Solvation free energy of small alkanes in polar and amphiphilic ambient

Baljkas, Marija

Master's thesis / Diplomski rad

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Split, University of Split, Faculty of science / Sveučilište u Splitu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:166:378082

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Repository / Repozitorij:

Repository of Faculty of Science

zir.nsk.hr

UNIVERSITY OF SPLIT

dabar

DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

University of Split
Faculty of Science

Solvation free energy of small alkanes in polar and amphiphilic ambient

Diplomski rad

Marija Baljkas

Split, rujan 2017.

Zahvalila bih se mentorici doc. dr. sc. Larisi Zoranić na ustrajnom vodstvu, pomoći i savjetima bez kojih ovaj rad ne bi bio moguć. Također se zahvaljujem i prof. dr. sc Franji Sokoliću i mr.sc. Martini Požar na pomoći i dugotrajnim diskusijama koje su iznimno pomogle u pisanju samog rada.

Temeljna dokumentacijska kartica

Sveučilište u Splitu
Prirodoslovno-matematički fakultet
Odjel za fiziku

Diplomski rad

Slobodna energija solvatacije malih alkana u polarnom i amfifilnom okruženju

Marija Baljkas

Sveučilišni diplomski studij Fizika, smjer Računarska fizika

Sažetak:

Sve brži napredak tehnologije doprinio je da se računalne simulacije prometnu u bitan aspekt znanstvenog rada koji se razvija usporedo teorijskim i eksperimentalnim metodama. Simulacije molekularne dinamike daju uvid u mikroskopsko ponašanje sustava koje u eksperimentu možemo samo ugrubo promatrati pomoću makroskopskih mjerljivih veličina. U ovom radu kao alat za simulacije molekularne dinamike korišten je programski paket Gromacs, a sve s ciljem razumijevanja procesa otapanja nepolarnih molekula u polarnom i amfifilnom mediju. Za otopljene tvari odabrani su mali alkani i to metan, etan i propan simulirani pomoću OPLS-AA polja sile. Kao otapalo smo promatrali vodu i metanol. Ova dva otapala su odabrana zbog svoje različite prirode, voda predstavlja polarno otapalo i simulirana je s dva različita modela vode: SPC/E modelom i TIP4P/2005 modelom. Budući da postoji iznimno velik broj različitih modela vode, postoje i brojni članci fokusirani na njihove usporedbe, no unatoč tome nismo pronašli literaturu koja uspoređuje modele na osnovu njihove reprodukcije slobodne energije solvatacije. Ova dva modela su uspoređena sa eksperimentalnim rezultatima za uvjete temperature 300K i tlaka od 1bar i na temelju rezultata smo došli do zaključka da TIP4P/2005 model bolje opisuje kvalitativno i kvantitativno ponašanje slobodne energije solvatacije. Ono što se pokazalo iznimno zanimljivim je da i simulacije i eksperimentalni rezultati ne pokazuju linearni rast ΔG sa porastom alkanskog lanca kao što literature predviđa.

Amfifilna otapala su predstavljena sa metanolom opisanim sa OPLS-UA poljem sile. Simulacije su napravljene za cijeli rang temperatura počevši sa 173K pa sve do 430K uz konstantni tlak od 1bar. Cilj ovih simulacija je bio uvidjeti kako se mijenja ΔG za sustav sa promjenom temperature. Predviđeno je bilo i da će alkani pokazivati bolju topljivost u metanolu upravo zbog njegove amfifilne prirode. Rezultati su ukazali na kompeticijske doprinose slobodnoj enegiji budući da krivulja ovisnosti energije solvatacije o temperaturi pokazuje maksimum. Rezultati ukazuju i na to da se veći alkani lakše otapaju u metanolu što se može objasniti formiranjem lančastih struktura metanola unutar samog sustava. Za ove simulacije nisu pronađeni eksperimentalni rezultati tako da su simulacije uspoređene sa rezultatima drugih znanstvenih grupa koje se bave istom tematikom i pronađeno je kvalitativno slaganje. Svi rezultati su pokazali dobro slaganje sa početnim pretpostavkama, no samo proučavanje simulacija je ukazalo na potrebu za korištenjem dodatnih metoda i programskih alata kako bi se dobio bolji uvod u promjene entalpije i entropije u procesu solvatacije što je dobra polazišna točka za daljni rad na ovoj tematici.

(66 stranica, 21 slika, 7 tablica, 45 literaturnih navoda, izvornik na engleskom jeziku)

Ključne riječi: slobodna energija, solvatacija, simulacije molekularne dinamike, modeli vode, hidrofobni efekt

Mentor/ica: doc. dr. sc. Larisa Zoranić

Povjerenstvo za obranu: doc. dr. sc. Larisa Zoranić,
prof. dr. sc. Franjo Sokolić,
prof. dr. sc. Paško Županović

Rad prihvaćen: 21. 9. 2017.

Rad obranjen: 26. 9. 2017.

Rad je pohranjen u Knjižnici Prirodoslovno-matematičkog fakulteta u Splitu, Ruđera Boškovića 33, 21000 Split, Hrvatska.

Basic documentation card

University of Split
Faculty of Science
Department of Physics

Master thesis

Solvation free energy of small alkanes in polar and amphiphilic ambient

Marija Baljkas

University graduate study programme Physics, orientation Computational Physics

Abstract:

Quick advancement of technology in the recent decades contributed to computer simulations becoming a major aspect of scientific research, developing parallel with theoretical and experimental methods. Simulations of molecular dynamics provide insight into the microscopic behaviour of systems that we cannot see in the experiments. In this work free energy of solvation was calculated by means of thermodynamic integration, with the aim of understanding solvation of nonpolar molecules in polar and amphiphilic solvent. Gromacs was used as a program package for molecular dynamics. As the solute we used four smallest alkanes and as a solvent water and methanol. We compared two different water models based on the accuracy of results and TIP4P/2005 model prevailed over SPC/E model. We also obtained interesting trends of behaviour for solvation of alkanes in the water. Solvation in methanol was done over a range of temperatures with results showing competition between different terms contributing to Gibbs free energy. For those simulations where we had access to the experimental results, comparison of qualitative and quantitative behaviour was made.

(66 pages, 21 figures, 7 tables, 45 references, original in English)

Keywords: solvation free energy, molecular dynamics, water models, hydrophobic effect

Supervisor: Assist. Prof. Dr. Larisa Zoranić

Reviewers: Assist. Prof. Dr. Larisa Zoranić,
Prof. Dr. Franjo Sokolić,
Prof. Dr. Paško Županović

Thesis accepted: September 21., 2017.

Thesis defended: September 26., 2017.

Thesis is deposed in the Library of the Faculty of Science, Split, Ruđera Boškovića 33, 21000 Split, Croatia.

Contents

1 Introduction **1**
1.1 Statistical mechanics and ensembles 1
1.1.1 Microcanonical ensemble 3
1.1.2 Canonical ensemble 3
1.1.3 Isothermal-isobaric ensemble 4
1.2 Thermodynamic potentials 6
1.2.1 Helmholtz free energy F 8
1.2.2 Enthalpy H 8
1.2.3 Grand potential Φ 9
1.3 Gibbs free energy G 11
1.3.1 Connecting the solvation with Gibbs free energy 12
1.3.2 Solvation of small alkanes 14
1.3.3 Ways of calculating free energy difference for solvation 16
1.3.4 BAR method 19

2 Methodology **22**
2.1 Molecular dynamics 22
2.2 Water models 24
2.2.1 SPC/E model 26
2.2.2 TIP4P/2005 model 27
2.2.3 Comparison of water models 28
2.3 OPLS forcefield 30
2.4 Short tutorial for free energy calculation 32

3 Results **35**
3.1 Setup of method 35
3.2 Solvation of alkanes in water 38
3.3 Solvation of alkanes in methanol 40
3.4 Structural properties 42

4 Analysis and discussion **46**
4.1 Alkanes in water 46
4.2 Alkanes in methanol 48
4.3 Hydrophobic effect 50

5 Conclusion **56**

A OPLA-AA parameters for alkanes **62**

B OPLA-UA parameters for methanol **64**

C Numerical results for all simulations **64**

D Parameters used in simulations **65**

1 Introduction

1.1 Statistical mechanics and ensembles

Statistical mechanics connects the observable variables or physical quantities we can measure on a system with microscopic behaviour of particles. In this work we are dealing with classical systems so their behaviour is governed by Newton's equations of motion. Real life systems, for example water in a glass, have tremendous number of particles and calculation of their exact positions and velocities is out of our reach, even with the help of the most advanced supercomputers. Based on this, one would think that we can't know much about microscopic world and the behaviour of atoms and molecules but that would be wrong. Large systems actually show signs of some regularities and predictiveness and we can follow Chandler on his saying that 'this distinctive regularities are consequences of statistical laws governing the behaviour of systems composed of many particles'. [1]. Some thermodynamic observable will be more easily explained with statistical mechanics if it can be described by smaller number of other macroscopic properties. For example, temperature is easily explained with looking at the velocities of particles, while it is also possible to connect it with volume and pressure by means of the equation of state.

System under consideration can be described with few macroscopic properties such as temperature, pressure and volume. If these three quantities have constant value we can say that we have one macroscopic realization of a system. But at the same time, this state can be achieved through numerous microscopic realizations of a system. Molecules in a system can have for example same average velocity in many different ways. If we want to take our research further we need to have some sort of representation of all those microscopic states that are constantly exchanging. For this reason, phase space is defined:

$$(\boldsymbol{r}^N, \boldsymbol{p}^N) \equiv (\vec{r_1}, \vec{r_2}, ..., \vec{r_N}, \vec{p_1}, \vec{p_2}, ..., \vec{p_N}) \tag{1.1}$$

where $\vec{r_i}$ and $\vec{p_i}$ represent coordinates and momenta for particles. Some point in phase space, given by combination of $\boldsymbol{r}^N$ and $\boldsymbol{p}^N$, represents one microscopic state of the system. Over the time system will pass through multiple points and it will draw a trajectory in phase space. This trajectory is something we could be able to calculate if only we had capabilities to solve $\vec{F} = m\vec{a}$ with adequate initial conditions for all particles under consideration. After long period of time, system will visit every point in phase space altough some of the points will be less accessible. To deal with all this we use ensembles. Their description that I like best is maybe the one given by Gibbs who states that "ensemble is an idealization consisting of a large number of virtual copies of a system, considered all at once, each of which represents a possible state that the real system might be in. In other words, a statistical ensemble is a probability distribution for the state of the system."

We can illustrate ensemble with one 'everyday' example. Let's imagine that we have huge closed hall full of random people dancing, and some of them know each other. People will dance all night, trying to avoid conflicts like bumping in others. They will interact in many different ways and we can monitor their behaviour. By taking snapshots of the hall every now and then and looking at the spots in the room they step onto after whole night we can come to different conclusions. If the monitoring time was long enough we can say that there is a huge probability that there will not be a spot in the room where nobody put his foot on. Also, in the end, we can have many different quantities to describe this gathering, like the average velocity with which people moved around or the time they spent talking to each other. This group of people can be described with some average quantities even though we don't know explicitly how each person acted all the time.

This is the exact thing we use ensembles for, we look at the particles and their group journey through the phase space. Measurements can be taken on a system for a long period of time (long from the perspective of the particles) and in the end we make averages which are called ensemble averages. Time average and ensemble average are usually taken as the same thing but that doesn't need to always be true. If a system has this property it's said to be ergodic, and equivalence of time and ensemble averages itself is called ergodicity. Ensemble is then assembly of all possible micro-states that can be achieved by a system under imposed macroscopic conditions. Ensembles can be divided according to macroscopic quantities that hold constant value. Some of the types are:

- microcanonical ensemble (N,V,E)
- canonical ensemble (N,V,T)
- grand canonical ensemble (μ,V,T)
- isothermal - isobaric ensemble (N,P,T)
- isoenthalpic - isobaric ensemble (N,P,H)

On the picture 1 we can see simple illustration of different ensembles and in the next few paragraphs we will go through some of them.

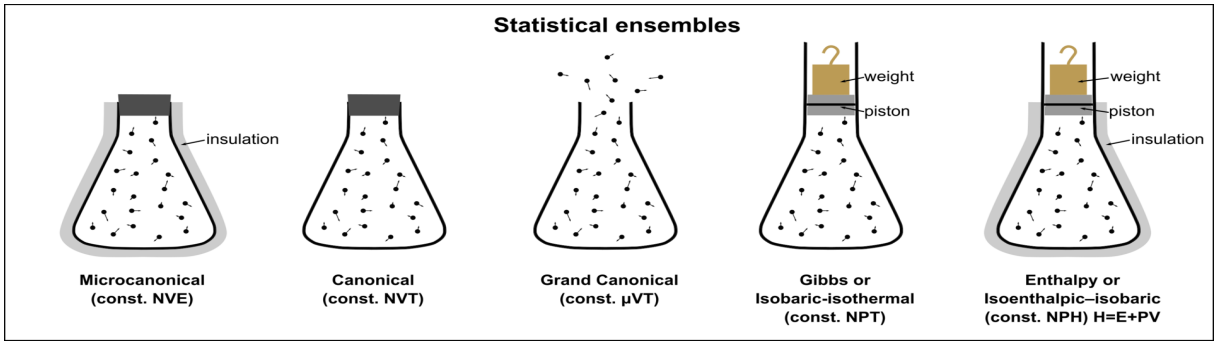


Figure 1: *Simple illustration of different statistic ensembles taken from [2]*

1.1.1 Microcanonical ensemble

This ensemble represents all states that have fixed total energy, number of particles and volume. Statistical postulate states that such system has microscopic states that are all equally likely at thermodynamic equilibrium. If $\Omega(N, V, E)$ is defined to be density of states such that $\Omega(N, V, E)dE$ is the number of energy states with energy between E and E+dE than the probability for some macroscopic state is given with $P_i = \frac{1}{\Omega(N,V,E)}$. For microcanonical ensemble all states with energy $E_i \neq E$ have probability equal to zero.

Connection with thermodynamics for this ensemble is made through entropy:

$$S = k_B ln\Omega(N, V, E) \tag{1.2}$$

Other thermodynamic quantities can be obtained from entropy with use of Maxwell's relations. For example, temperature is obtained as follows:

$$\Big(\frac{\partial S}{\partial E}\Big)_{N,V} = \frac{1}{T} \tag{1.3}$$

$$T = \frac{1}{(\frac{\partial S}{\partial E})_{N,V}} = \frac{1}{k_B(\frac{\partial ln\Omega}{\partial E})_{N,V}} \tag{1.4}$$

This ensemble represents some kind of isolated system, for example box full of water that is not able to exchange heat or particles with its environment.

1.1.2 Canonical ensemble

Canonical ensemble represents a system that has constant number of particles, size and temperature. We can look at it as a system that has immediate contact with huge heat bath that keeps its temperature constant by exchanging heat with it. Important thing for further discussion is the idea that heat bath and our system together form microcanonical system that has constant N,V and E. Canonical system is much smaller and we can say that $E_{tot} = E_b + E_{can} = constant$ and $E_{can} << E_b$. Number of energy states for the bath can be written as $\Omega(E_b) = \Omega(E_{tot} - E_{can})$. Probability for some energy state in canonical system is given by:

$$P_{can} \propto \Omega(E_{tot} - E_{can}) = exp[ln\Omega(E_{tot} - E_{can})] \tag{1.5}$$

Since we know that energy for canonical ensemble is much smaller from both the heat bath and the whole system we can write ln part of the previous equation like the series expansion.

$$ln\Omega(E_{tot} - E_{can}) = ln\Omega(E_{tot}) - E_{can}\frac{dln\Omega}{dE} = ln\Omega(E_{tot}) - E_{can}\beta \tag{1.6}$$

Partial derivation of $\frac{dln\Omega}{dE}$ is taken to be equal to β because the whole system is canonical. From all this we get that the probability for some state in canonical ensemble is proportional to the exponent of its negative energy:

$$P_{can} \propto exp(-\beta E_{can}) \tag{1.7}$$

The above expression is known as canonical distribution law or Boltzmann's distribution law. From it we can see that states with smaller energies will be more probable. The coefficient for proportionality in the above expression can be obtained from the fact that sum of all probabilities needs to be equal to one. Then it follows:

$$P_i = Q^{-1}exp(-\beta E_i) \qquad Q(N,V,T) = \sum_i e^{-\beta E_i} \tag{1.8}$$

Q is called canonical partition function and if we go to continuum limit, the sum will become integral and discrete energy levels E_i will become Hamiltonian for specific configuration of the system.

$$Q(N,V,T) = \frac{1}{N!h^{3N}} \int dr_i dp_i e^{-\beta H(r_i p_i)} \tag{1.9}$$

Connection with thermodynamics for this system is Helmholtz free energy:

$$F = -k_B T ln Q \tag{1.10}$$

and total energy of the system can also be expressed with the help of canonical distribution function:

$$E =< H >= -\frac{\partial}{\partial \beta} ln Q \tag{1.11}$$

1.1.3 Isothermal-isobaric ensemble

This ensemble depicts systems with constant pressure, temperature and number of particles. It is of great importance since these conditions are valid for a large number of systems and it is widely used. We can picture it as a small system inside a big reservoir that keeps its pressure and temperature constant. Constant pressure means that the volume of a system is changing through time. NPT system is much smaller than the whole system so it stands that:

$$E_{tot} = E_r + E_i = const. \quad V_{tot} = V_r + V_i = const. \quad N_{tot} = N_r + N_i = const. \tag{1.12}$$

The distribution of this ensemble is given by:

$$\Delta(N,P,T) = \int_0^\infty Q(N,V,T) e^{-\beta p V} dV \tag{1.13}$$

From the above expression we can see that distribution function for NPT ensemble is connected to canonical distribution function and to be more precise $\Delta(N, P, T)$ is Laplace transform of $Q(N, V, T)$. Thermodynamic quantity associated with this ensemble is Gibbs free energy given by:

$$G(N, P, T) = -k_B T ln \Delta(N, P, T) = F + PV \tag{1.14}$$

On the figure 2 we can see how values for pressure and temperature change during the simulation in NPT ensemble. Duration is 1500ps and we can see that running averages are very close to the values imposed on a system, or in this case temperature equal to 370K and pressure equal to 1bar.

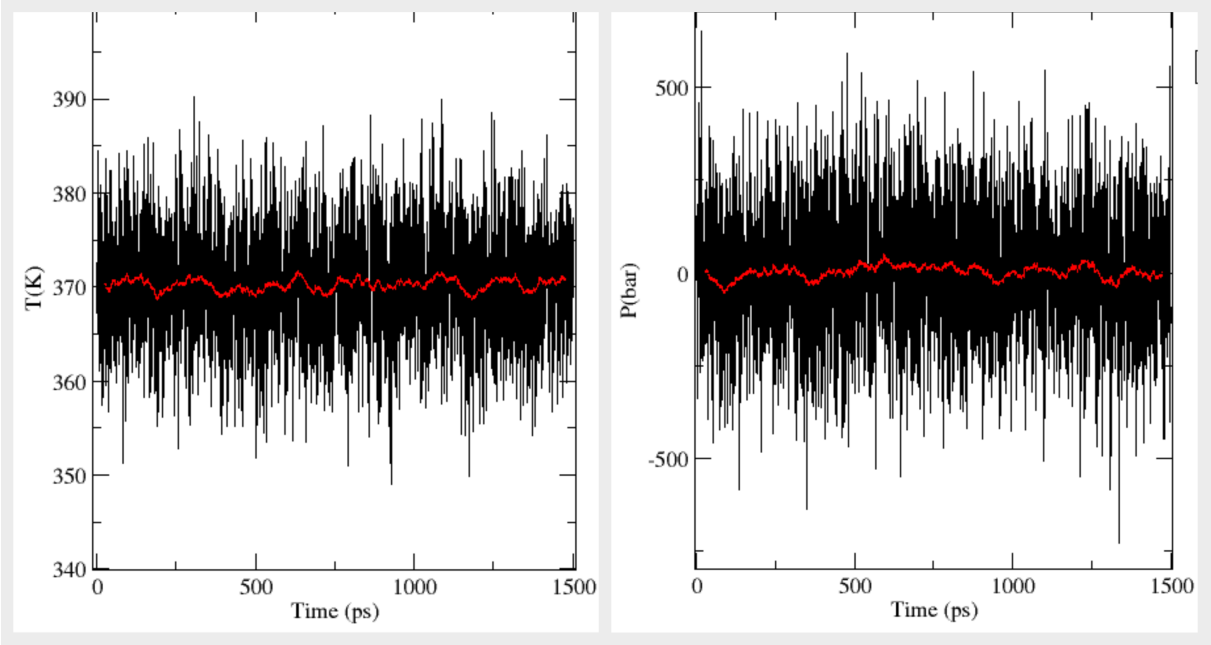


Figure 2: *Values for pressure and temperature alongside running averages for a simulation of methane in methanol at T=370K and P=1bar*

1.2 Thermodynamic potentials

On the contrary to the previous chapter where we discussed statistical mechanics, now we will focus on thermodynamics and already mentioned quantities like U, G, H, F, S that make thermodynamic potentials. General knowledge for this chapter was obtained through college and from literature listed under [1][3][4].

Thermodynamics is mostly about determination of macroscopic state of the system that is described by thermodynamic parameters. Those parameters are mostly known from everyday life and we can divide them in two major groups:

1. State functions
    - temperature, volume, pressure, energy, entropy, ...
    - it's generally valid that their value is the same whenever system ends in the same state. Their value is not dependent on the way system ended in some state.
    - from mathematical point of view, for some quantity f that is state function stands that $\oint df = 0$ and this means that if the system goes from some state A to state B difference in f depends only on the difference between the end points $\Delta f = f(B) - f(A)$ and in the case of reversible process A->B->A difference Δf is equal to zero.
2. Path functions / Functions of process
    - heat, work, ...
    - they are opposite of state functions in the way that their change in some process from state A to state B depends on the way this process happened and not only on end states. For them it stands that $\oint df \neq 0$.

For us to be able to fully understand thermodynamic potentials and other quantities used in this work we need to go over basic thermodynamic postulates or as they are most commonly called three laws of thermodynamics on which further cognitions are based. Newton established his laws way back in 17. century and they are valid for the whole classical physics. On the other hand, thermodynamic laws were put to our knowledge much later, in the middle of 19. century. This is understandable since it's in the human nature to notice and observe phenomena on macroscopic scale sooner. Although thermodynamic quantities, such as heat, energy or temperature, directly impact our lives they are based on the behaviour of microscopic systems with large number of particles. Basic quantities, we first learn about in thermodynamics, are heat, work and internal energy. Their connection is given through first law of thermodynamics (Mayer, Joule, Helmholtz) which states that the total energy of an isolated system remains constant. In other words, heat brought to the system will be used for work done by system and

for change in internal energy of the system.

$$Q = \Delta U + W \tag{1.15}$$

Second law (Thomson,Carnot,Clausius) states that for an isolated system, total entropy can only increase during some time. So for an isolated system Clausius inequality will be valid:

$$\oint \frac{d'Q}{T} \leq 0 \tag{1.16}$$

Entropy is defined as $dS = \frac{d'Q}{T}$ and based on its definition it is state function. Equality sign in the above expression stands for reversible processes. For thermally isolated system entropy will remain constant while in other cases it will enlarge its value until it reaches maximum. State with maximal entropy is equilibrium state for a system. Spontaneous irreversible processes will happen in an isolated system as long as they increase entropy. Third law (Nernst, Planck) came into the light of day with the beginning of 20. century when scientist have started to observe very low temperatures and the limit of absolute zero. It states that for a system at temperature of 0K there is only one microscopic realization and according to $S = k_B lnB$ entropy of a system will be equal to zero. As the temperature falls to the zero entropy will also decrease.

$$S \to 0 \qquad as \qquad T \to 0 \tag{1.17}$$

For a system under observation we can know that it reached equilibrium by looking at the behaviour of variables that we imposed as desired conditions. For example, we can look at the fluctuations of temperature or pressure and if the fluctuations are small and average value is equal to the required one we can say system is equilibrated. But how can we know if processes happening in the unequilibrated system lead system to the equilibrium? For this purpose we can look at the quantities that are called thermodynamic potentials. In this group we encounter internal energy U, enthalpy H, Helmholtz free energy F, Gibbs potential G, grand potential ϕ and entropy S. All of them can be calculated for all conditions imposed on a sytem but each has some specific conditions when it shows great utility. This was already seen in the previous chapter where each ensemble had specific thermodynamic potential associated with it. Common thing for all thermodynamic potentials is that their dimension is equal to the one for energy and they are extensive which means that their value is proportional to the number of particles in the system. We will now go over new quantities mentioned here and those are H,F and ϕ. Gibbs potential will be subject of the next chapter and internal energy and entropy are already mentioned. Before we start there is an expression which will be useful and we can get it by combining first and second law of thermodynamic together. This expression is then called basic relation of thermodynamics:

$$TdS \geq dU + PdV \tag{1.18}$$

1.2.1 Helmholtz free energy F

Helmholtz free energy is often called by free energy only or by Helmholtz potential. It is characteristic quantity for systems with constant temperature and volume or, as we have seen it before, for canonical ensemble. Definition of F can be derived from 1.18 by setting the condition $dV = 0$. From there we obtain:

$$d(U - TS) \leq 0 \Rightarrow F = U - TS \tag{1.19}$$

Its differential is given by:

$$dF = -SdT - pdV + \mu dN \tag{1.20}$$

From the differential of F we can get other thermodynamic variables by means of partial differentiation associated with imposing certain conditions like $dT = 0$. Some of these expressions are:

$$S = -\frac{\partial F}{\partial T}|_{V,N}, \qquad P = -\frac{\partial F}{\partial V}|_{T,N}, \qquad \mu = \frac{\partial F}{\partial N}|_{T,V} \tag{1.21}$$

What is the physical meaning behind F? If we imagine system in contact with big reservoir which keeps it temperature constant, change in F will represent the work done on a system or by a system in a reversible process. From the definition of F we can see that for irreversible processes stands that $dF \leq 0$. From this we can say that the system will lower its free energy in a way that it will try to maximize entropy and minimize internal energy. In an isothermal system spontaneous process will be the one that results in a lower free energy and because of this it can happen that endothermic process which requires energy happens spontaneously if it will reduce F in the end.

1.2.2 Enthalpy H

Enthalpy is characteristic quantity for systems with constant pressure and entropy and that's why we can define it from 1.18 as:

$$d(U + PV) \leq 0 \Rightarrow H = U + PV \tag{1.22}$$

Its differential is given with:

$$dH = dU + pdV + Vdp \tag{1.23}$$

And from differential we can get other quantities by assuming constant conditions:

$$T = \frac{\partial H}{\partial S}|_{p,N}, \qquad V = -\frac{\partial H}{\partial p}|_{S,N}, \qquad \mu = \frac{\partial H}{\partial N}|_{S,p} \tag{1.24}$$

It is useful for describing isobaric and adiabatic systems and processes. In the isobaric

changes enthalpy difference represents the amount of heat that was exchanged between system and surroundings plus the utilizable work. In the case of adiabatic process when system does not exchange heat with the ambience, the change in enthalpy represents only usable work (which does not include the volume work). For some isobar-adiabatic system spontaneous processes will happen as long as they result in the reduction of enthalpy. The equilibrium state of the system is described with minimum of enthalpy.

Enthalpy is of great importance in chemistry where it is used for description of chemical reactions as they are often happening under constant pressure. ΔH is in chemistry defined as $\Delta H = H_{products} - H_{reactants}$ and we can say if some reaction is probable to happen based on its ΔH. If $\Delta H \leq 0$ reaction will be exothermic and it will happen spontaneously and irreversibly. Enthalpy is function of state and its value depends only on difference between end points, no matter what happened in between. Because of this one chemical reaction will always have same ΔH associated with it even though in some cases there could have existed in-between products.

1.2.3 Grand potential Φ

This potential describes systems which have constant chemical potential as one of the imposed conditions. Chemical potential μ itself represents change in energy caused by change in number of particles (simply said change in U associated with adding one more particle in the system). μ is intensive quantity responsible for control of mass or particle equilibrium. Definition of grand potential Φ is:

$$\Phi = U - TS - \mu N \tag{1.25}$$

Its total differential is given by:

$$d\Phi = -SdT - pdV - Nd\mu \tag{1.26}$$

Connection with other quantities:

$$S = -\frac{\partial \Phi}{\partial T}|_{V,\mu}, \qquad p = -\frac{\partial \Phi}{\partial V}|_{T,\mu}, \qquad N = -\frac{\partial \Phi}{\partial \mu}|_{T,V} \tag{1.27}$$

To reach equilibrium this kind of system will need to minimize its grand potential until it reaches some minimal value. This system is maybe hard to visualize but we can describe it as a system in an immediate contact with particle reservoir which allows μ to be constant.

Now that we know basic features of thermodynamic potentials we need to find a way to facilitate our work. All thermodynamic potentials are functions of state, which means if we know their dependence on thermodynamic observables we can easily obtain expressions for other variables. Formulas for these connections are already given in 1.21, 1.24, 1.27. There are even more of them when we include Gibbs free energy. This is not easy to remember but there is a mnemonic diagram to help. Diagram is shown in figure 3. If we start from right side of the square and go in the clockwise direction letters read capital letters of sentence "Good Physicists Have Studied Under Very Fine Teachers". Usage is simple, we take one potential and its derivation with respect to the variable next to it is equal to the variable across the first one. If we go in the direction opposite to the arrows result must have additional minus sign. This way we can get all relations between thermodynamic potentials and other variables with assumption that the number of particles is constant in the system. These relations are called Maxwell's relations.

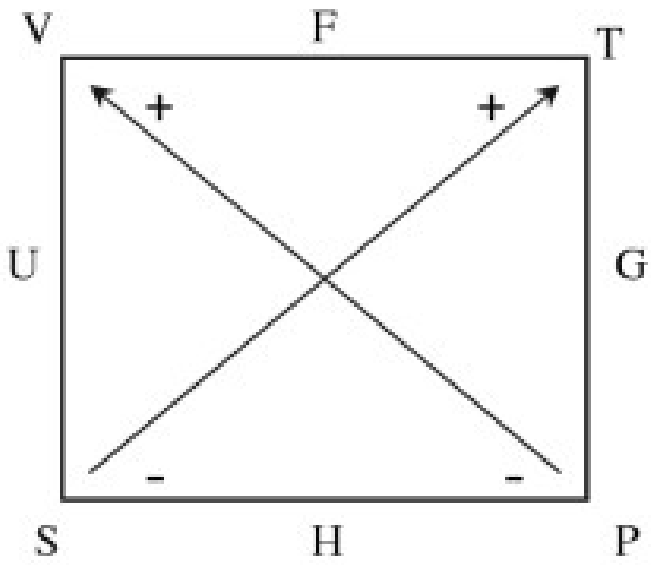


Figure 3: *Diagram of thermodynamic relations, taken from: http://www.eoht.info/page/Thermodynamic+square*

1.3 Gibbs free energy G

Gibbs free energy, also called free enthalpy or Gibbs potential, is the main quantity in this work and that's why it has whole chapter devoted to it. It has been introduced by J.W. Gibbs back in 1875. and it is defined by:

$$G = U - TS + pV \tag{1.28}$$

Its total differential is given by:

$$dG = -SdT + VdP + \mu dN \tag{1.29}$$

Maxwell relations which we can obtain from figure 3 are:

$$S = -\frac{\partial G}{\partial T}|_{P,N}, \qquad V = \frac{\partial G}{\partial p}|_{T,N}, \qquad \mu = -\frac{\partial G}{\partial N}|_{T,p} \tag{1.30}$$

From the definition of Gibbs free energy given by 1.28 we can see some resemblance to the Euler's equation $U = TS - pV + \sum_i \mu_i N_i$. This equation is valid for systems composed of different particle species. We can substitute Gibbs free energy into the Euler equation and we get next expression:

$$G = \sum_i \mu_i N_i \tag{1.31}$$

Gibbs free energy is, according to the above expression, equal to the sum of the product of chemical potential and number of particles for each specie contained in the system. In the case when we have simple one-component system this reduces to $G = \mu N$ which means that G will be proportional to the number of particles in the system and the proportionality factor will be μ. Chemical potential can then be described as free energy for a particle in this kind of a system. Gibbs free energy is obviously extensive since it grows bigger with more particles in the system. In this work we are dealing with solvation so our systems are never composed of only one type of particles. Also, in this work systems have constant number of particles and they are described with NPT ensemble. Because of this, we can ignore the change in the number of particles mentioned in 1.29. Difference in dG then becomes:

$$dG = -SdT + VdP \tag{1.32}$$

Gibbs free energy can be related to enthalpy of the system by inserting its total differential in the above expression. Then the change of Gibbs free energy can be expressed as:

$$\Delta G = \Delta H - T\Delta S \tag{1.33}$$

Physical meaning of Gibbs free energy would be that it represents the work that can be done by the system in isothermal isobaric reversible process. That's why it has 'free' in its name, free work means the energy accessible for some process. In NPT system equilibrium will be reached when free energy gets to its minimal value. ΔG is quantity of great importance in chemistry because big number of chemical reactions is happening under constant pressure (atmospheric pressure) and if these reactions are happening slowly we can assume that the temperature is held constant. Reactions inside a system will happen spontaneously if they result in a reduction of G. The reactions that are happening spontaneously and that end with $\Delta G \leq 0$ are called exergonic, and the ones that end with $\Delta G > 0$ are called endergonic and they will not happen spontaneously.

Now when we went over some of the basic properties of Gibbs free energy it's time to connect it to the solvation process which is the theme of this work.

1.3.1 Connecting the solvation with Gibbs free energy

First we need to look at the basic definition of solvation as stated by IUPAC: 'Any stabilizing interaction of a solute (or solute moiety) and the solvent or a similar interaction of solvent with groups of an insoluble material (i.e. the ionic groups of an ion-exchange resin). Such interactions generally involve electrostatic forces and van der Waals forces, as well as chemically more specific effects such as hydrogen bond formation' [5]. Since solvation is described as stabilizing interaction, ΔG shows as a good choice to describe stability of the system as well as the spontaneity of the reactions happening. It was said in the previous paragraph that in the NPT ensemble reaction will be spontaneous if they result in reduction of G and in the same way we will now look at solvation. If the solvation of some solute in a solvent results in smaller G than in the beginning we can say that this solute is soluble under imposed conditions. There are few ways to define Gibbs free energy of solvation and one of them is that free energy of solvation represents transfer free energy for a molecule from well defined referent state (like ideal gas state) to some other state like solvent. In this case interactions of a solute in the gas state are small so ΔG mostly takes in account interaction with the solvent. [6]. Second interpretation of ΔG for solvation is that it represents reversible work required for solvent to reorganize and to solvate the solute. [7] ΔG is also of great importance for solvation because of its connection with other physical properties used to describe the process of the solvation itself such as Henry's law constants, solubilities and distribution of chemical species between immiscible solvents or different phases. [8]

Change in ΔG is given already with 1.33 and we can see that it's immediately related to the change of entropy and enthalpy. ΔH represents an average measure of potential energy of molecule interaction and the PV work for cavity creation while ΔS is a measure of an arrangement and intermolecular correlations. Because there is a minus sign in 1.33 we can see that there will always be some kind of competition between these two contributions to ΔG. For a process that includes significant changes in a number of molecular interactions or in their strength, for example let's say breaking of hydrogen bonds, ΔG will be dominated by the enthalpy component. Breaking of bonds results in a smaller ΔH and that's why change in Gibbs free energy will decrease with higher temperature (smaller ΔH will also be diminished with the product of T and ΔS). On the other hand if we have a process that keeps the number of bonds and their strength same as before and only effects their arrangement we will have dominant entropic component. [7] If bonds now have strictly defined direction this means that the configurational space is reduced which reduces the entropy. If the change in entropy is negative this leads to greater ΔG with greater temperature.

It follows that a molecule can be soluble even if the process of solvation leads to the greater entropy or in some cases to greater internal energy. All it matters for some molecule to be soluble is the end change of Gibbs energy of solvation to be negative. Change in solvation free energy is independent of ensemble used or conditions imposed on a system.[9] If we were talking about NVT ensemble then we would have Helmholtz free energy to describe the solvation. Although ΔG of solvation gives the same result for different ensembles, ΔH and ΔS change their values in a way that in the end they give same ΔG. On the quest for the literature I found two articles by the same group of authors discussing decomposition of ΔG on entropy and enthalpy change.[9][10] As the authors say there is a great amount of research focused on free energy of solvation itself but very few of them discuss entropy and enthalpy. Their opinion is next: 'Accurate reproduction of entropies and enthalpies of solvation is ,in our opinion, essential if computer simulations are to become a more useful tool to obtain molecular insights into solvation phenomena'.[9] Entropy and enthalpy are both thermodynamic functions of state and as authors state they show more predictive trends opposite to free energy that can have really complex behaviour. They talk about something they call entropy-enthalpy compensation - the empiric observation that small change in free energy of solvation is a result of opposite big changes in ΔH and ΔS. Their research also showed that entropy and enthalpy are much more sensitive to the changes in forcefields used in the simulations and because of that it would be good idea to introduce these two variables as the modelling parameters in the forcefield creation. They also say that in some cases resulting free energy change can be different than $\Delta H - T\Delta S$, and that this observation in biological systems is often connected to the properties of weak interaction such as hydrogen bonding. [10]

1.3.2 Solvation of small alkanes

In this work simulations of solvation of alkanes in two different solvents are done. As a solvent we used water as a polar substance and methanol as amphiphilic substance. In the literature there are not many articles dealing with alkanes and the ones we found are mostly about solvating alkanes in water. For methanol as a solvent not even one article was found. Great thing is that there is Chandler's article published in Nature that deals with hydrophobic effect and its impact on solution of small molecules such as alkanes in water.[7] This section will go over main points made in this article that concern this work and all the ideas that follow are mostly made by author.

Since in this work we are dealing with first four alkanes, they all fall in the category of small solutes. According to the article, interactions between two molecules in homogeneous liquid are happening on distances smaller than 0.3nm-0.5nm and that's one of the reasons why thermodynamic cost of solvating small solute will be small, it's because reordering of the solvent will happen easily. First thing that needs to happen to solvate small molecule in a solvent is creation of a cavity with volume big enough to accommodate solute. For a small molecule this cavity won't disrupt existing hydrogen bonds and they will just reorder around the cavity with their average number held at a constant value. For small molecule, like methane, this cavity has approximate radius of $r \leq 0.5nm$. If a solute molecule is big then the solvation cavity will break some of the existing hydrogen bonds and water will get away from it forming some kind of interface similar to that between water and vapour. This interface has purpose of enabling maximum possible number of hydrogen bonds to form so that the cost of hydration is minimal. Beside the cavity formation, other kind of interaction between the alkane and the solvent will be weak Van der Waals forces for nonpolar molecules or strong attractive forces if we have polar molecules inserted.

For a small inert molecule like alkane we will have important entropic component in the free energy difference because of the rearrangement of hydrogen bonds. The author says that on the room temperature entropic component will be dominant and we can see that from ΔG growing as the temperature grows. This will be seen from the results for both solvents. There is a value of temperature for which this increase in ΔG will stop and ΔG will start to decrease. This is result of the high temperature which makes the solvent uninterested in the maintenance of hydrogen bond number since they are breaking because of the temperature even without the solute. The entropy change ΔS on the room temperature is negative and its value is fairly big but near the boiling temperature it becomes positive which leads to decrease in ΔG.

It is already mentioned couple of times that G is a function of state and because of that its change only depends on the values of G in the end points or, in this case, difference between free energy of clear solvent and free energy of solvent with solute inserted. Difference in free

energy can than be obtained from the partition function of those two states as:

$$\begin{aligned}\Delta G = G_2 - G_1 &= k_B T ln Z_2 - k_B T ln Z_1 = -k_B T ln(Z_2/Z_1) \\ &= -k_B T ln\Big(\frac{\int\int exp(-\beta H_2) dr^N dp^N}{\int\int exp(-\beta H_1) dr^N dp^N}\Big) \\ &= -k_B T ln\Big(\frac{\int\int exp(-\beta \Delta H exp(-\beta H_1)) dr^N dp^N}{\int\int exp(-\beta H_1) dr^N dp^N}\Big) \\ &= -k_B T ln < e^{-\Delta E/k_B T} >_1 \approx < \Delta E >_1\end{aligned} \tag{1.34}$$

Here $< \Delta E >_1$ is the difference between energies of microstates 1 and 2 averaged over ensemble. Basically, we calculate difference between microstates 1 and 2 and over long simulation time system will go over many microstates (in the ideal case it will go over all possible microstates, which means the whole phase space) and we will get true free energy difference between clear solvent and solvent mixed with solute. In the paper author also suggests other expression that leads us to the free energy difference of small alkane. For small molecule that is mostly inert and it doesn't have any kind of strong interaction with the solvent, we can approximate ΔG as free energy needed to form the cavity and energy associated with weak Van der Waals interactions. In this case ΔG is given as:

$$\Delta G = \Delta G_v + \rho \int d\vec{r} g_v(\vec{r}) u(\vec{r}) \tag{1.35}$$

In the above expression $g_v(\vec{r})$ is the average distribution function of the solvent and $u(\vec{r})$ is Van der Waals potential. Since $u(\vec{r})$ represents attractive interaction it always has negative values and that's why attractive forces result in decrease of ΔG. First term in the above expression represents free energy of cavity formation and there are means by which it can also be approximated. This free energy is associated with the probability of finding cavity of volume v inside the solvent. Variables that need to be introduced are: number of molecules found in a volume v in clear solvent $< N >_v = \rho v$ and χ_v or the average value of fluctuations of $< N >_v$. χ_v is given as:

$$\chi_v = \rho v + \rho^2 \int d\vec{r} \int d\vec{r'} [g(|\vec{r} - \vec{r'}) - 1] \tag{1.36}$$

We know that radial distribution function for large values of $\vec{r}$ goes to one so the above expression can be approximated as $\chi_v = \rho v$. With all this in mind the author suggests next expression for the free energy of cavity formation:

$$\begin{aligned}\Delta G &\approx k_B T \Big[\frac{\rho^2 v^2}{2\chi_v} + \frac{ln(2)\Pi\chi_v}{2}\Big] \\ &= k_B T [\frac{\rho v}{2} + \frac{ln(2)\Pi\rho v}{2}]\end{aligned} \tag{1.37}$$

The above expression combined with 1.35 serves as approximation of some analytic expression

for free energy of solvation for small molecule or, in our case, alkane. Until this point it is, I believe, obvious how this article showed useful for this work. The last part that we will go through is Chandler's analysis of values of solvation free energy for smallest alkanes which he addresses as 'the small alkane solubility paradox'. He calls it paradox because of the fact that ΔG for first three alkanes is almost the same and they are all $\approx 2kcalmol^{-1}$. If we assume that ΔG for methane is given with 1.35 than we can write it as $\Delta G_{Methane} = \Delta G_v - \epsilon$. Equation 1.37 gives free energy of cavity formation $\Delta G_v \approx 7kcalmol^{-1}$ and that combined with experimental result for full ΔG for methane gives interaction energy approximately equal to $\epsilon \approx 5kcalmol^{-1}$. Chandler observes ethane as two methane molecules attracted very closely on some distance L and then he states that its ΔG should be equal to free energy of two methane minus the work needed to pull them on such short distance $\Delta G_{Ethane} \approx 2\Delta G_{Methane} + \Delta W(L)$. If this work is again equal to $2kcalmol^{-1}$ this gives that ethane will have free energy of solvation similar to the methane. Same goes for propane if we approximate it as one methane and one ethane molecule coming together at the expanse of the work $\delta W(L)$. In the end author concludes with remark that alkane free energy of solvation in water should scale linerly with the size of alkane chain. We will see later in results that this is valid for some alkanes. Also it must be emphasized that similar ΔG for different alkanes in water is just a coincidence and the results will be very different in other solvent as will also be seen in the case of methanol.

1.3.3 Ways of calculating free energy difference for solvation

Solvation simulations were one of the starting points in the creation of free energy calculation because of their conceptual simplicity. They are computationally challenging since they require long times for sampling of the system. There are many methods for free energy calculation and the one we will use in this work is thermodynamic integration accompanied with BAR method we will cover later. These methods showed good accordance with the experimental results and that's why they were used for parametrization of forcefields, for example GROMOS forcefield had solvation free energies as one of its inputs when it was modelled. [6]

Before going to thermodynamic integration we will briefly go over one of the other methods commonly used in free energy calculation. This method is called Widom method or often particle insertion method. [11] It's based on chemical potential which is for some species a defined as:

$$\mu = \Big(\frac{\partial G}{\partial N_a}\Big)_{P,T,N_b} \tag{1.38}$$

With the previous expression, together with the assumption that N is very big, we can say that chemical potential for extra particle inserted in the system will be equal to:

$$\mu = -k_B T ln(\frac{Q_{N+1}}{Q_N}) \tag{1.39}$$

If we introduce Q as the partition function for the canonical ensemble it follows:

$$\begin{aligned}\mu &= -k_B T ln(\frac{V}{\lambda^d(N+1)}) - k_B T ln\Big(\frac{\int dr^{N+1} e^{-\beta U(r^{N+1})}}{\int dr^N e^{-\beta U(r^N)}}\Big) \\ &= \mu_{id(\rho)} + \mu_{ex}\end{aligned} \tag{1.40}$$

Chemical potential can obviously be divided in two parts, the one which is called ideal because it can easily by calculated analytically and the other one called excess part which is actually the chemical potential connected to the one particle we inserted in the system. Potential energy of the system can also be divided in two terms, the one representing the potential energy of the interactions of extra particle and potential energy of all the others. With this, change in potential energy because of the insertion can be written as $\Delta U = U(r^{N+1}) - U(r^N)$. Integral in the expression for the excess chemical potential can be shortened if we identify the ensemble average and then the expression of chemical potential becomes:

$$\mu_{ex} = -kTln\Big(\int d(r^{N+1}) < e^{-\beta \Delta U} >_N \Big) \tag{1.41}$$

μ_{ex} is equal to the change in free energy introduced with the insertion of a particle. It is clear that this method is suitable for solvation simulations since it is easy to insert one solute molecule in already existing solvent and then just calculate the averages under the integral sign. This method is often used and in this work we will show some results from other research groups that were obtained with this method. Now we will go over to the method we use for free energy.

Thermodynamic integration as a method for calculation of ΔG relies on the fact that ΔG depends only on difference between boundary values of free energy on some path through the phase space. Maybe most important property of this method is the one that this path between two ending points does not need to be physical at all when dealing with simulations. In real life experiments, some process will of course be physical but in the case of computer simulations we can take any imaginable path between two states and simulate it. Easy way to do this would be to take a look at Hamiltonian of the system and then choose one parameter that we will change in a way that ending points are exactly systems that we have for referent ones. The way it is often done is by introducing some coupling parameter λ which has boundary values $\lambda = 0$ and $\lambda = 1$. It must satisfy the condition that Hamiltonian with $\lambda = 0$ describes our referent system or in the case of solvation that is pure solvent with no solute inside. Also Hamiltonian with $\lambda = 1$ must be valid for our system under consideration or solvent with solute inside it. What do we control with this parameter in solvation simulation? λ parameter actually controls the strength of interactions between the solute and the solvent. In the case when λ is equal to zero, solute is not included in any kind of interaction (neither Coulomb or Lennard Jones) so in that case solvent doesn't have a clue that solute is actually in a system. With increasing λ we increase the strength of these interactions and they are in their full strength when $\lambda = 1$. If we

know that ΔG is related to partition function according to 1.14 we can write its differentiation with respect to λ as:

$$\begin{aligned}\Big(\frac{\partial G}{\partial \lambda}\Big)_{N,P,T} =& \frac{\partial}{\partial \lambda}\Big(\frac{-1}{\beta} ln\Delta(N,P,T)\Big) \\ &= \frac{-1}{\beta\Delta(N,P,T)}\frac{\partial\Delta(N,P,T)}{\partial\lambda} \\ &= -\frac{1}{\beta}\frac{\int Q(N,V,T)e^{-\beta pV}\frac{\partial H(\lambda)}{\partial\lambda}dV}{\int Q(N,V,T)e^{-\beta pV}dV} = \\ &= -\frac{1}{\beta}\Big\langle\frac{\partial H(\lambda)}{\partial\lambda}\Big\rangle_{\lambda}\end{aligned} \tag{1.42}$$

From the above expression we can see that difference in free energy od solvation can than be obtained through integration of the last expression. This method is useful because of the fact that it can be used for every possible Hamiltonian we can think of as long as it is differentiable and as it satisfies the boundary conditions for $\lambda = 0$ and $\lambda = 1$. Some calculations of liquids can take as a referent state the ideal gas phase because there are no strong interactions between molecules but sometimes some calculations for solvation free energy even take the solid phase as an initial referent state. These calculations then require two steps, one simulation where we look at the transition from solid to gas phase and then the one from gas to liquid. Thermodynamic integration can also depend on the direction we choose to go with, meaning that it would be better option to take in account both directions, from state 1 -> 2 and from 2 -> 1. This is why BAR method shows great accuracy, it uses information from both directions to calculate ΔG. It has bigger efficacy than other commonly used method called FEP (free energy perturbation, also known as exponential averaging or EXP). In the next few paragraphs we will shortly go over main properties of BAR method.

1.3.4 BAR method

BAR method is first mentioned back in 1976. by Bennett and its full name is Bennett's acceptance ratio method. [12] It is based on a fact that for some classical system we can divide its Hamiltonian to potential and kinetic part. Kinetic part is easy to solve:

$$Q_{kin} = \int\int exp\Big(\frac{-\beta(p_x^2+p_y^2+p_z^2)}{2m}\Big)dp^3dr^3 = (2\pi mkT)^{\frac{3}{2}}V \tag{1.43}$$

and with the above expression we are only left with the configuration part which is for canonical system given with:

$$Q = \int exp[-U(q_1,\ldots,q_N)]dq_1\ldots dq_N \tag{1.44}$$

where U stands for potential energy of the system divided by kT. This integration represents a problem in most of simulations because we need to sample big part of the phase space and we know that there are some points with higher energies and consequently with small probabilities for the system to reach them. Even if the sampling is done over big part of the phase space, remaining parts can lead to significant contribution to the integral. This problem is solved in this method with defining the ratio of two such integrals for potential functions U_1 and U_2 that are defined on the same configuration space. Bennett's method is in its original paper written for Monte Carlo simulations so following expressions will contain quantities specific for that kind of simulations. MC simulations give each trial step some kind of probability by means of Metropolis function. In the case when we want steps to have Boltzmann probability that function will look like $M(x) = min\{1, exp(-x)\}$. One step where system keeps its configuration and only switches its potential function from U_0 to U_1 needs to satisfy following condition:

$$M(U_1-U_0)exp(-U_0) = M(U_0-U_1)exp(-U_1) \tag{1.45}$$

If we integrate previous expression exponential parts will give configuration integrals and Metropolis function will be averaged over ensemble which will result in new expression:

$$\frac{Q_0}{Q_1} = \frac{\langle M(U_0-U_1)\rangle_1}{\langle M(U_1-U_0)\rangle_0} \tag{1.46}$$

Physical meaning behind this formula is that if we have MC simulation which will contain steps changing the configuration of the system but also, in some ratio, steps switching the potential, configurations of the unknown system 1 and configurations of the referent system 0 will be distributed according to the ratio of their configurational integrals. There also exists more general form of the above expression that includes the weight function $W(q_1, q_2, ..., q_N)$ which is defined as finite in the whole configurational space. Then we can write previous

expression as:

$$\frac{Q_0}{Q_1} = \frac{Q_0 \int W exp(-U_0 - U_1) dq^N}{Q_1 \int W exp(-U_1 - U_0) dq^N} = \frac{\langle W exp(-U_0) \rangle_1}{\langle W exp(-U_1) \rangle_0} \tag{1.47}$$

From here only thing we need to do to get to some usable expression for Q_0/Q_1 is to choose good weight function. If the choice falls on:

$$W(q_1, \ldots, q_N) = const * \left(\frac{Q_0}{n_0} exp(-U_1) + \frac{Q_1}{n_1} exp(-U_0) \right)^{-1} \tag{1.48}$$

then the ratio of configurational integrals and its error can be written with following expressions:

$$\frac{Q_0}{Q_1} = \frac{\langle f(U_0 - U_1 + C) \rangle_1}{\langle f(U_1 - U_0 + C) \rangle_0} exp(+C) \quad with \quad C = ln \frac{Q_0 n_1}{Q_1 n_0}, \quad f(x) = \frac{1}{1 + exp(x)} \tag{1.49}$$

$$\sigma^2 = \frac{2}{n} \left[\left(\int \frac{2\rho_0 \rho_1}{\rho_0 + \rho_1} dq^N \right)^{-1} - 1 \right] \tag{1.50}$$

symbol f in the above expressions is Fermi function, constant C is usually chosen arbitrarly but this is the value that minimizes the error. In the expression for error ρ stands for density $(1/Q)exp[-U(q_1 \ldots q_N)]$. With all this in mind, one asks himself what is the connection with free energy of solvation? We know that difference in free energy between two states is relatable to the ratio of partition functions. Since this article for BAR method is made for NVT ensemble, we can use Helmholtz free energy whose difference is given with $\Delta F = F_1 - F_0 = k_B T ln(Q_0/Q_1)$. We can now see the way how BAR method samples the system and calculates free energy difference at least for the Monte Carlo simulations.

In the article Bennett himself shows possible problems for this kind of simulation. According to him there are few important choices one need to make with great caution: what systems to choose for referent and mediate, which algorithm to use for sampling of the system and how to evaluate free energy with great precision. His method imposes few conditions on the ensembles used and those are that ensembles need to show big overlap among them and they need to be similar enough so that density of states can be smooth function of $\Delta U = U_1 - U_0$. Research that almost always goes along with the BAR method is the one done by Wu and Kofke on the overlap of phase space between two ensembles.[13][14] Their work takes in account that we have two systems A and B. Each of them has its configurations in phase space and we can narrow that space down to the 'important' phase space. By this we mean the part of the phase space that largely contributes to the partition function of the ensemble on which we judge by means of Boltzmann factor. The 'important' phase spaces of these two systems can satisfy one of the following relations:

- Subset relation - phase spaces are connected asymmetrically and B is subset of A. In this case if we want to have good sampling the best approach would be to sample system A and then use perturbation to sample system B. This way we will get all important

configurations for B. Still sampling needs to be long enough for all important points in phase space to be covered.

- Partial overlap relation - phase spaces of A and B are just partially overlapped and we will never be able to sample all important configurations of both of them, no matter which system we choose to be referent.

- Nonoverlap relation - if the important phase spaces don't show overlap at all every try of free energy calculation will result in catastrophic results.

The overlap of phase spaces can be measured in many different ways and one of the commonly used methods is measuring the relative entropy which was originally introduced in the information theory. Simply put with this method we sample both systems and we look at the number of configurations of A that also appeared in the sampling of B. Opposite way is also done. Sampling needs to be done for a long time for method to be considered accurate. In the end we watch at the overlap of obtained distributions and from there we make a decision on the overall overlap of two systems. Relative entropy is something that can be often seen as part of the output of free energy calculation.

2 Methodology

2.1 Molecular dynamics

Molecular dynamics simulation is a technique for computing the equilibrium and transport properties of a classical many-body system [15]. Many-body system is basically the same thing as N-body system with number N usually being more that hundred. In N-body problems we have bodies that interact with each other and we try to solve their equations of motion. When we say that molecular dynamics is classical technique we mean that bodies obey the laws of classical mechanics with no quantum effects.

Molecular dynamics started to develop in 1950's when researchers simulated behaviour of hard spheres whose potential was zero for all values of $r > \sigma$ and ∞ for all $0 < r < \sigma$. Although hard sphere potential wasn't realistic it marked the beginning of a long search for the best potential. Next stepping stone was implementation of first realistic potential in 1964. when Rahman simulated liquid argon. After that molecular dynamics was developing at constant rate and it became important tool in science. Many models and potentials were proposed but there is still no unique potential which shows good accordance with every property we want to simulate. Even 'popular' molecules, such as water, aren't fully described and we will talk more about it in the next sections. Today, simulations have important status in scientific method. They can be used as a control for theory when compared to experimental results but also to modify existing theories. Simulations are in many ways similar to 'normal' experiments. We have a system we want to perform our calculation upon and first we need to prepare it for simulation. System usually consists of big number of molecules and they need to placed under imposed conditions like temperature, pressure, density etc. depending on desired ensemble we wan't to simulate. Temperature of a system at moment *t* is expressed through velocities with expression:

$$k_B T(t) = \sum_{i=1}^{N} \frac{m v_i(t)^2}{N_f} \tag{2.1}$$

where $v_i(t)$ is velocity of particle i at t and N_f is number of degrees of freedom. As we can see, temperature depends on velocities so it's logical that thermostats work on principle of velocity scaling. Popular thermostats are: simple velocity scaling (used in this work), Berendsen thermostat and Nose-Hoover thermostat. Pressure of a system is expressed through momentum and forces acting on each particle by:

$$P_{int} = \frac{1}{3V}(\sum_{i=1}^{N}(\frac{p_i^2}{m_i} + r_i F_i)) \tag{2.2}$$

We can see that pressure depends on box vectors, so barostats will change the length of those vectors and with that change the pressure of a system. Commonly used barostats are Berendsen,

Hoover and Parrinello-Rahman (used in this work). In every step of simulation the forces are calculated for each particle. Particles act mutually so for a system with N particles we need to calculate $N(N-1)$ forces. Calculation of forces is the most time consuming part of simulation. It can be optimized in different ways but it is important to choose cut-off radii for long-range and short-range forces. Integration part, where we actually calculate new positions and velocities of particles, is done with algorithm of our choosing. Algorithms for integration, usually called integrators, differ in number of substeps needed for calculation of one time step and by propagation of error with time. Error (expressed as time step to some power, for example dt^2) is inversly proportional to the number of substeps, which means that if we calculate new positions with greater number of steps we will have smaller error. Of course, by choosing complex algorithm we need to be prepared that our simulation will require longer time to finish. Some of the most popular integrators are: Verlet algorithm($\mathscr{O}(dt)$), velocity Verlet ($\mathscr{O}(dt^2)$) and Leap-frog ($\mathscr{O}(dt^3)$). In this work Leap-frog is used.

From everything mentioned until now we can see that choosing best options for simulation is not easy or simple. It is possible to write new program for molecular dynamics from scratch but it would require thousands of lines of code to make it optimized, correct and of course, suitable for different calculations. There are many open source projects that can be found online but one that really stands out from the crowd and its name is Gromacs.

GROMACS stands for 'GROningen MAchine for Chemical Simulations' and it was developed as a tool for simulations of biological molecules such as proteins. Its development started back in the 90's at University of Groningen but as the project grew to bigger dimensions huge number of scientists joined and now Gromacs counts more than 1,5 milion lines of code [16]. One of its main advantages is certainly high performance. Simulations can be run on CPUs or GPUs since it supports parallelism. With the latest big update, version 5.0, the process of translating the code in C langauge started which makes Gromacs even more versatile. Gromacs handles every part of simulation process from system configuration to result calculation. It will be shown in one of the later chapters how free energy calculation looks like in Gromacs.

2.2 Water models

Although great number of molecules is nowadays simulated for different kind of research, there is no molecule that has had so much attention like water. One would expect that with so much effort put into determination of water properties, there is a unique model used for all kinds of simulations. Sadly, that is not true. Today we have over 50 models of water and in the next few sections it will be specified what are their most common differences and what are pros and cons for two models used in this work.

Development of water models started back in 1930's but until 90's it was going at a slow rate. Today we have models developing constantly as computer power is growing year by year just like the need for more precise results. We can divide models in separate groups according to the number od parameters they have, method of their development etc. By method of development there exist empirical models (their parameters are fitted to reproduce experimental results) or ab initio models (models built 'from scratch',without any assumptions, based on the basic laws of quantum mechanics). Number of parameters can vary from 3 in simple models to over 50 in precise, complicated models.

If we take something from nature and we try to describe it with a set of numbers it is logical that the greater the number of parameters our description will be better. Based on that we see that complex models give better results but one must bear in mind that complicated models require longer computational time. Because of this, we must adjust our need of accuracy with time we have at disposition for calculation. Today, simple models are used for long simulations of large number of molecules (even over 10 000 molecules), and complicated models for shorter simulations which require quantum mechanical or nuclear assumptions.

Simple models have few common parameters: O – H bond length, H-O-H angle, point charges whose number can vary, σ and ϵ for Lennard Jones interaction and mass. When empirical models are developed one can use already known experimental values for bond length or angle but in some cases those parameters are fitted according to results. Main principle is to choose some water property for which experimental data is available and then change simulation parameters all the way until agreement is met between simulation and experiment. There is a few properties that are often used like liquid density at ambient condition, vaporization enthalpy or temperature where water shows maximum density.

Simple models that are used in this work are rigid, empirical models. Rigid means that molecule is not flexible, it doesn't have internal vibrations. Simple rigid models can be divided in four groups shown in figure 4. Models shown in a, b and c are planar while the one shown in d is tetrahedral. Model a has three point charges in the places of hydrogen and oxygen atoms. Hydrogens carry negative charge whose value is half of oxygen charge. Main parameters in this model are charges, bond length, bond angle and σ. Models shown in b and c differ from a in the fact that they have extra dummy point that carries positive charge instead of oxygen. That

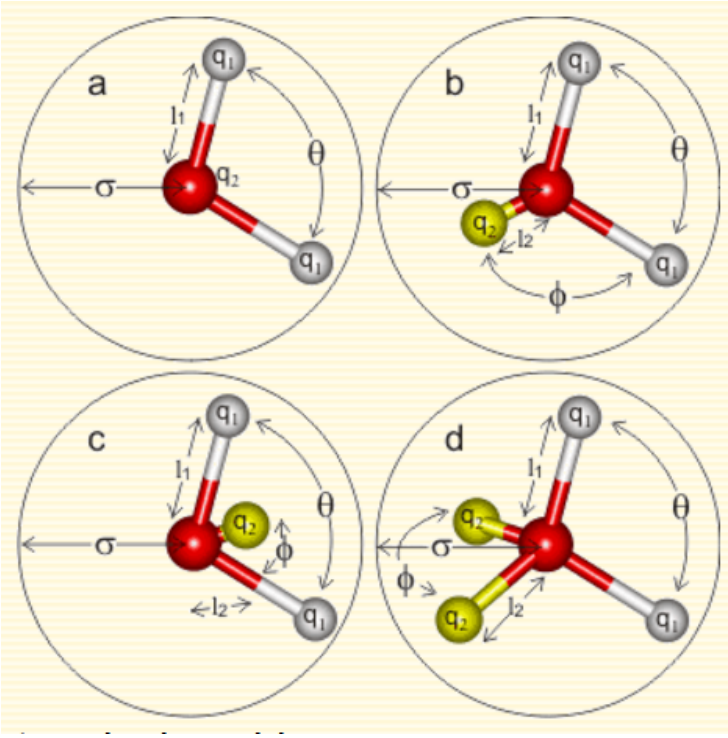


Figure 4: *Division of simple rigid water models, image taken from [17] in June,2017.*

dummy point doesn't have mass or Lennard Jones parameters so it only participates in Coulomb interaction. It can be placed at bisector of H-O-H angle as shown in model c or in middle of wider H-O-H angle. Obviously these types of models have extra parameters such as distance from oxygen to dummy point and angle at which it is placed. Last type of models (shown in d) has two dummy points each carrying half of the oxygen charge.

2.2.1 SPC/E model

SPC/E model (Simple Point Charge Extended) is first introduced back in 1987. by Berendsen, Grigera and Straasma [18]. It consists of three point charges so it falls in first category from picture 4. Authors developed SPC/E after seeing that SPC model doesn't give correct results for both energy and density at the same time. They were searching thouroughly for explanation why is that happening and they came to conclusion that effective pair potentials used in models don't include polarization. They ended up with the fact that the total energy of a molecule should be written as the sum of the full electrostatic interaction and the positive self energy E_{pol} which represents the energy the molecule needs to be distorted to its polarized state. If electrostatic interaction and polarization correction term are given by:

$$E_{el} = \frac{1}{2}\sum_i q_i v_i \tag{2.3}$$

$$E_{pol} = \frac{1}{2}\sum_i \frac{(\mu - \mu^0)^2}{\alpha_i} \tag{2.4}$$

then total energy of the molecule is given with $U = E_{el} + E_{pol}$. The term μ^0 is actually dipole moment of isolated molecule. What authors suggest is that all models should include correction in the future because it makes them closer to polarizable potentials. New enhanced parameters were obtained by fitting simulation results to experimental ones (density and pressure of liquid at room temperature, vaporization enthalpy). Simulated system contained only 216 molecules and main run in simulation took only 20ps. From here we can see how much computer power has changed in the last thirty years. One combination of parameters that gave best results became basis for SPC/E. This model improved few results such as diffusion constant and radial distribution function. In the table 1 are parameters for SPC/E water model extracted from Gromacs. This model does not use experimental values for bond length or angle in a molecule.

Table 1: *Gromacs parameters for SPC/E model of water*

SPCE water								
atoms								
	nr	name	type	mass	charge	ptype	sigma	epsilon
	1	Opls_116	OW	15,9940	-0,82	A	3,16557E-01	6,50E-01
	2, 3	Opls_117	HW	1,008	0,41	A	0	0
bonds								
	i	j	funct		length		force.c	
	1	2	1		0.1		345000	
	1	3	1		0.1		345000	
angles								
i	j	k	funct		length		force.c	
2	1	3	1		109.47		383	

2.2.2 TIP4P/2005 model

TIP4P/2005 model (Transferable Intermolecular Potential with 4 Points) was introduced in 2005. by Abascal and Vega [19] as an enhancement of TIP4P model. For water models from TIP family, number after the 'TIP' part depicts how many interaction points does the model have. From this fact we can conclude that TIP4P/2005 model has 4 interaction points and it falls in category c from picture 4. As an inspiration for this model served the fact that most of water models can't reproduce temperature at which water shows density maximum ($T_{md} = 4°C$). For this model T_{md} and several properties of some ice forms are used to parametrize new potential. Also for fine tuning of the model, intermediate model developed just for ice forms was used.

Authors say that their model is rigid, empirical model based on Bernal-Fowler geometry and functionality. This would be a good time to explain what are the main things about Bernal-Fowler model [20]. It was developed in 1933. from spectral and x-ray data and it is often mentioned as the first water model. Bernal and Fowler deduced that water forms some sort of tetragonal networks and that its liquid structure is quartz-like, whilst at higher temperatures structure is approaching some kind of ideal simple liquid. Their ideas became widely accepted and as one review article says: their model ' is in essence both conceptually and semi-quantitatively what is used today in some major simulations of water' [21].

Of four ineraction sites, three are subject to Coulomb interaction: two hydrogen sites and one dummy site which carries positive charge.Meanwhile, oxygen participates only in Lennard-Jones interaction. For bond length and angle experimental values are used $l = 0.9572\mathring{A}, \theta = 104.52°$.

Basic parameters of TIP4P/2005 model used in Gromacs simulations are shown in table 2.

Table 2: *Gromacs parameters for TIP4P/2005 model of water*

TIP4P2005 water								
atoms								
	nr	name	type	mass	charge	ptype	sigma	epsilon
	1		OW	15,9940	0	A	3,15890E-01	7,74526E-01
	2, 3		HW	1,008	0.5564	A	0	0
	4		MW	0	-1,1128	D	0	0
bonds								
	i	j	funct		length		force.c	
	1	2	1		0,09572		502416	
	1	3	1		0,09572		502416	
angles								
i	j	k	funct		length		force.c	
2	1	3	1		104,52		628,02	

2.2.3 Comparison of water models

Considering that up until this day there is no absolute model which would be suitable for every kind of simulation, it's no surprise that many comparisons have been made. Authors of TIP4P/2005 model made big review containing simulation results for 5 simple, rigid models including both models mentioned above [22]. They made an overview of 17 different water properties and they established a grading process for models. Grades go from 0 to 10, where 10 means biggest overlap between model and experimental results (to be more precise, 10 means that simulation results are within 0.5% tolerance limit from experimental results). Tolerance depends on the chosen water property, and it mostly has values of 0.5% or 2.5%. Picture 5 shows table extracted from [22] and it gives overall grades for all models and all properties.

First property mentioned is enthalpy of phase change, where results are taken for vaporization enthalpy and melting enthalpy of ice. If we focus only on SPC/E and TIP4P/2005, we can see that they are not reliable for these values. One must always take in account that models are fitted to some parameters so it's normal that they have grade 10 for those parameters. It's probably good to take a look at gas properties. We can see that none of the models make good prediction of gas properties of water and SPC/E and TIP4P/2005 are catastrophic with grade equal to zero. Possible explanation that is offered is exagarated dipole moment which is much bigger than that of the gas phase.

Difference between SPC/E and TIP4P/2005 can also be seen from T_{md}. TIP4P/2005 is modeled in a way to fit to T_{md} so it's normal that it has excellent accordance with experiment, but maybe that's the exact thing that makes this model superior to SPC/E in most of the comparisons. If we take a look at properties for which SPC/E is better, we can see that those are enthalpy of vaporization (with very small difference) and vapour pressure (for which both models don't work very well).

From all this we can conclude that in the reign of simple, rigid models TIP4P/2005 is the best choice since it gives better predictions of almost all properties taken into account. The problem here can be the fact that there is no comparison of models based on solvation energies and the only thing we can deduce is that TIP4P/2005 probably gives better results for that too. There is another great article about water models written back in 2002. that includes 46 different water models but of course without TIP4P/2005 model. It gives analysis of same properties but from a different perspective [23].

Property	TIP3P	SPC/E	TIP4P	TIP4P/2005	TIP5P
Enthalpy of phase change					
ΔH_{melt}	0	0	5	6	6
ΔH_{vap}	8	5	10	4	10
Critical point properties					
T_c	6	9	6	10	2
ρ_c	4	4	9	9	8
p_c	1	3	4	3	0
Surface tension					
σ_{300K}	0	5	3	9	0
σ_{450K}	0	4	0	9	0
Melting properties					
T_m	0	1	4	7	10
ρ_{liq}	6	8	9	9	8
ρ_{solid}	3	3	5	9	0
dp/dT	0	8	7	10	0
Orthobaric densities and TMD					
TMD	0	5	7	10	10
ρ_{298K}	7	9	8	9	6
ρ_{400K}	0	5	1	8	0
ρ_{450K}	0	3	0	7	0
Isothermal compressibility					
κ_T [1 bar; 298 K]	5	10	4	10	8
κ_T [1 bar; 360 K]	0	5	1	8	0
Gas properties					
p_v[350 K]	3	0	3	0	0
p_v[450 K]	5	2	1	0	0
B_2[450 K]	0	0	0	0	3
Heat capacity at constant pressure					
C_p[liq 298 K; 1 bar]	9	7	8	7	0
C_p[ice 250 K; 1 bar]	0	0	0	0	0
Static dielectric constant					
ε[liq; 298 K]	6	7	3	5	7
ε[I_h; 240 K]	0	0	0	0	0
Ratio	0	0	4	3	0
T_m–TMD–T_c ratios					
$T_m[I_h]/T_c$	2	6	9	9	5
TMD/T_c	5	8	10	10	5
TMD–T_m	4	6	7	6	10
Densities of ice polymorphs					
ρ[I_h 250 K; 1 bar]	0	5	6	10	0
ρ[II 123 K; 1 bar]	5	1	5	8	0
ρ[V 223 K; 5.3 kbar]	0	8	8	8	3
ρ[VI 225 K; 11kbar]	9	6	5	9	6
EOS high pressure					
ρ[373 K; 10 kbar]	8	8	8	10	6
ρ[373 K; 20 kbar]	7	8	7	10	5
Self-diffusion coefficient					
ln D_{278K}	0	7	2	9	7
ln D_{298K}	0	8	4	8	6
ln D_{318K}	0	10	4	7	4
E_a	1	7	7	8	1
Shear viscosity					
η[1 bar; 298 K]	0	6	1	9	6
η[1 bar; 373 K]	2	9	4	10	2
Orientational relaxation time					
τ_2^{HH}[1 bar; 298 K]	0	6	2	9	4
Structure					
$\chi^2(F(Q))$	4	6	7	8	8
$\chi^2(overall)$	4	6	5	7	7

Figure 5: *Comparison of water models with grade from 0 to 10 (10 shows simulation in good accordance with experiment). Table is reused from [22]*

2.3 OPLS forcefield

Since we took a glimpse at the water models, now would be a good time to go over OPLS forcefield used for describing solutes in this work. OPLS stands for 'Optimized Potentials for Liquid Simulations' and it was created by William L. Jorgensen in 1988. [24] [25] Originally OPLS was developed as a united atom forcefield (OPLS-UA) where simulation sites were placed on non-hydrogen atoms and those hydrogens attached to heteroatoms or carbons in aromatic rings. This type of forcefield yielded computational time roughly proportional to the number of sites squared. OPLS-UA is used in this work to describe methanol as a solvent in some simulations.

Further development of the forcefield (done at the Yale university) resulted in OPLS-AA version where all atoms are given role of simulation sites. Difference between these two versions is maybe best seen from the example of methanol molecule. Methanol (CH_3OH) has 6 atoms and with OPLS-AA it would take 6 simulation sites to picture it correctly while in the case of OPLS-UA one would need only 3 sites to simulate it, one site for CH_3, one for O and one for H. This number is obviously two times smaller for united atom version of the forcefield and that leads to 4 times shorter computational time. Analogous with the water models, forcefield will yield more accurate results if it simulates interaction using more parameters. It is stated in [26] that the comparison of simulation results for OPLS-AA and OPLS-UA was made and AA version showed better accordance with experiments. Authors bring up the example where results for hydration free energy of alkanes showed three times smaller error bar when OPLS-AA is used.

All atom OPLS is parametrized to meliorate simulations of many organic liquids such as alkanes or alcohols. Interaction between two molecules a and b is computed according to 2.5.

$$E_{ab} = \sum_{i}^{ona} \sum_{j}^{onb} [\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij}(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}})] f_{ij}] \tag{2.5}$$

In the above expression we can see Coulomb term and Lennard Jones term. Factor at the end, f_{ij}, is equal to one for all interactions except intramolecular 1.4 interactions for which is equal to 0.This value is used because in the case of small molecules these interactions can lead to wrong results in calculation of Gibbs free energy. Charges for all sites were obtained by fitting the results and charges for functional groups also have the advantage of transferability between the molecules. Other important expression are definitely those for bond stretching and angle bending inside the molecule. Those are described by:

$$E_{bond} = \sum_{allbonds} K_r (r - r_{eq})^2 \tag{2.6}$$

$$E_{angle} = \sum_{allangles} K_\theta (\theta - \theta_{eq})^2 \tag{2.7}$$

Expressions 2.6 and 2.7 describe stretching and bending of a molecule as a harmonic oscillator. Parameter for these energies for mostly inherited from AMBER forcefield [27] but alkanes turned out to be exception. Their bond angles needed to be enlarged for simulation to achieve desired results. Last important equation is the one for torsional energy given by 2.8.

$$E_{torsion} = \sum_i \frac{V_{1i}}{2}[1+cos(\phi_i + f_{1i})] + \frac{V_{2i}}{2}[1+cos(2\phi_i + f_{2i})] + \frac{V_{3i}}{2}[1+cos(3\phi_i + f_{3i})] \quad (2.8)$$

In the above expression ϕ_i is dihedral angle, V_i are coefficients from Fourier series and f_i are phase angles. Summation goes over all dihedral angles in a molecule. All these parameters used in OPLS-AA forcefield to describe desired molecule are shown in a figure 6. Full parameters are shown in the appendix A.

OPLS Methane

atoms

nr	name	type	mass	charge	ptype	sigma	epsilon
1	Opls_138	C	12,0110	-0.24	A	3.5E-01	2,76E-01
2	Opls_140	H1	1,008	0.06	A	2.5E-01	1,26E-01
3	Opls_140	H2	1,008	0.06	A	2.5E-01	1,26E-01
4	Opls_140	H3	1,008	0.06	A	2.5E-01	1,26E-01
5	Opls_140	H4	1,008	0.06	A	2.5E-01	1,26E-01

OPLS Ethane

atoms

nr	name	type	mass	charge	ptype	sigma	epsilon
1	Opls_135	C1	12,0110	-0.18	A	3.5E-01	2,76E-01
2	Opls_140	H1	1,008	0.06	A	2.5E-01	1,26E-01
3	Opls_140	H2	1,008	0.06	A	2.5E-01	1,26E-01
4	Opls_140	H3	1,008	0.06	A	2.5E-01	1,26E-01
5	Opls_135	C2	12,0110	-0.18	A	3.5E-01	2,76E-01
6	Opls_140	H4	1,008	0.06	A	2.5E-01	1,26E-01
7	Opls_140	H5	1,008	0.06	A	2.5E-01	1,26E-01
8	Opls_140	H6	1,008	0.06	A	2.5E-01	1,26E-01

OPLS Propane

atoms

nr	name	type	mass	charge	ptype	sigma	epsilon
1	Opls_135	C1	12,0110	-0.18	A	3.5E-01	2,76E-01
2	Opls_140	H1	1,008	0.06	A	2.5E-01	1,26E-01
3	Opls_140	H2	1,008	0.06	A	2.5E-01	1,26E-01
4	Opls_140	H3	1,008	0.06	A	2.5E-01	1,26E-01
5	Opls_136	C2	12,0110	-0.12	A	3.5E-01	2,76E-01
6	Opls_140	H4	1,008	0.06	A	2.5E-01	1,26E-01
7	Opls_140	H5	1,008	0.06	A	2.5E-01	1,26E-01
8	Opls_135	C3	12,0110	-0.18	A	3.5E-01	2,76E-01
9	Opls_140	H6	1,008	0.06	A	2.5E-01	1,26E-01
10	Opls_140	H7	1,008	0.06	A	2.5E-01	1,26E-01
11	Opls_140	H8	1,008	0.06	A	2.5E-01	1,26E-01

Figure 6: *OPLS-AA parameters for alkanes used in this work*

2.4 Short tutorial for free energy calculation

Free energy calculation using Gromacs is very similar to other kinds of simulations but there are few special things that one needs to know to perform it. Because of this reason we will go through all steps needed to get the simulation up and running and for example we will take solvation of methane in methanol. Basic ingredients of free energy calculation are solvent and solute. For solute we need its .pdb and .itp file, while for solvent we need to have box with molecules equilibrated at desired temperature, .top and .itp file. For simulations with SPC/E water as a solvent there is already equilibrated box in Gromacs library so we can just use command to solvate our desired molecule in it and later equilibrate it for some short time. Before proceeding any further, now would be a good time to go over file formats used by Gromacs. Linux operating system supports any file extension you can think of and it's easy to understand why Gromacs uses big number of them. Majority of the files used by Gromacs are in fact textual files but extensions are used for easier determination which files are used as an input with different commands. We will go over file formats used in this tutorial.

- structure files - commonly used are .pdb and .gro formats. .pdb stands for brookhaven Protein DataBank format and it's some kind of standardized file. On the other hand .gro is Gromacs format for structure of the molecules. .pdb files can easily be converted to .gro and other way around with pdb2gmx command implemented in Gromacs

- topology files - files used for specifying topology and molecules or atoms used in work. Here we say what atoms does the molecule consist of, how the bonds look like, basically what are the parameters of the forcefield. .top is the main topology file, .itp are include topology files or those that contain some specific information and we include them in .top. Also, there are .rtp files for residue topology

- parameter files - those are .mdp files used for specifying user input for the simulation itself. In this file we specify things like number of steps used, integrator, what will be our thermostat or barostat in a simulation. Some values need to be specified in this file but some will be set to their defaults if user doesn't enter some preferred value.

- output files - these files can be divided in those representing trajectory, energy and other files. Trajectories are written in different types of file depending on the desired accuracy and precision. They differ in the way that some contain only coordinates for system (.pdb or .xtc), some coordinates and velocities (.gro) and some even the forces (.trr). Energies are usually written in .edr files containing pressure, temperature, volume, density and virials. Other files can be very different depending on the type of simulation, one of them is .xvg format suitable for use in Xmgrace tool for drawing graphs and figures from data.

When we have initial files we need to insert one methane molecule in solvent, in this case methanol. There is Gromacs command for this which takes few arguments: number of solute

molecules (*-nmol*), how many times Gromacs will try to insert the solute (*-try*), solvent .gro file (*-f*), solute .pdb file (*-ci*) and output .gro file (*-o*). If we have dense solvent and big solute molecule there is a chance that Gromacs won't find adequate position for insertion in 10 tries and in that case we need to make number of tries larger.

gmx insert-molecules -nmol 1 -try 10 -ci methane.pdb -f methanol300.gro -o all.gro

As an output we have all.gro file that we can convert to .pdb file with *gmx editconf* command if we want to observe our system with Pymol. We need to include .itp file for solute in topol.top file either by *#include methane.itp* or by simply copying content of methane.itp to topol.top file. Also, we need to add 1 molecule of solute in the end of .top file. Now that we have whole system together, we need to do two energy minimizations with and without flexible bonds. Files needed for this are .mdp files in which we need to define integrator, number of steps, desired maximum force, type of interactions and constraints. Usually, for this simulation, it is enough to have 50000 steps for flexible minimization and 10000 for non-flexible. Input files for all types of run need to be preprocessed with *gmx grompp* command and then we actually run the simulation.

gmx grompp -f em_flexible.mdp -c all.gro -o em_flexible.tpr

gmx mdrun -v -deffnm em_flexible

Same commands go for non-flexible minimization with different input files. After minimization we need to equilibrate the system. For calculation of free energy it's best to have perfectly equilibrated system prior to dividing it according to lambda values. Equilibration time depends on system we have but if we had an equilibrated box of solvent at the beginning it is enough to let it run for 1 ns. There is a great variety of parameters that can be specified in .mdp file for equilibration and they need to be adjusted for current system. If we want NPT ensemble then we need to set referent temperature and pressure together with thermostat and barostat Gromacs will use in simulation. Periodic boundary conditions are also defined in .mdp. Time step that I always use is 0.001ps and with that time step I need one million steps in simulation to achieve total time of 1 ns. When we have defined .mdp file we preprocess it and run it with same commands as for energy minimization.

gmx grompp -equil.mdp -c em.gro -o equil.tpr

gmx mdrun -v -deffnm equil

To see if our system is equilibrated we can use generated .edr file that we give as an input to *gmx energy* command. With *gmx energy* we can see all averages for our system. If we specify output file we can get .xvg files suitable for use with Xmgrace to get charts. If equilibration succeeded, we can go to the interesting part which is special for free energy calculations. We need to adjust our main .mdp file which will be responsible for main simulation run. In this file we need to emphasize that free energy calculation will be performed by setting *free-energy*

parameter to value *yes*. After that we need to set which molecule we want to solvate with *couple-moltype* = <*name of solute from .top file*>. Other parameters are 'soft-core' parameters, interactions at $\lambda = 0$ and $\lambda = 1$ and of course λ values. If we use 20 different λ values we need to have 20 different folders with all files required for run. This is annoying to do 'by hands', so it's better to have bash script to do it for us. On the internet you can find similar tutorials with scripts. The one I always use is tutorial from Gromacs site where you have zipped scripts and file templates [28]. In the mklambdas.sh script you can set all λ values you want to use. This script takes as arguments three files: run.mdp, topol.top and equil.gro and it makes different folders containing those files. It also sets λ values in .mdp files. We run the script with:

bash mklambdas.sh run.mdp topol.top equil.gro

After this we copy .itp files to all folders if they are included in topol.top file. At this stage we are ready to run the simulation. We go to folders one by one and we run these commands:

gmx grompp

gmx mdrun -v

After this we just wait for every one of them to finish. At the end we need to calculate free energy difference for our solute. We do this by running next command from the parent folder (parent of lambda_* folders).

gmx bar -f lambda_*/dhdl.xvg

With this we will get our results written on terminal with total ΔG and errors for each λ step of simulation. *gmx bar* command has different options we can use to get histograms (-oh) or other information important for this calculation. Example of a figure that can be generated from output files from *gmx bar* is given on a figure 7.

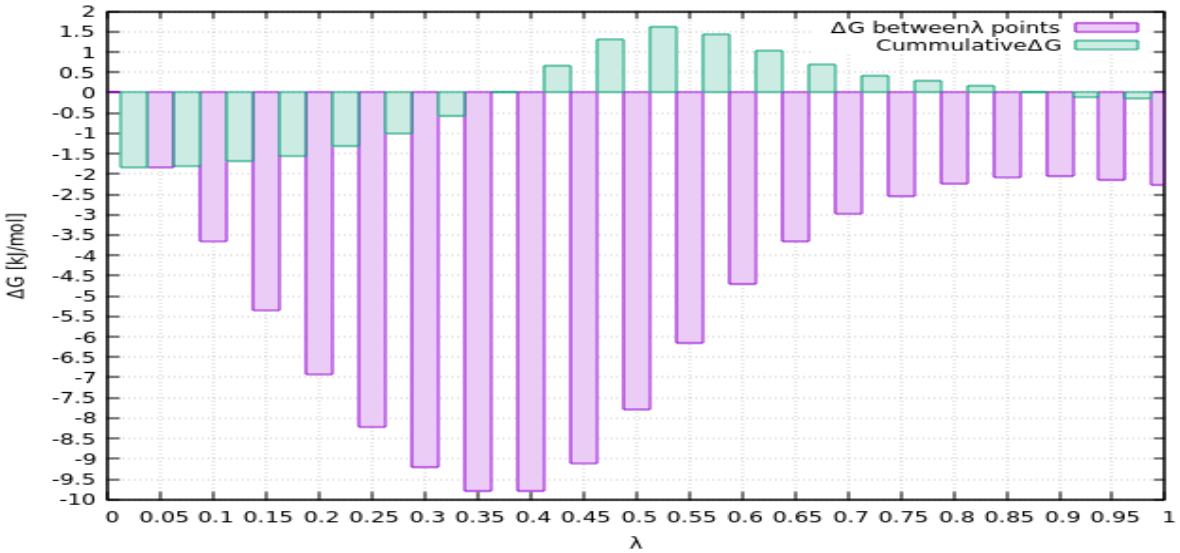


Figure 7: *Gromacs output for difference in ΔG and cumulative ΔG for different λ points for simulation of ethane in methanol at 300K*

3 Results

3.1 Setup of method

As has been mentioned in the previous chapter, success of Gromacs simulation depends on many different parameters such as equilibration and simulation time, chosen thermostats, barostats and so on. In the case of free energy calculations right selection is additionally weighted with extra parameters for λ points. Way of working used in this work was established through many tries and mistakes and in the next few paragraphs it will be shown in brief how some choices reflect on the final result.

First and most obvious parameter we need to check when talking about free energy calculation is the number of λ points used. It follows from basic logic that the greater the number of points greater the accuracy of simulation but question is how many is enough? On figure 8 we show results of methane solvation in SPCE water at temperature $T = 300K$. Free energy difference is plotted against number of λ points. Smallest number is only two points which means that we simulate clear solvent and solvent with 'full grown' solute molecule. It can be seen that for small number of points final result can vary and error is tremendously big , especially for the case of $N_\lambda = 3$ where it spans over area bigger than the value itself. With larger N_λ result becomes stable and error makes less than 5% of the free energy value.

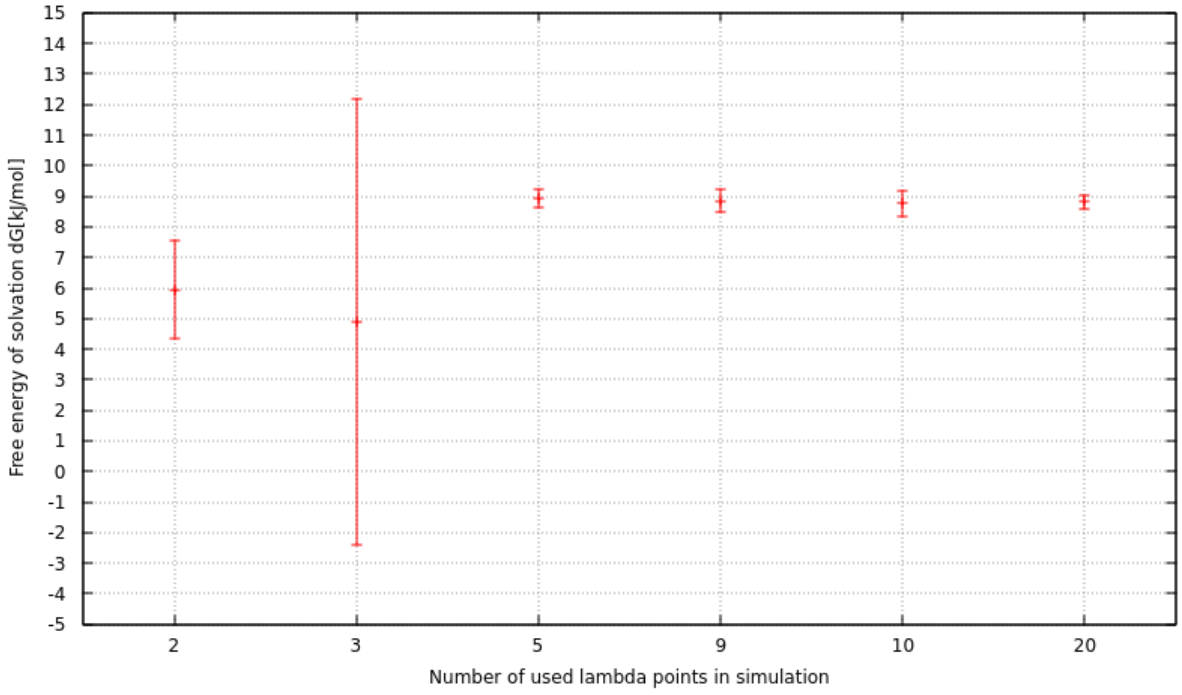


Figure 8: *Free energy difference for methane in SPCE water at 300K plotted against number of λ points used in simulation*

Equally important as the selection of N_λ is selection of equilibration and simulation time. Example for this is shown in figure 9. Simulation was made for three different combinations of equilibration and simulation time. All results were made for temperature equal to 173K and pressure equal to 1bar.

First result is extracted from output files after relatively short simulation, equilibration lasted only 200 ps and simulation lasted 600 ps. Here it's important to say that equilibration is made after insertion of solute in methanol box which is previously well equilibrated for a long time. This result has relatively big error especially when comparing to other temperatures for the same system. This will be seen in one of the next chapters where overall results for methane in methanol will be shown. Temperature 173K is very low for methanol and that's why big error bar was expected with such short times. Second combination was 300 ps for equilibration and 800 ps for main run but error bar is still very high. For third combination we wanted to be sure so equlibration time is enlarged to 500 ps and simulation time to 1500 ps. These results were finally satisfying as can be seen from figure below. Error bar is two times smaller from the one for first combination. One must be aware that different times for simulation can be needed depending on point in phase space we want to simulate. System will be easier equlibrated when it's simulated under 'normal' conditions for the liquid state. In this case, solvation of methane in methanol, even first combination was enough to get good results for temperature equal to 300K.

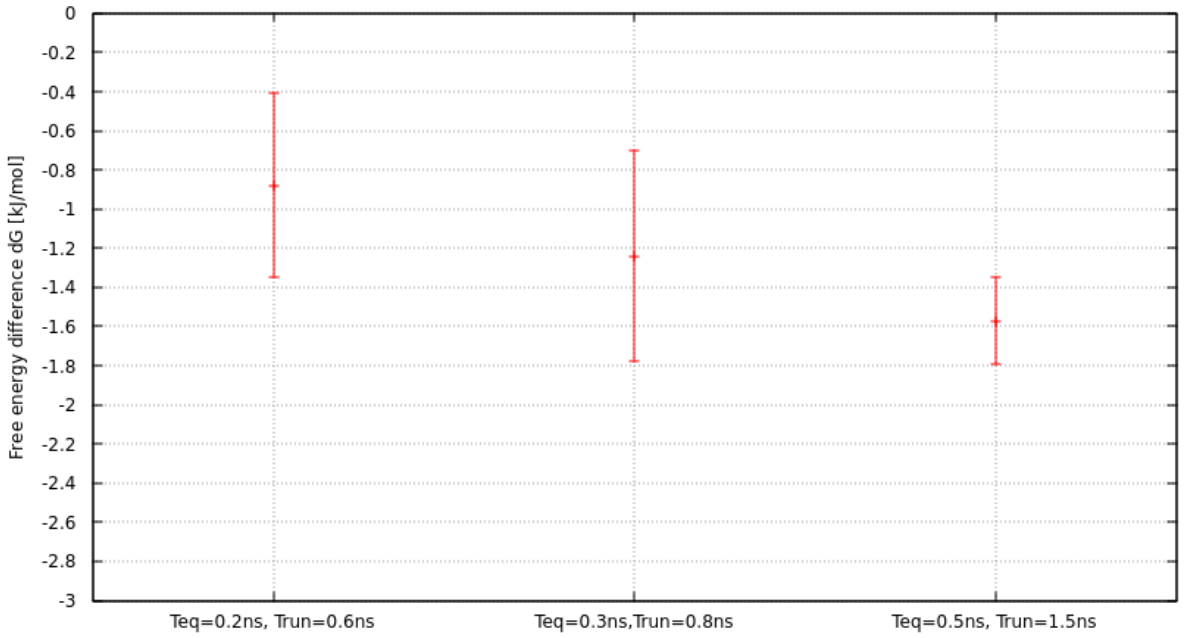


Figure 9: *Free energy difference for methane in methanol at 173K plotted against equilibration and simulation time*

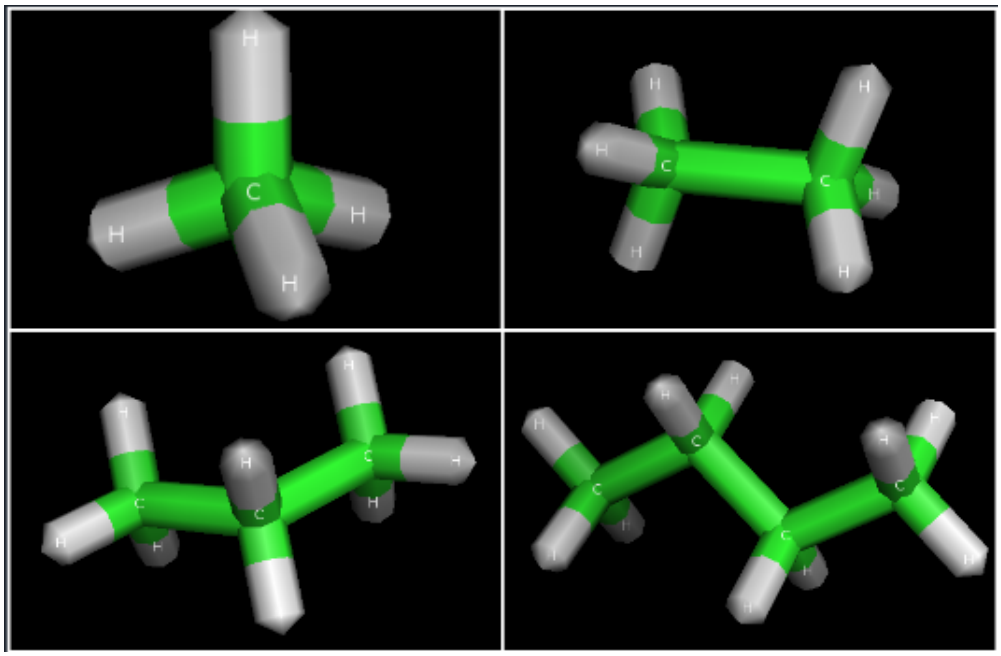


Figure 10: *Pymol snapshots of alkane models used in this work, from methane to butane*

Taking facts stated above in consideration we made final decision on number of λ points used in simulations as well as duration of simulations. Main goal is of course to have results that clearly represent the system, so all parameters are chosen with some 'safe space' left. All solvations are simulated with difference between λ points equal to 0.05 which gives total of 21 λ points. This should remove errors due to insufficient overlap between adjacent values of λ.

Equilibration of the whole system, containing solvent and solute, is for all results at least equal to 1ns. With this we can be sure that mix of solute and solvent reached desired conditions. Run time for simulation was 1ns or 1.5ns depending on a solvent used. Important option here is also number of steps after which Gromacs will compute energies and coordinates of atoms and it's mostly set to every 200 or 500 steps depending on the system under consideration.

In the next few sections we will go trough results made for this thesis, first solvation of alkanes in two different water models will be presented and afterwards solvation of alkanes in methanol. We chose methane, ethane, propane and butane for alkanes and their OPLS-AA models used in this work are shown in the figure 10. In the end of the results section short overview of structural properties of simulations will be given. Common parameters for all simulations are shown in the appendix D while non-common ones will be stated together with results.

3.2 Solvation of alkanes in water

First results we will focus on are those that depict solvation of alkanes in water. Two different water models are used: SPC/E and TIP4P/2005 and we will start with SPC/E. First simulations were made for temperature equal to 300K and pressure equal to 1bar. Equilibrated box of SPC/E water already exists in Gromacs libraries so it wasn't necessary to make separated box for these conditions. Standard tutorials for Gromacs state that since we insert solute in already well equilibrated solvent additional equilibration does not need to be long. It's stated that 500ps should be enough for system to adjust to imposed conditions. Just to be sure, after the insertion of alkane, system was equilibrated for 1ns with time step of 0.001ps. State of the system was checked by looking at running averages generated with g_energy command. Main run for every λ value had same duration of 1ns. Sampling for calculation of free energy difference was done every 500 steps which means that in the end there are 20000 sampling points. System consists of approximately 880 water molecules and 1 alkane places in a cubic box with volume $V \approx 27nm^3$. Simulation results are compared with experimental values obtained from Minnesota solvation database [29]. Results are shown in picture 11.

From figure 11 it can be seen that SPC/E water doesn't reproduce qualitative behaviour of experimental results. Experimental results show minimal value of free energy for ethane while simulation shows linear growth of free energy together with size of alkane chain. On the other hand, quantitative comparison of these results is not as bad as one would think. Error bar for simulation doesn't cover true value for none of the alkanes but value for methane is relatively close to experimental one. Butane shows big error bar and comment on that will be given after taking a look at the next water model.

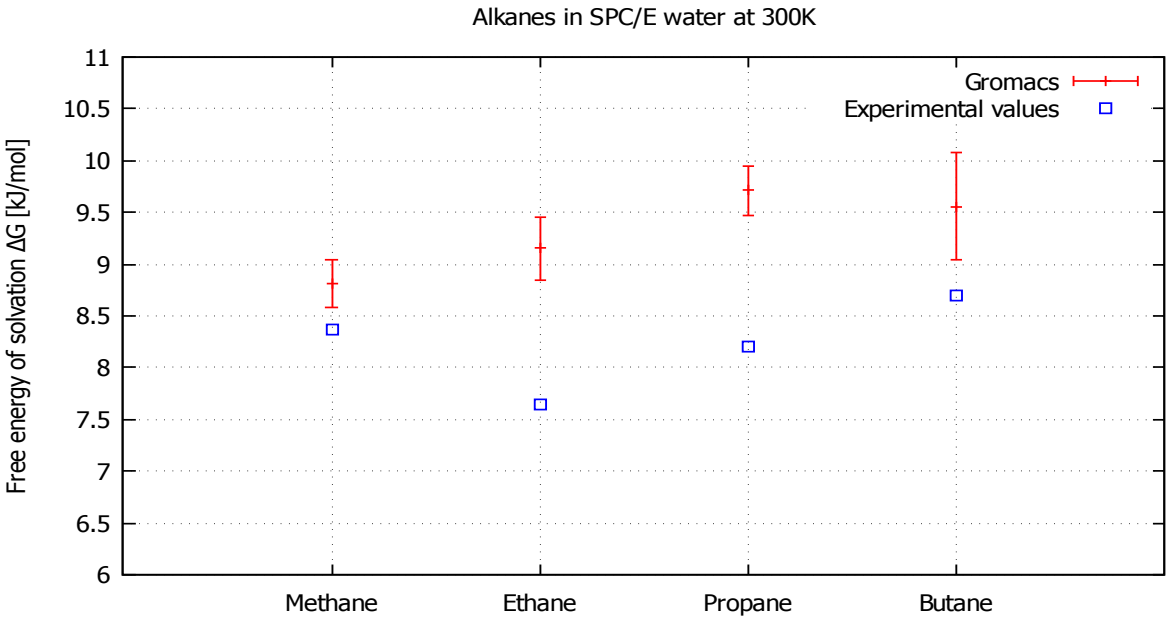


Figure 11: *Solvation free energy of alkanes in SPC/E water at T=300K*

Next step was to change water model from SPC/E to TIP4P/2005 and run the simulation again with the same parameters and same conditions. These results are shown in figure 12 and they are colored in red. Again we compare them with experimental database for same conditions. TIP4P/2005 model proved better for first three alkanes since it reproduces same qualitative behaviour as experiment. Minimum of free energy of solvation can be seen for ethane. Error bars are smaller in this case, especially for methane. One thing that stands out is butane free energy which is now shifted below experimental value. Also, same as with SPC/E butane shows biggest error bar among all alkanes. This was really troubling at first but it crossed our mind that butane has two conformations - anti and gauche conformation. Molecule simulated in this work, as shown in figure 10, has only anti conformation. Our conclusion is that, in order to get correct results for butane, one needs to simulate both of the conformations and then calculate free energy based on the portion each conformation takes in real life. Hence, we decided to exclude butane from further discussion.

From the overview of water models given in previous chapter it was clear that TIP4P/2005 model gives better results for most of the properties simulated by now. Free energy of solvation was not included in comparison in articles taken as reference. Our predictions that TIP model will win over SPC/E proved to be correct for qualitative and quantitative behaviour. For this reason only TIP4P/2005 model was used in simulation of the same systems at higher temperature $T = 350K$. Results are shown on the same figure 12 in green. Free energy takes higher values for higher temperature, and from picture it can be seen that there is no longer any minima. Error bars are visibly smaller for this temperature. ΔG grows linearly with size of alkane chain. There are no experimental results in our possession to compare free energies for alkanes in water under these conditions.

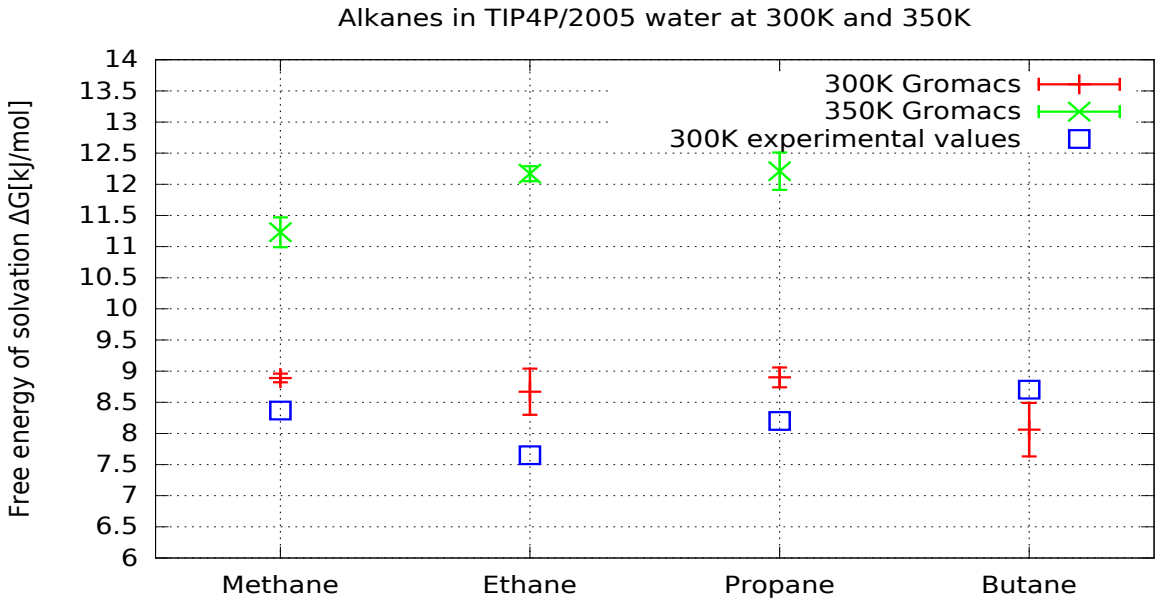


Figure 12: *Solvation free energy of alkanes in TIP4P/2005 water at T=300K and T=350K*

3.3 Solvation of alkanes in methanol

Second major part of this work were simulations of alkanes in methanol. Methanol was simulated using OPLS-UA forcefield with CH_3 part simulated as one site. Methanol is chosen because of its amphiphilic nature. CH_3 part is similar to alkanes, it's nonpolar and it should be non-soluble in water. On the other hand, hydroxyl group is polar and compensates for short carbon chain in methanol. Amphiphilic nature makes methanol interesting solvent because alkanes should show greater solubility in it than in water.

Alkanes are solvated in a wide range of temperatures from 173K to 430K. This range is supposed to give a better look at how free energy of solvation depends on temperature imposed on a system. All temperature are accompanied with pressure equal to 1bar, which means that, once again, we have NPT ensemble. Methanol as a solvent is not included in original Gromacs libraries that come with installation and because of that files for it were made manually. Cubic box was filled with thousand molecules and equilibrated for few nanoseconds. After this initial equilibration box had the volume of approximately $60nm^3$. In the light of unexpected butane behaviour that was discussed in previous section, in this part only first three alkane were used. Equilibration of the whole system containing solvent and solute lasted 500ps followed by main simulation run 1.5ns long. Results for all alkanes and all temperatures are shown in the figure 13 for one to get better feeling about ratio of solvation free energy for alkanes of different size.

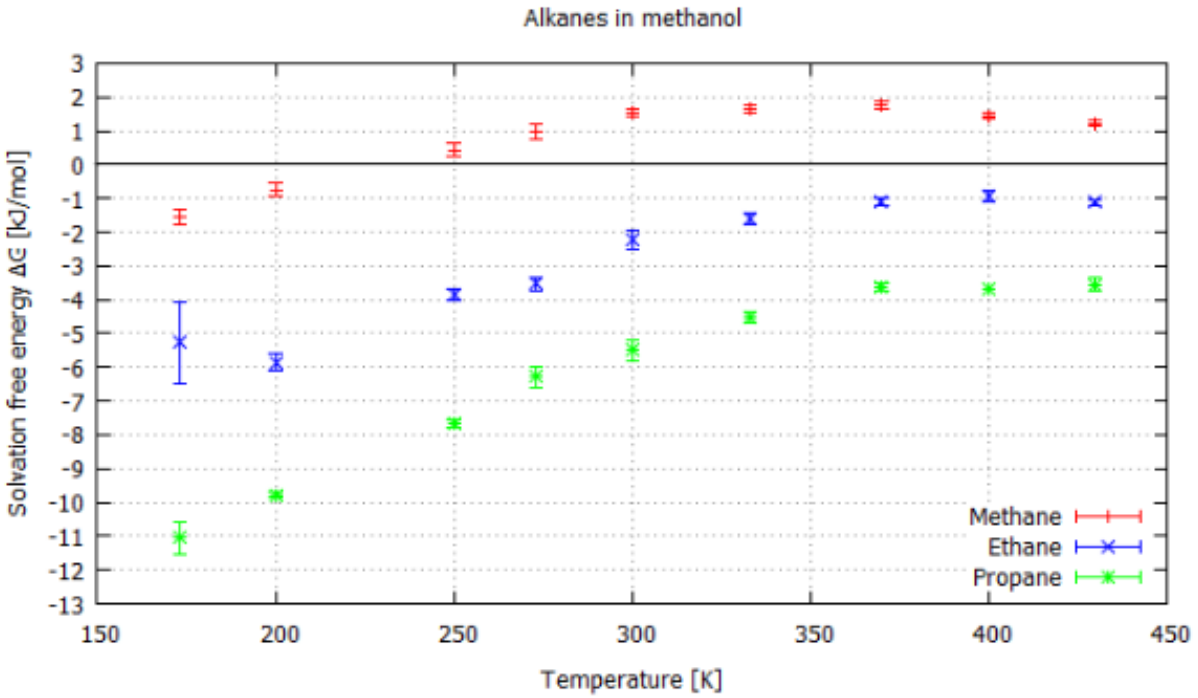


Figure 13: *Solvation free energy of alkanes in methanol from 173K to 430K after 1500ps run*

Most compelling thing about results presented in the previous figure is definitely the fact that ΔG becomes more negative as the alkane chain grows bigger. This is against our preliminary expectations that ΔG for this system mostly depends on the size of cavity needed to accommodate alkane molecule in the solvent. Moreover, if we focus on methane we can see that its free energy crosses the border $\Delta G = 0$ with temperature greater than 250K. This implies that methane 'prefers' to solvate in methanol only under some conditions which would be in this case only temperatures smaller than 250K accompanied by pressure of 1bar. At the same time, ethane and propane are soluble for all conditions the system was put in. Generally speaking, free energy in this case becomes more negative as the alkanes grow bigger. Difference between them is not so small if we consider the fact that for alkane hydration at 300K all values were placed between $\Delta G = 8.5$ kJ/mol and $\Delta G = 9$ kJ/mol. Same conditions here yield free energies in range from $\Delta G = -5.49$kJ/mol for propane to $\Delta G = 1.52$ kJ/mol for methane. Next property we can observe is the peak that shows for all alkanes at different temperatures. There is obviously some set of conditions which results in the smallest solubility for each of them. Further increase in temperature after that point results in better solubility. This peak is most obvious for methane while for propane it can be said that there is some kind of saturation after temperature of 370K.

Error bars for these results are relatively small and for all alkanes they are smaller for higher temperatures. Biggest errors are shown for temperature 173K and for ethane it is huge. This is probably due to some error in computation because when we look at gmx bar output only few λ points have low accuracy. Also, this can probably be fixed by equilibrating simulation box for a while longer since 173K is border line for methanol and it can be quite hard for system to reach perfect equilibrium under these conditions.

Since there is almost no literature dealing with methanol, it was hard to find any results to compare our simulations with. Despite this, there was need for finding any comparison at all so we decided to show our results via Henry's constant instead of 'usual' Gibbs free energy. Henry's constant can be expressed in many different ways depending on their connection to other quantities, but we will present our results with next equation:

$$H = \rho_v RT e^{\frac{\Delta G}{RT}} \tag{3.1}$$

In the above expression ρ_v is representing molar density of the solvent. [30]. Plot H-T is shown in figure 14.

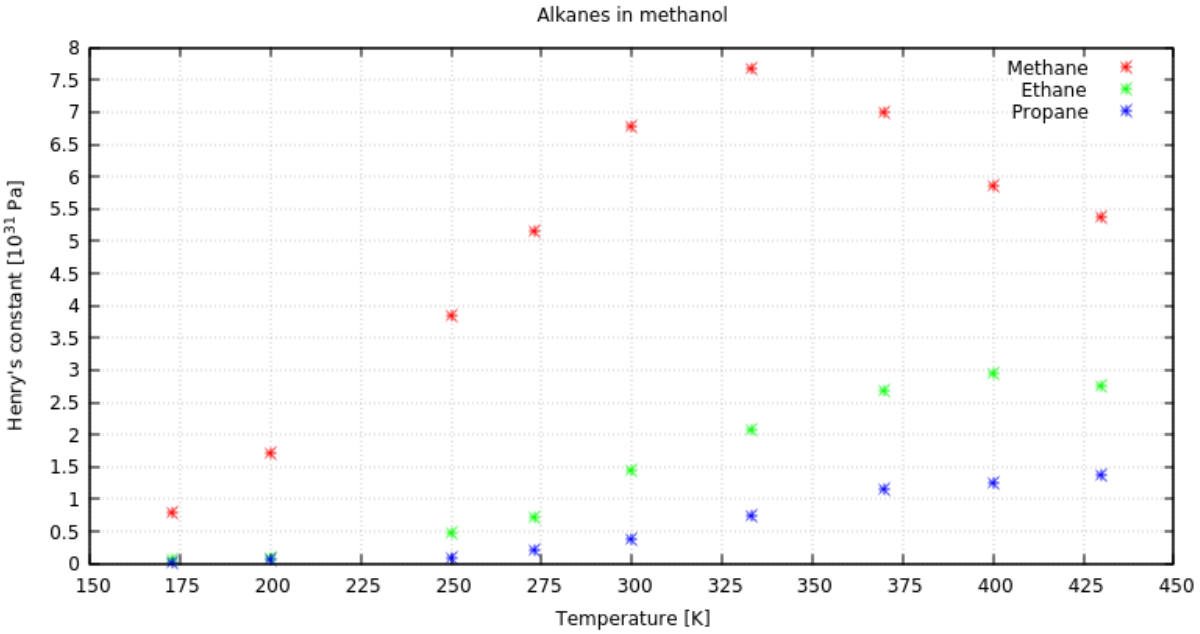


Figure 14: *Henry's constant for solvation of alkanes in methanol with temperatures from 173K to 430K after 1500ps run*

3.4 Structural properties

It's probably obvious by now that all results are showed in similar manner. Reason for this is the fact that g_bar method in Gromacs as an output gives free energy difference between adjacent λ points and associated error and total free energy difference between clear solvent and solvent and solute. It is possible to generate graphs and histograms directly from g_bar but they don't show anything interesting for this work.

Consequently, we decided to try some other ways to get a better perspective on what happens inside our system and how do solvent molecules and solute interact during the simulation. Alkane methanol radial distribution function was plotted from results of alkane solvation in methanol. We chose and compared two different temperatures (T=300K and T=370K) and all three alkanes. One must be aware that rdf is most accurate when averaged over large number of molecules and obtained from very large number of sampling points. In this case we only have one molecule of solute so for example methane-methanol rdf is not averaged at all. As a result rdfs showed in figures 15 and 16 are not smooth functions. Results are also shown for three different values of λ, the one representing clear solvent, one for solvent and fully simulated solute and one intermediate point for $\lambda = 0.5$ where solute molecule has its interactions diminished by half of their true value.

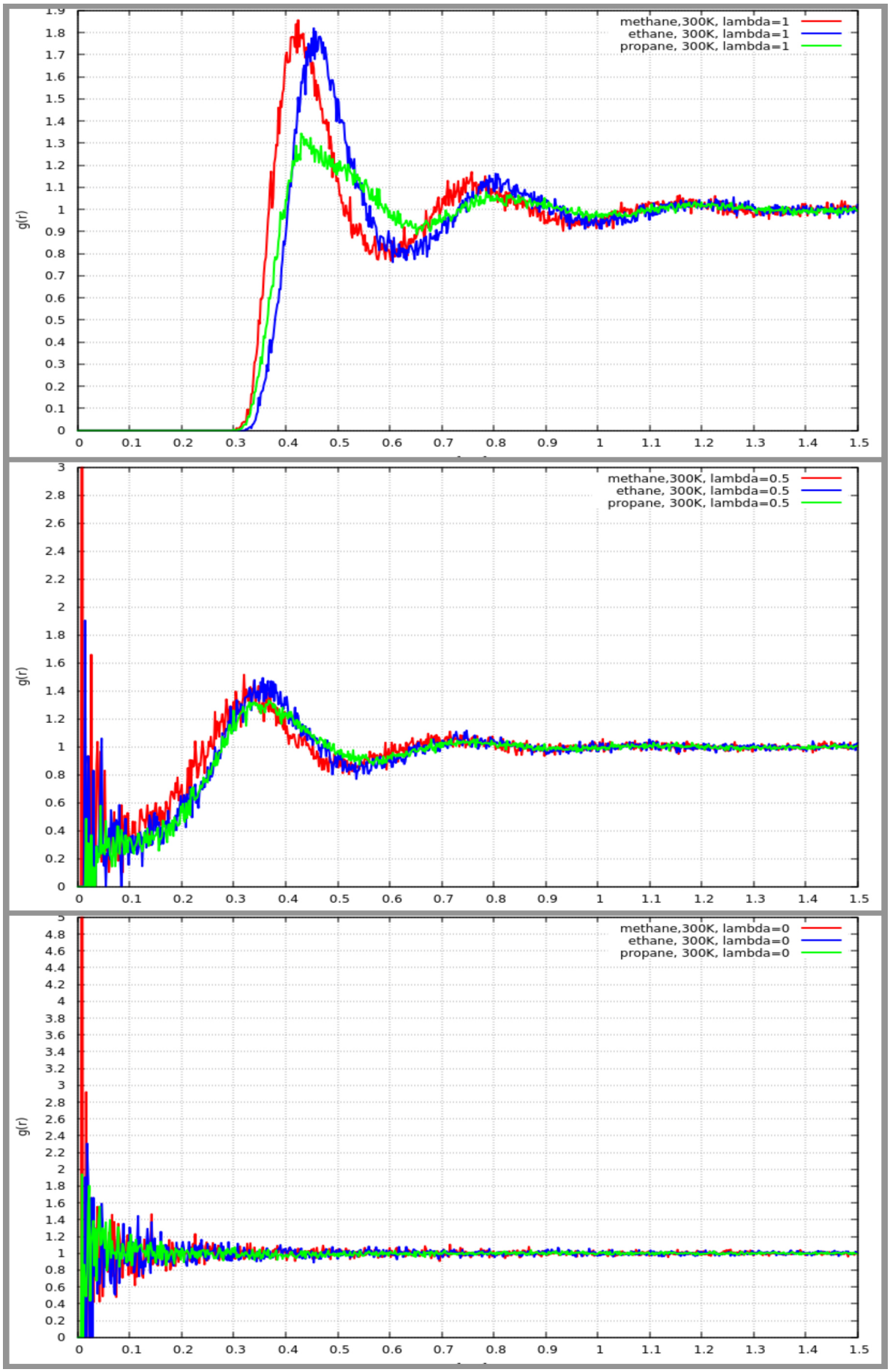


Figure 15: *Rdf for alkanes in methanol at temperature equal to 300K*

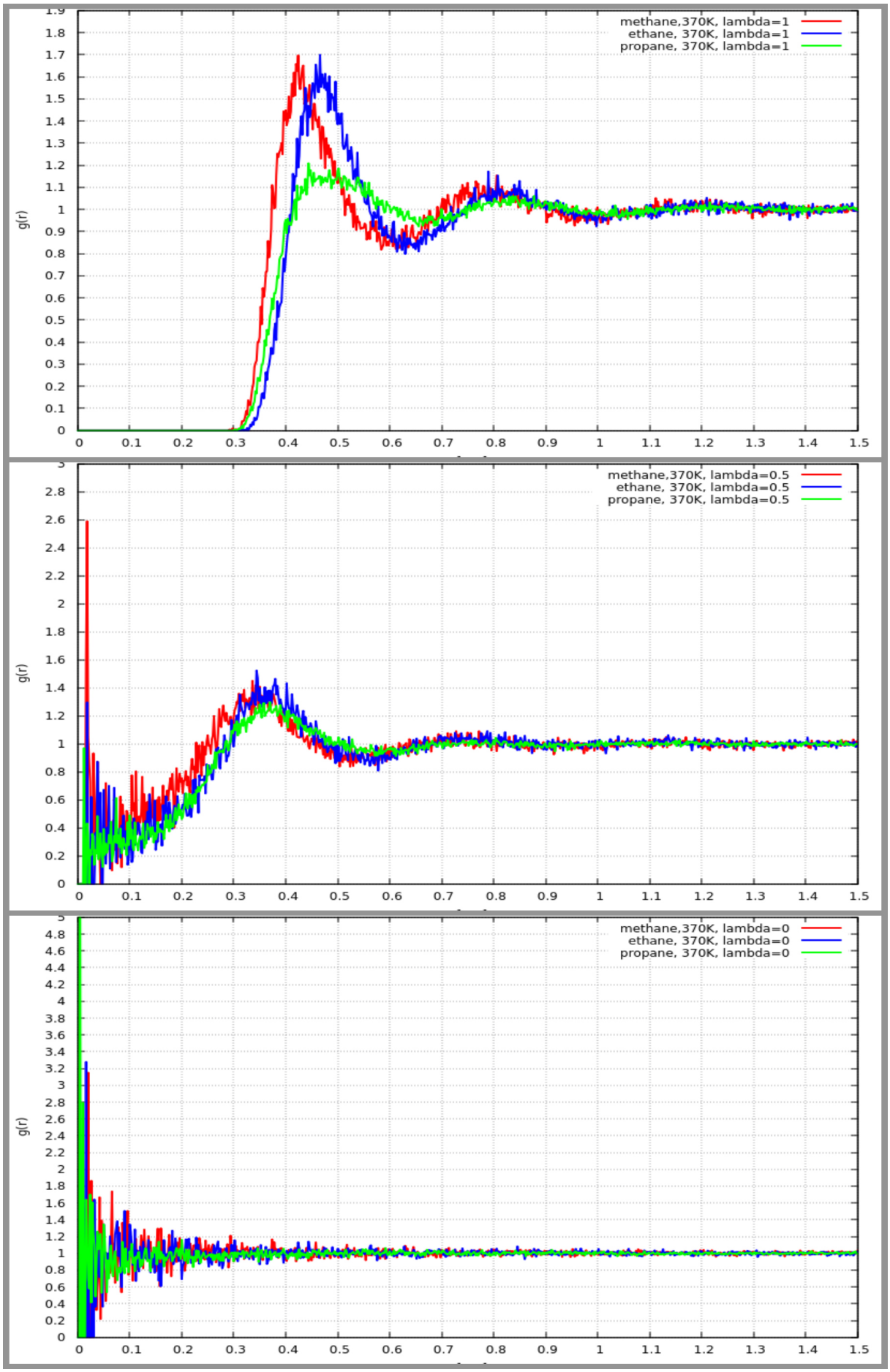


Figure 16: *Rdf for alkanes in methanol at temperature equal to 370K*

Let's first take a look at the rfd for $\lambda = 0$. It looks similar for both temperatures so the description stands for both. We can see that rdf doesn't even resemble the usual shape we are used to and that's the result of having all solute interactions equal to zero. In this case methanol molecules don't 'see' alkane at all and they are moving freely over it like it doesn't exist at all. This stands for all alkanes. When we get to half of the simulation that is to the point when $\lambda = 0.5$ solute has interactions with strength equal to half of their usual value so it's expected that rdf will show some kind of adjustment happening in the system. Rdf still doesn't have its usual shape since for all alkanes there are methanol molecules at immediate distance around them. Although, there are two small peaks forming for all of them at T=300K and one peak at 370K. From the middle part of both pictures we can see that in this case peaks are at approximately same distance for all alkanes.

Now we need to focus on the uppermost part of both pictures where rdfs for $\lambda = 1$ are shown. For temperature T=300K we can see that methane and ethane associated g(r) are quite similar. First, and most obvious peak, has value of approximately $g(r) = 1.8$ and it shows at distance $r = 0.42nm$ for methane and $r = 0.45$ for ethane. It is expected that methane will have its first neighbours closer than ethane since methane is smaller molecule. Second peaks are showing at $r = 0.75nm$ for methane and $r = 0.8nm$ for ethane. All in all, their rdfs are behaving very similar and it looks like ethane rdf has just been shifted little bit to the right of methane. Propane behaviour is on the other hand interesting. It has values much lower than methane and propane which is okay since it is bigger molecule. But interesting part is that its first peak shows at intermediate r value between the one for methane and one for propane.This would mean that there is smaller number of first neighbours for propane than for ethane but they are at closer distance.

Example for higher temperature, T=370K, has same qualitative properties as the one for T=300K but with all values shifter downwards a little bit. This means that at a higher temperature number of first neighbours is slightly smaller than at T=300K. All parts of both picture show that g(r) goes to one as r goes to higher values. This is normal behaviour for liquid. These results are not to be taken too seriously, or to be more precise, we can't take quantitative behaviour as really accurate because this method for determining rdf should be used on large number of molecules so that g(r) can be averaged. This can be done for free energy calculation by running all simulations many times until we have something to calculate average upon but it requires great hardware resources and long period of time since one simulation for specific alkane at a specific temperature can take up to three days even on cluster with 10 physical cores.

4 Analysis and discussion

To conclude this work, we will go over main results shown in this work in the previous chapter. We will try to find a bigger picture in which they can all fit. First we will focus on the water. Among two water models used we will focus on TIP4P/2005 model because we have already shown that it is superb when compared to SPC/E. After this we will go over to methanol as a solvent and comparison of the results with the work done by other research groups will be made to see how well do the results obtained with different methods fit together. Also it's important to state all limitations we became aware of while using Gromacs and molecular dynamics. For the end we will take a quick look at the future, what it brings us and what important questions do we still have to analyse and which became obvious during this work.

4.1 Alkanes in water

From what we can see on a picture 12 first thing that comes to our mind is that all values ΔG for T=300K are approximately the same. They are all somewhere from 8.5kJ/mol to 9kJ/mol with, we can say, small error. This is something that is in accordance with Chandler's article [7] but, as he stated in it, this is just a pure coincidence. In any other solvent values for different alkanes won't be so nearly placed. One thing that left us in wonder is solvation free energy for ethane that is smaller that the one for methane in the simulation and in the experiment. This is something we didn't expect since Chandler's article and other sources predict linear growth of ΔG with the size of alkane chain. This definitely calls for further discussion and more detailed look into it but we will try to see what can be said at this moment. ΔG for alkanes can be separated in two contributions - one from the cavity creation and the one from interactions (as can be seen in 1.35). Cavity creation for ethane must be connected with greater energy needed because, after all, ethane is bigger molecule than methane and it is less probable to find such cavity without solvent molecules inside the system. This can be put as $\Delta G_{v,methane} < \Delta G_{v,ethane}$. If we take 1.35 as an appropriate description for solvation free energy of alkanes, only way we can end with ΔG smaller for ethane is if interactions part of mentioned equation is much smaller for ethane since it also needs to compensate for difference in cavity energy. I, personally, don't see a way how this can be explained but it is of great importance to figure it out. This discussion would be a lot easier if we had more experimental results or at least simulations dealing with this problem. Chandler didn't go to great accuracy with numerical values in his article and all he said about them is that they are all around 2kcal/mol (that would be 8.368 kJ/mol) which is not sufficient since we got results that are in accordance with this but without 'perfect' linear growth.

Next thing we can go over is the growth of ΔG with temperature, also seen from the figure 12. For 350K values 'jumped' to over 11kJ/mol for methane and even to 12kJ/mol for ethane

and propane. It can still be said that different alkanes have similar values. For this temperature there is a linear growth of ΔG with the size of alkane chain and this makes the previous discussion even more complicated. Now we should focus on next thing, why is there a growth in ΔG when we increase the temperature imposed on a system? Alkanes are mostly inert molecules, especially when placed in a polar solvent such as water. It was discussed in the introduction that free energy can be divided in two major contributions - enthalpy and entropy (1.33). Since enthalpy is connected to the interactions between solvent and the solute it will not have an important role here. We don't have new interactions that are strong enough to introduce some important change in enthalpy. With this in mind, we can say that our change in G is mostly connected with entropy or reordering of hydrogen bonds. Water as a polar solvent has whole network of hydrogen bonds and it likes to keep it intact. By placing alkane in it we are perturbing this network so that water molecules need to reorganize themselves around the solvent in a particular way that makes possible for average number of bonds for one molecule to remain constant. Since hydrogen bonds are directional this leads to a smaller configurational space for water - molecules have small set of orientations they can take if they want to bond with their neighbour. Entropy grows as a system has greater number of microstates at its disposition (1.2), and in this case we can see that change in entropy will be negative. Let's simplify this discussion by denoting system on a temperature 300K as system 1, and the one on 350K as system 2. We already mentioned that change in enthalpy has minor influence on free energy in the case of alkanes and because of this, together with the fact that temperatures are in 'normal' range, we can approximate enthalpy change to be equal in these two cases: $\Delta H_1 = \Delta H_2$. Of course, for temperatures it stands that $T_2 > T_1$. Entropy change is negative in both cases, and ΔS_2 will have slightly smaller absolute value than ΔS_1 since molecules in a system with higher temperature move around little bit more quickly so they can get to some new parts of configurational space. With all this we have $\Delta S_i < 0$ and $\Delta S_1 < \Delta S_2$ but still it stands that $T_1 \Delta S_1 > T_2 \Delta S_2$ because of the difference in temperatures. All this leads to $\Delta G_1 < \Delta G_2$.

As for comparison of the results with some others, we can say that most important is maybe the one where we compared them with the experiments. There are many articles written about hydration of different molecules and some of them are also dealing with small alkanes. Problem here is the fact that there are many different methods used in free energy simulations and it's hard to find articles doing similar work. Still it's good to compare with other methods just to see what is the current state of the art in the field, and which simulations provide best results. One such article is dealing with solvation of alkanes and alcohols but at much lower level since they are using Hartree-Fock theory in combination with different program packages [31]. They made few comparisons of solvation free energy with different basis sets used and good thing about their work is that they separated ΔG on different energy contributions which is something impossible for us to do with Gromacs. Values they obtained differ with various basis sets but here their values that are closest to experimental values will be shown. For methane that is 9.163kJ/mol and for ethane 7.447kJ/mol. On the other side our values are 8.89kJ/mol

and 8.67kJ/mol, while experimental ones are 8.368 and 7.656kJ/mol for methane and ethane respectively. It can be seen that our result for methane is closer to 'real' value while they lead in the quest for ethane free energy. This is important because it can be seen that even with simulations done on a higher level one can achieve results comparable to those on low-level. Of course, it needs to be said that this research is from 2001. and in the meanwhile computational power improved together with Hartree-Fock theory so it's questionable could our results even be compared to some recent results obtained with their method.

4.2 Alkanes in methanol

Results for alkanes in methanol were shown in a figure 13. We can see that opposite to water as a solvent, alkanes are actually soluble in methanol (at least ethane and propane which show negative ΔG for all imposed conditions). Why is this the case? One possible reason is amphiphilic nature of the methanol itself. Methanol consists of hydroxyl group and short alkane chain and it has less developed hydrogen bond network since only hydroxyl group can participate in it. Smaller overall number of hydrogen bonds means that insertion of an alkane will result in a smaller perturbation compared to insertion in the water. Alkanes that we used are small, with short alkane chain, and they can easily find adequate place after the insertion. All this can result in a negative free energy of solvation.

Now we get to the question why larger alkanes show better solubility than the smaller ones? Methane, as the smallest solute used, should be the most soluble if we stick with the theory used in the previous section for explaining hydration because he requires creation of smallest cavity. Possible explanation for this would be formation of some sort of structures inside methanol itself. Because of its structural properties, hydroxyl group of one methanol will prefer to be close to hydroxyl group of the second methanol. This results in a formation of different structures which can be divided in two groups - those chain-like and those ring-like. [32] Structures that can exist in methanol are shown in a figure 17. From looking at that picture it's easy to see that there is a great probability that alkane will fit in one of these structures and that it would be his preferred state. Depending on a structure which formed inside the methanol it can easily happen that longer alkane chains fit better than methane resulting in a greater solubility.

Next thing we need to address is the qualitative behaviour of ΔG with temperature. We can see that ΔG grows until the peak placed on some temperature and after this it starts to decrease. Growth until this peak can be explained in the same way we explained behaviour for water in the previous section. For methanol we have bigger range of temperature so we are still left with the need to explain the peak and the fall after it.

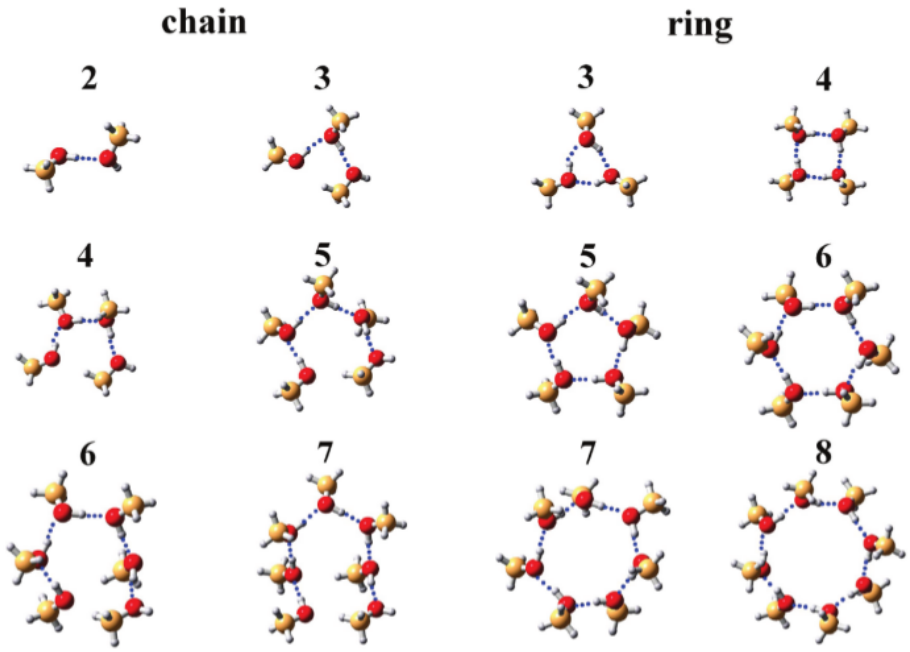


Figure 17: *Different structures inside liquid methanol. Picture taken from [32]*

With bigger temperature, molecules inside the solvent have bigger energies and when they reach a certain point they will have energy big enough to stop caring about preservation of hydrogen bonds. When this happens entropy change will suddenly become positive resulting in a negative change in free energy of solvation (according to equation 1.33). After this point molecules have great energy which puts at their disposition much bigger part of the configurational space resulting in a bigger entropy.

As for the comparison with the other results for this system, we will compare this work with the work done by another research group. They simulated solvation of alkanes in methanol via Widom insertion method which was briefly mentioned in the introduction. Difference in presentation of the results is that they are using Henry's constant to visualize solvation and solubility in various systems. Their work is shown on a figure 18 together with experimental results they obtained and we will compare it to ours shown on a figure 14. First thing we can see is that scales differ very much but this can be explained with the fact that Monte Carlo simulation uses reduced quantities. Also there are many different versions of Henry's constant and they mostly differ in the proportionality factor before the exponent. Important thing is that we have qualitative accordance between the results. We can say that our results maybe follow the experimental ones even better than the ones obtained with Widom method. This can be well seen in the case of methane where the growth of H with temperature is not so parabolic. Both figures show big gap between methane and ethane and also slower and less dominant growth of two bigger alkanes whose peaks are slightly shifted to higher temperatures. This comparison has great value because this way we see how well do these two methods describe the system.

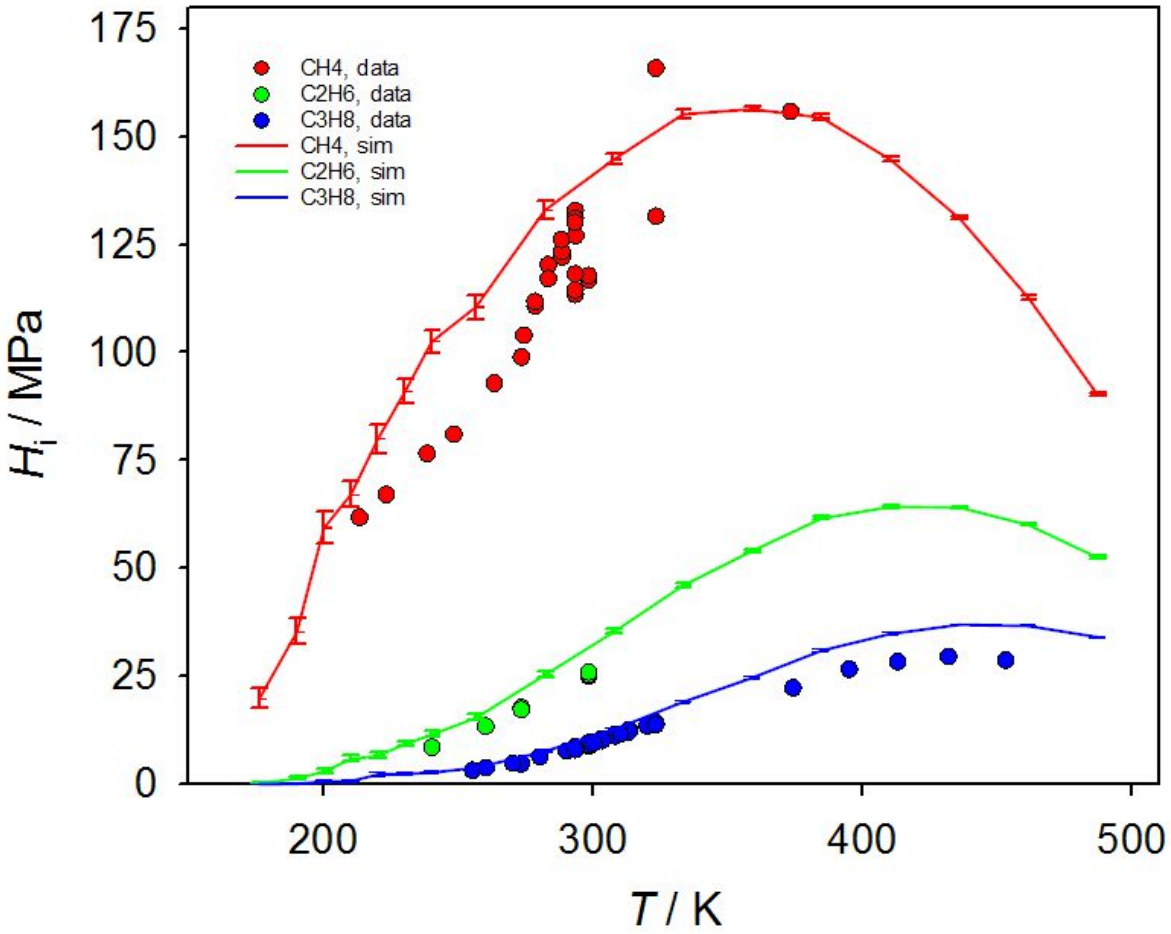


Figure 18: *Results for Henry's constant for alkanes in methanol together with experimental results. Picture taken with permission from authors*

4.3 Hydrophobic effect

We saved one final thing to discuss for the end of this work. Hydrophobic effect is something we all heard of, even as early as in elementary school, as an explanation of why water and oil don't mix perfectly. For a term so commonly used, it's interesting that there is still no unique explanation for its true nature. It is even hard to find good definition for hydrophobic effect although it can be given in various ways. Literature that proved best for this theme is definitely Israelachvili's book about the intramolecular and surface forces. [33] It has vast discussion about this effect and in the next few paragraphs we will go over most important parts of it.

Israelachvili gives next definition for the hydrophobic effect: "The immiscibility of inert substances with water and the mainly entropic nature of this incompatibility, is known as hydrophobic effect and such substances (e.g. hydrocarbons and fluorocarbons) are known as hydrophobic substances" [33]. I guess that now is pretty obvious how this relates to out work, especially the part about solvation of alkanes in water. We already stated in previous analysis that ΔG for alkanes must be mostly caused by the change in entropy. Author is introducing this

effect by firstly discussing the water and its interactions with other water molecules. Water is a molecule that can be simply described as egocentric - it likes itself very much. As discussed, water molecules bond over hydrogen bonds which are directional and very strong in their nature. Because of this fact and the 'egocentricity', water is not so happy when inert molecules are inserted. Inert solutes will not share any H-bonds with it and with this they will disturb already existing network. As we saw in our results for alkanes in water, the bigger the solute the bigger number of broken hydrogen bonds. In this book, author talks about this and about the structure water can form around the solute in order to maintain as big number of H-bonds as possible. These structures are shown on a figure 19. One thing we didn't include in the previous discussion is the fact that water can bend H-bonds from their preferred direction in order to connect with another molecule. The cost of this bending is not negligible but it's still okay if it serves the preservation of the bond. Useful diagram is shown on a figure 20 where we see the dependency of the strength of a bond on the angle two molecules close.

Author then argues the entropic contribution to the unfavourable solution of inert molecules. This gets really interesting because in the simple systems entropy is the one that will favor the mixing and solution unless there are strong repulsive forces. He states that entropy change is negative because of the fact that water needs to obey more ordered structure with the solute inside. This is the same thing we already said in the introduction and discussion of the results. But interesting part is his example for solvation of n-butane where we can see how much entropy really contributes to the final ΔG. As a remainder, we rejected n-butane as a solute because of

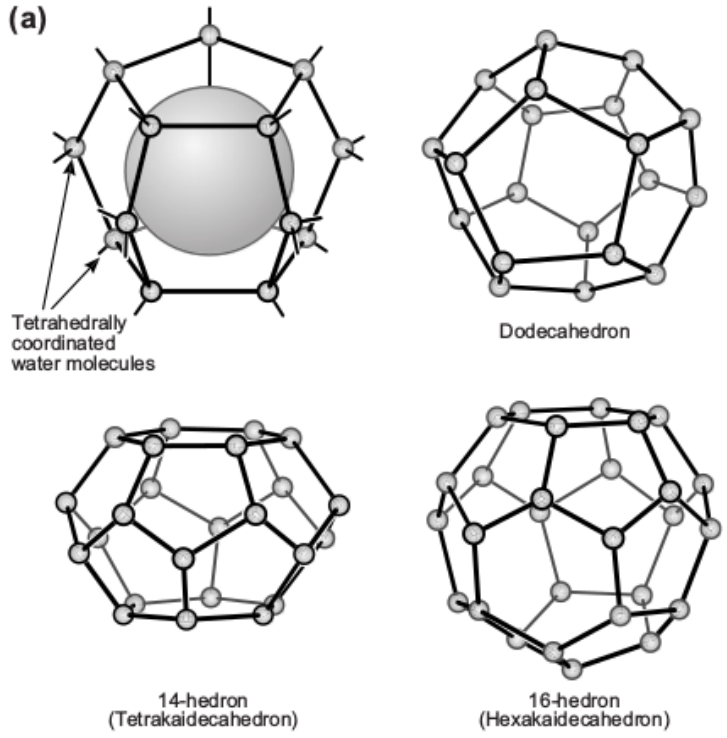


Figure 19: *Clathrate cages formed by water. Figure taken from [33]*

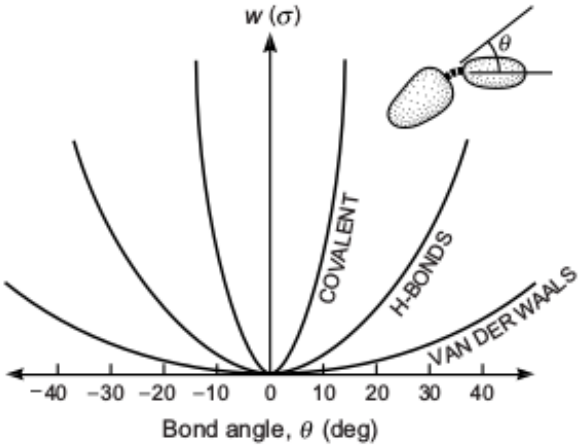


Figure 20: *Orientation-dependence of different types of bonds. Figure taken from [33]*

its two conformations we were unable to simulate. Free energy of solvation for this molecule is made of 85% of entropic contribution ($\Delta G = 24.5kJ/mol$ while $-T\Delta S = 28.7kJ/mol$). Also, author states that the price of solvation will depend on the surface around the molecule which is immiscible for H-bonding, that is the surface that encloses the solute cavity. There is one more thing we haven't thought of until now. What would free energy look like for a water molecule inserted as a solute in the nonpolar solvent? We can take author's word on this that free energy difference for water in, for example lets say, methane would be much bigger than in the opposite case.

There are many articles written on this theme and since hydrocarbons and noble gasses are most commonly used inert molecules, there are many articles dealing with simulations of alkanes in water. We will go over most interesting ones we found and we will present their conclusions. One of these is already discussed in great lengths in the free energy chapter and that is Chandler's article. [7] Next one we will mention is from the authors who parametrized the OPLS forcefield. Jorgensen et al. published way back in 1985. an article dealing with hydrophobic effect or to be more precise, hydration number (number of water molecules in a first hydration shell around the solute).[34] They simulated alkanes in water with Monte Carlo simulations and they have found correlations between thermodynamic quantities and coordination number. This is quite interesting to see since it shows that these quantities mostly depend linearly on C (where C is coordination number). They fitted the results to obtain next three equations:

$$\begin{aligned} \Delta G[kJ/mol] &= 0.17C + 22.04 \\ \Delta H[kJ/mol] &= -0.97C + 4.54 \\ \Delta S[J/moldeg] &= -3.83C - 59.03 \end{aligned} \tag{4.1}$$

We can see from the above equations that entropy decreases much faster with C than enthalpy.

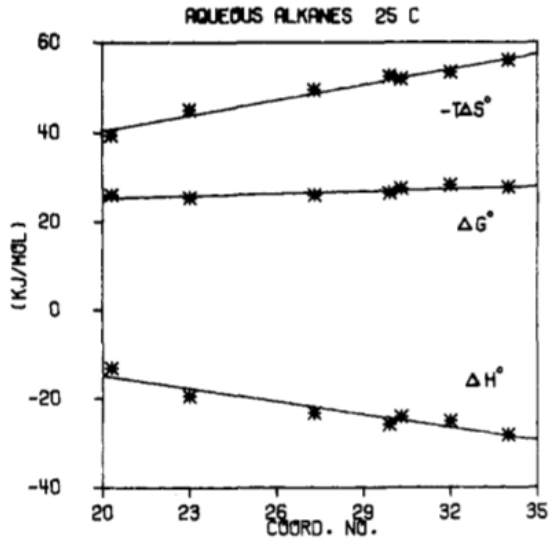


Figure 21: *Correlations between experimental data and coordination number for alkanes in water. Figure taken from [34]*

It shows that ordering of more molecules around the solute causes decrease in entropy which makes solvation unfavourable. Also, it is interesting to notice the entropy-enthalpy compensation which is not discussed in this article. Negative change in enthalpy and positive change in the product $T\Delta S$ produce only slight change in ΔG which is very close to constant when plotted in figure 21. One more thing from this article we might add is that authors noticed that C grows approximately by value of 3 for each new carbon atom added in alkane chain.

There is also a big number of review articles that give great view on all the research that is ongoing in this field. Baldwin published one relatively recent in 2013. [35] He discusses the problem in free energy calculations where scientists often simulated transfer of solute from liquid water to the liquid phase of the solute itself. In the case of alkanes, this would mean that we are not looking at the transfer of alkane from water to gas, but at the water-liquid alkane transfer. By making an unusual comparison of protein folding and solvation of alkane, author remainded that ΔG_{LL} (liquid-liquid transfer) is not negligible and it can be even twice as big as ΔG_{GL} (liquid-gas transfer). Example for ethane shows this with values $\Delta G_{GL} = 1.77 kcal/mol$ and $\Delta G_{LL} = 3.14 kcal/mol$. He also talks about cavity work related to hydrophobic effect and does this work really depends on surface area of cavity or its volume. This discussion is still ongoing in many articles and we can find many different theories. Mention of protein folding in this article will be much clearer later on, when we introduce something called hydrophobic interaction.

To continue, there is also a huge article by Southall et al. from 2001. giving historical background and most important findings on hydrophobic effect. [36] It also connects hydrophobical effect with its impact on large biomolecules and protein folding. One of the points is the formation of structures in water which they call icebergs (by now we already had cages, clathrates...

). Interesting fact they give is that these structures are important in some industries like for example extraction of natural gas. Pipelines that lead natural gas from the ground can often get clogged exactly because of these structures. Until the end of their work they go over MB 2D model of water and they discuss the hydrophobic interactions.

Decomposition of alkane free energy on enthalpy, entropy and cavity forming is obviously very popular idea for discussion. It is examined in an article by Gallichio, Kubo and Levy in 2001. [37] Computer simulations have been made using FEP method and they confirmed some of their ideas such as weak coupling of hydration entropy and solvent reorganization to solvent. They explained greater solubility of cyclic alkanes and the cavity formation accompanied with entropic effects. Their results are shown in a similar way as those from Jorgensen on figure 21. We can compare our results with theirs and for example we can see that for methane our prediction is closer to the experimental value (theirs is $2.404kcal/mol$). All of their energies for alkanes are slightly shifted to higher values. On the other hand, they have really nice view on decomposition of ΔG shown in a table with numerical values and errors.

Last work we will go over that's dealing with hydrophobic effect is the one from the 1988. where Abbraham and Mateolli discussed temperature variation of the hydrophobic effect. [38] They shown that for small temperatures structural effects have dominant influence on ΔG for alkanes. First hydration shell showed as a good candidate for the inspection. For 'normal' range of temperatures ΔH and ΔS turned out to be negative. They observed the compensation of very large enthalpies and entropies, resulting in small difference in free energy. Peak in alkane ΔG showed on a temperature dependent on the number of carbon atoms in the chain.

This short overview of the small part of literature dealing with hydrophobic effects can show us how it is all interconnected with the work we did in this thesis. Only thing is that hydrophobic effect goes to even further lengths because of its implications on bigger molecules such as proteins or even DNA. We never dealt with the fact what happens if we put two hydrophobic solutes in water, how would they interact mutually? Here we come to the term mentioned few times in the previous articles - hydrophobic interaction. It is the term describing unusually strong and attractive interaction between two hydrophobic solutes in water. Its oorigins are rooted deeply in the hydrophobic effect. As reported in [33], two methane molecules show few times stronger interaction when both in water than in air or vacuum. First case gives interaction of $-14 * 10^{-21}J$ and the second one 'only' $-2.5 * 10^{-21}J$. As Israelachvili said in his book, experiments that would deal with this are very hard to make since these compounds are so unlikely to solvate in water. This interaction is described as of very short range and it is said to be too complex to solve analytically. If hydrophobic molecules are attracted to each other when put in the polar medium, it can be seen that they will tend to form different kind of structures which will make possible for them to minimize the interaction surface with the solvent. These structures are often referred to as hydrophobic aggregates. Just like the hydrophobic effect itself, this interaction is very popular theme for researchers and there are many articles written.

We stumbled upon few of them and they deal with great variety of molecules influenced by these interactions. Some of those mentioned before included proteins in their discussion about hydrophobicity since their folding is of great importance in biology, chemistry and in the understanding of the human nature itself. Others go from small scale, where observation is done on only two methanes in polar ambient like in [39], to slightly bigger molecules like urea which is commonly used in scientific methods [40] and micelles [41]. Hydrophobic interaction is something very difficult to understand and we have only mentioned it to show how hydrophobic effect has large influence in different fields of science. It would require much longer time and deeper analysis of the literature to even get acquainted with it.

5 Conclusion

To conclude this work we will give short examination of everything we've done and learned so far. This work started with the research for one of the courses on college where we became aware of the fact that free energy calculations are very complex and that they require longer time to become acquainted with the methods they use. That's how we decided to put some effort to it and to make this the theme of this thesis. Since nobody did this kind of work before on our college, long time was needed to see pros and cons of the Gromacs package we decided to use for simulations. Solvation of small inert molecules became the theme of this work and we decided for two different solvents, water and methanol, because of the differences in their nature. Setup of methodology was done through multiple tries which ended with results followed by poor accuracy, but at the end we obtained some way of working which we followed for all results.

Water is a subject of worldwide research and consequently we had many articles to compare our results with. Still, we managed to have some interesting findings like the one where ethane shows smaller free energy difference than methane molecule even though it's bigger. Furthermore, we made a comparison of two different water models based on their capability to reproduce solvation free energy which is something we haven't found in literature for SPC/E and TIP4P/2005 models. Our results were satisfying, they were close enough to the experimental values. After we finished with water, we decided to move from polar solvent to amphiphilic one - methanol. This was captivating work since we haven't found any past work giving the free energy for alkanes in this solvent in a big range of temperatures. Because of this, we had long discussions to explain the behaviour obtained, which shows greater solubility of bigger alkanes in methanol. Great thing for us was that we became aware of other research groups doing similar calculations at the same time which made possible for us to compare the results with somebody. We ended up with the conclusion that these results are also satisfying.

After all this time, we can sum the advantages of doing this work together with the list of things we can improve in the future. From my point of view, biggest advantage of doing this kind of research is the knowledge it comes with it. This theme is something that is not done to great lengths on college and it was interesting to try to figure it all out. Search for an adequate literature was exhausting but in the end even those articles that weren't related to this work were helpful in gaining some knowledge about other research that is ongoing. Additionally, we saw what are the benefits of the method we used and how it compares to others. This brings us to the things that can be improved in the future. As is mentioned many times in this work, simulations of molecular dynamics require long time for computation. This reduces the work we are able to do in some period, lets say a year. Because of this it is of great importance to have adequate hardware that can withstand 100% utilization for long period of time. Likewise, there are many parts of free energy calculation we still need to learn. We became aware of

this when looking at the enthalpy-entropy decomposition of ΔG in the literature. Gromacs is terrific program package but it has its flaws, especially when one wants to change something in the calculation itself. Because of this if we want to take this work further in the future, we will need to find some other package or method to use jointly.

There are many directions we can take in the future. One of them would be to find a good method to make entropy-enthalpy decomposition for solvation of alkanes in methanol, just to see the system in a better light. Also we saw from the discussion on the hydrophobic effect that it is a vast subject and we can try to see what implications does it have on systems other than water. Even the work where we would solvate water in a non-polar environment could lead to some new findings. Possibilities are there and only thing left for us is to decide where we want to take it from here.

6 Bibliography

[1] Chandler, D. & Percus, J. K. Introduction to Modern Statistical Mechanics. Physics Today 41, 114 (1988).

[2] Wikipedia, Statistical ensemble (mathematical physics), access:August 2017., https://en.wikipedia.org/wiki/Statistical_ensemble_(mathematical_physics)\#/media/File:Statistical_Ensembles.png

[3] Šips,V., Uvod u statističku fiziku, Školska knjiga (1990).

[4] Greiner, W., Neise, L. & Stöcker, H. Thermodynamics and statistical mechanics. Nature. Springer. https://doi.org/10.1016/S0304-0208(08)70657-1

[5] IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997).

[6] Chipot, C. & Pohorille, A. in CompChemBook (2007). doi:10.1007/978-3-540-38448-9

[7] Chandler, D. Interfaces and the driving force of hydrophobic assembly. Nature 437, 640–647 (2005).

[8] Duarte Ramos Matos, G. et al. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. Journal of Chemical and Engineering Data 62, 1559–1569 (2017).

[9] Kubo, M. M., Gallicchio, E. & Levy, R. M. Thermodynamic decomposition of hydration free energies by computer simulation: Application to amines , oxides , and sulfides. J. Phys. Chem. B 101, 10527–10534 (1997).

[10] Gallicchio, E., Kubo, M. M. & Levy, R. M. Entropy - Enthalpy compensation in solvation and ligand binding revisited. Journal of the American Chemical Society 120, 4526–4527 (1998).

[11] Widom, B. Some Topics in the Theory of Fluids. The Journal of Chemical Physics 39, 2808–2812 (1963).

[12] Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. Journal of Computational Physics 22, 245–268 (1976).

[13] Wu, D. & Kofke, D. A. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. Journal of Chemical Physics 123, (2005).

[14] Wu, D., & Kofke, D. A. (2005). Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. Journal of Chemical Physics, 123(8). https://doi.org/10.1063/1.2011391

[15] Frenkel, D. & Smit, B. Understanding Molecular Simulation. Understanding Molecular Simulation 638 (2002). doi:10.1016/B978-012267351-1/50006-7

[16] Gromacs, About Gromacs, July 2017. http://www.gromacs.org/About_Gromacs

[17] Water models, access: June 2017. , http://www1.lsbu.ac.uk/water/water_models.html

[18] Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. The Journal of Physical Chemistry 91, 6269–6271 (1987).

[19] Abascal, J. L. & Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. The Journal of chemical physics 123, 234505 (2005).

[20] Bernal, J. D. & Fowler, R. H. A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. The Journal of Chemical Physics 1, 515–548 (1933).

[21] Finney, J. L. Bernal and the structure of water. Journal of Physics: Conference Series 57, 40–52 (2007).

[22] Vega, C. & Abascal, J. L. F. Simulating water with rigid non-polarizable models: a general perspective. Physical Chemistry Chemical Physics 13, 19663 (2011).

[23] Guillot, B. A reappraisal of what we have learnt during three decades of computer simulations on water. in Journal of Molecular Liquids 101, 219–260 (2002).

[24] Jorgensen, W. L. & Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. Journal of the American Chemical Society 110, 1657–1666 (1988).

[25] Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. Journal of the American Chemical Society 118, 11225–11236 (1996).

[26] Jorgensen, W. L. & Swenson, C. J. Optimized Intermolecular Potential Functions for Amides and Peptides. Hydration of Amides. Journal of the American Chemical Society 107, 1489–1496 (1985).

[27] Cornell, W. D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. Journal of the American Chemical Society 117, 5179–5197 (1995).

[28] Gromacs, Free energy tutorial, June 2017., http://www.gromacs.org/@api/deki/files/262/=gromacs-free-energy-tutorial.pdf

[29] Marenich, A.V., Kelly, C.P., Thompson, J.D., Hawkins, G.D., Chambers, C.C., Giesen, D.J., Winget, P., Cramer, C.J., Truhlar, D.G. Minnesota Solvation Database – version 2012, University of Minnesota, Minneapolis, 2012.

[30] Redmill, P. S., Capps, S. L., Cummings, P. T. & McCabe, C. A molecular dynamics study of the Gibbs free energy of solvation of fullerene particles in octanol and water. Carbon 47, 2865–2874 (2009).

[31] C. Pais, a. a. C., Sousa, A., Eusebio, M. E. & Redinha, J. S. Solvation of alkane and alcohol molecules. Energy contributions. Physical Chemistry Chemical Physics 3, 4001–4009 (2001).

[32] Lin, K., Zhou, X., Luo, Y. & Liu, S. The microscopic structure of liquid methanol from Raman spectroscopy. Journal of Physical Chemistry B 114, 3567–3573 (2010).

[33] Israelachvili, J. Intermolecular and Surface Forces. Intermolecular and Surface Forces (Elsevier Inc., 2011). doi:10.1016/C2009-0-21560-1

[34] Jorgensen, W. L., Gao, J. & Ravimohan, C. Monte Carlo simulations of alkanes in water: hydration numbers and the hydrophobic effect. The Journal of Physical Chemistry 89, 3470–3473 (1985).

[35] Baldwin, R. L. The new view of hydrophobic free energy. FEBS Letters 587, 1062–1066 (2013).

[36] Southall, N. T., Dill, K. A. & Haymet, A. D. J. A view of the hydrophobic effect. Journal of Physical Chemistry B 106, 521–533 (2002).

[37] Gallicchio, E., Kubo, M. M. & Levy, R. M. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: numerical results and implications for theories of hydrophobic solvation. Journal of Physical Chemistry B 104, 6271–6285 (2000).

[38] Abraham, M. H. & Matteoli E. The temperature variation of the hydrophobic effect. Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases 84.6 (1988): 1985-2000.

[39] Li, J.-L., Car, R., Tang, C. & Wingreen, N. S. Hydrophobic interaction and hydrogen-bond network for a methane pair in liquid water. Proceedings of the National Academy of Sciences 104, 2626–2630 (2007).

[40] Zangi, R., Zhou, R. & Berne, B. J. Urea's action on hydrophobic interactions. Journal of the American Chemical Society 131, 1535–1541 (2009).

[41] Maibaum, L., Dinner, A. R. & Chandler, D. Micelle Formation and the Hydrophobic Effect. J. Phys. Chem. B 108, 6778–6781 (2004).

[42] Goga, N., Rzepiela, A. J., De Vries, A. H., Marrink, S. J. & Berendsen, H. J. C. Efficient algorithms for langevin and DPD dynamics. Journal of Chemical Theory and Computation 8, 3637–3649 (2012).

[43] Ewald, P. P. (1921), Die Berechnung optischer und elektrostatischer Gitterpotentiale. Ann. Phys., 369: 253–287. doi:10.1002/andp.19213690304

[44] Parrinello, M. et al. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Chemical Physics 409, 7182–7190 (2013).

[45] Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. Journal of Chemical Physics 126, (2007).

A OPLA-AA parameters for alkanes

OPLS Methane								
atoms								
	nr	name	type	mass	charge	ptype	sigma	epsilon
	1	Opls_138	C	12,0110	-0.24	A	3.5E-01	2,76E-01
	2	Opls_140	H1	1,008	0.06	A	2.5E-01	1,26E-01
	3	Opls_140	H2	1,008	0.06	A	2.5E-01	1,26E-01
	4	Opls_140	H3	1,008	0.06	A	2.5E-01	1,26E-01
	5	Opls_140	H4	1,008	0.06	A	2.5E-01	1,26E-01
defaults								
nbfunc	comb-rule		gen-pairs		fudge LJ		fudge QQ	
1	3		yes		0,5		0,5	
bonds								
	i	j	funct		length		force.c	
	1	2	1		0,1090		284512	
	1	3	1		0,1090		284512	
	1	4	1		0,1090		284512	
	1	5	1		0,1090		284512	
angles								
i	j	k	funct		length		force.c	
2	1	3	1		107,80		276,144	
2	1	4	1		107,80		276,144	
2	1	5	1		107,80		276,144	
3	1	4	1		107,80		276,144	
3	1	5	1		107,80		276,144	
4	1	5	1		107,80		276,144	

OPLS Ethane										
atoms										
	nr	name	type	mass	charge	ptype	sigma	epsilon		
	1	Opls_135	C1	12,0110	-0.18	A	3.5E-01	2,76E-01		
	2	Opls_140	H1	1,008	0.06	A	2.5E-01	1,26E-01		
	3	Opls_140	H2	1,008	0.06	A	2.5E-01	1,26E-01		
	4	Opls_140	H3	1,008	0.06	A	2.5E-01	1,26E-01		
	5	Opls_135	C2	12,0110	-0.18	A	3.5E-01	2,76E-01		
	6	Opls_140	H4	1,008	0.06	A	2.5E-01	1,26E-01		
	7	Opls_140	H5	1,008	0.06	A	2.5E-01	1,26E-01		
	8	Opls_140	H6	1,008	0.06	A	2.5E-01	1,26E-01		
defaults										
nbfunc	comb-rule		gen-pairs		fudge LJ		fudge QQ			
1	3		yes		0,5		0,5			
bonds										
	i	j	funct		length		force.c			
	1	2	1		0,1090		284512			
	1	3	1		0,1090		284512			
	1	4	1		0,1090		284512			
	1	5	1		0,1529		224262,4			
	5	6	1		0,1090		284512			
	5	7	1		0,1090		284512			
	5	8	1		0,1090		284512			
angles										
i	j	k	funct		length		force.c			
2	1	3	1		107,80		276,144			
2	1	4	1		107,80		276,144			
2	1	5	1		110,70		313,8			
3	1	4	1		107,80		276,144			
3	1	5	1		110,70		313,8			
4	1	5	1		110,70		313,8			
1	5	6	1		110,70		313,8			
1	5	7	1		110,70		313,8			
1	5	8	1		110,70		313,8			
6	5	7	1		107,80		276,144			
6	5	8	1		107,80		276,144			
7	5	8	1		107,80		276,144			
dihedrals										
i	j	k	l	funct	c0	c1	c2	c3	c4	c5
2	1	5	6	3	0,62760	1,8828	0	-2,51040	0	0
2	1	5	7	3	0,62760	1,8828	0	-2,51040	0	0
2	1	5	8	3	0,62760	1,8828	0	-2,51040	0	0
3	1	5	6	3	0,62760	1,8828	0	-2,51040	0	0
3	1	5	7	3	0,62760	1,8828	0	-2,51040	0	0
3	1	5	8	3	0,62760	1,8828	0	-2,51040	0	0
4	1	5	6	3	0,62760	1,8828	0	-2,51040	0	0
4	1	5	7	3	0,62760	1,8828	0	-2,51040	0	0
4	1	5	8	3	0,62760	1,8828	0	-2,51040	0	0

Table 3: *Parameters for OPLS-AA methane and ethane*

OPLS Propane							
atoms							
nr	name	type	mass	charge	ptype	sigma	epsilon
1	Opls_135	C1	12,0110	-0.18	A	3.5E-01	2,76E-01
2	Opls_140	H1	1,008	0.06	A	2.5E-01	1,26E-01
3	Opls_140	H2	1,008	0.06	A	2.5E-01	1,26E-01
4	Opls_140	H3	1,008	0.06	A	2.5E-01	1,26E-01
5	Opls_136	C2	12,0110	-0.12	A	3.5E-01	2,76E-01
6	Opls_140	H4	1,008	0.06	A	2.5E-01	1,26E-01
7	Opls_140	H5	1,008	0.06	A	2.5E-01	1,26E-01
8	Opls_135	C3	12,0110	-0.18	A	3.5E-01	2,76E-01
9	Opls_140	H6	1,008	0.06	A	2.5E-01	1,26E-01
10	Opls_140	H7	1,008	0.06	A	2.5E-01	1,26E-01
11	Opls_140	H8	1,008	0.06	A	2.5E-01	1,26E-01
defaults							
nbfunc	comb-rule		gen-pairs	fudge LJ		fudge QQ	
1	3		yes	0,5		0,5	
bonds							
i	j		funct	length		force.c	
1	2		1	0,1090		284512	
1	3		1	0,1090		284512	
1	4		1	0,1090		284512	
1	5		1	0,1529		224262,4	
5	6		1	0,1090		284512	
5	7		1	0,1090		284512	
5	8		1	0,1529		224262,4	
8	9		1	0,1090		284512	
8	10		1	0,1090		284512	
8	11		1	0,1090		284512	
angles							
i	j	k	funct	length		force.c	
2	1	3	1	107,80		276,144	
2	1	4	1	107,80		276,144	
2	1	5	1	110,70		313,8	
3	1	4	1	107,80		276,144	
3	1	5	1	110,70		313,8	
4	1	5	1	110,70		313,8	
1	5	6	1	110,70		313,8	
1	5	7	1	110,70		313,8	
1	5	8	1	112,70		488,273	
6	5	7	1	107,80		276,144	
6	5	8	1	110,70		313,8	
7	5	8	1	110,70		313,8	
5	8	9	1	110,70		313,8	
5	8	10	1	110,70		313,8	
5	8	11	1	110,70		313,8	
9	8	10	1	107,80		276,144	
9	8	11	1	107,80		276,144	
10	8	11	1	107,80		276,144	
dihedrals							
i	j	k	l	funct	c0	c1	c2	c3	c4	c5
2	1	5	6	3	0,62760	1,8828	0	-2,51040	0	0
2	1	5	7	3	0,62760	1,8828	0	-2,51040	0	0
2	1	5	8	3	0,62760	1,8828	0	-2,51040	0	0
3	1	5	6	3	0,62760	1,8828	0	-2,51040	0	0
3	1	5	7	3	0,62760	1,8828	0	-2,51040	0	0
3	1	5	8	3	0,62760	1,8828	0	-2,51040	0	0
4	1	5	6	3	0,62760	1,8828	0	-2,51040	0	0
4	1	5	7	3	0,62760	1,8828	0	-2,51040	0	0
4	1	5	8	3	0,62760	1,8828	0	-2,51040	0	0
1	5	8	9	3	0,62760	1,8828	0	-2,51040	0	0
1	5	8	10	3	0,62760	1,8828	0	-2,51040	0	0
1	5	8	11	3	0,62760	1,8828	0	-2,51040	0	0
6	5	8	9	3	0,62760	1,8828	0	-2,51040	0	0
6	5	8	10	3	0,62760	1,8828	0	-2,51040	0	0
6	5	8	11	3	0,62760	1,8828	0	-2,51040	0	0
7	5	8	9	3	0,62760	1,8828	0	-2,51040	0	0
7	5	8	10	3	0,62760	1,8828	0	-2,51040	0	0
7	5	8	11	3	0,62760	1,8828	0	-2,51040	0	0

Table 4: *Parameters for OPLS-AA propane*

B OPLA-UA parameters for methanol

OPLS -UA methanol								
atoms								
	nr	name	type	mass	charge	ptype	sigma	epsilon
	1	Opls_080	C	15,0350	0,265	A	3,77500E-01	8,66E-01
	2	Opls_078	O	15,9994	-0,700	A	3,07E-01	7,11E-01
	3	Opls_079	H	1,008	0,435	A	0,00E+00	0,00E+00
defaults								
nbfunc	comb-rule		gen-pairs		fudge LJ		fudge QQ	
1	3		yes		0,5		0,5	
bonds								
	i	j	funct		length		force.c	
	1	2	1		0,1425		323004,8	
	2	3	1		0,0945		462750,4	
angles								
i	j	k	funct		length		force.c	
1	2	3	1		108,50		460,24	

Table 5: *Parameters for OPLS-UA methanol*

C Numerical results for all simulations

METHANOL						
	Methane		Ethane		Propane	
T[K]	ΔG [kJ/mol]	error	ΔG [kJ/mol]	error	ΔG [kJ/mol]	error
173	-1,57	0,22	-5,27	1,21	-11,05	0,47
200	-0,75	0,21	-5,85	0,27	-9,79	0,09
250	0,44	0,2	-3,85	0,16	-7,67	0,14
273	0,97	0,21	-3,54	0,18	-6,29	0,3
300	1,52	0,11	-2,23	0,26	-5,49	0,3
333	1,63	0,11	-1,61	0,16	-4,51	0,17
370	1,76	0,09	-1,1	0,11	-3,65	0,11
400	1,45	0,06	-0,93	0,16	-3,68	0,02
430	1,21	0,09	-1,12	0,07	-3,56	0,2
SPC/E WATER						
	Methane		Ethane		Propane	
T[K]	ΔG [kJ/mol]	error	ΔG [kJ/mol]	error	ΔG [kJ/mol]	error
300	8,81	0,23	9,15	0,31	9,71	0,24
TIP4P/2005 WATER						
	Methane		Ethane		Propane	
T[K]	ΔG [kJ/mol]	error	ΔG [kJ/mol]	error	ΔG [kJ/mol]	error
300	8,89	0,07	8,67	0,37	8,9	0,16
350	11,23	0,24	12,17	0,12	12,21	0,3

Table 6: *Results from all simulations mentioned in this work*

D Parameters used in simulations

Parameters used in simulations	
Integrator	sd (leap-frog stochastic dynamics integrator) (ref [42])
Time step	0.001 [ps]
Rlist	1
Nstlist	10
Coulombtype	pme (ref [43])
rvdw	1.5 [nm]
rcoulomb	1.5[nm]
Constraints	all-bonds
Thermostat	v-rescale (ref [45])
tau-t	0.2
Barostat	Parrinello-Rahman (ref [44])
tau-p	5
Periodic boundary conditions (pbc)	xyz
couple-intramol	no
couple-lambda1	vdwq (full interaction when lambda=1)
couple-lambda0	none (no interaction when lambda=0)

Table 7: *Parameters used in main run of simulations*