Lateral interactions between protein molecules adsorbed onto a planar interface

Lateral interactions between protein molecules adsorbed onto a planar interface

by Niels Kouwenhoven

LUND
UNIVERSITY

Master's Thesis for the degree of Organizing Molecular Matter
Thesis advisors: Mikael Lund, Peter Jönsson
Division of Theoretical Chemistry
24th September 2017

Lateral interactions between protein molecules adsorbed onto a planar interface

by Niels Kouwenhoven

LUND
UNIVERSITY

© Niels Kouwenhoven 2017

Theoretical Chemistry, Department of Chemistry

Printed in Sweden by Media-Tryck, Lund University, Lund 2017

Contents

1 Introduction 3

2 Theory 5
- 1 Introduction to the experiment . . . 5
- 2 Generalized Van der Waals . . . 6
- 3 Derivation from the partition function. . . . 7
- 4 Virial theory . . . 11
- 5 Simulation . . . 13
- 6 Widom particle insertion . . . 14

3 Model Development 19
- 1 Monte Carlo compared to Generalized van der Waals and Virial . 20

4 Theory compared with experimental data 23
- 1 Harddisk MC simulation and Virial Theory . . . 23
- 2 Table of Proteins . . . 25
- 3 Experimental data and Monte Carlo Simulations . . . 26

5 Interactive notebook 29
- 1 Technical details . . . 29

6 Discussion 31

A Protein details 33
- 1 Titration curves . . . 33

B Simple example of a Monte Carlo simulation 35

C Widom C++ Implementation 37

Bibliography 39

Chapter 1

Introduction

In this project we aim to describe the lateral interactions between protein molecules adsorbed onto a planar interface, such as membranes or mineral surfaces. These interactions have implications for both biological function and technical applications.

The strategy is to use statistical mechanical theory to create simplified models for protein interactions and, by gradually increasing the detail level, study the role of anisotropic interactions due to molecular shape and charge distribution.The project will be done in collaboration with experimentalists and we aim to create a computational tool for analysing experimental data using generalized Van der Waals (GvdW) theory. Moreover, we will explore the validity of the simplified models using Metropolis Monte Carlo (MC) simulations. Lastly, we aim to develop a web interface for the developed model which will be publically available and easy to copy or download to modify both online and offline.

Chapter 2

Theory

The goal is to model interactions between proteins that are studied in an *hydrodynamic trapping of molecules in lipid bilayers* by *Jönsson et al.*[1] using simple models and simulations. Below is a short summary of the methods in the paper mentioned above, followed by theory of the tools that will be used to model and simulate this experiment.

The simplest description of the experiment would be to model the proteins as hard disks on a flat surface with no pair interaction potential except for space exclusion. Such a system can be modelled by Virial expansion, the Van der Waals equation of state and using a Monte Carlo system. Chapter 4 compares the predictions of these three models. The theory is discussed below.

The simple hard disk model can be improved by also modelling an interaction between the particles. This model can also be captured by Van der Waals and Generalized Van der Waals theory. Again, results will be compared to the exact Monte Carlo simulations.

Even more interesting properties can be modelled with Monte Carlo simulations. For example, the orientation of the proteins with respect to the surface may change as the concentration increases. Using results from physical experiments allows to incorporate the effect of orientation into the model in Monte Carlo simulations.

1 Introduction to the experiment

The experiment features proteins attached to a bilayer. An external flow of liquid right above the bilayer creates a hydrodynamic force on molecules protruding from the bilayer, which in turn creates a concentration gradient along the bilayer. We are interested in the relation between the chemical potential gradient and the concentration gradient.

The bilayer is a supported lipid bilayer (SLB). Streptavidin was the protein used in the original study[1]. This protein is special because it binds strongly to the biotin

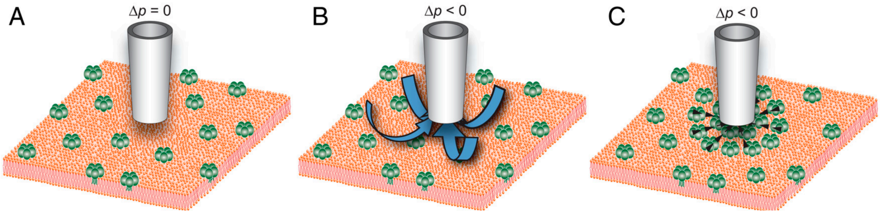


Figure 2.1: The experiment features proteins on a bilayer (A). By applying an external potential (B), the concentration of proteins varies along the bilayer (C).

receptor, which makes it useful for studies such as these. More experiments are currently being carried out for the proteins CD4 and CD2. For more data on these proteins, see section 2.

The bilayer and the proteins are submerged in an aqueous solution. A pipette was placed with the tip slightly above the surface. Siphoning the liquid from the solution into the pipette creates a force along the bilayer towards the tip of the pipette. This force in turn moves the proteins closer to the tip of the pipette creating an area of higher density.

To measure the density of the proteins on the surface, the proteins are labelled with a fluorescent marker. A picture of the intensity of the fluorescently labeled protein reveals the number density of the protein as a function of position. The force field created by the water flow can be simulated numerically. Knowledge of the density by location and force by location allows for a fitting of the chemical potential as a function of density.

2 Generalized Van der Waals

The simple Van der Waals equation of state was introduced in 1873[2] as an improvement to the ideal gas law. The assumptions made in this model are quite crude but works well at low concentrations. It takes into account the volume of particles (the excluded volume effect) and a binding energy (the net effect of attractive and repulsive forces). Although the approximation is crude, it does provide a way to analytically model a gas using its physical properties.

The well known Van der Waals equation of State in three dimensions is:

$$\left(p + \frac{n^2 a}{V^2}\right)(V - nb) = nRT \tag{2.1}$$

If the constants a and b would be zero the equation above would reduce to the ideal gas law, where p, V, n, R, and T are the pressure , volume, amount, gas constant and temperature. The constant a is the binding energy constant (with units J m^3 /

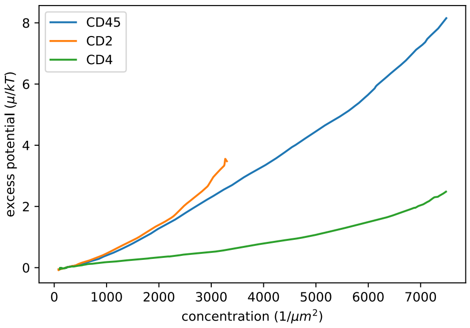


Figure 2.2: This data is result of the aggregation of many experiments, each of a shorter range of protein concentrations. These experiments were carried out with a salt concentration of 150 mM and a pH of 7. The larger proteins (CD2 has a higher excess chemical potential then the smaller CD4 proteins. These results will be further discussed in section 4

mol^2) which is a measure for how strongly the gas molecles attract each other. The constant b has units m^3 / mol and is a measure of the excluded volume per mole.

Van der Waals arrived at this equation by making several assumptions about the potential and density of the gas molecules. Equation 2.17 above is an expression for the pressure in terms of V, n, R, and T. The way that Van der Waals derived this equation is not shown here. Instead the same formula can be derived starting from the partition function and the same assumptions. This method derives an expression for the pressure (the equation of state and also for the chemical potential μ in terms of V, n, R, and T. This derivation will be shown below.

3 Derivation from the partition function.

Starting from the canonical partition function Q to describe all degrees of freedom in the system:

$$Q(V, T, N) = Q_{trans} Q_{int} \tag{2.2}$$

where Q_{trans} is the translational partition function:

$$Qtrans = \lambda^{-3N} \frac{V^N}{N!} \tag{2.3}$$

where λ is the thermal De Broglie wavelength given by:

$$\lambda = \frac{h}{\sqrt{2\pi m k_B T}} \tag{2.4}$$

where h is plack's constant, m is the mass of the particle, k_B is Boltzmann's constant and T is the temperature

The internal partition function Q_{int} is given by:

$$Q_{int} = q_{int}^N \tag{2.5}$$

where q_{int} is the internal partition function per molecule which is:

$$q_{int} = \exp\left(-\frac{E}{k_B T}\right) \tag{2.6}$$

where E is an internal energy

Combing equations 2.2 - 2.6 is the starting point of the derivation:

$$Q = \frac{1}{\lambda^{3N} N!} V^N \exp\left(\frac{-NE}{k_B T}\right) \tag{2.7}$$

From here Van der Waals made two assumptions:

- The volume of each particle reduces the volume available to other particles. Therefore, let the free volume V_{free} be the total volume minus the excluded volume: $V_{free} = V - V_{excluded}$
- The Mean field approximation: assume that the potential energy per particle only depends on the particle density n.

Or in formulas:

$$V_{free} = V - V_{excluded} = V - bN \tag{2.8}$$

where b is the excluded volume per molecule. With the mean field equation the interaction energy is given by:

$$E = \frac{1}{2} \int_0^\infty 4\pi r^2\, n(r)\, u(r)\, dr = \frac{1}{2} N/V \int_0^\infty 4\pi r^2\, g(r)\, u(r)\, dr = N/Va \tag{2.9}$$

where $u(r)$ is the pair interaction potential and $g(r)$ is the radial distribution function. The integral will result in a single value, a, which is the binding energy constant.

We now have a partition function from which the Van der Waals Equation of State can be derived:

$$Q = \frac{(V - bN)^N}{\lambda^{3N} N!} \exp\left(\frac{-N^2 a}{V k_B T}\right) \tag{2.10}$$

We'll need Stirlings' approximation to simplify the factorial term:

$$\ln(N!) \approx N\ln(N) - N \tag{2.11}$$

Or equivalently:

$$N! \approx \frac{N^N}{e^N} \tag{2.12}$$

Using the approximation gives:

$$Q = \left(\frac{(V-bN)e}{\lambda^3 N}\right)^N \exp\left(\frac{-N^2 a}{Vk_B T}\right) \tag{2.13}$$

From here we can calculate the Helmholtz free energy A:

$$\begin{aligned} A &= -kT\ln Q \\ &= kTN\ln\left(\frac{\lambda^3 N e^{-1}}{V-bN} + \frac{N^2 a}{V}\right) \end{aligned} \tag{2.14}$$

The pressure can now be calculated using the thermodynamical relation (at constant temperature T):

$$p = -\left(\frac{\delta A}{\delta V}\right)_T \tag{2.15}$$

which gives:

$$p = \frac{NkT}{V-bN} - \frac{N^2 a}{V^2} \tag{2.16}$$

which can be rewritten to

$$\left(p + \frac{n^2 a}{V^2}\right)(V - nb) = nRT \tag{2.17}$$

which is the same as equation 2.17 as we wanted to show.

Now, we are interested in the chemical potential μ instead of the pressure p and would like to work in 2 dimensions instead of 3. This derivation is shown in the next paragraph:

3.1 Van der Waals in two dimensions

We'll start from equation 2.13 above but modify it for two dimensions. The factor λ^3 is replaced by the factor λ^2 and the volume V is replaced by the surface Y. Now the constant b means excluded surface instead of excluded volume. The integrals for calculating a are slightly different but the idea is the same.

$$Q = \left(\frac{(Y - bN)e}{\lambda^2 N}\right)^N \exp\left(\frac{-N^2 a}{Y k_B T}\right) \tag{2.18}$$

Now the Helmotz free energy is:

$$\begin{aligned} A &= -kT \ln Q \\ &= kTN \ln\left(\frac{\lambda^2 N e^{-1}}{Y - bN}\right) + \frac{N^2 a}{Y} \end{aligned} \tag{2.19}$$

Now the chemical potential can be calculated with the relation:

$$\begin{aligned} \mu &= \frac{\delta A}{\delta N} \\ &= kT\left(+2\ln\lambda + \ln\left(\frac{N}{Y - bN}\right) - \frac{Y}{bN^2 - NY}\right) + \frac{2Na}{Y} \end{aligned} \tag{2.20}$$

To simplify this expression we define $\hat{a} = a/kT$. Furthermore, absolute values of the chemical potential are here unimportant, so we ignore the constant term $2\ln\lambda$:

$$\mu/kT = \ln\left(\frac{N}{Y - bN}\right) - \frac{bN}{bN - Y} + \frac{2N\hat{a}}{Y} \tag{2.21}$$

and introduce the quantity $n = N/Y$, the surface number density to simplify the expression:

$$\mu/kT = \ln\left(\frac{n}{1 - bn}\right) + \frac{bn}{1 - bn} + 2n\hat{a} \tag{2.22}$$

In an ideal gas, where the particles have no size and don't interact, the values for a and b would be zero. So for an ideal gas, the potential is:

$$\mu/kT = \ln n \tag{2.23}$$

4 Virial theory

Virial theory is another way to describe the equation of state for a gas, where the pressure of the gas is expressed as a polynomial of density:

$$\frac{p}{kT} = \rho + B_2(T)\rho^2 + B_3(T)\rho^3 + \ldots + B_m(T)\rho^m \quad = \rho + \sum_i B_i(T)\rho^i \tag{2.24}$$

where k is the Boltzmann constant and ρ is the density in either two or three dimensions. The derivations below are in terms of ρ and are thus valid for both two and three diminsions.

The first term ρ has no coefficient, that is the term for the ideal gas law. Higher terms B_i are called Virial coefficients. This expression is an approximation for the pressure at low densities.

To transform this into an expression for chemical potential μ we use this relation between the Gibbs free energy G, and the pressure p:

$$\frac{d\mu}{kT} = \frac{dG}{N} = \frac{Y}{N}dp = \frac{dp}{\rho} \tag{2.25}$$

The chemical potential can then be found through the following integral:

$$\mu = \int_{p_0}^{p} \frac{1}{\rho} dp' = \int_{\rho_0}^{\rho} \frac{1}{\rho}\frac{dp}{d\rho'} d\rho' \tag{2.26}$$

Plugging in equation 2.24 into the expression above yields:

$$\begin{aligned} \frac{\mu}{kT} &= \int_{n_0}^{\rho} dn' \frac{1}{\rho}\frac{d}{d\rho'}\left(\rho + \sum_{i=2} B_i \rho^i\right) \\ &= \int_{n_0}^{\rho} dn' \frac{1}{\rho}\left(1 + \sum_i B_{i=2}\, i\, \rho^{i-1}\right) \\ &= \int_{n_0}^{\rho} dn' \left(\frac{1}{\rho} + \sum_{i=2} B_i\, i\, \rho^{i-2}\right) \\ &= \ln \rho + \sum_{i=2} B_i \frac{i}{i-1}\rho^{i-1} \end{aligned} \tag{2.27}$$

The first few terms look like:

$$\frac{\mu}{kT} = \ln \rho + 2B_2\rho + \frac{3}{2}B_3\rho^2 + \ldots \tag{2.28}$$

4.1 Virial coefficients for the Van der Waals equation of state

Using Taylor expansion the Van der Waals equation of state can be written in the form of a Virial expansion to compare them.

From equation 2.16:

$$\frac{p}{kT} \approx n + (b - \hat{a})\, n^2 + b^2 n^2 + \ldots \tag{2.29}$$

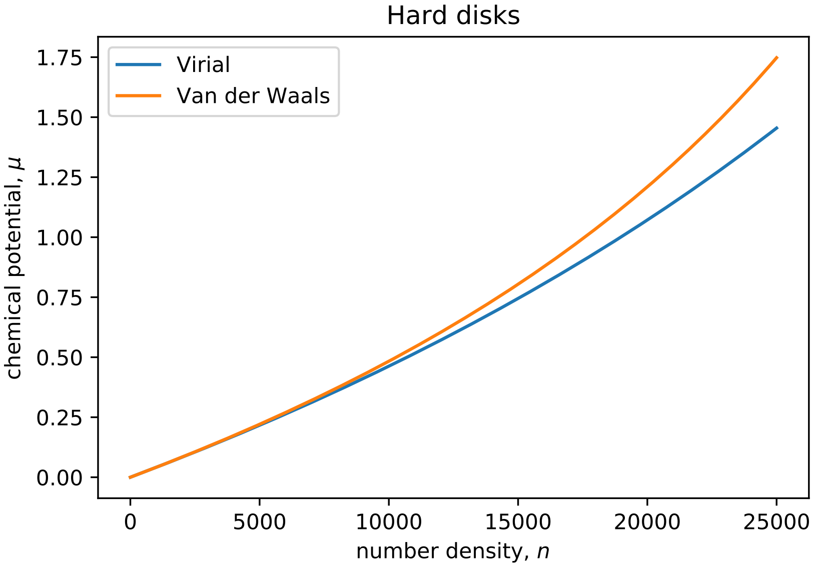


Figure 2.3: Figure The excess chemical potential for hard disks as expressed by Van der Waals theory and Virial expansion. The Virial coefficients for the plotted line are those up to B_9

4.2 Virial coefficients for hard disks

The Virial coefficient B_i represents the ith order interaction. For hard disks the interaction can be calculated analytically for the first 3 coefficients. From *Hemmer*[3]:

$$\begin{aligned} B_2 &= 2 \\ B_3 &\approx 3.128017947 \\ B_4 &\approx 4.257854656 \end{aligned} \tag{2.30}$$

Higher terms have been calculated in Monte Carlo simulations. From *Hu and Yu*[4]:

$$
\begin{aligned}
B_5 &= 5.33689664 \\
B_6 &= 6.363 \\
B_7 &= 7.352 \\
B_8 &= 8.319 \\
B_9 &= 9.272
\end{aligned}
\tag{2.31}
$$

5 Simulation

The analytical methods introduced above are useful for simple interactions but are limited for more complex interactions. The equation for the potential derived from the Van der Waals equation of state captures the binding energy in a single constant. A numerical simulation of particles allows interactions, charge distributions and steric interactions to be taken into account. In this paper we will use the software tool Faunus[5] for numerical simulations of gasses of particles. It uses Markov-chain Monte Carlo simulations for its simulations. This technique will be introduced below

5.1 Metropolis Monte Carlo simulation

The "Monte Carlo" method is numerical method for calculating quantities using random sampling. The Metropolis algorithm is an application of the Monte Carlo method that uses importance sampling.[6] It provides a straightforward algorithm for approximating integrals of the type:

$$
\langle E \rangle = \frac{\int d\mathbf{r}^N E(\mathbf{r}^N)\exp(-\beta u(\mathbf{r}^N))}{\int d\mathbf{r}^N \exp(-\beta u(\mathbf{r}^N))} \tag{2.32}
$$

where $\beta = 1/kT$ and $\exp(-\beta u(\mathbf{r}^N))$ is the Boltzman factor that weights the contribution of that energy state,

The idea is to favor the sampling of configurational states that are more likely and thus contribute more to the above integral. Instead of generating random states afresh, the Metropolis algorithm generates a new state from a perturbation from a previous state and favor it in a ratio based on their relative likeliness. That is, from the old state, make a trial move to a new state, and with some acceptance function acc(old → new), accept the new state, or reject and maintain the old state. In the equation above it is the Boltzmann factor that determines the likeliness of the state. Thus the likeliness of transitioning from an old state to a new $\pi(\text{old} \rightarrow \text{new})$ should satisfy:

$$
\frac{\pi(\text{old} \rightarrow \text{new})}{\pi(\text{new} \rightarrow \text{old})} = \frac{\text{acc}(\text{old} \rightarrow \text{new})}{\text{acc}(\text{new} \rightarrow \text{old})} = \exp\left(-\beta\left(u\left(\text{new}\right) - u\left(\text{old}\right)\right)\right) \tag{2.33}
$$

(The above only holds if the underlying Markov Matrix is symmetric). The standard acceptance function that satisfies the above is:

$$\text{acc}(\text{old} \to \text{new}) = \min\left(1, \exp\left(-\beta(u(\text{new}) - u(\text{old}))\right)\right) \tag{2.34}$$

Thus randomness is used in two circumstances. Firstly a trial move is generated by randomly translating, rotating or otherwise moving any part of the configuration. The acceptance for this transition is calculated, and a second random number is generated to decide whether the move is accepted or rejected.

The algorithm in words is:

1. make a trial move
2. calculate the acceptance probability a
3. generate a random number r on the interval $[0, 1\rangle$
4. accept the move if $a > r$ else reject

6 Widom particle insertion

The previous section describe how to the Monte Carlo effectively samples the relevant configuration of a system of particles. In this section we are interested in how we can calculate the chemical potential for such configurations. One such method is known as the "Widom insertion method", which was first introduced by B. Widom in his paper "Some Topics in the Theory of Fluids"[7].

In the section 3.1 above on Van der Waals theory we used the following formula for the chemical potential: (*i.e.* how does the energy change when the number of particles changes)

$$\mu = \frac{\partial A}{\partial N} \tag{2.35}$$

The same rationale can be applied to the numerical simulation: How does the energy of the configuration change if one particle were added to it? This is what the Widom insertion method does. In formula:

$$\mu_{\text{excess}} = -kT \ln \left\langle \exp\left(-\beta\psi(\mathbf{r})\right) \right\rangle_{\mathbf{r}} \tag{2.36}$$

Where $\beta = 1/kT$ and ψ is the interaction potential of the inserted particle with all the other particles, $\mathbf{r}$ is the coordinate vector and $\langle\,\rangle_{\mathbf{r}}$ signifies the average over all space.

The equations below aim to derive 2.36

Consider the configuration integral Z_N of a fluid.

$$Z_N = \int \cdots \int_V \exp\left(-\beta W_N(\mathbf{r}_1, \ldots, \mathbf{r}_N)\right)\, d\mathbf{r}_1 \ldots d\mathbf{r}_N \tag{2.37}$$

Where W_N is the potential interaction energy of all particles as a function of the positions of all particles.

Then consider adding a single particle. Let ψ be the interaction potential of the added particle with all the other particles.

$$\begin{aligned} Z_{N+1} &= \int \cdots \int_V \exp\left(-\beta W_N(\mathbf{r}_1, \ldots, \mathbf{r}_N)\right) \exp\left(-\beta \psi(\mathbf{r}_{N+1})\right)\, d\mathbf{r}_1 \ldots d\mathbf{r}_{N+1} \\ &= Z_N \int_V \exp(-\beta\psi(\mathbf{r}))\, d\mathbf{r} \\ &= Z_N V \langle \exp(-\beta\psi(\mathbf{r}))\rangle_{\mathbf{r}} \end{aligned} \tag{2.38}$$

The Helmoltz free energy, A, can be calculated from the configuration integral which in turn can be used to calculate the chemical potential:

$$\begin{aligned} A &= -kT \ln Z \\ \mu &= \frac{\partial A}{\partial N} = -kT \frac{\partial \ln Z}{\partial N} \end{aligned} \tag{2.39}$$

The derivative in 2.39 can be estimated directly using the result from [eqn:widom_conf].

$$\begin{aligned} \frac{d \ln Z}{dN} &\approx -\frac{\ln Z_N - \ln Z_{N+1}}{1} \\ &= \ln \frac{Z_{N+1}}{Z_N} \end{aligned} \tag{2.40}$$

The excess potential is then:

$$\mu_{\text{excess}} = -kT \ln \left\langle \exp\left(-\frac{\psi(\mathbf{r})}{kT}\right)\right\rangle_{\mathbf{r}} \tag{2.41}$$

6.1 Dilute hard spheres

As an example, we could use the Widom insertion method to calculate the potential for hard disks. The interaction potential for hard disks is either 0 or ∞ so the exponent term in 2.42 is either 1 or 0. Consider a dilute gas of harddisks as shown in figure \ref{fid{hardwidom} below. The gas is assumed to be so dilute that the chances of two particles being close enough to have overlapping excluded areas is assumed to be negligibly small.

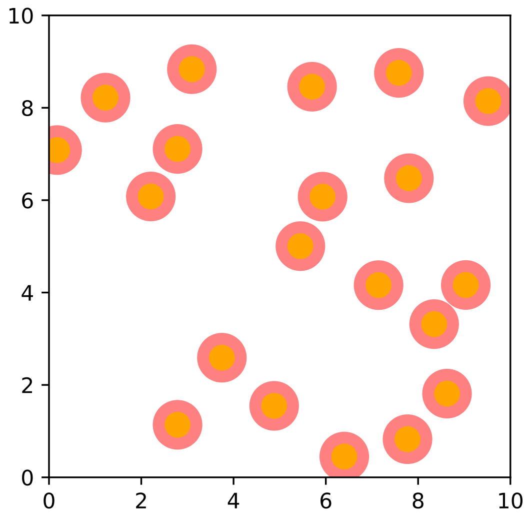


Figure 2.4: A box filled with hard disks. In this figure the particles are colored orange. Surrounding each particle there is an area where (the center) of other particles can not be: the excluded area. In this figure the excluded area of each particle is colored red.

Let the radius of disks with r. Now consider what happens if a ghost particle is inserted. There are only two options, either it overlaps with another particle, or it does not. An overlap would occur if the center of the particle is inserted anywhere in the excluded area of the particles in the box (the red areas in the figure). If the particles overlap the interaction potential is ∞, and if they do not the interaction potential is 0. Since there are only two values that the interaction potential can take we can expand the expectation value in equation 2.42:

$$\begin{aligned}\mu_{\text{excess}} &= -kT\ln\left\langle \exp\left(-\frac{\psi(\mathbf{r})}{kT}\right)\right\rangle_{\mathbf{r}} \\ &= -kT\ln\left(\exp(-\infty)p_{\text{overlap}} + \exp(-0)(1-p_{\text{overlap}})\right) \\ &= -kT\ln\left(0\cdot bn + 1\cdot(1-bn)\right) \\ &= -kT\ln\left(\cdot(1-bn)\right)\end{aligned} \tag{2.42}$$

where n is the surface particle density and b is half the excluded area per particle. (The definition of b is the same as in the section on van der Waals theory, see section 3.1)

The chemical potential is then (adding the ideal term $kT\ln n$):

$$\mu = kT\ln\left(\frac{n}{1-2bn}\right) \tag{2.43}$$

By taking a series expansion the expression above and for the expression for the chemical potential derived for hard disks using the Van der Waals theory in equation 2.22 for the variable n around 0 we see that the terms are the same up to the first order series expansion:

$$\frac{\mu}{kT} \approx \ln n + 2bn + O(n^2) \tag{2.44}$$

6.2 Monte Carlo simulations

In his paper B. Widom describes the averaging as follows:[7]

> Imagine that in a fluid which is in thermodynamic equilibrium the motion of the molecules suddenly ceases at some arbitrary instant, so that the molecules become rigidly fixed in one of the infinitely many possible configurations characteristic of the equilibrium fluid. Now add to the system another molecule, identical to those of which the fluid is composed, and let it wander among the fixed molecules, measuring, at each point, its energy of interaction ψ with the fixed molecules, or $\exp(-\psi/kT)$, or any other function of position. Then the average value of such a function of position, assigning equal weight to equal elements of volume in evaluating the average, is symbolized by $\langle\,\rangle$.

He has described exactly what we would do in a Monte Carlo simulation. At every step of the simulation we already have a "rigidly fixed configuration". We add a test particle and let it "wander" by repeatedly inserting and removing the same particle.

We have implemented the Widom insertion method into the Faunus[5] Monte Carlo simulation framework. The `C++` code of the implementation is included in appendix C

6.3 Weeks Chandler Andersen potential

The hard sphere potential is simple but the discontinuity of its step function makes it impossible to calculate forces. The Weeks Chandler Andersen (WCA) potential is a continuous potential which can approximate a hard sphere potential by choosing the parameters such that it becomes very steep.

The WCA potential is similiar to the Lennard-Jones (LJ) potential but it is *cut* and *shifted* at the minimum. See figure 2.5. This separates the repulsive part and the attractive part of the potential.

$$u(r) = 4\epsilon \begin{cases} \left(\frac{b}{r}\right)^{12} - \left(\frac{b}{r}\right)^{6} + 1/4 & r < r_{min} \\ \left(\frac{b}{r}\right)^{12} - \left(\frac{b}{r}\right)^{6} & r > r_{min} \end{cases} \tag{2.45}$$

By setting the ϵ to a very large value value and ignoring the attractive part, the

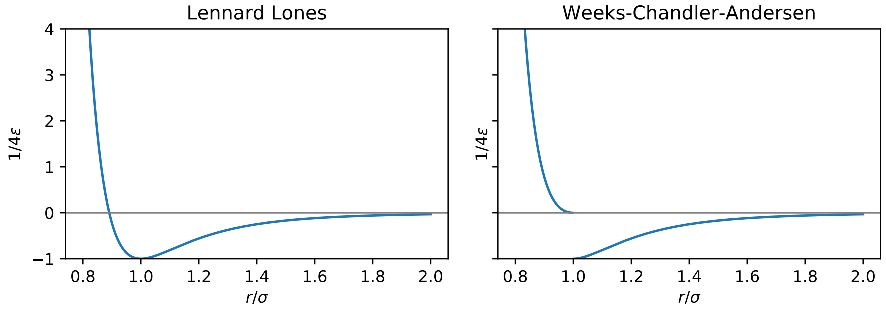


Figure 2.5: Lennard-Jones and Weeks-Chandler-Andersen potentials

potential effectively becomes a hard-sphere potential, but it remains continuous everywhere so that forces can be calculated.

6.4 Electrostatic potential

In Monte Carlo simulations, water molecules of the solution are not explicitly included in the simulation. Instead the water is modelled implicitly as a continuous dielectric medium. The presence of water effectively attenuates the strength of the electric field. The electric field is described with a screened coulomb potential. (or Yukawa potential)[8].

$$u(r) = \frac{\lambda_B z_i z_j}{r} \exp\left(-\frac{r}{\lambda_D}\right) \tag{2.46}$$

where z_i and z_j are the charges of the interacting particles, r is the distance between the particles. The Bjerrum length λ_B and the Debye length λ_D are defined below:

$$\begin{aligned} \lambda_B &= \frac{e^2}{4\pi\epsilon_0\epsilon_r k_B T} \\ \lambda_D &= \sqrt{\frac{4\pi q^2}{k_B T}} \end{aligned} \tag{2.47}$$

where e is the elementary charge, ϵ_0 is the vaccuum permittivity and ϵ_r is the relative dielectric constant. For water at room temperature, ϵ_r is about 80 which makes λ_B equal to about 0.7 nm.

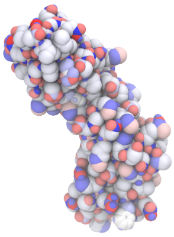
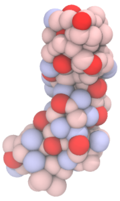
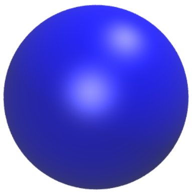
Chapter 3

Model Development

Running a simulation for a large surface full of proteins with an exact molecular structure and charge distribution is computationally expensive. Each additional particle could interact with all the existing particles, and the computation time thus increases quadratically with the number of particles. We could therefore reduce the number of molecules in the simulation or the number of atoms that these molecules are made out of. In the extreme case we can replace the entire protein by a single charged sphere. In this way we retain the effects of the volume exclusion and the monopole electrostatic interactions, but lose the effects of steric interactions and higher order electrostatic interactions (dipole, quadrupole, etc.). It might just be the case that these interaction are dominating the potential at least at low densities. An additional advantage of the simplification is that it allows for analytical solution to be found for the potential.

We can start with the most basic case: a hard disk. It has only repulsive interactions, no orientation, it only occupies space. This provides a good baseline. The chemical potential is defined as the amount of free energy it takes to add a single particle to the system. Hard disks do take space, and therefore it is harder to add a particle to a surface of high density of hard disks than a low density.

Table 3.1: Coarse-graining of CD4

		
The complete structure of the protein 1CDJ shown here with all of its 1384 relevant (excluding hydrogen) atoms.	Each amino acid has been replaced by a single sphere with the same size, charge and center of mass as the sum of its constituent atoms	A single sphere a represent the protein makes for the fastest calculations.

1 Monte Carlo compared to Generalized van der Waals and Virial

We are interested in the chemical potential of hard disks as a function of surface coverage. We can use several methods to plot this relation. Firstly, Generalized Van der Waals theory yields a simple analytical expression that approximates the chemical potential. Secondly, from the standard Virial expansion of the relation between compressibility and surface coverage we can obtain an expression for the chemical potential. Finally experimental data obtained from Peter.

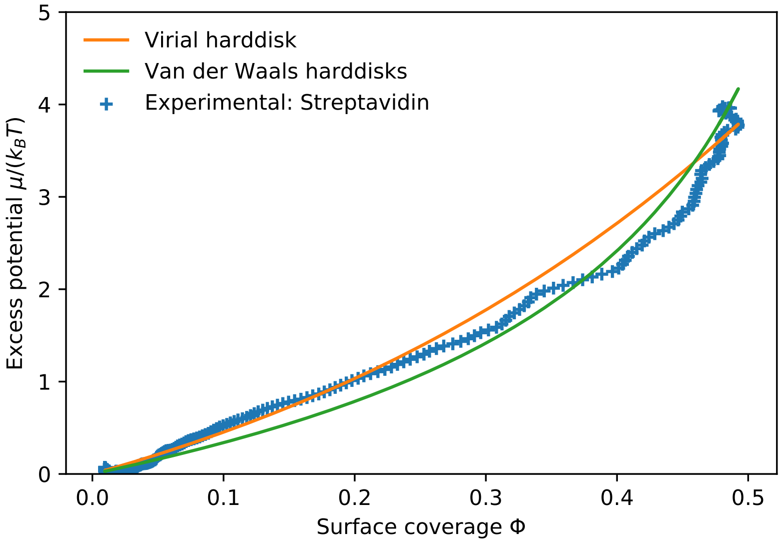


Figure 3.1: Virial and Experiment (Streptavidin 150 mM NaCl)

Chapter 4

Theory compared with experimental data

1 Harddisk MC simulation and Virial Theory

The potential of hard disks can be very accurately modelled by Virial theory. The expression for the potential expressed in terms of surface coverage in Virial theory is shown in (See eqn. (2.24). Using the coefficients for this expression (See eqn. 2.30, 2.31) we can plot this relation for hard disks. This relation is plotted below in Figure 4.1.

Since the Virial theory provides such a reliable estimation for the potential this reinforces that the Widom insertion method has been implemented correctly.

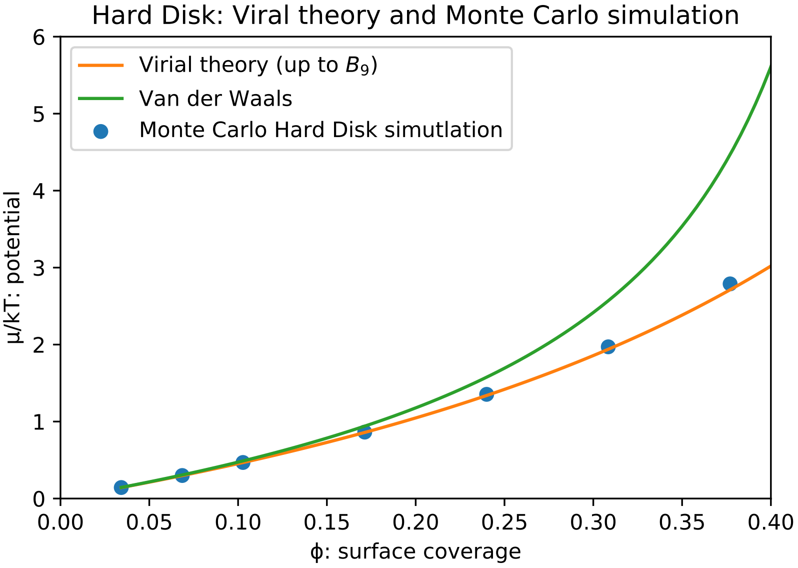


Figure 4.1: Figure Comparison of Monte Carlo simulation, Virial theory and Van der Waals theory

2 Table of Proteins

The original study on which this study is based[1] studies streptavidin. Current research by Peter Jönsson extends this research by studying a wider range of proteins which are similar in structure to streptavidin: CD2 and CD4. (the name CD stands for cluster of differentiation). Each of these proteins is found on the surface of T-cell and are thus involved in the immune response. These proteins cover a wide range of sizes from 178 residues for CD4. The structure data was obtained from the Protein Data Bank (PDB[9]) and UniProt[10] in a pdb format. This file was subsequently uploaded to the PDB2PQR Server to prepare the structure for continuum electrostatics calculations[11]. The service processes the uploaded `.pdb` file and estimates the titration states and protonating biomolecules in a manner consistent with favorable hydrogen bonding. This process yields a `.pqr` file which includes the estimated (partial) charges. The service also uses APBS (Adaptive Poisson-Boltzmann Solver)[12] to calculate the electrostatic potential around the molecule. The structural `.pqr` data along with the spatial electrostatic data were loaded into VMD[13] to produce the thumbnails below (red areas designate positive charge and blue negative).

The size of the proteins was determined numerically, by randomly sampling points in a box enclosing the molecule and checking if the point was contained in the molecule or not. This feature is available in `Faunus.geometry::calcVolume.`[5].

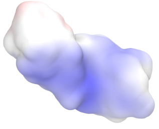
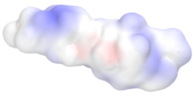
Name	CD4	CD2
PDB id	1CDJ	1HNF
UniProt	P01730	P06729
Thumb		
Residues	178	351
Charge (at pH=7)	4.27 - 4.42	5.48 - 5.66
Volume (nm^3)	18	18.6

Table : The method of calculating the charge and the volume is described in the text above.

3 Experimental data and Monte Carlo Simulations

One of the goals of the Monte Carlo simulations was to find the effect of charges on the chemical potential. Using a coarse-grained model of 1CDJ (amino acids as atoms) we set up two similar simulations. In one of the simulations all the charges were neutralized. Both of the simulations had the following setup: 20 proteins were put in a plane (a cuboid geometry with one coordinate constrained to be zero) with a Weeks-Chandler-Andersen potential for the interaction of neutral atoms and an additional Yukawa potential for the screened electrostatic interaction. Both setups were the same, but in one of the two all charges were set to 0. Using Widom insertion analysis the potential of both systems was calculated for both setups. The results are shown below in figure 4.2:

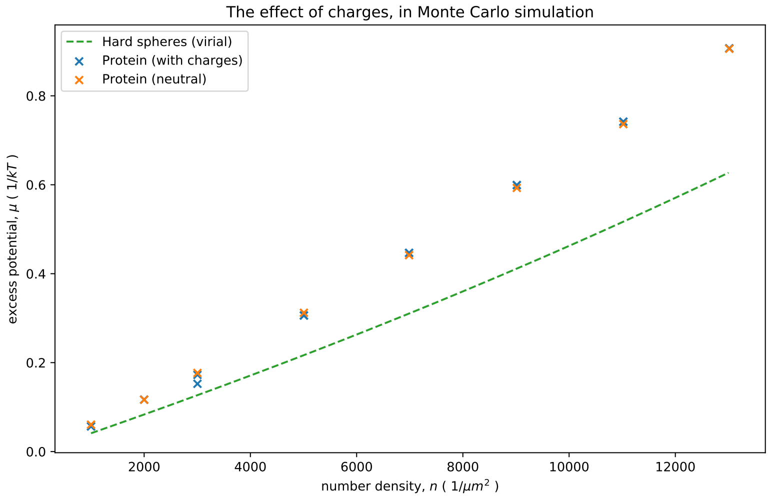


Figure 4.2

The proteins in this simulations (1CDJ) have a size of $18\ \text{nm}^3$. The dashed line in figure 4.2 plots the potential of hard disks of the same size using the Virial expansion (eq 2.28). The plot above shows that the shape of the proteins does have an influence on the potential, but the charges do not influence the potential much. It is known that the electrostatic interactions are shielded by the presence of salt in the solution. It could be that a stronger effect of the electrostatic interaction is found at lower salt concentrations.

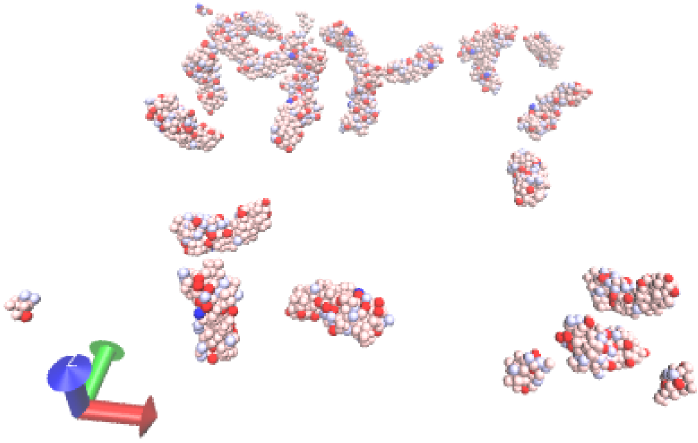


Figure 4.3: Snapshot of the simulation. This experiment has a square box of 400Å where the centers of mass of the 20 proteins are fixed to the x, y-plane. The proteins interact with Weeks-Chandler-Andersson potentials of strength $0.005kT$ and a Yukawa electrostatic interaction with a Debye-length of 10, which correspons to an electrolyte concentration of 0.1 M.

Chapter 5

Interactive notebook

One of the goals of this thesis is to make theory that it discusses more easily available. To this end I have created an online tool which enables interactive exploration of the Van der Waals interactions for proteins in two dimensions. The tool is aimed at experimentalists, and make it easy for them to upload their own data, compare their data with some already defined potentials, change the parameters for those potentials to see how they affect the potential, and finally to define their own potentials themselves.

The tool was created using a Jupyter notebook, which is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text.[kluyver2016jupyter]

The source code for the interactive notebook is available on Github, where it can be viewed, downloaded or copied by anyone.[1]

A free live hosted example of the notebook is available on the Microsoft Azure cloud[2]

1 Technical details

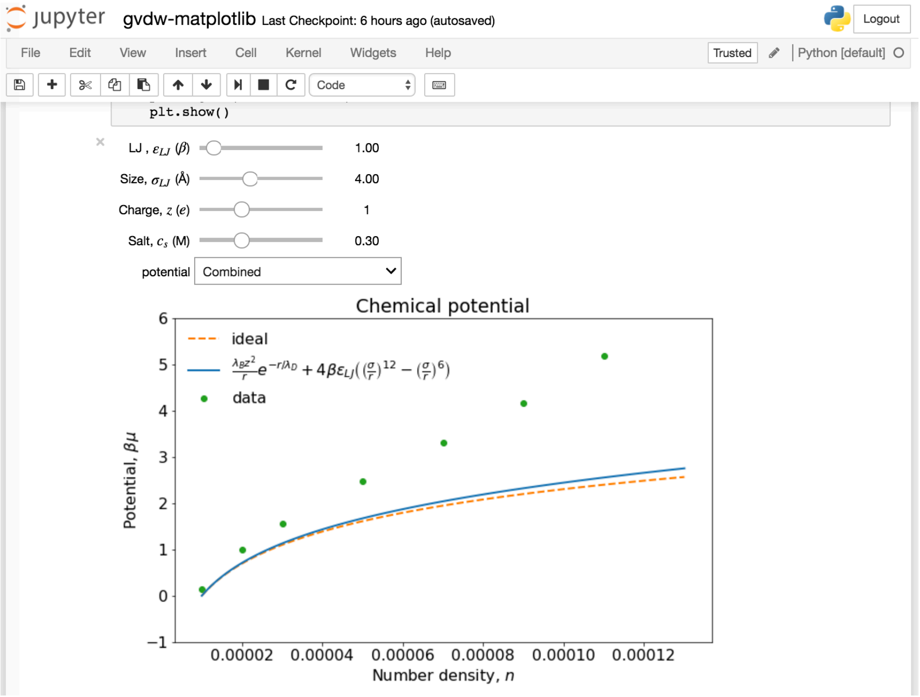
I have created a interactive example that uses the python plotting library `matplotlib` which renders high quality graphics. The interactivity is provided by the Jupyter notebook. When the value of the sliders changes a new png image is drawn by `matplotlib`. The old static image is quickly replaced by a new static image. The result is a high quality image

Additionally I have created a version of the notebook which uses the `bokeh` plotting library which produces a dynamically generated graph, which allows zooming

---

[1] The full link is https://github.com/Zweedeend/interactive-2d-gvdw . The url shortened link is goo.gl/uok86F

[2] The full link is notebooks.azure.com/kallgaard/libraries/interactive . The url shortened link is goo.gl/NZSvm2


jupyter
gvdw-matplotlib Last Checkpoint: 6 hours ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python [default]
Code
Chemical potential
ideal
$\frac{\lambda_B z^2}{r} e^{-r/\lambda_D} + 4\beta\varepsilon_{LJ}((\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^6)$
data
Potential, $\beta\mu$
6
5
4
3
2
1
0
−1
0.00002
0.00004
0.00006
0.00008
0.00010
0.00012
Number density, n

```
plt.show()
```

LJ, ε_{LJ} (β) 1.00

Size, σ_{LJ} (Å) 4.00

Charge, z (e) 1

Salt, c_s (M) 0.30

potential Combined

Figure 5.1: Interactive Matplotlib example

and panning within the graph. Since the data can be updated dynamically, the interaction is much more smooth.

Chapter 6

Discussion

In this project I aimed to describe the lateral interactions between protein molecules adsorbed onto a planar interface. I mainly focused on finding a effect of the protein density on the chemical potential. I have done this using theoretical considerations of Generalized Van der Waals theory and the Virial theory and numerical considerations using Monte Carlo simulations.

Considering the proteins as hard disks on a planar surface already gives a rough estimation of the chemical potential. Especially the Generalized Van der Waals theory is useful for exploring the effect of interaction potentials on the chemical potential. Since this relationship is a simple function it can be plotted interactively as I have done in a an interactive Jupyter notebook.

By using Monte Carlo simulations, a numerical attempt was made to explore the effect of the structure of and charge distribution of the protein on the chemical potential. To investigate the chemical potential the Widom insertion method was implemented for Faunus. I first performed a few simulations with hard disks to verify that it matches with Virial theory for hard disks which is a very good approximation for hard disks.

For a single protein (1CDJ) I have performed simulations using a coarse-grained model of the protein structure. The molecules were constrained in a 2D plane. The experiment was repeated for several box sizes to find the relation between the number density of proteins and the chemical potential. This led to a few preliminary results shown in figure 4.2. These results suggest that the charges are screened by the presence of salt.

The simulations that were performed were relatively simple because the space was restrained to 2 dimensions, the number of molecules was as low as 20, and the protein was coarse-grained to substitute amino acids for single 'atoms'.

To expand on this research it would be interesting to restict the rotational movement of the proteins. In the biological system the proteins are attached to the bilayer on one end of the protein. This tethering resticts the rotational freedom of the protein. This has not been taken into account yet in my model. Perhaps a simulation could be set up where not the protein center of mass but rather the tether point is restricted on the 2D plane. Also the center of mass would be restrained to stay above the plane. This would mimic the tethering.

To expand on the tethering there is an interesting paper *Glycosylation and Lipids Working in Concert Direct CD2 Ectodomain Orientation and Presentation*[14] which concerns exactly the 1CDJ protein (CD4). In the paper the researchers have enumerated the probabilties for angles of the protein relative to the surface. These probabilities can be translated into energies using the Boltzmann relation. These energies can then be incorporated in the Monte Carlo simulations.

Appendix A

Protein details

The images below show a high-resolution representation of the proteins listed in table 2.

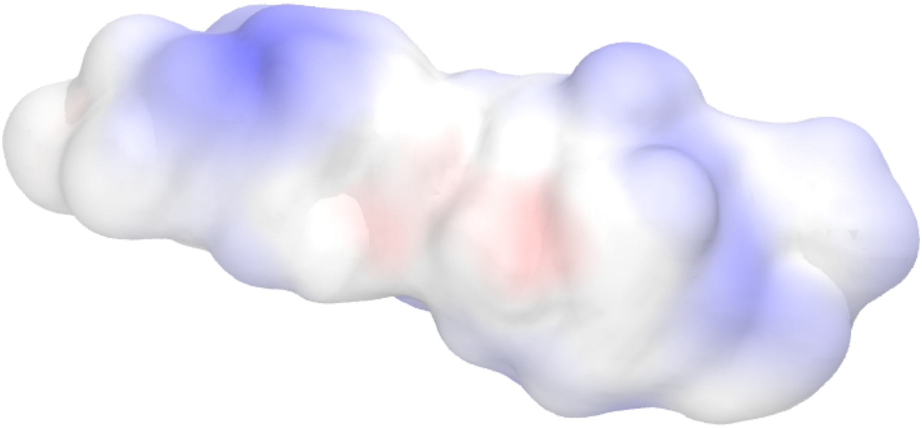


Figure A.1: 1HNF

1 Titration curves

One of the outputs that Propka provides is a titration curve. These are shown below. The Monte Carlo simulations in this experiment are run at neutral pH.

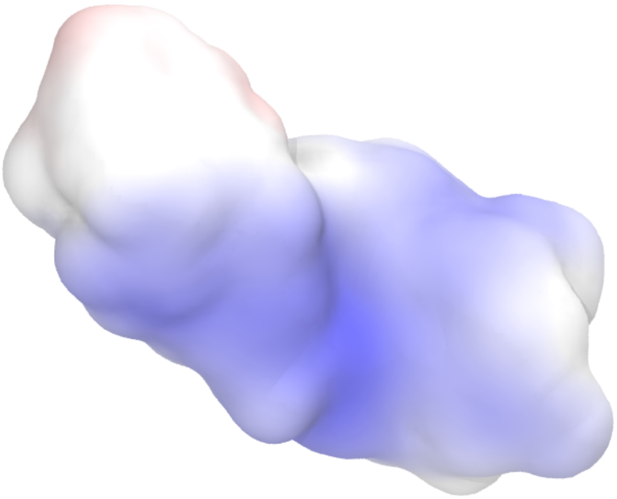
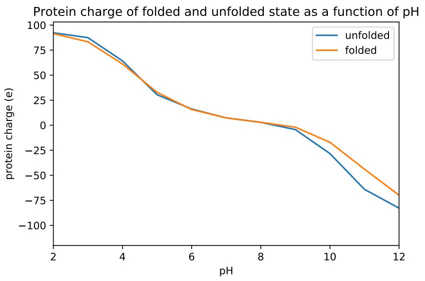
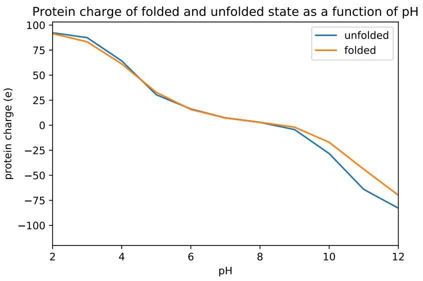


Figure A.2: 1CDJ





Appendix B

Simple example of a Monte Carlo simulation

To explore techniques of Monte Carlo simulations I created a small python script that illustrates the workings that has all of the basic parts.

This program simulates 100 particles in a one-dimensional harmonic potential. The particles are first uniformly distributed on the interval $[-1, 1]$. We take 100,000 steps in total, and every 1000 steps we take a sample of the total energy.

```python
import random
import math

# constants
BETA = 10  # beta = 1/kT
MACRO_STEPS = 1000
MICRO_STEPS = 1000
DX = 0.1  # displacement

positions = []
samples = []
for i in range(1000):
    positions.append(random.randrange(-1, 1))


def energy(x):
    "Harmonic potential"
    return x * x


def mcmove():
    "Attempt to move 1 randomly chosen particle by
     a random displacement"
    particle_id = random.randrange(len(positions))
    old = positions[particle_id]
    new = old + random.uniform(-0.5, 0.5) * DX
    energy_change = energy(new) - energy(old)
    if random.random() < math.exp(- BETA * energy_change):
        positions[particle_id] = new


def sample():
    "Calculate the total energy in the system"
    total_energy = 0
    for pos in positions:
        total_energy += energy(pos)
    samples.append(total_energy)

def main():
    "Main loop"
    for step in range(MACRO_STEPS):
        for micro_step in range(MICRO_STEPS):
            mcmove()
        sample()

# start the main loop
main()
```

Simple Monte Carlo Simulation

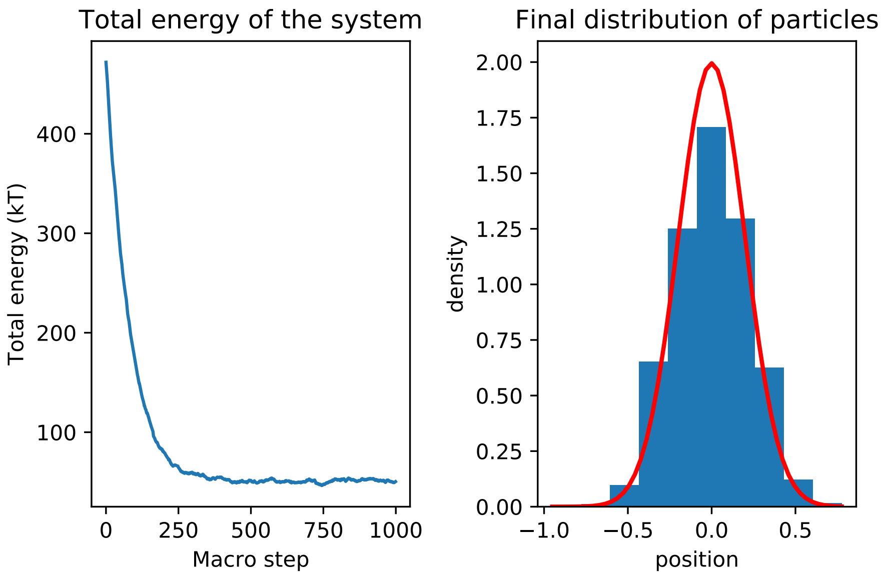


Figure B.1: The plot on the left shows the total energy in the system at each of the samples we took in each macro step. The histogram on the right shows the distribution of the particles at the end of the experiment

Appendix C

Widom C++ Implementation

Implemented in `Faunus/include/faunus/analysis.h` in the namespace `Faunus::Analysis`

```
template<typename Tspace>
class WidomMolecule : public AnalysisBase {
    private:
        typedef Energy::Energybase <Tspace> Tenergy;
        Tspace *spc;
        Energy::Energybase <Tspace> *pot;
        int ninsert, molid;
        string molecule;
        Point dir;
    public:
        Average<double> expu, rho;

    void _sample() override {
        typedef MoleculeData<typename
            Tspace::ParticleVector> TMoleculeData;
        auto rins = RandomInserter<TMoleculeData>();
        rho += spc->numMolecules(molid) / spc->geo.getVolume();
        rins.dir = dir;
        rins.checkOverlap = false;
        for ( int i = 0; i < ninsert; ++i ) {
            // ('spc->molecule' is a vector of molecules
            auto pin = rins(spc->geo, spc->p, spc->molecule[molid]);
            // energy between "ghost molecule" and system in kT
            double u = pot->v2v(pin, spc->p);
            expu += exp(-u); // widom average }  }  }
```

Bibliography

1. Jönsson, P. *et al.* Hydrodynamic trapping of molecules in lipid bilayers. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 10328–33 (2012).

2. Waals, J. *Over de continuiteit van den gas- en vloeistoftoestand.* (A. W. Sijthoff, 1873). at <https://www.lorentz.leidenuniv.nl/history/proefschriften/sources/vanderWaals_1873.pdf>

3. Hemmer, P. C. Virial Coefficients for the Hard-Core Gas in Two Dimensions. *The Journal of Chemical Physics* **42**, 1116 (1965).

4. Hu, J. & Yu, Y.-X. High-order virial coefficients and equation of state for hard sphere and hard disk systems. *Phys. Chem. Chem. Phys.* **11**, 9382–9390 (2009).

5. Stenqvist, B., Thuresson, A., Kurut, A., Vácha, R. & Lund, M. Faunus – a flexible framework for Monte Carlo simulation. *Molecular Simulation* **39**, 1233–1239 (2013).

6. Metropolis, N. & Ulam, S. The Monte Carlo Method. *Journal of the American Statistical Association* **44**, 335–341 (1949).

7. Widom, B. Some topics in the theory of fluids. *The Journal of Chemical Physics* **39**, 2808–2812 (1963).

8. Hill, T. *An introduction to statistical thermodynamics.* (Dover Publications, 1960). at <https://books.google.nl/books?id=pX4yxOHnWg8C>

9. Burley, S. K. *et al.* Protein data bank (pdb): The single global macromolecular structure archive. *Methods in molecular biology (Clifton, N.J.)* **1607**, 627—641 (2017).

10. UniProt: The universal protein knowledgebase. *Nucleic Acids Research* **45**, D158–D169 (2017).

11. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: An automated pipeline for the setup of poisson–boltzmann electrostatics calculations. *Nucleic Acids Research* **32**, W665–W667 (2004).

12. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings*

of the National Academy of Sciences **98**, 10037–10041 (2001).

13. Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).

14. Polley, A. *et al.* Glycosylation and lipids working in concert direct cd2 ecto-domain orientation and presentation. *The Journal of Physical Chemistry Letters* **8**, 1060–1066 (2017).