



PHD THESIS

Can approximate integral equation
theories accurately predict solvation
thermodynamics?

Author:
Maksim Misin

Supervisor:
Prof. Maxim V. FEDOROV
Dr. David S. PALMER

*A thesis submitted to the University of Strathclyde
in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy*

Department of Physics

January 5, 2017

Declaration of Authorship

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date: January 5, 2017

UNIVERSITY OF STRATHCLYDE

Abstract

Department of Physics

Doctor of Philosophy

Can approximate integral equation theories accurately predict solvation thermodynamics?

by Maksim Misin

The thesis focuses on the prediction of solvation thermodynamics using integral equation theories. Our main goal is to improve the approach using a rational correction. We achieve it by extending recently introduced pressure correction, and rationalizing it in the context of solvation entropy. The improved model (to which we refer as advanced pressure correction) is rather universal. It can accurately predict solvation free energies in water at both ambient and non-ambient temperatures, is capable of addressing ionic solutes and salt solutions, and can be extended to non-aqueous systems. The developed approach can be used to model processes in biological systems, as well as to extend related theoretical models further.

Acknowledgements

First of all, I would like to thank my supervisors, Prof. Maxim V. Fedorov and Dr. David S. Palmer, who made this Ph.D. an amazing experience. They offered me guidance and support, helping me not only to start but also to finish this project. I would like to acknowledge two visiting students Petteri A. Vainikka, and Samuel W. Coles, who turned out to be wonderful collaborators. Various Strathclyde postgraduate students: Ivor Kresic, Samiul M. Ansari, Benjamin R. Smith, Sean O'Connor, Rosemary Orr, helped me through stimulating discussions. My special thank you goes to Dr. Vladislav Ivanistsev and his group at the University of Tartu, who provided a wonderful and supportive environment during my time there.

I would also like to thank my mother and grandmother, without their love and support I would not have come this far. Most of all, I would like to thank my loving, encouraging, and patient Anna, whose support during this Ph.D. is so appreciated. Thank you.

Contents

Declaration of Authorship	i
Abstract	ii
Acknowledgements	iii
List of Figures	vi
1 Introduction	1
1.1 Structure	4
I Theoretical background	6
2 Solvation thermodynamics	7
2.1 Classical partition function	7
2.2 Free energies	9
2.3 Chemical potential	10
2.4 Solvation free energy	15
2.5 Decomposing solvation free energy	17
2.6 Numerical experiments	20
2.7 Ionic solvation	23
2.8 Applications	25
3 Theory of simple liquids	28
3.1 Particle densities and distributions	28
3.2 Free energy functionals	31
3.3 Functional derivatives and correlations	33
3.4 Ornstein-Zernike and mixtures	37
3.5 Linear response	40
3.6 Connection to experiment	43
4 Theory of molecular liquids	46
4.1 Hypernetted-chain approximation	46
4.2 Molecular liquids	52
4.3 Intramolecular correlation function	55

4.4	1D-RISM	58
4.5	3D-RISM as a density functional theory	61
4.6	3D-RISM as an integral equation theory	65
4.7	Partial molar volume	68
4.8	Solvation free energy from 3D-RISM	70
II	Results and Discussion	73
5	Pressure corrections	74
5.1	Dissecting HNC free energies	74
5.2	Hydrophobic solvation	78
5.3	Pressure corrections for solvation free energy	82
5.4	Free energy of charging	86
5.5	Effect of corrections on solvation thermodynamics	89
6	Applications	96
6.1	Neutral molecules	96
6.2	Predicting the temperature dependence of ΔG	102
6.3	Ionic solvation	107
7	Beyond pure water	113
7.1	Setschenow constant	113
7.2	Predicting solvation free energy in salt solutions	115
7.3	Non-aqueous solvents and the corresponding state principle	120
7.4	Solvation of model solutes in Lennard-Jones fluids	123
7.5	Comparison with experimental values	125
8	Conclusion	129
A	Methodology	131
A.1	1D-RISM	131
A.2	3D-RISM	132
A.3	Molecular dynamics	133
A.4	Force fields and geometry	134
B	Additional Results	135
B.1	Partial molar volume and grand canonical ensemble	135
B.2	Extra figures and tables	136
	List of original publications	141
	Bibliography	142

List of Figures

1.1	The solvation free energy is equal to the reversible work required to bring molecule from (ideal) gas phase to solvent. Alternatively, it can be computed from its equilibrium concentration ratio in two phases.	2
1.2	A particle with a spherically symmetric potential $\phi(r)$ is placed at the origin. The surrounding solvent rearranges, forming multiple solvation shells. The dependence of its average density $\rho(r)$ can be obtained using classical density functional theories. The units are arbitrary.	3
2.1	A schematic drawing demonstrating the dependence of electrostatic potential on the distance from the water-air interface.	24
2.2	Thermodynamic cycle relating solvation free energy of an ion A^- to the solvation free energies of acid AH , its dissociation free energies ΔG_r^o in both gas and solvent, and solvation free energy of the proton. Since free energy of the system is independent of the path, we have $\Delta G_s^*(AH) + \Delta G_{r,s}^o(AH) = \Delta G_{r,g}^o(AH) + \Delta G_s^*(A^-) + \Delta G_s^*(H^+)$, from which $\Delta G_s^*(A^-)$ can be deduced.	24
3.1	Radial distribution function of Ar near its boiling point obtained via molecular dynamics simulations using parameters from Ref. 73.	30
3.2	Total correlation function (left) and direct correlation function (right) of liquid argon near its boiling point. The correlation functions were extracted from molecular dynamics simulations performed with parameters taken from Ref. 73.	35
3.3	Fourier transforms of the total correlation function (left) and the direct correlation function (right) from figure 3.2.	38
3.4	Comparison of experimental and theoretical structure factors (left) and radial distribution functions (right) for argon at 85K (close to its boiling point). Experimental data is taken from Ref. 86. Theoretical predictions are obtained using molecular dynamics, with simulation parameters taken from Ref. 73.	44
4.1	Liquid argon local density near hard wall. Both singlet HNC calculations (red line) as well as molecular dynamics simulations (blue line) are performed at standard pressure and 85 K. The details of the calculations can be found in appendix A.	49
4.2	Comparison of liquid argon radial distribution functions (left figure) and direct correlation functions (right figure) predicted by molecular dynamics simulation (blue lines) and HNC calculation (red lines). For both simulation and theory interactions between argon atoms was approximated using Lennard-Jones potential from Ref. 73.	51

4.3	Water intramolecular site-site correlation function (left figure) and its functional inverse (right figure).	57
4.4	Fourier transforms of water intramolecular site-site correlation function (left figure) and its inverse (right figure).	58
4.5	Water oxygen-oxygen radial distribution functions (left figure) and partial structure factors (right figure) from RISM and molecular dynamics at standard conditions.	60
4.6	Distribution of hydrogens around caffeine, with regions of lower density coloured in yellowish colour and regions of higher in blue.	63
4.7	Distribution of water oxygens (blue) and hydrogens (red) in a large box with $u_O = \infty$ for $z < 20 \text{ \AA}$. You can see that some hydrogens can be found arbitrarily far away from the oxygens.	64
4.8	Solvation free energies of Lennard-Jones solutes with σ given on x axis and $\epsilon = 4 \text{ kcal/mol}$. On y axis you can see the difference in free energies by PSE-n and HNC closures.	67
4.9	A schematic depiction of Gibbs dividing surface for the case of planar interface. The surface, shown as a dashed line divides solvent local density (blue line) in such way that excess of solvent to the left of the surface (region shaded in red) is equal to the depletion of solvent to the right of the surface (region shaded in green).	69
4.10	Comparison of molecular dynamics and PSE-3 Lennard-Jones components of hydration free energies (left figure) and electrostatic components (right figure) for 504 molecules from Mobley dataset, which is discussed in detail in section 6.1. Molecular dynamics results are taken from Ref. 13. All values are in kcal/mol	71
5.1	Dependence of surface energy of hard spheres in water depending on their radius. Monte Carlo (MC) results are taken from Ref. 131.	80
5.2	Both figures demonstrates hard sphere–water oxygen radial distribution functions for spheres of various radii. Figure (a) is taken from Ref. 137 and was obtained using Monte Carlo simulations. Radial distribution functions in figure (b) were calculated using 3D-RISM/HNC.	83
5.3	This figure mirrors figure 5.1, except now instead of information theory and HNC we compare results by PC and PC+ models.	84
5.4	Solvation free energies of Lennard-Jones solutes with different radii and $\epsilon = 0.125 \text{ kcal/mol}$ in water. The estimate of MD solvation free energies for larger radii was done using equation 5.28.	85
5.5	Electrostatic potential inside the soft solute with $\sigma = 3.8 \text{ \AA}$ and $\epsilon = 0.125 \text{ kcal/mol}$ depending on its charge.	87
5.6	Top figures: charging free energy of Lennard-Jones sphere (left) and Lennard-Jones dipole (right) in water. The dipole consists of two Lennard-Jones spheres, separated by 2 \AA . Both the sphere and dipole sites have $\sigma = 3.8 \text{ \AA}$ and $\epsilon = 0.125 \text{ kcal/mol}$. Bottom figures: difference in charging free energy predictions between 3D-RISM models and molecular dynamics.	88
5.7	Dependence of free energy, entropy and enthalpy of linear alkanes on their surface area (number of atoms). The molecular dynamics results are taken from Ref. 151.	92

5.8	Solute-solvent components of alkanes solvation free energy. The definition of the quantities can be found in the main text. The molecular dynamics results are taken from Ref. 151.	93
5.9	Correlation between solvent reorganization energy and solute-solvent interaction energy for the first 6 linear alkanes. The units are in kcal/mol. . . .	95
6.1	Comparison of PC (top) and PC+ (bottom) free energies with molecular dynamics results, obtained by Mobley et al ¹³ . In the first column we compare nonpolar (Lennard-Jones) components, in the second, electrostatic contributions, and in the third, total hydration free energies. All values are in kcal/mol.	98
6.2	Hydration free energies by different models, compared to experimental data. Both molecular dynamics results and experimental values come from Ref. 13. All values are in kcal/mol.	99
6.3	The influence of different force fields on the accuracy of PC+ results when compared to experiment. All values are in kcal/mol.	101
6.4	Comparison experimental and computed hydration free energies at 298 K. The data is taken from the Chamberlin dataset. All values are in kcal/mol.	103
6.5	Comparison of predicted and experimental solvation entropies and heat capacities, evaluated using equation 6.1. Both entropy and heat capacity are in units of cal/mol/K.	104
6.6	Root mean square error of solvation free energy of 3D-RISM models on Chamberlin dataset depending on temperature.	105
6.7	Correlations between solute-solvent and solvent reorganization energies. R^2 stands for coefficient of determination. All values are in kcal/mol. . . .	106
6.8	Comparison of 3D-RISM/PC+ ionic hydration free energies with molecular dynamics results (left) and experimental values (right). Galvani potential is not taken into account. All values are in kcal/mol.	108
6.9	Schematic demonstration of interfacial potential jumps in solution.	108
6.10	The same comparison as in figure 6.8, but with model Galvani potentials taken into account. All values are in kcal/mol.	111
6.11	Left figure: 3D-RISM/PC+ ionic hydration free energies compared to experiment. The force field is OPLS/CM5. Right figure: SMD ionic hydration free energies; electronic structures are computed with MG3S/M06-2X level of theory. The values are in kcal/mol.	111
7.1	The change in hexanol hydration free energy depending on NaCl concentration. The experimental trend is estimated from the experimental Setschenow's constant.	114
7.2	Accuracies of Setschenow's constants predictions by COSMO-RS (left) and pp-LFER model (right). Both experimental and computed values are taken from Ref. 178. All values are in $l\ mol^{-1}$	115
7.3	Setschenow's constants by 3D-RISM/PC+ with Dang salt model compared to experimental measurements by Endo et al. Results on the left and right figures are obtained with OPLS/CM5 and GAFF/AM1-BCC force fields respectively. B-A stands for bisphenol A. The values are in l/mol	119

7.4	The optimised geometries of bisphenol A with and without Na^+ ion. The distances, shown in Angstroms, are measured between the centres of benzene rings and the ion. SMD model was used to take into account solvent effects.	120
7.5	The basic idea behind coarse graining used for 3D-RISM calculations. The real liquids are approximated as spheres interacting via Lennard-Jones potentials with parameters deduced using equations shown on the figure above.	122
7.6	Solvation free energy per unit area for solutes of radii r in the Lennard-Jones liquid with $T^* = 0.701$ and $\rho^* = 0.843$. LCW results are taken from Ref. 137.	124
7.7	Calculated Lennard-Jones solutes solvation free energies against MD data. Dichloroethane and CCl_4 models are defined in table 7.3. All values are in kcal/mol.	125
7.8	Computed solvation free energies of a number of compounds in apolar solvents (as defined in table 7.4) against experimental data. COSMO-RS results are taken from Ref. 219. All values are in kcal/mol.	126
7.9	Computed solvation free energies of a number of compounds in polar solvents (as defined in table 7.4) against experimental data. COSMO-RS results are taken from Ref. 219. All values are in kcal/mol.	127
A.1	We performed two sets of calculations on the Mobley dataset (described at the beginning of section 6.1) using standard and fine grid settings. The comparison of free energies (kcal/mol) is shown on the left, and in run time on the right figure.	133
B.1	Site-site radial distribution functions of SPC/E water (cSPC/E in case of DRISM calculations).	136
B.2	The figure mirrors figure 6.1, but compares UC and NgB models, instead of PC and PC+.	137
B.3	The figure mirrors figures 5.7 and 5.8, but instead of pressure corrected models uses HNC results. The molecular dynamics results are taken from Ref. 151.	138
B.4	Root mean square error of solvation free energy of 3D-RISM models on the Chamberlin dataset. Unlike figure 6.6, here we do not vary temperature in 3D-RISM calculations, using the value of ΔG at 298 K for all comparisons.	139
B.5	Comparison of computed (3D-RISM/HNC) and experimental solvation free energies in polar and apolar solvents (as defined in section 7.5). All values are in kcal/mol.	140

Chapter 1

Introduction

What is the amount of reversible work needed to bring a molecule from gas phase to solvent? It turns out that this question is not just a matter of scientific curiosity. Accurate predictions of phenomena such as solubility, partition coefficients, substrate binding, acid dissociation constants, all in one way or another depend on how accurately we can measure or predict this quantity¹⁻¹¹. The aim of this thesis is to develop a theory-based computational method for computing this amount of work.

Formally, the problem is to predict the free energy change, ΔG_A , occurring upon transfer of molecule from gas phase to liquid A ¹². This free energy tells us the probability P to find a molecule in a specific phase compared to gas: $P_A/P_{gas} = \exp[-\Delta G_A/(RT)]$, where T is the temperature and R is the gas constant.

There are multiple factors which make prediction of solvation free energies difficult^{13,14}. First of all, a good approximation of intermolecular potentials between molecule (solute) and liquid (solvent) is required¹⁵. Second, one has to take into account all possible conformations in which solute can exist in the solution and the gas phase¹⁶. Finally, and most importantly, one also has to consider all possible configurations of solvent molecules around the solute; after all, the affinity of a molecule to the phase is determined not only by the solute-solvent interactions, but also by the solvent-solvent ones¹⁷⁻¹⁹. The nature of all involved forces is quantum-mechanical; this puts *ab initio* prediction of ΔG_A into the category of (practically) unsolvable quantum many-body problems.

To make progress, we need to make some approximations. To an extent, each approximation we make is a trade-off between the speed of the model and its domain of applicability.

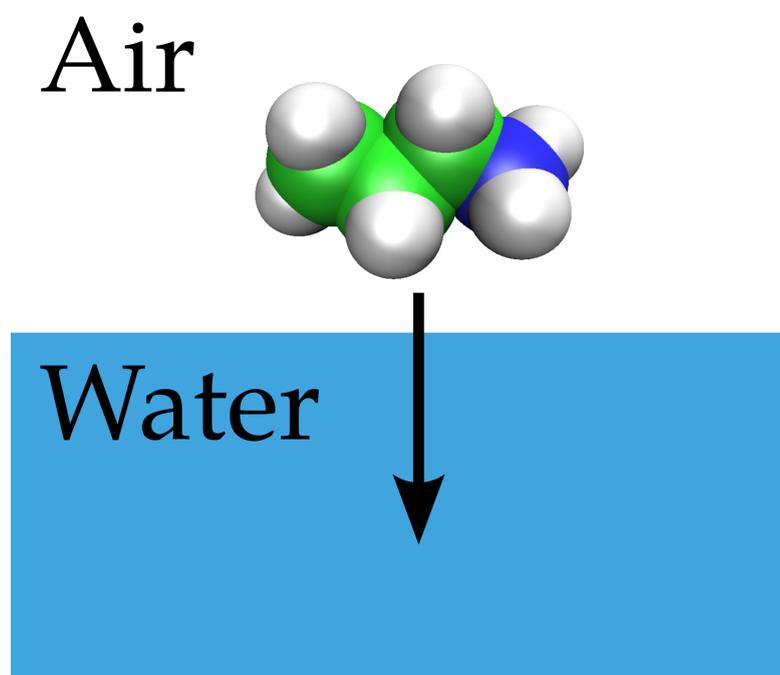


FIGURE 1.1: The solvation free energy is equal to the reversible work required to bring molecule from (ideal) gas phase to solvent. Alternatively, it can be computed from its equilibrium concentration ratio in two phases.

A classical molecular dynamics simulation, which uses only a few approximations, is a very general tool that can be applied to systems where quantum effects are not relevant to the motion of particles²⁰. At the same time, methods based on statistical learning, such as quantitative structure-property relationships (QSPR), are less generally applicable and are usually limited to systems and compounds that are sufficiently similar to the training data^{21,22}. As one would expect, molecular dynamics uses a large number of computational resources, while QSPR calculations are practically effortless.

A family of methods based on classical density functional theory and related integral equation theories offer an attractive balance between speed and generality^{6,23–25}. The idea behind these methods is to ignore unimportant degrees of freedom in a solvent and view it as a local density field $\rho(\mathbf{r})$. In the absence of an external potential, the solvent will be homogeneous with the value of local density in each point being equal to the bulk number density $\rho(\mathbf{r}) = \rho$. However, bringing a solute molecule in a solvent introduces an external field $\phi(\mathbf{r})$, which breaks the symmetry. As a result, solvent re-distributes itself around a solute, giving rise to a new density distribution $\rho(\mathbf{r}; \phi)$ that is uniquely determined by the external potential. Moreover, a new density field will be such that the

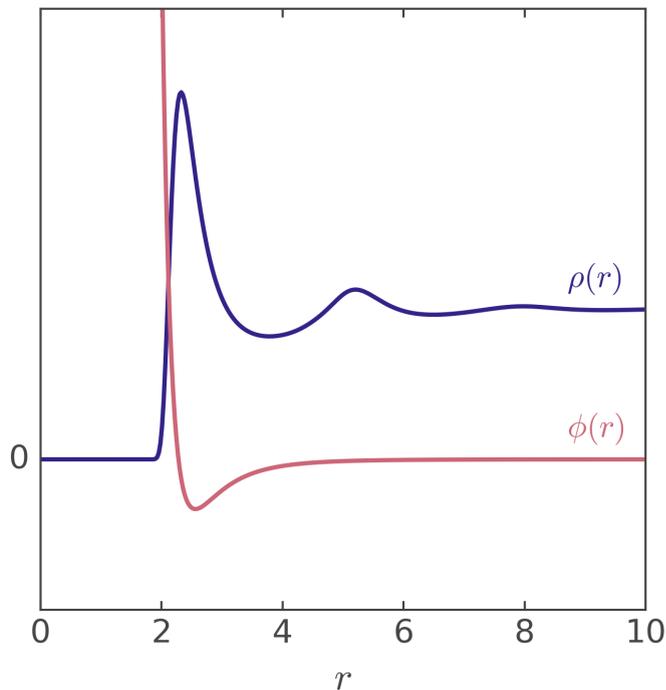


FIGURE 1.2: A particle with a spherically symmetric potential $\phi(r)$ is placed at the origin. The surrounding solvent rearranges, forming multiple solvation shells. The dependence of its average density $\rho(r)$ can be obtained using classical density functional theories. The units are arbitrary.

total free energy of the system will be minimised. The last two statements are the key results of density functional theory, known as the Hohenberg–Kohn–Mermin theorems²⁶.

Similarly to the electronic version of the density functional theory, the classical functional that relates the system’s density field $\rho(\mathbf{r})$ to its free energy is unknown. Therefore, one has to use approximations, tailored to a given problem. Moreover, often, such approximations lead to non-trivial results and are best understood through the applications of theory to specific problems²⁷.

In this thesis, we focus on one of the most popular approaches, called three-dimensional reference interaction site model or 3D-RISM^{6,28–30}. It is commonly categorised as an integral equation theory we will see that it is naturally derived and analysed from the viewpoint of classical density functional theory. The approach owes its popularity to its simplicity, robustness, speed, as well as the fact that it can be applied to biologically relevant solutions such as water with dissolved electrolytes. In 3D-RISM, the response of the solvent to external perturbation is essentially linear. This makes the approach

accurate for small perturbations, but it ultimately fails to describe larger ones, such as the creation of a cavity in a bulk solvent.

However, recently, some studies have shown that with a few empirical corrections, 3D-RISM can also be used to predict solvation free energies of small molecules with a good accuracy^{31–36}. Later, Sergiievskiy and co-workers demonstrated that these corrections are related to the overestimated bulk pressure of the liquid, found in 3D-RISM^{37,38}. They came up with a theory-based pressure correction to 3D-RISM (PC), which, while removing a need to introduce empirical adjustments to the model, was not as accurate as empirical alternatives. Building upon the work by Sergiievskiy et al., we introduced what is now called an advanced pressure correction (PC+)^{39–42}. While this correction was initially suggested based on empirical observation, it is now clear that it has a physical basis, as we discuss in the second half of the thesis.

The main goal of the thesis is to investigate the accuracy and the scope of 3D-RISM advanced pressure correction, PC+. We show that while PC+ suffers from a range of problems related to its approximations, ultimately, for molecular solutes, it can predict solvation free energies with a good accuracy in water as well as in a range of nonpolar solvents. Since the model does not have adjustable parameters, it can be applied to both pure liquids as well as mixtures with comparable accuracy. Additionally, it can be useful for prediction of the first derivatives of free energy such as entropy or enthalpy, although, the accuracy is lower. Through the thesis, we compare model performance with other approaches and discuss their advantages and downsides.

1.1 Structure

The thesis is split into two halves, with the first mostly dedicated to the review of the background theory and previous results, and the second concerned with the new findings, analyses, and discussion. Specifically, we dedicate the second chapter to solvation thermodynamics, third to the exact results in simple liquids theory, and fourth to approximations necessary for computing liquids structure and free energy. The second part of the thesis starts with the fifth chapter in which we introduce pressure corrections and present a number of analytical results that justify it and establish its limits. The sixth chapter is mostly concerned with the accuracy of the model when applied to water, while

the seventh chapter discusses non-aqueous solvents. We wrap everything up with the conclusion that summarizes the main findings and suggests directions for future research.

The thesis is supplemented by two appendices dedicated to the methodology of the calculations performed throughout the thesis as well as additional figures and tables that did not fit into the main part of the work.

Part I

Theoretical background

Chapter 2

Solvation thermodynamics

The first half of the chapter discusses a number of useful results from statistical mechanics with derivation based on the Ref. 12. The second half of the chapter is largely based on the recent research literature on solvation thermodynamics and contains a couple of original results.

2.1 Classical partition function

In this thesis, we will be mainly concerned with solvation in common solvents near room temperatures. Typically, the behaviour of atomic nuclei in such systems can be described sufficiently accurately using classical statistical mechanics^{12,26,43}. Electrons in these systems do behave quantum mechanically, but in many cases, their interactions can be reasonably well approximated by empirical potentials.

Equilibrium properties of the system can be conveniently described using partition functions. For simplicity, we assume that particle interactions do not depend on their orientation^a. Then the partition function of N indistinguishable classical particles in the canonical ensemble takes the following form:

$$Q_N = \frac{q^N}{h^{3N} N!} \iint \exp[-\beta \mathcal{H}(\mathbf{p}_1 \cdots \mathbf{p}_N, \mathbf{r}_1 \cdots \mathbf{r}_N)] d\mathbf{r}_1 \cdots d\mathbf{r}_N d\mathbf{p}_1 \cdots d\mathbf{p}_N, \quad (2.1)$$

^aThe orientation dependant potentials do not change much conceptually, but make notation more cumbersome.

where h is Planck's constant, \mathcal{H} is the Hamiltonian of the system, $\beta = 1/kT$, with k being the Boltzmann constant and T is temperature. \mathbf{r} is the position vector, in 3 dimensions given by $\mathbf{r} = [x, y, z]$ and \mathbf{p} is the momentum vector, given by $\mathbf{p} = [p_x, p_y, p_z]$. The integrals are taken over all possible positions and momenta for each particle. To simplify notation, whenever the limits on the integral sign are omitted, it is implied that integration is performed over the whole range of possible values.

Another quantity appearing in the equation, q is the single-particle partition function, containing degrees of freedom which we assume to be entirely independent of particle's position and interaction with other particles. For molecules, it can be typically factored into a product of the partition functions for each degree of freedom $q = q_{rot}q_{vib}q_{el}$, with subscripts *rot*, *vib*, and *el* representing rotational, vibrational, and electronic degrees of freedom. The exact form of these functions depends on the molecule in question⁴⁴.

For our systems of interest, kinetic and potential energy are independent so that Hamiltonian can be split into two parts

$$\mathcal{H}(\mathbf{p}_1 \cdots \mathbf{p}_N, \mathbf{r}_1 \cdots \mathbf{r}_N) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}_1 \cdots \mathbf{r}_N), \quad (2.2)$$

where m_i is the mass of the i -th particle and U is the potential energy of the whole system. Substituting the above expression into 2.1 and integrating over momenta we get

$$Q_N = \frac{q^N}{N! \Lambda^{3N}} Z_N. \quad (2.3)$$

Here we introduced the so-called thermal de Broglie wavelength, which is defined by

$$\Lambda = \sqrt{\frac{h^2}{2\pi m k T}}. \quad (2.4)$$

Another new quantity in the equation, Z , is called the configurational partition function and is defined as

$$Z_N = \int \exp[-\beta U(\mathbf{r}_1 \cdots \mathbf{r}_N)] d\mathbf{r}_1 \cdots d\mathbf{r}_N. \quad (2.5)$$

For the majority of systems, the configurational partition function cannot be evaluated analytically. One notable exception is an ideal gas for which $U = 0$ everywhere. The

ideal gas partition function is then

$$Q_N^{ig} = \frac{q^N}{N! \Lambda^{3N}} \int d\mathbf{r}_1 \cdots d\mathbf{r}_N = \frac{q^N V^N}{N! \Lambda^{3N}}, \quad (2.6)$$

where V is volume.

An average value of some observable quantity $A(\mathbf{p}_1 \cdots \mathbf{p}_N, \mathbf{r}_1 \cdots \mathbf{r}_N)$ can be conveniently expressed via a partition function. Note that the probability density Pr to find system in a state $\mathbf{r}_1 \cdots \mathbf{r}_N, \mathbf{p}_1 \cdots \mathbf{p}_N$ is

$$\text{Pr} = \frac{1}{Q_N} \exp(-\beta \mathcal{H}) d\mathbf{r}_1 \cdots d\mathbf{r}_N d\mathbf{p}_1 \cdots d\mathbf{p}_N, \quad (2.7)$$

where we omitted the dependence of \mathcal{H} on the phase space position for clarity. It follows that an average of some quantity A is

$$\langle A \rangle = \frac{1}{Q_N} \iint A(\mathbf{r}_1 \cdots \mathbf{r}_N, \mathbf{p}_1 \cdots \mathbf{p}_N) \exp(-\beta \mathcal{H}) d\mathbf{r}_1 \cdots d\mathbf{r}_N d\mathbf{p}_1 \cdots d\mathbf{p}_N, \quad (2.8)$$

where $\langle \cdots \rangle$ here denote ensemble average.

In addition to the canonical ensemble (which has constant N , V , and T), one can also define the isobaric-isothermal ensemble that has constant pressure P , number of particles N , and temperature T . For such systems the partition function becomes

$$\Delta_N = \frac{1}{V_0} \int_0^\infty \exp(-\beta PV) Q_N dV, \quad (2.9)$$

where V_0 is a unit volume that is used to make Δ dimensionless. In the grand canonical ensemble, the constant quantities are chemical potential μ , temperature T , and volume V . The number of particles, N , is allowed to vary. The corresponding partition function is

$$\Xi = \sum_{N=0}^{\infty} Q_N \exp(\beta N \mu). \quad (2.10)$$

2.2 Free energies

The second law of thermodynamics states that the equilibrium state of an isolated thermodynamic system maximizes its total entropy⁴⁵. However, in practice, we rarely

deal with isolated systems. For systems in contact with some external reservoirs this law can be reformulated in the following way: at equilibrium, the thermodynamic system minimizes its corresponding thermodynamic potential, which depends on the systems constraints. For systems subject to NVT conditions (canonical ensemble) the appropriate potential is Helmholtz free energy F , for NPT condition (isobaric-isothermal ensemble) it is Gibbs free energy G , and for μVT conditions (grand canonical ensemble) it is the grand potential Ω .

Importantly, the thermodynamic potentials (free energies) are linked to partition functions in the following way

$$\begin{aligned} F &= -kT \ln Q_N, \\ G &= -kT \ln \Delta, \\ \Omega &= -kT \ln \Xi. \end{aligned} \tag{2.11}$$

These relationships provide a link between thermodynamics and statistical mechanics.

Similarly to partition functions, free energies can be decomposed into contributions from kinetic energy (ideal gas) and from potential energy (usually called excess free energy). These parts can be expressed in terms of partition functions using equations 2.11, 2.6, and 2.3

$$F = -kT \ln Q_N^{ig} - kT \ln \frac{Z}{V^N} = F^{ig} + F^{ex}, \tag{2.12}$$

where superscript ig indicates ideal part and ex excess part of the free energy. F^{ig} can be readily evaluated using 2.6 and Stirling's approximation

$$\begin{aligned} F^{ig} &= -NkT \ln \frac{qV}{\Lambda^3} + kT \ln N! \\ &\approx -NkT \ln \frac{qV}{\Lambda^3} + kT(N \ln N - N) \\ &= NkT (\ln \Lambda^3 \rho q^{-1} - 1). \end{aligned} \tag{2.13}$$

2.3 Chemical potential

The chemical potential of the i -th type of particle in a multicomponent mixture is defined as:

$$\mu_i = \left(\frac{\partial F}{\partial N_i} \right)_{T, V, N_{j \neq i}}. \tag{2.14}$$

The above expression is perfectly valid for other thermodynamic potentials as long as appropriate constraints are chosen (for example, the derivative can be taken with respect to Gibbs free energy G if pressure instead of volume is fixed).

The chemical potential, despite often being considered as a somewhat mysterious quantity⁴⁶, actually, has a straightforward physical meaning. Similarly to how thermodynamic free energies can be viewed as generalizations of the potential of classical systems (hence the name thermodynamic potential), the derivatives of the said potentials are similar to forces. μ_i can be viewed as a force exerted by the system on the particle of type i , with negative and positive signs corresponding to particles being driven into and out of the system respectively.

Two, more formal definitions of chemical potentials within statistical mechanics, will be presented below. Both derivations will be presented in the canonical ensemble, but almost identical results hold in systems with other constraints. To simplify notation, we will be viewing solvation in a one-component system where the inserted molecule is identical to other particles (the same results hold true where the inserted molecule is distinct). This subject is given much attention since the main goal of the thesis is an estimation of chemical potentials in various systems.

It can be shown¹² that in the thermodynamic limit where $N \rightarrow \infty$ (or in other words, when the insertion of a new particle does not change the composition of the system):

$$\mu = F(N + 1) - F(N) \quad (2.15)$$

with all other variables kept constant. Combining this equation with 2.11 we get

$$\mu = -kT \ln \frac{Q_{N+1}}{Q_N} = -kT \ln \left[\frac{q^{N+1} N! \Lambda^{3N}}{q^N N! (N+1) \Lambda^{3(N+1)}} \frac{Z_{N+1}}{Z_N} \right]. \quad (2.16)$$

Then

$$\mu = -kT \ln \left[\frac{q}{(N+1) \Lambda^3} \frac{\int d\mathbf{r}_1 \cdots d\mathbf{r}_N d\mathbf{r}_{N+1} \exp(-\beta U_{N+1})}{\int d\mathbf{r}_1 \cdots d\mathbf{r}_N \exp(-\beta U_N)} \right]. \quad (2.17)$$

To move forward we need to split the potential energy of system with added particle U_{N+1} into two parts:

$$U_{N+1} = U_N(\mathbf{r}_1 \cdots \mathbf{r}_N) + E^{uv}(\mathbf{r}_1 \cdots \mathbf{r}_N, \mathbf{r}_{N+1}) \quad (2.18)$$

where E^{uv} is the interaction (binding) energy of the particle $N + 1$ to the rest of the system. The superscript uv denotes interactions between the solute, the particle being inserted, and solvent, the medium. Substituting the above back into equation 2.17 we get:

$$\mu = -kT \ln \left[\frac{q}{(N+1)\Lambda^3} \frac{\int d\mathbf{r}_1 \cdots d\mathbf{r}_N d\mathbf{r}_{N+1} \exp(-\beta E^{uv}) \exp(-\beta U_N)}{\int d\mathbf{r}_1 \cdots d\mathbf{r}_N \exp(-\beta U_N)} \right], \quad (2.19)$$

where dependencies of potential energies on positions of all particles were omitted for clarity. This equation can be further simplified by noting that in a homogeneous liquid the potential energy depends only on the relative positions of particles. By setting $\mathbf{r}'_i = \mathbf{r}_i - \mathbf{r}_{N+1}$ for position vectors $1 \cdots N$ and integrating out \mathbf{r}_{N+1} we get

$$\mu = -kT \ln \left[\frac{qV}{(N+1)\Lambda^3} \langle \exp(-\beta E^{uv}) \rangle_0 \right], \quad (2.20)$$

where $\langle \cdots \rangle_0$ denote the averaging over positions of all particles (the chemical potential does not depend on the moment, so their values are unimportant). In the thermodynamic limit $V/(N+1) \approx V/N$:

$$\mu = kT \ln (\rho \Lambda^3 q^{-1}) - kT \ln \langle \exp(-\beta E^{uv}) \rangle_0 \quad (2.21)$$

This important result was first discovered by Widom in 1963^{47,48}. In the above equation the chemical potential is clearly split into ideal and non-ideal parts, with

$$\mu^{ig} = kT \ln (\rho \Lambda^3 q^{-1}). \quad (2.22)$$

and

$$\mu^{ex} = -kT \ln \langle \exp(-\beta E^{uv}) \rangle_0. \quad (2.23)$$

There is an alternative and equally useful statistical mechanics expression for chemical potential that was first derived by Kirkwood in 1935^{49,50}. Imagine insertion of a single particle into the system as a continuous process during which particle-system interactions are slowly turned on. To characterise such a process we introduce a modified Hamiltonian $\mathcal{H}(\lambda)$, with $\lambda = 0$ representing an uncoupled state in which $N + 1$ -th particle does not interact with the rest of the system, and $\lambda = 1$ being a final state in which all particle-system interactions are turned on. The approach is quite general since any continuous

function $\mathcal{H}(\lambda)$ satisfying the above requirements would suffice.

The derivative of free energy with respect to coupling parameter is given by:

$$\begin{aligned} \frac{dF}{d\lambda} &= -kT \frac{d}{d\lambda} \ln \int \int \exp[-\beta\mathcal{H}(\lambda)] d\mathbf{r}_1 \cdots d\mathbf{r}_{N+1} d\mathbf{p}_1 \cdots d\mathbf{p}_{N+1} \\ &= -kT \frac{\int \int -\frac{d\mathcal{H}(\lambda)}{d\lambda} \exp[-\beta\mathcal{H}(\lambda)] d\mathbf{r}_1 \cdots d\mathbf{r}_{N+1} d\mathbf{p}_1 \cdots d\mathbf{p}_{N+1}}{\int \int \exp[-\beta\mathcal{H}(\lambda)] d\mathbf{r}_1 \cdots d\mathbf{r}_{N+1} d\mathbf{p}_1 \cdots d\mathbf{p}_{N+1}}. \end{aligned} \quad (2.24)$$

Expressing the above in terms of an ensemble average we get:

$$\frac{dF}{d\lambda} = \left\langle \frac{d\mathcal{H}(\lambda)}{d\lambda} \right\rangle_{\lambda}, \quad (2.25)$$

which after applying the fundamental theorem of calculus becomes

$$\Delta F = \int_0^1 \left\langle \frac{d\mathcal{H}(\lambda)}{d\lambda} \right\rangle_{\lambda} d\lambda. \quad (2.26)$$

In the case of a linear coupling of solute-solvent potential energy $\mathcal{H}(\lambda) = \mathcal{H}(0) + \lambda E^{uv}$, equation 2.26 reduces to

$$\mu^{ex} = \int_0^1 \langle E^{uv} \rangle_{\lambda} d\lambda, \quad (2.27)$$

where we necessarily get μ^{ex} , since the mass of the inserted particle, and thus, the total kinetic energy of the system was unchanged. Practically, the linear coupling is not always convenient, but from the theory perspective, it gives exact results.

Assuming a linear dependence of interaction energy on the coupling strength (linear response) $\langle E^{uv} \rangle_{\lambda} = \langle E^{uv} \rangle_{\lambda=1} \lambda + \langle E^{uv} \rangle_{\lambda=0} (1 - \lambda)$, we can obtain a useful rough estimate of the free energy change

$$\Delta F \approx \int_0^1 d\lambda \langle E^{uv} \rangle_{\lambda=1} \lambda = \frac{1}{2} \langle E^{uv} \rangle_{\lambda=1} + \frac{1}{2} \langle E^{uv} \rangle_{\lambda=0}, \quad (2.28)$$

which is just an average of initial and final particle system interaction energies. For chemical potential, this equation is not very useful, since $\langle E^{uv} \rangle_{\lambda=0}$ is not well defined, but for some smaller perturbations, such as adding electrostatic charge to the inserted formula, this equation is quite accurate⁵¹.

Comparing the two obtained equations for the chemical potential, it may appear that Widom's formula (2.21) is computationally more convenient as the interactions are computed only in one state, as opposed to equation 2.26 in which one has to compute $d\mathcal{H}(\lambda)/d\lambda$ in a number of systems with varying λ . However, in the case of dense systems, Kirkwood's equation is advantageous. For most liquids random insertions of a particle will cause overlap, resulting in a significant number of trials required for expression $\langle \exp(-\beta E^{uv}) \rangle_0$ to converge⁵². On the other hand, using Kirkwood's formula, one can start with a system in which particle is already fully coupled and then slowly decouple it, letting the surroundings relax. Such an approach is guaranteed to yield a good estimate of μ independent of the density of the system. For this reason, the majority of chemical potential calculations employ Kirkwood's formula.

Finally, it is useful to consider the relationship between the chemical potential and the ensemble from a thermodynamic perspective. Intuitively, it seems reasonable that as long as macroscopic thermodynamic parameters such as pressure or density are identical, and the system is sufficiently large, the chemical potential would be independent of the types of system constraints. However, if we write out the equations for insertion explicitly, we will notice that this independence is realised through the cancellation of ensemble specific contributions⁵³.

Let's denote the chemical potential obtained by differentiating Helmholtz free energy at constant volume as μ_V and the one obtained by differentiating Gibbs free energy at constant pressure as μ_P . To see the relationship between these quantities we right out

$$G(T, P, N) = F(T, V(T, P, N), N) + PV(T, P, N) \quad (2.29)$$

and then using a chain rule⁵³:

$$\begin{aligned} \mu_P &= \left(\frac{\partial G}{\partial N'} \right)_{T, P, N} \\ &= \left(\frac{\partial F}{\partial N'} \right)_{T, P, N} + P \left(\frac{\partial V}{\partial N'} \right)_{T, P, N} \\ &= \left(\frac{\partial F}{\partial N'} \right)_{T, V, N} + \left(\frac{\partial F}{\partial V} \right)_{T, N} \left(\frac{\partial V}{\partial N'} \right)_{T, P, N} + P \left(\frac{\partial V}{\partial N'} \right)_{T, P, N} \\ &= \left(\frac{\partial F}{\partial N'} \right)_{T, V, N} = \mu_V, \end{aligned} \quad (2.30)$$

where in the third equality we used the identity $P = -(\partial F/\partial V)_{T,N}$ and we used superscript N' to separate introduced particle from the rest of the system. The partial derivative appearing in the above expression $(\partial V/\partial N')_{T,P,N} = \bar{V}$ is called the partial molar volume. It indicates by how much system volume changes when we introduce a small number of new particles; we will discuss it in more detail in later section. At constant pressure, the extra work required to increase system size by $P\bar{V}$ is compensated by a decrease in free energy due to the system expansion.

An analogous procedure can be performed in the case of the Grand potential. We start by writing

$$\Omega(T, V, \mu) = F(T, V, N(T, V, \mu)) - \mu N(T, V, \mu). \quad (2.31)$$

Assuming the solvated particle is distinct from the rest

$$\begin{aligned} \mu_\mu &= \left(\frac{\partial \Omega}{\partial N'} \right)_{T,V,\mu} \\ &= \left(\frac{\partial F}{\partial N'} \right)_{T,V,\mu} - \mu \left(\frac{\partial N}{\partial N'} \right)_{T,V,\mu} \\ &= \left(\frac{\partial F}{\partial N'} \right)_{T,V,N} + \left(\frac{\partial F}{\partial N} \right)_{T,V} \left(\frac{\partial N}{\partial N'} \right)_{T,V,\mu} - \mu \left(\frac{\partial N}{\partial N'} \right)_{T,V,\mu} \\ &= \left(\frac{\partial F}{\partial N'} \right)_{T,V,N} = \mu_V, \end{aligned} \quad (2.32)$$

where $\mu = (\partial F/\partial V)_{T,V,N}$ was utilised. The quantity $(\partial N/\partial N')_{T,V,\mu}$ is directly related to partial molar volume. In appendix B.1 we show that $(\partial N/\partial N')_{T,V,\mu} = -\rho\bar{V}$. Then in a grand canonical ensemble, we obtain extra $\mu\rho\bar{V}$ energy from the particle bath, but it gets compensated due to a decrease of particle number inside the system.

2.4 Solvation free energy

To discuss solvation thermodynamics we first need to define a term solution. According to IUPAC "Gold Book"⁵⁴: "Solution is a liquid or solid phase containing more than one substance". A dominant component of the solution is usually called solvent, while minor components are referred to as solutes. In most of our discussion, we will be dealing with infinitely dilute solutions in which concentrations of solutes, as well as their mutual interactions, tend to zero.

According to Ben-Naim, the solvation may be defined as the process of transferring a solute from a fixed position in an ideal gas phase into a fixed position in the solvent^{6,12}. To express this definition analytically, we need to introduce the pseudo-chemical potential, which is a chemical potential associated with a stationary particle. It is typically denoted as μ^* and is given by

$$\mu^* = \mu - kT \ln \rho \Lambda^3. \quad (2.33)$$

The solvation free energy is defined as

$$\Delta G^* = \mu^{*l} - \mu^{*ig} = \Delta \mu^*, \quad (2.34)$$

where superscript l indicates solvent and $*$ refers to Ben-Naim's definition. As we saw in the previous section, the value of chemical potential is ensemble-independent and thus $\Delta G^* = \Delta F^* = \Delta \Omega^*$. However, in practice, the symbol ΔG is used to indicate that the system is connected to a constant pressure and temperature bath.

Expressing everything in terms of μ^{ex} (defined via the equation 2.21) we can rewrite the previous equation to get:

$$\Delta G^* = \mu^{ex} - kT \ln \frac{q^l}{q^{ig}}. \quad (2.35)$$

Clearly, if the internal partition function of the molecule is unaffected by the phase transfer, ΔG^* is equal to the coupling work of the solvent or μ^{ex} .

Ben-Naim's convention for solvation free energies is not the only one in use. Another commonly used way of expressing solvation free energies is based on standard states. The process of solvation is then described as a transfer of a compound at a standard gaseous state (a hypothetical state of pure substance at which it exhibits ideal gas behaviour and has standard pressure $P^o = 1$ bar) to the standard solution state (a hypothetical state of an ideal solution at standard pressure P^o and molality $b^o = 1$ mol kg⁻¹). Solvation free energies corresponding to this process are denoted as ΔG^o .

While the use of standard states has many advantages, for the process of solvation they are not very convenient. Within the standard-state approach, solute molecules change their density during the transfer, which leads to an artificial dependence of derivatives of solvation free energy on quantities such as thermal expansion or compressibility. Additionally, one molal standard state is far from infinite dilution; defining such state

as an ideal solution in which solute molecules do not interact with each other is not physically meaningful.

Due to the above reasons, in this thesis, we will be primarily using Ben-Naim's definition of solvation. For simplicity, we will also drop the unnecessary * and will simply denote corresponding solvation free energies as ΔG , ΔF , or $\Delta\Omega$, depending on the system. To convert from one type of definition to another, we can use equation 2.34. Then

$$\begin{aligned}\Delta G^o &= \Delta G^* - kT \ln \frac{\rho_l}{\rho_g} \\ &= \Delta G^* - kT \ln \frac{b_u^o M_v \rho_v kT}{P^o},\end{aligned}\tag{2.36}$$

where M_v is the molar mass and ρ_v is the density of the solvent. For water at 298.15 K and standard pressure, this corresponds to $\Delta G^* = \Delta G^o - 1.9$ kcal/mol.

2.5 Decomposing solvation free energy

A lot of insight can be obtained by examining various decompositions of solvation free energy into different components. In this section, we will demonstrate the separation of excess chemical potential into energetic and entropic components, following the approach that is commonly used in thermodynamic and statistical mechanics treatments of the subject, and in section 2.6 we will approach this task from the simulations perspective.

Ben-Amotz et al. have demonstrated that one can formally decompose μ^{ex} into two equivalent representations^{51,55}

$$\begin{aligned}\mu^{ex} &= \langle E^{uv} \rangle_{\lambda=1} + \beta \int_0^1 d\lambda \lambda \left[\langle (E^{uv})^2 \rangle_{\lambda} - \langle E^{uv} \rangle_{\lambda}^2 \right] \\ &= \langle E^{uv} \rangle_{\lambda=1} + \frac{1}{\beta} \ln \langle \exp[\beta(E^{uv} - \langle E^{uv} \rangle)] \rangle_{\lambda=1},\end{aligned}\tag{2.37}$$

where brackets $\langle \dots \rangle_{\lambda}$ denote averaging in the ensemble of interest at a particular coupling strength λ . The first term in the both equations is the strength of interactions between solvent and fully coupled solute, and it represents an enthalpic contribution to the solvation free energy. Both second terms represent an entropic contribution to the solvation free energy and will be denoted as $-TS^{uv}$; they are both proportional to fluctuations of solute-solvent interaction energy. Integrating the first of the above equation by parts we

get

$$\mu^{ex} = \langle E^{uv} \rangle_{\lambda=1} + \frac{\beta}{2} \left[\langle (E^{uv})^2 \rangle_{\lambda=1} - \langle E^{uv} \rangle_{\lambda=1}^2 \right] + \dots, \quad (2.38)$$

where \dots represent higher order cumulants that disappear if the fluctuations of solute-solvent energy are Gaussian.

From the previous paragraph, we can see that $\mu^{ex} = E^{uv} - TS^{uv}$, which is a very convenient decomposition from the theoretical point of view. The entropic contribution is always positive $-TS^{uv} \geq 0$ and energetic is negative for the absolute majority of solutes. Note that within the linear response approximation, discussed in section 2.3, $-TS^{uv} = -1/2E^{uv} + 1/2E_0^{uv}$.

We can also decompose the excess chemical potential of solvation μ^{ex} using more conventional definitions of solvation energy and entropy; however, these values have solvent-solvent contributions that will cancel each other out. We start by using the following definition of chemical potential

$$\mu^{ex} = \Delta U - T\Delta S, \quad (2.39)$$

where ΔS is the excess solvation entropy given by

$$\Delta S = - \left(\frac{\partial \mu^{ex}}{\partial T} \right)_V \quad (2.40)$$

and ΔU is the change in system's excess internal energy, given by

$$\Delta U = \left[\frac{\partial (\mu^{ex}/T)}{\partial (1/T)} \right]_V. \quad (2.41)$$

In the above equations, subscripts indicating that we are dealing with the excess quantities were dropped for clarity.

Let us first look at the change in excess internal energy ΔU . Similarly to equation 2.15, the following result holds $\Delta U = U^{ex}(N_v, N_u = 1) - U^{ex}(N_v, N_u = 0)$. Then we can formally write

$$\Delta U = \langle U^{vv} + E^{uv} \rangle_{\lambda=1} - \langle U^{vv} \rangle_{\lambda=0} = E^{uv} + \Delta \langle U^{vv}, \rangle \quad (2.42)$$

where U^{vv} is the interaction energy of solvent atoms and $\Delta U^{vv} = \langle U^{vv} \rangle_{\lambda=1} - \langle U^{vv} \rangle_{\lambda=0}$ is called the solvent reorganization energy.

The above equation contains a term accounting for the solute-solvent and the solvent-solvent interactions. Both terms can be either positive or negative. It also should be noted that the first term is relatively easy to compute, as for typical potentials $\langle E^{uv} \rangle_{\lambda=1}$ is short ranged. On the other hand, $\Delta \langle U^{vv} \rangle$ is the difference of two large interaction energies and is usually difficult to evaluate.

To obtain an expression for solvation entropy ΔS , we note that

$$T\Delta S = -T \left(\frac{\partial \mu^{ex}}{\partial T} \right)_V = \beta \left(\frac{\partial \mu^{ex}}{\partial \beta} \right)_V = \beta \int_0^1 d\lambda \frac{\partial \langle E^{uv} \rangle_\lambda}{\partial \beta}, \quad (2.43)$$

where we used the Kirkwood-Buff expression for chemical potential (equation 2.27) to obtain the final equality. We can simplify the integrand in the above by writing out its definition

$$\begin{aligned} \frac{\partial \langle E^{uv} \rangle_\lambda}{\partial \beta} &= \frac{\partial}{\partial \beta} \frac{\int d\mathbf{r}_1 \cdots d\mathbf{r}_N d\mathbf{r}_{N+1} E^{uv} \exp[-\beta U_{N+1}(\lambda)]}{Z(\lambda)} \\ &= \frac{-\langle U_{N+1}(\lambda) \rangle \langle E^{uv} \rangle Z^2(\lambda) + \langle E^{uv} U_{N+1}(\lambda) \rangle Z^2(\lambda)}{Z^2(\lambda)} \\ &= -\langle U_{N+1}(\lambda) \rangle \langle E^{uv} \rangle + \langle E^{uv} U_{N+1}(\lambda) \rangle \\ &= -\langle U^{vv} \rangle_\lambda \langle E^{uv} \rangle_\lambda - \lambda \langle E^{uv} \rangle_\lambda^2 + \langle E^{uv} U^{vv} \rangle_\lambda + \lambda \langle (E^{uv})^2 \rangle_\lambda \\ &= -\langle U^{vv} \rangle_\lambda \langle E^{uv} \rangle_\lambda - \langle E^{uv} U^{vv} \rangle_\lambda - \lambda \left[\langle E^{uv} \rangle_\lambda^2 - \langle (E^{uv})^2 \rangle_\lambda \right], \end{aligned} \quad (2.44)$$

where we used $\langle U_{N+1} \rangle_\lambda = \langle U^{vv} \rangle_\lambda + \lambda E^{uv}$ to obtain the fourth equality. Returning back to equation 2.43 we get

$$T\Delta S = -\beta \int_0^1 d\lambda \lambda \left[\langle E^{uv} \rangle_\lambda^2 - \langle (E^{uv})^2 \rangle_\lambda \right] - \beta \int_0^1 d\lambda \left[\langle U^{vv} \rangle_\lambda \langle E^{uv} \rangle_\lambda - \langle E^{uv} U^{vv} \rangle_\lambda \right]. \quad (2.45)$$

Ben-Amotz, and earlier Yu and Karplus, have shown that^{12,56}

$$\Delta U^{vv} = -\beta \int_0^1 d\lambda \left[\langle U^{vv} \rangle_\lambda \langle E^{uv} \rangle_\lambda - \langle E^{uv} U^{vv} \rangle_\lambda \right]. \quad (2.46)$$

This equation combined with the result above gives us the final expression

$$T\Delta S = T\Delta S^{uv} + \Delta U^{vv}. \quad (2.47)$$

Notice that both internal energy change ΔU and solvation entropy ΔS contain contributions due to solvent reorganization energy U^{vv} . However, it cancels out when we add individual derivatives $\mu^{ex} = \Delta U - T\Delta S = E^{uv} - T\Delta S^{uv}$.

Following Ben-Amotz, we can extend these results to other ensembles

$$\begin{aligned} \Delta G_s &= E_P^{uv} - TS_P^{uv} + P\bar{V}, \\ \Delta\Omega_s &= E_\mu^{uv} - TS_\mu^{uv} + \mu\rho\bar{V}. \end{aligned} \quad (2.48)$$

The rationale behind $\mu\rho\bar{V}$ term was given at the end of section 2.3. Note that since the averaging is done in different ensembles, generally $E_P^{uv} \neq E_V^{uv} \neq E_\mu^{uv}$, with the same holding true for entropy. Similarly, derivatives of free energy are not necessarily equal in different ensembles. Even though the chemical potential is ensemble-independent, its decompositions are not.

2.6 Numerical experiments

The dynamics of the majority of liquids at room temperature can be well approximated using Newton's equations of motion²⁶. Thus, in principle one could simulate the liquid by putting a sufficient number of molecules in a box, giving them initial velocities according to the Maxwell-Boltzmann distribution and then updating their positions and velocities using the force $\mathbf{F} = -\nabla U(\mathbf{r}_1 \cdots \mathbf{r}_N)$. This is the basic idea behind molecular dynamics (MD) simulations, which is an extremely powerful tool for studying liquids and their solutions.

The success of molecular dynamics simulation largely depends on the quality of the approximation of intermolecular potential U . In all of the simulations performed in this thesis, we assumed that U is pair decomposable:

$$U = \sum_{i,j} u_{ij}(r_{ij}), \quad (2.49)$$

where $\sum_{i,j}$ is the sum over all pairs of interacting sites (particles), r_{ij} is the distance between them, and u_{ij} is the pair potential.

The form of u_{ij} depends on whether sites i and j are part of the same or different molecules. If sites i and j are both located on the same molecules, the interaction between them will depend on the types of bonds present in the molecule. For i and j which are parts of different molecules, or are separated by a sufficiently large number of bonds, the pair potential is usually given by the sum of short-ranged and electrostatic potentials:

$$\begin{aligned} u_{ij}(r) &= u_{ij}^{LJ}(r) + u_{ij}^{el}(r), \\ u_{ij}^{LJ}(r) &= 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \\ u_{ij}^{el}(r) &= \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \end{aligned} \quad (2.50)$$

where ϵ_{ij} is the Lennard-Jones well depth, σ_{ij} is the Lennard-Jones diameter, q is the partial charge, and ϵ_0 is the vacuum permittivity. The 12-6 Lennard-Jones potential is the most commonly used model to approximate short-ranged repulsive forces that originate due to repulsion of electronic clouds as well as somewhat longer range (decaying as r^{-6}) attractive forces due to dispersion interactions.

The equation 2.50 defines the most standard and commonly used form of the intramolecular potential⁵⁷. Potentials of this form are robust and fast. However, they ignore a number of potentially significant effects such as polarization, charge transfer, multi-body interactions, etc.

The process of finding the interaction parameters describing each site is largely empirical. The Lennard-Jones parameters are usually fit to reproduce macroscopic parameters such as density or viscosity. In case of the water, there are a number of models with varying sophistication and number of sites. For a typical organic solute, one can take parameters from various force fields. More simple ones, such as general Amber force field (GAFF)⁵⁸, will have a single set of Lennard-Jones parameters per element, while more advanced ones, such as the optimised potential for liquid simulations (OPLS), have a number of different force field constants depending on the elements bonding⁵⁹⁻⁶¹.

To estimate Lennard-Jones interactions between different types of atoms, one can use various combination rules. For all simulations in the thesis, we will be using so-called

Lorentz-Berthelot rules⁵⁷

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \qquad \epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}, \qquad (2.51)$$

where σ_i and ϵ_i are input taken from the force fields, and σ_{ij} and ϵ_{ij} are inputs for equations 2.50.

An accurate set of partial charges should ideally reproduce an electrostatic potential field produced by the real molecule. Additionally, partial charges should be compatible with Lennard-Jones parameters. The solvent models typically come with specifically adjusted charges. For solutes, one commonly has to perform an electronic structure calculation to find the molecule's electron density distribution and then fit partial charges to it. In the case of GAFF, a typical procedure is to use Austin model one (AM1) method to estimate the distribution of valence electrons approximately and then correct it using semi-empirical bond charge corrections (BCC) scheme, abbreviated as AM1-BCC⁶². This approach has been well established and is known to yield good estimates of non-bonding parameters. More sophisticated schemes, such as charge model five (CM5), use full electronic structure calculations with large basis sets and hybrid electronic density functional theories.

Once parameters have been fit, the solvation free energy can be calculated by slowly coupling (or decoupling) the solute to the solvent. The procedure is split into two stages. At the first stage, one turns on the solute Lennard-Jones parameters, at the second – partial charges. Thus, the solvation free energy can be formally split into:

$$\Delta F = \Delta F^{LJ} + \Delta F^{el}. \qquad (2.52)$$

The above equation presents an alternative scheme for decomposing solvation free energies (besides thermodynamic decomposition into enthalpic-entropic parts), useful in computational chemistry. Since electrostatic free energy can be calculated relatively accurately by modelling solvent as a dielectric continuum, one can combine it with some empirical way of approximating ΔF^{LJ} to obtain a computationally cheap method for deducing solvation free energies without any solvent modelling. Simple models that use partial charges for the electrostatic part include the generalised Born and Poisson-Boltzmann models. Alternatively, models such as SMD or SM-12 combine an empirical ΔF^{LJ} term with continuum charge distributions derived from quantum mechanics.

2.7 Ionic solvation

In the preceding sections, we discussed solvation in the bulk phase of a solvent, ignoring any effects caused by the liquid-gas boundary. For situations when solutes are neutral such an approach is entirely justified since the range of the interfacial effects is small. However, in the case of the charged solutes (for which the sum of atomic charges is non-zero), the nature of the liquid interface actually does affect the solvation free energy, although, the forces inside the bulk are independent of it⁶³.

Any homogeneous bulk phase P will have a characteristic constant potential ϕ_G^P , called the Galvani potential. It can be expressed as a sum of the Volta potential ψ^P and interfacial potential χ^P :

$$\phi_G^P = \psi^P + \chi^P. \quad (2.53)$$

Typically, all three potentials would depend on both the other phases in contact with the bulk phase as well as the shape of the surface. However, an isolated phase with a homogeneous surface polarization has $\psi_P = 0$, so $\phi_P = \chi_P$. In this case, the Galvani potential is uniquely determined by properties of the phase and is independent of its shape or size⁶³. This is the situation we will be primarily concerned with, so the terms Galvani and surface potential are going to be used interchangeably.

In the bulk phase, the Galvani potential is constant and thus does not affect the forces between the particles in any way. However, it has a noticeable effect on the insertion of ions, shifting their intrinsic chemical potentials $\bar{\mu}$ by an amount proportional to the charge^b:

$$\mu = \bar{\mu} + q\phi_G^P \quad (2.54)$$

where q is the total charge of the ion. In the electrochemical literature, the quantity μ is often called the "real" or electrochemical potential and intrinsic chemical potential $\bar{\mu}$ is called the "chemical potential"⁶⁴. On the other hand, in the theory of liquids, the situation is reversed: the real chemical potential is called the "chemical potential" and is denoted as μ , whereas intrinsic chemical potential is usually given a special symbol. Throughout the thesis, we will be using the latter notation²⁶.

^bThe name intrinsic comes from the fact that this is a chemical potential arising purely due to the interactions within the system, without any contribution from the external field, in this case, ϕ_G^P .

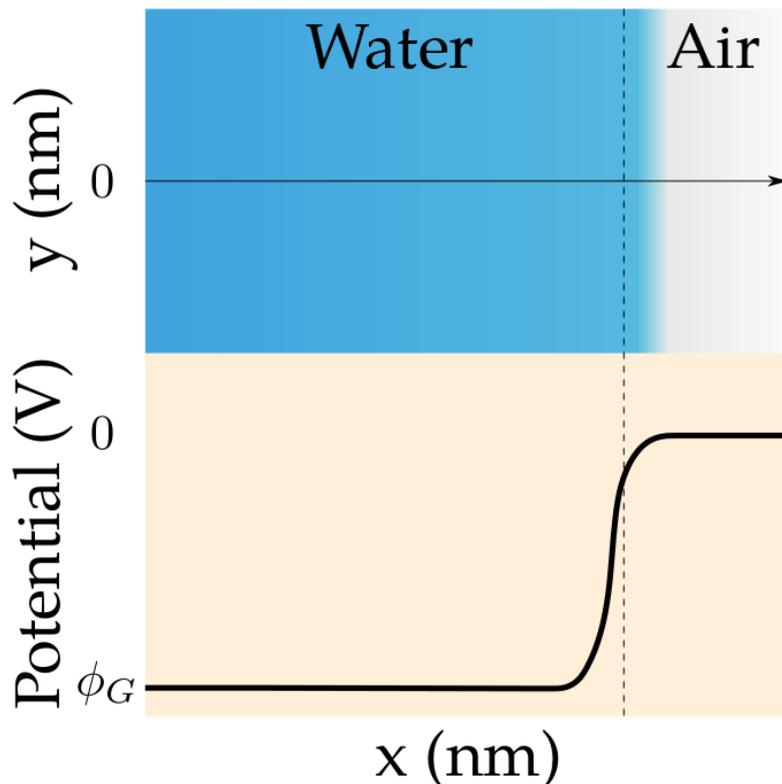


FIGURE 2.1: A schematic drawing demonstrating the dependence of electrostatic potential on the distance from the water-air interface.

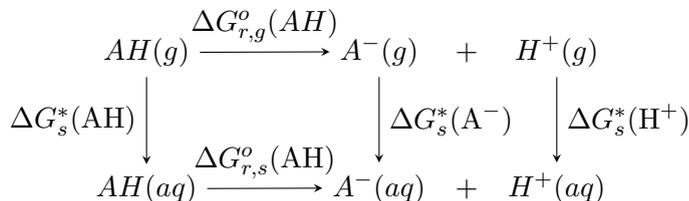


FIGURE 2.2: Thermodynamic cycle relating solvation free energy of an ion A^- to the solvation free energies of acid AH , its dissociation free energies ΔG_r^o in both gas and solvent, and solvation free energy of the proton. Since free energy of the system is independent of the path, we have $\Delta G_s^*(AH) + \Delta G_{r,s}^o(AH) = \Delta G_{r,g}^o(AH) + \Delta G_s^*(A^-) + \Delta G_s^*(H^+)$, from which $\Delta G_s^*(A^-)$ can be deduced.

The Galvani potential ϕ_G^P , and thus $\bar{\mu}$, are experimentally inaccessible quantities. Even though μ is in principle measurable using certain electrochemical techniques (for example by measuring absolute electrode potentials of the metal in solvent^{63,65,66}), a commonly involved approach to measuring chemical potentials (solvation free energies) of ions is done in a different way. One usually measures a hydration free energy of proton and then obtains solvation free energies of other ions from appropriate thermodynamic cycles involving the dissolution of neutral compounds. For example, the solvation free energy of an ion A^- can be obtained from the cycle shown on the figure 2.2. In case an ion B^+

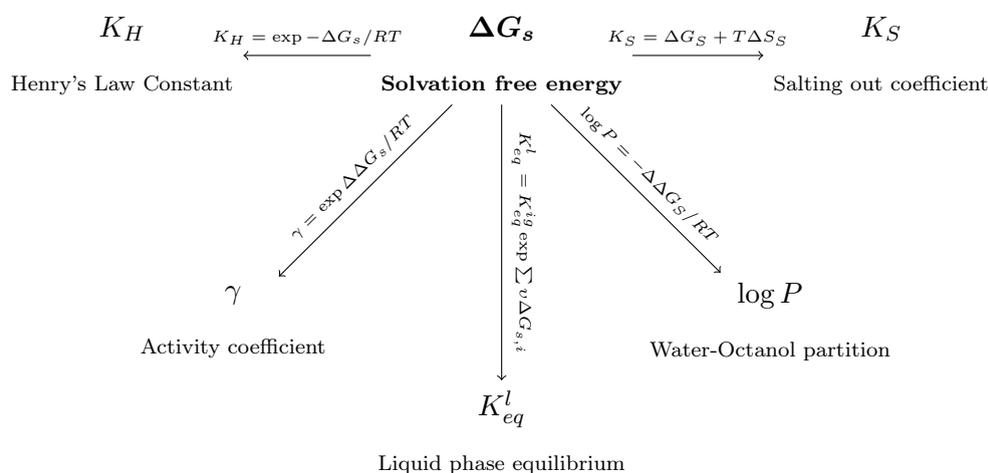
cannot be easily protonated or does not have protons, its solvation free energy can be deduced from the solvation free energy of the ionic pair AB, where A^- is some anion with a known free energy of transfer.

Note that the sum of solvation free energy of a pair of ions AB is independent of the phase's Galvani potential:

$$\begin{aligned}\mu(A^{+n}) + \mu(B^{-n}) &= \bar{\mu}(A^{+n}) + n\phi_G^P + \bar{\mu}(B^{-n}) - n\phi_G^P \\ &= \bar{\mu}(A^{+n}) + \bar{\mu}(B^{-n}) = \Delta G(\text{AB}),\end{aligned}\quad (2.55)$$

where n is the charge on the ion. Thus, the type (intrinsic or real) of solvation free energy of the ion is determined by the type of solvation free energy of a proton. A number of recent articles suggest that the commonly used value for the solvation free energy of proton in water, 265.9 kcal/mol by Tissandier et al.⁶⁷, contains a contribution from surface potential^{68–70}. It follows that hydration free energies of single ions, evaluated using Tissandier's value of proton's hydration free energy, are also "real".

2.8 Applications



In this section, we will briefly overview the relation of solvation free energy (chemical potential) to other solution thermodynamic quantities that present interest to chemical engineers, as well as environmental and life scientists.

The solvation free energy is directly related to Henry's law. Henry's law states that the amount of dissolved gas is proportional to its partial pressure above the solution¹². The

proportionality factor is called the Henry's law constant H . This constant is used in various areas of environmental research since it describes the distribution of species between air and liquid cloud droplets, rivers, wastewaters, as well as other naturally occurring liquid reservoirs. In the literature, a number of different definitions of Henry's law exist, with one of the most common involving dimensionless Henry constant $H^{cc} = c(\text{solution})/c(\text{gas})$, where c stands for the molar concentration⁷¹. It can be related to Ben-Naim's solvation free energy through:

$$\Delta G^* = -RT \ln H^{cc}, \quad (2.56)$$

where R is the universal gas constant. From the previous equation, it follows that solvation free energy and Henry's law constant are two different names for the same quantity.

Knowing solvation free energies at different concentrations allows one to compute solute activities, which are useful for understanding the properties of concentrated solutions. Assuming finite number density of solute ρ_u , we can recast expression 2.21 into the following form¹²

$$\mu = \mu^o + kT \ln \rho_u + kT \ln \gamma^{D,\rho}, \quad (2.57)$$

where $\gamma^{D,\rho}$ is the activity coefficient and μ^o is the standard chemical potential. Above quantities are measured on the number density ρ scale, but can be converted to more commonly used molar or molal scales¹². Relating this equation to our previous results one finds that

$$\mu_{\rho_u}^* - \mu_0^* = \Delta G_{\rho_u}^* - \Delta G_0^* = kT \ln \gamma^{D,\rho}, \quad (2.58)$$

where subscripts ρ_u and 0 correspond to finite and infinitely dilute concentrations of solute respectively.

Finally, solvation free energies can be used for predicting the equilibrium state of reactions and complex formations in different mediums. Thus, they present a considerable interest for areas of chemistry, biology, and material science that are concerned with the formation of various compounds in solutions. Consider the following reaction:



The equilibrium constant K and Gibbs free energy are related as $K^{ig} = \exp(-\beta \Delta G_r^{ig})$, where superscript ig indicates that reaction takes place in ideal gas phase. To obtain the equilibrium constant in the liquid phase we construct a thermodynamic cycle similar to

figure 2.2. Then:

$$K^l = \exp[-\beta(\Delta G_r^{ig} - \Delta G_A^* + \Delta G_B^*)] = K^{ig} \exp[-\beta(\Delta G_B^* - \Delta G_A^*)]. \quad (2.60)$$

Knowledge of solvation free energies allows one to compute reaction equilibria in any medium from the free energy of the gas phase reaction, which can often be computed relatively accurately using quantum chemistry based methods or estimated on the basis of the bond strengths.

Chapter 3

Theory of simple liquids

In this chapter, we overview the main aspects of classical density functional theory. The focus is on the exact results that can be obtained for simple liquids. Both the derivations and the structure of the chapter are based on Ref. 26.

3.1 Particle densities and distributions

Condensed matter systems such as liquids and solids are difficult to study. They consist of a large number of electrons and atomic nuclei interacting with each other in a complicated manner. To describe such systems one usually has to ignore unimportant degrees of freedom. For example, when studying molecules at room temperature vibrations of bonds and angles can be neglected.

A systematic approach towards the reduction of degrees of freedom is called coarse-graining. The basic idea is to replace values of a rapidly varying observable with its average value over certain volume⁷². As a result, we get a continuous and smoothly varying function that is called an order parameter field.

For the description of liquids, a natural order parameter is an ensemble average of the instantaneous density $\rho^I(\mathbf{r})$:

$$\rho(\mathbf{r}) = \langle \rho^I(\mathbf{r}) \rangle = \left\langle \sum_{k=1}^N \delta(\mathbf{r} - \mathbf{r}_k) \right\rangle, \quad (3.1)$$

where $\delta(\mathbf{r})$ is Dirac's delta function and N here is the total number of atoms in a system. $\rho(\mathbf{r})$ is usually called a single-particle density distribution, or local density, and it shows an average density of atoms in the volume element $d\mathbf{r}$. The product $\rho(\mathbf{r}) d\mathbf{r}$ thus shows an average number of particles at that position.

A related quantity, called the pair distribution function, describes the correlation between single-particle densities:

$$\begin{aligned}\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) &= \left\langle \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N \delta(\mathbf{r}_1 - \mathbf{r}_l) \delta(\mathbf{r}_2 - \mathbf{r}_m) \right\rangle \\ &= \langle \rho^I(\mathbf{r}_1) \rho^I(\mathbf{r}_2) \rangle - \rho(\mathbf{r}_1) \delta(\mathbf{r}_1 - \mathbf{r}_2).\end{aligned}\tag{3.2}$$

Note that due to the $m \neq l$ condition, two particles found in positions \mathbf{r}_1 and \mathbf{r}_2 must be different. Thus, $\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2$ can be interpreted as an average number of pairs formed by particles in volume element $d\mathbf{r}_1$ with particles in a volume element $d\mathbf{r}_2$ with a condition that two particles in a pair must be different¹².

It is useful to consider integrals of particle distribution functions. Using the property of the delta function

$$\int_V \rho(\mathbf{r}) d\mathbf{r} = \langle N \rangle_V,\tag{3.3}$$

where $\langle N \rangle_V$ is the total number of particles in the box V . In homogeneous liquids, local density should be constant by the definition. It follows that $\rho(\mathbf{r}) = N/V = \rho$. Similarly, we can take a double integral of pair distribution function to find

$$\iint_V \rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 = \langle N^2 \rangle_V - \langle N \rangle_V^2.\tag{3.4}$$

Then, the volume average of the pair distribution function is

$$\rho_{avg,V}^{(2)} = \frac{\langle N^2 \rangle_V - \langle N \rangle_V^2}{V^2} = \rho^2 \left(1 - \frac{1}{N} \right) \approx \rho^2.\tag{3.5}$$

Since in the ideal gas there are no correlations, the value of pair distribution function in it must be constant throughout the box, or in other words: $\rho_{id}^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \rho_{avg}^{(2)}$.

In real liquids, as the distance between \mathbf{r}_1 and \mathbf{r}_2 increases, the value of the pair distribution function tends to the ideal limit: $\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \rightarrow \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)$. Hence, it is useful to define a

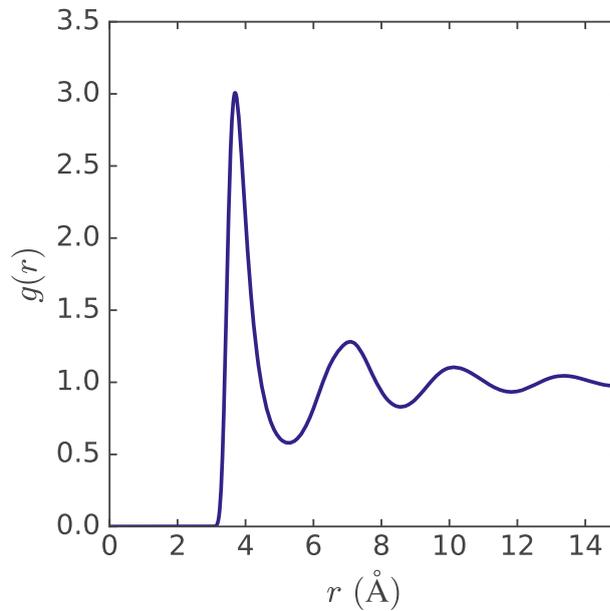


FIGURE 3.1: Radial distribution function of Ar near its boiling point obtained via molecular dynamics simulations using parameters from Ref. 73.

pair correlation function

$$g(\mathbf{r}_1, \mathbf{r}_2) = \frac{\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2)}{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}, \quad (3.6)$$

which quantifies deviation of $\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ from its large distance behaviour. In isotropic liquids, the pair distribution function only depends on the distance between two particles $g(\mathbf{r}_1, \mathbf{r}_2) = g(|\mathbf{r}_1 - \mathbf{r}_2|) = g(r)$, with $g(r) \rightarrow 1$ as $r \rightarrow \infty$. This spherically averaged pair distribution function is often referred to as the radial distribution function. An example of this function for argon near boiling point is shown in figure 3.1.

Finally, following Ben-Naim, we also define conditional local density:

$$\rho(\mathbf{r}_2/\mathbf{r}_1) = \frac{\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2)}{\rho(\mathbf{r}_1)}, \quad (3.7)$$

which shows the local density at \mathbf{r}_2 , given a particle in \mathbf{r}_1 . From equation 3.6 it follows that $g(\mathbf{r}_1, \mathbf{r}_2) = \rho(\mathbf{r}_2/\mathbf{r}_1)/\rho(\mathbf{r}_2)$ for isotropic systems $g(r) = \rho(\mathbf{r}_2/\mathbf{r}_1)/\rho$. These equations suggest an interpretation of the pair correlation function as a scaled local density of a system, in which a single particle is fixed at \mathbf{r}_1 and other particles are moving in its field. Using this interpretation, we can view both inhomogeneous and homogeneous liquids under the same framework.

3.2 Free energy functionals

The basic ideas behind the classical density functional theory of liquids can be illustrated in the following manner. Suppose we have (in general) an inhomogeneous system of interacting particles. Such system can be split into small volume elements containing $n_j = \Delta V_j \rho_j$ number of particles. As particles are free to move from one compartment to another, the free energy of each individual volume element is best described using the grand potential $\Omega_j = F_j - \mu n_j$, where μ is the chemical potential, constant in each compartment. To satisfy grand canonical ensemble condition, the whole system must be connected to external heat and particle reservoirs.

Assume that there is a spatially varying external potential ϕ_j that interacts with particles. Then, the Helmholtz free energy of each volume component is given by $F_j = \mathcal{F}^{ig} + \mathcal{F}^{ex} + n_j \phi_j$, where \mathcal{F} is an intrinsic Helmholtz free energy and superscripts *ig* and *ex* denote ideal and excess contributions. Taking the derivative of the above expression with respect to the number of particles in a box, we get:

$$\left(\frac{\partial F_j}{\partial n_j} \right)_{V,T} = \mu = \bar{\mu}_j^{ig} + \bar{\mu}_j^{ex} + \phi_j, \quad (3.8)$$

where $\bar{\mu}$ stands for the intrinsic chemical potential, familiar from the section 2.7. Notice that while the total chemical potential μ is constant throughout the system, the quantity $\bar{\mu}_j = \bar{\mu}_j^{ig} + \bar{\mu}_j^{ex}$ is spatially varying. Also, since $\mu^{ig} = \bar{\mu}^{ig}$, we will ignore the superscript.

The simple arguments described in the previous two paragraphs can be made formal by shrinking the volume of each box element to an arbitrarily small value and substituting ρ_j with single particle density, introduced in the previous section. The total grand potential becomes a functional of density²⁶

$$\Omega[\rho] = \mathcal{F}[\rho] + \int \rho(\mathbf{r}) \phi(\mathbf{r}) d\mathbf{r} - \mu \int \rho(\mathbf{r}) d\mathbf{r}, \quad (3.9)$$

with \mathcal{F} being an intrinsic free energy functional related to Helmholtz free energy F in the following way

$$\mathcal{F}[\rho] = F[\rho] - \int \rho(\mathbf{r}) \phi(\mathbf{r}) d\mathbf{r}. \quad (3.10)$$

The intrinsic Helmholtz free energy turns out to be a much more useful quantity than normal Helmholtz free energy for the description of these inhomogeneous systems.

Similarly to what we had before, \mathcal{F} can be split into two parts $\mathcal{F} = \mathcal{F}^{id} + \mathcal{F}^{ex}$, where \mathcal{F}^{id} is an ideal part, given by

$$\mathcal{F}^{id} = kT \int \rho(\mathbf{r}) \{ \ln [\Lambda^3 \rho(\mathbf{r})] - 1 \} d\mathbf{r}. \quad (3.11)$$

This is the same equation as we found in section 2.2 and \mathcal{F}^{ex} is an excess contribution (relative to an ideal gas). Notice that we dropped an internal partition function q since we are going to deal with liquids of particles without any internal structure in this section.

While the explicit form of $\mathcal{F}^{ex}[\rho]$ is usually unknown, it is possible to compute the change of \mathcal{F}^{ex} relative to some reference system. Similarly to the approach presented in section 2.3, one first splits the pair potential between densities into reference and perturbation parts:

$$u_\lambda(\mathbf{r}_1, \mathbf{r}_2) = u_0(\mathbf{r}_1, \mathbf{r}_2) + \lambda u(\mathbf{r}_1, \mathbf{r}_2). \quad (3.12)$$

By gradually increasing the perturbation part of the interaction between particles, we can find the change in the excess free energy:

$$\mathcal{F}^{ex}[\rho] - \mathcal{F}^{ex}[\rho_0] = \frac{1}{2} \int_0^1 d\lambda \iint \rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2; \lambda) u(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2. \quad (3.13)$$

The formula above forms the basis for various perturbation theories and simplifications.

Together with n-particle densities, free energy functionals form a useful set of tools to study liquid systems. The particle densities describe the structure of the liquid, while free energy functionals incorporate energetic information. A rigorous basis for these ideas, called density functional theory, is summarised in two results, called Hohenberg–Kohn–Mermin theorems^{26,74,75}.

The first theorem states that for a given μ , T and V , the equilibrium density distribution $\rho_{eq}(\mathbf{r})$ is uniquely determined by an external potential $\phi(\mathbf{r})$ acting on the system. Thus, the equilibrium particle distribution $\rho_{eq}(\mathbf{r}; \phi)$ is a unique functional of the external potential. As a result, it follows that the intrinsic free energy functional $\mathcal{F}[\rho]$ is a unique functional of the single particle density $\rho(\mathbf{r})$.

The second theorem states that equilibrium density ρ_{eq} minimizes the grand potential Ω for a given external field ϕ :

$$\left(\frac{\delta\Omega}{\delta\rho(\mathbf{r})} \right)_{\rho=\rho_{eq},\phi} = 0 \quad (3.14)$$

and

$$\Omega[\rho; \phi] \geq \Omega, \quad (3.15)$$

where equality only applies when $\rho(\mathbf{r}) = \rho_{eq}$.

These theorems and density functional theory, in general, can be applied to both quantum and classical systems. Since its original formulation in the 1960s, the theory has been mostly applied in many-electron systems^{76,77}, although quite a lot of work has also been done in the field of classical liquids^{25,27}. The field is too broad to cover completely. Thus, we will only cover the results associated with 3D-RISM and related theories, leaving more advanced approaches mostly for future work.

3.3 Functional derivatives and correlations

Particle densities and correlations can be naturally obtained as derivatives of free energy functionals. This approach provides more insight into their relation with each and with various system properties. In this section we define a number of useful functions that can be obtained by differentiating free energy functionals and relate them using Ornstein-Zernike equation.

We start by rewriting equation 3.9:

$$\begin{aligned} \Omega[\rho] &= \mathcal{F}[\rho] + \int \rho(\mathbf{r}) [\phi(\mathbf{r}) - \mu] d\mathbf{r} \\ &= \mathcal{F}[\rho] - \int \rho(\mathbf{r}) \bar{\mu}(\mathbf{r}) d\mathbf{r}. \end{aligned} \quad (3.16)$$

The intrinsic free energy functional does not explicitly depend on $\bar{\mu}$. It follows then that:

$$\frac{\delta\Omega[\rho]}{\delta\bar{\mu}(\mathbf{r})} = -\rho(\mathbf{r}), \quad (3.17)$$

where δ indicates a functional derivative. This equation can also be used as an alternative definition of a single particle density⁷⁸.

After lengthy algebraic manipulations, it is possible to show that the second derivative of the grand potential with respect to intrinsic chemical potential gives a correlation between density fluctuations²⁶:

$$\begin{aligned}
 -kT \frac{\delta^2 \Omega[\rho]}{\delta \bar{\mu}(\mathbf{r}_1) \delta \bar{\mu}(\mathbf{r}_2)} &= \chi(\mathbf{r}_1, \mathbf{r}_2) \\
 &= \langle [\rho^I(\mathbf{r}_1) - \rho(\mathbf{r}_1)] [\rho^I(\mathbf{r}_2) - \rho(\mathbf{r}_2)] \rangle \\
 &= \langle \rho^I(\mathbf{r}_1) \rho^I(\mathbf{r}_2) \rangle - \rho(\mathbf{r}_1) \rho(\mathbf{r}_2) \\
 &= \rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) + \rho(\mathbf{r}_1) \delta(\mathbf{r}_2 - \mathbf{r}_1) - \rho(\mathbf{r}_1) \rho(\mathbf{r}_2).
 \end{aligned} \tag{3.18}$$

The last equality in the above equation follows from equation 3.2. The function $\chi(\mathbf{r}_1, \mathbf{r}_2)$ is called a density-density correlation function. Note that using equation 3.17 we can rewrite the above functional derivative in a number of different ways that will be useful in the later chapters

$$\chi(\mathbf{r}_1, \mathbf{r}_2) = - \frac{\delta^2 \beta \Omega[\rho]}{\delta \beta \bar{\mu}(\mathbf{r}_1) \delta \beta \bar{\mu}(\mathbf{r}_2)} = \frac{\delta \rho(\mathbf{r}_1)}{\delta \beta \bar{\mu}(\mathbf{r}_2)}. \tag{3.19}$$

It is clear that at large separations where $r = |\mathbf{r}_2 - \mathbf{r}_1|$ goes to ∞ , the fluctuations of density become independent of one another and $\chi(\mathbf{r}_1, \mathbf{r}_2)$ should approach 0.

Let us define a total correlation function $h(\mathbf{r}_1, \mathbf{r}_2)$, given by

$$h(\mathbf{r}_1, \mathbf{r}_2) = \frac{\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) - \rho(\mathbf{r}_1) \rho(\mathbf{r}_2)}{\rho(\mathbf{r}_1) \rho(\mathbf{r}_2)} = g(\mathbf{r}_1, \mathbf{r}_2) - 1. \tag{3.20}$$

Then using equations 3.18 and 3.20

$$\chi(\mathbf{r}_1, \mathbf{r}_2) = \rho(\mathbf{r}_1) \rho(\mathbf{r}_2) h(\mathbf{r}_1, \mathbf{r}_2) + \rho(\mathbf{r}_1) \delta(\mathbf{r}_2 - \mathbf{r}_1) \tag{3.21}$$

In the absence of any inter-particle correlations (ideal gas), $\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \approx \rho(\mathbf{r}_2) \rho(\mathbf{r}_1)$ for any two points, and thus $h(\mathbf{r}_1, \mathbf{r}_2) = 0$. The total correlation function contains only non-ideal (excess) pair correlations.

An alternative family of correlation functions emerges if we take the functional derivatives of the intrinsic free energy with respect to density. Writing the first derivative we get

$$\frac{\delta \mathcal{F}}{\delta \rho(\mathbf{r})} = \frac{\delta \mathcal{F}^{id}}{\delta \rho(\mathbf{r})} + \frac{\delta \mathcal{F}^{ex}}{\delta \rho(\mathbf{r})} = \mu^{ig}(\mathbf{r}) + \bar{\mu}^{ex}(\mathbf{r}). \tag{3.22}$$

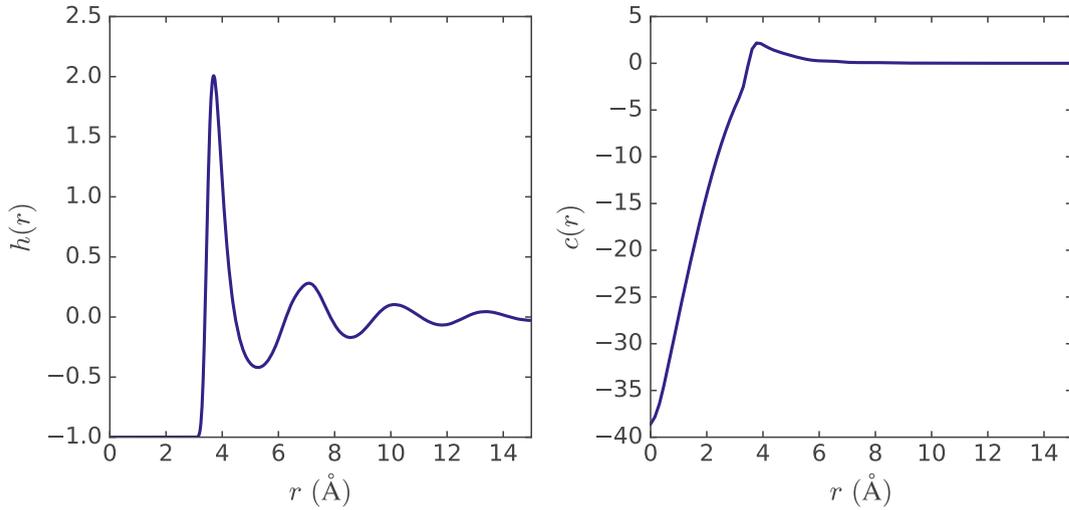


FIGURE 3.2: Total correlation function (left) and direct correlation function (right) of liquid argon near its boiling point. The correlation functions were extracted from molecular dynamics simulations performed with parameters taken from Ref. 73.

Ignoring the ideal part, we define the derivative of excess free energy as a direct correlation function:

$$c(\mathbf{r}) = -\beta \frac{\delta \mathcal{F}^{ex}[\rho]}{\delta \rho(\mathbf{r})} = -\beta \bar{\mu}^{ex}(\mathbf{r}). \quad (3.23)$$

The direct correlation function incorporates the effects of many-body interactions within the system. One can find its relationship with single particle density by considering derivative of grand potential $\frac{\delta \Omega}{\delta \rho(\mathbf{r})} = kT \ln [\Lambda^3 \rho(\mathbf{r})] - kT c(\mathbf{r}) + \phi(\mathbf{r}) = 0$. Rearranging the last result we get $\Lambda^3 \rho(\mathbf{r}) = \exp[\beta \phi(\mathbf{r}) + c(\mathbf{r})]$, which reveals the direct correlation function as a type of generalised potential.

Analogously, we can define a two-particle direct correlation function:

$$c^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = -\beta \frac{\delta^2 \mathcal{F}^{ex}}{\delta \rho(\mathbf{r}_1) \delta \rho(\mathbf{r}_2)} = \frac{\delta c(\mathbf{r}_1)}{\delta \rho(\mathbf{r}_2)}, \quad (3.24)$$

which reveals the effect of density change at \mathbf{r}_2 on the direct correlation function at \mathbf{r}_1 . The higher order direct correlation functions can be obtained in a similar way.

In isotropic liquids, the total and direct correlation functions depend only on the separation between particles. Both of them also tend to 0 as $r \rightarrow \infty$. An example of a typical behaviour of these functions for simple liquids is shown in figure 3.2.

Is it possible to relate the two types of correlation functions we defined above? The answer is yes. To obtain a meaningful relation, we start by writing down the identity that follows from the equation 3.19²⁶:

$$\int \chi(\mathbf{r}_1, \mathbf{r}_3) \chi^{-1}(\mathbf{r}_3, \mathbf{r}_2) d\mathbf{r}' = \int \frac{\delta\rho(\mathbf{r}_1)}{\delta\beta\bar{\mu}(\mathbf{r}_3)} \frac{\delta\beta\bar{\mu}(\mathbf{r}_3)}{\delta\rho(\mathbf{r}_2)} d\mathbf{r}_3 = \delta(\mathbf{r}_2 - \mathbf{r}_1), \quad (3.25)$$

where $\chi^{-1}(\mathbf{r}_1, \mathbf{r}_2)$ is the functional inverse of the density-density correlation function.

Using equation 3.22 we can express it via direct correlation function:

$$\begin{aligned} \chi^{-1}(\mathbf{r}_1, \mathbf{r}_2) &= \frac{\delta\beta\bar{\mu}(\mathbf{r}_1)}{\delta\rho(\mathbf{r}_2)} = \frac{\delta \ln \Lambda^3 \rho(\mathbf{r}_1)}{\delta\rho(\mathbf{r}_2)} - \frac{\delta c(\mathbf{r}_1)}{\delta\rho(\mathbf{r}_2)} \\ &= \frac{1}{\Lambda^3 \rho(\mathbf{r}_1)} \frac{\delta \Lambda^3 \rho(\mathbf{r}_1)}{\delta\rho(\mathbf{r}_2)} - \frac{\delta c(\mathbf{r}_1)}{\delta\rho(\mathbf{r}_2)} \\ &= \frac{1}{\rho(\mathbf{r}_1)} \delta(\mathbf{r}_1 - \mathbf{r}_2) - c(\mathbf{r}_1, \mathbf{r}_2). \end{aligned} \quad (3.26)$$

Plugging this result back into 3.25 we find:

$$\begin{aligned} &\int [\rho(\mathbf{r}_1)\rho(\mathbf{r}_3)h(\mathbf{r}_1, \mathbf{r}_3) + \rho(\mathbf{r}_1)\delta(\mathbf{r}_3 - \mathbf{r}_1)] \left[\frac{\delta(\mathbf{r}_3 - \mathbf{r}_2)}{\rho(\mathbf{r}_3)} - c(\mathbf{r}_3, \mathbf{r}_2) \right] d\mathbf{r}_3 \\ &= \int \left[\rho(\mathbf{r}_1)h(\mathbf{r}_1, \mathbf{r}_3)\delta(\mathbf{r}_3 - \mathbf{r}_2) - \rho(\mathbf{r}_1)\rho(\mathbf{r}_3)h(\mathbf{r}_1, \mathbf{r}_3)c(\mathbf{r}_3, \mathbf{r}_2) \right. \\ &\quad \left. + \frac{\rho(\mathbf{r}_1)}{\rho(\mathbf{r}_3)} \delta(\mathbf{r}_3 - \mathbf{r}_2)\delta(\mathbf{r}_3 - \mathbf{r}_1) - \rho(\mathbf{r}_1)\delta(\mathbf{r}_3 - \mathbf{r}_1)c(\mathbf{r}_3, \mathbf{r}_2) \right] d\mathbf{r}_3 \\ &= \rho(\mathbf{r}_1)h(\mathbf{r}_1, \mathbf{r}_2) - \rho(\mathbf{r}_1)c(\mathbf{r}_1, \mathbf{r}_2) - \int \rho(\mathbf{r}_1)\rho(\mathbf{r}_3)h(\mathbf{r}_1, \mathbf{r}_3)c(\mathbf{r}_3, \mathbf{r}_2) d\mathbf{r}_3 \\ &\quad + \delta(\mathbf{r}_2 - \mathbf{r}_1) = \delta(\mathbf{r}_2 - \mathbf{r}_1). \end{aligned} \quad (3.27)$$

Cancelling delta functions and dividing everything by $\rho(\mathbf{r}_1)$ we get the famous Ornstein-Zernike relation:

$$h(\mathbf{r}_1, \mathbf{r}_2) - c(\mathbf{r}_1, \mathbf{r}_2) = \int \rho(\mathbf{r}_3)h(\mathbf{r}_1, \mathbf{r}_3)c(\mathbf{r}_3, \mathbf{r}_2) d\mathbf{r}_3 \quad (3.28)$$

that will be the topic of next section.

3.4 Ornstein-Zernike and mixtures

The physical meaning of the Ornstein-Zernike equation can be understood by expressing the total correlation function in terms of direct correlation functions:

$$\begin{aligned}
 h(\mathbf{r}_1, \mathbf{r}_2) &= c(\mathbf{r}_1, \mathbf{r}_2) + \int c(\mathbf{r}_1, \mathbf{r}_3) \rho(\mathbf{r}_3) c(\mathbf{r}_3, \mathbf{r}_2) d\mathbf{r}_3 \\
 &+ \iint c(\mathbf{r}_1, \mathbf{r}_3) \rho(\mathbf{r}_3) c(\mathbf{r}_3, \mathbf{r}_4) \rho(\mathbf{r}_4) c(\mathbf{r}_4, \mathbf{r}_2) d\mathbf{r}_3 d\mathbf{r}_4 + \dots
 \end{aligned}
 \tag{3.29}$$

The total correlation of densities (particles) at point 1 and point 2 is given by direct correlation of densities plus correlation through one intermediate point, two intermediate points and so on.

For homogeneous and isotropic fluids the equation can be simplified:

$$h(r) = c(r) + \rho \int c(|\mathbf{r}' - \mathbf{r}|) h(r') d\mathbf{r}'.
 \tag{3.30}$$

We can see that as $\rho \rightarrow 0$, $h(r) \approx c(r)$, indicating that at low densities the correlation between particles is purely due to "direct" interactions between them, while as densities become larger, indirect interactions start playing a greater role.

To simplify equation 3.30 further, we need to use a Fourier transform, which we define in the next two paragraphs. In this thesis, we will use the following convention of a Fourier transform F

$$F\{f\}(\mathbf{k}) = \hat{f}(\mathbf{k}) = \int f(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r}) d\mathbf{r},
 \tag{3.31}$$

where \cdot stands for dot product. The inverse Fourier transform is given by

$$F^{-1}\{\hat{f}\}(\mathbf{r}) = f(\mathbf{r}) = \frac{1}{(2\pi)^3} \int \hat{f}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}) d\mathbf{k}.
 \tag{3.32}$$

Generally, $\hat{f}(\mathbf{k})$ is a complex valued function. However, the majority of transforms in this thesis are going to be performed on spherically symmetric functions for which $\hat{f}(\mathbf{k})$ is strictly real. Moreover, in that case, the Fourier transform simplifies and we have⁷⁹:

$$\hat{f}(k) = \frac{4\pi}{k} \int_0^{\infty} f(r) r \sin(kr) dr
 \tag{3.33}$$

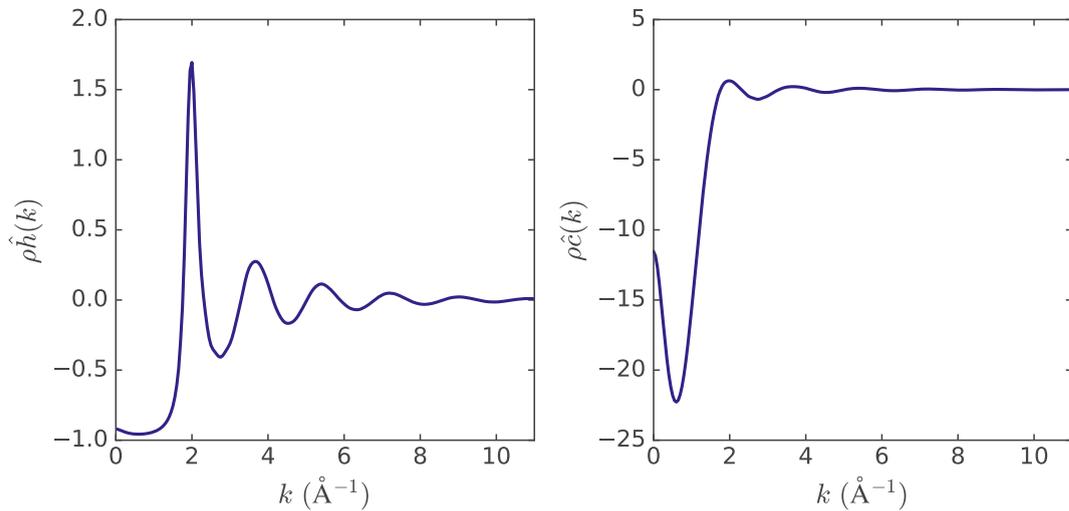


FIGURE 3.3: Fourier transforms of the total correlation function (left) and the direct correlation function (right) from figure 3.2.

and

$$f(r) = \frac{1}{2\pi r} \int_0^{\infty} \hat{f}(k) k \sin(kr) dk. \quad (3.34)$$

We will be mostly applying Fourier transforms on convolutions since they simplify considerably in k-space:

$$F \left\{ \int f(\mathbf{r}') g(\mathbf{r} - \mathbf{r}') d\mathbf{r}' \right\} (\mathbf{k}) = \hat{f}(\mathbf{k}) \hat{g}(\mathbf{k}). \quad (3.35)$$

Coming back to equation 3.30 and taking Fourier transform of both sides we can write down:

$$\hat{h}(k) = \hat{c}(k) + \rho \hat{c}(k) \hat{h}(k). \quad (3.36)$$

We obtain a simple algebraic equation, from which we can obtain expressions for $h(k)$ or $c(k)$

$$\hat{h}(k) = \frac{\hat{c}(k)}{1 - \rho \hat{c}(k)} \quad \hat{c}(k) = \frac{\hat{h}(k)}{1 + \rho \hat{h}(k)}. \quad (3.37)$$

The above results can be readily extended to mixtures. Consider a system containing n different types of particles labelled $i = 1 \dots n$. The average density of a particle of type i is given by

$$\rho_i = \frac{N_i}{V} = x_i \rho, \quad (3.38)$$

where N_i is the total number of particles of type i and x_i is their mole ratio.

While the definition of single particle density $\rho_i(\mathbf{r})$ in mixtures remains the same, the expression for two-particle density $\rho_{ij}^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$ depends on whether i and j are the same species. If $i = j$, equation 3.2 still holds, but for $i \neq j$ we no longer have to worry about the correlation of particle with itself:

$$\rho_{ij}^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \left\langle \sum_{l=1}^{N_i} \sum_{m=1}^{N_j} \delta(\mathbf{r}_1 - \mathbf{r}_l) \delta(\mathbf{r}_2 - \mathbf{r}_m) \right\rangle = \langle \rho_i^I(\mathbf{r}_1) \rho_j^I(\mathbf{r}_2) \rangle. \quad (3.39)$$

These two conditions can be summarised in a single equation using the Kronecker delta δ_{ij} :

$$\rho_{ij}^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \langle \rho_i^I(\mathbf{r}_1) \rho_j^I(\mathbf{r}_2) \rangle - \delta_{ij} \rho(\mathbf{r}_1) \delta(\mathbf{r}_1 - \mathbf{r}_2). \quad (3.40)$$

Similarly, for the density-density correlation function we have

$$\chi_{ij}(\mathbf{r}_1, \mathbf{r}_2) = \frac{\delta \rho_i(\mathbf{r}_1)}{\delta \mu_j(\mathbf{r}_2)} = h_{ij} \rho_i(\mathbf{r}_1) \rho_j(\mathbf{r}_2) + \delta_{ij} \delta(\mathbf{r}_1 - \mathbf{r}_2) \rho_i(\mathbf{r}_1). \quad (3.41)$$

The multicomponent Ornstein-Zernike equation is similar to its single component version:

$$h_{ij}(\mathbf{r}_1, \mathbf{r}_2) = c_{ij}(\mathbf{r}_1, \mathbf{r}_2) + \sum_{m=1}^n \rho_m(\mathbf{r}_3) c_{im}(\mathbf{r}_1, \mathbf{r}_3) h_{mj}(\mathbf{r}_3, \mathbf{r}_2) d\mathbf{r}_3, \quad (3.42)$$

where n is the number of components in mixture. The main difference is that in the multicomponent case we have to account for interactions through other types of species.

For a homogeneous mixture we have:

$$h_{ij}(r) = c_{ij}(r) + \sum_{m=1}^n \rho_m c_{im}(|\mathbf{r}_1 - \mathbf{r}_2|) h_{mj}(r_2) d\mathbf{r}_2 \quad (3.43)$$

in real space, and

$$h_{ij}(k) = c_{ij}(k) + \sum_{m=1}^n \rho_m c_{im}(k) h_{mj}(k) \quad (3.44)$$

in Fourier space. Summation over indices suggests a convenient matrix form:

$$\mathbf{H}(k) = \mathbf{C}(k) + \mathbf{DC}(k) \times \mathbf{H}(k), \quad (3.45)$$

where \mathbf{D} is the diagonal matrix of species densities:

$$\mathbf{D} = \begin{bmatrix} \rho_1 & 0 & 0 & \dots & 0 \\ 0 & \rho_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \rho_n \end{bmatrix}, \quad (3.46)$$

and correlation functions are grouped into matrices in which ij -th element describes the correlation between particles of type i and j : $\mathbf{C}(k) = [c_{ij}(k)]$, and $\mathbf{H}(k) = [h_{ij}(k)]$.

3.5 Linear response

In section 2.5 we discussed the linear response in the context of solvation free energy. The solute-solvent energy depended linearly on the coupling strength. Here we take a more general and microscopic approach, additionally considering the spatial dependence of the response.

Consider a uniform liquid exposed to a weak external potential $\delta\phi(\mathbf{r})$ (here we use δ to indicate that the field is small) that couples to local density in the usual way:

$$\Omega_{\delta\phi} = \Omega_0 + \int \delta\phi(\mathbf{r})\rho(\mathbf{r}) d\mathbf{r}. \quad (3.47)$$

For a weak perturbation we would expect that the density response can be described using the first order Taylor expansion:

$$\delta\rho(\mathbf{r}) = \int \left. \frac{\delta\rho(\mathbf{r})}{\delta\beta\phi(\mathbf{r}')} \right|_{\phi=0} \beta\delta\phi(\mathbf{r}') d\mathbf{r}', \quad (3.48)$$

where $\delta\rho(\mathbf{r}_1) = \rho_{\delta\phi}(\mathbf{r}_1) - \rho_0$. The density response is linear, but nonlocal. Notice that since $\phi(\mathbf{r}) = \mu - \bar{\mu}(\mathbf{r})$, we can use equation 3.19 to obtain

$$\begin{aligned} -\frac{\delta\rho(\mathbf{r}_1)}{\delta\beta\phi(\mathbf{r}_2)} &= \frac{\delta\rho(\mathbf{r}_1)}{\delta\beta\bar{\mu}(\mathbf{r}_2)} = \chi(\mathbf{r}_1, \mathbf{r}_2) \\ &= \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)h(\mathbf{r}_1, \mathbf{r}_2) + \rho(\mathbf{r}_1)\delta(\mathbf{r}_2 - \mathbf{r}_1), \end{aligned} \quad (3.49)$$

where the derivatives are assumed to be taken in the unperturbed system with $\phi = 0$. The above equation shows a connection between a density-density correlation and liquid

susceptibility to an external field. This result is quite general and is valid for all classical systems⁸⁰. Note that in statistical mechanics texts the static susceptibility is typically defined as $\chi = \beta \langle \delta A \delta A \rangle$, but in the reference interaction site model (RISM) literature the above definition is more widespread^{6,28,81}.

For simplicity, in the following, we focus on the response of an isotropic reference system. In that case liquid susceptibility depends only on the separation between two points $\chi(\mathbf{r}_1, \mathbf{r}_2) = \chi(|\mathbf{r}_2 - \mathbf{r}_1|) = \chi(r)$. Applying the convolution theorem (equation 3.35) we get

$$\widehat{\delta\rho}(\mathbf{k}) = -\beta\hat{\chi}(k)\widehat{\delta\phi}(\mathbf{k}), \quad (3.50)$$

where $k = |\mathbf{k}|$. This is a remarkable result that allows us to calculate the perturbation of liquid density due to a field with a certain periodicity. Since $\hat{\chi}(k) = \rho \left[\hat{h}(k)\rho + 1 \right]$, we can see that response is quite sensitive to the wavenumber, and can be both amplified or weakened (see figure 3.3).

Equation 3.50 suggests that there is a dual relationship between the density perturbation and the potential. We can rewrite the previous equation as

$$\beta\widehat{\delta\phi}(\mathbf{k}) = -\hat{\chi}^{-1}(k)\widehat{\delta\rho}(\mathbf{k}) \quad (3.51)$$

in which the inverse of susceptibility $\chi^{-1}(k)$ determines a field created by a periodic density modulation $\Delta\rho(\mathbf{k})$. Similarly, since $\delta\phi = -\delta\bar{\mu}$ we have

$$\beta\widehat{\delta\bar{\mu}}(\mathbf{k}) = \hat{\chi}^{-1}(k)\widehat{\delta\rho}(\mathbf{k}) \quad (3.52)$$

that describes the effect of density modulation on the intrinsic chemical potential.

Using equation 3.52 we can express the inverse of susceptibility as a functional derivative

$$\begin{aligned} \chi^{-1}(|\mathbf{r}_2 - \mathbf{r}_1|) &= -\frac{\delta\beta\phi(\mathbf{r}_1)}{\delta\rho(\mathbf{r}_2)} = \frac{\delta\beta\bar{\mu}(\mathbf{r}_1)}{\delta\rho(\mathbf{r}_2)} \\ &= \frac{1}{\rho(\mathbf{r}_1)}\delta(\mathbf{r}_2 - \mathbf{r}_1) - c(|\mathbf{r}_2 - \mathbf{r}_1|), \end{aligned} \quad (3.53)$$

where the last equality was obtained using equation 3.26. This result is similar to 3.49, providing an alternative interpretation of the total and the direct correlation functions as non-ideal components of the system response to a perturbation. Completing the parallel between susceptibilities, we express the inverse of susceptibility as a correlation function

between intrinsic chemical potentials

$$\chi^{-1}(\mathbf{r}_1, \mathbf{r}_2) = \langle \delta\beta\bar{\mu}(\mathbf{r}_1)\delta\beta\bar{\mu}(\mathbf{r}_2) \rangle, \quad (3.54)$$

where, as usual, $\delta\bar{\mu}(\mathbf{r}) = \bar{\mu}_{\delta\rho}(\mathbf{r}) - \bar{\mu}_0$. The above expression follows directly from the fluctuation-dissipation theorem⁸⁰.

These results can be readily generalised to multicomponent mixtures. The response of the local density of component i to an external field that couples to densities $1 \cdots N$ is given by

$$\delta\rho_i(\mathbf{r}) = \sum_{j=1}^N \int \frac{\delta\rho_j(\mathbf{r})}{\delta\beta\phi_j(\mathbf{r}')} \beta\delta\phi_j(\mathbf{r}') d\mathbf{r}', \quad (3.55)$$

or when written in terms of the density-density correlation functions (susceptibilities):

$$\delta\rho_i(\mathbf{r}) = -\beta \sum_{j=1}^N \int \chi_{ij}(\mathbf{r}, \mathbf{r}') \delta\phi_j(\mathbf{r}') d\mathbf{r}'. \quad (3.56)$$

As previously, it is easier to work with vectors and matrices when dealing with mixtures. Using vectors $\delta\boldsymbol{\rho}(\mathbf{r}) = [\delta\rho_1(\mathbf{r}) \cdots \delta\rho_n(\mathbf{r})]$, $\delta\boldsymbol{\phi}(\mathbf{r}) = [\delta\phi_1(\mathbf{r}) \cdots \delta\phi_n(\mathbf{r})]$, and matrix: $\mathbf{X}(\mathbf{r}_1, \mathbf{r}_2) = [\chi_{ij}(\mathbf{r}_1, \mathbf{r}_2)]$ we can rewrite the above equation as:

$$\delta\boldsymbol{\rho}(\mathbf{r}) = -\beta \int \mathbf{X}(\mathbf{r}, \mathbf{r}') \delta\boldsymbol{\phi}(\mathbf{r}') d\mathbf{r}'. \quad (3.57)$$

Assuming an isotropic system and taking the Fourier transform of all elements of matrices and vectors, we obtain the extensions of the previous results:

$$\begin{aligned} \widehat{\delta\rho}(\mathbf{k}) &= -\beta \widehat{\mathbf{X}}(k) \widehat{\delta\boldsymbol{\phi}}(\mathbf{k}), \\ \beta \widehat{\delta\boldsymbol{\phi}}(\mathbf{k}) &= -\widehat{\mathbf{X}}^{-1}(k) \widehat{\delta\rho}(\mathbf{k}), \\ \beta \widehat{\delta\bar{\mu}}(\mathbf{k}) &= \widehat{\mathbf{X}}^{-1}(k) \widehat{\delta\rho}(\mathbf{k}). \end{aligned} \quad (3.58)$$

Note that unlike equation 3.51, in which we were dealing with the algebraic inverse of susceptibility, in the above we have the matrix inverse of the susceptibility matrix. Similarly to the single component case, these results are valid only for small external fields $\delta\boldsymbol{\phi}$ and density perturbations $\delta\boldsymbol{\rho}$.

3.6 Connection to experiment

The integrals of correlation functions are related to a variety of thermodynamic properties of liquids and can be used to construct an equation of state. First, let us relate direct correlation function to the average fluctuations of particles.

For single and pair local densities we have the following normalization conditions $\int_V \rho(\mathbf{r}) d\mathbf{r} = \langle N \rangle_V$ and $\iint_V \rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 = \langle N^2 \rangle_V - \langle N \rangle_V^2$. Combining these two equations with 3.20 we find that:

$$\begin{aligned} \iint_V h(\mathbf{r}_1, \mathbf{r}_2) \rho(\mathbf{r}_1) \rho(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 &= \int \rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) - \rho(\mathbf{r}_1) \rho(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \\ &= \langle N^2 \rangle_V - \langle N \rangle_V^2 - \langle N \rangle_V^2 \\ &= \langle (\delta N)^2 \rangle_V - \langle N \rangle_V^2, \end{aligned} \quad (3.59)$$

where $\delta N = N - \langle N \rangle$ and $\langle (\delta N)^2 \rangle_V$ is the mean square fluctuation of a number of particles in volume V . In a translationally invariant medium the particle density is constant and the direct correlation function depends only on distance r , so $\iint \rho(\mathbf{r}_1) \rho(\mathbf{r}_2) h(|\mathbf{r}_2 - \mathbf{r}_1|) d\mathbf{r}_1 d\mathbf{r}_2 = N \rho \int h(r) dr$. Rearranging the two previous equations we get:

$$1 + \rho \int h(r) dr = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle} = kT \rho \chi_T, \quad (3.60)$$

where χ_T is the isothermal compressibility. The last equality can be proven using $\chi_T = (\partial \rho / \partial \mu)_{V,T} / \rho^2$ and the definition of the grand canonical partition function⁸².

We can also express compressibility using the direct correlation functions. Notice that since the integral over real space is equal to the value of the Fourier-transformed function at $k = 0$: $\int f(r) dr = \int f(r) e^{-0 \cdot r} dr = \hat{f}(k = 0)$, we can write:

$$kT \rho \chi_T = 1 + \rho \hat{h}(k = 0) = \frac{1}{1 - \rho \hat{c}(k = 0)}, \quad (3.61)$$

where the last equality follows from equation 3.36.

The above equation provides a connection between the integral of the direct correlation function and the pressure of liquid⁸³. Using the definition of isothermal compressibility

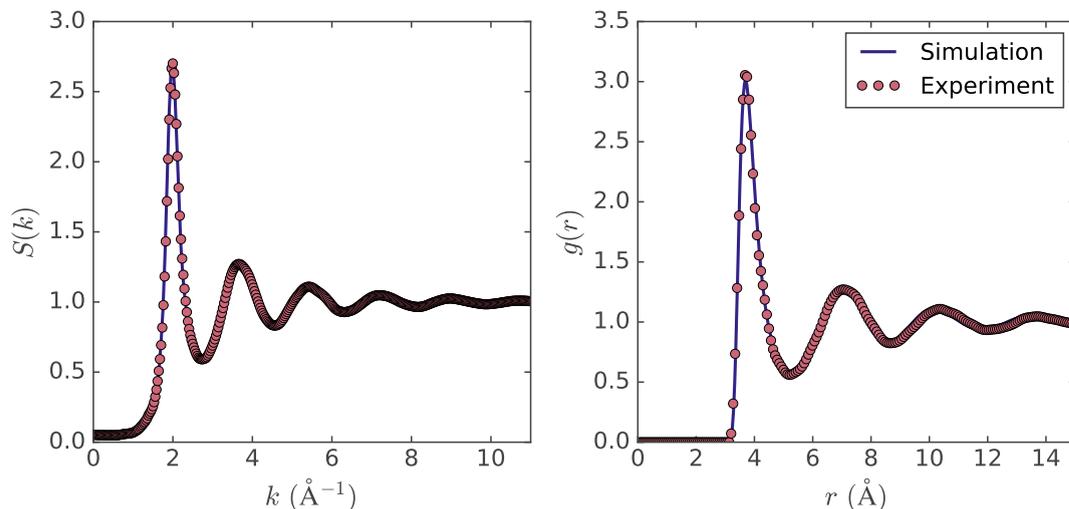


FIGURE 3.4: Comparison of experimental and theoretical structure factors (left) and radial distribution functions (right) for argon at 85K (close to its boiling point). Experimental data is taken from Ref. 86. Theoretical predictions are obtained using molecular dynamics, with simulation parameters taken from Ref. 73.

we get $\rho\chi_T = (\partial\rho/\partial P)_{T,V}$. Plugging this back into equation 3.61 we obtain

$$\left(\frac{\partial P}{\partial\rho}\right)_{T,V} = kT [1 - \rho\hat{c}(k=0; \rho)]. \quad (3.62)$$

Finally, integrating between initial and final densities ρ_1 and ρ_2 we obtain the equation of state

$$\Delta P = kT\Delta\rho - \int_{\rho_1}^{\rho_2} [\rho\hat{c}(k=0; \rho)] d\rho, \quad (3.63)$$

where $\Delta P = P_2 - P_1$ and $\Delta\rho = \rho_2 - \rho_1$. Thus, we have related experimentally measurable changes in pressure to the changes in the integral of the direct correlation function. The Kirkwood-Buff theory allows us to extend these results to multicomponent mixtures⁸⁴. These, more general relationships, have been used for validating both theoretical and experimental measurements of pressure in mixtures of various liquids⁸⁵.

Finally, we note that it is possible to measure correlation functions directly using spectroscopy²⁶. The idea is to perturb bulk liquid using small-angle neutron or x-ray scattering (radiation with longer wavelengths is too coarse to probe atomic structure of liquids and describes more macroscopic properties). The scattering cross section for the wave vector k is proportional to the liquid structure factor $S(k)$, with the proportionality constant depending on the experimental set-up⁸⁶. In the case of simple liquids, the structure factor

$S(k)$ is related to the total correlation function as:

$$S(k) = 1 + \rho \int h(r) e^{-i\mathbf{k}\cdot\mathbf{r}} d\mathbf{r} = 1 + \rho h(k). \quad (3.64)$$

Notice that the structure factor is directly related to the Fourier transform of susceptibility $S(k) = \chi(k)/\rho$.

Figure 3.4 shows the argon structure factor and radial distribution functions deduced from the former using the above equation. The experimental results, taken from Ref. 86, are compared to the molecular dynamics simulation. Both experimental measurements and theoretical simulations were performed at 85 K with argon density $\rho = 0.021 \text{ \AA}^{-3}$. The theory and experiment agree almost precisely, which might be expected, considering that the Lennard-Jones potential constants used in simulations were fitted to reproduce the experimental structure factor.

Chapter 4

Theory of molecular liquids

This chapter overviews approximations required to compute the free energy of molecular liquids in site-site formalism. The 3D-RISM equations and closures are derived in detail starting from the results from Refs. 81,87. We finish by introducing the main problem of the thesis.

4.1 Hypernetted-chain approximation

All results in the last chapter were exact. However, they do not provide a tractable way of computing free energies – to do so one needs to introduce some kind of approximation. Although a variety of approaches exist²⁶, we are going to consider one of the most basic approximations, built on the expansion of free energy functional around its equilibrium value. For clarity, in this section we are going to focus on its derivation from the viewpoint of simple liquids; the case of molecular liquids is described for example in Ref. 37.

In section 3.2 we obtained the expression for a grand canonical functional:

$$\Omega[\rho] = \mathcal{F}[\rho] + \int \rho(\mathbf{r})\phi(\mathbf{r}) d\mathbf{r} - \mu \int \rho(\mathbf{r}) d\mathbf{r}, \quad (4.1)$$

where the intrinsic free energy functional was given by

$$\begin{aligned} \mathcal{F}[\rho] = & kT \int \rho(\mathbf{r}) \{ \ln [\Lambda^3 \rho(\mathbf{r})] - 1 \} d\mathbf{r} + \mathcal{F}_0^{ex}[\rho] \\ & + \frac{1}{2} \int_0^1 d\lambda \iint \rho^{(2)}(\mathbf{r}, \mathbf{r}'; \lambda) u_\lambda(\mathbf{r}, \mathbf{r}') d\mathbf{r} d\mathbf{r}'. \end{aligned} \quad (4.2)$$

The coupling constant λ linearly interpolates between initial and final interaction strength between individual particles:

$$u_\lambda(\mathbf{r}, \mathbf{r}') = u_0(\mathbf{r}, \mathbf{r}') + \lambda u(\mathbf{r}, \mathbf{r}'), \quad (4.3)$$

similarly to how a single particle was slowly coupled to the rest of the system when computing a chemical potential (see section 2.3).

While the algebraic expression for the functional is relatively straightforward, it turns out that its application to realistic systems is essentially impossible. Indeed, to do it one would need to know the dependence of pair density $\rho(\mathbf{r}, \mathbf{r}'; \lambda)$ on λ , which is clearly very non-linear²⁶.

A simpler approach is to take the problematic excess part of free energy functional and expand it into Taylor series around an isotropic system with density ρ_0 . This approach is often referred to as the (functional) density expansion^{26,27} and it leads to

$$\begin{aligned} \mathcal{F}^{ex}[\rho] = & \mathcal{F}^{ex}[\rho_0] + \int \left. \frac{\delta \mathcal{F}^{ex}}{\delta \rho(\mathbf{r})} \right|_{\rho=\rho_0} \Delta \rho(\mathbf{r}) d\mathbf{r} \\ & + \frac{1}{2} \iint \left. \frac{\delta^2 \mathcal{F}^{ex}}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')} \right|_{\rho=\rho_0} \Delta \rho(\mathbf{r}) \Delta \rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}' + F^B[\rho], \end{aligned} \quad (4.4)$$

where F^B contains all higher order terms. Note that we defined partial derivatives of excess chemical potential in the section 3.3. Rewriting the expression using the definitions of the direct correlation function we obtain

$$\begin{aligned} \mathcal{F}^{ex}[\rho] = & \mathcal{F}^{ex}[\rho_0] + \mu_0^{ex} \int \Delta \rho(\mathbf{r}) d\mathbf{r} \\ & - \frac{1}{2} kT \iint \Delta \rho(\mathbf{r}) \Delta \rho(\mathbf{r}') c_0(|\mathbf{r}' - \mathbf{r}|) d\mathbf{r} d\mathbf{r}' + F^B[\rho], \end{aligned} \quad (4.5)$$

where we used the fact that in a homogeneous system $c(\mathbf{r})$ should be constant. Substituting the above expression back into the intrinsic free energy functional and using $\mu_0 =$

$\mu_0^{ex} + kT \ln \rho_0 \Lambda^3$ we get

$$\begin{aligned} \mathcal{F}[\rho] &= kT \int \rho(\mathbf{r}) \{ \ln [\Lambda^3 \rho(\mathbf{r})] - 1 \} d\mathbf{r} \\ &+ \mathcal{F}_0^{ex}[\rho_0] + \mu_0 \int \Delta\rho(\mathbf{r}) d\mathbf{r} - kT \int \Delta\rho(\mathbf{r}) \ln \rho_0 \Lambda^3 d\mathbf{r} \\ &- \frac{1}{2} kT \iint \Delta\rho(\mathbf{r}) \Delta\rho(\mathbf{r}') c_0(|\mathbf{r}' - \mathbf{r}|) d\mathbf{r} d\mathbf{r}' + F^B[\rho]. \end{aligned} \quad (4.6)$$

Finally, we cancel out the terms containing the de Broglie wavelength and plug the above expression into equation 4.1 to obtain

$$\begin{aligned} \Omega[\rho] &= \Omega_0 + kT \int \left[\rho(\mathbf{r}) \ln \frac{\rho(\mathbf{r})}{\rho_0} - \Delta\rho(\mathbf{r}) \right] d\mathbf{r} + \int \rho(\mathbf{r}) \phi(\mathbf{r}) d\mathbf{r} \\ &- \frac{1}{2} kT \iint \Delta\rho(\mathbf{r}) c_0(|\mathbf{r}' - \mathbf{r}|) \Delta\rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}' + F^B[\rho]. \end{aligned} \quad (4.7)$$

where $\Omega_0 = kT \int \rho_0 \ln \rho_0 \Lambda^3 d\mathbf{r} + \mathcal{F}^{ex}[\rho_0] - \mu_0 \int \rho_0 d\mathbf{r}$. The five terms in the above expression correspond to reference, ideal, external, second order, and higher order excess contributions to the grand canonical functional. As one may guess, the $F^B[\rho]$ term, incorporating all higher order functional derivatives is generally unknown³⁰.

We can find the density distribution $\rho(\mathbf{r})$ which minimizes equation 4.7. To do it, we take a functional derivative with respect to density

$$\begin{aligned} \frac{\delta\Omega[\rho]}{\delta\rho(\mathbf{r})} &= kT \ln \frac{\rho(\mathbf{r})}{\rho_0} + \phi(\mathbf{r}) - \frac{1}{2} kT \int c_0(|\mathbf{r}' - \mathbf{r}|) \Delta\rho(\mathbf{r}') d\mathbf{r}' \\ &- \frac{1}{2} kT \int \Delta\rho(\mathbf{r}) c_0(|\mathbf{r}' - \mathbf{r}|) d\mathbf{r} + \frac{\delta\mathcal{F}^B[\rho]}{\delta\rho(\mathbf{r})} = 0. \end{aligned} \quad (4.8)$$

It follows that

$$\rho(\mathbf{r}) = \rho_0 \exp \left[-\beta\phi(\mathbf{r}) + \int \Delta\rho(\mathbf{r}') c_0(|\mathbf{r}' - \mathbf{r}|) d\mathbf{r}' - \frac{\delta\mathcal{F}^B[\rho]}{\delta\rho(\mathbf{r})} \right]. \quad (4.9)$$

This is an integral equation for $\rho(\mathbf{r})$. One way of solving it is to make the following approximation:

$$\frac{\delta F^B[\rho]}{\delta\rho(\mathbf{r})} = B(\mathbf{r}) = 0, \quad (4.10)$$

where in the context of integral equation theories $B(\mathbf{r})$ is referred to as a bridge function. The above assumption is called the hypernetted-chain (HNC) approximation (the name comes from the original derivation⁸⁸, which considered cluster diagrammatic expansion of the configurational partition function). Extracting $c_0(r)$ from an equilibrium simulation,

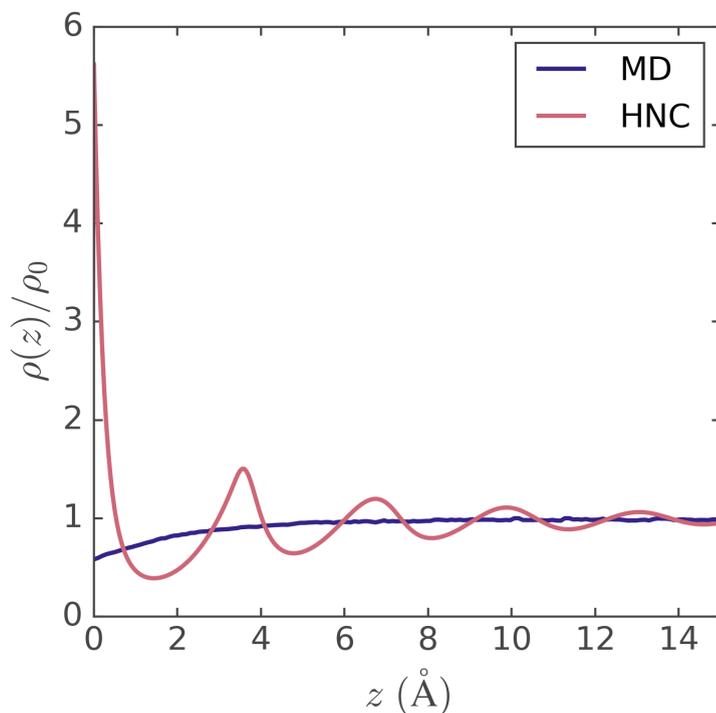


FIGURE 4.1: Liquid argon local density near hard wall. Both singlet HNC calculations (red line) as well as molecular dynamics simulations (blue line) are performed at standard pressure and 85 K. The details of the calculations can be found in appendix A.

we can find an equation with a single unknown, which is possible to solve iteratively for a given external potential $\phi(\mathbf{r})$. When combined with c_0 , the method is often referred to as the singlet HNC or HNC1 approximation, since both ρ and u depend only on a single coordinate.

An obvious question to ask is how good such an approximation is? The answer is not too good. While it can reproduce density oscillations of hard spheres near a wall, when it comes to liquids with both attractive and repulsive forces, HNC fails quite significantly²⁷. As a demonstration, consider the results obtained for liquid argon near a hard wall (figure 4.1). While molecular dynamics predicts a slight decrease in density near the wall, caused by the formation of an interface, HNC predicts oscillations similar to what is observed for liquids with purely repulsive interactions.

It is not hard to rationalize the failure of HNC to describe the formation of an interface. Since HNC only retains second order correlations in liquids, it captures its essential characteristic, namely, the dominant repulsive forces between particles with rather weak attractive components⁸⁹. As we shall see, while this does not usually present a problem

for describing a bulk liquid structure, the formation of interfaces is mostly due to long ranged attractive potentials which require a more sophisticated description. Additionally, a hard planar wall creates a huge region of excluded volume. This can hardly justify the use of a second order Taylor expansion that is valid only for small changes in density.

Before moving on it is useful to relate the singlet HNC approximation to the linear response approach, discussed in section 3.5. From the equation 4.7 we find

$$\begin{aligned} \Delta\mathcal{F}_{HNC}[\rho] &= kT \int \left[\rho(\mathbf{r}) \ln \frac{\rho(\mathbf{r})}{\rho_0} - \Delta\rho(\mathbf{r}) \right] d\mathbf{r} \\ &\quad - \frac{1}{2}kT \iint \Delta\rho(\mathbf{r}) c_0(|\mathbf{r}' - \mathbf{r}|) \Delta\rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}'. \end{aligned} \quad (4.11)$$

Then the change in chemical potential is given by

$$\Delta\bar{\mu}(\mathbf{r}) = \frac{\delta\mathcal{F}}{\delta\rho(\mathbf{r})} = kT \ln \frac{\rho(\mathbf{r})}{\rho_0} - kT \int c_0(|\mathbf{r} - \mathbf{r}'|) \Delta\rho(\mathbf{r}') d\mathbf{r}'. \quad (4.12)$$

Contrast this to a linear response change in intrinsic chemical potential obtained via equation 3.52:

$$\begin{aligned} \Delta\bar{\mu}(\mathbf{r}) &= kT \int \chi^{-1}(|\mathbf{r} - \mathbf{r}'|) \Delta\rho(\mathbf{r}') d\mathbf{r}' \\ &= kTh(\mathbf{r}) - kT \int c_0(|\mathbf{r} - \mathbf{r}'|) \Delta\rho(\mathbf{r}') d\mathbf{r}', \end{aligned} \quad (4.13)$$

where two terms in the second equality correspond to the local and non-local contributions to the change in intrinsic chemical potential. We can see that HNC model approximates the excess part of the chemical potential via the linear response approximation, while the ideal (local) contribution is exact. From this, it is reasonable to suggest that the singlet HNC should give somewhat more accurate results than the standard linear response approach, but they are not going to be significantly different. At the chapter 5 we will see that this is exactly what happens.

The failure to describe interfaces does not render HNC useless. Notably, it can quite accurately describe the bulk structure of simple liquids. To do so, we use an idea by Percus²⁶ and treat equation 4.9 as an expression for the pair correlation function $g(\mathbf{r}) = \rho(\mathbf{r})/\rho_0$ in the homogeneous liquid

$$g(r) = \exp \left[-\beta u(r) + \rho \int h(r') c(|r - r'|) dr' \right], \quad (4.14)$$

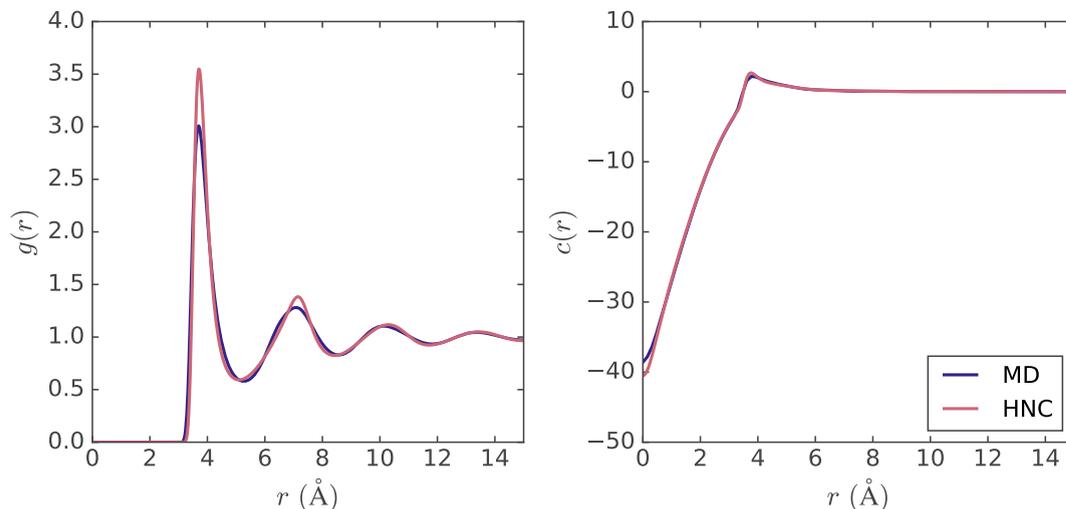


FIGURE 4.2: Comparison of liquid argon radial distribution functions (left figure) and direct correlation functions (right figure) predicted by molecular dynamics simulation (blue lines) and HNC calculation (red lines). For both simulation and theory interactions between argon atoms was approximated using Lennard-Jones potential from Ref. 73.

where we used $\Delta\rho(r) = \rho h(r)$ and the external potential takes a meaning of a pair potential $u(r)$ between particles. Using the Ornstein-Zernike equation (3.30) we obtain

$$g(r) = \exp[-\beta u(r) + h(r) - c(r)]. \quad (4.15)$$

When the above two equations are solved simultaneously, they are sometimes referred to as the pair HNC approximation (HNC2) to distinguish from its singlet form. The resulting solution gives direct and total correlation functions for isotropic liquids. Note that whenever we have equations that are solved for correlation functions (such as equations 4.15 and 4.9), we view them as an integral equation approach.

To test the accuracy of the model we apply the above equation to our model system of liquid argon. Both molecular dynamics simulation and HNC calculations were performed using the same Lennard-Jones potential and conditions. The results, shown in figure 4.2, demonstrate a surprisingly good agreement between the two. Compared to the hard wall, a single argon atom creates a much less significant perturbation. Additionally, an attractive r^{-6} part of the pair potential prevents any "drying" of the liquid close to the argon surface, making MD and HNC radial distribution functions quite similar.

4.2 Molecular liquids

So far, we have dealt only with simple (monoatomic) liquids. This made the discussion significantly easier and essentially cut the dimensionality of the problem by two. Indeed, to fully specify the position of a rigid molecule in space, one not only has to keep track of its spatial location \mathbf{r} but also of its orientation, which is typically specified using Euler angles ω ⁹⁰. Alternatively, one can keep track of the positions of individual atoms (or more generally, molecular sites) in the molecules that interact with each other. While this approach can significantly increase the dimensionality in the case of polyatomic liquids, it allows us to write down extensions of the equations used to describe simple monoatomic fluids. In this thesis, we will be mostly using the second approach.

While the interacting sites approach to predicting liquid structure and properties is relatively old⁹¹⁻⁹³, the first papers in which it was formalised as a part of density functional theory were written by Chandler, McCoy, and Singer (CMS) only in 1986^{81,87}. This theory provides a basis for various interaction site models such as reference interaction site model (RISM) or polymer reference interaction site model (PRISM).

In general, we have a mixture of M different molecules containing $n_1 + n_2 + \dots + n_M = N$ different sites. For some molecule α , each site has some local density $\rho_{i\alpha}(\mathbf{r})$ that couples to an external field $\phi_{i\alpha}(\mathbf{r})$. We will use Greek letters α, β, \dots to indicate different types of molecule in the system, and indices i, j, \dots to specify sites. For simplicity, we assume that each site in a molecule is unique (thus, water, for example, will have two distinct hydrogens) and has the same (macroscopic) density as the molecule itself $\rho_{1\alpha} = \rho_{2\alpha} = \dots = \rho_{n_\alpha\alpha} = \rho_\alpha$.

In CMS theory, we set the chemical potential of molecule α to be equal to the sum of chemical potentials of its individual sites:

$$\mu_\alpha = \sum_{i=1}^{n_\alpha} \mu_{i\alpha}. \quad (4.16)$$

This assumption can be regarded as a condition for chemical equilibrium. We also require all molecules to be rigid. This condition is not very problematic, as different conformations can be treated as different species.

The Grand canonical functional for such a system can be written in terms of the site densities⁹⁴

$$\begin{aligned}\Omega[\boldsymbol{\rho}] &= \mathcal{F}[\boldsymbol{\rho}] + \sum_{i=1}^N \int \rho_i(\mathbf{r}) \phi_i(\mathbf{r}) d\mathbf{r} - \sum_{i=1}^N \int \mu_i \rho_i(\mathbf{r}) d\mathbf{r} \\ &= \mathcal{F}[\boldsymbol{\rho}] - \sum_{i=1}^N \int \rho_i(\mathbf{r}) \bar{\mu}_i(\mathbf{r}) d\mathbf{r},\end{aligned}\tag{4.17}$$

where $\boldsymbol{\rho}$ is the vector of all site density distributions and \mathcal{F} is the intrinsic free energy functional for polyatomic liquids. For clarity, we suppressed molecule subscripts.

We can formally separate \mathcal{F} from equation 4.17 into an ideal gas part and excess. Even for a mixture of polyatomic molecules, an ideal gas is still defined to be the uniform mixture of all particles, with absolutely no correlations present⁸¹. Thus, the following holds

$$\begin{aligned}\mathcal{F}[\boldsymbol{\rho}] &= \mathcal{F}^{id}[\boldsymbol{\rho}] + \mathcal{F}^{ex}[\boldsymbol{\rho}] \\ &= \sum_{i=1}^N kT \int \rho_i(\mathbf{r}) \{ \ln [\Lambda_i^3 \rho_i(\mathbf{r})] - 1 \} d\mathbf{r} + \mathcal{F}^{ex}[\boldsymbol{\rho}].\end{aligned}\tag{4.18}$$

Here, \mathcal{F}^{ex} contains corrections not only for non-ideality, but also, for the fact that the sites belonging to the same molecule are actually bonded.

The intrinsic chemical potentials of each site are obtained by taking the derivative of grand potential with respect to the site densities:

$$\bar{\mu}_i(\mathbf{r}) = \frac{\delta \mathcal{F}}{\delta \rho_i(\mathbf{r})} = \mu_i^{id}(\mathbf{r}) + \bar{\mu}_i^{ex}(\mathbf{r}) = kT \ln [\Lambda_i^3 \rho_i(\mathbf{r})] - kT c_i(\mathbf{r}),\tag{4.19}$$

where we introduce the direct correlation function in the site formalism, defined as:

$$c_i(\mathbf{r}) = -\beta \frac{\delta \mathcal{F}^{ex}}{\delta \rho_i(\mathbf{r})}.\tag{4.20}$$

The density distributions of individual sites can be obtained by taking the functional derivative of the Grand potential

$$\rho_i(\mathbf{r}) = -\frac{\delta \Omega}{\delta \bar{\mu}_i(\mathbf{r})}.\tag{4.21}$$

Similarly to the simple fluid case, the second derivative of grand potential gives (site) density-(site) density correlation function or site-site susceptibility⁸¹:

$$\chi_{ij}(\mathbf{r}_1, \mathbf{r}_2) = -\frac{\delta\beta\Omega}{\delta\beta\bar{\mu}_i(\mathbf{r}_1)\delta\beta\bar{\mu}_j(\mathbf{r}_2)} = \frac{\delta\rho_i(\mathbf{r}_1)}{\delta\beta\bar{\mu}_j(\mathbf{r}_2)} = \langle\delta\rho_i(\mathbf{r}_1)\delta\rho_j(\mathbf{r}_2)\rangle, \quad (4.22)$$

with $\delta\rho(\mathbf{r}) = \rho^I(\mathbf{r}) - \rho(\mathbf{r})$. Repeating manipulations presented in the beginning of section 3.3, we can relate $\chi_{ij}(\mathbf{r}_1, \mathbf{r}_2)$ to the pair distribution function $\rho_{ij}^{(2)}(\mathbf{r}_1, \mathbf{r}_2)$:

$$\begin{aligned} \chi_{ij} &= \rho_{ij}^{(2)}(\mathbf{r}_1, \mathbf{r}_2) + \delta_{ij}\delta(\mathbf{r}_2 - \mathbf{r}_1)\rho(\mathbf{r}_1) - \rho_i(\mathbf{r}_1)\rho_j(\mathbf{r}_2) \\ &= \delta_{ij}\delta(\mathbf{r}_2 - \mathbf{r}_1)\rho_i(\mathbf{r}_1) + \rho_i(\mathbf{r}_1)\rho_j(\mathbf{r}_2)H_{ij}(\mathbf{r}_1, \mathbf{r}_2), \end{aligned} \quad (4.23)$$

where $H_{ij} = \rho_{ij}^{(2)}(\mathbf{r}_1, \mathbf{r}_2)/\rho_i(\mathbf{r}_1)/\rho_j(\mathbf{r}_2) - 1$, will be referred to as a whole total correlation function. Equation 4.23 suggests that H_{ij} should be similar to its simple fluid analogue: total correlation function h_{ij} . However, there is an important difference: while h_{ij} contains only *intermolecular* correlations between different particles, H_{ij} contains both *inter-* and *intramolecular* correlations.

Similarly to atomic liquids, $\chi_{ij}(\mathbf{r}_1, \mathbf{r}_2)$ determines the linear response of a system to a perturbing external field. Its functional inverse then characterizes fluctuations of chemical potential (section 3.5) $\chi_{ij}^{-1}(\mathbf{r}_1, \mathbf{r}_2) = \langle\delta\beta\bar{\mu}_i\delta\beta\bar{\mu}_j\rangle$. We can split the correlations into ideal/excess parts by defining an analogue of the two-particle direct correlation function, which we will call the *whole* site-site direct correlation function C_{ij} ⁸¹:

$$\begin{aligned} \chi_{ij}^{-1}(|\mathbf{r}_2 - \mathbf{r}_1|) &= \frac{\delta\beta\bar{\mu}_i(\mathbf{r}_2)}{\delta\rho_j(\mathbf{r}_1)} = \beta\frac{\delta\mu_i^{id}(\mathbf{r}_2)}{\delta\rho_j(\mathbf{r}_1)} + \beta\frac{\delta\bar{\mu}_i^{ex}(\mathbf{r}_2)}{\delta\rho_j(\mathbf{r}_1)} \\ &= \frac{\delta_{ij}\delta(\mathbf{r}_2 - \mathbf{r}_1)}{\rho_i} - C_{ij}(\mathbf{r}_1, \mathbf{r}_2). \end{aligned} \quad (4.24)$$

Here again, $C_{ij}(\mathbf{r}_1, \mathbf{r}_2)$ is similar to its simple fluid analogue, but in addition to intramolecular also contains intermolecular correlations.

Combining all definitions introduced in this section, we can arrive at an Ornstein-Zernike-like expression for site-site polyatomic fluids. Let us assume a homogeneous liquid and write the definition of the functional inverse for the site-site susceptibility⁹⁵:

$$\sum_{k=1}^N \int \chi_{ik}(\mathbf{r}_1 - \mathbf{r}')\chi_{kj}^{-1}(\mathbf{r}' - \mathbf{r}_2) d\mathbf{r}' = \delta_{ij}\delta(\mathbf{r}_1 - \mathbf{r}_2). \quad (4.25)$$

In the above equation, the left hand side contains a convolution of two functions. Taking the Fourier transform of both sides we get $\sum_{k=1}^N \hat{\chi}_{ik}(\mathbf{k}) \hat{\chi}_{kj}^{-1}(\mathbf{k}) = \delta_{ij}$, which can be readily expressed in a matrix form $\hat{\mathbf{X}}(\mathbf{k}) \hat{\mathbf{X}}^{-1}(\mathbf{k}) = \mathbf{I}$, where $\hat{\mathbf{X}} = [\hat{\chi}_{ij}(\mathbf{k})]$ is the N by N matrix of Fourier transformed site-site susceptibilities and $\hat{\mathbf{X}}^{-1}(\mathbf{k})$ is its matrix inverse.

Taking Fourier transform of expressions 4.23 and 4.24 we get

$$\hat{\chi}_{ij}(\mathbf{k}) = \delta_{ij} \rho_i + \rho_i \rho_j \hat{H}_{ij}(\mathbf{k}) \quad \hat{\chi}_{ij}^{-1}(\mathbf{k}) = \frac{\delta_{ij}}{\rho_i} - \hat{C}_{ij}(\mathbf{k}). \quad (4.26)$$

It follows that the equation 4.25 can be re-written as:

$$\left[\mathbf{D}\mathbf{I} + \mathbf{D}\hat{\mathbf{H}}(\mathbf{k})\mathbf{D} \right] \left[\mathbf{D}^{-1}\mathbf{I} - \hat{\mathbf{C}}(\mathbf{k}) \right] = \left[\mathbf{I} + \hat{\mathbf{H}}(\mathbf{k})\mathbf{D} \right] \left[\mathbf{I} - \hat{\mathbf{C}}(\mathbf{k})\mathbf{D} \right] = \mathbf{I}. \quad (4.27)$$

This is a site-site Ornstein-Zernike equation, written in terms of whole correlation functions⁹⁶.

Essentially, all equations in this section were definitions, based on few assumptions. These equations do not offer any insight into how to compute free energies or correlation. Similarly to the situation with simple liquids, to make any actual predictions we will have to make some guesses regarding free energy functionals.

4.3 Intramolecular correlation function

There is a special type of correlation that is not present in the case of simple liquids. These are correlations due to intramolecular bonding. In this section, we will briefly discuss them and explore their behaviour using the example of water.

Instead of combining all inter- and intramolecular correlations of $\chi_{ij}(\mathbf{r}_1, \mathbf{r}_2)$ in a single term, we can split them into two separate functions:

$$\chi_{ij}(\mathbf{r}_1, \mathbf{r}_2) = \rho_i(\mathbf{r}_1) \omega_{ij}(\mathbf{r}_1, \mathbf{r}_2) + \rho_i(\mathbf{r}_1) \rho_j(\mathbf{r}_2) h_{ij}(\mathbf{r}_1, \mathbf{r}_2), \quad (4.28)$$

where $\omega_{ij}(\mathbf{r}_1, \mathbf{r}_2)$ is called the intramolecular correlation function and contains correlations of a site with itself or with its bonded neighbours. $h_{ij}(\mathbf{r}_1, \mathbf{r}_2)$ is called the site-site total correlation function and contains site-site correlations between different molecules.

For the next couple of paragraphs we return to the full notation of sites and write $i\alpha$ to indicate site i in the molecule of type α .

Formally, $\omega_{i\alpha j\beta}(\mathbf{r}_1, \mathbf{r}_2)$ is defined via the following expression³⁰:

$$\omega_{i\alpha j\beta}(\mathbf{r}_1, \mathbf{r}_2) = \delta_{i\alpha j\beta} \delta(\mathbf{r}_2 - \mathbf{r}_1) + \delta_{\alpha\beta} (1 - \delta_{ij}) \left\langle \sum_{\alpha=1}^M \delta(\mathbf{r}_1 - \mathbf{r}_{i\alpha}) \delta(\mathbf{r}_2 - \mathbf{r}_{j\alpha}) \right\rangle, \quad (4.29)$$

where M stands for total number of distinct molecules in the system and $r_{i\alpha}$ is the position of the site i on the molecule α . For isotropic liquid we can simplify this equation to get

$$\omega_{i\alpha j\beta}(r) = \delta_{i\alpha j\beta} \delta(r) + \delta_{\alpha\beta} (1 - \delta_{ij}) \frac{\delta(r - L_{i\alpha j\alpha})}{4\pi L_{i\alpha j\alpha}^2}, \quad (4.30)$$

with $r = |\mathbf{r}_2 - \mathbf{r}_1|$, and $L_{i\alpha j\alpha}$ is the distance between sites i and j in a rigid molecule of type α . The $4\pi L_{i\alpha j\alpha}^2$ term in denominator ensures correct normalisation of the function.

In the absence of any intermolecular interactions $h_{ij} = 0$, thus $\chi_{ij}(r) = \rho_j \omega_{ij}(r)$, or in matrix notation $\mathbf{X}(r) = \boldsymbol{\omega}(r)\mathbf{D}$, where $\boldsymbol{\omega}(r) = [\omega_{ij}(r)]$. Then, from the equation 4.27 we have $\mathbf{I} - \hat{\mathbf{C}}(k)\mathbf{D} = \hat{\boldsymbol{\omega}}^{-1}(k)$, where $\hat{\boldsymbol{\omega}}^{-1}(k)$ is the matrix inverse of $\hat{\boldsymbol{\omega}}(k) = [\hat{\omega}_{ij}(k)]$. Conceptually, $\hat{\boldsymbol{\omega}}^{-1}(k)$ is the intramolecular part of $\mathbf{I} - \hat{\mathbf{C}}(k)\mathbf{D}$, similarly to how $\hat{\boldsymbol{\omega}}(k)$ contains the intramolecular part of $\mathbf{I} + \hat{\mathbf{H}}(k)\mathbf{D}$ ⁹⁶. We can then formally write

$$\mathbf{I} - \hat{\mathbf{C}}(k)\mathbf{D} = \hat{\boldsymbol{\omega}}^{-1}(k) - \hat{\mathbf{c}}(k)\mathbf{D}, \quad (4.31)$$

where $\hat{\mathbf{c}}(k) = [\hat{c}_{ij}(k)]$ is a matrix of site-site direct correlation function — an intermolecular part of $C_{ij}(r)$ defined via this equation.

The content of the last couple of paragraphs can be summarised by the following two equations:

$$\begin{aligned} \mathbf{D}^{-1}\mathbf{X}(r) &= \delta(r)\mathbf{I} + \mathbf{H}(r)\mathbf{D} = \boldsymbol{\omega}(r) + \mathbf{h}(r) \\ \mathbf{X}^{-1}(r)\mathbf{D} &= \delta(r)\mathbf{I} - \mathbf{C}(r)\mathbf{D} = \boldsymbol{\omega}^{-1}(r) - \mathbf{c}(r)\mathbf{D}. \end{aligned} \quad (4.32)$$

This decomposition of correlations allows us to conveniently separate bonding and intermolecular effects on the liquid structure.

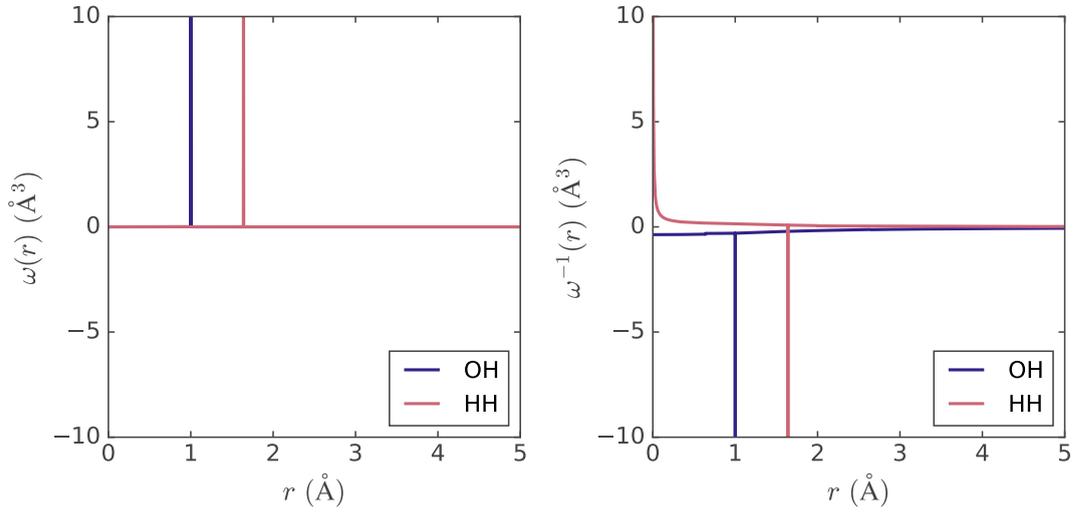


FIGURE 4.3: Water intramolecular site-site correlation function (left figure) and its functional inverse (right figure).

We illustrate the properties of $\omega(r)$ for the example of water. The matrix of intermolecular correlations in this case is given by:

$$\omega(r) = \begin{bmatrix} \delta(r) & \frac{\delta(r-L_{OH})}{4\pi L_{OH}^2} & \frac{\delta(r-L_{OH})}{4\pi L_{OH}^2} \\ \frac{\delta(r-L_{OH})}{4\pi L_{OH}^2} & \delta(r) & \frac{\delta(r-L_{HH})}{4\pi L_{HH}^2} \\ \frac{\delta(r-L_{OH})}{4\pi L_{OH}^2} & \frac{\delta(r-L_{HH})}{4\pi L_{HH}^2} & \delta(r) \end{bmatrix}. \quad (4.33)$$

The Fourier transform can be taken analytically and produces:

$$\hat{\omega}(k) = \begin{bmatrix} 1 & \frac{\sin(kL_{OH})}{kL_{OH}} & \frac{\sin(kL_{OH})}{kL_{OH}} \\ \frac{\sin(kL_{OH})}{kL_{OH}} & 1 & \frac{\sin(kL_{HH})}{kL_{HH}} \\ \frac{\sin(kL_{OH})}{kL_{OH}} & \frac{\sin(kL_{HH})}{kL_{HH}} & 1 \end{bmatrix}. \quad (4.34)$$

In principle, the matrix inverses of the above functions can be written analytically as well, but they do not have a simple form^{96–98}. Additionally, $\omega^{-1}(k)$ is undefined at $k \rightarrow 0$.

We can, however, still perform both matrix inversion and inverse Fourier transforms numerically and define $\omega^{-1}(k=0) = \lim_{k \rightarrow 0} \omega^{-1}(k)$. Figures 4.3 and 4.4 show the behaviour of all intramolecular correlation functions in both real and momentum spaces. As you can see, $\omega^{-1}(r)$ is similar to $-\omega(r)$, although it contains extra components that can be

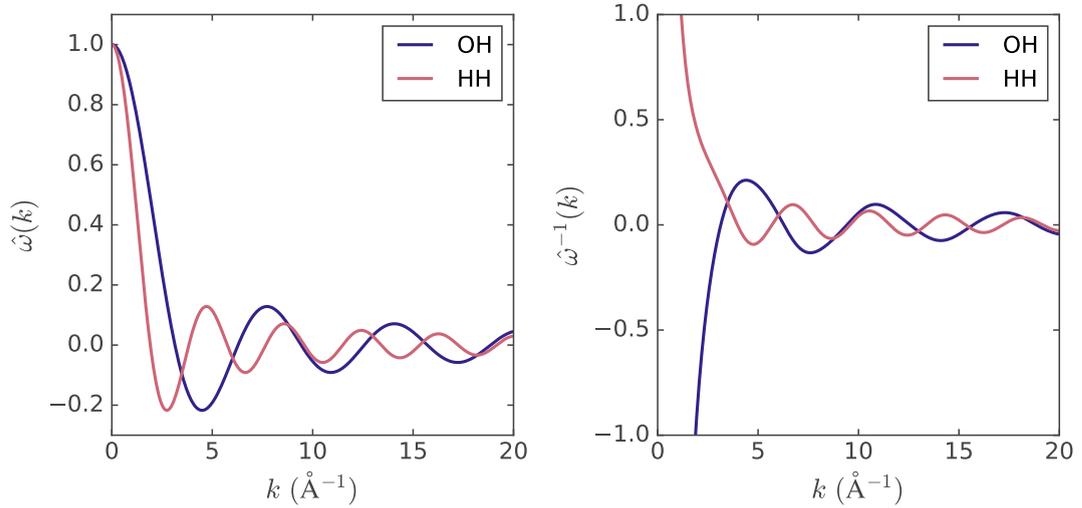


FIGURE 4.4: Fourier transforms of water intramolecular site-site correlation function (left figure) and its inverse (right figure).

defined by a sum of convolutions of delta functions⁹⁷. Its sharp spikes at $r = L_{OH}$ and $r = L_{HH}$ can be rationalised as a strongly favourable interaction which enforces a certain distance between bonded atoms:

$$\beta\Delta\mu_i \approx \Delta\rho_j(r)\omega_{ij}^{-1}(r). \quad (4.35)$$

In Fourier space intramolecular correlation functions oscillate with a constant frequency determined by bond distances in the molecule.

4.4 1D-RISM

After long preparations we are now ready to write the site-site Ornstein-Zernike equation for homogeneous liquid in a more common form

$$\left[\hat{\omega}(k) + \hat{\mathbf{h}}(k)\mathbf{D} \right] \left[\hat{\omega}^{-1}(k) - \hat{\mathbf{c}}(k)\mathbf{D} \right] = \mathbf{I}. \quad (4.36)$$

Opening brackets and rearranging terms results in $\hat{\mathbf{h}}(k)\mathbf{D}\hat{\omega}^{-1}(k) = \hat{\omega}(k)\hat{\mathbf{c}}(k)\mathbf{D} + \hat{\mathbf{h}}(k)\mathbf{D}\hat{\mathbf{c}}(k)\mathbf{D}$. Multiplying both by $\hat{\omega}(k)\mathbf{D}^{-1}$ from the right side gives us the usual form of the site-site Ornstein-Zernike equation²⁶

$$\hat{\mathbf{h}}(k) = \hat{\omega}(k)\hat{\mathbf{c}}(k)\hat{\omega}(k) + \hat{\mathbf{h}}(k)\mathbf{D}\hat{\mathbf{c}}(k)\hat{\omega}(k), \quad (4.37)$$

where we utilised $\mathbf{D}\hat{\omega}(k)\mathbf{D}^{-1} = \hat{\omega}(k)$. This equation can be re-written in real space to give

$$h_{ij}(|\mathbf{r}_1 - \mathbf{r}_2|) = \sum_{k=1}^N \sum_{l=1}^N \iint [\omega_{ik}(|\mathbf{r}_1 - \mathbf{r}'|)c_{kl}(|\mathbf{r}' - \mathbf{r}''|)\omega_{lj}(|\mathbf{r}'' - \mathbf{r}_2|) + h_{lj}(|\mathbf{r}'' - \mathbf{r}_2|)\rho_l c_{kl}(|\mathbf{r}' - \mathbf{r}''|)\omega_{ik}(|\mathbf{r}_1 - \mathbf{r}'|)] d\mathbf{r}' d\mathbf{r}'' \quad (4.38)$$

Although quite cumbersome, this is a direct extension of the Ornstein-Zernike equation for mixtures of simple liquids (3.43), with intramolecular correlation function $\omega(r)$ accounting for additional propagation of interactions through the intramolecular correlations.

To obtain a theory for bulk polyatomic liquids, one has to combine equation 4.38 with another suitable expression. One possibility is to use an HNC-like closure (eq. 4.15)

$$h_{ij}(r) + 1 = \exp[-\beta u_{ij}(r) + h_{ij}(r) - c_{ij}(r)], \quad (4.39)$$

where $u_{ij}(r)$ is a site-site interaction potential energy at separation r . One can also combine site-site Ornstein-Zernike equation with the Percus-Yevick (PY) closure:

$$h_{ij}(r) + 1 = \exp[-\beta u_{ij}(r)] [h_{ij}(r) - c_{ij}(r) + 1], \quad (4.40)$$

which is known to be reasonably accurate for hard sphere systems²⁶.

As we saw in the first section of the chapter, for simple liquids, HNC (as well as the PY closure^{26,99}) can be rationalised from the viewpoint of a density expansion of free energy. However, the use of these closures for liquids with site-site interactions is harder to justify. Indeed, the diagrammatic analysis shows that these closures lead to a number of unphysical interactions^{26,93,100}. Nevertheless, experience has shown that both HNC and PY approximations tend to produce relatively reasonable results even for molecular liquids. Usually, PY describes more accurately hard sphere systems, while HNC tends to be better for liquids interacting via Lennard-Jones and Coulomb potentials²⁶.

For strongly interacting systems, the convergence of HNC closure might become problematic due to the exponent on the right side of 4.39 becoming increasingly significant. This problem can be addressed by approximating the greater than one part of the exponential

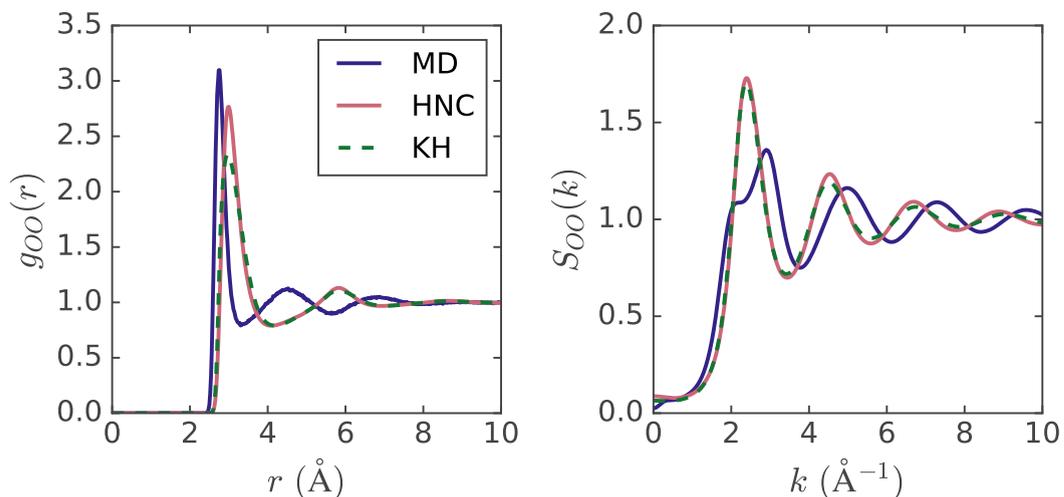


FIGURE 4.5: Water oxygen-oxygen radial distribution functions (left figure) and partial structure factors (right figure) from RISM and molecular dynamics at standard conditions.

function via a Taylor expansion. Defining $t_{ij}^*(r) = -\beta u_{ij}(r) + h_{ij}(r) - c_{ij}(r)$ we can write

$$g_{ij}(r) = \begin{cases} \exp [t_{ij}^*(r)] & \text{if } g_{ij}(r) \leq 1 \\ \sum_{k=0}^n \frac{1}{k!} t_{ij}^*(r)^k & \text{if } g_{ij}(r) > 1. \end{cases} \quad (4.41)$$

This closure is called a partial series expansion of order n (PSE- n)¹⁰¹. The case $n = 1$ is often referred to as Kovalenko-Hirata (KH) closure.

The approach, as we described it, can be successfully applied to liquids interacting only via short-ranged potentials such as oxygen, nitrogen, bromine,¹⁰² or CS_2 ¹⁰³. However, it makes incorrect predictions of an important characteristic of polar molecules: the dielectric constant. To address this issue, a dielectrically consistent reference interaction site model (DRISM) has been proposed, in which the dielectric constant becomes a fixed input parameter¹⁰⁴. The correction somewhat redefines the direct correlation function, but does not change equations 4.39 or 4.41. Thus, to decrease the number of acronyms, we will be referring to both standard as well as dielectrically consistent theory as 1D-RISM, but it will be assumed that polar liquids are treated via the DRISM approach.

Since water occupies a central theme in this thesis, it is useful to examine some failures of RISM theory when it is applied to it. Figure 4.5 demonstrates the site-site radial distribution functions of SPC/E water (cSPC/E in case of RISM¹⁰⁵), obtained at 298 K

using molecular dynamics (MD) and 1D-RISM with HNC and KH closures. While the oxygen radial distribution functions look quite similar, the RISM water is radically different from MD one. A comparison of oxygen partial structure factors $S_{OO}(k) = 1 + \rho \hat{h}_{OO}(k)$ shows that neither of the closures can predict the characteristic doublet structure of the first peak, which has been linked to the tetrahedral structure of water and voids existing in the actual liquid. Similarly, comparison of coordination numbers (table B.1) as well oxygen-hydrogen radial distribution functions (figure B.1) points towards a lack of hydrogen bonding and a much simpler structure of 1D-RISM water when compared to MD results.

Despite all of the mentioned shortcomings, 1D-RISM is still one of the most useful methods for quickly predicting susceptibilities χ of molecular liquids. This might come as a bit surprising, considering all the problems with 1D-RISM we have listed. However, while 1D-RISM fails to describe short order structure of a liquid ($k > 1 \text{ \AA}^{-1}$), it predicts longer wavelength responses relatively accurately. At the same time, obtaining accurate (and smooth) descriptions of these regions of correlation functions with MD proves to be quite problematic due to the very slow convergence. Also, most of the information regarding the electrostatic response of water is also connected to the small wavenumber part of the susceptibility. This makes the DRISM model (with its ability to take experimental dielectric constant as an input parameter) as an arguably better-suited approach for predicting the low-frequency dielectric response of liquid than conventional MD.

4.5 3D-RISM as a density functional theory

It is not difficult to extend the 1D-RISM model to the situations in which one of the components is present at infinite dilution⁶. This way one can apply the theory to model single molecule solvation. However, within this approach, all correlation functions are spherically symmetric, which makes applications to large, non-spherical solutes somewhat problematic.

The three-dimensional reference interaction site model (3D-RISM) provides a clearer picture of the solvation. The main idea is to treat the solute surrounded by the bulk solvent as an inhomogeneous system, in which external potential is produced by the solute.

The distribution of the solvent sites around the solute is determined by minimizing the total free energy of the system.

The derivation of 3D-RISM proceeds similarly to the derivation of HNC approximation, except that the Grand potential is substituted with its site-site version (equation 4.17). We start by writing down a second order expansion of excess intrinsic free energy functional, defined via equation 4.18:

$$\begin{aligned} \mathcal{F}^{ex}[\boldsymbol{\rho}] &= \mathcal{F}^{ex}[\boldsymbol{\rho}_0] + \sum_{i=1}^N \int \left. \frac{\delta \mathcal{F}}{\delta \rho_i(\mathbf{r})} \right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0} \Delta \rho_i(\mathbf{r}) d\mathbf{r} \\ &+ \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \iint \left. \frac{\delta^2 \mathcal{F}}{\delta \rho_i(\mathbf{r}) \delta \rho_j(\mathbf{r}')} \right|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0} \Delta \rho_i(\mathbf{r}) \Delta \rho_j(\mathbf{r}') d\mathbf{r} d\mathbf{r}' + F^B[\boldsymbol{\rho}], \end{aligned} \quad (4.42)$$

where ρ_i is the number density of site i , $\boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_N]$, subscript 0 indicates reference density, relative to which expansion is being taken, and $F^B[\boldsymbol{\rho}]$ collects neglected terms in the expansion.

Assuming a homogeneous reference state and using definitions of the whole site-site direct correlation function from equation 4.24, we get

$$\begin{aligned} \mathcal{F}^{ex}[\boldsymbol{\rho}] &= \mathcal{F}^{ex}[\boldsymbol{\rho}_0] + \sum_{i=1}^N \bar{\mu}_i^{ex} \int \Delta \rho_i(\mathbf{r}) d\mathbf{r} \\ &- \frac{kT}{2} \sum_{i=1}^N \sum_{j=1}^N \iint C_{ij}(|\mathbf{r}' - \mathbf{r}|) \Delta \rho_i(\mathbf{r}) \Delta \rho_j(\mathbf{r}') d\mathbf{r} d\mathbf{r}' + F^B[\boldsymbol{\rho}]. \end{aligned} \quad (4.43)$$

We write down the full intrinsic free energy functional using equations 4.18, 4.19, and cancel terms containing the thermal de Broglie wavelength

$$\begin{aligned} \mathcal{F}[\boldsymbol{\rho}] &= \sum_{i=1}^N kT \int \rho_i(\mathbf{r}) \ln \frac{\rho_i(\mathbf{r})}{\rho_{i,0}} d\mathbf{r} + \Omega[\boldsymbol{\rho}_0] + \sum_{i=1}^N \mu_{i,0} \int \rho(\mathbf{r}) d\mathbf{r} \\ &- \frac{kT}{2} \sum_{i=1}^N \sum_{j=1}^N \iint \Delta \rho_i(\mathbf{r}) C_{ij}(|\mathbf{r}' - \mathbf{r}|) \Delta \rho_j(\mathbf{r}') d\mathbf{r} d\mathbf{r}' + F^B[\boldsymbol{\rho}]. \end{aligned} \quad (4.44)$$

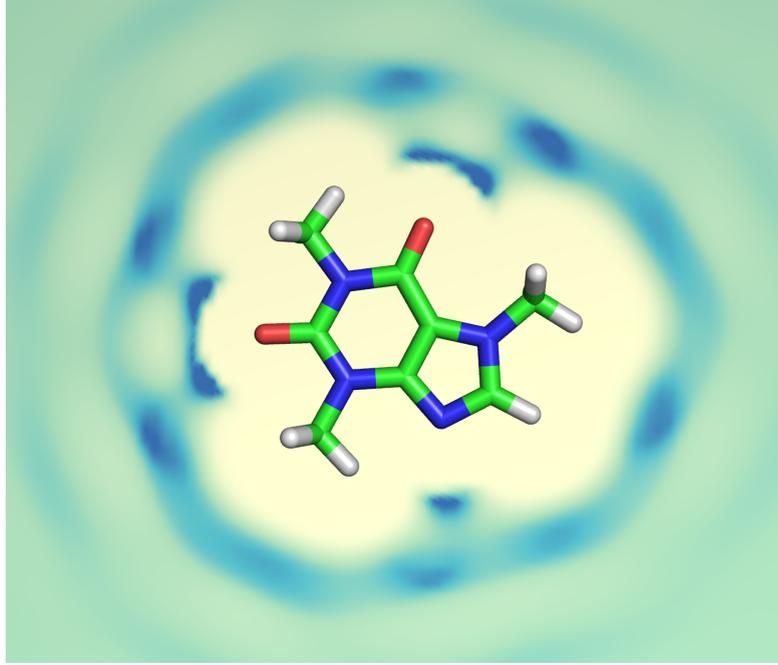


FIGURE 4.6: Distribution of hydrogens around caffeine, with regions of lower density coloured in yellowish colour and regions of higher in blue.

Defining the isotropic grand potential as $\Omega[\rho_0] = \sum_{i=1}^N [kT \int \rho_{i,0} \ln \rho_{i,0} \Lambda_i^3 d\mathbf{r} - \mu_{i,0} \int \rho_{i,0} d\mathbf{r}] + \mathcal{F}_0^{ex}[\rho_0]$ we can obtain an expression for the change of the grand potential

$$\begin{aligned} \Omega[\rho] = \Omega[\rho_0] &+ \sum_{i=1}^N kT \int \left[\rho_i(\mathbf{r}) \ln \frac{\rho_i(\mathbf{r})}{\rho_{i,0}} - \Delta\rho_i(\mathbf{r}) \right] d\mathbf{r} + \sum_{i=1}^N \int \rho_i(\mathbf{r}) \phi_i(\mathbf{r}) d\mathbf{r} \\ &- \frac{kT}{2} \sum_{i=1}^N \sum_{j=1}^N \iint \Delta\rho_i(\mathbf{r}) C_{ij}(|\mathbf{r}' - \mathbf{r}|) \Delta\rho_j(\mathbf{r}') d\mathbf{r} d\mathbf{r}' + F^B[\rho]. \end{aligned} \quad (4.45)$$

As we can see, this expression differs from the density expansion for simple liquids only via presence of intramolecular correlations, summarised by C .

The free energy change from the expression 4.45 is obtained by minimizing the Ω for a given external potential. Setting $F^B[\rho] = \text{const}$ results in a 3D-RISM/HNC approximation, which we will for simplicity call 3D-RISM.

As one would expect from the second order expansion, 3D-RISM is accurate only for small density changes and breaks down for the larger one. Unfortunately, placing a solute into the liquid makes a region of space inaccessible for solvent, making $\Delta\rho(\mathbf{r}) = 0$ inside the solute core. This does count as a significant perturbation and leads to poor solvation thermodynamics predictions.

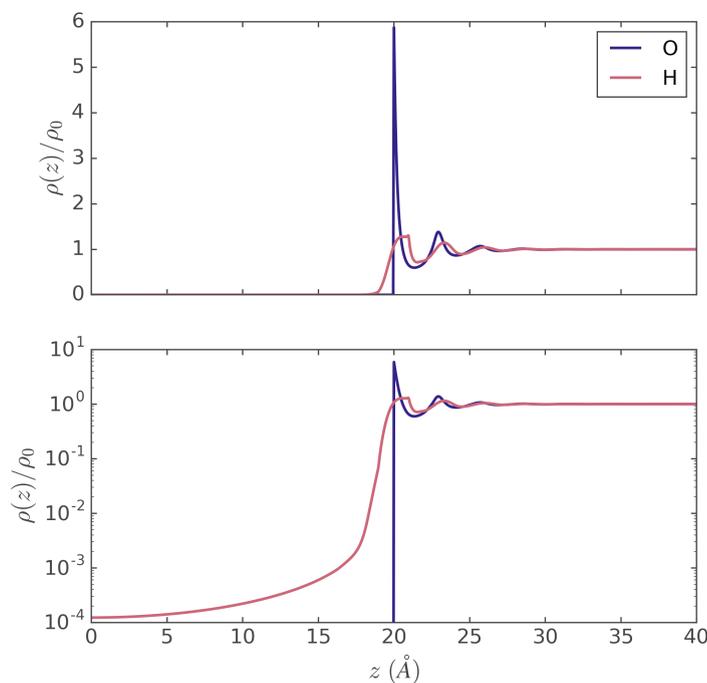


FIGURE 4.7: Distribution of water oxygens (blue) and hydrogens (red) in a large box with $u_O = \infty$ for $z < 20 \text{ \AA}$. You can see that some hydrogens can be found arbitrarily far away from the oxygens.

Interestingly, despite its shortcomings, the 3D-RISM model is still capable of producing relatively reasonable density distributions around small solutes. Figure 4.6 demonstrates a typical result of a 3D-RISM calculation. You can see a distribution of hydrogen sites around the caffeine molecule. It is easy to see hydrogen bonds as well as outlines of the first coordination shell. While it was difficult to compare these results directly to methods such as molecular dynamics, we believe that qualitative pictures are mostly identical. A number of studies confirm that for solutes described via Lennard-Jones and electrostatic potentials 3D-RISM predicts solvent distributions that are in good agreement with both other computational methods as well as experimental observations, even in the case of macromolecules^{106–110}.

One of the effects which 3D-RISM fails to account properly for is solvent bonding. Recall that just as intermolecular correlations, intramolecular correlations are also described only up to second order. Thus, we expect them to be inaccurate to some extent. On the figure 4.7 you can see a distribution of water oxygens and hydrogens near a hard wall, defined only for oxygens via $u_O(z) = \infty$ for $z < 20 \text{ \AA}$ and $u_O(z) = 0$ otherwise. Hydrogens are not restricted by any potential. You can see that the density of hydrogens

is non-zero at arbitrary distances from oxygen atoms, even though built-in intramolecular correlation functions ω requires them to be at exactly 1 Å from oxygens.

In the previous two sections, we described 1D and 3D-RISM approaches. While quite similar, it is important to emphasize a major difference between them. In 1D-RISM one starts with unknown total and direct correlation functions, but the system is assumed to be homogeneous. It is similar to pair HNC discussed at the beginning of the section. Within 3D-RISM, the direct correlation functions between solvent molecules are fixed, and the system density is varied. This approach is closely related to singlet HNC (HNC1). Note that while HNC1 is capable of describing anisotropic systems, it comes at a cost: the singlet Ornstein-Zernike approaches lose a power of density in accuracy compared to pair approximations⁹⁹.

4.6 3D-RISM as an integral equation theory

We can also treat 3D-RISM as an integral equation approach³⁰. A theoretical background of this approach can be found in Ref. 79 and Ref. 111. The main idea is to reduce the 6-dimensional Ornstein-Zernike equation⁹⁰ to get

$$\mathbf{D}\hat{\mathbf{h}}(\mathbf{k}) = \hat{\mathbf{X}}(k)\hat{\mathbf{c}}(\mathbf{k}), \quad (4.46)$$

where $\hat{\mathbf{h}}(\mathbf{k})$ is the vector of solute-solvent total correlation functions ($\Delta\rho_i = h_i\rho_i$), and $\hat{\mathbf{c}}(\mathbf{k}) = [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_N]$ is the vector of solute-solvent direct correlation functions. The above expression is typically called the solute-solvent Ornstein-Zernike equation and assumes that the solute is present at infinitely low concentration. Similarly to 1D-RISM or HNC theories, this equation needs to be supplemented with a proper closure.

Although one often introduces closures to the above equation by modifying analogous closures from simple liquid theories, we can formally derive a HNC-like closure starting from 3D-RISM free energy functional (equation 4.45). Recall the equilibrium condition in grand canonical ensemble $\delta\Omega/\delta\rho[\mathbf{r}] = 0$. Applying it to the 3D-RISM/HNC functional, we get:

$$\frac{\Omega[\rho]}{\rho_i(\mathbf{r})} = 0 = \ln g_i(\mathbf{r}) + \beta\phi_i(\mathbf{r}) - \sum_{j=1}^N \int \Delta\rho_j(\mathbf{r}') C_{ji}(|\mathbf{r}' - \mathbf{r}|) d\mathbf{r}'. \quad (4.47)$$

Rearranging this expression we get an integral equation for density distributions of solvent sites:

$$\rho_i(\mathbf{r}) = \rho_{i,0} \exp \left[-\beta\phi_i(\mathbf{r}) + \sum_{j=1}^N \int \Delta\rho_j(\mathbf{r}') C_{ji}(|\mathbf{r}' - \mathbf{r}|) d\mathbf{r}' \right]. \quad (4.48)$$

We use equation 4.24 to simplify the above and get

$$g_i(\mathbf{r}) = \exp \left\{ -\beta\phi_i(\mathbf{r}) + \sum_{j=1}^N \int \Delta\rho_j(\mathbf{r}') \left[\frac{\delta_{ij}\delta(|\mathbf{r}' - \mathbf{r}|)}{\rho_i} - \chi_{ji}^{-1}(|\mathbf{r}' - \mathbf{r}|) \right] d\mathbf{r}' \right\}. \quad (4.49)$$

Integrating the δ function we obtain

$$g_i(\mathbf{r}) = \exp \left[-\beta\phi_i(\mathbf{r}) + h_i(\mathbf{r}) - \sum_{j=1}^N \int \Delta\rho_j(\mathbf{r}') \chi_{ji}^{-1}(|\mathbf{r}' - \mathbf{r}|) d\mathbf{r}' \right], \quad (4.50)$$

which, with the aid of 4.46, is transformed into

$$g_i(\mathbf{r}) = \exp [-\beta\phi_i(\mathbf{r}) + h_i(\mathbf{r}) - c_i(\mathbf{r})], \quad (4.51)$$

a 3D-RISM equivalent of HNC closure.

For strongly attractive potentials, the 3D-RISM/HNC system of equations can be quite difficult to converge. Similarly to 1D-RISM theory, we can introduce a partial series expansion of order n (PSE- n) of HNC equation¹⁰¹. Defining $t_i^*(\mathbf{r}) = -\beta u_i(\mathbf{r}) + h_i(\mathbf{r}) - c_i(\mathbf{r})$ we write

$$g_i(\mathbf{r}) = \begin{cases} \exp [t_i^*(\mathbf{r})] & \text{if } g_i(\mathbf{r}) \leq 1 \\ \sum_{k=0}^n \frac{1}{k!} t_i^*(\mathbf{r})^k & \text{if } g_i(\mathbf{r}) > 1. \end{cases} \quad (4.52)$$

In practice combining the PSE-3 closure and equation 4.46 is the fastest and the most robust way to minimise the 3D-RISM/HNC functional.

3D-RISM equations, written in terms of the solute-solvent correlation functions, are often referred to as an integral equation theory or a molecular theory of liquids. Using this form, one can also write a somewhat simpler expression for the solvation free energy using thermodynamic integration (equation 2.26). In the case of 3D-RISM/HNC one arrives at³⁰

$$\Delta\Omega_{HNC} = kT \sum_{i=1}^N \rho_i \int \left[\frac{1}{2} h_i^2(\mathbf{r}) - \frac{1}{2} h_i(\mathbf{r}) c_i(\mathbf{r}) - c_i(\mathbf{r}) \right]. \quad (4.53)$$

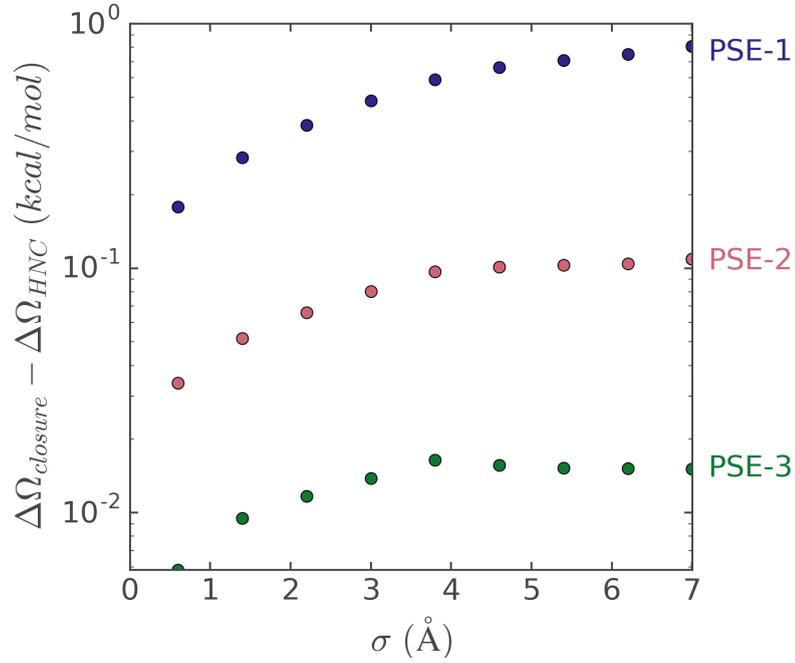


FIGURE 4.8: Solvation free energies of Lennard-Jones solutes with σ given on x axis and $\epsilon = 4$ kcal/mol. On y axis you can see the difference in free energies by PSE-n and HNC closures.

To verify that this expression is consistent with the 3D-RISM/HNC functional one needs to substitute $\ln g_i(\mathbf{r}) = -\beta\phi(\mathbf{r}) + h_i(\mathbf{r}) - c_i(\mathbf{r})$ into equation 4.45. Applying the same approach to PSE-n closure, one obtains¹⁰¹

$$\Delta\Omega_{PSE-n} = \Delta\Omega_{HNC} - kT \sum_{i=1}^N \rho_i \int \left\{ \Theta [h_i(\mathbf{r})] \frac{t_i^*(\mathbf{r})^{n+1}}{(n+1)!} \right\} d\mathbf{r}, \quad (4.54)$$

where Θ is a Heaviside step function:

$$\Theta(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0. \end{cases} \quad (4.55)$$

On figure 4.8 you can see how the differences in free energies computed by PSE-n and HNC closures depend on the solute size. The system is water with Lennard-Jones solutes of different radii and $\epsilon = 4$ kcal/mol. These results should not be generalised, as the precise magnitude of the disparity between closures depends on the type of solute and their size. However, in general, the differences between free energy changes from PSE-n closures and HNC becomes small quite fast, especially for $n \geq 3$.

For this thesis, the majority of calculations with realistic and larger solvents was done with the PSE-3 closure to avoid convergence issues, while calculations on model solutes were mostly done with HNC. However, it will be assumed that the conclusions which will be drawn for PSE-3, also hold for HNC and vice versa.

4.7 Partial molar volume

Partial molar volume is defined as a change in system's volume upon addition of an i -th component at constant pressure⁶:

$$\bar{V}_i = \left(\frac{\partial V}{\partial N_i} \right)_{P,T,N_{j \neq i}} = \left(\frac{\partial \mu_i}{\partial P} \right)_{P,T,N_{j \neq i}}, \quad (4.56)$$

where the second equality was obtained using Maxwell's relation⁴⁵. Usually, the partial molar volume is discussed as a part of solvation thermodynamics, but due to its relation to integral equation theories, we decided to present it here.

Similarly to other thermodynamic quantities, it can be split into ideal and excess contributions. Using equation 2.23 we get:

$$\begin{aligned} \bar{V}_i &= \frac{\partial \mu_i^{ex}}{\partial P} + \frac{\partial \mu_i^{ig}}{\partial P} \\ &= \frac{\partial \mu_i^{ex}}{\partial P} + kT \frac{\ln \rho}{\partial P} \\ &= \Delta V_i + kT \chi_T, \end{aligned} \quad (4.57)$$

where the last equality was obtained using equation 3.60, and ΔV_i stands for the excess part of partial molar volume. Thus, $kT \chi_T$ is the molecule volume arising due to its kinetic energy, while ΔV_i depends purely on intramolecular interactions¹².

The excess volume of a solute ΔV in a single-component solvent can be expressed through a total correlation function^{12,112}:

$$\Delta V = - \int [g(\mathbf{r}) - 1] d\mathbf{r} = - \int h(\mathbf{r}) d\mathbf{r}. \quad (4.58)$$

This relation can be readily interpreted for a solute that has $h(\mathbf{r}) = -1$ for points inside the solute and $h(\mathbf{r}) = 1$ for other regions of space. To obtain a more general interpretation we need to introduce the concept of the Gibbs dividing surface.

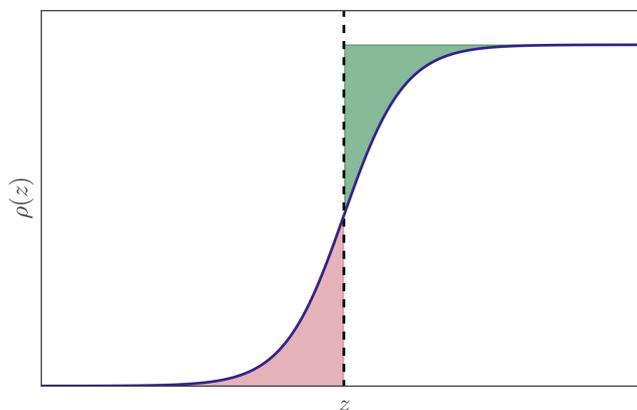


FIGURE 4.9: A schematic depiction of Gibbs dividing surface for the case of planar interface. The surface, shown as a dashed line divides solvent local density (blue line) in such way that excess of solvent to the left of the surface (region shaded in red) is equal to the depletion of solvent to the right of the surface (region shaded in green).

The Gibbs dividing surface is a two-dimensional boundary, dividing an interface in such a way that the excess of solvent in one phase is equal to its depletion from another (see figure 4.9)^{113,114}. For a spherical solute we can express it as

$$4\pi \int_0^{R_G} g(r)r^2 dr + 4\pi \int_{R_G}^{\infty} [g(r) - 1] r^2 dr = 0, \quad (4.59)$$

where R_G is a radius of the Gibbs dividing surface. Adding the above equation to equation 4.58 we find:

$$\begin{aligned} \Delta V &= -4\pi \int_0^{\infty} [g(r) - 1] r^2 dr + 4\pi \int_0^{R_G} g(r)r^2 dr + 4\pi \int_{R_G}^{\infty} [g(r) - 1] r^2 dr \\ &= -4\pi \int_0^{R_G} [g(r) - 1] r^2 dr + 4\pi \int_0^{R_G} g(r)r^2 dr \\ &= 4\pi \int_0^{R_G} r^2 dr = \frac{4}{3}\pi R_G^3. \end{aligned} \quad (4.60)$$

Thus, for a spheric solute, excess volume is equal to the volume enclosed by the Gibbs dividing surface. This result also holds for a solute with an arbitrary shape¹¹³. It gives us a convenient way to define both volume and surface of a solute that are consistent with each other.

For a multi-component solvent, solute partial molar volume becomes dependent on the solvent component molar volumes. For instance, in the case of a two-component solvent we have $\bar{V}_s = -\rho_A \bar{V}_A G_{sA} - \rho_B \bar{V}_B G_{sB}$, where G represents the Kirkwood-Buff integral $\bar{G}_{ij} = \int h_{ij}(\mathbf{r}) d\mathbf{r}$. A simpler expression can be obtained using direct correlation functions

$$\bar{V}_s = kT\chi_T \left[1 - \rho \sum_{i=1}^N \int c_i(\mathbf{r}) d\mathbf{r} \right], \quad (4.61)$$

with excess volume defined by

$$\Delta V_s = -kT\chi_T \rho \sum_{i=1}^N \int c_i(\mathbf{r}) d\mathbf{r}. \quad (4.62)$$

Partial molar volumes in this thesis were computed using the above equations.

4.8 Solvation free energy from 3D-RISM

Chemical potential (or solvation free energy) of a solute can be expressed as a change in the grand potential of a solvent due to the presence of a single molecule⁹⁴:

$$\mu^{ex} = \Delta F = \Omega[\rho] - \Omega[\rho_0], \quad (4.63)$$

where we will reserve the symbol ΔF to specifically denote solvation free energy in Ben-Naim's definition whenever we do not have to worry regarding the ensemble in which the process takes place. As was already discussed in section 2.8, the ability to accurately predict solvation free energies has much practical value. A major practical advantage of 3D-RISM models, when compared to molecular dynamics, comes from the possibility to obtain free energy changes from a single (or a couple of) end-point calculation, without the need to perform a thermodynamic integration. It is not surprising that much effort has been put into predicting solvation free energies by interaction site models. An overview of all proposed models is beyond the scope of the thesis; besides, a number of excellent articles have been published on this subject^{6,115–117}. Here we will only discuss a few corrections related to 3D-RISM.

Molecular dynamics simulations provide a straightforward way for evaluating the accuracy of 3D-RISM free energy functionals. A comparison of the results with experiment

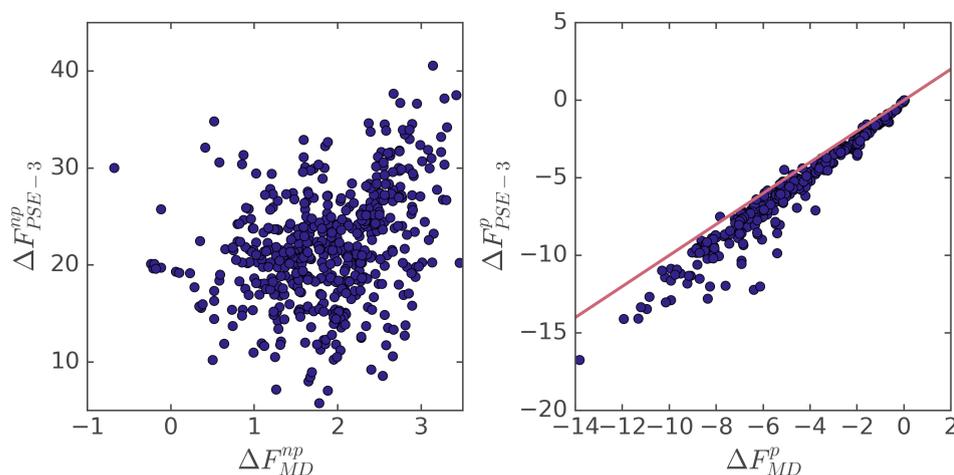


FIGURE 4.10: Comparison of molecular dynamics and PSE-3 Lennard-Jones components of hydration free energies (left figure) and electrostatic components (right figure) for 504 molecules from Mobley dataset, which is discussed in detail in section 6.1. Molecular dynamics results are taken from Ref. 13. All values are in kcal/mol

provides a somewhat less clear picture since, in addition to errors arising from 3D-RISM approximations, one has to take into account the accuracy of the potentials, the validity of the classical approximation, experimental errors, etc. On the other hand, a comparison of 3D-RISM solvation free energies with those from molecular dynamics allows us to directly assess the accuracy of 3D-RISM free energy expression, provided that the same potentials are used.

We have already discussed (section 2.6) that it is possible to split solvation free energy into Lennard-Jones and electrostatic contributions formally. To do so within 3D-RISM theory, we simply compute solvation free energies of a molecule with and without partial charges on atomic sites. Figure 4.10 demonstrates that while electrostatic components of free energy predicted by 3D-RISM are relatively accurate, the Lennard-Jones components practically do not correlate with molecular dynamics.

Recently, some studies have demonstrated that it is possible to significantly improve the accuracy of 3D-RISM solvation free energies using corrections of the form:

$$\Delta F_{UC} = \Delta F_{3D-RISM} + a\bar{V} + b, \quad (4.64)$$

where UC stands for universal correction, a and b are empirical coefficients that depend on the solvent and closure, and $\Delta F_{3D-RISM}$ is the 3D-RISM free energy, most typically

computed with KH closure^{31,35,116}. Another correction, called NgB, was developed specifically for water³⁵

$$\Delta F_{NgB} = \Delta F_{KH} + \frac{kT\rho_O}{2}(1 - \gamma) \int_{V_{in}} c_O^{LJ}(\mathbf{r}) d\mathbf{r}, \quad (4.65)$$

where ρ_O is the density of oxygen sites in water, γ is an empirical coefficient, c_O^{LJ} is the direct correlation function for water oxygens, evaluated without the solute charges, and V_{in} is the volume inside the solute, defined via the solute-solvent potential energy.

Both of the corrections introduced above significantly improve the accuracy of 3D-RISM solvation free energies (B.2). However, these corrections were introduced empirically and did not suggest reasons why they might be effective or how empirical coefficients might depend on the solvent or thermodynamic conditions. In the next chapter, we will introduce correction schemes that do not require prior parametrization and can be applied to a much larger variety of systems.

Part II

Results and Discussion

Chapter 5

Pressure corrections

The chapter is dedicated to the main theoretical results of the thesis. We first show how the HNC functional can be conveniently split into a couple of contributions. Identifying the overestimated component we eliminate it, introducing simple and advanced pressure corrections. The remaining sections are dedicated to discussing pressure corrected models in the context of water solvation. We are using simple model solutes to focus on physical insights instead of the individual peculiarities of realistic molecules.

5.1 Dissecting HNC free energies

For clarity we introduce the following notation ^a

$$\begin{aligned}\langle \mathbf{f} | \mathbf{g} \rangle &= \sum_{i=1}^N \int f_i(\mathbf{r}) g_i(\mathbf{r}) d\mathbf{r}, \\ \langle \mathbf{f} | \mathbf{K} | \mathbf{g} \rangle &= \sum_{i=1}^N \sum_{j=1}^N \iint f_i(\mathbf{r}) K_{ij}(\mathbf{r}, \mathbf{r}') g_j(\mathbf{r}') d\mathbf{r} d\mathbf{r}'\end{aligned}\tag{5.1}$$

where \mathbf{f} and \mathbf{g} are vectors of functions containing N elements, and \mathbf{K} is an N by N matrix of functions, also referred to as kernel.

Using the above notation we can rewrite the 3D-RISM/HNC functional (equation 4.45) as:

$$\beta \Delta \Omega[\boldsymbol{\rho}] = \langle \boldsymbol{\rho} | \ln \mathbf{g} \rangle - \langle \Delta \boldsymbol{\rho} | \mathbf{1} \rangle + \langle \boldsymbol{\rho} | \beta \mathbf{u} \rangle - \frac{1}{2} \langle \Delta \boldsymbol{\rho} | \mathbf{C} | \Delta \boldsymbol{\rho} \rangle\tag{5.2}$$

^aA similar notation can be found, for example, in Ref. 118.

where $\mathbf{1}$ represents a vector of functions equal to 1 everywhere and $\ln \mathbf{g} = [\ln g_1(\mathbf{r}) \cdots \ln g_N(\mathbf{r})]$. To simplify this expression we also rewrite the condition for equilibrium 4.47: $\ln \mathbf{g} + \beta \phi = \langle \mathbf{C} | \Delta \rho \rangle$. Plugging this into the equation 5.2 we obtain the change of grand potential at the equilibrium

$$\begin{aligned}
\beta \Delta \Omega[\rho] &= \langle \rho | \ln \mathbf{g} + \beta \mathbf{u} \rangle - \langle \Delta \rho | \mathbf{1} \rangle - \frac{1}{2} \langle \Delta \rho | \mathbf{C} | \Delta \rho \rangle \\
&= \langle \rho | \langle \mathbf{C} | \Delta \rho \rangle - \langle \Delta \rho | \mathbf{1} \rangle - \frac{1}{2} \langle \Delta \rho | \mathbf{C} | \Delta \rho \rangle \\
&= \frac{1}{2} \langle \rho | \mathbf{C} | \Delta \rho \rangle + \frac{1}{2} \langle \rho_0 | \mathbf{C} | \Delta \rho \rangle - \langle \Delta \rho | \mathbf{1} \rangle \\
&= \frac{1}{2} \langle \rho | \mathbf{C} | \rho \rangle - \langle \rho | \mathbf{1} \rangle - \frac{1}{2} \langle \rho_0 | \mathbf{C} | \rho_0 \rangle + \langle \rho_0 | \mathbf{1} \rangle,
\end{aligned} \tag{5.3}$$

where ρ_0 is the vector of the site densities in the reference system.

To interpret the above results we recall that $\Omega = -PV$. To connect the integrals of the correlation functions to the pressure, we can use the compressibility theorem, introduced in the section 3.6. Cummings and Stell have derived its extension for the interaction site fluids⁹⁶

$$\beta \frac{\partial P}{\partial \rho} = \sum_{i=1}^N \sum_{j=1}^N x_i \left[\delta_{ij} - \rho_j \hat{C}_{ij}(0) \right], \tag{5.4}$$

where $\rho_i = x_i \rho$. After a bit of algebra we can recast this result in the matrix form

$$\beta \left(\frac{\partial P}{\partial \rho} \right)_{V,T} = \rho \mathbf{x}^T \left[\mathbf{D}^{-1} - \hat{\mathbf{C}}(0) \right] \mathbf{x} = 1 - \rho \mathbf{x}^T \hat{\mathbf{C}}(0) \mathbf{x}, \tag{5.5}$$

where $\mathbf{x} = [x_1 \cdots x_N]$ is the vector of the mole fractions of the sites and superscript T denotes a transpose. To obtain the pressure we need to integrate the above expression. In general, it is not possible to do it in a straightforward manner, since \mathbf{C} depends on the density of the system. However, within the HNC approximation we assume direct correlation functions to be constant (the same as in the reference system), so the integration can be performed analytically from $\rho = 0$ to $\rho = \rho_0 = \sum_{i=1}^N \rho_{i0}$ to produce

$$\beta P_0 = \sum_{i=1}^N \rho_{i0} - \frac{1}{2} \rho_0^T \mathbf{C}(k=0) \rho_0, \tag{5.6}$$

where P_0 indicates the pressure (free energy density) of a homogeneous system. The grand potential is then

$$\begin{aligned}\beta\Omega_0 &= -\beta P_0 V = -\langle \boldsymbol{\rho}_0 | \mathbf{1} \rangle + \frac{1}{2} \langle \boldsymbol{\rho}_0 | \mathbf{C} | \boldsymbol{\rho}_0 \rangle \\ &= -V \sum_i \rho_i + \frac{V}{2} \sum_{ij} \rho_i \rho_j \hat{C}_{ij}(0).\end{aligned}\quad (5.7)$$

Note that an identical expression for homogeneous pressure in 3D-RISM was obtained by Sergiievskiy et al.³⁸ as well as by a number of others for singlet HNC in general^{99,119,120}.

To find the pressure in the case of an inhomogeneous system we can use the exact result obtained by Pozhar et al.¹²¹

$$P(\mathbf{r}; \rho) = \rho(\mathbf{r}) \left[\bar{\mu}(\mathbf{r}; \rho) - \int_0^1 d\lambda \bar{\mu}(\mathbf{r}; \lambda \rho) \right], \quad (5.8)$$

where $\bar{\mu}(\mathbf{r}; \rho)$ is the intrinsic chemical potential at \mathbf{r} and the λ parameter controls density $\rho_\lambda(\mathbf{r}) = \lambda \rho(\mathbf{r})$. The result can be readily extended to multicomponent systems

$$P(\mathbf{r}; \boldsymbol{\rho}) = \boldsymbol{\rho}(\mathbf{r}) \cdot \left[\bar{\boldsymbol{\mu}}(\mathbf{r}; \boldsymbol{\rho}) - \int_0^1 d\lambda \bar{\boldsymbol{\mu}}(\mathbf{r}; \lambda \boldsymbol{\rho}) \right], \quad (5.9)$$

in which scalars μ and ρ are substituted by vector analogues. Within the HNC approximation

$$\bar{\mu}_i = kT \ln \Lambda_i^3 \rho_i(\mathbf{r}) + \bar{\mu}_{i0}^{ex}(\mathbf{r}) - kT \langle \mathbf{C}_{i0} | \Delta \boldsymbol{\rho} \rangle_i, \quad (5.10)$$

where subscript 0 indicates that both excess quantities were evaluated at some reference system. Then, using $\langle \mathbf{C}_{i0} | \Delta \boldsymbol{\rho} \rangle = \langle \mathbf{C}_{i0} | \boldsymbol{\rho} \rangle - \langle \mathbf{C}_{i0} | \boldsymbol{\rho}_0 \rangle$ and taking the integral, we get

$$\begin{aligned}P(\mathbf{r}; \boldsymbol{\rho}) &= \sum_{i=1}^N \rho_i(\mathbf{r}) \left[kT \ln \Lambda_i^3 \rho_i(\mathbf{r}) + \bar{\mu}_{i0}^{ex}(\mathbf{r}) - kT \langle \mathbf{C}_{i0} | \Delta \boldsymbol{\rho} \rangle \right. \\ &\quad \left. - kT \ln \Lambda_i^3 \rho_i(\mathbf{r}) + kT - \bar{\mu}_{i0}^{ex}(\mathbf{r}) + \frac{kT}{2} \langle \mathbf{C}_{i0} | \boldsymbol{\rho} \rangle - kT \langle \mathbf{C}_{i0} | \boldsymbol{\rho}_0 \rangle \right] \\ &= kT \sum_{i=1}^N \rho_i(\mathbf{r}) \left[1 - \frac{1}{2} \langle \mathbf{C}_{i0} | \boldsymbol{\rho} \rangle \right].\end{aligned}\quad (5.11)$$

It follows that for an inhomogeneous system:

$$\beta\Omega[\boldsymbol{\rho}] = -\beta \int P(\mathbf{r}; \boldsymbol{\rho}) d\mathbf{r} = -\langle \boldsymbol{\rho} | \mathbf{1} \rangle + \frac{1}{2} \langle \boldsymbol{\rho} | \mathbf{C} | \boldsymbol{\rho} \rangle, \quad (5.12)$$

where we dropped the subscript from \mathbf{C} for consistency with previous results.

The findings of the past couple of paragraphs highlight that the HNC free energy corresponds to nothing else but $\Delta\Omega_{HNC} = -\Delta PV$, which perhaps is not very surprising. However, these results at least point out that the theory is internally consistent. Moreover, they readily highlight the problems with the approximation; indeed, using equation 5.6 one readily finds that liquid water at room temperature and normal density has a pressure of about 9500 bar, almost 9500 times larger than normal ^b. It is apparent that a single set of direct correlation functions evaluated for a bulk system cannot be used to describe regions with low (or high) liquid density (compared to the reference system), which is precisely what HNC does.

Note that the final result in equation 5.3 can be also expressed as

$$\beta\Delta\Omega[\rho] = \Delta N + \frac{1}{2} \sum_{ij} \iint C_{ij}(|\mathbf{r}_1 - \mathbf{r}_2|) [\rho_i(\mathbf{r})\rho_j(\mathbf{r}) - \rho_{i0}\rho_{j0}] d\mathbf{r}_1 d\mathbf{r}_2, \quad (5.13)$$

where $\Delta N = \sum_{i=1}^N \Delta\rho_i(\mathbf{r}) d\mathbf{r}$. We can obtain the same result from equation 24 in reference 94, by setting $F^B = \text{const}$ and using equation 4.32 to relate the whole and site-site direct correlation functions. We can see that 3D-RISM/HNC is essentially identical to the site density functional theory of Jianzhong Wu and coworkers, provided that one sets the bridge function to zero ^{24,94,122,123}.

Instead of expressing solvation free energy purely using the bulk solvent direct correlation functions, we can also split it into somewhat more familiar terms. Utilizing the HNC equilibrium condition 4.47 and the third equality in equation 5.3, we find

$$\beta\Delta\Omega[\rho] = \frac{1}{2} \langle \rho | \ln \mathbf{g} + \beta \mathbf{u} \rangle + \frac{1}{2} \langle \rho_0 | \mathbf{C} | \Delta \rho \rangle - \langle \Delta \rho | \mathbf{1} \rangle. \quad (5.14)$$

To simplify the expression further we note that

$$\iint f(x - x')g(x') dx' dx = \int g(x') \left[\int f(x - x') dx \right] dx' = \int g(x') dx' \int f(y) dy, \quad (5.15)$$

which follows from the Fubini–Tonelli theorem ¹²⁴. Using this result we can rewrite the second term as

$$\sum_{ij} \rho_{0i} \iint C_{ij}(|\mathbf{r} - \mathbf{r}'|) \Delta\rho_j(\mathbf{r}') d\mathbf{r}' d\mathbf{r} = \sum_{ij} \rho_i \rho_j \hat{C}_{ij}(0) G_j, \quad (5.16)$$

^bThe results are evaluated using experimentally measured water radial distribution functions, reported by Soper et al.

where we used the definition of Kirkwood-Buff integral $G_j = \int h_j(\mathbf{r}) d\mathbf{r}$ ¹². For multicomponent solvent we obtain

$$\Delta\Omega = \frac{kT}{2} \langle \rho | \ln \mathbf{g} + \beta \mathbf{u} \rangle - kT \sum_{i=1}^N \rho_i G_i \left(1 - \frac{1}{2} \sum_{j=1}^N \rho_j \hat{C}_{ij}(0) \right). \quad (5.17)$$

If solvent is a single component liquid with density ρ_0 and N sites, $G_j = G_i = -\Delta V$ and the expression can be simplified further

$$\begin{aligned} \Delta\Omega &= \frac{kT}{2} \langle \rho | \ln \mathbf{g} + \beta \mathbf{u} \rangle + N\rho_0 kT \Delta V - \frac{kT\rho_0^2}{2} \Delta V \sum_{ij} \hat{C}_{ij}(0) \\ &= \frac{kT}{2} \langle \rho | \ln \mathbf{g} + \beta \mathbf{u} \rangle + P_0 \Delta V. \end{aligned} \quad (5.18)$$

In the above equation one can readily identify entropic, enthalpic, and pressure terms, contributing to the total solvation free energy. However, it is important to note that since we are dealing with the grand potential, $P_0 \Delta V$ does not represent the familiar expansion work for the NPT system. The origin of this term is effectively entropic in nature.

Two above equations (5.17 and 5.18) are one of the main results of the thesis and can be readily used to both understand the failures of 3D-RISM theory and to formulate reasonable approximations.

5.2 Hydrophobic solvation

The hydrophobic effect is traditionally associated with (a) unusually high solvation free energies of apolar molecules in water, usually several kilocalories compared to organic solvents and (b) the tendency of apolar compounds in water to aggregate to minimize their surface area^{125,126}. Additionally, several other properties became associated with it such as negative solvation entropy, large system heat capacity increases upon solvation, and entropy convergence at higher temperature^{126,127}.

Here we focus on idealised situation in which our solutes are hard spheres with interaction given by

$$u_{hard-O}(r) = \begin{cases} \infty & \text{if } r < \sigma_{hard-O} \\ 0 & \text{otherwise.} \end{cases} \quad (5.19)$$

Then solvation is determined entirely by entropical and pressure effects. This makes hydrophobic solvation an ideal example using which we can better understand the problems of the 3D-RISM model.

Interestingly, a lot of insight into the failures of 3D-RISM model can be obtained by examining a significantly simpler model of hydrophobic solvation, called information theory (IT) ^{128,129}. Let us define $P_V(N)$, a probability that N water molecules can be found in volume v . We assume that water fluctuations can be described by Gaussian distribution

$$P_v(N) \approx \frac{1}{\sqrt{2\pi\sigma_v}} \exp \left[-\frac{(N - \langle N \rangle_v)^2}{2\sigma_v} \right], \quad (5.20)$$

with $\langle N \rangle_v = \rho v$ being an average number of particles, and $\sigma_v = \langle (\delta N)^2 \rangle_v$ a mean square fluctuation in volume v .

The probability of hard sphere solvation is equivalent to water molecules fluctuating and creating a large enough cavity for the sphere to fit. Then, using $\Delta F = -kT \ln P_v(0)$ we obtain

$$\Delta F_{IT} = kT \left[\frac{\rho^2 v^2}{2\sigma_v} + \frac{1}{2} \ln(2\pi\sigma_v) \right]. \quad (5.21)$$

Associating the position of water molecules only with the locations of oxygens, we obtain σ_v from the equation 3.59:

$$\sigma_v = \iint_v h_{OO}(|\mathbf{r}_1 - \mathbf{r}_2|) d\mathbf{r}_1 d\mathbf{r}_2 + \rho v. \quad (5.22)$$

From the same equation it follows that at the limit of $v \rightarrow \infty$, mean square fluctuations are linked to isothermal compressibility $\lim_{v \rightarrow \infty} \sigma_v = kT \rho^2 v \chi_T$ giving

$$\Delta F_{IT} = \frac{v}{2\chi_T} + \frac{kT}{2} \ln(2\pi kT \rho^2 v \chi_T). \quad (5.23)$$

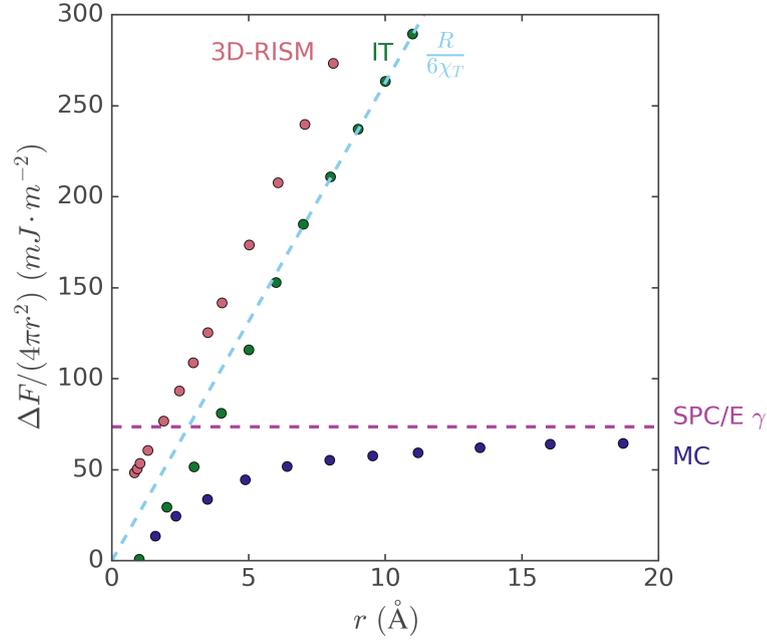


FIGURE 5.1: Dependence of surface energy of hard spheres in water depending on their radius. Monte Carlo (MC) results are taken from Ref. 131.

The corresponding expression for hard sphere hydration free energy in 3D-RISM is quite similar:

$$\begin{aligned}
 \Delta F_{HNC} &= \frac{kT}{2} \langle \rho | \ln \mathbf{g} \rangle - kT \langle \Delta \rho | \mathbf{1} \rangle + \frac{kT}{2} \langle \rho_0 | \mathbf{C} | \Delta \rho \rangle \\
 &= \frac{kT}{2} \langle \rho | \ln \mathbf{g} \rangle - \frac{kT}{2} \langle \Delta \rho | \mathbf{1} \rangle - \frac{kT}{2} \langle \rho_0 | \mathbf{X}^{-1} | \Delta \rho \rangle \\
 &= \frac{kT}{2} \langle \rho | \ln \mathbf{g} \rangle + \Delta V \left(\frac{kT}{2} \rho_0 + \frac{1}{2\chi_T} \right), \\
 &= \frac{kT}{2} \langle \rho | \ln \mathbf{g} \rangle + P_0 \Delta V,
 \end{aligned} \tag{5.24}$$

where to obtain the first equality we used equation 4.26, and the second followed from the result by Imai et al.¹³⁰ $\rho_0^T \hat{\mathbf{X}}^{-1}(0) \rho_0 = \frac{1}{kT\chi_T}$ and equation 5.16. You can see that while additional terms are different in two models, essentially both models suggest that hydrophobic solvation free energy scales proportionally to the volume of the solute. Additionally, in 3D-RISM the proportionality constant is simply the pressure of bulk liquid (equation 5.18), which using the above results can be conveniently expressed for a single component liquids as $P_0 = \frac{1}{2} (kT\rho_0 + 1/\chi_T)$.

The figure 5.1 shows predictions of the change of solvation free energy per unit area depending on the hard sphere radius, made by three different models. In principle, as

hard solute radius r gets larger and larger $\Delta\Omega/(4\pi r^2) \rightarrow \frac{Pr}{3} + \gamma$, where γ is the surface tension between hard solute and solvent. Thus, these type of figure allows us to evaluate both pressure and surface tension within the model.

The results of information theory agree well with Monte-Carlo simulations for small solutes ($r < 4 \text{ \AA}$), but become progressively worse for larger solutes. The 3D-RISM/HNC approach consistently predict surface energy values larger than the two models, but shows trends which are quite similar to information theory.

Monte-Carlo predictions provide a good insight on hydrophobic phenomena. Before diameter of a hard sphere reaches 1 nm, its solvation free energy scales with the volume of the sphere, while afterwards, with its surface area. The reason for this behaviour has been rationalised by Chandler, Weeks and co-workers in a number of important papers on the hydrophobic effect^{131–136}.

The hydration free energy of small molecules largely depends on the strength of hydrogen bonding. Bulk water forms a strong tetrahedral network, which, despite being quite dynamic, rarely breaks down substantially to form solute cavities. Whenever, cavities do occur, water tries to maintain its bonding network if possible; thus the structure of solvent around small cavities is quite similar to that of bulk water. This is the reason why approaches such as information theory are able to accurately describe solvation free energy for small molecules using mean square fluctuations obtained without the presence of a solute.

Near larger solutes, which resemble planar interfaces, the water surface layer undergoes substantial reorganization. The bulk-like hydrogen bonding network is substituted with an interfacial structure similar to water-air interface. The molecules are oriented with O-H bonds towards the solute and the density of water right next to the solute is lower than that found in the bulk. The decrease of density occurs due to the force imbalance: the interfacial water molecules do not experience a lot of attraction towards hydrophobic solute, but are strongly drawn in by the bulk water. These rearrangements help decrease the free energy of hydrophobic solvation, making the creation of larger cavities much more probable than what one would expect from simple Gaussian behaviour.

The dewetting transition is missed by both IT and 3D-RISM. The hydrophobic effect in these models occurs due to the linear increase of water chemical potential as a consequence

of solute excluded volume, and they cannot capture its more subtle details.

5.3 Pressure corrections for solvation free energy

As we saw in the previous section, 3D-RISM largely overestimates bulk solvent pressure (predicting water at normal conditions to have pressure of about 9500 bar). The high pressure arises because of the truncation of the free energy expansion at the second term. The dominant forces in water at equilibrium are mostly repulsive, with attractive forces being generally canceled out. We expect a low order expansion to capture general trends, which are repulsive, and to neglect more subtle attraction interactions, which are described by triplet and higher order correlation functions. This neglect becomes especially problematic when describing interfaces, which are dominated by collective, long distance interactions.

Notably, the large compressibility pressure of 3D-RISM (evaluated using 5.6) is quite consistent with the behaviour of radial distribution functions. The contact theorem^{26,138} tells us that in the limit of an infinitely large hard sphere, bulk pressure is related to the value of solvent density right next to the hard sphere

$$P_0 = \rho kT g(R), \quad (5.25)$$

where $g(R)$ is the value of oxygen radial distribution function at the surface of the hard sphere. Thus, at standard conditions, the 3D-RISM contact value of the oxygen radial distribution function with a hard sphere should approach $P_0/(\rho kT) = 7.2$. From figure 5.2 we can see the contact value approaches 6, which is close to what is predicted using compressibility route pressure (although, it is known that singlet HNC only satisfies the contact value theorem at low densities¹²⁰). Additionally, the shape of the 1D-RISM partial oxygen structure factor, shown in figure 4.5 resembles experimentally observed structure factors for bulk water at $P = 4000$ bar¹³⁹.

In view of the above, it is reasonable to try to correct 3D-RISM by subtracting the overestimated pressure work

$$\Delta F_{PC} = \Delta F_{3D-RISM} - P_0 \Delta V, \quad (5.26)$$

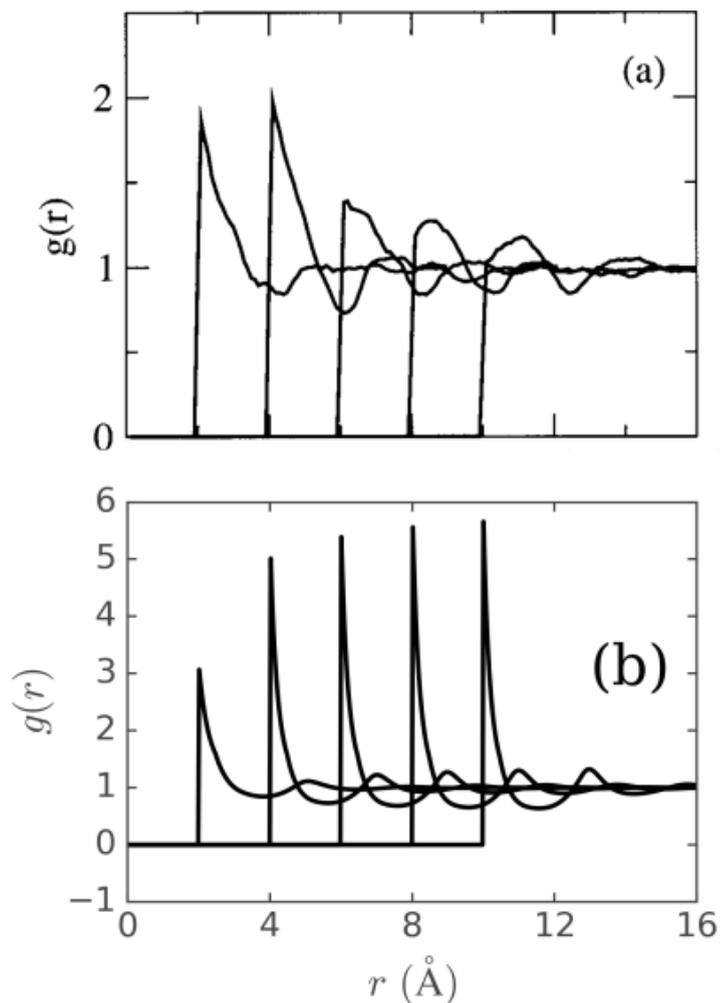


FIGURE 5.2: Both figures demonstrates hard sphere–water oxygen radial distribution functions for spheres of various radii. Figure (a) is taken from Ref. 137 and was obtained using Monte Carlo simulations. Radial distribution functions in figure (b) were calculated using 3D-RISM/HNC.

where P_0 is the 3D-RISM bulk pressure, ΔV is the excess volume of solvent, and PC stands for pressure correction.

We also introduce another way of correcting 3D-RISM results:

$$\Delta F_{PC+} = \Delta F_{3D-RISM} - P_0 \Delta V + P_{id} \Delta V, \quad (5.27)$$

where $P_{id} = \rho_{id} kT$ is an ideal gas pressure with $\rho_{id} = \sum_{\alpha=1}^M \rho_{\alpha}$ being the number density of solvent molecules, not sites. We call this equation the advanced pressure

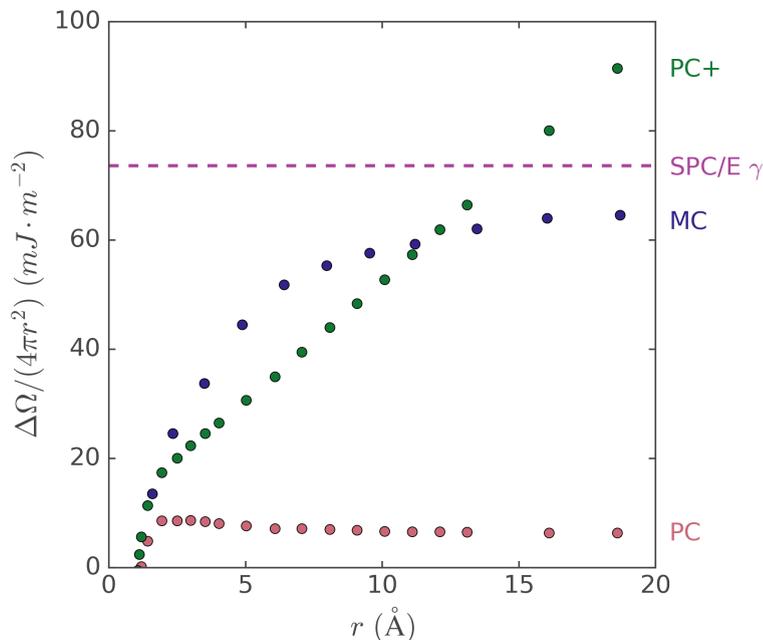


FIGURE 5.3: This figure mirrors figure 5.1, except now instead of information theory and HNC we compare results by PC and PC+ models.

correction (PC+), and this is one of the main results of the thesis ^c. For single component solutions both corrections can be conveniently defined as $\Delta F_{PC} = \frac{kT}{2} \langle \rho | \ln \mathbf{g} \rangle + \frac{1}{2} \langle \rho | \mathbf{u} \rangle$ and $\Delta F_{PC+} = \frac{kT}{2} \langle \rho | \ln \mathbf{g} \rangle + \frac{1}{2} \langle \rho | \mathbf{u} \rangle + P_{id} \Delta V$.

Figure 5.3 demonstrates the scaling of hydration free energies from PC and PC+ corrections. As can be seen, the $\langle \rho | \ln \mathbf{g} \rangle / 2$ term in the PC correction scales with solute surface area and defines its surface tension. Notice that PC+ effectively sets liquid pressure to its ideal value, which for liquid water at ambient conditions is $P_{id} = 1372$ bar. It is quite a bit larger than the pressure of water at standard conditions and leads to the overestimation of solvation free energy for larger volumes. However, the approach, at least for hard spheres, is relatively accurate up to $R \approx 1$ nm. This is twice the size of fullerene C₆₀ and covers much of the domain of conventional pharmaceutical, analytic and organic chemistries.

Since the beginning of the chapter we have been discussing hard spheres. While these solutes are quite convenient from a theoretical point of view, almost none of the actual

^cNote that the correction was discovered essentially by accident, and was initially referred to as initial state correction (ISc) ^{39,140}. Only after publication by Sergiievskiy et al. ³⁸ it was recognized that the correction was related to 3D-RISM pressure and the name PC+ became popular ^{40,42,115,116}.

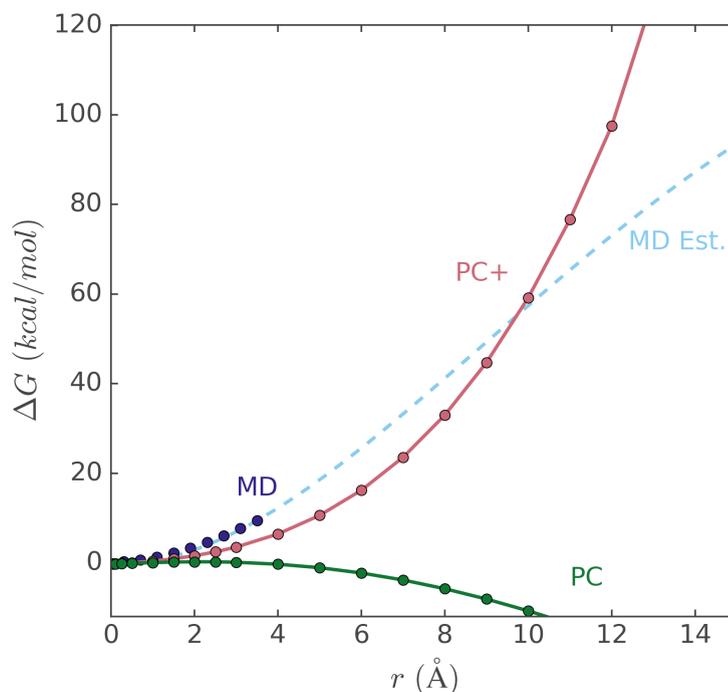


FIGURE 5.4: Solvation free energies of Lennard-Jones solutes with different radii and $\epsilon = 0.125$ kcal/mol in water. The estimate of MD solvation free energies for larger radii was done using equation 5.28.

molecules resembles them. Between pretty much any two materials there would exist dispersion interactions, typically approximated by Lennard-Jones potential.

The attractive forces are usually quite weak, on the order of $0.2kT$, but are enough to practically remove most dewetting effects between water interface and surface, making pair distribution function predicted by 3D-RISM and molecular dynamics much more similar. Figure 5.4 shows hydration free energies for a series of Lennard-Jones spheres, predicted by molecular dynamics, PC, and PC+. To extrapolate results of molecular dynamics to higher radii we estimated the contribution of dispersive interactions assuming that water structure is unperturbed by the sphere outside its exclusion radius (a reasonable approximation for these solutes^{136,141})

$$U_{disp} = \rho 4\pi \int_{\sigma_{LJ-O}}^{\infty} 4\epsilon_{LJ-O} \left[\left(\frac{\sigma_{LJ-O}}{r} \right)^{12} - \left(\frac{\sigma_{LJ-O}}{r} \right)^6 \right] r^2 dr, \quad (5.28)$$

where both σ_{LJ-O} and ϵ_{LJ-O} were computed using Lorentz-Bertholetz rules. The total solvation free energy was estimated via $\Delta F_{LR-LJ} = U_{disp} + \gamma A$, where we used $\gamma = 76 \text{ mJm}^{-2}$ ¹³⁷.

Overall, the results are similar to those that were obtained for hard spheres. PC+, due to its ideal pressure scales similarly to molecular dynamics up to $r \approx 1$ nm. Because dispersion interactions are relatively weak, their description within both molecular dynamics and 3D-RISM is similar, changing the picture little compared to that of hard spheres.

5.4 Free energy of charging

Consider a soft sphere with a charge q located at the centre. Its interaction potential with the surrounding solvent can be expressed as:

$$U_{uv}(q) = U_{LJ} + U_{el} = U_{LJ} + q\Phi, \quad (5.29)$$

where Φ is the solvent generated electrostatic potential in the centre. The charging free energy is the reversible work required to change the solute charge from 0 to q . It can be found using Kirkwood's charging formula:

$$\Delta F_{el} = \int_0^q \left\langle \frac{\partial U_{uv}(q')}{\partial q'} \right\rangle_{q'} dq' = \int_0^q \langle \Phi \rangle_{q'} dq', \quad (5.30)$$

where $\langle \Phi \rangle_{q'}$ is the electrostatic potential in the centre of the ion with charge q' .

It is commonly stated that the charging free energy in water can be well approximated using a linear response relationship. Technically speaking, it is a bit more complicated; a standard linear response implies $\langle \Phi \rangle_{q'} = q' \langle \Phi \rangle_1 + (1 - q') \langle \Phi \rangle_0$. However, for water (and other dipolar solvents) one typically encounters a piecewise-linear (piecewise-affine) response^{63,142}; that is:

$$\langle \Phi \rangle_{q'}^{PL} = \langle \Phi \rangle_0 - \begin{cases} q' C_+ & \text{if } q' \geq 0 \\ q' C_- & \text{if } q' < 0. \end{cases} \quad (5.31)$$

Both the potential in the uncharged cavity $\langle \Phi \rangle_0$ as well as constants C_+ and C_- depend on the size and "stickiness" of the solute.

Notably, theories which model water as a polarizable continuum cannot predict asymmetry of the solvent response, regardless of whether they take into account local or non-local polarizabilities¹⁴². Popular approaches such as the Born model or Poisson-Boltzmann

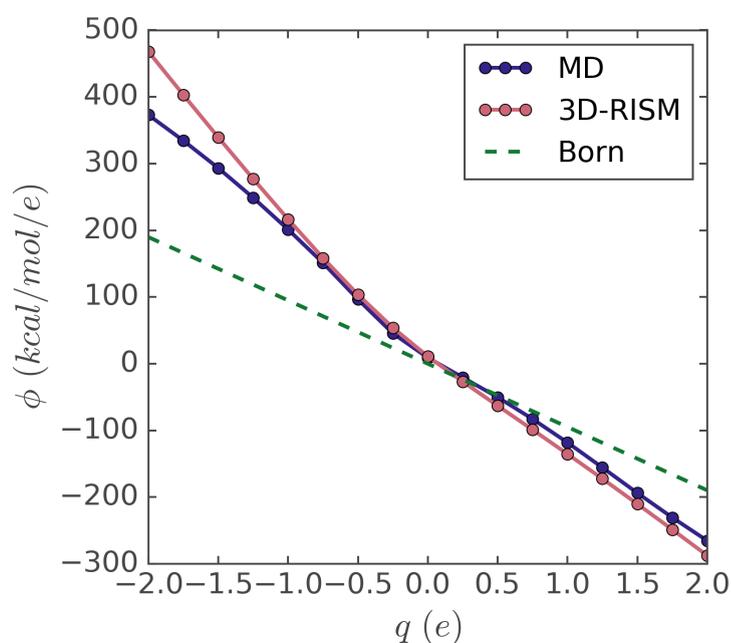


FIGURE 5.5: Electrostatic potential inside the soft solute with $\sigma = 3.8 \text{ \AA}$ and $\epsilon = 0.125 \text{ kcal/mol}$ depending on its charge.

and their modifications incorporate charge asymmetry by scaling the ion radius. This makes RISM models quite interesting, since they do predict piecewise linear response without any parametrization.

Evidently, the reason for RISM "awareness" of solute charge is due to the use of two separate fields: one for water oxygens and one for hydrogens¹⁴³. Figure 5.5 demonstrates the dependence of the water potential on the solute charge, predicted by three different models. We can see that while 3D-RISM correctly predicts the charge dependence of the water response, its scaling is predicted to be strictly linear, in agreement with the equation 5.31. At the same time, more precise molecular dynamics simulations do show deviations from it at the higher charges. The effect is due to dielectric saturation: in discrete solvent after certain point the polarization reaches maximum density and response becomes sublinear⁶³. Such saturation does not occur in 3D-RISM; similarly, in the case of hard sphere solvation 3D-RISM predicted response was always exactly proportional to the solute's volume.

Even though 3D-RISM does not predict saturation of the dielectric response, the accuracy of its approximation at moderate charges is more than enough to correctly predict charging free energies for the majority of common molecules. Figure 5.6 demonstrates predictions

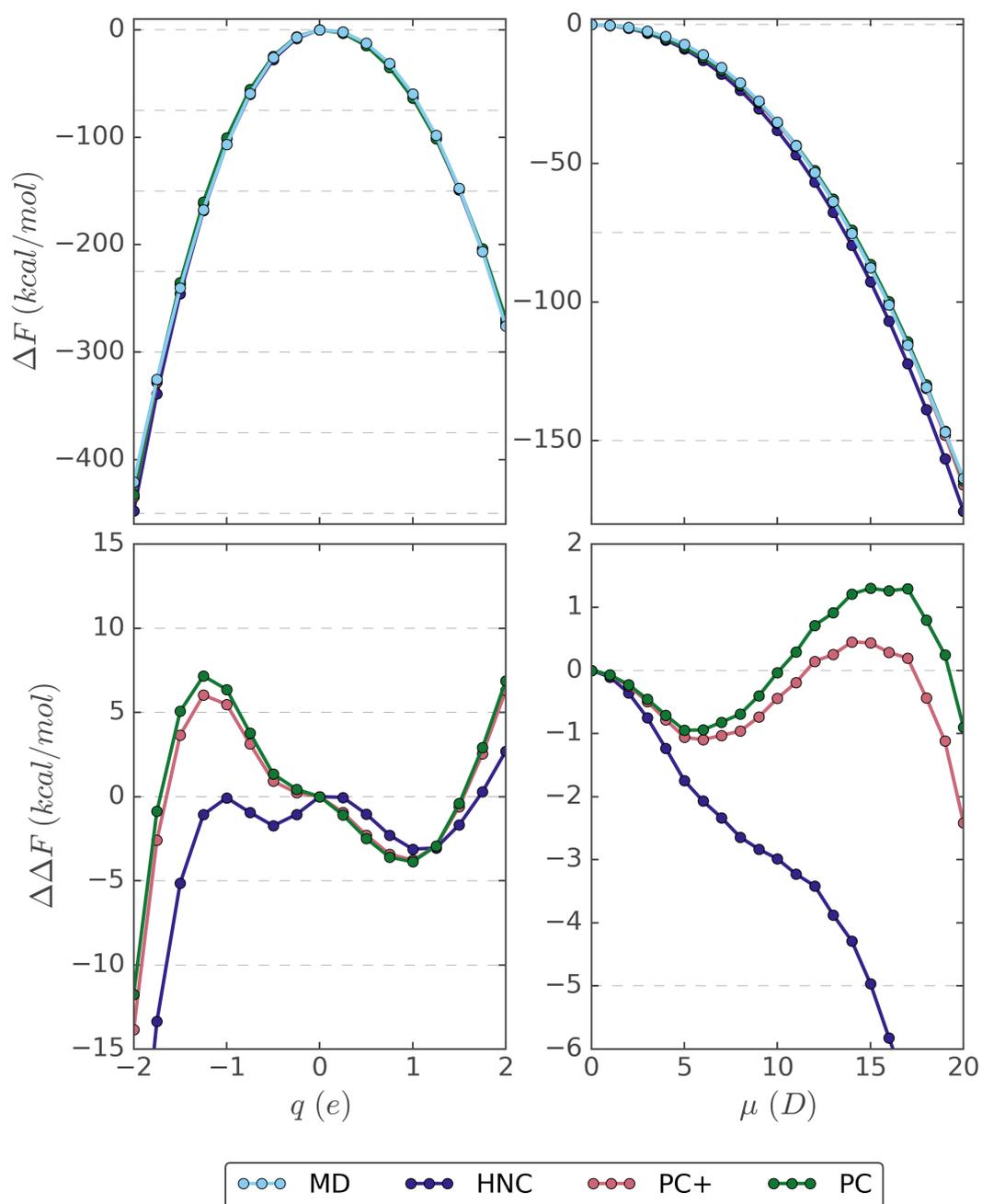


FIGURE 5.6: Top figures: charging free energy of Lennard-Jones sphere (left) and Lennard-Jones dipole (right) in water. The dipole consists of two Lennard-Jones spheres, separated by 2 \AA . Both the sphere and dipole sites have $\sigma = 3.8 \text{ \AA}$ and $\epsilon = 0.125 \text{ kcal/mol}$. Bottom figures: difference in charging free energy predictions between 3D-RISM models and molecular dynamics.

of charging free energies by 3D-RISM for a simple sphere and dipole. The volumes of van-der-Waals solutes do not depend on the charge, but their excess volumes still change due to the increased attraction. Thus, it makes sense to use pressure corrections even in the context of charging free energies. As the figure illustrates, this does not show considerable improvement for single ions, but improves solvation free energy predictions for dipoles.

5.5 Effect of corrections on solvation thermodynamics

In this section, we will discuss solvation thermodynamics of 3D-RISM/HNC and its pressure corrections. For simplicity and along with the main goal of the thesis, we restrict the discussion to single-component solvents, although extension to multicomponent mixtures should be relatively straightforward.

Previously (section 2.5) we have shown that solvation free energy can be split into energetic and entropic contributions. Specifically, for solvation in the grand canonical ensemble we had $\Delta\Omega_s = E_\mu^{uv} - TS_\mu^{uv} + \mu\rho\bar{V}$. Examining the derivation of 3D-RISM one can see that $\mu\rho\bar{V} = \mu\rho \int \Delta\rho(\mathbf{r}) d\mathbf{r}$ gets canceled out because of the way we define excess intrinsic chemical potential (section 4.5). Thus, one can split 3D-RISM solvation free energy into $\Delta F_{3D-RISM} = E^{uv} - TS^{uv}$, where E^{uv} takes a clear physical meaning, and TS^{uv} is simply defined using the above equation. We will see that such decomposition provides a sensible way of analyzing 3D-RISM and its pressure corrections.

We start by defining solute-solvent interactions in the usual way $E^{uv} = \langle \rho | \mathbf{u} \rangle$. Then solute-solvent entropy is given by $S_{HNC}^{uv} = -k \langle \rho | \ln \mathbf{g} \rangle + k \langle \Delta\rho | \mathbf{1} \rangle + k/2 \langle \Delta\rho | \mathbf{C} | \Delta\rho \rangle$. Using a rough estimate $\langle \Delta\rho | \mathbf{C} | \Delta\rho \rangle \approx \Delta\hat{\rho}(0)\hat{\mathbf{C}}(0)\Delta\hat{\rho}(0)$ and the fact that for most liquids $\hat{\mathbf{C}}(0) < 0$, we can see that S_{HNC}^{uv} is negative, consistent with the general result from section 2.5.

The ΔS_{HNC}^{uv} can be split into the ideal gas entropy change, given by $\Delta S_{HNC}^{id} = -k \langle \rho | \ln \mathbf{g} \rangle + k \langle \Delta\rho | \mathbf{1} \rangle$, as well the excess (or ring entropy, as it has been referred to by some authors^{144,145}) contribution $\Delta S_{HNC}^{ex} = k/2 \langle \Delta\rho | \mathbf{C} | \Delta\rho \rangle$. The minimization of the grand potential leads to

$$\Delta S_{HNC}^{uv} = -k/2 \langle \rho | \ln \mathbf{g} \rangle + \frac{1}{2T} \langle \rho | \mathbf{u} \rangle - \frac{1}{T} P_0 \Delta V. \quad (5.32)$$

The above result is interesting since we can readily interpret this entropy as a sum of logarithm of insertion probability for a hard solute within HNC model (equation 5.24) and a linear response entropy change occurring due to introduction of attractive interactions. Recall that $-T\Delta S_{LR}^{uv} = -1/2E^{uv} + 1/2E_0^{uv}$, where $1/2E_0^{uv}$ corresponds to solute-solvent energy without any coupling between solute and solvent; for a hard solute $E_0^{uv} = 0$ and we recover equation 5.32. Note that the presence of a linear response component in singlet HNC is not surprising, considering that in the section 4.1 we demonstrated that this model essentially treats ideal part of chemical potential exactly and excess part via linear response approximation.

A similar result has been obtained by Sanchez et al^{146,147} using a more general approach. They found that

$$\Delta S^{uv} = k \ln P_{ins} - k \ln \langle \exp [-\beta(E_I^{uv} - E^{uv})] \rangle_a - k \ln P_a, \quad (5.33)$$

where P_{ins} is the probability that a randomly inserted molecule will experience an attractive or zero interaction energy $\Delta E^{uv} \leq 0$; the second term is the familiar solute-solvent fluctuation energy, except the averaging is performed over the states where solute-solvent interactions are attractive. P_a is the probability that a fully inserted molecule will have an attractive interaction energy, which for the majority of normal molecules ≈ 1 . We can see that by setting $k \ln \langle \exp [-\beta(E_I^{uv} - E^{uv})] \rangle_a = -(1/2T)E^{uv}$ (this result is exact in the linear response regime¹⁴⁸) one recovers the HNC solute-solvent entropy.

Thus, after analyzing HNC entropy we found it to be consistent with other statistical mechanics theories. Its main problem is the overestimation of the hard solute insertion free energy, which as we have already discussed, stems from the failure to describe interface formation. Then we can readily interpret PC and PC+ corrections as adjustments to incorrect P_{ins} from the HNC approximation. For PC, entropy becomes

$$\Delta S_{PC}^{uv} = -k/2 \langle \rho | \ln \mathbf{g} \rangle + \frac{1}{2T} \langle \rho | \mathbf{u} \rangle \quad (5.34)$$

and for PC+

$$\Delta S_{PC+}^{uv} = -k/2 \langle \rho | \ln \mathbf{g} \rangle + \frac{1}{2T} \langle \rho | \mathbf{u} \rangle + k \langle \Delta \rho_0 | 1 \rangle, \quad (5.35)$$

where $\Delta \rho_0$ is the change in solvent density.

From the equations above it is not necessarily obvious which approximation should lead to a better estimate of solvation free energy. A clearer picture can be obtained if we rewrite the expression in terms of ideal/excess contributions. Both PC and PC+ models have identical excess entropies $\Delta S_{PC}^{ex} = \Delta S_{PC+}^{ex} = k/2 \langle \rho_0 | \mathbf{C} | \Delta \rho \rangle$. The ideal entropies are then $\Delta S_{PC}^{id} = -k \langle \rho | \ln \mathbf{g} \rangle$ and $\Delta S_{PC+}^{id} = -k \langle \rho | \ln \mathbf{g} \rangle + k \langle \Delta \rho_0 | 1 \rangle$. For a single component molecular solvent the solute-solvent entropy can be expanded in terms of n-particle correlation functions, with first terms given by^{149,150}

$$\begin{aligned} \Delta S^{uv} = & -k\rho_0 \int g(\mathbf{r}) \ln g(\mathbf{r}) d\mathbf{r} + k\rho_0 \int g(\mathbf{r}) - 1 d\mathbf{r} \\ & - k\rho \frac{V_i}{\Omega} \int g(\omega) \ln g(\omega) d\omega + \dots, \end{aligned} \quad (5.36)$$

where ρ is the solvent density, ω is the Euler angle, $\Omega = \int d\omega$, and V_i is the unit volume. From the expression above we can see that the PC+ model, unlike PC, contains both first terms in the expansion. Thus, one can expect it to reproduce solute-solvent entropy slightly better. In all site-site models, the orientational contribution (the second line of equation 5.36) is partially approximated by summation of $\rho \ln g$ terms over different solvent sites.

A more straightforward way of analyzing 3D-RISM thermodynamics is to simply compare it directly to molecular dynamics. As usual, we chose water as our solvent and the first six linear alkanes (methane to hexane) as our solutes due to availability of data. The molecular dynamics simulations were performed by Gallicchio et al.¹⁵¹ who used TIP4P water model and OPLS force field parameters for alkanes. We use the same force field, combined with cSPC/E water model for 3D-RISM.

Before we proceed it is important to discuss the way we compute solvation entropies and enthalpies within different models. Since 3D-RISM is formulated in the grand ensemble, that is, under conditions of constant temperature, volume and chemical potential, it can seem that we can only compute properties within this particular ensemble. However, it is possible to work around this problem by evaluating necessary derivatives numerically over multiple simulations in which only necessary thermodynamic variables are varied and others are kept constant. Recall that as long as macroscopic thermodynamic parameters are identical, the chemical potential will be independent of the ensemble. Thus, to evaluate, for example, temperature derivative of chemical potential under constant pressure, we run calculations at two separate temperatures, but identical pressures.

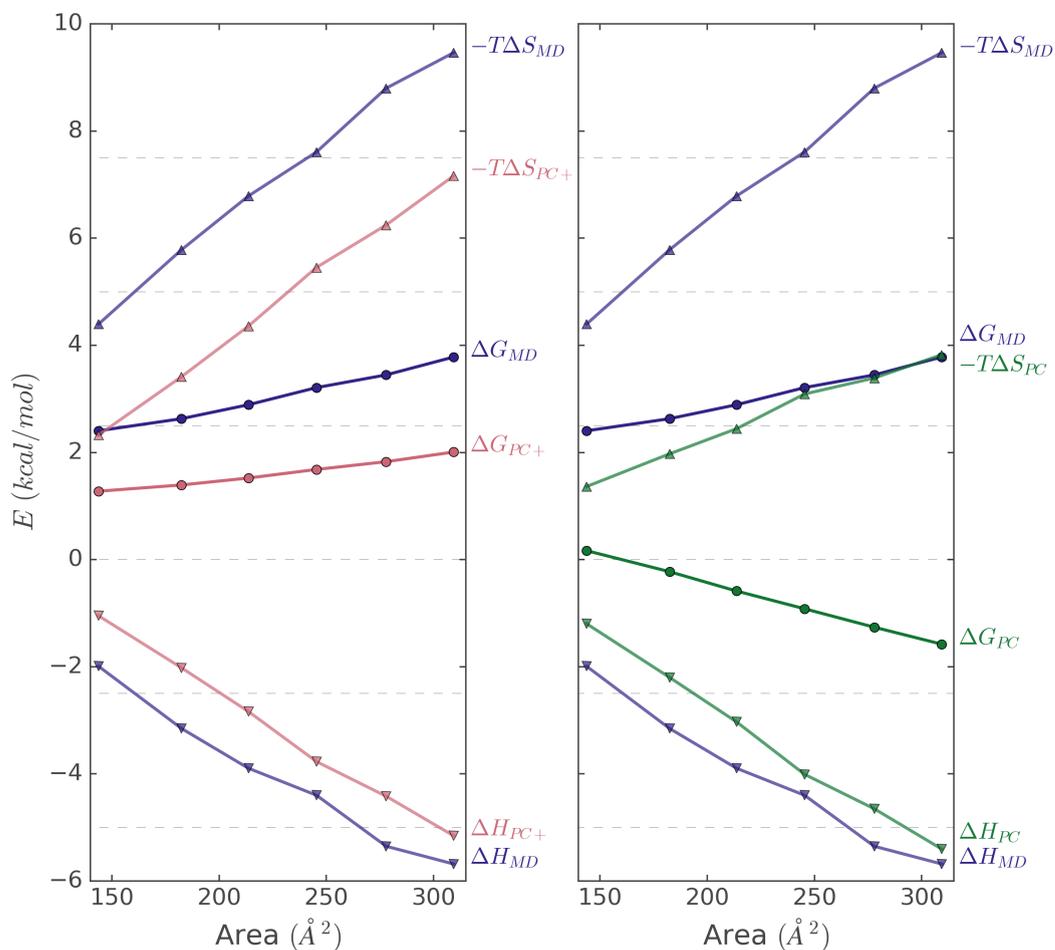


FIGURE 5.7: Dependence of free energy, entropy and enthalpy of linear alkanes on their surface area (number of atoms). The molecular dynamics results are taken from Ref. 151.

The approach described above will yield the best estimates we can get with RISM for constant pressure enthalpies, entropies as well as their higher derivatives; however, it will also lead to a conceptual problem. The set of temperatures and densities (input parameters for 3D-RISM calculation) corresponding to a constant pressure in real water actually leads to a variety of different values within the 3D-RISM approximation. Thus, any 3D-RISM entropies and enthalpies that we obtain via standard formulas will actually contain contributions from the derivatives of 3D-RISM pressure. This is not a very significant problem since essentially any water model will have a different phase diagram, making these dependences additional error contributions. For simplicity, we are going to use the symbol ΔG for both experimental (simulated) and 3D-RISM free energies, even

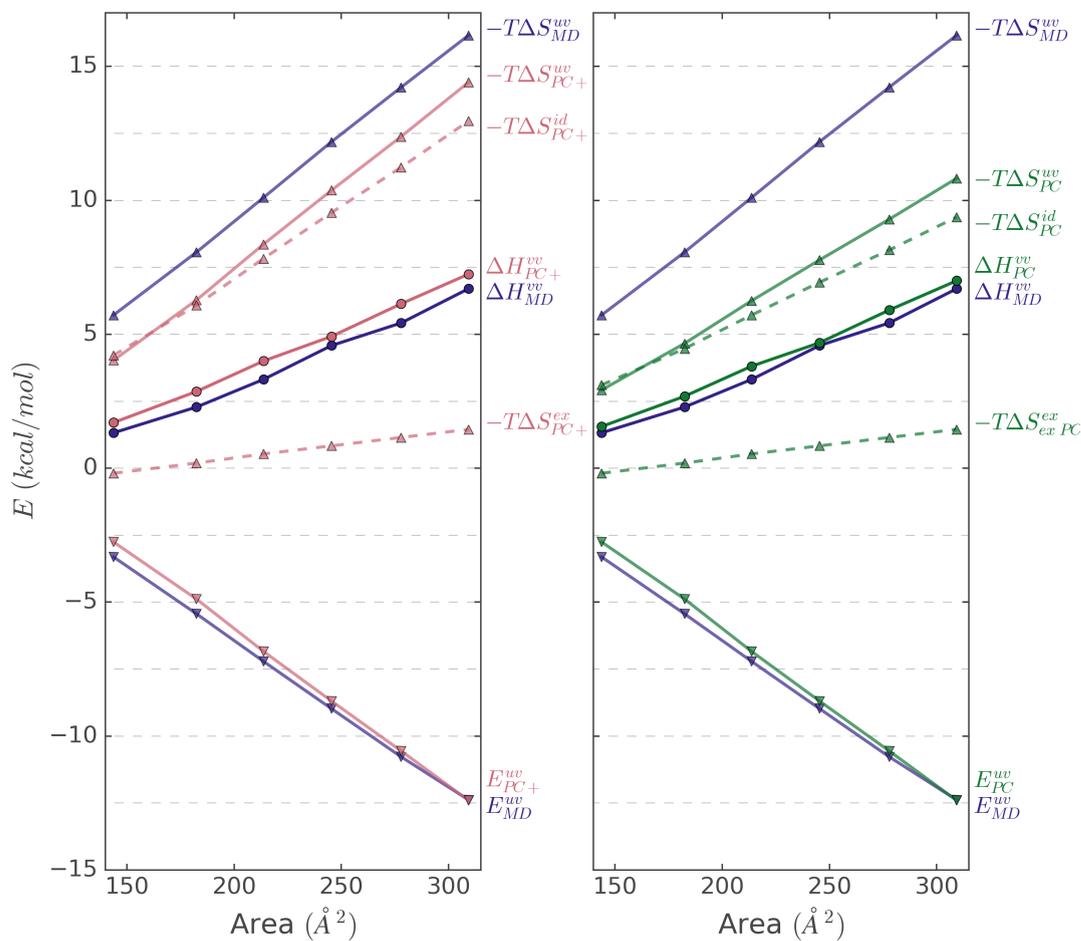


FIGURE 5.8: Solute-solvent components of alkanes solvation free energy. The definition of the quantities can be found in the main text. The molecular dynamics results are taken from Ref. 151.

though the later technically corresponds to $\Delta\Omega$.

Figure 5.7 demonstrates predictions of solvation free energy, enthalpy, and entropy obtained from molecular dynamics, PC and PC+. For reference, the same figure with the comparisons between uncorrected 3D-RISM and MD is included in appendix B (figure B.3). The solvation entropy from RISM models was calculated numerically via:

$$\Delta S(T) = -\frac{\Delta G(T + \Delta T) - \Delta G(T - \Delta T)}{\Delta T}, \quad (5.37)$$

where we used $\Delta T = 2.5 \text{ K}$; to get enthalpy we used

$$\Delta H = \Delta G + T\Delta S. \quad (5.38)$$

The density of bulk water at different temperatures was taken from Ref. 152. Note that these predictions are for thermodynamic entropies and enthalpies, related to their solute-solvent components as $\Delta S = \Delta S^{uv} + \Delta H^{vv}/T$ and $\Delta H = E^{uv} + \Delta H^{vv}$, where ΔH^{vv} is the solvent reorganization energy.

The alkanes are essentially chains of fused Lennard-Jones spheres. The behaviour of solvation free energy that we observed in the case of a single Lennard-Jones sphere (figure 5.4) is essentially identical for these solutes. In the case of PC, the dispersion interactions between solute and solvent dominate, making the free energy ΔG_{PC} become progressively negative with the increase of surface area. Conversely, the PC+ approximation of insertion probability results in a correct scaling of ΔG with solute size, although its value is smaller than the one predicted by molecular dynamics.

The predicted solvation enthalpy and entropy are smaller in magnitude for both 3D-RISM models when compared to molecular dynamics. The differences between PC+ and PC models is mostly due to solvation entropy ΔS . The enthalpies in the two models differ only slightly due to the temperature dependence of water density.

On figure 5.8 you can see a further decomposition of previous quantities into solute-solvent and solvent-solvent terms. The solvent reorganization is obtained from $\Delta H^{vv} = \Delta H - E^{uv}$. Perhaps surprisingly, the overall agreement of these quantities between 3D-RISM and MD is quite good, despite the fact that all solute-solvent energies and entropies do depend on the ensemble. Notably, E^{uv} is predicted essentially correctly, considering the fact that water models were not identical in 3D-RISM and MD. This indicates that for smaller solutes the solvent density distribution within 3D-RISM/HNC is in relatively good agreement with MD, and disagreement primarily comes from the entropic part. Additionally, as we expected from our analysis in the beginning of the chapter, PC+ model has a better S^{uv} estimate, mostly due to its ideal part.

The accuracy of ΔH^{vv} predictions is quite interesting, considering that this quantity is significantly overestimated in 3D-RISM/HNC (figure B.3). Within the linear response approximation one has $\Delta H^{vv} = -\frac{1}{2}E^{uv}$ 153. As figure 5.9 demonstrates, this result is

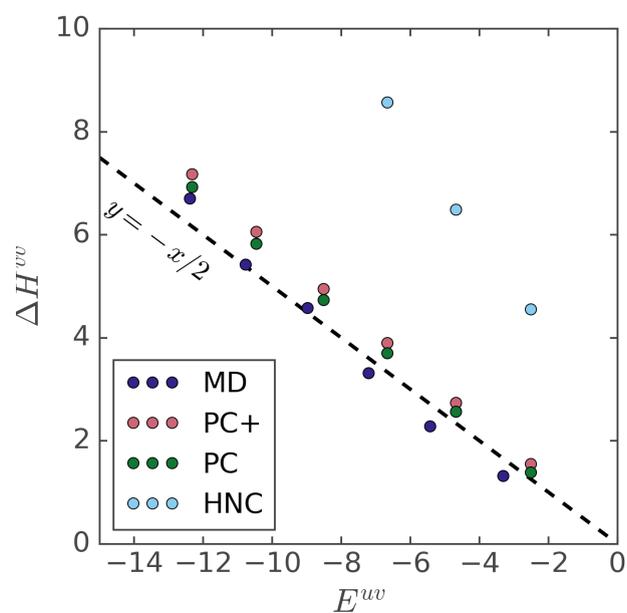


FIGURE 5.9: Correlation between solvent reorganization energy and solute-solvent interaction energy for the first 6 linear alkanes. The units are in kcal/mol.

relatively accurately satisfied by molecular dynamics, as well by PC and PC+ models. Once we get rid of overestimated insertion free energy contribution to 3D-RISM/HNC, the model starts giving a number of predictions that treat solute-solvent interaction via the linear response approximation.

Chapter 6

Applications

In contrast to the previous chapter, here we are primarily concerned with realistic solutes and comparison with experimental measurements. We discuss the solvation of neutral and charged molecules in water at various temperatures. The accuracy of pressure corrected 3D-RISM models is compared to other approaches. The chapter is based on two previously published articles: Refs. [39](#) and [40](#).

6.1 Neutral molecules

We start by comparing PC and PC+ predictions to molecular dynamics hydration free energies. The differences in the results then occur primarily due to the approximations in the model and not due to the inaccuracies of the force field, which would have been a major error source when comparing to experiment.

Unfortunately, the popular 3D-RISM water models are different from water models used in molecular dynamics. As we saw in section [4.5](#), the hydrogen atoms' density can behave quite independently from the oxygen one. Thus, leaving hydrogen atoms without Lennard-Jones parameters, which is commonly done in the water models used for molecular dynamics, will inevitably cause hydrogens to "spill" into the solutes. To avoid it, all 3D-RISM water models employ small Lennard-Jones parameters on hydrogen atoms.

One of the largest evaluations of hydration free energies using molecular dynamics was done by Mobley et al^{[13](#)}. They computed and published both Lennard-Jones and electrostatic

contribution to hydration free energies of 504 molecules, using quite long simulation runs to ensure low uncertainty. The set of solute molecules was quite diverse and included all the main functional groups such as alcohols, carboxylates, aromatic compounds, amines, etc. All solutes were described using general amber force field (GAFF) and AM1-BCC partial charges. Water was approximated using the standard TIP3P model⁵⁹.

For 3D-RISM calculations, we used the same solute potentials. We did not try to take into account their conformations and simply used a single minimised geometry. Water was described by the conventional cSPC/E model¹⁰⁵, which differs from ordinary SPC/E model by Lennard-Jones potentials on hydrogen atoms, mentioned earlier. More technical details are summarised in the appendix A.

The comparison between HNC results and molecular dynamics was already provided in figure 4.10. The performance of PC and PC+ models is presented in figure 6.1. As we can see, the PC model fails to predict nonpolar free energies ΔF^{np} (which correlate with cavity creation free energy), while PC+ does capture general trends. The polar part of hydration free energy ΔF^p is approximated with a reasonable accuracy by both models. The total hydration free energy is estimated by PC+ quite well.

Both models perform worse for polar molecules that have more negative solvation free energies. The larger outliers tend to be the molecules containing negatively charged oxygen or hydrogen atoms, such as 2-ethoxyethanol or hydrazine. It is quite likely that the presence of Lennard-Jones sites on hydrogen atoms becomes increasingly important and thus 3D-RISM describes them somewhat differently. Another potential source of error is a lack of conformational sampling, but it is unlikely to cause a large effect.

After comparing our corrections to MD, we decided to compare their accuracy against the actual experimental data. The molecules were taken from the already familiar dataset by Mobley et al. The uncertainty of experimental values in the set was estimated by authors to be around 0.2 kcal/mol. Figure 6.2 demonstrates hydration free energy predictions of MD, solvation model density (SMD)¹⁵⁴, PC+, NgB, and UC models to experimental measurements.

SMD is the most accurate model among those compared, with an error of 1.1 kcal/mol. When using SMD, we followed the recommended protocol, involving running two electronic structure calculations, which we performed in both vacuum and liquid phases¹⁵⁴. The

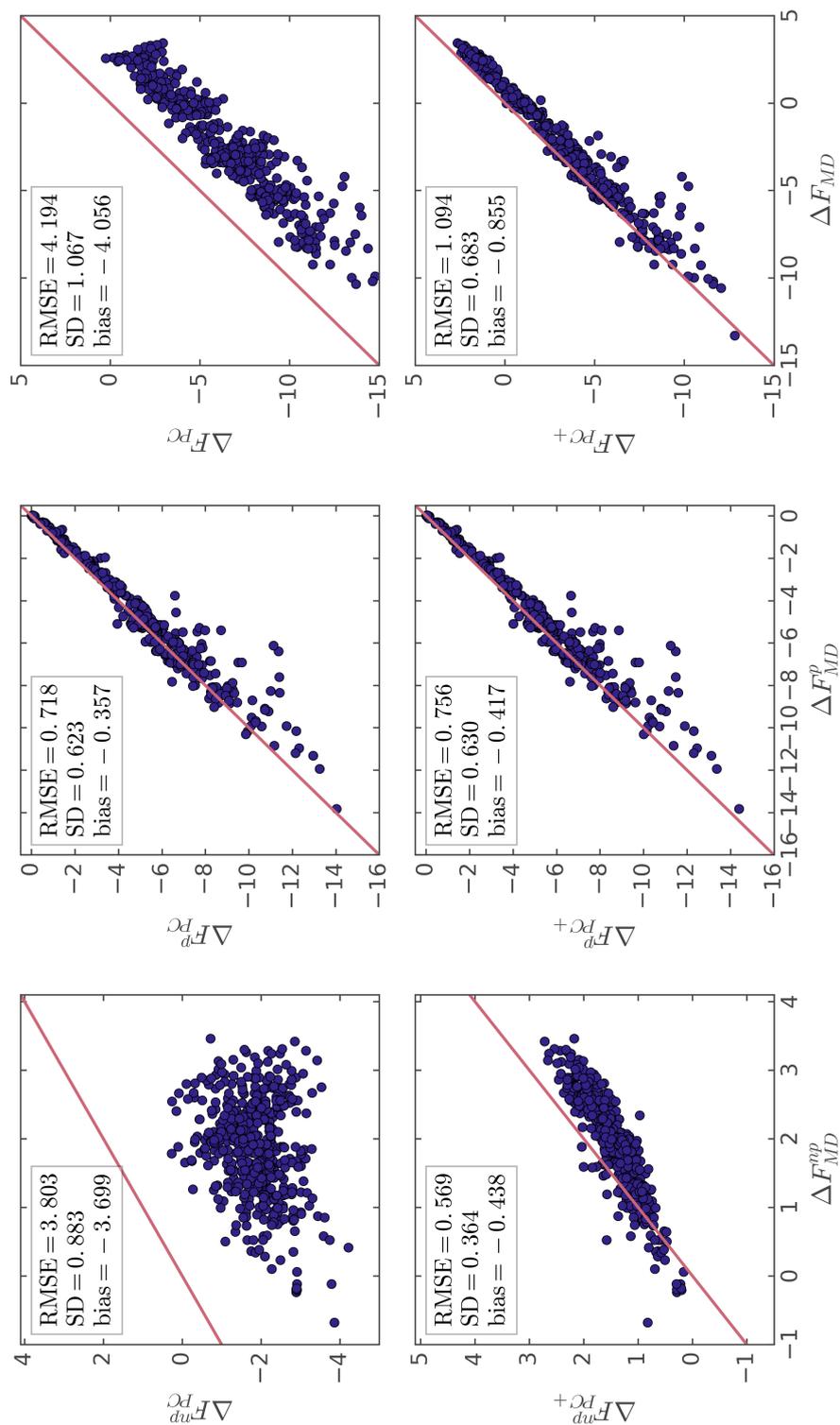


FIGURE 6.1: Comparison of PC (top) and PC+ (bottom) free energies with molecular dynamics results, obtained by Mobley et al.¹³. In the first column we compare nonpolar (Lennard-Jones) components, in the second, electrostatic contributions, and in the third, total hydration free energies. All values are in kcal/mol.

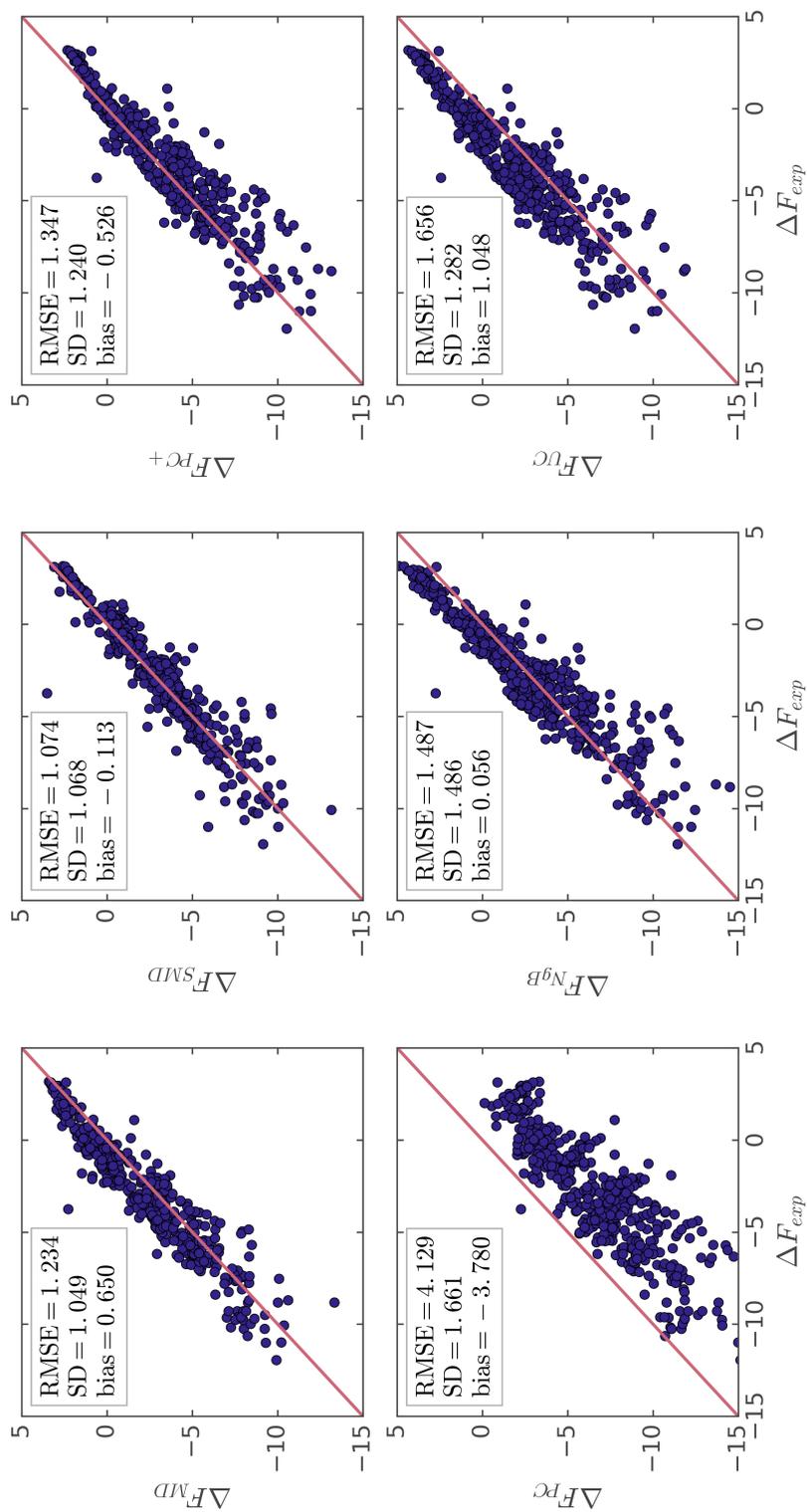


FIGURE 6.2: Hydration free energies by different models, compared to experimental data. Both molecular dynamics results and experimental values come from Ref. 13. All values are in kcal/mol.

geometries were optimised in both phases, using the M06-2X functional¹⁵⁵, combined with the MG3S basis set¹⁵⁶. The free energy was computed by subtracting the molecules energy in water from the energy in the vacuum. Thus, SMD is the only model which takes into the account polarization contribution to hydration free energy. We also expect the continuous charge distribution from quantum calculations to be more accurate than the point charges used in MD and RISM models.

A major shortcoming of SMD is that it estimates the non-polar part of free energy empirically, utilizing an equation based on solvent surface tensions, parametrized to fit experimental hydration free energies¹⁵⁴. This limits the applicability of the method to compounds without complicated structure and functional groups. However, for simple and small molecules such as those which were the part of the Mobley's dataset, SMD and related methods are probably the best choices since they can utilize accurate charge distributions from quantum mechanics and their empirical approaches to estimating cavity energies are not too inaccurate.

The PC+ model error is not much larger than the one seen from MD. The systematic errors due to its approximation of free energy as well as errors of utilised force field (GAFF/AM1-BCC) cancel out favourably, making the overall accuracy quite good. It performs better than 3D-RISM models with empirical corrections such as NgB and UC, but shows results which are poorer than MD and SMD. Overall, it seems that while PC+ is moderately accurate, it still performs worse than other common methods for predicting hydration free energies of neutral molecules.

Other 3D-RISM based models such as PC, UC, and NgB perform worse. We already discussed the issues with PC at the beginning of the section. The performance of UC on the other hand can likely be improved by a more careful choice of parameters used to fit the model (we used the values provided in section 4.8).

Before closing this section, it is interesting to discuss possible effects of the force field on the results. As molecular dynamics results show, a substantial part of the error can be attributed to the inaccurate solute-solvent interaction energy. Thus, a more accurate force field might in principle substantially improve the accuracy of the result.

We did not want to explore these possibilities in too much detail as, due to the number of force fields and water models available, such as investigation would warrant a separate

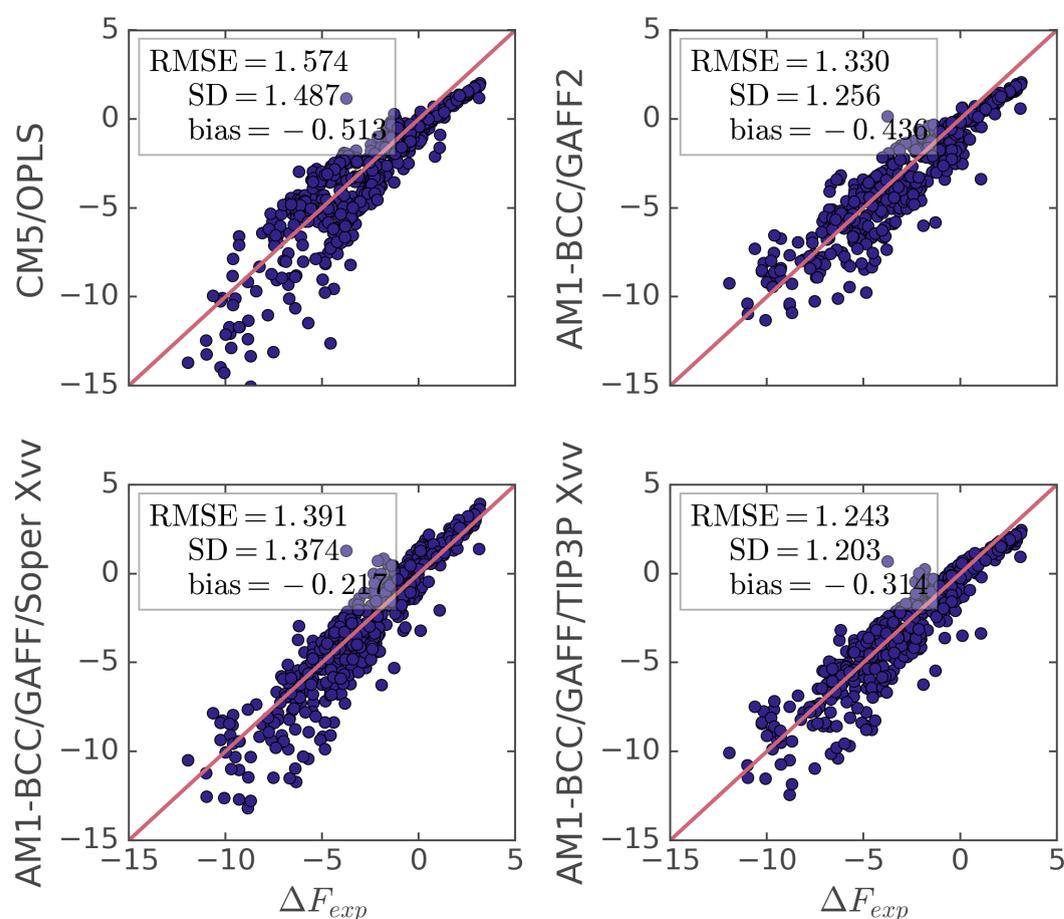


FIGURE 6.3: The influence of different force fields on the accuracy of PC+ results when compared to experiment. All values are in kcal/mol.

study. However, as a test we performed a few calculations investigating the sensitivity of PC+ results to the choice of interaction potential energy between solute and solvent (figure 6.3). In the first two tests (demonstrated on the first row) we used the standard cSPC/E water susceptibility functions computed with 1D-RISM, and varied Lennard-Jones and partial charges on solutes. Neither combination of CM5 and OPLS-2005, nor AM1-BCC and GAFF2 (an updated version of GAFF force field, introduced in AmberTools version 16¹⁵⁷) substantially improved the accuracy of free energy predictions, when compared to experiment. At the same time, standard parameters, combined with other water susceptibility functions, do slightly improve the accuracy of PC+ results (second row in the figure). We used standard 1D-RISM calculations with the cTIP3P¹⁰⁵ water model for one test and carefully smoothed experimental water radial distributions by Soper et al. for another¹⁵⁸. As these results show, more sophisticated or empirical water models might present an easy route towards further improving solvation free energy predictions

from PC+ approximation.

6.2 Predicting the temperature dependence of ΔG

It is difficult to predict hydration free energies at non-standard conditions or estimate their derivatives. First of all, the majority of the empirical/semi-empirical models are parametrized at 298 K and are only suited for computing solvation free energy and nothing else. Even approaches such as molecular dynamics, which can in principle access a wide range of conditions, usually require very extended calculation runs to estimate free energy derivatives with low uncertainty.

The above makes 3D-RISM based models interesting since they can be used at any thermodynamic conditions and can produce results with high numerical accuracy.

To test 3D-RISM at non-standard conditions, we used a set of experimental solvation free energies, measured at many different temperatures. For each compound, that dataset had at least five hydration free energy measurements, all made at different temperatures between 0 and 100 degrees Celsius. The data was compiled by Chamberlin et al., and presented in Refs. 159 and 160; for this reason, we will refer to this data as the Chamberlin dataset. To extract accurate solvation entropies and heat capacities from this data, we fit the following equation to all measurements, discarding the molecules to which this relationship fits poorly:

$$\Delta G(T) = \Delta G(T^*) - \Delta S(T^*)(T - T^*) + \Delta C_s(T^*) \left[T - T^* - T \ln \left(\frac{T^*}{T} \right) \right], \quad (6.1)$$

where T^* is an arbitrary temperature, which we set to 298.15 K, $\Delta G(T^*)$ is the solvation free energy at that temperature, $\Delta S(T^*)$ is solvation entropy, and $\Delta C_s(T^*)$ is the solvation heat capacity change, defined as:

$$\Delta C_s(T) = \frac{\partial \Delta H}{\partial T} = \frac{\partial \Delta G + T \Delta S}{\partial T}. \quad (6.2)$$

The equation 6.1 is known to accurately fit the temperature dependence of hydration free energy in quite large temperature ranges (since the dependence of solvation heat capacity on temperature is not very significant) and is used in experiments to measure solvation entropies and heat capacities.

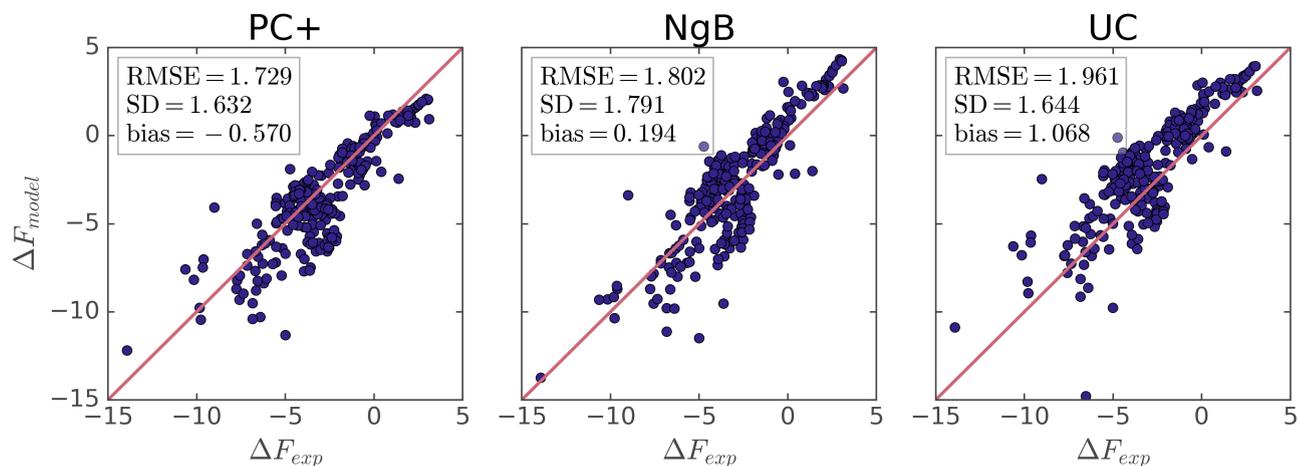


FIGURE 6.4: Comparison experimental and computed hydration free energies at 298 K. The data is taken from the Chamberlin dataset. All values are in kcal/mol.

For computational evaluation of solvation free energies we used geometries of solutes guessed using Openbabel software package^{161,162} and further optimised with OPLS_2005 force field⁶⁰. Each experimental solvation free energy at different temperature was matched by a corresponding 3D-RISM calculation. Water density and dielectric constants at each temperature were evaluated using interpolation functions provided in the Water Society manual¹⁶³ (the relative uncertainty of the density is around 0.0001% and for the dielectric constant is 0.01%). After performing all calculation we used equation 6.1 to extract solvation thermodynamic parameters from 3D-RISM calculations. For all solutes, it fit the data with practically perfect accuracy.

Before analyzing derivatives, we first checked how accurately PC+, NgB, and UC models could predict standard temperature solvation free energy on this dataset. The results are shown in figure 6.4. As you can see, the accuracy of these models on the Chamberlin dataset is lower than on the Mobley dataset. The reason for that is the larger diversity of the Chamberlin dataset, which covers a broader range of molecules and combinations of functional groups. Thus, the simple force field used for all 3D-RISM models (GAFF/AM1-BCC) might be somewhat poorly applicable to them.

Figure 6.5 demonstrates a comparison between predicted and experimental entropies and heat capacities. One can immediately see that free energy derivatives are predicted with much poorer accuracy than free energy itself. The average error in predicted entropies by PC+ accounts for about 20% of its total value. For NgB and UC it is about 70%

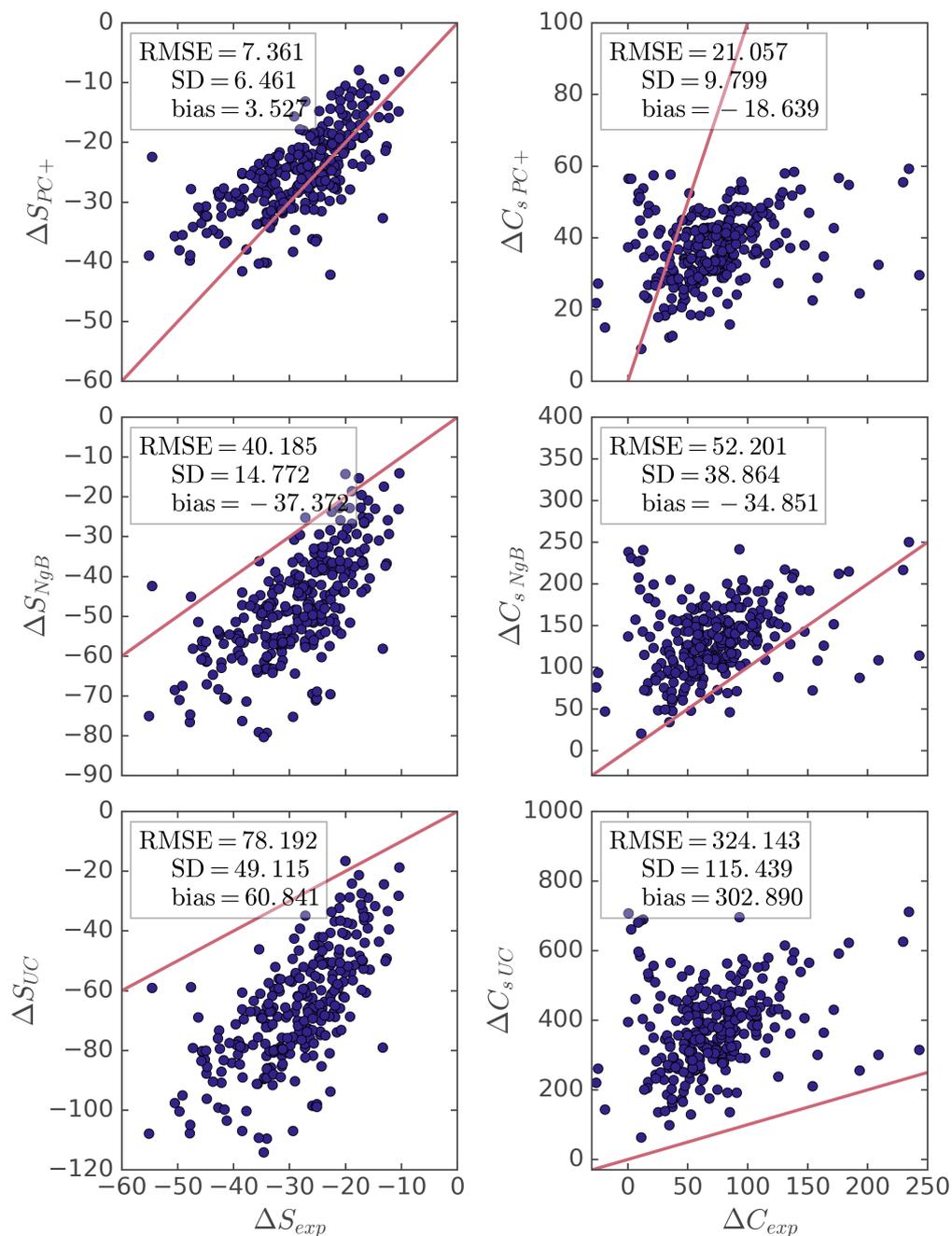


FIGURE 6.5: Comparison of predicted and experimental solvation entropies and heat capacities, evaluated using equation 6.1. Both entropy and heat capacity are in units of cal/mol/K.

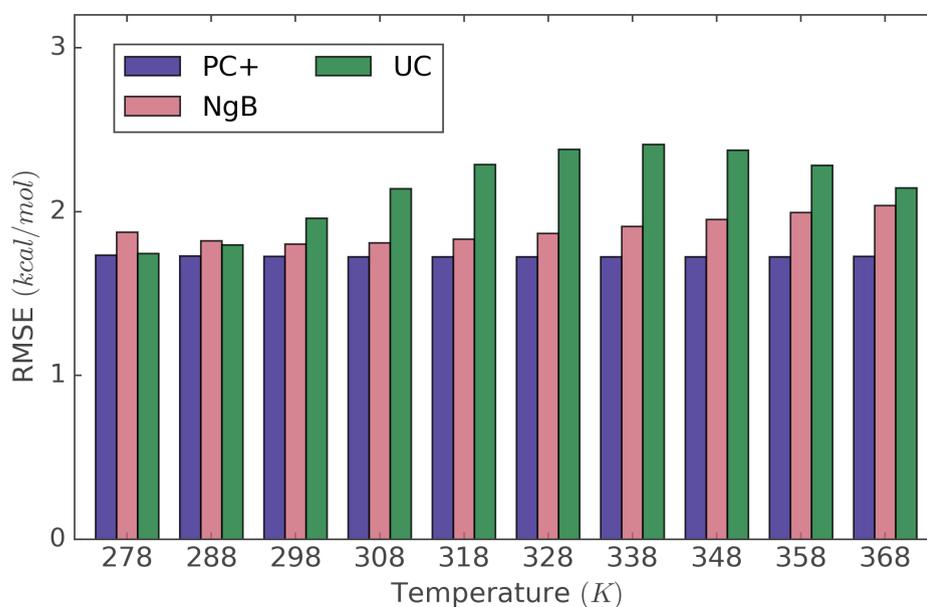


FIGURE 6.6: Root mean square error of solvation free energy of 3D-RISM models on Chamberlin dataset depending on temperature.

and 130% respectively. For solvation heat capacities, none of the models correlated with experiment.

It is not surprising for a model to give reasonable free energies while failing to predict its derivatives. Recall that enthalpic and entropic contributions tend to be larger than solvation free energy and have opposing signs (section 2.5). Other significant factors are the force fields that are parametrized specifically to reproduce free energies, ignoring both solvation entropies and enthalpies.

Despite giving poor estimates for solvation entropies and heat capacities, as figure 6.6 shows, the accuracy of free energies across the 273–373 K range remains similar for NgB, and almost constant for PC+. This is not surprising, considering that the absolute value of solvation free energies at this range, for smaller molecules, changes at most by 1–2 kcal/mol. Moreover, since the error in entropy predictions by PC+ is unbiased, half of the solvation free energy improves towards the higher temperatures (for instance, overestimated ΔG at 298 K combined with underestimated entropy will result in more accurate solvation free energy estimates at higher temperature). The larger problems in solvation heat capacities for all of the models are not significant enough to affect free energies at this temperature range. It is also worth noting that regardless of the somewhat inflated accuracy of solvation free energies at other temperatures, these predictions are

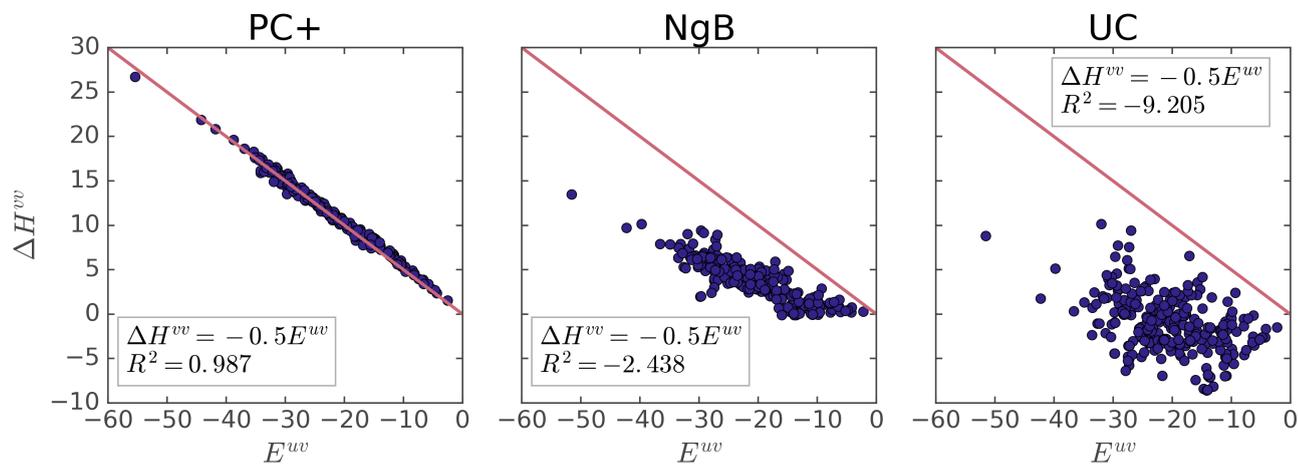


FIGURE 6.7: Correlations between solute-solvent and solvent reorganization energies. R^2 stands for coefficient of determination. All values are in kcal/mol.

still significantly better than "0 hypothesis" estimates, made under the assumption that solvation free energy does not depend on the temperature at all (figure B.2).

Concluding the discussion of solvation thermodynamics we go back to the linear response relationship $\Delta H^{vv} = -E^{uv}/2$, discussed in the section 5.5. As one would expect, the relation only holds for the PC+ model, with empirical model failing to reproduce it. It is possible that re-parametrization of both modes with the aim of recovering this relationship might improve the accuracy of their solvation entropy predictions.

To conclude this section, we analysed the performance of solvation free energy predictions by PC+, NgB, and UC on the extensive dataset of solvation free energies, covering temperatures from 0 to 100 degrees Celsius. Compared to empirically parametrized models, PC+ showed better results across all temperatures. However, even the model with more theoretical basis still cannot accurately predict the derivatives of free energy due to inherent difficulties associated with this task. It is also worth mentioning that very recently Johnson et al. extended both UC and NgB by introducing temperature dependence into empirical coefficients¹¹⁶. This approach might result in improved entropies for empirical models, but it remains to be tested on a larger dataset.

6.3 Ionic solvation

As we demonstrated in the first section of the chapter, using continuum models such as SMD, it is possible to predict solvation free energies of small neutral molecules with around 1 kcal/mol accuracy. The situation is quite different in the case of charged compounds, for which even the most accurate implicit models show relatively poor results (typical accuracies of about 5 kcal/mol for water solvation). The reason for these difficulties is associated with much larger interaction energies between solute and solvent, as well as some effects (such as charge asymmetry) that cannot be described by continuum models.

Considering the above information, applications of models such as PC+ to ionic solvation seems quite promising. In section 5.4 we demonstrated that 3D-RISM is capable of predicting reasonable charging free energy for Lennard-Jones solutes. To test its accuracy in a more realistic setting, we turned to polyatomic ions. Both experimental hydration free energies and solute geometries were taken from the 2012 version of the Minnesota solvation database^{70,164}. These values are based on the hydration free energy of the proton = 265.9 kcal/mol⁶⁷. When selecting compounds from the database, we avoided water clusters as well as ions that were structurally similar to other chosen molecules. We ended up selecting 70 compounds in total: 36 anions and 34 cations.

The non-bonding parameters for ions were derived from GAFF and were combined with AM1-BCC charges. We started by performing all free energy calculations with PC+ and then repeated calculations using molecular dynamics. The results are shown in figure 6.8.

As you can see, the agreement between MD and PC+ is quite good. However, neither of the methods agreed well with experimentally measured values. The reason for that is the lack of explicit water-air boundary in both MD and 3D-RISM. The whole system is schematically demonstrated at the figure 6.9. While in simulations the potential in the empty cavity is ϕ_0 , in experimental settings it has an extra contribution from the Galvani potential ϕ_G , making it $\phi_0 + \phi_G$.

To correct the solvation free energies from both models, we need to add an extra $q\phi_G$ term, accounting for the transfer from one phase to another. The question is, which ϕ_G to use? Simulations using various models have shown that it is extremely sensitive to the water representation. It has a value of 3000 mV when measured using ab initio

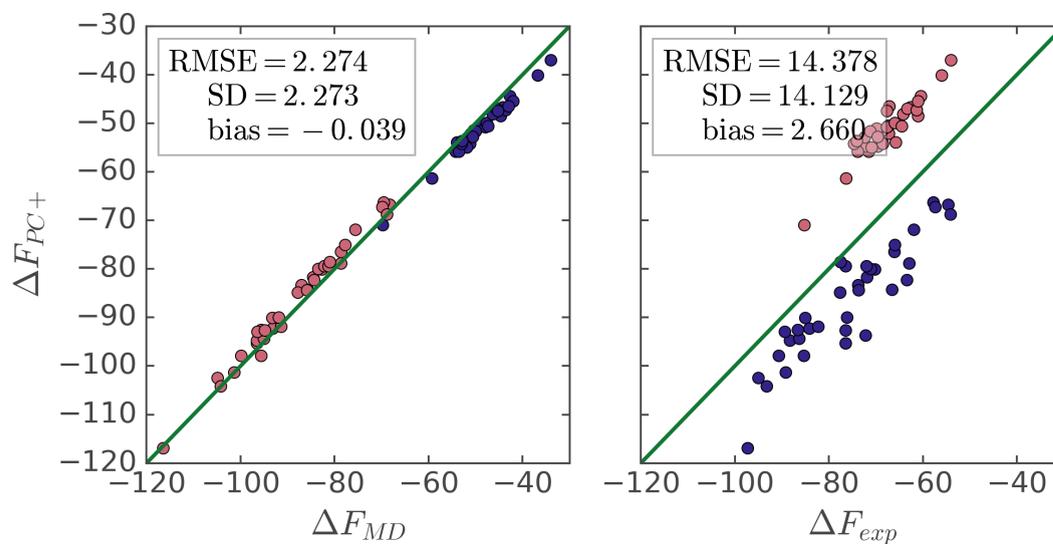


FIGURE 6.8: Comparison of 3D-RISM/PC+ ionic hydration free energies with molecular dynamics results (left) and experimental values (right). Galvani potential is not taken into account. All values are in kcal/mol.

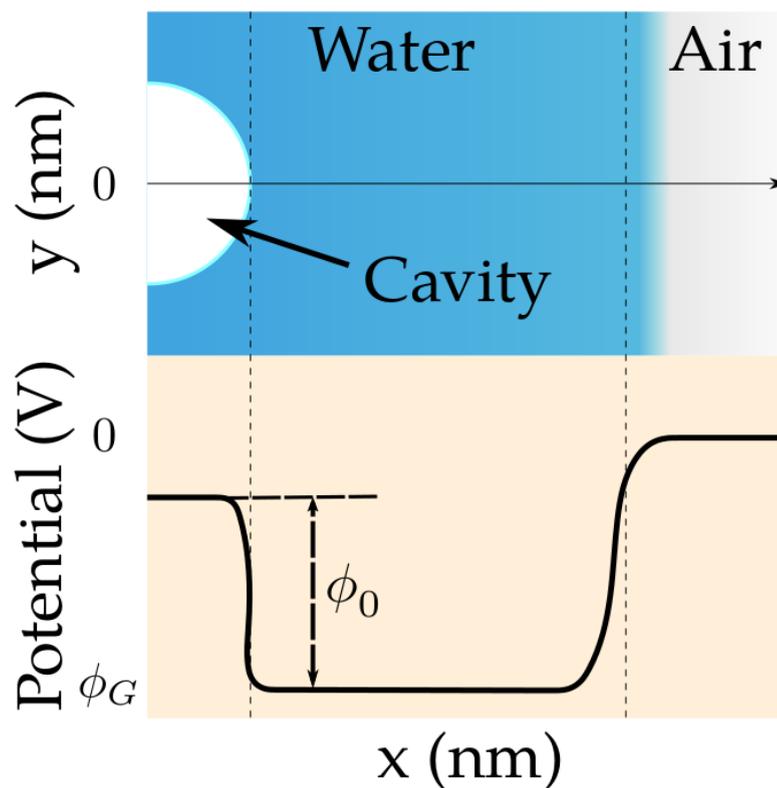


FIGURE 6.9: Schematic demonstration of interfacial potential jumps in solution.

approaches, while atomistic simulations give estimates close to 600 mV. The experimental data are also quite conflicting⁶³.

To understand the disparities in ϕ_G estimates it is important to consider different contributions towards the interfacial potential. It has been shown that electrostatic potential across any interface is given by^{165,166}

$$\begin{aligned}\phi_{vl} &= -\frac{1}{\epsilon_0} \int_{z_v}^{z_l} dz' z' \langle \rho_q(z') \rangle \\ &\approx -\frac{1}{\epsilon_0} \int_{z_v}^{z_l} \langle P_z(z') \rangle dz' + \frac{1}{\epsilon_0} [\langle Q_{zz}(z_l) \rangle - \langle Q_{zz}(z_v) \rangle] \\ &= \phi_D + \phi_Q,\end{aligned}\tag{6.3}$$

where $\phi_{vl} = \phi_l - \phi_v$ is the electrostatic potential difference between a liquid and vacuum, z is the direction perpendicular to the interface towards vacuum, z_l and z_v are positions sufficiently deep into the liquid and vacuum (we assume that liquid vapour contribution is negligible), ρ_q is the charge density, and ϵ_0 is the vacuum permittivity. In the above, the first line corresponds to the solution of the one-dimensional Poisson equation, while the second line is obtained by the Taylor expansion of the charge density in terms of molecular multipoles. The average polarization at z' is given by

$$\langle P_z(z') \rangle = \left\langle \sum_m \delta(z' - z_m) \left(\sum_i q_{im} z_{im} \right) \right\rangle,\tag{6.4}$$

with indices m and i indicating molecules and sites respectively, z_m being the z coordinate of molecular centre, q_{im} is the charge of the site i of molecule m , and z_{im} is the z component of distance \mathbf{r}_{im} from the molecular centre to site i . Similarly, the quadrupole contribution is

$$\langle Q_{zz}(z_l) \rangle = \left\langle \sum_m \delta(z' - z_m) \left(\frac{1}{2} \sum_i q_{im} z_{im}^2 \right) \right\rangle.\tag{6.5}$$

For clarity, we combined dipolar and quadrupolar contributions to the potential into ϕ_D and ϕ_Q respectively. Note that from equation 6.3 it follows that ϕ_Q does not depend on the structure of the interface and instead is only determined by the values of quadrupole contributions inside the liquid (since it is zero in the vacuum).

As was demonstrated by Rensing et al.¹⁶⁵, the ϕ_D is reasonably similar across models, and also is quite small for the water-vacuum interface, while ϕ_Q is a larger quantity and is

the one responsible for disparities across the models. However, it gets canceled out from the sum of the cavity and Galvani potentials since it does not depend on the structure of the interface

$$\phi_0 + \phi_G = -\phi_D^0 - \phi_Q^0 + \phi_D^{vl} + \phi_Q^{vl} = -\phi_D^0 + \phi_D^{vl}, \quad (6.6)$$

where superscripts 0 and vl denote cavity-liquid or liquid vacuum interfaces respectively. Thus, to compare solvent potentials as well as ionic solvation free energies across different models, one has to compare solvation free energies with Galvani potential contribution taken into account. Moreover, the Galvani potential has to be evaluated within the model so that quadrupole contribution to the solute-solvent potential is cancelled by the corresponding contribution to ϕ_G ⁶³.

While evaluation of the Galvani potential in MD does not present a lot of difficulties (one can simply use the Poisson equation defined in the first line of equation 6.3), it is hard to obtain it from 3D-RISM. In principle, one can construct a planar air-liquid interface in 3D-RISM; however, as we have seen in section 5.3, such an interface will have a density distribution quite different from the one observed in the experiment and will have quite a large dipole across it. An alternative approach was suggested by Reif and Hunenberger for molecular simulations⁶³. A potential inside a small cavity will be mostly determined by ϕ_Q , since $\phi_D \approx 0$. Then the quadrupole contribution can be evaluated directly via

$$\phi_Q \approx \phi_0 = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^N \int \frac{\rho_i(\mathbf{r})q_i}{r} d\mathbf{r}, \quad (6.7)$$

where $r = |\mathbf{r} - \mathbf{r}_0|$, with \mathbf{r}_0 being the center of cavity.

Since the MD calculations were done with SPC/E water, we used the results of Beck who found its Galvani potential to be -14.9 kcal/mol/e (-650 mV)⁶⁹. Within 3D-RISM, setting the size of hard sphere cavity to $r = 0.5$ Å (same as grid size), we found the cSPC/E water Galvani potential to be -13.43 kcal/mol/e (-606 mV).

The figure 6.10 demonstrates the comparison of real free energies (with Galvani contributions included) from PC+, MD, and experiment. Not only accounting for Galvani potential improves the agreement between RISM and MD, bringing it to the level of neutral particles, but it also dramatically improves the agreement of PC+ predictions with experiment.

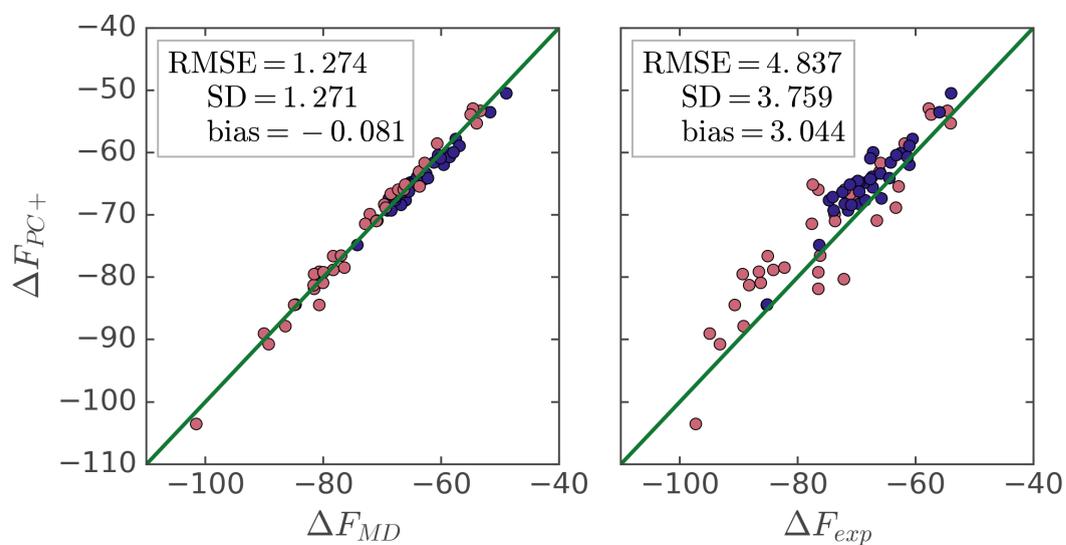


FIGURE 6.10: The same comparison as in figure 6.8, but with model Galvani potentials taken into account. All values are in kcal/mol.

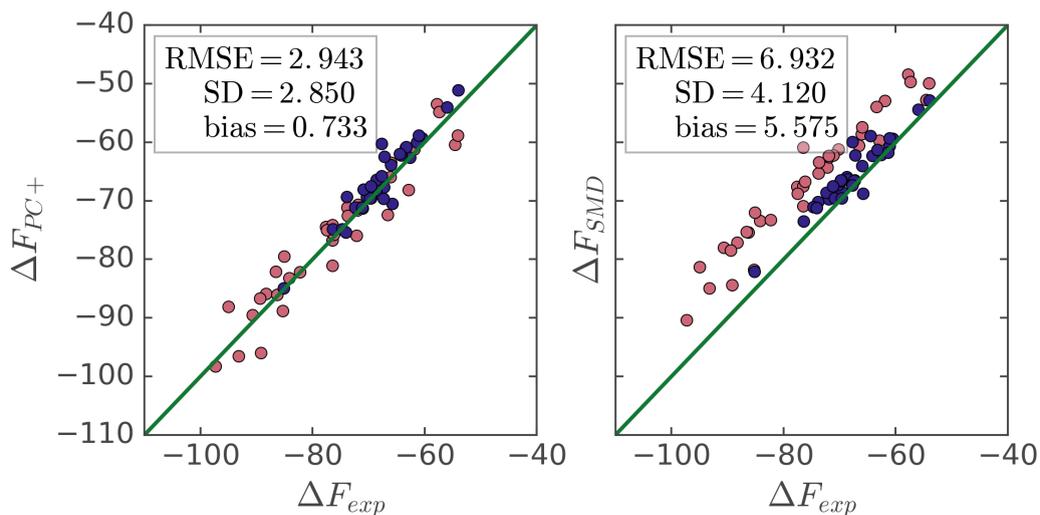


FIGURE 6.11: Left figure: 3D-RISM/PC+ ionic hydration free energies compared to experiment. The force field is OPLS/CM5. Right figure: SMD ionic hydration free energies; electronic structures are computed with MG3S/M06-2X level of theory. The values are in kcal/mol.

While the agreement between the PC⁺ and experimental ionic solvation free energies was relatively good, the error was still larger than in the case of neutral compounds. At least part of the error originates from the approximate solute charges that we obtained using the simple AM1-BCC scheme. While it does not introduce much of a problem in the case of neutral compounds (section 6.1), for charged molecules an accurate charge distribution is crucial since their solvation free energy is dominated by the charging free energy term. To accurately estimate the partial charges on solute atoms we used CM5 charges, obtained via electronic structure calculations. These partial charges in combination with OPLS non-bonded parameters, decreased the error of the PC⁺ model quite significantly (figure 6.11).

In the figure 6.11 we have also shown the prediction of the same ionic hydration free energies obtained by SMD. The SMD calculations were run with the same basis set and theory level as calculations we performed to evaluate CM5 charges. Thus, the charge distribution of solutes in both PC⁺ and SMD calculations were practically identical. The observed difference between the two models is then likely related to the asymmetry of water electrostatic response, which is missed by continuum methods such as SMD but is captured by 3D-RISM.

Chapter 7

Beyond pure water

In this chapter, we show a few applications of 3D-RISM/PC+ to systems other than pure water. The major difficulty here is not theoretical but rather finding an appropriate model of the solvent. In the case of salt solution, the problem is solved by essentially a brute-force approach, while for non-aqueous solvents we develop a coarse-grained approximation. The chapter extends Refs. [41](#) and [42](#).

7.1 Setschenow constant

Dissolving salts in water significantly affects its structure and polarity. Consequently, it changes solvation free energies, solubilities, activities, and other thermodynamic parameters of solutes [167–170](#). Understanding and modeling these effects is quite important since both the natural water reservoirs as well as water in biological tissues will have a considerable amount of ions. Thus, to accurately determine the environmental fate of compounds as well as their distribution in cellular environments one has to take into account effects of dissolved ions [171–174](#). Additionally, techniques such as purification, polymorph control, and yield improvement all utilize salt related effects [175–177](#).

In the context of solvation, the effects of salt on partition coefficient can be quantified using Setschenow's equation [178](#):

$$\log_{10} \left(\frac{K_{1/\text{water}}}{K_{1/\text{salt water}}} \right) = k_S C, \quad (7.1)$$

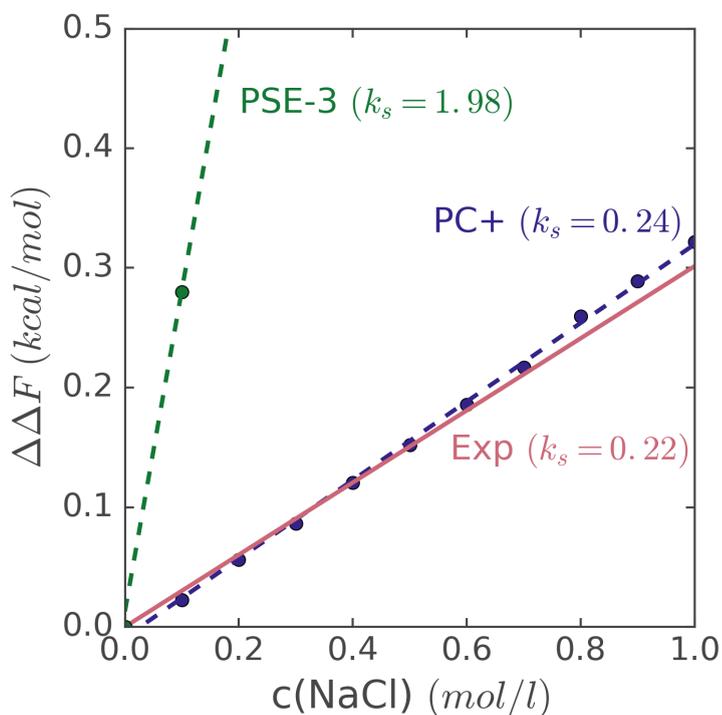


FIGURE 7.1: The change in hexanol hydration free energy depending on NaCl concentration. The experimental trend is estimated from the experimental Setschenow's constant.

where C is the molar concentration of salt in solution, k_S is the Setschenow's (or salting out) constant, and K is a partition coefficient of a compound between two phases, given by

$$K_{1/water} = [\text{solute}]_{\text{water}} / [\text{solute}]_1, \quad (7.2)$$

in which square brackets denote equilibrium concentrations. Setting phase 1 to a dilute gas, we can express the above equation in terms of corresponding solvation free energies¹² to get

$$\Delta G_C = \Delta G_0 + k_S R T C \ln(10), \quad (7.3)$$

where ΔG stands for solvation free energy, subscripts C and 0 denote salt concentrations in water, R is the universal gas constant, and T is temperature.

k_S is largely determined by molecular size. In sodium chloride solutions, the surface tension of water increases proportionally to the salt concentration¹⁷⁹, and thus, provides a positive contribution to the solvation free energy of a molecule. However, this is not the only factor contributing to the Setschenow's constant¹⁸⁰. Polar regions of molecules interact strongly with salts, and this can provide negative contributions to ΔG ¹⁸¹. Thus,

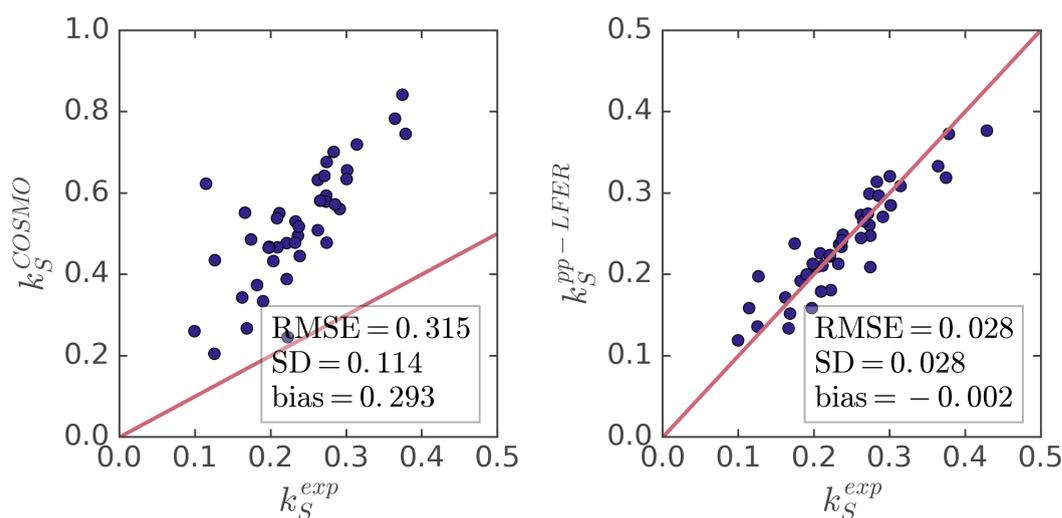


FIGURE 7.2: Accuracies of Setschenow's constants predictions by COSMO-RS (left) and pp-LFER model (right). Both experimental and computed values are taken from Ref. 178. All values are in l mol^{-1} .

to accurately predict k_S , one has to take into account the change of solvent surface tension, favourable interactions between solute and salts, and correlations between anions and cations.

The figure 7.1 demonstrates how much the solvation free energy of a compound is affected by the dissolved NaCl. It also demonstrates predictions by PSE-3 and PC+, which we will discuss in greater detail in the following section.

7.2 Predicting solvation free energy in salt solutions

Despite the importance of salt effects, the majority of computational approaches struggle to incorporate them. The difficulties stem from the fact that effects of salts on solvation free energy are largely non-electrostatic and arise mainly due to the changes in ΔF^{LJ} .

Figure 7.2 compares experimental values of Setschenow's constants to predictions from two computational models. The experimental measurements were performed by Endo et al., for a set of 42 environmentally relevant compounds. Estimates of k_S made with a conductor like screening model for real solvents¹⁸² (COSMO-RS) are relatively poor. The model's inaccuracy is likely caused by its assumption of piecewise surface interactions between surface elements. Better predictions are made by the polyparameter linear free

TABLE 7.1: Lennard-Jones parameters of NaCl models used in this study. The values of σ are in Angstroms and ϵ in kcal/mol.

Abbreviation	σ_{Na}	ϵ_{Na}	σ_{Cl}	ϵ_{Cl}	Ref.
da	2.584	0.100	4.401	0.100	184
jc	2.160	0.353	4.830	0.013	185
de	1.890	0.199	4.410	0.199	186
ho _a	2.130	1.540	4.400	0.100	187
ho _b	2.230	0.650	4.400	0.100	187

energy relationship (pp-LFER) approach, developed by Abraham and co-workers^{21,183}. The model typically uses an empirical equation of the type:

$$\log_{10} K = c + eE + sS + aA + bB + vV, \quad (7.4)$$

where K is the partition coefficient between two phases, E is solute excess molar refraction, S is polarizability, A is solute H-bond acidity, B is solute H-bond basicity, V is the solute molar volume, and lowercase letters are adjustable parameters that depend on the phases between which solutes are distributed. If one applies the above equation to the distribution of molecules between pure water and 1 M solution of sodium chloride, $\log_{10} K = k_S$. While the accuracy of the model is quite remarkable, the coefficients c , e , s , a , b , v were determined using this dataset. Thus, to at least some extent this agreement reflects the success of the linear regression.

To predict the Setschenow's constants using 3D-RISM we first need to obtain susceptibility functions χ for the bulk salt solutions. We decided to calculate them using the 1D-RISM approach with the HNC closure. As an input, the 1D-RISM calculations require site interaction potential energies. Pretty much all NaCl force fields use the same charges for ions: plus and minus one. However, there are quite a few options for Lennard-Jones parameters^{188–191}.

For simplicity, we limited our attention to only five sodium chloride models that were compatible with SPC/E water¹⁹² and developed with Lorentz-Berthelot combination rules⁵⁷ in mind (table 7.1). The Lennard-Jones parameters in different models were fit to different experimental observables. Dang's NaCl force-field parameters (da) were developed by fitting interaction energy, the first peak of the radial distribution function, and coordination number¹⁸⁴. Joung and Cheatham's model (jc) is based on fitting the experimental hydration free energies of ions, as well as lattice constants and energies¹⁸⁵.

TABLE 7.2: Accuracies of different models for predicting Setschenow’s constant. The units are l/mol.

Model	RMSE	SDE	bias	r^2
OPLS/CM5				
da	0.028	0.028	0.005	0.840
ho _b	0.036	0.029	-0.021	0.823
jc	0.051	0.034	0.038	0.789
ho _a	0.058	0.029	-0.050	0.818
de	0.085	0.037	-0.076	0.713
GAFF/AM1-BCC				
da	0.032	0.032	-0.004	0.800
jc	0.043	0.034	0.025	0.772
ho _b	0.050	0.035	-0.035	0.750
ho _a	0.078	0.042	-0.066	0.650
de	0.119	0.062	-0.102	0.300
Other models				
pp-LFER	0.028	0.028	-0.002	0.844
SEA ^a	0.051	0.035	0.037	0.706
MD/TIP3P ^a	0.120	0.029	0.116	0.848
COSMO-RS	0.315	0.114	0.293	0.670

^a The accuracy of the model was evaluated on a different dataset.

Deublein and co-workers (de) adjusted NaCl Lennard-Jones parameters to reproduce experimental density at a range of concentrations¹⁸⁶. Finally, Horinek et al. developed multiple force fields, by taking Dang’s Cl⁻ ion parameters and adjusting Na⁺ parameters to match the solvation free energy of the ion pair¹⁸⁷. Since this approach does not lead to a unique pair of ϵ and σ , the authors proposed models based on small ϵ (we could not converge this model in 1D-RISM), large ϵ (ho_a), and medium ϵ values (ho_b).

Table 7.2 compares accuracies of Setschenow’s constant predictions by different salt models. The results by the polyparameter linear free energy relationship (pp-LFER), semi-explicit assembly (SEA), molecular dynamics simulations with TIP3P water and Joung-Cheetham ions (MD/TIP3P), and COSMO-RS¹⁸² are taken from previous studies^{178,193}. In the literature, one can also find a few more chemoinformatics methods for Setschenow’s constant prediction based on other descriptors or various machine learning methods^{194–197}, but the accuracy of these models did not significantly exceed the accuracy of the pp-LFER approach. We found that predictions made with Dang’s salt force field (da), combined with OPLS/CM5 force field for solutes, had the best agreement with experimental data among the studied 3D-RISM models, both in terms of its accuracy and the correlation.

RISM calculations based on other salt models had similarly low random error, but larger biases.

In 3D-RISM/PC+ calculations we used two different sets of force fields to describe solutes: OPLS/CM5 and GAFF/AM1-BCC. Only the salt model by Joung-Cheetham (jc) showed better results when paired with GAFF/AM1-BCC solutes. In all other cases the use of OPLS/CM5 parameters improved predictions by various extents.

The results from PC+ with Dang's salt model are similar to the pp-LFER model that was fit on Endo's dataset using six adjustable descriptors. It outperforms both SEA and COSMO-RS models, which are both partially based on the idea that summing surface elements of a solute is a useful strategy for predicting solvation free energies. 3D-RISM, on the other hand, takes into account correlations between densities of solvent at the surface of solute, which most likely contributes to its better accuracy. Notice that accuracies of both SEA and MD/TIP3P models are evaluated on a different dataset for which 3D-RISM with Dang's NaCl force field has RMSE = 0.0381/mol, SDE = 0.0291/mol, bias = -0.0251/mol, and $r^2 = 0.798$. The slight decrease of accuracy is likely explained by the use of less reliable experimental data.

It is useful to note that we defined PC and PC+ models using equations 5.26 and 5.27. Then, for solvents consisting of multiple different species $\Delta G_{PC} \neq kT/2 \langle \rho | \ln \mathbf{g} \rangle + 1/2 \langle \rho | \mathbf{u} \rangle$. Instead using equation 5.17 we have

$$\begin{aligned} \Delta \Omega_{PC} = & \frac{kT}{2} \langle \rho | \ln \mathbf{g} + \beta \mathbf{u} \rangle - kT \sum_{i=1}^N \rho_i G_i \left(1 - \frac{1}{2} \sum_{j=1}^N \rho_j \hat{C}_{ij}(0) \right) \\ & - kT \Delta V \sum_{i=1}^N \rho_i \left(1 - \frac{1}{2} \sum_{j=1}^N \rho_j \hat{C}_{ij}(0) \right), \end{aligned} \quad (7.5)$$

which cannot be simplified a lot further due to $G_i \neq -\Delta V$. We can define the PC+ prime model as:

$$\Delta \Omega_{PC+'} = \frac{kT}{2} \langle \rho | \ln \mathbf{g} + \beta \mathbf{u} \rangle - kT \langle \Delta \rho | \mathbf{1} \rangle. \quad (7.6)$$

However, this definition of PC+ leads to worse results (see table B.2).

Molecular dynamics based predictions of Setschenow's constant, despite achieving impressive correlation with experimental data ($r^2 = 0.848$), have a large positive bias. We believe that the origin of this bias is likely related to the chosen salt model (jc) and force

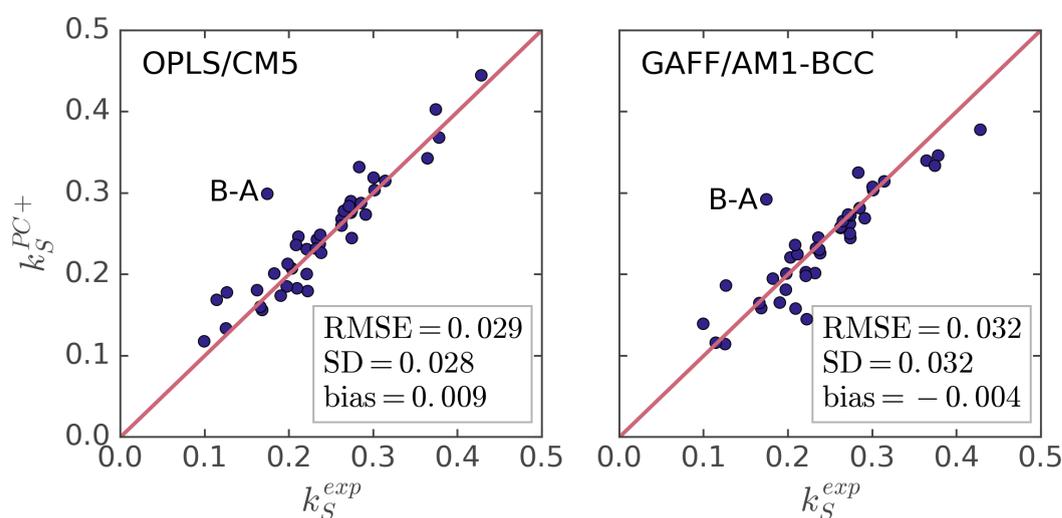


FIGURE 7.3: Setschenow’s constants by 3D-RISM/PC+ with Dang salt model compared to experimental measurements by Endo et al. Results on the left and right figures are obtained with OPLS/CM5 and GAFF/AM1-BCC force fields respectively. B-A stands for bisphenol A. The values are in l/mol.

fields: GAFF with TIP3P water. It is likely that a combination of Dang salt model and SPC/E water would reduce the bias in the prediction and make molecular dynamics simulations one of the most accurate ways of predicting Setschenow’s constant, although, quite time consuming.

Figure 7.3 has a comparison between Setschenow’s constants predicted by 3D-RISM with Dang’s NaCl model and those from experimental measurements. For both calculations made with OPLS/CM5 and GAFF/AM1-BCC parameters, one major outlier is bisphenol A (B-A). In both cases 3D-RISM calculations overestimate its k_S by 0.13 l/mol: more than four times greater than the average prediction error for Dang’s model. This molecule was also an outlier in 3D-RISM calculations with other salt models. While potentially this might be the result of measurement error, we believe that the reason for this lies in the fact that bisphenol A binds relatively strongly to sodium ions via π -cation interactions. We performed electronic density functional theory calculations to test this hypothesis. The optimization was done using the same level of theory and software as for the initial molecule preparation. Optimised geometries for bisphenol A with and without sodium atoms, shown in figure 7.4, indicate significant structural rearrangement as well as considerable bonding between Na^+ and both phenol rings. These type of interactions are difficult to characterize using conventional force fields¹⁹⁸, and would require a quantum mechanics

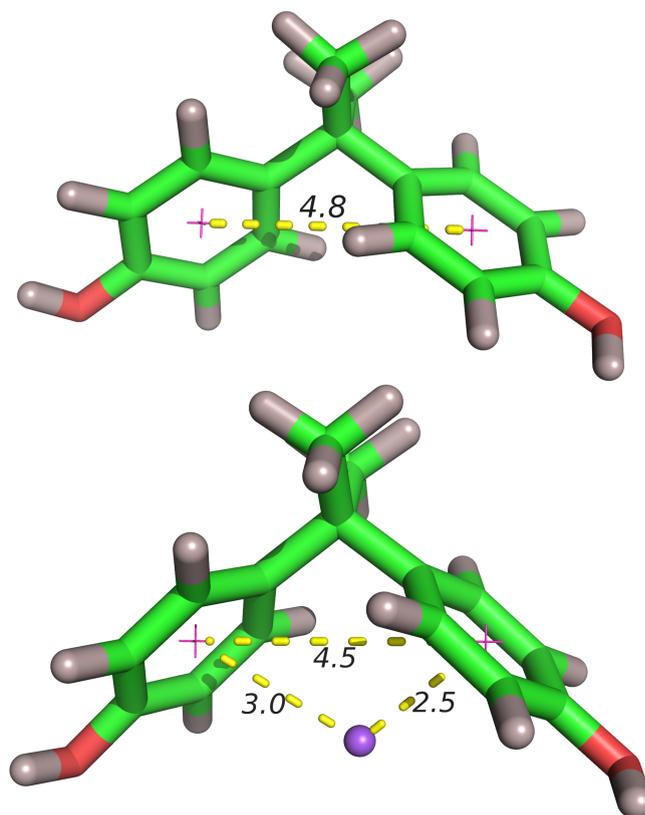


FIGURE 7.4: The optimised geometries of bisphenol A with and without Na^+ ion. The distances, shown in Angstroms, are measured between the centres of benzene rings and the ion. SMD model was used to take into account solvent effects.

approach to dispersion interactions. Additionally, conventional 3D-RISM operates with rigid solutes and does not capture salt-induced changes in solute conformation.

7.3 Non-aqueous solvents and the corresponding state principle

3D-RISM is generally poorly suited for a description of solvents with a large number of sites. As solvent molecules get larger, it gets harder to generate susceptibility files, as well as the speed and convergence of 3D-RISM calculations become slower. Additionally, since intramolecular correlations are also approximated in 3D-RISM, as the number of sites increases, the description of bonding within solute becomes poorer, which also negatively impacts 3D-RISM performance.

A solution to this problem is to coarse-grain organic solvents. In principle, there a number of conventional schemes available, but they all tend to be quite slow. However, when

reviewing the literature we discovered a scheme based on the corresponding state principle that lets one predict coarse-grained interaction parameters in a straightforward manner.

According to the corresponding states principle, reduced critical temperature

$$T_c^* = \frac{kT_c}{\epsilon} \quad (7.7)$$

and reduced critical density

$$\rho_c^* = \rho_c \sigma^3 \quad (7.8)$$

are constants for all classical fluids with orientation-independent interaction potentials^{199,200}. Here k is the Boltzmann constant, T_c and ρ_c are critical temperature and density, σ is effective particle diameter and ϵ is a constant that determines the strength of intramolecular interactions. This principle can be further extended to non-spherical molecules by assuming T_c^* and ρ_c^* are functions of molecular shape and electrostatic properties¹⁹⁹.

It follows that knowing T_c^* and ρ_c^* for a single reference fluid, one can easily obtain intermolecular interaction parameters ϵ and σ for many others from their critical properties. This idea has been used by a number of authors to construct coarse-grained models of real fluids and estimate their properties at a wide range of conditions^{201–211}. Most of them came to the conclusion that with the exception of a few simple fluids such as argon, nitrogen, or methane, the majority of the real fluids cannot be adequately described by just two simple parameters and require either additional fittings or more complicated interaction potentials.

However, a precise description of solvent behaviour and phase diagram is frequently not necessary for an accurate estimation of solvation free energy, as can be seen by a number of successful implicit solvation models^{154,182}. As we demonstrated in section 2.5, solvation free energy is independent of solvent reorganization energy, so it most likely can be estimated using a rather simple coarse-grained model. Thus, to construct an approximation for a number of organic solutes we simply used the reduced critical parameters of Lennard-Jones fluid: $T_{c,LJ}^* = 1.313$, $\rho_{c,LJ}^* = 0.304$ that were obtained by Okumura et al. using molecular dynamics²¹².

Since we were mainly interested in predicting solvation free energies, we decided to focus on popular solvents for which a large amount of data are available. The Minnesota solvation

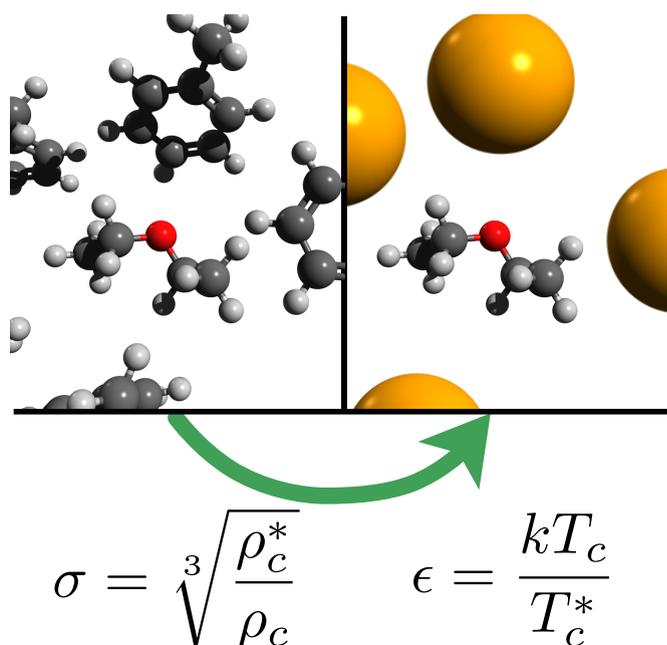


FIGURE 7.5: The basic idea behind coarse graining used for 3D-RISM calculations. The real liquids are approximated as spheres interacting via Lennard-Jones potentials with parameters deduced using equations shown on the figure above.

database contains a large collection of measurements made at standard conditions. From it we picked 17 non-associating solvents, listed in the table 7.3. The table lists critical properties liquids, taken from Ref. 213, as well as parameters of Lennard-Jones spheres approximating these liquids (obtained via equations 7.8 and 7.7). Note that the solvent xylene is a mixture of isomeric ortho-, meta- and para-forms of xylene.

Of course, a spherical Lennard-Jones fluid is a poor reference system for most of these solvents. For this reason, isooctane, heptane, and decane were also modelled as chains of Lennard-Jones spheres composed of m segments, each separated by a bond of length σ . The choice of m was motivated by an equation employed in Statistical Associating Fluid Theory (SAFT) and in some molecular dynamics studies^{210,214,215}

$$m = 1 + \frac{n(C) - 1}{3}, \quad (7.9)$$

where $n(C)$ is the number of carbons in the linear alkane. $m = 3$ for heptane and $m = 4$ for decane follow directly from the equation. We also assumed that $m = 2$ would be a reasonable choice for isooctane. The σ and ϵ parameters for chain beads were similarly obtained using equations 7.7 and 7.8, but using critical points for the 2-mer

TABLE 7.3: Critical properties and Lennard-Jones parameters.

Name	T_c [K]	ρ_c [nm ⁻³]	ϵ [kcal/mol]	σ [nm]
1,2-dichloroethane	561.60	2.74	0.85	0.48
acetonitrile	545.00	4.11	0.82	0.42
benzene	562.05	2.35	0.85	0.51
bromobenzene	670.15	1.86	1.01	0.55
carbon disulfide	552.00	3.76	0.84	0.43
carbon tetrachloride	556.35	2.18	0.84	0.52
chloroform	536.40	2.52	0.81	0.49
cyclohexane	553.80	1.96	0.84	0.54
diethyl ether	466.70	2.15	0.71	0.52
dimethyl sulfoxide	729.00	2.65	1.10	0.49
ethyl acetate	523.30	2.11	0.79	0.52
isooctane	543.80	1.29	0.82	0.62
isooctane (2-mer)			0.61	0.49
n-decane	617.70	1.07	0.93	0.66
n-decane (4-mer)			0.54	0.39
n-heptane	540.20	1.41	0.82	0.60
n-heptane (3-mer)			0.52	0.40
octanol	652.50	1.21	0.99	0.63
toluene	591.75	1.91	0.90	0.54
xylenes	624.57	1.59	0.95	0.58

($T_{c,LJC2}^* = 1.78$, $\rho_{c,LJC2}^* = 0.149$, Ref. 216), 3-mer ($T_{c,LJC3}^* = 2.063$, $\rho_{c,LJC3}^* = 0.088$, Ref. 217) and 4-mer ($T_{c,LJC4}^* = 2.26$, $\rho_{c,LJC4}^* = 0.0625$, Ref. 218) Lennard-Jones chain fluids.

7.4 Solvation of model solutes in Lennard-Jones fluids

Now that we have defined the solvents, we can start predicting solvation free energies of various solutes in them. However, before doing it, we first test the applicability of pressure corrections to these liquids by comparing results of HNC and PC+ approximations to molecular dynamics.

The figure 7.6 compares predictions of hard sphere solvation free energies in a Lennard-Jones fluid at reduced temperature $T^* = kT/\epsilon = 0.701$ and reduced density $\rho^* = \rho\sigma^3 = 0.005$. The 3D-RISM results are evaluated against estimations from the Lum-Chandler-Weeks (LCW) model that is known to be quite accurate and agrees well with MD predictions. The trends in errors of 3D-RISM models are similar to those observed for water (figures 5.1 and 5.3), however, now PC+ model underestimates the insertion free energy of the solute, while HNC gives relatively good predictions up to the point where

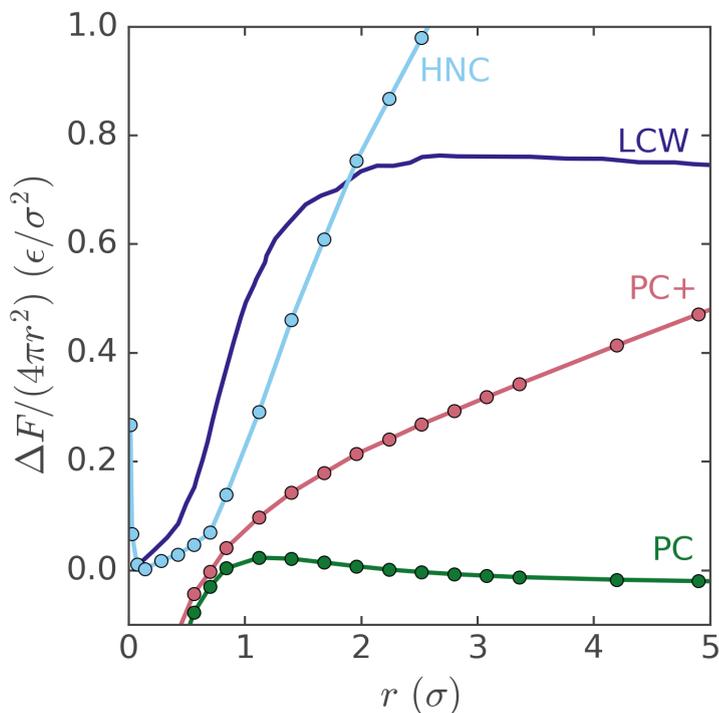


FIGURE 7.6: Solvation free energy per unit area for solutes of radii r in the Lennard-Jones liquid with $T^* = 0.701$ and $\rho^* = 0.843$. LCW results are taken from Ref. 137.

the interface forms. It seems that the compressibility-based estimate of insertion free energy that is employed in HNC works well for Lennard-Jones fluids.

We also wanted to check how well can 3D-RISM predict solvation free energies of Lennard-Jones solutes that offer a somewhat more accurate representation of typical molecules. We performed a number of molecular dynamics free energy simulations in model dichloroethane and tetrachloromethane. In total we used 90 different solutes that had all possible combinations of $\sigma = 0.6, 1.4, 2.2, \dots, 7.0 \text{ \AA}$, and $\epsilon = 2^{-7}, 2^{-6}, 2^{-5}, \dots, 2^2 \text{ kcal/mol}$. The simulations were performed at NVT conditions to make sure that the density of Lennard-Jones spheres matched the density of the real liquids at 298 K. Note that all molecules in the system are uncharged; we are primarily comparing the accuracy of insertion free energy estimates.

Comparison of MD solvation free energies with both PC+ and HNC predictions is shown in figure 7.7. When compared to water, the agreement of PC+ with MD was poorer, while agreement of HNC with molecular dynamics was significantly better. Interestingly, while absolute values were predicted slightly more accurately by HNC, the trends were captured more faithfully by PC+.

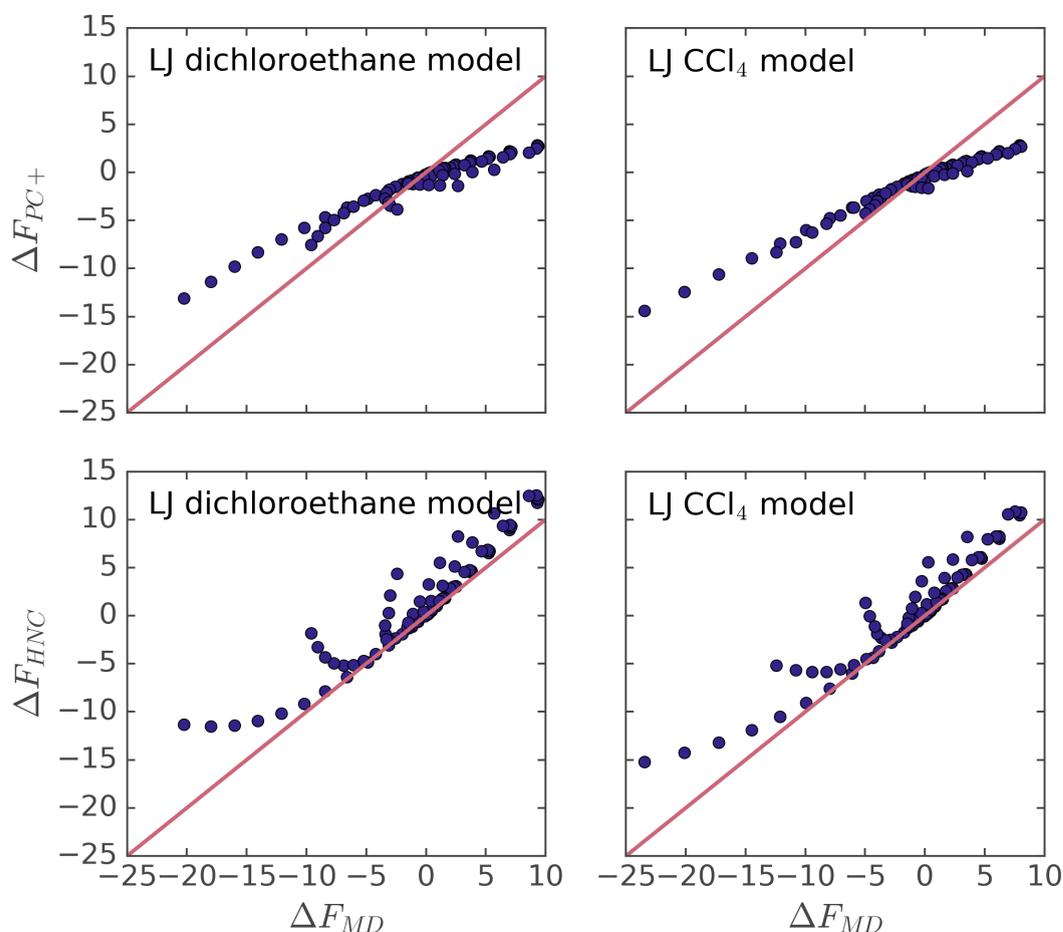


FIGURE 7.7: Calculated Lennard-Jones solutes solvation free energies against MD data. Dichloroethane and CCl_4 models are defined in table 7.3. All values are in kcal/mol.

These findings demonstrate that 3D-RISM, when applied to Lennard-Jones fluids, cannot predict interface formation. Similarly to what happens in the case of water, the results between two models agree only up to a certain solute size. Moreover, the PC+ correction seems to work significantly worse and does not approximate the insertion free energy too well.

7.5 Comparison with experimental values

As we discussed previously, we also evaluated the accuracy of our solvent approximation using experimental data from Minnesota solvation database^{154,164}. We selected data for 17 popular solvents, presented in table 7.1. Chosen solvents have 1247 associated experimental measurements for 482 unique solute molecules.

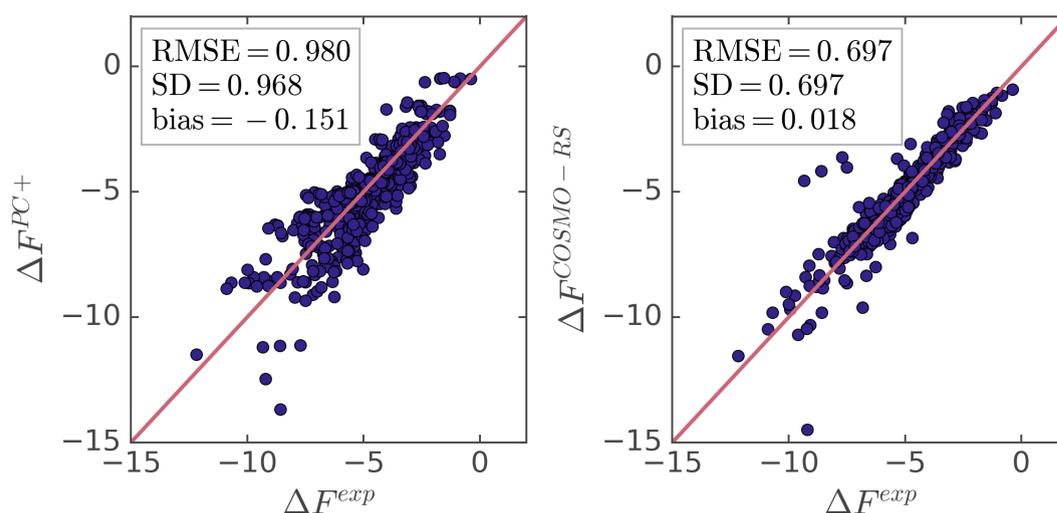


FIGURE 7.8: Computed solvation free energies of a number of compounds in apolar solvents (as defined in table 7.4) against experimental data. COSMO-RS results are taken from Ref. 219. All values are in kcal/mol.

The solvent susceptibility functions were generated using 1D-RISM, with a bulk density of model solvents set to experimental number densities at 298 K²¹³. The solute geometries were obtained from the Minnesota database; non-bonding parameters were taken from the OPLS-2005 force field. All partial charges were set to zero.

The figure 7.8 demonstrates the accuracy of PC+ predictions for apolar solvents, with table 7.2 showing a more detailed breakdown. For comparison, we also plotted COSMO-RS predictions made for the same set of solvents. You can see that while PC+ results are poorer than those made by more advanced model, they are still within acceptable 1 kcal/mol range.

The same approach is significantly less successful for polar solvents, which often interact with the solutes via strong specific interactions. As figure 7.9 demonstrates, the accuracy of PC+ is almost two times worse when compared to the predictions in polar solutes. This decrease of accuracy is not surprising since our coarse-grained models of solvents lack electrostatic charges. Still, the existence of any correlations between PC+ and experimental values suggests that this approximation still allows us to roughly estimate solvation free energies in these solvents.

In view of the results in section 7.4, the overall accuracy and correlation of PC+ predictions with experimental data seems surprising. The HNC model, which showed better agreement with MD values of solvation free energy for Lennard-Jones fluids, actually correlated

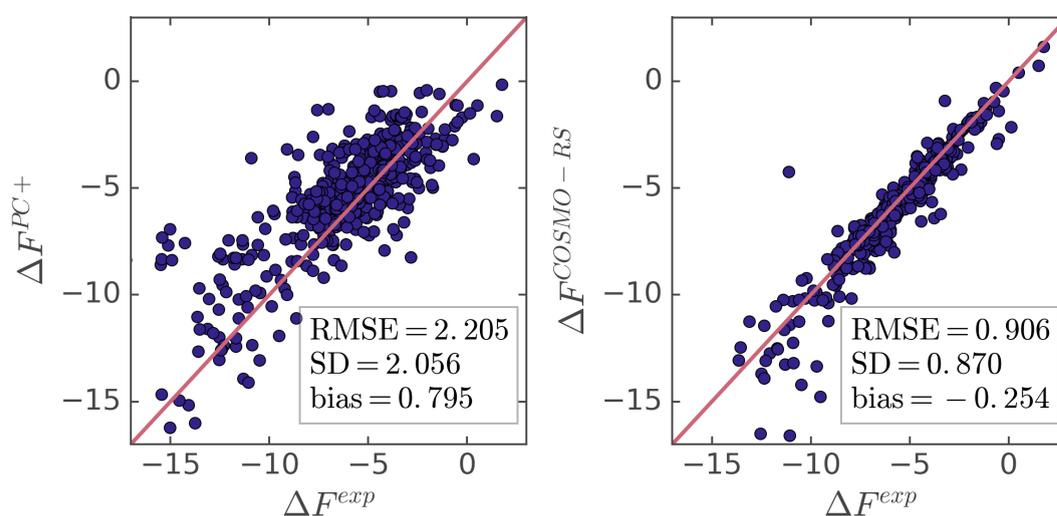


FIGURE 7.9: Computed solvation free energies of a number of compounds in polar solvents (as defined in table 7.4) against experimental data. COSMO-RS results are taken from Ref. 219. All values are in kcal/mol.

Solvent	N	RMSE	SDE	bias
Apolar				
1,2-dichloroethane	39	1.16	1.07	0.47
benzene	71	1.28	1.28	0.04
bromobenzene	27	1.17	1.15	-0.23
carbon disulfide	15	0.94	0.89	-0.30
carbon tetrachloride	79	0.85	0.84	-0.11
cyclohexane	103	1.01	0.75	-0.67
isooctane	32	0.98	0.68	-0.70
isooctane (2-mer)	32	0.63	0.60	-0.21
n-decane	39	1.70	1.23	-1.17
n-decane (4-mer)	39	0.68	0.56	-0.38
n-heptane	67	0.95	0.86	-0.42
n-heptane (3-mer)	67	0.74	0.74	0.05
olive oil	218	1.30	1.06	-0.75
toluene	51	1.00	0.99	0.08
xylenes	48	1.00	0.99	-0.10
Polar				
acetonitrile	7	2.23	2.13	0.67
chloroform	107	1.86	1.37	1.25
diethyl ether	70	2.21	1.65	1.47
dimethyl sulfoxide	7	2.65	2.65	0.01
ethyl acetate	22	3.02	2.18	2.09
octanol	245	2.24	2.22	0.31

TABLE 7.4: Accuracies of solvation free energy predictions by 3D-RISM/PC+ for various solvents. RMSE stands for root mean square error, SDE is standard deviation of error. Energies are in kcal/mol.

worse with experiment (figure B.5). The PC+ solvation free energy estimates made with coarse-grained solvents agree better with experimental values than predictions made with the same solvent and molecular dynamics. In other words, making an approximation in solvation free energy improves the result! It is not entirely clear why exactly this occurs; it is likely that a simplistic, mean-field estimate of solvent parameters made with the corresponding state principle, works best when combined with a simple, linear-response like free energies, given by pressure-corrected models.

Chapter 8

Conclusion

The main goal of this work was to develop a 3D-RISM-based advanced pressure correction model, PC+, and to investigate its scope of application and accuracy. The key findings of the thesis can be summarised as follows:

- From the theoretical point of view, the PC+ model is based on a linear response approximation to solvation free energy, combined with an estimate of cavity creation work. It was shown that the PC+ model could provide accurate predictions of the solvation free energy as long as both of these approximations hold. Chapter 5.
- For a pure aqueous solvent, the approach turns out to be quite useful; it predicts hydration free energies with an accuracy of about 1.3 kcal/mol and 3.0 kcal/mol for small neutral and charged molecules correspondingly. Chapter 6.
- The model can also be applied to aqueous NaCl solutions. It was shown that the model could significantly improve estimates of the Setschenow's constant for molecular compounds compared to commonly used models; with a properly chosen salt representation, pressure-corrected 3D-RISM can achieve a good accuracy in quantitative predictions of the Setschenow's constant. Chapter 7.
- Finally, we demonstrated that PC+ could also be applied to non-aqueous solvents. A major problem of such systems, a large amount of sites and flexibility, was solved by introducing a consistent coarse-grained approximation. This approach in combination with PC+ led to an accuracy of about 1 kcal/mol for a range of non-polar solvents. Chapter 7.

- Notable failures of the model include its relatively poor prediction of the solvation entropies and a failure to accurately estimate the changes in the heat capacity occurring due to the solute insertion. Additionally, the model provides poor performance for polar non-aqueous solvents such as DMSO or methanol. Chapters 6 and 7.

The work warrants further investigations. From the theoretical point of view, it is still not entirely clear why the pressure-correction approach works well for multicomponent mixtures. We also left for future investigations a possibility of defining a self-consistent pressure corrected functional and a detailed analysis of corresponding density distributions obtained via it. Additionally, it is also worthy to test a combination of pressure corrections with advanced free energy functionals such as anisotropic HNC or hydrostatic approximations. Even without focusing on the theory itself, we believe that extra improvements of the model can be achieved by using more sophisticated solute-solvent potentials, as well as testing more optimal models of solvents, designed specifically with 3D-RISM/PC+ in mind.

Appendix A

Methodology

Throughout the thesis, the majority of the calculations were performed using the same software and settings. For this reason, we decided to summarize the general methodology in this appendix, mentioning specifics of each calculation in appropriate parts of the main text.

A.1 1D-RISM

In the thesis, 1D-RISM calculations were primarily used to generate the susceptibility functions for 3D-RISM calculations. The calculation inputs are molecule geometry, its Lennard-Jones parameters, and partial charges, as well as bulk solvent density and dielectric constant, which we obtained from the experimental data.

The majority of the actual calculations were performed with the `rism1d` program^{30,105,220} included in the AmberTools 15 package²²¹. Calculations performed with AmberTools 14 or 16 versions only differed in the additional output, while the susceptibility functions remained the same. The 1D-RISM equations were solved with a tolerance set to 1×10^{-12} and grid spacing to 0.025 Å. Note that whenever solvent had partial charges, and thus an associated dielectric constant, we used the dielectrically consistent formulation of 1D-RISM (DRISM).

Most commonly we used either HNC or PSE-3 closures, which gave practically identical results. For solvents other than pure water it was often impossible to converge susceptibility

functions starting with these closures. Typically, we obtained an initial solution using the KH closure, then tried to converge using the PSE-2 closure, using the KH susceptibility function as the initial guess. Only after obtaining PSE-2 solution did we move to perform PSE-3 or HNC calculations. Quite often, we also had to adjust parameters such as force field, temperature or density to obtain a good initial guess. Weaker interacting systems provided reasonable starting guesses for further calculations.

A.2 3D-RISM

The input for 3D-RISM calculations includes external field specification (typically in the form of force field potential of a solute molecule) as well as bulk solvent susceptibility functions.

Most of the 3D-RISM calculation in the thesis were performed using `rism3d.snglpnt` program from AmberTools 15 package. Similarly to 1D-RISM, part of the calculations were done using different versions; however, it did not affect the results. For the majority of the calculations we used a grid spacing set to 0.5 Å, buffer to 25 Å, and tolerance to 1×10^{-5} . Calculations performed in the early stages of the thesis used finer grids with a 0.3 Å spacing, 30 Å buffer and 1×10^{-10} tolerance. However, we found that solvation free energy was largely unaffected by the change in grid settings and the additional precision obtained with the finer grid was offset by a significantly larger computational time (often about 10 longer CPU times).

To simplify the calculation of pressure corrections as well as solvation free energy workflow we created a small script, hosted on https://github.com/MTS-Strathclyde/PC_plus.

The majority of the 3D-RISM calculations were performed with the PSE-3 closure. The solvation free energies were almost unaffected by the choice of closure, with differences between HNC and PSE-3 often being smaller than the uncertainty due to the grid spacing. At the same time, PSE-3 calculations converged quicker and more reliably, which led us to prefer this particular closure.

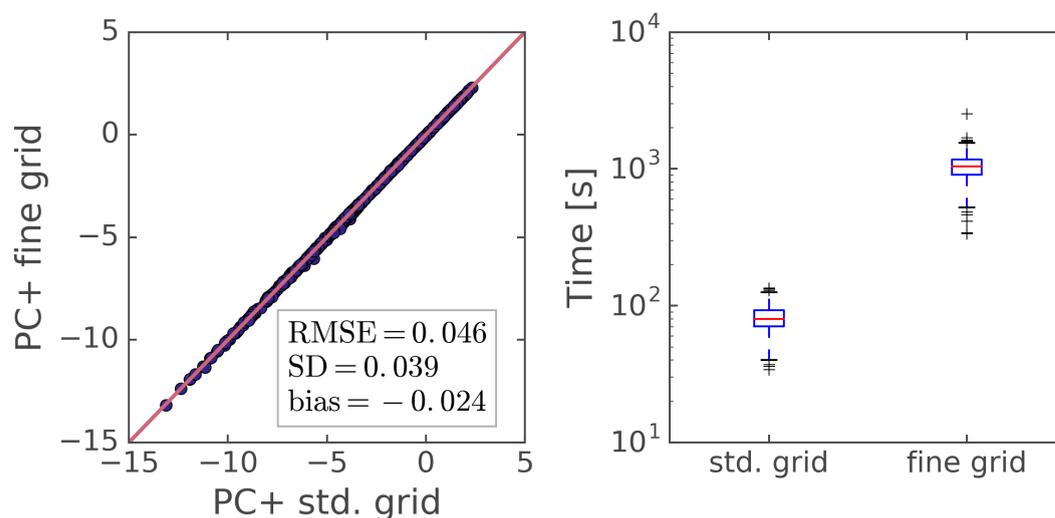


FIGURE A.1: We performed two sets of calculations on the Mobley dataset (described at the beginning of section 6.1) using standard and fine grid settings. The comparison of free energies (kcal/mol) is shown on the left, and in run time on the right figure.

A.3 Molecular dynamics

Molecular dynamics simulations in this thesis were performed using Gromacs 5.04 software²²². We used a cubic box with periodic boundary conditions. All bonds with hydrogens were kept rigid using LINCS algorithm of 12-th order. Dynamics was simulated using the Langevin integrator, with a reference temperature of 298.15 K and a friction constant of 1.0 ps⁻¹.

For short-range interactions, a pair list was generated using a Verlet cut-off scheme. Lennard-Jones interactions were smoothly switched off between 9 and 12 Å. The cut-off artifacts were accounted for using long-range pressure and dispersion corrections as implemented in Gromacs. Electrostatics interactions were treated using particle-mesh Ewald (PME) method²²³ with 12 Å real space cutoff, 1.2 Å Fourier spacing, 6-th order spline interpolation, and tolerance set to 10⁻⁶. For uncharged solvents, we used simple cut-off electrostatics.

The solvation free energy was typically computed using 20 separate calculations at each λ , decoupling first electrostatics and then Lennard-Jones interactions between solute and solvent. Intramolecular interactions within a solute were kept the same at all lambda values. Calculations with modified electrostatics interactions were performed at $\lambda = 0, 0.25, 0.5, 0.75, 1$. The decoupling of Lennard-Jones interactions was done using

calculations at $\lambda = 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0$. This setup has been shown to give good convergence¹⁷.

Before running MD simulations at each λ , we performed 5000 steps of steepest descent optimization. After that, we performed 200 ps equilibration and 1300 ps production runs. Typically, we used either 1 fs or 2 fs time steps. For NPT runs, the pressure was kept constant at 1 bar using Berendsen barostat²²⁴, with time constant set to 1 ps and compressibility to $4.5 \times 10^{-5} \text{ bar}^{-1}$.

After completing the simulations, the intrinsic hydration free energy was evaluated using Multistate Bennett Acceptance Ratio (MBAR)²²⁵. The actual calculation was performed using a python script `alchemical-analysis.py`¹⁷.

A.4 Force fields and geometry

To generate the initial geometry of molecules we typically used the Openbabel software package^{161,162}. For some molecules, further refinement of geometry was performed using the quantum chemical package Gaussian 09, Revision D.01²²⁶. The calculations were performed with the M06-2X functional¹⁵⁵, and MG3S basis set¹⁵⁶. Molecules that were not further optimized using the Gaussian package were simply optimized using assigned force field parameters.

For the majority of the solutes, we assigned either GAFF/AM1-BCC non-bonded parameters or the OPLS_2005 force field, combined with CM5 charges. Due to the amount of data we used only software which did the assignment automatically. The GAFF/AM1-BCC workflow mainly relied on the `antechamber` program found in the AmberTools package. The program was used to both assign parameters as well as to evaluate partial charges. For OPLS force field assignment we used the Maestro package²²⁷.

Similarly to quantum chemical geometry optimization, CM5 charges were evaluated with Gaussian 09 software at the MG3S/M06-2X level of theory. The solvent was represented using the SMD model¹⁵⁴ and charges were extracted from output files using the CM5PAC program²²⁸.

Appendix B

Additional Results

B.1 Partial molar volume and grand canonical ensemble

We want to represent the following partial derivative $(\partial N/\partial N')_{T,V,\mu}$ in terms of partial molar volume $\bar{V} = (\partial V/\partial N')_{T,P,N}$. For this we are going to rely on the following mathematical relationship

$$\left(\frac{\partial X}{\partial Y}\right)_Z = -\left(\frac{\partial Z}{\partial Y}\right)_X \left(\frac{\partial X}{\partial Z}\right)_Y. \quad (\text{B.1})$$

Applying it to the initial derivative we obtain

$$\left(\frac{\partial N}{\partial N'}\right)_{T,V,\mu} = -\left(\frac{\partial \mu}{\partial N'}\right)_{T,V,N} \left(\frac{\partial N}{\partial \mu}\right)_{T,V,N'}. \quad (\text{B.2})$$

To proceed further we use the Gibbs-Duhem relationship $\mu = -sdT + v dP$, where s is molar entropy and $v = 1/\rho$ is molar volume. Applying it to the first derivative in the above equation we get

$$\begin{aligned} \left(\frac{\partial \mu}{\partial N'}\right)_{T,V,N} &= -s \left(\frac{\partial T}{\partial N'}\right)_{T,V,N} + v \left(\frac{\partial P}{\partial N'}\right)_{T,V,N} \\ &= -v \left(\frac{\partial V}{\partial N'}\right)_{T,P,N} \left(\frac{\partial P}{\partial V}\right)_{T,N,N'}. \end{aligned} \quad (\text{B.3})$$

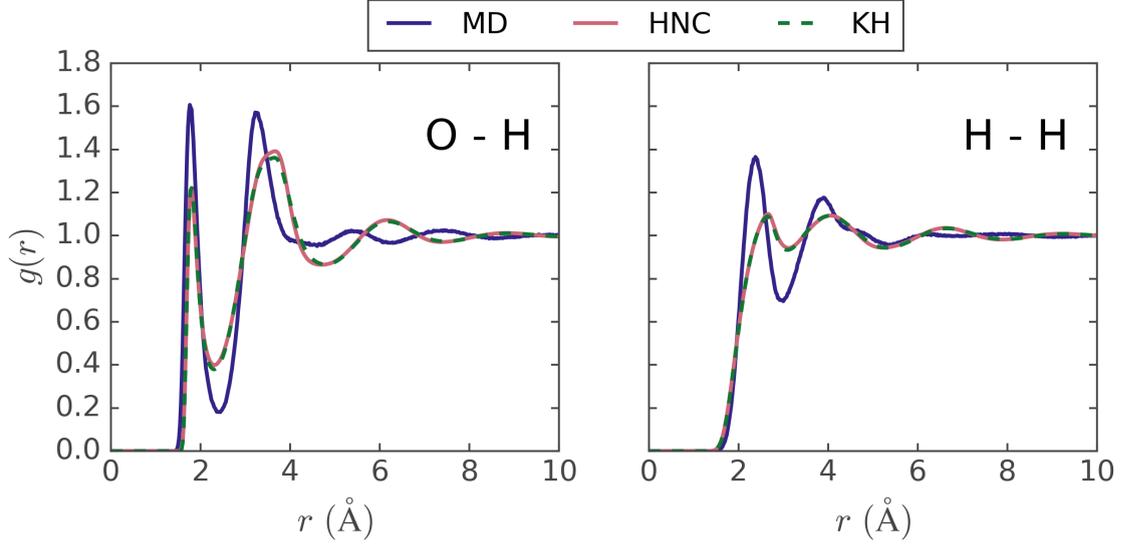


FIGURE B.1: Site-site radial distribution functions of SPC/E water (cSPC/E in case of DRISM calculations).

Applying the same procedure for the second derivative in equation B.2 we get

$$\begin{aligned}
 \left(\frac{\partial N}{\partial \mu}\right)_{T,V,N'} &= \left(\frac{\partial \mu}{\partial N}\right)_{T,V,N'}^{-1} = \left[-s \left(\frac{\partial T}{\partial N}\right)_{T,V,N'} + v \left(\frac{\partial P}{\partial N}\right)_{T,V,N'} \right]^{-1} \\
 &= \rho \left(\frac{\partial N}{\partial P}\right)_{T,V,N'} = -\rho \left(\frac{\partial V}{\partial P}\right)_{T,N,N'} \left(\frac{\partial N}{\partial V}\right)_{T,P,N'} = -\rho^2 \left(\frac{\partial V}{\partial P}\right)_{T,N,N'} ,
 \end{aligned} \tag{B.4}$$

since $(\partial N/\partial V)_{T,P,N'} = \rho$. Finally, we plug in the above two results into original equation to obtain

$$\left(\frac{\partial N}{\partial N'}\right)_{T,V,\mu} = -v \left(\frac{\partial V}{\partial N'}\right)_{T,P,N} \left(\frac{\partial P}{\partial V}\right)_{T,N,N'} \rho^2 \left(\frac{\partial V}{\partial P}\right)_{T,N,N'} = -\rho \bar{V}, \tag{B.5}$$

which is our final relationship.

B.2 Extra figures and tables

In this section, we included a few extra figures and tables that did not make it into the main thesis. All of them are referenced and discussed in the main text.

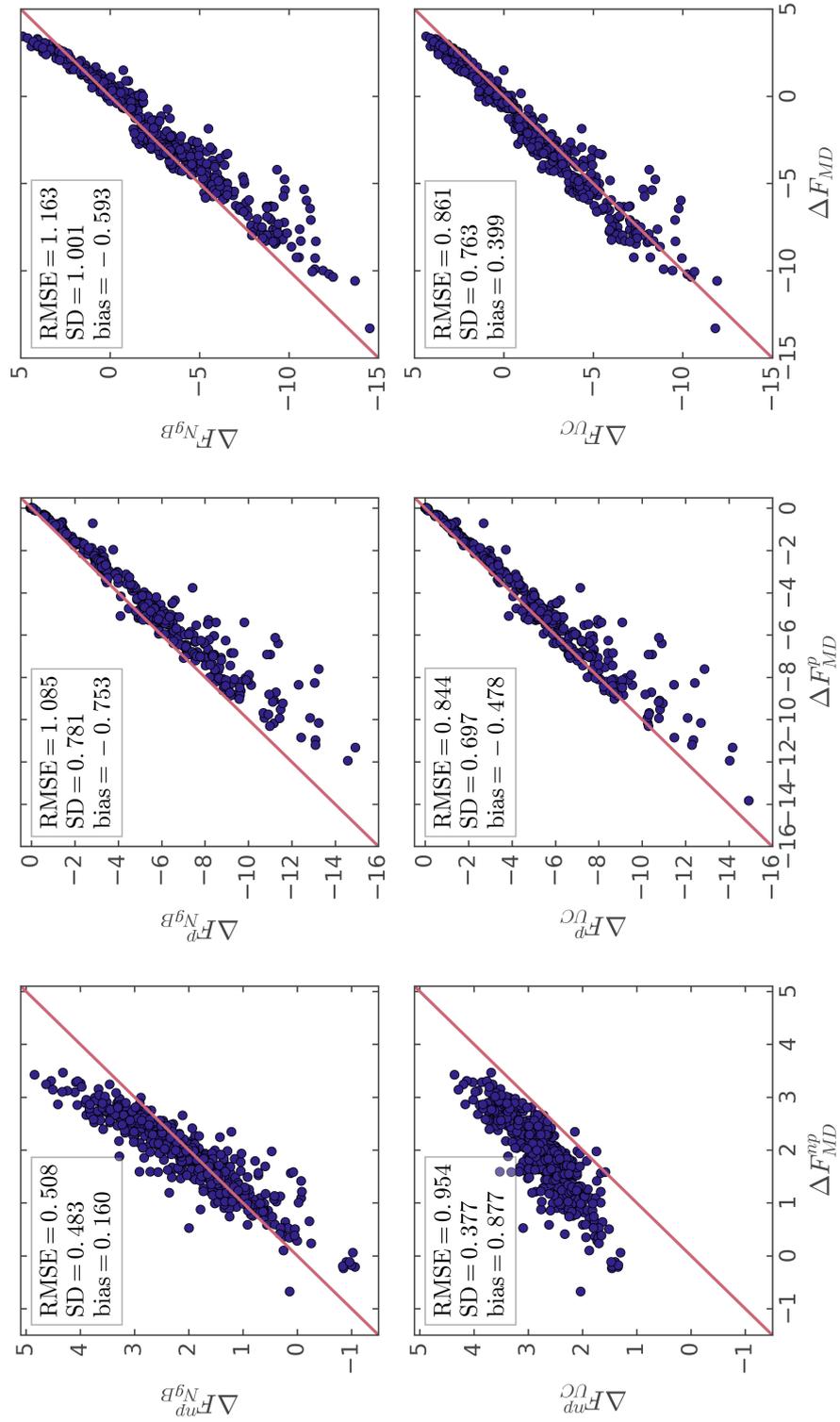


FIGURE B.2: The figure mirrors figure 6.1, but compares UC and NgB models, instead of PC and PC+.

Model	O - O			O - H			H - H		
	r	$g(r)$	CN	r	$g(r)$	CN	r	$g(r)$	CN
HNC	3.000	2.767	9.371	1.800	1.217	0.734	2.650	1.101	6.400
PSE-3	3.000	2.738	9.374	1.800	1.218	0.734	2.650	1.101	6.398
KH	2.975	2.338	9.988	1.800	1.221	0.728	2.650	1.089	6.155
MD	2.760	3.098	4.327	1.760	1.607	0.951	2.380	1.365	5.518

TABLE B.1: Location of the first peak in \AA r , the value of radial distribution function in the first peak $g(r)$, and coordination number CN for site-site radial distribution functions in SPC/E (cSPC/E in case of DRISM) water model.

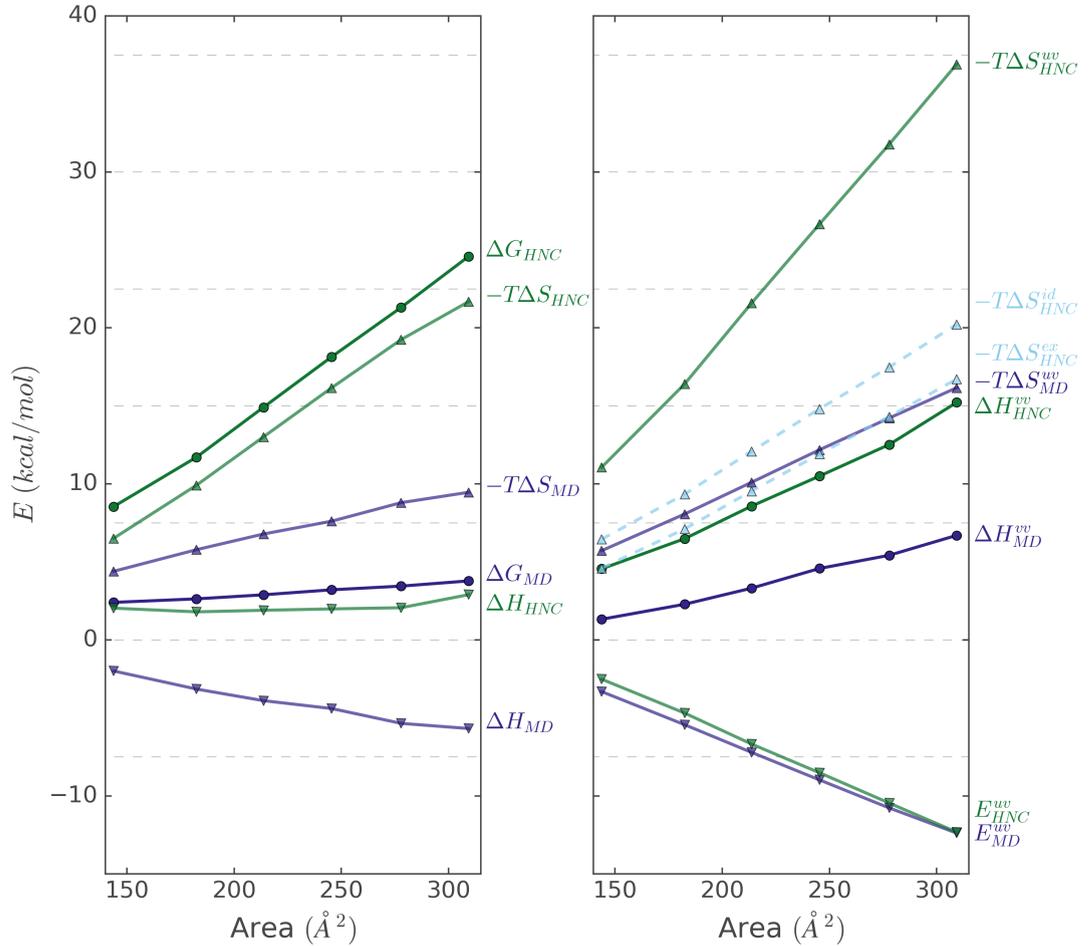


FIGURE B.3: The figure mirrors figures 5.7 and 5.8, but instead of pressure corrected models uses HNC results. The molecular dynamics results are taken from Ref. 151.

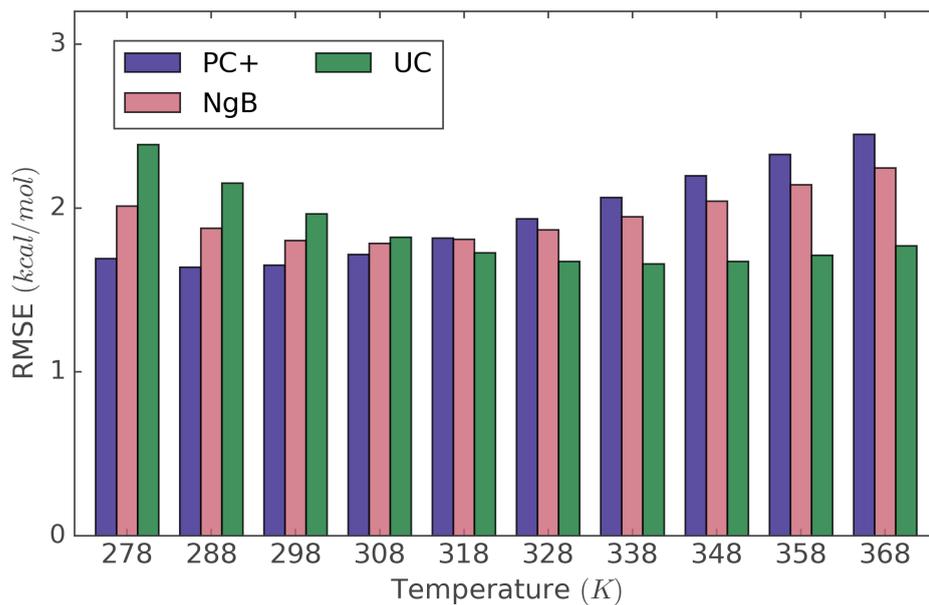


FIGURE B.4: Root mean square error of solvation free energy of 3D-RISM models on the Chamberlin dataset. Unlike figure 6.6, here we do not vary temperature in 3D-RISM calculations, using the value of ΔG at 298 K for all comparisons.

TABLE B.2: Accuracies of different models for predicting Setschenow’s constant using PC+’ correction (defined using equation 7.6). The units are 1/mol.

Model	RMSE	SDE	bias	r^2
OPLS/CM5				
da	0.225	0.061	0.217	0.722
ho _b	0.195	0.055	0.187	0.752
jc	0.366	0.091	0.354	0.617
ho _a	0.136	0.050	0.126	0.796
de	0.079	0.049	0.062	0.744

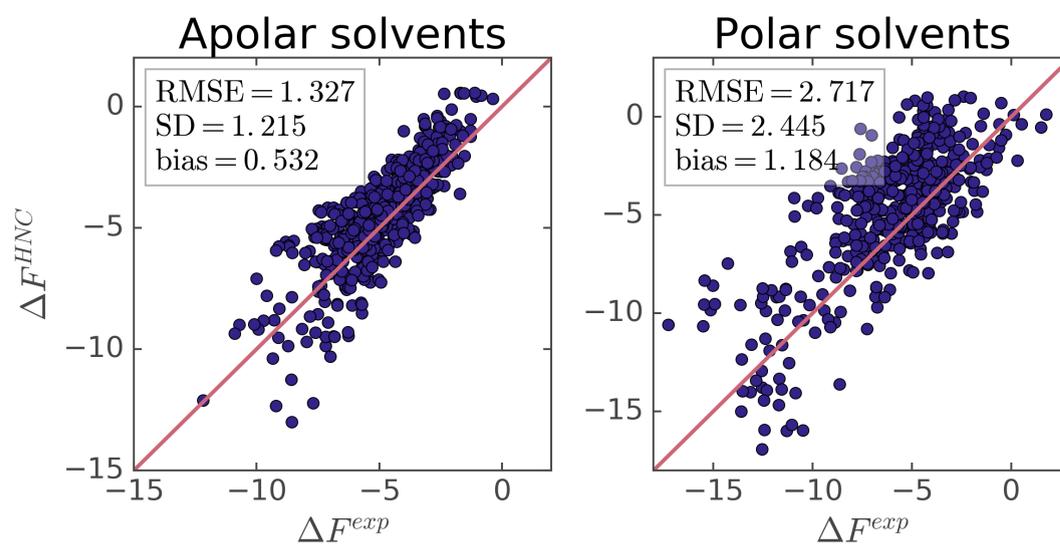


FIGURE B.5: Comparison of computed (3D-RISM/HNC) and experimental solvation free energies in polar and apolar solvents (as defined in section 7.5). All values are in kcal/mol.

List of original publications

1. **Misin, M.**, Fedorov, M.V., Palmer, D.S. Accurate Hydration Free Energies at a Wide Range of Temperatures from 3D-RISM. *J. Chem. Phys.*, 142(9):091105, March 2015.
2. Palmer, D.S., **Misin, M.**, Fedorov, M.V., Llinas, A. Fast and General Method To Predict the Physicochemical Properties of Druglike Molecules Using the Integral Equation Theory of Molecular Liquids. *Mol. Pharm.*, 12(9):3420–3432, August 2015.
3. **Misin, M.**, Fedorov, M.V., Palmer, D.S. Hydration Free Energies of Ionic Species by Molecular Theory and Simulation. *J. Phys. Chem. B*, 120(5):975–983, February 2016.
4. **Misin, M.**, Palmer, D.S., Fedorov, M.V. Predicting Solvation Free Energies Using Parameter-Free Solvent Models. *J. Phys. Chem. B*, 120(25):5724–57313, June 2016.
5. **Misin, M.**, Petteri, V., Fedorov M.V., Palmer, D.S. Salting-out effects by pressure-corrected 3D-RISM. *J. Chem. Phys.*, 145(19):194501, October 2016.

Bibliography

- [1] Jason D. Thompson, Christopher J. Cramer, and Donald G. Truhlar. Predicting aqueous solubilities from aqueous free energies of solvation and experimental or calculated vapor pressures of pure substances. *J. Chem. Phys.*, 119(3):1661–1670, July 2003.
- [2] David S. Palmer, Antonio Llinàs, Iñaki Morao, Graeme M. Day, Jonathan M. Goodman, Robert C. Glen, and John B. O. Mitchell. Predicting intrinsic aqueous solubility by a thermodynamic cycle. *Mol. Pharm.*, 5(2):266–279, April 2008.
- [3] David S. Palmer, James L. McDonagh, John B. O. Mitchell, Tanja van Mourik, and Maxim V. Fedorov. First-principles calculation of the intrinsic aqueous solubility of crystalline druglike molecules. *J. Chem. Theory Comput.*, 8(9):3322–3337, September 2012.
- [4] Rodrigo Casasnovas, Joaquin Ortega-Castro, Juan Frau, Josefa Donoso, and Francisco Muñoz. Theoretical pKa calculations with continuum model solvents, alternative protocols to thermodynamic cycles. *Int. J. Quantum Chem.*, 114(20):1350–1363, October 2014.
- [5] Nuno M. Garrido, António J. Queimada, Miguel Jorge, Eugénia A. Macedo, and Ioannis G. Economou. 1-Octanol/Water partition coefficients of n-alkanes from molecular simulations of absolute solvation free energies. *J. Chem. Theory Comput.*, 5(9):2436–2446, September 2009.
- [6] Ekaterina L. Ratkova, David S. Palmer, and Maxim V. Fedorov. Solvation thermodynamics of organic molecules by the molecular integral equation theory: Approaching chemical accuracy. *Chem. Rev.*, 115(13):6312–6356, July 2015.

- [7] Kristin S. Alongi and George C. Shields. Chapter 8 - theoretical calculations of acid dissociation constants: A review article. In Ralph A. Wheeler, editor, *Annual Reports in Computational Chemistry*, volume 6, pages 113–138. Elsevier, 2010.
- [8] Michael K. Gilson and Huan-Xiang Zhou. Calculation of protein-ligand binding affinities. *Annu. Rev. Bioph. Biom.*, 36(1):21–42, June 2007.
- [9] Munetaka Takeuchi, Nobuyuki Matubayasi, Yasuo Kameda, Babak Minofar, Shin-ichi Ishiguro, and Yasuhiro Umebayashi. Free-energy and structural analysis of ion solvation and contact ion-pair formation of Li^+ with BF_4^- and PF_6^- in water and carbonate solvents. *J. Phys. Chem. B*, 116(22):6476–6487, June 2012.
- [10] Ekaterina L. Ratkova and Maxim V. Fedorov. Combination of RISM and cheminformatics for efficient predictions of hydration free energy of polyfragment molecules: Application to a set of organic pollutants. *J. Chem. Theory Comput.*, 7(5):1450–1457, May 2011.
- [11] Jin Zhang, Badamkhatan Tuguldur, and David van der Spoel. Force field benchmark of organic liquids. 2. gibbs energy of solvation. *J. Chem. Inf. Model.*, 55(6):1192–1201, June 2015.
- [12] Arieh Ben-Naim. *Molecular Theory of Solutions*. OUP, Oxford, July 2006.
- [13] David L. Mobley, Christopher I. Bayly, Matthew D. Cooper, Michael R. Shirts, and Ken A. Dill. Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *J. Chem. Theory Comput.*, 5(2):350–358, February 2009.
- [14] David L. Mobley, Karisa L. Wymer, Nathan M. Lim, and J. Peter Guthrie. Blind prediction of solvation free energies from the SAMPL4 challenge. *J Comput Aided Mol Des*, 28(3):135–150, March 2014.
- [15] Devleena Shivakumar, Edward Harder, Wolfgang Damm, Richard A. Friesner, and Woody Sherman. Improving the prediction of absolute solvation free energies using the next generation OPLS force field. *J. Chem. Theory Comput.*, 8(8):2553–2558, August 2012.

- [16] Pavel V. Klimovich and David L. Mobley. Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. *J Comput Aided Mol Des*, 24(4):307–316, April 2010.
- [17] Pavel V. Klimovich, Michael R. Shirts, and David L. Mobley. Guidelines for the analysis of free energy calculations. *J Comput Aided Mol Des*, 29(5):397–411, March 2015.
- [18] Andrew Pohorille, Christopher Jarzynski, and Christophe Chipot. Good practices in free-energy calculations. *J. Phys. Chem. B*, 114(32):10235–10253, August 2010.
- [19] Michael R. Shirts and David L. Mobley. An introduction to best practices in free energy calculations. In Luca Monticelli and Emppu Salonen, editors, *Biomolecular Simulations*, number 924 in Methods in Molecular Biology, pages 271–311. Humana Press, January 2013.
- [20] Christopher J. Cramer. *Essentials of Computational Chemistry: Theories and Models*. Wiley, Chichester, West Sussex, England ; Hoboken, NJ, 2 edition edition, October 2004.
- [21] Angelika Stenzel, Kai-Uwe Goss, and Satoshi Endo. Experimental determination of polyparameter linear free energy relationship (pp-LFER) substance descriptors for pesticides and other contaminants: New measurements and recommendations. *Environ. Sci. Technol.*, 47(24):14204–14214, December 2013.
- [22] Paul C. M. van Noort, Joris J. H. Haftka, and John R. Parsons. Updated abraham solvation parameters for polychlorinated biphenyls. *Environ. Sci. Technol.*, 44(18):7037–7042, September 2010.
- [23] Guillaume Jeanmairat, Maximilien Levesque, Volodymyr Sergiievskyi, and Daniel Borgis. Classical density functional theory to tackle solvation in molecular liquids. In *Computational Trends in Solvation and Transport in Liquids*, volume 28, pages 187–208. Schriften des Forschungszentrums Jülich, Jülich, 2015.
- [24] Yu Liu, Jia Fu, and Jianzhong Wu. High-throughput prediction of the hydration free energies of small molecules from a classical density functional theory. *J. Phys. Chem. Lett.*, 4(21):3687–3691, November 2013.

- [25] Robert Evans, Martin Oettel, Roland Roth, and Gerhard Kahl. New developments in classical density functional theory. *J. Phys.: Condens. Matter*, 28(24):240401, 2016.
- [26] J.-P. Hansen and I. R. McDonald. *Theory of Simple Liquids, 4th ed.* Elsevier Academic Press, Amsterdam, The Netherlands, 2000.
- [27] Douglas Henderson, editor. *Fundamentals of Inhomogeneous Fluids.* CRC Press, New York, August 1992.
- [28] D. Beglov and B. Roux. An integral equation to describe the solvation of polar molecules in liquid water. *J. Phys. Chem.*, 101:7821–7826, 1997.
- [29] A. Kovalenko and F. Hirata. Three-dimensional density profiles of water in contact with a solute of arbitrary shape: A RISM approach. *Chem. Phys. Lett.*, 290(1-3):237–244, June 1998.
- [30] Fumio Hirata. *Molecular Theory of Solvation.* Kluwer Academic Publishers, New York, 2003.
- [31] David S. Palmer, Andrey I. Frolov, Ekaterina L. Ratkova, and Maxim V. Fedorov. Towards a universal method for calculating hydration free energies: a 3D reference interaction site model with partial molar volume correction. *J. Phys.: Condens. Matter*, 22(49):492101, December 2010.
- [32] D. S. Palmer, V. P. Sergiievskiy, F. Jensen, and M. V. Fedorov. Accurate calculations of the hydration free energies of druglike molecules using the reference interaction site model. *J. Chem. Phys.*, 133(4):044104, July 2010.
- [33] Ekaterina L. Ratkova, Gennady N. Chuev, Volodymyr P. Sergiievskiy, and Maxim V. Fedorov. An accurate prediction of hydration free energies by combination of molecular integral equations theory with structural descriptors. *J. Phys. Chem. B*, 114(37):12068–12079, September 2010.
- [34] D. S. Palmer, A. I. Frolov, E. L. Ratkova, and M. V. Fedorov. Toward a universal model to calculate the solvation thermodynamics of druglike molecules: The importance of new experimental databases. *Mol. Pharm.*, 8(4):1423–1429, May 2011.

- [35] Jean-François Truchon, B. Montgomery Pettitt, and Paul Labute. A cavity corrected 3D-RISM functional for accurate solvation free energies. *J. Chem. Theory Comput.*, 10(3):934–941, March 2014.
- [36] David S. Palmer, Maksim Mišin, Maxim V. Fedorov, and Antonio Llinas. Fast and general method to predict the physicochemical properties of druglike molecules using the integral equation theory of molecular liquids. *Mol. Pharmaceutics*, 12(9):3420–3432, September 2015.
- [37] Volodymyr P. Sergiievskiy, Guillaume Jeanmairet, Maximilien Levesque, and Daniel Borgis. Fast computation of solvation free energies with molecular density functional theory: Thermodynamic-ensemble partial molar volume corrections. *J. Phys. Chem. Lett.*, 5(11):1935–1942, June 2014.
- [38] Volodymyr Sergiievskiy, Guillaume Jeanmairet, Maximilien Levesque, and Daniel Borgis. Solvation free-energy pressure corrections in the three dimensional reference interaction site model. *J. Chem. Phys.*, 143(18):184116, November 2015.
- [39] Maksim Misin, Maxim V. Fedorov, and David S. Palmer. Communication: Accurate hydration free energies at a wide range of temperatures from 3D-RISM. *J. Chem. Phys.*, 142(9):091105, March 2015.
- [40] Maksim Misin, Maxim V. Fedorov, and David S. Palmer. Hydration free energies of molecular ions from theory and simulation. *J. Phys. Chem. B*, 120(5):975–983, February 2016.
- [41] Maksim Misin, David S. Palmer, and Maxim V. Fedorov. Predicting solvation free energies using parameter-free solvent models. *J. Phys. Chem. B*, June 2016.
- [42] Maksim Misin, Petteri Vainikka, Maxim V. Fedorov, and David S. Palmer. Salting-out effects by pressure-corrected 3D-RISM. *J. Chem. Phys.*, 145(18):Manuscript Accepted, November 2016.
- [43] M. Diraison, G. J. Martyna, and M. E. Tuckerman. Simulation studies of liquid ammonia by classical ab initio, classical, and path-integral molecular dynamics. *J. Chem. Phys.*, 111(3):1096–1103, July 1999.
- [44] Donald A. McQuarrie and John D. Simon. *Physical Chemistry : A Molecular Approach*. University Science Books, Sausalito, Calif, August 1997.

- [45] Herbert B. Callen. *Thermodynamics and an introduction to thermostatistics*. Wiley, New York, 2nd ed edition, 1985.
- [46] Ralph Baierlein. The elusive chemical potential. *AJP*, 69(4):423–434, April 2001.
- [47] B. Widom. Some topics in the theory of fluids. *J. Chem. Phys.*, 39(11):2808–2812, December 1963.
- [48] B. Widom. Potential-distribution theory and the statistical mechanics of fluids. *J. Phys. Chem.*, 86(6):869–872, March 1982.
- [49] Dor Ben-Amotz, Fernando O. Raineri, and George Stell. Solvation thermodynamics: Theory and applications†. *J. Phys. Chem. B*, 109(14):6866–6878, April 2005.
- [50] J. G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3:300–313, 1935.
- [51] Dor Ben-Amotz and Robin Underwood. Unraveling water’s entropic mysteries: A unified view of nonpolar, polar, and ionic hydration. *Acc. Chem. Res.*, 41(8):957–967, August 2008.
- [52] Andrew R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education, 2001.
- [53] Hong Qian and J. J. Hopfield. Entropy-Enthalpy compensation: Perturbation and relaxation in thermodynamic systems. *J. Chem. Phys.*, 105(20):9292–9298, November 1996.
- [54] A. D. McNaught and A. Wilkinson. *IUPAC. Compendium of Chemical Terminology*. Blackwell Scientific Publications, Oxford, 2 edition, 1997.
- [55] Dor Ben-Amotz. Global thermodynamics of hydrophobic cavitation, dewetting, and hydration. *J. Chem. Phys.*, 123(18):184504, November 2005.
- [56] Hsiang-Ai Yu and Martin Karplus. A thermodynamic analysis of solvation. *J. Chem. Phys.*, 89(4):2366–2379, August 1988.
- [57] M. P. Allen and D. J. Tildesley, editors. *Computer Simulation of Liquids*. Clarendon Press, Oxford, UK, 1987.

- [58] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–1174, July 2004.
- [59] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffrey D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, July 1983.
- [60] Jay L. Banks, Hege S. Beard, Yixiang Cao, Art E. Cho, Wolfgang Damm, Ramy Farid, Anthony K. Felts, Thomas A. Halgren, Daniel T. Mainz, Jon R. Maple, Robert Murphy, Dean M. Philipp, Matthew P. Repasky, Linda Y. Zhang, Bruce J. Berne, Richard A. Friesner, Emilio Gallicchio, and Ronald M. Levy. Integrated modeling program, applied chemical theory (IMPACT). *J. Comput. Chem.*, 26(16):1752–1780, December 2005.
- [61] Edward Harder, Wolfgang Damm, Jon Maple, Chuanjie Wu, Mark Reboul, Jin Yu Xiang, Lingle Wang, Dmitry Lupyan, Markus K. Dahlgren, Jennifer L. Knight, Joseph W. Kaus, David S. Cerutti, Goran Krilov, William L. Jorgensen, Robert Abel, and Richard A. Friesner. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.*, 12(1):281–296, January 2016.
- [62] Araz Jakalian, David B. Jack, and Christopher I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *J Comput Chem*, 23(16):1623–1641, December 2002.
- [63] Philippe Hunenberger and Maria Reif. *Single-Ion Solvation: Experimental and Theoretical Approaches to Elusive Thermodynamic Quantities*. Royal Society of Chemistry, Cambridge, April 2011.
- [64] Jun Cheng and Michiel Sprik. Alignment of electronic energy levels at electrochemical interfaces. *Phys. Chem. Chem. Phys.*, 14(32):11245–11267, 2012.
- [65] Ns Hush. The free energies of hydration of gaseous ions. *Aust. J. Chem.*, 1(4):480, 1948.
- [66] J. E. B. Randles. The real hydration energies of ions. *Trans. Faraday Soc.*, 52(0):1573–1581, January 1956.

- [67] Michael D. Tissandier, Kenneth A. Cowen, Wan Yong Feng, Ellen Gundlach, Michael H. Cohen, Alan D. Earhart, James V. Coe, and Thomas R. Tuttle. The proton's absolute aqueous enthalpy and gibbs free energy of solvation from cluster-ion solvation data. *J. Phys. Chem. A*, 102(40):7787–7794, October 1998.
- [68] D. Asthagiri, Lawrence R. Pratt, and H. S. Ashbaugh. Absolute hydration free energies of ions, Ion–Water clusters, and quasichemical theory. *J. Chem. Phys.*, 119(5):2702–2708, August 2003.
- [69] Thomas L. Beck. The influence of water interfacial potentials on ion hydration in bulk water and near interfaces. *Chem. Phys. Lett.*, 561–562:1–13, March 2013.
- [70] Casey P. Kelly, Christopher J. Cramer, and Donald G. Truhlar. Aqueous solvation free energies of ions and Ion-Water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. *J. Phys. Chem. B*, 110(32):16066–16081, August 2006.
- [71] R. Sander. Compilation of henry's law constants (version 4.0) for water as solvent. *Atmos. Chem. Phys.*, 15(8):4399–4981, April 2015.
- [72] A. N. Gorban. Basic types of coarse-graining. In Alexander N. Gorban, Ioannis G. Kevrekidis, Constantinos Theodoropoulos, Nikolaos K. Kazantzis, and Hans Christian Öttinger, editors, *Model Reduction and Coarse-Graining Approaches for Multiscale Phenomena*, pages 117–176. Springer Berlin Heidelberg, 2006.
- [73] Gergely Tóth. Determination of pair-potential parameters from experimental structure factors. *J. Chem. Phys.*, 115(10):4770–4775, September 2001.
- [74] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev. B*, 136(3B):B864–&, 1964.
- [75] N. David Mermin. Thermal properties of the inhomogeneous electron gas. *Phys. Rev.*, 137(5A):A1441–A1443, March 1965.
- [76] J. K. Labanowski and J. W. Andzelm. *Density functional methods in chemistry*. Springer-Verlag New York, Inc, 1991.
- [77] Feliciano Giustino. *Materials Modelling using Density Functional Theory: Properties and Predictions*. Oxford University Press, Oxford, July 2014.

- [78] Jean-Pierre Hansen and Enrico Smargiassi. Density functional methods at finite temperature. In *Conference Proceedings-Italian Physical Society*, volume 49, page 573–600. EDITRICE COMPOSITORI, 1996.
- [79] V. P. Sergiievskiy. *Modelling of Solvation Thermodynamics by Using a Combination of Reference Interaction Site Model Theory and Multi-grid Numerical Methods*. Doctoral thesis, University of Strathclyde, Glasgow, 2013.
- [80] Robert Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University Press, Oxford ; New York, April 2001.
- [81] D. Chandler, J. D. McCoy, and S. J. Singer. Density functional theory of nonuniform polyatomic systems. 1. general formulation. *J. Chem. Phys.*, 85(10):5971–5976, 1986.
- [82] H. C. Andersen. The structure of liquids. *Annu. Rev. Phys. Chem.*, 26:145–166, 1975.
- [83] Paul E. Smith, Enrico Matteoli, and John P. O’Connell. *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering, and Biophysics*. CRC Press, February 2013.
- [84] John P. O’Connell and Jens Abildskov. 9. molecular thermodynamic modeling of fluctuation solution theory properties. In *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering, and Biophysics*, page 225. CRC Press, 2016.
- [85] Vladimir Diky, John P. O’Connell, Jens Abildskov, Kenneth Kroenlein, and Michael Frenkel. Representation and validation of liquid densities for pure compounds and mixtures. *J. Chem. Eng. Data*, 60(12):3545–3553, December 2015.
- [86] J. L. Yarnell, M. J. Katz, R. G. Wenzel, and S. H. Koenig. Structure factor and radial distribution function for liquid argon at 85 $^{\circ}\text{K}$. *Phys. Rev. A*, 7(6):2130–2144, June 1973.
- [87] D. Chandler, J. D. McCoy, and S. J. Singer. Density functional theory of nonuniform polyatomic systems. 2. rational closures for integral-equations. *J. Chem. Phys.*, 85(10):5977–5982, November 1986.

- [88] Tohru Morita. Theory of classical fluids: Hyper-netted chain approximation, i formulation for a one-component system. *Prog. Theor. Phys.*, 20(6):920–938, December 1958.
- [89] B. Widom. Intermolecular forces and the nature of the liquid state. *Science*, 157(3787):375–382, July 1967.
- [90] K. E. Gubbins. *Theory of Molecular Fluids: Fundamentals Volume I*. Oxford University Press, Oxford ; New York, 1 edition edition, January 1985.
- [91] D. Chandler and H. C. Andersen. Optimized cluster expansions for classical fluids. 2. theory of molecular liquids. *J. Chem. Phys.*, 57(5):1930–1937, 1972.
- [92] David Chandler. Derivation of an integral equation for pair correlation functions in molecular fluids. *J. Chem. Phys.*, 59(5):2742–2746, September 1973.
- [93] Branka M. Ladanyi and David Chandler. New type of cluster theory for molecular fluids: Interaction site cluster expansion. *J. Chem. Phys.*, 62(11):4308–4324, June 1975.
- [94] Yu Liu, Shuangliang Zhao, and Jianzhong Wu. A site density functional theory for water: Application to solvation of amino acid side chains. *J. Chem. Theory Comput.*, 9(4):1896–1908, April 2013.
- [95] James P. Donley, John G. Curro, and John D. McCoy. A density functional theory for pair correlation functions in molecular liquids. *J. Chem. Phys.*, 101(4):3205–3215, August 1994.
- [96] PT Cummings and G. Stell. Interaction site models for molecular fluids. *Mol. Phys.*, 46(2):383–426, 1982.
- [97] George Stell and Yaoqi Zhou. Analytical approach to molecular liquids. i. site–site interaction model using an extended mean-spherical approximation. *J. Chem. Phys.*, 91(8):4861–4868, October 1989.
- [98] Yu. V. Kalyuzhnyi and P. T. Cummings. 6 equations of state from analytically solvable integral equation approximations. In C. J. Peters J.V. Sengers, R. F. Kayser and H. J. White, editors, *Experimental Thermodynamics*, volume 5 of *Equations of State for Fluids and Fluid Mixtures*, pages 169–254. Elsevier, 2000.

- [99] Phil Attard. *Thermodynamics and Statistical Mechanics: Equilibrium by Entropy Maximisation*. Academic Press, San Diego, Calif, 1 edition edition, August 2002.
- [100] David Chandler. Cluster diagrammatic analysis of the RISM equation. *Mol. Phys.*, 31(4):1213–1223, April 1976.
- [101] S. M. Kast and T. Kloss. Closed-form expressions of the chemical potential for integral equation closures with certain bridge functions. *J. Chem. Phys.*, 129(23):236101, December 2008.
- [102] C. S. Hsu, David Chandler, and L. J. Lowden. Applications of the RISM equation to diatomic fluids: the liquids nitrogen, oxygen and bromine. *Chemical Physics*, 14(2):213–228, May 1976.
- [103] Branka M. Ladanyi, Thomas Keyes, Dominic J. Tildesley, and William B. Streett. Structure and equilibrium optical properties of liquid CS₂. *Mol. Phys.*, 39(3):645–659, February 1980.
- [104] John Perkyns and B. Montgomery Pettitt. A site–site theory for finite concentration saline solutions. *J. Chem. Phys.*, 97(10):7656–7666, November 1992.
- [105] Tyler Luchko, Sergey Gusarov, Daniel R. Roe, Carlos Simmerling, David A. Case, Jack Tuszynski, and Andriy Kovalenko. Three-dimensional molecular theory of solvation coupled with molecular dynamics in amber. *J. Chem. Theory Comput.*, 6(3):607–624, March 2010.
- [106] Dragan Nikolić, Nikolay Blinov, David Wishart, and Andriy Kovalenko. 3D-RISM-Dock: a new fragment-based drug design protocol. *J. Chem. Theory Comput.*, 8(9):3356–3372, September 2012.
- [107] M. C. Stumpe, N. Blinov, D. Wishart, A. Kovalenko, and V. S. Pande. Calculation of local water densities in biological systems: A comparison of molecular dynamics simulations and the 3D-RISM-KH molecular theory of solvation. *J. Phys. Chem. B*, 115(2):319–328, January 2011.
- [108] Takashi Imai, Koji Oda, Andriy Kovalenko, Fumio Hirata, and Akinori Kidera. Ligand mapping on protein surfaces by the 3D-RISM theory: Toward computational fragment-based drug design. *J. Am. Chem. Soc.*, 131(34):12430–12440, September 2009.

- [109] Takashi Imai, Ryusuke Hiraoka, Andriy Kovalenko, and Fumio Hirata. Locating missing water molecules in protein cavities by the three-dimensional reference interaction site model theory of molecular solvation. *Proteins*, 66(4):804–813, March 2007.
- [110] Daniel J. Sindhikara, Norio Yoshida, and Fumio Hirata. Placevent: An algorithm for prediction of explicit solvent atom distribution—Application to HIV-1 protease and f-ATP synthase. *J. Comput. Chem.*, 33(18):1536–1543, July 2012.
- [111] C. M. Cortis, P. J. Rossky, and R. A. Friesner. A three-dimensional reduction of the ornstein-zernicke equation for molecular liquids. *J. Chem. Phys.*, 107:6400–6414, 1997.
- [112] Nobuyuki Matubayasi and Ronald M. Levy. Thermodynamics of the hydration shell. 2. excess volume and compressibility of a hydrophobic solute. *J. Phys. Chem.*, 100(7):2681–2688, January 1996.
- [113] Seishi Shimizu and Nobuyuki Matubayasi. Preferential solvation: Dividing surface vs excess numbers. *J. Phys. Chem. B*, 118(14):3922–3930, April 2014.
- [114] A. W. Neumann, Robert David, and Yi Zuo, editors. *Applied Surface Thermodynamics, Second Edition*. CRC Press, Boca Raton, 2 edition edition, October 2010.
- [115] Tyler Luchko, Nikolay Blinov, Garrett C. Limon, Kevin P. Joyce, and Andriy Kovalenko. SAMPL5: 3D-RISM partition coefficient calculations with partial molar volume corrections and solute conformational sampling. *J Comput Aided Mol Des*, pages 1–13, September 2016.
- [116] J. Johnson, D. A. Case, T. Yamazaki, S. Gusarov, A. Kovalenko, and T. Luchko. Small molecule hydration energy and entropy from 3D-RISM. *J. Phys.: Condens. Matter*, 28(34):344002, 2016.
- [117] Nicolas Tielker, Daniel Tomazic, Jochen Heil, Thomas Kloss, Sebastian Ehrhart, Stefan Güssregen, K. Friedemann Schmidt, and Stefan M. Kast. The SAMPL5 challenge for embedded-cluster integral equation theory: solvation free energies, aqueous pKa, and cyclohexane–water log d. *J Comput Aided Mol Des*, 30(11):1035–1044, November 2016.

- [118] M. V. Vener, I. V. Leontyev, Yu. A. Dyakov, M. V. Basilevsky, and M. D. Newton. Application of the linearized MD approach for computing equilibrium solvation free energies of charged and dipolar solutes in polar solvents. *J. Phys. Chem. B*, 106(50):13078–13088, December 2002.
- [119] R. Evans, P. Tarazona, and U. Marini Bettolo Marconi. On the failure of certain integral equation theories to account for complete wetting at solid-fluid interfaces. *Mol. Phys.*, 50(5):993–1011, December 1983.
- [120] Phil Attard, D. R. Bérard, C. P. Ursenbach, and G. N. Patey. Interaction free energy between planar walls in dense fluids: An ornstein-zernike approach with results for hard-sphere, lennard-jones, and dipolar systems. *Phys. Rev. A*, 44(12):8224–8234, December 1991.
- [121] Liudmila A. Pozhar, Keith E. Gubbins, and Jerome K. Percus. Generalized compressibility equation for inhomogeneous fluids at equilibrium. *Phys. Rev. E*, 48(3):1819–1822, September 1993.
- [122] Shuangliang Zhao, Yu Liu, Honglai Liu, and Jianzhong Wu. Site-site direct correlation functions for three popular molecular models of liquid water. *J. Chem. Phys.*, 139(6):064509, August 2013.
- [123] Shijie Sheng and Jianzhong Wu. Direct correlation functions for three-site and four-site water models. *Mol. Phys.*, 114(16-17):2351–2363, September 2016.
- [124] Emmanuele DiBenedetto. *Real Analysis*. Birkhäuser Advanced Texts Basler Lehrbücher. Springer New York, New York, NY, 2016.
- [125] C. Tanford. The hydrophobic effect and the organization of living matter. *Science*, 200(4345):1012–1018, June 1978.
- [126] Wilfried Blokzijl and Jan B. F. N. Engberts. Hydrophobic effects. opinions and facts. *Angew. Chem. Int. Ed. Engl.*, 32(11):1545–1579, November 1993.
- [127] Shekhar Garde, Gerhard Hummer, Angel E. García, Michael E. Paulaitis, and Lawrence R. Pratt. Origin of entropy convergence in hydrophobic hydration and protein folding. *Phys. Rev. Lett.*, 77(24):4966–4968, December 1996.

- [128] G. Hummer, S. Garde, A. E. García, M. E. Paulaitis, and L. R. Pratt. Hydrophobic effects on a molecular scale. *J. Phys. Chem. B*, 102(51):10469–10482, December 1998.
- [129] David Chandler and Patrick Varilly. Lectures on molecular- and nano-scale fluctuations in water. *arXiv*, January 2011.
- [130] Takashi Imai, Masahiro Kinoshita, and Fumio Hirata. Theoretical study for partial molar volume of amino acids in aqueous solution: Implication of ideal fluctuation volume. *J. Chem. Phys.*, 112(21):9469–9478, June 2000.
- [131] David Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640–647, September 2005.
- [132] John D. Weeks, Robin L. B. Selinger, and Jeremy Q. Broughton. Self-consistent treatment of repulsive and attractive forces in nonuniform liquids. *Phys. Rev. Lett.*, 75(14):2694–2697, October 1995.
- [133] John D. Weeks, Kirill Katsov, and Katharina Vollmayr. Roles of repulsive and attractive forces in determining the structure of nonuniform liquids: Generalized mean field theory. *Phys. Rev. Lett.*, 81(20):4400–4403, November 1998.
- [134] Ka Lum, David Chandler, and John D. Weeks. Hydrophobicity at small and large length scales. *J. Phys. Chem. B*, 103(22):4570–4577, June 1999.
- [135] David M. Huang and David Chandler. Cavity formation and the drying transition in the lennard-jones fluid. *Phys. Rev. E*, 61(2):1501–1506, February 2000.
- [136] David M. Huang and David Chandler. The hydrophobic effect and the influence of Solute-Solvent attractions. *J. Phys. Chem. B*, 106(8):2047–2053, February 2002.
- [137] David M. Huang, Phillip L. Geissler, and David Chandler. Scaling of hydrophobic solvation free energies. *J. Phys. Chem. B*, 105(28):6704–6709, July 2001.
- [138] Frank H. Stillinger. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *J Solution Chem*, 2(2-3):141–158, 1972.
- [139] Alan K. Soper and Maria Antonietta Ricci. Structures of high-density and low-density water. *Phys. Rev. Lett.*, 84(13):2881–2884, March 2000.

- [140] Bo Li, Alexei V. Matveev, and Notker Rösch. Three-dimensional reference interaction site model solvent combined with a quantum mechanical treatment of the solute. *Comput. Theor. Chem.*, 1070:143–151, October 2015.
- [141] J. Dzubiella, J. M. J. Swanson, and J. A. McCammon. Coupling nonpolar and polar solvation free energies in implicit solvent models. *J. Chem. Phys.*, 124(8):084905, February 2006.
- [142] Maria M. Reif and Philippe H. Hünenberger. Origin of asymmetric solvation effects for ions in water and organic solvents investigated using molecular dynamics simulations: The swain Acidity–Basicity scale revisited. *J. Phys. Chem. B*, 120(33):8485–8517, August 2016.
- [143] M. V. Fedorov and A. A. Kornyshev. Unravelling the solvent response to neutral and charged solutes. *Mol. Phys.*, 105(1):1–16, January 2007.
- [144] M. R. Bush, M. J. Booth, A. D. J. Haymet, and A. G. Schlijper. Integral equation approximations for inhomogeneous fluids: functional optimization. *Mol. Phys.*, 95(3):601–619, October 1998.
- [145] Kevin A. T. Silverstein, Ken A. Dill, and A. D. J. Haymet. Hydrophobicity in a simple model of water: Entropy penalty as a sum of competing terms via full, angular expansion. *J. Chem. Phys.*, 114(14):6303–6314, April 2001.
- [146] Tugba A. Özal and Nico F. A. van der Vegt. Confusing cause and effect: Energy-Entropy compensation in the preferential solvation of a nonpolar solute in dimethyl Sulfoxide/Water mixtures. *J. Phys. Chem. B*, 110(24):12104–12112, June 2006.
- [147] Isaac C. Sanchez and Thomas M. Truskett. Configurational properties and corresponding states in simple fluids and water. *J. Phys. Chem. B*, 103(24):5106–5116, June 1999.
- [148] Dor Ben-Amotz. Interfacial solvation thermodynamics. *J. Phys.: Condens. Matter*, 28(41):414013, 2016.
- [149] Michael E. Paulaitis, Henry S. Ashbaugh, and Shekhar Garde. The entropy of hydration of simple hydrophobic solutes. *Biophysical Chemistry*, 51(2):349–357, August 1994.

- [150] Henry S. Ashbaugh and Michael E. Paulaitis. Entropy of hydrophobic hydration: Extension to hydrophobic chains. *J. Phys. Chem.*, 100(5):1900–1913, January 1996.
- [151] E. Gallicchio, M. M. Kubo, and R. M. Levy. Enthalpy-Entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. *J. Phys. Chem. B*, 104(26):6271–6285, July 2000.
- [152] The International Association for the Properties of Water and Steam. Revised supplementary release on properties of liquid water at 0.1 MPa, September 2011.
- [153] Dmitry V. Matyushov and Gregory A. Voth. A perturbation theory for solvation thermodynamics: Dipolar–quadrupolar liquids. *J. Chem. Phys.*, 111(8):3630–3638, August 1999.
- [154] Aleksandr V. Marenich, Christopher J. Cramer, and Donald G. Truhlar. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B*, 113(18):6378–6396, May 2009.
- [155] Yan Zhao and Donald G. Truhlar. The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theor Chem Account*, 120(1-3):215–241, July 2007.
- [156] Benjamin J. Lynch, Yan Zhao, and Donald G. Truhlar. Effectiveness of diffuse basis functions for calculating relative energies by density functional theory. *J. Phys. Chem. A*, 107(9):1384–1388, March 2003.
- [157] D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham, T. A. Darden III, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, and P. A. Kollman. AMBER 2016, 2016.

- [158] A. K. Soper. The radial distribution functions of water as derived from radiation total scattering experiments: Is there anything we can say for sure? *Int. Sch. Res. Notices*, 2013:e279463, February 2013.
- [159] Adam C. Chamberlin, Christopher J. Cramer, and Donald G. Truhlar. Predicting aqueous free energies of solvation as functions of temperature. *J. Phys. Chem. B*, 110(11):5665–5675, March 2006.
- [160] Adam C. Chamberlin, Christopher J. Cramer, and Donald G. Truhlar. Extension of a temperature-dependent aqueous solvation model to compounds containing nitrogen, fluorine, chlorine, bromine, and sulfur. *J. Phys. Chem. B*, 112(10):3024–3039, March 2008.
- [161] Noel M. O’Boyle, Michael Banck, Craig A. James, Chris Morley, Tim Vandermeersch, and Geoffrey R. Hutchison. Open babel: An open chemical toolbox. *J Cheminform*, 3(1):33, October 2011.
- [162] Noel M. O’Boyle, Chris Morley, and Geoffrey R. Hutchison. Pybel: a python wrapper for the OpenBabel cheminformatics toolkit. *Chemistry Central Journal*, 2(1):1–5, December 2008.
- [163] Karol Daucik. Revised supplementary release on properties of liquid water at 0.1 MPa, September 2011.
- [164] A.V. Marenich, C.P. Kelly, J.D. Thompson, G.D. Hawkins, C.C. Chambers, D.J. Giesen, P. Winget, C.J. Cramer, and D.G. Truhlar. *Minnesota Solvation Database – version 2012*. University of Minnesota, Minneapolis, 2012.
- [165] Richard C. Remsing, Marcel D. Baer, Gregory K. Schenter, Christopher J. Mundy, and John D. Weeks. The role of broken symmetry in solvation of a spherical cavity in classical and quantum water models. *J. Phys. Chem. Lett.*, 5(16):2767–2774, August 2014.
- [166] Yen-Lin Lin, Alexey Aleksandrov, Thomas Simonson, and Benoît Roux. An overview of electrostatic free energy computations for solutions and proteins. *J. Chem. Theory Comput.*, 10(7):2690–2709, July 2014.
- [167] K. J. Tielrooij, N. Garcia-Araez, M. Bonn, and H. J. Bakker. Cooperativity in ion hydration. *Science*, 328(5981):1006–1009, May 2010.

- [168] Yizhak Marcus. Effect of ions on the structure of water: Structure making and breaking. *Chem. Rev.*, 109(3):1346–1370, March 2009.
- [169] Dwight R. Robinson and William P. Jencks. The effect of concentrated salt solutions on the activity coefficient of acetyltetraglycine ethyl ester. *J. Am. Chem. Soc.*, 87(11):2470–2479, June 1965.
- [170] W. F. McDevit and F. A. Long. The activity coefficient of benzene in aqueous salt solutions. *J. Am. Chem. Soc.*, 74(7):1773–1777, April 1952.
- [171] M. G. Cacace, E. M. Landau, and J. J. Ramsden. The hofmeister series: salt and solvent effects on interfacial phenomena. *Q. Rev. Biophys.*, 30(03):241–277, August 1997.
- [172] J. W. Readman, R. F. C. Mantoura, and M. M. Rhead. The physico-chemical speciation of polycyclic aromatic hydrocarbons (PAH) in aquatic systems. *Z. Anal. Chem.*, 319(2):126, 1984.
- [173] W. Mabey and T. Mill. Critical review of hydrolysis of organic compounds in water under environmental conditions. *J. Phys. Chem. Ref. Data*, 7(2):383–415, April 1978.
- [174] Richard A. Park, Jonathan S. Clough, and Marjorie Coombs Wellman. AQUATOX: modeling environmental fate and ecological effects in aquatic ecosystems. *Ecological Modelling*, 213(1):1–15, April 2008.
- [175] Brett A. Holmberg, Huanting Wang, Joseph M. Norbeck, and Yushan Yan. Controlling size and yield of zeolite γ nanocrystals using tetramethylammonium bromide. *Microporous and Mesoporous Materials*, 59(1):13–28, April 2003.
- [176] Shigeo Sasaki, Shogo Koga, Ryota Imabayashi, and Hiroshi Maeda. Salt effects on the volume phase transition of ionic gel induced by the hydrophobic counterion binding. *J. Phys. Chem. B*, 105(25):5852–5855, June 2001.
- [177] Wayne Melander and Csaba Horváth. Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: An interpretation of the lyotropic series. *Archives of Biochemistry and Biophysics*, 183(1):200–215, September 1977.

- [178] Satoshi Endo, Andrea Pfennigsdorff, and Kai-Uwe Goss. Salting-out effect in aqueous NaCl solutions: Trends with size and polarity of solute molecules. *Environ. Sci. Technol.*, 46(3):1496–1503, February 2012.
- [179] Yizhak Marcus. Individual ionic surface tension increments in aqueous solutions. *Langmuir*, 29(9):2881–2888, March 2013.
- [180] Mark A Schlautman, Soobin Yim, Elizabeth R Carraway, Ji Hoon Lee, and Bruce E Herbert. Testing a surface tension-based model to predict the salting out of polycyclic aromatic hydrocarbons in model environmental solutions. *Water Research*, 38(14–15):3331–3339, August 2004.
- [181] Weifeng Li and Yuguang Mu. Hydration patterns and salting effects in sodium chloride solution. *J. Chem. Phys.*, 135(13):134502, October 2011.
- [182] Andreas Klamt. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*. Elsevier Science, Amsterdam, July 2005.
- [183] Michael H. Abraham, Adam Ibrahim, and Andreas M. Zissimos. Determination of sets of solute descriptors from chromatographic measurements. *Journal of Chromatography A*, 1037(1–2):29–47, May 2004.
- [184] Liem X. Dang. Mechanism and thermodynamics of ion selectivity in aqueous solutions of 18-crown-6 ether: A molecular dynamics study. *J. Am. Chem. Soc.*, 117(26):6954–6960, July 1995.
- [185] In Suk Joung and Thomas E. Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112(30):9020–9041, July 2008.
- [186] Stephan Deublein, Jadran Vrabc, and Hans Hasse. A set of molecular models for alkali and halide ions in aqueous solution. *J. Chem. Phys.*, 136(8):084501, February 2012.
- [187] Dominik Horinek, Shavkat I. Mamatkulov, and Roland R. Netz. Rational design of ion force fields based on thermodynamic solvation properties. *J. Chem. Phys.*, 130(12):124507, March 2009.

- [188] Filip Moučka, Ivo Nezbeda, and William R. Smith. Molecular force fields for aqueous electrolytes: SPC/E-compatible charged LJ sphere models and their limitations. *J. Chem. Phys.*, 138(15):154102, April 2013.
- [189] Berk Hess, Christian Holm, and Nico van der Vegt. Osmotic coefficients of atomistic NaCl (aq) force fields. *J. Chem. Phys.*, 124(16):164509, April 2006.
- [190] Michael Patra and Mikko Karttunen. Systematic comparison of force fields for microscopic simulations of NaCl in aqueous solutions: Diffusion, free energy of hydration, and structural properties. *J. Comput. Chem.*, 25(5):678–689, April 2004.
- [191] Aldi Asmadi, Tom Kirchner, Wael Abdallah, Maxim V. Fedorov, and Mikhail R. Stukan. Influence of the drude charge value on the performance of polarisable water model: A test for microscopic and macroscopic parameters. *J. Mol. Liq.*, 188:245–251, December 2013.
- [192] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, November 1987.
- [193] Libo Li, Christopher J. Fennell, and Ken A. Dill. Small molecule solvation changes due to the presence of salt are governed by the cost of solvent cavity formation and dispersion. *J. Chem. Phys.*, 141(22):22D518, December 2014.
- [194] Nina Ni and Samuel H. Yalkowsky. Prediction of setschenow constants. *International Journal of Pharmaceutics*, 254(2):167–172, March 2003.
- [195] Yajun Li, Qinghua Hu, and Chongli Zhong. Topological modeling of the setschenow constant. *Ind. Eng. Chem. Res.*, 43(15):4465–4468, July 2004.
- [196] Jie Xu, Lei Wang, Luoxin Wang, Xiaolin Shen, and Weilin Xu. QSPR study of setschenow constants of organic compounds using MLR, ANN, and SVM analyses. *J. Comput. Chem.*, 32(15):3241–3252, November 2011.
- [197] Xinliang Yu and Ruqin Yu. Setschenow constant prediction based on the IEF-PCM calculations. *Ind. Eng. Chem. Res.*, 52(32):11182–11188, August 2013.
- [198] Fabien Archambault, Christophe Chipot, Ignacio Soteras, F. Javier Luque, Klaus Schulten, and François Dehez. Polarizable intermolecular potentials for water and benzene interacting with halide and metal ions. *J. Chem. Theory Comput.*, 5(11):3022–3031, November 2009.

- [199] Thomas W. Leland and Patsy S. Chappellear. The corresponding states principle - a review of current theory and practice. *Ind. Eng. Chem.*, 60(7):15–43, July 1968.
- [200] E. A. Guggenheim. The principle of corresponding states. *J. Chem. Phys.*, 13(7):253–261, July 1945.
- [201] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird. *Molecular theory of gases and liquids*. Chapman & Hall, New York, 1954.
- [202] Lawrence W. Flynn and George Thodos. Lennard-jones force constants from viscosity data: Their relationship to critical properties. *AIChE J.*, 8(3):362–365, June 1962.
- [203] Ting Horng Chung, Lloyd L. Lee, and Kenneth E. Starling. Applications of kinetic gas theories and multiparameter correlation for prediction of dilute gas viscosity and thermal conductivity. *Ind. Eng. Chem. Fund.*, 23(1):8–13, February 1984.
- [204] J. J. van Loef. Transport properties of dense monatomic and molecular fluids and their mixtures, and the corresponding states principle i. shear viscosity and thermal conductivity. *Physica B+C*, 124(3):305–326, June 1984.
- [205] F. Cuadros, I. Cachadiña, and W. Ahumada. Determination of lennard-jones interaction parameters using a new procedure. *Mol Eng*, 6(3):319–325, September 1996.
- [206] Dor Ben-Amotz, Alan D. Gift, and R. D. Levine. Improved ccorresponding states scaling of the equations of state of simple fluids. *J. Chem. Phys.*, 117(10):4632–4634, September 2002.
- [207] Yu Zhu, Xiaohua Lu, Jian Zhou, Yanru Wang, and Jun Shi. Prediction of diffusion coefficients for gas, liquid and supercritical fluid: Application to pure real fluids and infinite dilute binary solutions based on the simulation of Lennard–Jones fluid. *Fluid Phase Equilib.*, 194–197:1141–1159, March 2002.
- [208] Guillaume Galliéro, Christian Boned, and Antoine Baylaucq. Molecular dynamics study of the Lennard-Jones fluid viscosity: Application to real fluids. *Ind. Eng. Chem. Res.*, 44(17):6963–6972, August 2005.
- [209] Guillaume Galliero, Carlos Nieto-Draghi, Christian Boned, Josep B. Avalos, Allan D. Mackie, Antoine Baylaucq, and François Montel. Molecular dynamics simulation of

- acid gas mixtures: A comparison between several approximations. *Ind. Eng. Chem. Res.*, 46(15):5238–5244, July 2007.
- [210] Guillaume Galliero and Christian Boned. Shear viscosity of the lennard-jones chain fluid in its gaseous, supercritical, and liquid states. *Phys. Rev. E*, 79(2):021201, February 2009.
- [211] Andrés Mejía, Carmelo Herdes, and Erich A. Müller. Force fields for coarse-grained molecular simulations from a corresponding states correlation. *Ind. Eng. Chem. Res.*, 53(10):4131–4141, March 2014.
- [212] Hisashi Okumura and Fumiko Yonezawa. Liquid–vapor coexistence curves of several interatomic model potentials. *J. Chem. Phys.*, 113(20):9162–9168, November 2000.
- [213] Carl L. Yaws. *Thermophysical Properties of Chemicals and Hydrocarbons*. Gulf Professional Publishing, Amsterdam Boston, 2 edition, July 2014.
- [214] Thomas Lafitte, David Bessieres, Manuel M. Piñeiro, and Jean-Luc Daridon. Simultaneous estimation of phase behavior and second-derivative properties using the statistical associating fluid theory with variable range approach. *J. Chem. Phys.*, 124(2):024509, January 2006.
- [215] Guillaume Galliero. Equilibrium, interfacial and transport properties of n-alkanes: Towards the simplest coarse grained molecular model. *Chem. Eng. Res. Des.*, 92(12):3031–3037, December 2014.
- [216] Girija S. Dubey, Seamus F. O’Shea, and Peter A. Monson. Vapour-liquid equilibria for two centre lennard-jones diatomics and dipolar diatomics. *Mol. Phys.*, 80(4):997–1007, November 1993.
- [217] Felipe J. Blas and Lourdes F. Vega. Improved Vapor–Liquid equilibria predictions for lennard-jones chains from the statistical associating fluid dimer theory: Comparison with monte carlo simulations. *J. Chem. Phys.*, 115(9):4355–4358, September 2001.
- [218] Fernando A. Escobedo and Juan J. De Pablo. Simulation and prediction of vapour-liquid equilibria for chain molecules. *Mol. Phys.*, 87(2):347–366, February 1996.
- [219] Andreas Klamt and Michael Diedenhofen. Calculation of solvation free energies with DCOSMO-RS. *J. Phys. Chem. A*, 119(21):5439–5445, May 2015.

- [220] Andriy Kovalenko and Fumio Hirata. Potentials of mean force of simple ions in ambient aqueous solution. i. three-dimensional reference interaction site model approach. *J. Chem. Phys.*, 112(23):10391–10402, June 2000.
- [221] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, and A. Roitberg. AMBER 13, 2012.
- [222] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R. Shirts, Jeremy C. Smith, Peter M. Kasson, David van der Spoel, Berk Hess, and Erik Lindahl. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, April 2013.
- [223] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An $n \cdot \log(n)$ method for ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, June 1993.
- [224] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, October 1984.
- [225] Michael R. Shirts and John D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, 129(12):124105, September 2008.
- [226] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg,

S. Dapprich, A. D. Daniels, Ö Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian~09 revision d.01, 2009.

[227] Maestro 9.7, schrödinger, LLC, 2014.

[228] A.V. Marenich, C.J. Cramer, and D.G. Truhlar. CM5PAC, 2013.