

Structure of Liquid Linear Alcohols

J. Lehtola,* M. Hakala, and K. Hämäläinen

Department of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland

Received: October 15, 2009; Revised Manuscript Received: April 9, 2010

The properties of linear alcohols in the liquid phase are studied by molecular dynamics simulations. We analyze the effects of the use of bond length constraints on the simulation density, self-diffusion constant, and hydrogen-bonding characteristics of the alcohol series. We find that the densities are well-reproduced in each of the cases but that the constraints have clear effects on the value of the diffusion constant and hydrogen-bonding properties, which is probably caused by the use of a gas-phase reference value in the OH bond length constraint. Although finite size effects are found to be present in the hydrogen bond networks, the networks are determined to be composed of chain-type structures that are well-converged. The results indicate that liquid alcohols consist of hydrogen-bonded chains of molecules. This finding can likely be tested experimentally with inelastic X-ray techniques at modern synchrotron radiation sources.

1. Introduction

Molecular dynamics simulations can be sped up by removing vibrational degrees of freedom by the use of constraints. One of the first general algorithms able to do this, called SHAKE, was developed by Ryckaert et al.¹ in 1977, and many others such as LINCS² have thereafter been introduced. The fast vibrational degrees of freedom are irrelevant for certain macroscopic properties when their corresponding excitation energies are much larger than the temperature (classically, this corresponds to high-frequency oscillations with a small amplitude). Tironi et al. noted³ that when typical vibrational energies are much larger than the temperature, the introduction of flexible bond lengths in the molecular dynamics simulation may lead to wrong results and that in the case of liquid water, flexible models do not add any value compared to bond-length-constrained models parametrized using the same procedures and experimental data. Bond angle constraints, on the other hand, combined with bond length constraints have been noted to perturb the dynamics of systems greatly.⁴ Generally speaking, even though constraints should only be used in combination with force fields parametrized using them, it is rather normal to use bond length constraints with the prevalent force fields of today that were originally parametrized for flexible molecules.

In chemical simulations, the fastest degrees of freedom are often related to hydrogen atoms, due to their small mass. Thus, one would expect effects arising from the use of constraints to be seen most clearly in hydrogen-rich compounds, such as hydrocarbons. In this paper, we have performed systematical benchmark calculations of a series of linear alcohols from methanol (CH_3OH) to tridecanol ($\text{CH}_3(\text{CH}_2)_{12}\text{OH}$) for 1000 molecules using the all-atom version of Jorgensen et al.'s Optimized Potentials for Liquid Simulations (OPLS).⁵ As OPLS has been parametrized without the use of constraints, we can thus examine the effects of using bond length constraints on the dynamics and structural properties of the system.

Stephenson et al. have performed⁶ molecular dynamics simulations for alcohols from pentanol to dodecanol using the united atom (UA) version of OPLS⁷ and analyzed the correlation

between the oxygen–oxygen coordination number in the alcohols and their water and ethanol capacities.

On the experimental side, Vahvaselkä et al. have studied⁸ the structure of linear alcohols of the series from methanol to butanol and of octanol using wide-angle X-ray scattering (WAXS) and have suggested that the molecules are arranged in linear chains of about 10 molecules, with their arrangements and lengths varying. A more recent simulation study on the same data performed by Chen and Siepmann⁹ suggests a rather more complicated picture of aggregation in octanol. Tomšič et al. have studied¹⁰ the series of linear alcohols from ethanol to hexanol using small-angle X-ray scattering (SAXS). In the same paper, they performed configurational bias Monte Carlo simulations with the TraPPE-UA force field¹¹ and found that the simulations reproduced well the SAXS scattering curves.

WAXS and SAXS are sensitive only to average structures. Local information around individual elements can be obtained by the use of other techniques, such as X-ray absorption spectroscopy (XAS), X-ray emission spectroscopy (XES), or inelastic X-ray scattering – Compton scattering (CS) or X-ray Raman scattering (XRS). All of the aforementioned methods are indirect structural probes as they depend on the electronic structure, which in turn depends on the atomic structure of the system. Thus, it is clear that the calculation of these spectra requires the use of structural models. The differences between the spectra of different model structures make it possible to interpret the experimental spectra and identify the structures that occur in the experimental system.

Guo et al. have used XES to study liquid water, methanol, and water–methanol solutions.¹² They have suggested that liquid methanol comprises hydrogen-bonded chains and rings with six and/or eight molecules of equal abundance. The paper by Wilson et al.¹³ employing XAS points to liquid methanol comprising long chains and rings of hydrogen-bonded (H-bonded) molecules. Very recently, Hakala et al. have studied¹⁴ isomer effects in the charge density distributions in liquid propanol and butanol by the use of CS.

The layout of this paper is as follows: first, we shall present the methods used in performing the simulations and the analysis (section 2). Second, we will discuss the criteria that are used in determining the existence of HBs (section 2.1). Next, in section

* To whom correspondence should be addressed. E-mail: jussi.lehtola@helsinki.fi.

2.2, we introduce the method of nodal analysis, which is useful in studying structuring in weakly bonded systems. We discuss the effects of the constraints on the intramolecular bond lengths in section 3.1 and continue by determining the time needed for equilibration of the simulations in section 3.2. We examine the densities in section 3.3 and self-diffusion constants in section 3.4. Finally, we study the number of hydrogen bonds (HBs) per molecule of the three models in section 3.5 and the hydrogen-bonded aggregates themselves in section 3.6. The conclusions of the study are given in section 4.

2. Methods

We performed molecular dynamics simulations for linear alcohols from methanol to tridecanol in the *NTP* ensemble at 298 K and 1 bar for 1000 molecules in a cubic box using GROMACS 4.0.5^{15–18} with the OPLS all-atom force field.⁵ In order to perform a systematical benchmark of the force field, we exceptionally have used the alkane parametrization for the atoms in the methyl group in methanol. This differs from the parametrization of methanol made by Jorgensen et al., in which a charge of 0.04e (*e* denoting the elementary charge) is placed on the methyl group hydrogens in methanol, whereas in the alkane parametrization, the charge is 0.06e. This is reflected in the charge of the carbon (0.085e in our model versus 0.145e of Jorgensen et al.). The other parameters of the model, such as the Lennard-Jones constants ϵ and σ , are the same for both models.

A leapfrog integrator was used with a 0.5 fs time step for all models. In each run, 2×10^7 MD time steps were taken, resulting in run lengths of 10 ns. The energies and particle coordinates were stored every 50 and 250 fs, respectively.

In order to study the effects of bond length constraints, three models were considered, (i) molecules without constraints (flexible), (ii) molecules with the intramolecular CH and OH bond lengths fixed to the OPLS equilibrium values (h-bonds), and (iii) molecules with all intramolecular (also CC and CO bonds) bond lengths fixed (all-bonds) to the OPLS equilibrium values. The constraints were implemented with the PLINCS algorithm.¹⁹

The neighbor list was updated every 10 fs with a neighbor list cutoff of 9 Å. Coulombic interactions were calculated using the particle mesh Ewald²⁰ method with a real-space cutoff of 9 Å and a 1.1 Å grid spacing with fourth-order spline interpolation. A twin-range method was used to calculate the Lennard-Jones interactions, with the Lennard-Jones cutoff being 14 Å. Standard long-range corrections to the virial and potential energy for the Lennard-Jones interaction were applied starting from the Lennard-Jones cutoff.

The temperature and pressure were kept constant using the velocity rescaling thermostat²¹ and the Berendsen barostat²² with the time constants being $\tau_T = 50$ fs and $\tau_p = 10$ ps, respectively. The pressure scaling was performed isotropically. Runs with the Berendsen thermostat²² were also performed for shorter simulation times, but the results were found to be statistically identical to those obtained with the velocity rescaling thermostat.

The molecular self-diffusion constants were calculated using a least-squares line fit to the mean-square displacement (MSD) according to the Einstein relation²³

$$\text{MSD}(t) = \langle |\mathbf{r}_{\text{CM}}(t_0 + t) - \mathbf{r}_{\text{CM}}(t_0)|^2 \rangle \quad (1)$$

$$D = \lim_{t \rightarrow \infty} \frac{1}{6t} \text{MSD}(t) \quad (2)$$

in which $\text{MSD}(t)$ is the MSD at time difference t , averaged over all the molecules in the system and all reference times t_0 , $r_{\text{CM}}(t)$ is the center-of-mass coordinate of the molecule, and D is the self-diffusion constant.

2.1. Hydrogen Bonding. In this paper, a H-bonded aggregate is defined as a group of molecules that are H-bonded together. A molecule without HBs forms by itself an aggregate. For further analysis, we define the diameter $d_{\mathcal{A}}$ of an aggregate \mathcal{A} as

$$d_{\mathcal{A}} = \max_{i,j \in \mathcal{A}} \left| \mathbf{r}_i - \mathbf{r}_j \right|_{\text{PBC}} \quad (3)$$

where i and j are atoms that are in the aggregate and PBC denotes that periodic boundary conditions are used in calculating the distances. This is a useful tool for analyzing finite size effects (FSEs) as according to the minimum image convention the components of the displacement vector $\Delta \mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ satisfy

$$-\frac{L}{2} < (\Delta \mathbf{r}_{ij})_\alpha \leq \frac{L}{2} \quad \alpha = x, y, z \quad (4)$$

in a cubic box with the edge length L . The index α in eq 4 refers to the x , y , or z component of the displacement vector $\Delta \mathbf{r}_{ij}$. If the displacement is only in one dimension, the corresponding maximal diameter is $d_1 = L/2$. In the case of two dimensions, the maximal diameter is $d_2 = L/2^{1/2}$, and in three, it is $d_m = 3^{1/2}L/2$, where m indicates the diameter is the maximal possible in the simulation system.

To examine H-bonding characteristics, we naturally need first a criterion for the existence of a HB. Unfortunately, no unambiguous one exists;^{24–26} thus, in the following, we examine three different (geometric) definitions for a HB.

The first HB criterion (criterion 1) is the classical criterion of Haughney et al.²⁷ (parametrized for liquid methanol), in which two oxygens are considered H-bonded if their distance r_{OO} is less than 3.5 Å, the acceptor–donor hydrogen distance is $r_{OH} < 2.6$ Å, and the hydrogen donor–acceptor (HDA) angle θ is less than 30°, the donor and acceptor being oxygen atoms and the hydrogen being chemically bonded to the donor oxygen.

The second HB criterion (criterion 2) is the one by Wernet et al.²⁴ (parametrized for liquid water), in which the existence of a bond is determined by a distance–angle criterion; the bond exists if the oxygen–oxygen distance satisfies the relation $r_{OO} \leq (3.3 - 0.00044\theta^2)$ Å, the HDA angle θ being given in degrees.

The third HB criterion (criterion 3) is the “STRICT” criterion of Chen and Siepmann⁹ (parametrized for liquid octanol–water systems), which is the same as the Haughney criterion, except with values of $r_{OO} < 3.3$ Å, $r_{OH} < 2.5$ Å, and $\cos \theta > 0.4$ (meaning $\theta \lesssim 66.4^\circ$).

Chen and Siepmann have devised²⁸ an unambiguous way of classifying the topologies of H-bonded aggregates into chains, rings, and branched structures. By itself, this classification does not reveal any detailed information about the scale of the structuring. In the case of Figure 1a, one would see that the structure is branched and consists of 16 molecules, but information about the relevant length scales (elementary chains of 2–4 molecules) would not be obtained. In our study, we found

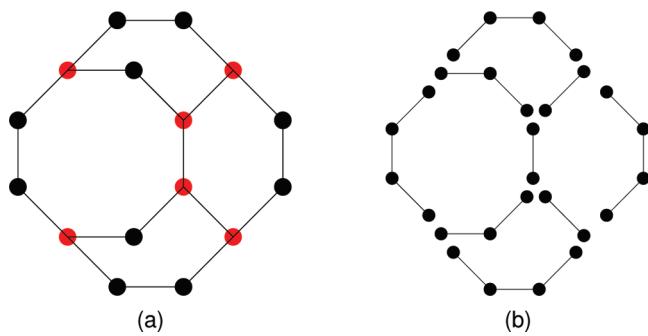


Figure 1. Illustration of the method of nodal analysis. (a) The original branched aggregate. (b) The nodal decomposition of the original aggregate into nodal aggregates.

aggregates consisting of more than half of the system; thus, a need for supporting information emerged.

2.2. Nodal Analysis. In order to study the relevant length scales associated with the H-bonded networks, we have devised an alternative scheme to decompose the aggregates into their constituent parts—nodal analysis. In this scheme, the aggregate is cut into pieces at the nodal molecules, that is, at every molecule with more than two H-bonds (see Figure 1). The constituent unit is then taken to be the partial chain thus extracted. The nodal molecule is counted in all of the partial chains connected to it. The partial chains are called nodal aggregates. The convergence of the size distribution of the nodal aggregates as a function of the system size can be easily checked and thus offers a good way to analyze the FSEs. Note that since nodal molecules are counted multiple times, one needs to be careful when comparing the sizes of the original and the nodal aggregates. In the following, the application of nodal analysis is called nodal decomposition.

The method of nodal analysis can, in principle, be applied to any system, but it is at its most useful in systems that can be considered weakly bonded. By weakly bonded, we refer to systems in which most particles have up to two bonds with its neighbors. The structures created by this type of bonding can be meaningfully cut into subchains, which may link together to form more complicated structures, such as rings and lassos. One is then able to extract the size distributions of these elementary chains. Although molecules with three H-bonds exist in liquid alcohols, the mean number of H-bonds per molecule is less than two. In this respect, H-bonded networks in alcohols can be considered weakly bonded.

In strongly bonded systems, however, the particles often form bonds with more than two neighbors, and thus, most particles will be classified as nodal ones. As a result, the nodal decomposition of the system will consist mainly of pairs of nodal particles; no information about the bonding on a larger scale is obtained. For example, water molecules in the condensed phase have a mean number of H-bonds per molecule that is greater than three; thus, the use of nodal analysis is not helpful.

When the method of nodal analysis is combined with the classification of structures suggested by Chen and Siepmann,²⁸ one sees straightaway that only branched aggregates need nodal decomposition as the linear and ring-type aggregates already form nodal aggregates. The nodal decomposition of the branched aggregates into nodal branches reveals the hitherto elusive relevant length scales of structuring in the system.

3. Results and Discussion

3.1. Bond Lengths.

As we are comparing models with and without bond length constraints, we devote the first section of

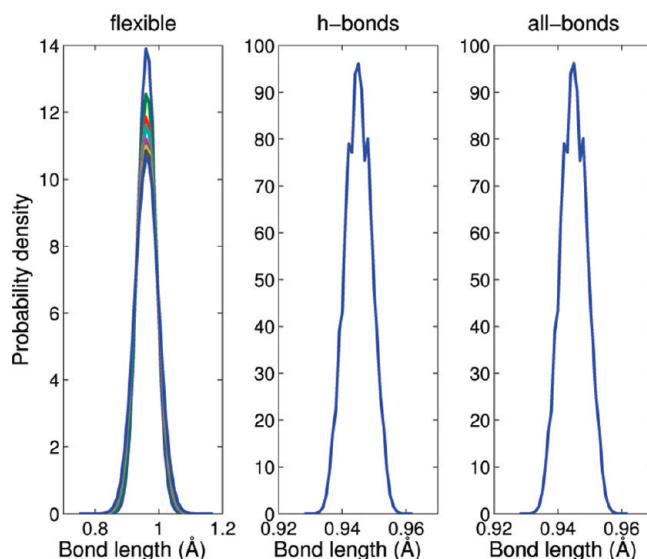


Figure 2. OH bond lengths for the different models. The highest peak in the left part of the figure is the OH bond length distribution in methanol, and the lowest one is that in octanol, the peak height behaving monotonically in the length of the alkyl chain.

the discussion to the study of bond length distributions. As OPLS uses harmonic bond-stretching potentials, the bond length distributions are Gaussian, even in the case of the constrained models.

For all models, the CC, CH, and CO bond length distributions are centered on the OPLS equilibrium value, with standard deviations of roughly 0.03 Å for nonconstrained bonds (all bonds in the flexible model and CC and CO bonds in the h-bonds model) and 0.004 Å for constrained bonds (all bonds in the all-bonds model and CH bonds in the h-bonds model).

A large difference can be seen, however, between the OH bond length distributions of the nonconstrained flexible model and those of the constrained h-bonds and all-bonds models. The distributions of the latter models are centered on the OPLS equilibrium value (determined from the gas phase) of 0.945 Å, with a standard deviation of roughly 0.004 Å, whereas the distribution of the flexible model has a mean bond length of 0.962 Å and a standard deviation of roughly 0.04 Å. Why this happens can be easily understood; in the OPLS model, the hydroxyl group is strongly polar. In the formation of a HB, the hydrogen (which has a positive partial charge) feels the attraction of the oxygen (which has a negative partial charge) of a neighboring molecule. This Coulombic interaction stretches out the OH bond from the gas-phase equilibrium value. The used constraint value for the OH bond length is thus incorrect in the case of liquid alcohols.

The OH bond lengths are shown in Figure 2. Slight differences can be seen between the OH bond lengths of the different alcohols in the flexible model; methanol has the narrowest OH bond length distribution (and the highest peak value, as shown in the figure), the spread increasing monotonically as a function of the length of the alkyl chain. In the constrained models, the bond length distributions are very similar across the whole alcohol series.

3.2. Equilibration. In the case of the alcohols up to octanol, thermalization with respect to the density and potential energy was found to occur within the first 1 ns of run time. The convergence of the HB networks was inspected by dividing the full trajectory into pieces of 10 ps. The analysis of H-bonded aggregates was performed on these pieces, giving the aggregate

TABLE 1: Densities in kg/m³ of the Simulated Alcohols^a

compound	flexible	h-bonds	all-bonds	experiment ²⁹
methanol	800 ± 3	785 ± 3	786 ± 3	791.4
ethanol	800 ± 3	785 ± 3	786 ± 3	789.3
propanol	811 ± 3	798 ± 2	801 ± 2	799.7
butanol	810 ± 2	799 ± 2	803 ± 2	809.8
pentanol	814 ± 2	805 ± 2	809 ± 2	814.4
hexanol	819 ± 2	811 ± 2	816 ± 2	813.6
heptanol	824 ± 2	817 ± 2	822 ± 2	821.9
octanol	830 ± 2	822 ± 2	829 ± 2	826.3

^a The numbers indicated with ± are standard deviations.

size (amount of molecules in the aggregate) and diameter (as given by eq 3) histograms for each type of aggregate (linear, ring, branched, and nodal branches). The used bin width in the diameter histograms was 0.2 Å.

For each histogram, an error measure was calculated with

$$e_{t_i} = \sum_{n=1}^N (h_{t_i}(n) - h_{t_e}(n))^2$$

where $h_{t_i}(n)$ is the number of entries in the n th histogram bin at time slice t_i , t_e the last time slice of the simulation used for comparison, and the sum over the bin index n runs over all of the bins. With this error measure, we could see that also the distributions of the HB networks converge within the first 1 ns of run time. Statistics were gathered from the remaining 9 ns.

The drift in the density over the thermalized part of the trajectory was found to be on the order of a few kg/m³ for all of the alcohols from methanol to octanol. The drift over the last 9 ns of the trajectory for nonanol was already several kg/m³, for decanol, it was 10–20 kg/m³, and for undecanol, dodecanol, and tridecanol, it was over 20 kg/m³. Although linear alcohols are liquids at the simulation temperature up to dodecanol (tridecanol with a melting point²⁹ of 32.5 °C is a solid), the evolution of the densities of the simulations shows that with the simulation parameters that have been used in this study, the OPLS model does not describe liquids starting from nonanol; the long time scale associated with the change of density accompanied by a slowly decreasing potential energy suggests that a phase change is going on, that is, that the system is solidifying. Thus, in the following, we will not discuss alcohols heavier than octanol.

As the performed simulations were computationally very demanding, we were unable to perform many simulations for each system studied. Instead, we verified the sampling of phase space by another, limited set of simulations (the verification set) from a different starting point. The verification simulations were performed for the flexible and all-bonds models of methanol, propanol, pentanol, and heptanol. The results of the second set of runs fit within the error limits of the results of the whole set of simulations that are reported below.

3.3. Density. The density is an easily measurable macroscopic quantity, which is trivial to calculate microscopically and is the first quantity that should be reproduced by a molecular dynamics simulation. The densities obtained from the simulations for the various models are given in Table 1 with the experimental data obtained from ref 29. The values are also shown graphically in Figure 3. The experimental densities are computationally reproduced to an accuracy of a couple of percent.

In practice, the density is a macroscopic quantity. This means that a huge phase space is sampled in experiment, and what

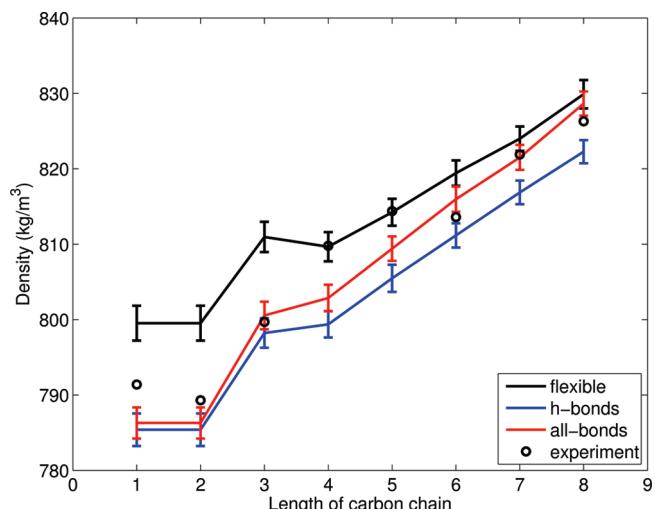


Figure 3. Densities of the simulated liquid alcohols; the error bars represent standard deviations. The experimental data have been taken from ref 29.

one actually measures is the average of the density over the phase space. Although the thermal fluctuations of the density can be significant, we are mostly interested in how its average, which represents the measurable quantity, behaves.

From Figure 3, it is clearly seen that the h-bonds model always produces the lowest density, the all-bonds model the second lowest density, and the flexible model the largest density. It can be speculated that the densities of the constrained models are lower than that of the flexible model due to the lower attraction of molecules caused by the use of a gas-phase value for the OH bond length constraints, as discussed in section 3.1. What causes the higher density of the all-bonds model compared to the h-bonds model is unclear.

On average, the all-bonds model describes the liquids best, with a mean error of −1.9 kg/m³ and a rms error of 1.4 kg/m³ compared to experiment, compared to the −5.2 and 2.1 kg/m³ of the h-bonds model and the 5.1 and 2.3 kg/m³ of the flexible model, respectively.

In the parametrization of the force field, Jorgensen et al. obtained⁵ densities of 779, 799, and 790 kg/m³ for methanol, ethanol, and propanol. They performed MC simulations without constraints (also, bond-stretching moves were included); thus, the flexible model should correspond to their results. The difference in their results can probably be explained by differences in the simulation system size and simulation parameters as Jorgensen et al. used cutoffs for Coulombic interactions. Furthermore, we used a different molecular topology for methanol. Using a cutoff of 11 Å for the (nonbonded) Coulombic and Lennard-Jones interactions with a cubic spline that makes the potential vanish smoothly during the last 0.5 Å, we obtained a density of (815 ± 1) kg/m³ for methanol with the alkane parametrization and (809 ± 1) kg/m³ for that with the Jorgensen parametrization. Using a plain cutoff of 11 Å with long-range corrections for Lennard-Jones interactions, the results were (775 ± 3) and (764 ± 3) kg/m³, respectively.

3.4. Diffusion Constant. The self-diffusion constant is an experimentally determinable quantity that can be used as a measure of the dynamics in the system. The MSD was calculated using the thermalized trajectory, resulting in 9 ns worth of data. A log–log plot of the MSDs is shown in Figure 4, where the curves of the different models have been displaced by a constant factor.

To be able to meaningfully fit diffusion constants to the simulation data, the diffusion limit has to have been reached.

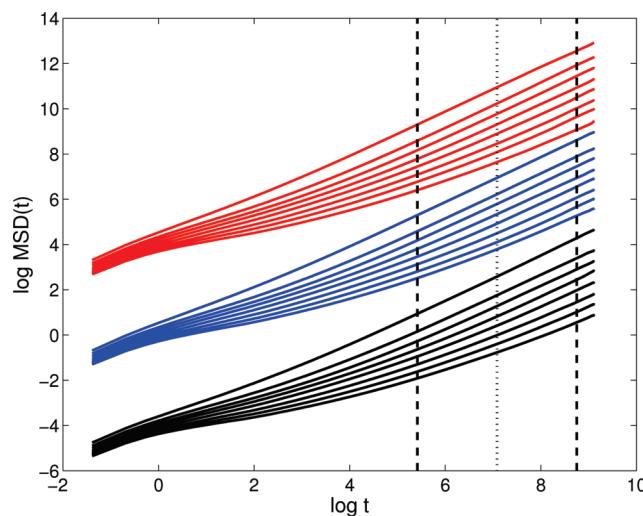


Figure 4. Plots of the MSDs. The full black lines correspond to the flexible model, and the blue and the red ones correspond to the h-bonds and all-bonds models, respectively. In each model, the highest line corresponds to methanol, the MSD decreasing as a function of increasing alkyl length.

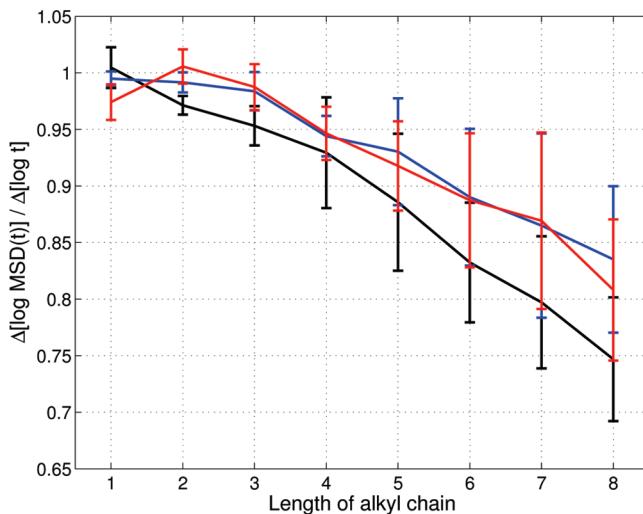


Figure 5. Fitted slopes of the MSD curves in the $(\log t, \log \text{MSD}(t))$ plane for the different models as a function of the length of the alkyl chain. The error bars denote the difference of the slope in the left and the right half of the fitting region as shown in Figure 4. The colors are the same as those in Figure 4.

To verify that the Einstein relation (eq 2) holds, we fit straight lines in the plot and see if their slopes equal unity.

As the beginning of the MSD is chaotic and the sampling of the MSD decays with growing time difference t , we only consider the region of $t = 225–6300$ ns (indicated with dashed vertical lines in Figure 4) for the fit. To obtain error estimates for the slopes, the fit region was divided into two equal parts (in the logarithm of time, as indicated by the dotted line in Figure 4), in which the fits were performed separately. The slopes of the fitted lines are shown in Figure 5.

Although the simulation time was rather long, 10 ns, the diffusion limit was not reached, strictly speaking, for alcohols heavier than butanol as the slopes of the MSDs in the log–log plots (Figure 5) were not close to unity. Even though the theoretical justification is not fulfilled, it may still be worthwhile to fit diffusion constants to the whole series.

To obtain the diffusion constants, the aforementioned fitting region was divided into 135 intervals of 45 ps, in which the

TABLE 2: Diffusion Constants in $10^{-9} \text{ m}^2/\text{s}$ for the Different Models^a

compound	flexible	h-bonds	all-bonds	experiment
methanol	1.91 ± 0.01	2.65 ± 0.05	2.5 ± 0.3	2.30^{30} 2.32^{31} 2.42^{32}
ethanol	0.78 ± 0.04	1.30 ± 0.06	1.34 ± 0.06	0.946^{30} 1.01^{31} 1.075^{32}
propanol	0.48 ± 0.03	0.83 ± 0.03	0.83 ± 0.04	0.552^{34}
butanol	0.30 ± 0.02	0.49 ± 0.04	0.49 ± 0.02	0.416^{33} 0.426^{34}
pentanol	0.20 ± 0.02	0.34 ± 0.02	0.32 ± 0.03	0.296^{33}
hexanol	0.12 ± 0.03	0.21 ± 0.02	0.20 ± 0.03	0.218^{33}
heptanol	0.08 ± 0.02	0.14 ± 0.02	0.14 ± 0.02	0.172^{33}
octanol	0.05 ± 0.02	0.09 ± 0.02	0.08 ± 0.02	0.138^{34} 0.142^{33}

^a The results and the errors are half of the sum and of the difference of the maximum and minimum values obtained from the fits over the 135 fit intervals.

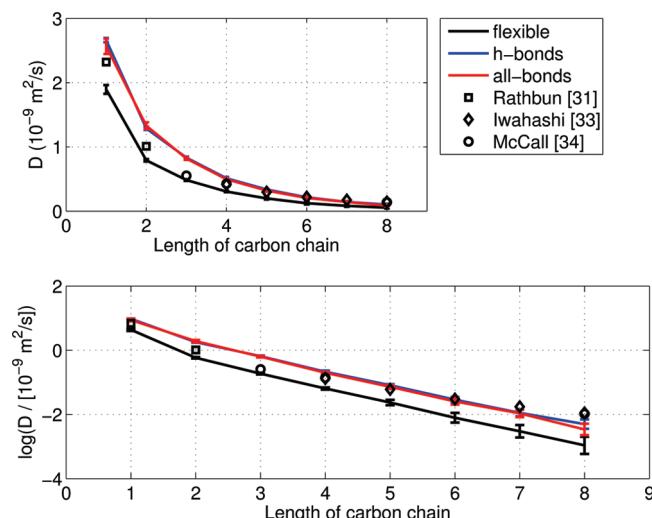


Figure 6. Diffusion constants for the three models plotted on normal (upper part) and semilog scales (lower part). The symbols represent various experimental values in the literature, as indicated in the figure.

fits to the Einstein relation were performed. The values and the error estimates for the diffusion constants were obtained by taking the mean and half of the difference, respectively, of the maximum and the minimum of the 135 individual fits over the separate intervals. The diffusion constants are given in Table 2 and plotted with experimental data from refs 30–34 in Figure 6.

The diffusion constants fitted to the simulations are quite well in agreement with experiment for alcohols up to hexanol, after which the compatibility breaks down. This is likely an artifact of the diffusion limit not having been reached in the case of these heavier alcohols.

Constrained systems exhibit higher diffusion constants than nonconstrained systems, which may be linked to the weaker interaction between molecules arising from the gas-phase reference OH bond length used in the constraints, as discussed in section 3.1.

3.5. Number of Hydrogen Bonds. The mean number of HBs is a useful quantity in the benchmarking of a force field as it is easily obtainable (given a definition for the existence of a HB) from a simulation and can be compared to ab initio results. The

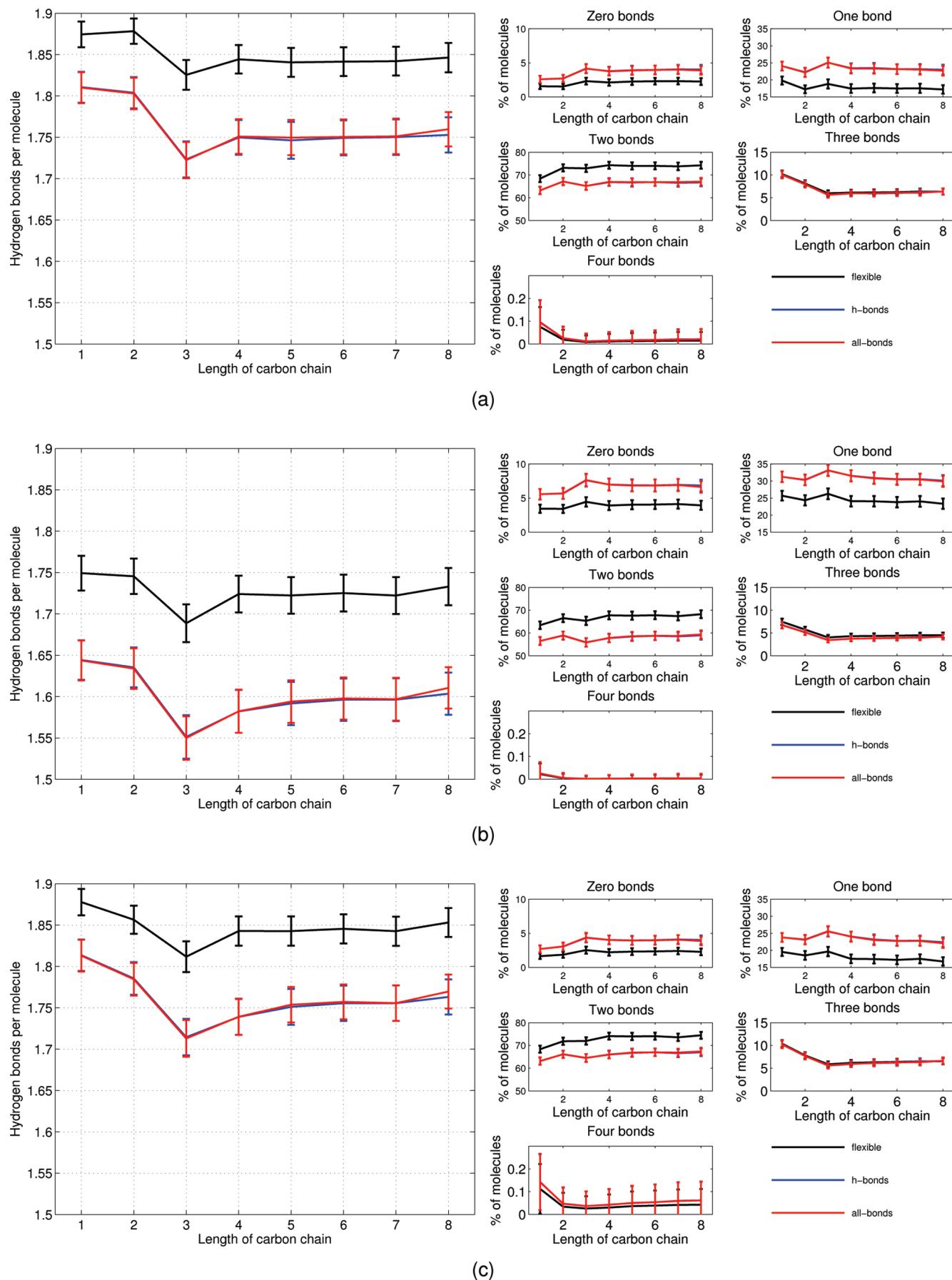


Figure 7. Mean number of HBs per molecule (left column) and distribution of the number of HBs (right column) as a function of the length of the carbon chain, with three different HB criteria, (a) Haughney,²⁷ (b) Wernet,²⁴ and (c) Chen and Siepmann.⁹ The error bars represent standard deviations.

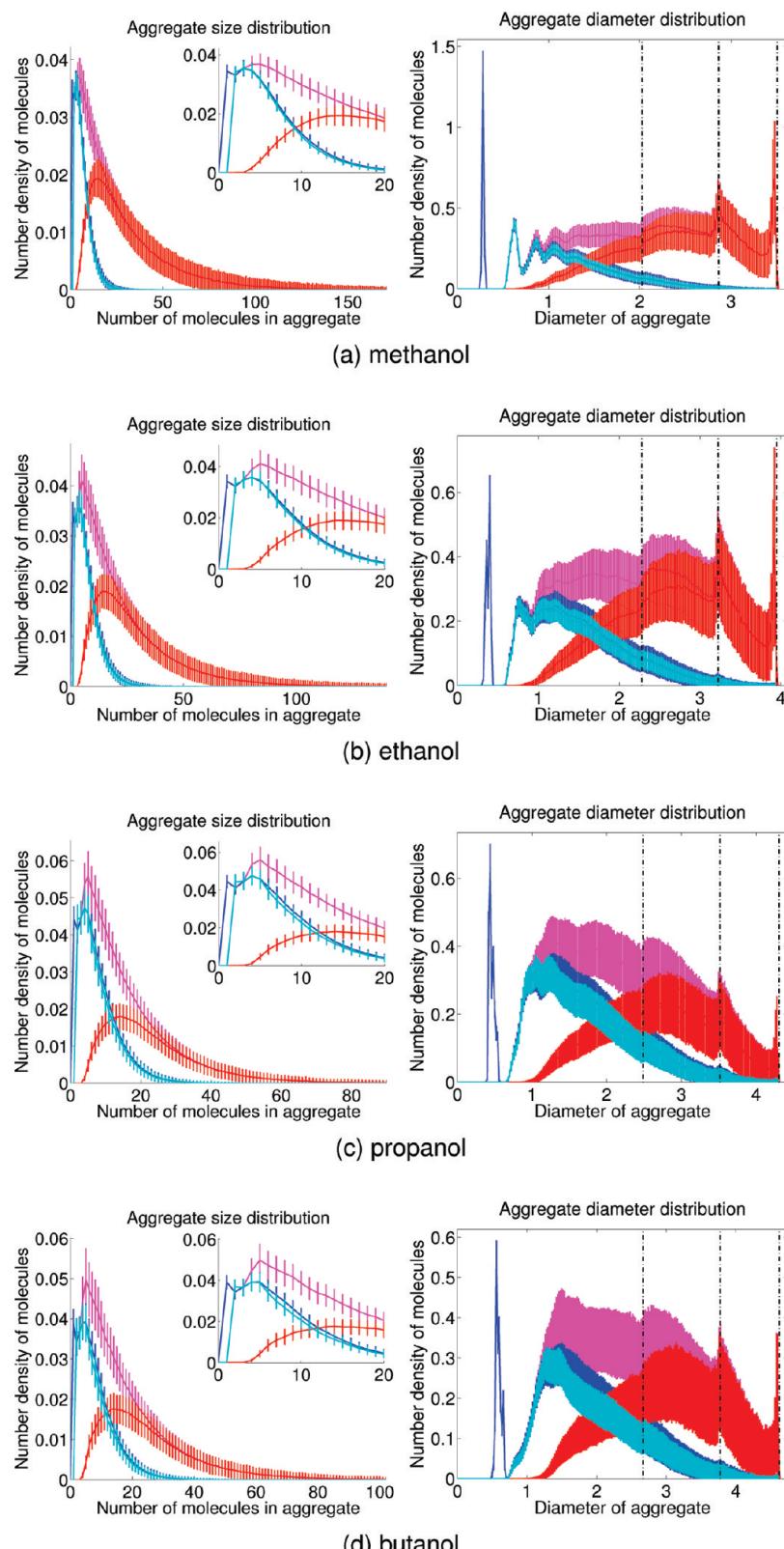


Figure 8. Aggregate size (left column) and diameter distributions (right column) from methanol (a) to butanol (d). The colors are as follows: linear aggregates, blue line; branched, red; total, magenta; nodal branches, cyan. The vertical lines in the diameter distributions represent d_1 , d_2 , and d_m (see section 2.1). The error bars represent standard deviations from the mean, calculated over the 900 histograms of the thermalized trajectory, each collected over a time period of 10 fs.

mean number of HBs per molecule according to the three HB criteria presented in section 2.1 are plotted in the left column of Figure 7.

Distinctively, methanol and ethanol have the largest number of HBs. Propanol is also exceptional in its bonding behavior,

having fewer HBs than ethanol and butanol. This feature is equally well visible with all of the H-bonding criteria considered. Liquid propanol has fewer molecules with two HBs and more with zero or one HB than ethanol or butanol. The lower mean number of HBs per molecule in propanol also appears in the

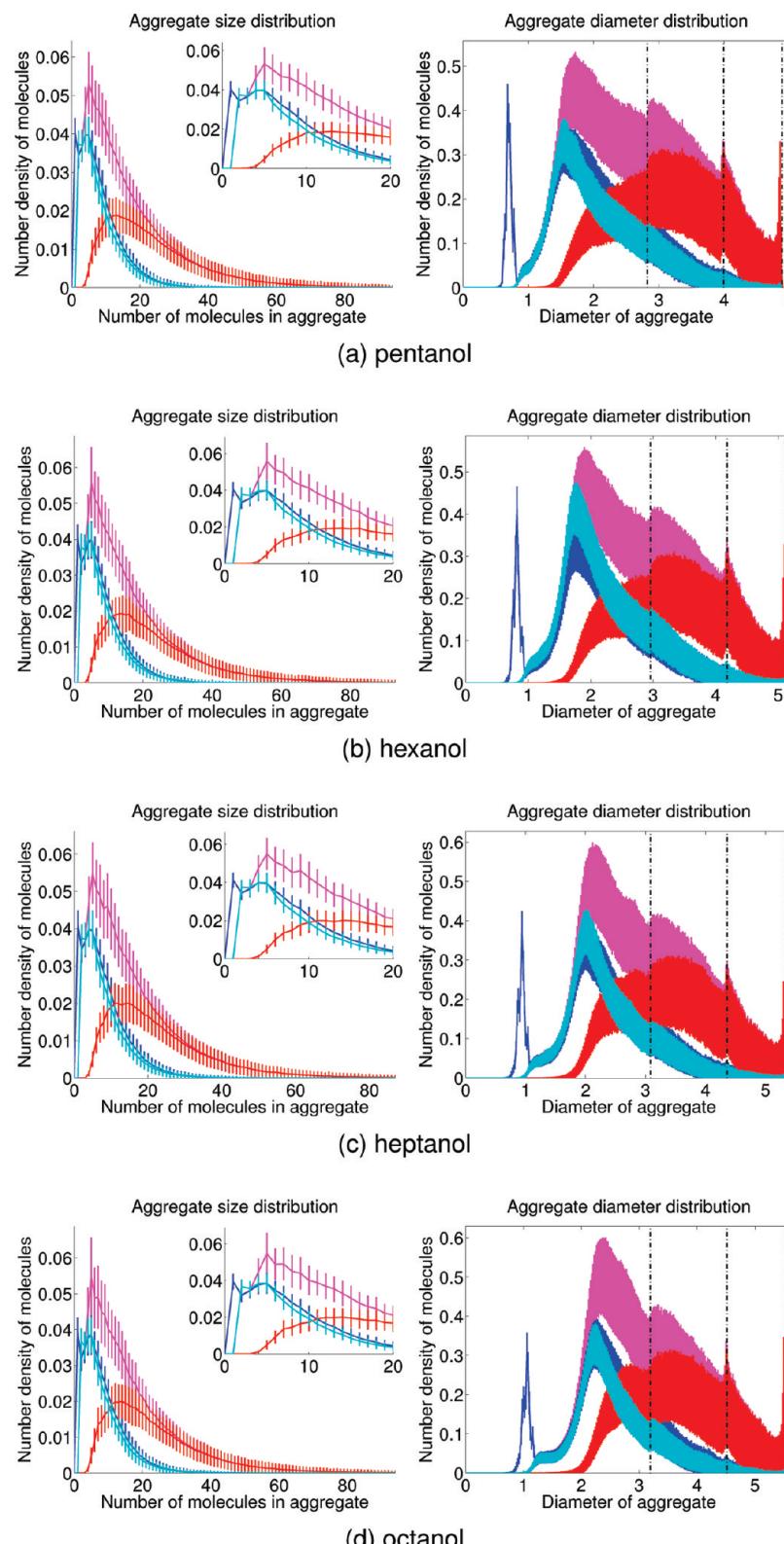


Figure 9. Aggregate size (left column) and diameter distributions (right column) from pentanol (a) to octanol (d). The colors are the same as those in Figure 8.

configurational bias Monte Carlo³⁵ simulations using the united atom transferable potential for phase equilibria (TrPPE-UA)¹¹ made by Tomšič et al.¹⁰ As such, the propanol anomaly seems to be an emergent physical feature, which is likely connected to the geometry of the propanol molecule.

Given their difference in molecular size, the alcohols from butanol to octanol are surprisingly similar in their bonding

characteristics as all of the HB distributions are practically constant through the series.

The HB criteria 1 and 3 give very similar results for all alcohols; the mean number of HBs per molecule differs by less than 1.5%, the difference being the greatest in the cases of ethanol and propanol. Criterion 2 is much stricter, having 6–8% fewer bonds than what exist under criteria 1 or 3.

TABLE 3: Fractions of Molecules in Percent in Linear, Ring-Type, and Branched Aggregates^a

compound	linear aggregates (%)	ring-type aggregates (%)	branched aggregates (%)
methanol	32 ± 2	0.3 ± 0.2	68 ± 2
ethanol	38 ± 2	0.9 ± 0.4	61 ± 2
propanol	52 ± 3	1.7 ± 0.6	47 ± 3
butanol	47 ± 3	1.9 ± 0.7	51 ± 3
pentanol	47 ± 3	2.8 ± 0.8	51 ± 3
hexanol	46 ± 3	3.1 ± 0.9	51 ± 3
heptanol	46 ± 3	3.3 ± 1.0	51 ± 3
octanol	45 ± 3	3.6 ± 1.3	52 ± 4

^a The error estimates represent standard deviations.

The number of molecules that fulfill the criterion for the existence of five HBs is $\lesssim 2 \times 10^{-3}\%$ in all alcohols according to criterion 3 and $\lesssim 6 \times 10^{-5}\%$ according to criterion 1. In the case of the criterion 2, such molecules only exist in methanol in small quantities ($10^{-5}\%$).

As shown in section 3.1, the used OH bond length constraint in the h-bonds and all-bonds models uses a gas-phase reference value, weakening the intermolecular interactions. This shows as the flexible model having more HBs per molecule than the constrained h-bonds and all-bonds models. The results for the h-bonds and all-bonds models are statistically identical.

The number of molecules having two HBs is reduced in the constrained models in comparison to the flexible model, increasing the number of molecules with zero or one HB. The use of constraints has little effect on the number of molecules with more than two HBs.

3.6. Analysis of Hydrogen-Bonded Aggregates. As the constrained models differ significantly in their H-bonding behavior (see section 3.5) from the unconstrained flexible model, which corresponds to the parametrization of the OPLS force field, for the rest of this paper, we will only consider the flexible model. Furthermore, we will only consider HB criterion 2.

The analysis of the aggregates was performed in time slices of 10 ps (see section 3.2). The error estimates represent standard deviations from the mean over the full thermalized trajectory. The fractions of molecules in linear, ring-type, and branched aggregates are given in Table 3. The size and diameter distributions of the aggregates are presented in Figures 8 (methanol–butanol) and 9 (pentanol–octanol). The right column presents the distribution of the diameters of the aggregates, whereas the left columns of the figures present the distribution of aggregate sizes (the number of molecules contained by the aggregate), with the insets displaying the behavior at small aggregate sizes.

The distributions represent the probability of finding a molecule in an aggregate of a given size or diameter. The distributions of the ring-type aggregates are not shown in the figures as the fraction of ring-type aggregates is very small. In effect, the total distribution (in magenta) is the sum of that of linear (blue) and branched aggregates (red). Also, the distributions of the nodal decompositions of the branched aggregates (nodal branches) are plotted in the figures (cyan).

Note that singletons are classified as linear aggregates. The minimal size for nodal branches is two molecules. Branched aggregates contain, by definition, at least one molecule that has at least three HBs; therefore, the minimal size of branched aggregates is four molecules.

What can be noted is the similarity of the hydrogen-bonding characteristics of the alcohols; the forms of the size distributions are the same, even though the distributions shift to favor smaller aggregates as the alkyl chain grows in length. This is an expected

feature as the distributions of the number of HBs were found to be similar in each of the alcohols. The propanol anomaly is rather clearly visible in Figure 8; the aggregates in propanol contain fewer molecules than those in ethanol or butanol.

The classification of aggregates as linear versus branched is sensitive to the choice of HB criterion. Criteria 1 or 3 result in the fraction of branched aggregates being much larger and in extremely pronounced peaks in the diameter distribution of branched aggregates. Still, the distributions of the nodal branches remain looking very much as those of linear aggregates.

As the fraction of ring-type aggregates is very small, it can be deduced that liquid alcohols consist of chains of molecules (that sometimes link together to form branched aggregates) in the OPLS model since the distributions of the linear aggregates and those of the nodal branches are extremely similar.

The aggregates that form in the system are quite large. The structure in the histograms around the diameters d_1 , d_2 , and d_m are clear indications of finite size effects (FSEs) caused by percolation, which shall be discussed later. However, the majority of the linear aggregates, ring-type aggregates, and nodal branches are smaller than the dimensions of the system, which indicates that even though there is percolation, the elementary structures are well-converged.

Percolation phenomena have been extensively studied in the literature. For a recent paper examining percolation threshold parameters in fluids, see, for example, the paper by Škvor and Nezbeda³⁶ that showed that the probability of percolation depends strongly on the temperature, the nature of the system studied, and also the used definition for a cluster (which we could see with the different HB definitions).

Even if the hydrogen bond network percolates through the system, according to the nodal decomposition, the percolating network is mostly formed of nodal chains that are smaller than the system size. Moreover, additional simulations show that the distributions of the linear aggregates and nodal branches are largely independent of the simulation system size; only the tails of the distributions are affected by the FSEs, and the artifacts vanish smoothly as the system size is increased.

What still remains to be studied is if the percolation of the network of chains itself is a FSE; percolation is present at the limit of an infinite system size only if the density is large enough (over the percolation threshold ρ_c).

The distributions in the constrained models are very similar to those in the unconstrained model. The distributions of linear aggregates and nodal branches are almost the same; however, due to the smaller mean number of HBs, there are fewer branched aggregates, in favor of more linear aggregates, in the constrained models. This is again reflected in the diameter distributions, in which the FSE peaks become smaller, although still visible.

4. Conclusions

We have studied systematically linear alcohols in the liquid phase in order to investigate the effects of the use of bond length constraints on the density, diffusion constant, and hydrogen-bonding properties. Furthermore, we have analyzed the topologies of the forming H-bonded networks in the alcohols.

We have presented an alternative method of analyzing clustering—nodal analysis—which is useful in studying structuring in weakly bonded systems. The application of nodal analysis, the nodal decomposition, breaks structures into elementary chains, giving information about the relevant length scales of the structures that are formed in the simulations.

To investigate the effects of bond length constraints, we have used three models in our study, the flexible model, in which no constraints are used, the h-bonds model, in which the CH and OH bond lengths have been fixed to the OPLS equilibrium values, and the all-bonds model, where, in addition to the constraints of the h-bonds model, also the CC and CO bond lengths have been fixed.

The bond length distributions of the models were examined, and it was found that with constraints, the mean bond length remains the same as that in the unconstrained case, except for OH bonds, where the interaction between molecules in the liquid stretches the OH bond length out of the gas-phase reference value that was used as the constraint value. This was found to have implications for the dynamical and structural properties of the system. If bond length constraints are used in simulations, the reference values obviously need to be chosen with care. As the equilibrium bond lengths depend on the environment, they are system-specific. Algorithms able to tune the reference bond lengths toward the system equilibrium values should be developed.

Although the flexible model corresponds to the parametrization of the OPLS force field, the experimental densities are better reproduced by the constrained h-bonds and all-bonds models, the latter having the smallest errors on average. The differences to the experimental values and the ones obtained in the parametrization of the force field can likely be explained by differences in the simulation methods.

Self-diffusion constants were calculated and found to roughly coincide with the experimental values for the alcohols for which the diffusion limit was reached. A clear difference between the unconstrained and the constrained models could be seen in the constants; diffusion in the constrained models was significantly faster than that in the unconstrained model, which we interpreted as a consequence of the weaker intermolecular interactions caused by the gas-phase reference value of the OH bond length constraint.

Three geometrical criteria for the existence of a hydrogen bond were considered, the criterion of Haughney et al.,²⁷ the one of Wernet et al.,²⁴ and the one of Chen and Siepmann.⁹ We demonstrated that also the formation of HBs is sensitive to the OH bond length constraints that were used, which is likely another effect of the use of the gas-phase reference value.

We studied the topologies of the hydrogen-bonded aggregates using the HB criterion of Wernet et al. in the unconstrained flexible model. The fractions of linear, ring-type, and branched aggregates were reported for each of the alcohols. The sizes and the diameters of the forming aggregates were studied, indicating finite size effects that are present in the HB network topologies.

After nodal decompositions were performed on the branched aggregates, it was found that the elementary structures of the HB networks are well-converged and that the (nodal) chains forming the branched aggregates closely resemble linear aggregates. This in combination with the small fraction of ring-type aggregates was taken as evidence for liquid alcohols consisting of H-bonded molecule chains, which may link together to form branched aggregates.

In the case of methanol, this conclusion is at variance with the findings of Guo et al.,¹² who suggested that liquid methanol consists of chains and rings of six or eight molecules in equal abundance, and those of Wilson et al.,¹³ who proposed long chains of more than six molecules and H-bonded rings. Not only is the fraction of rings very small in the molecular dynamics simulations that we have performed, but also, a large fraction of the chains comprise less than seven molecules.

Our hypothesis of the structure of liquid alcohols, which is based on the simulations performed in this paper with the OPLS

force field,⁵ can be tested experimentally at modern synchrotron radiation sources, using Compton or X-ray Raman scattering as indirect structural probes.

Acknowledgment. This work has been supported by the Academy of Finland [Contracts 110571, 1127462, its Centers of Excellence Program (2006–2011)] and the Jenny and Antti Wihuri Foundation, to which the authors express gratitude. The computing resources of the Center for Scientific Computing (Espoo, Finland) are gratefully acknowledged. We thank Empu Salonen for insightful comments on the manuscript.

Appendix

Generating initial configurations with the equilibrium density can be rather difficult, especially when the molecules are large and constraints are used since this makes the energy minimization nontrivial. For instance, in this work, the generation of cubic configurations of the all-bonds model of tridecanol proved troublesome. It is much easier to generate starting configurations with small densities since they can be energy-minimized nicely with no molecular overlap.

However, the optimization of simulation parameters depends on the physical extent of the system via the grid used in particle mesh Ewald.²⁰ When constraints are used, the constraint algorithm (or its implementation) can also have some limitations related to the size of the system. It is thus important that the size of the system not change radically during the simulation, as otherwise some of the calculations end up being performed on an unoptimal grid. This means that the initial density needs to be close to the equilibrium value.

This equilibration can of course be performed by the use of molecular dynamics, but it is inefficient since the pressure coupling cannot be very strong. However, when no dynamics are in place, more abrupt methods can be used. A combined method of energy minimizations and deformations of the system allows for generation of (cubic) configurations from (noncubic) initial configurations as presented in Algorithm 1, which was motivated by the method of simulated (pressure) annealing. The algorithm should be quite easy to extend to triclinic cells.

Algorithm 1 Generation process of optimal initial configurations in cubic box.

1. Set the desired density ρ_0 and the value of the scaling constant α .
2. Generate an initial configuration and minimize its energy.
3. Transform the system towards the optimal one
 - (a) determine the scaling factor s
 - i. if $\rho < \rho_0(1 - \alpha)^3$, set $s = 1 - \alpha$.
 - ii. if $\rho_0(1 - \alpha)^3 \leq \rho \leq \rho_0(1 + \alpha)^3$, set $s = \sqrt[3]{\rho/\rho_0}$.
 - iii. if $\rho > \rho_0(1 + \alpha)^3$, set $s = 1 + \alpha$.
 - (b) calculate partial deviations t_i from cubic box

$$t_x = 1 - 2\alpha \left(\frac{L_x}{\sqrt{L_y L_z}} - 1 \right)$$

$$t_y = 1 - 2\alpha \left(\frac{L_y}{\sqrt{L_x L_z}} - 1 \right)$$

$$t_z = 1 - 2\alpha \left(\frac{L_z}{\sqrt{L_x L_y}} - 1 \right)$$
 - (c) scale system towards a cubic box

$$L_x \rightarrow \frac{s t_x}{\sqrt{t_y t_z}} L_x$$

$$L_y \rightarrow \frac{s t_y}{\sqrt{t_x t_z}} L_y$$

$$L_z \rightarrow \frac{s t_z}{\sqrt{t_x t_y}} L_z$$
4. Minimize the energy of the configuration.
5. If density is not the desired one or the box is not cubic return to step Item 3. Else write out final starting configuration.

What can be furthermore noted is that the minimization does not need to be exact. At least in the initial stages, it is not worthwhile to use costly, elaborate methods for long-range interactions, but cheaper cut-offs can be used instead.

The algorithm scales in each step the density ρ by a factor of $(1 \pm \alpha)^{-3}$ until the desired value ρ_0 is attained. Also, the form of the simulation box can be deformed during the minimization process; in principle, any geometry can be transformed into any other geometry by the use of this method. A small value of α makes the minimization step easier but, on the other hand, increases the necessary number of minimization steps.

We generated the initial configurations by placing randomly rotated molecules on a cubic grid, with an initial density on the order of $\rho \approx 100 \text{ kg/m}^3$, and optimizing the configuration to the density $\rho_0 = 800 \text{ kg/m}^3$ with the scaling constant value of $\alpha = 0.04$.

References and Notes

- (1) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. *Comput. Phys.* **1977**, *23*, 327–341.
- (2) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. J. *Comput. Chem.* **1997**, *18*, 1463–1472.
- (3) Tironi, I. G.; Brunne, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1996**, *250*, 19–24.
- (4) van Gunsteren, W. F.; Karplus, M. *Macromolecules* **1982**, *15*, 1528–1544.
- (5) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. J. *Am. Chem. Soc.* **1996**, *118*, 11225.
- (6) Stephenson, S. K.; Offeman, R. D.; Robertson, G. H.; Orts, W. J. *Chem. Eng. Sc.* **2006**, *61*, 5834–5840.
- (7) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276–1284.
- (8) Vahvaselkä, K. S.; Serimaa, R.; Torkkeli, M. *J. Appl. Crystallogr.* **1995**, *28*, 189–195.
- (9) Chen, B.; Siepmann, J. I. *J. Phys. Chem. B* **2006**, *110*, 3555–3563.
- (10) Tomšič, M.; Jammik, A.; Fritz-Popovski, G.; Glatter, O.; Vlček, L. *J. Phys. Chem. B* **2007**, *111*, 1738–1751.
- (11) Chen, B.; Potoff, J. J.; Siepmann, J. I. *J. Phys. Chem. B* **2001**, *105*, 3093–3104.
- (12) Guo, J.-H.; Luo, Y.; Augustsson, A.; Kashtanov, S.; Rubensson, J.-E.; Shuh, D. K.; Ågren, H.; Nordgren, J. *Phys. Rev. Lett.* **2003**, *91*, 157401.
- (13) Wilson, K. R.; Cavalleri, M.; Rude, B. S.; Schaller, R. D.; Catalano, T.; Nilsson, A.; Saykally, R. J.; Pettersson, L. G. M. *J. Phys. Chem. B* **2005**, *109*, 10194–10203.
- (14) Hakala, M.; Nygård, K.; Vaara, J.; Itou, M.; Sakurai, Y.; Hämäläinen, K. *J. Chem. Phys.* **2009**, *130*, 034506.
- (15) Berendsen, H.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (16) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
- (17) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (18) Hess, B.; Kurtzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (19) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
- (20) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. *J. Chem. Phys.* **1995**, *103*, 8577–8592.
- (21) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- (22) Berendsen, H.; Postma, J.; van Gunsteren, W.; Dinola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (23) Einstein, A. *Ann. Phys.* **1905**, *17*, 549–560.
- (24) Wernet, P.; Nordlund, D.; Bergmann, U.; Cavalleri, M.; Odelius, M.; Ogasawara, H.; Näslund, L. Å.; Hirsch, T. K.; Ojamäe, L.; Glatzel, P.; Pettersson, L. G. M.; Nilsson, A. *Science* **2004**, *304*, 995–999.
- (25) Smith, J. D.; Cappa, C. D.; Wilson, K. R.; Messer, B. M.; Cohen, R. C.; Saykally, R. J. *Science* **2004**, *306*, 851–853.
- (26) McGrath, M. J.; Siepmann, J. I.; Kuo, I.-F.; Mundy, C. J.; VandeVondele, J.; Hüttner, J.; Mohamed, F.; Krack, M. *J. Phys. Chem. A* **2006**, *110*, 640–646.
- (27) Haughney, M.; Ferrario, M.; McDonald, I. R. *J. Phys. Chem.* **1987**, *4934*–4940.
- (28) Chen, B.; Siepmann, J. I. *J. Phys. Chem. B* **2000**, *104*, 8725–8734.
- (29) *Handbook of Chemistry and Physics*; Lide, D. R., Ed.; CRC Press LLC: Boca Raton, FL, 2001.
- (30) Karger, N.; Vardag, T.; Lüdemann, H.-D. *J. Chem. Phys.* **1990**, *93*, 3437–3444.
- (31) Rathbun, R. E.; Babb, A. L. *J. Phys. Chem.* **1961**, *65*, 1072–1074.
- (32) Hurle, R. L.; Easteal, A. J.; Woolf, L. A. *J. Chem. Soc., Faraday Trans. 1* **1985**, *81*, 769–779.
- (33) Iwahashi, M.; Ohbu, Y.; Kato, T.; Suzuki, Y.; Yamauchi, K. *Bull. Chem. Soc. Jpn.* **1986**, *59*, 3771–3774.
- (34) McCall, D. W.; Douglass, D. C. *J. Chem. Phys.* **1960**, *32*, 1876–1877.
- (35) Siepmann, J. I.; Frenkel, D. *Mol. Phys.* **1992**, *75*, 59–70.
- (36) Škvor, J.; Nezbeda, I. *Phys. Rev. E* **2009**, *79*, 041141.

JP909894Y