A KIRKWOOD-BUFF FORCE FIELD FOR AROMATIC AMINO ACIDS

by

ELIZABETH ANNE PLOETZ

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Graduate Biochemistry Group
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2010

Approved by:

Major Professor
Paul E. Smith
Department of Chemistry

Copyright

Abstract

We are developing a force field (FF) for molecular dynamics (MD) simulations of peptides and small proteins that is grounded in the Kirkwood-Buff theory of solutions. Here we present the Kirkwood-Buff Force Field (KBFF) parameters for the aromatic amino acids, based upon simulations of binary mixtures of small molecules representative of these amino acids over their entire composition ranges (excluding Histidine). Many aromatics are not fully soluble in water, so they have instead been studied in solvents of methanol or toluene. The parameters were developed by studying the following binary solutions: Phenylalanine– benzene + methanol, toluene + methanol, and toluene + benzene; Tyrosine– toluene + phenol and toluene + *p*-Cresol; Tryptophan– pyrrole + methanol and indole + methanol; Histidine– pyrrole + methanol, pyridine + methanol, pyridine + water, histidine + water (at 0.25 molal), and histidine monohydrochloride + water (at 0.3 molal and 0.6 molal). Our simulations reproduce the Kirkwood-Buff integrals, which guarantees that the KBFF provides an adequate balance of solute-solvent, solute-solute, and solvent-solvent interactions. Additionally, we show that the KBFF does not sacrifice reproduction of other solution properties in order to achieve this improved description of intermolecular interactions. We present these results as validating evidence for the future use of the KBFF in simulations of peptides and small proteins.

Table of Contents

List of Figures vi

List of Tables viii

Acknowledgements ix

Dedication xii

Chapter 1: Introduction 1

Fundamentals 1
Force Fields 2
The Biological FF Menagerie: Approaches and Parameter Sources 3
General Functional Forms Used in Protein Force Fields 5
The Need for an Improved Force Field 10
On the Effects of the Gas to Liquid Phase Transition 11
Our Guide, The Kirkwood-Buff (KB) Theory of Solutions 17
Justifying Our Dependence on KB Theory 17
Derivation 21
On Modeling 28
Aims of This Work 29

Chapter 2: The Kirkwood-Buff Derived Force Field for Phenylalanine, Tyrosine, and Tryptophan 30

Abstract 30
Introduction 31
Methods 33
Molecular Dynamics Simulations 33
Parameter Development: Lennard Jones Parameters 36
Kirkwood-Buff Analysis of Simulated Data 38
Kirkwood-Buff Analysis of Experimental Data 39
Experimental Sources for Composition and Activity Data 42
Parameter Development: Partial Charges 45
Results and Discussion 45
The rdf, KBI, and N_{ij} 45
Other Solution Properties 55
Conclusions and Future Direction 58
Appendix A 59

Chapter 3: The Kirkwood-Buff Derived Force Field for Histidine 60

Abstract 60
Introduction 60
Approach 61
Theory 62
Salts 63
Methods 64
Molecular Dynamics Simulations 64
Experimental Sources for Composition and Activity Data 65
Results 66
Conclusion 69
Appendix B 76

Summary and Future Direction 77

References 79

List of Figures

Figure 1.1: A. Benzene Model with C_{ar} atom types labeled. B. V_{bond}, C. V_{angle}, D. V_{vdw}, E. $V_{Coulomb.}$ 9

Figure 1.2: A. HistidineH Model with atom types labeled. B. $V_{torsion}$ C. $V_{improper}$ 10

Figure 1.3: Relationship between the rdf and total energy, pressure, and chemical potential for a pure liquid in the *NVT* ensemble 19

Figure 1.4:Web of Science search results for "Kirkwood-Buff AND fluctuation theory of solutions" 21

Figure 1.5: Illustrative radial distribution function 25

Figure 1.6: Illustrative Kirkwood-Buff integral 26

Figure 1.7: Illustrative excess coordination number 27

Figure 2.1: Systems of Interest. Phenylalanine, Tyrosine, and Tryptophan with atom types labeled and systems studied to develop their charge distributions 34

Figure 2.2: Sensitivity of the KBIs to the charge distribution 36

Figure 2.3: Sensitivity of the MOH-MOH aggregation to the Ben charge distribution from a visual inspection of the Ben-MOH system at $x_{Ben} = 0.500$ 37

Figure 2.4: Unit cell dimensions of the benzene crystal structure 38

Figure 2.5: Center of mass based radial distribution function for Ben + MOH system as a function of distance 47

Figure 2.6: Kirkwood-Buff integrals for Ben + MOH 47

Figure 2.7: Excess coordination numbers for Ben + MOH 49

Figure 2.8: Representative snapshots of a methanol chain and ring observed in simulations of Ben + MOH at 308 K with $x_{Ben} = 0.9$ 50

Figure 2.9: Center of mass based radial distribution functions for six systems 50

Figure 2.10: Kirkwood-Buff integrals for six systems 51

Figure 2.11: Excess coordination numbers for six systems 51

Figure 2.12: Representative snapshots of phenol and *p*-cresol cyclic ring structures observed in simulations of Tol + PhOH or Tol + pCr at 333 K and $x_{Tol} = 0.9$ 52

Figure 2.13: Simulated dielectric constants of binary mixtures as a function of composition 56

Figure 2.14: Enthalpy of mixing for seven systems 57

Figure 3.1: HistidineH and the two tautomers of neutral Histidine with atom types labeled and systems studied to develop their charge distributions 62

Figure 3.2: Center of mass based rdf for Pyridine + Methanol and Pyridine + Water 70

Figure 3.3: Center of mass based rdf for Histidine + Water 71

Figure 3.4: Center of mass based rdf for HistidineHCl + Water 71

Figure 3.5: Kirkwood-Buff integrals for Pyridine + Methanol and Pyridine + Water 72

Figure 3.6: Kirkwood-Buff integrals for Histidine + Water 72

Figure 3.7: Kirkwood-Buff integrals for HistidineHCl + Water 73

Figure 3.8: Excess coordination numbers for Pyridine + Methanol and Pyridine + Water 73

Figure 3.9: Excess coordination numbers for Histidine + Water 74

Figure 3.10: Excess coordination numbers for HistidineHCl + Water 74

Figure 3.11: Enthalpy of Mixing for Pyridine + Methanol and Pyridine + Water 75

Figure 3.12: Snapshot of 0.3 *m* HisHCl 75

List of Tables

Table 1.1: Comparison of Select Force Fields Used in Protein Simulations. 4

Table 1.2: (Experimental) ΔG_{solv}, (Experimental) ΔH_{vap}, (Experimental) δ Solubility Parameter, (Experimental) Molecular Polarizability, α. 13

Table 1.3: Select Solvent Scales: Dielectric Constant, π* scale, α scale, β scale. 15

Table 1.4: Pauling & Mulliken Electronegativity Scales for H, C, and O. 15

Table 1.5: Summary of Smith Group's Progress Towards Completion of KBFF v.1 29

Table 2.1: Non-bonded Parameters for Phe, Tyr, and Trp. 46

Table 2.2: Comparison Between Simulated and Experimental Properties of Pure Liquids with Temperatures as Noted and Pressures at 1bar. 54

Table A.1: Simulation Details for Ben + MOH, Tol + MOH, and Tol + Ben Solution Mixtures. 59

Table A.2: Simulation Details for Tol + PhOH, Tol + pCr, Pyr + MOH, and Ind + MOH Solution Mixtures. 59

Table 3.1: Non-bonded Parameters for KBFF Models of Pyridine, His, and HisH. 68

Table 3.2: Comparison of Experimental and Simulated Properties of Liquid Pyridine at 298 K. 69

Table B.1: Simulation Details Pyrd + MOH, Pyrd + HOH, HisHCl + HOH, and His + HOH Solution Mixtures. 76

Acknowledgements

Professor Paul E. Smith: This work is your inspiration and your daily investments have made the results a reality. The longer I have worked for you, the more my respect for you has grown -- "hand on heart". For one, you are not impressed with fancy words and elaborate figures; I am. You focus on concisely conveying accurate, distilled messages based upon rigorous preliminary work; I tend to make things unnecessarily complicated in their presentation. My approach can create a message that isolates rather than engages (*e.g.*, "Explicating?"). I am trying to let your style rub off on me, even if just to a degree. *Thank you for guiding and correcting me with patience, and for being a wealth of wisdom and a strong source of encouragement.* I am truly honored that I get to "sit at your feet" as a doctoral student; my hope is to soak up every minute and to have fun helping you reach your goals.

The Smith group (Samantha Weerasinghe, Myungshim Kang, Feng Chen, Moon Bae Gee, Yuanfang Jiao, Jin Zou, Chester McDowell, and Shu Dai): You have created an atmosphere of camaraderie in the lab. Thank you for helping me learn the strange ways of molecular dynamics and for letting me "steal" your machines in the pre-cluster era.

Frederick W. and Sandra K. Ploetz: I am blown away by your wholehearted interest in my pursuits, no matter how foreign they are to you. Had I not experienced the nurturing environment you created in my most formative years, I do not know where I

would be today. I have witnessed sobering events, including supervising the last visit between a mother and her children before they were permanently separated due to criminal activity in the home. In stark contrast to those children, my life has been blessed without warrant. You have given me immeasurable advantages by your sacrifices. *Thank your for your constant prayers, all encompassing support, and wise perspective.*

Christopher D. Ploetz: I remember the tone of your voice in 2008 after I asked if you had access to *Fluid Phase Equilibiria*. You responded, "*What* are *you* doing?" At the time, I had no idea that I was diving into a world so closely related to your daily work. It is fun to share this peculiar interest with my big brother; thank you especially for our conversations (and your calculations) about bubble-point VLE.

Riley County EMS: We are bombarded with the limitations of science and medicine and we see how life -- this grand, philosophical topic -- is also but an easily toppled, precarious balance of measurable quantities. Reflect on the following recent calls:

- Our CVA-alert pt. missed the "golden hour" for thrombolytics by *x* min. too many
- Our entrapped pt. was under a *p x* psi too high or his extrication took *x* min. too long
- Our EtOH pt., found unresponsive in the snow, had a core T *x* degrees too low
- Our pediatric Code Blue was shaken *x* times too many with a force *x* N too great
- Our Code Black OD had taken *x* mg/kg too much or *x* hits too many

To be sure, I love flying down the road while hitting the sirens, tearing through the map books, and contacting dispatch, but I also savor each moment of peace and quiet (*i.e.*, the

raging cluster) in the lab and the normalcy of it all. As author Mary Jean Irion so eloquently wrote,

> "Normal day, let me be aware of the treasure you are . . . let me not pass you by in quest for some rare and perfect tomorrow. One day I shall dig my nails into the earth, or bury my head in my pillow . . . or raise my head to the sky, and want, more than all in the world, your return."

Thank you all for the constant banter and laughter, for being my home-away-from-home, and for occasionally, albeit begrudgingly, letting me be studious during downtime. Special thanks to **Cpt. David Adams** (for holding me accountable in my studies), former **Lt. Bill Hartloff** (for understanding my desire for extensive map training), **Cpt. Joel Reimer (**for nightly ice-cream and for your contagious giggling during *The Office*), **Cpt. David Vance** (for entrusting me with our trucks), and **Lt. Galen Slough** (for your wonderfully ridiculous stories -- but I want to fall asleep!). Lastly, to my unlucky OOT transfer partner **John Griffin** for our ritual gas station breakfasts and our record breaking transfer times so that we can indeed get off of work at 0700 hrs.

National Institutes of Health, National Institute of General Medical Sciences: Your funding is appreciated.

Dedication

To Nick Deneault,
Action cures fear.

Chapter 1: Introduction

"It seems to me that the test of, 'Do we not understand a particular subject in physics?' is, 'Can we make a mechanical model of it?'"

William Thomson, Lord Kelvin (1824 – 1907) Baltimore Lecture XI

Fundamentals

The usage of the term *molecular dynamics* (MD) is not unique. In this work, it means the use of computer simulations to create pictures of how molecules move in time *i.e.*, how they traverse the expanse of their container while rotating and colliding with each other and their container's walls. MD, one in a number of computer simulation methods, numerically integrates Newton's equations of motion to describe the momentum, position, and orientation of each particle in a container once per step in "time" to generate a trajectory. The nature of MD is therefore deterministic, meaning that a given input demands a certain output, in contrast with the purely stochastic Metropolis Monte Carlo simulation technique.

The careful reader may be alarmed by the use of Newton's equations of motion; clearly, molecules are more than a mere collection of balls (atoms) held together by springs (bonds). Although a classical mechanics (CM) approach has limitations (*e.g.*, an inability to simulate charge transfer or chemical reactions), the adoption of a quantum mechanical (QM) approach would be, as yet, inconceivable due to the large systems (boxes < 10 nm in length containing $O(10^4\text{-}10^6)$ atoms) and relatively long timescales (ns - μs) that we wish to study.

In a simulation, the system of interest is described by a Hamiltonian (H) that can be divided into kinetic energy (K) and potential energy (V) terms, $H(\mathbf{p},\mathbf{r};m,s) = K(\mathbf{p};m) + V(\mathbf{r};s)$, in which the K is independent of the atomic coordinates, $\mathbf{r}_i$, but is a function of the momenta, $\mathbf{p}_i \equiv m_i\mathbf{v}_i = \frac{\partial H}{\partial \dot{\mathbf{r}}_i}$ (with $\mathbf{v} = \dot{\mathbf{r}}_i = \frac{\partial H}{\partial \dot{\mathbf{p}}_i}$), and the masses, m_i, and is summed over the N atoms in the system to give,[8-10]

$$K(\mathbf{p};m) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} = \sum_{i=1}^{N} \frac{1}{2} m_i \mathbf{v}_i^2 \,. \qquad (1.1)$$

Different atom types describe the masses based upon the hybridization; the atom type concept will be illustrated later. The V term depends upon the atomic coordinates, $\mathbf{r}_i$, and the FF parameters, s. The s will be the main focus of this discussion.

Force Fields

MD simulations evaluate a system's configurations as a function of time based upon the position and momentum of each particle. Thus the simulation maps, point-by-point, the portions of the potential energy surface (PES) that have been accessed. The input for these calculations is a potential energy function. Together the extent of sampling achieved in a simulation and the quality of the FF that is used serve as the primary determinants for whether a simulation's results will be trustworthy or meaningless.

The PE function is assumed to be separable into a sum of terms for a given molecule,

$$V(\mathbf{r};\, s) = V_{\text{bond}} + V_{\text{angle}} + V_{\text{torsion}} + V_{\text{improper}} + V_{\text{vdW}} + V_{\text{Coulomb}}, \qquad (1.2)$$

which is taken to be pair-wise additive

$$V(\mathbf{r};s) = \sum_i \sum_j V_{ij}(\mathbf{r}_{ij};s)\,,\ i < j \qquad (1.3)$$

for a collection of $i + j$ molecules, where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$. As noted earlier, taking the negative gradient of the potentials provides the forces, F, on the particles and is used to generate the phase-space trajectory. Precisely because $F = -\nabla V(\mathbf{r};s)$, it is necessary that the potential energy function be continuously differentiable. Furthermore, to minimize computational expense, preference is given to the simplest algebraic or trigonometric form that captures the salient features of the potential. Within these confines, different research groups have taken liberties to choose slightly different functional forms to model the same interactions and, with far greater variance, different parameters for those functions. We are now among them. This work will describe our attempts to create an advanced united-atom, non-polarizable, explicit-solvent FF in which carefully developed parameters were created to improve the description of non-bonded interactions for simulations of peptides and small proteins.

The Biological FF Menagerie: Approaches and Parameter Sources

With so many FFs already available for the simulation of biological systems, it can sometimes be confusing to remember their differences and similarities, and difficult to fathom why another one should be created. Among the most popular are AMBER,[1,2] CHARMM[3-5], ECEPP-05[6,7], GROMOS[8-10], and OPLS[11,12]. In Table 1.1 the underlying approaches used to create these FFs are provided to distinguish the key similarities and differences among them. This list is neither intended to be exhaustive in the number of

Table 1.1: Comparison of Select Force Fields Used in Protein Simulations.

FF	Developer	Versions: 1st/Latest Protein FF	Water Model	Bonds	Angles	Torsions	Improper Dihedrals	vdW	Coulomb	Current Work
AMBER[1,2]	Late **P. Kollman (UCSF)**, **D. Case (Rutgers)**, **T. Cheatham (U of UT)**, **T. Darden (NIEHS)**, **K. Merz (U of FL)**, **C. Simmerling (Stony Brook)**, **R. Luo (UC Irvine)**, **J. Wang (Encysive Pharm.)**	1986/1995	TIP3P	Normal mode freq. calcs. or fitting to structural data/vib. freq.	Normal mode freq. calcs. or fitting to structural data/vib. freq.	QM rotational energy profiles	*ab initio* to obtain the PES (gas phase)	LJ. Density, ΔH_{vap} of pure liquids (some adopted from OPLS)	Restrained-ESP (places a hyperbolic restraint on bonds to heavy atoms) (6-31G*)	Improve QM/MM approach and polarizable models
CHARMM[3-5]	**Martin Karplus (Harvard)**	1983(Pre-CHARMM)/1993 (CHARMM22)	Modified TIP3P	Microwave, and electron diffraction or x-ray crystal structure + *ab initio*. UB term only used in select cases.	Vib. freq. or *ab initio* calculation to obtain the PES (gas phase)	QM rotational energy profiles	Vib. freq. or *ab initio* to obtain the PES (gas phase)	LJ. *ab initio* interaction calculations on rigid monomers	**Scale gas phase charges from ab initio calculations of the minimum interaction energies and geometries between dimers by factor of 1.16**	Polarizable model
ECEPP[6,7]	**Harold A. Scheraga (Cornell)**	1975/2006 (ECEPP-05)	NA	Fixed. X-ray and neutron diffraction.	Fixed. X-ray and neutron diffraction.	QM rotational energy profiles	NA	Buckingham. Crystal structures and QM calculations of dimers.	Restrained-ESP (places a hyperbolic restraint on bonds to heavy atoms) (6-31G*)	Create a hydration model to treat aqueous solutions of proteins
GROMOS[8-10]	**Wilfred van Gunsteren (ETH)**	1987/2004	SPC	Anharmonic for comput. efficiency. X-ray diffraction.	Cosine based for comput. efficiency. X-ray diffraction.	QM rotational energy profiles	Spectroscopic and X-ray diffraction data	LJ. Atomic polarizabilities	Thermo. properties of pure liquids and free enthalpies of solvation in cyclohexane and in water.	Polarizable model
KBFF	**Paul E. Smith (KSU)**	To be released/NA	SPC/E	From GROMOS	From GROMOS	QM rotational energy profiles	From GROMOS	LJ. Density, scaling rules for pure polar liquids, ΔH_{vap} for pure non-polar liquids	**KBIs**	Finalize and test version 1
OPLS[11,12]	**William Jorgensen (Yale)**	1988/1996	TIP3P, TIP4P	From AMBER	From AMBER	QM rotational energy profiles	NA	LJ. MC SM simulations to produce thermo./structural properties for pure liquids	MC SM simulations to produce thermo./structural properties for pure liquids	Polarizable model and continuum solvent models

FFs or the description of those that are mentioned. Additionally, we only focus on non-polarizable versions of these FFs, although some of these FFs also support polarizable versions. Figures 1.1 – 1.2 provide examples of these potentials with functional forms and parameters taken from the force field we are currently developing, which we refer to as the Kirkwood-Buff Force Field (KBFF). The bond, angle, vdW, and coulomb parameters are those of benzene (Figure 1.1) while the torsional and improper dihedral parameters are from HisH (Figure 1.2). Next, we will compare the functional forms.

General Functional Forms Used in Protein Force Fields:

AMBER:

$$\sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} \frac{k_\varphi}{2}[1+\cos(n\varphi-\delta)]$$
$$+ \sum_{nonbonded} 4\varepsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad (1.4)$$

CHARMM:

$$\sum_{bonds} k_b(b-b_0)^2 + \sum_{UB} k_{UB}(S-S_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2$$
$$+ \sum_{dihedrals} k_\varphi[1+\cos(n\varphi-\delta)] + \sum_{impropers} k_\xi(\xi-\xi_0)^2$$
$$+ \sum_{nonbonded} 4\varepsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \sum_{residues} V_{CMAP}(\varphi,\psi) \qquad (1.5)$$

ECEPP-05:

$$\sum_{torsions} k_\theta^1[1+\cos(\theta)] + k_\theta^2[1-\cos(2\theta)] + k_\theta^3[1+\cos(3\theta)]$$

$$+ \sum_{nonbonded} [-A_{ij} r_{ij}^{-6} + (1/k_{14})]B_{ij} \exp(-C_{ij} r_{ij})] + \frac{332 q_i q_j}{k_{14}^{el} \varepsilon_0 r_{ij}} \qquad (1.6)$$

GROMOS/KBFF:

$$\sum_{bonds} \frac{k_b}{4}(b^2 - b_0^2)^2 + \sum_{angles} \frac{k_\theta}{2}(\cos\theta - \cos\theta_0)^2$$

$$+ \sum_{dihedrals} k_\varphi[1+\cos(\delta)\cos(n\varphi)] + \sum_{impropers} \frac{k_\xi}{2}(\xi - \xi_0)^2$$

$$+ \sum_{nonbonded} 4\varepsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad (1.7)$$

OPLS:

$$\sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2$$

$$+ \sum_{torsions} \frac{V_1}{2}[1+\cos(\varphi + f_1)] + \frac{V_2}{2}[1+\cos(2\varphi + f_2)] + \frac{V_3}{2}[1+\cos(3\varphi + f_3)]$$

$$+ \sum_{nonbonded} 4\varepsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

(1.8)

To aid in the reader's understanding of the functional forms, the following points are clarified:

- **Bonds and angles:** Bond lengths are given by $b = r_{ij} = (\mathbf{r}_{ij} \cdot \mathbf{r}_{ij})^{1/2}$. Unless otherwise noted k is the force constant, b the bond length, θ the bond angle between atoms with the sequence *i-j-k*, n the multiplicity, φ the torsional angle, δ the phase shift, and S the Urey-Bradley 1-3 distance. All terms with a corresponding subscript of zero are the equilibrium values. Note that ECEPP-05 uses both fixed bond lengths and bond angles, so that the intramolecular energy is a function only of torsional angles.[3]
- **Torsional dihedral angles:** δ is the phase shift (restricted to 0 or π *i.e.*, $\cos(\delta) = \pm 1.0$), m the multiplicity of the torsional dihedral angle, and φ the value of the dihedral angle between atoms with the sequence *i-j-k-l*. In the OPLS torsions, V_1, V_2, and V_3 are the coefficients in the Fourier series, and f_1, f_2, and f_3 are phase angles. All fs are taken to be zero.
- **Improper dihedral angles:** ξ is the actual value of the dihedral angle between atoms with the sequence *i-j-k-l*, and can be calculated as $\xi = \text{sign}(\xi)\arccos(\mathbf{r}_{mj} \cdot \mathbf{r}_{qk}/r_{mj}r_{qk})$, $0 \leq \arccos \leq \pi$, where $\mathbf{r}_{mj} \equiv \mathbf{r}_{ij} \times \mathbf{r}_{kj}$, $\mathbf{r}_{qk} \equiv \mathbf{r}_{kj} \times \mathbf{r}_{kl}$, $\text{sign}(\xi) = \text{sign}(\mathbf{r}_{ij} \cdot \mathbf{r}_{qk})$, and ξ is undefined if $\mathbf{r}_{mj} = 0$ or $\mathbf{r}_{qk} = 0$.
- **van der Waals interactions:** In the ECEPP-05 FF the Buckingham potential is used instead of the Lennard-Jones potential.[3] In this case A_{ij}, B_{ij}, and C_{ij} are nonbonded parameters. For the remaining FFs, the LJ potential is used in which ε is the magnitude of the most favorable LJ energy of interaction, σ is the distance at which the LJ energy of interaction is zero. Combination rules:

AMBER and CHARMM use the Lorentz-Berthelot combination rules (ε_{ij} is the geometric mean of ε_i and ε_j, $\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$, and σ_{ij} is the arithmetic mean between σ_i and σ_j, $\sigma_{ij} =(\sigma_i + \sigma_j)/2)$ whereas OPLS, GROMOS, and KBFF use geometrical means for both σ_{ij} and ε_{ij}. ECEPP-05 uses a geometric mean for A_{ij} and B_{ij} and an arithmetic mean for C_{ij}.[3] The ECEPP-05 also scales the interactions separated by three bonds by the k_{14} and k_{14}^{el} parameters.[3] Interactions separated by more than three bonds have these parameters set equal to one.

- **Coulomb interactions:** ε_0 is the electric constant and $1/(4\pi\varepsilon_0)$ is a conversion factor from Coulomb^2/nm to energy (138.9354 kJmol^{-1}e^{-2}nm).[7]

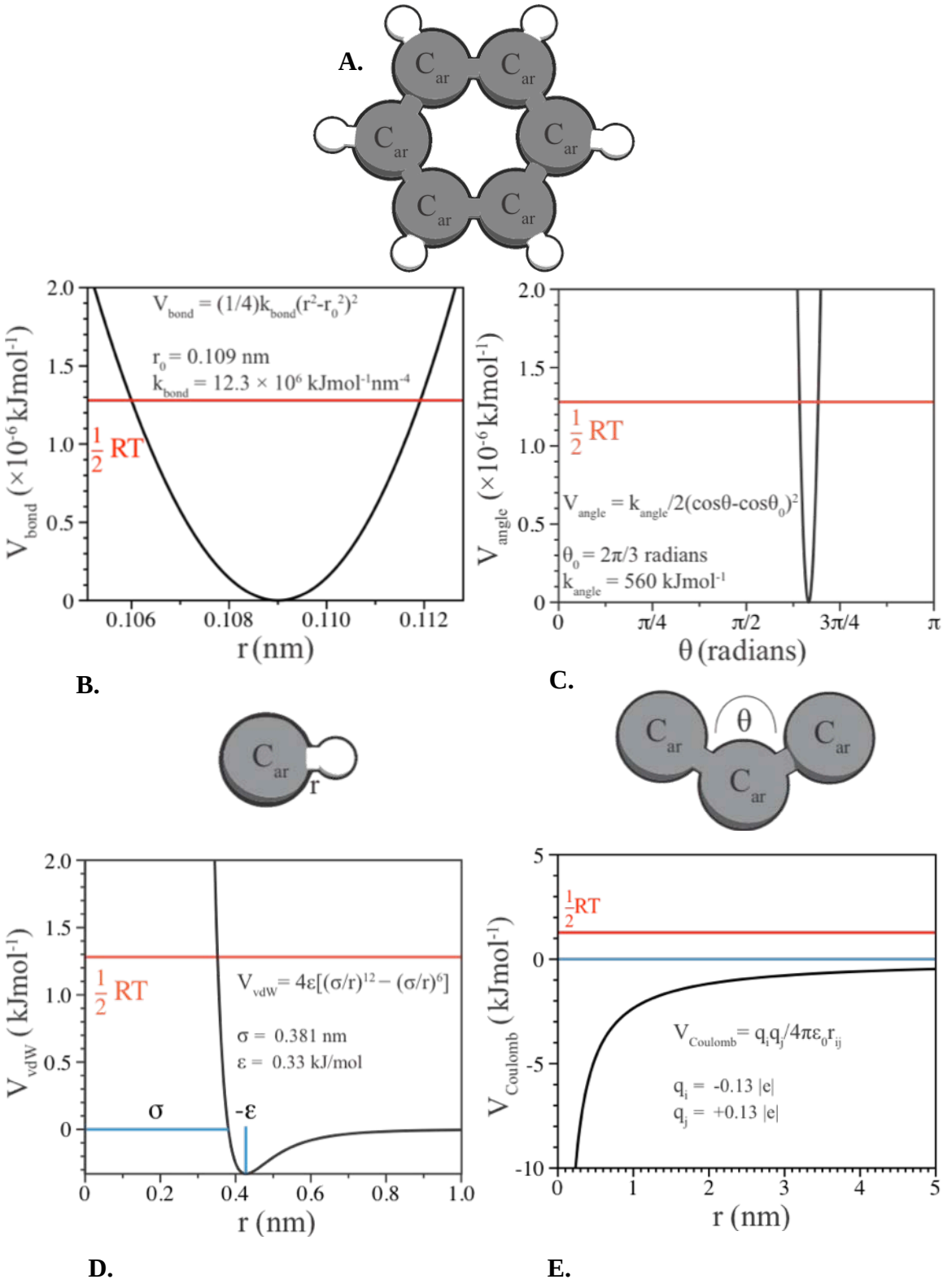


Figure 1.1: A. Benzene Model with C_{ar} atom types labeled. **B.** V_{bond}, **C.** V_{angle}, **D.** V_{vdw}, **E.** $V_{Coulomb}$. The thermal energy accessible to one degree of freedom, ½ RT, is shown to provide perspective.

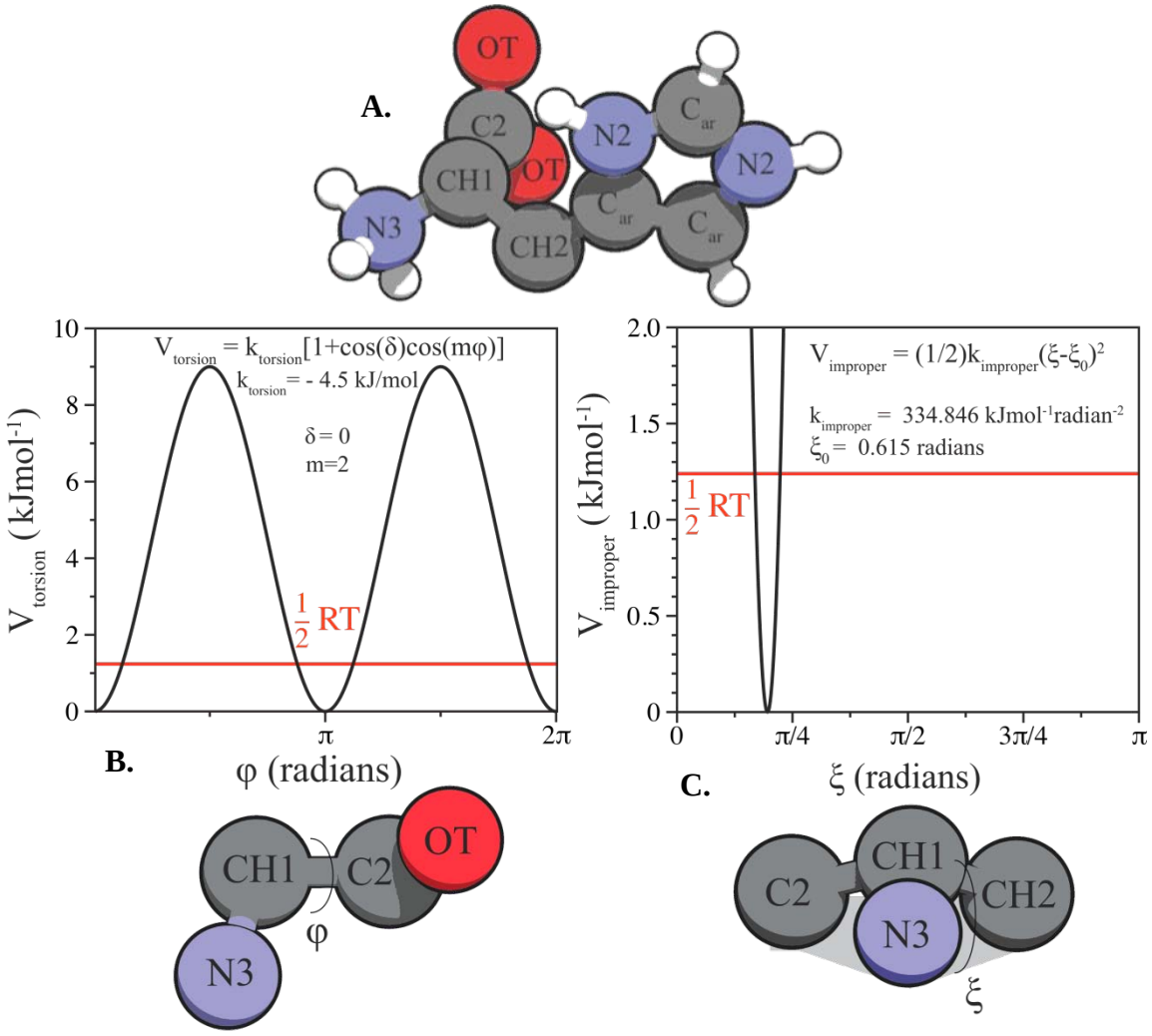


Figure 1.2: A. HistidineH Model with atom types labeled. **B.** $V_{torsion}$ **C.** $V_{improper}$. The thermal energy accessible to one degree of freedom, ½ RT, is shown to provide perspective.

The Need for an Improved Force Field

Several comparisons of FFs have been made with an emerging consensus that available FFs generally tend to over stabilize secondary structural elements, particularly helices.[10] Additionally, reduced solvation (overestimation of solute-solute and solvent-solvent interactions) has been noticed frequently.[11-15] Sometimes this leads to obviously erroneous results, such as spontaneous formation of aggregates[16] or phase separation in

systems which should be fully miscible,[17] but other times the errors do not make themselves apparent from a visual inspection. Blame is often placed on the description of electrostatic interactions provided by the force field parameters.[16,18] Why is this so?

On the Effects of the Gas to Liquid Phase Transition

Since our model is classical, we do not have a description of the wave character of electrons. We deal only with the particulate atom, a simplification that allows the electron distribution to be modeled by a partial charge located on each sphere representing the atoms in a molecule.

As noted in Table 1.1, most non-polarizable FFs choose the partial atomic charges for atoms such that they reproduce the gas phase quantum mechanical (QM) electron density of a molecule. This approach takes into account the permanent multipole moment of the molecule, but egregiously neglects the effects of bulk solvation on the polarization of a molecule, which determines its induced multipole moment. Recalling that the multipole moment measures the difference in the distribution of positive and negative charge within a molecule *i.e.*, its molecular polarity, consider briefly the effects a transition from the gas to the liquid phase will have on this value for a molecule.

The strength of the always-favorable intermolecular interactions (*i.e.*, dipole-induced dipole/London dispersion, dipole-dipole, and hydrogen bonding) is increased in the condensed phase. For a water molecule, this makes interactions with neighboring water molecules more favorable due to an increased separation in the positive and negative partial charges on the oxygen and two hydrogens. The degree to which this occurs will differ for different molecules. For example, the dipole moment of water is

6.17×10^{-30} Cm (1.85 D) in the gas phase[19] but between 8.34×10^{-30} Cm and 1.17×10^{-29} Cm (between 2.5 and 3.5 D) in the liquid phase,[19] whereas the quadrupole moment of benzene is -2.90×10^{-39} Cm^2 in the gas phase[20] but -5.68×10^{-29} Cm^2 in the liquid phase[21]. A negative quadrupole moment indicates that it is an oblate spheroid (the anti-parallel dipoles are oriented along the minor axis) as opposed to a prolate spheroid (the anti-parallel dipoles are oriented along the major axis). Because the models created in classical MD simulations only consist of point charges located on spheres, they do not take into account the outer plane quadrupole of benzene arising from the π-electrons.

Polarizable FFs have been developed which, in principle, should be better than non-polarizable FFs; however, they are still very computationally expensive and are not without their own set of implementation difficulties. Thus, non-polarizable FFs are still very popular. To address the differences between gas and solution phase charge distributions, non-polarizable FF developers generally either scale the gas phase charges, or use a basis set that is known *a priori* to create erroneously large charge distributions for a gas phase molecule *e.g.*, 6-31G*.[22] These modified gas phase charges are then adopted for the liquid phase charge distribution.

Is this a good approach?

When molecule *i* passes from the gas phase into solution, physical changes occur which contribute to the free energy of solvation. ΔG_{solv} is always negative if the solute dissolves in the solvent. The enthalpic contribution to ΔG_{solv} is often thought about in the reverse direction *i.e.*, (*l*) to (*g*), and is due to a reduction in favorable solvent-solvent interactions during this transition. The more cohesive a solvent is (quantified by a large δ parameter, see Table 1.2), the more unfavorable the evaporation process is. These costs

Table 1.2: (Experimental) ΔG_{solv}: The Gibbs energy of solvation has an unfavorable ΔH_{vap} contribution and a favorable enthalpy/entropy solute-solvent interaction contribution. **(Experimental) ΔH_{vap}:** Heat/Enthalpy of Vaporization is the amount of energy required to vaporize solvent/gram or mole of solvent. Those listed are the ΔH_{vap} at the boiling point (B.P.) and 1.0 atm. The ΔH_{vap} measures the energy required to overcome attractive forces. **(Experimental) δ Solubility Parameter:** Square of the energy of cohesion per unit density of the solvent $(\Delta H_{vap}/V)^2$; measures difficulty to make a bubble in a solvent. **(Experimental) Molecular Polarizability, α:** Magnitude of the dipole induced by one unit field of gradient (which is one unit of volume); exhibits a positive correlation with the volume occupied by electrons.

Molecule	ΔG_{solv}[23] (kcal/mol)	ΔH_{vap}[24,25] (kcal/g)	B.P.[26] (K)	δ[24,25]	α[26] ($cm^3/10^{-24}$)
Water	N/A	0.540	373	23.4	1.45
Methanol	-5.1	0.263	338	14.3	3.23
Benzene	-0.9	0.94	353	9.2	10.32

are compensated due to new solute-solvent interactions that are created in the condensed phase (the enthalpy of mixing, ΔH_{mix}) and the presence of a now less ordered solution (entropy of mixing, ΔS_{mix}). Refer to Table 1.2 to compare the ΔG_{solv} and ΔH_{vap} values for water, methanol, and benzene.

The above terms will vary greatly depending upon the constituents of the mixture. We now consider the intermolecular interactions present in the condensed phase in an attempt to differentiate between aspects that traditional FFs have or have not captured due to the approaches used in their determinations of the charge distributions.

Firstly, we must consider the electronegativity differences between atoms that define their bond dipole moments. Collectively, these bond dipole moments create the group polarities that define the permanent multipole moment of the molecule. This phenomenon results primarily from the withdrawing of electrons through sigma bonds to the more electronegative atom or group and is often termed the *inductive effect*. This is

taken into account in FF parameterization that is based on QM calculations in the gas phase, since it is a property of a molecule irrespective of its environment.

Secondly, we must consider *field effects*, which polarize a molecule resulting from charges/partial charges that interact through space (in contrast to through sigma bonds). This includes dipole-induced-dipole/London dispersion, ion-induced-dipole, and ion-ion interactions. In the gas to liquid phase transition, electrostatic interactions will become more favorable because (1) the multipole moments of the solvent will align themselves to oppose the multipole moment of *i*. This will cause (2) *i*'s polarity to increase (*i.e.*, an increased multipole moment) and (3) the solvent to "respond" by polarizing in kind. Refer to Table 1.3 to compare select solvent scales for water, methanol, and benzene. Because field effects can influence the polarization of various parts of a molecule to different extents, we believe solution charge distributions should not be determined from a simple scaling of the QM calculated electrostatic potential (ESP), because that approach has only accounted for the permanent multipole moments.

Thirdly, since we are designing a FF for the simulation of peptides and small proteins, we must take into account the different environments an amino acid can experience in a protein. One common example is the variability in the strength of salt bridges. The solvent exposed Asp•••Lys salt bridge has been attributed a strength of 0-8 kJ/mol while that of a buried Asp•••Lys salt bridge has been shown to be worth up to 13 kJ/mol.[27] Thus we believe that it is important to study systems over a range of compositions, and find the optimal charge distribution for a molecule when in a variety of environments.

Table 1.3: Select Solvent Scales[27]: Dielectric Constant: Bulk property that measures fluctuations in the dipole moment (large molecule dipoles, large molecular polarizabilities, and H-bonding all contribute to a large dielectric constant). Water has a large dielectric, second only to formamide. **π^* scale:** A measure of the extent to which the solvent can stabilize ionic or polar species; a measure of non-specific electrostatic solvation. Water has the highest value. **α scale:** A measure of the solvent's ability to act as a H-bond donor to a solute; water is the best and therefore has the highest value. **β scale:** A measure of a solvent's ability to act as a H-bond acceptor from a solute. Many solvents are better than water at being a H-bond acceptor (e.g., DMSO).

Solvent	ε	π^*	α	β
Water	78	1.1	1.17	0.47
Methanol	33	0.60	0.93	0.66
Benzene	2	0.59	0.00	0.10

Table 1.4: Pauling Electronegativity Scale: Measured by examining bond dissociation energies of molecules, therefore it depends upon molecular properties and is not an intrinsic property of the atom. **Mulliken Electronegativity Scale:** Solely an atomic property because it is derived from the average of the ionization potential and the electron affinity.

Atom	Pauling[28]	Mulliken[29]
H	2.1	3.01
C	2.5	2.67
O	3.5	3.22

The C-H bond is not typically considered polar; however, as noted in Table 1.4, the electronegativity scales of Pauling and Mulliken actually swap the ordering of the electronegativities of carbon and hydrogen. This seeming disagreement is assuaged if we recall that the more *s*-character a hybrid orbital has, the more electron density will be located at the nucleus. Thus, hybridization explains the experimentally observed trend of electronegativity *i.e.*, $\equiv CH > =CH_2 > -CH_3$ or equivalently $sp > sp^2 > sp^3$. This is important when considering the prototypical aromatic, benzene, which owes its large quadrupole moment to the symmetric addition of six $C^{\delta-}-H^{\delta+}$ bond dipoles.

The aromatic amino acids additionally exhibit *π-effects* (cation-π, polar-π, and π donor-acceptor), a topic of considerable attention due to its role in molecular recognition

(see *e.g.*, J. C. Ma and D. A. Dougherty's 1997 *Chemical Review* article, "The Cation-π Interaction" pp. 1303 or E. A. Meyer, R. K. Castellano *et al*'s 2003 *Angewandte Chemie* article, "Interactions with Aromatic Rings in Chemical and Biological Recognition" pp. 1210). When considering benzene, the ESP is negative on the surface of the ring and positive along the edges, which allows for a favorable interaction between cations or δ^+ charges on the face of the ring.[27] As already mentioned, these effects are not explicitly taken into account in any CM FF due to a lack of orbital descriptions.

In summary, since the solvation process distorts the solute's electronic structure, with different parts of a molecule becoming polarized to different extents, we believe it is not appropriate to simply scale gas phase charges, as is done in most FFs. Instead we propose that it is better to scale and redistribute the gas phase charges empirically to fit experimental data by studying mixtures of model compounds. We use this redistribution approach to find a charge distribution that reproduces the Kirkwood-Buff integrals (KBIs).

Biochemical simulators do not need another FF unless it has been developed with a markedly different approach, with a principle focus on balancing solute-solute, solute-solvent, and solvent-solvent interactions by studying these interactions far beyond the infinitely dilute range. This is the role the Kirkwood-Buff Force Field (KBFF) seeks to fill. The KBFF for amino acids and their analogs has adopted the GROMOS V_{bond} and V_{angle} parameters. The new parameters of V_{torsion} for all rotatable bonds were developed to reproduce the gas phase rotation energy profiles,[30] V_{vdW} to reproduce the density and enthalpy of vaporization of pure liquids (excluding the CH, CH2, and CH3 united atoms

parameters which were adopted from GROMOS) and, of principle interest, V_{Coulomb} to reproduce the KBIs of binary mixtures.

Our Guide, the Kirkwood-Buff (KB) Theory of Solutions

Justifying Our Dependence on KB Theory

Essentially all biochemical processes, and likewise many of industrial interest, take place in solution mixtures. Attempts to understand solutions have led researchers to accept and later disband many theories and models over the last two centuries.[31] Despite much advancement, the community is still far from satisfied with their present understanding of solution mixtures. One way to measure our progress in this area is by evaluating how well we can make models of solutions that produce experimental properties. Additionally, accurate modeling allows for the prediction of behavior in situations where the experimental work is prohibitive on some grounds, be it too time intensive, costly, or difficult. Our goal is to create a model that would require only input of the composition and state of the system (a set of *NVE*, *NVT*, *NpT*, or μVT) and which could then provide output of thermodynamic properties. The Smith group is specifically interested in applying these models to obtain thermodynamic and structural information about peptides and small proteins in various physiological environments.

To understand solutions on the microscopic level, we must turn to statistical mechanics (SM), which is a crucial tool because it serves to bridge the microscopic and macroscopic properties of a system. The goal in SM is to determine the partition function because, much like the Schrödinger equation in QM, it completely defines the state of the system.[32] From the partition function we can derive any macroscopic property of

interest.[32] SM requires information of the microscopic energy levels, however to solve for these we need to know the PE between molecules.[31] This could be obtained from QM, however we are interested in large systems where the Schrödinger equation is only solvable if the PE is assumed to be zero everywhere, which is of no use for our purposes.[31]

How then may we proceed? Since the spacing between translational energy levels is very small, we can think of them as continuous and not discrete.[33] This is not true for rotation and vibration, in general, and certainly not for nuclear or electronic energy levels. Thus, we can separate internal from translational motions. We will represent the center of mass translational motion with a configurational partition function, Z_N, given by[33]

$$Z_N = \int_{Volumes} e^{-\beta U}\, dr_1 \cdots dr_N , \tag{1.9}$$

which takes into account the interaction between molecules (here we denote the total potential energy as U instead of V to distinguish it from the volume, V). We will represent the internal partition function, which includes the contribution from the internal motions of the molecules, with the internal partition function of an ideal gas under the same conditions, which is a solvable case.[33] This approximation is acceptable unless we are interested in very light molecules *e.g.*, helium or hydrogen gases. We are left with a semi-classical partition function (Q_{CM}), but it still contains a pervasive Boltzmann factor with its unknown PE function,[33]

$$Q_{CM} = \frac{Z_N}{N!\Lambda^{3N}}, \tag{1.10}$$

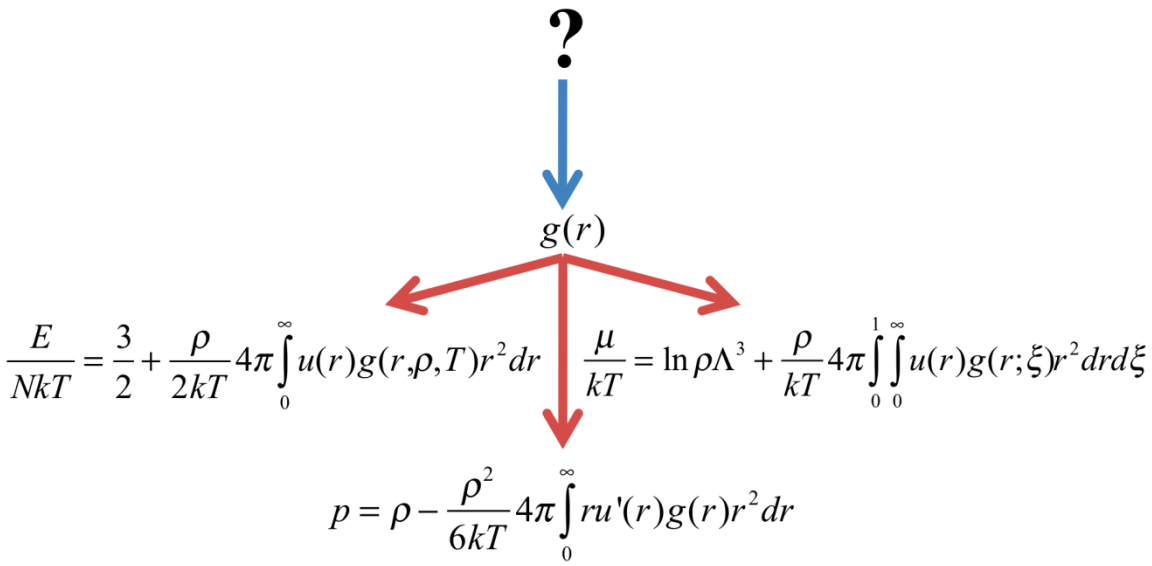


Figure 1.3: Relationship between the rdf, if obtainable, and the total energy, pressure, and chemical potential for a pure liquid in the *NVT* ensemble. From *E*, *p*, and μ other thermodynamic properties may be obtained. In the calculation of the chemical potential, ξ is a coupling parameter, which varies from zero to one and effectively controls whether or not a central molecule is included in the system. In the pressure equation u' is taken to mean u after a change of variables due to differentiating Z_N with respect to V while assuming that the volume is large the container is cubic. The reader is referred to D. A. McQuarrie's *Statistical Mechanics* Chapter 13: Distribution Functions in Classical Monatomic Liquids for the full derivation.[33]

where the ideal contribution is $1/N!\Lambda^{3N}$ and Λ is the thermal de Broglie wavelength, $\Lambda = (h^2/2\pi mkT)^{1/2}$. Thus, we have reached another impasse because we cannot evaluate Z_N without making crude approximations.

In 1951 Kirkwood and Buff published their theory of solutions in the *Journal of Chemical Physics*.[34] Instead of relying solely on molecular information to obtain thermodynamic data, KB theory allows for molecular distribution functions to act in concert with molecular information to obtain thermodynamic properties.[35] The theory was primarily untouched for almost twenty-five years however, because molecular distribution functions themselves can only be obtained experimentally from x-ray or

neutron diffraction (which is only possible for simple molecules), solved using integral equations, or solved numerically in a computer simulation.[35] Returning to SM we see that the difficulty in obtaining the pair correlation function originates, again, from the presence of the PE function in the Boltzmann factor,[33]

$$g^{(2)}(r_1,r_2)=\frac{V^2N!}{N^2(N-2)!}\frac{\int\cdots\int e^{-\beta U_N}\,dr_3\,dr_4\cdots dr_N}{\mathrm{Z}_N}=\underset{radial}{g(r)}. \qquad (1.11)$$

If we did know the PE function, and we assumed pairwise additivity in which[33]

$$U_N(r_1,\cdots,r_N)=\sum_{i<j}u(r_{ij}) \qquad (1.12)$$

(where the summation is over all *i*-*j* pairs), we could obtain any thermodynamic property of interest from $g(r)$ as shown in Figure 1.3 for a one component system in the *NVT* (canonical) ensemble.[33]

While Equation 1.11 can be solved for a simple system such liquid argon, if the system of interest contains molecules, then we do not know the PE function (which is not only a function of distance but of orientation as well). Thus, since distribution functions were no easier to obtain than the very thermodynamic properties for which KB theory was formulated to solve, KB theory rested for years with relatively little use.

In 1977 Arieh Ben-Naim established the inversion of the KB theory, a route to obtain local properties, in the form of KBIs, from thermodynamic properties.[36] In both directions the formulation is exact and rigorously dependent upon the same statistical mechanical tools.[35] By using this inversion, we circumvent our reliance on the Boltzmann factor because we are going from thermodynamic to microscopic properties.

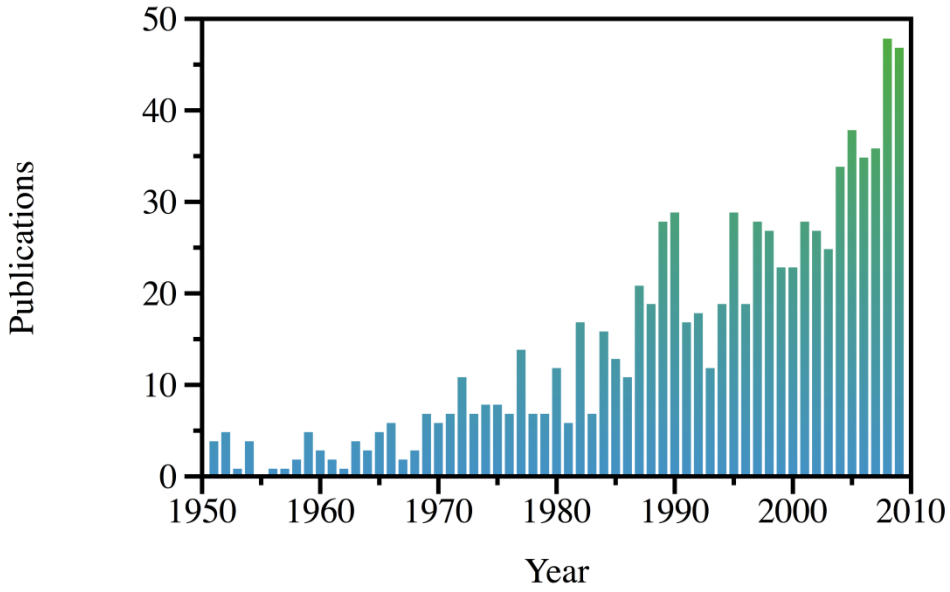


Figure 1.4: Web of Science search results for "Kirkwood-Buff AND fluctuation theory of solutions" as of March 2010 with results for 2010 excluded.

KB theory has many wonderful features, as is evident by the rapidly growing number of publications using the theory since Ben-Naim developed the inversion procedure (Figure 1.4). In addition to being exact, KB theory is general in the sense that it can be used to describe mixtures with any number of components of any complexity, as long as the solution is stable wrt composition.[37] Additionally, unlike many other theories, it does not rely on an assumption of pairwise additivity when going in the direction of [Thermodynamics Properties] $\Rightarrow$ [KBIs], but instead takes into account the true many-body interactions in solutions.[38]

Derivation

The formal derivation of the KB theory can be found elsewhere, beginning in the μVT (grand canonical) ensemble and then employing thermodynamic transformations to

obtain the ensemble of most frequent experimental relevance, *NpT* (isothermal-isobaric).[34,36,39] Here we will focus on the actual process by which we extract the KBIs from experimental data and from the results of an MD simulation in order to compare experiment to simulation and parameterize molecular charge distributions so that the two are in agreement.

All expressions provided in this discussion only apply to binary systems. The expressions are not the same for mixtures with more components, and although the theory can be applied to mixtures with any number of components, the terms become increasingly complex and, more importantly, it becomes increasingly difficult to find the necessary experimental data for systems with more than two components.[34,36,39] The brave reader is encouraged to pore over a recent JCP article in which M. Kang and P. E. Smith formulate KB theory for systems with up to five components.[40]

The KB inversion procedure requires experimental knowledge of the chemical potential; however, this information can be expressed in the literature in many ways, in the form of activity coefficients (γ) through $\mu_i^E = RT \ln \gamma_i$ or extracted from the excess molar Gibbs energy, g^E, since $g^E / RT = x_i \ln \gamma_i + x_j \ln \gamma_j$.

We need an analytical expression for g^E because we must take its first and second derivatives. A commonly used expression for fitting any excess function is the Redlich – Kister power series equation,[41]

$$X^E / \text{kJmol}^{-1} = x_1 x_2 \sum_n A_n (x_1 - x_2)^{n-1}, \qquad (1.13)$$

where X^E is any excess function and A_n is a parameter. When g^E or activity coefficient data is explicitly provided, such as through the Redlich-Kister equation, the analysis is

relatively straightforward; however vapor-liquid equilibria (VLE) and liquid-liquid equilibria (LLE) data are often presented in terms of models of g^E which have been designed primarily for chemical engineering purposes. Several of these models include the Wilson, Non-Random Two Liquid (NRTL), Margules, modified Margules, van Laar, UNIversal Functional Activity (UNIFAC), and UNIversal QUAsi-Chemical (UNIQUAC) equations. Each model has different strengths and areas of applicability. The activity data used in Chapters 2 and 3 relies on two of the g^E models, Wilson's equation and Renon's NRTL equation, along with the use of the Redlich-Kister power series equation.

The Wilson equation provides g^E for a system with any number of components.[42] It is particularly useful for representing g^E for a system of polar or associating solutions in nonpolar solvents, however it should only be used for completely miscible systems or in a range of complete miscibility[41] and it is only applicable to vapor-liquid (not liquid-liquid) systems.[43] Renon's Non-Random Two Liquid (NRTL) equation, in contrast, is applicable to liquid-liquid systems and partially miscible systems.[43] Therefore, the NRTL model is especially useful when a system is strongly non-ideal.[41]

The key equation which combines experimental knowledge of the isothermal compressibilities, partial molar volumes, and derivatives of the chemical potential to provide the three KBIs for a binary system from thermodynamic data is,[35]

$$G_{ij} = kT\kappa_T - \frac{\delta_{ij}}{\rho_i} + \rho kT \frac{(1-\rho_i \bar{V}_i)(1-\rho_j \bar{V}_j)}{\rho_i \rho_j \mu_{ij}}, \qquad (1.14)$$

where δ_{ij} is the Kroenecker delta function, $\rho = \rho_i + \rho_j$, $\mu_{ij} = (\partial\mu_i/\partial N_j)_{p,T,N_i}$, and all other symbols have their usual meaning.

Additionally, to extract the KBIs from a simulation, all that is needed is a trajectory (how the coordinates of all the atoms in the system change with time) and the mass of each atom. This is trivial in a simulation. From this information the center of mass (COM) radial distribution function (rdf) can be calculated, which is a standard procedure in MD suites. The COM rdf measures the probability of finding the center of mass of one molecule a specified distance *r* away from the center of mass of a central molecule. For a binary mixture there are three radial distribution functions, one for the probability of finding *i r* away from *i*, one for the probability of finding *j r* away from *i*, and the third for the probability of finding *j r* away from *j*. The rdf must be zero at the origin due to the volume occupied by the central molecule.

Peaks in the rdf correspond to solvation shells and dips correspond to depletions of molecules at that distance. Thus the rdf provides information on the residual structure of a solution. When the structure of the solution reaches that of a bulk solution, the rdf has a value of one, indicating that the probability of finding a certain type of molecule a distance *r* away from the central molecule is what would be obtained from a random distribution of molecules. As illustrated by the magnification of Figure 1.5 between 2.0 nm – 2.5 nm, rdfs typically exhibit long-range oscillatory behavior. When integrating the rdf to obtain the KBI (as discussed next), it is important that the rdf has reached unity so that the KBI will reach a plateau and an appropriate average value of the KBI can be obtained.

The expressions for the KBIs are then given by[44]

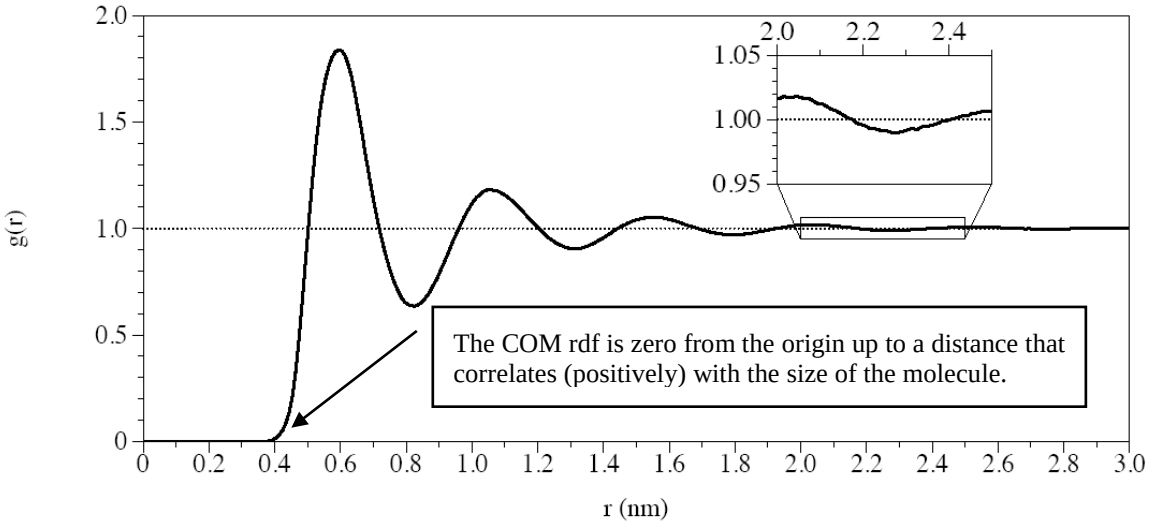


Figure 1.5. Illustrative radial distribution function.

$$G_{ij} = G_{ji} = \int_0^{\infty} [g_{ij}^{\mu VT}(r) - 1] r^2 \, dr \approx \int_0^{R_{cut}} [g_{ij}^{NpT}(r) - 1] r^2 dr \qquad (1.15)$$

in which R_{cut} is the distance at which the rdf has reached unity. The KBI measures the probability of finding a molecule *j* within a certain volume away from a central *i* molecule. Just as for the rdf, large positive numbers indicate association of *j* around *i*. For a binary mixture there are three KBIs, G_{ii}, $G_{ij} = G_{ji}$, and G_{jj}. In practice, a value for the KBI in Figure 1.6 would be obtained by averaging over a distance in which the rdf was approximately unity, such as between 2.0 and 2.5 nm. This value for the KBI (which has units of cm^3/mol) could then be compared to the KBI that was extracted from the experimental data at that composition.

As noted in the lhs integral in Equation 1.16, the experimental data at constant *T* and *p* correspond to the KB integrals in the μVT (grand canonical) ensemble. It is assumed that there is a close relationship between this distribution of molecules in an

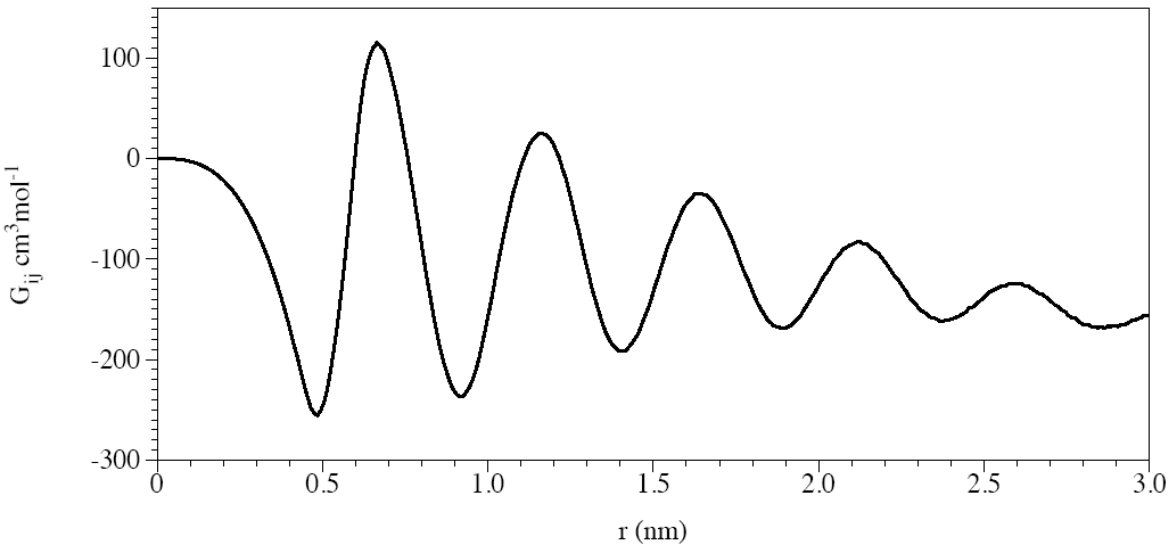


Figure 1.6. Illustrative Kirkwood-Buff integral. In practice, a value is obtained for comparison with the experimental KBI by averaging the integral over a distance in which the corresponding COM rdf was approximately unity.

open system and the distribution in a closed *NpT* ensemble at the same density and *T* since the rdfs are primarily determined by short-ranged intermolecular interactions that are on the order of the first few molecular diameters.[35]

The reader should note that at low concentration of species *j*, the simulated and experimental G_{jj} value is statistically unreliable due to a small number of *j* molecules. To reduce the effect of the noise in this number, the KBI can be multiplied by the number density, $\rho_j = N_j/V$, such that $\rho_j G_{ij} = N_{ij}$. N_{ij} is called the excess coordination number and it represents the number of *j* molecules a certain volume away from molecule *i* over that which would be present in the absence of the central *i*.[45] Again, large positive numbers indicate association of *i* around *j*. Although the KB integrals are symmetric with respect to a change in the indices *i* and *j i.e.*, $G_{ij} = G_{ji}$, this symmetry is lost when considering N_{ij} *i.e.*, $N_{ij} \neq N_{ji}$ since $\rho_j G_{ij} = N_{ij}$ and $\rho_i G_{ij} = N_{ji}$.

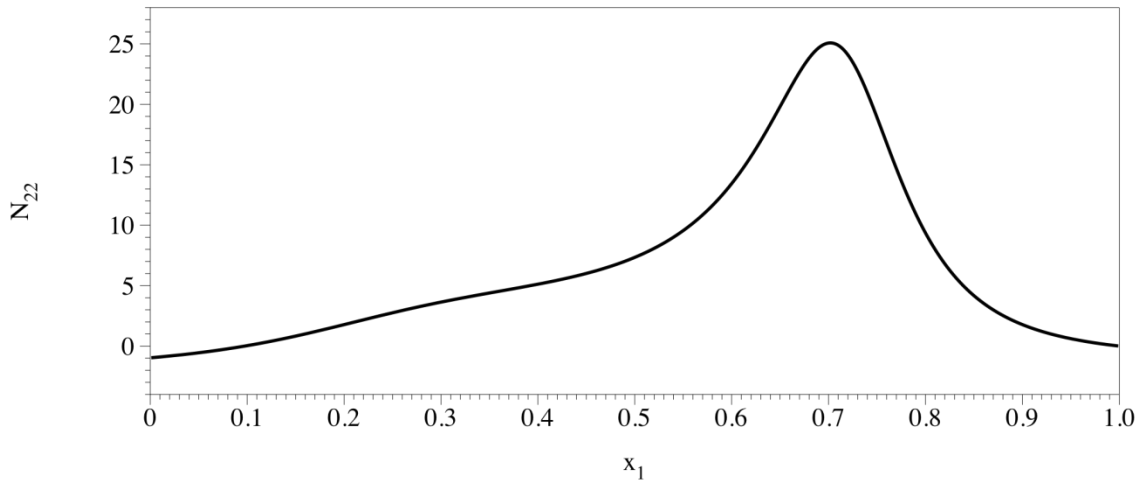


Figure 1.7. Illustrative excess coordination number. This system shows a clustering of molecule 2 around molecule 2 *i.e.*, self-aggregation of 2, at x_1 ~0.7.

Following the calculation of the KBIs for a binary system, the following thermodynamic properties may be extracted,[35]

$$\kappa_T = \frac{\zeta}{kT\eta},\ \bar{V}_i = \frac{1+\rho_j(G_{jj}-G_{ij})}{\eta},\ \bar{V}_j = \frac{1+\rho_i(G_{ii}-G_{ij})}{\eta}, \tag{1.16}$$

$$\mu_{ii} = \frac{\rho_j kT}{\rho_i V\eta},\ \mu_{jj} = \frac{\rho_i kT}{\rho_j V\eta},\ \text{and}\ \mu_{ij} = \mu_{ji} = \frac{-kT}{V\eta}. \tag{1.17}$$

In these equations k is the Boltzmann constant, $\rho_i = N_i/V$ (*i.e.*, the average number density of species i), $\mu_{ii} = \partial\mu_i/\partial \ln x_i$, and η and ζ are auxiliary equations with[35]

$$\eta = \rho_i + \rho_j + \rho_i\rho_j(G_{ii}+G_{jj}-2G_{ij}), \tag{1.18}$$

$$\zeta = 1+\rho_i G_{ii} + \rho_j G_{jj} + \rho_i\rho_j(G_{ii}G_{jj}-G_{ij}^2), \tag{1.19}$$

and $\Delta_{ij} = G_{ii} + G_{jj} - 2G_{ij}$. Δ_{ij} is equal to zero if the mixture is symmetric ideal.[35]

Because we are able to extract the thermodynamic properties from the KBIs of a microscopic, simulated system and directly compare them to the experimental,

thermodynamic properties, we have a two-way bridge between [Macroscopic/Bulk/Thermodynamic Properties] $\Leftrightarrow$ [Microscopic/Local Properties].[35]

Additionally, because the KBIs are very sensitive to the parameters describing the PE function in a simulation (illustrated in Chapter 2), there is a link of [FF parameters] $\Rightarrow$ [Thermodynamic properties of mixtures over a range of compositions], which we believe can be used as a stringent test of the accuracy of FF parameters.

On Modeling

In the Results sections of Chapters 2 and 3, we will claim to "reproduce thermodynamic properties of binary mixtures of small molecules." We need to be careful in our diction and in our thoughts. What have we really done? We have identified binary mixtures to study *e.g.*, toluene in methanol, and several observables that we believe to be relevant to understanding those mixtures *e.g.*, densities and KBIs. We then create a model of toluene (*i.e.*, its FF), simulate that model, and compare the output observables with their experimental values. Note then that we do not simulate toluene, but instead a model of toluene. The model mimics real toluene only to the extent that it incorporates whichever features determine the observables under investigation (be they the atomic partial charges and Lennard-Jones parameters, as we believe to be of principle import, or something more).

It naturally follows that it is possible to create a model that produces certain properties of toluene well (at a minimum, those upon which the FF was parameterized), but poorly reproduces other properties. Therefore simulators should carefully choose a FF, and should be wary of results obtained when using a FF that is not known to reliably

Table 1.5: Summary of Smith group's Progress Towards Completion of KBFF v.1. The ϕ, ψ, χ, and ω dihedral potential parameters were developed by F. Chen and P. E. Smith.[30]

Solute	Solvent	Analog For	Reference
Acetone	Water	Cosolvent	[16]
Urea	Water	Cosolvent	[15]
NaCl	Water	Cosolvent	[17]
Guanidinium chloride	Water	Cosolvent, Arg	[46]
Amides	Water	Asn, Gln, Peptide group	[19]
Methanol	Water	Ser, Thr, Tyr	[47]
Thiols, Sulfides	Methanol	Met, Cys, Disulfide	[48]
Amine salts, carboxylates	Water	Lys, Asp, Glu, Termini	[49]
Benzene, Toluene	Methanol, each other	Phe	Chapter 2
Phenol, *p*-Cresol	Toluene	Tyr	Chapter 2
Pyrrole, Indole	Methanol	Trp	Chapter 2
HistidineHCl	Water	His+	Chapter 3
Pyridine, Pyrrole	Methanol, Water	His	Chapter 3
Histidine	Water	His	Chapter 3

produce the feature they are attempting to probe. The KBFF has been shown to produce the KBIs without sacrificing agreement with properties that are well produced with currently available FFs.[15-17,19,45-48]

Aims of This Work

We have explained the need for an improved non-polarizable FF for the simulation of biological systems and how we hope to fill that need by using the KB theory of solutions as a guide. The Smith group's progress to date is summarized in Table 1.5. In this work the non-bonded parameters for the remaining amino acids, the aromatics, are presented. We developed the amino acid parameters by studying binary solutions of small molecules that are analogous to the side chains of these aromatic amino acids and, in the case of histidine, the amino acid itself. In Chapter two we will discuss the development of parameters for Phe, Tyr, and Trp. In Chapter three we will discuss the development of parameters for His and HisH.

Chapter 2:

The Kirkwood-Buff Derived Force Field for Phenylalanine, Tyrosine, and Tryptophan

Abstract

We are developing a force field for the description of peptides and proteins that is specifically designed to reproduce the experimental Kirkwood-Buff (KB) integrals (KBIs) for mixtures of small solutes, representing amino acid sidechains, in solution. This Kirkwood-Buff Force Field (KBFF) is intended to reproduce the thermodynamic properties of mixtures due to an improved description of intermolecular interactions. Here we focus on the development of the KBFF models for phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp). Aromatics representative of these sidechains (phenol, *p*-cresol, and indole, respectively) are only sparingly soluble in water. This limits the study of aqueous solutions to dilute concentrations, so we have switched to solvents of either methanol or toluene in order to study mixtures over their full range of compositions. We provide the KBFF parameters for Phe, Tyr, and Trp sidechains and the results of our simulations for the following binary mixtures: benzene + methanol, benzene + toluene, toluene + methanol, phenol + toluene, *p*-cresol + toluene, pyrrole + methanol, and indole + methanol, as validating evidence for the future use of the KBFF models in simulations of biological systems.

Introduction

There are multiple force fields (FF) available today for simulations of biological systems.[1-12] The Smith group is designing yet another alternative, but we do so by taking an unconventional approach for the parameterization of our molecules. The Kirkwood-Buff Force Field (KBFF) is based on the Kirkwood-Buff (KB) theory of solutions proposed in 1951,[34] and the theory's inversion procedure introduced by Ben-Naim in 1977.[36] KB theory is a rigorous, zero-approximation method that provides a bidirectional link between the global, thermodynamic properties of solutions and the microscopic properties *i.e.*, local structure.[35]

The impetus for development of the KBFF grew naturally from observations of previous simulations that often produced incorrect descriptions of intermolecular interactions. Examples include a simulation by A. C. Vaiana in which rapid, erroneous KCl aggregates formed around DNA[18] and Kang and P. E. Smith's aqueous *N*-methyl-acetamide (NMA) simulations which produced phase separation at $x_{NMA} = 0.1$.[45] These and other simulations prompted an investigation of the ability to accurately simulate intermolecular interactions.

The relevance and severity of the above problem is magnified as advancing computational power reduces the expense of large, long simulations, thereby minimizing the usual concerns associated with system size and sampling. Binding studies and other simulations that probe energetics are increasingly common, but errors due to inappropriate force field parameters can be problematic since the analysis of these simulations relies on the ranking of energetics.

As we discussed in Chapter 1, in traditional FF development the parameters are often chosen to reproduce properties such as the density, diffusivity, excess enthalpy of vaporization, compressibility, dielectric constants, and thermal expansion primarily of *pure* liquids. It is then assumed that if the pure liquids' properties are well reproduced, their *mixtures* will be as well. As mentioned above, this is not always the case. The KBFF approach additionally bases its parameterization on agreement with the KBIs over a range of compositions, thus taking into consideration the reproduction of mixture properties, without sacrificing traditionally reproduced physical and thermodynamic properties of pure liquids.[15] This method makes it possible to obtain an improved balance between solute-solute, solute-solvent, and solvent-solvent interactions.

As we continue to develop a FF specifically designed to reproduce the experimental KBIs for condensed phase solution mixtures applicable to the simulation of peptides and small proteins, models for the aromatic amino acid sidechains are needed. In this work, toluene (Tol) is used as a model for the Phe sidechain, *p*-cresol (pCr) as a model for the Tyr sidechain, and pyrrole (Pyrr) and indole (Ind) as models of the Trp sidechain. Before simulating models of toluene, *p*-cresol, pyrrole, and indole however, the simplest aromatic molecule, benzene (Ben), was studied in order to develop parameters for a fully conjugated carbon atom (C_{ar}). The C_{ar} atom type was then used to create models for the other aromatics. The systems under study and their atom types are shown in Figure 2.1.

It is clear that our approach relies strongly on the assumption that a protein is simply a sum of its parts (amino acids) and likewise, that an amino acid can be described by a combination of small molecule models. For example, we assume that by modeling

toluene solutions well and *N*-methylacetamide (a model for the peptide backbone) solutions well, that we have a good description of phenylalanine solutions. Extending beyond this, we assume that by linking models for the amino acids together, we will have a good description of a protein. This is a standard assumption in MD, which is termed *additivity*.

Previously the Smith group has developed models for the sidechains of other amino acids by studying their analogues solvated in water (Table 1.5). The aromatics representative of the Phe, Tyr, and Trp side chains are only sparingly soluble in water however, making the study of aqueous solutions problematic. To avoid this problem we have switched to solutions of aromatics in methanol (or in toluene when experimental methanol data was not available). The KBFF for methanol was previously developed by S. Weerasinghe and P. E. Smith[47] and aqueous solutions employed the SPC/E water model.[53] We are then assuming that our FF is *transferable e.g*., a good description of benzene + methanol solutions and a good description of methanol + water solutions will provide for a good description of benzene + water solutions.

Methods

Molecular Dynamics Simulations

All binary mixtures were simulated with classical molecular dynamics techniques using the Gromacs program[54] in the isothermal isobaric (*NpT*) ensemble. All simulations

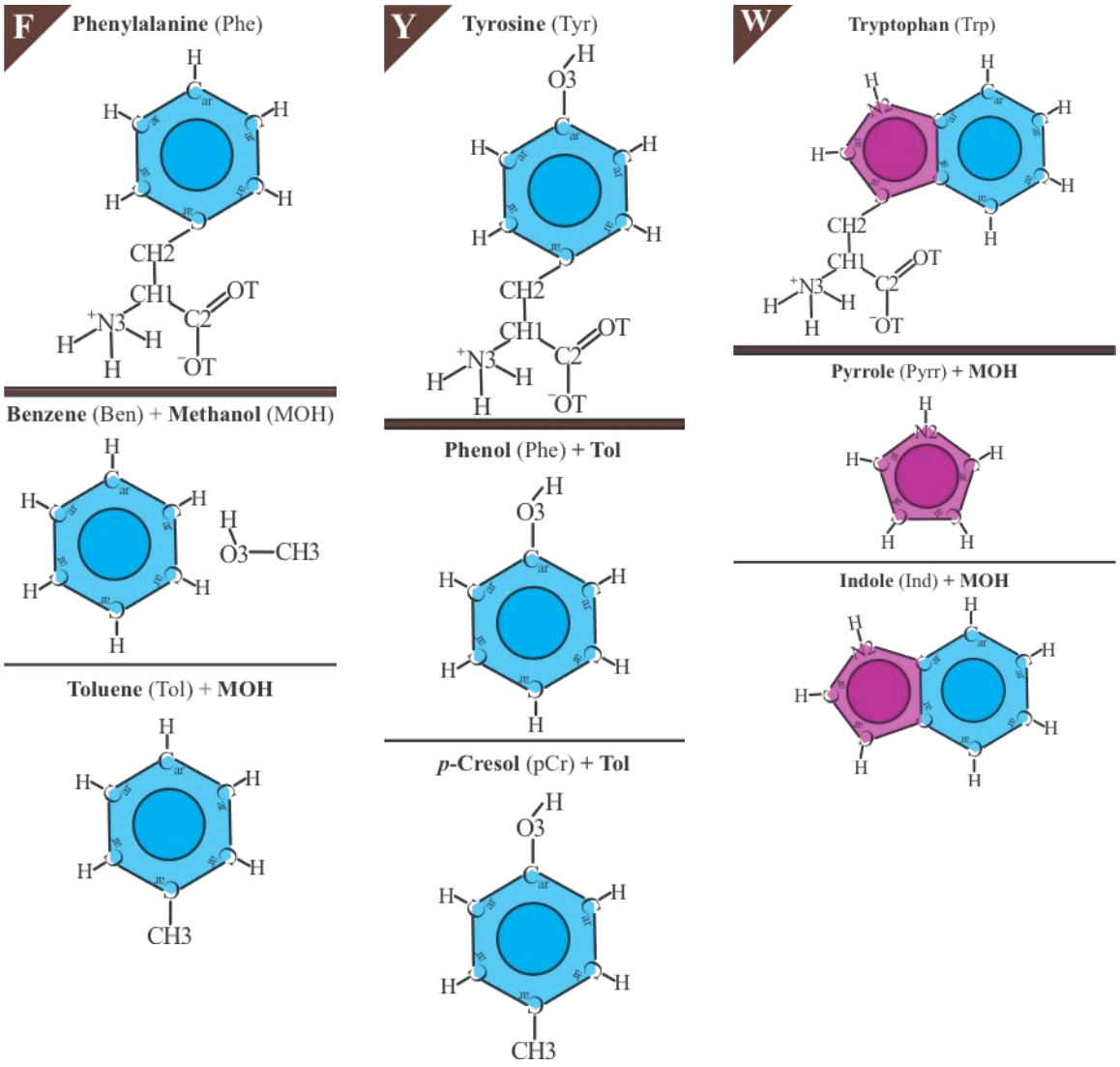


Figure 2.1: Systems of Interest. **Left:** *Top***:** Phenylalanine with atom types labeled. *Middle/Bottom:* Systems studied to develop Phe charge distribution. **Center:** *Top:* Tyrosine with atom types labeled. *Middle/Bottom:* Systems studied to develop Tyr charge distribution. **Right:** *Top:* Tryptophan with atom types labeled. Middle/Bottom: Systems studied to develop Trp charge distribution.

were performed at a pressure of 1 atm. The Berendsen weak coupling technique[55] was used to modulate the temperature and pressure with relaxation times of 0.1 ps and 5.0 ps, respectively, with a 4.5×10^{-5} bar^{-1} compressibility. All bonds were constrained using

the LINCS[56] algorithm and a relative tolerance of 0.1 fs. The bond constraints allowed for a 2 fs timestep for the integration of the equations of motion, which was performed using the Leap-Frog algorithm.[57] The particle-mesh-Ewald technique[58] was used to calculate the electrostatic interactions, employing a twin range cutoff of 1.2 and 1.5 nm for the real space electrostatic and van der Waals interactions, respectively. The neighbor list was updated every 10 steps. Initial molecular configurations were generated from random in cubic boxes. The steepest descent method was then used to perform ≥1000 steps of energy minimization. This was followed by extensive equilibration that was continued until intermolecular potential energy contributions and rdfs displayed no drift with time. Configurations were saved every 1.0 ps for analysis. A summary of the simulations is presented in Appendix A.1. All mixture simulations were run for at least 20 ns to ensure that the time history of both the density and KBIs no longer displayed systematic variations with time. We found that relatively long simulation time were necessary to yield reasonably precise data, presumably due to the large size of aromatics when compared to our previously studied molecules.

Initially, we intended to use the same Lennard Jones (LJ) parameters previously developed by Kang and Smith for the amide group for the carbon in benzene and to only adjust the charge distribution on the C-H bonds of benzene in order to reproduce the experimental KBIs.[19] We found that the KBIs were indeed very sensitive to the charge distribution (Figure 2.2) and in a binary mixture of Ben + methanol (MOH) the degree of MOH-MOH aggregation was noticeably different even from a visual inspection of the trajectories (Figure 2.3), however that the density and enthalpy of vaporization of pure benzene were far from the experimental values. We found that is was necessary to

develop a new atom type for a fully conjugated carbon, C_{ar}, in order to get better agreement for these properties as well.

Parameter Development: Lennard Jones Parameters

Starting with an initial guess for the charge distribution (C_{ar}= -0.14 |e|, taken from Gromos) the benzene LJ sigma and epsilon of the sp^2 amide carbon type were adjusted to best reproduce the experimental liquid density[59] of 0.8629 g/cm^3 at 308 K and ΔH_{vap}[60] of 33.85 kJ/mol at 298 K. Our final simulated density is 0.862 g/cm^3 and ΔH_{vap} is 32.88 kJ/mol. The benzene crystal dimensions[61] are also well reproduced (Figure 2.4) which

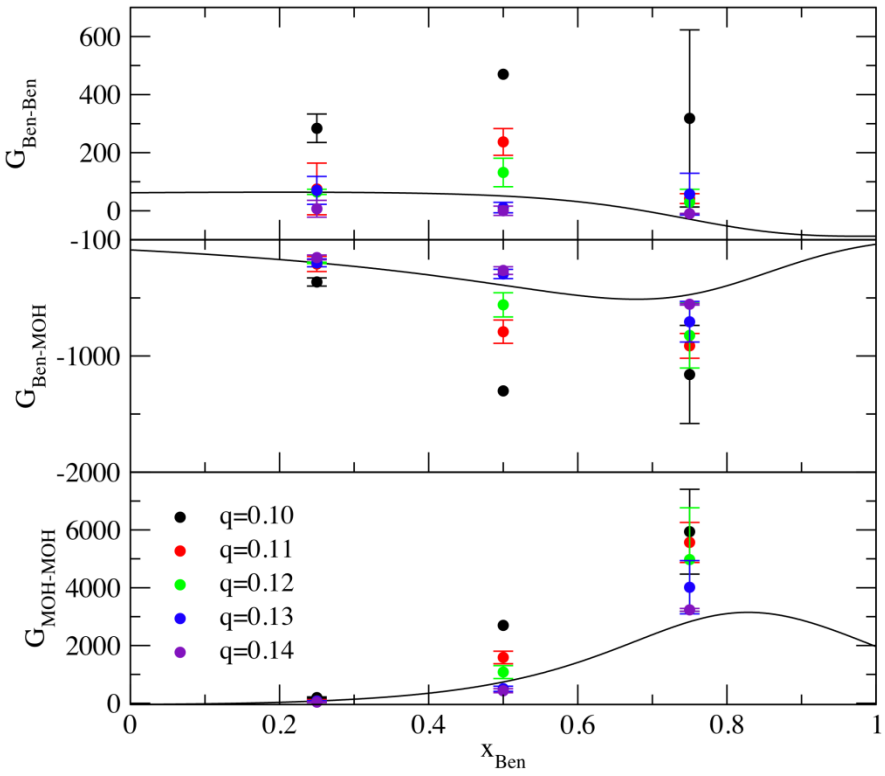


Figure 2.2. Sensitivity of the KBIs to the charge distribution. G_{ij} units are cm^{3}mol^{-1}. This figure is strictly to illustrate the systematic trends observed as q is adjusted. The KBIs corresponding to q = +/- 0.13 do not match the final KBIs because the sigma, epsilon, angle, and torsional parameters were not the same as the final parameters. Lines: Experiment, Points: Simulation.

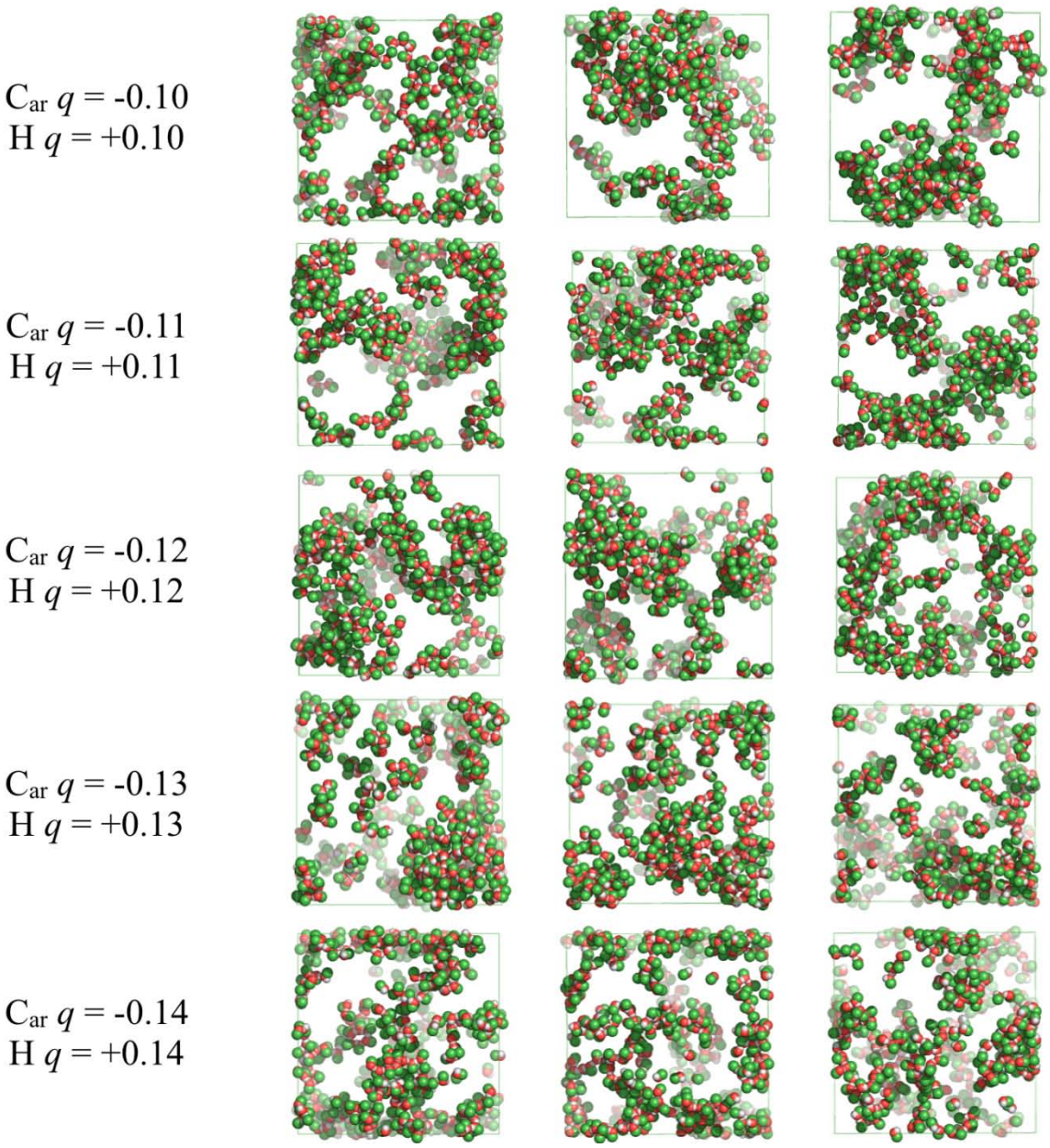
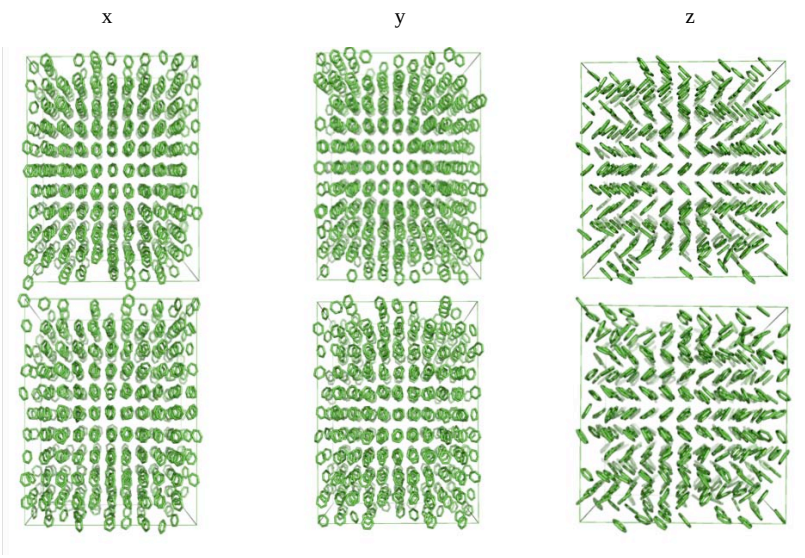


Figure 2.3 Sensitivity of MOH-MOH aggregation to the Ben charge distribution from a visual inspection of the Ben-MOH system at x_{Ben} = 0.500. MOH molecules were rendered as spheres while Ben molecules were excluded from view. Green: united CH3, Red: oxygen, White: hydrogen.

provides confidence that the parameter describing the size of C_{ar}, σ, and the atom type's energy of most favorable interaction, ε, were appropriately chosen. The final C_{ar} parameters were found to be σ = 0.381 nm and ε = 0.33 kJ/mol. Next, the charge



	x (Å)	y (Å)	z (Å)	V (nm^3)
Simulation	37.271	48.955	35.475	64.728
Experiment[61]	37.049	48.664	35.264	63.579

Figure 2.4: Unit cell dimensions of the benzene crystal structure after 1 ns MD simulation at 270 K (bottom) compared to experimental dimensions (top). The edges appear to show more deviations from the experimental structure since a snapshot only captures one unit cell of a periodic system. Molecules crossing the dimensions of this cell, even if due to only small translations, will appear on the opposite side of the cell.

distribution was developed by comparing the simulated KBIs with those extracted from experimental data.

Kirkwood-Buff Analysis of Simulated Data

Simulated KBIs (G_{ij}) are calculated by integrating the radial distribution functions (rdfs), and are approximated for a closed system by the rhs of Equation 1.16. G_{ij} signifies deviations in molecular distributions from that of a bulk solution due to the effects of species-specific intermolecular interactions.

In this work the KBI values were obtained by carefully averaging over a range in which the rdfs were converged and the KBIs had reached a plateau. This small range of distances, corresponding to one solvation shell, was not the same for each system or for each of the three KBIs within a specific system. The simulation results in which the KBI (or correspondingly the $N_{ij} = \rho G_{ij}$) were not converged have not been reported because they are statistically unreliable. Error estimates were obtained from the averages of multiple 5 ns runs.

KB theory allows for thermodynamic properties to be extracted from the molecular information available in simulations.[35] These properties include the mole fraction derivative of the chemical potential (μ_{ii}), partial molar volume ($\bar{V}_i$) and isothermal compressibility (κ_T) of binary mixtures at constant pressure (p) and temperature (T), and are obtained according to the following relationships in Equations 1.16-1.17.

Kirkwood-Buff analysis of Experimental Data

Due to Ben-Naim's development of an inversion procedure, the KB theory can also be used to extract information about the molecular structure from thermodynamic data. The KBIs obtained by analysis of the experimental data, which correspond to integrals over rdfs in the grand canonical (μVT) ensemble, are given by[35]

$$G_{ij} = \lim_{r\to\infty} G_{ij}(r) = 4\pi \int_0^\infty [g_{ij}^{\mu VT}(r) - 1] r^2 \, dr \qquad (2.1)$$

and are calculated from the μ_{ij}, $\bar{V}_i$, and κ_T of binary mixtures according to Equation 1.14.

Partial molar volumes were calculated from experimental data of the solution's mass density or volume as a function of the system's composition according to[48]

$$\bar{V}_i = V_m + x_j \frac{dV_m}{dx_i};\ \bar{V}_j = (V_m - x_i \bar{V}_i) / x_j \tag{2.2}$$

where

$$\frac{dV_m}{dx_i} = \frac{dV^E}{dx_i} + V_i^0 - V_j^0, \tag{2.3}$$

$$V_m = x_i \bar{V}_i + x_j \bar{V}_j, \tag{2.4}$$

and

$$V^E = V_m - x_i V_i^0 - x_j V_j^0. \tag{2.5}$$

Here V^E is the excess molar volume and V_i^0 is the molar volume of pure component *i*.

For systems in which there was no experimental data available for the density or partial molar volumes as a function of composition, it was assumed based on precedence that $\bar{V}_i = V_i^0$, leading to $V_m = x_i V_i^0 + x_j V_j^0$ and $V^E = 0$.[62]

It has also been shown that the isothermal compressibility has an insignificantly small effect on the calculation of the KBIs.[62] Thus, for all systems the solution isothermal compressibility can simply be calculated according to $\kappa_T = \rho_i \bar{V}_i \kappa_{T,i} + \rho_j \bar{V}_j \kappa_{T,j}$, where ρ_i is the number density of *i* and $\kappa_{T,i}$ is the isothermal compressibility of pure *i*.

All systems were studied as binary mixtures with methanol except for PhOH and pCr because isothermal activity data was not found for those systems. This led us to study PhOH and pCr in solution with toluene, for which isothermal activity data was available.

As discussed in Chapter 1, experimental activity data is often obtained from models of the excess molar Gibbs energy, g^E. The Wilson equation provides g^E for a system with any number of components according to[41]

$$\frac{g^E}{RT} = -\sum_i x_i \ln(1 - \sum_j x_j \Lambda_{j/i}) \qquad (\Lambda_{i/i} = 0,\ \Lambda_{j/i} \neq \Lambda_{i/j}), \tag{2.6}$$

where x_i is the mole fraction of component i and $\Lambda_{j/i}$ is an adjustable parameter. This equation leads to Raoult's law at $x_i = 1$ and Henry's law at $x_i = 0$.[41] The parameters $\Lambda_{i/j}$ and $\Lambda_{j/i}$ are positive if the deviation from ideality is positive and negative if the deviation is negative. If one Λ is positive and the other Λ is negative, then the deviation may change signs over the composition range and the sign of the deviation will depend upon the dominating Λ. The Λ parameters are related to the pure-component molar volumes and the energies of interactions between molecules i and j, denoted by λ_{ij}, according to[41]

$$\Lambda_{ij} = \frac{V_j^0}{V_i^0} \exp\left(-\frac{\lambda_{ij} - \lambda_{ii}}{RT} \right) \tag{2.7}$$

$$\Lambda_{ji} = \frac{V_i^0}{V_j^0} \exp\left(-\frac{\lambda_{ij} - \lambda_{jj}}{RT} \right). \tag{2.8}$$

Λ must be less than or equal to one to ensure that g^E is real over the full composition range and a Λ of zero indicates no interaction between molecules.

Another model for g^E is Renon's Non-Random Two Liquid (NRTL) equation, which is given by[41]

$$\frac{g^E}{RT} = x_i x_j \left(\frac{\tau_{ji} G_{ji}}{x_i + x_j G_{ji}} + \frac{\tau_{ij} G_{ij}}{x_j + x_i G_{ij}} \right) \tag{2.9}$$

where

$$\tau_{ij} = \frac{g_{ij} - g_{jj}}{RT} \qquad \tau_{ji} = \frac{g_{ji} - g_{ii}}{RT}, \tag{2.10}$$

and

$$G_{ij} = \exp(-\alpha\tau_{ij}) \qquad G_{ji} = \exp(-\alpha\tau_{ji}). \tag{2.11}$$

Here g_{ij} is similar to Wilson's λ_{ij} parameter; it characterizes the interaction energy between molecules *i* and *j*. α characterizes the non-randomness of the mixture *i.e.*, if α is equal to zero the mixture is completely random. From these equations the activity coefficients can be expressed as[41]

$$\ln\gamma_i = x_j^2\left[\tau_{ji}\left(\frac{G_{ji}}{x_i + x_j G_{ji}}\right)^2 + \frac{\tau_{ij}G_{ij}}{(x_j + x_i G_{ij})^2}\right] \tag{2.12}$$

$$\ln\gamma_j = x_i^2\left[\tau_{ij}\left(\frac{G_{ij}}{x_j + x_i G_{ij}}\right)^2 + \frac{\tau_{ji}G_{ji}}{(x_i + x_j G_{ji})^2}\right]. \tag{2.13}$$

The notation used in Equations 2.9 – 2.13 is repeated for emphasis. g^E is the excess molar Gibbs energy, the g_{ij} parameter describes the interaction energy between particles *i* and *j*, and the G_{ij} is the negative exponential of the multiplied NRTL alpha and tau parameters. These are the standard notations of the field, but here they could unfortunately be confused with radial distribution functions, g_{ij}, or Kirkwood-Buff integrals, G_{ij}. The reader is cautioned to take care when following this notation to avoid mishap.

Experimental Sources for Composition and Activity Data

Benzene + Methanol. Partial molar volumes were calculated from experimental density versus composition data at 303 K and 1 bar.[59] g^E was fit using the Wilson equation that is expressed for a binary mixture as[42]

$$\frac{g^E}{RT} = -x_i \ln(1-\Lambda_{2/1}x_2) - x_2 \ln(1-\Lambda_{1/2}x_1) \tag{2.14}$$

where $\Lambda_{1/2}$ = 0.6570 and $\Lambda_{2/1}$ = 0.91226 for a mixture of Ben (1) and MOH (2) at 308 K.[42] The activity coefficients were obtained from fitting the total pressure data for this mixture in which the parameters were used to fit the data according to the relationships[42]

$$\ln\gamma_1 = -\ln(1-\Lambda_{2/1}x_2) + x_2\left[\frac{x_2\Lambda_{1/2}}{1-\Lambda_{1/2}x_1} - \frac{x_1\Lambda_{2/1}}{1-\Lambda_{2/1}x_2}\right] \tag{2.15}$$

$$\ln\gamma_2 = -\ln(1-\Lambda_{1/2}x_1) + x_1\left[\frac{x_2\Lambda_{1/2}}{1-\Lambda_{1/2}x_1} - \frac{x_1\Lambda_{2/1}}{1-\Lambda_{2/1}x_2}\right]. \tag{2.16}$$

Toluene + Methanol. Partial molar volumes were calculated from V^E data for this mixture at 318 K and 1 bar.[63] Activity coefficients were reported directly in an experimental isothermal VLE study[64] at 318 K and 1 bar. From the activity data, g^E was calculated as a function of composition and fit using the non-linear least squares method with three parameters.

Toluene + Benzene. In this system, the partial molar volumes were set equal to the molar volumes of the pure components. Activity data was obtained from experimental isothermal VLE data at 313 K using the Wilson equation, with $\Lambda_{1/2} = 4.2 \times 10^{-5}$ and $\Lambda_{2/1} = -4.3 \times 10^{-5}$.[65] These Λ values are significantly smaller than those in the Ben + MOH system; this is a reflection of the ideality of this system.

Toluene + Phenol / Toluene + *p*-Cresol. Isothermal VLE data[66] provided a fit for V^E using the Redlich – Kister equation (Equation 1.14) in which the coefficients for Tol + PhOH were given as A_1 = -1.2413, A_2 = 0.3561, and A_3 = 0.0792 and the coefficients for Tol + pCr were given as A_1 = -0.9125, A_2 = +0.1173, and A_3 = +0.0479.

The NRTL g^E model was used to calculate the activity coefficients in which $\tau_{ij}T = (g_{ij} - g_{jj}) / R = C_{ij}$, where C_{ij} is a temperature dependent parameter given by C_{ij} / K $= C_{ij}^C + C_{ij}^T (T = 273.15 \text{ K})$.[66] The parameters for Tol + PhOH were given as α = 0.20, C_{12}^C(K) = +857.14, C_{21}^C(K) = -308.41, C_{12}^T(K) = -4.3775, and C_{21}^T(K) = +2.8430. Tol + pCr parameters were given as α = 0.20, C_{12}^C(K) = +443.83, C_{21}^C(K) = -50.11, C_{12}^T(K) = +1.6730, and C_{21}^T(K) = -1.6607. These systems are soluble over the full composition range and both systems exhibit a negative V^E over the full composition range.

Pyrrole + Methanol. Activity data was taken from experimental LLE data[67] at 298 K for the ternary system methanol (1) + hexadecane (2) + pyrrole (3), for which the NRTL equation has the form[67]

$$\frac{g^E}{RT} = \sum_{i=1}^{n} \frac{\sum_{j=1}^{n} \tau_{ji} G_{ji} x_j}{\sum_{i=1}^{n} G_{li} x_i}, \tag{2.17}$$

where

$$\ln \gamma_i = \frac{\sum_{j=1}^{n} \tau_{ji} G_{ji} x_j}{\sum_{l=1}^{n} G_{li} x_l} + \sum_{j=1}^{n} \left(\frac{x_j G_{ij}}{\sum_{l=1}^{n} G_{li} x_l} \right) \left(\tau_{ji} - \frac{\sum_{k=1}^{n} \tau_{kj} G_{kj} x_k}{\sum_{l=1}^{n} G_{lj} x_l} \right), \tag{2.18}$$

and x_2 was simply set equal to zero. The Pyrr + MOH NRTL parameters were given as $(g_{ji} - g_{ii})/R = 524.80$, $(g_{ij} - g_{jj})/R = -666.43$, and $\alpha = 0.010$.

Indole + Methanol. Activity data at 298 K was taken from the same source as for the pyrrole + MOH system.[67] The Ind + MOH NRTL parameters were given as $(g_{ji} - g_{ii})/R = 967.32$, $(g_{ij} - g_{jj})/R = -1016.99$, and $\alpha = 0.009$. Pure indole is a solid at 298 K.[68]

Parameter Development: Partial Charges

Once both simulated and experimental KBIs (G_{ii}, $G_{ij} = G_{ji}$, and G_{jj}) were obtained, they were compared to provide information for the parameterization of the charge distributions for the molecules of interest. The benzene model has explicitly defined hydrogens in which the charge is distributed evenly between each carbon and hydrogen pair as a simple model of electrostatic interactions. Charges on the atoms were adjusted to best reproduce the experimental ΔH_m and KBIs for solution mixtures over a range of compositions. The systems of Ben + MOH, Tol + MOH, and Tol + Ben were studied simultaneously to develop the C_{ar}-H charge distribution. Despite the initial use of a two parameter fit to develop the sigma and epsilon based upon the density and enthalpy of vaporization of liquid benzene, perfect agreement was not achieved due to the iterative nature of our approach whereby subsequent adjustments to the charge distribution created small changes in the simulated density and enthalpy of vaporization.

Results and Discussion

The rdf, KBI, and N_{ij}

Table 2.1: Non-bonded parameters for Phe, Tyr, and Trp. Bonded parameters were taken from Gromos.[6] Parameters are listed for the atom type in bold. Plain typeface atom types are included to illustrate the change in *q* depending upon the neighboring atom type.

Model	Atom type	ε (kJ/mol)	σ (nm)	q (|e|)
Aromatics				
	C_{ar}−H (sp^2 aromatic)	0.3300	0.3810	-0.130
	C_{ar}−CH3 (sp^2 aromatic)	0.3300	0.3810	-0.130
	C_{ar}−O3 (sp^2 aromatic)	0.3300	0.3810	+0.240
	C_{br} (sp^2 bridging carbon)	0.4170	0.3770	0.000
	H−C_{ar}	0.0880	0.1580	+0.130
	H−O3	0.0880	0.1580	+0.416
	CH3−C_{ar}(united methyl)	0.8672	0.3748	+0.130
	C_{ar}−**O3**−H (hydroxyl)	0.6506	0.3192	-0.656
	N2-H (sp^2)	0.5000	0.3110	-0.450
	N2-**H**	0.0880	0.1580	+0.450
Methanol[47]				
	O	0.6506	0.3192	-0.820
	H	0.0880	0.1580	+0.520
	CH3	0.8672	0.3748	+0.300

The iterative procedure used to obtain the optimal charge distribution resulted in the choice of the final parameters as shown in Table 2.1. The results for the Ben + MOH system will be discussed in the greatest detail to illustrate the mindset that may be used when looking at the results for the remaining systems. For all systems excluding Ben + MOH, the experimental and simulated KB integrals are compared in Figure 2.10 and shown in Figure 2.11 after a multiplication by the number density. $\rho_j G_{ij}$ is the excess coordination number and is denoted N_{ij}. Discussions of N_{ij} diminish the noise in the KBIs by effectively minimizing the uncertainty that is inherent at low concentrations of *i* or *j* in both experimental and simulated KBI data. This uncertainty is attributed to both sampling error (due to fewer *j* molecules at high *i* concentration, and *vice versa*) and, at high concentration of the aromatic molecules, the relative difficulty of large aromatic

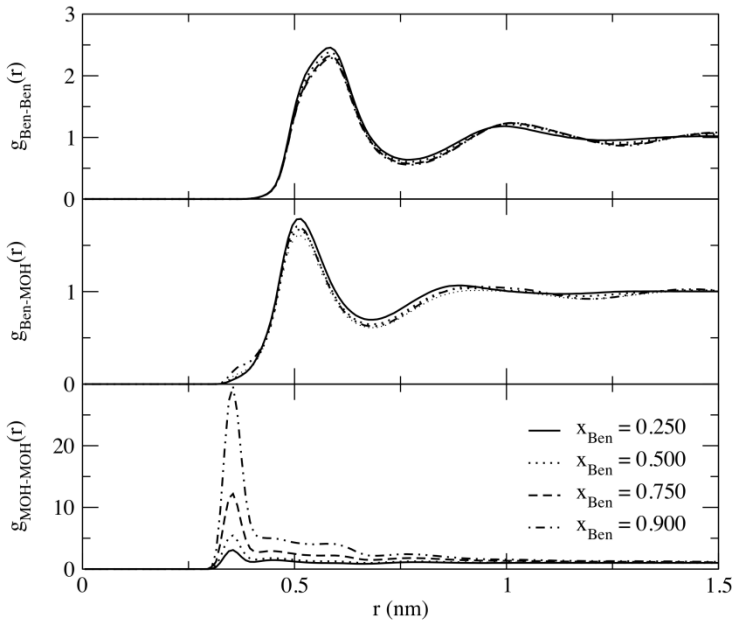


Figure 2.5. Center of mass based radial distribution function for Ben (1) + MOH (2) system as a function of distance. As the concentration of MOH decreases, the $g_{MOH-MOH}$ peak increases dramatically, whereas the other two rdfs remain virtually unchanged.

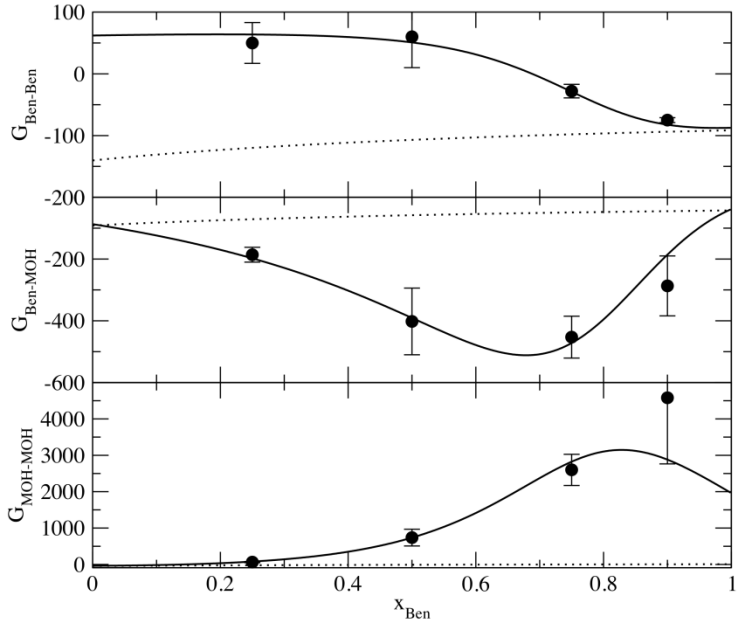


Figure 2.6. The KBFF's reproduction (points) of the experimental (lines) Ben + MOH KBIs as a function of composition. As the concentration of MOH decreases, the large $G_{MOH-MOH}$ attests to the self-aggregation of MOH. The calculated SI KBIs are show as dotted lines to quantify the deviation of this system from symmetric ideality.

molecules to move past each other. Additionally, the N_{ij} has a physical interpretation of being the excess or deficit in the number of *j* molecules within a certain volume of space away from a central *i* molecule when compared to the number that would have been found had the central *i* not been present.

Benzene (1) + Methanol (2), 308 K. The center of mass rdfs are displayed in Figure 2.5 as a function of composition. As x_{MOH} decreases, $g_{22}(r)$ greatly increases, illustrating that the probability of finding one MOH *r* away from another MOH increases dramatically at low x_{MOH}. Noticeably, g_{11} and g_{12} remain almost constant as the concentration changes. The calculated experimental and simulated KBIs (N_{ij}) are compared in Figure 2.6 (Figure 2.7). A large, positive G_{ij} (N_{ij}) such as that observed from the experimental peak of ~3100 cm^3mol^{-1} (~9) for G_{22} (N_{22}) at x_{Ben} ~0.8 (~0.7), can be interpreted to mean that there is a clustering of MOH molecules at this composition.

It is interesting to note that while $g_{Ben\text{-}Ben}$ remains almost constant as the composition of the mixture changes, $G_{Ben\text{-}Ben}$ does not; $G_{Ben\text{-}Ben}$ actually changes sign as the concentration varies. This is one example of an important characteristic of the KBIs in any mixture *i.e.*, subtle features of rdfs are magnified in the corresponding KBIs due to the r^2 weighting factor present in Equation 1.17.

If Ben + MOH were a symmetric ideal (SI) system, the G_{22} (N_{22}) would be 2 cm^3mol^{-1} (0),[35]

$$G_{ij}^{SI} = RT\kappa_T - V_i - V_j + \rho_i V_i^2 + \rho_j V_j^2 . \qquad (2.19)$$

The deviations from these SI values serve to quantify the deviation of this mixture from an ideal mixture.

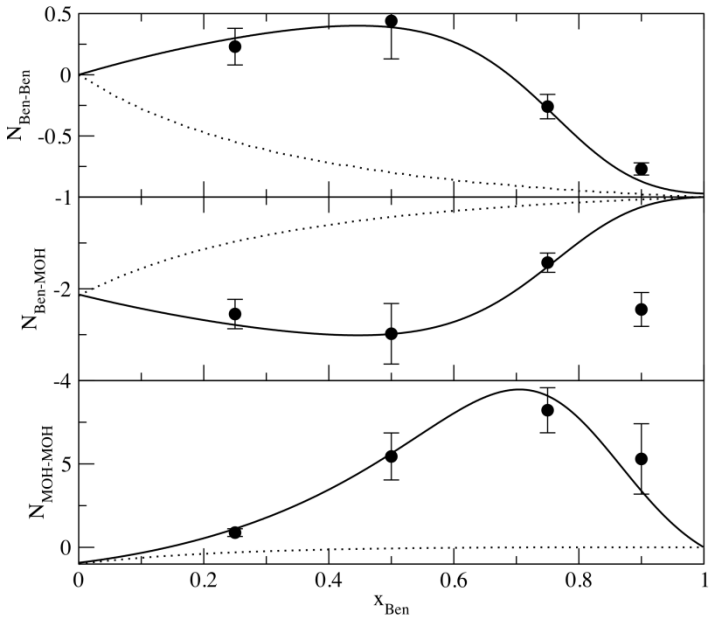


Figure 2.7. The KBFF's reproduction (points) of the experimental (lines) Ben + MOH excess coordination numbers, N_{ij}, as a function of composition. As the concentration of MOH decreases, the large $N_{\text{MOH-MOH}}$ suggests self-aggregation of MOH. The calculated SI N_{ij} values are show as dotted lines to quantify the deviation of this system from ideality.

Figure 2.8 shows representative MOH clusters observed at $x_{\text{Ben}} = 0.900$. A visual inspection revealed that MOH predominantly forms MOH-MOH chain structures over MOH-MOH cyclic structures which is an observation that is in agreement with other simulations and experiments,[69] however a rigorous analysis was not performed of the percentages or lifetimes of these structures.

As introduced in Figure 2.2, it is interesting to note that while benzene is not considered a polar molecule, large differences were observed in the simulated KBIs as changes were made to the charge distribution between the C_{ar} and H atoms. We believe this is because any change in the point charges represents a cumulatively large change in the overall charge distribution of the molecule due to the six C_{ar}-H bonds per benzene.

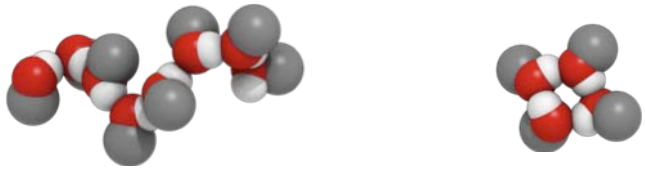


Figure 2.8. Representative snapshots of methanol chains (left) and rings (right) observed in simulations of Ben + MOH at 308 K with x_{Ben} = 0.9.

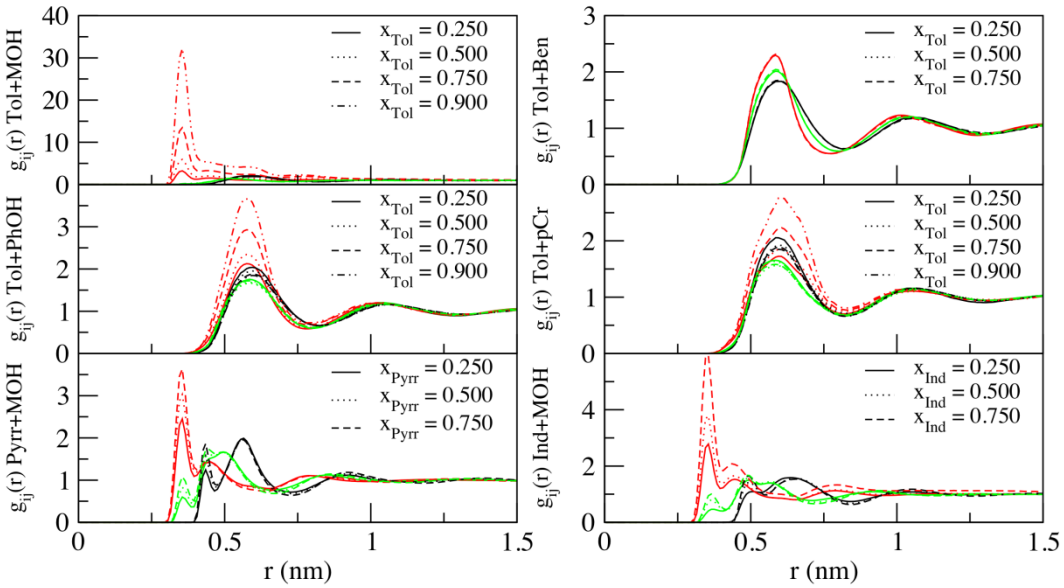


Figure 2.9. Center of mass based radial distribution functions as a function of distance. Black: g_{11}, Red: g_{22}, Green: g_{12}. In all cases in which g_{22} is an alcohol (every mixture except Tol + Ben), g_{22} is highly concentration dependent and it increases as the amount of alcohol decreases.

Additionally in Ben–Ben interactions the charge parameters' significance is even more pronounced. As the charge distribution was increased, there was a systematic reduction in the aggregation of MOH-MOH as was evident both from a decreasing G_{jj} and from a visual inspection of snapshots of these mixtures (Figure 2.3) in which the number of MOH that are free in solution increases as the charge distribution is increased.

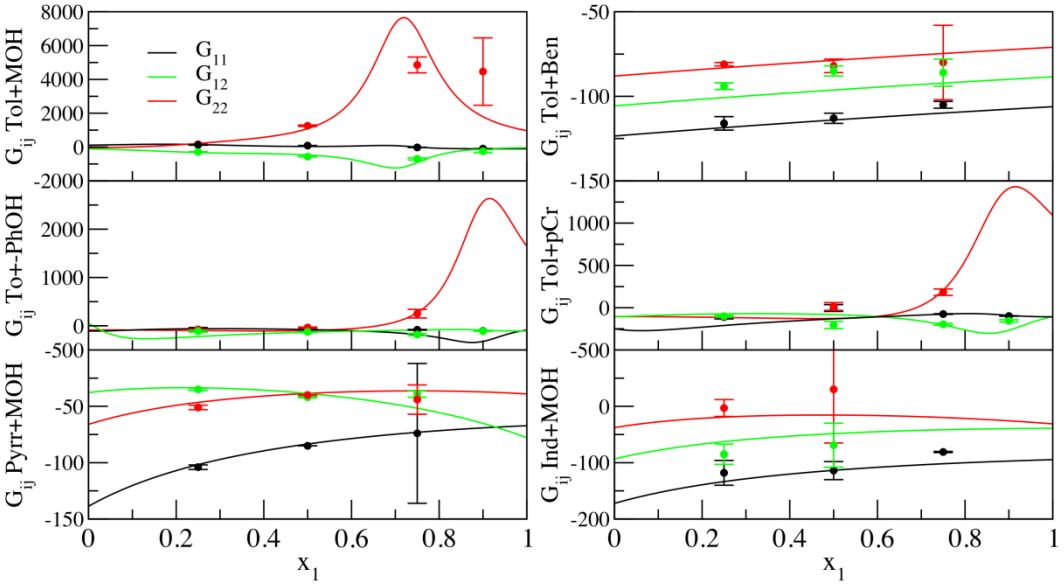


Figure 2.10. Kirkwood-Buff Integrals (in cm^3mol^{-1}) as a function of composition. Lines: Experiment, Points: Simulation.

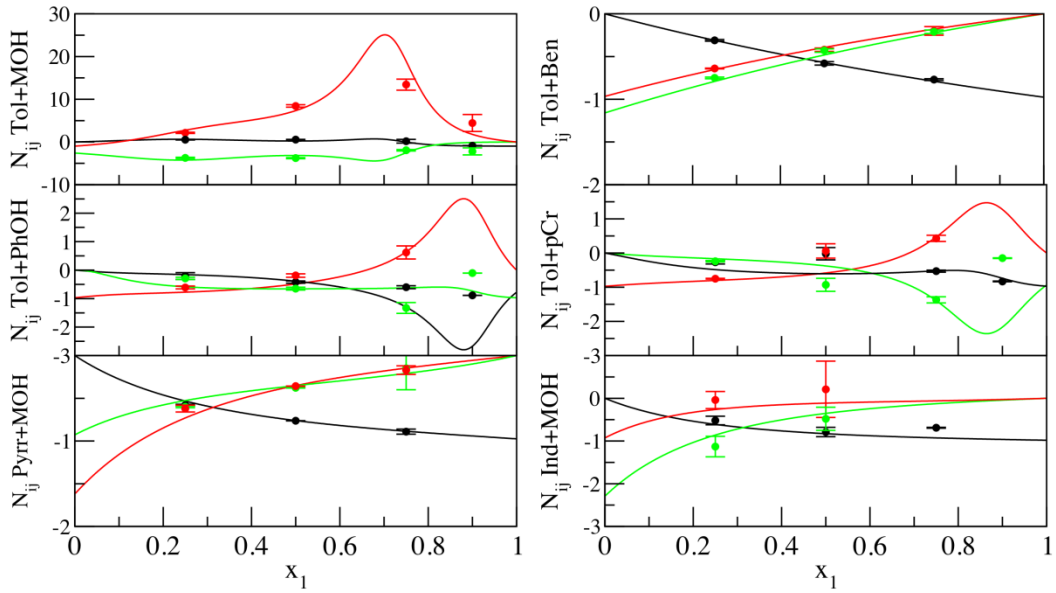


Figure 2.11. Excess Coordination Numbers as a function of composition. Lines: Experiment, Points: Simulation.

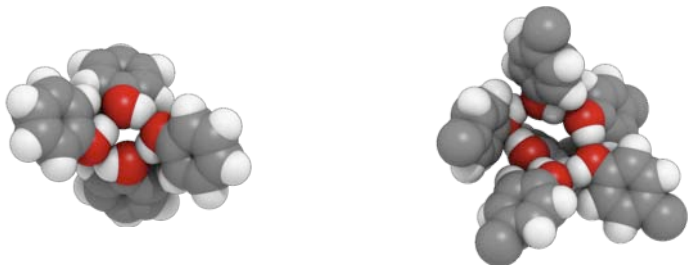


Figure 2.12. Representative snapshots of phenol (left) and *p*-cresol (right) cyclic ring structures observed in simulations of Tol + PhOH or Tol + pCr at 333 K and $x_{Tol} = 0.9$.

As we have developed the KBFF, we have repeatedly seen similar sensitivities of the KBIs to the charge distributions in other systems. We believe that the responsiveness of the KBIs to changes in charge distributions is a major advantage of our approach.

For the sake of brevity, the remaining mixtures will be discussed in less detail. The rdfs for these mixtures are shown in Figure 2.9, the KBIs in Figure 2.10, and the excess coordination numbers in Figure 2.11.

Toluene (1) + Methanol (2), 318 K. Although there is only a small difference between the atomic construction of benzene and toluene, the experimental N_{22} displays a significantly greater peak, over twice the magnitude of that in the Ben + MOH system, at $x_{TOL} \approx 0.7$. The peak is in the same position, but much narrower than the peak in Ben + MOH. Our simulation does not quite produce this peak; it provides a N_{22} value closer to that seen in the Ben + MOH system.

The charge distribution for the C_{ar}-CH3 on Tol was tested in order to determine if there should be polarity across the bond or if the KBIs would be better reproduced with charges of zero across this bond. Based upon fundamental chemical principles, it seems reasonable that one could argue either way. The electronegativity difference between C_{ar} and CH3 is minimal, supporting charges of 0.00 and 0.00 on the C_{ar} and CH3. On the

contrary, our simple model of aromaticity on the benzene ring would be disturbed under this arrangement. We tested Tol + MOH (and Ben + Tol) with these charges set equal to zero against Tol + MOH (and Ben + Tol) using the same charges as those across the C_{ar}-H of the benzene ring, and the results showed a clear improvement when the polarity across the bond was retained. Thus, the only way in which our toluene and benzene models differ is in their sigma and epsilon LJ parameters; they have identical coulombic parameters.

Benzene (1) + Toluene (2), 313 K. Ben + Tol is one of the quintessential ideal systems. We have discussed the excellent simulation of this ideality previously.{Ploetz, 2010 #637} The Tol + Ben $g_{ij}(r)$ shows a first solvation shell at ~0.6 nm, however g_{ii} and g_{jj} have slightly different magnitudes with a slightly higher probability of finding a Ben ~0.6 nm away from a central Ben than a Tol ~0.6 nm away from a central Tol. This is due to the tighter packing that is possible between benzene molecules due to the absence of a methyl group when compared to Tol. The nearly straight lines of the G_{ij} and N_{ij} also show the ideality of the system.

Phenol (1) + Toluene (2) / *p*-Cresol (1) + Toluene (2), 333 K. We first attempted to use the MOH charges for the C_{ar}-O3-H charge, however this resulted in poor reproduction of the KBIs. The final charges correspond to the MOH charges scaled by a factor of 0.80. The differences in the partial charges are attributed to both the aromaticity of PhOH and pCr and the large size of PhOH and pCr when compared to MOH, which allows for fewer H-bonding partners to be formed in these solutions. A large peak in the g_{22} is observed as x_{TOL} approaches one. This is illustrated in snapshots of these systems (Figure 2.12) at $x_{TOL} = 0.9$, in which cyclic clusters of phenol or *p*-cresol molecules are

Table 2.2. Comparison between simulated and experimental properties of pure liquids with temperatures (K) equal for simulation and experiment and pressures at 1 bar. Symbols are as follows: ρ is the density, ΔH_{vap} the enthalpy of vaporization, κ_T the isothermal compressibility, and ε the dielectric constant.

Model	T_{melt}[71]	ρ (g/cm^3)			ΔH_{vap} (kJ/mol)			κ_T (× 10^{+5} bar^{-1})			ε		
		T	Exp	Sim	T	Exp	Sim	T	Exp	Sim	T	Exp	Sim
Benzene	278	308	0.863[72]	0.862	298	33.85[60]	32.88	298	9.67[73]	7.5	298	2.2[74]	1
Toluene	178	303	0.857[59]	0.881	298	38.01[71]	39.16	298	9.14[73]	6.2	298	2.4[75]	1
Phenol	314	333	1.042[76]	1.034	333	55.4[76]	60.18	319	5.61[71]	4.1	333	10.3[76]	9
p-Cresol	308	318	1.014[77]	1.028	333	--	67.29	323	6.07[78]	4.1	314	16.4[99]	4
Pyrrole	250	298	0.973[67]	0.973	298	--	--	298	--	5.5	291	7.5[68]	4
Indole	326	298	1.22[68]	1.135	298	--	85.6	330	--	2.6	330	--	2

each observed in an arrangement vaguely reminiscent of reverse micelles. A more thorough discussion of this microstructure can be found in a recent *Fluid Phase Equilibria* conference proceeding.[70]

Pyrrole (1) + Methanol (2) / Indole (1) + Methanol (2), 298 K. As mentioned in the experimental section, indole is a solid at 298 K; however, indole was simulated at 298 K and the PE was used to calculate the enthalpy of mixing in indole + MOH mixtures. Kokubo and Pettitt *et al* set precedence for this approach in their 2007 *Biophysical Journal* article, "Molecular Basis of the Apparent Near Ideality of Urea Solutions" (pp. 3392). This approach is acceptable because indole is in an amorphous state on the ns timescales of our simulations and it does not crystallize into the solid phase, which allows for this metastable state to be characterized.

The rdfs for these systems show secondary solvation shells and a significant increase in the first solvation shell in all three rdfs as the concentration of Pyrr or Ind is increased. The scale is much smaller than that used for systems of Ben + MOH and Tol + MOH, so this does not necessarily indicate that these systems will be highly aggregating. Indeed, the KBIs do not show an association of species in these solutions. The small scale of these rdfs also gives the appearance that the rdfs behave very

differently from Ben + MOH and Tol + MOH in which only $g_{\text{MOH-MOH}}$ changes dramatically with an increase in the aromatic species. This is not the case. Magnification of the Ben-MOH and Tol-MOH rdfs to a scale similar to that of Pyrr + MOH and Ind + MOH exposes that similar changes in the g_{ii} and g_{ij} are observed in these systems as well.

Other Solution Properties

The KBFF is designed to reproduce solution properties, with particular interest in the Kirkwood-Buff integrals; however, the KBFF models for previous systems have been shown to be competitive with other FFs in their ability to reproduce both other mixture properties and the properties of pure liquids. Traditionally, solution properties are not necessarily well reproduced with the mainstream FFs unless, of course, a certain property was the focus of a FF developer's parameterization efforts.[15,17] Here we explore the ability of the KBFF to reproduce (where data is available) the experimental enthalpy of vaporization of the pure liquids, dielectric constants over the full composition range, enthalpies of mixing over the full composition range, and isothermal compressibilities of the pure liquids.

Enthalpy of Vaporization. The calculated enthalpy of vaporization is shown in Table 2.2 and was calculated for the *N i* species according to

$$\Delta H_{\text{vap}} = \frac{-[E_{tot}(l) - E_{intra}(l)]}{N(l)} + RT$$

$$= \frac{-[E_{tot}(l) - E_{angles}(l) - E_{dihedrals}(l) - E_{impropers}(l)]}{N(l)} + RT\,, \qquad (2.20)$$

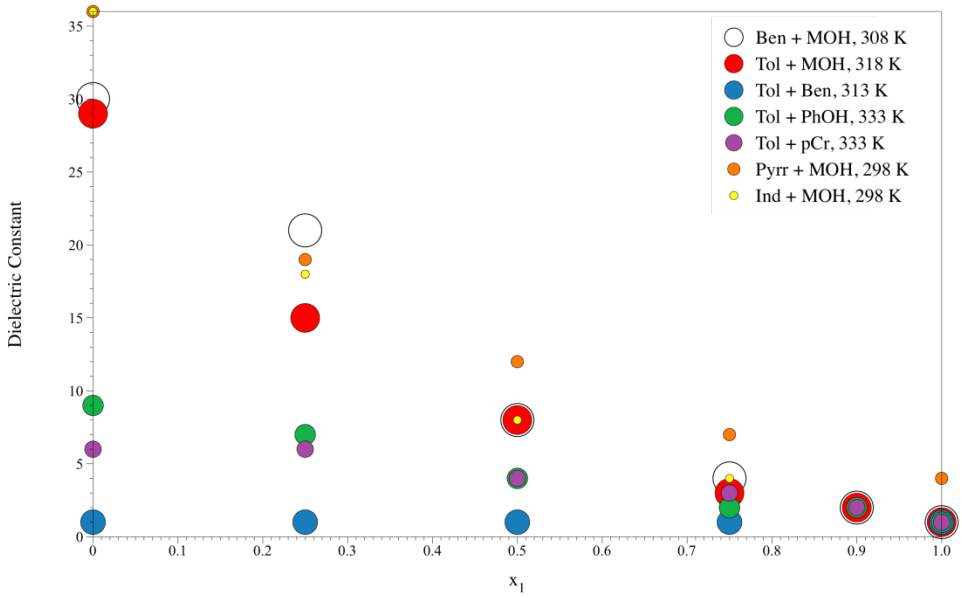


Figure 2.13. Simulated dielectric constants, ε, (error approximated at ±1) of binary mixtures as a function of composition. The size of the circles has no meaning and was used simply to allow the reader to distinguish between different systems that exhibit the same ε at a given composition.

assuming that the intermolecular interactions in the gas phase are negligible. Here E_{tot} is the total potential energy of the N i molecules, the other potential energy contributions are explicitly labeled, $N(l)$ is the number of i molecules in the liquid phase, RT is the pV work contribution from the gas phase (ideal gas approximation), and the pV work term is assumed to be zero for the liquid phase (incompressible liquid approximation).[32,79] As mentioned previously, the iterative nature of the approach used for obtaining the parameters of C_{ar} in which the charge distribution was adjusted after the choice of sigma and epsilon was made, resulted in an imperfect enthalpy of vaporization especially important for polar molecules such as phenol, and quantum corrections for benzene. Other errors in our results could be due in part to the omission of two factors

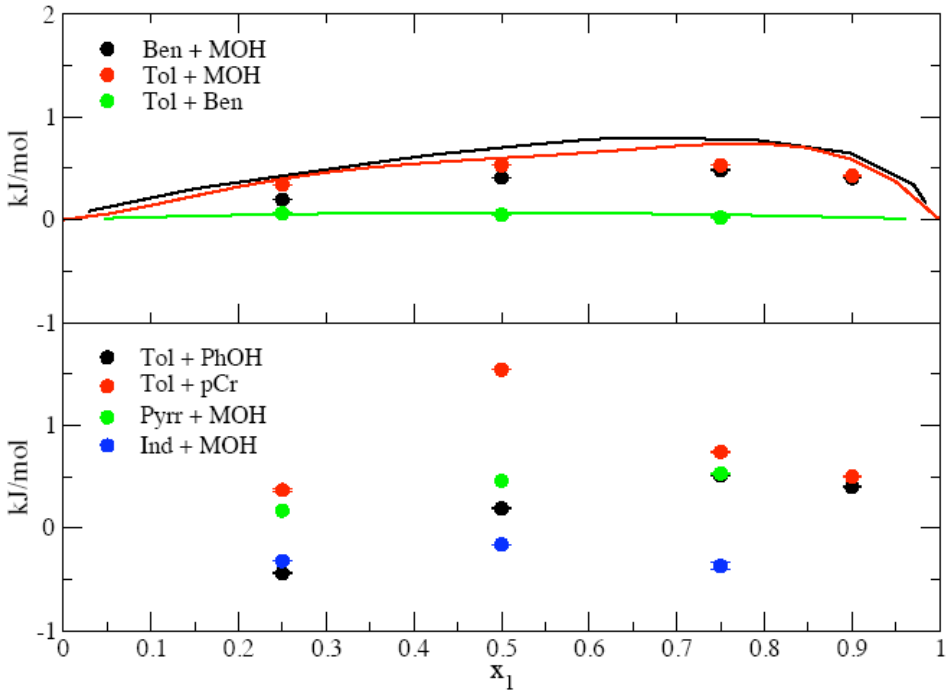


Figure 2.14. Enthalpy of Mixing. *Top:* Comparison of simulated (lines) and experimental (points) enthalpies of mixing as a function of composition for Ben + MOH,[81] Tol + MOH,[82] Tol + Ben.[83] *Bottom:* Simulated enthalpies of mixing as a function of composition for Tol + PhOH, Tol + pCr, Pyrr + MOH, and Ind + MOH.

that contribute to the true enthalpy of vaporization: polarizability, which is vibrational energies.

Dielectric Constant. The relative permittivity was determined from the dipole moment fluctuations[80] using a reaction field permittivity of infinity.[47] Values for the pure liquids are shown in Table 2.2. These values agree moderately well with experiment excluding *p*-cresol in which the appropriate trend in the values from phenol to *p*-cresol is not observed. Dielectric constants as a function of composition are shown in Figure 2.12.

Enthalpy of Mixing. For liquids, which are nearly incompressible, $\Delta V \approx 0$ thus $\Delta H \approx \Delta E$.[32] The ΔH_m for each system can then be calculated according to[79]

$$\Delta H_{\mathrm{m}} = \frac{E_{tot,mix}}{N_i + N_j} - x_i \frac{E_{tot,i}}{N_i} - x_j \frac{E_{tot,j}}{N_j} \tag{2.21}$$

and the results are shown in Figure 2.13. The results do not agree very well with experiment for Ben + MOH, however they were not sensitive to changes in the charge distribution. It is reasonable to then assume that a portion of the errors in the enthalpy of mixing of all of the solutions containing aromatics (which are all small numbers less than 1 kJ/mol in magnitude) could be due to the imperfect enthalpy of vaporization (that is almost 1 kJ/mol too low) since all the aromatics use the C_{ar} atom type.

Isothermal Compressibility. Finite difference[79] isothermal compressibilities were obtained by performing additional 1 ns long simulations at a *p* of 250 and 500 bar. The values are consistently lower than the experimental κ_T; the cause of this deviation is, as yet, unknown.

Conclusions and Future Directions

We plan to compare additional properties of the simulated solutions to their experimental values *e.g.*, diffusivity and thermal expansitivity. However, it is gratifying to have developed non-polarizable models for benzene, toluene, phenol, *p*-cresol, pyrrole, and indole that quantitatively reproduce Kirkwood-Buff integrals and other physical and thermodynamic properties of solutions of these aromatics. The Lennard Jones parameters presented here for the benzene carbon and used for toluene, phenol, *p*-cresol, pyrrole, and indole, could also be used as the atom type for the description of other fully conjugated carbon atoms. We will use them again in Chapter 3 for the development of the KBFF's histidine charge distribution parameters.

Appendix A

Table A.1 Simulation Details for Ben + MOH, Tol + MOH, and Tol + Ben Solution Mixtures.

1	*2*	x_1	N_1	N_2	V (nm^3)	ρ (g/cm^3)	E_{pot} (kJ/mol)
Ben	MOH						
T_{sim}= 308 K							
		0.00	0	3177	224.22	0.754	-42.19
		0.25	611	1834	221.69	0.798	-33.97
		0.50	993	993	220.42	0.824	-25.79
		0.75	1255	418	219.54	0.843	-17.71
		0.90	1300	144	206.51	0.854	-13.00
		1.00	1444	0	217.24	0.862	-10.29
Tol	MOH						
T_{sim} = 318 K							
		0.00	0	3177	226.9	0.745	-41.64
		0.25	611	1834	238.62	0.801	-34.96
		0.50	993	993	246.22	0.832	-28.47
		0.75	1255	418	251.30	0.853	-22.17
		0.90	1300	144	239.60	0.862	-18.49
		1.00	1444	0	254.23	0.869	16.40
Tol	Ben						
T_{sim} = 313 K							
		0.00	0	1444	218.31	0.858	-9.62
		0.25	361	1083	227.18	0.862	-11.42
		0.50	722	722	235.86	0.865	-13.26
		0.75	1083	361	244.51	0.869	-15.12
		1.00	1444	0	253.14	0.873	-17.00

Table A.2 Simulation Details for Tol + PhOH, Tol + pCr, Pyr + MOH, and Ind + MOH Solution Mixtures.

1	*2*	x_1	N_1	N_2	V (nm^3)	ρ (g/cm^3)	E_{pot} (kJ/mol)
Tol	PhOH						
T_{sim} = 333 K							
		0.00	0	1425	215.34	1.034	-32.12
		0.25	361	1083	227.46	9.869	-28.16
		0.50	722	722	237.31	9.410	-23.14
		0.75	1083	361	247.61	8.971	-18.44
		0.90	1282	143	250.60	8.719	-15.90
		1.00	1444	0	258.18	8.558	-14.56
Tol	pCr						
T_{sim} = 333 K							
		0.00	1444	0	254.75	1.018	-40.38
		0.25	361	1083	254.60	0.981	-33.52
		0.50	722	722	264.75	0.907	-25.89
		0.75	361	1083	256.45	0.899	-20.26
		0.90	144	1300	257.43	0.873	-16.63
		1.00	0	1444	258.18	8.558	-14.56
Pyrr	MOH						
T_{sim}= 298							
		0.00	0	3177	224.22	0.754	-42.15
		0.25	611	1834	199.05	8.322	-26.07
		0.50	993	993	184.37	8.866	-9.33
		0.75	1255	418	174.15	9.305	+7.19
		1.00	1444	0	165.38	9.727	+23.10
Ind	MOH						
T_{sim}= 298							
		0.00	0	3177	224.22	0.754	-42.15
		0.25	611	1834	229.35	9.437	-45.68
		0.50	993	993	236.69	1.039	-48.74
		0.75	1255	418	242.36	1.099	-51.91
		1.00	1444	0	247.74	1.134	-54.61

Chapter 3:

The Kirkwood-Buff Derived Force Field for Histidine

Abstract

In a continuation of our efforts to develop an advanced, united atom, non-polarizable protein force field based upon the solution theory of Kirkwood and Buff, we present the non-bonded parameters for the histidine sidechain. These parameters were developed based upon studies of binary mixtures of pyrrole + methanol, pyridine + methanol, pyridine + water, histidine + water, and histidine monohydrochloride + water in an effort to ensure that the force field potentials appropriately balance solute-solvent, solute-solute, and solvent-solvent interactions.

Introduction

The accuracy of computer simulations is dependent upon simulation time and the quality of the force field (FF) used.[45,84] As discussed in Chapters 1 and 2, current FFs often yield simulations with too little solvation of peptide groups, producing self-aggregation of peptides beyond that which is experimentally observed. Since FFs describe the intermolecular and intramolecular interactions in a system and since the Kirkwood-Buff (KB) theory of solutions provides a way to quantify solute-solvent, solute-solute, and solvent-solvent interactions, a force field designed to reproduce the experimental KB integrals (KBIs) observed in solution mixtures of amino acid analogs

should create an improved balance between solvation and aggregation in computer simulations of biologically relevant systems.[45]

Indeed, this is the approach the Smith group has adopted, and a complete first version of this Kirkwood-Buff Force Field (KBFF) is lacking only in the non-bonded parameters for histidine (His). Here we present a model for the two tautomers of neutral His (HisA and HisB) and a model for doubly protonated His (HisH).

Approach

Experimental data for His and HisH mixtures is sparse because histidine is only sparingly soluble in water.[85] Thus, the non-bonded parameters are developed based upon the studies of pyrrole (Pyrr) + methanol (MOH) discussed in Chapter 2 and additional studies of pyridine (Pyrd) + MOH and Pyrd + water (HOH) over the full composition range, as well as His + HOH at 0.25 m_{His} and HisHCl + HOH at 0.3 and 0.6 m_{HisHCl}. The N2-H pyrrole charge distribution obtained from Chapter 2 for use in Trp was used as an initial guess of the charges on the acidic, protonated sp^2 nitrogen (N2) present in His and HisH. The charges on HisH were then adjusted to achieve better reproduction of the experimental KBIs. Pyridine was studied to model the charges of neutral histidine's basic, unprotonated sp^2 nitrogen (N_{py}) and the charges on the two C_{ar}-H that flank this nitrogen. The pyrrole and pyridine results were then combined to develop a crude approximation for the charges in His. His + HOH was then simulated to ensure that it did not aggregate excessively. The systems studied are illustrated in Figure 3.1.

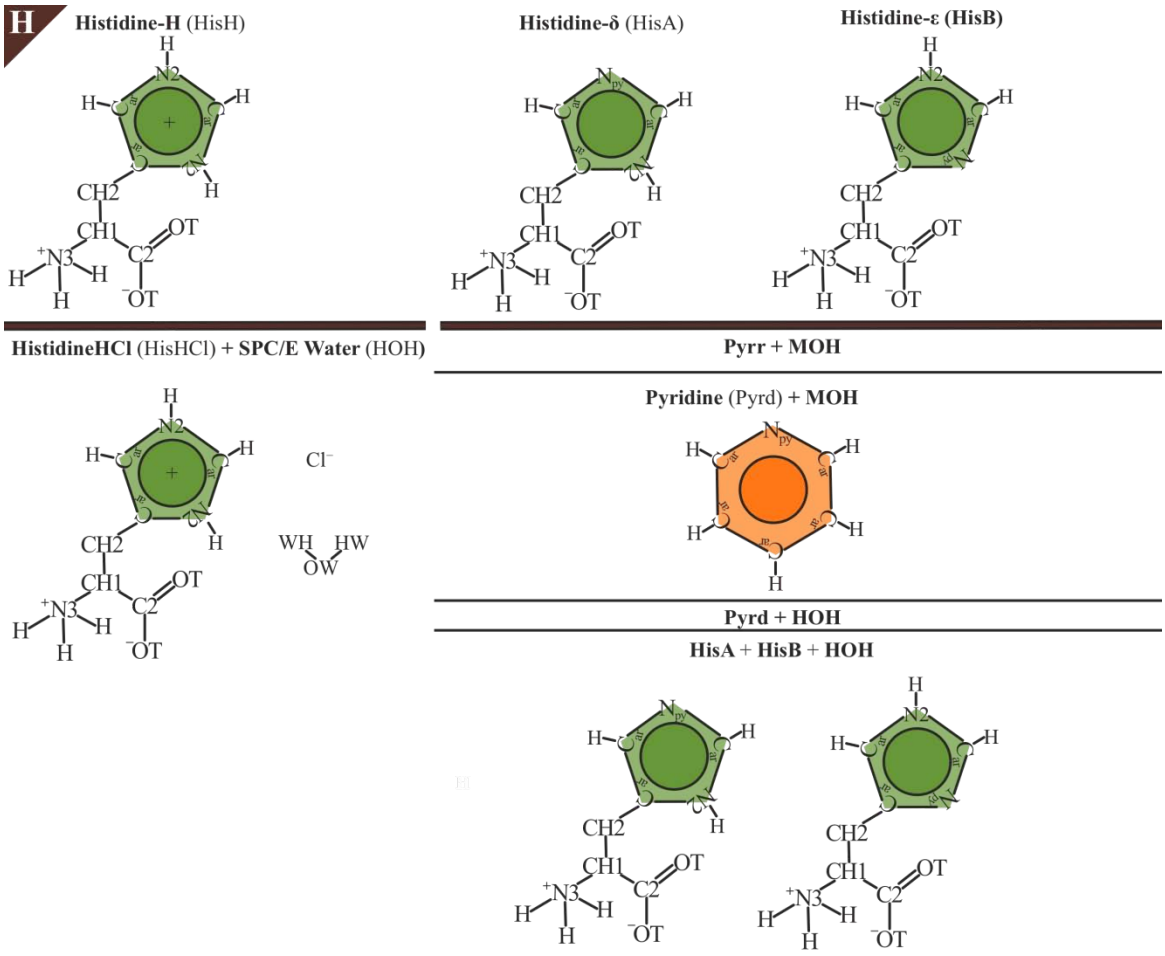


Figure 3.1: *Top:* HisH and the two tautomers of His with atom types labeled. *Middle/Bottom:* Systems studied to develop HisH and His charge distributions.

Theory

Following the Kirkwood-Buff inversion procedure developed by Ben-Naim,[36] we analyzed experimental data from the mixtures under study. The Kirkwood-Buff integrals (G_{ij}) are defined by Equation 1.17 and can be calculated from the experimental derivative of the chemical potential (μ_{ii}), partial molar volume ($\bar{V}_i$) and isothermal compressibility (κ_T) of the binary mixtures at constant pressure (p) and temperature (T) according to[35]

$$G_{ij} = RT - \frac{\bar{V}_i \bar{V}_j}{(1+ f_{cc})V_m} = kT\kappa_T - \rho \bar{V}_i \bar{V}_j / D \qquad (3.1)$$

and

$$G_{ii} = G_{ij} + \frac{1}{x_i}\left(\frac{\bar{V}_j}{1+f_{cc}} - V_m\right) = kT\kappa_T - \frac{1}{\rho_i} + \frac{\rho_j \bar{V}_j^2 \rho}{\rho_i D} \tag{3.2}$$

where R is the gas constant, x_i is the mole fraction of i, $V_m = V/(N_i + N_j)$ is the molar volume,

$$1+f_{cc} = 1+\left(\frac{\partial \ln f_c}{\partial \ln x_c}\right)_{p,T} = \beta\left(\frac{\partial \mu_c}{\partial \ln x_c}\right)_{p,T} \tag{3.3}$$

with (β=1/RT) and f_c equal to the cosolvent activity coefficient on the mole fraction scale with the pure cosolvent solution as the standard state, and

$$D = \frac{x_i}{kT}\left(\frac{\partial \mu_i}{\partial x_i}\right)_{p,T}. \tag{3.4}$$

Salts

Adjustments must be made when applying KB theory to electrolyte solutions due to the restrictions of electroneutrality.[17,35,37,46] For an aqueous solution of a completely dissociating salt (*s*) *i.e.*, water (*w*) + cation (*c*) + anion (*a*), electroneutrality demands the following conditions:[35]

(1) The charge around a solvent molecule must equal zero, so ρG_{wc} must equal ρG_{wa}, where ρ is the number density of the salt and $\rho = \rho_s = \rho_c = \rho_a$. Thus, $G_{wc} = G_{wa}$.

(2) There must be total conservation of charge around the *c* and *a*, leading to $1 + \rho G_{cc} = \rho G_{ac}$ and $1 + \rho G_{aa} = \rho G_{ac}$. Subsequently, $G_{cc} = G_{aa}$.

The above conditions make it impossible to obtain the individual cation or anion activity derivatives, but this problem can be avoided by assuming that the cations and anions are indistinguishable.[86] This allows the binary form of the KB equations that have been used repeatedly in this work for neutral solutions to also be used for electrolyte solutions.[36,86] Here we have adopted this approach to study HisHCl. The mole fraction of salt in water is then defined as $x_s = 2m_s/(2m_s + m_w)$,[35] where the multiplication of m_s by two indicates there are two ions obtained by the dissociation of one molecule of salt. In studies of HisHCl in water, we will be interested in three KBIs, $G_{\text{HisHCl-HisHCl}}$, $G_{\text{HisHCl-HOH}}$, and $G_{\text{HOH-HOH}}$. Expressions for $G_{\text{HisH-HisH}}$, $G_{\text{HisH-Cl}}$, and $G_{\text{Cl-Cl}}$ are also extractable from the conditions of electroneutrality, but were not investigated.[35]

Methods

Molecular Dynamics Simulations

Following the same procedure outlined in Chapter 2, all mixtures were simulated with classical MD techniques using the Gromacs program[54] in the *NpT* ensemble at 298 K and 1 atm. The Berendsen weak coupling technique[55] was used to modulate the temperature and pressure with relaxation times of 0.1 ps and 5.0 ps, respectively, with a 4.5×10^{-5} bar^{-1} compressibility. All bonds were constrained using the LINCS[56] algorithm and a relative tolerance of 0.1 fs. The bond constraints allowed for a 2 fs timestep for the integration of the equations of motion, which was performed using the Leap-Frog algorithm.[57] The particle-mesh-Ewald technique was used to calculate the electrostatic interactions,[58] employing a twin range cutoff of 1.2 and 1.5 nm for the real space electrostatic and van der Waals interactions, respectively. The neighbor list was updated

every 10 steps. Initial molecular configurations were generated from random in cubic boxes. The steepest descent method was then used to perform ≥1000 steps of energy minimization. This was followed by extensive equilibration that was continued until intermolecular potential energy contributions and rdfs displayed no drift with time. Configurations were saved every 1.0 ps for analysis. A summary of the simulations is presented in Appendix B.1. All mixture simulations were run for at least 20 ns to ensure that the time history of both the density and KBIs no longer displayed drifts. Bonded parameters (excluding the torsional parameters for His) were taken from the Gromos FF[6] and the Lennard Jones σ and ε parameters were taken from the KBFF including the fully conjugated carbon (C_{ar}) parameters developed in Chapter 2. For HisA, HisB, and HisH, the torsional parameters for all-rotatable bonds were those developed by the F. Chen and P. E. Smith.[30]

Experimental Sources for Composition and Activity Data

Pyrrole (1) + Methanol (2), 298 K. See Chapter 2.

Pyridine (1) + Methanol (2), 298 K: This mixture exhibits small, regular, negative deviations from ideality over the whole concentration range at 298 K and 1 bar.[87] In addition to $g^E < 0$, the V^E is also less than zero over the full composition range.[87 88] Nakanishi and Touhara *et al* obtained V^E and g^E from measurements of vapor pressures and densities using the Redlich – Kister power series equation.[87] The coefficients for g^E were given as A_1 = -.39702, A_2 = -0.26636, A_3 = 0.13640, A_4 = 138.76 and the coefficients for V^E were given as $A_1 = -1.931\times10^{-3}$, $A_2 = -0.876\times10^{-3}$, $A_3 = -0.176\times10^{-3}$. Nakanishi suggests that the small negative deviation is due to a favorable

enthalpy decrease by the formation of OH•••N hydrogen bonds and an unfavorable loss of entropy due to the orientation that is requisite for hydrogen bonding.

Pyridine (1) + Water (2), 298 K: This binary mixture exhibits $g^E > 0$ and $V^E < 0$ over the full composition range at 298 K and 1 bar with a maximum in density at $x_1 \sim 0.2$.[88] Following the same procedure as for Pyridine + Methanol and using experimental data from the same authors,[88] the coefficients for Equation 1.14 are $A_1 = +2.379$, $A_2 = -0.867$, $A_3 = +1.259$, $A_4 = -1.512$ for g^E and $A_1 = -2.927\times10^{-3}$, $A_2 = 0.969\times10^{-3}$, $A_3 = 0.849\times10^{-3}$ for V^E.[88] Nakanishi notes that a large, negative excess entropy term dominates over a negative enthalpy term to yield a positive g^E.[88]

Histidine (1) + Water (2), 298 K: Composition data was obtained from a density versus composition study at 298 K and 1 bar.[85] Activity coefficients were obtained from a vapor pressure osmometry study at 298 K and 1 bar with a maximum $m_{His} = 0.25$.[89] We studied only one composition, $m_{His} = 0.25$, composed of 15% HisA and 85% HisB corresponding to 11 HisA + 65 HisB + 16,764 HOH molecules in a box ~80 Å in length.[90-95]

Histidine Monohydrochloride (1) + Water (2): Composition data was approximated from $V_m = x_i V_i^0 + x_j V_j^0$. Activity coefficients were obtained from the same study as for His + Water at 298 K and 1 bar with a maximum $m_{HisHCl} = 0.63$.[89] We studied two compositions, $m_{HisHCl} = 0.3$ and 0.6.

Results

The final non-bonded parameters for all systems under study are presented in Table 3.1 and select properties of pure pyridine are presented in Table 3.2. The rdfs are

shown in Figures 3.2-3.4, KBIs in Figures 3.5-3.7, excess coordination numbers in Figures 3.8-3.10, and enthalpies of mixing for the two pyridine systems in Figure 3.11.

Pyridine. When considering liquid pyridine, the simulated density (Table 3.1) is slightly too high, but the enthalpy of vaporization is quite good. Other physical properties are currently under investigation including the dielectric constant as a function of composition, the diffusivity, and the thermal expansitivity.

In the comparison of the Pyrd + MOH and Pyrd + HOH systems, the first solvation shell of $g_{\text{HOH-HOH}}$ exhibits a higher peak than $g_{\text{MOH-MOH}}$ at a slightly smaller radius due to the smaller excluded volume of HOH when compared to MOH.

We were unable to reproduce both the Pyrd + MOH and Pyrd + HOH KBIs with a single charge distribution. We chose the charge distribution that best produced the Pyrd + HOH KBIs because HOH is the solvent of primary interest for biological simulations. The Pyrd + HOH KBIs are reproduced quite well, except at $x_{\text{Pyrd}} = 0.1$. Experimentally there is self-aggregation of water at this composition, however our simulations do not capture this aggregation. In contrast, the Pyrd + MOH system exhibits too much self-aggregation at each composition.

As would be expected based upon the results of the KBIs, the enthalpy of mixing for Pyrd + MOH shows that the mixing of the solute and solvent is too unfavorable, while the enthalpy of mixing for Pyrd + HOH is well reproduced across the full composition range.

Histidine. The His + HOH and HisHCl + HOH systems both exhibit similar $g_{\text{HOH-HOH}}$ rdfs with a sharp first solvation shell and well-structured second and third solvation shells. $g_{\text{His-HOH}}$ does not have any strong peak while $g_{\text{HisHCl-HOH}}$ has a sharp first

Table 3.1. Non-bonded Parameters for KBFF Models of Pyridine, Histidine, and HistidineH.

Model	σ (nm)	ε (kJ/mol)	q (|e|)
Pyridine			
H-C_{ar}-**N_{py}**	0.316	0.453	-0.575
H-**C_{ar}**-N_{py}	0.381	0.33	0.1575
H-C_{ar}-N_{py}	0.158	0.088	0.13
H-**C_{ar}**-C_{ar}	0.381	0.33	-0.13
H-C_{ar}-C_{ar}	0.158	0.088	0.13
Histidine			
N3	0.337	0.562	0.5
H	0.158	0.088	0
H	0.158	0.088	0
H	0.158	0.088	0
CH1	0.5019	0.0949	0.5
CH2	0.407	0.4105	0.13
C_{ar}	0.381	0.33	-0.13
N2	0.311	0.5	-0.45
H	0.158	0.088	0.45
C_{ar}	0.381	0.33	0.1575
H	0.158	0.088	0.13
C_{ar}	0.381	0.33	0.1575
H	0.158	0.088	0.13
N_{py}	0.316	0.453	-0.575
C2	0.336	0.33	1
OT	0.35	0.6047	-1
OT	0.35	0.6047	-1
HistidineH			
N3	0.337	0.562	0.5
H	0.158	0.088	0
H	0.158	0.088	0
H	0.158	0.088	0
CH1	0.5019	0.0949	0.5
CH2	0.407	0.4105	0.13
C_{ar}	0.381	0.33	-0.13
N2	0.311	0.5	0.15
H	0.158	0.088	0.35
Car	0.381	0.33	-0.13
H	0.158	0.088	0.13
C_{ar}	0.381	0.33	-0.13
H	0.158	0.088	0.13
N2	0.311	0.5	0.15
H	0.158	0.088	0.35
C2	0.336	0.33	1
OT	0.35	0.6047	-1
OT	0.35	0.6047	-1

Table 3.2. Comparison of Experimental and Simulated Properties of Liquid Pyridine at 298 K.

Model	T_{melt} [71]	ρ (g/cm^3)		ΔH_{vap} (kJ/mol)		κ_T (× 10^{+5}bar^{-1})	
		Exp[96]	Sim	Exp[97]	Sim	Exp	Sim
Pyridine	231	0.978	0.993	40.21	40.52	--	4.40

solvation shells. $g_{\text{His-HOH}}$ does not have any strong peak while $g_{\text{HisHCl-HOH}}$ has a sharp first solvation shell at a distance of ~0.25 nm with a peak slightly less than 2.25. $g_{\text{His-His}}$ and $g_{\text{HisHCl-HisHCl}}$ both show first solvation shells at a distance of ~0.7 nm with broad shoulders, however the $g_{\text{HisHCl-HisHCl}}$ peak at $m_{\text{HisHCl}} = 0.3$ is slightly greater than 3.5 and the $g_{\text{His-His}}$ peak is ~2. Even in a relatively large simulation box, with a length of ~80 Å, neither $g_{\text{His-His}}$ or the 0.3 m $g_{\text{HisHCl-HisHCl}}$ have reached unity. This is partly due to the very small number of His molecules in these compositions. Thus the errors in $G_{\text{His-His}}$ and $G_{\text{HisHCl-HisHCl}}$ will be high and they have not been reported due to their unreliability. In the His-HOH system, $G_{\text{His-HOH}}$ and $G_{\text{HOH-HOH}}$ (and likewise $N_{\text{His-HOH}}$ and $N_{\text{HOH-HOH}}$) reproduce the experimental values well. Although we have only considered one composition, we believe that this shows that we have a modest description of neutral histidine in our simulations. The HisHCl + HOH KBIs are also quite good, especially when considering the error on the $G_{\text{HisHCl-HOH}}$. Although we could not gain information on the $G_{\text{HisHCl-HisHCl}}$ due to unconverged $g_{\text{HisHCl-HisHCl}}$, a visual inspection of the mixture does not reveal excessive self-aggregation of the amino acid (Figure 3.12).

Conclusion

With only a small amount of experimental data available for the testing of a model for Histidine, we have been forced to design the Histidine FF based upon the side

chain's analogous atomic construction to pyrrole or pyridine, depending upon the protonation state of the nitrogen under consideration. Certainly this design is only a rough approximation, however a more elaborate approach would be futile without a richer source of experimental data for testing of the model.

Since the charge distribution has been carefully parameterized, even this simple design is able to quantitatively reproduce the Kirkwood-Buff integrals extracted from experimental data of the activity of His and HisHCl in water. Thus we are confident that we may proceed with the release of the first version of the Kirkwood-Buff Force Field to begin test simulations on peptides and small proteins.

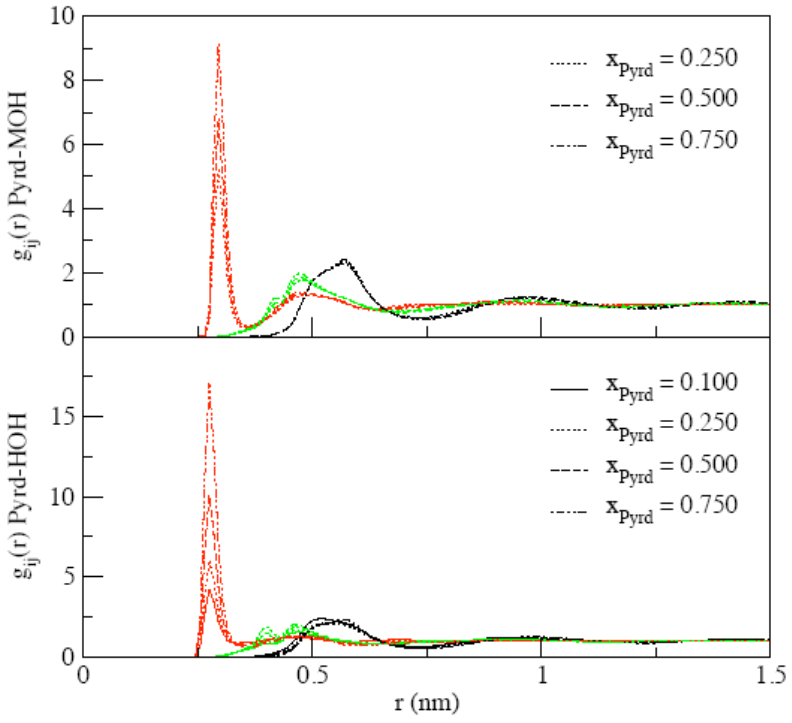


Figure 3.2. Center of mass based rdf for (top) pyridine (1) + methanol (2) and (bottom) pyridine (1) + water (2) as a function of distance. Black: g_{11}, Red: g_{22}, Green: g_{12}.

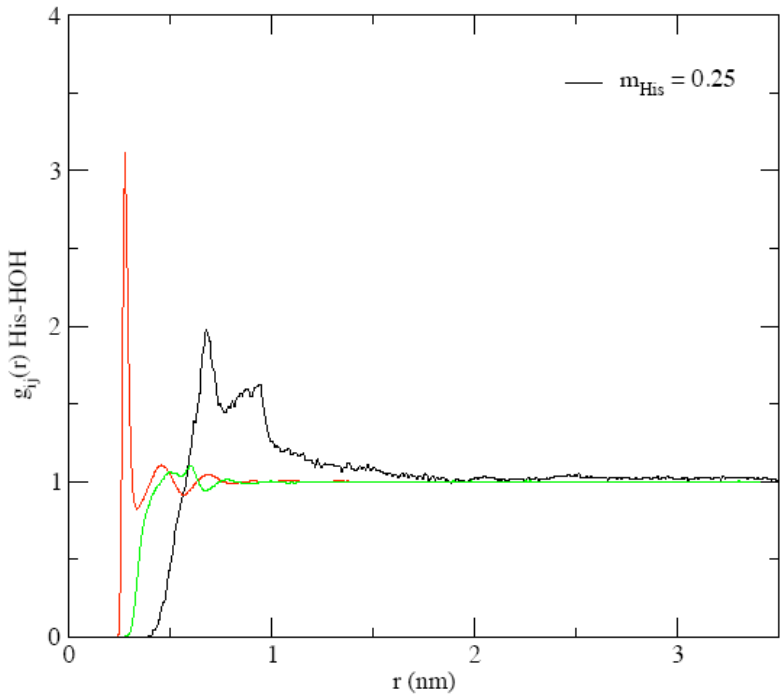


Figure 3.3. Center of mass based rdf for histidine (1) + water (2) as a function of distance. Black: g_{11}, Red: g_{22}, Green: g_{12}.

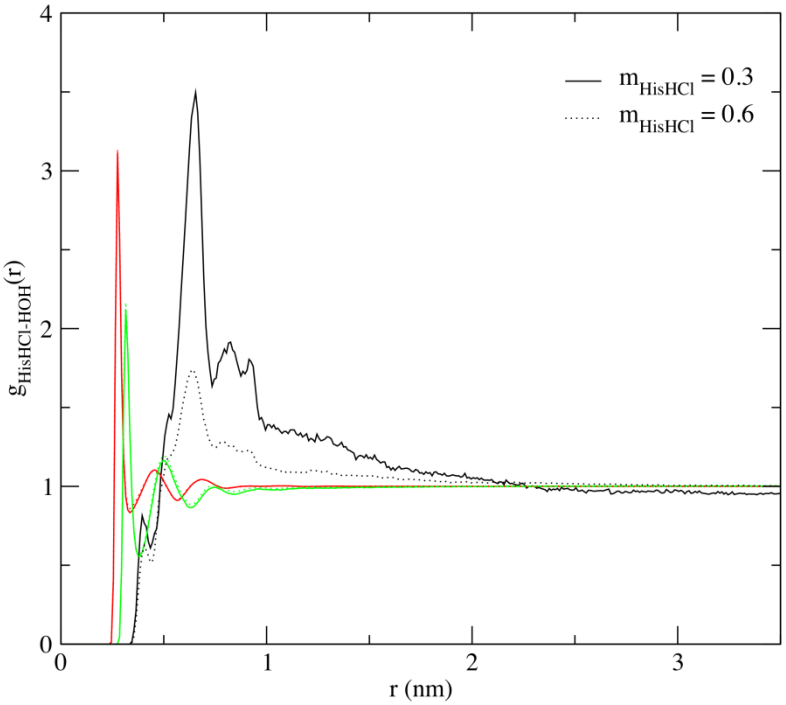


Figure 3.4. Center of mass based rdf for histidineHCl (1) + water (2) as a function of distance. Black: g_{11}, Red: g_{22}, Green: g_{12}.

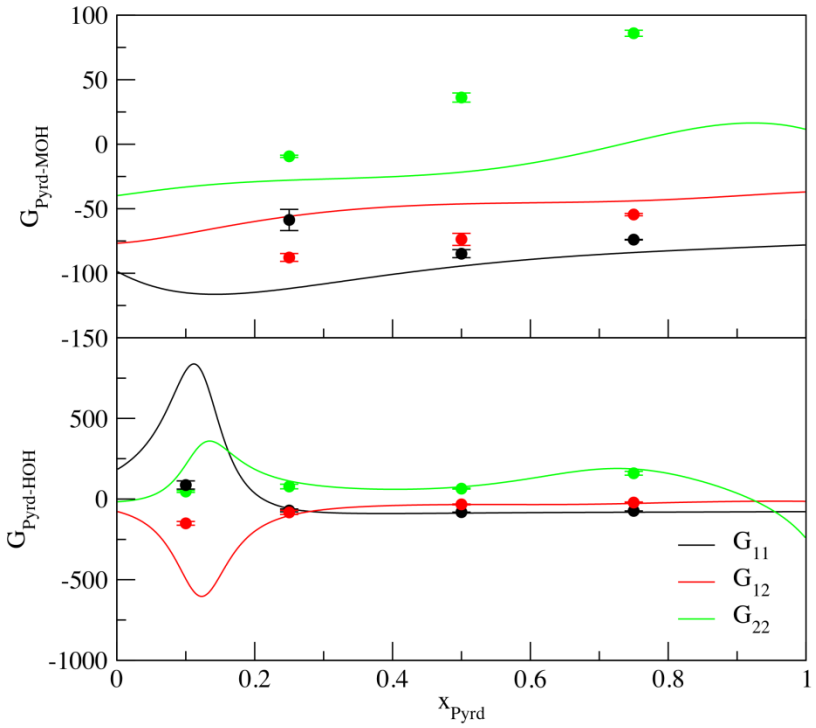


Figure 3.5. KBIs for (top) pyridine (1) + methanol (2) and (bottom) pyridine (1) + water (2) as a function of pyridine mole fraction. Line: experiment, Points: simulation.

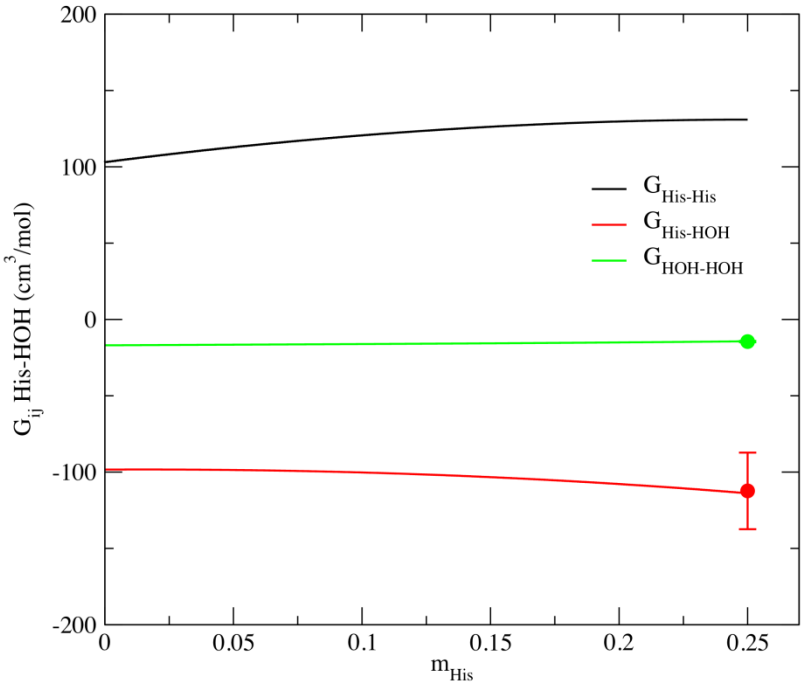


Figure 3.6. KBIs for histidine + water at 0.25 *m* His. Line: experiment, Points: simulation.

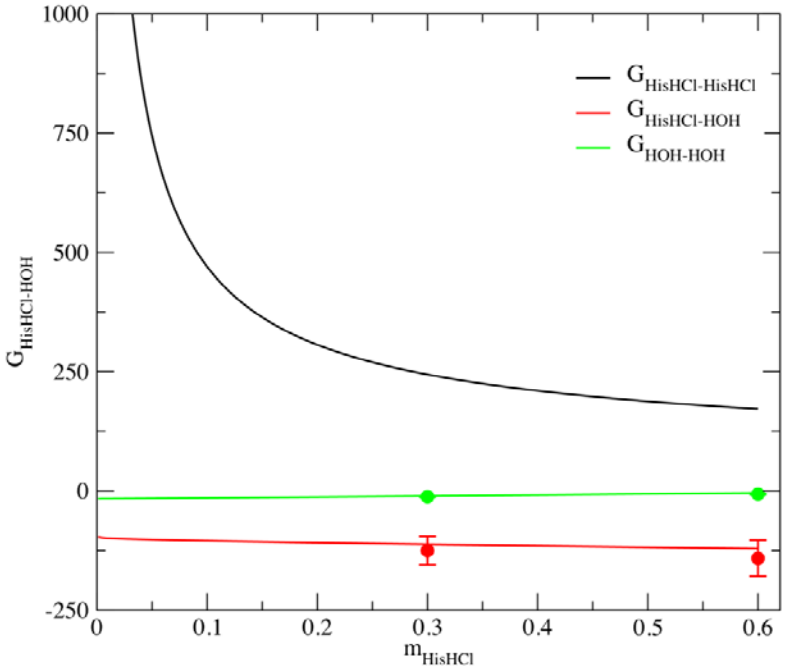


Figure 3.7. Kirkwood-Buff integrals for histidineHCl + water at 0.3 and 0.6 *m* HisHCl. Experiment (line) compared to simulation (points).

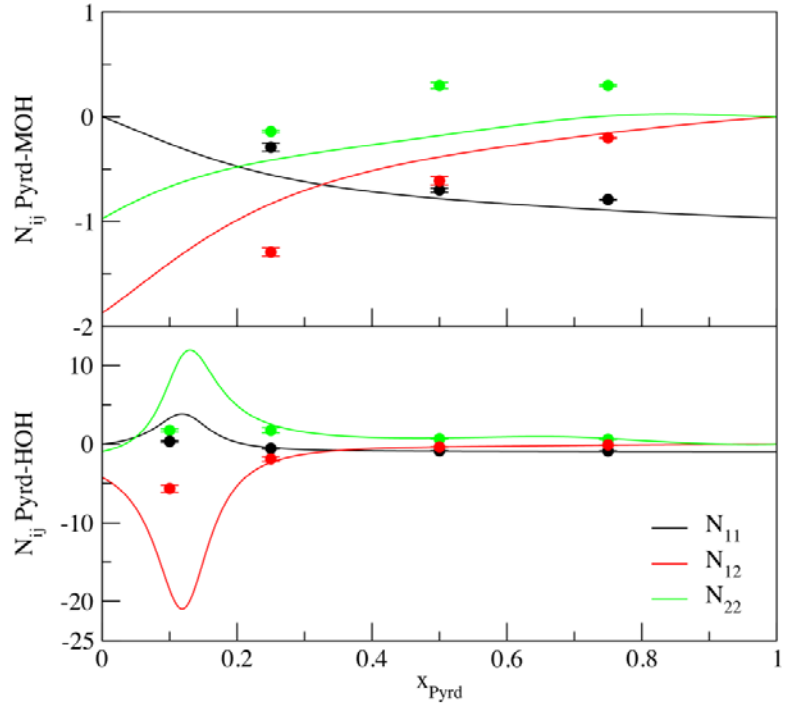


Figure 3.8. Excess coordination numbers for (top) Pyridine (1) + Methanol (2) and Pyridine (1) + Water (2) as a function of Pyridine mole fraction. Experiment (line) compared to simulation (points).

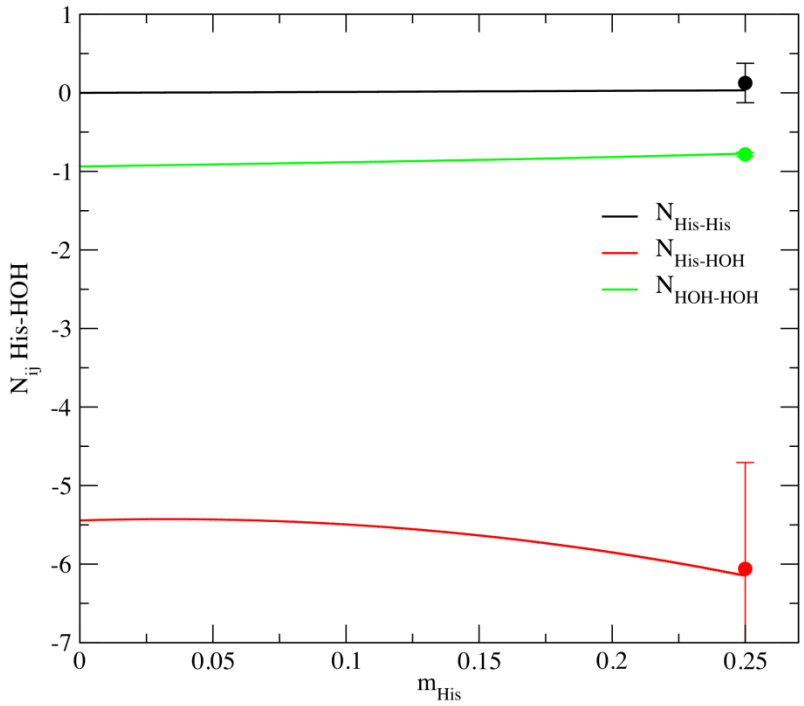


Figure 3.9. Excess coordination numbers for histidine + water at 0.25 *m* His. Experiment (line) compared to simulation (points).

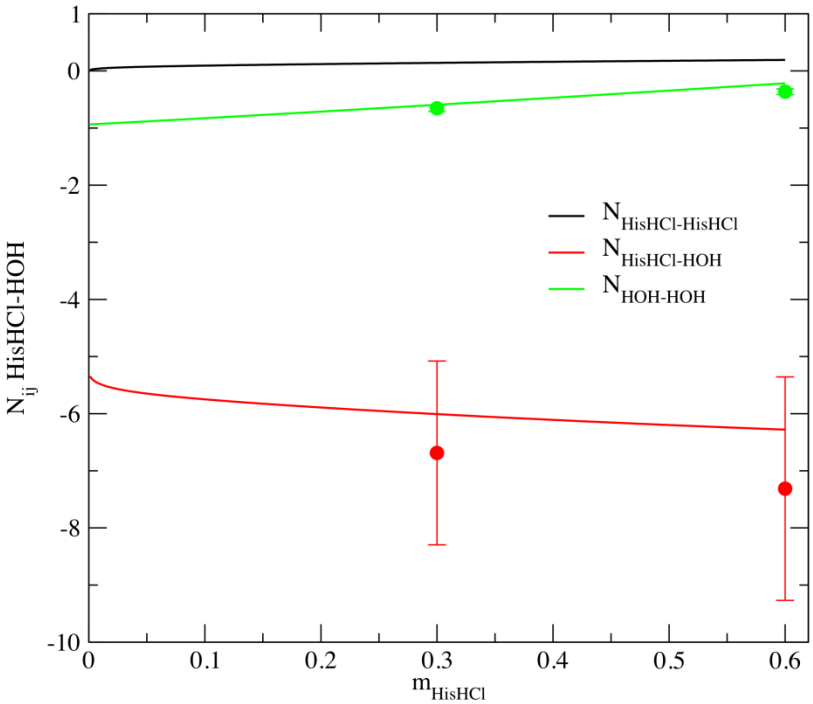


Figure 3.10. Excess coordination numbers for HistidineHCl + water at 0.3 and 0.6 *m* HisHCl. Experiment (line) compared to simulation (points).

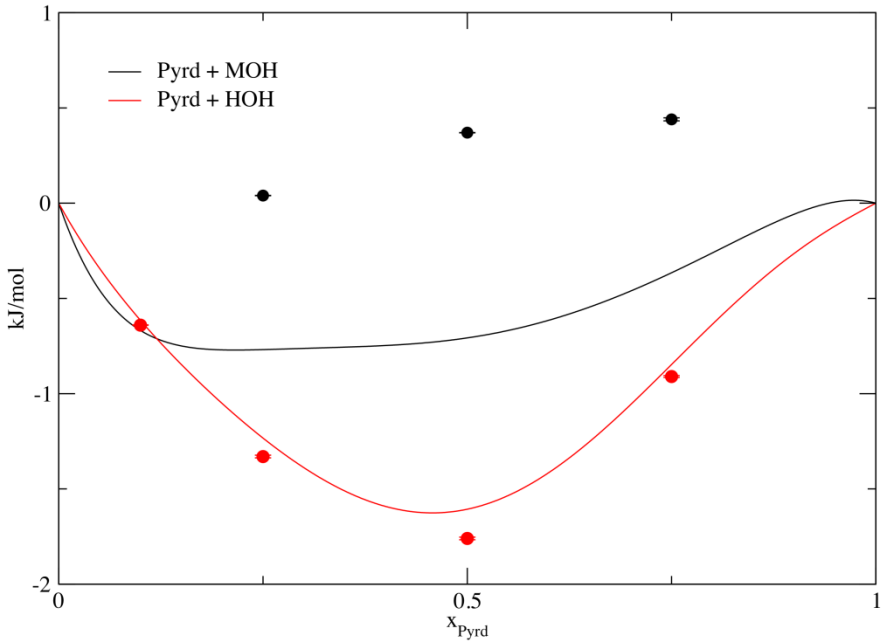


Figure 3.11. Enthalpy of Mixing for Pyridine + Methanol and Pyridine + Water. Experiment (line) compared to simulation (points).

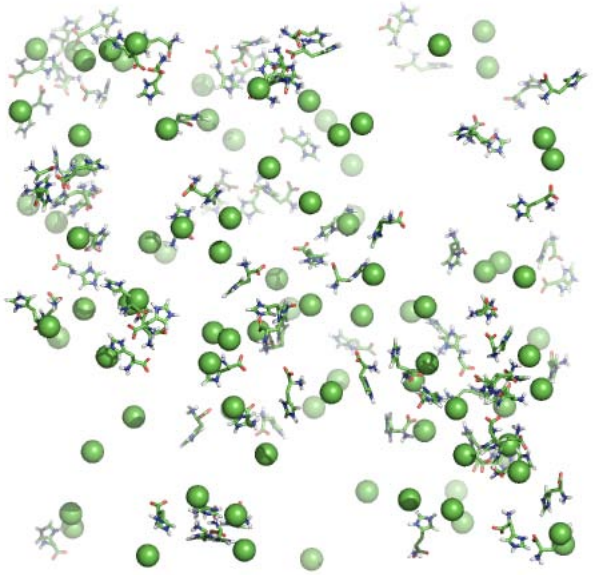


Figure 3.12. Snapshot of 0.3 *m* HisHCl with the final parameters. Water has been restricted from view. HisH are rendered as sticks and the chloride ions as spheres. Despite an unconverged HisHCl-HisHCl rdf, a visual inspection of the solutions confirms that there is not excessive self-aggregation of HisH, and that there is not complete ion pairing of $HisH^{+}Cl^{-}$.

Appendix B

Table B.1: Simulation Details Pyrd + MOH, Pyrd + HOH, HisHCl + HOH, and His + HOH Solution Mixtures.

1	*2*	**Composition**	N_1	N_2	V (nm^3)	ρ (g/cm^3)	E_{pot} (kJ/mol)
Pyrd	MOH	x_1					
T_{sim}= 298 K							
		0.00	0	3177	224.22	0.754	-42.15
		0.25	611	1834	207.96	0.855	-36.86
		0.50	993	993	220.42	0.916	-31.28
		0.75	1255	418	195.32	0.958	-25.96
		1.00	1444	0	191.07	0.993	-21.11
Pyrd	HOH	x_1					
T_{sim} = 298 K							
		0.00	0	7200	216.29	0.996	-46.70
		0.10	534	4824	213.17	1.006	-44.79
		0.25	966	2897	211.7	1.008	-41.64
		0.50	1255	1255	201.53	1.004	-35.69
		0.75	1444	81	204.75	0.996	-35.91
		1.00	1444	0	191.07	0.993	-21.11
HisHCl	HOH	m_1					
T_{sim} = 298 K							
		0.00	0	7200	216.29	0.996	-46.70
		0.30	90	16558	515.39	1.0167	-53.04
		0.60	174	16023	514.71	1.0388	-58.95
His	HOH	m_1					
T_{sim} = 298 K							
		0.00	0	7200	216.29	0.996	-46.70
		0.25	76	16764	515.79	1.0103	-50.34
			(11 HisA + 65 HisB)				

Summary and Future Direction

The overall message of this work is not meant to be that traditional force fields are fundamentally inferior to the KBFF or that they should no longer be used in MD simulations. It must be emphasized, however, that the choice of a FF should depend upon which properties it is that a researcher wishes to probe. We would be greatly amiss not to recognize that in many areas biological simulations using traditional FFs have certainly been successful at reproducing experimental properties and even providing new hypotheses for experimentalists to subsequently test.

Where we believe we have made an important improvement over traditional FFs is in the balance of intermolecular interactions. Thus, if a simulator seeks to correctly determine the equilibrium between a protein in the folded and unfolded state or to rank preferential interactions, then current FFs may be unreliable. It would be prudent for this simulator to compare results from other FFs with results obtained using the KBFF.

With the addition of the models for the aromatic amino acids described in this work, we have now finished the first version of the Kirkwood-Buff Force Field. Before we can fully gauge the improvements we have made, we need to perform tests on polypeptides and small proteins to see if the KBFF yields the experimental percentages of different secondary structures. Although we have already begun working toward this aim, many more tests need to be performed. Our next step is to release the first generation of the KBFF so that the community may help in this process.

In force field development, consistency is key. Lessons that are learned during the development process often cannot be immediately incorporated because they would

compromise internal consistency. For us, some of these lessons could potentially inspire a second version of the KBFF. A few of these lessons are briefly summarized below.

We have a united atom model, *e.g.*, a methyl group is represented by only one ball. We may be able to see improvements in our densities, and subsequently gain better agreement in the Kirkwood-Buff integrals, if we instead had an all-atom description of our systems.

The addition of new atom types could also create better results; however, this is a very time intensive task and it is usually desirable to keep the number of atom types as low as possible in a FF. Regardless, one case where it may need to be considered is hydrogen. Currently we have only two types of hydrogens, one type for those in water molecules (SPC/E water) and a second type to model the hydrogens in every other molecule.

Technical details, such as the twin-range cut-offs used in the calculation of non-bonded interactions, may need more careful consideration. We have chosen a coulomb cut-off distance of 1.2 nm here because the aromatics are relatively large molecules; however, previous KBFF models were created with this cut-off set equal to 0.8 nm. We may wish to revisit this aspect of our design in the future to create improved consistency.

Recently a temperature coupling procedure that rescales velocities was included in the MD program that we use for our simulations (Gromacs). Since it can be shown that this temperature coupling method produces the correct ensemble according to the Boltzmann distribution, whereas the temperature coupling technique we have employed here (Berendsen weak coupling) does not,[98] we may want to investigate if any differences are observed in the results with the velocity rescaling technique.

References

(1) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force-Field for Simulations of Proteins and Nucleic-Acids *Journal of Computational Chemistry* **1986**, *7*, 230.

(2) Roitberg, G. S.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California: San Francisco, **2008**.

(3) Arnautova, Y. A.; Jagielska, A.; Scheraga, H. A. A New Force Field (ECEPP-05) for Peptides, Proteins, and Organic Molecules *Journal of Physical Chemistry B* **2006**, 5025.

(4) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. Energy Parameters in Polypeptides. 7. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen-Bond Interactions, and Intrinsic Torsional Potentials for Naturally Occuring Amino-Acids *Journal of Physical Chemistry* **1975**, 2361.

(5) Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; Van Gunsteren, W. F. The GROMOS Software for Biomolecular Simulation: GROMOS05 *Journal of Computational Chemistry* **2005**, 1719.

(6) van Gunsteren, W. F.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Hochschulverlag AG an der ETH Zurich and BIOMOS b.v.: Zurich, Groningen, **1996**.

(7) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual*; BIOMOS b.v.: Groningen, **1987**.

(8) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins: Energy Minimizations for Crystals of Cyclic-Peptides and Crambin *Journal of the American Chemical Society* **1988**, *110*, 1657.

(9) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids *Journal of the American Chemical Society* **1996**, *118*, 11225.

(10) Cramer, C. *Essentials of Computational Chemistry: Theories and Models*, Second ed.; John Wiley & Sons Ltd.: West Sussex, **2004**.

(11) Okur, A.; Strockbine, B.; Hornak, V.; Simmerling, C. Using PC Clusters to Evaluate the Transferability of Molecular Mechanics Force Fields for Proteins *Journal of Computational Chemistry* **2003**, 21.

(12) Mu, Y. G.; Kosov, D. S.; Stock, G. Conformational Dynamics of Tri-Alanine in Water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS Force Fields to NMR and Infrared Experiments *Journal of Physical Chemistry B* **2003**, 5064.

(13) Mazur, A. K. Titration in Silico of Reversible B to A Transitions in DNA *Journal of the American Chemical Society* **2003**, 7849.

(14) Perera, A.; Sokolic, F. Modeling Nonionic Aqueous Solutions: The Acetone-Water Mixture *Journal of Chemical Physics* **2004**, 11272.

(15) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Mixtures of Urea and Water *Journal of Physical Chemistry B* **2003**, *107*, 3891.

(16) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Mixtures of Acetone and Water *Journal of Chemical Physics* **2003**, *118*, 10663.

(17) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Sodium Chloride in Water *Journal of Chemical Physics* **2003**, *119*, 11342.

(18) Auffinger, P.; Cheatham, T. E.; Vaiana, A. C. Spontaneous Formation of KCl Aggregates in Biomolecular Simulations: A Force Field Issue? *Journal of Chemical Theory and Computation* **2007**, 1851.

(19) Kang, M.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Amides *Journal of Computational Chemistry* **2006**, 1477.

(20) Kemp, D. A.; Gordon, M. S. An Interpretation of the Enhancement of the Water Dipole Moment Due to the Presence of Other Water Molecules *Journal of Physical Chemistry A* **2008**, 4885.

(21) Williams, J. H. The Molecular Electric Quadrupole Moment and Solid-State Architecture *Accounts of Chemical Research* **1993**, 593.

(22) Baker, C. M.; Grant, G. H. The Structure of Liquid Benzene *Journal of Chemical Theory and Computation* **2006**, 947.

(23) Kuyper, L. F.; Hunter, R. N.; Ashton, D.; Merz, K. M.; Kollman, P. A. Free-Energy Calculations on the Relative Solvation Free Energies of Benzene, Anisole, and 1,2,3-Trimethoxybenzene: Theoretical and Experimental Analysis of Aromatic Methoxy Solvation *Journal of Physical Chemistry* **1991**, 6661.

(24) Burgess, M. A. *Metal Ions in Solution*; John Wiley & Sons: New York, **1978**.

(25) Atkins, P. *Physical Chemistry*, Sixth ed.; W. H. Freeman and Company: New York, **1998**.

(26) Abraham, M. H. *Progress in Physical Organic Chemistry* **1974**, 1.

(27) Anslyn, E. V. D.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science Books: Sausalito, **2006**.

(28) Pauling, L. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: Introduction to Modern Structural Chemistry*, Third ed.; Cornell University Press: Ithaca, **1960**.

(29) Allen, L. C. Electronegativity is the Average One-Electron Energy of the Valence-Shell Electrons in Ground-State Free Atoms *Journal of the American Chemical Society* **1989**, 9003.

(30) Chen, F.; Smith, P. E. To be submitted. **2010**.

(31) Mansoori, G. A.; Matteoli, E. Fluctuation Theory of Mixtures (a Statistical Mechanical Background). In *Fluctuation Theory of Mixtures*; Matteoli, E. M.,

Mansoori, G. A., Eds.; Taylor & Francis Inc.: New York, **1990**; Vol. Second; pp 1.

(32) Seddon, J. M.; Gale, J. D. *Thermodynamics and Statistical Mechanics*; The Royal Society of Chemistry: Cambridge, **2001**; Vol. 10.

(33) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, **2000**.

(34) Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions. 1. *Journal of Chemical Physics* **1951**, 774.

(35) Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press, Inc.: New York, **2006**.

(36) Ben-Naim, A. Inversion of Kirkwood-Buff Theory of Solutions - Application to Water-Ethanol System *Journal of Chemical Physics* **1977**, *67*, 4884.

(37) Newman, K. E. Applications of Kirkwood-Buff Theory to Electrolyte Solutions. In *Fluctuation Theory of Mixtures*; Matteoli, E. M., Mansoori, G. A., Eds.; Taylor & Francis Inc.: New York, **1990**; Vol. Second; pp 373.

(38) Ben-Naim, A. Inversion of the Kirkwood-Buff Theory of Solutions and Its Applications. In *Fluctuation Theory of Mixtures*; Mansoori, G. A., Matteoli, E., Eds.; Taylor & Francis Inc.: New York, **1990**; Vol. Second; pp 211.

(39) *Fluctuation Theory of Mixtures*; Taylor & Francis Inc.: New York, **1990**; Vol. 2.

(40) Kang, M.; Smith, P. E. Kirkwood-Buff Theory of Four and Higher Component Mixtures *Journal of Chemical Physics* **2008**.

(41) Prausnitz, J. M.; Lichtenthaler, R. N.; deAzevedo, E. G. *Molecular Thermodynamics of Fluid-Phase Equilibria*, Second ed.; Prentice-Hall, Inc.: Englewood Cliffs, **1986**.

(42) Wilson, G. M. Vapor-Liquid Equilibrium. 11. New Expression for Excess Free Energy of Mixing *Journal of the American Chemical Society* **1964**, *86*, 127.

(43) Renon, H.; Prausnitz, J. M. Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures *AIChE Journal* **1968**, 135.

(44) Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. Recent Applications of Kirkwood-Buff Theory to Biological Systems *Cell Biochemistry and Biophysics* **2008**, *50*, 1.

(45) Weerasinghe, S.; Gee, M. B.; Kang, M.; Bentenitis, N.; Smith, P. E. Developing Force Fields from the Microscopic Structure of Solutions: The Kirkwood-Buff Approach. In *Modeling Solvent Environments*; Feig, M., Ed.; Wiley-VCH Verlag GmbH & Co. KBaA: Verlag, **2010**; pp 55.

(46) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for the Simulation of Aqueous Guanidinium Chloride Solutions *Journal of Chemical Physics* **2004**, *121*, 2180.

(47) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Methanol and Aqueous Methanol Solutions *Journal of Physical Chemistry B* **2005**, *109*, 15080.

(48) Bentenitis, N.; Cox, N. R.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Thiols, Sulfides, and Disulfides *Journal of Physical Chemistry B* **2009**, 12306.

(49) Gee, M. B.; Smith, P. E. To be submitted. **2010**.

(50) MacKerell, A. D.; Brooks, B.; Brooks, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. In *The Encyclopedia of Computational Chemistry*; John Wiley & Sons: Chichester, **1998**; Vol. 1.

(51) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins *Journal of Physical Chemistry B* **1998**, 3586.

(52) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program *Journal of Computational Chemistry* **2009**, 1545.

(53) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials *Journal of Physical Chemistry* **1987**, 6269.

(54) van der Spoel, D.; Lindahl, E.; Hess, B.; Kutzner, C.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tieleman, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual Version 4.0*; The GROMACS development team: Groningen, **2006**.

(55) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath *Journal of Chemical Physics* **1984**, 3684.

(56) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations *Journal of Computational Chemistry* **1997**, 1463.

(57) Hockney, R. W.; Goel, S. P.; Eastwood, J. W. Quiet High-Resolution Computer Models of a Plasma *Journal of Computational Physics* **1974**, 148.

(58) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - An Nlog(N) Method for Ewald Sums in Large Systems *Journal of Chemical Physics* **1993**, 10089.

(59) Sumer, K. M.; Thompson, A. R. Refraction, Dispersion, and Densities for Methanol Solutions of Benzene, Toluene, Aniline, and Phenol *Journal of Chemical and Engineering Data* **1967**, 489.

(60) Jorgensen, W. L.; Laird, E. R.; Nguyen, T. B.; Tirado-Rives, J. Monte-Carlo Simulations of Pure Liquid Substituted Benzenes with OPLS Potential Functions *Journal of Computational Chemistry* **1993**, 206.

(61) Jeffrey, G. A.; Ruble, J. R.; McMullan, R. K.; Pople, J. A. The Crystal Structure of Deuterated Benzene *Proceedings of the Royal Society of London (Series A: Mathematical, Physical, and Engineering Sciences)* **1987**, 47.

(62) Matteoli, E.; Lepori, L. Solute Solute Interactions in Water. 2. An Analysis Through the Kirkwood-Buff Integrals for 14 Organic Solutes *Journal of Chemical Physics* **1984**, 2856.

(63) Wanchoo, R. K.; Narayan, J. Excess Properties of (Methanol + Toluene or *p*-Xylene) Binary Liquid Mixture *Physics and Chemistry of Liquids* **1992**, 15.

(64) Nagata, I. Isothermal (Vapor + Liquid) Equilibria of (Methanol + Toluene) and of (Methanol + Acetonitrile + Toluene) *Journal of Chemical Thermodynamics* **1988**, 467.

(65) Kassmann, K. D.; Knapp, H. Vapor-Liquid-Equilibria for Binary and Ternary Mixtures of Benzene, Toluene, and Normal-Butyraldehyde *Berichte der Bunsen-Gesellschaft (Physical Chemistry Chemical Physics)* **1986**, 452.

(66) Klauck, M.; Grenner, A.; Taubert, K.; Martin, A.; Meinhardt, R.; Schmelzer, J. Vapor-Liquid Equilibria in Binary Systems of Phenol or Cresols Plus Water, Plus Toluene, and Plus Octane and Liquid-Liquid Equilibria in Binary Systems of Cresols Plus Water *Industrial and Engineering Chemistry Research* **2008**, 5119.

(67) Park, S. J.; Kim, H. H.; Han, K. J.; Won, D. B.; Lee, S. B.; Choi, M. J. Isothermal Vapor-Liquid Equilibria and Excess Molar Volumes for 2-Methyl Pyrazine (2MP) Containing Binary Mixtures *Fluid Phase Equilibria* **2001**, *180*, 361.

(68) Lide, D. R. *CRC Handbook of Chemistry and Physics*; CRC Press, Inc.: Boca Raton, **1990**.

(69) Buck, U.; Huisken, F. Infrared Spectroscopy of Size-Selected Water and Methanol Clusters *Chemical Reviews* **2000**, 3863.

(70) Ploetz, E. A.; Bentenitis, N.; Smith, P. E. Developing Force Fields from the Microscopic Structure of Solutions *Fluid Phase Equilibria* **2010**, *290*, 43.

(71) Lide, D. R. *CRC Handbook of Chemistry and Physics*, Eighty-ninth ed.; Taylor & Francis Group: Boca Raton, **2008**.

(72) Moravkova, L.; Wagner, Z.; Linek, J. (p, Vm, T) Measurements of (Octane + Benzene) at Temperatures From (298.15 to 328.15) K And at Pressures up to 40 MPa *Journal of Chemical Thermodynamics* **2008**, 607.

(73) Tardajos, G.; Aicart, E.; Costas, M.; Patterson, D. Liquid Structure and 2nd-Order Mixing Functions for Benzene, Toluene, Para-Xylene, and Normal-Alkanes *Journal of the Chemical Society-Faraday Transactions I.* **1986**, 2977.

(74) Cappelli, C.; Mennucci, B.; Cammi, R.; Rizzo, A. Quantum Mechanical Polarizable Continuum Model Approach to the Kerr Effect of Pure Liquids *Journal of Physical Chemistry B.* **2005**, 18706.

(75) Wypych, G. *Knovel Solvents-A Properties Database*; ChemTech Publishing: Toronto, New York, **2000**.

(76) Mooney, D. A.; Muller-Plathe, F.; Kremer, K. Simulation Studies for Liquid Phenol: Properties Evaluated and Tested Over a Range of Temperatures *Chemical Physics Letters* **1998**, 135.

(77) Yang, C. S.; Liu, Z. G.; Lai, H. X.; Ma, P. S. Excess Molar Volumes and Viscosities of Binary Mixtures of *p*-Cresol with Ethylene Glycol and Methanol at Different Temperature and Atmospheric Pressure *Journal of Chemical and Engineering Data* **2006**, 457.

(78) Faizullaev, S. F. Study of the Fine Structure of a Rayleigh Line and Determination of the Hypersound Propagation Rate in Some Associated Disubstituted Benzene Derivatives as a Function of Temperature *Opt.-Akust., Elektr. Magn. Issled. Konde* **1982**, 8.

(79) Walser, R.; Mark, A. E.; van Gunsteren, W. F.; Lauterbach, M.; Wipff, G. The Effect of Force-Field Parameters on Properties of Liquids: Parametrization of a Simple Three-Site Model for Methanol *Journal of Chemical Physics* **2000**, 10450.

(80) Smith, P. E.; van Gunsteren, W. F. Consistent Dielectric Properties of the Simple Point Charge and Extended Simple Point Charge Water Models at 277 and 300 K *Journal of Chemical Physics* **1994**, 3169.

(81) Rajendran, M.; Renganarayanan, S.; Madhavan, P. R.; Srinivasan, D. Effect of Dissolved Salts on the Heat of Mixing of Three Binary Systems *Journal of Chemical and Engineering Data* **1989**, 375.

(82) Letcher, T. M.; Prasad, A. K.; Mercerchalmers, J. Excess Molar Enthalpies and Excess Molar Volumes on Mixing an Aromatic Compound with an Alcohol at 298.15 K *South African Journal of Chemistry* **1991**, 17.

(83) Boublik, T.; Lam, V. T.; Murakami, S.; Benson, G. C. Excess Thermodynamic Functions of Cyclopentane-Carbon Tetrachloride Mixtures *Journal of Physical Chemistry* **1969**, 2356.

(84) van Gunsteren, W. F.; Dolenc, J.; Mark, A. E. Molecular Simulation as an Aid to Experimentalists *Current Opinion in Structural Biology* **2008**, *18*, 149.

(85) Yasuda, Y.; Tochio, N.; Sakurai, M.; Nitta, K. Partial Molar Volumes and Isentropic Compressibilities of Amino Acids in Dilute Aqueous Solutions *Journal of Chemical and Engineering Data* **1998**, 205.

(86) Chitra, R.; Smith, P. E. 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* **2002**, 1491.

(87) Nakanishi, K.; Ashitani, K.; Touhara, H. Excess Gibbs Free-Energy and Excess Volume of Methanol + Pyridine, + 2-Methylpyridine, and +2,6-Dimethylpyridine *Journal of Chemical Thermodynamics* **1976**, 121.

(88) Abe, J. I.; Nakanishi, K.; Touhara, H. Thermodynamic Properties of Aqueous-Solutions of Hydrophilic Compounds. 1. Pyridine and Methylpyridines *Journal of Chemical Thermodynamics* **1978**, 483.

(89) Tsurko, E. N.; Neueder, R.; Kunz, W. Water Activity and Osmotic Coefficients in Solutions of Glycine, Glutamic Acid, Histidine and their Salts at 298.15 K and 310.15 K *Journal of Solution Chemistry* **2007**, 651.

(90) Farr-Jones, S.; Wong, W. Y. L.; Gutheil, W. G.; Bachovchin, W. W. Direct Observation of the Tautomeric Forms of Histidine in N-15 NMR Spectra at Low Temperatures: Comments on Intramolecular Hydrogen Bonding and on

Tautomeric Equilibrium Constants *Journal of the American Chemical Society* **1993**, 6813.

(91) Henry, B.; Tekely, P.; Delpuech, J.-J. pH and pK Determinations by High-Resolution Solid-State C-13 NMR: Acid-Base and Tautomeric Equilibria of Lyophilized L-Histidine *Journal of the American Chemical Society* **2002**, 2025.

(92) Cheng, F.; Sun, H. H.; Zhang, Y.; Mukkamala, D.; Oldfield, E. Solid State C-13 NMR, Crystallographic, and Quantum Chemical Investigation of Chemical Shifts and Hydrogen Honding in Histidine Dipeptides *Journal of the American Chemical Society* **2005**, 12544.

(93) Huang, Z. J.; Lin, Z. J.; Song, C. Protonation Processes and Electronic Spectra of Histidine and Related Ions *Journal of Physical Chemistry A* **2007**, 4340.

(94) Blomberg, F.; Maurer, W.; Ruterjans, H. Nuclear Magnetic Resonance Investigation of N-15 Labeled Histidine in Aqueous Solution *Journal of the American Chemical Society* **1977**, 8149.

(95) Alei, M.; Morgan, L. O.; Wageman, W. E.; Whaley, T. W. pH-Dependence of N-15 NMR Shifts and Coupling Constants in Aqueous Imidazole and 1-Methylimidazole: Comments on Estimation of Tautomeric Equilibrium Constants for Aqueous Histidine *Journal of the American Chemical Society* **1980**, 2881.

(96) Jorgensen, W. L.; McDonald, N. A. Development of an All-Atom Force Field for Heterocycles. Properties of Liquid Pyridine and Diazenes *Theochem-Journal of Molecular Structure* **1998**, 145.

(97) Mayer, V. *IUPAC Data Series No. 32: Enthalpies of Vaporization of Organic Compounds*; Blackwell Scientific Publications: Oxford, **1985**.

(98) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling *The Journal of Chemical Physics* **2007**, *126*, 014101.

(99) Petrescu, V. La Variation de la Constante Dielctricque de Quelques Derives du Benzene et de la Naphthaline au Voisinage du Point de Fusion, **1940,** 233.