

Washington University in St. Louis

## Washington University Open Scholarship

---

All Theses and Dissertations (ETDs)

---

January 2009

### Advances in Computational Solvation Thermodynamics

Matthew Wyczalkowski

*Washington University in St. Louis*

Follow this and additional works at: <https://openscholarship.wustl.edu/etd>

---

#### Recommended Citation

Wyczalkowski, Matthew, "Advances in Computational Solvation Thermodynamics" (2009). *All Theses and Dissertations (ETDs)*. 385.

<https://openscholarship.wustl.edu/etd/385>

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact [digital@wumail.wustl.edu](mailto:digital@wumail.wustl.edu).

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Science

Department of Biomedical Engineering

Thesis Examination Committee:

Rohit V. Pappu, Chair

Nathan A. Baker

Anders E. Carlsson

Donald L. Elbert

Lev D. Gelb

Garland R. Marshall

ADVANCES IN COMPUTATIONAL SOLVATION THERMODYNAMICS

by

Matthew A. Wyczalkowski

A dissertation presented to the Graduate School of Arts and Sciences  
of Washington University in partial fulfillment of the  
requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2009  
Saint Louis, Missouri

copyright by  
Matthew A. Wyczalkowski  
2009

## ABSTRACT OF THE DISSERTATION

Advances in Computational Solvation Thermodynamics

by

Matthew A. Wyczalkowski

Doctor of Philosophy in Biomedical Engineering

Washington University in St. Louis, 2009

Research Advisor: Professor Rohit V. Pappu

The aim of this thesis is to develop improved methods for calculating the free energy, entropy and enthalpy of solvation from molecular simulations.

Solvation thermodynamics of model compounds provides quantitative measurements used to analyze the stability of protein conformations in aqueous milieus. Solvation free energies govern the favorability of the solvation process, while entropy and enthalpy decompositions give insight into the molecular mechanisms by which the process occurs. Computationally, a coupling parameter  $\lambda$  modulates solute-solvent interactions to simulate an insertion process, and multiple lengthy simulations at a fixed  $\lambda$  value are typically required for free energy calculations to converge; entropy and enthalpy decompositions generally take 10-100 times longer.

This thesis presents three advances which accelerate the convergence of such calculations:

1. Development of entropy and enthalpy estimators which combine data from multiple simulations;
2. Optimization of  $\lambda$  schedules, or the set of parameter values associated with each simulation;
3. Validation of Hamiltonian replica exchange, a technique which swaps  $\lambda$  values between two otherwise independent simulations.

Taken together, these techniques promise to increase the accuracy and precision of free energy, entropy and enthalpy calculations. Improved estimates, in turn, can be used to investigate the validity and limits of existing solvation models and refine force field parameters, with the goal of understanding better the collapse transition and aggregation behavior of polypeptides.

# Acknowledgments

Thank you, Rohit, for your fundamental generosity and kindness in things large and small. Under your tutelage I learned of scholarship, science and humanity, and am a better person for it. My career as a scientist will be, in ways both subtle and profound, a reflection of your mentorship.

My lab mates – fellow travelers on this voyage – have challenged, informed, cheered and comforted me. Thank you Alan, Albert, Andreas, Hoang, Nick, Scott, Tim and Xiaoling. Thank you, also, to all my friends in the CCB and beyond who shared this journey with me. I will remember you always.

Of the many individuals who have graced and blessed my life, several are instrumental to this accomplishment, having lit my path by their example and propelled me with their support. Zé Ayala kindled my passion for science, validating a nascent vision and encouraging me in bold ventures. Andrew Szeri remains a guiding star in my firmament, a model of scientific elegance, dedication to students and a life deliberately lived. Linda and Fred Prior have been constant allies whose friendship, advice and aid over the years provided me with remarkable opportunities and helped me navigate life's vagaries. Thank you.

My greatest mentors, above all others, are my parents Halina and Wojtek. Whatever success I may enjoy in life is the flower of seeds they planted long ago, and nurtured ever since. To you I dedicate this work.

My family is my strength. Thank you Krzysiek and Alex, and also Lindsay, for your unabating support, generosity, love and friendship.

I would not be here today without Katie. She is my advocate and confidante, a source of inspiration and courage. Your patience, kindness and love have sustained me through difficult times, and with you I joyfully share this accomplishment. I love you dearly.

Matthew A. Wyczalkowski

*Washington University in Saint Louis*  
*December 2009*

---

This work was supported by grants MCB - 0416766 and MCB - 0718924 from the National Science Foundation.

This dissertation is dedicated to my parents  
Halina and Wojtek.

Your strength, courage and kindness  
are my bedrock and inspiration.

Jestem Wam wiecznie wdzięczny.

# Contents

<b>Abstract</b> . . . . .	<b>ii</b>
<b>Acknowledgments</b> . . . . .	<b>iv</b>
<b>List of Tables</b> . . . . .	<b>ix</b>
<b>List of Figures</b> . . . . .	<b>x</b>
<b>1 Introduction</b> . . . . .	<b>1</b>
1.1 Free Energy of Solvation . . . . .	2
1.1.1 Motivation . . . . .	2
1.1.2 Definition . . . . .	2
1.1.3 Relevance to Biology . . . . .	5
1.2 Driving Forces in Protein Folding . . . . .	6
1.2.1 The Hydrophobic Effect . . . . .	6
1.2.2 Hydrogen Bonding . . . . .	7
1.2.3 Entropy of Unfolding . . . . .	8
1.3 Solvation Thermodynamics . . . . .	9
1.3.1 Model Compound Experimental Studies . . . . .	10
1.3.2 Hydrophobic Solvation Theory . . . . .	12
1.3.3 Polar Solvation Theory and Additivity . . . . .	16
1.4 Collapse of Polar Peptides . . . . .	19
1.4.1 Why Is Water a Poor Solvent for Polyamides? . . . . .	21
1.4.2 The Promise of Simulations . . . . .	22
1.5 Synopsis of Thesis . . . . .	23
1.5.1 Summary Description of Solvation Calculations . . . . .	23
1.5.2 Advances Presented in Thesis . . . . .	24
<b>2 Background</b> . . . . .	<b>26</b>
2.1 Simulation Methodology . . . . .	27
2.1.1 Molecular Mechanics Force Fields . . . . .	27
2.1.2 Sampling Techniques . . . . .	29
2.1.3 Solvation Calculations . . . . .	32
2.1.4 Replica Exchange . . . . .	36
2.2 Free Energy Methods . . . . .	37
2.2.1 Equilibrium techniques . . . . .	38



2.2.2	Slow Growth and Nonequilibrium Techniques . . . . .	40
2.2.3	Reaction Coordinate Techniques . . . . .	42
2.3	Entropy and Enthalpy Decompositions . . . . .	44
2.3.1	The Direct Estimator and Entropy/Enthalpy Convergence . . . . .	45
2.3.2	Finite Difference Methods . . . . .	46
2.3.3	Analytical Methods . . . . .	47
2.4	Error Measurements . . . . .	47
2.4.1	Block Averaging . . . . .	48
2.4.2	Bootstrap Method . . . . .	49
2.4.3	Hysteresis Error . . . . .	50
2.4.4	Absolute Error . . . . .	51
2.5	Computational Requirements . . . . .	51
<b>3</b>	<b>Acceptance Ratio Methods for Solvation Entropy and Enthalpy Calculations . . . . .</b>	<b>53</b>
3.1	Introduction . . . . .	53
3.2	Methods . . . . .	54
3.2.1	Free Energy and Enthalpy Energy Estimators . . . . .	55
3.2.2	Calculation of Free Energies of Solvation . . . . .	59
3.2.3	Error Estimates . . . . .	68
3.3	Results . . . . .	71
3.3.1	Evaluation of Estimators . . . . .	71
3.3.2	Application to Model Compounds . . . . .	78
3.4	Discussion . . . . .	80
3.4.1	Thermodynamic Estimators . . . . .	80
3.4.2	Convergence Rates . . . . .	83
3.4.3	Swap Probabilities . . . . .	84
3.4.4	Solvation Calculations . . . . .	85
3.5	Derivations . . . . .	86
3.5.1	Definitions and General Identities . . . . .	86
3.5.2	BAR . . . . .	88
3.5.3	MBAR . . . . .	89
3.5.4	Sun model . . . . .	90
<b>4</b>	<b>Replica Exchange in Free Energy Calculations . . . . .</b>	<b>91</b>
4.1	Preamble . . . . .	91
4.2	Introduction and Overview . . . . .	92
4.3	Theory . . . . .	95
4.3.1	Background . . . . .	95
4.3.2	The Hysteresis Error . . . . .	96
4.3.3	Replica Exchange . . . . .	97
4.3.4	Swap Probability . . . . .	100
4.3.5	Swap Probability and the Hysteresis Error Convergence Rate . . . . .	102

4.4	Methods . . . . .	103
4.5	Results . . . . .	105
4.5.1	Acetamide Free Energy of Hydration . . . . .	105
4.5.2	Hysteresis Error and Replica Exchange . . . . .	106
4.5.3	Average Swap Probability . . . . .	108
4.6	Discussion . . . . .	109
4.6.1	Physical Interpretation of $C_\lambda$ Profile . . . . .	109
4.6.2	Optimal $\lambda$ Schedule for Free Energy Calculations . . . . .	110
4.6.3	Replica Exchange . . . . .	110
4.7	Summary and Conclusion . . . . .	111
4.8	Derivations . . . . .	112
4.8.1	Fluctuation Theorem Derivation . . . . .	112
4.8.2	Fluctuation Theorem and Hysteresis Error . . . . .	114
4.8.3	Inter-Replica Equilibrium and Hysteresis Error . . . . .	115
4.8.4	Linearized Average Swap Probability . . . . .	115
4.9	$V_{LJ}$ and $V_C$ Functional Forms . . . . .	118
4.10	Addendum . . . . .	119
4.10.1	Bootstrap Method Analysis . . . . .	120
4.10.2	Hysteresis Error and Replica Exchange Swap Rate . . . . .	123
<b>5</b>	<b>Conclusion . . . . .</b>	<b>125</b>
5.1	Summary of Results and Discussion . . . . .	126
5.1.1	Improved Entropy and Enthalpy Estimators . . . . .	126
5.1.2	Improved $\lambda$ Schedule . . . . .	127
5.1.3	Improved Sampling with Hamiltonian Replica Exchange . . . . .	129
5.2	Future Directions . . . . .	131
5.2.1	Sampling of Flexible Chains . . . . .	131
5.2.2	Phase Transitions . . . . .	137
5.2.3	Thermodynamic Length . . . . .	149
	<b>References . . . . .</b>	<b>153</b>
	<b>Curriculum Vitae . . . . .</b>	<b>169</b>

# List of Tables

3.1	OPLS-AA and <i>m</i> OPLS-AA force field parameters for model compounds.	63
3.2	OPLS-AA NMA results with different estimators. . . . .	74
3.3	Summary of experimental quantities for model compounds. . . . .	80
3.4	Summary of computational quantities for model compounds. . . . .	81
3.5	Details of computational quantities for model compounds. . . . .	82
4.1	The hydration free energy of acetamide. . . . .	107

# List of Figures

1.1	The molecular structure of ubiquitin. . . . .	3
1.2	A definition of the transfer process. . . . .	5
1.3	A definition of the Tanford transfer model. . . . .	17
2.1	Schematic of solvation calculations. . . . .	26
3.1	Representation of molecules used in solvation calculations. . . . .	62
3.2	The Sun model potential function. . . . .	69
3.3	Plots of Sun model errors for different block sizes and estimators. . . . .	72
3.4	$\delta F/\delta\lambda$ and $\delta U/\delta\lambda$ obtained with various estimators. . . . .	75
3.5	$\delta F$ errors and swap probabilities for different $\lambda$ schedules. . . . .	76
3.6	$\delta U$ errors and swap probabilities for different $\lambda$ schedules. . . . .	77
3.7	Comparison of thermodynamic quantities calculated with OPLS-AA vs. <i>m</i> OPLS-AA parameters. . . . .	78
3.8	Comparison of calculated vs. experimental thermodynamic quantities for model compounds. . . . .	79
4.1	A graphical representation of replica exchange. . . . .	99
4.2	Hysteresis error with and without replica exchange. . . . .	107
4.3	Average swap probability and $C_\lambda$ . . . . .	108
4.4	Water density around acetamide versus $\lambda$ . . . . .	109
4.5	Scatter plot of 43 independent $\Delta F$ and $\Delta U$ calculations. . . . .	121
4.6	Expected error calculated with bootstrap. . . . .	122
4.7	$\Delta F$ and $\Delta U$ hysteresis error versus simulation length for various replica exchange swap rates. . . . .	124
5.1	Polyglycine ensembles in poor and good solvent. . . . .	131
5.2	A thermodynamic cycle illustrating polypeptide transfer and growth processes. . . . .	133
5.3	Good-to-poor solvent transition for glycine 15-mer as solute-solute dispersions are scaled. . . . .	137
5.4	Convexity of entropy in large and small systems. . . . .	140
5.5	A plot of inverse temperature versus energy in a microscopic phase transition. . . . .	142
5.6	Observation of a phase transitions with respect to $\lambda$ . . . . .	148

# Chapter 1

## Introduction

Protein folding and binding is the *sine qua non* of biological life. The myriad tasks which proteins perform in an organism – binding and recognition of macromolecules, catalysis of reactions, conformational switching in response to stimuli, and structural support and motility – are based on the three dimensional structure of folded proteins and their specific associations with other molecules. Understanding the principles of protein folding and binding is central to comprehending the molecular basis of life.

The pioneering work of Anfinsen, for which he was awarded the 1972 Nobel Prize in Chemistry, led to the “thermodynamic hypothesis” for protein folding.

This hypothesis states that the three dimensional structure of a native protein in its normal physiological milieu . . . is the one in which the Gibbs free energy of the whole system is lowest; that is, that the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment (Anfinsen, 1973).

The free energy of a protein’s conformation in an aqueous milieu – the focus of our interest – is then the fundamental determinant of its structure. The entropic and enthalpic constituents of that free energy, furthermore, are the fingerprints of the underlying molecular processes.

The central focus of this thesis work is the development of improved techniques for calculating free energies of solvation, and their entropy and enthalpy components, based on computational molecular simulations.

## 1.1 Free Energy of Solvation

### 1.1.1 Motivation

Proteins fold spontaneously to their native states. All the information necessary for folding is encoded in the amino acid sequence and consequently, in the appropriate milieu, a protein folds because the free energy of the protein and solvent system is minimized upon folding. If the solution conditions are altered, then a protein will unfold if the unfolded state minimizes the free energy of the entire system. Therefore, in protein folding, the relevant quantity is the free energy of folding,  $\Delta G_{fold}^{\circ}$ , defined as the difference in the standard state partial molar Gibbs free energies of the folded and unfolded states. For a given set of solution conditions, the folded state is the preferred state when  $\Delta G_{fold}^{\circ} < 0$ , and  $\Delta G_{fold}^{\circ}$  can change in magnitude and sign as solution conditions vary.

In the folded state, hydrophobic groups are shielded from the aqueous milieu, resulting in compact structures with hydrophobic interiors and hydrophilic (polar or charged) groups on the surface (see Fig. 1.1). Hence, protein folding is characterized by the transfer of specific groups from an aqueous milieu to a different microenvironment (a “greasy” interior in the case of hydrophobic groups). This transfer process is connected fundamentally to the free energy of solvation, a quantity which will be defined below.

### 1.1.2 Definition

Ben-Naim (1987) has provided a formal definition for the solvation process. According to this definition, solvation refers to the “transfer of a solute molecule from a fixed position in the ideal gas phase into a fixed position in the liquid phase, and the process is carried out at fixed temperature and pressure.” In this definition, the term solvation conceptually refers to a measure of the interaction between the solute and its surroundings. With constant temperature  $T$  and pressure  $P$  as thermodynamic constraints, and subscripts  $s$  and  $l$  referring to solute and solvent, respectively, the

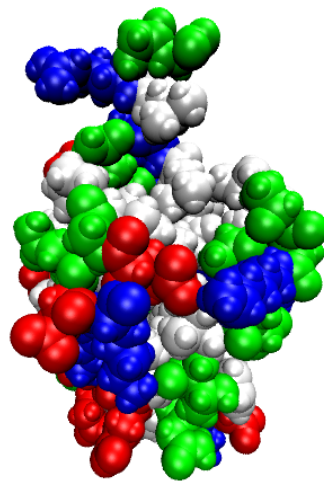


Figure 1.1: The molecular structure of ubiquitin, a globular protein (PDB code 1D3Z). Hydrophobic residues, in white, tend to partition to the interior of the protein, while the colored hydrophilic residues are on the exterior (with polar, basic and acidic residues in green, blue and red, respectively). Visualization with VMD (Humphrey et al., 1996).

Gibbs free energy of solvation is written as,

$$\Delta G_s = \Delta G_s^* + k_B T \ln(q\rho_s\Lambda^3). \quad (1.1)$$

Here,  $\Delta G_s^*$  is the pseudochemical potential which captures the free energy of transfer of an immobile solute from the gas phase to the solvated liquid phase. It is defined as,

$$\Delta G_s^* = G^*(T, P, N_l + 1) - G(T, P, N_l) \quad (1.2)$$

where  $N_l$  is the number of solvent molecules and the asterisk (\*) indicates that the position of the solute is fixed. The pseudochemical potential captures the interactions between the solute and the solvent, and may be calculated on the basis of such interactions as,

$$\Delta G_s^* = -k_B T \ln\langle \exp\{-[V_{N_l+1}(\Gamma_l, \Gamma_s^*) - V_{N_l}(\Gamma_l)]/(k_B T)\} \rangle. \quad (1.3)$$

where the ensemble average is taken over all possible conformations ( $\Gamma_l$ ) of the  $N_l$  solute molecules while the solvent configuration ( $\Gamma_s^*$ ) is held fixed.  $V_{N_l+1}$  and  $V_{N_l}$  are the internal energies of the system at a given configuration with and without the solute molecule, respectively. Their difference is the binding energy of the solute with the solvent, or equivalently, the work required to insert the solute molecule.

The second term in Eq. (1.1) is the liberation free energy, which captures the contribution from releasing the solute position restraint.  $k_B$  is the Boltzmann constant,  $\rho_s$  is the number density of solute  $s$  in the liquid,  $\Lambda$  is the momentum partition function and  $q$  is the internal partition function (Ben-Naim and Marcus, 1984; Ben-Naim, 1987). It is discussed further in chapter 3.

Eq. (1.3) forms the operational basis of evaluating solvation free energies computationally, and is effectively the free energy perturbation technique which will be discussed in chapters 2, 3 and 4.



### 1.1.3 Relevance to Biology

Driving forces for a range of phenomena are tied to the precise values of  $\Delta G_s$  at the appropriate physiological conditions. Such phenomena include (i) protein folding – the transfer of a protein from a solvent-exposed macrostate (unfolded) to a folded state characterized by the partitioning of some groups into a solvent protected interior; (ii) transfer of ions from water into the interior of an ion channel; (iii) transfer of a protein from an aqueous milieu into a hydrophobic lipid environment (membrane protein folding); (iv) self-association of proteins and aggregation/phase separation; and (v) binding, where a ligand and protein lose interactions with the surroundings and gain interactions with each other.

Such processes are all governed by transfer free energies from one solvent into another, and are related to solvation free energies by the relationship (see Fig. 1.2),

$$\Delta G_t = \Delta G_s^{(2)} - \Delta G_s^{(1)} \quad (1.4)$$

with  $\Delta G_s^{(i)}$  the free energy of solvation in solvent  $i$ .  $\Delta G_t$  is then the transfer free energy from solvent 1 to solvent 2.

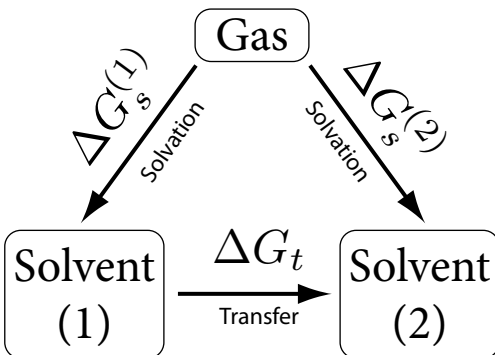


Figure 1.2: The transfer process of a solute from solvent 1 to solvent 2 can be thermodynamically decomposed into a pair of gas→solvent solvation processes.

## 1.2 Driving Forces in Protein Folding

As described in section 1.1,  $\Delta G_t$  and the relevant  $\Delta G_s^{(i)}$  values provide the quantitative basis for understanding the stability and spontaneity for a range of self-assembly and molecular recognition processes involving proteins. In this section, we review our understanding of the driving forces for protein folding from the perspective of solvation thermodynamics.

The conformations of a protein in solution are governed by a number of forces, including electrostatic, hydrogen bond and van der Waals forces, as well as the hydrophobic effect and conformational and solvent entropy (Dill, 1990).  $\Delta G_{fold}^\circ$  is governed by a balance of large opposing forces, with three forces dominant: the hydrophobic effect tends to stabilize compact states; chain entropy favors swollen conformations; and hydrogen bonding stabilizes specific native conformations, although its overall contribution to the relative stability of compact and swollen conformations is uncertain, and is the subject of ongoing debate (Pace et al., 1996). As a consequence of this balance, the stability of native structures of proteins is typically marginal, with small perturbations in solvent conditions or sequence able to tip the balance and lead to denaturation (Bolen and Rose, 2008).

### 1.2.1 The Hydrophobic Effect

The free energy change associated with the solvation of a nonpolar molecule such as a hydrocarbon is positive (Tanford, 1973). This is intuitively understood from the observation that water and oil do not mix and remain in separate phases, although mixing does occur at very low concentrations. The magnitude and nature of this driving force can be quantified by measuring the transfer free energy of a hydrophobic model compound from a nonpolar solvent into water (Baldwin, 2005). Alternatively,  $\Delta G_s$  can be measured directly by determining the vapor pressure of the compound over dilute aqueous solutions, for instance, from the equilibrium constant for the transfer of the compound from water to the vapor phase may be calculated (Wolfenden et al., 1981). The hydrophobic effect provides a driving force for the collapse of hydrophobic polypeptides, favoring compact, globular structures which

minimize protein interfaces with water. The nature and basis of the hydrophobic effect will be discussed in greater detail in section 1.3.

## 1.2.2 Hydrogen Bonding

Hydrogen bonds are directional interactions involving atoms of different electronegativity. The hydrogen ( $H$ ) is covalently attached to a donor atom ( $D$ ), with the hydrogen bond ( $\cdots$ ) forming between this pair and an acceptor ( $A$ ),  $D - H \cdots A$ , an effect which may be modeled as a dipole-charge interaction. Donor-acceptor distances range between  $2.7\text{\AA}$  and  $3.2\text{\AA}$ , and the interaction strength depends on the angle  $\angle_{DHA}$ , leading to the linearity of hydrogen bonds. It has been proposed that quantum mechanical effects such as charge transfer and classical, higher order electronic polarization also play an important role in determining the strengths of hydrogen bonds. Gas phase calculations show that the strength of multiple hydrogen bonds is greater than the sum of the bonds individually, implying a cooperativity associated with hydrogen bonding (Jeffrey, 2003; Ponder and Case, 2003).

Hydrogen bonds play a central role in determining the properties of water and ice, such as the large heat capacity of liquid water, its strong surface tension and broad temperature range in which it is liquid. These properties are the result of the open tetrahedral structure which derives from intermolecular hydrogen bonding (Dill and Bromberg, 2003).

Hydrogen bonding also occurs in proteins. The polypeptide backbone contains both hydrogen bond donor and acceptor groups, and various side chains contain donors and/or acceptors as well. As a result, in a protein system, the backbone, polar side chains and water can all participate in hydrogen bonding. However, the effect of hydrogen bonding on the stability of a collapsed polypeptide remains unclear, and remains mired in controversy (Pace, 2009; Dill, 1990; Bolen and Rose, 2008). The  $\alpha$  helix (Pauling et al., 1951) and  $\beta$  sheets (Pauling and Corey, 1951) as originally constructed were understood to be stabilized by intramolecular hydrogen bonds. Soon after, however, Kauzmann argued that hydrogen bonding, while certainly important, cannot be the dominant driving force for the stabilization of the folded state of the protein, since there was no basis for believing that the intrachain hydrogen bonds in

the folded state would have a lower free energy than those of the unfolded chain to water (Dill, 1990).

Much effort has been made to quantify the energetic difference between a hydrogen bond to water in the unfolded chain as compared to an intramolecular hydrogen bond in a folded chain (Stickle et al., 1992; Pace, 2009). In order to evaluate this difference, a number of estimates must be made, including (a) the enthalpy of hydrogen bond formation in water and in the protein, (b) desolvation penalty in transferring polypeptide from aqueous to protein environment, (c) changes in conformational entropy of both polypeptide chain and water molecules, (d) conformational dependence of intramolecular hydrogen bond strength and (e) hydrogen bond cooperativity (Bolen and Rose, 2008). None of these estimates are straightforward, and all are subject to considerable uncertainty. Dill (1990) argues that the overall contribution of hydrogen bonding to the stability of a protein is negligible. More recently, however, consensus appears to be moving toward recognizing the importance of hydrogen bonding in maintaining protein stability. Pace et al. (1996) argue that hydrogen bonding and the hydrophobic effect make comparable contributions to the stability of globular proteins. Bolen and Rose (2008) conclude that the energetically favorable hydrogen bond stabilizes compact polypeptide structures. Recently, Gao et al. (2009) found that the energetics of hydrogen bonds are sequence dependent, and can be up to 1.2 kcal/mol stronger in a hydrophobic environment, thus contributing significantly to the stability of a protein native state.

### 1.2.3 Entropy of Unfolding

There are fewer ways for a protein to be collapsed than swollen, and fewer yet to be in the native state; as a result, absent the other driving forces, a protein is overwhelmingly more likely to assume swollen conformations (Dill, 1990). An additional contribution is the entropy change of the solvent water upon protein unfolding. Whereas the entropy of the chain increases when it swells, interactions of the solvent with the newly exposed residues – both polar and hydrophobic – decrease solvent entropy. On balance, conformational entropy dominates and drives the chain to swell (Makhatadze, 2005).

Quantifying such entropy changes is challenging, although consideration of the individual components of the system allows some progress to be made. Simple models of chain entropy based on excluded volume arguments provide a useful starting point for theoretical estimates (Dill, 1990). Pursuing the problem in more detail, the protein entropy can be divided into backbone, side chain, and rotation and translational components (Amzel, 2000). Upon collapse, both the backbone and side chains experience a loss of conformers accessible to each dihedral bond, as well as a reduction in the magnitude of their fluctuations. The change in solvent entropy can be estimated from the change in solvent exposed polar and hydrophobic areas (Lee et al., 1994). Based on such arguments, the entropy loss upon collapse and folding can be obtained.

Alternatively, estimates of the entropy change upon unfolding can be based on experimental measurements of heat capacity changes, together with assumptions about the effect that burial of hydrophobic and polar polypeptide groups has on the heat capacity (Pace et al., 1996). Several different estimates of the magnitude of the entropic penalty associated with collapse have been made. The actual value for a given protein will likely be context specific, and dependent on the conformations populated in the unfolded ensemble (Makhatadze, 2005).

### 1.3 Solvation Thermodynamics

Instead of evaluating the stability of a given protein conformation by considering the total sum of forces acting on it, an alternative approach is to estimate the transfer free energy change associated with removing a protein group – be it a side chain or the entire residue – from the aqueous solvent and “burying” it in the interior of the protein (Baldwin, 2005). The relative affinity with which polypeptide units interact with water, as compared to their interaction with the rest of the polypeptide, governs the stability of given polypeptide conformations. If such a transfer is favorable, the polypeptide will tend adopt a compact structure, whereas unfavorable transfer free energies stabilize the swollen protein conformation.

### 1.3.1 Model Compound Experimental Studies

Our understanding of the solvation process, and in particular of the hydrophobic effect, is based on the thermodynamic analysis of experiments of the solvation of model compounds. These solvation free energies, and their entropy/enthalpy decompositions, inform the construction of solvation theories. Here we present the basic results of such experiments and their interpretations.

#### Hydrophobic Model Compounds

The Gibbs free energy of solvation,  $\Delta G_s$ , can be decomposed into the solvation enthalpy  $\Delta H_s$  and entropy  $\Delta S_s$  components as,

$$\Delta G_s = \Delta H_s - T\Delta S_s, \quad (1.5)$$

where  $T$  is the temperature. With  $\Delta G_s$  measured at a variety of temperatures, the enthalpy of solvation may be calculated from experiment as

$$\Delta H_s = d[\Delta G_s/T]/d[1/T]. \quad (1.6)$$

The following observations are made from experimental data of the solvation of ethane, a nonpolar solute, in water at and around room temperature (Tanford, 1973):

- Since  $\Delta G_s > 0$ , the solvation process is unfavorable.
- Further,  $\Delta G_s$  decreases with increasing temperatures. According to Eq. (1.6), this implies that  $\Delta H_s < 0$ . That is, the solvation process is energetically favored, but only by a modest amount.
- Given the prior two observations, the solvation of non-polar model compounds must be entropically unfavorable:  $\Delta S_s < 0$ .
- The heat capacity of the system increases upon the addition of a solute, a phenomenon known as anomalous heat capacity.

These observations hold at vanishingly small solute concentrations, and so cannot be explained by solute-solute interactions (such as clustering). Rather, they must be interpreted as the result of changes in the state of water molecules induced by the presence of hydrocarbons.

Since nonpolar compounds cannot form hydrogen bonds, they must perturb the water hydrogen bond network. Water hydrogen bonds are not simply severed, however, as that would require the enthalpy of the system to increase, whereas it is observed to slightly decrease. Given that the free energy of the system increases, its entropy must significantly decrease upon hydrophobic solvation. Apparently, the waters restructure and lose degrees of freedom in the vicinity of the solute (Tanford, 1973).

This leads to the idea of clathrates or “icebergs” of water around hydrophobic solutes, originally proposed by Frank and Evans (1945). That view, however, is inconsistent with the anomalous heat capacity. Rather, it appears that there are competing arrangements of water. One model consistent with these observations postulates two types of water phases in the vicinity of the solute: one phase has high entropy and enthalpy, the other low entropy and enthalpy. Both have the same free energy, being in equilibrium. As the temperature increases the low entropy/enthalpy phase melts out, and in so doing leads to the observed heat capacity anomalies (Gill et al., 1985). The two-state model is not free of controversy, and other models have been proposed (Sharp and Madan, 1997).

## **Polar Model Compounds**

For polar compounds,  $\Delta S_s$  is likewise negative, but a much more negative  $\Delta H_s$  makes the solvation process favorable. Also, the temperature dependence of  $\Delta H_s$  is of the opposite direction, with  $\Delta C_P < 0$  (Sharp and Madan, 1997). This suggests that unlike hydrophobic solvation, where the waters around the solute are more easily displaced than in bulk, around a hydrophilic compound the waters are less labile and more restrained (Tanford, 1973). Instead of fluctuating phases, the solvation shell waters likely participate in hydrogen bonding with the solute and are less mobile.

The free energy associated with the solvation of a model compound defines the favorability of the process. Its decomposition into entropy and enthalpy, moreover,

gives insight into, and imposes constraints upon, molecular models of the mechanism. Theories of solvation are based upon these results.

### 1.3.2 Hydrophobic Solvation Theory

The generalized descriptions of the solvation process presented so far are helpful in illustrating the mechanism of solvation, but are of limited use in predicting the solvation free energies of arbitrary solutes. To this end, one needs a more detailed and quantitative description of the solvation mechanism. As a first conceptual step, the solvation process is divided into two legs: first a cavity of the size and shape of the solute is created in the solvent, and then the solute is inserted so that it interacts with the solvent (Lee, 1995). The first step can be understood in terms of the fluctuations of the solvent alone. The second step involves van der Waals interactions, hydrogen bonding and electrostatic interactions, and this step is in general much more difficult to model accurately. The various solvation theories, differ in: 1) the nature of the model used for the solvent, and 2) the models for interactions between the solute and the solvent in the second step of the solvation process.

#### Hard sphere solutes

In hard sphere models of solvation, the solute and solvent particles both interact like billiard balls, with infinitely strong repulsion at distances smaller than the sum of their radii, and no attraction or repulsion at larger distances. For such systems the free energy associated with solvation is entirely entropic and related to the probability of spontaneous fluctuations making a cavity the size and shape of the solute. Two types of theories are frequently used to predict these free energies. The first, scaled particle theory (SPT) (Reiss et al., 1959; Postma et al., 1982; Heying and Corti, 2004), uses strictly geometric arguments based on the size, shape and bulk density of the solute. From these the theory makes deductions about the average density of the fluid some distance away from the cavity, as well as the work required to change the size of the cavity. These are then used to make predictions about the thermodynamic properties of the solute as well as solute density radial distribution functions.



Hard sphere models, while not entirely realistic, are nevertheless a useful and important starting point for understanding solvation phenomena. They predict the entropic penalty associated with the formation of a cavity in a solvent, a process common to all types of solvation, as well as the magnitude of the hydrophobic effect between two hard sphere particles (Berne, 1996).

Information theory (Hummer et al., 1998) is a generalization of the SPT model described above relates the probability of finding a molecule-sized cavity to the solvation free energy. Based on computer simulations of liquid water, the probability of finding  $n$  solvent particles in a solute-shaped cavity is evaluated. Maximum entropy methods are then used to obtain the most likely probability of finding a void of a given size, which leads to an estimate of the free energy of solvation. Unlike scaled particle theory, where both the solute and solvent are hard sphere particles, information theory utilizes molecular models of the solvent to make predictions about solvation properties of hard sphere solutes. They utilize the properties of bulk water alone, and allow comparison to other, nonpolar solvents in order to understand how liquid water responds to the presence of small spherical cavities.

### **Cavity Size Effects**

The discussion thus far has focused implicitly on relatively small solutes (3-5 Å). As first described by Stillinger (1973) using scaled particle theory, the behavior of water around a small cavity differs fundamentally from its behavior next to a hydrophobic wall. Whereas the water density immediately around a small solute is large, a vapor barrier with greatly decreased water density is predicted to form adjacent to the wall. Small cavities and walls are, in effect, the limiting cases of a family of spherical cavity sizes, from water molecule-sized cavities to cavities of infinite size (whose boundary is a planar wall). As the cavity size increases, presumably there is a transition from the small cavity behavior, with increased water density at its boundary, to large cavity behavior, where a vapor barrier separates the cavity from bulk water.

This observation has been formalized by the Lum-Chandler-Weeks theory of solvation (Lum et al., 1999), which predicts the transition from the small to large solute

regime. For small cavities the probability distribution for finding a spherical observation volume with  $n$  solvent molecules (as analyzed in the information theory model) is Gaussian with respect to  $n$ , and the solvation free energy of a hard sphere solute is dominated by entropic effects (ten Wolde, 2002). For larger cavities, however, this distribution is no longer Gaussian. As cavity size increases, the water hydrogen bond network ruptures, close proximity of water to the void becomes energetically unfavorable, and a vapor barrier is formed immediately next to the cavity. This drying transition is a collective effect and can be interpreted as a microscopic manifestation of a phase transition (ten Wolde, 2002). For such large cavities, the free energy of solvation is enthalpically dominated. The transition from the small to large solute regime occurs for spherical cavity radii on the order of 1 nm (Chandler, 2005); below this size, the solvation free energy is proportional to the volume of the cavity, whereas for larger sizes, it is proportional to the surface area.

### **Accessible Surface Area models**

The free energy of solvation for linear alkanes, which are hydrophobic, is found to scale linearly with their accessible surface area (Sharp et al., 1991). There is some debate as to whether such scaling is proportional to surface area or to volume, with some experiments suggesting that a linear relationship with volume more is appropriate (Baldwin, 2005). As discussed, some hydrophobic theories predict that volume scaling holds for small volumes and surface area scaling for large ones (Chandler, 2005). For solvation of polar model compounds, however, the linear relationship between solute size (whether surface area or volume) does not hold, and two compounds with very different surface areas can have nearly equal polar contributions to the solvation free energy (Baldwin, 2005).

In the protein folding field, accessible surface area (ASA) models, which quantify free energy changes of protein conformations based on changes in the protein surface area accessible to water (Lee and Richards, 1971) are widespread. When a hydrophobic side chain is transferred from the surface of a protein into its interior, hydrophobic interactions with the surface waters are lost. The hydrophobicity of side chains is found to scale linearly with the accessible surface area (Chothia, 1974), so the change

in this area provides a measure of the hydrophobic free energy change upon protein collapse.

Polar residues also interact with water through the hydrophobic effect, in addition to hydrogen bonding. It is assumed that upon burial of a polar group the hydrogen bonds lost with water are replaced by hydrogen bonds with the protein backbone or side chains, and so is energetically neutral (Chothia, 1974). As a result, the hydrophobic energy contribution is proportional to the ASA change for all components of the protein, and may be used in the analysis of protein stability as well as binding (Jones and Thornton, 1996; Halperin et al., 2002).

### **Collapse of Hydrophobic Chains**

Liquid water at ambient conditions is argued by some researchers to lie close to phase coexistence with vapor (Chandler, 2005), and transient vapor cavities spontaneously flit into and out of existence. Continuum models which take such density fluctuations into account can reproduce the hydrophobic interaction between two hydrophobic hard spheres (Willard and Chandler, 2008).

This phenomenon is also believed to be responsible for the collapse of chains composed of small hard spheres (ten Wolde and Chandler, 2002). Water interactions around the extended conformations of such chains are in the small solute regime, making the conformations transiently stable. Collapse begins only when a vapor bubble forms spontaneously in the vicinity of the chain. A kink may then form, creating a critical nucleus which transitions the chain from the small to the large solute regime. Now, instead of being volume dependent (and hence conformation independent), the free energy of solvation is proportional to the surface area; as a result, the chain collapses further in order to minimize the solvent interface. Hydrophobic collapse is then driven by the transition from the small to large solute regimes. Whether this model, confirmed by computer simulations for hard sphere models, holds for solutes with dispersive (attractive) interactions is subject to debate (Athawale et al., 2007; Ashbaugh and Paulaitis, 2001).

### 1.3.3 Polar Solvation Theory and Additivity

The theory of hydrophobic solvation, as discussed previously, is relatively well developed from a theoretical perspective and the predictions made by such models correlate well with experimental results. The same cannot be said for the solvation mechanisms for polar compounds. Here, the hydrogen bond plays a dominant role and its effect on the solvation process is significant and complicated.

Unlike hydrophobic solvation, where the solute is well approximated as a cavity which does not interact with the water, polar solutes participate in and affect the hydrogen bond network of liquid water. In general, such interactions tend to be favorable enthalpically, and polar compounds tend to be soluble as well (hence, hydrophilic). The central approximation of hydrophobic theories – that solutes are simply holes in water – simply does not hold. As a result, theories of polar solvation tend to be more heuristic and empirical.

#### The Tanford Transfer Model

Transfer models utilize the free energy changes associated with transferring a compound – typically a model of the protein backbone or side chain – from water to a liquid which mimics the conditions in the interior of the protein. This liquid-liquid transfer free energy is in turn obtained from  $\Delta G_s$  for model compounds (Baldwin, 2005).

Figure 1.3 illustrates the Tanford transfer model formalism (Aune and Tanford, 1969), used to quantify the change in the stability of a protein upon transfer to a chemical denaturant; the same theory also applies to the process of transferring a protein (or its constituent groups) from water into a nonpolar solvent approximating the interior of a protein. Based on the thermodynamic cycle illustrated in Fig. 1.3, where the free energy change is independent of the path, we can write,

$$\Delta G_D^\circ + \Delta G_t^{(N)} = \Delta G_{H_2O}^\circ + \Delta G_t^{(D)}. \quad (1.7)$$

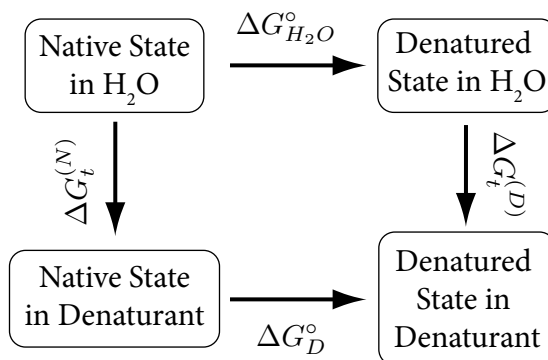


Figure 1.3: The change in stability of the native state of protein upon transferring from an aqueous solution to denaturant,  $\Delta G_D^\circ - \Delta G_{H_2O}^\circ$ , is given by the difference in transfer free energies,  $\Delta G_t^{(D)} - \Delta G_t^{(N)}$ . By the Tanford transfer model, this quantity can be approximated by the sum of the individual group transfer free energies.

Rearranging, we obtain an expression for the change in the free energy of folding upon transfer from water into the denaturant,

$$\Delta G_D^\circ - \Delta G_{H_2O}^\circ = \Delta G_t^{(D)} - \Delta G_t^{(N)}. \quad (1.8)$$

From experimental evidence it is observed (Auton and Bolen, 2004; Bolen and Rose, 2008) that the difference in the transfer free energies between the aqueous and the denaturing solution is proportional to the number of groups newly exposed upon denaturation and the transfer free energies of these groups into the denaturing solvent. Thus, Eq. (1.8) can be rewritten as,

$$\Delta G_D^\circ - \Delta G_{H_2O}^\circ = \sum_i^M n_i \alpha_i \delta g_{t,i}. \quad (1.9)$$

Here  $M$  is the number of distinct group types,  $n_i$  is the number of groups of type  $i$  in the polypeptide, and  $\delta g_{t,i}$  is the free energy of transfer of a group  $i$  from water to a given denaturing solution. Finally,  $\alpha_i$  is the fraction of groups of type  $i$  that are exposed in the denatured state and not in the native state; it accounts for groups which are exposed to solvent in the folded conformation that do not change their solvent exposure.

In summary, the Tanford transfer model, Eq. (1.9), predicts the change in the stability of the folded state of a protein as it is transferred from an aqueous to a denaturing (or hydrophobic) solvent, based on the individual group transfer free energies which are derived from model compound solvation studies.

### **Limits of additivity**

In spite of the utility of transfer models and the widespread use of ASA techniques, there are concerns about the general soundness of such additivity methods. A protein is not just the sum of transfers of small molecule side chains, but a polymer (Dill, 1990). Aside from conformational entropy, a driving force which cannot be captured at all by model compounds, the view that free energies, entropies and enthalpies of solvation are group-wise additive is not a fundamental principle but an empirical approximation, valid only when it is demonstrated to be so, but at times erroneous (Dill, 1997).

For instance, chain termini are much more exposed to solvent water than polypeptide groups in the central portions of a chain (Bolen and Rose, 2008) and this modulation of the local environment will presumably influence solvation thermodynamics. In fact, free energies of transfer (from water to 1 M urea) of short segments of the polypeptide backbone are additive only if the contributions of the end groups are considered separately (Auton and Bolen, 2004), and a linear relationship between chain length and transfer free energy holds for chains past a given length.

The general validity of using model compounds to measure the partitioning of side chains has also been called into question (Roseman, 1988). This is because the polar polypeptide group – typically absent in a side chain model partition experiment – modifies the local aqueous environment. As a result, the hydrophobicity of the polar group is markedly increased, and the transfer of the solute into water becomes much less favorable than is predicted from the transfers of the individual components. In addition, a phenomenon known as the proximity effect increases the hydrophobicity of polar functional groups when they are separated by one or two carbon atoms, again suggesting limits to simple additivity models (Roseman, 1988).

In short, extrapolation of thermodynamic quantities from model compounds must be done with care; additivity principles are heuristic and not supported by any rigorous theory. On the other hand, statistical thermodynamics does not suffer from such limitations (Dill, 1997), and models of solvation based on this theory, along with molecular dynamics simulations, may offer a sound way to evaluate the validity and limits of additivity models.

## 1.4 Collapse of Polar Peptides

The preceding section detailed the limits of additivity assumptions based on formal considerations (Dill, 1997) and empirical observations (Roseman, 1988). Recent investigations which consider the conformational equilibria of polypeptide systems rich in polar amino acids provide a direct test of the validity of the transfer model and its implicit additivity assumptions.

The notion of solvent quality from the field of polymer physics is helpful in understanding and categorizing disordered states of proteins. A polymer chain which interacts more favorably with itself than with the solvent will collapse, and is said to be in a poor solvent. Conversely, if chain-solvent interactions are preferred, the system is in a good solvent regime. These two states can be characterized by considering how the average radius of gyration  $\langle R_g \rangle$  of the polymer scales with chain length  $N$ . For long enough chains the relationship  $\langle R_g \rangle \propto N^\nu$  is found to generally hold, with the parameter  $\nu$  indicating solvent quality. For poor solvent  $\nu = 0.33$ , whereas for a good solvent  $\nu = 0.59$  (Rubinstein and Colby, 2003).

Intrinsically disordered proteins (IDPs) are functional proteins that do not fold into well-defined, unique three-dimensional structures under physiological conditions (Fink, 2005). IDPs are ubiquitous *in vivo* and their intrinsic disorder is implicated in a range of regulatory functions, such as signaling, molecular switching, protein trafficking, and protein turnover (Wright and Dyson, 1999; Dunker et al., 2001, 2002a,b; Uversky, 2002; Dyson and Wright, 2005). Typical IDP sequences have a combination of low overall hydrophobicity, high mean net charge (Uversky et al., 2000), and in some cases, low sequence complexity (Sim and Creamer, 2002; Weathers et al., 2007).

Uversky et al. (2000) argued that low overall hydrophobicity of IDPs must imply the lack of a driving force for formation of ensembles with compact structures, suggesting that IDPs in water will behave as chains in a good solvent. These predictions have been questioned by recent spectroscopic studies which characterize the conformational ensembles for sequences that have many of the traits of IDPs, including low hydrophobicity and low sequence complexity. Experiments on polyglutamine (Crick et al., 2006) and glycine-serine block copolypeptides (Möglich et al., 2006) show that, contrary to expectation, such polypeptides prefer to form collapsed structures in aqueous solutions. Mukhopadhyay et al. (2007) obtained similar results for the glutamine / asparagine rich N-terminal domain of the yeast prion protein Sup35. Tran et al. (2008) then asked if these observations were attributable, at least partially, to intrinsic preferences of polypeptide backbones in water. Their studies, based on molecular simulation data, showed that polyglycine chains spontaneously form collapsed structures in water. These results are inconsistent with the predictions of the transfer model, as explained below.

The free energy of solvation for N-methylacetamide (NMA), a model compound mimic of the polypeptide unit, is about -10 kcal/mol at 298 K (Wolfenden, 1978) (see also chapter 3). Extrapolation from the transfer free energy model suggests that polyglycine – essentially, a concatenation of repeating NMA units – should prefer structures that maximize the interface with the aqueous solvent. That is, according to this theory, water should be a good solvent for this generic polypeptide backbone.

The simulations of Tran et al. (2008) suggest that water behaves like a poor solvent for polyglycine, despite the fact that NMA and other secondary amides are highly miscible as small molecules in water. Clearly, the transfer model does not anticipate the correlations imposed by chain connectivity, which in turn leads to liquid-liquid demixing that causes the polyamides to collapse on themselves to facilitate the formation of water deficient collapsed structures. While the repeating units are miscible in water across a wide concentration range, the polyamides become immiscible as chain length increases. The accumulated experimental and molecular simulation data suggest that polymers of polar and hence hydrophilic model compounds behave like polymers of hydrophobic model compounds. Concatenation of polar moieties introduces interactions and effects on new length scales that are only realizable in the polymeric forms, suggesting a new twist to the hydrophobic effect that defies explanation via simple



extrapolations of the transfer model. These observations have precedent in the polymer solubility literature and the open question that remains unanswered pertains to the driving force for collapse transitions of polymers of polar molecules in aqueous milieus.

### 1.4.1 Why Is Water a Poor Solvent for Polyamides?

Tran et al. (2008) proposed several hypotheses to explain their observations regarding the collapse of polyglycine in water and the reversal of this preference for collapsed states in 8 M urea. These hypotheses fall into three categories.

1. Collapse is the direct result of amide-amide and water-water interactions being preferred to amide water interactions. If this is true, then weakening the intra-solute hydrogen bonds without perturbing the model compound hydrophobicity, the model compound van der Waals interactions, and only weak perturbation of the solvent-model compound hydrogen bonds should promote the preference for more swollen states. This perturbation can be achieved through systematic replacement of amides with esters, an exercise that has been carried out in the Pappu lab (unpublished data). The results demonstrate that the constructs with amide-to-ester substitutions as well as polyesters remain collapsed, with greater stabilization of the collapsed states. This suggests that the preference of amide-amide hydrogen bonds, while providing a seemingly simple explanation, cannot be the sole reason for the observations regarding polyamides.

2. It is conceivable that the free energy of solvation per amide becomes less favorable as the lengths of polyamides increase, suggesting a negative cooperativity in the solvation process. Unfortunately, this proposal has been difficult to test pending the availability of robust methods for calculating free energies of solvation in the context of long polymers in aqueous milieus. Preliminary calculations carried out in the Pappu lab (unpublished data) for polyglycine N-mers up to a  $N = 3$  suggest a very weak negative cooperativity in the  $\Delta G_s$  per amide.

**3.** A linear relationship between  $\Delta G_s$  and  $N$  may still mask the details of the solvation process. A finer dissection of the preceding proposal comes from parsing the data for NMA into the enthalpy and entropy of solvation. The favorable free energy of hydration ( $\Delta G_s \simeq -10$  kcal/mol) at 25°C for NMA is the result of a balance between highly favorable enthalpy ( $\Delta H_s \simeq -20$  kcal/mol) and negative entropy ( $T\Delta S_s \simeq -10$  kcal/mol) (Makhatadze et al., 1997). The large negative entropy offsets at least half the favorable enthalpy. Graziano has proposed that this “negentropic” term derives mainly from the excluded volume penalty associated with creation of a solute-sized cavity in water (Graziano, 2000). One working hypothesis is that the negentropic term becomes increasingly unfavorable for hydration of long, intrinsically flexible chains, and the work done to create solute-sized cavities for expanded conformations will also be significant. The entropic penalties associated with cavitation for this heterogeneous ensemble of swollen conformations might increase nonlinearly with chain length. Consequently, one can postulate that longer chains collapse to minimize the entropic penalties of solvent organization around swollen, loosely packed conformations.

### 1.4.2 The Promise of Simulations

In the preceding discussion we presented examples calling into question the validity of the additivity assumptions underlying the transfer model. The hypotheses presented above require the computation of free energies of solvation and their entropy-enthalpy decompositions for flexible polymers. Such measurements, inaccessible to experimental probes, are, in principle, directly accessible with computer simulations.

Given the fundamental importance of solvation free energy calculations in testing the validity of additivity assumptions and solvation models in general, the work presented here focuses on issues related to the efficiency and accuracy of such calculations and their entropy/enthalpy decomposition. The overall goal is to enable calculations that can provide precise answers through computational tests of specific hypotheses. Increasing the precision and accuracy of such calculations will enable us to address the nature of the driving forces for the collapse of polar tracts as well as to understand how this new manifestation of the hydrophobic effect fits into the general framework

for the hydrophobic effect being pursued by other research groups (Wagoner and Baker, 2006).

## 1.5 Synopsis of Thesis

The focus of this thesis is on the methodological aspects of computational thermodynamics. The overall goal is to improve the accuracy and precision of the thermodynamic quantities – the free energy, entropy and enthalpy – obtained from computer simulations of the solvation process.

While the motivation for such studies, as described in detail in the previous section, is to understand the thermodynamic stability of protein conformations, the work here will focus on the solvation of small model compounds. The reasons for this are threefold. First, model compounds are important conceptual stepping stones for understanding the solvation process and thermodynamic stabilities of proteins. Also, model compound solvation calculations are important in refining force field parameters; being able to reproduce both the experimental free energies as well as their decompositions lends confidence to the general accuracy of simulations. Finally, techniques developed here can be applied directly to protein systems. Larger flexible molecules are considerably more difficult to sample accurately, but there are no apparent conceptual difficulties in applying these techniques to protein systems.

### 1.5.1 Summary Description of Solvation Calculations

In order to contextualize the main contributions of this thesis, we outline here the basic concepts behind solvation calculations. These will be fleshed out in chapter 2.

Solvation calculations as employed here consist of multiple independent equilibrium molecular simulations. All such simulations have the same constituents: one solute molecule bathed by a relatively large number of solvent water molecules. A coupling parameter  $\lambda$  governs the interactions between the solute and the solvent: when  $\lambda = 0$  the solute and solvent do not interact, and the simulation is effectively of the neat solvent phase. With  $\lambda = 1$  the solute and solvent interact fully.  $\lambda$  can assume

intermediate values, resulting in an unphysical but methodologically useful system of scaled solute-solvent interactions.

Periodically over the course of the simulations the internal energy of the system, corresponding to the sum of all the pairwise atomic interactions, is saved to disk. Foreign energies, or the internal energy of a given simulation evaluated at some other  $\lambda$  value, may also be saved, along with the derivative of the internal energy with respect to  $\lambda$ . These energy values are the raw data from which the thermodynamic solvation quantities  $\Delta F$ ,  $\Delta S$  and  $\Delta U$  are calculated, given that the simulation is in the canonical ensemble. These can then be converted to the NPT ensemble quantities  $\Delta G$ ,  $\Delta S$  and  $\Delta H$ , respectively.

## 1.5.2 Advances Presented in Thesis

Three methodological and conceptual advancements for calculating the solvation thermodynamic quantities –  $\Delta F$ ,  $\Delta S$  and  $\Delta U$  – are presented in this thesis.

### 1. Improved Estimators

There are a number of techniques by which free energy changes can be obtained from simulation output. The Bennett acceptance ratio estimator (BAR) and the recently published multistate Bennett acceptance ratio (MBAR) estimators are designed to yield the free energy ( $\Delta F$ ) estimates with the lowest possible statistical error. In chapter 3, we present two new entropy ( $\Delta S$ ) estimators, derived from the BAR and MBAR  $\Delta F$  estimators. We show that like their  $\Delta F$  versions, these methods have markedly superior performance as measured by statistical error.

### 2. Improved $\lambda$ Schedule

The specific  $\lambda$  values at which equilibrium simulations are performed have traditionally been chosen in an *ad hoc* fashion. In chapter 4, we derive a relationship, based on recent results in the field of nonequilibrium statistical mechanics, which prescribes the

formal requirements for an optimal  $\lambda$  schedule. A procedure is described for how test simulation results can be used to obtain improved  $\lambda$  schedules for the full simulation, which results in improved convergence of thermodynamic estimators.

### **3. Improved Sampling through Hamiltonian Replica Exchange**

Replica exchange techniques implement a Monte Carlo move which may exchange a parameter between two otherwise independent simulations in order to improve equilibration. Such simulations typically swap the temperature, although in Hamiltonian replica exchange the  $\lambda$  parameter is exchanged. While this technique has been used for  $\Delta F$  calculations in the literature, its efficacy has not been demonstrated and the the fundamental principles underlying Hamiltonian replica exchange are obscure. In chapter 4 we describe how Hamiltonian replica exchange helps to satisfy a multicanonical equilibrium condition, and demonstrate that it significantly speeds the convergence of both free energy and entropy calculations.

Taken together, these advances provide techniques by which thermodynamic quantities may be estimated with computer simulations more rapidly and accurately.

# Chapter 2

## Background

The goal of computational solvation thermodynamics is to calculate the free energies of solvation, and their entropy and enthalpy decompositions, from a series of molecular simulations. The physicochemical process of solute transfer and solvation is simulated by a gradual “growing in” of a solute molecule in a simulation cell of hundreds or thousands of explicitly represented water molecules. This “growing in” is accomplished by scaling the strength of the interactions of the water with the solute by means of a coupling parameter  $\lambda$ ; when  $\lambda = 0$  the solute and solvent do not interact, and for  $\lambda = 1$  they interact fully (see Fig. 2.1). Based on the statistics of the internal energies of the system accumulated over the course of the simulations at different  $\lambda$  values, the free energy of solvation, and its entropy/enthalpy decomposition, can be calculated.

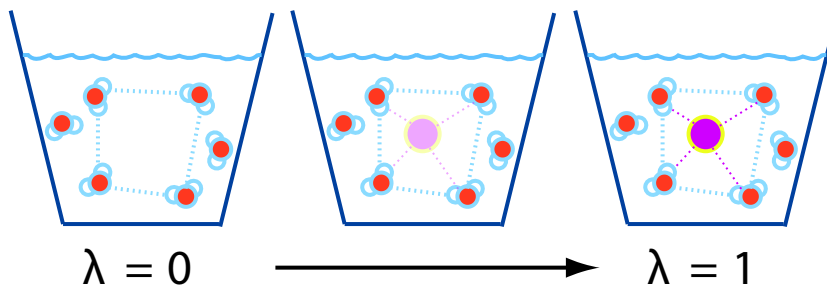


Figure 2.1: A schematic of solvation calculations. The parameter  $\lambda$  controls the solute-solvent interactions. For  $\lambda = 0$  the solute does not interact with solvent, and the simulation is effectively of the neat liquid. At  $\lambda = 1$  solute and solvent interact fully, and at intermediate  $\lambda$  values the interactions are gradually scaled.

## 2.1 Simulation Methodology

At a fundamental level, molecular mechanics models consist of some number of point mass particles representing atoms and their interactions through various forces. In the most straightforward case of molecular dynamics simulations, the positions and velocities of these particles evolve according to Newton’s equation of motion,  $\mathbf{a} = \mathbf{F}/m$ . The resulting behavior of the system then reflects and predicts the processes which take place at the atomic level in physical system. Here we briefly review the techniques and concepts relevant to solvation calculations; standard texts include Allen and Tildesley (1987), Frenkel and Smit (2002) and Leach (2001).

### 2.1.1 Molecular Mechanics Force Fields

A force field defines the potential energy of a system as a function of its atomic coordinates, and the force acting on an atom is the gradient of this potential. Most common force fields employed in the simulation of proteins use a potential energy function of the form (Ponder and Case, 2003),

$$V(\Gamma) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi[\cos(n\phi + \delta) + 1] + \sum_{nonbond\ pairs} \left\{ \frac{q_i q_j}{r_{ij}} + 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right\}. \quad (2.1)$$

Here,  $\Gamma$  refers to all the atomic coordinates collectively,  $r_{ij}$  to the distance between atoms  $i$  and  $j$ , and  $b$ ,  $\theta$  and  $\phi$  to the bond distance, bending angle and torsion angles between atoms.

The first three terms specify interactions between atoms adjacent to each other in a molecule (*bonds*), those separated by one atom (*angles*) and those separated by two atoms (*torsions*). The non-bonded terms govern the interactions of more distant atoms in a molecule as well as atoms in different molecules. Electrostatic interactions are proportional to the product of the charges of two atoms,  $q$ . In most force fields, hydrogen bonding is modeled as an electrostatic effect with two (or more) opposite

charges in a neutral molecule. These charges form a dipole which then simulates the linear hydrogen bond.

Lennard-Jones interactions are a second type of non-bonded force. They consist of two components: the repulsive term, whose potential decays as  $r^{-12}$ , simulates the hard repulsive core of atoms, while the dispersive  $r^{-6}$  term accounts for the attractive van der Waals interactions between atoms.

The most common force fields, such as AMBER (Case et al., 2005), CHARMM (MacKerell et al., 1998) and OPLS (Jorgensen et al., 1996), all have the same (or similar) functional form as in Eq. (2.1). They differ in the specific parameter values ( $\sigma$ ,  $\epsilon$ ,  $q$ , etc.) for a given molecular system, as well as in the methodologies by which such parameters are determined (Ponder and Case, 2003). For instance, the OPLS approach has emphasized the testing of the force field on thermodynamic properties of pure organic liquids, especially heats of vaporization and densities (Jorgensen et al., 1996). Similarly, the water models used in this work, TIP3P and TIP4P, were developed to reproduce the structural features as well as the energy and density of liquid water (Jorgensen et al., 1983).

It should be recognized that, force fields of the type in Eq. (2.1) involve severe tradeoffs between accuracy and computational efficiency, and have well recognized and significant shortcomings (Ponder and Case, 2003). For instance, the use of fixed charges means that the model is unable to respond directly to the molecular environment. Polarization, where an electric field induces a dipole in a molecule, is affected by the presence of other charges; this is inherently a multi-body effect and cannot be captured by potentials of the form (2.1), which consider strictly pair-wise interactions. Polarizable force fields iteratively solve for the induced electric field and are able to model this effect, albeit at a high computational cost (Grossfield et al., 2003). Force fields of the type in Eq. (2.1) have also been demonstrated to be limited in their ability to reproduce the hydrogen bond geometry found in high resolution PDB structures (Kuster, 2009). Despite their shortcomings, non-polarizable force fields are acceptably accurate in many situations, particularly those involving neutral molecules, and are widely used (Leach, 2001).



### 2.1.2 Sampling Techniques

Given the potential of the system as a function of atomic coordinates,  $V(\Gamma)$ , the task of molecular simulations is to obtain the equilibrium probability of observing a given system configuration,  $\rho(\Gamma)$ . Formally, this can then be used to derive the expectation value of any observable  $\mathcal{A}$ , as will be discussed later. The probability  $\rho(\Gamma)$  is given at thermal equilibrium by the Boltzmann distribution,

$$\rho(\Gamma) = \exp[-\beta V(\Gamma)]/Z \quad (2.2a)$$

where  $\beta = 1/k_B T$  and

$$Z = \int d\Gamma \exp[-\beta V(\Gamma)] \quad (2.2b)$$

is the configuration integral or partition function.  $Z$  is generally inaccessible to simulations directly, as it involves an integral over all configurations. Instead, various computational sampling techniques are used to generate configurations  $\Gamma$  which are sampled from the Boltzmann distribution. With enough sampling, the normalized histogram of such observations converges to the probability distribution  $\rho(\Gamma)$ .

The two major techniques by which configurations are generated with probabilities given by the Boltzmann distribution are molecular dynamics (MD) and Monte Carlo (MC) techniques. With either method, system statistics (e.g.  $\mathcal{A}(\Gamma)$ ) and possible entire configurations  $\Gamma$  are saved periodically over the course of the simulation and analyzed at its conclusion.

#### Molecular Dynamics

Molecular dynamics techniques integrate Newton's laws of motion,  $\mathbf{F} = m\mathbf{a}$ , in order to obtain the trajectory of all the atoms in the system as a function of time. The force is obtained from the gradient of the potential function,  $\mathbf{F} = -\nabla V$ . Starting from some arbitrary initial condition, after an initial equilibration period the system reaches thermal equilibrium, and successive of samples of the system configurations  $\Gamma$  are drawn from the equilibrium probability distribution.

Langevin dynamics (LD), a variant of molecular dynamics techniques, is a type of stochastic dynamics algorithm commonly used to simulate the effect of solvent in simulations where it is not explicitly represented. In an LD simulation, all degrees of freedom are implicitly coupled to a heat bath, and there is assumed to be an underlying continuum which provides frequent collisions to create the stochastic and friction forces. The dynamics are again governed by Newton’s laws of motion, but with two extra force terms. In addition to the force  $-\nabla V$  from Eq. (2.1), a force proportional to velocity,  $\gamma\mathbf{v}$  simulates the friction caused by the motion of the particle through the solvent, with  $\gamma$  the friction coefficient. Also, a random fluctuation force  $\mathbf{R}(t)$  captures the force on the particle due to random fluctuations caused by interactions with solvent molecules; this force is typically Gaussian in distribution with a zero mean (Leach, 2001; Skeel and Izaguirre, 2002). With these forces added, the simulation proceeds as described for molecular dynamics.

## Monte Carlo

Monte Carlo techniques obtain the probability distribution of atomic configurations by randomly generating trial moves (or configuration changes) and then accepting or rejecting them based on energetic criteria. Here we briefly present the theory of Monte Carlo simulations, describing along the way an alternate formulation which forms the basis of a set of results in chapter 4.

Monte Carlo simulations are based on the concept of a Markov chain of states (Newman and Barkema, 1999). Such a chain satisfies two conditions: (1) the outcome of each trial move depends only on the state of the previous trial, and (2) each trial belongs to a finite set of  $N$  possible outcomes.

Suppose that the probability of the system having a configuration  $\Gamma_m$  is  $\tilde{\rho}_m$ , with the tilde representing a non-stationary (i.e. not necessarily equilibrium) state. We can then construct a combined probability vector,

$$\tilde{\boldsymbol{\rho}} = (\tilde{\rho}_1, \tilde{\rho}_2, \dots, \tilde{\rho}_m, \tilde{\rho}_n, \dots, \tilde{\rho}_N). \tag{2.3}$$

which represents the probability of the system being in any give state, with  $\sum_i \tilde{\rho}_i = 1$ .

If the system is in state  $m$ , then the probability that it is in state  $n$  after the trial move is given by the transition probability  $\pi_{mn}$ .  $\boldsymbol{\pi}$  is then the  $N \times N$  matrix of all such transition probabilities.

Assume that at step 1 the system has the probability distribution  $\tilde{\boldsymbol{\rho}}^{(1)}$ . The probability distribution at the next step is then given as,

$$\tilde{\boldsymbol{\rho}}^{(2)} = \tilde{\boldsymbol{\rho}}^{(1)} \boldsymbol{\pi}. \quad (2.4)$$

At equilibrium, the probability distributions are stationary and do not change; representing such a distribution as  $\boldsymbol{\rho}$ , we require the following as a condition for equilibrium,

$$\boldsymbol{\rho} = \boldsymbol{\rho} \boldsymbol{\pi}. \quad (2.5)$$

In terms of vector components, this can be written as,

$$\sum_m \rho_m \pi_{mn} = \rho_n. \quad (2.6)$$

We proceed by imposing the condition of microscopic reversibility, which states that at equilibrium the transitions between two states (i.e.  $m \rightarrow n$  and  $m \leftarrow n$ ) occur with the same frequency. That is,

$$\rho_m \pi_{mn} = \rho_n \pi_{nm}. \quad (2.7)$$

Equilibrium probabilities are given by the Boltzmann distribution, Eq. (2.2a). While the partition function  $Z$  remains unknown, it cancels in the ratio  $\rho_n/\rho_m$ , which then becomes,

$$\frac{\pi_{mn}}{\pi_{nm}} = \exp\{-\beta[V(\Gamma_n) - V(\Gamma_m)]\} \quad (2.8)$$

There are a number of ways in which the ratio in Eq. (2.8) may be satisfied. The classic way, used in most Monte Carlo simulations, is with the Metropolis acceptance criterion, which defines the transition probability  $\pi_{mn}$  as (Metropolis et al., 1953),

$$\pi_{mn}^{Metropolis} = \min(1, \exp\{-\beta[V(\Gamma_n) - V(\Gamma_m)]\}). \quad (2.9)$$

Other possibilities exist, however, and in chapter 4 we make use of an alternative acceptance criterion, the Fermi criterion (see also Bennett (1976)),

$$\pi_{mn}^{Fermi} = \frac{1}{1 + \exp\{-\beta[V(\Gamma_n) - V(\Gamma_m)]\}}. \quad (2.10)$$

Both choices, Eq. (2.9) and Eq. (2.10), will satisfy Eq. (2.8), and will lead to a Boltzmann distribution of configurations. The Metropolis criterion typically leads to faster convergence in a simulation (Newman and Barkema, 1999), but it is not analytical. The Fermi criterion is more mathematically convenient, and its use as a replacement for the Metropolis criterion in the analysis of replica exchange swap probabilities (discussed in section 2.1.4) paves the way for one of the major results of chapter 4.

In practice, the Metropolis Monte Carlo algorithm is implemented as follows. Starting with system configuration  $\Gamma_m$ ,

1. Choose a trial move  $\delta\Gamma$  to obtain the trial configuration  $\Gamma_n = \Gamma_m + \delta\Gamma$ .
2. Evaluate  $\delta V = V(\Gamma_n) - V(\Gamma_m)$ .
3. If  $\text{rand}[0,1] < \min[1, \exp(-\beta\delta V)]$  then,
  - *Move accepted.*
  - Replace  $\Gamma_m$  with  $\Gamma_n$ .
4. Collect statistics on observables  $\mathcal{A}(\Gamma_m)$ , if necessary.
5. Repeat.

As in MD simulations, consecutive configurations are highly correlated, and system statistics are saved at some less frequent intervals.

### 2.1.3 Solvation Calculations

As illustrated in Fig. 2.1, solvation calculations scale the solute-solvent interactions by the parameter  $\lambda$  (Mezei and Beveridge, 1986). That is, the strength of the non-bonded forces between a solute and solvent atom is modulated; at  $\lambda = 0$  these forces

are strictly zero, and assume regular, unscaled values at  $\lambda = 1$ . The forces between solvent atoms are unaltered by  $\lambda$ , and solvent-solvent interactions may or may not be modulated, depending on the implementation, as will be discussed later.

## Scaling of Potentials

The most straightforward implementation is to scale solute-solvent non-bonded interactions linearly using an equation such as,

$$V_{solute\ solute}(r_{ij}, \lambda_{LJ}) = \lambda \left\{ \frac{q_i q_j}{r_{ij}} + 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right\} \quad (2.11)$$

While formally valid, this approach can lead to numerical instabilities when the distance between solute and solvent atoms is very small and  $\lambda \simeq 0$  (Shirts et al., 2003). For instance, unphysical effects like “nuclear fusion” can occur with the Lennard-Jones forces unable to counteract the electrostatic attractions. Some authors (e.g. Frenkel and Smit (2002)) suggest avoiding simulations with  $\lambda = 0$  and using extrapolation techniques to evaluate the end point along the  $\lambda$  schedule.

A more rigorous approach is to control the Lennard-Jones and Coulomb terms individually, and divide the insertion process into two legs: first, Lennard-Jones interactions are increased with the parameter  $\lambda_{LJ}$ , followed by the Coulomb interactions, governed by  $\lambda_C$  (Pitera and van Gunsteren, 2002). For this technique, linear scaling of the Coulomb interactions using

$$V_C(r_{ij}, \lambda_C) = \lambda_C \frac{q_i q_j}{r_{ij}} \quad (2.12)$$

is acceptable. This is because the electrostatic interactions become active only after the repulsive Lennard-Jones interactions are fully in place. However, linearly scaling of the Lennard-Jones interactions with the equation

$$V_{LJ}^{linear}(r_{ij}, \lambda_{LJ}) = 4\lambda_{LJ}\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.13)$$

remains problematic for very small values of  $r$ . The problem is that with  $\lambda = 0$  this potential is strictly zero for overlapping atoms, whereas for any  $\lambda$  infinitesimally greater than zero, such atoms will experience a very strong repulsive force. This is sometimes called the “fence post effect” (Shirts et al., 2003) can also lead to numerical instabilities, especially in molecular dynamics simulations (Beutler et al., 1994), and results in the ensemble changing discontinuously as  $\lambda$  increases past 0.

A number of methods have been introduced to scale the non-bonded interactions smoothly with  $\lambda$  and prevent the problems discussed above (Beutler et al., 1994; Pitera and van Gunsteren, 2002). One effective approach is the soft-core Lennard-Jones potential, given as (Beutler et al., 1994; Shirts et al., 2003),

$$V_{LJ}^{soft\ core}(r, \lambda_{LJ}) = 4\lambda_{LJ}^4 \epsilon \left\{ \frac{1}{[\alpha_{LJ}(1 - \lambda_{LJ})^2 + (r/\sigma)^6]^2} - \frac{1}{\alpha_{LJ}(1 - \lambda_{LJ})^2 + (r/\sigma)^6} \right\}, \quad (2.14)$$

where  $\alpha_{LJ}$  is a parameter which governs the the soft core term. Variants of this functional form are employed in chapters 3 and 4.

Whether solute-solute interactions should scale with  $\lambda$  is an open question, and both approaches have been employed in this thesis work. In most published work as well as in chapter 4, solute-solute non-bonded interactions are not scaled, whereas in chapter 3 solute-solute interactions scale just as solute-solvent interactions do. Each approach has its advantages and complications. Implementing advanced electrostatic techniques, such as Ewald summations (not discussed here) is considerably more difficult if solute-solvent interactions are not scaled. If such interactions are scaled, then there is a spurious contribution to the free energy of solvation stemming from changing solute-solute interactions. This must be accounted for and corrected by “growing in” a solute in vacuo, calculating the  $\Delta F$ ,  $\Delta U$  and  $\Delta S$  of this process, and subtracting these values from the results obtained in solvent. (This is the source of the self correction term discussed in chapter 3.) The choice of solute-solute interaction scaling will also affect the conformational ensemble of flexible solute for  $\lambda < 1$ . The molecules considered here are generally small and inflexible, however, so this difference should not be significant.

## The $\lambda$ Schedule

The  $\lambda$  schedule refers to the specific  $\lambda$  values for each equilibrium simulation in a solvation calculation. For the case of non-equilibrium simulations, discussed in section 2.2.2, the rate at which the  $\lambda$  value varies during the course of the simulation holds analogies to the  $\lambda$  schedule; the focus here, however, is on equilibrium simulations.

In general, we wish to find the free energy change between  $\lambda = 0$  and  $\lambda = 1$ ; that is, we wish to calculate  $\Delta F$ , defined as,

$$\Delta F = F(\lambda = 1) - F(\lambda = 0). \quad (2.15)$$

$\Delta U$  and  $\Delta S$  are defined similarly. Since the  $\lambda$  schedule is divided up into multiple simulations, and we typically obtain the free energy change between simulations adjacent to each other on that schedule; we call these neighboring simulations. The free energy change between neighboring simulations is,

$$\delta F = F_{i+1} - F_i, \quad (2.16)$$

where  $F_i = F(\lambda_i)$ , and with  $M$  simulations in the schedule,  $\Delta F = \sum_i^{M-1} \delta F_i$ . Again,  $U_i$  and  $S_i$  are defined similarly, and  $V_i(\Gamma) = V(\Gamma, \lambda_i)$ .

Researchers often use a linear  $\lambda$  schedule, with equal  $\lambda$  spacing across the range  $\lambda = 0 - 1$  (e.g. Jiang et al. (2009)). Shirts et al. (2003), using the accuracy of the trapezoidal rule in thermodynamic integration as a guide, placed additional  $\lambda$  values in regions of the schedule where the curvature of  $\langle \partial V / \partial \lambda \rangle$  is large. Lu and Kofke (1999) emphasize the importance of a judicious  $\lambda$  schedule (or “staging” in their parlance), and argue that it should be selected such that the entropy change per neighboring  $\lambda$  pair is constant. Other authors adjust the  $\lambda$  schedule such that the free energy change for neighboring  $\lambda$  values should be equal (Pearlman and Kollman, 1989). In chapter 4, we argue that the replica exchange swap probability, defined below, should be equal between neighboring simulations. This quantity pertains to the rate at which errors in free energy calculations between neighboring simulations decrease. Very recent published work (Shenfeld et al., 2009) supports this approach, and will be discussed in chapter 5.

## 2.1.4 Replica Exchange

The Hamiltonian replica exchange method (Sugita and Okamoto, 1999; Fukunishi et al., 2002) is a computational technique used to enhance sampling for a series of simulations, or replicas, running in parallel. Each replica has a unique  $\lambda$  parameter associated with it<sup>1</sup>, and the simulation proceeds independently in each replica for the duration of a round. At the end of a round, a series of swap attempts takes place.

For each swap attempt, an exchange of  $\lambda$  parameters is attempted between two replicas chosen at random. Assigning the labels 0 and 1 to the two replicas chosen and with  $\Gamma_i$  the ending configuration for simulation with  $\lambda_i$ , we evaluate the change in the total potential energy of a system upon a swap,  $\Delta V_{swap}$ , as,

$$\Delta V_{swap} = V_0[\Gamma_1] + V_1[\Gamma_0] - V_0[\Gamma_0] - V_1[\Gamma_1]. \quad (2.17)$$

The probability of a swap succeeding is given by Metropolis swap probability,

$$P_{swap} = \min[1; \exp(-\beta\Delta V_{swap})]. \quad (2.18)$$

This swap probability satisfies the microscopic reversibility criterion in the multi-canonical ensemble (Sugita and Okamoto, 1999) and is formally a Monte Carlo move in the “super-system” of all concurrent replicas. Provided that this “super-system” is at equilibrium to begin with, a swap, if it occurs, will preserve that equilibrium. Chapter 4 discusses the significance of the swap probability in greater detail.

If a swap succeeds,  $\lambda_0$  becomes associated with replica  $\Gamma_1$  and vice versa. With  $K$  replicas in a simulation, we attempt  $(K-1)^2$  swaps between rounds. At the conclusion of these swap attempts a new round of simulation begins, with each replica resuming with the final coordinates of the previous round and possibly a new  $\lambda$  parameter.

**Non-neighbor swaps** Traditionally, replica exchange moves are typically permitted only between replicas which are neighbors along the  $\lambda$  schedule (Jiang et al.,

---

<sup>1</sup>In the more popular temperature replica exchange techniques, temperature takes the place of the  $\lambda$  parameter, but operationally the implementation is very similar.



2009). In our implementation of replica exchange in chapter 4, we permit swaps between any two replicas along the  $\lambda$  schedule. Such non-neighbor swaps are believed to increase the efficiency of replica exchange, particularly in cases where the  $\lambda$  schedule is relatively dense, by allowing replicas to traverse the entire range of  $\lambda$  from 0 to 1 more quickly than if only neighbor swaps were permitted (Predescu et al., 2005).

The algorithm of the multiple-swap replica exchange algorithm is sketched out below. At the end of a round, each of the  $K$  replicas have associated with them configuration  $\Gamma_i$  and parameter  $\lambda_m$ ; the mapping between them is given by the array  $f$ , with  $i = f(m)$ . The inverse map gives  $m = f^{-1}(i)$ . For each swap attempt, two distinct replicas  $i$  and  $j$  are randomly chosen, with the corresponding  $\lambda$  indices  $m = f^{-1}(i)$  and  $n = f^{-1}(j)$ .  $\Delta V_{swap}$  in this context is given as,

$$\Delta V_{swap} = V_m[\Gamma_j] + V_n[\Gamma_i] - V_m[\Gamma_i] - V_n[\Gamma_j],$$

If the swap succeeds (according to the swap probability (2.18)), the array  $f$  is updated as,

$$f(m) = j, \quad f(n) = i.$$

This modified array is then used for all succeeding swap attempts. At the start of the following round, each replica  $i$  has the initial configuration  $\Gamma_i$  and parameter  $\lambda_{f^{-1}(i)}$ .

## 2.2 Free Energy Methods

The free energy of a simulation at a given value of  $\lambda$  can be defined formally as,

$$\exp(-\beta F_\lambda) = \int d\Gamma \exp[-\beta V_\lambda(\Gamma)] \quad (2.19)$$

The right side is an integral over all possible configurations of the system, and all portions of phase space contribute. Such a thermal quantity (Frenkel and Smit, 2002) can formally be evaluated only by an exhaustive integration over all of phase space, an unfeasible proposition for all but trivial systems. Instead, we calculate the free energy differences  $\Delta F = F_1 - F_0$  associated with a process in which  $\lambda$  varies

(Bennett, 1976). A variety of techniques for evaluating such free energy changes have been developed, and the field remains active.

### 2.2.1 Equilibrium techniques

Equilibrium techniques are perhaps the most common types of free energy calculations. Here, the  $\lambda$  parameter is held fixed over the course of a simulation and, after an initial equilibration period, the probability of a given configuration can be assumed to be given by the Boltzmann distribution. There are several such estimators, discussed below; see section 3.2.1 for additional details.

#### Thermodynamic Integration

Thermodynamic integration (TI) is based on the identity,  $\partial F/\partial\lambda = \langle\partial V/\partial\lambda\rangle$ . Integrating this derivative from  $\lambda = 0$  to  $\lambda = 1$  yields (Leach, 2001),

$$\Delta F = \int_0^1 d\lambda \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda. \quad (2.20)$$

In practice, equilibrium simulations are performed at discrete values of  $\lambda$ , and the integral  $\Delta F$  is typically approximated by the trapezoidal rule. Implementation of thermodynamic integration requires that derivatives of  $V$  with respect to  $\lambda$  be output during the course of a simulation.

#### Free Energy Perturbation

The free energy perturbation (FEP) method, attributed to Zwanzig (1954), can be derived from Eq. (2.19) and gives the pair of estimators,

$$\delta F^F = -\ln\langle\exp(-\beta V_1 + \beta V_0)\rangle_0, \quad (2.21a)$$

$$\delta F^R = +\ln\langle\exp(+\beta V_1 - \beta V_0)\rangle_1. \quad (2.21b)$$

The two formulae correspond to the forward and reverse estimates in which simulations take place at  $\lambda_0$  and  $\lambda_1$ , respectively (as discussed in chapter 4). The method extrapolates information from one equilibrium simulation to obtain the free energy difference at another  $\lambda$  value. In its implementation, it requires (for instance) the foreign energy  $V_1$  to be evaluated for a configuration observed in the  $\lambda_0$  simulation. The free energy perturbation method is formally exact and can calculate  $\Delta F$ , in principle, from one simulation. In practice, multiple  $\delta F$ 's are calculated along a  $\lambda$  schedule and summed to obtain the  $\Delta F$  of the entire process.

This technique suffers from slow convergence, and is prone to statistical errors (Zuckerman and Woolf, 2004). Furthermore, the forward and reverse estimators converge at different rates (Lu et al., 2003b), and knowing which one is more accurate is in general difficult. The underlying basis behind the convergence problems can be understood in terms of the distinction between the system configurations which are most frequently sampled, and those which must be observed in order for the estimator to converge (Jarzynski, 2006). In essence, the important configurations are rarely seen, and correspond to *transient* violations of the second law of thermodynamics (Ritort, 2003).

### Bennett Acceptance Ratio

The Bennett acceptance ratio (BAR) method (Bennett, 1976) makes use of two simulations to obtain  $\delta F$ , rather than one as for FEP. It was derived to minimize the statistical error (or variance) associated with an estimate based on a finite number of samples,  $N_0$  and  $N_1$ , drawn from two independent simulations, respectively. Here,

$$\delta F = \ln \frac{\sum_1 f(U_0 - U_1 + C)}{\sum_0 f(U_1 - U_0 - C)} + C - \ln(N_1/N_0), \quad (2.22)$$

$$\delta F = C - \ln(N_1/N_0), \quad (2.23)$$

where the Fermi function  $f(x) = 1/[1 + \exp(x)]$ , and the set of equations are solved iteratively until convergence. To estimate  $\delta F$ , output from two simulations, each with foreign energies, is needed. Although this method was developed in the mid-1970's, it was largely ignored for 30 years, and only relatively recently have its advantages been appreciated and its use increased (Shirts and Pande, 2005b).

In order for the BAR estimator to converge, configurations which are common to both simulations must be observed (Hahn and Then, 2009), a much less stringent requirement than that needed of the FEP estimators. As a result, the BAR estimator converges more quickly than FEP (see also Shirts and Pande (2005b)).

### Multistate Bennett Acceptance Ratio

In a typical free energy of solvation calculation (e.g. Shirts et al. (2003)) simulations take place at a number of  $\lambda$  values. To obtain the free energy change between adjacent  $\lambda$  values, the FEP technique makes use of data from one simulation, and BAR uses two. The free energy changes between neighboring simulations are then summed to obtain the free energy change across the entire  $\lambda$  schedule. The multistate Bennett acceptance ratio (MBAR) method (Shirts and Chodera, 2008) makes use of data from all simulations to obtain a free energy estimate across the entire schedule. Like BAR, it was derived to minimize the statistical error stemming from a finite number of samples, and is given as,

$$\beta F_i = -\ln \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{\exp[-\beta V_i(x_{jn})]}{\sum_{k=1}^K N_k \exp[\beta F_k - \beta V_k(x_{jn})]} \quad \text{for } i = 1, 2, \dots, K, \quad (2.24)$$

where  $K$  is the number of simulations and  $N_j$  is the number of observations drawn from the  $j^{\text{th}}$  simulation. In fact, for  $K = 2$ , MBAR reduces to the BAR method (Shirts and Chodera, 2008). The MBAR technique is relatively new and has not been widely used, although results presented in chapter 3 suggest that it yields estimates with the lowest variance of all the estimators. To implement this method, each simulation at every  $\lambda$  value needs to output foreign energies at all the other  $\lambda$  values.

### 2.2.2 Slow Growth and Nonequilibrium Techniques

Not all free energy techniques require the use of equilibrium simulations with fixed  $\lambda$  values. Slow growth techniques (Straatsma et al., 1986) utilize one simulation for which  $\lambda$  is slowly and continuously modified from  $\lambda = 0$  to  $\lambda = 1$ . In principle, if  $\lambda$  is varied slowly enough, the system will remain in equilibrium at all times, in

analogy to a quasi-static process in thermodynamics. Then, the fundamental formula for thermodynamic integration, Eq. (2.20), can be evaluated directly, rather than through piecewise trapezoidal integration between multiple equilibrium simulations. For a time, this technique enjoyed some popularity and refinement (e.g. Hunter et al. (1993)).

It has been shown, however, that slow growth techniques suffer from unavoidable systematic error due to a phenomenon known as Hamiltonian lag (Pearlman and Kollman, 1989): the ensemble average of  $\partial V/\partial\lambda$  trails the actual  $\lambda$  value due to a time lag between the changing of  $\lambda$  and a reflection of this change in the ensemble average. As a result, simulation protocols which continually adjust the  $\lambda$  parameter have fallen out of favor and attention focused on equilibrium simulations, which do not suffer from this problem.

Jarzynski (1997) described a relationship between the work performed along a path and the free energy change which is not dependent on the system being at (or near) equilibrium:

$$\exp(-\beta\Delta F) = \langle \exp(-\beta W) \rangle_{path}. \quad (2.25)$$

Here, the ensemble average is taken over an ensemble of simulations in which  $\lambda$  is continuously varied between 0 and 1 at some rate (which defines the path). This relationship was later generalized by Crooks (2000) to relate the probabilities of observing given work distributions far from equilibrium, a result which forms the theoretical underpinning for chapter 4. The Jarzynski equality has been employed for free energy calculations (Hendrix and Jarzynski, 2001) and continues to be under active development (e.g. Oberhofer and Dellago (2009)). In general, the convergence rate is dependent on the speed with which  $\lambda$  is varied, with more slowly varying simulations converging more quickly. Nevertheless, it is frequently found that its convergence properties are poor, and it is not clear that it is superior to equilibrium techniques for computer simulations (Cossins et al., 2009). It has perhaps been most useful in the experimental field, where nonequilibrium analysis techniques have been used to estimate the folding free energy of RNA using AFM and optical tweezer experiments (Collin et al., 2005).

### 2.2.3 Reaction Coordinate Techniques

In the techniques described thus far, the coupling parameter  $\lambda$  is an intensive parameter, established and held fixed by the simulation protocol. In a related but distinct set of techniques, we consider instead an extensive parameter  $\xi(\Gamma)$  which is a function of the coordinates  $\Gamma$ .  $\xi$  is frequently called the reaction coordinate, examples of which include the separation between two particles or the radius of gyration of a protein. We then wish to find the free energy profile of the system as a function of  $\xi$ , called the potential of mean force, which is related to the probability of observing some given value of  $\xi$  in a simulation. Unlike  $\lambda$ ,  $\xi$  generally fluctuates as the system evolves in time.

In calculating the probability of a specific  $\xi$  value, it is necessary to observe this and other coordinate values with sufficient frequency in order to accumulate good statistics – that is, the system must be well sampled. It often happens that the reaction coordinates of interest are infrequently observed, making the accumulation of adequate statistics time consuming or practically impossible. The umbrella sampling technique was developed by Torrie and Valleau (1977) in order to improve sampling of infrequently-observed configurations. By adding an additional “umbrella” potential  $W(\xi)$ , otherwise rarely observed values of  $\xi$  are stabilized. With  $W(\xi)$  known, its contribution to the final biased probability distribution can be backed out and the unbiased probability distribution  $\rho(\xi)$  recovered. A significant complication with this technique, however, is that the optimal umbrella potential needed to explore all the relevant  $\xi$  efficiently is not *a priori* known. A number of techniques have been developed to construct umbrella potentials and simplify the use of umbrella sampling.

One such technique is the metadynamics method (Laio and Parrinello, 2002). Here, an umbrella potential is constructed incrementally by adding an energetic penalty to frequently observed reaction coordinates. Over time, as such “penalties” accumulate, the common configurations are destabilized and the system explores other regions of phase space. This bias is itself the umbrella potential, and is used to construct the free energy surface over the reaction coordinate.

A common method for improving sampling along reaction coordinates is the Weighted Histogram Analysis Method (WHAM) (Kumar et al., 1992; Souaille and Roux, 2001). Here, an entire series of (typically parabolic) umbrella potentials is constructed, with one such potential per simulation. Together, these umbrella potentials cover the entire range of reaction coordinates, so that all relevant values of  $\xi$  are sampled in one simulation or another. The key aspect of WHAM is the reconstruction of the unbiased potential from a multitude of biased simulations. There is no unique way to remove this bias; WHAM provides one technique for doing so which is optimal in the sense of minimizing the variance (and maximizing the reproducibility) of the resulting PMF. In that sense, WHAM is similar to MBAR, with both estimating the free energy profiles with the least variance based on limited sampling. The MBAR technique can in fact be understood as a histogram-free extension of WHAM for obtaining the free energy as a function of the coupling parameter  $\lambda$  (Shirts and Chodera, 2008).

Up to this point, the fluctuating extensive reaction coordinate  $\xi$  and the imposed intensive control parameter  $\lambda$  were considered distinct. In techniques such as  $\lambda$ -dynamics (Kong and Brooks III, 1996),  $\lambda$  is itself a dynamical variable which evolves in time. Essentially,  $\lambda$  becomes a “particle” with its own mass, inertia and restraining umbrella potential  $V(\lambda)$ . These terms associate with  $\lambda$  a kinetic and potential energy, so that it evolves in time just as the atomic coordinates do. The observed probability distribution of  $\lambda$  is used, along with the  $\lambda$  umbrella potential, to construct an unbiased potential of mean force along  $\lambda$ . Just as in traditional umbrella sampling, however, the biasing potential  $V(\lambda)$  is not *a priori* known, and its specific shape is important to obtaining good sampling across the entire  $\lambda$  range. In addition, the potential must rise steeply at  $\lambda \leq 0$  and  $\lambda \geq 1$  to prevent the sampling of unphysical (and potentially numerically destabilizing) configurations.

## 2.3 Entropy and Enthalpy Decompositions

The free energy of a system can be divided into its entropy and enthalpy<sup>2</sup> components as,

$$F = U - TS. \quad (2.26)$$

Formally, in the context of statistical mechanics,  $U = \langle V \rangle$  and  $S = -k_B \langle \ln \rho \rangle$ . The enthalpy measures the average internal energy of the system, while the entropy quantifies the volume of phase space accessible to it.

The fundamental importance of entropy and enthalpy decomposition rests on the fact that, with the temperature as a controllable experimental parameter, the free energy change associated with any physical process can be experimentally resolved, at least in theory, into its entropy and enthalpy contribution.

Unlike free energy changes, which report on the spontaneity and stability of a process, entropies and enthalpies can yield information about the molecular mechanisms associated with that process. As an example, two solutes may be equally soluble under given conditions for entirely different reasons. For one the interaction with water may be energetically favorable, whereas solvation of the other solute may increase the configuration space available to itself and to water, making the process entropically driven (Levy and Gallicchio, 1998). Such different solvation mechanisms are discernible only when the free energy is decomposed into its entropic and enthalpic constituents.

In addition to elucidating mechanism, entropy and enthalpy decompositions can also provide an additional point of reference in the parameterization of atomic force fields (Gallicchio et al., 2000). Being able to computationally reproduce the entropy and enthalpy of an experimental process, in addition to its free energy, provides evidence that the underlying physical phenomena are correctly captured in a simulation and that the force field is transferable to other conditions and systems.

---

<sup>2</sup>In the context of estimators, where the ensemble is not specified, we use the terms enthalpy and energy interchangeably. The distinction is immaterial here, but is discussed in detail in chapter 3.



Most entropy/enthalpy estimators, aside from the direct method (below) are based on the thermodynamic relationship (Chandler, 1987),

$$S = -\frac{\partial F}{\partial T}. \quad (2.27)$$

Other methods to calculate absolute entropy changes exist, based for instance on the analysis of the covariance matrix of atomic fluctuations (Carlsson and Aqvist, 2005). These are not decompositions of the free energy change associated with a process, however, and will not be considered here.

In spite of its utility, progress in the decomposition of solvation processes into entropy and enthalpy components has been slow. This is likely, in large measure, due to the large errors and slow convergence of quantities associated with these decompositions, as discussed below.

### 2.3.1 The Direct Estimator and Entropy/Enthalpy Convergence

At first glance, obtaining the entropy and enthalpy decomposition of a free energy seems straightforward. After all, the enthalpy is the average internal energy in the canonical ensemble, and its change can be evaluated with the direct estimator,

$$\Delta U = \langle V_K \rangle_K - \langle V_0 \rangle_0. \quad (2.28)$$

where  $\lambda_0 = 0$  and  $\lambda_K = 1$ . It turns out, however, that for typical solvation calculations the convergence properties of this estimator are poor, 10 to 100 times slower than that of free energies (Lu et al., 2003a). Consequently its statistical error is frequently larger than the estimate  $\Delta U$  itself, making it uninformative for even a qualitative interpretation of the mechanism.

The reason for this is that the internal energy of a typical system is numerically large, with correspondingly large fluctuations. The energy difference from the solvation process, however, is frequently rather small. As a result, the signal  $\Delta U$  is simply swamped by the noise of the thermal fluctuations, a situation which gets worse with

increasing system size. By contrast, free energies evaluate quantities such as  $\langle \partial V / \partial \lambda \rangle$  or  $\langle \exp(V_1 - V_0) \rangle$ ; in the former case only the  $\lambda$ -dependent contribution is considered in the average, and in the latter any common (non  $\lambda$ -dependent) energy components tend to cancel each other (Lu et al., 2003a). As a result, in general, free energy calculations converge much more quickly, and the direct estimator is useful only when the enthalpy changes of a process are large with respect to thermal fluctuations. Nevertheless, perhaps due to its simplicity, the direct estimator is still commonly employed (e.g. Horinek et al. (2009)).

### 2.3.2 Finite Difference Methods

Finite difference methods approximate Eq. (2.27) with the finite difference approximation (Levy and Gallicchio, 1998),

$$\Delta S \simeq - \frac{\Delta F(T + \Delta T) - \Delta F(T - \Delta T)}{2\Delta T}, \quad (2.29)$$

and obtain an entropy estimate by performing free energy estimates – using any of the free energy estimators described previously – at two different temperatures.

In principle, such techniques can take advantage of the superior convergence qualities of free energy estimators to obtain the entropy. However, they require a judicious choice of  $\Delta T$ . The statistical uncertainty of this method is inversely proportional to  $\Delta T$  (Kubo et al., 1997) and as  $\Delta T$  gets smaller, increasingly long simulations are required in order to minimize the error of the free energy – and hence entropy – estimates. If  $\Delta T$  is too large, errors are introduced from the implicit assumption that the heat capacity is constant over the temperature range  $(T - \Delta T)$  to  $(T + \Delta T)$ . This assumption is reasonably valid near room temperature for  $\Delta T = 30K$  and up to 50K (Wan et al., 2004). One result of such an assumption, however, is that finite difference methods cannot accurately measure the variation of entropy with temperature. Also, three free energy simulations at temperatures  $T$ ,  $T + \Delta T$  and  $T - \Delta T$  have to be performed to estimate  $\Delta F$  and  $\Delta S$  at the temperature  $T$ .

Other finite difference methods, based on the relationship  $\Delta U = \partial(\beta\Delta F)/\partial\beta$  exist, although in general the differences in performance among such techniques are small (Lu et al., 2003a).

### 2.3.3 Analytical Methods

Rather than approximating Eq. (2.27) with finite differences, one can take analytical temperature derivatives of the free energy estimators directly. In fact, each of the free energy estimators discussed in section 2.2 has a corresponding entropy or enthalpy estimator, and such derivations form a central aspect of the work presented in chapter 3. We provide a brief outline of the known estimators, and defer further discussion and derivation to that chapter.

The TI and FEP (also known as Thermodynamic Perturbation) entropy/enthalpy estimators are frequently used in the literature (Smith et al., 1992; Wan et al., 2004). They are given as, for TI (Levy and Gallicchio, 1998),

$$T\Delta S = \beta \int_0^1 d\lambda \left[ \left\langle V \frac{\partial V}{\partial \lambda} \right\rangle_\lambda - \langle V \rangle_\lambda \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda \right], \quad (2.30)$$

and for FEP (Levy and Gallicchio, 1998),

$$T\delta S^F = + \frac{\langle V_1 \exp(-\beta V_1 + \beta V_0) \rangle_0}{\langle \exp(-\beta V_1 + \beta V_0) \rangle_0} - \langle V_0 \rangle_0 - \delta F^F, \quad (2.31a)$$

$$T\delta S^R = - \frac{\langle V_0 \exp(+\beta V_1 - \beta V + 0) \rangle_1}{\langle \exp(+\beta V_1 - \beta V_0) \rangle_1} + \langle V_1 \rangle_1 - \delta F^R. \quad (2.31b)$$

The entropy estimators corresponding to the BAR and MBAR methods are new to this work, and will be presented in chapter 3.

## 2.4 Error Measurements

Measuring the errors associated with solvation calculations is challenging. Thermodynamic quantities tend to converge slowly, with typical simulations, consisting of

several thousand atoms, requiring days or even weeks per single  $\lambda$  value. The exact values of these quantities are generally unknown, so the absolute error of a simulation is unknown. Beyond random statistical errors, which can themselves be large, thermodynamic estimators are also susceptible to systematic bias errors. Whereas statistical errors deal with precision and reproducibility of results, bias errors relate to accuracy and the convergence of a simulation to an incorrect value, and are typically not detectable with standard error analysis techniques.

Yet the evaluation of errors is of central importance to this work, both to compare novel estimation methods, and to evaluate the efficacy of improved  $\lambda$  schedules and advanced sampling techniques such as Hamiltonian replica exchange. For this reason, we have explored a variety of error estimation methods, and in some cases moved beyond the standard techniques utilized in the literature. Here we describe the four different error estimation techniques which are utilized in chapters 4 and 3.

### 2.4.1 Block Averaging

One of the most common techniques for evaluating the statistical error associated with a simulation is block averaging (Allen and Tildesley, 1987). Here, the entire simulation is divided into  $n_b$  consecutive blocks of equal length. The block size is chosen to be large enough such that the quantity of interest  $\mathcal{A}$  (for instance the free energy) evaluated in one block is independent of that quantity in another block. In our work we take a common approach used in the literature and typically set  $n_b = 10$ . By the central limit theorem (Kreyszig, 1993) the expected variance of the quantity  $\mathcal{A}$  over the entire simulation,  $\sigma^2(\langle \mathcal{A} \rangle_{sim})$ , is obtained from the variance of the mean of  $\mathcal{A}$  across all  $n_b$  blocks as,

$$\sigma^2(\langle \mathcal{A} \rangle_{sim}) = \sigma^2(\langle \mathcal{A} \rangle_{block})/n_b. \quad (2.32)$$

The square root of Eq. (2.32) is the expected standard deviation of  $\langle \mathcal{A} \rangle_{sim}$ , and indicates the confidence interval of the quantity. That is, we expect that future simulations using the same parameters will yield an average value of  $\mathcal{A}$  within one standard error of the current result about 68% of the time.

One disadvantage of the block averaging method is that the choice of  $n_b = 10$  leaves the error estimate itself susceptible to a fair amount of statistical error. Increasing the number of blocks over which to take the average will help this, and there are sophisticated ways to evaluate the optimum block size which maximize the number of blocks while ensuring that averages across these blocks remain statistically independent of one another (Flyvbjerg and Petersen, 1989). Such an approach introduces another complication when the quantity  $\mathcal{A}$  is the free energy. Certain thermodynamic estimators – notably, the free energy perturbation methods – have size-dependent bias errors. As a result, averages of many small blocks do not equal the average of a fewer number of larger blocks, even when the same dataset is used. This type of bias error stems from the fact that rare events play a dominant role in nonlinear averages, and are discussed in greater detail by Zuckerman and Woolf (2004) and Lu and Kofke (2001a). As a result, shortening blocks in and of itself introduces an error. Nevertheless, block averages remain popular and widely used, in part due to their simplicity, and we utilize them to estimate statistical errors in chapter 3.

## 2.4.2 Bootstrap Method

A second technique to estimate statistical error associated with a simulation is the bootstrap method (Efron and Tibshirani, 1993). For the case of a dataset of  $N$  unbiased observations, bootstrap proceeds by drawing  $n = N$  observations from the original dataset at random with replacement<sup>3</sup> to create a single bootstrap sample. The quantity of interest  $\mathcal{A}$  is then calculated from the bootstrap sample. This process is repeated many times, so that typically thousands of bootstrap estimates of  $\mathcal{A}$  are obtained. The mean and variance of the bootstrap estimates of  $\mathcal{A}$  then predict the mean and variance of the entire simulation.

When the observations in a dataset are correlated, as is generally the case in computer simulations, the number of observations  $n$  used for the bootstrap sample must correspond to the number of independent observations in the dataset. This number is usually obtained from the correlation time  $\tau$  of  $V$  (or  $\partial V/\partial\lambda$ ) as  $n = N/2\tau$ . We obtain  $\tau$  as the time at which the autocorrelation function of the quantity of interest

---

<sup>3</sup>With replacement indicates that each observation in the original dataset has an equal probability of being selected, regardless of whether it was selected previously.

decays to  $1/e$ , although other methods to estimate  $\tau$  exist, including integrated correlation times (Newman and Barkema, 1999) and measures of statistical inefficiency (Allen and Tildesley, 1987). All these techniques yield somewhat different values of  $\tau$  and affect the statistical error estimate proportionally to  $1/\sqrt{n}$ .

A distinct advantage that the bootstrap method has over block averaging is that the average is taken over a much larger sample; we typically use  $n_b = 10$  for block averaging, but obtaining 10,000 bootstrap samples from the same dataset is easy to do, making the estimate less susceptible to statistical error. We use bootstrap methods in chapter 4. At the conclusion of that chapter, however, we also present unpublished data relating the observed spread of 43 independent free energy simulations and the projected error obtained from a bootstrap analysis of one such simulation. We find that the bootstrap method underestimates the statistical error of such simulations rather severely, possibly as a result of inadequate sampling. The bootstrap method also suffers from some of the same problems as the block average method, including an inability to detect bias errors and length-dependent bias for some estimators.

### 2.4.3 Hysteresis Error

Given the difficulty in quantifying errors with the block averaging and bootstrap techniques, we developed an error measure called hysteresis error, which is defined simply as the difference between the forward and reverse free energy perturbation estimators (see chapter 4). The hysteresis error has several attractive features. First, it incorporates bias as well as statistical errors; since the forward and reverse FEP estimators tend to be biased in opposite directions (Lu et al., 2003b) bias error tends to be emphasized. This error is also obtained from the entire simulation rather than subsets or blocks of the entire data, minimizing length-dependent bias error. It has a solid theoretical basis and is mathematically convenient. These properties are exploited in chapter 4 to demonstrate the efficacy of advanced sampling techniques and lend insight to their mechanism. The hysteresis error indicates areas of the  $\lambda$  schedule where accurate sampling is problematic and where all estimators experience convergence difficulties, albeit to differing degrees (see Fig. 4.2).

One disadvantage of the hysteresis error is that it does not quantify the error of a simulation in a readily interpretable way. While it is strictly zero when the simulation is converged, it does not yield error bars or other measures that allow one to interpret the magnitude of the bias or statistical error. Finally, it is based on FEP techniques and does not quantify the error associated with more advanced estimators such as BAR and MBAR (discussed in chapters 3).

#### 2.4.4 Absolute Error

A final error measure introduced in chapter 3 is the absolute error. There, the Sun model provides a system for which the values of both free energy and entropy/enthalpy changes are known exactly. As a result, we can calculate the exact error, and plot its root-mean-square average over 100 independent simulations. The absolute error allows an unambiguous measure of the performance of estimators and includes both the statistical and bias errors. It is, however, restricted to simulations of simple systems where analytical results are available.

## 2.5 Computational Requirements

In order to implement the calculations described in this section as well as the sampling improvements from Hamiltonian replica exchange, molecular modeling software packages require:

1. Scaling of non-bonded interactions with  $\lambda$ , ideally with soft-core scaling. This is a basic requirement for free energy of solvation calculations.
2. The ability to output  $\partial V/\partial\lambda$ . This is required for the thermodynamic integration family of estimators and is available in many software packages.
3. The ability to write foreign energies  $V_{j\neq i}(\Gamma_i)$  – that is, the energy using some  $\lambda_j$  of a configuration taken from a system running with  $\lambda_i$ .
4. The facility to implement Hamiltonian replica exchange, which requires foreign energies.

As of this writing, relatively few publicly available molecular simulation programs have all the features described above. The work in chapter 4 was performed using the publicly available MCCC Towhee (Martin and Siepmann, 1999) Monte Carlo simulation package which was modified to implement all of the features described above (see <http://towhee.sourceforge.net>. Modifications available under version maw4.17.4 [maw-dev branch]). The work in chapter 3 was performed using the CAMPARI software package (Vitalis *et al.*, in prep.), which has implemented items 1 and 3.

Other popular packages, such as GROMACS (Hess et al., 2008), have only item 1 implemented. It is, in fact, possible to evaluate foreign energies in GROMACS during postprocessing: save the full-resolution trajectory (TRR) file and perform the command `mdrun -rerun` on the TRR file, with the new  $\lambda$  value in the run configuration MDP file; the energies written to the EDR file are then the foreign energies. This technique requires a relatively large amount of disk space to store the trajectories, but allows techniques like MBAR to be utilized.



# Chapter 3

## Acceptance Ratio Methods for Solvation Entropy and Enthalpy Calculations

### 3.1 Introduction

There are two primary motivations for investigating the decomposition of the free energy of solvation into entropy and enthalpy components. The first is to investigate in detail the mechanisms of solvation (Durell and Wallqvist, 1996; Smith and Haymet, 1993), so as to validate and refine existing solvation models (see also section 1.3). A second motivation is the development of improved parameters for molecular mechanics force fields (Kubo et al., 1997; Horinek et al., 2009). Entropies and enthalpies of solvation are experimentally accessible quantities, and reproducing them accurately in a simulation provides an additional point of validation for force field parameters (Levy and Gallicchio, 1998).

Both goals were recognized in early work in the field (Fleischman and Brooks III, 1987), yet progress has been slow. While solvation free energy calculations have advanced to the point where computational errors are on par with experimental ones (Shirts et al., 2003), obtaining entropy and enthalpy decompositions with similar errors has proven challenging (Gallicchio et al., 2000). Part of the problem is the inherent difficulty of such calculations, which suffer from seemingly unavoidable statistical errors 10 - 100 times larger than free energy calculations (Lu et al., 2003a). Another problem, however, is methodological. While all estimators are formally correct and

will yield the same converged value, some are better able to take advantage of limited amounts of data, so that they yield more converged quantities from shorter simulations. Free energy calculations can take advantage of advanced estimators such as BAR (Bennett, 1976) and MBAR (Shirts and Chodera, 2008), which utilize bridging estimators to take advantage of information from multiple simulations to improve the reliability of estimates made from limited simulation data. Currently available entropy and enthalpy estimators, on the other hand, utilize information from only one simulation at a time, in effect discarding data that could be used to improve the estimate.

We address this problem by developing BAR and MBAR entropy estimators which, like their free energy versions, utilize information from multiple simulations at once in order to reduce the error of the estimates. We demonstrate the validity of these methods on a simple model whose free energy, entropy and energy changes are known exactly. Next, we demonstrate the improved performance of the estimators under various simulation protocols for the solvation of NMA in explicit water. We conclude with an analysis of eleven model compounds and consider the effect of modified Lennard-Jones parameters on the free energy of solvation and its entropy and enthalpy decompositions.

## 3.2 Methods

In calculations designed to estimate the free energy of solvation, the solute molecule is effectively transferred from the gas phase into the solvent. Introduction of the Kirkwood coupling parameter  $\lambda$  (Mezei and Beveridge, 1986) into potential functions allows one to vary the degree of coupling between specific molecules in a dense fluid.  $\lambda$  modulates solute-solvent interactions, with the limits  $\lambda = 0$  and  $\lambda = 1$  corresponding to the pure solvent and solvent plus inserted solute, respectively, and intermediate values of  $\lambda$  interpolating smoothly between these limits. In our approach, independent equilibrium simulations are performed for a series of  $\lambda$  values drawn from a prescribed  $\lambda$  schedule. These simulations were performed in the canonical NVT ensemble. Different estimators were used to utilize data gathered from independent simulations to compute the Helmholtz free energies of solvation,  $\Delta F$ , and their decompositions into

entropies ( $\Delta S$ ) and energies ( $\Delta U$ ). Operationally, we define

$$\delta F = F(\lambda_1) - F(\lambda_0) \tag{3.1}$$

as the change in free energy between two adjacent  $\lambda$  values,  $\lambda_1$  and  $\lambda_0$ , and  $\Delta F$  is the cumulative sum of the  $\delta F$  across the  $\lambda$  schedule.  $\delta U$ ,  $\Delta U$ ,  $\delta S$  and  $\Delta S$  are defined similarly.

### 3.2.1 Free Energy and Enthalpy Energy Estimators

There exist a number of methods to estimate the free energy changes across a  $\lambda$  schedule. For each of these, an analytical temperature derivative of the  $\delta F$  equation yields a formula for either  $\delta U$  or  $\delta S$ . With two of these thermodynamic quantities known, the third is determined from the relationship,

$$\delta F = \delta U - T\delta S. \tag{3.2}$$

The exception to this is the direct method, which estimates only  $\Delta U$  and which has no corresponding free energy estimator.

The estimators in this section are written in terms of dimensional quantities, with the same temperature assumed for all simulations. Section 3.5.1 derives expressions for the BAR and MBAR estimators for the more general case of arbitrary temperatures per replica. The potential energy  $V(x, \lambda_i)$ , with  $x$  the positions of all the atoms of the system, is written as  $V_i(x)$ , and we write the work associated with changing the scaling parameter from  $\lambda_0$  to  $\lambda_1$  for a given  $x$  as (Ritort, 2003),

$$\delta V(x) \equiv V_1(x) - V_0(x). \tag{3.3}$$

#### Direct Energy Estimator

The simplest estimator for  $\Delta U$  is the direct estimator, which obtains the thermodynamic energy difference of two states based on the relationship  $U = \langle V \rangle$ . Specifically,

with  $\lambda_0 = 0$  and  $\lambda_K = 1$ ,

$$\Delta U = \langle V_K \rangle_{\lambda_K} - \langle V_0 \rangle_{\lambda_0}. \quad (3.4)$$

Unlike the other estimators, the direct estimator does not utilize information obtained from intermediate  $\lambda$  values. Where we wish to consider the energy profile as a function of  $\lambda$ , we define  $\delta U(\lambda_i) = \langle V_{i+1} \rangle_{i+1} - \langle V_i \rangle_i$ .

### Free Energy Perturbation

The free energy perturbation (FEP) method (Zwanzig, 1954) extrapolates information obtained in one simulation to calculate the free energy difference at another  $\lambda$  value. Because this estimator is directional, two independent free energy estimates – forward and reverse – can be constructed for a pair of simulations:

$$\delta F^F = -\beta^{-1} \ln \langle \exp(-\beta \delta V) \rangle_0, \quad (3.5a)$$

$$\delta F^R = +\beta^{-1} \ln \langle \exp(+\beta \delta V) \rangle_1. \quad (3.5b)$$

A temperature derivative of these equations yields the forward and reverse FEP (also known as thermodynamic perturbation) energy estimators (Levy and Gallicchio, 1998),

$$\delta U^F = + \frac{\langle V_1 \exp(-\beta \delta V) \rangle_0}{\langle \exp(-\beta \delta V) \rangle_0} - \langle V_0 \rangle_0, \quad (3.6a)$$

$$\delta U^R = - \frac{\langle V_0 \exp(+\beta \delta V) \rangle_1}{\langle \exp(+\beta \delta V) \rangle_1} + \langle V_1 \rangle_1. \quad (3.6b)$$

### Thermodynamic Integration

The thermodynamic integration (TI) estimator is based on the identity,

$$\frac{\partial F}{\partial \lambda} = \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda}, \quad (3.7)$$

whose temperature derivative leads to the expression (Levy and Gallicchio, 1998),

$$-T \frac{\partial S}{\partial \lambda} = \beta \left( \left\langle V \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda} - \langle V \rangle_{\lambda} \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda} \right) = \beta \operatorname{cov} \left( V, \frac{\partial V}{\partial \lambda} \right) \Big|_{\lambda}. \quad (3.8)$$

Both  $\langle V \rangle$  and  $\langle \partial V / \partial \lambda \rangle$  are accumulated from some number of equilibrium simulations, and typically the trapezoid rule is used to estimate the difference in the thermodynamic quantities between adjacent simulations, so that,

$$\delta F = \frac{\delta \lambda}{2} \left( \frac{\partial F}{\partial \lambda} \Big|_{\lambda_0} + \frac{\partial F}{\partial \lambda} \Big|_{\lambda_1} \right) \quad (3.9a)$$

and

$$T \delta S = \frac{\delta \lambda}{2} \left( T \frac{\partial S}{\partial \lambda} \Big|_{\lambda_0} + T \frac{\partial S}{\partial \lambda} \Big|_{\lambda_1} \right). \quad (3.9b)$$

### Bennett Acceptance Ratio

The Bennett acceptance ratio free energy estimator for two simulations of equal length and temperature may be written as,

$$\langle g_+^* \rangle_0 = \langle g_-^* \rangle_1, \quad (3.10a)$$

with

$$g_+^* \equiv [1 + \exp(+\beta \delta V - \beta \delta F)]^{-1}, \quad (3.10b)$$

$$g_-^* \equiv [1 + \exp(-\beta \delta V + \beta \delta F)]^{-1}. \quad (3.10c)$$

These equations are an implicit formula for  $\delta F$ , and are typically solved iteratively. Details of this and subsequent derivations are found in section 3.5, as are formulae for arbitrary simulation lengths  $N_i$  and temperatures  $T_i$ . Here, the \* symbol indicates that all  $N_i$  and  $T_i$  are equal. A temperature derivative of (3.10) yields an explicit equation for the entropy,

$$T \delta S = \frac{\alpha_0^* - \alpha_1^*}{\langle g_+^* g_-^* \rangle_0 + \langle g_+^* g_-^* \rangle_1} - \delta F, \quad (3.11a)$$

where

$$\alpha_0^* = \langle g_+^* V_0 \rangle_0 - \langle g_+^* \rangle_0 \langle V_0 \rangle_0 + \langle g_+^* g_-^* \delta V \rangle_0, \quad (3.11b)$$

$$\alpha_1^* = \langle g_-^* V_1 \rangle_1 - \langle g_-^* \rangle_1 \langle V_1 \rangle_1 - \langle g_+^* g_-^* \delta V \rangle_1, \quad (3.11c)$$

and the converged BAR  $\delta F$  estimate from Eq. (3.10) is used in Eq. (3.11a).

### Multistate Bennett Acceptance Ratio

Shirts and Chodera (2008) recently introduced the multistate Bennett acceptance ratio (MBAR) free energy estimator which uses information from all  $K$  simulations along a  $\lambda$  schedule to construct an estimate of the free energies  $F_i$  all at once, rather than calculating each pairwise  $\delta F$  individually. With an equal number of observations drawn from all simulations, and all with the same temperature, the MBAR estimator may be written as (see section 3.5),

$$\sum_j^K \langle \mu_i^* \rangle_j = 1 \quad \text{for } i = 1, 2, \dots, K, \quad (3.12a)$$

where for convenience we defined

$$\mu_i^*(x) \equiv \frac{\exp[\beta F_i - \beta V_i(x)]}{\sum_k^K \exp[\beta F_k - \beta V_k(x)]}. \quad (3.12b)$$

Equation (3.12) is a set of  $K$  equations which are solved simultaneously for all  $F_i$ , with  $F_0$  held fixed at 0.

Upon taking the temperature derivative of Eq. (3.12) (see section 3.5) we obtain the entropy estimator,

$$\begin{aligned} TS_i &= -F_i + \sum_j^K \langle \mu_i V_i \rangle_j + \sum_j^K (\langle \mu_i V_j \rangle_j - \langle \mu_i \rangle_j \langle V_j \rangle_j) \\ &+ \sum_k^K (F_k + TS_k) \sum_j^K \langle \mu_i \mu_k \rangle_j - \sum_k^K \sum_j^K \langle \mu_i \mu_k V_k \rangle_j \quad \text{for } i = 1, 2, \dots, K. \end{aligned} \quad (3.13)$$

Like Eq. (3.12), (3.14) is a set of  $K$  equations which yield  $S_i$ . In practice, the  $F_i$  are obtained first, then held fixed as the  $S_i$  are calculated iteratively.

### 3.2.2 Calculation of Free Energies of Solvation

We used our homegrown CAMPARI package (Vitalis and Pappu, 2009b; Vitalis et al., 2009) for calculating the free energies of solvation for different model compounds. Two sets of calculations were carried out for each of the model compounds. The first set used parameters from the OPLS-AA molecular mechanics force field (Jorgensen et al., 1996; Kaminski et al., 2001). The second set of calculations used the *m*OPLS-AA parameters, which utilize the partial charges from the OPLS-AA force field and modified Lennard-Jones parameters as detailed below (Vitalis and Pappu, 2009a).

The MBAR analysis for both the free energies and its decompositions was performed with a modified version of the freely available PyMBAR package (Shirts and Chodera, 2008).

#### Scaled Atomic Potentials

The solvation calculation takes place in two legs: first  $\lambda_{LJ}$  is scaled from 0 to 1 with  $\lambda_C = 0$ , and then  $\lambda_C$  increases from 0 to 1 with  $\lambda_{LJ} = 1$ . For convenience, in the context of the solvation calculations, we will refer to a single combined  $\lambda = \lambda_{LJ} + \lambda_C$  parameter increasing from 0 to 2.

The potential energy of the entire atomic system  $V(x, \lambda)$  may be written as,

$$\begin{aligned} V_{tot}(x, \lambda_{LJ}, \lambda_C) &= V_W^{bonded}(x) + V_S^{bonded}(x) + V_{W-W}^{nonbonded}(x) \\ &+ V_{W-S}^{nonbonded}(x, \lambda_{LJ}, \lambda_C) + V_{S-S}^{nonbonded}(x, \lambda_{LJ}, \lambda_C), \end{aligned} \quad (3.14)$$

where  $W$  and  $S$  refer to solvent and solute atoms, respectively. The bonded and solvent-solvent interactions are not modified by  $\lambda$  and are given as described in section

2.1.1. The scaled solute-solvent and solute-solute potentials are given by,

$$V_{W-S}^{nonbonded}(x, \lambda_{LJ}, \lambda_C) = \sum_i^{N_S} \sum_j^{N_W} [V_{LJ}(r_{ij}, \lambda_{LJ}) + V_C(r_{ij}, \lambda_C)], \quad (3.15a)$$

$$V_{S-S}^{nonbonded}(x, \lambda_{LJ}, \lambda_C) = \sum_i^{N_S} \sum_{j \neq i}^{N_S} [V_{LJ}(r_{ij}, \lambda_{LJ}) + V_C(r_{ij}, \lambda_C)]. \quad (3.15b)$$

Here,  $r_{ij}$  is the distance between atoms  $i$  and  $j$ . The index  $j$  in the  $V_{S-S}$  term iterates over all solute atoms that participate in nonbonded interactions with atom  $i$ , typically those which are separated by four or more covalent bonds. Scaling of solute-solute interactions allows for a simplified implementation of reaction field electrostatics, as discussed in the next section.

$V_{LJ}$  is a modified soft-core Lennard-Jones potential (Beutler et al., 1994) given by

$$V_{LJ}(r, \lambda_{LJ}) = 4\epsilon_{ij}\lambda_{LJ} (R^{-2} - R^{-1}), \quad (3.16)$$

where

$$R = 0.5(1 - \lambda_{LJ}^2) + r^6/\sigma_{ij}^6, \quad (3.17)$$

and

$$V_C(r, \lambda_C) = \lambda_C \frac{q_i q_j}{r}. \quad (3.18)$$

The combined Lennard-Jones parameters  $\sigma_{ij}$  and  $\epsilon_{ij}$  are constructed with geometric combination rules ( $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$ ,  $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ ) for the OPLS-AA parameters and Lorentz-Berthelot rules ( $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ ,  $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ ) for the *m*OPLS-AA parameters. These parameters, along with the atomic partial charges  $q_i$ , are listed for both force fields in table 3.1 and are discussed in section 3.2.2. The solvent of interest, water, was modeled using the rigid, three-site TIP3P model of Jorgensen et al. (1983).

## Force Field Parameters

Eleven model compounds were used this study, ten of which are analogs of amino acid side chains; the eleventh, NMA, is an analog of the repeating unit in a peptide backbone. For all of these solutes experimental data are available for free energies



of solvation as well as their decompositions into entropies and enthalpies. These compounds constitute a subset of those studied by the Pande group (Shirts et al., 2003; Shirts and Pande, 2005a), the free energies of solvation and their decompositions obtained from experiment are listed in table 3.3.

Figure 3.1 illustrates the molecular structures of the 11 model compounds used in this work. Many, although not all, of the molecules in the set are small, rigid solutes. Bond lengths and angles were held fixed in the equilibrium values prescribed by the all-atom OPLS-AA force field. We used an internal coordinate stochastic dynamics integrator to sample conformational space for the solute-solvent system, details of which are presented in section 3.2.2.

Vitalis and Pappu (2009a) followed Tran et al. (2005) and developed the *mOPLS-AA* set of Lennard-Jones parameters for Monte Carlo simulations of polypeptides in implicit solvent, where sampling is carried out in torsional space using fixed bond lengths and bond angles. Table 3.1 details the Lennard-Jones parameters and charges associated with each atom type. In general the  $\sigma$  parameters in *mOPLS-AA* are smaller than in OPLS-AA, while the well-depth parameters  $\epsilon$  on the whole tend to be larger for *mOPLS-AA*. In the OPLS paradigm, parameters for Lennard-Jones potentials are derived to reproduce the properties of neat liquids, with the  $\sigma$  parameters aiming to capture liquid densities and the  $\epsilon$  parameters heats of vaporization (Kaminski et al., 2001). The modified  $\sigma$  and  $\epsilon$  *mOPLS-AA* parameters are based on the original estimates of Pauling (1970), which reproduce crystal packing and heats of fusion of neutral model compounds.

## Simulation Details

For each model compound, we performed two sets of solvation calculations, one based on the OPLS-AA and the other on the *mOPLS-AA* force fields. These comparative calculations allow us to demonstrate how free energies of solvation and their decompositions into entropies and energies are useful for assessing the validity of different force field parameters.



Atom	OPLS-AA		<i>m</i> OPLS-AA		q
	$\sigma$	$\epsilon$	$\sigma$	$\epsilon$	
C1	3.500	0.066	3.300	0.100	0.020
C2	3.500	0.066	3.300	0.100	0.145
C3	3.500	0.066	3.300	0.100	0.085
C4	3.500	0.066	3.300	0.100	-0.180
C5	3.500	0.066	3.300	0.100	-0.120
C6	3.500	0.066	3.300	0.100	-0.060
C7	3.500	0.066	3.300	0.100	-0.240
C8	3.500	0.066	3.300	0.100	-0.065
C9	3.550	0.070	3.000	0.100	0.150
C10	3.550	0.070	3.000	0.100	-0.115
C11	3.750	0.105	3.000	0.100	0.500
H1	2.000	0.000	2.000	0.025	0.380
H2	2.000	0.000	2.000	0.025	0.300
H3	2.000	0.000	2.000	0.025	0.418
H4	2.000	0.000	2.000	0.025	0.435
H5	2.500	0.030	2.000	0.025	0.060
H6	2.420	0.030	2.000	0.025	0.115
N1	3.250	0.170	2.700	0.150	-0.760
N2	3.250	0.170	2.700	0.150	-0.500
O1	2.960	0.210	2.700	0.200	-0.500
O2	3.120	0.170	3.000	0.150	-0.683
O3	3.070	0.170	3.000	0.150	-0.585

Table 3.1: Lennard-Jones parameters for the OPLS-AA and *m*OPLS-AA force fields, as well as Coulomb parameter  $q$  shared by both. Atom types are defined in Fig. 3.1. Units:  $\sigma$  in [Å],  $\epsilon$  in [kcal/mol] and  $q$  in [ $e$ ].

**Integrator** Stochastic dynamics simulations based on integration of Langevin equations of motion were used to sample the system conformational space. The center-of-mass translations and solid body rotations of all molecules, as well as torsion angles for the flexible solutes, were sampled using the impulse integrator of Skeel and Izaguirre (2002). For all of the degrees of freedom we used  $\gamma = 5 \text{ ps}^{-1}$  for the frictional coefficient and a time step of 2.0 fs. Integrator stability was successfully tested by assessing the convergence of a system of TIP3P water to a target temperature of 298 K, as well as the reproduction of accurate distributions of energy fluctuations and the heat capacity. Details of how the algorithm of Skeel and Izaguirre was adapted for the case internal coordinate degrees of freedom are being prepared for publication (Vitalis and Pappu, in preparation).

An equilibration period of 1.6 ns was followed by a 8 ns production run, with energy output every 0.2 ps, for a total of 40,000 energy observations. The autocorrelation time of the potential energy is less than 0.4 ps.

**Cutoffs** The simulations were performed using periodic boundary conditions in cubic cells 32 Å to a side. Spatial distance cutoffs of 14 Å were employed for both the Lennard-Jones and electrostatic interactions. The reaction field method (Onsager, 1936) was used evaluate corrections due to truncation of long-range electrostatic interactions. As all of our solutes are neutral, use of the reaction field is appropriate (Garde et al., 1998). Analytical corrections were applied to account for the Lennard-Jones effects beyond this distance, as discussed later in section.

We employed twin-range cutoffs for both the Lennard-Jones and electrostatic interactions. The full set of interactions was computed if the pairwise distance was below 10 Å. For distances in the 10-14 Å range, we computed the interactions once every four time steps using neighbor lists that were also updated once every four steps.

**Ensemble and System Preparation** All solvation simulations were performed in the NVT ensemble, as stochastic dynamics based on the Langevin dynamics formalism reliably yields converged statistics from the canonical ensemble (Hünenberger, 2005). Experimental data, specifically for the enthalpy and entropy decompositions, were corrected for the NVT ensemble prior to comparison to the results from simulations. Details of how these corrections were made are discussed in the next section.

For neat water with a target density of 1 gm/cc, the central simulation cell comprises of 1,086 TIP3P water molecules. In preparing simulations with the solute added, the number of water molecules removed was based on the Van der Waals volume of the single solute and neat water density. We evaluated the error associated with this method by performing two additional simulations on the NMA system, one with an additional water molecule and another with one fewer water molecule. It was found that the errors for both  $\Delta F$  and  $\Delta U$  associated with this pressure perturbation were less than the standard errors associated with the simulation.

## The $\lambda$ Schedule

The  $\lambda$  schedule for the free energy of solvation calculations was constructed as described previously (Wyczalkowski and Pappu, 2008). The rate at which the error of a free energy estimate between two equilibrium simulations decreases – in particular the error associated with FEP estimators – is governed by the Fermi swap probability, defined as,

$$\langle p_{swap} \rangle = \left\langle \left\langle \frac{1}{1 + \exp(\beta\delta V[x_0] - \beta\delta V[x_1])} \right\rangle_0 \right\rangle_1, \quad (3.19)$$

where  $x_0$  and  $x_1$  are drawn from the  $\lambda_0$  and  $\lambda_1$  ensembles, respectively, and  $[\delta V(x_0) - \delta V(x_1)]$  (see Eq. (3.3)) is the change in the combined potential energy of both systems due to a replica exchange swap. Like the Metropolis transition probability typically used in Hamiltonian replica exchange calculations (Sugita et al., 2000), the Fermi swap probability satisfies detailed balance and leads to a multi-canonical Boltzmann probability distribution (Bennett, 1976; Newman and Barkema, 1999), although no actual swaps need to take place to calculate this quantity. A significant advantage of the Fermi swap probability is that it is analytical, and its Taylor series expansion in terms of  $\delta\lambda$  leads to the expression,

$$\langle p_{swap} \rangle \simeq \frac{1}{2} - \frac{\beta^2(\delta\lambda)^2}{4} \text{var} \left( \left. \frac{\partial V}{\partial \lambda} \right|_{\lambda_0} \right). \quad (3.20)$$

Free energy calculations, as well as their thermodynamic decompositions, converge more slowly and have larger errors if the swap probability is low. An optimum  $\lambda$  schedule is one where the errors between all simulations decrease uniformly, implying a uniform swap probability. Either Eq. (3.19) or Eq. (3.20) can be used to design a  $\lambda$  schedule which equalizes  $\langle p_{swap} \rangle$  across all replica pairs. This is done by adding additional simulations where the swap probability, based on preliminary simulations, is low or equivalently, where  $\text{var}(\partial V/\partial\lambda)$  is large. Shenfeld et al. (2009) have obtained a similar relationship between the replica exchange swap probability and the convergence rate of  $\Delta F$ , and related both to a measure of thermodynamic length (Crooks, 2007).

For this work, the  $\lambda$  schedule was refined by adding simulations in the interval  $0.2 \leq \lambda \leq 1.0$ , based on the analysis of a preliminary set of simulations for NMA, giving a 29

$\lambda$  schedule. This schedule, used for all model compounds unless indicated otherwise, consists of simulations at  $\lambda$  values (0.0, 0.1, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0).

## Correction Terms and Ensemble Conversions

The estimates of the thermodynamic quantities obtained from a simulation are not immediately comparable to experimental results. Corrections are needed to account for methodological artifacts, and results obtained from different ensembles must be converted to a common ensemble to permit comparison. Also, corrections may need to be applied to experimental results to bring them into line with standard states.

**Self Corrections** In our implementation, the interactions between scaled atoms of a solute are also scaled, and the contributions that these self-interactions make to the free energy must be taken into account. This is done by performing the same experiment in the absence of solvent, which yields self-correction terms which are subtracted from the thermodynamic quantities obtained in explicit solvent. Note that this is methodology specific: if solute-solute interactions are not scaled, then no such corrections are necessary. Self correction terms for all simulations are listed in table 3.5.

**Long Range Lennard-Jones Corrections** These corrections account for errors in the potential energy of the system imposed by cutoffs in the Lennard-Jones potential. Taking into account only the slower-decaying  $r^{-6}$  term, the correction is calculated for the entire solute molecule as (Allen and Tildesley, 1987),

$$V_{lr} = \sum_i \frac{8\pi N}{3Vr_c} \epsilon_{iW} \sigma_{iW}^6 \quad (3.21)$$

for each solute atom  $i$ , where  $\epsilon_{iW}$  and  $\sigma_{iW}$  represent the combined Lennard-Jones parameters between atom  $i$  and TIP3P water oxygen;  $N$  is the number of water molecules and  $V$  the volume of the box. With  $r_c = 1.4$  nm, the correction term for

each entire solute molecule is relatively small, varying between  $-0.02$  kcal/mol for Absinth methane and  $-0.13$  kcal/mol for OPLS toluene. This correction is added to both  $\Delta F$  and  $\Delta U$ , and its values are listed in table 3.5.

**Standard State Corrections** Experimentally obtained free energies of solvation and their decompositions are reported in a variety of standard states. To make them comparable to computational values, they must be converted to the Ben-Naim standard state convention (Ben-Naim and Marcus, 1984). Such corrections generally need to be applied to the free energies, entropies and enthalpies (Kubo et al., 1997).

In practice, we find that free energies reported in the literature frequently have the corrections already applied, or else the transfer process is defined such that no such corrections are necessary. For the case of Cabani et al. (1981), a particularly useful reference for decompositions, the free energies are already in standard state and only an additional term of  $\beta^{-1}(1-T\alpha_{water})=0.549$  kcal/mol needs to be added to tabulated  $\Delta H$  and  $T\Delta S$  values (Gallicchio et al., 2000).

**Ensemble Corrections** The solvation calculations, performed in the NVT ensemble, yield values for the Helmholtz free energy  $\Delta F$  as well as the energy  $\Delta U$  and the entropy, denoted in this ensemble as  $(\Delta S)_V$ . Experimental values are typically given for the NPT ensemble, yielding the Gibbs free energy  $\Delta G$ , the enthalpy  $\Delta H$  and the constant pressure entropy,  $(\Delta S)_P$ . In order to compare computational results to experimental values ensemble corrections, in addition to standard state corrections, must be performed.

The Helmholtz free energy obtained in an NVT simulation at a given volume is numerically equal to the Gibbs free energy obtained in an NPT simulation whose average volume (and pressure) is the same as the NVT simulation (Levy and Gallicchio, 1998), so the comparison between  $\Delta F$  and  $\Delta G$  can be made directly. The decomposition of the free energies is ensemble dependent, however. The NVT energy and entropy can

be obtained from the NPT enthalpy and entropy (Kubo et al., 1997),

$$\Delta U = \Delta H - (\Delta V)_P \frac{T\alpha}{\kappa}, \quad (3.22a)$$

$$T(\Delta S)_V = T(\Delta S)_P - (\Delta V)_P \frac{T\alpha}{\kappa}, \quad (3.22b)$$

where  $(\Delta V)_P$  is the partial molar volume of the solute,  $\alpha = V^{-1}(\partial V/\partial T)_{P,N}$  is the thermal expansion coefficient and  $\kappa = -V^{-1}(\partial V/\partial T)_{T,N}$  is the isothermal compressibility.

In principle, it is equally valid to either convert the computational quantities into the NPT ensemble, or the experimental values into NVT. However, the calculated TIP3P properties  $\alpha$  and  $\kappa$  needed to convert computational quantities from NVT to NPT are subject to large errors (Jorgensen et al., 1983; Jorgensen and Jenson, 1998; Mahoney and Jorgensen, 2000), making the conversion factors unreliable. Also, the partial molar volume  $(\Delta V)_P$  is only known experimentally, making its use in a computational conversion factor another possible source of error.

On the other hand, the values of  $\alpha$  and  $\kappa$  for liquid water are known to a relatively high degree of accuracy, and we take advantage of this reduced error by converting experimental results to the NVT ensemble for comparison to calculated values. For liquid water,  $\alpha = 25.7 \times 10^{-5}$  and  $\kappa = 45.8 \times 10^{-6}$  (Jorgensen and Jenson, 1998).

### 3.2.3 Error Estimates

Estimates of thermodynamic quantities based on simulation of fine length are subject to two types of error (Lu and Kofke, 2001b). The first is statistical error, which concerns issues of precision and reproducibility. This error quantifies the expected distribution of results given repeated simulations. In this work, statistical errors are estimated by block averaging (Allen and Tildesley, 1987).

The second type of error is bias error, and relates to the accuracy or correctness of the estimate. In certain instances (such as nonlinear averages) bias error is observed as the systematic dependence of the mean on the data size (Zuckerman and Woolf, 2004). In the context of evaluating estimator performance the bias error is a much more



informative quantity, but calculating the bias error requires that the exact solution be known, which is generally not the case.

### Absolute Error and the Sun Model

The Sun model is a dimensionless potential of a single coordinate, given as (Sun, 2003),

$$V_{sun}(x, \lambda) = x^4 - 16(1 - \lambda)x^2. \quad (3.23)$$

Figure 3.2 plots  $V_{sun}$  for various  $\lambda$ . This model is of interest as it can be analytically integrated to obtain the exact values of  $\Delta F$ ,  $\Delta U$  and  $T\Delta S$  associated with changing  $\lambda$  from 0 to 1 (see section 3.5.4). With the exact values of the thermodynamic quantities known, bias error can be accessed directly, allowing an unambiguous assessment of the accuracy of various estimators.

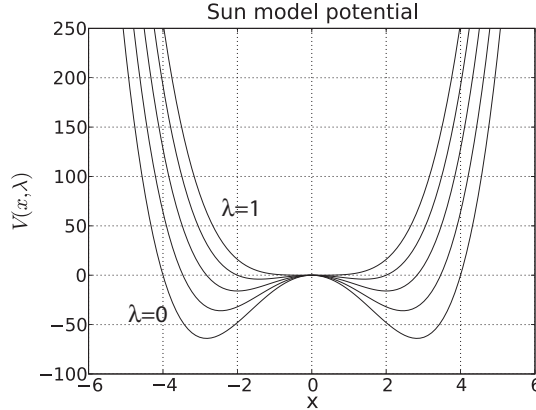


Figure 3.2: The Sun model potential, Eq. (3.23), as a function of coordinate  $x$  for various  $\lambda$  between 0 and 1. The potential is symmetric about  $x = 0$ .

We quantify the error of Sun  $\Delta F$  and  $\Delta U$  estimates using the  $\epsilon_{sun}$  error measure. The error  $\epsilon_{sun}^F$  is defined for  $M$  independent estimates  $\Delta F_i$  as,

$$\epsilon_{sun}^F = \sqrt{\sum_i^M (\Delta F_i - \Delta F_{exact})^2 / M}, \quad (3.24)$$

where the exact solution  $\Delta F_{exact}$  is known; the corresponding  $\Delta U \epsilon_{sun}^U$  error is defined analogously. Since it measures error with respect to the exact solution,  $\epsilon_{sun}$  incorporates both statistical and bias errors, and the averaging over multiple ( $M$ ) estimates reduces the statistical noise.

A simple Monte Carlo sampler was used to construct a dataset of 1 million observations for each of eleven  $\lambda$  values, spaced uniformly between 0 and 1. The inverse nondimensional temperature  $\beta$ , used in the sampler as well as the thermodynamic estimates, has the value of 0.02. For each observation at a given  $\lambda$  value, foreign energies corresponding to all other  $\lambda$  values were also computed. A second dataset, consisting of 3  $\lambda$  values ( $\lambda = 0, 0.5, 1$ ) was constructed by discarding data from the 11- $\lambda$  dataset. Being one dimensional, sampling is rapid and sequential observations, output every 10 Monte Carlo steps, are fully decorrelated based on autocorrelation analysis.

For both  $\Delta F$  and  $\Delta U$  the error  $\epsilon_{sun}$  was calculated as a function of dataset size. 100 consecutive samples, each of size ranging from 1 to  $10^4$  observations, were drawn from the entire dataset of 1 million observations.  $\Delta F_i$  and  $\Delta U_i$ , with  $i = 1..100$ , were calculated for each sample using different estimators, and  $\epsilon_{sun}$  is calculated with Eq. (3.24). The performance of the estimators in terms of this error was compared for a range of sample sizes.

## Standard Errors

With the exact values of the thermodynamic quantities unavailable for the solvation calculations, we utilized block averaging to obtain the standard error of the solvation calculation estimates (Allen and Tildesley, 1987). Simulation data were divided into ten consecutive blocks of 4000 observations each and the thermodynamic quantities were calculated independently for each block. With the variance of the 10 estimates given by  $\sigma^2$ , the standard error is  $\sqrt{\sigma^2/10}$ . Were the same solvation calculation to be repeated multiple times, we would expect that about 68% of the time those results would be within a standard error of the results of the present calculation. For simplicity, we calculated the block averages of  $\Delta F$ ,  $\Delta U$  and  $T\Delta S$  as if they

were independent quantities. However, only two of the three are independent and in practice, the standard errors of  $\Delta U$  and  $T\Delta S$  are very similar.

## 3.3 Results

### 3.3.1 Evaluation of Estimators

Section 3.2.1 provided an overview of several different estimators for calculating free energies based on data from equilibrium simulations for every  $\lambda$  value along a given schedule. Known entropy/energy estimators were presented, and we derived two new entropy estimators based on temperature derivatives of the BAR and MBAR free energy estimators. In this section, we evaluate the accuracy of these estimators for two different systems of differing complexity. Since the quantities  $\Delta S$  and  $\Delta U$  are not independent, we analyze the estimators in terms of  $\Delta F$  and  $\Delta U$  only.

#### Sun Model

The Sun model (Eq. 3.23) is a simple, one dimensional system which permits an analytical evaluation of the free energy and its decomposition. For this system  $\Delta F$  quantifies the free energy change associated with a transition from a bi-stable state to a state with a single minimum (see Fig. 3.2). With  $\beta = 0.02$ , analytical integration (see section 3.5.4) yields the nondimensional thermodynamic quantities  $\Delta F = 65.8878$ ,  $\Delta U = 53.1957$ , and  $T\Delta S = -12.6921$ .

Numerical simulation data were analyzed using all of the estimators discussed in section 3.2.1. We assessed the accuracy of the estimators by evaluating the error measures  $\epsilon_{sun}^F$  and  $\epsilon_{sun}^U$  as defined in Eq. (3.24). The errors report on the root-mean-square absolute error of 100 samples and are a function of sample size, or the number of independent observations in each sample. We consider two different  $\lambda$  schedules: the 11  $\lambda$  schedule simulations take place at  $\lambda = (0, 0.1, \dots, 0.9, 1)$ , and the 3  $\lambda$  calculations utilize a subset of these data for  $\lambda = (0, 0.5, 1)$ . Figure 3.3 plots the error as a function of sample size for  $\Delta F$  and  $\Delta U$  at the two different schedules.

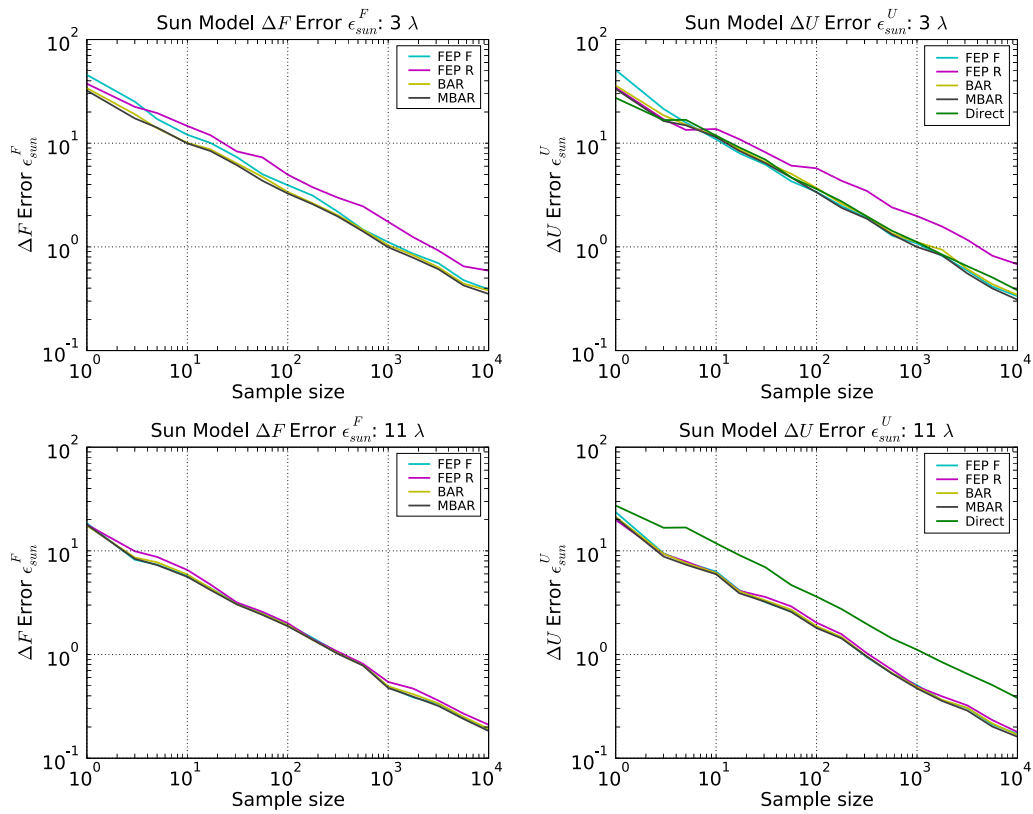


Figure 3.3: The Sun model errors  $\epsilon_{sun}^F$  and  $\epsilon_{sun}^U$  (see Eq. 3.24), which quantify the absolute error of the various estimators for different  $\lambda$  schedules. The 3- $\lambda$  schedule is (0.0, 0.5, 1.0), the 11- $\lambda$  schedule is (0.0, 0.1, ... 0.9, 1.0).

Most estimators perform similarly, yielding a series of coincident lines of a uniform slope on the log-log plot. As a result, BAR, MBAR and the FEP estimators yield approximately the same  $\epsilon_{sun}$  error for the same sample size  $N$ . Increasing the number of  $\lambda$  values from 3 to 11 results in lower errors, so that a Sun model simulation with the finer  $\lambda$  schedule needs to be only about 1/3 as long to achieve the same absolute error.

Even for this simple one-dimensional model we find differences in the performance of the different estimators. For the 3  $\lambda$  simulation, TI estimators for both  $\Delta F$  and  $\Delta U$  converge to an incorrect value because the trapezoidal rule is inaccurate for such a coarse schedule. The TI estimate performs on par with other methods for the finer  $\lambda$  schedule. Both FEP estimators, particularly the FEP R variant, tend to have larger errors than either BAR or MBAR, especially for the 3  $\lambda$  schedule. While the Direct  $\Delta U$  method yields errors on par with others for 3  $\lambda$ , its errors are significantly larger for the 11  $\lambda$  case. In fact, the Direct  $\Delta U$  results are identical in both plots, as it only uses data from the  $\lambda = 0$  and  $\lambda = 1$  simulations. Both BAR and MBAR estimators perform uniformly well, although under closer examination the MBAR method tends to have somewhat lower errors than BAR.

In order to establish consistency between the  $\epsilon_{sun}$  and standard errors, these same data were analyzed with block averaging (with 100 blocks). We find that the standard and  $\epsilon_{sun}$  errors are virtually identical; the only significant difference is that the standard error for TI for the 3  $\lambda$  system decreases together with the other estimators, whereas the  $\epsilon_{sun}$  error plateaus as the trapezoidal integration converges to an incorrect value. Aside from this, all the conclusions drawn from the Sun model  $\epsilon_{sun}$  error analysis hold for a block average analysis as well.

The one-dimensional Sun model was used to test the accuracy of different free energy and entropy/energy estimators. It also demonstrates that, in theory, it is possible to obtain  $\Delta S$  and  $\Delta U$  estimates that are as accurate as  $\Delta F$ , at least for simple one-dimensional systems. Next, we turn to assessing the performance of the new estimators in the context of solvation calculations.

## Solvation Calculations

Solvation calculations are considerably more complicated than the one-dimensional example discussed above, and we tested the performance of the estimators in this context by investigating the solvation process for N-methylacetamide (NMA). Table 3.2 lists the free energy of solvation for NMA and its decomposition using the various estimators, with errors estimated by block averaging. Note that TI estimates are not listed as  $\partial V/\partial\lambda$  were not computed.

	$\Delta F$ [kcal/mol]	$\Delta U$ [kcal/mol]	$T\Delta S$ [kcal/mol]
Direct	—	$-20.751 \pm 0.631$	—
FEP F	$-6.422 \pm 0.014$	$-22.167 \pm 0.933$	$-15.745 \pm 0.935$
FEP R	$-6.453 \pm 0.019$	$-22.251 \pm 1.541$	$-15.798 \pm 1.542$
BAR	$-6.430 \pm 0.015$	$-22.542 \pm 0.987$	$-16.111 \pm 0.987$
MBAR	$-6.431 \pm 0.015$	$-22.396 \pm 0.798$	$-15.965 \pm 0.798$

Table 3.2: N-methylacetamide results calculated with OPLS-AA parameters for various estimators, along with standard errors calculated with block averaging. Self and long range Lennard-Jones corrections applied.

All  $\Delta F$  estimators are relatively well converged, with statistical errors smaller than 0.02 kcal/mol. The errors in  $\Delta U$ , however, are 50-80 times larger. The Direct method appears to have the lowest statistical error, but its energy estimate is significantly different from the consensus of the other estimators, suggesting that significant bias error may underlie the smaller statistical error observed for this estimator.

The quantities  $\Delta F$  and  $\Delta U$  are constructed by the summation of  $\delta F$  and  $\delta U$  estimates, and the errors of the former are some unknown combination of the errors of the latter. To better understand the nature of the different estimators, and to identify regions of the  $\lambda$  schedule which contribute most to the collective error, we consider  $\delta F$  and  $\delta U$  as a function of  $\lambda$ . Figures 3.4 (a) and (b) plot  $\delta F/\delta\lambda$  and  $\delta U/\delta\lambda$  versus  $\lambda$ , respectively. The quantity  $\delta F/\delta\lambda$  is the finite difference approximation to  $\partial F/\partial\lambda$ , the derivative of the potential of mean force (Roux and Simonson, 1999), and  $\delta U/\delta\lambda$  is its energy decomposition.  $\Delta F$  and  $\Delta U$  then correspond to the areas underneath the curves of Fig. 3.4.

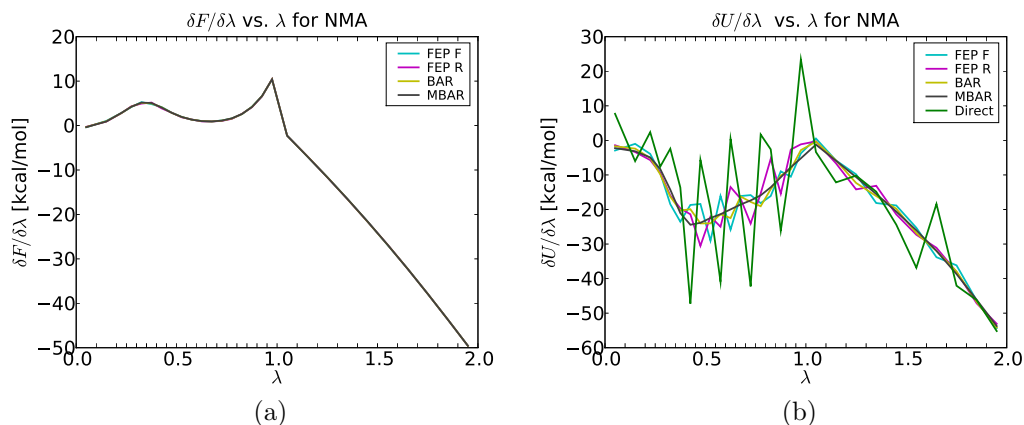


Figure 3.4: Plots of  $\delta F/\delta\lambda$  and  $\delta U/\delta\lambda$  for the solvation of NMA as obtained with various estimators. These quantities are effectively the derivatives of the potential of mean force, and its energy decomposition, along the  $\lambda$  schedule. For the free energies (panel a) all the estimators are consistent and the differences are nearly imperceptible. Significant differences among the estimators are observed for the energies (panel b), where the direct estimator is subject to large statistical error while the MBAR estimator yields a smooth profile.

In Fig. 3.4(a) we find that for all of the estimators,  $\delta F$  is converged to the point that differences between them are not discernible by eye, an observation consistent with the results of table 3.2. This is not the case for the energy plot, Fig. 3.4b, where there are large differences in the smoothness of the  $\delta U/\delta\lambda$  profile. The direct estimator has a serrated contour, suggesting a large degree of statistical error. The FEP estimators are somewhat smoother, followed by BAR, and the  $\delta U/\delta\lambda$  profile for the MBAR estimator is by far the smoothest of all.

The “smoothness” observed in Fig. 3.4 is quantified for  $\delta F$  in Fig. 3.5 and for  $\delta U$  in Fig. 3.6, where the statistical error of the estimates, obtained by block averaging, is plotted vs.  $\lambda$ . These figures illustrate the performance of estimators, as quantified by statistical error, versus  $\lambda$ . Four different schedules are considered: the 29  $\lambda$  schedule, as defined previously, has a uniform spacing of  $\delta\lambda = 0.1$  plus additional simulations in the region  $0.2 < \lambda < 1.0$ , so that  $\delta\lambda = 0.05$  in that region. The remaining plots have a uniform schedule across the  $\lambda$  range:  $\delta\lambda = 0.1$  for 21  $\lambda$ ;  $\delta\lambda = 0.2$  for 11  $\lambda$  and  $\delta\lambda = 0.5$  for 5  $\lambda$ . The latter three schedules are constructed from a subset of data obtained in the full 29  $\lambda$  simulation.

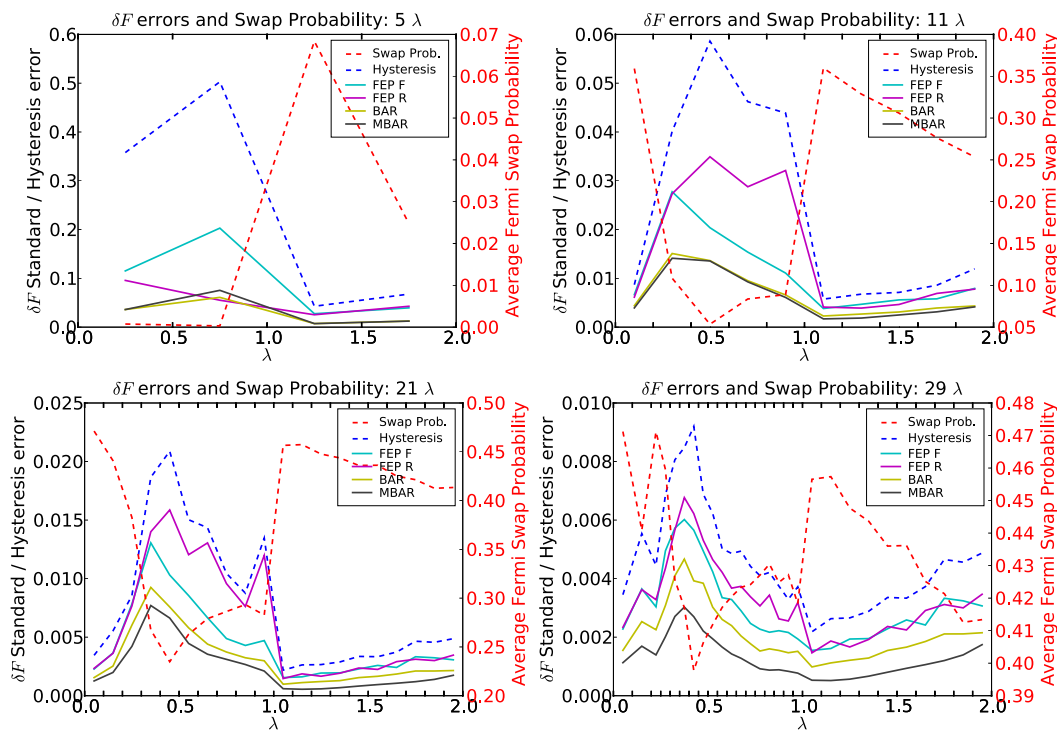


Figure 3.5:  $\delta F$  error diagnostics for the solvation of NMA under different  $\lambda$  schedules. Estimator standard errors are from 100-block averages, and the root-mean-square hysteresis error is obtained from these same blocks. The average Fermi swap probability is calculated from all data, and plotted using red scale on right. The different  $\lambda$  schedules, detailed in the text, are indicated with tick marks on top and bottom axes. All errors, including the hysteresis error, tend to be large in regions of the  $\lambda$  schedule where the swap probability is low.

Consistent with the results of Fig. 3.4, in Fig. 3.5 the MBAR  $\delta F$  estimator generally has the lowest standard error, followed in increasing order by BAR, FEP F and FEP R. The error of all estimators increases considerably with a coarsening of the  $\lambda$  schedule. The same observations hold for the  $\delta U$  results in Fig. 3.6, even though the statistical error of the  $\delta U$  estimates is roughly 100 times larger than for  $\delta F$ . Here, the direct method is considerably worse than all other estimators for the 29  $\lambda$  schedule, but it improves in relation to the other methods as the schedule coarsens. The error of the direct method in fact stays constant while other techniques deteriorate with fewer  $\lambda$  points.

In addition to the estimator standard error, Figs. 3.5 and 3.6 also plot the root-mean-square hysteresis error (calculated using 100 blocks per  $\lambda$  pair) as well as the average



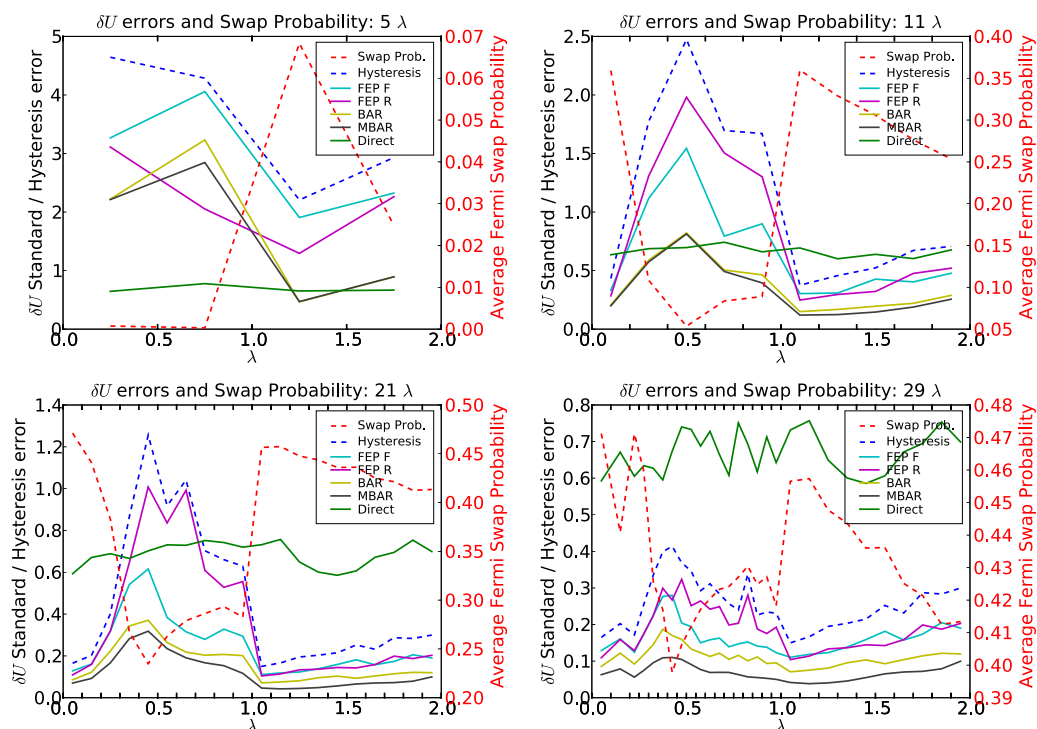


Figure 3.6:  $\delta U$  error diagnostics for the solvation of NMA under different  $\lambda$  schedules; see Fig. 3.5 for details. Like the  $\delta F$  errors in Fig. 3.5,  $\delta U$  errors for all estimators (except direct) are large in regions of the  $\lambda$  schedule where the swap probability is low. Additional simulations in these regions increase the swap probability and significantly reduce estimator errors. The direct estimator is independent of the schedule and becomes worse in relation to the other estimators as additional simulations are added to the  $\lambda$  schedule.

Fermi swap probability. Both quantities, described in section 3.2.2 and discussed in chapter 4, are useful diagnostics of convergence problems. The hysteresis error – defined as the difference between the forward and reverse FEP estimates – is a general indicator of a lack of equilibrium between two simulations. Its rate of convergence with simulation length is given by the average Fermi swap probability: when the swap probability is large the hysteresis error is low, and vice versa. This relationship is borne out in Fig. 3.5. Moreover, the swap probability is seen to be predictive not only of the hysteresis error, but of the statistical error of all estimators, for both  $\delta F$  and  $\delta U$ . All estimators converge more quickly in regions of the  $\lambda$  schedule where the swap probability is large.

### 3.3.2 Application to Model Compounds

Having established the utility and performance of the newly developed estimators on NMA, we carried out calculations to quantify  $\Delta F$ ,  $\Delta S$ , and  $\Delta U$  for 11 model compounds using the OPLS-AA and *m*OPLS-AA force fields. We also carried out comparisons between these calculated values and experimental data for the model compounds, allowing us to dissect the differences between the two different force field paradigms.

Table 3.4 shows the values obtained for  $\Delta F$ ,  $\Delta U$  and  $T\Delta S$  for each of the eleven model compounds using the two force fields, and table 3.3 gives the experimental data for these compounds. Figure 3.7 provides a visual comparison of the results for the two force fields, whereas Fig. 3.8 compares the computational values to experimental data.

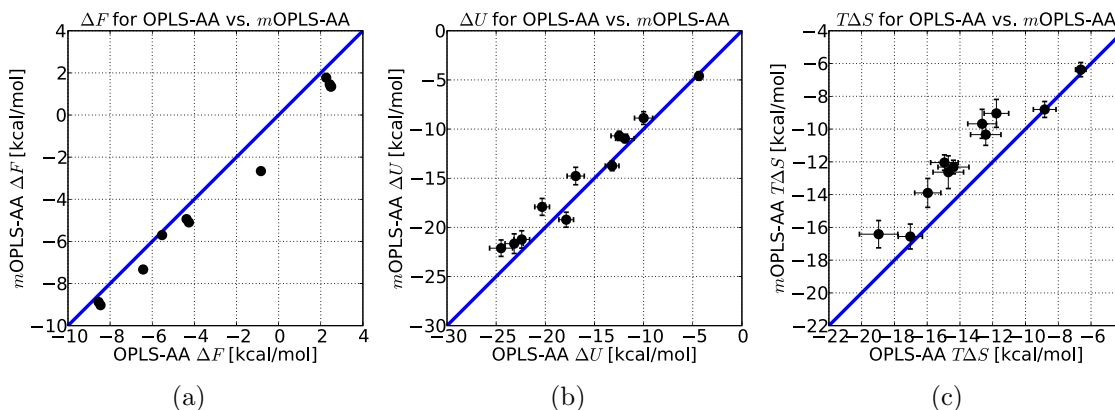


Figure 3.7: Comparison of thermodynamic quantities calculated using OPLS-AA and *m*OPLS-AA parameters for all compounds listed in table 3.4 using the MBAR estimators, with error bars indicating standard errors.

Data in table 3.4 and Fig. 3.7(a) indicates that compounds using the OPLS-AA parameters systematically have more positive – or less favorable – free energies of solvation than solutes with the *m*OPLS-AA parameters. As a result, agreement between computational  $\Delta F$  values and experimental data improves with the *m*OPLS-AA parameters for the polar compounds (acetamide, ethanol, methanol, p-cresol, N-methylacetamide, and propionamide). For nonpolar compounds, specifically linear

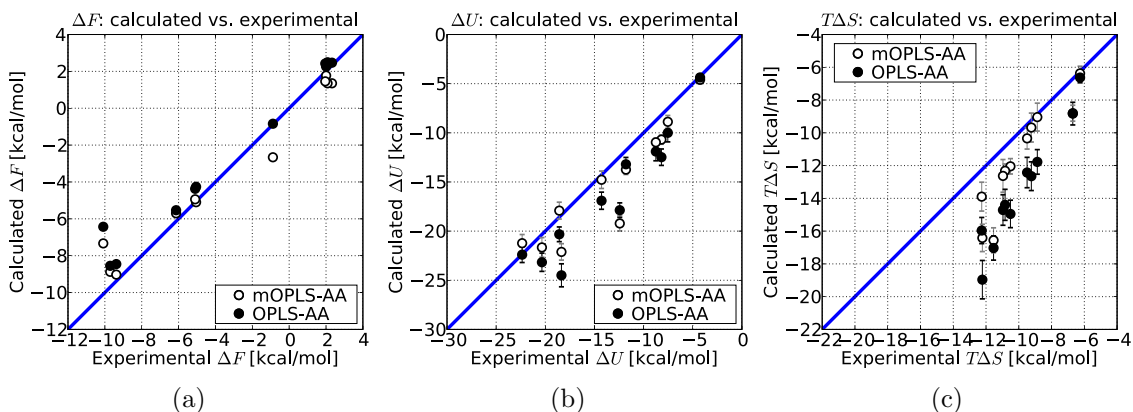


Figure 3.8: Comparison of calculated thermodynamic quantities to experimental values. Both OPLS-AA and *m*OPLS-AA results were calculated with MBAR.

hydrocarbons, the computational  $\Delta F$  values obtained using the *m*OPLS-AA parameters are consistently more negative than experimental values, whereas with OPLS-AA parameters they are generally more positive. Finally, the  $\Delta F$  value for toluene is too negative with the *m*OPLS-AA parameters while the OPLS-AA values approximately match experiment.

Panels (b) and (c) of Figs. 3.7 and 3.8 show the decompositions of  $\Delta F$  into  $T\Delta S$  and  $\Delta U$ . As with the case of NMA discussed previously, the statistical errors for the decompositions are about two orders of magnitude larger than they are for  $\Delta F$ , complicating the interpretation of results. Nevertheless, a few systematic trends are revealed in the analysis of the computational and experimental results. With the exception of toluene, the  $\Delta U$  values are systematically more negative using the OPLS-AA parameters as compared to values obtained using the *m*OPLS-AA parameters. Conversely, in all cases, the OPLS-AA parameters make more negative entropic contributions to the free energy of solvation.

Model Compound	Name	AA	Experimental				NVT Correction		
			$\Delta G$	$\Delta H$	$T\Delta S$	$(\Delta V)_P$	NVT	$\Delta U$	$T\Delta S$
acetamide	ACA	Asn	-9.71 <sup>a</sup>	-16.32	-6.61	55.82 <sup>c</sup>	2.259	-18.58	-8.87
ethyl alcohol	EOH	Thr	-5.05 <sup>b</sup>	-12.05	-7.00	55.12 <sup>c</sup>	2.231	-14.28	-9.23
isobutane	IBU	Leu	2.32 <sup>b</sup>	-4.83	-7.15	83.10 <sup>f</sup>	3.363	-8.19	-10.51
methane	MET	Ala	2.00 <sup>b</sup>	-2.75	-4.75	37.30 <sup>c</sup>	1.510	-4.26	-6.26
methyl alcohol	MOH	Ser	-5.10 <sup>b</sup>	-10.25	-5.16	38.25 <sup>c</sup>	1.548	-11.80	-6.70
n-butane	NBU	Ile	2.08 <sup>b</sup>	-5.66	-7.74	76.60 <sup>f</sup>	3.100	-8.76	-10.84
N-methylacetamide	NMA	BB	-10.09 <sup>d</sup>	-19.36	-9.27	74.04 <sup>c</sup>	2.996	-22.36	-12.26
4-methyl phenol	PCR	Tyr	-6.14 <sup>c</sup>	-14.18	-8.05	103.23 <sup>e</sup>	4.178	-18.36	-12.22
propionamide	PPA	Gln	-9.38 <sup>a</sup>	-17.45	-8.07	71.54 <sup>c</sup>	2.895	-20.35	-10.97
n-propane	PRP	Val	1.96 <sup>b</sup>	-4.83	-6.78	67.00 <sup>c</sup>	2.711	-7.54	-9.49
toluene	TOL	Phe	-0.88 <sup>b</sup>	-8.10	-7.21	106.86 <sup>g</sup>	4.325	-12.43	-11.54

Table 3.3: Free energy, enthalpy and entropy ( $\Delta G$ ,  $\Delta H$  and  $T(\Delta S)_P$ ) in kcal/mol, partial molar volume  $(\Delta V)_P$  in cm<sup>3</sup>/mol. Model compound names, abbreviations (see Fig. 3.1) and corresponding amino acids (“BB” is backbone mimic). All values reported in Ben-Naim convention (Ben-Naim and Marcus, 1984). NVT corrections are applied to obtain  $\Delta U$  and  $T(\Delta S)_V$  values, and are given by  $(\Delta V)_P T\alpha/\kappa$  (see Eq. 3.22), with the correction factor  $T\alpha/\kappa = 0.04047[\text{kcal} / \text{cm}^3]$  based on experimental water properties from Jorgensen and Jenson (1998).

<sup>a</sup> Avbelj et al. 2000

<sup>b</sup> Ben-Naim and Marcus 1984

<sup>c</sup> Cabani et al. 1981

<sup>d</sup> Graziano 2000

<sup>e</sup> Hnedkovsky et al. 1998

<sup>g</sup> Moore et al. 1982

<sup>h</sup> Wilhelm et al. 1977

## 3.4 Discussion

### 3.4.1 Thermodynamic Estimators

The simplicity of the direct estimator is appealing, and its statistical error measures in table 3.2 are small. However, it deviates significantly from the consensus of all other estimators, making its performance difficult to gauge by this analysis alone. The deficits of the direct energy estimator are more apparent in Figs. 3.4 – 3.6. Since the direct estimator is independent of the schedule, its error for all  $\delta U$  is about the same as the  $\Delta U$  error. Where  $\Delta U$  of a process is large, the direct estimator may be a good choice, particularly if the  $\lambda$  schedule is coarse. For energy changes on the

**Computational**

Name	FF	$\Delta F$	$\Delta U$	$T\Delta S$
ACA	OPLS-AA	$-8.549 \pm 0.015$	$-20.324 \pm 0.747$	$-11.776 \pm 0.750$
ACA	<i>m</i> OPLS-AA	$-8.875 \pm 0.016$	$-17.916 \pm 0.857$	$-9.042 \pm 0.851$
EOH	OPLS-AA	$-4.260 \pm 0.013$	$-16.908 \pm 0.872$	$-12.648 \pm 0.876$
EOH	<i>m</i> OPLS-AA	$-5.098 \pm 0.016$	$-14.777 \pm 0.886$	$-9.679 \pm 0.883$
IBU	OPLS-AA	$2.471 \pm 0.011$	$-12.480 \pm 0.835$	$-14.951 \pm 0.838$
IBU	<i>m</i> OPLS-AA	$1.353 \pm 0.015$	$-10.691 \pm 0.469$	$-12.045 \pm 0.469$
MET	OPLS-AA	$2.259 \pm 0.011$	$-4.368 \pm 0.324$	$-6.627 \pm 0.319$
MET	<i>m</i> OPLS-AA	$1.772 \pm 0.006$	$-4.599 \pm 0.431$	$-6.371 \pm 0.434$
MOH	OPLS-AA	$-4.371 \pm 0.012$	$-13.200 \pm 0.694$	$-8.829 \pm 0.691$
MOH	<i>m</i> OPLS-AA	$-4.941 \pm 0.014$	$-13.744 \pm 0.494$	$-8.803 \pm 0.486$
NBU	OPLS-AA	$2.493 \pm 0.011$	$-11.902 \pm 0.944$	$-14.396 \pm 0.946$
NBU	<i>m</i> OPLS-AA	$1.353 \pm 0.017$	$-10.966 \pm 0.413$	$-12.319 \pm 0.415$
NMA	OPLS-AA	$-6.431 \pm 0.015$	$-22.396 \pm 0.798$	$-15.965 \pm 0.798$
NMA	<i>m</i> OPLS-AA	$-7.333 \pm 0.011$	$-21.230 \pm 0.883$	$-13.896 \pm 0.881$
PCR	OPLS-AA	$-5.533 \pm 0.021$	$-24.497 \pm 1.169$	$-18.964 \pm 1.171$
PCR	<i>m</i> OPLS-AA	$-5.700 \pm 0.017$	$-22.116 \pm 0.835$	$-16.416 \pm 0.833$
PPA	OPLS-AA	$-8.455 \pm 0.015$	$-23.167 \pm 0.929$	$-14.712 \pm 0.931$
PPA	<i>m</i> OPLS-AA	$-9.027 \pm 0.012$	$-21.658 \pm 0.998$	$-12.631 \pm 0.999$
PRP	OPLS-AA	$2.426 \pm 0.009$	$-9.999 \pm 0.924$	$-12.425 \pm 0.927$
PRP	<i>m</i> OPLS-AA	$1.458 \pm 0.013$	$-8.881 \pm 0.646$	$-10.339 \pm 0.653$
TOL	OPLS-AA	$-0.841 \pm 0.016$	$-17.872 \pm 0.745$	$-17.031 \pm 0.739$
TOL	<i>m</i> OPLS-AA	$-2.661 \pm 0.013$	$-19.220 \pm 0.763$	$-16.559 \pm 0.760$

Table 3.4: Results of the solvation calculations as calculated using the MBAR method with long range Lennard-Jones and self corrections applied.  $\Delta U$  and  $T(\Delta S)_V$  columns give the NVT results, and  $\Delta H$  and  $T(\Delta S)_P$  columns are the NPT values (with NVT corrections applied). Helmholtz ( $\Delta F$ ) and Gibbs ( $\Delta G$ ) free energies do not differ numerically.

Name	FF	Uncorrected			Self Correction			LJLR
		$\Delta F$	$\Delta U$	$T\Delta S$	$\Delta F$	$\Delta U$	$T\Delta S$	
ACA	OPLS-AA	-37.828	-49.612	-11.784	-29.356	-29.364	-0.008	-0.077
ACA	<i>m</i> OPLS-AA	-9.202	-18.244	-9.042	-0.384	-0.384	-0.000	-0.056
EOH	OPLS-AA	-2.961	-15.653	-12.692	1.240	1.196	-0.044	-0.059
EOH	<i>m</i> OPLS-AA	-5.198	-14.913	-9.714	-0.160	-0.195	-0.035	-0.059
IBU	OPLS-AA	4.102	-11.183	-15.285	1.544	1.210	-0.334	-0.087
IBU	<i>m</i> OPLS-AA	0.827	-11.241	-12.068	-0.600	-0.624	-0.024	-0.074
MET	OPLS-AA	2.286	-4.341	-6.627	0.000	0.000	0.000	-0.027
MET	<i>m</i> OPLS-AA	1.793	-4.577	-6.371	0.000	0.000	0.000	-0.021
MOH	OPLS-AA	0.777	-8.052	-8.829	5.109	5.109	-0.000	-0.039
MOH	<i>m</i> OPLS-AA	-4.974	-13.777	-8.803	-0.068	-0.068	-0.000	-0.034
NBU	OPLS-AA	4.825	-9.921	-14.746	2.244	1.894	-0.350	-0.087
NBU	<i>m</i> OPLS-AA	1.105	-11.403	-12.507	-0.322	-0.511	-0.189	-0.074
NMA	OPLS-AA	-22.052	-38.062	-16.009	-15.722	-15.766	-0.045	-0.100
NMA	<i>m</i> OPLS-AA	-7.487	-21.368	-13.882	-0.227	-0.212	0.015	-0.073
PCR	OPLS-AA	-6.259	-25.126	-18.867	-0.863	-0.766	0.097	-0.137
PCR	<i>m</i> OPLS-AA	-5.942	-22.358	-16.416	-0.347	-0.347	-0.000	-0.104
PPA	OPLS-AA	-38.154	-52.924	-14.770	-29.795	-29.853	-0.058	-0.097
PPA	<i>m</i> OPLS-AA	-9.085	-21.906	-12.822	-0.132	-0.322	-0.191	-0.073
PRP	OPLS-AA	4.983	-7.580	-12.563	2.490	2.352	-0.138	-0.067
PRP	<i>m</i> OPLS-AA	1.180	-9.163	-10.344	-0.334	-0.339	-0.004	-0.056
TOL	OPLS-AA	4.442	-12.590	-17.032	5.157	5.157	-0.000	-0.126
TOL	<i>m</i> OPLS-AA	-2.904	-19.462	-16.559	-0.334	-0.334	0.000	-0.091

Table 3.5: Raw data, corrections and final values for free energy and entropy/enthalpy, calculated using MBAR. LJLR is the long range Lennard-Jones correction.

order of thermal fluctuations – typical for solvation calculations – this estimator’s slow convergence makes its use inefficient. This is especially true for the decomposition of the mean force along a fine  $\lambda$  schedule, as in Fig. 3.4b, where  $\delta U$  is necessarily small between neighboring simulations; here, the direct estimator is nearly useless and other methods, especially MBAR, perform exceptionally well.

The FEP estimators for both  $\Delta F$  and  $\Delta U$  generally suffer from significant errors which decrease with an improved  $\lambda$  schedule. In general, FEP R is worse than FEP F, a known result which stems from the fact that the insertion of a particle (forward direction) constrains the phase space of a system and leads to faster convergence. This is true for both  $\Delta F$  as well as  $\Delta U$  calculations (Lu et al., 2003a; Wu and Kofke, 2004).

Finally, both the BAR and the MBAR methods perform consistently well for both  $\Delta F$  and  $\Delta U$  calculations, with MBAR outperforming BAR in all cases. The distinction between the two methods becomes more pronounced with an increasing number of simulations along the  $\lambda$  schedule, where the MBAR estimator can take advantage of information from non-neighbor simulations to improve its estimate. For a coarse schedule, non-neighboring ensembles are sufficiently different that little additional information can be gleaned from them and BAR, which considers only neighboring simulations, performs similarly to MBAR.

The  $\Delta F$  MBAR estimator was derived to minimize the variance of the estimate given a finite number of samples from simulations at multiple  $\lambda$  values. The present MBAR  $\Delta U$  estimator is a temperature derivative of the  $\Delta F$  formula, with no explicit variance minimization. It is encouraging that the  $\Delta U$  estimator likewise has a low variance, although conceivably a distinct  $\Delta U$  estimator, which minimizes the energy variance explicitly, may exist and yield even better performance.

### 3.4.2 Convergence Rates

In the solvation calculations, the standard errors of  $\Delta U$  are 50-100 times larger than  $\Delta F$ , consistent with observations in the literature (Trzesniak and van Gunsteren,

2006). For the Sun model, however, the errors of the two quantities are nearly identical. We can understand the difference between the systems by considering the source of the distinct  $\Delta F$  and  $\Delta U$  convergence rates (Lu et al., 2003a).

The vast majority of the interactions in the solvation calculations are between pairs of water molecules, which dominate the potential energy of the system and are independent of  $\lambda$ . The thermal fluctuations of these interactions swamp out the relatively small energy changes due to the scaled potential, requiring long simulations to make out the signal from the scaling process. By contrast, free energy calculations generally consider ensemble averages of  $\delta v$ , where the contributions which are independent of  $\lambda$  tend to cancel. Consequently, the free energies for the solvation calculations converge much more quickly than do the energies. (This remains true for all the estimators considered here, including MBAR.) In the case of the Sun model, however, this distinction does not hold; there, the  $\lambda$  scaling directly affects the sole degree of freedom, and both the free energy and the energy converge at the same rate.

### 3.4.3 Swap Probabilities

As discussed in chapter 4, the average replica exchange swap probability – and in particular, the Fermi swap probability – indicates the relative rate of convergence of the hysteresis error. When the swap probability is low, free energy calculations converge slowly and the hysteresis error is large, whereas for high swap probabilities, convergence is more rapid. While the hysteresis error is based on the free energy perturbation estimators, we find that it is an indicator of convergence problems in all estimators, both  $\Delta F$  and  $\Delta U$ , as illustrated in Figs. 3.5 and 3.6.

The swap probability is a particularly useful diagnostic of simulation convergence. It is computed directly from a simulation without the need for statistical error analysis, and has a ready interpretation as the overlap between two ensembles. Two simulations are required in order to calculate the swap probability, but we can approximate it by considering the variance of  $\partial V/\partial\lambda$  obtained from only one simulation. As a result, it is possible to improve a  $\lambda$  schedule, based on preliminary simulations, by placing additional simulations at  $\lambda$  values either where the swap probability is low or where  $\text{var}(\partial V/\partial\lambda)$  is large. As an example, the 21  $\lambda$  and 29  $\lambda$  schedules in Figs. 3.5 and



3.6 differ by the addition of extra simulations where the swap probability (in the 21  $\lambda$  case) is low. This increases the swap probability considerably in this region and results in a significant decrease of the errors of all the estimators.

While no Hamiltonian replica exchange swaps need to take place in order to evaluate the swap probabilities, actually performing such swaps is an additional effective way of accelerating the convergence, and decreasing the errors, of thermodynamic estimators, as discussed in chapter 4.

### 3.4.4 Solvation Calculations

Computational values for  $\Delta F$ , particularly those based on the MBAR estimator, are generally statistically reliable and can be used to make quantitative comparisons to experimental data (Shirts and Pande, 2005a). The thermodynamic decompositions into entropies and energies, on the other hand, are typically two orders of magnitude larger, irrespective of the specific model compound (Lu et al., 2003a). Despite the lower reliability of the thermodynamic decompositions, we can discern trends to rationalize the  $\Delta F$  comparisons between experimental data and the computational values.

Based on Fig. 3.7 and table 3.4, we find that

- $\Delta F_{OPLS-AA} > \Delta F_{mOPLS-AA}$  (*mOPLS-AA* more favorable),
- $\Delta U_{OPLS-AA} < \Delta U_{mOPLS-AA}$  (*OPLS-AA* more favorable), and
- $T\Delta S_{OPLS-AA} < T\Delta S_{mOPLS-AA}$  (*mOPLS-AA* more favorable).

Thus it appears that the source of the improved overall agreement between experimental data and computational values obtained using the *mOPLS-AA* force field is due to the reduced entropic penalty associated with the modified force field.

These trends can be partly understood on the basis of the force field Lennard-Jones parameters. The  $\sigma$  parameters are smaller for the *mOPLS-AA* parameter set, resulting in smaller atomic volumes from which the water must be displaced. This results in

a less negative solvation entropy, which is governed by the cavity size. This difference between the OPLS-AA and *m*OPLS-AA Lennard-Jones parameters is consistent with previous work (Gallicchio et al., 2000) and theoretical predictions (Pratt and Chandler, 1977). Larger  $\epsilon$  parameters increase the attractive dispersion forces between the solute and water and result in more negative energy changes, which are dominated by dispersive interactions. Depending on the atom, the  $\epsilon$  parameters for *m*OPLS-AA can be larger or smaller than for OPLS-AA, with most molecules a mix of increased and reduced dispersion forces; this prevents a straightforward interpretation of the observation that  $\Delta U$  for *m*OPLS-AA tends to be more positive and less favorable.

An illustrative example is that of NMA, a model compound of relevance for polypeptide backbones. While the computational  $\Delta F$  values for both force fields disagree with experimental data, the disagreement is smaller for the *m*OPLS-AA force field. The entropy-energy decomposition shows that both force fields yield similar estimates, within error, for  $\Delta U$ , and that this result is close to the experimental value of -22.4 kcal/mol. The main contribution to the difference in  $\Delta F$  values comes from the entropic terms: for OPLS-AA  $T\Delta S = -16.0 \pm 0.8$  kcal/mol, whereas for *m*OPLS-AA  $T\Delta S = -13.9 \pm 0.9$  kcal/mol. Both these values are more negative than the experimental value of -12.3 kcal/mol, but the *m*OPLS-AA is closer, resulting in more accurate  $\Delta F$  estimate, and suggesting directions for further force field refinement. This example highlights the need for a more holistic approach to parameter optimization, and provides a cautionary note regarding efforts that focus primarily on the reparameterization of partial charges (Udier-Blagović et al., 2004).

## 3.5 Derivations

### 3.5.1 Definitions and General Identities

The free energy, energy, entropy and potential energy of simulation  $i$  –  $F_i$ ,  $U_i$ ,  $S_i$  and  $V_i$ , respectively – are normalized by the inverse temperature  $\beta_i = (k_B T_i)^{-1}$ , to yield the nondimensional forms of these quantities,

$$f_i = \beta_i F_i, \quad u_i = \beta_i U_i, \quad s_i = \beta_i S_i, \quad v_i = \beta_i V_i. \quad (3.25a)$$

With  $\partial F_i/\partial T = -S_i$ , temperature derivatives of these include,

$$\frac{\partial f_i}{\partial T} = -\frac{f_i}{T_i} - s_i, \quad \frac{\partial v_i}{\partial T} = -\frac{v_i}{T_i}, \quad (3.25b)$$

and with  $\delta f = f_1 - f_0$ ,

$$\frac{\partial \delta f}{\partial T} = -\frac{1}{T_0}(\tau f_1 - f_0 + T_0 \delta s), \quad (3.25c)$$

where  $\tau = T_1/T_0$ .

The Boltzmann probability distribution,

$$\rho_i(x) = \frac{\exp(-v_i(x))}{\int dy \exp(-v_i(y))} \quad (3.26a)$$

has the temperature derivative,

$$\frac{\partial \rho_i(x)}{\partial T} = \frac{\rho_i(x)}{T_i} (v_i - \langle v_i \rangle_i) \quad (3.26b)$$

with  $\langle A \rangle_i = \int dx \rho_i(x) A(x)$ .

The Fermi function is defined (Bennett, 1976) as,

$$g(a) = 1/[1 + \exp(a)] \quad (3.27a)$$

and its derivative is,

$$\frac{\partial g(a)}{\partial T} = -\frac{1}{1 + \exp(a)} \frac{\exp(a)}{1 + \exp(a)} \frac{\partial a}{\partial T} = -g(a)g(-a) \frac{\partial a}{\partial T} \quad (3.27b)$$

### 3.5.2 BAR

Bennett (1976) derived the following free energy estimator (eq. (9) in that reference),

$$\frac{Q_0}{Q_1} = \frac{\langle g(v_0 - v_1 + C) \rangle_1}{\langle g(v_1 - v_0 - C) \rangle_0} \exp(C) \quad (3.28a)$$

$$C = \ln \frac{Q_0 N_1}{Q_1 N_0} \quad (3.28b)$$

with  $\delta f = \ln(Q_0/Q_1)$ , which holds for an arbitrary number of observations per simulation  $N_i$  and temperatures  $T_i$ . Equations (3.28) are typically solved iteratively until  $C$  converges. Here, however, we substitute (3.28b) into (3.28a) directly to obtain,

$$N_0 \langle g_+ \rangle_0 = N_1 \langle g_- \rangle_1 \quad (3.29a)$$

with

$$g_+(x) \equiv [1 + \exp(+v_1(x) - v_0(x) - \delta f - \ln N_1/N_0)]^{-1}, \quad (3.29b)$$

$$g_-(x) \equiv [1 + \exp(-v_1(x) + v_0(x) + \delta f + \ln N_1/N_0)]^{-1}. \quad (3.29c)$$

Equations (3.10) are specific to  $N_0 = N_1$  and  $T_0 = T_1$ , as indicated by the \* symbol.

We obtain the entropy equation by taking the temperature derivative of (3.29a) and multiplying by  $T_0$ ,

$$T_0 \frac{\partial}{\partial T} \left\{ N_0 \int dx \rho_0(x) g_+(x) = N_1 \int dx \rho_1(x) g_-(x) \right\}, \quad (3.30)$$

which, with identities from section 3.5.1, becomes

$$\begin{aligned} & N_0 [\langle g_+ v_0 \rangle_0 - \langle g_+ \rangle_0 \langle v_0 \rangle_0 + \langle g_+ g_- (\tau v_1 - v_0) \rangle_0 - (\tau f_1 - f_0 + T_0 \delta s) \langle g_+ g_- \rangle_0] \\ & = N_1 [\tau \langle g_- v_1 \rangle_1 - \tau \langle g_- \rangle_1 \langle v_1 \rangle_1 - \langle g_+ g_- (\tau v_1 - v_0) \rangle_1 + (\tau f_1 - f_0 + T_0 \delta s) \langle g_+ g_- \rangle_1]. \end{aligned} \quad (3.31)$$

Solving for  $\delta s$  yields,

$$T_0 \delta s = \frac{N_0 \alpha_0 - N_1 \alpha_1}{N_0 \langle g_+ g_- \rangle_0 + N_1 \langle g_+ g_- \rangle_1} + f_0 - \tau f_1, \quad (3.32a)$$

where

$$\alpha_0 = \langle g_+ v_0 \rangle_0 - \langle g_+ \rangle_0 \langle v_0 \rangle_0 + \langle g_+ g_- (\tau v_1 - v_0) \rangle_0, \quad (3.32b)$$

$$\alpha_1 = \tau \langle g_- v_1 \rangle_1 - \tau \langle g_- \rangle_1 \langle v_1 \rangle_1 - \langle g_+ g_- (\tau v_1 - v_0) \rangle_1. \quad (3.32c)$$

Equations (3.32) reduce to Eqs. (3.11) for  $N_0 = N_1$  and  $T_0 = T_1$ .

### 3.5.3 MBAR

The MBAR free energy estimator is defined as Eq. (11) in (Shirts and Chodera, 2008),

$$f_i = -\ln \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{\exp[-v_i(x_{jn})]}{\sum_{k=1}^K N_k \exp[f_k - v_k(x_{jn})]} \quad \text{for } i = 1, 2, \dots, K, \quad (3.33)$$

where  $K$  is the number of simulations and  $N_j$  is the number of observations drawn from the  $j^{\text{th}}$  simulation. Substituting the empirical estimator  $\sum_n^{N_j} A_n = N_j \langle A \rangle_j$  and defining,

$$\mu_i(x) \equiv \frac{\exp(f_i - v_i(x))}{\sum_k^K N_k \exp(f_k - v_k(x))}, \quad (3.34a)$$

Eq. (3.33) can be written,

$$\sum_j^K N_j \langle \mu_i \rangle_j = 1 \quad \text{for } i = 1, 2, \dots, K, \quad (3.34b)$$

with  $x$  is drawn from probability distribution corresponding to  $v_j(x)$ .

We obtain the MBAR entropy estimator from the temperature derivative of (3.34b),

$$T_i \frac{\partial}{\partial T} \left\{ \sum_j^K N_j \int dx \rho_j(x) \mu_i(x) = 1 \right\}. \quad (3.35)$$

With the identity

$$\frac{\partial \mu_i}{\partial T} = -\frac{1}{T_i} \left[ (f_i + T_i s_i) \mu_i - \mu_i v_i - \sum_k^K N_k \frac{T_i}{T_k} \{ (f_k + T_k s_k) \mu_i \mu_k - \mu_i \mu_k v_k \} \right], \quad (3.36)$$

we find

$$\begin{aligned}
T_i s_i &= -f_i + \sum_j^K N_j \langle \mu_i v_i \rangle_j + \sum_j^K N_j \frac{T_i}{T_j} (\langle \mu_i v_j \rangle_j - \langle \mu_i \rangle_j \langle v_j \rangle_j) \\
&+ \sum_j^K \sum_k^K N_j N_k \frac{T_i}{T_k} [(f_k + T_k s_k) \langle \mu_i \mu_k \rangle_j - \langle \mu_i \mu_k v_k \rangle_j] \quad \text{for } i = 1, 2, \dots, K,
\end{aligned} \tag{3.37}$$

Equations (3.12) and (3.14) in the body of the paper correspond to equations (3.34) and (3.38) for all  $N_i$  and  $T_i$  the same.

### 3.5.4 Sun model

Given a one-dimensional potential  $V(x, \lambda)$  free energy is given as,

$$F(\lambda) = -\beta^{-1} \ln \int_{-\infty}^{\infty} dx \exp[-\beta V(x, \lambda)] \tag{3.38}$$

and the free energy change  $\Delta F = F(1) - F(0)$ . For the Sun model (3.23), we can integrate Eq. (3.38) and obtain  $\Delta F$  analytically (Oberhofer et al., 2005; Nummela et al., 2008),

$$\Delta F = \beta^{-1} \ln \left( \sqrt{2\pi} e^{32\beta} \left( I_{-\frac{1}{4}}(32\beta) + I_{\frac{1}{4}}(32\beta) \right) \right) - \beta^{-1} \ln \left( \frac{2\Gamma\left(\frac{5}{4}\right)}{\sqrt[4]{\beta}} \right) \tag{3.39}$$

where  $\Gamma$  is the gamma function and  $I_n$  is a modified Bessel function of the first kind. We can obtain the energy change by differentiating,

$$\begin{aligned}
\Delta U &= \frac{\partial}{\partial \beta} \beta \Delta F \\
&= \frac{32 \left( I_{-\frac{3}{4}}(32\beta) + I_{-\frac{1}{4}}(32\beta) + I_{\frac{1}{4}}(32\beta) + I_{\frac{3}{4}}(32\beta) \right)}{I_{-\frac{1}{4}}(32\beta) + I_{\frac{1}{4}}(32\beta)}
\end{aligned} \tag{3.40}$$

For  $\beta = 0.02$ ,  $\Delta F = 65.8878$ ,  $\Delta U = 53.1957$ , and  $T\Delta S = -12.6921$ . The analytical derivations for the Sun model were performed with Mathematica (Wolfram Research, 2008).

# Chapter 4

## Replica Exchange in Free Energy Calculations

### 4.1 Preamble

In this chapter we show how Hamiltonian replica exchange and an improved  $\lambda$  schedule can speed the convergence of free energy calculations. Two concepts – the Crooks fluctuation theorem and the hysteresis error – are central to the discussion.

The Crooks fluctuation theorem (Crooks, 1999), defined in the next section, is a relationship between two equilibrium simulations at different  $\lambda$  values. Much like the Boltzmann distribution upon which it is based, finite simulations satisfy this relationship only approximately. We use the Crooks fluctuation theorem as a measure of how well two simulations are equilibrated with respect to one another, as it detects convergence problems – e.g. one simulation stuck in a local minimum – which are not detectable from an analysis of one simulation alone.

The hysteresis error is one way to characterize the divergence of two simulations from the ideal expected by the Crooks fluctuation theorem; it is, essentially, a measure of the departure of a pair of simulations from mutual equilibrium (see also section 2.4.3). We show how Hamiltonian replica exchange introduces a Monte Carlo move to directly minimize this error, and that replica exchange swap probabilities between two simulations characterize the rate at which the hysteresis error decreases with increasing simulation length. Based on this, a prescription for finding the optimal

$\lambda$  schedule – one for which the hysteresis error decreases uniformly quickly for all neighboring simulation pairs – is constructed.

In the context of optimizing simulations, this chapter focuses on the hysteresis error without discussing free energy calculations *per se*. The hysteresis error is in fact operationally defined as the difference between the forward and reverse free energy perturbation  $\Delta F$  estimates. Moreover, as discussed herein and illustrated in section 3.3.1, all  $\Delta F$  and  $\Delta U$  estimators, not only FEP, suffer from convergence problems in specific sections of the  $\lambda$  schedule for the same underlying reason – a rapid change with respect to  $\lambda$  of the equilibrium configuration distributions. Consequently, reducing the hysteresis error is tantamount to reducing errors of all  $\Delta F$  and  $\Delta U$  estimators.

Sections 4.2 through 4.9 have been published previously as Wyczalkowski and Pappu 2008. We conclude this chapter with two unpublished results stemming from this work: a critical evaluation of the bootstrap technique for estimating standard errors based on data from only one simulation, and an illustration of the effect of Hamiltonian replica exchange swap rate on  $\Delta F$  and  $\Delta U$  errors.

## 4.2 Introduction and Overview

Free energies of solvation provide quantitative assessments of driving forces for spontaneous processes such as protein folding, binding, self-assembly, and solubility. Formally, the free energy of solvation in the canonical ensemble is the free energy change  $\Delta F$  associated with the transfer of a solute from the gas phase to a fixed position in the solvent (Ben-Naim, 1987). Operationally, one has access to a range of techniques to obtain estimates for  $\Delta F$  (Ytreberg et al., 2006; Levy and Gallicchio, 1998). Kirkwood (1935) showed that one could introduce arbitrary parameters into potential functions and continuously vary the degree of coupling between specific molecules in a dense fluid. The device of coupling parameters leads to simple expressions for chemical potentials of any component of the fluid. If the component is the solute molecule, which is transferred from the gas phase into the solvent, then a single coupling parameter  $\lambda$ , where  $0 \leq \lambda \leq 1$ , modulates solute-solvent interactions in the system’s potential function. The limits  $\lambda = 0$  and  $\lambda = 1$  correspond to the pure solvent and solvent plus fully grown solute, respectively. Intermediate values of  $\lambda$  correspond to



potential functions that include only a part of the solute-solvent interactions. The Kirkwood coupling parameter plays a central role in equilibrium methods for calculating  $\Delta F$ . One carries out a series of independent canonical simulations where each simulation is associated with a distinct potential function, characterized by a specific  $\lambda$  value. As it samples the equilibrium ensemble, each simulation generates a series of work values, which are then used to estimate the free energy change across the entire  $\lambda$  schedule.

The multicanonical approach described above takes advantage of the simple formalism developed by Kirkwood for calculating  $\Delta F$ . However, in practice, standard free energy calculations based on multicanonical simulations are plagued by slow convergence and inaccurate estimates of  $\Delta F$  (Mobley et al., 2007). Errors may be divided into statistical and bias (or finite sampling) errors (Lu and Kofke, 2001a). The former stem from the fluctuations of the free energy estimator, and can be estimated by block averaging or bootstrap methods (Frenkel and Smit, 2002; Efron and Tibshirani, 1993). Since the statistical error decreases as the inverse square root of simulation length, it is frequently used as an indicator of the convergence of the multicanonical simulation. While statistical errors are random fluctuations of short simulation results about some mean value, the bias error is an error of the mean value itself, and it changes with simulation length. As discussed by Zuckerman and Woolf (2004), bias errors have two causes: the free energy estimates are nonlinear averages; and, the work distributions on which such estimates are based will typically have long tails which are rarely sampled, and yet these are important to the average. The latter point is important: rare events dominate free energy estimates, and one seldom observes these events in short simulations. As a result, the average drifts with simulation length, resulting in inaccurate estimates for  $\Delta F$  from bias error even when the statistical error is small. The magnitude of the bias error is difficult to quantify directly, as it requires knowledge of the actual free energy difference, the very quantity we wish to determine. Furthermore, small fluctuations in the estimate for  $\Delta F$  may not be indicative of convergence, but rather of inadequate sampling of the rare but important configurations. To address these problems, we develop an alternate measure of free energy error, one based on deviations from equilibrium distributions.

Crooks (1999) derived a fluctuation theorem (section 4.8.1) valid for stochastic, microscopically reversible dynamics, which relates the distribution of dissipated work

values along a forward and reverse path as,

$$\exp(\beta W_D) = \frac{P_F(\beta W_D)}{P_R(-\beta W_D)}. \quad (4.1)$$

Here,  $\beta = (k_B T)^{-1}$ ,  $P_F(\beta W_D)$  is the probability distribution for dissipated work associated with switching  $\lambda$  from  $\lambda_0$  to  $\lambda_1$ , and  $P_R(-\beta W_D)$  is the corresponding distribution for the reverse process. If the canonical simulations for each value of  $\lambda$  sample the equilibrium ensemble adequately, then the distributions of dissipated work obtained over the course of free energy calculations will satisfy Eq. (4.1).

In this work, we develop a readily measured error estimate, the hysteresis error  $\epsilon_H$ , which quantifies the degree to which observed work distributions obey the Crooks fluctuation theorem. Hamiltonian replica exchange, a multicanonical equilibration technique, effectively reduces the hysteresis error. We relate the average replica exchange swap probability to the degree of overlap between equilibrium ensembles, as well as to the rate at which  $\epsilon_H$  falls. Based on this, we may construct an optimized  $\lambda$  schedule to further minimize the hysteresis error for an entire simulation.

The remainder of this presentation is organized as follows: the theory section introduces the hysteresis error in the context of the Crooks fluctuation theorem followed by a formal illustration of how Hamiltonian replica exchange minimizes  $\epsilon_H$ ; the definition of swap probability as a measure of the overlap between different equilibrium ensembles; and a connection between the amount of overlap and minimization of  $\epsilon_H$ . We calculate the free energy of hydration for acetamide to demonstrate how to estimate  $\epsilon_H$  and minimize this error using replica exchange coupled to standard multicanonical simulations. We conclude with a summary and a discussion of the features of our methodology.

## 4.3 Theory

### 4.3.1 Background

The free energy of replica  $i$  in the canonical ensemble at temperature  $T$ , whose potential  $V_i(\Gamma) = V(\Gamma, \lambda_i)$  is a function of system configuration  $\Gamma$  and the parameter  $\lambda_i$ , is formally given as (Chandler, 1987),

$$F_i = -\beta^{-1} \ln \left\{ \int d\Gamma \exp[-\beta V_i(\Gamma)] \right\}. \quad (4.2)$$

At equilibrium, the probability of observing configuration  $\Gamma$  is given as,

$$\rho_i(\Gamma) = \exp\{\beta[F_i - V_i(\Gamma)]\}. \quad (4.3)$$

To calculate the free energy change  $\delta F$  associated with switching the Hamiltonian from  $V_0$  to  $V_1$  we perform simulations at  $\lambda_0$  and  $\lambda_1$ , and calculate the forward and reverse work as,

$$W^F(\Gamma) = V_1(\Gamma) - V_0(\Gamma), \quad (4.4a)$$

$$W^R(\Gamma) = V_0(\Gamma) - V_1(\Gamma). \quad (4.4b)$$

For the forward and reverse work values the configuration  $\Gamma$  is typically drawn from the equilibrium ensemble of  $V_0$  and  $V_1$ , respectively. The Free Energy Perturbation (FEP) method (Zwanzig, 1954) utilizes forward and reverse work distributions to provide two independent estimators for  $\delta F$ ,

$$\delta F_{FEP}^F = -\beta^{-1} \ln \langle \exp(-\beta W^F) \rangle_0, \quad (4.5a)$$

$$\delta F_{FEP}^R = +\beta^{-1} \ln \langle \exp(-\beta W^R) \rangle_1, \quad (4.5b)$$

where the forward estimator  $\delta F_{FEP}^F$  utilizes forward work values from the simulation at  $V_0$ , and the reverse estimator the reverse work from  $V_1$ . Note that in both cases  $\delta F$  is associated with the process of switching  $\lambda_0 \rightarrow \lambda_1$ . These two estimators have different convergence rates (Lu and Kofke, 2001a). Therefore, while in practice the

two estimates should be equal, in simulations with finite sampling they are generally different.

Another free energy estimator, the Bennett acceptance ratio (Bennett, 1976), uses both the  $W^F$  and  $W^R$  distributions to obtain a free energy estimate. It is generally more accurate (Shirts and Pande, 2005b) and is employed later in this paper for numerical free energy estimates, but will not be considered for theoretical development.

### 4.3.2 The Hysteresis Error

The hysteresis error  $\epsilon_H$  is defined as the difference between the forward and reverse  $\delta F_{FEP}$  estimates,

$$\epsilon_H \equiv \delta F_{FEP}^F - \delta F_{FEP}^R. \quad (4.6)$$

$\epsilon_H$  has contributions from both the statistical and bias error of the FEP estimators (Lu and Kofke, 2001a; Zuckerman and Woolf, 2004). The bias error of the two estimators is typically in the opposite direction. While the statistical error may dominate the  $\epsilon_H$  for a given simulation, in averages over multiple short simulations the dominant contribution to the average hysteresis error is the sum of the forward and reverse FEP bias.

We take  $\epsilon_H$  as a measure of sampling quality and aim to minimize its magnitude between all pairs of neighboring replicas. The validity of using  $\epsilon_H$  as a general sampling error is based on a relationship between it and the fluctuation theorem of Crooks(4.1), derived below.

Switching the parameter  $\lambda_0 \rightarrow \lambda_1$  (and vice versa) is equivalent to performing non-equilibrium work; the difference between the work performed and the free energy change of the system is the dissipated work, defined in the forward and reverse direction as,

$$W_D^F(\Gamma) = W^F(\Gamma) - \delta F, \quad (4.7a)$$

$$W_D^R(\Gamma) = W^R(\Gamma) + \delta F. \quad (4.7b)$$

Crooks (1999) equates  $W_D^F$  and  $W_D^R$  to the entropy production caused by changing  $\lambda_0 \rightarrow \lambda_1$  and  $\lambda_1 \rightarrow \lambda_0$ , respectively, for the given configuration.

The distributions  $P_F(W_D)$  and  $P_R(W_D)$  give the probability of realizing a specific value for the dissipated work in the forward and reverse directions, respectively. The distributions are related to each other by the fluctuation theorem shown in Eq. (4.1), which we have re-derived in section 4.8.1 for the specific case of instantaneous switching between configurations with different  $\lambda$  values. In practice, Eq. (4.1) will not be satisfied exactly because of errors due to finite sampling. To take simulation errors into account, we rewrite Eq. (4.1) with an arbitrary error term  $\epsilon_{FT}^*$  and with observed (rather than ideal) dissipated work distributions  $P_F^*$  and  $P_R^*$ ,

$$\exp[\beta W_D + \beta \epsilon_{FT}^*(W_D)] = \frac{P_F^*(\beta W_D)}{P_R^*(-\beta W_D)}. \quad (4.8)$$

Eq. (4.8) is constructed such that the Crooks fluctuation theorem is recovered and  $\epsilon_{FT}^* = 0$  when the observed work distributions match the correct distributions. The hysteresis error  $\epsilon_H$  and the fluctuation error  $\epsilon_{FT}^*$  are related to each other as, (see section 4.8.2),

$$\epsilon_H = -\beta^{-1} \ln \langle \exp(-\beta \epsilon_{FT}^*) \rangle_0^*, \quad (4.9)$$

where  $\langle \cdot \rangle^*$  is defined as the average obtained from a finite simulation. The more closely a simulation obeys the relationship (4.1), the smaller the hysteresis error  $\epsilon_H$ , and vice versa. In the next section, we will discuss methods to reduce  $\epsilon_H$ , which in turn leads to the satisfaction of the Crooks fluctuation theorem.

### 4.3.3 Replica Exchange

In a Hamiltonian replica exchange (Sugita and Okamoto, 1999; Fukunishi et al., 2002) simulation, Monte Carlo moves are employed to exchange configurations  $\Gamma$  (or equivalently, parameters  $\lambda$ ) between two replicas with the probability,

$$P_{swap} = \min[1, \exp(-\beta \Delta V_{swap})], \quad (4.10)$$

where,

$$\begin{aligned} \Delta V_{swap} &= V_0(\Gamma_1) + V_1(\Gamma_0) \\ &\quad - V_0(\Gamma_0) - V_1(\Gamma_1), \end{aligned} \tag{4.11a}$$

$$= W^F + W^R, \tag{4.11b}$$

$$= W_D^F + W_D^R. \tag{4.11c}$$

$\Gamma_0$  and  $\Gamma_1$  denote configurations drawn at random from the equilibrium ensembles of  $V_0$  and  $V_1$ , respectively. For convenience, we write  $\gamma = (\Gamma_0, \Gamma_1)$  as a pair of such configurations, and  $\gamma' = (\Gamma_1, \Gamma_0)$  is the swapped configuration pair.

Since  $\Gamma_0$  and  $\Gamma_1$  are independent configurations, we can consider the probability of sampling  $\Gamma_0$  in the equilibrium ensemble of  $V_0$  and sampling  $\Gamma_1$  in the equilibrium ensemble of  $V_1$ ; this is the native probability  $\rho_N(\gamma)$ . Analogously, the joint probability of sampling the swapped configurations,  $\Gamma_1$  from  $\rho_0$  and  $\Gamma_0$  from  $\rho_1$  is given as  $\rho'_N(\gamma)$ :

$$\rho_N(\gamma) = \rho_0(\Gamma_0)\rho_1(\Gamma_1), \tag{4.12a}$$

$$\rho'_N(\gamma) = \rho_0(\Gamma_1)\rho_1(\Gamma_0) = \rho_N(\gamma'). \tag{4.12b}$$

Replica exchange swaps are conveniently visualized by plotting the independent configurations  $\Gamma_0$  and  $\Gamma_1$  along orthogonal axes and the equilibrium ensemble of the system as an isocontour of  $\rho_N$ , illustrated in Fig. 4.1(a).

At equilibrium, the relative probability of observing a pair of replicas in their swapped versus native configurations is,

$$\frac{\rho'_N}{\rho_N} = \exp(-\beta\Delta V_{swap}), \tag{4.13}$$

which is derived with definitions (4.12), (4.3) and (4.11a). We will refer to this as an inter-replica equilibrium relationship.

In an infinitely long simulation, (4.13) will be satisfied exactly, but this will generally not be the case for finite simulations, where inadequate sampling of configuration space will result in inaccurate probability estimates. However, in simulations with

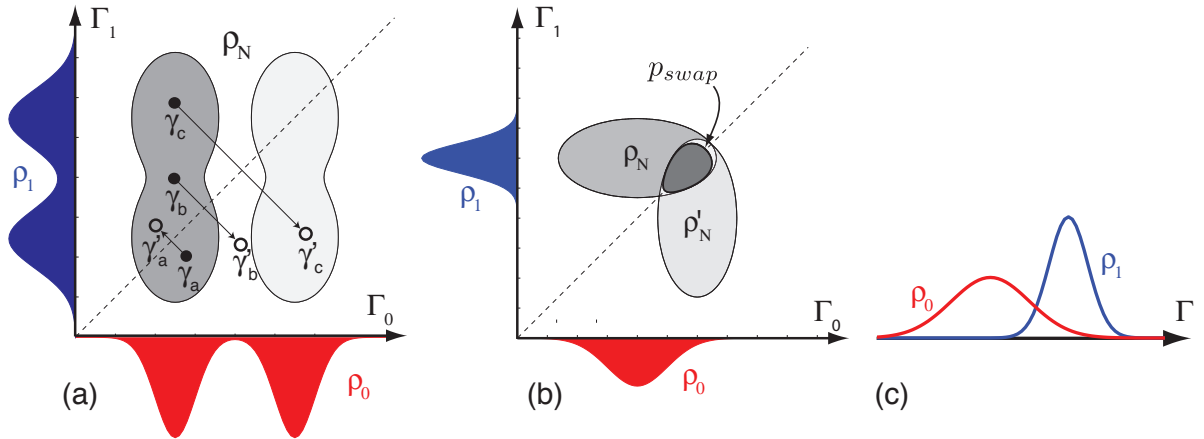


Figure 4.1: A graphical representation of replica exchange. **(a)** The independent (high dimensional) configuration spaces  $\Gamma_0$  and  $\Gamma_1$  have probability distributions  $\rho_0$  and  $\rho_1$ , respectively, and the joint equilibrium ensemble  $\rho_N$  is drawn over this domain. The  $\rho_0$  system has a kinetic barrier (represented by the two disconnected lobes) and with no replica exchange the system explores only the configurations of the shaded domain. A replica exchange swap is a reflection of the configuration pair  $\gamma$  about the  $\Gamma_0 = \Gamma_1$  diagonal axis, and three swap attempts are shown: the configuration pair  $\gamma_a$  swaps successfully and becomes  $\gamma'_a$ , but it does not sample otherwise inaccessible regions; a swap of  $\gamma_b$  fails because  $\gamma'_b$  is not in the equilibrium ensemble; and the swap of  $\gamma_c$  succeeds and allows the system to explore otherwise inaccessible regions of phase space. **(b)** The equilibrium domain  $\rho_N$  and its swapped image  $\rho'_N$  are drawn. Swaps are feasible only for configuration pairs which belong to both  $\rho_N$  and  $\rho'_N$ . This overlap region, labeled  $p_{swap}$ , is the domain where the integrand of Eq. (4.17b) is large, and its integral corresponds to the average swap probability  $\langle p_{swap} \rangle$ . **(c)** The overlap of the  $\rho_0$  and  $\rho_1$  distributions along the common configuration  $\Gamma_0 = \Gamma_1$ . For the hysteresis error to converge, the  $\lambda_0$  simulation must observe configurations where  $\rho_1 > \rho_0$ , and the  $\lambda_1$  simulation must adequately sample the region  $\rho_0 > \rho_1$ . The frequency with which this occurs is approximately given by  $\langle p_{swap} \rangle$ .

replica exchange we expect the inter-replica equilibrium relationship to be satisfied more closely than in simulations without replica exchange, because the swap move distributes configuration pairs in such a way as to satisfy Eq. (4.13). To illustrate, consider the system in Fig. 4.1(a) where the  $V_0$  replica is presumed to be stuck in the left lobe of the  $\rho_0$  distribution because of a kinetic barrier. Without replica exchange, only the heavily shaded region of  $\rho_N$  will be sampled accurately. The simulation will not have a correct estimate for  $\rho'_N(\gamma_c) = \rho_N(\gamma'_c)$ , since  $\rho_0$  for the swapped configuration, never having been observed, will be inaccurate. Consequently, Eq. (4.13) will not hold. Replica exchange directly populates swapped configurations (e.g.,  $\gamma'_c$ ), thereby improving the statistics of  $\rho'_N$  and allowing inter-replica equilibrium to be achieved more quickly for all configurations in  $\rho_N$ .

The degree to which Eq. (4.13) is satisfied determines the magnitude of the hysteresis error. To illustrate this, suppose that the distribution  $\rho'_N$  has some small error  $\rho_\epsilon(\Gamma_0, \Gamma_1)$  due to finite sampling, so that we write  $(\rho'_N + \rho_\epsilon)$  as the numerator in Eq. (4.13). In section 4.8.3 we show, by integrating over all configuration pairs, that the relationship between the hysteresis error and the error of sampling the swapped distribution,  $\rho_\epsilon$  is,

$$\epsilon_H \simeq -\beta^{-1} \int d\Gamma_0 d\Gamma_1 \rho_\epsilon. \quad (4.14)$$

The hysteresis error, then, will be minimized when the estimated swapped configuration probabilities  $\rho'_N$  are consistent with the equilibrium distribution. Since replica exchange populates the swapped configurations directly, it provides an efficient route to minimizing  $\epsilon_H$ .

### 4.3.4 Swap Probability

Analysis of the average replica exchange swap probability is complicated by the fact that the Metropolis function (Eq. (4.10)) is not analytical. For the purposes of interpreting this quantity, we will instead consider the Fermi swap probability,

$$p_{swap} = f(\beta\Delta V_{swap}),$$



where  $f(x)$  is defined as,

$$f(x) = 1/[1 + \exp(x)]. \quad (4.15)$$

(See Bennett (1976) for discussion). We use  $p_{swap}$  to denote the Fermi swap probability and  $P_{swap}$  for the Metropolis swap probability; while the theoretical development uses  $p_{swap}$ , replica exchange moves are accepted/rejected using  $P_{swap}$ . A simulation with either the Metropolis or Fermi swap probability will yield a Boltzmann distribution of swapped and unswapped configurations (Eq. (4.13)). While the exact numerical values of the Fermi and Metropolis swap probabilities will differ somewhat, their qualitative behavior and the conclusions drawn here will hold for both.

The average Fermi swap probability for two systems evolving independently is,

$$\langle p_{swap} \rangle \equiv \langle \langle f(\beta \Delta V_{swap}) \rangle \rangle_0, \quad (4.16a)$$

$$= \int d\Gamma_0 d\Gamma_1 \rho_N f(\beta \Delta V_{swap}), \quad (4.16b)$$

which can be written as,

$$\langle p_{swap} \rangle = \left\langle \left\langle \frac{\rho'_N}{\rho_N + \rho'_N} \right\rangle \right\rangle_1, \quad (4.17a)$$

$$= \int d\Gamma_0 d\Gamma_1 \frac{\rho_N \rho'_N}{\rho_N + \rho'_N}. \quad (4.17b)$$

The integrand of (4.17b) is a normalized probability of observing a given configuration pair, and the average swap probability is then the overlap of  $\rho_N$  and  $\rho'_N$ . See Fig. 4.1(b) for a graphical interpretation. Thus, a large average swap probability implies a large overlap between the equilibrium distributions of the two replicas, and a low  $\langle p_{swap} \rangle$  indicates that the configurations these replicas adopt are distinct.

We can expand (4.16a) in a Taylor series about  $\lambda = \lambda_0 + \delta\lambda$ . To leading order in  $\delta\lambda$ , we find that in the neighborhood of  $\lambda_0$  the average swap probability is, (see section 4.8.4),

$$\langle p_{swap} \rangle \simeq \frac{1}{2} - \frac{\beta^2 \delta\lambda^2}{4} C_\lambda, \quad (4.18)$$

where

$$C_\lambda \equiv \text{var} \left( \frac{\partial V}{\partial \lambda} \right) = \langle (\partial V / \partial \lambda)^2 \rangle_0 - \langle \partial V / \partial \lambda \rangle_0^2.$$

$C_\lambda$ , then, determines the rate at which the average swap probability declines as the difference in  $\lambda$  between the two replicas,  $\delta_\lambda$ , increases, although this linear analysis is accurate only for small  $\delta_\lambda$ .

### 4.3.5 Swap Probability and the Hysteresis Error Convergence Rate

We now demonstrate that the average swap probability between two replicas gives a measure of how quickly the hysteresis error decreases, on average, over the course of a simulation. The hysteresis error is the difference between the forward and reverse  $\delta F_{FEP}$ , and since the forward and reverse FEP estimators do not converge at equal rates (Lu and Kofke, 2001a), it is the slower of these which governs the convergence of  $\epsilon_H$ .

We may rewrite Eq. (4.5a) as,

$$\langle \exp(-\beta W_D^F) \rangle_0 = 1. \quad (4.19)$$

For this to hold, we must sample configurations where  $W_D^F < 0$ ; since the dissipated work is on average greater than zero by the second law of thermodynamics, such configurations tend to be rare (Jarzynski, 2006). As a result, the convergence rate of  $\delta F_{FEP}^F$  is governed by the probability of observing negative dissipated forward work values. Likewise, the convergence of  $\delta F_{FEP}^R$  is dictated by observations of  $W_D^R < 0$ . We can understand this criterion graphically with the relationships, (see section 4.8.1),

$$\frac{\rho_0(\Gamma_0)}{\rho_1(\Gamma_0)} = \exp[\beta W_D^F(\Gamma_0)], \quad (4.20a)$$

$$\frac{\rho_1(\Gamma_1)}{\rho_0(\Gamma_1)} = \exp[\beta W_D^R(\Gamma_1)]. \quad (4.20b)$$

In the context of Fig. 4.1(c), observing  $W_D^F < 0$  corresponds to sampling configurations from the  $\rho_0$  distribution where  $\rho_1 > \rho_0$ , and for  $W_D^R < 0$  we require  $\rho_0 > \rho_1$  when sampled from the  $\rho_1$  distribution.

Turning our attention to the average swap probability, we note that  $\Delta V_{swap}$ , which is the sum of  $W_D^F$  and  $W_D^R$ , is negative whenever  $\rho'_N > \rho_N$  (by Eq. (4.13)). Configurations for which this is the case are sampled by a simulation only in the lower-right half of the domain labeled  $p_{swap}$  in Fig. 4.1(b). The larger this domain, whose size is given by the average swap probability, the more frequently negative values of  $W_D^F$  and  $W_D^R$  are observed, and the more quickly the hysteresis error converges. A numerical confirmation of this argument, that low swap probabilities correspond to large hysteresis errors and vice versa, is demonstrated in the results section.

## 4.4 Methods

The computational system consists of 21 replicas, each with a different  $\lambda$ , which are simulated independently to obtain equilibrium statistics. The parameter  $\lambda$  controls the non-bonded interactions between an acetamide (ACE) solute and the water molecules. Two independent sets of simulations were performed, with and without replica exchange, in order to investigate the effect of this technique.

The Lennard-Jones and Coulomb interactions between the water and ACE molecules are scaled by  $\lambda_{LJ}$  and  $\lambda_C$ , respectively. We scaled both parameters simultaneously, such that  $\lambda_{LJ} = \lambda_C$ ; the single parameter  $\lambda$  then refers to both terms. This choice, while not commonplace in free energy calculations, was made to simplify the replica exchange implementation, and since the free energy is a state function, any path through  $(\lambda_{LJ}, \lambda_C)$  space is valid (Chialvo and Haile, 1987). The specific way in which the Lennard-Jones and Coulomb terms scale with  $\lambda$  is described in section 4.9.  $\lambda$  varies across the 21 replicas from 0 to 1 in increments of 0.05.

Each replica consists of 343 water molecules and one ACE molecule, which is rigid and whose position is fixed in the central box. All simulations were performed at constant temperature (298K) and volume (21.8Å cubic box) using Metropolis Monte Carlo sampling. Parameters from the OPLS-AA force field (Jorgensen et al., 1996) and 4-site TIP4P water model (Jorgensen et al., 1983) were used to model the solute and solvent, respectively. Minimum image boundary conditions and spherical cutoffs were employed for the Coulomb and Lennard-Jones potentials. The cutoff radius was 10.5Å for electrostatic interactions and 10Å for van der Waals interactions. Cutoffs were

group-based for the former, and atom-based for the latter. No long-range corrections were employed. All simulations were carried out using the MCCCS Towhee (Martin and Siepmann, 1999) Monte Carlo simulation package <sup>4</sup>.

The initial configurations for all replicas were identical and correspond to the end-point of a pre-equilibration run with ACE in water. For each replica, simulations consisted of 2 million cycles, where a cycle corresponds to 343 Monte Carlo moves; each move combines rotations and translations of a randomly chosen individual water molecule. The initial  $10^5$  cycles were discarded for equilibration. The average acceptance rate for all replicas was 31%.

The replica exchange simulation consists of a number of simulation rounds, where each replica evolves independently, separated by swap rounds, when a number of swap attempts take place. The length of the simulation round was drawn from a normal distribution with a mean of 500 and standard deviation of 50 cycles. 500 cycles is the approximate energy autocorrelation “time”. The swap round consists of  $21^2$  swap attempts between randomly selected replica pairs. Allowing swaps beyond neighboring replicas increases the efficiency of replica exchange, by allowing a replica to traverse the entire range of  $\lambda$  from 0 to 1 more quickly than if only neighbor swaps were permitted (Predescu et al., 2005).

During the course of the simulation, the native ( $V_i(\Gamma_i)$ ) and foreign ( $V_{j \neq i}(\Gamma_i)$ ) potential energies, as well as values for  $dV/d\lambda_C$  and  $dV/d\lambda_{LJ}$  (where  $dV/d\lambda = dV/d\lambda_{LJ} + dV/d\lambda_C$ ), were saved every 10 cycles. These were then post-processed to obtain the free energies, the hysteresis error, swap probabilities, and  $C_\lambda$ , regardless of whether actual replica exchange swaps took place. The total free energy of hydration,  $\Delta F$ , is the sum of all free energy changes  $(\delta F)_i$  between neighboring replicas  $i$  and  $i + 1$ , calculated using the Bennett acceptance ratio method (Bennett, 1976),

$$\Delta F \equiv \sum_i^{M-1} (\delta F)_i$$

---

<sup>4</sup>See <http://towhee.sourceforge.net>. Version maw4\_17.4 (maw-dev branch) was used for this work.

where  $M$  is the total number of replicas. Similarly,  $\epsilon_{RMS}$  is the root-mean-square of the hysteresis error  $(\epsilon_H)_i$  between neighboring replicas,

$$\epsilon_{RMS} \equiv \sqrt{\sum_i^{M-1} (\epsilon_H)_i^2 / M} \quad (4.21)$$

Statistical errors for  $\Delta F$  were estimated using the bootstrap method (Efron and Tibshirani, 1993). With the simulation dataset consisting of  $N$  observations, we drew  $n^*$  observations at random and with replacement to create one bootstrap estimate,  $\Delta F^*$ . This process was repeated 10,000 times, and the standard deviation among all the  $\Delta F^*$  is the estimated error of  $\Delta F$ .  $n^*$  is the expected number of independent observations in the dataset; here,  $n^* = 1900$  with the assumption that there is one independent observation per two internal energy autocorrelation “times” (Newman and Barkema, 1999).

## 4.5 Results

### 4.5.1 Acetamide Free Energy of Hydration

The hydration free energies we calculate for acetamide are in line with results obtained by other researchers, as shown in Table 4.1. All numerical results differ somewhat from experimental values due to differences in force field parameters. Our calculations were carried out in the canonical ensemble. Therefore, we obtain estimates for the Helmholtz free energy  $\Delta F$ , whereas the experimental and other computational values obtain estimates for the Gibbs free energy,  $\Delta G$ . Since the box volume at  $\lambda = 0$  was adjusted to correspond to 1 atmosphere, the distinction between these two values should be negligible (Qian and Hopfield, 1996) even with some “pressurizing” due to insertion of the acetamide (Shirts et al., 2003); test calculations of  $\Delta G$  in the NPT ensemble confirm this assertion (data not shown). The general consistency between our results and those of others serves to verify our implementation and sampling technique.

The methods to calculate  $\Delta G$  in the NPT ensemble do not differ from those for calculating  $\Delta F$  in the NVT ensemble; in particular, the replica exchange swap probability (4.10) does not change, since the pressure-volume work is reversible and does not contribute to the dissipated work.

Table 4.1 shows differences between results obtained with and without replica exchange. As expected from our theoretical considerations, we find that the root-mean-square hysteresis error is lowered by an order of magnitude when replica exchange is coupled to the multicanonical sampling protocol. However, it should be noted that the statistical error estimated using bootstrap remains unaffected. This is not an artifact of the bootstrap method used to estimate statistical errors. Instead, fluctuations in estimates for  $\delta F$  originate in fluctuations of the underlying work distribution, shown in Eq. (4.1). So long as both simulations sample the work distribution adequately, they will have similar statistical error associated with them. As a cautionary note, low statistical errors can also be caused by inadequate sampling of the appropriate work distributions. The statistical error between two replicas can be reduced by decreasing the  $\lambda$ -distance between them, and an optimal  $\lambda$  schedule can reduce it for an entire simulation.

### 4.5.2 Hysteresis Error and Replica Exchange

For a fixed  $\lambda$  schedule, the hysteresis error may be reduced with either an improved sampling methodology like replica exchange, or longer simulations per replica. The effects of both approaches are illustrated in Fig. 4.2.

Panel (a) shows  $\epsilon_H$  for each neighboring replica pair. The hysteresis error is not uniform across all pairs, with spikes in the region  $\lambda = 0.1 - 0.3$ . Replica exchange systematically reduces the hysteresis error for all pairs of replicas.

Panel (b) illustrates how both longer sampling and replica exchange affect the hysteresis error. Block averaging shows that the average root-mean-square hysteresis error declines consistently with longer simulations. This reduction can be improved with replica exchange; in fact, a simulation with replica exchange will achieve the same magnitude of  $\epsilon_{RMS}$  about 5 times more quickly than one without replica exchange.

**(a) Acetamide Free Energy of Hydration: Current Work**

	$\Delta F$ (kcal/mol)	$\epsilon_{RMS}$ (kcal/mol)
No Replica Exchange	$-8.35 \pm 0.051$	0.120
Replica Exchange	$-8.14 \pm 0.053$	0.023

**(b) Acetamide Free Energy of Hydration: Literature**

	$\Delta G$ (kcal/mol)	Details
MacCallum and Tieleman (2003)	$-8.25 \pm 0.26$	TIP4P, TI
Shirts et al. (2003)	$-8.20^5 \pm 0.03$	TIP3P, TI
Chang et al. (2007)	$-8.54 \pm 0.1 - 0.3$	TIP4P, BAR
Udier-Blagović et al. (2004)	$-9.65 \pm 0.3 - 0.5$	TIP4P, FEP
Experimental (Wolfenden, 1978)	-9.54	

Table 4.1: The hydration free energy of acetamide. **(a)** The Helmholtz hydration free energy  $\Delta F$  for the current work, as calculated by the Bennett acceptance ratio, and the root-mean-square hysteresis error. The  $\Delta F$  statistical errors are calculated by the bootstrap method. **(b)** Published values of the Gibbs free energy  $\Delta G$ , obtained both computationally and experimentally. All computational results utilize the OPLS-AA force field for the solute acetamide. Also noted are the water model and free energy estimator (TI: thermodynamic integration; FEP: free energy perturbation; BAR: Bennett acceptance ratio)

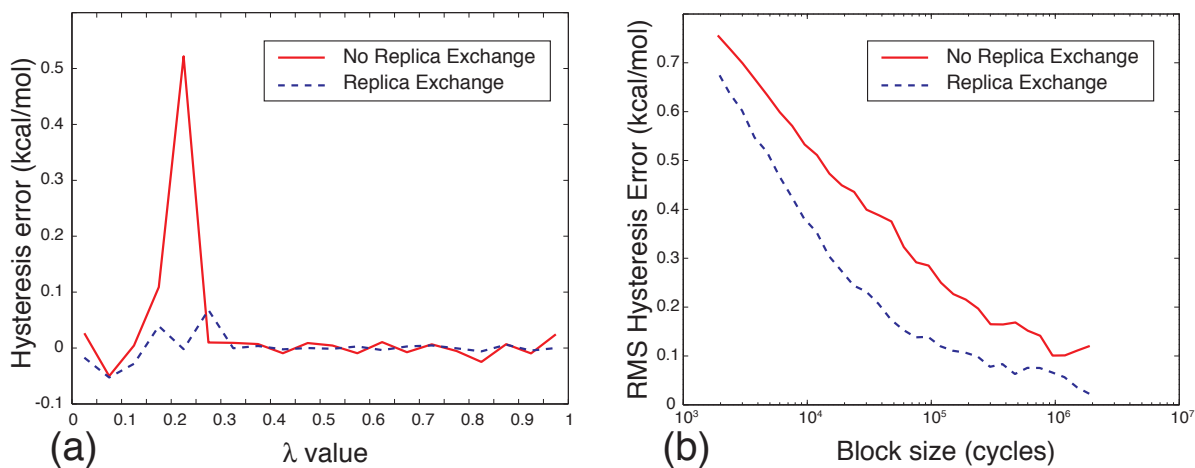


Figure 4.2: **(a)** The hysteresis error between neighboring replicas. Replica exchange effectively reduces the hysteresis error for replica pairs. **(b)** Block averages of the root-mean-square hysteresis error, showing that the hysteresis error falls with increasing block size. Replica exchange increases the rate at which hysteresis error is lowered, thereby achieving the same magnitude error with simulations which are on average 4-8 times shorter.

### 4.5.3 Average Swap Probability

Fig. 4.3 shows downward spikes in the swap probability for values of  $\lambda$  where the hysteresis error is large in Fig. 4.2(a). These results are consistent with the proposal that swap probability between two replicas is an indicator of the rate at which  $\epsilon_H$  is minimized. The same region is characterized by a positive spike in  $C_\lambda$ , which is expected based on the relationship between the swap probability and  $C_\lambda$  in Eq. (4.18). However, while the swap probability calculation requires two separate simulations, estimates of  $C_\lambda$  can be obtained from just one. Moreover,  $\langle p_{swap} \rangle$  varies as the distance between the replicas changes, complicating the interpretation if the  $\lambda$  schedule is not uniform. Evaluation of  $C_\lambda$  as a function of  $\lambda$  using a preliminary, coarse  $\lambda$  schedule can identify regions where the swap probability is expected to be low, and can be used to construct optimal  $\lambda$  schedules, as discussed in Sec. 4.6.2.

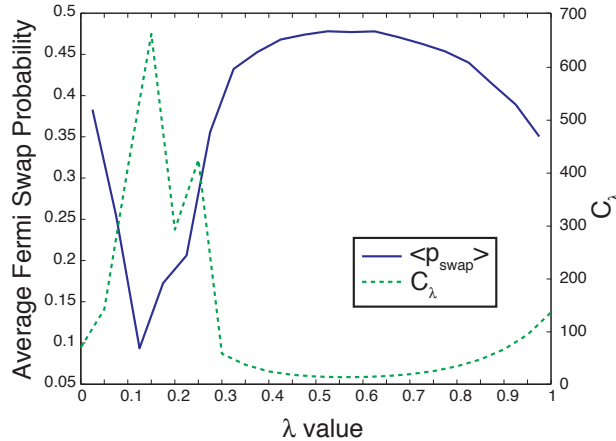


Figure 4.3: The average swap probability between adjacent replicas and  $C_\lambda = \text{var}(\partial V / \partial \lambda)$  evaluated for each replica (from the replica exchange simulation; simulation with no replica exchange is not significantly different). Spikes in  $C_\lambda$  indicate regions of low swap probability.



## 4.6 Discussion

### 4.6.1 Physical Interpretation of $C_\lambda$ Profile

To gain a physical interpretation of the profile for  $C_\lambda$  shown in Fig. 4.3, we plot in Fig. 4.4 the average water density in a  $2.5\text{\AA}$  sphere surrounding the carbonyl carbon of acetamide. The plot shows that water occupancy around the growing solute decreases rapidly in the range of  $\lambda \sim 0.15$ . The expulsion and rearrangement of water molecules during cavitation leads to a large shift in the equilibrium ensemble, giving rise to a pronounced spike in  $C_\lambda$ . (Smaller shifts in  $C_\lambda$  near  $\lambda = 1$  reflect electrostatic effects and are not observed for simulations where  $\lambda_C = 0$ , data not shown.) Thus,  $C_\lambda$  profiles serve as useful probes for detecting large shifts in equilibrium ensembles. Regions where the equilibrium ensembles change most rapidly are the regions that contribute the largest errors in free energy calculations.

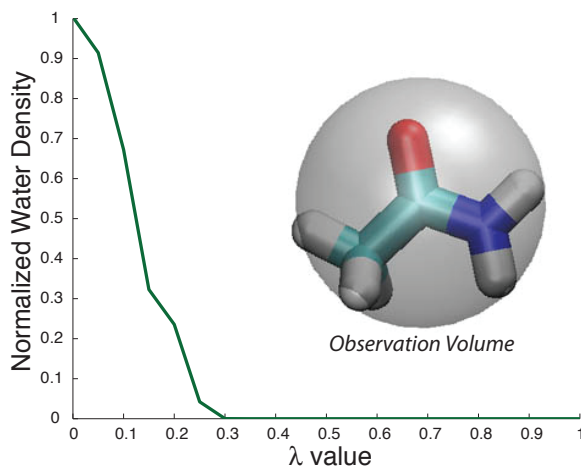


Figure 4.4: Water density within  $2.5\text{\AA}$  of the acetamide carbonyl carbon as  $\lambda$  varies. The inset illustrates the position and size of the observation volume with respect to an acetamide molecule. Density is normalized by the bulk density. As  $\lambda$  increases, waters are expelled by the growing cavity.

## 4.6.2 Optimal $\lambda$ Schedule for Free Energy Calculations

For given computational resources, with the number of replicas and the simulation length fixed, the root-mean-square hysteresis error of a simulation may be decreased by optimizing the  $\lambda$  schedule, or the distribution of  $\lambda$  across the replicas. The swap probability gives the rate at which the average hysteresis error falls between two replicas, and in an optimized simulation it would be uniform across all replica pairs. In practice it is difficult to obtain the  $\lambda$  schedule which makes the swap probability exactly uniform, but reasonable approximations can be made by using the linearized swap probability, given by Eq. (4.18).

First, it is necessary to perform some number of preliminary simulations to obtain  $C_\lambda$  along a coarse  $\lambda$  schedule. These initial simulations need not be as long as the final production runs, since  $C_\lambda$  converges more quickly than  $\delta F$  and is more tolerant of error. With a rough estimate of  $C_\lambda(\lambda)$  in hand, the  $\lambda$  schedule can be adjusted to ensure that the linear swap probability is uniform between all replicas. Alternatively, one might simply shift replicas from where  $C_\lambda$  is small to where it is large. Both approaches are only approximate, and break down when the linear response assumption in Eq. (4.18) ceases to be valid. They may be applied iteratively as  $C_\lambda$  is evaluated for new  $\lambda$  schedules.

The aim of an optimal  $\lambda$  schedule is to place replicas close together in regions where the  $C_\lambda$  profile shows spikes. This ensures reasonable swap probabilities and minimal hysteresis errors in regions that are problematic. Preliminary investigations show that even when the schedule is improved in an *ad hoc* manner, hysteresis as well as statistical errors decrease.

## 4.6.3 Replica Exchange

Replica exchange provides a Monte Carlo move which may allow a replica to access a distant part of its equilibrium ensemble in one step. It is no substitute for conformational exploration within a replica. This point, while obvious, must be emphasized in the context of the hysteresis error, which does not report on the quality of intra-replica sampling. As an extreme but illustrative case, consider a system of some number of

frozen replicas, each with a different configuration, which undergo replica exchange moves but no conformational changes. With just a modest number of swaps, these configurations attain the probability distribution described by Eq. (4.13), and the hysteresis error is zero. The system has achieved inter-replica equilibrium, but the intra-replica probability distribution has not been obtained. In practice, the majority of Monte Carlo moves must be within a replica. The optimal frequency of swap moves remains an open question, although preliminary simulations suggest that more frequent swaps reduce the hysteresis error more quickly.

It is worthwhile to relate replica exchange-based free energy calculations presented here to other generalized ensemble techniques, particularly  $\lambda$ -dynamical methods (Kong and Brooks III, 1996; Bitetti-Putzer et al., 2003; Abrams et al., 2006). There,  $\lambda$  is a dynamical variable which evolves in time according to the conjugate force  $\partial V/\partial\lambda$ , with additional terms typically added to the Hamiltonian to associate with  $\lambda$  a momentum and to restrict its range. Since  $\lambda$  fluctuates, equilibrium distributions associated with a fixed  $\lambda$  cannot be calculated, and measures like the swap probability and the hysteresis error are no longer defined. Nonetheless, both classes of techniques attempt to distribute atomic configurations across a range of  $\lambda$  values according to a Boltzmann distribution;  $\lambda$ -dynamical generalized ensemble techniques generate this distribution dynamically, whereas replica exchange techniques utilize a Markov chain to the same effect. Sampling difficulties in both cases are associated with phase changes and a large variance of  $\partial V/\partial\lambda$ , or  $C_\lambda$ . This results in low swap probabilities (for replica exchange) or regions in  $\lambda$  space not easily traversed (for  $\lambda$  dynamics), and these problems may be overcome with improved  $\lambda$  schedules or modified biasing potentials, respectively. One practical advantage of replica exchange techniques is that they are readily parallelizable across a number of computers, a trait not shared by all  $\lambda$ -dynamical methods.

## 4.7 Summary and Conclusion

In a simulation of multiple replicas, each sampling the equilibrium ensemble of a different Hamiltonian, swapping configurations between replicas is a nonequilibrium

work process. Accordingly, the work needed to perform such swaps has a distribution of values, as described by the Crooks fluctuation theorem. The hysteresis error  $\epsilon_H$  developed here measures how closely a given simulation reproduces these work distributions between a pair of replicas.

The hysteresis error is particularly useful in the context of free energy calculations. It reports on the combined bias of the forward and reverse free energy perturbation techniques, and it measures how completely individual replicas sample their equilibrium ensemble. The root-mean-square hysteresis error, which reports on  $\epsilon_H$  for the whole  $\lambda$  schedule, may be decreased by running a longer simulation, employing replica exchange, utilizing an improved  $\lambda$  schedule, or all of these approaches.

The average swap probability is another useful measure and can be calculated whether or not replica exchange is employed. Since it determines the rate at which the hysteresis error decreases with simulation length, the swap probability can be used to optimize the  $\lambda$  schedule. With a uniform average swap probability the hysteresis error falls evenly between all replica pairs. This maximizes the efficiency of simulations with fixed computational resources, avoiding unnecessary replicas where the hysteresis is low and preventing excessive errors from regions where the hysteresis error is large.

The swap probability, along with a related measure  $C_\lambda$ , yields insight into the microscopic behavior of a system. The swap probability is low and  $C_\lambda$  is large when the equilibrium ensemble changes rapidly with  $\lambda$  – for instance, during phase changes. Slow convergence and bias errors in free energy calculations arise when there are spikes in the  $C_\lambda$  profile along the  $\lambda$  schedule, resulting in large hysteresis errors.

## 4.8 Derivations

### 4.8.1 Fluctuation Theorem Derivation

We derive the Crooks fluctuation theorem (4.1) in the context of instantaneously switching  $\lambda_0 \rightarrow \lambda_1$  (forward) and  $\lambda_1 \rightarrow \lambda_0$  (reverse). Expanding the ratio  $\rho_0/\rho_1$  with

(4.3) for an arbitrary configuration  $\Gamma$ ,

$$\begin{aligned}\frac{\rho_0(\Gamma)}{\rho_1(\Gamma)} &= \exp[\beta(F_0 - F_1) - \beta(V_0 - V_1)], \\ &= \exp(-\beta\delta F + \beta W^F), \\ &= \exp(\beta W_D^F),\end{aligned}\tag{4.22a}$$

and similarly,

$$\frac{\rho_1(\Gamma)}{\rho_0(\Gamma)} = \exp[\beta W_D^R(\Gamma)]\tag{4.22b}$$

where the definitions of work (4.4) and dissipated work (4.7) were used.

We integrate  $\rho_1$  from (4.22a) over all configurations, but consider contributions only from those  $\Gamma$  for which the forward dissipated work value takes on a specific value,  $W_D$ :

$$\begin{aligned}\int d\Gamma \rho_0(\Gamma) \exp[-\beta W_D^F(\Gamma)] \delta[\beta W_D - \beta W_D^F(\Gamma)] \\ = \int d\Gamma \rho_1(\Gamma) \delta[\beta W_D - \beta W_D^F(\Gamma)].\end{aligned}\tag{4.23}$$

Since, from (4.22a) and (4.22b),

$$W_D^F(\Gamma) = -W_D^R(\Gamma)$$

(4.23) becomes,

$$\begin{aligned}\int d\Gamma \rho_0(\Gamma) \exp[-\beta W_D^F(\Gamma)] \delta[\beta W_D - \beta W_D^F(\Gamma)] \\ = \int d\Gamma \rho_1(\Gamma) \delta[\beta W_D + \beta W_D^R(\Gamma)].\end{aligned}\tag{4.24}$$

We define  $P^F(W_D)$  as the probability of observing a given dissipated work value in the forward switching process, and it can be expressed as an integral over all configurations which yield this value,

$$P^F(W_D) = \int d\Gamma \rho_0(\Gamma) \delta[\beta W_D - \beta W_D^F(\Gamma)]\tag{4.25a}$$

Likewise, the probability of observing a given dissipated work value in the reverse switching process is,

$$P^R(W_D) = \int d\Gamma \rho_1(\Gamma) \delta[\beta W_D - \beta W_D^R(\Gamma)] \quad (4.25b)$$

With these definitions, (4.24) may be written as,

$$\exp(-\beta W_D) P^F(\beta W_D) = P^R(-\beta W_D),$$

which is equivalent to (4.1).

## 4.8.2 Fluctuation Theorem and Hysteresis Error

The relationship between some arbitrary deviation of a simulation from the Crooks fluctuation theorem and the hysteresis error is derived by first rewriting Eq. (4.8) as,

$$P_R^*(-\beta W_D) \exp(\beta W_D) = P_F^*(\beta W_D) \exp(-\beta \epsilon_{FT}^*). \quad (4.26)$$

Inserting the  $\delta F_{FEP}^R$  definition (4.5b) into the definition of the hysteresis error (4.6), expanding the reverse work with (4.7b) and using the  $\delta F_{FEP}^F$  estimate for  $\delta F$ ,

$$\begin{aligned} \epsilon_H &= \delta F_{FEP}^F - \beta^{-1} \ln \langle \exp(-\beta W^R) \rangle_1^*, \\ &= \delta F_{FEP}^F - \beta^{-1} \ln \left[ \langle \exp(-\beta W_D^R) \rangle_1^* \exp(\beta \delta F_{FEP}^F) \right], \\ &= -\beta^{-1} \ln \left[ \langle \exp(-\beta W_D^R) \rangle_1^* \right]. \end{aligned}$$

We now expand the estimated ensemble average as an integral over all values of  $\beta W_D^R$ , with  $P_R^*$  the normalized histogram of  $\beta W_D^R$  obtained from a simulation,

$$\epsilon_H = -\beta^{-1} \ln \left[ \int_{-\infty}^{+\infty} d[\beta W_D^R] P_R^*(\beta W_D^R) \exp(-\beta W_D^R) \right].$$

As  $\beta W_D^R$  is a dummy variable, we change it to  $-\beta W_D$ ,

$$\epsilon_H = -\beta^{-1} \ln \left[ \int_{-\infty}^{+\infty} d[\beta W_D] P_R^*(-\beta W_D) \exp(\beta W_D) \right],$$

where we implicitly multiplied the integrand by  $-1$  to preserve the limits of integration. With (4.26) the above can be written as,

$$\epsilon_H = -\beta^{-1} \ln \left[ \int_{-\infty}^{+\infty} d(\beta W_D) P_F^*(\beta W_D) \exp(-\beta \epsilon_{FT}^*) \right],$$

which reduces to (4.9).

### 4.8.3 Inter-Replica Equilibrium and Hysteresis Error

We can relate an small arbitrary error in the calculated distribution  $\rho'_N$  to the hysteresis error by considering a small error  $\rho_\epsilon(\Gamma_0, \Gamma_1)$  in the otherwise correctly estimated  $\rho'_N$ . Rewriting (4.13),

$$\rho'_N + \rho_\epsilon = \rho_N \exp(-\beta \Delta V_{swap}),$$

we integrate over all configuration pairs and rewrite  $\Delta V_{swap}$  with (4.11b),

$$\begin{aligned} \int d\Gamma_0 d\Gamma_1 \rho'_N &+ \int d\Gamma_0 d\Gamma_1 \rho_\epsilon \\ &= \int d\Gamma_0 d\Gamma_1 \rho_0(\Gamma_0) \rho_1(\Gamma_1) \exp[-\beta W^F(\Gamma_0)] \\ &\times \int d\Gamma_1 \rho_1(\Gamma_1) \exp[-\beta W^R(\Gamma_1)]. \end{aligned} \quad (4.27)$$

With the sampling error contained in  $\rho_\epsilon$ , the  $\rho'_N$  term (expanded with (4.12b)) is identically one. Taking the logarithm and dividing by  $\beta$ , (4.27) becomes,

$$-\beta^{-1} \ln \left[ 1 + \int d\Gamma_0 d\Gamma_1 \rho_\epsilon \right] = \delta F_{FEP}^R - \delta F_{FEP}^F, \quad (4.28)$$

where we have used the  $\delta F_{FEP}$  definitions (4.5). With the approximation  $\ln(1+x) \simeq x$  for small  $x$  and the definition of  $\epsilon_H$  (4.6), we obtain Eq. (4.14).

### 4.8.4 Linearized Average Swap Probability

Here we consider the average Fermi swap probability between two replicas whose  $\lambda$  parameters differ by a small amount,  $\delta$  (written as  $\delta_\lambda$  in the text). For convenience

we define

$$\begin{aligned}\mu &\equiv \beta\Delta V_{swap}, \\ &= \beta[V_\delta(\Gamma_0) - V_0(\Gamma_0) + V_0(\Gamma_\delta) - V_\delta(\Gamma_\delta)],\end{aligned}$$

where  $\Gamma_0$  and  $\Gamma_\delta$  are configurations drawn from the equilibrium distributions  $V_0$  and  $V_\delta$  parameterized by  $\lambda_0$  and  $\lambda_0 + \delta$ , respectively. We expand  $V_\delta$  as a Taylor series about  $\lambda_0$ ,

$$V_\delta(\Gamma) = V_0(\Gamma) + \delta A_0(\Gamma) + \frac{\delta^2}{2} B_0(\Gamma) + O(\delta^3),$$

with

$$\begin{aligned}A_0 &\equiv \left. \frac{\partial V}{\partial \lambda} \right|_{\lambda=\lambda_0}, \\ B_0 &\equiv \left. \frac{\partial^2 V}{\partial \lambda^2} \right|_{\lambda=\lambda_0}.\end{aligned}$$

$\mu$  can then be written as,

$$\mu = \beta\delta[A_0(\Gamma_0) - A_0(\Gamma_\delta)] + \frac{\beta\delta^2}{2}[B_0(\Gamma_0) - B_0(\Gamma_\delta)].$$

Note that  $\mu$  is small ( $O(\delta)$ ); thus, with the identities,

$$\exp(x) = 1 + x + x^2/2 + \dots, \quad (4.29a)$$

$$\frac{1}{1+x} = 1 - x + x^2 - \dots, \quad (4.29b)$$

we may write the Fermi swap probability between configurations  $\Gamma_0$  and  $\Gamma_\delta$  as,

$$\begin{aligned}p_{swap} &= \frac{1}{1 + \exp \mu}, \\ &= \frac{1}{2} \left( \frac{1}{1 + \mu/2 + \mu^2/4 + O(\mu^3)} \right), \\ &= \frac{1}{2} [1 - (\mu/2 + \mu^2/4) + (\mu/2 + \mu^2/4)^2 + O(\mu^3)], \\ &= \frac{1}{2} - \frac{1}{4}\mu + O(\mu^3).\end{aligned}$$



The average swap probability is the ensemble average over all configuration pairs,

$$\begin{aligned}
\langle\langle p_{\text{swap}} \rangle_0 \rangle_\delta &= \frac{1}{2} - \frac{1}{4} \langle\langle \mu \rangle_0 \rangle_\delta, \\
&= \frac{1}{2} - \frac{1}{4} \left( \beta\delta \langle A_0 \rangle_0 + \frac{\beta\delta^2}{2} \langle B_0 \rangle_0 \right. \\
&\quad \left. - \beta\delta \langle A_0 \rangle_\delta - \frac{\beta\delta^2}{2} \langle B_0 \rangle_\delta \right) \\
&\quad + O(\delta^3).
\end{aligned} \tag{4.30}$$

To evaluate  $\langle \cdot \rangle_\delta$ , we first obtain  $Q_\delta$ , the partition function at  $(\lambda_0 + \delta)$ ,

$$\begin{aligned}
Q_\delta &\equiv \int d\Gamma \exp(-\beta V_\delta), \\
&= \int d\Gamma \exp(-\beta V_0) [1 - \beta\delta A_0 + O(\delta^2)], \\
&= Q_0 [1 - \beta\delta \langle A_0 \rangle_0 + O(\delta^2)],
\end{aligned}$$

and its reciprocal,

$$Q_\delta^{-1} = Q_0^{-1} [1 + \beta\delta \langle A_0 \rangle_0 + O(\delta^2)].$$

We can now evaluate  $\langle A_0 \rangle_\delta$  and  $\langle B_0 \rangle_\delta$ , retaining only terms which will remain  $O(\delta^2)$  or larger in (4.30):

$$\begin{aligned}
\langle A_0 \rangle_\delta &\equiv Q_\delta^{-1} \int d\Gamma \exp(-\beta V_\delta) A_0, \\
&= Q_0^{-1} (1 + \beta\delta \langle A_0 \rangle_0) \int d\Gamma (1 - \beta\delta A_0) \exp(-\beta V_0) A_0, \\
&= (1 + \beta\delta \langle A_0 \rangle_0) (\langle A_0 \rangle_0 - \beta\delta \langle A_0^2 \rangle_0), \\
&= \langle A_0 \rangle_0 + \beta\delta (\langle A_0 \rangle_0^2 - \langle A_0^2 \rangle_0),
\end{aligned}$$

and

$$\begin{aligned}
\langle B_0 \rangle_\delta &\equiv Q_\delta^{-1} \int d\Gamma \exp(-\beta V_\delta) B_0, \\
&= Q_0^{-1} (1 + O(\delta)) \int d\Gamma \exp(-\beta V_0) B_0 [1 - O(\delta)], \\
&= \langle B_0 \rangle_0 + O(\delta).
\end{aligned}$$

Finally, (4.30) becomes,

$$\langle\langle p_{swap} \rangle\rangle_0 \delta = \frac{1}{2} - \frac{\beta^2 \delta^2}{4} (\langle A_0^2 \rangle_0 - \langle A_0 \rangle_0^2) + O(\delta^3), \quad (4.31)$$

equivalent to Eq. (4.18), which is valid for small  $\delta$ .

## 4.9 $V_{LJ}$ and $V_C$ Functional Forms

The functional forms of both the Coulomb and Lennard-Jones potentials were developed for this work based on three criteria:

1. Configurations where the solute and solvent overlap may be observed for  $\lambda = 0$ . For such configurations, we require:
  - That swaps be permitted with reasonable frequency for small  $\lambda$  (e.g.  $\lambda = 0.1$ ).
  - That swap probabilities falls off quickly thereafter; in particular, we wish to avoid the situation where the swap probability declines only very near  $\lambda = 1.0$ .
2. We require that  $\partial V / \partial \lambda$  is not always zero for  $\lambda = 0$  to avoid complications with the thermodynamic integration (TI) estimator. While, we do not report results using TI in this work, we wish to construct a  $\lambda$  schedule that works with all estimators.
3. In this work,  $\lambda_{LJ} = \lambda_C$ . Therefore, Lennard-Jones repulsion must dominate Coulombic attraction at very small atomic separations.

While various ways to scale the potential have been discussed in the literature (Beutler et al., 1994; Pitara and van Gunsteren, 2002; Shirts and Pande, 2005a), none of these satisfied all of our requirements. It should be noted that condition 3 is somewhat arbitrary, and more common scaled potentials may be used if the insertion process scales the Lennard-Jones prior to the Coulomb potential. The specific profiles of Figs. 4.2(a), 4.3 and 4.4 are dependent on the choice of the Coulomb and Lennard-Jones functional forms, as well as the relationship between  $\lambda_C$  and  $\lambda_{LJ}$ .

**Coulomb scaling** We employ a modified version of the linear soft-core scaling (Pitera and van Gunsteren, 2002); for two atoms of charges  $q_i$  and  $q_j$  distance  $r$  apart, the potential energy is  $\lambda_C$  as,

$$V_C(r, \lambda_C) = \lambda_C \frac{q_i q_j}{\alpha_C(1 - \lambda_C) + r}, \quad (4.32)$$

$\alpha_C$  controls the “soft core” term, and for small  $\lambda_C$  imposes a minimum effective atomic separation.  $\alpha_C = 1.5\text{\AA}$  for all simulations in this work.

**Lennard-Jones scaling** The Lennard-Jones potential between two particles may be written generally as,

$$V_{LJ}(r, \lambda_{LJ}) = BA(A - 1), \quad (4.33)$$

where, for unscaled Lennard-Jones,

$$A(r) = \left(\frac{\sigma}{r}\right)^6, \quad B = 4\epsilon.$$

Simple linear scaling by  $\lambda_{LJ}$  of the Lennard-Jones potential is known to be unsatisfactory, and a number of alternate forms have been introduced. We developed the exponential soft-core,

$$A(r, \lambda_{LJ}) = 1 / \left[ \alpha_{LJ}(1 - \lambda_{LJ})^b + \left(\frac{r}{\sigma}\right)^6 \right], \quad (4.34a)$$

$$B(\lambda_{LJ}) = 4\epsilon \frac{1 - e^{-k\lambda_{LJ}}}{1 - e^{-k}}, \quad (4.34b)$$

with  $a = 4$ ,  $k = 1$  and  $\alpha_{LJ} = 0.5\text{\AA}$ . The precise position along the  $\lambda$  coordinate of the swap probability trough (see Fig. 4.3) is specific to this Lennard-Jones potential.

## 4.10 Addendum

Here we present two additional results which were not included in the original paper of relevance in the context of the dissertation. The first is a critical evaluation of the error estimates provided by the bootstrap estimator, and the second an analysis

of the hysteresis error – both free energy and energy – and how it decreases with increasing replica exchange swap rate.

### 4.10.1 Bootstrap Method Analysis

The statistical error as calculated with bootstrap or block averaging techniques attempts to estimate the range of results one would expect were the simulation to be repeated multiple times (see section 2.4.2). The advantage of such methods is that they obtain such information from an analysis of one simulation only. Yet they rely on a number of assumptions, such as that the original simulation is well sampled and representative of a population of simulations, which may not hold but are difficult to verify. In the course of the work described in this chapter, we critically evaluated the bootstrap technique against a series of independent simulations.

Figure 4.5 illustrates the  $\Delta F$  and  $\Delta U$  values of 43 independent simulations, 20 with and 23 without replica exchange. Here, we used a simplified version of the system presented in the body of this chapter, evaluating the solvation free energy for an acetamide molecule with no coulomb interactions ( $\lambda_C = 0$ ). The  $\lambda_{LJ}$  schedule was constructed to try to equalize the swap probability (based on preliminary simulations), with  $\lambda_{LJ}$  values of 0.0, 0.08, 0.137, 0.192, 0.27, 0.38, 0.54, 0.755 and 1.0; the  $\lambda$  schedule is finer in the region where rapid changes in the equilibrium ensemble occur and where swap probabilities are low. All the simulations started with the same configuration and an equilibration period of 10,000 cycles, followed by a production run of 450,000 cycles. Round lengths for the replica exchange system were given by a Gaussian distribution with a mean of 500 and standard deviation of 50 cycles; for comparison, the autocorrelation time of the internal energy was about 450 cycles.

The values of  $\Delta F$  and  $\Delta U$  in Fig. 4.5 were both calculated with the Bennett acceptance ratio. The standard deviation of  $\Delta F$  is 0.18 and 0.30 kcal/mol without and with replica exchange, respectively. The corresponding  $\Delta U$  standard deviations are 4.91 and 4.04 kcal/mol. The source of this difference is unclear, but may suggest that improved sampling due to replica exchange may increase, for short simulations, the statistical error of free energy calculations, perhaps because such simulations can “jump out” of local energy minima and explore conformational space more effectively.

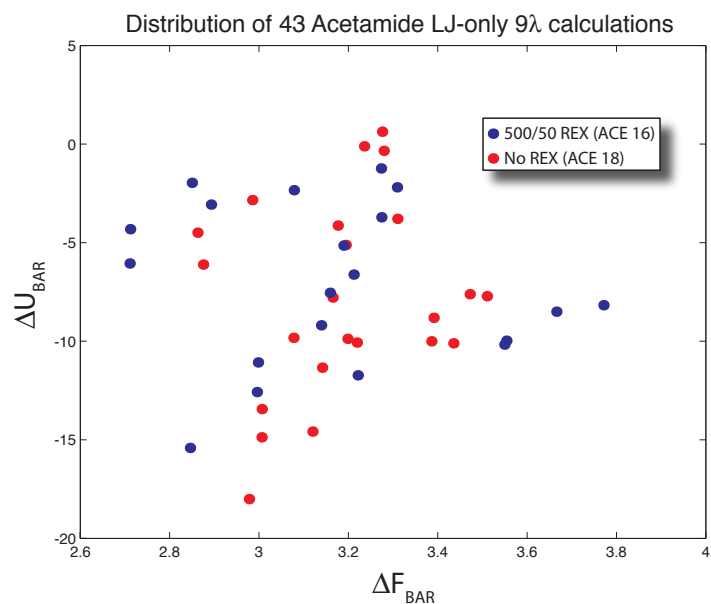


Figure 4.5: Summary of results of 43 independent simulations, 20 with replica exchange (“REX”, blue) and 23 without (“NoREX”, red), to independently characterize the statistical error of such simulations. The standard deviation of  $\Delta F$  is 0.18 kcal/mol for NoREX and 0.30 kcal/mol for REX, respectively;  $\Delta U$  standard deviation is 4.91 kcal/mol NoREX and 4.04 kcal/mol REX, respectively. Estimates performed with BAR.

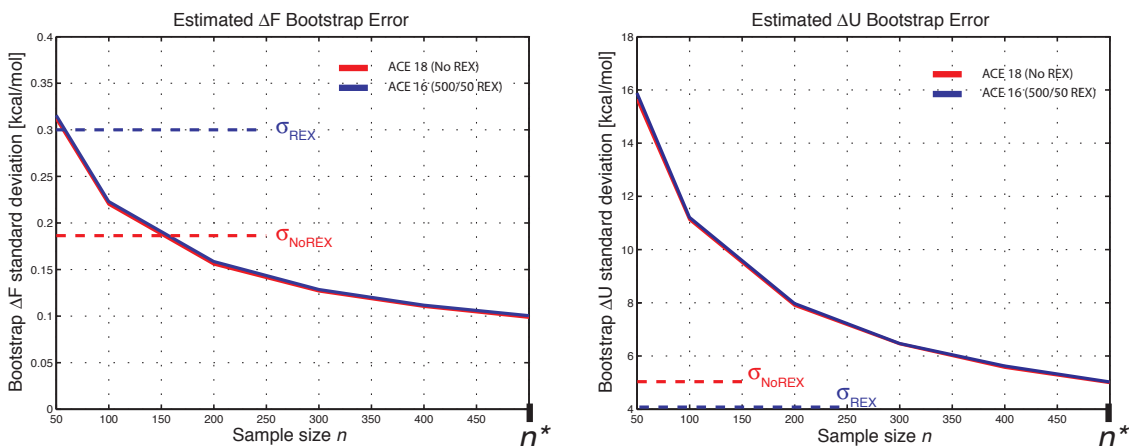


Figure 4.6: Predicted standard deviation of  $\Delta F$  and  $\Delta U$  based on bootstrap analysis of one simulation as a function of sample size (with 5000 such bootstrap samples drawn). Identical analyses of two different simulations, with (“REX”) and without replica exchange (“No REX”), are shown (blue and red lines, respectively), and are nearly indistinguishable. The expected number of independent observations per simulation,  $n^*$ , is 500 based on an autocorrelation analysis, yielding from this plot an expected statistical error of 0.10 kcal/mol for both REX and NoREX. The observed  $\Delta F$  errors (Fig. 4.5), indicated by “ $\sigma_{REX}$ ” and “ $\sigma_{NoREX}$ ” are significantly larger than this bootstrap prediction. Observed  $\Delta U$  errors are roughly consistent with bootstrap estimates.

Since replica exchange reduces the error of free energy calculations as measured by other error techniques (see e.g. Fig. 4.2 and the next section), this result suggests that variance-based error measures should be used with caution. Nevertheless, our aim here is not to analyze the source of this difference, but rather to use these independent results to validate the statistical error calculated by the bootstrap analysis of one simulation.

The number of independent observations during the simulation,  $n^*$ , is given by the total simulation length divided by twice the system correlation time (see section 2.4.2). With the correlation time of the internal energy  $\tau = 450$  cycles, we find for these simulations  $n^* = 500$ . According to theory, analysis of one simulation using the bootstrap technique with a sample size  $n^*$  yields the simulation standard error.

Figure 4.6 critically evaluates this prediction, plotting the bootstrap error (standard deviation  $\sigma$ ) versus sample size  $n$ . We find that bootstrap does not discern between

simulations with and without replica exchange, and in both cases significantly underestimates the expected errors of the simulations. A bootstrap analysis reports an expected  $\Delta F$  error of 0.10 kcal/mol for  $n = 500$ , well below the observed errors of 0.18 and 0.30 kcal/mol. The bootstrap prediction of the  $\Delta U$  error is roughly consistent with that observed in Fig. 4.6, although the error for the replica exchange simulation is somewhat overestimated.

In conclusion, we find that the bootstrap analysis of one simulation severely underestimates the statistical error of free energy calculations, yielding a standard deviation two to three times smaller than observed by running the same simulation multiple times. The source of this discrepancy is unclear, but may stem from basing the bootstrap analysis on a simulation which is too short to be representative of a population of simulations. For this simulation protocol, which is relatively short with few  $\lambda$  values, replica exchange also increases the standard deviation of independent  $\Delta F$  calculations. That better sampling leads to increased  $\Delta F$  variance suggests that sampling in the individual simulations is inadequate. In conclusion, it is clear that statistical error estimates may themselves be prone to error, particularly in the case of slowly converging quantities such as free energies, and should be analyzed critically.

#### 4.10.2 Hysteresis Error and Replica Exchange Swap Rate

An important parameter in replica exchange simulations is the frequency with which swaps are attempted. In the work described in the body of this chapter, swaps occur every 500 cycles or so (with a length given by a Gaussian distribution with a standard deviation of 50 cycles); for comparison, the autocorrelation time of the internal energy is 450 cycles. Figure 4.7 illustrates how the hysteresis error decreases with the swap rate as a function of simulation length. As in Fig. 4.2, we plot the RMS hysteresis error, averaged over all blocks of a given length, for both the free energy  $\Delta F$  and the energy  $\Delta U$ .  $\Delta U$  hysteresis error is defined analogously to the  $\Delta F$  hysteresis error as the difference of the forward and reverse FEP  $\Delta U$  estimates (see section 3.2.1).

Consistent with results from Fig. 4.2, we find that the  $\Delta F$  hysteresis errors fall more quickly with replica exchange; we also find that the more rapid the swaps, the more quickly the errors decrease. Interestingly, the corresponding error for  $\Delta U$  rises

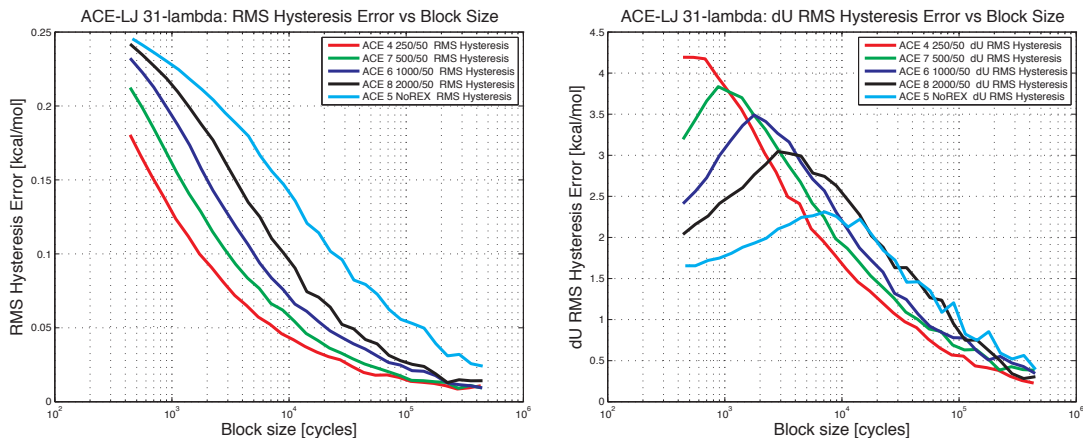


Figure 4.7: The  $\Delta F$  (left panel) and  $\Delta U$  (right panel) RMS hysteresis errors as a function of simulation length and replica exchange swap rate. The legend indicates swap rates, with e.g. “250/50” representing swaps on average every 250 cycles, and “NoREX” indicating no swaps. More frequent replica exchange attempts uniformly result in a faster decrease in errors. The system is the solvation of uncharged acetamide with a 31- $\lambda$  schedule.

at short block lengths before declining. This rise and fall shifts to shorter blocks and steepens with faster replica exchange rates. The reason for rise and fall of  $\Delta U$  hysteresis error is not entirely clear, but may have to do with the fact that measures of  $\Delta U$  are closely related to the covariance of two quantities,  $V$  and  $\partial V/\partial\lambda$  (see section 3.2.1). The inter-replica equilibration which is accelerated by replica exchange may then affect such covariance measures.

The increased convergence rates due to replica exchange may be quantified from Fig. 4.7; compared to a simulation with no replica exchange, simulations with the fastest (250/50) swap rates converge to the same error value  $> 10$  times more quickly for  $\Delta F$ , and about 3 times more quickly for  $\Delta U$ . This is in line with results in Fig. 4.2, where  $\Delta F$  a 4-8 convergence rate increase was observed for a 500/50 swap rate.

These results, together with those presented in the body of this chapter, support the notion that replica exchange is an effective way to speed the convergence of solvation thermodynamics calculations, and the more frequent the swaps the greater the improvement.



# Chapter 5

## Conclusion

The aim of this thesis is to improve the methodology of free energy of solvation calculations and their entropy/enthalpy decompositions, so that such thermodynamic quantities may be calculated with greater accuracy and precision with given computational resources. We accomplish this goal with three primary advancements:

1. Improved estimators for entropy/enthalpy decompositions.
2. Methods to improve the  $\lambda$  schedule, so as to accelerate both free energy and entropy/enthalpy calculations.
3. Improved sampling techniques – specifically, Hamiltonian replica exchange – which likewise improve simulation convergence.

These techniques, along with the conceptual improvements upon which they are based, are summarized and discussed in the following section.

We conclude with a discussion of various paths that such improvements present. First, we consider possible simulation protocols by which the motivating problem of polar collapse can be explored. Next we describe a technique by which phase transition-like events can be detected unambiguously without reference to any specific order parameter. Finally, we discuss a connection between the our analysis of Hamiltonian replica exchange and recent findings in the field of nonequilibrium statistical mechanics, which suggest further ways in which thermodynamic paths may be improved.

## 5.1 Summary of Results and Discussion

### 5.1.1 Improved Entropy and Enthalpy Estimators

There are two major families of equilibrium free energy estimators. The sole member of the first family is the thermodynamic integration method. This estimator is based on a truncated Taylor series expansion of the free energy change between two systems (Bennett, 1976).

The second family consists of the perturbation estimators, which can be categorized by the number of systems from which they simultaneously draw their information. Free energy perturbation uses information from one simulation to extrapolate the free energy difference to another  $\lambda$  value. The Bennett acceptance ratio uses data from two simulations and interpolates their difference to estimate the free energy change. Finally, the multistate Bennett acceptance ratio estimator uses data from  $K$  simulations to obtain, at once, the free energy changes between each of them and across the entire  $\lambda$  schedule. Both BAR and MBAR are derived so as to minimize the variance of the estimate  $\Delta F$ , given that only a finite number of samples are drawn from each simulation (Bennett, 1976; Shirts and Chodera, 2008). The MBAR method reduces to BAR for the case of two simulations, and BAR reduces to FEP in the case that all the data are drawn from only one simulation.

The crux of chapter 3 involves taking the temperature derivative of the BAR and MBAR estimators in order to obtain the entropy estimator formulae. One apparent complication of this approach is that there are in fact no explicit formulae for  $\Delta F$  in either of these methods; rather, both are implicit equations requiring an iterative solution, where estimators for  $\Delta F$  appear on both sides of the equals sign (c.f. (3.10) and (3.12)). This is in fact not a problem, and we can take a derivative of both sides, yielding an implicit equation for  $\Delta S$ . A slightly more complicated issue is that the MBAR  $\Delta F$  equation is written in terms of an empirical average, which needs to be cast as an ensemble average before the temperature derivative can be performed (cf. section 3.5).

While both BAR and MBAR were derived to minimize the variance of  $\Delta F$ , there is no *a priori* reason to believe that the  $\Delta S$  estimator likewise has the lowest possible

variance. It may be possible to derive a  $\Delta S$  estimator *de novo* which will formally minimize the variance of  $\Delta S$  and which differs from the one described here. Nevertheless, the  $\Delta S$  estimates obtained from the BAR and MBAR derivatives are observed to have favorable convergence properties. As illustrated in Fig. 3.6, the MBAR estimator in all cases has a standard error lower than BAR, which is in turn superior to the FEP estimates.

The remaining entropy estimator, the direct method, is a common but naïve choice for performing entropy/enthalpy decompositions. It can be effective when the enthalpy change is large, and is trivial to implement. For  $\Delta U$  changes on the order of  $k_B T$ , however, it is a poor choice because of its poor convergence qualities. For estimating the enthalpic contribution to the mean force profile, which is typically calculated for a fine  $\lambda$  schedule with correspondingly small enthalpy changes, the direct estimator is so noisy as to be practically useless. It is in such situations where the MBAR estimator, which makes maximal use of information from all simulations, excels.

### 5.1.2 Improved $\lambda$ Schedule

The insertion of a solute into solvent, particularly if the solute is large, involves considerable reorganization of the solvent molecules. Splitting the insertion process into many substeps by means of multiple simulations along the  $\lambda$  schedule is a classic technique used to aid convergence (see Mezei and Beveridge (1986) for a historical perspective).

A variety of different ways have been presented in the literature to construct improved  $\lambda$  schedules (Pearlman and Kollman, 1989; Lu and Kofke, 1999; Shirts et al., 2003), although simple linear  $\lambda$  schedules are still commonly employed. We construct our  $\lambda$  schedule so that it directly minimizes a specific measure of the error associated with free energy calculations.

The configuration probability distributions of two systems with different  $\lambda$  values are not independent of one another. Just like the relative probability of observing two different configurations in one simulation is given by the Boltzmann distribution, so the probability of observing given work values when swapping configurations between the

independent simulations is given by the Crooks fluctuation theorem (Crooks, 1999). That is, when one simulation is well sampled, its internal probability distributions satisfy the Boltzmann distribution; when two independent simulations are both well sampled, configuration swaps between them satisfy the Crooks fluctuation theorem.

The hysteresis error presented in chapter 4 quantifies how closely the Crooks fluctuation theorem is satisfied; it can be understood as a measure of inter-replica equilibration. Operationally, however, it is defined as the difference between the forward and reverse FEP  $\Delta F$  estimates. If the estimates agree then the hysteresis error is zero, and the simulations are well converged.

In chapter 4, we propose that the rate at which the hysteresis error decreases is governed by the replica exchange swap probability. This swap probability – technically, the Fermi swap probability, which has the same essential qualities as the Metropolis swap probability (see section 2.1.2) – is a measure of the overlap in phase space of the ensembles of the two simulations. The larger the swap probability the more similar the two systems are, and the more likely they are to sample the regions of phase space which lead to the convergence of the FEP estimators. Thus, the larger the swap probability the more quickly the hysteresis error decreases, the FEP calculations converge and the Crooks fluctuation theorem is satisfied.

An optimal  $\lambda$  schedule is one for which the free energy estimate for the entire calculation converges most quickly. Since the error of the whole calculation is governed by the error of the least accurate section along the  $\lambda$  schedule, we define the optimal  $\lambda$  schedule as one for which the hysteresis error decreases uniformly quickly everywhere. That is, the optimal  $\lambda$  schedule will have a uniform swap probability between all replicas.

The swap probability, however, needs two simulations in order to be evaluated, and varies nonlinearly with  $\delta\lambda$ , defined as the distance along the  $\lambda$  schedule between two simulations. To permit its evaluation based on the results of only one simulation, we expand the swap probability in a Taylor series and find that it is proportional to  $\delta\lambda^2 \text{var}(\partial V/\partial\lambda)$  (cf. Eq. (4.18)). This leads to a protocol by which the results of a series of trial simulations are used as the basis for constructing an improved  $\lambda$  schedule.

One concern is that this methodology is specific only to FEP  $\Delta F$  estimators, as they are the basis for the hysteresis error formulation. Would a schedule which is optimal for FEP  $\Delta F$  also be optimal for, say, MBAR  $\Delta U$  calculations? Figures 3.5 and 3.6 address this question by plotting the standard errors of various  $\Delta F$  and  $\Delta U$  estimators alongside the swap probability. It is clear that “difficult” regions of the  $\lambda$  schedule – where the ensemble is changing most quickly – correspond to areas of relatively large errors for *all* estimators as well as low swap probabilities. Improving the  $\lambda$  schedule by adding additional simulations in this region (compare, for instance the  $21\lambda$  and  $29\lambda$  results) reduces all errors as it increases the swap probability.

### 5.1.3 Improved Sampling with Hamiltonian Replica Exchange

Replica exchange (also known as parallel tempering) employs a Monte Carlo move to swap system parameters between multiple concurrent simulations, so that kinetic bottlenecks can be overcome more quickly and sampling improved. The original and most common implementations use temperature as the parameter (Hukushima and Nemoto, 1996; Hansmann, 1997; Nymeyer et al., 2004). Hamiltonian replica exchange (Sugita and Okamoto, 1999; Fukunishi et al., 2002) instead swaps some parameter which modifies the Hamiltonian. The natural choice is the same  $\lambda$  parameter used in solvation calculations, and this approach has been used in the context of free energy calculations previously (Sugita et al., 2000; Woods et al., 2003a,b).

The contribution of this thesis is in the analysis of Hamiltonian replica exchange (although the non-neighbor swap algorithm, section 2.1.4, is to our knowledge novel). By casting the problem of independent replicas in the context of the Crooks fluctuation theorem, as described in the previous section, we clarify and quantify the concept of inter-replica equilibration. Then, just as a Monte Carlo move of, say, the displacement of a solvent molecule in a simulation leads to the accumulation of samples which obey the Boltzmann distribution, so a replica exchange move works to satisfy the work distributions required by the Crooks fluctuation theorem. By allowing the “jumping” over barriers, Hamiltonian replica exchange leads to the Crooks fluctuation theorem being satisfied more quickly.

Also, by rewriting the replica exchange swap probability in terms of the Fermi rather than the Metropolis criteria, we can formally interpret the average of the swap probability as the overlap of two ensembles.

Previous work on Hamiltonian replica exchange free energy calculations (Sugita et al., 2000; Woods et al., 2003a,b) focused on the variance of the resulting estimates as a measure of the improvement of sampling. This can be misleading, as improved sampling can in fact lead to increased variance under some conditions (e.g. a system escaping local minima; see also section 4.10.1). The hysteresis error is a much less ambiguous error measure and leads to a clear demonstration of the benefits of replica exchange. We show in Fig. 4.7 that both free energy and entropy estimates are improved by Hamiltonian replica exchange, and that faster swaps lead to accelerated convergence, a result otherwise demonstrated only for temperature replica exchange (Sindhikara et al., 2008).

The optimal replica exchange swap frequency is not clear, and is dependent on the measure of efficiency one wishes to maximize as well as on the specifics of the replica exchange implementation. For instance, in the Towhee simulation engine (used in chapter 4) each round requires a restart of the program, which incurs significant overhead; even if this were avoided, swaps typically require the synchronization of the different replicas (no swaps occur until the last one finishes) and inter-process communication. Such restrictions limit the efficiency of the technique at very high swap rates if one measures efficiency by the “wall clock time” spent performing the simulation. In general, however, it would appear that the less overhead is incurred by Hamiltonian replica exchange swaps, the more frequently they should be performed to maximize sampling efficiency.

Another open question is the balance between the number of replicas versus the simulation time per replica, if the product of these two is held fixed such that the total computer simulation time is constant. One could envision, for instance, a simulation consisting of hundreds or even thousands of replicas along a  $\lambda$  schedule, each performing rapid Hamiltonian replica exchange swaps (perhaps asynchronously) and running for a relatively short period of time. Would this approach be more effective than the current model of a dozen or two simulations performing lengthy simulations? Once

again, the tradeoffs are closely tied to the algorithm implementation, but it is plausible that denser schedules increase sampling efficiency more quickly than increased simulation time.

## 5.2 Future Directions

### 5.2.1 Sampling of Flexible Chains

The motivation for this work, as discussed in section 1.4, is to understand why polar peptides – in contradiction to expectations based on transfer models – form compact globules in water. Figure 5.1 illustrates the behavior of glycine 15-mers in water and in the excluded volume limit, the latter simulating chain behavior in a good solvent. The histogram of two quantities, the radius of gyration and asphericity (a measure of shape), is plotted for both solvent conditions. This plot makes clear that chains in water tend to collapse into compact, spherical globules, whereas in the EV limit the chains are swollen and elongated (Tran et al., 2008).

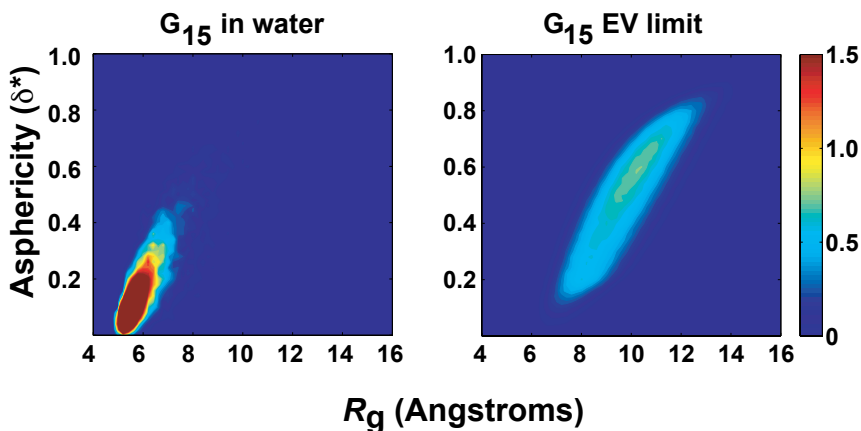


Figure 5.1: The radius of gyration ( $R_g$ ) and asphericity ( $\delta^*$  is a measure of shape, with  $\delta^* \approx 0$  for spherical shapes) for glycine 15-mers in different solvent conditions. It is observed that the polypeptides collapse in water but swell and elongate in the EV limit. We define the compact ensemble as that observed in water, and the swollen ensemble as the EV limit. Figure from Tran et al. (2008).

We can consider Fig. 5.1 as defining two ensembles: polypeptide conformations ( $\Gamma$ ) drawn from the water simulation belong to the ensemble  $\gamma_{compact}$ , and those taken from the EV simulation are members of the  $\gamma_{swollen}$  ensemble. These two ensembles overlap very little. Since in water the compact conformations are preferred, by definition they have a lower free energy. That is,

$$G_{water}(\gamma_{compact}) < G_{water}(\gamma_{swollen}), \quad (5.1)$$

where the free energy of an ensemble is defined as the expectation value of the free energy for a population of fixed conformations drawn from that ensemble. For instance,

$$G_{water}(\gamma_{compact}) = \langle G_{water}(\Gamma) \rangle_{compact}. \quad (5.2)$$

We now consider two related but distinct questions:

1. How to measure,

$$\Delta G_{collapse} = G_{water}(\gamma_{compact}) - G_{water}(\gamma_{swollen})? \quad (5.3)$$

This is the free energy difference between the swollen and compact conformations in water (akin to  $\Delta G_{fold}^{\circ}$  defined in section 1.1) and quantifies the stability of the collapsed state. We can define  $\Delta S_{collapse}$  and  $\Delta U_{collapse}$  similarly, and these give insight into the nature of the thermodynamic driving force which leads to collapse.

2. How does the total free energy of a flexible glycine N-mer,  $G(N)$ , vary with chain length N? That is, we wish to quantify

$$\Delta G_{grow}(N) = G_{water}(N + 1) - G_{water}(N), \quad (5.4)$$

and its decomposition into  $\Delta S_{grow}(N)$  and  $\Delta U_{grow}(N)$ . This question addresses the nonlinearity of the concatenation process and the failure of the additivity model.

The solvation calculation techniques presented in this thesis provide a methodological way forward but do not, in and of themselves, directly address these questions. In



particular, solvation calculations of flexible chains cannot calculate  $\Delta G_{collapse}$ , since the chain will collapse when fully coupled with the water and never explore the swollen ensemble. The fundamental problem is that we wish to evaluate free energy changes with respect to changes in the conformational ensemble, but we have no control of this ensemble. In their current form, solvation calculations will report on the free energy difference between a polypeptide *in vacuo* and a collapsed chain in water.

Solvation calculations may be able to approach the second problem of calculating  $\Delta G_{grow}(N)$ , but not directly and not without assumptions. Figure 5.2 illustrates the issue in terms of a thermodynamic cycle. We wish to calculate  $\Delta G_{grow}(N)$ , the free energy change of increasing chain length by one unit in water. At our disposal are calculations which report on  $\Delta G_s(N)$  and  $\Delta G_s(N+1)$ , the free energy change associated with transferring a flexible chain of a given length from vacuum into water. Since this transfer process is path independent, we can obtain the quantity of interest as,

$$\Delta G_{grow}(N) = \Delta G_s(N+1) - \Delta G_s(N) + \Delta G_{grow}^{vac}(N). \quad (5.5)$$

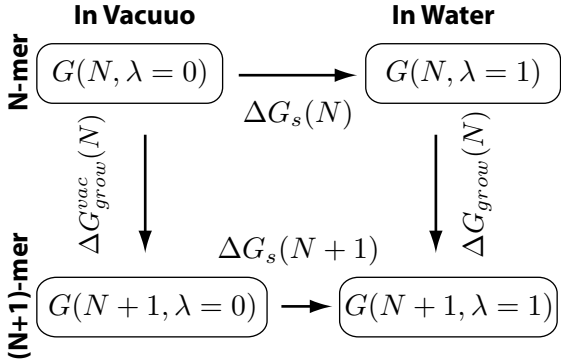


Figure 5.2: A thermodynamic cycle illustrating polypeptide transfer and growth processes.

To evaluate  $\Delta G_{grow}(N)$ , we need to obtain  $\Delta G_{grow}^{vac}(N)$ , the free energy change associated with increasing the length of the chain *in vacuo*, that is, with  $\lambda = 0$ . It may be possible to estimate this quantity, although the conformation of the chain at  $\lambda = 0$  would have to be determined. The conformation ensemble for a flexible molecule at  $\lambda = 0$  is in fact dependent on the particulars of the choice of the scaling potential (see section 2.1.3). If solute-solute interactions are not scaled, then the polypeptide

will assume a globular conformation, whereas if solute-solute interactions are scaled, then the chain will likely be in the EV limit. An estimate of  $\Delta G_{grow}^{vac}(N)$  would need to take these details into account. Another problem with this approach is that it is very inefficient; two solvation processes need to be performed with high accuracy in order to calculate  $\Delta G_{grow}(N)$ , even though information at  $\lambda \neq 1$  is of no interest. Thus while solvation calculations may be able to address question 2 above, they are ill designed to do so.

Fortunately, the techniques presented here can estimate free energy changes and their decompositions for any process parameterized by  $\lambda$ , not just the insertion of a whole particle into solvent, and we have a great deal of latitude in choosing this process. In whatever process we design, we wish to perturb the system as little as possible as  $\lambda$  varies, so as to maximize ensemble overlap and the rate of convergence.

The approaches described below can be implemented using current technology. Preliminary work has been performed on some of these approaches using the GROMACS (Hess et al., 2008) simulation engine, which is well suited to these tasks, principally because of its speed and flexibility. The main drawback of GROMACS is that, at present (version 4.0.5), it lacks the ability to output foreign energies. As discussed in section 2.5 this can to some extent be overcome in postprocessing for calculations involving the MBAR methods. Foreign energies must be available at run time to implement Hamiltonian replica exchange, however, and this method is currently not possible in GROMACS.

## Frozen Conformation Solvation

A straightforward way to estimate  $\Delta G_{collapse}$  is to solvate some number of frozen polypeptide conformations which are representative of the  $\gamma_{compact}$  and  $\gamma_{swollen}$  ensembles. That is, we can choose conformations at random from each of the ensembles illustrated in Fig. 5.1, hold them immobile, and proceed with the solvation process just as for the small model compounds to calculate the solvation free energy and its entropy/enthalpy decomposition. In effect, this technique estimates the quantity  $G_{water}(\gamma_{swollen/collapsed})$  in Eq. (5.2) by an empirical average over a finite number of fixed conformations, and the greater the number of such conformations are sampled,

the better converged the estimate will be. This approach is in the spirit of Ben-Naim (1987), who generally focuses on the solvation of compounds with no conformational degrees of freedom.

It is not clear *a priori* how many such calculations would be needed and what level of accuracy is required to distinguish, beyond statistical error, differences between the ensembles. If the differences are clear and distinct, then just a few calculations would suffice. If the differences are subtle, or very sensitive to the details of the conformation, then performing tens or hundreds of calculations may be necessary. This would likely be a poor investment of resources, and other techniques would likely be preferred. Its simplicity, however, make this approach an obvious first choice.

## Chain Growth Simulations

To calculate the quantity  $\Delta G_{grow}(N)$  and address directly chain length-dependent effects, we can consider a simulation where the  $\lambda$  parameter grows in only a terminal amino acid in a peptide chain. Here, for  $\lambda = 0$  the chain is effectively of length  $N$ , and for  $\lambda = 1$  it is of length  $N + 1$ , so that both the  $\lambda = 0$  and  $\lambda = 1$  systems are physically relevant.

More specifically, at  $\lambda = 0$  the N-mer and water would interact normally, with a terminal ghost residue which does not interact with either. Interactions of the terminal residue with the rest of the system would then scale with  $\lambda$ , as if simulating the solvation of the terminal residue onto the tail of a chain. Such simulations could be performed for any  $N$ , and this technique could access length-dependent effects directly. Since growing in one residue would presumably perturb the system only modestly, swap probabilities should be reasonable with a  $\lambda$  schedule which is coarser than that required to grow, say, a whole chain in from scratch.

One complication is that the ghost terminal group would need to be identical to the  $N^{th}$  group. Typically, simulations of polyglycine (and other polypeptides) have acetyl and n-methyl amide capping groups to preserve chemical accuracy and ensure neutral charge groups. In this simulation protocol, however, it would be advantageous to avoid mutating a capping group into an internal group while adding a new terminal group. It seems reasonable that an appropriate synthetic system of  $N$  identical repeating

units in a chain (with no additional capping groups) could be constructed while approximating the behavior of a physical polypeptide.

Such a system could address directly the linear additivity of peptide groups into a chain, interpolating all the way from solvating a single amide group to the long chain limit, and could readily be implemented in e.g. the GROMACS simulation engine.

## Inducing Coil-Globule Transitions

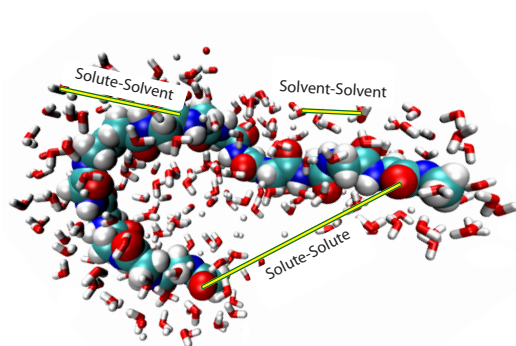
Neither of these approaches can investigate the transition from poor to good solvent directly. The first method considers conformations which are by construction in one ensemble or the other, while the second evaluates the free energies of the conformations naturally assumed by the solute. It would be helpful, for the investigation of the collapse of polar peptides, to be able to access the coil-to-globule transition directly in explicit solvent calculations.

It is, in fact, possible to populate the coil state in an explicit solvent model, and recent work has demonstrated two different parameters which can induce a coil-to-globule solvent transition for a glycine 15-mer. The scaling of either solute-solute or solute-solvent dispersions<sup>6</sup> will induce a transition from the swollen to the collapsed state; others (Polson and Moore, 2005) have used similar approaches. The  $\lambda$  parameter governs the magnitude of such dispersions, with  $\lambda = 1$ ,  $\lambda < 1$  and  $\lambda > 1$  representing a system with normal, reduced and strengthened dispersions, respectively. No other interactions in the system are modified. The radius of gyration order parameter in Fig. 5.3 captures the coil-to-globule transition of a polyglycine chain.

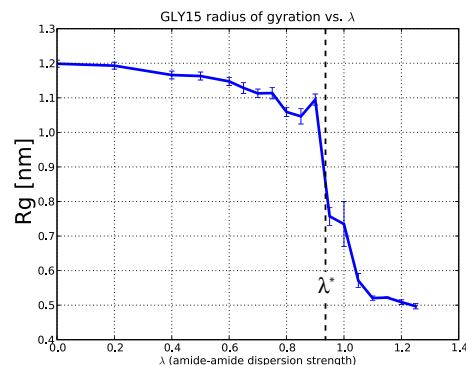
Given that we can induce a coil-to-globule transition with the parameter  $\lambda$ , how can we use this methodology to investigate the driving force for polar collapse? A first step would be to observe how  $\Delta F$ ,  $\Delta U$  and especially  $T\Delta S$  change during the putative phase transition, and perhaps parse these system-wide changes into contributions from the solvent and solute individually (Peter et al., 2004). A more formal procedure might be based on the  $m$ -value analysis common in denaturation experiments (Pace, 1986). With the assumption that the good-to-poor transition is a two-state process, the average radius of gyration at a given  $\lambda$  can be used to estimate the free

---

<sup>6</sup>Dispersion forces refer to the attractive component of Lennard-Jones interactions.



(a) Glycine 15-mer in water with solute, solvent combination pairs labeled.



(b) Radius of gyration of glycine 15-mer as solute-solute dispersion interactions are scaled. The value  $\lambda^*$  indicates the mid-point of the coil-to-globule transition.

Figure 5.3: A coil-to-globule transition can be induced in a glycine 15-mer system with explicit water by scaling the strength of solute-solute dispersion forces with the parameter  $\lambda$ .

energy as well as the entropy and enthalpy of folding, independent of the systematic perturbations induced by the modification of the Hamiltonian.

Both approaches seem reasonable and should be informative. In the next section we will consider another approach, one based on analyzing the coil-to-globule collapse as a phase transition, which may yield quantitative estimates of the free energy penalty associated with the peptide collapse.

## 5.2.2 Phase Transitions

We can induce a coil-to-globule transition in a glycine 15-mer in explicit water by modulating the solute-solute dispersion forces with the parameter  $\lambda$ , as illustrated in Fig. 5.3b. Due to their cooperative nature coil-to-globule transitions, as well as the protein folding process itself, are frequently modeled as phase transitions (Grosberg and Kuznetsov, 1992; Ivanov et al., 1998; Pande et al., 1998; Tcherkasskaya and Uversky, 2001; Shakhnovich, 2006). In this section, we will de-emphasize the specific problem of polypeptide collapse and consider phase transitions more generally, with an emphasis on how they can be detected in computer simulations in general, and return to the issue of coil-to-globule at the conclusion of this section.

## Polypeptide Collapse as a Phase Transition

Whether the coil-to-globule transition is a phase transition in the sense of classical thermodynamics (Callen, 1985) or simply a useful analogy (Mirny and Shakhnovich, 2001) is open to debate, as the precise nature of the collapse transition has proven difficult to establish (Shakhnovich and Finkelstein, 1989; Finkelstein and Shakhnovich, 1989; Kuznetsov and Timoshenko, 1999; Ivanov et al., 1998). The study of phase transitions in general has long been a topic of great interest in a variety of scientific fields and plays a central role in many areas of physics. Formally, such phenomena occur only in the thermodynamic limit of systems of infinitely many particles (Borrmann et al., 2000). Proteins, on the other hand, are necessarily of finite length, as are computer simulations. Still, mathematical models which describe phase transition-like behavior in small systems have been constructed, and these models converge to classical definitions of phase transitions in the thermodynamic limit (Borrmann et al., 2000; Gross, 2001; Pleimling and Behringer, 2005; Proykova and Berry, 2006). Our aim here is not to review concepts of phase transitions or to commit to its nature in small systems, but rather to discuss a set of techniques by which cooperative, phase transition-like phenomena can be observed in computer simulations. In the following, we use the term “phase transition” according to the definition of Gross (2001) (which will be discussed), with the understanding that by other definitions the term “phase transition-like” may be more appropriate.

Fundamentally, first order phase changes are caused by an instability in the underlying thermodynamic potential and are marked by discontinuous changes in the molar entropy and enthalpy of the system (Callen, 1985). In practice – that is, experimentally and typically in computer simulations – phase changes are indicated by a sudden change in an order parameter which characterizes some aspect of the system (Landau and Binder, 2000), for instance the density (in a fluid system) or radius of gyration (in a polypeptide system). In general a whole host of quantities changes simultaneously and there is no single unique order parameter by which to characterize the phase transition (Proykova and Berry, 2006). Theory aside, one problem with such an approach to small systems, such as the collapse of polypeptides, is that sharp transitions in order parameters occur only in the thermodynamic limit, and observing them gets increasingly difficult as system size decreases. Also, specifying the appropriate order parameter can be difficult.

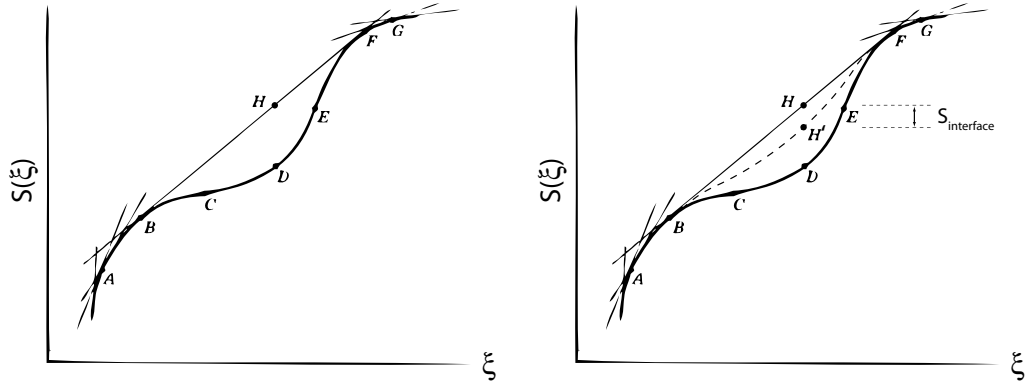
Nevertheless, the instabilities in the thermodynamic potentials which are at the root of phase transitions still exist in small systems. These form the basis of a technique to detect phase transitions in small systems which is independent of the choice of order parameters. The aim of the remainder of this section is to introduce the concepts behind this technique, and then to outline how these ideas may be used in the case of polypeptide collapse like the one illustrated in Fig. 5.3b. The motivation is the work of Junghans et al. (2006, 2008), who describe the thermodynamics of the aggregation of a simple polypeptide model, and the theory is largely based on Gross (2001). The extension to the case of the parameter  $\lambda$  is novel to this work.

### Thermodynamics of Small Systems

The entropy of a system in the microcanonical ensemble,  $S(\xi)$ , is maximized for a system at thermodynamic equilibrium (Callen, 1985). Here,  $\xi$  is any extensive variable, for instance the total internal energy  $U$ .

Suppose that the thermodynamic potential  $S(\xi)$  is concave in a section of its domain, like illustrated in Fig. 5.4a. In the region marked by the point  $A$ , the system will exist in a single phase, as a homogeneous system will maximize the entropy for the given quantity  $\xi_A$ . In the region marked by point  $D$ , however, the situation is different. The system will not stay in a homogeneous state with the entropy indicated by point  $D$ , as this situation is not thermodynamically stable.

To illustrate, imagine that we divide the system into two parts and transfer some quantity  $\Delta\xi$  from one part to the other, so that (on a per mole basis) one part has  $\xi_B = \xi_D - \Delta\xi$  and the other  $\xi_F = \xi_D + \Delta\xi$ . Given that for large systems entropy is extensive (Touchette et al., 2004), the total system entropy is given by the point  $H$  on the line drawn from  $B$  to  $F$  according to the lever rule (Callen, 1985). The entropy of point  $H$  is larger than for  $D$ , and the partitioning we proposed would occur spontaneously. In fact, anytime  $\xi$  is between  $\xi_B$  and  $\xi_F$  the system will separate into two phases.



(a) In the thermodynamic limit, a convex entropy surface is never observed and the total system entropy will lie on the line  $BHF$ .

(b) In small systems, the entropic penalty associated with the interface ( $S_{interface}$ ) is significant, and the observed system entropy will lie on the curve  $BH'F$ . As a result, a convex entropy surface is observed.

Figure 5.4: Convex regions of the entropy  $S$  versus an extensive quantity  $\xi$  lead to first order phase transitions. Figure adapted from Callen (1985).

This relationship between the shape of  $S(\xi)$  and the stability of the phases can be expressed as,

$$\frac{\partial^2 S}{\partial \xi^2} \leq 0. \quad (5.6)$$

If, at a given value of  $\xi$  this is not the case, the system will phase separate and restore this relationship. For large systems, violations of Eq. (5.6) are not observed.

In small systems the situation is different, because entropy is no longer extensive. The interface which occurs between phases incurs an entropy cost which is not negligible (Janke, 1998; Junghans et al., 2008; Proykova and Berry, 2006; Gross, 2001; Gross and Kenney, 2005). Bulk energy terms scale with the number of particles as  $N$ , whereas interface terms grow as  $N^{2/3}$ . As a result, in the thermodynamic limit the bulk terms dominate and interface effects can be neglected. For small systems, however, the interfaces form a significant part of the energy and cannot be neglected (Hill, 1994; Sheehan and Gross, 2006). For a small system separated into two phases, the total system entropy is given as,

$$S_{system} = S_{phase A} + S_{phase B} + S_{interface}. \quad (5.7)$$



The presence of the interface entropy term has important consequences on the stability criterion, Eq. (5.6). Figure 5.4b illustrates what the observed relationship between  $S$  and  $\xi$  may look like. The additional entropy term shifts the state  $H$  predicted by the lever rule line down by an amount  $S_{interface}$ , such that the combined, phase separated system has a decreased entropy marked by  $H'$ . As a result, an apparent violation of the stability criterion occurs, and a convex entropy relationship,

$$\frac{\partial^2 S}{\partial \xi^2} > 0 \quad (5.8)$$

is observed in this region (Gross, 2001). This condition is an *indicator* of a first order phase transition in a small system, and is the central conceptual result of this section.

Gross (2001) *defines* a phase transition for a finite system by regions of non-negative curvature of the entropy surface as a function of conserved extensive quantities. For the general case of entropy as a function of multiple extensive quantities,  $S(\xi_1, \xi_2, \dots)$ , we define the Hessian  $\mathcal{H}(S)$  as a matrix of second derivatives of  $S$  with respect to  $\xi_i$  (Hale and Koçak, 1996),

$$\mathcal{H}(S) = \frac{\partial^2 S}{\partial \xi_i \partial \xi_j}, \quad (5.9)$$

and consider its eigenvalues  $\nu_i$ . A first order phase transition occurs when one of the  $\nu_i > 0$ . Second order transitions and critical points can also be defined in terms of the eigenvalues of  $\mathcal{H}$  (Gross, 2001).

An important consequence of phase transitions in small systems is that the micro-canonical and canonical ensembles are no longer equivalent, so that phenomena which can be observed in the former are obscured in the latter (Touchette et al., 2004). Formally, this comes from the failure of the Legendre transform whenever Eq. (5.8) holds true (Callen, 1985). Because the slope of the curve  $(\partial S / \partial \xi)$  in Fig. 5.4b does not vary monotonically with  $\xi$ , it ceases to be a useful control parameter in the vicinity of the convex region.

An example from Junghans et al. (2006) helps to illustrate this point. The authors analyze a model system of two model amphiphilic peptides in implicit solvent and sample it exhaustively to obtain the density of states. From this the entropy  $S$  as a function of the internal energy  $U$  is obtained. Effectively, the authors consider a

system like in Fig. 5.4b where the extensive parameter  $\xi = U$ , the energy of the system; then  $\partial S/\partial U = 1/T$  defines the intensive parameter, the temperature.

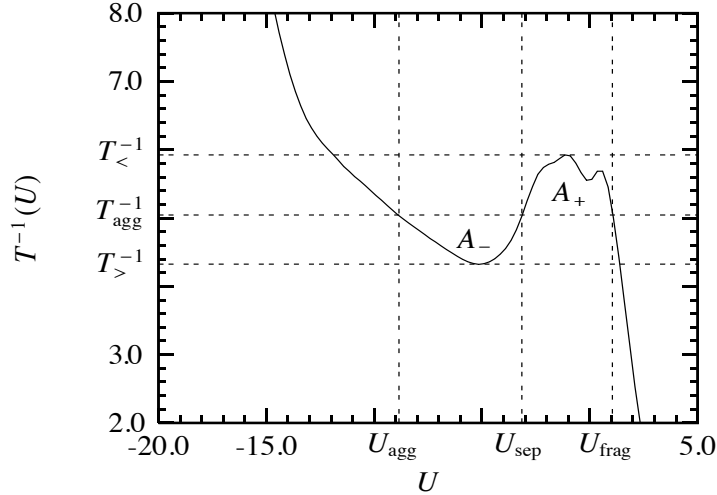


Figure 5.5: A plot of the the inverse temperature versus energy for a system of two model peptides, from Junghans et al. (2006). The region with a positive slope of  $T^{-1}$  indicates a phase transition. Here, the temperature temporarily decreases with increasing energy. The equal areas  $A_+$  and  $A_-$  quantify  $\Delta S_{interface}$ . Note that a phase transition is observed unambiguously on the basis of a thermodynamic instability, with no reference to any specific order parameter.

Fig. 5.5 plots  $T^{-1}$  versus  $U$  and illustrates the “backbending”, resulting from a concavity of  $S(U)$ , characteristic of a phase transition. Here, the temperature ceases to be a suitable external control parameter. To illustrate, suppose that we hold the temperature constant at  $T^{-1} = T_{agg}^{-1}$ ; this temperature corresponds to the coexistence of two phases, one characterized by  $U = U_{agg}$  and the other by  $U = U_{frag}$ . The state of the system is some uncontrolled combination of the two phases (e.g. it may be history dependent), and the macrostate of the system cannot be adjusted by fixing the temperature. Instead, the phase transition is more favorably analyzed in the microcanonical ensemble where the extensive variable ( $U$ , in this instance) is held fixed (Junghans et al., 2006, 2008; Bachmann et al., 2005).

A fair amount of insight into the behavior of the system can be obtained from an analysis of Fig. 5.5. Below  $U_{agg}$  the system is in an aggregated state, and above  $U_{frag}$  the system fragments and monomers dissociate; the phase transition occurs in the

region  $U_{agg} < U < U_{frag}$ . The positive slope of  $T^{-1}$  indicates a decreasing temperature with increasing energy: additional energy leads to cooperative rearrangements of monomers in the aggregate in order to reduce surface tension; as a consequence, kinetic energy is transferred into potential energy and the temperature decreases. That is, the aggregate becomes colder, although the total energy increases (Junghans et al., 2006).

Finally, the entropy penalty associated with the interface,  $S_{interface}$ , can be obtained from Fig. 5.5 as the areas  $A_-$  and  $A_+$ . These areas are defined by the Maxwell construction (Chandler, 1987), a horizontal line which divides the lobes of the “S” curve into equal areas. Either area  $A_+$  or  $A_-$  quantifies  $S_{interface}$ .

Similar results have long been observed in computer simulations of melting systems (see e.g. Allen and Tildesley (1987), Fig. 11.2a). Such phenomena are a general result of convexity in the entropy function in small systems stabilized by interfaces (Wales and Berry, 1994).

To summarize, plotting the intensive parameter (e.g. the temperature) while holding the extensive parameter (e.g. the energy) fixed will allow phase transitions in small systems to be identified as a “backbending” or S curve. This backbending will not be observed when the intensive parameter is controlled, and may be used to quantify the free energy penalty associated with the creation of the interface.

### **Observing Phase Transitions with Respect to $\lambda$**

The proposal is that a similar approach can be used to investigate the coil-to-globule transition of polypeptides in explicit solvent with variable solute-solute dispersions. Simply replicating the approach of Junghans et al. (2006) is not possible, however. First, we cannot exhaustively enumerate the density of states, a central feature of their approach, since doing so in explicit solvent would be computationally prohibitive. In addition, our control parameter is the intensive parameter  $\lambda$ , whereas we need to control an extensive variable to observe the “backbending” upon which the analysis rests.

To preview our approach, we find that it is not necessary to estimate the density of states of a system in order to proceed with the analysis of microscopic phase transitions, as thermodynamic instabilities can be observed by considering only the derivative of the thermodynamic potential (Wales and Berry, 1994). Given that, we perform calculations at multiple fixed  $\lambda$  values and obtain the multicanonical potential as a function of  $\lambda$  and its conjugate force  $\xi$  in post-processing. Then we switch ensembles and obtain details of the system as if the extensive variable were held fixed. This then allows us to proceed with an analysis like that in Junghans et al. (2006).

The rest of this section describes this procedure in detail, and concludes with a prescription and a demonstration of how such a calculation would proceed. The mathematics and concepts utilized here are based on standard Legendre transforms (Callen, 1985) and potentials of mean force calculations (see e.g. Roux (1995)).

**Derivation** The thermodynamic potential associated with the microcanonical ensemble is the entropy,  $S(U, \xi)$ , which we have written as a function of two extensive variables. In physical systems the entropy is a monotonically increasing function of the energy and hence invertible, so that we can write  $U(S, \xi)$ . This is called the energy convention (Callen, 1985).

So that we can keep the temperature rather than the entropy constant we switch ensembles by means of a Legendre transform to obtain a new thermodynamic potential,

$$L(T, \xi) = U - TS, \tag{5.10}$$

in which the temperature and  $\xi$  are controlled.

In our simulations it is  $\lambda$  rather than  $\xi$  which is controlled, so we switch ensembles again with another Legendre transform we obtain,

$$F(T, \lambda) = L - \lambda\xi. \tag{5.11}$$

This is the free energy we calculate in canonical simulations at a fixed  $\lambda$ .

To identify  $\xi$ , we perform a full derivative of  $F$ ,

$$dF(T, \lambda) = \left( \frac{\partial F}{\partial T} \right)_\lambda dT + \left( \frac{\partial F}{\partial \lambda} \right)_T d\lambda, \quad (5.12a)$$

$$= -SdT + \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda, \quad (5.12b)$$

$$= -SdT + \langle \xi \rangle_\lambda d\lambda. \quad (5.12c)$$

That is, we identify  $\xi = \partial V / \partial \lambda$  as the extensive conjugate force to the constraint  $\lambda$  we impose on the system.

We will focus on simulations at a fixed temperature and ignore any phase changes with respect to  $T$ . Therefore, the instability condition (5.8) which indicates a phase transition can be written in terms of the thermodynamic potential  $L$  as,

$$\frac{\partial^2 L}{\partial \xi^2} < 0, \quad (5.13)$$

where the sign has switched due to the energy convention. At a fixed temperature,

$$dL(\xi) = \left( \frac{\partial L}{\partial \xi} \right)_T d\xi, \quad (5.14a)$$

$$= \langle \lambda \rangle_\xi d\xi, \quad (5.14b)$$

so that the instability condition (5.13) is,

$$\frac{\partial \langle \lambda \rangle_\xi}{\partial \xi} < 0. \quad (5.15)$$

To detect phase transitions, then, we plot  $\langle \lambda \rangle$  versus  $\xi$ ; a negative slope then signals a phase transition.

Since, our simulations are performed at fixed  $\lambda$ , not  $\xi$ , we need to switch ensembles from  $(N, V, T, \lambda)$  to  $(N, V, T, \xi)$  in post-processing. The remainder of this section describes how to do so with computer simulation data.

**Implementation** We define a multi-canonical potential  $W(\lambda, \xi)$  as,

$$\exp[-\beta W(\lambda, \xi)] \equiv \int d\Gamma \exp[-\beta V_\lambda(\Gamma)] \delta \left[ \frac{\partial V}{\partial \lambda}(\Gamma, \lambda) - \xi \right], \quad (5.16)$$

where  $\delta$  is equal to 1 when  $\partial V/\partial \lambda = \xi$  and 0 otherwise. Thus,  $W(\lambda, \xi)$  is a potential of mean force in the two coordinates, which are treated independently. We may then define the relative probability of observing  $\xi$  at a fixed  $\lambda$ ,  $\rho_\lambda(\xi)$ , as well as  $\rho_\xi(\lambda)$ , as

$$\rho_\lambda(\xi) = \frac{\exp[-\beta W(\lambda, \xi)]}{\int d\xi \exp[-\beta W(\lambda, \xi)]}, \quad (5.17a)$$

$$\rho_\xi(\lambda) = \frac{\exp[-\beta W(\lambda, \xi)]}{\int d\lambda \exp[-\beta W(\lambda, \xi)]}. \quad (5.17b)$$

$W$  relates the thermodynamic potentials as,

$$\exp[-\beta F(\lambda)] = \int d\xi \exp[-\beta W(\lambda, \xi)], \quad (5.18a)$$

$$\exp[-\beta L(\xi)] = \int d\lambda \exp[-\beta W(\lambda, \xi)], \quad (5.18b)$$

and

$$\frac{\partial F}{\partial \lambda} = \left\langle \frac{\partial W}{\partial \lambda} \right\rangle_\lambda = \langle \xi \rangle_\lambda, \quad (5.19a)$$

$$\frac{\partial L}{\partial \xi} = \left\langle \frac{\partial W}{\partial \xi} \right\rangle_\xi = \langle \lambda \rangle_\xi. \quad (5.19b)$$

Finally, we can obtain the average  $\lambda$  at a given  $\xi$  as,

$$\langle \lambda \rangle_\xi = \frac{\int d\lambda \lambda \rho_\lambda(\xi) \exp[-\beta F(\lambda)]}{\int d\lambda \rho_\lambda(\xi) \exp[-\beta F(\lambda)]}. \quad (5.20)$$

Room for improvement remains. In particular, the formula for calculating the potential  $W$ , Eq. (5.16), could no doubt be improved; it is, on inspection, an analog of the free energy perturbation technique, and could likely be improved by making use of information obtained from simulations at other  $\lambda$  values. That is, it should be possible to extend the methodology of the MBAR method to obtain  $W$ , thus obtaining a more accurate estimate of that quantity.

**Example** A proof of concept implementation of this technique is illustrated in Fig. 5.6. The data are from the solvation of an uncharged acetamide molecule with 31  $\lambda$  values ( $\lambda_{LJ}$  varies,  $\lambda_C = 0$ ); this is the same system (“ACE 4 250/50”) analyzed in Fig. 4.7. The multicolored lines indicate a contour map of  $W(\lambda, \xi)$ , the red line is  $\langle \xi \rangle_\lambda$  vs.  $\lambda$ , and the black line is  $\langle \lambda \rangle_\xi$  vs.  $\xi$ .

It is clear that there is a region of negative slope in the black  $\langle \lambda \rangle_\xi$  curve, indicating a phase transition. It is hypothesized that this feature corresponds to the creation of an acetamide-sized cavity in neat water and the resulting rearrangement of water at the interface. The free energy associated with this interface can be estimated from this plot by a Maxwell construction, indicated by the horizontal gray dashed line; the two equal areas formed by this construction give the free energy penalty associated with the interface.

## Summary of Phase Transition Calculations

In practice, the procedure to investigate the existence of phase transitions in a simulation where  $\lambda$  parameterizes some aspect of the Hamiltonian proceeds as follows:

1. Perform simulations at various fixed  $\lambda$  values and obtain statistics on  $\xi = \partial F / \partial \lambda$ .
2. Calculate  $F(\lambda)$  using any of the techniques described in chapter 3.
3. Calculate  $W(\lambda, \xi)$ , Eq. (5.16), as a histogram.
4. From  $W$  obtain  $\rho_\lambda$ , Eq. (5.17a).
5. Calculate  $\langle \lambda \rangle_\xi$  as in Eq. (5.20), possibly approximating the integrals as sums using the trapezoidal rule.
6. Plot  $\langle \lambda \rangle_\xi$  versus  $\xi$ , and investigate whether this quantity decreases at any point over the domain. Such a decrease indicates first order phase transitions.
7. Draw a horizontal line (a Maxwell construction), as illustrated in Fig. 5.5, to divide the “backbending” region into two equal areas. This area quantifies the free energy penalty associated with creating the interface.

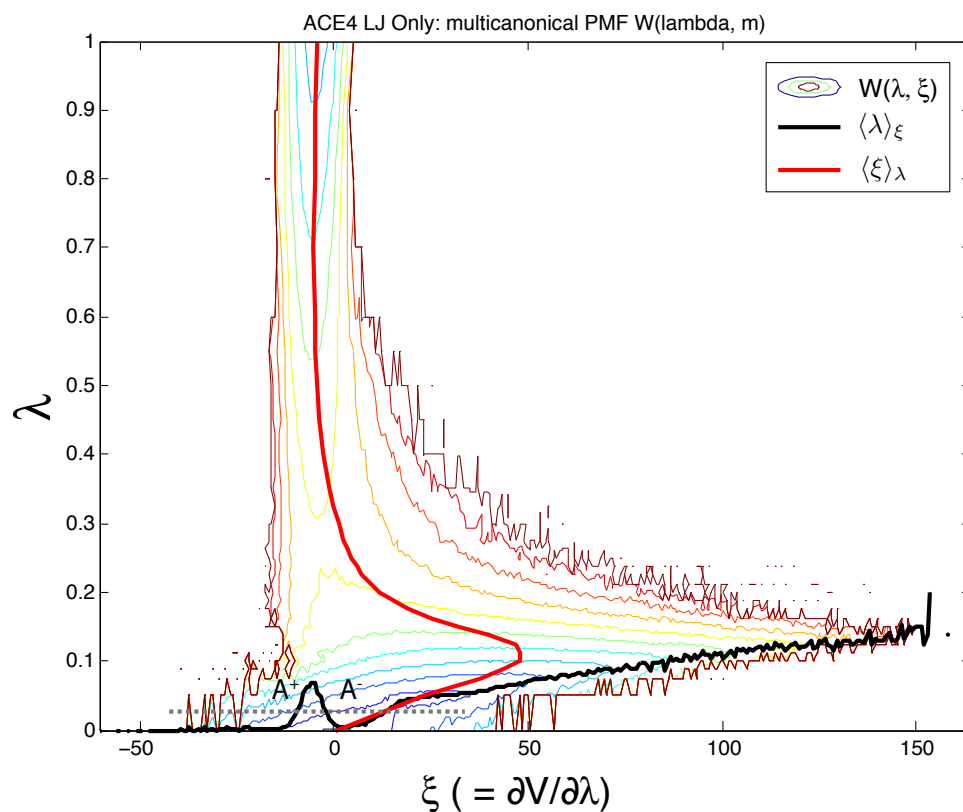


Figure 5.6: An observation of a phase transition in a solvation calculation. Data are from the insertion of uncharged acetamide into solvent, as described chapter 4. The contour map corresponds to  $W(\lambda, \xi)$ , obtained histograms of  $\xi = \partial V / \partial \lambda$  from 31 simulations at fixed  $\lambda$ . From this, the black line  $\langle \lambda \rangle_\xi$  is obtained, and its negative slope in the vicinity of  $\xi = -10$  indicates a phase transition in the system. This phase transition is possibly due to the formation of a cavity, with the interface penalty due to the rearrangement of water at the cavity boundary. The dashed gray line is the Maxwell construction, and defines the equal areas  $A^+$  and  $A^-$ , each of which is equal to the free energy penalty associated with the creation of the interface.



8. The middle crossing of the Maxwell line with the  $\langle\lambda\rangle$  curve indicates  $\lambda^*$ , the critical  $\lambda$  value at which the phase transition occurs (analogous to  $U_{sep}$  in Fig. 5.5).

In this manner phase transitions can be observed generically in molecular simulations performed at fixed  $\lambda$  values. No ad hoc order parameter need be defined, and the thermodynamic limit is not invoked. The method rests on a well established observation that the interface between two phases in a small system induces a significant entropic cost, and it is this penalty which allows us to detect the phase transition. Furthermore, this penalty can be quantified, yielding a measurement of the free energy cost associated with the coexistence of two phases.

In the context of glycine N-mer simulations, such an approach would be able to unambiguously determine whether the coil-to-globule phase transition in explicit solvent, induced by varying the solute-solute dispersions, is a first order phase transition, regardless of any specific order parameters. Provided that it is, the Maxwell construction could then quantify the free energy associated with an interface between the two phases. This quantity,  $\Delta F_{interface}$ , is the increase in the free energy of the system due to the coexistence of both a swollen and collapsed state simultaneously, under equilibrium conditions when both states are equally probable. It would be worthwhile to investigate the nature of this quantity in more detail, and to clarify its relationship with the quantity  $\Delta G_{collapse}$ .

### 5.2.3 Thermodynamic Length

Aside from addressing the motivating topic of polar collapse and techniques to accelerate free energy calculations, the work done in chapter 4 has bearing on contemporaneous issues in nonequilibrium statistical mechanics. One of the major conclusions of chapter 4 is the relationship between the swap probability between adjacent replicas on a  $\lambda$  schedule and the rate of convergence of free energy calculations between the replicas. This relationship can then be used to construct an optimal  $\lambda$  schedule for which the convergence rate for  $\Delta F$  between all replicas is uniform, and which for the whole system is optimal.

Other authors have recently reached similar conclusions using different approaches, drawing connections between such calculations and the geometrical foundations of statistical thermodynamics. Shenfeld et al. (2009) show that a  $\lambda$  schedule with replicas equidistant along a measure called the thermodynamic length will both minimize the variance in the calculated free energy and provide an “almost optimal” replica exchange schedule. Nulton et al. (1985) reached a similar conclusion, finding that a  $\lambda$  schedule which minimizes the dissipation of a thermodynamic process will have steps of equal thermodynamic length.

Thermodynamic length is a measure of a distance between equilibrium thermodynamic states. Unlike the free energy or entropy, which are state functions independent of the path taken, thermodynamic length is a function of that path (Weinhold, 1975; Crooks, 2007). This length can be understood in terms of the number of equilibrium thermal fluctuations the system undergoes along that path; the larger the fluctuations, the closer in thermodynamic space two equilibrium states are (Wootters, 1981). More generally, thermodynamic length provides for a geometric interpretation for statistical thermodynamics (Mrugala et al., 1990; Nulton and Salamon, 1985). In the context of nonequilibrium free energy calculations simulations, a path of the shortest thermodynamic distance will minimize the dissipation for slow, finite time transformations (Feng and Crooks, 2009).

The thermodynamic length for a single controllable parameter  $\lambda$  parameterized by the path variable  $s$  ( $s \in [0, 1]$ ) is given as (Feng and Crooks, 2009),

$$\mathcal{L} = \int_0^1 ds \frac{\partial \lambda}{\partial s} \sqrt{\mathcal{I}(\lambda)}, \quad (5.21)$$

where

$$\mathcal{I} = \beta^2 \text{var} \frac{\partial V}{\partial \lambda}. \quad (5.22)$$

How to actually obtain the thermodynamic length for a given path from a series of simulations at discrete  $\lambda$  values, however, remains an open question. Crooks (2007) obtains a bound on this quantity using an overlap measure derived from the Bennett acceptance ratio technique. It is possible that the Fermi swap probability may yield a better estimate of thermodynamic length, based on two lines of reasoning.

The Fermi swap probability, as discussed in chapter 4 and illustrated in Fig. 4.1(b), quantifies the overlap of two ensembles as an integral over configurations drawn from both:

$$\langle p_{swap} \rangle = \int d\Gamma_0 \int d\Gamma_1 \frac{\rho_0(\Gamma_0)\rho_1(\Gamma_1)\rho_1(\Gamma_0)\rho_0(\Gamma_1)}{\rho_0(\Gamma_0)\rho_1(\Gamma_1) + \rho_1(\Gamma_0)\rho_0(\Gamma_1)} \quad (5.23)$$

The BAR overlap measure, on the other hand, is an integral over a subset of such configurations; from Eq. (3.10), we can obtain the BAR overlap function  $q_{BAR}$  as,

$$q_{BAR} = \int d\Gamma \frac{\rho_0(\Gamma)\rho_1(\Gamma)}{\rho_0(\Gamma) + \rho_1(\Gamma)} \quad (5.24)$$

Graphically, we can consider the Fermi swap probability as the shaded area in 4.1(b), while the BAR overlap measure considers only the dashed diagonal line in that figure. That is, both the BAR and the Fermi swap probabilities consider the overlap of ensembles, but over a different domain; it may be that considering the latter measure may lead to a superior estimate thermodynamic length.

Supporting this notion is the observation that the average swap probability can be written, to second order in  $\delta\lambda$ , as (see Eq. (4.18)),

$$\langle p_{swap} \rangle \simeq \frac{1}{2} - \frac{\delta\lambda^2}{4} \mathcal{I} \quad (5.25)$$

That is, the linearized Fermi swap probability includes the same factor as does the definition of thermodynamic length, Eq. (5.21). This suggests that further analysis of the Fermi swap probability may lead to a superior estimate of the thermodynamic length between two  $\lambda$  values.

A better understanding of thermodynamic length could lead to significant improvements in free energy calculations and their decompositions. The particular functional form with which Hamiltonians are scaled (see section 2.1.3) remain ad hoc with plenty of room for improvement. Many different approaches can be used to construct such scaled potentials (e.g. those described in section 4.9), and there is at present no theoretical foundation upon which to base their improvements.

An approach based on thermodynamic length would offer a way forward. The insertion of a solute into a solvent should formally follow a path which minimizes this

distance. A calculus of variation approach, like the one employed in e.g. Wyczalkowski and Szeri (2003), could be used, together with simulation data, to obtain a functional form of the scaling function which minimizes the thermodynamic length. Insertion processes using this scaling potential should result in insertion paths superior to those used at present.

# References

- J. B. Abrams, L. Rosso, and M. E. Tuckerman. Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics. *J. Chem. Phys.*, 125:074115, 2006.
- M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, New York, 1987.
- L. M. Amzel. Calculation of entropy changes in biological processes: folding, binding, and oligomerization. *Meth Enzymol*, 323:167–77, 2000.
- C. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096): 223–230, 1973.
- H. S. Ashbaugh and M. E. Paulaitis. Effect of solute size and solute-water attractive interactions on hydration water structure around hydrophobic solutes. *J Am Chem Soc*, 123:10721–10728, 2001.
- M. V. Athawale, G. Goel, T. Ghosh, T. M. Truskett, and S. Garde. Effects of lengthscales and attractions on the collapse of hydrophobic polymers in water. *Proc Natl Acad Sci USA*, 104(3):733–8, 2007.
- K. C. Aune and C. Tanford. Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. i. dependence on ph at 25 degrees. *Biochemistry*, 8(11): 4579–85, 1969.
- M. Auton and D. W. Bolen. Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry*, 43(5):1329–42, 2004.
- F. Avbelj, P. Luo, and R. L. Baldwin. Energetics of the interaction between water and the helical peptide group and its role in determining helix propensities. *Proc Natl Acad Sci USA*, 97(20):10786–91, 2000.
- M. Bachmann, H. Arkin, and W. Janke. Multicanonical study of coarse-grained off-lattice models for folding heteropolymers. *Phys. Rev. E*, 71:031906, 2005.
- R. L. Baldwin. Weak interactions in protein folding. In Johannes Buchner and Thomas Kiefhaber, editors, *Protein Folding Handbook, Volume 1*, pages 127–162, Weinheim, 2005. Wiley-VCH Verlag GmbH & Co.

- A. Ben-Naim. *Solvation Thermodynamics*. Plenum Press, New York, 1987.
- A. Ben-Naim and Y. Marcus. Solvation thermodynamics of nonionic solutes. *Journal of Chemical Physics*, 81(4):2016–2027, 1984.
- C. H. Bennett. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.*, 22:245–268, 1976.
- B. J. Berne. Inferring the hydrophobic interaction from the properties of neat water. *P Natl Acad Sci Usa*, 93(17):8800–8803, 1996.
- T. C. Beutler, A. E. Mark, R. C. van Scheik, P. R. Gerber, and W. F. van Gunsteren. Avoiding singularities and numerical instabilities in free energy calculations based on molecular mechanics simulations. *Chem. Phys. Lett.*, 222:529–539, 1994.
- R. Bitetti-Putzer, W. Yang, and M. Karplus. Generalized ensembles serve to improve the convergence of free energy simulations. *Chem. Phys. Lett.*, 377:633–641, 2003.
- D. W. Bolen and G. D. Rose. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annual Review of Biochemistry*, 77:339–362, 2008.
- P. Borrmann, O. Mülken, and J. Harting. Classification of phase transitions in small systems. *Physical Review Letters*, 84(16):3511–3514, 2000.
- S. Cabani, P. Gianni, V. Mollica, and L. Lepori. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous-solution. *J Solution Chem*, 10(8):563–595, 1981.
- H. B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. John Wiley and Sons, New York, second edition, 1985.
- J. Carlsson and J. Aqvist. Absolute and relative entropies from computer simulation with applications to ligand binding. *J Phys Chem B*, 109:6448–6456, 2005.
- D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. The amber biomolecular simulation programs. *J Comput Chem*, 26(16):1668–1688, 2005.
- D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, New York, 1987.
- D. Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437:640–647, 2005.
- J. Chang, A. M. Lenhoff, and S. I. Sandler. Solvation free energy of amino acids and side-chain analogues. *J. Phys. Chem. B*, 111:2098–2106, 2007.

- A. A. Chialvo and J. M. Haile. Determination of excess gibbs free energy from computer simulation: Multiple-parameter charging approach. *Fluid Phase Equilib.*, 37:293–303, 1987.
- C. Chothia. Hydrophobic bonding and accessible surface area in proteins. *Nature*, 248(446):338–9, 1974.
- D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, 437(7056):231–4, 2005.
- B. P. Cossins, S. Foucher, C. M. Edge, and J. W. Essex. Assessment of nonequilibrium free energy methods. *J Phys Chem B*, 113(16):5508–5519, 2009.
- S. L. Crick, M. Jayaraman, C. Frieden, R. Wetzell, and R. V. Pappu. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *P Natl Acad Sci Usa*, 103:16764–16769, 2006.
- G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60(3):2721–2726, 1999.
- G. E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, 61(3):2361–2366, 2000.
- G. E. Crooks. Measuring thermodynamic length. *Physical Review Letters*, 99(10):100602, 2007.
- K. A. Dill. Dominant forces in protein folding. *Biochemistry-U.S.*, 29:7133–7155, 1990.
- K. A. Dill. Additivity principles in biochemistry. *J Biol Chem*, 272(2):701–4, 1997.
- K. A. Dill and S. Bromberg. *Molecular Driving Forces*. Garland Science, New York, 2003.
- A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic. Intrinsically disordered protein. *J Mol Graph Model*, 19(1):26–59, 2001.
- A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, and Z. Obradović. Intrinsic disorder and protein function. *Biochemistry*, 41(21):6573–82, 2002a.
- A. K. Dunker, C. J. Brown, and Z. Obradovic. Identification and functions of usefully disordered proteins. *Advances in Protein Chemistry*, 62:25–49, 2002b.

- S. R. Durell and A. Wallqvist. Atomic-scale analysis of the solvation thermodynamics of hydrophobic hydration. *Biophys J*, 71(4):1695–706, 1996.
- H. J. Dyson and P. E. Wright. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Bio*, 6:197–208, 2005.
- B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, New York, 1993.
- E. H. Feng and G. E. Crooks. Far-from-equilibrium measurements of thermodynamic length. *Physical Review E*, 79(1):012104, 2009.
- A. L. Fink. Natively unfolded proteins. *Curr Opin Struc Biol*, 15(1):35–41, 2005.
- A. V. Finkelstein and E. I. Shakhnovich. Theory of cooperative transitions in protein molecules .2. phase-diagram for a protein molecule in solution. *Biopolymers*, 28(10):1681–1694, 1989.
- S. H. Fleischman and C. L. Brooks III. Thermodynamics of aqueous solvation – solution properties of alcohols and alkanes. *Journal of Chemical Physics*, 87(5):3029–3037, 1987.
- H. Flyvbjerg and H. G. Petersen. Error-estimates on averages of correlated data. *J Chem Phys*, 91:461–466, 1989.
- H. S. Frank and M. W. Evans. Free volume and entropy in condensed systems III. entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *The Journal of Chemical Physics*, 13:507, 1945.
- D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, New York, 2002.
- H. Fukunishi, O. Watanabe, and S. Takada. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.*, 116(20):9058–9067, 2002.
- 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*, 104:6271–6285, 2000.
- J. Gao, D. A. Bosco, E. T. Powers, and J. W. Kelly. Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. *Nat Struct Mol Biol*, 16(7):684–90, 2009.
- S. Garde, G. Hummer, and M. E. Paulaitis. Free energy of hydration of a molecular ionic solute: Tetramethylammonium ion. *Journal of Chemical Physics*, 108(4):1552–1561, 1998.



- S. Gill, S. Dec, G. Olofsson, and I. Wadsoe. Anomalous heat capacity of hydrophobic solvation. *The Journal of Physical Chemistry*, 89:3758–3761, 1985.
- G. Graziano. Hydration thermodynamics of N-methylacetamide. *J. Phys. Soc. Jpn.*, 69(11):3720–3725, 2000.
- A. Y. Grosberg and D. V. Kuznetsov. Quantitative theory of the globule-to-coil transition .1. link density distribution in a globule and its radius of gyration. *Macromolecules*, 25(7):1970–1979, 1992.
- D. H. E. Gross. *Microcanonical Thermodynamics: Phase Transitions in “Small” Systems*. John Wiley and Sons, New Jersey, 2001.
- D. H. E. Gross and J. F. Kenney. The microcanonical thermodynamics of finite systems: The microscopic origin of condensation and phase separations, and the conditions for heat flow from lower to higher temperatures. *J Chem Phys*, 122: 224111, 2005.
- A. Grossfield, P. Ren, and J. W. Ponder. Ion solvation thermodynamics from simulation with a polarizable force field. *J Am Chem Soc*, 125(50):15671–82, 2003.
- A. M. Hahn and H. Then. Characteristic of Bennett’s acceptance ratio method. *Physical Review E*, 80(3):031111, 2009.
- J. K. Hale and H. Koçak. *Dynamics and Bifurcations*. Springer-Verlag, New York, 1996.
- I. Halperin, B. Ma, H. Wolfson, and R. Nussinov. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, 47(4):409–43, 2002.
- U. H. E. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chem Phys Lett*, 281:140–150, 1997.
- D. A. Hendrix and C. Jarzynski. A “fast growth” method of computing free energy differences. *Journal of Chemical Physics*, 114(14):5974–5981, 2001.
- B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.
- M. Heying and D. S. Corti. Scaled particle theory revisited: New conditions and improved predictions of the properties of the hard sphere fluid. *J Phys Chem B*, 108:19756–19768, 2004.
- T. L. Hill. *Thermodynamics of Small Systems*. Dover, New York, 1994.

- L. Hnedkovsky, S. Degrange, and I. Cibulka. Partial molar volumes of organic solutes in water. I. O-, m-, and p-cresol at temperatures 298K to 573K. *J Chem Thermodyn*, 30(5):557–569, 1998.
- D. Horinek, S. I. Mamatkulov, and R. R. Netz. Rational design of ion force fields based on thermodynamic solvation properties. *The Journal of Chemical Physics*, 130(12):124507, 2009.
- K. Hukushima and K. Nemoto. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.*, 65(6):1604–1608, 1996.
- G. Hummer, S. Garde, A. E. Garcia, M. E. Paulaitis, and L. R. Pratt. Hydrophobic effects on a molecular scale. *J Phys Chem B*, 102:10469–10482, 1998.
- W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- P. Hunenberger. Thermostat algorithms for molecular dynamics simulations. *Adv Polym Sci*, 173:105–147, 2005.
- J. E. Hunter, W. P. Reinhardt, and T. F. Davis. A finite-time variational method for determining optimal paths and obtaining bounds on free-energy changes from computer-simulations. *J Chem Phys*, 99:6856–6864, 1993.
- V. A. Ivanov, W. Paul, and K. Binder. Finite chain length effects on the coil-globule transition of stiff-chain macromolecules: A Monte Carlo simulation. *Journal of Chemical Physics*, 109(13):5659–5669, 1998.
- W. Janke. Canonical versus microcanonical analysis of first-order phase transitions. *Nucl Phys B-Proc Sup*, 63:631–633, 1998.
- C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78(14):2690–2693, 1997.
- C. Jarzynski. Rare events and the convergence of exponentially averaged work values. *Phys. Rev. E*, 73:046105, 2006.
- G. Jeffrey. Hydrogen-bonding: an update. *Crystallography Reviews*, 9(2-3):135–176, 2003.
- W. Jiang, M. Hodoscek, and B. Roux. Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics. *Journal of Chemical Theory and Computation*, 5(10):2583–2588, 2009.
- S. Jones and J. M. Thornton. Principles of protein-protein interactions. *P Natl Acad Sci Usa*, 93(1):13–20, 1996.

- W. Jorgensen, J. Chandrasekhar, J. Madura, R.W. Impey, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.
- W. L. Jorgensen and C. Jenson. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *J Comput Chem*, 19(10):1179–1186, 1998.
- W.L. Jorgensen, D.S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.
- C. Junghans, M. Bachmann, and W. Janke. Microcanonical analyses of peptide aggregation processes. *Physical Review Letters*, 97:218103, 2006.
- C. Junghans, M. Bachmann, and W. Janke. Thermodynamics of peptide aggregation processes: An analysis from perspectives of three statistical ensembles. *J. Chem. Phys.*, 128(8):085103, 2008.
- G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B*, 105(28):6474–6487, 2001.
- J. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3(5):300–313, 1935.
- X. Kong and C. L. Brooks III.  $\lambda$ -dynamics: A new approach to free energy calculations. *J. Chem. Phys.*, pages 2414–2423, 1996.
- E. Kreyszig. *Advanced Engineering Mathematics*. John Wiley & Sons, New York, seventh edition, 1993.
- M. M. Kubo, E. Gallicchio, and R. M. Levy. Thermodynamic decomposition of hydration free energies by computer simulation: Application to amines, oxides and sulfides. *J. Phys. Chem. B*, 101:10527–10534, 1997.
- S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules .1. The method. *Journal of Computational Chemistry*, 13:1011–1021, 1992.
- D. J. Kuster. *A Unifying Model for Helices in Proteins*. PhD thesis, Washington University in St. Louis, 2009.
- Y. A. Kuznetsov and E. G. Timoshenko. On the conformational structure of a stiff homopolymer. *Journal of Chemical Physics*, 111(8):3744–3752, 1999.

- A. Laio and M. Parrinello. Escaping free-energy minima. *P Natl Acad Sci Usa*, 99(20):12562–12566, 2002.
- D. P. Landau and K. Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, Cambridge, 2000.
- A. R. Leach. *Molecular Modelling: principles and applications*. Prentice Hall, New York, second edition, 2001.
- B. Lee. Analyzing solvent reorganization and hydrophobicity. *Method Enzymol*, 259:555–576, 1995.
- B. Lee and F. M. Richards. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*, 55(3):379–400, 1971.
- K. H. Lee, D. Xie, E. Freire, and L. M. Amzel. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins*, 20(1):68–84, 1994.
- R. M. Levy and E. Gallicchio. Computer simulations with explicit solvent: Recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annu. Rev. Phys. Chem.*, 49:531–567, 1998.
- N. Lu and D. A. Kofke. Optimal intermediates in staged free energy calculations. *J. Chem. Phys.*, 111(10):4414–4423, 1999.
- N. Lu and D. A. Kofke. Accuracy of free-energy perturbation calculations in molecular simulation. I. modeling. *J. Chem. Phys.*, 114(17):7303–7311, 2001a.
- N. Lu, D. A. Kofke, and T. B. Woolf. Staging is more important than perturbation method for computation of enthalpy and entropy changes in complex systems. *J Phys Chem B*, 107(23):5598–5611, 2003a.
- N. Lu, J. K. Singh, and D. A. Kofke. Appropriate methods to combine forward and reverse free-energy perturbation averages. *J. Chem. Phys.*, 118(7):2977–2984, 2003b.
- N. D. Lu and D. A. Kofke. Accuracy of free-energy perturbation calculations in molecular simulation. I. modeling. *Journal of Chemical Physics*, 114(17):7303–7311, 2001b.
- K. Lum, D. Chandler, and J. D. Weeks. Hydrophobicity at small and large length scales. *J Phys Chem B*, 103:4570–4577, 1999.
- J. L. MacCallum and D. P. Tieleman. Calculation of the water-cyclohexane transfer free energies of neutral maino acid side-chain analogs using the OPLS all-atom force field. *J. Comput. Chem.*, 24(15):1930–1935, 2003.

- A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*, 102(18): 3586–3616, 1998.
- M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J Chem Phys*, 112:8910–8922, 2000.
- G. I. Makhatadze. Thermal unfolding of proteins studied by calorimetry. In Johannes Buchner and Thomas Kiefhaber, editors, *Protein Folding Handbook, Volume 1*, pages 70–98, Weinheim, 2005. Wiley-VCH Verlag GmbH & Co.
- G. I. Makhatadze, M. M. Lopez, and P. L. Privalov. Heat capacities of protein functional groups. *Biophysical Chemistry*, 64(1-3):93–101, 1997.
- M. G. Martin and J. I. Siepmann. Novel configurational-bias Monte Carlo method for branched molecules. transferable potentials for phase equilibria. 2. united-atom description of branched alkanes. *J. Phys. Chem. B*, 103:4508–4517, 1999.
- N. Metropolis, A. Rosenbluth, M. Rosenbluth, and A. H. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1953.
- M. Mezei and D. L. Beveridge. Free energy simulations. *Annals of the New York Academy of Sciences*, 482(1):1–23, 1986.
- L. Mirny and E. Shakhnovich. Protein folding theory: From lattice to all-atom models. *Annu Rev Bioph Biom*, 30:361–396, 2001.
- D. L. Mobley, E. Dumont, J. D. Chodera, and K. A. Dill. Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B*, 111:2242–2254, 2007.
- A. Möglich, K. Joder, and T. Kiefhaber. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc Natl Acad Sci USA*, 103(33):12394–9, 2006.
- J. C. Moore, R. Battino, T. R. Rettich, Y. P. Handa, and E. Wilhelm. Partial molar volumes of gases at infinite dilution in water at 298.15K. *J Chem Eng Data*, 27(1): 22–24, 1982.
- R. Mrugala, J. D. Nulton, J. C. Schon, and P. Salamon. Statistical approach to the geometric structure of thermodynamics. *Physical Review A*, 41(6):3156–3160, 1990.

- S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, and A. A. Deniz. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *P Natl Acad Sci Usa*, 104(8):2649–54, 2007.
- M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Clarendon Press, Oxford, UK, 1999.
- J. Nulton, P. Salamon, B. Andresen, and Q. Anmin. Quasistatic processes as step equilibrations. *Journal of Chemical Physics*, 83(1):334–338, 1985.
- J. D. Nulton and P. Salamon. Geometry of the ideal-gas. *Physical Review A*, 31(4):2520–2524, 1985.
- J. Nummela, F. Yassin, and I. Andricioaei. Entropy-energy decomposition from nonequilibrium work trajectories. *The Journal of Chemical Physics*, 128(2):024104, 2008.
- H. Nymeyer, S. Gnanakaran, and A. E. García. Atomic simulations of protein folding, using the replica exchange algorithm. *Methods Enzymol.*, 383:119–149, 2004.
- H. Oberhofer and C. Dellago. Efficient extraction of free energy profiles from nonequilibrium experiments. *J Comput Chem*, 30(11):1726–1736, 2009.
- H. Oberhofer, C. Dellago, and P. L. Geissler. Biased sampling of nonequilibrium trajectories: can fast switching simulations outperform conventional free energy calculation methods? *The Journal of Physical Chemistry B*, 109(14):6902–15, 2005.
- L. Onsager. Electric moments of molecules in liquids. *J Am Chem Soc*, 58:1486–1493, 1936.
- C. N. Pace. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Meth Enzymol*, 131:266–80, 1986.
- C. N. Pace. Energetics of protein hydrogen bonds. *Nat Struct Mol Biol*, 16(7):681–2, 2009.
- C. N. Pace, B. A. Shirley, M. McNutt, and K. Gajiwala. Forces contributing to the conformational stability of proteins. *FASEB J*, 10(1):75–83, 1996.
- V. S. Pande, A. Y. Grosberg, T. Tanaka, and D. S. Rokhsar. Pathways for protein folding: is a new view needed? *Curr Opin Struc Biol*, 8(1):68–79, 1998.
- L. Pauling. *General Chemistry*. W.H. Freeman Press, San Francisco, CA, 1970.
- L. Pauling and R. B. Corey. The pleated sheet, a new layer configuration of polypeptide chains. *P Natl Acad Sci Usa*, 37(5):251–6, 1951.

- L. Pauling, R. B. Corey, and H. R. Branson. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *P Natl Acad Sci Usa*, 37(4):205–11, 1951.
- D. A. Pearlman and P. A. Kollman. The lag between the Hamiltonian and the system configuration in free-energy perturbation calculations. *Journal of Chemical Physics*, 91(12):7831–7839, 1989.
- C. Peter, C. Oostenbrink, A. van Dorp, and W. F. van Gunsteren. Estimating entropies from molecular dynamics simulations. *The Journal of Chemical Physics*, 120(6):2652–61, 2004.
- J. W. Pitera and W. F. van Gunsteren. A comparison of non-bonded scaling approaches for free energy calculations. *Mol. Simulat.*, 28:45–65, 2002.
- M. Pleimling and H. Behringer. Microcanonical analysis of small systems. *Phase Transit*, 78:787–797, 2005.
- J. M. Polson and N. E. Moore. Simulation study of the coil-globule transition of a polymer in solvent. *Journal of Chemical Physics*, 122(2):024905, 2005.
- J. W. Ponder and D. A. Case. Force fields for protein simulations. *Adv Protein Chem*, 66:27–85, 2003.
- J. P. M. Postma, H. J. C. Berendsen, and J. R. Haak. Thermodynamics of cavity formation in water – a molecular-dynamics study. *Faraday Symp Chem S*, 17:55–67, 1982.
- L. R. Pratt and D. Chandler. Theory of the hydrophobic effect. *Journal of Chemical Physics*, 67:3683, 1977.
- C. Predescu, M. Predescu, and C. V. Ciobanu. On the efficiency of exchange in parallel tempering Monte Carlo simulations. *J. Phys. Chem. B*, 109:4189–4196, 2005.
- A. Proykova and R. S. Berry. Insights into phase transitions from phase changes of clusters. *J Phys B-At Mol Opt*, 39:R167–R202, 2006.
- H. Qian and J. J. Hopfield. Entropy-enthalpy compensation: Perturbation and relaxation in thermodynamic systems. *J. Chem. Phys.*, 105(20):9292–9298, 1996.
- H. Reiss, H. L. Frisch, and J. L. Lebowitz. Statistical mechanics of rigid spheres. *The Journal of Chemical Physics*, 31(2):369, 1959.
- F. Ritort. Work fluctuations, transient violations of the second law and free-energy recovery methods: Perspectives in theory and experiment. In J. Dalibard, editor, *Poincaré Seminar 2003: Bose-Einstein Condensation-Entropy*, pages 195–229, Basel, 2003. Birkhäuser Verlag.

- M. A. Roseman. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J Mol Biol*, 200(3):513–22, 1988.
- B. Roux. Calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, 91:275–282, 1995.
- B. Roux and T. Simonson. Implicit solvent models. *Biophys Chem*, 78:1–20, 1999.
- M. Rubinstein and R. H. Colby. *Polymer Physics*. Oxford University Press, New York, 2003.
- E. Shakhnovich. Protein folding thermodynamics and dynamics: Where physics, chemistry, and biology meet. *Chemical Reviews*, 106(5):1559–1588, 2006.
- E. I. Shakhnovich and A. V. Finkelstein. Theory of cooperative transitions in protein molecules .1. why denaturation of globular protein is a 1st-order phase-transition. *Biopolymers*, 28(10):1667–1680, 1989.
- K. A. Sharp and B. Madan. Hydrophobic effect, water structure, and heat capacity changes. *J. Phys. Chem. B*, 101:4343–4348, 1997.
- K. A. Sharp, A. Nicholls, R. Friedman, and B. Honig. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry*, 30(40):9686–97, 1991.
- D. P. Sheehan and D. H. E. Gross. Extensivity and the thermodynamic limit: Why size really does matter. *Physica A*, 370(2):461–482, 2006.
- D. K. Shenfeld, H. Xu, M. P. Eastwood, R. O. Dror, and D. E. Shaw. Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations. *Physical Review E*, 80(4):046705, 2009.
- M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *Journal of Chemical Physics*, 129(12):124105, 2008.
- M. R. Shirts and V. S. Pande. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.*, 122:134508, 2005a.
- M. R. Shirts and V. S. Pande. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.*, 122:144107, 2005b.
- M. R. Shirts, J. W. Pitner, William C. Swope, and V. S. Pande. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.*, 119(11):5740–5761, 2003.



- K. L. Sim and T. P. Creamer. Abundance and distributions of eukaryote protein simple sequences. *Mol Cell Proteomics*, 1(12):983–95, 2002.
- D. Sindhikara, Y. Meng, and A. E. Roitberg. Exchange frequency in replica exchange molecular dynamics. *Journal of Chemical Physics*, 128(2):024103, 2008.
- R. D. Skeel and J. A. Izaguirre. An impulse integrator for Langevin dynamics. *Mol Phys*, 100(24):3885–3891, 2002.
- D. E. Smith and A. D. J. Haymet. Free energy, entropy and internal energy of hydrophobic interactions: Computer simulations. *J. Chem. Phys.*, 98(8):6445–6454, April 1993.
- D. E. Smith, L. Zhang, and A. D. J. Haymet. Entropy of association of methane in water – a new molecular-dynamics computer-simulation. *J Am Chem Soc*, 114:5875–5876, 1992.
- M. Souaille and B. Roux. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput Phys Commun*, 135:40–57, 2001.
- D. F. Stickle, L. G. Presta, K. A. Dill, and G. D. Rose. Hydrogen-bonding in globular-proteins. *J Mol Biol*, 226(4):1143–1159, 1992.
- F. H. Stillinger. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *Journal of Solution Chemistry*, 2(2-3):141–158, 1973.
- T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma. Free-energy of hydrophobic hydration – a molecular-dynamics study of noble-gases in water. *Journal of Chemical Physics*, 85(11):6720–6727, 1986.
- Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314:141–151, 1999.
- Y. Sugita, A. Kitao, and Y. Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J Chem Phys*, 113:6042–6051, 2000.
- S. X. Sun. Equilibrium free energies from path sampling of nonequilibrium trajectories. *Journal of Chemical Physics*, 118(13):5769–5775, 2003.
- C. Tanford. *The Hydrophobic Effect*. John Wiley & Sons, New York, 1973.
- O. Tcherkasskaya and V. N. Uversky. Denatured collapsed states in protein folding: example of apomyoglobin. *Proteins*, 44(3):244–54, 2001.
- P. R. ten Wolde. Hydrophobic interactions: an overview. *J Phys-Condens Mat*, 14:9445–9460, 2002.

- P. R. ten Wolde and D. Chandler. Drying-induced hydrophobic polymer collapse. *P Natl Acad Sci Usa*, 99(10):6539–6543, 2002.
- G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187 – 199, 1977.
- H. Touchette, R. Ellis, and B. Turkington. An introduction to the thermodynamic and macrostate levels of nonequivalent ensembles. *Physica A: Statistical Mechanics and its Applications*, 340(1-3):138–146, 2004.
- H. T. Tran, X. L. Wang, and R. V. Pappu. Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. *Biochemistry*, 44(34):11369–11380, 2005.
- H. T. Tran, A. Mao, and R. V. Pappu. Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J Am Chem Soc*, 130(23):7380–7392, 2008.
- D. Trzesniak and W. F. van Gunsteren. Pathway dependence of the efficiency of calculating free energy and entropy of solute-solute association in water. *Chem Phys*, 330(3):410–416, 2006.
- M. Udier-Blagović, P. M. De Tirado, S. A. Pearlman, and W. L. Jorgensen. Accuracy of free energies of hydration using CM1 and CM3 atomic charges. *J. Comput. Chem.*, 25(11):1322–1332, 2004.
- V. N. Uversky. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*, 11(4):739–56, 2002.
- V. N. Uversky, J. R. Gillespie, and A. L. Fink. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins*, 41(3):415–27, 2000.
- A. Vitalis and R. V. Pappu. Absinth: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *J Comput Chem*, 30(5):673–699, 2009a.
- A. Vitalis and R. V. Pappu. Methods for Monte Carlo simulations of biomacromolecules. *Annual Reports in Computational Chemistry*, 5:49–76, 2009b.
- A. Vitalis, A. Steffen, N. Lyle, A. Mao, and R. V. Pappu. Campari: A software suite for modeling and computational analysis of macromolecular properties across resolutions and interfaces. *In Preparation*, 2009.
- J. A. Wagoner and N. A. Baker. Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. *P Natl Acad Sci Usa*, 103: 8331–8336, 2006.

- D. J. Wales and R. S. Berry. Coexistence in finite systems. *Phys. Rev. Lett.*, 73: 2875–2878, 1994.
- S. Wan, R. H. Stote, and M. Karplus. Calculation of the aqueous solvation energy and entropy, as well as free energy, of simple polar solutes. *J. Chem. Phys.*, 121: 9539–9548, 2004.
- E. A. Weathers, M. E. Paulaitis, T. B. Woolf, and J. H. Hoh. Insights into protein structure and function from disorder-complexity space. *Proteins*, 66(1):16–28, 2007.
- F. Weinhold. Metric geometry of equilibrium thermodynamics. *The Journal of Chemical Physics*, 63(6):2479–2483, 1975.
- E. Wilhelm, R. Battino, and R. Wilcock. Low-pressure solubility of gases in liquid water. *Chemical Reviews*, 77:219, 1977.
- A. P. Willard and D. Chandler. The role of solvent fluctuations in hydrophobic assembly. *The Journal of Physical Chemistry B*, 112(19):6187–92, 2008.
- R. Wolfenden. Interaction of the peptide bond with solvent water: A vapor phase analysis. *Biochemistry*, 17(1):201–204, 1978.
- R. Wolfenden, L. Andersson, P. M. Cullis, and C. C. B. Southgate. Affinities of amino-acid side-chains for solvent water. *Biochemistry*, 20(4):849–855, 1981.
- Wolfram Research. *Mathematica 7.0*. Wolfram Research Inc., Champaign, Illinois, 2008.
- C. J. Woods, J. W. Essex, and M. A. King. The development of replica-exchange-based free-energy methods. *J. Phys. Chem. B*, 107:13703–13710, 2003a.
- C. J. Woods, J. W. Essex, and M. A. King. Enhanced configurational sampling in binding free-energy calculations. *J Phys Chem B*, 107:13711–13718, 2003b.
- W. K. Wootters. Statistical distance and Hilbert space. *Phys. Rev. D*, 23(2):357–362, 1981.
- P. E. Wright and H. J. Dyson. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J Mol Biol*, 293(2):321–331, 1999.
- D Wu and DA Kofke. Asymmetric bias in free-energy perturbation measurements using two hamiltonian-based models. *Physical Review E*, 70(6):066702, 2004.
- M. A. Wyczalkowski and R. V. Pappu. Satisfying the fluctuation theorem in free-energy calculations with Hamiltonian replica exchange. *Phys. Rev. E*, 77(2):11, 2008.

- M. A. Wyczalkowski and A. J. Szeri. Optimization of acoustic scattering from dual-frequency driven microbubbles at the difference frequency. *J Acoust Soc Am*, 113(6):3073–3079, 2003.
- F. M. Ytreberg, R. H. Swendsen, and D. M. Zuckerman. Comparison of free energy methods for molecular systems. *J. Chem. Phys.*, 125:184114, 2006.
- D. M. Zuckerman and T. B. Woolf. Systematic finite-sampling inaccuracy in free energy differences and other nonlinear quantities. *J. Stat. Phys.*, 114(5-6):1303–1323, 2004.
- R. W. Zwanzig. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.*, 22(8):1420–1426, 1954.

# Curriculum Vitae

Matthew A. Wyczalkowski  
4106 Wyoming St.  
St. Louis, MO 63116  
matt@wyczalkowski.com

- Born** November 30, 1973  
Wrocław, Poland
- Degrees** B.S. Engineering Science and Mechanics, August 1996  
*Pennsylvania State University, State College, PA USA*
- M.S. Mechanical Engineering, May 2000  
*University of California, Berkeley, CA USA*
- Ph.D. Biomedical Engineering, December 2009  
*Washington University, Saint Louis, MO USA*
- Employment** Programmer Analyst, 9/2002 - 9/2004  
*Mallinckrodt Institute of Radiology,  
Washington University School of Medicine, St. Louis, MO*
- Test Engineer, 12/2000 - 8/2002  
*Onix Microsystems, Richmond, CA*
- Co-founder and programmer, 3/2000 - 12/2000  
*Maze Technologies, San Francisco, CA*
- Research Assistant, Summers 1993-1995, 6/1996 – 8/1997  
*Dept. of Radiology, Hershey Medical Center, Hershey, PA*
- Professional Societies** Biomedical Engineering Society  
Biophysical Society

- Select Honors and Awards** University Scholars Program, Pennsylvania State University, 1992-1996
- Peer-reviewed publications**
- Fuentes, A., M. A. Wyczalkowski and K. C. MacKinnon. Niche construction through cooperation: A nonlinear dynamics contribution to modeling resilience and evolutionary history in the genus Homo. *Current Anthropology*, in press.
- Wyczalkowski, M. A. and R. V. Pappu. Satisfying the fluctuation theorem in free-energy calculations with Hamiltonian replica exchange. *Physical Review E* 2008; 77; 026104.
- Wang, X., A. Vitalis, M. A. Wyczalkowski and R. V. Pappu. Characterizing the conformational ensemble of monomeric polyglutamine. *Proteins: Structure, Function, and Bioinformatics* 2006; 63; 297-311.
- Wyczalkowski, M. A., and A. J. Szeri. Optimization of acoustic scattering from dual-frequency driven microbubbles at the difference frequency. *Journal of the Acoustical Society of America* 2003; 113:6; 3073-3079.
- Wyczalkowski, M. A. Adequacy of the linearized local electric field determination in a cubically nonlinear composite. *Optik* 1997; 105:1; 41-44.
- Manuscripts in preparation** Wyczalkowski, M. A., A. Vitalis and R. V. Pappu. Acceptance ratio methods for solvation entropy and enthalpy calculations.

December 2009

**Computational Solvation Thermodynamics, Wyczalkowski, Ph.D. 2009**