to 423.15 K over 10 ns at a constant pressure of 1 bar. Then the temperature was held at 423.15 K for another 20 ns

for further equilibration and to obtain the average volume. This provided the starting configuration for simulations of another 300 ns at 423.15 K and constant volume with or without electric field. The first 20 ns of these final simulations with electric field were discarded.

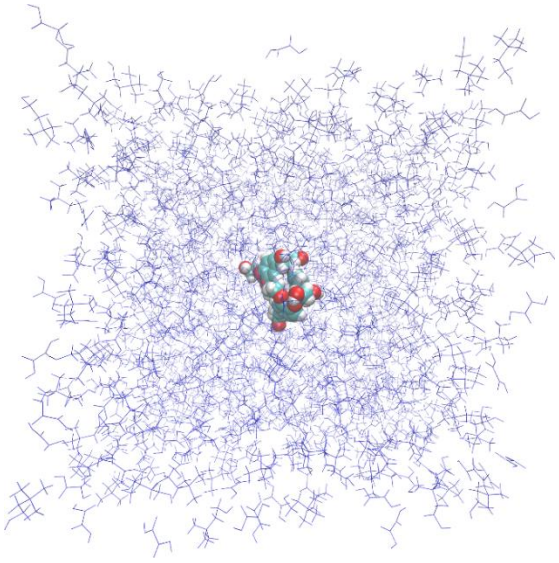


Figure 5.1. A simulation box containing one (β-O-4, β-O-4, β-O-4) G tetramer (spheres) and DES solvent molecules (purple lines).

5.3. Results and discussion

5.3.1. Effect of microwave field on DES

Figure 5.2 shows the total energy as a function of time for different field strengths. The slopes of total energy as a function of time are 34.45, 105.12, and 192.67 kJ/mol-ns for E = 0.05, 0.075, and 0.1 V/nm, respectively. As expected, these slopes approximately scale with E^2. Temperature increases were 0.01, 3.32, and 5.89 K for E = 0.05, 0.075, and 0.1 V/nm, respectively.

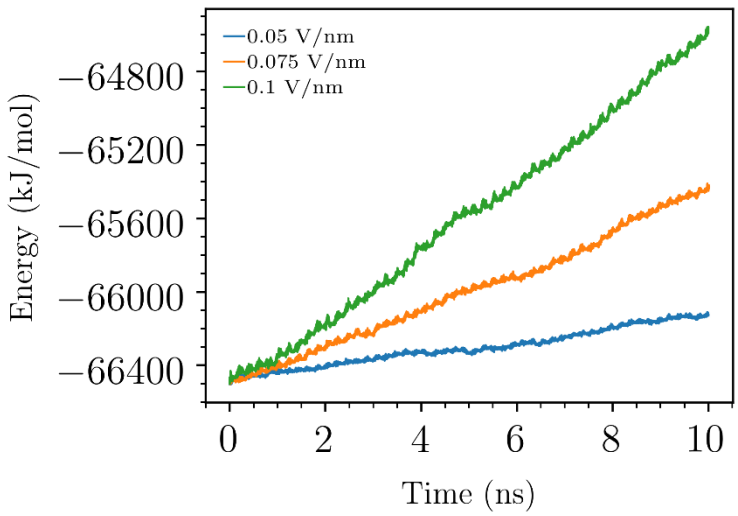


Figure 5.2. Total energy versus time in NVE simulations of DES.

Time autocorrelation functions (ACFs) were calculated for the angles between choline molecule N-O vectors and the x direction and for the x component of the dipole moment of all choline molecules (Figure 5.3). The normalized ACFs are then fitted to a sum of two exponential terms where a and τ are constants and t represents time.

$$ACF_{fit}(t) = ae^{-\frac{t}{\tau_1}} + (1-a)e^{-\frac{t}{\tau_2}} \quad (5.1)$$

An example fit is shown in Figure 5.4. We use the time integrals of ACF_{fit} from zero to infinity as a means of comparing the ACFs with a smaller number indicating a faster decay of the ACF. The ACFs decay faster with higher electric field intensity as shown in Table 5.1 and Table 5.2.

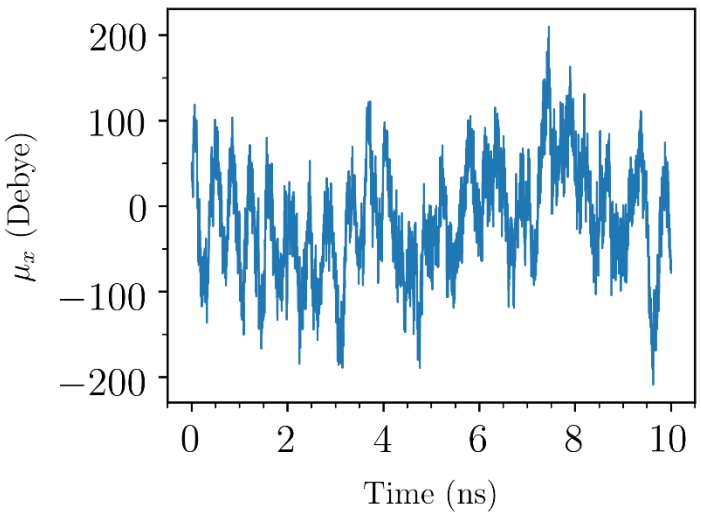


Figure 5.3. Dipole moment x component (μ_x) of all choline molecules in the simulation box versus time.

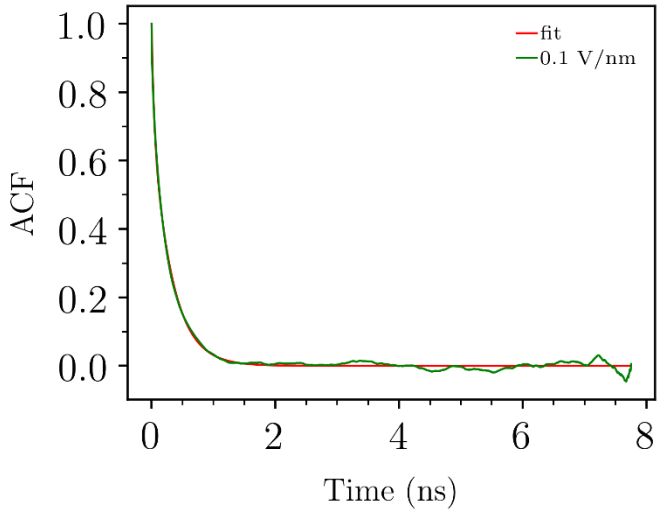


Figure 5.4. ACF and the corresponding fit for the angle between the choline N-O vectors and the x direction.

Table 5.1. Integrals of ACF fits (ns) for the angles between choline N-O vectors and the x direction. Config1, Config2, and Config3 refer to simulations with different starting configurations. SE is the standard error of the mean.

Field strength	0.0 V/nm	0.05 V/nm	0.075 V/nm	0.1 V/nm
Config1	256.8	254.8	258.5	242.6
Config2	-	246.6	233.6	235.1
Config3	-	236.0	239.5	219.9
Mean	-	245.8	243.9	232.5
SE	-	5.4	7.5	6.6

Table 5.2. Integrals of ACF fits (ns) for the dipole moment of all choline molecules. Config1, Config2, and Config3 refer to simulations with different starting configurations. SE is the standard error of the mean.

Field strength	0.0 V/nm	0.05 V/nm	0.075 V/nm	0.1 V/nm
Config1	263.72	266.3	270.7	261
Config2	-	253.5	232.5	218.1
Config3	-	245.6	253.3	224.6
Mean	-	255.1	252.2	234.6
SE	-	6.0	11.0	13.3

5.3.2. Effect of microwave field on lignin bond lengths

The tetramer molecules which were considered are shown in Figure 5.5. The structure of the two G tetramers with all β-O-4 linkages and different chirality are shown in Figure 2(a). Specifically, each β-O-4 linkage contains two chiral centers. The difference between the two tetramers was that the central linkage had either S, S chirality or R, S chirality. In addition, one G tetramer with β-O-4, β-5, β-O-4 linkages was also considered (see Figure 2(b)). Bonds which have differences in lengths greater than their uncertainties with and without electric field are marked on the structures. Note that all differences are only on the order of 10^{-4} nm or smaller which is expected since forces due to the electric field are small compared to thermal fluctuations. The bond with the lowest dissociation energy [148] in the β-O-4 linkages is the C_{β}-O bond which is bonds

3, 7, and 11 in Figure 5.5(a) and bonds 3 and 13 in Figure 5.5(b). Bond 7 in the all β-O-4 R,R,S,S,R,S isomer was significantly longer with electric field (Figure 5.6), but none of the other C_β-O bond lengths were increased significantly by the electric field. Other bonds with increased lengths were bonds 9 and 10 in Figure 5.5(a) for the R,R,R,S,R,S isomer, and bonds 6 and 8 Figure 5.5(b) for the β-5 linkage. Bonds with decreased lengths were bonds 1 and 11 in Figure 5.5(a) for the R,R,S,S,R,S isomer, bonds 4 and 5 in Figure 5.5(a) for the R,R,R,S,R,S isomer, and bonds 1, 5, 9, and 10 in Figure 5.5(b) for the β-5 linkage. Results for mean bond length differences and differences in probabilities of the bond lengths being greater than cutoff distances equal to the 99^{th} percentiles in the zero field cases for all bonds are discussed.

Files are included with the values of bond length differences for all bonds considered. The "System" column refers to the type of lignin tetramer where bo4-RRSSRS and bo4-RRRSRS are the isomers with all β-O-4 linkages and b5 is the tetramer with β-O-4, β-5, β-O-4 linkages. The "Bond num" column refers to the bond numbers shown in Figure 8 in the paper. The "Mean bond len diff (nm)" column refers to the difference in the mean bond length (E = 0.1 V/nm – E = 0.0 V/nm). The "Mean bond len diff unc (a=0.05)" column is half the width of the 95% confidence intervals for the mean difference. In addition to the difference between mean bond lengths, the right tail of the distributions was also analyzed since a significant difference in the means may not necessarily lead to a significant difference in the tails. The right tail with the longest bond lengths is the most important part when considering whether bonds will more frequently break or not. For analyzing the right tail, we found the bond length for the 99^{th} percentile (d_{cut}) of the distribution with E = 0.0 V/nm. This cutoff is in the "Bond length cutoff (nm)" column. We then found the probability of the bond length being greater than d_{cut} for E = 0.1 V/nm, $P(d > d_{cut} \mid E = 0.1\ V/nm)$. The difference between the probabilities for the two field strengths is just $P(d > d_{cut} \mid E = 0.1\ V/nm)$

– 0.01. This difference is in the “P > cutoff diff” column and half the width of the 95% confidence intervals is in the “P > cutoff diff unc (a=0.05)” column. In all cases, we see that if there is a significant difference in the means, then there is also a significant difference in the right tails. The distributions are mostly just shifted; their shapes do not change significantly.

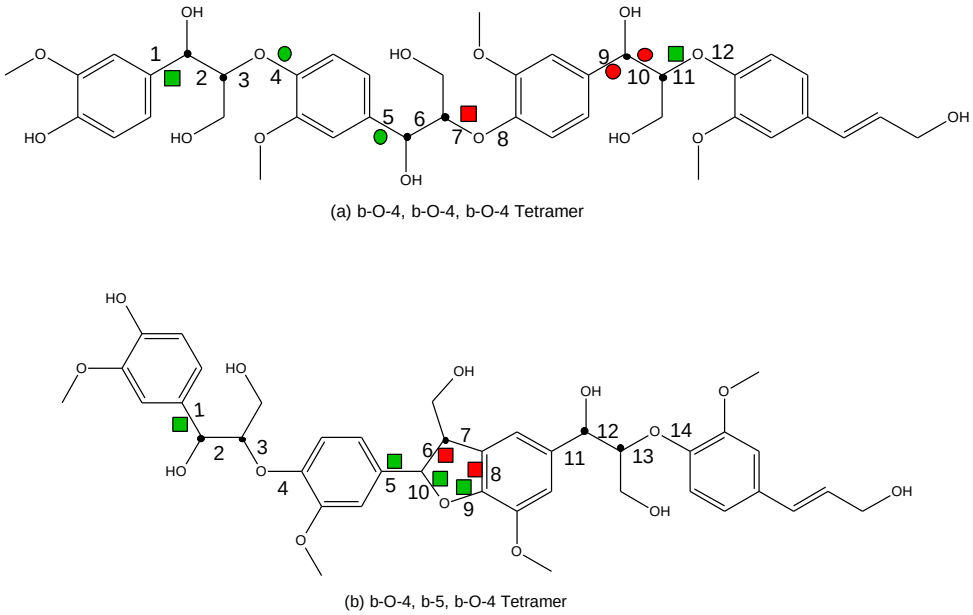


Figure 5.5. Tetramers considered. Chiral centers are marked with a black dot. Bonds which were significantly longer with electric field are marked in red and bonds which were significantly shorter with electric field are marked in green. (a) All β-O-4 linkages. Isomers considered were R, R, S, S, R, S (bonds with significant changes marked with squares) or R, R, R, S, R, S (bonds with significant changes marked with circles) starting from the left. (b) β-O-4, β-5, β-O-4 linkages.

Supplemental files are provided in Github repositorie https://github.com/pandatt66/deep-eutectic-solvent. The bond_lengths.csv file contains all the bonds. The bond_lengths_sig_chng.csv file contains only bonds where the magnitude of the difference in mean bond length is greater than the uncertainty. The bond_lengths_sig_inc.csv file contains only bonds where there is an increase in the mean bond length with the electric field turned on and the difference in mean bond length is greater than the uncertainty. The uncertainties were computed with a method based on the one

in the appendix [149]. Instead of using a constant prefactor, we allow the prefactor to vary to obtain better fits to the standard error as a function of block length. In addition, instead of using only one set of blocks for each length, we use the stationary bootstrap method [150] to sample 200 sets of blocks for each block length considered. This reduces the noise in the data significantly and therefore improves the estimate of the infinite block length standard error. The python code used for the analysis is in confidence_interval.py.

An example of using this code is:

```
cutoff=`cat ../analysis/bo4-RRSSRS/0.0/1/cutoff.dat`
python ../scripts/confidence_interval.py npy_files/bo4-RRSSRS/0.0/bond_dist_time.npy 1 -op ../analysis/bo4-RRSSRS/0.0/1/bond_dist_ -od ../analysis -sk 3 -vp -mb 24 -bsn 100 -nb 200 -cf "lambda x: np.hstack((np.mean(x), np.sum(x > $cutoff)/x.shape [0]))"
```

The mean bond lengths and differences in mean bond lengths are also shown in Table 5.3, Table 5.4 and Table 5.5. The bond length probability density bond 6 in the b5 system is plotted in Figure 5.6, showing a small shift towards longer bond lengths with E = 0.1 V/nm relative to E = 0.0 V/nm.

Table 5.3. βO4-RRSSRS bond length means and differences in means.

Bond #	Bond length difference	Bond length mean E = 0.0 V/nm	Bond length mean E = 0.1 V/nm
1	-2.1E-04	0.15389	0.15368
2	-5.4E-05	0.15450	0.15445
3	1.3E-05	0.14871	0.14872
4	-1.3E-04	0.13885	0.13871
5	7.0E-05	0.15375	0.15382
6	1.7E-04	0.15448	0.15465
7	2.0E-04	0.14862	0.14882
8	-1.9E-05	0.13870	0.13868
9	8.6E-05	0.15348	0.15357
10	-2.4E-04	0.15468	0.15444
11	-9.1E-05	0.14870	0.14861
12	-3.5E-05	0.13860	0.13856

Table 5.4. βO4-RRRSRS bond length means and differences in means.

Bond #	Bond length difference	Bond length mean E = 0.0 V/nm	Bond length mean E = 0.1 V/nm
1	-6.1E-05	0.15377	0.15371
2	9.0E-05	0.15452	0.15461
3	-3.2E-05	0.14870	0.14867
4	-2.0E-04	0.13895	0.13875
5	-2.0E-04	0.15360	0.15340
6	4.8E-05	0.15449	0.15454
7	-7.8E-05	0.14871	0.14863
8	-1.4E-04	0.13871	0.13857
9	2.9E-04	0.15365	0.15394
10	2.4E-04	0.15440	0.15464
11	7.7E-05	0.14874	0.14882
12	3.3E-06	0.13868	0.13868

Table 5.5. β5 bond length means and differences in means.

Bond #	Bond length difference	Bond length mean E = 0.0 V/nm	Bond length mean E = 0.1 V/nm
1	-4.1E-04	0.15383	0.15342
2	-3.2E-05	0.15434	0.15431
3	3.4E-05	0.14863	0.14866
4	-9.4E-05	0.13867	0.13857
5	-3.9E-04	0.15260	0.15221
6	3.6E-04	0.15622	0.15658
7	5.6E-05	0.15151	0.15156
8	1.1E-04	0.13942	0.13953
9	-1.0E-04	0.13853	0.13843
10	-1.4E-04	0.14663	0.14649
11	-3.0E-05	0.15407	0.15404
12	5.3E-06	0.15451	0.15452
13	-7.8E-05	0.14877	0.14869
14	-1.1E-05	0.13867	0.13866

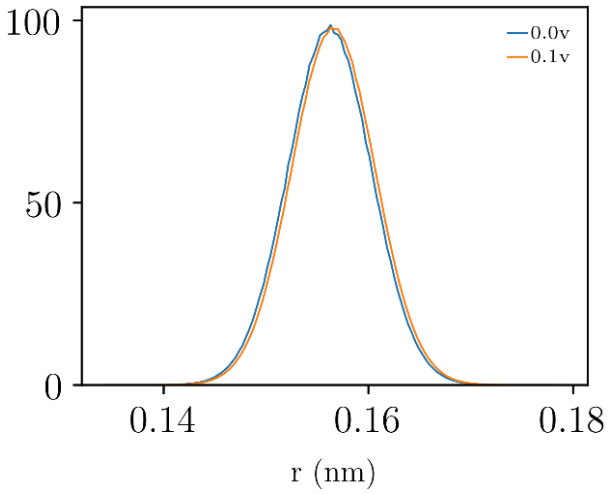


Figure 5.6. Bond length probability density distributions for b5 bond 6 with E = 0.0 V/nm and E = 0.1 V/nm.

5.4. Conclusions

The conclusion based on these limited simulations is that the electric field does increase some of the bond lengths including C_β-O in the β-O-4 linkage, but which bonds are affected seems to be a function of the overall conformation of the molecule which is a function of its stereochemistry. Changing one chiral center does not only change the bond length behavior for its linkage, but for the whole molecule. Based on our simulations it seems likely that some of the bonds in real lignin are stretched on average by the microwaves leading to an increased probability of breaking and therefore an increased rate of breakdown of the polymer.

CHAPTER 6. INTERACTIONS OF LIGNIN DIMERS AND ITS DERIVATIVES WITH BETA CYCLODEXTRINS

6.1. Introduction

6.1.1. Lignin biomass conversion

Lignin products have arisen increasing attention due to their abundance in nature as well as high production in industry process. The products are promising fuels or antibacterial ingredients, which are of great market value. Separation of lignin compounds, however, is still an intriguing work because of the structural complexity and stable intramolecular bonds. Chemical and enzymatic conversion included processes such as pyrolysis, hydrotreating, gasification and hydrolysis [135-137, 151]. Cleavage of C–O or C–C linkages are focused as a strategy of lignin depolymerization [152, 153]. The reductive or oxidative cleavage upon fractionation is a mixture of aromatic monomers, dimers, and larger oligomers [153, 154], which can be identified through Gas Chromatography, Mass Spectrometry (GC-MS) and Matrix-assisted laser spectrometry (MALDI-TOF MS) [155]. Effective method of high selectivity to allow separation, purification, and valorization to defined products is challenging.

6.1.2. Cyclodextrins (CDs)

Cyclodextrins (CDs) are natural molecules discovered from starch degradation in 1891 by Dr. Villiers. Three natural CDs composed of six, seven or eight D-glucopyranose units linked by α-(1, 4) glucosidic bonds are called $\alpha, \beta, \text{and } \gamma$–cyclodextrin respectively. These cone-shaped oligosaccharides have remarkable ability to capture guest molecules inside their internal hydrophobic cavity. Meanwhile, hydrophilic exterior improves inclusion complexes water solubility and increases their bioavailability and stability. Also due to their nontoxicity, CDs has contributed a wide application in the field of pharmaceutical, food industry, cosmetics, textiles,

agricultural and chemical industry. Uses of CDs or its relative polymers as sorbents in the separation of pollutants, organic molecules, proteins are also largely investigated [18, 156, 157]. Danquah et.al assessed the adsorption of phenol pollutants to linear and globular β-CD polymer, and it was likely to improve the adsorptive uptake by changing morphology [158].

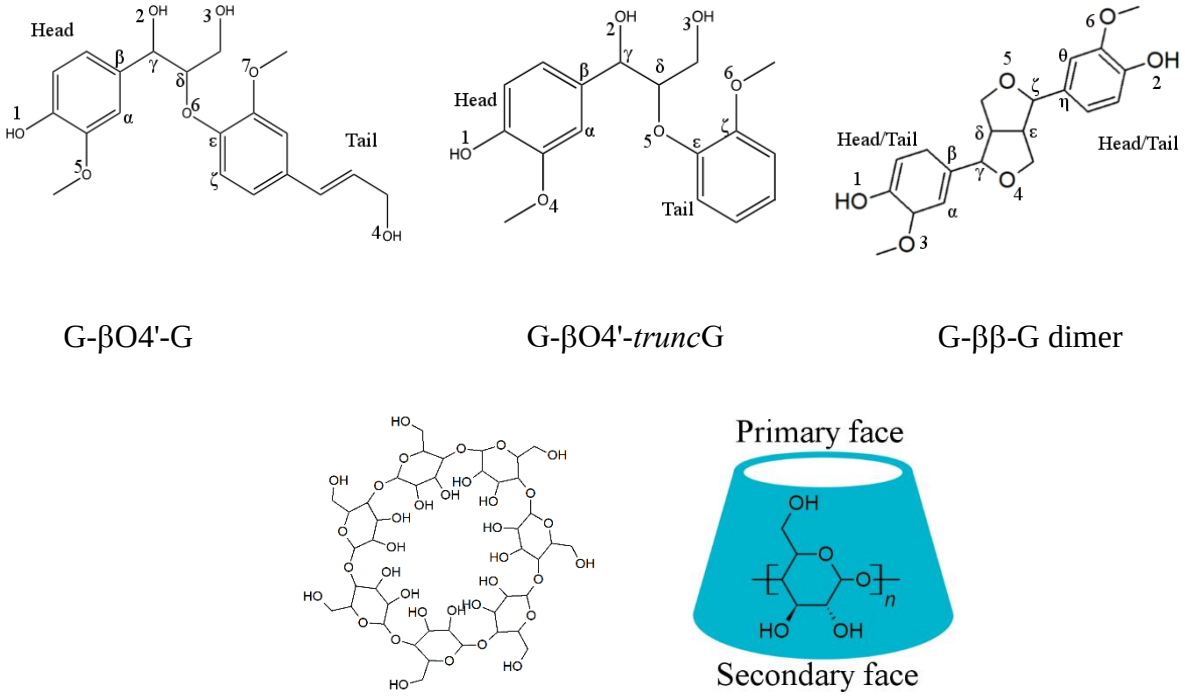
Molecular simulation is trendy and in massive studies of calculations on supramolecular systems involving CDs. Cesar et.al proposed that β-CD are able to remove cholesterol from cholesterol monolayer which favors at a high concentration [159]. Alvira studied the separation of isoleucine enantiomers by β-CD via MD and their binding affinity decreased when solvent dielectric constant increased [160]. Theoretical study on inclusion complexes of molecules with natural and modified β-CD via MD simulation also helps to understand the inclusion mode and mechanism and binding free energy [19, 161, 162].

6.1.3. Aim of the present work

The present work aims to study binding mode in complex of β-CD to guest molecules. Three types of lignin dimers and derivatives of β-O-4' linked dimer based on coniferyl alcohol (GG) was taken into consideration. The stoichiometry of guest/β-CD complex formed in aqueous solution is frequently 1:1, 1:2 or 2:1. [15, 163] Understanding the complex formation and dynamical behavior is of great importance. Therefore, the study of interactions between lignin dimers and β-CDs would provide insights into the potential application of selective separation from lignin heterogeneous mixture.

6.2. Methodology

6.2.1. Systems



G-βO4'-G G-βO4'-*trunc*G G-ββ-G dimer

β–cyclodextrins, n=7

Figure 6.1. Structure of molecules investigated in this work.

Table 6.1. Definition of four types of dihedral angles in dimer.

	G-βO4'-G	G-βO4'-truncG	G-ββ-G
Dihea1	$C_\zeta - C_\varepsilon - O_6 - C_\delta$	$C_\zeta - C_\varepsilon - O_5 - C_\delta$	$C_\alpha - C_\beta - C_\gamma - C_\delta$
Dihea2	$C_\varepsilon - O_6 - C_\delta - C_\gamma$	$C_\varepsilon - O_5 - C_\delta - C_\gamma$	$C_\beta - C_\gamma - C_\delta - C_\varepsilon$
Dihea3	$O_6 - C_\delta - C_\gamma - C_\beta$	$O_5 - C_\delta - C_\gamma - C_\beta$	$C_\gamma - C_\delta - C_\varepsilon - C_\zeta$
Dihea4	$C_\delta - C_\gamma - C_\beta - C_\alpha$	$C_\delta - C_\gamma - C_\beta - C_\alpha$	$C_\varepsilon - C_\zeta - C_\eta - C_\theta$

6.2.1. Simulation Details

Parametrization of β -CD and lignin dimers

Atomic geometry was extracted from crystal structure (ID 3LD6 [164]) in Protein Data Bank [165]. Cyclic connection was corrected using TopoTools [166] in VMD. [167] CHARMM-GUI Glycan Reader & Modeler [168, 169] was employed to generate CHARMM input files by detecting sugar-like molecules. Lignin natural G-βO4'-G dimer and G-ββ-G dimer structure and topology were created via LigninBuilder [145]. G-βO4'-truncG dimers coordinates were modified based on G-βO4'-G dimer by reducing a few atoms in tail and adding one hydrogen atom. Similar atom types were adopted for it, so the topology parameters were from the same library except some charges were redistributed to make whole molecule neutral. Crystal properties of G-βO4'-truncG were compared with the Cambridge Structural Database [170] so as to validate these parameters (see details in Supplementary material).

System set-up

All simulations were performed with the GROMACS 2018.3 package. Initial configurations were energy minimized by steepest descent algorithm, which followed by NPT simulation with leap-frog algorithm. The Lennard-Jones and electrostatic short range cut off was 1.2nm, and Force-switch modifier was applied on Van der Waals. Particle mesh Ewald (PME) was applied on long range Coulomb interactions. Berendsen barostat and v-rescale thermostat were used for pressure coupling and temperature coupling respectively. Water molecules were using TIP3P water model. MD simulations were run with a 2fs time step at 350K, 1bar. Periodic boundary conditions were applied in all directions.

Unbiased MD simulation

For inclusions of dimer into two cyclodextrin, the most preferable binding mode of dimer head inserting β-CD secondary face was initiated. One dimer was inserted in one β-CD cavity and the other β-CD was placed sort of parallel to the host β-CD. This starting construction will not artificially make a higher possibility of dimer being captured by both cyclodextrins, which can be proven by checking the stability of the complex. But it only accelerates the formation of that possible composition considering the time of diffusion. Two hosts capture one guest has exhibited regularly as well as can be conjectured from PMF curves.

Low concentrations simulation was prepared by using Packmol [146]. Reference group contains 10 dimers with 9000 water molecules. It showed that each dimer put into water by VMD [167] replaced about two dozens of water molecules. When added 20 β-CDs, a few hundred water particles were reduced, so the total volume were about the same.

Two phases simulation contained 200 packed dimers with 7000 water molecules which was prepared via Packmol [146]. By plotting density profile of a 8ns equilibrated simulation, bulk water depth was estimated at about 6nm. Unfolded dimer length and an inclusion complex size of one dimer into two β-CDs was measured about 1.5nm and 3.1 nm respectively. So, the amount of water should be enough. But because of the large spatial size of β-CDs, 200dimers and 30 β-CDs was hydrated with 9000 water. The simulations were run about 300ns.

Biased MD simulation

Free energy profiles were estimated by Umbrella Sampling method which was implemented through pulling code in Gromacs. A series of configuration was firstly generated based on distance between center of mass of dimers and β-CD. There are four possible insertion direction of dimers

to β-CD faces: Head-Secondary, Tail-Secondary, Head-Primary and Tail-Primary. Because of high orientational freedom of β-CD and dimers, pulling along the vector connecting them was set without an absolute direction. To avoid clutter provoked by samplings across the center of β-CD, the first window contained a dimer near the center of β-CD. Eleven windows were simulated starting from d = 0.1nm to d =1.2 nm with a spacing of 0.1nm. Each configuration had a biased potential with a harmonic force constant of 3000 kJ mol^{-1} nm^{-2}. Biased energy was recovered by weighted histogram analysis method (WHAM) and each window was analyzed from 60ns simulation with an equilibration time of 25ns. A correction on PMF afterward is necessary because distance restraint between two groups still have rotational freedom which contribute to entropic term [57]

$$\mathrm{PMF}_{\mathrm{corr}} = \mathrm{PMF}_{\mathrm{wham}} + 2 * \log(d) \, . \qquad (6.1)$$

MD dockings

Molecular docking was carried out via AutoDock Vina 1.1.2 [ref]. Energy minimized PDB structure from unbiased simulation was used to prepare PDBQT file through MGL Tools 1.5.6. Grid box with dimension 22 Å x 22 Å x 26 Å was centered at Cartesian coordinates: x=20.217 Å, y=20.008 Å, and z=20.514 Å, and the gird spacing was 0.375 Å.

Binding energy and entropy

$$\Delta G = -RTln(K) = \Delta H - T\Delta S \, , \qquad (6.2)$$

Where free energy ΔG is calculated from potential of mean force. R is gas constant. T is reference temperature.

Enthalpy change ΔH was estimated from the enthalpy difference of bound and unbound states $H_{bound} - H_{unbond}$. Unbound enthalpy was estimated from biased simulation where dimer and β-CD were far apart. Bound enthalpy was obtained from the unbiased simulation.

Eccentricity

Moment of inertia of a body is a sum of mass (m_i) multiple by squared distance (r_i^2) from rotation axis for each particle

$$I = \sum_i m_i r_i^2 \quad . \tag{6.3}$$

Radius of gyration (R_g) around principal axis is determined as:

$$R_{g,x} = \sqrt{\frac{\sum_i m_i \left(r_{i,y}^2 + r_{i,z}^2\right)}{\sum_i m_i}} \quad . \tag{6.4}$$

The Eccentricity based on R_g is averaged from over simulation frames, where $R_{g,x}$ is smaller than $R_{g,y}$

$$\mathrm{e_{gr}} = \sqrt{1 - \left(\frac{R_{g,x}}{R_{g,y}}\right)^2} . \tag{6.5}$$

Inclusions analysis

We used 'mdtraj.compute_neighbors' module from MDTraj tool 30 to calculate the number of oxygen atoms that connecting sugar units in β-CDs whose distance to one carbon atom in each dimer are smaller than the cutoff. The selected carbon atom(SCA) of dimer should be the one near the center of β-CD when it's included. We investigated each β-CD separately to avoid the occurrence that one dimer with multiple β-CDs around. This way if a dimer is trapped inside of

ring, the number of neighbors should be exactly seven. The cutoff distance of neighbors was determined by one dimer with one β-CD inclusion system. As we know the binding mode is stable over time, we are able to decide the cutoff until the neighbor's number are steadily seven.

6.3. Results and discussion

6.3.1. Binding free energy

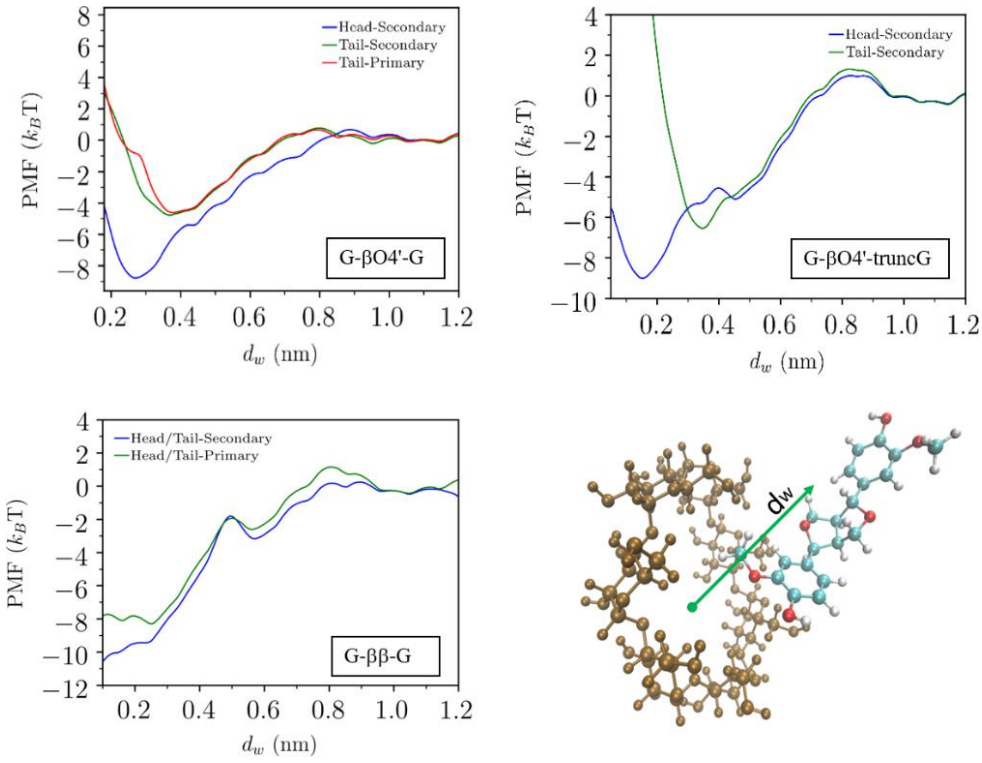


Figure 6.2. Potential of mean force for a dimer as a function of distance (d_w) from center mass of dimer to β-CD.

Table 6.2. PMF minima estimation

	PMF (k_BT) well depth		
	G-βO4'-G dimer	G-βO4'-*trunc*G	G-ββ-G dimer
Head-Secondary	8.8	9.4	10.6
Tail-Secondary	4.8	6.5	10.6
Head-Primary	-	-	8.9
Tail-Primary	4.6	-	8.9

PMF profiles for the formation of dimer into β-CD inclusion complexes in head-primary, tail-secondary and tail-primary modes are shown in Figure 6.2. The lowest energy is head-secondary binding mode for all types of dimers. As dimer approach to β-CD from solution, a drop of PMF curve showed a more favorable thermodynamic state. When dimer located inside the hydrophobic cavity, the hydrophilic moiety contains hydroxyl group are left outside.

Head-primary or Tail-Primary mode has not been explicitly run because dimer would come out in a few picoseconds from the cavity and then entered the secondary face shortly. Because of the difficulty of deciding whether dimer has passed the center of cyclodextrin or not, samplings obtained right near the middle will not be trusted. Therefore, the distance to center mass of β-CD in PMF started around 0.1nm. As seen in Figure 6.2, the PMF minima is missed for G-ββ-G dimer. Affinity obtained from docking best mode for G-βO4'-G dimer, G-βO4'-truncG, G-ββ-G dimer are -8.3, -8.0 and -9.33 k_BT respectively. These results lined up with free energy calculation from MD. Then we extracted the best mode from docking PDBQT file and measured the center mass between dimer and β-CD. There showed 0.26, 0.21 and 0.15 nm for G-βO4'-G dimer, G-βO4'-truncG and G-ββ-G dimer respectively. It has the same trend as we observed from PMF. Therefore, we estimated that the lowest energy of G-ββ-G dimer insertion located near the center of the β-CD.

6.3.2. Inclusions of lignin dimers into β-CDs

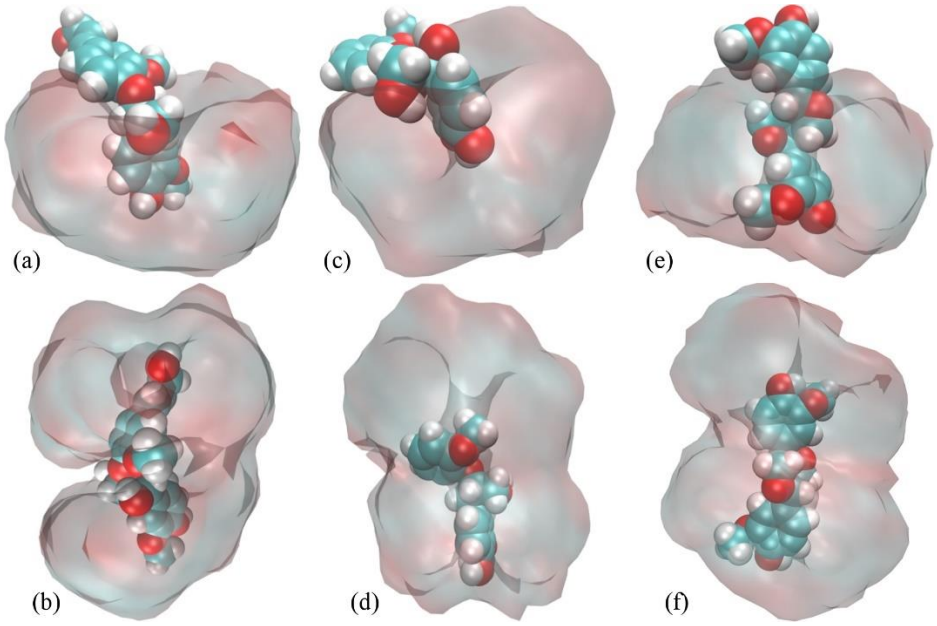


Figure 6.3. Inclusion complexes of (a-b): G-βO4'-G ;(c-d) G-βO4'-truncG ; (e-f) G-ββ-G with of one β-CD (upper) and two β-CDs (lower).

Dimer into one β-CD has a most preferable head to secondary face binding mode. When one G-βO4'-G dimer and one β-CD was randomly put in a small hydrated box, it could not form a complex even after 20ns. Both particles diffusing fast in the water brought about their arduous encounter along with a perfect insertion pose. Therefore, lignin dimer was initialized located near the opening of β-CD to accelerate the procedure and it took only a few picoseconds for dimer to move innermost part. These positions of formed complex are stable more than 10ns. As a high temperature of the simulation, captured dimer was able to overcome a 4 to 5 $k_B T$ barrier to leave the ring for a short time.

From Figure 6.2, the lowest energy of inclusion complex for both first and secondary face are near the center of dimer. Thus, for G-ββ-G to get rid of one host could happen at a much higher frequency than the other two dimers. A confliction between β-CDs is very strong as they

encapsulated a part of G-ββ-G but not the optimal position. Compared to the native βCD, the circle shape for β-CDs has been distorted about 7% with dimers included. It does not make too much difference when dimers inserted in one or two β-CDs. Angles between major principal axes of two β-CDs with one included dimer shows no preferences as the distribution for which greater than 15 degrees are about the same. Even run at a lower temperature of 300K, the results are similar. So, it is probably not a matter of sampling and β-CDs are flexible enough to rotate with dimers inside.

The inlusion of 1:2 dimer/ β-CD is not stable as the evolution of two β-CDs distance becomes large and fluctuating in time, which is hint of a separation for the two β-CDs(). Even there shows two β-CDs are close, it could also possibly that one attached to the other from sideway. G-ββ-G dimers shows a faster equilibration and dimer can only be included in one β-CD. Because there are competitions between two β-CDs as the most preferable part of dimer of capture is not a wide region. G-ββ-G dimer most likely have a stable inclusion when the center of mass of dimer located near the center of β-CD based on PMF calculations. Also due to a shortest size in length , it's the most difficult for another β-CD to steal from an exsisted inclusion complex. A longer simulation was performed for G-βO4'-G and G-βO4'-truncG (See Appendix C). As the results indicates, a 1:2 ratio of dimer/ β-CDs is still not stable over time even they can form once in a while.

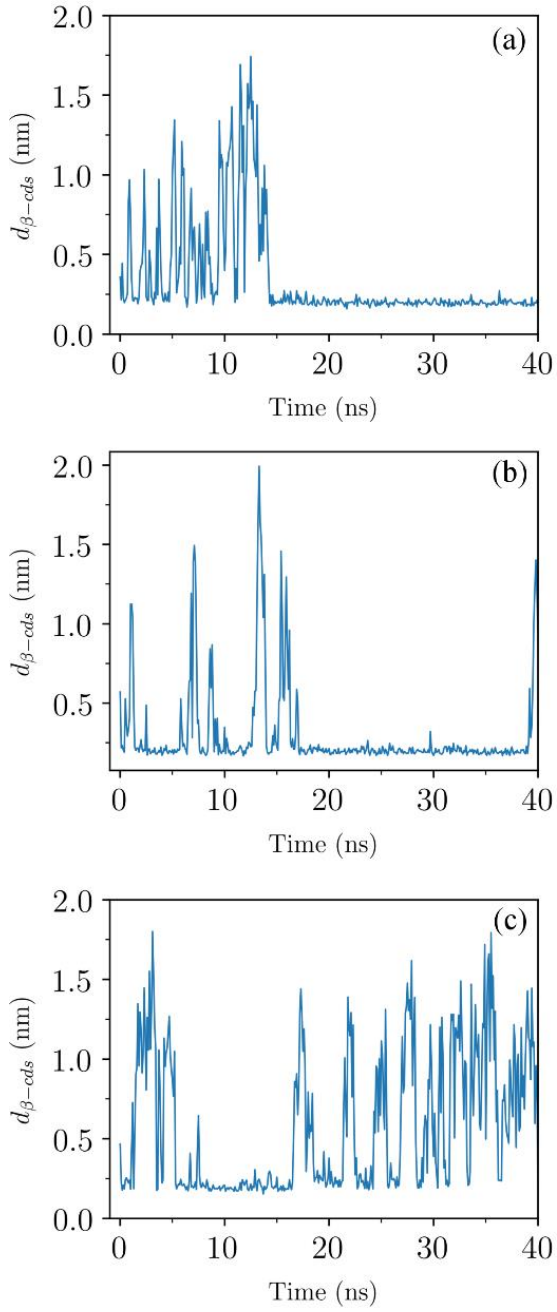


Figure 6.4. Distance between two β-CDs based on their center of mass with a inclusion of (a) G-βO4'-G (b) G-βO4'-truncG (c) G-ββ-G.

Table 6.3. Eccentricity of one or two β-CDs with one included dimer; and angles between major principal axes of two β-CDs with one included dimer. Eccentricity of a native β-CD is 0.14.

	One β-CD	Two β-CD		
	Eccentricity	Eccentricity1	Eccentricity2	Angle(°)
G-βO4'-G	0.39±0.03	0.37±0.01	0.38±0.03	54.0±1.6
G-βO4'-truncG	0.37±0.02	0.39±0.01	0.39±0.01	48.2±2.8
G-ββ-G	0.39±0.01	0.40±0.01	0.43±0.01	46.5±4.2

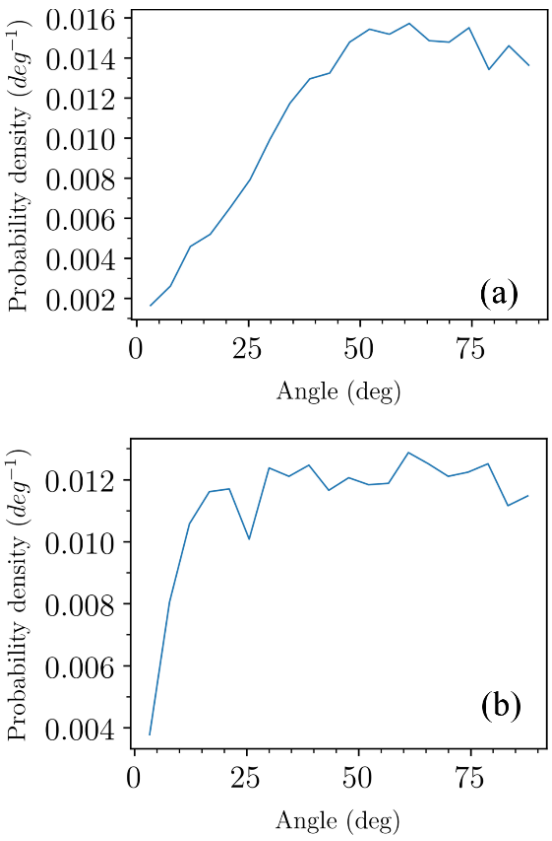


Figure 6.5. Probability distribution of angles between principal axes of two β-CDs. The complex contains one (a) G-βO4'-G (b) G-βO4'-truncG dimer.

6.3.3. Hydrogen bonds (HBs) and thermodynamic property

Table 6.4. Hydrogen bond number in inclusion of lignin dimers into one β-CD

	Number of Hydrogen bonds		
Bond type	G-βO4'-G	G-βO4'-truncG	G-ββ-G
Dimer-water	5.44±0.14	5.50±0.18	4.84±0.15
(β-CD)-water	25.94±0.34	28.3±0.33	26.07±0.30
(β-CD) -dimer	1.07±0.08	0.40±0.05	0.58±0.07
Water-water	3234.57±2.78	3242.41±2.21	3241.59±2.58
Water-water/N_w*	1.54	1.54	1.54
Donor/acceptor from dimer	Number of Hydrogen bonds with water		
OH-1	1.23± 0.07	1.49± 0.07	1.59±0.07
OH-2	0.24 ±0.05	1.13± 0.08	1.55±0.09
OH-3	1.06 ±0.07	1.46± 0.07	-
OH-4	2.14 ±0.08	-	-
O-3	-	-	0.5±0.06
O-4	-	0.60 ±0.06	0.28±0.05
O-5	0.15 ±0.04	0.40± 0.04	0.25±0.04
O-6	0.34 ±0.05	0.43± 0.04	0.67±0.06
O-7	0.29 ±0.05	-	-

*N_w is the number of water molecules (2099, 2106, 2105) in the systems containing inclusion of G-βO4'-G, G-βO4'-truncG and G-ββ-G into one β-CD respectively.

Hydrogen bonding (HB), hydrophobic interaction and pi-pi interaction are recognized as important factor of complex formation [169, 170]. Number of hydrogen bonds were calculated by Gromacs built-in function 'gmx hbond'. The cut-offs for a hydrogen bonding donor (HBD)−acceptor(HBA) distance and acceptor−donor−hydrogen angle was 0.35 nm and 30° respectively. Hydrogen bonding types of dimer-water, water-water, and dimer donor/acceptor-water were analyzed for inclusions complex of β-CD. The analysis of hydrogen bonding between dimer and water has been split into the contributions from hydroxyl groups and ester oxygen at varied locations. One dimer with 2100 water was used as a reference which is listed in Table 6.5. G-βO4'-G, G-βO4'-truncG and G-ββ-G has 11,9 and 7 HBAs/HBDs respectively. Even with less HBAs/HBDs, G-ββ-G has about the same number of hydrogen bonding with water compare with G-βO4'-truncG. This perhaps due to the conformational preference of G-ββ-G is mostly unfolded

and relaxing in water. From the dihedral angle distribution (Figure 6.6), G-βO4'-G and G-βO4'-truncG are likely to fold in the middle of around 60˚. Therefore, encompass of ester oxygen in the middle led to less hydrophobic interactions than unfolded structure. Two benzene rings of G-βO4'-G stacking more than G-βO4'-truncG as the hydroxyl group in the tail can also form hydrogen bonding with the one in head. Dimer-water hydrogen bonding numbers decreased once dimers were inside hosting as the hydrophobic cavity protected HBA/HBD moiety from contacting with water. Number of HB between OH-4 and water is about the same as in pure water for G-βO4'-G. It is probably located in hydrophilic region outside the cavity. In addition, it can also form hydrogen bonding with rim of β-CD. Two -OH in the middle of dimer therefore will have less freedom to rotate than G-βO4'-truncG.Water-water hydrogen bonding decreased a little with the presence of one more β-CD.

Table 6.5. Hydrogen bond number of lignin dimers in water

	Number of Hydrogen bonds with water		
Bond type	G-βO4'-G	G-βO4'-truncG	G-ββ-G
Dimer-water	9.12±0.13	6.90±0.13	7.10±0.13
Water-water	3243.14±2.23	3249.64±2.03	3247.66±2.15
Water-water/N_w*	1.54	1.55	1.55
Donor/acceptor from dimer	Number of Hydrogen bonds with water		
OH-1	1.71± 0.07	1.71± 0.06	1.69±0.06
OH-2	1.71 ±0.05	1.63± 0.07	1.72±0.07
OH-3	1.64 ±0.07	1.64± 0.06	-
OH-4	2.18 ±0.08	-	-
O-3	-	-	0.78±0.06
O-4	-	0.81 ±0.05	0.97±0.05
O-5	0.89 ±0.06	0.37± 0.04	1.07±0.04
O-6	0.46 ±0.04	0.74± 0.05	0.86±0.06
O-7	0.53 ±0.04	-	-

*N_w is the number of water molecules (2100) in the systems containing inclusion of G-βO4'-G, G-βO4'-truncG and G-ββ-G into one β-CD.

Thermodynamics property for the formation of dimer/β-CD complex is shown in Table 6.6. The enthalpy changes for G-βO4'-truncG is close to zero and molar entropies are positive and

similar for the G-ββ-G. As dimer is bound into β-CD, hydrophobic area for water networking has been reduced. The large positive value of ΔS means the inclusion processes for G-βO4'-truncG and G-ββ-G into β-CD are exothermic and is driven by entropy. In contrary, enthalpy change dominates inclusion formation process for G-βO4'-G, and entropy has decreased after the formation of inclusion complex. This enthalpy-entropy compensation relationship has also been observed in other literatures [171, 172]. As seen in Figure 6.6(a), the dihedral angle in the middle of G-βO4'-G dimer has been straighten up and the dimer hydrophobic carbonyl tails stick out may contribute higher energy to water.

Table 6.6. Thermodynamic property of inclusion complex of dimer into one β-CD.

	G-βO4'-G	G-βO4'-truncG	G-ββ-G
H_{bound}(kJ/mol)	-58096.8	-58324	-58361.6
$H_{unbound}$(kJ/mol)	-58066.8	-58323.2	-58352.2
ΔG(kJ/mol)	-25.6	-27.4	-30.8
ΔH (kJ/mol)	-30	-0.8	-9.4
ΔS(J/K.mol)	-12.6	75.8	61.1

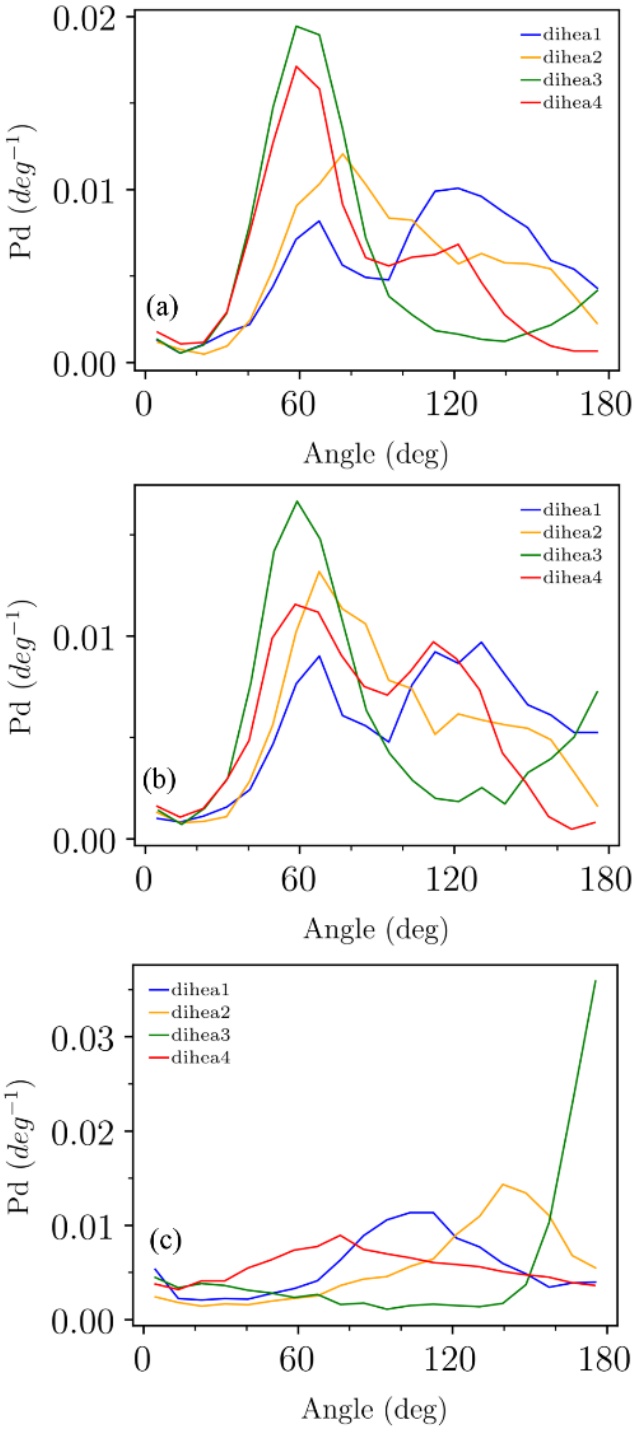


Figure 6.6. Dihedral angles distribution of one (a) G-βO4'-G (b) G-βO4'-truncG (c) G-ββ-G dimer with water.

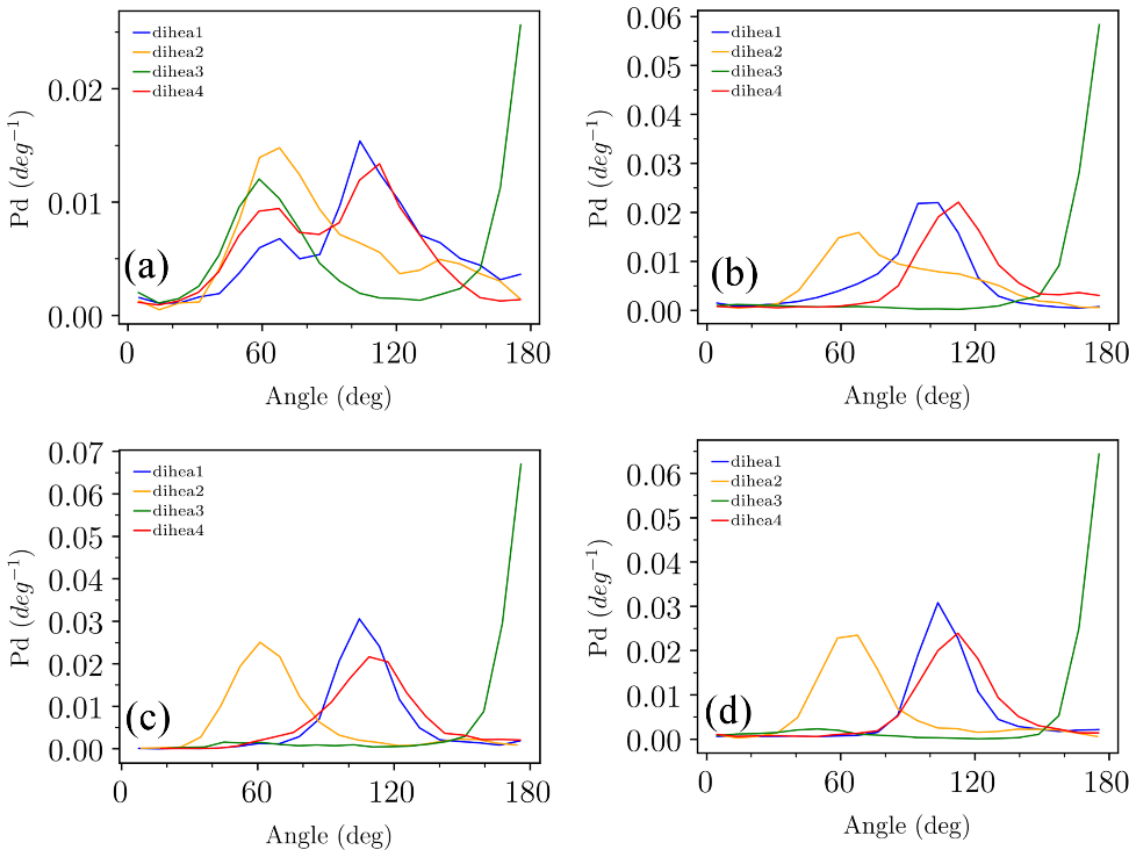


Figure 6.7. Dihedral angles distribution of one (a-b) G-βO4'-G (c-d) G-βO4'-truncG dimer with one β-CD(left) and two β-CDs(right).

6.3.4. Low concentrations of dimer with β-CDs

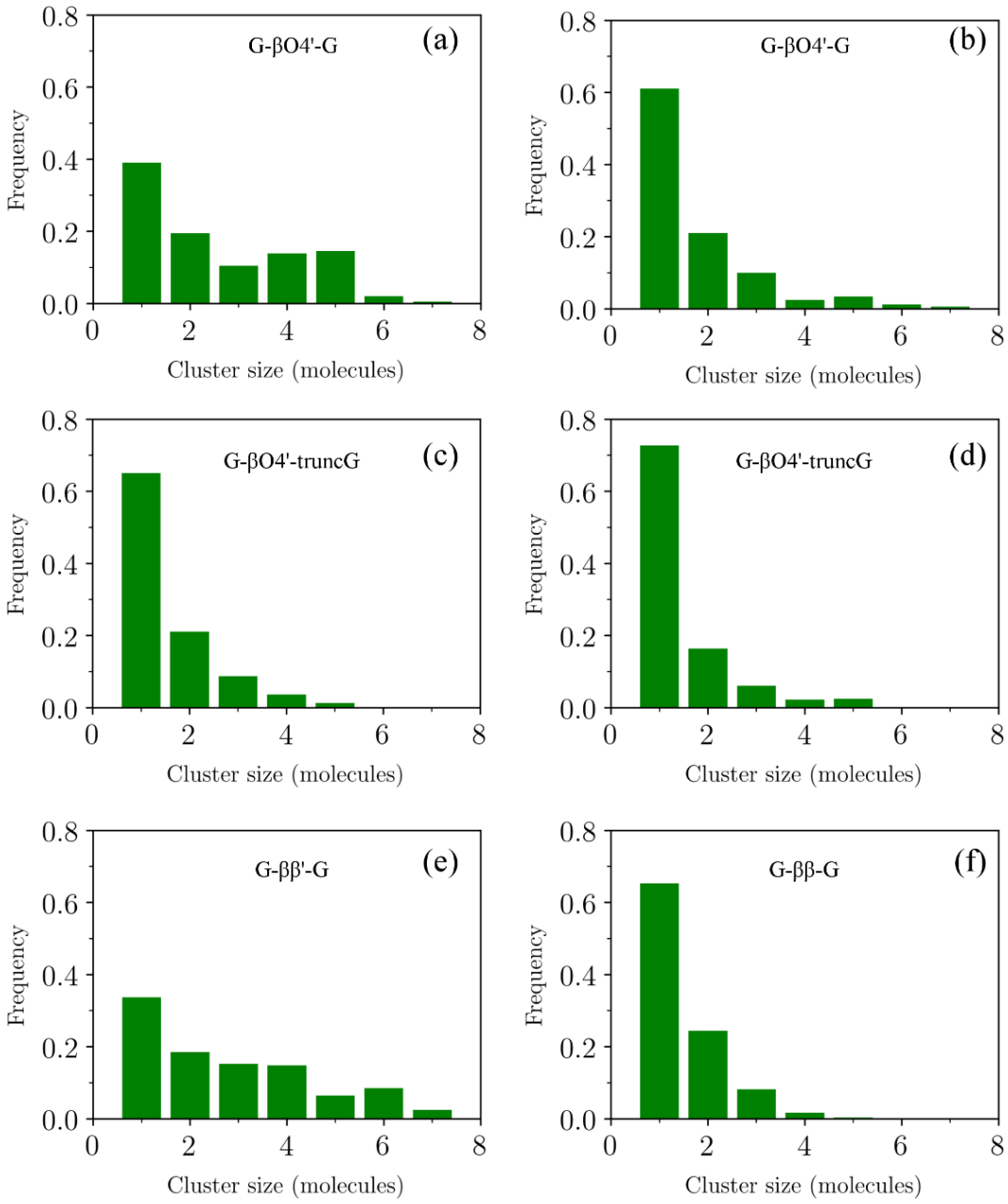


Figure 6.8. Frequency versus cluster size from statistical analysis over dynamical frames. System contains dimers with water only (left column) and with system contains 10 dimer and 20 β-CDs (right column).

A mixture of N_{dimer}: $N_{\beta-CD}$ =1:2 was simulated in water to investigate the interaction states at low concentration. Figure 8 compared the cluster size with dimers only in water.

Cluster size was determined by the number of molecules who are in contact with a cutoff of 0.35nm. As indicated from MD simulation, G-βO4'-G and G-ββ–G dimer aggregated more in water than G-βO4'-truncG. With presence of β-CDs, the cluster composed of one dimer has been increased 50%,7% and 90% for G-βO4'-G, G-βO4'-truncG and G-ββ-G respectively. Also, two to seven dimers in one cluster has diminished. However, it depends largely on the cutoff and the cluster size does not reflect weather the cluster included dimers captured by β-CDs or not. Once a dimer is included in β-CDs, if they are connected side by side which would still be counted as a cluster.

Therefore, further analysis of the number of formed inclusions is shown in Figure 6.9. We investigated each pair of β-CD with one dimer and if the selected carbon of dimer has selected atoms of β-CD ring as neighbors, it was estimated as an included mode. C_β was selected for G-βO4'-G and G-βO4'-truncG, which is near the head of dimer. C_δ was selected of G-ββ-G which is near the center of the dimer. These positions are most likely to have a minimal energy based on PMF. All seven oxygen atoms that connecting sugar units were used as reference of β-CD ring. The cutoff distance obtained from single dimer into β-CD inclusion was 1.0, 0.9 and 1.15 nm for G-βO4'-G, G-βO4'-truncG and G-ββ-G respectively. This cutoff offered 99%, 99% and 100% accuracy which assumes dimers are bound all the time during simulation. It probably covered most of the dynamical positions of dimer inside a β-CD. In the cases of dimer on the edge of β-CD may be slightly outside the cutoff or contributed a large distance to SCAs in order to be considered as bounded. Because the cutoff distance is sort of large and may include all the 1:1 binding mode . G-ββ-G are mostly separated by β-CDs and the mixture reached an equilibrium in around 6 ns. It seems like to take longer to equilibrate inclusion of G-βO4'-G and G-βO4'-truncG into cyclodextrin. There exist more competitions between dimer-dimer, dimer-CD and CD-CD in the

fusion. For 1:2 inclusions, a dimer was traced when it has two β-CDs as neighbors at the same time. Although the results are not completely accurate, as the complexity of possible states, what's done from the approach provided a rough estimation of bindings components during the free simulation at low concentration of a mixture of dimers and β-CDs.

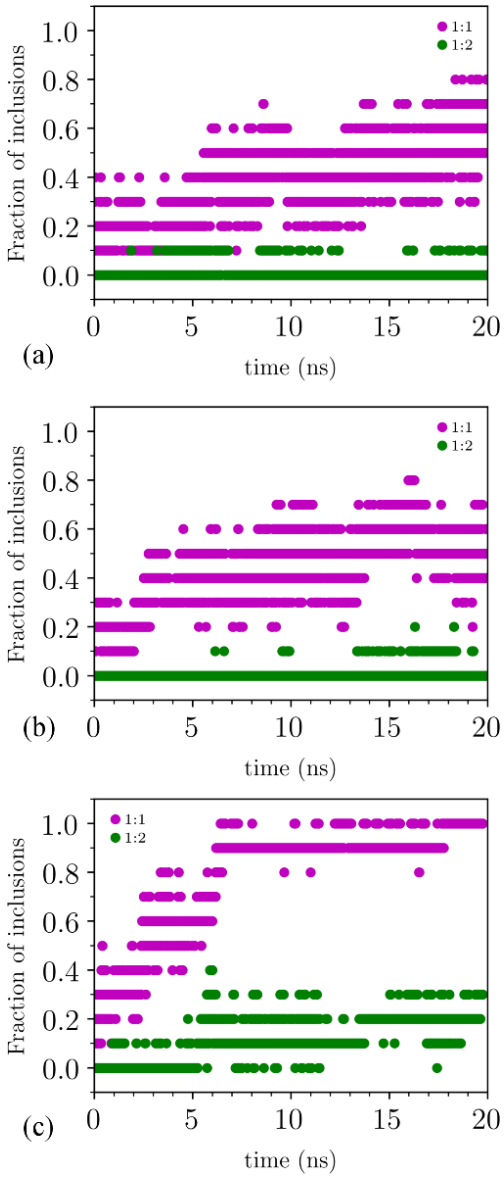


Figure 6.9. Fraction of 1:1 and 1:2 inclusions versus time(ns). System contains 10 dimer and 20 β-CDs with 9000 water for (a) G-βO4'-G (b) G-βO4'-truncG and (c) G-ββ-G.

6.3.5. Two phases unbiased simulation

Coordinates of each dimer was analyzed to estimate required free energy to dissolve lignin dimers in water:

$$F_i = G_{dimer\_phase} - G_{water\_phase} \quad (6.6)$$

$$F_i = -\ln(\frac{p_{zi}}{p_0}) , \quad (6.7)$$

where p_{zi} and p_0 is the probability distribution of dimers positions in z direction and in bulk water.

The free energy cost for a dimer to move from dimer phase to water phase are shown in

Table 6.7 and Figure 6.10. It shows that β-CDs decreased the solubility of G-βO4'-G dimer. β-CDs covered the interface and somewhat prohibited dimers from dissolving. β-CDs improved the solubility of G-βO4'-*trunc*G and G-ββ–G dimer but only a small amount of about 1 $k_B T$. Formed inclusion complexes which diffused into water phase went back after some time as observed. Thus, the dimer-dimers interactions are much stronger than dimer with β-CD and the required energy to pull a dimer out of its phase is large. This could also due to a relative low concentration of β-CDs. With more β-CDs per dimer, more disruption may occur to break down the original arrangement. On the other hand, β-CDs has a much lower solubility(2% w/w) comparing with α-CDs(13%), γ-CDs(26%) and 2-hydroxypropyl-β-cyclodextrin (HP-β-CD)(60%) [19]. Hence with a guest inside, β-CD intramolecular hydrogen bonding network can cause less interaction with water and absorbed back to dimers phase.

Table 6.7. Free energy converted from dimer-water phase

	F (k_BT) minima		
	G-βO4'-G dimer	G-βO4'-*trunc*G	G-ββ-G dimer
Water-dimer	-5.06	-5.10	-6.24
(Water+30 β-CDs)-dimer	-5.39	-4.16	-5.10

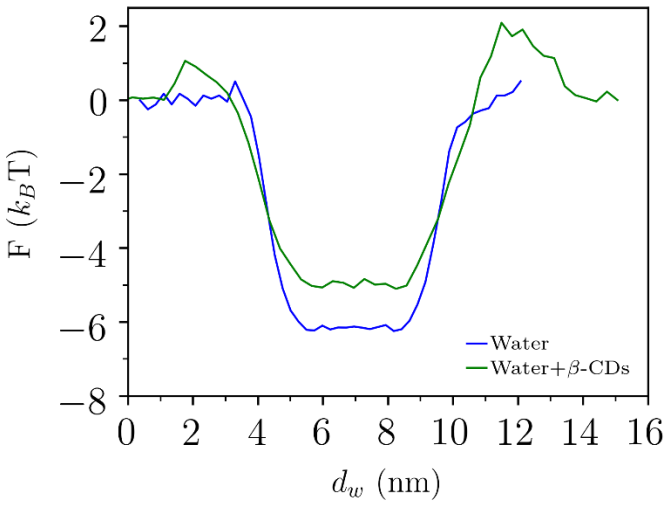


Figure 6.10. Estimated free energy of dissolving lignin dimers into water. Blue curve shows 200dimers +7000 water in system. The green one shows 200 dimers + 30 β-CDs + 9000water. The flat region in the middle (dw = 5-9 nm) is for dimers phase. The two ends of curve (0-4nm, 10-16nm) regions are for water phase.

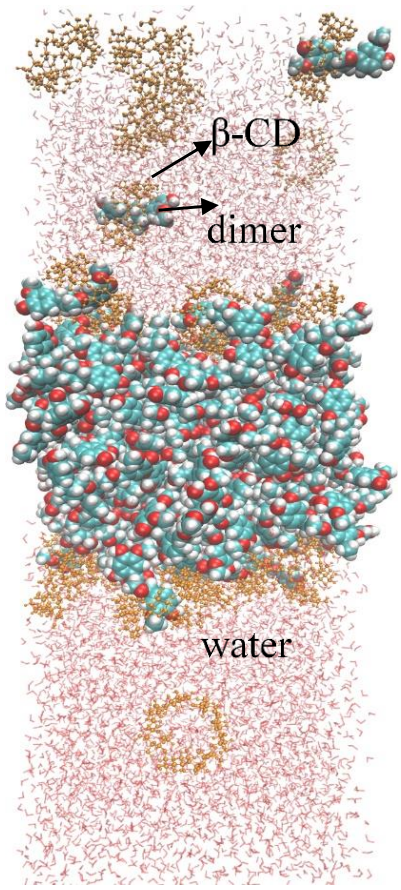


Figure 6.11. A simulation box contains G-ββ-G dimer phase and water phase.

6.4. Conclusions

In this work, free energy profiles for the inclusion of one lignin dimer into β-CD were obtained using umbrella sampling method. The results indicated a most preferable binding mode is head to secondary surface. Furthermore, dynamical studies showed the stability of the inclusion complex of 1:1 dimer/β-CD. For an inclusion complexes of 1:2 dimer/ β-CD, G-βO4'-G dimer and G-βO4'-truncG will go for the ratio back and forth but will not stable overall. G-ββ-G dimer has a more obvious reluctance with that ratio because of its size and minima energy located near to the center of the molecule. Furthermore, dynamical property of hydrogen bonding and thermodynamics properties were analyzed for the formation of 1:1 inclusion complex. The

formation of β-CD with G-βO4'-truncG and G-ββ-G dimer is entropy driven and with G-βO4'-G dimer is enthalpy driven. At low concentration of a mixture of N_{dimer}: $N_{\beta-CD}$ =1:2, the aggregation of dimers has been improved to some extent. Around 80% G-βO4'-G dimer and G-βO4'-truncG dimer will find a host and each individual G-ββ-G dimer could be captured by one or two β-CDs. Two phases simulation indicated a difficulty of dimer to be extracted by β-CDs. Also, as the β-CDs approached the dimer phase, a coverage of interface on the contrary impeded dimers dissolving. Even formed inclusion complexes carrying dimers into water may have a retreat over time. This may due to a low concentration and low solubility of β-CDs. Modified CDs with different size and better solubility can be simulated to compare with the current results in the future. In summary, cyclodextrin is a promising material to assist short lignin oligomers separation from a heterogenous mixture as it showed a strong affinity to β-CDs as well as a stable 1:1 inclusion complex. The selectivity should be improved extensively if a spaced β-CDs can be created according to the desired molecule size.

CHAPTER 7. SUMMARY

In this work, using MD simulations, we investigated the interactions of various small molecules with lipid bilayers and surfaces. The study of interaction of DNA mono-nucleotides with self-assembled monolayers with various terminal groups demonstrates that the phenoxy terminated surfaces are the best choice that provide optimized nanoslits characteristics for their use in DNA sequencing devices. Overall the simulation findings have the potential of providing valuable information critical to development of a new DNA sequencing technique based on time of flight approach. In another study we investigated the interaction of lignin oligomers with lipid bilayers. By focusing on three lignin dimers with different structure and hydrophobicity, our MD simulations provided atomistic insight into how the lignols affect the gel to fluid phase transition temperature in DPPC bilayers. The values of the phase transition temperature shifts obtained from MD simulations are in good agreement with those obtained experimentally from differential scanning calorimetry studies. These findings may play an important role in the development of a comprehensive understanding of lignin oligomers antibacterial properties and their potential for use in pharmaceutical applications. The investigation of the interaction of lignin dimer with cyclodextrin provides insights and values of their affinity and complex formation which may play an important role in the development of separation technologies of small lignin fragment from heterogeneous mixtures. We are confident that in the future the MD simulations will continue to serve as a powerful research tool that will expand and facilitate the advancement and new discoveries in science, engineering, and industry.

APPENDIX A. SUPPLEMENTAL MATERIAL FOR CHAPTER 3

Potential of mean force

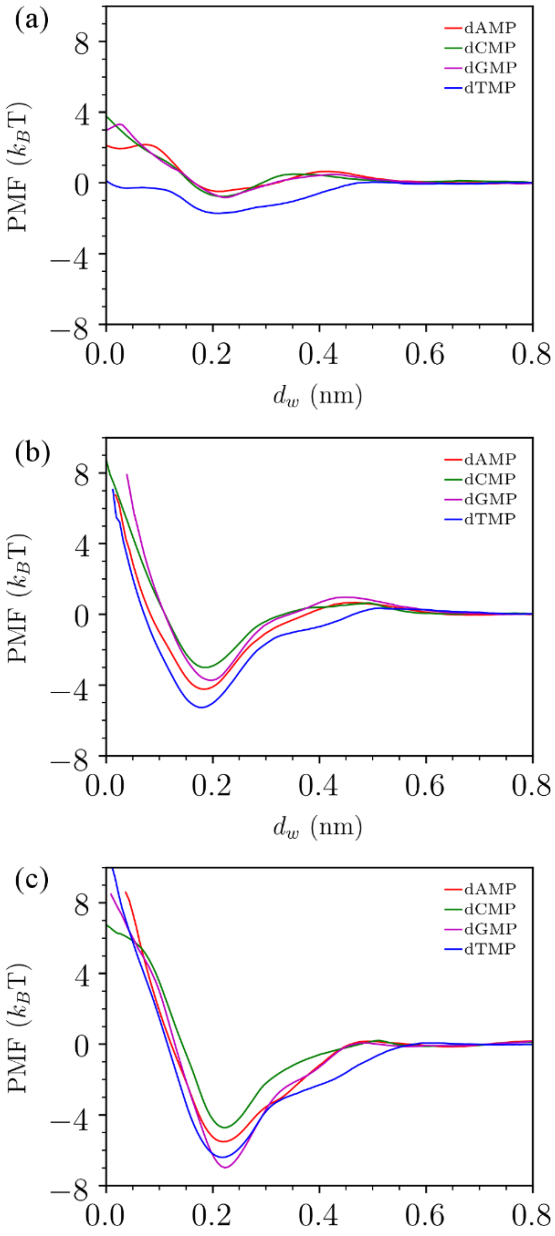


Figure A.1. Comparison of PMFs in k_BT as function of distances (nm) from the dNMP nucleobases to the wall surface for (a) COOCH3-SAM, (b) CH3-SAM, and (c) OC6H5-SAM.

Positioning

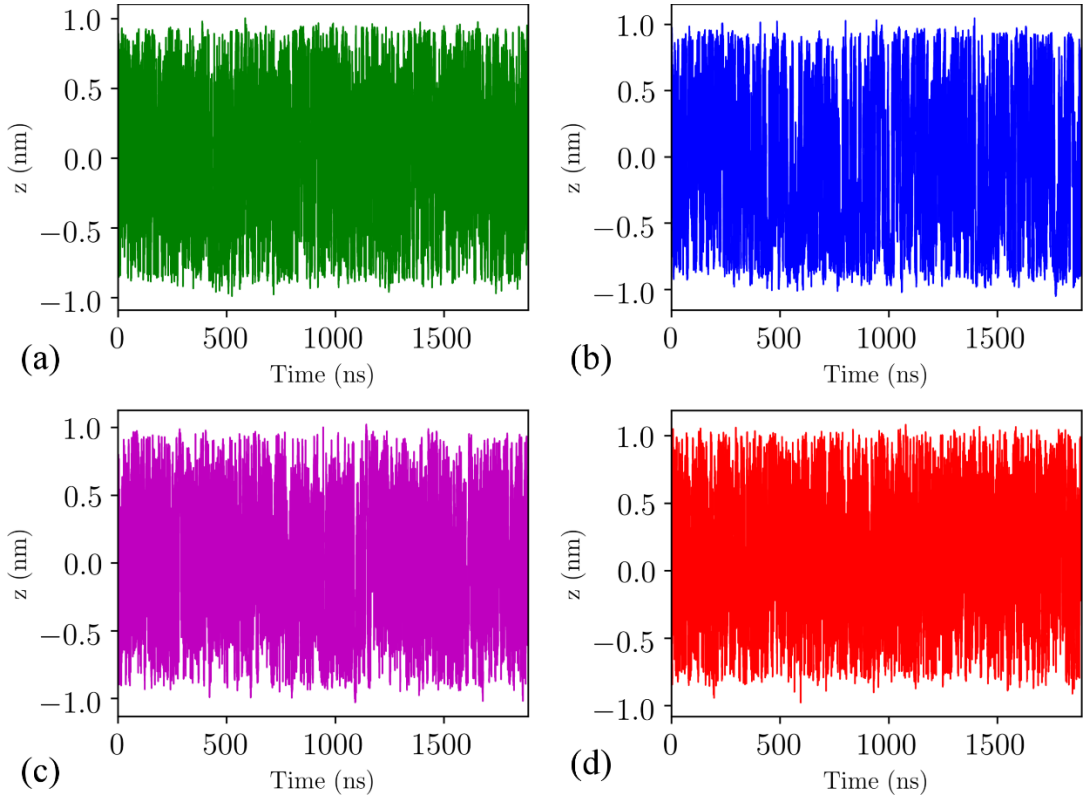


Figure A.2. dNMP position, *z*, relative to the slit centerline as a function of time for the COOCH3-SAM walls: (a) dCMP, (b) dTMP, (c) dGMP, (d) dAMP.

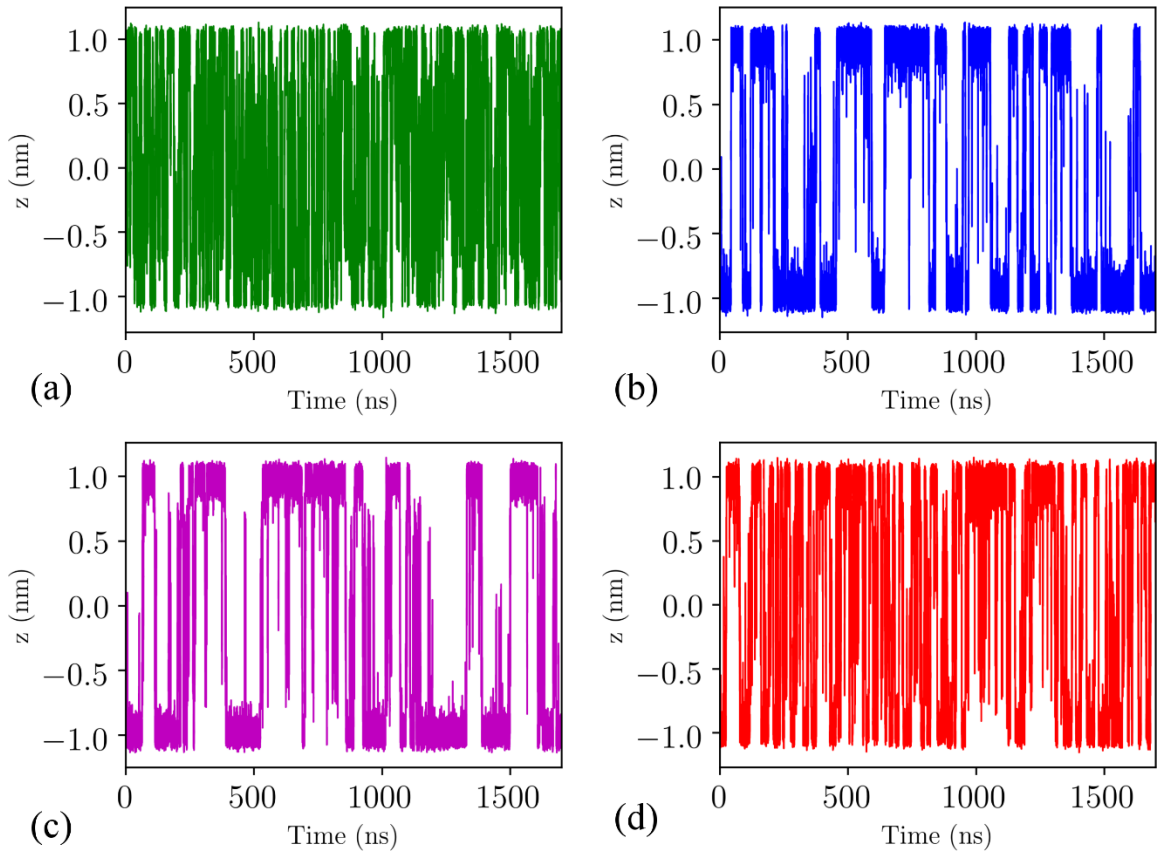


Figure A.3. dNMP position, *z*, relative to the slit centerline as a function of time for the CH3-SAM walls: (a) dCMP, (b) dTMP, (c) dGMP, (d) dAMP.

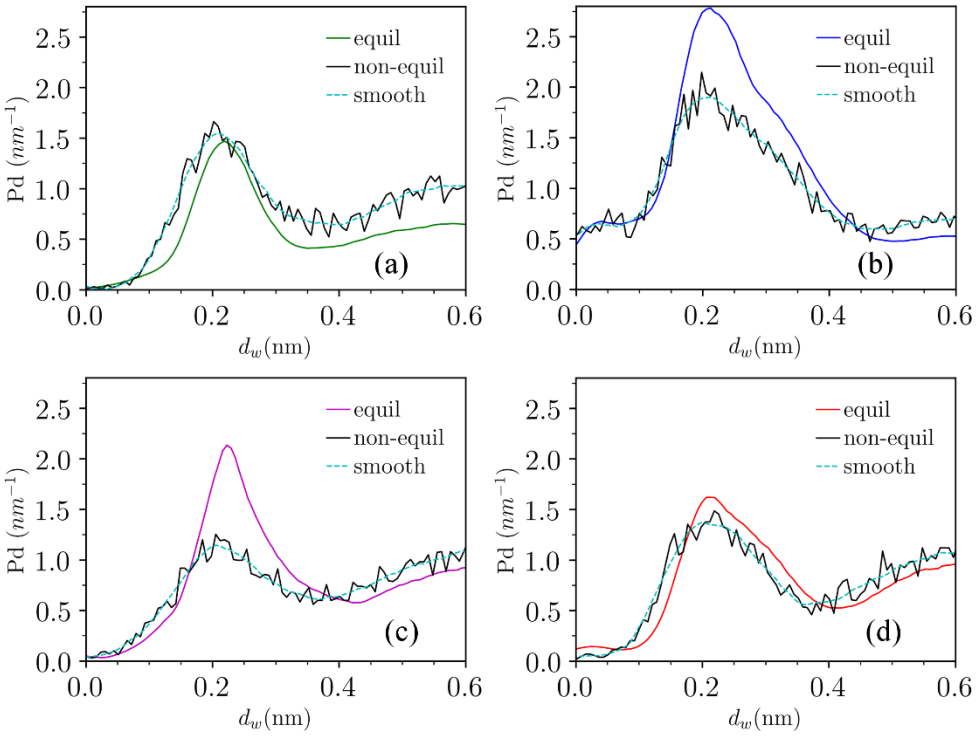


Figure A.4. Comparison of probability distributions as a function of d_w (nm) for dNMPs in the COOCH3-SAM system. Blue curves are from non-equilibrium simulation while the orange colors are obtained from equilibrium PMF simulations. (a) dCMP (b)dTMP (c) dGMP (d) dAMP

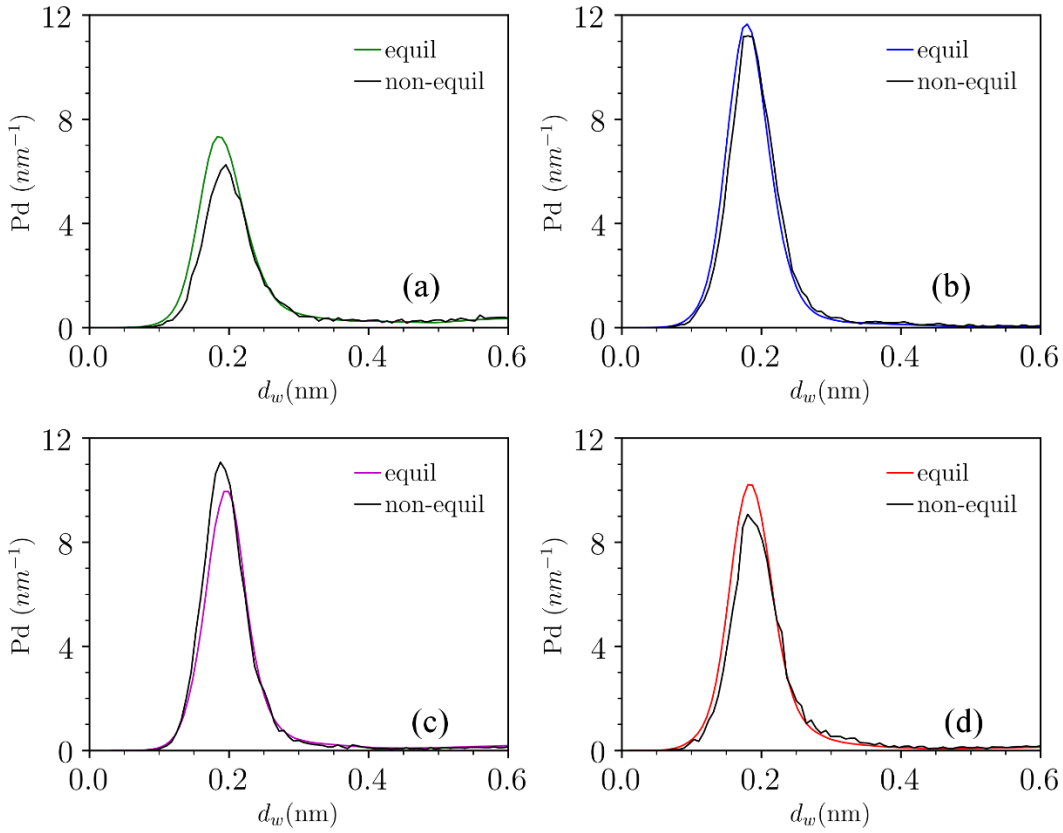


Figure A.5. Comparison of probability distributions as a function of d_w (nm) for dNMPs in the CH3-SAM system. Blue curves are from non-equilibrium simulation while the orange colors are obtained from equilibrium PMF simulations. (a) dCMP (b)dTMP (c) dGMP (d) dAMP

Sodium ion-dNMP association

When nucleotide is surrounded by ions, positive charged sodium ions are likely to seek for negatively charged phosphate. Most of the time one sodium ion was associated, sometimes two or three. Sodium ions can be called counter ions because of its opposite charge to phosphate while chloride ion is called co-ions since it repels negative particles. Associated ions can weaken the repulsion of negative charged particles thus nucleotides are more compact when surrounded by salt solutions. Similarly, the effect of electric field can be weakened as an increasing of associated ions number. Average association number between negatively charged phosphate and sodium ions in solution were calculated in table S1.

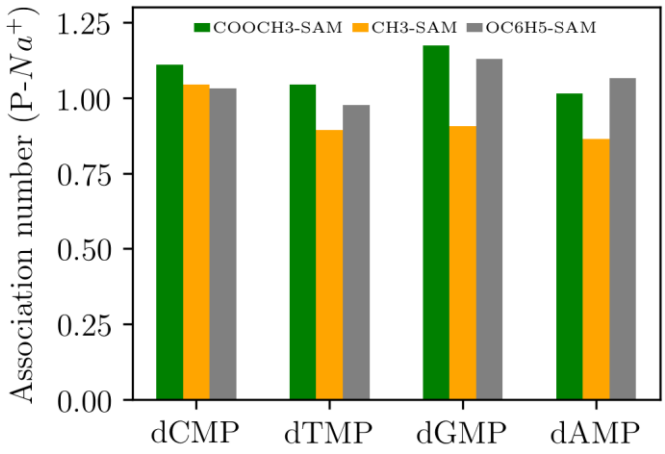


Figure A. 6: Average association number of between negatively charged phosphate in dNMPs and sodium ions in solution.

Hydrogen bonds

The number of hydrogen bonds between water and dNMPs are in the order of dGMP > dAMP > dCMP >dTMP for all systems. In general, dTMP adsorbed to the wall most of the time. Undoubtedly less water is around dTMP because of more contact with wall surface. dTMP base is the most hydrophobic and dGMP is most hydrophilic, which means when they are adsorbed to the wall, least and most OH-O , NH-O or OH-N bonding will be formed for dTMP and dGMP respectively.

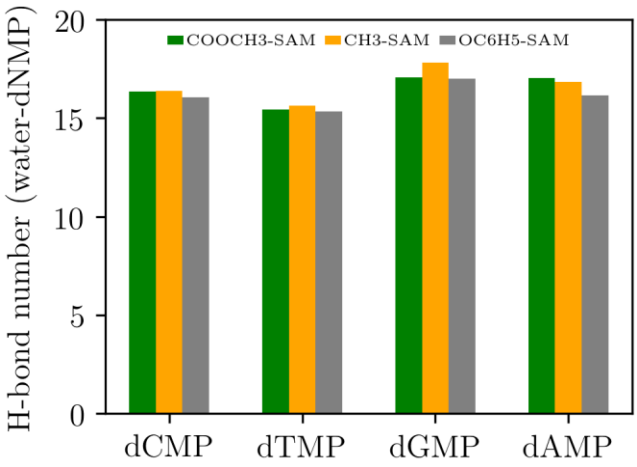


Figure A. 7: Average hydrogen bond number between water and dNMPs for three systems, where hydrogen bonds were determined by the cutoff (0.35nm) for distance of donor(OH or NH) and acceptor(O or N).

Table A.1. Average association number between negatively charged phosphate on dNMPs bases and sodium ions in solution; average hydrogen number between water and dNMPs, and between SAMs walls and dNMPs.

Number on average	dNMP type	COOCH3-SAM	CH3-SAM	OC6H5-SAM
Association (P-Na+)	dCMP	1.10	1.04	1.03
	dTMP	1.04	0.89	0.98
	dGMP	1.17	0.91	1.13
	dAMP	1.01	0.86	1.07
Hydrogen bond (water-dNMP)	dCMP	16.4	16.4	16.1
	dTMP	15.4	15.6	15.3
	dGMP	17.1	17.8	17.0
	dAMP	17.0	16.8	16.2
Hydrogen bond (wall-dNMP)	dCMP	0.06	0	0
	dTMP	0.14	0	0
	dGMP	0.09	0	0
	dAMP	0.07	0	0

APPENDIX B. SUPPLEMENTAL MATERIAL FOR CHAPTER 4

MD force field validation

Download EPD.cif from website http://scripts.iucr.org/cgi-bin/paper?ac6164#suppinfoanchor. PYMOL 1.7 [173] was used to create a supercell of 4*4*4 unit cell and the plugin can be found at https://pymolwiki.org/index.php/Supercell. Residue names were changed to be consistent with our topologies and extra hydrogens were removed (there were six atoms bonded to one carbon atom which was incorrect from the downloaded file). Added box size 4 times the original box and saved as EPD.pdb. NPT simulation was run at 173 K for 18 ns. As seen from cell dimensions, the lattice parameters remained close to experimental values [98]. Three average dihedral angles were examined: d1 = C6-C1-O7-C8, d2 = C1-O7-C8-C16 and d3 =C18-C16-C8-O7. By comparing with the experimental data and reported simulation data [174], MD results from our GROMOS force field parameters are reasonable (Table B.1). The topology files can be downloaded from github: https://github.com/pandatt66/Lignin-Gromos.

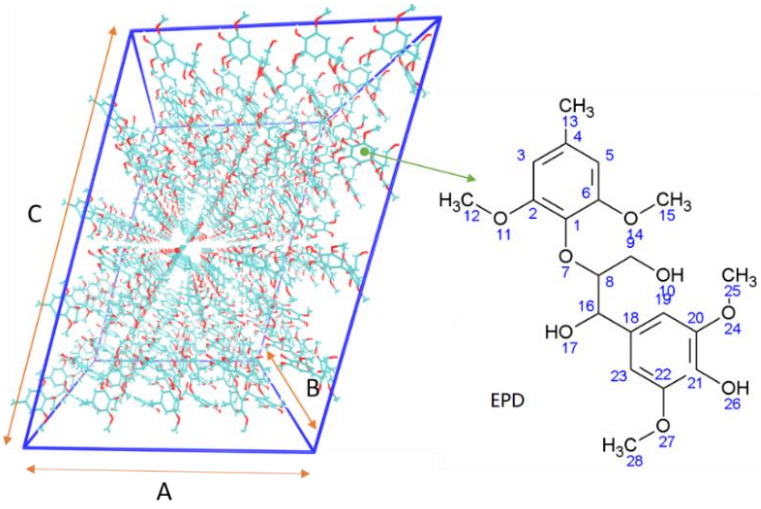


Figure B.1. Simulation box with crystalline EPD molecules.

Table B.1. Unit cell properties of EPD crystal from experiment and MD

Cell dimension	Experiment	MD[1]	MD[*]
A (Å)	8.69	8.693	8.63
B (Å)	8.90	8.832	9.04
C (Å)	13.11	13.583	13.32
α (deg)	73.85	75.00	72.5
β (deg)	86.15	86.42	84.7
γ (deg)	83.06	83.06	83.3
Cell Volume ($Å^3$)	966	1043	983
d_1 (deg)	80.0	77.9 ± 6.3	82.45 ± 0.007
d_2 (deg)	152.8	148.5 ± 5.5	150.8 ± 0.006
d_3 (deg)	70.8	-	72.7 ± 0.005

MD[1] is from reported literature [174] and MD[*] is from our investigation

Lignin monomer interacting with DOPC lipid bilayer

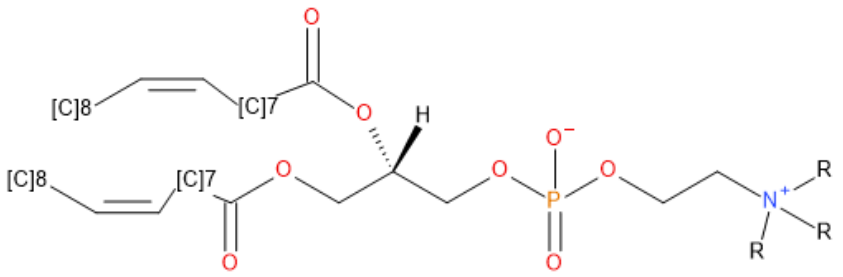


Figure B.2. 1,2-Dioleoyl-sn-glycero-3-phosphocholine (DOPC) lipid, R = CH3

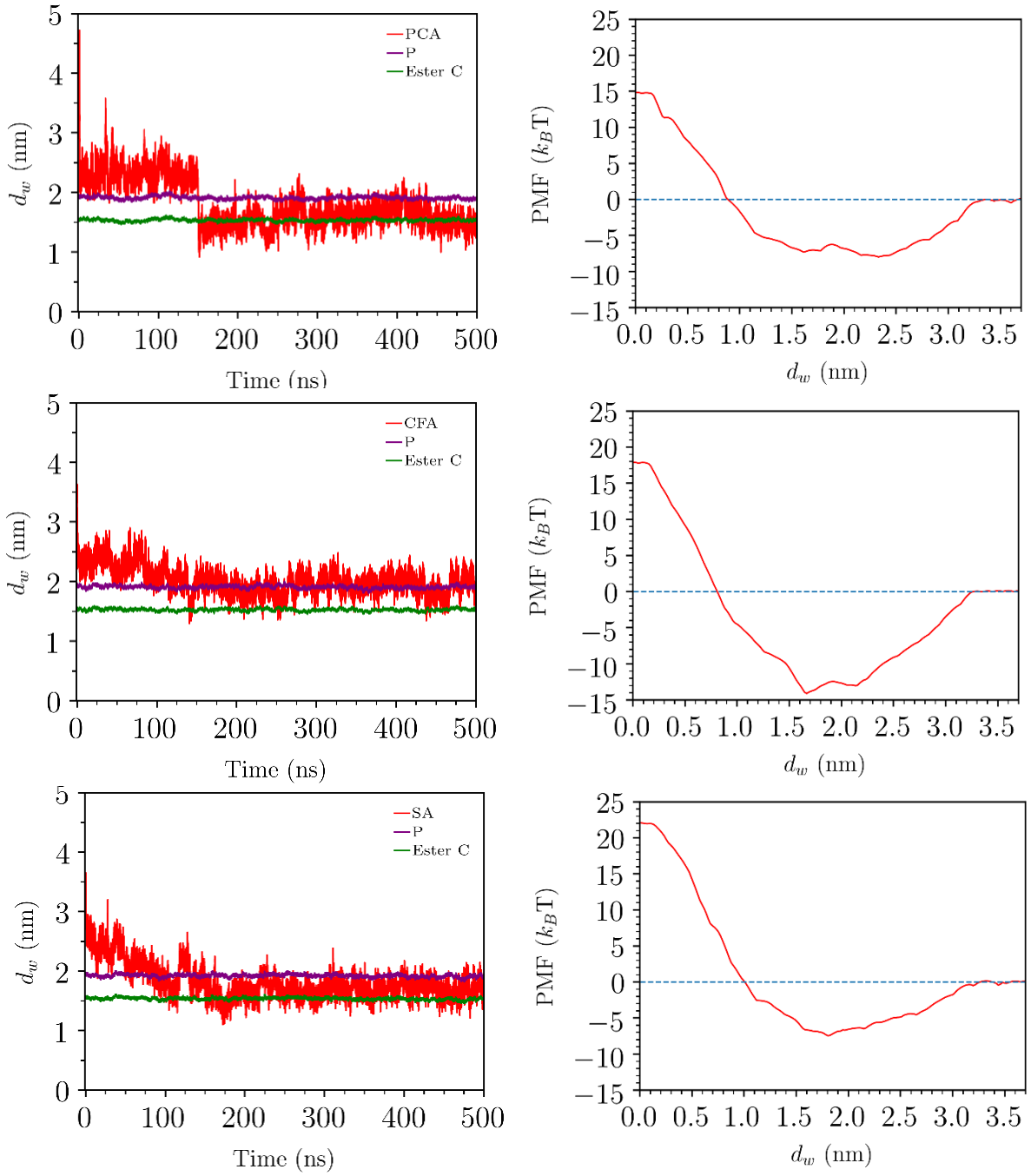


Figure B.3. Left column: Single monoligonol trajectory with DOPC lipid bilayers verse time in nanoseconds. Right column: Potential mean force shown in k_BT as function of distances (dw) from center mass of monolignols (a) PCA (b) CFA (c) SA to DOPC lipid bilayer center.

Potential of mean force indicates a binding interaction between the lipid bilayer and lignin molecules. A single molecule with lipid bilayer is considered as a diluted system since the ratio of

dimer to lipid is 1/128, which is small. In general, lignin molecules are hydrophobic due to their phenyl and methyl group. All monoligonols or dimers are able to reach the surface of lipid bilayer in less than 25ns.

From the result, we can conclude that it is very unlikely for monolignols to flip because of the large energy barrier for PCA, SA and CFA which are about 22 k_BT, 29 k_BT and 31 k_BT respectively. This is in consistent with trajectories of monoligonols. From the positioning, monoligonols are barely adsorb to the surface of DOPC lipid bilayer. It seems PCA went across the phosphate and ester carbons level more often than the SA and CFA. Because PCA has a small size which makes it easier to go through the surface of DOPC. SA has two side groups of OCH3 which makes it a largest molecule among them. However, it has an even smaller PMF than CFA. Besides, the width around PMF minima of PCA and SA are obviously bigger than CFA. The barrier from water to minima of PCA and SA are also very close which are about 7 k_BT while of CFA is about double of that. These phenomena may be explained by the following reasons. First, SA is more hydrophobic than CFA which may dominate other than the molecule size. Second, CFA is kind of special because of the asymmetric geometry. For other type of lipids, for example, dioleoylphosphatidylethanolamine(DOPE), whose headgroups are even smaller and tightly packed would prevent lignin molecules going inside.

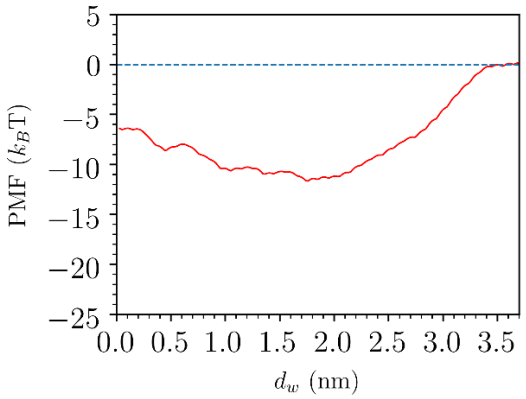


Figure B.4. PMF profile for a lign as a function of the distance (d_w) from bilayer center to terminated –OH group of $X_{benzG\text{-}\beta O4'\text{-}G}$ = 0.286.

PMFs for distance from the bilayer center when adding lignin dimers outside gel phase

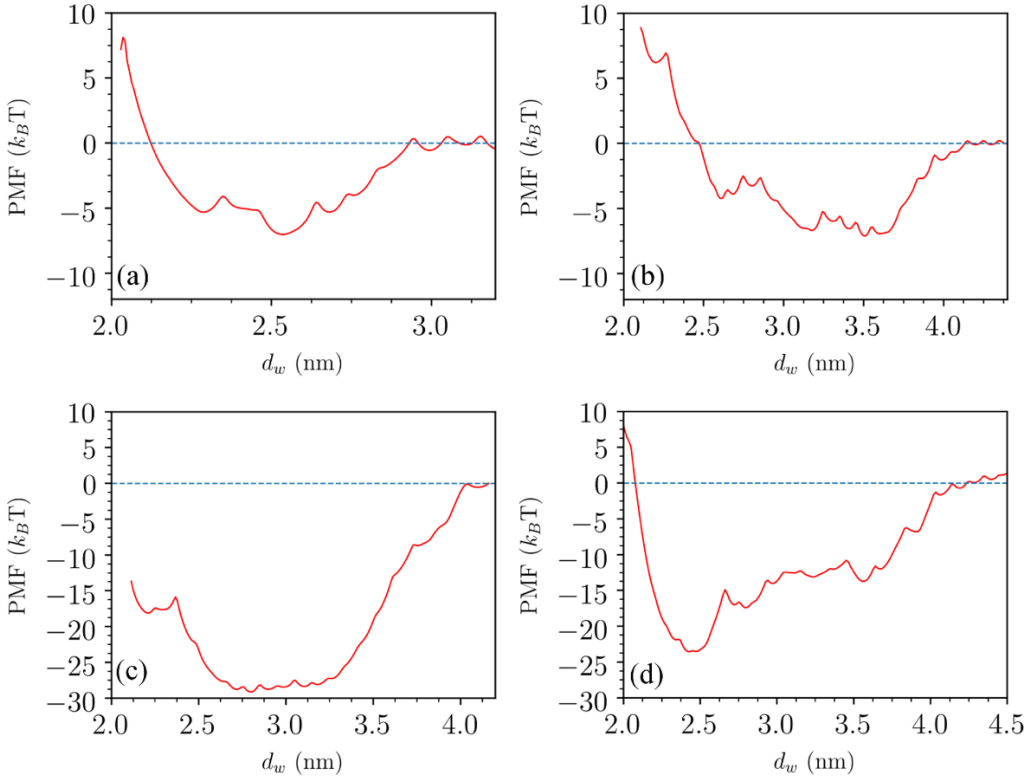


Figure B.5. PMF curves for each of the individual lignin dimers as a function of the distance from the dimer center of mass to the center of the lipid bilayer (d_w). The

(caption cont'd)

reference state is the dimer in bulk water, where the PMF is set to zero. Dimers were put outside gel phase DPPC and the temperature was 250K. (a) and (c) with $X_{G\text{-}\beta O4'\text{-}truncG}$ = 0.158 and 0.289 respectively; (b) and (d) with $X_{benzG\text{-}\beta O4'\text{-}G}$ = 0.158 and 0.289 respectively.

System size effects on lipid transition temperature

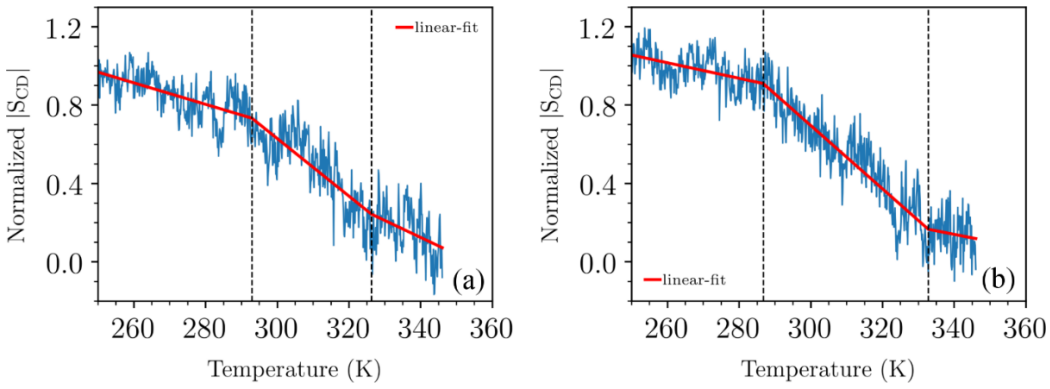


Figure B.6. Normalized deuterium order parameter for selected carbons in the lipid tails of DPPC. (a) G-βO4'-*truncG*, X_{dimer}= 0.158 with 128 DPPC lipids (b) *benz*G-βO4'-G, X_{dimer} with 128 DPPC lipids. The vertical dashed lines indicate the turning points; the transition occurs between these lines and the transition temperature is determined to be midway between them.

Table B.2. Transition temperatures (T_m) and transition widths (ΔT_{trans}) from MD simulations. N_{DPPC} is the number of DPPC lipid molecules. Transition widths are determined by temperature difference between turning points. The transition temperature is defined to be midway between the turning points T_0 and T_1; $T_m = (T_1\text{-}T_0)/2$.

N_{DPPC}	128		512			
X_{dimer}	0.158		0.158		0.289	
dimer type	G-βO4'-*truncG*	*benz*G-βO4'-G	G-βO4'-*truncG*	*benz*G-βO4'-G	G-βO4'-*truncG*	*benz*G-βO4'-G
T_m (K)	309.60±3.8	309.73±2.2	309.8±1.8	309.3±1.0	291.5±1.6	292.2±2.0
ΔT_{trans} (K)	33.3	46.1	30.0	35.1	38.6	52.3

* Coexistence method.

For a 128 DPPC lipids system with X_{dimer} = 0.158, the order parameter plots were nosier and more sensitive to the initial guess for the fitting parameters. We were not confident fitting the lines

to get a transition temperature because different initial guesses lead to varied results. For a 128 DPPC lipids system with X_{dimer} = 0.289, the noise was even worse, and in some case the transition temperature could not be determined at all. Using A 512 DPPC lipid system improved the data a lot and the turning points determined from a linear fit were not sensitive to initial guesses for the fits. Therefore larger simulation system size leads to better estimation of transition temperature of the lipid bilalyer.

Transition temperature results when lignin dimers were added outside and equilibrated with the lipid bilayer in the fluid phase

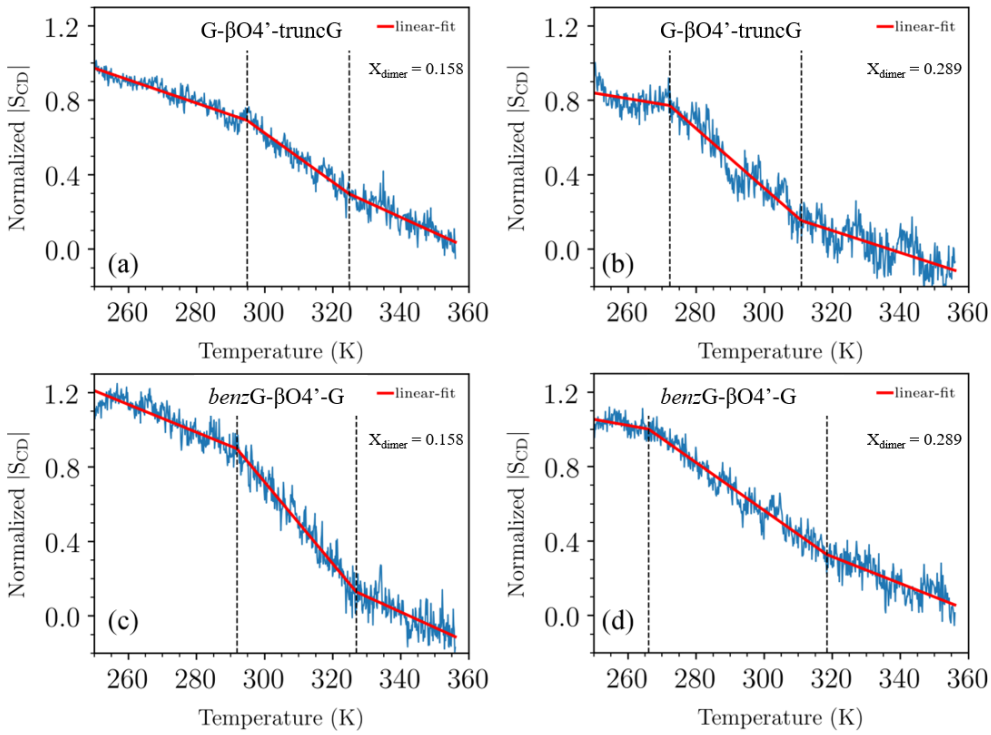


Figure B.7. Normalized deuterium order parameter, $|S_{CD}|$, versus temperature and 3-

(caption cont'd)

piece linear fits. (a) 128 DPPC lipids with G-βO4'-truncG dimers of concentration X_{dimer} = 0.158. (b) and (c) 512 DPPC lipids with G-βO4'-truncG dimers of concentration X_{dimer} = 0.158 and 0.289, respectively, (d) and (e) 512 lipids with benzG-βO4'-G dimers of concentration X_{dimer} = 0.158 and 0.289 respectively. The vertical dashed lines indicate the turning points; the transition occurs between these lines and the transition temperature is determined to be midway between them. All these results are from dimers started inside the gel phase bilayer.

Procedure for determination of turning points in the variation of the averaged normalized deuterium order parameter with temeprature

Normalized deuterium order parameter was linearly fit into three lines using python's Scipy package named 'optimize.curve_fit'. Code for a fit is shown below where x is temperature and y is the order parameter. One needs to make initial guesses for x0, x1, b, k1, k2, k3 before the fitting, but any reasonable set of initial guesses will give the same result as long as the data is basically piecewise linear with small fluctuations.

```
import numpy as np
import uncertainties

def piecewise_linear3(x, x0, x1, b, k1, k2, k3):
     condlist =  [x < x0, (x >= x0) & (x < x1), x >= x1]
     funclist =  [lambda x: k1*x + b,
                  lambda x: k1*x + b + k2*(x-x0),
                  lambda x: k1*x + b + k2*(x-x0) + k3*(x - x1)]
     return np.piecewise(x, condlist, funclist)

p, e = optimize.curve_fit(piecewise_linear3, x, y,
                          p0= [290, 320, 1, -1, -10, -1]
(T1,T2) = uncertainties.correlated_values( [p [0],p [1]],e [:2, :2])
Tm = (T1+T2)/2
```

Order prameters for gel phase DPPC with and without lignin dimers added outside the bilayer

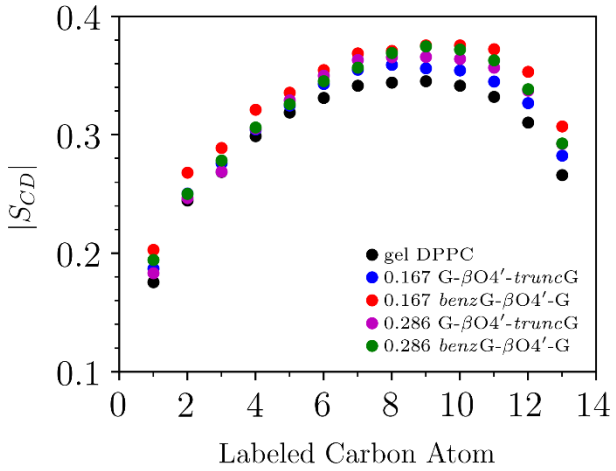


Figure B.8. Deuterium order parameter $|S_{CD}|$ for lipid tail sn1 in gel phase(250K). Black spheres represent the pure gel phase DPPC bilayer and the others show gel phase DPPC with different concentrations of dimers added outside bilayer after running a 200ns of equilibration.

$|S_{CD}|$ of pure gel phase DPPC and gel phase DPPC with dimers in solution do not differ much. With higher concentration of dimer, the gel phase DPPC is even more ordered. This is probably due to a dehydration of lipids when lignin dimers are covering the bilayer surface. [175] This supports the idea that starting dimers outside of lipids bilayer is a reasonable approximation since the order parameter was not largely affected by different concentrations of lignin dimers on the surface.

APPENDIX C. SUPPLEMENTAL MATERIAL FOR CHAPTER 6

A modified force field parameter is validated by comparing the unit cell properties with from experiment as show in Table C.1. Tuncated dimer crystal unit structure can be downloaded from Cambridge structural database [170]. Then we had the unit 10, 2 and 5 times replicated in x,y and z direction for a larger system followed by energy minimization and NPT simulation. After equilibration, the simulation box dimentions as well as three average dihedral angles were examined: d1= $C_\zeta - C_\varepsilon - O_5 - C_\delta, d2 = C_\varepsilon - O_5 - C_\delta - C_\gamma$ and d3 = $O_5 - C_\delta - C_\gamma - C_\beta$(labeled atoms see Figure 6.1) We tested with parameters generated from Swisspara [144] too. But dihedral angles have a better approximation from modified parameter based on Ligninbuilder for natural dimer.

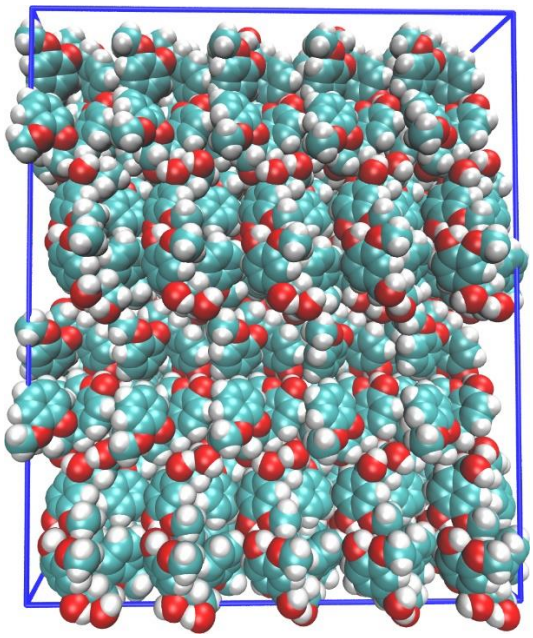


Figure C.1. Simulation box with crystalline G-βO4'-*trunc*G dimer molecules.

Table C.1. Unit cell properties of G-βO4'-*trunc*G crystal from experiment and MD

Cell dimension	Experiment	MD-swisspara	MD-lignbuilder(modified)
A (Å)	5.08	5.11	5.14
B (Å)	30.11	29.88	30.9
C (Å)	10.40	10.59	10.2
α (deg)	90.0	90.0	90
β (deg)	96.16	95.30	96.42
γ (deg)	90.0	90.0	90.0
Cell Volume ($Å^3$)	1581.59	1610.04	1609.8
d_1 (deg)	72.14	69.2±0.01	78.1±0.02
d_2 (deg)	141.90	148.6 ± 0.006	145.9±0.01
d_3 (deg)	50.52	60.06± 0.005	50.3±0.02

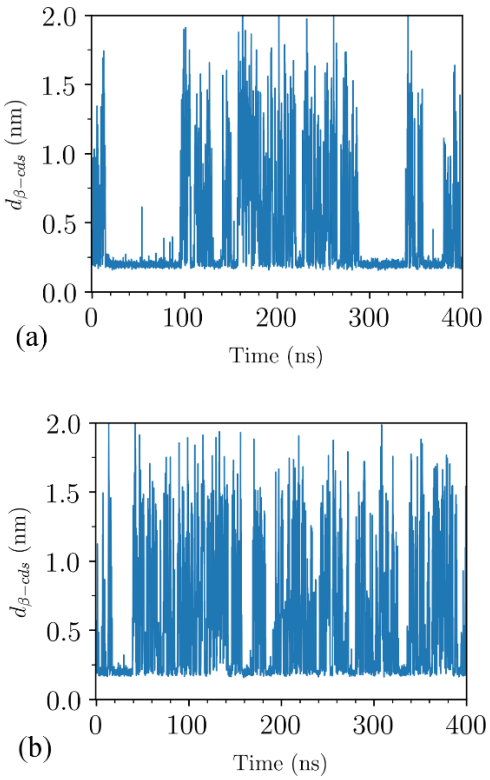


Figure C.2. Distance between two β-CDs based on their center of mass with a inclusion of (a) G-βO4'-G (b) G-βO4'-truncG over 400ns simulation.

APPENDIX D. PERMISSIONS

Title: Experimental and Molecular Dynamics Simulation Study of the Effects of Lignin Dimers on the Gel-to-Fluid Phase Transition in DPPC Bilayers

Author: Xinjie Tong, Mahsa Moradipour, Brian Novak, et al

Publication: The Journal of Physical Chemistry B

Publisher: American Chemical Society

Date: Oct 1, 2019

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

REFERENCES

1. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins.* Nature, 1977. **267**(5612): p. 585.

2. Fersht, A.R., *Profile of martin karplus, michael levitt, and arieh warshel, 2013 nobel laureates in chemistry.* Proceedings of the National Academy of Sciences, 2013. **110**(49): p. 19656-19657.

3. Karplus, M. and R. Lavery, *Significance of molecular dynamics simulations for life sciences.* Israel Journal of Chemistry, 2014. **54**(8-9): p. 1042-1051.

4. Alonso, H., A.A. Bliznyuk, and J.E. Gready, *Combining docking and molecular dynamic simulations in drug design.* Medicinal research reviews, 2006. **26**(5): p. 531-568.

5. Karplus, M. and J. Kuriyan, *Molecular dynamics and protein function.* Proceedings of the National Academy of Sciences, 2005. **102**(19): p. 6679-6685.

6. Selvaraj, C., et al., *Molecular dynamics simulations and applications in computational toxicology and nanotoxicology.* Food and Chemical Toxicology, 2018. **112**: p. 495-506.

7. Vologodskii, A.V. and N.R. Cozzarelli, *Supercoiling, knotting, looping and other large-scale conformational properties of DNA.* Current Opinion in Structural Biology, 1994. **4**(3): p. 372-375.

8. Chial, H., *DNA sequencing technologies key to the Human Genome Project.* Nature Education, 2008. **1**(1): p. 219.

9. Branton, D., et al., *The potential and challenges of nanopore sequencing*, in *Nanoscience and technology: A collection of reviews from Nature Journals*. 2010, World Scientific. p. 261-268.

10. Boija, E. and G. Johansson, *Interactions between model membranes and lignin-related compounds studied by immobilized liposome chromatography.* Biochimica Et Biophysica Acta-Biomembranes, 2006. **1758**(5): p. 620-626.

11. Vinardell, M.P. and M. Mitjans, *Lignins and their derivatives with beneficial effects on human health.* International journal of molecular sciences, 2017. **18**(6): p. 1219.

12. Ugartondo, V., M. Mitjans, and M.P. Vinardell, *Comparative antioxidant and cytotoxic effects of lignins from different sources.* Bioresource technology, 2008. **99**(14): p. 6683-6687.

13. Huijgen, W., et al., *Characteristics of wheat straw lignins from ethanol-based organosolv treatment.* Industrial Crops and Products, 2014. **59**: p. 85-95.

14. Maazaoui, R. and R. Abderrahim, *Applications of cyclodextrins: formation of inclusion complexes and their characterization.* International Journal of Advanced Research, 2015. **3**(2): p. 1030-1030.

15. Rajewski, R.A. and V.J. Stella, *Pharmaceutical applications of cyclodextrins. 2. In vivo drug delivery.* Journal of pharmaceutical sciences, 1996. **85**(11): p. 1142-1169.

16. Tiwari, G., R. Tiwari, and A.K. Rai, *Cyclodextrins in delivery systems: Applications.* Journal of Pharmacy and Bioallied Sciences, 2010. **2**(2): p. 72.

17. Astray, G., et al., *A review on the use of cyclodextrins in foods.* Food Hydrocolloids, 2009. **23**(7): p. 1631-1640.

18. Hinze, W.L., *Applications of cyclodextrins in chromatographic separations and purification methods.* Separation and Purification methods, 1981. **10**(2): p. 159-237.

19. Tidemand, K.D., et al., *Computational Investigation of Enthalpy–Entropy Compensation in Complexation of Glycoconjugated Bile Salts with β-Cyclodextrin and Analogs.* The Journal of Physical Chemistry B, 2014. **118**(37): p. 10889-10897.

20. Zhang, H., et al., *Cooperative binding of cyclodextrin dimers to isoflavone analogues elucidated by free energy calculations.* The Journal of Physical Chemistry C, 2014. **118**(13): p. 7163-7173.

21. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.* SoftwareX, 2015. **1**: p. 19-25.

22. Grubmüller, H., et al., *Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions.* Molecular Simulation, 1991. **6**(1-3): p. 121-142.

23. Mudi, A. and C. Chakravarty, *Effect of the Berendsen thermostat on the dynamical properties of water.* Molecular Physics, 2004. **102**(7): p. 681-685.

24. Hünenberger, P.H., *Thermostat algorithms for molecular dynamics simulations*, in *Advanced computer simulation*. 2005, Springer. p. 105-149.

25. Evans, D.J. and B.L. Holian, *The nose–hoover thermostat.* The Journal of chemical physics, 1985. **83**(8): p. 4069-4074.

26. Berendsen, H.J., et al., *Molecular dynamics with coupling to an external bath.* The Journal of chemical physics, 1984. **81**(8): p. 3684-3690.

27. Oostenbrink, C., et al., *A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6.* Journal of computational chemistry, 2004. **25**(13): p. 1656-1676.

28. Hess, B., et al., *LINCS: a linear constraint solver for molecular simulations.* Journal of computational chemistry, 1997. **18**(12): p. 1463-1472.

29. Kästner, J., *Umbrella sampling.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2011. **1**: p. 932–942.

30. Venkatesan, B.M. and R. Bashir, *Nanopore sensors for nucleic acid analysis.* Nature nanotechnology, 2011. **6**: p. 615.

31. Laszlo, A.H., et al., *Decoding long nanopore sequencing reads of natural DNA.* Nature biotechnology, 2014. **32**(8): p. 829.

32. Wang, Y., Q. Yang, and Z. Wang, *The evolution of nanopore sequencing.* Frontiers in genetics, 2015. **5**: p. 449.

33. Haque, F., et al., *Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA.* Nano Today, 2013. **8**(1): p. 56-74.

34. Movileanu, L., et al., *Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore.* Nature biotechnology, 2000. **18**(10): p. 1091.

35. Derrington, I.M., et al., *Subangstrom single-molecule measurements of motor proteins using a nanopore.* Nature biotechnology, 2015. **33**(10): p. 1073.

36. Laszlo, A.H., I.M. Derrington, and J.H. Gundlach, *MspA nanopore as a single-molecule tool: From sequencing to SPRNT.* Methods, 2016. **105**: p. 75-89.

37. Feng, J., et al., *Identification of single nucleotides in MoS 2 nanopores.* Nature nanotechnology, 2015. **10**(12): p. 1070.

38. dela Torre, R., et al., *Fabrication and characterization of solid-state nanopore arrays for high-throughput DNA sequencing.* Nanotechnology, 2012. **23**: p. 385308.

39. Feng, Y., et al., *Nanopore-based fourth-generation DNA sequencing technology.* Genomics, proteomics & bioinformatics, 2015. **13**(1): p. 4-16.

40. Liu, S., et al., *Boron Nitride Nanopores: Highly Sensitive DNA Single-Molecule Detectors.* Advanced Materials, 2013. **25**: p. 4549-4554.

41. Garaj, S., et al., *Graphene as a subnanometre trans-electrode membrane.* Nature, 2010. **467**(7312): p. 190-U73.

42. Farimani, A.B., K. Min, and N.R. Aluru, *DNA Base Detection Using a Single-Layer MoS2.* Acs Nano, 2014. **8**(8): p. 7914-7922.

43. Lim, B.N., et al., *Directed evolution of nucleotide-based libraries using lambda exonuclease.* Biotechniques, 2012. **53**: p. 357–364.

44. Novak, B.R., et al., *Distinguishing single DNA nucleotides based on their times of flight through nanoslits: A molecular dynamics simulation study.* Journal of Physical Chemistry B, 2013. **117**(12): p. 3271-3279.

45. Sampath, G., *DNA sequencing with stacked nanopores and exonuclease: A simulation-based analysis.* ELECTROPHORESIS, 2016. **37**(17-18): p. 2429-2434.

46. Hosseini, S., et al., *Recent advances in surface functionalization techniques on polymethacrylate materials for optical biosensor applications.* Analyst, 2014. **139**(12): p. 2933-43.

47. Wochnowski, C., M.S. Eldin, and S. Metev, *UV-laser-assisted degradation of poly (methyl methacrylate).* Polymer degradation and stability, 2005. **89**(2): p. 252-264.

48. Ulman, A., *Formation and structure of self-assembled monolayers.* Chemical reviews, 1996. **96**(4): p. 1533-1554.

49. Devi, J.M., *A simulation study on the thermal and wetting behavior of alkane thiol SAM on gold (111) surface.* Progress in Natural Science: Materials International, 2014. **24**(4): p. 405-411.

50. Blackburn, G.M. and M.J. Gait, eds. *Nucleic Acids in Chemistry and Biology*. 1990, IRL Press: Oxford.

51. Latour Jr, R.A. and C.J. Rini, *Theoretical analysis of adsorption thermodynamics for hydrophobic peptide residues on SAM surfaces of varying functionality.* Journal of Biomedical Materials Research: An Official Journal of The Society for Biomaterials, The Japanese Society for Biomaterials, and The Australian Society for Biomaterials and the Korean Society for Biomaterials, 2002. **60**(4): p. 564-577.

52. Jorgensen, W.L. and C. Jenson, *Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density.* Journal of Computational Chemistry, 1998. **19**: p. 1179–1186.

53. O'Boyle, N.M., et al., *Open Babel: An open chemical toolbox.* Journal of Cheminformatics, 2011. **3**: p. 33.

54. Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.* Journal of computational chemistry, 2010. **31**: p. 671–690.

55. Devi, J.M., *A simulation study on the thermal and wetting behavior of alkane thiol SAM on gold (111) surface.* Progress in Natural Science: Materials International, 2014. **24**: p. 405–411.

56. Love, J.C., et al., *Self-assembled monolayers of thiolates on metals as a form of nanotechnology.* Chemical reviews, 2005. **105**: p. 1103–1170.

57. Abraham, M.J., et al., *GROMACS User Manual version 5.1*. 2015.

58. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling.* Journal of Chemical Physics, 2007. **126**(1).

59. Xia, K., et al., *Electrophoretic transport of single DNA nucleotides through nanoslits: a molecular dynamics simulation study*. The Journal of Physical Chemistry B, 2015. **119**(35): p. 11443-11458.

60. Shih, P., et al., *Hydrophobicities of the nucleic acid bases: distribution coefficients from water to cyclohexane*. Journal of molecular biology, 1998. **280**(3): p. 421-430.

61. Efron, B., *Bootstrap methods: another look at the jackknife,* in *Breakthroughs in statistics*. 1992, Springer. p. 569-593.

62. Schneider, G.F., et al., *Tailoring the hydrophobicity of graphene for its use as nanopores for DNA translocation.* Nature communications, 2013. **4**: p. 2619.

63. Yablonski, J., *Plant and Animal Cells: Understanding the Differences Between Plant and Animal Cells*. 2005: The Rosen Publishing Group.

64. Darnell, J.E., H.F. Lodish, and D. Baltimore, *Molecular cell biology*. Vol. 2. 1990: Scientific American Books New York.

65. Cooper, G.M. and R.E. Hausman, *The cell: Molecular approach*. 2004: Medicinska naklada.

66. Eze, M.O., *Phase transitions in phospholipid bilayers: lateral phase separations play vital roles in biomembranes.* Biochemical education, 1991. **19**(4): p. 204-208.

67. Pouteau, C., et al., *Antioxidant properties of lignin in polypropylene.* Polymer Degradation and Stability, 2003. **81**(1): p. 9-18.

68. García, A., et al., *Lignin as natural radical scavenger. Effect of the obtaining and purification processes on the antioxidant behaviour of lignin.* Biochemical engineering journal, 2012. **67**: p. 173-185.

69. Dong, X., et al., *Antimicrobial and antioxidant activities of lignin from residue of corn stover to ethanol production.* Industrial Crops and Products, 2011. **34**(3): p. 1629-1634.

70. Barber, M.S., V.S. McConnell, and B.S. DeCaux, *Antimicrobial intermediates of the general phenylpropanoid and lignin specific pathways.* Phytochemistry, 2000. **54**(1): p. 53-56.

71. Sato, S., et al., *Effect of lignin-derived lignophenols on hepatic lipid metabolism in rats fed a high-fat diet.* Environmental Toxicology and Pharmacology, 2012. **34**(2): p. 228-234.

72. Witzler, M., et al., *Lignin-Derived Biomaterials for Drug Release and Tissue Engineering.* Molecules, 2018. **23**(8): p. 1885.

73. Figueiredo, P., et al., *In vitro evaluation of biodegradable lignin-based nanoparticles for drug delivery and enhanced antiproliferation effect in cancer cells.* Biomaterials, 2017. **121**: p. 97-108.

74. Baurhoo, B., C. Ruiz-Feria, and X. Zhao, *Purified lignin: Nutritional and health impacts on farm animals—A review.* Animal Feed Science and Technology, 2008. **144**(3-4): p. 175-184.

75. Pan, X., et al., *Organosolv ethanol lignin from hybrid poplar as a radical scavenger: relationship between lignin structure, extraction conditions, and antioxidant activity.* Journal of agricultural and food chemistry, 2006. **54**(16): p. 5806-5813.

76. Azadfar, M., et al., *Structural characterization of lignin: A potential source of antioxidants guaiacol and 4-vinylguaiacol.* International journal of biological macromolecules, 2015. **75**: p. 58-66.

77. Xü, Y., et al., *Protein-mediated lipid transfer. The effects of lipid-phase transition and of charged lipids.* Biochemical Journal, 1983. **213**(1): p. 21-24.

78. Jain, M.K. and N.M. Wu, *Effect of small molecules on the dipalmitoyl lecithin liposomal bilayer: III. Phase transition in lipid bilayer.* The Journal of Membrane Biology, 1977. **34**(1): p. 157-201.

79. Saija, A., et al., *Flavonoids as antioxidant agents: importance of their interaction with biomembranes.* Free Radical Biology and Medicine, 1995. **19**(4): p. 481-486.

80. Hendrich, A.B., et al., *Differential interaction of Sophora isoflavonoids with lipid bilayers.* European journal of pharmaceutical sciences, 2002. **16**(3): p. 201-208.

81. Ojogun, V., et al., *Partitioning of homologous nicotinic acid ester prodrugs (nicotinates) into dipalmitoylphosphatidylcholine (DPPC) membrane bilayers.* Colloids Surf B Biointerfaces, 2010. **78**(1): p. 75-84.

82. Zhao, L., et al., *DSC and EPR investigations on effects of cholesterol component on molecular interactions between paclitaxel and phospholipid within lipid bilayer membrane.* 2007. **338**(1-2): p. 258-266.

83. Leekumjorn, S. and A.K. Sum, *Molecular studies of the gel to liquid-crystalline phase transition for fully hydrated DPPC and DPPE bilayers.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2007. **1768**(2): p. 354-365.

84. Kowalik, B., et al., *Combination of MD simulations with two-state kinetic rate modeling elucidates the chain melting transition of phospholipid bilayers for different hydration levels.* The Journal of Physical Chemistry B, 2015. **119**(44): p. 14157-14167.

85. Qin, S.S., Z.W. Yu, and Y.X. Yu, *Structural characterization on the gel to liquid-crystal phase transition of fully hydrated DSPC and DSPE bilayers.* The journal of physical chemistry. B, 2009. **113**(23): p. 8114-8123.

86. Coppock, P.S. and J.T. Kindt, *Determination of phase transition temperatures for atomistic models of lipids from temperature-dependent stripe domain growth kinetics.* The Journal of Physical Chemistry B, 2010. **114**(35): p. 11468-11473.

87. Waheed, Q., R. Tjörnhammar, and O. Edholm, *Phase Transitions in Coarse-Grained Lipid Bilayers Containing Cholesterol by Molecular Dynamics Simulations.* Biophysical journal, 2012.

88. Zhang, Y., et al., *A coarse-grained molecular dynamics investigation of the phase behavior of DPPC/cholesterol mixtures.* Chemistry and physics of lipids, 2015. **185**: p. 88-98.

89. Wang, Y., et al., *DPPC-cholesterol phase diagram using coarse-grained Molecular Dynamics simulations.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2016. **1858**(11): p. 2846-2857.

90. Duncan, S.L., I.S. Dalal, and R.G. Larson, *Molecular dynamics simulation of phase transitions in model lung surfactant monolayers.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2011. **1808**(10): p. 2450-2465.

91. Liu, C.-J., Y.-C. Miao, and K.-W. Zhang, *Sequestration and Transport of Lignin Monomeric Precursors.* Molecules, 2011. **16**(1): p. 710-727.

92. Miao, Y.-C. and C.-J. Liu, *ATP-binding cassette-like transporters are involved in the transport of lignin precursors across plasma and vacuolar membranes.* Proceedings of the National Academy of Sciences, 2010. **107**(52): p. 22728-22733.

93. Perkins, M., R.A. Smith, and L. Samuels, *The transport of monomers during lignification in plants: anything goes but how?* Current Opinion in Biotechnology, 2019. **56**: p. 69-74.

94. Sibout, R. and H. Höfte, *Plant Cell Biology: The ABC of Monolignol Transport.* Current Biology, 2012. **22**(13): p. R533-R535.

95. Besombes, S. and K. Mazeau, *The cellulose/lignin assembly assessed by molecular modeling. Part 1: adsorption of a threo guaiacyl β-O-4 dimer onto a Iβ cellulose whisker.* Plant Physiology and Biochemistry, 2005. **43**(3): p. 299-308.

96. Demetzos, C., *Differential Scanning Calorimetry (DSC): a tool to study the thermal behavior of lipid bilayers and liposomal stability.* J Liposome Res, 2008. **18**(3): p. 159-73.

97. Malde, A.K., et al., *An automated force field topology builder (ATB) and repository: version 1.0.* Journal of chemical theory and computation, 2011. **7**(12): p. 4026-4037.

98. Langer, V., K. Lundquist, and G.E. Miksche, *erythro-2-(2, 6-Dimethoxy-4-methylphenoxy)-1-(4-hydroxy-3, 5-dimethoxyphenyl) propane-1, 3-diol.* Acta Crystallographica Section E, 2005. **61**(4): p. o1001-o1003.

99. Lindahl, E., B. Hess, and D. Van Der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* Molecular modeling annual, 2001. **7**(8): p. 306-317.

100. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling.* The Journal of chemical physics, 2007. **126**(1): p. 014101.

101. Schmid, N., et al., *Definition and testing of the GROMOS force-field versions 54A7 and 54B7.* European biophysics journal, 2011. **40**(7): p. 843.

102. Berendsen, H.J., et al., *Interaction models for water in relation to protein hydration*, in *Intermolecular forces*. 1981, Springer. p. 331-342.

103. Kästner, J., *Umbrella sampling.* Wiley Interdisciplinary Reviews: Computational Molecular Science, 2011. **1**(6): p. 932-942.

104. Souaille, M. and B.t. Roux, *Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations.* Computer physics communications, 2001. **135**(1): p. 40-57.

105. Zocher, F., et al., *Local partition coefficients govern solute permeability of cholesterol-containing membranes.* Biophysical journal, 2013. **105**(12): p. 2760-2770.

106. Jakobtorweihen, S., et al., *Predicting solute partitioning in lipid bilayers: Free energies and partition coefficients from molecular dynamics simulations and COSMOmic.* The Journal of chemical physics, 2014. **141**(4): p. 07B622_1.

107. Feller, S.E. and R.W. Pastor, *Constant surface tension simulations of lipid bilayers: the sensitivity of surface areas and compressibilities.* The Journal of chemical physics, 1999. **111**(3): p. 1281-1287.

108. Allen, M.P. and D.J. Tildesley, *Computer simulation of liquids*. 2017: Oxford university press.

109. Vermeer, L.S., et al., *Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with 2 H NMR experiments.* European Biophysics Journal, 2007. **36**(8): p. 919-931.

110. Vega, C., E. Sanz, and J. Abascal, *The melting temperature of the most common models of water.* The Journal of chemical physics, 2005. **122**(11): p. 114507.

111. Neale, C. and R. Pomès, *Sampling errors in free energy simulations of small molecules in lipid bilayers.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2016. **1858**(10): p. 2539-2548.

112. Doktorova, M., et al., *A new computational method for membrane compressibility: Bilayer mechanical thickness revisited.* Biophysical journal, 2019. **116**(3): p. 487-502.

113. Buchoux, S., *FATSLiM: a fast and robust software to analyze MD simulations of membranes.* Bioinformatics, 2016. **33**(1): p. 133-134.

114. Tong, X., et al., *Experimental and Molecular Dynamics Simulation Study of the Effects of Lignin Dimers on the Gel-to-Fluid Phase Transition in DPPC Bilayers.* The Journal of Physical Chemistry B, 2019. **123**(39): p. 8247-8260.

115. Biltonen, R.L. and D. Lichtenberg, *The use of differential scanning calorimetry as a tool to characterize liposome preparations.* Chemistry and physics of lipids, 1993. **64**(1-3): p. 129-142.

116. Awasthi, N. and J.S. Hub, *Simulations of pore formation in lipid membranes: reaction coordinates, convergence, hysteresis, and finite-size effects.* Journal of chemical theory and computation, 2016. **12**(7): p. 3261-3269.

117. Mirjalili, V. and M. Feig, *Density-biased sampling: a robust computational method for studying pore formation in membranes.* Journal of chemical theory and computation, 2014. **11**(1): p. 343-350.

118. Lin, S.Y. and C.W. Dence, *Methods in lignin chemistry*. 2012: Springer Science & Business Media.

119. Janshekar, H., C. Brown, and A. Fiechter, *Determination of biodegraded lignin by ultraviolet spectrophotometry.* Analytica Chimica Acta, 1981. **130**(1): p. 81-91.

120. Harton, S.E., et al., *Evidence for complex molecular architectures for solvent-extracted lignins.* ACS Macro Letters, 2012. **1**(5): p. 568-573.

121. Charlier, L. and K. Mazeau, *Molecular modeling of the structural and dynamical properties of secondary plant cell walls: influence of lignin chemistry.* The Journal of Physical Chemistry B, 2012. **116**(14): p. 4163-4174.

122. Brebu, M. and C. Vasile, *Thermal degradation of lignin—a review.* Cellulose Chemistry & Technology, 2010. **44**(9): p. 353.

123. Cannatelli, M.D. and A.J. Ragauskas, *Conversion of lignin into value-added materials and chemicals via laccase-assisted copolymerization.* Applied microbiology and biotechnology, 2016. **100**(20): p. 8685-8691.

124. Beste, A., *ReaxFF study of the oxidation of softwood lignin in view of carbon fiber production.* Energy & Fuels, 2014. **28**(11): p. 7007-7013.

125. Nimz, H., *The Chemistry of Lignin. Von I. A. Pearl. Edward Arnold (Publishers) Ltd., London 1967. Marcel Dekker, Inc., New York 1. Aufl., XI, 339 S., 24 Abb., geb. $ 15.75.* Angewandte Chemie, 1968. **80**(8): p. 328-328.

126. Cazacu, G., M. Capraru, and V.I. Popa, *Advances concerning lignin utilization in new materials*, in *Advances in natural polymers*. 2013, Springer. p. 255-312.

127. Espinoza-Acosta, J.L., et al., *Antioxidant, antimicrobial, and antimutagenic properties of technical lignins and their applications.* BioResources, 2016. **11**(2): p. 5452-5481.

128. Lora, J.H. and W.G. Glasser, *Recent industrial applications of lignin: a sustainable alternative to nonrenewable materials.* Journal of Polymers and the Environment, 2002. **10**(1-2): p. 39-48.

129. Gellerstedt, G. and G. Henriksson, *Lignins: major sources, structure and properties*, in *Monomers, Polymers and Composites from Renewable Resources*. 2008, Elsevier. p. 201-224.

130. Lora, J., *Industrial commercial lignins: sources, properties and applications*, in *Monomers, polymers and composites from renewable resources*. 2008, Elsevier. p. 225-241.

131. Kawamoto, H., T. Watanabe, and S. Saka, *Strong interactions during lignin pyrolysis in wood – A study by in situ probing of the radical chain reactions using model dimers.* Journal of Analytical and Applied Pyrolysis, 2015. **113**: p. 630-637.

132. Kotake, T., H. Kawamoto, and S. Saka, *Pyrolytic formation of monomers from hardwood lignin as studied from the reactivities of the primary products.* Journal of Analytical and Applied Pyrolysis, 2015(113): p. 57-64.

133. Lindner, B., et al., *Solvent-driven preferential association of lignin with regions of crystalline cellulose in molecular dynamics simulation.* Biomacromolecules, 2013. **14**(10): p. 3390-3398.

134. Goring, D.A.I. *The lignin paradigm*. 1989. Oxford University Press.

135. Rößiger, B., G. Unkelbach, and D. Pufky-Heinrich, *Base-catalyzed depolymerization of lignin: History, challenges and perspectives.* Lignin-Trends and Applications, 2018: p. 99-120.

136. Dutta, A., et al., *Techno-economics for conversion of lignocellulosic biomass to ethanol by indirect gasification and mixed alcohol synthesis.* Environmental Progress & Sustainable Energy, 2012. **31**(2): p. 182-190.

137. Sun, Z., et al., *Bright side of lignin depolymerization: toward new platform chemicals.* Chemical reviews, 2018. **118**(2): p. 614-678.

138. George, A., et al., *Design of low-cost ionic liquids for lignocellulosic biomass pretreatment.* Green Chemistry, 2015. **17**(3): p. 1728-1734.

139. Abbott, A.P., et al., *Deep eutectic solvents formed between choline chloride and carboxylic acids: versatile alternatives to ionic liquids.* Journal of the American Chemical Society, 2004. **126**(29): p. 9142-9147.

140. Francisco, M., A. van den Bruinhorst, and M.C. Kroon, *New natural and renewable low transition temperature mixtures (LTTMs): screening as solvents for lignocellulosic biomass processing.* Green Chemistry, 2012. **14**(8): p. 2153-2157.

141. Liu, Y., et al., *Efficient Cleavage of Lignin–Carbohydrate Complexes and Ultrafast Extraction of Lignin Oligomers from Wood Biomass by Microwave-Assisted Treatment with Deep Eutectic Solvent.* ChemSusChem, 2017. **10**(8): p. 1692-1700.

142. Jorgensen, W.L., D.S. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids.* Journal of the American Chemical Society, 1996. **118**(45): p. 11225-11236.

143. Doherty, B. and O. Acevedo, *OPLS force field for choline chloride-based deep eutectic solvents.* The Journal of Physical Chemistry B, 2018. **122**(43): p. 9982-9993.

144. Zoete, V., et al., *SwissParam: a fast force field generation tool for small organic molecules.* Journal of computational chemistry, 2011. **32**(11): p. 2359-2368.

145. Vermaas, J.V., et al., *Automated Transformation of Lignin Topologies into Atomic Structures with LigninBuilder.* ACS Sustainable Chemistry & Engineering, 2018.

146. Martínez, L., et al., *PACKMOL: a package for building initial configurations for molecular dynamics simulations.* Journal of computational chemistry, 2009. **30**(13): p. 2157-2164.

147. Essmann, U., et al., *A smooth particle mesh Ewald method.* The Journal of chemical physics, 1995. **103**(19): p. 8577-8593.

148. HUANG, J.-b., et al., *Theoretical study of bond dissociation energies for lignin model compounds.* Journal of Fuel Chemistry and Technology, 2015. **43**(4): p. 429-436.

149. Hess, B., *Determining the shear viscosity of model liquids from molecular dynamics simulations.* The Journal of chemical physics, 2002. **116**(1): p. 209-217.

150. Politis, D.N. and J.P. Romano, *A circular block-resampling procedure for stationary data.* Exploring the limits of bootstrap, 1992. **2635270**.

151. Muley, P.D., et al., *Rapid microwave-assisted biomass delignification and lignin depolymerization in deep eutectic solvents.* Energy Conversion and Management, 2019. **196**: p. 1080-1088.

152. Guadix-Montero, S. and M. Sankar, *Review on catalytic cleavage of C–C inter-unit linkages in lignin model compounds: towards lignin depolymerisation.* Topics in Catalysis, 2018. **61**(3-4): p. 183-198.

153. Yan, N., et al., *Selective degradation of wood lignin over noble-metal catalysts in a two-step process.* ChemSusChem: Chemistry & Sustainability Energy & Materials, 2008. **1**(7): p. 626-629.

154. Azadi, P., et al., *Liquid fuels, hydrogen and chemicals from lignin: A critical review.* Renewable and Sustainable Energy Reviews, 2013. **21**: p. 506-523.

155. Liu, Q., et al., *Lignin depolymerization to aromatic monomers and oligomers in isopropanol assisted by microwave heating.* Polymer degradation and stability, 2017. **135**: p. 54-60.

156. Morin-Crini, N., et al., *Water-insoluble β-cyclodextrin–epichlorohydrin polymers for removal of pollutants from aqueous solutions by sorption processes using batch studies: A review of inclusion mechanisms.* Progress in Polymer Science, 2018. **78**: p. 1-23.

157. Crini, G. and M. Morcellet, *Synthesis and applications of adsorbents containing cyclodextrins.* Journal of Separation Science, 2002. **25**(13): p. 789-813.

158. Danquah, M.K., R.C. Aruei, and L.D. Wilson, *Phenolic pollutant uptake properties of molecular templated polymers containing β-cyclodextrin.* The Journal of Physical Chemistry B, 2018. **122**(17): p. 4748-4757.

159. López, C.A., A.H. de Vries, and S.J. Marrink, *Molecular mechanism of cyclodextrin mediated cholesterol extraction.* PLoS computational biology, 2011. **7**(3): p. e1002020.

160. Alvira, E., *Molecular Simulation of the Separation of Isoleucine Enantiomers by β-Cyclodextrin.* Molecules (Basel, Switzerland), 2019. **24**(6): p. 1021.

161. Nutho, B., et al., *Binding mode and free energy prediction of fisetin/β-cyclodextrin inclusion complexes.* Beilstein journal of organic chemistry, 2014. **10**(1): p. 2789-2799.

162. Schönbeck, C., P. Westh, and R. Holm, *Complexation thermodynamics of modified cyclodextrins: extended cavities and distorted structures.* The Journal of Physical Chemistry B, 2014. **118**(34): p. 10120-10129.

163. Martin, J., E.J. Díaz-Montaña, and A.G. Asuero, *Cyclodextrins: Past and Present.* Cyclodextrin: A Versatile Ingredient, 2018: p. 1.

164. Strushkevich, N., S.A. Usanov, and H.-W. Park, *Structural basis of human CYP51 inhibition by antifungal azoles.* Journal of molecular biology, 2010. **397**(4): p. 1067-1078.

165. Berman, H.M., et al., *The protein data bank.* Acta Crystallographica Section D: Biological Crystallography, 2002. **58**(6): p. 899-907.

166. Kohlmeyer, A., *The TopoTools VMD plugin*. 2017.

167. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* Journal of molecular graphics, 1996. **14**: p. 33–38.

168. Jo, S., et al., *CHARMM-GUI: a web-based graphical user interface for CHARMM.* Journal of computational chemistry, 2008. **29**(11): p. 1859-1865.

169. Jo, S., et al., *Glycan Reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins.* Journal of computational chemistry, 2011. **32**(14): p. 3135-3141.

170. Allen, F.H., *The Cambridge Structural Database: a quarter of a million crystal structures and rising.* Acta Crystallographica Section B: Structural Science, 2002. **58**(3): p. 380-388.

171. Chen, W., C.-E. Chang, and M.K. Gilson, *Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design.* Biophysical Journal, 2004. **87**(5): p. 3035-3049.

172. Schönbeck, C. and R. Holm, *Exploring the Origins of Enthalpy–Entropy Compensation by Calorimetric Studies of Cyclodextrin Complexes.* The Journal of Physical Chemistry B, 2019. **123**(31): p. 6686-6693.

173. DeLano, W.L., *Pymol: An open-source molecular graphics tool.* CCP4 Newsletter On Protein Crystallography, 2002. **40**(1): p. 82-92.

174. Petridis, L. and J.C. Smith, *A molecular mechanics force field for lignin.* Journal of computational chemistry, 2009. **30**(3): p. 457-467.

175. Golovina, E.A., et al., *Water replacement hypothesis in atomic detail—factors determining the structure of dehydrated bilayer stacks.* Biophysical journal, 2009. **97**(2): p. 490-499.

VITA

Xinjie Tong obtained B.S. and M.S. degree from the Department of Chemical Engineering at Wuhan Institute of Technology, China, in 2010 and 2012 respectively. She obtained another M.S. degree from the Department of Chemical Engineering at Illinois Institute of Technology in 2014. After that, she joined Dr. Dorel Moldovan's group in January 2015 for a Ph.D. program at the Department of Mechanical Engineering. Her research interest focuses on Molecular dynamics simulation of interaction of DNA nucleotides and lignin oligomers with small molecules and interfaces.