Simulating a DPPC bilayer in polarizable water using a hybrid set-up

Bachelor Thesis Chemistry

Floor van Elsacker
s2321629
05-02-2016

First supervisor:　　dr. A.H. De Vries
Second supervisor:　dr. R.W.A. Havenith

Summary

A simple hybrid model was developed with the goal of simulating a DPPC bilayer in polarizable water (PW) [1] in a realistic manner. In the model the MARTINI [1] and Gromos [2] forcefield are combined. The solute (DPPC) was scaled while the solvent (water) remains coarse grained (CG). The forcefields were mixed using a scaling factor λ, whiche means that at λ =1 the lipids are represented atomistically and at λ =0 the lipids are represented coarse grained. The CG character of the lipids was represented by virtual sites (VS). [3] It was difficult to obtain the correct area per lipid using this simple model. Furthermore, the system could not equilibrate within a reasonable time. The model was altered slightly by disabling the atomistic (AA) – CG interactions and this lead to more stable results, but the right area per lipid was not be obtained. Furthermore the lipid bilayer seemed to be partially in the gel phase if the atomistic character was dominant. Due to the difficulties in pinpointing the errors in the system, the adjusted model was tested on a simpler system. It was used to simulate pure isopropanol and the results were compared with standard gromos and temperature scaling simulations [4]. The system was characterised using the diffusion coefficient D, the radial distribution function (RDF) and the conformational entropy. The diffusion coefficient was similar for all three methods, but the D for the hybrid model was slightly lower at all times even when λ =1 which means that the system was completely described on an atomistic level. The results should be similar for this value of λ. It is unclear what caused this. The results from the RDFs showed that the model could give reasonable results for a pure substance, although only high values of λ (> 0,6) could be used otherwise the atomistic character was lost. Furthermore it showed that the system seems to become more ordered if more coarse grained character is introduced. The results for the conformational entropy showed that the speed-up was minor but higher than the speedup of the temperature scaling method. The next step would be that the model could be tested on a more complex system, such as a mixture of two components, to see what the limits are to this simple set-up.

Introduction

Molecular dynamics (MD) have been used for several years to accurately describe several biomolecular systems [3,5], such as membranes. The main advantage of MD simulations is the high resolution which enables the user to study processes on an (semi-) atomistic level and review the dynamics [4].

The simulations can roughly be divided into two different categories. The first category is the atomistic (AA) MD simulations which allow the user to simulate a system in high detail, but which are computationally costly due to the amount of detail which leads to a limited time scale and a limited size of the system. The other category is coarse grained (CG) models. In this type of model a group of atoms is represented by one interaction site which saves a lot of computational time. However due to this approximation, the system loses part of its accuracy.

The problem lies in the fact that a lot of biomolecular processes require high accuracy at specific domains and a long time scale to be able to study the process properly[6]. For this reason a combination of the FG and CG forcefield is necessary. To be able to use the best of both worlds, several hybrid models were proposed the last few years. The aim of these models is to gain a certain speedup without losing too much atomic details. There are different categories of models: back mapping methods, constant scaling methods, interface methods and fixed resolution methods.[3] With backmapping, the system is first represented at a coarse grained level and is switched to a higher resolution. This is usually done for systems where equilibration is too slow to be simulated at an atomistic level. [6] For the constant scaling method, all the particles are represented at both resolutions at the same time. The ratio between the FG and CG character is given by a mixing parameter λ. [4]When the interface method is used, the resolution of the molecules is dependent on its location in the system. The system consists of domains of different resolution which can be divided into an atomistic, hybrid and CG domain.[7] The last category is the fixed resolution method. For this method all particles are represented either in FG or CG representation throughout the entire simulation independent of the location of the particle in the system. [3,5]

In this report a new model is proposed which is a combination of the fixed resolution and constant scaling method.The aim of the present research is to investigate whether the proposed model is able to simulate a lipid bilayer using polarizable water and give realistic results. For this model the lipid will be scaled while the solvent will remain fully coarse grained. The reason for this is to gain maximum speed-up by only using CG solvent. This decision was made under the assumption that the interaction between solute and solvent requires less accuracy and that at the same time it would still be possible to (partially) describe the solute-solute interactions at an atomistic level. This model is quite similar to models that were proposed earlier [3,5], except for the fact that in these models the solute was not scaled, but was kept at an atomistic level. In the model proposed by Wassenaar et al. [5] the atomistic (AA)-CG interaction was allowed for Coulomb interactions and virtual sites (VS) are used to describe the Lennard-Jones interaction. The VS are placed at the centre of mass of the molecule and interact with the CG beads. The forces acting upon the VS are then distributed among the atomistic particles. [3] In the model of Rzepiela et al. [3] virtual sites (VS) were used to handle all the AA-CG interactions. Two variations of the proposed model were made. In both cases the solute was scaled which differs from both models described earlier. The first resembles the model proposed by Wassenaar et al.[5] and handles Lennard-Jones interactions via VS and charges in a direct AA-CG interaction. The second model has a similar set-up compared to the model from Rzepiela et al.[3] and handles all non-bonded interactions between the solvent and solute on a CG level.First of all the necessary theory will be explained including the details about the proposed model. Then the results for the separate simulations will be given. Afterwards the result of a second simulation of pure isopropanol will be discussed. This simulation was done to check the results of the model on a simpler system. A conclusion and recommendations for further testing end this report.

Theory

In this section the theory used to define the hybrid model will be explained. The hybrid model is a combination of an atomistic forcefield (gromos) and a coarse grained forcecield (MARTINI). First necessary information about the separate forcefields will be mentioned and afterwards the suggested hybrid model will be described.

Forcefields

For the hybrid model the atomistic gromos forcefield is combined with the coarse grained MARTINI model. In this section the focus lies on the interactions that are scaled in the hybrid model.

GROMOS forcefield

The GROMOS forcefield is an atomistic forcefield which means that the molecule is represented on an atomistic scale. The exception to this rule are hydrogens connected to a carbon atom. These are not represented separately and are combined with the carbon, so CH_n particles are used.[2] The interactions can be divided into bonded and non-bonded interaction.

Bonded interactions

The bonded interactions are a sum of the bond, bond angle, harmonic (improper) dihedral angle and trigonometric (torsional) dihedral angle terms[8].

Bond stretching

The Gromos forcefield uses equation (1) to describe the potential for the bond stretching[8]:

$$V^{bond}(\vec{r}; K_b, b_o) = \sum_{n=1}^{N_b} \frac{1}{4} K_{b_n}[b_n^2 - b_o^2]^2 \qquad (1)$$

where the force constant K_b and the ideal bond length b_0 are dependent on the chosen bond type. These values for these parameters are essentially based on experimental data.

Bond angle

The potential energy of the covalent bond-angle interactions is given by equation(2): [8]

$$V^{angle}(\vec{r}; K_\theta, \theta_0) = \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} (\cos\theta_n - \cos\theta_{0_n})^2 \qquad (2)$$

The force constant $K_{\theta n}$ and the ideal bond length θ_0 are derived from experimental data. The actual bond angle is represented by θ_n.

Improper Dihedral-Angle interactions
Improper dihedral-angle interactions are used to keep a set of four atoms in a certain conformation. In this case these interactions are used to keep a tetrahedral configuration around carbon atoms and to keep the ester carbonyl group in the lipids flat. Equation (3) is used[8]:

$$V^{har}(\vec{r}; K_\xi, \xi_0) = \sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n} (\cos\xi_n - \cos\xi_{0_n})^2 \qquad (3)$$

where again K_ξ and ξ_{0n} are constants and ξ_n represents the actual dihedral angle.

Torsional Dihedral -Angle Interactions
The potential energy for this type of interaction can be described as[8]:

$$V^{trig}(\vec{r}; s) = V^{trig}(\vec{r}; K_\phi, \delta, m) = \sum_{n=1}^{N_\phi} K_{\phi_n} (1 + \cos\delta_n \cos m_n \phi_n) \qquad (4)$$

The value of δ is restricted to 0 or π, m is the mulitplicity of the torsional dihedral angle and φ is the actual value of the dihedral angle between the ijk and jkl planes for the dihedral between atoms i, j, k and l.

Non-bonded interactions
In Gromos the non-bonded interactions are the sum of the Lennard-Jones interactions and the electrostatic (Coulomb) interactions between all pairs of atoms[8] within the cut-off distance. For the Gromos forcefield the cut-off distance is set at 1.4 nm. The first and second neighbours are excluded from the non-bonded interactions and special parameters are used to describe the (1,4)-interactions.[8]

Coulomb interaction with reaction field
The standard equation for the Coulomb interaction between the two particles i and j is given as:

$$V_c(r_{ij}) = f \frac{q_i q_j}{\varepsilon_r r_{ij}} \qquad (5)$$

where q_i and q_j are the charges of the particles i and j respectively, r_{ij} is the distance between the particles and ε_r is the dielectric constant of the substance and f is given by equation (6):

$$f = \frac{1}{4\pi\varepsilon_0} \qquad (6)$$

The constant ε_0 is the dielectric constant of vacuum. The Coulomb interaction can be adjusted if it is assumed that the system is homogeneous and there is a constant dielectric constant for the environment (ε_{rf}) beyond the cut-off distance. This field will then also influence the particles within the cut-off distance. This type of interaction is called a Coulombic interaction with a reaction field. This means that a particle experiences a field from another particle and a general reaction field. Equations (7), (8) and (9) are used to describe the potential and the force.[8]

$$V_{crf} = \frac{q_i q_j}{\varepsilon_r 4\pi\varepsilon_0} \left[\frac{1}{r_{ij}} + k_{rf} r_{ij}^2 - c_{rf} \right] \qquad (7)$$

$$k_{rf} = \frac{1}{r_c^3} \frac{\varepsilon_{rf} - \varepsilon_r}{(2\varepsilon_{rf} + \varepsilon_r)} \qquad (8)$$

$$c_{rf} = \frac{1}{r_c} + k_{rf} r_c^2 = \frac{1}{r_c} \frac{3\varepsilon_{rf}}{(2\varepsilon_{rf} + \varepsilon_r)} \qquad (9)$$

The potential becomes zero at the cut-off distance, so for $r_{ij} = r_c$.
Since the force is described as

$$F_i(r_{ij}) = \frac{-dV_{crf}}{dr_{ij}}$$

(10)

It is described by equation (11):

$$F_i(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r}[\frac{1}{r_{ij}^2} - 2k_{rf}r_{ij}]$$

(11)

However, if this equation is used, the force will not become zero at the cut-off distance and will therefore not be continuous when used. Therfore the equation is slightly adjusted:

$$F_i(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r}[\frac{1}{r_{ij}^2} - 2k_{rf}r_{ij} - f_{shift}]$$

(12)

The f_{shift} factor is added to make the force zero at the cut-off distance. The f_{shift} factor is calculated by filling in r_c for every r_{ij} and to see what is necessary to make the function zero. It is given by:

$$f_{shift} = \frac{3\varepsilon_r}{r_c^2(2\varepsilon_{rf} + \varepsilon_r)}$$

(13)

In the Gromacs manual the force is represented as a vector[8]:

$$\vec{F}_i(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r}[\frac{1}{r_{ij}^2} - 2k_{rf}r_{ij}]\frac{\vec{r}_{ij}}{r_{ij}}$$

(14)

This equation also does not contain the f_{shift} factor and will therefore not go to zero smoothly at the cut-off distance. However, the GROMACS code does implement the shift internally.

The charges can be defined on the atomistic level and on the coarse grained level. The r_c, the ε_r and ε_{rf} can be varied for each different type of interaction between atomistic particles, coarse grained particles and virtual sites (AA-AA, AA-CG, CG-CG, CG-VS, VS-VS).

Lennard-Jones interaction
For the Lennard-Jones potential equation (15) is used:

$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{12}}{r_{ij}^{12}} - \frac{C_{ij}^6}{r_{ij}^6}$$

(15)

Since again the force is described using equation (10), it can be described as:

$$F_i(r_{ij}) = (12\frac{C_{ij}^{12}}{r_{ij}^{13}} - 6\frac{C_{ij}^6}{r_{ij}^7})$$

(16)

All the C6 and C12 terms can be specified separately for each interaction between two particles. Beyond the cut-off distance the Lennard-Jones interactions are defined as zero. A shift function to ensure that the function is continuous at the cut-off distance was not used

MARTINI forcefield
The MARTINI model is an example of a CG model. [1] Within the MARTINI model a four-to-one mapping is used, which means that four heavy atoms (so not hydrogen) are represented by a single interaction centre, also known as a bead. There are four different types of beads: polar, nonpolar, apolar and charged. Within these categories there are sublevels. [1] The difference in these beads can be seen from the parameters in the Lennard-Jones interactions and of course in the charge of the particles. The same distinction between bonded and non-bonded interactions can be made for the MARTINI forcefield.

Bonded interactions
The bonded interaction within the MARTINI forcefield are described by bond stretching and the

bond angle interaction. There were no dihedrals used in this model.

Bond stretching
The covalent bonds are described in the MARTINI model by a harmonic potential:

$$V^{bond}\left(\vec{r}\,;K_{bond},R_0\right)=\frac{1}{2}K_{bond}\left(R-R_0\right)^2 \tag{17}$$

The force constant K_{bond} is for the MARTINI model set at 1250 kJmol^{-1}nm^{-2} and the equilibrium distance r_{bond} at 0.47 nm, but the values can be altered if necessary.[1]

Bond   *angle*
For the bond angle interactions the potential energy can be described as[1]:

$$V^{angle}\left(\vec{r}\,;K_{angle},\theta_0\right)=\frac{1}{2}K_{angle}\left(\cos\left(\theta\right)-\cos\left(\theta_0\right)\right)^2 \tag{18}$$

For aliphatic chains the force constant K_{angle} is set at 25 kJmol^{-1} and the equilibrium angle θ_0 is set at 180°. For cis double bonds, the values become 45 kJmol^{-1} for K_{angle} and 120° for θ_0.

Non-bonded interactions
In the MARTINI model only the nearest neighbour is excluded from non-bonded interactions.[1] The cut-off distance is set at 1.2 nm

In the MARTINI forcefield shift functions are used for both the Coulomb interactions and the Lennard-Jones interactions. These shift functions have the same general structure. Therefore, first the shift function will be explained. The shift function for the non-bonded potentials is given by the equations (19), (20), (21) and (22)[9]:

$$\Phi_{sn}\left(r\right)=\frac{1}{r^{\alpha_n}}-\alpha_n\frac{A}{3}\left(r-r_1\right)^3-\alpha_n\frac{B}{4}\left(r-r_1\right)^4-C \tag{19}$$

$$A=\frac{-\left(\alpha_n+4\right)r_c-\left(\alpha_n+1\right)r_1}{r_c^{\left(\alpha_n+2\right)}\left(r_c-r_1\right)^2} \tag{20}$$

$$B=\frac{\left(\alpha_n+3\right)r_c-\left(\alpha_n+1\right)r_1}{r_c^{\left(\alpha_n+2\right)}\left(r_c-r_1\right)^3} \tag{21}$$

$$C=\frac{1}{r_c^{\alpha_n}}-\alpha_n\frac{A}{3}\left(r_c-r_1\right)^3-\alpha_n\frac{B}{4}\left(r_c-r_1\right)^4 \tag{22}$$

It is clear that the C term makes the potential go to zero at the cut-off distance.
The r_1 in the equations is the r_{shift}, so that means that from that distance on a shift function will be used.

The force is thus represented by equation (23):

$$F_s\left(r\right)=\frac{1}{r^{\left(\alpha_n+1\right)}}+\left[\alpha_n A\left(r-r_1\right)^2+B\left(r-r_1\right)^3\right] \tag{23}$$

Coulomb interaction
The Coulomb potential is in this case given by a shifted Coulomb potential: [8]

$$V_{s,C}\left(r_{ij}\right)=f\,\Phi_{sl}\left(r_{ij}\right)q_i q_j \tag{24}$$

For the coulomb interaction $\alpha = 1$, so the Coulomb force is represented by:

$$F_{s,C}\left(r\right)=\frac{1}{r^2}+\left[A\left(r-r_1\right)^2+B\left(r-r_1\right)^3\right] \tag{25}$$

For these interactions the $r_1=0$, so either the shift function is applied or the potential (and force) are

set to be zero.

Lennard-Jones interaction
For the Lennard-Jones interaction the r_{shift} is not set at zero. This means that the function used to describe the potential (and thus force) varies with the r. If the r is bigger than the r_{shift}, but smaller than the r_c , equation (26) applies.

$$V_{s,LJ}(r_{ij}) = \Phi_{s3}(r_{ij})C_{12} - \Phi_{s2}(r_{ij})C_6 \qquad (26)$$

The shift functions have the same definition as in equation (20), (21), (22) and (23), but $\alpha_2 = 6$ and $\alpha_3 = 12$. For the forces the same definition as in (19) and (20) apply with the different values for α. The total Lennard-Jones force is then:

$$F_{s,LJ}(r) = F_{s,3}(r) - F_{s,2}(r) \qquad (27)$$

If the r is smaller than the r_{shift} the standard Lennard Jones potential is used (see equation (15) and (16)). The C6 and C12 parameters are defined based on the types of beads involved in the interaction.[8]

Hybrid model
The proposed hybrid model describes all solutes constantly at both resolutions and always represents the solvent at a coarse grained level. The goal was to use DPPC as a solute and polarizable water as solvent. To mix the different resolutions the scalingfactor λ is introduced. The potential energy is a sum of the fine grained (FG) interactions and the CG interactions:

$$V^{Total} = \lambda V^{FG} + (1-\lambda)V^{CG} \qquad (28)$$

The bonded and non-bonded interactions were scaled using this scalingfactor λ. For the bonded interactions this means that the force constants used in the different equations for the interactions were scaled. The non-bonded interactions were scaled as well, but the exact interactions depend on the exact model that is used as is explained later on.

Two different models were proposed. The first model described the charges always on the atomistic level which means that the fully unscaled atomistic charges are used for all interactions. An overview of the non-bonded interactions within the proposed model can be found in table 1.

	AA	CG	VS
AA	Gromos	Gromos with repulsive LJ	DISABLED
CG	Gromos with repulsive LJ	MARTINI	MARTINI without charges
VS	DISABLED	MARTINI without charges	MARTINI

Table 1: An overview of the non-bonded interactions in hybrid model 1

The repulsive LJ for the AA-CG was introduced to make sure there was no overlap between the charged AA and CG particles if polarizable water was used. The charged satelites of the PW normally do not have a short-ranged Lennard-Jones repulsion with other particles in a normal CG set-up, since they are well hidden within the volume of the whole particle, but this leads to problems in combination with the much smaller atomistic representation of the charged groups.[1]

The second model was proposed due to unrealistic results from the first model as will be discussed later on. In this model the interactions between the CG water and the atomistic representation of the lipid was disabled and the CG-AA interaction was treated entirely at CG level through the VS. The charge should now always be described at a coarse grained level. The non-bonded interactions within the new system are summarized in table 2:

	AA	CG	VS
AA	Gromos	DISABLED	DISABLED
CG	DISABLED	MARTINI	MARTINI
VS	DISABLED	MARTINI	MARTINI

Table 2: An overview of the non-bonded interactions in hybrid model 2

This also implies that only the AA-AA and VS-VS interactions were scaled. The other interactions (CG-VS and CG-CG) were fixed. Since the AA-CG interactions are no longer enabled the extra repulsive Lennard-Jones interactions are no longer necessary.

Materials & Method

First the default settings will be explained and afterwards the specific setting of each simulation will be mentioned. All simulations were run using the GROMACS simulation package (version 4.6.7)[8].In all simulations the Gromos 53a6 parameter set was used for the atomistic parts of the system. For the coarse grained force field MARTINI v2.2 was used with PW when PW was used. In all other cases the normal MARTINI v2.2 was used. The Berendsen method was used for temperature coupling at a temperature of 327 K and for the pressure coupling at a pressure of 1 bar. The pressure coupling type was semiisotropic. The non-bonded interactions were described using user tabulated potentials. The scalingfactor λ was introduced in the user tabulated potentials to enable the scaling of the different interactions.

Several simulations were performed. For the DPPC simulations 144 DPPC were dissolved in 1280 PW beads. The simulations were run at a constant pressure for varying values of λ and varying values of the dielectric constant ε_r of the Coulomb interactions for the AA-AA and the AA-CG interactions. The default value of ε_r is 1.0. When the ε_r of one interaction was varied, the ε_r of the other interaction was kept at the default value. The dielectric constant of the reaction field, ε_{rf}, was set at 62. The ε_r for the all the interactions at coarse grained level was kept at 2.5.

To test the second model, simulations were run of pure isopropanol. For this system, there is no difference between solvent and solute and only the groups AA and VS are used. In this case 500 isopropanol molecules were simulated in a cubic box at a constant volume after the system was equilibrated using the berendsen pressure coupling for the full AA simulation. In this model no constraints were used. The system was tested for varying values of λ. The ε_r was kept at a constant value of 1.0 and the ε_{rf} was kept at the constant value of 62. The temperature was kept at 327 K. For reference the system was also simulated using Gromos 53a6 parameter set and was compared to another hybrid model, the temperature scaling method[2](T-Scaling). These simulations were done at the same constant volume.

Discussion & Results

The results of the various experiments will be discussed seperately starting with the results for the DPPC bilayer in PW using both models followed by the results of a pure isopropanol system using the second hybrid model.

DPPC lipid bilayer in PW

For these simulations the area per lipid was used as a criterium to see whether the simulation could give realistic results. The area per lipid was chosen, since this is seen as an important structural quantity in the characterization of the bilayer.[10]

Hybrid model 1

First the results of the simulation run at λ = 1 will be discussed. In this system the lipids are described on an atomistic level and are surrounded by CG water. For this value of λ the set-up is similar to the model of Wassenaar et al.[5] It is known from literature that the area per lipid should

have a value of 0.631 nm^2 with a variance of 0.013 nm^2 at a temperature of 50 °C (323K) and a value of 65.0 nm^2 with a variance of 0.013 nm^2 at a temperature of 60 °C (333 K). [10] The area per lipid was determined by calculating the area of the box by multiplying the box size in the X- and Y-direction and dividing this by 72 (half the number of lipids). A representative graph is shown in figure 1. The area per lipid for ε_r (AA-AA) and ε_r (AA-CG) were set both at 1.0 for the first experiment. As can be seen from table 3, the correct area per lipid could not be achieved. For the DPPC system the area per lipid can be influenced by the AA-CG interaction, which describes the lipid-solvent interaction and the AA-AA interaction, which describes the lipid-lipid interaction. It is known from previous work from Wassenaar et al. that it is a challenge to describe the AA-CG resolution due to the difference in resolution, especially when it concerns the Coulomb interactions. The study showed that varying the ε_r for the AA-CG interaction could improve results for a combination of polarizable water and (charged) amino acids.[5] This option was also investigated for the DPPC system to investigate how lipids might respond to the changing dielectric constant for the AA-CG interactions. The ε_r for the AA-AA interaction was also varied to investigate the effect of the lipid-lipid interactions on the area per lipid. The graphs for the area per lipid can be found in the appendix. It should be noted that the time scales for the different simulations differ and are not always representative for the (total) run of the simulations.
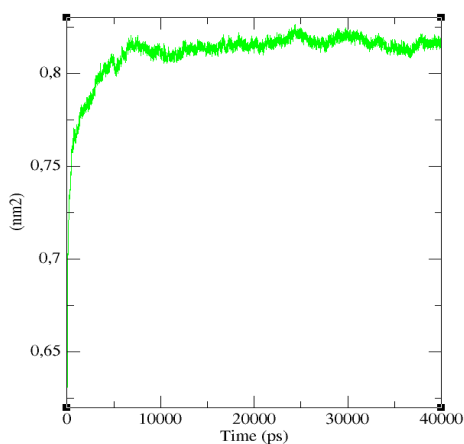
The results for the different simulations are shown in table 3. The results were obtained using the g_analyze tool. The area per lipid are shown with the standard deviation, calculated from the moment that the system was equilibrated. If the ε_r was varied for one of the interactions, it was kept constant at a value of 1.0 for the other interaction. Not all simulations were run until equilibration which can be seen from the slowly changing area per lipid over time, due to lack of time. This should be done to get more accurate results.

Figure 1: The area per lipid (ε_r is 1.0 for AA-AA and AA-CG interactions)

Interaction	ε_r	Area per lipid (nm^2)	
AA-AA	1.0	0.82 ± 0.01	
	1.45	0.74 ± 0.01	
	1.8	0.76 ± 0.01	
	2.0	0.72 ± 0.01	$\varepsilon_r (AA\text{-}CG) = 1.0$
	2.5	0.71 ± 0.01	
	2.75	0.71 ± 0.02	
AA-CG	1.0	0.82 ± 0.01	
	1.45	0.71 ± 0.01	$\varepsilon_r (AA\text{-}AA) = 1.0$
	2.5	0.56 ± 0.01	

Table 3: Area per lipid for varying values of ε_r for the AA-AA and AA-CG interactions.

Futhermore figure 1 shows that the equilibrium is reached after approximately 10000 s, so it is uncertain whether the other simulations have equilibrated or have just reached a local equilibrium. This could also indicate instabilities within the system. The data were plotted and showed the following trends:
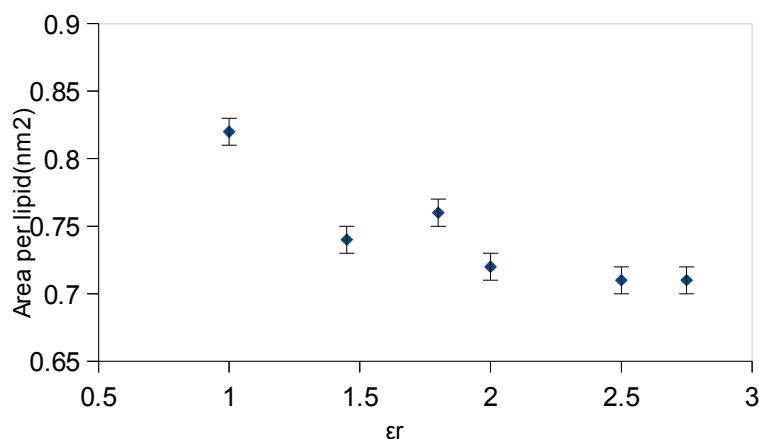
Figure 2: The influence of ε_r (AA-AA) on the area per lipid

Figure 2 shows that the varying of the ε_r influences the area per lipid. The area per lipid seems to decrease as the ε_r increases, except for the highest value of ε_r. An explanation for the general trend can be that a higher ε_r implies a larger screening of the polar head groups from one another and makes it therefore possible that the headgroups are more closesly packed. However the area per lipid remains stable for ε_r = 2.75 which contradicts the given explanation, unless a certain maximum is reached. Furthermore, it is clear that for none of the values of ε_r the correct area per lipid is achieved, since all the area per lipids are too high, given that the experimental results predict an area per lipid of approximately 0.64 nm^2.

The results for the AA-CG interactions follow a similar trend as can be seen in figure 3. This should however be said with a certain precaution, since there are only three data points. The ε_r now represents how well the water and lipids are screened from one another. The decrease in area per lipid can be explained due to the decreasing attractive forces between the polar water and the charged headgroups. This implies that the headgroups will be less hydrated which well decrease the area per lipid. From the work of Wassenaar et al. it was concluded that a value for ε_r of 1.45 would lead to the best result where a simlar hybrid set-up was used and tested for a system containing amino acids in PW[5]. However this was not the case in this system. The best value for ε_r for the AA-CG interaction would be higher in this system, although further testing would be required to find the exact value of the ε_r. However this option was no further explored due to instabilities in the system.
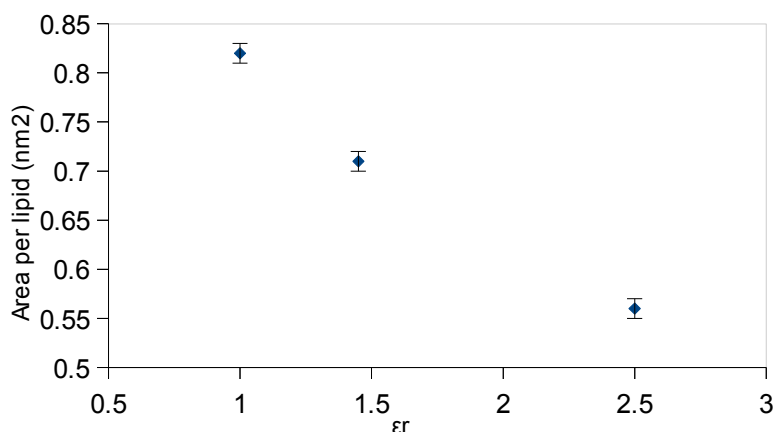
Figure 3: The influence of ε_r (AA-CG)on the area per lipid

To test the effect of the amount of solvent in the system the amount of water molecules was doubled. This increased the area per lipid drastically to 1.0 nm^2 and the area per lipid kept rising and could not equilibrate within a reasonable amount of time. The head groups are surrounded by water as could be seen from the simulation. A snapshot is shown in figure 4. The higher area per lipid is not surpising when it is compared with the results of changing the ε_r
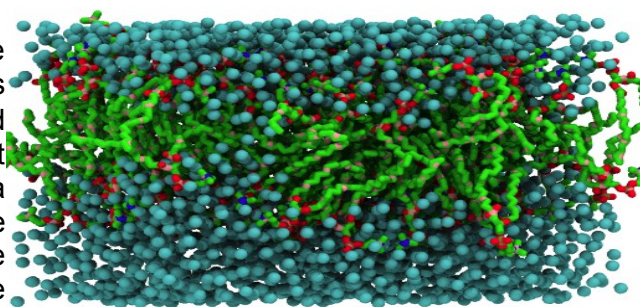
Figure 4: Snapshot of simulation with extra solvent

10

for the AA-CG interactions. However, the high increase is, since the value has nearly doubled. Furthermore, there was even water present in the hydrophobic part of the bilayer and a pore was visible in the bilayer which might indicate that this model is not feasible.

Hybrid model 2

One of the possible reasons for the difficulties with the first model might be that the difference in resolution in the AA-CG interactions is not manageable. Therefore an alteration to the first model was made. In the second model the interactions between the atomistic representation of the lipids and the coarse grained solvent is handled via virtual sites. Therefore it is no longer possible to vary the ε_r for the AA-CG interactions. From the hybrid model 1 it can be seen that varying the ε_r for the AA-AA interactions did not improve the results significantly. Therefore the λ value of the system was varied instead. The area per lipid was determined in the same manner as for the first model. The results for the simulations were generated using the g_analyze tool and are given in table 4:

λ	Area per lipid (nm^2)
0	NO BILAYER
0.1	0.75 ± 0.02
0.2	0.78 ± 0.02
0.5	0.69 ± 0.01
0.9	0.54 ± 0.01
1.0	0.50 ± 0.01

Table 4: Area per lipid for varying values of λ

These simulations turned out to be more stable and show more promise in terms of the area per lipid. However, it still took the system relatively long to equilibrate. The system only crashed for $\lambda = 0$. The trend is shown in figure 5:
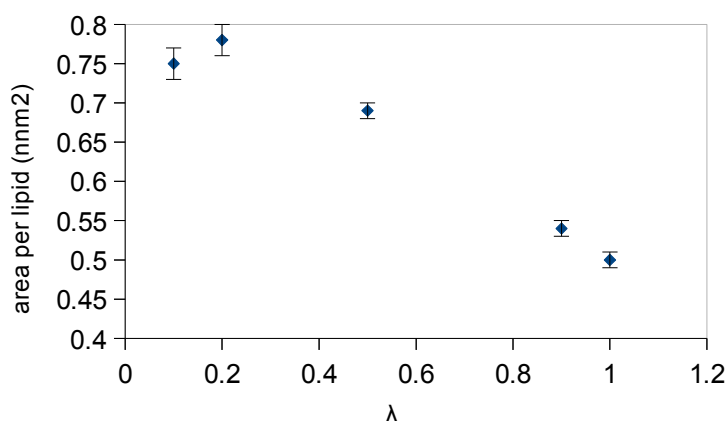
Figure 5 :The influence of varying λ on the area per lipid

The results indicate that the area per lipid decreases when λ is increased. Only the results at $\lambda = 0.1$ and 0.2 do not follow this trend exactly, but the difference between the two values of the area per lipid is within the error margins.

A full CG simulation was performed using the MARTINI forcefield (version 2.2P) which gave an area per lipid of 0.63 nm^2 which is similar to the reference value[10] which means that the hybrid model should be able to reproduce this result at λ=0 and give similar results for low values of λ, but as can be seen from the results the area per lipid is too big for values of λ = 0.1-0.5.

The area per lipid is however also too low for several values of λ. One explanation might be that the membrane is partially in the gel phase instead of the liquid ordered phase as is the case for the simulations at low values of λ. The main difference between these phases is the lateral diffusion in the membrane, that is one hundred times smaller in the gel phase than in the liquid ordered phase. These phases are known to coexist in biological systems.[11] From the simulations it could be seen that the system is nearing the gel phase for high values of λ. This observation was made based on the simulations without quantative analysis. A picture from the simulation for $\lambda = 1$ is shown in figure 6. This is an interesting result, since this phenomenon is only visible when the atomistic character in the lipids is dominant. The occurrence of the gel or liquid ordered phase depends on

the lipid composition and the temperature.[11] Both parameters should be the same for every value of λ, so there is no apparent reason for the occurrence of the gel phase, unless there is a temperature difference within the system. The use of thermostats for the different subsystems (solute and solvent) did not seem to solve this issue.
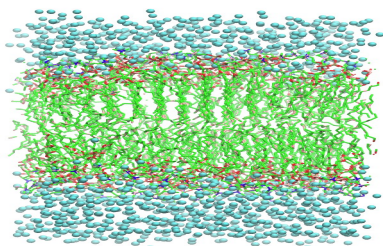

Figure 6: Snapshot of the simulation for λ =1

The coupling between the CG and AA is handled via virtual sites which are excluded from the normal temperature measurements due to the fact that VS have no mass. This means that the kinetic energy cannot be calculated.To get a full picture of possible temperature differences in the system, it would be interesting to get an indication of the temperature of the VS to see whether this is consistent. For this reason, the distribution of the velocities of the VS was measured to see whether they would match the Maxwell distribution of speeds which is represented by equation (29)[12]:

$$f(v) = 4\pi \left(\frac{M}{2\pi RT} \right)^{3/2} v^2 e^{-Mv^2/2RT}$$

(29)

In this equation T is the temperature, v is the velocity and M is the molar mass. An example of such a distribution is shown here and is compared to the distribution for the full CG system. The other distributions can be found in the appendix.
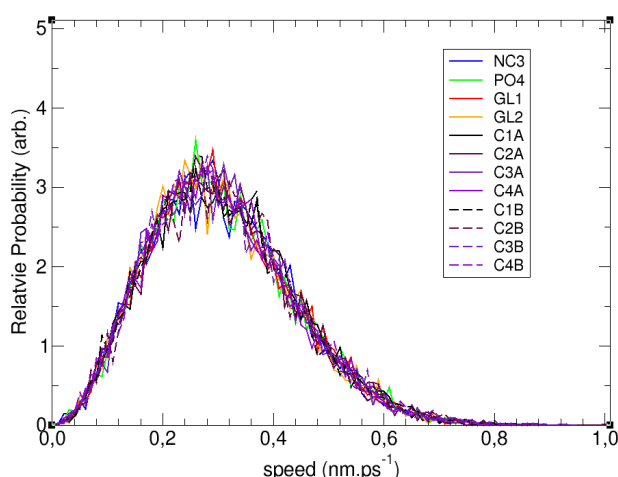
Figure 7: The distribution of velocities for a full CG system
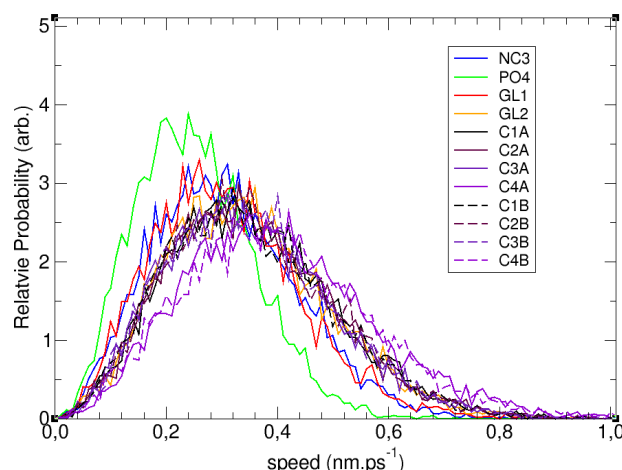
Figure 8: The distribution of velocities of the VS for λ = 0,1

To be able to make this distribution the masses of the groups of atomistic particles that the VS represent were used. This difference in mass also explains the difference between the different graphs in figure 8. In the MARTINI simulation all beads have the same mass, so this explains the difference between figure 7 and 8. It should be noted that the distributions were calculated over relatively small datasets so there was still a lot of noise visible. This makes a quantitative analysis of the distribution difficult. In general it can be seen that the distributions for the different systems are similar from a qualitatively point of view, which indicates that the temperature is consistent for varying values of λ and is also similar to the temperature in a full CG system.

This model shows more promise than the first one, but it is still unable to reproduce the correct area per lipid for the system. Due to the complexity of the current system it becomes quite difficult to pinpoint where the error lies. Therefore the model was used for a much simpler system, namely pure isopropanol.

Isopropanol

Hybrid model 2 for isopropanol was tested for three different values of λ. All particles in the system were scaled using the scalingfactor due to the absence of solvent. The system was tested on a number of properties, namely radial distribution functions (RDF), diffusion coefficient (D) and conformational entropy. After equilibrating the system for λ =1, the density of the system was found to be 811 kg/m^3. It is known from experiment that the density of isopropanol is 885 kg/m^3 at 298K[2] and 740 kg/m^3 at 298 K if the Gromos parameter set 53a5 was used. These values show that the calculated density is a reasonable value, certainly when it is taken into account that it is difficult to get a correct value for the density, even when using the standard gromos forcefield.

Diffusion

The value of the diffusion coefficient D is characteristic for a substance and is therefore deemed a good parameter to test the hybrid model for. The diffusion coefficient in one direction can be calculated using equation (30) [12]:

$$\langle x^2 \rangle = 2Dt \tag{30}$$

Where <x^2> is the mean square displacement (MSD) in the x-direction and t is the time. Only the linear part of the MSD graph should be used to calculate D. The MSD was calculated using the center of mass of the molecules, except for the gromos simulation. In that case the central carbon atom was used, because it was deemed to be a good enough approximation.

It has to be noted that different methods were used to calculate D. The normal GROMACS tool g_msd showed an artefact in the MSD curve. There was a certain amount of noise in the graph which should not appear and was only visible if the temperature scaling method was used. Therefore a different method was used to calculate the MSD for the temperature scaling. The GROMACS tools g_traj and g_analayze were used subsequently to calculate the MSD for all molecules separately and then the average was used to calculate the average MSD.

The D was calculated in all three directions for all different systems. The results are shown in table 5, 6 and 7.

T-Scaling	D (10^{-9} m^{2}s^{-1})		
System	X-direction	Y-direction	Z-direction
λ= 1.00	0.38 ± 0.04	0.40 ± 0.10	0.30 ± 0.04
λ= 0.75	1.65 ± 0.15	1.60 ± 0.20	1.40 ± 0.15
λ= 0.60	2.00 ± 0.15	2.20 ± 0.20	1.90 ± 0.15

Table 5: Diffusion contant in the x, y and z-direction for the temperature scaling method

Model 2	D (10^{-9} m^{2}s^{-1})		
System	X-direction	Y-direction	Z-direction
λ= 1.00	0.23 ± 0.01	0.25 ± 0.00	0.31 ± 0.01
λ= 0.75	1.38 ± 0.01	1.37 ± 0.06	1.34 ± 0.14
λ= 0.60	1.93 ± 0.09	1.85 ± 0.10	1.71 ± 0.02

Table 6: Diffusion contant in the x, y and z-direction for hybrid model 2

Gromos	D (10^{-9} m^{2}s^{-1})		
System	X-direction	Y-direction	Z-direction
λ= 1.00	0.33 ± 0.01	0.36 ± 0.03	0.38 ± 0.02

Table 7: Diffusion contant in the x, y and z-direction for the gromos method

The overall diffusion can be calculated by using the Einstein relation [8]:

$$\lim_{t \to \infty} \langle \| r_i(t) - r_i(0) \|^2 \rangle_{i \in A} = 6\, D_A t \tag{31}$$

Where it is stated that the MSD of each molecule in all three direction (x,y,z) of substance A is equivalent to the diffusion constant multiplied by 6 and the time t. This was done for the hybrid model and the all-atomistic simulation. For the temperature scaling the average of the x, y and z-direction were used to be able to compare the results.

D ($10^{-9}\,m^2s^{-1}$)			
System	T- scaling	Hybrid model 2	Gromos 53a6
λ= 1.00	0.36 ± 0.08	0.26 ± 0.01	0.35 ± 0.02
λ= 0.75	1.55 ± 0.06	1.38 ± 0.01	XX
λ= 0.60	2.02 ± 0.01	1.88 ± 0.04	XX

Table 8 : Overall diffusion coefficients for the three different methods

From the results it can be seen that there are no large deviations in the value of D for the different directions. The values for the second hybrid model seem to be slightly lower than for the temperature scaling method and the temperature scaling method gives approximately the same value for D as the atomistic model. This means that the model 2 cannot reproduce the same results as methods that are already validated, although there are minor differences between the two methods. The difference in D for λ = 1 is the most worrying, since this set-up for the hybrid model should be the same as the gromos simulation and should therefore also result in the same diffusion coefficient. A possible difference is the use of user tabulated potentials in the hybrid model, but this should not lead to such a big difference. It is not clear what causes this discrepancy.

Radial distribution function (RDF)
The RDF describes the probability of finding particle B at a certain distance from particle A. In this case it is used to see how the hydroxy groups of the isopropanol interact and to investigate the amount of order in the liquid by calculating the RDF for the centre of mass (COM) of the isopropanol molecules. For the RDFs for the hydrogen bonds the oxygen atom was used as reference and the RDFs for the oxygen and hydrogen atoms from the other molecules were calculated. Two representative graphs are shown in figure 9 and 10. The other graphs can be found in the appendix.
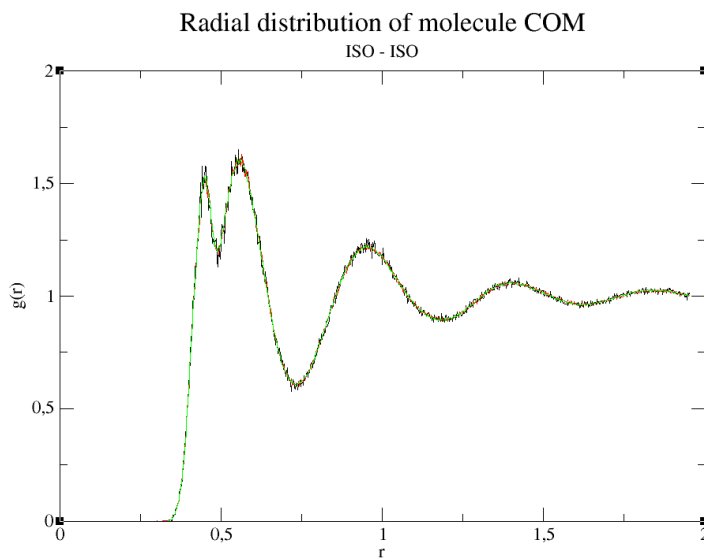
Figure 9: RDF COM-COM for λ = 1 for all three methods
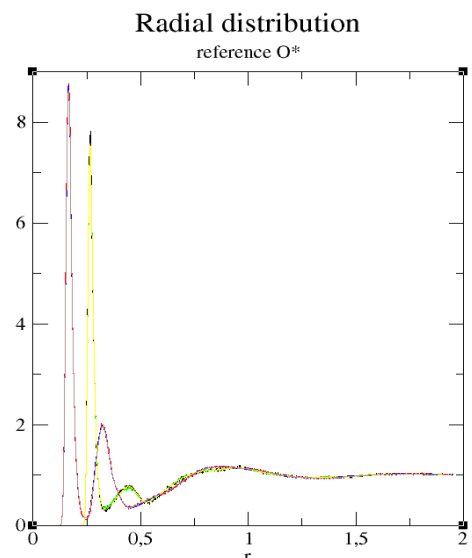
Figure 10: RDF OH for λ = 1 for all three methods

The RDFs for the COMs for λ = 1 are the same for all three methods which shows that the same amount of order can be found in the liquid. This shows that the second hybrid model is able to represent the system well for this value of λ. There seems to be a double peak at r =0.5 which is only visible for λ =1. A possible explanation for this behaviour might be that there are two different conformations in which isopropanol can stack and where one conformation has a slightly smaller distance between the center of masses. The difference between the RDFs for the hydrogen bonds between the T-scaling and hybrid model 2 become larger with a decreasing λ . The difference between the two models at λ = 0.75 might still be acceptable, but at λ = 0.6 the hybrid model 2 has lost most of the atomistic details. This also shows that the hydrogen bonds within the solvent are losing directionality which is deemed an important property of the system. This tests shows that the model can only be used for high values of λ if atomistic detail is to be preserved.

The RDFs for the COM remain the same for both methods. The system seems to become more ordered for increasing values of λ. It was to be expected that the system would become more disordered if the CG character of the system would be increased due to the loss of the directionality of the hydrogen bonds. It is also interesting to see the RDFs for hybrid model 2 and the T-scaling method are the same for the COM-COM RDF, although this was not the case for the hydrogen bond RDFs, which implies that the directionality of the hydrogen bonds is not a determining factor for the overall amount of order in the system. An explanation for the increasing order might be that the molecule loses atomistic character and is represented more by a sphere which might improve the stacking. Overall it can be concluded that the hybrid model 2 behaves the same as the temperature scaling method for the COM-COM RDFs for all tested values of λ and is similar for the hydrogen bond RDF for high values of λ.

Conformational entropy
The conformational entropy is an indication for the spread of different conformations that can be visited by the different molecules. If there is a large difference in the conformational entropy this could indicate unreliable thermodynamic properties. Furthermore it can be used to see what the gained speed-up is by the time necessary to reach the final value of the entropy. [4] The conformational entropy is calculated using the approximation proposed by Schlitter [4]:

$$S_{true} \leqslant S' = \frac{k_B}{2} \ln det \left(I + \frac{k_B T e^2}{\hbar} D \right) \tag{32}$$

where D is the mass-weighted covariance matrix. The graphs can be found in the appendix. The analysis of the graphs can be found in table 9. All the results are for the conformational entropy at fine grained level.

System	Final S ($Jmol^{-1}K^{-1}$)			T (ps, 98% final value)		
	T-scale	Hybrid model	Gromos	T-scale	Hybrid model	Gromos
λ= 1.00	43.8 ± 0.1	43.5 ± 0.1	43.7 ± 0.1	342	335	341
λ= 0.75	44.0 ± 0.1	43.2 ± 0.1	XX	322	316	XX
λ= 0.60	44.0 ± 0.1	43.2 ± 0.1	XX	350	307	XX

Table 9: Values of the conformational entropy for hybrid model 2 and the temperature scaling method

The results show that there is a minor speed-up if the value of λ is increased for the hybrid model 2, which means introducing a more coarse grained character to the system, as was to be expected based on previous results for the temperature scaling method. Previous results for the temperature scaling method showed a speedup for hexadecane. The conformational entropy was calculated on both fine grained and coarse grained level. The speedup was calculated at FG-level with respect to the full FG case for λ = 0-0.75 (same scaling method used as in this report). The results showed maximum speed-up for λ = 0.25[4]. In this case the maximum speed-up for the hybrid model seems to be at lower values of λ than were tested here, so a larger range of λ should be tested to investigate the maximum speedup possible.

15

For the temperature scaling method the maximum speedup seems to be at λ = 0.75. For the other values of λ there seems to be no speed-up and for λ = 0.60 the speedup seems to become negative. The goal of hybrid simulation is to gain a certain speedup, but instead it seems to become slower. This is contradictory to previous results for hexadecane, since there was significant speedup[4]. It is uncertain what causes these results. One of the explanations for the limited speedup might be that this is a very simple molecule as it is with a limited spread in configurations. However this does not explain why there is a minor speedup for the hybrid model. A logical next step would be to test the hybrid model 2 for hexadecane to make a good comparison between both methods. Furthermore there is a difference in final value of the entropy, since these differ between the hybrid model and the temperature scaling method. The temperature scaling method seems to be more accurate if the results are compared to the gromos simulation. This difference is minor, but might indicate unreliable thermodynamic properties for the hybrid model 2.

Conclusion

From the results for the DPPC bilayer it may be concluded that a simple model like this is unable to reproduce realistic results for such a complex system. It seemed that too many assumptions were made and that for a complex system more factors need to be taken into account. One of the issues with the first hybrid model could be the difference in resolution between the lipid and the PW for high values of λ which leads to unreallistic interactions. The results already improved when the AA-CG interactions were disabled, but the system still had difficulty equilibrating which is worrying, because this might indicate several instabilities in the system. Furthermore the lipids became too ordered for low values of λ which is something that is not supposed to happen at the set temperature.

However, the model showed more promise for the pure isopropanol simulations. The results deviated from the reference simulations in some cases, but overall it can be said that the results were promising. There were no issues of equilibrating with these simulations. The diffusion coefficient for the hybrid model was somewhat lower when compared to the reference simulations which is deemed reasonable, except for the fact that at λ =1, the system should be the same as the gromos simulation. It is not clear where this difference in D comes from. The only difference between the gromos simulation and the hybrid model 2 is the use of user tabulated potentials, but these should not cause this discrepancy. The RDFs for the hydrogen bonds of the simulation and the reference simulation were quite similar except for the lowest value of λ where the hybrid model 2 lost most of the atomistic character. The COM RDFs showed great overlap for all values of λ for the different methods and they showed an increase in the amount of order in the system if λ was decreased which might be explained by the more efficient stacking of the CG sphers. The conformational entropy was nearly the same for the different values of λ, but not for the different methods. The hybrid model showed a small speedup when the value of λ is increased, which was to be expected from previous results[4]. However, the difference is small and lower values of λ should be tested to see whether there is a trend. The simulation for the temperature scaling methods only showed a speedup for λ = 0.75 and a small decrease in speed for λ =0,60. This seems contradictory to previous work and defeats the purpose of using a hybrid model. It is not certain what caused this odd behaviour. The hybrid model should be used for hexadecane to be able to compare the results to results from the temperature scaling method to ensure no mistakes were made. One of the reasons for the partially improved results could be that all particles were scaled in the isopropanol simulations which means that there was no difference in resolution between particles. Furthermore this was a pure substance instead of a mixture of molecules which also lowers the amount of different interactions within a system. Overall the results show that the hybrid model can only be used for high values of λ if atomistic detail is to be preserved, but that the speedup will then be minor. Further testing is of course required. The next step would be to test a mixture of small molecules, for example isopropanol and water and have one component fixed at CG resolution and scale the other or to test the hybrid model on a more complex pure substance to investigate the possible speedup.

References

(1) Marrink, S.J.;Risselada, H.J.; Yefimov, S.; Tielemand, D.P.; de Vries, A.H. *J. Phys. Chem. B* **2007**, *111*, 7812-7824

(2) Oostenbrink, C.;Villa, A.;Mark, A. E.; van Gunsteren, W.F. **2004**, *25*, 1656-1674

(3) Rzepiela, A.J.; Louhivuori, M.; Peter, C.; Marrink, S.J. *Phys. Chem. Chem. Phys.,* **2011**, *13*, 10437-10448

(4) Goga, N.; Melo, M. N.; Rzepiela, A.J.; de Vries, A.H. ; Hadar, A.; Marrink, S.J.; Berendsen, H.J. *J. Chem. Theory Comput.* **2015**, *11*, 1389-1398

(5) Wassenaar, T. A.; Ingólfsson, H. I.; Prieβ, M.; Marrink, S.J.; Schäfer, L.V. *J. Comput. Chem.* **2013**, *117,* 3516-3530

(6) Wassenaar, T.A.; Pluhackova, K.; Böckmann, R.A.; Marrinnk, S.J.; Tieleman, D. P. *J. Chem. Theory Comput.* **2014***, 10,* 676-690

(7) Zavadlav, J,; Melo, N. M.; Cunha, A. V.; de Vries, A. H.; Marrink, S.J.; Prapotnik, M. *J. Chem. Theory Comput.* **2014***, 10,* 2591-2598

(8) Van der Spoel, D.; Lindahl, E.; Hess, B. et al. *GROMACS User Manual version 4.6.7*, **2014** www.gromacs.org

(9) Baron, R.; Trzesniak, D.; de Vries, A.H.; Elsener, A.; Marrink, S.J.; van Gunsteren, W.F. *ChemPhysChem* **2007**,*8,* 452-461

(10)Kučerka, N.; Nieh, M.; Katsaras, J. *Biochimica et Biophysica Acta,* **2011**, *1808*, 2761-2771

(11) Risselada, H.J. *Fascinating vesicles?*, Rijksuniversiteit Groningen, Groningen, **2009**, p.2-5

(12) Atkins, P.; De Paula, *J. Atkins' Physical Chemistry,* Oxford University Press, Oxford, **2010**, p. 748, 773

Appendix
Area per lipid
The calculation of the area per lipid can be found in the results and discussion section of this report. The area per lipid has the unit nm^2 instead of nm as is pictured in the graphs.
Hybrid model 1
Note: figure 13-16 show a system which is not fully equilibrated and should be run over a longer period of time. Due to lack of time to run the simulations these preliminary results were put in.
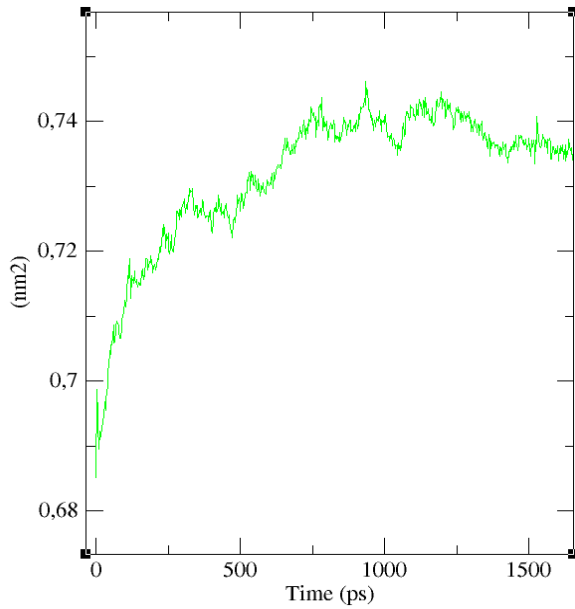
Figure 11: Area per lipid ε$_r$ (AA-AA) = 1.45 ε$_r$ (AA-CG) = 1.0
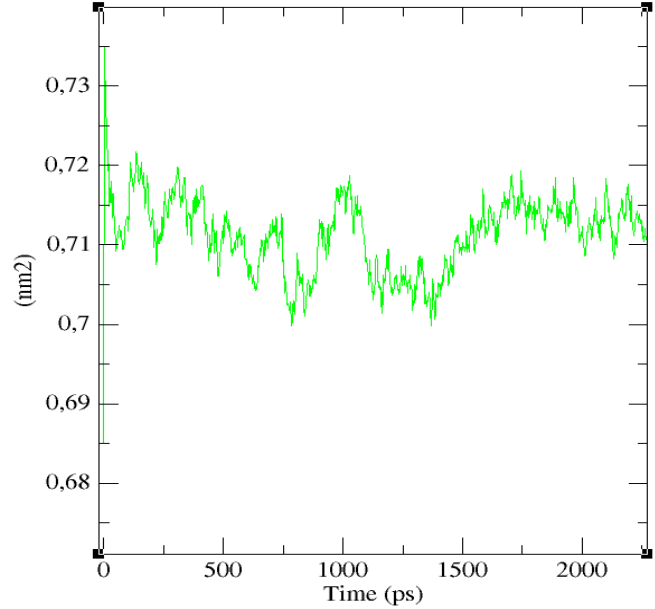
Figure 12: Area per lipid ε$_r$ (AA-AA) = 1.0 ε$_r$ (AA-CG) = 1.45
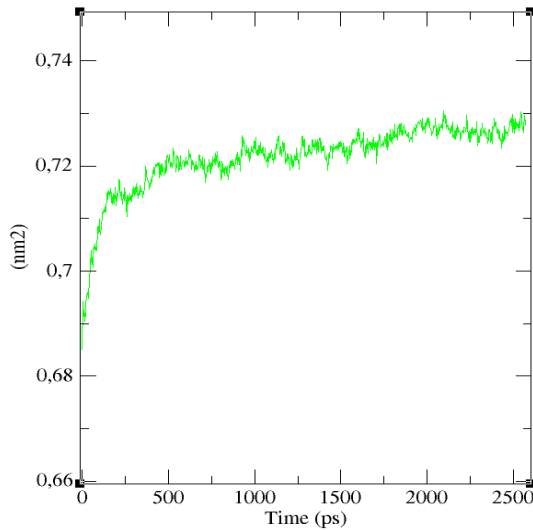
Figure 13: Area per lipid ε$_r$ (AA-AA) = 2.0 ε$_r$ (AA-CG) = 1.0
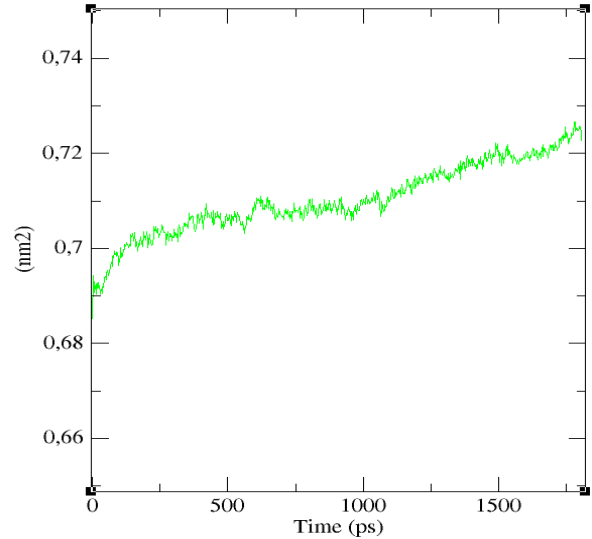
Figure 14: Area per lipid ε$_r$ (AA-AA) = 2.5 ε$_r$ (AA-CG) = 1

18
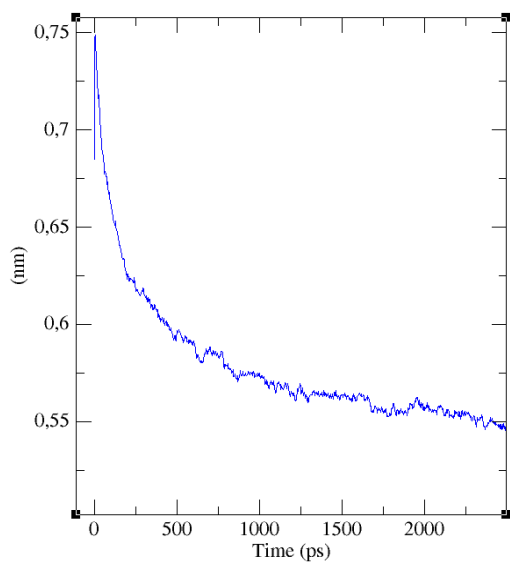
Figure 15: Area per lipid ε_r (AA-AA) = 1 ε_r (AA-CG) = 2.5
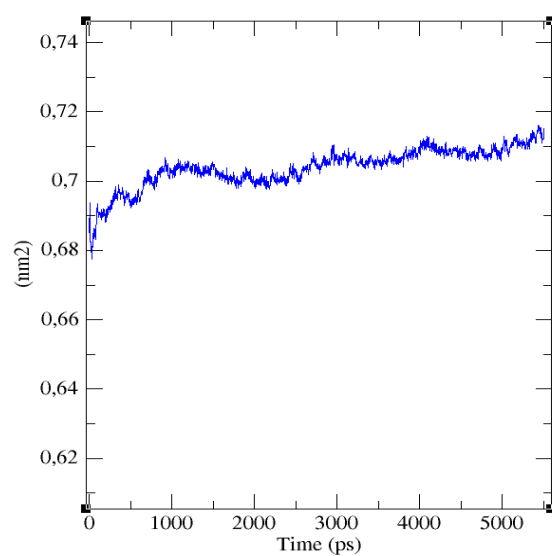
Figure 16: Area per lipid ε_r (AA-AA) = 2.75 ε_r (AA-CG) = 1.0
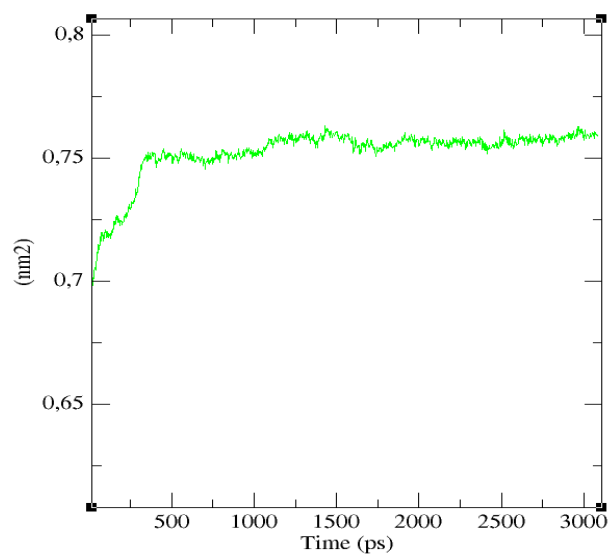
Figure 17: Area per lipid ε_r (AA-AA) = 1.8 ε_r (AA-CG) = 1.0

Hybrid model 2
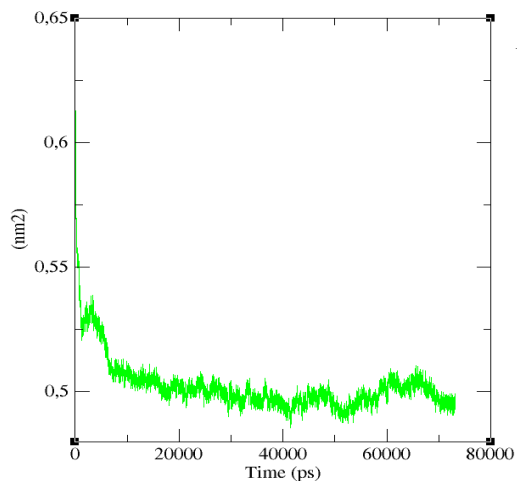
Figure 18: Area per lipid for λ = 1
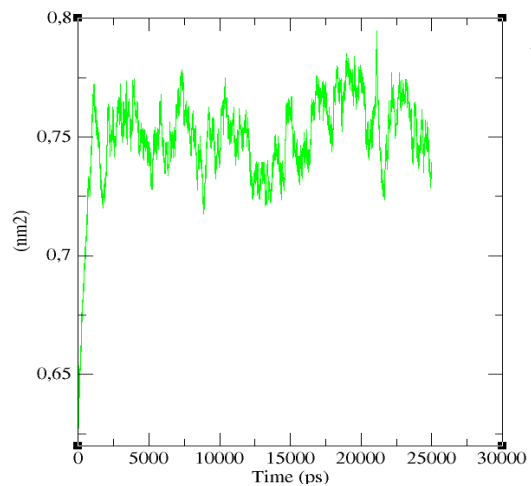
Figure 19: Area per lipid for λ = 0.1
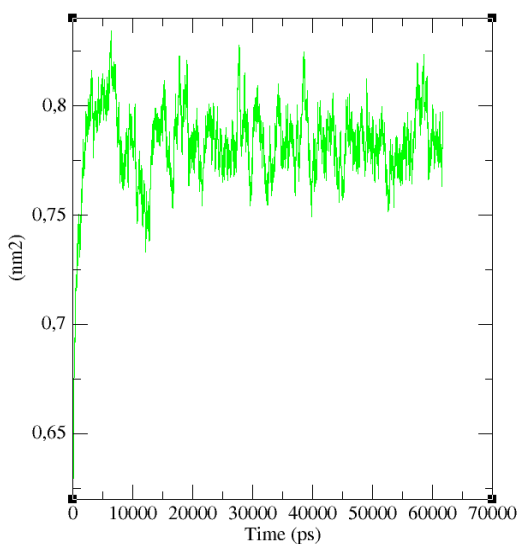
Figure 20: Area per lipid for λ = 0.2
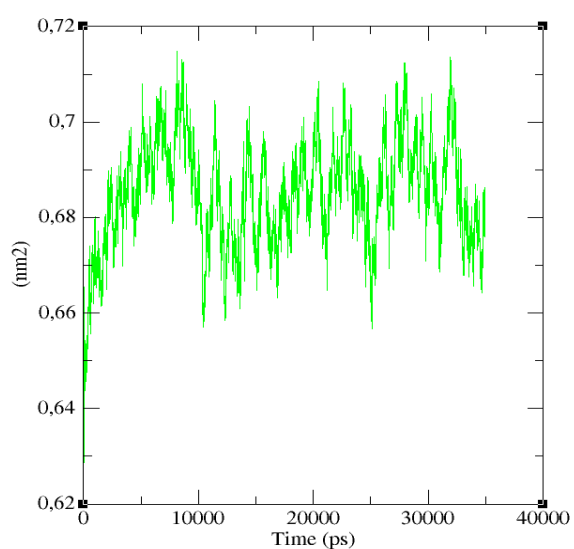
Figure 21: Area per lipid for λ = 0.5
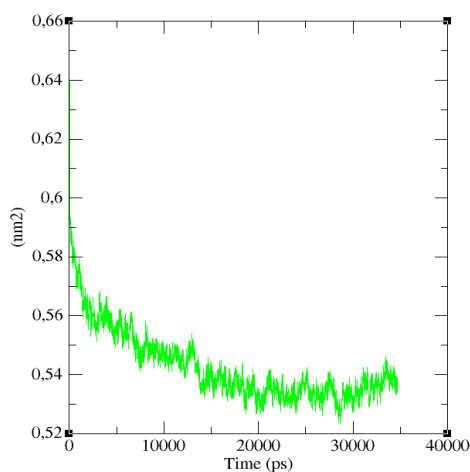
Figure 22: Area per lipid for λ = 0.9

Full coarse grained simulation
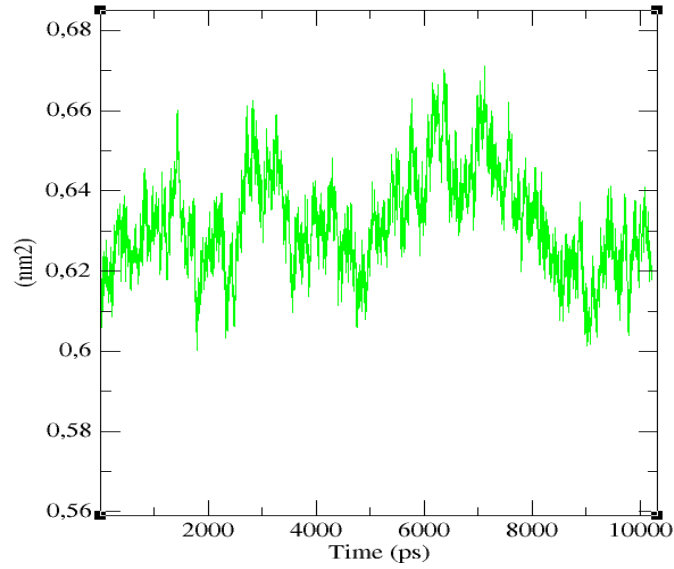
Figure 23: Area per lipid of a CG simulation using the MARTINI model

Velocity distribution
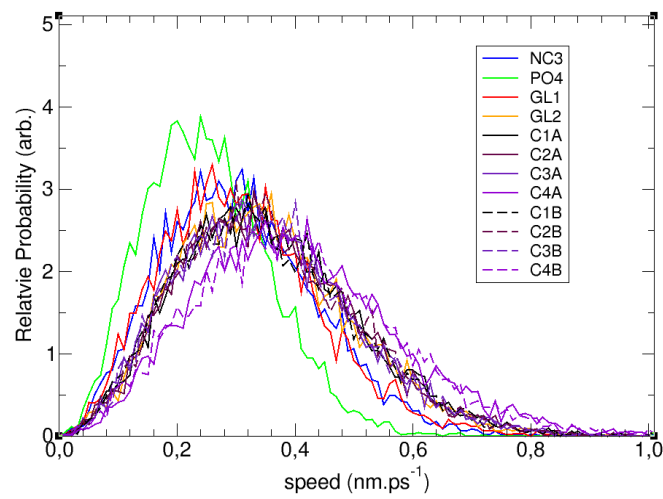
Figure 24: Velocity distribution of the VS for λ = 0.1
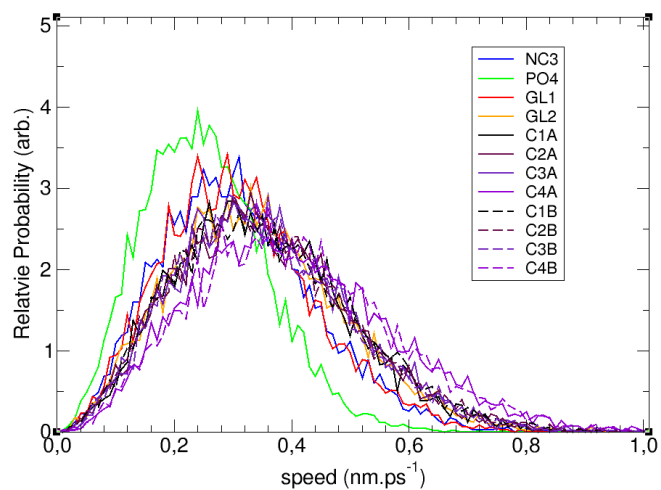
Figure 25: Velocity distribution of the VS for λ = 0.2
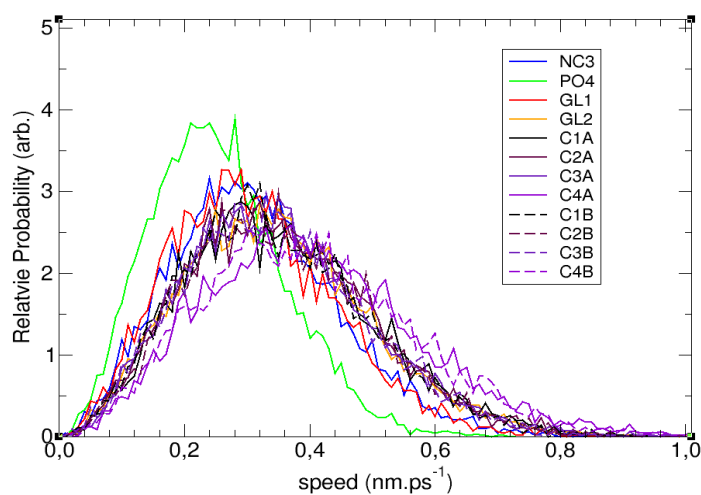
Figure 26: Velocity distribution of the VS for λ = 0.5

Radial distribution function
Hydrogen bonds
Colour coding: H/O Atomistic: Grey/Yellow; Hybrid model: Red/Black; T-scaling: Blue/green
Note: same for all graphs
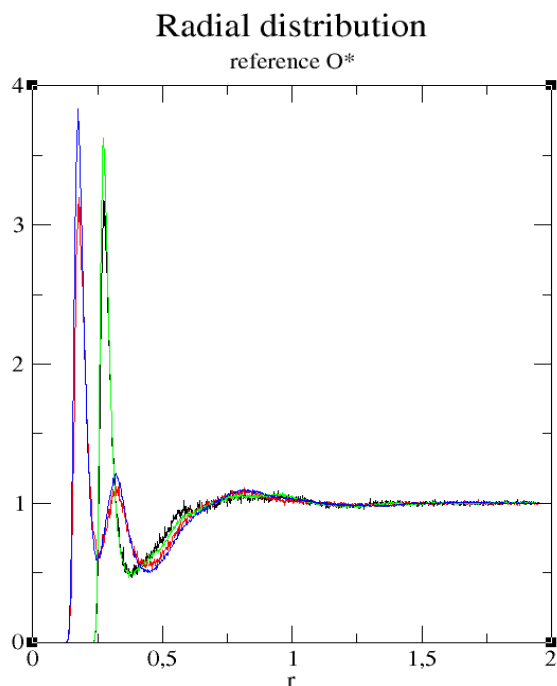
Figure 27: RDF for λ = 0.75
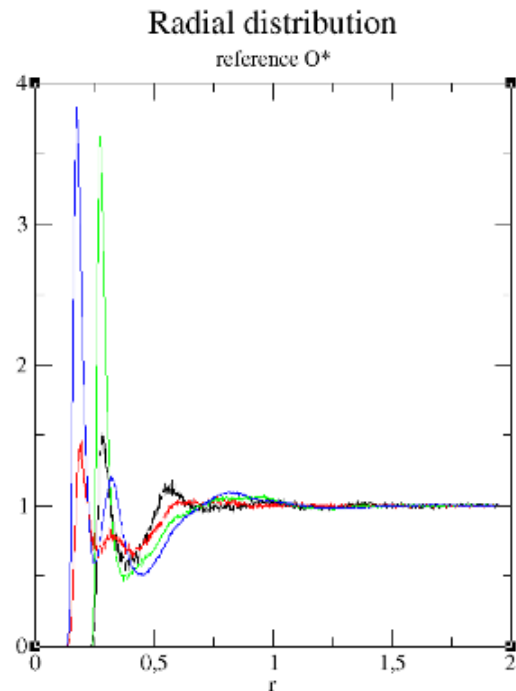
Figure 28: RDF for λ = 0.6

Centre of mass
Colour coding: H/O Atomistic: Green; Hybrid model: Black; T-scaling: Red
Note: same for all graphs
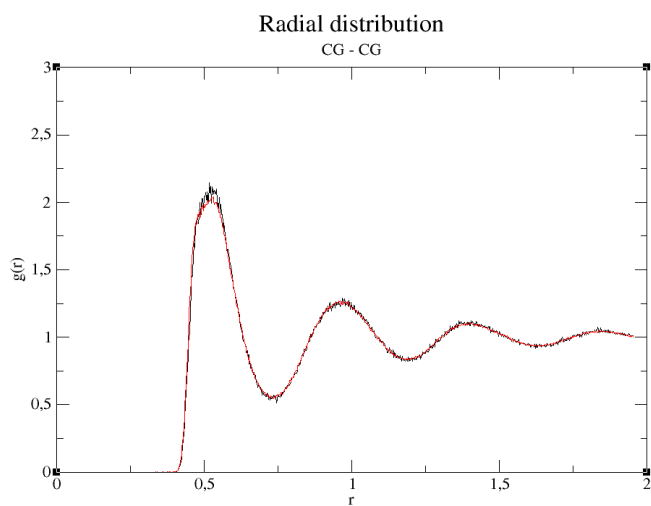
Figure 29: RDF for λ = 0.75
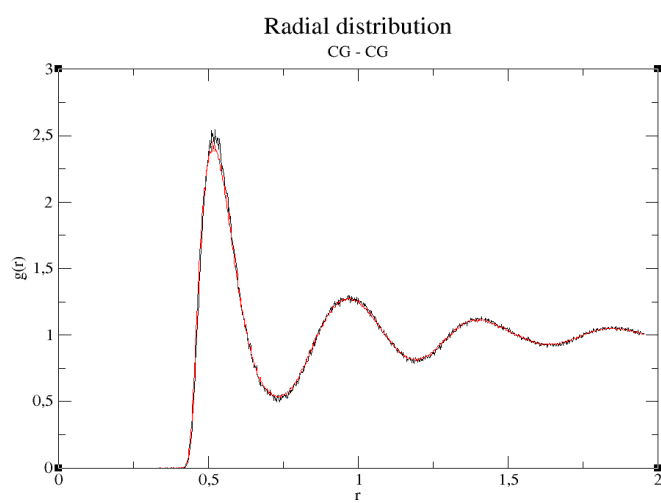
Figure 30: RDF for λ = 0.6

Conformational entropy

In the first graph the upper graph belongs to the atomistic simulation, the lower to the hybrid model 2 and the graph in the middel belongs to the T-scaling method. In all other cases the upper graph belongs to the temperature scaling method and the lower graph to hybrid model two. All graphs consist of three lines: the average value and the variance.
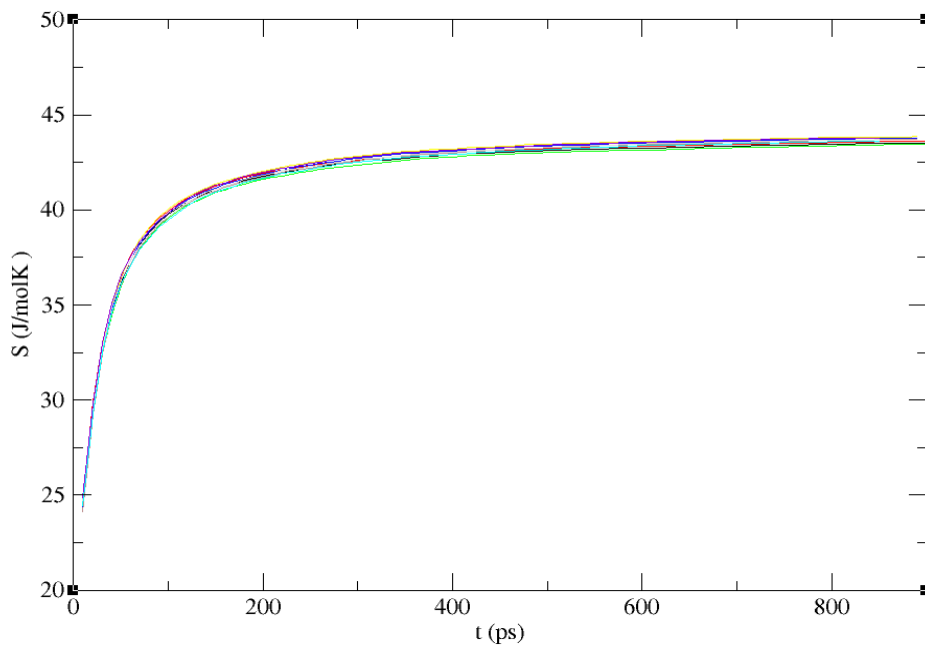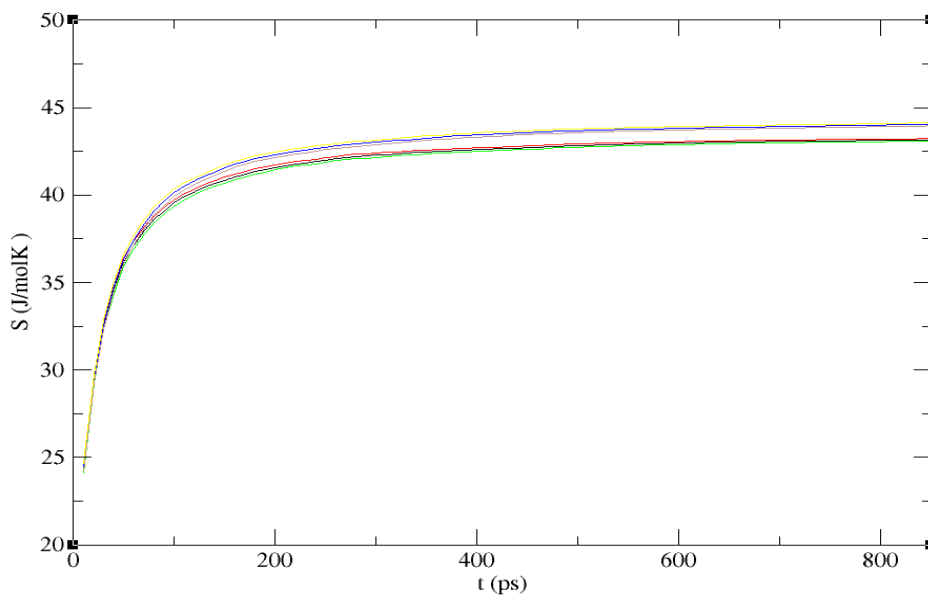

Figure 31: Conformational entropy for λ = 1.00
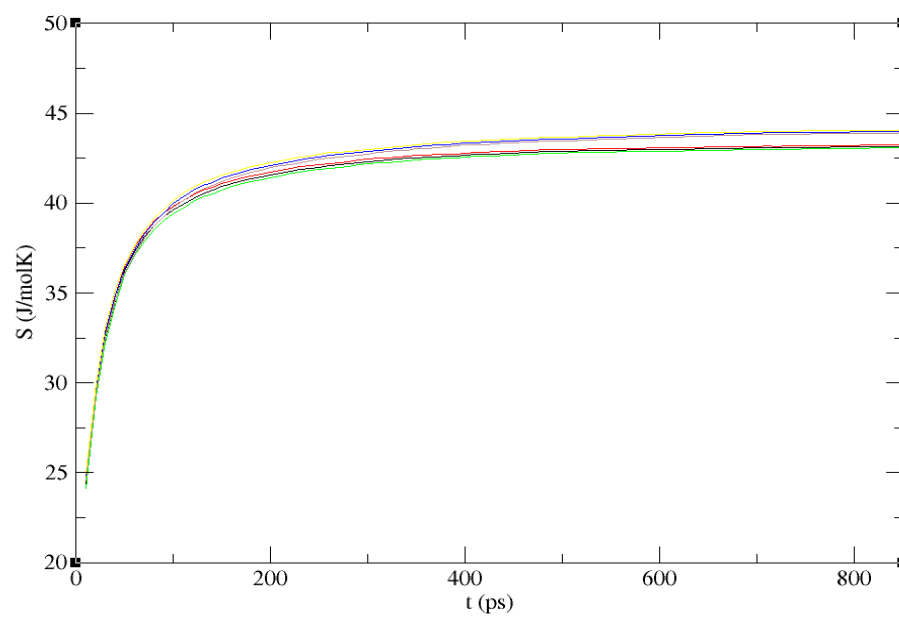

Figure 32: Conformational entropy for λ = 0.75

Figure 33: Conformational entropy for λ = 0.6