1-Octanol / Water Partition Coefficients of *n*-Alkanes from Molecular Simulations of Absolute Solvation Free Energies

Nuno M. Garrido[1,2], António J. Queimada[1], Miguel Jorge[1], Eugénia A. Macedo[1] and Ioannis G. Economou[2,*]

1. LSRE - Laboratory of Separation and Reaction Engineering, Departamento de Engenharia Química, Faculdade de Engenharia, Universidade do Porto, Rua do Dr. Roberto Frias, 4200 - 465 Porto, Portugal

2. Molecular Thermodynamics and Modeling of Materials Laboratory, Institute of Physical Chemistry, National Center for Scientific Research "Demokritos", GR-153 10, Aghia Paraskevi Attikis, Greece

[*]Author to whom all correspondence should be addressed at: economou@chem.demokritos.gr

For submission to *Journal of Chemical Theory and Computation*

April 2009

Abstract

The 1-octanol / water partition coefficient is an important thermodynamic variable usually employed to understand and quantify the partitioning of solutes between aqueous and organic phases. It finds widespread use in many empirical correlations to evaluate the environmental fate of pollutants as well as in the design of pharmaceuticals. The experimental evaluation of 1-octanol / water partition coefficients is an expensive and time consuming procedure, and thus theoretical estimation methods are needed, particularly when a physical sample of the solute may not yet be available, such as in pharmaceutical screening. 1-Octanol / water partition coefficients can be obtained from Gibbs free energies of solvation of the solute in both the aqueous and octanol phases. The accurate evaluation of free energy differences remains today a challenging problem in computational chemistry. In order to study the absolute solvation Gibbs free energies in 1-octanol, a solvent that can mimic many properties of important biological systems, free energy calculations for *n*-alkanes in the range C_1-C_8 were performed using molecular simulation techniques, following the thermodynamic integration approach.

In the first part of this paper, we test different force-fields, by evaluating their performance in reproducing pure 1-octanol properties. It is concluded that all-atom force fields can provide good accuracy, but at the cost of a higher computational time compared to united-atom force fields. Recent versions of united-atom force fields, such as Gromos and TraPPE, provide satisfactory results, and are thus useful alternatives to the more expensive all-atom models. In the second part of the paper, the Gibbs free energy of solvation in 1-octanol is calculated for several *n*-alkanes using three force-fields to describe the solutes, namely Gromos, TraPPE and OPLS-AA. Generally, the results obtained are in excellent agreement with the available experimental data, and of similar accuracy to commonly used QSPR models. Moreover, we have estimated the Gibbs free energy of hydration for the different compounds with the three force-fields, reaching average deviations from experimental data of less than 0.2 kcal/mol, for the case of the Gromos force-field. Finally, we systematically compare different strategies to obtain the 1-octanol / water partition coefficient from the simulations. It is shown that a fully predictive method combining the Gromos force-field in the aqueous phase and the OPLS-AA / TraPPE force-field for the organic phase can give excellent predictions for *n*-alkanes up to C_8 with absolute average deviation of 0.1 log P units to experimental data.

Keywords: Solvation Free Energy, 1-Octanol / Water Partition Coefficient, Molecular Simulation

1. Introduction

In several biochemical processes and for successful drug design strategies in the pharmaceutical industry, a correct understanding of the interactions of a given solute in both aqueous (hydrophilic) and biological (lipophilic) media is necessary[1-4]. Together with the Gibbs free energy of solute transfer, the corresponding partition coefficient between 1-octanol and water phases is probably the most important input parameter used in quantitative structure-property relationships (QSPR) to correlate and predict many solute properties[5]. Especially in the pharmaceutical industry, the prediction of drug partitioning, hydrophobicity and even the prediction of pharmacokinetic characteristics in biological systems can be quantified by expressions based on the 1-octanol / water partition coefficient (commonly known as P or even log P)[2,3,6]. Furthermore, log P is also used as a measure of activity of agrochemicals, degree of purity in metallurgy, and hydrophobicity in environmental problems. Partition coefficient data are also useful to estimate the solubility of a solute in a solvent[7,8].

The partition coefficient of a solute between 1-octanol and water was first introduced in 1964 by Hansch and Fujita[9] and since then, many different approaches have been developed in an attempt to estimate this property. In the beginning, mostly semi-empirical approaches based on the sum of fragment contributions or atom-derived group equivalents were proposed[1-3,10]. Nowadays, fragment additive schemes remain a standard method to estimate solvation free energies and partition coefficients[11], but the most common methods to estimate solvation properties are procedures based on QSPR that (cor)relate partition coefficients or solvation properties with other calculated or available molecular properties[12-14]. Although these methods are considerably fast and applicable to large databases of molecular structures, they require large multi-parameter tables having the disadvantage that whenever new molecules/compounds are under study, these need to be similar to the ones contained in the training set. This is evidenced by the lack of existing parameters to calculate log P for new chemical groups[15-17]. In short, we can conclude that QSPR methods are statistically rather than physically-based. Simulations based on linear response theory and molecular descriptors to derive empirical relationships for estimating log P values have been carried out by Duffy and Jorgensen[18]. Finally, approaches based on continuum models have also been investigated[15,16,19].

Besides the above mentioned estimation methods, the partition coefficient can also be obtained from experiments, by applying e.g. the shake-flask method[20-22] for generating the saturated liquid phases, followed by sampling and quantitative solute analysis (e.g. high-performance liquid chromatography[23]). Still, this can be a very expensive and time-consuming procedure and, thus it has limited practical use for product design, such as in pharmaceutical

screening. A different approach to all of the above is to use information on the free energy of solvation in water and in octanol to estimate the partition coefficient. Essex *et al.* [24,25] have shown that from Gibbs free energies of solvation in two different phases at temperature *T*, one can calculate the corresponding partition coefficient, according to the expression:

$$\log P^{octanol/water} = \frac{\Delta_{wat}G - \Delta_{solv}G}{2.303RT} \quad (1)$$

where $\Delta_{wat}G$ is the hydration free energy and $\Delta_{solv}G$ is the Gibbs free energy of solvation in 1-octanol. With recent developments in simulation methods and computing power, it is now possible to calculate solvation free energies of complex molecules, such as amino acid analogues or drugs, directly from molecular simulations[26-30]. Thus, we propose here an innovative approach to predict the 1-octanol / water partition coefficient without (or at least with a minimum) experimental information, based on the estimation of absolute solvation energies in water and 1-octanol, obtained from molecular simulation.

Regarding solvation, the majority of previously published studies focused on aqueous media (e.g. see a review paper by Tomasi and Persico[31]), but nowadays, computer simulation methods can also provide useful tools to model and understand molecular-level interactions of biological membranes, proteins and lipids. It is now possible to simulate the interactions of small solutes with complex biological membranes, by explicit simulation of the lipid-bilayers[32], an approach that has the disadvantage of being very computationally expensive[33]. Therefore, alternatives are sought to mimic the fundamental characteristics of biological systems using simpler molecules. Numerous solvents, such as oils[1], chloroform[5-9] or alkanes[34], have been tested to study and reproduce the hydrophobic properties of organic systems, but 1-octanol remains today the most important reference solvent for this kind of study. The amphiphilic nature of the 1-octanol molecule (a polar head group attached to a flexible non-polar tail) gives to this molecule similar characteristics to the main constituents of lipid bio-membranes. 1-octanol molecules can also mimic the complex behavior of the soil and thus play an important role in the prediction of solute partitioning in environmental fate and in toxicological processes[35]. Although 1-octanol cannot form stable structures such as bilayers or micelles[36], which are typical of lipid solutions, it can successfully mimic many of the properties of biologically relevant systems, and has been widely used for this purpose.

Several simulation studies related to 1-octanol systems have been reported in the literature. In the work of Debolt and Kollman[33], pure 1-octanol and water-saturated 1-octanol physical properties were also studied in detail. More recently, MacCallum and Tieleman[36] investigated 1-

octanol mixtures at different hydration levels, including the calculation of pure 1-octanol physical properties using various force-fields. In that study, formation of hydrogen-bonded chains in 1-octanol / water systems were observed, which interestingly become more spherical with increasing water concentration. On the contrary, in pure 1-octanol these clusters are long and thin. Chen and Siepmann[35] identified these microscopic structural differences in the aggregate/micelle formation between dry and water-saturated 1-octanol using configurational-bias Metropolis Monte Carlo simulations in the Gibbs ensemble. Regarding free energy calculations, most studies in 1-octanol have reported only relative free energy changes, i.e., the free energy associated with a mutation from one solute into another solute of similar structure, a different approach than the one followed here. Studied systems include: benzene to phenol[33]; ethylbenzene to phenol, pyridine to benzene, cyclopentane to tetrahydrofuran, methanol to methylamine, iso-propanol to iso-propane, acetamide to acetone and phenol to benzene[15]. Finally, Gibbs free energies of transfer of *n*-alkanes and primary alcohols between water and (dry or wet) 1-octanol were obtained by Chen *et al.*[37].

Our starting point in the present study is to evaluate / predict the Gibbs free energy of solvation of *n*-alkanes up to C_8 in 1-octanol. The availability of free energy data can be used to understand the behavior of complex systems and has the potential to revolutionize several scientific and technological fields[38,39], particularly in the pharmaceutical industry[40]. Solvation free energy can also be an important input parameter in order to predict solubility[17,41]. Several investigations regarding free energy calculations in aqueous systems have been reported in the literature[26-28,42-47], and it is now well established that accurate results can be obtained directly from molecular simulation methods. However, for non-aqueous solvents, and for 1-octanol in particular, there is a clear lack of data. We propose to fill this gap by presenting calculations for absolute solvation free energies of alkanes in 1-octanol. Initially, a comparison is made between several force-fields (FF), including all-atom (AA) and united-atom (UA) descriptions, in reproducing pure 1-octanol physical properties. Afterwards, we present a comparison of three popular FF, TraPPE, Gromos and OPLS-AA, to represent solute molecules by analyzing their performance in predicting the 1-octanol absolute Gibbs free energy of solvation for n-alkanes up to C_8. Finally, calculation of the hydration free energies and 1-octanol / water partition coefficient by molecular simulation is discussed.

The remainder of this paper is organized as follows: in Section 2 we describe the computational methods used for the Gibbs free energy calculation, particularly the thermodynamic integration, the molecular dynamics details and the force fields tested; in Section 3.1 results for the pure 1-octanol physical properties predicted using different FF are shown, while the capability of molecular simulation methods in predicting solvation free

energies and 1-octanol / water partition coefficients are discussed in Sections 3.2-3.4. The main conclusions of this work are summarized in Section 4.

2. Computational Methods

2.1 – Thermodynamic Integration

The solvation process consists of the transfer of a compound from a well-defined state (gas / vacuum) to another state (solution), and the solvation free energy may be defined as the free energy difference given by the total reversible work associated with changing the Hamiltonian of the system from the gas to the liquid state[48]. Solvation can be measured experimentally or calculated using an appropriate model and methodology. Experimental free energies are commonly estimated from solute concentration measurements in two-phase systems (vapor and liquid solution) in which, after reaching equilibrium, one evaluates the transfer of molecules between the two phases (see references [49] and [50] for equations and details). From the theoretical point of view, in the ideal gas approximation, the interaction of a solute with its environment in the gas state is effectively zero, and only the interactions of the solute with a particular solvent environment need to be considered. Free energy is a state function, and can thus be calculated by molecular simulation based on the construction of a thermodynamic cycle that may include non-physical transformations necessary to make the calculation feasible. Thus, the 1-octanol solvation free energy at temperature T and pressure P, $\Delta_{solv}G\left(P,T\right)$, can be calculated using the following thermodynamic cycle[51]:

$$\begin{array}{ccc} \text{Solute (1-octanol)} & \xrightarrow{\Delta_{oct}G} & \textit{Dummy} \text{ (1-octanol)} \\ \Delta_{solv}G \uparrow & & \downarrow \Delta_{dummy}G \\ \text{Solute (Vacuum)} & \xrightarrow{\Delta_{vac}G} & \textit{Dummy} \text{ (Vacuum)} \end{array}$$

where, $\Delta_{oct}G$ is the free energy associated with the mutation of the solute molecules into *dummy* molecules in 1-octanol media, $\Delta_{vac}G$ is the free energy associated with the same process in vacuum, and finally $\Delta_{dummy}G$ can be seen as the hypothetical solvation free energy of *dummy* species. *Dummy* molecules do not interact with their environment. In practice, these molecules have no electrostatic or van der Waals interactions, but their intramolecular bonded interactions are the same as in the solute molecules. As a consequence, $\Delta_{dummy}G$ is equal to zero and we can write the following equation for the thermodynamic cycle:

$$\Delta_{solv}G = \Delta_{vac}G - \Delta_{oct}G - \Delta_{dummy}G = \Delta_{vac}G - \Delta_{oct}G \quad (2)$$

The separate calculation in vacuum is necessary to compensate for changes in solute-solute intramolecular non-bonded interactions that take place when the intermolecular interactions are switched off[27]. For each case (solvent and vacuum), the associated free energy (expressed in terms of ΔG for the NPT ensemble) is estimated here using the thermodynamic integration method[48,52], whose algorithm is as follows: let us consider two generic well defined states, an initial reference state (state 0) and a final target state (state 1), with Hamiltonians $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. A coupling parameter[48], λ, can be added to the Hamiltonian, $\mathcal{H}(\boldsymbol{p},\boldsymbol{q};\lambda)$, where p is the linear momentum and q the atomic position, and used to describe the transition between the two states: $\mathcal{H}(\boldsymbol{p},\boldsymbol{q};0) \rightarrow \mathcal{H}(\boldsymbol{p},\boldsymbol{q};1)$. Considering several discrete and independent λ values between 0 and 1, equilibrium averages can be used to evaluate derivatives of the free energy with respect to λ. One then integrates the derivatives of the free energy along a continuous path connecting the initial and final states in order to obtain the energy difference between them[51]:

$$\Delta G = \int_0^1 \left\langle \frac{\partial \mathcal{H}(\boldsymbol{p},\boldsymbol{q},\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (3)$$

In practice, the solvation free energy can be estimated as follows: i) simulate the system in 1-octanol at different λ values; ii) simulate the system in vacuum at different λ values; iii) compute the solvation free energy from equation (4):

$$\Delta_{solv}G = \int_0^1 \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_\lambda^{vac} d\lambda - \int_0^1 \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_\lambda^{oct} d\lambda \quad (4)$$

Notice that because we are using thermodynamic integration, which involves equilibrium runs at independent λ values, the direction of the process is irrelevant and the results for $\Delta_{solv}G$ are free of hysteresis. This is an important advantage relatively to other methods (e.g. slow growth) where the results depend on the direction of the calculation[26,44]. As a final remark, one should notice that since we are studying non-polar molecules, *n*-alkanes, the Coulombic contribution to the free energy is negligible, and does not need to be accounted for separately in equation (4).

2.2 – Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were performed with the GROMACS[53] simulation package. The integration of Newton's equations of motion was carried out using the leap-frog dynamic algorithm[54] with a time step of 2 fs. Langevin (stochastic) dynamics[55] were used to control the temperature, with a frictional constant of 1 ps^{-1} and the reference temperature of 298 K. This approach eliminates several problems that may arise from the use of conventional thermostats in free energy calculations[43]. For constant pressure simulations, the Berendsen barostat[56] with a time constant of 0.5 ps and an isothermal compressibility of 4.5×10^{-5} bar^{-1} was used to enforce pressure coupling, where the box size was scaled at every time step. The reference pressure was always set to 1 bar. Each simulation box was cubic, with periodic boundary conditions in all directions, and contained 200 1-octanol molecules. Simulations of systems with different number of molecules revealed this to be the optimum system size: larger systems yielded statistically similar results but at a higher computational cost, while smaller systems exhibited finite-size effects.

The initial configuration for the pure 1-octanol simulations was generated by randomly placing 200 molecules in a large cubic box. We then run an energy minimization, followed by a constant volume equilibration of 100 ps and finally a 5 ns long NPT production stage. Two minimization procedures were employed: first, minimization was performed using the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm of Nocedal[57], for 5000 steps, followed by a steepest descent minimization for 500 steps. Analysis of several observables ensured that the simulations were properly equilibrated during the NPT run. Average properties were computed by discarding the time steps pertaining to the equilibration period.

To calculate solvation free energies it is necessary to carry out several independent simulations of each solute (from methane to *n*-octane) in each solvent (1-octanol and water), for different values of the coupling parameter as described in section 2.1. The starting configuration for each of these simulations was obtained by immersing each solute molecule into an equilibrated box of 200 1-octanol solvent molecules or 500 water solvent molecules (the equilibrated 1-octanol box was obtained from the pure-liquid simulations, described above, while the water box was obtained from our previous work on hydration free energies[27]). In these simulations an energy minimization was initially performed using the same protocol as for the pure liquid simulations, followed by a constant volume equilibration of 100 ps, a constant pressure equilibration of 1 ns (enough to fully equilibrate the box volume), and finally an NVT production run of 5 ns. This procedure was repeated for each of the following 16 λ values:

$$\lambda \in \{0.0, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00\}$$

where $\lambda = 0$ refers to a fully interacting solute and $\lambda = 1$ to a non-interacting solute. We have used such a large number of intermediate λ states because in thermodynamic integration, the accuracy of the $\Delta_{solv}G$ estimates depends mostly on the smoothness of the $\partial\mathcal{H}/\partial\lambda$ vs. λ curve, where a smooth profile is necessary in order to minimize numerical integration errors. In the present work the reported statistical uncertainties were obtained from block averaging[54] and integrals were computed via the trapezoidal rule[58]. Finally, it should be noted that in the transformation process between states with different λ values, the λ-dependence of the Lennard-Jones (LJ) potential was interpolated between the neighboring states via soft-core interactions. The soft-core expression of Beuler et al.[59] eliminates singularities in the calculation as the LJ interactions are turned off[60]. As suggested in the literature,[26,43] the soft-core parameter used was 0.5, which is the optimized value when the power for λ in the soft-core function is 1, and the soft-core σ value used was 0.3 nm.

2.3 – Force Fields

MD simulations for pure 1-octanol were performed using six different force fields. The force fields examined included Gromos (versions 43A2[61], 53A5[29] and 53A6[29]), OPLS-UA[62,63], OPLS-AA[64] and TraPPE[65-67]. We have decided to test three different versions of the Gromos force field since they were parameterized for different purposes, all relevant to this work. Version 43A2 was parameterized in order to reproduce only pure solvent properties. More recently, the Gromos parameter set 53A5 was optimized to reproduce thermodynamic properties of pure liquids and the solvation Gibbs free energy of amino acid analogs in cyclohexane, while parameter set 53A6 was optimized to reproduce free energies in water[29]. The TraPPE force field was also chosen because it was optimized to provide accurate descriptions of pure liquids and liquid-vapor phase equilibria[65-67]. It should be noted that, contrary to the original version of TraPPE where all bonds were fixed, bond stretching was modeled in our studies by a harmonic potential with force constants taken from CHARMM[68], except for bonds involving hydrogen atoms that were constrained using LINCS[69]. Finally, we have tested the popular OPLS force fields, which are designed to be transferrable to a wide range of organic molecules in the liquid phase. We have compared united-atom (UA) against all-atom (AA) force fields because the former are expected to be computationally much cheaper.

In this work, the Modified Extended Simplified Point Charge (MSPC/E)[70] model was used for the simulation of water. MSPC/E is an accurate force field for pure water and water – hydrocarbon thermodynamic properties and it was chosen over other popular force fields for water. This force-field also includes a polarization correction expected to improve the hydration predictions[28].

Several of the above force fields (in particular OPLS-AA,[64] TraPPE[65-67] and Gromos 53A6[29]) were also used to model the alkane molecules, solvated in either 1-octanol or water. Different combinations of solute-solvent force fields were tested in order to assess the influence of this choice on the free energy and partition coefficient predictions. *Dummy* molecules were considered to be identical to *real* solute molecules in terms of mass, while their LJ interaction parameters were set to zero. In all cases, electrostatic interactions were calculated using the Reaction Field[71] method with $\varepsilon_{rf,oct}$ = 10.3 (the dielectric constant for pure 1-octanol[72]) or $\varepsilon_{rf,wat} = 80$ (the dielectric constant for pure water[72]). Tests performed with the more computationally demanding particle mesh Ewald method yielded similar results. The cut-off radii used were 1 nm for the electrostatic interactions, 1 nm for the short-range neighbor list and 0.8–0.9 nm switched cut-off for the LJ interactions. Long range corrections for energy and pressure were used[26]. Detailed van der Waals parameters, point charges, bond stretching, bond angle bending and torsional force constants are provided in Supporting Information for all compounds and force fields. Coordinate and topology files were built manually or with the help of the Molden[73] and PRODRG[74] software.

3. Results and Discussion

3.1 - Pure 1-octanol physical properties

The accuracy of different force fields for the prediction of pure 1-octanol properties was initially evaluated. The calculated 1-octanol densities over a wide temperature range from NPT MD and the heat of vaporization at 298 K are shown in Table 1. Densities were directly obtained from the GROMACS suite using the g_energy[54] tool, while heats of vaporization were estimated by taking the difference of enthalpy in the vapor and liquid phases:

$$\Delta_{vap}H = E_g - E_L + RT \quad (5)$$

where, E_g is the total energy in the gas phase and E_L is the total energy per mole in the liquid phase.

From the results obtained, we can observe that Gromos generally overestimates the 1-octanol densities, which is a known deficiency[44] of this force field, and also significantly

underestimates the vaporization enthalpy. As expected, version 43A2 of Gromos performed better that the other two versions, because it was optimized to reproduce pure-liquid properties. The TraPPE FF provides good accuracy for the density over this wide temperature range (with a slight underestimation), but also significantly underestimates the enthalpy of vaporization. Conversely, OPLS-UA overestimates the density at all temperatures, but does an excellent job at predicting the enthalpy of vaporization. OPLS-AA is the most accurate of all force fields examined here, yielding good predictions of both density and vaporization enthalpy, but at the cost of an increased computational time. In fact, computational production times, included in Table 1, show that this AA FF is about 7 times more expensive compared to the UA approaches. One should also notice that for higher temperature OPLS-AA accuracy decreases. Generally speaking, it is preferable to use an AA model for pure 1-octanol, provided one can afford the additional computational cost. For simulations where this is an important issue, such as in the highly demanding free energy calculations performed in this work, it is reasonable to use a UA approximation. In this case, TraPPE is perhaps the better option, since it performs well for pure-liquid properties and is also able to accurately describe vapor-liquid equilibrium[65].

3.2 - Free Energies of Solvation in 1-octanol

The Gibbs free energy of solvation of *n*-alkanes in 1-octanol at 298 K was calculated from MD as described above. Simulations were performed using three different force fields for the representation of both 1-octanol and *n*-alkane molecules, namely Gromos 53A6[29], TraPPE[65-67,75-80] and OPLS-AA.[64] The 53A6 version of Gromos was preferred over the other two versions as it was parameterized to reproduce solvation properties in a polar solvent. Preliminary calculations using the OPLS-AA force field to model both alkanes and octanol showed that the computational time required for the accurate estimation of $\Delta_{solv}G$ was very high. As shown in section 3.1, this is due to the high cost associated to an AA description of 1-octanol. Consequently, 1-octanol molecules were modeled with the TraPPE force field instead, and these calculations are referred to as OPLS-AA / TraPPE in the remainder of this paper. We have also tested a combination of OPLS-AA for the solutes with OPLS-UA for the solvent for consistency. Unfortunately, differences to experimental data in a preliminary test with propane were as high as 1 kcal/mol, and this combination of FF was not pursued further. It should be noted that deficiencies of the OPLS-UA force field in reproducing hydration free energies and hydrocarbon solubilities in water were also reported by MacCallum and Tieleman[36].

Thermodynamic integration was performed using the three force fields for the solutes both in vacuum and in solvent media. Representative results for the integrand of equation (4) in the

octanol phase are shown in Figure 1 based on the Gromos 53A6 force field, while the complete data set for all force fields is given in Supporting Information. Furthermore, MD calculations of $\Delta_{vac}G$, $\Delta_{oct}G$ and $\Delta_{solv}G$ from the different force fields are shown in Table 2, with experimental data reported for comparison, while the different data sets for $\Delta_{solv}G$ are shown in Figure 2. It should be noted that the experimental values in Table 2 and Figure 2 represent solvation free energies of *n*-alkanes in water-saturated 1-octanol solutions, since there are no data available for anhydrous 1-octanol, which is used in the simulations. However, the difference between the free energy of solvation determined in pure and water-saturated 1-octanol is typically small, on the order of 0.2 – 0.4 kcal/mol[16] (and references[81-83]). Moreover, for the case of propane, *n*-butane and *n*-pentane there are no available experimental data. To allow for a better comparison of our simulations with experimental results, experimental values presented in Table 2 marked with ** were estimated from:

$$\Delta_{solv}G^{octanol} = \Delta_{solv}G^{water} - \log P^{octanol/water} \times 2.303 \times RT \quad (6)$$

where $\Delta_{solv}G^{water}$ are experimental data from Michielan *et al.*[84] and $\log P^{octanol/water}$ are the 1-octanol / water partition coefficients suggested by Sangster[3].

In general, the calculated $\Delta_{solv}G$ decrease with increasing chain length, which is consistent with the experimental data. Calculations based on OPLS-AA / TraPPE force fields provide the best agreement with experimental data, while Gromos predicts lower $\Delta_{solv}G$ values and TraPPE higher $\Delta_{solv}G$ than experiments. The average deviation between experimental data and simulations is 0.1 kcal/mol for OPLS-AA / TraPPE, 0.8 kcal/mol for Gromos and 0.4 kcal/mol for TraPPE. In the pharmaceutical industry, accuracies of 0.5 – 1.0 kcal/mol are required for predicting affinities in drug binding[42]. In this respect, the polarizable continuum model MST, originally developed by Miertus et al.[85], was recently re-parameterized[16] for reproducing solvation free energies in 1-octanol and differences of 0.4 – 0.6 kcal/mol were observed for *n*-alkanes from C_6-C_8. Even more, this (re)parameterization required the knowledge of the solvation experimental data, which for complex molecules is a clear disadvantage. Indeed, the methodology used in this work can provide molecular level details and insights that cannot be obtained using continuous models, since solvent molecules are modeled explicitly. The AAD observed in this work for the organic phase are considerably smaller than the typical AAD published in literature for aqueous systems (see Section 3.3).

In short, the accuracy of the OPLS-AA / TraPPE combination of force fields for solute/solvent to describe the Gibbs energy of solvation in 1-octanol is clearly better in

comparison with other published studies. These calculations also verify that an AA description of the solute molecules clearly improves the accuracy in the prediction of solvation energies.

3.3 - Free Energies of Hydration of *n*-Alkanes

Contrary to the case of 1-octanol, there are many experimental data and simulation studies available in the literature concerning $\Delta_{hyd}G$ of *n*-alkanes. In Table 3, a compilation of such data is presented (last two columns). In our simulations, the same molecular models as above were used for *n*-alkanes. Simulation results for $\Delta_{vac}G$, $\Delta_{wat}G$ and $\Delta_{hyd}G$ from the various force fields are presented in Table 3. A graphical comparison of simulation results with experimental data for $\Delta_{hyd}G$ is shown in Figure 3. We can observe that while in 1-octanol, solvation free energies are negative and decrease with the chain length so that the solubility in octanol increases, the opposite is found in water and the solubility decreases with the chain length. These facts are supported both by experiments and simulation.

For the hydration calculations, the deviation between experimental data and our MD results is larger than in the case of 1-octanol, although in the same accuracy range of previously published studies for these systems.[26,28,42-44] Typical average absolute deviations for hydration Gibbs energy calculations available in the literature range from 0.8-1.5 kcal/mol, as can be found in the study of Shirts *et al.*[26] for 15 amino acid side chain analogs: 1.2 kcal/mol for AMBER, 1.1 kcal/mol for CHARMM and 0.8 kcal/mol for OPLS-AA. As another example, for the hydration of alkanes (up to C_5) average deviations of 0.5 kcal/mol[44-47] were reported.

Gromos provides the better agreement to experimental data, with an average absolute deviation lower than 0.3 kcal/mol, while OPLS-AA / TraPPE predictions deviate by an average of 1.2 kcal/mol and TraPPE by an average of 0.9 kcal/mol from experimental data. This good performance of the Gromos force-field is to be expected *a priori* since this force field was parameterized to reproduce free energies of hydration. Interestingly, the use of an AA description of the solute in hydration free energy calculations seems to be less important than the optimization of the interaction parameters. This is in marked contrast to the case of solvation free energies in 1-octanol, as described above. Thus, it appears that it is important to take hydration free energies into consideration during the parameterization of a force field, if accurate predictions of this property are desired. Previous simulation studies have also revealed the importance of the force field used for water in the description of the hydration free energy [42,28].

3.4 - 1-octanol / water partition coefficients

The 1-octanol / water partition coefficient at 298 K for the various *n*-alkanes can be readily estimated from eq. (1) using the Gibbs free energies of solvation calculated from our MD simulations. In Table 4, simulation predictions are shown for the different force fields employed together with literature experimental data for comparison.

The overall average absolute deviation (AAD) between experimental data and simulation results for log P is equal to 0.4 (in log P units) for Gromos, 0.4 for TraPPE and 0.9 for OPLS-AA / TraPPE. Interestingly, the TraPPE FF provides accurate log P predictions, while the corresponding solvation energies are not so accurately estimated, and this can be attributed to cancellation of errors between the two phases – the high overestimation of the hydration free energy (Figure 3) is partially compensated by an overestimation of the octanol solvation free energy (Figure 2). A similar effect occurs in the Gromos predictions, but from the opposite direction – underestimation of both water and octanol free energies. On the other hand, the OPLS-AA / TraPPE FF combination is much more accurate in the organic phase than in the aqueous phase, leading to larger deviations in log P.

However, if one calculates log P using the most accurate simulation predictions for $\Delta_{hyd}G$ (from Gromos) and the most accurate simulation predictions for $\Delta_{solv}G$ (from OPLS-AA / TraPPE), then an AAD of 0.14 is obtained. Clearly, this approach provides a very accurate prediction, within the experimental uncertainty. Comparing accuracies of different methods can be merely qualitative since the method performance is highly dependent on the validation set used, which may vary on size, complexity or the overlap of information used in the training set/model correlation. Even so, similar calculations using a continuous model resulted in an AAD of 0.75 log P units[16] verifying that our predictions should be considered very satisfactory. Another published work[86] reports deviations of 0.6 log P units using a continuum method based on a continuous electrostatic model using atomic point charges combined with a non-electrostatic term function of surface tension for a set of 2116 molecules.

A final remark should be made regarding the accuracy of the available experimental data. As previously explained, log P and Gibbs free energy of solvation data are estimated following different experimental methodologies. At the same time, equation (1) provides a means to check the consistency between different data. A compilation of different data results in deviations up to 0.8 log P units with AAD of 0.24 log P units.

4. Conclusions

In order to predict the partition coefficient of a solute between 1-octanol and water MD absolute free energy calculations were performed in 1-octanol and water systems for different *n*-alkanes up to *n*-octane, using thermodynamic integration. The absolute free energies of solvation were estimated by fully decoupling the solute from the solvent, which must be distinguished from previous studies where relative free energies were calculated from mutations between two solutes. The method we used here is more flexible and not limited to mutations between similar structures. However, this complete decoupling requires large changes in the Hamiltonian and potentially higher errors are introduced in the calculations as more intermediate states are required. It is also worthwhile to notice that contrary to many other methodologies presented in the literature, we do not need the knowledge of the solvation experimental data in advance, which is a clear advantage.

Our method is capable to predict solvation free energies of non-polar solutes such as n-alkanes in 1-octanol with good accuracy. A comparison between different force fields permitted to conclude that the OPLS-AA FF for the solute in combination with the TraPPE FF for 1-octanol produces the most accurate results, with differences to experimental data of 0.1 kcal/mol, which is approximately the precision of the experimental methods. The results are much improved by using an AA model for the *n*-alkanes, relative to UA models, with very little increase in computational cost. Arguably, the predictions could be further improved by adopting an AA description of the 1-octanol solvent as well, since this yielded a better representation of pure-liquid properties. However, the associated high computational cost currently precludes this approach.

Moreover, we reproduced experimental hydration free energies of the same *n*-alkanes with average deviations of 0.3 kcal/mol, using the Gromos FF. For hydration free energies, a correct parameterization of the interaction potentials seems to be more important than using an AA description of the solute. For this reason, Gromos, which included hydration free energies in its parameterization, performed better than OPLS-AA.

Combining the simulated values of solvation free energy of the *n*-alkanes in water and in 1-octanol, we were able to predict the corresponding partition coefficients with an accuracy that is within the experimental uncertainty. All force field combinations that were tested here performed well, in some cases due to cancellation of errors in both solvation free energies. The most accurate log P predictions are afforded by the combination of the Gromos FF in the water phase with the OPLS-AA / TraPPE FF in the organic phase, reaching absolute deviations to

experimental data of 0.1 log P units which can be comparable to the widely used QSPR statistical methods.

5. Acknowledgments

The authors are grateful for the support provided by *Fundação para a Ciência e a Tecnologia* (FCT) Portugal, through projects FEDER/POCI/2010 and REEQ/1164/EQU/2005. NMG acknowledges his FCT Ph.D. scholarship SFRH/BD/47822/2007 and financial support for his visit to NCSR "Demokritos". Financial support from the Greek Secretariat of Research and Technology was provided for NMG stay in NCSR "Demokritos". AJQ acknowledges financial support from POCI/N010/2006 and MJ from Ciência 2008.

6. Supporting Information Description

Detailed bonded and non-bonded potential parameters for all the compounds and for the different force-fields under study as well as plots of the derivatives of the Hamiltonian with respect to the coupling parameter for all the case studies are available as supporting information.

7. References

(1) Leo, A.; Hansch, C.; Elkins, D. *Chem. Rev.* **1971**, *71*, 525-616.

(2) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Hydrophobic, Electronic and Steric Constants*; American Chemical Society: Washington DC, 1995.

(3) Sangster, J. *Octanol-Water Partitioning Coefficients: Fundamentals and Physical Chemistry*; John Willey & Sons: Chichester, U.K., 1997.

(4) Perlovich, G. L.; Kurkov, S. V.; Kinchin, A. N.; Bauer-Brandl, A. *AAPS Pharmsci* **2004**, *6*, 1-9.

(5) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Fundamentals and applications in chemistry and biology*; American Chemical Society: Washington DC, 1995.

(6) Betageri, G. V.; Rogers, J. A. *Pharm. Res.* **1989**, *6*, 399-403.

(7) Yalkowsky, S. H. *Solubility and Solubilization in Aqueous Media*; Oxford University: Oxford, 1999.

(8) Pinho, S. P.; Macedo, E. A. In *Developments and Applications in Solubility*; Letcher, T. M., Ed.; Royal Society of Chemistry: Cambridge, 2003, p 309-326.

(9) Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616-1626.

(10) Leo, A. J. *Chem. Rev.* **1993**, *93*, 1281-1306.

(11) Viswanadhan, V. N.; Ghose, A. K.; Singh, U. C.; Wendoloski, J. J. 1999, p 405-412.

(12) Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley - Interscience: New York, 2000.

(13) Bodor, N.; Buchwald, P. *J. Phys. Chem. B* **1997**, *101*, 3404-3412.

(14) Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Marcus, Y.; Taft, R. W. *J. Phys. Chem.* **1988**, *92*, 5244-5255.

(15) Best, S. A.; Merz, K. M.; Reynolds, C. H. *J. Phys. Chem. B* **1999**, *103*, 714-726.

(16) Curutchet, C.; Orozco, M.; Luque, F. J. *J. Comp. Chem.* **2001**, *22*, 1180-1193.

(17) Westergren, J.; Lindfors, L.; Hoglund, T.; Luder, K.; Nordholm, S.; Kjellander, R. *J. Phys. Chem. B* **2007**, *111*, 1872-1882.

(18) Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 2878-2888.

(19) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187-4225.

(20) Bergstrom, C. A. S.; Norinder, U.; Luthman, K.; Artursson, P. *Pharm. Res.* **2002**, *19*, 182-188.

(21) Glomme, A.; Marz, J.; Dressman, J. B. *J. Pharm. Sci.* **2005**, *94*, 1-16.

(22) Loftsson, T.; Hreinsdottir, D. *Aaps Pharmscitech* **2006**, *7*, E1-E4.

(23) Chen, X. Q.; Venkatesh, S. *Pharm. Res.* **2004**, *21*, 1758-1761.

(24) Essex, J. W.; Reynolds, C. A.; Richards, W. G. *Chem. Commun.* **1989**, 1152-1154.

(25) Essex, J. W.; Reynolds, C. A.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 3634-3639.

(26) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740-5761.

(27) Garrido, N. M.; Jorge, M.; Queimada, A. J.; Economou, I. G.; Macedo, E. A. *Submitted for Publication* **2009**.

(28) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616-17626.

(29) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comp. Chem.* **2004**, *25*, 1656-1676.

(30) Shivakumar, D.; Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 919-930.

(31) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027-2094.

(32) Tieleman, D. P.; Marrink, S. J.; Berendsen, H. J. C. *Biochimica Et Biophysica Acta-Reviews on Biomembranes* **1997**, *1331*, 235-270.

(33) Debolt, S. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5316-5340.

(34) Shih, P.; Pedersen, L. G.; Gibbs, P. R.; Wolfenden, R. *J. Molec. Biol.* **1998**, *280*, 421-430.

(35) Chen, B.; Siepmann, J. I. *J. Phys. Chem. B* **2006**, *110*, 3555-3563.

(36) MacCallum, J. L.; Tieleman, D. P. *J. Am. Chem. Soc.* **2002**, *124*, 15085-15093.

(37) Chen, B.; Siepmann, J. I. *J. Am. Chem. Soc.* **2000**, *122*, 6464-6467.

(38) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395-2417.

(39) Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461-469.

(40) Jorgensen, W. L. *Science* **2004**, *303*, 1813-1818.

(41) Palmer, D. S.; Llinas, A.; Morao, I.; Day, G. M.; Goodman, J. M.; Glen, R. C.; Mitchell, J. B. O. *Mol. Pharm.* **2008**, *5*, 266-279.

(42) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(43) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007**, *111*, 2242-2254.

(44) Wescott, J. T.; Fisher, L. R.; Hanna, S. *J. Chem. Phys.* **2002**, *116*, 2361-2369.

(45) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. *J. Phys. Chem.* **1994**, *98*, 13077-13082.

(46) Ashbaugh, H. S.; Kaler, E. W.; Paulaitis, M. E. *Biophys. J.* **1998**, *75*, 755-768.

(47) Slusher, J. T. *J. Phys. Chem. B* **1999**, *103*, 6075-6079.

(48) Christophe Chipot; Pohorille, A. *Free Energy Calculations - Theory and Applications in Chemistry and Biology*; Springer: Berlin, 2007.

(49) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849-855.

(50) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016-2027.

(51) Leach, A. *Molecular Modeling: principles and applications*; Prentice-Hall, 2001.

(52) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300-313.

(53) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701-1718.

(54) Spoel, D. L., E; Hess, B; Buuren, A; Apol, E; Meuulenhof, P; Tieleman, D; Sijbers, A; Feenstra, K; Drunen, R; Berendsen, H; *Gromacs User Manual - version 3.3* The Netherlands, 2006.

(55) van Gunsteren, W.; Berendsen, H. *Mol. Sim.* **1988**, *1*, 173-185.

(56) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

(57) Liu, D. C.; Nocedal, J. *Mathematical Programming* **1989**, *45*, 503-528.

(58) Chapra, S.; Canale, R. *Numerical Methods for Engineers*; 5th ed.; McGraw-Hill, 2006.

(59) Beuler, T. M., R; van Schaik, RC; Gerber, PR; van Gunsteren, WF *Chem. Phys. Lett.* **1994**, *222*, 529-539.

(60) Pitera, J. W.; van Gunsteren, W. F. *Mol. Sim.* **2002**, *28*, 45-65.

(61) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: GROMOS96 Manual and User Guide;*; Vdf Hochschulverlag AG an der ETH Zürich: Zürich, 1996.

(62) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638-6646.

(63) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276-1284.

(64) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.

(65) Chen, B.; Potoff, J. J.; Siepmann, J. I. *J. Phys. Chem. B* **2001**, *105*, 3093-3104.

(66) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, *102*, 2569-2577.

(67) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1999**, *103*, 4508-4517.

(68) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187-217.

(69) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. *J. Comp. Chem.* **1997**, *18*, 1463-1472.

(70) Boulougouris, G. C.; Economou, I. G.; Theodorou, D. N. *J. Phys. Chem. B* **1998**, *102*, 1029-1035.

(71) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100-3107.

(72) Lide, D. R. *CRC handbook of chemistry and physics*; 85 ed.; CRC Press: Boca Raton, FL, 2005.

(73) Schaftenaar, G.; Noordik, J. H. *J. Comput. Aided Mol. Des.* **2000**, *14*, 123-134.

(74) Schuettelkopf, A. W.; Aalten, D. M. F. v. *Acta Cryst. D* **2004**, *60*, 1355--1363.

(75) Martin, M. G.; Siepmann, J. I. *Journal of Physical Chemistry B* **1998**, *102*, 2569-2577.

(76) Martin, M. G.; Siepmann, J. I. *Journal of Physical Chemistry B* **1999**, *103*, 4508-4517.

(77) Wick, C. D.; Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **2000**, *104*, 8008-8016.

(78) Chen, B.; Potoff, J. J.; Siepmann, J. I. *Journal of Physical Chemistry B* **2001**, *105*, 3093-3104.

(79) Stubbs, J. M.; Potoff, J. J.; Siepmann, J. I. *J. Phys. Chem. B* **2004**, *108*, 17596-17605.

(80) Wick, C. D.; Stubbs, J. M.; Rai, N.; Siepmann, J. I. *J. Phys. Chem. B* **2005**, *109*, 18974-18982.

(81) Bernazzani, L.; Cabani, S.; Conti, G.; Mollica, V. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 649-655.

(82) Berti, P.; Cabani, S.; Conti, G.; Mollica, V. *J. Chem. Soc., Faraday Trans. 1* **1986**, *82*, 2547-2556.

(83) Dallas, A. J.; Carr, P. W. *J. Chem. Soc., Perkin Trans. 2* **1992**, 2155-2161.

(84) Michielan, L.; Bacilieri, M.; Kaseda, C.; Moro, S. *Bioorg. Med. Chem.* **2008**, *16*, 5733-5742.

(85) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117-129.

(86) Bordner, A. J.; Cavasotto, C. N.; Abagyan, R. A. *J. Phys. Chem. B* **2002**, *106*, 11009-11015.

(87) Hales, J. L.; Ellender, J. H. *J. Chem. Therm.* **1976**, *8*, 1177-1184.

(88) Smith, B. D.; Srivastava, R. *Thermodynamic Data for Pure Compounds. Part B. Halogenated Hydrocarbons and Alcohols*; Elsevier: Amsterdam, 1986.

(89) The DIPPR information and data evaluation manager for the design institute for Physical properties version 4.1.1.

(90) Rytting, E.; Lentz, K. A.; Chen, X. Q.; Qian, F.; Venkatesh, S. *AAPS J.* **2005**, *7*, E78-E105.

Table 1: 1-octanol density and heat of vaporization at 1 bar from MD simulations and experimental measurements. Computational production times per node (Intel Xeon at 3.0 GHz) for each FF are also included.

	T (K)								Production times (hr/ns)
	280		340		400		298		
Force Field	ρ (kg/m^3)	Dev (%)	ρ (kg/m^3)	Dev (%)	ρ (kg/m^3)	Dev (%)	$\Delta_{vap}H$ (kJ/mol)	Dev (%)	
G43A2	864.4 ± 0.9	3.4	822.3 ± 0.3	3.9	779.7 ± 0.7	5.0	64.4	-10.5	1.09
G53A5	867.9 ± 0.8	3.8	827.0 ± 0.9	4.5	785.3 ± 0.6	5.7	59.5	-17.3	1.09
G53A6	868.0 ± 0.7	3.8	827.1 ± 1.4	4.5	785.3 ± 0.8	5.7	59.5	-17.3	1.09
OPLS-UA	859.5 ± 0.7	2.8	818.8 ± 0.6	3.5	773.5 ± 0.6	4.1	72.3	0.4	1.19
OPLS-AA	841.8 ± 0.9	0.7	781.2 ± 1.3	-1.3	719.5 ± 1.2	-3.1	70.7	-1.8	8.00
TraPPE	819.7 ± 0.9	-2.0	775.7 ± 0.5	-2.0	726.8 ± 0.8	-2.2	61.9	-14.0	1.15
Experimental	836.26[a]		791.39[a]		742.75[a]		71.98[b]		-

[a] data from refs.[87-89]

[b] data from ref.[72]

Table 2: Comparison of $\Delta_{vac}G$, $\Delta_{oct}G$ and $\Delta_{solv}G$ predictions for n-alkanes in 1-octanol using TraPPE, Gromos and OPLS-AA / TraPPE FF against available experimental data at 298 K[16].

Solute	TraPPE			Gromos			OPLS-AA / TraPPE			Exp.
	$\Delta_{vac}G$	$\Delta_{oct}G$	$\Delta_{solv}G$	$\Delta_{vac}G$	$\Delta_{oct}G$	$\Delta_{solv}G$	$\Delta_{vac}G$	$\Delta_{oct}G$	$\Delta_{solv}G$	
methane	0*	-0.5 ± 0.2	**0.5 ± 0.1**	0*	-0.4 ± 0.1	**0.4 ± 0.1**	0*	-0.2 ± 0.1	**0.2 ± 0.1**	0.5
ethane	0*	0.4 ± 0.2	**-0.4 ± 0.2**	0*	0.9 ± 0.2	**-0.9 ± 0.2**	-0.0 ± 0.1	0.5 ± 0.2	**-0.5 ± 0.2**	-0.6
propane	0*	1.0 ± 0.2	**-1.0 ± 0.2**	0*	1.9 ± 0.2	**-1.9 ± 0.2**	-0.6 ± 0.1	0.6 ± 0.2	**-1.2 ± 0.2**	-1.2**
n-butane	-2.5 ± 0.1	-1.0 ± 0.2	**-1.5 ± 0.3**	0.0 ± 0.1	2.9 ± 0.2	**-2.9 ± 0.2**	-1.3 ± 0.1	0.6 ± 0.2	**-1.9 ± 0.2**	-1.8**
n-pentane	-5.1 ± 0.1	-3.3 ± 0.2	**-1.8 ± 0.1**	-0.1 ± 0.1	3.3 ± 0.2	**-3.4 ± 0.2**	-2.1 ± 0.1	0.7± 0.2	**-2.8 ± 0.2**	-2.3**
n-hexane	-7.5 ± 0.1	-5.1 ± 0.2	**-2.4 ± 0.1**	-0.1 ± 0.1	4.4 ± 0.2	**-4.5 ± 0.2**	-2.9 ± 0.1	0.5 ± 0.2	**-3.4 ± 0.2**	-3.3
n-heptane	-10.0 ± 0.1	-7.1 ± 0.2	**-2.9 ± 0.1**	-0.1 ± 0.1	4.7 ± 0.2	**-4.8 ± 0.2**	-3.8 ± 0.1	0.2 ± 0.2	**-4.0 ± 0.2**	-4.1
n-octane	-12.5 ± 0.1	-9.0 ± 0.2	**-3.5 ± 0.1**	-0.1 ± 0.1	6.0 ± 0.2	**-6.1 ± 0.2**	-4.9 ± 0.2	-0.1 ± 0.3	**-4.7 ± 0.3**	-4.6

* In the definition of the potential model, non-bonded intramolecular interactions which are separated by less than three bonds are excluded. As a consequence, the values in these cells are strictly zero.

** Values estimated from equation (6)

Table 3: $\Delta_{vac}G$, $\Delta_{wat}G$, and $\Delta_{hyd}G$ predictions for *n*-alkanes in MSPC/E water using TraPPE, Gromos and OPLS-AA FF against available experimental data at 298 K[84]. For comparison, additional values representing the range of results obtained by molecular simulations taken from the literature are also included.

Solute	TraPPE			Gromos			OPLS-AA			Exp.	Simulation
	$\Delta_{vac}G$	$\Delta_{wat}G$	$\Delta_{hyd}G$	$\Delta_{vac}G$	$\Delta_{wat}G$	$\Delta_{hyd}G$	$\Delta_{vac}G$	$\Delta_{wat}G$	$\Delta_{hyd}G$		
methane	0*	-2.3 ± 0.1	**2.3±0.1**	0*	-2.0 ± 0.1	**2.0 ± 0.1**	0*	-2.4 ± 0.1	**2.4 ± 0.1**	1.98	2.0-2.6[26,28,42-47]
ethane	0*	-2.1± 0.1	**2.1±0.1**	0*	-1.8 ± 0.1	**1.8 ± 0.1**	-0.0 ± 0.1	-2.6 ± 0.1	**2.6 ± 0.1**	1.81	1.7-2.6[44-47]
propane	0*	-2.8 ± 0.1	**2.8±0.1**	0*	-1.9 ± 0.1	**1.9 ± 0.1**	-0.6 ± 0.1	-3.7 ± 0.1	**3.1 ± 0.1**	2.02	1.9-2.7[26,28,42-47]
n-butane	-2.5 ± 0.1	-5.6 ± 0.2	**3.1±0.2**	-0.0 ± 0.1	-1.7 ± 0.2	**1.7 ± 0.2**	-1.3 ± 0.1	-4.7 ± 0.2	**3.4 ± 0.2**	2.18	1.9-3.5[26,28,42-47]
n-pentane	-5.1 ± 0.1	-8.4 ± 0.2	**3.3±0.1**	-0.1 ± 0.1	-2.1 ± 0.2	**2.0 ± 0.2**	-2.1 ± 0.1	-5.6 ± 0.2	**3.5 ± 0.2**	2.36	2.7-3.7[44,46]
n-hexane	-7.5 ± 0.1	-11.2 ± 0.2	**3.7±0.1**	-0.1 ± 0.1	-2.3 ± 0.2	**2.2 ± 0.2**	-2.9 ± 0.1	-7.1 ± 0.2	**4.2 ± 0.2**	2.58	n.a.
n-heptane	-10.0 ± 0.1	-14.2 ± 0.2	**4.2±0.1**	-0.1 ± 0.1	-2.4 ± 0.2	**2.3 ± 0.2**	-3.8 ± 0.1	-7.9 ± 0.2	**4.2 ± 0.2**	2.65	n.a.
n-octane	-12.5 ± 0.1	-16.7 ± 0.2	**4.3±0.1**	-0.1 ± 0.1	-2.4 ± 0.2	**2.3 ± 0.2**	-4.9 ± 0.2	-9.7 ± 0.2	**4.8 ± 0.2**	2.93	n.a.

* In the definition of the potential model, non-bonded intramolecular interactions which are separated by less than three bonds are excluded. As a consequence, the values in these cells are strictly zero.

Table 4: Experimental data[2,3,90] and simulation predictions for the 1-octanol / water partition coefficient using different force field combinations. The absolute average deviations between experiment and simulation are also included.

	log P				
Solute	Gromos	TraPPE	OPLS-AA / TraPPE	Gromos + OPLS-AA/TraPPE	Exp.
methane	1.2	1.2	1.6	1.3	1.1
ethane	2.0	1.3	2.3	1.7	1.8
propane	2.8	2.8	3.2	2.3	2.4
n-butane	3.4	3.4	3.8	2.6	2.9
n-pentane	4.0	3.7	4.6	3.5	3.4
n-hexane	4.9	4.5	5.6	4.1	3.9
n-heptane	5.2	5.2	6.0	4.6	4.7
n-octane	6.2	5.7	7.0	5.1	5.2
AAD	0.4	0.4	0.9	0.1	-

Figure Captions

Figure 1: Derivative of the Hamiltonian with respect to λ for *n*-alkanes in 1-octanol using the Gromos force field.

Figure 2: $\Delta_{solv}G$ for *n*-alkanes in 1-octanol at 298 K as a function of carbon number: Experimental data and MD simulations.

Figure 3: $\Delta_{hyd}G$ for *n*-alkanes as a function of carbon number at 298 K: Experimental data and MD simulations.

Figure 4: Comparison of log P predictions using different force fields against experimental data.

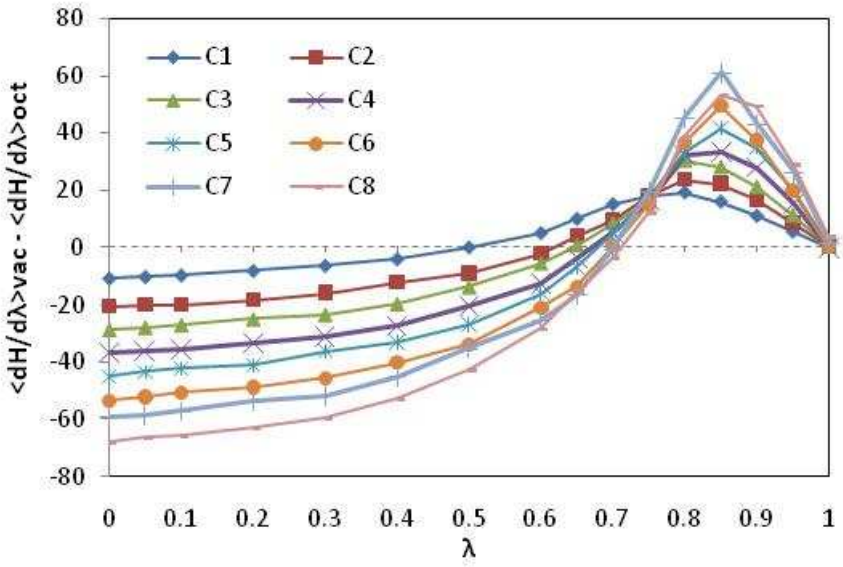

Figure 1

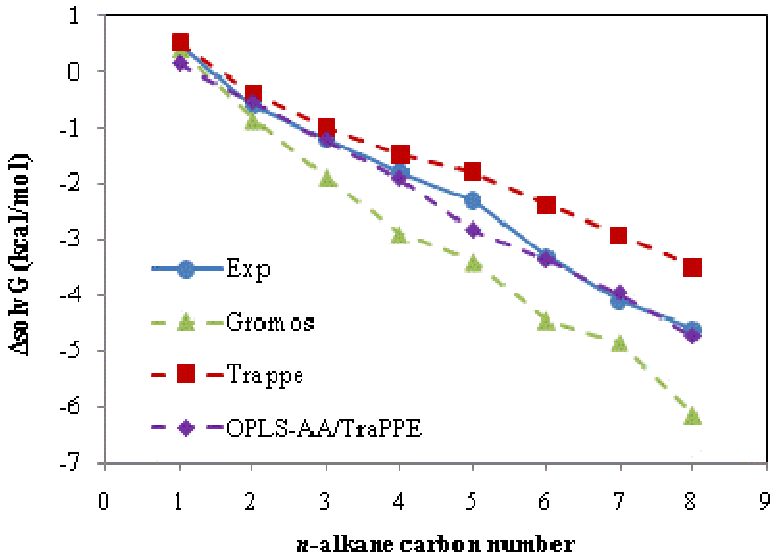

Figure 2

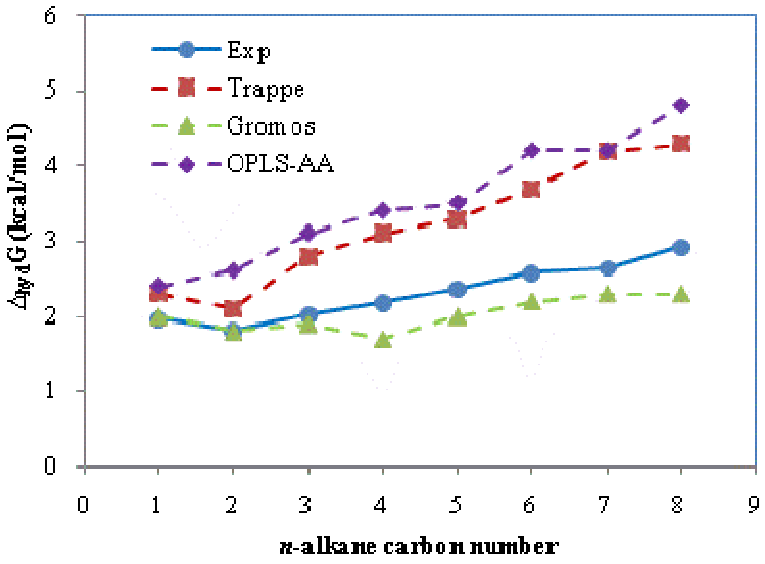

Figure 3

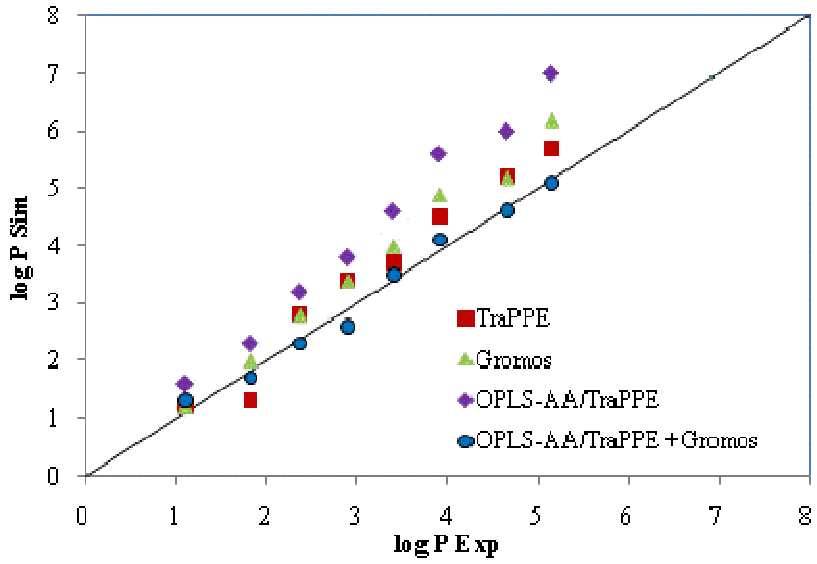

Figure 4