PCCP

PAPER

Cite this: *Phys. Chem. Chem. Phys.*, 2015, **17**, 9885

Simple and complex disorder in binary mixtures with benzene as a common solvent

Martina Požar,[a] Jean-Baptiste Seguier,[b] Jonas Guerche,[b] Redha Mazighi,[b] Larisa Zoranić,[a] Marijana Mijaković,[a] Bernarda Kežić-Lovrinčević,[a] Franjo Sokolić[a] and Aurélien Perera*[b]

Received 19th December 2014,
Accepted 5th March 2015

DOI: 10.1039/c4cp05970k

www.rsc.org/pccp

Substituting benzene for water in computer simulations of binary mixtures allows one to study the various forms of disorder, without the complications often encountered in aqueous mixtures. In particular, we study the relationship between the local order generated by different types of molecular interactions and the nature of the global disorder, by analyzing the relationship between the concentration fluctuations and the correlation functions and the associated structure factors. Alkane–benzene mixtures are very close to ideal mixtures, despite appreciable short range shape mismatch interactions, acetone–benzene mixtures appear as a good example of regular mixtures, and ethanol–benzene mixtures show large micro-segregation. In the latter case, we can unambiguously demonstrate, unlike in the case of water, the appearance of domain–domain correlations, both in the correlation functions and the structure factor calculated in computer simulations. This finding helps to confirm the existence of a pre-peak in the structure factor associated with the micro-heterogeneity, which was speculated from several of our previous simulations of aqueous–alcohol mixtures. The fact that benzene as a solvent allows us to solve some of the problems that could not be solved with water points towards some of the particularities of water as a solvent, which we discuss herein. The concept of molecular emulsion put forward in our earlier work is useful in formulating these differences between water and benzene through the analogy with direct and inverse micellar aggregates.

1 Introduction

Liquids are textbook examples of disordered systems[1] (excluding here ordered forms of liquids such as nematic or smectic liquid crystalline phases). They are usually almost as dense as solids, but unlike the latter, they possess only a local order, which strongly depends on the nature of the molecular interactions.[2] In the case of mixtures, the result of the competition between the local ordering and the thermal agitation is the presence of concentration fluctuations.[3] One interesting question is whether it is possible or not to guess the nature of the local ordering by monitoring the concentration fluctuations. In other words, is it possible to classify liquid mixtures according to the nature of their concentration fluctuations, which can be measured through the Kirkwood–Buff integrals (KBIs)?[4] Such a question was first asked in a seminal paper by Matteoli and Lepori,[5] who experimentally investigated several binary aqueous mixtures and displayed a rich array of behaviour, hinting that an underlying classification was possible, possibly based on the hydrophilic/hydrophobic nature of the solutes. However, water is a complex liquid and several of its properties are still not fully understood[6] and their explanations are subject to controversies.[7] Therefore, in the present work we would like to revisit this question through computer simulations, by using a simpler solvent such as benzene. Other choices such as carbon tetrachloride or toluene are equally good candidates. As shown herein, this type of system presents much less of the problems encountered when studying aqueous mixtures, and moreover allows better insights into the nature of the disorder in liquids.

One of the interesting problems about concentration fluctuations is that they are primarily a static quantity, as measured by the Kirkwood–Buff integrals,[5] and as such they should not be dependent upon their lifetime. Indeed, static quantities are measured by statistical averages over many independent configurations, which are not necessarily successive in time.[8] To illustrate the difficulty posed by this problem, let us consider a mixture of disk-like and rod-like molecules, such as benzene and heptane, for example. Clearly, we expect that at any point of time, molecules of similar shape will tend to group themselves for entropic reasons. This is exactly what we observe in our simulations.

[a] *Department of Physics, Faculty of Sciences, University of Split, Nikole Tesle 12, 21000, Split, Croatia*
[b] *Laboratoire de Physique Théorique de la Matière Condensée (UMR CNRS 7600), Université Pierre et Marie Curie, 4 Place Jussieu, F75252, Paris cedex 05, France. E-mail: aup@lptmc.jussieu.fr*

Yet, despite these obvious concentration fluctuations induced by shape mismatch, the corresponding KBIs are surprisingly ideal. In other words, the persistence of clusters induced by shape compatibility does not seem to reflect the existence of any particular correlations in this system. Conversely, when considering mixtures of ethanol and benzene, we expect and observe the existence of hydrogen bonded ethanol clusters. In this case, however, the corresponding concentration fluctuations give rise to strongly non-ideal KBIs, suggesting that H bonded clusters produce stronger static correlations, independently of any cluster considerations about their lifetime. These two systems differ through the magnitude of the interaction energies involved, as we will demonstrate below. Yet, the second system shows a feature totally absent from the first: the existence of a prepeak in the oxygen–oxygen structure factor. While this alcohol pre-peak is already present in the many neat alcohols, as known from experiments,[9,10] the striking feature that we find here is that it increases in magnitude as ethanol concentration diminishes, suggesting that ethanol supramolecular cluster structures are more visible when immersed in benzene than in neat ethanol. The experimental observation of such a pre-peak by X-ray or neutron scattering experiments meets the problem of relating unambiguously the atom–atom structure factors found in simulations with the global intensity measured in scattering experiments.[11,12]

The micro-heterogeneity of aqueous mixtures has been progressively made apparent by the problems they posed in the statistical analysis of computer simulations.[13] Indeed, since these labile structures have a larger scale and slower time dynamics, they require larger system sizes and longer statistics. It is not possible to use multiscale simulation strategies because these cluster structures are not like independent large molecules (proteins) for which these techniques are adapted: these clusters exchange molecules constantly, such that their intrinsic spatio-temporal scale is not separable from that of the constituent molecules. The relation between clustering, KBIs and cellular crowding was recently pointed out by Smith.[14] While the statistical problems posed by these structures seem to be formidable in aqueous mixtures, the present study shows that it is possible to study similar structures in benzene, and at a reasonable computational cost.

Mixtures of benzene with alkanes and alcohols have been experimentally studied by very many authors. The association of ethanol in benzene has been studied through various thermodynamical models[15,16] and the resulting micro-heterogeneous structure was studied by light scattering techniques.[17] To our knowledge, there are no computer simulation studies of the benzene–ethanol mixtures with focus on the cluster formation and related KBIs. Conversely, methanol benzene mixtures have been studied by computer simulations, with focus on aggregation of the alcohol,[19] and with focus on the KBIs.[18] Similarly, the association of terbutyl alcohol in benzene was analyzed by infrared spectroscopy and quantum calculations.[20] While these studies show that alcohol molecules aggregate in benzene, much like they do in their respective neat systems, these do not address the link between local fluctuations and aggregation. This link can be efficiently found by studying the structure factors, as we show in the present study.

2 Models and simulation details

We have conducted molecular dynamics simulations of several binary mixtures involving benzene as a common solvent component. We used the OPLS[21] force field for all the solute components, namely pentane, heptane, acetone, ethanol and benzene. In order to check for possible artefacts due to the choice of any particular ethanol force field, we have also briefly considered the TraPPE model for ethanol.[22] The force field parameters are displayed in Table 1.

All simulations were done with the Gromacs 4.6 package.[23] We systematically used system sizes of N = 2048 molecules, except for the two cases of ethanol–benzene mixtures, where we used N = 16 384 molecules, corresponding to doubling the box size, which allowed us to better explore the strong micro-heterogeneity present in these system mixtures. These two cases correspond to the benzene mole fraction x = 0.5 and x = 0.8. The system box sizes vary between different cases since we keep the number of particles fixed in a constant pressure simulation. In average, for the N = 2048 particle system, the box size was about 60–67 Å, and for the N = 16 384 it was around 130 Å. The simulations were performed in the isothermal isobaric (constant NPT) ensemble, with temperature T = 300 K maintained constant using a modified Benrendsen thermostat and pressure maintained constant using the Parrinello–Rahman barostat (both with a time constant of 0.1 ps). The leap-frog integrator time step was fixed at 1 fs. We followed the same protocol for all of our simulations. All initial configurations were generated using the packmol package,[24] with appropriate pdb files for each molecule. The system was then energy minimized, followed by constant *NVT* simulations of 500 ps performed to obtain an initial equilibrium configuration. A 500 ps run was subsequently performed in the *NPT* ensemble, which allowed us to reach the 1 atm pressure in all the cases. Production runs were performed over 4 ns runs, for collecting the site–site correlation functions. The site–site structure factor reported here is calculated by direct Fourier transform of the site–site correlation functions by standard numerical methods.[25]

The present work aims at showing the correlations and the underlying structure. Therefore we will not focus our attention on comparing the models to the usual experimental results

Table 1 OPLS force field parameters for the molecules used in the paper (the values in parenthesis correspond to the TraPPE force field for ethanol)

Molecule	Atom/united atom	Sigma (Å)	Epsilon (kJ mol^{-1})	Charge (e)
Benzene	CH	3.75	0.460	0
Alkanes	CH_3	3.91	0.732	0
	CH_2	3.91	0.494	0
Acetone	CH_3	3.78	0.866	0
	C	3.75	0.439	0.470
	O	2.96	0.879	−0.470
Ethanol	O	3.07(3.19)	0.711(0.65)	−0.674(−0.82)
	H	0.00(1.58)	0.00(0.088)	0.408(0.52)
	CH_3	3.78(3.75)	0.866(0.867)	0(0)
	CH_2	3.91(4.07)	0.494(0.40)	0.266(0.30)

concerning thermodynamics such as densities and excess enthalpies, or diffusion constants. We defer to a subsequent report[26] such comparisons as well as the detailed comparisons with other force fields such as the TraPPE force field for example. We conducted few calculations with this force field and while results for thermodynamic properties can be somewhat different depending on systems, the rather small differences between the correlations[26] do not affect the conclusions drawn here concerning the local order in different systems.

2.1 The Lebowitz–Percus correction

The site–site structure factors $S_{ij}(k)$ are defined as:[1]

$$S_{ij}(k) = 1 + \rho\sqrt{x_i x_j}\tilde{h}(k) \quad (1)$$

where x_i is the mole fraction of the component to which site i belongs, ρ is the number density $\rho = N/V$ (N is the total number of particles and V the volume), $h_{ij}(r) = g_{ij}(r) - 1$, and the tilde indicates a Fourier transform. The $S_{ij}(k)$ were computed by direct Fourier transform of the site–site correlation functions $g_{ij}(r)$. It was first necessary to correct the asymptotes of the site–site functions. Indeed, as reported in our previous studies,[13,27] in systems of finite extent, such as that encountered in a computer simulation, even when made pseudo-infinite through periodic boundary conditions, the exact limit in eqn (3) is to be replaced by one involving explicitly the finite size of the system with finite total number N of molecules:

$$\lim_{r\to\infty} g_{ij}(r) = 1 - \frac{\varepsilon_{ij}}{N} \quad (2)$$

with

$$\varepsilon_{ij} = \frac{1}{\rho x_i x_j}\left(\frac{\partial \rho_i}{\partial \beta \mu_j}\right)_{TV\mu_k} \quad (3)$$

where ρ_i and μ_i are the number density and chemical potential of species i, respectively, and $\beta = 1/k_B T$ is the Boltzmann factor (k_B the Boltzmann constant and T the temperature). It is clear that this expression reduces to the expected limit $\lim_{r\to\infty} g_{ij}(r) = 1$ in the thermodynamic limit $N \to \infty$. Eqn (2) was derived for the first time by Lebowitz and Percus in 1961.[28] We give a simpler intuitive derivation of this LP asymptote correction in the appendix. We note here that, in many textbooks, the LP correction is often quoted only for the ideal mixture case, where one has $g_{ij}(r) \to 1 - 1/N$, since $\varepsilon_{ij} = 1$ for the ideal mixture. This ideal limit is properly mentioned in textbooks such as ref. 1, but often incorrectly quoted in many others. For large systems of $N \geq 10^3$ particles, the above correction is nearly irrelevant when computing quantities such as the excess internal energy which involve multiplying with the pair interaction which is often very short ranged (the Coulomb interaction is not an issue either because of the cancellation of the long range contribution that occurs because of global electroneutrality). However, direct integration of the bare simulation $g_{ij}(r)$ such as in eqn (11) will always lead to an erroneous contribution $-(\varepsilon_{ij}/N)\int_B d\vec{r}$ where the integral is over half the simulation box size. We proposed to correct this problem by multiplying $g_{ij}(r)$ by the factor $1\Big/\left(1 - \frac{\varepsilon_{ij}}{N}\right) \approx 1 + \varepsilon_{ij}/N$, which ensures the correct asymptote of 1. On the practical side, it is far easier to control for the asymptote shift by looking at the distortions near $k = 0$. Minimizing such distortions corresponds to the best asymptote shift. Finally, we note that the expression eqn (3) holds only in simulations in the constant NVT canonical ensemble, while eqn (2) holds equally in the isobaric ensemble that we use in our simulations. Since we use a numerical method to shift the asymptote, the differences for the asymptote correction between the canonical and isobaric ensemble are implicitly addressed.

3 Results

3.1 Asymptote corrections for mixtures

In Fig. 1, we illustrate the requirements for the LP correction by displaying some typical site–site correlation functions of the benzene–pentane equimolar mixture, namely the correlations between the carbon atoms of benzene $g_{CC}(r)$, between the CH_3 united-atoms of the pentane molecule $g_{MM}(r)$, and that between the cross-correlation function of these 2 sites $g_{CM}(r)$. The short range features show typical dense liquid correlations. The inset shows the asymptotes, which clearly do not converge to 1.

The top panel of Fig. 2 shows RBKI $G_{CC}(r)$, $G_{MM}(r)$ and $G_{CM}(r)$ corresponding to the three functions shown in Fig. 1. We display both the LP corrected results and the original functions (in dashes). It illustrates the need for the LP correction if proper asymptotical limits of the KBIs are to be obtained this way.

The lower panel of Fig. 2 shows the corresponding structure factors $S_{CC}(k)$, $S_{MM}(k)$ and $S_{CM}(k)$, both for the corrected and uncorrected results. Here too, the correction affects clearly the very small-k values of these functions, leading to high and

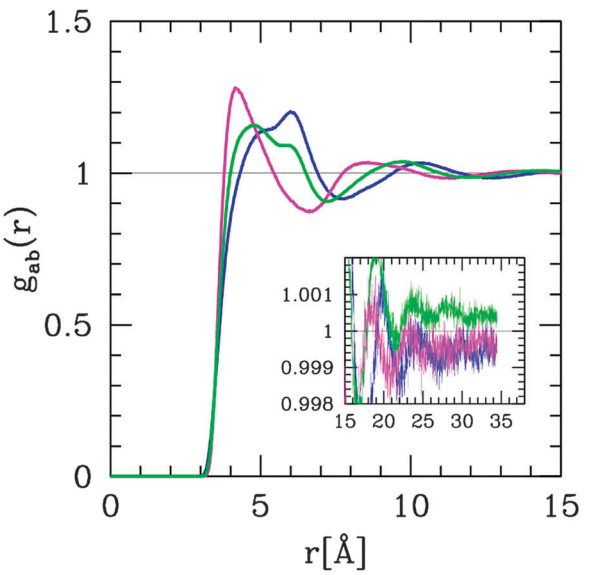

Fig. 1 Illustration of the asymptote correction for mixtures in the case of the site–site distribution functions of the benzene–pentane system. $g_{BB}(r)$ in blue, $g_{PP}(r)$ in magenta and $g_{BP}(r)$ in green. The inset shows that the asymptotes of the $g_{AB}(r)$ do not go to 1.

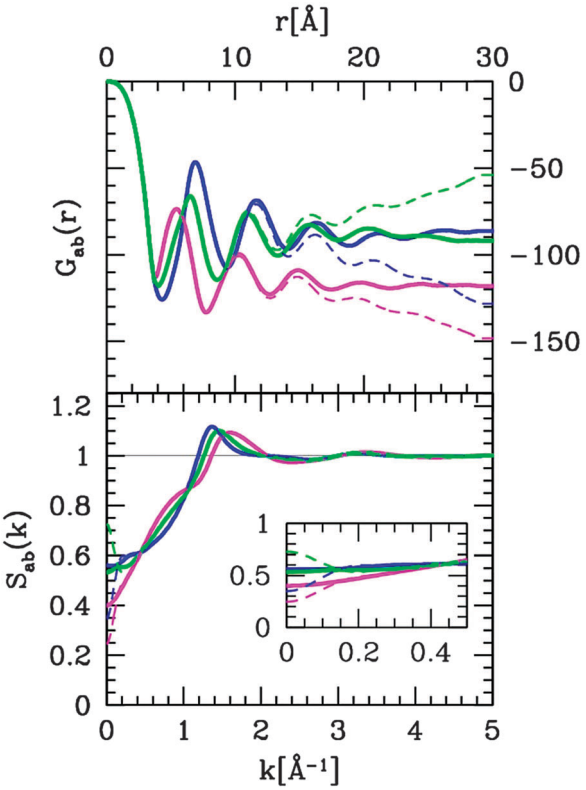

Fig. 2 (top) Running Kirkwood–Buff integrals $G_{AB}(r)$ for the system shown in Fig. 1 with same color conventions. The dashed lines show the data uncorrected for the asymptote problem, and full lines for the data after correction. (bottom) Partial structure factors associated with the functions shown in Fig. 1, with same color conventions. The dashed lines correspond to the uncorrected asymptotes, and the full lines for the data after correction.

misleading $k = 0$ values. We also note the presence of a weak shoulder in $S_{MM}(k)$, which shows the dual short range ordering of the pentane molecules: parallel and cross, which is typical of small rods.

3.2 Snapshots

We first show typical snapshots from all the systems we have studied, such that the heterogeneity of the various systems is clearly displayed before any specific studies. We show typically 3 concentrations of the largest molecule in each binary system, which are $x = 0.2, 0.5, 0.8$. This way one can observe the clustering tendencies at low concentration of either species, as well as the bi-continuous (or bi-percolating) distribution tendencies of equimolar mixtures. These 3 types of clustering tendencies have been previously reported for other aqueous mixtures.[29–31] Interestingly, they mimic those found in aqueous micro-emulsions, where they are named Winsor phases I, III and II.[32,33] Winsor I corresponds to direct micelles, Winson III to the bi-continuous phase and Winsor II to the inverse micelles. This appealing analogy has enforced us recently to name the aqueous mixtures as "molecular emulsions". Interestingly, it would seem that such a name could well apply to the non-aqueous mixture studied here, in par with similar behaviour found in non-aqueous micro-emulsions.[34] Fig. 3 shows the benzene–heptane mixtures,

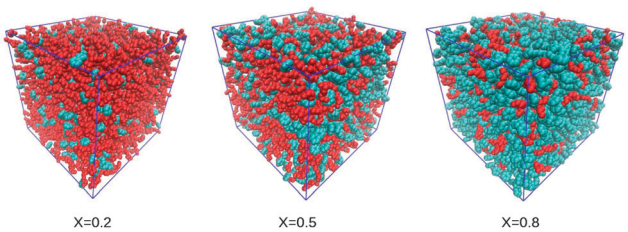

Fig. 3 Snapshots of the benzene–heptane mixtures, at three typical benzene concentrations of $x = 0.2$, $x = 0.5$ and $x = 0.8$. Benzene is shown with atoms in cyan and heptane has atoms colored in red.

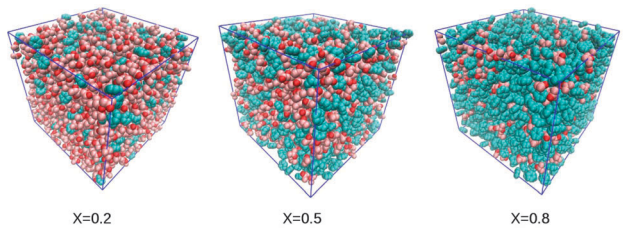

Fig. 4 Snapshots of the benzene–acetone system. The acetone oxygen atom is shown in red and the carbon groups in orange.

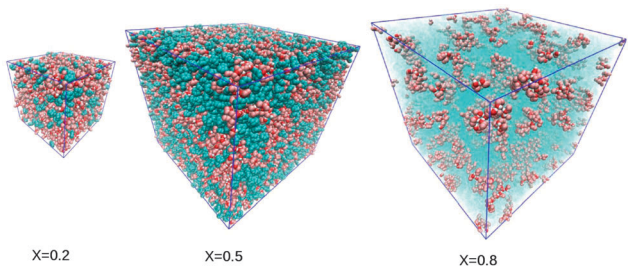

Fig. 5 Snapshots of the benzene–ethanol system. The snapshots for $x = 0.5$ and $x = 0.8$ are for the $N = 16\,384$ systems. The ethanol hydrogen atom is shown in white, the oxygen in red and the carbon group in orange.

Fig. 4 is for the acetone–benzene systems and Fig. 5 for the ethanol–benzene mixtures. Except for the 50% and 80% ethanol–benzene mixtures, for which we show snapshots of the $N = 16\,386$ systems, all other snapshots are for the $N = 2048$ system. The carbon atoms of benzene are shown in cyan, all of the methyl united atoms of the alkanes are in red, the oxygen atoms of acetone and ethanol are in red and their carbon and methyl groups are in orange. The hydrogen atoms of the ethanol are in white.

The notable first feature is how similarly heterogeneous all these snapshots look like. This is because all the systems are precisely chosen to exhibit such heterogeneity, either due to shape mismatch, or because of specific interactions. A closer look at the ethanol–benzene system shows that ethanol tends to form Hbonded clusters. This is particularly obvious in the 80% benzene–ethanol snapshot in Fig. 5, where the benzene molecules are shown as semi-transparent in order to enhance the 3 dimension effect. One can see that all ethanol clusters are grouped with the red oxygens together, forming small "circular micellar aggregates". Larger aggregates appear as a collection

of such small micelles. This is not so visible in the 50% snapshot, but the bi-continuous distribution is clear.

By looking at these snapshots, it is not really possible to tell concentration fluctuations, even "enhanced" by specific interaction, from specific clusters that could modulate the molecular distribution. These concentration fluctuations can be computed through the Kirkwood–Buff integrals.

3.3 Kirkwood–Buff integrals

The KBIs are defined as the integrals over the pair correlation functions

$$G_{ij} = \int \mathrm{d}\vec{r}\left[g_{ij}(r) - 1\right] \tag{4}$$

where i, j are the species index, and which can be expressed in terms of various equilibrium thermodynamical quantities.[4] It is worth reminding that this quantity can be equally computed from site–site functions, in place of molecular center-of-mass correlations, since the integral from is invariant from the choice of the origin.[1] We recall the expressions for the KBIs of a binary mixture[5]

$$\begin{aligned} G_{12} &= k_{\mathrm{B}}T\kappa_{\mathrm{T}} - \frac{\bar{V}_1\bar{V}_2}{VD} \\ G_{aa} &= G_{12} + \frac{1}{x_a}\left(\frac{\bar{V}_b}{D} - V\right) \end{aligned} \tag{5}$$

with $a = 1, 2$ and $b = 2, 1$, and the factor D being related to the concentration fluctuations:

$$D = x_a\left(\frac{\partial\beta\mu_a}{\partial x_a}\right)_{TP} \tag{6}$$

Due to the Gibbs–Duhem equality, D is indifferently related to both partial derivatives with respect to partial densities. The generic statistical mechanics expression for the chemical potentials involves the ideal and the excess contributions as:

$$\beta\mu_a = \ln\rho_a + \beta\mu_{a;\mathrm{ex}} \tag{7}$$

such that we have:

$$D = 1 + x_a\frac{\partial\beta\mu_{a;\mathrm{ex}}}{\partial x_a} \tag{8}$$

In the case of hypothetical "ideal" mixtures, there would be no excess contributions and one has

$$D_{\mathrm{ideal}} = 1 \tag{9}$$

The so-called regular mixture corresponds to truncating the above expression at second virial expression, leading to

$$D_{\mathrm{reg}} = 1 + \alpha x_1 x_2 \tag{10}$$

with $\alpha = 2\rho^2 B_{2;12}$, involving the cross species second virial coefficient.

In Fig. 6–8 we show the KBIs from our simulations, as obtained by direct integration of the corrected correlation functions of all 4 mixtures. We equally show the ideal KBIs ($D = 1$), or those obtained by tentatively fitting the coefficient α in the term D in eqn (10), in order to get through the simulation data. The reported analytical KBIs have been obtained by using a linear expression for the volume $V(x) = (1 - x)V_1^{(0)} + xV_2^{(0)}$ where the $V_a^{(0)}$ are the molar volumes of the neat liquids and x the the mole fraction of the largest solute in the mixture. From this relation, the partial molar volumes in eqn (5) are simply the molar volumes of the neat liquids $\bar{V}_a = V_a^{(0)}$. Our previous study

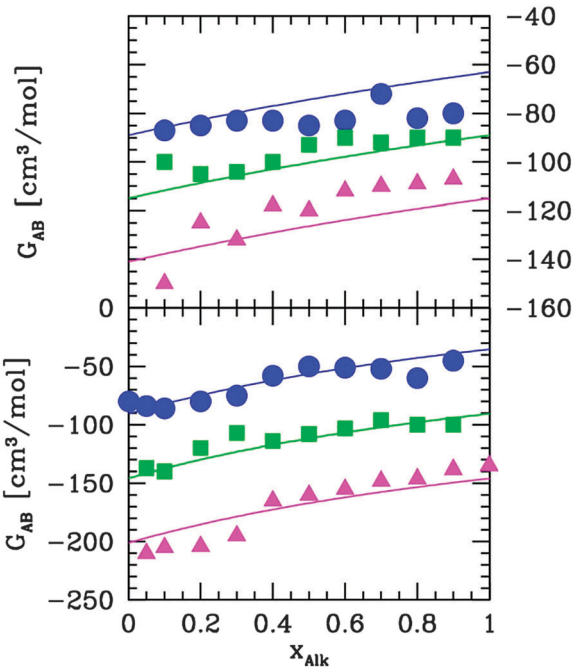

Fig. 6 Kirkwood–Buff integrals G_{AB} for the benzene–pentane system (upper panel) and benzene heptane system (lower panel), with same color conventions as in Fig. 1 and 2: blue dots for G_{SS} (S for solute), magenta triangles for benzene G_{BB} and cross functions G_{BS} in green squares. The lines are the theoretical results assuming ideal mixtures (see text).

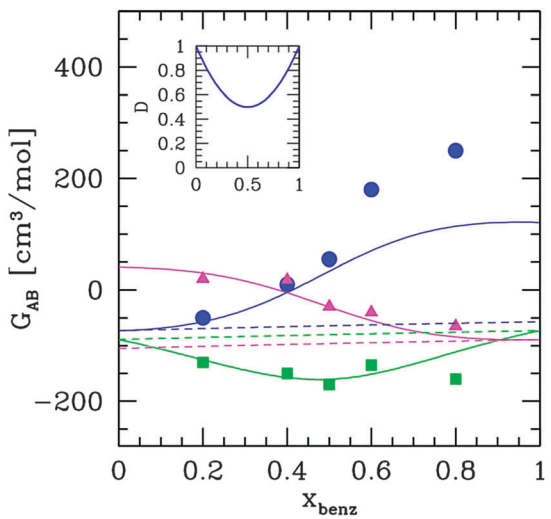

Fig. 7 Kirkwood–Buff integrals for the benzene–acetone system with same color conventions as Fig. 6. Dashed lines represent the ideal behaviour and full lines the KBIs obtained through the expression for the regular mixture given in the text and shown in the inset.

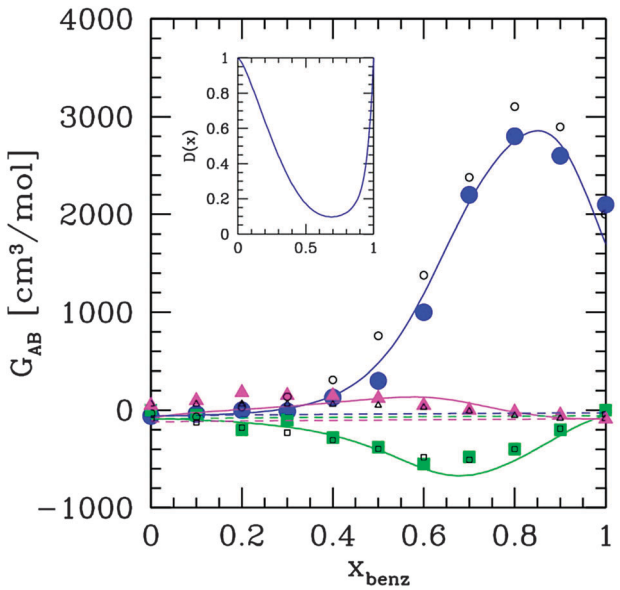

Fig. 8 Kirkwood–Buff integrals for the benzene–ethanol system with same color conventions as Fig. 6. Dashed lines for the ideal behaviour and full lines for the expression for D given in the text and shown in the inset. The open symbols are experimental KBIs for the benzene–methanol system given in ref. 18.

of the KBIs of several aqueous alcohol mixtures[35] has shown that is sufficient. We have also neglected the compressibilty term, which is usually very small for dense liquids. However, in the case of mixing benzene with alkanes this may not be true, as discussed below.

The most striking feature, in total contrast with the apparent similarities between the snapshots, is the wide differences between the 4 systems. The benzene alkane system looks almost ideal, with rather small KBIs. In contrast, the KBIs of the ethanol–benzene mixture look so much different, with values larger by almost 2 orders of magnitude. The fitting through eqn (5) indicates that since all volumes are more or less similar, then only the D term affects so greatly all the KBIs. Indeed, since D is in the denominator, the KBIs can get very large when $D \to 0$.

The Kirkwood–Buff theory of fluctuations allows one to relate the correlation functions to the species number fluctuations, which are the concentration fluctuations $\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle$ in the system,[1,4] through a key relation of statistical mechanics of liquids:

$$S_{ij}(0) = 1 + \rho\sqrt{x_i x_j}G_{ij} = \frac{\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\sqrt{\langle N_i\rangle\langle N_j\rangle}} = \sqrt{x_i x_j}\varepsilon_{ij} \quad (11)$$

where ε_{ij} is given by eqn (3).

From this equation, the large KBIs for G_{OO}, as seen in Fig. 8 at large x, seem to correspond to the underlying large concentration fluctuations $\langle N_O{}^2\rangle - \langle N_O\rangle^2$, which are also compatible with the fluctuations seen in the snapshot in Fig. 5. To gain a better insight into this problem we need to look at the details of the underlying correlation functions.

3.3.1 The benzene–alkane system: an example of shape induced clustering. In view of the fact that, from snapshots, benzene molecules tend to group with themselves and alkane acting similarly, it is very surprizing that a fit with $D = 1$, that is ideal mixture case, would near perfectly match the simulation data. Surely, such mixtures are not ideal. First of all, the fact that $(\partial\mu_{a;ex}/\partial\rho_a) = 0$ does not imply that $\mu_{a;ex} = 0$. So the denomination "ideal mixtures" found in the literature is incorrect. In order to gain a better insight at this problem, we show in Fig. 9 the site–site correlations $g_{MM}(r)$ and $g_{BB}(r)$ between the end methyl groups of heptane and the carbon atoms of benzene, for 3 different mole fractions of benzene in the benzene–heptane mixtures, namely $x_B = 0.2, 0.5, 0.8$. One sees that there are nearly no differences in the correlations between these 3 concentrations, indicating that there are very little structural changes when the concentrations are varied. This enforces the idea that the species are indeed micro-segregated, each one seeing a similar environment despite the change in concentration. However, we note that the correlation functions look very similar. This is because the site–site Lennard-Jones interactions are also very similar. So, it is more the similarity of the interactions, rather than the absence of interactions, that is the reason for the apparent ideality of such mixtures.

The noticeable differences between the theoretical curves and the simulations can have several origins. The first is the fact that the volume of the simulated neat pentane is not exactly that of the real system. Next, from the snapshots in Fig. 3 these mixtures appear more fluid than dense liquids,

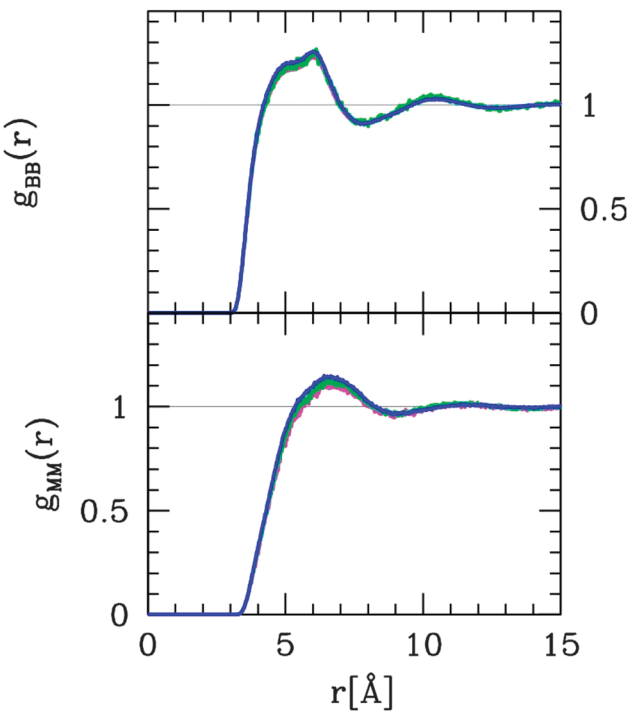

Fig. 9 Site–site correlations (top) between the carbon atoms of benzene and (bottom) between the end methyl groups of heptane, and for 3 benzene mole fractions of $x = 0.2$ (magenta), $x = 0.5$ (green) and $x = 0.8$ (blue).

since both types of liquids are rather volatile. Therefore the compressibility κ_T that is neglected in the theoretical expressions eqn (5) may be important. These 2 reasons explain the lack of a better agreement between the simulated and the analytical KBIs.

3.3.2 The benzene–acetone mixtures: an example of regular mixtures. The acetone molecule force field has both the oxygen and the central carbon atoms that wear partial charges, with total electro–neutral contributions. The oxygen atom of one acetone molecule will be attracted by the opposite charge on a carbon atom of another molecule, making the two acetone molecules to lie anti-parallel, since the bulky methyl groups avoid themselves that way. This mutual orientation is also compatible with the near neighbour dipolar interaction resulting from the charge distribution on each acetone molecule.[36] The Coulomb interaction, at a contact distance, is about a factor of 6×10^2 larger than the Lennard-Jones interaction between the sites. However, when two acetone molecules are in this configuration, no other acetone molecule can come any closer in a favorable way. So the preferred association is dimer formation. Therefore, in the benzene–acetone mixtures, acetone molecules will tend to stay in segregated pockets, where they can randomly associate into dimers.[36] The internal energy in such pockets will be in average greater (more negative) than if the mixture was strictly random. So acetone association is energy bound. When we look at the KBIs in Fig. 8, we notice that they display a small non-ideality, as depicted by the behaviour of the fitted D in the inset. We used the expression from eqn (10), $D(x) = 1 - \alpha x(1 - x)$, with $\alpha = 2$. This is entirely compatible with a weak non-ideality, limited to the second virial coefficient correction only. It confirms the enthalpic nature of the clusters that form. It also brings the question about the exact nature of such clusters: are they pseudo-particles or are they concentration fluctuations? We seek an answer that involves only equilibrium thermodynamic static quantities, and is free of any dynamical or kinetic considerations. It is known from statistical mechanics[1,8] that fluctuations are related to the $k = 0$ behaviour of the structure factor. Our own investigations relate the clusters to the existence of a pre-peak in the structure factors.[13,37,38]

To answer this question we look at the correlations in Fig. 10 for the equimolar mixture case. First, we note from the upper panel, which shows the acetone oxygen–oxygen correlations $g_{OO}(r)$, the benzene carbon–carbon $g_{CC}(r)$ and the cross correlations $g_{OC}(r)$. The inset shows the corresponding running KBI (RKBI), defined as:

$$G_{ij}(r) = 4\pi \int_0^r \mathrm{d}s\, s^2 \left[g_{ij}(s) - 1\right] \qquad (12)$$

where i,j represent atom indexes for given species, and $g_{ij}(r)$ are corrected for their asymptotes as mentioned in Section 2.1. This quantity tends asymptotically to the KBIs as defined in eqn (3). We again emphasize that, while various $G_{ij}(r)$ may depend on specific sites and species index at short range, their asymptote depend only on species pairs. This is because the total integral does not depend on the choice of the molecular centers. One can notice that the OO correlations are somewhat higher in magnitude than all others, confirming the enthalpic

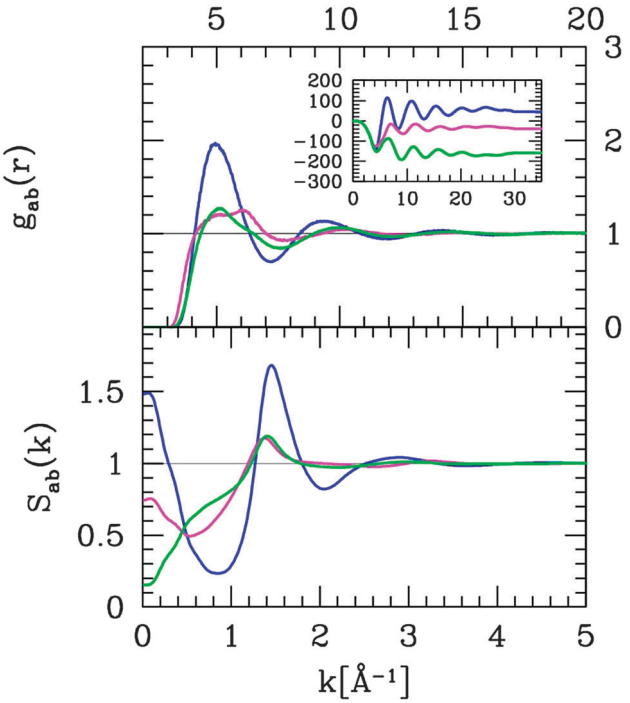

Fig. 10 Site–site correlations (top) for the equimolar benzene–acetone mixtures; blue lines for $g_{OO}(r)$ correlations between oxygen molecules on acetone, magenta lines for $g_{CC}(r)$ carbon–carbon correlations in benzene and green for cross correlations $g_{OC}(r)$. Corresponding structure factors in the lower panel. The inset in the upper panel shows the RKBI and the corresponding asymptotical values for the KBIs.

association picture, from eqn (1). The lower panel shows the corresponding structure factors. We notice that the acetone $S_{OO}(k)$ has a strong $k = 0$ raise, indicating large concentration fluctuations. Since there is no specific wave dependent peak other than that at $k = 1.4$ Å^{-1}, we conclude that the raise is due to concentration fluctuations. We note that benzene has less fluctuations than acetone, which again confirms that it is acetone that drives the self-segregation.

Finally, we observe from the KBIs in Fig. 7 that the simulation results for G_{OO} for large acetone concentrations are somewhat higher than that predicted by the analytical expression. This is due to an overestimation of G_{OO} from the simulations due to the fact that the tail of $G_{OO}(r)$ is not stabilized at a perfectly horizontal value. It is likely that a simulation of a larger system will provide a smaller value of the KBIs, more compatible with the theoretical prediction.

3.3.3 The benzene–ethanol mixtures: an example of micro-heterogeneous mixtures. The force field for ethanol shows that the charge distribution is between the oxygen and the hydrogen atoms. However, in contrast to the case of acetone, the disposition of the charges allows for a favorable direct pairing between the oxygen and hydrogen atoms of two different acetone molecules. This way, the Coulomb ordering of the ethanol molecules is not induced by the dipolar ordering, such as in the case of acetone, but by a classical equivalent of hydrogen bonding: ethanol molecules can form chains of the form O–H–O–H–O$\cdots$, with the methyl groups randomly disposed outside

the Hbonded chain. In pure ethanol, this order is permanently destroyed by both the many bonding possibilities and the thermal agitation, just like in the case of water: the strength of the Coulomb interaction is weakened by the proximity of the binding possibilities, hence allowing the bond to be easily broken by thermal agitation. In a benzene mixture, however, the presence of surrounding benzene diminishes the binding choices, and previously bonded clusters are quasi-permanently stabilized. This is the key to the interpretation of the high ethanol KBIs that are observed for this mixture, when the benzene KBIs stay comparatively small. The dashed lines in Fig. 8 show the ideal KBIs, and they are very small compared to the actual values. From the arguments developed above, one could expect that the benzene KBIs would be similar to the ideal KBIs, since benzene is not affected by charge-induced clustering. It turns out that one cannot have inhomogeneity in the distribution of ethanol without influencing that of the benzene. Hence the fluctuations in the distribution of benzene are forced to follow those of ethanol, even though they are not concerned by Coulomb induced clustering. These remarks show that the KBIs do not only reflect the concentration fluctuations, as might be deduced from eqn (11), but also by the nature of the interaction, mainly through the small-k behaviour of the various structure factors. So we need to look at the correlations to have a better insight at the underlying physics of the KBIs.

Fig. 11 shows various features of the correlation functions $g_{OO}(r)$ for the ethanol molecules, and for 4 different benzene concentrations, x = 0, 0.2, 0.5, 0.8. The main panel and the top inset show a dramatic increase of OO correlations as ethanol concentration is rarefied. This supports the idea of the formation of Hbonded specific clusters when ethanol becomes isolated. The second inset shows the long range oscillatory structure $g_{OO}(r)$, and it is seen that, as the concentration of ethanol is decreased, there is a clear long range modulation that sets it. However, in order to be able to see this domain modulation clearly, it is necessary to do simulation of N = 16 384 particles instead of N = 2048 particles, which is twice the standard size we used in most of our simulations. The last inset shows the evolution of the correlations for benzene for the concentrations x = 0.2, 0.5, 0.8 and pure benzene. It is clearly seen here that the structure of benzene evolves very little compared to Hbonded ethanol.

Fig. 12 shows RKBI $G_{OO}(r)$ for the benzene concentrations x = 0.5 (top panel) and x = 0.8 (lower panel), and for the two different system sizes. It is clearly seen that the long range part of $G_{OO}(r)$ builds a domain modulation. The small size simulations do not allow us to obtain good values of the KBIs because the tail is not stabilized for small system sizes. We also note the presence of a very large contribution at short range, which can be mistakenly confused with the actual lower value of the asymptote. This transient large value is due to the large contribution coming from the ''intra-cluster'' – or first cluster neighbour – correlations. This is akin to the large value of the first peak in the correlation of any simple liquid, which should not be confused with the lower asymptotical value that settles at

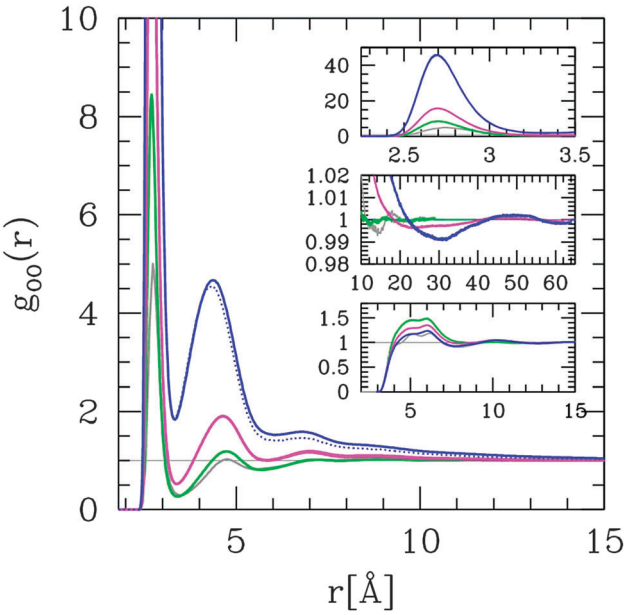

Fig. 11 (Main panel) Ethanol oxygen–oxygen correlations $g_{OO}(r)$ in benzene–ethanol mixtures, for benzene concentration x = 0 (grey curve), x = 0.2 (green), x = 0.5 (magenta) and x = 0.8 (dotted blue N = 2048, thick blue N = 16 384). Insets: (top inset) zoom on the first peak of $g_{OO}(r)$; (middle inset) zoom on the large separation behaviour of $g_{OO}(r)$ of the main panel: (lower inset) $g_{CC}(r)$ for benzene–benzene correlations for the same concentrations x as the main panel and the same color code.

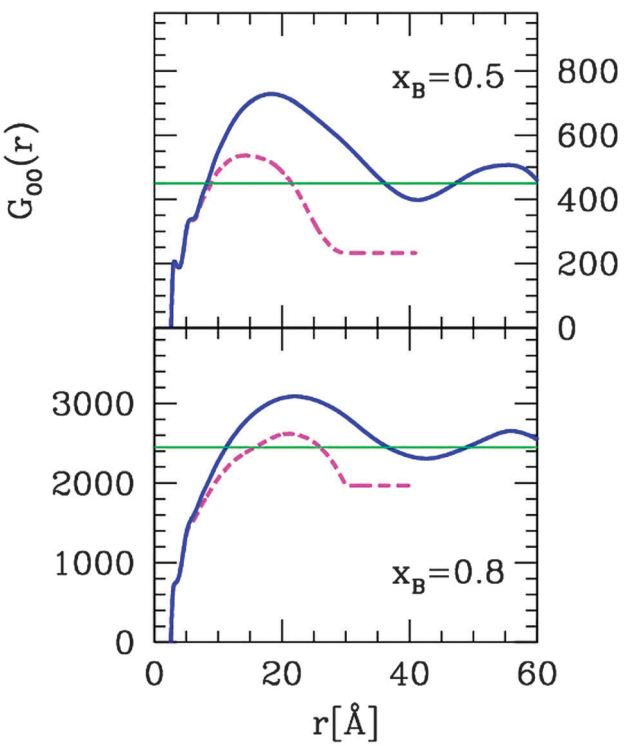

Fig. 12 Running KBI $G_{OO}(r)$ for the $g_{OO}(r)$ data at x = 0.5 shown in Fig. 11 for x = 0.5 (top) and x = 0.8 (bottom). Blue line for N = 16 385 and dashed magenta for N = 2048.

larger separations. In the simulation of simple liquids it is necessary to have system sizes of more than 100 particles as in very first simulations.[39] The situation is comparable when supramolecular assemblies are present, except that now system sizes of 10^4 are required in place of 10^3. This type of problem has considerably plagued the interpretation of the simulated KBIs in aqueous mixtures, as we discuss in the next section.

Why labile clusters would modulate the random distribution of molecules in a disordered liquid? Indeed, one expects that long range modulation in pair correlations would be due to some global ordering such as the lamellar phase for example. Fig. 13 shows the structure factors $S_{OO}(k)$ (top panel) and $S_{CC}(k)$ (lower panel) for various benzene concentrations. $S_{OO}(k)$ shows a feature that is not seen in previous reports: the presence of a clear pre-peak centered at the ethanol cluster domain size $k_P \approx 0.1$ Å^{-1}, which corresponds to a domain size of $d = 2\pi/k_P \approx 60$ Å^{-1}, which corresponds to about 10 ethanol molecules. We note that this pre-peak position does not change so much between $x = 0.5$ and $x = 0.8$, witnessing the presence of very similar ethanol aggregates in the two cases. A close look at the isolated aggregates shows the presence of rings of 4 Hbonded ethanol molecules. Larger aggregates seem to be made of piles of such smaller elementary aggregates. One notes that neat ethanol has also a pre-peak at $k_P \approx 0.8$ Å^{-1}, which is about $d \approx 7.5$ Å. This would correspond to the size of the elementary ring.

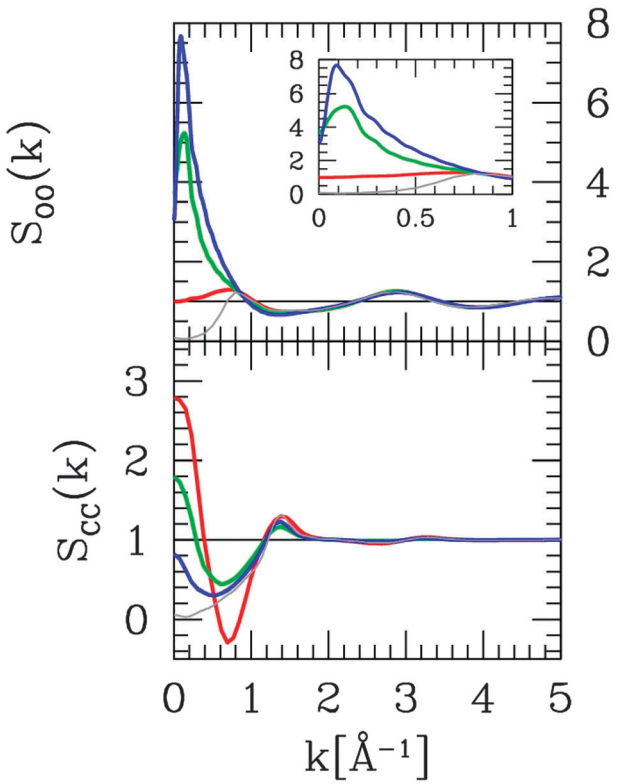

Fig. 13 Structure factors for the benzene–ethanol system: (top) oxygen–oxygen $S_{OO}(q)$ for the benzene concentrations $x = 0$, 0.2, 0.5 and 0.8 with same color code as in Fig. 11; (bottom) benzene carbon–carbon $S_{CC}(q)$ with same color codes as for above, except for $x = 1$ (pure benzene) in thin grey line.

The interpretation of these pre-peak features is that such elementary rings exist in neat ethanol, but they are likely to be very labile. As an inert solvent is progressively added, such clusters acquire a "particle" status, and help forming larger clusters.

The structure factor of benzene equally shows large concentration fluctuations at $k = 0$ when benzene is in low concentration $x = 0.2$. However, $S_{CC}(k)$ does not show a pre-peak. This is clearly related to the absence of specific benzene clusters. This finding illustrates the radical difference between concentration fluctuations and micro-heterogeneity due to specific clusters. In particular, it is seen that at $x = 0.8$, there is very little concentration fluctuations in benzene, despite the presence of large ethanol clusters, most of them being very labile.

Returning to the KBIs in Fig. 8, we observe that these are very similar to those of the methanol–benzene mixtures, for which experimental KBIs are available.[18] The D term is obtained by a fitting procedure through the simulation data based on eqn (5). We empirically found that $D(x) = 1 - x(1 - x)[3.9x^{0.4} + 2.72(\exp(1.8x^{6.5}) - 1)]$ would allow us to reasonably fit all 3 KBIs. The principle in picking this particular expression was to control both the behaviour near $x = 0$ and $x = 1$ through the 2 terms in the brackets.

Fig. 14 shows a comparison of the structural properties $g_{OO}(r)$ and $S_{OO}(k)$ for the ethanol–benzene mixture at $x = 0.8$ for $N = 2048$ particles, but with the TraPPE force field for ethanol.[22] Although some differences are clearly seen, these essentially concern only the long range part of $g_{OO}(r)$ and the small-k parts of $S_{OO}(k)$. Otherwise, the results are indistinguishable from one another.

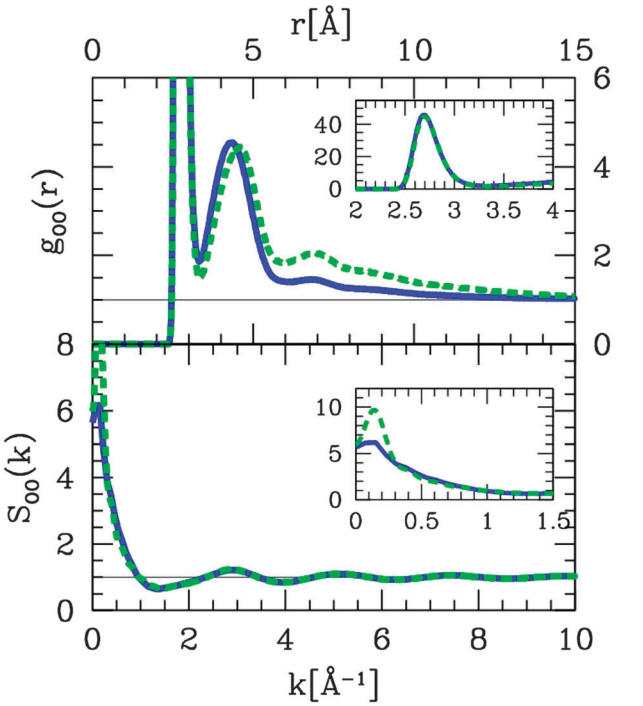

Fig. 14 A comparison of the ethanol oxygen–oxygen correlations, $g_{OO}(r)$ (top panel) and $S_{OO}(k)$ (lower panel) between the OPLS (full line in blue) and TraPPE (dashed lines in green) force fields.

These differences show that the small differences in the force fields seen in Table 1 influence more the cluster properties than the short range ethanol–ethanol correlations. The most important feature is that, despite these difference, the presence of the pre-peak at $k \approx 0.1$ Å is confirmed by both force fields. Indeed, a visual inspection of the snapshots reveals the existence of clear ethanol ''micelles'' in benzene. A more thorough study of the influence of the force fields on the clustering will be reported elsewhere.[26]

4 Discussion

4.1 Interactions, correlations and complexity of disorder

The relationship between site–site interactions $v_{ij}(r)$ and corresponding correlations can be summarized by the exact equation[1,40]

$$g_{ij}(r) = \exp\left[-\frac{v_{ij}(r)}{k_B T} + W_{ij}(r)\right] \quad (13)$$

where $W_{ij}(r) = g_{ij}(r) - 1 - c_{ij}(r) + b_{ij}(r)$, where $c_{ij}(r)$ is the site–site direct correlation function and $b_{ij}(r)$ embeds combinations of high order rank direct correlations. Because of the presence of $g_{ij}(r)$ in the exponential, eqn (13) shows how both the interactions and the correlations control the form of the correlations at short range and also at long range. It is the interplay between these two contributions that seems to control the nature of the disorder in liquid state mixtures.[1,40]

In the case of the benzene–alkane mixtures, even though short range correlations exist, they do not acquire enough persistence to build into aggregate entities that would persist through the system. This is because all the interactions are nearly of the same order of magnitude. Such a mixture is dominated by entropy.

In the case of the acetone–benzene mixtures, the dipole–dipole pairing interactions facilitate the formation of aggregates. However, even in the neat acetone, such aggregates do not produce a pre-peak in the OO structure factor. This is because oxygen sites cannot bind to each other directly, even mediated by the carbon site. So, the presence of elementary dimer clusters is not enough to build up larger clusters, typically because of dipolar frustration in the case of acetone. So acetone–benzene mixtures are strictly limited to showing large concentration fluctuations, which reflects the tendency to form dimers in acetone pockets inside benzene.

In the case of the ethanol–benzene mixtures, ethanol O–H–O Hbonding connectivity is open. This is perhaps because the hydrogen site is very small (or zero diameter in the usual force fields), and allows direct O–O interactions and correlations. As a result, ethanol cluster chains can appear. This is the main reason for a pre-peak in the O–O correlations. Such clusters grow in strength when solvent molecules do not compete with their formation, such as in the benzene–ethanol mixtures.

It is very important to note that no considerations about the cluster life times or kinetics of the clustering are ever taken into consideration here. This is because we only analyze static quantities, so cluster dynamics should not be involved. The presence of a cluster–particle pre-peak is static observation, and the lifetime of the Hbonded clusters does not play any role in this. If one wants to study the dynamics of such clusters, then one should compute the dynamical structure factor $S_{ab}(k,t)$, the so called van Hove function,[1] and study the time evolution of the pre-peak.

4.2 Comparison between aqueous ethanol and benzene–ethanol systems

In previous studies, we studied aqueous mixtures in search of the cluster-prepeaks in $S_{OO}(k)$ but could not find this feature straight out of the simulations, contrary to the fact that we so easily found it in the ethanol–benzene systems. In the case of aqueous mixtures we had to resort to a theoretical model based on the Teubner–Strey structure factor for micro-emulsions, in order to explain the clustering feature of alcohol–water mixtures.[38] We designated by the term ''molecular emulsions'' aqueous mixtures with labile cluster formation, which tend to show a prepeak. By comparing the $S_{OO}(k)$ structure factors of the neat ethanol and neat water, we find that water is missing the clear pre-peak that ethanol and most higher alcohol have.[10] Yet, water is a strong Hbonding liquid. This is perhaps because water does not form any particular specific clusters, unlike alcohols. Nevertheless, we do observe well defined water pockets in water poor aqueous mixtures.[38,41–44] It would seem that the ability of a mixture to exhibit pre-peak is related to the presence of inert methyl sites in the molecules, which steers the formation of specific clusters. Since water lacks such sites, it is more difficult to characterize the existence of a pre-peak within simulations. This point is open for future investigations.

In our previous work on aqueous alcohol,[13,27,38] we noticed that the long range oscillatory nature of the correlation would make the determination of a truly equilibrium value of the KBIs depending on very large system sizes and run lengths in the 10 ns range. In the case of the ethanol–benzene system, we reported differences between 2048 and 16 000 particle systems, but at fixed 4 ns run lengths. Longer runs in both systems show that there is a slow kinetics of the clusters, which affects the statistics of the correlations. These are essentially visible in the long range oscillatory part of $g_{OO}(r)$ and the small-k part of the $S_{OO}(k)$. This issue is again related to the problem of having a full statistics of these labile clusters, which seem to have a life time around several ns. It also seems that system sizes larger than 16 386 particles might be required to obtain a better statistics of the cluster distribution. Despite these numerical issues, the principal point of the present study is that simulation studies of the same type (system size and run length) permit us to observe an unambigous manifestation of the Hbond pre-peak in solvent benzene, while this is more elusive and subject to theoretical extrapolations[38] in the case of solvent water.

It is also interesting to compare ethanol clustering at low ethanol concentrations in both water and benzene. Our previous studies of this clustering for the case of aqueous mixtures for ethanol mole fraction $x = 0.2$ showed that there was some ethanol clustering, but nothing as marked as that shown in Fig. 8. If we pursue the molecular emulsion picture, this

difference could be understood in analogy with direct and inverse micelles. Direct micelles form in a water rich environment, when the surfactant molecules gather their oily tails in the interior of the micelle, whereas inverse micelles form in an oil-rich environment, when the polar heads are buried inside the micelle. In the present case, the size of the fatty tail of ethanol is so small that it is unlikely that direct micelles as such can be formed in a water rich environment. This might explain why we did not see interesting ethanol aggregates. Conversely, in a benzene rich environment, the OH groups of ethanol can group themselves in specific clusters that are stabilized by the strong underlying electrostatic interactions. So, even if we cannot really speak of micelles, the aggregation picture in both solvent seems to follow a pattern connected to the corresponding real micelle systems. There is another point worth mentioning. Water is a more fluctuating environment than benzene: the fluctuations of the Hbonds induce strong energy and density fluctuations. Therefore, it is probably harder to stabilize a micelle-like shape in water than in benzene. This appears to be like a second argument that might explain that we could obtain well defined pre-peaks in $S_{OO}(k)$ in the case of benzene as a solvent.

5 Conclusion

In this work, we have examined how different types of interactions that lead to strong local ordering control the nature of the fluctuations that occur in the midst of the corresponding systems. We show that many such systems lead to near ideal concentration fluctuations when interactions are essentially dominated by excluded volume effects. Interactions that favour long range ordering of one of the components, such as dipolar interactions in the case of acetone, lead to a weak deviation from ideality. Both such systems, we put into the simple disorder category. Stronger interactions such as Hbond interaction, which lead to extended clustering and domain formation, such as various aqueous mixtures, we put them into complex disorder. The microscopic signature of such systems is the slow spatial and temporal fluctuations of the domains, which then induce statistical problems in computer simulation of the corresponding systems. Such problems have been previously reported for *a priori* more complex mixtures, involving in particular water as one of the components. Here we show that mixture of one simple liquid with one complex one, namely benzene–ethanol mixture, can lead to statistical problems similar to that of aqueous mixtures, particularly in the determination of the KBIs. In general, experimental studies of the hydrogen bonding in hydrophobic environments[45,46] or in addition to water[47,48] provide only indirect information about the resulting mesoscopic structures. Self-hydrogen bonding in neat alcohols is a typical example, with cluster formation[49,50] and an associated pre-peak in the structure factor.[51] The important point that we would like to underline here is that, even though on macroscopic spatial and temporal scales, problems related to these heterogeneity might be unnoticed or unnoticeable, for example at the scale of thermodynamical experiments, they nevertheless pose important practical problems at microscopic scales of investigation. The large KBIs seem to be a signature that such difficulties are present. The second point that we underline here is that KBIs are only the $k = 0$ part of the structure factors, and that their small-k part gives a better insight into the microscopic structure of these liquids. Both these point provide a sound justification to the classification of disorder that we provide here. From an ontological point of view, lowering the definition of complex disorder down to systems such as benzene–ethanol permits an unification of complexity to the appearance of microscopic structures that start to depart from mere concentration fluctuations. It is not the presence of complicated molecules, such as in soft matter systems, which induces complex disorder, but the way persistent and non-invasive local structures are formed. By non-invasive we mean not leading to a full phase separation and phase transition. Finally, complex disorder deserves its denomination because, although it would seem that only the time scale of the lifetime of aggregate would matter, it is precisely this time scale that would permit other faster processes, such as chemical reactions, for example, to occur before significant reorganisation of the local environment, is the key factor to distinguish complex disorder from mere random fluctuation induced disorder. To conclude, organized fluctuations *versus* random fluctuations would seem to be the key to complex organisation that occurs in soft matter and biomaterials.

Appendix: derivation of the LP asymptote correction

We give here a somewhat simple and intuitive derivation of the Lebowitz–Percus asymptote correction,[28] for a single component system and mixtures in the constant *NVT* Canonical ensemble.

One component system

We start from the fundamental relation linking fluctuations in the number of particles N to the correlation function and the thermodynamical property the isothermal compressibility $\kappa_T = (1/\rho)(\partial\rho/\partial P)_T$:[1]

$$S(0) = 1 + \rho\tilde{h}(0) = \kappa_T{}^* = \left(\frac{\partial\rho}{\partial\beta P}\right)_T = \rho k_B T \kappa_T = \frac{\langle N^2\rangle - \langle N\rangle^2}{\langle N\rangle} \quad (14)$$

where $h(r) = g(r) - 1$, $S(k) = 1 + \rho\tilde{h}(k)$ is the structure factor defined as in eqn (1), the averages $\langle\cdot\rangle$ are computed in the Grand Canonical Ensemble (GCE).[1] The reduced isothermal compressibilty is defined as $\kappa_T{}^* = \kappa_T/\kappa_T^{(0)}$ where $\kappa_T^{(0)} = 1/(\rho k_B T)$ is the compressibility of the ideal gas. We note that $\tilde{h}(0) = \int d\vec{r}(g(r) - 1)$ is the integral of the radial distribution function $g(r)$.

If we write that the exact $g(r)$, as obtained in the GCE, is the sum of $g_{NVT}(r)$ evaluated in finite size constant *NVT* ensemble simulations with a function $t(r)$ which is essentially a smooth step function which is zero inside the core and raises to the LP

correction value ε_{LP} around the first peaks and stays at that value until infinity:

$$g(r) = g_{NVT}(r) + t(r)$$

which amounts to say that

$$\lim_{r\to\infty} g_{NVT}(r) = 1 - \varepsilon_{LP}$$

then we can re-rewrite eqn (1) as

$$\tilde{h}_{NVT}(0) = \frac{\kappa_T{}^*}{\rho} - \frac{1}{\rho} - D$$

where

$$D = \int \mathrm{d}\vec{r} t(r) \approx V\varepsilon_{LP}$$

where V is the volume, and the approximation is obtained by neglecting the small contribution due to the core integration. Using $\rho = N/V$, we rewrite the contribution from the simulations as

$$\tilde{h}_{NVT}(0) = \frac{\kappa_T{}^*}{\rho} - \frac{1}{\rho} - \frac{N\varepsilon_{LP}}{\rho} \quad (15)$$

On the other hand, the integral of the $h_{MD}(r)$ obtained from any N-constant ensembles should be exactly $-1/\rho$ from the definition of the canonical ensemble. The proof is easy for the canonical ensemble.

The definition of the n-body correlation function $\rho^{(n)}(1, 2,\ldots,n)$ (written $\rho^{(n)}(n)$ for a short-hand notation) in a N-constant ensemble is

$$\rho^{(n)}(n) = \frac{N!}{(N-n)!}\frac{1}{Z_N}\int \mathrm{d}(n+1)\ldots\mathrm{d}N \exp[-\beta V(N)]$$

where $Z_N = \int \mathrm{d}1\ldots\mathrm{d}N\exp[-\beta V(N)]$ is the canonical ensemble partition function and $V(N) = V(1, 2,\ldots,N)$ is the total interaction energy in the system. The integral above runs over the remaining $(N - n)$ particles. The n-body distribution function is defined from the n-body correlation function as

$$\rho^{(n)}(n) = g^{(n)}(n)\left[\prod_{i=1}^{n}\rho^{(1)}(i)\right]$$

which measures the deviation from the uncorrelated case at large separation. We see that the normalisation is such that

$$\int \mathrm{d}1\ldots\mathrm{d}n\rho^{(n)}(n) = \frac{N!}{(N-n)!}$$

For the pair correlation function one obtains

$$\int \mathrm{d}1\,\mathrm{d}2\rho^{(2)}(1,2) = N(N-1)$$

It the system is transitionally invariant, then $\rho^{(1)}(1) = \rho$ the bulk density, and the pair correlation function depends on the radial distance between particles $r = r_{12}$. Then we have

$$\rho^{(2)}(1,2) = \rho^{(2)}(r) = \rho^2 g(r)$$

where $g(r) = g^{(2)}(r)$ is the usual pair distribution function. The normalisation condition translates into

$$\rho^2\int \mathrm{d}1\,\mathrm{d}2 g(r) = \rho^2 V\int \mathrm{d}\vec{r} g(r) = N(N-1)$$

which is simply (using $\rho = N/V$)

$$\int \mathrm{d}\vec{r} g(r) = V\left(1 - \frac{1}{N}\right)$$

Using $g(r) = h(r) + 1$, we obtain

$$\int \mathrm{d}\vec{r} h(r) = -\frac{1}{\rho}$$

While this result is exact in the canonical ensemble with the NVT constant, it is not obvious how it holds in the isobaric ensemble with the NPT constant.

This last equality gives us the value of the LP correction from the relation obtained above in eqn (15)

$$\varepsilon_{LP} = \frac{1}{N}\kappa_T{}^* = \frac{1}{N}\left(\frac{\partial\rho}{\partial\beta P}\right)_T$$

From this relation, we obtain the asymptotic form of the LP correction for $g_{NVT}(r)$

$$\lim_{r\to\infty} g_{NVT}(r) = 1 - \frac{\alpha_{LP}}{N}$$

which is the desired final result. We observe that the following thermodynamic relation allows one to make a direct link with the corresponding expression for mixtures:

$$\frac{\partial\beta P}{\partial\rho} = \rho\frac{\partial\beta\mu}{\partial\rho}$$

Mixtures

The same procedure can be followed for a mixture. In the N-fluctuating μVT grand-canonical ensemble, one has from eqn (9) of the Kirkwood–Buff paper[4]

$$B_{ij} = \sqrt{\rho_i\rho_j}\left(\delta_{ij} + \sqrt{\rho_i\rho_j}\tilde{h}_{ij}(0)\right) = \sqrt{\rho_i\rho_j}\tilde{S}_{ij}(0)$$

$$= \frac{\sqrt{\langle N_i\rangle\langle N_j\rangle}}{V}\tilde{S}_{ij}(0)$$

and the inverse matrix to B is the matrix A, such that (eqn (8)) (“cof” designates the cofactor) reads

$$B_{ij} = \frac{\mathrm{cof}\,A_{ij}}{\det A} = \langle N_iN_j\rangle - \langle N_i\rangle\langle N_j\rangle$$

Combining these equations we get the analog of eqn (14)

$$\tilde{S}_{ij}(0) = V\frac{\langle N_iN_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\sqrt{\langle N_i\rangle\langle N_j\rangle}}$$

where we have used $\rho_i = \langle N_i\rangle/V$. We remind the following definitions, in the Grand Canonical ensemble: the total number of particle $N = \sum_i \langle N_i\rangle$, the mole fractions $x_i = \langle N_i\rangle/N$ and the total density $\rho = N/V = \sum_i \rho_i$. We note that there is a factor V difference which should NOT be there. This is because eqn (9) in the KB paper is erroneous to a factor V precisely. Indeed, since A is dimensionless, B is also, and it cannot be defined by

the equation above which shows that B has a factor $1/V$ precisely. So we redefine the B_{ij} as

$$B_{ij} = V\sqrt{\rho_i\rho_j}\left(\delta_{ij} + \sqrt{\rho_i\rho_j}\tilde{h}_{ij}(0)\right) = V\sqrt{\rho_i\rho_j}\tilde{S}_{ij}(0)$$

such that we now get

$$\tilde{S}_{ij}(0) = \frac{\langle N_iN_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\sqrt{\langle N_i\rangle\langle N_j\rangle}}$$

which is the analog of eqn (14) in Section 1. KB also find the following result (eqn (8))

$$A_{ij} = \left(\frac{\partial\beta\mu_i}{\partial N_j}\right)_{TVN_k} = \frac{1}{V}\left(\frac{\partial\beta\mu_i}{\partial\rho_j}\right)_{TVN_k}$$

which suggests that the inverse matrix is

$$B_{ij} = V\left(\frac{\partial\rho_i}{\partial\beta\mu_j}\right)_{TV\mu_k}$$

such that we now get the full relation equivalent to eqn (14)

$$\tilde{S}_{ij}(0) = \frac{B_{ij}}{\sqrt{\langle N_i\rangle\langle N_j\rangle}} = \frac{1}{\sqrt{\rho_i\rho_j}}\left(\frac{\partial\rho_i}{\partial\beta\mu_j}\right)_{TV\mu_k}$$

from which we get

$$\tilde{h}_{ij}(0) = \frac{1}{\rho_i\rho_j}\left(\frac{\partial\rho_i}{\partial\beta\mu_j}\right)_{TV\mu_k} - \frac{\delta_{ij}}{\sqrt{\rho_i\rho_j}} \qquad (16)$$

If we now express the exact $h_{ij}(r)$ in terms of the simulation results as for the case of the 1 component system above:

$$h_{ij}(r) = h_{NVT;ij}(r) + t_{ij}(r)$$

this leads to

$$\tilde{h}_{ij}(0) = \tilde{h}_{NVT;ij}(0) + V\varepsilon_{ij} \qquad (17)$$

with the same assumption as in the one component case that the core part contribution to the integral over the $\delta_{ij}(r)$ function is negligible (α_{ij} being the asymptotical value).

Now we demonstrate the exact limit of $\tilde{h}_{NVT;ij}(0)$ for the canonical ensemble. For a mixture of n-species, the partition function is defined as

$$Z_N = Z_N(N_1\ldots N_n) = \int \mathrm{d}N_1\mathrm{d}N_2\ldots\mathrm{d}N_n \times \exp(-\beta V(N_1, N_2, ..., N_n))$$

where $\mathrm{d}N_k = \mathrm{d}1_k\mathrm{d}2_k\ldots\mathrm{d}N_k$ is the short hand notation for the differential elements running over all the coordinate i_k of molecules of species k. Then, one can define pair correlation functions between 2 species i and j as

$$\rho_{ij}^{(2)}(1,2) = N_i(N_j - \delta_{ij})\frac{1}{Z_N}\int \mathrm{d}N_1\ldots\mathrm{d}N_i^{(-1)}\ldots\mathrm{d}N_j^{(-1)}\ldots\mathrm{d}N_n \times \exp(-\beta V(N_1, N_2, \ldots, N_n))$$

where we have taken care to account for the case where both species are the same with a δ_{ij} and the notation $\mathrm{d}N_k^{(-p)}$ means that the integration for species k runs over all the variables except those p variables that are selected in the correlation function. The pair distribution function is defined as the deviation from uniformity as:

$$\rho_{ij}^{(2)}(1,2) = \rho_i^{(1)}(1)\rho_j^{(1)}(2)g_{ij}(1,2)$$

For an homogeneous, hence translationally invariant mixture, one has $\rho_k^{(1)}(i) = \rho_k = N_k/V$ which is simply the bulk density of species k. In addition, translational invariance implies that the distance dependence $\vec{r}_1$, $\vec{r}_2$ can be replaced by $\vec{r} = \vec{r}_{12}$. If we now integrate over variables 1 and 2, and use the translational invariance, we obtain

$$\int \mathrm{d}1\,\mathrm{d}2\rho_{ij}^{(2)}(1,2) = \rho_i\rho_j V\int \mathrm{d}\vec{r}g_{ij}(r) = N_i(N_j - \delta_{ij})$$

Using $g_{ij}(r) = h_{ij}(r) + 1$ we get:

$$\sqrt{\rho_i\rho_j}\tilde{h}_{NVT;ij}(0) = \sqrt{\rho_i\rho_j}\int \mathrm{d}\vec{r}h_{ij}(r) = -\delta_{ij} \qquad (18)$$

Then, combining eqn (16)–(18), and using the mole fractions defined as $x_i = \langle N_i\rangle/N$, we get

$$\varepsilon_{ij} = \frac{1}{N\rho}\frac{1}{x_ix_j}\left(\frac{\partial\rho_i}{\partial\beta\mu_j}\right)_{TV\mu_k}$$

which gives the results in Section 2.1 eqn (2) and (3).

Acknowledgements

J-B. S. and J. G. thank the Laboratoire de Physique Théorique de la Matière Condensée for allowing the completion of their Master graduation thesis work. A. P. and B. K-L. thank Luc Belloni for privately pointing out incorrections in the form of the LP asymptote correction eqn (3) in previous articles[13,38,42] of the group. This work has been supported in part by the Croatian Science Foundation under the project 4514 "Multi-scale description of meso-scale domain formation and destruction".

References

1 J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, Academic, London, 1986.
2 P. M. Chaikin and T. C. Lubensky, *Principles of condensed matter physics*, Cambridge University Press, 1995.
3 S. Rowlinson and F. L. Swinton, *Liquids and Liquids mixtures*, Butterworths Scientific, London, 1982.
4 J. G. Kirkwood and F. Buff, *J. Chem. Phys.*, 1950, **19**, 774.
5 E. Matteoli and L. Lepori, *J. Chem. Phys.*, 1984, **80**, 2856.
6 P. Ball, *Nature*, 2008, **452**, 291.
7 For an insightful review about water science, visit http://www1.lsbu.ac.uk/water/.
8 M. P. Allen and D. J. Tildesley, *Computer simulations of liquids*, Oxford University Press, 1987.
9 A. K. Karmakar, P. S. R. Krishna and R. N. Joarder, *J. Phys. Chem.*, 1995, **99**, 16501.

10 A. K. Karmakar, P. S. R. Krishna and R. N. Joarder, *Phys. Lett. A*, 1999, **253**, 207.
11 C. J. Pinga and J. Waser, *J. Chem. Phys.*, 1968, **48**, 3016.
12 A. R. Abdel Hamid, R. Lefort, Y. Lechaux, A. Moreac, A. Ghoulfi, C. Alba-Simionesco and D. Morineau, *J. Phys. Chem.*, 2013, **117**, 10221.
13 A. Perera, B. Kežić, F. Sokolić and L. Zoranić, in *Molecular Dynamics*, ed. L. Wang, InTech, Rijeka, 2012, vol. 2.
14 M. Bae Gee and P. E. Smith, *J. Chem. Phys.*, 2009, **131**, 165101.
15 C. J. Wormald and C. J. Sowden, *J. Chem. Thermodyn.*, 1997, **29**, 1223.
16 I. Nagata, K. Tamura, N. Kishi and K. Tada, *Fluid Phase Equilib.*, 1997, **135**, 227.
17 V. M. Byakov, S. V. Stepanov, P. M. Zorkii, L. V. Lanshina and O. P. Stepanova, *Russ. J. Phys. Chem.*, 2007, **81**, 638.
18 E. A. Ploetz and P. E. Smith, *Phys. Chem. Chem. Phys.*, 2014, **13**, 18154.
19 Y. Adachi and K. Nakanishi, *Fluid Phase Equilib.*, 1993, **83**, 69.
20 G. Larsen, Z. K. Ismael, B. Herrero and R. D. Parra, *J. Phys. Chem. A*, 1998, **102**, 4734.
21 W. L. Jorgensen, J. D. Madura and C. J. Swenson, *J. Am. Chem. Soc.*, 1984, **106**, 6638; W. L. Jorgensen, *J. Phys. Chem.*, 1986, **90**, 1276.
22 B. Chen, J. J. Potoff and J. I. Siepmann, *J. Phys. Chem.*, 2001, **105**, 3093.
23 D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701.
24 J. M. Martínez and L. Martínez, *J. Comput. Chem.*, 2003, **24**(7), 819–825; L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, 2009, **30**(13), 2157–2164.
25 E. Oran Brigham, *The Fast Fourier Transform*, Prentice Hall, 1988.
26 M. Požar, L. Zoranić, M. Mijaković, B. Kežić-Lovrincević, F. Sokolić and A. Perera, in preparation.
27 A. Perera, L. Zoranić, F. Sokolić and R. Mazighi, *J. Mol. Liq.*, 2011, **159**, 52.
28 J. L. Lebowitz and J. K. Percus, *Phys. Rev.*, 1961, **122**, 1675.
29 L. Dougan, S. P. Bates, R. Hargreaves, J. P. Fox, J. Crain, J. L. Finney, V. Reat and A. K. Soper, *J. Chem. Phys.*, 2004, **121**, 6456.
30 J. T. Towey, A. K. Soper and L. Dougan, *Faraday Discuss.*, 2013, **167**, 159.
31 L. Zoranic, R. Mazighi, F. Sokolic and A. Perera, *J. Chem. Phys.*, 2009, **130**, 124315.
32 S. Komura, *J. Phys.: Condens. Matter*, 2007, **19**, 463101.
33 R. Nagarajan, *Langmuir*, 2000, **16**, 64000.
34 J. Hao and T. Zemb, *Curr. Opin. Colloid Interface Sci.*, 2007, **12**, 129.
35 A. Perera, F. Sokolic, L. Almasy and Y. Koga, *J. Chem. Phys.*, 2006, **124**, 124515.
36 M. Musso, M. G. Giorgini, H. Torii, R. Dorka, D. Schiel, A. Asenbaum, D. Keutel and K.-L. Oehme, *J. Mol. Liq.*, 2006, **125**, 115.
37 A. Perera, in *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering and Biophysics*, ed. P. E. Smith, J. P. O'Connell and E. Matteoli, CRC Press Taylor and Francis, 2012, ISBN 9781439899229.
38 B. Kežić and A. Perera, *J. Chem. Phys.*, 2012, **137**, 014501.
39 M. Ferrario, M. Haughney, I. R. McDonald and M. L. Klein, *J. Chem. Phys.*, 1990, **93**, 5156.
40 C. Caccamo, *Phys. Rep.*, 1996, **274**, 1.
41 B. Kežić and A. Perera, *J. Chem. Phys.*, 2012, **137**, 134502.
42 A. Perera and B. Kežić, *Faraday Discuss.*, 2013, **167**, 145.
43 A. Perera, R. Mazighi and B. Kežić, *J. Chem. Phys.*, 2012, **136**, 174516.
44 B. Kežić, R. Mazighi and A. Perera, *Physica A*, 2013, **392**, 567.
45 P. Sassi, F. Palombo, R. S. Cataliotti, M. Paolantoni and A. Morresi, *J. Phys. Chem. A*, 2007, **111**, 6020.
46 N. Asprion, H. Hasse and G. Maurer, *Fluid Phase Equilib.*, 2003, **208**, 23.
47 P. N. Perera, K. R. Fega and C. Lawrence, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 12230.
48 S. Young and E. C. Fortey, *J. Chem. Soc.*, 1902, **81**, 739.
49 J.-H. Guo, Y. Luo, A. Auggustsson, S. Kashtanov, J.-E. Rubensson, D. K. Shuh, H. Agren and J. Nordgren, *Phys. Rev. Lett.*, 2003, **91**, 157401.
50 R. Ludwig, *ChemPhysChem*, 2005, **6**, 1369.
51 L. Zoranic, F. Sokolic and A. Perera, *J. Chem. Phys.*, 2007, **127**, 024502.

