

Bernadette Mohr

Multiscale Simulations of cardiolipin-binding Small Molecules

Thesis

to achieve the university degree of

Diploma

Course programme: Biology

submitted to

Johannes Gutenberg University Mainz

Supervisor

Dr. Tristan Bereau



Mainz, September 2018

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used.

Date

Signature

Acknowledgements

I would first like to thank my advisor Dr. Tristan Bereau from the Max Planck Institute for Polymer Research for giving me the opportunity to work at this great research project and for always being helpful and encouraging.

I would also like to thank Dr. Roberto Menichetti for introducing me to the fine points of free-energy calculations.

I want to thank Marius Bause for all the impromptu physics lessons. He was always on hand if I encountered topics that exceeded the curriculum of biology.

I have to thank Christian Hoffmann for the coarse-grained parameterization of 10-N nonyl-acridine orange.

I would also like to thank Kiran Kanekal for the suggestion of the writing club, a sorrow shared is a sorrow halved.

Last, but not least, I want to thank Max Kaiser for providing me with informative, high quality structural formulae of acridine orange and the lipids I used in my project.

Contents

| | |
|--|------------|
| Acknowledgements | v |
| Contents | vii |
| 1. Introduction | 1 |
| 1.1. Context | 1 |
| 1.2. Biological Model System | 2 |
| 1.2.1. Mitochondria | 3 |
| 1.2.2. Lipids | 4 |
| 1.2.3. Drug-like Molecules | 8 |
| 2. Physical Principles | 9 |
| 2.1. Statistical Mechanics | 9 |
| 2.1.1. Hamiltonian Dynamics | 9 |
| 2.1.2. Statistical Mechanics Ensembles | 11 |
| 2.2. Free Energy | 16 |
| 2.2.1. Gibbs Free Energy | 16 |
| 2.3. Potential of mean force | 18 |
| 3. Materials and Methods | 21 |
| 3.1. Molecular Models | 21 |
| 3.1.1. Classical Force Fields | 21 |
| 3.1.2. Atomistic Resolution | 22 |
| 3.1.3. Coarse Graining | 23 |
| 3.2. Molecular Dynamics Simulations | 24 |
| 3.2.1. MD Workflow | 25 |
| 3.3. Free Energy Calculations | 28 |
| 3.3.1. Alchemical Transformations | 29 |
| 3.3.2. Umbrella Sampling | 32 |
| 3.3.3. Histogram Reweighting | 34 |
| 3.4. Simulation Setup | 38 |
| 3.4.1. Coarse-grained Resolution | 39 |
| 3.4.2. Atomistic Resolution | 41 |

Contents

| | |
|---|-----------|
| 4. Results | 45 |
| 4.1. Coarse-Grained Resolution | 45 |
| 4.1.1. Alchemical Transformations | 45 |
| 4.1.2. Potentials of Mean Force | 50 |
| 4.2. Atomistic Resolution | 53 |
| 4.2.1. Alchemical Transformations | 54 |
| 5. Discussion | 59 |
| 5.1. Conclusion | 59 |
| A. Additional Data | 63 |
| A.1. Free energies of the coarse-grained Alchemical Transformations | 63 |
| A.2. Relevant points of the PMF curves | 66 |
| A.3. Free energies from the atomistic Alchemical Transformations | 68 |
| Glossary | 71 |
| Bibliography | 73 |

List of Figures

| | | |
|-------|--|----|
| 1.1. | EM image and illustration of a mitochondrion | 3 |
| 1.2. | Schematic drawing of a bilayer and a micelle | 5 |
| 1.3. | Structural formula of Cardiolipin | 6 |
| 1.4. | Structural formula of Phosphatidylglycerol | 7 |
| 1.5. | Structural formula of Phosphatidylcholine | 7 |
| 1.6. | Structural formula of 10-N Nonyl-acridine orange | 8 |
| 2.1. | A PMF curve superimposed on a lipid membrane. | 19 |
| 3.1. | A general MD simulation workflow. | 26 |
| 3.2. | The singularity at $r \rightarrow 0$ of the Lennard-Jones potential. | 30 |
| 3.3. | Thermodynamic cycle for the binding affinity of different molecules to a lipid membrane. | 31 |
| 3.4. | Umbrella windows along a reaction coordinate. | 34 |
| 3.5. | NAO molecule with original bead types and transformations | 40 |
| 4.1. | Calculation of relative free binding energies | 46 |
| 4.2. | $\Delta\Delta G$ of the top center aromatic bead transformations | 47 |
| 4.3. | $\Delta\Delta G$ of the side bead transformations | 49 |
| 4.4. | PMFs of the different molecules in PC compared to CL | 51 |
| 4.5. | PMFs of the different molecules in PG compared to CL | 52 |
| 4.6. | Transformed molecules on atomistic level | 54 |
| 4.7. | Position density plot: PG and original NAO compared to NAO with added hydroxy group | 55 |
| 4.8. | Position density plot: CL and original NAO compared to NAO with added hydroxy group | 55 |
| 4.9. | Alignment of NAO in the membrane | 56 |
| 4.10. | Alignment of NAO with added hydroxy group | 57 |
| 4.11. | Alignment of NAO with two added amine groups | 57 |
| A.1. | Position density plot: two added hydroxy groups | 68 |
| A.2. | Position density plot: one added amine group | 69 |
| A.3. | Position density plot: two added amine groups | 69 |

List of Tables

| | | |
|------|--|----|
| 3.1. | Atomistic intermediate states | 43 |
| 4.1. | Free energy differences with transformed beads on top of the NAO ring system | 48 |
| 4.2. | Free energy differences with transformed beads on the sides of the NAO ring system | 49 |
| 4.3. | Free energy barriers between the different regions of the membranes | 53 |
| A.1. | Free energy differences from coarse-grained alchemical transfromations . . | 63 |
| A.2. | PMF regions of interest | 66 |
| A.3. | Free energy change of atomistic alchemical transforamtions | 68 |

1. Introduction

Rational design of materials with specific functions requires an understanding of the underlying fundamental physical mechanisms. Structure-property relationships are of interest for a number of fields, from materials science to drug discovery. For the targeted modification of a molecule's properties, the respective influence of its components have to be defined. The basic assumption for molecule based hypotheses is that molecules with similar structures have similar properties. Therefore, significant differences have to be defined on a molecular level, since each kind of interaction can depend on different factors. To efficiently filter out structures that share desired properties, high throughput screening methods are used. They allow the assaying of a large number of potential active substances against a set of defined targets. In an experimental context, high throughput methods include a high level of expertise in robotic automation, analysis technology and high content imaging, which means the overhead is big compared to a success rate of about 2% of the total number of molecules screened. This makes the development and improvement of virtual high throughput screening methods extremely interesting.

To improve the performance of virtual screening, methods are needed that increase the efficiency of screening the chemical compound space without causing loss of relevant information. The results then have to be transferred to a considerably reduced number of candidate molecules that can be investigated in more detail. An approach to a solution of this problem in a drug development context is investigated in this work.

1.1. Context

Virtual screening is widely used in drug discovery and it proves valuable in eliminating inactive compounds. However, it is often insufficiently effective in discovering molecules that interact with a specific biochemical target in a desired way [1]. Current approaches reduce the computational cost of screening the chemical compound space by using simplified representations and approximations, speed is given priority over accuracy. As a result, statistical mechanical effects such as conformational entropy, binding modes or solvent behavior are deliberately excluded.

Potential drug candidates have to be highly selective, ideally cause no side effects and need to be easily delivered to the targeted biochemical mechanisms. Those requirements rely mainly on the processes not considered in a majority of the currently available virtual screening

1. Introduction

methods. Therefore, applying physics-based models to virtual screening is a promising approach to yield fast but still realistic results. Coarse-grained molecular simulation models are one such method, that can provide general insight into the complex behavior of a biochemical system.

This thesis investigates the capacity of adding a preliminary, coarse-grained step to increase the efficiency of virtual screening. The coarse-graining technique is based on reducing the complexity of simulated molecules and interactions and allows the effective screening of a large number of molecules, but suffers from the loss of information caused by this reduction of the degrees of freedom of the system. As yet, the capabilities of physics based molecular models for virtual screening have not been comprehensively evaluated.

Fully atomistic simulation models contain more realistic information about molecular structures and interactions, but this vastly increases the number of degrees of freedom that have to be considered. Therefore, they are too computationally expensive to be used from the very beginning of the virtual screening process.

The experimentally shown binding affinity of a fluorescent cationic dye to a specific lipid is used as a starting point of the optimization process. The physical and chemical properties of this molecule are to be optimized to increase its selectivity for the target lipid. A large number of possible modifications are tested on the coarse-grained level, which reduces the computational cost by reducing the number of unknowns to take into account. The promising results of this coarse-grained simulations are then transferred back to atomistic resolution and verified. The model system chosen for the simulations is described in the following sections.

1.2. Biological Model System

One possible mode of action for drugs is influencing the structure of cellular membranes. Depending on their target, they have to be able to permeate the membranes of specific cells or intracellular compartments.

Biological membranes [2] are leaf-like, 6 to 10 nm wide structures consisting of a large number of different, spatially orientated lipids and proteins. The separate membrane molecules are connected by non-covalent interactions, therefore membranes are liquid structures. They can be deformed by external stress or thermal fluctuations, making them a soft-matter system. Membranes consist of lipid bilayers and create enclosed compartments forming either whole cells or the organelles within the cell. A simplified illustration of membrane structures can be seen in figure 1.2. The lipid composition of the two layers forming a membrane is always different. Lipids can diffuse freely along the membrane plane, but rarely along the membrane normal.

Membrane permeability is highly selective, the lipid bilayers being nearly impermeable for ions and most polar molecules. Transmembrane proteins control the ion and molecular composition inside the membrane compartments and facilitate signal transduction between

cells.

Eucaryotic cells contain several membrane enclosed organelles. Of those, the nucleus and mitochondria are surrounded by two different membranes.

1.2.1. Mitochondria

Mitochondria [3] play a prominent part in a number of physiological processes and are therefore an important target for drug development. They occupy a substantial amount of the cytoplasmic volume of eucaryotic cells, and they have been essential for the evolution of complex organisms.

Mitochondria are semiautonomous organelles existing in a symbiotic relationship with the host cell. They hold their own DNA coding for proteins and lipids specific for the mitochondrial membranes, but other molecules have to be imported from the host cell. Cells containing mitochondria are dependent on the Adenosine triphosphate (ATP) molecules provided by oxidative phosphorylation of glucose in the mitochondria, a co-enzyme involved in intracellular energy transfer. This is explained by symbiogenesis, an independent monacellular organism being incorporated into the cells of another organism.

Mitochondria generate the vast majority of the ATP required in an aerobic cell. They are mobile, plastic organelles and can accumulate in regions with increased energy demand. Each mitochondrion is bounded by two specialized membranes, an outer and a vast, strongly folded inner membrane, as illustrated in figure 1.1.

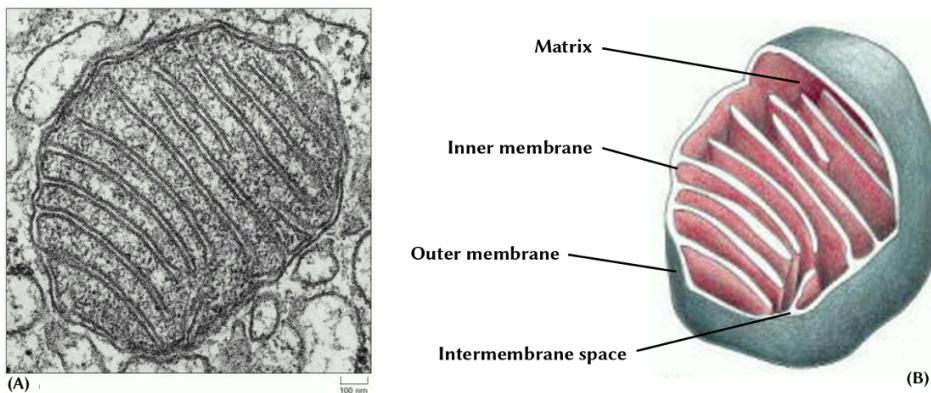


Figure 1.1.: (A) A cross section of a mitochondrion under the electron microscope. (B) Schematic drawing of a mitochondrion showing the three-dimensional structure. [4]

They have highly specialized functions and create two separate, aqueous compartments, the narrow intermembrane space and the internal mitochondrial matrix. The matrix contains a mixture of enzymes for oxidation of pyruvate and fatty acids and for the citric acid cycle,

1. Introduction

and the mitochondrial DNA. The total surface area of the inner membrane (red) is increased by its tight folding into structures called cristae. It is highly specialized and its lipid bilayer contains a high concentration of the phospholipid cardiolipin, which is found nowhere else in an eucaryotic organism. The outer membrane contains porin, a large channel-forming protein, and is permeable for all molecules with an atomic mass lower than 5000 Dalton. The intermembrane space (white) contains enzymes that use ATP to phosphorylate other nucleotides.

Most of the citric acid cycle and the fatty acid oxidation reactions are located in the outer membrane, while the oxidative phosphorylation takes place exclusively in the inner mitochondrial membrane and inter-membrane space. Several diseases are caused by mutations in the mitochondrial genome inhibiting steps in the specific metabolic pathways located in the inner mitochondrial membrane.

Accumulating mutations over years can contribute to the aging process, degenerative diseases and cancer [5], and mitochondria have been found to control apoptosis [6].

1.2.2. Lipids

Lipids [7] constitute about 50% of the mass of eukaryotic cell membranes. A $1\mu m \times 1\mu m$ are of lipid bilayer contains approximately 5×10^6 lipid molecules. The lipids in biological membranes are amphiphilic, they have a hydrophilic, polar head group and hydrophobic, nonpolar tails.

Phospholipids form the biggest group of the membrane lipids, they have a polar head group and two hydrophilic fatty acid tails normally containing 14 to 24 carbon atoms. One tail usually has one or more cis-double bonds creating small kinks in the tail, while the other tail is saturated. The differences in length and saturation of the tails influence the packing ability of the phospholipid molecules and therefore the membrane fluidity.

In aqueous environments, lipids spontaneously self assemble into larger structures, because the free energy cost of the hydrophobic effect is reduced if the contact of their hydrophobic tails to water molecules is minimal. Depending on the shape of the lipids, they form spherical micelles or bilayers as shown in figure 1.2.

The hydrophilic head groups arrange themselves in a polar phase (red, P) in contact with water molecules, whereas the hydrophobic tails get placed inwards to form a nonpolar phase (blue, N) to minimize the increase in free energy caused by the formation of a hydration shell.

The same driving forces also cause the self-healing property of membranes. Tears or defects create a free edge that is energetically unfavorable, and are therefore closed by spontaneous self assembly of the lipids molecules. The energetic prohibition of free edges leads to lipid bilayers always closing in on themselves to form sealed compartments.

The lipids used in this work are described in the following sections.

1.2. Biological Model System

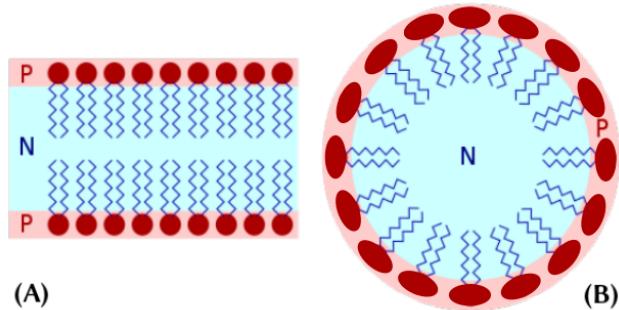


Figure 1.2.: Packing arrangements of lipid molecules in an aqueous environment [8]. (A) Cylinder-shaped phospholipids form bilayers. (B) Wedge-shaped lipids form micelles.

Cardiolipin

Cardiolipin [9] (CL) constitutes up to 20% of the inner mitochondrial membrane and therefore is an important target for possible drug substances.

CL is a phospholipid which has four fatty acids chains, as shown in figure 1.3, and therefore makes the membrane especially impermeable to ions. It can be found in most bacteria and all eucaryotic cells.

In mammalian cells, CL is exclusively located in the inner mitochondrial membrane where it is predominantly clustered at highly curved parts of the cristae. CL is synthesized from cytidinediphosphate-diacylglycerol and phosphatidylglycerol. This reaction is catalyzed by cardiolipin synthase, an evolutionary highly conserved enzyme. The composition of the four fatty acid chains is specific for tissues and cell types of an organism. This is achieved by additional remodeling steps made possible because of the unspecific substrate acceptance of the cardiolipin synthase.

Recent studies have shown that CL is doubly deprotonated at physiological pH [10, 11, 12, 13]. Its charge state influences the headgroup charge repulsion and controls the molecular geometry. The phase behavior and curvature of CL depend on both pH and ionic concentration [14]. Therefore, CL plays an important role in supporting the structural integrity of the inner mitochondrial membrane [15, 16].

Additionally, CL serves as a cofactor for several mitochondrial enzymes, is associated with different enzymes of the respiratory chain and helps stabilize the pH in the mitochondrial intermembrane space by serving as a proton trap for oxidative phosphorylation [17].

Another important function of CL is the mediation of apoptotic processes in association with cytochrome c [18]. Also, CL is suspected to have a role in mitochondrial membrane fusion and fission, which is important for normal cellular function and would also add to the influence of CL in apoptosis [19].

1. Introduction

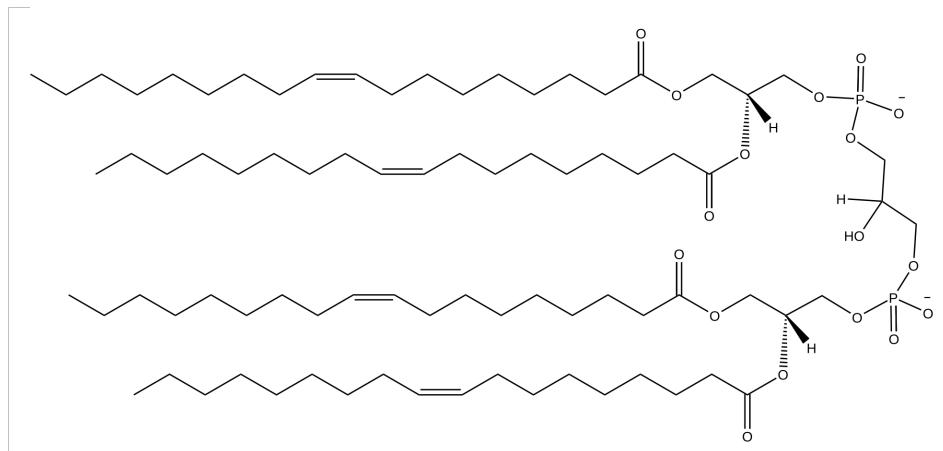


Figure 1.3.: Structural formula of Cardiolipin

Deficiencies in CL have been linked to diseases like Barth syndrome [20], a disorder caused by a mutation of the gene coding for tafazzin, an enzyme involved in the biosynthesis of CL. The mutation of tafazzin causes insufficient CL remodeling which results in reduced ATP production.

The CL content in the brain decreases with aging because of lipid peroxidation due to free radical stress in mitochondria. Oxidative stress and lipid peroxidation are believed to be contribution factors to Parkinson's disease and the development of Alzheimer's disease [21]. CL deficiencies in the heart tissue has been found to be the cause of premature heart diseases in patients suffering from diabetes [22, 23, 24]. The mitochondrial dysfunction in non-alcoholic fatty liver disease and heart failure corresponds with decreased CL levels and changes in the fatty acid chain compositions [25].

As well as CL deficiencies, increased production rates also result in disordered metabolic functions [26].

Phosphatidylglycerol

Phosphatidylglycerol [27] (PG) is a phospholipid present in all organisms. In the membranes of plants and animals, its concentration is generally low at a level of 1-2%, but it often has essential functions. PG is an anionic lipid at neutral pH. Because its glycerol headgroup gets enlarged by hydration, it has an overall cylindrical shape.

PG is a key intermediate in the biosynthesis of CL [28], which makes it an important basis for the mitochondrial morphology and function [29].

Further on, it takes part in the regulation of the innate immune response in the lungs, both by attenuating inflammation induced by bacterial lipopolysaccharides and by blocking viral infections.

Palmitoyl-oleoyl-PG is the main molecular species in human lungs, where its acidic head

1.2. Biological Model System

group plays an important role for its surfactant function. It prevents alveolar epithelial apoptosis and blocks the effects of inflammatory agents that cause acute lung injury, therefore it can be applied as a treatment for respiratory distress syndrome.

Additionally, in mammalian cells PG activates a nuclear protein kinase C, is present in the active site of the enzyme cytochrome c oxidase, the last enzyme of the respiratory chain of mitochondria, and is also known to bind non-covalently to several membrane transporter and channel proteins.

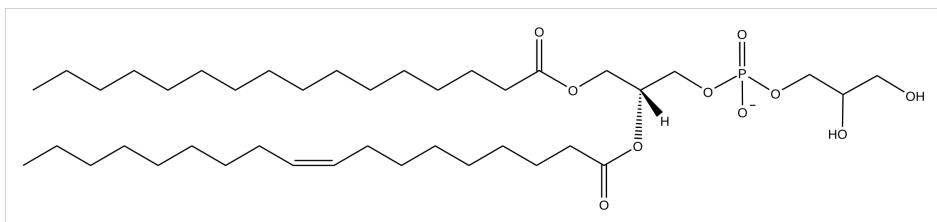


Figure 1.4.: Structural formula of Phosphatidylglycerol

Phosphatidylcholine

Phosphatidylcholine [30] (PC) is a phospholipid that is neutral or zwitterionic over a pH range from strongly acidic to strongly alkaline and has been found to show virtually no interaction with the fluorescent dye Nonyl-acridine orange [31], therefore it has been used as a scientific control of the molecular model in this work. It occurs with the highest concentration in eukaryotes, constituting up to 50% of the total complex membrane lipids. PC is mainly found in the outer leaflet of the plasma membrane and is also an integral component of high-density-lipoprotein (HDL).

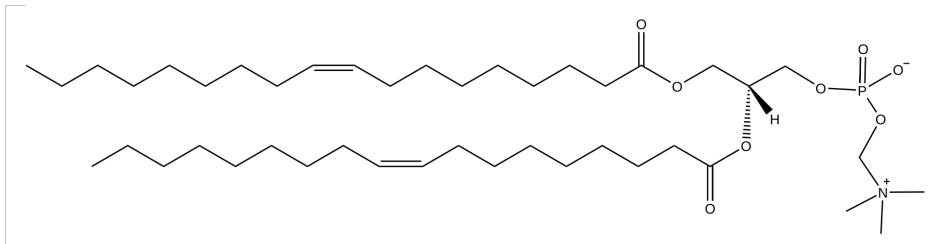


Figure 1.5.: Structural formula of Phosphatidylcholine

1. Introduction

1.2.3. Drug-like Molecules

In the search for drug candidates, large databases of chemical compounds are screened for drug-like molecules.

A molecule is drug-like [32], if it has certain properties that increase its probability to become a successful active substance. The structural properties of a molecule, like the presence of hydrogen bonding moieties, lipophilicity, molecular weight, polar surface area, shape, reactivity and dissociation constants characterize its physicochemical properties by interaction with the physical environment. The physicochemical properties are for example solubility, permeability and chemical stability. Interaction of the structural properties with the drug target characterize the biochemical properties of a molecule. This includes drug metabolism, protein and tissue binding, and transport modality. At the highest level, the interactions between physicochemical and biochemical properties and a living system characterize the pharmacokinetics and toxicity of a molecule. Therefore, it is possible to control the drug-like properties of a molecule through the modification of its structure. The drug-like molecule used in this thesis is described in the following paragraph.

10-N Nonyl-Acridine Orange

The fluorescent dye acridine orange (NAO) shown in figure 1.6 has a high affinity for cardiolipin [33] and is therefore used as a mitochondrial probe. The targeting potential is believed to be induced by three factors. The regions of the NAO molecule experimentally shown to be interacting with CL are the electrostatic interaction between the NAO quaternary ammonium and the ionized phosphate residues of CL, hydrophobic interactions between the chromophores of adjacent NAO molecules, and the insertion of the hydrophobic nonyl chain into the hydrophobic environment created by the four fatty acid chains of CL [33].

Because of those properties, NAO, together with CL, PG and PC, was chosen as the starting point of a structure based drug design approach implementing the coarse-grained screening step.

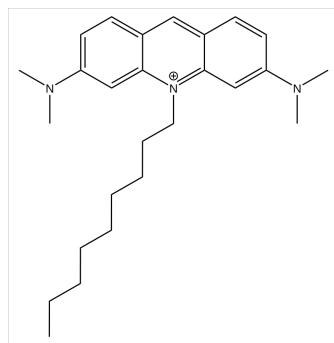


Figure 1.6.: Structural formula of 10-N Nonyl-acridine orange

2. Physical Principles

2.1. Statistical Mechanics

Statistical mechanics [34] allows the interpretation of the equilibrium properties of macroscopic systems. It provides a correct description of processes such as the magnitudes and time scales of spontaneous energy fluctuations by reducing the number of degrees of freedom. Macroscopic systems consist of a large number of particles and therefore cannot be completely characterized and a deterministic prediction of their time-evolution is computationally infeasible.

Statistical mechanics describes the microstates of a system and allows an estimation of macroscopic observables like energy, volume or pressure.

2.1.1. Hamiltonian Dynamics

The Hamiltonian formulation [35] of classical mechanics is widely used in molecular dynamics simulations, because it can be generalized to quantum mechanics and the Hamiltonian equations can be efficiently solved by suitable integrators [36].

In a classical physical system containing N particles, the particle positions are described by coordinates q on the $3N$ -dimensional configuration space with $q = 1, \dots, 3N$, and their velocities \dot{q} . Those coordinates have to satisfy the constraints acting on the system.

Constraints defining the relationships among the coordinates can be used to reduce the set of coordinates, because only the independent degrees of freedom of the system are relevant. This leaves only the generalized coordinates independent of the constraints to be evaluated. The Hamiltonian $\mathcal{H} = (q, p, t)$ is a function of the generalized coordinates q_i , the Lagrangian definition of the conjugate momenta $p_i = \frac{\delta L}{\delta \dot{q}_i}$, and time. It is given by the Legendre transform of the Lagrangian:

$$\mathcal{H}(q_i, p_i, t) = \sum_i p_i \dot{q}_i - L(q, \dot{q}, t)$$

2. Physical Principles

Hamilton's equations of motion

The equations of motion of Hamiltonian mechanics can be derived from the Lagrangian equation of motion,

$$\frac{\delta L}{\delta q_i} - \frac{d}{dt} \frac{\delta L}{\delta \dot{q}_i} = 0,$$

via Legendre transforms, using the conjugate momenta p_i and the definition for the derivative of the conjugate momenta $\dot{p}_i = \frac{\delta L}{\delta \dot{q}_i} = \frac{dp_i}{dt}$. This leads to Hamilton's equations of motion:

$$\begin{aligned}\dot{p}_i &= -\frac{\delta \mathcal{H}}{\delta q_i} \\ \dot{q}_i &= \frac{\delta \mathcal{H}}{\delta p_i}\end{aligned}$$

Physical Interpretation

The Hamilton \mathcal{H} is the total energy of the system. If, for example, the generalized coordinates q_i are defined as the Cartesian coordinates of the particles, from the Lagrangian

$$L(q_i, \dot{q}_i) = \sum_i \frac{1}{2} m \dot{q}_i^2 - U(q_i),$$

the Lagrangian relation between kinetic energy K and potential energy U ,

$$L(q_i, \dot{q}_i) = K - U,$$

and the definition of the conjugate momenta

$$p_i = \frac{\delta L}{\delta \dot{q}_i} = m \dot{q}_i$$

follows:

$$\begin{aligned}\mathcal{H} &= \sum_i p_i \dot{q}_i - L = \sum_i m \dot{q}_i^2 - \frac{1}{2} m \dot{q}_i^2 + U(q_i) \\ &= \sum_i \frac{1}{2} m \dot{q}_i^2 + U(q_i) \\ &= \sum_i \frac{p_i^2}{2m} + U(q_i)\end{aligned}\tag{2.1}$$

$$\mathcal{H} = K + U$$

2.1. Statistical Mechanics

For other sets of conjugate coordinates and with the corresponding definition of the kinetic energy K , the derivation in equation (2.1) also holds true.

Force is defined as the rate of change of momentum with respect to time, so the time-derivative of the momentum p is the Newtonian force: $F = \frac{dp}{dt} = m\frac{dv}{dt} = ma$. For a system described by Cartesian coordinates, Hamilton's mechanics is an alternative description of the system's dynamic, equivalent to Newton's mechanics.

2.1.2. Statistical Mechanics Ensembles

In statistical mechanics, an ensemble [34] is the assembly of all possible microstates consistent with the constraints characterizing the system macroscopically. This makes a statistical ensemble a probability distribution for the states of the system.

Most systems satisfy the ergodic hypothesis: Over a sufficiently long timescale, they will move through all their possible microstates. The time averages over experimental measurements and the statistical mechanics ensemble averages are equivalent:

$$F_{\text{observed}} = \sum_v \underset{\substack{\text{time} \\ \text{average}}}{P_v} \times \underset{\substack{\text{probability of finding} \\ \text{the system in microstate } v}}{F_v} \equiv \langle F \rangle \underset{\substack{\text{value of property } F \\ \text{in microstate } v}}{\underset{\substack{\text{ensemble} \\ \text{average}}}{}}$$

For molecular dynamics simulations, this means that statistically averaged properties can be determined from simulated trajectories, and for sufficiently long trajectories, the time-averaged properties become independent of the initial conditions.

To reproduce the ensemble constraints of a simulated system, thermostats (section 2.1.2) and barostats (section 2.1.2) are used.

Phase Space

The phase space Γ [37] of a system represents all possible states it can adopt. Each possible state corresponds to one unique point in phase space.

In classical mechanics, the phase space represents all possible values of positions (q_i) and conjugate momenta (p_i). The state of a system at any time t consisting of N particles is completely specified by a $6N$ -dimensional phase vector,

$$\vec{v} = (q_1, q_2, \dots, q_N; p_1, p_2, \dots, p_N)$$

as each particle is associated with three position variables and three momentum variables. The trajectories of all N particles $\vec{v}_i(t)$ are specified by Hamilton's equations of motion,

$$\vec{v}_i(t) = \left(\frac{\delta \mathcal{H}}{\delta p_i}; -\frac{\delta \mathcal{H}}{\delta q_i} \right)$$

2. Physical Principles

and are equivalent to the motion of a point in the $6N$ -dimensional phase space Γ . If the particles are distinguishable and can be assigned definite values for the positions and momenta, a point in phase space represents a microstate of the system. The phase space Γ is not equivalent to the configurational space of all the positions q_i .

Entropy

Entropy is the measure of a system's thermal energy that is not accessible for doing mechanical work.

The entropy S of a system is given by Boltzmann's equation,

$$S = k_B \ln \Omega_{NVE} \quad (2.2)$$

with Boltzmann's constant k_B and the density of states Ω_{NVE} for a given number of particles N , volume V and energy E .

The Boltzmann distribution is a probability distribution of all particles in the system over all possible states. It applies only to individual configurations of the system. In the form

$$\rho(i) = e^{-\frac{U}{k_B T}}$$

it describes the relative probability of a particular state of a system.

The Boltzmann factor

$$e^{-\frac{U}{k_B T}}$$

gives the relative probability of a particular microstate of a system for a given energy.

Because work is obtained from ordered molecular motion, the amount of entropy is also a measure of the molecular disorder, or randomness, of a system.

Isothermal-Isobaric Ensemble

The isothermal-isobaric (NpT) ensemble [38] describes systems in contact with an external thermostat and barostat.

The number of particles N , the pressure p and the temperature T of each system are conserved, the respective volume V and internal energy U can fluctuate.

Therefore, the macrostate probability distribution function has to include the overall volume V_{tot} as a variable to represent the number of possibilities the thermostat and barostat can readjust to conserve V_{tot} and the overall energy $E_{tot} = H(q_i, p_i)$,

$$\rho(q_i, p_i; V) = \frac{1}{Q_{NpT}} e^{-\beta[\mathcal{H}(q_i, p_i; V) + pV]}, \quad (2.3)$$

with the inverse thermodynamic temperature β , $\beta = \frac{1}{k_B T}$. The isothermal-isobaric partition function $Q_{(N,p,T)}$ is related to the canonical partition function [37, 34] and the density of

2.1. Statistical Mechanics

states Ω_{NVE} .

For systems with N unique particles, it can be given as a discrete sum

$$\begin{aligned} Q_{NpT} &= \sum_V e^{-\beta PV} Q_{NVT} = \sum_V \sum_{\mathcal{H}} e^{-\beta \mathcal{H}(q_i, p_i; V) - \beta PV} \Omega_{NVE} \\ &= \sum_V \sum_N e^{-\beta [\mathcal{H}_n(q_i, p_i; V) + PV]}. \end{aligned} \quad (2.4)$$

The classical formulation for a continuous system with N indistinguishable particles is:

$$\begin{aligned} Q_{NpT} &= \frac{1}{N! h^{3N}} \int_0^\infty dV \int e^{-\beta [\mathcal{H}(q_i, p_i; V) + PV]} dp_i dq_i \\ &= \frac{1}{N! h^{3N}} \int_0^\infty dV e^{-\beta PV} Q_{NVT}. \end{aligned}$$

Partition functions are sums over weights over all microstates states the system can adopt. They describe how the probabilities are partitioned by energy among the different microstates, so they yield the density of states, the number of microstates of a system at a certain energy level.

The corresponding free energy for the NpT ensemble is Gibbs free energy,

$$G(N, p, T) \equiv -k_B T \ln Q_{NpT} \quad (2.5)$$

The constraints of the NpT match experimental conditions of constant temperature and pressure, it is widely used in molecular dynamics simulations.

Thermostats

To generate a statistical ensemble at constant temperature, a thermostat has to be added. It can be used to match experimental conditions, manipulate the temperature in algorithms and avoid energy drifts caused by the accumulation of numerical errors.

Velocity Rescaling

The variable factor of the system connected to temperature is the velocity of it's particles. A multiplication factor β corresponding to a target temperature T_{target} and the current temperature of the system T_{old} is added to each velocity at every timestep Δt :

$$\begin{aligned} \beta &= \sqrt{\frac{T_{target}}{T_{old}}} \\ v_{rescale} &= \beta v_{old} \end{aligned}$$

The velocity rescaling method [39] is straightforward to implement, but in its simplest form, it is non-deterministic and time-irreversible. Therefore, it should only be used for preprocessing, not in an actual simulation.

2. Physical Principles

Langevin Thermostat

Langevin dynamics [40] can be used for molecular dynamics simulations of solute-solvent systems. If only the behavior of the solute is of interest, the influences of the solvent on the solute's dynamics via random collisions and a frictional drag force are modeled by Langevin's equation of motion.

At each time step Δt the Langevin thermostat updates the system so that the change in momenta is

$$\Delta p_i = \left(\frac{\delta U(q)}{\delta q_i} - \gamma p_i + \delta \xi \right) \Delta t,$$

where γ is a damping factor for the momenta p_i and $\delta \xi$ is a Gaussian distributed random number with the probability

$$\rho(\delta \xi) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|\delta \xi|^2}{2\sigma^2}}$$

and standard deviation $\sigma^2 = 2\gamma m_i k_B T$. The random fluctuating force $\delta \xi$ represents the thermal collisions from the solvent particles. The combination of the damping factor and the random force keep the system at constant temperature.

The Langevin thermostat reduces the computational cost by treating solvent atoms implicitly by stochastic terms and thus allowing larger time steps because only the larger degrees of freedom have to be considered explicitly.

Drawbacks for the Langevin thermostat are the loss of solvent effects and size constraints of the solvent molecules compared to the smallest solute molecule. The relation of the friction coefficient to the particle radius, $\gamma_i = \frac{6\pi\eta r_i}{m_i}$, which is only defined for spherical particles, is another problem of the Langevin thermostat complicating the implementation of non-spherical particles.

Barostats

To keep the system at constant pressure, the simulation cell volume is allowed to fluctuate in order to equilibrate the system to an external stress state.

Berendsen Pressure Coupling

The Berendsen pressure coupling approach [41] employs a scaling factor μ to the volume of the simulation box, with the isothermal compressibility β and a relaxation constant τ :

$$\mu = \left[1 - \beta \frac{\delta t}{\tau} (\mathcal{P} - P) \right]^{\frac{1}{3}}.$$

Where \mathcal{P} is the instantaneous pressure of the system and P is the target pressure. The scaling factor μ is added to the equations of motion, :

$$\begin{aligned} \dot{q}_i &= p_i + \mu q_i, \\ V &= 3\mu V. \end{aligned}$$

2.1. Statistical Mechanics

The pressure change is related to the isothermal compressibility β ,

$$\frac{dP}{dt} = -\frac{1}{\beta V} \frac{dV}{dt} = -\frac{3\mu}{\beta}$$

The modified equation of motion is:

$$\dot{q}_i = p_i - \frac{\beta(\mathcal{P} - P)}{3\tau} q_i$$

Parrinello-Rahman Coupling

A variable box size is useful for simulations of solid substances, because it allows phase changes occurring which involve changes in the unit cell dimension and angles.

For a simulation box with variable side lengths r and volume V , the box sides have to be independent, and nonorthogonal box side vectors have to be acceptable [42, 43]. This kind of box can be described by a 3×3 matrix h , the columns of which are the box side vectors \vec{a} , \vec{b} and \vec{c} ,

$$h = (\vec{a}, \vec{b}, \vec{c})$$

$$V = \det h = \vec{a} \cdot (\vec{b} \times \vec{c}).$$

The position r_i of a particle i can be written in terms of the matrix a and a column vector s_i ,

$$\vec{r}_i = h s_i = l_i \vec{a} + m_i \vec{b} + n_i \vec{c},$$

with the components l, m, n under the condition that $l \leq 0; m, n \leq 1$.

The pressure component is constituted by a piston with mass Q_p that gets included into the kinetic energy of the variable box:

$$K_V = \frac{1}{2} Q_p \sum_{\alpha} \sum_{\beta} \dot{h}_{\alpha\beta}^2.$$

The equation of motion becomes

$$Q_p \ddot{h} = (\mathcal{P} - P) V (h^{-1})^T,$$

where the difference between the instantaneous pressure \mathcal{P} acting on the box at the current time step and the target pressure P has to be minimized [44]. The instantaneous pressure is given by the pressure tensor of the fluctuating box:

$$\mathcal{P} = \frac{1}{V} \left(\sum_i (h s_i)_{\alpha}^T (h s_i)_{\beta} + \sum_i \sum_{j>i} (h s_{ij})_{\alpha} (f_{ij})_{\beta} \right).$$

2. Physical Principles

2.2. Free Energy

The free energy is the thermodynamic potential, a fundamental quantity containing information about all states of a system. It is at its minimum if the system is at equilibrium with the environment.

Free energy also is the energy that is released from, or absorbed by, a process that is slow enough for the system to stay at equilibrium.

2.2.1. Gibbs Free Energy

Gibbs free energy G represents the amount of energy available to do work in a physical system within the constraints of the NpT ensemble. The change of state in a system is characterized by the change of free energy, ΔG .

Biochemical reactions take place under constant temperature and pressure. Therefore, Gibbs free energy is preferred for molecular dynamics simulations of biomolecules, keeping the theoretical results comparable to experimental data. The sign of ΔG shows the possibility of a process happening spontaneously:

$\Delta G < 0$ The reaction is exotherm

$\Delta G > 0$ The reaction is endotherm

$\Delta G = 0$ The system is at equilibrium, its macroscopic properties will remain unchanged.

Gibbs free energy is widely used in molecular dynamics simulations. It can be derived from the enthalpy H describing the tendency of a system to move toward a lower energetic state.

Derivation

Gibbs free energy depends on the entropy S , and the enthalpy H . In the context of Gibbs free energy, the entropy S describes the fact that a system always wants to reach the energetically most favorable state, thermodynamic equilibrium. It is derived from the thermodynamic equation for the internal energy U :

$$\begin{aligned} dU &= TdS - pdV \\ TdS &= dU + pdV. \end{aligned}$$

The enthalpy H has to change if the internal energy U of the system changes:

$$\begin{aligned} H &= U + pV, \\ dH &= TdS + Vdp. \end{aligned} \tag{2.6}$$

2.2. Free Energy

Mathematically, Gibbs free energy is defined as

$$\begin{aligned} G &= U + pV - TS \\ &= H - TS. \end{aligned} \tag{2.7}$$

The fundamental thermodynamic equation for G follows from its definition and the fundamental equation for enthalpy (2.6),

$$\begin{aligned} dG &= dH - d(TS) \\ &= dH - TdS - SdT \\ &= TdS + Vdp - TdS - SdT \\ &= Vdp - SdT \end{aligned}$$

The natural variables of Gibbs free energy are the pressure p and the temperature T .

Statistical Significance

The relationship between the energy functions and the probabilities of microscopic configurations of a system is described by statistical mechanics.

In a NpT system, the enthalpy H is the relevant potential, and the Boltzmann factor becomes $e^{-\frac{H}{k_B T}}$. The overall probability of the system can be calculated by multiplying the Boltzmann probability by the number of possible microstates Ω at a given energy.

Solving Boltzmann's equation (2.2) for Ω ($\Omega = e^{\frac{S}{k_B}}$) and multiplying it with the Boltzmann factor leads to

$$\begin{aligned} \Omega e^{-\frac{U}{k_B T}} &= e^{\frac{S}{k_B}} \cdot e^{-\frac{H}{k_B T}} \\ &= e^{\frac{TS}{k_B T}} \cdot e^{-\frac{H}{k_B T}} \\ &= e^{-\frac{(H-TS)}{k_B T}}, \end{aligned}$$

so the overall probability of the system is proportional to Gibbs free energy (2.7):

$$P(i) \propto e^{-\frac{G}{k_B T}}$$

The Boltzmann distribution solved for Gibbs free energy yields the average free energy along a reaction coordinate z :

$$\begin{aligned} \rho(z) &= e^{-\frac{G(z)}{k_B T}} \\ \ln \rho(z) &= -\frac{G(z)}{k_B T} \\ G(z) &= -k_B T \ln \rho(z). \end{aligned} \tag{2.8}$$

Since free energies can be defined in terms of the respective partition functions, they contain information about the energy and entropy of the system.

2. Physical Principles

2.3. Potential of mean force

A potential of mean force [45] (PMF) describes the free energy surface over all the possible configurations of a given system. Its derivatives yield the average force along the direction of a chosen reaction coordinate z at equilibrium.

A reaction coordinate is an abstract coordinate representing progress along a reaction pathway. It is defined on some order parameter such as distance, torsion angles or the root mean square deviation of particles from reference states.

PMF curves can show the free-energy barrier between two conformations, the favorable position of a molecule in a system or the free energy difference between bound and unbound states 2.1.

The PMF function includes the potential energy effects of all the interactions in the system. A PMF along a reaction coordinate is defined by a partial integration of the partition function. For the NpT ensemble, it is given by

$$\begin{aligned} G(z) &= G(N, p, T; z) \\ &= -k_B T \ln Q_{NpT}(z) \\ &= -k_B T \ln \frac{1}{N! h^{3N}} \int_0^\infty dV \int e^{-\beta[U(q)+pV]} \delta[z - z(q^N)] dq \\ &= -k_B T \ln \langle \rho(z) \rangle. \end{aligned} \tag{2.9}$$

The probability distribution $\rho(z)$ along the reaction coordinate z is defined as

$$\langle \rho(z) \rangle = \frac{\int_0^\infty dV \int e^{-\beta[U(q)+pV]} \delta[z - z(q)] dq}{\int_0^\infty dV \int e^{-\beta[U(q)+pV]} dq}, \tag{2.10}$$

with U being the internal energy of state at the position q . Potentials of mean force can be calculated from simulations or experiments run at equilibrium conditions by creating histograms of the values of the chosen reaction coordinate to obtain an approximation of the probability distribution.

2.3. Potential of mean force

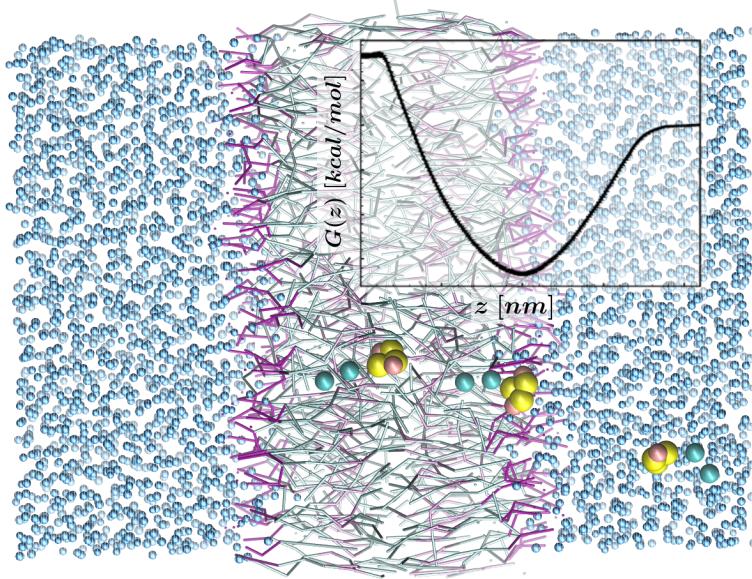


Figure 2.1.: The extrema of a PMF curve relative to the corresponding positions of the small molecule in the lipid bilayer. The energy minimum lies at the interface between the hydrophilic head groups of the lipids and the water phase, whereas the maximum is at the hydrophobic center of the bilayer.

The logarithmic relationship between the PMF and the distribution function visible from (2.8) and (2.9) shows that a small change in the free energy may lead to the probability changing by an order of magnitude from its most likely value.

In a MD simulation, configurations along a coordinate z around the energy minimum are sampled well, configurations around higher energies are rarely reached, and if the energy barrier separating two configurations becomes bigger than the amount of entropy per molecule $E = k_B T$, the simulation will either remain in the local minimum it started in or cross to other minima, but the transition state is unlikely to be reached.

This incomplete sampling means the simulation time needed to obtain accurate results for the PMF would increase to an impractical length. A way to overcome those obstacles is using biased ensembles with a configurational probability distribution that allows to sample states unlikely to be sampled in classical mechanical ensembles.

3. Materials and Methods

3.1. Molecular Models

In molecular mechanics, forces and energies are modeled as force fields of intramolecular interactions in a molecule. A force field defines the energy function of the system, the corresponding forces are given by the negative gradients of the energy function. The energy expressions in MD force fields are defined only up to typical bond lengths. A distinction is drawn between bonded (covalent) and non-bonded interactions.

3.1.1. Classical Force Fields

Force fields [38] consist of terms modeling for different kinds of interactions and describing different types of strain possible in a molecular system. The potential energy term contains a set of predefined parameters and is constructed of two parts,

$$U_{total} = U_{bonded} + U_{nonbonded}.$$

The bonded interactions make up the internal or intramolecular part of the energy,

$$U_{bonded} = U_{bonds} + U_{angles} + U_{dihedrals} + U_{impropers} + \dots,$$

and the non-bonded interactions represent the external or intermolecular energy component

$$U_{nonbonded} = U_{Van\;der\;Waals} + U_{electrostatic}.$$

The different kinds of interactions in the force field are represented by mathematical models like Hooke's law, harmonic oscillators, anharmonic oscillators like the Morse potential or analytical potentials like the Lennard-Jones potential (LJ) and Coulomb's law, that can also be fitted to experimental data.

3. Materials and Methods

3.1.2. Atomistic Resolution

All-atom force fields include explicit parameters for every atom type present in a system, including hydrogens. Molecules are modeled as mechanically connected systems of atoms. Electrons are treated together with the nuclei as effective atoms, therefore reactions or transformations involving bond breaking or bond formation are difficult to model. While they capture a high level of physical and chemical detail, because of their high computational cost, all-atom models are inadequate to simulate the large system sizes and long time scales of cellular processes.

CHARMM

An example of the atomistic approach is the CHARMM force field [46, 47], which is mostly aimed at biomolecular simulations. It is used for example to model fluid membranes by explicitly defining every atom type for all lipids. The CHARMM potential energy function contains a term for the bonded interactions,

$$U_{bonded} = \sum_{bonds} K_r(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \sum_{n=1}^6 K_{\phi,n} (1 + \cos(n\phi - \delta_n)) \\ + \sum_{impropers} K_\varphi(\varphi - \varphi_0)^2 + \sum_{Urey-Bradley} K_{UB}(r_{1,3} - r_{1,3;0})^2 + \sum_{\Phi,\Psi} U_{CMAP},$$

with two groups of corresponding force field parameters. The first is the equilibrium terms for bonds b_0 , angles θ_0 , the dihedral phase δ_n , impropers φ_0 and Urey-Bradley interactions $r_{1,3;0}$ for improved angle representation.

The second group is formed by the respective force constants K . The Urey-Bradley term is a harmonic potential acting on atoms 1 and 3 of the angles and was introduced to optimize their vibrational spectra. The CMAP procedure [48] is used to improve the representation of pairs of dihedral angles. It is mainly used for modeling the conformational properties of protein backbones.

The part of the potential energy function representing the non-bonded interactions is

$$U_{nonbonded} = \sum_{VdW} \epsilon_{ij} \left[\left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{min}}{r_{ij}} \right)^6 \right] + \sum_{elec.} \frac{q_i q_j}{\epsilon r_{ij}}.$$

The atomic Lennard-Jones parameters ϵ and r^{min} are combined with mixing rules for a pair of atoms i, j , $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ (geometric mean) and $r_{ij}^{min} = (r_i^{min} + r_j^{min}) / 2$ (arithmetic mean). Both non-bonded terms are modulated by a shifting or switching function or the use of periodic boundary conditions [38].

3.1.3. Coarse Graining

In coarse-grained (CG) force fields [49, 50], molecules are composed of pseudo-atoms, or beads. They represent groups of atoms, monomers, or even several monomer units. The interactions between those beads are modeled by interaction functions that take into account the average effect of the omitted degrees of freedom.

The main goal of coarse-grained models is the simulation of long time scales and large system sizes of biophysical processes such as protein folding, protein-ligand complex formation, nucleic acid structures or lipid bilayer organization.

The effectiveness of a CG model depends on its resolution, the mapping of the beads onto the corresponding atoms, the coarse grained potential function and the data used to parameterize the force field parameters.

There are essentially two different procedures followed to generate realistic and transferable CG force fields. In the bottom-up approach, CG interactions are extracted from atomistic reference simulations. Top-down approaches aim at reproducing experimental data, typically of thermodynamic properties, by starting with simple analytical interaction potentials and iteratively optimizing their parameters. Since both methods have their own benefits and disadvantages, most CG force fields rely on a combination of both.

While there is an inevitable loss of information inherent in the coarse graining process, it also provides powerful advantages. Because of the reduced computational cost and therefore increased time scales and system sizes, coarse graining enables the simulation of slow processes on a micro- to millisecond range and can be used for high-throughput screening and provides insights into the relevant physical details of the fundamental driving forces.

The Martini Force Field

The Martini force field [51, 52] is a coarse grained force field suitable for molecular dynamics simulations of biomolecular systems. It was originally developed for MD simulations of lipids [53], but versions for nucleic acids and proteins have since been added and it can be extended to include other chemical compounds

The Martini force field is parameterized hierarchically using a building block approach. Starting from the smallest increment of chemical groups, which can be parameterized using all-atom and experimental data, bigger subsystems are assembled and again tested against reference data from atomistic simulations and experiments. In the end, whole macromolecules can be assembled with realistic thermodynamic properties.

The Martini model uses a combination of the bottom up and the top down approaches, non-bonded interactions of beads are parameterized based on the reproduction of thermodynamic data, and bonded interactions are parameterized to match conformations of all-atom simulations or structural databases.

The complexity of real molecules is reduced by considering groups of atoms as building blocks. A four-to-one mapping is used for coarse graining, four heavy (non-hydrogen) atoms

3. Materials and Methods

are mapped to one coarse grained bead. Four general bead types are distinguished by their interaction types:

- C** Apolar, increasing polarity from subtypes 1 to 5,
- N** Intermediate, distinguished by hydrogen-bond donor or acceptor qualities,
- P** Polar, decreasing polarity from subtypes 1 to 5, and
- Q** Charged, again classification by hydrogen-bonding behavior.

This results in 18 different beads of the same size and different interactions. Special bead types have been added to enable the modeling of rings or the stacking forces of DNA molecules. Because of this simplified representation, the Martini model provides a $\sim 10^3$ speed-up factor of the sampling of the phase space.

The potentials for the non-bonded interactions are generally the same as for atomistic force fields, electrostatic interactions for beads with full charges and Lennard-Jones interactions for dispersion and repulsion processes.

The Lennard-Jones interactions depend on the nature of the beads represented by eight subtypes of decreasing hydrophobicity and adapted parameters for the charged beads to model for the increased repulsive forces of charged groups.

There is a polarizable water model [54] improving the behavior of charged beads if the environment changes.

Because several different sets of atoms can map to the same bead, coarse-grained models such as the Martini force field allow the fast screening of the chemical compound space to determine promising reactive groups with desired chemical interactions. On the other hand, this also means that results from coarse/grained simulations map to more than one atomistic system.

Another limitation of the Martini force field is the reduced contribution of entropy caused by the reduced degrees of freedom, which means that the temperature dependence and the driving forces of the model are unrealistic and have to be correlated to experimental or atomistic data.

3.2. Molecular Dynamics Simulations

Molecular dynamics (MD) is a computer simulation method for studying the movements of atoms and molecules at a finite temperature over a limited period of time.

The atoms and their connecting bonds and interactions, represented by the chosen force field, are modeled as a data structure. This initial structure has to be optimized before the actual MD simulation, by solving Newton's second equation of motion and finding a point in configurational space where the overall inter-atomic force on each atom converges to zero.

3.2. Molecular Dynamics Simulations

3.2.1. MD Workflow

The MD simulation approach is very similar to experimental procedures. A sample in the form of a system of N particles has to be prepared and brought into accordance with the appropriate ensemble conditions, and then the actual measurements have to be performed. The sources of problems affecting MD simulations are also very similar to those occurring in experimental setups, for example insufficient data collection, the wrong treatment of the sample system or the use of an inappropriate measuring protocol. The MD simulation steps will be described in the following sections.

Structure Minimization

In MD simulations, the systems have to be in thermal equilibrium, therefore their potential energy has to be minimized.

Using an adequate force field, the positions of the respective atoms can be adjusted to minimize the overall energy of the molecule. There are three spatial dimensions, therefore the number of degrees of freedom of a molecule with N atoms is $3N$. The energy results from mapping the spaces $V : \mathbb{R}^{3N} \mapsto \mathbb{R}$, which corresponds to a hypersurface in $3N + 1$ dimensional space. This hypersurface is called a potential energy surface (PES). The minima of the PES correspond to energetically favorable conformers of the molecule and can be found by minimization algorithms. Global minima are difficult to determine algorithmically, so the local minima are calculated and the smallest one is chosen.

Equilibration

To initiate a molecular dynamics simulation, the system has to be equilibrated to the simulation conditions. This is achieved by assigning a specific starting value T_0 to the temperature in the initial minimization and gradually heating the system.

The initial values are obtained by using a random number generator using the constraint of the Maxwell-Boltzmann distribution

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

The velocities are initialized at a temperature $t_0 + \Delta t$ and a MD simulation is run. This process is repeated, slowly increasing the temperature. All the velocities are multiplied by a common factor to obtain a new temperature. The equilibration is terminated if the system has reached the required temperature for the actual MD simulation.

MD simulations iteratively calculate interactions between atoms or molecules and the resulting spatial movement. Complex systems are generally modeled using force fields or

3. Materials and Methods

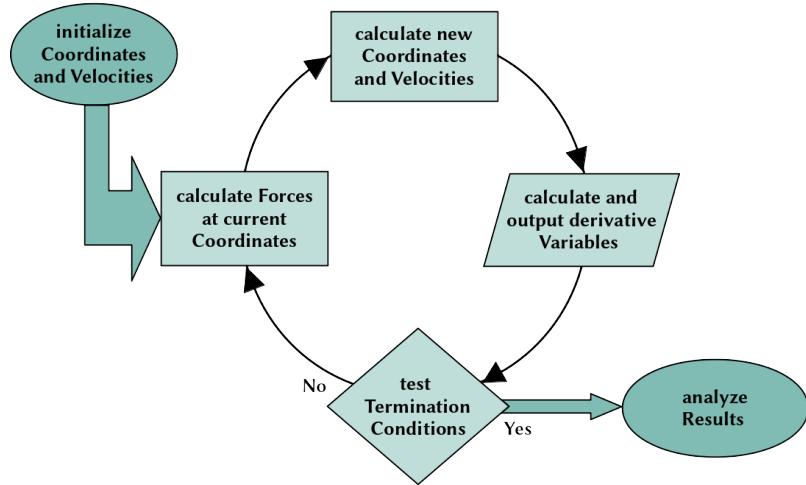


Figure 3.1.: Flowchart of the general MD simulation workflow

semi-empirical methods. With increasing computer performance available, quantum mechanics methods can also be used for medium sized systems. A MD simulation creates configurations corresponding to the chosen thermodynamic ensemble.

MD Simulation Cycle

The basic MD simulation process as shown in 3.1 generally follows five steps:

1 Initializing of Coordinates and Velocities

Values for initial coordinates and velocities have to be chosen depending on the system, from similar simulations or experimental data. Molecular structures have to be revised in terms of environmental parameters like pressure and temperatures, the applied experimental method and the completeness of the available data. Solvents and counter ions generally have to be added manually.

2 Calculation of the Forces

The selected force field defines the potential energy function, the negative gradient of which equates to the forces.

Non-bonded interactions are pairwise forces affecting both of the involved atoms with equal force in opposite directions. Bond stretches and bends have a pairwise effect, improper torsions are appropriately distributed, proper torsions affect the terminal atoms reflectively. According to Newton's third law, only half of the interactions have to be explicitly calculated.

Because the absolute value and the gradient of the force of non-bonded interactions

3.2. Molecular Dynamics Simulations

decrease with increasing distance, they have to be calculated only up to a preassigned distance. After that, they can either be set to zero (cut off) or continuously led to zero (switch off).

Atoms can also not be packed infinitely close, so with a suitably small cut off, the number of interactions that have to be calculated can be reduced to a constant number.

Periodic boundary conditions [38] are used to virtually enlarge the system making the filtering more efficient.

3 Calculation of new Coordinates and Velocities

The new coordinates and velocities for the next simulation step are given by the numerical time derivative of the equations of motion. The available methods are based on different discretizations of the differential operators and are varyingly complex and stable. In actual applications, complex predictor-corrector methods are implemented. In this work, the Leap Frog method was used to generate the trajectories. This basic algorithm with an explicit velocity function uses Taylor expansions up to an accuracy $\mathcal{O}(\delta t^2)$ for both positions and velocities.

Leap frog implements a mid-step in determining the velocities, thereby increasing its accuracy and stability:

$$\begin{aligned} r_i(t + \delta t) &= r_i(t) + v_i \left(t + \frac{\delta t}{2} \right) \\ v_i \left(t + \frac{\delta t}{2} \right) &= v_i \left(t - \frac{\delta t}{2} \right) + a_i(t) \delta t. \end{aligned} \tag{3.1}$$

The leap-frog algorithm performs the following steps:

- 1) Use the initial positions to calculate the forces acting on the particles and their accelerations $a_i(t)$ from the forces.
- 2) Use these accelerations and the velocities $v_i \left(t - \frac{\delta t}{2} \right)$ from the previous mid-step to derive estimates for the velocities $v_i \left(t + \frac{\delta t}{2} \right)$ at the next mid-step.
- 3) Obtain the new positions $r_i(t + \delta t)$ from the previous positions $r_i(t)$ and the velocities $v_i \left(t + \frac{\delta t}{2} \right)$.

Because the velocity calculation has an offset of $\frac{\delta t}{2}$ to the determination of the positions, the velocity $v_i(t)$ at the time step at which the positions are known has to be estimated from

$$v_i(t) = \frac{v_i \left(t + \frac{\delta t}{2} \right) - v_i \left(t - \frac{\Delta t}{2} \right)}{2}.$$

4 Calculation and Output of derived Parameters

Derived parameters of a system, for example absolute temperature, absolute volume, molecular mass, distance between atoms or functional groups or the frequency of oscillation, can be represented as functions of its trajectory. They can be calculated

3. Materials and Methods

by partial integration of the initial or previously calculated coordinates and velocities. Often, the absolute average over all time steps is given as an end result.

5 Testing of the Termination Conditions

A limited number of steps or the convergence of derived parameters are commonly used as termination conditions. The absolute time of the trajectory is given by the number and size of the time steps. Common time steps are on the order of 10^{-15} s, common trajectory lengths lie between 10^{-9} s and 10^{-3} s.

3.3. Free Energy Calculations

Free energy calculations provide a way to incorporate the statistical mechanical effects into virtual screening that are excluded by other methods. In this way, they facilitate the generation of results that are reasonably close to experimental data.

There are two general classes of free energy calculation approaches, one focused on computing the free energy difference between two discrete states of a molecular system, and one aiming at extracting a PMF profile along a reaction coordinate as described in section 2.3. For estimating free energy differences there exists a number of methods grouped together under the term Alchemical Transformations, see section 3.3.1, the most widely used method for calculating a PMF is umbrella sampling, described in section 3.3.2.

It is infeasible to directly compute the partition function of the chosen ensemble, because it is too computationally expensive to sample all the degrees of freedom of a multi-particle system [55]. But in biological experiments, all observed quantities can be classified as free energy differences between two thermodynamic states [56]. The relation of the NpT partition function and Gibbs free energy in equation (2.5) shows that calculating free-energy differences is equivalent to estimating the value of the partition function. Therefore, the difference in free energy ΔG between two states A and B is the quantity of interest in free energy calculations:

$$\Delta G = -\beta^{-1} \ln \frac{Q_B}{Q_A}. \quad (3.2)$$

The two required macroscopic states of the system have to be defined through the volume in phase space accessible to the respective system.

This can describe for example different conformations of a flexible molecule, complexes and unbound states of protein-ligand interactions, or different ligands bound to the same target molecule. The differences between the states of interest can be expressed by different values of state points or different Hamiltonians.

The results of free-energy methods are statistical estimates of the correct free energy of a state change, depending on the statistical mechanics ensemble used for the calculations.

3.3. Free Energy Calculations

3.3.1. Alchemical Transformations

Alchemical Transformations [57] are a subset of free energy methods, where the free-energy difference of physical processes is estimated by transforming one state into another via a pathway of unphysical intermediate states $\{0, \dots, N\}$,

$$\Delta G = -\beta^{-1} \ln \frac{Q_N}{Q_0} = -\beta^{-1} \ln \frac{Q_1}{Q_0} + \ln \frac{Q_2}{Q_1} \dots \ln \frac{Q_N}{Q_{N-1}} = \sum_{n=1}^{N-1} \Delta G_{n \rightarrow n+1}.$$

This is achieved by gradually transforming the Hamiltonian with an alchemical variable λ . Since the difference in free energies between two states A and B is relevant and the free energy is a state function and therefore does not depend on the path the system took to reach it, the free energy difference can be efficiently calculated along a closed cycle of reactions, as described in section 3.3.1.

With alchemical transformations, the conformational complexity of biochemical reactions, like protein folding or enzyme-ligand binding, can be covered. Using intermediate states facilitates the efficient computation of free energy differences between Hamiltonians with little phase space overlap, where configurations sampled with one end state Hamiltonian are very improbable with the other end state Hamiltonian.

Equilibrium Methods

A widely used approach to alchemical transformations are the equilibrium methods, where in their simplest form [58], the change in free energy is calculated by averaging over two points in phase space Γ ,

$$\Delta G = -\beta^{-1} \ln \langle e^{\mathcal{H}_1(\Gamma) - \mathcal{H}_0(\Gamma)} \rangle,$$

and is averaged over the equilibrium ensemble of the end states. The equilibrium methods generally follow three steps:

1. Selection of alchemical intermediates

The efficiency of alchemical free energy calculations depends on an optimal choice of intermediate states that aims at increasing the phase space overlap. The simplest approach is to follow a linear path in the two end point Hamiltonians:

$$\mathcal{H}(\lambda) = \lambda \mathcal{H}_1 + (1 - \lambda) \mathcal{H}_0.$$

If the alchemical transformation involves the creation or annihilation of particles, this approach causes a singularity (an infinite potential), at $r = 0$ because of the repulsive term in the LJ potential as shown in figure 3.2. This leads to the calculations to converge very slowly and/or become numerically unstable.

A way to overcome this problem is to use a soft-core approach, see section 3.3.1. This

3. Materials and Methods

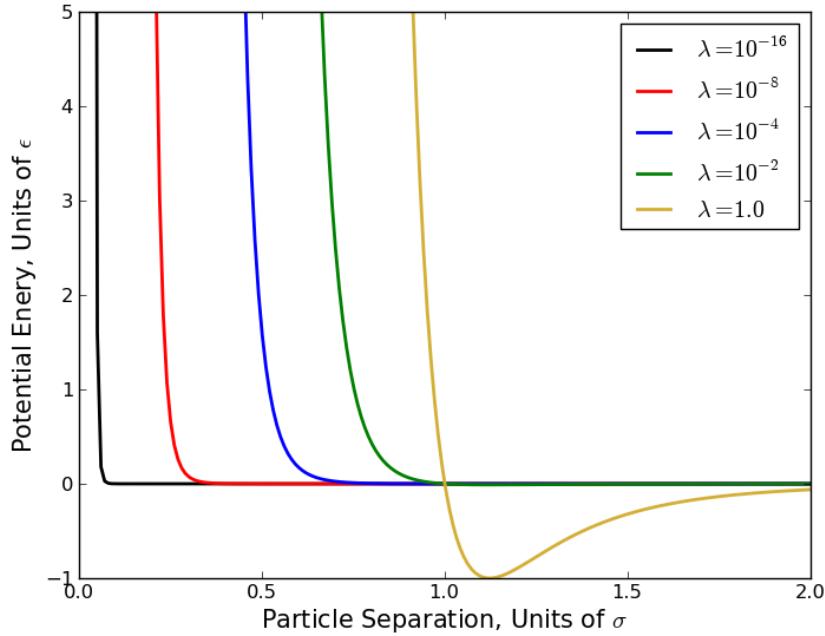


Figure 3.2.: Linearly scaled Lennard-Jones potential, the singularity is still present even with very small λ . [61]

avoids the appearance of singularities and still yields comparable results, because the free energy of the end states is independent of the path taken from one end state to another.

2. Generation of uncorrelated samples at each intermediate

For a correct estimate of the free energy, a sufficiently large number of uncorrelated configurations at every intermediate Hamiltonian is needed.

Different alchemical intermediates lead to different correlation times, because for example uncharged or uncoupled states imply lower energy barriers than states where charges or bonds have to be transformed. To ensure comparable sampling times for all intermediates, a biasing weight can be applied that reduces the correlation time of each λ . The results have to be evaluated using a reweighting method.

3. Estimation of the free energy difference between the end states

There are different analysis methods, grouped by their general principles. Thermo-dynamic Integration [59] methods integrate over the derivatives of the free energy with respect to λ . In Free Energy Perturbation [60, 59, 58] approaches, the free energy difference between two states is estimated using the total energy of the initial system and a perturbation function.

In this work, a histogram method based on the probability distributions of the intermediate states was used, as described in section 3.3.1.

3.3. Free Energy Calculations

Soft-Core Potentials

Soft-core potentials are a standard approach to keep alchemical transformations numerically stable. They soften the potential at the end points by coupling the configurational variable r to λ . A generalized form of the LJ soft-core potential [62, 63] is represented by:

$$U_{ij}(\lambda, r_{ij}) = 4\epsilon_{ij}\lambda^n \left[\left(\frac{1}{\alpha(1-\lambda)^m + (\frac{r}{\sigma})^l} \right)^{12/l} - \left(\frac{1}{\alpha(1-\lambda)^m + (\frac{r}{\sigma})^l} \right)^{6/l} \right]$$

where α is a constant and $l, m, n \in \mathbb{Z}^+$. Different values for α, l, m and n can significantly improve the variance [64, 65].

Thermodynamic Cycle

The free energy depends on the internal energy, the enthalpy and the entropy, which are state quantities of a thermodynamic system describing its equilibrium state independently of the path the system took to reach equilibrium. Therefore, the sum of the free energy differences along a closed cycle of reactions will always be zero. This means that the double difference in free energy $\Delta\Delta G$ of the observed process between two states of a system can be calculated.

For the relative binding affinity of different druglike molecules to a lipid, the thermodynamic cycle is given in figure 3.3, with

$$\Delta\Delta G = \Delta G_{binding}^B - \Delta G_{binding}^A = \Delta G_{mutation}^B - \Delta G_{mutation}^A$$

Because systematic errors cancel out when the free energy differences are subtracted, $\Delta\Delta G$ is more accurate than the ΔG values used to calculate it.

Using a thermodynamic cycle, relative free energies can be calculated without having to simulate the corresponding reaction.

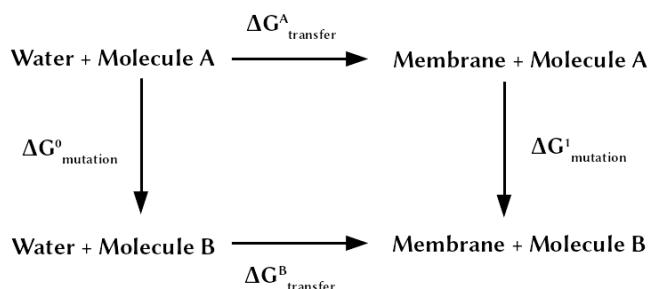


Figure 3.3.: Thermodynamic cycle for the insertion of different states of a molecule into a lipid membrane.

3. Materials and Methods

Histogram Methods

The alchemical transformations approach called “probability distribution” or “histogram method” [45] is based on estimating the probability density functions of the intermediates. The connection between the probabilities of states and the partition function is given by equation (2.3), so for example the probability of a system having a certain potential energy U is

$$P(U) = \frac{e^{-\beta U} \Omega(U)}{Q_{NpT}}.$$

An initial System A can be transformed into a system B through the continuous change of λ , so that λ_0 belongs to system A and λ_1 to system B. The alchemical variable λ can stand for a macroscopic variable like the temperature T , a parameter transforming the Hamiltonian \mathcal{H}_0 to \mathcal{H}_1 or the generalized coordinates of the system that represent a torsion angle or an intermolecular distance.

The probabilities of the different intermediates and, because of the equality (2.3), the relative partition functions become

$$P_i = P(\lambda_i) = \frac{1}{N!h^{3N}} \int_0^\infty dV \int d\lambda e^{-\beta[\mathcal{H}(q_i, p_i; \lambda) + pV]} \delta(\lambda - \lambda_i) dq_i dp_i = Q_{NpT;i}. \quad (3.3)$$

Here, the Dirac delta function [35] $\delta(\lambda - \lambda_i)$ means that the value of λ remains unchanged during the integration: $\lambda_i = \lambda$.

Combining equations (3.3) and (3.2) leads to the probability density term for Gibbs free energy,

$$\Delta G = -\beta^{-1} \ln \frac{P_B}{P_A}.$$

Probability distribution methods generate the probability distribution function $P(\lambda)$ for $\lambda \in \{\lambda_0 \dots \lambda_1\}$ in form of a histogram that can be processed by the reweighting method described in section 3.3.3, which can be evaluated using methods like MBAR, explained in 3.3.3.

3.3.2. Umbrella Sampling

Umbrella Sampling [45, 66, 67] is a biased MD simulation method where the potential function is modified so that the energetically unfavorable conformation states are sampled adequately. The weighting function used to modify the potential function will become larger with increasing distance from the equilibrium state of the system, so that a simulation with a biased potential will be guided along a predetermined reaction coordinate away from the equilibrium state.

There are two general approaches to the biasing potential. The first is the adaptive bias potential, where the whole range of the reaction coordinate z is covered in one simulation

3.3. Free Energy Calculations

starting out from an initial guess for the bias potential $w(z)$, which is iteratively adapted to achieve uniform distribution. The second potential is the harmonic bias potential described in section 3.3.2 used in this work, where multiple separate simulations are performed along the reaction coordinate. The probability distribution varies less inside each window than it would along the whole reaction coordinate, leading to more efficient calculation of the free energy.

A biased ensemble is created by adding a biasing potential w_i for each window i ,

$$-\beta(U(q) + pV)^w = -\beta((U(q) + pV + w_i(z))) \quad (3.4)$$

which only depends on the reaction coordinate z . MD simulations of a biased system result in a biased histogram, the biased probability distribution $\rho_i^w(z)$ along the reaction coordinate,

$$\begin{aligned} \rho_i^w(z) &= \frac{\int_0^\infty dV \int e^{-\beta[(U(q)+pV)+w_i(z'(q))]} \delta[z'(q) - z] dq}{\int_0^\infty dV \int e^{-\beta[(U(q)+pV)+w_i(z'(q))]} dq} \\ &= e^{[-\beta w_i(z)]} \times \frac{\int_0^\infty dV \int e^{-\beta(U(q)+pV)} \delta[z'(q) - z] dq}{\int_0^\infty dV \int e^{-\beta[(U(q)+pV)+w_i(z'(q))]} dq}, \end{aligned} \quad (3.5)$$

with the bias only dependent on z and the enumerator integrated over all degrees of freedom but z .

Equating the biased and the unbiased probability distributions defined in equations (3.5) and (2.10), leads to

$$\begin{aligned} \rho_i(z) &= \rho_i^w(z) e^{[\beta w_i(z)]} \times \frac{\int_0^\infty dV \int e^{-\beta[(U(q)+pV)+w_i(z(q))]} dq}{\int_0^\infty dV \int e^{-\beta(U(q)+pV)} dq} \\ &= \rho_i^w(z) e^{[\beta w_i(z)]} \langle e^{[-\beta w_i(z)]} \rangle. \end{aligned}$$

With the unbiased potential, an exact reconstruction of the total free energy of an umbrella simulation is possible from the $\rho_i^w(z)$ provided by a MD simulation of the biased system and the analytically derived $w_i(z)$,

$$G_i(z) = -\beta^{-1} \ln \rho_i^w(z) - w_i(z) + f_i,$$

with the free energy f_i introduced by adding the biasing potential to the energy given by equation (3.4), $f_i = -\beta^{-1} \ln \langle e^{[-\beta w_i(z)]} \rangle$.

To combine the free energies of several windows, the forces f_i connecting them have to be calculated. They depend on the bias potential and connect the different free energies $G_i(z)$ from the different windows,

$$\begin{aligned} e^{(-\beta f_i)} &= \langle e^{[-\beta w_i(z)]} \rangle \\ &= \int \rho(z) e^{[-\beta w_i(z)]} dz \\ &= \int_0^\infty dV \int e^{-\beta[G_i(z)+w_i(z)]} dz. \end{aligned} \quad (3.6)$$

3. Materials and Methods

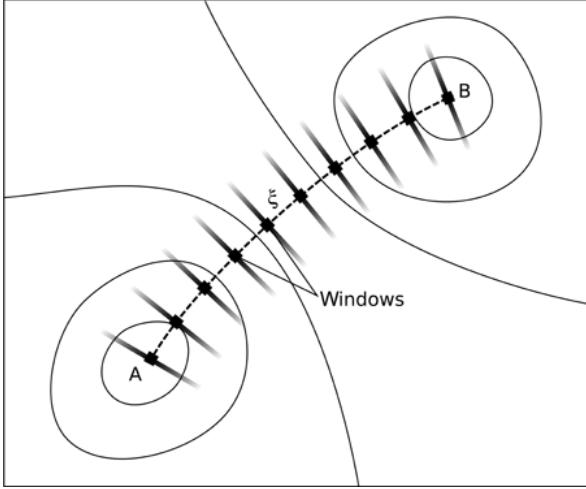


Figure 3.4.: Partition of the reaction coordinate into evenly spaced windows [67]

The f_i depend on the global unbiased distribution $\rho(z)$ and therefore can not be directly obtained from biased sampling. One possible way to do so is by using histograms 3.3.3, which have to be unweighted and combined to obtain the underlying free energy function.

Harmonic Bias Potentials

The reaction coordinate is divided into a series of windows as shown in figure 3.4, and a constraint is applied to keep the system close to the center of the window i at z_i . To achieve this, a series of bias potentials $w(z)$ are used which constrain the reaction coordinate to specific values. Generally, a simple harmonic bias of the strength k is sufficient:

$$w_i(z) = \frac{k}{2}(z - z_i)^2,$$

with z_i being the center of the window and k the force constant leading to the weighted potential. The constant k has to be initially chosen and has to be large enough for the system to cross the energy barrier but small enough to guarantee a broad enough probability distribution. The probability distributions of the different windows have to overlap sufficiently for the sampling results to yield an accurate PMF.

3.3.3. Histogram Reweighting

In histogram reweighting [45, 38], macrostate probabilities are adjusted along microstate probabilities by using the equilibrium probability distribution obtained from a simulation at one state point to estimate the probability distribution of a neighboring state point.

3.3. Free Energy Calculations

Histograms give a graphical representation of the probability distribution of numerical data. In thermodynamics, they relate to the fluctuations of observables in the ensemble, and contain information about the underlying statistical-mechanical properties of a system. Histograms are represented as arrays or matrices with indices corresponding to variations in a parameter like the number of particles, energy or distances.

The partition function as the sum over macroscopic states of a system given by equation (2.4) consists of individual terms representing individual states. Those terms are proportional to the probabilities of the individual states in the associated ensemble, and a histogram calculated from a simulation measuring the sampling frequency of the macrostates reflects those probabilities.

In a simulation in the NpT ensemble with a given temperature and pressure, the general probability to find a microstate with an energy U and a volume V is given by equation (2.3). Since in the NpT ensemble the volume and the energy fluctuate, the number of times that a certain volume and a certain energy are observed is stored in a two-dimensional histogram $H(V, U)$. For a simulation performed at temperature $\beta_0 = \frac{1}{k_B T_0}$ and a pressure p_0 , the probability that a system is sampled with a certain combination of V and U becomes

$$\rho(p_0, \beta_0; V, U) = \frac{H(V, U)}{K}, \quad (3.7)$$

with $K = \sum_V \sum_U H(V, U)$ being the total number of configurations sampled during the simulation. The probability distribution for a NpT ensemble to have a specific energy and volume at some initial conditions p_0 and β_0 is given by

$$\rho(p_0, \beta_0; V, U) = \frac{e^{-\beta_0(U+p_0V)} \Omega(V, U)}{Q_{NpT}(p_0, \beta_0)}. \quad (3.8)$$

This means that the NpT partition function for the ensemble at the conditions p_0 and β_0 can be estimated by

$$Q_{NpT}(p_0, \beta_0) = \sum_V \sum_U e^{-\beta_0(U+p_0V)} \Omega(V, U).$$

From equations (3.7) and (3.8) follows an estimate for the density of states with the specific energy U and a volume V :

$$\begin{aligned} \Omega(V, U) &= \frac{\rho(p_0, \beta_0; V, U) Q_{NpT}(p_0, \beta_0)}{e^{-\beta_0(U+p_0V)}} \\ &= \frac{Q_{NpT}(p_0, \beta_0)}{K} \frac{H(V, U)}{e^{-\beta_0(U+p_0V)}}, \end{aligned} \quad (3.9)$$

where $\frac{Q_{NpT}(p_0, \beta_0)}{K}$ is a proportionality constant independent of temperature and pressure. The probability to observe the system with the volume V and energy U at another state point with pressure p and temperature T is given by the general definition of the probability

3. Materials and Methods

(2.3). Inserting equation (3.9) in equation (2.3) shows that the probability $\rho(p, \beta; V, U)$ can be estimated by reweighting the probability $\rho(p_0, \beta_0; V, U)$ from the simulation at p_0 and β_0 :

$$\rho(p, \beta; V, U) = \rho(p_0, \beta_0; V, U) \frac{Q_{NpT}(p_0, \beta_0)}{Q_{NpT}(p, \beta)} e^{[-(\beta - \beta_0)U - (p\beta - p_0\beta_0)V]}.$$

With this relation, the ensemble average of any property $F(p, \beta)$ can be estimated,

$$\langle F \rangle = \sum_V \sum_U F(V, U) \rho(p, \beta, V, U)$$

without having to simulate the system at p and β , if the states are similar enough so the probability distributions of p_0, β_0 and p, β overlap.

The Weighted Histogram Analysis Method

The Weighted Histogram Analysis Method (WHAM) [68, 69] is an extension to the Ferrenberg - Swendsen [70, 71] method, generalizing it to simulations with different sets of coupling parameters used to enhance the conformational sampling in applications such as biomolecular systems.

The statistical errors of a simulation increase at the margins of the measured distribution. WHAM minimizes the statistical error by using a multiple histogram approach, where the data collected in multiple simulations is combined. The contribution of each individual simulation to the combined estimate is weighted according to the magnitude of errors in the corresponding histogram, so that simulations with results closer to the reweighting conditions have a bigger contribution to the estimated property averages.

WHAM is often used to analyze umbrella sampling simulations, where for each window, the respective energy has been replaced by a biased potential. From each of these simulations, a biased probability distribution $\rho_i^w(xi)$ can be stored as a normalized histogram of the values of z occurring during the simulation i . The corresponding unbiased probability distribution $\rho_i(z)$ from simulation i is defined as

$$\rho_i(z) = e^{-\beta[w_i(z) - f_i]} \rho_i^w(xi).$$

WHAM provides a method for recombining the individual unbiased probability distributions $\rho_i(z)$. The global distribution function is calculated by a weighted average of the distributions of the individual windows:

$$\rho(z) = \sum_i^R p_i(\rho_i(z)),$$

where R is the total number of windows. The weights p_i have to be normalized to $\sum_{i=1}^R p_i = 1$ and are chosen in order to minimize the statistical error of $\rho(z)$:

$$\frac{\delta \sigma^2 \rho(z)}{\delta p_i} = 0,$$

3.3. Free Energy Calculations

therefore they are defined as

$$p_i = \frac{K_i e^{[-\beta w_i(z) + \beta f_i]}}{\sum_j^R K_j e^{[-\beta w_j(z) + \beta f_j]}},$$

where K is the total number of steps sampled for simulation i . Inserting the weights P_i , the global probability distribution becomes the first WHAM equation:

$$\begin{aligned} \rho(z) &= \sum_{i=1}^R \frac{H_i(V, U) e^{-\beta(w_i(z) - f_i)}}{\sum_{j=1}^R K_j e^{-\beta(w_j(z) - f_j)}} \rho_i(z) \\ &= \sum_{i=1}^R \frac{H_i(V, U)}{\sum_{j=1}^R K_j e^{-\beta(w_j(z) - f_j)}} \rho_i^w(z). \end{aligned} \quad (3.10)$$

The free energies f_i can be calculated using the second WHAM equation derived from (3.6):

$$\begin{aligned} e^{-\beta f_i} &= \int \rho(z) e^{(-\beta w_i(z))} dz \\ &= \int dz \sum_{k=1}^R \frac{H_k(V, U) e^{(-\beta w_i(z))}}{\sum_{j=1}^R K_j e^{-\beta(w_j(z) - f_j)}} \rho_k^w(z). \end{aligned} \quad (3.11)$$

The free energies f_i are initially unknown, but they can be solved self-consistently by starting from an initial guess and iterating between equations (3.10) and (3.11) until they converge. Since the WHAM equations only provide relative free energies, the initial calculation is generally done with $f_0 = 0$. For a large number of histogram bins or poor overlap, this convergence can be slow.

Multistate Bennett Acceptance Ratio

The Multistate Bennett Acceptance Ratio (MBAR) method [72] estimates the free energy difference by considering all intermediate states simultaneously. It is able to predict the free energy at an unsampled state and gives the statistical uncertainty of the calculated free energies. It reduces to WHAM 3.3.3 where bin widths of zero are considered.

MBAR is based on a set $K \times K$ weighting functions $\alpha_{AB}(\vec{q})$ to minimize the variance during the global reweighting of all K states. Keeping the same configuration of the coordinates, but switching potential functions between the states A and B [73] leads to the relation

$$Q_A < \alpha_{AB} e^{-\beta U_B} >_A = Q_B < \alpha_{AB} e^{-\beta U_A} >_B. \quad (3.12)$$

The optimal weighting functions α_{AB} can be determined using extended bridge sampling [74], which leads to

$$\alpha_{AB} = \frac{N_B \hat{c}_B^{-1}}{\sum_{k=1}^K N_k \hat{c}_k^{-1} e^{-\beta U_k}}. \quad (3.13)$$

3. Materials and Methods

Substituting α_{AB} in (3.12) with (3.13) leads to the equation for the estimated Gibbs free energies,

$$G_i = -\beta^{-1} \ln \sum_{j=1}^K \sum_{n=1}^{N_k} \frac{e^{[-\beta(U_i(p_{j,n}) + pV_{j,n})]} }{\sum_{k=1}^K N_k e^{[\beta G_k - \beta(U_k(p_{k,n}) + pV_{k,n})]}}, \quad (3.14)$$

where the G_i are the free energies of each state, $p_{k,n}$ is the n 'th sample from the k 'th state and the U_i are the potential of the K states. The reason why only a single free energy is estimated here, is that free energies for a given set of states are only unique up to an additive constant, so one energy must be taken as a reference. The result is still a free energy difference. MBAR is valid for any generalized potential as a function of the applied thermodynamic constraints. Therefore, the appropriate subset of the additive terms in the reduced potential equation [72]

$$u_i(x) \equiv \beta_i [U_i(q) + p_i V(q) + \mu_i N(q)]$$

can be selected to define the reduced potential $u_i(x)$ for the thermodynamic state i in the ensemble of choice. The reduced potential generalizes the MBAR equations to the computation of free energy differences $\Delta f_{AB} \equiv \beta_B F_B - \beta_A F_A$ between adjacent states. This transforms (3.14) to the estimating equations for the dimensionless free energies

$$f_i = -\ln \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{e^{[-u_i(p_{j,n})]} }{\sum_{k=1}^K N_k e^{[f_k - u_k(x_{j,n})]}},$$

where again $p_{j,n}$ is the n 'th sample from the j 'th state. The resulting free energies will depend on the selected subset of constraints. Since in this expression the origin of the samples is irrelevant, it is equivalent to the sum over all $N = \sum_{j=1}^K N_j$, which leads to

$$f_i = -\ln \sum_{n=1}^{N_j} \frac{e^{[-u_i(p_n)]}}{\sum_{k=1}^K N_k e^{[f_k - u_k(x_n)]}} = -\ln \left\langle \frac{e^{[-u_i(p_n)]}}{\sum_{k=1}^K \frac{N_k}{N} e^{[f_k - u_k(x_n)]}} \right\rangle.$$

The forces f_i can be given as the ensemble average over the ensemble consisting of samples taken from each state i at a $\frac{N_i}{N}$ th of the overall simulation time.

3.4. Simulation Setup

The members of each lipid family chosen for the simulations were *tetraoleoyl cardiolipin* (TOCL2), which constitutes the largest portion of all CL types found in the inner mitochondrial membrane, it's precursor *1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol* (POPG) and *1,2-dioleoyl-sn-glycero-3-phosphocholine* (DOPC), which has the same hydrocarbon chains as TOCL2. NAO is available in several versions, *10-N Nonyl-acridine orange* was used in the simulations because it showed the highest binding affinity to CL [33].

3.4. Simulation Setup

3.4.1. Coarse-grained Resolution

All coarse-grained simulations were run in Gromacs 5 [75] using the Martini force field version 2.0, 06/2015, described in section 3.1.3. The membranes were solvated with the Martini polarizable water model [52, 54].

Periodic boundary conditions [38] were applied in all simulations, the Electrostatic interactions were controlled with Particle-mesh Ewalds, The Van der Waals interactions were set to “cutoff”. The cutoff distance for both Van der Waals and Coulomb interactions was 1.1 nm. For temperature coupling, a reference temperature of 300 K and a time constant of 1.0 ps was used. Pressure coupling was done to a reference pressure of 1 bar, the time constant was set to 12.0 ps and the compressibility to $3e - 4 \text{ bar}^{-1}$. The temperature and pressure of all parts of the system were coupled, the NAO variants, the lipid membrane, the solvent and the ions. Constraints were imposed on the bonds using Lincs, with the accuracy set to 1 and a maximum allowed bond rotation of 30° per simulation step. If not stated otherwise, the algorithms used are implemented in Gromacs and can be looked up in the manual [76]. All settings not explicitly mentioned are Gromacs 5 default settings.

Systems

The coarse-grained lipid structures for TOCL2, POPG, and DOPC were obtained from the CHARMM-GUI Martini Maker [77], which was also used for creating homogeneous bilayers for each lipid.

The initial box size for the CL was 8 nm in x and y direction and 11 nm in z direction, containing 49 CL molecules in every membrane layer. The number of water molecules was automatically set to 3609 by selecting a water thickness of 32.5 Å. The overall charge of the system was equilibrated by adding 196 sodium ions, two for each unprotonated CL molecule. For the PG and PC simulation boxes, the same procedure was used. The box dimensions for each of those two lipids were 6 nm in x and y , and 11 nm in z direction. Each layer of both lipid membranes contained 59 molecules. For PG, the number of water molecules was automatically set to 1957 and the charge of the system was equilibrated by adding 118 sodium ions. To the PC box, 2237 water molecules were added. Because PC is zwitterionic, its charge is intrinsically at equilibrium. So two sodium and two chloride ions were added to enable later equilibration of the net charge after the NAO molecule was added.

The coarse-grained representation of NAO has been parameterized in our group, the bead types from this original parameterization are shown in figure 3.5 (A). The variations for the free energy calculations were obtained by manually changing selected Martini beads (figure 3.5 (B) and (C), the asterisks standing for any member of the respective bead group).

3. Materials and Methods

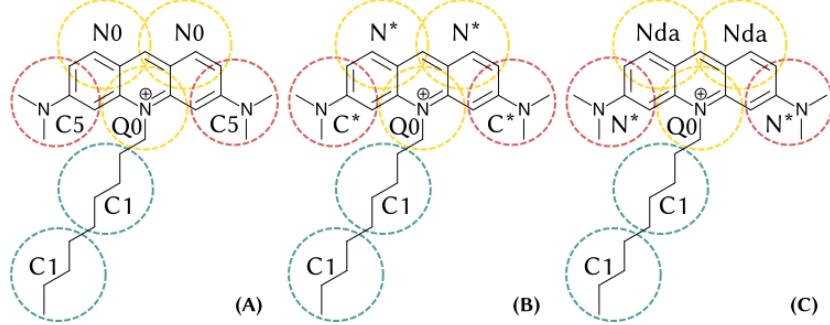


Figure 3.5.: (A) The NAO molecule with the bead types from the original parameterization. The beads in the aromatic system are S-types for representing rings, the letter S is omitted for clarity. (B) The changes to the bead types in the first alchemical transformation step. (C) The changes made in the second alchemical transformations step. For the function of the different bead types, refer to the section on the Martini force field: 3.1.3.

Alchemical Transformations

Each variety of the NAO molecule was placed once at the membrane–water interface and once in bulk water. From each simulation box, one sodium ion was removed to keep the overall charge of the system neutral. Both the positions of the molecule and the lipid bilayer were fixed using the Gromacs pull code. The reference distance at timestep $t = 0$ was set to the desired position of the molecule in the membrane and not modified during the simulations. Only the Van der Waals interactions had to be transformed, which was done in 24 evenly distributed states, using a soft-core potential (described in section 3.3.1) with Gromacs’ default value for α (0.5) to avoid singularities. No constraints were imposed on the molecules. All lambda points were written out and analyzed with pymbar [78], which implements MBAR described in 3.3.3.

The systems were minimized using steepest descent for a maximum number of 500 steps or until the cutoff of 10 kJ/(mol nm) was reached.

The equilibration was performed using the Gromacs implementation of the Leap Frog algorithm shown in equation (3.1) for integrating Newton’s equation of motion. The time step was set to 0.01 ps and the equilibration ran for 8 ns. The temperature was adjusted using velocity rescaling (see section 2.1.2) with an additional stochastic term ensuring a correct kinetic energy distribution, refer to the Gromacs manual [76]. For controlling the temperature during equilibration, semiisotropic Berendsen pressure coupling (section 2.1.2) was used.

The MD simulation was done again using Leap Frog, with a time step of 0.02 ps, for a duration of 40 ns. The Gromacs implementation of velocity rescaling was used to adjust the temperature of the system, and the pressure was controlled by the Parrinello-Rahman barostat described in section 2.1.2, with semiisotropic coupling.

3.4. Simulation Setup

Umbrella Sampling

Two copies of a NAO derivatives were inserted into the membrane system and two sodium ions were removed. The small molecules were pulled along the bilayer normal (z), using the Gromacs pull code, to facilitate complete sampling of the system. The positions of the two molecules are evenly distributed along the box coordinate z , each position corresponds to one umbrella window as one simulation system. Umbrella pulling with a harmonic bias potential was applied with the geometry setting “distance”. The pull groups were the two copies of the small molecules and the lipid membrane.

The systems were minimized with steepest descent for a maximum number of 1000 steps or until the cutoff energy of 10 kJ/(mol nm) was reached.

The equilibration step was conducted with Leap Frog and a time step of 0.01 ps, for a duration of 10 ns. The temperature was controlled with the velocity rescaling thermostat, the pressure was adjusted by semiisotropic Berendsen coupling.

A simulated annealing [79] step was included to increase the probability of the system to reach the global energy minimum without getting trapped in a local minimum. The process ran for 10 ns with a time step of 0.01 ps, using the Leap Frog algorithm.

The annealing was done with a single sequence and two annealing points per group. The small molecules, the lipid membrane, the ions and the water were each cooled down to 0 K at 0 ps, and then heated up to reach 300 K after 2.5 ns.

The initial temperature of the system was set by velocity rescaling, the pressure was controlled by semiisotropic Berendsen coupling.

The MD trajectory was calculated over 200 ns with a time step of 0.02 ns, using the Leap Frog algorithm. Temperature and pressure were adjusted again using velocity rescaling and the Parrinello-Rahman barostat with semiisotropic scaling.

The results from the different umbrella windows were combined using WHAM, described in section 3.3.3.

3.4.2. Atomistic Resolution

The atomistic simulations were performed with Gromacs 5 and using CHARMM version 36 for lipids, refer to section 3.1.2. As solvent, the polarizable TIP3P water model [80] was used. All simulations were done using periodic boundary conditions.

The electrostatic interactions were simulated using Particle-mesh Ewalds. The interpolation order for PME was set to 6 and the relative strength of the shifted potential at the coulomb interaction cutoff distance was $1e - 6$. The setting for the van der Waals interactions was “cutoff”, combined with the modifier “force-switch”. The cutoff distance for both Van der Waals and Coulomb interactions was 1.2 nm.

For temperature coupling, a reference temperature of 300 K and a time constant of 1.0 ps was used. Pressure coupling was done to a reference pressure of 1 bar, the time constant was set to 0.5 ps and the compressibility to $4.5e - 5 \text{ bar}^{-1}$. The temperature and pressure of all

3. Materials and Methods

parts of the system were coupled, the NAO variants, the lipid membrane, the solvent and the ions. All settings and algorithms not explicitly mentioned are Gromacs 5 defaults.

Systems

For the atomistic simulations, homogeneous membrane systems of PG and CL were created using the Bilayer Builder of the CHARMM-GUI web interface [81]. The box dimensions, the numbers of lipids per box and the numbers of neutralizing ions for each lipid were the same for each lipid type as with the coarse-grained simulations. The CL simulation box contained 15340 water molecules, the PG membrane was solvated in 8790 water molecules.

The atomistic representations for the NAO molecule and its different variations for the alchemical free energy calculations were automatically parameterized using CGenFF [82, 83, 84].

Alchemical Transformations

The same approach for placing the molecules at the interface or in bulk water was used as in the coarse-grained simulations, and one sodium ion was again removed from each box to keep the net charge neutral. As recommended for the use of CHARMM with Gromacs, constraints were imposed on hydrogen bonds. The constraint algorithm used was Lincs, with the accuracy set to 12.

A soft-core potential was used for the van der Waals interactions, again setting the default value for α . For the Coulomb interactions, the generally more computationally efficient linear interpolation was chosen. The van der Waals and Coulomb interactions of the small molecules were decoupled both in state A and state B.

Because atoms were transformed, new atoms were added and the partial charges of the affected parts of the molecule changed, the corresponding interactions had to be transformed. This was done distributed over 24 intermediate states, as shown in table 3.1. All lambda values were again written out and analyzed with MBAR using the python tool pymbar [78]. The minimization of the systems representing the different states was done in two steps. First using steepest descent with a maximum force cutoff of 100 kcal/mol and a maximum number of 5000 steps. The second minimization round was done with limited-memory BFGS, a quasi-Newton algorithm for energy minimization. The second minimization ran for a maximum number of again 5000 steps or until the force cutoff of 10 kcal/mol was reached. The systems were equilibrated first for 100 ps with a time step of 0.002 ps in a NVT ensemble, with an integrator based on Langevin dynamics. The temperature coupling was handled by this integrator using the Langevin thermostat described in section 2.1.2. In a second step, the systems were equilibrated under constant temperature and pressure with the same time step for another 100 ps. The temperature was again controlled by the Langevin Thermostat, the pressure was adjusted by the Parrinello-Rahman Barostat with a time constant of 5.0 ps as

3.4. Simulation Setup

Table 3.1.: The λ values for the intermediate states used to transform each interaction type.

| Interaction Type | States | | | | | | |
|-------------------------|---------------|------|------|------|------|------|------|
| Van der Waals | | | | | | | |
| λ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| | 0.00 | 0.17 | 0.33 | 0.50 | 0.67 | 0.83 | 1.00 |
| Coulomb | | | | | | | |
| λ | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | 0.00 | 0.17 | 0.33 | 0.50 | 0.67 | 0.83 | 1.00 |
| Bonded | | | | | | | |
| λ | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| | 0.00 | 0.17 | 0.33 | 0.50 | 0.67 | 0.83 | 1.00 |
| Mass | | | | | | | |
| λ | 18 | 19 | 20 | 21 | 22 | 23 | |
| | 0.00 | 0.20 | 0.40 | 0.60 | 0.80 | 1.00 | |

used by the CHARMM Bilayer Builder.

The actual free energy calculation was done for 1 ns, with a timestep of 0.02 ps. For temperature and pressure coupling, the same settings as in the NpT equilibration step were used.

4. Results

The objective of this thesis was to redesign an existing molecule to improve its selective binding affinity towards CL. This is too computationally expensive to be accomplished directly on the atomistic level. Therefore, a preliminary coarse-graining approach was taken to screen a vast number of possible interactions. The best results were transferred to the atomistic level by adding reactive groups to the original NAO molecular structure. The selected reactive groups had the same properties that had improved the binding affinity on the coarse grained level.

4.1. Coarse-Grained Resolution

The coarse-grained representation of the original NAO molecule was the starting point of the simulations, it was automatically parameterized for the Martini force field using auto-martini [85]. The three rings in the molecular structure of NAO initially led to an exaggerated hydrophobicity of the coarse-grained molecule, which is a general problem with the auto-martini scheme, see [85], figure 7. Therefore, four beads belonging to the ring system were removed to bring the hydrophobicity closer to atomistic data. The bead types of the original parameterization are shown in figure 3.5 (A).

4.1.1. Alchemical Transformations

Relative binding free energies of trial molecules in different lipid membranes were calculated, starting from the original NAO molecule. The relative binding free energy of a molecule inserted into different environments shows whether it would bind more stable to one target or another.

Relative binding free energies are obtained in several steps, as illustrated in figure 4.1. First, as shown in steps 1a) and 1b), the free-energy difference between the two end states of a transformed molecule are calculated both in bulk water and the membrane environment. This is repeated in steps 2a) and 2b) for the same states A and B of the molecule, but in a different simulation box containing a membrane that consists of a different lipid.

The two free-energy values from each system are then used to calculate the transfer free

4. Results

energy, as shown in step 3). The transfer free energy is a double free-energy difference, calculated from the difference between the free energy differences of the molecule transformed in the membrane and the same molecule transformed in bulk water. Again, this is done for both membrane systems, as illustrated by steps 3a) and 3b).

The resulting free energies of transferring the molecule into different membranes are used to determine the relative binding free energy between the molecule and the two compared lipids in step 4).

The results of steps 3a) and 3b) are plotted against each other in figures 4.2 and 4.3. The position of each data point relative to the diagonal black helper line shows the relative binding affinity of the corresponding molecule to the compared lipids. If the data point is above the line, it has a higher binding affinity toward the lipid plotted on the x axis, if it is below the diagonal line, its binding affinity towards the lipid plotted on the y axis is higher. So the data point of a molecule that has a high binding affinity towards the target lipid CL, in this work always plotted on the x axis, has to be placed as close to the point of origin at the x axis and as high on the y axis as possible.

An obvious preference of NAO towards binding to the anionic phospholipids CL and PG compared to PC is clearly visible in all simulations. All the free-energy values given are free-energy differences $\Delta\Delta G$ between the membrane-water interface region and bulk water, as explained in section 3.3.1. The thermodynamic cycle used in this thesis is shown in figure 3.3.

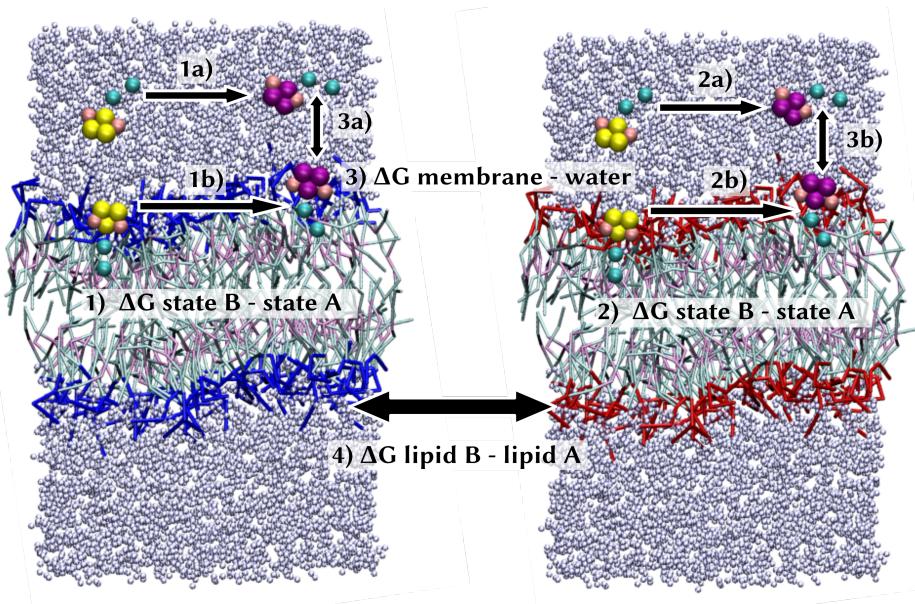


Figure 4.1.: The calculation of relative binding free energies follows several steps. In steps 1) and 2), the free energies of transforming a molecule from state A (illustrated with yellow ring beads), to state B (purple ring beads), are calculated both in bulk water and in the interface region of different membranes (headgroups colored blue and red). In step 3), the transfer free energy between the molecules in the membrane and in bulk water are calculated. Finally, step 4) yields the relative binding free energy between the molecules in different environments.

4.1. Coarse-Grained Resolution

Transformation of the Top Beads of the Aromatic Ring System

The only difference between the molecular structure of CL and PG is between the headgroups, as is clearly visible comparing the structural formulae shown in figures 1.3 and 1.4. From the positioning of NAO in the interface region of the membrane, the top of the molecule was identified as the starting point for the optimization.

The two top beads of the aromatic ring system were transformed according to their hydrogen bonding capacities and the side beads were iterated over C bead types that, according to the Martini interaction matrix [52], interact more favorably with the CL headgroup bead (GL0) compared to the PG headgroup bead (P4), see figure 3.5 (B).

If only the polarity of the beads on the sides of the NAO ring system, representing the two amine groups, was changed, the overall relative binding affinity of NAO was only slightly higher towards CL than towards PG or PC (table 4.1, figure 4.2, green triangles).

Transforming the beads on the top center of the NAO molecule (N0) to other nonpolar types showed a decrease in preference for CL for hydrogen bond acceptor beads (Na), as visible in figure 4.2, red and blue triangles.

The most positive effect on the relative binding affinity was found with beads with both hydrogen bond donor and acceptor qualities (Nda), followed by pure donor bead (Nd), figure 4.2, purple triangles and orange diamonds. The beads on the sides influenced the free energy in the desired way if left as the weakest apolar type (C5) from the original NAO parameterization.

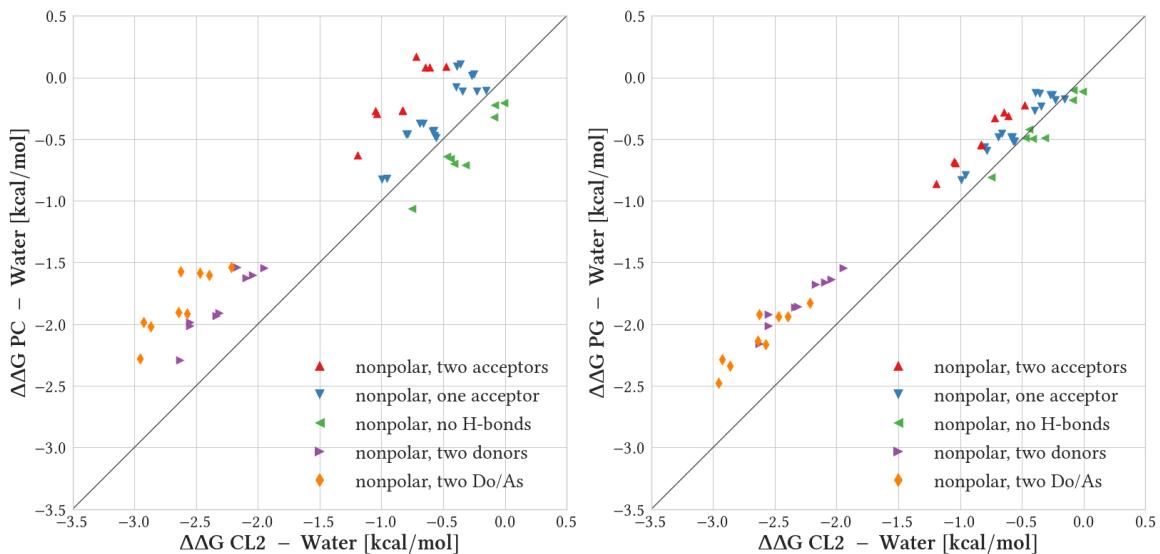


Figure 4.2.: Double free energy differences between the original NAO molecule and different end states with transformed top beads, inserted into the different membranes. The changes in free energy are plotted against each other, PC is compared to CL on the left and PG to CL on the right. Do/A refers to donor-acceptor here and further on in the work.

4. Results

Table 4.1.: The differences in free energy change (in kcal/mol) between PG and CL compared to PC and CL. The beads on top of the NAO aromatic ring system were transformed. By way of example, only the free-energy values for the end states with the same kind of C bead on the sides of the NAO molecule are shown. The free-energy differences of end states with mixed outer bead types lie in between the corresponding same-bead molecules, refer to table A.1.

| Apolarity | $\Delta\Delta G_{CL} - \Delta\Delta G_{PC}$ | | | $\Delta\Delta G_{CL} - \Delta\Delta G_{PG}$ | | |
|-------------------------|---|--------------|--------|---|--------------|--------|
| | strong | intermediate | weak | strong | intermediate | weak |
| Hydrogen bonding | | | | | | |
| two acceptors | -0.567 | -0.567 | -0.889 | -0.331 | -0.256 | -0.391 |
| one acceptor | -0.166 | -0.118 | -0.483 | -0.163 | -0.050 | -0.270 |
| no bonding | 0.309 | 0.230 | 0.197 | 0.055 | 0.088 | 0.103 |
| two donors | -0.341 | -0.405 | -0.633 | -0.470 | -0.403 | -0.492 |
| two donor-acceptors | -0.678 | -0.678 | -1.053 | -0.479 | -0.385 | -0.706 |

Transformation of the Side Beads of the Aromatic Ring System

Another, smaller structural difference between CL and PG is that CL has four fatty acid chains and two phosphate groups, twice the number of PG. Therefore, the sides of the NAO molecule were transformed to add possible interactions to each of the two phosphate groups of CL.

Starting from the results of the first alchemical transformations simulations, two nonpolar donor-acceptor beads (Nda) were selected for the top center of the NAO ring system. Because the phosphate groups of CL can act as hydrogen bond acceptors and the hydroxy group of its glycerol bridge can be both a donor and an acceptor, the side beads of NAO were then changed to nonpolar beads with different hydrogen bonding capacities (figure 3.5 (C)). As shown in figure 4.3 and table 4.2, further decreasing the apolarity of the side beads from the original C5, but not adding any hydrogen bonding capacity (bead type N0, yellow pentagon) led to a negative result compared to the previous C bead types at those positions (purple triangles, orange diamonds).

When hydrogen bonding capacities are introduced, there is an overall trend visible of an increase in binding affinity towards CL from adding an acceptor (Na) bead (red triangles) over a donor-acceptor (Nda) bead (blue triangles), to a hydrogen bond donor (green triangles), always referring to the same part of the molecule. The binding affinity of those NAO varieties to PC compared to CL decreases significantly.

The biggest difference in free energy change among the tested bead combinations between PG and CL is found with the combination of two donor-acceptor beads at the top of the

4.1. Coarse-Grained Resolution

aromatic system and two donor beads at the sides. In this case, the negative change in free energy in the CL membrane is bigger than the one in the PG membrane by -1.101 kcal/mol , see table 4.2.

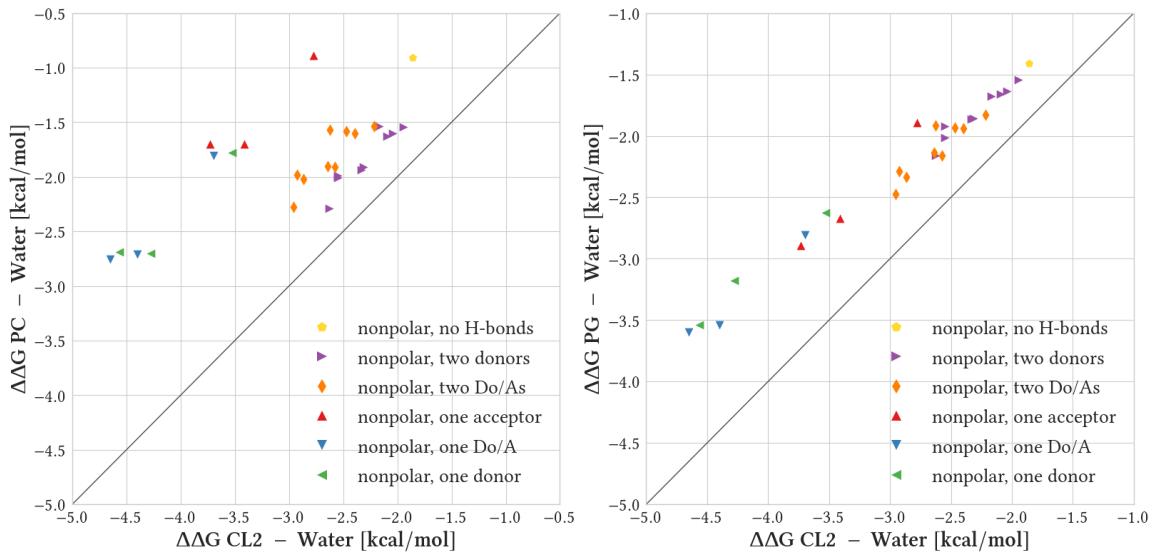


Figure 4.3.: Double free energy differences between the original NAO molecule and different end states with transformed side beads, inserted into the different membranes. The molecules with two hydrogen bond donors (purple triangles) and hydrogen bond donor-acceptors (orange diamonds) are the same as in figure 4.2 and have been kept on as a reference.

Table 4.2.: The differences in free energy change (kcal/mol) between PG and CL compared to PC and CL. The rows represent the amine bead on one side of the NAO molecule, the columns show the bead type of the other side. Both top center beads are Nda.

| Hydrogen bonding | $\Delta\Delta G_{\text{CL}} - \Delta\Delta G_{\text{PC}}$ | | | $\Delta\Delta G_{\text{CL}} - \Delta\Delta G_{\text{PG}}$ | | |
|-------------------------|---|--------|--------|---|--------|--------|
| | acceptor | donor | Do/A | acceptor | donor | Do/A |
| Hydrogen bonding | | | | | | |
| acceptor | -1.886 | -1.710 | -2.035 | -0.887 | -0.740 | -0.840 |
| Do/A | -1.896 | -1.694 | -1.897 | -0.893 | -0.859 | -1.059 |
| donor | -1.750 | -1.574 | -1.875 | -0.900 | -1.101 | -1.024 |
| no bonding | | | | | | |
| no bonding | | | -0.953 | | | -0.453 |

4. Results

4.1.2. Potentials of Mean Force

The transformed molecules which showed the greatest negative change in free energy in the alchemical transformations were used to reconstruct complete PMFs across the interface between water and membrane.

Binding Affinities to PC and PG compared to CL

The PMF curves calculated from umbrella sampling show three areas of interest, the membrane midplane (M), the interface region (I) between the lipid headgroups and water, and the bulk water (W) phase. The corresponding free energy values are given in table A.2 in the appendix.

The maximum energies at the membrane midplane ($z = 0$) show a significant free energy barrier for NAO and the molecules developed from it to cross the hydrophobic center of the membrane bilayer, as shown in table 4.3. There is also a considerable free energy difference between the minimal energy wells at the interface region and the bulk water, it is more favorable for the small molecules to align with the lipid head groups than to remain in an aqueous environment.

The biggest energy barriers are found with the molecule having two donor-acceptor beads (Nda) at the top and another two at the sides, followed by the one with Nda beads at the top and donor (Nd) beads at the sides.

The PMF curves in figures 4.4 and 4.5 visualize the bigger preference of NAO and its derivatives towards CL and PG compared to PC. The depth of the minimal energy wells grow progressively deeper from the original NAO molecule over the variants with hydrogen bonding capacity at the top to the ones with both hydrogen bonding capacity at the top and the sides. The colors refer to:

- blue** The original NAO molecule.
- green** Two Nd beads (adding hydrogen bond donor capacities) at the top and two weakly apolar C5 beads at the sides, among the purple triangles in figures 4.2 and 4.3.
- red** Two Nda beads with both hydrogen bond donor and acceptor properties (referred to as “Do/A” in the plots) at the top and two weakly apolar C5 beads at the sides, member of the orange triangle group in figures 4.2 and 4.3.
- orange** Two Nda beads at the top and two Nda beads at the sides, adding hydrogen bond donor-acceptor properties to all adapted parts of the molecule. This transformed molecule is part of the blue triangles in figure 4.3.
- purple** Two Nda beads at the top and two Nd beads at the sides, with green triangles in figure 4.3.

While the improvement of the binding affinities of the different NAO varieties along the performed optimization steps is generally true for both the insertion into the PG and CL

4.1. Coarse-Grained Resolution

membranes, there is one exception. The molecule with two donor-acceptor beads at the top of the ring system and two donor beads at the sides (purple curves) shows a visibly lower binding affinity to PG than the one with donor-acceptor beads both at the top and the sides (orange curve). The minimal energy well is less deep than with the former molecule in the PG membrane-water interface region, whereas it stays almost the same with CL. For the free energy values, refer to tables A.2 and 4.3.

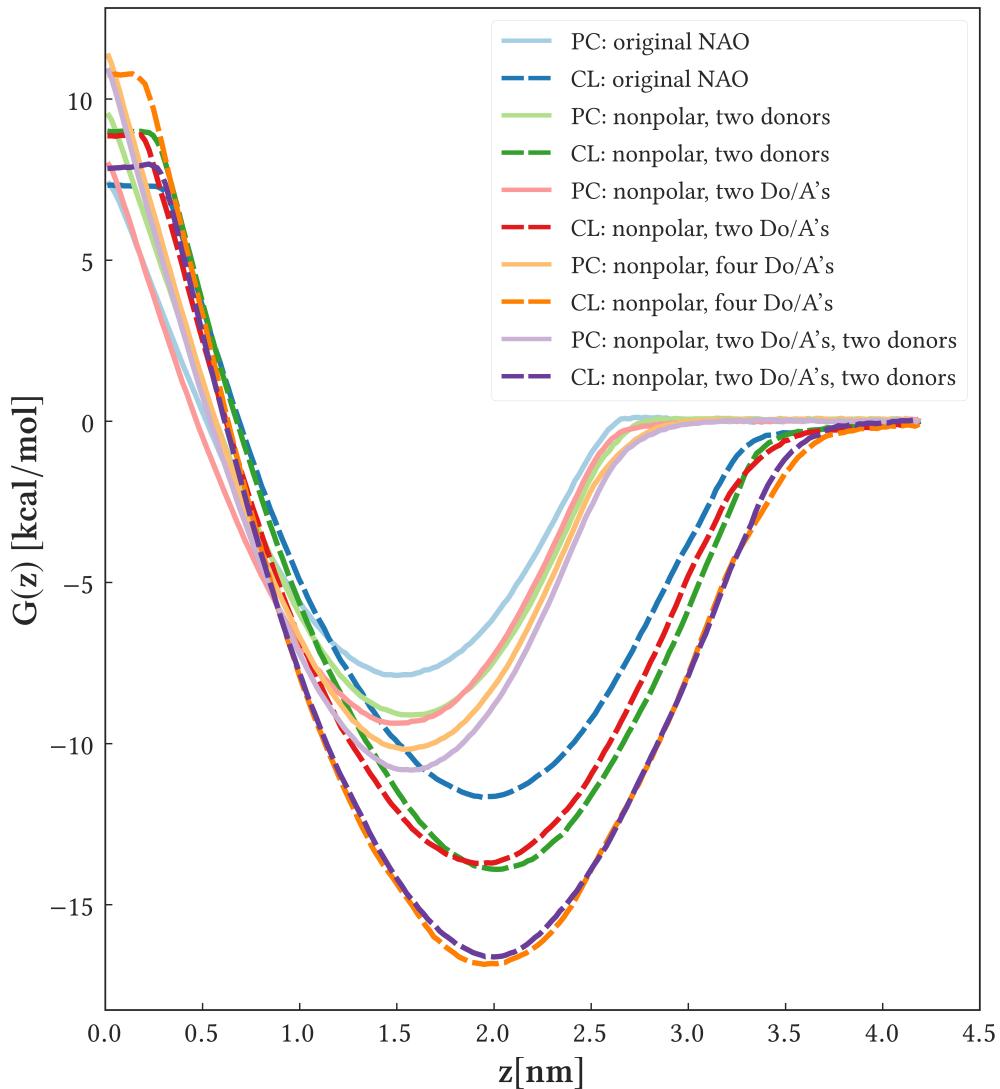


Figure 4.4.: The PMF curves of the different end state varieties of the alchemical transformation simulations. The pastel and vivid tones of each color represent the same molecule in PC and CL membranes respectively.

4. Results

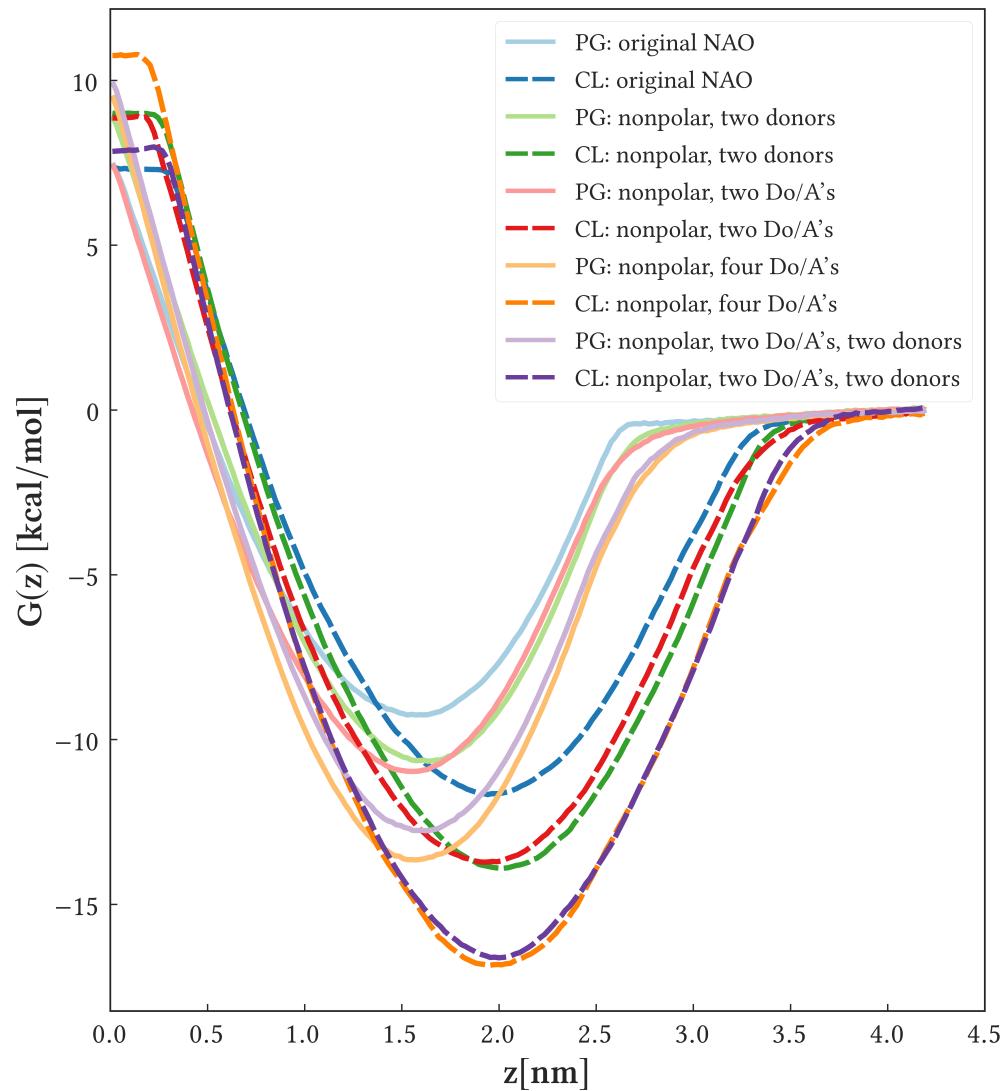


Figure 4.5.: The PMF curves of the different end state varieties of the alchemical transformation simulations in PG (pastel) and CL (vivid).

4.2. Atomistic Resolution

Table 4.3.: The free energy barriers between the different membrane regions, the bilayer midplane (M), the interface region (I) and the water phase (W), show a clear preference of the NAO difference to position themselves at the membrane-water interface, and a bigger binding affinity to CL than to the other two lipids. For the respective positions of the beads on the coarse grained NAO molecule, refer to figure 3.5

| | molecule | $\Delta G_{I - M}$ [kcal/mol] | $\Delta G_{I - W}$ [kcal/mol] |
|-----------|-------------------------------|-------------------------------|-------------------------------|
| PC | original NAO | -15.32 | -7.88 |
| | C1-C1-SQ0-SNd-SNd-SC5-SC5 | -18.68 | -9.11 |
| | C1-C1-SQ0-SNda-SNda-SC5-SC5 | -17.40 | -9.37 |
| | C1-C1-SQ0-SNda-SNda-SNda-SNda | -21.59 | -10.18 |
| PG | C1-C1-SQ0-SNda-SNda-SNd-SNd | -11.92 | -10.83 |
| | original NAO | -16.75 | -9.26 |
| | C1-C1-SQ0-SNd-SNd-SC5-SC5 | -19.57 | -10.66 |
| | C1-C1-SQ0-SNda-SNda-SC5-SC5 | -18.42 | -10.97 |
| CL | C1-C1-SQ0-SNda-SNda-SNda-SNda | -23.20 | -13.65 |
| | C1-C1-SQ0-SNda-SNda-SNd-SNd | -22.72 | -12.77 |
| | original NAO | -18.99 | -11.67 |
| | C1-C1-SQ0-SNd-SNd-SC5-SC5 | -22.91 | -13.90 |
| | C1-C1-SQ0-SNda-SNda-SC5-SC5 | -22.58 | -13.72 |
| | C1-C1-SQ0-SNda-SNda-SNda-SNda | -27.60 | -16.85 |
| | C1-C1-SQ0-SNda-SNda-SNd-SNd | -24.47 | -16.62 |

4.2. Atomistic Resolution

The reduction of the degrees of freedom in the coarse-graining process means that the results from the coarse-grained simulations do not have the information depth of atomistic simulations. Nevertheless, they did give strong indications as to what interactions should be introduced into the NAO molecule. A molecule that has a highly selective binding affinity towards CL compared to PG should have reactive groups with both hydrogen bond donor and acceptor qualities at the top of the aromatic ring system and hydrogen bond donors at the sides. Attempts have been made to transfer the results from the coarse-grained simulations to the atomistic resolution.

4. Results

4.2.1. Alchemical Transformations

Proceeding from the coarse-grained results, the same first step of the alchemical transformations calculations was performed with atomistic systems.

Transformed Molecules

Since the most promising candidates for an improved binding affinity towards CL had hydrogen bond donor-acceptor beads on the top of the aromatic ring system, functional groups with donor-acceptor properties were added to the atomistic structure of NAO. Coarse-grained simulations give no information as to where exactly a reactive group should be positioned on atomistic level, therefore, two different donor-acceptor groups were placed at two different positions.

The hydroxy group added in figure 4.6 (A) can act as hydrogen bond acceptor via the free electron pair on the oxygen atom as well as as hydrogen bond donor on the hydrogen atom. The same mechanism applies for the primary amines added to the molecule in figure 4.6 (B).

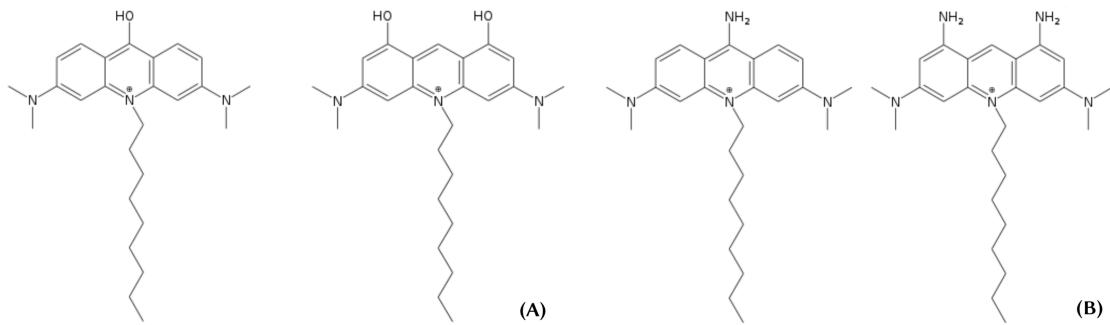


Figure 4.6.: Two different reactive group with hydrogen bond donor-acceptor qualities were added to the original NAO structure.

The atomistic alchemical transformations calculations did show an increase in binding affinity towards both membranes with state B. However, the binding affinity of all four NAO varieties increased stronger towards PG than towards CL, see table A.3.

Positions of the modified NAO in the Membranes

To investigate the influence of the added hydrogen bond donor-acceptor groups, position density plots were created. The number of atoms per \AA^3 shows in which area of the membrane selected groups of atoms were found with the highest frequency. As an example, the NAO molecule with one added hydroxy group illustrated on the left of figure 4.6 (A) is shown below.

4.2. Atomistic Resolution

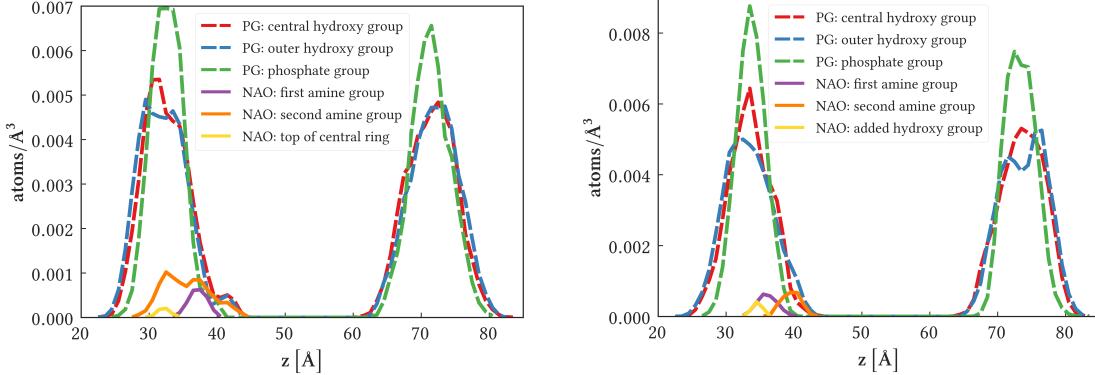


Figure 4.7.: The positions of the selected parts of the NAO molecule show it's preferred position in the membrane. The left plot shows the alignment of the original molecule in the membrane, the right plot the molecule with added hydrogen.

In the plot on the left side of figure 4.7, it is apparent that the different parts of the PG headgroups represented by the red, green and blue lines align themselves in a horizontal plain along the membrane water interface. The atoms of the NAO amine groups shown in orange and yellow and the atoms at the top of the NAO ring system represented by the purple line are found with the highest frequency horizontally aligned to the PG headgroup atoms and slightly shifted down towards the membrane midplane. The added hydroxy group does not significantly change the positioning of the molecule in the PG membrane, as visible in the plot on the right side of figure 4.7.

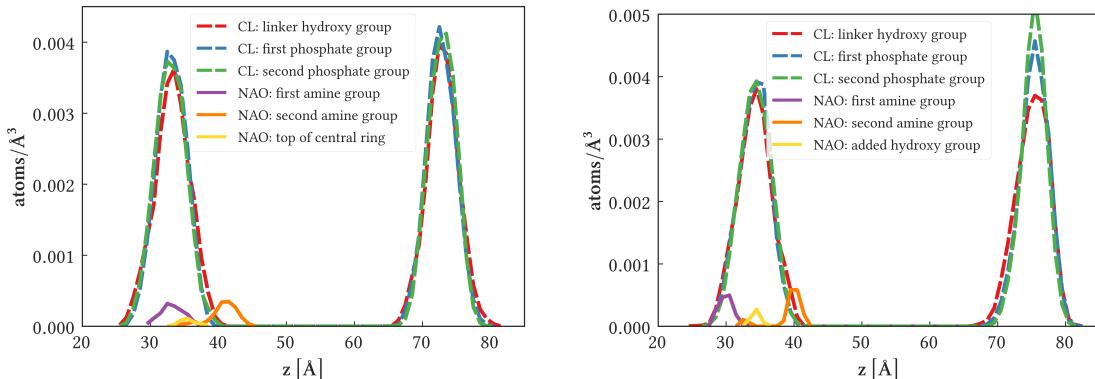


Figure 4.8.: The left plot shows the alignment of the original molecule in the membrane, the right plot shows the NAO molecule with added hydroxy group.

The headgroup atoms of CL are aligned in a plane along the membrane-water interface, the phosphate groups of CL are shown in blue and green, the bridging hydroxy group is

4. Results

represented by the red line. The added hydroxy group on the top of the NAO ring system is shown in purple, and the amine groups are again represented by the orange and yellow lines. The most frequently taken positions of the amine group and ring system atoms of the original NAO molecule show that it takes up a slightly titled position relative to the CL headgroup atoms, as visible in the left part of figure 4.8.

The right part of figure 4.8 shows that, with the added hydroxy group, this gets even more pronounced. There, the NAO molecule arranges itself so that the aromatic ring system is orthogonal to the headgroup plane and pulled up and away from the bilayer midplane and closer to the bulk water.

The positions of the other modified NAO molecules showed similar changes, as can be seen in the corresponding plots A.1, A.2 and A.3 in the appendix.

The alignment of the original and the modified NAO molecules are shown in figures 4.9, 4.10 and 4.11. The NAO molecule is colored blue, and the headgroups of the lipids are shown in red and amber. The lipid tails are represented by gray-green lines. The black lines indicate the course of the membrane-water interface region and the lower edge of the aromatic ring system of the NAO molecule versions.

As examples, the original NAO molecule, the molecule with one added hydroxy group and the version with two added amine groups are shown.

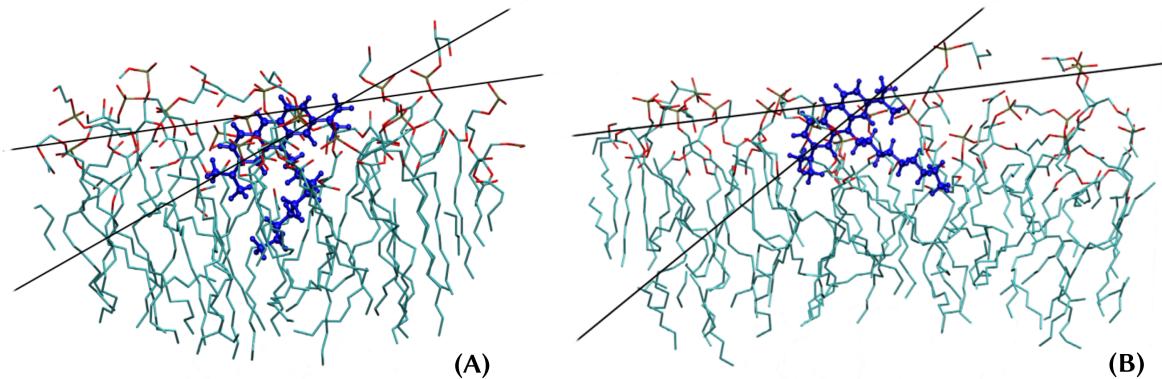


Figure 4.9.: Alignment of the original NAO molecule in: (A) A PG membrane and (B) A CL membrane. the black lines indicate the membrane interface region and the ring system of NAO

The original NAO molecule is inserted into the membrane slightly below the interface region, and it aligns itself almost parallel to the interface plane, as visible in figure 4.9. The hydrocarbon chain is inserted in the hydrophobic environment generated by the fatty acid chains of both phospholipids. The angle between the ring system of the NAO molecule and the headgroups seems to be slightly bigger with the CL membrane than with the PG membrane.

4.2. Atomistic Resolution

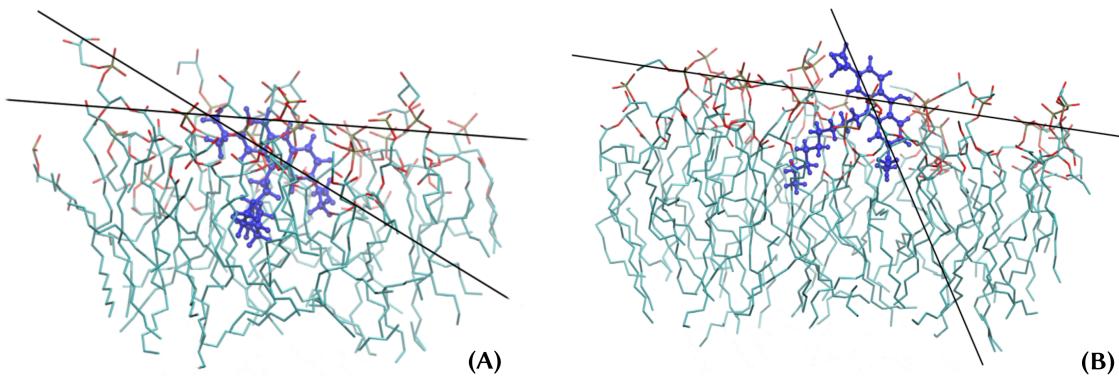


Figure 4.10.: Angles between the ring system of the NAO molecule with an added hydroxy group and the interface regions of the PG membrane in (A) and the interface region of the CL membrane in (B).

Adding one hydroxy group causes no or very little change in the alignment of the NAO ring system into the PG membrane, see figure 4.10 (A). In the CL membrane, the NAO molecule with the hydroxy group stays higher up and closer to the interface region, shown in figure 4.10 (B). The angle between the ring system of this NAO version and the lipid headgroup plane is visibly bigger with the CL membrane than with the PG membrane.

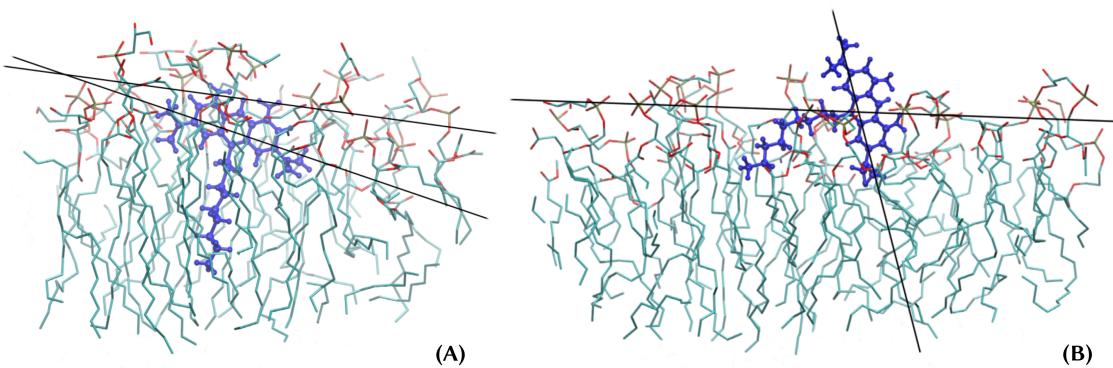


Figure 4.11.: Insertion of the NAO molecule with two added amine groups into the PG membrane in figure (A) and into the CL membrane in figure (B).

The introduction of two amine groups to the top of the NAO molecule shows next to no influence on the insertion into the PG membrane, compare figure 4.11 (A) to figures 4.9 (A) and 4.10 (A). But with the CL membrane, the modified NAO molecule aligns itself almost vertically to the interface of the membrane. The nonyl chain is pulled almost completely out of the membrane and into the polar headgroup region, a configuration that is energetically less favorable than the insertion of the nonyl chain into the hydrophobic region of the lipid tails.

5. Discussion

The data from the coarse-grained simulations suggested that hydrogen bonding was a starting point to improve the binding affinity of the NAO molecule towards CL. From the alchemical transformations, it was apparent that the molecules with two Nda beads at the top and two Nd or two Nda beads at the sides showed the biggest preference towards the headgroups of the two anionic phospholipids and had the highest binding affinity to CL compared to the other two lipid types. This was confirmed on the coarse-grained level with the umbrella sampling simulations.

Because CL has one hydroxy group at the bridge in the headgroup and PG has two hydroxy groups in the headgroup, adding hydrogen bonding properties to NAO increases its binding affinity towards both lipids.

The increasing free energy differences that become apparent if one side bead is fixed to an acceptor, donor or donor-acceptor type indicates that the angle at which the aromatic ring system of the NAO molecule aligns itself relative to the headgroup influences the binding affinity. Since the molecule is symmetrical along the vertical axis, there likely is some mechanism that fixes the position of NAO relative to the lipid headgroups. This may be due to the electrostatic interactions between the quarternary ammonium of NAO, represented by the Q0 bead in the coarse-grained representation, and the ionized phosphate groups of CL. There is a clear positive effect of placing two donor beads in the positions of the NAO amine groups to the binding affinity towards all three lipids. This is likely caused by the formation of hydrogen bonds to the phosphate groups of the lipids.

The first modification step of the coarse-grained alchemical transformations was repeated on the atomistic level, but yielded opposite results. Adding functional groups with hydrogen bond donor-acceptor properties to the top of the aromatic ring system of NAO increased its binding affinity more towards PG than towards CL. Also, the placement of the donor-acceptor groups had a clear adverse effect on the alignment of NAO with the targeted CL head group. The reason for this may be that the ideal placement for the donor-acceptor groups is not on the top of the NAO ring system, but more to the sides and lower towards the nonyl chain.

5.1. Conclusion

Up to this point in time, the benefit of coarse graining as a preliminary filtering step for virtual high throughput screening has not been thoroughly investigated. In this work, this

5. Discussion

has been attempted for the case of a suitable model system. Promising candidates for the optimization of a molecule towards increasing its specificity for a target lipid were preselected from coarse-grained simulations. Those optimizations were subsequently evaluated on the atomistic level.

In the atomistic simulations that are part of this thesis, it was not possible to recreate the coarse-grained results. The significant reduction of chemical space by the coarse-graining process means that the results are averaged over a large number of possible molecules and interaction types. This means that coarse graining can only give indications of promising interaction types in a general area of a molecule. The transfer of the changes made to the NAO molecule on the coarse-grained level did not yield the same results, but it did considerably change the stability of NAO in the molecule. This showed that an effect found in coarse-grained simulations has a counterpart on the atomistic level, even if it may not have exactly the same consequences.

Further calculations will be needed to successfully map the results from coarse-grained simulations to an atomistic representation of a new molecule.

Apart from adapting the positions of the donor-acceptor groups, another possible starting point for ongoing research is for example the removal of the tertiary amines at the sides of the NAO molecule in exchange for ammonium ions (-NH₃). This will greatly increase the capability for electrostatic interactions of the transformed NAO molecule with the phosphate groups of the lipids. It will also improve the hydrogen bonding capability of this part of the NAO molecule. Since there is no hydrogen covalently bound to the nitrogen atom, hydrogen bonding with tertiary amines is impossible. Moreover, the actual influence of the quarternary ammonium has to be evaluated with respect to its suspected influence on the positioning of the NAO molecule in relation to the lipid headgroups.

Maybe the most important constraint for the further optimization of NAO is the preservation of the symmetry of the molecule. Because there is next to no chemical difference between CL and PG, the capability of the molecule to bond to both the hydroxy group as well as the phosphate groups of CL is most important. The difference in free energy between the binding to CL and PG will likely not become big enough to provide sufficient selectivity for CL. The interaction of NAO with one CL molecule causes a lower loss of entropy than NAO hydrogen bonding to two PG molecules, so this factor has to be retained.

Further work has to be done to increase the efficiency of preselecting coarse-grained candidates and relating them to actual, possibly unknown, atomistic molecular structures. This may also provide information about possible common chemical properties of related reactive groups or molecular conformations.

Overall, it has been shown that valuable insight can be gained from coarse-grained free energy calculations to provide a clear direction for the ongoing work. It has allowed for efficiently covering a large chemical space in a short period of time and provided clear information as to the required chemical properties of a specific probe for Cardiolipin. Therefore, implementing coarse graining as a preliminary step of high throughput screening has proved valuable for speeding up the process and increasing the efficiency of discovering new molecules with highly specific properties.

Appendix

Appendix A.

Additional Data

A.1. Free energies of the coarse-grained Alchemical Transformations

Table A.1.: The free energy differences $\Delta\Delta G$ from the coarse-grained alchemical transformations simulations.

| Molecules | ΔG PC - Water | ΔG PG - Water | ΔG CL - Water |
|---------------------------|-----------------------|-----------------------|-----------------------|
| C1-C1-SQ0-SN0-SN0-SC1-SC1 | -1.061 | -0.807 | -0.752 |
| C1-C1-SQ0-SN0-SN0-SC1-SC2 | -0.712 | -0.488 | -0.319 |
| C1-C1-SQ0-SN0-SN0-SC1-SC5 | -0.655 | -0.418 | -0.445 |
| C1-C1-SQ0-SN0-SN0-SC2-SC1 | -0.698 | -0.495 | -0.414 |
| C1-C1-SQ0-SN0-SN0-SC2-SC2 | -0.322 | -0.18 | -0.092 |
| C1-C1-SQ0-SN0-SN0-SC2-SC5 | -0.223 | -0.1 | -0.086 |
| C1-C1-SQ0-SN0-SN0-SC5-SC1 | -0.64 | -0.491 | -0.474 |
| C1-C1-SQ0-SN0-SN0-SC5-SC2 | -0.207 | -0.113 | -0.01 |
| C1-C1-SQ0-SN0-SNa-SC1-SC1 | -0.82 | -0.788 | -0.959 |
| C1-C1-SQ0-SN0-SNa-SC1-SC2 | -0.44 | -0.481 | -0.585 |
| C1-C1-SQ0-SN0-SNa-SC1-SC5 | -0.372 | -0.455 | -0.663 |
| C1-C1-SQ0-SN0-SNa-SC2-SC1 | -0.471 | -0.519 | -0.569 |
| C1-C1-SQ0-SN0-SNa-SC2-SC2 | -0.106 | -0.177 | -0.152 |
| C1-C1-SQ0-SN0-SNa-SC2-SC5 | 0.024 | -0.14 | -0.252 |
| C1-C1-SQ0-SN0-SNa-SC5-SC1 | -0.462 | -0.564 | -0.8 |
| C1-C1-SQ0-SN0-SNa-SC5-SC2 | -0.08 | -0.267 | -0.395 |
| C1-C1-SQ0-SN0-SNa-SC5-SC5 | 0.106 | -0.129 | -0.359 |
| C1-C1-SQ0-SNa-SN0-SC1-SC1 | -0.828 | -0.831 | -0.994 |
| C1-C1-SQ0-SNa-SN0-SC1-SC2 | -0.491 | -0.516 | -0.56 |
| C1-C1-SQ0-SNa-SN0-SC1-SC5 | -0.459 | -0.595 | -0.786 |
| C1-C1-SQ0-SNa-SN0-SC2-SC1 | -0.433 | -0.486 | -0.577 |

Continued on next page

Appendix A. Additional Data

Table A.1 – continued from previous page

| Molecules | ΔG PC - Water | ΔG PG - Water | ΔG CL - Water |
|------------------------------|-----------------------|-----------------------|-----------------------|
| C1-C1-SQ0-SNa-SN0-SC2-SC2 | -0.112 | -0.18 | -0.23 |
| C1-C1-SQ0-SNa-SN0-SC2-SC5 | -0.111 | -0.234 | -0.345 |
| C1-C1-SQ0-SNa-SN0-SC5-SC1 | -0.376 | -0.482 | -0.689 |
| C1-C1-SQ0-SNa-SN0-SC5-SC2 | 0.017 | -0.139 | -0.268 |
| C1-C1-SQ0-SNa-SN0-SC5-SC5 | 0.091 | -0.122 | -0.392 |
| C1-C1-SQ0-SNa-SNa-SC1-SC1 | -0.626 | -0.862 | -1.193 |
| C1-C1-SQ0-SNa-SNa-SC1-SC2 | -0.266 | -0.548 | -0.831 |
| C1-C1-SQ0-SNa-SNa-SC1-SC5 | -0.29 | -0.691 | -1.038 |
| C1-C1-SQ0-SNa-SNa-SC2-SC1 | -0.271 | -0.541 | -0.828 |
| C1-C1-SQ0-SNa-SNa-SC2-SC2 | 0.09 | -0.221 | -0.477 |
| C1-C1-SQ0-SNa-SNa-SC2-SC5 | 0.086 | -0.28 | -0.647 |
| C1-C1-SQ0-SNa-SNa-SC5-SC1 | -0.267 | -0.681 | -1.05 |
| C1-C1-SQ0-SNa-SNa-SC5-SC2 | 0.083 | -0.307 | -0.61 |
| C1-C1-SQ0-SNa-SNa-SC5-SC5 | 0.169 | -0.329 | -0.72 |
| C1-C1-SQ0-SNd-SNd-SC1-SC1 | -2.288 | -2.159 | -2.629 |
| C1-C1-SQ0-SNd-SNd-SC1-SC2 | -1.907 | -1.859 | -2.315 |
| C1-C1-SQ0-SNd-SNd-SC1-SC5 | -2.011 | -1.922 | -2.551 |
| C1-C1-SQ0-SNd-SNd-SC2-SC1 | -1.934 | -1.861 | -2.335 |
| C1-C1-SQ0-SNd-SNd-SC2-SC2 | -1.542 | -1.544 | -1.947 |
| C1-C1-SQ0-SNd-SNd-SC2-SC5 | -1.627 | -1.662 | -2.095 |
| C1-C1-SQ0-SNd-SNd-SC5-SC1 | -1.986 | -2.015 | -2.552 |
| C1-C1-SQ0-SNd-SNd-SC5-SC2 | -1.602 | -1.634 | -2.042 |
| C1-C1-SQ0-SNd-SNd-SC5-SC5 | -1.535 | -1.676 | -2.168 |
| C1-C1-SQ0-SNda-SNda-SC1-SC1 | -2.278 | -2.477 | -2.956 |
| C1-C1-SQ0-SNda-SNda-SC1-SC2 | -1.902 | -2.134 | -2.64 |
| C1-C1-SQ0-SNda-SNda-SC1-SC5 | -2.02 | -2.337 | -2.865 |
| C1-C1-SQ0-SNda-SNda-SC2-SC1 | -1.912 | -2.162 | -2.575 |
| C1-C1-SQ0-SNda-SNda-SC2-SC2 | -1.536 | -1.829 | -2.214 |
| C1-C1-SQ0-SNda-SNda-SC2-SC5 | -1.602 | -1.939 | -2.396 |
| C1-C1-SQ0-SNda-SNda-SC5-SC1 | -1.982 | -2.286 | -2.924 |
| C1-C1-SQ0-SNda-SNda-SC5-SC2 | -1.582 | -1.935 | -2.47 |
| C1-C1-SQ0-SNda-SNda-SC5-SC5 | -1.571 | -1.918 | -2.624 |
| C1-C1-SQ0-SNda-SNda-SN0-SN0 | -0.909 | -1.409 | -1.862 |
| C1-C1-SQ0-SNda-SNda-SNa-SNa | -0.891 | -1.89 | -2.777 |
| C1-C1-SQ0-SNda-SNda-SNa-SNd | -1.703 | -2.673 | -3.413 |
| C1-C1-SQ0-SNda-SNda-SNa-SNda | -1.697 | -2.892 | -3.732 |
| C1-C1-SQ0-SNda-SNda-SNd-SNa | -1.776 | -2.626 | -3.526 |

Continued on next page

A.1. Free energies of the coarse-grained Alchemical Transformations

Table A.1 – continued from previous page

| Molecules | ΔG PC - Water | ΔG PG - Water | ΔG CL - Water |
|-------------------------------|-----------------------|-----------------------|-----------------------|
| C1-C1-SQ0-SNda-SNda-SNd-SNd | -2.704 | -3.177 | -4.278 |
| C1-C1-SQ0-SNda-SNda-SNd-SNda | -2.69 | -3.541 | -4.565 |
| C1-C1-SQ0-SNda-SNda-SNda-SNa | -1.803 | -2.806 | -3.699 |
| C1-C1-SQ0-SNda-SNda-SNda-SNd | -2.706 | -3.541 | -4.4 |
| C1-C1-SQ0-SNda-SNda-SNda-SNda | -2.757 | -3.595 | -4.654 |

Appendix A. Additional Data

A.2. Relevant points of the PMF curves

Table A.2.: Regions of the PMF curves corresponding to the membrane midplane, membrane-water interface and bulk water region of the simulation box

| | molecule | Midplane | Interface | Water |
|-----------------------------|------------------|-----------------|------------------|--------------|
| original NAO | $z[nm]$ | 0.0 | 1.49 | 3.20 |
| | $G(z)[kcal/mol]$ | 7.43 | -7.88 | 0.0 |
| C1-C1-SQ0-SNd-SNd-SC5-SC5 | $z[nm]$ | 0.0 | 1.57 | 3.19 |
| | $G(z)[kcal/mol]$ | 9.56 | -9.11 | 0.0 |
| PC | $z[nm]$ | 0.0 | 1.47 | 3.19 |
| | $G(z)[kcal/mol]$ | 8.03 | -9.37 | 0.0 |
| C1-C1-SQ0-SNda-SNda-SC5-SC5 | $z[nm]$ | 0.0 | 1.53 | 3.21 |
| | $G(z)[kcal/mol]$ | 11.41 | -10.18 | 0.0 |
| C1-C1-SQ0-SNda-SNda-SNd-SNd | $z[nm]$ | 0.0 | 1.57 | 3.21 |
| | $G(z)[kcal/mol]$ | 10.95 | -10.83 | 0.0 |
| original NAO | $z[nm]$ | 0.0 | 1.53 | 3.50 |
| | $G(z)[kcal/mol]$ | 7.49 | -9.26 | 0.0 |
| C1-C1-SQ0-SNd-SNd-SC5-SC5 | $z[nm]$ | 0.0 | 1.63 | 3.50 |
| | $G(z)[kcal/mol]$ | 8.91 | -10.66 | 0.0 |
| PG | $z[nm]$ | 0.0 | 1.55 | 3.49 |
| | $G(z)[kcal/mol]$ | 7.46 | -10.97 | 0.0 |
| C1-C1-SQ0-SNda-SNda-SC5-SC5 | $z[nm]$ | 0.0 | 1.57 | 3.51 |
| | $G(z)[kcal/mol]$ | 9.54 | -13.65 | 0.0 |
| C1-C1-SQ0-SNda-SNda-SNd-SNd | $z[nm]$ | 0.0 | 1.62 | 3.50 |
| | $G(z)[kcal/mol]$ | 9.95 | -12.77 | 0.0 |
| CL | $z[nm]$ | 0.0 | 1.94 | 4.19 |
| | $G(z)[kcal/mol]$ | 7.32 | -11.67 | 0.0 |
| C1-C1-SQ0-SNd-SNd-SC5-SC5 | $z[nm]$ | 0.0 | 2.02 | 4.21 |
| | $G(z)[kcal/mol]$ | 9.00 | -13.90 | 0.0 |

Continued on next page

A.2. Relevant points of the PMF curves

Table A.2 – continued from previous page

| | molecule | Midplane | Interface | Water |
|----|-------------------------------|------------------|------------------|--------------|
| CL | C1-C1-SQ0-SNda-SNda-SC5-SC5 | $z[nm]$ | 0.0 | 1.92 |
| | | $G(z)[kcal/mol]$ | 8.86 | -13.72 |
| | C1-C1-SQ0-SNda-SNda-SNda-SNda | $z[nm]$ | 0.0 | 1.95 |
| | | $G(z)[kcal/mol]$ | 10.75 | -16.85 |
| | C1-C1-SQ0-SNda-SNda-SNd-SNd | $z[nm]$ | 0.0 | 2.00 |
| | | $G(z)[kcal/mol]$ | 7.84 | -16.62 |

Appendix A. Additional Data

A.3. Free energies from the atomistic Alchemical Transformations

Table A.3.: The free energy change of alchemical transformations simulations of the molecules shown in figure 4.6. The varieties shown in figure 4.6 (A) are referred to as NAO+OH and NAO+2OH, the molecules shown in 4.6 (B) are called NAO+NH₂ and NAO+2NH₂

| | $\Delta G_{PG} - \Delta G_{Water}$ | $\Delta G_{PG} - \Delta G_{Water}$ | $\Delta\Delta G_{CL} - \Delta\Delta G_{PG}$ |
|----------------------|------------------------------------|------------------------------------|---|
| NAO+OH | -1.236 | -0.662 | 0.574 |
| NAO+2OH | -1.421 | -1.392 | 0.029 |
| NAO+NH ₂ | -0.682 | -0.400 | 0.282 |
| NAO+2NH ₂ | -1.925 | -0.617 | 1.308 |

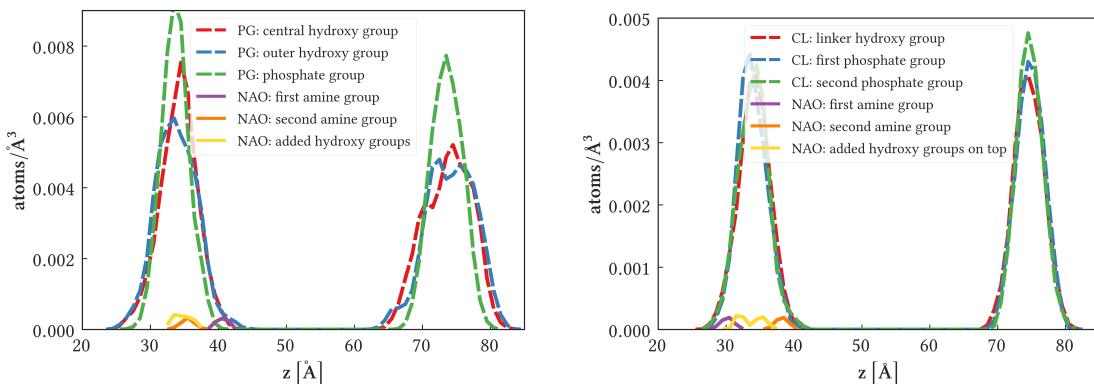


Figure A.1.: The two hydroxy groups added to the NAO at the top of the aromatic ring system, as illustrated on the right side of figure 4.6 (A), had the same effect of the positioning of the molecule in PG and CL membranes as adding only one hydroxy group (shown in figure 4.6 (A), left structure), see plots 4.7 and 4.8.

A.3. Free energies from the atomistic Alchemical Transformations

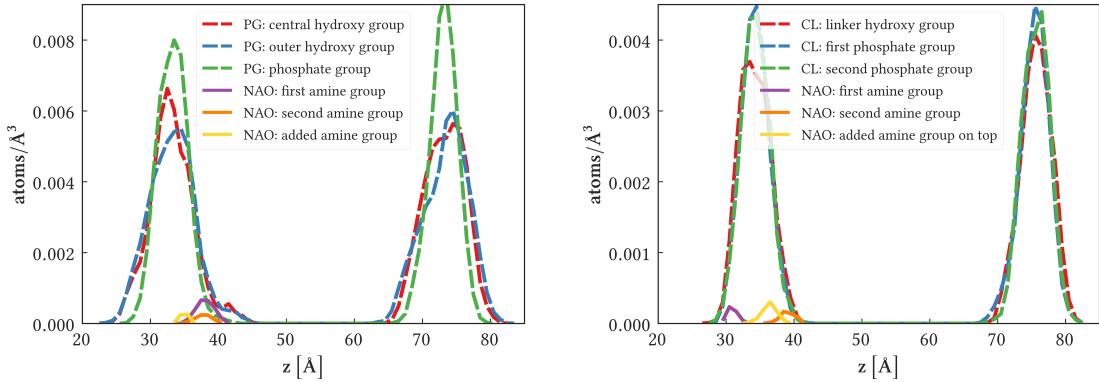


Figure A.2.: The addition of one amine group to the top of the molecular structure (as in figure 4.6 (B), left structure) showed the same effects on the positioning of NAO in the membrane as the added hydroxy group in figures 4.7 and 4.8.

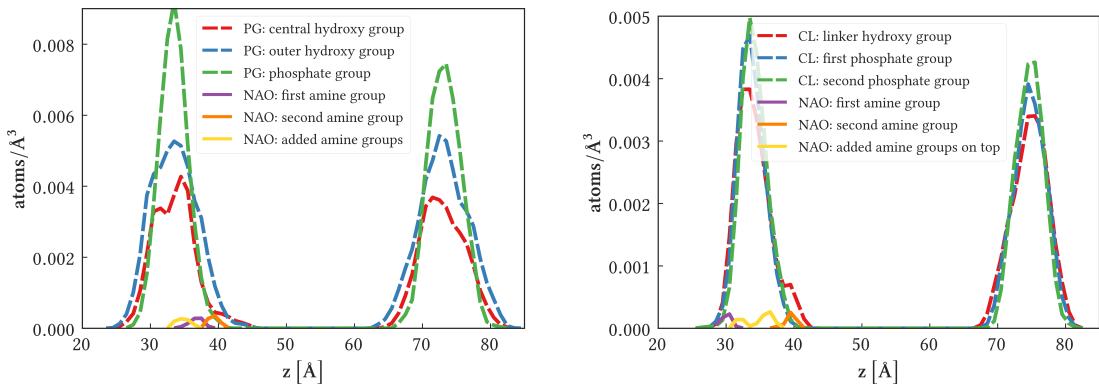


Figure A.3.: Adding two amine groups to the top of the rings of NAO (figure 4.6 (B) on the right) again led to their vertical tilting in the CL membrane.

Glossary

- k_B** The Boltzmann constant. A physical constant relating the average kinetic energy of particles in a gas with the temperature of the gas. It occurs in Boltzmann's entropy formula and in Planck's law of black-body radiation. It was introduced by Max Planck, but named after Ludwig Boltzmann. 12
- β** Thermodynamic beta. It expresses the response of entropy to an increase in energy. If an amount of energy is added to a system, then beta describes the amount by which the entropy of the system will increase. 12
- ATP** Adenosine Triphosphate. A coenzyme that transports chemical energy within cells for metabolism. 3, 4
- CG** Coarse-grained modeling, coarse-grained models, aim at simulating the behavior of complex systems using their coarse-grained (simplified) representation. 23, 71
- CHARMM** Chemistry at Harvard Macromolecular Mechanics, a widely used set of force fields for molecular dynamics. 22, 43
- CL** Cardiolipin. Lipid occurring exclusively in the inner mitochondrial membrane. ix, 5, 6, 8, 38, 39, 42, 45–57, 59, 60, 63–69
- Do/A** Abbreviation for donor-acceptor. A reactive group or CG bead that can act both as hydrogen bond donor and acceptor. 47, 49
- LJ** The Lennard-Jones potential is a mathematically simple model that approximates the interaction between a pair of neutral atoms or molecules. A form of this interatomic potential was first proposed in 1924 by John Lennard-Jones. 21, 29, 31
- Martini** A coarse-grained force field developed by Marrink and coworkers at the University of Groningen, initially developed for molecular dynamics simulation of lipids, extended to various other molecules. 23, 24, 39, 45
- MBAR** Multistate Bennett Acceptance Ratios. Evaluation method for molecular dynamics simulations, accesses data from multiple different states. 32, 37, 38, 40, 42
- MD** Molecular dynamics. Molecular dynamics (MD) is a computer simulation method for studying the physical movements of atoms and molecules. The atoms and molecules are allowed to interact for a fixed period of time, giving a view of the dynamic evolution of the system. ix, 19, 21, 23–26, 33, 40, 41

Glossary

- NAO** Nonyl acridine orange. Fluorescent dye used as a mitochondrial probe. ix, xi, 8, 38–42, 45–50, 53–57, 59, 60, 66, 68, 69
- NpT** Thermodynamic or statistical mechanical ensemble. Number of particles, temperature and pressure are fixed, volume and energy can fluctuate. 12, 13, 16–18, 28, 35, 43
- NVT** Thermodynamic or statistical mechanical ensemble. Number of particles, temperature and volume are fixed, energy is allowed to fluctuate. 42
- PC** Zwitterionic lipid. ix, 7, 8, 39, 46–51, 63–66
- PES** Potential energy surface. A PES describes the energy of a system, especially a collection of atoms, in terms of certain parameters, normally the positions of the atoms. 25
- PG** Phosphatidylglycerol. Lipid abundant in eukaryotic cells, precursor of Cardiolipin. ix, 6–8, 39, 42, 46–57, 59, 60, 63–66, 68
- PMF** Potential of Mean Force. The free energy surface along a reaction coordinate. Shows the free energy change as a function of this reaction coordinate. ix, 18, 19, 28, 34, 50
- WHAM** Weighted Histogram Analysis Method. A method to integrate information from multiple histograms. 36, 37, 41

Bibliography

- [1] John D Chodera et al. “Alchemical free energy methods for drug discovery: progress and challenges”. In: *Current opinion in structural biology* 21.2 (2011), pp. 150–160 (cit. on p. 1).
- [2] Jeremy M Berg, John L Tymoczko, and Lubert Stryer. “Lipids and cell membranes”. In: *Biochemistry* 5 (2002), pp. 319–344 (cit. on p. 2).
- [3] Bruce Alberts et al. “The Mitochondrion”. In: *Molecular Biology of the Cell*. Garland Science, 2002. Chap. 14 (cit. on p. 3).
- [4] Bruce Alberts et al. “Genetic Information in Eucaryotes”. In: *Molecular Biology of the Cell*. Garland Science, 2002. Chap. 1, Figure 1-34 (cit. on p. 3).
- [5] William J. Valente and Jason H. Bielas. “Mitochondrial Mutagenesis in Cancer”. In: *Mitochondria and Cell Death*. Ed. by David M. Hockenberry. Springer New York, 2016 (cit. on p. 4).
- [6] Guy A. Perkins and Mark H. Ellisman. “Remodeling of Mitochondria in Apoptosis”. In: *Mitochondria and Cell Death*. Ed. by David M. Hockenberry. Springer New York, 2016 (cit. on p. 4).
- [7] Bruce Alberts et al. “The lipid bilayer”. In: *Molecular Biology of the Cell*. Garland Science, 2002 (cit. on p. 4).
- [8] Wikimedia Commons. *File:Lipid bilayer and micelle.svg* — *Wikimedia Commons, the free media repository*. [Online; accessed 3-September-2018]. 2014. URL: %5Curl%7Bhttps://commons.wikimedia.org/w/index.php?title=File:Lipid_bilayer_and_micelle.svg&oldid=124880425%7D (cit. on p. 5).
- [9] William (Bill) W. Christie. *Cardiolipin (Diphosphatidylglycerol)*. URL: <http://www.lipidhome.co.uk/lipids/complex/dpg/index.htm> (visited on 09/03/2018) (cit. on p. 5).
- [10] EE Kooijman et al. “Magic angle spinning 31 P NMR spectroscopy reveals two essentially identical ionization states for the cardiolipin phosphates in phospholipid liposomes”. In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1859.1 (2017), pp. 61–68 (cit. on p. 5).
- [11] Dmitry Malyshka, Leah A Pandiscia, and Reinhard Schweitzer-Stenner. “Cardiolipin containing liposomes are fully ionized at physiological pH. An FT-IR study of phosphate group ionization”. In: *Vibrational Spectroscopy* 75 (2014), pp. 86–92 (cit. on p. 5).

Bibliography

- [12] Gerd Olofsson and Emma Sparr. “Ionization constants pKa of cardiolipin”. In: *PLoS one* 8.9 (2013), e73040 (cit. on p. 5).
- [13] Murugappan Sathappa and Nathan N Alder. “The ionization properties of cardiolipin and its variants in model bilayers”. In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1858.6 (2016), pp. 1362–1372 (cit. on p. 5).
- [14] Kevin J Boyd, Nathan N Alder, and Eric R May. “Molecular Dynamics Analysis of Cardiolipin and Monolysocardiolipin on Bilayer Properties”. In: *Biophysical journal* 114.9 (2018), pp. 2116–2127 (cit. on p. 5).
- [15] Ruthven NAH Lewis and Ronald N McElhaney. “The physicochemical properties of cardiolipin bilayers and cardiolipin-containing lipid membranes”. In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1788.10 (2009), pp. 2069–2079 (cit. on p. 5).
- [16] Susanne E Horvath and Günther Daum. “Lipids of mitochondria”. In: *Progress in lipid research* 52.4 (2013), pp. 590–614 (cit. on p. 5).
- [17] Giuseppe Paradies et al. “Functional role of cardiolipin in mitochondrial bioenergetics”. In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1837.4 (2014), pp. 408–417 (cit. on p. 5).
- [18] RH Houtkooper and FM Vaz. “Cardiolipin, the heart of mitochondrial metabolism”. In: *Cellular and Molecular Life Sciences* 65.16 (2008), pp. 2493–2506 (cit. on p. 5).
- [19] Alexander M van der Bliek, Qinfang Shen, and Sumihiro Kawajiri. “Mechanisms of mitochondrial fission and fusion”. In: *Cold Spring Harbor perspectives in biology* 5.6 (2013), a011072 (cit. on p. 5).
- [20] Nikita Ikon and Robert O Ryan. “Barth syndrome: connecting cardiolipin to cardiomyopathy”. In: *Lipids* 52.2 (2017), pp. 99–108 (cit. on p. 6).
- [21] Geon Ha Kim et al. “The role of oxidative stress in neurodegenerative diseases”. In: *Experimental neurobiology* 24.4 (2015), pp. 325–340 (cit. on p. 6).
- [22] Yuguang Shi. “Emerging roles of cardiolipin remodeling in mitochondrial dysfunction associated with diabetes, obesity, and cardiovascular diseases”. In: *Journal of biomedical research* 24.1 (2010), pp. 6–15 (cit. on p. 6).
- [23] Zheni Shen et al. “The role of cardiolipin in cardiovascular health”. In: *BioMed research international* 2015 (2015) (cit. on p. 6).
- [24] Quan He and Xianlin Han. “Cardiolipin remodeling in diabetic heart”. In: *Chemistry and physics of lipids* 179 (2014), pp. 75–81 (cit. on p. 6).
- [25] Giuseppe Paradies et al. “Oxidative stress, cardiolipin and mitochondrial dysfunction in nonalcoholic fatty liver disease”. In: *World journal of gastroenterology: WJG* 20.39 (2014), p. 14205 (cit. on p. 6).
- [26] Manfred Fobker et al. “Accumulation of cardiolipin and lysocardiolipin in fibroblasts from Tangier disease subjects”. In: *FEBS letters* 500.3 (2001), pp. 157–162 (cit. on p. 6).

Bibliography

- [27] William (Bill) W. Christie. *Phosphatidylglycerol and Related Lipids*. URL: <http://www.lipidhome.co.uk/lipids/complex/pg/index.htm> (visited on 09/03/2018) (cit. on p. 6).
- [28] Wei-Wei Chen et al. “Phosphatidylglycerol Incorporates into Cardiolipin to Improve Mitochondrial Activity and Inhibits Inflammation”. In: *Scientific reports* 8.1 (2018), p. 4919 (cit. on p. 6).
- [29] Lucia Pokorná et al. “Specific degradation of phosphatidylglycerol is necessary for proper mitochondrial morphology and function”. In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1857.1 (2016), pp. 34–45 (cit. on p. 6).
- [30] William (Bill) W. Christie. *Phosphatidylcholine and Related Lipids*. URL: <http://www.lipidhome.co.uk/lipids/complex/pc/index.htm> (visited on 09/03/2018) (cit. on p. 7).
- [31] Jean-Michel PETIT et al. “10N-nonyl acridine orange interacts with cardiolipin and allows the quantification of this phospholipid in isolated mitochondria”. In: *European journal of biochemistry* 209.1 (1992), pp. 267–273 (cit. on p. 7).
- [32] Gerhard Klebe. *Drug Design: Methodology, concepts, and mode-of-action*. Springer Publishing Company, Incorporated, 2013 (cit. on p. 8).
- [33] Myriam E Rodriguez et al. “Targeting of mitochondria by 10-N-alkyl acridine orange analogues: role of alkyl chain length in determining cellular uptake and localization”. In: *Mitochondrion* 8.3 (2008), pp. 237–246 (cit. on pp. 8, 38).
- [34] David Chandler. “Introduction to modern statistical mechanics”. In: *Introduction to Modern Statistical Mechanics*, by David Chandler, pp. 288. Foreword by David Chandler. Oxford University Press, Sep 1987. ISBN-10: 0195042778. ISBN-13: 9780195042771 (1987), p. 288 (cit. on pp. 9, 11, 12).
- [35] Matthew J Benacquista and Joseph D Romano. *Classical mechanics*. Springer, 2018 (cit. on pp. 9, 32).
- [36] Jesus M Sanz-Serna. “Symplectic integrators for Hamiltonian problems: an overview”. In: *Acta numerica* 1 (1992), pp. 243–286 (cit. on p. 9).
- [37] Terrell L Hill. *An introduction to statistical thermodynamics*. Courier Corporation, 1986 (cit. on pp. 11, 12).
- [38] Gabriele Raabe. *Molecular Simulation Studies on Thermophysical Properties*. Springer, 2017 (cit. on pp. 12, 21, 22, 27, 34, 39).
- [39] L-V_ Woodcock. “Isothermal molecular dynamics calculations for liquid salts”. In: *Chemical Physics Letters* 10.3 (1971), pp. 257–261 (cit. on p. 13).
- [40] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*. Vol. 1. Elsevier, 1992 (cit. on p. 14).
- [41] Herman JC Berendsen et al. “Molecular dynamics with coupling to an external bath”. In: *The Journal of chemical physics* 81.8 (1984), pp. 3684–3690 (cit. on p. 14).

Bibliography

- [42] M Parrinello and A Rahman. “Crystal structure and pair potentials: A molecular-dynamics study”. In: *Physical Review Letters* 45.14 (1980), p. 1196 (cit. on p. 15).
- [43] Michele Parrinello and Aneesur Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method”. In: *Journal of Applied physics* 52.12 (1981), pp. 7182–7190 (cit. on p. 15).
- [44] Michael P Allen and Dominic J Tildesley. *Computer simulation of liquids*. Oxford university press, 2017 (cit. on p. 15).
- [45] Christophe Chipot and Andrew Pohorille. *Free energy calculations - Theory and applications in Chemistry and Biology*. Springer, 2007 (cit. on pp. 18, 32, 34).
- [46] A. D. MacKerell et al. “All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins”. In: *The Journal of Physical Chemistry B* 102.18 (1998), pp. 3586–3616 (cit. on p. 22).
- [47] Jeffery B Klauda et al. “Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types”. In: *The journal of physical chemistry B* 114.23 (2010), pp. 7830–7843 (cit. on p. 22).
- [48] Alexander D. Mackerell, Michael Feig, and Charles L. Brooks. “Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations”. In: *Journal of Computational Chemistry* 25.11 (2004), pp. 1400–1415 (cit. on p. 22).
- [49] Helgi I Ingólfsson et al. “The power of coarse graining in biomolecular simulations”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 4.3 (2014), pp. 225–248 (cit. on p. 23).
- [50] Riccardo Baron et al. “Comparison of Thermodynamic Properties of Coarse-Grained and Atomic-Level Simulation Models”. In: *ChemPhysChem* 8.3 (2007), pp. 452–461 (cit. on p. 23).
- [51] Siewert J Marrink et al. “The MARTINI force field: coarse grained model for biomolecular simulations”. In: *The journal of physical chemistry B* 111.27 (2007), pp. 7812–7824 (cit. on p. 23).
- [52] Siewert J Marrink and D Peter Tieleman. “Perspective on the Martini model”. In: *Chemical Society Reviews* 42.16 (2013), pp. 6801–6822 (cit. on pp. 23, 39, 47).
- [53] Siewert J Marrink, Alex H De Vries, and Alan E Mark. “Coarse grained model for semiquantitative lipid simulations”. In: *The Journal of Physical Chemistry B* 108.2 (2004), pp. 750–760 (cit. on p. 23).
- [54] Semen O Yesylevskyy et al. “Polarizable water model for the coarse-grained MARTINI force field”. In: *PLoS computational biology* 6.6 (2010), e1000810 (cit. on pp. 24, 39).
- [55] Vytautas Gapsys et al. “Calculation of binding free energies”. In: *Molecular Modeling of Proteins*. Springer, 2015, pp. 173–209 (cit. on p. 28).

Bibliography

- [56] Michael R Shirts and David L Mobley. “An introduction to best practices in free energy calculations”. In: *Biomolecular Simulations: Methods and Protocols*. Springer, 2013, pp. 271–311 (cit. on p. 28).
- [57] Michael R Shirts, David L Mobley, and John D Chodera. “Alchemical free energy calculations: ready for prime time?” In: *Annual reports in computational chemistry* 3 (2007), pp. 41–59 (cit. on p. 29).
- [58] Robert W Zwanzig. “High-temperature equation of state by a perturbation method. I. Nonpolar gases”. In: *The Journal of Chemical Physics* 22.8 (1954), pp. 1420–1426 (cit. on pp. 29, 30).
- [59] John G Kirkwood. “Statistical mechanics of fluid mixtures”. In: *The Journal of Chemical Physics* 3.5 (1935), pp. 300–313 (cit. on p. 30).
- [60] Max Born. “Volumen und hydratationswärme der ionen”. In: *Zeitschrift für Physik* 1.1 (1920), pp. 45–48 (cit. on p. 30).
- [61] Thomas C Beutler et al. “Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations”. In: *Chemical physics letters* 222.6 (1994), pp. 529–539 (cit. on p. 30).
- [62] Tri T Pham and Michael R Shirts. “Identifying low variance pathways for free energy calculations of molecular transformations in solution phase”. In: *The Journal of chemical physics* 135.3 (2011), p. 034114 (cit. on p. 31).
- [63] Tri T Pham and Michael R Shirts. “Optimal pairwise and non-pairwise alchemical pathways for free energy calculations of molecular transformation in solution phase”. In: *The Journal of chemical physics* 136.12 (2012), p. 124120 (cit. on p. 31).
- [64] Jed W Pitera and Wilfred F van Gunsteren. “A comparison of non-bonded scaling approaches for free energy calculations”. In: *Molecular Simulation* 28.1-2 (2002), pp. 45–65 (cit. on p. 31).
- [65] Thomas Steinbrecher, David L Mobley, and David A Case. “Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations”. In: *The Journal of chemical physics* 127.21 (2007), p. 214108 (cit. on p. 31).
- [66] Glenn M Torrie and John P Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *Journal of Computational Physics* 23.2 (1977), pp. 187–199 (cit. on p. 32).
- [67] Johannes Kästner. “Umbrella sampling”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.6 (2011), pp. 932–942 (cit. on pp. 32, 34).
- [68] Shankar Kumar et al. “The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method”. In: *Journal of computational chemistry* 13.8 (1992), pp. 1011–1021 (cit. on p. 36).

Bibliography

- [69] Marc Souaille and Benoît Roux. “Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations”. In: *Computer physics communications* 135.1 (2001), pp. 40–57 (cit. on p. 36).
- [70] Alan M Ferrenberg and Robert H Swendsen. “New Monte Carlo technique for studying phase transitions”. In: *Physical review letters* 61.23 (1988), p. 2635 (cit. on p. 36).
- [71] Alan M Ferrenberg and Robert H Swendsen. “Optimized monte carlo data analysis”. In: *Computers in Physics* 3.5 (1989), pp. 101–104 (cit. on p. 36).
- [72] Michael R Shirts and John D Chodera. “Statistically optimal analysis of samples from multiple equilibrium states”. In: *The Journal of chemical physics* 129.12 (2008), p. 124105 (cit. on pp. 37, 38).
- [73] Charles H Bennett. “Efficient estimation of free energy differences from Monte Carlo data”. In: *Journal of Computational Physics* 22.2 (1976), pp. 245–268 (cit. on p. 37).
- [74] Zhiqiang Tan. “On a likelihood approach for Monte Carlo integration”. In: *Journal of the American Statistical Association* 99.468 (2004), pp. 1027–1036 (cit. on p. 37).
- [75] Mark James Abraham et al. “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”. In: *SoftwareX* 1 (2015), pp. 19–25 (cit. on p. 39).
- [76] MJ Abraham et al. *GROMACS User Manual*. Version 5.1.2. 2016 (cit. on pp. 39, 40).
- [77] Yifei Qi et al. “CHARMM-GUI martini maker for coarse-grained simulations with the martini force field”. In: *Journal of chemical theory and computation* 11.9 (2015), pp. 4486–4494 (cit. on p. 39).
- [78] Beauchamp, Kyle A and Chodera, John D and Naden, Levi N and Shirts, Michael R. *Python implementation of the multistate Bennett acceptance ratio (MBAR)*. Version 3.0.3. Published under the MIT license. URL: <https://github.com/choderalab/pymbar> (cit. on pp. 40, 42).
- [79] John H Kalivas. *Adaption of simulated annealing to chemical optimization problems*. Vol. 15. Elsevier, 1995 (cit. on p. 41).
- [80] Yijin Mao and Yuwen Zhang. “Thermal conductivity, shear viscosity and specific heat of rigid water models”. In: *Chemical Physics Letters* 542 (2012), pp. 37–41 (cit. on p. 41).
- [81] Sunhwan Jo et al. “CHARMM-GUI: a web-based graphical user interface for CHARMM”. In: *Journal of computational chemistry* 29.11 (2008), pp. 1859–1865 (cit. on p. 42).
- [82] Kenno Vanommeslaeghe et al. “CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields”. In: *Journal of computational chemistry* 31.4 (2010), pp. 671–690 (cit. on p. 42).
- [83] Kenno Vanommeslaeghe and Alexander D MacKerell Jr. “Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing”. In: *Journal of chemical information and modeling* 52.12 (2012), pp. 3144–3154 (cit. on p. 42).

Bibliography

- [84] Kenno Vanommeslaeghe, E Prabhu Raman, and Alexander D MacKerell Jr. “Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges”. In: *Journal of chemical information and modeling* 52.12 (2012), pp. 3155–3168 (cit. on p. 42).
- [85] Tristan Bereau and Kurt Kremer. “Automated parametrization of the coarse-grained Martini force field for small organic molecules”. In: *Journal of chemical theory and computation* 11.6 (2015), pp. 2783–2791 (cit. on p. 45).