Evaluation of Kirkwood-Buff integrals via finite size scaling: a large scale molecular dynamics study

W Dednam[1,2], A E Botha[2]

1 Departamento de Fisica Aplicada, Universidad de Alicante, San Vicente del Raspeig, E-03690 Alicante, Spain
2 Department of Physics, Science Campus, University of South Africa, Private Bag X6, Florida Park 1710, South Africa

E-mail: `bothaae@unisa.ac.za`

Abstract. Solvation of bio-molecules in water is severely affected by the presence of co-solvent within the hydration shell of the solute structure. Furthermore, since solute molecules can range from small molecules, such as methane, to very large protein structures, it is imperative to understand the detailed structure-function relationship on the microscopic level. For example, it is useful know the conformational transitions that occur in protein structures. Although such an understanding can be obtained through large-scale molecular dynamic simulations, it is often the case that such simulations would require excessively large simulation times. In this context, Kirkwood-Buff theory, which connects the microscopic pair-wise molecular distributions to global thermodynamic properties, together with the recently developed technique, called finite size scaling, may provide a better method to reduce system sizes, and hence also the computational times. In this paper, we present molecular dynamics trial simulations of biologically relevant low-concentration solvents, solvated by aqueous co-solvent solutions. In particular we compare two different methods of calculating the relevant Kirkwood-Buff integrals. The first (traditional) method computes running integrals over the radial distribution functions, which must be obtained from large system-size NVT or NpT simulations. The second, newer method, employs finite size scaling to obtain the Kirkwood-Buff integrals directly by counting the particle number fluctuations in small, open sub-volumes embedded within a larger reservoir that can be well approximated by a much smaller simulation cell. In agreement with previous studies, which made a similar comparison for aqueous co-solvent solutions, without the additional solvent, we conclude that the finite size scaling method is also applicable to the present case, since it can produce computationally more efficient results which are equivalent to the more costly radial distribution function method.

1. Introduction

The solvation thermodynamics of solutes in water are extremely sensitive to the presence and concentration of co-solvents such as alcohols, osmolytes and salts [1, 2, 3, 4]. Their presence in the hydration shell of the solute in an aqueous co-solvent solution usually determines how well the solute is solvated by the solution [1]. In general, the free energy of solvation depends on the relative proportions of co-solvent and water in the mixture, and the more negative the free energy of solvation, the greater the extent to which the solute is solvated by the solution.

Understanding the effects of co-solvents on the solvation of solutes in aqueous co-solvent solutions, via calculations of the solvation free energy, for instance, may shed light on why proteins tend to unfold or denature in the presence of alcohols and urea [3]. It may also help

to explain the salting in and out effects on hydrophobic solutes, due to favourable interactions between the solute and large monovalent ions of low charge density on the one hand, and unfavourable interactions between the solute and small ions of high charge density, on the other [2, 4].

In a recent study, Ganguly and Van der Vegt [5] have compared the efficiency of the so-called finite-size scaling method, first introduced by Schnell et al. [6], to the more traditional method of calculating the Kirkwood-Buff integrals (KBIs) via the radial distribution functions. Ganguly and Van der Vegt [5] consider the full range of concentrations of methanol-water and urea-water systems and conclude that the finite size scaling method (also called the fluctuation or sub-box method, see Refs. [7, 8, 5]) can provide accurate KBIs for system sizes as small as 2000 molecules, which is about a factor of ten smaller than the largest system size used in the present work, which is the minimal size required for the accurate computation of the radial distribution functions in the limit of large distances [9].

In this paper we further explore the validity of the two, above-mentioned methods by performing molecular dynamics (MD) simulations of another very important test case; namely, that of the solvation of methane (a hydrophobic solute) in a methanol (co-solvent) and water (solvent) mixture. In agreement with the previous results [5], which were obtained for for aqueous co-solvent solutions only, our results demonstrate that the finite size scaling method may also be used to improve the efficiency in aqueous co-solvent systems with the addition of a solute, such as methane.

2. Theory
2.1. Kirkwood-Buff theory of solutions
Macroscopic thermodynamic properties of solute-water-co-solvent systems can be directly obtained from microscopic molecular distributions. More specifically, the Kirkwood-Buff theory of solutions [10, 11] connects fluctuations in the grand canonical ensemble to macroscopic thermodynamic properties through the so-called Kirkwood-Buff integrals (KBIs) between components i and j of the solution

$$G_{ij} = 4\pi \int_0^\infty \left(g_{ij}^{\mu VT}(r) - 1 \right) r^2 \mathrm{d}r \tag{1}$$

$$= V \left(\frac{\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - \frac{\delta_{ij}}{\langle N_j \rangle} \right) \tag{2}$$

where $\langle \ \rangle$ denotes the time and ensemble average, V is the volume of the simulation domain, N_i is the number of particles of component i, and $g_{ij}^{\mu VT}(r)$ is the radial distribution function in the grand canonical (μVT) ensemble. G_{ij}, the KBI between component i and j, is a local quantity that measures the deviation of the intermolecular distribution from that of a random one, i.e., an ideal gas. Hence, it is also a measure of the affinity of these two components for each other in the solution environment.

The macroscopic quantity that is of interest in this work is the free energy of solvation of the solute in the methanol-water mixture. This quantity can be expressed in terms of the G_{ij} and the solution component number densities ρ_i as

$$\left(\frac{\partial \Delta \mathrm{G_c}}{\partial \rho_c} \right)_{p,T} = \frac{-RT \left(G_{cc} - G_{cw} \right)}{1 + \rho_c \left(G_{cc} - G_{cw} \right)} \tag{3}$$

for a binary system of water (w) and co-solvent (c), and as

$$\left(\frac{\partial \Delta \mathrm{G_s}}{\partial x_c} \right)_{p,T} = \lim_{\rho_s \to 0} \frac{RT \left(\rho_w + \rho_c \right)^2}{\eta} \left(G_{sw} - G_{sc} \right) \tag{4}$$

for a ternary system in the limit of infinite dilution of the solute [10]. Here ΔG_s is the free energy of solvation of the solute, x_c is the co-solvent mole fraction, R is the universal gas constant, T the temperature, and $\eta = \rho_w + \rho_c + \rho_w \rho_c (G_{ww} + G_{cc} - 2G_{cw})$.

This paper outlines two independent methods for computing the right-hand side of Eq. (1) and Eq. (2) and compares the results obtained from each method. The first method involves generating the $g_{ij}^{\mu V T}(r)$ at every concentration by MD simulation of a system which is large enough to permit the accurate evaluation of the integral on the right-hand side of Eq. (1). The second method, that of finite size scaling, permits the evaluation of the right hand side of Eq. (2), as described in the next section.

2.2. Finite Size Scaling

The KBI between component i and j for a given methanol concentration can be calculated efficiently by an innovative particle counting and extrapolation scheme referred to as finite size scaling [6, 7, 8]. This scheme involves computing the KBI by extrapolation to a system of theoretically infinite size which produces more accurate values for G_{ij}, even at relatively small system sizes. The first step involves repeating a count of the number of particles of type i and j that fall within a certain sub-volume, also called a volumetric pixel (voxel). A large number of voxels of a fixed volume are randomly positioned within the simulation domain, without touching the boundary. By counting the numbers of particles in each voxel at each time step, the averages required in Eq. (2) are obtained. This process is repeated for a range of different voxel volumes. In theory, the ideal voxel volume should be larger than the correlation length of the system, and much smaller than the overall volume of the simulation domain, so that the simulation domain can still act as a particle reservoir to ensure that the ensemble of voxels do approximate a grand canonical ensemble. This means that there is an optimal range of voxel volumes for which the KBI will scale linearly with $1/L$, where L is a linear dimension of the voxel (in the case of a sphere, L would be the diameter). In practice, a good approximation to a grand canonical ensemble can be obtained by using a simulation domain with approximately 20000 molecules or more. (See, for example, Fig. 2 of Ref. [9].)

For a given concentration, a plot of G_{ij} as a function of $1/L$ will show the expected linear relationship over the optimal range of voxel volumes. The KBI value corresponding to a (theoretically) infinite system size may then be obtained by linearly extrapolating the value of G_{ij} back to the ordinate ($1/L = 0$) of the graph, where L is infinite. The KBIs thus determined can then be compared to those obtained directly from Eq. (1).

3. Methodology

Five mixtures of methanol and water were prepared with the concentrations of methanol given by $x_c = 0.00, 0.24, 0.50, 0.75$ and 1.00 mole fraction. The solvent co-solvent system always contained a total of 20000 molecules. An additional 250 methane molecules were added as solute.

In order to simulate the mixtures, the GROMACS molecular dynamics simulation package was used [12]. The methanol and water molecules were modeled using the GROMOS43a1 and SPC water force fields, respectively. All atom simulations were carried out in the NpT ensemble with pressure maintained by the Berendsen Barostat at a pressure of 1 atm and a coupling time of 0.5 ps [13]. The integration time step used was 1 fs for all concentrations. All the simulations were allowed to equilibrate for 15 ns, after which the trajectory was extended by a further 15 ns. Only the last 15 ns of the trajectories were used to calculate the KBIs. The electrostatic interactions in the all atom simulations were handled by means of the particle mesh Ewald.

Calculations of the KBIs, based on the particle fluctuation method (Eq. (2)), were implemented in the Python programming language. Several Python scripts were developed to extract the coordinates of each atom from the system trajectories and then to count the number of each atomic species within the voxels.

4. Simulation results

Initially, to test the water and methanol force fields, a detailed comparison of the computed KBIs was made with the experimentally determined KBIs from Refs. [14, 15]. Figure 1 shows the KBIs, as calculated from the RDFs (solid lines) for a 50% methanol-water mixture. The experimental values are given by the horizontal dashed and dotted lines for EXP 1 (Ref. [14]) and EXP 2 (Ref. [15]), respectively. In this simulation there were 10000 methanol and 10000 water molecules. Figure 1 shows that the KBIs reach a reasonably well-defined plateau beyond
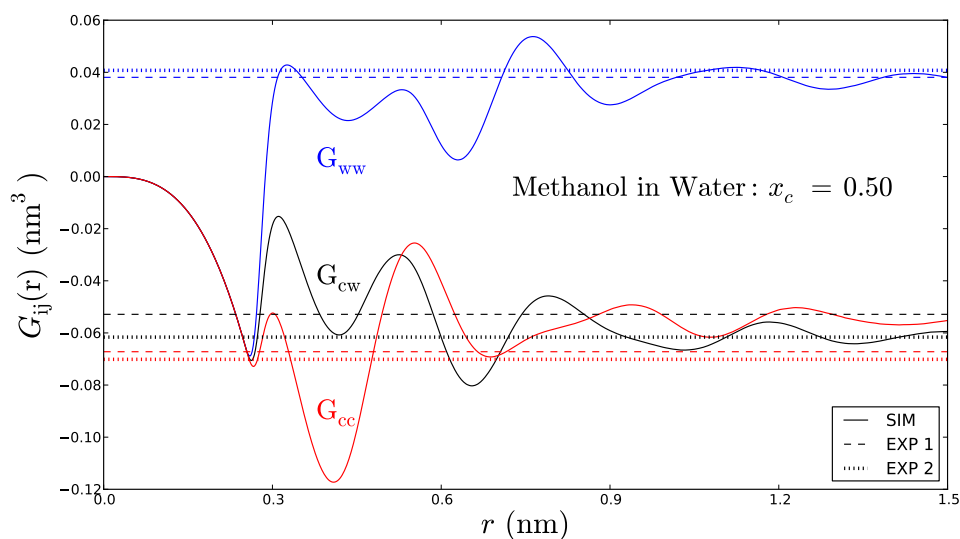

Figure 1. (Color online) The Kirkwood-Buff Integrals (KBIs) calculated via the radial distribution functions (solid curves) compared to the experimental values in references [14] (EXP 1) and [15] (EXP 2).

$r = 1.5\,\text{nm}$, which is a good indication that the chosen system size for this work (20000 molecules in total) is sufficiently large. We arrived at this size by starting with a smaller system size and then doubling the system size until convergence in the KBIs was observed. Thus our system is sufficiently large to approximate the thermodynamic limit. In spite of this, however, we could not obtain agreement (within the experimental uncertainties) for concentrations of methanol below about 15% and we attribute this discrepancy to the well-known inaccuracies in the force fields employed here [16, 17, 18]. For the present work, in which our main aim is to compare the KBIs obtained from the same force field via two different methods, such inaccuracies in the force fields should not matter, and in view of the difficulties associated with re-parameterization of the force fields [18], we have chosen to ignore this problem in the present work.

In Fig. 2 we illustrate how the fluctuation method was used to obtain the KBIs via Eq. (2). In Fig. 2 the diameter L of the sample sub-volume (in this case a sphere) was varied from about half the total simulation box size, down to one twentieth of the simulation box size. As expected from theory [8], the KBI value (in this case for water-water) scales linearly over a certain range of sample system voxels. The KBI for a system of infinite size is therefore well approximated for each concentration by extrapolating to best line fit to infinite L, i.e. to the intercept corresponding to $1/L = 0$. Figures 5 (a) to (e) show comparisons of the KBIs obtained via the fluctuation method (Eq. (2)) and the radial distribution functions (Eq. (1)) for five different concentrations of the methane, methanol and water system. In Fig 5 (a), which is for pure water, the agreement between the two methods is satisfactory. Even though the system
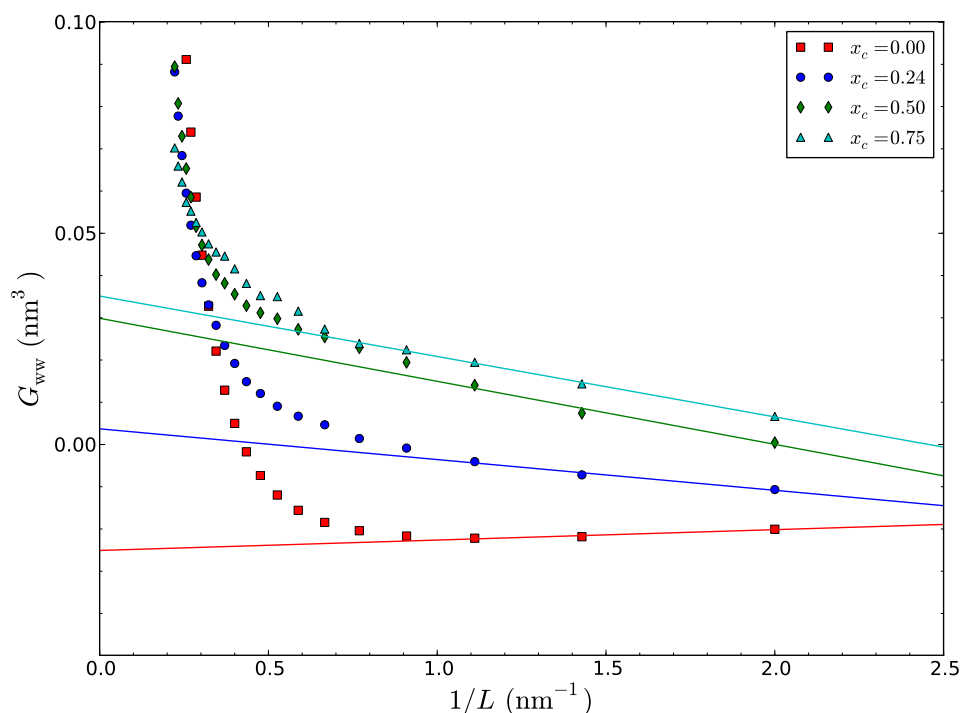
Figure 2. (Color online) Illustration of the method of finite size scaling for four different concentrations of co-solvent for the methane (250 molecules), methanol (co-solvent) and water system.

contains only 250 methane molecules, the RDF method still produces a KBI which shows a well-defined plateau, which indicated that in the calculation of the KBI, via the RDF method, the 15 ns trajectory was sufficiently long for the purposes of these calculations. In Fig. 5 (b) the two methods do not agree very well and the reason for this is not clear at present. One possibility could be that the inaccuracies in the methanol force field at lower concentrations produce unphysical clustering. The resolution of such clustering effects via the fluctuation method may require the use of much longer system trajectories (longer than the 15 ns used here).

5. Discussion and conclusion

We have developed a new, biologically relevant test case in order to compare two important methods for the calculation of Kirkwood-Buff integrals from molecular dynamic simulations. Specifically, we have compared the two methods for simulations of the hydration of methane (a hydrophobic solute) in an aqueous co-solvent solution. Unlike the traditional method, which requires the computation of radial distribution functions in the limit of large distances, the main advantage of the finite-size scaling method is that it can make use of relatively small systems to obtain results that are comparable to those of larger systems that would require substantially longer simulation times. This advantage is due to the analytical finite-size scaling relation [6], which allows the results from successively smaller sub-boxes to be extrapolated to systems which approach the thermodynamic limit.

   We note that the finite size scaling method which we have employed here made use of cubic

sub-volumes to determine the particle fluctuations. This geometry becomes problematic when the linear dimensions of the sub-volume voxel are on the order of the sizes of the molecules themselves [8]. These so-called nook and corner effects can be avoided by using spherical sub-volumes instead [6, 7, 8]. Although we have not done so in the present work, we thus expect that the agreement between the two methods discussed here could be improved by using spherical sub-volumes instead of cubes. Finally we note that problems with the employed force fields
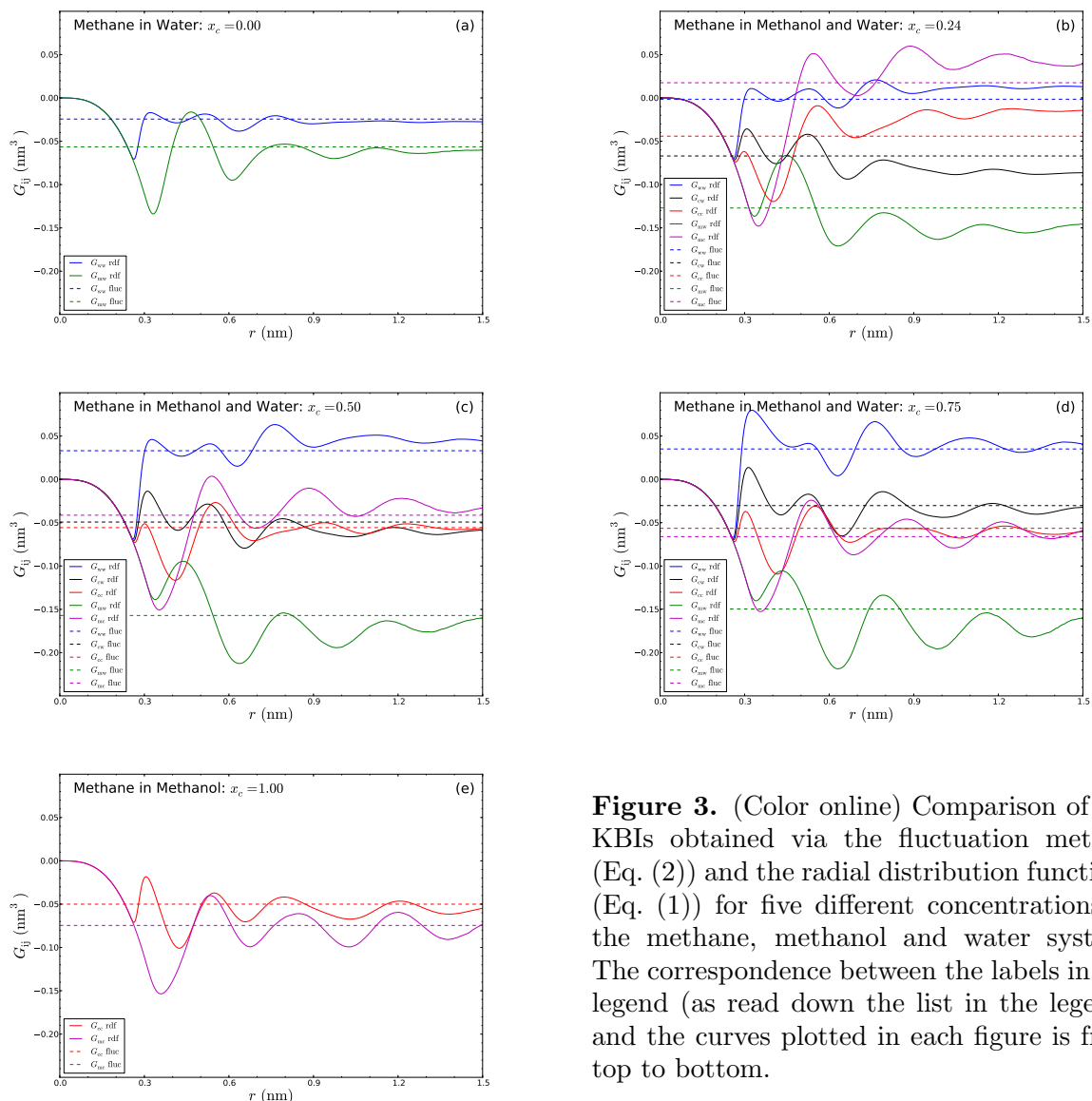

Figure 3. (Color online) Comparison of the KBIs obtained via the fluctuation method (Eq. (2)) and the radial distribution functions (Eq. (1)) for five different concentrations of the methane, methanol and water system. The correspondence between the labels in the legend (as read down the list in the legend) and the curves plotted in each figure is from top to bottom.

have prevented us from making a meaningful comparison of the solvation free energies with the experimental results. The difficulties associated with obtaining force field parameters that are applicable over a wide range of concentrations and to a variety of different systems are non-trivial [16, 17, 18], and since chemical potentials (and hence solvation free energies) are very sensitive to force field parameters, further refinement of the force fields would be required to facilitate a quantitative comparison of these quantities obtained from MD simulations and experimental results. However, the present work clearly demonstrates that finite size scaling can also be used

in the new context of an aqueous co-solvent solution with the addition of a solute, to reduce simulation times.

Acknowledgments

References

[1] Roccatano D, Colombo G, Fioroni M and Mark A E 2002 *Proceedings of the National Acadamy of Sciences* **99** 12179
[2] Sadeghi R and Jahani F 2012 *J. Phys. Chem. B* **116** 5234
[3] Bennion B J and Daggett V 2003 *Proceedings of the National Acadamy of Sciences* **100** 5142
[4] van der Vegt N F A and van Gunsteren W F 2004 *J. Phys. Chem. B* **108** 1056
[5] Ganguly P and van der Vegt N F A 2013 *J. Chem. Theory Comput.* **9** 1347
[6] Schnell S K, Liu X, Simon J M, Bardow A, Bedeaux D, Vlugt T J H and Kjelstrup S 2011 *J. Phys. Chem. B* **115** 10911
[7] Liu X, Schnell S K, Simon J M, Bedeaux D, Kjelstrup S, Bardow A and Vlugt T J H 2011 *J. Phys. Chem. B* **115** 12921
[8] Schnell S K, Vlugt T J H, Simon J M, Bedeaux D and Kjelstrup S 2012 *Molecular Physics* **110** 1069
[9] Mukherji D, van der Vegt N F A, Kremer K and Site L D 2012 *J. Chem. Theory Comput.* **8** 375
[10] Ben-Naim A 2006 *Molecular Theory of Solutions* (New York: Oxford University Press)
[11] Kirkwood J G and Buff F P J 1951 *Chem. Phys.* **19** 774
[12] Lindahl E, Hess B and van der Spoel D 2001 *J. Mol. Mod.* **7** 306
[13] Berendsen H J C, Postma J P M, van Gunsteren W F, Dinola A and Haak J R 1984 *J. Phys. Chem.* **81** 3684
[14] Perera A, Sokolić F, Almásy L and Koga Y 2006 *J. Phys. Chem.* **124** 124515
[15] Matteoli E and Lepori L 1984 *J. Chem. Phys.* **80** 2856
[16] Weerasinghe S and Smith P E 2003 *J. Chem. Phys.* **118** 5901
[17] Weerasinghe S and Smith P E 2005 *J. Phys. Chem. B* **109** 15080
[18] Hawlicka E and Rybicki M 2012 *Molecular Dynamics - Theoretical Developments and Applications in Nanotechnology and Energy* ed Wang L (Rijeka: InTech) p 399