FLUCTUATION SOLUTION THEORY

by

ELIZABETH ANNE PLOETZ

B.S., Kansas State University, 2010
M.S., Kansas State University, 2010

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Abstract

The Kirkwood-Buff (KB) theory of solutions, published in 1951, established a route from integrals over radial (pair) distribution functions (RDFs) in the grand canonical ensemble to a set of thermodynamic quantities in an equivalent closed ensemble. These "KB integrals" (KBIs) can also be expressed in terms of the particle-particle (*i.e.*, concentration or density) fluctuations within grand canonical ensemble regions. Contributions by Ben-Naim in 1977 provided the means to obtain the KBIs if one already knew the set of thermodynamic quantities for the mixture of interest; that is, he provided the inversion procedure. Thus, KB theory provides a two-way bridge between local (microscopic) and global (bulk/thermodynamic) properties. Due to its lack of approximations, its wide ranging applicability, and the absence of a competitive theory for rigorously understanding liquid mixtures, it has been used to understand solution microheterogeneity, solute solubility, cosolvent effects on biomolecules, preferential solvation, *etc*. Here, after using KB theory to test the accuracy of pair potentials, we present and illustrate two extensions of the theory, resulting in a general Fluctuation Solution Theory (FST). First, we generalize KB theory to include two-way relationships between the grand canonical ensemble's particle-energy and energy-energy fluctuations and additional thermodynamic quantities. This extension allows for non-isothermal conditions to be considered, unlike traditional KB theory. We illustrate these new relationships using analyses of experimental data and molecular dynamics (MD) simulations for pure liquids and binary mixtures. Furthermore, we use it to obtain conformation-specific infinitely dilute partial molar volumes and compressibilities for proteins (other properties will follow) from MD simulations and compare the method to a non-FST method for obtaining the same properties. The second extension of KB theory involves moving beyond doublet particle fluctuations to additionally consider triplet and quadruplet particle fluctuations, which are related to derivatives of the thermodynamic properties involved in regular KB theory. We present these higher order fluctuations obtained from experiment and simulation for pure liquids and binary mixtures. Using the newfound experimental third and fourth cumulants of the distribution of particles in solution, which can be extracted from bulk thermodynamic data using this extension, we also probe particle distributions' non-Gaussian nature.

FLUCTUATION SOLUTION THEORY

by

ELIZABETH ANNE PLOETZ

B.S., Kansas State University, 2010

M.S., Kansas State University, 2010

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry

College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2014

Approved by:

Major Professor

Professor Paul E. Smith

Abstract

The Kirkwood-Buff (KB) theory of solutions, published in 1951, established a route from integrals over radial (pair) distribution functions (RDFs) in the grand canonical ensemble to a set of thermodynamic quantities in an equivalent closed ensemble. These "KB integrals" (KBIs) can also be expressed in terms of the particle-particle (*i.e.*, concentration or density) fluctuations within grand canonical ensemble regions. Contributions by Ben-Naim in 1977 provided the means to obtain the KBIs if one already knew the set of thermodynamic quantities for the mixture of interest; that is, he provided the inversion procedure. Thus, KB theory provides a two-way bridge between local (microscopic) and global (bulk/thermodynamic) properties. Due to its lack of approximations, its wide ranging applicability, and the absence of a competitive theory for rigorously understanding liquid mixtures, it has been used to understand solution microheterogeneity, solute solubility, cosolvent effects on biomolecules, preferential solvation, *etc*. Here, after using KB theory to test the accuracy of pair potentials, we present and illustrate two extensions of the theory, resulting in a general Fluctuation Solution Theory (FST). First, we generalize KB theory to include two-way relationships between the grand canonical ensemble's particle-energy and energy-energy fluctuations and additional thermodynamic quantities. This extension allows for non-isothermal conditions to be considered, unlike traditional KB theory. We illustrate these new relationships using analyses of experimental data and molecular dynamics (MD) simulations for pure liquids and binary mixtures. Furthermore, we use it to obtain conformation-specific infinitely dilute partial molar volumes and compressibilities for proteins (other properties will follow) from MD simulations and compare the method to a non-FST method for obtaining the same properties. The second extension of KB theory involves moving beyond doublet particle fluctuations to additionally consider triplet and quadruplet particle fluctuations, which are related to derivatives of the thermodynamic properties involved in regular KB theory. We present these higher order fluctuations obtained from experiment and simulation for pure liquids and binary mixtures. Using the newfound experimental third and fourth cumulants of the distribution of particles in solution, which can be extracted from bulk thermodynamic data using this extension, we also probe particle distributions' non-Gaussian nature.

Table of Contents

List of Figures

List of Tables

List of Supplementary Tables

List of Symbols

<u>Greek</u>

α_i	Polarizability of site i
α_p	Thermal expansivity (isobaric coefficient of volume expansion)
β	$1/RT$ or $1/k_BT$
Γ_{21}	Excess surface adsorption / Preferential binding parameter
$\gamma_{\pm}$	Stoichiometric mean molal (practical activity coefficient)
γ	Surface tension / Molal activity coefficient
Δ	Isothermal-isobaric (Gibbs) partition function
ΔG_{ij}	$G_{ii} + G_{jj} - 2G_{ij}$
Δ_m	Energy-Energy fluctuation
$\Delta_N^D X = X_D - X_N$	Change in property X upon unfolding/denaturation
δ_{ij}	Kronecker delta function
δX_i	Instantaneous fluctuation, $X_i - \langle X_i \rangle$
$\langle (\delta X)^2 \rangle$	Mean square fluctuation of a property X
σ	LJ size-parameter / standard deviation
ε	Dielectric constant / LJ interaction-parameter / Excess internal energy
ε_0	Permittivity of vacuum / free space
ξ_V	$\sum_i N_i \bar{V}_i$
ξ_i	$N_i - m_i N_1$
ξ_H	$E - \sum_i N_i \bar{H}_i = \varepsilon - \sum_i N_i \bar{H}_i^{Ex}$
ζ	$1 + \rho_i G_{ii} + \rho_j G_{jj} + \rho_i \rho_j (G_{ii}G_{jj} - G_{ij}^2)$
η	$\rho_i + \rho_j + \rho_i \rho_j (G_{ii} + G_{jj} - 2G_{ij})$
κ_T	Isothermal compressibility
Λ	Thermal de Broglie wavelength
λ_i	Absolute activity, $\exp(\mu_i/RT)$
$\boldsymbol{\mu}$	Dipole moment
μVT	Grand canonical ensemble
μ_i	Chemical potential of component i
μ_{ij}	Chemical potential derivative, $\beta(\partial\mu_i/\partial x_j)_{p,T}$
Ξ	Grand canonical partition function
ρ	Mass density / Sum of number densities for all components
ρ_i	Number density of i
$\rho_i^{(n)}$	Probability that any particle is within $d\{r\}$ at $\{r\}$
ϕ_i	Volume fraction of i

<u>Latin</u>

$_+$	Cation
$_-$	Anion
*	Pseudo (as in chemical potential) / Interface / Reduced quantity
^{o}	Pure / Standard state
$^\infty$	Infinitely dilute (limiting)
$\langle \rangle$	Average over ensemble or over time
$\{\}$	Set notation
1	Solvent
$1/RT$	β
2	(Biomolecular) solute
$a_{\pm}$	Mean activity of electrolyte in solution
a_i	Activity of i
(aq)	Aqueous solution

B_{ij}	Particle Doublet Fluctuations
bulk	Bulk
C_p	Constant pressure heat capacity
C_{ijk}	Particle Triplet Fluctuations
C_V	Constant volume heat capacity
D	Denatured
D	Activity derivative
D_i	Diffusion constant
D_{ijkl}	Particle Quadruplet Fluctuations
E	Energy (In CM, KE + PE; In thermodynamics, the internal energy, heat + work)
$\boldsymbol{E}$	Electric field
e	Elementary charge
eq	Equilibrium
F_i	Particle(Number)-Energy fluctuation
f_i	Lewis-Randall (mole fraction) activity coefficient
G	Gibbs free energy
g	Gas or vapor
$g_i(z)$	Relative probability
g_{ij} / $g_{ij}^{(2)}$	Radial (pair / two-body) distribution function
$g_{ijk}^{(3)}$	Three-body distribution function
$g_{ijkl}^{(4)}$	Four-body distribution function
G_{ij}	Integral over the pair (doublet) probability distribution, RDF integral, KBI
G_{ijk}	Integral over the triplet probability distribution
G_{ijkl}	Integral over the quadruplet probability distribution
H	Enthalpy
h	Planck's constant
I	Indistinguishable ion
id	Ideal (mole fraction)
K	Equilibrium constant
k_B	Boltzmann constant
l	Liquid
L	Box length
m_i	N_i/N_1 (Dimensionless) molality
mix	Mixing process
N	Native
n_c	Number of components in the system
N_A	Avogadro's number
N_i	Number of entities (usually molecules, atoms, or ions)
n	Number of cations/anions, $n = n_+ + n_-$
n_i	Amount of substance i
N_{ij}	Excess coordination number = $\rho_j G_{ij}$
NpT	Isothermal-isobaric (Gibbs) ensemble
NVE	Microcanonical ensemble
NVT	Canonical ensemble
$P(\cdot)$	Probability
$P_{\alpha\beta}$	Element of the pressure tensor
p	Momentum / Pressure
Q	Canonical partition function
q, q_i	Generalized coordinates / Atomic partial charge
q^{int}	Internal partition function
R	Gas constant
r	$|\boldsymbol{r}_1 - \boldsymbol{r}_2|$, distance between COM of molecules
S	Entropy
T	Temperature

T_m	Melting temperature
U	Internal energy
V	Volume
v	vapor
X^*	Reduced quantity X
$\bar{X}_i$	Partial molar property of X
$\bar{X}_i^{\infty}$	Infinitely dilute partial molar property of X
X^{Ex}	X excess
X_m	Molar quantity
$\langle X_i \rangle_j$	Average X for species i in the presence of j
$\langle X_i \rangle_\mathrm{o}$	Average X for species i in the absence of j, *i.e.*, in pure solvent
x_i	Liquid phase composition
x, y, z	Cartesian space coordinates, vector $\boldsymbol{r}$
y_i	Gas phase composition / Molar activity coefficient
$z_{+/-}$	Charge of cation/anion

Acronyms/Abbreviations

AA	All-atom
AMBER	Assisted Model Building with Energy Refinement, a set of force fields
Ben	Benzene
BPTI	Basic (often Bovine) Pancreatic Trypsin Inhibitor, a protein
CHARMM	Chemistry at HARvard Molecular Mechanics, a set of force fields
CM	Classical Mechanics
COM	Center of mass
COS	Charge-On-Spring
COS/B2	COS water model (B2 meaning unknown)
COS/G2	COS water model, based upon COS/B2 (G2 meaning unknown)
COS/M	COS Methanol model
CPC	Condensed Phase COS, a COS-based Methanol model
DCF	Direct Correlation Function
DFT	Density Functional Theory
DO	Drude-Oscillator
EOS	Equation of State
FF	Force Field
FST	Fluctuation Solution Theory (interchangeable with FT or the Fluctuation Theory of Solutions)
FT	Fluctuation Theory (interchangeable with FST or the Fluctuation Theory of Solutions)
GD	Gibbs-Duhem
GROMACS	GROningen MAchine for Chemical Simulations, an MD package
GROMOS	GROningen MOlecular Simulation, a set of FFs
HEWL	Hen Egg White Lysozyme, a protein
HOH	Water
HS	Hard Sphere
KB	Kirkwood-Buff
KBFF	Kirkwood-Buff Force Field
KBI	Kirkwood-Buff integral
LINCS	LINear Constraint Solver
LJ	Lennard-Jones
MC	Monte Carlo
MD	Molecular Dynamics
MM	McMillan-Mayer / Molecular Mechanics
MOH	Methanol
MSD	Mean Square Displacement
NMA	*N*-methylacetamide
NMR	Nuclear Magnetic Resonance
OPLS	Optimized Potentials for Liquid Simulations, a set of force fields

PDB	Protein Data Bank
PME	Particle-Mesh Ewald
PMV	Partial molar volume
PT	Perturbation Theory
QM	Quantum Mechanics
RDF	Radial distribution function
RK	Redlich-Kister
RMS	Root Mean Square
RMSD	Root Mean Squared Deviation
SI	Symmetric ideal
SPC	Simple Point Charge, a water model
SPC/E	Simple Point Charge / Extended, a water model based on SPC water
SPT	Scaled particle theory
TIP3P	Transferable Intermolecular Potential 3-Point, a water model
TIP4P	Transferable Intermolecular Potential 4-Point, a water model
UA	United-atom
VDW	van der Waals

Acknowledgements

Dr. Smith: It has been a blast (and a wild ride). There is no way I could possibly distill here how grateful and thrilled I am that I have been blessed to be able to do my graduate studies under your guidance and mentoring. Thank you.

I might describe you by saying, "He epitomizes the messaging of the Life is good® company." Actually, your easy-going, not critical of others, *etc*. personality almost seems incompatible with how rigorous, principled, and thoughtful you are in your own research. I think it is a rare (great!) combination. The depth of KB / FST and the well thought out extensions that you have put forth and are exploring astound me. I love *trying* to keep up. You have stretched me and kept me on my toes every day. My brain hurts most days because of it!

You have always made me feel competent and like I could learn to do whatever I needed to do to accomplish a project – like I could do things. That is powerful. My natural self-talk is usually, "I'm just a girl/student/*etc*., what do I know about this? Why am I lucky enough to be the one that gets to work on this project? Someone else would be way better/faster." I appreciate you believing in me and not giving up on me even though I make so many mistakes, still do not understand so many things, and forget so many other things. Thanks also for letting me express my true opinion and argue with you and others. I know I can be annoying and disagreeable. Thanks for putting up with me.

At the risk of coming across as inappropriate, the best way I can summarize my sentiments would be to say, "Watch the 1967 movie, 'To Sir, With Love.'" I have loved this movie/song since I was a young girl. A teacher I worked with told me that every teacher loves this movie and wants this type of impact to be their legacy. [Some find this movie/theme song creepy. For the record, I mean this in a completely proper way. I respect you very much and you have made a big impact on my life, like Mr. Thackeray made as a teacher on the lives of a class of unruly teenagers.]

Let me close with a story about Alice Stewart, a 1950s physician and epidemiologist in Oxford, who discovered that the (then) rising incidence of childhood cancers was due to mothers being X-rayed while pregnant.

> "So for 25 years Alice Stewart had a very big fight on her hands. So, how did she know she was right? Well, she had a fantastic model for thinking. She worked with a statistician named George Kneale, and George was pretty much everything that Alice wasn't. So Alice was very outgoing and sociable, and George was a recluse. Alice was warm, very empathetic with her patients. George, frankly, preferred numbers to people, but he said this fantastic thing about their working relationship. He said, 'My job is to prove Dr. Stewart wrong.' He actively sought disconfirmation: different ways of looking at her models, at her statistics, different ways of crunching the data in order to disprove her. He saw his job as creating conflict around her theories. Because it was only by not being able to prove that she was wrong, that George could give Alice the confidence she needed to know she was right… It's a fantastic model of collaboration – thinking partners who aren't echo chAmbers. I wonder how many of us have, or dare to have, such collaborations. Alice and George were very good at conflict. They saw it as thinking." – "Conflict as Thinking" by Margaret Heffernan at TEDGlobal 2012

I am not that good yet, but I will strive to be more like George for the rest of the time that I have left. I am envious of your currently remaining and future students. They are in for a fully rewarding journey. I am excited to see what comes of all of this. The more deeply I examine FST, the better it gets. Few things in life possess this quality.

Tobe Eggers: Thank you so much for creating the custom wiring harness for our Backblaze® inspired Storage Pod. Thanks also for your general help with the rest of the build. I now have much greater peace of mind about our data.

1. Introduction

1.1 Motivation

Liquids and liquid mixtures have turned out to be much more difficult to model than have gases or solids, due to their strong intermolecular interactions and high degree of orientational and translational disorder. The desire to interpret and be able to predict the properties of liquids and liquid mixtures at both the microscopic and macroscopic levels is so general and encompasses so many areas of research that it threatens to lose any potential reader's interest by its very vagueness. However, if a theory were found that could bridge the knowledge gap on how microscopic properties give rise to thermodynamic properties, and *vice vera*, and do so in a way that could be applied to essentially any liquid mixture, we students of chemistry may be wise to hold our specific application areas in the back of our minds for a moment and to investigate a bit further until we find a catch or a snag in the theory that renders it useless to us.

This dissertation is an attempt to do just that for the Kirkwood-Buff (KB) theory of solutions and generalizations/extensions of KB theory called Fluctuation Solution Theory (FST). KB theory was first published in 1951, and yet many readers may be unfamiliar with it. This is, in and of itself, enough to cast doubt upon its usefulness and/or credibility. Indeed, many solution theories have risen and fallen over the decades, and most of today's chemists probably do not feel that they have time to learn the nuances of them all.

What sets KB theory apart is two things. First, it does not make approximations [beyond the usual statistical mechanical (SM) approximation of being in the thermodynamic limit] or simplifications. The KB theory equations arise from the manipulation and rearrangement of statistical mechanical and thermodynamic relationships without making any assumptions and without introducing any parameters. It is exact; meaning, it is exact just as one would consider *e.g.*, $\Delta G = \Delta H - T\Delta S$ to be exact. Indeed, the KB/FST theory is the *only* theory available for understanding liquid mixtures that can make such claims. Second, it can be applied to any solution regardless of what type of molecules it contains, how many components it consists of, or the concentrations of those components. The components could be as simple as Lennard-Jones (LJ) spheres, or as complicated as biomolecules; both would be treated identically by the theory.

There is one other theory that is a very limiting case of the more general KB theory, namely, McMillan-Mayer (MM) theory (McMillan and Mayer 1945). In MM theory, the osmotic pressure of a liquid mixture is expanded in terms of the solute concentration and the coefficients of the expansion are called the osmotic virial coefficients. MM theory is entirely analogous to the virial expansion for imperfect gases. It is only applicable to cases of (very) dilute solutions. Very little is known about the osmotic virial coefficients higher than the second coefficient and the second virial coefficient involves cluster integrals of exactly two solute molecules at infinite dilution in pure solvent. Thus, the potential range of applications for MM theory is only a very small subset of those for KB theory and we will not be considering it further.

1.2 What is KB Theory?

As published in 1951 (Kirkwood and Buff 1951), KB theory provided a one-way route from a set of integrals over center of mass based radial distribution functions (RDFs) to a set of thermodynamic properties that one could measure in the wet laboratory. For binary mixtures the relationships are

$$G_{ij} = G_{ji} \equiv 4\pi \int_0^\infty [g_{ij}^{\mu VT}(r) - 1] r^2 dr \approx 4\pi \int_0^R [g_{ij}^{closed}(r) - 1] r^2 dr$$

$$\kappa_T = \zeta / RT\eta$$

$$\bar{V}_i = \frac{1 + \rho_j (G_{jj} - G_{ij})}{\eta}$$

$$\mu_{ii} = \frac{\rho_j RT}{\rho_i V \eta}, \qquad \mu_{ij} = \mu_{ji} = -\frac{RT}{V\eta}$$

$$\eta = \rho_i + \rho_j + \rho_i \rho_j (G_{ii} + G_{jj} - 2G_{ij})$$

$$\zeta = 1 + \rho_i G_{ii} + \rho_j G_{jj} + \rho_i \rho_j (G_{ii} G_{jj} - G_{ij}^2)$$

1.1

where the G_{ij}s are called the Kirkwood-Buff integrals (KBIs), g_{ij} are the center of mass based RDFs, κ_T is the isothermal compressibility, R is the gas constant, T is the temperature, $\bar{V}_i$ is the partial molar volume of component i, $\rho_i = <N_i>$ / V is the number density of species i, and μ_{ij} is the derivative of the chemical potential of species i with respect to the number of molecules of species j while holding fixed the number of molecules of species i. The RDFs are labeled with a superscript μVT to denote that the KBIs are defined in terms of RDFs in the grand canonical ensemble (constant set of chemical potentials, constant V, and constant T). If one calculates the KBIs within closed systems by integrating up to some distance R beyond which the RDFs have all converged to unity instead of integrating over all space (as will be done in this dissertation), then they will be calculating a closed KBI that differs from the open KBI by on the order of one over the number of molecules, which is considered a trivial difference.

Since RDFs are short-ranged functions that converge to unity on the order of a few molecular diameters and since thermodynamic properties are clearly bulk, global properties, KB theory therefore provided a means to understand how the microscopic properties of a specific mixture at a specific state point give rise to that mixture's unique macroscopic properties.

Note that the KBIs may also be written in terms of particle fluctuations

$$G_{ij} = G_{ji} \equiv V \frac{\langle \delta N_i \delta N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - \frac{\delta_{ij}}{\rho_i}$$

1.2

Where V is the volume of the grand canonical ensemble, $<>$ denotes time or ensemble averages, and $\delta N_i = N_i - \langle N_i \rangle$ is the instantaneous number of molecules of species i within the fixed volume minus the average number present.

KB theory has not achieved the level of popularity that one might assume based upon the long list of strengths it seems to have. One reason for this is that it requires RDFs as input. At the time of its publication, computers were not available for the simulation of RDFs, and many research endeavors at the time were instead specifically focused on

how to calculate an RDF. This diminished the attractiveness of the theory immensely, since the input was generally unavailable. Another reason for the lack of interest initially was that there was controversy for some time as to whether or not KB theory could be applied to salt solutions. It can, but it was not immediately obvious to the research community and it took some time to sort this out.

In 1977, Arieh Ben-Naim established the inversion of KB theory, meaning that if one knew the relevant thermodynamic properties of the mixture of interest, one could go "backwards" to obtain the integrals over the RDFs (Ben-Naim 1977). The RDFs themselves could not be determined, only the KBIs. This means that one does not obtain distance dependent information from the inversion procedure, but does obtain the net affinity of i for j at the local level by measuring the local composition near a central i molecule relative to the bulk composition. A large positive G_{ij} thus means that there must have been a net ordering of j around i, or equivalently aggregation of j and i, whereas negative values for G_{ij} indicate that j was not strongly associated with i. For a binary mixture (composed of species 1 and 2), the inversion procedure is given by

$$G_{ij} = RT\kappa_T - \frac{\delta_{ij}}{\rho_i} + \rho RT \frac{\left(1-\rho_i\bar{V}_i\right)\left(1-\rho_j\bar{V}_j\right)}{\rho_i\rho_j\mu_{ij}} \qquad 1.3$$

where δ_{ij} is the Kronecker delta function ($\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$) and $\rho = \rho_1 + \rho_2$.

1.3 A Very Brief History of Solution Theories

Prior to MM and KB theories, there were really no competitive theories of liquid mixtures. Figure 1.1 provides a timeline of some of the key developments in solution theory. Recently, Mazo described solution theory history in terms of an "Ancient" period, an "Early Modern" period, and a "Modern" period (Smith *et al.* 2013). Mazo marks the start of the Ancient period with the discovery of Henry's law, due to it being the first quantitative law explaining the properties of solutions (Henry 1803). During this time van der Waal's thesis was also published and the Lorentz and Berthelot combination rules were developed (Gubbins 2013), which were an attempt to improve intermolecular potentials when they were applied to mixtures (van der Waals 1873). Mazo suggests that the Ancient period ended with the discovery of Raoult's law (Raoult 1878, 1882, 1887, 1888). This was approximately the same time that van't Hoff was deriving the equation relating the concentration of a solute to a solution's osmotic pressure and just after he had discovered three of the colligative properties (Van't Hoff 1887, 1894).

The Early Modern period involved the work of Gibbs, who began developing SM in the early 1900s (Gibbs 1902; Gibbs 1948). During this period. SM proved to be very successful for gases, but less so for liquids and liquid mixtures (Smith *et al.* 2013). The first attempt to approxicate the excess Gibbs free energy of mixing was also published (by van Laar) during this period (Van Laar 1906), but it was tied to the VDW EOS and, therefore, had quite limited accuracy. In the late 1920s, Hildebrand's Regular Solution Theory was developed (Hildebrand and Sharma 1929), which broke away for the VDW EOS but has many assumptions itself, including that the RDFs expressed as a function of reduced separation distances (σ_{ij}) are independent of the species (Gubbins 2013). During this period, success was also being achieved for the understanding of solids (Gubbins 2013). Development of the Ising model led to attempts to use lattice

theories as models for liquids as well (Ising 1925). Ultimately this was considered very unsuccessful, since we now know that liquids are very unlike solids (Gubbins 2013). In the 1930s, Mayer derived the virial equation for imperfect gases, which would later lead to his and Mayer's development of the analog for solutions, MM theory (McMillan and Mayer 1945; Smith *et al.* 2013).

The ultimate goal in solution theory would be that all one would need to use as input to a theory would be the molecular properties of the constituent molecules in a mixture (as well as the conditions such as the temperature and pressure) and that from that input, one could explain and predict all of the mixture's properties (Ben-Naim 2006). The above attempts from the Ancient and Early Modern periods largely fall into that category, but were generally unsuccessful. However, during the Early Modern period Kirkwood (and others) were developing the correlation function approach (Kirkwood 1935; Smith *et al.* 2013). Ultimately these would be used in less ambitious (but more successful) theories of solutions (Ben-Naim 2006). Instead of only requiring information about the properties of the molecules that make up a solution mixture, now additional input in the form of the correlation functions could be used.

The development of KB (and MM) theory is what marks the beginning of the Modern period of solution theory according to Mazo (Smith *et al.* 2013). Unlike all of the previous theories (and those that were not mentioned here), KB theory does not make assumptions such as stating that certain excess properties of mixing are zero, that the components of the mixture must be atomic or that they must be nonpolar, *etc.* (see *e.g*, (Hildebrand and Sharma 1929)). It requires absolutely no knowledge about how the molecules interact with each other (the intermolecular potentials), not even that the interactions be pairwise additive. It says nothing about the shape of the molecules (such as, that they must be spherical) and it makes no requirements that the must be rigid, *etc.* In this sense, it is not appropriate to group KB theory with the other types of theories mentioned above. It is simply done here in an attempt to provide a bit of historical perspective.

A short while later, the very first computer simulations of hard spheres (HS) and Lennard-Jones (LJ) spheres were reported (Smith and Lea 1960; Singer 1969). It was not until the development of sufficiently sophisticated computers and computational techniques that would allow for the simulation of RDFs that KB theory could be used without the inversion procedure. In 1990 a monograph was published filled with examples of KB theory's uses (Matteoli and Mansoori 1990), followed by a second monograph in 2013 that was filled with updates such as applications to biomolecules (Smith *et al.* 2013), which were nonexistant in the 1990 monograph.

One modern theory that is experiencing a lot of success is the Statistical Associating Fluid Theory (SAFT) (Chapman *et al.* 1989, 1990), which is an EOS for mixtures that has been used widely in the chemical engineering community and has been applied to over one hundred pure fluids and their mixtures, including molecular and polar liquids. SAFT is based upon Wertheim's off-lattice theory called "Fluids with Highly Directional Attractive Forces" and it involves an expansion in the activities of a mixture's components (Wertheim 1984b, 1984a; Gubbins 2013). Numerous

extensions to SAFT have been proposed. It is an active area of ongoing work, the ultimate success of which it is probably too early to decide; however, it really falls into a separate category from KB/FST theory.

Traditional KB theory is an isothermal theory. Building upon previous work of Buff and Brout and of Debenedetti (Buff and Brout 1955; Debenedetti 1986, 1987a, 1987b, 1988; Debenedetti 1989), in 2011 Paul E. Smith published the analog of the KB inversion procedure with the removal of the isothermal restriction. This dissertation will illustrate the first attempts to use this new extension (see Chapters 4 and 5).

To our knowledge, no new theories of solution mixtures have appeared that offer as many advantages as does KB/FST theory. FST has now been widely used to understand the basic properties of solutions (Matteoli and Mansoori 1990), to understand the effects of additives on the solubility of solutes (from small hydrocarbons to proteins) and biomolecular equilibria (O'Connell 1971a; Chialvo *et al.* 1996; Chitra and Smith 2001a; Ruckenstein and Shulgin 2002; Shimizu and Smith 2004; Shulgin and Ruckenstein 2005; Mazo 2006), to investigate the local composition of solutions in the context of preferential solvation (Marcus 1989), to study the effects of additives on the surface tension of liquids (Chen and Smith 2008), to study critical behavior (Cochran *et al.* 1987; Debenedetti 1987a; O'Connell and Liu 1998; Baird and Kim 2003), to interpret computer simulation data (Smith 1999; Chitra and Smith 2001a; Weerasinghe and Smith 2003a; Ganguly *et al.* 2011), to develop models for many of the above effects (Smith 2004b), and to provide expressions for the volume, enthalpy, compressibility and heat capacity corresponding to simple association equilibria (Gee and Smith 2009; Jiao and Smith 2011). This list is not exhaustive.
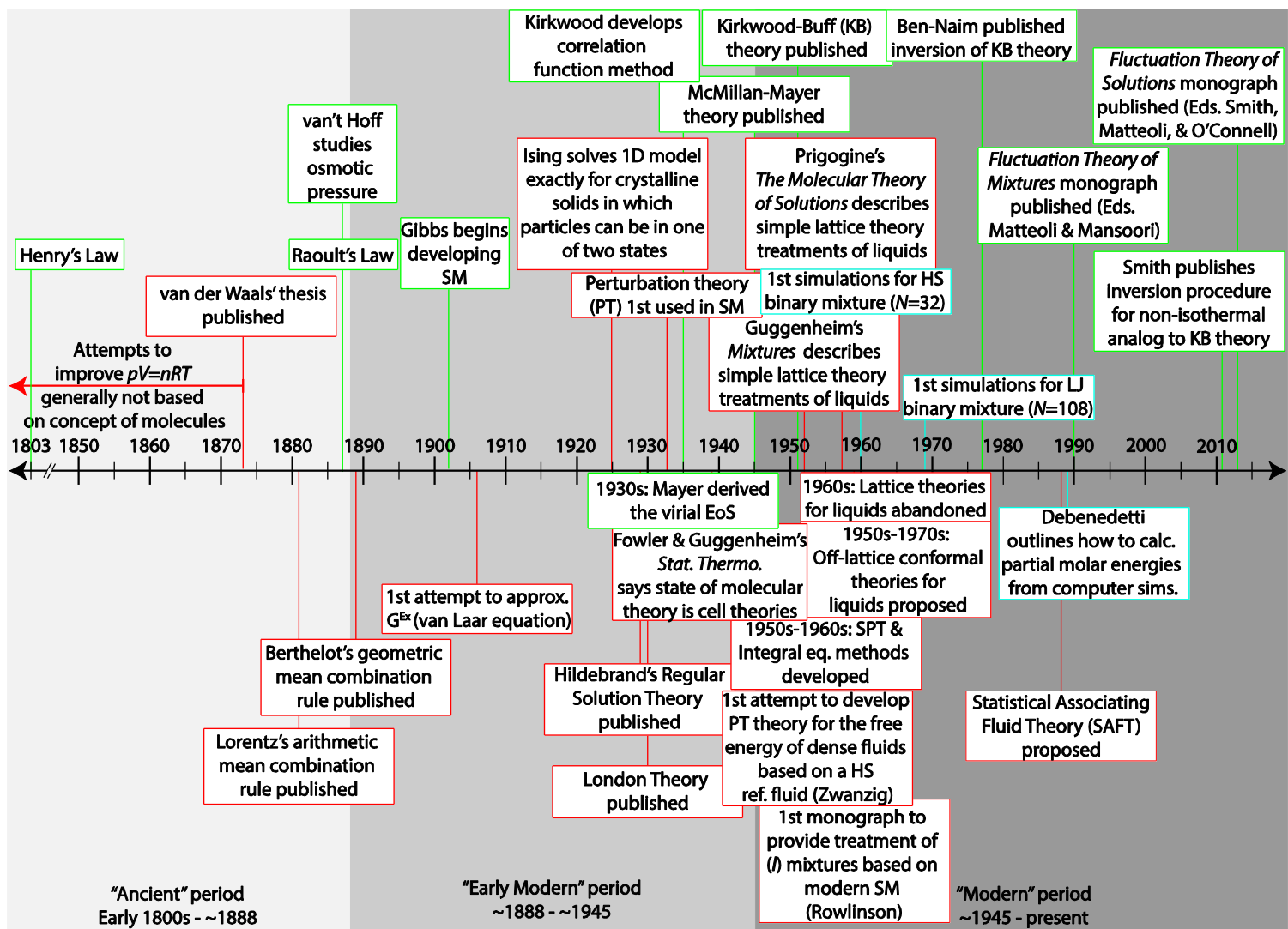
Figure 1.1 Timeline of selected advancements in solution theory. References within the timeline that are not included in the text include the following: (London 1930; Fowler and Guggenheim 1939; Prigogine 1957; Rowlinson and Swinton 1982). The references for the Lorentz and Berthelot combination rules may be found within (Gubbins 2013). For conformal theories see *e.g.*, (Longuet-Higgins 1951). For PT see *e.g.*, (Zwanzig 1954; Buff and Schindler 1958; Mazo 1958).

1.4 Problems with KB/FST Theory

It is never a good idea to list all of the strengths of a thing and to not discuss its weaknesses just as plainly. The issues with KB/FST Theory are not theoretical, but there are often practical issues with its use. These are enumerated here.

When going in the direction of the {RDFs} to the {KBIs}, one still has the question of how to obtain the input RDFs. Experimental RDFs are available for some liquids and very few liquid mixtures, but generally not the liquid mixtures in which we are actually interested (mixtures composed of molecular and/or polar liquids). Therefore, we perform molecular dynamics simulations of the systems we are interested in and use the simulated trajectories to obtain the RDFs. This means that the accuracy of the simulated RDFs that we obtain will depend upon the accuracy of the force field that was used to generate the trajectories of the molecules within the system. The force fields that are used are simple models that have been created to balance the need for accuracy with the need for computational efficiency and are themselves filled with approximations. However, one can develop force fields intelligently to reproduce the KBIs, which gives increased hope that the intermolecular potentials used in the simulation are of increased accuracy and the RDFs will be as well. The method to do so, pioneered by Smith, is discussed in Chapters 2 and 3.

Assuming that one has an accurate description of the intermolecular potentials, the second issue with the simulation of the RDFs/KBIs is their relatively slow convergence with respect to time and distance. This means that one may need to simulate systems for on the order of tens of nanoseconds in order for the KBIs to converge and that one may need to use relatively large simulation boxes. For binary mixtures of small molecules, examples might be on the order of tens of nanoseconds and ~6-10 nm in length simulation boxes. The exact requirements will depend upon the specific system studied. Less ideal systems and systems with larger molecules will generally need comparatively larger boxes and longer simulation times than those of more ideal systems and/or smaller molecules. Additionally, solutes that are at low concentrations require larger system sizes and longer simulation times to converge than would a mixture of the same components at more equally balanced compositions. This is simply an issue of poor statistics for the species with fewer molecules present. These issues are discussed in Chapter 2 and are especially well illustrated in Chapters 3 and 6. We generally have been able to overcome the issues in this paragraph; it simply requires a bit of extra simulation effort.

Lastly, it is largely unknown how precisely one can obtain the KBIs from the experimental thermodynamic data. This is addressed to some extent in this dissertation (see Chapters 3 and 6); however, more work in this area is certainly worth pursuing. The KBIs are sensitive to the experimental excess Gibbs free energy (G^{Ex}) data from which the chemical potential composition derivatives shown in Equation 1.3 are obtained. In addition to having experimental error in the raw *T-p-x-y* Vapor-Liquid Equilibria data (*p* is the pressure, *x* is the liquid phase mole fraction and *y* is the gas phase mole fraction) that is used to obtain the Gibbs free energy, there is a second potential kind of noise/error, which arises from the requirement for an analytical expression to be used to go from G^{Ex} to μ_{ij}. The researcher often has a few different correlating equations to choose from. Their decision of a specific equation will impact, to some

extent, the value of the KBIs obtained from the inversion procedure. The degree to which this is an important issue versus a potential problem is almost entirely unknown at this point in time.

With these practical issues in plain view, we still view KB/FST theory as the most powerful theory of liquid mixtures available. The simulation issues mentioned above will probably greatly diminish with time due to continuing increases in computational power.

It is our view, then, that it is better to use the principled KB/FST theory, with its current but diminishing practical implementation issues, than it would be to use a theory that could be solved for analytically, but contained significant approximations and/or simplifications.

1.5 Organization of Dissertation

Chapters 2 and 3 deal with traditional KB theory. In Chapter 2 we test modern biomolecular force fields' abilities to reproduce the experimental KBIs for a few important mixtures. We argue that if a force field cannot reproduce the experimental KBIs well, then it is not correctly describing the association of molecules in solution, nor the thermodynamic properties of the solution. We then show how force fields can be parameterized to reproduce KBIs over a full range of compositions in binary mixtures and suggest that this gives more confidence that the force field will then describe both structural and thermodynamic properties of simulated systems with improved accuracy. In Chapter 3 we then compare a force field that explicitly accounts for electronic polarization to a non-polarizable force field that was parameterized to reproduce the KBIs. Polarizable force fields are theoretically superior to non-polarizable force fields. In Chapter 3 we test if this theoretical superiority necessarily leads to a practical superiority.

Chapters 4 and 5 deal with the above mentioned extension of KB theory to make it a non-isothermal theory. This extension necessarily involves considering other fluctuations that occur in the grand canonical ensemble besides fluctuations in the number of particles within a fixed volume (those appearing in Equation 1.2). Specifically, it involves particle – energy and energy – energy fluctuations, which leads to our use of the term FST instead of KB theory. Chapter 4 only presents the results for the inversion procedure *i.e.*, where we have analyzed bulk thermodynamic data and from it extracted the local, fluctuating properties. Chapter 5 pairs the analysis of experimental data with results from molecular dynamics simulations.

In Chapter 6 we investigate how the particle-particle fluctuations depend on composition and pressure. Their dependence on temperature is of great interest to us as well, but that will appear elsewhere. (Furthermore, the analogous dependence of the particle-energy and energy-energy fluctuations on temperature, pressure, and composition are of great interest, but will not be presented here.) The pressure and composition derivatives of the particle-particle fluctuations involve triplet particle-particle-particle fluctuations that are non-zero. These triplet fluctuations are the third cumulant (or third central moment) of the distribution of particles in solution. The fact that they are non-zero allows us to draw some conclusions about the way particles are distributed in solution (*i.e.*, it cannot

be a symmetric distribution such as the Gaussian distribution). We also test whether or not it is possible to simulate these higher fluctuations precisely, since it is sometimes difficult to achieve adequate precision for the simulation of the traditional KBIs.

Chapter 7 is actually not about KB theory or FST. Here we test the ability to precisely simulate infinitely dilute partial molar quantities of proteins using a thermodynamically rigorous approach, the (experimentally well known) apparent molar approach. We are unaware of prior applications of this general approach to computer simulations. Traditional computational methods to calculate these properties are scarce for all but the infinitely dilute partial molar volume, which has typically been calculated using subjective approaches. The reason for the inclusion of Chapter 7 in a disseration on FST is that we are in the process of testing an FST-based approach to also calculate these same properties. The FST-based approach has several advantages over the apparent molar approach presented here, and is briefly described at the end of Chapter 7.

Chapter 8 provides a summary of what we have learned so far and of our future directions.

2. Accurate Force Fields for Molecular Simulation[*][†]

Abstract

The analysis of thermodynamic data concerning solutions using Fluctuation Solution Theory (FST) has provided a deeper understanding of local composition in solution mixtures. Over the past decade, we have used this data to help improve the description of intermolecular interactions implemented in common force fields for molecular simulation, with an emphasis on solutes of biological interest. A comparison between experimental and simulation data for small solute-solvent mixtures provides a clear indication of the quality of the force field, provides access to solute (and solvent) chemical potential derivatives with respect to composition, and thereby helps to ensure that the correct balance between solute-solute, solute-solvent, and solvent-solvent distributions is attained. Here we discuss the advantages and disadvantages of using such an approach for force field design and parameterization, and provide simulation results for a series of representative solution mixtures obtained with a variety of currently available biomolecular force fields.

[†] Designed by PES and EAP, written by EAP, edited by EAP and PES. Simulations were similar to those previously performed and analyzed by MK that had remained unpublished and used shorter simulation times, smaller simulation systems, and fewer systems.

2.1 Introduction

Computer simulation is used for a wide range of purposes, from exploring experimentally inaccessible phenomena, to providing an alternative when the corresponding experiment would be prohibitively costly, time-consuming, difficult, dangerous, or controversial. From a biomolecular perspective, molecular modeling techniques are commonly used for rational drug design, for protein structure prediction, and in a host of other scenarios where the atomic level resolution they provide can lend valuable insights into the systems of interest. Using computer simulation, one can even study processes that are impossible or unphysical in real life (*e.g.,* computational alchemy). Thus, the potential of biomolecular computer simulation is undeniable.

Despite the many exciting uses of molecular simulation, there is no formal guarantee that the results will reflect reality. The quality of a Molecular Dynamics (MD) or Monte Carlo (MC) simulation depends on the degree of sampling achieved during the simulation, which we shall not consider here, together with a satisfactory description of the intra- and intermolecular interactions in the system, that is, an accurate force field (FF). Decades have passed since the first MD simulations were performed, but the effort to continuously improve the quality of the underlying FFs continues to this day (Cheatham and Brooks 1998; Wang *et al.* 2001; Mu *et al.* 2003; Mackerell 2004; Allison *et al.* 2011; Schmid *et al.* 2011; Zhu *et al.* 2012). It has been expressed that both the general philosophy and the parameters for biomolecular FFs are probably converging (Wang *et al.* 2001). However, we argue and demonstrate here that there is still significant room for improvement and that a markedly different parameterization philosophy is probably necessary.

2.2 Currently Identified Force Field (FF) Challenges

Two of the main biomolecular FF challenges are: (*i*) The generation of accurate torsional potentials for peptides and proteins and (*ii*) improved parameters for the description of nonbonded interactions. A commonly cited example of the former is Simmerling's (and others) exposure of a bias toward the α-helical secondary structure in the AMBER ff94 and ff99 FFs (Okur *et al.* 2003). Simmerling was able to achieve an improved balance between the possible secondary structure propensities through modification of the torsional potentials, resulting in the AMBER99sb FF (Hornak *et al.* 2006). While incorrect torsional potentials may have contributed to other reported artifacts (Mu *et al.* 2003; Allison *et al.* 2011), they are not addressed here because FST will not help with this part of the parameterization. Our focus will only be on the use of FST/KB theory to help improve intermolecular nonbonded interactions for simple, effective charge, nonpolarizable FFs.

One strict test of the quality of the nonbonded parameters is the ability to reproduce appropriate protein-ligand association interactions and binding free energies. In FF-based computational drug design, it is notoriously difficult to predict binding free energies quantitatively or to rank ligands based upon their binding affinities, even in cases where the correct ligand-receptor pose has been predicted (Lazaridis 2002; Bonnet and Bryce 2004). Consequently, it has been suggested that the nonbonded parameters may be deficient (Lazaridis *et al.* 2002). This issue extends further to question the quantitative features concerning assembly and aggregation equilibria.

Furthermore, because many proteins are only marginally stable (Dill *et al.* 2011), a correct description of nonbonded interactions is necessary for attaining quantitative conformational equilibria, melting temperatures, and other thermodynamic properties associated with protein folding. For example, it was reported that a 28-residue miniprotein that adopts a ββα motif had a simulated melting temperature 84 K higher than experiment using OPLS-AA (Feng *et al.* 2009), while the native structure of the Trp-cage miniprotein has been shown to be less stable than the unfolded state when using the OPLS-AA FF (Juraszek and Bolhuis 2008). The melting temperature of the Trp-cage protein was also overestimated by 130 K using AMBERff94 (Paschek *et al.* 2008), but was in reasonable agreement with experiment when using the AMBER99sb FF due to an apparent cancellation of errors in which the enthalpy and entropy contributions were both a factor of two or more smaller than experiment (Day *et al.* 2010; Zhang and Ma 2010; Paschek *et al.* 2011). The Trpzip2 miniprotein simulated with AMBER99sb has also been reported to have an enthalpy and entropy of unfolding about half the magnitude of the experimental values, and a heat capacity change about $1/10^{th}$ the experimental value (Nymeyer 2009; Zhang and Ma 2010). Interestingly, we have not found a simulated folding equilibrium where both the melting temperature and the underlying entropy and enthalpy contributions agreed with experiment. Considering that protein folding is currently a popular area of biomolecular simulation (Bowman *et al.* 2011), it seems prudent to place the nonbonded interactions under further scrutiny.

Our interest in the quality of FFs grew naturally from a longstanding effort to understand the effects of cosolvents on biomolecules. Examples from the 1990s include Smith and Pettitt's observations of interesting salt-biomolecule interactions in light of the Hofmeister series (Smith and Pettitt 1991, 1992; Smith *et al.* 1993). Unfortunately, it was unclear if some of the results described in these early studies represented real physical behavior or were simply artifacts of the FF (Smith *et al.* 1993).

Consequently, in the early 2000s the Smith group analyzed several solution mixtures from a KB point of view seeking to decipher if the mainstream biomolecular FFs could be used to quantify the preferential interactions occurring in ternary systems of a solvent, biomolecule, and cosolvent. Specifically, Chitra and Smith simulated aqueous mixtures of a series of cosolvents including 2,2,2-trifluoroethanol, urea, guanidinium chloride, sodium chloride, and ammonium sulfate with several commonly available FFs. These studies indicated a general inability of the available FFs to reproduce the experimental KBIs (and hence the chemical potential composition derivatives), resulting from a tendency for the effective solute-solute and solvent-solvent interactions to be too favorable (Chitra and Smith 2001a, 2001c, 2001b). These studies also implied that established small molecule FFs would not be able to model the correct distributions of common cosolvents around biomolecules (Kang and Smith 2007). The general conclusion from these studies was that, while it was common for different FFs to provide similar accuracy for properties such as densities, diffusion constants, dielectric properties, and so forth, the resulting KBIs usually varied greatly (Chitra and Smith 2000, 2001b).

Others have confirmed the general inability of biomolecular FFs to consistently reproduce the experimental KBIs. For example, Perera and coworkers studied acetone-water mixtures with both OPLS and the acetone FF of Klein (Ferrario *et al.* 1990) and found that, while both FFs were able to acceptably reproduce the enthalpies, densities, and diffusion coefficients of the mixtures, neither of the models could reproduce the solution KBIs. Indeed, both models phase separated in the acetone mole fraction range of ~0.3-0.7 (Perera and Sokolic 2004). Furthermore, alternative force fields for urea have been observed to exhibit different quantitative behavior concerning their interactions with small peptides (Horinek and Netz 2011). One of the main advantages of studying solutions using KB theory is the access to chemical potentials, in the form of their composition derivatives.

We note that instead of using KB theory, thermodynamic integration or particle insertion methods could be used to calculate the chemical potentials (Kokubo *et al.* 2007; Joung and Cheatham 2008, 2009). However, these methods often do not have the precision required to detect the typically small changes in the chemical potential with composition, which can often be less than a kJ/mol, without a significant computational investment. Additionally, we emphasize that the chemical potential derivatives with respect to composition, not the chemical potentials themselves, are directly related to the preferential interactions generally used to quantify cosolvent effects on biomolecular systems (Record *et al.* 1998; Timasheff 1998). Finally, while infinite dilution free energies or enthalpies of solvation probe solute–solvent interactions, they do not provide a check of solute–solute interactions. Thus, we prefer to use KB theory as a guide, because it provides a framework to determine the chemical potential changes for solutes of any size throughout a range of compositions.

Current FFs typically have very similar bonded terms. The VDW parameters are usually developed to reproduce the density and enthalpy of vaporization of pure liquids and the crystal structure dimensions. Most of the variability and uncertainty lies in the Coulomb term. Usually, gas phase charges from *ab initio* calculations of the geometries and minimum interaction energies between dimers are scaled to mimic condensed phase charges, and are then tested to ensure that they reproduce the experimental data for pure liquids. Alternatively, condensed phase partial charge distributions are simply adopted from gas phase values that were calculated using a basis set that created erroneously large gas phase charge distributions. We have argued that the primary reason that many current force fields struggle to reproduce the experimental KBIs lies in the approximate nature of the effective charge distributions used to describe solute polarity, that is, the unknown degree of polarization of a solute in a polar solvent (Weerasinghe and Smith 2003a). Consequently, instead of seeking effective pair potentials that reproduce the KBIs, one could pose the legitimate argument that explicitly polarizable FFs should naturally reproduce the KBIs better than any nonpolarizable FF. We agree with this argument in theory. In practice, the increased computational demand of explicit polarization may prohibit those investigators seeking to reach long simulation timescales (a significant subset of the community) from adopting polarizable FFs (Freddolino *et al.* 2010). This would be especially true for those interested in implicit solvent simulations where explicit polarization is more difficult to include. Polarizable FF are currently available and examples include Ponder's Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) FF,

which is currently only available for organic molecules (Ren *et al.* 2011), and the CHARMM polarizable FF for proteins, nucleic acids, and lipids, which is nearly completed (Zhu *et al.* 2012).

2.3 Simulated Binary Mixture Kirkwood–Buff Integrals for Popular Biomolecular Force Fields: A Case Study

Our previous studies have investigated a range of FFs for simple cosolvents that often interact with peptides and proteins. Here, we ask the question: how well do current biomolecular FFs reproduce the binary solution properties for small solutes representative of amino acid sidechains? Biomolecular force fields are designed such that a protein is simply a sum of its parts, that is, small solutes can be studied to develop the parameters required for description of functional groups commonly found in proteins. Hence, parameters are assumed to be additive and transferable. To this end, we present the results of a case study showing the KBIs for four binary mixtures that were each simulated using some of the most commonly used biomolecular FFs—AMBER99sb (Hornak *et al.* 2006), CHARMM27 (Mackerell *et al.* 2004; Bjelkmar *et al.* 2010), OPLS-AA (Jorgensen and Tirado-Rives 1988; Kaminski *et al.* 2001; Jorgensen and Tirado-Rives 2005), and GROMOS54a7 (Schmid *et al.* 2011). Thus, for a given FF we will interpret the results from these mixtures as a general indicator of how successful the FF may be at reproducing solute–solute, solute–solvent, and solvent–solvent distributions, which presumably relates to the quality of the underlying interaction potentials.

It is worth noting when the most recent nonbonded updates were published for each of these FFs. The AMBER99sb release only involved updates to the torsional parameters of AMBER99ff (Hornak *et al.* 2006). Among other changes, AMBER99ff (Wang *et al.* 2000), based upon AMBER94ff (Cornell *et al.* 1995), did include updates to the partial charges and the addition of a few new atom types. CHARMM's nonbonded parameters were most recently updated with the release of the CHARMM22 protein FF (Mackerell *et al.* 1998). GROMOS modified their N–H, C=O repulsion with the release of GROMOS 54A7 (Schmid *et al.* 2011) and, prior to that, refined their nonbonded parameters for oxygen-containing functional groups with the release of the GROMOS 53A6OXY parameter set (Horta *et al.* 2011). Following the extensive nonbonded OPLS-AA parameterization for all amino acids (Jorgensen and Tirado-Rives 1988), the cysteine and methionine nonbonded parameters were updated in the latest OPLS-AA protein FF (Kaminski *et al.* 2001).

The systems studied here were methanol (MOH) + water (HOH) at $x_{MOH} = 0.5$ at 300 K, benzene (Ben) + MOH at $x_{Ben} = 0.5$ and $x_{Ben} = 0.75$ at 308 K, *N*-methylacetamide (NMA) + HOH at $x_{NMA} = 0.1$ and $x_{NMA} = 0.2$ at 313 K, and 1 *m* and 3 *m* aqueous zwitterionic glycine (Gly) at 300 K. The aqueous simulations were performed with the appropriate water models: TIP3P (Jorgensen *et al.* 1983) for AMBER and CHARMM, TIP4P for OPLS (Jorgensen *et al.* 1983); SPC (Berendsen *et al.* 1981) for GROMOS; and SPC/E (Berendsen *et al.* 1987) for the KBFF models. The MOH + HOH system was chosen as an example of a solution containing two polar molecules. Benzene is representative of aromatic amino acid sidechains, and mixtures of Ben + MOH display interesting features in the KBIs as a function of composition, namely aggregation of MOH at low x_{MOH}. Here, we also argue that FFs that reproduce the composition

dependence of the KBIs for Ben + MOH mixtures are also more likely to correctly model the behavior of both solvent-exposed and buried phenylalanine. NMA was chosen because it is often used as a model for the peptide group. Low NMA compositions were studied because they are more representative of physiological protein backbone concentrations than high NMA compositions would appear. The current compositions may very roughly mimic peptide–peptide interactions in the presence of large amounts of water and may be very loosely analogous to the peptide interactions present in the denatured state. Finally, zwitterionic Gly (*aq*) was chosen because, despite being electrically neutral as a whole, Gly contains features similar to a mixture of salt-bridge forming ions. All of these choices were also driven by the availability of the required experimental data.

2.4 Analysis of Experimental Data and Molecular Dynamics Simulations

The analysis of the experimental data has been presented by us before (Weerasinghe and Smith 2005; Kang and Smith 2006; Gee 2010; Ploetz and Smith 2011a; Karunaweera *et al.* 2012). All simulations were performed using classical MD techniques and the GROMACS 4.5.3 simulation package (Hess *et al.* 2008). Molecules were randomly inserted in 10 nm in length cubic simulation boxes to create the desired mole fraction based on the experimental density. Each system was minimized for >1,000 steps using the steepest descent algorithm and then equilibrated for 100 ps in the *NVT* ensemble followed by 10 ns in the *NpT* ensemble and 100 ns of production in the *NpT* ensemble. All *NpT* simulations were at a *p* of 1 bar. MOH + HOH and Gly (*aq*) were simulated at 300 K, Ben + MOH at 308 K, and NMA + HOH at 313 K. These *T*s were chosen to correspond with the experimental activity data. The Berendsen weak-coupling technique was used to modulate the *T* and *p*, with relaxation times of 0.1 ps and 0.5 ps, respectively (Berendsen *et al.* 1984). A 4.5×10^{-5} bar^{-1} compressibility was used for all systems. Periodic boundary conditions and the minimum image convention were employed. All bond lengths were constrained using the Settle (Miyamoto and Kollman 1992) and Linear Constraint Solver (Hess *et al.* 1997) algorithms for water and non-water molecules, respectively. The use of bond constraints allowed for a two fs time step to be used for the integration of the equations of motion, which was performed using the Leap-Frog algorithm (Hockney *et al.* 1974). The particle-mesh Ewald (PME) technique was used to calculate electrostatic interactions (Darden *et al.* 1993; Lange *et al.* 2010). van der Waals interactions were calculated with the Lennard-Jones 6-12 functional form for all FFs. Cutoff distances for the real space electrostatic/van der Waals interactions were 0.8 nm/0.8 nm for AMBER (Hornak *et al.* 2006), 1.0 nm /1.5 nm for CHARMM, 0.8 nm /1.4 nm for GROMOS (Schmid *et al.* 2011), and 1.0 nm /1.5 nm for OPLS, respectively. CHARMM and OPLS do not use a consistent set of cut-offs, thus these values were chosen somewhat arbitrarily. The PME convergence parameter was 3.12 nm^{-1} for CHARMM and OPLS and 3.90 nm^{-1} for AMBER and GROMOS. Cubic interpolation and tinfoil boundary conditions (Darden *et al.* 1993) were used. The maximum fast Fourier transform grid spacing was 0.12 nm for the reciprocal space sum. The neighbor list was updated every ten steps.

The KBIs are defined as integrals over RDFs in the μVT ensemble, but μVT ensemble simulations are nontrivial to perform. We circumvent this issue by performing simulations in the *NpT* ensemble and approximating the μVT RDF by the *NpT* RDF (they differ by a term on the order of N^{-1}). We then truncate the KBIs at some distance R_c beyond which the RDFs are assumed to be unity. If the KBIs are not truncated they cannot be used, because they will be equal

to zero for unlike pairs and minus one for like pairs. This is discussed in more detail elsewhere (Chitra and Smith 2001b). One can check the appropriateness of both the specific choice of R_c and the truncation approximation in general such that internally consistent partial molar volumes (PMVs) are achieved by calculating the PMVs from the simulated KBIs and comparing them with the PMVs obtained from the composition dependent system volume data (Chitra and Smith 2001c). Here, the simulated values of the KBIs that have been plotted in comparison with experiment were obtained by averaging the simulated KBIs between 0.95 and 1.2 nm for MOH + HOH; between 2.0 and 2.5 nm for Ben + MOH; and between 1.2 and 1.6 nm for NMA + HOH and Gly + HOH based upon our previous studies of these same systems (Weerasinghe and Smith 2005; Kang and Smith 2006; Ploetz and Smith 2011a).Note that perfect agreement with our previous results is not achieved here due to the current studies use of much larger simulation boxes and much longer simulation times compared to our previous publications.

There are technical issues surrounding the determination of KBIs from simulations of closed systems (Schnell *et al.* 2011; Smith *et al.* 2013). We do not contend that the truncation (with averaging) type of approach described above is the most accurate. However, in the absence of more formal and accurate approaches, which were not available when we first investigated these integrals, we believe it has provided reasonable approximations to the true integrals. Certainly, it has allowed us to distinguish major flaws in many force fields, flaws that cannot simply be explained by inaccuracies in the integration procedure. Some examples of these flaws are described later.

Further technical details can be found in our previous work (see Table 2.2 for references) and will be supplemented in a future publication, which will additionally expand the set of thermodynamic and physical properties compared across the FFs and the mole fractions studied.

2.5 FF Implementation

The number of excluded interactions for each molecule is listed in Supplementary Table 2.1. The parameters for each molecule for each FF are provided in Supplementary Table 2.2 (non-bonded) and Supplementary Table 2.3 (bonded). Supplementary Table 2.4 shows the combination rules and the 1-4 pair interaction scaling rules used for each FF. Although each FF was developed with different methods to truncate non-bonded interaction potentials, we have used a straight cut-off for van der Waals interactions for each FF and PME (Darden *et al.* 1993) for the treatment of long-range electrostatics. It has been shown that all FFs generally perform better with PME (Lange *et al.* 2010).

2.5.1 AMBER99sb

The AMBER99sb implementation in GROMACS has been validated (Sorin and Pande 2005). A building block for NMA was constructed based upon the blocked termini (acetyl and *N*-methyl) groups' parameters. A building block for zwitterionic Gly was constructed based upon the charged *C*- and *N*-terminal Gly entries provided in the residue database. The TIP3P (Jorgensen *et al.* 1983) model was used for all water simulations. Because AMBER FFs have not been developed with the charge group concept, atomic charge groups were used.

2.5.2 CHARMM27

MOH (by A. D. MacKerell), Ben (by A. D. MacKerell), and NMA (by Louis Kuchnir) parameters were adopted directly from the CHARMM22 toppar_all22_prot_model.str file. Zwitterionic Gly was constructed from the provided database entries for the Gly residue and the charged termini. The TIP3P model (Jorgensen *et al.* 1983) was used for all water simulations. Because CHARMM FFs have not been developed with the charge group concept, atomic charge groups were used.

2.5.3 OPLS-AA

We used the MOH topology file provided in the OPLS-AA force field folder of GROMACS-4.5.3 (Jorgensen *et al.* 1996). The Ben topology was built based upon the toluene topology built for use in GROMACS (Jorgensen and Tirado-Rives 2005; Caleman *et al.* 2012). The NMA topology was taken directly from the literature (Jorgensen and Tirado-Rives 2005; Caleman *et al.* 2012). Zwitterionic Gly was constructed from the provided database entries for the Gly residue and the charged termini. The TIP4P model (Jorgensen *et al.* 1983) was used for all water simulations.

2.5.4 GROMOS54a7

The GROMOS54a7 FF is not supported in GROMACS-4.5.3. The complete set of GROMACS-compatible FF files for GROMOS 54A7 was obtained from the Automated Topology Builder and Repository (Malde *et al.* 2011). These files include a topology for MOH (rigid), which was used. Ben was constructed based upon the phenylalanine topology. A building block for NMA was constructed based upon the blocked *N*-terminus (acetyl) parameters and the N-H parameters used in the residue building blocks, which then created only one option for the CH3 (charge) parameter. Zwitterionic Gly was constructed from the provided database entries for the Gly residue and the charged termini. The SPC model (Berendsen *et al.* 1981) was used for all water simulations.

2.6 Assessment of KB Integrals from Established Biomolecular FFs

Figure 2.2 displays the excess coordination numbers $N_{ij} = \rho_j G_{ij}$ for the four systems and the four FFs. All of the models performed reasonably well for MOH + HOH, but there was a wide variety of results for the other systems.

N_{11} or N_{22} values of ~100 typically indicate a mixture that is approaching immiscibility (Matteoli and Lepori 1984). The GROMOS NMA + HOH N_{11} values are, indeed, very large for the compositions studied. Furthermore, snapshots from the simulations gave a visual indication that the GROMOS NMA + HOH systems appeared to phase separate at both NMA compositions studied. Hence, the RDFs for the GROMOS NMA + HOH mixtures could not be considered as converged. Therefore, the values reported in Figure 2.1 are only shown for comparison purposes.
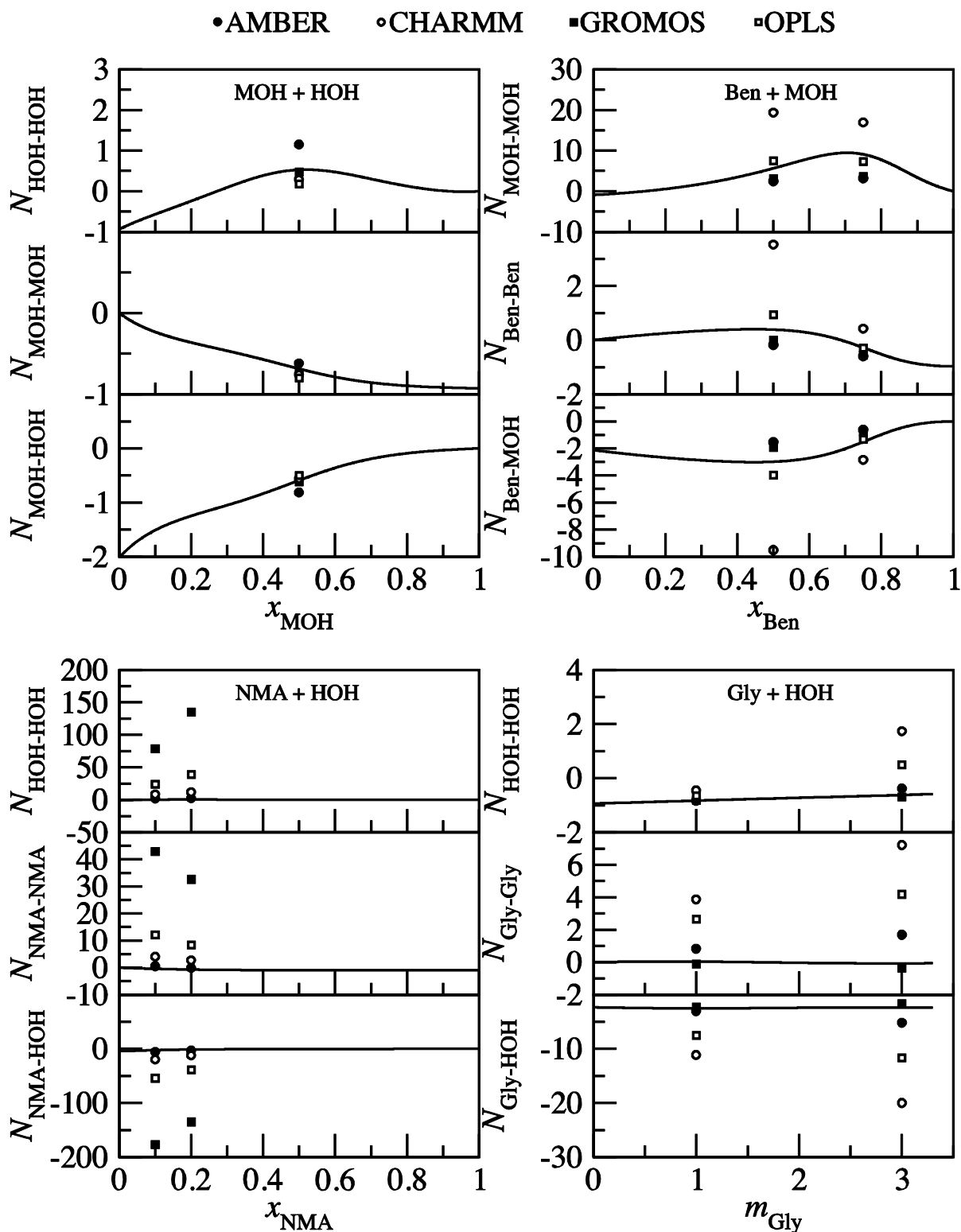
Figure 2.1 Comparisons of the excess coordination numbers for mixtures of methanol + water, benzene + methanol, N-methylacetamide + water, and zwitterionic glycine + water using four commonly used biomolecular force fields. Error bars are not shown for clarity. The GROMOS NMA + HOH points were calculated using unconverged RDFs.

It is interesting to note that, while some FFs appeared to reproduce the N_{ij}s reasonably well for certain systems, none of the FFs were able to reproduce the N_{ij}s for all four systems. Additionally, in each aqueous solution, all deviations from experiment were in the direction of too much self-association between the solute molecules, that is, large positive N_{22} values. In many cases, the deviations from experiment were quite large. Clearly, in these cases the models are not capturing the correct balance of solute–solvent, solute–solute, and solvent–solvent interactions. It was exactly these types of results, albeit for different systems, that prompted us to investigate what the reasons were for the lack of consistency among the FFs, why the results were sometimes in such poor agreement with experiment, and if FFs could be systematically improved.

Due to the issues mentioned previously surrounding the unknown polarization of molecules in condensed phases, we believe condensed phase charge distributions cannot, in general, be determined from a simple scaling of the QM calculated electrostatic potential. In our opinion, optimization of the effective charge distribution should be performed when studying the properties of liquid mixtures, not from quantum calculations for molecules in the gas phase. This is rarely attempted. Biomolecular FFs are usually designed to reproduce the properties of pure liquids and/or infinitely dilute solutes, with the assumption that mixtures of these molecules, in any ratio, will produce reasonable solution properties. For example, the most recent GROMOS partial charges were parameterized to reproduce the hydration free enthalpies of amino acid analogs in water (Oostenbrink *et al.* 2004). However, it is clear that this FF does not produce the correct balance of solute–solute, solute–solvent, and solvent–solvent interactions for NMA + HOH at the compositions studied. We believe that it is more desirable to base the partial charge parameterization on properties of mixtures across a large range of compositions. This is not a particularly new concept, but finding data that is sensitive to changes in composition is just as important and has been difficult.

2.7 The Kirkwood–Buff Derived Force Field Approach

We have above seen that current FFs do not necessarily reproduce the KBIs, and that generally the properties of solution mixtures are not typically taken into account during the parameterization procedure. Because different FFs produce such widely varying KBIs, and because the greatest variation in their parameters comes in the form of their charge distributions (Weerasinghe and Smith 2003a), it seems that the simulated KBIs are very sensitive to the choice of the charge distribution and that this distribution could be optimized to reproduce the KBIs. Armed with this philosophy, we set out to create a peptide / protein FF, known as the *Kirkwood-Buff derived force field* (KBFF), which attempts to reproduce the KBIs for small molecules representative of the building blocks of proteins and their mixtures with solvents such as water and methanol. The focus is centered on determining reasonable effective condensed phase charge distributions for polar solutes. More details of the philosophy can be found in Table 2.1 and elsewhere (Ploetz *et al.* 2010a; Weerasinghe *et al.* 2010). Table 2.2 shows the models we have parameterized to date.

Recently, approaches relying on KB theory have also been used by other groups for the parameterization or testing of several small molecule FFs (Lee and van der Vegt 2005; Zhong and Patel 2010; Horinek and Netz 2011). Additionally, outside of the biological FF community, there has been significant work to develop FFs that reproduce solution

activities, which is a key aim of the KBFF models. Notably the TraPPE models, developed by the Siepmann group, are specifically designed to model vapor-liquid coexistence curves (Maerzke *et al.* 2009). TraPPE has been developed primarily for thermodynamic and engineering purposes, and it cannot be directly applied to biomolecular systems in its present form.

Before indicating the type of results one can obtain using the KBFF philosophy, we briefly summarize the advantages and disadvantages of the approach.

Table 2.1 Strategy behind the KBFF

Philosophy	Action
The functional form of the FF should be simple and efficient to allow for the simulation of large systems for long timescales.	Develop a non-polarizable, united atom FF. Adopt the GROMOS functional form (geometric combination rules for σ and ε). Adopt the best (at the time the work was initiated) three-site water model, SPC/E.
Bond and angle parameters are reasonably well known.	Adopt the GROMOS bond and angle parameters.
Dihedral potentials can be obtained from gas-phase QM calculations.	Modify dihedral potentials to fit gas-phase QM calculations.
Because hydrocarbons are non-polar, their charge distributions play a relatively minor role.	Adopt the GROMOS nonbonded parameters for electrically neutral united atom hydrocarbons.
A correlation has been determined between atomic polarizabilities and atomic sizes (Miller and Savchik 1979).	Assign LJ 6-12 parameters for polar atoms using a simple scaling scheme between atomic size and atomic polarizabilities (Miller and Savchik 1979; Weerasinghe and Smith 2003c).
The charge distribution of polar groups is highly variable among FFs and is the major deficiency.	Iteratively parameterize the charge distributions of a solute to reproduce its experimental KBIs.
The FFs generated should provide accurate descriptions of the changes in the cosolvent chemical potentials as a function of cosolvent concentration.	Perform simulations of mixtures over as wide a range of compositions as possible. Consider all these compositions when fitting the KBIs; however, be aware of areas prone to statistical uncertainty experimentally and in the simulations (G_{11} at low x_1 and G_{22} at low x_2).

Table 2.2 Completed KBFF models

Solute	Solvent	Relevant species	Reference
Acetone	Water	Cosolvent	Weerasinghe and Smith (2003b)
Urea	Water	Cosolvent	Weerasinghe and Smith (2003c)
NaCl	Water	Cosolvent	Weerasinghe and Smith (2003d)
Guanidinium chloride	Water	Cosolvent, Arg	Weerasinghe and Smith (2004)
Methanol	Water	Ser	Weerasinghe and Smith (2005)
Amides	Water	Asn, Gln, peptide group, blocked termini	Kang and Smith (2006)
Thiols, sulfides	Methanol	Met, Cys, disulfide	Bentenitis *et al.* (2009)
Aromatics, aromatic alcohols	Methanol	Phe, Tyr	Ploetz and Smith (2011a)
Heterocycles	Methanol or Water	Trp, His (charged and neutral)	Ploetz and Smith (2011a)
$1°$, $2°$, $3°$ alcohols	Water	Thr, cosolvent	Jiao *et al.* (2014)
Amine salts, carboxylates	Water	Lys^{+1}, Asp^{-1}, Glu^{-1}, $termini^{+1/-1}$	Gee and Smith (2012)
Amines, carboxylic acids	Water	Cosolvent, Lys^0, Asp^0, Glu^0, $termini^0$	Dai *et al.* (2014)

2.7.1 Advantages of using KB Theory in FF parameterization

1. It is an exact solution theory, that is, it has no free parameters, requires no approximations, and makes no assumptions except that the system is in the thermodynamic limit.

2. Due to the complexities of the condensed phase, FF developers often must resort to using gas phase data to parameterize condensed phase FFs. KB theory allows condensed phase experimental data to be used.

3. The theory can be applied to solutions composed of any number of molecules of any size and complexity as long as the solution is stable with respect to composition.

4. It does not assume that the interactions are pairwise additive when expressing thermodynamic properties in terms of KBIs, but instead it takes into account the true many-body interactions in solutions.

5. One can decompose KBIs into distance dependent effects, which is difficult to do experimentally.

6. KBIs are a sensitive probe of the molecular distributions observed for different solutions and can be used to distinguish between models with otherwise similar characteristics.

2.7.2 Disadvantages of using KB theory in the parameterization approach

1. Relatively large simulation boxes (≥ 6 nm) may be necessary for systems with large molecules and/or in aggregating systems. The need for a large box may be anticipated if the experimental G_{ij}s show large deviations from their symmetric ideal values (Ploetz *et al.* 2010b).

2. Relatively long simulation times (tens of nanoseconds) are needed in order for the RDFs and G_{ij}s to converge.

3. The search for a charge distribution is more or less random. Our models have been developed based upon ~30 or fewer iterations (Bentenitis *et al.* 2009).

4. For some mixtures, the experimental data needed to conduct the KB inversion procedure is not available or is often unreliable.

The KBFF simulations followed the same protocol as the simulations for the other four FFs. The water model used in KBFF is SPC/E (Berendsen *et al.* 1987). Cutoff distances for the real space electrostatic/van der Waals interactions were 1.0 nm/1.5 nm. The PME convergence parameter was 3.12 nm^{-1}. The KBIs for each system were averaged over the same distances as were used for the other FFs.

Figure 2.2 displays the KBFF results compared to the results for the FF that came closest to reproducing the experimental N_{ij}s for a given system. To a certain extent the superior ability of the KBFF models to reproduce the KBIs is not surprising, as the parameters of the other FFs have not been optimized to reproduce this data, while the KBFF models were specifically parameterized to do so. But, we use it here as an indication of how this type of data can be improved with a reasonable degree of effort.We note that for MOH + HOH, GROMOS appears to have a
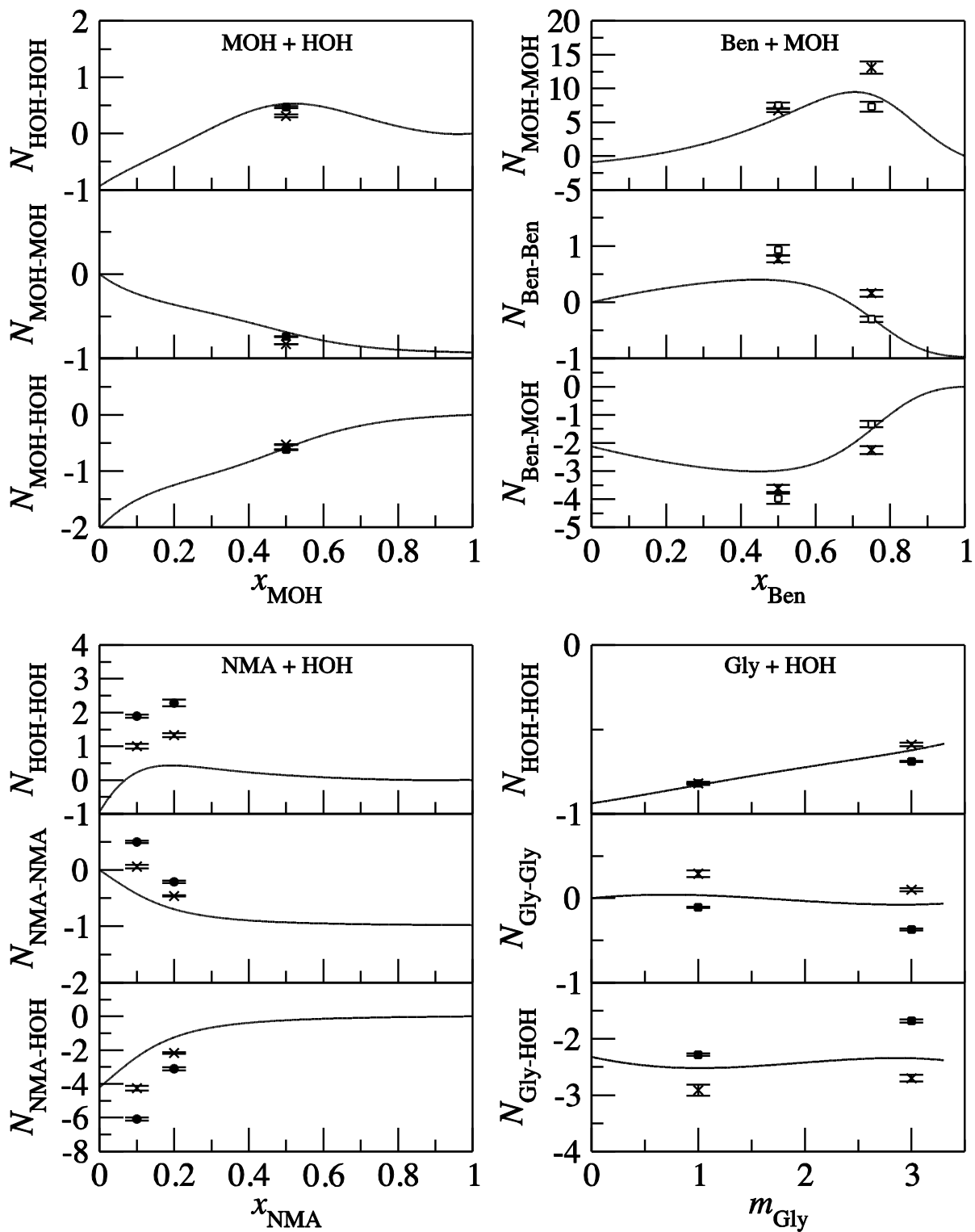
Figure 2.2 Comparisons of the excess coordination numbers using the Kirkwood-Buff Force Field and whichever force field best reproduced the excess coordination numbers. Error bars show the standard deviation on five twenty-nanosecond subaverages.

slightly better agreement than KBFF. However, this may not be true across the full composition range. For the current comparison, we have picked the composition ($x_{MOH} = 0.5$) in which the KBFF model shows the largest deviation from experiment (Weerasinghe and Smith 2005).

The results indicate that several of the FFs are competitive with the KBFF models for one of the systems, but none exhibited the overall balance between solute–solute, solute–solvent, and solvent–solvent distributions for all the solutes as provided by the KBFF models. The NMA and water results are particularly concerning due to the central role of peptide–peptide interactions in determining the balance between folded and unfolded protein conformations. The observation that all the FFs display a significantly more positive solute–solute excess coordination number compared to experiment also suggests a possible explanation for the imperfect thermodynamic behavior mentioned in Section 2.2.

Figure 2.3 shows the G_{ij}s as a function of distance for the NMA + HOH $x_{NMA} = 0.1$ system. While our choice of where to average the G_{ij}s is somewhat arbitrary, it is clear that the conclusions concerning the accuracy of the FFs are independent of the method used to obtain the exact values. For all FFs, the value of G_{11} and G_{22} are positive, indicating a degree of aggregation; however, the magnitude varies widely depending upon the FF. The values of the KBIs also tend to a constant limiting value for the systems in which the solute self-association was not too excessive. The absence of issues surrounding the integration of the RDFs is probably due to the large systems adopted here (involving between 10,000 and 33,000 molecules). The patterns displayed in Figure 2.3 for NMA and HOH was similar for all the systems studied here.

Figure 2.4 shows the solute-soute and solute-solvent RDFs for all five FFs at one composition for each mixture. Similar results were obtained for the other compositions. For MOH + HOH and Ben + MOH, there appears to be qualitative agreement among the RDFs for different FFs, despite their quantitative differences. This agreement is observed for both the short-ranged and the long-ranged structure. The NMA + HOH and Gly + HOH RDFs do not show qualitative agreement among the FFs. For NMA + HOH, similarities are evident in the shape of the RDFs for all FFs excluding GROMOS. However, the height of the first maximum in the RDFs differs among all FFs. Note that the GROMOS NMA + HOH $g_{NMA-NMA}$ does not tend toward unity due to the excessive self-aggregation of NMA in this simulation. For the Gly + HOH system, the solute-solute RDF does not appear to display a similar shape among the five FFs. This indicates that the Gly-Gly pairing differs substantially among the FFs, although the FFs do provide the same excluded Gly-Gly distance. The long-range structural behavior is also different among the FFs for the Gly + HOH system, although not to the degree exhibited for NMA + HOH mixtures.

The simulated KBIs are dependent on the choice of water model used with each solute FF, but the effect is minimal compared to the differences observed for different solute FFs (Chitra and Smith 2001a; Weerasinghe and Smith 2003d, 2003c, 2003b, 2004, 2005; Kang and Smith 2006). Based upon our previous studies of the KBI dependence on the water model, we predict that the results would not be significantly different on switching to a different model.
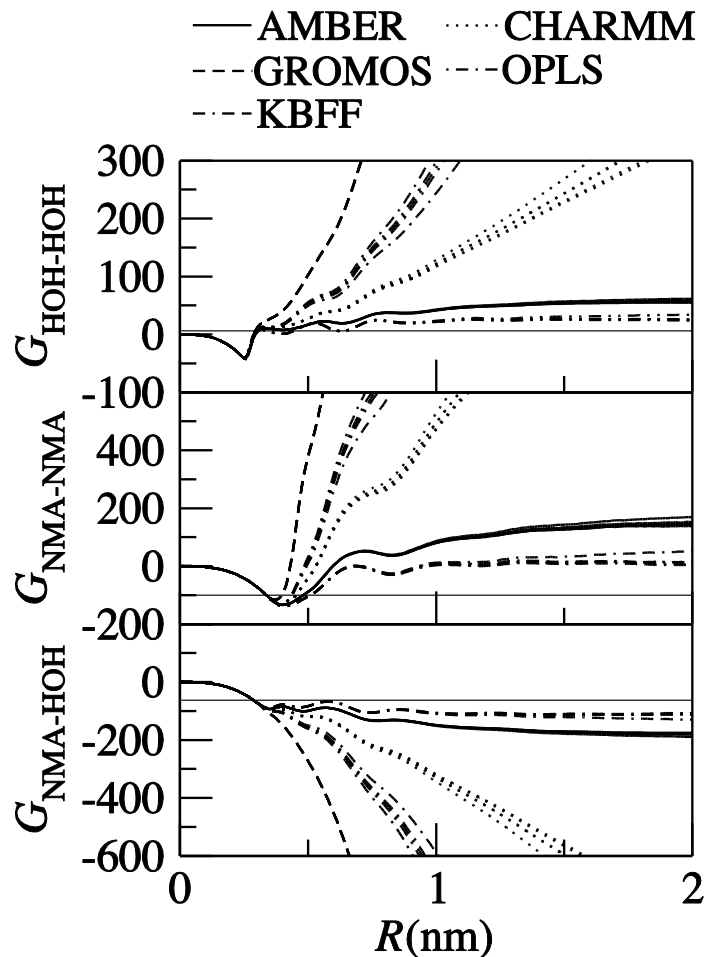
Figure 2.3 The Kirkwood-Buff integrals (G_{ij}, units of cm^3/mol) as a function of distance for the *N*-methylacetamide (NMA) + water system at $x_{NMA} = 0.1$ for the five force fields tested. The thin horizontal line is the experimental value. For each force field, the multiple lines correspond to five twenty-nanosecond subaverages. The GROMOS NMA + HOH G_{ij}s were calculated using unconverged RDFs.

2.8 Summary and Conclusions

The key conclusions we have drawn from our present and past work include the following:

1. A KB analysis provides more data for testing and development of FFs. The KBIs are very sensitive to the charge distributions. The major advantage of the KBFF approach is that by developing an FF that reproduces the KBIs, it is assured that the corresponding solution thermodynamics (including the activity of the solute) will be well described. This then further ensures the correct balance between solvation and self-association of solutes.
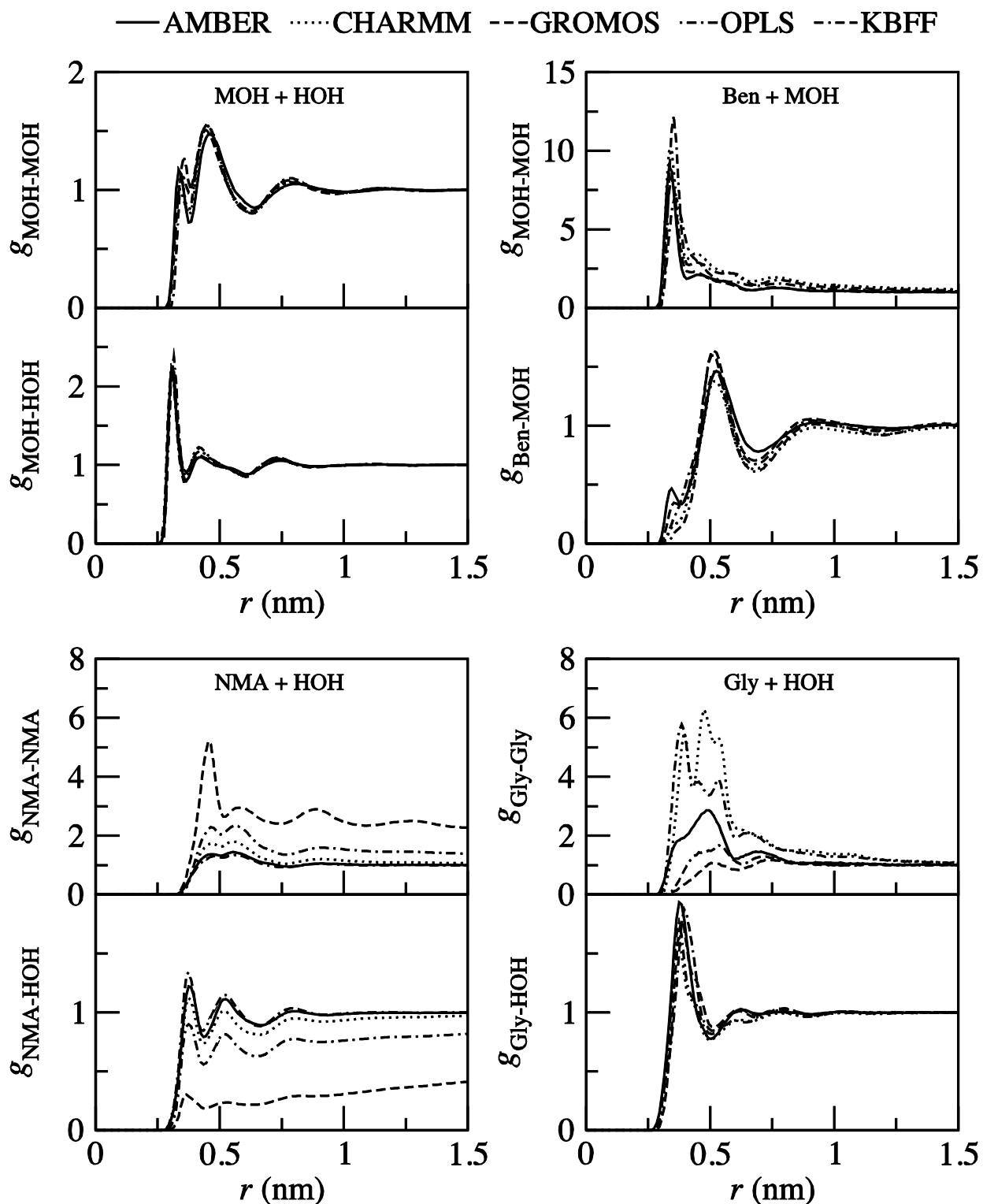
Figure 2.4 The center of mass based radial distribution functions (RDF) for five force fields for the mixtures methanol + water at x_{MOH} =0.5, benzene + methanol at x_{Ben} = 0.75, N-methylacetamide + water at x_{NMA} =0.1, and 1 m zwitterionic glycine (aq). Please note the different y-axis scales. The data represent the average RDF from 80 to 100 nanoseconds of production simulation after 10 nanoseconds of equilibration. The RDFs are converged with respect to time.

2. The KBFF approach is designed to specifically reproduce the experimental KBIs for small solutes in solution mixtures. This is an important property to model correctly, but if it were only attainable by sacrificing agreement with experiment for other properties that are commonly reproducible, it would be of little use. Fortunately, this does not appear to be the case (Weerasinghe and Smith 2003c, 2003d, 2003b, 2004, 2005; Kang and Smith 2006; Bentenitis *et al.* 2009; Gee *et al.* 2011; Ploetz and Smith 2011a).

3. We consider the improvements made to be essential if one wants to quantitatively understand phenomena such as conformational or association equilibria in quantitative detail, despite the disadvantages of needing relatively long simulation times and relatively large system sizes during the parameterization procedure.

4. The actual charge distribution, not merely the solute polarity (dipole moment), appears to be what balances the solvation, thus it is not enough to simply scale up gas-phase charge distributions to mimic condensed-phase charge distributions (Weerasinghe and Smith 2003a).

5. In all aqueous solutions we have tested (excluding MOH + HOH, which was reasonable for all the FFs tested), whenever there has been a large deviation from the experimental KBIs, it has always been toward an overfavoring of the solute-solute interactions, and thus too much self-aggregation. For FFs that overstabilize peptide-peptide group interactions, this likely leads to protein simulations that have low RMSDs from experimentally determined structures, and thus may give a false appearance of being reliable. However, they would not be able to reproduce conformational equilibria accurately, resulting in a tendency to exaggerate the stability (melting temperature).

6. The KBIs are most sensitive to the accuracy of the activity data used during the analysis, and least sensitive to the accuracy of the compressibility data. Often it is even assumed that the compressibility of the solution is ideal, and this is a reasonable assumption in the absence of real compressibility data as is the use of isentropic compressibility data (Matteoli and Lepori 1984). Compressibility uncertainties as high as 50% have been shown to have insignificant effects on the values of the KBIs (Matteoli and Lepori 1984).

7. The KBIs should reach a plateau value at the distance that the RDFs tend toward unity. Although the procedure for choosing a distance over which to take the average of the KBIs is somewhat arbitrary, it is able to capture the significant differences in the degree of molecular association between different FFs.

8. Although, in theory, there should be multiple RDFs that, upon integration, could provide the same value of G_{ij}, and thus multiple solutions to the charge distribution, we have never found this to be a practical concern. In fact, it has been difficult to find one set of partial charges (or one set of RDFs) that reasonably reproduce the three KBIs.

In our opinion, the results presented here clearly show that conventional FFs can possess a significant imbalance of solute solvation and self-association in solutions. Fortunately, this problem can be overcome by adjusting the effective

molecular charge distributions of the solutes during the parameterization procedure. In our opinion, the KBFF models represent an often dramatic improvement over the currently used protein FFs and should be seriously considered if one wishes to study conformational or association equilibria with reasonable accuracy.

2.9 Supplementary Materials

Supplementary Table 2.1 Number of exclusions

	All-atom			United-atom	
Molecule	AMBER	CHARMM	OPLS	GROMOS	KBFF
HOH	All	All	All	All	All
MOH	All+pairs	All+pairs	All+pairs	All	All
BEN	3+pairs	3+pairs	3+pairs	4	All
NMA	3+pairs	3+pairs	3+pairs	All+pairs	All+pairs
GLY	3+pairs	3+pairs	3+pairs	3+pairs	3+pairs

Supplementary Table 2.2 Non-bonded parameters

Water	TIP3P		TIP4P	SPC	SCP/E
σ (Å)					
O	3.151		3.154	3.167	3.166
H	0.000		0.000	0.000	0.000
M			0.000		
ε (kJmol^{-1}) × 10					
O	6.364		6.485	6.502	6.506
H	0.000		0.000	0.000	0.000
M			0.000		
q (|e|) × 10					
O	-8.340		0.000	-8.200	-8.476
H	4.170		5.200	4.100	4.238
M			-10.400		

Methanol	AMBER	CHARMM	OPLS	GROMOS	KBFF
σ (Å)					
O	3.066	3.154	3.120	3.177	3.192
H(O)	0.000	0.400	0.000	0.000	1.580
C	3.400	4.054	3.500		
H(methyl)	2.650	2.352	2.500		
CH3				3.601	3.748
ε (kJmol^{-1}) × 10					
O	8.803	6.364	7.113	5.500	6.506
H(O)	0.000	1.925	0.000	0.000	0.880
C	4.577	0.837	2.761		
H(methyl)	0.657	0.921	1.255		
CH3				10.172	8.672
q (|e|) × 10					
O	-6.497	-6.600	-6.830	-6.740	-8.200
H(O)	4.215	4.300	4.180	4.080	5.200
C	1.166	-0.400	1.450		
H(methyl)	0.372	0.900	0.400		

27

				2.660	3.000

Benzene	AMBER	CHARMM	OPLS	GROMOS	KBFF
σ (Å)					
C	3.400	3.550	3.550	3.581	3.810
H	2.600	2.420	2.420	2.373	1.580
ε (kJmol^{-1}) × 10					
C	3.598	2.929	2.929	2.774	3.300
H	0.628	1.255	1.255	1.184	0.880
q (|e|) × 10					
C	-1.305	-1.150	-1.150	-1.400	-1.300
H	1.305	1.150	1.150	1.400	1.300

N-Methylacetamide	AMBER	CHARMM	OPLS	GROMOS	KBFF
σ (Å)					
H(methyl)	2.650	2.352	2.500		
C(methyl)	3.400	3.671	3.500		
CH3				3.747	3.748
C(O)	3.400	3.564	3.750	3.581	3.360
O	2.960	3.029	2.960	2.760	3.100
N	3.250	3.296	3.250	3.137	3.110
H(N)	1.069	0.400	0.000	0.000	1.580
C'(methyl)	3.400	3.671	3.500		
H'(methyl)	2.471	2.352	2.500		
CH3'				3.747	3.748
ε (kJmol^{-1}) × 10					
H(methyl)	0.657	0.920	1.255		
C(methyl)	4.577	3.347	2.761		
CH3				8.672	8.672
C(O)	3.598	4.602	4.393	2.774	3.300
O	8.786	5.021	8.786	12.791	5.600
N	7.113	8.368	7.113	6.398	5.000
H(N)	0.657	1.925	0.000	0.000	0.880
C'(methyl)	4.577	3.347	2.761		
H'(methyl)	0.657	0.920	1.255		
CH3'				8.672	8.672
q (|e|) × 10					
H(methyl)	1.123	0.900	0.600		
C(methyl)	-3.662	-2.700	-1.800		
CH3				0.000	0.000
C(O)	5.972	5.100	5.000	4.500	6.200
O	-5.679	-5.100	-5.000	-4.500	-6.200
N	-4.157	-4.700	-5.000	-3.100	-7.000
H(N)	2.719	3.100	3.000	3.100	3.600
C'(methyl)	-1.490	-1.100	0.200	0.000	
H'(methyl)	0.976	0.900	0.600		
CH3'				0.00	3.400

Zwitterionic Glycine	AMBER	CHARMM	OPLS	GROMOS	KBFF
σ (Å)					
N	3.250	3.296	3.250	3.137	3.370
H(N)	1.069	0.400	0.000	0.000	1.580
C(methylene)	3.400	3.875	3.500		
H(methylene)	1.960	2.352	2.500		
CH2				4.070	4.070
C(O)	3.400	3.564	3.750	3.581	3.360
O	2.960	3.029	2.960	2.626	3.500
ε (kJmol^{-1}) × 10					
N	7.113	8.368	7.113	6.398	5.620
H(N)	0.657	1.925	0.000	0.000	0.880
C(methylene)	4.577	2.301	2.761		
H(methylene)	0.657	0.920	1.255		
CH2				4.105	4.105
C(O)	3.598	2.929	4.393	2.774	3.300
O	8.786	5.021	8.786	17.250	6.047
q (|e|) × 10					
N	2.943	-3.000	-3.000	1.290	5.000
H(N)	1.642	3.300	3.300	2.480	0.000
C(methylene)	-1.341	1.300	0.900		
H(methylene)	0.976	0.900	0.600		
CH2				1.270	5.000
C(O)	7.231	3.400	7.000	2.700	1.000
O	-7.855	-6.700	-8.000	-6.350	-1.000

Supplementary Table 2.3 Bonded parameters

Methanol	AMBER	CHARMM	OPLS	GROMOS (rigid)	KBFF (rigid)
Bonds	*Harmonic*	*Harmonic*	*Harmonic*	*Fourth power*	*Fourth power*
	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)
C-H(C)	0.1090	0.1111	0.1090		
C-O	0.1410	0.1420	0.1410		
O-H(O)	0.0960	0.0960	0.0945	0.1000	0.0945
CH3-O				0.1530	0.1430
CH3-H(O)				0.2077	0.1948
Angles	*Harmonic*	*Urey-Bradley*	*Harmonic*	*Cosine based*	*Cosine based*
	θ_{eq} (deg)/k_θ (kJmol^{-1}rad^{-2})	θ_{eq} (deg)/k_θ (kJmol^{-1})/r_{13} (nm)/k_{UB} (kJmol^{-1})	θ_{eq} (deg)/k_θ (kJmol^{-1}rad^{-2})	θ_{eq} (deg)/k_θ (kJmol^{-1})	θ_{eq} (deg)/k_θ (kJmol^{-1})
H-C-H	109.5/292.9	108.4/297.1/0.2/4518.7	107.8/276.1		
H-C-O	109.5/292.9	108.9/384.1/0.0/ 0.0	109.5/292.9		
C-O-H	108.5/460.2	106.0/481.2/0.0/ 0.0	108.5/460.2		
Proper Dihedrals	*Periodic, Mult. Simult. Fns.*	*Periodic, Mult. Simult. Fns.*	*Ryckaert-Bellemans*	*Periodic*	*Periodic*
	φ_s (deg)/k_φ (kJmol^{-1})/mult.	φ_s (deg)/k_φ (kJmol^{-1})/mult.	C_0/C_1/C_2/C_3/C_4/C_5 (kJmol^{-1})	φ_s (deg)/k_φ (kJmol^{-1})/mult.	φ_s (deg)/k_φ (kJmol^{-1})/mult.
H-C-O-H	0/2.1/3	0/0.6/3	0.9/2.8/0/-3.8/0/0		

Benzene	AMBER	CHARMM	OPLS	GROMOS	KBFF
Bonds	*Harmonic*	*Harmonic*	*Harmonic*	*Fourth power*	*Fourth power*
	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)
C-H	0.1080	0.1080	0.1080	0.1090	0.1090
C-C	0.1400	0.1375	0.1400	0.1390	0.1390
Angles	*Harmonic*	*Urey-Bradley*	*Harmonic*	*Cosine based*	*Cosine based*
	θ_{eq} (deg)/k_θ (kJmol^{-1}rad^{-2})	θ_{eq} (deg)/k_θ (kJmol^{-1})/r_{13} (nm)/k_{UB} (kJmol^{-1})	θ_{eq} (deg)/k_θ (kJmol^{-1}rad^{-2})	θ_{eq} (deg)/k_θ (kJmol^{-1})	θ_{eq} (deg)/k_θ (kJmol^{-1})
H-C-C	120/418.4	120/251.0/0.2/18409.6	120/292.9	120/505	120/505
C-C-C	120/527.2	120/334.7/0.2/29288.0	120/527.2	120/560	120/560
Proper Dihedrals	*Periodic, Mult. Simult. Fns.*	*Periodic, Mult. Simult. Fns.*	*Ryckaert-Bellemans*	*Periodic*	*Periodic*
	φ_s (deg)/k_φ (kJmol^{-1})/mult.	φ_s (deg)/k_φ (kJmol^{-1})/mult.	C_0/C_1/C_2/C_3/C_4/C_5 (kJmol^{-1})	φ_s (deg)/k_φ (kJmol^{-1})/mult.	φ_s (deg)/k_φ (kJmol^{-1})/mult.
C-C-C-C	180/15.2/2	180/13.0/2	30.3/0.0/-30.3/0.0/0.0/0.0		
H-C-C-C	180/15.2/2	180/17.6/2	30.3/0.0/-30.3/0.0/0.0/0.0		
H-C-C-H	180/15.2/2	180/10.0/2	30.3/0.0/-30.3/0.0/0.0/0.0		
Improper Dihedrals	*Periodic*	*Harmonic*	*Periodic Proper*	*Harmonic*	*Harmonic*
	φ_s (deg)/k_φ (kJmol^{-1})/mult.	ξ_{eq} (deg)/k_ξ (kJmol^{-1}rad^{-2})	φ_s (deg)/k_φ (kJmol^{-1})/mult.	ξ_{eq} (deg)/k_ξ (kJmol^{-1}rad^{-2})	ξ_{eq} (deg)/k_ξ (kJmol^{-1}rad^{-2})
C-C-H-C	180/4.6/2		180/4.6/ 2.0	0/167.4	
H-C-C-H	180/4.6/2				
C-C-C-C			180/4.6/ 2.0	0/167.4	0/167.4

N-Methylacetamide	AMBER	CHARMM	OPLS	GROMOS	KBFF
Bonds	*Harmonic*	*Harmonic*	*Harmonic*	*Fourth power*	*Fourth power*
	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)
C(methyl)-H	0.1090	0.1111	0.1090		
C(methyl)-C(O)	0.1522	0.1490	0.1522		
C(O)-O	0.1229	0.1230	0.1229	0.1230	0.1224
C(O)-N	0.1335	0.1345	0.1335	0.1330	0.1386
N-H	0.1010	0.0997	0.1010	0.1000	0.1000
N-C(methyl)	0.1449	0.1430	0.1449		
CH3-C(O)				0.1530	0.1520
N-CH3				0.1470	0.1468
Angles	*Harmonic*	*Urey-Bradley*	*Harmonic*	*Cosine based*	*Cosine based*
	θ_{eq} (deg)/k_θ (kJmol⁻¹rad⁻²)	θ_{eq} (deg)/k_θ (kJmol⁻¹)/r_{13} (nm)/k_{UB} (kJmol⁻¹)	θ_{eq} (deg)/k_θ (kJmol⁻¹rad⁻²)	θ_{eq} (deg)/k_θ (kJmol⁻¹)	θ_{eq} (deg)/k_θ (kJmol⁻¹)
H-C(methyl)-H	109.5/292.9	108.4/297.1/0.2/ 4518.7	107.8/276.1		
H-C(methyl)-C(O)	109.5/418.4	109.5/276.1/0.2/25104.0	109.5/292.9		
C(methyl)-C(O)-O	120.4/669.4	121.0/669.4/0.0/ 0.0	120.4/669.4		
C(methyl)-C(O)-N	116.6/585.8	116.5/669.4/0.0/ 0.0	116.6/585.8		
O-C(O)-N	122.9/669.4	122.5/669.4/0.0/ 0.0	122.9/669.4	124/730	121.8/502.1
C(O)-N-H	120.0/418.4	123.0/284.5/0.0/ 0.0	119.8/292.9	123/415	110.4/292.9
C(O)-N-C(methyl)	121.9/418.4	120.0/418.4/0.0/ 0.0	121.9/418.4		
H-N-C(methyl)	118.0/418.4	117.0/292.9/0.0/ 0.0	118.4/318.0		
N-C(methyl)-H(methyl)		109.5/431.0/0.0/ 0.0	109.5/292.9		
CH3-C(O)-O				121/685	124.1/502.1
CH3-C(O)-N				115/610	114.1/502.1
C(O)-N-CH3				122/700	119.6/502.1
H-N-CH3				115/460	130.0/376.6
Proper Dihedrals	*Periodic, Mult. Simult. Fns.*	*Periodic, Mult. Simult. Fns.*	*Ryckaert-Bellemans*	*Periodic*	*Periodic*
	φ_s (deg)/k_φ (kJmol⁻¹)/mult.	φ_s (deg)/k_φ (kJmol⁻¹)/mult.	$C_0/C_1/C_2/C_3/C_4/C_5$ (kJmol⁻¹)	φ_s (deg)/k_φ (kJmol⁻¹)/mult.	φ_s (deg)/k_φ (kJmol⁻¹)/mult.
H(methyl)-C(methyl)-C(O)-O	180/ 0.3/3 and 0/3.4/1	180/ 0.0/3	0.0/ 0.0/ 0.0/0.0/0/0		
H(methyl)-C(methyl)-C(O)-N	0/ 0.0/0	0/ 0.0/3	0.0/ 0.0/ 0.0/0.0/0/0		
C(methyl)-C(O)-N-H	180/10.5/2	180/10.5/2	20.5/ 0.0/-20.5/0.0/0/0		
C(methyl)-C(O)-N-C'(methyl)	180/10.5/2	0/ 6.7/1	30.3/-4.8/-25.5/0.0/0/0		
O-C(O)-N-H	180/10.5/2	180/10.5/2	20.5/ 0.0/-20.5/0.0/0/0		
O-C(O)-N-C(methyl)	180/10.5/2	180/10.5/2	25.5/ 0.0/-25.5/0.0/0/0		
C(O)-N-C(methyl)-H(methyl)	0/ 0.0/0	0/ 0.0/3	-0.3/-0.9/ 0.0/1.2/0/0		
H-N-C(methyl)-H(methyl)	0/ 0.0/0	0/ 0.0/3	0.0/ 0.0/ 0.0/0.0/0/0		
C(O)-C(methyl)-N-O			180.0/ 4.6/2		
N-H-C(O)-C'(methyl)			180.0/ 4.6/2		
CH3-C(O)-N-CH3'				180/33.5/2	180/33.5/2
Improper Dihedrals	*Periodic*	*Harmonic*	*Periodic Proper*	*Harmonic*	*Harmonic*
	φ_s (deg)/k_φ (kJmol⁻¹)/mult.	ξ_{eq} (deg)/k_ξ (kJmol⁻¹rad⁻²)	φ_s (deg)/k_φ (kJmol⁻¹)/mult.	ξ_{eq} (deg)/k_ξ (kJmol⁻¹rad⁻²)	ξ_{eq} (deg)/k_ξ (kJmol⁻¹rad⁻²)
C(O)-C(methyl)-N-O	180/4.6	0/1004.2			
N-H-C(O)-C'(methyl)	180/4.6	0/ 167.4			
C(O)-CH3-N-O				0/167.4	0/167.4
N-C(O)-CH3'-H				0/167.4	0/167.4

Zwitterionic Gly	AMBER	CHARMM	OPLS	GROMOS	KBFF
Bonds	*Harmonic*	*Harmonic*	*Harmonic*	*Fourth power*	*Fourth power*
	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)	r_{eq} (nm)
N-H	0.1010	0.1040	0.1010	0.1000	0.1000
N-C(methylene)	0.1471	0.1480	0.1471		
C(methylene)-H	0.1090	0.1080	0.1090		
C(methylene)-C(O)	0.1522	0.1522	0.1522		
C(O)-O	0.1250	0.1260	0.1250	0.1250	0.1250
N-CH2				0.1470	0.1470
CH2-C(O)				0.1530	0.1530
Angles	*Harmonic*	*Urey-Bradley*	*Harmonic*	*Cosine based*	*Cosine based*
	θ_{eq} (deg)/k_θ (kJmol^{-1}rad^{-2})	θ_{eq} (deg)/k_θ (kJmol^{-1})/r_{13} (nm)/k_{UB} (kJmol^{-1})	θ_{eq} (deg)/k_θ (kJmol^{-1}rad^{-2})	θ_{eq} (deg)/k_θ (kJmol^{-1})	θ_{eq} (deg)/k_θ (kJmol^{-1})
H-N-C(methylene)	109.5/418.4	109.5/251.0/0.2/16736	109.5/292.9		
N-C(methylene)-H	109.5/418.4	107.5/431.0/0.0/ 0	109.5/292.9		
N-C(methyl)-C(O)	111.2/669.4	110.0/365.7/0.0/ 0	111.2/669.4		
H-C(methylene)-H	109.5/292.9	115.0/301.2/0.0/ 0	107.8/276.1		
H-C(methylene)-C(O)	109.5/418.4	109.5/418.4/0.0/ 0	109.5/292.9		
C(methylene)-C(O)-O	117.0/585.8	118.0/334.7/0.2/41840	117.0/585.8		
O-C(O)-O	126.0/669.4	124.0/836.8/0.2/58576	126.0/669.4	126.0/770	126.0/770
H-N-CH2				109.5/425	109.5/425
N-CH2-C(O)				109.5/520	109.5/520
CH2-C(O)-O				117.0/635	117.0/635
Proper Dihedrals	*Periodic, Mult. Simult. Fns.*	*Periodic, Mult. Simult. Fns.*	*Ryckaert-Bellemans*	*Periodic*	*Periodic*
	φ_s (deg)/k_φ (kJmol^{-1})/mult.	φ_s (deg)/k_φ (kJmol^{-1})/mult.	C_0/C_1/C_2/C_3/C_4/C_5 (kJmol^{-1})	φ_s (deg)/k_φ (kJmol^{-1})/mult.	φ_s (deg)/k_φ (kJmol^{-1})/mult.
H-N-C(methylene)-H(methylene)	0/0.7/3	0/ 0.4/3	0.5/1.6/0/-2.2/0/0		
H-N-C(methylene)-C(O)	0/0.7/3	0/ 0.4/3	0.7/2.2/0/-2.9/0/0		
N-C(methylene)-C(O)-O	0/0.0/0	180/13.4/2	0/0/0/0/0/0		
H-C(methylene)-C(O)-O	0/0.0/0	180/ 0.2/6	0/0/0/0/0/0		
H-N-CH2-C(O)				0/3.8/3	0/1.8/3
N-CH2-C(O)-O				0/0.4/6	0/5.7/1
Improper Dihedrals	*Periodic*	*Harmonic*	*Periodic Proper*	*Harmonic*	*Harmonic*
	φ_s (deg)/k_φ (kJmol^{-1})/mult.	ξ_{eq} (deg)/k_ξ (kJmol^{-1}rad^{-2})	φ_s (deg)/k_φ (kJmol^{-1})/mult.	ξ_{eq} (deg)/k_ξ (kJmol^{-1}rad^{-2})	ξ_{eq} (deg)/k_ξ (kJmol^{-1}rad^{-2})
C(O)-C(methylene)-O-O	180/43.9/2	0/803.3	180/43.9/2		
C(O)-CH2-O-O				0/167.4	0/167.4

Supplementary Table 2.4 Combination rules and 1-4 interactions

Force Field	Combination Rule	Generate Pairs	Scale LJ 1-4	Scale qq 1-4
AMBER	Lorentz-Berthelot	Yes	50%	83.33%
CHARMM	Lorentz-Berthelot		*Non-uniform scaling*	
OPLS-AA	Geometric	Yes	50%	50%
GROMOS	Geometric		*Non-uniform scaling*	
KBFF	Geometric	Yes	10%	50%

Lorentz-Berthelot combination rules: $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$, $\varepsilon_{ij} = (\varepsilon_i + \varepsilon_j)^{1/2}$

Geometric combination rules: $\sigma_{ij} = (\sigma_i + \sigma_j)^{1/2}$, $\varepsilon_{ij} = (\varepsilon_i + \varepsilon_j)^{1/2}$, or $C_{ij} = (C_i + C_j)^{1/2}$, $C_{ij} = (C_i + C_j)^{1/2}$ for GROMOS

3. To Polarize or Not To Polarize? Charge-On-Spring (COS) versus KBFF Force Fields for Water and Methanol Bulk and Vapor-Liquid Interfacial Mixtures[*]

Abstract

Simulations of water and methanol mixtures using polarizable force fields (FFs) for methanol (COS/M and CPC) and water (COS/G2) were performed and compared to experiment and also to a nonpolarizable methanol (KBFF) model with SPC/E water in an effort to quantify the importance of explicit electronic polarization effects in bulk liquid mixtures and vapor-liquid interfaces. The bulk liquid mixture properties studied included the center of mass radial distribution functions, Kirkwood-Buff integrals (KBIs), volumetric properties, isothermal compressibility, enthalpy of mixing, dielectric constant, and diffusion coefficients. The vapor-liquid interface properties investigated included the relative surface probability distributions, surface tension, excess surface adsorption, preferred surface molecule orientations, and the surface dipole. None of the three FFs tested here was clearly superior for all of the properties examined. All the force fields typically reproduced the correct trends with composition for both the bulk and interfacial system properties; the differences between the force fields were primarily quantitative. The overall results suggest that the polarizable FFs are not, at the present stage of development, inherently better able to reproduce the bulk and interfacial properties. In fact, the specific parameterization of the FF appears to be just as important as the class of FF.

3.1 Introduction

Computer simulation has become one of the most popular theoretical approaches available for the study of molecular interactions in solution (Allen and Tildesley 1987; Frenkel and Smit 2002). In particular, biomolecular simulations have provided a wealth of information concerning complicated systems of biological interest (Orozco and Luque 2000; van Gunsteren *et al.* 2006). At the center of all simulation studies lies the collection of equations and parameters used to quantify the interactions within the system; collectively known as the force field (Hünenberger and van Gunsteren 1997; Mackerell 2004). The quality of the force field, together with the degree of sampling achieved, determines the accuracy of the simulation results (van Gunsteren *et al.* 2006). For biomolecular systems, where long time scales are prevalent, the force field is typically empirical in nature to ensure a computationally efficient approach (Cornell *et al.* 1995; Jorgensen *et al.* 1996; van Gunsteren *et al.* 1996; Mackerell *et al.* 1998;). Hence, one often has to strike a balance between force field accuracy and efficiency.

When a molecule is taken from the gas phase to the liquid phase, the electron density in the molecule becomes polarized, especially if the solution contains polar molecules like water. This polarization effect is well known and generally leads to an increase in the polarity of the molecule in question according to the polarizability of the molecule and the solution environment. Traditionally, this effect is accounted for in biomolecular force fields by using "effective" fixed charges on the atoms to model the polarized electron density in solution (Cornell *et al.* 1995; Jorgensen *et al.* 1996; van Gunsteren *et al.* 1996; Mackerell *et al.* 1998). These charges are typically larger than the values obtained from quantum mechanical calculations performed in the gas phase, in order to mimic the expected polarization effects in solution (Yu and van Gunsteren 2005). Alternatively, the polarization effect can be treated explicitly by using induced dipoles, or other similar approaches, to solve the electrostatic contribution to the interaction energy (Warshel and Levitt 1976; Vesely 1977; Van Belle *et al.* 1987; Straatsma and McCammon 1990; Rappé and Goddard 1991; Rick *et al.* 1994; Halgren and Damm 2001; Rick and Stuart 2002; Yu and van Gunsteren 2005; Warshel *et al.* 2007). In this way, the molecular interactions are dependent on the particular environment of the molecules. It should be noted that there are three types of polarization: electronic, geometric, and orientational. The only one that is not accounted for in nonpolarizable FFs is the electronic polarization, since the geometry of the molecules is (generally) not fixed (although it is common for small molecules, such as those studied here, to be kept rigid) and the orientation of molecules is generally not fixed. Efforts to develop polarizable force fields have been pursued over the last few decades (see for example references 26-74 in reference (Geerke and van Gunsteren 2007a)). More recently, polarizable FF parameters have been developed and optimized for a wide variety of building blocks relevant to proteins and other biomolecules (Kaminski *et al.* 2002; Ren and Ponder 2002; Baucom *et al.* 2004; Patel and Brooks 2004; Patel *et al.* 2004; Anisimov *et al.* 2005; Maple *et al.* 2005; Vorobyov *et al.* 2005; Babin *et al.* 2006; Yang and Zhang 2006; Anisimov *et al.* 2007; Xie *et al.* 2007).

The (often stated) advantages of polarizable force fields are the expected increase in accuracy and better transferability of force field parameters (Dang *et al.* 1991; Meng *et al.* 1994; Jorgensen *et al.* 1995; Borodin *et al.* 2001; Grossfield *et al.* 2003; Noskov *et al.* 2005; Patel and Brooks 2005; Patel and Brooks 2006; Rivera *et al.* 2006; Geerke and van

Gunsteren 2007a; Harder *et al.* 2008). The (often stated) disadvantages are generally the increased computationl cost and the fact that polarizalbe FFs are less established and have not, therefore, been as thoroughly tested as popular nonpolarizable FFs. Nonpolarizable, fixed-charge, force fields have the advantage that they are more efficient, are well established, and are inherently pairwise additive. The primary disadvantage is an inability to mimic changes in the molecular charge distribution due to changes in the local environment. Several studies have attempted to compare the properties of both polarizable and nonpolarizable force fields (Dang *et al.* 1991; Meng *et al.* 1994; Jorgensen *et al.* 1995; Borodin *et al.* 2001; Grossfield *et al.* 2003; Noskov *et al.* 2005; Patel and Brooks 2005; Patel and Brooks 2006; Rivera *et al.* 2006; Geerke and van Gunsteren 2007a, 2007b; Harder *et al.* 2008). Generally these studies favored the use of polarizable approaches.

Recently (see Table 2.2), a series of simple, nonpolarizable force fields for small solutes in water have been developed in an effort to reproduce the properties of solution mixtures (Weerasinghe and Smith 2003a, 2003b, 2003c, 2004, 2005; Kang and Smith 2006). These differ from common approaches in that the charge distributions are not obtained from quantum calculations, but are varied to improve the agreement with the liquid mixture properties. The major property of interest is the chemical potential composition derivatives, which are monitored by determining the Kirkwood-Buff (KB) integrals between species pairs (Ben-Naim 1987). The advantage of this approach is that the parameterization is directed by the liquid mixture properties, an not by quantum calculations in the gas phase (albeit corrected to mimic the condensed phase) or the commonly targeted data of only pure liquid and/or infinitely dilute properties. The resulting KB derived force fields (KBFF) also perform well for more traditionally parameterized liquid properties, so one does not have to sacrifice agreement for one set of properties with agreement for another set (Weerasinghe and Smith 2003a, 2003b, 2003c, 2004, 2005; Kang and Smith 2006; Smith 2006). Inherent in the KBFF approach is that the degree of polarization at each atomic site is the determining (unknown) factor. In particular, comparison of the KBFF partial charges with other established nonpolarizable force fields suggests that quantum calculations generally do not accurately mimic the degree of polarization at each specific site.

In light of these developments one may ask the question, "When is explicit polarizability required for computer simulations?" The results for dielectric properties of liquids, solution interfaces, and small clusters of polar or ionic molecules are typically considered to improve when treated by polarizable approaches. Alternatively, it is possible that many mixtures of similar molecules may be safely treated with nonpolarizable force fields. However, the exact systems for which the inclusion of explicit polarization is absolutely required to obtain accurate results are still unknown.

The major aim of the current study is to compare a variety of results of simulations using well-parameterized polarizable force fields (based on the Charge-On-Spring model (Straatsma and McCammon 1990; Yu *et al.* 2003), a derivative of the Drude Oscillator, DO, or Shell polarization approach) with their corresponding nonpolarizable KBFF models. It should be noted that the parameterization of the applied COS model was primarily based on the properties of pure liquids, whereas aqueous mixture properties, specifically the KBIs, were explicitly included in the

parameterization of the KBFF model. As an initial attempt to answer the above question, we have simulated bulk liquid and vapor-liquid interfacial methanol and water mixtures. At the outset, we hypothesized that the polarizable force fields would perform better for all systems, but most especially for the interfacial systems in which the cosolvent molecules are exposed to extreme differences in the dielectric environment, ranging from the bulk aqueous enironment to a vacuum.

3.2 Force Fields

Table 3.1 shows the FF parameters used for the molecules simulated. The KBFF rationale is explained in Chapter 2. The KBFF methanol model is a rigid, united atom model consisting of three atomic sites and three point charges.

Table 3.1 Force field parameters for the water models

		Water	
Property	**Units**	**SPC/E**	**COS/G2**
q_O	e	-0.84760	0.00000
q_H	e	0.42380	0.52650
$q_{virtual, core}$	e	--	6.94700
$q_{COS, shell}$	e	--	-8.00000
d_{O-H}	nm	0.10000	0.09572
$d_{O-virtual}$	nm	--	0.02200
d_{H-H}	nm	0.16333	0.15139
α_{mol}	$10^{-3}nm^3$	--	1.25500
σ_O	nm	0.31660	0.31960
ε_O	kJ/mol	0.65060	0.76105

q_i is the partial charge of site i, O is oxygen, H is hydrogen, COS is the Charge-On-Spring, d is the bond distance, α_i is the polarizability of site i, σ_i and ε_i are the Lennard-Jones VDW parameters of site i.

Table 3.2 Force field parameters for the methanol models

		Methanol		
Property	**Units**	**KBFF**	**COS/M**	**CPC**
q_O	e	-0.82000	−0.53000	−0.55000
$q_{O,core}$	e	--	7.47000	7.45000
$q_{O,shell}$	e	--	-8.00000	-8.00000
q_H	e	0.52000	0.36000	0.34000
q_{CH3}	e	0.30000	0.17000	0.21000
$q_{CH3,core}$	e	--	--	8.21000
$q_{CH3,shell}$	e	--	--	-8.00000
d_{O-H}	nm	0.09450	0.10000	0.10000
d_{CH3-O}	nm	0.14300	0.14300	0.14300
d_{H-CH3}	nm	0.19480	0.19880	0.19880
α_O	$10^{-3}nm^3$	--	1.32000	1.10000
α_{CH3}	$10^{-3}nm^3$	--	--	0.95000
σ_O	nm	0.31920	0.32373	0.31498
σ_H	nm	0.15800	--	**--**
σ_{CH3}	nm	0.37480	0.36012	0.37069
ε_O	kJ/mol	0.65060	0.49206	0.57033
ε_H	kJ/mol	0.08800	--	--
ε_{CH3}	kJ/mol	0.86720	1.01724	1.20030

q_i is the partial charge of site i, O is oxygen, H is hydrogen, CH3 is the united methyl group COS is the Charge-On-Spring, d is the bond distance, α_i is the polarizability of site i, σ_i and ε_i are the Lennard-Jones VDW parameters of site i.

The water model that KBFF has been parameterized with is the SPC/E model. SPC/E is rigid and consists of three interaction sites and three point charges.

The Charge-On-Spring approach is a derivative of the Drude Oscillator (DO) or Shell polarization approach. Other than DO, the two additional main classes of polarization implementations are Fluctuating Charge and polarizable Point Dipole approaches. In the DO approach, a point charge (shell) is attached *via* a harmonic potential (spring) to another interaction site (core) (Vosmeer *et al.* 2012). The force constant for the harmonic potential depends upon the polarizability parameter, α_i. The displacement of the shell with respect to the core particle due to the presence of the electric field $\boldsymbol{E}$ arising from the surrounding molecular environment determines the instantaneous magnitude and direction of the induced dipole moment of the core particle. In the COS scheme the inducible dipole moment $\boldsymbol{\mu}_i$ is assigned to heavy atoms i through $\boldsymbol{\mu}_i = \alpha_i(4\pi\varepsilon_0)\boldsymbol{E}(\boldsymbol{r}_i)$, where ε_0 is the permittivity of free space (Vosmeer *et al.* 2012). This relationship has assumed a linear response between the induced dipole moment and the electric field. All standard electrostatic treatments, such as Particle Mesh Ewald (PME), may still be employed with COS since the electrostatic interactions all still arise from point charges (Vosmeer *et al.* 2012). To implement COS models, an iterative energy minimization is performed at each point in time during which the shell particle is moved. For the COS models, this results in a 2-3 fold increase in the computational cost compared to a comparable non-polarizable simulation (Vosmeer *et al.* 2012).

3.2.1 COS/G2 water model

COS/G2 (Yu and van Gunsteren 2004) was a reparameterization of an earlier polarizable water model also by van Gunsteren, COS/B2 (Yu *et al.* 2003). The focus of the reparameterization was to correct COS/B2's overestimation of the dielectric permittivity. The new model moved the core and shell off of the oxygen atom and used a massless virtual site as the core site, in order to reproduce the water monomer quadrupole moments. The specific hydrogen atom partial charges and position of the virtual site, which bisects the H-O-H angle, were parameterized to reproduce the molecular dipole and quadrupole moments of the monomer in the gas phase. The molecular polarizability and the oxygen Lennard-Jones parameters were parameterized to reproduce the experimental values for the heat of vaporization and the density of liquid water under ambient conditions. COS/G2 is a rigid, 5 site water model, with a single polarizable center located on the virtual site connected to the oxygen atom and with Lennard-Jones parameters only on the oxygen atom. It was parameterized to reproduce properties of liquid water well and gas phase water reasonably. The COS/G2 water model has a geometry taken from quantum mechanical calculations and uses gas phase charges. COS/G2 uses a condensed phase, isotropic polarizability parameter obtained from QM fitting of α, because gas phase polarizability leads to molecular dipole moments that are too large and fluctuate too much (overpolarization).

3.2.2 COS/M methanol model

COS/M (Yu *et al.* 2006) was developed by van Gunsteren and coworkers using the Charge-On-Spring technique to be compatible with the COS/G2 water model. COS/M is a rigid, three atomic interaction site (united methyl) model for methanol, which has one polarizable center located on the oxygen atom. The oxygen Lennard–Jones parameters

together with the molecular polarizability were parameterized to reproduce the experimental heat of vaporization and density of liquid methanol at 298K and 1 atm. The partial charges were assigned to reproduce the gas phase dipole moment. The parameterized molecular polarizability is less than the experimentally known gas phase value.

3.2.3 CPC methanol model

The development of the CPC model by Geerke and coworkers (Vosmeer *et al.* 2012) was based upon a combined QM/MM effort in which an ensemble of MM configurations were sampled. Post MM simulation, the trajectory was reanalyzed with one of the molecules being treated quantum mechanically, specifically using DFT. The response of the QM solute's electron configuration (electrostatic potential) to the electric field generated by the surrounding MM solvent point charges was calculated for each configuration. The inducible atomic dipole moment was then fitted to reproduce the electrostatic potential. From the fitted dipole moment the atomic polarizability parameters were determined, using the knowledge of the electric field at each atomic center and under the assumption of linear response. This process was performed for an infinitely dilute methanol molecule in pure water, in pure methanol, and in pure chloroform and also for an infinitely dilute water molecule in pure water.

The CPC model uses the COS/M geometry (rigid) and gas-phase estimated partial charges, but has two polarizable centers, one on the oxygen and one on the carbon. The polarizability of the carbon center and the van der Waals parameters were parameterized to reproduce the experimental density and enthalpy of vaporization for pure methanol. The CPC model has three atomic interaction sites.

3.3 Methods

All molecular dynamics simulations were carried out using Gromacs 4 (Hess *et al.* 2008). Newton's equations of motion were integrated based on the Leap-Frog algorithm (Hockney 1970), with a time step of 2 fs. Bond lengths of methanol were constrained to their minimum-energy values using the LINear Constraint Solver (LINCS) (Hess *et al.* 1997) algorithm. Water bond lengths were constrained using Settle (Miyamoto and Kollman 1992). A grid-based neighbor list was used and updated every ten steps for the KBFF simulations. The polarizable simulations did not use a list. All simulations used three dimensional periodic boundary conditions. The van der Waals interactions were truncated at 1.5 nm.

In the polarizable simulations, the energy minimization tolerance for the shell relaxation was 10 kJ/mol/nm and the maximum number of shell relaxation iterations was 20. For the polarizable simulations, the energy was computed and summed over all processors every step ('ntcalcenergy = 1'), because global communication is needed when the energy is optimized with respect to the shell positions.

Although the polarizable models were parameterized using reaction field electrostatics, the particle-mesh Ewald (PME) technique was used here to calculate electrostatic interactions with a real space cutoff distance of 1.0 nm for the nonpolarizable simulations and 1.5 nm for the polarizable simulations. To simulate a vapor-liquid interfacial

system using a reaction field, which uses a fixed bulk dielectric constant for the whole system, would be inappropriate. The bulk mixture results for the polarizable simulations, obtained using both reaction field and PME electrostatics, were found to be identical to within the statistical errors (results not shown). The PME implementation cubic interpolation, a maximum fast Fourier transform grid spacing of 0.12 nm for the reciprocal space sum, and tinfoil boundary conditions (Darden *et al.* 1993).

After steepest descent energy minimization and an equilibration period of at least 10 ns using the appropriate force field, production simulations of 40 ns were carried out for the nonpolarizable simulations, 20 ns for the bulk polarizable, and 10 ns for the interfacial polarizalbe. Energies and atomic coordinates were written out every 5 ps and diagonal elements of the pressure tensor every 0.1 ps. The bulk simulations were ran in the *NpT* ensemble while the interfacial simulations were ran in the *NVT* ensemble. The temperature and (in the simulations of the bulk mixtures) the pressure were maintained at an average value of 300 K and 1 bar using the weak-coupling method of Berendsen (Berendsen *et al.* 1984), with coupling times of the temperature and pressure baths set to 0.1 and 0.5 ps, respectively. Separate temperature coupling groups were used for methanol and for water. In the *NpT* simulations the isothermal compressibility was set to 4.5×10^{-5} bar^{-1} and the pressure coupling was isotropic. No other differences exist in the simulation parameters for the bulk and interfacial system. Unless otherwise noted, the estimated error bars shown in the Results section were obtained from the standard devation of 10-ns block-averages (thus, error bars are not reported for the polarizable interfacial systems). Based upon previous simulations of the polarizable simulations using slightly different simulation settings and much longer simulation timescales (results not shown), this is an acceptable assumption.

3.3.1 Bulk Systems

The bulk systems consisted of 6 nm in length cubic boxes in which the molecules were initially assigned random positions. Simulations of the bulk aqueous mixtures were performed at nine different mole fractions methanol (x_{MOH} = 0.00, 0.125, 0.250, 0.375, 0.50, 0.625, 0.750, 0.875, and 1.00).

3.3.2 Interfacial Systems

As shown in Figure 3.1, the interfacial systems consisted of $6 \times 6 \times 36$ nm simulation boxes in which molecules were initially assigned to random positions within the cental $6 \times 6 \times 12$ nm box leaving the flanking $6 \times 6 \times 12$ nm boxes on both ends of this region to serve as the vapor regions. The compositions studied were x_{MOH} = 0.000, 0.125, 0.250, and 1.000. No long-range corrections to the surface tension were included. These corrections are non-trivial for mixtures, and we consider the cutoffs to form an integral part of the force field model. As a test, we increased the van der Waals cut-off from 1.5 nm to 2.5 nm for the pure SPC/E water and pure KBFF methanol systems. The surface tension increased by 2.5% and 12.5%, respectively, and provided improved agreement with experiment in the case of water, but worse agreement with experiment for methanol..
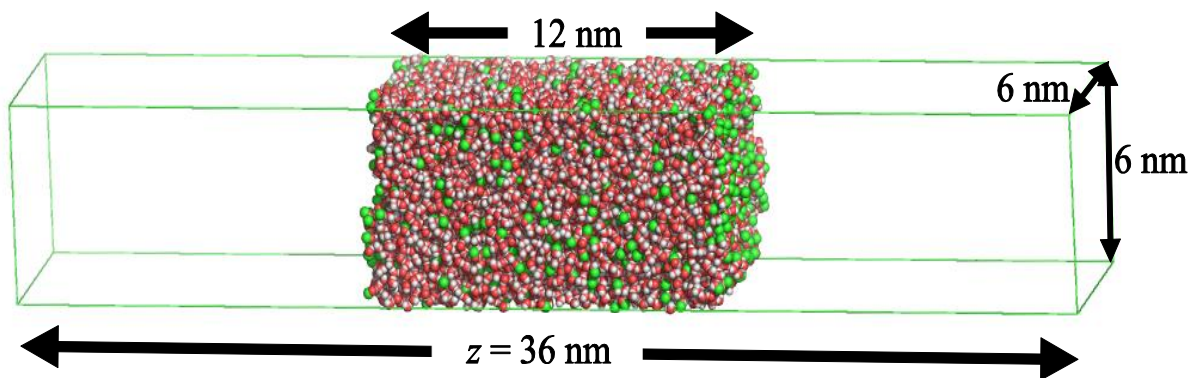
Figure 3.1 Diagram of the simulated interfacial systems. The surface lies in the *xy* plane perpendicular to the *z*-axis. In the KBFF simulations, only negligible evaporation occurred, while in the polarizable simulations, one or two large droplets (not shown) repeatedly formed in the vapor phase and re-condensed containing on the order of ten to one hundred molecules out of the total system of on the order of 10,000 molecules. The droplets contained both methanol and water molecules.

3.3.3 Determination of bulk properties

The experimental KBIs were calculated from the KB inversion procedure. The simulated KBIs were obtained using the Schnell approach in which the linear region was taken at $1/L = 0.5\text{-}1.7$ nm^{-1} (Schnell *et al.* 2011). Briefly, this involves using the particle fluctuation definition of the Kirkwood-Buff integrals (see Equation 1.2) instead of the integrated RDF definition (see Equation 1.1). Computationally, Equation 1.2 is implemented by picking a random coordinate within the simulation box and keeping track of how many molecules of each type are found as a function of distance (linearly or radially) away from the random point. For each configuration, a user-specified number of random origins for this calculating can be chosen. In this respect, it offers a potential statistical advantage over the "integrated RDF" definition of the KBIs, because when using the integrated RDFs, the number of origins for the calculation is fixed at the number of molecules in the system. Another advantage of the particle fluctuation determination of the KBIs is that they plateau with distance sooner than do the integrated RDFs, which necessarily have oscillations at short distances and eventually converge/plateau at larger radii. Conversely, this is also a disadvantage of the approach, because the microstructural information about the mixture is lost using the particle fluctuation equation.

After calculating the local KBIs using the particle fluctuation method, the Schnell approach involves the linear fit of the fluctuations as a function of local box length, L, according to the expression, $G_{ij}(L^{-1}) = mL^{-1} + G_{ij}(0)$, where m is the slope. The fit involves a subset of the points that are deemed to be linear, followed by extrapolation to the infinite system (thermodynamic) limit to provide, $G_{ij}(0)$, as the appropriate value for the KBI. The points that deviate from linearity, and must therefore be dropped, include those at both small and large L. At small L, deviations from linearity can be attributed to having only a few molecules within the local volume. At large L, not enough bath particles exist to allow for correct sampling of a local μVT ensemble from within the closed simulation system.

For situations where the KBIs have not completely converged, the Schnell approach has been touted as being more rigorous than averaging the KBIs over a range of distances corresponding to approximately one solvation shell (Schnell *et al.* 2011). Our experience, however, has been that there is an equal amount of ambiguity in determining which points correspond to the "linear" region and which do not as there is in choosing where to average when using the integrated RDF approach. In this work, we chose the linear region that gave internally consistent values for the partial molar volumes and isothermal compressibilities when calculated from the simulated $(\partial V/\partial N_i)_{T,p,\{N\}'}$ (where $\{N\}'$ is the set of N_j excluding N_i) or from $(-1/V_m)(\partial V_m/\partial p)_{T,p}$, respectively.

The simulated excess (Ex) volume of mixing was calculated from $V^{Ex} = V_m - V_1^o x_1 - V_2^o x_2$ where V_m is the molar volume of the mixture and the o superscript denotes the pure liquid. The simulated partial molar volumes ($\bar{V}_i$) were calculated from $(\partial V/\partial N_i)_{T,p,\{N\}'}$ and from

$$\bar{V}_2 = \frac{1 + N_{11} - N_{21}}{\rho_1(1 + N_{22} - N_{12}) + \rho_2(1 + N_{11} - N_{21})} \qquad N_{ij} = \rho_j G_{ij} \qquad \rho_i = N_i/V \qquad \text{3.1}$$
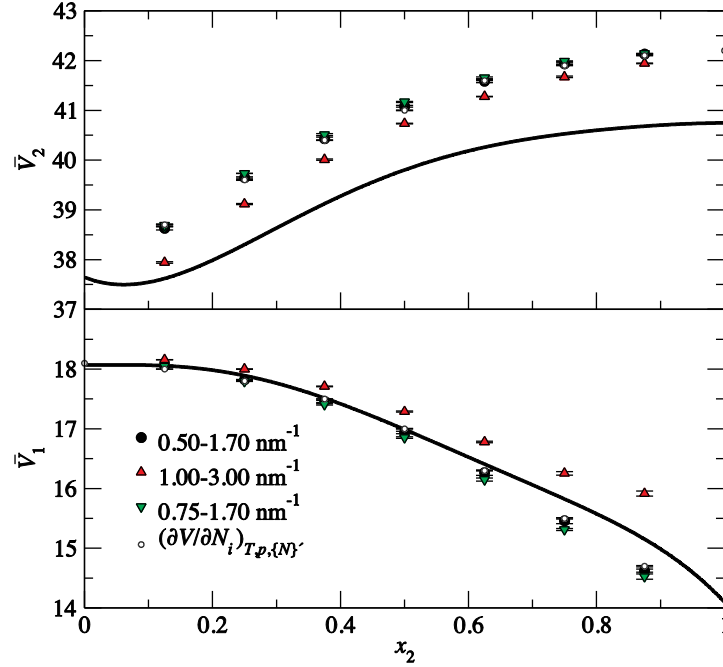
Figure 3.2 Internal consistency of partial molar volumes guided choice of linear region for Schnell approach determination of the Kirkwood-Buff integrals. The linear region taken as 0.50-1.70 nm^{-1}. $\{N\}'$ is the set of all N_j not equal to N_i.

where the partial molar volume of species 1 can be obtained by a simple change of indices. For the KBFF simulations, the agreement between these two methods is shown in Figure 3.2. The linear region used gave the best agreement for all models, not just the KBFF models. Similarly, the isothermal compressibility (κ_T) was obtained from $(-1/V_m)(\partial V_m/\partial p)_{T,p}$ for $x_{MOH} = 0.00$, 0.25, 0.50, 0.75, and 1.00 by performing additional simulations at pressures of 250, 500, 750, and 1,000 bar for 1 ns (polarizable simulations and pure compositions of nonpolarizable simulations) or 10 ns (nonpolarizable mixture simulations) of production after 100 ps of equilibration. The V versus p data was fitted to a quadratic polynomial to obtain κ_T. κ_T was also from the KBIs *via*

$$\kappa_T = \beta \frac{1 + N_{11} + N_{22} + N_{11}N_{22} - N_{12}N_{21}}{\rho_1(1 + N_{22} - N_{12}) + \rho_2(1 + N_{11} - N_{21})} \qquad 3.2$$

It was confirmed that these two methods were in very resonable agreement, further justifying the choice of the linear region used in the Schnell approach (shown for KBFF only in Figure 3.3).

The simulated excess heat of mixing was calculated from $H^{Ex} = H_m - H_1^o x_1 - H_2^o x_2$ where H was actually taken to be the potential energy instead of the enthalpy, which is reasonable since the pV term is negligible at 1 bar and assuming the kinetic energy contributions cancel in the mixture and the pure liquids. The simulated diffusion coefficients were obtained using the Gromacs 'g_msd' analysis tool, which calculates the mean square displacement (MSD) of molecules from their starting positions and then fits the linear region of the MSD versus time plot and uses the Einstein relation to determine the diffusion coefficients.
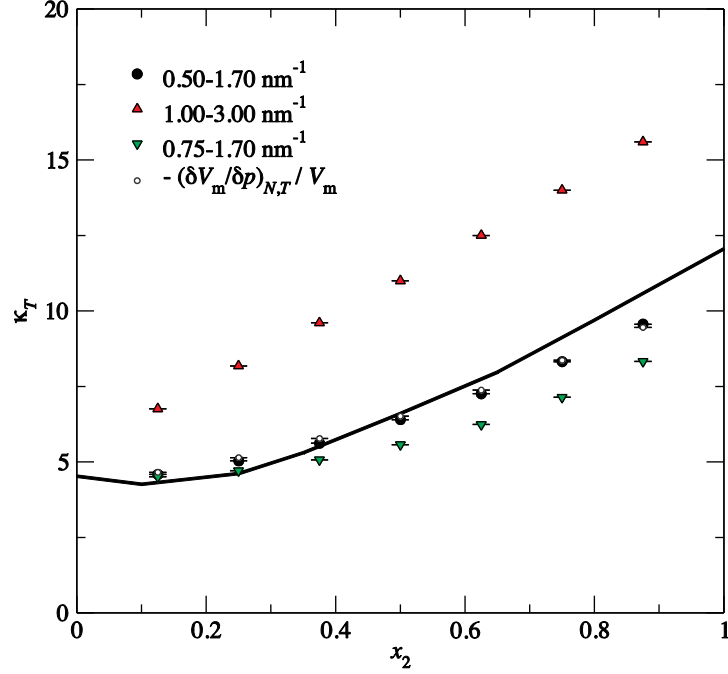
Figure 3.3 Internal consistency of isothermal compressibility, $\kappa_T \times 10^5$ (bar^{-1}), guided choice of linear region for Schnell approach determination of the Kirkwood-Buff integrals. The linear region taken as 0.50-1.70 nm^{-1}.

3.3.4 Determination of interfacial properties

The surface relative probability plots, $g_i(z)$, are plateaued at large z (see Results), indicating that the simulation boxes were large enough that a that a central bulk mixture region was obtained. The surface tension γ was calculated from the simulations according to

$$\gamma = \frac{1}{2}L_z[P_{zz} - \frac{1}{2}(P_{xx} + P_{yy})]$$

3.3

where $L_z = \sim 36$ nm is the box dimension in the extended (z) dimension and $P_{\alpha\alpha}$ are the diagonal elements of the pressure tensor.

The expression for the excess surface adsorption of species 2 (methanol) over species 1 (water), $\Gamma_{2,1}$, was derived from summing the Gibbs-Duhem equations for the interface (*), bulk gas phase, and bulk liquid phase at constant T,

$$N_1^* d\mu_1 + N_2^* d\mu_2 = -Ad\gamma$$
$$N_1^{\text{bulk } (v)} d\mu_1 + N_2^{\text{bulk } (v)} d\mu_2 = 0$$
$$N_1^{\text{bulk } (l)} d\mu_1 + N_2^{\text{bulk } (l)} d\mu_2 = 0$$

3.4

where A is the surface area of the interface. No Vdp term appears in the above equation, because the volume of the interface is negligible compared to the volume of the bulk phases. The above equation then leads to

$$N_1 d\mu_1 + N_2 d\mu_2 = -Ad\gamma$$

3.5

where $N_i = N_i^* + N_i^{\text{bulk}(v)} + N_i^{\text{bulk}(l)} = N_i^* + N_i^{\text{bulk}}$. The $d\mu_1$ term may then be eliminated by summing the second and third Gibbs-Duhem equations in Equation 3.4 leading to $d\mu_1 = -N_2^{\text{bulk}} d\mu_2 / N_1^{\text{bulk}}$, which can then be substituted into Equation 3.5 to yield

$$A\Gamma_{2,1}(Z) = N_2(Z) - N_1(Z)\frac{N_2^{\text{bulk}}}{N_1^{\text{bulk}}} \qquad\qquad 3.6$$

$N_i(Z)$ is the total number of molecules of species i found between $z = 0$ and $z = Z$. The bulk liquid phase region was taken to be the central $6 \times 6 \times 6$ nm region of the condensed phase; the bulk vapor phase region was taken as $z = 0 - 10$ nm and $z = 26 - 36$ nm nm (Figure 3.11). Bulk liquid and vapor densities were determined from averaging over these respective regions.

Equation 3.6 requires the determination of the simulated bulk densities before the calculation of the $\Gamma_{2,1}$ and it requires a decision of the extent of the bulk liquid and gas phases. Having to make the decision of where the bulk phases start and stop is not ideal, but it is not prohibitive as long as one avoids getting too near to the interfacial region. If significant evaporation does not occur, then the deviation of the bulk liquid composition from the total composition may be accounted for on the fly without a user defined definition of the extent of the bulk regions (Chen and Smith 2008).

The experimental $\Gamma_{2,1}$ is calculated from the composition derivative of the experimental surface tension (see below). Thus, an analytical form must be chosen for the correlation of surface tension with composition. The semi-empirical Connors-Wright equation (Connors and Wright 1989) was chosen due to its simple, two-parameter functional form and its established success and continued use for binary aqeous – organic liquid mixtures (Alvarez *et al.* 2003; Blanco *et al.* 2013). Many determinations of the surface tension for water-methanol mixtures exist in the literature. All sources at 298 K and 1 bar that were deemed to be without spurious results and had data points over the full composition range were used, in an attempt to measure the experimental error on the surface tension and how that error propagates into the experimental $\Gamma_{2,1}$ values. The Connors-Wright equation is (Connors and Wright 1989):

$$\gamma^* = (\gamma_1^o - \gamma) / (\gamma_1^o - \gamma_2^o) = x_2[1 + (ax_1) / (1 - bx_1)] \qquad\qquad 3.7$$

where the superscript o denotes the pure liquid, γ denotes the surface tension of the mixture, and thus the superscript * denotes the reduced surface tension. The values of a and b are given in Table 3.3 for the six different raw surface tension data sets used. From our comparison of raw data sets and possible correlating equations, we found that the $\Gamma_{2,1}$ values depended more on the functional form used to fit $\gamma(x_2)$ than they did on the raw data set itself, at least for this system.

Table 3.3 Connors-Wright parameters for fitting the surface tension dependence on composition

Reference	*a*	*b*
Maximino (2009)	0.878	0.799
Vazquez *et al.* (1995)	0.871	0.830
Cheong and Carr (1987)	0.905	0.748
Hu *et al.* (2007)	0.913	0.798
Teitelbaum *et al.* (1951)	0.894	0.766
Bandyopadhyay *et al.* (2010)	0.916	0.743

Upon fitting $\gamma(x_2)$, the experimental $\Gamma_{2,1}$ is then provided by

$$\Gamma_{2,1} = -\frac{\beta\eta x_2}{\rho}\left(\frac{\partial\gamma}{\partial x_2}\right)_{T,p}$$

3.8

where $\eta = \rho_1 + \rho_2 + \rho_1\rho_2\Delta G_{12}$, $\Delta G_{12} = G_{11} + G_{22} - 2G_{12}$, $\beta=1/RT$, and $\rho = N/<V> = \Sigma_i\rho_i$. η, ρ, and G_{ij} all correspond to experimental properties from the bulk solution, not properties of the vapor-liquid interface.

Using the KBFF simulations, we tested for convergence of the results with respect to system size and simulation time by comparing the (KBFF) 40 ns of production results to simulations of 100 ns in length for both the bulk and interfacial systems and by extending the interfacial systems to sytems of size $10 \times 10 \times 60$ nm (results not shown). It was determined that the system sizes and simulation lengths chosen were adequate.
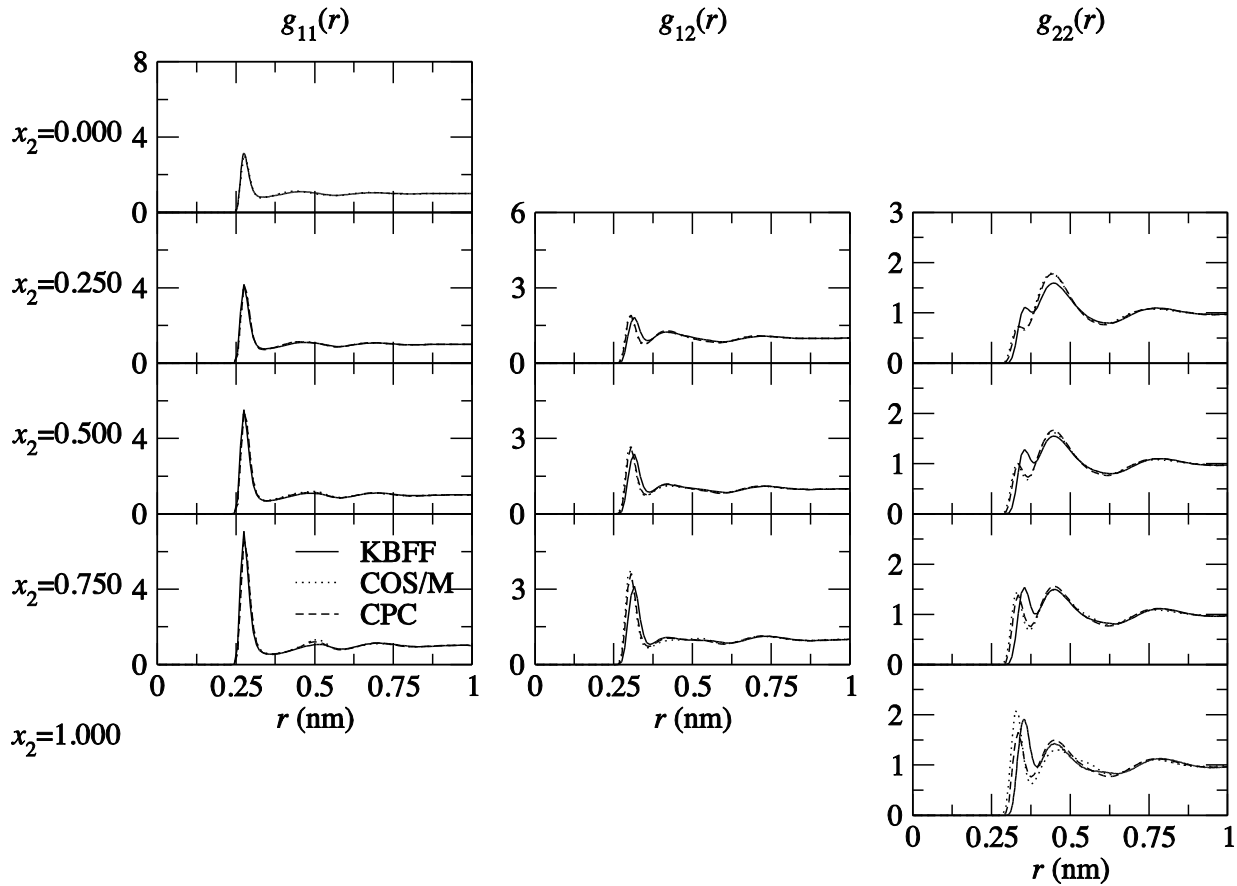

Figure 3.4 Center of mass based radial distribution functions (RDFs).

3.4 Results and Discussion

3.4.1 Simulations of the bulk water-methanol mixtures

The center of mass based radial distribution functions (RDFs) are shown in Figure 3.4 for a subset of the compositions studied. As the mole fraction of methanol increases, all three RDFs become more structured, indicating a less random distribution of molecules for x_2=0.750 than for x_2=0.250. The relative height of the first solvation shell increases with composition for g_{11} and g_{12}, while for g_{22} a peak grows in as the composition of methanol increases. Most of the differences occur when comparing the nonpolarizable results to the polarizable results, whereas the two polarizable results give relatively consistent results. All three force fields produce the same general features and the same trends with composition.

Figure 3.5 presents experimental and simulated values for the KBIs. The dashed lines show the results for the mixture if it were ideal on the mole fraction concentration scale (symmetric ideality), in which the excess properties are all zero and the activity coefficients are unity at all concentrations. It is important for FFs to be able to reproduce KBIs, because these quantities sensitively probe the intermolecular interactions within mixtures *e.g.*, they allow one to determine if the appropriate balance is obtained between solute-solute, solute-solvent, and solvent-solvent interactions. If the KBIs are well reproduced, then one has increased confidence in the simulated microstructural and thermodynamic properties.
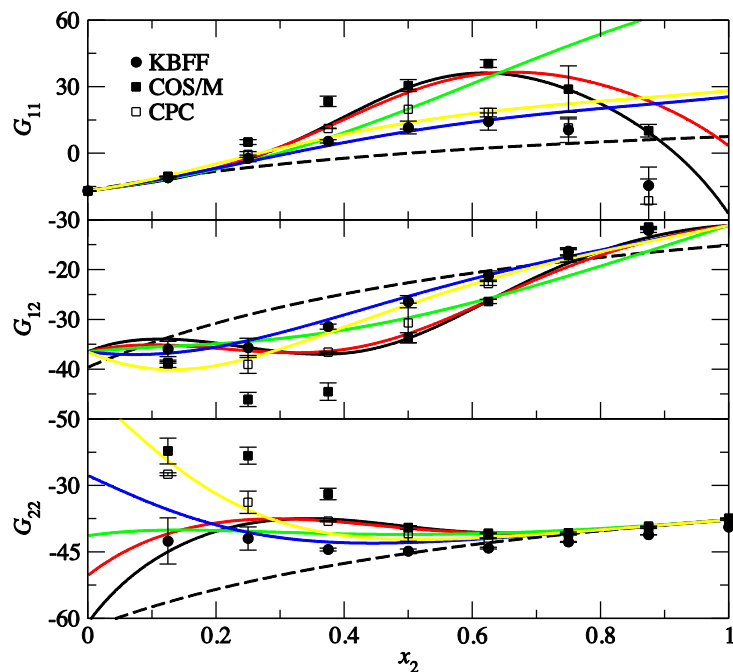
Figure 3.5 Bulk KBIs (cm³/mol). Symbols correspond to the simulated data. Experimental data were taken from the literature. All analysis was performed with the same density data (Douheret *et al.* 1989). Several data sets and fitting functions for G^{Ex} were explored: Black lines – Redlich-Kister equation (Butler *et al.* 1933); Red lines – modified Redlich-Kister equation (Hu *et al.* 2003); Green lines – Wilson equation (Soujanya *et al.* 2010); Blue lines – Wilson equation (Gmehling *et al.* 1991); and Yellow lines – Wilson equation (Gmehling *et al.* 1991). Dashed lines represent the corresponding symmetric ideal results.

Without the context provided by comparing the results of Figure 3.5 to those of other binary mixtures, it may appear that there is tremendous dispersion in the experimental KBIs depending upon the raw excess Gibbs free energy data set and/or the correlating equation used (Redlich-Kister, Wilson, *etc.*) (and similarly the force field used). While the dispersion does confound the comparison of the FFs, it should be noted that the water-methanol mixture KBIs are relatively small and relatively ideal compared to many of the other systems we have examined (Ploetz and Smith 2011a; Jiao *et al.* 2014). Additionally, it is well known that the experimental and simulated G_{ii} will be statistically more noisy at low mole fraction of i since fewer i molecules are present. This explains why the spread in the simulated and experimental results for G_{11} are greatest at low x_1 and for G_{22} are greatest at low x_2.
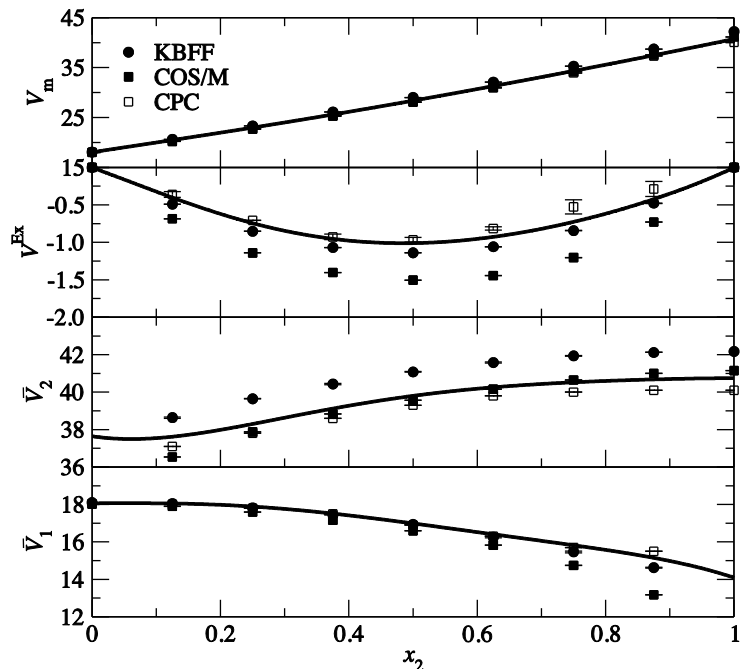
Figure 3.6 Molar volumes, excess molar volumes, and partial molar volumes (cm³/mol). Solid lines: experimental values, points: simulation values.

From Figure 3.5, it can be seen that the CPC and KBFF models were always within the spread of the different experimental determinations, excluding the values at $x_2 = 0.875$. The slightly too positive G_{11} and G_{22} values, leading to the too negative G_{12} values, in the simulations of the COS/G2 + COS/M mixtures at $x_2 \leq 0.375$ indicate that the net methanol-methanol and water-water interactions are slightly too attractive or water-methanol interactions are slightly too repulsive. It should be noted that all three models give very reasonable results for the KBIs even though, in contrast to the COS-based models, the KBFF models were specifically parameterized to reproduce the KBIs. However, one would expect that the polarizable FFs might inherently perform better at reproducing the KBIs. The principle focus of the KBFF philosophy has been to develop effective partial charges that can best describe the composition dependent properties from one extreme (*e.g.*, pure water) to another (*e.g.*, pure methanol). As this flexibility is inherent to polarizable FFs, it is not too surprising that reasonable results are also obtained for the KBIs from polarizable models.

Figure 3.6 presents experimental and simulated volumetric data. KBFF and CPC both reproduce the excess volume of mixing well, whereas COS/M values are too negative, which is consistent with previous results (Vosmeer *et al.* 2012). The absolute value of the partial molar volume of methanol is best reproduced by the polarizable models, with KBFF reproducing the correct trend with composition but being consistently too large. The partial molar volume of water as a function of composition was best reproduced by COS/G2 + CPC and SPC/E + KBFF, while COS/G2 + COS/M, gave partial molar volumes of water that were consistently too small.

Figure 3.7 shows that KBFF gave very reasonable, but consistently too small, values for the isothermal compressibility, whereas COS/M was too high and CPC was too low as the concentration of methanol increased.

Figure 3.8 (top panel) shows the enthalpy of mixing results. Both the KBFF and the CPC models agree well with experiment, with the KBFF results being slightly closer to experiment. In the COS/M simulations, the enthalpy of mixing values were too positive. Note that Patel and co-workers (Zhong *et al.* 2008) found too negative values (by 1-2 kJ mol^{-1}) for the heat of mixing when using the polarizable CHARMM-based fluctuating charge force field.

The second panel in Figure 3.8 indicates that the KBFF model best reproduced the experimental dielectric constant, while both polarizable FFs reported dielectric constants that were systematically too large at high mole fractions of water. This is in contrast to the dielectric constant for pure COS/G2 water reported in the literature. The time dependence of the dielectric constant was investigated and, when using short (ps) time scales like those reported in the publication of the COS/G2 model, agreement with the literature value of 88 was approximately achieved. However, that lower dielectric constant was not converged with respect to time in the current simulations.
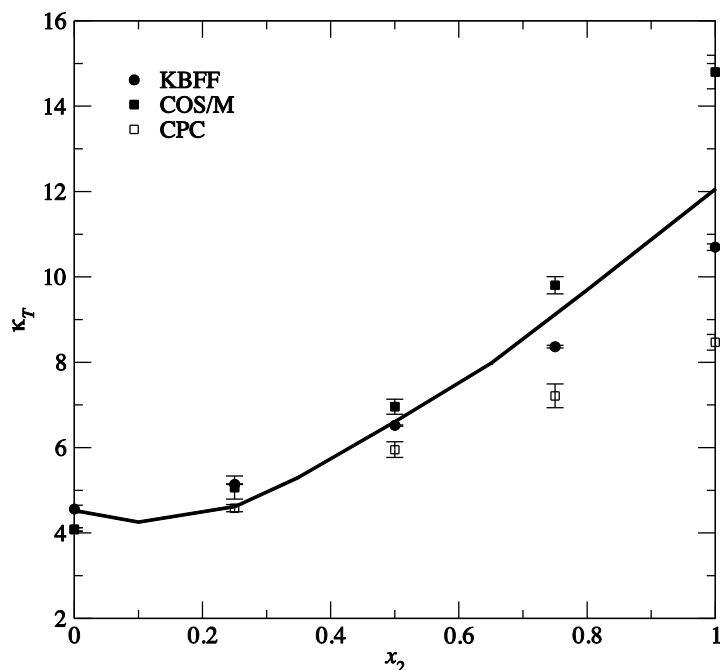

Figure 3.7 Isothermal compressibility, $\kappa_T \times 10^5$ (bar^{-1}), obtained from pressure dependence of the system volume. Solid lines: experimental values, points: simulation values.
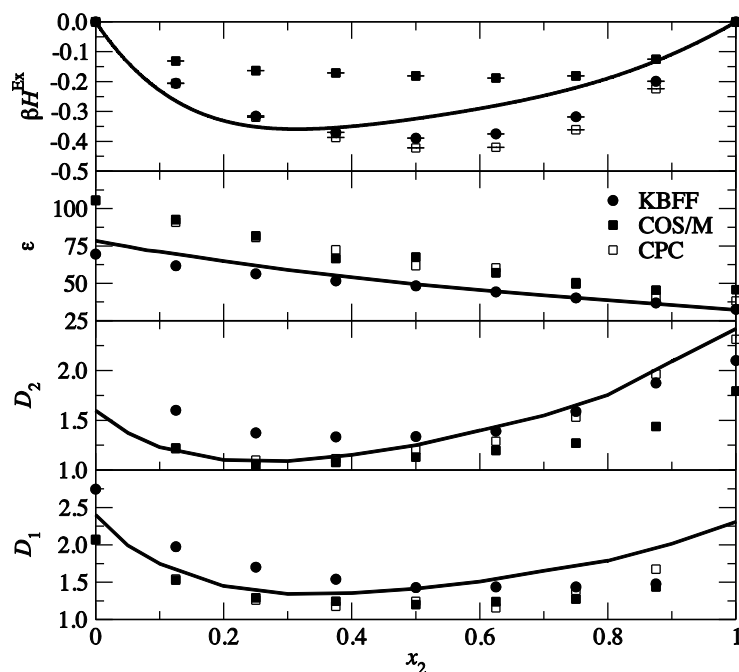
Figure 3.8 Enthalpy of mixing (dimensionless), static dielectric constant (dimensionless), and translational self-diffusion constants × 10^9 (m^2s^{-1}). Solid lines: experiments, points: simulation values (typical error on diffusion constant is < $3.5×10^{-11}$ m^2s^{-1}).

The bottom two panels in Figure 3.8 indicate that CPC yielded the best diffusion coefficient values, while KBFF and COS/M both struggled to reproduce the correct trends with composition, but still gave very reasonable magnitudes.

3.4.2 Simulations of the water-methanol mixtures at the vapor-liquid interface

The polarizable simulations of the vapor-liquid interface led to a significant amount of evaporation, whereas evaporation in the KBFF simulations was negligible. Many more methanol and water molecules evaporated with the polarizable FFs than would be expected from an ideal gas treatment of the vapor phase. Using a vapor phase volume corresponding to that of our simulation box, $V_{vapor} = 24 × 6 × 6$ nm, $T = 298$ K, and $N_A = 6.022×10^{23}$ molecules/mole, $N_{i,vapor} = \beta p_{i,vapor}V_{vapor}N_A$ was calculated for each composition. Vapor partial pressures, $p_{i,vapor}$, for this calculation were taken from the 298 K isothermal vapor liquid equilibria data of Hu *et al.* (2003). Table 3.4 shows the number of molecules in the vapor phase averaged over the full production run versus the number expected using the above ideal gas treatment of the vapor phase. The simulation that exhibited the most evaporation, COS/G2 + CPC at $x_{MOH} = 0.125$, corresponds to an evaporation of less than 4% of the total molecules in the system.

Table 3.4 Average number of water(1) or methanol(2) molecules in the vapor phase.

	$x_2 = 0.000$	$x_2 = 0.125$		$x_2 = 0.250$		$x_2 = 1.000$
	H_2O	H_2O	CH_3OH	H_2O	CH_3OH	CH_3OH
COS/M	349	27	16	26	25	10
CPC	349	301	102	241	106	6
KBFF	0	0	0	0	0	1
Expected	1	1	1	1	1	4
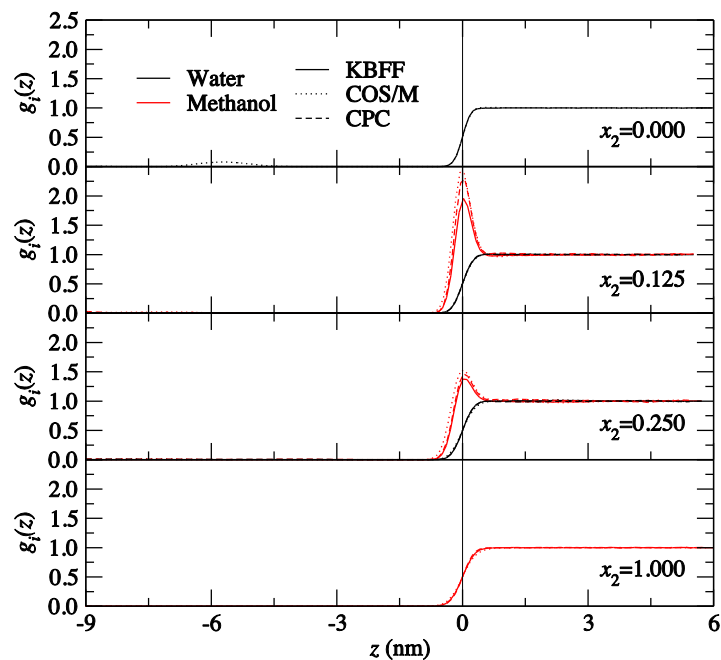
Figure 3.9 Relative surface probability $g_i(z)$ as a function of z (nm) for the interfacial systems. Values were set to $g_1(z=0.0nm)=0.5$ (all compositions except pure methanol) or $g_2(z=0.0nm)=0.5$ (pure methanol only).
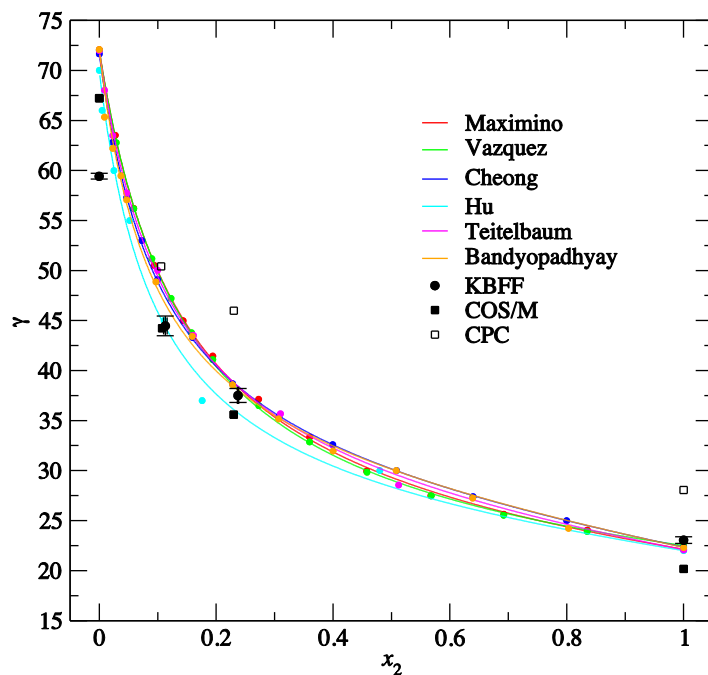
Figure 3.10 Raw experimental (points without error bars), fitted experimental (lines), and simulated (points with error bars) surface tension, γ (dynes/cm = mN/m), versus mole fraction of methanol, x_2. All experimental determinations were at 298 K and 1 bar (Teitelbaum *et al.* 1951; Cheong and Carr 1987; Vazquez *et al.* 1995; Hu *et al.* 2007; Maximino 2009; Bandyopadhyay *et al.* 2010). The non-linear decrease in the surface tension as x_2 increases indicates adsorption of methanol in the interface region.

In Figure 3.9 the relative surface probability $g_i(z)$ plots are displayed as a function of z. For the pure liquids, no peak is observed, but in the mixtures there is a single peak in the methanol distribution, located at the interface. The peak is located at smaller z values than is the rise in the water probability, which indicates that the methanol preferentially adsorbs at the interface. No further structure (second or subsequent peaks) is observed, indicating that the ordering of methanol molecules at the interface is not long-ranged. At large negative z-values (vapor phase), non-zero $g_i(z)$ values for the polarizable force fields are due to evaporation.

The surface tension results are shown in Figure 3.10. We have plotted six raw experimental determinations (points without error bars) at 298 K and 1 bar. We chose to fit the surface tension data (lines) with a functional form that is established in the literature, the Connors-Wright equation (Connors and Wright 1989). From our observations, differences in the surface excess (see Figure 3.12) were more dependent upon the functional form of the analytical equation used to describe the dependence of the surface tension on composition than they were on the raw surface tension data, at least for water-methanol mixtures.

The calculated surface tension value (Figure 3.10) from the pure COS/G2 simulation is much closer to the experimental value than is the surface tension of SPC/E. However, the surface tension for the KBFF simulations are quantitatively in better agreement with experiment than are the polarizable simulations for both mixture compositions studied, especially when compared to the CPC simulations. For pure methanol, KBFF has a slightly too high surface tension and COS/M has a slightly too low surface tension, while the CPC surface tension is comparatively much too high.

The computed excess surface adsorption, $\Gamma_{2,1}$ as a function of distance is shown in Figure 3.11. $Z = 0.0$ nm denotes the edge of the vapor phase furthest from the liquid, with increasing Z approaching the interface and passing through it, entering the liquid phase, and eventually reaching the center-most point of the bulk liquid region at $Z = 18.0$ nm (see 25% opacity underlay of the system in Figure 3.11). Non-zero values of $\Gamma_{2,1}$ at low Z were observed for both polarizable FFs, but not for the nonpolarizable simulations, due to the evaporation that occurred in the polarizable simulations. The values of $\Gamma_{2,1}$ increase until there is no longer any excess of molecule 2 (methanol) over molecule 1 (water) (after normalizing for the overall composition), at which point $\Gamma_{2,1}$ plateaus. The simulated values have not plateaued perfectly, but tests of the KBFF simulations using a larger interface system of $10 \times 10 \times 60$ nm showed no statistically significant differences (not shown), indicating the system size is probably large enough. Similarly, extension of the simulation times to 100 ns (KBFF simulations only) also showed no statistically significant differences, suggesting that the simulation lengths used here were probably sufficient. Our tests with KBFF indicate that extensions to longer simulation times or larger system sizes may change the quantitative numbers slightly, but would not change the qualitative differences observed among the force fields when comparing them to each other or to experiment. The values of $\Gamma_{2,1}$ at $Z = 18.0$ nm were taken for comparison with the experimental $\Gamma_{2,1}$ values (as shown in Figure 3.12).
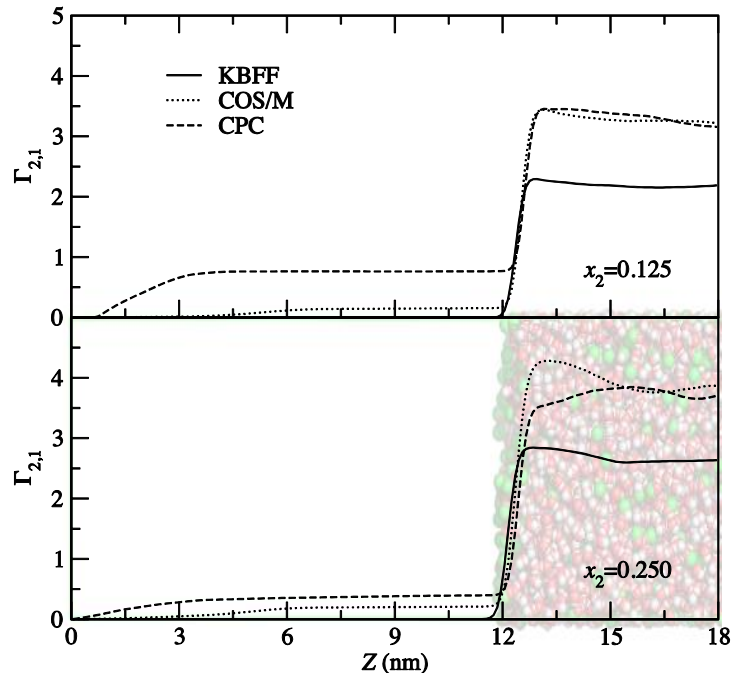
Figure 3.11 Simulated excess surface adsorption, $\Gamma_{2,1}$ (molecules/nm^2), as a function of Z (nm) for x_2=0.125 (top) and x_2=0.250 (bottom) vapor-liquid interface systems. Due to the symmetry of these systems, in this analysis the systems were folded in half about the center (z = 18 nm) of the liquid phase as illustrated by the 25% opacity schematic of the system in the bottom panel. Non-zero values at small Z are due to evaporated molecules in the polarizable simulations.
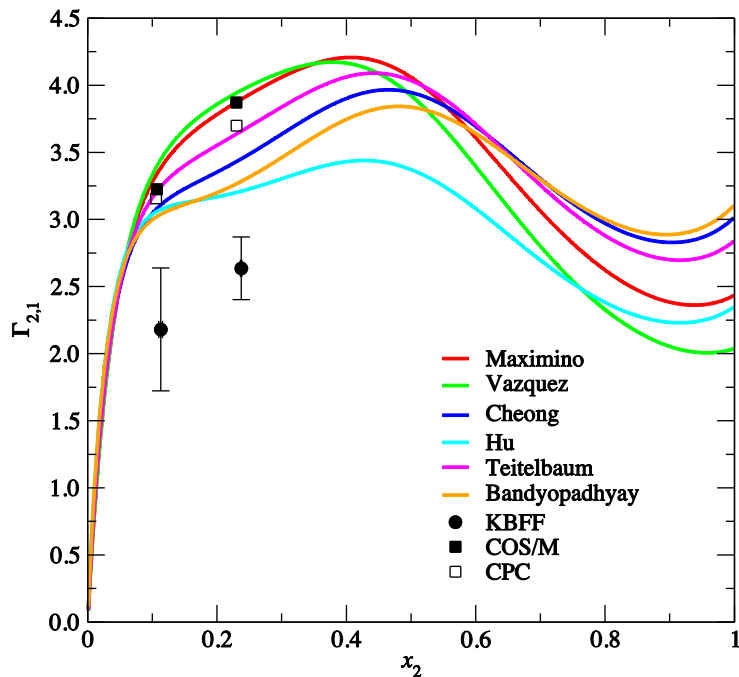
Figure 3.12 Excess surface adsorption, $\Gamma_{2,1}$ (molecules/nm^2), *vs*. x_2. Lines: experiment, points: simulation.

.

In Figure 3.11 it is clear to see that the values of $\Gamma_{2,1}$ were the lowest at both mole fractions in the KBFF simulations. This explains why the compositions in Figure 3.10 are the closest to the bulk values in the KBFF simulations (they are only slightly less than 0.125 and 0.250), whereas the simulated bulk mole fractions of methanol are smaller for CPC and COS/M.

In Figure 3.12 we show the experimental excess surface adsorption as a function of composition for the six experimental surface tension data sets correlated with the Connors-Wright equation and the simulated $\Gamma_{2,1}$ values. We did not propagate the experimental dispersion in the bulk KBIs (as shown in Figure 3.5 into the measure of the spread in the excess surface adsorption. Due to the combination of KBIs that appear in the calculation of $\Gamma_{2,1}$ (Equation 3.8), it is likely that some of the dispersion would cancel when propogated into $\Gamma_{2,1}$.

KBFF did not show as high of an excess surface adsorption of methanol molecules at the vapor-liquid interface as was observed experimentally. The KBFF simulated values reported here agree with the previous results for the KBFF model (Chen and Smith 2008). It appears that the low KBFF $\Gamma_{2,1}$ values may be due to the low surface tension of pure SPC/E water more than a problem with the KBFF methanol or even the mixtures, which complicates any possible conclusion about the need for polarization. The CPC and COS/M methanol models are both within the spread of the experimental determinations.
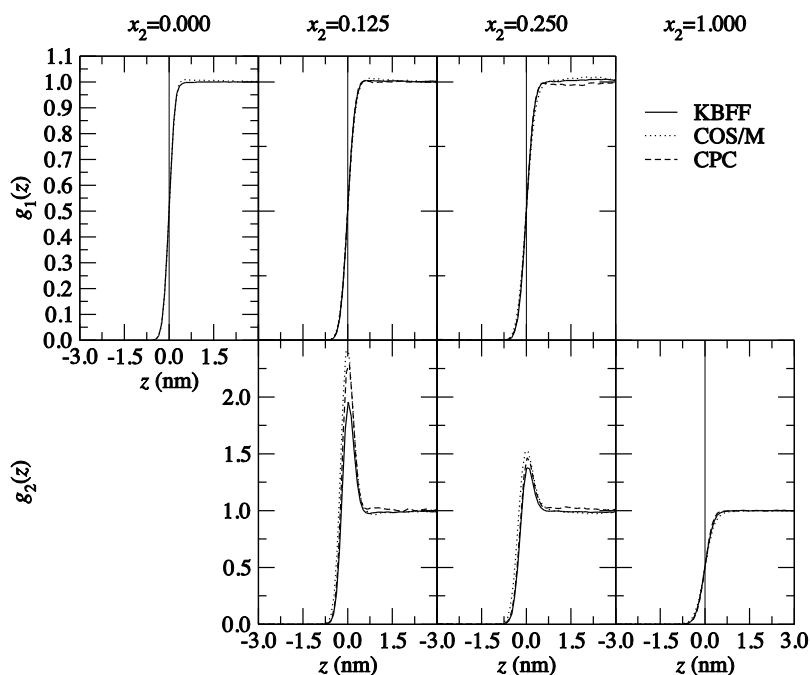
Figure 3.13 Relative surface probabilities, $g_i(z)$, in the interface region. Values were shifted to $g_1(z=0.0\text{nm})=0.5$ (all compositions except pure methanol) or $g_2(z=0.0\text{nm})=0.5$ (pure methanol only).
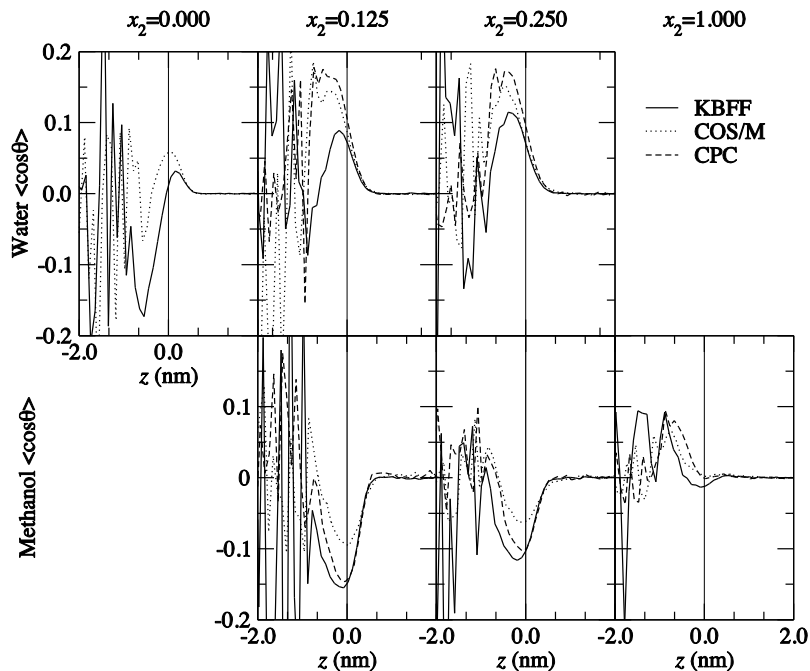
Figure 3.14 Cosine of the angle of the dipole moment to the positive *z*-axis, <cosθ>, for methanol and water as a function of distance from the interfacial region (located at *z* = 0.0 nm). $0 \leq \theta \leq \pi$ radians. Values are averaged over the full production simulation and over all molecules of species *i*. The *z*-axis was shifted so that $g_1(z=0.0nm)=0.5$ (all compositions except pure methanol) or $g_2(z=0.0nm)=0.5$ (pure methanol only). Non-zero values for *z*-values less than ~-1 nm are noise from the relatively few evaporated molecules. There appears to be a change in angle for water and for methanol on going from the neat liquids to the mixtures.

Figure 3.13 shows the relative probability plots of methanol and of water in the interface region for the four compositions studied. The adsorption of methanol molecules gives rise to a strong peak in the relative probablity at the interface in the mixture and the peak occurs at smaller *z* values than the plateau in the water probablity plot, showing the adsorption of the methanol. The peak was greater for the lower mole fraction of methanol, but upon integration, it gives rise to a lower excess surface adsorption.

In Figure 3.14 the cosine of the average angle, $0 \leq \theta \leq \pi$ radians, that the dipole moment of each molecule is making with the positive *z*-axis is plotted after averaging over all molecules of a given species and over the full production simulation. All the force fields indicated the same qualitative changes with respect to composition.

Figure 3.15 Total system charge (units of the elementary charge, *e*) as a function of *z*. The *z*-axis was shifted so that $g_1(z=0.0\text{nm})=0.5$ (all compositions except pure methanol) or $g_2(z=0.0\text{nm})=0.5$ (pure methanol only).

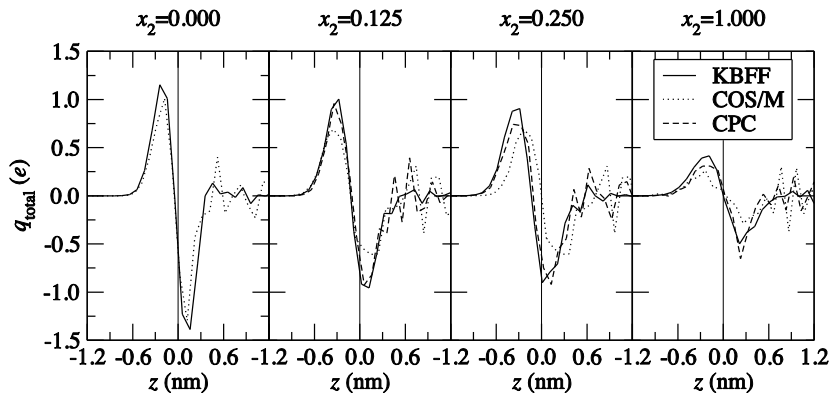The most remarkable (but well-known) feature from these plots is that neat water ordering is different from the ordering of water in the presence of the amphipathic methanol molecule. It is also known, and our results agree, that neat water has two layers with different structures. On the vapor side of the interface, the dipoles are tilted towards the vapor phase (negative $<\cos\theta>$). On the liquid side of the interface, the dipoles are (subtly) tilted towards the liquid phase (small, positive $<\cos\theta>$). Pure SPC/E exhibited more (less) ordering than pure COS/G2 on the vapor (liquid) side of the interface.

For neat methanol, our results showed that the dipole moment was pointed toward the bulk on the vapor side of the interface and very subtly towards the vapor on the liquid side of the interface. KBFF methanol was the most ordered on the liquid side of the interface and CPC was the most ordered on the vapor side of the interface.

In the mixtures, the ordering of each component slightly increased as its concentration decreased, as has been observed before. Considering the water orientation in the mixtures, our results showed that the dipole moment of the interfacial waters was pointing toward the bulk (positive $<\cos\theta>$), as has also been observed. The most ordered water distribution was observed for COS/G2 when mixed with CPC, followed by COS/G2 when mixed with COS/M, making SPC/E mixed with KBFF methanol the least ordered water distribution. Considering the methanol orientation from the mixtures, our results indicated that the dipole moment of methanol was orientated towards the vapor phase (negative $<\cos\theta>$). The most ordered methanol distribution from the mixtures was observed for the KBFF simulations, whereas the COS/M model exhibited the least ordered methanol distribution.

In Figure 3.15 the total charge of the system, q_{total}, is plotted in units of the elementary charge (*e*) as a function of *z*. Again, the FFs all indicate the same trends with composition, but have different quantitative magnitudes The results show the presence of a surface dipole with the water hydrogens or methanol methyl group closest to the interface (presumably the smallest penalty to pay) to maximize the number of hydrogen bonding partners in the liquid phase.

57

While our results for bulk mixture and interfacial properties do not indicate that polarizable FFs are fundamentally better than nonpolarizable FFs, our study is certainly not the first to attempt to answer this question, nor is it the first to end up concluding that the results were mixed. As a few examples, results from calculations on the surface tension of liquid ethanol (Patel and Brooks 2005) and at the hexane-water interface (Patel and Brooks 2006) were found to be better reproduced using polarizable force fields instead of nonpolarizable ones, and a simulation study of Rivera *et al.* (Rivera *et al.* 2006) pointed out the importance of polarization contributions to the surface tension of liquid water. However, this conclusion was debated by Voth and co-workers (Iuchi *et al.* 2007), who found nearly identical values for the surface tension of water in simulations using the polarizable TTM2-R model or a nonpolarizable model parameterized based on bulk TTM2-R water simulations. A quantitative comparison of computed surface tensions for water-methanol mixtures in previous simulation studies (Matsumoto *et al.* 1993; Chang and Dang 2005; Paul and Chandra 2005; Partay *et al.* 2008) with nonpolarizable or polarizable force fields did not give a clear picture of the importance of explicitly including electronic polarization effects. Two of the studies (Matsumoto *et al.* 1993; Paul and Chandra 2005) that made use of nonpolarizable models gave too negative values for the surface tension when compared to experiment, whereas another one (Partay *et al.* 2008) showed values that were significantly too large for the surface tension of the pure liquids and values close to experiment at intermediate compositions. Chang and Dang (Chang and Dang 2005) reported slightly too large values for the surface tension of aqueous methanol solutions when using a polarizable force field based on the inducible dipole model.

3.5 Conclusions

We have compared two polarizable force fields for the water-methanol mixture to a nonpolarizable force field and looked at both bulk properties and those at the vapor-liquid interface for a few compositions. It is difficult to compare the FFs and determine an overall "winner" and/or "loser." The differences in the results between the two polarizable methanol molecules, which were in some cases just as large as the differences between the polarizable and the nonpolarizable results, indicate that there is no fundamental guarantee that a polarizable FF will be inherently better that a non-polarizable FF. The results seem to depend first and foremost upon the quality of the individual FF parameterization. As shown here, at least some nonpolarizable FFs are capable of producing reasonable results for the vapor-liquid interface. This may be because a nonpolarizable simulation still accounts for the first-order effect of having a lack of molecules on one side of the interface and an abundance of molecules on the other side of the interface.

We, as a community, still need to test more systems and more FFs to truly answer the question of when polarization needs to be explicitly included. The force field comparison presented here, and most of those in the literature, are confounded by the fact that the force fields under scrutiny were all parameterized differently. Depending upon the quality of the individual parameterizations and what properties were specifically parameterized, it is difficult to draw any strong conclusions. In the future, we plan to investigate the ability of a polarizable variant of TIP3P water that reproduces the properties of nonpolarizable (regular) TIP3P water under ambient conditions (Leontyev and Stuchebrukhov 2012). This would remove the afore mentioned confounding variables; however, this study would not allow for the testing of mixture properties and is, therefore, greatly limited in its own way.

4. Local Fluctuations in Solution Mixtures[*][†]

Abstract

An extension of the traditional Kirkwood-Buff (KB) theory of solutions is outlined that provides additional fluctuating quantities that can be used to characterize and probe the behavior of solution mixtures. Particle-energy and energy-energy fluctuations for local regions of any multicomponent solution are expressed in terms of experimentally obtainable quantities, thereby supplementing the usual particle-particle fluctuations provided by the established KB inversion approach. The expressions are then used to analyze experimental data for pure water over a range of temperatures and pressures, a variety of pure liquids, and three binary solution mixtures – methanol and water, benzene and methanol, and aqueous sodium chloride. In addition to providing information on local properties of solutions it is argued that the particle-energy and energy-energy fluctuations can also be used to test and refine solute and solvent force fields for use in computer simulation studies.

4.1 Introduction

Kirkwood-Buff (KB) theory, more generally known as the Fluctuation Theory (FT) of solutions, has provided a wealth of data concerning the properties of solution mixtures (Kirkwood and Buff 1951; O'Connell 1971b; Matteoli and Mansoori 1990; Ben-Naim 2006; Pierce *et al.* 2008). Specific examples include studies of preferential solvation in binary solutions (Matteoli and Lepori 1984; Matteoli and Mansoori 1990; Marcus 2001; Ben-Naim *et al.* 2008), transfer free energies and the solubility of solutes (O'Connell 1971a; Chialvo 1993; Mazo 2006), surface tension and free energy changes due to the addition of additives (Chen and Smith 2008; Smith 2010), cosolvent effects on both small solutes and proteins (Chitra and Smith 2001b; Aburi and Smith 2004; Shimizu 2004; Smith 2004a; Ruckenstein and Shulgin 2006), and the investigation of protein-protein interactions (Pjura *et al.* 1995). The approach has also been used as a basis for the generation of improved force fields for computer simulation (Ploetz *et al.* 2010a; Weerasinghe *et al.* 2010). In all these examples KB theory provides a rigorous statistical mechanical formulism used to provide a link between thermodynamic data concerning the system of interest, and integrals over the underlying molecular distributions corresponding to local regions within the system.

The primary quantities of interest which help to characterize and quantify the above effects are the KB integrals (KBIs) between the different species present in solution. The KBIs (G_{ij}'s) can be expressed in terms of either radial distribution functions (g_{ij}) between species i and j, or as particle-particle fluctuations corresponding to local regions within the system of interest (Kirkwood and Buff 1951),

$$G_{ij} = G_{ji} = 4\pi \int_0^\infty \left[g_{ij}(r) - 1 \right] r^2 dr = V \left[\frac{\langle \delta N_i \delta N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - \frac{\delta_{ij}}{\langle N_i \rangle} \right] \qquad 4.1$$

where δ_{ij} is the Kronecker delta function, and $\delta N_i = N_i - <N_i>$ is the deviation in the number of i particles from the average number of i particles in the fixed volume (V) for each member of the grand canonical ensemble. Combinations of the KBIs and number densities then provide expressions for thermodynamic properties - such as chemical potential derivatives, partial molar volumes and isothermal compressibilities - in essentially any ensemble with the same average properties as the open system. Alternatively, given an appropriate set of experimental data as a function of composition one can invert the whole procedure and extract the experimental KBIs (Ben-Naim 1977), thereby providing information concerning the local distribution of species in solution. KB theory is particularly attractive in this regard as the expressions are exact, they can be applied to any stable solution mixture involving any type of molecule at any composition, and the resulting KBIs can be obtained relatively easily from computer simulation data.

The vast majority of experimental data analyzed by the KB inversion procedure involve binary or ternary systems at constant pressure (p) and temperature (T) where the results are presented in terms of local particle-particle correlations. Far less attention has been paid to the study of local particle-energy and energy-energy fluctuations in mixtures which naturally arise for thermodynamic temperature derivatives such as the thermal expansion and constant pressure (or volume) heat capacities. The main studies of relevance to the present work include the development of expressions for the partial molar energies and constant volume heat capacity in terms of molecular distribution functions provided by Buff and Brout (Buff and Brout 1955), and a series of papers by Debenedetti outlining the calculation of partial
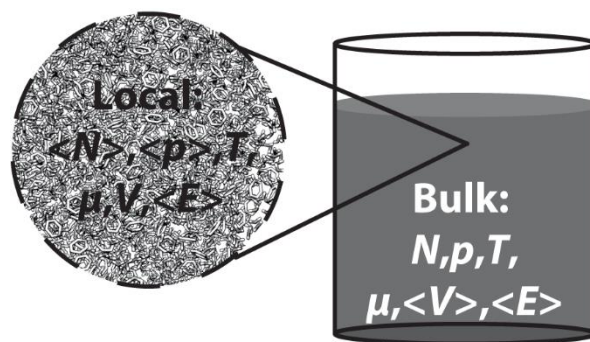
Figure 4.1 Local fluctuations in solution mixtures. The local region of interest is surrounded by the macroscopic bulk solution and therefore has the same intensive thermodynamic parameters (pressure, temperature, molar energy, molar volume, number densities, chemical potentials, *etc*.) as the bulk solution. Fluctuations in the number of particles and energy occur within the fixed local volume, which can be considered to be under the thermodynamic constraints associated with the grand canonical (μ*VT*) ensemble. The extent of the local region will depend on the pressure, temperature, composition and the nature of the solution components. However, the exact size of this region is irrelevant as long as it is large enough that the region displays the same intensive properties as the bulk solution – typically 3-5 solvation shells.

molar energies (and thereby enthalpies) from computer simulation (Debenedetti 1987b, 1988; Debenedetti 1989). However, to our knowledge, the corresponding local fluctuating quantities for mixtures have not been extracted from available experimental data. This is the major aim of this study.

In the following sections we outline the theory behind a general fluctuation description of solutions. The current approach is somewhat different from previous approaches (Buff and Brout 1955; Debenedetti 1987b), but is used here as it provides a simple formulation of the inversion process {experimental data} → {fluctuating properties} for multicomponent systems. The overall approach is then illustrated using available experimental data for a variety of solutions and solution mixtures, including salt solutions.

4.2 Theory

The basic aim is to develop expressions for properties of local regions within solution mixtures by reference to the equations of the grand canonical ensemble. Hence, the angular brackets used here signify an ensemble average in the grand canonical ensemble unless stated otherwise. The resulting expressions, involving fluctuations in the energy and number of particles, can then be related to local (μ*VT*) regions within systems corresponding to a variety of different ensembles possessing the same average properties, although the *NpT* ensemble will be the one of primary concern.

In our previous and subsequent discussions we continually refer to the concept of local fluctuations within the solution mixture of interest. Hence, it is important to be clear exactly what is meant by the term "local". Clearly, Equation 4.1 does not appear to represent a property characterizing a local region in the solution. However, the integration to infinity actually represents the extent of the open system. In the case of closed systems, one can consider finite microscopic regions within the solution to represent the open system of interest, the extent of which is dictated by the requirement

that all g_{ij} are unity at large distances. The fluctuations observed within this microscopic region are the "local" fluctuations of interest here. This is illustrated in Figure 4.1. It should be noted, however, that the extent of this microscopic region is generally unknown. For the majority of solutions under ambient conditions the local region may merely extend over a few nanometers. Alternatively, for solutions approaching a critical point the extent of the region will become macroscopic in size and therefore may no longer be considered as "local". As the majority of solutions fall into the former category, and to help distinguish the present fluctuations from the bulk system fluctuations, we have used "local" to describe the fluctuations provided by the current analysis.

In the grand canonical ensemble the average number of i molecules ($< N_i >$) is a function of the set of chemical potentials ($\{\beta\mu\}$), the volume, and the temperature ($\beta = 1/RT$), where R is the Gas constant. Hence, one can write the following differential for these independent variables,

$$d\langle N_i \rangle = \left(\frac{\partial \langle N_i \rangle}{\partial V}\right)_{\{\beta\mu\},\beta} dV + \left(\frac{\partial \langle N_i \rangle}{\partial \beta}\right)_{\{\beta\mu\},V} d\beta + \sum_j \left(\frac{\partial \langle N_i \rangle}{\partial \beta\mu_j}\right)_{\{\beta\mu\}',V,\beta} d\beta\mu_j \qquad 4.2$$

where the summation is over all j components in the mixture and the prime in the subscript indicates that all $\beta\mu$ except for the one of interest are held constant. The above thermodynamic derivatives can be related to their corresponding fluctuating quantities using the equations for a classical grand canonical ensemble (Davidson 1962),

$$pV = RT \ln \Xi(\{\beta\mu\},V,\beta)$$
$$d(pV) = SdT + pdV + \sum_i N_i d\mu_i \qquad 4.3$$

$$\Xi(\{\beta\mu\},V,\beta) = \sum_{\{N\}=0}^{\infty} e^{\beta\mu \cdot N} Q(\{N\},V,\beta)$$

where Q is the classical canonical partition function provided by,

$$Q(\{N\},V,\beta) = \left[\prod_i \frac{1}{N_i!}\left(\frac{q_i^{int}}{h^3}\right)^{N_i}\right] \iint_{-\infty}^{+\infty} e^{-\beta E'} \{dp\}\{dq\} \qquad 4.4$$

and $\{dp\}$ and $\{dq\}$ are the generalized momenta and coordinates associated with the molecules, respectively. The internal partition function (q_i^{int}) of molecule i contains the contribution from all the intramolecular degrees of freedom, which are assumed to be decoupled from the intermolecular degrees of freedom. Using the above equations one can show that the required derivatives are given by the following expressions (Buff and Brout 1955; Debenedetti 1987b),

$$\left(\frac{\partial \langle N_i \rangle}{\partial \beta\mu_j}\right)_{\{\beta\mu\}',V,\beta} = \langle \delta N_i \delta N_j \rangle = \langle N_i \rangle(\delta_{ij} + N_{ij})$$

$$\left(\frac{\partial \langle N_i \rangle}{\partial V}\right)_{\{\beta\mu\},\beta} = \frac{\langle N_i \rangle}{V} = \rho_i \qquad 4.5$$

$$\left(\frac{\partial \langle N_i \rangle}{\partial \beta}\right)_{\{\beta\mu\},V} = -\langle \delta N_i \delta E \rangle$$

Here, E is the total internal energy for each member of the ensemble, and includes any intramolecular contributions from the internal partition function. Hence, combining Equations 4.2 and 4.5 one finds that,

$$d\ln\rho_i = -F_{\mu,i}d\beta + \sum_j \left(\delta_{ij} + N_{ij}\right)d\beta\mu_j \qquad 4.6$$

for any species in any multicomponent mixture and where we have defined the energy-number fluctuations by,

$$F_{\mu,i} = \frac{\langle \delta N_i \delta E \rangle}{\langle N_i \rangle} . \qquad 4.7$$

The F_μ's are intensive quantities that represent correlations between the energy and the number of i particles in the region of interest. The $N_{ij} = \rho_j G_{ij}$ quantities can be considered as either particle number fluctuations for the same local region, or as an excess coordination number, *i.e.* the change in the number of j particles resulting from the introduction of an i particle to the reference volume compared to the number of j particles observed in the same volume of the bulk solution. If the temperature is kept constant the above set of equations reduce to a series of simultaneous equations which, after taking the appropriate derivatives, generate the expressions obtained from the traditional KB theory approach for solution mixtures (Kang and Smith 2008).

A more convenient form of Equation 4.6 for the analysis of experimental data can be obtained by defining an excess internal energy (ε) for each member of the ensemble such that,

$$\varepsilon = E - \sum_j N_j E_j^\circ \qquad 4.8$$

where E_j° is a convenient reference energy per particle. For completely miscible mixtures it is natural to take this reference energy as the average molar internal energy in the pure liquid j at the temperature and average pressure of interest. However, for solid solutes it may be more convenient to use the internal energy of the solute at infinite dilution in the primary solvent. In either case one can write,

$$d\ln\rho_i = -F_i d\beta + \sum_j \left(\delta_{ij} + N_{ij}\right)(d\beta\mu_j - E_j^\circ d\beta) \qquad 4.9$$

where,

$$F_i = \frac{\langle \delta N_i \delta\varepsilon \rangle}{\langle N_i \rangle} \qquad 4.10$$

Equation 4.9 represents a series of source equations which can be used to obtain expressions for various properties of solution mixtures in terms of number-number, number-energy, and energy-energy correlations characterizing local microscopic regions within the solution of interest.

Taking derivatives with respect to T with p and $\{N\}$ constant one obtains the following expression,

$$RT^2\alpha_p = RT^2\left(\frac{\partial\ln V}{\partial T}\right)_{p,\{N\}} = -F_i + \sum_j \left(\delta_{ij} + N_{ij}\right)(\bar{H}_j - E_j^\circ) \qquad 4.11$$

where $\bar{H}_j$ is the partial molar enthalpy of species j, and α_p is the thermal expansion coefficient. Some rearrangement provides an expression (inversion formula) for each F_i in terms of observable experimental quantities,

$$F_i = pRT\kappa_T - RT^2\alpha_p + \sum_j \left(\delta_{ij} + N_{ij}\right)\bar{U}_j^{Ex} \qquad 4.12$$

and κ_T is the isothermal compressibility provided by Equation 4.9,

63

$$RT\kappa_T = -RT\left(\frac{\partial \ln V}{\partial p}\right)_{T,\{N\}} = \sum_j \left(\delta_{ij} + N_{ij}\right)\bar{V}_j \qquad 4.13$$

with $\bar{V}_j$ indicating the partial molar volume of j, and we have used the fact that $\bar{H}_j = \bar{U}_j + p\bar{V}_j$ and $\bar{U}_j^{\text{Ex}} = \bar{U}_j - E_j^{\circ}$.

The required excess partial molar energies $\bar{U}_j^{\text{Ex}}$ can be obtained from the relevant experimental enthalpy of mixing and density data. The contributions to F_i and $\bar{U}_j^{\text{Ex}}$ from the intramolecular degrees of freedom will cancel in Equation 4.12, even when the intramolecular energy levels corresponding to the mixture are different from those in the reference states.

Before leaving this section we note that the thermal expansion coefficient can also be expressed in terms of all the F's by multiplying Equation 4.12 by the volume fraction ($\phi_i = \rho_i \bar{V}_i$), summing over all i, and then using Equation 4.13 to provide,

$$RT^2\alpha_p = \frac{RT\kappa_T}{V_m}(U_m^{\text{Ex}} + pV_m) - \sum_i \phi_i F_i \qquad 4.14$$

where U_m^{Ex} is the excess molar internal energy of mixing and V_m is the molar volume. The above expression will be used later. It also provides a route to the excess energy of mixing from a single simulation when the thermal expansion is known, although this issue will not be considered here.

The constant pressure heat capacity can be used to provide information on the local energy fluctuations. Given the thermodynamic relationship,

$$C_p = \left(\frac{\partial H}{\partial T}\right)_{p,\{N\}} = \left(\frac{\partial U}{\partial T}\right)_{p,\{N\}} + pV\alpha_p \qquad 4.15$$

and the derivative of the average energy obtained by treating $U = <E>$ as a function of the chemical potentials, volume and β in a similar manner to before, and to the approach of Buff and Brout (Buff and Brout 1955), one finds,

$$RT^2\left(\frac{\partial U_m}{\partial T}\right)_p = RT^2 U_m \alpha_p + \frac{\left\langle (\delta E)^2 \right\rangle}{\langle N \rangle} - \sum_i x_i F_{\mu,i} \bar{H}_i. \qquad 4.16$$

Manipulation of Equation 4.16 indicates that the local energy fluctuations can be written in terms of experimentally obtainable properties as,

$$\Delta_m = \frac{\left\langle (\delta\varepsilon)^2 \right\rangle}{\langle N \rangle} = RT^2 C_{p,m} + \sum_i x_i (F_i - RT^2\alpha_p)(\bar{U}_i^{\text{Ex}} + p\bar{V}_i) \qquad 4.17$$

where the F's are given by Equation 4.12. The value of $\beta^2\Delta_m$ obtained from the above expression includes contributions of ½ from each molecular translational degree of freedom, and additional contributions of ½ from each classical intramolecular potential and kinetic quadratic degree of freedom.

Equations 4.12 and 4.17 contribute to a general theory of local solution behavior in terms of experimentally observable properties. Alternatively, the calculation of the number-number, energy-number, and energy-energy fluctuations, from theory or computer simulation, provides a route to a series of experimental properties. For completion we include the corresponding expression obtained for the constant volume heat capacity previously presented by Buff and Brout (Buff and Brout 1955),

$$\Delta_m = RT^2 C_{V,m} + \sum_i F_i (x_i \bar{U}_i^{Ex} - \phi_i U_m^{Ex}) + \frac{V_m}{RT\kappa_T} \left(\sum_i \phi_i F_i \right)^2 \qquad 4.18$$

The two previous expressions can be shown to satisfy the known thermodynamic relationship, $C_{p,m} - C_{V,m} = TV_m \alpha_p^2 / \kappa_T$ after some algebra and the use of Equation 4.14.

Using Equations 4.12 and 4.17 we have investigated the local fluctuations for a variety of solutions as obtained from the relevant experimental data. These fluctuations can be used to either characterize the properties of solutions, as demonstrated by the traditional KB approach, to help rationalize changes in properties such as the compressibility, thermal expansion and heat capacities, or as target data for the testing and evaluation of force fields developed for molecular simulation.
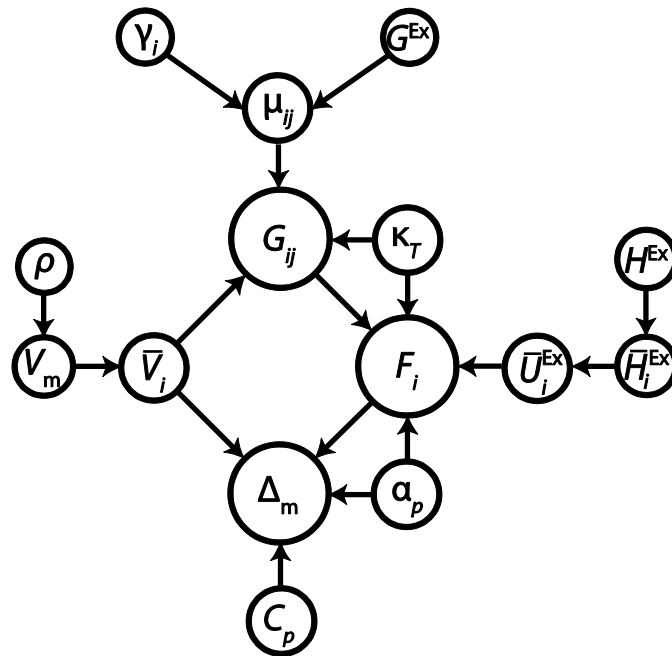

Figure 4.2 The general scheme for obtaining local fluctuating properties from the available experimental data. Here ρ is the solution mass density and the definitions of the other symbols are provided in the main text.

4.3 Methods

Pure water was analyzed at six pressures (1, 200, 400, 600, 800, and 1,000 bar) over the subset of temperatures between 273-1,073 K for which a given p,T point was in the liquid region of the water phase diagram (Wagner and Kretzschmar 2008). Experimental data was analyzed for several n-alcohols (methanol through decanol) (Kiyohara and Benson 1979; Lide 1990; Garcia *et al.* 1996; Segatin and Klofutar 2004; Han *et al.* 2006; Uosaki *et al.* 2006; Zorebski *et al.* 2010), linear alkanes (hexane through tetradecane) (Diazpena *et al.* 1979; Aicart *et al.* 1981; Garcia *et al.* 1988; Lide 1990; Aicart *et al.* 1995; Valencia *et al.* 2009), 1-alkenes (hexene through octene) (Lide 1990; Aicart *et al.* 1995), alkylamines (propy-, butyl-, hexyl-, and octylamine, diethylamine through dibutylamine, and triethylamine through tributylamine) (Oswal *et al.* 2004), and aromatics (benzene and methylbenzene through propylbenzene) (Aicart *et al.* 1995; Fujii *et al.* 1995; Cibulka and Takagi 1999; Watanabe and Kato 2004). For all liquids, excluding pure water, the T and p were fixed at 298 K and 1 bar, respectively. The molar volume of each liquid was calculated from the molecular weights and densities (Lide 1990).

The general outline for the analysis of solution mixtures is provided in Figure 4.2. We have followed a sequential approach whereby the usual KB analysis is performed first to obtain the KBIs, this data is then combined with the thermal expansion and excess partial molar energy data to provide the F's, and finally the heat capacity data is used to obtain Δ_m values. Experimental data for the methanol + water system at a T of 298 K and p of 1 bar was analyzed over the full composition range. The Redlich-Kister power series equation was used to fit the density data (four parameters) (Douheret *et al.* 1989), excess Gibbs energy (G^{Ex}) data (three parameters) (Butler *et al.* 1933; Hu *et al.* 2003), enthalpy of mixing (H^{Ex}) data (four parameters) (Benjamin and Benson 1963), α_p data (six parameters) (Reis *et al.* 2008), and C_p data (four parameters) (Reis *et al.* 2008). κ_T was approximated to display ideal behavior. Data for the benzene + methanol system at 308 K and 1 bar was analyzed over the full composition range. G^{Ex} was fit using the Wilson equation and the enthalpy of mixing was fit using the Redlich-Kister power series equation with six parameters (Wilson 1964; Rajendran *et al.* 1989). All other properties were approximated to be ideal (Gorbunova *et al.* 1982; Hruby *et al.* 1993; Vercher *et al.* 2007; Tripathi *et al.* 2010). Aqueous sodium chloride data at 298 K and 1 bar was studied up to $4m$ NaCl due to a lack of α_p data at higher salt concentrations. The solution V_m was fit to a cubic polynomial (Heyrovska 1996). Experimental activity data was fit to the form $\ln\gamma_{\pm} = -(A\sqrt{m_s})/(1+B\sqrt{m_s}) - \ln(1-Cm_s)$ in which m_s is the salt molality, $\gamma_{\pm}$ is the molal salt activity coefficient, and the other symbols are fitting constants (Robinson and Stokes 1959; Smith 2004a). Polynomials were used to model the enthalpy of mixing (cubic) (Beggerow 1976), α_p (quadratic) (Rogers and Pitzer 1982), C_p (quartic) (Perron *et al.* 1981), and κ_T (quadratic) (Rogers and Pitzer 1982). After a fit to the data for each solution was obtained, the first derivatives of V_m were determined to provide the partial molar volumes for use in the calculation of the KBIs and F's, while second derivatives of G^{Ex} were used for use in the calculation of the KBIs, and first derivatives of the enthalpy of mixing were taken to obtain $\bar{H}_i^{Ex}$. From $\bar{H}_i^{Ex}$, the excess partial molar energies of mixing were calculated according to $\bar{U}_i^{Ex} = \bar{H}_i^{Ex} - p\bar{V}_i^{Ex}$ for use in the calculation of the F's. The solution α_p, C_p, and κ_T were used directly in the KBI, F_i, and Δ_m calculations.

66

4.4 Results

4.4.1 One Component Solutions

The fluctuation quantities for a one component fluid (1) are provided by the following equations where the superscript (o) denotes the pure liquid values,

$$
\begin{aligned}
1 + N_{11}^{o} &= \rho_{1}^{o} RT \kappa_{T}^{o} \\
F_{1}^{o} &= pRT\kappa_{T}^{o} - RT^{2}\alpha_{p}^{o} \\
\Delta_{m}^{o} &= RT^{2}C_{p,m}^{o} + pV_{m}^{o}\left(pRT\kappa_{T}^{o} - 2RT^{2}\alpha_{p}^{o}\right)
\end{aligned}
\qquad 4.19
$$

and are in agreement with previous literature values (Schofield 1966; Adams 1975). We have evaluated the corresponding local fluctuating properties for water over a range of temperatures and pressures. The results are shown in Figure 4.3. As expected the curves peak in the vicinity of the critical temperature (647 K) and critical pressure (220 bar) (Mishima and Stanley 1998). The βF curves are dominated by the thermal expansion term. This also results in all the pressure curves adopting the same value of $\beta F = -0.06$ at ≈ 292 K, in a similar manner to the raw thermal expansion data (Bridgman 1913; Navia *et al.* 2010). The fact that $dF/dp = 0$ at this temperature implies a relationship between the higher local fluctuating moments of the liquid. These moments can be determined (Jiao and Smith 2011), but this is beyond the scope of the present work.

An analysis of the data concerning simple pure organic liquids at 298 K and 1 bar is presented in Figure 4.4. All liquids studied here displayed a systematic decrease in the number fluctuations (compressibility) (Ben-Naim 1977), and a corresponding increase in F and Δ_m values, as the number of carbons is increased. The energy fluctuations display essentially a linear dependence on the number of carbons, which is to be expected from the relationships observed for the heat capacities of hydrocarbons (Hadden 1970; Matulis 2001), and the dominating contribution from $C_{p,m}$ at this temperature and pressure.
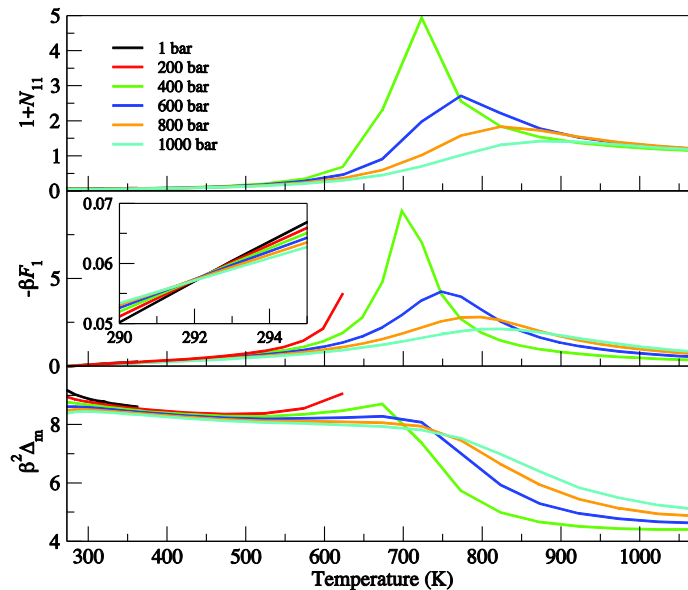
Figure 4.3 Local fluctuating properties for pure liquid water over a range of *T* and *p*.

67
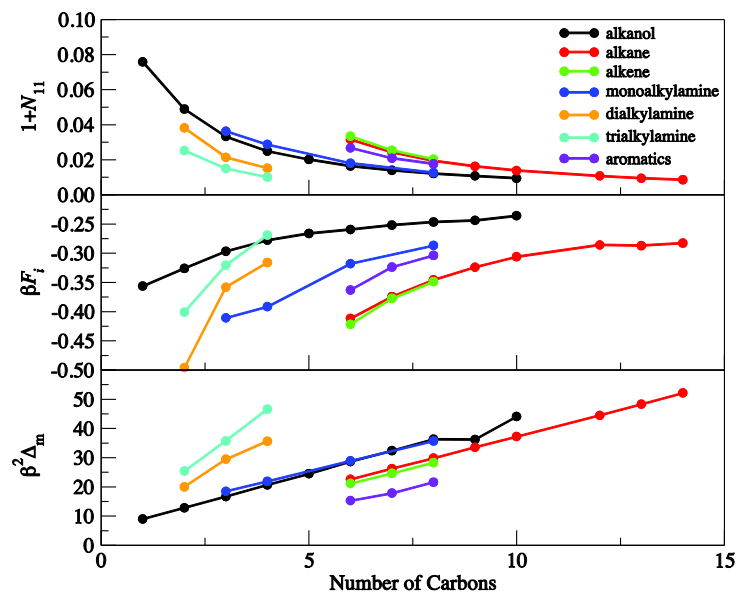
Figure 4.4 Local fluctuating properties of pure liquids at 298K and 1 bar. The number of carbons for the aromatic molecules was taken as the total number of carbons including the ring atoms, while the number of carbons for the di and trialkyl amines was taken as the number of carbons in a single alkyl chain.
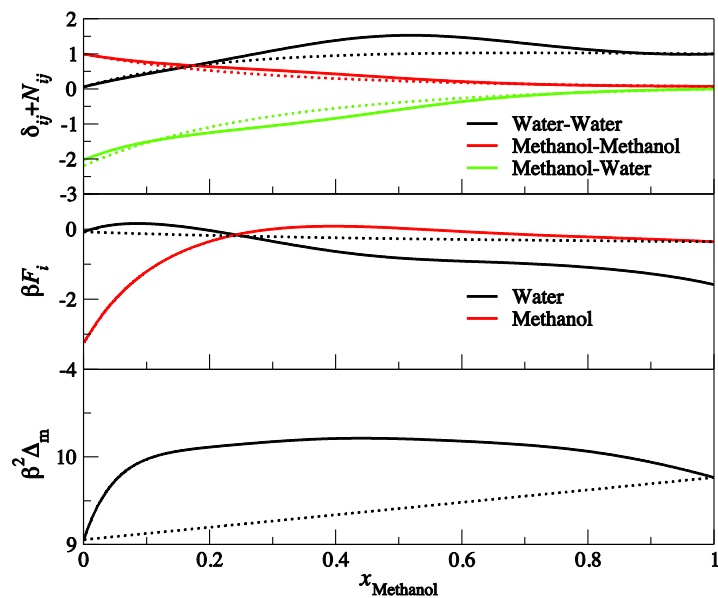
Figure 4.5 A fluctuation analysis of water (1) and methanol (2) mixtures at 298 K and 1 bar. The solid lines correspond to the experimental data, while the dotted lines represent the SI solution values.

4.4.2 Binary Solutions

One natural application of the above theory is to deepen our understanding of the local microstructure of binary mixtures. For a binary mixture of a solvent (1) and solute (2) one finds the following expressions apply,

$$\mu_{22} = \beta\left(\frac{\partial\mu_2}{\partial\ln m_2}\right)_{T,p} = x_1\beta\left(\frac{\partial\mu_2}{\partial\ln x_2}\right)_{T,p}$$

$$1 + N_{11} = \rho_1 RT\kappa_T + \rho_1\frac{\phi_2\overline{V_2}}{\mu_{22}}$$

$$1 + N_{22} = \rho_2 RT\kappa_T + \frac{\phi_1^2}{\mu_{22}} \qquad\qquad 4.20$$

$$N_{12} = \rho_2 RT\kappa_T - \frac{\phi_1\phi_2}{\mu_{22}}$$

$$F_i = \overline{U}_i^{Ex} + pRT\kappa_T - RT^2\alpha_p + N_{i1}\overline{U}_1^{Ex} + N_{i2}\overline{U}_2^{Ex}$$

$$\Delta_m = RT^2 C_{p,m} - RT^2\alpha_p\left(U_m^{Ex} + pV_m\right) + x_1 F_1 u\overline{U}_1^{Ex} + x_2 F_2 u\overline{U}_2^{Ex} + pV_m(\phi_1 F_1 + \phi_2 F_2)$$

where m_i and x_i are the molality and mole fraction of i, respectively. The first four expressions correspond to the traditional KB inversion approach. The N's and F's can be eliminated from the last two expressions if desired, but we retain them here for simplicity. Figures 4.5-4.7 display the results from a fluctuation analysis of three binary solutions. It is known that the KBIs for solutions under ambient conditions are most sensitive to the solution activities, relatively insensitive the partial molar volumes, and essentially unaffected by the precise value of the isothermal compressibility (Matteoli and Lepori 1984). The F values presented here were sensitive to the partial molar enthalpies (energies), with a significant but relatively constant contribution from the thermal expansion coefficient, and an essentially negligible contribution from the compressibility term.

The results from an analysis of methanol and water mixtures at 298 K and 1 bar are presented in Figure 4.5. A KB analysis of this system has already been presented (Matteoli and Lepori 1984; Weerasinghe and Smith 2005). At this temperature and pressure the experimentally extracted F's closely mirror the partial molar excess enthalpies (data not shown). The F's and Δ_m functions display nontrivial variations with mole fraction. Hence, they provide insights into the behavior of solution mixtures which can be used to characterize the solution in terms of local fluctuating properties. However, to obtain a deeper insight into these variations with composition one probably requires a similar analysis of computer simulation data.

In contrast to methanol and water mixtures an analysis of methanol and benzene solutions at 308 K and 1 bar indicates a substantial degree of methanol self association at low methanol compositions. This is illustrated in Figure 4.6. The F's no longer closely mimic the partial molar enthalpies, as was the case for methanol and water, primarily due to the relatively large values of the KBIs (see Equation 4.20). The energy-energy fluctuations are essentially the same as the ideal values as, at these temperatures and pressures, the Δ_m term is dominated by the heat capacity contribution and this was taken to be ideal in nature. The large values of F for methanol at low methanol mole fractions coincides with the increased self association of methanol molecules (increase in N_{22}) albeit at a lower benzene mole fraction for the
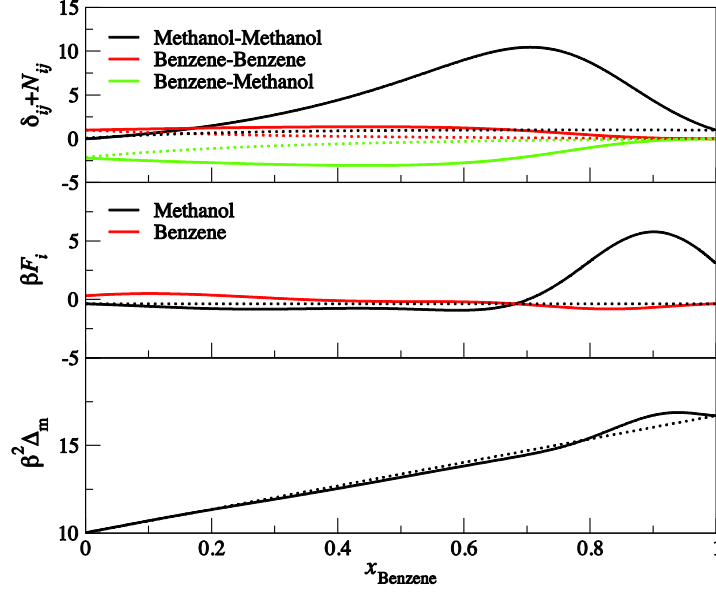
Figure 4.6 A fluctuation analysis of methanol (1) and benzene (2) mixtures at 308 K and 1 bar. The solid lines correspond to the experimental data, while the dotted lines represent the SI solution values.

latter (Ploetz *et al.* 2010a).

Many potential applications involve systems where one component (2) is at infinite dilution (∞), as is the case with many biological problems for example. Under these conditions F_1 is given by Equation 4.19 and F_2 by the expression,

$$F_2^\infty = pRT\kappa_T^o - RT^2\alpha_p^o + \bar{U}_2^{\text{Ex},\infty} \qquad 4.21$$

Hence, there is a direct link between F_2 and the excess partial molar energy at infinite dilution, *i.e.* between a local fluctuating property and a thermodynamic derivative of the solution. The value of F_2 at infinite dilution expressed in terms of fluctuations reduces to,

$$F_2^\infty = \langle\varepsilon\rangle_2 - \langle\varepsilon\rangle_o = \langle E\rangle_2 - \langle N_1\rangle_2 E_1^o - E_2^o \qquad 4.22$$

where the subscripts 2 and o refer to ensemble averages obtained for regions containing the solute molecule, or for regions in the pure solvent, respectively. For classical pairwise additive potentials this can be expressed as,

$$F_2^\infty = \langle E_{22} + E_{21} + E_{11}\rangle_2 - \langle N_1\rangle_2 E_1^o - E_2^o \qquad 4.23$$

where E_{22} is the intramolecular (potential and kinetic) contribution to the energy from the solute. E_{21} is the solute-solvent potential energy, and E_{11} is the solvent-solvent potential and kinetic energy. The E_1^o term represents the total energy that the same number of local solvent molecules would have in bulk solution. Hence, the solute F has been reduced to a series of simple ensemble averages for a region in the vicinity of the solute under these conditions. Furthermore, these averages quantify the energy of the solute and the interaction with the solvent, and how the solute perturbs the solvent-solvent interactions from those observed in bulk solvent, all relative to the solute reference energy.

4.4.3 Salt Solutions

The study of electrolyte solutions by fluctuation theory is slightly complicated by the electroneutrality constraints that are often invoked for these solutions (Friedman and Ramanathan 1970). If we consider a salt $(M_{n_+}^{z_+} X_{n_-}^{z_-})$ which completely dissociates to provide n_+ cations and n_- anions, and therefore n total ions, one can relate fluctuations involving any indistinguishable ion (I) to that of the cations and anions *via* the expressions,

$$G_{II} = \left(\frac{n_+}{n}\right)^2 G_{++} + \left(\frac{n_-}{n}\right)^2 G_{--} + \frac{n_+ n_-}{n^2}(G_{+-} + G_{-+})$$

$$G_{II} = G_{1I} = \frac{n_+}{n}G_{+1} + \frac{n_-}{n}G_{-1} \qquad 4.24$$

$$F_I = \frac{n_+}{n}F_+ + \frac{n_-}{n}F_-$$

which were obtained from a simple substitution of $N_I = N_+ + N_-$ into Equations 4.1 and 4.10 (Gee and Smith 2009). Further imposing the electroneutrality conditions (Friedman and Ramanathan 1970; Kusalik and Patey 1987; Chitra and Smith 2002),

$$z_+ + z_+\rho_+ G_{++} + z_-\rho_- G_{+-} = 0$$
$$z_- + z_+\rho_+ G_{+-} + z_-\rho_- G_{--} = 0$$
$$z_+\rho_+ G_{+1} + z_-\rho_- G_{-1} = 0 \qquad 4.25$$
$$z_+\rho_+ F_+ + z_-\rho_- F_- = 0$$
$$n_+ z_+ + n_- z_- = 0$$

provides the following relationships,

$$G_{II} = -\frac{1}{\rho_I} + G_{+-}$$

$$G_{+-} = \frac{1}{\rho_+} + G_{++}$$

$$\frac{1}{\rho_+} + G_{++} = \frac{1}{\rho_-} + G_{--} \qquad 4.26$$

$$G_{1I} = G_{+1} = G_{-1}$$

$$F_I = F_+ = F_-$$

and indicates that there is only one unique (independent) solute KBI and F value for a salt solution. Hence, in practice it is often easier to treat the solution as a mixture of a solvent and a collection of indistinguishable ions (2 = I) in which case the following relationships should be used,

$$m_2 = nm_s \qquad\qquad \rho_2 = n\rho_s \qquad\qquad n\overline{V_2} = \overline{V_s}$$

$$\mu_{22} = 1 + \left(\frac{\partial \ln\gamma_\pm}{\partial \ln m_s}\right)_{T,p} \qquad\qquad n\overline{U}_2^{Ex} = \overline{U}_s^{Ex} \qquad 4.27$$

in Equation 4.20 to provide the required G_{II}, G_{1I}, and F_I values. Here, the subscript s refers to the traditional salt concentration, volume or partial molar excess energy, and $\gamma_I = \gamma_\pm$ is the mean molal ion activity coefficient.

The results from a fluctuation analysis of NaCl solutions at 298 K and 1 bar are provided in Figure 4.7. We have included the results obtained from two different reference states for the salt: namely, the salt crystal and an infinitely dilute solute. The choice of reference state does not significantly affect the values of F for water or the energy-energy fluctuations, while the salt F value is significantly shifted and displays a slightly different dependence on composition. The trend in the ion F with composition is the same for both reference states and indicates a decrease in the ion-energy correlation with increasing ion concentration. The local energy-energy fluctuation displays a similar decrease with salt concentration.

4.4.4 Ideal Solutions

Ideal solutions represent interesting reference states for comparison to real solutions. Symmetrical ideal (SI) solutions are defined by the fact that $\mu_i = RT \ln x_i$ and $\bar{U}_i^{\mathrm{Ex}} = 0$ for all i species. The local fluctuations exhibited by symmetric ideal solution mixtures involving any number of n_c components are then given by (Ben-Naim 2006; Ploetz *et al.* 2010b),

$$G_{ij}^{\mathrm{SI}} = RT\kappa_T^{\mathrm{SI}} - V_i^{\mathrm{o}} - V_j^{\mathrm{o}} + S_{n_c} \qquad\qquad S_{n_c} = \sum_{k=1}^{n_c} \rho_k (V_k^{\mathrm{o}})^2$$

$$F_i^{\mathrm{SI}} = pRT\kappa_T^{\mathrm{SI}} - RT^2\alpha_p^{\mathrm{SI}}$$

$$\Delta_{\mathrm{m}}^{\mathrm{SI}} = RT^2 C_{p,\mathrm{m}}^{\mathrm{SI}} + pV_{\mathrm{m}}^{\mathrm{SI}}(pRT\kappa_T^{\mathrm{SI}} - 2RT^2\alpha_p^{\mathrm{SI}})$$

$$V_{\mathrm{m}}^{\mathrm{SI}} = \sum_{k=1}^{n_c} x_k V_k^{\mathrm{o}} \qquad\qquad\qquad\qquad\qquad 4.28$$

$$\kappa_T^{\mathrm{SI}} = \sum_{k=1}^{n_c} \phi_k \kappa_{T,k}^{\mathrm{o}}$$

$$\alpha_p^{\mathrm{SI}} = \sum_{k=1}^{n_c} \phi_k \alpha_{p,k}^{\mathrm{o}}$$

$$C_{p,\mathrm{m}}^{\mathrm{SI}} = \sum_{k=1}^{n_c} x_k C_{p,\mathrm{m},k}^{\mathrm{o}}$$

In this case the F values vary with composition, but are the same for each component (as would be expected). The SI results have been included on Figures 4.5 and 4.6 to help indicate deviations from ideal behavior.

4.4.5 Bulk System Fluctuations

It is well known that fluctuations in the properties of bulk systems can be related to various experimental properties. For systems at constant $\{N\}$, pressure and temperature the experimental properties are the compressibility, thermal expansion and constant pressure heat capacity. The appropriate relationships can be obtained directly from the isothermal-isobaric partition function (Allen and Tildesley 1987),

$$\left\langle (\delta V)^2 \right\rangle_{NpT} = VRT\kappa_T$$

$$\left\langle \delta V \delta(E + pV) \right\rangle_{NpT} = VRT^2\alpha_p \qquad\qquad 4.29$$

$$\left\langle \left(\delta\left(E + pV\right)\right)^2 \right\rangle_{NpT} = RT^2 C_p$$
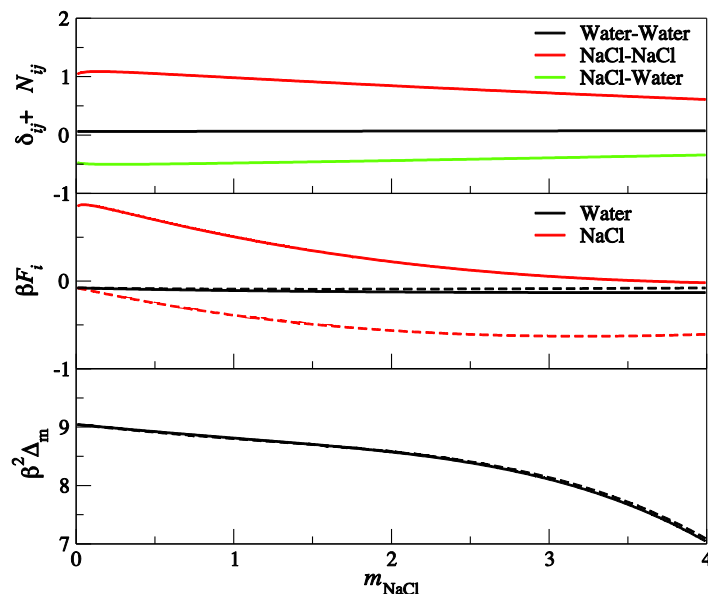
Figure 4.7 A fluctuation analysis of water (1) and sodium chloride (2) mixtures at 298 K and 1 bar. The solid lines correspond to the experimental data using the crystal reference state, while the dashed lines correspond to the experimental data using the infinitely dilute salt reference state.

These relationships are different from the ones presented here as they represent fluctuations in the properties of *bulk closed systems* and not *local open regions* within these systems. It is interesting to note that the local fluctuation relationships (Equations 4.13, 4.14 and 4.17) can be generated from the corresponding bulk expressions by use of the following transformations,

$$\delta V \rightarrow -\sum_i \bar{V}_i \delta N_i \qquad \delta\left(E + pV\right) \rightarrow \delta E - \sum_i \bar{H}_i \delta N_i \qquad 4.30$$

followed by some algebra and a subsequent transformation of E to ε using Equation 4.8.

4.5 Conclusions

We have outlined a method for extracting information concerning local energy-particle and energy-energy fluctuations in solution mixtures from experimental data. The resulting local fluctuations can be used to help characterize the properties of solution mixtures in much the same way as the traditional KB/FT approach. Another possible use of the experimentally derived fluctuations is for the evaluation and development of accurate force fields for molecular simulation, where the fluctuating quantities serve as additional experimental data for the parametrization procedure (Ploetz *et al.* 2010a; Weerasinghe *et al.* 2010). In addition, the same type of approach can be applied to understand the effects of temperature on chemical equilibria in terms of energy-particle and energy-energy fluctuations. Studies along these lines will be published separately (Jiao and Smith 2011; Weerasinghe and Smith 2014).

5. Local Fluctuations in Solution: Theory and Applications[*][†]

Abstract

We derive a general fluctuation theory of solutions that relates fluctuating quantities in the grand canonical ensemble to thermodynamic properties of solutions under isothermal isobaric conditions and apply the theory to a variety of binary systems, which we investigate by analyzing available experimental data and by molecular dynamics simulations. We additionally analyse the local fluctuations in terms of a multivariate Gaussian distribution with examples from computer simulation and present the role that higher moments of the distributions play in understanding the effects of temperature, pressure, and composition on the fluctuating quantities.

5.1 Introduction

Liquids are complicated. Liquid mixtures are even more complicated (Prigogine 1957). While our knowledge of gases and solids is quite extensive, and in many cases can be considered essentially complete, a deep understanding of the properties of solutions remains elusive. The major problems involve our simplistic understanding of how molecules interact, especially at high densities, and how these interactions are affected by changes in the conditions – usually temperature, pressure and composition. Even in the absence of strong intermolecular interactions, our ability to rationalize the packing of molecules in solution remains rather rudimentary. One possible solution to this problem involves the use of molecular simulation (Allen and Tildesley 1987). Computer simulations have now advanced to a stage where they have provided a wealth of information using reasonably accurate models of liquid solutions. The ability to investigate the properties of solution mixtures is clearly aided by computer simulation and research in this direction will continually evolve. However, consistent comparison with experiment is still required in order to validate many of the results. Furthermore, our ability to relate the properties of solution mixtures to the underlying interactions between the molecules is still in its infancy. For instance, when a particular model (or force field) provides an incorrect description of a given thermodynamic property, for example the isothermal compressibility or the enthalpy of mixing, it is often extremely difficult to trace this to a specific incorrectly modeled interaction between molecules. Conversely, it is also difficult to predict the thermodynamic consequences arising from variations in the intermolecular interactions. Both issues limit the effectiveness of computer driven research.

Another, more subtle, problem arises when attempting to analyze experimental data concerning solution mixtures. Experimental data is typically obtained under isothermal isobaric conditions where the system is closed to matter exchange. This is clearly the most convenient and relevant set of conditions for most experiments and real systems. However, it is well known that the thermodynamic constraints placed on a system affect the behavior of the system (Munster 1970). The statistical thermodynamics of large systems ensures that ensemble averages between systems under the same average conditions are equal (Hill 1956). In fact, many theories of solutions have traditionally been developed starting from the grand canonical ensemble where the required manipulations are often much easier to perform. However, this equivalence does not hold for fluctuating quantities. Fluctuations depend on the ensemble. We will see here that fluctuations characteristic of open systems can be used to help understand solution behavior in closed systems. In this type of approach the fluctuations then represent the fluctuations associated with a relatively small local volume of solution. We will argue that these local fluctuations should be much easier to understand and probe compared to fluctuations of the system as a whole. One then simply needs to relate these local fluctuations to thermodynamic properties of the solution under the appropriate thermodynamic constraints. This is the role that the Fluctuation Theory (FT) of solutions seeks to play.

Fluctuations play an important role in the properties of systems. This was well established by the work of Einstein and others (Callen 1960). However, this work has primarily focused on the fluctuations in bulk systems under different thermodynamic constraints. A major step forward was provided by Kirkwood and Buff when they related particle number fluctuations and integrals over molecular distribution functions to the properties of closed systems -

specifically the changes in chemical potentials, partial molar volumes and isothermal compressibility - and to changes in the osmotic pressure for semi-open systems (Kirkwood and Buff 1951). This approach is generally known as the Kirkwood-Buff (KB) theory of solutions. However, the theory lay relatively unused for over 25 years. Presumably, this initial lack of interest was because the theory appeared to be unsuited for the study of salt solutions (Friedman and Ramanathan 1970; Kusalik and Patey 1987), and/or the theory required radial distribution functions, which are generally unknown, as input data.

More recently, KB theory has developed into a powerful tool for probing the microstructure of solution mixtures (Matteoli and Mansoori 1990; Ben-Naim 2006; Pierce *et al.* 2008; Smith *et al.* 2013). This can be attributed to two main factors. First, the development of the KB inversion procedure by Ben-Naim (Ben-Naim 1977), whereby the experimental data can be used to obtain the integrals over the RDFs, the so called KB integrals (KBIs). Second, the ability of modern computer simulations to provide realistic RDFs as input for the theory has improved significantly. These two factors have prompted the development of approaches to investigate the local microstructure in solutions, *i.e.* deviations in the local composition from the bulk distribution, which has in turn led to a rigorous description of preferential solvation (Ben-Naim 2006). Fluctuation Theory has now been widely used to understand the basic properties of solutions (Matteoli and Mansoori 1990), to understand the effects of additives on the solubility of solutes (from small hydrocarbons to proteins) and biomolecular equilibria (O'Connell 1971a; Chialvo *et al.* 1996; Chitra and Smith 2001a; Ruckenstein and Shulgin 2002; Shimizu and Smith 2004; Shulgin and Ruckenstein 2005; Mazo 2006), to investigate the local composition of solutions in the context of preferential solvation (Marcus 1989), to study the effects of additives on the surface tension of liquids (Chen and Smith 2008), to study critical behavior (Cochran *et al.* 1987; Debenedetti 1987a; O'Connell and Liu 1998; Baird and Kim 2003), to interpret computer simulation data (Smith 1999; Chitra and Smith 2001a; Weerasinghe and Smith 2003a; Ganguly *et al.* 2011), to develop models for many of the above effects (Smith 2004b), and to provide expressions for the volume, enthalpy, compressibility and heat capacity corresponding to simple association equilibria (Gee and Smith 2009; Jiao and Smith 2011). The general formulation and application of KB theory has been outlined in detail by several workers (Matteoli and Mansoori 1990; Ben-Naim 2006; Pierce *et al.* 2008; Smith *et al.* 2013).

KB theory is restricted to the use of particle number fluctuations. It was recognized quite early that one could extend this type of approach to include particle-energy and energy-energy fluctuations (Buff and Brout 1955). Further work by Debenedetti has also highlighted the additional insight this provides for computer simulations (Debenedetti 1986, 1987b, 1988). However, a direct link to experimental data was absent until very recently when Ploetz and Smith provided a framework for the analysis of additional solution properties - specifically the excess partial molar enthalpies, the thermal expansion coefficient and the constant pressure heat capacity - all as a function of solution composition (Ploetz and Smith 2011b). Furthermore, all the fluctuating properties can also be obtained directly from available experimental data. The approach has since been extended to treat molecular association and conformational equilibria (Jiao and Smith 2011).

Here, we outline a practical general theory of solution fluctuations, representing local regions within a solution, which relates the local fluctuations to thermodynamic properties of bulk closed isothermal isobaric systems. We argue that this type of approach leads to a much simpler, though certainly not trivial, description of the behavior of solution mixtures. Furthermore, this appears to be the most appropriate and convenient approach for analyzing a variety of computer simulation data.

5.2 Outline, Notation and General Remarks

The outline of this review is as follows. Some basic background material in the areas of thermodynamics (Section 5.3) and statistical thermodynamics (Section 5.4) is presented with an emphasis on the Gibbs (NpT) and grand canonical (μVT) ensembles of the latter. This is followed by the derivation of a general fluctuation theory of solutions (Sections 5.5-5.7) involving any number of components, which relates fluctuating quantities in the grand canonical ensemble to thermodynamic properties of solutions under isothermal isobaric conditions. The theory is essentially complete at the end of Section 5.7, but the specific forms relating to various applications are not always obvious. Thus, this leads to several sections outlining the application to a variety of binary (Section 5.8) systems. This is illustrated by an analysis of experimental data corresponding to representative binary mixtures (Section 5.8). This is followed by an analysis and discussion of the local fluctuations in terms of a multivariate Gaussian distribution with examples from computer simulation (Section 5.9). The role of higher moments of the distributions in understanding the effects of temperature, pressure, and composition on the fluctuating quantities provided by FT is then presented (Section 5.10). Finally, we discuss some technical issues surrounding the implementation of the theory for the analysis of computer simulation data. We conclude with a summary and future outlook.

In presenting the expressions below it quickly becomes apparent that a general matrix formulation is possible in many cases, but this often suffers from a lack of transparency. Hence, in many places we provide expressions involving a combination of local fluctuating and bulk thermodynamic quantities. We believe this provides clarity to the situation and often helps during subsequent manipulations that might be of interest, *e.g.* infinitely dilute conditions.

Throughout this work we refer to a collection of species (1, 2, 3, . . .) in a system of interest. We consider this to represent a primary solvent (1), a solute of interest (2), and a series of additional cosolutes or cosolvents (3, 4, …) which may also be present in the solution. It should be noted that when we refer to solutions we are primarily concerned with liquid mixtures. FT can, however, be applied to mixtures of gases. The summations appearing here involve sums over all thermodynamically independent components (n_c) in the mixture unless stated otherwise. The composition of the solution is denoted by either mole fractions ($x_i = N_i/N$), dimensionless molalities ($m_i = N_i/N_1$), or number densities ($\rho_i = N_i/V$), where N_i and N are the number of molecules of i and the total number of molecules, respectively. The classic definitions of molalities and molarities can be obtained by a simple scaling of the dimensionless molalities and number densities, respectively, but this does not affect the final relationships provided here. Our formulation will focus on the use of $\beta = (k_B T)^{-1}$ rather than the absolute temperature as an independent variable, where k_B is the Boltzmann constant. The former is much more convenient for statistical thermodynamic derivatives involving

temperature. The temperature derivatives can be obtained by simple substitution, $-k_B T^2 \, d\beta = dT$, although second derivatives are more involved.

Derivatives of the chemical potentials with respect to composition form a central component of the theory. Unfortunately, these derivatives can take many forms depending on the concentration scale of interest. This is further complicated if one switches to a particular activity coefficient. Hence, we choose to use chemical potentials and define,

$$\mu_{ij} = \left(\frac{\partial \beta \mu_i}{\partial \ln N_j} \right)_{p,T,\{N\}'} = \left(\frac{\partial \beta \mu_i}{\partial \ln m_j} \right)_{p,T,\{m\}'}$$

5.1

as our primary derivative of interest. Using this definition we also have $N_i \, \mu_{ij} = N_j \, \mu_{ji}$. Our focus in this work will be on fluctuating properties of solutions. A general fluctuation in a property X is written as $\delta X = X - <X>$, where the angular brackets denote an ensemble or time average. All ensemble averages will refer to the grand canonical ensemble unless stated otherwise. We use the symbol E to represent the instantaneous internal energy for each member of the ensemble, or each configuration of a simulation trajectory, whose average value is the thermodynamic internal energy (U). In many applications we are interested in the expressions for an infinitely dilute solute. These are denoted with an infinity superscript, while the pure solvent properties are indicated by a zero superscript. Finally, the Kronecker delta function (δ_{ij}) is used consistently during the theoretical development and is equal to unity when $i = j$, and zero otherwise.

5.3 Thermodynamic Background

In this section we define many of the thermodynamic properties of interest to FT, and also present several useful relationships which aid in the development and application of the theory. The experienced reader may skip this section. Our focus will be on two types of systems. In the first type of system the independent variables are the number of particles ($\{N\}$), the absolute temperature (T), and the pressure (p), while in the second type of system the independent variables are the chemical potentials ($\{\beta\mu\}$), the volume (V), and temperature. Starting with the entropy formulation of Gibbs' fundamental equation and then transforming to the above sets of variables provides (Munster 1970),

$$d\beta G = Hd\beta + \beta Vdp + \sum_i \beta \mu_i dN_i$$
$$d\beta pV = -Ud\beta + \beta pdV + \sum_i N_i d\beta \mu_i$$

5.2

The first set of variables corresponds to the most common situation of experimental interest, while the second set refer to a system that is open to matter exchange with its surroundings. The above equations dictate the response of the system at equilibrium to changes in the independent variables. The enthalpy ($H = U + pV$), volume, and chemical potentials are therefore provided by first derivatives of the Gibbs free energy ($G = H - TS$) for our closed system, while the internal energy, the pressure, and the average number of particles are provided by derivatives of the thermodynamic function (pV) characterizing our open system.

Several global properties of solution mixtures will be relevant to FT. These include the isothermal compressibility (κ_T), the isobaric thermal expansion coefficient (α_p), and the constant pressure molar heat capacity ($C_{p,\mathrm{m}}$) given by the following derivatives,

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_{T,\{N\}} = -\frac{1}{V_\mathrm{m}}\left(\frac{\partial V_\mathrm{m}}{\partial p}\right)_T = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial p}\right)_T$$

$$\alpha_p = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_{p,\{N\}} = \frac{1}{V_\mathrm{m}}\left(\frac{\partial V_\mathrm{m}}{\partial T}\right)_p = -\frac{1}{\rho}\left(\frac{\partial \rho}{\partial T}\right)_P \qquad 5.3$$

$$C_{p,\mathrm{m}} = \left(\frac{\partial H_\mathrm{m}}{\partial T}\right)_p$$

where ρ is the total particle number or mass density and X_m is the corresponding molar quantity, $X_\mathrm{m} = <X>/N$. All three properties can be obtained directly from experiment. They involve second derivatives of the Gibbs free energy. It is the second derivatives of the Gibbs free energy shown above that one can express in terms of second derivatives of the thermodynamic function (pV) for open systems. Second derivatives are typically related to fluctuations in the system as we shall see below.

Chemical potentials are central to our understanding of material and phase equilibrium and phase stability. However, the vast majority of the studies using the FT of solutions involve a single stable phase with multiple components. Here, we are concerned with the relationships between the chemical potentials (and their derivatives) and the solution properties. Thermodynamically, we have from Equation 5.2 that,

$$\mu_i = \bar{G}_i = \left(\frac{\partial G}{\partial N_i}\right)_{p,T,\{N\}'} \qquad 5.4$$

Hence, the chemical potential quantifies the change in the Gibbs free energy on addition of a single particle of species i while keeping p, T, and the number of all other species constant (indicated by the prime superscript). There are additional useful relationships between the partial molar quantities which can be obtained from interchanging the second derivatives of the Gibbs free energy. These are provided in Table 5.1.

Table 5.1 Schematic relationship between second derivatives of the Gibbs free energy.

βG			
	$\beta\mu_i = \partial/\partial N_i$	$V = \partial/\partial p$	$H = \partial/\partial\beta$
$\partial/\partial N_j$	μ_{ij}	$\bar{V}_j$	$\bar{H}_j$
$\partial/\partial p$	$\bar{V}_i$	κ_T	α_p
$\partial/\partial\beta$	$\bar{H}_i$	α_p	C_p

The application of Euler's Theorem to the extensive properties ($X = U$, S, V, H, G and C_p) of a system at constant T and p indicates that (Davidson 1962),

$$X = \sum_i N_i \left(\frac{\partial X}{\partial N_i} \right)_{p,T,\{N\}'} = \sum_i N_i \bar{X}_i \qquad\qquad X_m = \sum_i x_i \bar{X}_i \qquad\qquad 5.5$$

where $\bar{X}_i$ is known as a partial molar quantity. All the extensive quantities depend on T, p, and composition and hence one can write differentials of each X quantity which, when subtracting the differential obtained from Equation 5.5, provides a series of relationships between partial molar quantities. The most useful being when $X = G$,

$$-\beta V dp - H d\beta + \sum_i N_i d\beta \mu_i = 0 \qquad\qquad 5.6$$

which is known as the Gibbs-Duhem (GD) equation. At constant T and p the above expressions indicate that changes in the partial molar quantities are not independent.

In practice, many of the extensive functions (U, S, H, G), and their corresponding partial molar quantities, can only be determined up to an additive constant. Hence, their values are expressed with respect to a set of well-defined reference states. These are typically taken as the pure solutions of each component at the same T and p. One can then define a series of mixing quantities such that,

$$\Delta_{mix} X_m = X_m - \sum_i x_i X_m^o = \sum_i x_i (\bar{X}_i - X_m^o) \qquad\qquad 5.7$$

This concept is also often applied to the volume of the solution even though absolute molar volumes are clearly available. A series of excess quantities may then be defined using the properties of ideal solutions such that,

$$X_m^{Ex} = \Delta_{mix} X_m - \Delta_{mix} X_m^{id} = \sum_i x_i \bar{X}_i^{Ex} \qquad\qquad 5.8$$

where the excess mixing quantities are expressed in terms of the corresponding excess partial molar quantities. It is these excess quantities ($X = G$, V, H, S) that are normally available experimentally and are typically represented as a function of composition by a suitable fitting function. First derivatives of the fitting function with respect to composition then provide the corresponding excess partial molar volumes, enthalpies, and chemical potentials, while second derivatives of the excess Gibbs energy with respect to composition then provide derivatives of the chemical potentials.

In many cases it is necessary to transform the derivatives of the chemical potentials between different concentration units. This can be achieved by use of the following relationships,

$$
\begin{aligned}
&\left(\frac{\partial \ln \rho_i}{\partial \ln m_j} \right)_{p,T} = \delta_{ij} - \phi_j
&&\left(\frac{\partial \ln x_i}{\partial \ln m_j} \right)_{p,T} = \delta_{ij} - x_j
&&\left(\frac{\partial \ln \rho_i}{\partial \ln x_j} \right)_{p,T} = \frac{\delta_{ij} - \phi_j}{1 - x_j} \\[2mm]
&\left(\frac{\partial \ln \rho_i}{\partial \ln \rho_j} \right)_{p,T} = \frac{\delta_{ij} - \phi_j}{1 - \phi_j}
&&\left(\frac{\partial \ln x_i}{\partial \ln x_j} \right)_{p,T} = \frac{\delta_{ij} - x_j}{1 - x_j}
&&\left(\frac{\partial \ln m_i}{\partial \ln m_j} \right)_{p,T} = \delta_{ij}
\end{aligned}
\qquad 5.9
$$

which apply to solutions with any number of components and where $\phi_i = \rho_i \bar{V}_i$ are the volume fractions.
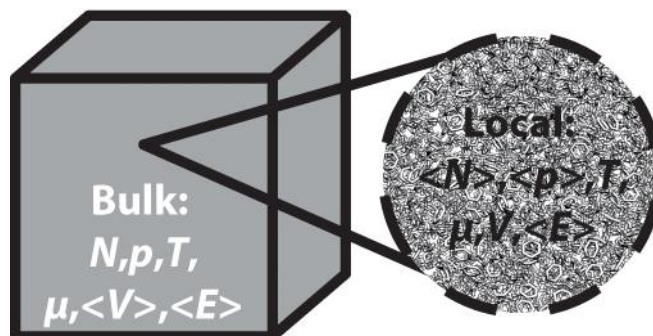
Figure 5.1 A schematic illustration of the local versus bulk representations of solution behavior. Both the local and bulk regions are characterized by the same average pressure and temperature. However, in the bulk region the number of particles is fixed and the chemical potentials are obtained as ensemble averages. In the relatively small local region molecules are free to enter or leave in an effort to maintain a set of constant chemical potentials, determined by the bulk composition, which acts as a matter bath.

The above expressions are all that we require for the manipulation of the available experimental data. Most of the relationships correspond to closed systems which are of primary interest to the vast majority of researchers. FT will be used to provide expressions for many of these closed system quantities in terms of fluctuations in an equivalent open system. This open system can then be considered as a local region of fixed volume within the bulk system that possesses the same average pressure and temperature, but allows matter exchange in order to maintain the same chemical potentials. This is illustrated in Figure 5.1. The spatial extent of the local region will depend on the nature of the components, the solution composition, and the particular values of temperature and pressure concerned.

5.4 Statistical Thermodynamics Background

Statistical thermodynamics is used to provide the characteristic functions in terms of the appropriate partition functions, which will formally involve a sum over the microscopic states available to the system. It is also used to provide relevant expressions for the fluctuations under one set of variables, which can then be used to rationalize the thermodynamic properties of a system characterized by a different set of variables (Hill 1956). Again, the experienced reader may skip this section.

Our application of statistical thermodynamics will focus on the grand canonical ensemble, where we shall develop expressions for the local fluctuations, and the isothermal isobaric (or Gibbs) ensemble, which is representative of most experimental conditions. The characteristic functions provided in Equation 5.2 can be expressed in terms of the corresponding classical multicomponent partition functions such that (Hill 1956),

$$\beta G = -\ln \Delta \left(\{ N \}, p, \beta \right) \qquad \qquad \beta pV = \ln \Xi \left(\{ \beta \mu \}, V, \beta \right)$$

$$\Delta \left(\{ N \}, p, \beta \right) = V_o^{-1} \int_0^\infty Q \left(\{ N \}, V, \beta \right) e^{-\beta pV} dV$$

$$\Xi \left(\{ \beta \mu \}, V, \beta \right) = \sum_{\{ N \} = 0}^\infty Q \left(\{ N \}, V, \beta \right) e^{\beta \mu \cdot N} \qquad \qquad 5.10$$

$$Q \left(\{ N \}, V, \beta \right) = \prod_i \frac{1}{N_i !} \left(\frac{q_i^{\text{int}}}{h^3} \right)^{N_i} \int\!\!\int_{-\infty}^{\infty} e^{-\beta E} \{ dr \} \{ dp \}$$

where E is the internal energy for a particular set of molecular positions $\{r\}$ and momenta $\{p\}$, q_i^{int} is the internal molecular partition function, h is the Planck constant, and we have used the shorthand notation $\mu \cdot N = \mu_1 N_1 + \mu_2 N_2 \cdots$ for simplicity. It should be noted that the focus on classical systems does not imply that FT is invalid for quantum systems, but is merely an acknowledgement that these are the most common systems of interest.

Using these expressions in Equation 5.3 and evaluating the derivatives leads to expressions for the compressibility, thermal expansion, and heat capacity in the Gibbs ensemble. The results are,

$$k_B T \kappa_T = \frac{\langle \delta V \delta V \rangle}{\langle V \rangle}$$

$$k_B T^2 \alpha_p = \frac{\langle \delta H \delta V \rangle}{\langle V \rangle} \qquad \qquad 5.11$$

$$k_B T^2 C_{p,\text{m}} = \frac{\langle \delta H \delta H \rangle}{N}$$

where all the ensemble averages correspond to the *NpT* ensemble. It is important to note that the above formulas represent fluctuations in the properties of the whole system, *i.e.* bulk fluctuations. Expressions for the chemical potentials, partial molar enthalpies and volumes, and also derivatives of the chemical potentials can (in principle) also be provided in terms of properties of the whole system. They are valuable expressions, but they provide little or no information concerning fluctuations in the local vicinity of atoms or molecules. It is these local quantities that will prove more useful.

If we turn to the grand canonical ensemble, one finds the expressions provided in Table 5.2. Several of the second derivatives involve particle-particle, particle-energy, or energy-energy fluctuations, and will play a central role in the FT of solutions. We emphasize that the expressions in Table 5.2 can be used to describe an ensemble of systems in the grand canonical ensemble, in which case they correspond to bulk properties, or to small regions of any bulk system in any ensemble at constant T, in which case they represent local fluctuations in a much larger bulk system.

Finally, before leaving this section we briefly discuss the use of the pseudo chemical potential and associated quantities developed by Ben-Naim (Ben-Naim 2006). The pseudo chemical potential (pcp) formulation offers certain advantages over other traditional approaches to understanding the properties of solutions. The reader is referred to the detailed discussions of Ben-Naim for further information (Ben-Naim 2006). The pseudo chemical potential μ_i^* of species i is

Table 5.2 First and second derivatives in the grand canonical ensemble.

	$\beta p V$		
	$<N_i> = \partial/\partial\beta\mu_i$	$-<E> = \partial/\partial\beta$	$\beta p = \partial/\partial V$
$\partial/\partial\beta\mu_j$	$<\delta N_i\delta N_j>$	$-<\delta E\delta N_j>$	$<N_j>/V$
$\partial/\partial\beta$	$-<\delta E\delta N_i>$	$<\delta E\delta E>$	$-<E>/V$
$\partial/\partial V$	$<N_i>/V$	$-<E>/V$	0

defined by,

$$\beta\mu_i = \beta\mu_i^* + \ln\Lambda_i^3\rho_i \qquad 5.12$$

where Λ is the thermal de Broglie wavelength. The major advantage of this definition is that one separates the total chemical potential into contributions from the translational degrees of freedom (the ln term) and contributions from specific interactions with the other molecules in the system (the pseudo chemical potential term). Using the above definition and taking the appropriate derivatives one also finds,

$$\bar{V}_i = V_i^* + k_B T\kappa_T$$
$$\bar{H}_i = H_i^* + k_B T^2\alpha_p + \frac{3}{2}k_B T \qquad 5.13$$
$$\mu_{ij} = \mu_{ij}^* + \delta_{ij} - \phi_j$$

which provide the pseudo volume and enthalpy, and derivatives of the pseudo chemical potential. Derivatives of the pcp are most closely related to changes in the molar activity coefficient. Standard thermodynamic transformations then provide relationships to Lewis-Randall (mole fraction) activity coefficients and the rational (molal) activity coefficients. The pcp is also closely related to the excess chemical potential often determined from computer simulation using, for example, the Widom particle insertion approach (Allen and Tildesley 1987). Here, we find,

$$\beta\mu_i^* = \beta\mu_i^{Ex} - \ln q_i^{int} \qquad 5.14$$

In many applications the changes to the internal partition function are neglected, in which case changes to the pcp are equal to the changes in the excess chemical potential.

5.5 A General Fluctuation Theory of Solutions

5.5.1 Theory

Many derivations of the principle expressions involving only particle-particle fluctuations have appeared in the literature (Kirkwood and Buff 1951; Hall 1971; O'Connell 1971b; Newman 1994; Ben-Naim 2006; Nichols *et al.* 2009; Blanco *et al.* 2011). They all start with expressions valid for the grand canonical ensemble, followed by a series of transformations to provide expressions for properties representative of the Gibbs ensemble. Here, we will present a relatively simple derivation that attempts to reduce many of the thermodynamic manipulations appearing in other approaches. The general approach for multicomponent systems will involve the manipulation of a few matrices. However, for binary and ternary systems the expressions can be obtained directly from a few simultaneous equations. This offers greater transparency for those less familiar with matrix operations and their properties.

The average number of particles of species i and the average internal energy can be considered a function of the temperature, volume and set of chemical potentials. Using the expressions provided in Table 5.2 the following set of differentials can be written,

$$d\langle X \rangle = -\langle \delta X \delta E \rangle d\beta + \frac{\langle X \rangle}{V} dV + \sum_j \langle \delta X \delta N_j \rangle d\beta \mu_j \qquad 5.15$$

where $X = N_i$ or E. In the traditional formulation of KB theory, T remains constant (Kirkwood and Buff 1951). Hence, the $d\beta$ term disappears and one has a series of simultaneous differentials, one for each component, which can be solved after taking the appropriate derivatives. Here, we wish to include temperature related effects and therefore our expressions will involve energy fluctuations. We now define a set of fluctuating quantities and, in order to make these more compatible with experimentally available data, we choose a reference energy equal to the enthalpy of the pure components such that,

$$\delta_{ij} + N_{ij} = \frac{\langle \delta N_i \delta N_j \rangle}{\langle N_i \rangle} \qquad F_i = \frac{\langle \delta N_i \delta \varepsilon \rangle}{\langle N_i \rangle} \qquad \Delta_m = \frac{\langle \delta \varepsilon \delta \varepsilon \rangle}{\langle N \rangle} \qquad \varepsilon = E - \sum_j N_j H_j^\circ \qquad 5.16$$

where ε corresponds to an excess energy for each member of the ensemble. We note that the above definitions are slightly different from the ones used in the original study (Ploetz and Smith 2011b). However, the new definitions lead to a significant simplification of the resulting expressions. With these steps our differentials can be written,

$$d \ln \rho_i = -F_i d\beta + \sum_j (\delta_{ij} + N_{ij})(d\beta \mu_j - H_j^\circ d\beta)$$

$$d\langle H \rangle = -\left[\langle \delta E \delta \varepsilon \rangle + k_B T H^{Ex} \right] d\beta + \frac{\langle H \rangle}{V} dV + \sum_j \left[\langle \delta E \delta N_j \rangle + \langle N_j \rangle \right] (d\beta \mu_j - H_j^\circ d\beta) \qquad 5.17$$

The last expression for the enthalpy was obtained by a combination of Equations 5.2 and 5.15. We note that the following relationships follow directly from the above expressions,

$$k_B T \kappa_T = \sum_j (\delta_{ij} + N_{ij}) \bar{V}_j$$

$$\delta_{ik} - \rho_k \bar{V}_k = \sum_j (\delta_{ij} + N_{ij}) \mu_{jk}$$

$$k_B T^2 \alpha_p = -F_i + \sum_j (\delta_{ij} + N_{ij}) \bar{H}_j^{Ex} = \rho k_B T \kappa_T H_m^{Ex} - \sum_j \phi_i F_i \qquad 5.18$$

$$k_B T^2 C_{p,m} = \Delta_m + k_B T^2 \alpha_p H_m^{Ex} - \sum_j x_j F_j \bar{H}_j^{Ex}$$

where $\rho = N/V$ is the total number density. Equations 5.17 and 5.18, coupled with the Gibbs-Duhem equation, are all that is required to derive the basic theory and provide a series of relationships between fluctuating quantities and thermodynamic properties. All the properties on the right hand side of the above expressions involve fluctuations or properties that can also be expressed in terms of fluctuations.

A general formulation for the chemical potential derivatives and partial molar volumes is most easy if one first considers the constant temperature case and transforms to molalities. Using component 1 as the primary solvent and noting that $d \ln m_i = d \ln \rho_i - d \ln \rho_1$ provides,

$$d \ln m_i = -\left(F_i - F_1\right)d\beta + \sum_j [\delta_{ij} + N_{ij} - (\delta_{1j} + N_{1j})](d\beta\mu_j - H_j^\circ d\beta) \tag{5.19}$$

Using this expression, the GD relation at constant T and p, and the expressions provided in Equation 5.18, it is a simple process to obtain relationships for μ_{22}, $\overline{V}_2$, and κ_T using a few simple steps. However, for ternary or higher solutions a series of matrices becomes more convenient. We write the number fluctuations in the above expression in matrix form where we also include the number densities (GD expression) in the first row,

$$\begin{pmatrix} \rho_1 & \rho_2 & \cdots & \rho_n \\ -(1+N_{11}-N_{21}) & 1+N_{22}-N_{12} & \cdots & N_{2n}-N_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ -(1+N_{11}-N_{n1}) & N_{n2}-N_{12} & \cdots & 1+N_{nn}-N_{1n} \end{pmatrix} \tag{5.20}$$

This will prove to be a central matrix for the development of FT. Therefore, if we signify the above matrix as M, then take derivatives of Equation 5.19 with respect to molality at constant p (and T), and also take derivatives with respect to p with composition (and T) constant, the results can be expressed in the form,

$$\begin{pmatrix} \overline{V}_1 & \mu_{12} & \cdots & \mu_{1n} \\ \overline{V}_2 & \mu_{22} & \cdots & \mu_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{V}_n & \mu_{n2} & \cdots & \mu_{nn} \end{pmatrix} = M^{-1} \qquad \overline{V}_i = \frac{M^{i1}}{|M|} \qquad \mu_{ij>1} = \frac{M^{ij}}{|M|} \tag{5.21}$$

i.e. the expressions or values for the partial molar volumes and activity derivatives are just the elements of the inverse of the M matrix, which can also be expressed in terms of cofactors of the M matrix. This also leads to an expression for the compressibility,

$$k_B T \kappa_T = \frac{|NN|}{|M|} \tag{5.22}$$

where NN is the matrix of $\delta_{ij} + N_{ij}$ elements. At this point we have a matrix formulation for the chemical potential derivatives, partial molar volumes, and isothermal compressibility. This is essentially the theory of Kirkwood and Buff.

To complete the general theory we require expressions for the partial molar excess enthalpies. The excess enthalpies can be obtained from the expressions provided above (Equation 5.18), or in matrix form as $H^{Ex} = M^{-1} F'$, where F' is a vector with elements $\delta_{1i} H_m^{Ex} + F_i - F_1$. A general expression for the enthalpies then becomes,

$$\overline{H}_i^{Ex} = \rho \overline{V}_i H_m^{Ex} + \sum_j F_j \mu_{ij} \tag{5.23}$$

after the use of Equation 5.21 and some rearrangement *via* the GD expression. The fluctuation expressions for the partial molar volumes and chemical potential derivatives could be inserted into the above equation. However, the expression becomes rather complicated. We prefer the above form as the limiting expressions at low concentrations
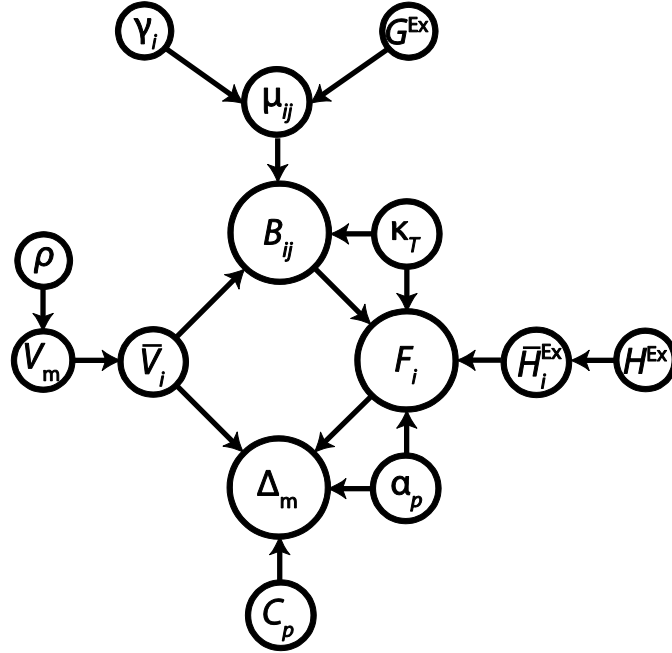
Figure 5.2 A general scheme for obtaining local fluctuating quantities from bulk thermodynamic properties. See text for the definition of the symbols. All properties are composition dependent.

are easily obtained. The fluctuation expressions for the thermal expansion and heat capacity then follow from Equation 5.18.

In this section we have related a series of properties for solution mixtures to fluctuations in local regions of the solution. The expressions are somewhat complicated, especially as the number of components increases. However, this is primarily due to closure of the bulk system to matter exchange. The corresponding fluctuations for local regions are much simpler in form and do not change with the number of components – there are simply more fluctuating quantities. This suggests that a focus on the local properties may prove more beneficial to our understanding of solution mixtures than a focus on bulk properties.

5.5.2 Inversion of FT

The KB inversion procedure is the process of obtaining expressions for the particle number fluctuations in terms of experimentally available (isothermal isobaric) data (Ben-Naim 1977). Again, there are multiple approaches to the inversion procedure (Ben-Naim 1977; Smith 2008). Analysis of Equation 5.18 indicates that an inversion to obtain particle-energy and energy-energy fluctuations is rather trivial once one has the particle-particle formulas. Arguably, the simplest approach to obtain these involves the pseudo chemical potential and partial molar volumes. First, we note that one can add the expressions for $\delta_{ik} - \rho_k \bar{V}_k$ and $\rho_k k_{\mathrm{B}} T \kappa_T$, provided in Equation 5.18, and then convert to pcp derivatives using Equation 5.13 to give,

$$\delta_{ik} - \rho_i V_k^* = \sum_j (\delta_{ij} + N_{ji})(\delta_{kj} + \mu_{kj}^*) \qquad 5.24$$

This can then be expressed in matrix form,

$$
\begin{pmatrix}
1+N_{11} & \cdots & N_{n1} \\
\vdots & \ddots & \vdots \\
N_{1n} & \cdots & 1+N_{nn}
\end{pmatrix}
=
\begin{pmatrix}
1-\rho_1 V_1^* & \cdots & -\rho_1 V_n^* \\
\vdots & \ddots & \vdots \\
-\rho_n V_1^* & \cdots & 1-\rho_n V_n^*
\end{pmatrix}
\begin{pmatrix}
1+\mu_{11}^* & \cdots & \mu_{n1}^* \\
\vdots & \ddots & \vdots \\
\mu_{1n}^* & \cdots & 1+\mu_{nn}^*
\end{pmatrix}^{-1}
$$

5.25

to provide a simple expression for the elements of the transpose of the *NN* matrix (and thereby the KBIs) in terms of a matrix containing just the pseudo partial molar volumes and the inverse of a matrix containing just the pseudo chemical potentials, both of which can be expressed in terms of experimental data. Other approaches are also available (Smith 2008). The particle-energy and energy-energy fluctuations are then easily obtained from Equation 5.18. The overall inversion process is outlined in Figure 5.2.

In summary, FT provides a link between fluctuating properties in the grand canonical ensemble and thermodynamic properties of systems in the Gibbs ensemble. This is illustrated schematically as,

$$
\left\{ \left\langle \delta N_i \delta N_j \right\rangle \right\}, \left\{ \left\langle \delta N_i \delta \varepsilon \right\rangle \right\}, \left\langle \delta \varepsilon \delta \varepsilon \right\rangle \qquad \leftrightarrow \qquad \left\{ \mu_{ij} \right\}, \left\{ \bar{V}_i \right\}, \left\{ \bar{H}_i^{Ex} \right\}, \kappa_T, \alpha_p, C_{p,m}
$$

For an n_c component solution there are $n_c(n_c + 1)/2$ unique particle number fluctuations, n_c unique particle-excess energy fluctuations and one excess energy-energy fluctuation. These can be used to provide, or can be extracted from, $n_c(n_c - 1)/2$ independent chemical potential derivatives, $n_c - 1$ independent partial molar volumes, $n_c - 1$ independent excess partial molar enthalpies, together with the compressibility, thermal expansion and heat capacity. Alternatively speaking, $(n_c^2 + 3n_c)/2 + 1$ fluctuating quantities are related to the same number of independent thermodynamic properties. All of these properties vary with composition.

5.6 Fluctuations in Terms of Molecular Distribution Functions

The previous expressions involve particle number (and energy) fluctuations. It is more common, and totally equivalent, to use distribution functions to replace the number fluctuations. However, in doing so one has to remember that these distributions correspond to a system volume that is open to all species. Consequently, integration over the system volume does not lead to the trivial results observed for closed systems (Ben-Naim 2006). In the grand canonical ensemble the probability that any N_1 molecules of species 1, and N_2 molecules of species 2, and so on, are within $d\{r\}$ at $\{r\}$ is given by $\rho^{(n)}(\{r\})\{dr\}$ where (Hill 1956),

$$
\int \rho^{(n)}(\{r\})\{dr\} = \left\langle \prod_i \frac{N_i!}{(N_i - n_i)!} \right\rangle
$$

5.26

the product is over the different species in the mixture, and n_i is the number of i molecules in the n particle probability distribution. Hence, we can evaluate the following integrals,

$$
\int \rho_i^{(1)}(r_1) dr_1 = \left\langle N_i \right\rangle
$$

$$
\iint \rho_{ij}^{(2)}(r_1, r_2) dr_1 dr_2 = \left\langle N_i N_j \right\rangle - \delta_{ij} \left\langle N_i \right\rangle
$$

$$
\iiint \rho_{ijk}^{(3)}(r_1, r_2, r_3) dr_1 dr_2 dr_3 = \left\langle N_i N_j N_k \right\rangle - \delta_{ik} \left\langle N_i N_j \right\rangle - \delta_{ij} \left\langle N_i N_k \right\rangle - \delta_{jk} \left\langle N_i N_j \right\rangle + 2\delta_{ij}\delta_{jk} \left\langle N_i \right\rangle
$$

5.27

where the integrals are over all possible positions of the particles. Various combinations of the above integrals are generally encountered in statistical thermodynamic theories of solutions (Hill 1956; Davidson 1962). The two most common integrals are,

$$\rho_i \rho_j V G_{ij} = \iint [\rho_{ij}^{(2)}(r_1, r_2) - \rho_i^{(1)}(r_1) \rho_j^{(1)}(r_2)] dr_1 dr_2$$

$$\rho_i \rho_j \rho_k V G_{ijk}$$
$$= \iiint \left[\rho_{ijk}^{(3)}(r_1, r_2, r_3) - \rho_{ij}^{(2)}(r_1, r_2) \rho_k^{(1)}(r_3) - \rho_{ik}^{(2)}(r_1, r_3) \rho_j^{(1)}(r_2) - \rho_{jk}^{(2)}(r_2, r_3) \rho_i^{(1)}(r_1) + 2\rho_i^{(1)}(r_1) \rho_j^{(1)}(r_2) \rho_k^{(1)}(r_3) \right] dr_1 dr_2 dr_3$$

$$5.28$$

The required integrals expressed as fluctuations are therefore given by,

$$\rho_i \rho_j V G_{ij} = \langle \delta N_i \delta N_j \rangle - \delta_{ij} \langle N_i \rangle$$

$$\rho_i \rho_j \rho_k V G_{ijk} = \langle \delta N_i \delta N_j \delta N_k \rangle - \delta_{ij} \langle \delta N_i \delta N_k \rangle - \delta_{ik} \langle \delta N_i \delta N_j \rangle - \delta_{jk} \langle \delta N_i \delta N_j \rangle + 2\delta_{ij} \delta_{jk} \langle N_i \rangle$$

$$5.29$$

These fluctuations can then be expressed in terms of a set of distribution functions $g^{(n)}$ for species i, j and k defined with reference to the bulk distributions by,

$$g_i^{(1)}(r_1) = \frac{\rho_i^{(1)}(r_1)}{\rho_i} = 1 \qquad g_{ij}^{(2)}(r_1, r_2) = \frac{\rho_{ij}^{(2)}(r_1, r_2)}{\rho_i \rho_j} \qquad g_{ijk}^{(3)}(r_1, r_2, r_3) = \frac{\rho_{ijk}^{(3)}(r_1, r_2, r_3)}{\rho_i \rho_j \rho_k} \qquad 5.30$$

The integrals can then be related to the distribution functions where we have integrated over the position of the central particle and only consider the inter-particle distance $r_{12} = |r_2 - r_1|$,

$$G_{ij} = G_{ji} = 4\pi \int_0^\infty [g_{ij}^{(2)}(r) - 1] r^2 dr$$

$$G_{ijk} = \iint [g_{ijk}^{(3)}(r_{12}, r_{13}) - g_{ij}^{(2)}(r_{12}) - g_{ik}^{(2)}(r_{13}) - g_{jk}^{(2)}(r_{23}) + 2] dr_{12} dr_{13}$$

$$5.31$$

At this point we have related the particle number fluctuations to integrals over radial distribution functions in the grand canonical ensemble. FT does not require information on the angular distributions for pairs of molecules as these are averaged out in the above expressions. The RDFs correspond to distributions obtained in a solution at the composition of interest, after averaging over all the remaining molecular degrees of freedom. The G_{ij}'s are known as Kirkwood-Buff integrals (KBIs) and are the central components of KB theory.

The KBIs quantify deviations from a random distribution surrounding a central molecule summed over all space. In this respect they are more informative than the particle fluctuation formulation as they can then be decomposed and interpreted in terms of spatial contributions – using computer simulation data, for example. They clearly resemble the integrals one encounters in the study of imperfect gases and the McMillan-Mayer theory of solutions (McMillan and Mayer 1945). However, the KBIs do not form part of a virial expansion and are relevant at any solution composition – not just infinitely dilute solutes in the gas phase or a primary solvent.
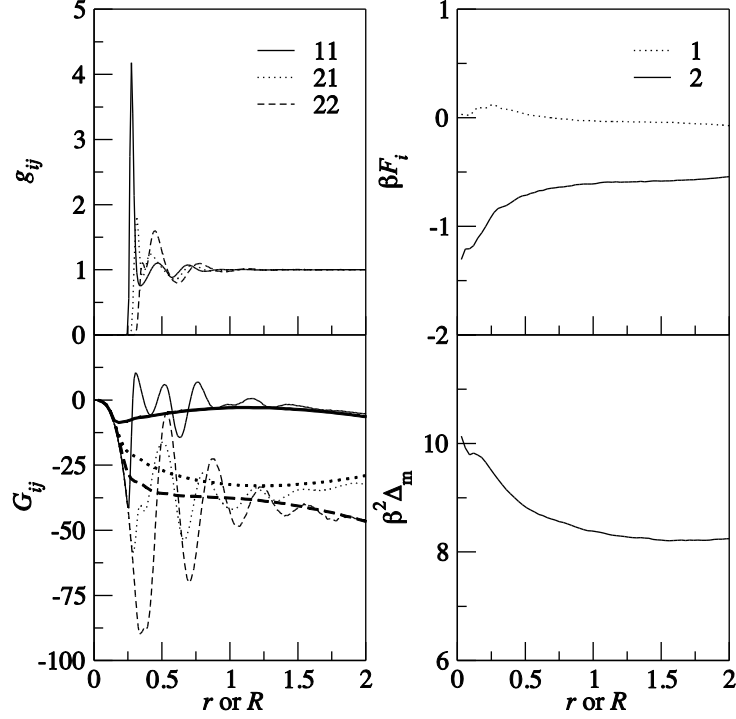
Figure 5.3 Radial distribution functions, Kirkwood-Buff integrals, particle-excess energy and excess energy-energy fluctuations obtained from a simulation of methanol (2) and water (1) mixtures ($x_2 = 0.25$) at 300 K and 1 bar. The properties are displayed as a function of distance (r) and integration distance or sphere radius (R) in nm. The KBIs obtained from integration of the RDFs correspond to the oscillating (thin) curves in the bottom left panel, while those from the particle-particle fluctuations correspond to the smooth (thick) lines.

Multiplying the KBIs by number densities (concentrations) one obtains an alternative picture of the KBIs. The quantities, $N_{ij} = \rho_j G_{ij}$, are referred to as excess coordination numbers and quantify the change in the number of j molecules observed in an open volume of solution on introduction of a central i molecule, from that observed for the same volume of bulk solution in the absence of the i molecule (Hall 1971). The integrals can be related to the fluctuation densities *via*,

$$B_{ij} = \rho_i \left(\delta_{ij} + N_{ij} \right) = \frac{\langle \delta N_i \delta N_j \rangle}{V}$$

$$\rho_i \left(\delta_{ij} \delta_{ik} + \delta_{jk} N_{ij} + \delta_{ij} N_{ik} + \delta_{ik} N_{kj} + \rho_j \rho_k G_{ijk} \right) = \frac{\langle \delta N_i \delta N_j \delta N_k \rangle}{V}$$

5.32

The triplet correlations will become important when discussing third derivatives of the thermodynamic potentials that arise when we consider the variations in the KBIs and chemical equilibria with temperature, pressure and composition. Hence, there are two complimentary viewpoints. The first is provided by integrals over molecular distribution functions and quantifies any deviation from a random distribution (Equation 5.31). The second is provided by particle number fluctuations and describes the degree of correlation between the various particle numbers (Equation 5.32).

An example of the fluctuating quantities obtained from computer simulation is provided in Figure 5.3. The results are specifically for a mixture of methanol and water at 300 K and 1 bar at a methanol mole fraction of 0.25, but are merely used here to illustrate the typical behavior observed for many solutions. The RDFs display the usual oscillatory

behavior, which is dampened as one moves further away from the central molecule until all RDFs approach unity as the random bulk distribution is reached. The resulting KBIs also display a similar oscillatory behavior as the integration distance is increased. This is both an advantage and disadvantage. Clearly, any structure arising from the different solvation shells is apparent in the KBIs. However, these oscillations extend over many solvation shells, often more than appears in the original RDF due to the r^2 weighting in Equation 5.31, which can hinder the assignment of a final limiting value. The same oscillatory behavior is not observed by direct simulation of the fluctuations in a spherical region of space of the same volume, but not centered on a particular molecule. The values appear to converge much more rapidly with distance and reasonable estimates can be obtained for spherical radii of as small as 0.5 nm or so. As required, the KBI and fluctuation approaches tend to the same values.

Finally, we note that particle number fluctuations increase, and many properties diverge, as one approaches a critical point. O'Connell and others have shown that one can still apply FT under these circumstances by adopting integrals over direct correlation functions (DCF) (O'Connell 1971b; Chialvo 1993), and not integrals over the total correlation function as adopted in the standard KBIs (O'Connell and Liu 1998). The advantages of the DCF approach include the ability to treat systems close to critical points and a general simplification of the expressions provided above. The disadvantages are that the DCFs cannot be calculated directly, only through the defining Ornstein-Zernike expression, and that they appear to be more difficult to interpret than the KBIs. We will not consider the DCF approach here.

5.7 Limiting Expressions for the Fluctuating Quantities

In many cases one is interested in a particular fluctuating quantity as the concentration of a species tends to infinite dilution. The limiting behavior of the chemical potential derivatives is such that $\mu_{ij} \rightarrow \delta_{ij}$ as $\rho_j \rightarrow 0$. The behavior of the N_{ij}'s and F_i's under these conditions is the easiest to follow and is one of the reasons we adopted this particular formulation. It is clear that N_{ij} tends to zero as ρ_j tends to zero. However, the same is not true for N_{ij} and F_i when ρ_i tends to zero. Under these circumstances the following limiting expressions are found,

$$\delta_{ij} + N_{ij}^{\infty} = \lim_{N_i \rightarrow 1} \frac{\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle}{\langle N_i \rangle} = \langle N_j \rangle_i - \langle N_j \rangle \qquad 5.33$$

The interpretation of the above expression is as follows. The first term represents the average number of j molecules observed in a given local volume centered on a single infinitely dilute i particle, compared to the number of j particles observed in the same volume of solution in the absence of the i particle. The exact size of the local volume is unimportant as long as it is large enough to extend beyond the influence of the central i particle. When this condition is met, the difference between the two extensive quantities should be intensive with respect to the local volume. Correspondingly, for the F_i's we have,

$$F_i^\infty = \lim_{N_i \to 1} \frac{\langle N_i \varepsilon \rangle - \langle N_i \rangle \langle \varepsilon \rangle}{\langle N_i \rangle} = \langle \varepsilon \rangle_i - \langle \varepsilon \rangle \qquad \text{5.34}$$

which quantifies the difference between the average excess energy in the local region surrounding the i particle and the average excess energy in the same volume of solution in the absence of the i particle. The above expression can be expanded further to give,

$$F_i^\infty = \langle E_i \rangle - \langle E \rangle - \sum_j \left(\langle N_j \rangle_i - \langle N_j \rangle \right) H_j^\circ = \langle E_i \rangle - \langle E \rangle - \sum_j (\delta_{ij} + N_{ij}^\infty) H_j^\circ \qquad \text{5.35}$$

which is more convenient and useful for the analysis of computer simulations. Hence, the fluctuations have essentially disappeared and we find only the difference between average properties in various regions of the solution. The value of Δ_m remains a fluctuation representative of the bulk solution under these conditions.

The triplet correlations that appear in later sections have a more interesting limiting behavior. If we define the following fluctuating quantities,

$$N_{ijk} = \frac{\langle \delta N_i \delta N_j \delta N_k \rangle}{\langle N_i \rangle} \qquad N_{ij\varepsilon} = \frac{\langle \delta N_i \delta N_j \delta \varepsilon \rangle}{\langle N_i \rangle} \qquad N_{i\varepsilon\varepsilon} = \frac{\langle \delta N_i \delta \varepsilon \delta \varepsilon \rangle}{\langle N_i \rangle} \qquad \text{5.36}$$

and then use a similar approach as above one finds that as ρ_j tends to zero we have,

$$\begin{aligned}
N_{jik}^\infty &= N_{jki}^\infty = 0 \\
N_{ijk}^\infty &= \langle \delta N_j \delta N_k \rangle_i - \langle \delta N_j \delta N_k \rangle \\
N_{iik}^\infty &= \langle N_k \rangle_i - \langle N_k \rangle \\
N_{iii}^\infty &= 1
\end{aligned} \qquad \text{5.37}$$

where $\delta N_j = N_j - \langle N_j \rangle$, *etc*. The limiting expressions for the fluctuations involving the excess energy can be obtained by a simple substitution of N_j or N_k with ε.

5.8 Application to Binary Mixtures

5.8.1 Bulk Thermodynamic Properties in Terms of Local Fluctuating Quantities

Binary mixtures are the most common type of system studied using FT. The expressions for the bulk properties in terms of local fluctuations are of both conceptual interest and also useful for the practical analysis of theoretical or simulation results. If we consider a binary mixture of a solvent (1) and a solute (2) and then define two additional variables,

$$\begin{aligned}
\eta_{12} &= |M| = \rho_1 \left(1 + N_{22} - N_{12} \right) + \rho_2 \left(1 + N_{11} - N_{21} \right) \\
\zeta_2 &= |NN| = \left(1 + N_{11} \right) \left(1 + N_{22} \right) - N_{12} N_{21}
\end{aligned} \qquad \text{5.38}$$

one finds the following results,

$$\mu_{22} = \frac{\rho_1}{\eta_{12}} \qquad \left(\frac{\partial \beta \mu_2}{\partial \ln x_2}\right)_{T,p} = \frac{\rho}{\eta_{12}} \qquad \left(\frac{\partial \beta \mu_2}{\partial \ln \rho_2}\right)_{T,p} = \frac{1}{1 + N_{22} - N_{12}}$$

$$\bar{V}_i = \frac{1 + N_{jj} - N_{ij}}{\eta_{12}}$$

$$\bar{H}_i^{Ex} = \rho \bar{V}_i H_m^{Ex} + \frac{\rho_j}{\eta_{12}}\left(F_i - F_j\right) \tag{5.39}$$

$$k_B T \kappa_T = \frac{\zeta_2}{\eta_{12}}$$

$$k_B T^2 \alpha_p = \rho k_B T \kappa_T H_m^{Ex} - \phi_1 F_1 - \phi_2 F_2$$

$$k_B T^2 C_{p,m} = \Delta_m + k_B T^2 \alpha_p H_m^{Ex} - x_1 F_1 \bar{H}_1^{Ex} - x_2 F_2 \bar{H}_2^{Ex}$$

Other chemical potential derivatives can be found by application of the GD expression. All the quantities on the right hand side involve local fluctuations or properties that can be expressed in terms of local fluctuations using a previous expression. In principle, they can be obtained from a single simulation. Hence, one can determine, given a set of fluctuating quantities, all the properties on the left hand side by a stepwise process. The above expressions can be manipulated further to provide relationships for changes in the various activity coefficients,

$$\left(\frac{\partial \ln \gamma_2}{\partial \ln m_2}\right)_{T,p} = -\frac{\rho_1\left(N_{22} - N_{12}\right) + \rho_2\left(1 + N_{11} - N_{21}\right)}{\eta_{12}}$$

$$\left(\frac{\partial \ln f_2}{\partial \ln x_2}\right)_{T,p} = -\frac{\rho_1\left(N_{22} - N_{12}\right) + \rho_2\left(N_{11} - N_{21}\right)}{\eta_{12}} \tag{5.40}$$

$$\left(\frac{\partial \ln y_2}{\partial \ln \rho_2}\right)_{T,p} = -\frac{N_{22} - N_{12}}{1 + N_{22} - N_{12}}$$

which may be of more interest for certain applications.

At this point it is worth pausing to remind ourselves what we have accomplished here. The above expressions correspond to thermodynamic properties of solutions that are amenable to experiment. The expressions involve fluctuating quantities for an equivalent open system in which the average particle numbers are equal to the fixed number of particles per unit volume used in the experiment. The same argument holds for the chemical potentials, except in the opposite direction, and for the volume and the pressure, *i.e.* we are invoking the equivalence of ensembles approach. In addition, the open regions can be thought of as finite local regions in an essentially infinite bulk solvent. Hence, the approach provides information on local structure and fluctuations within the system. These local fluctuations are different from the bulk fluctuations characteristic of isothermal isobaric systems.

Examination of the above expressions clearly demonstrates that, even for binary mixtures, one requires combinations of multiple fluctuating quantities to provide the thermodynamic properties. This illustrates that the thermodynamic properties considered here are rather complex and will be difficult to relate to the underlying microscopic picture of molecular interactions and association. In contrast, focusing on the fluctuating quantities representative of local regions in the solution provides a simpler picture of solution behavior. It is more probable that one can explain the

fluctuating quantities – in terms of molecular distributions or interactions, hydrogen bonds, hydrophobic interactions, *etc.* – than one could explain the trends in the excess partial molar enthalpies, for instance. This is the argument we presented in the introduction.

In many cases one is interested in the limiting behavior when the solute concentration approaches zero. The limiting expressions are then,

$$\left(\frac{\partial \beta \mu_1}{\partial m_2}\right)^{\infty}_{T,p} = -1 \qquad \left(\frac{\partial \beta \mu_1}{\partial x_2}\right)^{\infty}_{T,p} = -1 \qquad \left(\frac{\partial \beta \mu_1}{\partial \rho_2}\right)^{\infty}_{T,p} = -\frac{1}{\rho_1^{\circ}}$$

$$\left(\frac{\partial \beta \mu_2}{\partial m_2}\right)^{\infty}_{T,p} = \frac{1}{m_2} \qquad \left(\frac{\partial \beta \mu_2}{\partial x_2}\right)^{\infty}_{T,p} = \frac{1}{x_2} \qquad \left(\frac{\partial \beta \mu_2}{\partial \rho_2}\right)^{\infty}_{T,p} = \frac{1}{\rho_2}$$

$$k_B T \kappa_T^{\circ} = \frac{1+N_{11}^{\circ}}{\rho_1^{\circ}} \qquad k_B T^2 \alpha_p^{\circ} = -F_1^{\circ} \qquad k_B T^2 C_{p,m}^{\circ} = \Delta_m^{\circ} \qquad \text{5.41}$$

$$\bar{V}_1^{\circ} = \frac{1}{\rho_1^{\circ}} \qquad \bar{V}_2^{\infty} = \frac{1+N_{11}^{\circ}-N_{21}^{\infty}}{\rho_1^{\circ}} = k_B T \kappa_T^{\circ} - G_{12}^{\infty}$$

$$\bar{H}_1^{\text{Ex,o}} = 0 \qquad \bar{H}_2^{\text{Ex,}\infty} = F_2^{\infty} - F_1^{\circ} = k_B T^2 \alpha_p^{\circ} + F_2^{\infty}$$

$$G_{12}^{\infty} = \left(\langle N_1 \rangle_2 - \langle N_1 \rangle\right)V_1^{\circ} \qquad F_2^{\infty} = \langle \varepsilon \rangle_2 - \langle \varepsilon \rangle$$

The limiting activity coefficient derivatives are provided by,

$$\left(\frac{\partial \ln \gamma_2}{\partial m_2}\right)^{\infty}_{T,p} = -\left[1 + \rho_1^{\circ}\left(G_{11}^{\circ} + G_{22}^{\infty} - 2G_{12}^{\infty}\right)\right]$$

$$\left(\frac{\partial \ln f_2}{\partial x_2}\right)^{\infty}_{T,p} = -\rho_1^{\circ}\left(G_{11}^{\circ} + G_{22}^{\infty} - 2G_{12}^{\infty}\right) \qquad \text{5.42}$$

$$\left(\frac{\partial \ln y_2}{\partial \rho_2}\right)^{\infty}_{T,p} = -(G_{22}^{\infty} - G_{12}^{\infty})$$

and can be used to analyze Henry's law, for instance (Shulgin and Ruckenstein 2002). The above expressions represent all the thermodynamic quantities typically provided by FT that are relevant for binary solutions. If one has access to the fluctuating quantities then it is possible to predict or rationalize the experimental properties. The main sources of the fluctuations would appear to be computer simulation and integral equation methods (Imai *et al.* 2000). Hence, one can determine the fluctuating quantities and use this information to make contact with real experimental data. This can thereby provide added confidence in the theoretical results.

5.8.2 Local Fluctuating Quantities in Terms of Bulk Thermodynamic Properties

Most investigations of closed binary systems have involved the determination and examination of the three KBIs, possibly followed by some description of local composition or preferential solvation, as a function of composition. The fluctuating quantities in terms of experimentally available properties can be obtained from Equation 5.25. It is more common, however, to express the particle fluctuations in terms of partial molar volumes, the compressibility
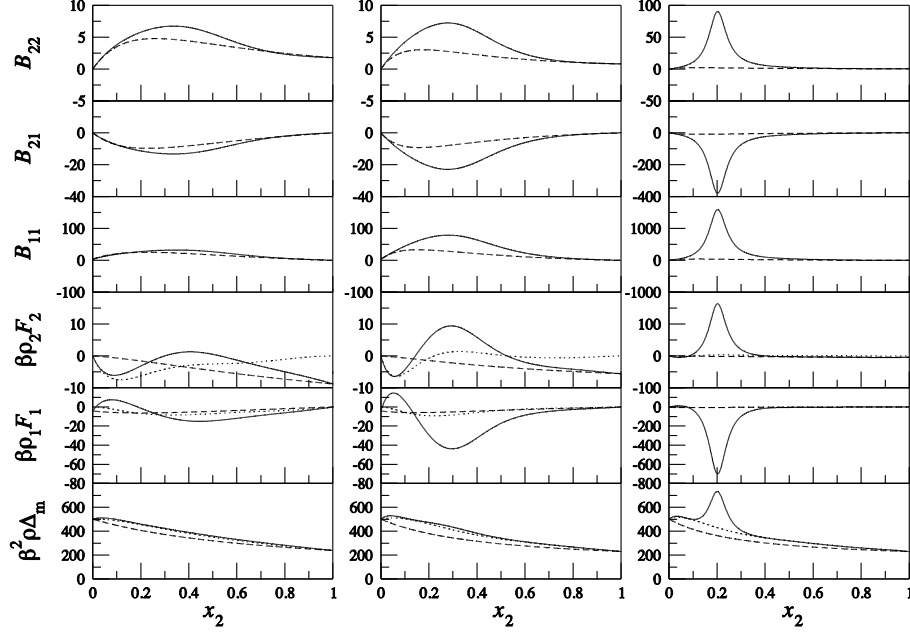
Figure 5.4 A fluctuation theory analysis of water (1) and alcohol (2) mixtures as a function of composition at 298 K and 1 bar. The panels correspond to methanol (left), ethanol (center), and *n*-propanol (right) mixtures, respectively. Experimental data were taken from the literature (Benjamin and Benson 1963; Douheret *et al.* 1989; Hu *et al.* 2003; Liltorp *et al.* 2005; Reis *et al.* 2008), and analyzed using Equation 5.43. The data are displayed as number-number (B_{ij}), number-excess energy ($\rho_i F_i$), and excess energy-energy ($\rho\Delta_m$) fluctuation densities in units of mol/L. Solid lines correspond to the experimental analysis and the dashed lines to the symmetric ideal values (see Section 4.4.4). The dotted lines appearing on the number-excess energy and excess energy-energy fluctuation plots correspond to the experimental $\beta\rho_i \bar{H}_i^{\mathrm{Ex}}$ and $\rho C_{p,\mathrm{m}} k_{\mathrm{B}}^{-1}$ data, respectively.

and chemical potential derivatives, rather than the pcp approach (Ben-Naim 1977). Hence, we provide the more common relationships here. The final expressions are,

$$\delta_{ij} + N_{ij} = \rho_j k_{\mathrm{B}} T \kappa_T + \frac{x_j^2}{x_1 x_2}\frac{(1-\phi_i)(1-\phi_j)}{\mu_{ij}}$$

$$F_i = -k_{\mathrm{B}} T^2 \alpha_p + (\delta_{i1} + N_{i1})\bar{H}_1^{\mathrm{Ex}} + (\delta_{i2} + N_{i2})\bar{H}_2^{\mathrm{Ex}} \qquad 5.43$$

$$\Delta_{\mathrm{m}} = k_{\mathrm{B}} T^2 C_{p,\mathrm{m}} - k_{\mathrm{B}} T^2 \alpha_p H_{\mathrm{m}}^{\mathrm{Ex}} + x_1 F_1 \bar{H}_1^{\mathrm{Ex}} + x_2 F_2 \bar{H}_2^{\mathrm{Ex}}$$

All the data on the right hand side of the above expressions can be obtained from experiment or involve fluctuating quantities which can be expressed in terms of experimental data.

5.8.3 Fluctuation Theory Analysis of Experimental Data

To illustrate the usefulness of the above expressions we have analyzed the experimental data for several binary solution mixtures. The results are presented in Figure 5.4 for a series of small primary alcohols in water at 298 K and 1 bar. The fluctuating quantities obtained from the experimental data using Equation 5.43 are displayed as fluctuation densities as a function of composition. The three top panels provide KBI related quantities and have been investigated before. The bottom three panels, which all involve excess energy fluctuations, have only become available recently.

The data indicate increasing values of B_{22} and B_{11} for the central regions, with a corresponding decrease in the B_{12} value, and an increasing trend with alcohol chain length. The trend with chain length is to be expected, as an increase in hydrophobic character would tend to promote local solute-solute and solvent-solvent self association or aggregation. This is manifested in an increase in the B_{ij} values and, therefore, an increase in the width of the solute-solute particle distribution. The B_{ij} values for water and propanol mixtures display a large peak or inverted peak around $x_2 = 0.2$ (note the scale change in this graph). The next alcohol in the series (*n*-butanol) phase separates in this region. What is most interesting is that no similar behavior is observable in the raw experimental data. This is usually the case. However, on extracting the particle number fluctuations, one immediately sees that the solution behavior in this composition region is the most noticeable feature of the mixture and, presumably, deserves the most attention.

The energy fluctuations display similar features. The excess energy-energy fluctuations (Δ_m) for methanol and ethanol mixtures with water are rather simple. The fluctuations vary slowly from those of pure water to those of pure methanol. They are somewhat similar to the values expected for ideal mixtures, and very similar to the values obtained from just the heat capacity term in Equation 5.43. However, this is not true when one examines the propanol and water mixtures. Clearly, the large particle number fluctuations significantly influence the energy-energy fluctuations in the region of 0.2 mole fraction. The energy fluctuations increase in this region. This is to be expected as the particle fluctuations are the ultimate source of the energy fluctuations, although the energy of each particle can vary significantly. The excess energy-particle number fluctuations also exhibit the same behavior. The values for the mixture of methanol and water are somewhat similar to the ideal values, and to the excess partial molar enthalpies. However, as one moves to ethanol and then propanol the local fluctuations again display interesting features that are not apparent in the experimental data used for extracting these values.

The data in Figure 5.4 indicate changes in the local properties of solution mixtures. In many cases, this can be expressed in terms of changes in the local composition of the mixture. This has proved to be a very successful approach to understanding solution structure and has led to the concept of preferential solvation (Ben-Naim 1988; Matteoli and Lepori 1995; Marcus 2001). Preferential solvation can be defined by use of the KBIs for both binary and ternary mixtures (Ben-Naim 1989, 1990), although only qualitative results are typically reliable as the size of the local region is generally unknown when examining experimental data. It may be possible to consider the energy fluctuations in a similar way, but we know of no such approach.

Another series of interest is the poly alcohols. A FT analysis of methanol, 1,2-ethanediol, and glycerol with water is presented in Figure 5.5. Clear trends in the particle number fluctuations are observed as the number of hydroxyl groups is increased. The general trend is towards solutions that mix "better" with water at the local level, as evidenced by the decrease in B_{ii} values. A shift in the positions of the maxima and minima to lower alcohol compositions also occurs. The excess energy-particle number fluctuation densities are more varied and specific
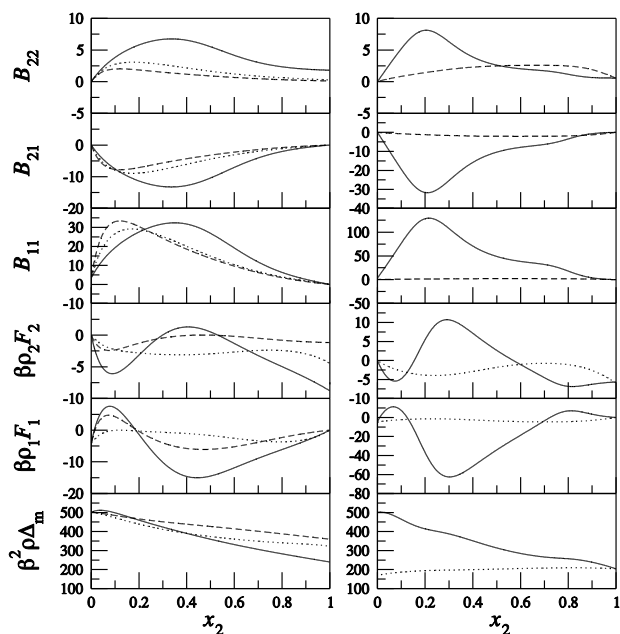
Figure 5.5 A fluctuation theory analysis of water (1) and alcohol (2) mixtures (left panel) and solvent (1) and acetone (2) mixtures (right panel) as a function of composition at 298 K and 1 bar. The left panel contains data for methanol (solid lines), 1,2-ethanediol (dotted lines), and glycerol (dashed lines). The right panel contains data for the solvents water (solid lines) and chloroform (dashed lines). Experimental data was taken from the literature (Mueller and Kearns 1958; Benjamin and Benson 1963; Coomber and Wormald 1976; Murthy and Subrahmanyam 1977; Apelblat *et al.* 1980; Villamanan *et al.* 1984; Villamanan and van Ness 1984; Akamatsu *et al.* 1987; Vannhu *et al.* 1987; Huot *et al.* 1988; Douheret *et al.* 1989; Grolier *et al.* 1994; Auslander *et al.* 1995; Marcus 2002; Hu *et al.* 2003; Reis *et al.* 2008), and analyzed using Equation 5.43. The data are displayed as number-number (B_{ij}), number-excess energy ($\rho_i F_i$), and excess energy-energy ($\rho\Delta_m$) fluctuation densities in units of mol/L.

trends depend on the composition of interest. The values for 1,2-ethanediol are relatively insensitive to composition, whereas the methanol solutions displayed the largest variation.

A final example is provided by mixtures of acetone as a solute with water or chloroform as a solvent. Again, this probes the effects of changes in the solvent polarity on the fluctuating quantities and their composition dependence. The results are also displayed in Figure 5.5. A significant degree of local water self association (a large positive B_{11}) is observed for most compositions with acetone. In contrast, mixtures of chloroform and acetone display no such behavior. The acetone and chloroform mixtures are clearly more ideal, whereas aggregation is observed in acetone and water mixtures due to an inability of acetone to donate hydrogen bonds. A similar trend is observed for the excess energy related fluctuations. While these trends are not particularly surprising, FT provides a clear and meaningful way to quantify these effects.

The analysis of experimental data in this form is well established for the KBIs and their related number fluctuation densities (Matteoli and Mansoori 1990). Alternatively, one can use the KBIs to define and quantify local composition and the degree of preferential solvation of one species by another in solution (Ben-Naim 1989; Marcus 2001; Ben-Naim 2006). In contrast, the analysis of the excess energy fluctuations is relatively new and the exact meaning of the

variations in the fluctuations still needs to be explored. What is clear from Figure 5.4 and Figure 5.5 is that features not present in the bulk properties become clear in the local fluctuations. The fluctuations in the excess energy provide a measure of the movement of particles in and out of the local region, coupled with how energetic these particles are. The shape of the F_1 and F_2 data curves presented in the previous examples are essentially mirror images of each other. This probably reflects the fact that (on average) a fluctuation of one of the molecules will tend to increase the excess energy of the local region, while a change in the number of the other type of particle will therefore decrease the excess energy of the local region. This is simply related to the differences in the single particle excess energies. This issue will become clearer following the results presented in Section 10.

5.8.4 Cosolvent Effects on Surface Tension

The distribution of solutes at the solution/vapor interface has important consequences in chemistry and chemical engineering (Richmond *et al.* 1988; Jungwirth and Tobias 2006). Consequently, a variety of experimental and theoretical approaches have been used to help understand surface adsorption or exclusion. It is well established that an increase in the surface tension of a solution due to the addition of a solute indicates exclusion of that solute from the interface region, and *vice versa*. Unfortunately, it has been difficult to fully quantify such structural and thermodynamic changes using computer simulation data so that they may be compared with experimental data on surface adsorption. A recent application of FT to this problem has provided a solution to this problem (Chen and Smith 2008).

Analysis of the GD expressions for the surface and bulk solution regions provides a differential for the surface tension (γ_s) of a liquid-vapor interface (Tronel-Peyroz *et al.* 1989). From this one can obtain the standard Gibb absorption expression for the change in surface tension on addition of a cosolvent,

$$\left(\frac{\partial \gamma_s}{\partial x_i}\right)_{\beta,p} = -\sum_{j>1}\Gamma_{sj}\left(\frac{\partial \mu_j}{\partial x_i}\right)_{\beta,p} \qquad 5.44$$

where the sum is over all components except for the primary solvent, and one could use any concentration derivative on both sides of the expression. The chemical potential derivative is a property of the bulk solution away from the interface, while the Γ_{sj} values quantify the surface adsorption of j molecules relative to the primary solvent. For a simple binary mixture this reduces to,

$$\left(\frac{\partial \gamma_s}{\partial x_2}\right)_{\beta,p} = -\Gamma_{s2}\left(\frac{\partial \mu_2}{\partial x_2}\right)_{\beta,p} \qquad 5.45$$

The activity derivative must be positive for stable solutions and hence the direction of the surface tension change is dictated by the sign of Γ_{s2}. The activity derivative is provided by FT as indicated in Equation 5.39. The surface adsorption can be related to the number of molecules within the interface region, or in terms of surface probability distributions by,

$$\Gamma_{sj} = A^{-1}\left(\langle N_j\rangle_s - m_j\langle N_1\rangle_s\right) = \rho_j\int_0^\infty [g_j^s(z) - g_1^s(z)]dz \qquad 5.46$$

where z is the direction perpendicular to the surface of area (A). The integration over $g_j - g_1$ should be performed to a distance at which the integral remains unchanged, *i.e.* $g_j = g_1 = 1$, which may involve many solvation shells away from the surface. This type of formulation has proved useful in obtaining a consistent picture of solute effects on surface tension as provided by computer simulation (Chen and Smith 2008).

5.8.5 Force Fields for Molecular Simulation

Fluctuation theory provides information on local fluctuations in solution and may be used to characterize the properties of mixtures. The fluctuations or integrals provided by FT are also relatively easy to obtain from computer simulation. Clearly, the results obtained from a computer simulation should agree with the experimentally extracted fluctuations providing that the description of the intermolecular interactions, *i.e.* the force field, is sufficiently accurate and one has adequate sampling of the solution distributions. Therefore, one can use FT results to test and improve the quality of force fields used for molecular simulation (Weerasinghe and Smith 2003b; Lee and van der Vegt 2005). We have taken this approach to help develop new, hopefully improved, force fields for the simulation of liquid mixtures (Weerasinghe and Smith 2003b, 2003d).

Force fields, especially in the biomolecular simulation area, have traditionally been developed with the aid of quantum calculations (Kaminski *et al.* 2001; Hornak *et al.* 2006). The calculations involve properties of molecules in the gas phase. These properties change in the condensed phase. This is, of course, a rather well known effect. However, exactly how one has to adjust the description of intermolecular interactions to account for the role of molecular polarization in liquids is still unknown. Furthermore, many force fields only test the resulting models using properties of pure liquids. This is a necessary, but not sufficient test, to establish the quality of a force field. More recently, thermodynamic data such as the solvation free energies of solutes have been used to improve the quality of many force fields (Schmid *et al.* 2011). Unfortunately, this quantity does not depend on an accurate description of the solute-solute interactions. We have found that relatively poor agreement with the KBIs is observed for many solution mixtures (Chitra and Smith 2001a; Weerasinghe and Smith 2003d; Kang and Smith 2006). This clearly indicates an incorrect balance between solute-solute, solute-solvent, and solvent-solvent distributions which arises due to an incorrect description of the interactions. The main problem appears to be related to the choice of effective charges characterizing the polarity of a molecule or ion.

In an attempt to improve current force fields, we have designed a set of solute-solvent models that specifically aim to reproduce the experimentally derived KBIs, and thereby ensure a reasonable balance in the molecular distributions. These are termed Kirkwood-Buff derived Force Field (KBFF) models. Several recent reviews highlight our progress in this area (Ploetz *et al.* 2010a; Weerasinghe *et al.* 2010). In developing these new models our main focus has been on modifying the charge distributions assigned to the solutes and solvents. This has led to a series of models for small solutes representative of the functional groups found in peptides and proteins, together with various cosolvents of biological interest. The result is a complete force field for the simulation of peptides and proteins based on the KBFF approach (http://kbff.chem.k-state.edu). It is hoped that these models provide a
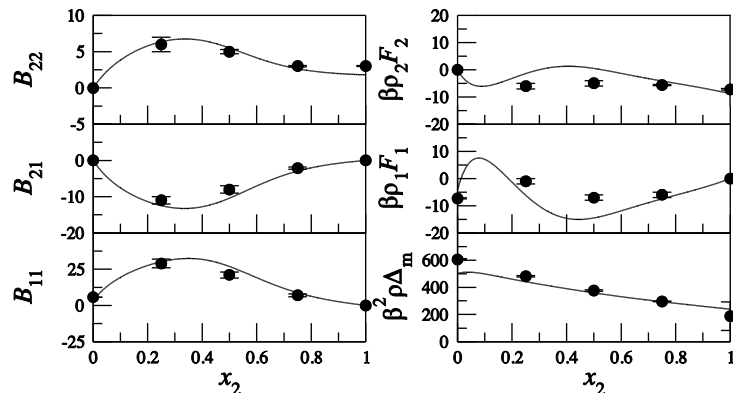
Figure 5.6 A fluctuation theory and computer simulation analysis of water (1) and methanol (2) mixtures as a function of composition at 298 K and 1 bar. Experimental data (solid lines) was taken from the literature and analyzed using Equation 5.43 (see Figure 5.4 caption). The simulated data (closed circles) were determined from simulations using the KBFF model for methanol (Weerasinghe and Smith 2005), together with the SPC/E water model. The data are displayed as number-number (B_{ij}), number-excess energy ($\rho_i F_i$), and excess energy-energy ($\rho\Delta_m$) fluctuation densities in units of mol/L.

significant degree of improvement for the accurate computer simulation of biomolecular systems. Others have taken a similar approach for small solutes (Lee and van der Vegt 2005; Klasczyk and Knecht 2010).

As an example of the quality of data that can be obtained from a computer simulation, we present the results of a FT analysis of methanol and water mixtures in Figure 5.6. The simulations employed the KBFF model for methanol (Weerasinghe and Smith 2005), the SPC/E model for water (Berendsen *et al.* 1987), and were performed in periodic unit cells of 6 nm for simulation times of 50 ns with the same protocols as our other recent studies (Ploetz and Smith 2011a). The simulations were then analyzed to determine the fluctuating quantities and compared with the analysis performed for Figure 5.3. The particle number and energy fluctuations are very well reproduced with all the same observed trends. The agreement is quantitative for most compositions. The models did not perform so well for the particle-energy fluctuations. This is somewhat surprising considering the excess enthalpy of mixing is very reasonable, although not perfect (Weerasinghe and Smith 2005), compared to many other methanol models. It appears that the partial molar enthalpies, which are related to derivatives of the excess enthalpy of mixing are more sensitive indicators of the quality of a particular model. Nevertheless, the particle-energy fluctuations are of the appropriate magnitude. This provides some confidence in the quality of the distributions provided in Section 5.9.

5.9 Particle Number and Energy Distributions in Solution

The fluctuasting quantities outlined in the previous sections are related to the moments of the particle number and energy probability distribution $P = P(N_1, N_2, \ldots, E$ or $\varepsilon)$. For systems far away from a critical point a multivariate Gaussian distribution in the energy and particle numbers is often assumed. Hence, one can write,

$$P(Y) = \frac{1}{(2\pi)^{(n_c+1)/2} \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(Y - \langle Y \rangle)^{\mathrm{T}} \Sigma^{-1} (Y - \langle Y \rangle)\right) \qquad 5.47$$

where Y is a vector of the form $Y = (N_1, N_2, \ldots, E$ or $\varepsilon)$ and Σ is a symmetric correlation matrix of the form,

$$
\begin{pmatrix}
\langle \delta N_1 \delta N_1 \rangle & \langle \delta N_1 \delta N_2 \rangle & \cdots & \langle \delta N_1 \delta E \rangle \\
\langle \delta N_2 \delta N_1 \rangle & \langle \delta N_2 \delta N_2 \rangle & \cdots & \langle \delta N_2 \delta E \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle \delta E \delta N_1 \rangle & \langle \delta E \delta N_2 \rangle & \cdots & \langle \delta E \delta E \rangle
\end{pmatrix}
\qquad 5.48
$$

This matrix is clearly related to the matrices provided earlier. In this section we investigate the above multidimensional distribution in more detail. It should be noted that various previous studies along this line have appeared in the literature. Kirkwood and Goldberg provided an elegant analysis of light scattering in multicomponent systems, which focused on composition fluctuations using a multivariate Gaussian approach (Kirkwood and Goldberg 1950). More recently, Mazo has analyzed the eigenvalues of the B matrices for a series of binary systems and observed that most systems, over large composition ranges, can be described by B matrices that contain a large eigenvalue and a small eigenvalue, to the extent that the smaller eigenvalue can almost be neglected when attempting to understand the system properties (Mazo 2008; Mazo *et al.* 2009). We will relate the current findings to these results. We could find no such analysis that included the particle-energy and energy-energy distributions.

Using computer simulations one can extract not only the number of particles of each type in a local region of space, but also the energy of that region. This latter quantity requires a pairwise decomposition of intermolecular interactions. Fortunately, this is the approach adopted by most (even polarizable) force fields for liquid simulations.

In Figure 5.7 we provide the natural logarithm of the number and energy distributions for methanol and water mixtures. It is immediately clear that the distributions are not Gaussian in nature. While the energy distribution appears to be essentially symmetrical, the N_1 and N_2 probability distributions display significant asymmetry. The fact that there are differences between the assumed Gaussian distribution and the true distribution was recognized some time ago (Greene and Callen 1951). Fortunately, it appears that the second moments are the same in both cases.
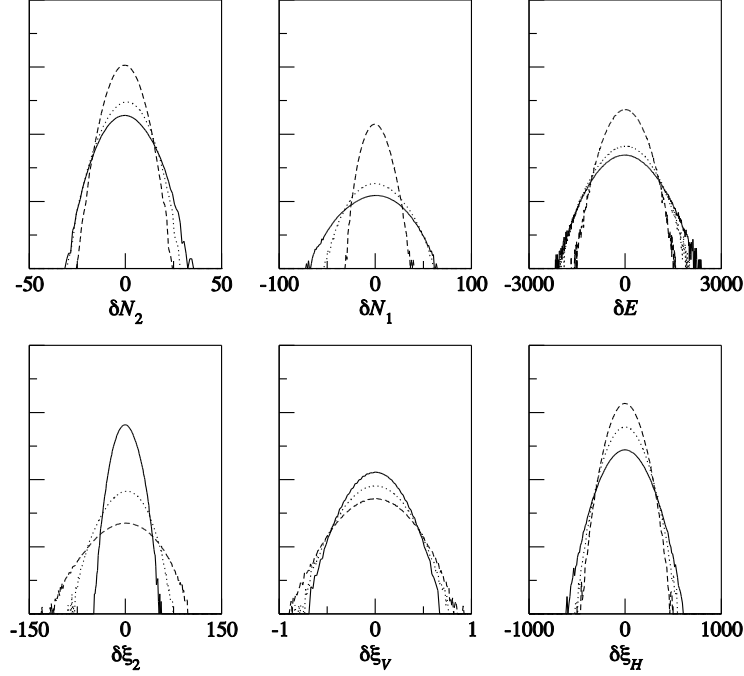
Figure 5.7 Natural logarithm of the number, energy, and excess energy probability distributions for methanol (2) and water (1) mixtures at 300 K and 1 bar obtained from computer simulation. The distributions correspond to local spherical regions with a radius of 1.5 nm. The raw data are displayed in the top panel, while the transformed quantities of Equation 5.50 are displayed in the bottom panel. The probabilities are in arbitrary units. Energies and ξ_H are in units of kJ/mol, with ξ_V in units of L/mol. Only the intermolecular energies are used in the analysis. The solid, dotted, and dashed lines correspond to methanol mole fractions of 0.25, 0.50 and 0.75, respectively.

In Figure 5.8 we provide scatter plots for the particle number and energy distributions obtained from simulations of methanol and water mixtures. It is clear that the data are correlated, *i.e.* when N_1 is smaller than the average then N_2 is larger than the average. This leads to particle number correlations with negative slopes. This correlation also appears in the energy probability distributions, but the slope can have either sign. In this system the energy of the region (E) tends to become lower in energy when fluctuations increase the number of water molecules, and *vice versa*. This is a result of the more favorable average energy per molecule for water compared to methanol.

The results follow the trends described by Mazo for his analysis of particle-particle fluctuations (Mazo 2008). In particular, fluctuations in the various particle numbers are related through Equation 5.18. This indicates that the fluctuation in the number of N_1 and N_2 particles are related to the respective partial molar volumes and the compressibility. For liquids approaching zero compressibility, albeit an impossible situation, the fluctuations are simply given by,

$$k_B T \kappa_T = \langle \delta N_1 \delta N_1 \rangle \langle N_1 \rangle^{-1} \overline{V}_1 + \langle \delta N_1 \delta N_2 \rangle \langle N_1 \rangle^{-1} \overline{V}_2 \approx 0 \qquad 5.49$$

and therefore suggests that $\delta N_1 \overline{V}_1 \approx -\delta N_2 \overline{V}_2$ on the average. This simply states that, for an incompressible fluid, exchanges of particle numbers occur in order to maintain the same total volume.
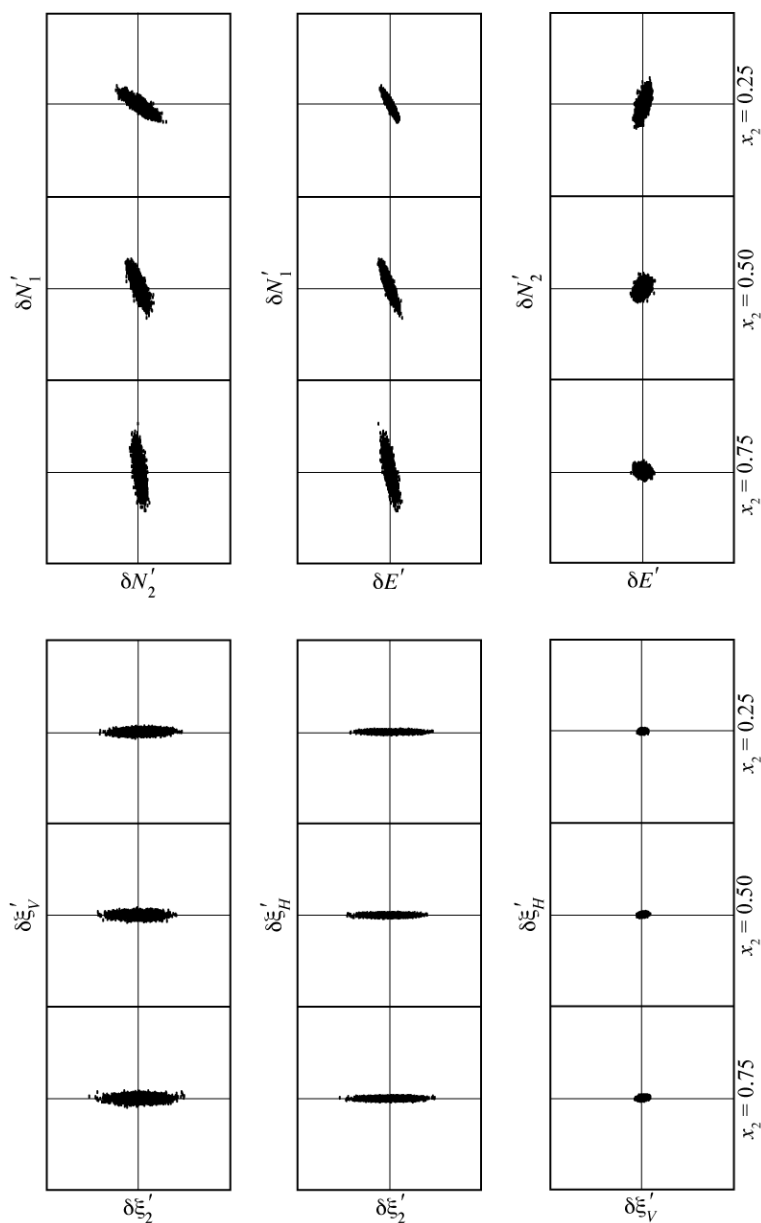
Figure 5.8 Number-number and number-energy distributions for methanol (2) and water (1) mixtures at 300 K and 1 bar obtained from computer simulation. The distributions correspond to local spherical regions with a radius of 1.5 nm. In the top panel, the reduced number, $\delta N'_i = \delta N_i / <N_i>$, and energy, $\delta E' = \delta E / |E|$, fluctuations are provided. In the bottom panel, the transformed quantities of Equation 5.50 are displayed as reduced properties $\delta \xi'_2 = \delta \xi_2 / <N_2>$, $\delta \xi'_V = \delta \xi_V / V$, and $\delta \xi'_H = \delta \xi_H / |E|$. All axes range from -1 to +1 with lines drawn through $x = 0$ and $y = 0$. Only the intermolecular energies are used in the analysis.

Diagonalization of the above matrix provides a series of eigenvalues and eigenvectors – three in the case of a binary solution. The eigenvalues correspond to the magnitude of the principle components of the various fluctuations. The eigenvectors form the components of a 3×3 rotation matrix, which transform the orthogonal fluctuating quantities into the observed results (Equation 5.48). To our knowledge, general expressions for the eigenvectors and eigenvalues

have not been obtained. However, for binary systems one can diagonalize the number fluctuation matrix as illustrated below, and thereby partially diagonalize the full fluctuation matrix.

The above type of analysis was generalized by Kirkwood and Goldberg for particle number fluctuations (Kirkwood and Goldberg 1950). They observed that one could decouple the composition and density fluctuations by transformations of the form,

$$\xi_V = \sum_i N_i \bar{V}_i$$
$$\xi_i = N_i - m_i N_1$$
$$\xi_H = E - \sum_i N_i \bar{H}_i = \varepsilon - \sum_i N_i \bar{H}_i^{\mathrm{Ex}}$$

5.50

The final transformation was prompted by our recent work (Jiao and Smith 2011; Ploetz and Smith 2011b). Using these transformations one can show that,

$$\langle \xi_V \rangle = V \qquad \langle \xi_2 \rangle = 0 \qquad \langle \xi_H \rangle = -pV$$
$$\langle \delta\xi_V \delta\xi_V \rangle = Vk_{\mathrm{B}}T\kappa_T \qquad \langle \delta\xi_H \delta\xi_H \rangle = k_{\mathrm{B}}T^2 C_p$$
$$\langle \delta\xi_2 \delta\xi_2 \rangle = \frac{\langle N_2 \rangle}{\mu_{22}} \qquad \langle \delta\xi_V \delta\xi_H \rangle = -Vk_{\mathrm{B}}T^2 \alpha_p$$
$$\langle \delta\xi_V \delta\xi_2 \rangle = \langle \delta\xi_H \delta\xi_2 \rangle = 0$$

5.51

It should be noted that the first and last transformations in Equation 5.50 can also be used to transform the Gibbs ensemble expressions for the compressibility, thermal expansion and heat capacity (Equation 5.11) into the expressions provided in Equation 5.51 (Ploetz and Smith 2011b). For binary systems the transformations provided in Equation 5.50 generate a diagonal correlation matrix for the particle number fluctuations. One of the eigenvalues is related to the compressibility of the system, while the other is related to a single chemical potential derivative (Kirkwood and Goldberg 1950). The former is the small eigenvalue from the analysis performed by Mazo (Mazo 2008). However, the matrix displayed in Equation 5.48 is not diagonal using the above transformations, due to the cross correlation in the energy and volume transformations. Furthermore, even the number fluctuation matrices for ternary or higher component systems are not diagonal with the above transformations.

Application of the above transformations to computer simulation data is also provided in Figure 5.7 and Figure 5.8. The ξ_H and ξ_V probability distributions provided in Figure 5.7 are very close to symmetrical. Furthermore, the ξ_V distribution in Figure 5.7 is no closer to being symmetrical than the individual particle number distributions. In Figure 5.8 one observes that the collective particle number (ξ_2) and density (ξ_V) fluctuations are clearly orthogonal, as are the collective particle number and energy fluctuations (ξ_H), although the excess energy and volume fluctuations still present a degree of correlation related to the thermal expansion coefficient. In principle, one can fully diagonalize the above matrix of fluctuations. However, the corresponding transformations are unknown, even for binaries, and it is not clear that this is even required. The partially diagonalized matrix of number fluctuations was satisfactorily used by Kirkwood and Goldberg to study light scattering in multicomponent systems
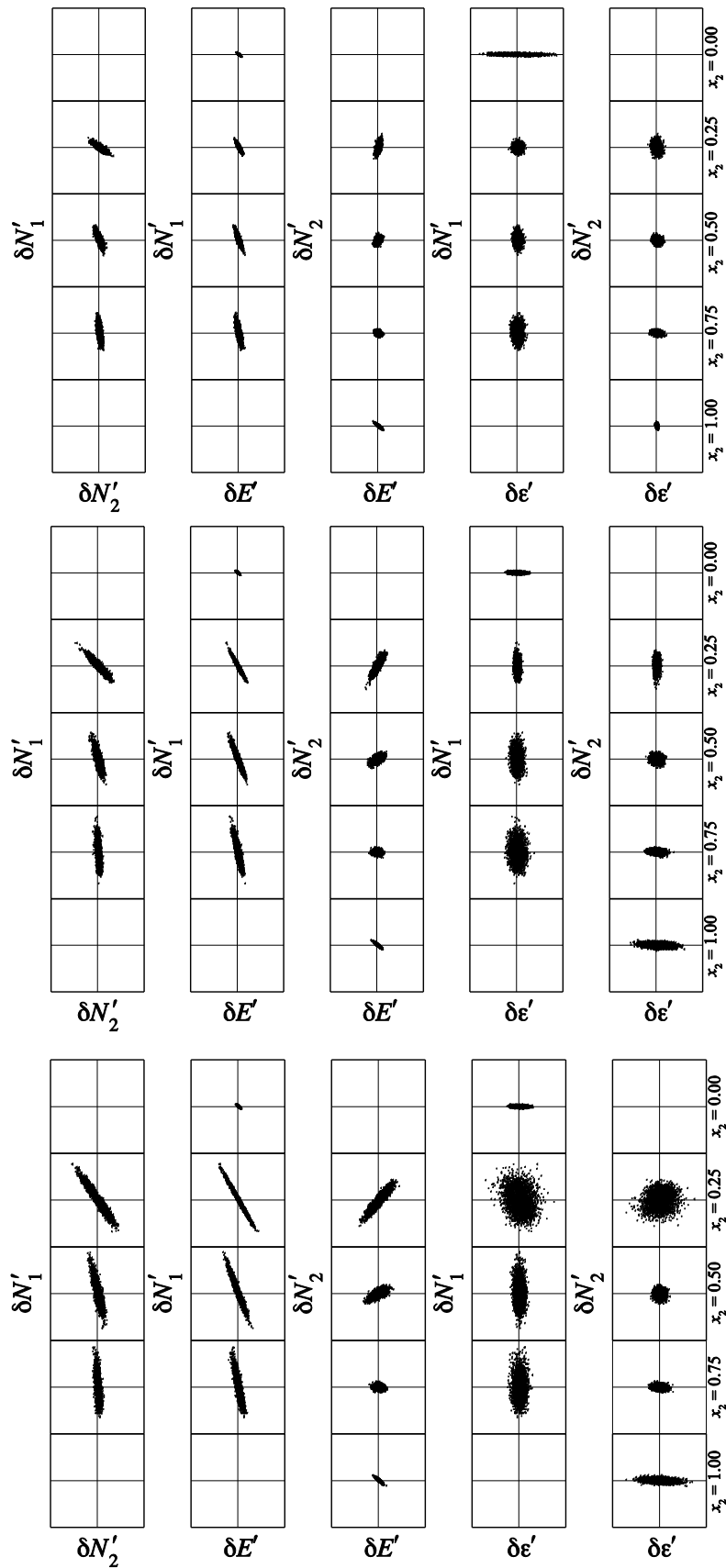
Figure 5.9 Number-number and number-energy distributions for the alcohol (2) and water (1) mixtures at 300

K and 1 bar obtained from computer simulation. The distributions correspond to local spherical regions with a radius of 1.5 nm. Data are displayed for mixtures of water with methanol (top panel), ethanol (center panel) and *n*-propanol (bottom panel) as a function of alcohol mole fraction. The reduced number, $\delta N'_i = \delta N_i / <N_i>$, energy $\delta E' = \delta E / |E|$, and excess energy $\delta\varepsilon' = \delta\varepsilon / |\varepsilon|$ fluctuations are provided. All axes range from -1 to +1, with the exception of the $\delta\varepsilon'$ axis which ranges from -10 to 10 for methanol and from -30 to 30 for ethanol and *n*-propanol. Lines are drawn through $x = 0$ and $y = 0$ for reference. Only the intermolecular energies are used in the analysis.

(Kirkwood and Goldberg 1950), and hence it appears that other potential applications may be viable with a partial diagonalization.

In Figure 5.9 we compare the fluctuations in particle, energy, and excess energy obtained from simulation for the linear alcohols in water. The number-number fluctuations all display a negative slope, as indicated above, the magnitude of which increases with alcohol size and composition. The former trend is also a consequence of the approximate relationship provided by Equation 5.49. The reason for the latter trend is less obvious. The energy-number distributions also vary in magnitude and shape as one changes composition and system. The change in slope for the energy-alcohol number distributions is understandable. On average the local energy will become more unfavorable as the number of alcohol molecules in the region increases (pure water has a higher energy density than the pure alcohols in our models), while the pure solutions will display a negative slope if there is a net cohesive energy.

Also shown in Figure 5.9 are the fluctuations in the excess energy of the region. It is clear that distributions involving the excess energy and the particle number are almost orthogonal. This is due to the fact that the transformation to the excess energy is numerically very close to the transformation of Equation 5.50. The difference involves the excess enthalpies, which are typically small in magnitude compared to the pure enthalpies. Interestingly, it is only in the excess energy distributions that one observes features that may help explain the changes observed in the experimental fluctuations in the region of $x_2 = 0.2$ for propanol. The change in shape and character of the simulated distributions appears to suggest that, at low propanol compositions, the energy fluctuations increase substantially and become much less correlated with either particle number fluctuation. It is precisely these types of features which may help to deepen our understanding of changes in solution properties with composition. Clearly, additional work needs to be performed on more systems to fully appreciate this type of behavior. In our opinion, the combination of computer simulation and FT will play a major role in this area.

5.10 The Effects of Temperature, Pressure and Composition on Local Fluctuations

The local fluctuations or KBIs are dependent on temperature and pressure in addition to composition, although the composition dependence appears to be the most significant. In Section 5.1 we illustrated how one can use derivatives of the average particle numbers and energy to provide expressions involving fluctuating quantities. The same approach can be applied to the resulting fluctuations. Substituting the fluctuations for X in Equation 5.15 one can determine the required derivatives (Jiao and Smith 2011). The simplest expressions involve the B_{ij}'s and provide,

$$\left(\frac{\partial B_{ij}}{\partial \beta}\right)_{p,\{N\}} = -\frac{\langle \delta N_i \delta N_j \delta \xi_H \rangle}{V}$$

$$\left(\frac{\partial B_{ij}}{\partial p}\right)_{\beta,\{N\}} = \beta \frac{\langle \delta N_i \delta N_j \delta \xi_V \rangle}{V} \qquad 5.52$$

$$\left(\frac{\partial B_{ij}}{\partial x_l}\right)_{p,\beta} = \frac{1}{x_l(1-x_l)V} \sum_k \langle \delta N_i \delta N_j \delta N_k \rangle \mu_{kl}$$

for any number of components. The first two expressions have been simplified by using the transformations for ξ_V and ξ_H provided in Equation 5.50. Some of the chemical potential derivatives can also be eliminated using the GD expression if desired. All the above derivatives involve triplet correlations. It is immediately apparent, therefore, that if the distribution of particle numbers and excess energy were truly characterized by a multivariate Gaussian, then the above derivatives would all be zero. Hence, the observed variations of the B_{ij}'s with temperature, pressure and composition probe the nature of the underlying true (non-Gaussian) distribution.

In principle, the experimental data required to compute the above derivatives is available for many systems. However, we know of very few (experimental or simulation) studies in this direction for all but infinitely dilute solutes (Matteoli and Lepori 1984; Huot *et al.* 1988). The most relevant is a recent analysis of the composition dependence of the osmotic pressure for binary systems (Karunaweera *et al.* 2012), where the expressions are somewhat simpler to develop. It is highly probable that such an analysis is numerically challenging. Given the sensitivity of the KBIs to the quality of the activity data, the additional derivatives required for the triplet correlations could be problematic. Furthermore, it is unlikely that the KBIs vary dramatically with pressure, unless one examines very high pressures. The interpretation of the above derivatives is also complicated by the presence of many terms, although all the terms could be determined.

The last expression in Equation 5.52 does, however, provide some insight in to the nature of the composition dependence of the B_{ij}'s. For many binary mixtures one observes maxima and minima in the B_{ij}'s as a function of composition. Using the GD expression it is possible to show that this maxima or minima occurs when $N_{1ij} = N_{2ij}$. The significant changes of the B_{ij}s with composition observed in many systems suggest the particle number fluctuations cannot be Gaussian or symmetric in nature.

Matteoli and Lepori have investigated the experimental data for aqueous methanol and ethanol mixtures as a function of temperature (Matteoli and Lepori 1984). The alcohol KBIs (G_{22}) are essentially independent of temperature until one reaches low solute concentrations. In this region the G_{22} values systematically increase with temperature for both alcohols. This clearly indicates a positive sign for the triplet fluctuation at low alcohol concentrations according to Equation 5.52. However, the small changes of the properties with temperature suggests that the distribution must be close to $\langle \delta N_2 \delta N_2 \delta \xi_H \rangle$ symmetric in nature.

The other primary fluctuating quantities can also affected by temperature, pressure and composition. The variations in the F_i's are provided by,

$$\left(\frac{\partial \rho_i F_i}{\partial \beta}\right)_{p,\{N\}} = -\frac{\langle \delta N_i \delta \varepsilon \delta \xi_H \rangle}{V} + \frac{k_B T^2}{V}\sum_j \langle \delta N_i \delta N_j \rangle C^\circ_{p,m,j}$$

$$\left(\frac{\partial \rho_i F_i}{\partial p}\right)_{\beta,\{N\}} = \beta \frac{\langle \delta N_i \delta N_j \delta \xi_V \rangle}{V} - \frac{1}{V}\sum_j \left[\langle \delta N_i \delta N_j \rangle V^\circ_{m,j}(1-T\alpha^\circ_{p,j})\right]$$

$$\left(\frac{\partial \rho_i F_i}{\partial x_l}\right)_{p,\beta} = \frac{1}{x_l(1-x_l)V}\sum_j \langle \delta N_i \delta N_j \delta \varepsilon \rangle \mu_{jl}$$

5.53

whereas the variation in Δ_m is provided by,

$$\left(\frac{\partial \rho \Delta_m}{\partial \beta}\right)_{p,\{N\}} = -\frac{\langle \delta \varepsilon \delta \varepsilon \delta \xi_H \rangle}{V} + \frac{2k_B T^2}{V}\sum_i \langle \delta N_i \delta \varepsilon \rangle C^\circ_{p,m,i}$$

$$\left(\frac{\partial \rho \Delta_m}{\partial p}\right)_{\beta,\{N\}} = \beta \frac{\langle \delta \varepsilon \delta \varepsilon \delta \xi_V \rangle}{V} - \frac{2}{V}\sum_i \left[\langle \delta N_i \delta \varepsilon \rangle V^\circ_{m,i}(1-T\alpha^\circ_{p,i})\right]$$

$$\left(\frac{\partial \rho \Delta_m}{\partial x_l}\right)_{p,\beta} = \frac{1}{x_l(1-x_l)V}\sum_i \langle \delta \varepsilon \delta \varepsilon \delta N_i \rangle \mu_{il}$$

5.54

However, we know of no studies that have attempted to characterize these variations.

For binaries solutions it is possible to introduce the definition of ξ_2 to simplify the above expressions. A further simplification is provided if one considers the activity derivative. The final expressions are then,

$$\left(\frac{\partial B_{ij}}{\partial \beta \mu_2}\right)_{p,\beta} = \frac{\langle \delta N_i \delta N_j \delta \xi_2 \rangle}{V}$$

$$\left(\frac{\partial \rho_i F_i}{\partial \beta \mu_2}\right)_{p,\beta} = \frac{\langle \delta N_i \delta \varepsilon \delta \xi_2 \rangle}{V}$$

$$\left(\frac{\partial \rho \Delta_m}{\partial \beta \mu_2}\right)_{p,\beta} = \frac{\langle \delta \varepsilon \delta \varepsilon \delta \xi_2 \rangle}{V}$$

5.55

Hence, the activity dependence of the fluctuating quantities probes the nature of the correlation between the transformed particle number fluctuation and the original fluctuating quantities.

5.11 Analysis of Computer Simulation Data

It is clear that computer simulations represent one of the better approaches to determining the local fluctuations. There are, however, some technical difficulties which can arise during a typical analysis. We discuss some of these issues here. Most evaluations of the KBIs have used the integration approach, in contrast to the actual particle number fluctuations. Furthermore, as the vast majority of simulations are performed for closed periodic systems, one is naturally limited to performing the integration out to some cutoff distance from the particle of interest. It is then assumed that that the RDFs are essentially unity beyond this distance. In practice, deciding exactly where to truncate

the integral can be rather subjective. In our previous studies we usually plot the KBIs as a function of integration or cutoff distance. Hence, one can define distance dependent KBIs (and even distance dependent thermodynamic functions) such that,

$$G_{ij}(R) \approx 4\pi \int_0^R [g_{ij}^{(2)}(r) - 1]r^2 dr \qquad\qquad 5.56$$

A plot of $G_{ij}(R)$ can be very informative. In favorable cases, the integral converges and one observes a limiting constant value for the KBI. More typically, there can be significant statistical noise that obscures the real limiting behavior. The situation can be improved by simulating relatively large systems (>6 nm box length) for times in excess of 10 ns. However, this approach is not always productive.

Several studies have investigated this problem with a variety of suggested solutions (Nichols *et al.* 2009; Wedberg *et al.* 2010; Schnell *et al.* 2011). The problem seems to be particularly acute when determining the isothermal compressibility, especially if this property is plotted as a function of the integration distance. Slowly converging oscillatory behavior is commonly observed. The isothermal compressibility obtained from simple integration should be considered questionable at best. Fortunately, the situation is significantly improved when examining the partial molar volumes and chemical potential derivatives. This is probably related to the fact that the partial molar volume and chemical potential derivative expressions only involve differences in the KBIs. In the majority of our previous simulations, we have chosen to extract the final KBI values after averaging the integrals over a short range of distances (typically a full solvation shell) where the integrals appear to be converging (Weerasinghe and Smith 2003c). This is a subjective approach. However, we have tested the values of the resulting properties with other approaches – typically finite difference compressibilities or partial molar volumes obtained from the simulated densities – and found good agreement (Weerasinghe and Smith 2005). This provides some confidence in the above approximation. An investigation of the activity of urea and water mixtures has also been performed using thermodynamic integration techniques (Kokubo *et al.* 2007), and provides excellent agreement with our own KBI determined values. However, care should always be taken to ensure that reasonable estimates for the KBIs are obtained.

A slightly different approach is usually adopted for the particle-energy (F_i) and energy-energy (Δ_m) fluctuating quantities. First, to evaluate these properties one has to determine the energy of a given volume of the solution. This is possible if one can assign energies to each particle in the system. Single particle potential energies can be determined by assigning one half of the potential energy for the interaction between each pair of molecules (Debenedetti 1987b). This is even possible for models that include explicit polarization. In principle, the internal energy should include all intramolecular contributions and the kinetic energy of the molecules. In practice, this may be somewhat awkward to achieve. It may be safe to assume the intramolecular potential energies cancel in determining the value of ε, *i.e.* they are the same in the mixture and the pure liquids. The kinetic energy contribution to the local energy cancels in the calculation of the F's such that one can use just the potential energies for both the instantaneous internal energies (E) and the reference pure solvent values. The kinetic energy contribution to the $\beta^2 \Delta_m$ values does not cancel. For the nonlinear molecules studied here this contribution is simply 3 and arises from the three translational and three

rotational degrees of freedom. The intramolecular kinetic energy contribution is more problematic as high frequency vibrations will not contribute to the heat capacity at normal temperatures. However, heavier molecules at higher temperatures could provide a significant contribution. Classical simulations would provide the equipartition values at all temperatures for all non-constrained vibrational degrees of freedom. A series of random origins can then be chosen and the instantaneous energy and number of particles determined as a function of the volume of a local region centered at each origin. The fluctuations observed after averaging over the trajectory provide values of F_i and Δ_m as a function of local volume (sphere radius). A graph of F_i or Δ_m as a function of sphere radius should tend to a constant value for radii large enough to approach the bulk solution distributions (see Figure 5.3).

Another issue arises when determining the KBIs, or local distributions, around infinitely dilute solutes. An example of this includes our study of the solvent and cosolvent distributions around a single protein solute (Kang and Smith 2007). As the simulations involve finite systems, it is sometimes necessary to adjust the bulk distribution (m_3) during the analysis. For instance, the ratio of cosolvent to solvent molecules in the bulk region might not be equal to the ratio of total molecules used in the simulation when cosolvent and solvent molecules can exchange with other molecules in the local vicinity of the protein. This is especially true for small systems and/or large values of Γ_{23}. In this case the value of the preferential binding should be adjusted so that,

$$\Gamma_{23}^{\infty} = \langle N_3 \rangle_2 - m_3 \langle N_1 \rangle_2 = \langle N_3 \rangle_2 - \frac{N_3^o - \langle N_3 \rangle_2}{N_1^o - \langle N_1 \rangle_2} \langle N_1 \rangle_2 \qquad 5.57$$

where the zero superscript indicates the total number of molecules in the system. Hence, as molecules move in or out of the local region the bulk ratio is correspondingly adjusted. This adjustment may appear small but is magnified by the fact that $\langle N_1 \rangle_2$ may be very large for biological solutes.

Finally, we note that the problem of choosing a reference energy/enthalpy for the energy fluctuations is not actually required for the analysis of computer simulation data. One can, in fact, determine the $\bar{H}_i$ values themselves. This has the advantage that the expressions for many of the properties described here become much simpler with this formulation. More details can be found in the literature (Jiao and Smith 2011).

5.12 Pseudo Chemical Potentials, Volumes and Enthalpies

As mentioned previously, the pseudo chemical potential and related properties can be used to analyze experimental data concerning a variety of processes – transfer free energies, and so on (Ben-Naim 2006). For many applications it is advantageous to use general fluctuation based expressions for the pseudo chemical potential, volume and enthalpy directly. From Equation 5.18 one finds,

$$\mu_{ik}^* = -\sum_j N_{ij}\mu_{jk} = -\sum_{j>1}(N_{ij} - m_j N_{i1})\mu_{jk} = -\sum_{j>1}\Gamma_{ij}\mu_{jk}$$

$$V_i^* = -\sum_j N_{ij}\bar{V}_j \qquad\qquad 5.58$$

$$H_i^{*,\text{Ex}} = H_i^* - H_i^{*,\text{o}} = k_\text{B}T^2(\alpha_p - \alpha_{p,i}^\text{o}) + F_i - \sum_j N_{ij}\bar{H}_j^{\text{Ex}}$$

after taking the appropriate derivatives. These expressions can also be used directly for the study of changes in chemical equilibria (see Section 11.2).

5.13 Summary and Future Directions

In the previous sections we have attempted to provide a relatively comprehensive summary of the theory and applications of a general fluctuation approach to understanding the behavior of solutions. It is quite clear that many of the above expressions for "normal" thermodynamic quantities appear to be quite complex. We are arguing that this complexity arises due to the constraints implied by the typical thermodynamic conditions of constant temperature, pressure, and composition. In contrast, a simpler picture of solutions emerges when one isolates the particle number and energy fluctuations corresponding to local regions of the solution, *i.e.* a microscopic constant volume and temperature region open to all matter exchange, for which the average pressure, energy, and number of particles corresponds to that of the experimental macroscopic conditions. It seems clear to us that attempting to understand the characteristics of solutions using properties of these open local regions is much simpler than focusing on the much larger bulk properties. Nevertheless, a simple picture of solution mixtures remains elusive.

At this stage it appears that a general FT of solutions, together with association and conformational equilibria, is essentially complete. The expressions provided here are exact and can be applied to any stable solution mixture containing any number of components at any concentration. In particular, we envision the application of FT to biological systems to continue, especially in the area of conformational equilibria. The natural coupling of FT and computer simulation holds great promise for future study. The ability to express changes in the equilibrium constant in terms of local properties has led to expressions for the enthalpy, heat capacity changes, *etc.*, which have previously been treated in a somewhat *ad hoc* manner, but are now available through a rigorous approach. Hopefully, this will provide significant insight into biological equilibria such as protein denaturation and peptide aggregation.

It is well known that many non-equilibrium effects that are observed in the presence of a perturbation can be related to fluctuations observed in the absence of that perturbation, *i.e.* to equilibrium properties. It seems that the local fluctuations described here could be of future use in understanding many perturbing effects on solutions. We anticipate future work in this area which will further advance the applications of a general fluctuation theory of solutions.

6. Experimental and Simulated Triplet and Quadruplet Particle Fluctuations and Probability Distribution Integrals in Multicomponent Solution Mixtures[*]

Abstract

The Kirkwood-Buff/Fluctuation Solution Theory (FST) is commonly used to relate particle-particle correlations in a grand canonical ensemble to thermodynamic properties in an isothermal-isobaric ensemble. This is potentially useful because the analysis of the local, fluctuating properties of an open system that is embedded within a much larger, closed system may allow for the elucidation of how the molecular interactions give rise to the bulk properties. Interestingly, we present here that, although the probability of observing a particular number of molecules is often assumed to follow a Gaussian distribution, relationships from FST show that the composition and pressure dependence of the particle-particle fluctuations both involve the third cumulants of the particle number probability distributions. Specifically, if the third cumulants were zero (meaning the distribution was symmetric), the particle-particle fluctuations would be independent of both composition and pressure. Clearly, this is not the case. To probe the form of the underlying, true distribution, we compare the results from computer simulations and an analysis of experimental, thermodynamic data. Specifically, we illustrate (1) how to obtain the uni- and bivariate particle number probability distributions obtained from computer simulations, (2) how to extract those same cumulants from an analysis of the bulk, thermodynamic experimental data, (3) how to calculate the same thermodynamic properties from simulation results, and (4) how to obtain experimental probability distributions using the maximum entropy principle and the experimentally extracted cumulants. Examples are provided for real systems such as water at a variety of temperatures and pressures and binary mixtures over their full composition ranges. A comparison is also made of the results for the real systems with their equivalent symmetric ideal systems.

6.1 Introduction

Fluctuation Solution Theory (FST) has provided an alternative view of many solution properties in terms of particle number fluctuations (Kirkwood and Buff 1951). The particle number fluctuations can also be related to integrals of the corresponding two body distribution functions between molecular pairs in order to provide a more physical picture of solutions. Here, we extend this type of approach to derive equations for higher order (triplet and quadruplet) fluctuations, and thereby integrals over the corresponding distribution functions, both of which can be obtained from existing experimental data. This is essentially the inverse of several expansions which have expressed higher order derivatives of thermodynamic properties in terms of higher order fluctuations and their distribution functions (Kirkwood and Buff 1951; Buff and Brout 1955; O'Connell 1971b). All expressions are exact. The results are compared to the same quantities obtained from simulation.

6.2 Theory

As is traditional with FST, we start with the equations of the grand canonical ensemble and use various thermodynamic transformations to provide properties corresponding to either semi-open osmotic systems, or fully closed isothermal isobaric systems. The resulting fluctuating properties can then be considered to represent local fluctuations within these systems. In this way, the theory is different from the study of bulk system fluctuations. The fluctuating quantities are then related to integrals over distribution functions.

The thermodynamic potential and partition function in the grand canonical ensemble are given by,

$$\beta pV = \ln \Xi \qquad \Xi = \sum_{\{N\}} \sum_i e^{\beta(\boldsymbol{\mu} \cdot \mathbf{N} - E_i)} \qquad\qquad 6.1$$

where $\beta = (k_B T)^{-1}$ and $\boldsymbol{\mu} \cdot \mathbf{N} = \mu_1 N_1 + \mu_2 N_2 \cdots$. The corresponding differential is,

$$d\beta pV = -Ud\beta + \beta pdV + \sum_\alpha N_\alpha d\beta\mu_\alpha \qquad\qquad 6.2$$

and a grand canonical ensemble average property (X) can be written,

$$\left\langle X \right\rangle = \Xi^{-1} \sum_{\{N\}} \sum_i X(\{N\}, V, E_i) e^{\beta(\boldsymbol{\mu} \cdot \mathbf{N} - E_i)} \qquad\qquad 6.3$$

This leads directly to the following relationships for the internal energy, the pressure, and the number of particles of each species,

$$U = -\left(\frac{\partial \ln \Xi}{\partial \beta}\right)_{V, \{\beta\mu\}} = \left\langle E \right\rangle \qquad\qquad \beta p = \left(\frac{\partial \ln \Xi}{\partial V}\right)_{\beta, \{\beta\mu\}} = \frac{\ln \Xi}{V}$$

$$\qquad\qquad 6.4$$

$$N_\alpha = \left(\frac{\partial \ln \Xi}{\partial \beta\mu_\alpha}\right)_{\beta, V, \{\beta\mu\}'} = \left\langle N_\alpha \right\rangle$$

for any multicomponent system (Hill 1956).

The ensemble averages are, therefore, a function of the thermodynamic variables associated with the grand canonical ensemble and hence one can write the following general differential,

$$d\langle X\rangle=\left(\frac{\partial\langle X\rangle}{\partial\beta}\right)_{V,\{\beta\mu\}}d\beta+\left(\frac{\partial\langle X\rangle}{\partial V}\right)_{\beta,\{\beta\mu\}}dV+\sum_{\alpha}\left(\frac{\partial\langle X\rangle}{\partial\beta\mu_{\alpha}}\right)_{\beta,V,\{\beta\mu\}'}d\beta\mu_{\alpha}$$

6.5

where X is a particle number or energy, or a particle number or energy fluctuation. If we restrict ourselves to the study of isothermal changes, the results relate to particle number fluctuations only (energy fluctuations will be included at a later date). The final expressions are considerably simplified by defining the following fluctuating quantities,

$$V\rho_{\alpha}=\langle N_{\alpha}\rangle \qquad VB_{\alpha\beta}=\langle\delta N_{\alpha}\delta N_{\beta}\rangle \qquad VC_{\alpha\beta\gamma}=\langle\delta N_{\alpha}\delta N_{\beta}\delta N_{\gamma}\rangle$$

6.6

$$VD_{\alpha\beta\gamma\delta}=\langle\delta N_{\alpha}\delta N_{\beta}\delta N_{\gamma}\delta N_{\delta}\rangle-\langle\delta N_{\alpha}\delta N_{\beta}\rangle\langle\delta N_{\gamma}\delta N_{\delta}\rangle-$$
$$\langle\delta N_{\alpha}\delta N_{\gamma}\rangle\langle\delta N_{\beta}\delta N_{\delta}\rangle-\langle\delta N_{\alpha}\delta N_{\delta}\rangle\langle\delta N_{\beta}\delta N_{\gamma}\rangle$$

where $\delta X=X-\langle X\rangle$ denotes a fluctuation in the value of X. The rhs of the above expressions correspond to the cumulants of the multivariate particle number probability distribution. Alternatively, they can be viewed as the mean, covariance, coskewness and excess cokurtosis of the same distribution. We note that there are several different definitions of skewness and excess kurtosis in the literature. The definition of skewness and excess kurtosis referred to here are those provided by $VC_{\alpha\beta\gamma}$ and $VD_{\alpha\beta\gamma\delta}$ expressions, respectively.

The required partial derivatives appearing in Equation 6.5 are given by Equations 6.3 and 6.4 such that,

$$\left(\frac{\partial\langle Y\rangle}{\partial V}\right)_{\beta,\{\beta\mu\}}=\frac{\langle Y\rangle}{V} \qquad\qquad \left(\frac{\partial\langle N_{\alpha}\rangle}{\partial\beta\mu_{\beta}}\right)_{\beta,V,\{\beta\mu\}'}=VB_{\alpha\beta}$$

6.7

$$\left(\frac{\partial\langle\delta N_{\alpha}\delta N_{\beta}\rangle}{\partial\beta\mu_{\gamma}}\right)_{\beta,V,\{\beta\mu\}'}=VC_{\alpha\beta\gamma} \qquad\qquad \left(\frac{\partial\langle\delta N_{\alpha}\delta N_{\beta}\delta N_{\gamma}\rangle}{\partial\beta\mu_{\delta}}\right)_{\beta,V,\{\beta\mu\}'}=VD_{\alpha\beta\gamma\delta}$$

where $Y=N_{\alpha}$ or $\delta N_{\alpha}\delta N_{\beta}$ or $\delta N_{\alpha}\delta N_{\beta}\delta N_{\gamma}$. Therefore, using the above expressions in Equation 6.5 and noting that $d\langle X\rangle/V=V^{-1}d\langle X\rangle-\langle X\rangle V^{-2}dV$ leads to the following differentials,

$$d\ln\Xi/V=\sum_{\alpha}\rho_{\alpha}d\beta\mu_{\alpha} \qquad\qquad d\rho_{\alpha}=\sum_{\beta}B_{\alpha\beta}d\beta\mu_{\beta}$$

6.8

$$dB_{\alpha\beta}=\sum_{\gamma}C_{\alpha\beta\gamma}d\beta\mu_{\gamma} \qquad\qquad dC_{\alpha\beta\gamma}=\sum_{\delta}D_{\alpha\beta\gamma\delta}d\beta\mu_{\delta}$$

The above expressions are valid for any multicomponent system in any ensemble at constant T. Taking derivatives with respect to pressure (indicated with a p superscript), or composition (indicated with an i superscript) provides a series of relationships. From the first expression one obtains the common relationships between the partial molar volumes and derivatives of the chemical potential (GD relation) at constant temperature and pressure,

$$1=\beta^{-1}\left(\frac{\partial\ln\Xi/V}{\partial p}\right)_{\beta,\{N\}}=\sum_{\alpha}\rho_{\alpha}\bar{V}_{\alpha} \qquad\qquad 0=\left(\frac{\partial\ln\Xi/V}{\ln N_{i}}\right)_{\beta,p,\{N\}'}=\sum_{\alpha}\rho_{\alpha}\mu_{\alpha i}$$

6.9

The second differential provides,

$$\rho_\alpha^p = \beta^{-1}\left(\frac{\partial \rho_\alpha}{\partial p}\right)_{\beta,\{N\}} = \sum_\beta B_{\alpha\beta}\bar{V}_\beta \qquad\qquad \rho_\alpha^i = \left(\frac{\partial \rho_\alpha}{\partial \ln N_i}\right)_{\beta,p,\{N\}'} = \sum_\beta B_{\alpha\beta}\mu_{\beta i} \qquad 6.10$$

and leads to the traditional KB theory of solutions (Kirkwood and Buff 1951), given that $\rho_\alpha^p = \rho_\alpha k_B T \kappa_T$ and $\rho_\alpha^i = \rho_\alpha(\delta_{\alpha i} - \phi_i)$, and one can thereby relate the two body fluctuations (B's) to the relevant thermodynamic properties,

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_{\beta,\{N\}} \qquad \bar{V}_\alpha = \left(\frac{\partial V}{\partial N_\alpha}\right)_{\beta,p,\{N\}'} = \left(\frac{\partial \mu_\alpha}{\partial p}\right)_{\beta,\{N\}} \qquad \mu_{\alpha\beta} = \left(\frac{\partial \mu_\alpha}{\partial \ln N_\beta}\right)_{\beta,p,\{N\}'} \qquad 6.11$$

as done previously. The final two differentials provide,

$$B_{\alpha\beta}^p = \beta^{-1}\left(\frac{\partial B_{\alpha\beta}}{\partial p}\right)_{\beta,\{N\}} = \sum_\gamma C_{\alpha\beta\gamma}\bar{V}_\gamma \qquad\qquad B_{\alpha\beta}^i = \left(\frac{\partial B_{\alpha\beta}}{\partial \ln N_i}\right)_{\beta,p,\{N\}'} = \sum_\gamma C_{\alpha\beta\gamma}\mu_{\gamma i}$$

$$6.12$$

$$C_{\alpha\beta\gamma}^p = \beta^{-1}\left(\frac{\partial C_{\alpha\beta\gamma}}{\partial p}\right)_{\beta,\{N\}} = \sum_\delta D_{\alpha\beta\gamma\delta}\bar{V}_\delta \qquad\qquad C_{\alpha\beta\gamma}^i = \left(\frac{\partial C_{\alpha\beta\gamma}}{\partial \ln N_i}\right)_{\beta,p,\{N\}'} = \sum_\delta D_{\alpha\beta\gamma\delta}\mu_{\delta i}$$

and describe the pressure and composition dependence of the two and three body particle number fluctuations or distributions. It is clear from these expressions that if the multivariate particle number probability distribution for a local region of the solution was Gaussian in nature, then the C's and D's would be zero and the B's would therefore be independent of pressure and composition (see Equation 6.6). This is clearly not the case, as noted previously (Greene and Callen 1951). We also note that the Central Limit Theorem does not apply to the particle number fluctuations, with an exception for the case of an infinitely dilute solute, as they are not random in nature but driven by the molecular interactions.

Our aim here is to invert this whole procedure to express the B's, C's and D's, together with the associated molecular distribution functions, in terms of experimental data. This is analogous to the standard KB inversion procedure (Ben-Naim 1977). To do this we define the following matrix,

$$\mathbf{Z} = \begin{bmatrix} \bar{V}_1 & \bar{V}_2 & \cdots & \bar{V}_n \\ \mu_{12} & \mu_{22} & \cdots & \mu_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{1n} & \mu_{2n} & \cdots & \mu_{nn} \end{bmatrix} \qquad 6.13$$

which allows one to express the relationships in Equation 6.12 in matrix form and then invert to provide,

$$\boldsymbol{\rho} = \mathbf{Z}^{-1}\mathbf{1}' \qquad \mathbf{B}_\alpha = \mathbf{Z}^{-1}\boldsymbol{\rho}_\alpha' \qquad \mathbf{C}_{\alpha\beta} = \mathbf{Z}^{-1}\mathbf{B}_{\alpha\beta}' \qquad \mathbf{D}_{\alpha\beta\gamma} = \mathbf{Z}^{-1}\mathbf{C}_{\alpha\beta\gamma}' \qquad 6.14$$

where the vectors on the lhs are the required fluctuating quantities, $\boldsymbol{\rho} = (\rho_1, \rho_2, \cdots)$, $\mathbf{B}_\alpha = (B_{\alpha 1}, B_{\alpha 2}, \cdots)$, $\mathbf{C}_{\alpha\beta} = (C_{\alpha\beta 1}, C_{\alpha\beta 2}, \cdots)$, and $\mathbf{D}_{\alpha\beta\gamma} = (D_{\alpha\beta\gamma 1}, D_{\alpha\beta\gamma 2}, \cdots)$, while the vectors on the rhs contain only experimental properties and are given by, $\mathbf{1}' = (1, 0, 0, \cdots)$, $\boldsymbol{\rho}_\alpha' = (\rho_\alpha^p, \rho_\alpha^2, \rho_\alpha^3, \cdots)$, $\mathbf{B}_{\alpha\beta}' = (B_{\alpha\beta}^p, B_{\alpha\beta}^2, B_{\alpha\beta}^3, \cdots)$, and $\mathbf{C}_{\alpha\beta\gamma}' = (C_{\alpha\beta\gamma}^p, C_{\alpha\beta\gamma}^2, C_{\alpha\beta\gamma}^3, \cdots)$. The above expressions could be made more general by using multidimensional array algebra, but we prefer the simpler (more transparent) expressions provided here. The above equations can therefore be solved to provide expressions for the fluctuating quantities in terms of experimental data. The explicit expressions are quite cumbersome even for binary

114

solutions. However, a general matrix formulation is possible which involves just a single matrix of experimental properties and its derivatives (see Section 6.5.1).

Before discussing some systems of common interest we note that the fluctuating quantities can be related to the corresponding distribution functions *via* the expressions (see Section 6.5),

$$B_{\alpha\beta} = \rho_\alpha(\delta_{\alpha\beta} + \rho_\beta G_{\alpha\beta})$$

$$C_{\alpha\beta\gamma} = \rho_\alpha(\delta_{\alpha\beta}\delta_{\beta\gamma} + \delta_{\beta\gamma}\rho_\beta G_{\alpha\beta} + \delta_{\alpha\beta}\rho_\gamma G_{\alpha\gamma} + \delta_{\alpha\gamma}\rho_\beta G_{\beta\gamma} + \rho_\beta\rho_\gamma G_{\alpha\beta\gamma})$$

$$
\begin{aligned}
D_{\alpha\beta\gamma\delta} = \rho_\alpha (&\delta_{\alpha\beta}\delta_{\alpha\gamma}\delta_{\alpha\delta} + \delta_{\alpha\gamma}\delta_{\alpha\delta}\rho_\beta G_{\alpha\beta} + \delta_{\alpha\delta}\delta_{\beta\gamma}\rho_\gamma G_{\alpha\gamma} + \delta_{\alpha\gamma}\delta_{\beta\delta}\rho_\delta G_{\alpha\delta} + \\
&\delta_{\beta\gamma}\delta_{\gamma\delta}\rho_\delta G_{\alpha\delta} + \delta_{\alpha\beta}\delta_{\beta\delta}\rho_\gamma G_{\beta\gamma} + \delta_{\alpha\beta}\delta_{\gamma\delta}\rho_\delta G_{\beta\delta} + \delta_{\alpha\beta}\delta_{\beta\gamma}\rho_\delta G_{\gamma\delta} + \\
&\delta_{\alpha\delta}\rho_\beta\rho_\gamma G_{\alpha\beta\gamma} + \delta_{\alpha\gamma}\rho_\beta\rho_\delta G_{\alpha\beta\delta} + \delta_{\gamma\delta}\rho_\beta\rho_\delta G_{\alpha\beta\delta} + \\
&\delta_{\beta\gamma}\rho_\gamma\rho_\delta G_{\alpha\gamma\delta} + \delta_{\beta\delta}\rho_\gamma\rho_\delta G_{\alpha\gamma\delta} + \delta_{\alpha\beta}\rho_\gamma\rho_\delta G_{\beta\gamma\delta} + \rho_\beta\rho_\gamma\rho_\delta G_{\alpha\beta\gamma\delta})
\end{aligned}
$$

6.15

which involve the following integrals over the *n*-body distribution functions $g_{\alpha\beta...}^{(n)}(r_1, r_2, ...)$,

$$G_{\alpha\beta} = V^{-1}\int\left[g_{\alpha\beta}^{(2)} - 1\right]dr_1 dr_2$$

$$G_{\alpha\beta\gamma} = V^{-1}\int\left[g_{\alpha\beta\gamma}^{(3)} - 1 - (g_{\alpha\beta}^{(2)} - 1) - (g_{\alpha\gamma}^{(2)} - 1) - (g_{\beta\gamma}^{(2)} - 1)\right]dr_1 dr_2 dr_3$$

6.16

$$G_{\alpha\beta\gamma\delta} = V^{-1}\int\left[\begin{array}{l}g_{\alpha\beta\gamma\delta}^{(4)} - 1 - (g_{\alpha\beta\gamma}^{(3)} - 1) - (g_{\alpha\beta\delta}^{(3)} - 1) - (g_{\alpha\gamma\delta}^{(3)} - 1) - (g_{\beta\gamma\delta}^{(3)} - 1) \\ -(g_{\alpha\beta}^{(2)} - 1)(g_{\gamma\delta}^{(2)} - 1) - (g_{\alpha\gamma}^{(2)} - 1)(g_{\beta\delta}^{(2)} - 1) - (g_{\alpha\delta}^{(2)} - 1)(g_{\beta\gamma}^{(2)} - 1) \\ +(g_{\alpha\beta}^{(2)} - 1) + (g_{\alpha\gamma}^{(2)} - 1) + (g_{\alpha\delta}^{(2)} - 1) + (g_{\beta\gamma}^{(2)} - 1) + (g_{\beta\delta}^{(2)} - 1) + (g_{\gamma\delta}^{(2)} - 1)\end{array}\right]dr_1 dr_2 dr_3 dr_4$$

Hence, one can obtain the fluctuating quantities, and thereby the integrals over the two, three and four body correlation functions, in terms of experimental data. The integrals over the distribution functions correspond to a particular solution composition and are obtained after averaging over all the other molecules in the system. Fluctuation based expressions for the *G*'s are provided in Section 6.5.

The above integrals and probability distributions are often used to provide insight into the "structure" of liquids and solution mixtures. The quantity $\rho_\beta G_{\alpha\beta}$ provides the excess number of β molecules in a given local volume of solution surrounding a central α molecule over and above the number of β molecules found in the same volume of bulk solution. Hence, a significantly positive value of $G_{\alpha\beta}$ suggests an overall net affinity (attraction) between the two types of molecules. A similar picture should emerge for the $G_{\alpha\beta\gamma}$ and $G_{\alpha\beta\gamma\delta}$ values.

6.3 Results

6.3.1 One Component Systems

Some of the resulting expressions for pure systems have appeared previously. For instance, it is well known that the pressure dependence of the radial distribution function is related to the triplet distribution (Egelstaff 1973). However,

we found no relevant quantitative experimental data concerning the local fluctuations in the literature. For single component systems the relationships between the fluctuating quantities and the corresponding distribution functions are provided by,

$$B_{11} = \rho_1(1+\rho_1 G_{11})$$

$$C_{111} = \rho_1(1+3\rho_1 G_{11} + \rho_1^2 G_{111}) \qquad 6.17$$

$$D_{1111} = \rho_1(1+7\rho_1 G_{11} + 6\rho_1^2 G_{111} + \rho_1^3 G_{1111})$$

with,

$$G_{11} = V^{-1}\int \left[g_{11}^{(2)} -1\right] dr_1 dr_2$$

$$G_{111} = V^{-1}\int \left[g_{111}^{(3)} -1 -3(g_{11}^{(2)}-1)\right] dr_1 dr_2 dr_3 \qquad 6.18$$

$$G_{1111} = V^{-1}\int \left[g_{1111}^{(4)} -1 -4(g_{111}^{(3)}-1) -3(g_{11}^{(2)}-1)(g_{11}^{(2)}-1) +6(g_{11}^{(2)}-1)\right] dr_1 dr_2 dr_3 dr_4$$

which can also be expressed in terms of particle number fluctuations,

$$\rho_1^2 V G_{11} = \langle \delta N_1 \delta N_1\rangle - \langle N_1\rangle$$

$$\rho_1^3 V G_{111} = \langle \delta N_1 \delta N_1 \delta N_1\rangle - 3\langle \delta N_1 \delta N_1\rangle + 2\langle N_1\rangle \qquad 6.19$$

$$\rho_1^4 V G_{1111} = \langle \delta N_1 \delta N_1 \delta N_1 \delta N_1\rangle - 3\langle \delta N_1 \delta N_1\rangle^2 - 6\langle \delta N_1 \delta N_1 \delta N_1\rangle + 11\langle \delta N_1 \delta N_1\rangle - 6\langle N_1\rangle$$

The fluctuating quantities are related to derivatives of the molar volume (V_1), or the solution (number) density, as indicated in Equation 6.12 and are given by,

$$B_{11} = -\frac{k_B T}{V_1^3} V_1'$$

$$C_{111} = \frac{(k_B T)^2}{V_1^5}[3(V_1')^2 - V_1 V_1''] \qquad 6.20$$

$$D_{1111} = \frac{(k_B T)^3}{V_1^7}[-15(V_1')^3 + 10 V_1 V_1' V_1'' - V_1^2 V_1''']$$

where the prime indicates a derivative with respect to pressure. The corresponding *G*'s can then also be obtained.
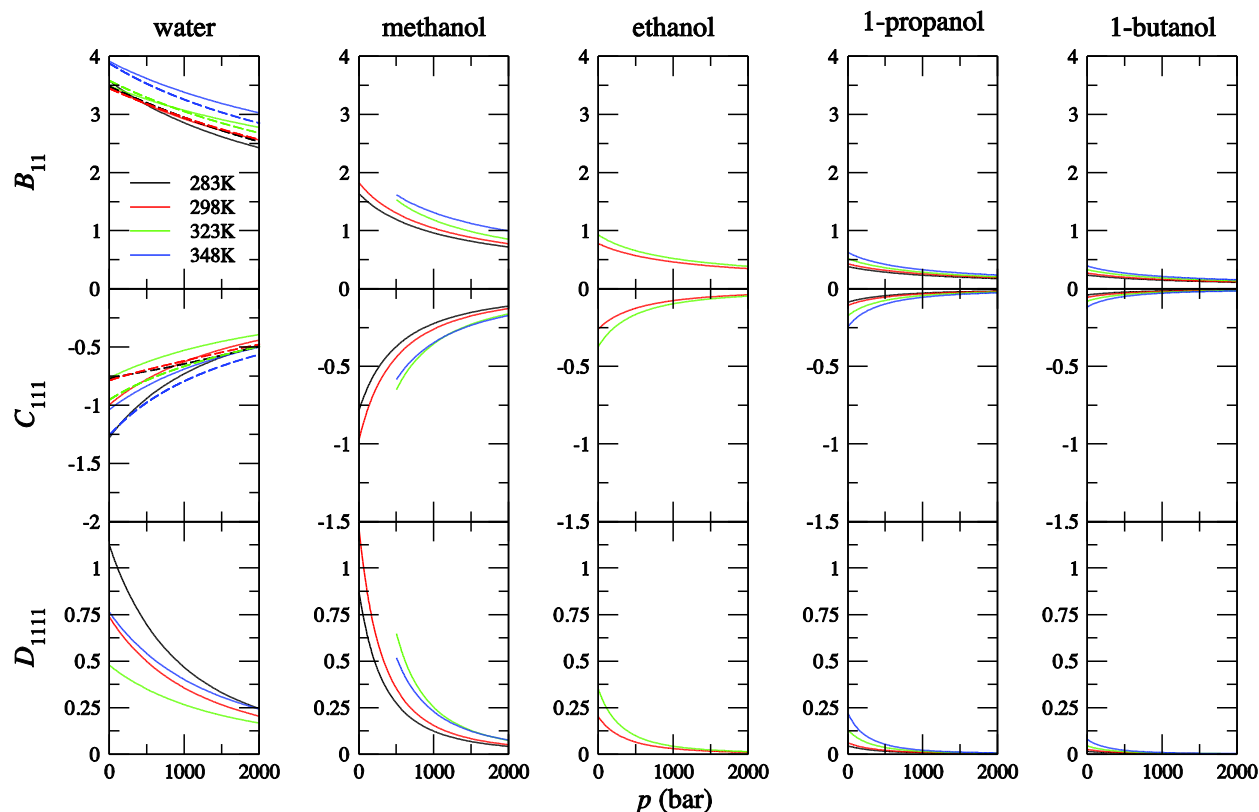
Figure 6.1 Local fluctuations (in *M*) as a function of pressure and temperature for pure water and several pure linear alcohols. Experimental density data were taken from the literature (Kubota *et al.* 1987) and then fitted to the Tait Equation of State (EOS), $V_1 = V_{1,o}(1\text{-}C\ln[B+p]/(B+p_o))$, where $p_o = 1$ bar (solid lines). The low pressure–high temperature data for methanol provided a poor fit in this region so these points have been dropped. We have also included the data for water obtained from the Wagner & Pruss EOS (dashed lines) for comparison (Wagner and Pruss 2002).

The results for pure water and some linear alcohols as a function of pressure and temperature are provided in Figure 6.1 and Figure 6.2. Some initial observations from Figure 6.1 include the following. The moments alternate in sign. As expected, the two body fluctuations (B_{11}) generally increase with increasing temperature and decreasing pressure for the alcohols. Water displays somewhat unusual behavior in that the number fluctuations do not vary systematically with temperature, although the corresponding G's do. The skewness of the distributions is always negative (a depletion of particles in the sample volume is favored over an excess) and the excess kurtosis is always positive (the actual distribution is more peaked than a normal distribution). The distributions tend to the normal distribution (B_{11} is constant, $C_{111} = D_{1111} = 0$) as the pressure increases and/or the alcohol chain length is increased. Presumably, this indicates a weaker correlation between alcohol molecules as the role of hydrogen bonding decreases. Some initial observations from Figure 6.2 include the following. The integrals alternate in sign and increase in magnitude as the pressure decreases and/or temperature increases. In contrast to the fluctuations, they tend to increase in magnitude as the alcohol chain increases.
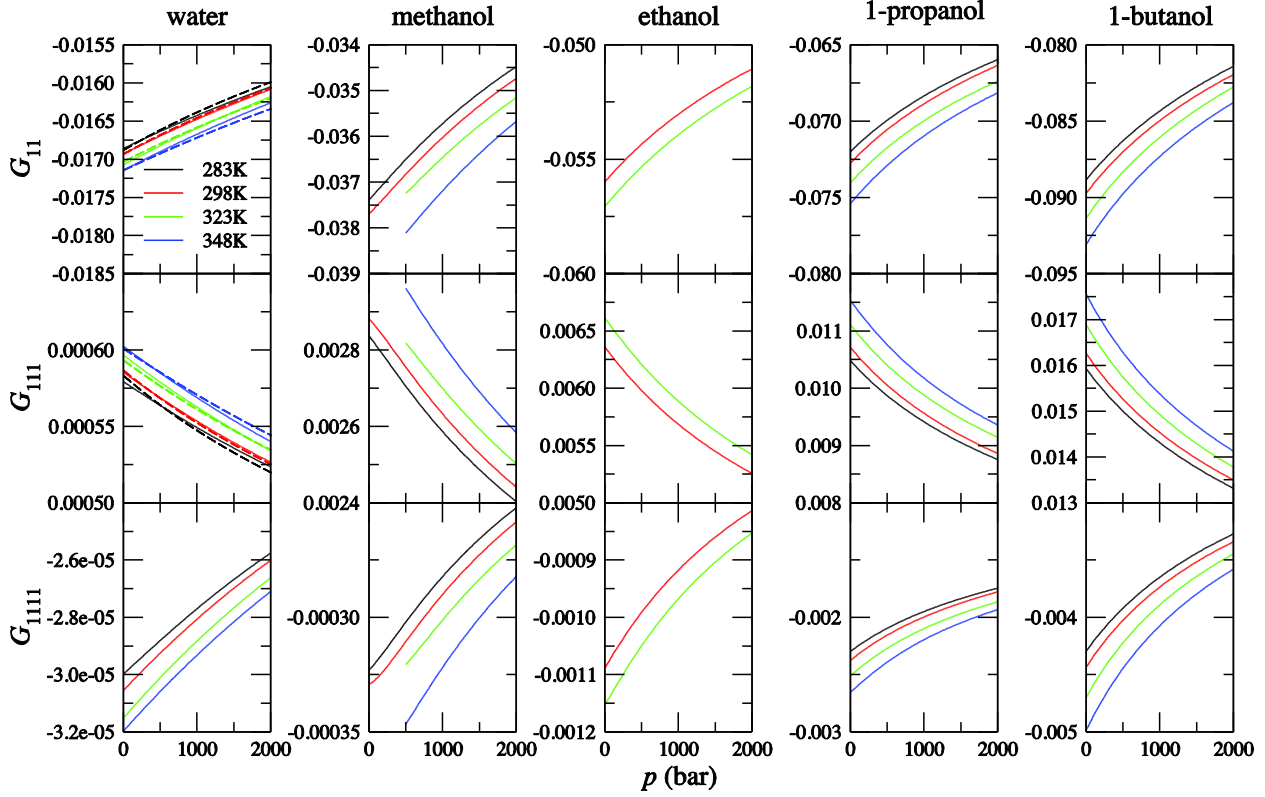
Figure 6.2 Integrals over the pair (G_{22} in M^{-1}), triplet (G_{222} in M^{-2}), and quadruplet (G_{2222} in M^{-3}) probability distributions as a function of pressure and temperature for pure water and several pure linear alcohols. See Figure 6.1 for experimental details.

In Figure 6.3 we provide the values of the two and three body fluctuations and distribution function integrals for liquid water over a larger region of the phase diagram using the Equation of State (EOS) provided by Wagner & Pruss. This analysis is also included in Figure 6.1 and Figure 6.2 and indicates that the fluctuations are somewhat sensitive to the EOS used, while the G's are fairly consistent between the data sets. Figure 6.3 again illustrates that the particle fluctuations tend to Gaussian behavior at low temperatures and high pressures. As expected, B_{11} and G_{11} tend to positive infinity as one approaches the critical point, whereas the C_{111} and G_{111} values approach negative infinity in this region.

We know of no reason why a similar analysis could not be performed for pure gases, or mixtures of gases, and pure solids. For solids we would expect the fluctuations to be very close to Gaussian in nature. Gases would approach ideal behavior at low pressure and high temperature, where $B_{\alpha\alpha} = C_{\alpha\alpha\alpha} = D_{\alpha\alpha\alpha\alpha} = \rho_\alpha$ and all the G's and other B's, C's and D's are zero, which corresponds to a series of independent Poisson distributions for the particle number fluctuations. It should be noted that the Poisson distribution is often considered to approach a Gaussian distribution for large sample sizes ($\langle N_\alpha \rangle$ in this case). This is true when characterizing the distribution using the standardized skewness and excess kurtosis, but not when using the non-standardized definitions presented in Equation 6.6.

118
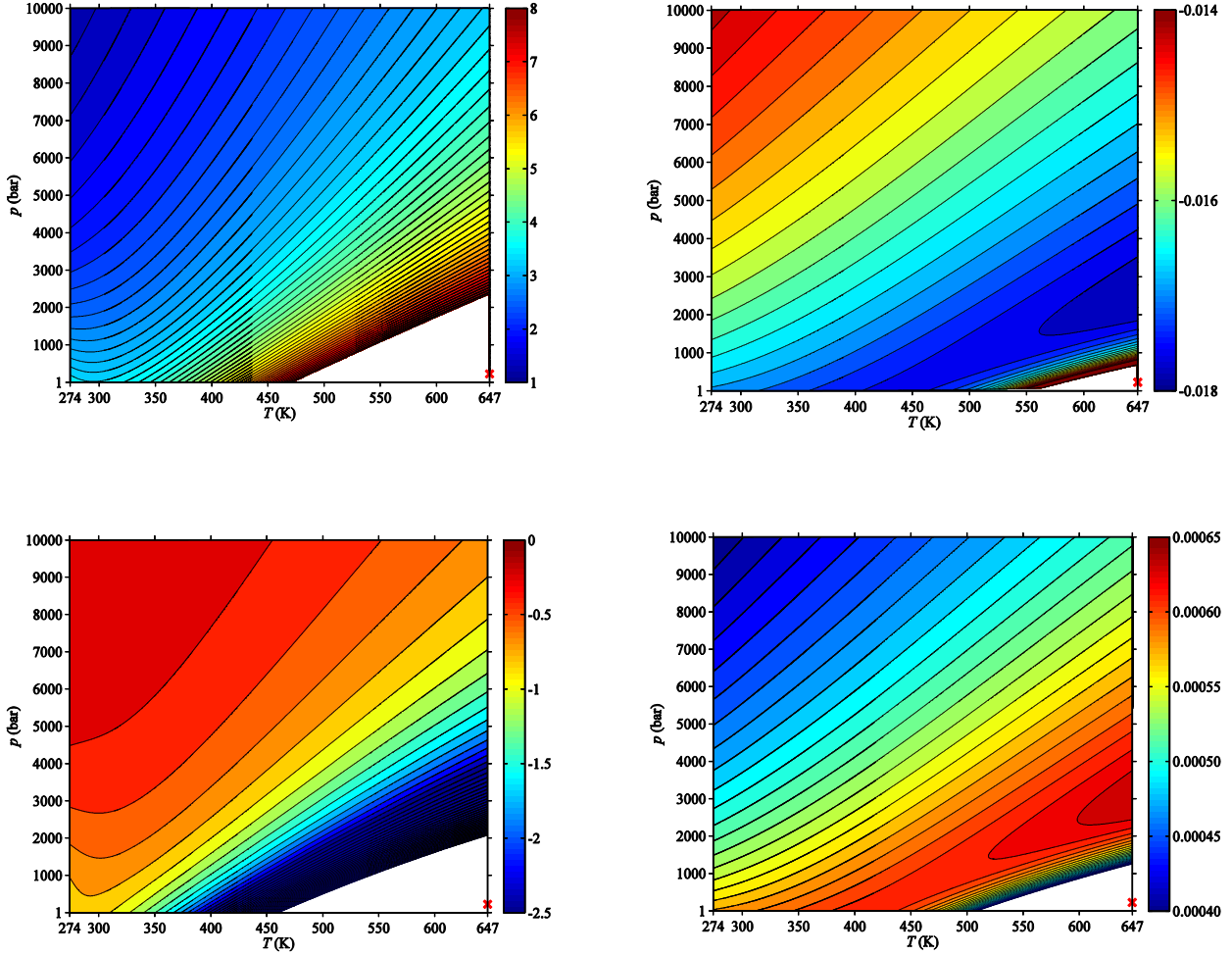
Figure 6.3 Contour plots for water obtained from the Wagner & Pruss EOS (Wagner and Pruss 2002). The fluctuations are displayed for B_{11} (top left) and C_{111} (bottom left) in units of M, while the pair G_{11} (top right) and triplet G_{111} (bottom right) integrals are displayed in units of M^{-1} and M^{-2}, respectively. The critical point is indicated with a small red cross. Data in the vicinity of the critical point have been removed to allow for better contouring. We are still in the process of obtaining the quadruplet terms.

The pressure dependence of the G's can be obtained by taking derivatives of Equation 6.17 and comparing with Equation 6.12. One finds,

$$k_B T \left(\frac{\partial G_{11}}{\partial p} \right)_{\beta, N_1} = G_{111} - 2G_{11}^2 \qquad k_B T \left(\frac{\partial G_{111}}{\partial p} \right)_{\beta, N_1} = G_{1111} - 3G_{111}G_{11} \qquad 6.21$$

The first expression has appeared before (Egelstaff 1973; Gubbins *et al.* 1978).

6.3.2 Closed Two Component Systems

Binary solutions are the most common system of interest. The inverse of the **Z** matrix for binary solutions takes the form,

$$\mathbf{Z}^{-1} = \begin{bmatrix} \rho_1 & \dfrac{\phi_2}{\mu_{12}} \\ \rho_2 & \dfrac{\phi_1}{\mu_{22}} \end{bmatrix}$$

6.22

The use of this matrix in Equation 6.14 and analysis of the results leads to the following general expressions for the fluctuating quantities,

$$B_{\alpha\beta} = \rho_\beta \rho_\alpha^p + \frac{1-\phi_\beta}{\mu_{\beta\beta}} \rho_\alpha^\beta$$

$$C_{\alpha\beta\gamma} = \rho_\gamma B_{\alpha\beta}^p + \frac{1-\phi_\gamma}{\mu_{\gamma\gamma}} B_{\alpha\beta}^\gamma$$

6.23

$$D_{\alpha\beta\gamma\delta} = \rho_\delta C_{\alpha\beta\gamma}^p + \frac{1-\phi_\delta}{\mu_{\delta\delta}} C_{\alpha\beta\gamma}^\delta$$

The relationships can be solved sequentially or, alternatively, using the matrix approach provided in Section 6.5.1. For binary mixtures there are 3 unique B's, 4 unique C's and 5 unique D's.

The results for methanol (2) and water (1) mixtures are shown in Figure 6.4 and Figure 6.5. The fluctuating quantities are provided in Figure 6.4. It is important to test the degree to which the experimental data provide reproducible values of the cumulants. There are many determinations of G^{Ex} (which we assume contains the major source of error in the present analysis) for this system. Some examples are included in Figure 6.4 and Figure 6.5. There are clearly quantitative differences between the experimental data sets – especially for those using different fitting functions. However, the trends appear to be reasonably clear, and consistent results are obtained for the B's and C's for most compositions, while the D's clearly show the largest variations between data sets. It does appear that the present type of analysis can provide reasonably reliable data for the B's and C's, while care would have to be used when drawing conclusions concerning the D's. A similar story is found for the corresponding G's provided in Figure 6.5. Clearly, the G's also become less reliable as the relevant species concentration approaches infinite dilution.

The expressions provided in Equation 6.12 also indicate the conditions for which there is a maximum or minimum in the B's or C's as a function of composition, as is commonly observed. Using these expressions followed by application of the GD expression provides,

$$B_{\alpha\beta}^i = C_{\alpha\beta1}\mu_{1i} + C_{\alpha\beta2}\mu_{2i} = (C_{\alpha\beta2} - m_2 C_{\alpha\beta1})\mu_{2i}$$

6.24

where $m_2 = \rho_2/\rho_1$. The chemical potential derivative has to be non-zero for stable solutions and hence the derivative of $B_{\alpha\beta}$ can only be zero when,

$$\langle \delta n_\alpha \delta n_\beta \delta n_2 \rangle = \langle \delta n_\alpha \delta n_\beta \delta n_1 \rangle$$

6.25

120
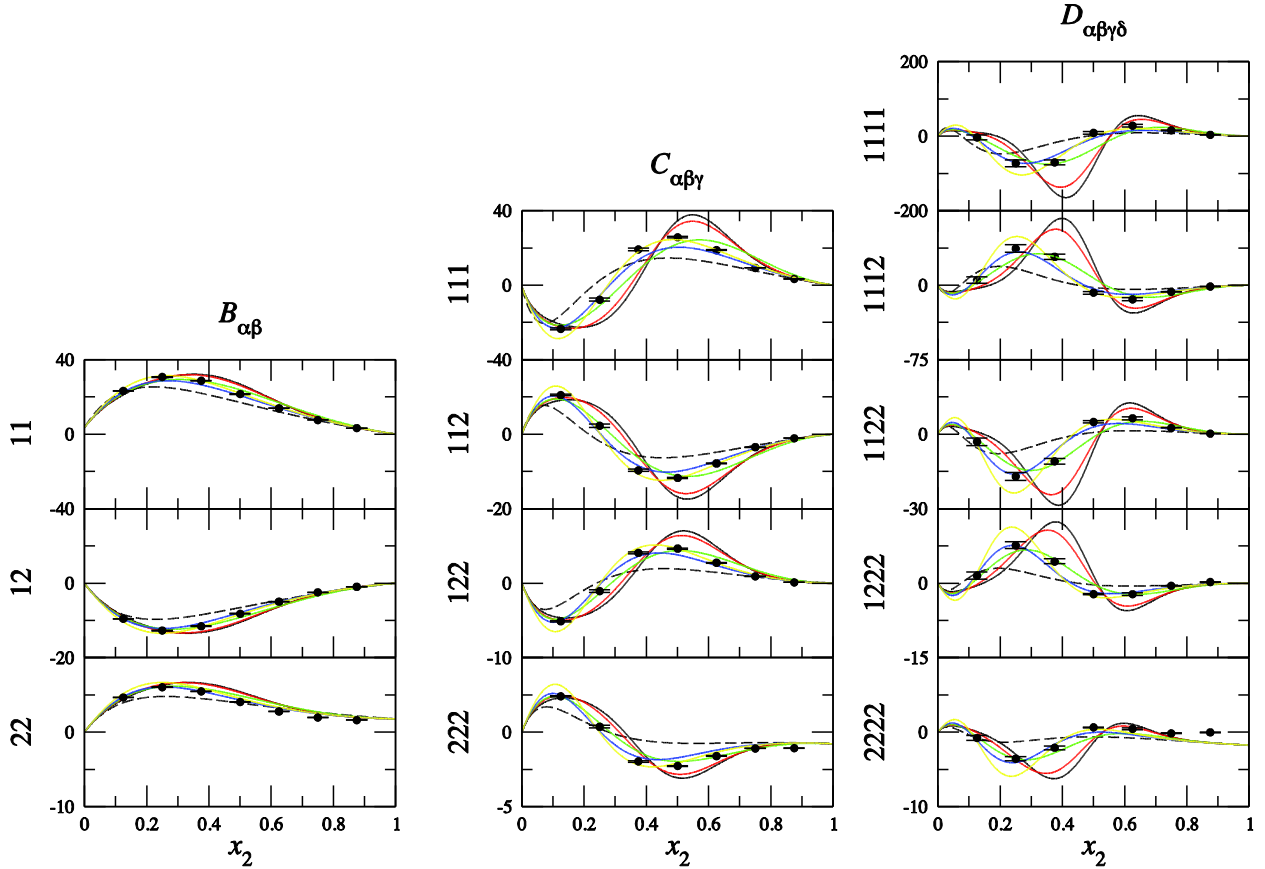
Figure 6.4 Local fluctuations (in *M*) as a function of solute mole fraction obtained from experimental data on methanol (2) and water (1) mixtures at 298 K and 1 bar. Dashed lines represent the corresponding SI results, while symbols correspond to the simulated data. For this system, the condition $\phi_2 = \phi_1$ occurs when $x_2 = 0.31$, see Equation 6.26 and subsequent discussion. Experimental data were taken from the literature. All analysis was performed with the same density data (Douheret *et al.* 1989; Papaioannou and Panayiotou 1995; Wagner and Kretzschmar 2008). Several data sets and fitting functions for G^{Ex} were explored: Black lines – Redlich-Kister (RK) (Butler *et al.* 1933); Red lines – modified RK (Hu *et al.* 2003); Green lines – Wilson (Soujanya *et al.* 2010); Blue lines – Wilson (Gmehling *et al.* 1991); and Yellow lines – Wilson (Gmehling *et al.* 1991).

where we have made the substitution, $n_i = N_i/\langle N_i \rangle$. Hence, the maximum or minimum in $B_{\alpha\beta}$ occurs when the corresponding (reduced) triplet fluctuations are equal for both components. Similar relationships can be obtained for maxima or minima for the *C*'s in terms of the *D*'s. Conditions that lead to a maximum or minimum in the *G*'s can also be developed but are more complicated and also appear less informative.

A somewhat reasonable physical picture of the two body (and higher) particle number fluctuations can be developed for binary systems. In Figure 6.4 the experimental B_{11} and B_{22} values appear correlated, while the B_{12} value appears to be anti-correlated with the other *B*'s. In pure solutions of component 1, the B_{11} values describe the fluctuations of N_1 particles in and out of a fixed local volume of solution. The magnitude of these fluctuations is determined by the size of the region and the compressibility of the pure solution, which is normally small under ambient conditions. On the other hand, when component 1 appears at low dilution the value of B_{11} tends to zero simply because there are very few N_1 molecules in a given region of solution. For intermediate compositions one can observe particle number
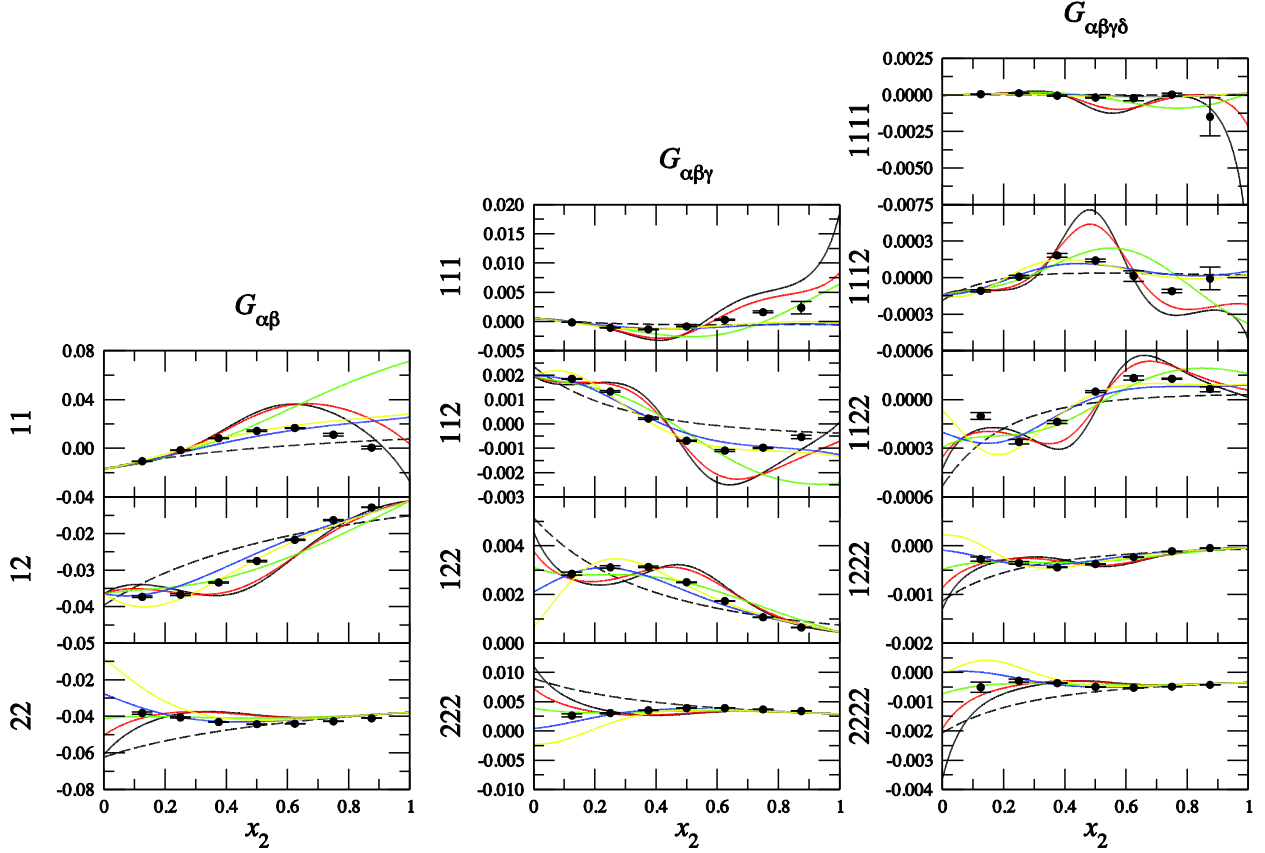
Figure 6.5 Integrals over the pair ($G_{\alpha\beta}$ in M^{-1}), triplet ($G_{\alpha\beta\gamma}$ in M^{-2}), and quadruplet ($G_{\alpha\beta\gamma\delta}$ in M^{-3}) probability distributions as a function of solute mole fraction obtained from experimental data on methanol (2) and water (1) mixtures at 298 K and 1 bar. Dashed lines represent the corresponding SI results, while symbols correspond to the simulated systems. See Figure 6.4 for experimental details.

fluctuations due to the compressibility of the solution, and also due to exchange of 1 and 2 molecules such that the volume of solution remains essentially constant for all compositions (Mazo *et al.* 2009),

$$\delta N_1 \bar{V}_1 + \delta N_2 \bar{V}_2 \approx 0 \qquad\qquad 6.26$$

Hence, the magnitude of the fluctuations can increase due to this additional mechanism. This also helps to explain why the B_{11} (and B_{22}) fluctuations are essentially anti-correlated with the B_{12} values. This correlation/anti-correlation relationship also persists for the higher order fluctuations involving pairs of C's and D's that only differ by a single species index (C_{111} and C_{112}, for example). Furthermore, if one multiplies Equation 6.26 by $\delta n_\alpha \delta n_\beta$ and then takes the ensemble average one finds the approximate relationship $\langle \delta n_\alpha \delta n_\beta \delta n_2 \rangle \phi_2 + \langle \delta n_\alpha \delta n_\beta \delta n_1 \rangle \phi_1 \approx 0$, which is true when $B_{\alpha\beta}^p$ is small. Using the condition given by Equation 6.25 implies that (at the maxima or minima) not only are the relevant fluctuating quantities equal, but they are both close to zero. This type of behavior is observed in Figure 6.4. We have also observed that the maxima and minima in the B's appear close to the composition for which $\phi_1 \approx \phi_2$ ($x_2 = 0.31$ in Figure 6.4 and $x_2 = 0.5$ in Figure 6.9), although it is not clear from the previous expressions why this should be the case.

6.3.3 Closed Three Component Systems

The explicit expressions for three component systems are significantly more complicated and we only provide some of the basic relationships. However, the expressions can be solved numerically using the approach provided in Section 6.5.1. The inverse of the $\mathbf{Z}$ matrix is given by,

$$\mathbf{Z}^{-1} = \begin{bmatrix} \rho_1 & \dfrac{\bar{V}_3\mu_{23} - \bar{V}_2\mu_{33}}{|Z|} & \dfrac{\bar{V}_2\mu_{32} - \bar{V}_3\mu_{22}}{|Z|} \\[3mm] \rho_2 & \dfrac{\bar{V}_1\mu_{33} - \bar{V}_3\mu_{13}}{|Z|} & \dfrac{\bar{V}_3\mu_{12} - \bar{V}_1\mu_{32}}{|Z|} \\[3mm] \rho_3 & \dfrac{\bar{V}_2\mu_{13} - \bar{V}_1\mu_{23}}{|Z|} & \dfrac{\bar{V}_1\mu_{22} - \bar{V}_2\mu_{12}}{|Z|} \end{bmatrix} \qquad 6.27$$

The use of the matrix in Equation 6.14 followed by significant rearrangement using the GD expressions for ternary systems leads to the following relationships,

$$B_{\alpha\beta} = \rho_\beta \rho_\alpha^p - \sum_{k \neq \beta} K(k,\beta)\rho_\alpha^k$$

$$C_{\alpha\beta\gamma} = \rho_\gamma B_{\alpha\beta}^p - \sum_{k \neq \gamma} K(k,\gamma)B_{\alpha\beta}^k$$

$$D_{\alpha\beta\gamma\delta} = \rho_\delta C_{\alpha\beta\gamma}^p - \sum_{k \neq \delta} K(k,\delta)C_{\alpha\beta\gamma}^k$$

$$K(i,j) = \frac{(1-\phi_j)(\mu_{ij} + \mu_{ii}) + \rho_j \bar{V}_i(\mu_{ji} + \mu_{jj})}{\mu_{ii}\mu_{jj} - \mu_{ij}\mu_{ji}}$$

6.28

In principle, these expressions can be solved by substitution. In practice, it will be much easier to use the matrix formulation provided in Section 6.5.1.

6.3.4 Symmetric Ideal Systems

Symmetric Ideal (SI) solutions provide an interesting reference solution for the investigation of many thermodynamic properties of closed systems. SI solutions are characterized by $d\beta\mu_\alpha = d\ln x_\alpha$, with a zero excess volume of mixing and an ideal compressibility. The higher order correlations also require knowledge of the pure densities as a function of pressure. Expressions for any number of components are possible in this case. The inverse of the $\mathbf{Z}$ matrix becomes,

$$\mathbf{Z}^{-1} = \begin{bmatrix} \rho_1 & -1+\rho_1(V_1^o - V_2^o) & -1+\rho_1(V_1^o - V_3^o) & \cdots & -1+\rho_1(V_1^o - V_n^o) \\ \rho_2 & 1+\rho_2(V_1^o - V_2^o) & \rho_2(V_1^o - V_3^o) & \cdots & \rho_2(V_1^o - V_n^o) \\ \rho_3 & \rho_3(V_1^o - V_2^o) & 1+\rho_3(V_1^o - V_3^o) & \cdots & \rho_3(V_1^o - V_n^o) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_n & \rho_n(V_1^o - V_2^o) & \rho_n(V_1^o - V_3^o) & \cdots & 1+\rho_n(V_1^o - V_n^o) \end{bmatrix} \qquad 6.29$$
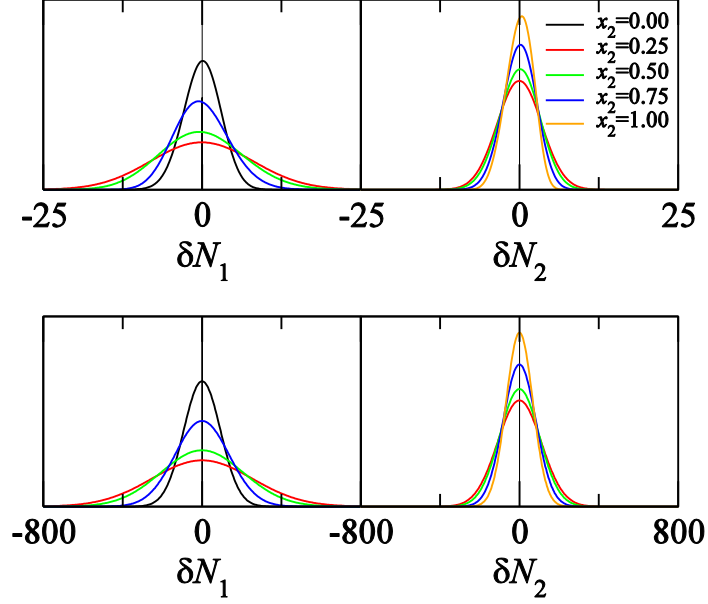
Figure 6.6 Particle number probability distributions for SI mixtures of methanol (2) and water (1) at 298 K and 1 bar as a function of methanol mole fraction. Experimental density data for the pure liquids was taken from the literature (see legend to Figure 6.4). The distributions were generated using the maximum entropy method ensuring that the first four cumulants of the distribution are correct. The distributions are provided for two sample spherical volumes of the solution with radii of 1 nm (upper) and 10 nm (lower). The corresponding probability distributions obtained for real mixtures of methanol and water also display a similar behavior.

where V_i° is the molar volume of pure component i. This leads to the following general results,

$$B_{\alpha\beta} = \rho_\beta \rho_\alpha^p + \rho_\alpha^\beta - \rho_\beta \sum_k V_k^\circ \rho_\alpha^k$$

$$C_{\alpha\beta\gamma} = \rho_\gamma B_{\alpha\beta}^p + B_{\alpha\beta}^\gamma - \rho_\gamma \sum_k V_k^\circ B_{\alpha\beta}^k \qquad 6.30$$

$$D_{\alpha\beta\gamma\delta} = \rho_\delta C_{\alpha\beta\gamma}^p + C_{\alpha\beta\gamma}^\delta - \rho_\delta \sum_k V_k^\circ C_{\alpha\beta\gamma}^k$$

The SI results are also included in Figure 6.4 and Figure 6.5 for comparison. The SI fluctuating quantities oscillate with composition. However, the corresponding G's display a monotonic and a much less profound dependence on composition.

As mentioned previously, the B's, C's and D's can be considered as the cumulants of the multiparticle distribution function for a given local region of the solution. It is informative to generate probability distributions that possess these characteristics. In Figure 6.6 we provide the probability distributions for spherical volumes of solution (radii of 1 and 10 nm) for a SI mixture of water (1) and methanol (2) generated using the maximum entropy method
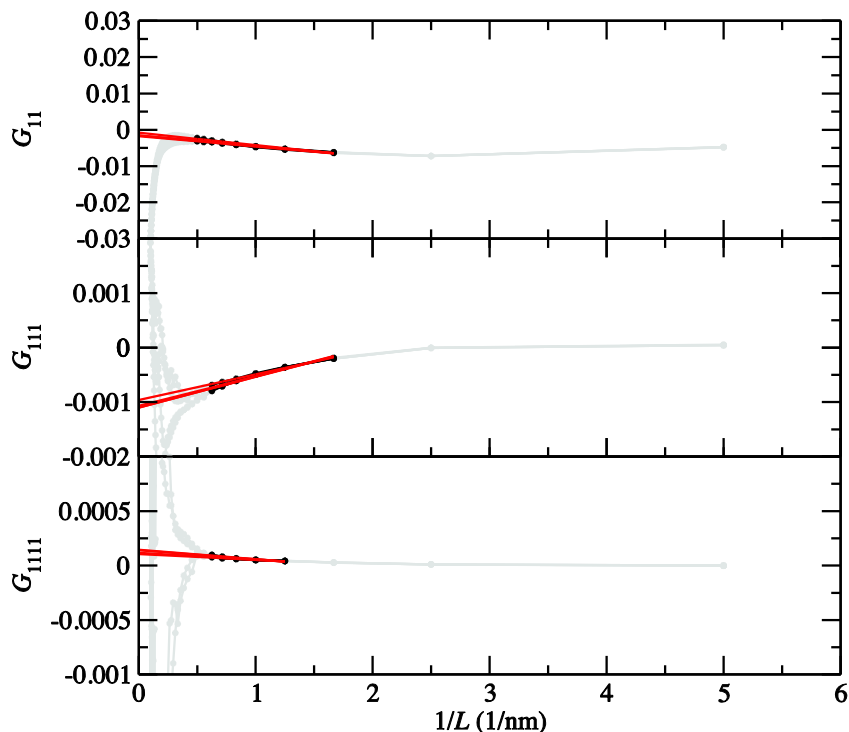
Figure 6.7 The linear extrapolation method of Schnell *et al*. applied to the probability distributions. Sample results are shown for simulated data obtained from a mixture of methanol (2) and water (1) at 300 K and 1 bar and a solute mole fraction of $x_2 = 0.25$. Data corresponding to the pair, triplet and quadruplet distributions are displayed as a function of the sample volume (cubic box length, L). The range of data points considered for the linear region is highlighted with bold symbols, together with the extrapolated (red) lines obtained from a fit to the linear region. The multiple lines correspond to data obtained from the analysis of four 20 ns long sub-averages of the trajectory.

(Abramov 2007). We restrict our analysis to a description of the single particle distributions (characterized by $B_{\alpha\alpha}$, $C_{\alpha\alpha\alpha}$, $D_{\alpha\alpha\alpha\alpha}$) for simplicity. The deviations from Gaussian behavior are small but clearly visible for the smaller reference volume. It is also clear that the distributions (variance and skewness) change significantly with composition even for a (presumably) simple model of a solution mixture. Interestingly, as the volume of solution under consideration increases, the distribution becomes *visually* indistinguishable from a normal (Gaussian) distribution. However, the *values* of C and D for this distribution are not zero and are indeed the same as for the smaller volume of solution. The fluctuations are significantly larger for water compared to methanol due to the approximate relationship provided by Equation 6.26.

6.3.5 Computer Simulation of the Fluctuations

To make full use of the current type of analysis we envision that computer simulation data will be used to aid in many interpretations. Hence, it is important to determine if reliable fluctuations can be obtained from simulation data using current computational resources. To do this one needs to have a system where the answers are known for
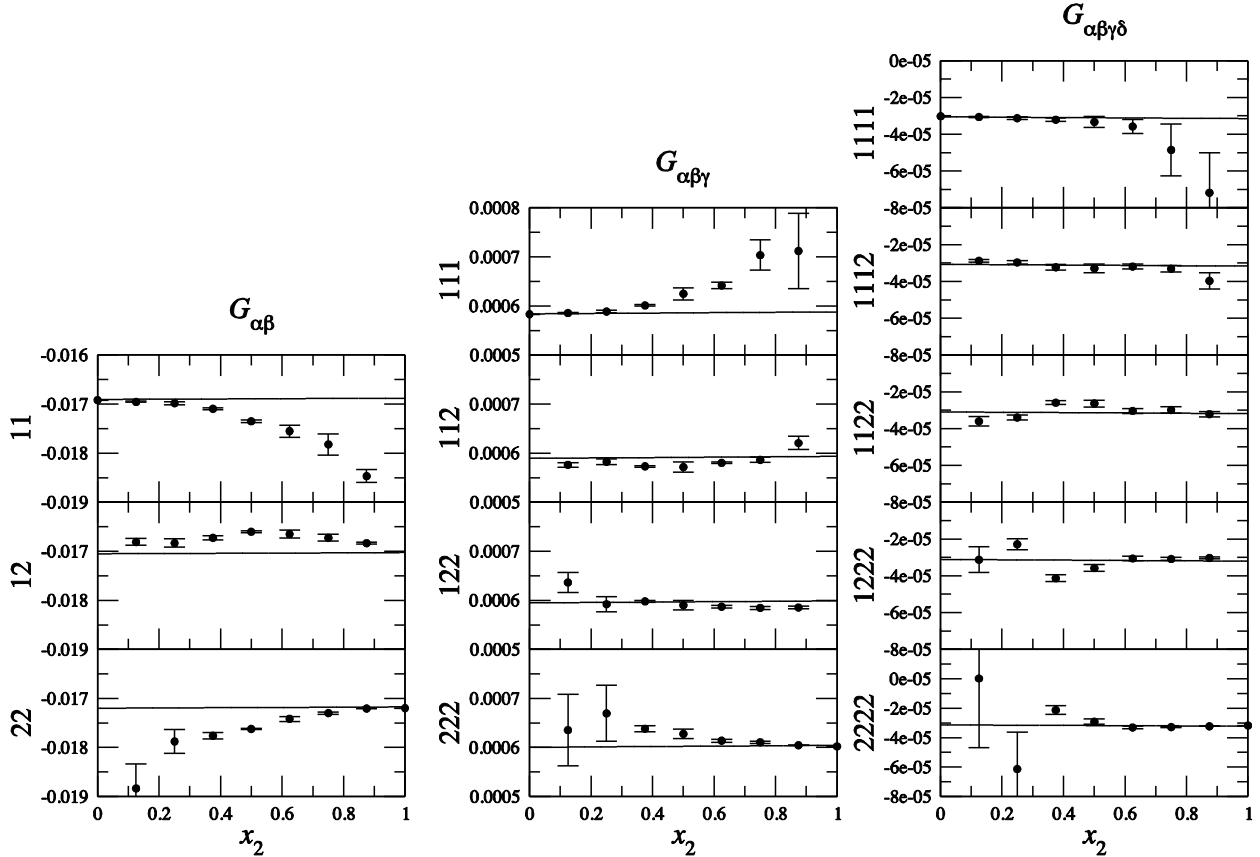
125
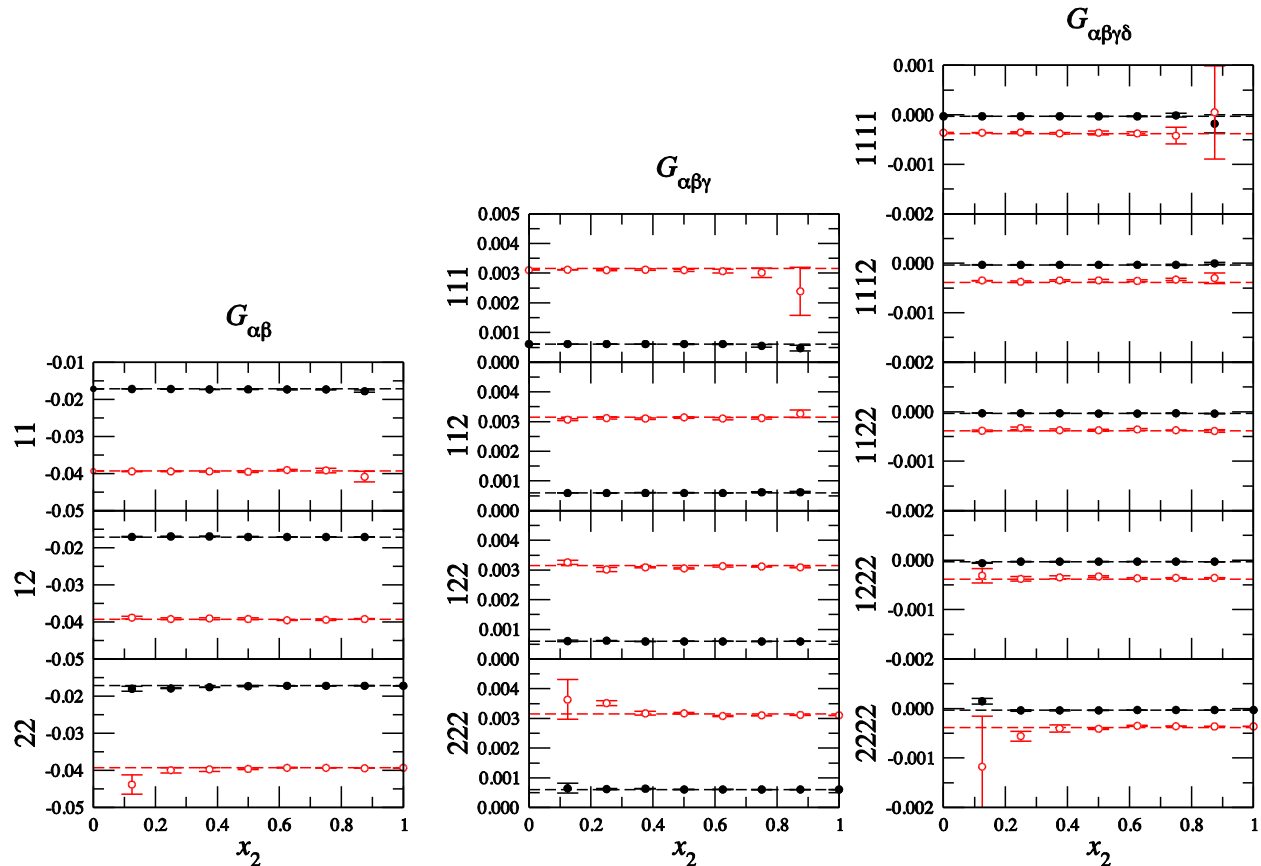
Figure 6.8 Simulated integrals over the pair ($G_{\alpha\beta}$ in M^{-1}), triplet ($G_{\alpha\beta\gamma}$ in M^{-2}), and quadruplet ($G_{\alpha\beta\gamma\delta}$ in M^{-3}) probability distributions for two pseudo binary SI solutions as a function of "solute" mole fraction obtained from simulations at 298 K and 1 bar. Dashed lines represent the corresponding SI results, while symbols correspond to the simulated systems. Data are shown for an analysis of pure SPC water (black) and pure KBFF methanol (red).

the model. Unfortunately, to our knowledge, there are none available. Hence, our first test was to investigate whether we could reproduce the fluctuations and integrals for a pseudo SI binary system, *i.e.* the analysis of a pure liquid as if it was composed of two components with identical properties. Simulations were performed of pure water (SPC) and pure methanol (KBFF) at 300 K and 1 bar, and at additional pressures of 250, 500, 750 and 1000 bar. Each system was then analyzed as a binary mixture of components at different compositions. The G's are clearly the same for both components and correspond to the G's for pure water or pure methanol according to the FF model.

To determine the fluctuations, and thereby the G's, we used the recent approach of Schnell *et al.* in which one plots the particle number fluctuations as a function of local volume within the solution (Schnell *et al.* 2011). This approach was also extended to the triplet and quadruplet fluctuations. An example of the type of results one observes is provided in Figure 6.7. The fluctuating values are then obtained after linear extrapolation to the $1/L$ limit. The choice of linear region remains somewhat subjective as very small (presumably because small volumes do not sample the larger δN_α fluctuations in a reasonable manner) and very large (as the open region is no longer
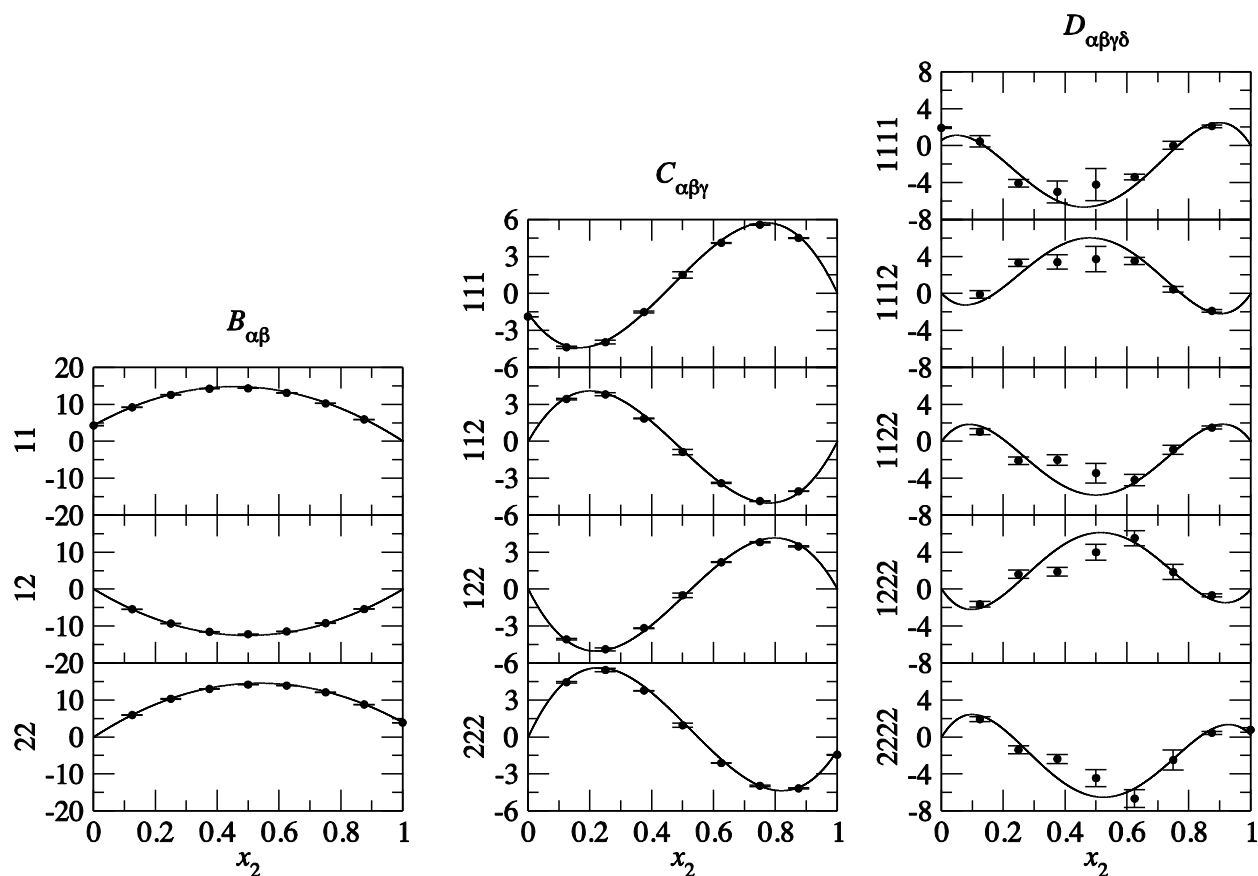
Figure 6.9 Local fluctuations (in *M*) for a binary mixture of the SPC (2) and TIP3P (1) water models at 300 K and 1 bar as a function of composition. The solid lines represent the SI results, while the symbols correspond to the simulated data.

surrounded by sufficient bulk solution) solution volumes do not display this linearity. The main advantage of this approach is that the fluctuations determined for the different sample volumes appear to converge with size much faster than the corresponding integrated distribution functions centered on a particular molecule. However, all the local composition information around a specific molecule is then lost.

All the simulation data provided here was obtained from simulations of molecules in periodic boxes of length 10 nm simulated for 110 ns in the isothermal (300 K) isobaric (1 bar) ensemble. Four 20 ns sub-averages from the final 80 ns of trajectory were used to estimate the statistical significance in the data. The estimated standard deviations from the four sub-averages were then divided by ½ to estimate the error for the full 80 ns trajectory.

Using the above approach we obtained the integrals displayed in Figure 6.8 for two pseudo binary mixtures. The simulated data clearly agree with the known SI results within the error bars. Again, the errors increase significantly as one of the components tends to low concentrations. It was somewhat surprising to us that one could obtain the triplet and quadruplet *G*'s considering that there has been significant discussion concerning the precision of the pair *G*'s obtained from closed system simulations (Wedberg *et al.* 2010; Schnell *et al.* 2011). We assume that the use of

Figure 6.10 Integrals over the pair ($G_{\alpha\beta}$ in M^{-1}), triplet ($G_{\alpha\beta\gamma}$ in M^{-2}), and quadruplet ($G_{\alpha\beta\gamma\delta}$ in M^{-3}) probability distributions for a binary mixture of the SPC (2) and TIP3P (1) water models at 300 K and 1 bar as a function of composition. The solid lines represent the SI results, while the symbols correspond to the simulated data.

large (L = 10 nm) size systems, for long (100 ns) simulation times, followed by the extrapolation analysis has dramatically increased the precision of these calculations. We have also analyzed the SPC simulation as a pseudo ternary and quaternary SI system. The results (not shown) are in agreement with the known pure solution values and are statistically significant for compositions which do not approach infinite dilution. Furthermore, the SI values displayed in Figure 6.8 were obtained from the pressure dependence of the pure solution density, and thereby provide a model consistent check of the underlying theory.

A slightly less trivial system, but one where the results should also closely follow the SI values, involves a mixture of the SPC and TIP3P water models. These models display a minimum in the enthalpy of mixing at -0.02 kJ/mol, clearly indicating the solution ideality. In Figure 6.9 we present the simulated fluctuations for this system compared to the results expected if this was a true SI solution. The results suggest we can obtain meaningful (precise) simulated values for the B's, C's, and D's, although the statistical error does depend on the exact composition of interest. However, to a large extent the fluctuating quantities are dominated by the number densities. The corresponding G's are more sensitive to the system properties. The G's are displayed in Figure 6.10. Here, one can capture the (very) small deviations from SI behavior. The respective integrals systematically approach the correct limiting values, even though there are very small differences between the two pure solutions. We should note that this is a trend we have observed

in simulations before. That is, the precision of the simulated data appears to be related to the degree of ideality of the solution, with the precision increasing as the solution becomes more ideal.

In Figure 6.4 and Figure 6.5 we have also included the simulated moments and G's for a binary mixture of SPC/E water and the KBFF model for methanol. Again, the results suggest that one can obtain reasonably precise values for these systems. Of course, the accuracy of the results will ultimately depend on the quality of the FF models. For almost all compositions the deviation from SI behavior exhibited by the experimental data is displayed by the simulations. In fact, the statistical uncertainty displayed by the simulations is much less than that of the different experimental data sets. Unfortunately, the statistical noise in the simulated integrals increases as the molecules approach infinite dilution, and yet these values are often of principle interest from a theoretical perspective.

6.4 Conclusions

We have illustrated how triplet and quadruplet fluctuations, or integrals over triplet and quadruplet distribution functions, can be obtained from existing experimental and simulation data. The results could help to provide a deeper understanding of solutions and solution mixtures. The only other experimental technique that we know of which provides such data for solutions is that of solution scattering studies (Egelstaff 1973). However, these are limited to triplet correlations and to relatively simple mixtures. No such limitation is found with the current approach. It appears that both the experimental extraction and simulation of these correlations is viable for the triplet correlations, while the quadruplet correlations may be more problematic, especially at low concentrations. Fortunately, while the current approach does not provide the experimental values of $g^{(3)}$ or $g^{(4)}$ (we know of no such general approach), the thermodynamics of the solution are directly related to the integrals over these correlation functions, which are provided by the current approach.

Some areas of future study involve: *i*) extending the current approach to include energy fluctuations; *ii*) developing the expressions for third derivatives of the Gibbs free energy (*i.e.*, partial molar compressibilities, expansivities, and heat capacities) in terms of fluctuating quantities; *iii*) deriving a fluctuation formula that corresponds to the superposition approximation; *iv*) examining more systems using available experimental data; *v*) deriving expressions for the limiting behavior of the G's in binary or ternary systems; *vi*) comparing the results for systems with and without explicit polarization to results obtained with and without the superposition approximation; and *vii*) obtaining $g^{(3)}$ from our simulations for comparison with the integrated values.

6.5 Appendices

6.5.1 Analysis of Experimental Data

The expressions provided above were developed using the chemical potential and **Z** matrix definitions provided in Equations 6.11 and 6.13, respectively. We found these definitions to be most useful for developing the expressions that appear in Equation 6.23, for example. However, a more general approach, valid for any number of components,

is possible using matrix algebra. To simplify the notation we will redefine the central $\mathbf{Z}$ matrix and chemical potential derivative such that,

$$\mathbf{Z} = \begin{bmatrix} \mu_{11} + \bar{V}_1 & \mu_{21} + \bar{V}_2 & \cdots & \mu_{n1} + \bar{V}_3 \\ \mu_{12} + \bar{V}_1 & \mu_{22} + \bar{V}_2 & \cdots & \mu_{n2} + \bar{V}_3 \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{1n} + \bar{V}_1 & \mu_{2n} + \bar{V}_2 & \cdots & \mu_{nn} + \bar{V}_3 \end{bmatrix} \qquad \mu_{\alpha\beta} = \left(\frac{\partial \mu_\alpha}{\partial x_\beta} \right)_{\beta, p} \qquad 6.31$$

We note that these definitions are restricted to this section only. Furthermore, we will express all the required derivatives in terms of derivatives of the $\mathbf{Z}$ matrix at the expense of additional matrix operations. Hence, this approach will be much less suitable for the development of explicit expressions for the fluctuating quantities in terms of experimental data, but will be more general and relatively easy to implement on a computer.

Before proceeding we note that there are several useful properties of this new $\mathbf{Z}$ matrix which will be utilized later. These include,

$$\mathbf{Z\rho} = \mathbf{1} \qquad\qquad (\mathbf{Z}^{-1})^{\mathbf{a}} = -\mathbf{Z}^{-1}\mathbf{Z}^{\mathbf{a}}\mathbf{Z}^{-1} \qquad 6.32$$

where the superscript a (and later b and c) indicates a derivative with respect to pressure or a specific mole fraction x_i. The above equations can be used to provide expressions for first, second and third derivatives of the number densities with respect to pressure or composition. The results are,

$$\boldsymbol{\rho}^{\mathbf{a}} = -\mathbf{Z}^{-1}\mathbf{Z}^{\mathbf{a}}\boldsymbol{\rho} \qquad\qquad \boldsymbol{\rho}^{\mathbf{ab}} = -\mathbf{Z}^{-1}(\mathbf{Z}^{\mathbf{b}}\boldsymbol{\rho}^{\mathbf{a}} + \mathbf{Z}^{\mathbf{ab}}\boldsymbol{\rho} + \mathbf{Z}^{\mathbf{a}}\boldsymbol{\rho}^{\mathbf{b}})$$

$$6.33$$

$$\boldsymbol{\rho}^{\mathbf{abc}} = -\mathbf{Z}^{-1}(\mathbf{Z}^{\mathbf{c}}\boldsymbol{\rho}^{\mathbf{ab}} + \mathbf{Z}^{\mathbf{b}}\boldsymbol{\rho}^{\mathbf{ac}} + \mathbf{Z}^{\mathbf{a}}\boldsymbol{\rho}^{\mathbf{bc}} + \mathbf{Z}^{\mathbf{bc}}\boldsymbol{\rho}^{\mathbf{a}} + \mathbf{Z}^{\mathbf{ac}}\boldsymbol{\rho}^{\mathbf{b}} + \mathbf{Z}^{\mathbf{ab}}\boldsymbol{\rho}^{\mathbf{c}} + \mathbf{Z}^{\mathbf{abc}}\boldsymbol{\rho})$$

which can be solved sequentially. There will be a series of these vectors corresponding to derivatives with respect to to each component. Again, these are relatively easy to manipulate on a computer.

Using the above definition of the $\mathbf{Z}$ matrix the expressions provided in Equation 6.12 can be written,

$$\rho_\alpha^p \mathbf{1} + \boldsymbol{\rho}_\alpha^{\mathbf{I}} = \mathbf{Z}\mathbf{B}_\alpha$$

$$B_{\alpha\beta}^p \mathbf{1} + \mathbf{B}_{\alpha\beta}^{\mathbf{I}} = \mathbf{Z}\mathbf{C}_{\alpha\beta} \qquad 6.34$$

$$C_{\alpha\beta\gamma}^p \mathbf{1} + \mathbf{C}_{\alpha\beta\gamma}^{\mathbf{I}} = \mathbf{Z}\mathbf{D}_{\alpha\beta\gamma}$$

Note here that $\boldsymbol{\rho}_\alpha^{\mathbf{I}} = (\rho_\alpha^1, \rho_\alpha^2, ...)$, where the i superscript indicates a derivative with respect to x_i. This is not the same as the derivatives indicated in Equation 6.33, but this vector can be generated directly from those results. The inversion approach then leads to the following expressions for the B's, C's and D's,

$$\mathbf{B}_{\alpha} = \rho_{\alpha}^{p}\boldsymbol{\rho} + \mathbf{Z}^{-1}\boldsymbol{\rho}_{\alpha}^{\mathbf{I}}$$

$$\mathbf{B}_{\alpha}^{\mathbf{a}} = \rho_{\alpha}^{pa}\boldsymbol{\rho} + \rho_{\alpha}^{p}\boldsymbol{\rho}^{\mathbf{a}} - \mathbf{Z}^{-1}(\mathbf{Z}^{\mathbf{a}}\mathbf{Z}^{-1}\boldsymbol{\rho}_{\alpha}^{\mathbf{I}} - \boldsymbol{\rho}_{\alpha}^{\mathbf{Ia}})$$

$$\mathbf{B}_{\alpha}^{\mathbf{ab}} = \rho_{\alpha}^{pab}\boldsymbol{\rho} + \rho_{\alpha}^{pa}\boldsymbol{\rho}^{\mathbf{b}} + \rho_{\alpha}^{pb}\boldsymbol{\rho}^{\mathbf{a}} + \rho_{\alpha}^{p}\boldsymbol{\rho}^{\mathbf{ab}}$$
$$- \mathbf{Z}^{-1}\left[\mathbf{Z}^{\mathbf{ab}}\mathbf{Z}^{-1}\boldsymbol{\rho}_{\alpha}^{\mathbf{I}} + \mathbf{Z}^{\mathbf{a}}\mathbf{Z}^{-1}(\boldsymbol{\rho}_{\alpha}^{\mathbf{Ib}} - \mathbf{Z}^{\mathbf{b}}\mathbf{Z}^{-1}\boldsymbol{\rho}_{\alpha}^{\mathbf{I}}) + \mathbf{Z}^{\mathbf{b}}\mathbf{Z}^{-1}(\boldsymbol{\rho}_{\alpha}^{\mathbf{Ia}} - \mathbf{Z}^{\mathbf{a}}\mathbf{Z}^{-1}\boldsymbol{\rho}_{\alpha}^{\mathbf{I}}) - \boldsymbol{\rho}_{\alpha}^{\mathbf{Iab}}\right]$$

$$\mathbf{C}_{\alpha\beta} = B_{\alpha\beta}^{p}\boldsymbol{\rho} + \mathbf{Z}^{-1}\mathbf{B}_{\alpha\beta}^{\mathbf{I}}$$

$$\mathbf{C}_{\alpha\beta\gamma}^{\mathbf{a}} = B_{\alpha\beta}^{pa}\boldsymbol{\rho} + B_{\alpha\beta}^{p}\boldsymbol{\rho}^{\mathbf{a}} - \mathbf{Z}^{-1}(\mathbf{Z}^{\mathbf{a}}\mathbf{Z}^{-1}\mathbf{B}_{\alpha\beta}^{\mathbf{I}} - \mathbf{B}_{\alpha\beta}^{\mathbf{Ia}})$$

$$\mathbf{D}_{\alpha\beta\gamma} = C_{\alpha\beta\gamma}^{p}\boldsymbol{\rho} + \mathbf{Z}^{-1}\mathbf{C}_{\alpha\beta\gamma}^{\mathbf{I}}$$

6.35

which can again be solved in an sequential manner. A general expression for the B's can be written,

$$\mathbf{B} = k_{\mathrm{B}}T\kappa_{T}\boldsymbol{\rho}\boldsymbol{\rho}^{\mathbf{T}} + \mathbf{Z}^{-1}\boldsymbol{\Phi} \qquad 6.36$$

where the $\boldsymbol{\Phi}$ matrix has elements of $\Phi_{\alpha\beta} = \rho_{\beta}^{\alpha} = \rho_{\beta}(\delta_{\alpha\beta} - \phi_{\alpha})/[x_{\alpha}(1 - x_{\alpha})]$.

Hence, from derivatives of the $\mathbf{Z}$ matrix one can generate all the derivatives of the number densities and thereby the required derivatives of the B's and C's. To determine derivatives of the $\mathbf{Z}$ matrix one requires values of G^{Ex} and V^{Ex} as a function of composition, together with the pure volumes and V^{Ex} (or the density) as a function of pressure. The ideal contributions $\beta\mu_{\alpha}^{\mathrm{id}} = \beta\mu_{\alpha}^{\circ} + \ln x_{\alpha}$ to the chemical potential derivatives are given by,

$$\mu_{\alpha\beta}^{\mathrm{id}} = \frac{\delta_{\alpha\beta} - x_{\beta}}{x_{\beta}(1 - x_{\beta})} \qquad\qquad \mu_{\alpha\beta\gamma}^{\mathrm{id}} = -\mu_{\alpha\beta}^{\mathrm{id}}\mu_{\alpha\gamma}^{\mathrm{id}}$$

6.37

$$\mu_{\alpha\beta\gamma\delta}^{\mathrm{id}} = 2\mu_{\alpha\beta}^{\mathrm{id}}\mu_{\alpha\gamma}^{\mathrm{id}}\mu_{\alpha\delta}^{\mathrm{id}} \qquad\qquad \mu_{\alpha\beta\gamma\delta\varepsilon}^{\mathrm{id}} = -6\mu_{\alpha\beta}^{\mathrm{id}}\mu_{\alpha\gamma}^{\mathrm{id}}\mu_{\alpha\delta}^{\mathrm{id}}\mu_{\alpha\varepsilon}^{\mathrm{id}}$$

for all components using the mole fractions scale $\beta\mu_{\alpha} = \beta\mu_{\alpha}^{\mathrm{id}} + \ln f_{\alpha}$, where f is the mole fraction scale activity coefficient. Derivatives of each f and V^{Ex} can be obtained from the usual fitting functions.

In the present analysis of the experimental data for binary solutions we have made one minor approximation. The pressure dependence of the solution volume as a function of composition is rarely available in the literature. Therefore, we have assumed that,

$$\left(\frac{\partial^{n}V_{\mathrm{m}}}{\partial p^{n}}\right)_{T} = \sum_{\alpha} x_{\alpha}\left(\frac{\partial^{n}V_{\alpha}^{\circ}}{\partial p^{n}}\right)_{T} \qquad 6.38$$

for all solution mixtures. This is equivalent to neglecting the pressure dependence of the excess molar volume or excess partial molar volumes. A comparison of the data in Figure 6.1 and Figure 6.4 indicates that the magnitude of the C's obtained for solution mixtures is typically much larger than for the pure solutions. Hence, this approximation should be very reasonable under ambient conditions.

131

6.5.2 Distribution Functions

In the grand canonical ensemble the probability that any N_α molecules of species α, and N_β molecules of species β, *etc.* are within $d\{r\}$ at $\{r\}$ is given by $\rho^{(n)}(\{r\})d\{r\}$ where (Hill 1956),

$$\int \rho^{(n)}\left(\{r\}\right)d\{r\} = \left\langle \prod_s \frac{N_s!}{(N_s - n_s)!} \right\rangle \qquad 6.39$$

Here, the product involves the different species (s) present in the solution, while n_s is the number of molecules of each species in the (n) particle distribution. We require integrals up to and including the four body distribution for a general mixture of any number of components. These involve integrals over the following probability density distributions,

$$\int \rho_\alpha^{(1)}\left(r_1\right)dr_1 = \left\langle N_\alpha \right\rangle$$

$$\int \rho_{\alpha\beta}^{(2)}\left(r_1, r_2\right)dr_1 dr_2 = \left\langle N_\alpha(N_\beta - \delta_{\alpha\beta}) \right\rangle$$

$$\int \rho_{\alpha\beta\gamma}^{(3)}\left(r_1, r_2, r_3\right)dr_1 dr_2 dr_3 = \left\langle N_\alpha(N_\beta - \delta_{\alpha\beta})(N_\gamma - \delta_{\alpha\gamma} - \delta_{\beta\gamma}) \right\rangle \qquad 6.40$$

$$\int \rho_{\alpha\beta\gamma\delta}^{(4)}\left(r_1, r_2, r_3, r_4\right)dr_1 dr_2 dr_3 dr_4 = \left\langle N_\alpha(N_\beta - \delta_{\alpha\beta})(N_\gamma - \delta_{\alpha\gamma} - \delta_{\beta\gamma})(N_\delta - \delta_{\alpha\delta} - \delta_{\beta\delta} - \delta_{\gamma\delta}) \right\rangle$$

In analogy with the theory of imperfect gases we define the following integrals (Kirkwood and Buff 1951; Hill 1956),

$$\rho_\alpha \rho_\beta V G_{\alpha\beta} = \int \left[\rho_{\alpha\beta}^{(2)} - \rho_\alpha^{(1)}\rho_\beta^{(1)} \right] dr_1 dr_2$$

$$\rho_\alpha \rho_\beta \rho_\gamma V G_{\alpha\beta\gamma} = \int \left[\rho_{\alpha\beta\gamma}^{(3)} - \rho_{\alpha\beta}^{(2)}\rho_\gamma^{(1)} - \rho_{\alpha\gamma}^{(2)}\rho_\beta^{(1)} - \rho_{\beta\gamma}^{(2)}\rho_\alpha^{(1)} + 2\rho_\alpha^{(1)}\rho_\beta^{(1)}\rho_\gamma^{(1)} \right] dr_1 dr_2 dr_3 \qquad 6.41$$

$$\rho_\alpha \rho_\beta \rho_\gamma \rho_\delta V G_{\alpha\beta\gamma\delta} = \int \begin{bmatrix} \rho_{\alpha\beta\gamma\delta}^{(4)} - \rho_{\alpha\beta\gamma}^{(3)}\rho_\delta^{(1)} - \rho_{\alpha\beta\delta}^{(3)}\rho_\gamma^{(1)} - \rho_{\alpha\gamma\delta}^{(3)}\rho_\beta^{(1)} - \rho_{\beta\gamma\delta}^{(3)}\rho_\alpha^{(1)} \\ -\rho_{\alpha\beta}^{(2)}\rho_{\gamma\delta}^{(2)} - \rho_{\alpha\gamma}^{(2)}\rho_{\beta\delta}^{(2)} - \rho_{\alpha\delta}^{(2)}\rho_{\beta\gamma}^{(2)} \\ +2\rho_{\alpha\beta}^{(2)}\rho_\gamma^{(1)}\rho_\delta^{(1)} + 2\rho_{\alpha\gamma}^{(2)}\rho_\gamma^{(1)}\rho_\delta^{(1)} + 2\rho_{\alpha\delta}^{(2)}\rho_\beta^{(1)}\rho_\gamma^{(1)} \\ +2\rho_{\beta\gamma}^{(2)}\rho_\alpha^{(1)}\rho_\delta^{(1)} + 2\rho_{\beta\delta}^{(2)}\rho_\alpha^{(1)}\rho_\gamma^{(1)} + 2\rho_{\gamma\delta}^{(2)}\rho_\alpha^{(1)}\rho_\beta^{(1)} \\ -6\rho_\alpha^{(1)}\rho_\beta^{(1)}\rho_\gamma^{(1)}\rho_\delta^{(1)} \end{bmatrix} dr_1 dr_2 dr_3 dr_4$$

where we have removed the spatial dependencies for clarity. These integrals arise during the common expansions of the partition function (see Equation 6.3) employed in gas and solution theory. The first two integrals also appeared in the original KB paper (Kirkwood and Buff 1951). The above integrals can be expressed in terms of probability distributions *via* their definitions,

$$g_\alpha^{(1)}(r_1) = \frac{\rho_\alpha^{(1)}\left(r_1\right)}{\rho_\alpha} = 1 \qquad\qquad g_{\alpha\beta}^{(2)}(r_1, r_2) = \frac{\rho_{\alpha\beta}^{(2)}\left(r_1, r_2\right)}{\rho_\alpha \rho_\beta}$$

$$g_{\alpha\beta\gamma}^{(3)}(r_1, r_2, r_3) = \frac{\rho_{\alpha\beta\gamma}^{(3)}\left(r_1, r_2, r_3\right)}{\rho_\alpha \rho_\beta \rho_\gamma} \qquad\qquad g_{\alpha\beta\gamma\delta}^{(4)}(r_1, r_2, r_3, r_4) = \frac{\rho_{\alpha\beta\gamma\delta}^{(4)}\left(r_1, r_2, r_3, r_4\right)}{\rho_\alpha \rho_\beta \rho_\gamma \rho_\delta}$$

<div align="right">6.42</div>

A combination of Equations 6.41 and 6.42 followed by some minor rearrangement provides the integrals given in Equation 6.16. Furthermore, a combination of Equations 6.40 and 6.41 provides the G's in terms of fluctuating quantities,

$$\rho_\alpha \rho_\beta V G_{\alpha\beta} = \langle \delta N_\alpha \delta N_\beta \rangle - \delta_{\alpha\beta} \langle N_\alpha \rangle$$

$$\rho_\alpha \rho_\beta \rho_\gamma V G_{\alpha\beta\gamma} = \langle \delta N_\alpha \delta N_\beta \delta N_\gamma \rangle - \delta_{\beta\gamma} \langle \delta N_\alpha \delta N_\beta \rangle - \delta_{\alpha\beta} \langle \delta N_\alpha \delta N_\gamma \rangle$$
$$- \delta_{\alpha\gamma} \langle \delta N_\beta \delta N_\gamma \rangle + 2\delta_{\alpha\beta}\delta_{\alpha\gamma} \langle N_\alpha \rangle$$

$$
\begin{aligned}
\rho_\alpha \rho_\beta \rho_\gamma \rho_\delta V G_{\alpha\beta\gamma\delta} =\ & \langle \delta N_\alpha \delta N_\beta \delta N_\gamma \delta N_\delta \rangle - \langle \delta N_\alpha \delta N_\beta \rangle \langle \delta N_\gamma \delta N_\delta \rangle \\
& - \langle \delta N_\alpha \delta N_\gamma \rangle \langle \delta N_\beta \delta N_\delta \rangle - \langle \delta N_\alpha \delta N_\delta \rangle \langle \delta N_\beta \delta N_\gamma \rangle \\
& - (\delta_{\alpha\delta} + \delta_{\gamma\delta}) \langle \delta N_\alpha \delta N_\beta \delta N_\gamma \rangle - (\delta_{\alpha\gamma} + \delta_{\beta\gamma}) \langle \delta N_\alpha \delta N_\beta \delta N_\delta \rangle \\
& - (\delta_{\alpha\beta} + \delta_{\beta\delta}) \langle \delta N_\alpha \delta N_\gamma \delta N_\delta \rangle \\
& + (\delta_{\alpha\gamma}\delta_{\alpha\delta} + \delta_{\alpha\delta}\delta_{\gamma\delta} + \delta_{\beta\delta}\delta_{\gamma\delta} + \delta_{\beta\gamma}\delta_{\gamma\delta}) \langle \delta N_\alpha \delta N_\beta \rangle \\
& + (\delta_{\alpha\beta}\delta_{\alpha\gamma} + \delta_{\alpha\beta}\delta_{\gamma\delta} + \delta_{\alpha\gamma}\delta_{\beta\gamma} + \delta_{\alpha\gamma}\delta_{\beta\delta}) \langle \delta N_\alpha \delta N_\delta \rangle \\
& + (\delta_{\alpha\beta}\delta_{\alpha\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma} + \delta_{\alpha\delta}\delta_{\beta\delta}) \langle \delta N_\alpha \delta N_\gamma \rangle - 6\delta_{\alpha\beta}\delta_{\alpha\gamma}\delta_{\alpha\delta} \langle N_\alpha \rangle
\end{aligned}
$$

6.43

These expressions can then be rearranged to provide the equivalent fluctuating quantities given in Equation 6.15.

7. Apparent Molar Approach for the Calculation of Infinitely Dilute Partial Molar Properties of Proteins from Computer Simulation[*]

Abstract

A detailed understanding of temperature and pressure effects on an infinitely dilute protein's conformational equilibrium requires knowledge of the corresponding infinitely dilute partial molar properties. Established molecular dynamics methodologies generally have not provided a way to calculate these properties without either a loss of thermodynamic rigor, the introduction of non-unique parameters, or a loss of information about which solute conformations specifically contributed to the output values. Here we implement a simple method that is thermodynamically rigorous and possesses none of the above disadvantages and we report on the method's feasibility and computational demands. We calculate infinitely dilute partial molar properties for two proteins and attempt to distinguish the thermodynamic differences between a native and a denatured conformation of a designed mini-protein. We conclude that simple ensemble average properties can be calculated with very reasonable amounts of computational power. In contrast, properties corresponding to fluctuating quantities are computationally demanding to calculate precisely from a single simulation, although they can be obtained relatively easily by following the temperature and/or pressure dependence of the corresponding ensemble averages.

---

[*] Design, derivation, advising, and editing by PES. Simulations, analysis, figures, and manuscript by EAP under the guidance of PES excluding the Appendix text, which was almost entirely written by PES.

7.1 Introduction

Equilibrium constants (and their derivatives) that describe the delicate balance between the native and denatured states of proteins are frequently measured to understand if and why certain proteins are more stable than others and how mutations alter this stability. However, under physiological conditions, the ratio of [denatured]/[native] is very small, making it difficult to measure accurately. Often the equilibrium must be perturbed, for example by changing the temperature or pressure, in order to achieve an equilibrium constant that is larger in magnitude and can, therefore, be calculated more precisely. Afterwards, the results are usually extrapolated back to physiological conditions. Thus, it is important to understand the effects of temperature and pressure on the equilibrium constant. One way of learning more about these effects is through computer simulation. However, structural properties are usually the sole or primary focus of most simulations studies, and the corresponding thermodynamic properties are often ignored. When thermodynamic properties are determined from simulation, it is typically by way of a method that does not retain information concerning which conformation gave rise to the specific values - meaning the simulation does not really provide any more insight than an experiment would - or the properties are calculated from phenomenological relationships rather than rigorously derived thermodynamic expressions.

7.1.1 Thermal and pressure effects on conformational equilibria

Although multiple environmental variables affect a proteins' conformational equilibrium (Davis-Searles *et al.* 2001), here we limit our scope to thermal and pressure effects. Hence, the effects of cosolvents and pH on protein equilibrium are not discussed here, although a theoretical framework for understanding their effects on equilibrium has recently been advanced (Smith 2004b, 2004a; Jiao and Smith 2011).

To understand the effects of the environment on a protein's thermodynamic properties, it is common to study an infinitely dilute protein in a single solvent. To keep the situation tractable, we will assume that a protein molecule adopts one of only two forms, the native, N, or the denatured, D.[*] When considering the equilibrium between N and D, N$\rightleftharpoons$D, the standard state Gibbs free energy of unfolding ($\Delta_N^D G^o$, hereafter ΔG^o) determines the populations of the protein conformations through the usual relationship to the thermodynamic equilibrium constant $K=a_D/a_N$, where a_i is the equilibrium activity of form i, through

$$\Delta G^o = -RT\ln K = \mu_D^o - \mu_N^o \qquad\qquad 7.1$$

where R is the gas constant, T is the temperature, μ_i is the chemical potential of the protein in conformation i, and the superscripted o denotes the standard state (Ben-Naim 1992). The activity coefficients at infinite dilution are assumed to be one, which leads to the replacement of the activities with their concentrations in the equilibrium constant expression.

---

[*] The process of classifying a specific conformation as being N or D is subjective, but addressing this subjectivity is beyond the focus of this chapter. When studying a specific system's equilibrium by computer simulation, it would be wise to use the same definition as was used experimentally *e.g.*, measuring tryptophan fluorescence, if possible.

While ΔG° quantifies the ratio of species at equilibrium, derivatives of ΔG° reveal increasingly more detailed information the higher their order. The bivariate Taylor expansion of the standard state free energy change for unfolding, $-\beta\Delta\Delta G^\circ(\beta, p) = \ln(K/K_{\text{ref}})$, up to second order around a reference (ref) inverse temperature ($\beta = 1/RT$) and pressure (p) provides an avenue to interpret experimental thermal and pressure denaturation data *via*

$$-\beta\Delta\Delta G^\circ(\beta, p) \approx \mathbf{k}^{\mathrm{T}}\mathbf{d} + \frac{1}{2!}\mathbf{d}^{\mathrm{T}}\mathbf{K}\mathbf{d} \qquad 7.2$$

where $\mathbf{k}$ is the column vector whose elements are the first derivatives of $\ln K$ with respect to β and/or p evaluated at a reference β and p and $\mathbf{K}$ is the matrix whose elements are the corresponding second partial derivatives of $\ln K$. $\mathbf{d}$ is the column vector of deviations from the reference β and p,

$$\mathbf{d} = \begin{bmatrix} \Delta\beta \\ \Delta p \end{bmatrix} \qquad 7.3$$

where $\Delta X = X - X_{\text{ref}}$.

It can be shown that

$$\mathbf{k} = \begin{bmatrix} -\Delta H^\circ & -\beta\Delta V^\circ \end{bmatrix}$$
$$\mathbf{K} = \begin{bmatrix} RT^2\Delta C_p^\circ & RT^2\Delta\alpha_p^\circ \\ RT^2\Delta\alpha_p^\circ & \beta\Delta\kappa_T^\circ \end{bmatrix} \qquad 7.4$$

where H denotes the enthalpy, V the volume, C_p the heat capacity, α_p the thermal expansion, and κ_T the isothermal compressibility. The changes are given by $\Delta H^\circ \equiv H_{\mathrm{D}}^\circ - H_{\mathrm{N}}^\circ$, $\Delta V^\circ \equiv V_{\mathrm{D}}^\circ - V_{\mathrm{N}}^\circ$, $\Delta C_p^\circ \equiv C_{p,\mathrm{D}}^\circ - C_{p,\mathrm{N}}^\circ$, $\Delta\alpha_p^\circ \equiv V_{\mathrm{D}}^\circ\alpha_{p,\mathrm{D}}^\circ - V_{\mathrm{N}}^\circ\alpha_{p,\mathrm{N}}^\circ$, and $\Delta\kappa_T^\circ \equiv V_{\mathrm{D}}^\circ\kappa_{T,\mathrm{D}}^\circ - V_{\mathrm{N}}^\circ\kappa_{T,\mathrm{N}}^\circ$.

First order Taylor expansions provide the following temperature and pressure dependencies of ΔH°, the standard state entropy change (ΔS°), and ΔV° *via*,

$$\Delta H^\circ(\beta) \approx \Delta H_{\text{ref}}^\circ - RT_{\text{ref}}^2\Delta C_{p,\text{ref}}^\circ\Delta\beta$$
$$\Delta S^\circ(\beta) \approx \Delta S_{\text{ref}}^\circ - RT_{\text{ref}}\Delta C_{p,\text{ref}}^\circ\Delta\beta$$
$$\Delta V^\circ(\beta) \approx \Delta V_{\text{ref}}^\circ - RT_{\text{ref}}^2\Delta\alpha_{p,\text{ref}}^\circ\Delta\beta \qquad 7.5$$
$$\Delta V^\circ(p) \approx \Delta V_{\text{ref}}^\circ - \Delta\kappa_{T,\text{ref}}^\circ\Delta p$$

Thus, to use MD simulations to understand how temperature and pressure affect the conformational equilibrium of proteins in significant detail, *i.e.*, to map $\Delta\Delta G^\circ$ as a function of β and p, one requires ΔH°, ΔV°, ΔC_p°, $\Delta\alpha_p^\circ$, and $\Delta\kappa_T^\circ$. Similarly, to map ΔH°, ΔS°, and ΔV° as a function of β and p, one requires ΔC_p°, $\Delta\alpha_p^\circ$, and $\Delta\kappa_T^\circ$.

The second derivative properties can also be expressed as

$$C^o_{p,i} = \left(\partial H^o_i / \partial T\right)_p$$
$$\kappa^o_{T,i} = -(1/V^o_i)\left(\partial V^o_i / \partial p\right)_\beta \qquad 7.6$$
$$\alpha^o_{p,i} = (1/V^o_i)\left(\partial V^o_i / \partial T\right)_p$$

The relationships in Equation 7.6 will be used later. All of the above properties have been standard state properties. Standard state properties are usually associated with those of an infinitely dilute protein (for which the activity is approximated by the concentration) in either the native or denatured forms (Majer *et al.* 2004), which is helpful since most molecular dynamics simulations model the infinitely dilute case. In fact, the single protein molecules studied by simulation should really be considered as *pseudo* infinitely dilute (but we will drop the use of pseudo for the sake of brevity), since the amount of protein in a given volume (the formal or effective concentration) will vary depending upon the total system size even though the number of protein molecules is fixed at one.

Most simulations do not readily provide the above infinitely dilute partial molar properties, because the expressions relating the output simulation results to these properties have generally been considered unknown. In contrast, it is well known that the system fluctuations are related to the compressibility, thermal expansion, and heat capacity of the whole system. In the *NpT* ensemble the corresponding expressions are

$$VRT\kappa_T = <(\delta V)^2 >$$
$$VRT^2\alpha_p = <\delta V\delta H > \qquad 7.7$$
$$RT^2C_p = <(\delta H)^2 >$$

The angled brackets indicate time or ensemble averages. $\delta X = X - \langle X \rangle$ and $\langle \delta X \delta Y \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle = \text{cov}[X,Y]$, where cov is the covariance. We emphasize that these are all bulk system properties, which means that when applied to a system containing an infinitely dilute solute, the properties obtained are those of the solvent, perturbed to some extent by the presence of the solute molecule. For a binary mixture in which both species 1 and 2 are at finite concentrations, the process of extracting partial molar properties is straightforward. One needs only to fit the property of interest as a function of composition and then take the partial derivatives with respect to each component (Lewis 1907). However, for the infinitely dilute case, an appropriate method for determining partial molar properties has not been clear or computationally intuitive.

7.1.2 Current computer simulation methodologies for determining partial molar properties

Computationally speaking, we are only aware of one method in the literature that has been used to obtain all of the infinitely dilute partial molar properties and can be considered thermodynamically rigorous, namely, the enhanced sampling technique of replica exchange molecular dynamics (REMD). REMD simulations map out the phase space and allow one to obtain *K* by determining the fraction of protein molecules that are folded or unfolded as a function of temperature. Once this is known, the free energy of unfolding as a function of temperature can be computed according to (Canchi *et al.* 2010)

$$\Delta_{N}^{D}G^{\circ}\left(\beta, p\right) = -\beta^{-1}\ln\left(\frac{\rho_{D}}{\rho_{N}}\right) \qquad\qquad 7.8$$

where ρ_i is the number density of i ($N_i/\langle V \rangle$). After the free energy of unfolding has been computed, the difference (unfolded minus folded) in all of the thermodynamic properties may be fitted simultaneously according to the approach illustrated by *e.g*., Garcia (Canchi *et al.* 2010). Note, however, that all that is obtained using REMD is the difference between the thermodynamic properties in the N and D forms, not the values of the thermodynamic properties for each of the native and the many denatured states that collectively gave rise to that difference.

Thus, the REMD approach provides an objective way to calculate the change in an infinitely dilute partial molar quantity, $\Delta\bar{X}_{2}^{\infty}$, and is extremely useful from a force field validation or a property prediction point of view. What remains lacking, however, is the ability to assign thermodynamic properties to specific conformations *i.e*., the ability to rank different conformations according to their $\bar{X}_{2}^{\infty}$ values. Many approaches for the determination of partial molar quantities of individual biomolecule conformations have been proposed in the literature.

7.1.3 Protein volume and compressibility

From a computational viewpoint, many subjective definitions of the volume of a protein exist. Levy and co-workers used three possible definitions to calculate the volume of fifteen proteins (Murphy *et al.* 1998). The most striking volume variability was, naturally, for the smallest protein they studied, the insulin monomer (51 residues). They reported the van der Waals (VDW) volume to be 5.564 nm^3, the molecular volume (calculated using a sphere with probe radius of 0.14 nm) to be 6.848 nm^3, and the excluded volume (also calculated using a sphere with probe radius of 0.14 nm) to be 11.142 nm^3 (Murphy *et al.* 1998). This corresponds to a 100% increase in the volume on going from a VDW volume definition to an excluded volume definition. Even the largest protein they studied, Carboxypeptidase A (307 residues), had a 17% increase in the volume on going from a vdW to an excluded volume definition (reported volumes: VDW=33.571 nm^3, molecular=42.098 nm^3, excluded=58.055 nm^3) (Murphy *et al.* 1998). While there is nothing wrong with calculating these quantities and using them to probe the properties of proteins, it is unclear which approach best corresponds to the rigorously defined partial molar volume. Unfortunately, for those interested in further calculations that depend upon the protein volume, it is unclear which volume definition they should use. This has been an issue in some publications (Dadarlat and Post 2001, 2006).

We have found examples where an objective definition of the protein volume has been used *via* the apparent molar approach *i.e*. (Shing and Chung 1988; Devane *et al.* 2003; Mitra *et al.* 2006; Sarupria *et al.* 2010),

$$\bar{V}_{2}^{\infty} = \langle V \rangle - V_{1}^{*}N_{1} \qquad\qquad 7.9$$

Where $\langle V \rangle$ is the average system volume from an *NpT* simulation, V_{1}^{*} is the molar volume of pure 1, N_{1} is the number of molecules of species 1, and the angled brackets, again, refer to time or ensemble averages. [*]

---

[*] In all other chapters, pure properties are denoted by the superscript o. This was not used here so that the standard state could be denoted by the superscript o, since that is the usual notation.

Another rigorous $\bar{V}_2^\infty$ definition comes from the Kirkwood-Buff theory of solutions *via* the expression

$$\bar{V}_2^\infty = RT\kappa_T^* - G_{21}^\infty \qquad 7.10$$

where the Kirkwood-Buff integral G_{21}^∞ is given by

$$G_{21}^\infty = \int_0^\infty 4\pi r^2 [g_{21}^{\infty,\mu VT}(r)\text{-}1]dr \qquad 7.11$$

and where $g_{21}(r)$ is the radial distribution function of the solvent (1) around the solute (2) defined in the grand canonical (μVT) ensemble. Equation 7.10 has been used by several authors (Yu *et al.* 2010; Jiao and Smith 2011). Due to the non-spherical geometry of proteins, the integration in Equation 7.11 is more difficult to perform if one is interested in a surface based approach. Similarly, Hirata and collegues have derived statistical mechanical expression for the infinitely dilute partial molar volumes using the Reference Interaction Site Model (RISM) integral equation theory coupled to KB theory.

It is worth noting that while traditional KB theory provides a rigorous expression for the volume only, the generalization of KB theory to Fluctuation Solution Theory (FST) provides rigorous expressions for all of the thermodynamic properties considered in this work. However, FST is still in its infancy and therefore, while the expressions have appeared, they have not yet been applied to obtain te full set of possible thermodynamic properties (Jiao and Smith 2011; Ploetz and Smith 2013).

Most calculations of $\bar{\kappa}_{T,2}^\infty$ we have found cannot be considered rigorous for one of two main reasons. First, they may have relied on subjective definitions of the protein volume. Second, they often assume that the equation that relates the bulk system compressibility to fluctuations in the volume of the entire system can be applied to a component of the system (*i.e.*, the protein) to obtain its compressibility *i.e.*, that $\bar{\kappa}_{T,2}^\infty = \beta < (\delta V_2)^2 > / < V_2 >$, (Høiland and Hedwig; Cooper 1976; Gekko and Hasegawa 1986; Dadarlat and Post 2001, 2006) where V_2 is some measure of the protein volume. This may or may not provide a measure of the intrinsic compressibility of the protein, but it is unlikely that it corresponds to the thermodynamically defined infinitely dilute partial molar compressibility of the protein. This idea has been used in the literature many times, and may have originated from the work of Cooper (1976). It has been further supported by Hinz, but has also been strongly contested by Hirata. Furthermore, while many researchers have been able to successfully obtain reasonable agreement with the experimental $\bar{\kappa}_{T,2}^\infty$ by breaking the simulated compressibility into different contributions, such as the "intrinsic," "hydrational," and/or "thermal" components (which involves the use of non-unique volume definitions), here we seek to calculate the thermodynamically defined property, $\bar{\kappa}_{T,2}^\infty$.

A few noteworthy exceptions to the above compressibility studies exist, such as the work of Yu and coworkers mentioned in the previous paragraphs explaining protein volume calculation methods. Yu used Equation 7.10 and then

looked at the pressure dependence of the volume to obtain the corresponding compressibility. Likewise, Hirata also used the pressure derivative of the volume that was obtained from RISM combined with KB theory to obtain the isothermal compressibility. In contrast to these studies, we seek a method that can be used to analyze computer simulations, does not involve integral equations, and directly provides any first or second derivative of the free energy – rather than only providing the volume (obtained from the first pressure derivative of the free energy).

7.1.4 Protein enthalpy, heat capacity, and thermal expansion

In a computer simulation, the absolute partial molar enthalpy is available, whereas experiments only provide the excess partial molar enthalpy. When a single protein conformation is simulated, it is not obvious that there would be any value in quoting the absolute enthalpy of that protein form. This may explain why we have found considerably fewer attempts to extract $\bar{H}_2^\infty$ or $\bar{C}_{p,2}^\infty$ from simulations than the preceding properties, outside of REMD studies. For interesting explanations of the dearth of simulations of $\bar{C}_{p,2}^\infty$, the reader is referred to a review by Prabhu and Sharp (2005) and an article by Cooper (2010).

Shing and Chung (1988) did calculate the internal energy of an infinitely dilute LJ sphere in a LJ binary mixture by the finite difference of the system internal energy in the absence or presence of the infinitely dilute solute. However, the method was not generalized to all properties or other types of systems such as biomolecules. As we will show here, it is trivial to make these generalizations, but the literature from the intervening years suggests that this approach has not been popular. In the same vein, Smit and coworkers recently calculated the enthalpy of a lipid bilayer, with respect to the enthalpy of the solvent background, in order to make a connection to differential scanning calorimetry experiments (Rodgers *et al.* 2012). Unlike nucleic acids, proteins, and polysaccharides, lipids are not polymeric. Therefore, Smit calculated the (extensive) enthalpy of a lipid aggregate, not an (intensive) infinitely dilute partial molar lipid enthalpy. However, their method is in the same spirit as the method used in this work and it could be applied to any infinitely dilute solute (considering the lipid aggregate as a single object) and to any property, not just the enthalpy.

Lastly, very few examples of protein thermal expansion values calculated from computer simulations studies can be found. They have all used non-unique protein volume definitions and/or decomposed the thermal expansion into different contributions, such as the "intrinsic," "hydrational," and/or "thermal" components, which introduces several additional subjective parameters (Mitra *et al.* 2006; Mitra *et al.* 2008; Brovchenko *et al.* 2010).

7.1.5 Aim of this study

The purpose of this chapter is to investigate the computational feasibility of using a general method for extracting the thermodynamically defined $\bar{X}_2^\infty$ that does not result in a loss of information about the conformational states of the protein that gave rise to these properties. The significance of this work is that it allows MD simulations to better fulfill

their claim of providing exquisite details about systems, often unavailable from experiment, and to do so using thermodynamically exact relationships and approaches.

As mentioned in the introduction, obtaining thermodynamically rigorous partial molar properties for species whose concentrations are finite is trivial by computer simulation. However, finite concentrations are typically not feasible for biological system simulations. Thus, it is the infinitely dilute limit that is of primary interest for this work. The approaches we have found in the literature to calculate all of the above properties are reasonable and are often rather involved. In comparison, we show here that it is often simpler to use a thermodynamically rigorous apparent molar approach.

7.2 Theory

As discussed in the Introduction, obtaining partial molar properties is, in principle, straightforward when all species are at finite concentrations. For a multi-component system of $1 \equiv$ solvent, $2 \equiv$ solute, $3, \ldots, n \equiv$ cosolvents, a molar property is defined as $X_m = X/N$, where X is the corresponding extensive thermodynamic property of interest and N is the total number of molecules in the system. Here we will be focused on $X = H, V, C_p, V\kappa_T,$ and $V\alpha_p$ only, where we have multiplied the κ_T and α_p by the system volume. In addition, we have also investigated results for the kinetic energy, K, but only as a test of the procedure, so as to ensure that the equipartition results were achieved. The partial molar property of 2 is then obtained from

$$\bar{X}_2\left(\{N\}, p, T\right) = \left(\frac{\partial X}{\partial N_2}\right)_{\{N\}', p, T} \qquad 7.12$$

where the $\{N\}'$ indicates all N other than N_2.

Now consider that 2 is a protein plus its counter-ions. Generally, it would not be feasible to perform simulations in which one varies the number of protein molecules, because the system sizes would become elephantine rather quickly. Furthermore, researchers are often more interested in the infinitely dilute system where the inclusion of protein-protein interactions would be undesirable. A finite difference approach can be used if N_2 is set equal to zero and to one in two separate simulations where $\{N\}'$ is held fixed. As mentioned in the Introduction, this finite difference calculation has been used for calculations of a LJ sphere's volume and internal energy, for determining protein volumes, and for determining lipid aggregate enthalpies. It has also been noted that such an approach can suffer from large statistical errors.

We instead have constructed a series of infinitely dilute systems in which we vary x_2 by changing N_1, where N_2 is always one. If we choose systems such that even the smallest system contains water that is beyond the "sphere of influence" of the protein *i.e.*, some waters exhibit the thermodynamic properties of bulk water, then the addition of more waters will correspond to a simple dilution process, and the corresponding plot of X_m *vs.* x_2 will be linear. An example plot (created using real data), which illustrates the method, is shown in Figure 7.1. Due to the linearity, it is easy to fit X_m. The line is given by

$$X_m = X_m^* + ax_2 \qquad\qquad 7.13$$

where X_m^* is the molar property of pure 1 (always water here), and a is the slope. After taking the partial derivative of Equation 7.13 with respect to N_2, multiplying both sides by N, and then taking $\lim\limits_{x_2 \to 0}$, one finds that

$$a = \bar{X}_2^\infty - X_m^* \qquad\qquad 7.14$$

Thus, a linear regression of X_m provides a and subsequently $\bar{X}_2^\infty$, the thermodynamicall rigorous infinitely dilute partial molar property of the protein plus its counter-ions. Rearrangement of Equation 7.13, after substituting Equation 7.14 in for a, indicates that this method is equivalent to the apparent molar approach often used experimentally,

$$\bar{X}_2^\infty = \langle X \rangle - X_m^* N_1 \qquad\qquad 7.15$$

Although we have seen a few molecular dynamics studies that use this approach to calculate $\bar{V}_2^\infty$ (Devane *et al.* 2003; Mitra *et al.* 2006; Sarupria *et al.* 2010; Yu *et al.* 2010; Jiao and Smith 2011), to our knowledge it has never been presented generally or applied for the determination of higher derivatives.

In using Equation 7.13, one has a choice of fitting both X_m^* and a, or of fixing the y-intercept and only fitting the slope. We have chosen to fix the y-intercept for the following two reasons: (1) So that we could use the same values of X_m^* regardless of the protein under study and (2) because we have performed the pure water simulation for four times longer than the simulations of systems containing proteins, in an effort to increase the precision on the molar properties of water. It should be noted that one could use this method with any number of components as long as the ratio of $1:3:\ldots:n$ is held constant and only the ratio of $2:(N\text{-}1)$ varies.

The advantages of the approach include the following:

- It involves an interpolation between pure water and some value of x_2, not an extrapolation.
- It is thermodynamically exact.
- No subjective definitions are required.
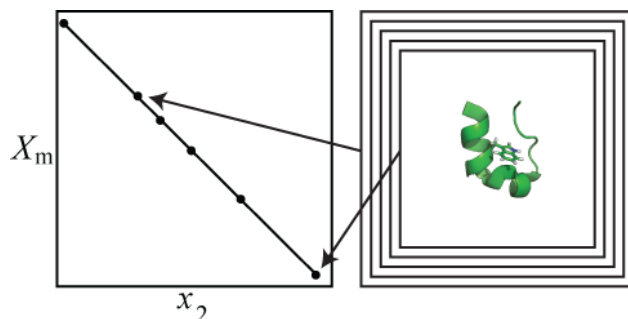
Figure 7.1 Illustration of apparent molar approach used to calculate infinitely dilute partial molar properties. A protein is simulated in various box sizes so that a plot of X_m *vs.* x_2 may be achieved from which $\bar{X}_2^\infty$ may be calculated. (Simulation boxes not to scale.)

- It can be applied to extract any partial molar property for any solute in any solvent system, regardless of the number of components.
- Partial molar properties can be assigned to specific conformations.

However, regardless of the quality of the approach, in computer simulations the ultimate arbiter remains the quality of the force field and the need for sufficient sampling of the relevant phase space. Errors in our results will be due to these factors. Specifically, there will be errors in our results due to a classical treatment of those X which have significant quantum mechanical contributions, such as the constant pressure heat capacity.

The disadvantage of the approach is that, as alluded to above, multiple simulations are required to obtain a good fit for the calculation of $\overline{X}_2^\infty$. Since it is a linear fit, in principle, only two compositions are required, namely a pure water box and one protein system. However if any noise is present in the results, it can be better identified when multiple compositions are used. Furthermore, systematic deviations from linearity in the plot of X_m *vs.* x_2 would indicate that a simulation box is too small, such that bulk water is not present. Figure 7.1 schematically illustrates the use of multiple simulation boxes.
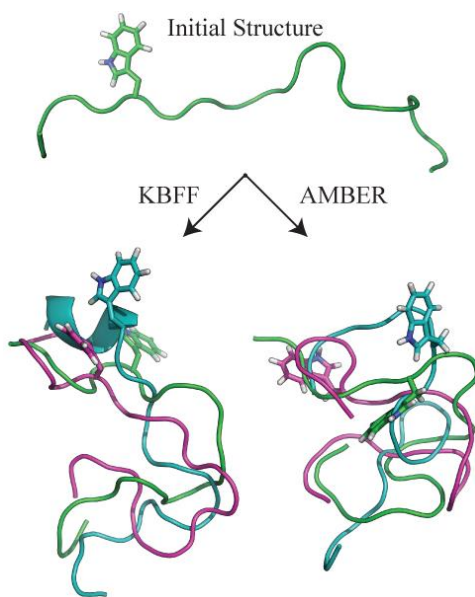
Figure 7.2 Trp-cage denatured conformations simulated. The denatured conformation was obtained after clustering a 100 ns, 500 K, *NVT* simulation using the AMBER99sb force field. Subsequent simulations of this structure using KBFF and AMBER99sb resulted in the denatured conformations shown. The three structures for each force field correspond to the structures from the three compositions (*i.e.*, the three system sizes) studied.

7.3 Methods

7.3.1 Systems Studied and Molecular Dynamics Simulations

We illustrate the method using basic pancreatic trypsin inhibitor (58 amino acids, PDB ID 5pti), abbreviatd as BPTI, and hen egg white lysozyme (129 amino acids, PDB ID 4lzt), abbreviated as HEW lysozyme. These proteins are relatively small, which allows smaller simulation boxes to be used, and experimental data is available for several of the partial molar properties for comparison with the simulated results.

The initial structures for BPTI and HEW lysozyme were each centered in a series of truncated octahedron boxes where the distances between any two parallel box faces were 8, 9, 10, 11, and 12 nm. The titratable residues were protonated to correspond to neutral pH and counter-ions were added to create neutral systems (6 Cl$^-$ for BPTI systems, 9 Cl$^-$ for HEW lysozyme systems). A pure water system, consisting of a truncated octahedron box (with the distance between any two parallel box faces equal to 7 nm) was used to calculate the corresponding molar properties of the pure solvent.

All simulations were performed using classical MD techniques at a temperature of 300 K and pressure of 1 bar unless otherwise noted. In an effort to ensure that the simulations would sample from the *NpT* ensemble and the corresponding system fluctuations would be correct, the Nosé-Hoover (chain length of one) and Parrinello-Rahman *T* and *p* baths were used. The protein and ions were modeled using the Kirkwood-Buff Force Field (http://kbff.chem.k-state.edu, KBFF) (Pierce *et al.* 2008; Ploetz *et al.* 2010a; Weerasinghe *et al.* 2010) in explicit solvent with the SPC/E water model (Berendsen *et al.* 1987). Following 25 ns of production simulation for each system size, the average Cα root mean squared deviation (RMSD) for the final BPTI (lysozyme) structures, after a translational and rotational fit to the initial structure, was 0.22 nm (0.27 nm). These values correspond to the average over the final structure from each of the different system sizes. Additional details of the simulations are provided in the Supporting Information.

Due to the promising results for BPTI and HEW lysozyme, and due to our overarching questions, we also asked whether the method could be used to distinguish between different conformations of a protein *i.e.*, if it could be used to determine $\Delta_N^D \bar{X}_2^\infty$. To test this, we simulated a native and an extended denatured conformation of trp-cage (PDB ID 2jof, construct TCb10) using the KBFF models (see Supporting Information). Trp-cage is a twenty amino acid designed mini-protein. The system setup was the same as for BPTI and HEW lysozyme, except that the distance between any two parallel box faces was 7, 8, 9, and 10 nm for the native conformation, and 8, 9, and 10 nm for the denatured conformation. The simulation protocol for trp-cage was similar to that for BPTI and HEW lysozyme. The specific trp-cage construct used here is neutral, so no counter-ions were added to the systems. Following 25 ns of production simulation, the average C$^\alpha$ RMSD for the final trp-cage native structures, after a translational and rotational fit to the initial structure, was 0.17 nm.

In addition to modeling trp-cage using the KBFF in explicit solvent with the SPC/E water model, we performed identical trp-cage simulations with the AMBER99sb force field and the TIP3P water model to allow for a force field comparison (Hornak *et al.* 2006). A TIP3P truncated octahedron box, with the distance between any two parallel box faces equal to 7 nm, was used to calculate the properties of pure water for the AMBER99sb simulations. Following 25 ns of production simulation, the average C^α RMSD for the final trp-cage native structures, after a translational and rotational fit to the initial structure, was 0.09 nm using AMBER99sb.

The initial and final denatured conformations are shown in Figure 7.2. Since the experimental $\Delta_N^D \overline{X}_2^\infty$ corresponds to a difference between properties of an ensemble of native and denatured states, our goal here was not necessarily to match the experimental value, although certainly that is of long-term interest. In this study, we instead specifically sought to establish how precisely the differences could be calculated, and to see if we achieved the correct sign and order of magnitude for the values using only two conformations.

7.3.2 Extracting X_m from Simulations

The properties U_m, K_m, H_m, and V_m were extracted from the Gromacs energy file since they simply correspond to time averages of system properties. The other properties ($C_{p,m}$, κ_T, and α_p) were calculated from their bulk system fluctuations definitions in Equation 7.7. The standard deviation for each property was calculated using block averages where the block size was 5 ns.

The system fluctuations had a large standard deviation (see Results). In an effort to increase the signal to noise ratio, we additionally calculated these properties using the expressions in Equation 7.6 (where the standard state superscript should be removed). To do this, each composition was run at four additional temperatures (290 K, 295 K, 305 K, 310 K) and four additional pressures (250 bar, 500 bar, 750 bar, 1000 bar) for 10 ns per new state point. The $C_{p,m}$, κ_T, and α_p were then obtained by fitting the simulated molar volumes or enthalpies to a polynomial function and taking the derivatives of the polynomial at the temperature or pressure of interest (300 K or 1 bar). Quadratic fits were used for the calculation of κ_T and α_p. Linear fits were used for the calculation of $C_{p,m}$. The standard deviations for each property using this "polynomial fitting method" were calculated using block averages where the block size was again 5 ns.

Although the production simulation lengths for the system fluctuation approach (25 ns) and polynomial fitting approach (10 ns per state point) are not equal and correspond to 25 ns per property for the system fluctuation approach and 50 ns per property for the polynomial fitting approach, the reported error estimates (see Results) indicate that the polynomial fitting method greatly reduced the noise beyond the type of noise reduction one would see by doubling the simulation length alone.
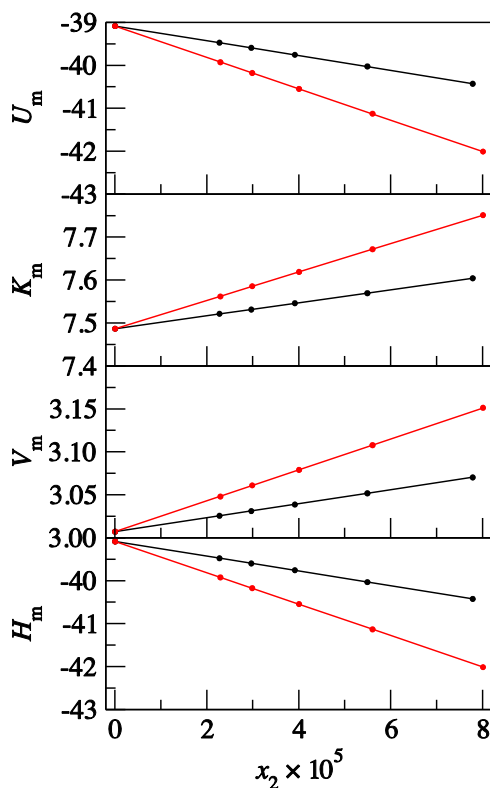
Figure 7.3 System properties used to calculate the infinitely dilute partial molar properties of basic pancreatic trypsin inhibitor (BPTI) and hen egg white (HEW) lysozyme in pure water at 300 K and 1 bar according to Equation 7.14. Black: BPTI. Red: HEW lysozyme. Circles: The property at each composition. Lines: Linear fits (fixed *y*-intercept) through the set of points. The properties shown correspond to time averages from the simulations: the molar internal energy, U_m (kJmol^{-1}), kinetic energy, K_m (kJmol^{-1}), volume, V_m (nm^3×10^2), and enthalpy, H_m (kJmol^{-1}).

7.4 Results/Discussion

Figure 7.3 shows that ensemble average properties were well behaved for both BPTI and HEW lysozyme. In contrast, there was significant noise in the three properties that are obtained from fluctuating properties. This is clearly illustrated by the scattering of the filled points around the solid lines in Figure 7.4.

It is well known that the isothermal compressibility of proteins is less than that of pure water. Lines with negative slopes in the top panel of Figure 7.4 correspond to protein partial molar isothermal compressibilities that are less than the isothermal compressibility of pure water. If the slope is too negative (see *e.g.*, the system fluctuation result for BPTI), then the compressibility of the protein will actually become negative (Table 7.2). It is known that the isothermal compressibilities of zwitterionic amino acids are negative. So, in general, this is not an impossible result. However, most reported isothermal compressibilities of proteins are positive.
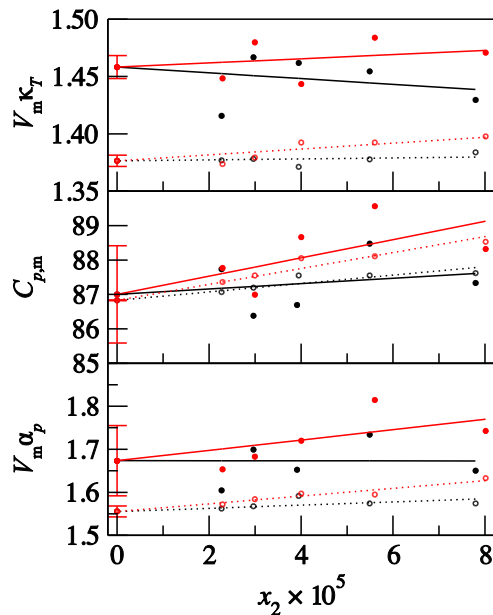
Figure 7.4 System properties used to calculate the infinitely dilute partial molar properties of basic pancreatic trypsin inhibitor (BPTI) and hen egg white (HEW) lysozyme in pure water at 300 K and 1 bar according to Equation 7.14. Black: BPTI. Red: HEW lysozyme. Results from the bulk system fluctuations are shown as solid lines with filled circles, while results from the polynomial fitting method are shown as dotted lines with open circles. The property at each composition is shown as a circle, and the linear fits through the set of points, in which the y-intercept was fixed, is shown as a line. The top panel corresponds to $V_m\kappa_T$ (nm^{3}bar$^{-1}\times10^6$), the middle panel to $C_{p,m}$ (Jmol^{-1}K^{-1}), and the bottom panel to $V_m\alpha_p$ (nm^{3}K$^{-1}\times10^5$).

Tables 7.1 and 7.2 show the values and standard deviations of the $\overline{X}_2^\infty$ obtained for BPTI and HEW lysozyme. While the internal energy, kinetic energy, and enthalpy of the proteins are not comparable to experiments, it may be of interest to study trends in these properties with *e.g.*, protein mass for a given force field. One application of such a study might be that these trends could provide an additional metric for the characterization of protein force fields. It is possible that such a characterization, for a given force field, could be used to distinguish between folded and unfolded conformations using thermodynamic, rather than structural, criteria. The simulated protein kinetic energy agreed with the predicted equipartition result in every case, calculated as $\overline{K}_2^\infty = \frac{1}{2}k_BT(3N_{atoms}-N_{bonds})$, where N_{atoms} is the number of protein + ion atoms and N_{bonds} is the number of constrained bonds in the protein.

The simulated BPTI volume, 8.164(8) nm^3, was larger than the value calculated based upon the amino acid composition, 7.8 nm^3 (Squire and Himmel 1979). This difference of 0.364 nm^3 corresponds to approximately 12 waters (the volume of an SPC/E water molecule is ~0.03 nm^3). The simulated HEW lysozyme volume, 18.00(3) nm^3, was larger than the experimental volume, 16.92(2) nm^3 (experimental measurement was of isoionic lysozyme in completely deionized and distilled water at 25 °C) (Gekko and Noguchi 1979). This difference of 1.08 nm^3 for HEW lysozyme corresponds to approximately 36 waters. For both BPTI and HEW lysozyme, the disagreement with the literature results is partly due to the inclusion of counter-ion volumes in our simulated values (6 for BPTI, 9 for HEW lysozyme). For HEW lysozyme, this is clearly not the only issue.

Table 7.1 $\bar{K}_2^\infty$, $\bar{H}_2^\infty$, and $\bar{V}_2^\infty$ of native BPTI and Lysozyme in pure water at 300 K and 1 bar

		BPTI		Lysozyme	
		Sim.	Exp./Pred.*	Sim.	Exp./Pred.*
$\bar{K}_2^\infty$	kJ/mol×10⁻³	1.515(8)	1.511*	3.309(8)	3.306*
$\bar{H}_2^\infty$	kJ/mol×10⁻⁴	-1.725(1)		-3.655(5)	
$\bar{V}_2^\infty$	nm³	8.164(8)	7.8	18.00(3)	16.92(2)

Error estimates are reported in parentheses and correspond to the standard deviation from 5 ns block averages. The "experimental" BPTI volume is actually a calculation based upon BPTI's amino acid composition (Chiou *et al.* 1979). The experimental HEW lysozyme volume was measured using a solution of isoionic lysozyme in distilled and deionized water (Gekko and Noguchi 1979). Simulations were performed using the KBFF force field. For reference, the volume of a single SPC/E water molecule is ¬0.03 nm³ and 1 nm³/molecule is 602.2 cm³/mol.

Table 7.2 $\bar{C}_{p,2}^\infty$, $\bar{\kappa}_{T,2}^\infty$, and $\bar{\alpha}_{p,2}^\infty$ of native BPTI and Lysozyme in pure water at 300K and 1 bar

		BPTI			Lysozyme		
		S.F.	P.F.	Exp.	S.F.	P.F.	Exp.
$\bar{C}_{p,2}^\infty$	kJ/mol/K	7(15)	11.2(4)	9.5	33(38)	23.39(9)	17.97, 21.22
$\bar{\kappa}_{T,2}^\infty$	bar⁻¹×10⁶	-25(32)	5(2)	*O*(1-10)	8(13)	12(4)	7.73
$\bar{\alpha}_{p,2}^\infty$	K⁻¹×10⁴	0(22)	4(4)	6	6(12)	5(1)	4.26

S.F.: System Fluctuations; P.F.: Polynomial Fitting. Error estimates are reported in parentheses and correspond to the standard deviation from 5 ns block averages. Experimental BPTI heat capacity value from Makhatadze *et al.* (1993), and thermal expansion value from Lin *et al.* (2002). Experimental Lysozyme heat capacity values from Gekko and Noguchi (1979) and Yang and Rupley (1979), isothermal compressibility value from Gekko and Hasegawa (1986), and thermal expansion value from Gekko and Noguchi (1979). Simulations were performed using the KBFF force field.

Using the system fluctuation method, the fluctuating properties for both proteins displayed too much noise to determine if there was good agreement with experiment or not (Table 7.2). To increase the signal to noise ratio, we used a second approach, in which we looked at the system volume and enthalpy as a function of pressure and or temperature and calculated the κ_T, $C_{p,m}$, and α_p using their partial derivative definitions instead of their fluctuation definitions (see Methods). This greatly reduced the error (shown visually in Figure 7.4 and tabulated in Table 7.2).

The BPTI and HEW lysozyme $C_{p,m}$ over-predict the experimental values. Since our simulations were classical and there are significant quantum corrections to the heat capacity, perfect agreement with experiment, if achieved, may have been for the wrong reasons. Specifically, real molecules store energy in low frequency bonds, whereas all bonds were constrained in our simulations. Secondly, in real molecules only the low frequency bond angles would store energy, whereas all angles (except those in water) are flexible and store energy in our simulations. Quantum corrections could be made (Berens *et al.* 1983), but they are difficult for proteins and have not been made here.

The thermal expansion and isothermal compressibility were in reasonable agreement with experiment, but displayed large error bars. The magnitude of the errors on the fluctuating properties is a downside of our approach that has not been observed using more subjective methods (Dadarlat and Post 2001, 2006).

Literature values are available for all of the pure water (SPC/E) properties. However, since it is critical that all of the simulation conditions are consistent, we simulated a pure SPC/E system instead of simply quoting previously reported SPC/E properties. The reason for the difference between the pure SPC/E values obtained using the system fluctuation

method versus the polynomial fitting method for both the isothermal compressibility and thermal expansivity is currently unknown (Figure 7.4). However, the value obtained using the polynomial fitting method is in better agreement with experiment [polynomial fitting: $4.48(2) \times 10^{-5}$ bar^{-1}, system fluctuation: $4.75(3) \times 10^{-5}$ bar^{-1}, experiment: 4.525×10^{-5} bar^{-1}].

The results for the trp-cage simulations with both force fields and with both methods (the system fluctuation and the polynomial fitting method) are summarized in Tables 7.3 and 7.4. As shown in Table 7.3, the KBFF results suggest that the native state has a larger volume than the denatured state by a value of approximately one water molecule or ~1% of the protein volume, regardless of whether or not the denatured conformation was position restrained. This is the typical sign observed for the volume change of a protein upon denaturation. In contrast, the AMBER99sb simulations produced no statistically significant volume change.

The KBFF simulation produced the correct sign for the change in enthalpy upon denaturation. The enthalpy of the denatured state was more positive, or unfavorable, than the enthalpy of the native state, as expected for heat denaturation. The average KBFF result was in better agreement with experiment when the position restraints were removed, although the difference was not statistically significant and the native state was slightly over-stabilized in both cases. Since the starting conformation for both simulations was taken from a 500 K simulation using the AMBER99sb force field (see SI text), better agreement with experiment may be achieved if the KBFF simulations were extended to allow for more conformational rearrangements. The AMBER99sb result was within the error of the experimental value with or without position restraints on the denatured conformation.

Globular protein heat capacity changes upon denaturation are typically positive; a positive (negative) change is considered to be due to apolar (polar) solvation upon denaturation. Since trp-cage is considered globular and its name stems from its buried tryptophan side chain, one might predict that the sign of $\Delta_N^D \bar{C}_{p,2}^\infty$ would be positive. Indeed, the experimental $\Delta_N^D \bar{C}_{p,2}^\infty$ for the original trp-cage construct (TC5b) was reported to be small and positive (0.3 ± 0.1 kJmol^{-1}K^{-1}). However, the experimental results of Barua *et al.* were small and negative ($\Delta_N^D \bar{C}_{p,2}^\infty = -0.2 \pm 0.1$ kJmol^{-1}K^{-1}) for the trp-cage construct simulated here (TC10b), indicating that considering tryptophan exposure is not enough to predict the correct sign.

While the statistical uncertainty in our results was too great to determine whether or not the correct sign was obtained, our results may suggest that the KBFF and AMBER99sb simulations provide opposite signs (negative and positive, respectively). It is encouraging that the KBFF simulations may be able to detect this atypical $\Delta_N^D \bar{C}_{p,2}^\infty$ sign. The positive sign for the AMBER results reported here agrees with the recent 1 μs per replica REMD simulations of the TC10b construct by English and Garcia, who also used the AMBER99sb/TIP3P force fields and reported $\Delta_N^D \bar{C}_{p,2}^\infty = 0.35 \pm 0.51$ kJmol^{-1}K^{-1}.

The results for the isothermal compressibility and thermal expansion changes upon denaturation are too noisy to determine the simulated sign; however, the error was greatly reduced using the polynomial fitting method.

Visualization of the denatured trp-cage trajectories showed that the KBFF and AMBER99sb denatured simulations without position restraints behaved differently from each other. The KBFF denatured trp-cage conformation remained more fully extended throughout the 25 ns, whereas the AMBER denatured trp-cage collapsed quickly and remained collapsed for most of the simulation (although it became more extended at the very end of the simulation). The final snapshot of the denatured simulations is shown in Figure 7.2. Despite this, the two force fields gave similar enthalpy differences [Experiment: $\Delta_N^D \bar{H}_2^\infty = 65(2)$ kJmol^{-1}, KBFF: $\Delta_N^D \bar{H}_2^\infty = 90(20)$ kJmol^{-1}, AMBER99sb: $\Delta_N^D \bar{H}_2^\infty = 50(20)$ kJmol^{-1}]. Certainly, the simulations did not access all the denatured conformations with only 25 ns of production, so it was remarkable that the $\Delta_N^D \bar{H}_2^\infty$ values were so close to the experimental value.

To test whether or not the different denatured conformations that were sampled in the simulations of trp-cage with these two different force fields accounted for their different enthalpy differences, we ran a new set of denatured simulations for both FFs using soft harmonic position restraints (100 kJmol^{-1}nm^{-2}) on the C$^\alpha$ atoms of trp-cage. The average position restraint energy was small, approximately 2 Jmol-1, so we did not repeat the simulations of the native trp-cage with position restraints. Thus, we assumed that any artifacts from the addition of the position restraints were negligible when calculating the $\Delta_N^D \bar{X}_2^\infty$ using simulations of the native structure without position restraints. The $\Delta_N^D \bar{X}_2^\infty$ results were the same with and without position restraints to within the statistical error. This is consistent with a two-state folding pathway in which the energy landscape is relatively flat for the (high free energy) unfolded conformations and drops down into one minimum for the folded conformation. Remarkably, the enthalpy differences remained essentially correct when calculated from difference between the native conformation and just one restrained denatured conformation.

Table 7.3 $\bar{H}_2^\infty$ and $\bar{V}_2^\infty$ of native (N) and denatured (D) Trp-Cage in pure water at 300 K and 1 bar

		KBFF $\bar{X}_2^\infty$			AMBER $\bar{X}_2^\infty$			Exp.
		N	D	Δ_N^D	N	D	Δ_N^D	$\Delta_N^D \bar{X}_2^\infty$
No Position Restraints:								
$\bar{H}_2^\infty$	kJ/mol×10^{-3}	-5.120(8)	-5.03(2)	0.09(2)	-1.817(9)	-1.77(2)	0.05(2)	0.065(2)
$\bar{V}_2^\infty$	nm^3	2.524(3)	2.496(7)	-0.028(8)	2.375(8)	2.377(6)	0.00(1)	
Position Restraints on Denatured Conformation:								
$\bar{H}_2^\infty$	kJ/mol×10^{-3}		-5.00(1)	0.12(1)		-1.74(2)	0.08(2)	0.065(2)
$\bar{V}_2^\infty$	nm^3		2.48(2)	-0.04(2)		2.37(1)	0.00(1)	

Error estimates are reported in parentheses and correspond to the standard deviation from 5 ns block averages. The kinetic energy contributions agreed with the equipartition results and were the same, within statistical uncertainty, for the native and denatured conformations. Experimental $\Delta_N^D \bar{H}_2^\infty$ from Barua *et al.* (2008).

Table 7.4 $\bar{C}_{p,2}^{\infty}$, $\bar{\kappa}_{T,2}^{\infty}$, and $\bar{\alpha}_{p,2}^{\infty}$ of native (N) and denatured (D) Trp-Cage in pure water at 300 K and 1 bar

		KBFF $\bar{X}_2^{\infty}$			AMBER $\bar{X}_2^{\infty}$		
		N	D	Δ_N^D	N	D	Δ_N^D
$\bar{C}_{p,2}^{\infty}$	kJmol⁻¹K⁻¹						
Exp.				-0.21			-0.21
S.F.		14(21)	23(18)	9(28)	19(6)	18(16)	-1(17)
P.F.		3.6(1)	3.5(5)	-0.1(5)	5.8(5)	6.4(1)	0.6(5)
P.F. P.R.D.		3.6(1)	3.23(5)	-0.4(2)	5.8(5)	6.1(4)	0.3(6)
P.F.(1 μs)					6.2(1)		
$\bar{\kappa}_{T,2}^{\infty}$	bar⁻¹×10⁶						
S.F.		9(72)	-100(100)	-100(100)	-78(168)	65(160)	144(233)
P.F.		3(8)	33(11)	29(13)	-16(10)	-12(9)	-9(10)
$\bar{\alpha}_{p,2}^{\infty}$	K⁻¹×10⁴						
S.F.		30(40)	20(40)	-10(60)	-10(50)	40(70)	50(90)
P.F.		6(6)	8(15)	2(16)	7(2)	6.7(2)	-1(2)
P.F. P.R.D.		6(6)	8(5)	2(8)	7(2)	6(3)	2(4)
P.F.(1 μs)					8.6(4)		

Error estimates are reported in parentheses and correspond to the standard deviation from 5 ns block averages, excluding the 1 μs simulation, which corresponds to the standard deviation on ~333 ns block averages. For the 1 μs simulations, the error on the y-intercept (pure water value) was taken as zero. S.F.: System Fluctuations, P.F.: Polynomial Fitting. P.R.D.: Position Restrained Denatured conformation. Experimental $\Delta_N^D\bar{C}_{p,2}^{\infty}$ from Barua *et al* (2008).

7.4.1 Error Analysis

Lin *et al*. and references therein have reported that typical magnitudes for the volume change associated with protein unfolding were <0.5% of the protein volume (Lin *et al.* 2002). Using pressure perturbation calorimetry, Lin determined the $\Delta_N^D\bar{V}_2^{\infty}$ was -0.06% of the volume of native BPTI in water at pH 4 and -0.08% of the volume of native HEW lysozyme in water at pH 2.5 (Lin *et al.* 2002). The errors on our $\Delta_N^D\bar{V}_2^{\infty}$ for BPTI and HEW lysozyme were 0.1% and 0.2%, thus simulations longer than ours would be necessary to calculate $\Delta_N^D\bar{V}_2^{\infty}$ with precision for these proteins. The corresponding $\Delta_N^D\bar{\alpha}_{p,2}^{\infty}$ values reported by Lin were 0.5×10^{-4} K⁻¹ and 1.4×10^{-4} K⁻¹ for BPTI and HEW lysozyme, respectively. To calculate $\Delta_N^D\bar{\alpha}_{p,2}^{\infty}$ with precision using the polynomial fitting method, we would need to reduce the error on $\bar{\alpha}_{p,2}^{\infty}$ by a factor of eight for BPTI, which would require simulations approximately 64 times as long *i.e.*, 640 ns at each temperature, assuming that the error $\propto 1/t_{\text{sim}}^{1/2}$. With the current simulation time, the $\bar{\alpha}_{p,2}^{\infty}$ value for HEW lysozyme was at the minimal precision necessary for detection of $\Delta_N^D\bar{\alpha}_{p,2}^{\infty}$, if simulations of denatured conformations were performed.

In a compendium of protein thermodynamic properties, Makhatadze reported that BPTI $\Delta_N^D\bar{H}_2^{\infty} = 130$ kJmol⁻¹ and $\Delta_N^D\bar{C}_{p,2}^{\infty} = 3.0$ kJK⁻¹mol⁻¹ at 298 K (Makhatadze 2007). These numbers are larger than the uncertainty on the native enthalpy and native heat capacity calculated here, 10 kJmol⁻¹ and 0.4 kJmol⁻¹K⁻¹, respectively. The same source reports

that HEW lysozyme $\Delta_N^D \bar{H}_2^\infty = 242$ kJmol^{-1} and $\Delta_N^D \bar{C}_{p,2}^\infty = 9.1$ kJmol^{-1}K^{-1} at 298 K (Makhatadze 2007). Again, these numbers are larger than the uncertainty on the native HEW lysozyme enthalpy and heat capacity calculated here, 50 kJmol^{-1} and 0.09 kJmol^{-1}K^{-1}, respectively. Thus, we predict that similar length simulations of either of these proteins in their denatured conformations would allow for $\Delta_N^D \bar{H}_2^\infty$ and $\Delta_N^D \bar{C}_{p,2}^\infty$ to be calculated with statistical precision. The challenge in these situations, heightened as the number of residues in the protein of interest increases, would be in choosing protein conformations that were "representative" of the denatured state. Although the trp-cage results from these simulations surprisingly showed that the values were not sensitive to the protein conformation, much more work on different proteins would be necessary before confirming that behavior as a general conclusion. Indeed, one would anticipate that it is not generally true. However, Shaw and coworkers recently reported similar observations for ubiquitin (Piana *et al.* 2013).

We could have simulated all of the systems for even longer periods of time. However, even a multiplication of the production simulation length by 4 would only result in a noise reduction by a factor of 2 (assuming that the error $\propto t_{sim}^{-1/2}$), which would not be sufficient in most cases to determine these properties with precision. So, in an effort to estimate how long one would need to simulate trp-cage in order to obtain the heat capacity and thermal expansion with sufficient precision, we only extended the AMBER99sb native trp-cage simulations to 1 μs at temperatures of 290 K, 295 K, 300 K, 305 K, and 310 K. Since we had already established the linear behavior of X_m *vs.* x_2 for the 7, 8, 9, and 10 nm box sizes, we only extended the simulations for the 8 nm box. Extension of a simulation from 10 ns to 1 μs is an increase of t_{sim} by a factor of 100, thus we would have predicted a reduction in the noise by a factor of 10. As shown in Table 7.4, the errors were actually reduced by only a factor of 5 for $\bar{C}_{p,2}^\infty$ and 2.5 for $\bar{\alpha}_{p,2}^\infty$.

7.5 Conclusion

We have investigated the ability of a simple method to extract infinitely dilute partial molar properties of biomolecules from molecular dynamics simulations using basic pancreatic trypsin inhibitor and hen egg white lysozyme as our test case proteins. We then used the method to distinguish between the protein volume and enthalpy of a native and a denatured trp-cage conformation with sufficient precision. The method is most similar to that used recently by Smit to determine the enthalpy of a lipid aggregate, and to Shing and Chung's work in the late 1980s for the infinitely dilute partial molar volume and enthalpy of a Lennard-Jones solute, but to our knowledge it has never been used or presented for any general solute and for any general partial molar property. The strengths of the approach are that it is thermodynamically exact, no ambiguous parameters are introduced, and knowledge of the specific conformations that contributed to the output values is retained. The method is general and this approach may be used to calculate any partial molar property of any solute in any solution. It could be employed in the future as a test of the quality of a FF.

The current limitation of the approach is the high level of noise in the isothermal compressibility and thermal expansion coefficient results, and the moderate level of noise in the isobaric heat capacity results. To increase the signal to noise ratio, a polynomial fitting method was used. Clearly, the system fluctuation approach was not the

method of choice for the calculation of the fluctuating properties, and the polynomial fitting method provided higher precision results. However, we were still unable to precisely distinguish between native and denatured trp-cage conformations' heat capacities, thermal expansions, and isothermal compressibilities.

Based upon our results, we predict that for a protein the size of HEW lysozyme (129 residues), the following amounts of time are needed to calculate specific thermodynamic properties with statistical precision: $\Delta_N^D \bar{V}_2^\infty$, ~200 ns; $\Delta_N^D \bar{H}_2^\infty$, achieved necessary precision using 25 ns simulation; $\Delta_N^D \bar{C}_{p,2}^\infty$, achieved necessary precision using 10 ns simulations at each of 5 temperatures; $\Delta_N^D \bar{\kappa}_{T,2}^\infty$, difficult to predict due to large noise; $\Delta_N^D \bar{\alpha}_{p,2}^\infty$, achieved necessary precision using 10 ns simulations at each of 5 temperatures. Currently the methods that rely upon subjective definitions, if they can be trusted, are more competitive than our approach for the isothermal compressibility due to the high amounts of noise in the current method. Continued increases in computational power should make the method presented here more usable, with ease, over time.

Lastly, the method presented here does not provide a means of directly decomposing the thermodynamic properties into specific contributions, from specific surface groups for example. The solute-solute and solute-solvent energy terms decay as a function of distance away from the solute, which makes them directly calculable. Since the solvent-solvent contribution does not, it must be solved for indirectly. This is a commonly encountered issue see *e.g.*, McCammon and co-workers (Setny *et al.* 2010). We are currently working to develop and test a method to calculate all of the above properties that uses expressions from Fluctuation Solution Theory and is decomposable, analogous to the group contribution for partial molar volumes already achieved using traditional Kirkwood-Buff theory.

This second approach is also thermodynamically exact and will allow for distance dependent infinitely dilute partial molar properties to be determined from a single simulation of a protein instead of the multiple compositions recommended here to determine linearity in the X_m *vs.* x_2 plots. The approach and preliminary results using this alternative method are discussed for the volume and compressibility only in Section 7.7. The other thermodynamic properties are our ongoing work and will be presented later; however it should be noted that the expressions have already been derived by Smith (Jiao and Smith 2011; Ploetz and Smith 2013).

7.6 Supporting Information: Additional details of the simulation methods

7.6.1 BPTI and HEW lysozyme

All the BPTI and HEW lysozyme simulations were performed using the GROMACS 4.0.5 simulation package (Hess *et al.* 2008). Each system was minimized for over 1,000 steps using the steepest descent algorithm and then equilibrated for 100 ps at $p = 1$ bar and $T = 100$ K, 100 ps at 1 bar and 200 K, and 100 ps at 1 bar and 300 K, with position restraints on all the heavy atoms during each equilibration. 25 ns of production simulation (100 ns for pure water) without position restraints was performed for each system at 300 K and 1 bar in the isothermal-isobaric (*NpT*) ensemble. In an effort to ensure that the simulations would sample from the *NpT* ensemble and the system fluctuations

would be correct, the Nosé-Hoover (chain length of one) and Parrinello-Rahman T and p baths were used, with relaxation times of 0.5 ps and 2.5 ps, respectively, as advised by the Gromacs 4.5 manual. A 4.5×10^{-5} bar^{-1} compressibility was used for all systems. Periodic boundary conditions and the minimum image convention were employed. All bond lengths were constrained using the Settle (Miyamoto and Kollman 1992) and LINear Constraint Solver (LINCS) (Hess *et al.* 1997) algorithms for water and non-water molecules, respectively. The use of bond constraints allowed for a two fs time step to be used for the integration of the equations of motion, which was performed using the Leap-Frog algorithm (Hockney *et al.* 1974). The particle-mesh Ewald technique was used to calculate electrostatic interactions with cutoff distances of 1.0 nm and 1.5 nm for the real space electrostatic and van der Waals interactions, respectively, a convergence parameter of 3.123 nm^{-1}, cubic interpolation, a maximum fast Fourier transform grid spacing of 0.12 nm for the reciprocal space sum, and tinfoil boundary conditions (Darden *et al.* 1993). The grid-based neighbor list was updated every ten steps.

7.6.2 Trp-cage

We simulated two conformations of trp-cage using KBFF. We chose the first NMR structure in PDB ID 2jof, a stabilized mutant of the original sequence, as our initial structure for the native conformation. The starting denatured conformation for the KBFF simulation was selected from a 100 ns canonical (*NVT*) simulation at 500 K ran using the AMBER99sb force field. The reason for using AMBER99sb was that this trajectory had previously been generated for a separate project. The conformations were clustered using the Gromacs program g_cluster and, specifically, the algorithm of (Daura *et al.* 1999). The initial structure was taken from a cluster that was nearly fully extended and had little secondary structure.

The denatured trp-cage conformations were equilibrated for 100 ps at 300 K followed by 25 ns of production. The native conformations were equilibrated for 100 ps at p = 1 bar and T = 100 K, 100 ps at 1 bar and 200 K, and 5 ns at 1 bar and 300 K, with position restraints on all the heavy atoms during each equilibration. They were then equilibrated for 5 ns at 300 K and 1 bar with position restraints on the α-carbons and 5 ns at 300 K and 1 bar without position restraints, followed by 25 ns of production simulation without position restraints.

For the AMBER99sb + TIP3P trp-cage simulations, Gromacs version 4.5.3 was used, the particle-mesh Ewald cutoff distance was 1.2 nm for both the real space electrostatic and van der Waals interactions, and the convergence parameter was 2.603 nm^{-1}. The pure TIP3P production simulation was 100 ns. All other details were the same as those for the KBFF + SPC/E simulations.

7.7 Appendix: Towards a Fluctuation Solution Theory-Based Approach for the Calculation of Infinitely Dilute Partial Molar Properties of Proteins from Computer Simulation[*]

7.7.1 Fluctuation Solution Theory (FST)

Here we provide progress towards a combination of Fluctuation Solution Theory (FST) and computer simulation studies to access the thermodynamic properties of infinitely dilute proteins. Just like the apparent molar approach presented in the main text of this chapter, the FST-based approach is exact; however, using the FST-based approach properties may be obtained from a single simulation instead of multiple simulations, which was recommended for the apparent molar approach. Furthermore, the FST-based expressions allow for a decomposition of the properties into spatial contributions, whereas the apparent molar approach provided no such decomposition. Like the apparent molar approach, the FST-based approach may be used to obtain conformation-specific properties (not shown here).

FST was described thoroughly in Chapters 4 and 5 and will not be completely repeated here. Our previous work has involved the use of FST to understand the properties of solution mixtures (Gee and Smith 2009; Jiao and Smith 2011; Ploetz and Smith 2011b). In particular, the combination of FST and computer simulation provides a very powerful analysis tool. The two most important advances related to the present work involve the inclusion of local energy fluctuations in a manner that promotes comparison with experiment, and the development of expressions for the study of conformational transitions and molecular association (Pierce *et al.* 2008; Gee and Smith 2009; Jiao and Smith 2011). In all cases, the aim has been to relate second and third derivatives of the Gibbs free energy to second and third derivatives in the grand canonical ensemble, which then provide local descriptions of the particle and energy distributions. Recall that the open distributions and fluctuations relevant to FST are considered to represent local regions of a much larger bulk solution (see Figure 7.5). These approaches have recently formed the basis of a comprehensive review (Ploetz and Smith 2013), and an edited book outlining the details and applications of the current approach (Smith *et al.* 2013).

7.7.2 Expressions for the Pseudo Volume and Compressibility

In this Appendix, we specifically focus on the infinitely dilute partial molar volume and isothermal compressibility to the neglect of the other properties of interest (enthalpy, heat capacity, thermal expansion), but all of the expressions have recently been established (Jiao and Smith 2011; Ploetz and Smith 2013) and calculations of the other properties from computer simulations are ongoing and will be presented in the future.

FST can be applied to any number of components at any composition. However, if we consider an infinitely dilute biomolecule (2) in a primary water solvent (1) that undergoes a transition from a native (N) to a denatured (D) state with an equilibrium constant $K = \rho_D/\rho_N$, where $\rho_i = <N_i>/V$, then the derivatives of the equilibrium constant under the

---

[*] For this Appendix, we recommence with the notation used throughout the rest of this dissertation that was disbanded solely for the main text of Chapter 7. That is, a superscript asterisk will again indicate a pseudo property and a superscript o will return to denoting a pure property. Standard state properties are not necessary for this Appendix, due to the use of the pseudo properties in FST.
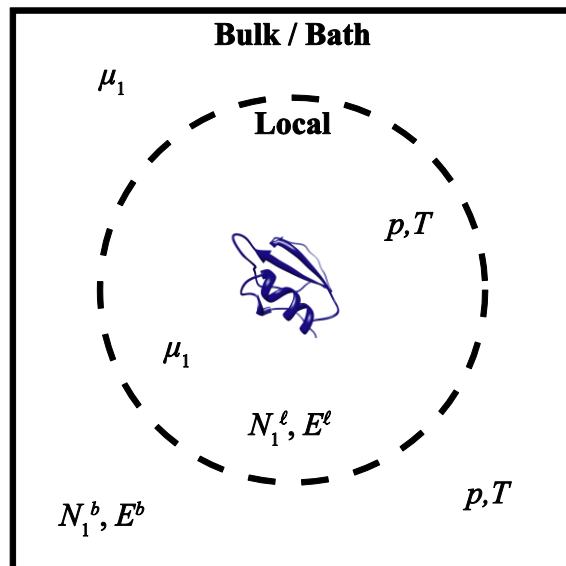
Figure 7.5 An illustration of the FST approach for the analysis of simulation data for an infinitely dilute solute (2) in a mixture of solvent (1) and cosolvent (3). FST quantifies the local distribution of particles and energy in the vicinity of the protein for the conformation of interest. The local particle and energy distributions (denoted by the *l* superscript) differ from the bulk (*b*) distributions due to the influence of the solute. Here we use a protein solute, but the solute could be of any size or shape. The range of influence of the solute is indicated by the virtual dashed boundary and is generally unknown; however, this does not affect the results as long as one includes any and all deviations from the bulk distribution. The local region can be considered to be a virtual grand canonical region of the solution for which the energy and number of particles can fluctuate. These fluctuations are then related to the thermodynamic properties of the original closed system.

influence of changes in pressure (*p*) can be written as (Jiao and Smith 2011; Ploetz and Smith 2013)

$$\left(\frac{\partial \ln K}{\partial p}\right)_{\beta,m_2} = -\beta(V_D^{*,\infty} - V_N^{*,\infty}) \qquad 7.16$$

The asterisk superscript indicates a pseudo volume (*V*) of the biomolecule and refers to the concept of the pseudo chemical potential developed by Ben-Naim (Ben-Naim 2006). The above expressions are essentially standard thermodynamic relationships. The quantities appearing in Equation 7.16 can be obtained from FST and can be expressed in terms of a series of fluctuating quantities, or integrals over radial distribution functions (RDFs), $g_{ij}(r)$, defined as,

$$G_{ij} = V\left[\frac{\langle \delta N_i \delta N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - \frac{\delta_{ij}}{\langle N_i \rangle}\right] = 4\pi \int_0^\infty [g_{ij}(r) - 1]r^2 dr \qquad 7.17$$

where δ_{ij} is the Kronecker delta function and $\delta X = X - \langle X \rangle$ represents a fluctuation in the property *X*. The RDFs correspond to distributions obtained in a solution, at the composition of interest, after averaging over all the remaining molecular degrees of freedom. The G_{ij}'s are known as Kirkwood-Buff integrals (KBIs) and are the central components of KB theory.

The KBIs quantify deviations from a random distribution surrounding a central molecule summed over all space. In this respect they are more informative than the particle fluctuation formulation as they can then be decomposed and interpreted in terms of spatial contributions – using computer simulation data, for example. They clearly resemble the integrals one encounters in the study of imperfect gases and the McMillan-Mayer theory of solutions (McMillan and Mayer 1945). However, the KBIs do not form part of a virial expansion and are relevant at any solution composition – not just infinitely dilute solutes in the gas phase or a primary solvent. After multiplying the KBIs by number densities (molar concentrations), one obtains an alternative picture of the KBIs. The quantities, $N_{ij}=\rho_j G_{ij}$, are referred to as excess coordination numbers and quantify the change in the number of j molecules observed in an open volume of solution on introduction of a central i molecule, from that observed for the same volume of bulk solution in the absence of the i molecule (Hall 1971; Newman 1994). They can be used to define and develop the concepts of local solution composition and preferential interactions (Ben-Naim 1988; Marcus 2002). It should be noted that FT does not require information on the angular distributions for pairs of molecules, as these are averaged out in the above expressions.

The RDF formulation in Equation 7.17 is quite cumbersome for irregular shaped objects such as proteins. Fortunately, the fluctuations do not suffer from this drawback. In fact, at infinite dilution of a protein solute the above fluctuations become simple averages such that,

$$V_N^{*,\infty} = -G_{N1}^{\infty} = -\left(\left\langle N_1\right\rangle_N - \left\langle N_1\right\rangle_{\circ}\right)V_1^{\circ}$$

7.18

where the subscript N or o represents an ensemble average in the vicinity of the native protein, or in bulk solvent, respectively. Similar expressions can be written for the denatured state. It should be noted that each of the ensemble averages in the above expressions is actually extensive, depending on the size of the local volume of interest, but the difference between the two ensemble averages is independent of the size of the local volume as long as it is large enough that the solvent distribution is beyond the influence of the protein solute.

The above expression deserves more discussion. It is exact. It provides a thermodynamically rigorous description of the protein volumes associated with the conformational equilibrium in terms of the local distribution of water molecules, which can be determined as a function of distance from the protein center of mass (or protein surface) using computer simulation data. This volume is "measured" by determining the average number of water molecules in the local vicinity surrounding the protein and then subtracting the number of water molecules one would observe in the same volume of bulk solution in the absence of the protein, both scaled by the molar volume of pure water (V_1°). Hence, the pseudo volume of the protein is determined by the volume of water molecules surrounding it. This can also be related to the partial molar volume using standard expressions obtained from the application of FST to binary solutions (Ben-Naim 2006). This expression has been used before to help understand peptide and protein volume data (Imai *et al.* 2007; Yu *et al.* 2010). Higher derivatives of the equilibrium constant are also provided by FST. Second derivatives of ln K correspond to third derivatives of the Gibbs free energy and can be expressed in terms of third derivatives in the grand canonical ensemble. These then involve triplet particle correlations. For an infinitely dilute solute they become pair fluctuating quantities in the local vicinity of the protein and are given by (Ploetz and Smith 2013),

157

$$\left(\frac{\partial V_N^{*,\infty}}{\partial p}\right)_{\beta,m_2} = -V_N^*\kappa_{T,1}^o - \beta(V_1^o)^2\left[\left\langle\delta N_1\delta N_1\right\rangle_N - \left\langle\delta N_1\delta N_1\right\rangle_o\right] \tag{7.19}$$

where $<\delta X\delta X>_N =< X^2>_N - < X>_N^2$ represents a local fluctuation in the property X. The isothermal compressibility of the pure solvent $\left(\kappa_{T,1}^o\right)$ can also be expressed in terms of local fluctuations (Ploetz and Smith 2011b). However, this term is very small compared to the term contained in the square brackets, and could even be safely neglected under ambient conditions (Jiao and Smith 2011). Similar expressions can be written for the denatured state. Hence, the second derivative of ln K with respect to p is related to the relative fluctuations, compared to the same fluctuations in bulk solution, in the number of local water molecules, as provided by the last term in the above expression. This expression clearly resembles the bulk system expression for compressibility (Jiao and Smith 2011). They are, however, different. More importantly, the bulk expression is approximate when applied to proteins at infinite dilution, whereas the local expression is exact.

7.7.3 Initial Pressure Denaturation Studies

As the application of FST to pressure denaturation is relatively new, it is important to demonstrate that the results are indeed reasonable. Unfortunately, the simulation results will also depend on the force field employed, and there are no known results for proteins using formally correct expressions; however, we have been able to illustrate reasonable behavior. Our initial studies have focused on the compressibility of a native protein, which can be obtained directly from the pressure derivative provided in Equation 7.19.

The water distribution and local water fluctuations around native lysozyme have been analyzed and the results for the volume and compressibility of native lysozyme are presented in Figure 7.6 as a function of radial distance from the center of mass of the protein. The water distribution, g_{21} is relatively featureless and increases steadily with distance (r) until the bulk distribution is reached between 2.0-2.5 nm. This rather large distance is probably due to the spherical nature of the analysis performed here. The distribution from the surface of the protein may converge much faster and still needs to be investigated. It should be noted, however, that both types of analysis, center of mass or surface based, should produce identical overall results. However, it is safe to say that large systems are required to ensure a significant bath region beyond the influence of the protein. The simulated protein volume as a function of local volume radius (R) is also displayed in Figure 7.6. A constant value for the volume is obtained for distances beyond which the the solvent distribution reaches the bulk random distribution. The corresponding water
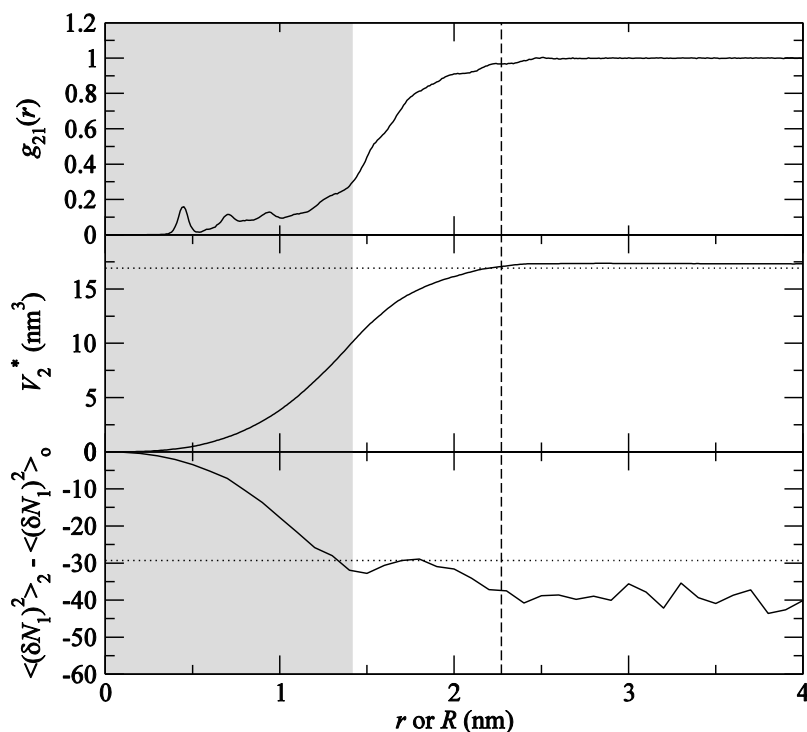
Figure 7.6 Simulated water distribution and fluctuations around native lysozyme at 300 K and 1 bar obtained using the AMBER99SB-ILDN force field. Top: The water RDF as a function of distance from the protein center of mass (r). Middle: the protein pseudo volume as a function of local solution volume radius (R). Bottom: The difference in water fluctuations as a function of local solution volume radius (R). The shaded area represents distances less than the radius of gyration, while the vertical dashed line indicates the distance of the furthest backbone atom from the center of mass. The dotted horizontal lines indicate the experimental values.

fluctuations also converge to a reasonably constant "plateau" value at a similar distance. The time convergence of the local water fluctuations is indicated in Figure 7.7. For this system it appears that at least 100-200 ns of simulation time are required to obtain converged results.

The experimental and preliminary simulated compressibilities of water and lysozyme are displayed in Figure 7.8. First, we note that the compressibility of the water model used here is quite large compared to the experimental value. This could affect the quality of the results. Nevertheless, the simulated volume and compressibility of lysozyme is in very good agreement with experiment. The low (compared to pure water) experimentally observed protein compressibility is reproduced by the current approach and simulations. We have also investigated the effect of applying restraints to the protein atoms in an attempt to elucidate the possible reasons for the low compressibility of proteins in general. The compressibilities obtained after restraining all protein atoms (PR) and all protein C^{α} atoms (PRC) are also displayed in Figure 7.8. The application of all atom restraints, which removes the protein volume fluctuations, caused a significant drop in compressibility. This difference, 1.8×10^{-5} bar^{-1}, corresponds to the intrinsic compressibility of the protein, which has been estimated between 1.0-3.0×10^{-5} bar^{-1} with a most probable value of 2.5×10^{-5} bar^{-1} (Taulier and Chalikian 2002). Hence, we are in very reasonable agreement with these estimates. Furthermore, by applying restraints to just the C^{α} atoms, one can eliminate the protein volume fluctuations while retaining the dynamics of surface sidechains. This effect seemed to be relatively small, suggesting it is probably the
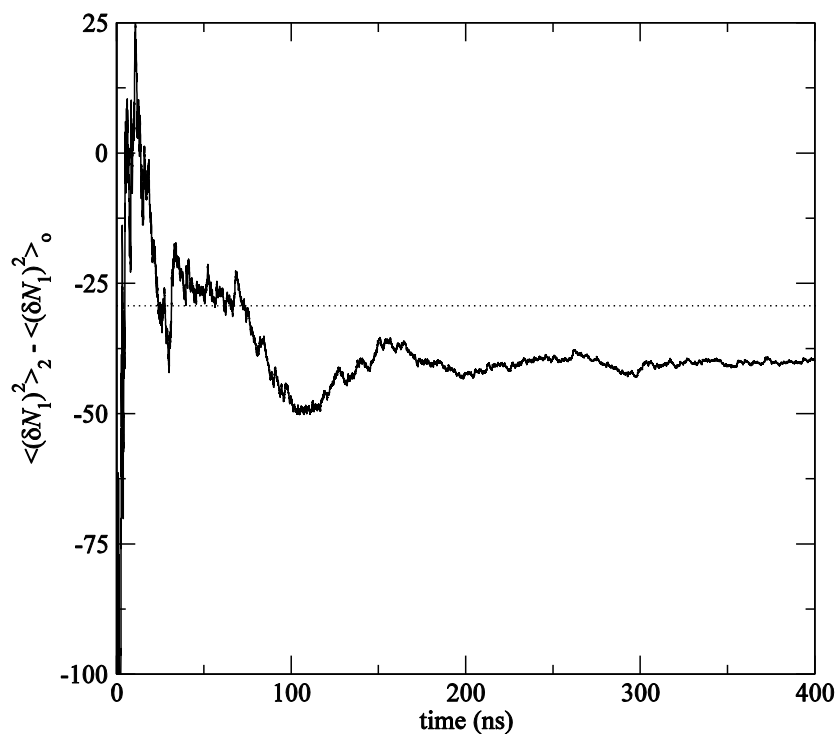
Figure 7.7 The convergence (running average) of the fluctuation difference with time for a local spherical volume of radius r = 2.5 nm. The horizontal dashed line indicates the experimental value.
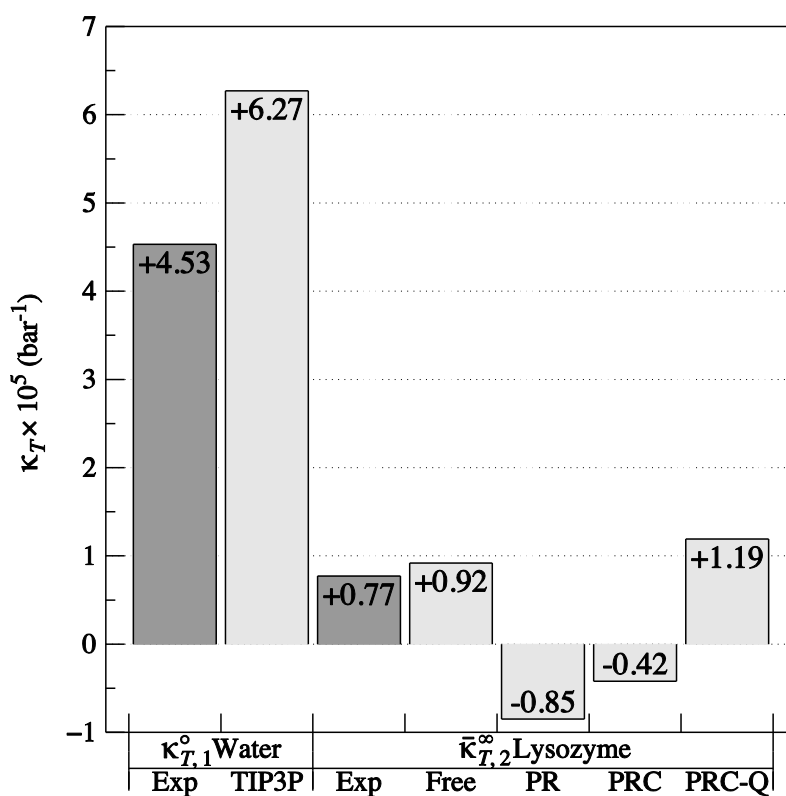
Figure 7.8 Simulated and experimental partial molar compressibilities. The simulations of lysozyme were performed using no restraints (free), with all atoms restrained (PR), just C^{α} atoms restrained (PRC), and with C^{α} atoms restrained and the charged groups neutralized (PRC-Q).

global protein volume fluctuations that dominate the intrinsic compressibility value. Finally, we also determined the effect of neutralizing all the charged residues of the protein, while applying restraints to the C^α atoms to avoid significant changes in structure (labelled PRC-Q). The change in compressibility compared to the PRC simulation suggests a contribution from the charged groups of around -1.6×10^{-5} bar^{-1}. Hence, even though we still have to perform a more rigorous statistical analysis, the picture that emerges from this combination of simulation and FT is a significant positive compressibility from the protein volume dynamics, with a counter balancing negative contribution from surface charges due to electrostriction effects. These few simulations illustrate the potential power of the approach.

We have also simulated zwitterionic glycine at infinite dilution in water. This represents an interesting system as the experimental partial molar compressibility is actually negative with a value of -4.9×10^{-5} bar^{-1}. Our simulations provided a value of $-14.4(4) \times 10^{-5}$ bar^{-1}. Hence, although the magnitude of the compressibility is significantly different from experiment, the correct sign for the compressibility was obtained. This is important as many phenomenological models often struggle to permit even the possibility of negative compressibilities, especially for small solutes. It also illustrates that the current method can be applied to molecules of any size. (As an aside, many phenomenological models would also struggle to be able to calculate the volume and/or compressibility of a denatured conformation, due to the subjective volume calculations used, whereas the FST-based method has no difficulty with the denatured state.)

The FT expressions are exact, but whether one can obtain the required properties from simulation with sufficient precision (and hopefully accuracy) still has to be established. The results provided here suggest that they can. The first derivative properties should not be a major problem as these are simple ensemble averages. However, the second derivative properties could converge slowly with simulation time as they correspond to local fluctuations. Fortunately, one can (if necessary) change the pressure and investigate how the first derivative properties respond. Furthermore, it is possible to decompose these differences into distance dependent contributions (not shown). We believe these possibilities demonstrate the power of the approach.

A series of simulations are ongoing to investigate the nature of the water and energy distributions surrounding an interesting test set of infinitely dilute proteins in water and mixed solvents. Additional simulations are being performed of different denatured conformations of ubiquitin, using the denatured state ensemble generated by Shaw and coworkers (Piana *et al.* 2013). We argue that analysis of all of these simulations will provide insights into the complete set of thermodynamic properties of interest – partial molar volumes, enthalpies, compressibilities, heat capacities, and thermal expansivities – and how these properties relate to the nature of proteins in general. The results in Figure 7.6 and Figure 7.8 indicate that these simulations are feasible. The results in Figure 7.7 suggest the simulations have to extend to 100 ns or more. The requirement of a significant bath region, as depicted in Figure 7.5, implies that these simulations need to be larger than normal to ensure the properties of a solvent bath are fully realized. We argue that this FST-based approach will be superior to the apparent molar approach presented in the body of this chapter.

161

8. Conclusions & Future Directions

Chapters 2 and 3 showed that popular biomolecular force fields are often unable to reproduce the Kirkwood-Buff integrals for important liquid mixtures. These chapters also showed that force fields may be parameterized (KBFF) to reproduce the Kirkwood-Buff integrals that are competitive with the standard force fields. Two polarizable models were compared to the (nonpolarizable) KBFF water and methanol models and were not found to be necessarily superior to KBFF. Ongoing work related to Chapter 2 involves the simulation of additional compositions for the mixtures presented, as well as simulations of the guanidinium-acetate ion pair in water for all of the force fields, followed by a KB analysis. We are additionally in the process of comparing a number of other properties for these mixtures. Future work directly related to Chapter 3 is not planned beyond the comparisons of polarizable and non-polarizable water in the ability to reproduce accurate local fluctuating quantities.

This dissertation has investigated two key advancements to KB/FST theory, housed in Chapters 4-6. First (Chapters 4 and 5), it showed how to extract particle-energy and energy-energy local fluctuations from bulk, experimental, thermodynamic data and how to simulate these quantities as well. The resulting local fluctuations can be used to help characterize the properties of solution mixtures in much the same way as the traditional KB approach. Another possible use of the experimentally derived fluctuations is for the evaluation and development of accurate force fields for molecular simulation, where the fluctuating quantities serve as additional experimental data for the parametrization procedure. Ongoing work will continue to quantify how well current force fields perform at reproducing these new properties, in an effort to see if further force field parameterization is necessary or unrequired. In addition, the approach presented here may be applied to understand the effects of temperature on chemical equilibria in terms of particle-energy and energy-energy fluctuations. We envision the application of FST to biological systems to continue, especially in the area of conformational equilibria. The natural coupling of FST and computer simulation holds great promise for future study.

Second (Chapter 6), this dissertation showed how to extract pressure and composition derivatives of the particle-particle fluctuations from bulk, experimental, thermodynamic data, and how to simulate the resulting triplet and quadruplet fluctuating quantites. In so doing, new information was obtained about the way molecules are distributed in solution. The results could help to provide a deeper understanding of solutions and solution mixtures. The only other experimental technique that we know of that provides such data for solutions is that of solution scattering studies. However, these are limited to triplet correlations and to relatively simple mixtures. No such limitation is found with the current approach. It appears that both the experimental extraction and simulation of these correlations is viable for the triplet correlations, while the quadruplet correlations may be more problematic, especially at low concentrations. Fortunately, while the current approach does not provide the experimental values of $g^{(3)}$ or $g^{(4)}$ (we know of no such approach), the thermodynamics of the solution are directly related to the integrals over these correlation functions, which are provided by the current approach. Some areas of future study involve: i) extending the current approach to include energy fluctuations; ii) developing the expressions for third derivatives of the Gibbs free energy (*i.e.*, partial

molar compressibilities, expansivities, and heat capacities) in terms of fluctuating quantities; *iii*) deriving a fluctuation formula that corresponds to the superposition approximation; *iv*) examining more systems using available experimental data; *v*) deriving expressions for the limiting behavior of the *G*'s in binary or ternary systems; *vi*) comparing the results for systems with and without explicit polarization to results obtained with and without the superposition approximation; and *vii*) obtaining $g^{(3)}$ from our simulations for comparison with the integrated values.

We additionally presented and tested a way to calculate infinitely dilute partial molar properties of solutes (here proteins) using the apparent molar approach (Chapter 7). The strengths of the apparent molar approach as compared to literature approaches are that it is thermodynamically exact, no ambiguous parameters are introduced, knowledge of the specific conformations that contributed to the output values is retained, and the method is general. It may be used to calculate any partial molar property of any solute in any solution. The current limitation of the approach is the high level of noise in the isothermal compressibility and thermal expansion coefficient results, and a moderate level of noise in the isobaric heat capacity results. Continued increases in computational power should make the method presented here more usable over time. However, the apparent molar approach does not provide a means of directly decomposing the thermodynamic properties into specific contributions and requires multiple simulations.

We are currently working on an FST-based alternative to the apparent molar approach, discussed preliminarily at the conclusion of Chapter 7. The FST-based alternative promises to offer advantages over the apparent molar approach. This second approach is also thermodynamically exact. Furthermore, it will allow for distance dependent infinitely dilute partial molar properties to be determined from a single simulated system instead of requiring multiple compositions.

9. References

Abramov, R. V. 2007. An improved algorithm for the multidimensional moment-constrained maximum entropy problem. *Journal of Computational Physics.* 226, 621-644.

Aburi, M. and P. E. Smith. 2004. A combined simulation and Kirkwood-Buff approach to quantify cosolvent effects on the conformational preferences of peptides in solution. *Journal of Physical Chemistry B.* 108, 7382-7388.

Adams, D. J. 1975. Grand canonical ensemble Monte-Carlo for a Lennard-Jones fluid. *Molecular Physics.* 29, 307-311.

Aicart, E., E. Junquera, and T. M. Letcher. 1995. Isobaric thermal expansivity and isothermal compressibility of several nonsaturated hydrocarbons at 298.15 K. *Journal of Chemical and Engineering Data.* 40, 1225-1227.

Aicart, E., G. Tardajos, and M. D. Pena. 1981. Isothermal compressibility of cyclohexane-normal-decane, cyclohexane-normal-dodecane, and cyclohexane-normal-tetradecane. *Journal of Chemical and Engineering Data.* 26, 22-26.

Akamatsu, Y., H. Ogawa, and S. Murakami. 1987. Molar excess-enthalpies, molar excess volumes and molar isentropic compressions of mixtures of 2-propanone with heptane, benzene and trichloromethane at 298.15-K. *Thermochimica Acta.* 113, 141-150.

Allen, M. P. and D. J. Tildesley. 1987. *Computer Simulation of Liquids*. New York: Oxford University Press.

Allison, J. R., M. Bergeler, N. Hansen, and W. F. van Gunsteren. 2011. Current computer modeling cannot explain why two highly similar sequences fold into different structures. *Biochemistry.* 50, 10965-10973.

Alvarez, E., A. Cancela, R. Maceiras, J. M. Navaza, and R. Taboas. 2003. Surface tension of aqueous binary mixtures of 1-amino-2-propanol and 3-amino-1-propanol, and aqueous ternary mixtures of these amines with diethanolamine, triethanolamine, and 2-amino-2-methyl-1-propanol from (298.15 to 323.15) K. *Journal of Chemical and Engineering Data.* 48, 32-35.

Anisimov, V. M., G. Lamoureux, I. V. Vorobyov, N. Huang, B. Roux, and A. D. Mackerell. 2005. Determination of electrostatic parameters for a polarizable force field based on the classical drude oscillator. *Journal of Chemical Theory and Computation.* 1, 153-168.

Anisimov, V. M., I. V. Vorobyov, B. Roux, and A. D. Mackerell. 2007. Polarizable empirical force field for the primary and secondary alcohol series based on the classical Drude model. *Journal of Chemical Theory and Computation.* 3, 1927-1946.

Apelblat, A., A. Tamir, and M. Wagner. 1980. Thermodynamics of acetone-chloroform mixtures. *Fluid Phase Equilibria.* 4, 229-255.

Auslander, D., I. Lenart, and A. Ciupe. 1995. Contributions to the ultrasonic study of the internal-pressure of liquids. *Acustica.* 81, 75-78.

Babin, V., J. Baucom, T. A. Darden, and C. Sagui. 2006. Molecular dynamics simulations of DNA with polarizable force fields: Convergence of an ideal B-DNA structure to the crystallographic structure. *Journal of Physical Chemistry B.* 110, 11571-11581.

Baird, J. K. and Y. W. Kim. 2003. Fluctuation-dissipation theorem for chemical reactions near a critical point. *Journal of Physical Chemistry A.* 107, 10241-10242.

Bandyopadhyay, G., S. Dutta, and S. C. Lahiri. 2010. Determination of surface tension, structural and related properties of aquo-alcoholic mixtures at 298 K. *Zeitschrift für Physikalische Chemie.* 224, 729-742.

Barua, B., J. C. Lin, V. D. Williams, P. Kummler, J. W. Neidigh, and N. H. Andersen. 2008. The trp-cage: Optimizing the stability of a globular miniprotein. *Protein Engineering Design & Selection.* 21, 171-185.

Baucom, J., T. Transue, M. Fuentes-Cabrera, J. M. Krahn, T. A. Darden, and C. Sagui. 2004. Molecular dynamics simulations of the d(ccaacgttgg)(2) decamer in crystal environment: Comparison of atomic point-charge, extra-point, and polarizable force fields. *Journal of Chemical Physics.* 121, 6998-7008.

Beggerow, G. 1976. *Landolt-Boernstein. Group IV. Macroscopic and technical properties of matter. Vol 2. Heats of mixing and solution*. Edited by K.-H. Hellwege. Berlin: Springer-Verlag.

Ben-Naim, A. 1977. Inversion of the Kirkwood-Buff theory of solutions: Application to the water-ethanol system. *Journal of Chemical Physics.* 67, 4884-4890.

Ben-Naim, A. 1987. *Solvation Thermodynamics*. New York: Plenum Press.

Ben-Naim, A. 1988. Theory of preferential solvation of nonelectrolytes. *Cell Biophysics.* 12, 255-269.

Ben-Naim, A. 1989. Preferential solvation in 2-component systems. *Journal of Physical Chemistry.* 93, 3809-3813.

Ben-Naim, A. 1990. Preferential solvation in 2-component and in 3-component systems. *Pure and Applied Chemistry.* 62, 25-34.

Ben-Naim, A. 2006. *Molecular Theory of Solutions*. New York: Oxford University Press.

Ben-Naim, A., A. M. Navarro, and J. M. Leal. 2008. A Kirkwood-Buff analysis of local properties of solutions. *Physical Chemistry Chemical Physics.* 10, 2451-2460.

Ben-Naim, A. Y. 1992. *Statistical Thermodynamics for Chemists and Biochemists*. New York: Plenum Press.

Benjamin, L. and G. C. Benson. 1963. A deuterium isotope effect on the excess enthalpy of methanol-water solutions. *Journal of Physical Chemistry.* 67, 858-861.

Bentenitis, N., N. R. Cox, and P. E. Smith. 2009. A Kirkwood-Buff derived force field for thiols, sulfides, and disulfides. *Journal of Physical Chemistry B.* 113, 12306-12315.

Berendsen, H. J. C., J. R. Grigera, and T. P. Straatsma. 1987. The missing term in effective pair potentials. *Journal of Physical Chemistry.* 91, 6269-6271.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak. 1984. Molecular-dynamics with coupling to an external bath. *Journal of Chemical Physics.* 81, 3684-3690.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. Interaction models for water in relation to protein hydration. *Intermolecular Forces.* 331-342.

Berens, P. H., D. H. J. Mackay, G. M. White, and K. R. Wilson. 1983. Thermodynamics and quantum corrections from molecular-dynamics for liquid water. *Journal of Chemical Physics.* 79, 2375-2389.

Bjelkmar, P., P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl. 2010. Implementation of the CHARMM force field in Gromacs: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *Journal of Chemical Theory and Computation.* 6, 459-466.

Blanco, A., A. Garcia-Abuin, D. Gomez-Diaz, J. M. Navaza, and O. L. Villaverde. 2013. Density, speed of sound, viscosity, surface tension, and excess volume of n-ethyl-2-pyrrolidone + ethanolamine (or diethanolamine or triethanolamine) from T = (293.15 to 323.15) K. *Journal of Chemical and Engineering Data.* 58, 653-659.

Blanco, M. A., E. Sahin, Y. Li, and C. J. Roberts. 2011. Reexamining protein-protein and protein-solvent interactions from Kirkwood-Buff analysis of light scattering in multi-component solutions. *Journal of Chemical Physics.* 134, 225103.

Bonnet, P. and R. A. Bryce. 2004. Molecular dynamics and free energy analysis of neuraminidase-ligand interactions. *Protein Science.* 13, 946-957.

Borodin, O., R. L. Bell, Y. Li, D. Bedrov, and G. D. Smith. 2001. Polarizable and nonpolarizable potentials for K+ cation in water. *Chemical Physics Letters.* 336, 292-302.

Bowman, G. R., V. A. Voelz, and V. S. Pande. 2011. Taming the complexity of protein folding. *Current Opinion in Structural Biology.* 21, 4-11.

Bridgman, P. W. 1913. Thermodynamic properties of twelve liquids between 20 degrees and 80 degrees and up to 12,000 kgm. per sq. cm. *Proceedings of the American Academy of Arts and Sciences.* 49, 3-114.

Brovchenko, I., M. N. Andrews, and A. Oleinikova. 2010. Volumetric properties of human islet amyloid polypeptide in liquid water. *Physical Chemistry Chemical Physics.* 12, 4233-4238.

Buff, F. P. and R. Brout. 1955. Molecular formulation of thermodynamic functions encountered in solution theory. *Journal of Chemical Physics.* 23, 458-465.

Buff, F. P. and F. M. Schindler. 1958. Small perturbations in solution theory. *Journal of Chemical Physics.* 29, 1075-1081.

Butler, J. A. V., D. W. Thomson, and W. H. Maclennan. 1933. The free energy of the normal aliphatic alcohols in aqueous solution. Part I. The partial vapour pressures of aqueous solutions of methy, n-propyl, and n-butyl alcohols. Part II. The solubilities of some normal aliphatic alcohols in water. Part III. The theory of binary solutions, and its application to aqueous-alcoholic solutions. *Journal of the Chemical Society.* 674-686.

Caleman, C., P. J. van Maaren, M. Hong, J. S. Hub, L. T. Costa, and D. van der Spoel. 2012. Force field benchmark of organic liquids: Density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. *Journal of Chemical Theory and Computation.* 8, 61-74.

Callen, H. B. 1960. *Thermodynamics*. New York: John Wiley & Sons.

Canchi, D. R., D. Paschek, and A. E. Garcia. 2010. Equilibrium study of protein denaturation by urea. *Journal of the American Chemical Society.* 132, 2338-2344.

Chang, T.-M. and L. X. Dang. 2005. Liquid-vapor interface of methanol-water mixtures: A molecular dynamics study. *Journal of Physical Chemistry B.* 109, 5759-5765.

Chapman, W. G., K. E. Gubbins, G. Jackson, and M. Radosz. 1989. SAFT - Equation-of-state solution model for associating fluids. *Fluid Phase Equilibria.* 52, 31-38.

Chapman, W. G., K. E. Gubbins, G. Jackson, and M. Radosz. 1990. New reference equation of state for associating liquids. *Industrial & Engineering Chemistry Research.* 29, 1709-1721.

Cheatham, T. E. and B. R. Brooks. 1998. Recent advances in molecular dynamics simulation towards the realistic representation of biomolecules in solution. *Theoretical Chemistry Accounts.* 99, 279-288.

Chen, F. and P. E. Smith. 2008. Theory and computer simulation of solute effects on the surface tension of liquids. *Journal of Physical Chemistry B.* 112, 8975-8984.

Cheong, W. J. and P. W. Carr. 1987. The surface-tension of mixtures of methanol, acetonitrile, tetrahydrofuran, isopropanol, tertiary butanol and dimethylsulfoxide with water at 25-degrees-C. *Journal of Liquid Chromatography.* 10, 561-581.

Chialvo, A. A. 1993. Solute solute and solute solvent correlations in dilute near-critical ternary mixtures - mixed-solute and entrainer effects. *Journal of Physical Chemistry.* 97, 2740-2744.

Chialvo, A. A., Y. V. Kalyuzhnyi, and P. T. Cummings. 1996. Solvation thermodynamics of gas solubility at sub- and near-critical conditions. *AIChE Journal.* 42, 571-584.

Chiou, S. H., P. Azari, M. E. Himmel, and P. G. Squire. 1979. Isolation and physical characterization of bovine lens crystallins. *International Journal of Peptide and Protein Research.* 13, 409-417.

Chitra, R. and P. E. Smith. 2000. Molecular dynamics simulations of the properties of cosolvent solutions. *Journal of Physical Chemistry B.* 104, 5854-5864.

Chitra, R. and P. E. Smith. 2001a. A comparison of the properties of 2,2,2-trifluoroethanol and 2,2,2-trifluoroethanol/water mixtures using different force fields. *Journal of Chemical Physics.* 115, 5521-5530.

Chitra, R. and P. E. Smith. 2001b. Preferential interactions of cosolvents with hydrophobic solutes. *Journal of Physical Chemistry B.* 105, 11513-11522.

Chitra, R. and P. E. Smith. 2001c. Properties of 2,2,2-trifluoroethallol and water mixtures. *Journal of Chemical Physics.* 114, 426-435.

Chitra, R. and P. E. Smith. 2002. Molecular association in solution: A Kirkwood-Buff analysis of sodium chloride, ammonium sulfate, guanidinium chloride, urea, and 2,2,2-trifluoroethanol in water. *Journal of Physical Chemistry B.* 106, 1491-1500.

Cibulka, I. and T. Takagi. 1999. P-rho-T data of liquids: Summarization and evaluation. 5. Aromatic hydrocarbons. *Journal of Chemical and Engineering Data.* 44, 411-429.

Cochran, H. D., L. L. Lee, and D. M. Pfund. 1987. Application of the Kirkwood-Buff theory of solutions to dilute supercritical mixtures. *Fluid Phase Equilibria.* 34, 219-234.

Connors, K. A. and J. L. Wright. 1989. Dependence of surface tension on composition of binary aqueous-organic solutions. *Analytical Chemistry.* 61, 194-198.

Coomber, B. A. and C. J. Wormald. 1976. Stirred flow calorimeter - excess enthalpies of acetone water and of acetone and some normal alcohols. *Journal of Chemical Thermodynamics.* 8, 793-799.

Cooper, A. 1976. Thermodynamic fluctuations in protein molecules. *Proceedings of the National Academy of Sciences of the United States of America.* 73, 2740-2741.

Cooper, A. 2010. Protein heat capacity: An anomaly that maybe never was. *Journal of Physical Chemistry Letters.* 1, 3298-3304.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society.* 117, 5179-5197.

Dadarlat, V. M. and C. B. Post. 2001. Insights into protein compressibility from molecular dynamics simulations. *Journal of Physical Chemistry B.* 105, 715-724.

Dadarlat, V. M. and C. B. Post. 2006. Decomposition of protein experimental compressibility into intrinsic and hydration shell contributions. *Biophysical Journal.* 91, 4544-4554.

Dai, S., S. Weerasinghe, and P. E. Smith. 2014. A Kirkwood-Buff derived force field for amines and carboxylic acids. *In preparation.*

Dang, L. X., J. E. Rice, J. Caldwell, and P. A. Kollman. 1991. Ion solvation in polarizable water: Molecular dynamics simulations. *Journal of the American Chemical Society.* 113, 2481-2486.

Darden, T., D. York, and L. Pedersen. 1993. Particle Mesh Ewald: An n.Log(n) method for ewald sums in large systems. *Journal of Chemical Physics.* 98, 10089-10092.

Daura, X., K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark. 1999. Peptide folding: When simulation meets experiment. *Angewandte Chemie, International Edition in English.* 38, 236-240.

Davidson, N. 1962. *Statistical Mechanics*. New York: McGraw-Hill.

Davis-Searles, P. R., A. J. Saunders, D. A. Erie, D. J. Winzor, and G. J. Pielak. 2001. Interpreting the effects of small uncharged solutes on protein-folding equilibria. *Annual Review of Biophysics and Biomolecular Structure.* 30, 271-306.

Day, R., D. Paschek, and A. E. Garcia. 2010. Microsecond simulations of the folding/unfolding thermodynamics of the trp-cage miniprotein. *Proteins: Structure Function and Bioinformatics.* 78, 1889-1899.

Debenedetti, P. G. 1986. Derivation of operational definitions for the computer calculation of partial molar properties in multicomponent mixtures. *Chemical Physics Letters.* 132, 325-330.

Debenedetti, P. G. 1987a. Clustering in dilute, binary supercritical mixtures - a fluctuation analysis. *Chemical Engineering Science.* 42, 2203-2212.

Debenedetti, P. G. 1987b. Fluctuation-based computer calculation of partial molar properties. 1. Molecular-dynamics simulation of constant volume fluctuations. *Journal of Chemical Physics.* 86, 7126-7137.

Debenedetti, P. G. 1988. Fluctuation-based computer calculation of partial molar properties. 2. A numerically accurate method for the determination of partial molar energies and enthalpies. *Journal of Chemical Physics*. 88, 2681-2684.

Debenedetti, P. G. 1989. Fluctuation simulations and the calculation of mechanical partial molar properties. *Molecular Simulation*. 2, 33-53.

Devane, R., C. Ridley, R. W. Larsen, B. Space, P. B. Moore, and S. I. Chan. 2003. A molecular dynamics method for calculating molecular volume changes appropriate for biomolecular simulation. *Biophysical Journal*. 85, 2801-2807.

Diazpena, M., G. Tardajos, R. L. Arenosa, and C. Menduina. 1979. Isothermal compressibility of benzene + normal-undecane, + normal-dodecane, + normal-tetradecane, and + normal-hexadecane. *Journal of Chemical Thermodynamics*. 11, 951-957.

Dill, K. A., K. Ghosh, and J. D. Schmit. 2011. Physical limits of cells and proteomes. *Proceedings of the National Academy of Sciences of the United States of America*. 108, 17876-17882.

Douheret, G., A. Khadir, and A. Pal. 1989. Thermodynamic characterization of the water + methanol system, at 298.15K. *Thermochimica Acta*. 142, 219-243.

Egelstaff, P. A. 1973. The structure of simple liquids. *Annual Review of Physical Chemistry*. 24, 159-187.

Feng, J. W. A., J. Kao, and G. R. Marshall. 2009. A second look at mini-protein stability: Analysis of FSD-1 using circular dichroism, differential scanning calorimetry, and simulations. *Biophysical Journal*. 97, 2803-2810.

Ferrario, M., M. Haughney, I. R. Mcdonald, and M. L. Klein. 1990. Molecular-dynamics simulation of aqueous mixtures - methanol, acetone, and ammonia. *Journal of Chemical Physics*. 93, 5156-5166.

Fowler, R. H. and E. A. Guggenheim. 1939. *Statistical Thermodynamics: A Version of Statistical Mechanics for Students of Physics and Chemistry*. Cambridge, England: The University Press.

Freddolino, P. L., C. B. Harrison, Y. X. Liu, and K. Schulten. 2010. Challenges in protein-folding simulations. *Nature Physics*. 6, 751-758.

Frenkel, D. and B. Smit. 2002. *Understanding Molecular Simulation*. San Diego: Elsevier.

Friedman, H. L. and P. S. Ramanathan. 1970. Theory of mixed electrolyte solutions and application to a model for aqueous lithium chloride-cesium chloride. *Journal of Physical Chemistry*. 74, 3756-3765.

Fujii, S., K. Tamura, and S. Murakami. 1995. Thermodynamic properties of (an alkylbenzene plus cyclohexane) at the temperature 298.15 K. *Journal of Chemical Thermodynamics*. 27, 1319-1328.

Ganguly, P., P. Schravendijk, B. Hess, and N. F. A. van der Vegt. 2011. Ion pairing in aqueous electrolyte solutions with biologically relevant anions. *Journal of Physical Chemistry B.* 115, 3734-3739.

Garcia, B., F. J. Hoyuelos, R. Alcalde, and J. M. Leal. 1996. Molar excess volumes of binary liquid mixtures: 2-pyrrolidinone with C-6-C-10 n-alkanols. *Canadian Journal of Chemistry-Revue Canadienne De Chimie.* 74, 121-127.

Garcia, M., C. Rey, V. P. Villar, and J. R. Rodriguez. 1988. Excess volumes of (normal-nonane + normal-undecane) between 288.15-K and 308.15-K. *Journal of Chemical and Engineering Data.* 33, 46-48.

Gee, M. B. 2010. Computer simulation and theory of amino acid interactions in solution, Chemistry, Kansas State University, Manhattan, Kansas.

Gee, M. B., N. R. Cox, Y. Jiao, N. Bentenitis, S. Weerasinghe, and P. E. Smith. 2011. A Kirkwood-Buff derived force field for aqueous alkali halides. *Journal of Chemical Theory and Computation.* 7, 1369-1380.

Gee, M. B. and P. E. Smith. 2009. Kirkwood-Buff theory of molecular and protein association, aggregation, and cellular crowding. *Journal of Chemical Physics.* 131, 165101.

Gee, M. B. and P. E. Smith. 2012. A Kirkwood-Buff derived force field for amine salts and carboxylates. *In preparation.*

Geerke, D. P. and W. F. van Gunsteren. 2007a. Calculation of the free energy of polarization: Quantifying the effect of explicitly treating electronic polarization on the transferability of force-field parameters. *Journal of Physical Chemistry B.* 111, 6425-6436.

Geerke, D. P. and W. F. van Gunsteren. 2007b. The performance of non-polarizable and polarizable force-field parameter sets for ethylene glycol in molecular dynamics simulations of the pure liquid and its aqueous mixtures. *Molecular Physics.* 105, 1861-1881.

Gekko, K. and Y. Hasegawa. 1986. Compressibility structure relationship of globular proteins. *Biochemistry.* 25, 6563-6571.

Gekko, K. and H. Noguchi. 1979. Compressibility of globular proteins in water at 25-degrees-C. *Journal of Physical Chemistry.* 83, 2706-2714.

Gibbs, J. W. 1902. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*. New York: Charles Scribner's Sons.

Gibbs, J. W. 1948. Elementary principles in statistical mechanics. In *The Collected Works of J. Willard Gibbs*. New Haven: Yale Univ. Press.

Glättli, A., X. Daura, and W. F. van Gunsteren. 2003. A novel approach for designing simple point charge models for liquid water with three interaction sites. *Journal of Computational Chemistry.* 24, 1087-1096.

Gmehling, J., U. Onken, W. Arlt, P. Grenzheuser, U. Weidlich, B. Kolbe, and J. Rarey. 1991. Aqueous-organic systems. In *The DECHEMA chemistry data series*. Frankfurt: DECHEMA.

Gorbunova, N. I., V. A. Grigoriev, V. M. Simonov, and V. A. Shipova. 1982. Heat capacity of liquid benzene and hexafluorobenzene at atmospheric pressure. *International Journal of Thermophysics.* 3, 1-15.

Greene, R. F. and H. B. Callen. 1951. On the formalism of thermodynamic fluctuation theory. *Physical Review.* 83, 1231-1235.

Grolier, J. P. E., G. Rouxdesgranges, M. Berkane, and E. Wilhelm. 1994. Heat-capacities and densities of mixtures of very polar substances. 3. Mixtures containing either trichloromethane or 1,4-dioxane or diisopropylether. *Journal of Solution Chemistry.* 23, 153-166.

Grossfield, A., P. Y. Ren, and J. W. Ponder. 2003. Ion solvation thermodynamics from simulation with a polarizable force field. *Journal of the American Chemical Society.* 125, 15671-15682.

Gubbins, K. E. 2013. The theory of non-electrolyte solutions: An historical review. *Molecular Physics.* 111, 3666-3697.

Gubbins, K. E., C. G. Gray, and P. A. Egelstaff. 1978. Thermodynamic derivatives of correlation-functions. *Molecular Physics.* 35, 315-328.

Hadden, S. T. 1970. Heat capacity of hydrocarbons in the normal liquid range. *Journal of Chemical and Engineering Data.* 15, 92-98.

Halgren, T. A. and W. Damm. 2001. Polarizable force fields. *Current Opinion in Structural Biology.* 11, 236-242.

Hall, D. G. 1971. Kirkwood-Buff theory of solutions - Alternative derivation of part of it and some applications. *Transactions of the Faraday Society.* 67, 2516-2524.

Han, K. J., J. H. Oh, and S. J. Park. 2006. Densities and viscosities for the ternary systems of methyl tert-butyl ether plus methanol plus benzene and methyl tert-butyl ether plus methanol plus toluene and their sub-binary systems at 298.15 K. *Journal of Chemical and Engineering Data.* 51, 1339-1344.

Harder, E., V. M. Anisimov, T. Whitfield, A. D. Mackerell, and B. Roux. 2008. Understanding the dielectric properties of liquid amides from a polarizable force field. *Journal of Physical Chemistry B.* 112, 3509-3521.

Henry, W. 1803. Experiments on the quantity of gases absorbed by water, at different temperatures, and under different pressures. *Philosophical Transactions of the Royal Society of London.* 93, 29-274.

Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry.* 18, 1463-1472.

Hess, B., C. Kutzner, D. van der Spoel, and E. Lindahl. 2008. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation.* 4, 435-447.

Heyrovska, R. 1996. Physical electrochemistry of strong electrolytes based on partial dissociation and hydration - quantitative interpretation of the thermodynamic properties of NaCl(aq) from "zero to saturation". *Journal of the Electrochemical Society.* 143, 1789-1793.

Hildebrand, J. H. and J. N. Sharma. 1929. The activities of molten alloys of thallium with tin and with lead. *Journal of the American Chemical Society.* 51, 462-471.

Hill, T. L. 1956. *Statistical Mechanics: Principles and Selected Applications*, *The McGraw-Hill Series in Advanced Chemistry*. New York: McGraw-Hill Book Company, Inc.

Hockney, R. W. 1970. The potential calculation and some applications. *Methods in Computational Physics.* 9, 136-211.

Hockney, R. W., S. P. Goel, and J. W. Eastwood. 1974. Quiet high resolution computer models of a plasma. *Journal of Computational Physics.* 14, 148-158.

Høiland, H. and G. R. Hedwig. 6 Compressibilities of amino acids, peptides and proteins in aqueous solution. *SpringerMaterials-The Landolt-Börnstein Database.* 1-17.

Horinek, D. and R. R. Netz. 2011. Can simulations quantitatively predict peptide transfer free energies to urea solutions? Thermodynamic concepts and force field limitations. *Journal of Physical Chemistry A.* 115, 6125-6136.

Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure Function and Bioinformatics.* 65, 712-725.

Horta, B. A. C., P. F. J. Fuchs, W. F. van Gunsteren, and P. H. Hünenberger. 2011. New interaction parameters for oxygen compounds in the GROMOS force field: Improved pure-liquid and solvation properties for alcohols, ethers, aldehydes, ketones, carboxylic acids, and esters. *Journal of Chemical Theory and Computation.* 7, 1016-1031.

Hruby, J., J. Klomfar, and O. Sifner. 1993. (T, p, ρ) relation of liquid methanol at temperatures from 205-K to 321-K and pressures up to 50-MPa. *Journal of Chemical Thermodynamics.* 25, 1229-1242.

Hu, B. J., A. W. Nienow, E. H. Stitt, and A. W. Pacek. 2007. Bubble sizes in agitated water-hydrophilic organic solvents for heterogeneous catalytic reactions. *Industrial and Engineering Chemistry Research.* 46, 4451-4458.

Hu, J. H., C. A. Haynes, A. H. Y. Wu, C. M. W. Cheung, M. M. Chen, E. G. M. Yee, T. Ichioka, K. Nishikawa, P. Westh, and Y. Koga. 2003. Chemical potential and concentration fluctuation in some aqueous alkane-mono-ols at 25 degrees C. *Canadian Journal of Chemistry.* 81, 141-149.

Hünenberger, P. H. and W. F. van Gunsteren. 1997. Empirical classical interaction functions for molecular simulation. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, edited by W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Huot, J. Y., E. Battistel, R. Lumry, G. Villeneuve, J. F. Lavallee, A. Anusiem, and C. Jolicoeur. 1988. A comprehensive thermodynamic investigation of water-ethylene glycol mixtures at 5, 25, and 45-degrees-C. *Journal of Solution Chemistry.* 17, 601-636.

Imai, T., M. Kinoshita, and F. Hirata. 2000. Theoretical study for partial molar volume of amino acids in aqueous solution: Implication of ideal fluctuation volume. *Journal of Chemical Physics.* 112, 9469-9478.

Imai, T., S. Ohyama, A. Kovalenko, and F. Hirata. 2007. Theoretical study of the partial molar volume change associated with the pressure-induced structural transition of ubiquitin. *Protein Science.* 16, 1927-1933.

Ising, E. 1925. The theory of ferromagnetism. *Zeitschrift für Physik.* 31, 253-258.

Iuchi, S., S. Izvekov, and G. A. Voth. 2007. Are many-body electronic polarization effects important in liquid water? *Journal of Chemical Physics.* 126, 124505.

Jiao, Y., F. Chen, E. A. Ploetz, S. Weerasinghe, and P. E. Smith. 2014. A Kirkwood-Buff derived force field for alcohols in water. *In preparation.*

Jiao, Y. F. and P. E. Smith. 2011. Fluctuation theory of molecular association and conformational equilibria. *Journal of Chemical Physics.* 135, 014502.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics.* 79, 926-935.

Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society.* 118, 11225-11236.

Jorgensen, W. L., N. A. McDonald, M. Selmi, and P. R. Rablen. 1995. Importance of polarization for dipolar solutes in low-dielectric media - 1,2-dichloroethane and water in cyclohexane. *Journal of the American Chemical Society.* 117, 11809-11810.

Jorgensen, W. L. and J. Tirado-Rives. 1988. The OPLS potential functions for proteins - Energy minimizations for crystals of cyclic-peptides and crambin. *Journal of the American Chemical Society.* 110, 1657-1666.

Jorgensen, W. L. and J. Tirado-Rives. 2005. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences of the United States of America.* 102, 6665-6670.

Joung, I. S. and T. E. Cheatham. 2008. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B.* 112, 9020-9041.

Joung, I. S. and T. E. Cheatham. 2009. Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *Journal of Physical Chemistry B.* 113, 13279-13290.

Jungwirth, P. and D. J. Tobias. 2006. Specific ion effects at the air/water interface. *Chemical Reviews.* 106, 1259-1281.

Juraszek, J. and P. G. Bolhuis. 2008. Rate constant and reaction coordinate of trp-cage folding in explicit water. *Biophysical Journal.* 95, 4246-4257.

Kaminski, G. A., R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B.* 105, 6474-6487.

Kaminski, G. A., H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. X. Cao, R. B. Murphy, R. H. Zhou, and T. A. Halgren. 2002. Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *Journal of Computational Chemistry.* 23, 1515-1531.

Kang, M. and P. E. Smith. 2006. A Kirkwood-Buff derived force field for amides. *Journal of Computational Chemistry.* 27, 1477-1485.

Kang, M. and P. E. Smith. 2007. Preferential interaction parameters in biological systems by Kirkwood-Buff theory and computer simulation. *Fluid Phase Equilibria.* 256, 14-19.

Kang, M. and P. E. Smith. 2008. Kirkwood-Buff theory of four and higher component mixtures. *Journal of Chemical Physics.* 128, 244511.

Karunaweera, S., M. B. Gee, S. Weerasinghe, and P. E. Smith. 2012. Theory and simulation of multicomponent osmotic systems. *Journal of Chemical Theory and Computation.* 8, 3493–3503.

Kirkwood, J. G. 1935. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*. 3, 300-313.

Kirkwood, J. G. and F. P. Buff. 1951. The statistical mechanical theory of solutions. I. *Journal of Chemical Physics*. 19, 774-777.

Kirkwood, J. G. and R. J. Goldberg. 1950. Light scattering arising from composition fluctuations in multi-component systems. *Journal of Chemical Physics*. 18, 54-57.

Kiyohara, O. and G. C. Benson. 1979. Ultrasonic speeds and isentropic compressibilities of normal-alkanol + normal-heptane mixtures at 298.15-K. *Journal of Chemical Thermodynamics*. 11, 861-873.

Klasczyk, B. and V. Knecht. 2010. Kirkwood-Buff derived force field for alkali chlorides in simple point charge water. *Journal of Chemical Physics*. 132, 024109.

Kokubo, H., J. Roesgen, D. W. Bolen, and B. M. Pettitt. 2007. Molecular basis of the apparent near ideality of urea solutions. *Biophysical Journal*. 93, 3392-3407.

Kubota, H., Y. Tanaka, and T. Makita. 1987. Volumetric behavior of pure alcohols and their water mixtures under high-pressure. *International Journal of Thermophysics*. 8, 47-70.

Kusalik, P. G. and G. N. Patey. 1987. The thermodynamic properties of electrolyte-solutions - some formal results. *Journal of Chemical Physics*. 86, 5110-5116.

Lange, O. F., D. van der Spoel, and B. L. de Groot. 2010. Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophysical Journal*. 99, 647-655.

Lazaridis, T. 2002. Binding affinity and specificity from computational studies. *Current Organic Chemistry*. 6, 1319-1332.

Lazaridis, T., A. Masunov, and F. Gandolfo. 2002. Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins: Structure Function and Bioinformatics*. 47, 194-208.

Lee, M. E. and N. F. A. van der Vegt. 2005. A new force field for atomistic simulations of aqueous tertiary butanol solutions. *Journal of Chemical Physics*. 122, 114509.

Leontyev, I. V. and A. A. Stuchebrukhov. 2012. Polarizable mean-field model of water for biological simulations with AMBER and CHARMM force fields. *Journal of Chemical Theory and Computation*. 8, 3207-3216.

Lewis, G. N. 1907. Outlines of a new system of thermodynamic chemistry. *Proceedings of the American Academy of Arts and Sciences*. 43, 259-293.

Lide, D. R. 1990. *CRC Handbook of Chemistry & Physics*. Boca Raton: CRC.

Liltorp, K., P. Westh, and Y. Koga. 2005. Thermodynamic properties of water in the water-poor region of binary water plus alcohol mixtures. *Canadian Journal of Chemistry-Revue Canadienne De Chimie.* 83, 420-429.

Lin, L. N., J. F. Brandts, J. M. Brandts, and V. Plotnikov. 2002. Determination of the volumetric properties of proteins and other solutes using pressure perturbation calorimetry. *Analytical Biochemistry.* 302, 144-160.

London, F. 1930. Theory and systematics of molecular forces. *Zeitschrift für Physikalische Chemie.* 63, 245-279.

Longuet-Higgins, H. C. 1951. The statistical thermodynamics of multicomponent systems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences.* 205, 247-269.

Mackerell, A. D. 2004. Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry.* 25, 1584-1604.

Mackerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-Mccarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B.* 102, 3586-3616.

Mackerell, A. D., M. Feig, and C. L. Brooks. 2004. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry.* 25, 1400-1415.

Maerzke, K. A., N. E. Schultz, R. B. Ross, and J. I. Siepmann. 2009. TraPPE-UA force field for acrylates and Monte Carlo simulations for their mixtures with alkanes and alcohols. *Journal of Physical Chemistry B.* 113, 6415-6425.

Majer, V., J. Sedlbauer, and R. H. Wood. 2004. Calculation of standard thermodynamic properties of aqueous electrolytes and non-electrolytes. In *Aqueous Systems at Elevated Temperatures and Pressures: Physical Chemistry in Water, Steam, and Hydrothermal Solutions*, edited by D. A. Palmer, R. Fernandez-Prini, and A. H. Harvey. Amsterdam: Elsevier.

Makhatadze, G. I. 2007. Thermodynamic properties of proteins. In *Physical Properties of Polymers Handbook*, edited by J. E. Mark. New York: Springer.

Makhatadze, G. I., K. S. Kim, C. Woodward, and P. L. Privalov. 1993. Thermodynamics of BPTI folding. *Protein Science.* 2, 2028-2036.

Malde, A. K., L. Zuo, M. Breeze, M. Stroet, D. Poger, P. C. Nair, C. Oostenbrink, and A. E. Mark. 2011. An automated force field topology builder (ATB) and repository: Version 1.0. *Journal of Chemical Theory and Computation.* 7, 4026-4037.

Maple, J. R., Y. X. Cao, W. G. Damm, T. A. Halgren, G. A. Kaminski, L. Y. Zhang, and R. A. Friesner. 2005. A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions. *Journal of Chemical Theory and Computation.* 1, 694-715.

Marcus, Y. 1989. Preferential solvation of ions in mixed-solvents. 4. Comparison of the Kirkwood Buff and Quasi-Lattice Quasi-Chemical approaches. *Journal of the Chemical Society-Faraday Transactions I.* 85, 3019-3032.

Marcus, Y. 2001. Preferential solvation in mixed solvents X. Completely miscible aqueous co-solvent binary mixtures at 298.15 K. *Monatshefte Fur Chemie.* 132, 1387-1411.

Marcus, Y. 2002. *Solvent Mixtures: Properties and Selective Solvation*. New York: Marcel Dekker.

Matsumoto, M., Y. Takaoka, and Y. Kataoka. 1993. Liquid-vapor interface of water-methanol mixture. I. Computer simulation. *Journal of Chemical Physics.* 98, 1464-1472.

Matteoli, E. and L. Lepori. 1984. Solute-solute interactios in water. II. An analysis through the Kirkwood-Buff integrals for 14 organic solutes. *Journal of Chemical Physics.* 80, 2856-2863.

Matteoli, E. and L. Lepori. 1995. Kirkwood-Buff integrals and preferential solvation in ternary nonelectrolyte mixtures. *Journal of the Chemical Society-Faraday Transactions.* 91, 431-436.

Matteoli, E. and G. A. Mansoori. 1990. *Fluctuation Theory of Mixtures*. New York: Taylor & Francis.

Matulis, D. 2001. Thermodynamics of the hydrophobic effect. III. Condensation and aggregation of alkanes, alcohols, and alkylamines. *Biophysical Chemistry.* 93, 67-82.

Maximino, R. B. 2009. Surface tension and density of binary mixtures of monoalcohols, water and acetonitrile: Equation of correlation of the surface tension. *Physics and Chemistry of Liquids.* 47, 475-486.

Mazo, R. M. 1958. Statistical mechanical theory of solutions. *Journal of Chemical Physics.* 29, 1122-1128.

Mazo, R. M. 2006. A fluctuation theory analysis of the salting-out effect. *Journal of Physical Chemistry B.* 110, 24077-24082.

Mazo, R. M. 2008. Concentration fluctuations in fluid mixtures. *Journal of Chemical Physics.* 129, 154101.

Mazo, R. M., E. Matteoli, and P. E. Smith. 2009. Concentration fluctuations in fluid mixtures. II. *Journal of Chemical Physics.* 130, 234508.

McMillan, W. G. and J. E. Mayer. 1945. The statistical thermodynamics of multicomponent systems. *Journal of Chemical Physics.* 13, 276-305.

Meng, E. C., P. Cieplak, J. W. Caldwell, and P. A. Kollman. 1994. Accurate solvation free energies of acetate and methylammonium ions calculated with a polarizable water model. *Journal of the American Chemical Society.* 116, 12061-12062.

Miller, K. J. and J. A. Savchik. 1979. New empirical-method to calculate average molecular polarizabilities. *Journal of the American Chemical Society.* 101, 7206-7213.

Mishima, O. and H. E. Stanley. 1998. The relationship between liquid, supercooled and glassy water. *Nature.* 396, 329-335.

Mitra, L., A. Oleinikova, and R. Winter. 2008. Intrinsic volumetric properties of trialanine isomers in aqueous solution. *ChemPhysChem.* 9, 2779-2784.

Mitra, L., N. Smolin, R. Ravindra, C. Royer, and R. Winter. 2006. Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution: Experiments and theoretical interpretation. *Physical Chemistry Chemical Physics.* 8, 1249-1265.

Miyamoto, S. and P. A. Kollman. 1992. Settle: An analytical version of the Shake and Rattle algorithm for rigid water models. *Journal of Computational Chemistry.* 13, 952-962.

Mu, Y. G., D. S. Kosov, and G. Stock. 2003. Conformational dynamics of trialanine in water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments. *Journal of Physical Chemistry B.* 107, 5064-5073.

Mueller, C. R. and E. R. Kearns. 1958. Thermodynamic studies of the system acetone and chloroform. *Journal of Physical Chemistry.* 62, 1441-1445.

Münster, A. 1970. *Classical Thermodynamics*. Bristol: Stonebridge Press.

Murphy, L. R., N. Matubayasi, V. A. Payne, and R. M. Levy. 1998. Protein hydration and unfolding: Insights from experimental partial specific volumes and unfolded protein models. *Folding & Design.* 3, 105-118.

Murthy, N. M. and S. V. Subrahmanyam. 1977. Behavior of excess heat-capacity of aqueous non-electrolytes. *Indian Journal of Pure & Applied Physics.* 15, 485-489.

Navia, P., J. Troncoso, and L. Romani. 2010. Isobaric thermal expansivity of highly polar nitrogen compounds at temperatures from (278.15 to 348.15) H and at pressures from (5 to 55) MPa. *Journal of Chemical and Engineering Data.* 55, 1537-1541.

Newman, K. E. 1994. Kirkwood-Buff solution theory - Derivation and applications. *Chemical Society Reviews.* 23, 31-40.

Nichols, J. W., S. G. Moore, and D. R. Wheeler. 2009. Improved implementation of Kirkwood-Buff solution theory in periodic molecular simulations. *Physical Review E.* 80, 051205.

Noskov, S. Y., G. Lamoureux, and B. Roux. 2005. Molecular dynamics study of hydration in ethanol-water mixtures using a polarizable force field. *Journal of Physical Chemistry B.* 109, 6705-6713.

Nymeyer, H. 2009. Energy landscape of the Trpzip2 peptide. *Journal of Physical Chemistry B.* 113, 8288-8295.

O'Connell, J. P. 1971a. Molecular thermodynamics of gases in mixed solvents. *AIChE Journal.* 17, 658-663.

O'Connell, J. P. 1971b. Thermodynamic properties of solutions based on correlation functions. *Molecular Physics.* 20, 27-33.

O'Connell, J. P. and H. Q. Liu. 1998. Thermodynamic modelling of near-critical solutions. *Fluid Phase Equilibria.* 144, 1-12.

Okur, A., B. Strockbine, V. Hornak, and C. Simmerling. 2003. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. *Journal of Computational Chemistry.* 24, 21-31.

Oostenbrink, C., A. Villa, A. E. Mark, and W. F. van Gunsteren. 2004. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53a5 and 53a6. *Journal of Computational Chemistry.* 25, 1656-1676.

Orozco, M. and F. J. Luque. 2000. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chemical Reviews.* 100, 4187-4225.

Oswal, S. L., P. Oswal, R. L. Gardas, S. G. Patel, and R. G. Shinde. 2004. Acoustic, volumetric, compressibility and refractivity properties and reduction parameters for the eras and flory models of some homologous series of amines from 298.15 to 328.15 K. *Fluid Phase Equilibria.* 216, 33-45.

Papaioannou, D. and C. Panayiotou. 1995. Viscosity of binary-mixtures of propylamine with alkanols at moderately high-pressures. *Journal of Chemical and Engineering Data.* 40, 202-209.

Partay, L. B., P. Jedlovsky, A. Vincze, and G. Horvai. 2008. Properties of free surface of water-methanol mixtures. Analysis of the truly interfacial molecular layer in computer simulation. *Journal of Physical Chemistry B.* 112, 5428-5438.

Paschek, D., R. Day, and A. E. Garcia. 2011. Influence of water-protein hydrogen bonding on the stability of trp-cage miniprotein. A comparison between the TIP3P and TIP4P-EW water models. *Physical Chemistry Chemical Physics.* 13, 19840-19847.

Paschek, D., S. Hempel, and A. E. Garcia. 2008. Computing the stability diagram trp-cage miniprotein of the. *Proceedings of the National Academy of Sciences of the United States of America.* 105, 17754-17759.

Patel, S. and C. L. Brooks. 2004. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *Journal of Computational Chemistry.* 25, 1-15.

Patel, S. and C. L. Brooks. 2005. Structure, thermodynamics, and liquid-vapor equilibrium of ethanol from molecular-dynamics simulations using nonadditive interactions. *Journal of Chemical Physics.* 123, 164502.

Patel, S., A. D. Mackerell, and C. L. Brooks. 2004. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *Journal of Computational Chemistry.* 25, 1504-1514.

Patel, S. A. and C. L. Brooks, III. 2006. Revisiting the hexane-water interface via molecular dynamics simulations using nonadditive alkane-water potentials. *Journal of Chemical Physics.* 124, 204706.

Paul, S. and A. Chandra. 2005. Hydrogen bond properties and dynamics of liquid-vapor interfaces of aqueous methanol solutions. *Journal of Chemical Theory and Computation.* 1, 1221-1231.

Perera, A. and F. Sokolic. 2004. Modeling nonionic aqueous solutions: The acetone-water mixture. *Journal of Chemical Physics.* 121, 11272-11282.

Perron, G., A. Roux, and J. E. Desnoyers. 1981. Heat-capacities and volumes of NaCl, MgCl$_2$, CaCl$_2$, and NiCl$_2$ up to 6 molal in water. *Canadian Journal of Chemistry-Revue Canadienne De Chimie.* 59, 3049-3054.

Piana, S., K. Lindorff-Larsen, and D. E. Shaw. 2013. Atomic-level description of ubiquitin folding. *Proceedings of the National Academy of Sciences of the United States of America.* 110, 5915-5920.

Pierce, V., M. Kang, M. Aburi, S. Weerasinghe, and P. E. Smith. 2008. Recent applications of Kirkwood-Buff theory to biological systems. *Cell Biochemistry and Biophysics.* 50, 1-22.

Pjura, P. E., M. E. Paulaitis, and A. M. Lenhoff. 1995. Molecular thermodynamic properties of protein solutions from partial specific volumes. *AIChE Journal.* 41, 1005-1009.

Ploetz, E. A., N. Bentenitis, and P. E. Smith. 2010a. Developing force fields from the microscopic structure of solutions. *Fluid Phase Equilibria.* 290, 43-47.

Ploetz, E. A., N. Bentenitis, and P. E. Smith. 2010b. Kirkwood-Buff integrals for ideal solutions. *Journal of Chemical Physics.* 132, 164501.

Ploetz, E. A. and P. E. Smith. 2011a. A Kirkwood-Buff force field for the aromatic amino acids. *Physical Chemistry Chemical Physics.* 13, 18154-18167.

Ploetz, E. A. and P. E. Smith. 2011b. Local fluctuations in solution mixtures. *Journal of Chemical Physics.* 135, 044506.

Ploetz, E. A. and P. E. Smith. 2013. Local fluctuations in solution: Theory and applications. In *Advances in Chemical Physics*: John Wiley & Sons, Inc.

Prabhu, N. V. and K. A. Sharp. 2005. Heat capacity in proteins. *Annual Review of Physical Chemistry.* 56, 521-548.

Prigogine, I. 1957. *The Molecular Theory of Solutions*. New York: Interscience.

Rajendran, M., S. Renganarayanan, P. R. Madhavan, and D. Srinivasan. 1989. Effect of dissolved salts on the heat of mixing of 3 binary-systems. *Journal of Chemical and Engineering Data.* 34, 375-382.

Raoult, F.-M. 1878. Vapour-tension and solidifying point of saline solutions. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences.* 87, 167-169.

Raoult, F.-M. 1882. Law of freezing of solvents. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences.* 95, 1030-1033.

Raoult, F.-M. 1887. General law of the vapor pressure of solvents. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences.* 104, 1430-1433.

Raoult, F.-M. 1888. Vapour-tensions of ethereal solutions. *Zeitschrift für Physikalische Chemie, Stöchiometrie und Verwandtschaftslehre.* 2, 353-373.

Rappé, A. K. and W. A. Goddard. 1991. Charge equilibration for molecular-dynamics simulations. *Journal of Physical Chemistry.* 95, 3358-3363.

Record, M. T., W. T. Zhang, and C. F. Anderson. 1998. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: A practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. *Advances in Protein Chemistry.* 51, 281-353.

Reis, J. C. R., G. Douheret, M. I. Davis, I. J. Fjellanger, and H. Hoiland. 2008. Isentropic expansion and related thermodynamic properties of non-ionic amphiphile-water mixtures. *Physical Chemistry Chemical Physics.* 10, 561-573.

Ren, P. Y. and J. W. Ponder. 2002. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *Journal of Computational Chemistry.* 23, 1497-1506.

Ren, P. Y., C. J. Wu, and J. W. Ponder. 2011. Polarizable atomic multipole-based molecular mechanics for organic molecules. *Journal of Chemical Theory and Computation.* 7, 3143-3161.

Richmond, G. L., J. M. Robinson, and V. L. Shannon. 1988. 2nd harmonic-generation studies of interfacial structure and dynamics. *Progress in Surface Science.* 28, 1-70.

Rick, S. W. and S. J. Stuart. 2002. Potentials and algorithms for incorporating polarizability in computer simulations. In *Reviews in Computational Chemistry.* Volume 18 (Eds K. B. Lipkowitz and D. B. Boyd), John Wiley & Sons, Inc., Hoboken, New Jersey, USA.

Rick, S. W., S. J. Stuart, and B. J. Berne. 1994. Dynamical fluctuating charge force-fields: Application to liquid water. *Journal of Chemical Physics.* 101, 6141-6156.

Rivera, J. L., F. W. Starr, P. Paricaud, and P. T. Cummings. 2006. Polarizable contributions to the surface tension of liquid water. *Journal of Chemical Physics.* 125, 094712.

Robinson, R. A. and R. H. Stokes. 1959. *Electrolyte solutions.* London: Butterworths.

Rodgers, J. M., J. Sorensen, F. J. M. de Meyer, B. Schiott, and B. Smit. 2012. Understanding the phase behavior of coarse-grained model lipid bilayers through computational calorimetry. *Journal of Physical Chemistry B.* 116, 1551-1569.

Rogers, P. S. Z. and K. S. Pitzer. 1982. Volumetric properties of aqueous sodium chloride solutions. *Journal of Physical and Chemical Reference Data.* 11, 15-81.

Rowlinson, J. S. and F. L. Swinton. 1982. *Liquids and Liquid Mixtures.* 3rd ed. London: Butterworth Scientific.

Ruckenstein, E. and I. Shulgin. 2002. Salting-out or -in by fluctuation theory. *Industrial & Engineering Chemistry Research.* 41, 4674-4680.

Ruckenstein, E. and I. L. Shulgin. 2006. Effect of salts and organic additives on the solubility of proteins in aqueous solutions. *Advances in Colloid and Interface Science.* 123, 97-103.

Sarupria, S., T. Ghosh, A. E. Garcia, and S. Garde. 2010. Studying pressure denaturation of a protein by molecular dynamics simulations. *Proteins: Structure Function and Bioinformatics.* 78, 1641-1651.

Schmid, N., A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. van Gunsteren. 2011. Definition and testing of the GROMOS force-field versions 54a7 and 54b7. *European Biophysics Journal with Biophysics Letters.* 40, 843-856.

Schnell, S. K., X. Liu, J.-M. Simon, A. Bardow, D. Bedeaux, T. J. H. Vlugt, and S. Kjelstrup. 2011. Calculating thermodynamic properties from fluctuations at small scales. *Journal of Physical Chemistry B.* 115, 10911-10918.

Schofield, P. 1966. Wavelength-dependent fluctuations in classical fluids. I. Long wavelength limit. *Proceedings of the Physical Society of London.* 88, 149-170.

Segatin, N. and C. Klofutar. 2004. Limiting partial molar volumes of water in 1-hexanol, 1-octanol, 1-decanol, and cyclohexanol at 298.15 k. *Monatshefte Fur Chemie.* 135, 161-172.

Setny, P., R. Baron, and J. A. McCammon. 2010. How can hydrophobic association be enthalpy driven? *Journal of Chemical Theory and Computation.* 6, 2866-2871.

Shimizu, S. 2004. Estimating hydration changes upon biomolecular reactions from osmotic stress, high pressure, and preferential hydration experiments. *Proceedings of the National Academy of Sciences of the United States of America.* 101, 1195-1199.

Shimizu, S. and D. J. Smith. 2004. Preferential hydration and the exclusion of cosolvents from protein surfaces. *Journal of Chemical Physics.* 121, 1148-1154.

Shing, K. S. and S. T. Chung. 1988. Calculation of infinite-dilution partial molar properties by computer simulation. *AIChE Journal.* 34, 1973-1980.

Shulgin, I. and E. Ruckenstein. 2002. Henry's constant in mixed solvents from binary data. *Industrial & Engineering Chemistry Research.* 41, 1689-1694.

Shulgin, I. L. and E. Ruckenstein. 2005. A protein molecule in an aqueous mixed solvent: Fluctuation theory outlook. *Journal of Chemical Physics.* 123, 054909.

Singer, K. 1969. Monte Carlo calculations of the thermodynamic properties of mixtures of Lennard-Jones liquids. *Chemical Physics Letters.* 3, 164-166.

Smith, E. B. and K. R. Lea. 1960. Monte-Carlo equation of state for mixtures of hard-sphere molecules. *Nature.* 186, 714-714.

Smith, P. E. 1999. Computer simulation of cosolvent effects on hydrophobic hydration. *Journal of Physical Chemistry B.* 103, 525-534.

Smith, P. E. 2004a. Cosolvent interactions with biomolecules: Relating computer simulation data to experimental thermodynamic data. *Journal of Physical Chemistry B.* 108, 18716-18724.

Smith, P. E. 2004b. Local chemical potential equalization model for cosolvent effects on biomolecular equilibria. *Journal of Physical Chemistry B.* 108, 16271-16278.

Smith, P. E. 2006. Chemical potential derivatives and preferential interaction parameters in biological systems from Kirkwood-Buff theory. *Biophysical Journal.* 91, 849-856.

Smith, P. E. 2008. On the Kirkwood-Buff inversion procedure. *Journal of Chemical Physics.* 129, 124509.

Smith, P. E. 2010. The effect of urea on the morphology of NaCl crystals: A combined theoretical and simulation study. *Fluid Phase Equilibria.* 290, 36-42.

Smith, P. E., G. E. Marlow, and B. M. Pettitt. 1993. Peptides in ionic-solutions - A simulation study of a bis(penicillamine) enkephalin in sodium-acetate solution. *Journal of the American Chemical Society.* 115, 7493-7498.

Smith, P. E., E. Matteoli, and J. P. O'Connell. 2013. *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering and Biophysics*. Boca Raton: Taylor & Francis.

Smith, P. E. and B. M. Pettitt. 1991. Effects of salt on the structure and dynamics of the bis(penicillamine) enkephalin zwitterion - a simulation study. *Journal of the American Chemical Society.* 113, 6029-6037.

Smith, P. E. and B. M. Pettitt. 1992. Amino-acid side-chain populations in aqueous and saline solution - bis-penicillamine enkephalin. *Biopolymers.* 32, 1623-1629.

Sorin, E. J. and V. S. Pande. 2005. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophysical Journal.* 88, 2472-2493.

Soujanya, J., B. Satyavathi, and T. E. V. Prasad. 2010. Experimental (vapour + liquid) equilibrium data of (methanol + water), (water + glycerol) and (methanol + glycerol) systems at atmospheric and sub-atmospheric pressures. *Journal of Chemical Thermodynamics.* 42, 621-624.

Squire, P. G. and M. E. Himmel. 1979. Hydrodynamics and protein hydration. *Archives of Biochemistry and Biophysics.* 196, 165-177.

Straatsma, T. P. and J. A. McCammon. 1990. Molecular dynamics simulations with interaction potentials including polarization development of a noniterative method and application to water. *Molecular Simulation.* 5, 181-192.

Taulier, N. and T. V. Chalikian. 2002. Compressibility of protein transitions. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology.* 1595, 48-70.

Teitelbaum, B. Y., T. A. Gortalova, and E. E. Sidorova. 1951. Politermicheskoe issledovanie poverkhnostnogo natyazheniya vodnykh rastvorov nizshikh spirtov. *Zhurnal Fizicheskoi Khimii.* 25, 911-919.

Timasheff, S. N. 1998. Control of protein stability and reactions by weakly interacting cosolvents: The simplicity of the complicated. *Advances in Protein Chemistry.* 51, 355-432.

Tripathi, B. S., A. Awasthi, P. K. Pandey, and A. Awasthi. 2010. Thermal expansion coefficient of ternary liquid mixture using hard sphere models and flory's statistical theory. *Indian Journal of Physics.* 84, 449-458.

Tronel-Peyroz, E., J. M. Douillard, R. Bennes, and M. Privat. 1989. Application of the Kirkwood-Buff theory of solutions to the surface phase between the water ethanol binary-liquid mixture and its vapor. *Langmuir.* 5, 54-58.

Uosaki, Y., K. Ito, M. Kondo, S. Kitaura, and T. Moriyoshi. 2006. Effect of pressure on the static relative permittivities of alkan-1-ols at 298.15 k. *Journal of Chemical and Engineering Data.* 51, 1915-1921.

Valencia, J. L., D. Gonzalez-Salgado, J. Troncoso, J. Peleteiro, E. Carballo, and L. Romani. 2009. Thermophysical characterization of liquids using precise density and isobaric heat capacity measurements as a function of pressure. *Journal of Chemical and Engineering Data.* 54, 904-915.

Van't Hoff, J. H. 1887. Die rolle des osmotischen druckes in der analogie zwischen lösungen und gasen. *Zeitschrift für Physikalische Chemie, Stöchiometrie und Verwandtschaftslehre.* 1, 481-508.

Van't Hoff, J. H. 1894. How the theory of solutions arose. *Berichte der Deutschen Chemischen Gesellschaft.* 27, 6-19.

Van Belle, D., I. Couplet, M. Prevost, and S. J. Wodak. 1987. Calculations of electrostatic properties in proteins. Analysis of contributions from induced protein dipoles. *Journal of Molecular Biology.* 198, 721-735.

van der Waals, J. D. 1873. On the continuity of the gaseous and liquid states, Doctoral Thesis in Mathematics and Physics, University of Leiden.

van Gunsteren, W. F., D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glattli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu. 2006. Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie, International Edition in English.* 45, 4064-4092.

van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. 1996. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Zürich, Switzerland: Vdf Hochschulverlag at the ETH Zürich.

Van Laar, J. J. 1906. *Six lectures over the thermodynamic potential and its applications on chemical and physical equilibrium problems, introduced with two lectures over not-diluted solutions and the osmotic pressure. [machine translation]*.

Vannhu, N., S. N. Bhat, and F. Kohler. 1987. Calculation of isothermal compressibilities of mixtures - the example acetone + chloroform. *Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics*. 91, 525-528.

Vazquez, G., E. Alvarez, and J. M. Navaza. 1995. Surface tension of alcohol + water from 20-degrees-C to 50-degrees-C. *Journal of Chemical and Engineering Data*. 40, 611-614.

Vega, C. and J. L. F. Abascal. 2011. Simulating water with rigid non-polarizable models: A general perspective. *Physical Chemistry Chemical Physics*. 13, 19663-19688.

Vercher, E., A. V. Orchilles, P. J. Miguel, and A. Martinez-Andreu. 2007. Volumetric and ultrasonic studies of 1-ethyl-3-methylimidazolium trifluoromethanesulfonate ionic liquid with methanol, ethanol, 1-propanol, and water at several temperatures. *Journal of Chemical and Engineering Data*. 52, 1468-1482.

Vesely, F. J. 1977. N-particle dynamics of polarizable stockmayer-type molecules. *Journal of Computational Physics*. 24, 361-371.

Villamanan, M. A., C. Gonzalez, and H. C. Van Ness. 1984. Excess thermodynamic properties for water ethylene-glycol. *Journal of Chemical and Engineering Data*. 29, 427-429.

Villamanan, M. A. and H. C. van Ness. 1984. Excess thermodynamic properties for water acetone. *Journal of Chemical and Engineering Data*. 29, 429-431.

Vorobyov, I. V., V. M. Anisimov, and A. D. Mackerell. 2005. Polarizable empirical force field for alkanes based on the classical Drude oscillator model. *Journal of Physical Chemistry B*. 109, 18988-18999.

Vosmeer, C. R., A. S. Rustenburg, J. E. Rice, H. W. Horn, W. C. Swope, and D. P. Geerke. 2012. QM/MM-based fitting of atomic polarizabilities for use in condensed-phase biomolecular simulation. *Journal of Chemical Theory and Computation*. 8, 3839-3853.

Wagner, W. and H.-J. Kretzschmar. 2008. *Part B: Tables of the Properties of Water and Steam, International Steam Tables: Properties of Water and Steam Based on the Industrial Formulation IAPWS-IF97*. Berlin: Springer.

Wagner, W. and A. Pruss. 2002. The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *Journal of Physical and Chemical Reference Data*. 31, 387-535.

Wang, J. M., P. Cieplak, and P. A. Kollman. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*. 21, 1049-1074.

Wang, W., O. Donini, C. M. Reyes, and P. A. Kollman. 2001. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annual Review of Biophysics and Biomolecular Structure.* 30, 211-243.

Warshel, A., M. Kato, and A. V. Pisliakov. 2007. Polarizable force fields: History, test cases, and prospects. *Journal of Chemical Theory and Computation.* 3, 2034-2045.

Warshel, A. and M. Levitt. 1976. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology.* 103, 227-249.

Watanabe, H. and H. Kato. 2004. Thermal conductivity and thermal diffusivity of twenty-nine liquids: Alkenes, cyclic (alkanes, alkenes, alkadienes, aromatics), and deuterated hydrocarbons. *Journal of Chemical and Engineering Data.* 49, 809-825.

Wedberg, R., J. P. O'Connell, G. H. Peters, and J. Abildskov. 2010. Accurate Kirkwood-Buff integrals from molecular simulations. *Molecular Simulation.* 36, 1243-1252.

Weerasinghe, S., M. B. Gee, M. Kang, N. Bentenitis, and P. E. Smith. 2010. Developing force fields from the microscopic structure of solutions: The Kirkwood–Buff approach. In *Modeling Solvent Environments*: *Applications to Simulations of Biomolecules*, Wiley-VCH Verlag GmbH & Co. KGaA.

Weerasinghe, S. and P. E. Smith. 2003a. Cavity formation and preferential interactions in urea solutions: Dependence on urea aggregation. *Journal of Chemical Physics.* 118, 5901-5910.

Weerasinghe, S. and P. E. Smith. 2003b. Kirkwood-Buff derived force field for mixtures of acetone and water. *Journal of Chemical Physics.* 118, 10663-10670.

Weerasinghe, S. and P. E. Smith. 2003c. A Kirkwood-Buff derived force field for mixtures of urea and water. *Journal of Physical Chemistry B.* 107, 3891-3898.

Weerasinghe, S. and P. E. Smith. 2003d. A Kirkwood-Buff derived force field for sodium chloride in water. *Journal of Chemical Physics.* 119, 11342-11349.

Weerasinghe, S. and P. E. Smith. 2004. A Kirkwood-Buff derived force field for the simulation of aqueous guanidinium chloride solutions. *Journal of Chemical Physics.* 121, 2180-2186.

Weerasinghe, S. and P. E. Smith. 2005. A Kirkwood-Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B.* 109, 15080-15086.

Weerasinghe, S. and P. E. Smith. 2014. In preparation.

Wertheim, M. S. 1984a. Fluids with highly directional attractive forces. 1. Statistical thermodynamics. *Journal of Statistical Physics.* 35, 19-34.

Wertheim, M. S. 1984b. Fluids with highly directional attractive forces. 2. Thermodynamic perturbation-theory and integral-equations. *Journal of Statistical Physics.* 35, 35-47.

Wilson, G. M. 1964. Vapor-liquid equilibrium. 11. New expression for excess free energy of mixing. *Journal of the American Chemical Society.* 86, 127-130.

Xie, W., J. Pu, A. D. Mackerell, Jr., and J. Gao. 2007. Development of a polarizable intermolecular potential function (PIPF) for liquid amides and alkanes. *Journal of Chemical Theory and Computation.* 3, 1878-1889.

Yang, P. H. and J. A. Rupley. 1979. Protein-water interactions: Heat capacity of the lysozyme-water system. *Biochemistry.* 18, 2654-2661.

Yang, Z.-Z. and Q. Zhang. 2006. Study of peptide conformation in terms of the ABEEM/MM method. *Journal of Computational Chemistry.* 27, 1-10.

Yu, H. B., D. P. Geerke, H. Y. Liu, and W. F. van Gunsteren. 2006. Molecular dynamics simulations of liquid methanol and methanol-water mixtures with polarizable models. *Journal of Computational Chemistry.* 27, 1494-1504.

Yu, H. B., T. Hansson, and W. F. van Gunsteren. 2003. Development of a simple, self-consistent polarizable model for liquid water. *Journal of Chemical Physics.* 118, 221-234.

Yu, H. B. and W. F. van Gunsteren. 2004. Charge-on-spring polarizable water models revisited: From water clusters to liquid water to ice. *Journal of Chemical Physics.* 121, 9549-9564.

Yu, H. B. and W. F. van Gunsteren. 2005. Accounting for polarization in molecular simulation. *Computer Physics Communications.* 172, 69-85.

Yu, I., T. Tasaki, K. Nakada, and M. Nagaoka. 2010. Influence of hydrostatic pressure on dynamics and spatial distribution of protein partial molar volume: Time-resolved surficial Kirkwood-Buff approach. *Journal of Physical Chemistry B.* 114, 12392-12397.

Zhang, C. and J. P. Ma. 2010. Enhanced sampling and applications in protein folding in explicit solvent. *Journal of Chemical Physics.* 132, 244101.

Zhong, Y. and S. Patel. 2010. Nonadditive empirical force fields for short-chain linear alcohols: Methanol to butanol. Hydration free energetics and Kirkwood-Buff analysis using charge equilibration models. *Journal of Physical Chemistry B.* 114, 11076-11092.

Zhong, Y., G. L. Warren, and S. Patel. 2008. Thermodynamic and structural properties of methanol-water solutions using nonadditive interaction models. *Journal of Computational Chemistry.* 29, 1142-1152.

Zhu, X., P. E. M. Lopes, and A. D. Mackerell, Jr. 2012. Recent developments and applications of the CHARMM force fields. *Wiley Interdisciplinary Reviews-Computational Molecular Science.* 2, 167-185.

Zorebski, E., M. Geppert-Rybczynska, and B. Maciej. 2010. Densities, speeds of sound, and isentropic compressibilities for binary mixtures of 2-ethyl-1-hexanol with 1-pentanol, 1-heptanol, or 1-nonanol at the temperature 298.15 K. *Journal of Chemical and Engineering Data.* 55, 1025-1029.

Zwanzig, R. W. 1954. High-temperature equation of state by a perturbation method. 1. Nonpolar gases. *Journal of Chemical Physics.* 22, 1420-1426.

Appendix A - Copyright Clearance

Taylor and Francis Group LLC Books LICENSE
TERMS AND CONDITIONS

This is a License Agreement between Elizabeth Ploetz ("You") and Taylor and Francis Group LLC Books ("Taylor and Francis Group LLC Books") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Taylor and Francis Group LLC Books, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3356770431752
License date	Mar 26, 2014
Licensed content publisher	Taylor and Francis Group LLC Books
Licensed content title	Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering, and Biophysics
Licensed content author	Paul E. Smith, Enrico Matteoli, John P. O' Connell
Licensed content date	02/22/2013
Type of Use	Thesis/Dissertation
Requestor type	Author of requested content
Format	Print, Electronic
Portion	chapter/article
Number of pages in chapter/article	15
Title or numeric reference of the portion(s)	Chapter 5
Title of the article or chapter the portion is from	Accurate Force Fields for Molecular Simulation
Editor of portion(s)	Paul E. Smith, Enrico Matteoli, and John P. O'Connell
Author of portion(s)	Elizabeth A. Ploetz, Samantha Weerasinghe, Myungshim Kang, and Paul E. Smith
Volume of serial or monograph.	N/A
Issue, if republishing an article from a serial	N/A
Page range of the portion	Chapter 5, pages 117-131
Publication date of portion	02/22/2013
Rights for	Main product
Duration of use	Life of current and all future editions

Creation of copies for the disabled	no
With minor editing privileges	yes
For distribution to	United States
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of new product	0 to 499
Made available in the following markets	Academia
The requesting person/organization is:	Elizabeth A. Ploetz
Order reference number	
Author/Editor	Elizabeth A. Ploetz
The standard identifier	Dissertation
Title	Fluctuation Solution Theory
Publisher	Elizabeth A. Ploetz
Expected publication date	Apr 2014
Estimated size (pages)	400
Total (may include CCC user fee)	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

The following terms are individual to this publisher:

Taylor and Francis Group and Informa healthcare are division of Informa plc. Permission will be void if material exceeds 10% of all the total pages in your publication and over 20% of the original publication. This includes permission granted by Informa plc and all of its subsidaries.

Other Terms and Conditions:

Please make sure the appropriate source line is credited under the requested material.

STANDARD TERMS AND CONDITIONS

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the

Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

3.1 All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

3.2 General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, Dept 001, P.O. Box 843006, Boston, MA 02284-3006. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3 Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4 In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible

for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5 Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6 User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS,

GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. **Miscellaneous.**

8.1 User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2 Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here: http://www.copyright.com/content/cc3/en/tools/footer/privacypolicy.html.

8.3 The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

8.4 No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5 The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to

the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to info@copyright.com.

v 1.1

Step 3: Order Confirmation

Confirmation Number: 11168162
Order Date: 03/18/2014

Payment Information

Elizabeth Ploetz
ploetz@ksu.edu
+1 (913)7080567
Payment Method: n/a

Order Details

The Journal of chemical physics

Order detail ID:	64531907
Order License Id:	3352101351794
Article Title:	Local fluctuations in solution mixtures
Author(s):	Ploetz, Elizabeth A. ; Smith, Paul E.
DOI:	10.1063/1.3615718
Date:	Jan 01, 2011
ISSN:	0021-9606
Publication Type:	Journal
Volume:	135
Issue:	4
Start page:	044506
Publisher:	AMERICAN INSTITUTE OF PHYSICS.
Author/Editor:	AMERICAN INSTITUTE OF PHYSICS

Permission Status: ✅ **Granted**

Permission type: Republish or display content
Type of use: reuse in a thesis/dissertation

Requestor type	Author (original article)
Format	Print and electronic
Portion	Excerpt (> 800 words)
Will you be translating?	No
Title of your thesis / dissertation	Fluctuation Solution Theory
Expected completion date	Apr 2014
Estimated size (number of pages)	400

Note: This item will be invoiced or charged separately through CCC's **RightsLink** service. More info

$ 0.00

Advances in Chemical Physics

With incidental promotional use	no
Lifetime unit quantity of new product	0 to 499
Made available in the following markets	Academia
The requesting person/organization	Elizabeth A. Ploetz
Order reference number	
Author/Editor	Elizabeth Anne Ploetz
The standard identifier	Dissertation
Title	Fluctuation Solution Theory
Publisher	Elizabeth Anne Ploetz
Expected publication date	Apr 2014
Estimated size (pages)	400

Note: This item will be invoiced or charged separately through CCC's **RightsLink** service. More info   **$ 0.00**

Proteins

Order detail ID: 64531954
Order License Id: 3352101355778

ISSN: 1097-0134
Publication Type: e-Journal
Volume:
Issue:
Start page:
Publisher: JOHN WILEY & SONS, INC.

Permission Status: ✅ **Granted**

Permission type: Republish or display content
Type of use: Republish in a thesis/dissertation

Requestor type	Author of requested content
Format	Print, Electronic
Portion	chapter/article
Number of pages in chapter/article	11
Title or numeric reference of the portion(s)	pages 2758-2768

Title of the article or chapter the portion is from	Flexible connection of the N-terminal domain in ClpB modulates substrate binding and the aggregate reactivation efficiency
Editor of portion(s)	N/A
Author of portion(s)	Ting Zhang, Elizabeth A. Ploetz, Maria Nagy, Shannon M. Doyle, Sue Wickner, Paul E. Smith, and Michal Zolkiewski
Volume of serial or monograph	80
Issue, if republishing an article from a serial	12
Page range of portion	2758-2768
Publication date of portion	15 SEP 2012
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	yes
For distribution to	United States
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	0 to 499
Made available in the following markets	Academia
The requesting person/organization	Elizabeth A. Ploetz
Order reference number	

Author/Editor	Elizabeth A. Ploetz
The standard identifier	Dissertation
Title	Fluctuation Solution Theory
Publisher	Elizabeth A. Ploetz
Expected publication date	Apr 2014
Estimated size (pages)	400

Note: This item will be invoiced or charged separately through CCC's **RightsLink** service. More info   **$ 0.00**

This is not an invoice.

Total order items: 3   **Order Total: $ 0.00**

Confirmation Number: 11168162

Special Rightsholder Terms & Conditions

The following terms & conditions apply to the specific publication under which they are listed

The Journal of chemical physics
Permission type: Republish or display content
Type of use: reuse in a thesis/dissertation

Advances in Chemical Physics
Permission type: Republish or display content
Type of use: Republish in a thesis/dissertation

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC�s Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state�s conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

Other terms and conditions:

REPUBLICATION IN MAIN PRODUCT, PRODUCT FAMILY, COMPILATIONS OR COLLECTIVE WORKS

Limitations on reuse of Wiley Materials:

Where the following additional rights have been granted as part of this Agreement, please note that in no instance may the total amount of Wiley Materials used in any Main Product, Compilation or Collective work comprise more than 5% (if figures/tables) or 15% (if full articles/chapters) of the Main Product, Compilation or Collective Work.

Product Family: main product AND any product related to the main product

Main Product and related products which are created to supplement or add value to the Main Product, and in which the overall content of the Main Product remains substantially the same with relatively minor additions or variations. Examples include: ancillaries, instructor guides, testing materials, student subject-driven resources, abridgements, and custom editions that are substantially based on a single Main Product.

Other: main product AND other compilations or collective works

Main Product and new collective works produced by requestor that do not consist of substantially the same material as an individual Main Product. For example, licensees may wish to acquire rights for use in the Main Product AND for use in separate projects within the same subject discipline as the Main Product. This would include, for example, custom editions or compilations which incorporate chapters from multiple underlying works within or across disciplines.

All: main product, AND Product Family AND Other compilations or collective works

This combines the right to republish in the following components: 1) main product, 2) products related to the main product (i.e. Product Family) and 3) other compilations/collective works.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses:: Creative Commons Attribution (CC-BY) license, Creative Commons Attribution Non-Commercial (CC-BY-NC) license and Creative Commons Attribution Non-Commercial-NoDerivs (CC-BY-NC-ND) License. The license type is clearly identified on the article.

Copyright in any research article in a journal published as Open Access under a Creative Commons License is retained by the author(s). Authors grant Wiley a license to publish the article and identify itself as the original publisher. Authors also grant any third party the right to use the article freely as long as its integrity is maintained and its original authors, citation details and publisher are identified as follows: [Title of Article/Author/Journal Title and Volume/Issue. Copyright (c) [year] [copyright owner as specified in the Journal]. Links to the final article on Wiley�s website are encouraged where applicable.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-commercial re-use of an open access article, as long as the author is properly attributed.

The Creative Commons Attribution License does not affect the moral rights of authors, including without limitation the right not to have their work subjected to derogatory treatment. It also does not affect any other rights held by authors or third parties in the article, including without limitation the rights of privacy and publicity. Use of the article must not

republication.2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.3. Scope of License; Limitations and Obligations.3.1 All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.3.2 General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, Dept 001, P.O. Box 843006, Boston, MA 02284-3006. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.3.3 Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).3.4 In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.3.5 Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.3.6 User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use

plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.8. Miscellaneous.8.1 User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.8.2 Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here: http://www.copyright.com/content/cc3/en/tools/footer/privacypolicy.html.8.3 The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.8.4 No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.8.5 The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to info@copyright.com.v 1.1


Proteins
Permission type: Republish or display content
Type of use: Republish in a thesis/dissertation

TERMS AND CONDITIONSThe following terms are individual to this publisher:

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a �Wiley Company�) or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively �WILEY�).   By clicking �accept� in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., (�CCC�s Billing and Payment terms and conditions�), at the time that you opened your Rightslink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this licence must be completed within two years of the date of the grant of this licence (although copies prepared before the end date may be distributed thereafter).   The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license.   Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material.   Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto.

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC�s Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC�s Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC�s Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state�s conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

Other terms and conditions:

REPUBLICATION IN MAIN PRODUCT, PRODUCT FAMILY, COMPILATIONS OR COLLECTIVE WORKS

Limitations on reuse of Wiley Materials:

Where the following additional rights have been granted as part of this Agreement, please note that in no instance may the total amount of Wiley Materials used in any Main Product, Compilation or Collective work comprise more than 5% (if figures/tables) or 15% (if full articles/chapters) of the Main Product, Compilation or Collective Work.

Product Family: main product AND any product related to the main product

Main Product and related products which are created to supplement or add value to the Main Product, and in which the overall content of the Main Product remains substantially the same with relatively minor additions or variations. Examples include: ancillaries, instructor guides, testing materials, student subject-driven resources, abridgements, and custom editions that are substantially based on a single Main Product.

Other: main product AND other compilations or collective works

Main Product and new collective works produced by requestor that do not consist of substantially the same material as an individual Main Product. For example, licensees may wish to acquire rights for use in the Main Product AND for use in separate projects within the same subject discipline as the Main Product. This would include, for example, custom editions or compilations which incorporate chapters from multiple underlying works within or across disciplines.

All: main product, AND Product Family AND Other compilations or collective works

This combines the right to republish in the following components: 1) main product, 2) products related to the main product (i.e. Product Family) and 3) other compilations/collective works.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses:: Creative Commons Attribution (CC-BY) license, Creative Commons Attribution Non-Commercial (CC-BY-NC) license and Creative Commons Attribution Non-Commercial-NoDerivs (CC-BY-NC-ND) License. The license type is clearly identified on the article.

Copyright in any research article in a journal published as Open Access under a Creative Commons License is retained by the author(s). Authors grant Wiley a license to publish the article and identify itself as the original publisher. Authors also grant any third party the right to use the article freely as long as its integrity is maintained and its original authors, citation details and publisher are identified as follows: [Title of Article/Author/Journal Title and Volume/Issue. Copyright (c) [year] [copyright owner as specified in the Journal]. Links to the final article on Wiley�s website are encouraged where applicable.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-commercial re-use of an open access article, as long as the author is properly attributed.

The Creative Commons Attribution License does not affect the moral rights of authors, including without limitation the right not to have their work subjected to derogatory treatment. It also does not affect any other rights held by authors or third parties in the article, including without limitation the rights of privacy and publicity. Use of the article must not assert or imply, whether implicitly or explicitly, any connection with, endorsement or sponsorship of such use by the author, publisher or any other party associated with the article.

For any reuse or distribution, users must include the copyright notice and make clear to others that the article is made available under a Creative Commons Attribution license, linking to the relevant Creative Commons web page.

To the fullest extent permitted by applicable law, the article is made available as is and without representation or warranties of any kind whether express, implied, statutory or otherwise and including, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of defects, accuracy, or the presence or absence of errors.

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC) License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by non-commercial users

For non-commercial and non-promotional purposes, individual users may access, download, copy, display and redistribute to colleagues Wiley Open Access articles, as well as adapt, translate, text- and data-mine the content subject to the following conditions:

- The authors' moral rights are not compromised. These rights include the right of "paternity" (also known as "attribution" - the right for the author to be identified as such) and "integrity" (the right for the author not to have the work altered in such a way that the author's reputation or integrity may be impugned).

- Where content in the article is identified as belonging to a third party, it is the obligation of the user to ensure that any reuse complies with the copyright policies of the owner of that content.

- If article content is copied, downloaded or otherwise reused for non-commercial research and education purposes, a link to the appropriate bibliographic citation (authors, journal, article title, volume, issue, page numbers, DOI and the link to the definitive published version on Wiley Online Library) should be maintained. Copyright notices and disclaimers must not be deleted.

- Any translations, for which a prior translation agreement with Wiley has not been agreed, must prominently display the statement: "This is an unofficial translation of an article that appeared in a Wiley publication. The publisher has not endorsed this translation."

Use by commercial "for-profit" organisations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee. Commercial purposes include:

- Copying or downloading of articles, or linking to such articles for further redistribution, sale or licensing;

- Copying, downloading or posting by a site or service that incorporates advertising with such content;

- The inclusion or incorporation of article content in other works or services (other than normal quotations with an appropriate citation) that is then available for sale or licensing, for a fee (for example, a compilation produced for marketing purposes, inclusion in a sales pack)

- Use of article content (other than normal quotations with appropriate citation) by for-profit organisations for promotional purposes

- Linking to article content in e-mails redistributed for promotional, marketing or educational purposes;

- Use for the purposes of monetary reward by means of sale, resale, licence, loan, transfer or other form of commercial exploitation such as marketing products

- Print reprints of Wiley Open Access articles can be purchased from: corporatesales@wiley.com

Further details can be found on Wiley Online Library http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:NoneSTANDARD TERMS AND CONDITIONS1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.3. Scope of License; Limitations and Obligations.3.1 All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.3.2 General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, Dept 001, P.O. Box 843006, Boston, MA 02284-3006. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.3.3 Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions

and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).3.4 In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.3.5 Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.3.6 User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.8. Miscellaneous.8.1 User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.8.2 Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here: http://www.copyright.com/content/cc3/en/tools/footer/privacypolicy.html.8.3 The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.8.4 No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.8.5 The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation.

The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to info@copyright.com.v 1.1

Close

Confirmation Number: 11168162

Citation Information

Order Detail ID: 64531907

The Journal of chemical physics by AMERICAN INSTITUTE OF PHYSICS Reproduced with permission of AMERICAN INSTITUTE OF PHYSICS. in the format reuse in a thesis/dissertation via Copyright Clearance Center.

Order Detail ID: 64531945

Advances in Chemical Physics by Wiley. Reproduced with permission of Wiley in the format Republish in a thesis/dissertation via Copyright Clearance Center.

Order Detail ID: 64531954

Proteins by JOHN WILEY & SONS, INC.. Reproduced with permission of JOHN WILEY & SONS, INC. in the format Republish in a thesis/dissertation via Copyright Clearance Center.

Close