Study of water dynamics around ATP

Alexandra Krause

May 26, 2017

Bachelor thesis

Freie Universität Berlin

Department of Physics

Matrikel number : 4681990

Field of study : Physics mono bachelor

First supervisor: Prof. Dr. Ana-Nicoleta Bondar

Second supervisor: Prof. Dr. Johannes Reuther

Contents

1 Introduction

Adenosine triphosphate, ATP for short, is the universal energetic currency in living things ranging from plants and bacteria up to humans. It is used in the metabolism to provide energy for ion pumping processes, transport energy to mitochondria and as a power source for enzyme folding. [3] The principal reaction providing this energy is ATP hydrolysis during which one or two phosphates are split from ATP. Hydration of phosphate then gives an enthalpy of $\frac{-30.5kJ}{mol}$ per split phosphate. [5]

$$ATP + H_2O \Longleftrightarrow ADP + P \tag{1}$$

$$\Delta G = -30.5\frac{kJ}{mol} \tag{2}$$

With a net charge of -4 in solution, ATP is highly negative. This charge can polarize the water molecules around ATP in the cells and change the orientation of dipole moments as well as allow for hydrogen bonding between phosphate and oxygen atoms in ATP, for example on the phosphate groups. The following work tries to evaluate the range of induced polarization of water around ATP as well as the orientation of dipole moments through a numerical simulation of ATP in water. For this purpose I wrote with the support of Professor Bondar short simulations of ATP in a water box of 50 Angstrom using water models TIP3P and SPC/E and comparing the results from both simulations.

2 Theory

2.1 Water Models

Water is one of the most complex systems to be studied in theoretical biophysics. For that purpose, researchers created different water models to be used in molecular dynamics simulations. Unfortunately, no water model is able to correctly reproduce all physical properties of real water so that the choice of the water model is a critical factor for the success in obtaining usable results from a simulation. Three categories of water models can be distinguished : rigid models such as TIP3P or SPC/E and flexible models. Rigid models have the advantage of properly reproducing the real geometry of water and facilitating calculations by reducing the degrees of freedom in the system. That reduction comes from introducing a constraint to the system : The bond lengths are kept constant. Numerically this is done with the SHAKE algorithm introducing a virtual bond between both hydrogen atoms. While a normal water molecule has 9 degrees of freedom without constraints, with 3 rigid bonds this number is reduced to 6. In flexible water models the bonds are approximated to behave like an oscillator, allowing the water molecule to have vibrational degrees of freedom (angle bending and bond stretching). They correctly reproduce vibrational spectra. [8] [13] [12]. In the following work I will use the standard model for water in MD-analysis, the TIP3P model. It will then be compared to the SPC/E (extended simple point charge) water model which better displays water properties of the second hydration shell. This can be seen in a comparison for the radial distribution function between water oxygens where SPC/E correctly displays the characteristic second peak belonging to that hydration shell while TIP3P does not. [13] I will later verify this property in Figure 17.

2.2 Molecular Dynamics Simulation

Having mainly huge systems with more than 1000 atoms, most systems of interest in biophysics cannot be treated with pure quantum mechanics. For this purpose they are

mainly treated classically by solving the Newtonian equations of motion. For a system with N particles these are

$$m_i \frac{dr_i}{dt} = -\nabla_i V \tag{3}$$

with $i \in [1, N]$, m_i being the mass of particle i and V the potential of the whole system. The potential is dependent of the force field being used, in my case CHARMm. During the simulation the equations of motion are numerically integrated to obtain the trajectories of all atoms. The simulation begins with the initial coordinate structure without kinetic energy in the atoms, corresponding to 0 K. During the heating period, in every step every atom receives a short amount of kinetic energy followed by an equilibration phase of some picoseconds until the desired temperature (in my case 300 K) is reached. After the heating period follows a short equilibration period before data can be collected in the production phase.

2.2.1 Force Fields

Most water models were originally developed for a specific force field, e.g. SPC and SPC/E for GROMOS and TIP3P for AMBER, later modified to work well with CHARMm. [8] The force field contains the bounded potentials within the molecule as well as those between different molecules, also called non-bounded or unbounded potential terms. The total potential V can be written as

$$V = V_b + V_\phi + V_\theta + V_\omega + V_{vdW} + V_{el} \tag{4}$$

Here, V_b, V_ϕ, V_θ and V_ω are the bounded potential terms and V_{vdW} and V_{el} are the non-bounded ones. [1]

$$V_b = \sum_{bonds} k_b(l_b - l_0) \tag{5}$$

V_b is the potential term from the bond stretching with a force constant k_b and the bond length difference $l_b - l_0$ indicating the deflection from equilibrium. [1]

$$V_\phi = \sum_{angles} k_\phi(\phi - \phi_0) \tag{6}$$

V_ϕ collects the potential terms of the angle stretching analogous to V_b. k_ϕ again denotes the force constant and $\phi - \phi_0$ the change of the angle with respect to the equilibrium. [1]

$$V_\theta = \sum_{proper-dihedrals} k_\theta[1 + cos(n\theta - \alpha)] \tag{7}$$

$$V_\omega = \sum_{improper-dihedrals} k_\omega(\omega - \omega_0)^2 \tag{8}$$

Equation 7 and Equation 8 include the potential energy from proper and improper dihedrals (torsion angle and out-of-plane bending) with k_θ and k_ω being the respective force constants, $\omega - \omega_0$ being the angular difference with respect to the equilibrium for the out-of-plane-bending, α the phase shift of the torsion, n the multiplicity of the torsion function and θ the torsion angle. [1]

$$V_{vdW} = \sum_{i,j|i<j} 4\epsilon \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right] \tag{9}$$

The Van-de-Waal-forces are modeled by the Lennard-Jones potential, a polynominal potential for uncharged atoms and molecules which combines attracting and repelling forces. Indices i and j count over the atoms in the sum, $r_{i,j}$ is defined as the modulus of the position vectors ($\vec{r_i}$) and ($\vec{r_j}$) for atoms i and j. ϵ denotes the y-coordinate of the minimum in the potential and $\sigma_{i,j}$ is the x-coordinate of the position where $V_{vdW} = 0$. [1]

$$V_{el} = \sum_{i,j|i<j} \frac{q_i q_j}{4\pi\epsilon_0 r_{i,j}} \tag{10}$$

The strongest term in the potential is V_{el}, the electrostatic Coulomb interaction for atoms i and j with charges q_i and q_j and their position vector difference $r_{i,j}$ as in the Lennard-Jones potential. ϵ_0 denotes the dielectric constant.

2.2.2 Particle Mesh Ewald Method

A possible way to calculate the Coulomb potential is to use the Particle Mesh Ewald Method (PME) which splits the potential V_{el} into two terms, $V_{fourier}$ which converges fast in the reciprocal space and V_{real} which converges fast in real space. To change the coordinates of the potential part $V_{fourier}$ into the reciprocal space, the algorithm uses the Fast Fourier Transform and then calculates the part in the reciprocal space. After that, coordinates are changed back to the real space by the inverse Fourier transform. This allows for the algorithm to decrease the computational effort from $O(n^2)$ to $O(nlog(n))$. [14] For the application of PME with simultaneous use of boundary conditions the water box has to be neutral. This is done by putting counter ions into the solution, normally Magnesium for ATP, as magnesium ions are the ions mostly found in nature in a ligand bonding with ATP. [20]

2.2.3 NVE And NPT

Molecular dynamics simulations can be run using either the isothermal-isobaric (NPT) or the canonical (NVE) ensemble. NVE means that the total number of atoms in the system N the volume V and the total energy are kept constant while the temperature can vary. In an NPT ensemble the system is coupled to an external energy source, mostly the Langevin thermal bath to keep the temperature constant. In that case the energy will not remain constant. Also the volume is not kept constant, but the pressure P is. [18] As the coupling to the thermal bath can distort water properties, it is best to perform a NVE simulation to study water dynamics around a molecule. [17]

2.2.4 Periodic Boundary Conditions

When studying a closed system undesired surface effects will disturb the dynamics of the molecules within the system. This can be prevented by using periodic boundary conditions. In that case the water box is infinitely often replicated in all three dimensions. If a water molecule leaves the system on one side of the water box, its image within the other boxes moves exactly in the same way, causing the water molecule to reenter on the opposite side. Therefore no surface effects will occur. [19]

3 Simulations

3.1 Finding The Right PDB File

The *.pdb*-file contains the initial positions of all atoms in the system as well as information about the bonds between them. Usually, these files can be obtained from the webpage www.rcsb.org, which contains a database of most molecules used in biophysics simulations. As ATP is too small to be classified on its own, it is only contained as a ligand in files. To obtain a *.pdb*-file for ATP, I searched in the database for proteins having ATP as a ligand and found the structure $4x35$ from [7], a viral protein with a crystal structure resolution of 1.5 Angstrom.

3.2 Errors During The Simulation

The first problem encountered was a problem with the psf-file generated by the pdb reader which could not be read by NAMD, the software used for integrating the equations of motion. To circumvent that problem, I used the VMD software (Visual Molecular Dynamics) to generate both the water box around ATP with 4 sodium ions for neutralization as well as the $.psf$-file belonging to that water box with ATP. Also this $.psf$-file could not be read giving the error that a certain atom type ("P3") could not be recognized. A solution was to separate the water box as well as the ions from the ATP using VMD and reunite them using a CHARMm script which would also create a new $.psf$. This made the simulation run but giving physically unrealistic results. Most angles in the ATP molecule such as the angle between the phosphate and the two oxygen atoms in the phosphate groups were unrealistically small - which would not occur like that in reality as the oxygens have a partial negative charge which normally leads to a rejecting force and therefore to a widening of the angle.

3.3 Finally Used Simulations

Taking into account these problems and the time it would take to fix them, Professor Bondar finally prepared the TIP3P simulation with three runs providing three trajectories to be used by me for the analysis. The simulation was prepared with the CHARMm software using the actual version of the CHARMm forcefield v. 36. For the first two runs the system was coupled to a Langevin thermal bath to equilibrate the system at a constant temperature. During the third run the coupling was turned of for the NVE production phase. The system was neutralized using two magnesium ions. The simulation using the SPC/E model was done with the NAMD [4] software and the pdb as well as psf files from the previous simulation. Furthermore here the CHARMm forcefield v. 36 was used but due to missing data for the SPC/E water in the forcefield those data was added to the parameter files using parameters from [13]. Details for the three simulation runs can be found in the following table :

Table 1: Details for the simulation run

Run Number	Time step duration	Number of time steps	total duration	NPT or NVE
1	1 ps	100	100 ps	NPT
2	1 ps	1000	1 ns	NPT
3	10 fs	100000	1 ns	NVE

All analysis was done on all three trajectories except for the calculation of dipole moment projection. As a higher number of frames delivers better statistics results presented

in the following work will all be derived from the third trajectory.

4 Analysis And Discussion Of Results

Data analysis was performed using tcl-scripts to be run with VMD. The data obtained was then analysed using Python together with specific libaries to treat scientific data (numpy and scipy) and to create plots (pylab.pyplot). Pictures of molecules from the simulation were created with the Visual Molecular Dynamics (VMD) software [9]. For all data analysis PBC wrapping was included, except in the case for calculating dihedral angles over time. In that case no differences occured between results with prior PBC wrapping and without, as only the internal degrees of freedom from the ATP molecule were concerned. Therefore PBC wrapping was unnecessary. As PBC wrapping increased the computational effort, it was left out in this case. During the PBC wrapping the coordinates of all molecules of the water box were recentered around a chosen center, in this case the ATP molecule. This prevented them from leaving the water box which would have given wrong data with the trajectory progressing. This was accomplished using the PBCTools Plugin v. 2.7 with the ATP molecule as center.

4.1 RMSD Evolution Of ATP Over Time

A stable RMSD (Root mean square derivation) value is an indicator for the convergence of the simulation. The RMSD of a molecule as a function of time can be calculated as

$$RMSD(A,B) = \sqrt{\frac{1}{n}\sum_{i\in[1,n]} (a_{i,x}-b_{i,x})^2 + (a_{i,y}-b_{i,y})^2 + (a_{i,z}-b_{i,z})^2} \tag{11}$$

with

$$a_i \in A, b_i \in B \tag{12}$$

where A denotes the set of the initial atom positions determined from the *pdb* file and B contains the atom positions at time t which are read from the *dcd* file. This equation can be derived from the expression for the MSD of a molecule found in [16] which appears there in the context of the self diffusion coefficient of a liquid, in that case water.

The RMSD was first measured with the VMD tool "RMSD Visualizor" giving very high rates for the RMSD. This was due to a missing realignment of the ATP to account for the rotational and translational degrees of freedom. Usually we want to know the deviation from the initial conformation of the molecule (in my case ATP), so that we have to account for the translation of the molecule. Alignment of the ATP does that by moving the molecule back to its initial position after each frame before calculating the RMSD. To prevent that error, I used a script written by Suliman Adams in TCL which also used VMD to calculate the RMSD. This script took care of the realignment of the structure and gave smaller values between 1 Angstrom and 1.8 Angstrom for the first analyzed trajectory, approximately 0.7 Angstrom to 2.7 Angstrom for the second trajectory and between 0.7 - 2.8 Angstrom for the third trajectory. RMSD values can be seen as a function of time in Figure 1 and Figure 2. For better estimation of which parts of ATP contribute to the RMSD values I then also computed the RMSD values for the adenine as well as for the phosphate groups. The blue plot shows the RMSD evolution over time for the whole ATP molecule, green denotes the phosphate group RMSD and red was chosen for the adenine. One can see that the RMSD value for the phosphate groups remains relatively stable during all three simulations with fluctuations of approximately 0.3 Angstrom while the

adenine shows sudden jumps between 0.1 Angstrom and 0.6 Angstrom within the second and third trajectory. The first trajectory is too short for a jump to occur.

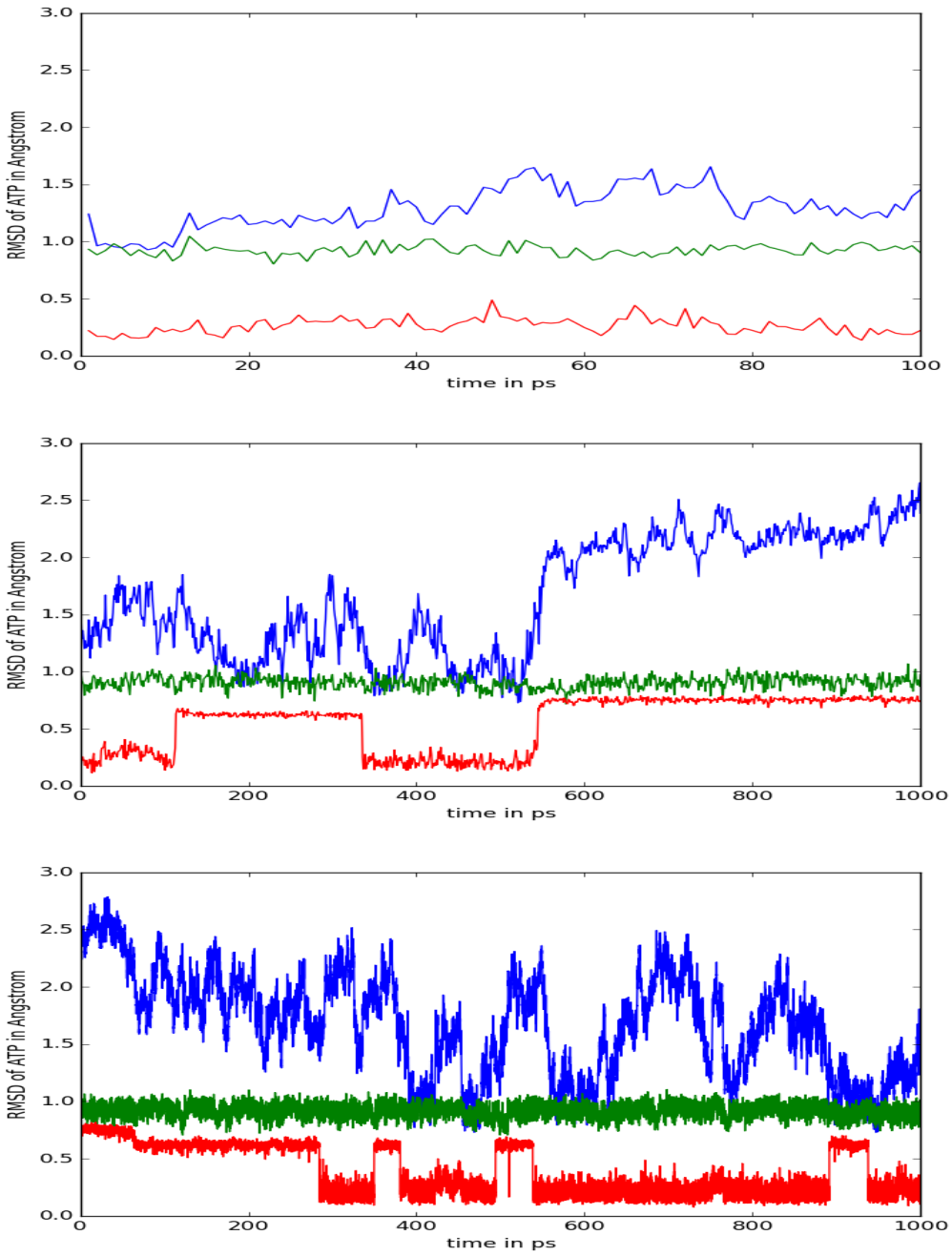

Figure 1: RMSD evolution over time for all three trajectories, TIP3P water model

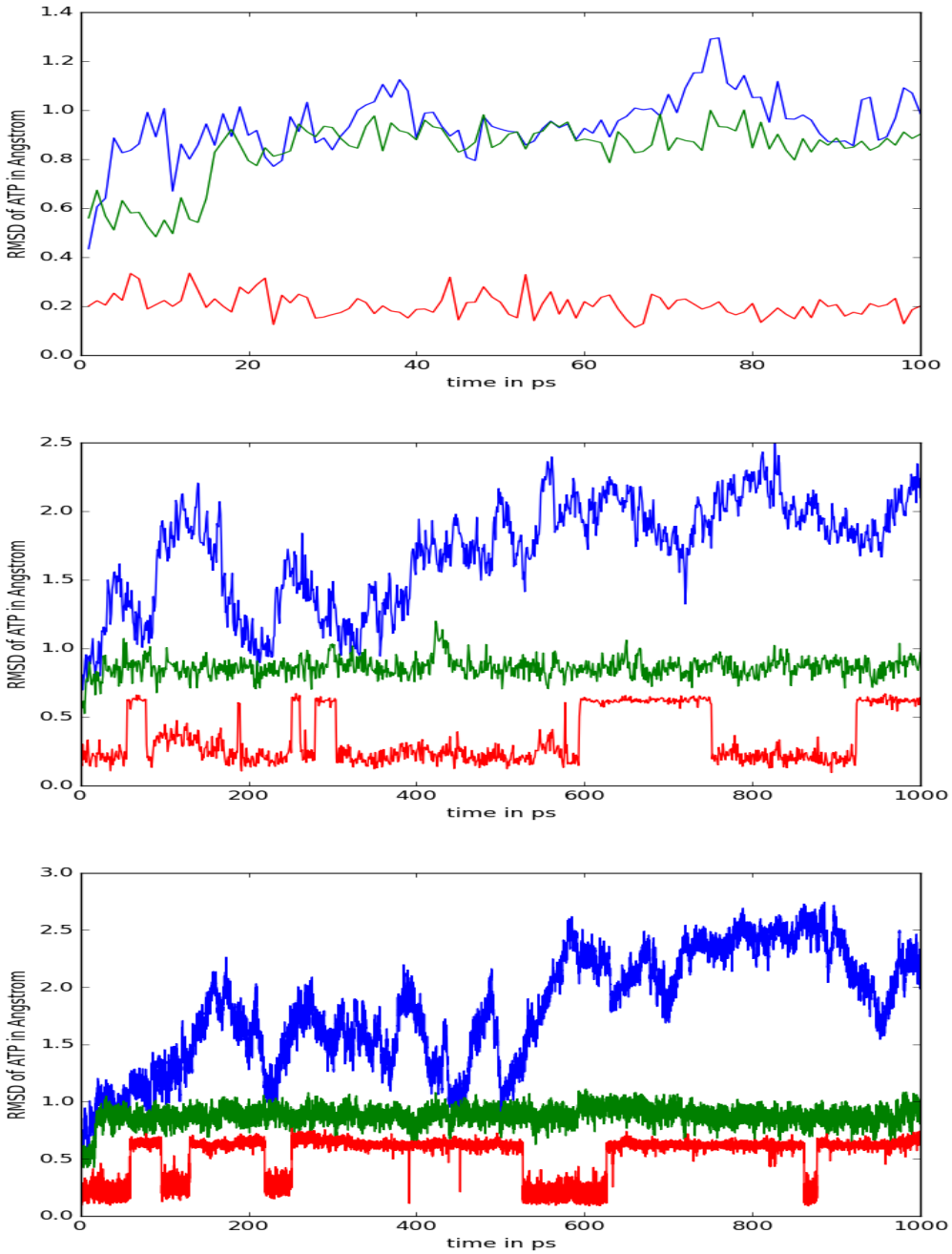

Figure 2: RMSD evolution over time for all three trajectories, SPC/E water model

A possible reason for those differences can be seen in the following picture showing water molecules in a distance of 1.9 Angstrom around ATP. As the phosphate groups form a long chain they have six rotational degrees of freedom and will therefore achieve a higher value for overall motions. As they also attract a lot of water molecules coming close forming a hydrogen bonded network, the motion is stabilized by the network and remains stable during the simulation. On the other hand the adenine has only two rotational degrees of freedom. One of them, specifically the rotation of the NH2 group, can be neglected because the RMSD value of two hydrogen atoms is tiny compared to the motion of the atoms in the nucleobase. The atoms in the nucleobase all lie within one plane with

motions being restricted by bonds as constraints. Therefore high energy levels are needed to shift the nucleobase from one configuration to another.

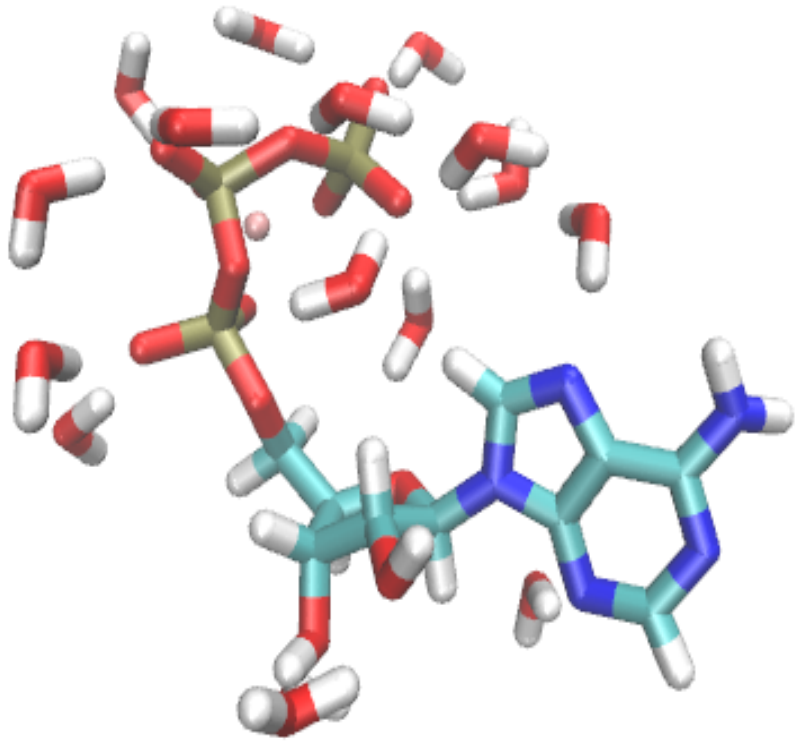

Figure 3: Water molecules in a distance of 1.9 Angstrom around ATP

4.2 Dihedral Angles

Having four atoms, one can always arrange three of them in a plane and thereby obtain two planes. The angle between those two planes is called dihedral angle. Its time development can be an indicator for how two parts of a molecule move relatively to each other. In the case of ATP, the dihedral angle involving molecules from the phosphate group and the sugar base can describe, how the phosphates move relative to the sugar base. The same can be done for atoms linking adenine to the sugar base. The observed dihedral angles can be seen in Figure 4 with bonds connecting the atoms belonging to the dihedral angle linking phosphate groups to the sugar base (PDA) highlighted in green, and those connecting the adenine to the sugar base (ADA) highlighted in black.

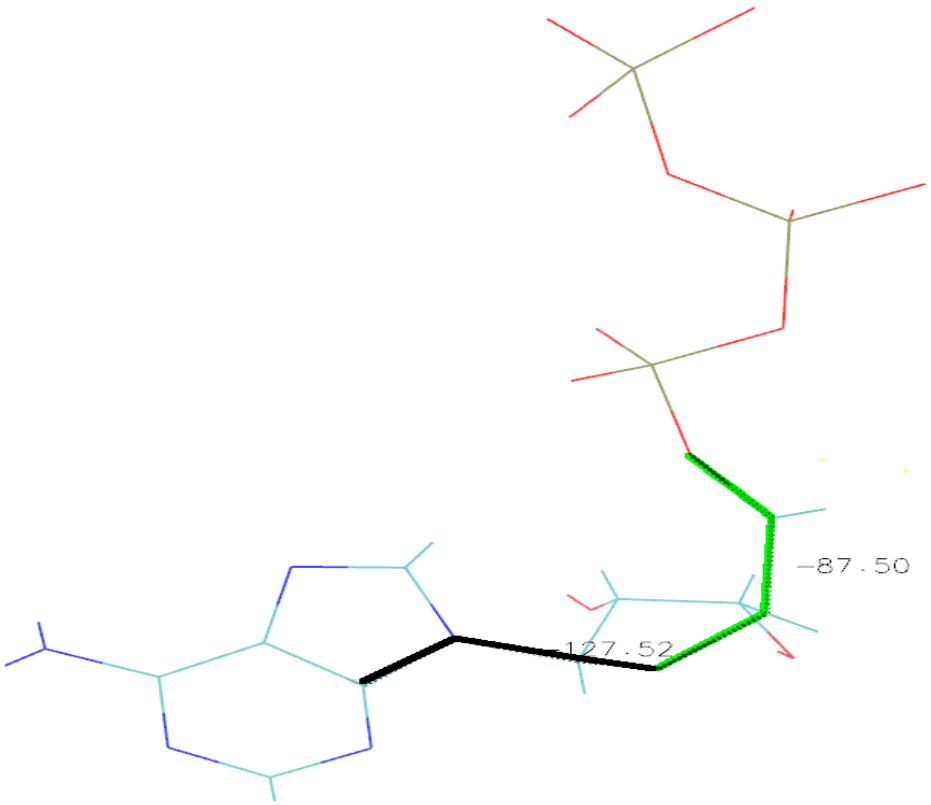

Figure 4: Bonds connecting atoms chosen for the calculation of the dihedral angles

Grouping the so obtained angles in a histogram then allowed me to determine the mean force potential according to Equation 13 where y describes the counts of the respective angle and y_{max} the counts of the angle appearing the most often. The mean force potential describes the potential belonging to a force averaged over all possible configurations of the system along a specific reaction coordinate [15], in my case the torsion around the dihedral angle.

$$PMF = -k_b T \log_e \left(\frac{y}{y_{max}} \right) \tag{13}$$

The dihedral angles were measured with the VMD software. After the import of pdb, psf and trajectory files, determination of the atoms involved in the dihedral angles was necessary. Using the "Pick atom" function in VMD, determination of the number of those atoms was possible. The command *set dihedlist* [*measure dihed* $atom_1$ $atom_2$ $atom_3$ $atom4$ *frame all*] gave a list including all dihedral angles over time. Here $atom_i$ denotes the number of the respective atom. Dihedral angles were tracked over time and ordered in a histogram that was used to calculate the mean force potential for each angle. As the sampling rate was not sufficiently high around the whole spectrum from -180 to 180 degrees, the mean force potential was only calculated around the peaks of the histogram where enough data could be gathered.

4.2.1 Phosphate Linking To Sugar Base (PDA)

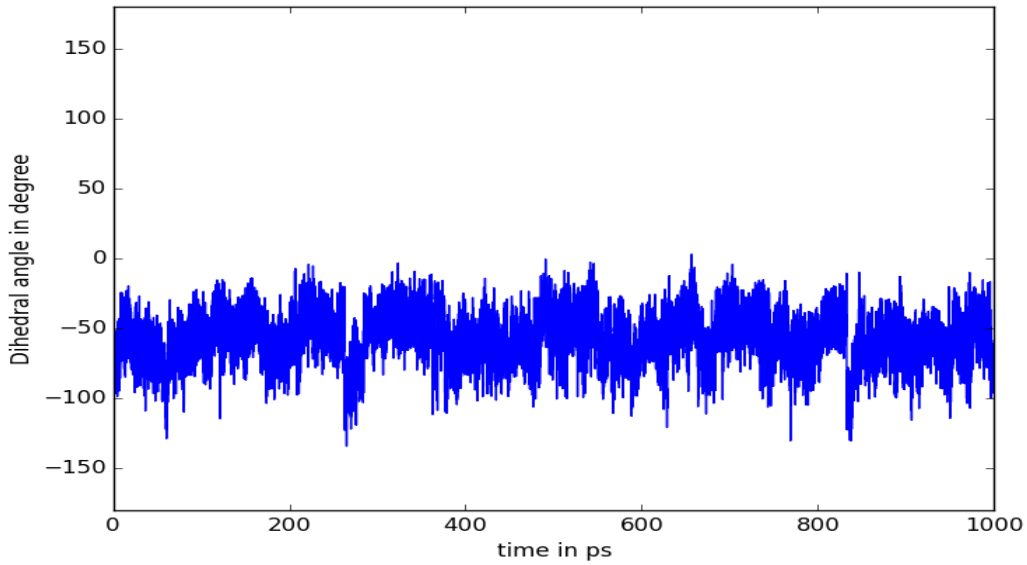

Figure 5: Time evolution for the dihedral angle, TIP3P model

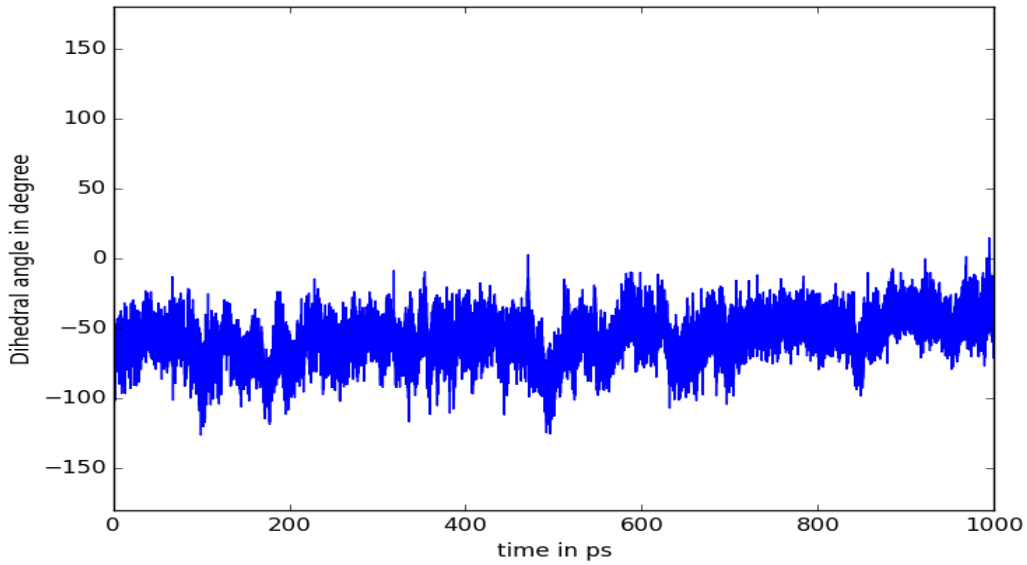

Figure 6: Time evolution for the dihedral angle, SPC/E model

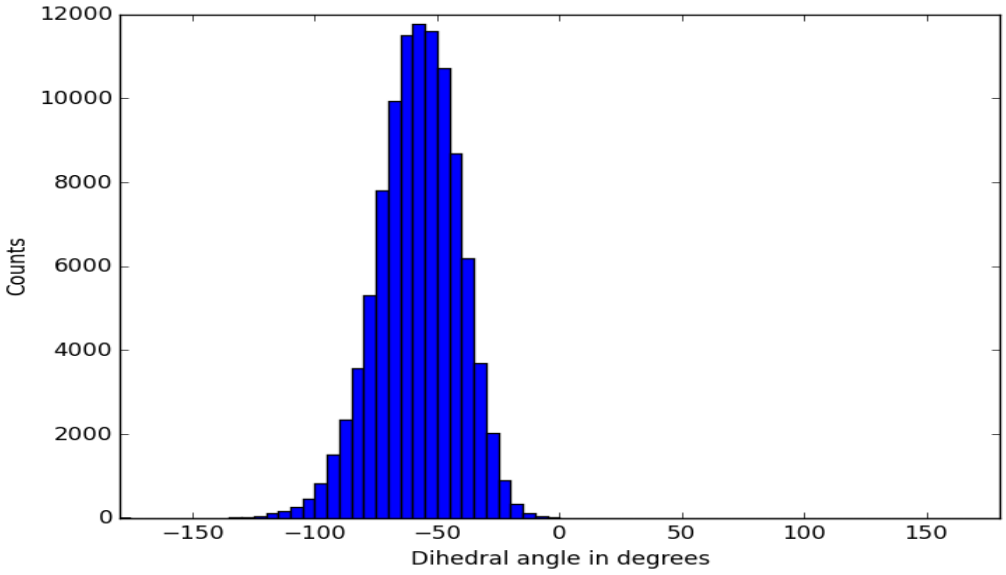

Figure 7: Histogram for the dihedral angle; TIP3P model

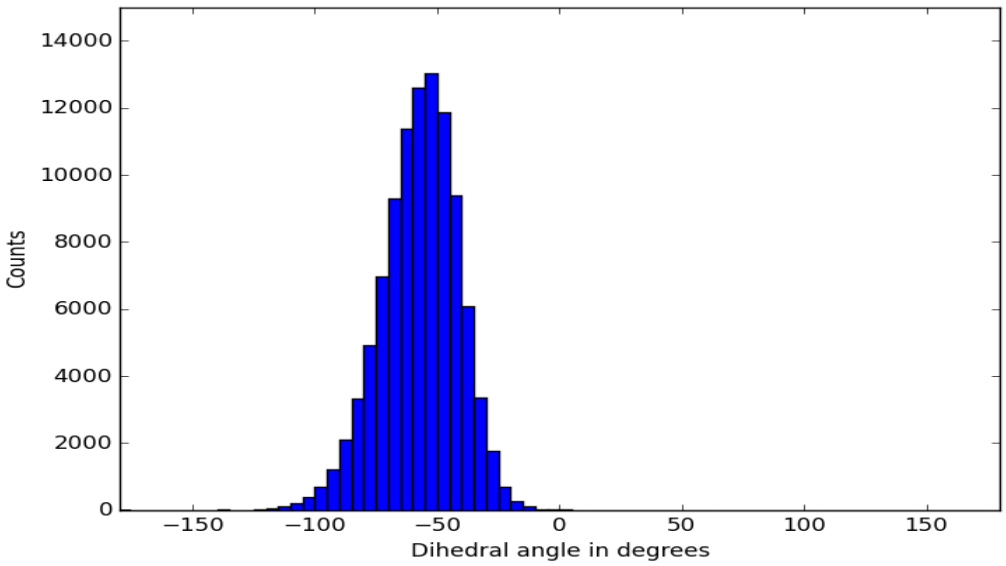

Figure 8: Histogram for the dihedral angle; SPC/E model

Both histograms have the same Gaussian shape which is characteristic for data obtained from statistical processes. Standard deviation and arithmetic mean are nearly identical, which can also be seen in Figure 9. The minimum of the potential equals the peak in the Gaussian distribution from the histogram. For both water models this peak is at approximately -59 degrees. It seems like changing the parameters for SPC/E in the CHARMm forcefield did not affect this dihedral angle much.

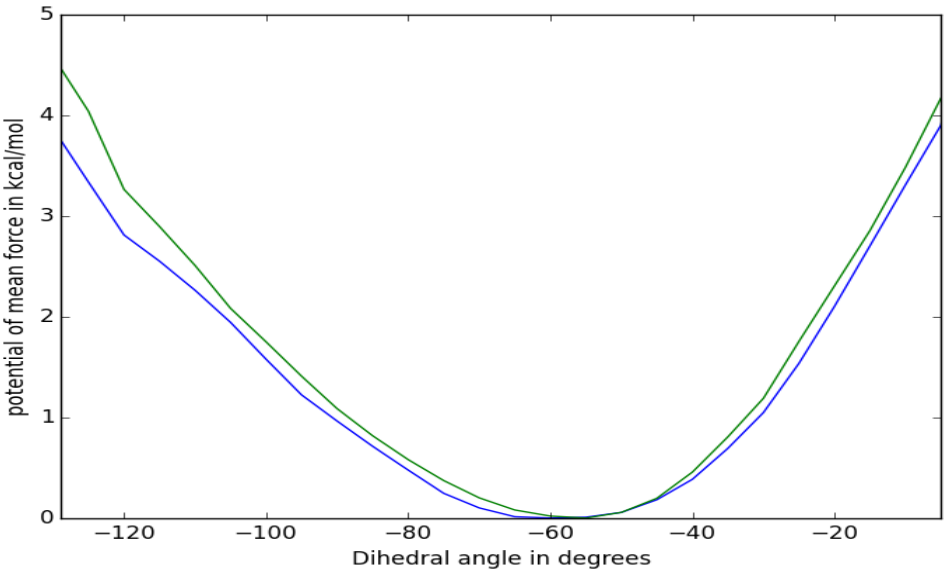

Figure 9: Mean force potential for both water models in comparison (blue : TIP3P potential, green: SPC/E potential)

For the mean force potential, both potential functions look like a square function being described by the form $a(x-m)^2$. The plot shows even more precisely than for the Gaussian distribution that the average dihedral angle is the same. The coefficient a seems to be almost the same. With a bigger data set allowing for better convergence, one would not notice that difference.

4.2.2 Sugar Base Linking To Adenine (ADA)

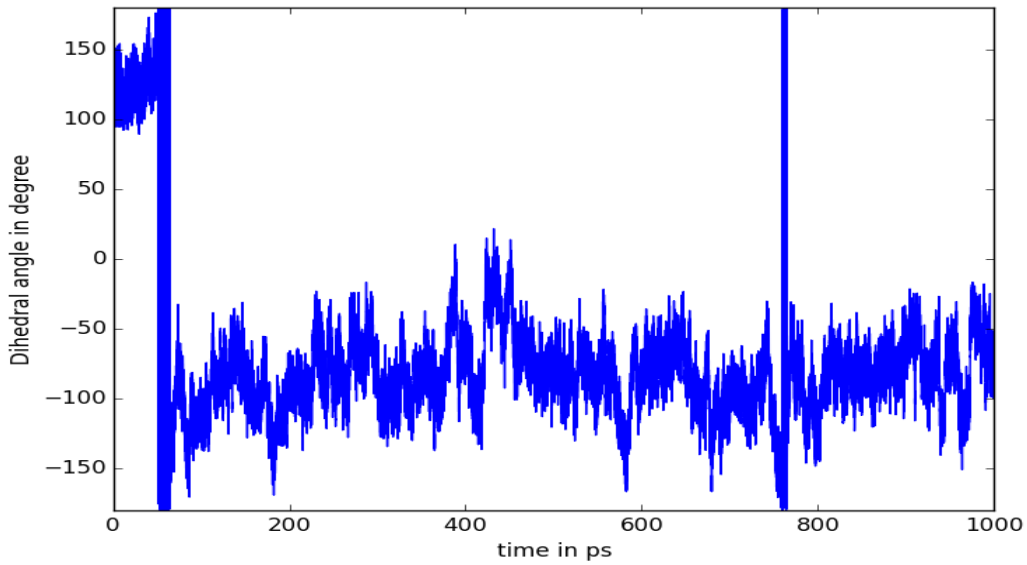

Figure 10: Time evolution for the dihedral angle, TIP3P model

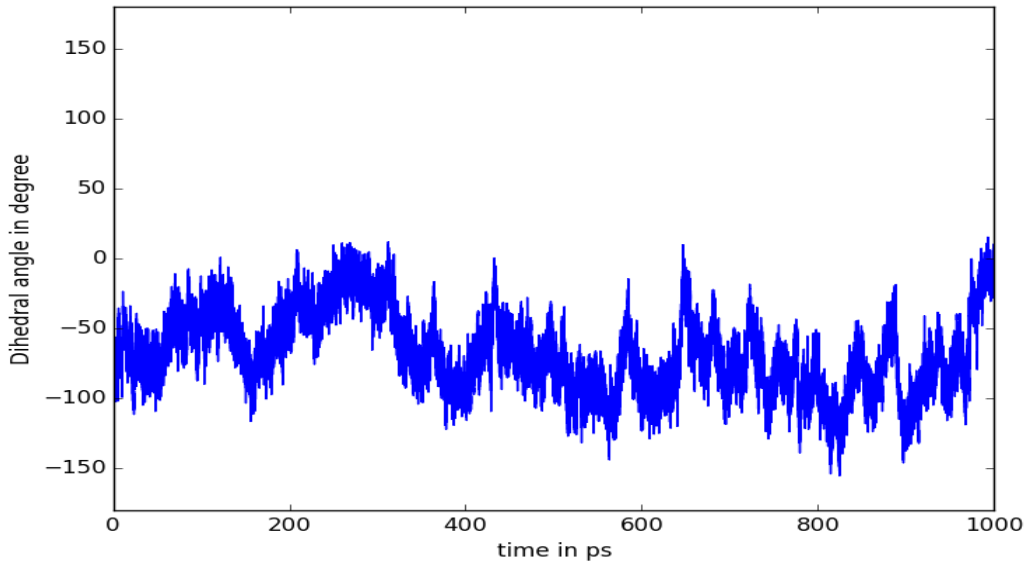

Figure 11: Time evolution for the dihedral angle, SPC/E model

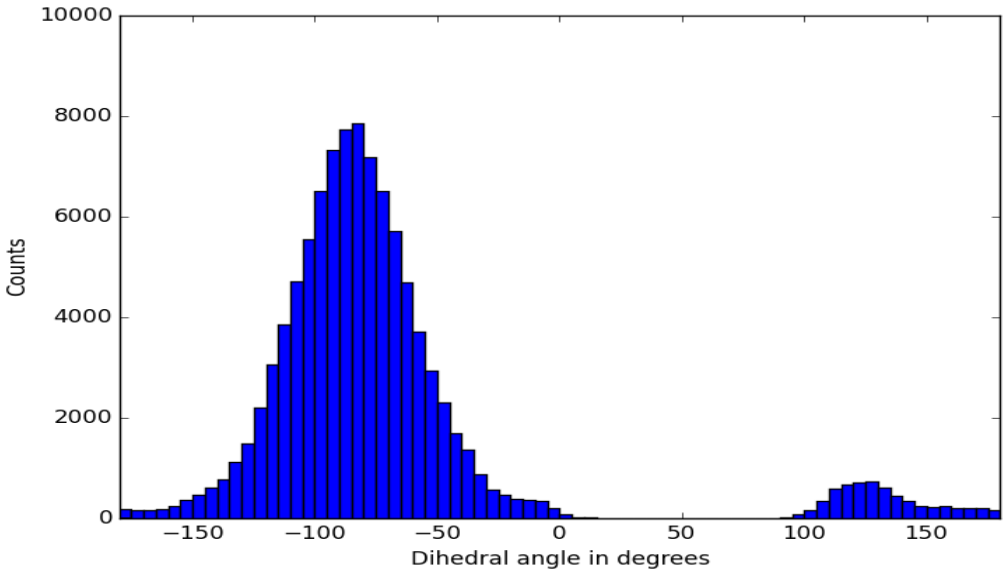

Figure 12: Histogram for the dihedral angle; TIP3P model

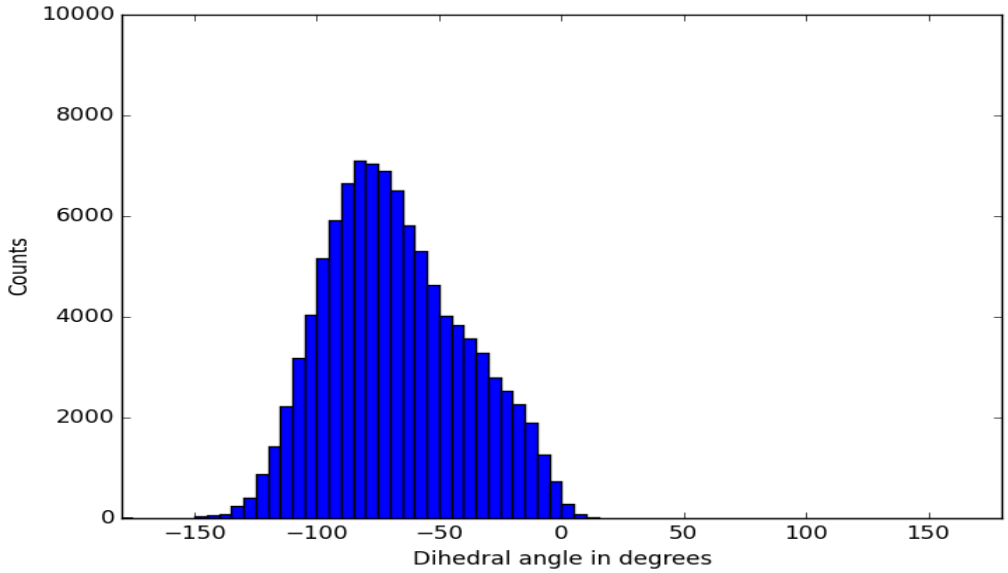

Figure 13: Histogram for the dihedral angle; SPC/E model

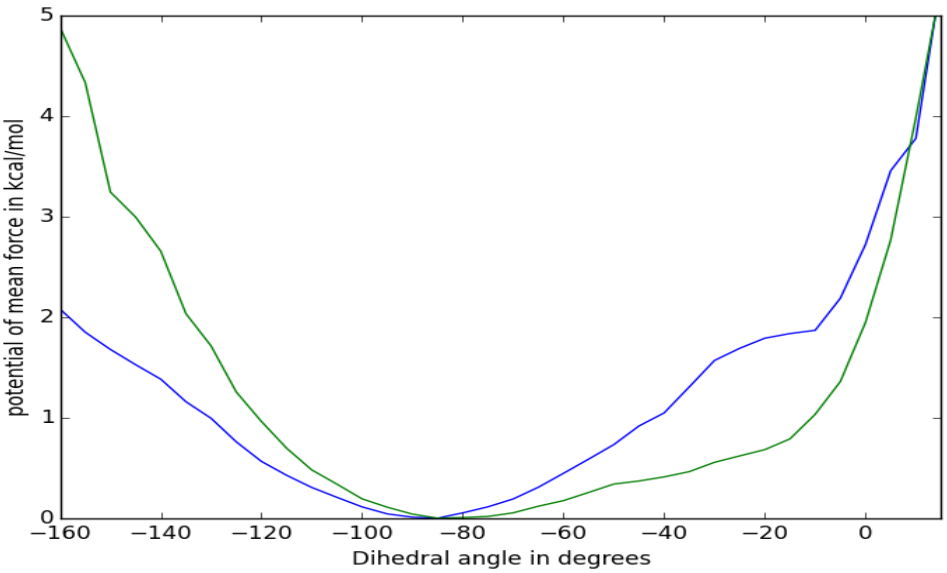

Figure 14: Mean force potential for both water models in comparison (blue: TIP3P potential, green : SPC/E potential)

Here things are quite different. While for the phosphate-sugarbase dihedral angle no significant differences can be seen, for the adenine-sugarbase dihedral angle the TIP3P histogram gives a bimodal shape with a big peak around -80 degrees and a small peak at approximately 125 degrees. The big peak can also be found at approximately -80 degrees using the SPC/E water model within the simulation. But in that case an unimodal shape with a bulge at the right side is found. This affects the calculation of the mean force potential. For the bimodal distribution, to gain precision I removed the data of the second

peak to improve the comparison of both potentials. Averages and square functions for both water models do not match well. A possible reason might be the disturbance caused by strong water fluctuation around the molecule. While the phosphate groups are kept stable by the water network this does not apply for the adenine. A calculation of successive averages in time could have been more appropriate to obtain an equilibrium distribution but was beyond the scope of this bachelor thesis.

4.3 Number Of Water Molecules Bonding To ATP

The number of water molecules bonding to ATP was calculated by using the number of hydrogen bonds around the molecule. Using the VMD plugin HBonds v. 1.2 written by J.C. Gumbart and D. Luo [11], I was able to track the average number of hydrogen bonds to the oxygen atoms of the phosphate groups. As a comparison the average number of hydrogen bonds to the oxygen atoms in the sugar base ring was taken. Denoting with D the donator atom and the acceptor atom with A, a hydrogen bond between D and A occurs if and only if the distance between D and A is inferior to a cutoff distance of 3 Å and the angle between the atoms $\measuredangle(D, A) \leq 30°$. The average number of hydrogen bonds for an oxygen atom in bulk water lies between 1.5 and 2. Similar numbers can therefore be expected for the oxygen atoms belonging to the phosphate groups. Every phosphate group except for the terminal γ phosphate group has two oxygen atoms exposed to the hydration shell as well as two oxygen atoms connecting two phosphate groups. The γ phosphate group is the only group with three oxygen atoms exposed to the hydration shell and will therefore bond to more water molecules in average. As for three oxygen atoms exposed to the hydration shell the average number of hydrogen bonds was significantly lower than for others, I also measured the average distance to the magnesium ion close to ATP which was one of two magnesium ions used for neutralization. Those difference can be explained by ligand bonding of the three oxygens to the magnesium ion. Here a distance of 1.9 Angstroms was defined as the ligand bonding distance. The obtained numbers can be seen in table Table 3 for the atom names, the letter O stands for the element oxygen, letters A, B and G denote the phosphate group the oxygen atoms belong to (A for the α group linking to the sugar base, B for the β group and G for the terminal γ group). Oxygens with the number 3 are those linking two groups and are less exposed to the hydration shell. Therefore they are less likely to bond to water molecules. The same is true for Oxygen $O5'$ connecting the α phosphate group with the sugar base as well as $O4'$ the oxygen atom in the sugar base ring taken as a reference.

Table 2: Comparison TIP3P and SPCE for H-bond analysis

Atom	Bonds (TIP3P)	Bonds(SPC/E)	SV	distance MG TIP3P	distance MG SPCE	SV
O1A	1.6 ± 0.7	1.7 ± 0.7	yes	4.3 ± 0.1	3.9 ± 0.1	no
O2A	0.1 ± 0.1	0.2 ± 0.4	yes	1.9 ± 0.1	1.9 ± 0.1	yes
O3A	0.1 ± 0.3	0.1 ± 0.3	yes	3.8 ± 0.2	3.5 ± 0.1	yes
O1B	1.5 ± 0.8	0.4 ± 0.5	no	3.9 ± 0.2	1.9 ± 0.1	no
O2B	1.9 ± 0.8	1.9 ± 0.7	yes	5.3 ± 0.1	4.4 ± 0.1	no
O3B	0.4 ± 0.5	0.6 ± 0.6	yes	3.5 ± 0.1	3.5 ± 0.2	yes
O1G	2.6 ± 0.8	2.8 ± 0.7	yes	3.7 ± 0.1	3.7 ± 0.2	yes
O2G	1.1 ± 0.6	2.8 ± 0.7	no	1.9 ± 0.1	4.3 ± 0.1	no
O3G	0.8 ± 0.5	0.9 ± 0.7	yes	1.9 ± 0.1	1.9 ± 0.1	yes
O5'	0.1 ± 0.4	0.2 ± 0.4	yes	-	-	-
O4'	0.2 ± 0.3	0.1 ± 0.3	yes	-	-	-

Table 3: Average number of hydrogen bonds to oxygen atoms and standard deviations for TIP3P and SPC/E water, SV indicates whether the values are similar for both water models or not. Values are defined as similar, when their standard deviations overlap.

It can be seen that the number of hydrogen bonds differs strongly for the oxygen atoms. This can be explained by structure and by the distance to the magnesium atoms. For atoms between two functional groups (O3A, O3B, O5') or within a ring (O4' in the ring of the sugar base) the average number of hydrogen bonds is lower than the average for oxygens of bulk water. This can be explained by structure as those oxygen atoms are less exposed to the hydration shell. For atoms exposed to the hydration shell one expects values similar to those of bulk water. This is true for the atoms of the α and β phosphate groups except for O2A for the simulation using TIP3P model and atoms O2A and O1B for the SPC/E simulation. Those atoms are in a distance of 1.9 Angstrom to the magnesium ion, a distance characteristic for ligand bonding. [6] This ligand bond between the magnesium ion and the oxygen prevents further intermolecular bonding to the oxygens and therefore explains the low average number of hydrogen bonds. For the terminal γ phosphate group, average values between 2.6 and 2.8 Angstrom are higher than the average for bulk water. The only exception are atoms O2G and O3G for the TIP3P simulation and O3G for the SPC/E simulation. Those atoms are again ligand bound to magnesium which explains the lower number of hydrogen bonds. Standard deviations for the average number of hydrogen bonds are very high thus indicating bad statistics. For statistical analysis, a possible way to go further and gain precision would be to use the box plot. It allows to see not only the average, but also median and quartiles on the same plot and to understand where most data points are. This provides for a better understanding about the distribution of data far around the average.

4.4 Radial Distribution Function Of Water Around ATP

The radial distribution function gives information about the structure of liquid solutions. Given an initial position of an atom r_0 it denotes the probability of finding other atoms within a distance r around this atoms. If both atoms are of the same type, e. g. oxygen atoms in a solution of water, we call it the radial distribution g_{OO}. For atoms of different types or belonging to different types of molecules it is called the pair distribution function. Figure 15 explains how the radial distribution function can be calculated. From the initial position one can look within a spherical shell with radius r and thickness dr whether there are other atoms of the desired type present and count those. This provides a histogram with distance bins of length dr from r_0 to a chosen maximal value r_{max}. Averaging the counts over the whole simulation then gives the radial distribution function $g(r)$. [10]

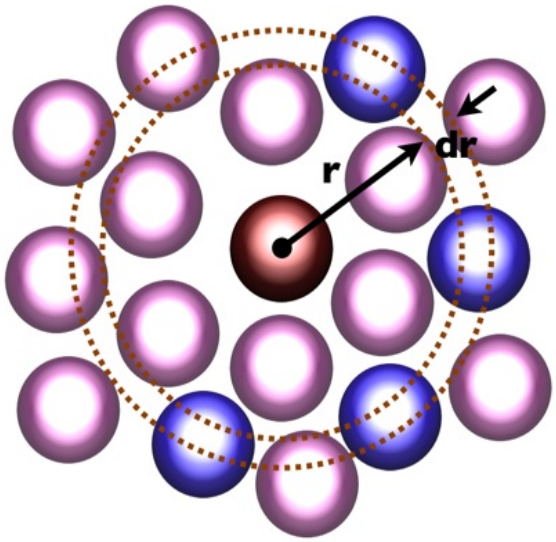

Figure 15: Schematic explaining the radial distribution function [2]

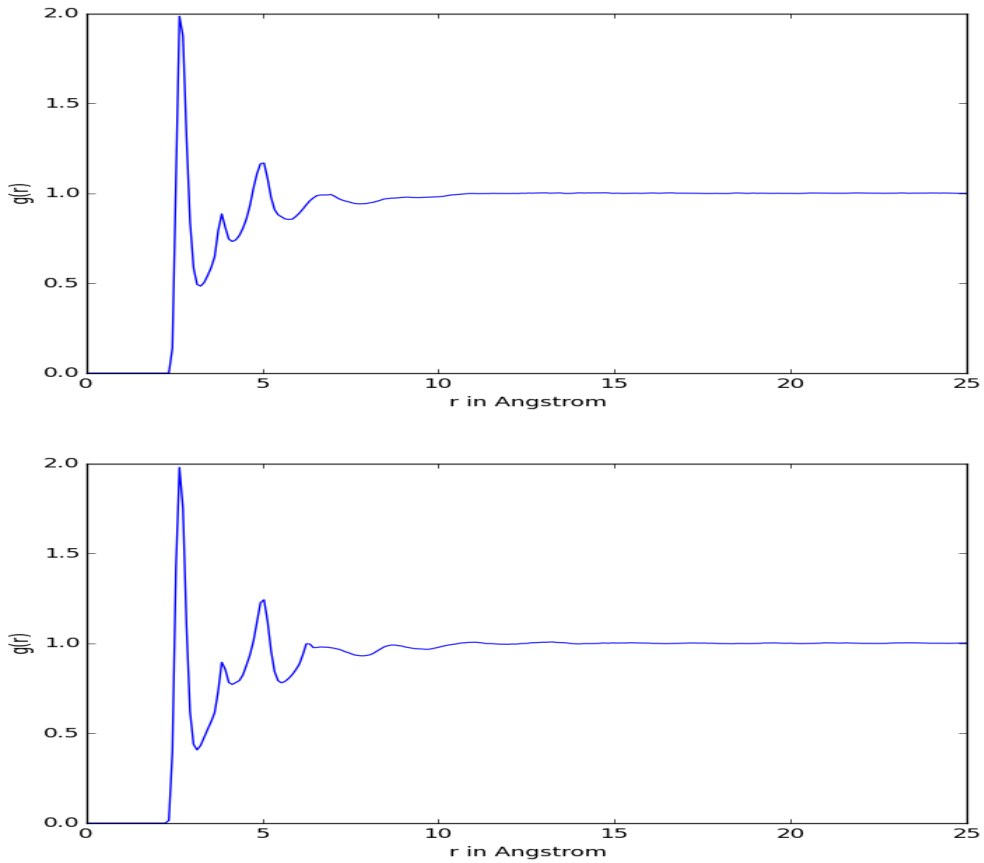

Figure 16: Average oxygen to oxygen radial distribution function for water around ATP (TIP3P model above, SPC/E model below)

Figure 16 shows the pair distribution function of oxygens belonging to the phosphate groups with the oxygens of all water molecules in the system. The pair distribution

function was obtained with the VMD command *measure gofr* using a maximal radius of 25 Angstrom around the ATP molecule and a selection update after every processed frame of the simulation. Oxygen atoms from the phosphate groups were chosen with the command *atomselect type ON2 or type ON3*, oxygen atoms of the water with *atomselect type OH2* where the atomtypes were first identified from the psf file. As all trajectories delivered the same course of the graph only plots for the third trajectory are shown. This trajectory was chosen as it has the best sampling rate, having the highest number of frames. Both radial distribution functions show a similar course. The SPC/E course shows slightly more pronounced peaks for the second and third peak at 3.9 Angstrom and 5 Angstrom. Those peaks belong to the second hydration shell which is better modeled by the SPC/E water model. The sharp peak at 3.2 Angstrom belongs to the first hydration shell. Those radial distribution functions help to define the hydration shells for the later analysis of dipole moment orientation around ATP. I define the region between 0 Angstrom and 3.2 Angstrom around ATP as first hydration shell, the region between 3.2 Angstrom and 5 Angstrom as second hydration shell and the region in a distance of 15 Angstrom to 25 Angstrom around ATP as bulk water which will be used as a reference.

Using the same definitions as above except for defining both selections with *atomselect type OH2* I was then able to compute the oxygen to oxygen radial distribution function for both water models in the bulk according to the definition above and to compare it to the given plots from the literature. Those plots can be seen in Figure 17.

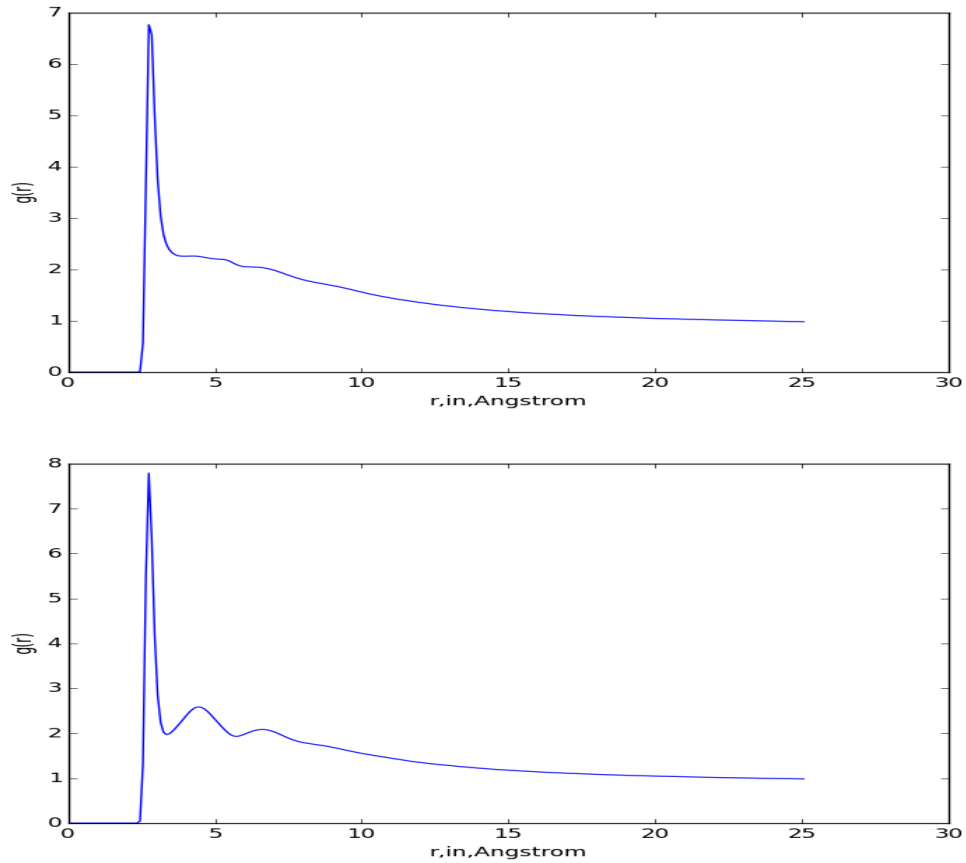

Figure 17: Average oxygen to oxygen radial distribution function for bulk water (above TIP3P, below SPC/E water models)

The obtained radial distribution functions for TIP3P and SPC/E bulk water are in good agreement with the literature [13]. For short distances to r_0, $g(r)$ equals zero due to strong repulsive forces from the Lennard-Jones potential. [10] The first peak at 2.8 angstrom is characteristic for the first hydration shell. While SPC/E also shows a second peak at around 4.5 Angstrom and a small third peak at approximately 6 Angstrom, the curve for TIP3P does not. This shows the SPC/E water model's ability to correctly model the second and third hydration shell which the TIP3P model has not. For long distances $g(r)$ decays to one for both models indicating the absence of a long range order within the structure. [10]

5 Dipole Moments Of Water Relative To The Phosphate Groups

Given all this knowledge about water dynamics around ATP and the internal ATP motion I was now able to compute the dipole moments of water relative to the phosphate groups to determine the polarization range of ATP. For that purpose I chose 5 reference axes and averaged over all of these axes. The reference axes are chosen to be those between a phosphate atom and the oxygen atoms connecting both phosphate groups. For symmetry reasons the average over all these axes will always be taken with orientation of the axes from minus to plus meaning the arrow shows always towards the phosphate atom. As the water

molecules move on a much smaller timescale than ATP, a higher time resolution of the trajectory is needed for a correct analysis. For this purpose only the third trajectory with a time resolution of 10 ns per integration step was used for the analysis of dipole moments. I calculated the projection of the dipole moments onto the distance vectors between the phosphates and the oxygen atoms connecting them before averaging over all projections. The internal motion of ATP should remain stable during the simulations to minimize the computational effort and errors during the rmsfit performed before calculating the dipole moments. For that purpose I looked into the RMSD plots for the third trajectory which showed the phosphate group RMSD value remaining stable during the simulation while the adenine motion showed sudden jumps. After those jumps also the adenine motion remained stable around 0.1 Angstrom or 0.6 Angstrom. As fluctuations around the 0.6 Angstrom scale were smaller than those around 0.1 Angstrom, I chose the part of the trajectory where the adenine RMSD lies around 0.6 Angstrom. The chosen parts of the third trajectory range from 65 to 380 ps for TIP3P water and 627 ps to 861 ps for SPC/E water. Different regions around ATP were first chosen for analysis. As water within a distance of up to 3 Angstrom around the phosphate groups only gave non convergent results, water within a distance of up to 3.2 Angstrom around the whole ATP was chosen. The divergence was detected with the help of the time correlation function which shows exponential decays with fluctuations around zero for converged data. In the case of water close to the phosphate groups only the time correlation function decreased linearily which can be explained by strong hydrogen bonding forces of water molecules to the oxygens of the phosphate groups. A test with water ligand-bound to the magnesium ion showed the same linear decrease in the time correlation function which leads to the assumption that strong bondings of water to an atom indicate long remaining times and therefore a slower decrease in the time correlation function.

5.1 Calculating Dipole Moments From The Trajectory

To obain data I wrote a script in tcl to be used with VMD. A first part consists of importing the psf file and creation of the output file as well as selecting a reference structure for the ATP molecule without its hydrogen atoms (also called heavy atom structure) and picking an arbitrary water molecule and calculating the modulus of its dipole moment vector p_0. As all bond lengths of water are fixed in both water molecules the vector can only change its orientation, therefore p_0 remains constant for all water molecules. This calculation was only done to obtain the most precise value of p_0 of my actual system. After that the script loops over each frame in the trajectory to save memory during computation. For each frame PBC wrapping was done with a subsequent rmsfit. The rmsfit works as follows: The current position of the ATP heavy atom structure is compared to the reference structure, the transition matrix needed to move from the current position to the reference structure is computed and after that the whole system is moved by the transistion matrix. Then the projections of dipole moment vectors to the reference axes were computed. For the computation of those projections first all water molecules within a chosen distance X to the ATP molecules (hydrogen atoms included) are selected using VMD's built-in command *atomselect*. After that the number of water molecules in the selection denoted with B was calculated by calling the number of atoms in the selection and dividing by three. With that I obtain a normalization quotient QN defined by

$$QN = \frac{1}{Np_0} \tag{14}$$

After that the overall dipole moment vector of the selection P is calculated and then resized to

$$\vec{P_X} = QN \cdot \vec{P} = \frac{\vec{P}}{Np_0} \tag{15}$$

We are left with calculating the reference axes. For that the script gets the position vectors of atoms PA, PB, PG, O5', O3B and O3A, calculates the reference axes from the difference of the current position vectors and normalizes those reference axes to modulus one. After that the projection k is calculated as arithmetic average over the individual projections of P on reference axes a_i as

$$k = \frac{1}{5} \sum_{i \in [1,5]} \vec{P}_X \cdot \vec{a}_i \tag{16}$$

and put to the output file. After the loop is finished, the output file is closed and the total running time printed to the screen.

Using linearity of the standard scalar product in the real Euclidean space, it can be proven that k equals the projection of all individual water molecule dipole moments onto chosen axes $\vec{a}_i$ after averaging over the axes and the number of water molecules. Let us denote the dipole moment of the $j-th$ water molecule with $\vec{p}_j$ and choose an arbitrary axis in the system to be denoted with $\vec{a}$. Taking the mean projection of all dipole moments in the system we get

$$Projection = \frac{1}{N} \sum_{i \in [1,N]} \frac{\vec{p}_i}{|\vec{p}_i|} \cdot \vec{a} \tag{17}$$

Using the fact that $\vec{a}$ is constant and also $|\vec{p}_i| = p_0 \forall i$ we can write this as

$$\frac{1}{Np_0} \vec{a} \cdot \sum_{j \in [1,N]} \vec{p}_i \tag{18}$$

where $\sum_{i \in [1,N]} \vec{p}_j = \vec{P}$ by definition of $\vec{P}$. With that we get, that

$$Projection = \frac{\vec{P}}{Np_0} \cdot \vec{a} \tag{19}$$

As the sum of my chosen axes will give again a vector and $\vec{a}$ was chosen arbitrarily, we can just denote $\vec{a} = \frac{1}{5} \sum_{i \in [1,5]} \vec{a}_i$ and get that $Projection = k$.

Therefore the calculated projection is equal to the projection I first wanted to obtain. Calculating the overall dipole moment of a selection requires less computational effort than looping over all water molecules to determine whether they are in a distance X to ATP and then calculating their individual dipole moment as well as the projection. In an earlier version of the script this loop was implemented and took a running time of 4 minutes for 100 frames of the 100000 frames trajectory, equaling a total running time of the script of 4000 minutes (approximately three days). The actual version delivers the same result as proven above with a running time of 45 minutes for the whole trajectory which is a drastic reduction in computation time.

5.2 Analysis Of Obtained Data

The analysis of the obtained data was done using Python with libraries *numpy* and *scipy*. To see whether the trajectory converges , the normalized autocorrelation function for the dipole moment projection was calculated using the *correlate* function from *scipy.signal*. Figures for all chosen domains show a typical exponential decrease to zero with fluctuations

around the asymptote. An indicator for the velocity of memory loss as a function of distance is the decrease time from one to zero. Around the ATP one expects a longer decrease time as dipole moment orientation towards the ATP should remain fixed by Coulomb interaction. This decrease time was calculated with a linear fit of the first 1000 data points of the autocorrelation function. As a comparison time correlation functions using the TIP3P model together with the linear fit can be seen in Figure 18

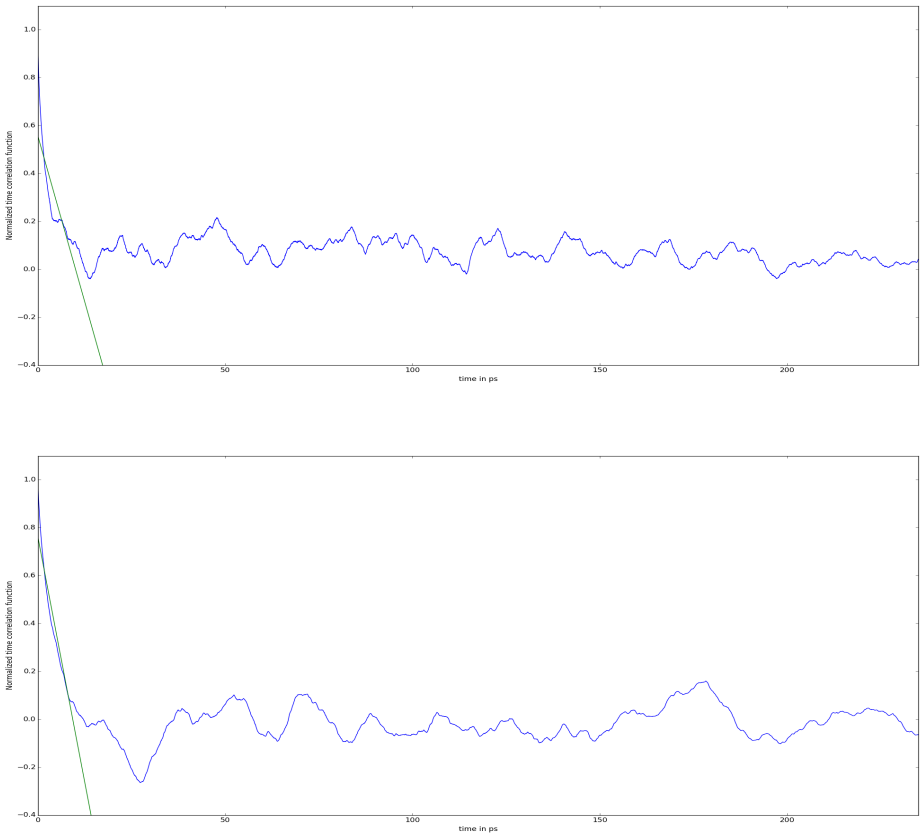

Figure 18: Time correlation functions of the first hydration shell and the bulk, TIP3P model with linear fit

Even though the function is exponential, I decided to use a linear least square fit for multiple reasons. The precision of exponential fits depends strongly on the choice of initial parameters while the least square fit uses the solution from an optimization problem to get the best parameters. The equivalent for an exponential fit would have been to choose arbitrary initial parameters, calculate the error between the fit and the data points and iterate over a high number of initial parameters to minimize the error. As the least square fit uses the optimization problem it will converge faster to a good approximation of the data than randomly guessing initial parameters. The most intuitive approach would have been using a logarithmic y-axis and then use a linear fit to calculate the time constant of the exponential function from the slope. Unfortunately the fluctuations around zero created too much noise for this method to be successful. The slope of the linear fit would give a percentual time loss per frame. Knowing that the time between two frames is $10fs$ one can then easily calculate the decrease time from the slope. Slope and decrease time can be found in table Table 4.

Table 4: Slope and decrease time

[H]

Model	Distance to ATP in A	Slope in $\frac{\%}{10fs}$	decrease time in ps
TIP3P	0 - 3.2	-5.5	0.2
TIP3P	3.2 - 5	-6.3	0.2
TIP3P	15 - 25	-8.1	0.1
SPC/E	0 - 3.2	-4.5	0.2
SPC/E	3.2 - 5	-7.1	0.1
SPC/E	15 - 25	-4.7	0.2

For the TIP3P water it can be seen that the decrease time decreases with rising distance to ATP. This can be attributed to stronger memory effects within the hydration shells due to polarization of water near the ATP molecule.

To determine the influence of the ATP on water structure concerning dipole orientation, I created histograms for the projection of dipole moments by using *numpy.histogram*. It can be seen from those histograms in Figure 19 that the center is shifted slightly to the positive domain with decreasing distance to ATP and that the standard deviation decreases with increasing distance. All histograms have a Gaussian shape and are centered around zero. Here only the example for TIP3P water is shown. Histograms looked the same for SPC/E water.

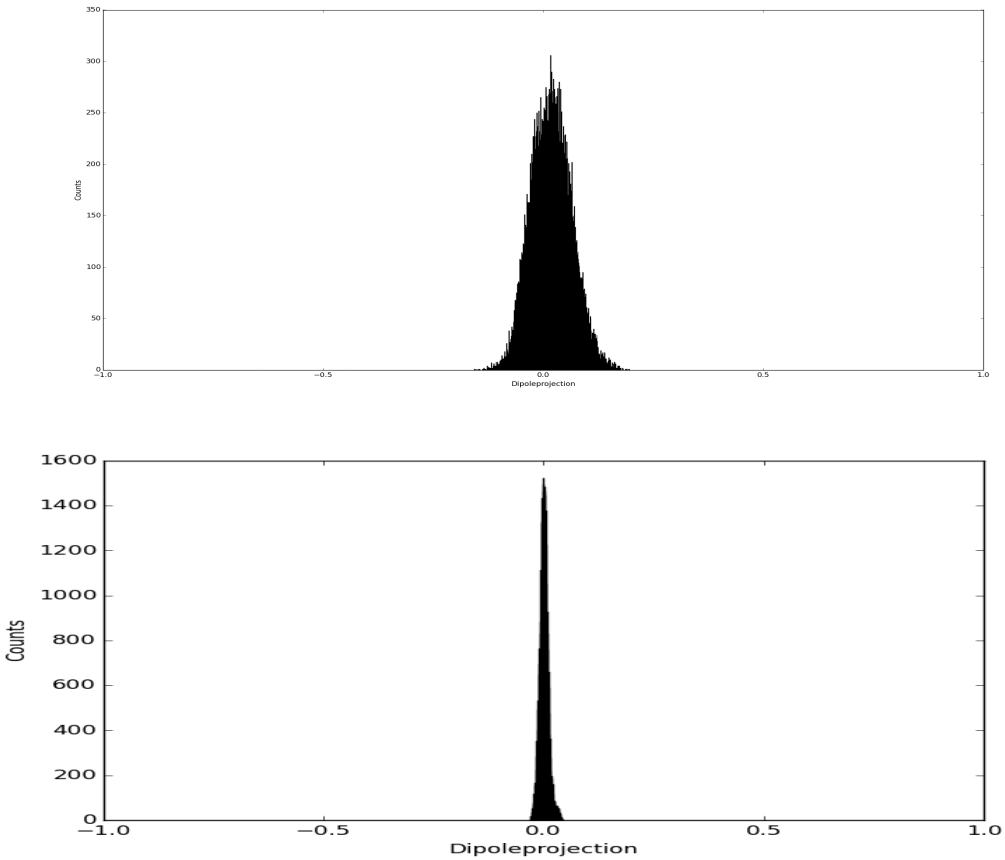

Figure 19: Histograms of the first hydration shell and the bulk

To determine the relation between polarization and distance, an analysis of the standard deviation and centering of the Gaussian histograms was done. For that purpose standard deviations and arithmetic mean, which is equal to the center of the Gaussian were calculated as a function ofthe distance to ATP. This was done by using overlapping spherical shells around ATP with a radius of 2 Angstrom each, e. g. water in a distance from 1-3 Angstrom, 2-4 Angstrom etc. Values determined from a shell were assigned to the distance value in the middle, e. g. values for the 2-4 Angstrom distance shell to a 3 Angstrom distance value. The values obtained in that way can be seen in Figure 20.

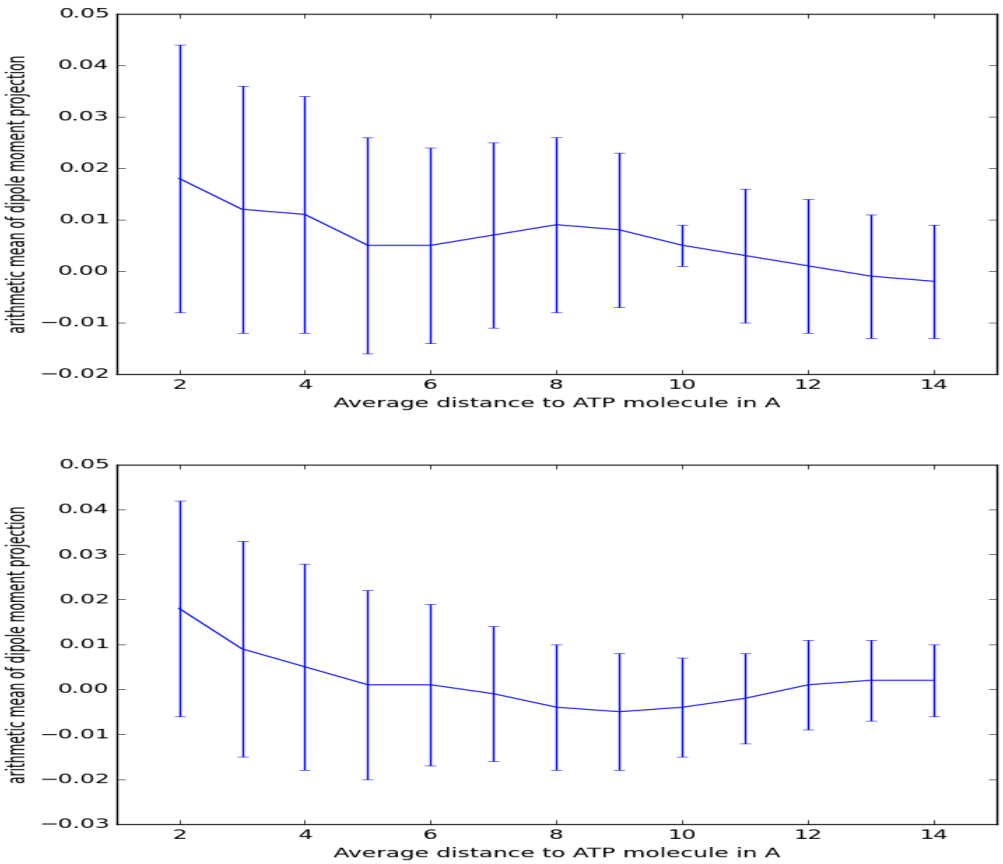

Figure 20: Plot of average mean as a function of distance to ATP using TIP3P water (above) and SPC/E water (below)

Both figures show the decrease in the arithmetic mean of the dipole moment projection as well as the decrease in standard deviation. For the former, differences between the individual values are too small to be meaningful. This can be attributed to bad statistics as water around the whole molecule was included in the analysis, which shows strong fluctuations around the adenine part of the molecule. A first analysis involving water around the phosphate groups only showed quite different results. In that case the water residence time around the phosphate groups is longer due to hydrogen bonding to the oxygens and the center of the Gaussian distribution of the projection is clearly shifted to the negative range with centering around -0.29. It shifted towards zero with increase in the distance. Unfortunately this strong bonding prevents the system from being equilibrated with the autocorrelation function decaying linear, not exponential, indicating longer times for memory loss.The decrease in the standard deviation can be observed in both cases and

can be attributed to a higher number of water molecules for the average. The spherical shell has a volume increase proportional to the cube of the radius. Introducing the constraint for the difference between two spheres being constant with 2 Angstrom, this gives an increase proportional to the square of the radius. Assuming that the number of water molecules in a shell increases with the volume this explains the increase in the number of water molecules per shell with the increasing distance and therefore the decrease in standard deviations. The standard deviation only accounts for statistical errors. For a discussion of the systematic error more than one data sample would have been needed for comparison . This could have been done by repeating the whole analysis for more than one piece of the trajectory.

6 Conclusion

Stable RMSD values indicate for all simulations to converge. Differences in the RMSD evolution over time for different parts of the ATP-molecule could be explained by water from the hydration shell intramolecularily bonding to the polar part of the molecule and thus stabilizing its motion. A similar trend can be observed for the dihedral angles with different water models. While the phosphate motion showed similar results for both water models, the adenine motion shows significant differences for both water models. Whether this is due to the different parameters in the force field or due to the adenine being more dynamic in its motion cannot be clearly answered. The radial distribution function of oxygen atoms from bulk water molecules could be reproduced for both water models and is in good agreement with the literature. On the other hand, water structuring around the ATP molecule as indicated by the pair distribution function is the same for both water models. The average number of water molecules bonding to ATP clearly shows how the polar part affects the structure of the water of the first hydration shell by water bonding to it. This could also be seen within the dipole moment analysis where the strong hydrogen bonds hold water molecules longer in place around ATP and therefore time of total memory loss as indicated by the time correlation function is longer. Most water molecules are binding to the terminal γ-phosphate group. To determine whether this also affects the hydrolysis would require a simulation using quantum mechanics to model the reaction. Still this would be an interesting question for further analysis. The distance dependance of dipole moment orientation around ATP was shown through the projection on the chosen axes. The distance dependance of memory decrease time determined trough the slope of the linear fit can be clearly seen for the TIP3P model with slope values ranging from -5.5 percent of memory loss per timestep to -8.1 percent per timestep. On the other hand the SPC/E model gave similar values for the first hydration shell (-4.5 percent) and the bulk (-4.7 percent). Here the outlier is seen for the second hydration shell (-7.1 percent). As only three values per water model are compared and also the uncertainty from the linear fit have to be considered these results are to be treated with caution. Still the polarization effects of ATP can also be seen from analysis of the arithmetic mean of dipole moment projections. While water near the ATP shows a small deviation from zero in the projection this average decreased with increasing radius to zero for bulk water. Unfortunately the effects from water around the adenine part of the molecule are necessary to obtain convergent data but also disturb the data. While water stronger bound to the phosphate groups showed a clear negative projection in polarization but a divergent time correlation function, the obtained convergent data showed a much weaker positive polarization effect. The question whether ATP is able to polarize the water around can be thus clearly answered by yes, but it is still unknown to what extend. Doing all those

analysis on a polarizable water model like TIP4P or a QM simulation could provide further information and would in my opinion be an interesting research question.

7 Acknowledgements

I want to thank Professor Ana-Nicoleta Bondar for her support and helpful advices during my bachelor thesis as well as for giving me the opportunity to learn about molecular dynamics simulation and statistical biophysics. Furthermore I want to thank Konstantina Karathanou for her feedback and help with the simulation setup and the analysis.

8 Notes

All online sources were last accessed May 25th 2017.

References

[1] The CHARMM force field. http://www.ks.uiuc.edu/Training/Tutorials/science/forcefield-tutorial/forcefield-html/node5.html.

[2] Cri. Abbildung 1: Schema zur bestimmung der rdf. https://de.wikipedia.org/wiki/Radiale_Verteilungsfunktion#/media/File:Rdf_schematic.jpg.

[3] J. C. Liao et al. The conformational states of mg.atp in water. *Eur Biophys. J.*, pages 29–37, 2004.

[4] J. C. Phillips et al. Scalable molecular dynamics with namd. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005.

[5] K. Meyer-Rogge et al. Bausteine der nucleinsäuren - atp als energiespeicher. http://www.chemgapedia.de/vsengine/vlu/vsc/de/ch/5/bc/vlus/atp.vlu/Page/vsc/de/ch/5/bc/atp/funktionen/energiespeicher/energiespeicher.vscml.html.

[6] M. Dittrich et al. On the mechanism of atp hydrolysis in f1-atpase. *Biophys. J 85(4)*, 2003.

[7] P. Roedig et al. A micro-patterned silicon chip as sample holder for macromolecular crystallography experiments with minimal background scattering. *Sci Rep 5*, 2015.

[8] M. Fyta. Water models in classical simulations. https://www.icp.uni-stuttgart.de/~icp/mediawiki/images/3/35/SimmethodsII_ss13_lecture6_watermodels.pdf.

[9] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[10] A. Hsu, J. Li. Molecular dynamics simulations of hydrophobic solutes in liquid water. http://w3.iams.sinica.edu.tw/lab/jlli/thesis_andy/node14.html.

[11] J. C. Gumbart, D. Luo. Hbonds plugin, version 1.2. http://www.ks.uiuc.edu/Research/vmd/plugins/hbonds/.

[12] H. Ohtaki. Structure and dynamics of hydrated ions. *Chem. Rev. E.93*, pages 1157–1204, 1993.

[13] L. Nilsson P. Mark. Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *J. Phys. Chem. A 105*, pages 9954–9960, 2001.

[14] T. Darden, D. York, L. Pedersen. Particle mesh ewald: An n log(n) method for ewald sums in large systems. *J. Chem. Phys. 98*, page 10089–10092, 1993.

[15] X. P. Resina. Potential of mean force.

[16] C. Mukhopadhyay S. G. Dastidar. Structure, dynamics, and energetics of water at the surface of a small globular protein: A molecular dynamics simulation. *Phys. Review E. 68*, page 021921, 2003.

[17] A.-N. Bondar S. Lorch, S. Capponi, F. Pieront. Dynamic carboxylate/water networks on the surface of the psbo subunit of photosystem ii. *J. Phys. Chem. B 119*, page 12172–12181, 2015.

[18] M. P. Allen, D. J. Tildesley. *Computer Simulation of Liquids.* Oxford University Press, 1987, edition from 1991.

[19] M. P. Allen, D. J. Tildesley. *Computer Simulation of Liquids.* Oxford University Press, 1987, edition from 1991.

[20] G. T. Ibragimova, R. C. Wade. Importance of explicit salt ions for protein stability in molecular dynamics simulation. *Biophysical Journal V. 74*, page 2906–2911, 1998.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Abschlussarbeit selbstständig angefertigt habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen (Abbildungen, direkte oder indirekte Zitate) habe ich unter Benennung des Autors / der Autorin und der Fundstelle als solche kenntlich gemacht. Mir ist bekannt, dass die wörtliche oder nahezu wörtliche Wiedergabe von fremden Texten oder Textpassagen ohne Quellenangabe als Täuschungsversuch gewertet wird. Ich erkläre weiterhin, dass die vorliegende Arbeit noch nicht im Rahmen eines anderen Prüfungsverfahrens eingereicht wurde.

I hereby declare that I wrote the bachelor thesis on my own and did not use other sources than those cited. The content from the used sourced I referenced as such by naming the author and the source. I am aware that citing of texts or passages of texts which were not written by me without citing the source will be seen as fraud. Furthermore I hereby declare that the following work has not been used within the frame of another exam.