Evaluating the Transferability of Coarse-Grained, Density Dependent Implicit Solvent Models to Mixtures and Chains

Erik C. Allen, Gregory C. Rutledge*
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA

***Corresponding author:** rutledge@mit.edu

Keywords: Coarse-graining, implicit solvent, density dependence, mixing rule

Abstract:

Previously, we described a coarse-graining method for creating local density-dependent implicit solvent (DDIS) potentials that reproduces the radial distribution function (RDF) and solute excess chemical potential across a range of particle concentrations [E.C. Allen and G.C. Rutledge, J. Chem. Phys. **128**, 154115 (2008)]. In this work, we test the transferability of these potentials, derived from simulations of monomeric solute in monomeric solvent, to mixtures of solutes and to solute chains in the same monomeric solvent. For this purpose, "transferability" refers to the predictive capability of the potentials without additional optimization. We find that RDF transferability to mixtures is very good, while RDF errors in systems of chains increase linearly with chain length. Excess chemical potential transferability is good for mixtures at low solute concentration, chains, and chains of mixed composition; at higher solute concentrations in mixtures, chemical potential transferability fails due to the unique nature of the DDIS potentials. With these results, we demonstrate that DDIS potentials derived for pure solutes can be

used effectively in the study of many important systems including those involving mixtures, chains and chains of mixed composition in monomeric solvent.

I. Introduction

Coarse-graining techniques have received growing interest as methods to extend the time and length scales of molecular simulations. In its most essential form, a coarse-graining algorithm is a statistical fitting process that systematically reduces the number of degrees of freedom from that of an all-atom simulation. This reduction is achieved either by aggregating multiple atomic coordinates into a single coarse-grained (CG) particle or by the deletion of particles, as in an implicit solvent simulation. A coarse-graining scheme usually starts with a short simulation in which every particle is explicitly included; this is used to generate descriptive "data" about the exact system. This simulation is followed by a matching procedure in which the coarse-grained potential is created and applied to a system involving a reduced number of particles to reproduce the data of the underlying all-particle simulation. Various types of data have been used for CG potential fitting, including forces[1-7], reversible work[8], radial distribution functions[9-12] (RDFs), and experimental results[13-15]. Additionally, a wide variety of fitting procedures have been proposed, including simplex optimization[16-18], radial distribution function inversion[19-31], wavelet transform[32,33], and semi-grand canonical Monte Carlo[13].

The value of a coarse-grained potential is determined in large part by its utility, or "transferability", to situations outside of the one to which it was fit, because transferability is the only way that the upfront cost of performing the fit can be recouped.

As suggested by Johnson *et al.*[34], transferability can be classified as one of two types: "observable transferability", and "state point transferability". Observable transferability (called "representability" by Johnson *et al*) refers to the ability of a CG potential that is fit to one set of simulation observables to reproduce accurately another set of simulation observables. State point transferability, by contrast, refers to ability of a CG potential that is fit at one thermodynamic state point (temperature, density, system size, and composition) to predict the same observables at other thermodynamic state points. Both types of transferability have been previously addressed in the literature.[36-40]

Lyubartsev[35] examined the observable transferability of a CG potential for lipid molecules in water. The potential utilized RDF matching, and was subsequently shown to reproduce reasonably well the density profile of a lipid bilayer. Noid *et al.*[6] showed that for isotropic, homogeneous materials a force-matched potential will also reproduce the system RDF. Frequently, however, CG algorithms are used to simulate systems on time scales that are inaccessible via all-atom simulation[35]. In these cases, demonstrating observable transferability with respect to an underlying molecular simulation is impossible, since the all-atom simulation cannot be performed for purposes of comparison. Further, theoretical study by Louis[36] demonstrated that observable transferability breaks down even in simple cases; for example, he showed that potentials fit to system RDFs do not reproduce system energy, and *vice-versa*. As a result, he suggested that only more complex potential forms, incorporating density dependence or many-body interactions, may overcome these representability problems. Indeed, Merabia and Pagonabarraga[37] have demonstrated that density dependent potentials avoid some

theoretical representability problems. In previous work[38] we demonstrated that a coarse-grained implicit solvent potential containing density dependent one-body and two-body interactions can replicate both the system RDF and a measure of system energy across a range of solute concentrations.

State point transferability has been studied many times as well. Henderson's[39] seminal theoretical work indicates that the pairwise potential that generates a given RDF is unique to within an additive constant for systems with only pairwise interactions. However, the RDF of a system depends on the state point; it does not follow that the potential obtained by RDF inversion at one state point will generate the correct RDF at other state points. Recent simulations of polymers[40] by this approach have exhibited limited state-point transferability of such potentials. Louis[36] has suggested that state point transferability may be improved in systems containing a local density dependence, since these systems contain an extra parameter based on the local environment that a CG particle experiences. It is worth noting that improving one type of transferability is no guarantee of improving the other.

This paper examines the state point transferability of the density-dependent implicit solvent (DDIS) potentials reported in our previous work[38], hereafter referred to as Paper 1. There, the relative merits of density-dependent potentials parameterized on either the local or global solute density were explored; in this work, we focus primarily on the local density-dependent implicit solvent potential, abbreviated as DDIS unless noted otherwise. The fitting process used in generating the DDIS potentials guarantees the

transferability across a range of solute concentrations. This paper examines their performance in other situations. Specifically, we create DDIS potentials generated from simulations of pure monomeric solute particles in monomeric solvent, and then examine the transferability of these potentials to two other cases – mixtures involving more than one type of solute particle and solutes comprising chains particles – in the same implicit solvent. These cases are of particular relevance for the study of surfactants, because transferability would imply that one need parameterize CG potentials only for the head and tail solute particles individually, thereby greatly extending their utility. Therefore, we examine here the limits of transferability of density dependent potentials for these cases.

II. Theory

Density Notation Conventions

In this work, ρ_S refers to the *total* density of solute particles, where the subscript "S" stands for solute, and may include contributions from different solute types. ρ_A, ρ_B, and ρ_C, respectively, are the densities of A, B, and C-type solute particles only. Finally, ρ without subscript refers to the state point density, considering all particles (solute and solvent). A superscript L indicates that the density is the *local* density of solute particles; otherwise, the density refers to the global average density.

DDIS Potential Review

In Paper 1, we proposed a density-dependent, implicit solvent (DDIS) potential with the form

$$E_i = \mu\left(\rho_{S,i}^L\right) + \frac{1}{2}\sum_{j\neq i} V^{EFF}\left(r_{ij}, \rho_{S,i}^L\right) \quad (1)$$

where E_i is the effective energy of particle i, V^{EFF} is a pairwise potential between solute particles that is an explicit function of solute particle density in the vicinity of particle i, $\rho_{S,i}^L$, and μ is a "self-interaction" term that is also a function of solute particle density. We also proposed a method to fit such potentials, such that the solute-solute RDF and solute excess chemical potential, μ_A^{ex}, are reproduced across all solution compositions. We present here only a brief sketch of the solution algorithm. For further details, the reader is referred to Paper 1.[38]

The algorithm first involves generating pairwise potentials as a function of the global solute density by performing RDF-inversion[22] for a number of solute compositions. Assuming that the distribution of local densities is centered about the global density in these first simulations, the RDF-inversion potential obtained for a given global solute density can then be taken as representative of the potential to be applied for a particle that experiences a comparable local solute density, regardless of the actual global density of the system in which the particle is found. Once the pairwise potentials are determined, an iterative procedure is used to determine the self-interaction as a function of local solute density such that the solute excess chemical potential is reproduced across all compositions. The method does not guarantee fit to an arbitrary accuracy.

The measure of error in the RDF is given by the solute-solute energy, defined as:

$$E_{L,AA} = \rho_A \int_0^{r_C} V_{AA}(r) g_{AA}(r, \rho_A, \rho_B) 4\pi r^2 dr \quad (2)$$

where ρ_A is the density of solute type A, V_{AA} is the exact interaction potential between A particles, and g_{AA} is the A-A RDF. The difference between E_L for a coarse-grained system and that for the all-atom simulation provides a measure of the error in $g(r,\rho_A,\rho_B)$, relative to the exact $g(r,\rho_A,\rho_B,\rho)$ for the explicit system, in units of energy. The error in the excess chemical potential, μ_A^{ex}, is the difference (in units of ε_Z, the Lennard-Jones parameter of the solvent) between the target (all-atom) and measured (coarse-grain) values. In this text, both error measurements are presented in terms of the standard error over all simulations. Additionally, we supply figures where appropriate comparing the all-atom and coarse-grained chemical potentials and RDFs, as these offer an intuitive sense of the closeness of fit.

Solute Enhancement Ratio

One way to characterize the local environment is to calculate the local number of solute particles, $<N_L>$, defined as:

$$< N_L >= \rho_S \int_0^{R_C} g(r;\rho_S) 4\pi r^2 dr \quad (3)$$

We define the solute enhancement ratio as the number of solute particles within a sphere of radius R_c divided by the average number of particles in such a volume. A ratio near

1.0 indicates that the local solute environment is very similar to the global average environment. We find this metric particularly useful in the discussion of chain molecules.

Mixing Rule

Simulations of solute particle types A and B in solvent Z require a mixing rule to govern A-B interactions. We propose here a simple yet physically intuitive mixing rule for the two-body portion of the DDIS potential, which has the benefit of requiring no additional simulation. The one-body portion of the potential is left unmodified. There are certainly more complex mixing rules that one could propose; some of these choices are discussed in Appendix A. We begin by decomposing the A-A coarse-grained potential, V_{AA}^{EFF}, into two terms:

$$V_{AA}^{EFF}\left(\frac{r}{\sigma_A};\rho_A^L\right) = V_{AA}\left(\frac{r}{\sigma_A}\right) + \Delta V_{AA}^{EFF}\left(\frac{r}{\sigma_A};\rho_A^L\right) \tag{4}$$

where V_{AA} is the all-atom potential, and ΔV_{AA}^{EFF} incorporates all modifications to the all-atom potential. A similar equation can be written for the B-type solute. We then assume that the modification of the A-B potential upon coarse-graining follows a simple mixing rule, such that

$$V_{AB}^{EFF}\left(\frac{r}{\sigma_{AB}};\rho_S^L\right) = V_{AB}\left(\frac{r}{\sigma_{AB}}\right) + \frac{\varepsilon_{AB}}{2}\left(\frac{\Delta V_{AA}^{EFF}\left(r/\sigma_A;\rho_S^L\right)}{\varepsilon_A} + \frac{\Delta V_{BB}^{EFF}\left(r/\sigma_B;\rho_S^L\right)}{\varepsilon_B}\right) \tag{5}$$

where ρ_S^L is the local solute density comprising both A and B, and V_{AB} is the A-B all-atom potential. Whereas in the single solute system, V_{AA}^{EFF} and the one-body term are both parameterized on the local density of A particles, in the combined system both are now assumed to be parameterized based on the total solute density ρ_S^L. Equation (5) includes explicit reference to V_{AB}, and therefore does not assume a particular mixing rule for the all-atom potential.

Limits of Transferability

Even if the mixing rules presented above were to produce the correct RDFs for every system composition, this does not in general guarantee the transferability with respect to the excess chemical potential. To see this, we extend the analytical framework of Paper 1, demonstrated in Figure 1, to mixtures of two solute particles. Here, the excess chemical potential of particle type A in an implicit solvent simulation with global solute densities ρ_A and ρ_B (ρ_S=ρ_A+ρ_B), is given by:

$$\mu_A^{ex}\left(\rho_A,\rho_B\right) = \Delta F_{1,A}\left(\rho_A,\rho_B\right) + \Delta F_{2,A}\left(\rho_A,\rho_B\right) \tag{6}$$

where $\Delta F_{1,A}$ is the free energy change associated with pairwise interactions between the inserted particle(s) and the system,

$$\Delta F_{2,A}(\rho_A,\rho_B) = \mu_A\left(\rho_S^L = \rho_S\right) + (n_A - 1)\frac{\partial \mu_A\left(\rho_S^L\right)}{\partial n_A} + n_B \frac{\partial \mu_B\left(\rho_S^L\right)}{\partial n_A} + \frac{(n_A - 1)^2}{2}\left[AA\left(\rho_S^L\right)\right] + (n_A - 1)n_B\left[AB\left(\rho_S^L\right)\right] + \frac{{n_B}^2}{2}\left[BB\left(\rho_S^L\right)\right] \quad (7)$$

$n = n_A + n_B$, and

$$\left[AA\left(\rho_S^L = \rho_S = \frac{n}{V}\right)\right] = \frac{1}{V}\int_0^{r_c} g_{AA}(r,\rho_A,\rho_B)\frac{\partial V_{AA}^{EFF}\left(r,\rho_S^L\right)}{\partial n} 4\pi r^2 dr \quad (8)$$

$AB\left(\rho_S^L\right)$ and $BB\left(\rho_S^L\right)$ are defined by equations similar to Equation (8). The excess chemical potentials calculated by particle insertion for both the all-atom and coarse-grained potentials contain free energy contributions associated with the pairwise interactions between the inserted solute particle and the other particles in the system (denoted ΔF_1' and ΔF_1, respectively, in Figure 1). The correction term, Equation (7), is unique to the DDIS framework, and arises because insertion of a solute particle changes the density (and hence energetic interactions) of the system. This small change in solute density introduces a differential change in the interactions between existing particles in the system, which nonetheless produces a significant impact on the excess chemical potential for high solute densities. The first three terms on the right-hand of Equation (7) address one-body energy changes, while the last three terms (AA, AB, and BB) address the change in existing pairwise interactions due to the differential change in density. Further details can be found in Paper 1.[38]

Because Equation (7) has a dependence on $\mu_B^{ex}(\rho_S)$, transferability cannot be assured, since the A potential was fit without knowledge of B-type particles. There are, however, situations in which transferability with respect to μ^{ex} can be expected, as will be discussed next. These special cases cover a large number of relevant simulations.

Special Cases

Transferability with respect to excess chemical potential is more likely for solutions at low total solute concentration. Equation (7) for $\Delta F_{2,\mathrm{A}}$ shows that as n_B becomes small, those terms involving B-type particles will vanish. As these terms vanish, μ_A^{ex} depends only on terms involving A particles. Since these are the conditions under which the A potential was fit, μ_A^{ex} should be accurately reproduced. A similar argument can be made for B-type particles.

A second special case arises when A-type and B-type particles are inserted simultaneously, *in proportion to their existing ratios in the system*. The most common example of such a system is insertion of an A_iB_j chain into a system of identical chains. We assume here that the free energies of insertion are additive. In that case, the average chemical potential of each inserted particle is:

$$\mu^{ex} = \frac{n_A}{n}\mu_A^{ex} + \frac{n_B}{n}\mu_B^{ex} \tag{9}$$

where n_A and n_B are the number of A-type and B-type particles in the system, respectively, and $n = n_A + n_B$. By analogy to Equation (6)-(8), Equation (9) can be written as:

$$\mu^{ex} = \frac{n_A}{n}\Delta F_{1,A}(\rho_A,\rho_B) + \frac{n_B}{n}\Delta F_{1,B}(\rho_A,\rho_B) + \frac{n_A}{n}\Delta F_{2,A}(\rho_A,\rho_B) + \frac{n_B}{n}\Delta F_{2,B}(\rho_A,\rho_B) \tag{10}$$

with

$$\Delta F_{2,A}(\rho_A,\rho_B) = \left[\mu_A(\rho_S) + (n_A - 1)\frac{\partial \mu_A(\rho_S)}{\partial n} + n_B\frac{\partial \mu_B(\rho_S)}{\partial n}\right] + \left[\frac{(n_A-1)^2}{2}\left[AA\left(\frac{n}{V}\right)\right] + (n_A-1)n_B\left[AB\left(\frac{n}{V}\right)\right] + \frac{{n_B}^2}{2}\left[BB\left(\frac{n}{V}\right)\right]\right] \tag{11}$$

and

$$\Delta F_{2,B}(\rho_A,\rho_B) = \left[\mu_B(\rho_S) + (n_B - 1)\frac{\partial \mu_B(\rho_S)}{\partial n} + n_A\frac{\partial \mu_A(\rho_S)}{\partial n}\right] + \left[\frac{(n_B-1)^2}{2}\left[BB\left(\frac{n}{V}\right)\right] + (n_B-1)n_A\left[AB\left(\frac{n}{V}\right)\right] + \frac{{n_A}^2}{2}\left[AA\left(\frac{n}{V}\right)\right]\right] \tag{12}$$

If we assume first that $AA(n/V)$ and $BB(n/V)$ are not greatly different from their pure component values, and second that $AB(n/V)$ is an average of $AA(n/V)$ and $BB(n/V)$, then Equations (10)-(12) can be simplified to

$$\begin{aligned}\mu^{ex} &= \frac{n_A}{n}\Delta F_{1,A}(\rho_A,\rho_B) + \frac{n_B}{n}\Delta F_{1,B}(\rho_A,\rho_B) + \\ &\frac{n_A}{n}\left[\mu_A(\rho_S) + (n-1)\frac{\partial \mu_A(\rho_S)}{\partial n} + \frac{(n-1)^2}{2}\left[AA\left(\frac{n}{V}\right)\right]\right] + \\ &\frac{n_B}{n}\left[\mu_B(\rho_S) + (n-1)\frac{\partial \mu_B(\rho_S)}{\partial n} + \frac{(n-1)^2}{2}\left[BB\left(\frac{n}{V}\right)\right]\right]\end{aligned} \tag{13}$$

The bracketed terms in Equation (13) are simply the correction terms derived for pure solute system, $\Delta F_{2,A}(\rho_S)$ and $\Delta F_{2,B}(\rho_S)$. Therefore, in this special case, the average excess chemical potential will be correct.

III. Simulation Protocol

Simulation protocols for all-atom and coarse-grained simulations of monomeric solute are given in Paper 1. . All-atom simulations were performed using the GROMACS (Version 3.3)[41] molecular dynamics simulation software, whereas the coarse-grained (DDIS) simulations were performed using a Monte Carlo code of our own design. As in that paper, all simulation were carried out in the *NVT* ensemble at $T^*=k_BT/\varepsilon_Z=1.35$ and $\rho^*=\rho\sigma_Z^3=0.55$, where ε_Z and σ_Z are the Lennard-Jones parameters of the solvent Z. Each simulation consisted of 1000 total particles. The cutoff radius for interactions was set to $5\sigma_Z$ unless otherwise indicated

All-atom simulations of solute mixtures and of chains were carried out in the same manner as the molecular dynamics simulations of Paper 1, with the exception of intramolecular degrees of freedom for the chain simulations. Bond lengths and angles were held fixed at the minimum energy values using the LINCS[42] algorithm, while torsion angles were allowed to rotate freely. Chains of length two and four were simulated. Free energy calculations in the all-atom simulation were performed using thermodynamic integration (TI)[43]. The standard GROMACS λ switching function was used to integrate between an initial state in which the test particle interactions with the

remaining system were switched off and a final state in which the test particle interactions were fully enabled. Switching was applied to nonbonded interactions only, with soft core interactions to avoid singularities and using a soft core interaction parameter α=0.51. A total of 31 λ values were used (λ = [0.00 0.03 0.07 0.10 …. 0.93 0.97 1.00]).

Implicit solvent Monte Carlo simulations were carried out in the same manner as the monomeric density dependent, implicit solvent simulations. Nearest bonded neighbor particles were included in the calculation of local density. In addition to the chain translation moves used for monomeric simulations, simulations of dimers included rigid body rotation moves as well, in the proportion 80% translation:20% rotation. For tetramers, we also included rotation moves about individual bonds, in the proportion 50% translation:20% rigid body rotation:30% bond rotation. Simulations of monomeric solutes were equilibrated for 10^4 cycles, followed by sampling for 3×10^4 cycles. Sampling of dimers was performed for twice as many cycles, or 6×10^4, and chains of length four were sampled for 1.2×10^5 MC cycles. Free energy was calculated using the Bennett Acceptance Ration method[44]. Two λ values were used (λ = [0.00,1.00]) with an initial state comprising a non-interacting test particle and a final state having a fully interacting test particle. Sampling was performed every two MC cycles, which was sufficient to generate statistically independent samples, as determined by the autocorrelation function of the measured energy difference. RDF sampling was performed every 100 MC cycles.

IV. Results and Discussion

Particle Types Used in this Work

In the all-atom simulations described in this work, particles interact via the truncated and shifted Lennard-Jones potential, where the potential between particles *i* and *j* is described by:

$$V_{ij}^{TS}\left(r_{ij};\varepsilon_{ij},\sigma_{ij},R_{c,ij}\right)=\begin{cases}V_{ij}^{LJ}\left(r_{ij};\varepsilon_{ij},\sigma_{ij}\right)-V_{ij}^{LJ}\left(R_{C,ij};\varepsilon_{ij},\sigma_{ij}\right) & r_{ij}\le R_{C,ij}\\ 0 & r_{ij}>R_{C,ij}\end{cases} \tag{14}$$

with

$$V_{ij}^{LJ}\left(r_{ij};\varepsilon_{ij},\sigma_{ij}\right)=4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{15}$$

where ε_{ij} and σ_{ij} are the Lennard-Jones parameters for *ij* interactions. Equations (14) and (15) allow for the possibility of different cut-off radii ($R_{C,ij}$) for interaction between different particle types. ε and σ values of unlike particles are governed by Lorentz-Berthelot mixing rules.

In Paper 1, we examined the behavior of a single solute type (hereafter referred to as "A") in solvent ("Z") at reduced temperature and density T^*=1.35 and ρ^*=0.55. Because the A-type particles were identical to the solvent Z, the behavior of A-Z mixtures was identical for all compositions of A. For the work presented here, we find it useful to

introduce two additional solute types, which display "solvent-philic" and "solvent-phobic" behavior, respectively.

A solute's relative preference for the solvent can be measured by the free energy of transfer, $\Delta G_S(Z \rightarrow S)$, defined as the free energy change associated with transferring a single solute particle from a bath of solvent particles Z to a bath of solute particles S. The free energy of transfer can be calculated from the excess chemical potential of solute particles:

$$\Delta G_S(Z \rightarrow S) = \mu_S^{ex}\left(\frac{\rho_S}{\rho} = 1\right) - \mu_S^{ex}\left(\frac{\rho_S}{\rho} = 0\right) \tag{16}$$

By varying particle ε_{ii}'s and $R_{C,ij}$'s, we created a "solvent-philic" particle (type B) and a "solvent-phobic" particle (type C), with interaction parameters given in Table 1. Solute type B is distinguished by its reduced ε_B of 0.5, compared to ε_Z of 1.0. Solute type C interacts with the solvent Z via a reduced cutoff radius of $1.84\sigma_Z$, with C-C and Z-Z interactions maintaining the usual $5.0\sigma_Z$ cutoff.

Table 1 shows the solute enhancement ratio for all three types of solute particles at ρ_S/ρ=0.5. The specific ε_{ii}'s and $R_{C,ij}$'s used were selected such that B-type and C-type solute enhancement ratios were close to 1.0 for all compositions. This indicates that both solute types, while expressing relative preference for like or unlike interactions, are completely miscible in solvent at all compositions, and validates the use of Equations (6)-(8).

DDIS Potentials

Paper 1 previously reported the DDIS potential for the coarse-graining of A-type particles in solvent Z, where the A and Z particles were identical. Here, we followed the same fitting procedures to produce DDIS potentials for B-type and C-type particles in the same solvent Z. Figure 2 shows the fitted values of the excess chemical potential compared to the all-atom simulations. Figure 3 shows the worst fit RDFs for each system. In general, the accuracy of fit achieved for the B-type particles is comparable to the results for A-type particles reported in Paper 1, and both are superior to that obtained for the C-type potentials. The standard error in μ^{ex} for B-type particles was $0.04\varepsilon_Z$, versus $0.11\varepsilon_Z$ for C-type particles. Similarly, the E_L standard error for B-type particles was $0.005\varepsilon_Z$, versus $0.06\varepsilon_Z$ for C-type particles. The worst case RDFs for the B-type and C-type particles are shown in Figure 3. The lower quality of fit attained for the C-type particles is perhaps because of the large discrepancy in cut-off radii between like and unlike particles. In both cases, however, the visual examination suggests the fit is quite good.

Figure 4 compares the two-body portion of the coarse-grained potentials for the A-type, B-type, and C-type potentials at a local density of ρ_S^L/ρ=0.5. All three potentials display the same general form, with a secondary local maximum in the potential. The "solvent-philic" B-type particles show a shallower attractive well when compared to the A-type particles, while the "solvent-phobic" C particles have a deeper well. This is consistent with the form of the pairwise implicit solvent potentials used in the Effective Energy Function[45].

Figure 5 shows the one-body portion of the DDIS potential for A,- B-, and C-type particles. As with the two-body portion, all the potentials share certain general characteristics. As solute density goes to zero, the value of the one-body term approaches $\mu_S^{ex}(\rho_S/\rho=0)$. The profile is relatively flat for low solute densities, before rising rapidly as the solute fraction approaches 1.

The potentials used in this work can be found in tabular format in the Supplemental Information section.

Mixtures

We investigated the behavior of A/B and A/C mixtures in implicit solvent Z. The analysis of this system is simplified by the fact that the A and Z particles are identical, which means that the system can be expressed solely as a function of ρ_B for the A/B system, and of ρ_C for the A/C systems. Considering A/X/Z systems, where X=B or C, is equivalent to considering A/X systems with partially coarse-grained A.

In this work we performed simulations in increments of 10% in solute concentration. As a result, there were 55 possible combinations of A/X/Z mixtures. To reduce computational time, we simulated only a representative fraction of these systems. That subset consisted of systems in which the total solute density, ρ_S, was 20%, 50%, or 90% of the total particle density, and constitutes 13 of the 55 mixture fractions possible.

Additionally, we measured only $\mu_X^{ex}(\rho_A,\rho_X)$ (and not $\mu_A^{ex}(\rho_A,\rho_X)$) for each system, to further reduce computational time.

The ability of the DDIS potential to reproduce the all-atom RDF's for both systems was very good. For the A/B/Z mixtures, the standard errors in E_L were $E_{L,AA}$=0.013ε_Z and $E_{L,BB}$=0.005ε_Z, which compare favorably with the pure-component errors of $E_{L,AA}$=0.015ε_Z and $E_{L,BB}$=0.005ε_Z. In other words, the RDF accuracy of the coarse-grained mixtures is comparable to the accuracy of the coarse-grained pure systems. Results were similar for the A/C/Z mixtures. Here, the standard errors were $E_{L,AA}$=0.010ε_Z and $E_{L,CC}$=0.06ε_Z, compared to the pure component errors of $E_{L,AA}$=0.015ε_Z and $E_{L,CC}$=0.06ε_Z. Figure 6 shows the worst reproductions of all-atom RDFs by the DDIS potentials for the A/B/Z and A/C/Z potentials, which demonstrates visually the similarity of RDFs.

Figure 7 examines the transferability with respect to excess chemical potential for the A/B/Z and A/C/Z mixtures at the three total solute concentrations. As expected, the transferability with respect to excess chemical potential is not as good as that with respect to the RDF, for the reasons laid out in the theory section. The standard error in $\mu_B^{ex}(\rho_A,\rho_B)$ for the A/B/Z system is 0.47ε_Z, and the standard error in $\mu_C^{ex}(\rho_A,\rho_B)$ for the A/C/Z system is 0.63ε_Z, which are both substantially higher than the pure-component errors of 0.04ε_Z and 0.11ε_Z. Nevertheless, the performance of the DDIS potentials at low solute concentration is quite good. The chemical potential errors for ρ_S/ρ=0.2 are 0.01ε_Z and 0.21ε_Z, which are comparable to the error in the pure component systems.

Chains

To study the transferability of DDIS potentials to systems of chains, we examined the behavior of dimers and tetramers of A, B, and C solutes in solvent Z. These simulations were performed over a range of solute densities from ρ_S/ρ=[0.1,1.0] in increments of 0.1, which were the same solute densities at which the potential was fit. Bond lengths were held fixed at $1.22\sigma_Z$, which is equal to the average separation between particles in the monomeric simulations. We selected this bond length so that the total local density of particles (solute and solvent) remained close to the fitting density. For shorter chain lengths, total local density was higher than in the monomeric case. Angles were held fixed at 112 degrees. Errors in μ^{ex} were measured on a per-particle basis for chain simulations, to permit comparison to the errors measured in monomeric solute simulations.

We first examined the all-atom and DDIS errors in E_L for monomers, dimers and tetramers of A, B, and C. These results are summarized in Table 2. For all three particle types, the errors in the reproduction of all-atom RDFs increased with increasing chain length. It appears that, for the systems studied here, each doubling of the chain length results in an approximate doubling in the error of E_L, although the progression is not entirely smooth. For example, the accuracy appears nearly equal for dimers and tetramers of B. Figure 8 shows the worst case reproductions of all-atom RDFs by the DDIS potentials for tetramers for each particle type. At the tetramer level, one can begin to see visual disagreement between the all-atom and coarse-grained RDFs. In general,

the locations of the peaks in the RDF remain correct. Errors arise in maxima and minima of g(r) that are systematically too high or too low.

Figure 9 compares the all-atom and DDIS results for μ^{ex} per particle for monomers, dimers and tetramers of A, B, and C. The standard errors are given in Table 3. For all particle types, the errors increased with increasing chain length. As is the case in mixtures, chemical potential transferability is better at low solute concentrations. The reason that transferability breaks down at higher solute concentrations is related to the two-body correction factor, Equation (8), which contains the solute-solute RDF. As chain length increases, the solute-solute RDF becomes less similar to the RDF at the fitting conditions, and as a result transferability degrades. At low solute densities, the influence of the two-body correction is smaller, and μ^{ex} transferability is better.

In Paper 1, we demonstrated that a potential with local density dependence performed better in reproducing the RDF behavior of a system with solute aggregation than did a potential with a global density dependence. We hypothesized that this was because the local density dependence permitted the potential to be responsive to local variations in the environment around the solute at every time point of the simulation. In this work, we tested if this result was applicable to other systems with solute aggregation, namely dimers and tetramers of C-type solute in solvent Z. As discussed previously, C-type particles are completely miscible in solvent Z in the monomer state. However, longer chains of C-type particles aggregate substantially. At ρ_S/ρ=0.5, the solute enhancement ratio for monomers is 1.01, indicating a nearly homogeneous dispersion. For dimers, the

solute enhancement ratio rises to 1.09, and to 1.21 for tetramers. This solute enhancement cannot be attributed to the presence of solute particles held in close proximity by bonding, as chains of A-type particles do not experience an equivalent increase in solution enhancement ratio. Instead, the local composition is more concentrated in solute than is the system in which the coarse-graining was performed due to chain aggregation. Aggregation is expected because increasing the chain length reduces molecular entropy, and thus the driving force for solute particle dispersion in solution.

Figure 10 compares the performance of the local DDIS potential of C-type solutes dispersed in Z to the equivalent global density-dependent potential. The global potential was tuned to fit the RDF of monomeric solute particles at a given solute density, and then tested on chains at the same solute density. The local potential shows superior RDF transferability when compared with the global potential. The slope of the standard error line as a function of chain length is smaller for the local potential, and for tetramers the absolute value of the error is smaller than for the global potential.

Figure 11 examines the all-atom RDF at a given solute density, and compares the results to those obtained using the local and global potentials. The results demonstrate that the global density-dependent potential does not correctly capture the solute aggregation effect. The value of g(r) is systematically too high for all r. For the composition shown in Figure 11, the all-atom ratio is 1.47, and the local potential yields a ratio of 1.43. By contrast, the global potential yields a solute enhancement ratio of 1.76.

Chains of Mixed Composition

Finally, we examined a system of chains of mixed composition. This system invokes a combination of mixture and chain transferability. The surfactants studied here are A_1B_1 molecules in solvent Z. The bond length is set to $1.22\sigma_Z$, as in the foregoing section on chains, and is held constant throughout the simulation.

Transferability with respect to μ^{ex} was quite effective, with a standard error in μ^{ex} of $0.13\varepsilon_Z$, as shown in Figure 12. This value is substantially below the errors found in mixtures of monomeric A and B ($0.47\varepsilon_Z$) and validates our predictions from the theoretical section. The RDF transferability is also good, with standard errors in $E_{L,AA}$ and $E_{L,BB}$ of $0.012\varepsilon_Z$ and $0.008\varepsilon_Z$, respectively. These results are comparable to the standard error at the state point of fit.

V. Conclusion

In this work, we developed coarse-grained DDIS potentials for pure solute particles, and investigated their transferability to solute mixtures and chains in the same solvent. For mixtures, transferability with respect to the solute-solute RDF is very good. For the mixture systems studied in this work, the RDF standard errors were equal to those at the state point of fitting, indicating no loss of accuracy. The transferability with respect to μ^{ex} in mixtures is good for low solute concentrations, but at high solute concentrations the transferability breaks down. This breakdown is due to the unique nature of the DDIS potential, in which particle insertion causes a change in local density for some fraction of

the existing system particles. The parameters of the DDIS potential are carefully tuned to account for this energy change at the state point of fitting. At other compositions, the energy change is different; as a consequence, the predictions of excess chemical potential are less accurate.

For the systems of chains studied, we found that errors in RDF fitting increased linearly with chain length. Increasing chain length was also associated with increases in the μ^{ex} error. Additionally, we find that the local DDIS potential produced superior RDF transferability when compared to comparable global density-dependent implicit solvent potentials for a chain system with solute aggregation. The most likely reason for this improved transferability is that the local density dependence captures solute enhancement effects more accurately than the global potential. Finally, we found that chains of mixed composition had μ^{ex} transferability that was superior to mixtures of monomers of equivalent composition. This is because the simultaneous insertion of *both* particle types removes the biases inherent in mixed systems.

The transferability characteristics described in this paper indicate that DDIS potentials can function effectively for a wide variety of systems, including mixed composition chain molecules at low concentrations. Given these results, we feel confident in suggesting that the potentials would be useful in the simulation of surfactant systems. We intend to report results for such systems in the near future.

Acknowledgements

This work is supported by funding from the Department of Energy's Computational Science Graduate Fellowship.

Supplemental Information

The Supplemental Information section contains all the forcefields used in this word, presented in table format.

Appendix

The performance of the basic mixing rule described in this work is generally good. The lone exception is in its performance with respect to μ^{ex} in systems of mixtures. We discuss here a more complex mixing rule that can improve transferability of potentials.

Equations (6)-(8) provide a framework to predict the excess chemical potential, $\mu_A^{ex}(\rho_A,\rho_B)$, based on the self-interaction energies, μ_A and μ_B. Therefore, Equations (6)-(8) can be used to optimize the values of the self-interaction energies based on any number of parameterizations. Here, we show that $\mu_A(\rho_A^L,\rho_B^L)$ and $\mu_B(\rho_A^L,\rho_B^L)$ can be parameterized as a two-dimensional function of both ρ_A^L and ρ_B^L. To do so, we propose adding a cubic correction term to the pure component one-body terms, so that

$$\mu_A\left(\rho_A^L,\rho_B^L\right) = \mu_A^{pure}\left(\rho_A^L + \rho_B^L\right) + \mu_A^{cub}\left(\rho_A^L,\rho_B^L\right) \tag{A1}$$

with

$$\mu_A^{cub}\left(\rho_A^L,\rho_B^L\right)=c_1\rho_A^L+c_2\rho_B^L+c_4\left(\rho_A^L\right)^2+c_4\rho_A^L\rho_B^L+c_5\left(\rho_B^L\right)^2+c_6\left(\rho_A^L\right)^3+c_7\left(\rho_A^L\right)^2\rho_B^L+c_8\rho_A^L\left(\rho_B^L\right)^2+c_9\left(\rho_B^L\right)^3 \quad \text{(A2)}$$

A similar equation can be written for the μ_B's. If values of $\mu_{ex}{}^A$, $\mu_{ex}{}^B$, $\Delta F_{1,A}$, and $\Delta F_{1,B}$ can be measured or estimated across as range of densities (ρ_A,ρ_B); then the resulting system of 18 unknowns can then be optimized without the use of additional molecular simulations. We elected to use a cubic correction instead of the simpler quadratic because optimization over Equations (6)-(8) indicated a slightly superior fit with the cubic correction. However, quadratic correction would likely provide adequate results in many cases.

We examined the performance of this new algorithm in A/B/Z mixtures, and compared the results to the original mixing rule. A subset of the results is shown in Figure A1. It indicates that the two-dimensional density dependence improves transferability substantially. The standard error with respect to μ^{ex} is $0.18\varepsilon_Z$ for the 13 points studied, compared to $0.47\varepsilon_Z$ for the original mixing rule.

References

1 S. Izvekov and G. A. Voth, J. Chem. Phys. **123**, 134105 (2005).

2 Q. Shi, S. Izvekov, and G. A. Voth, J. Phys. Chem. B **110**, 15045 (2006).

3 S. Izvekov and G. A. Voth J. Chem. Phys. **125**, 151101 (2006)

4 G. S. Ayton, W. G. Noid, and G. A. Voth, Mrs Bull. **32**, 929 (2007).

5 P. Liu, S. Izvekov, and G. A. Voth, J. Phys. Chem. B **111**, 11566 (2007).

6 W. G. Noid, J. W. Chu, G. S. Ayton, and G. A. Voth, J. Phys. Chem. B **111**, 4116 (2007).

7 J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth, Biophys. J. **92**, 4289 (2007).

8 J.D. McCoy and J.G. Curro, Macromolecules **31**, 9362 (1998).

9 K. Kremer and F. Muller-Plathe, Mol. Simulat. **28**, 729 (2002).

10 N. B. Wilding, J. Chem. Phys. **119**, 12163 (2003).

11 M. Bathe and G. C. Rutledge, J. Comput. Chem. **24**, 876 (2003).

12 M. Bathe, G. C. Rutledge, A. J. Grodzinsky, and B. Tidor, Biophys. J. **88**, 3870 (2005).

13 G. C. Rutledge, Phys. Rev. E **63** 021111 (2000).

14 S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein, J. of Chem. Phys. **119**, 7043 (2003).

15 W. Shinoda, R. Devane, and M. L. Klein, Mol. Simulat. **33**, 27 (2007).

16 R. Faller, H. Schmitz, O. Biermann, and F. Muller-Plathe, J. Comput. Chem. **20**, 1009 (1999).

17 H. Meyer, O. Biermann, R. Faller, D. Reith, and F. Muller-Plathe, J. Chem. Phys. **113**, 6264 (2000).

18 D. Reith, H. Meyer, and F. Muller-Plathe, Macromolecules **34**, 2335 (2001).

19 D. Reith, M. Putz, and F. Muller-Plathe, J. Comput. Chem. **24**, 1624 (2003).

20 G. Milano, S. Goudeau, and F. Muller-Plathe, J. Polym. Sci. Pol. Phys. **43** (8), 871 (2005).

21 G. Milano and F. Muller-Plathe, J. Phys. Chem. B **109**, 18609 (2005).

22 W. Schommers, Phys. Lett. A **43**, 157 (1973).

23 W. Schommers, Phys. Rev. A **28**, 3599 (1983).

24 D. Levesque, J.J. Weis, and L. Reatto, Phys. Rev. Lett. **54**, 454 (1985).

25 G. Toth and A. Baranyai, Mol. Phys. **97** (3), 339 (1999).

26 G. Toth and A. Baranyai, J. Mol. Liq. **85**, 3 (2000).

27 G. Toth and A. Baranyai, J. Phys.: Condens. Matter **17**, S159 (2005).

28 A. P. Lyubartsev and A. Laaksonen, J. Phys. Chem. **100**, 16410 (1996).

29 A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **55**, 5689 (1997).

30 A. P. Lyubartsev and A. Laaksonen, Comput. Phys. Comm. **122**, 57 (1999).

31 A. P. Lyubartsev and A. Laaksonen, Chem. Phys. Lett. **325**, 15 (2000).

32 A.E. Ismail, G.C. Rutledge, G. Stephanopoulos, J. Chem. Phys. **118**, 4414 (2003).

33 A.E. Ismail, G. Stephanopoulos, G.C. Rutledge, J. Chem. Phys. **118**, 4424 (2003).

34 M.E. Johnson, T. Head-Gordon, and A.A. Louis, J. Chem. Phys. **126**, 144509 (2007).

35 A. P. Lyubartsev, Eur. Biophys. J. Biophy. **35**, 53 (2005).

36 A.A. Louis, J. Phys.: Condens. Matter **14**, 9187 (2002).

37 S. Merabia and I. Pagonabarraga, J. Chem. Phys. **127**, 054903 (2007).

38 E.C. Allen and G.C. Rutledge, J. Chem. Phys. **128**, 154115 (2008).

39 R.L. Henderson, Phys. Lett. A **49**, 197 (1974).

40 P.G. Bolhuis, A.A. Louis, J.P. Hansen, and E.J. Meijer, J. Chem. Phys. **114**, 4296 (2001).

41 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J.C. Berendsen, J. Comput. Chem. **26**, 1701 (2005).

42 B. Hess, H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije, J. Comp. Chem. **18**, 1463 (1997).

43 T.P. Straatsma and J.A. McCammon, J. Chem. Phys. **95**, 1175 (1991).

44 C.H. Bennett, J. Comput. Phys. **22**, 245 (1976).

45 T. Lazaridis and M. Karplus, Prot. Struct. Funct. Genet. **35**, 133 (1999).

Figure 1: Comparison of free energy changes upon particle insertion in the (a) all-particle and (b) density-dependent, implicit solvent cases. The density-dependent potential introduces a secondary free energy change due to the change in energy models associated with a change in global average solute density of the system.

Figure 2: Comparison of excess chemical potential in all-atom (solid line) and coarse-grained (circles) simulations. Left: B-type particles in solvent Z. Right: C-type particles in solvent Z. The dashed lines demarcate errors of $0.06\varepsilon_Z$, and are provided as a guide to compare the relative errors between the two coarse-grained potentials.

Figure 3: Comparison of worst-fit solute RDF in all-atom (solid line) and coarse-grained (circles) simulations. Left: B-type particles in solvent Z. Right: C-type particles in solvent Z.

Figure 4: Coarse-grained two-body term for local solute density ρ_S^L/ρ=0.5 (—A,--B,···C)

Figure 5: Coarse-grained one-body term as a function of local solute density (Circles: A-type particles, Squares: B-type particles, Crosses: C-type particles)

Figure 6: Worst case RDF reproduction for mixtures. Left Side: A/B/Z mixture, Right Size: A/C/Z mixture (Solid Line – all-atom results; Circles –DDIS results).

Figure 7: Comparison of coarse-grained and all-atom excess chemical potential for mixtures, broken down by total solute density, ρ_S. Left Side: $\mu_B^{ex}(\rho_A,\rho_B)$ in A/B/Z mixtures. Right Side: $\mu_C^{ex}(\rho_A,\rho_C)$ in A/C/Z mixtures. (Solid Line – All-atom values; Circles – ρ_S/ρ=0.2; Squares – ρ_S/ρ=0.5; Crosses – ρ_S/ρ=0.9).

Figure 8: Worst case RDF reproduction for tetramers of solute particles (a) A-type, (b) B-type, (c) C-type. (Solid Line – All-atom; Circles – DDIS results).

Figure 9: Comparison of coarse-grained and all-atom excess chemical potential for monomers, dimers, and tetramers as a function of solute density, ρ_S. (a) A-type, (b) B-type, (c) C-type. (Solid Line – All-atom; Circles – Monomers; Squares – Dimers; Crosses – Tetramers).

Figure 10: E_L standard error (units of ε_Z) as a function of chain length for local DDIS and global density-dependent potentials. (Circles – Local potential; Squares – Global potential).

Figure 11: Comparison of C-C particle RDFs for all-atom, local potential, and global potential for C-type tetramers in solvent Z at ρ_C/ρ=0.2 (Solid Line – all-atom; Circles – global potential; Squares – local potential).

Figure 12: All-atom (solid line) and coarse-grained (circles) $\mu^{ex}(\rho_S)$ for A_1B_1 molecule.

Figure A1: Comparison of measured $\mu_{ex}^{B}(\rho_A,\rho_B)$ transferability with and without mixture correction function for ρ_S/ρ=0.9. Results indicate that parameterizing the self-interaction energy on the density of both particle types improves transferability, and that allowing the pure-component values to vary improves transferability even more (Straight line – All-atom values; Squares – original mixing rule; Circles – modified mixing rule).

Table 1: Parameters for solute types used in this work. B-type solute is "solvent-philic", as indicated by the positive free energy of transfer. C-type is "solvent-phobic", as indicated by a negative free energy of transfer.

Table 2: Values of E_L for systems of monomers (N=1), dimers (N=2), and tetramers (N=4) for chains of A, B, and C type solute particles.

Table 3: Values of μ^{ex} for systems of monomers (N=1), dimers (N=2), and tetramers (N=4) for chains of A, B, and C type solute particles.

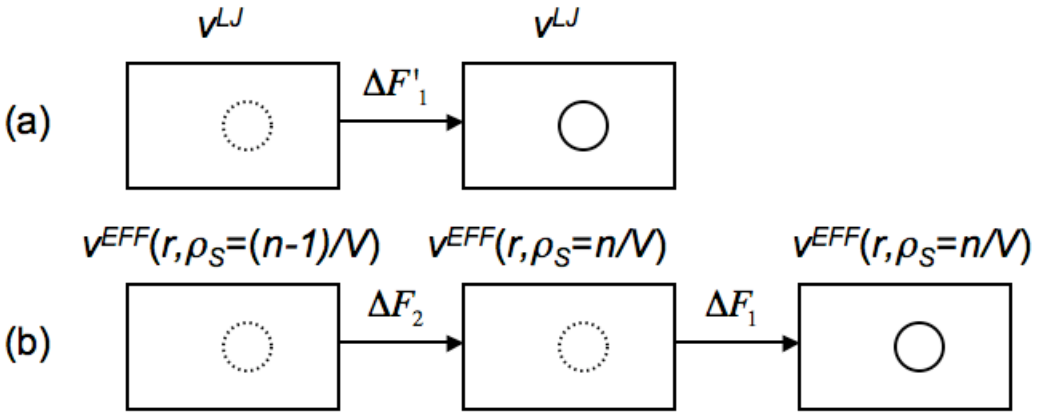
(a)
v^{LJ}
v^{LJ}
$\Delta F'_1$
(b)
$v^{EFF}(r,\rho_S=(n-1)/V)$
$v^{EFF}(r,\rho_S=n/V)$
$v^{EFF}(r,\rho_S=n/V)$
ΔF_2
ΔF_1

Figure 1

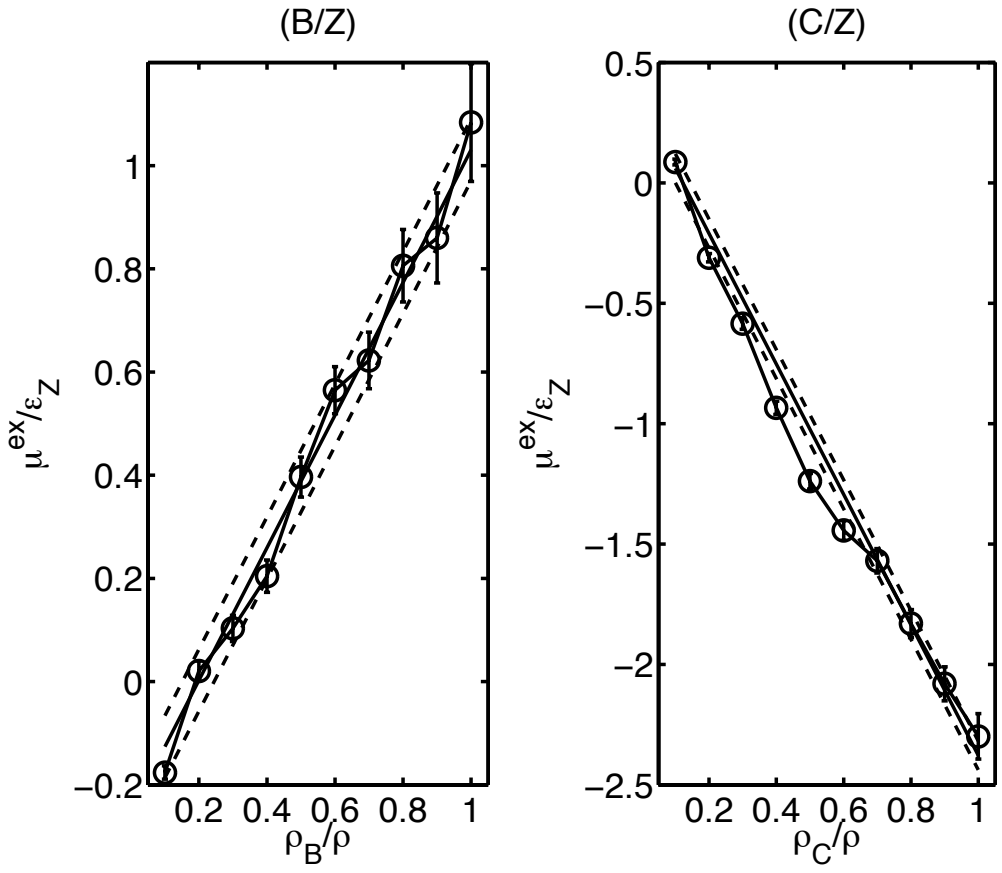
(B/Z)
(C/Z)
μ^{ex}/ε_Z
ρ_B/ρ
ρ_C/ρ
1
0.8
0.6
0.4
0.2
0
−0.2
0.5
−0.5
−1
−1.5
−2
−2.5

Figure 2

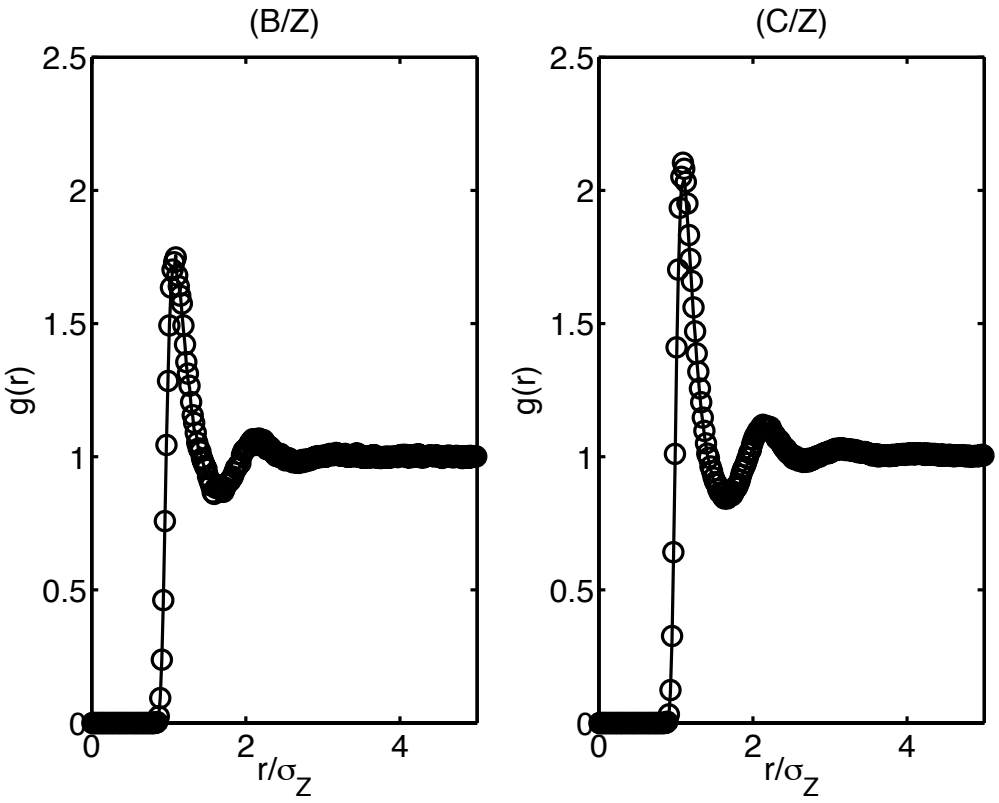

Figure 3

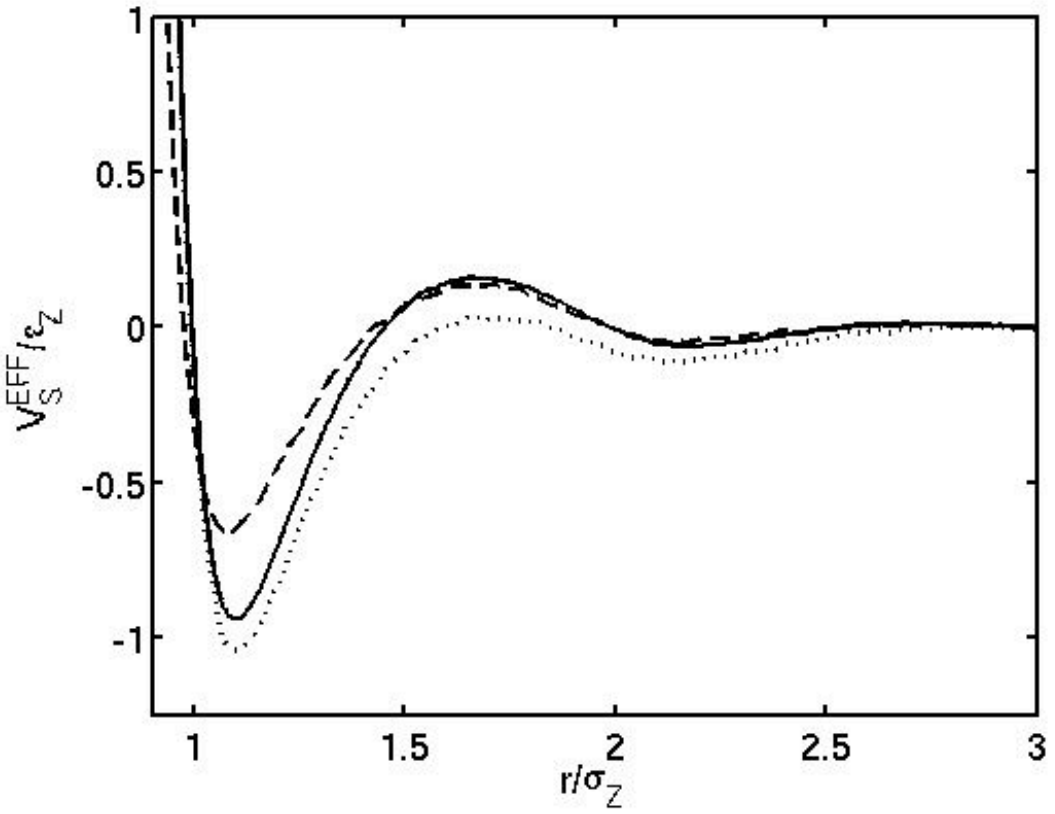

Figure 4

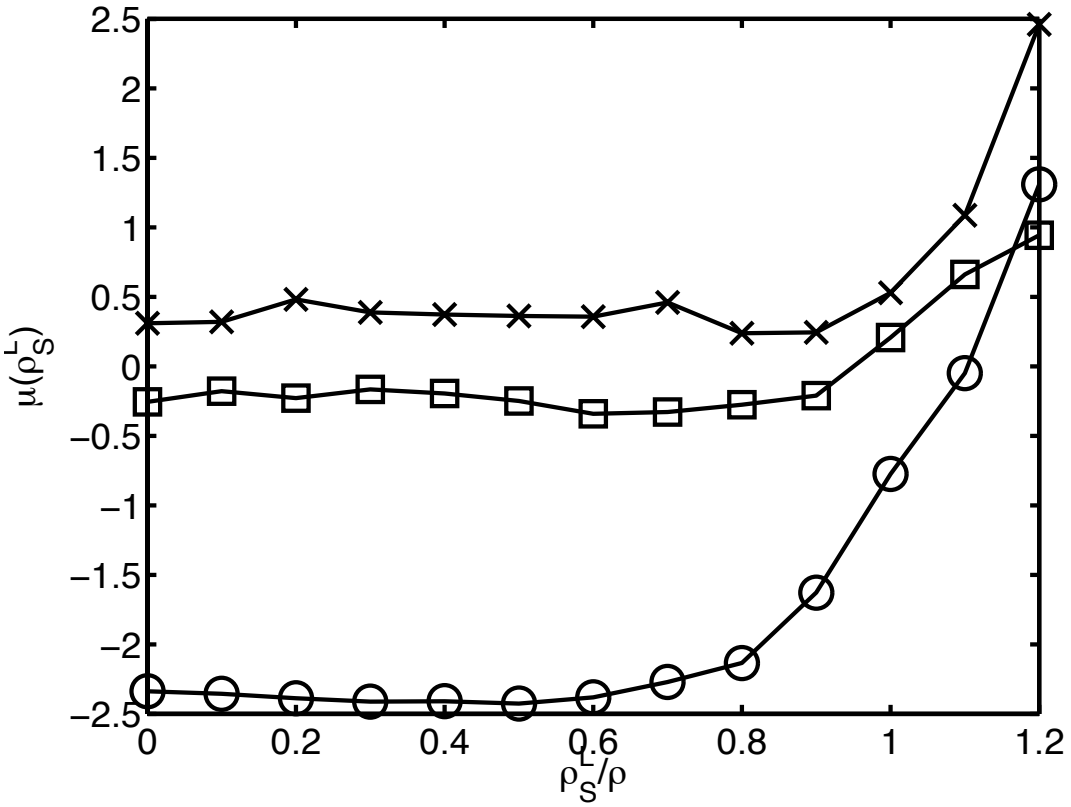

Figure 5

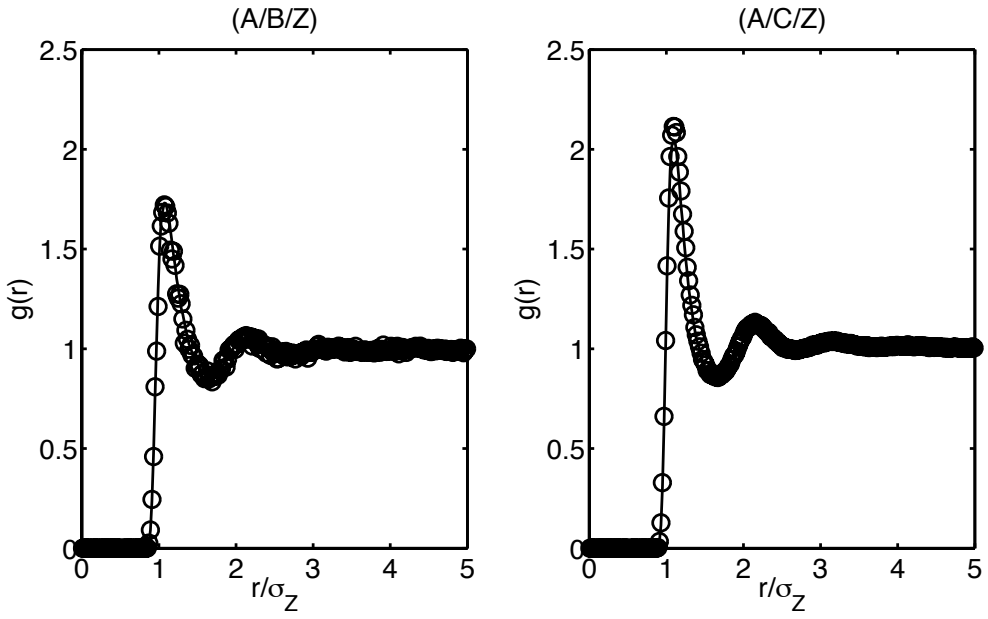

Figure 6

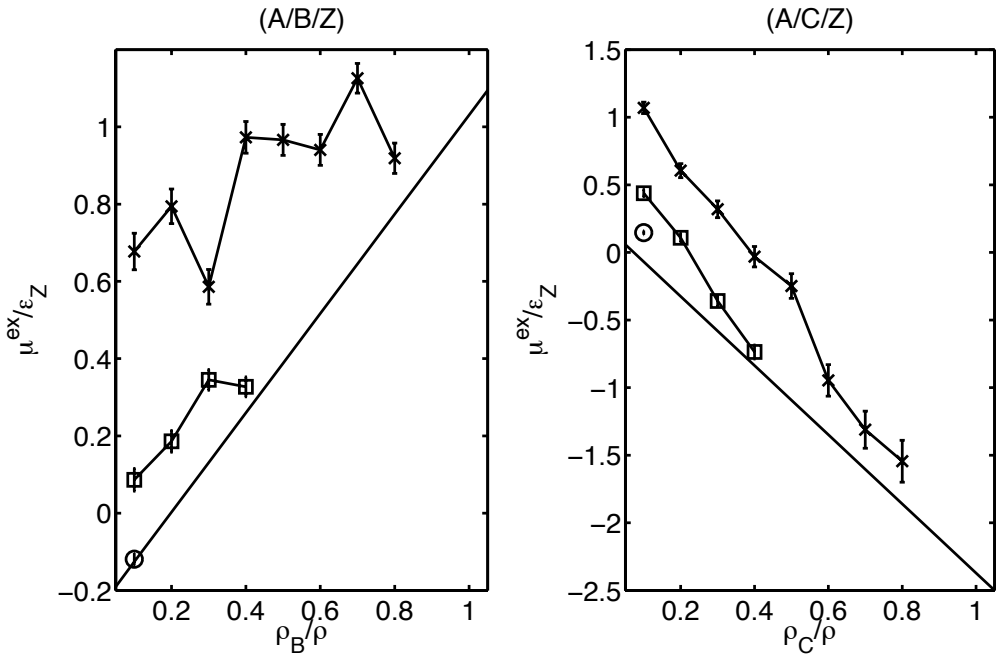

Figure 7

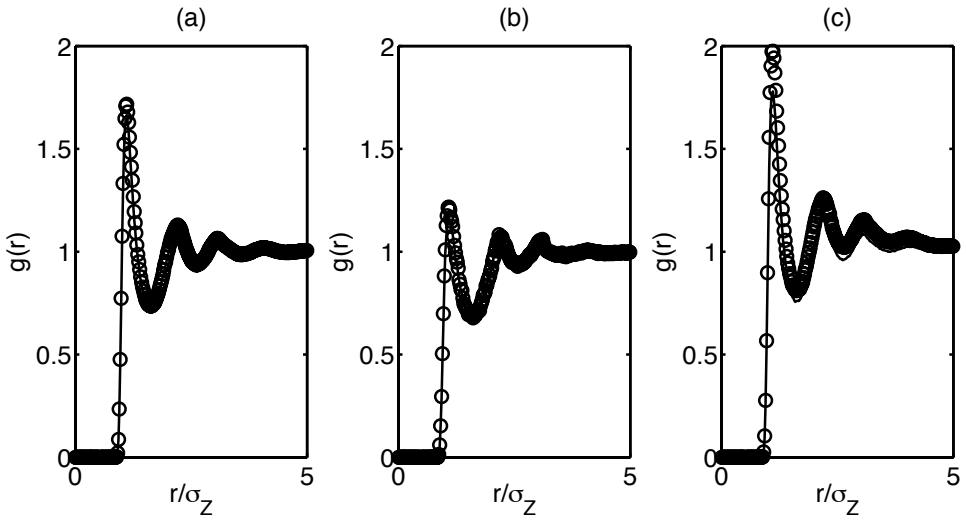

Figure 8

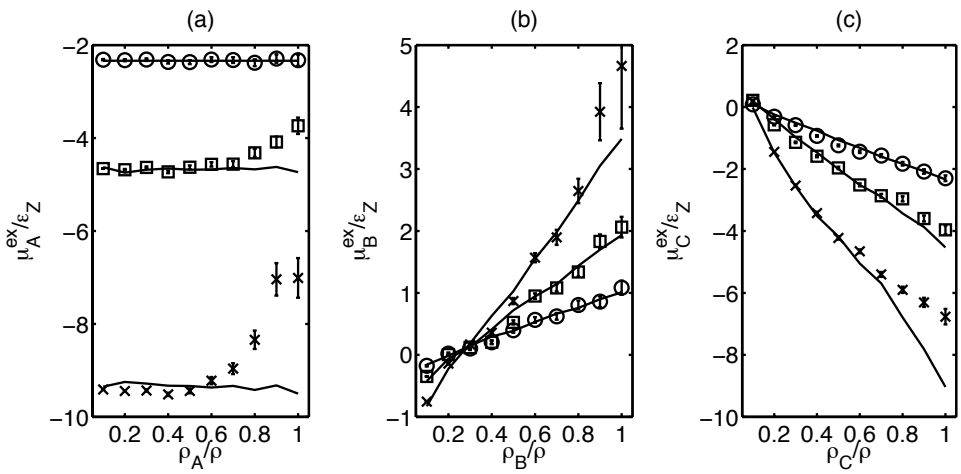
(a)
(b)
(c)
μ_A^{ex}/ε_Z
μ_B^{ex}/ε_Z
μ_C^{ex}/ε_Z
ρ_A/ρ
ρ_B/ρ
ρ_C/ρ
0.2 0.4 0.6 0.8 1

Figure 9

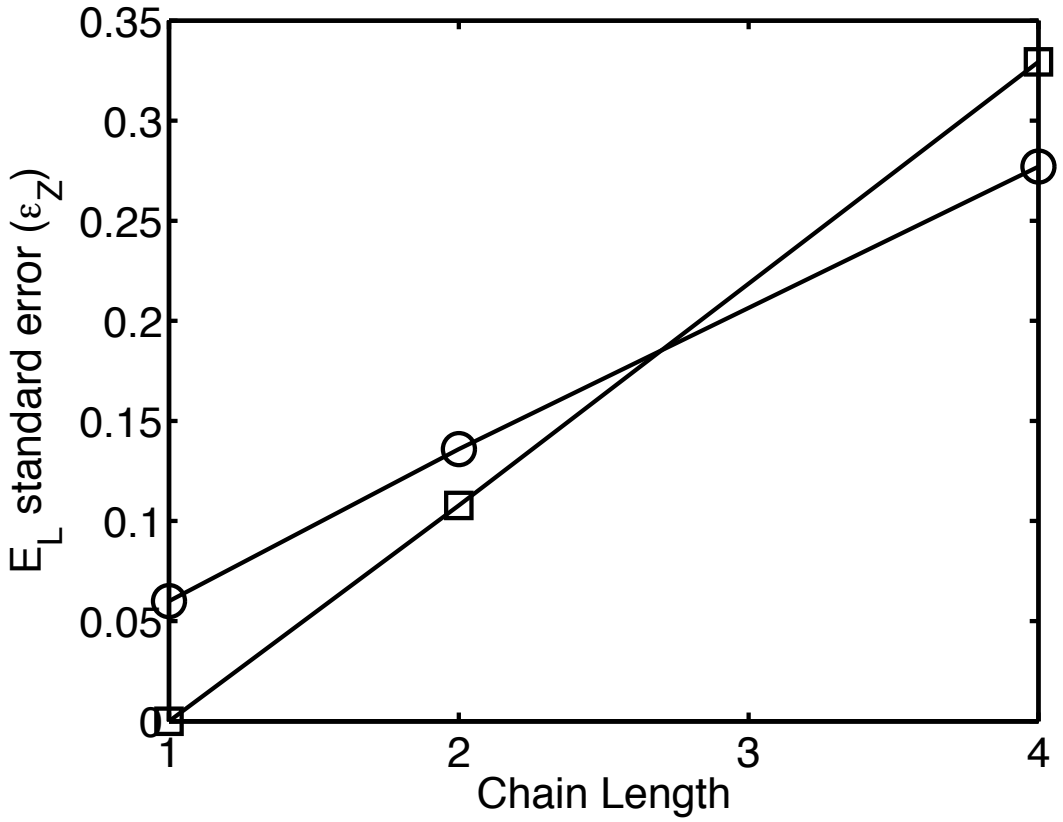
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
Chain Length
E_L standard error (ε_Z)

Figure 10

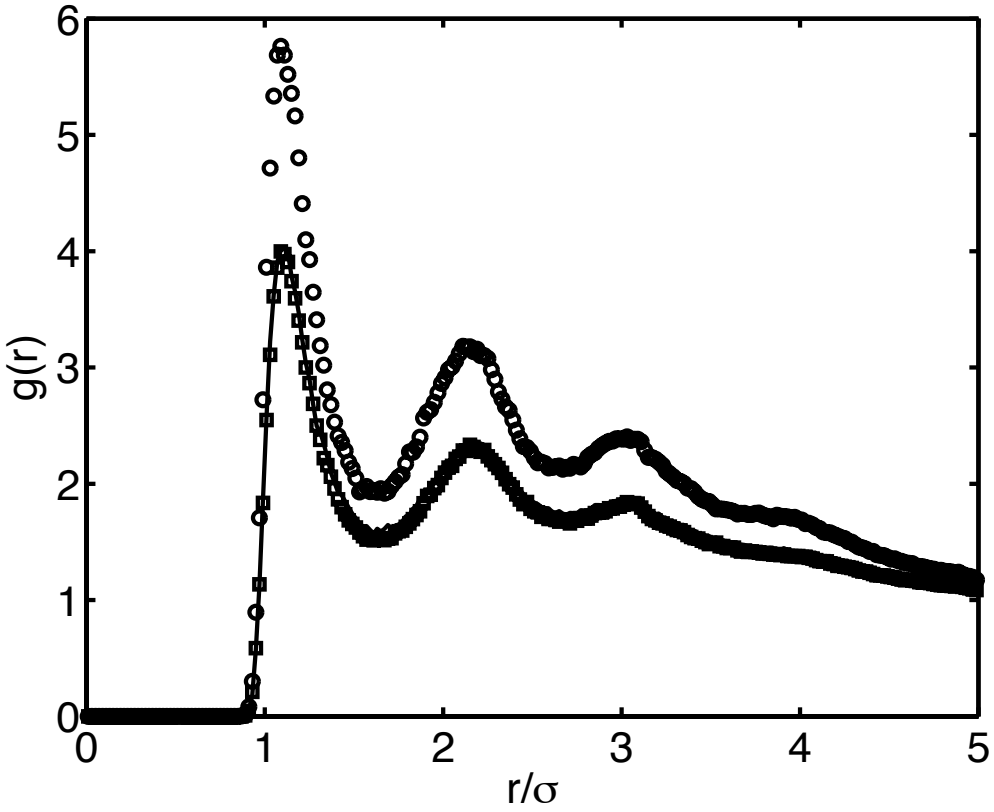

Figure 11

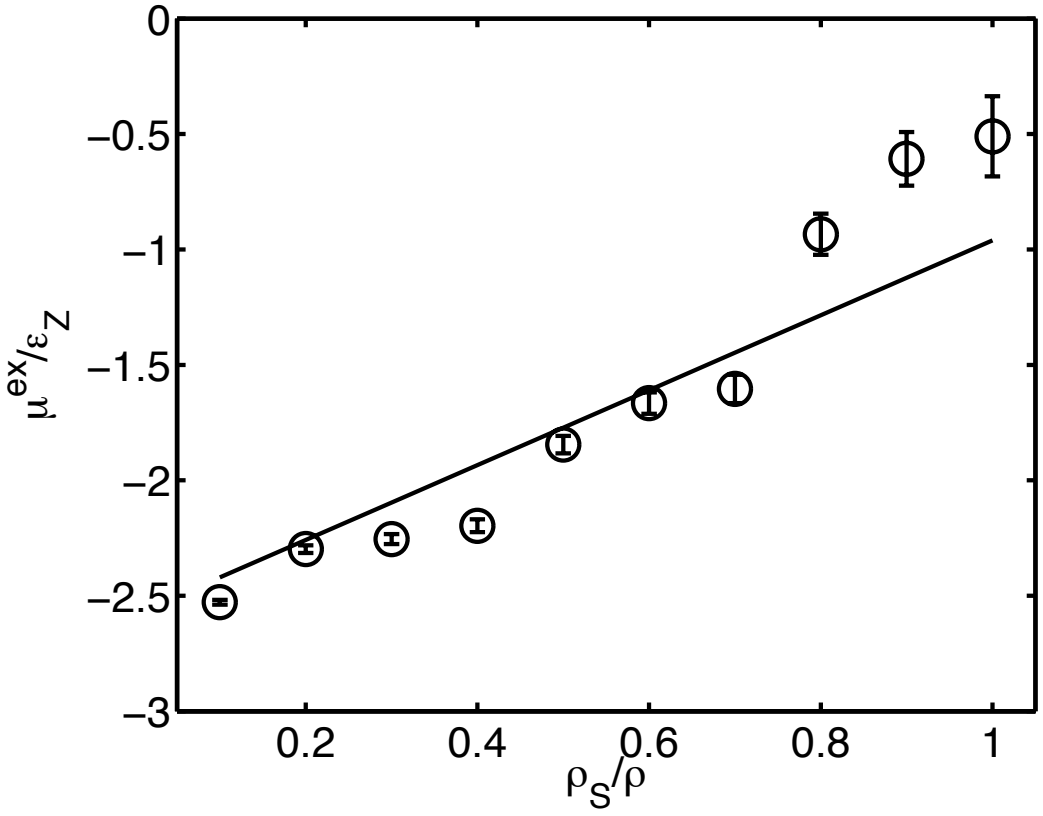

Figure 12

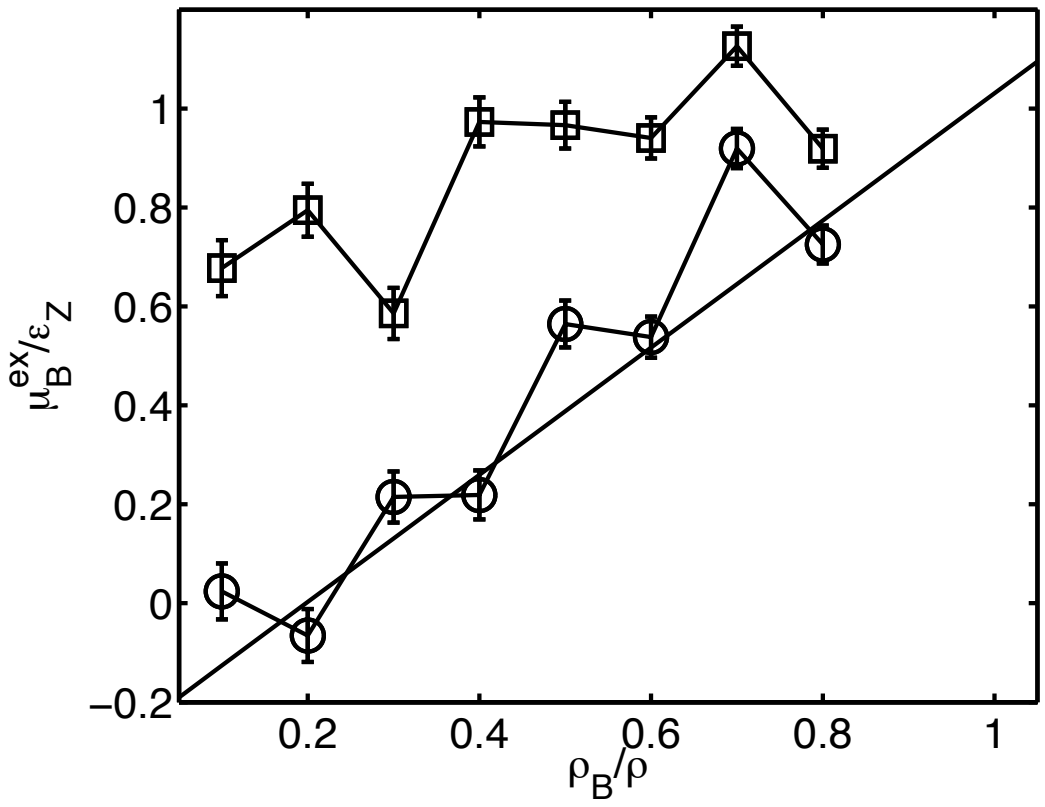

Figure A1

Particle Type	ε_{ii}	σ_{ii}	R_C/σ_{ii} ***(Like)***	R_C/σ_{ij} ***(Unlike)***	$\Delta G_S(Z \rightarrow S)$	$\langle N_L \rangle / \langle N_L \rangle^{IDEAL}$
A	1.0	1.0	5.0	5.0	0.0	1.00
B	0.5	1.0	5.0	5.0	$1.29 \pm 0.05\varepsilon_Z$	1.00
C	1.0	1.0	5.0	1.84	$-2.71 \pm 0.08\varepsilon_Z$	1.01

Table 1

E_L/ε_Z	*N=1*	*N=2*	*N=4*
A	0.015	0.04	0.06
B	0.005	0.010	0.011
C	0.06	0.14	0.28

Table 2

μ^{ex}/ε_Z	*N=1*	*N=2*	*N=4*
A	0.04	0.19	0.27
B	0.04	0.06	0.11
C	0.09	0.13	0.23

Table 3