TÉCNICO LISBOA

Water + hydrocarbons mixtures and the oil extraction industry.

A computer simulation study.

Pedro José do Vale Duarte

Theses to obtain the Master of Science Degree in

Chemical Engineering

Supervisors: Doctor Eduardo Jorge Morilla Filipe and
Doctor Pedro Jorge Rodrigues Morgado

Examination Committee

Chairperson: Doctor Sebastião Manuel Tavares Silva Alves
Supervisor: Doctor Eduardo Jorge Morilla Filipe
Members of the Committee: Doctor Luís Filipe Guerreiro Martins

December 2015

Acknowledgements

Human beings are built to be social and out of that comes our ability to learn from each other. Each of us can only advance a small fraction of the collective knowledge we are given by all the other member of our species. This work is yet another tiny step in broadening our understanding of nature, supported on the work of countless others.

In particular I would like to thank two very important people without whom this work would never have been done. The first is Professor Eduardo Filipe who has given me a place to find new things and be a part of something exciting for the last three years of my college studies. He is a teacher at heart, with a love for what he does that perspires in every conversation and inspires all his students. The second person is Pedro Morgado, someone that I could describe as Professor Eduardo's right hand. He is an incredible researcher and academic, with intricate knowledge on all the working apparatuses and experiments at play within our laboratory. To him I owe the tireless guidance throughout all the work, from the first simulation to the last word written here.

I would also like to mention some of my colleges, Tiago Cruz, João Marques and Pedro Couto, that were always present to engage in conversation about what we learned and did throughout this past years, and hope we can continue doing so for many more.

Finally, but not least relevant, I would like to thank my family for their support in my *non-linear* progression through college.

Abstract

The solubility of water in liquid alkanes and their mixtures is of vital importance to the gas and petroleum industries. Preventing corrosion and hydrate crystal formation in subsea and terrestrial pipelines are examples of operating and safety problems, which require a precise knowledge of water content in hydrocarbons. Given the importance of the subject, a large number of results can be found in the literature. However, significant disagreement is found between these results, surely reflecting the difficulty of these measurements.

In this work, a molecular-level interpretation of experimental results of water solubility in alkanes was performed using all-atom molecular dynamics simulations (MD). This approach helps to rationalize the observed trends in solubility and derived solvation enthalpy, with particular insight on the differences in the structuration of the solvents around the water molecules and on the associated interaction energies.

Keywords – Water, Alkanes, Hydrocarbons, Liquid structure, Molecular Dynamics, Solvation.

Resumo

A solubilidade da água em alcanos e misturas de alcanos líquidos é de importância vital para a indústria petroquímica. A prevenção de corrosão e a formação de cristais hidratados em *pipelines* terrestres e submarinos são exemplos de problemas de operação e segurança, que requerem conhecimento preciso da quantidade de água presente nos hidrocarbonetos. Dada a importância deste tema é possível encontrar uma vasta quantidade de resultados na literatura. No entanto, estes apresentam disparidades significativas, que certamente refletem a dificuldade na sua obtenção.

Neste trabalho, foi feita uma interpretação molecular de um conjunto de dados experimentais de solubilidade de água em alcanos, usando simulações *all-atom* de dinâmica molecular. Esta abordagem permite aprofundar o conhecimento sobre as tendências observadas na solubilidade e entalpia de solvatação (derivada da solubilidade), com particular foco nas diferenças da estrutura do solvente em torno da molécula de água (soluto) e nas energias de interação.

Palavras-chave – Água, Alcanos, Hidrocarbonetos, Estrutura líquida, Dinâmica Molecular, Solvatação.

Contents

List of tables

List of figures

Chapter 1 Introduction

Water is used in all industrial plants, whether directly in the process or just as an utility (ex. as a heat transfer fluid). When hydrocarbons are present somewhere along a factory's process, even the most advanced technologies might not be enough to prevent contamination, on both sides (water in hydrocarbons and vice-versa).

In no other industry is the water and alkane system knowledge more important than the gas and petroleum industry. From the very start of the process water is a relevant component. It is always present in reservoirs in equilibrium with hydrocarbons, and throughout the production life of the reservoir its production increases, being sometimes greater than the production of hydrocarbons at the end. [1]

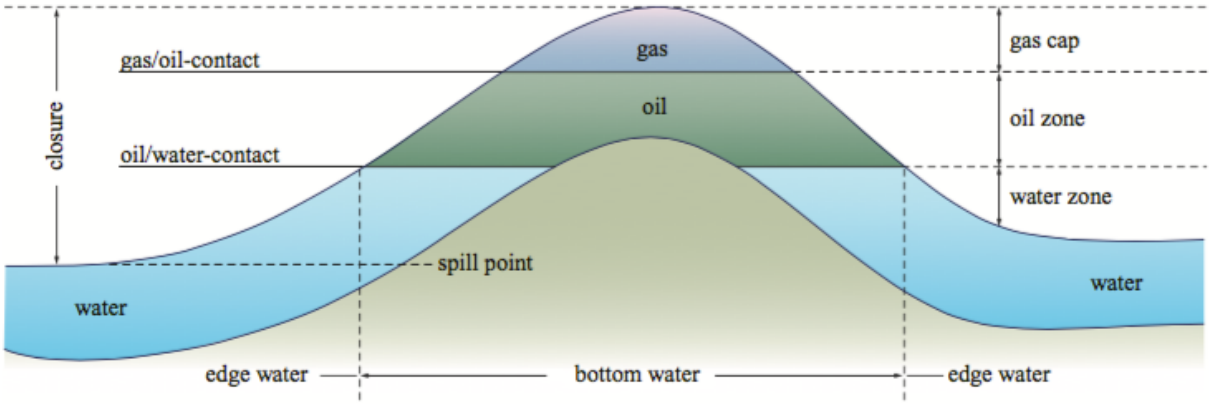

Figure 1 – Diagram showing a folded sandstone layer representing a reservoir trap. At the apex of this anticline, natural gas and oil below has accumulated. In the pore space of the gas cap and the oil zone, the original pore water was displaced by gas and oil respectively, while below the oil/water-contact the sandstone remains water-saturated. [2]

Considering this, being able to understand and predict de behaviour of these systems becomes of the utmost importance for the process design and operation, by being able to predict the phase equilibrium. Also, it is possible to mitigate hydrate crystal formation inside the reservoirs and on transfer pipelines and model pollutant dispersion in the environment. [3]

1.1 Water and Hydrocarbon Mixtures

Water and hydrocarbon mixtures are very non-ideal, since water and hydrocarbons are very different substances, with different interactions when pure. Water is a polar molecule and has the capacity to establish hydrogen bonds and dipole-dipole interactions (Keesom force) that have great importance for its properties and behaviour. Each molecule can be part of, as much as, 4 hydrogen bonds at the same time, by donating two hydrogen bonds through the two protons (hydrogen atoms), and accepting two more hydrogen bonds through the two sp^3-hybridized lone pairs (on the oxygen atom). This creates a 3D network of interactions between water molecules. On the other hand, hydrocarbons are non-polar

making the most important forces different from those of water. In hydrocarbons, dispersion interactions dominate (London dispersion force) which are a result of polar interactions between an instantaneous dipole and an induced dipole. They increase with the molar mass, causing a higher boiling point.

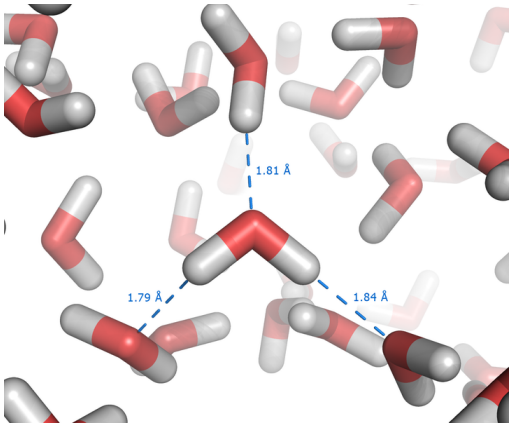

Figure 2 – Hydrogen bonds in liquid water from molecular dynamics simulation. [4]

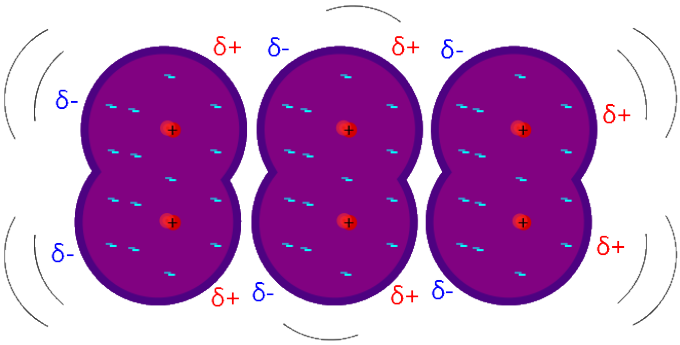

Figure 3 – Induced dipoles, London's dispersion force in hydrocarbon–hydrocarbon interactions. [5]

Given their non-ideal profile, these mixtures exhibit limited miscibility over a wide range of temperatures, thus giving rise to two distinct phases: a hydrocarbon rich phase containing a very small concentration of water; and a water-rich phase containing an even smaller concentration of dissolved hydrocarbon.

1.1.1 Water rich Phase

Hydrocarbon molecules don't have the ability to form hydrogen bonds. Hence they will interact with water molecules through induction (Debye) and dispersion (London) interactions. When a hydrocarbon molecule is inserted into water, a cavity is required to house the molecule, which leads to a disruption of the 3D network of hydrogen bonds. This forces the water molecules that are on the surface of the hydrocarbon to reorient tangentially to that same surface, in such a way as to make as many hydrogen bonds with neighbouring molecules as possible. These water molecules at the surface have reduced mobility and form a structured water "cage" around the non-polar molecule. [6] The structuring of water along the surface of the hydrocarbon reduces the mixture's entropy proportionally to the cavity's size. This is known as the hydrophobic effect. [7,8]

Since every system tends to have as much entropy as possible, when more than one molecule of hydrocarbon is present, and near another, they tend to "join" in order to reduce the surface area of the non-polar aggregate, reducing the amount of structured water molecules and increasing the entropy (relative to what it would be if the molecules of hydrocarbon were separate).

Most of the published experimental data indicates that the solubility of hydrocarbons in water is highly dependent on the hydrocarbon, with larger molecules being less soluble [9]. This follows from the behaviour described above, where larger molecules mean larger surface area and greater loss of entropy. In some cases enthalpy may add to this effect, i.e. some high-energy hydrogen bonds are replaced by

the weaker dipole-induced dipole interactions, accounting for positive solvation enthalpies. However, this is not the key factor for the observed immiscibility. [10]

1.1.2 Hydrocarbon rich Phase

On the other phase, a single water molecule dissolved in a hydrocarbon has lost all its hydrogen bonds, and now interacts essentially through induction (Debye) and dispersion (London) forces, which depend on the polarizability and density of the solvent. In other words, whereas the behaviour of water is dominated by its polarity and ability to form hydrogen bonds, the behaviour of hydrocarbons is dictated by their essentially nonpolar and flexible nature, where the dispersive interactions prevail. The main challenge for any theoretical or computational modelling of these systems that attempts to calculate or predict the phase equilibrium lies in the need to simultaneously account for the highly asymmetric nature of the two coexisting phases.

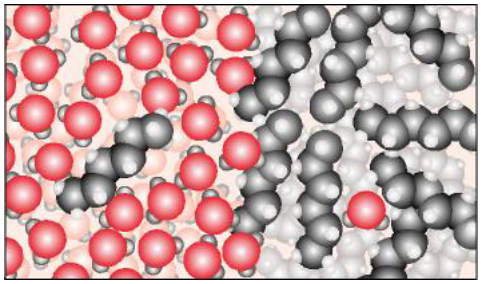

Figure 4 – Molecular representation of a mixture of water with n-hexane. From left to right we see: a water rich phase; the interface; a hydrocarbon rich phase.

1.2 State of the art

As mentioned, the importance of understanding water-hydrocarbon mixtures is great, whether for environmental sustainability or to improve efficiency in extraction and handling of petroleum reserves, to which we are still largely dependent even with the most recent advancements in renewable energy sources. The design of equipment that processes these hydrocarbon mixtures is dependent on accurate data for the behaviour of these mixtures.

Despite the technological and fundamental importance of accurate data on *n*-alkane + water mutual solubilities, the available literature data are widely scattered, certainly reflecting the difficulty of the measurements; it is not uncommon to find data from different authors that differ by more than 100% or even a full order of magnitude. Besides, an accurate and coherent set of data that could serve as a reference has not yet been established.

Moreover, the available data is, in most cases, not for the same conditions in which the process takes place resulting in reduced efficiencies and losses, due to inaccurate dimensioning. Data on such extreme conditions is usually difficult to get. Therefore, accurate models that describe the behaviours of these systems become essential in order to get reliable, or at least more accurate, thermodynamic properties.

The International Union of Pure and Applied Chemistry (IUPAC) and the National Institute of Standards and Technology (NIST) published an exhaustive compilation [9] of mutual hydrocarbon + water solubility data, which included a critical evaluation based on the use of a cubic equation of state with an additional term that accounts for hydrogen bonding. [11] Tsonopoulos proposed correlations [12] for the mutual *n*-alkane + water solubilities, both as a function of the alkane chain length at ambient temperature and for some of the systems as a function of temperature. Very recently, Fouad *et al*. [13] proposed a generalized correlation for the solubility of water in *n*-alkanes, as a function of chain length and temperature; their approach was based on the combined use of the very accurate IAPWS empirical equation of state for water [14] and the PC-SAFT equation of state [15] for the alkane-rich phase.

The properties of liquids and liquid mixtures are known to depend largely on the organization of the fluid, for which molecular shape (i.e. repulsion forces) is a key factor. Nevertheless, and in spite of considerable efforts, modelling and predicting the structure of liquids remains a major challenge even to state of the art theories of fluids and to detailed computational models. Thermodynamic data, although largely used for this purpose, is unable to provide direct structural information at the molecular level. Therefore, combining thermodynamic studies with microscopic information obtained, for example, from spectroscopic techniques can be an important step towards the elucidation of the structure of liquids. The use of computer simulations is another way to obtain evidence on the structure of liquids. In the following work, Molecular Dynamics (MD) simulations were employed to gain such an insight.

1.3 Computational Chemistry and Molecular Modeling [16,17]

In this dissertation several systems were studied using a *Computational chemistry* and *Molecular modelling* approach. *Computational Chemistry* is, essentially, applying computers and computational techniques in chemistry, with a focus that can go from the quantum mechanics of molecules to the dynamics of large complex molecular systems. *Molecular modelling* is the process by which complex systems are described using realistic atomic models, aiming to understand and predict the macroscopic properties through a detailed knowledge on an atomic scale. Often molecular modelling is employed on the design of new materials, for which accurate prediction of their macroscopic properties is required.

Macroscopic physical properties can be distinguished in two types. (1) Static equilibrium properties, such as the system's density, vapour pressure, the average potential energy or any radial distribution function, and (2) dynamic properties – also referred as non-equilibrium properties – such as the viscosity of a liquid or diffusion processes. The choice of computational technique depends on the feasibility of the method to deliver results that can be considered reliable, at the present state of the art. Ideally, each system would be treated using the (relativistic) time dependent Schrödinger equation, which, at least conceptually, yields the highest accuracy in describing the molecular systems and, consequently, the best properties. But, for systems that are more complex than a few atoms this *ab initio* approach (as it is called) becomes too expensive, in terms of computational power, and hence unfeasible. Thus, the use of approximations becomes a necessity; the more complex a system is and the longer it needs to run, for better statistical accuracy of the results, the more severe the approximations must be.

At a given point, the *ab initio* approach must be augmented or replaced by an empirical parameterization of the model used. And, where simulations using, only, atomic interaction fail, due to the system's complexity, the molecular modelling approach based entirely on a similarity analysis of known molecular structures, is the most feasible.

Since most properties are ensemble averages over a representative statistical ensemble of molecular configurations, this implies for molecular modelling that the knowledge of a single structure, even if it is the global energy minimum, is not enough. It is necessary to generate, at a given temperature, a representative ensemble from which to compute macroscopic properties.

Also, while molecular simulations, in principle, provide details on atomic structures and motions, these details are often not fundamental for the macroscopic properties of interest. Given this excess information it is possible to simplify, based on the science of statistical mechanics as a framework, the description models and average over the irrelevant details, when obtaining the properties of interest.

To obtain a representative equilibrium ensemble two methods are available: (1) *Monte Carlo simulations* (not used in the following work) and (2) *Molecular Dynamics simulations* (MD).

1.3.1 Molecular Dynamics Simulations

While Monte Carlo simulations are simpler to implement than Molecular Dynamics (MD), because they don't require the computation of forces, they do not yield significantly better statistics than MD considering a certain computation time. Because it allows the calculation of transport properties, MD can be considered a more universal technique.

MD simulations usually start with an energy minimisation step, since the starting configuration might be very far from equilibrium, resulting in the computation of excessively large forces and a consequent simulation failure.

Molecular Dynamics simulations solve Newton's equations of motion for a given system of N interacting particles,

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i, \qquad i = 1 \dots N \tag{1}$$

where m_i is the mass of the particle, $\frac{\partial^2 r_i}{\partial t^2}$ is the acceleration, F_i the resulting force and N is the total number of particles.

The equations are solved simultaneously in small time steps (whose span is related to the interactions and forces at play in the system). The system's progress is followed for as long as it is deemed required for the calculation of a given property, taking care to maintain it in the required thermodynamic ensemble and at the specified conditions, while recording a system "snapshot" (particle coordinates and/or relevant properties) at regular intervals. The recording of positional coordinates as a function of time gives a trajectory of the system, which can be considered that of a succession of equilibrium configurations, after an initial equilibration time. The macroscopic physical properties can be extracted from the output file (or files), by averaging over this equilibrium trajectory.

1.3.1.1 Limitations

As with any method, there are limitations; in this case, they arise from approximations in the workings of the simulation and from those assumed by the model. Each limitation has, almost always, a mitigation measure that can be employed while doing the simulations.

1.3.1.1.1 Simulations are classical

The simulations are, as the use of Newton's equations implies, classical mechanical simulations. This approximation is valid for most atoms at normal temperature, save some exceptions, most notably, hydrogen atoms. Hydrogen atoms are quite light and the motion of protons (hydrogen nuclei) is sometimes of essential quantum mechanical character. For instance, a proton may *tunnel* through a potential barrier in the course of a transfer over a hydrogen bond. As expected, such processes (related to quantum mechanics) cannot be treated by classical mechanics.

The statistical mechanics of a classical harmonic oscillator differs appreciably from that of a real quantum oscillator when the resonance frequency ν approximates or exceeds $k_B T/h$. At room temperature this gives a wavenumber of approximately 200 cm^{-1}. Thus, all wave numbers higher than about 100 cm^{-1} may misbehave in classical simulations. This means that practically all bond and bond-angle vibrations are suspect, even hydrogen-bonded motions as translational or librational H-bond vibrations are beyond the classical limit, where a quantum oscillator would be more accurate to describe them (see Table 1).

Table 1 – Typical vibrational frequencies (wavenumbers) in molecules and hydrogen-bonded liquids.

Type of bond	Type of vibration	Wavenumber / cm^{-1}
C–H, O–H, N–H	Stretch	3000-3500
C=C, C=O	Stretch	1700-2000
H–O–H	Bending	1600
C–C	Stretch	1400-1600
H_2CX	Rock	1000-1500
CCC	Bending	800-1000
O–H · · · O	Libration	400-700
O–H · · · O	Stretch	50-200

To solve this problem, apart from doing quantum-dynamical simulations, two solutions exist: (1) a correction factor can be included for the calculation of the system's properties; (2) the bonds and bond angles can be fixed as constraints in the equation of motion – the rationale behind this is that a quantum oscillator in its ground state resembles a constrained bond more closely than a classical oscillator. As a result of this approximation the algorithm can use larger time steps, since the highest frequencies (higher wave number) are *frozen*.

1.3.1.1.2 Electrons are in the ground state

In MD we use a *conservative* force field that is a function of the positions of atoms only. This means that the electronic motions are not considered: the electrons are supposed to adjust their dynamics instantly when the atomic positions change (the *Born-Oppenheimer* approximation), and remain in their ground state. This makes electron transfer processes and electronically excited states out of the applicability scope for these simulations. Also, chemical reactions can't be properly treated.

1.3.1.1.3 Boundary conditions are unnatural

Given that the systems are usually small (less than a million atoms), when compared to experimental samples, due to computational restrictions, a cluster of particles would have a lot of unwanted boundary with its environment (vacuum) if the simulations were conducted with real phase boundaries. To simulate a bulk system periodic boundary conditions are used to avoid real phase boundaries. Since liquids are not crystals, something unnatural remains. For large systems, the errors are small, but for small systems

with a lot of internal spatial correlation, the periodic boundaries may enhance that same internal correlation. In case that may be happening it is possible to test the influence of the system's size.

1.3.1.1.4 Long-range interactions are cut off

A cut-off radius is used for the Lennard-Jones interactions and for the Coulomb interactions. The "minimum-image convention" requires that only one image of each particle in the periodic boundary conditions is considered for a pair interaction, hence the cut-off radius cannot exceed half the box size. This yields a missing energy contribution for interactions that would occur at larger distances than the cut-off. To account for these interactions corrections can be included using methods like the Ewald summation, that estimate the interactions each particle would get from virtual particles up to infinite distance, and analytical tail corrections can be added for the dispersion energies.

1.3.1.2 Definitions

1.3.1.2.1 Thermostat

The system is coupled to a heat bath to ensure that its average temperature is maintained close to the requested temperature, T_{ext}. When this is done the equations of motion are modified and the system no longer samples the microcanonical ensemble*. Instead trajectories in the canonical (*NVT*) ensemble† are generated.

1.3.1.2.2 Barostat

The size and shape of the simulation cell may be dynamically adjusted by coupling the system to a barostat in order to obtain a desired average pressure, p_{ext}. This implies that the volume is not fixed, such as in an *NpT* simulation, so that it can change to maintain the pressure.

In this work the Nosé-Hoover thermostat and barostat is used, which alters the Newton's equations of motion of the particles and scales the size of the system, keeping the Helmholtz free energy constant.

1.3.1.2.3 Force Field

The outcome of the simulations is primarily controlled by the expressions for the total energy, which are collectively referred to as the force field. A force field is built up from two components:

- The set of equations (called the *potential functions*) used to generate the potential energies and their derivatives, the forces.
- The parameters used in this set of equations.

---

* In statistical mechanics, a microcanonical (or NVE) ensemble is the statistical ensemble that is used to represent the possible states of a mechanical system that has an exactly specified total energy.

† In statistical mechanics, a canonical ensemble is the statistical ensemble that represents the possible states of a mechanical system in thermal equilibrium with a heat bath at some fixed temperature. Sometimes called NVT ensemble.

Within one set of equations various sets of parameters can be used. The combination of equations and parameters form a consistent set. The force field used throughout this work is a variation, regarding the parameters for long hydrocarbon groups, of the Optimized Potentials for Liquid Simulations All-Atom [18] (OPLS-AA), the L-OPLS [19].

Chapter 2 Simulation Details

Simulations were performed for systems consisting of a liquid linear alkane (*n*-hexane, *n*-heptane, *n*-nonane, *n*-undecane, *n*-tridecane or *n*-hexadecane) with a single dissolved water molecule.

The alkanes were modelled with an optimized version [19] of the well-known OPLS-AA [18] force field. This version, designated by L-OPLS, was developed to improve the description of alkanes with six or more carbon atoms, and for these compounds it achieves a very good agreement with properties such as density, vaporization enthalpy, self-diffusion coefficient, viscosity and gauche-to-trans ratio. The L-OPLS is a fully atomistic force field, where each atom interacts through a Lennard-Jones potential and is assigned a partial electrostatic charge, and the intramolecular structure explicitly includes bond stretching, angle bending and dihedral torsions.

The water molecule was represented by the SPC/E force field, [20] a rigid 3-centre model with the intermolecular interactions mediated by a single Lennard-Jones centre on the oxygen atom, plus three partial electrostatic charges on the oxygen and hydrogen atoms. This model has previously been used in other studies where water played the role of solute. [21,22,23]

Following the OPLS framework, the cross interaction dispersion parameters were obtained using the geometric mean rule. The molecular dynamics simulations were performed using the *DL_POLY Classic* [16] code, with studied systems consisting of one water molecule and between 45 to 100 solvent molecules, with periodic boundary conditions in all directions. The initial liquid box sizes were established according to the experimental densities. Systems were equilibrated in the *NpT* ensemble for 0.5 ns, and then 10 ns production runs were performed to accumulate averages. For both runs, a time step of 2 fs was used. In alkanes, the vibrations of bonds involving hydrogen atoms have been constrained to their equilibrium distances using the SHAKE algorithm, whereas water was treated as a rigid body using the Fincham Implicit Quaternion Algorithm, as implemented in the *DL_POLY* program.

A cut-off distance of 13 Å was used for both non-bonded Lennard-Jones and electrostatic potentials. The Ewald summation technique was used to account for the electrostatic interactions beyond the cut-off, and standard analytic tail corrections for the energy and pressure dispersion terms were added. A neighbour list, with a radius of 14.3 Å, was used, which is updated around every 20 time steps.

Simulations were done at atmospheric pressure and at two temperatures for each solvent: 298.15 K and the reduced temperature of 0.538. In Table 2, the temperature and the number of solvent molecules for each simulation are collected. In all the simulations, temperature and pressure were controlled by the Nosé-Hoover thermostat and barostat with coupling constants of 0.5 ps for temperature and 2.0 ps for pressure.

The systems' densities were calculated and compared to the experimental densities of *n*-alkanes [24] (also collected in Table 2) as an element of validation of the force field used. The calculated densities were

obtained from the simulations with the solute molecule, which was considered to have a negligible effect on the density of the systems.

Table 2 – Temperatures (T_1 and T_2), number of solvent molecules (N) used in simulations, simulated densities (ρ_1 and ρ_2) and the deviations between simulation and experimental densities (σ_1 and σ_2) [25].

Solvent	T_1 / K	T_2 / K	N	ρ_1 / kg.m^{-3}	ρ_2 / kg.m^{-3}	σ_1 / %	σ_2 / %
n-hexane	298.15	273.2	100	645.9	673.7	-1.37	-0.50
n-heptane	298.15	290.8	90	673.6	678.5	-0.87	-1.05
n-nonane	298.15	320.1	70	707.6	687.3	-1.44	-1.35
n-undecane	298.15	344.0	65	731.4	690.8	-1.16	-1.65
n-tridecane	298.15	362.9	55	750.1	692.8	-0.78	-1.78
n-hexadecane	298.15	388.7	45	769.0	693.1	-0.13	-1.95

Chapter 3 Results and Discussion

3.1 Radial Distribution Functions

Radial distribution functions (rdf) were obtained from the simulations, in order to analyse the local structure around the solute molecule. In Figure 5 are presented the rdf between the oxygen atom of water and the methyl (CH_3) and methylene (CH_2) carbons of the *n*-alkanes, calculated from the simulations at 298.15 K and on Figure 6 are the same rdf at a constant reduced temperature of 0.538.

Given the structure of the water molecule, the point for the centre of mass is practically the same as the oxygen centre, which is also the van de Waals sphere centre. This makes the results for the oxygen atom representative of the whole molecule. However, on Appendix A contains the analogous rdf with hydrogen as the reference atom.

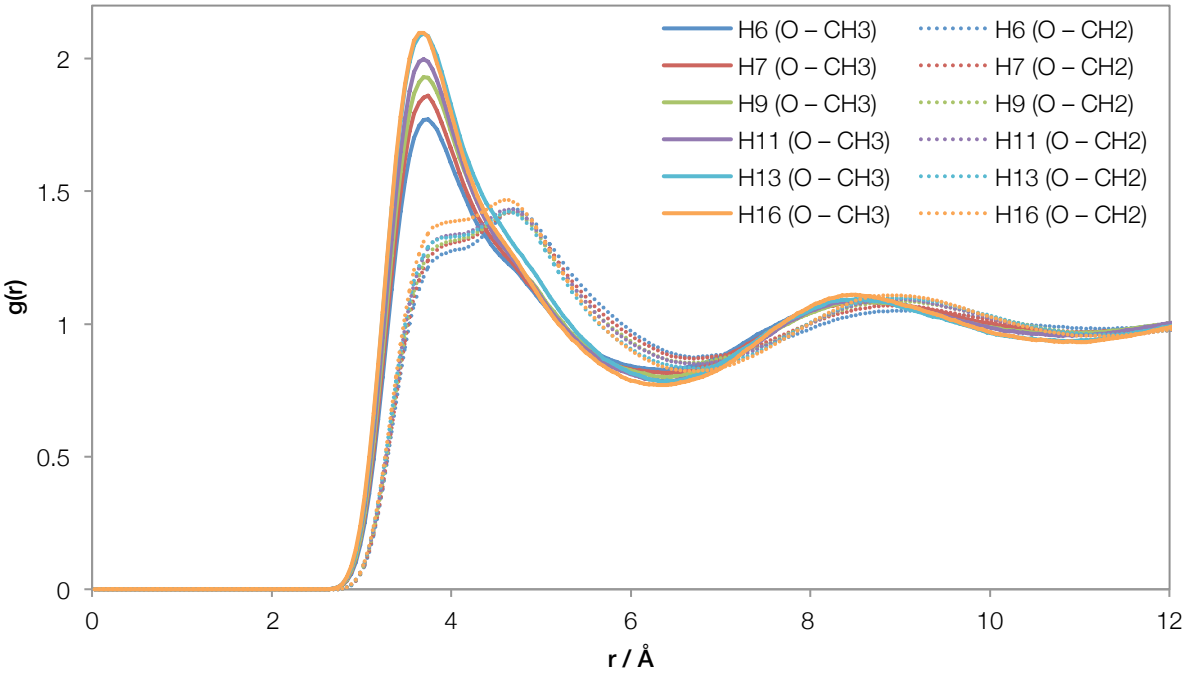

Figure 5 – Radial distribution functions between CH_3 (solid lines) and CH_2 (dotted lines) and water's oxygen atom in linear alkanes, at 298.15 K.

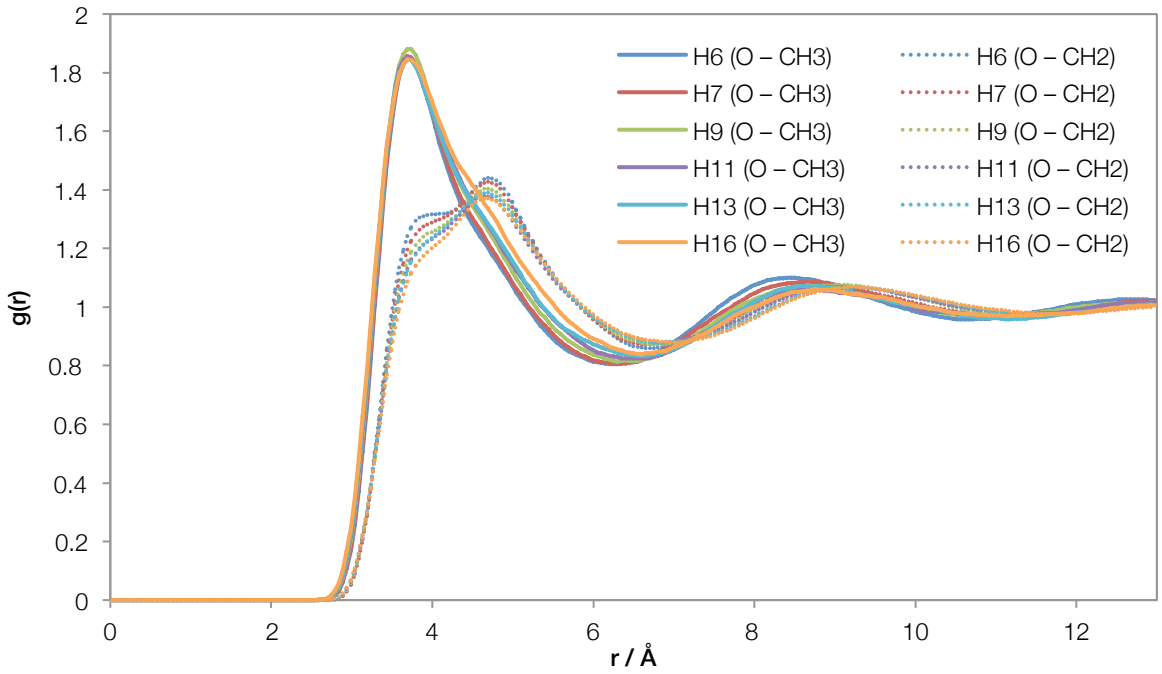

Figure 6 – Radial distribution functions between CH_3 (solid lines) and CH_2 (dashed lines) and water's oxygen atom in linear alkanes, at 0.538 reduced temperature.

As can be seen on Figure 5 (and Figure 6), the height of the water–CH_3 peak clearly increases with the length of the *n*-alkane solvent, whereas the water–CH_2 peak seems to be less sensitive. The CH_3 groups can always approach the water molecule at closer distances, in spite of their larger volume, and at T=298.15 K the corresponding peaks are systematically more intense than the water–CH_2 for distances under ~4.5 Å. The combination of these effects suggests that the water molecules have a preferential tendency to be dissolved in the vicinity of methyl groups and that this tendency might increase with chain length at the T=298.15 K.

To further check this hypothesis, we have integrated the radial distribution functions, thus obtaining the number of interaction sites (*N*) of each type in a coordination shell around the reference site as:

$$N = 4\pi \int_0^r g(r) r^2 \rho \qquad \mathbf{2}$$

where *r* is the radius of the coordination shell and ρ is the segment bulk density. Given the uncertainty in defining the limits of the coordination sphere, we have calculated the number of sites of a given group around water, as a function of the distance from its centre.

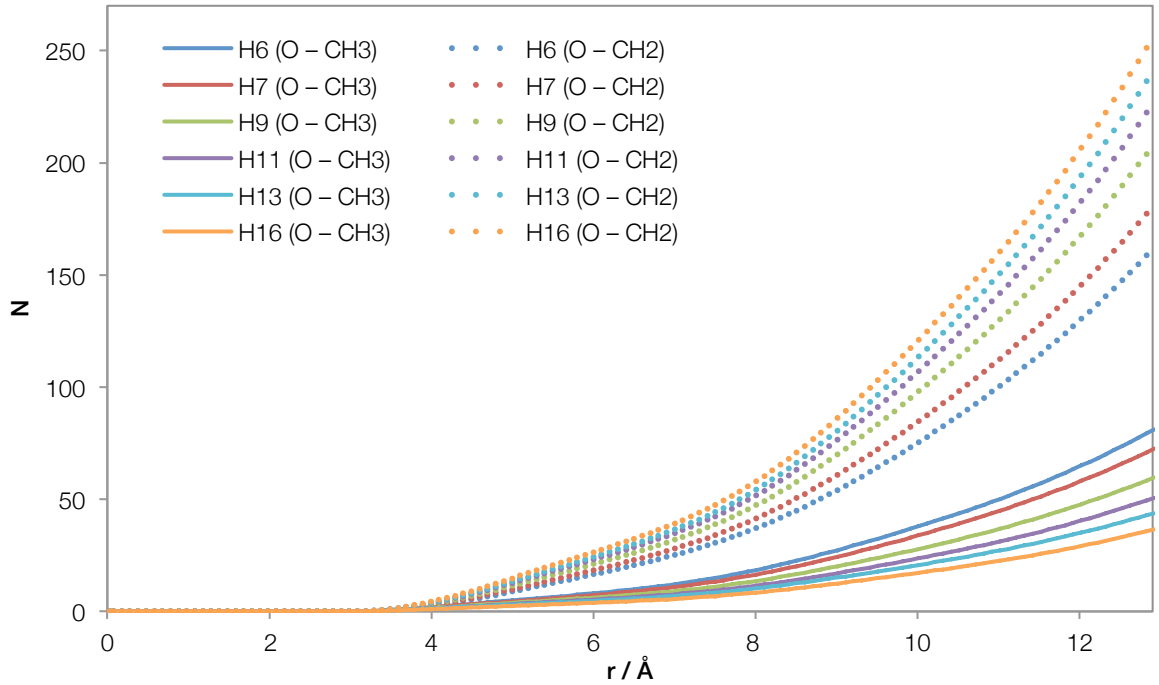

Figure 7 – Number of carbon sites of methyl and methylene groups around water's oxygen atom as a function of distance for linear alkanes at T=298.15 K.

As expected, the number of methyl and methylene segments in the vicinity of the water molecule increases with the respective radius and, in general, the ratio between the two essentially reflects the proportion in the solvent molecule. However, if we represent the ratio between the local molar fraction of CH_3 groups around water and their bulk molar fraction, as a function of the radius *r* of the coordination sphere, it becomes clear that the close vicinity of water is always enriched in methyl groups. The result is similar for both temperatures and for the reduced temperature the figure can be found on Append A, Figure 16. In Figure 8, at T=298.15 K, it can be seen that the fraction of CH_3 groups is always higher than the bulk value at low *r*, tending to the bulk value for large *r*. For almost all *n*-alkanes (save tridecane), the fraction of CH_3 groups becomes smaller than the bulk fraction for intermediate values of *r*. This is an indication of a local enrichment in CH_3 around the solute over the global composition at short distances, and that this enrichment is more pronounced as the solvent chain length increases. The same kind of structural effect was already observed in a previous work on the interaction between xenon and alkanes [26], suggesting that a small solute will preferentially be solvated in a region of the liquid *n*-alkane that is enriched in terminal chain groups.

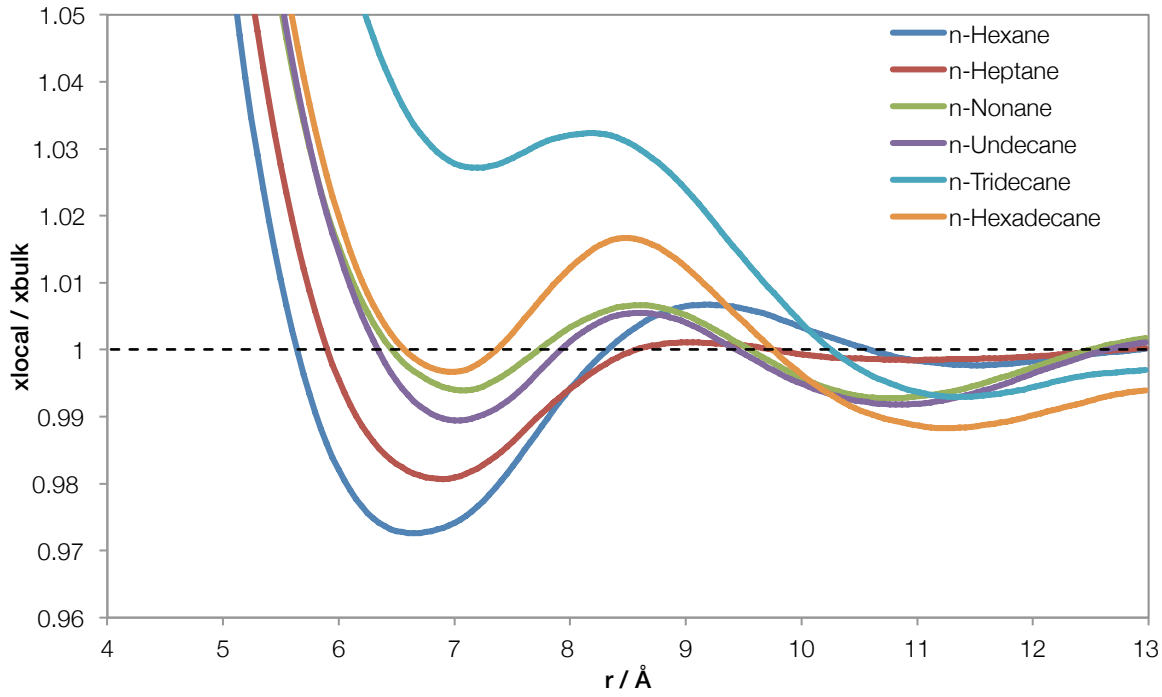

Figure 8 – Ratio between the local fraction of CH_3 groups around water's oxygen atom and the bulk fraction, as function of distance in linear alkanes (at a temperature of 298.15 K).

The apparent switch in order for the curves of some of the bigger alkanes, when compared to the order observed for the rdf at T=298.15 K, prompted us to check for any issues with the entry files and, in particular, the motion of the particles inside in the systems. To evaluate this motion, self-diffusion coefficients were obtained for all the solvents with the results indicating particle motion consistent with experimental values. However, this apparent shift may most likely be due to the radius interval where it occurs in the rdf, because with increasing radius it becomes more difficult to take information of such fine detail.

3.1.1 Self-Diffusion Coefficients

From the simulations it is possible to obtain an average of the square of the displacement of a given type of particle (mean square displacement – MSD) as a function of time ($MSD(t)$), and the diffusion coefficient can be obtained from the slope of the linear part of this curve, multiplied by a constant (equation **3**) [27].

$$D = \frac{1}{6}\frac{dMSD(t)}{dt} \qquad \textbf{3}$$

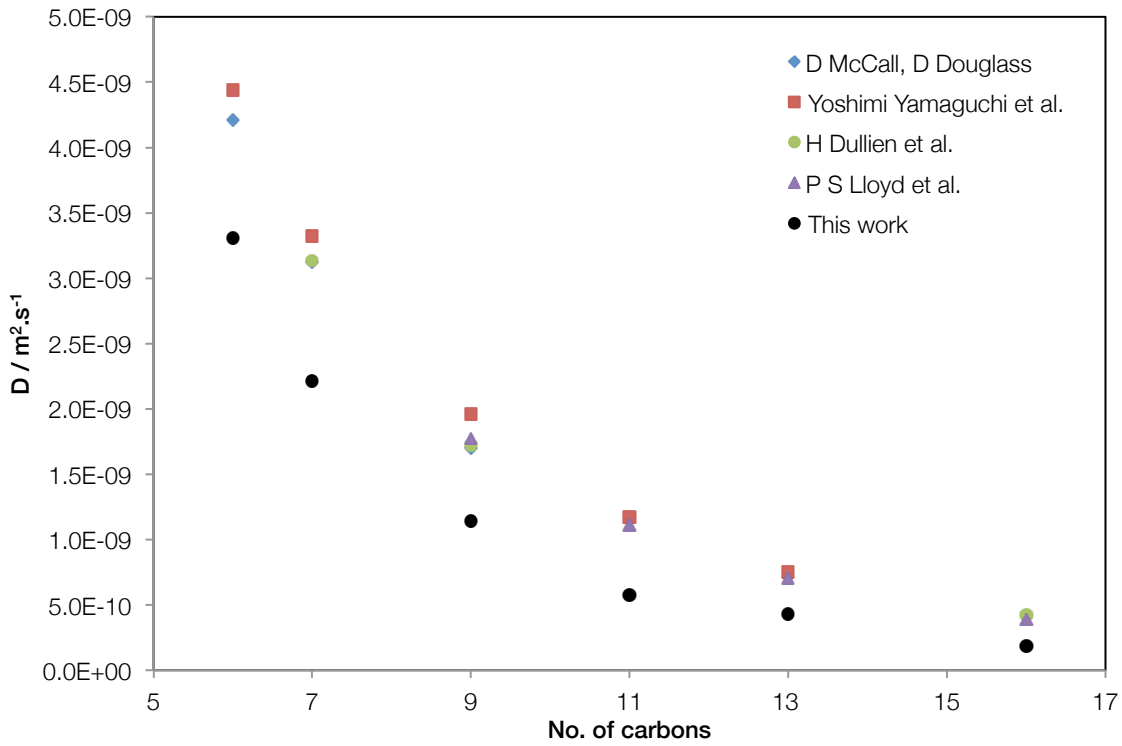

Figure 9 - Self-diffusion coefficients obtained from the simulations (black dots) and from the literature [28] [29] [30] [31], for each *n*-alkane at a temperature of 298 K.

Observing the figure above we see that the self-diffusion coefficient values for the solvents at 298 K differ no more than 56.6 % from the experimental values found in the literature. Moreover, the data displays a similar trend to that of the experimental results. This indicates that the systems are not stuck on a dynamic loop, especially for the longer *n*-alkanes with a fusion point closer to 298 K.

3.1.2 Terminal Carbon RDFs

Additional RDFs were calculated around the terminal carbon (CT) of the solvent molecules. The aim was to see if the *n*-alkane solvents already have statistically relevant concentrations of CH_3 groups near each other, higher than the bulk concentration. This analysis was performed on the same simulated systems, and it was considered that the presence of the solute molecule has a negligible effect on the global

structure of the solvent, at least for a first, qualitative analysis. On Figure 10 we can see a profile that mirrors the shape of the one seen on Figure 11, indicating that the CH_3 groups in the solvent seem to be preferentially surrounded by similar groups, at the short distances, even without the presence of the solute.

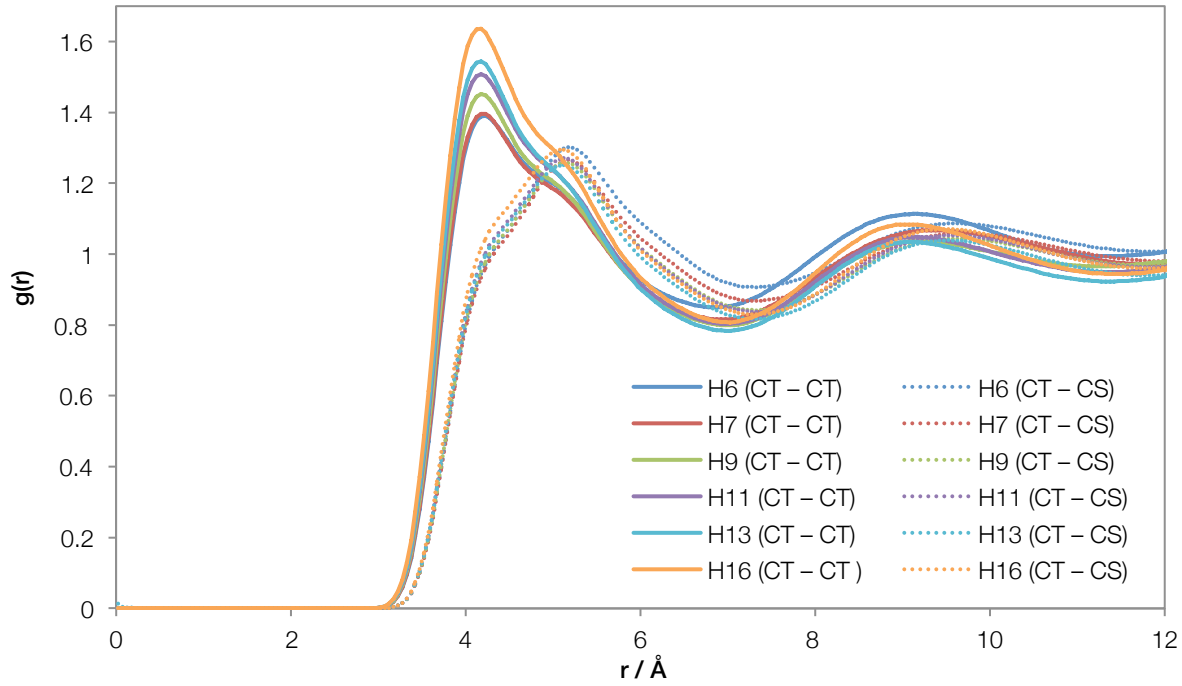

Figure 10 – Radial distribution functions between CH_3 (solid lines) and CH_2 (dotted lines) and each solvent's CH_3 carbon atom, at 298.15 K.

A similar analysis was employed to obtain the ratio between the local and bulk fractions of CH_3.

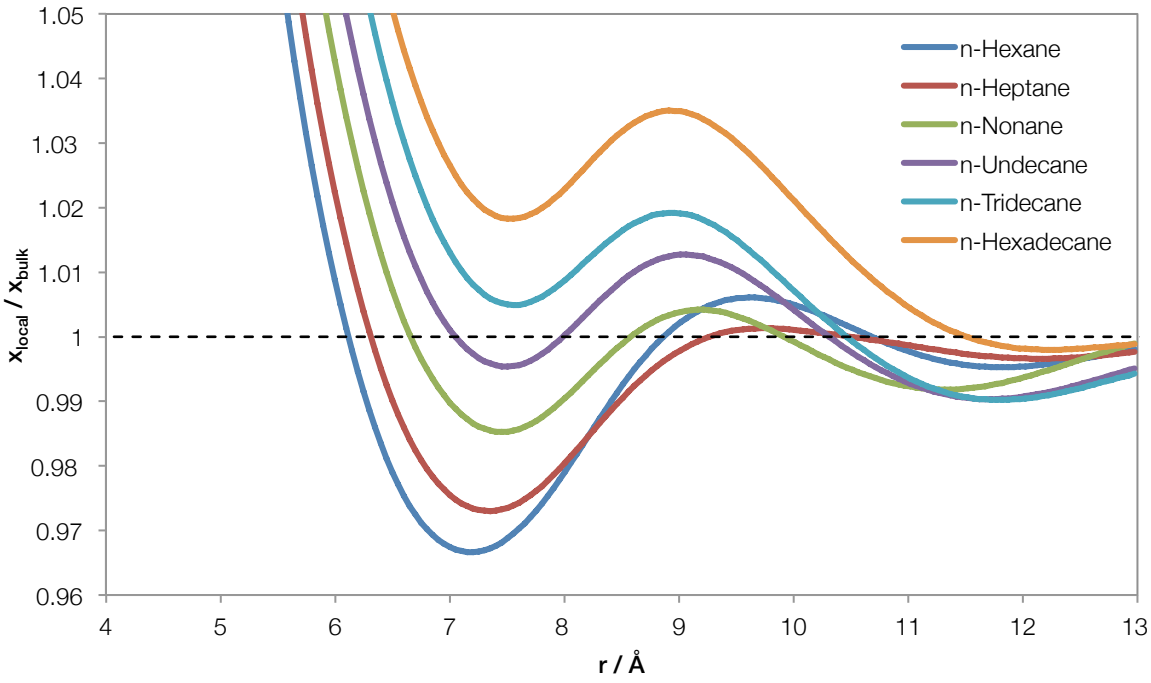

Figure 11 – Ratio between the local fraction of CH_3 groups around the solvent's CH_3 carbon atom and the bulk fraction, as function of distance in linear alkanes (at a temperature of 298.15 K).

In the figure, we can see that the enrichment is present in the solvent's default liquid structuring. The presence of these rich "pockets" of CH_3 groups influence the position of the solute molecule in the liquid, since they have, not just, a slight increase in their interaction with the solvent molecules, but less impact on the solvent's structure. It can be argued that creating space for the solute between chains of solvent molecules is energetically less favourable than opening space near the CH_3 groups. In a sense, the solute becomes an extension of the solvent's long chain molecules, fitting in with the dominant structure.

3.2 Enthalpy of Solvation

The quantity that is directly derived from the experimental solubility values is the enthalpy of solution, which corresponds to the released/absorbed energy when mixing the solute with the solvent. However, the solvation enthalpy is, particularly for this work, a property with greater interest. Thermodynamically, it is the energy required to transfer the solute from the perfect gas state to the liquid solvent, and so it differs from the enthalpy of solution by the solute's vaporization enthalpy. Table 3 collects both the enthalpy of solution, obtained experimentally, and the enthalpy of solvation obtained from the first, for some of the solvents in this work. [24]

Table 3 – Standard molar enthalpy of solution ($\Delta_{sol}H_m^0$) and solvation ($\Delta_{svt}H_m^0$) at 298.15 K for the experimentally studied alkane systems. [24]

Solvent	$\Delta_{sol}H_m^0$ / kJ.mol^{-1}	$\Delta_{svt}H_m^0$ / kJ.mol^{-1}
n-Hexane	32.7 ± 1.0	-11.3 ± 1,0
n-Heptane	33.0 ± 0.2	-11.0 ± 0.2
n-Undecane	30.6 ± 0.5	-13.4 ± 0.5
n-Hexadecane	29.9 ± 0.3	-14.1 ± 0.3

3.2.1 Interaction Energies

As can be seen in table 3, the enthalpy of solution of water in *n*-alkanes decreases with the chain length of the solvent, suggesting that the solute-solvent interaction is more favourable for the longer alkanes. To further explore this experimentally observed trend from the point of view of the simulations, we have decomposed the total intermolecular (dispersion + electrostatic) potential energy of the simulated systems in solvent-solvent, solute-solute and solute-solvent contributions. It should be noted that, since the simulations only have one solute molecule, the solute-solute contributions are essentially null (apart from very small values that stem from the Ewald summation of the electrostatic potential and from the analytic tail correction for the dispersion term). Figure 12 shows a plot of the total solute-solvent interaction energies with increasing chain length, for both the constant temperature of 298.15 K and the reduced temperature of T_r=0.538 (these values, and the respective dispersive and electrostatic parcels, can be found in Appendix C, Table 5 and Table 6).

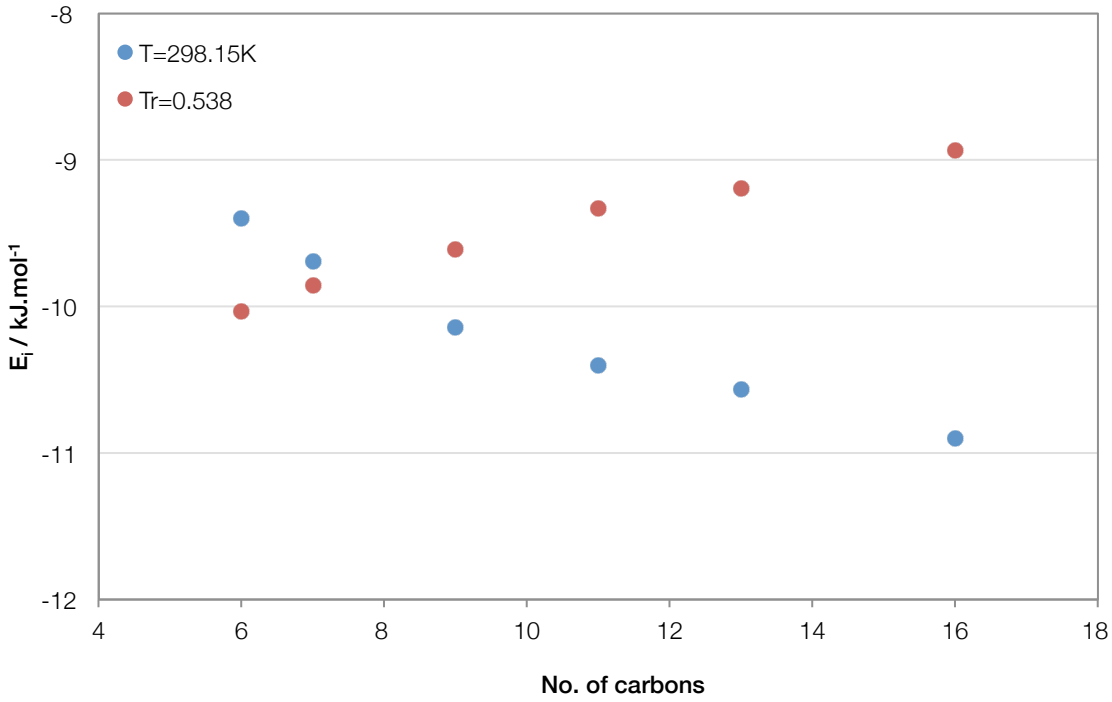

Figure 12 – Solute-solvent interaction energies at constant temperature T=298.15 K and at constant reduced temperature of 0.538, with varying chain length.

In Figure 12 we observe (blue dots) that, at constant temperature, the interaction energy between water and the *n*-alkane solvent increases with the alkane chain length, despite the fact that the concentration of the more interactive CH_3 group is decreasing. This suggests that the dominating effect in this case is the global density of the solvent, giving larger interaction energies between water and the longer (denser) *n*-alkanes.

On the other hand, at constant reduced temperature there is a decreasing interaction with increasing chain length. Given that these systems (red dots) are now at similar thermodynamic conditions, the order of solute-solvent interactions is now, most likely, determined by the relative concentration of the more interactive CH_3 groups. Furthermore, this density effect can be noticed in the differences between the RDFs for both temperatures, with the first peaks at constant reduced temperature overlapping each other (Error! Reference source not found.).

The solvation process can be conceptually decomposed in two steps, each contributing a given amount of energy to the final result. First, a cavity large enough to house the solute molecule must be created in the solvent – the formation of this cavity always represents an increase in the system's energy. Afterwards, the solute molecule is put inside the cavity, and the solute-solvent interactions represent the second (and negative) contribution. Considering the experimentally determined enthalpy of solvation on Table 3, the solute-solvent interaction enthalpy can be obtained by removing the energy required to form the cavity for the solute. This result can be compared directly to that obtained from the simulation results.

3.2.2 Enthalpy of Cavity Formation [32]

The main issue with calculating the energy associated with the formation of a cavity in a liquid is the thermodynamic description of the process that leads to the formation of a *reference cavity*, which has the same size and shape as a single molecule in a pure liquid. The formation of this reference cavity is not associated with any changes in molecular order in the liquid. The method used to calculate the cavity formation enthalpy for all systems is a semi-empirical method that takes these considerations into account.

Following the procedure for the method as described in reference [32], and considering a spherical solute (water) in a cylindrical solvent (*n*-alkane), the following equation was used to calculate the enthalpy of cavity formation (H_c) in each system.

$$H_c = \frac{S}{S_0} H_{c0} = \frac{V^{2/3}}{0.140V_0 + 11.3509} 5.365(1+\omega)RT_c \qquad \textbf{4}$$

where H_c and H_{c0} are, respectively, the solute's and reference enthalpy of cavity formation, $\frac{S}{S_0}$ is the ratio between the surface area of the solute's cavity and the reference cavity, V and V_0 are, respectively, the molar volume for the solute and solvent, ω is the Pitzer acentric factor of the solvent, R is the ideal gas constant and T_c the critical temperature of the solvent.

Given the nature of the method the approximations made might be too far from the physical reality. Even considering that the approximations are valid, the results depend, in this case, on the value used for the molar volume of the dissolved solute, which was approximated by the molar volume of pure water. In the particular case of this work, where the studied systems consist of the same solute dissolved in a homologous series of solvents, it is expected that the results obtained from this method are internally consistent and at least qualitatively valid.

The following Table 4 collects the required properties and parameters, the resulting enthalpy for the formation of each cavity and the solute-solvent interactions obtained from the experimental enthalpies of solvation. It can be seen that the energy required for the opening of a cavity increases with the size of the *n*-alkane.

Table 4 – Solute cavity enthalpy of formation (H_c), critical temperature (T_c), Pitzer acentric factor (ω) and molar volumes for each pure solvent (V_0), solute-solvent interaction for each system ($E_{i,exp}$), plus water's pure molar volume (V). H_c, V_0 and V relate to a temperature of 298 K. [32]

Solvent	ω	T_c / K	V_0 / $cm^3.mol^{-1}$	H_c / $kJ.mol^{-1}$	$E_{i,exp}$ / $kJ.mol^{-1}$
n-hexane	0.299	507.5	131.61	6.802	-18.10
n-heptane	0.349	540.3	147.47	6.995	-18.00
n-undecane	0.535	638.8	211.23	7.362	-20.76
n-hexadecane	0.742	722	294.07	7.356	-21.46
V / $cm^3.mol^{-1}$			18.07		

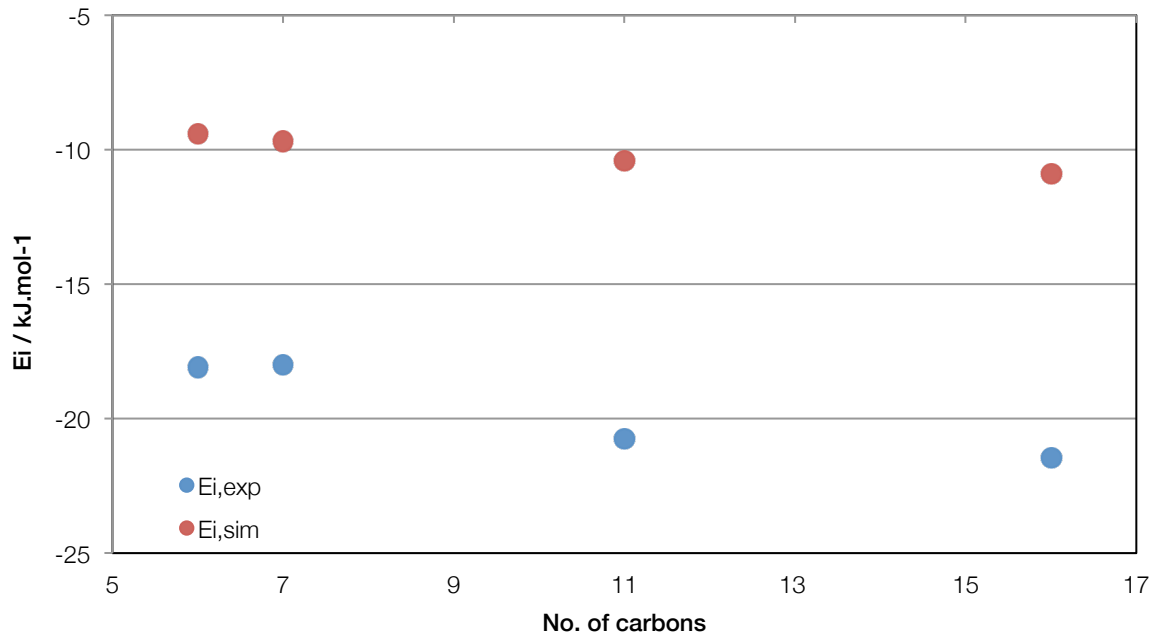

Figure 13 – Energy of solute-solvent interaction obtained from the simulations (red dots) and from the enthalpy of solvation (blue dots), at 298 K.

From the analysis of Figure 11, it can be seen that the interaction energies obtained from the simulations and from the experimental results follow the same trend, both becoming more negative with the increase in alkane chain length. The discrepancy in value between both sets may stem from two sources: the fact that the simulations use force fields which are not optimized to give interaction energies; and the values obtained for the enthalpy of cavity formation may be overestimated. A possible reason for the later is the solvent's intermolecular space, that may be bigger or smaller than the solute, which would eliminate or reduce the energy required to open a cavity (the first step in the solvation process).

Chapter 4 Conclusions and Further Developments

The results from the radial distribution functions and energy decomposition analysis seem to indicate that, like with xenon [26], the interactivity between water and each *n*-alkane solvent is seen to depend on the number and proportion of methyl/methylene groups, and also on the local structure of the solvent around water. Hence, water is not randomly distributed throughout the liquid *n*-alkanes, being more probably located near the terminal groups of the solvent molecules. This results in the enrichment in CH_3 groups around the water molecule when compared with the bulk proportion.

By taking a preliminary look at what happens around the methyl groups, with the radial distributions functions around the terminal carbon, it seems that there is already some spatial correlation between the CH_3 groups of the *n*-alkanes, which tend to cluster together. Still, further work should be done, by analysing the rdf for the pure systems (as the results above were obtained with the presence of the solute). Furthermore, it could be of interest to repeat some of the simulations in order to recheck the apparent anomalies in the order of the curves for the solvents in Figure 8 and Figure 11.

The solute-solvent interaction energies obtained from the simulations, at a constant temperature of 298.15 K, display an increase in interaction with growing chain length. These results qualitatively agree with the ones obtained experimentally on reference [24], and represent further evidence for the observed increase in interaction.

From a fundamental point of view it would be interesting, in future work, to see what happens with other solute sizes or shapes, and with the effect of changing the polarity of the solute. Solutes like carbon dioxide, methane or hydrogen sulphide would be particularly interesting from the perspective of the oil and gas applications.

References

[1] Istituto Della Enciclopedia Italiana Fondata, "Encyclopeadia of Hydrocarbons: Exploration, Production and Transport", in *Petroleum Fluid Properties*.: Marchesi Grafiche Editoriali S.p.A., 2006, vol. I, pp. 504-505.

[2] Istituto Della Enciclopedia Italiana Fondata, "Encyclopedia of Hydrocarbons: Exploration, Production and Transport", in *Origin, Migration and Accumulation of Petroleum*.: Marchesi Grafiche Editoriali S.p.A., 2006, vol. I, pp. 77-79.

[3] Istituto Della Enciclopedia Italiana Fondata, "Encyclopaedia of Hydrocarbons: Refining and Petrochemicals", in *Environmental Management in Refineries*.: Marchesi Grafiche Editoriali S.p.A., 2006, vol. II, pp. 398-401.

[4] Splette I, Liquid water hydrogen bond, Licensed under CC BY-SA 3.0 via Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Liquid_water_hydrogen_bond.png#/media/File:Liquid_water_hydrogen_bond.png.

[5] Riccardo Rovinetti, Forze di London, Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Forze_di_London.png#/media/File:Forze_di_London.png.

[6] Kota Saito, Masatoshi Kishimoto, Tanaka Ryo, and Ryo Ohmura, "Crystal growth of clathrate hydrate at the interface between hydrocarbon gas mixture and liquid water", *Crystal Growth and Design*, vol. 11, no. 1, pp. 295-301, 2011.

[7] Janamejaya Ladanyi, Branka M Chowdhary, "Hydrogen Bond Dynamics at the Water / Hydrocarbon Interface Hydrogen Bond Dynamics at the Water / Hydrocarbon Interface", pp. 4045-4053, Novembre 2008.

[8] Janamejaya Ladanyi, Branka M. Chowdhary, "Water/hydrocarbon interfaces: Effect of hydrocarbon branching on single-molecule relaxation", *Journal of Physical Chemistry B*, vol. 112, no. 19, pp. 6259-6273, 2008.

[9] Andrzej Maczynski, David G. Shaw, Marian Goral, and Wisniewska-Goclowska, "IUPAC-NIST Solubility Data Series. 81. Hydrocarbons with Water and SeawaterPart 4. C 6 H 14 Hydrocarbons with Water", *Journal of Physical and Chemical*, 2005.

[10] Silverstein T, "The Real Reason Why Oil and Water Don't Mix", *Journal of chemical education*, vol. 75, no. 1, pp. 116-118, 1998.

[11] M. Góral, "Cubic equation of state for calculation of phase equilibria in association systems", *Fluid Phase Equilib*, pp. 27-59, 1996.

[12] C. Tsonopoulos, "Tsonopoulos, C. Thermodynamic Analysis of the Mutual Solubilities of Normal Alkanes and Water", *Fluid Phase Equilibria*, pp. 21–33, 1999.

[13] W. A. Fouad, D. Ballal, K. R. Cox, and W. G. Chapman, "Examining the Consistency of Water Content Data in Alkanes Using the Perturbed-Chain Form of the Statistical Associating Fluid Theory Equation of State", *J. Chem. Eng. Data*, no. 59, pp. 1016–1023, 2014.

[14] W. Wagner and A. Pruβ, "The IAPWS Formulation for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use", *J. Phys. Chem.*, no. 31, p. 387, 1995, Ref. Data 2002.

[15] J. Gross and G. Sadowski, "Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules", *Ind. Eng. Chem. Res.*, no. 40, p. 1244–1260, 2001.

[16] W. Smith, T. R. Forester, and I. T. Todorov, *The DL_POLY Classic User Manual*. Daresbury Laboratory, UK, 2012.

[17] Berk Hess, David van der Spoel, and Erik Lindahl, *GROMACS User Manual*. Groningen , 2010.

[18] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids", *J. Am. Chem. Soc.*, no. 118, pp. 11225–11236, 1996.

[19] S. W. I. Siu, K. Pluhackova, and R. A. Böckmann, "Optimization of the OPLS-AA Force Field for Long Hydrocarbons", *J. Chem. Theory Comput.*, no. 8, pp. 1459–1470, 2012.

[20] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The Missing Term in Effective Pair Potentials", *J. Phys. Chem.*, no. 91, pp. 6269–6271, 1987.

[21] J. N. Canongia Lopes, M. F. Costa Gomes, and A. A. H. Pádua, "Nonpolar, Polar, and Associating Solutes in Ionic Liquids", *J. Phys. Chem. B*, no. 110, pp. 16816–16818, 2006.

[22] E. Johansson, K. Bolton, D. N. Theodorou, and P. Ahlström, "Monte Carlo Simulations of Equilibrium Solubilities and Structure of Water in N-Alkanes and Polyethylene", *J. Chem. Phys.*, no. 126, p. 224902, 2007.

[23] D. Ballal, P. Venkataraman, W. A. Fouad, and K. R. Cox, "Chapman, W. G. Isolating the Non-Polar Contributions to the Intermolecular Potential for Water-Alkane Interactions", *J. Chem. Phys.*, no.

141, p. 064905, 2014.

[24] Pedro Morgado, *Semifluorinated Alkanes – Structure – Properties Relations, PhD Thesis Pag 154-163*. Lisboa: IST Lisboa, 2011.

[25] Eds. P. J. Linstrom and W. G. Mallard. NIST Chemistry Webbook, NIST Standard Reference Database Number 69. [Online]. http://webbook.nist.gov/chemistry

[26] P. Morgado, R. P. Bonifácio, L. F. G. Martins, and E. J. M. Filipe, "Probing the Structure of Liquids with 129Xe NMR Spectroscopy: N-Alkanes, Cycloalkanes, and Branched Alkanes", *J. Phys. Chem. B*, no. 117, pp. 9014–9024, 2013.

[27] Luís F G Parreira, M Cristina B Ramalho, João P Prates Morgado, Pedro Filipe, Eduardo J M Martins, "Prediction of diffusion coefficients of chlorophenols in water by computer simulation", *Fluid Phase Equilibria*, vol. 396, pp. 9-19, 2015.

[28] D McCall, D Douglass, "Diffusion in Paraffin Hydrocarbons", *Journal of Physical Chemistry*, vol. 62, no. 9, pp. 1102-1107, 1958.

[29] Yoshimi Yamaguchi, Yoshio Ogura and Mazau Suzuki Makio Iwahashi, "Dynamical Structures of Normal Alkanes, Alcohols, and Fatty Acids in the Liquid State as Determined by Viscosity, Self-Diffusion Coefficient, Infrared Spectra, and 13C NMR Spin-Lattice Relaxation Time Measurements", *The Chemical Society of Japan*, vol. 63, no. 8, pp. 2154-2158, 1990.

[30] H Dullien, A L Ertl, "Self-Diffusion and Viscosity of Some Liquids as a Function of Temperature", *AIChE Journal*, vol. 19, no. 6, pp. 1215-1223, 1973.

[31] P S Lloyd, D Clark, C a Barker, G J Parker, G J McConville, P Baldock, C Pope, J M Tofts, "Test liquids for quantitative MRI measurements of self-diffusion coefficient in vivo.", *Magnetic Resonance in Medicine*, vol. 43, no. 3, pp. 368-374, 2000.

[32] Joaquim J. Moura Ramos and Raquel M. Gonçalves Madalena S. Dionísio, "The enthalpy and entropy of cavity formation in liquids and Corresponding States Principle", *Canadian Journal of Chemistry*, vol. 68, pp. 1937-1949, 1990.

Appendixes

Appendix A. Radial Distribution Functions

The following figures are rdf for hydrogen as the reference atom (for both temperatures – 298.15K and 0.538 reduced temperature), and also the ratios between local and bulk fraction of methyl groups for each respective rdf.

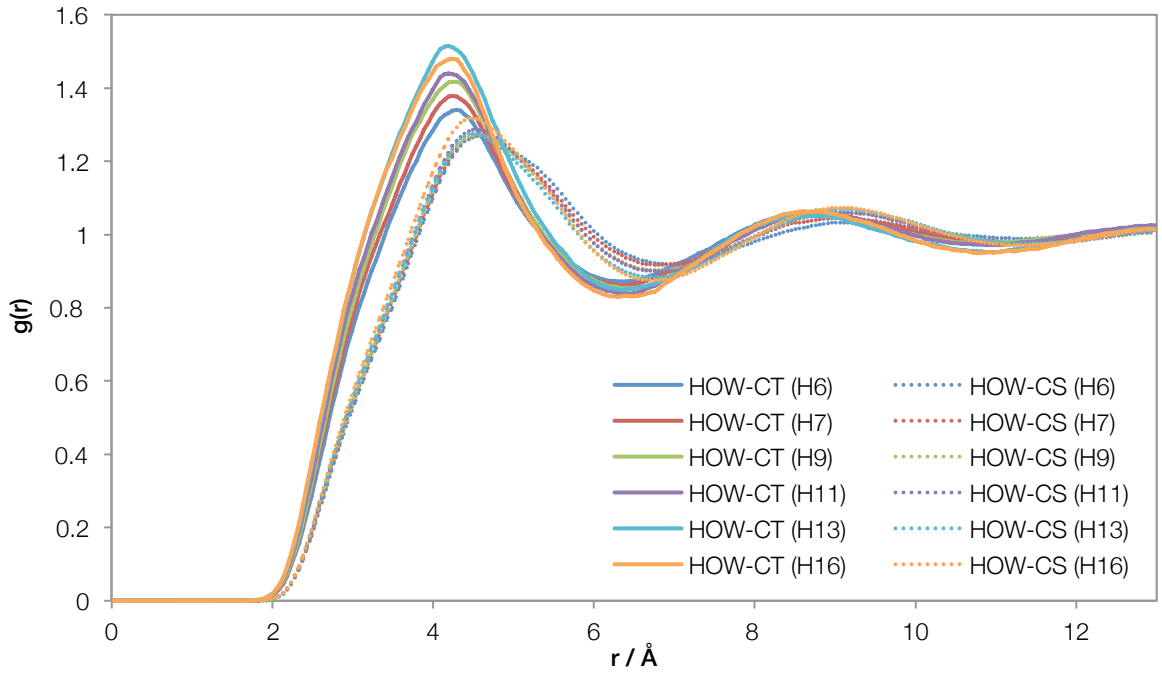

Figure 14 – Radial distribution functions between CH_3 (solid lines) and CH_2 (dotted lines) and water's hydrogen atom in linear alkanes, at 298.15 K.

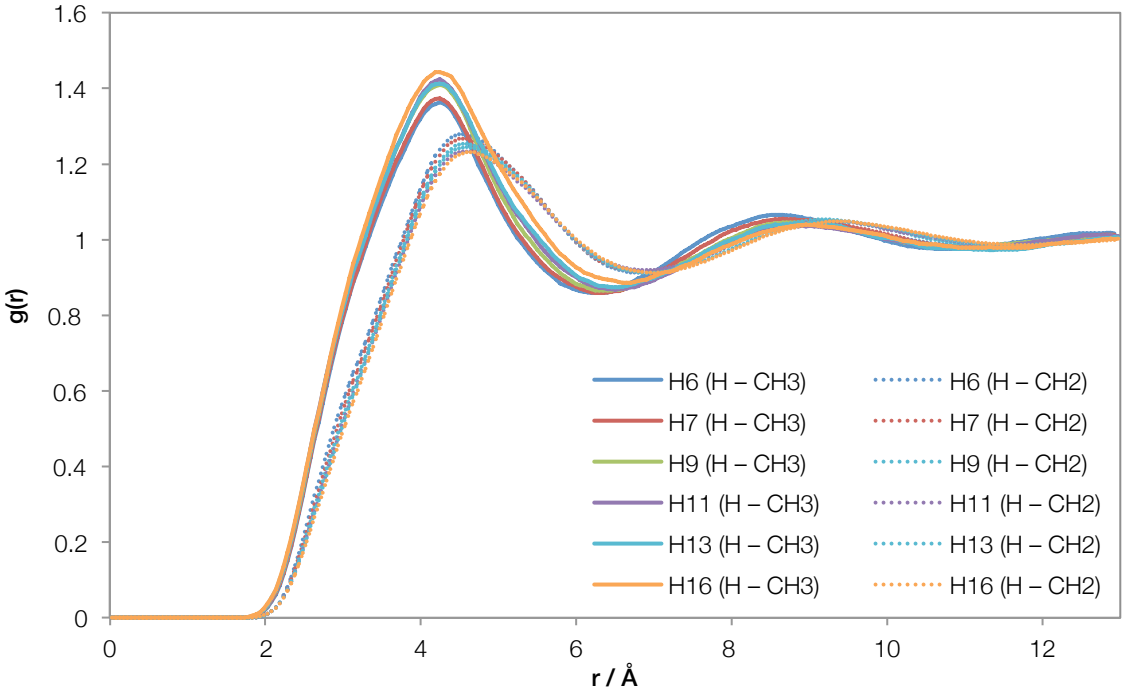

Figure 15 – Radial distribution functions between CH_3 (solid lines) and CH_2 (dashed lines) and water's hydrogen atom in linear alkanes, at 0.538 reduced temperature.

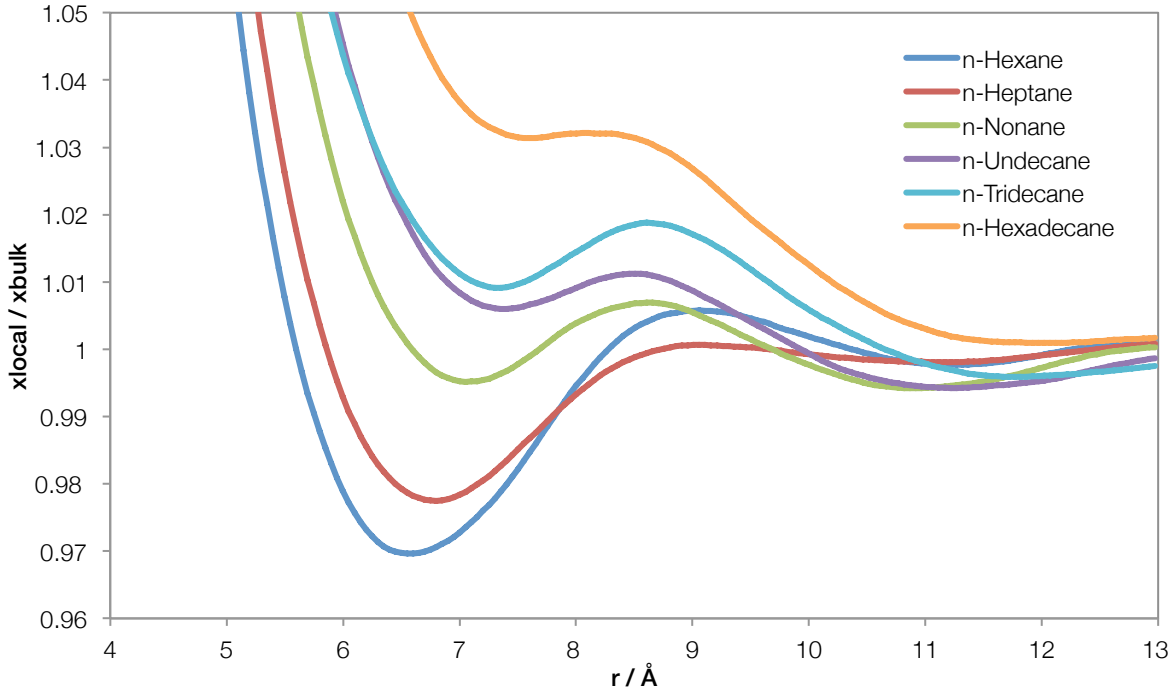

Figure 16 – Ratio between the local fraction of CH_3 groups around water's oxygen atom and the bulk fraction, as function of distance in linear alkanes (at a reduced temperature of 0.538).

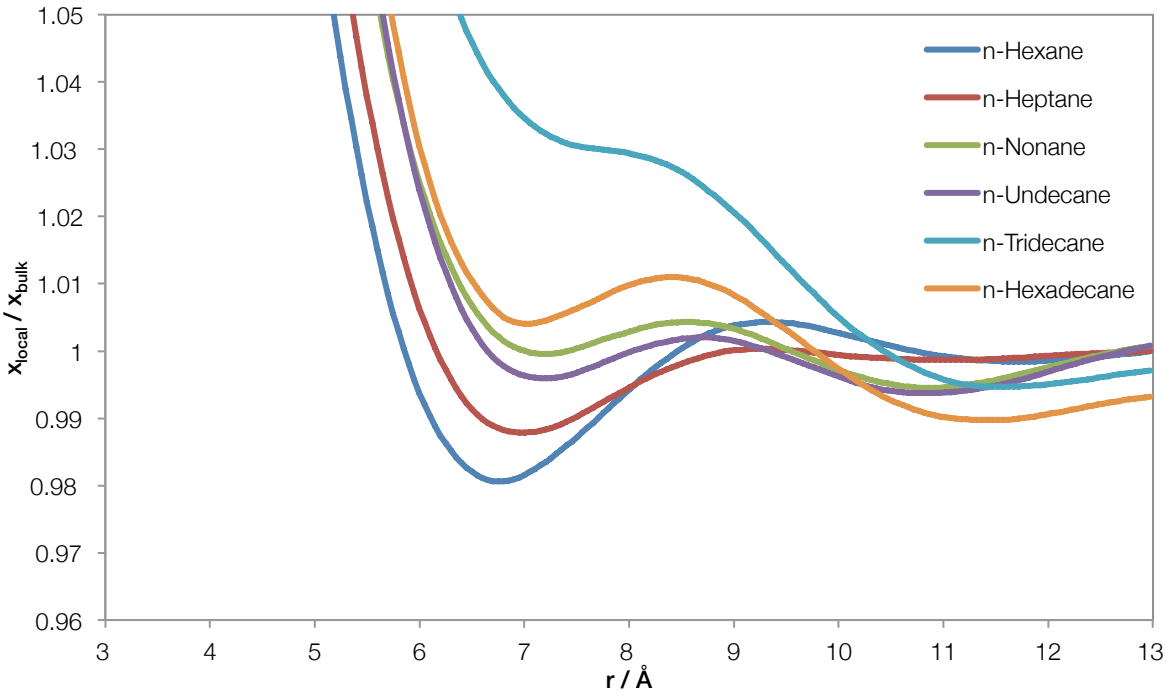

Figure 17 – Ratio between the local fraction of CH_3 groups around water's hydrogen atom and the bulk fraction, as function of distance in linear alkanes (at a temperature of 298.15 K).

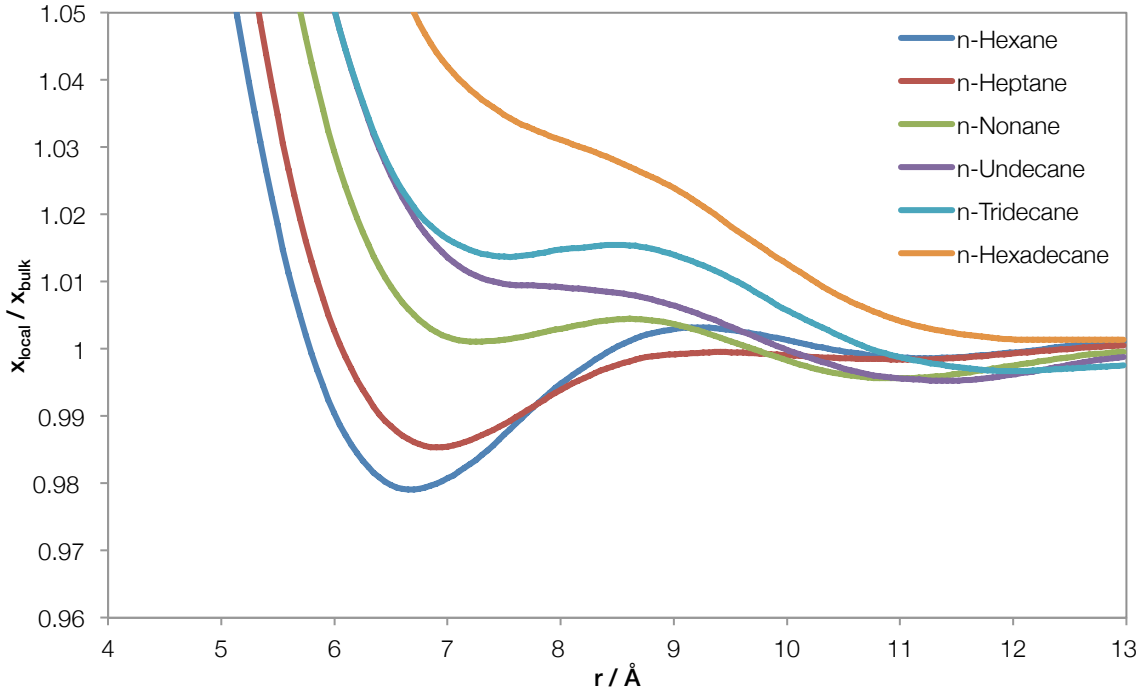

Figure 18 – Ratio between the local fraction of CH_3 groups around water's hydrogen atom and the bulk fraction, as function of distance in linear alkanes (at a reduced temperature of 0.538).

Appendix B. Energy Decomposition

Table 5 – Energy decomposition values for cross species interactions at 298.15 K: Coulomb energy (*Coul.*), Van der Waals energy (*VdW*) and their sum (*Sum*).

Solvent	*Coul / kJ.mol^{-1}*	*VdW / kJ.mol^{-1}*	*Sum / kJ.mol^{-1}*
n-hexane	-0.874	-8.526	-9.400
n-heptane	-0.906	-8.786	-9.692
n-nonane	-0.966	-9.178	-10.144
n-undecane	-1.004	-9.399	-10.403
n-tridecane	-1.019	-9.545	-10.564
n-hexadecane	-1.045	-9.853	-10.898

Table 6 – Energy decomposition values for cross species interactions at a reduced temperature of 0.538: Coulomb energy (*Coul.*), Van der Waals energy (*VdW*) and their sum (*Sum*).

Solvent	*Coul. / kJ.mol^{-1}*	*VdW / kJ.mol^{-1}*	*Sum / kJ.mol^{-1}*
n-hexane	-1.003	-9.032	-10.035
n-heptane	-0.927	-8.931	-9.858
n-nonane	-0.853	-8.759	-9.612
n-undecane	-0.793	-8.536	-9.329
n-tridecane	-0.745	-8.448	-9.193
n-hexadecane	-0.684	-8.251	-8.935