On the calculation of solvation free energy from Kirkwood-Buff integrals: A large scale molecular dynamics study

Wynand Dednam and André E. Botha

Department of Physics, University of South Africa, P.O. Box 392, Pretoria, South Africa

E-mail. bothaae@unisa.ac.za

Abstract. Solvation of (bio)molecules in water is severely effected 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 large protein structures, it is imperative to understand the detailed structure-function relationship on the microscopic level. Although such an understanding can be obtained through molecular dynamic simulations, excessively large system sizes and simulation times are required in order to obtain meaningful results. In this context, Kirkwood-Buff (KB) 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 method to reduce computational times. In this paper, we present a molecular dynamics test simulation to calculate and compare the KB integrals, for the solvation of methane in methanol in water, calculated via two different methods: via the radial distribution functions and via the fluctuation method. In the results reported here we demonstrate that the latter method can produce equivalent results by using a relatively small system size. In future work we thus intend using the fluctuation method and finite size scaling to study the conformational transitions of large (bio)macromolecules.

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 [1-4]. Their presence in the hydration shell of the solute in aqueous solution usually determines how well the solute is solvated by the solution [1]. In other words, 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 mixture.

Understanding the effects of co-solvents on the solvation of solutes in aqueous solutions and on the magnitude of the solvation free energy sheds light on, for instance, why proteins tend to unfold or "denature" in the presence of alcohols and urea [3], and 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].

This paper presents the results of molecular dynamics (MD) simulations of an important initial test, namely the solvation of methane (hydrophobic solute) in a methanol (co-solvent) and water (solvent) mix. Our results demonstrate that by applying the method of finite size scaling [5-7] to relatively small systems (which are computationally less expensive to simulate), results equivalent to much larger systems can be obtained.

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 [8-9] 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:

$$G_{ij} = V\left[\frac{\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - \frac{\delta_{ij}}{\langle N_j \rangle} \right] \tag{1}$$

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

where quantities in $\langle \cdot \rangle$ represent averages in the grand canonical ensemble, V is the volume of the simulation domain, N_i is the number of particles in the simulation domain and $g_{ij}^{\mu VT}(r)$ is the radial distribution function in the grand canonical (μVT) ensemble. G_{ij}, the Kirkwood-Buff integral 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 of interest for 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 solution component number densities ρ_i as [8]

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

for a binary system of water and cosolvent, and as [8]

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

for a ternary system in the limit of infinite dilution of the solute.

Here ΔG_s is the free energy of solvation of the solute, x_c is co-solvent mole fraction, R is the universal gas constant, T the temperature, and $\eta = \rho_w + \rho_c + \rho_w \rho_c (G_{ww} + G_{cc} - 2G_{cw})$ is known as the preferential solvation factor [reference].

This paper outlines two independent methods for computing eq. (3) that permits a comparison of the results obtained and a means by which to assess their robustness. The first involves generating the $g_{ij}^{\mu VT}(r)$ at every concentration by simulation. This, in turn, enables numerical evaluation of the second expression in eq. (1) and hence the G_{ij} needed for eq. (3). The second method, Finite Size Scaling, is described in the section that follows.

2.2. Finite Size Scaling

The KBI between component i and j for a given methanol concentration can also be calculated by an innovative particle counting and extrapolation scheme referred to as finite size scaling [5-7]. This involves computing KBIs for a system of theoretically infinite size which ensures that an accurate value for the G_{ij} at the concentration in question is obtained. The first step involves repeating a count of the number of particles of type i and j that fall within a sphere, randomly positioned within the simulation domain but not touching its boundaries, and then taking the average of these quantities by dividing by the number of times counted. This process is repeated using spheres of increasing radius of the order of the correlation length (between 0.2 and 4.0 nm, roughly the range over which the radial distribution function for this system size exhibits large fluctuations before settling down to its asymptotic value of 1). The results of this process and eq. (1) can then be used to calculate the G_{ij} for every sphere size. This is possible because the numbers of molecules counted in this way approximate the average quantities in Eq. (1) in the grand canonical ensemble for a reasonably large system size, which could be for as few as 20000 molecules [10].

For the concentration in question, plotting the G_{ij} divided by the sphere volumes versus the inverse of the sphere radii permits an estimation of the KBI for a sphere of infinite radius. In practice this is achieved by carrying out a linear interpolation of the data nearest to the abscissa and extrapolating to the ordinate where the radius is infinite.

The KBIs thus determined can then be used to evaluate Eq. (2) for every concentration, and the results compared with those obtained by the method outlined in the previous section.


3. Methodology

Five mixtures of methanol and water were prepared with the concentrations of methanol given by x_c = 0, 0.24, 0.5, 0.75 and 1 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 [11]. 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 [12]. 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 15ns. Only the last 15ns of the trajectories were used to calculate the Kirkwood Buff integrals. The electrostatic interactions in the all atom simulations were handled by means of the particle mesh Ewald.

Calculations of the KBI, based on the particle fluctuation method (Eq. 1), 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 each atomic species within small sub-volumes (cubes of side length L) of the simulation box (of side length L_t). In order to speed up these calculations these python scripts were parallelized to make use of 8 CPUs simultaneously, by using the multiprocessing module. This produced a fourfold increase in speed, which allowed the calculations to be run overnight, rather than for two to three days.

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. [13-14]. 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. [13]) and EXP 2 (Ref. [14]) respectively. In this simulation there were 10000 methanol and 10000 water molecules.
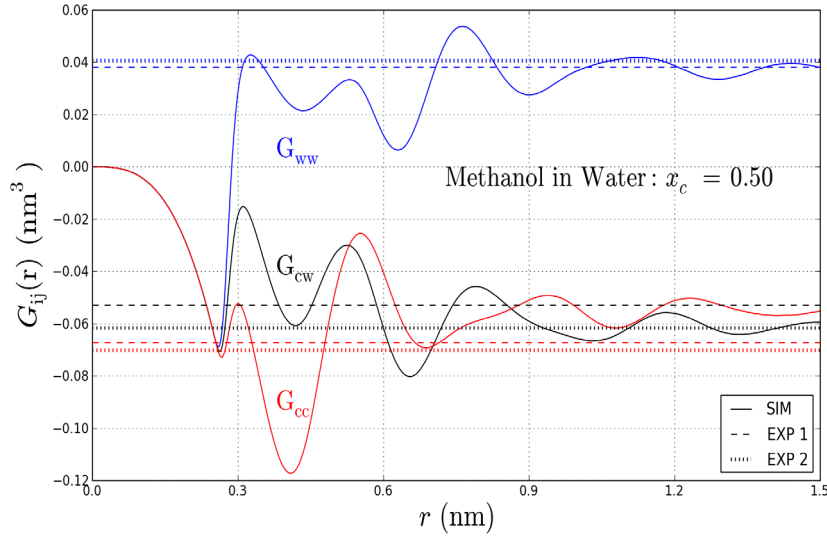
Figure 1: The Kirkwood Buff Integrals (KBIs) calculated via the radial distribution functions (solid curves) compared to the experimental values in references [13] (EXP 1) and [14] (EXP 2).

Figure 1 shows that he KBI reach a reasonably well-defined plateau beyond r=1.5 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 [15-17]. 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 [17], 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. (1). In Fig. 2 the side length L of the sample volume (in this case a cube) was varied from about half the total simulation box size, down to one twentieth of the simulation box size. As expected from theory [7], the KBI value (in this case for water-water) scales linearly over a certain range of sample system volumes. 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.
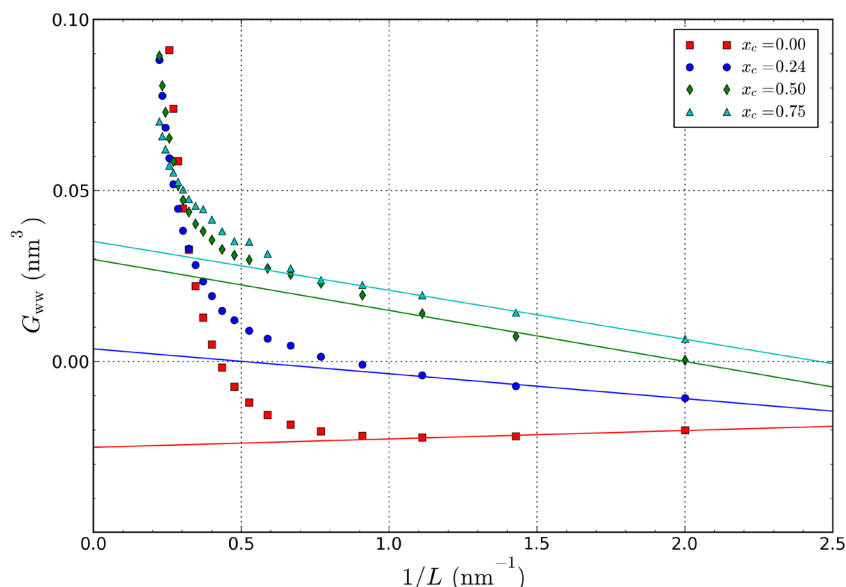
Figure 2: Illustration of the method of finite size scaling for four different concentration of co-solvent for the methane (250 molecules), methanol (co-solvent) and water system.

Figures 3 (a) to (e) show comparisons of the KBIs obtained via the fluctuation method (Eq. 1) and the radial distribution functions (Eq. 2) for five different concentrations of the methane, methanol and water system. In Fig 3 (a), which is for pure water, the agreement between the two methods is satisfactory. Even though the 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 15ns trajectory was sufficiently long for the purposes of this calculations. In Fig. 3 (b) the two methods do no 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 15ns used here).

5. Discussion and conclusion

We have developed a test case in order to compare two methods for calculating KBIs from simulations of the hydration of Methane (hydrophobic solute) in a water-co-solvent mixture. Unlike the traditional method, which makes use of the RDF, the main advantage of the newer fluctuation method is that it can make use of relatively small systems to obtain results that are comparable to those which would require substantially longer simulation times. This advantage is due to the finite size scaling method, which allows the results from finite size systems to be extrapolated to systems which approach the thermodynamic limit.

The fluctuation method we employed made use of cubic sub-volumes to determine the particle fluctuations. This geometry becomes problematic when the linear dimensions of the sub-volume are on the order of the sizes of the molecules themselves [7]. These so-called nook and corner effects can be avoided by using spherical sub-volumes instead [5-7]. 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 subvolumes instead of cubes.
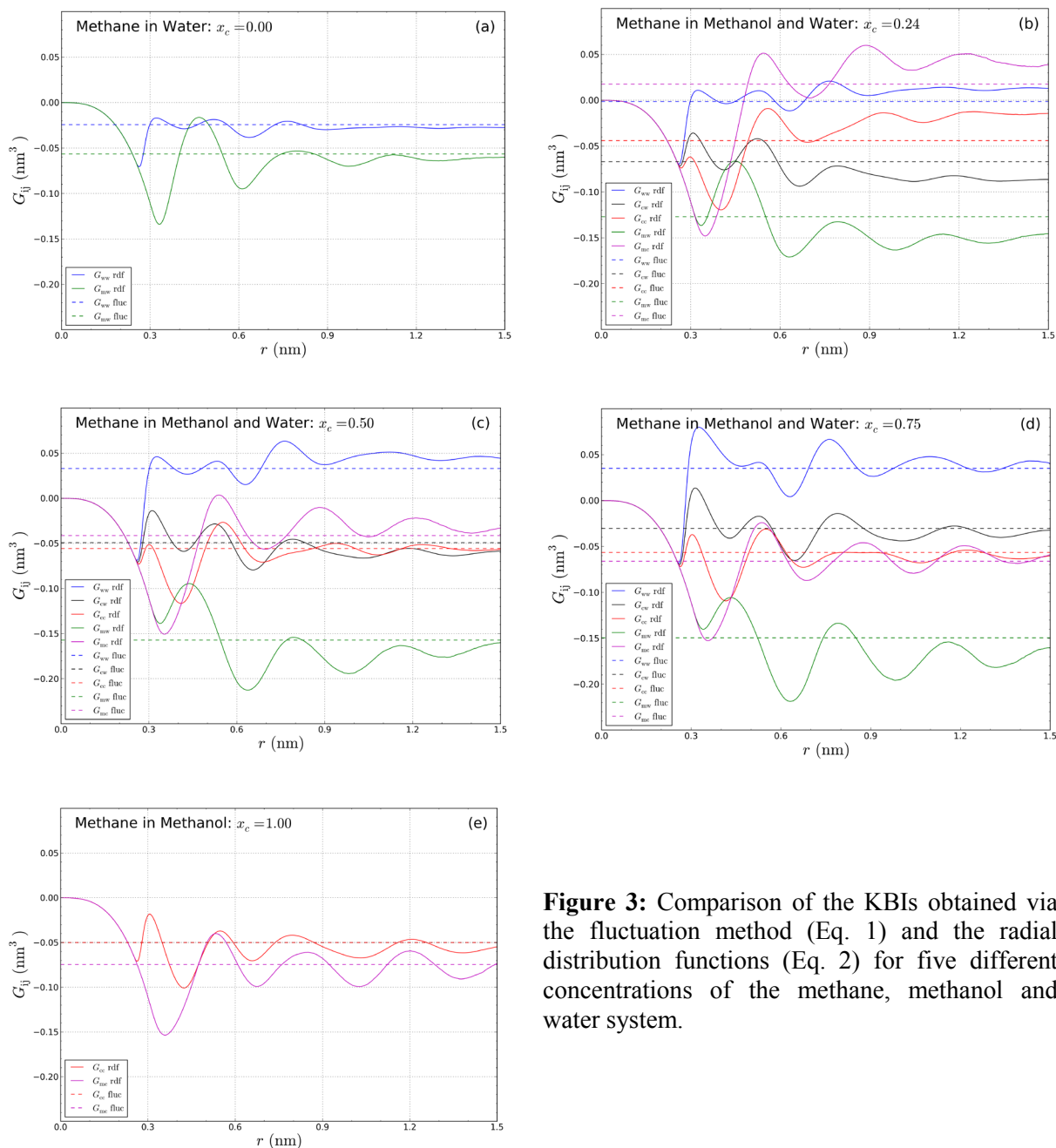
Figure 3: Comparison of the KBIs obtained via the fluctuation method (Eq. 1) and the radial distribution functions (Eq. 2) for five different concentrations of the methane, methanol and water system.

Problems with force field 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 [15-17] 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* be used in the above context to reduce simulation times.

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]   Schnell, S. K. *et al*. 2011 J. Phys. Chem. B **115** 10911
[6]   Liu X *et al*. 2011 J. Phys. Chem. B **115** 12921
[7]   Schnell, S. K. *et al*. 2012 Molecular Physics **110** 1069
[8]   Ben-Naim, A. Molecular Theory of Solutions; Oxford University Press: New York, 2006.
[9]   Kirkwood, J G and Buff F P J 1951 Chem. Phys. **19** 774
[10]   Mukherji D *et al*. 2012 J. Chem. Theory Comput., **8** 375
[11]   Lindahl E, Hess B, and Van der Spoel D 2001 J. Mol. Mod. **7** 306
[12]   Berendsen H J C *et al*. 1984 J. Chem. Phys. **81** 3684
[13]   Perera A, Sokolić F, Almásy L, and Koga Y 2006 J. Chem. Phys. **124** 124515
[14]   Matteoli E and Lepori L 1984 J. Chem. Phys. **80** 2856
[15]   Weerasinghe S and Smith P E 2003 J. Chem. Phys. **118** 5901
[16]   Weerasinghe S and Smith P E 2005 J. Phys. Chem. B **109** 15080
[17]   Hawlicka E and Rybicki M *Molecular Dynamic Simulation of the Ion Solvation in Methanol-Water Mixtures*, in Molecular Dynamics - Theoretical Developments and Applications in Nanotechnology and Energy, ed. Lichang Wang (InTech, 2012)