University of Southampton

The Development of Free Energy Methods for Protein-Ligand Complexes

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Southampton.

Christopher J. WOODS
Department of Chemistry
University of Southampton

September 2003

Supervisor : Dr. J. W. Essex
Adviser : Prof. G. A. Attard
Industrial Supervisor : Dr. M. A. King

Abstract

UNIVERSITY OF SOUTHAMPTON
FACULTY OF SCIENCE
CHEMISTRY

Doctor of Philosophy

THE DEVELOPMENT OF FREE ENERGY METHODS FOR PROTEIN-LIGAND COMPLEXES

by Christopher Woods

Methods for the calculation of the relative binding free energies of ligands to a protein are investigated and developed. The aim of these investigations was to improve the reliability and speed of free energy methods, such that they become practical tools for commercial rational drug design. To this end, the relative hydration free energy of water and methane, and the relative binding free energies of halides to a calix[4]pyrrole derivative were investigated by three established free energy methods (Free Energy Perturbation (FEP), Finite Difference Thermodynamic Integration (FDTI) and Adaptive Umbrella WHAM (AdUmWHAM)). The results of these applications showed that inconsistencies in sampling led to unreliable free energy predictions. To overcome these problems, a series of four new free energy methods were developed (Bivariate Multicanonical WHAM (BMW), Parallel Tempering Thermodynamic Integration (PTTI), Replica Exchange Free Energy Perturbation (REFEP) and Replica Exchange Thermodynamic Integration (RETI)). These approaches all combined traditional free energy methods with generalised ensembles. Testing of these methods revealed that Replica Exchange Thermodynamic Integration was the superior of all seven methods. FDTI and RETI were then tested by calculating the relative binding free energies of a group of SB1-like ligands to p38 MAP kinase. The results of this test showed that RETI was still the superior method. This test also revealed that there were still sampling issues that needed to be resolved. A new Monte Carlo code was developed to run the tests on p38. The optimised data structure of the code led to a ten to twelve fold speed up compared to an established MC code. This, combined with the use of a large Linux Beowulf cluster, enabled each protein-ligand free energy calculation to be run within 1.5 days. We predict that, using the latest computers, these calculations could take less than 12 hours.

Dedicated to Erica.

I spoke about wings. You just flew.

Acknowledgements

I would like to thank my supervisor, Jon, for the freedom and support I enjoyed throughout this research, which enabled many ideas to be played out. I thank Mike for his enthusiasm and understanding, and the group at Celltech, Slough for the three enjoyable months I spent with them. I have had the good fortune of working with a large, helpful, and above all fun group of people. The unique mix of talents and personalities within the Essex group has led to many great discussions, and lots of opportunity for bouncing around ideas. I would especially like to thank Steve and Rob, who have kept the computers running for the last three and a bit years, and Oz, who keeps Iridis going! Rob deserves extra thanks for writing and supporting the 'histplot' program which I have used for all of the colour histograms in this thesis.

My family are important to me, and I have drawn strength from their support and encouragement. It is their faith in me that provides the confidence and belief necessary to keep pursuing my ambitions, and never settling for anything but my best.

Work such as this cannot be performed without adequate software. Today, we are fortunate enough to be experiencing a revolution in the way software is developed and distributed. This Open Source, of Free Software movement means that enterprise class software is available for *anyone* to use, develop, modify and learn from. This is all free from financial cost or intellectual restrictions. This research almost exclusively used open source software, and, with the exception of the use of ChemDraw, this thesis has been written entirely using open source tools. I have been able to write tens of thousands of lines of code, run on hundreds of computers, automate billions of calculations, all because of open source software. I, and the field of computational chemistry, owe a huge debt to the writers of Linux, GCC, glibc, g77, Perl (Huge Thanks!), Python, KDE, LaTeX, PyMol, Xmgrace, and countless other programs that we all now take for granted. We should all take action to protect this freedom from the threat of software patents (see http://petition.eurolinux.org), and all do our bit to repay this debt through being generous with code and scientific ideas that we develop ourselves. For my part, once it is complete, I intend to release the Monte Carlo code developed during this work under an open source license.

I would like to thank the University of Southampton for providing the computing facilities made available to me. Through the use of the Iridis, a large Linux Beowulf cluster, nearly 74 billion(!) Monte Carlo steps of simulation were performed, nearly 11 billion of these on the p38 protein-ligand system. I would also like to thank the BBSRC and Celltech R&D for providing the funding for this work. Their generous contributions make exploratory science possible.

Finally, I would like to thank my wife, Erica. Her love, support, presence and inspiration gives me the strength to keep doing this, even when things get tough. She shares in my joy when I enthuse about my latest success, and listens to my ranting when things go wrong. Erica, you are everything to me. I love you.

Contents

Chapter 1

Introduction

1.1 Aims

The aim of this work is to allow the reliable and accurate calculation of the relative binding affinities of a range of ligands to a protein. For these calculations to become generally useful and accepted, they must be able to proceed with knowledge of only the crystal structures of the bound complexes, and be performed under the time and user constraints that apply within the pharmaceutical industry. That is to say that the calculations should take days rather than weeks or months to complete, and should require as little user intervention as possible.

1.2 Rational Drug Design

Drug design today is far from the random "hit and miss" of the early twentieth century. Today, effort may be applied to understand how and why particular ligands bind to proteins, and by using this knowledge, to design further molecules that can bind more strongly. This process is generally called *rational drug design.*[1] The process can be represented as a sequence of steps;

1. The structure of the target protein is obtained.

2. The binding mode of a potential ligand in the protein must be determined. This could either be via an experimental structure of the complex, or via computational docking[2] of the ligand into the active site.

3. The affinity of the ligand for the protein is then estimated. This can either be via a simple scoring function, or via simulations aimed at calculating the binding constant.

4. The information provided by the simulations is used to propose modifications to the ligand. These modified ligands are also docked and scored against the target protein.

5. The most promising ligands are synthesised, and assays performed to experimentally determine the binding constants.[3] The information from these experiments is then fed back into the simulations.

6. More simulations are run. These should be of a higher quality, as they have access to more experimental data on the ligands (e.g. the protein-ligand crystal structures). They may also be run at a higher level of theory.

7. The information from the simulations is then fed back into step 4. This cycle is repeated until a sufficiently potent ligand is developed.

A key part of this process is the accurate and reliable prediction of the binding affinity of the potential ligands to the target protein. In particular, this process requires the dependable prediction of the effect of ligand modifications on the binding affinity. The development of methods which can do this is a primary goal of this field of study, and the achievement of this goal, in the absence of extensive experimental data, has so far eluded current workers. This is despite the potential rewards that a solution to the *binding problem* could bring to the pharmaceutical industry. The achievment of this goal requires the development of methods that can accurately and reliably predict the binding affinities of a range of ligands to a single protein. While a 'correct' solution to this problem lies in the equations of statistical thermodynamics, the difficulties involved in solving these equations have left many to pursue alternative, less rigorous routes. Rigour and reliability are a primary goal of this work, so current methods that are correctly derived from the underlying statistical thermodynamics will be investigated. The strengths and weaknesses of these methods will be assessed on a range of challenging test cases, and new methods will be developed.

1.3 Roadmap

First, this thesis will provide a background to molecular modeling, and will explore how biomolecular systems may be represented in the computer. Statistical thermodynamics will be introduced as a method to explore these models and estimate real thermodynamic properties. The most important thermodynamic property examined will be *free energy*, and rigorous procedures to calculate relative free energies of binding will be discussed. The huge expense of free energy simulations has led some workers to abandon theoretical rigour, and pursue less expensive, yet less exact methods. Some of these methods will be discussed.

The aim of this thesis is to identify the problems that exist in current, rigorous free energy simulations. Through identification, ideas to overcome any problems or inefficiencies may be developed. The first experimental section will introduce a conceptually simple, yet physically challenging test case, the relative hydration free energy of water and methane. Three of the established free energy methods will be examined in the context of this test case, and their strengths and weaknesses will be highlighted. The second experimental section will take the lessons learned, and use them to develop four further free energy methods. These too will be evaluated on the water-methane system, and their results compared. The third experimental section will introduce a simple binding free energy test case, the binding of halide ions to a calix[4]pyrrole derivative. All of the established and new methods will be applied to this system, and their efficiency and reliability compared. The final experimental section will then take the best established, and best new free energy method, and will use them to investigate a real protein-ligand system, namely the binding of SB1-like ligands to p38 MAP kinase. The results from these applications will then be used as an indicator of the success of this research.

1.4 Conclusion

One of the great challenges of rational drug design is the solution of the binding problem, namely the reliable and efficient calculation of the relative affinities of a range of potential ligands to a protein. A solution to this problem lies in the

equations of statistical thermodynamics, so this work will evaluate established free energy methods, and will develop and test new free energy methods. These methods will be tested on three challenging systems, the relative hydration free energy of water and methane, the binding of halide ions to a calix[4]pyrrole derivative, and the binding of ligands to p38 kinase.

Chapter 2

Biomolecular Simulations

2.1 Introduction

Computer simulations rely on good models for the systems to which they are being applied. Such models must be sufficiently detailed to represent accurately the problem being investigated, yet not overly elaborate, lest they are too expensive for a computer to calculate. The best models in this field rely on *Quantum Mechanics* (QM) to predict charge densities and intermolecular interactions.[4] The solution of the QM equations for large biomolecular systems is not trivial. While techniques such as divide and conquer[4] may allow a single large system to be divided into several connected small systems, these methods are not sufficiently fast to allow the full configurational exploration that was necessary for the work presented in this thesis. Instead, simplified models, based on point charges and springs were used. Such models are typically referred to as *Molecular Mechanics* (MM) models. A *forcefield* is the complete set of molecular mechanics terms needed to model a system. The complexity of such terms depends on the system under investigation, and what properties the forcefield is trying to predict.

2.2 Molecular Mechanics Forcefields

Molecular mechanics forcefields provide simple models of a system which computers can calculate in a reasonable time. There are many forcefields that may

be used to represent biomolecules, for example OPLS,[5,6] AMBER,[7] MM3[8] and CHARMM22.[9] The OPLS forcefield, which evolved from AMBER, was specifically parameterised to reproduce condensed phase properties, and as such, was chosen for this work.

The OPLS forcefield models the protein-ligand system as a collection of atoms with charges, held together with springs. Partial atomic charges, q, are placed on each atom, and the intermolecular interaction energy, $E^{coul}(r)$, between charges q_i and q_j on different molecules, separated by distance r, is calculated via the Coulomb equation,[6]

$$E^{coul}(r) = \frac{q_i q_j}{4\pi\varepsilon_0 r}, \tag{2.1}$$

where ε_0 is the permittivity of free space.

In addition to an intermolecular coulombic energy, a term is also needed to represent the intermolecular electron dispersion forces. In reality, these interactions depend on the positions of many atoms. However, such *many body* interactions are difficult to evaluate, as they involve the sum over all triplets of atoms, quadruplets of atoms etc. To avoid this expense, the dispersion forces are approximated by *effective pair potentials*. One such approximation is the *Lennard-Jones* (LJ) potential,

$$E^{LJ}(r) = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right). \tag{2.2}$$

This equation returns the pair potential, $E^{LJ}(r)$, for a homonuclear pair of atoms separated by distance r. The parameters, σ and ε are determined empirically. This determination implicitly includes the effects of many-body interactions. The OPLS forcefield uses the LJ equation to calculate the intermolecular dispersion energy,[6]

$$E^{LJ}(r) = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right) \tag{2.3}$$

$$\text{where } \sigma_{ij} = \sqrt{(\sigma_i \sigma_j)} \text{ and } \varepsilon_{ij} = \sqrt{(\varepsilon_i \varepsilon_j)}.$$

Since the LJ function was parameterised on homonuclear pairs, combining rules must be used to obtain the values of σ and ε for heteronuclear pairs. In the case of OPLS, σ_{ij} and ε_{ij} are obtained by the geometric average shown. Other combining

rules are used in other forcefields, e.g. AMBER uses the arithmetic mean to obtain σ_{ij}.[7] The total intermolecular energy between molecules is thus given by the sum over all pairs of atoms between molecules, of the Coulomb and Lennard-Jones energies.

The OPLS forcefield models the intramolecular interactions via *bond*, *angle*, *dihedral* and *non-bonded* terms.[6] All the bonds between atoms are determined, and all of the directly bonded (1-2), directly angled (1-3) and dihedral (1-4) sets of atoms are identified. The interactions between bonded, 1-2 pairs, are then simply modelled as springs, using a harmonic potential, around the equilibrium bond length, r_0. The strength of the bond is assigned from a parameter, k_{bond}, and the bond energy, $E^{bond}(r)$, as a function of bond-length, r, is given by,[6]

$$E^{bond}(r) = k_{bond}(r - r_0)^2. \tag{2.4}$$

The angles, which act between 1-3 atoms, are treated in a similar fashion, with a harmonic potential acting around the equilibrium angle, θ_0. The strength of the angle is taken from a parameter, k_{angle}, and the angle energy, $E^{angle}(\theta)$, as a function of angle, θ, is given by,[6]

$$E^{angle}(\theta) = k_{angle}(\theta - \theta_0)^2. \tag{2.5}$$

The dihedral angles, which act over 1-4 atoms, are slightly more complicated, and their energy is evaluated via three terms. The first term is a Fourier series over cosine functions. This is used to represent the broad shape of the dihedral energy surface. The next two terms allow for angle bending in the eclipsed conformation, and are represented by the Coulomb and Lennard-Jones interaction energies between the 1 and 4 atoms in the dihedral. These last two energies are normally scaled by a factor, s_{14}, which typically is equal to two, though some implementations use different values for the Coulomb and Lennard-Jones terms. The parameters used in this evaluation are the amplitudes, v_0 to v_3, and phases, p_1 to p_3, of the cosines, and the charges and LJ parameters of the atoms. The total dihedral energy, $E^{dihedral}(r, \phi)$, for 1-4 atoms separated by distance, r, and dihedral angle,

ϕ, is given by,[5,10]

$$E^{dihedral}(r,\phi) = v_0 + \frac{1}{2}\sum_{m=1}^{3} v_m\Big(1+(-1)^{m+1}\cos(m\phi+p_m)\Big) + \frac{1}{s_{14}}\Big(E^{coul}(r)+E^{LJ}(r)\Big). \tag{2.6}$$

The remaining intramolecular interactions, between 1-5 and above pairs of atoms, are modelled identically to the intermolecular interactions. This *non-bonded* term in the OPLS forcefield, $E^{nb}(r)$, between a pair of intramolecular atoms, separated by distance r, is given by,

$$E^{nb}(r) = E^{coul}(r) + E^{LJ}(r). \tag{2.7}$$

The total energy of the system is taken as the sum over all inter- and intramolecular terms. As it stands, this forcefield may still become too expensive for use on biomolecules, so the intermolecular terms are normally truncated such that interactions between atoms separated by more than a cut-off distance are ignored.[11,12] This cutoff may be applied between pairs of atoms, or it may be based on the distance between pairs of groups, e.g. if the closest distance between two residues of molecules is greater than the cutoff distance, then all of the pair-pair interactions between the two groups are ignored.[12] This truncation of the non-bonded terms can lead to discontinuities in the potential energies and forces associated with the interaction.[12] To overcome this problem, the non-bonded terms, $E_{NB}(r)$, may be scaled by multiplying by a *switching function*, $S(r)$,[12]

$$E^{nb'}(r) = S(r) \times E^{nb}(r), \tag{2.8}$$

where r is the distance between atoms. The aim of the switching function is to preserve the nature of the non-bonded interaction at low r, while gradually smoothing the energy to zero by the cutoff distance. This switching function may be applied over the entire range of distances, or only for a short range of distances before the

cutoff,[12]

$$E^{nb'}(r) = E^{nb}(r) \text{ for } r < r_{feather}$$

$$E^{nb'}(r) = S(r) \times E^{nb}(r) \text{ for } r_{feather} \leq r \leq r_{cut} \quad (2.9)$$

$$E^{nb'}(r) = 0 \text{ for } r > r_{cut},$$

where r_{cut} is the cutoff distance, and $r_{feather}$ is the distance beyond which the switching function feathers the non-bonded interaction down to zero. The OPLS forcefield, as implemented in MCPRO,[13] uses an harmonic switching function that quadratically scales down the non-bonded interaction energy to zero between $r_{feather}$ and r_{cut}. When using group-based cutoffs, it is important to ensure that the switching function has the same value for each pair of atoms between the interacting groups.[12] This may be achieved by calculating a single value of the switching function for the interacting groups, and multiplying it by the total non-bonded interaction energy between the groups.

This forcefield does not explicitly take polarisation effects into account. Real physical systems polarise substantially when placed in a high-dielectric medium, such as water.[14] Such polarisation could affect the energies and structures adopted by the system.[14] Fixed charge forcefields, such as OPLS, implicitly include the effects of polarisation by increasing the partial charges such that molecular or fragment dipoles are approximately 10-20% larger than those in the gas phase.[14] Forcefields have been developed that explicitly include terms to account for polarisation effects.[14] Inducible dipole forcefields place point dipoles over the system to represent the polarisable groups.[14,15] In these models, the size of the dipole moment is determined by minimising the total interaction energy between all of the dipoles, and all of the fixed charges.[15] This represents a 'self-consistent field' determination of dipole moments,[15] and is performed for every configuration of the system. Induced charge models take a different approach. In these methods, the partial charges themselves are allowed to fluctuate in response to their environment.[16] Point dipoles are still placed on polarisable groups, and the magnitude of each induced dipole is calculated. This dipole can then be explicitly represented by a set of point charges.[16] These induced charges are added to the fixed charges

on the atom sites. This allows the partial charges on each atom to fluctuate in response to their environment. The benefit of the using a point charge representation of the induced dipole is that it avoids the costly calculation of the dipole-dipole and charge-dipole interaction energies.[16]

2.3 Statistical Thermodynamics

Given a model, or forcefield, it then becomes possible to calculate the total energy of the system for every single possible configuration of that system. The collection of all possible configurations and momenta for a system is called *phase space*, and it has a dimension for every single degree of freedom of the system.[17] Plotting the total energy for each point within phase space yields a hypersurface, called the *energy surface*. Such surfaces for proteins are known to be rugged and frustrated,[18] filled with many, near-isoenergetic minima, separated by mid-level transition states. As the system vibrates and moves, it will tend to spend more time in certain regions of phase space than others. To calculate the probability that the system will be at a particular point in phase space, we need to use statistical thermodynamics, and in particular, the Boltzmann distribution.

2.3.1 The Boltzmann Distribution

The Boltzmann distribution lies at the heart of statistical thermodynamics. It was initially derived for a closed system of identical, independent atoms, which could adopt a range of discrete energy levels (see appendix A). The atoms were allowed to exchange energy through elastic collisions, and the entire system was subjected to the constraints that the total energy, and total number of atoms, were a constant. The derivation uses statistical theory to show that, for a given temperature, T, there will be one overwhelmingly probable configuration of this system. Assuming that the system adopts this configuration, it becomes possible to calculate the probability, p_i, that a particular energy level, ε_i, is occupied, via,

$$p_i = \frac{\exp(-\varepsilon_i / k_B T)}{q}, \tag{2.10}$$

where k_B is the Boltzmann constant, and q is the molecular partition function, given by the sum over all of the accessible energy levels,

$$q = \sum_j \exp(-\varepsilon_j / k_B T). \tag{2.11}$$

The molecular Boltzmann distribution was derived to operate on a collection of identical particles, which could only share energy in a specific manner, under the constraints of constant total energy and numbers of particles. To apply this to an entire system of molecules, all interacting through complex inter- and intramolecular potentials, requires the introduction of the concept known as the *ensemble*. An ensemble is a collection of identical *replicas* of the system. Each system is allowed to evolve independently, though they are allowed to share defined properties, subject to specific constraints. The properties shared, and constraints applied, determine the type of the ensemble. For example, in the NVT, or *canonical ensemble*, the total number of atoms (N), volume (V) and temperature (T) are constrained to be constant. This means that each of the replicas are constrained to a constant volume, number of particles and temperature. The temperature is constant as the replicas are placed in thermal contact with one another, and are allowed to exchange energy.[19] Each of the replicas are allowed to equilibrate under these conditions. At equilibrium, the probabilities that a replica will adopt particular configurations will become constant. At this steady-state, for temperature T, the probability, $p_T(i)$, that a replica adopts a particular configuration, i, with associated energy, E_i, is given by the canonical Boltzmann equation,[12]

$$p_T(i) = \frac{exp(-E_i/k_B T)}{Q_{NVT}}, \tag{2.12}$$

where Q_{NVT} is the canonical partition function. Many ensembles exist, e.g. the *isothermal-isobaric ensemble* (constant Number of particles, Pressure and Temperature, NPT), or the *microcanonical ensemble* (constant Number of particles, Volume and Energy, NVE). Each ensemble has its own partition function, the importance of which will become apparent in the next section.

2.3.2 Free Energy

The thermodynamic quantity known as "free energy" is perhaps one of the most important of all the thermodynamic quantities. This is because free energy points in the direction of spontaneous change. The binding free energy is the change in free energy associated with the binding of a guest to a host, and is a direct measure of the strength of that binding. The more negative the binding free energy, the stronger the binding between the host and guest. The free energy of a system, G can be related directly to its partition function,[11]

$$G = -k_B T \ln Q. \tag{2.13}$$

If the partition function was calculated over the canonical ensemble, then G represents the Helmholtz free energy, while if the partition function was calculated over the NPT ensemble, then G represents the Gibbs free energy. This is because the partition function for any ensemble, Q_{ens}, can be related to the thermodynamic potential for the ensemble, Ψ_{ens},[11]

$$\Psi_{ens} = -\ln Q_{ens}. \tag{2.14}$$

The thermodynamic potential for an ensemble has a minimum value at thermodynamic equilibrium.[11] For the canonical ensemble, the thermodynamic function is $G_{Helmholtz}/k_B T$, while for the NPT ensemble the function is $G_{Gibbs}/k_B T$.

Equation 2.13 shows that the calculation of the absolute free energy of a system requires the calculation of its partition function. The canonical partition function, Q, is formed as a sum over all possible configurations, Γ, of the system, of the exponential of the energy of each configuration, $E(\Gamma)$, normalised by the temperature, T,

$$Q = \sum_{\Gamma} \exp\left(\frac{-E(\Gamma)}{k_B T}\right). \tag{2.15}$$

At the classical limit, the partition function for an atomic system can be expressed as an integral[11] over all possible configurations,

$$Q = \frac{1}{N!}\frac{1}{h^{3N}} \int \exp\left(\frac{-E(\Gamma)}{k_B T}\right) d\Gamma. \tag{2.16}$$

In this equation, N is the number of atoms, h is Planck's constant, introduced to define the volume of the system, and the factor of $\frac{1}{N!}$ is used to account for the indistinguishability of particles.[11]

The total energy of the system is formed as a sum of the potential energy, E_p, and the kinetic energy, E_k. The potential energy is a function of the coordinates, q, of the system, while the kinetic energy is a function of the momenta, p, and thus the partition function may be expressed as an integral over all coordinates and momenta,

$$Q = \frac{1}{N!}\frac{1}{h^{3N}}\int_q\int_p \exp\left(\frac{-E_p(q) - E_k(p)}{k_B T}\right) dp\,dq. \tag{2.17}$$

The coordinates and momenta of the system are independent, and thus the kinetic and potential energies of the system may be separated. This allows the partition function to be expressed as a product of kinetic and potential parts,[11]

$$\begin{aligned} Q &= \frac{1}{N!}\frac{1}{h^{3N}}\int_p\int_q \exp\left(\frac{-E_p(q) - E_k(p)}{k_B T}\right) dp\,dq \\ &= \frac{1}{N!}\frac{1}{h^{3N}}\int_p\int_q \exp\left(\frac{-E_p(q)}{k_B T}\right)\exp\left(\frac{-E_k(p)}{k_B T}\right) dp\,dq \\ &= \frac{1}{N!}\frac{1}{h^{3N}}\int_p \exp\left(\frac{-E_k(p)}{k_B T}\right) dp \int_q \exp\left(\frac{-E_p(q)}{k_B T}\right) dq \\ &= Q_k \cdot Q_p, \end{aligned} \tag{2.18}$$

where the factors, $\frac{1}{N!}$ is absorbed into Q_k, and $\frac{1}{h^{3N}}$ is absorbed into Q_p.

The integral over momenta, known as the *ideal gas* part, can be solved analytically via quantum mechanics using the "particle in the box" model,[11,17]

$$Q_k = \frac{V^N}{N!\Lambda^{3N}} \text{ where } \Lambda = (h^2/2\pi m k_B T)^{\frac{1}{2}}, \tag{2.19}$$

where V is the volume of the system, and m is the mass of each atom.

The remaining potential energy partition function (the *excess part*) cannot be solved analytically, due to the large number of coupled inter- and intramolecular energy terms that would need to be evaluated. The evaluation of this integral over all possible configurations of the system must be performed numerically. In doing

this, many workers drop the factor of $\frac{1}{h^{3N}}$, and instead work with the *configuration integral*, Z,

$$Z = \int_q \exp\left(\frac{-E_p(q)}{k_B T}\right) dq. \tag{2.20}$$

where Z has units of V^N.[11]

Unfortunately, the numerical integration of the configuration integral converges very slowly for all but the most simple of systems.[11] The absolute free energy of a system depends directly on the configurational integral, and is thus too difficult to calculate for protein sized systems.

2.3.3 Relative Free Energies

It is too difficult to calculate the absolute free energy of large systems. However, it is possible to calculate the relative free energy of two different systems. This was first realised in the derivation by Zwanzig,[20]

$$\begin{aligned}
\Delta G_{A\to B} &= G_B - G_A \\
&= (-k_B T \ln Q_B) - (-k_B T \ln Q_A) \\
&= -k_B T \ln\left[\frac{Q_B}{Q_A}\right] \\
&= -k_B T \ln\left[\frac{\int \exp(-E_B(q)/k_B T)\, dq}{\int \exp(-E_A(q)/k_B T)\, dq}\right]
\end{aligned}$$

multiply by $1 = \exp(-E_A(q)/k_B T)\exp(E_A(q)/k_B T)$ gives,

$$\begin{aligned}
&= -k_B T \ln\left[\frac{\int \exp(-E_B(q)/k_B T) \times \exp(-E_A(q)/k_B T)\exp(E_A(q)/k_B T)\, dq}{\int \exp(-E_A(q)/k_B T)\, dq}\right] \\
&= -k_B T \ln\left[\frac{\int \exp(-E_A(q)/k_B T) \times \exp\left(-(E_B(q) - E_A(q)/k_B T\right) dq}{\int \exp(-E_A(q)/k_B T)\, dq}\right] \\
&= -k_B T \ln\left[\int \frac{\exp(-E_A(q)/k_B T)}{Q_A} \times \exp(-\Delta E_{AB}(q)/k_B T)\, dq\right] \\
&= -k_B T \ln\left[\int p_A(q) \times \exp(-\Delta E_{AB}(q)/k_B T)\, dq\right],
\end{aligned} \tag{2.21}$$

where $P_A(q)$ is the Boltzmann probability of configuration q in the ensemble of state A, and ΔE_{AB} is the difference in energy between system A and system B.

This equation shows that the relative free energy is the integral over all configurations of system A, of the Boltzmann weighted exponential of the difference in energy between the two states. The vast majority of the configurations of system A will be high in energy, and will have a very low Boltzmann probability. Because of this, only the low energy, thermally significant configurations of A should be needed to obtain the relative free energy of the two states. The question now arises as to how these highly probable configurations are generated.

2.4 Sampling Methods

The biomolecular system has been modelled using a molecular mechanics forcefield. From this forcefield, a potential energy surface can be generated. The Boltzmann equation can be used to calculate the probability of various points on the energy surface, thus giving information about which configurations a single system would prefer. However, while we can only look at a single system of molecules (10 K to 100 K atoms) on a computer, in the "real World" we experience unimaginably huge numbers of these systems. To bridge the macroscopic and molecular worlds, statistical thermodynamics assumes that the thermodynamically observable property of a system, A_{obs}, is equal to the time average of the corresponding molecular property, A, calculated from a single system, providing it has evolved for a sufficiently long time,[11]

$$A_{obs} = \langle A(\Gamma(t)) \rangle_{time} = \lim_{t_{obs} \to \infty} \frac{1}{t_{obs}} \int_0^{t_{obs}} A(\Gamma(t)) dt, \tag{2.22}$$

where $A(\Gamma(t))$ is the value of A calculated for configuration Γ at time t, and $\langle ... \rangle_{time}$ represents an average over time.

If the time trajectory of the system can visit every point in phase space which has a non-zero Boltzmann probability, then the trajectory is termed *ergodic*. If this were the case, then we could imagine a large ensemble of systems, all following this time trajectory. Each system would be at a different point along the trajectory.

While all of the systems are dynamically moving along the trajectory, at thermal equilibrium, the density of systems at each region of phase space will become constant. The probability of finding a system in a particular configuration along the trajectory will thus be a constant, and, from the work in the previous sections, will be equal to the Boltzmann probability for that configuration in the ensemble. The *ergodic hypothesis* thus proposes that the time average of a property for a single system is equal to the Boltzmann weighted property calculated over a whole ensemble of systems,[11]

$$\begin{aligned} A_{obs} &= \langle A(\Gamma(t)) \rangle_{time} \\ &= \langle A(\Gamma) p(\Gamma) \rangle \\ &= \langle A \rangle_{ens}, \end{aligned} \tag{2.23}$$

where $p(\Gamma)$ is the Boltzmann probability for configuration Γ in the ensemble, and $\langle ... \rangle_{ens}$ represents the average of A calculated over the whole ensemble of configurations. The ergodic hypothesis thus proposes that the observable thermodynamic properties of the macroscopic world are equal to the ensemble average of the corresponding molecular property. Sampling methods exist which can generate ensembles of structures with the correct Boltzmann probabilities. Two such methods are Molecular Dynamics[12] and Metropolis Monte Carlo.[21]

2.4.1 Molecular Dynamics

Molecular Dynamics[11,12] is a method to locate and sample the significant regions of the energy surface. It takes the most obvious route available to generate the ensemble of configurations, namely that of evolving the time trajectory. The system is placed in a starting configuration, at a particular point on the energy surface. The gradient at that point on the surface can be evaluated. The negative of this gradient is a force, which can be converted into an acceleration via Newton's laws of motion. The direct solution of Newton's laws requires the solution of 3N coupled, second order differential equations, where N is the number of atoms in the system. This is too difficult to solve analytically, so numerical, finite difference approaches must be used.[12] These numerically integrate the forces over time to yield a trajec-

tory. Since Newton's laws conserve energy, the resulting trajectories sample from the NVE, microcanonical ensemble. To sample at constant temperature, the system must be connected to a thermostat. In addition, a piston may be used to make the trajectory sample at constant pressure. Assuming that the trajectory has evolved for sufficiently long, the ergodic hypothesis[11] states that the time trajectory will have formed the complete and correct ensemble.

2.4.2 Monte Carlo

Molecular Dynamics samples the energy surface by taking one system, and evolving from it, a trajectory over time. Monte Carlo takes the opposing view, and randomly generates many configurations of the system, and then weights each one according to its probability within the desired ensemble.[11] It does this by forming a *Markov chain*, by making random changes of configuration, and accepting or rejecting these changes via a test. The form of the test depends on how the Monte Carlo equations are solved, and many such solutions exist. Appendix B details some MC tests, and how they are obtained from the underlying Markov chain. The most important solution to these equations is the Metropolis solution.[21] Indeed, this solution is so important, that the many workers equate Monte Carlo with Metropolis Monte Carlo.

2.4.3 Metropolis Monte Carlo

Normal Monte Carlo randomly generates configurations of the system, and weights each one according to its probability within the ensemble. However, these probabilities are rarely known *a priori*. Metropolis Monte Carlo[21] cleverly solves this by randomly generating each configuration such that it appears with its correct ensemble probability, and then weighting each configuration in the set equally. The Metropolis solution can achieve this, as it includes the Boltzmann equation (equation 2.12, and see appendix B). Metropolis Monte Carlo takes a configuration of the system, and calculates its energy. A random move is made, and the energy recalculated.[21] A Monte Carlo test uses this change in energy, ΔE, to evaluate the

move, via,[21]

$$\exp(-\Delta E/k_BT) \geq rand(0,1). \tag{2.24}$$

If this test is passed, then the new configuration is accepted, otherwise the old configuration is restored, and is recounted in the average. Using this method, at the end of the MC run, the thermally important regions of the energy surface have been explored with the correct probability. This particular Monte Carlo test will sample from the canonical, NVT ensemble. Different tests exist which sample from different ensembles, e.g. isothermal-isobaric, NPT. Some of these tests are discussed in appendix B.

2.5 Calculating Relative Free Energies

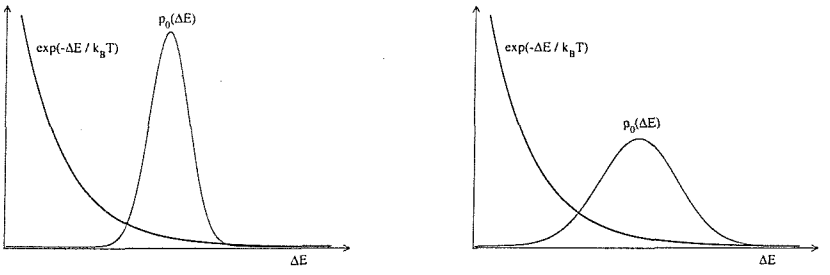
In equation 2.21, the relative free energy between two different states was seen to be,

$$\Delta G_{A\rightarrow B} = -k_BT \ln\left[\int_q p_A(q) \times \exp(-\Delta E_{AB}(q)/k_BT)\, dq\right]. \tag{2.25}$$

This integral represents the ensemble average of the exponential of the difference in energy between states A and B, formed over the ensemble of states of system A. The equation can be rewritten as,[20]

$$\Delta G_{A\rightarrow B} = -k_BT \ln\langle \exp(-\Delta E_{AB}(q)/k_BT)\rangle_A, \tag{2.26}$$

where $\langle ... \rangle_A$ represents an average taken over the ensemble of structures of system A. This is known as the Zwanzig equation,[20] and it shows that the relative free energies between two systems can be calculated over an MD or MC sampled trajectory. An MC or MD simulation is performed on system A. This is called the *reference state*. At each step of the simulation, the difference in energy between the configuration in system A, and the same configuration in system B is calculated, and averaged according to the Zwanzig equation. System B is called the *perturbed state*. The difference in free energy between the two systems will only converge well if they are very similar, and thus the fluctuations of ΔE are small. The reason for this is made clear by recasting the Zwanzig equation in terms of an integral



(a) Low fluctuations in ΔE (b) Large fluctuations in ΔE

Figure 2.1: Plots of $\exp(-\Delta E/k_BT)$ and $p_0(\Delta E)$ for two hypothetical free energy simulations. Plot (a) shows the probability density two systems with good overlap, while plot (b) shows the density for two systems which have poor overlap.

over ΔE,[11]

$$\begin{aligned}\Delta G_{A\to B} &= -k_BT \ln\left[\int_q p_A(q) \times \exp(-\Delta E(q)/k_BT)\, dq\right] \\ &= -k_BT \ln\left[\int_{-\infty}^{\infty} p_0(\Delta E) \times \exp(-\Delta E/k_BT)\, d(\Delta E)\right],\end{aligned} \tag{2.27}$$

where $p_0(\Delta E)$ is the normalised number of configurations of system A that have a difference in energy with system B that is equal to ΔE, i.e. $E_B(q) - E_A(q) = \Delta E$.[11] p_0 includes the ensemble probabilities of system A, and also contains the factors necessary to go from the integral over the $3N$ spatial coordinates to the integral over a single ΔE coordinate.[11]

The Zwanzig equation represents an integral over all ΔE, of the product of the probability of configurations which have that value of ΔE, and $\exp(-\Delta E/k_BT)$. These two terms are shown in figure 2.1. This figure shows these terms for two different free energy simulations. In the first simulation, the two systems are very similar. The fluctuations in the difference in energy between the systems, ΔE, are small, and thus the probability density along ΔE is tightly peaked, with small tails. In the second simulation, there are larger differences between the systems, and thus the fluctuations in ΔE are larger. The probability density is more spread out, with a wider peak, and longer tails. The MC or MD simulation will generate values of ΔE according to the probability density $p_0(\Delta E)$. If the simulation was run for

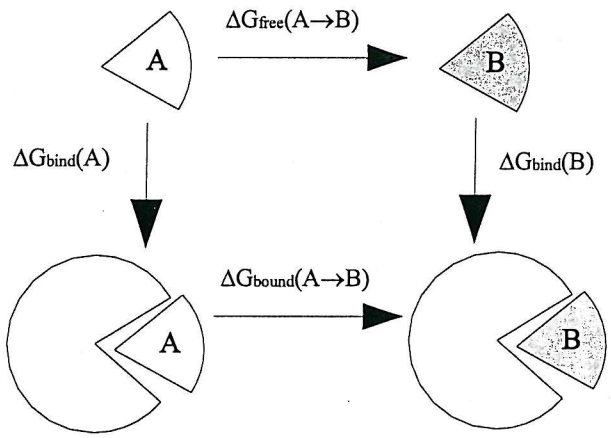
an infinite time, then the histogram of values of ΔE collected from the simulation would be exactly equal to the underlying probability distribution. However, for a finite time simulation, the histogram of values of ΔE will only approximate the true underlying probability density.[22] The errors in this approximation, per unit ΔE, will be larger for the tails than they would be for the peak of the distribution. Since this tail is more significant when there are larger fluctuations in ΔE, then the error associated with the numerical sampling of this tail will also be more significant.[22] These errors are magnified by the exponential function that, in the product of the two terms, greatly increases the contribution from the left tail of the probability distribution. This suggests that, for a finite time simulation, the fluctuations in ΔE should be as small as possible to allow the Zwanzig equation to converge in a reliable and rapid manner.[22] These fluctuations can be minimised by ensuring that the two systems are as similar as possible, i.e. there is high similarity and overlap of their potential energy surfaces. If there is poor overlap, then this analysis suggests that the Zwanzig equation could require an unattainably long simulation to sample adequately the tails of the probability distribution.

2.5.1 The λ-coordinate

The Zwanzig equation relates the difference in free energy between two systems, with differences between their respective energies. This equation will only work well when there is a high degree of overlap between the two systems. To ensure that this is the case, it is possible to use the technique of *morphing* to gradually change one system into the other. A λ-coordinate is used to gradually morph the forcefield such that at $\lambda = 0.0$, the forcefield represents system A, and at $\lambda = 1.0$ it represents system B. At λ-values in-between, the system is a non-physical hybrid of A and B. This morphing may be achieved by linearly changing the OPLS parameters between those for $\lambda = 0.0$, and those for $\lambda = 1.0$, e.g.[13]

$$\begin{aligned} k_{bond_\lambda} &= \lambda k_{bond_1} + (1-\lambda) k_{bond_0} \\ \text{or } \; q_\lambda &= \lambda q_1 + (1-\lambda) q_0. \end{aligned} \tag{2.28}$$

The use of λ allows two related systems to be slowly morphed from one to the



$$\Delta G_{bind}(B) - \Delta G_{bind}(A) = \Delta G_{bound}(A \rightarrow B) - \Delta G_{free}(A \rightarrow B) \qquad (2.29)$$

Figure 2.2: The free energy cycle used to calculate relative binding free energies of ligands *A* and *B* to a protein (shown in grey). The relative binding free energy of ligands *A* and *B* to a protein is equal to the perturbation free energy of *A* to *B* while bound to the protein, minus the perturbation free energy of the free ligands in solvent.

other. This allows the relative free energy between the two systems to be calculated as a potential of mean force along the λ-coordinate that connects them.

2.5.2 Relative Binding Free Energies

The relative binding free energy between two ligands for the same protein can be calculated by morphing the first ligand into the second. From the free energy cycle[23] in figure 2.2 it is seen that this perturbation must be performed while the ligands are bound to the protein (the *bound leg*), and while the ligands are free in solvent (the *free leg*). This is due to the fact that any host-guest binding can be viewed from the perspective of a competition between the host and solvent for the guest. Thus when we ask the question of which of a pair of ligands binds best to a protein, we are really asking which ligand has the greater affinity for the protein, and the lower affinity for the solvent. However the calculation of the free energy changes associated with these legs can be expensive, and there are many ways in which it can be achieved. These methods broadly fall into two groups; rigorous techniques that rigidly stick to derivations from statistical thermodynamics, and

non-rigorous methods that sacrifice rigour in return for assumptions that speed up the convergence of the calculations. This work is only interested in rigorous free energy methods, although both types will now be discussed.

2.6 Rigorous Free Energy Methods

Rigorous free energy methods adhere to the derivations from statistical thermodynamics. That is to say that the methods use a λ-coordinate to connect different systems, and use exact equations to calculate the change in free energy with respect to λ across the entire perturbation. The methods used to calculate the change in free energy with respect to λ are however different. Three such methods will be discussed, *Free Energy Perturbation* (FEP), *Thermodynamic Integration* (TI), and *Adaptive Umbrella WHAM* (AdUmWHAM). The methods discussed here are also described in various reviews.[23,24] In addition, the recently developed, *Fast Growth* method will also be discussed.[25,26]

2.6.1 Free Energy Perturbation

Free Energy Perturbation[20] (FEP) is a rigorous free energy method that has been used to calculate binding free energies in many successful studies, i.e. calculating the specificities of various ligands for the COX-1 or COX-2 enzymes,[27] investigating the enantioselective binding of peptide based ligands to a small host,[28] the binding of ligands to SH2,[29,30] or FK506 binding protein,[31] and the binding of alkali metal cations to spherands.[32] FEP calculates the free energy change along the λ-coordinate through direct use of the Zwanzig equation (equation 2.26). To ensure that this equation converges, the λ-coordinate is split into a series of windows (figure 2.3). The width between each window must be sufficiently small to ensure good overlap between the reference and perturbed states.[12] An MC or MD simulation is then run within each window, and the Zwanzig equation applied to calculate the free energy change between each window and its neighbour. The relative free energy along the λ-coordinate can then be achieved by summing each of the individual free energy differences between each window (figure 2.3). The free

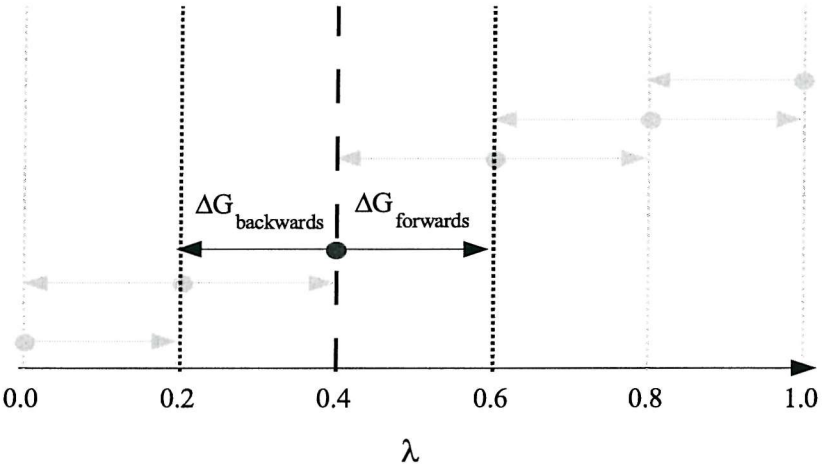


Figure 2.3: The use of windows in FEP simulations. In this figure, the λ-coordinate is split up into 6 windows, spaced evenly every 0.2 λ-units. A simulation (represented by a filled circle) is run in each window. The $\lambda = 0.4$ window is highlighted. The forwards free energy to the next window, and backwards free energy to the previous window are calculated during the simulation. The sum of all the forwards free energies yields the forwards estimate of the relative free energy, while the negative of the sum of all of the backwards free energies yields the backwards estimate. Obviously, both estimates should be the same.

energy differences could be calculated between each window and the next neighbour. Summing these values would yield the free energy change for the forwards perturbation from state *A* to state *B*. Conversely, the free energy differences could be calculated between each window and its previous neighbour. Summing these would yield the free energy change for the backwards perturbation from state *B* to state *A*. If the calculation has converged, then the forwards and backwards free energies should be equal. Any difference between them is known as *hysteresis*, and examination of where hysteresis occurs can be used to position better the λ-windows for any subsequent calculations. To enable this analysis, FEP simulations typically calculate both the forwards and backwards perturbations, using a technique known as *double-wide sampling*.[33] Using this technique, a free energy simulation is performed at each λ-window, and the free energy difference simultaneously estimated between the next and previous window.

FEP accumulates the exponential of the *difference* in energy between neighbouring windows. If the λ-windows are well positioned, then these differences in energy will be well-behaved, and their fluctuations will be small. However, the initial positioning of the windows is difficult, as there is no *a priori* knowledge

of the shape of the *potential of mean force* (PMF) along λ. More windows are needed in regions where the PMF changes rapidly, while fewer are needed in regions where the PMF is flat. The technique of dynamically modified windows[34] tries to alleviate this problem by using the free energy change in one window to estimate the optimum width of the next window. It achieves this by fitting a straight line through previously calculated free energies to estimate the current gradient of the PMF. The value of this gradient can then be used to decide where to place the next λ-value. A problem with this method is that it was developed at a time when computer resources required that each λ-window were run serially, and thus the gradient information from the previous λ-windows was readily available. However, the advent of cheap, yet powerful *Beowulf clusters* (see appendix D) means that all λ-windows can now be run simultaneously in parallel, and thus the spacing between all λ-windows must be determined before the simulation starts.

2.6.2 Thermodynamic Integration

Thermodynamic Integration (TI) is another rigorous free energy method with a significant history of successful applications, e.g. the calculation of the relative binding free energies of ligands to p38,[35] the estrogen receptor ligand-binding domain[36] or acetylcholinesterase,[37] the relative hydration free energy of n-alkanes,[38] the binding of ions to a calix[4]pyrrole derivative,[39] and investigating the interactions between amino acid residues in the binding site of trypsin.[40] While FEP directly uses the Zwanzig equation to calculate the difference in free energy along the λ-coordinate, TI takes a different approach. The method still looks at discrete λ-values along the coordinate, and generates an MC or MD trajectory at each λ-value. However, instead of calculating the difference in energy between neighbouring λ-values, it calculates the rate of change of free energy, with respect to λ at each point. TI thus avoids the problems of low overlap experienced in FEP, as this free energy gradient, $(\frac{\partial G}{\partial \lambda})_\lambda$, is a property of the system at each value of λ only. Once all of these free energy gradients are obtained, they may be integrated

to yield the relative free energy along the λ-coordinate.

$$G_{\lambda=1} - G_{\lambda=0} = \int_0^1 \left(\frac{\partial G}{\partial \lambda}\right)_\lambda d\lambda \tag{2.30}$$

This integral can be evaluated numerically, e.g. via the trapezium rule.[35] The free energy gradients themselves may be obtained analytically or numerically. The analytical route uses a modified forcefield to calculate the gradient of each forcefield term directly with respect to λ. The ensemble average of the gradient of forcefield, $\langle \frac{\partial E}{\partial \lambda} \rangle_\lambda$, is equal to the free energy gradient (as shown in appendix C.1).

$$\int_0^1 \left(\frac{\partial G}{\partial \lambda}\right)_\lambda d\lambda = \int_0^1 \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{2.31}$$

The numerical route approximates the gradient, $(\frac{\partial G}{\partial \lambda})_\lambda$, via the finite difference, $(\frac{\Delta G}{\Delta \lambda})_\lambda$. This free energy difference can be calculated via the Zwanzig equation, with the reference state at λ, and the perturbed state at $\lambda + \Delta\lambda$. This would give a forwards estimate of the free energy gradient. A perturbed state of $\lambda - \Delta\lambda$ yields the backwards estimate. These two estimates should of course be equal if $\Delta\lambda$ were sufficiently small, and the trajectory ran until the Zwanzig equation had converged. This method is normally referred to as *Finite Difference Thermodynamic Integration*[41] (FDTI), and again, there is a significant body of literature that demonstrates its successful application. These include its application to the relative binding free energy of thrombin inhibitors[42–44] and ligands to DHFR.[45] Most workers who use FDTI position their windows, and integrate the results, through the use of a Gaussian quadrature.[46] This is a technique that was developed by Gauss in the early nineteenth century to integrate definite functions. It works by recasting the integral of the function into the class of integrals known as 'polynomials times a known weighting function'. Given a function, $f(x)$, the method allows the optimal choice of the points along x to evaluate the function, and the optimal weights to give each of those points. This can be achieved by multiplying $f(x)$ by a known weighting function, $W(x)$. If $f(x)$ is a polynomial, then, given an integer number of evaluation points, N, it is possible to find a set of weights, w_j, and abscissas, x_j,

such that the approximation,

$$\int_a^b W(x)f(x)dx \approx \sum_{j=1}^{N} w_j f(x_j), \tag{2.32}$$

is exact. The values of w_j and x_j depend only on the choice of the weighting function, and the number of evaluation points.[46] Most workers use Gauss-Legendre quadrature, which uses a weighting function of $W(x) = 1$. The values of x_j and w_j are obtained from a lookup table, and are not in any way based on the shape of the PMF along λ. It is our opinion that this method of integration is not best suited for application to free energy calculations, as it was designed to integrate definite functions, and only has high accuracy for functions that are well approximated by polynomials. The gradients obtained along λ will have associated errors, and the underlying PMF could contain sharp peaks or troughs. The positioning of the λ-windows should thus reflect the shape of the PMF, with more λ-windows in places where there are large changes in the free energy gradient. All gradients should also be weighted equally, so as to avoid the possibility of points with large errors being highly weighted, while reducing the contribution from points with low errors. In addition, the use of Gaussian quadrature has, in our opinion, led to the use of far too few λ-windows along the perturbation. Previous studies[42–44] have used as few as 6 λ-windows, while another study[45] stated that increasing the number of λ-windows from 6 to 8 reduced the quality of the results. Such a reduction in quality would have been the result of poor positioning of λ windows, not an increase in their number. Our application of FDTI will thus use the same integration methods that have been used successfully in standard TI simulations,[35] namely trapezium rule integration over many, closely spaced λ-windows. Using this scheme, FDTI is very similar to FEP, and can be run with the same system conditions and reference states. In the case of FEP, the perturbed states are the neighbouring windows, while in the case of FDTI, the perturbed states are $\Delta\lambda$ above and below each window. FDTI and FEP become identical in the limit of the window width becoming equal to $\Delta\lambda$.

2.6.3 Fast Growth

One of the most recently developed free energy methods is the so-called 'Fast Growth' method.[25,26] This is an evolution of the Slow Growth method, which is the subject of a recent review.[47] The slow growth method estimates the free energy change between two systems within a single simulation. It achieves this by slowly increasing the value of λ by a constant amount, $\delta\lambda$, at each simulation step, such that at the start of the simulation, $\lambda = \lambda_0$, and by the end of the simulation, $\lambda = \lambda_1$. If the simulation consists of M steps, then $\delta\lambda$ is given by,[47]

$$\delta\lambda = \frac{\lambda_1 - \lambda_0}{M}. \tag{2.33}$$

The system is constantly being perturbed at every step of the simulation. This perturbation requires an amount of work. The work required to perform the entire perturbation, W, is formed as a sum over all of the simulation steps,[47]

$$W = \sum_{i=1}^{M} \delta\lambda \left(\frac{\partial E}{\partial \lambda}\right)_{\lambda=\lambda_0+i\delta\lambda}. \tag{2.34}$$

If $\delta\lambda$ were infinitesimally small, then this perturbation would occur infinitely slowly. This would mean that the system would stay in thermodynamic equilibrium throughout the mutation, and the perturbation would occur reversibly. If this were the case, then the work required to perform this change would be equal to the free energy associated with the change, i.e. $W = \Delta G$. However, if the change occurred in a finite time, then the response of the system would lag behind the perturbation,[23] and the simulation would move out of equilibrium. The resulting change would not be reversible, and some of the work would be dissipated. The amount of work required would be larger than the free energy change,[25] giving the slow growth inequality,

$$W \geq \Delta G. \tag{2.35}$$

The problem with the slow growth method is that the system is constantly being moved out of equilibrium. This leads to the inequality in equation 2.35. Recently, Jarzynski examined this inequality, and derived a remarkable equality,[25,26]

now known as the *Jarzynski identity*.[48] Jarzynski realised that the problem with the slow growth method was that the individual system moved away from equilibrium in an unpredictable manner.[49] It is quite possible that there exists no general formula that describes the nonequilibrium distribution of the system at the end of the perturbation.[49] Jarzynski then showed that if an ensemble of systems were perturbed away from equilibrium, then the behaviour of the ensemble as a whole was predictable. Thus the ensemble of non-equilibrium statistics could be related to an average over the equilibrated ensemble. Using these ideas, the Jarzynski identity relates the change in free energy of a perturbation, to the average of the work calculated for a slow growth simulation for each member of the original equilibrated ensemble,[25]

$$\overline{\exp(-W/k_BT)} = \exp(-\Delta G/k_BT). \tag{2.36}$$

The overbar in this equation denotes an average over an ensemble of slow growth simulations. The beauty of this equality is that it is independent of the speed of the perturbation.[25]

To use this equality, an equilibrated ensemble of structures at $\lambda = \lambda_0$ must be generated. A slow growth simulation should be performed for each member of this ensemble, although it can be performed with a fewer number of steps, and thus a larger $\delta\lambda$ than normal slow growth. The work necessary for each simulation should be calculated, and the average of the exponential obtained. This average will then equal $\exp(-\Delta G/k_BT)$. Because the rate of change of λ is higher than for slow growth, this method is referred to as fast growth.[50] This method has been tested on the calculation of the excess chemical potential of a Lennard Jones fluid,[50] the potential of mean force between a pair of methane molecules in water,[48] and the charging of a sodium ion in water.[51] These tests demonstrated that similar results were obtained via the fast growth method compared with other free energy methods, using comparable amounts of processor time. This was despite the applications only averaging the results from a small subset of starting points from the initial ensemble (between 10 and 3334). The main benefit of the method appears to be its huge potential for coarse level parallelisation over a very large Beowulf cluster. Once the initial generation of the equilibrated ensemble is complete, each

fast growth simulation could be performed in parallel on independent nodes.[48,50] The drawback of the method is that the average will only converge reliably if the fluctuations in the work are not too large.[25] This reasons for this are very similar to those used for the Zwanzig equation, which also collects the ensemble average of an exponential. In practice, this means that applications of this method also need to split up the λ-coordinate into a series of windows, and apply fast growth between neighboring windows.[48]

2.6.4 Adaptive Umbrella WHAM

Methods have been presented that treat λ as a simulation parameter. Through special treatment of λ, these methods are able to integrate the free energy along the λ-coordinate, and in so doing, produce the potential of mean force (PMF) across λ.[52] A totally different approach becomes apparent when it is realised that λ is just another coordinate of the system. λ does not have to be treated specially, and thus it is possible to make dynamic changes in λ throughout a simulation. The calculation of the free energy along λ then becomes equivalent to the calculation of the PMF along a normal structural coordinate, for which many methods have been derived. One such method is *Adaptive Umbrella WHAM* (AdUmWHAM),[53,54] a method that combines adaptive umbrella sampling[55,56] with the Weighted Histogram Analysis Method (WHAM).[57]

AdUmWHAM is typically used to derive the potential of mean force along structural coordinates, e.g. for dihedral angles in a small peptide.[53] AdUmWHAM can be applied to perturbations by realising that λ can be treated as a dynamic coordinate, and that it is possible to make moves in λ throughout a trajectory. This realisation was first made in a precursor to AdUmWHAM, λ-dynamics.[58]

λ-dynamics

λ-dynamics is another rigorous free energy method that was designed to achieve enhanced sampling of ligand configurations and orientations within a binding free energy calculation.[58–60] The method treats λ as a dynamic coordinate, and allows

motion along the λ-coordinate during normal configurational sampling. In this way, the ligands are dynamically morphing between each other during a single trajectory. This has the advantage that the ligands spend most of their time between $\lambda = 0.0$ and $\lambda = 1.0$, and thus the forcefield is much softer for the perturbing atoms. This should enhance the many configurational changes that are necessary to move through the λ-coordinate.[58] A potential disadvantage of the method is that λ could be changed too rapidly for the rest of the system to respond. The configurational sampling will thus lag behind the λ-sampling. This *Hamiltonian lag* is exactly the problem that is addressed by the Jarzynski equality in section 2.6.3. If the λ-sampling is too rapid, then the system will move out of equilibrium, and the change in free energy will contain systematic error. Unfortunately, *a priori* knowledge of the system's relaxation time is not possible, so the λ-sampling must be performed as slowly as possible within the constraints of the simulation.

λ-dynamics uses a variable λ-coordinate to calculate the free energy. To ensure that the entire λ-coordinate is sampled fully, the motion along λ must be encouraged through the use of an *umbrella potential.*

Umbrella Sampling

Umbrella sampling was first developed in the late 1970s, and originally applied to a small LJ fluid test system.[61] The aim of the method is to direct the sampling along a reaction coordinate to unfavourable regions. This is achieved through biasing the simulation through use of an *umbrella potential.* The umbrella is implemented as an additional term to the forcefield of the system, and acts to penalise or encourage particular configurations. This has the effect of making the system sample from a biased distribution. For example, figure 2.4 shows the potential of mean force, $P(\lambda)$, for a reaction coordinate, λ, for a hypothetical system. The PMF has two important minima, A and B, separated by a transition state, C. An umbrella potential, $U(\lambda)$, can be added to this system. This term is added to the forcefield, and has the effect of encouraging sampling where $U(\lambda)$ is low, and discouraging sampling where $U(\lambda)$ is high. In this case, the umbrella potential will encourage sampling of the transition state, thus allowing conformations A and B to interconvert freely.

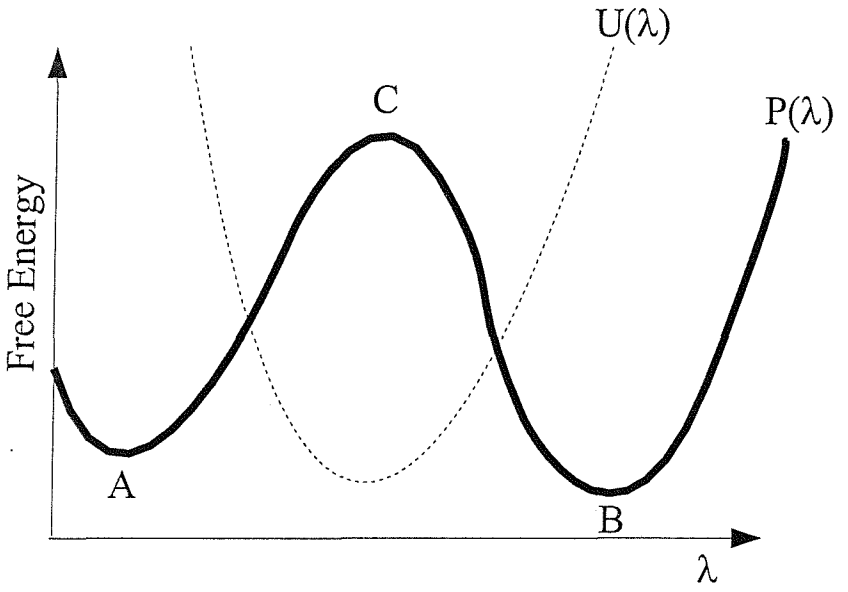


Figure 2.4: A hypothetical umbrella potential which could be used to enhance sampling of the reaction coordinate, λ.

The umbrella potential forces sampling from a biased probability distribution. If, in the general case, the reaction coordinate, λ, is a function of the system coordinates, r, then the biased probability, P_b, for the particular point $\lambda = \lambda_0$ is given by,[56]

$$P_b(\lambda_0) = \frac{\int \exp\left[-\beta\left(E(r)+U(\lambda)\right)\right]\delta(\lambda-\lambda_0)dr}{\int \exp\left[-\beta\left(E(r)+U(\lambda)\right)\right]dr}, \tag{2.37}$$

where $\beta = 1/k_BT$, $E(r)$ is the normal energy of the system for coordinates r, $U(\lambda)$ is the value of the umbrella potential, and δ is the Dirac δ function, which is equal to one for $\delta(0)$, and zero for all other values. The use of δ ensures that only configurations of r that map to $\lambda = \lambda_0$ contribute to the probability at $P_b(\lambda_0)$. Since the form of $U(\lambda)$ is known, it is possible to re-weight this biased probability distribution to return a Boltzmann weighted distribution along λ.[56] Looking at the specific

case of the point $\lambda = \lambda_0$,

$$P_b(\lambda_0) = \frac{\int \exp\left[-\beta\left(E(r)+U(\lambda)\right)\right]\delta(\lambda-\lambda_0)dr}{\int \exp\left[-\beta\left(E(r)+U(\lambda)\right)\right]dr}$$

Multiplying both sides by $\exp\left(\beta U(\lambda_0)\right)$,

$$P_b(\lambda_0)\exp\left(\beta U(\lambda_0)\right) = \frac{\int \exp\left[-\beta\left(E(r)+U(\lambda)\right)\right]\delta(\lambda-\lambda_0)dr \times \exp\left(\beta U(\lambda_0)\right)}{\int \exp\left[-\beta\left(E(r)+U(\lambda)\right)\right]dr}$$

Since $\exp\left(\beta U(\lambda_0)\right)$ is a constant, it can be moved inside the integral,

$$P_b(\lambda_0)\exp\left(\beta U(\lambda_0)\right) = \frac{\int \exp\left[-\beta\left(E(r)+U(\lambda)\right)\right]\delta(\lambda-\lambda_0)\exp\left(\beta U(\lambda_0)\right)dr}{\int \exp\left[-\beta\left(E(r)+U(\lambda)\right)\right]dr}$$

The δ function has eliminated all points where $\lambda \neq \lambda_0$. $U(\lambda_0)$ now cancels,

$$P_b(\lambda_0)\exp\left(\beta U(\lambda_0)\right) = \frac{\int \exp\left[-\beta E(r)\right]\delta(\lambda-\lambda_0)dr}{\int \exp\left[-\beta\left(E(r)+U(\lambda)\right)\right]dr}$$

using $Q_{bias} = \int \exp\left[-\beta\left(E(r)+U(\lambda)\right)\right]dr$

and $Q_{Boltz} = \int \exp\left[-\beta E(r)\right]dr$ gives

$$P_b(\lambda_0)\exp\left(\beta U(\lambda_0)\right) \times \frac{Q_{bias}}{Q_{Boltz}} = \frac{\int \exp\left[-\beta E(r)\right]\delta(\lambda-\lambda_0)dr}{Q_{bias}} \times \frac{Q_{bias}}{Q_{Boltz}}$$

$$P_b(\lambda_0)\exp\left(\beta U(\lambda_0)\right) \times \frac{Q_{bias}}{Q_{Boltz}} = \frac{\int \exp\left[-\beta E(r)\right]\delta(\lambda-\lambda_0)dr}{Q_{Boltz}}$$

The biased probability for $\lambda = \lambda_0$ is thus related to its

Boltzmann probability, P_{Boltz}

$$P_b(\lambda_0)\exp\left(\beta U(\lambda_0)\right) \propto P_{Boltz}(\lambda_0).$$

In the general case for any λ,

$$P_{Boltz}(\lambda) \propto P_b(\lambda)\exp\left(\beta U(\lambda)\right). \tag{2.38}$$

The unbiased probability distribution can be used to calculate the potential of mean

force along λ,[52]

$$\begin{aligned} G(\lambda) &= -\frac{1}{\beta}\ln\left[P_{Boltz}(\lambda)\right] \\ &= -\frac{1}{\beta}\ln\left[P_b\exp\left(\beta U(\lambda)\right)\times\frac{Q_{bias}}{Q_{Boltz}}\right] \\ &= -\frac{1}{\beta}\ln\left[P_b\exp\left(\beta U(\lambda)\right)\right]-\frac{1}{\beta}\ln\left[\frac{Q_{bias}}{Q_{Boltz}}\right] \\ &= -\frac{1}{\beta}\ln\left[P_b\exp\left(\beta U(\lambda)\right)\right]+C. \end{aligned} \tag{2.39}$$

The constant, C, is undetermined, though its only effect is to shift the entire PMF up or down in free energy. The value of C does not affect the shape of the PMF nor the values of any relative free energies.

The form of the umbrella potential is not *a priori* known, so some workers use many sequential umbrella potentials, e.g. harmonic potentials, to encourage the sampling to scan successive windows along the reaction coordinate.[62] The sampling within each window can be re-weighted, and as long as the windows are overlapping, the resulting probabilities may be combined to form the PMF along the entire reaction coordinate. The combination of each of the small pieces of PMF is based on changing the values of C from equation 2.39, such that the overlap between neighbouring windows is maximised. The optimal way to accomplish this is through the use of the Weighted Histogram Analysis Method.[57]

The Weighted Histogram Analysis Method

The Weighted Histogram Analysis Method[57] (WHAM) represents the optimal method of combining the statistics of multiple umbrella simulations into a single, self-consistent PMF. It achieves this by maximising the overlap in the PMF by weighting the statistics for each of the individual umbrella simulations.

The method works by dividing the reaction coordinate into a series of bins. The number of times that the sampling falls within each bin during simulation j, $n_j(i)$, is collected for each of the umbrella simulations. The value of the umbrella, for simulation j, at the centre of each bin i, $U_j(i)$ is also collected. The complete, unbiased probability for each histogram bin, i, across the reaction coordinate, $P_0(i)$,

is then estimated via the self-consistent solution to the WHAM equations,[63]

$$
\begin{aligned}
P_0(i) &= \kappa(i) \times \sum_j n_j(i) \\
\kappa(i) &= \frac{1}{\sum_j N_j f_j c_j(i)} \\
\text{where } f_j &= \frac{1}{\sum_i c_j(i) P_0(i)} \\
c_j(i) &= \exp\left(-U_j(i)/k_B T\right) \\
N_j &= \sum_i n_j(i).
\end{aligned}
\tag{2.40}
$$

A solution to these equations is obtained via an iteration. An initial estimate of $P_0(i)$ is made for each bin along the reaction coordinate. This estimate is used to calculate the weighting factor, f_j, for each simulation, which can then be used to estimate the unbiasing factor, $\kappa(i)$, for each bin in the histogram. This is then placed in the first equation to unbias the collected statistics along the reaction coordinate, and return a new estimate of $P_0(i)$. This iteration is repeated until the differences between the estimates of $P_0(i)$ are sufficiently small.

The use of these equations may be extended over multiple reaction coordinates, thus allowing the use of multidimensional umbrellas.[63,64]

Adaptive Umbrella Sampling

The use of the WHAM equations allows the facile unbiasing and combination of multiple umbrella sampling simulations. The method does not however solve the main problem of umbrella sampling, namely that of identifying the best umbrella potential. The best umbrella potential to use would be the negative of the PMF,

$$U(\lambda) = k_B T \ln P_{Boltz}(\lambda), \tag{2.41}$$

where $P_{Boltz}(\lambda)$ is the Boltzmann probability for each value of λ. This is the optimal umbrella, as it yields even sampling of the reaction coordinate.[52] Recalling that equation 2.38 shows that the biased probability along λ, $P_b(\lambda)$, is proportional

to the Boltzmann probability, $P_{Boltz}(\lambda)$,

$$P_b(\lambda)\exp\big(U(\lambda)/k_BT\big) \propto P_{Boltz}(\lambda)$$

Multiply both sides by $\exp\big(-U(\lambda)/k_BT\big)$

$$P_b(\lambda) \propto P_{Boltz}(\lambda) \times \exp\big(-U(\lambda)/k_BT\big)$$

Using the 'best' umbrella potential from equation 2.41

$$P_b(\lambda) \propto P_{Boltz}(\lambda) \times \exp\big(-k_BT \ln P_{Boltz}(\lambda)/k_BT\big)$$

$$P_b(\lambda) \propto P_{Boltz}(\lambda) \times \exp\big(-\ln P_{Boltz}(\lambda)\big)$$

$$P_b(\lambda) \propto P_{Boltz}(\lambda)/P_{Boltz}(\lambda)$$

Thus the biased sampling along λ will be even,

$$P_b(\lambda) \propto 1. \tag{2.42}$$

Adaptive Umbrella Sampling uses iterative simulations to refine an initial estimate of the umbrella potential until it is equal to the negative of the PMF.[55,56] A modification to adaptive umbrella sampling uses the WHAM equations to combine the statistics of each iteration. The combination of these two techniques is known as *Adaptive Umbrella WHAM*[53,54] (AdUmWHAM). The simulations are performed using the following protocol;[53,65]

1. An initial simulation is performed using a null, or zero umbrella. The system is free to sample the λ reaction coordinate. This coordinate has been divided into a series of histogram bins, i. The number of times that the system spends in each of these bins, $n_0(i)$, is recorded throughout the simulation.

2. The probability density for each bin along the reaction coordinate, $P_0(i)$, is estimated from the sampling histogram via,

$$P_0(i) = n_0(i)/N, \tag{2.43}$$

where N is the total number of simulation steps.

3. A new umbrella potential for the next simulation, $U_1(i)$, is estimated from the estimated probability density via,

$$U_1(i) = k_B T \ln P_0(i) \text{ for } P_0(i) \neq 0$$
$$U_1(i) = U_{min} \text{ for } P_0(i) = 1, \tag{2.44}$$

where U_{min} is the minimum value of the umbrella from all of the occupied bins. To prevent discontinuities, the umbrella potential can be processed via a smoothing function. A suitable smoothing function may replace each value of the umbrella in each bin, $U(i)$, by,[53]

$$U(i) = \frac{1}{3}\left(-0.3U(i-2) + 1.3U(i-1) + U(i) + 1.3U(i+1) - 0.3U(i+2)\right). \tag{2.45}$$

Multiple passes of this function may be used, and a continuous umbrella may be returned via fitting to a series of functions. The number of functions should not exceed the number of bins, and typically a combination of simple polynomials, sine and cosine functions are used.[53,63,65]

4. This umbrella is then used to bias a new simulation. The statistics along the reaction coordinate are collected into a new histogram.

5. This histogram, together with the umbrella potential, are processed via the WHAM equations. These unbias the sampling, and combine it with all previous iterations to estimate a new probability density along the reaction coordinate, $P_1(i)$. This is used to obtain a new refinement of the umbrella potential, $U_2(i)$. This umbrella may be processed and smoothed as in step 3.

6. This new umbrella is used to bias a new simulation, which leads to a new probability histogram. This can be processed via WHAM to unbias it and combine it with all previous iterations, returning a new refinement of the umbrella potential. This sequence is then repeated until even sampling along the reaction coordinate has been achieved.

7. Once even sampling has been achieved, the umbrella equals the negative of the PMF.[65] Relative free energies along the PMF can be obtained by taking

differences of points along the umbrella potential.[65]

2.7 Fast "Free Energy" Methods

Rigorous free energy simulations have the benefit that given sufficient time, they will converge onto the correct result for the forcefield representation of the system. However, "sufficient time" can mean millenia, depending on the complexity of the free energy change under investigation! Recent studies have shown that even simple free energy changes, for example the relative hydration free energy of acetaldehyde and acetone, require significant amounts of simulation (over 510 ps of MD).[66] Most of this simulation is on partially morphed, non-physical intermediates. To avoid spending so much simulation effort on non-physical systems, some workers have tried to develop methods that approximate the *absolute* binding free energies of ligands to proteins. Two such methods will be discussed; those based on the Linear Response approximation, and those based on the use of implicit solvents.

2.7.1 Linear Response Methods

The *Linear Interaction Energy* method (LIE)[67] allows the estimation of the absolute binding free energy of a ligand to a protein based on just two simulations. One is of the ligand free in solution, and the other is of the ligand bound to the protein. The absolute binding free energy, ΔG, is estimated from the simulation average of the difference in electrostatic energy between the ligand and the environment in the two simulations, $\langle \Delta U_{elec} \rangle$, and a similar average for the van der Waals ligand-environment energy, $\langle \Delta U_{vdw} \rangle$,

$$\Delta G = 0.5 \langle \Delta U_{elec} \rangle + \alpha \langle \Delta U_{vdw} \rangle. \tag{2.46}$$

Justification for this equation comes from the *Linear Response* approximation,[67,68] which provides the factor of 0.5. The linear relationship with the van der Waals energy was justified by the linear relationship between the number of carbons in an n-alkane, and its free energy of solvation. To obtain the value of α, equation 2.46

was fitted to the experimental binding free energies of a series of four endothiapepsin inhibitors. This yielded a value of 0.161, which has since been applied with varying degrees of success to a glucose/galactose receptor[69] and HIV-1 proteinase inhibitors.[70]

LIE has since been extended by other workers,[71–73] who have retrained equation 2.46 on new experimental binding data. These workers modified the original equation such that the electrostatic parameter is also optimised, and a third term is included to account for changes in solvent accessible surface area (SASA),

$$\Delta G = \beta\langle \Delta U_{elec} \rangle + \alpha\langle \Delta U_{vdw} \rangle + \gamma\langle \Delta SASA \rangle. \tag{2.47}$$

Other workers have expanded this equation even further, and developed models which include terms that relate the change in the number of hydrogen bonds to the solute, or the change in the aromatic surface area, to the absolute binding free energy.[71] The problem with the LIE method is that coefficients derived from one system and model are not readily transferable to another.[74] This means that the equation should be retrained for each protein-ligand set that it is applied to. This is an expensive procedure, and requires a large number of experimental protein-ligand binding free energies to have been previously determined. Wall *et al.* extensively studied the LIE equation,[74] and conclude that the version shown in equation 2.47 ideally requires a training set of 27 binding free energies. Also in this study, advanced statistical methods were used to investigate the correlation between the binding free energy and many individual forcefield components. The conclusion of this analysis was that, while the electrostatic and van der Waals parameters were the most important components to determine ligand binding, other forcefield components could also be used. There thus exists no generally applicable model equation on which to base any LIE model, and a complete statistical analysis should be performed for any system on which the method would be used.

2.7.2 Implicit Solvent Methods

The Linear Interaction Energy method suffers from the need to train the equation on existing experimental data. Implicit solvent methods avoid this problem, as they

use continuum electrostatics theory to calculate the free energy of solvation of ligands and proteins directly. The most rigorous implicit solvent models are based on the Poisson equation.[12] This equation describes the change in electrostatic potential, $\nabla\phi(r)$, with respect to the change in dielectric constant, $\nabla\varepsilon(r)$, and the underlying charge density, $\rho(r)$,

$$\nabla\varepsilon(r).\nabla\phi(r) = -4\pi\rho(r). \tag{2.48}$$

This equation is valid under conditions of zero ionic strength.[4] If mobile electrolytes are present in solution, then the Poisson-Boltzmann equation is used instead,[12]

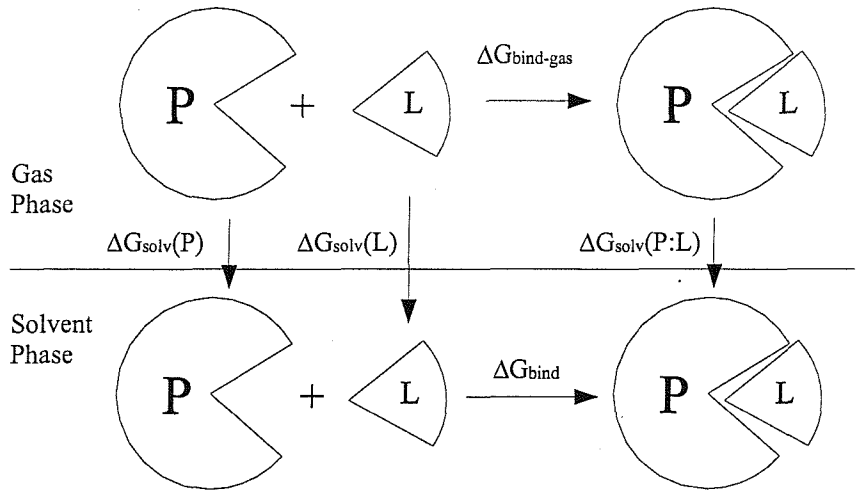
$$\nabla\varepsilon(r).\nabla\phi(r) - \kappa' \sinh[\phi(r)] = -4\pi\rho(r), \tag{2.49}$$

where κ' is related to the Debye-Hückel inverse length, κ, by,[4,12]

$$\kappa^2 = \frac{\kappa'^2}{\varepsilon} = \frac{8\pi N_A e^2 I}{\varepsilon k_B T}, \tag{2.50}$$

where e is the electronic charge, I is the ionic strength of the solution and N_A is Avogadro's number.

Since the Poisson equation can be considered a special case of the Poisson-Boltzmann (PB) equation, it is common for articles in the literature to describe the use of the PB equation when only the Poisson equation has been used.[4] The application of these equations to biomolecular systems has been well described.[75,76] Such applications must solve these equations numerically, most frequently through the use of a finite difference grid.[12] The molecule is represented on a cubic grid. Partial charges are spread over grid points, and the dielectric constant between grid points determined based on whether the points are inside the protein ($\varepsilon = \varepsilon_{cavity}$), or in the solvent ($\varepsilon = \varepsilon_{bulk}$). Having defined the system on the grid, the potential at each grid point can be determined through the finite difference equations. Two PB calculations are needed to calculate the solvation free energy of a molecule. One of these is the solvent phase calculation, with $\varepsilon_{bulk} = \varepsilon_{solvent}$, and the other is the gas-phase calculation, with $\varepsilon_{bulk} = \varepsilon_{cavity}$. From the potential at each point, $\phi(i,j,k)$, where i, j and k are the indices of the grid point, and the total atomic



$$\Delta G_{bind} = \Delta G_{bind-gas} + \Delta G_{solv}(P:L) - \Delta G_{solv}(P) - \Delta G_{solv}(L) \qquad (2.53)$$

Figure 2.5: The free energy cycle used by the PB/SA method to estimate the absolute binding free energy of a ligand (L) to a protein (P), ΔG_{bind}, from the solvation free energies of the ligand, protein and protein-ligand complex, ΔG_{solv}, and the binding free energy of the complex in the gas phase, $\Delta G_{bind-gas}$.

charge, $\rho(i,j,k)$, at that grid point, the electrostatic contribution to the solvation free energy, ΔG^{elec}_{solv} may be calculated via,[77]

$$\Delta G^{elec}_{solv} = \frac{1}{2}\sum_i\sum_j\sum_k \rho(i,j,k)\left(\phi_{solvent}(i,j,k) - \phi_{gas}(i,j,k)\right). \qquad (2.51)$$

The non-electrostatic contributions to the solvation free energy, ΔG_{solv}, may be estimated via a solvent accessible surface area (SASA) term, ΔG^{SASA}_{solv},

$$\Delta G_{solv} = \Delta G^{elec}_{solv} + \Delta G^{SASA}_{solv}. \qquad (2.52)$$

This combination of PB with a SASA calculation is normally termed PB/SA.[78] The absolute binding free energy of a ligand to a protein may be obtained from the free energy cycle in figure 2.5.[77,79] This equation may be used by taking a single protein-ligand structure, and using PB/SA to calculate the solvation free energies of each of the protein, ligand and complex, while they are in the bound configuration. The gas phase single structure binding free energy is then just the coulomb association energy between the protein and ligand.[77,79] While this method is quick

and easy to implement, it neglects the effect of different protein-ligand configurations, and the loss of ligand entropy upon binding. Woods *et al.*[77] have shown that this method can predict widely different binding free energies for protein-ligand complexes that differ only slightly in configuration. This configurational dependence, of up to 10 kcal mol^{-1}, means that it is difficult to even qualitatively rank the binding of a group of ligands to a protein using this method.

The MM/PBSA method was developed to overcome this problem.[80–83] This technique combines molecular mechanics (MM) energies with PB/SA by averaging the PB/SA analysis over multiple snapshots from a single molecular dynamics trajectory. The MD trajectory is run in explicit solvent. Periodic snapshots are taken from the simulation, and a PB/SA analysis is performed. The absolute binding free energy is then estimated via,[80]

$$\Delta G_{bind} = \langle E_{mm} \rangle + \Delta G_{solv} - T\Delta S. \tag{2.54}$$

The configurational energies are taken into account through $\langle E_{mm} \rangle$, which represents the difference in the average molecular mechanics energies between the complex, and free protein and ligand. ΔG_{solv} also represents the average difference between the solvation free energy of the complex, and that of the free protein and ligand. The change in entropy upon binding, ΔS, is estimated from the final snapshot, using a normal mode analysis.[82,83]

While the MM/PBSA method attempts to account for configurational and entropy effects, it does so in an approximate fashion. Its use of snapshots from an explicit MD trajectory means that the different configurations are not properly weighted, and in any case, are generated using a different forcefield to the one used during analysis. The main problem with the method is that it requires the taking of differences of very large numbers to return the final free energy. The molecular mechanics term represents the average difference in molecular mechanics energy between the complex, and free protein and ligand. These energies are of the order of thousands, to tens of thousands of kcal mol^{-1}.[80] The difference of the solvation free energies also asks for differences between numbers which are of the order of thousands of kcal mol^{-1}.[77,80] Since the final result is of the order of

one to ten kcal mol^{-1}, the potential for error is large.

2.8 Conclusion

Protein-ligand free energy simulations represent the system using simple molecular mechanics forcefields. The energy surfaces generated by these forcefields can be explored using sampling techniques such as Monte Carlo, or Molecular Dynamics, which preferentially sample thermally significant configurations. Through the use of a perturbing λ-coordinate, the techniques of FEP, FDTI or AdUmWHAM can be used to calculate relative free energies between different systems, assuming that the systems are not too dissimilar. These relative free energies may be combined via a cycle to obtain the relative binding free energies of different ligands to the same protein. While these methods are exact, and will eventually return the correct results for the forcefield, they could take a *very* long time to converge. To avoid this problem, non-rigorous free energy methods can avoid the sampling of non-physical intermediates, by attempting to estimate the absolute binding free energies of each ligand to the target protein directly. Such estimation is not trivial, and requires many assumptions to be taken. These assumptions could lead to problems of reliability and transferability.

Chapter 3

The Testing of Established Free Energy Methods

3.1 Introduction

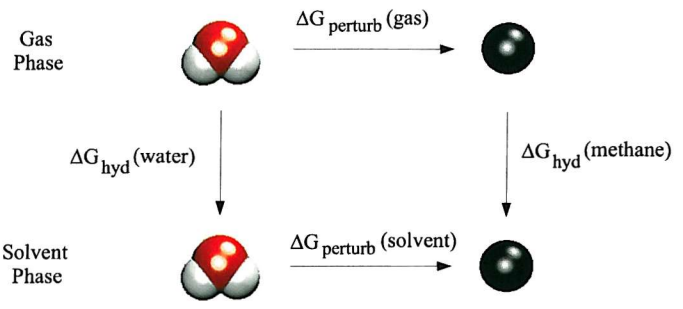
The aim of the work presented in this thesis is to develop theoretically rigorous methods to calculate the relative binding free energies of several ligands to a target protein. Ideally, these methods must be sufficiently fast to be used in commercial drug discovery, as in such an environment they could be applied to dozens, or even hundreds, of different ligands. The use of such large numbers of ligands means that the amount of user intervention with the method should be as low as possible. By this, we mean that the free energy calculation should work first time, and not require endless analysis of results, tweaking of parameters, or rerunning of the simulation. Thus a single, blind application of the method should reliably and reproducibly return the true result. It should do so with little prior knowledge of the system, other than the crystal structure, or any prior exploratory simulation on the system. To develop such methods, the behaviour of established free energy techniques needs to be investigated under these strict conditions. To this end, three such methods will be applied to a simple test system, and their reliability, and suitability for robust use will be appraised. For this evaluation to be useful, the test system must be sufficiently small to allow rapid simulations, yet sufficiently complex to represent the major features found in protein-ligand binding. To this

end, the solvation of methane in water was chosen, as it involves disruption of hydrogen bond networks, and large changes in solvent structure.

3.2 The Water-Methane System

The water-methane system, chosen as a test case in this work, has been the subject of much experimental and simulation study by other workers. The primary motivation for such study is to gain a better understanding of the hydrophobic effect, the tendency for hydrophobic solutes to self-aggregate in solutions of water. The hydrated methane system represents such cases in their simplest form, and thus permits the study of the experimental thermodynamics,[84] solvation shell structure[85] and the running of detailed molecular simulations.[86] The results of these studies reveal that the hydration of methane is thermodynamically unfavourable, with an experimental absolute free energy of +2.00 kcal mol^{-1} at 298 K.[87] Neutron diffraction experiments suggest that the first solvation shell contains 19 ± 2 water molecules,[85] with a peak in both the hydrogen-methane and oxygen-methane radial distribution functions occurring at 3.5 Å. Owing to the symmetrical nature of methane, no attempt was made to deconvolute these functions into hydrogen-carbon or oxygen-carbon RDFs.[85] The hydrogen peak is broader than the oxygen peak, suggesting that the hydrating water molecules point their O-H bonds towards the bulk.[85] These properties can be compared to those of pure water, which has an experimental absolute hydration free energy of -6.31 kcal mol^{-1} at 298 K.[87] Recent X-ray and neutron scattering results[88] show a single, sharp peak in the oxygen-oxygen RDF of height 2.8, at 2.8 Å. A second solvation shell is seen at about 4.5 Å, and a third at about 6.8 Å. The oxygen-hydrogen RDF shows a double peak, at 1.8 Å and 3.5 Å.[88]

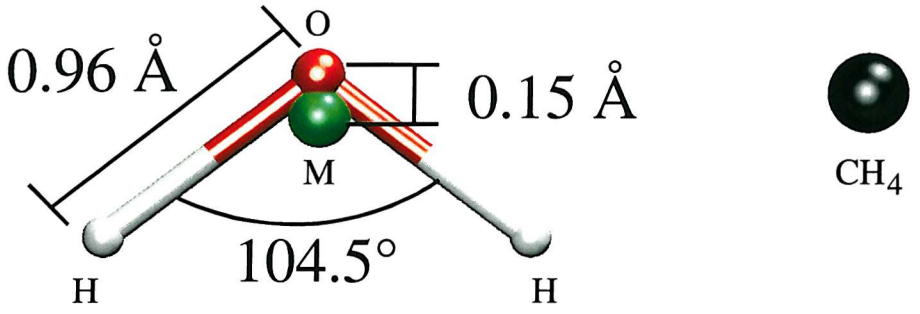
Previous studies have attempted to calculate the absolute hydration free energy of methane.[86] However, such simulations still represent a significant challenge, and are not suitable as test cases at this stage. It was decided that the best test case would be the calculation of the relative hydration free energy of water and methane. The reasons behind this decision are made clear in the free energy cy-



$$\begin{aligned} \Delta\Delta G_{hyd} &= \Delta G_{hyd}(methane) - \Delta G_{hyd}(water) \\ &= \Delta G_{perturb}(solvent) - \Delta G_{perturb}(gas) \end{aligned} \tag{3.1}$$

Figure 3.1: The free energy cycle used to calculate the relative hydration free energy of water and methane.

cle for this calculation (figure 3.1). The calculation of the absolute hydration free energy of methane requires the creation of a methane molecule into an already densely packed solution of liquid water. This is difficult, as a cavity would first have to be made in the solution to accommodate the methane. Fortunately, it can be seen from figure 3.1 that the relative hydration free energy of water and methane is equal to the non-physical perturbation of a single water molecule into methane in solvent, minus the same perturbation in the gas phase. If a rigid water model, and a rigid methane model were used, then the gas phase perturbation would have a free energy change of zero. The entire relative hydration free energy calculation could thus be achieved through a single perturbation of a hydrated water molecule into a hydrated methane molecule. This would mean that the change in solute dipole, and solvent rearrangement, hydrophobic effect and change in hydrogen bonding, would all occur in a single simulation. This simulation would produce a result that could be compared with the experimental value of 8.31 kcal mol^{-1}. This value is obtained from the difference of the absolute hydration free energy of methane[87] and the absolute hydration free energy of water.[87]



Atom	Charge	Sigma	Epsilon
O	0.000	3.15365	0.1550
M	-1.040	0.00000	0.0000
H	0.520	0.00000	0.0000
CH_4	0.000	3.730	0.294

Figure 3.2: The TIP4P water, and united-atom methane models.

3.2.1 Simulation Model

Good simulations rely on good models of the system in question. In this case, the methane was modelled as a single OPLS united atom CH_4 particle,[5] while water was modelled as a rigid OPLS TIP4P molecule[89] (figure 3.2). This model enhances the dipole of water by offsetting the oxygen's charge onto an "M" atom. The perturbation was designed to morph the oxygen atom of the water into the united-atom CH_4 particle. The remaining hydrogens and "M" atom were morphed into dummy atoms, by gradually scaling their charge and LJ parameters to zero. To improve convergence, and prevent abrupt changes as the hydrogen atoms disappeared, the hydrogens were gradually pulled into the oxygen as the perturbation progressed.[90] The O-H bond lengths started at 0.96 Å, and were linearly scaled down to 0.2 Å by $\lambda = 1.0$.

A single water-methane particle was then placed in an orthorhombic box of 1679 TIP4P waters, of initial dimension 37.3 Å × 37.9 Å × 37.4 Å. This system was then equilibrated for 2 M Monte Carlo (MC) steps, according to the parameters in table 3.1. The aim of this equilibration was to rattle out any bad contacts in the constructed solvent box, and its quality was ensured by monitoring the total energy of the system. The equilibration, and all subsequent simulations, were

Parameter	Value
Ensemble	Isothermal-Isobaric (NPT)
Simulation Temperature	25 °C
Simulation Pressure	1 atm
Boundary Type	Orthorhombic periodic boundaries
Solute / solvent move ratio	1 / 1600
Non-bonded cutoff	15 Å
Maximum solute translation / rotation	0.1 Å / 5.0 °
Maximum solvent translation / rotation	0.1 Å / 2.5 °
Frequency of volume moves	1 every 10375 MC moves
Maximum volume change	830 $Å^3$
Preferential sampling centre	Oxygen / CH_4 of the solute
Preferential sampling parameter (WKC)	200
Number of MC moves per simulation block	250 K
Number of simulation blocks	8

Table 3.1: Parameters used to control the simulations on the water-methane system.

performed by a modified version of MCPRO 1.5.[13] The modifications were to allow the code to run all of the free energy methods that are under investigation in this work, and were checked to ensure that they did not interfere with the normal running of the code. All simulations also used preferential sampling of the solvent,[91,92] as implemented within MCPRO (see appendix B). This scheme allowed the solvent molecules nearest to the perturbing solute to be moved more frequently than those further away, and its use has been shown to improve the rate of convergence of liquid phase MC simulations.[92] The use of these parameters gave a solute move acceptance rate of about 65% to 70%, a solvent move rate of about 65% to 70%, and a volume move rate of about 20% to 30%. Unless otherwise stated, these simulation conditions were used for all subsequent simulations on the water-methane system, and the final equilibrated structure was used as the starting structure.

3.3 Free Energy Perturbation

The water-methane system was investigated using *Free Energy Perturbation* (FEP). This method is described in detail in chapter 2 (section 2.6.1). The aim of this investigation was to see how FEP copes when it has no prior knowledge of the system. The FEP windows were thus spaced evenly over the entire λ-coordinate, 0.05 λ-units apart. This spacing represents a balance between using sufficiently small FEP windows, while keeping their number, 21, sufficiently low such that all windows may be run in parallel over a Beowulf cluster. 10 M steps of simulation were run within each window, with statistics collected over 20 blocks of 500 K steps. These 20 blocks can be divided into *equilibration* and *production* at the end of the simulation, when the predicted free energy change from each block can be calculated, and the convergence of that prediction ascertained. The complete simulation therefore totals 210 M Monte Carlo steps. This is in addition to the 2 M steps used during general equilibration to provide the starting structure for each λ-window. To investigate the reliability of the calculation, four copies were run, each starting from the same equilibrated structure. Each copy used a different random number seed, thus giving rise to four different sets of trajectories.

3.3.1 Results

The results for each simulation were generated by calculating the individual forwards and backwards free energies from each block of 500 K steps. The free energies were seen to converge after the first 3 M steps (figure 3.3). The first 3 M steps of sampling within each λ-window were therefore discarded as equilibration, and the data from the remaining 7 M steps per window used to calculate the free energy averages. The average free energies from each λ-window were then summed to yield the total forwards and backwards free energy across the entire λ-coordinate. The standard error was also estimated for each average, and then propagated across λ to estimate a maximum error on the calculation. A full description of the error analysis is given in appendix E. The results from the four FEP simulations are shown in table 3.2, and the predicted potentials of mean force are shown in figure 3.4.

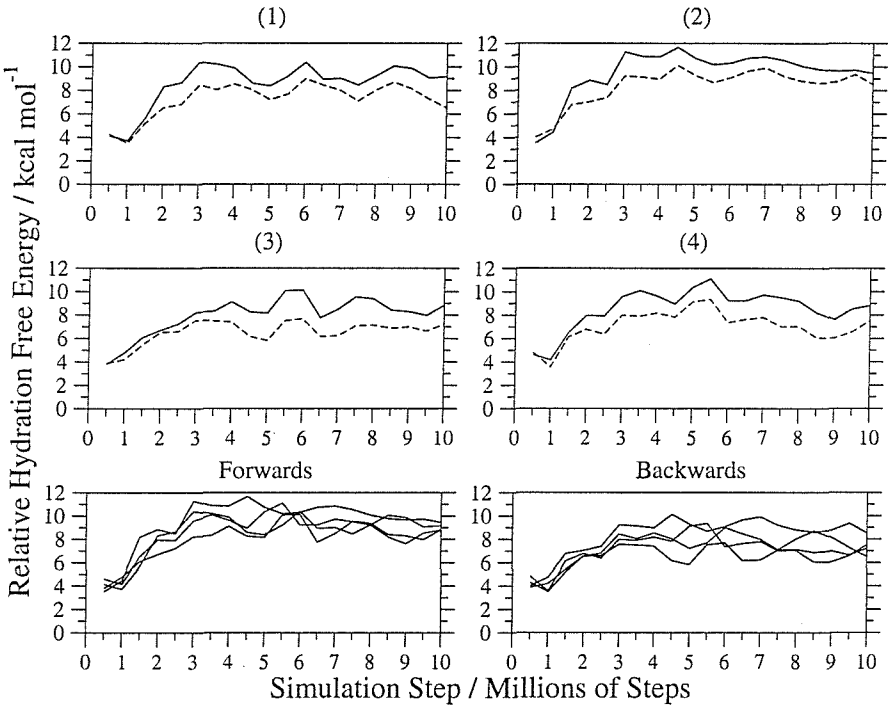


Figure 3.3: The convergence of the predicted relative hydration free energy of water and methane, as predicted by four independent FEP simulations, as a function of simulation step. The forwards (solid) and backwards (dashed) values at each step are based on their average calculated from the previous 500 K steps. The convergence for both the forwards and backwards predictions are also collected together into single plots.

Simulation	Forwards Free Energy / kcal mol^{-1}	Backwards Free Energy / kcal mol^{-1}
1	8.8 (0.7)	8.3 (0.7)
2	9.9 (0.7)	9.5 (0.6)
3	8.3 (0.6)	7.3 (0.6)
4	8.7 (0.7)	8.1 (0.7)
Average	9.0 (0.6)	8.3 (0.8)

Table 3.2: The relative hydration free energy of water and methane as predicted by four independent FEP simulations. One standard error is shown in parenthesis. The averages of the four forwards, and four backwards free energy results are also shown. The standard deviations of these averages are shown in parenthesis.

3.3.2 Analysis

Within error, the results from the four FEP simulations are in agreement with each other and with the experimental value of 8.31 kcal mol^{-1}. However, the spread of the results, from 7.3 to 9.9 kcal mol^{-1} is quite large. In addition, the results

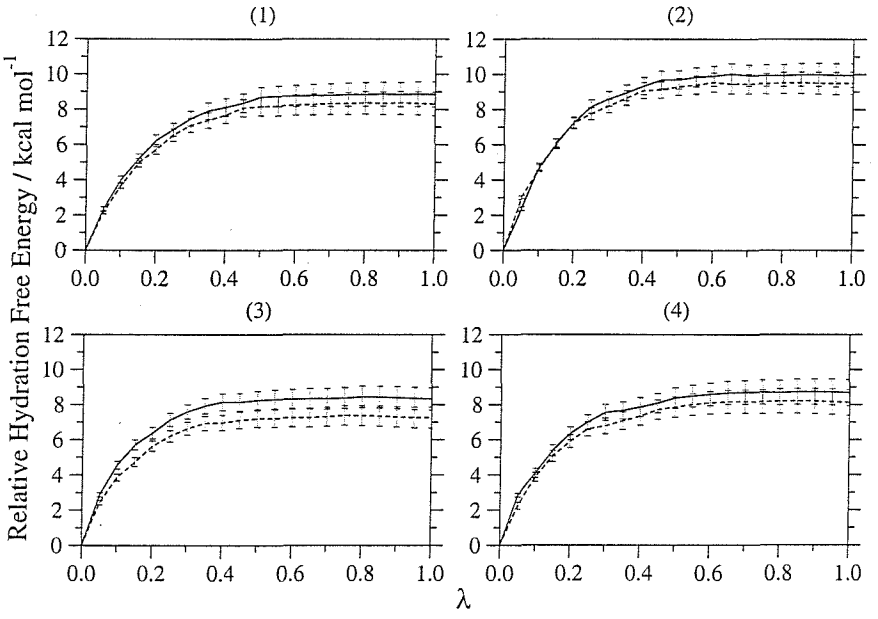


Figure 3.4: The potential of mean force for the relative hydration free energy of water and methane, as calculated by four independent FEP simulations. The error bars represent one standard error. Both the forwards (solid line) and backwards (dashed line) free energies are shown.

show significant difference in the forwards and backwards predicted free energies. This hysteresis, of 0.4 to 1.0 kcal mol^{-1}, shows that some of the λ-windows are not sufficiently close to converge the FEP simulations. Section 2.5 discussed how poor overlap between the reference and perturbed states could lead to a poor convergence of the Zwanzig equation. The shape of the PMF suggests that the overlap between the reference and perturbed states at low λ is not as good as the overlap at high λ. This is revealed by the high gradient of the PMF in this region, and by the growing hysteresis of the forwards and backwards free energies. This indicates that the parts of the simulation at low λ, between $\lambda = 0.0$ and $\lambda = 0.3$, should be rerun with a larger number of windows. However this would be against the spirit of this test, which required that the simulation proceeds with no prior knowledge of the system. These results could also have been improved through methods which optimise the window widths throughout the simulation, for example through the use of dynamically modified windows.[34] However, this technique requires that the simulations within each λ-window are run serially. Current computer technology mitigates against this approach.

Error Analysis

Three methods have been used to analyse the errors on these simulations. The first was through taking a simple average and standard deviation of the results of the four simulations. As the four simulations were run independently, their results should not be correlated, and thus the size of any random errors should be correctly determined. This analysis suggests that the error on the results of the simulation are between 0.6 kcal mol^{-1} and 0.8 kcal mol^{-1}. The second form of error analysis involved calculating the standard error for each individual trajectory, over blocks of 500 K steps. This standard error was then added to the free energy of each λ-window, and propagated across λ to get a total error on the relative hydration free energy. The blocks of 500 K steps were not independent, as they were calculated within single trajectories. This analysis suggests that the error on the simulation results is between 0.6 kcal mol^{-1} and 0.7 kcal mol^{-1}. The final method of error analysis employed during these simulations was the use of double-wide sampling to calculate both the forwards and backwards free energies. This suggests that the error was between 0.4 kcal mol^{-1} and 1.0 kcal mol^{-1}. It is useful to note that the magnitudes of all three error analyses are similar, and that the analysis over blocks of 500 K steps returns an error that is comparable to the standard deviation over four independent simulations.

Fluctuations in Interaction Energy

The results from all four FEP simulations show significant hysteresis between the forwards and backwards free energies. The analysis in section 2.5 suggests that this is the result from insufficient overlap between adjacent λ-windows. This insufficient overlap should result in large fluctuations in the difference in energy between the reference and perturbed states. To investigate whether this were the case, the difference between the perturbed state, average solute-solvent energy, and the reference state average solute-solvent energy was calculated for each value of λ, over blocks of 500 K steps. This analysis was performed for the last 7 M steps of simulations within each λ-window, and the results are shown in figure 3.5. The magnitude of the difference in interaction energy is at its largest between $\lambda = 0.0$

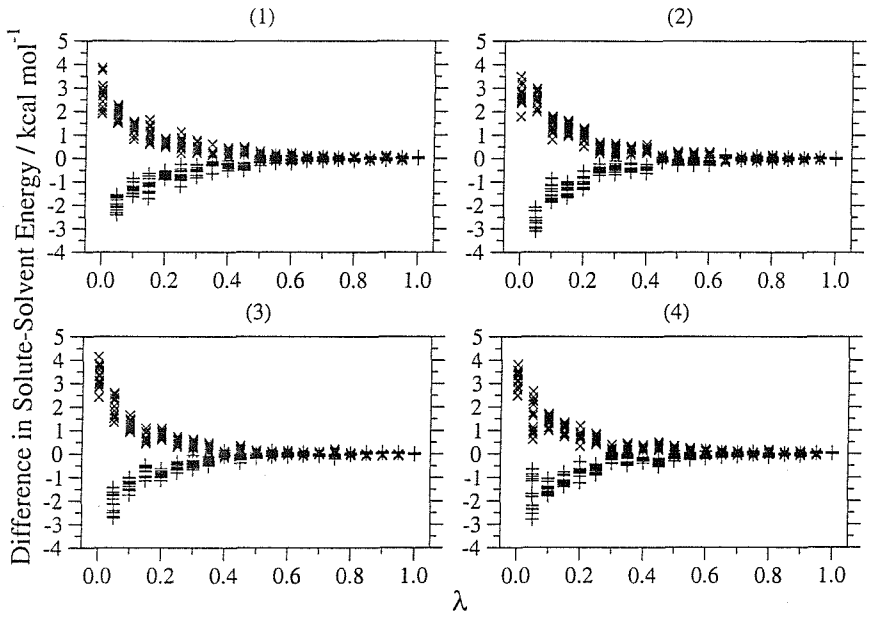


Figure 3.5: The difference between the average solute-solvent energy for the perturbed state, and reference state, as calculated over blocks of 500 K steps for the last 7 M steps of each λ-window, for each for the four simulations. The differences to the forwards perturbed state (crosses) and backwards perturbed state (pluses) are shown. There is no backwards perturbed state for $\lambda = 0.0$, nor forwards state for $\lambda = 1.0$.

and $\lambda = 0.3$. The fluctuations in this difference are significant in this range, giving a spread of interaction energies at $\lambda = 0.0$ of approximately 2.0 kcal mol^{-1}. This confirms the earlier hypothesis that poor overlap at low λ-values was the cause of the hysteresis in the free energy results.

Figure 3.5 shows that the forwards and backwards differences for each λ-value approximately mirror each other. This is against expectation, as the forwards difference represents the difference in interaction energy between the system at λ and $\lambda + \Delta\lambda$, while the backwards difference is between the system at λ and $\lambda - \Delta\lambda$. The reason for this observation is most likely because both energy differences were calculated over structures collected over the same reference state. Since the change in solute-solvent energy is approximately linear over small $\Delta\lambda$, the magnitude of the average will be determined by the range of configurations that have contributed.

Radial Distribution Function

Analysis of the results of four independent FEP simulations has demonstrated that poor overlap between the reference and perturbed states at low λ-values is responsible for the hysteresis between the forwards and backwards free energies. The source of this poor overlap is likely to be in the range of solvent structures sampled within each λ-window. To probe the solvent structures, the *radial distribution function*[93] (RDF) for each λ-window was accumulated. The RDF was derived from structures of the system taken every 50 K steps, for the last 7 M steps of simulation. The RDF between the solute oxygen and solvent oxygen, and solute oxygen and solvent hydrogen was calculated. At $\lambda = 0.0$ this yields $g_{(O-O)}$ and $g_{(O-H)}$. At $\lambda = 1.0$, where the solute oxygen has been fully perturbed into the united-atom methane, it yields $g_{(C-O)}$ and $g_{(C-H)}$. Since the RDFs were only evaluated between the central solute and the remaining solvent, they are quite noisy. They are, however, sufficiently detailed to yield information about the effect of the mutation on the primary hydration shell. The RDFs were collected with a histogram bin size of 0.1 Å. The RDFs for $\lambda = 0.0$ and $\lambda = 1.0$ are shown in figure 3.6. The RDFs produced from the simulations are seen to agree with those reported by other workers.[85,86,88] The water-water RDFs display a double-peak in $g_{(O-H)}$, at about 1.9 Å and 3.3 Å, while $g_{(O-O)}$ shows a sharp single peak of height 2.8 at about 2.8 Å. The methane-water RDFs show a broader single peak in $g_{(C-O)}$, at about 3.5 Å, while the double peak in $g_{(O-H)}$ has become a very broad single peak in $g_{(C-H)}$ at about 3.5 - 4.0 Å. These RDFs thus clearly demonstrate a large difference in solvent structure between water in water, and methane in water.

To see how the RDFs changed throughout the perturbation, they are plotted as a function of λ (figure 3.7). This figure shows that the majority of the change in solvent structure occurs at low λ-values, from $\lambda = 0.0$ to about $\lambda = 0.25$. This is characterised by the loss of a double peak in the hydrogen RDF, and a shift of the primary hydration shell in the oxygen RDF. The λ-range that contains the most structural change almost exactly corresponds to the range where most of the free energy change occurs in the PMF (figure 3.4). The large change in solvent struc-

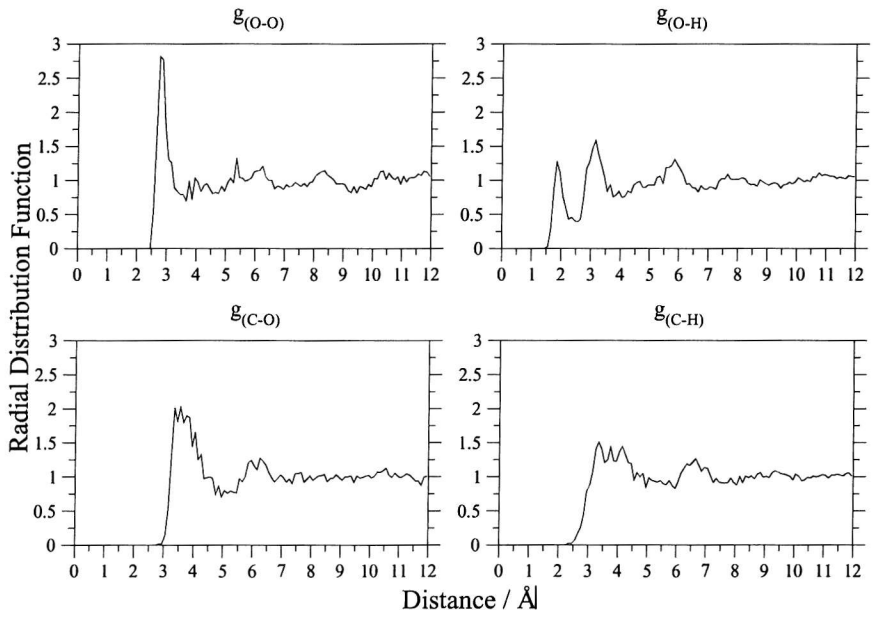
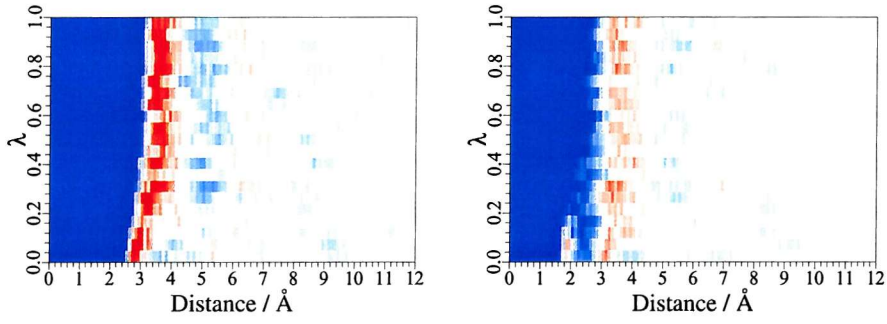


Figure 3.6: The radial distribution functions between the central solute atom and the solvent oxygens and hydrogens. Both the RDFs for $\lambda = 0.0$, which corresponds to water ($g_{(O-H)}$ and $g_{(O-O)}$), and $\lambda = 1.0$, which corresponds to methane ($g_{(C-H)}$ and $g_{(C-O)}$) are shown.



(a) Oxygen (b) Hydrogen

Figure 3.7: The radial distribution functions between the central solute atom and (a) solvent oxygens, and (b) solvent hydrogens, as a function of λ. The scale runs from blue (0.0), through white (1.0) up to red (greater than 2.0).

ture is thus the reason why there is little overlap between neighbouring windows at low λ-values. This explains the hysteresis in the FEP results, however it does not explain their wide variation. This large spread of predicted results, from 7.3 to 9.9 kcal mol^{-1}, is however explained by the noisiness of the RDFs. The statistics to generate these plots are derived from a single simulation trajectory at each

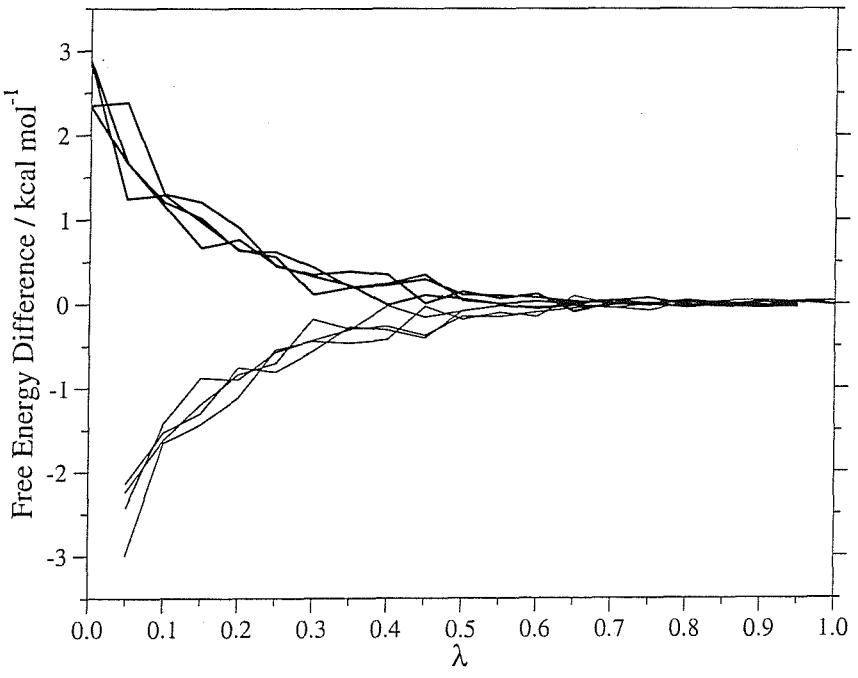


Figure 3.8: The individual average forwards (dark line) and backwards (light line) free energies for each λ-window, from four independent applications of FEP to calculate the relative hydration free energy of water and methane. There is no backwards free energy for $\lambda = 0.0$, nor forwards free energy for $\lambda = 1.0$.

λ-window. Each trajectory samples only a subset of all possible solute-solvent configurations. The subset sampled determines the RDF for that value of λ, and it is seen in figure 3.7 that the exact detail of the RDF varies from one value of λ to another. For example, since the perturbation over λ involves smooth, linear changes in potential functions, then the change in position and width of the second peak in the hydrogen RDF is also expected to change smoothly with λ. This is not seen, with the second peak starting at random between 2.5 Å and 3.0 Å for different λ-values. This *random sampling error* is also manifest in the individual free energies between neighbouring windows predicted by each of the four simulations. Figure 3.8 shows the average forwards and backwards free energies for each λ-window, from each of the four FEP simulations. Each free energy in this plot is generated from only a single trajectory. It is seen that each of the four different trajectories give rise to a slightly different predicted free energy. This effect is particularly acute at low λ-values, where the change of the solvent structure is at its greatest. In this region, the spread of predictions can be up to 1.0 kcal mol^{-1}. The effect

of random sampling error is thus easily observed through the running of multiple simulations. However, while it may be desirable to reduce this error through averaging the results of multiple trajectories, there is no mechanism through which this may be correctly achieved. To see why this is the case, let us consider the free energies predicted by two independent simulations, A and B, within a single λ-window. The most obvious method of averaging the free energy results of the two trajectories, ΔG_A and ΔG_B, is through a mean average,

$$\Delta G_{avg} = \frac{\Delta G_A + \Delta G_B}{2}. \tag{3.2}$$

However, the individual free energies are calculated via the Zwanzig equation,

$$\Delta G_A = -k_B T \ln\langle \exp(-\Delta E / k_B T) \rangle_A, \tag{3.3}$$

where ΔE is the difference in energy between the reference and perturbed states, and $\langle ... \rangle_A$ represents an average over the subset of configurations generated by simulation A. The Zwanzig equation forms the relative free energy over an average of the exponentials of the differences in energy. This suggests that the average free energy over multiple trajectories should also be formed as an average of exponentials,

$$\Delta G_{avg} = \ln\left(\frac{\exp\Delta G_A + \exp\Delta G_B}{2}\right). \tag{3.4}$$

Substituting the Zwanzig equation into this average, and using $\langle ... \rangle_A$ to represent the subset of configurations sampled during simulation A, and $\langle ... \rangle_B$ to represent the subset sampled during simulation B, gives,

$$\begin{aligned}
\Delta G_{avg} &= \ln\left[\frac{1}{2}\left(\exp\left(-k_B T \ln\langle \exp(-\Delta E/k_B T)\rangle_A\right)\right.\right. \\
&\qquad \left.\left. + \exp\left(-k_B T \ln\langle \exp(-\Delta E/k_B T)\rangle_B\right)\right)\right] \\
&= \ln\left[\frac{1}{2}\left(\langle \exp(-\Delta E/kt)\rangle_A + \langle \exp(-\Delta E/kt)\rangle_B + 2\exp(-k_B T)\right)\right] \\
&= -k_B T \ln\left[\frac{1}{2}\left(\langle \exp(-\Delta E/kt)\rangle_A + \langle \exp(-\Delta E/kt)\rangle_B\right]\right.
\end{aligned} \tag{3.5}$$

The average over the set of configurations sampled during simulation A represents the sum over all configurations of the system, Γ, of the product of the probability of that configuration, $p_A(\Gamma)$, with the exponential of the energy difference from that configuration, $\exp(-\Delta E(\Gamma)/k_BT)$. The probability, $p_A(\Gamma)$ is the probability of configuration Γ within the subset of configurations sampled by simulation A. The same correspondence can be made for the average over the subset of configurations from simulation B. Using this representation, the above equation can be represented as,

$$\begin{aligned}\Delta G_{avg} = -k_BT \ln\left[\frac{1}{2}\left(\sum_{\Gamma} p_A(\Gamma)\exp(-\Delta E(\Gamma)/k_BT)\right.\right.\\ \left.\left.+\sum_{\Gamma} p_B(\Gamma)\exp(-\Delta E(\Gamma)/k_BT)\right)\right] \qquad (3.6)\\ = -k_BT \ln\left[\sum_{\Gamma} \frac{p_A(\Gamma)+p_B(\Gamma)}{2}\exp(-\Delta E(\Gamma)/k_BT)\right].\end{aligned}$$

This equation shows that ΔG_{avg} will only equal the correct value of ΔG if the average of the probabilities from each simulation, for each configuration, is equal to the Boltzmann probability for that configuration, p_{Boltz}, i.e.

$$\frac{p_A(\Gamma)+p_B(\Gamma)}{2} = p_{Boltz}(\Gamma) \quad \text{for all } \Gamma. \qquad (3.7)$$

This will only be true in two cases. The first is when $p_{Boltz}(\Gamma) = p_A(\Gamma) = p_B(\Gamma)$, i.e. when both simulations A and B are fully converged. The other is when simulations A and B are both subsets of the same simulation trajectory. This conclusion is best illustrated by considering the following limiting case. The free energy calculated over a simulation of 5 M steps would not equal the average free energy calculated over 5 M trajectories each consisting of a single step, as this would represent an average over the 5 M starting configurations, and not an average over a correctly weighted ensemble. The average over 5 M steps will however equal the average free energy over 5 M trajectories, each consisting of 1 step taken in sequence from the ensemble of configurations generated via the 5 M step simulation. In this case, each of the 5 M trajectories represents a subset of configurations from the original 5 M step simulation, and thus the complete set of 5 M trajectories forms a complete and consistent ensemble that is equal to that generated by

the 5 M step trajectory. This analysis shows that there will be an error associated with the forming of a free energy average over multiple independent trajectories. While this error will be very small if the independent trajectories are long, it has the potential to become significant if the individual trajectories are short.

3.3.3 Summary

Free Energy Perturbation was applied to the calculation of the relative hydration free energy of water and methane. While the results were, within error, consistent and in agreement with experiment, the simulations produced a wide spread of results with a significant hysteresis. The source of these problems was traced to the challenges posed by the changing solvent configuration. The solvent configuration changed dramatically at low λ-values, leading to poor overlap in this region, and thus large hysteresis. The use of only a single trajectory at each λ-value led to a significant random sampling error. Each trajectory sampled only a subset of all possible configurations, leading to a random error in the individual free energies. FEP provides no mechanism for averaging this random error, as it provides no method to combine the statistics of multiple independent trajectories correctly.

The results of these simulations suggest that FEP is a good method to use if there is already a good understanding of the nature of the free energy change, though it may perform poorly if it is applied blindly to a new system. These simulations also suggest that multiple applications of FEP would be necessary to quantify the effect of random sampling error.

3.4 Thermodynamic Integration

The water-methane system was next investigated via the method of Thermodynamic Integration, (TI). This method is described in detail in chapter 2 (section 2.6.2). Unlike FEP, which calculates differences in free energy between two different λ-values, TI calculates merely the free energy gradient with respect to λ, for a particular value of λ, $(\frac{\partial G}{\partial \lambda})_\lambda$. The free energy gradients for a collection of λ-values can then be evaluated and integrated to return the PMF across λ, and thus the value of the free energy change. Because TI only needs to sample a sin-

gle λ-value, it should not suffer from the problems of poor overlap and hysteresis that reduced the quality of the FEP results on water-methane. However, unlike standard TI, the free energy gradients for this study were evaluated numerically, as described in section 2.6.2. This was to prevent substantial modifications of the simulation code, and to allow an almost identical simulation setup to the FEP simulations in the previous section. A reference state was defined for the value of λ for which the gradient was desired, and perturbed states were placed Δλ above and below this value. In this way, the Zwanzig equation[20] from FEP could be used to approximate $(\frac{\partial G}{\partial \lambda})_\lambda$ by $(\frac{\Delta G}{\Delta \lambda})_\lambda$. Since Δλ was very small, the forwards and backwards calculations of the gradient should be equal, and any deviation between the two could be used a sign of error. In addition, the small value of Δλ means that the reference and perturbed states are so similar, that there should be little problem from poor overlap. This approach of obtaining free energy gradients is typically referred to as *Finite Difference Thermodynamic Integration* (FDTI).[41]

3.4.1 Simulation Conditions

The application of FDTI to the water-methane system was almost identical to that used for the FEP simulations (sections 3.2.1 and 3.3). The same starting structure as in the FEP simulations was taken, and four identical FDTI simulations were run with different random number seeds. The value of Δλ used was 0.001, chosen to be sufficiently small to give good agreement between the forwards and backwards estimated gradients, and yet sufficiently large to prevent numerical precision problems during the running of the simulations. All other parameters and simulation lengths were identical to the FEP runs.

3.4.2 Results

The results were generated by calculating $(\frac{\Delta G}{\Delta \lambda})_\lambda$ for each block of 500 K steps of simulation per λ-value. The forwards and backwards gradients were seen to converge after about 3 M steps (figure 3.9). The values from the last 7 M steps were thus averaged and integrated via the trapezium rule to obtain the forwards and backwards predicted potentials of mean force. To deal with the ends of the PMF,

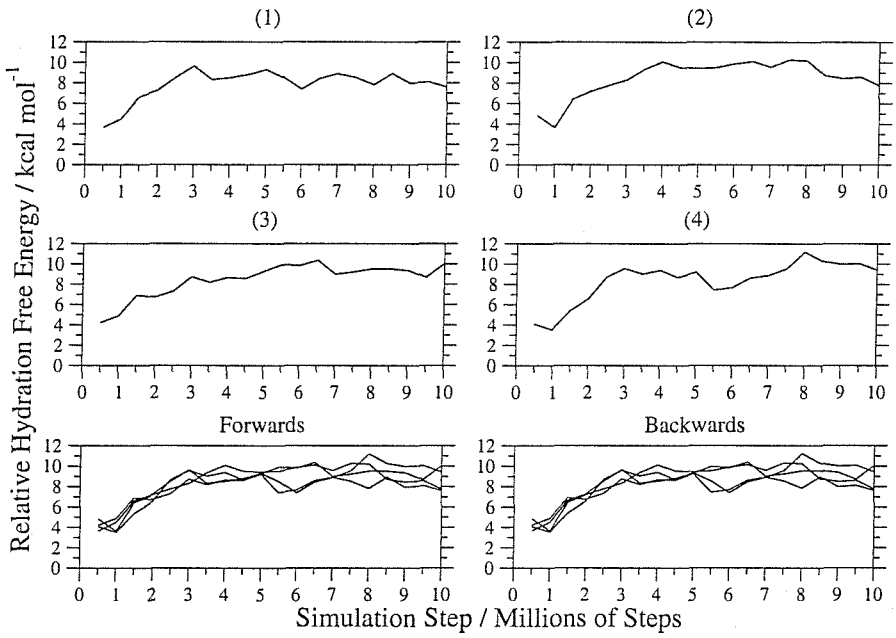


Figure 3.9: The convergence of the predicted relative hydration free energy of water and methane, as predicted by four independent FDTI simulations, as a function of simulation step. The forwards (solid) and backwards (dashed) values at each step are based on their average calculated from the previous 500 K steps, and are almost equal for each step. The convergence for both the forwards and backwards predictions are also collected together into single plots.

Simulation	Forwards Free Energy / kcal mol^{-1}	Backwards Free Energy / kcal mol^{-1}
1	8.4 (0.6)	8.4 (0.6)
2	9.4 (0.7)	9.4 (0.7)
3	9.3 (0.7)	9.3 (0.7)
4	9.2 (0.7)	9.2 (0.7)
Average	9.1 (0.4)	9.1 (0.4)

Table 3.3: The relative hydration free energy of water and methane as predicted by four independent FDTI simulations. The standard error is given in parenthesis. The averages of the four forwards, and four backwards free energy results are also shown. The standard deviations of these averages are shown in parenthesis.

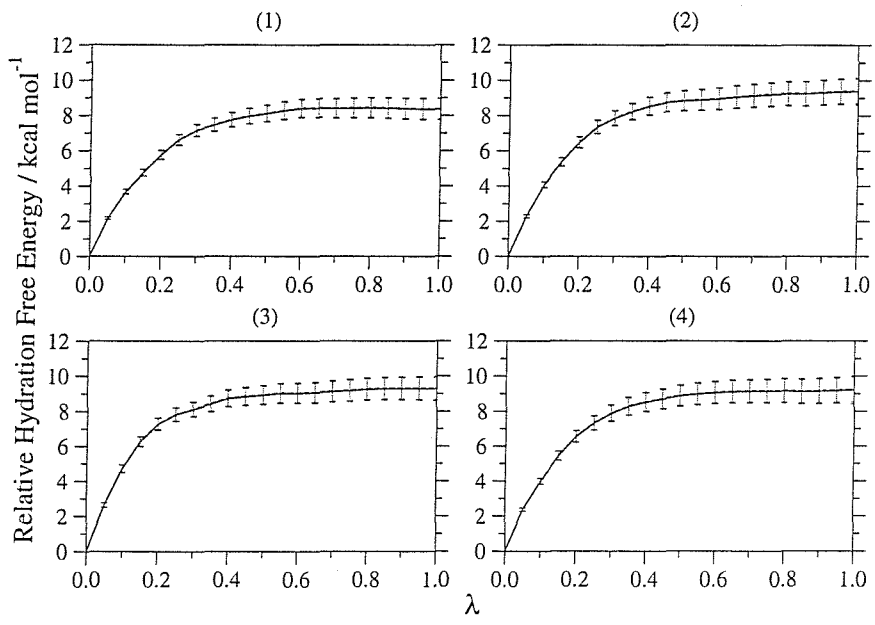


Figure 3.10: The potential of mean force for the relative hydration free energy of water and methane, as calculated by four independent FDTI simulations. The error bars represent one standard error. Both the forwards (solid line) and backwards (dashed line) free energies are shown.

the backwards predicted gradient for $\lambda = 0.0$ was obtained from the negative of the forwards gradient at that value. A similar scheme was used to get the forwards gradient at $\lambda = 1.0$. The error on the calculations were obtained in an identical manner to FEP. The results of the four FDTI simulations are shown (table 3.3), as are the PMFs (figure 3.10).

3.4.3 Analysis

Within error, the results from all four FDTI simulations agree with each other, the four FEP simulations, and with experiment. Unlike FEP, the results show almost no hysteresis, due to the very small $\Delta\lambda$ used between the reference and perturbed states. The values from the four simulations are very consistent, though simulation 1 predicts a free energy which is about 1 kcal mol^{-1} lower than simulations 2 to 4. While the spread of results is much improved over FEP, the standard error on the individual results is of the same magnitude.

The radial distribution functions of the FDTI simulations are essentially the same as those generated by the FEP simulations, as both FEP and FDTI shared

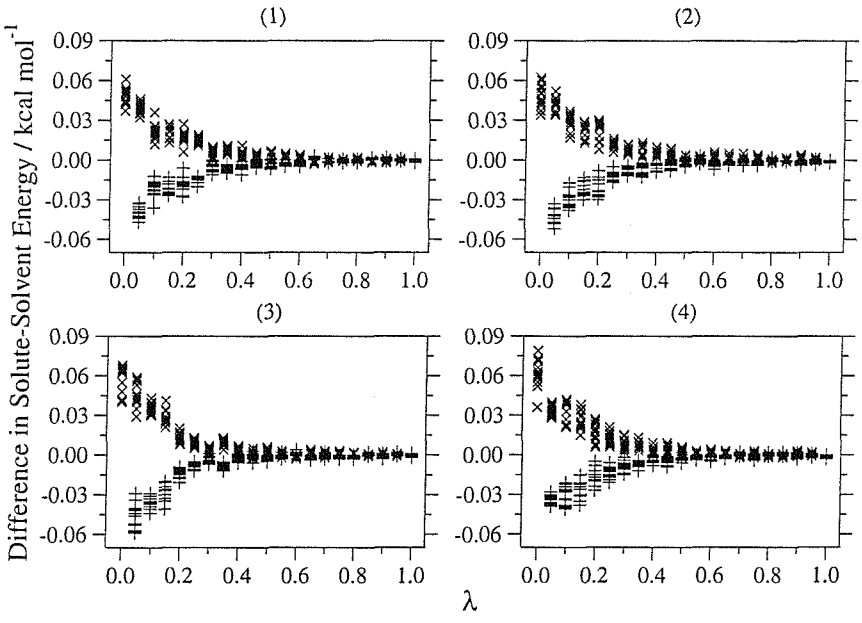


Figure 3.11: The difference between the average solute-solvent energy for the perturbed state, and reference state, as calculated over blocks of 500 K steps for the last 7 M steps of each λ-window, for each for the four simulations. The differences to the forwards perturbed state (crosses) and backwards perturbed state (pluses) are shown. There is no backwards perturbed state for $\lambda = 0.0$, nor forwards state for $\lambda = 1.0$.

the same reference states. However, the perturbed states for both methods were different, with FDTI using perturbed states which had a much smaller $\Delta\lambda$ to the reference states (0.001 compared to 0.05). The difference between the perturbed state, and reference state average interaction energies were calculated as in section 3.3.2 for each λ-value of each simulation. These are shown in figure 3.11. These differences show very similar patterns to FEP, with mirroring of the forwards and backwards differences, and the absolute values being larger at low λ-values. However, the magnitude of the differences is significantly lower, being around 50 to 60 times smaller than those from FEP. This corresponds to the use of a $\Delta\lambda$ which was also 50 times smaller than FEP. The reduced magnitudes of the free energy differences are translated into reduced fluctuations of the differences in energy. This is the reason behind the almost non-existant hysteresis on the results from the FDTI simulations.

Unlike FEP, the FDTI simulations do not show any hysteresis. They do however exhibit random sampling error. This is clearly seen in the average free energies

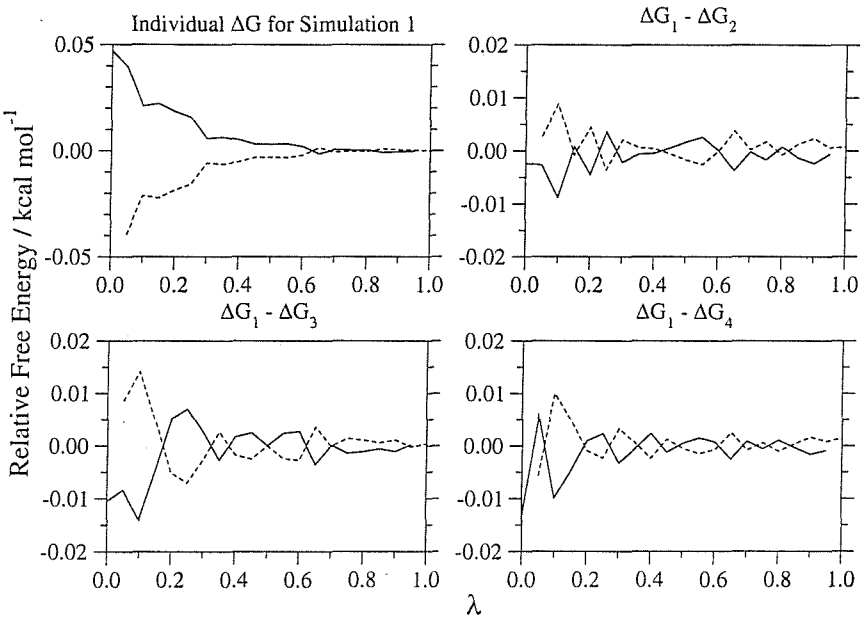


Figure 3.12: The average free energies from four FDTI simulations calculated between the reference and perturbed states for each value of λ. Both the forwards (solid line) and backwards (dashed line) are shown. The first plot shows the actual free energies, while the remaining plots show the difference between the free energies from the first simulation and each of the other three simulations.

calculated by each simulation at each λ-value (figure 3.12). This figure shows the individual free energies calculated between the reference and perturbed states for each value of λ, for simulation 1. It also shows the differences between the free energies calculated for simulation 1, with those calculated for each of the other 3 simulations. It is clear from this figure that there is much variation between the free energies calculated from each simulation at each λ-value. This variation is at a maximum at low λ-values, where its magnitude is about 0.01 kcal mol^{-1}. While this may seem small, it must be remembered that the numerical integration of these free energies involves their multiplication by $\frac{\delta\lambda}{\Delta\lambda}$, where $\delta\lambda$ is the difference in λ between adjacent windows, and $\Delta\lambda$ is the difference in λ between the reference and perturbed states. Since, for these simulations, $\Delta\lambda = 0.001$, and $\delta\lambda = 0.05$, this variation of 0.1 kcal mol^{-1} is multiplied into a random sampling error of approximately 0.5 kcal mol^{-1}. The repetition of this error for the first two values along λ is enough to explain the 1.0 kcal mol^{-1} random sampling error in the final free energy results. The reason for this random sampling error is the same as for

FEP, namely that only a single trajectory is run at each λ-value. Each trajectory will have independently covered a particular subset of configurations, leading to a slightly different free energy result. As in the case of FEP, FDTI provides no mechanism for correctly averaging the results of several trajectories to allow this random sampling error to be reduced.

3.4.4 Summary

FDTI was applied to calculate the relative hydration free energies of hydration of water and methane. The results show less spread than FEP, though have a similar standard error. The spread of predicted results was seen to arise from a spread of estimates of the free energy values at low λ-values. This spread highlights the main problem of FDTI, namely that an individual free energy is estimated from only a single simulation trajectory. This free energy will therefore be strongly dependent on the peculiarities of the subset of configurations covered by that single trajectory. As in the case of FEP, FDTI provides no framework for averaging out this random sampling error by averaging multiple trajectories at each λ-value.

3.5 Adaptive Umbrella WHAM

The relative hydration free energy of water and methane has been investigated via the established free energy methods of FEP and FDTI. Both methods estimate the potential of mean force across a λ-coordinate by placing trajectories at individual points along λ. The use of a single trajectory at each λ-point means that the values calculated are subject to a random sampling error. This random error cannot be reduced by averaging over multiple independent trajectories, as the method provides no mechanism to correctly combine their statistics. *Adaptive Umbrella WHAM* (AdUmWHAM) is a method that overcomes this problem as it uses the Weighted Histogram Analysis Method[57] (WHAM) to correctly combine the statistics of multiple trajectories. It is described in detail in chapter 2, section 2.6.4. The method uses a dynamic λ-coordinate, by allowing each trajectory to make moves along λ, as well as in normal coordinate space. Sampling along the λ-coordinate is enhanced by an umbrella potential,[61] which is gradually refined throughout the

simulation. Once converged, the umbrella will encourage even sampling of the λ-coordinate. At this point, the umbrella will equal the negative of the potential of mean force along λ. Multiple trajectories could be run in parallel, and the results from each trajectory, and each iteration, combined via WHAM.[57] In this way, multiple trajectories visit each point along the λ-coordinate, and their contributions to the free energy change are correctly weighted.

3.5.1 Simulation Conditions

The application of AdUmWHAM to the calculation of the relative hydration free energy of water and methane was designed to be as similar to the applications of FEP and FDTI as possible. This was to ensure that valid comparisons could be made between the three methods. Four identical simulations were run, each starting from the same initial structure as used by the FEP and FDTI simulations. The simulation parameters were identical to those used in FEP and FDTI (table 3.1), and the simulation protocol is described in section 2.6.4. Each of the four simulations were identical, except for the use of a different random number seed. A λ-move was made every 500 normal MC moves. This value was chosen to maximise the number of steps between λ-moves, thus minimising the affect of Hamiltonian lag discussed in section 2.6.4. The size of each λ-move was randomly chosen to be between -0.05 and +0.05. This move size was chosen to allow good sampling of λ, without moving λ too quickly for the system to respond. Each simulation comprised two trajectories running in parallel. Both trajectories were generated simultaneously, and when each one finished, it submitted its λ-sampling statistics to a custom program that performed the WHAM analysis. The WHAM analysis was performed according to the protocol described in section 2.6.4. The resulting umbrella was then smoothed, twice, according to the protocol in section 2.6.4, and fitted to the sum of 3 polynomial functions, 50 sine functions and 50 cosine functions,

$$U(\lambda) = \sum_{i=1}^{3} c_i \times \lambda^{i-1} + \sum_{i=1}^{50} c_{i+3} \times \sin(2\pi i\lambda) + \sum_{i=1}^{50} c_{i+53} \times \cos(2\pi i\lambda), \qquad (3.8)$$

where the 103 coefficients, c_i, were obtained through a single value decomposition algorithm.[46] This umbrella was immediately returned to continue the simulation that submitted the trajectory, which then continued sampling. In this way, two trajectories were run in parallel, with both contributing statistics to refine the umbrella.

Each trajectory ran in blocks of 200 K steps, starting with a new, random value of λ at the start of each block, but otherwise using the coordinates from the end of the previous block. Each block used 40 K steps of equilibration, and 160 K steps of data collection. When a block completed its 200 K steps, it returned the results to a server, which processed these results via WHAM, and returned a new estimate of the umbrella potential. The new umbrella was immediately taken by the trajectory to start a new block, and thus both trajectories ran asynchronously in parallel. The simulation was run for 500 completed WHAM iterations. This corresponded to a total simulation length of 100 M MC steps, which was about half of the 210 M used in FEP and FDTI. However, FEP and FDTI ran 21 trajectories in parallel, with each trajectory only covering 10 M steps. AdUmWHAM is not as coarsely parallel, as each trajectory must cover the entire λ-coordinate, and share its results via WHAM with the other trajectory. The simulation thus took five times longer than FEP or FDTI, as it used 50 M steps per trajectory.

3.5.2 Results

The results of the four AdUmWHAM simulations are shown (table 3.4), as are the predicted PMFs (figure 3.13). The results were obtained by taking the negative of the umbrella from the final iteration. The error on these results is difficult to quantify, and requires knowledge of the underlying λ-sampling. Because of this, the presentation and discussion of errors will wait until after the analysis and discussion of the underlying λ-sampling.

3.5.3 Analysis

The results from the four AdUmWHAM simulations on water-methane show poor agreement with FEP, FDTI or experiment. The spread of results is also quite large (5.6 kcal mol^{-1} to 7.2 kcal mol^{-1}). A possible reason for the poor results is that

Simulation	Free Energy / kcal mol^{-1}
1	6.2
2	7.1
3	6.4
4	5.6
Average	6.4 (0.5)

Table 3.4: The relative hydration free energy of water and methane as predicted by four independent AdUmWHAM simulations. The average over the four results is also shown, together with the standard deviation in parenthesis.

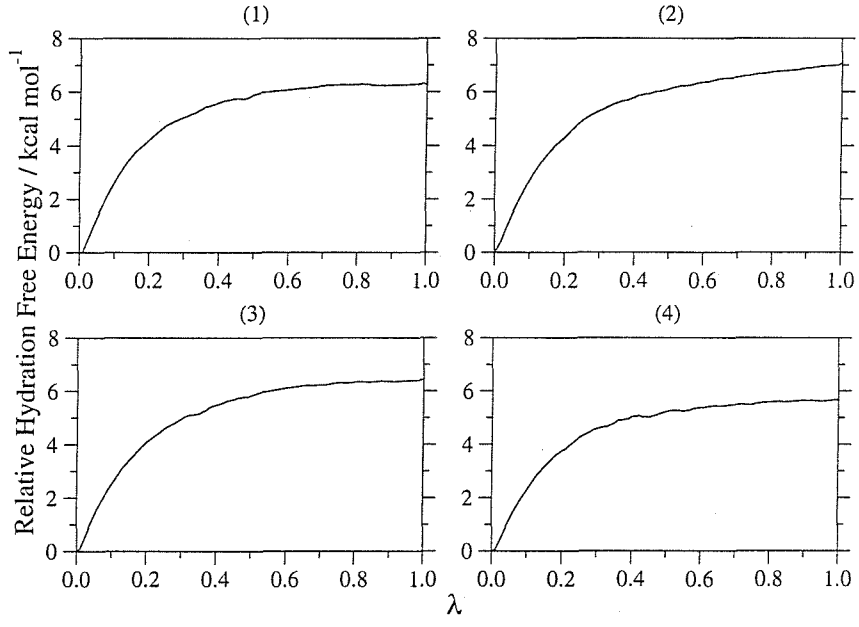


Figure 3.13: The predicted PMF of the relative hydration free energy of water and methane, as predicted by four equivalent AdUmWHAM simulations. The PMF is taken as the negative of the final iteration's umbrella.

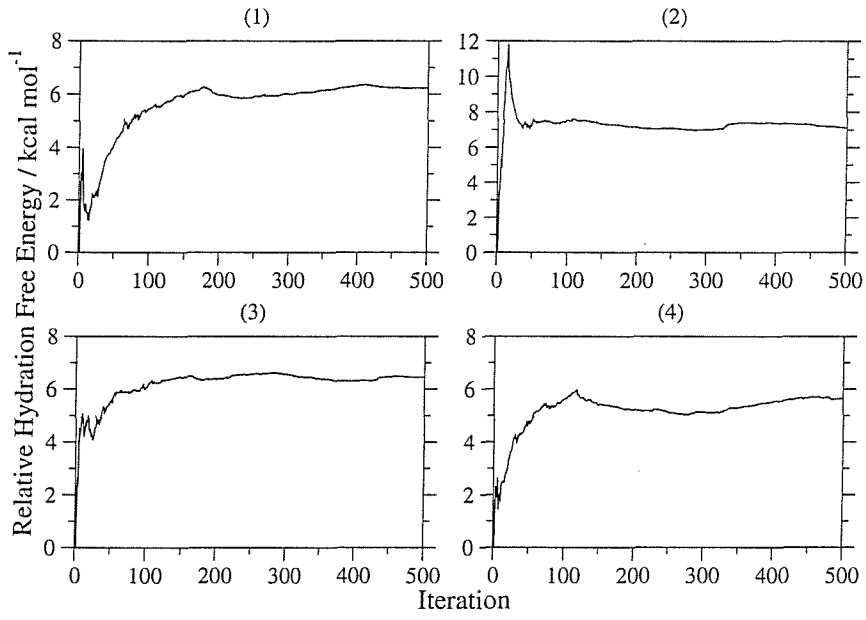


Figure 3.14: The convergence of the predicted relative hydration free energy of water and methane as a function of AdUmWHAM iteration. The relative free energy is taken as the difference between the ends of the umbrella for each iteration.

the method has not converged after 500 iterations. To investigate whether this were the case, the predicted relative hydration free energy was plotted as a function of iteration (figure 3.14). This plot was obtained from the difference between the ends of the umbrella produced by each iteration. It shows that the predicted free energy changes greatly in the first 50 to 100 iterations, and then settles down, slowly oscillating around a mean value. This implies that the AdUmWHAM simulations all converged rapidly. This view is reinforced through investigation of the actual sampling along λ, which shows that even sampling had been obtained by the end of the four simulations.

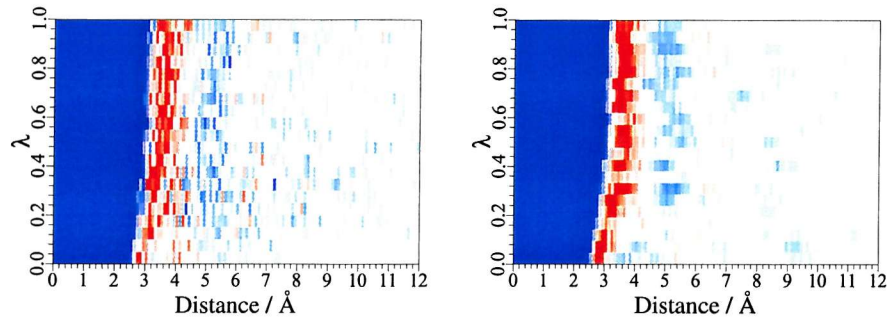
Since the results seem to suggest that the AdUmWHAM simulations have converged, another reason must be sought to explain why the four simulations give such varied results, and disagree with FEP and FDTI. One possible avenue of exploration is the amount and type of solvent sampling. The radial distribution function was calculated for the system structure at the end of each iteration, and histogrammed according to the value of λ at the end of the iteration. This procedure produced an RDF with respect to λ, and the use of a λ-histogram bin size of 0.05 leads to comparable plots to figure 3.6 on page 54 from the first FEP simulation.

Comparisons between these RDFs and those calculated via FEP should be made with caution, as the total number of data points within the FEP RDFs is larger than for those calculated via AdUmWHAM. Also, the FEP results are evenly spread over λ, while the dynamic λ-coordinate means that the AdUmWHAM results are less evenly distributed. The RDF with respect to λ was calculated for all four simulations. All four sets of plots were similar, so only those from the first simulation are shown (figure 3.15).

While the oxygen RDF showing the same main features as those seen in the RDF from the FEP simulations (figure 3.6), the hydrogen RDF fails to show the distinct double-peak at low λ-values. While a double peak can be seen, it is very poorly defined, and the main peak is quite similar over the entire λ-coordinate, remaining broad and noisy. There is no sharp peak at 3.0 Å at $\lambda = 0.0$ in this hydrogen RDF. Instead there is a broad and noisy peak from 3.0 to 4.0 Å. This suggests that AdUmWHAM is failing to sample the key solvent configurational change for this calculation. Since the methods appears to have converged, the reason for this must lie within the simulation protocol. Given that the results of the simulations show some form of systematic error, an analysis of the data in terms of random errors would be misleading.

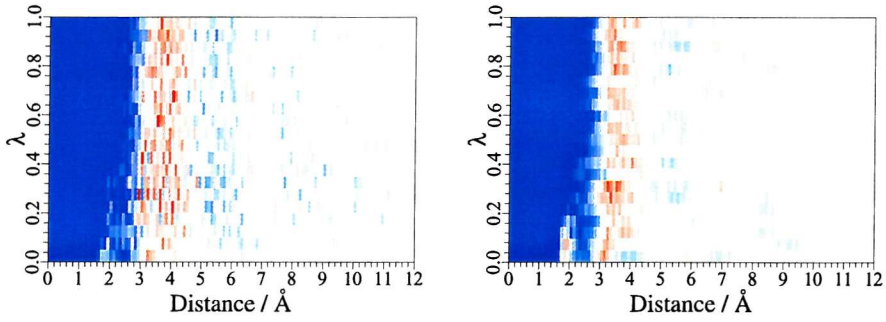
3.6 Changing the AdUmWHAM Protocol

Four AdUmWHAM simulations were applied to calculate the relative hydration free energy of water and methane. The results produced were contradictory and failed to agree with either experiment, or previous simulations. Analysis of the results suggested that the λ-sampling had become even, and that the umbrellas had converged. However, investigation of the RDFs showed that the system was not sampling solvent configurations correctly. The source of this problem is likely to reside in the protocol used to run the AdUmWHAM simulations. A key part of this protocol is the randomisation of λ after each iteration, i.e. that the value of λ was re-assigned at random after each iteration. This was done to ensure that the value of λ did not become stuck between iterations, and thus hinder the convergence of the calculation. At the start of each iteration, the system was given 40 K steps to adjust



(a) AdUmWHAM Oxygen

(b) FEP Oxygen



(c) AdUmWHAM Hydrogen

(d) FEP Hydrogen

Figure 3.15: The oxygen and hydrogen RDFs calculated from the first FEP and first AdUmWHAM simulation. The colour scale is the same as in figure 3.7.

to the new λ-value. In retrospect, this was insufficient, and indeed the uniform randomisation of λ could bias the statistics to make it appear that the λ-sampling was more even than it actually was. This would also explain the poor RDFs, as a predominantly 'methane' solvent structure may suddenly find that it is hydrating a water molecule, as λ is randomly changed from near 1.0 to near 0.0. The solvent is unlikely to respond in 40 K steps, and thus the statistics collected will be incorrect, and the RDF will show a methane-like structure at low λ-values. To see if this were the case, four further AdUmWHAM simulations were run which did not randomise λ after each iteration. The value of λ at the end of one simulation block was preserved for the next block. In all other respects, these four AdUmWHAM simulations were identical to the original four.

Simulation	Free Energy/ kcal mol^{-1}
1	9.7
2	8.0
3	7.7
4	7.9
Average	8.3 (0.8)

Table 3.5: The relative hydration free energy of water and methane as predicted by four independent AdUmWHAM simulations which preserved λ after each iteration. The average over the four results is also shown, together with the standard deviation in parenthesis.

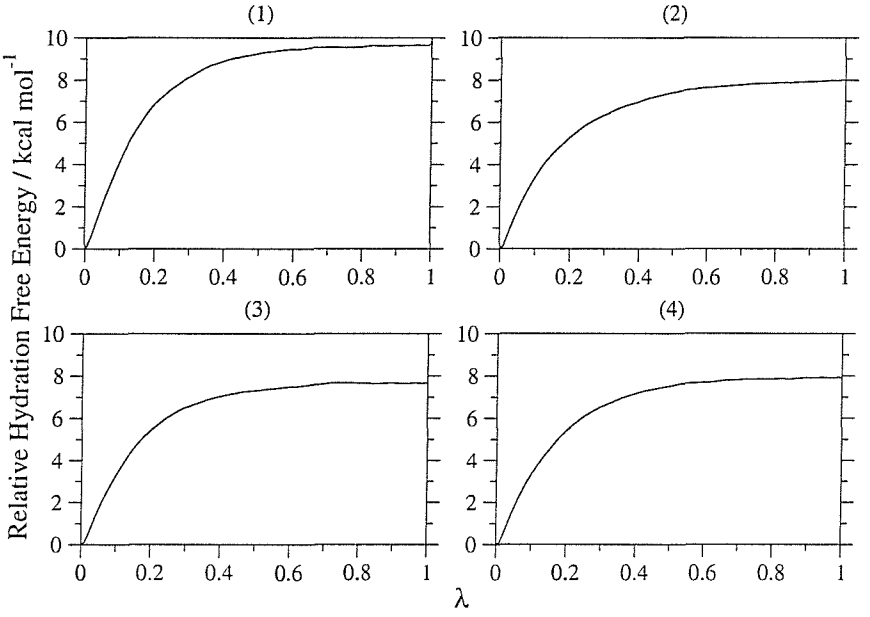


Figure 3.16: The predicted PMF of the relative hydration free energy of water and methane, as predicted by four equivalent AdUmWHAM simulations. The PMF is taken as the negative of the last iteration's umbrella, and the value of λ is preserved between each iteration.

3.6.1 Results

The results of the four new AdUmWHAM simulations are shown (table 3.5), as are the PMFs (figure 3.16). As in the previous simulations, the results were obtained from the negative of the final iteration's umbrella, and a discussion of the error on the calculation will be postponed until after a deeper analysis of the sampling.

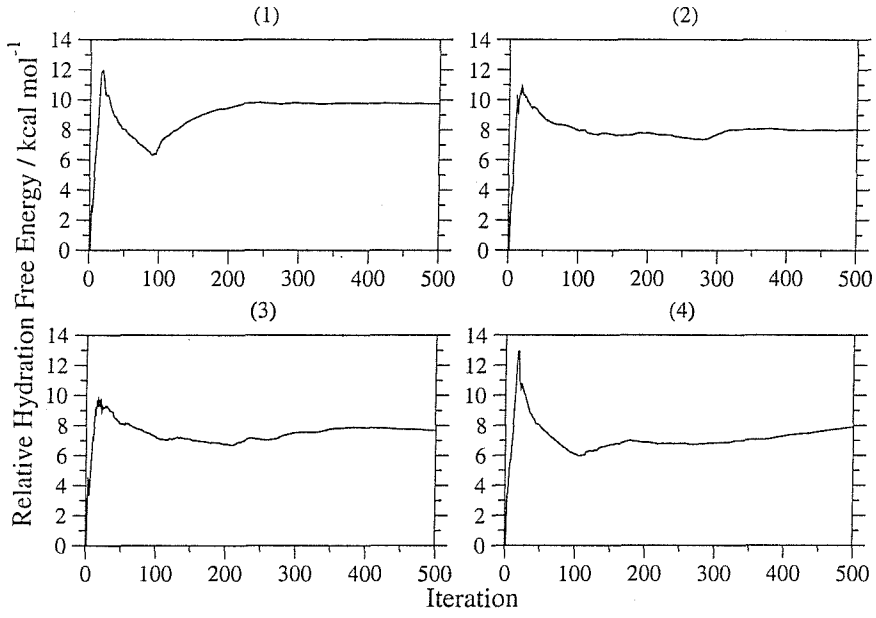
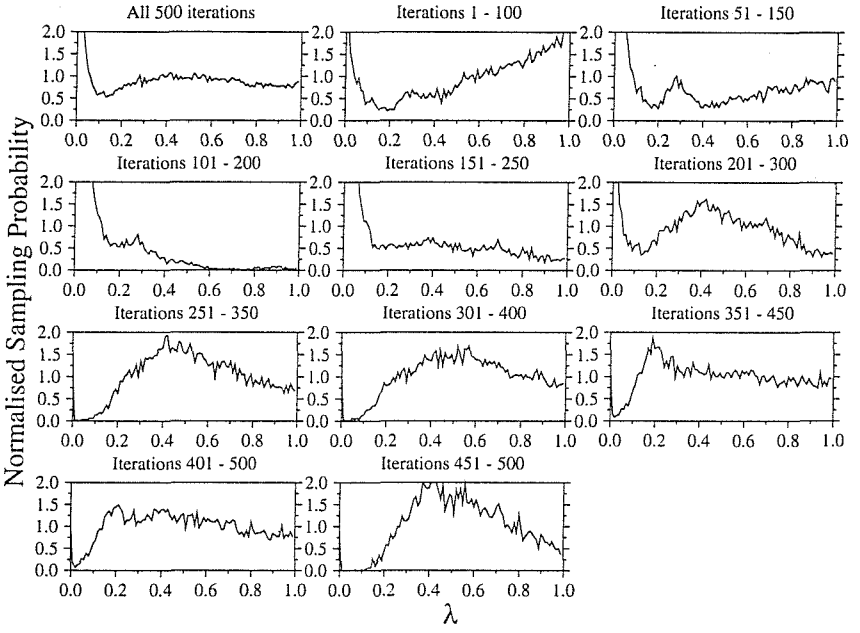


Figure 3.17: The convergence of the predicted relative hydration free energy of water and methane as a function of AdUmWHAM iteration for four independent AdUmWHAM simulations. The value of λ was preserved between iterations.

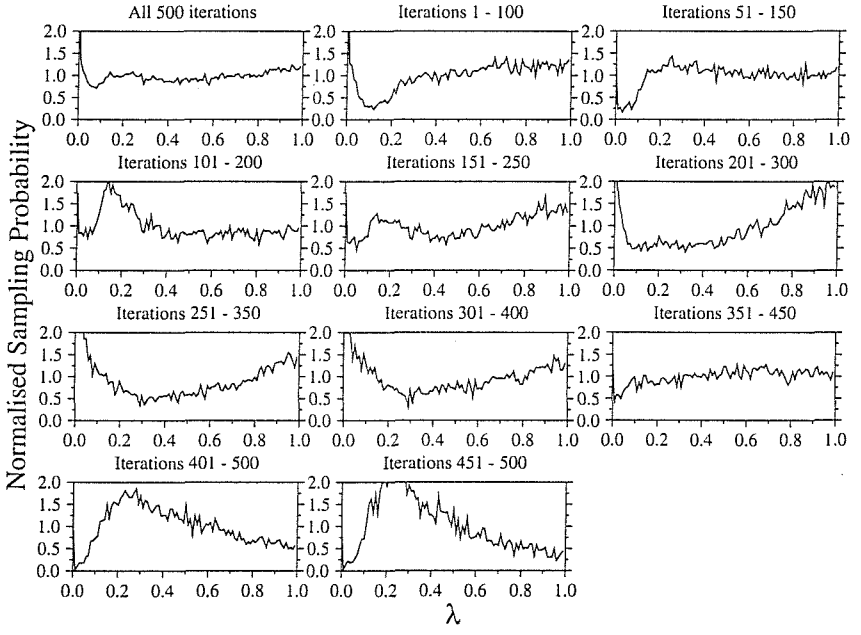
3.6.2 Analysis

The results from the four new applications of AdUmWHAM show much better agreement with experiment, FEP and FDTI. The spread of results is quite good, although the results from simulation 1 are around 2 kcal mol^{-1} higher than the other three simulations. The convergence of these calculations was plotted (figure 3.17) using the identical method as that used to obtain figure 3.14. The four simulations are seen to converge in a similar manner up to around iteration 100. At this point, the predicted result from simulation 1 rises quickly to nearly 10 kcal mol^{-1}, and stays there for the remaining 250 iterations. The predicted values from all four simulations appear stable for the last 200 to 300 iterations, implying that the umbrella has converged. To see if the simulations had truly converged, the λ-sampling was plotted over sets of iterations. The sampling for all four simulations are quite similar, so only those from simulations 1 and 2 are shown (figure 3.18). These figures show that the total λ-sampling was reasonably even. This is what the simulations should give if the umbrella is the negative of the PMF.

An explanation for the pattern of convergence of the AdUmWHAM simula-



(a) Simulation 1



(b) Simulation 2

Figure 3.18: The normalised sampling of λ for ranges of iterations, for both simulation 1 and simulation 2.

tions may be found through the comparison of figures 3.17 and 3.18. There was no umbrella at the start of the simulation. The λ-sampling was thus biased towards low values of λ, as water in water has a lower free energy than methane in water. This is apparent in the set of iterations from 1 to 100, which shows a sharp peak around $\lambda = 0.0$. As the data is processed by subsequent iterations of the WHAM equations, the umbrella is refined to encourage broader sampling of λ. The difference between the ends of the umbrella is increased to discourage sampling of low λ, and encourage sampling of high λ. This quickly works, and the sampling becomes more even by the end of the first 100 iterations. The increase in the difference between the ends of the umbrella is seen as a sharp peak in figure 3.17, at approximately 9 to 13 kcal mol^{-1}, during the first 50 iterations. Unfortunately, this is an overestimate of the free energy difference between water and methane. If the free energy difference is overestimated, then the umbrella would be biasing the λ-sampling too much, and thus the sampling would be biased towards high λ. This is seen between iterations 150 and 250 for simulation 2, and is especially acute for simulation 1, which shows a bias towards high λ before iteration 100. The iterations of the WHAM equations now seek to reduce the difference between the ends of the umbrella, and thus to correct this over-bias towards high λ. The difference between the ends of the umbrella is thus reduced, and the sampling becomes more even. However, this correction is also overestimated, and now the predicted free energy drops to as low as 6 kcal mol^{-1}. This underestimate is the result of an umbrella that is too small, and thus the sampling returns to being biased towards low λ. Continued iterations of the WHAM equations now seek to increase the magnitude of the umbrella, to correct for this overcorrection. This process continues throughout the simulation, and the umbrella "see-saws" around the correct PMF. In theory, this should mean that the umbrella potential should converge in an oscillatory fashion, as each overcorrection is corrected. However, this profile is complicated by the contribution from previous iterations. As the simulation progresses, the weight of the current iteration's data decreases compared to the weight of all previous iteration's data. The contribution of sampling from the current iteration becomes small compared to the history of sampling from the pre-

vious $n-1$ iterations. Thus the influence of the current λ-sampling statistics on the umbrella will become increasingly small. This explains why the rate of change of the predicted free energy seems to slow as the number of iterations increases. This hypothesis is reinforced by the observation that despite the changing λ-sampling throughout sets of iterations, the total λ-sampling across the entire simulation is almost completely even. This would also explain why the results from simulation 1 are so poor compared to the other three simulations. Since the λ-sampling during the initial part of simulation 1 was strongly biased to low λ, the resulting oscillations had a large amplitude. By iteration 200, the weight of total sampling was then too great to allow the following iterations to correct this overestimate, and thus simulation 1 appears to converge onto a result which is too high. The other three simulations also have this oscillatory convergence, but the amplitude of the swings are smaller, and thus by iteration 200, the predicted result is close to the actual result. Iterations 351 to 450 for simulation 2 (figure 3.18) shows sampling that is more or less even, reinforcing its predicted result of 8.0 kcal mol^{-1}. Efforts have been made to reduce the weight of the preceding iterations, either via discarding the statistics from the initial iterations, or applying the WHAM analysis to a moving window of the 50 previous iterations, though neither of these techniques proved to be successful.

Because of this oscillatory convergence, the sampling of λ does not become totally even within the 500 iterations, and thus any long simulation based on a single iteration's umbrella is unlikely to fully sample the entire λ-coordinate. This means that it is difficult to use a statistical method to estimate the error on the umbrella, since the simulation has not fully converged.

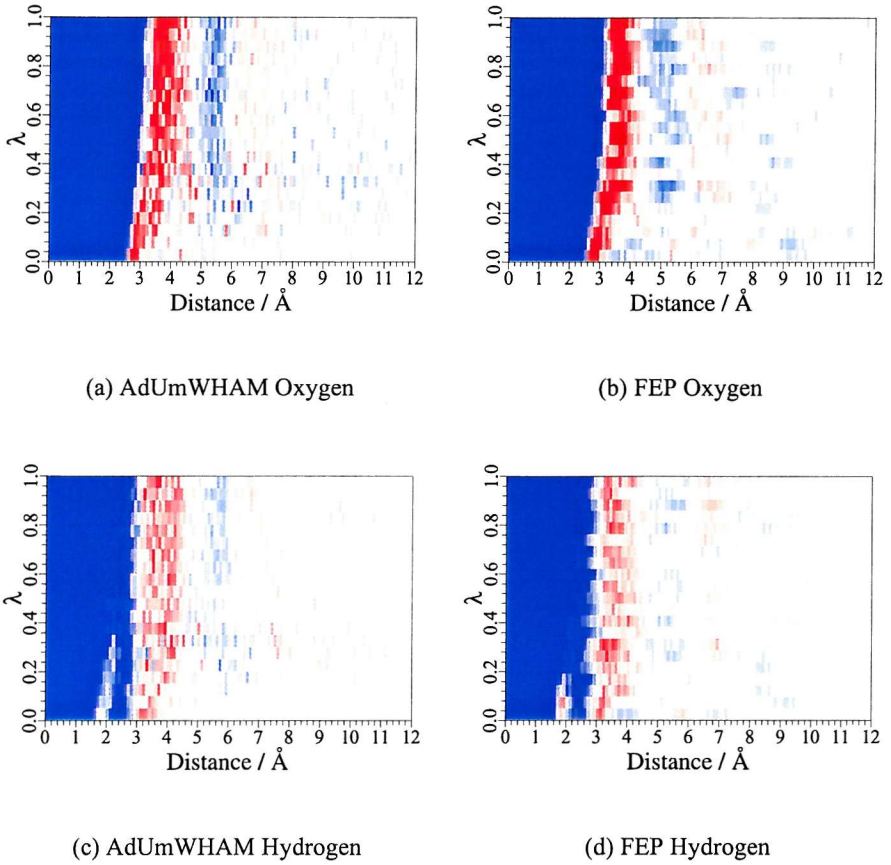
Radial Distribution Functions

The error in the protocol for the original application of AdUmWHAM to the water-methane system was exposed via plots of the radial distribution function with respect to λ (figure 3.15). Similar plots were made to see if the revised protocol gave better results. These were all very similar, so only those from simulation 1 are shown (figure 3.19). These plots show much better agreement with those for FEP and FDTI (figure 3.6), and clearly show the double-peak in the hydrogen RDF.

The second peak in the hydrogen RDF is also much sharper than in the RDFs from the original protocol (figure 3.15). However, the double peak in the hydrogen RDF persists at $\lambda = 0.3$, while it disappears more quickly in the RDFs from FEP, at around $\lambda = 0.2$. This difference hints that there may still be a problem with the solvent not responding sufficiently quickly to the solute as it changes between water and methane. These results may be improved by reducing the frequency of λ-moves from one every 500 MC steps to perhaps one every 1000. However, this would increase the overall run-time of the simulation, which is already five times that of FEP or FDTI. Also, it is difficult to know in advance what the relaxation time of the system with respect to λ-moves would be, without first running exploratory simulations. A blind application can thus only use the lowest possible frequency of λ-moves that will collect sufficient data to run the simulation in a reasonable time.

3.6.3 Summary

AdUmWHAM was applied to calculate the relative hydration free energy of water and methane. The initial application of the method failed, due to the randomisation of λ between iterations. A further application which preserved λ between iterations was successful, and predicted results that were in agreement with experiment, FEP and FDTI. However, the AdUmWHAM simulations failed to converge properly, and while they estimated the free energy well near the start of the iteration, the weight of statistics collected at the beginning reduced the rate of change of the umbrella, and hence reduced the rate of convergence. The method appears very sensitive to sampling in the early stages of the simulation, and any peculiarities there, as occurred in the case of simulation 1, can destabilise the oscillatory convergence, and lead to poor quality results. In addition, the difference in the hydrogen RDF between AdUmWHAM and FEP implies that the solvent may not be responding sufficiently quickly to the dynamically changing λ-value. While this problem may be overcome through the use of less frequent λ-moves, this would increase the time required to run a simulation which already takes five times longer than FEP or FDTI. In addition, it is not possible for a blind application of AdUmWHAM to know the relaxation time of the system with respect to λ-moves,



(a) AdUmWHAM Oxygen

(b) FEP Oxygen

(c) AdUmWHAM Hydrogen

(d) FEP Hydrogen

Figure 3.19: The oxygen and hydrogen RDFs calculated from the first FEP and first modified AdUmWHAM simulation. The colour scale is the same as in figure 3.7.

and thus know in advance whether λ would be changing too quickly. Despite these problems, AdUmWHAM still managed to produce results in agreement with the other methods, and it produced a good estimate of the results within the first fifth of the simulation. This ability to give a good, yet quick estimate of the free energy could perhaps be exploited through the use of AdUmWHAM as a free energy scoring function.

3.7 Conclusion

Free Energy Perturbation, Finite Difference Thermodynamic Integration and Adaptive Umbrella WHAM have been applied to calculate the relative hydration free energy of water and methane. While the results from all three methods gave good agreement with each other and experiment, they all suffered from poor precision.

There was a large spread of results within each method. In the case of FEP and FDTI this was caused by random sampling error due to the use of only a single trajectory at each λ-value. In the case of AdUmWHAM, which allows multiple trajectories to visit each λ-value, the variability appears to be a result of poor convergence in the latter half of the simulation due to the weight of statistics in the initial stages. There was also the suggestion of Hamiltonian lag, due to the change in the value of λ throughout the trajectory. The results presented in this chapter can be summarised via a comparison of the advantages and disadvantages of each method;

1. Free Energy Perturbation

   (a) Advantages

      i. Calculated free energy agrees with the experimental value.

      ii. Use of a fixed λ-value minimises problems due to Hamiltonian lag.

   (b) Disadvantages

      i. The quality of the results was reduced by problems converging the Zwanzig equation. This resulted in significant hysteresis.

      ii. The use of a single trajectory at each λ-value led to random sampling errors.

      iii. Improvement in the quality of the results requires the rerunning of windows with different widths. The original data for these windows would be discarded.

2. Finite Difference Thermodynamic Integration

   (a) Advantages

      i. Has the same advantages as FEP.

      ii. Additionally, the use of a small $\Delta\lambda$ leads to small fluctuations in the difference in energy between the reference and perturbed states, and thus the calculation has very small hysteresis.

iii. Improvement in the quality of the results may be achieved by calculating the free energy gradient at extra λ-values. No previously collected data needs to be discarded.

(b) Disadvantages

i. Again, the use of a single trajectory at each λ-value leads to random sampling errors.

3. Adaptive Umbrella WHAM

(a) Advantages

i. Calculated free energy agrees with the experimental value.

ii. The relative free energy is quickly estimated. This could lead to use of the method as a simple free energy score.

iii. The use of multiple trajectories at each λ-value helps to reduce the random sampling error seen in FEP and FDTI.

(b) Disadvantages

i. The method fails to fully converge in a reasonable time on the water-methane system.

ii. While the number of simulation steps is lower than FEP or FDTI, the method is not as parallisable, so the total simulation time is much longer.

iii. The changing value of λ may lead to errors resulting from Hamiltonian lag.

The advantages of Thermodynamic Integration over FEP have been noted in previous studies.[90,94] These studies compared the two methods on a variety of small test systems, and conclude that while FEP and TI produce comparable results for equivalent amounts of simulation, if the overlap between the two systems is poor, then the quality of the FEP results is reduced compared to those from FDTI. The results presented in this chapter demonstrates that FDTI has the same advantages over FEP as TI. The benefit of FDTI is that it may be performed in any FEP-capable simulation code, without the need to program the modifications necessary to generate the analytical gradients required for TI.

In conclusion, the results presented in this chapter demonstrate the problems faced by the application of established free energy methods to the calculation of the relative hydration free energy of water and methane. The root cause of these problems is that the water-methane system is deceptively complicated, with the hydration free energy change strongly depending on a significant solvent rearrangement. It can thus be concluded from these results that the main drawback of current free energy methods is their reliability. Research must now be undertaken to improve the consistency of results, and in so doing, reduce the statistical errors on the calculations. These results suggest that the route to achieving this aim is to generally improve the sampling of the system, either by removing random sampling error through the correct combination of multiple trajectories, or through increasing the rate or quality of sampling.

Chapter 4

The Development of Advanced Free Energy Methods

4.1 Introduction

The aim of this work is to develop free energy methods that may be reliably and routinely applied to calculate the relative binding free energies of a group of ligands to a protein. In the last chapter, established free energy methods (Free Energy Perturbation, Finite Difference Thermodynamic Integration and Adaptive Umbrella WHAM) were tested through the calculation of the relative hydration free energy of water and methane. This proved to be a challenging test case, as the perturbation involved the elimination of the solute dipole, and the significant re-arrangement of the solvent configuration. All three free energy methods predicted relative hydration free energies that were, within error, in agreement with each other and with the experimental value of 8.31 kcal mol^{-1}.[87] However, it was found that repeated application of the same method on the same system could sometimes yield significantly different results, and thus the precision of the methods was a major problem. Significant random sampling error was present in the free energies calculated via FEP and FDTI. This was a result of the use of only a single trajectory at each λ-value. This error could be estimated and reduced through the running of multiple FEP or FDTI simulations, though this of course increases the expense of the methods. AdUmWHAM avoids this problem as it uses a dynamic λ-

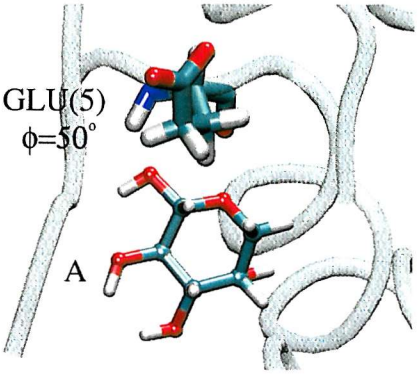
coordinate, and thus effectively allows multiple trajectories to visit each λ-value. While AdUmWHAM enhanced sampling of λ, it did not enhance the sampling of the solvent. The solvent's significant configurational change was thus insufficiently relaxed as the value of λ changed and the solute was dynamically morphed back and forth between water and methane. This resulted in Hamiltonian lag, and a possible error in the results. In addition, the rate of convergence of the method seemed to slow as the iteration progressed, due to the weight of statistics already accumulated. This meant that any peculiarities in the initial convergence of the calculation were not corrected, and that one of the four simulations failed to converge in a reasonable time.

The aim of this chapter is to take the lessons learned from the application of the established free energy methods and use them to develop methods which are more reliable and reproducible. To achieve this, sampling error will need to be reduced through the correct averaging of multiple simulation trajectories. In addition, the rate of sampling of the entire system, as well as λ, needs to be improved.

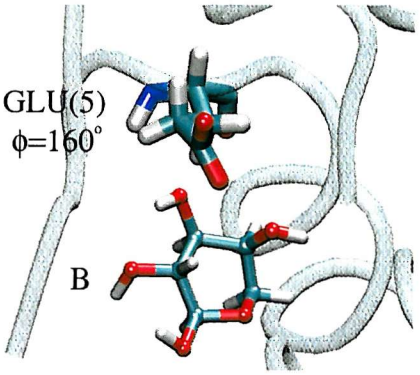
4.2 Bivariate Multicanonical WHAM

Adaptive Umbrella WHAM solves one of the two main problems of current free energy methods, namely it reduces random sampling error by treating the perturbing λ-coordinate as a normal degree of freedom. λ is changed dynamically throughout the simulation, and the sampling of λ is enhanced through an umbrella potential. The umbrella potential is created through a sequence of iterations, throughout which multiple trajectories refine the umbrella potential until it encourages even sampling of the λ-coordinate. At this point, the umbrella potential is equal to the negative of the potential of mean force (PMF), and the relative free energy is simply read from the difference in energy between the umbrella's endpoints. While AdUmWHAM successfully enhances the sampling of the λ-coordinate, it does not enhance the sampling of the rest of the system. It does nothing to enhance the sampling of an associated configurational change, like for example the configurational change in the solvent between hydrated water and hydrated methane. AdUmWHAM relies on the rate of sampling of the system be-

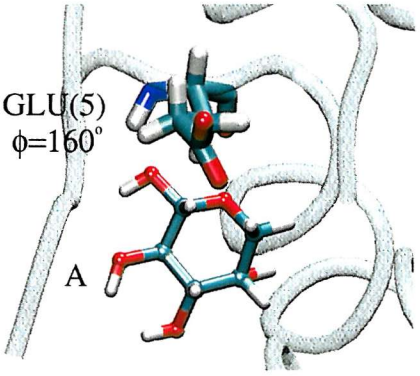
ing sufficiently high such that any associated configurational changes can occur. Unfortunately, if the configurational change is large, then the rate of sampling of that change will be slow, and the umbrella can have the effect of driving the λ-sampling through unfavourable system configurations. While such short-cuts may lead to a converged AdUmWHAM simulation, and an even λ-distribution, they are also likely to put an unquantifiable error into the free energy calculation. To better illustrate this point, let us consider a simulation that aims to calculate the relative binding free energy of two binding modes, *A* and *B*, of the same ligand to a protein (figure 4.1). The relative binding free energy in this case could be calculated through the use of a dual topology model.[95] Two non-interacting copies of the ligand are placed in the active site of the protein, one in each binding mode. The interactions between each ligand and the rest of the system are scaled as a function of λ. At $\lambda = 0.0$, the ligand in binding mode *A* is fully on, and the ligand in binding mode *B* is fully off, while at $\lambda = 1.0$, the situation is reversed, and *B* is fully on, while *A* is fully off. In this way, the motion along λ acts to switch between the two binding modes. The protein configuration may be slightly different for the two binding modes. For example, the value of the dihedral angle of the GLU(5) residue is 50 ° for binding mode *A*, and 160 ° for binding mode *B* (figure 4.1). This dihedral angle has a high barrier to rotation, so it is unlikely to be properly sampled during the course of the AdUmWHAM simulation. The dihedral angle will thus not change conformation readily as λ changes the ligand between the two binding modes. If the dihedral angle starts in the *A* conformation, then since the umbrella potential can only directly affect the sampling along λ, it will keep pushing the sampling of λ until the ligand is in the *B* binding mode, regardless of the value of this important dihedral angle. It may be the case that the change in the dihedral angle is sampled so poorly that the umbrella pushes binding mode *A* all the way to binding mode *B*, while the dihedral angle is still in conformation *A*. This example represents an extreme form of Hamiltonian lag, where the relaxation time of the system is significantly longer than the rate of motion along λ. To solve this problem, a new method must be developed which not only enhances the sampling of λ, but also enhances the sampling of the rest of the system, and in so doing, improves the relaxation time of such important configurational changes.



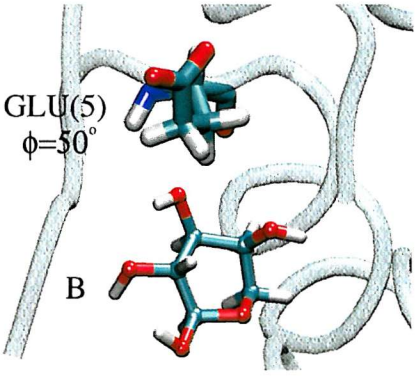
(a) Binding mode *A* with correct dihedral conformation.



(b) Binding mode *B* with correct dihedral conformation.



(c) Binding mode *A* with incorrect dihedral conformation.



(d) Binding mode *B* with incorrect dihedral conformation.

Figure 4.1: (a) and (b) show the two different binding modes of the ligand to the protein. The dihedral angle of the GLU(5) residue has a different conformation for the two binding modes ($\phi = 50\ ^\circ$ for binding mode *A*, and $\phi = 160\ ^\circ$ for binding mode *B*). This dihedral angle is poorly sampled, thus the effect of the umbrella is to drive λ-sampling such that in (c) and (d), the binding modes may see the incorrect conformation.

4.2.1 Generalised Ensembles

The rate of configurational sampling within a system depends on the ensemble conditions for that system, e.g. there will be more sampling for a higher temperature ensemble than for that of a lower temperature. However, the free energies produced depend on the ensemble conditions, and thus the results from high temperature simulations are irrelevant if the interest is in ambient temperature behaviour. *Generalised* or *extended* ensembles have been developed to overcome this problem. They are reviewed in detail by Iba,[96] and comparative simulations have been

performed by Hansmann and Okamoto.[97,98] One such generalised ensemble is the *multicanonical ensemble*.[99]

4.2.2 Multicanonical Ensemble

The multicanonical ensemble is a generalised ensemble over energy.[96] This means that the algorithm attempts to form an ensemble where there is even sampling of potential energy. Once this is achieved, there is a random walk over the energy coordinate, and all potential energy barriers are removed. At this limit, it is possible to re-weight the probabilities to obtain the canonical ensemble for a wide range of temperatures.[98] The method is applied iteratively.[100]

1. A conventional Monte Carlo simulation is run, and the energies sampled are saved in a histogram. This histogram can be used to estimate the probability of sampling each energy, $P_0(E)$, from which the *density of states*, $n_0(E)$ can be estimated via,[100]

$$n_0(E) = P_0(E)\exp(E/k_BT). \tag{4.1}$$

   The temperature for this simulation is set sufficiently high such that the system can overcome the potential energy barriers of interest. This temperature can be found through a process of trial and error.

2. Now the multicanonical simulation starts. This uses a modified Monte Carlo acceptance test, which evaluates the energy of the new configuration, E_{new}, versus the old configuration, E_{old}, based on the estimated density of states of two energies.

$$\frac{n_0(E_{old})}{n_0(E_{new})} \geq rand(0,1) \tag{4.2}$$

   This test has the effect of pushing the sampling to regions of energy that are comparatively under-sampled. This test is derived in appendix C.2. The energies sampled during the run are again stored in a histogram, which is used to estimate the new probability of sampling each energy, $P_1(E)$.

3. A new estimate of the density of states, $n_1(E)$ is made from the probability estimate from the previous run, and the previous estimate of the density of

states, $n_0(E)$,

$$n_1(E) = P_1(E)n_0(E). \tag{4.3}$$

4. A new multicanonical simulation is run, using $n_1(E)$ for the acceptance test. This cycle is then repeated until the density of states converges, i.e. $n_i(E) = n_{i-1}(E)$.

5. Once the density of states has converged, the ensemble can be re-weighted to obtain the probability of an energy, $P_T(E)$, at a particular temperature, T, via,[100]

$$P_T(E) \propto n(E)\exp(-E/k_BT) \tag{4.4}$$

This method was first developed to look at phase transitions,[99] though it has since been applied to peptide-protein docking,[101] investigation of the interaction between an amino-acid side chain and a DNA base pair,[102] and peptide configurational sampling.[103–105] The method can also be extended into multiple dimensions, as in the technique of bivariate multicanonical sampling.[100] This works in an identical manner to normal multicanonical, with the additional collection of statistics along a second reaction coordinate, r. In this way, the probability density along both E and r can be histogrammed, $P_0(E,r)$, and used to obtain a two-dimensional density of states, $n_0(E,r)$. The acceptance test is then based on the new density of states, and the simulations are run until it has converged in both dimensions.

4.2.3 Comparison with AdUmWHAM

Multicanonical sampling is very similar to adaptive umbrella sampling, in that the multicanonical method aims to sample the potential energy evenly, while adaptive umbrella sampling aims to sample an arbitrary reaction coordinate evenly. This similarity was first realised by Bartels and Karplus[106] who showed that adaptive umbrella sampling over potential energy was equivalent to multicanonical sampling.[106] They went on to develop multidimensional adaptive umbrella sampling over potential energy and another reaction coordinate,[63] which is equivalent to bivariate multicanonical sampling. The equivalence of the two methods is shown in

appendix C.2. The protocol for umbrella sampling version of the multicanonical simulation is as follows;

1. An initial simulation is run at the temperature that is of interest for the study, e.g. room temperature.

2. The energies, E, and the values of the reaction coordinate, r, which are sampled during this simulation, are stored in a two-dimensional histogram.

3. The probability of being in a particular histogram bin, $p(E,r)$, is estimated, and used to obtain an umbrella potential, $U(E,r)$, over energy and r,

$$U(E,r) = k_B T \ln(p(E,r)) \tag{4.5}$$

4. As in the case of AdUmWHAM, the value of the umbrella where there was no sampling is obtained through extrapolation,[53,63] by setting their values to the smallest umbrella potential that has been sampled. This results in some discontinuities, so the umbrella has to be smoothed. This is achieved by replacing the umbrella potential for bin (i,j), $U(i,j)$, with,[53]

$$\begin{aligned} U(i,j) = \frac{1}{3}\Big(& -0.3U(i-2,j) + 1.3U(i-1,j) \\ & + U(i,j) + 1.3U(i+1,j) - 0.3U(i+2,j) \Big). \end{aligned} \tag{4.6}$$

   This smoothing is applied sequentially in strips to each dimension. It has been shown that this smoothing works best when it is applied twice.[53]

5. The smoothed umbrella is used to bias a new umbrella sampling simulation, again run at the temperature of interest. The sampling along energy and r is recorded in another histogram.

6. At the end of the iteration, the sampling is unbiased and combined with all previous statistics via the WHAM equations, to obtain a new estimate of the umbrella potential, $U(E,r)$.

7. This cycle is repeated until the umbrella has converged, and even sampling along energy and r has been obtained.

8. At this point, the umbrella that would be needed to obtain even sampling of r for a normal AdUmWHAM simulation at temperature T, $U_T(r)$, can be obtained from the probability, $p(r)$, of each point along r. This is calculated from the probability, $p(E,r)$, of each of the histogram bins along energy and r, which is obtained from the final umbrella potential, $U(E,r)$, across the histogram,

$$p(E,r) = \exp(U(E,r)/k_B T_{sim})$$

where T_{sim} is the temperature that the simulation was run at,

$$p(r) = \sum_E p(E,r) \tag{4.7}$$

$$U_T(r) = k_B T \ln(p(r)).$$

This umbrella is the negative of the PMF along r at temperature T, which can be used to obtain relative free energies in an identical way to AdUmWHAM. Note that the temperature for this umbrella, T, does not have to be the same as the temperature of the umbrella simulations, T_{sim}. As in the case of standard multicanonical, it is possible to reweight the statistics to a range of temperatures. It is our experience that the appearance of exponential terms in these equations means that this reweighting is only practical for a small range of temperatures, perhaps 5 to 10 °C, around T_{sim}.

The AdUmWHAM versions of the multicanonical algorithm have been applied to configurational sampling of small systems.[63,106] They have not, to our knowledge, been applied in the field of free energy calculations, though such application represents the trivial extension of either bivariate multicanonical or multidimensional adaptive umbrella sampling to using λ as the second reaction coordinate. To simplify nomenclature, and to recognise the fact that multidimensional adaptive umbrella sampling over potential energy and another reaction coordinate is equivalent to bivariate multicanonical sampling, the application of this method over potential energy and λ shall be referred to as *Bivariate Multicanonical WHAM* (BMW). While BMW and bivariate multicanonical are equivalent, and should converge on the same result, we believe the form of convergence of the two methods will be

very different. In the case of bivariate multicanonical, the use of the modified acceptance test should mean that there will be an initial random walk in energy space. This random walk will occur in an energy range between the global minimum of the system and the energy range of the initial, high temperature simulation. The sampling should thus fill this range in a reasonably even manner. In contrast, BMW starts off as a normal NVT or NPT simulation, running at the desired simulation temperature. This means that the simulation will initially sample from the energy range for that temperature. As the umbrella potential is refined, it has the effect of discouraging the sampling of energy-λ regions that have already been sampled. The umbrella potential acts to push the sampling to regions outside that for the desired simulation temperature. Since the MC test is automatically accepted for changes which reduce the total energy, the umbrella will effectively push the sampling down to an energy range below that which is expected for the simulation temperature. Only when the umbrella has been refined for all lower temperatures will it then act to push the sampling above the expected energy range. In this sense, we expect BMW and bivariate multicanonical to converge in opposite directions, i.e. BMW will converge from low energies up to high energies, while bivariate multicanonical will converge from high energies down to low energies.

4.2.4 Simulation Conditions

The application of AdUmWHAM to the calculation of the relative hydration free energy of water and methane was described in the previous chapter. This method failed to sample the solvent configurational change correctly as the solute was rapidly morphed between the water and methane states. The aim of this section is to use BMW to calculate the relative hydration free energy, and examine whether this multicanonical extension of the AdUmWHAM method is able to overcome the problems presented by slow system configurational changes. To allow comparison with the AdUmWHAM simulations in section 3.5, the simulation protocol was kept as similar as possible. Thus an identical system and set of starting coordinates were used, and 500 WHAM iterations were performed in an asynchronous, parallel fashion over two independent trajectories. Each iteration consisted of 200 K steps of simulation, split into 40 K steps of equilibration, and 160 K steps of data

collection. Each trajectory preserved the value of λ between iterations. The simulations were run in the NPT ensemble, at a temperature of 25 °C, and a pressure of 1 atm. The λ-coordinate was split up into 100 histogram bins, of width 0.01. Investigation of the energies sampled during the FEP and FDTI simulations from the previous chapter suggested that the energy range sampled at 25 °C was between -16400 kcal mol^{-1} and -16200 kcal mol^{-1}. Owing to the limitations of computer memory, this energy range could only be spanned by 100 bins, as the WHAM equations required the storage of each of the 10 000 sampling and umbrella histograms produced during the simulation. The iterative solution to the WHAM equations was implemented within a Perl script. This script required over 500 megabytes of memory to store all the information necessary to process the 500 iterations. The use of a compiled language, such as C or Fortran, would reduce these memory requirements. Such languages, however, are not best suited to the additional roles of the Perl script, namely the management of the submission and processing of each of the individual iterations. Owing to the limited number of bins along the energy coordinate, different energy ranges and resolutions were investigated. To check for reliability, some of these ranges were simulated multiple times.

It is important to restrict the range of energy sampling to fall within the energy-λ histogram. This is because the umbrella implementation of the multicanonical ensemble will cause the system to seek out the global minimum on the energy surface, and thus the energy sampling would be pushed far below the values expected at the desired simulation temperature.[106] In the case of the water-methane system, the umbrella potential could be used to freeze the water into ice. To prevent this from happening, some workers set the umbrella potential to zero outside the range of the energy-λ histogram.[106] Our initial investigations using this method have instead led us to use an harmonic potential that penalises the system when it falls outside this specified range. These same investigations have also suggested that the smoothing of the umbrella should be performed over more than just the neighbouring 4 bins. We have thus implemented a smoothing function that works over

the neighbouring 10 bins,

$$\begin{aligned} U(i,j) =& 0.0222U(i-5,j)+0.0455U(i-4,j)+0.0798U(i-3,j) \\ &+0.1191U(i-2,j)+0.1513U(i-1,j)+0.1640U(i,j) \\ &+0.1513U(i+1,j)+0.1191U(i+2,j)+0.0798U(i+3,j) \\ &+0.0455U(i+4,j)+0.0222U(i+5,j). \end{aligned} \tag{4.8}$$

The coefficients to this equation, for the position a bins away from the central bin, are calculated via a Gaussian kernel,

$$c(a) = \exp\left(\frac{-a^2}{0.5 \times n^2}\right), \tag{4.9}$$

where n is the number of bins to the right of the central bin, which in this case is 5. The coefficients are then normalised by dividing by their sum. This is just a linear filter[46] using a gaussian function with $\sigma = n$ to obtain the individual coefficients.

4.2.5 Results

To investigate the effect of the energy range on the results, multiple energy ranges and resolutions were tested. Some of these ranges were repeated multiple times to check for reliability. The predicted relative hydration free energies, as calculated from the final umbrella at 25 °C via equation 4.7, are shown in table 4.1.

4.2.6 Analysis

These results are disappointing, with a range of predicted free energies between 4.1 and 8.0 kcal mol^{-1}. What makes the result worse, is that the spread of results within a set of identical conditions (e.g. set 1 or set 5) is equally poor. Indeed, the variability of results is so large that it is not possible to derive any pattern with the changing energy histogram resolution or size. Despite the incorrect free energies, the potentials of mean force do have approximately the correct shape. Those from set 5 are shown (figure 4.2).

Set	Bin Width / kcal mol^{-1}	Energy Range / kcal mol^{-1}	Relative hydration free energy / kcal mol^{-1}
1	2.0	-16600 to -16400	4.1, 6.9, 5.8, 6.5
2	2.0	-16400 to -16200	8.0, 6.9
3	2.0	-16300 to -16100	6.3
4	2.0	-16200 to -16000	7.5
5	1.0	-16400 to -16300	6.2, 6.7, 7.1, 5.6
6	1.0	-16350 to -16250	7.1
7	1.0	-16300 to -16200	7.5
8	0.5	-16400 to -16350	7.2
9	0.5	-16350 to -16300	7.3
10	0.5	-16300 to -16250	4.9

Table 4.1: Effect of energy histogram parameters on the predicted relative hydration free energy of water and methane.

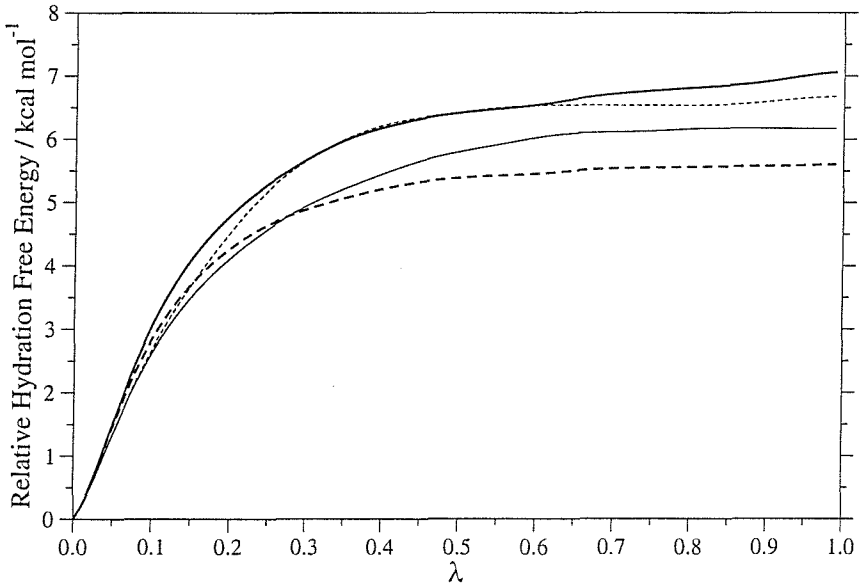


Figure 4.2: The predicted PMFs for four applications of BMW from set 5, to the calculation of the relative hydration free energy of methane. (1) Solid line, (2) dashed line, (3) thick solid line, (4) thick dashed line.

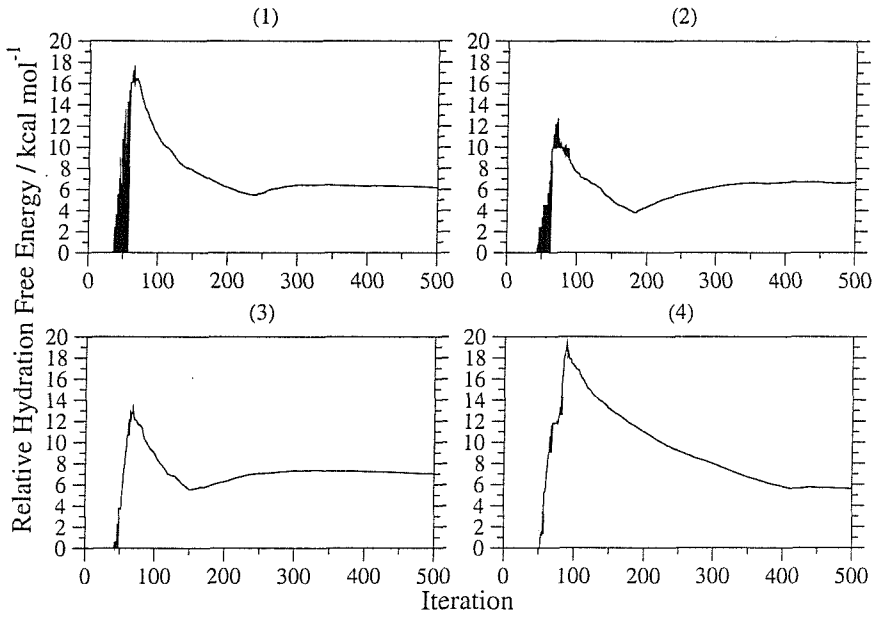
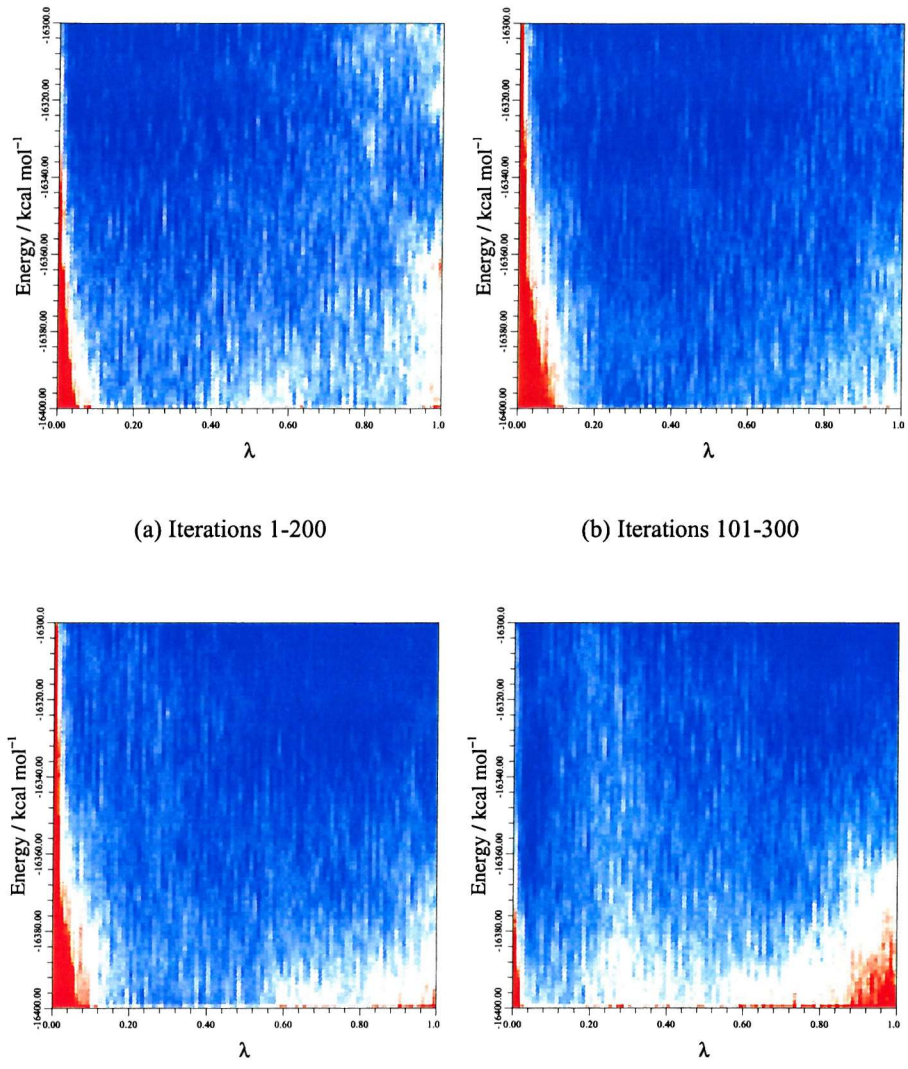


Figure 4.3: The convergence of the predicted relative hydration free energy of water and methane as a function of BMW iteration, for four applications of BMW from set 5.

Sampling and Convergence

The results of the application of BMW to calculate the relative hydration free energy of water and methane are very poor. Figure 4.3 shows the convergence of four simulations from set 5, as a function of WHAM iteration. The convergence of these simulations show the same main features as those using AdUmWHAM (figure 3.17), namely a spike in the estimated free energy near the beginning of the simulation, followed by a slow fall to the converged value. The main difference between the convergence of BMW and the convergence of AdUmWHAM is that BMW takes many more iterations to follow this pattern. While AdUmWHAM spikes at about iteration 25, and is converging around iteration 200, BMW spikes around iterations 75 to 100, and does not converge properly until iterations 200 to 300. Simulation 4 does not appear to converge until iteration 400. It was expected that BMW would take longer to converge than AdUmWHAM, as the umbrella has an extra dimension, and should thus require more sampling. The slow rate of convergence leads to the suspicion that despite the results appearing to have converged, as was the case of AdUmWHAM, the underlying sampling has not yet become even. Investigation of the sampling demonstrates that this was indeed the



(a) Iterations 1-200

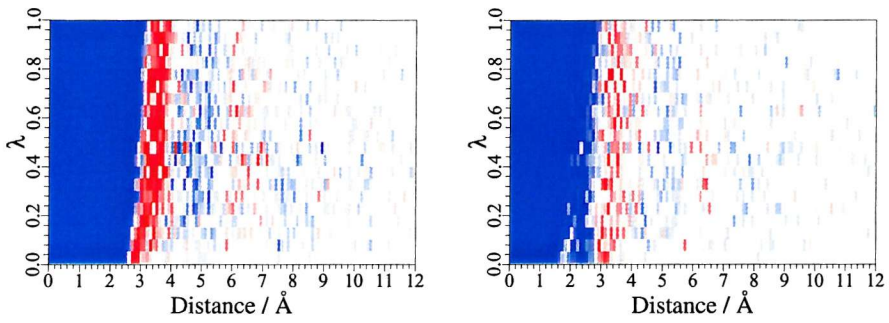
(b) Iterations 101-300

(c) Iterations 201-400

(d) Iterations 301-500

Figure 4.4: The total sampling over windows of iterations for the third BMW simulation from set 4. The colour scale runs from no sampling (blue), though average sampling (white), up to high sampling (red).

case. Figure 4.4 shows the λ-energy sampling for windows of iterations from simulation 3 of set 5. The sampling is clearly not uniform, and is concentrated at the bottom of the energy range. The simulation does go some way to pushing the sampling over the entire λ-coordinate, though more simulation would be needed before it has properly converged. Unfortunately, the slow rate of convergence, the huge memory requirements, and the increasing processor overhead to actually solving the WHAM equations after iteration 100 to 200, makes the running of yet more



(a) Oxygen (b) Hydrogen

Figure 4.5: The RDF as a function of λ for BMW simulation 3, set 5. The scale runs from blue (0.0), through white (1.0) up to red (greater than 2.0).

iterations completely impractical.

To see if the use of a generalised ensemble has improved the problem of Hamiltonian lag encountered during the AdUmWHAM simulations, the RDFs with respect to λ were plotted. Figure 4.5 shows the oxygen and hydrogen RDFs for simulation 3 of set 5. Compared to the RDFs from FEP and AdUmWHAM (figure 3.19), these RDFs are poorly defined, and show definite signs of Hamiltonian lag. The double-peak in the hydrogen RDF is visible up to $\lambda = 0.4$ compared to $\lambda = 0.3$ for AdUmWHAM, and $\lambda = 0.2$ for FEP. Since the umbrella has not converged, the random walk in energy space promised by the multicanonical ensemble has not occurred, and thus it has not been able to solve this fundamental problem of AdUmWHAM.

4.2.7 Summary

A new technique, Bivariate Multicanonical WHAM, formed as a synthesis between the bivariate multicanonical ensemble, and adaptive umbrella WHAM, has been developed and applied to calculate the relative hydration free energy of water and methane. Unfortunately, this method performed poorly on the water-methane system. The large memory and processor requirements of the method prevented it from being run to convergence. A full multicanonical ensemble was thus not achieved, and the method failed to overcome the problems of Hamiltonian lag that were apparent in the AdUmWHAM results.

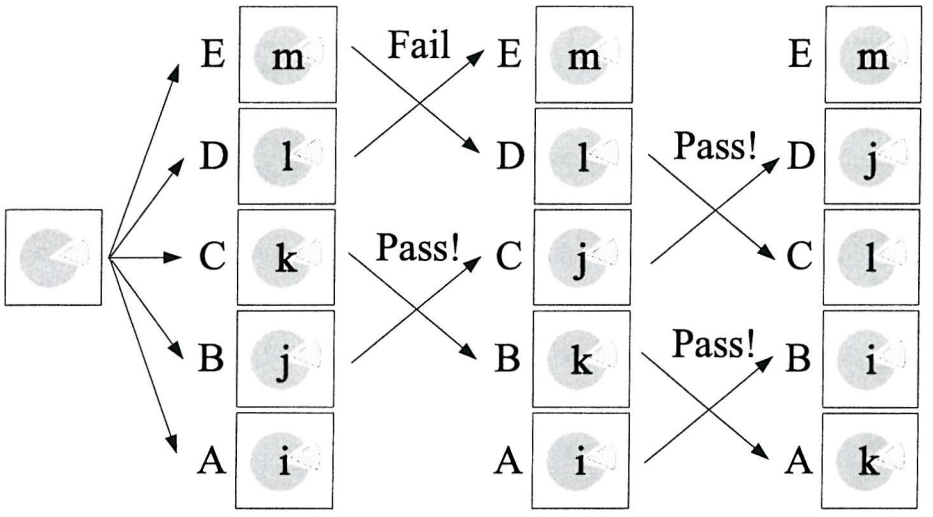
4.3 Parallel Tempering

The two main problem areas apparent from the application of the established free energy methods were the consistency of the results, and the range and speed of configurational sampling. BMW attempted to solve these problems through the use of AdUmWHAM's multiple trajectories to average out random sampling error, and the multicanonical ensemble to enhance general system sampling and overcome problems of Hamiltonian lag. BMW failed, as it was unable to converge on the challenging water-methane test system. This was because of the inherent convergence problems of AdUmWHAM, and the large CPU and memory requirements needed to store and process the two dimensional umbrellas.

The main idea behind BMW, the use of generalised ensembles to enhance configurational sampling, is a good one. A method could be developed to apply a generalised ensemble in the context of an FEP- or FDTI-like simulation. One possible candidate would be the combination of FDTI with the generalised ensemble known as *Parallel Tempering*[107] (PT).

4.3.1 Background

Parallel Tempering[107,108] forms a generalised ensemble over temperature.[96] That is to say that temperature is treated as a dynamic variable, and an ensemble is constructed that samples over many temperatures. At the end of the simulation, the probabilities can be re-weighted to obtain the correct statistics for the desired temperature. Parallel tempering forms this generalised ensemble through the running in parallel of multiple replicas of the system (figure 4.6). Each replica has a different temperature, within which normal NVT or NPT sampling is performed. Periodically, neighbouring replicas are tested according to a special Monte Carlo test. If this test is passed, then the coordinates of the two replicas are swapped. The sampling then continues until convergence. The major benefit of this technique is that while a generalised ensemble has been formed over the complete set of replicas, the complete set of configurations at each temperature still forms a correct NVT or NPT ensemble (figure 4.6). There is thus no need to perform an iteration to find the relative weights of the trajectories, which was the case for



1
Get original system. Create replicas of that system (i to m) and distribute over a set of temperatures (A to E). Perform normal NVT or NPT sampling for each replica for each temperature in parallel.

2
Perform a *Parallel Tempering* test between neighbouring replicas. Swap the coordinates of pairs that pass. Do nothing with the pairs that fail. Then run some more NVT or NPT sampling.

3
Keep repeating this test / sampling cycle until the trajectories have converged. At this point, the set of coordinates for each temperature (e.g. 'i, i, k...' for temperature 'A', or 'j, k, i...' for temperature 'B') form a correct ensemble distribution for that temperature.

Figure 4.6: The Parallel Tempering algorithm.

AdUmWHAM or BMW, as the individual trajectories are already correctly Boltzmann weighted. The implementation of PT is also very simple, as the exchange of coordinates between two temperatures is equivalent to the exchange of temperatures between two trajectories. The method is thus well-suited to application over multiple nodes of a Beowulf cluster, as the only information that needs to be exchanged between nodes is the temperatures and final energies of each replica.

The parallel tempering test has different forms depending on the type of the individual sub-ensembles. For example, if replica i, with temperature A, was swapping with replica j, with temperature B, then the test for a set of NVT ensembles would be,[107]

$$\exp\left[\left(\frac{1}{k_B T_B} - \frac{1}{k_B T_A}\right)\left(E(j) - E(i)\right)\right] \geq rand(0,1), \tag{4.10}$$

while in the NPT ensemble, the test would also depend on the volumes of the two

replicas, $V(i)$ and $V(j)$, giving,[109]

$$\exp\left[\left(\frac{1}{k_B T_B}-\frac{1}{k_B T_A}\right)\left(E(j)-E(i)+P(V(j)-V(i))\right)\right] \geq rand(0,1). \quad (4.11)$$

Details of the derivation of these, and other tests are given in appendix C.3.

The probability of accepting the PT tests depends strongly on the difference in temperature between replicas, and the number of degrees of freedom, f.[107,110] The probability of accepting a temperature-swap move is directly related to the overlap in energies of the two replicas, which is directly related to the fluctuations in energies of the replicas.[107,110] Since the energy fluctuations scale with approximately $O(f^{-1/2})$,[17] the difference in temperatures between replicas should also scale with $O(f^{-1/2})$. This would mean that the number of replicas required to span a given temperature range would grow with $O(f^{1/2})$.[107,110]

Parallel Tempering improves the sampling of the system as it allows configurational changes that occur at high temperature to "rain down" to low temperatures through successive PT moves. A low temperature replica could have its temperature raised such that an important configurational change occurred, upon which it could then return to the temperature whence it came. The benefit of this method over simulated annealing[4] is that replicas are continually moving up and down in temperature, and most importantly, the relative weighting of the configurations are correct for each of the sub-ensembles. Applications of PT are predominantly to enhance configurational sampling, either for a polymer melt,[111] for small peptide or protein systems.,[107,108,112,113] or for a simplified protein-ligand system.[114]

4.3.2 PTTI

Parallel Tempering Thermodynamic Integration (PTTI) represents the combination of PT with FDTI. Parallel Tempering is applied independently to each window along the λ-coordinate. There will thus be multiple replicas and temperatures for $\lambda = 0.0$, multiple replicas and temperatures for $\lambda = 0.05$ etc. The immediate benefit of this scheme is that there will be multiple trajectories at each λ-value and thus the effect of random sampling error should be reduced. However, a disad-

vantage of this method will be that the total number of trajectories will be equal to the number of λ-windows *multiplied* by the number of temperatures. With 21 windows along λ, this could reach a very large number, and stretch the resources of even the biggest Beowulf clusters.

4.3.3 Simulation Conditions

The application of the PTTI method to the water-methane system will be as similar as possible to the application of FDTI from section 3.4, thus enabling a fair comparison between methods. Initial investigations suggested that a difference in temperature between replicas of 5 °C gave a sufficiently high acceptance rate of the PT moves. To keep the number of replicas reasonable, only 16 were used, spaced evenly between 10 °C and 85 °C. Since the λ-coordinate was split into 21 windows, this gave 336 individual trajectories per simulation. Four copies of the simulation were run, each starting from the same structure as the four FDTI simulations. Owing to the expense of the calculations, the first two simulations only ran for 5 M steps per trajectory (1.68 billion (B) steps per simulation), while the last two simulations ran for the full 10 M steps (3.36 B steps per simulation). Parallel tempering moves were attempted between neighbouring replicas, every 50 K steps. The pairs chosen followed an *even-odd* pattern, with even replicas swapping with the neighbour above on one move, and odd replicas swapping with the neighbour above on the next PT move. The trajectory run at each λ-value and temperature originally started from the same initial structure as used in the FDTI simulations. However, to enable each system to equilibrate correctly, a further 900 K MC steps were performed for each temperature and λ-value before the start of data collection.

4.3.4 Results

The results at 25 °C for the four simulations are shown in table 4.2. Simulations 1 and 2 consisted of only 5 M steps, while simulations 3 and 4 had the full 10 M. The results of 3 and 4 are thus shown after 5 M steps, and also after 10 M steps.

Simulation	Total number of MC steps / M	Relative free energy at 25 °C / kcal mol^{-1}	
		Forwards	Backwards
1	5	8.3 (0.9)	8.3 (0.9)
2	5	7.3 (1.0)	7.4 (1.0)
3	5	7.6 (0.8)	7.7 (0.8)
4	5	8.2 (1.1)	8.2 (1.1)
3	10	7.8 (0.5)	7.9 (0.5)
4	10	8.1 (0.7)	8.2 (0.7)

Table 4.2: The results of 4 independent applications of PTTI to calculate the relative hydration free energy of water and methane. The first 2 M steps of each trajectory were discarded as equilibration. The standard error is shown in parenthesis, and is calculated over blocks of 500 K steps, in an identical manner to the FDTI simulations from section 3.4.2.

4.3.5 Analysis

These results show good agreement with the results from FDTI in section 3.4.2. Indeed, the spread of results and magnitude of the errors are very similar. Only the results from simulations 3 and 4 after 10 M steps show slight improvement over standard FDTI. The PMFs (not shown) are also similar, as are the plots of the convergence of the simulations (figure 4.7). However, the results do not show the marked improvement necessary to justify the sixteen-fold increase in resources needed for these simulations.

Temperature Exploration

The results from four application of PTTI to calculate the relative hydration free energy of water and methane are of almost the same quality as those from standard FDTI. This is despite the PTTI simulations using sixteen times the number of processors / simulation steps as standard FDTI. The reason for this apparent lack of improvement can be seen in the temperatures that each replica adopted throughout the simulation. These *temperature trajectories* all look very similar for all values of λ, and all four simulations, so only those from for $\lambda = 0.0$ for simulation 3 are shown (figure 4.8). To make this figure clearer, only 7 of the 21 replicas are shown. There is not much motion in temperature, and each replica lingers at the temperature from which it started. While there is a satisfactory acceptance rate

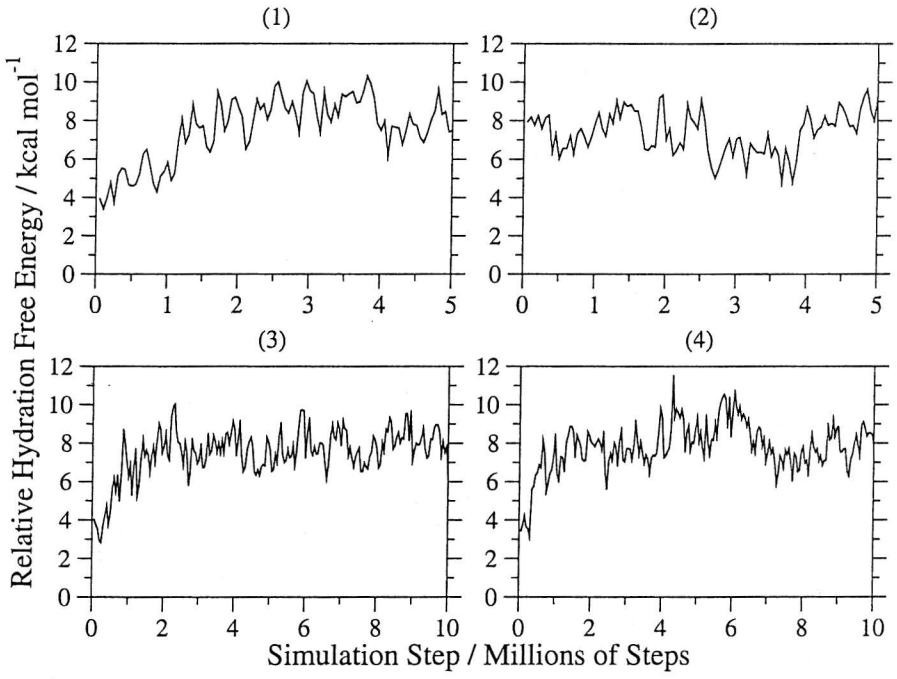
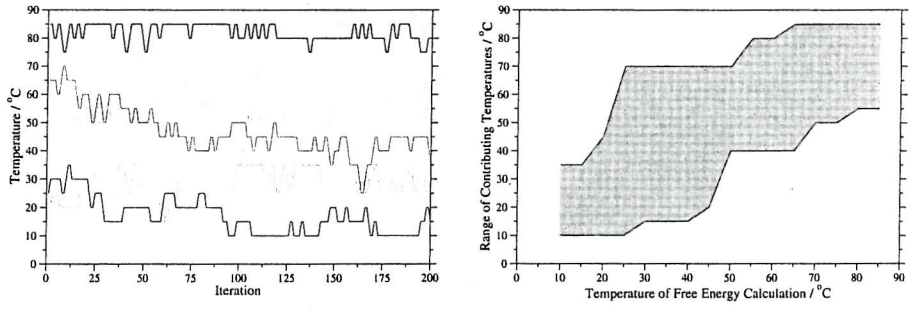


Figure 4.7: The predicted relative hydration free energy of water and methane as a function of simulation step at 25 °C, as calculated by four independent applications of PTTI. Simulations 1 and 2 ran for 5 M steps, while 3 and 4 ran for 10 M steps. The relative hydration free energy was calculated every 50 K steps over the sampling from the previous 50 K steps.



(a) Temperature Exploration (b) Travel

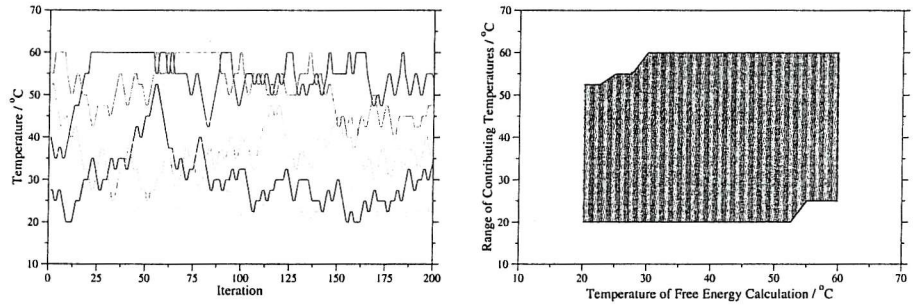
Figure 4.8: (a) Temperatures visited by seven of the sixteen replicas, for $\lambda = 0.0$ from simulation 3, as a function of iteration. (b) The travel for each temperature for this simulation. This shows the range of temperatures previously visited by replicas that contributed data for the free energy calculation at each temperature.

for these simulations (approximately 40%), most of the accepted moves are pair-swaps, where replicas repeatedly swap back-and-forth between neighbouring temperatures (e.g. the top two replicas in figure 4.8). This is because the method does not drive the temperature swapping. Instead, it swaps the temperatures of replicas if their trajectories are ready to swap, e.g. if a low temperature trajectory has a higher than normal potential energy, or a high temperature replica has a lower than normal potential energy. The trajectories thus drive the temperature swapping, not the other way around.

To visualise this temperature exploration better, we can define the *travel* of a temperature as the range of temperatures that contribute to its statistics. For example, the travel at 25 °C may be 10 °C to 45 °C. This would mean that no replica swapped to 25 °C that had *previously* been at a temperature of above 45 °C, or below 10 °C. The free energy calculated at 25 °C would thus only contain statistics that had been collected between 10 °C and 45 °C. Ideally, the free energy calculated at each temperature would contain statistics that were collected at all available temperatures. The travel plot in this case would show a filled square over the temperature range. At this point the travel of the simulation would be total, and the PT ensemble is likely to have converged.

The travel for all values of λ and all simulations were calculated, and were all very similar. The travel for $\lambda = 0.0$ for simulation 3 is shown in figure 4.8. The travel plots for all of the simulations show the same key features, namely a near linear increase in the values of travel with respect to temperature, and a near constant range of travel for each temperature.

To investigate whether the poor travel of the four PTTI simulations was responsible for the lack of improvement of results, two further PTTI simulations were run which used smaller temperature separations. Fifteen replicas used temperatures from 20 °C to 55 °C with a gap of 2.5 °C, and a final sixteenth replica was used at 60 °C. Each simulation was otherwise identical to those in section 4.3.3, and both ran for the full 10 M steps per trajectory. The temperature exploration was much improved, with an acceptance rate of approximately 65%, and a travel that was almost total (figure 4.9). Despite the increase in travel, the range of temperatures sampled was comparable to the previous simulations, as a smaller



(a) Temperature Exploration (b) Travel

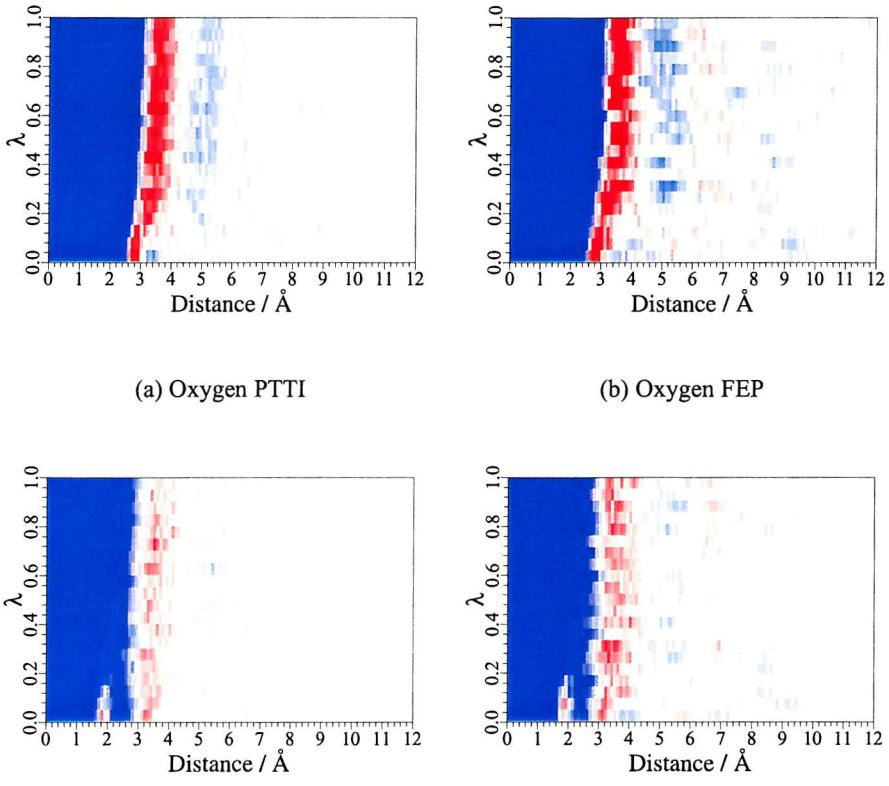
Figure 4.9: (a) Temperatures visited by seven of the sixteen replicas, for $\lambda = 0.0$ from simulation 5, as a function of iteration. (b) The travel for each temperature from this simulation.

Simulation	Relative free energy at 25 °C / kcal mol^{-1}	
	Forwards	Backwards
5	8.5 (0.5)	8.6 (0.5)
6	8.4 (0.5)	8.5 (0.5)

Table 4.3: The results of 2 further applications of PTTI to calculate the relative hydration free energy of water and methane, using more closely spaced temperatures. The standard error is shown in parenthesis.

range of temperatures were available. The results of these simulations are shown in table 4.3.

These two results are much better than those from FDTI, showing a very low spread and a smaller standard error. This improvement is most likely the result of the averaging out of random sampling error through the use of multiple trajectories at each λ-value. To see if the sampling was improved over FEP or FDTI, the RDFs with respect to λ were calculated for simulations 5 and 6. Those calculated for simulation 5 are shown in figure 4.10, together with those calculated in section 3.3.2 from the first FEP simulation. The amount of data contributing to all of these RDFs were identical, so valid comparison can be made. These RDFs show slight improvement over those from FEP or FDTI, and show less noise in the second solvation peaks. However there is still some noise at the beginning of the main peak of the hydrogen solvation shell, and like FEP, the RDFs do not change smoothly across the λ-coordinate.



(a) Oxygen PTTI

(b) Oxygen FEP

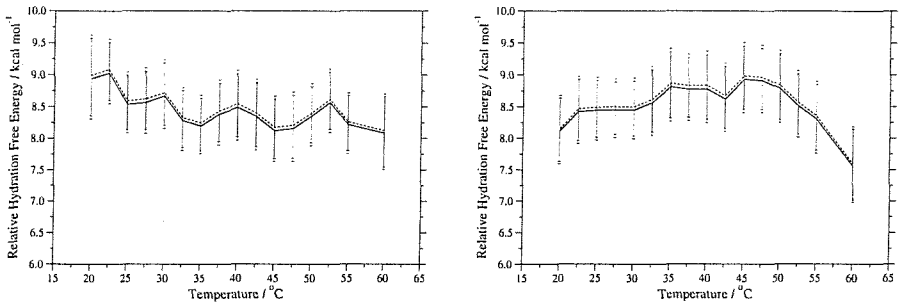
(c) Hydrogen PTTI

(d) Hydrogen FEP

Figure 4.10: The hydrogen and oxygen RDF with respect to λ as calculated from the fifth PTTI simulation, and for comparison, from the first FEP simulation. The scale runs from blue (0.0), through white (1.0) up to red (greater than 2.0).

4.3.6 Free Energies with Respect to Temperature

PTTI is a computationally expensive free energy method, as it runs so many more trajectories than standard FDTI. The extra expense could be justified if the method could reliably return the temperature dependence of relative free energies, thus providing a route to the relative entropies and enthalpies (since entropy is the gradient of the free energy with respect to temperature). The temperature dependence of the relative hydration free energy predicted by simulations 5 and 6 is shown in figure 4.11. While simulation 5 shows a pronounced trend in free energy with temperature, this is not seen in simulation 6. Indeed, the trend in results from simulation 6 from 20 °C to 45 °C is the exact opposite of that seen in simulation 5. These results demonstrate that the free energies are not yet sufficiently converged to extract reliable trends with respect to temperature.



(a) Simulation 5 (b) Simulation 6

Figure 4.11: The temperature dependence of the relative hydration free energy of water and methane, as calculated by simulations 5 and 6. The error bars represent a single standard deviation. Both the forwards (solid line) and backwards (dashed line) free energies are shown.

4.3.7 Summary

Parallel Tempering Thermodynamic Integration was successfully applied to calculate the relative hydration free energy of water and methane. The method allows multiple trajectories to visit each λ-value, thus minimising random sampling error, and also enhances sampling through the use of higher temperature simulations. However, large numbers of replicas and small differences in temperature were needed to ensure adequate temperature exploration. This led to a highly expensive calculation, which is not justified by the only modest improvements in the spread of results and their associated error. These runs required sixteen times the CPU than the FDTI simulations, and thus would need to be running on 336 processors to run all 21 values of λ simultaneously. Also, as 336 trajectories are generated, the data storage requirements become significant (each simulation requires several gigabytes), and the total number of MC steps becomes huge (3.3 *billion* per simulation). Disappointingly, the resulting free energies are too noisy to reliably obtain their temperature dependence, and thus the vast majority of the results, 3.1 B, of the 3.3 B steps of simulation, are effectively discarded. The only modest improvement in the results for this system is probably due to the lack of any significant energy barriers encountered as water is perturbed into methane. This system presents more of an entropic challenge, as it seems to be the number and range of solvent configurations accessible that determines the free energy. PTTI partially rises to

this challenge through the use of multiple trajectories at each λ-value, though its expense suggests that its application should be limited to systems which present enthalpic challenges, and where high temperature sampling is required to overcome large energy barriers, e.g. in the case of the dihedral angle barrier discussed in section 4.2.

4.4 Hamiltonian Replica Exchange

Parallel Tempering is a special case of general replica exchange (see appendix C.3). There are other special cases, for example the technique of *Hamiltonian Replica Exchange*.[115]

4.4.1 Background

Hamiltonian Replica Exchange is the special case of general replica exchange, where trajectories are set up with different Hamiltonians, and swap moves are made between different Hamiltonians.[115] This method is very similar to Parallel Tempering, and both methods share a common derivation (see appendix C.3). In the original application of this method,[115] the potential of mean force was derived along a structural coordinate for an alanine tripeptide in implicit solvent. This was achieved through the use of umbrella sampling, where a harmonic umbrella potential was applied in sections along the structural coordinate, r. These umbrella potentials, $U_m(r)$, modified the potential energy of the system, $E_0(q,r)$ via,

$$E_m(q,r) = E_0(q,r) + U_m(r). \tag{4.12}$$

Fourteen replicas of the system were then spread over the reaction coordinate, each replica using a different umbrella potential, which were centred on successive windows along the structural coordinate. A replica exchange simulation was then performed, which ran each of the fourteen replicas in parallel. Periodically, neighbouring replicas, i and j, in neighbouring umbrella potentials, m and n, were tested via,

$$\exp\left[\frac{1}{k_B T}\left(U_n(r_j) - U_n(r_i) + U_m(i) - U_m(j)\right)\right] \geq rand(0,1). \tag{4.13}$$

If this test was successful, then the coordinates of the replicas were swapped, otherwise the replicas were left where they were. Once the simulations had converged, the statistics within each umbrella potential window were unbiased and recombined via WHAM to construct the PMF along the reaction coordinate.

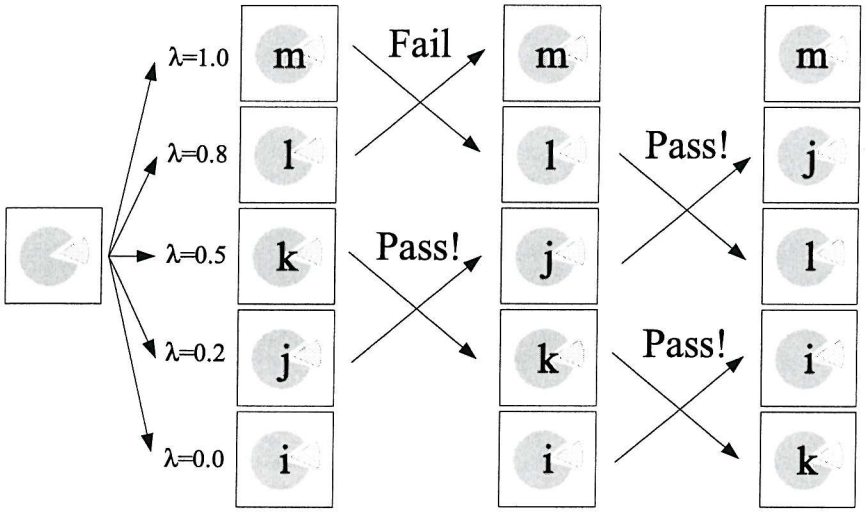
Despite the obvious potential of this method, its application seems to be limited to only this original test case; the structural sampling of blocked alanine tripeptide using a distance dependent dielectric representation of the solvent. The only other application of Hamiltonian Replica Exchange has been in the simulation of protein folding,[110] where multiple Hamiltonians change the hydrophobicity of the solvent in a simplified forcefield. This allows the protein folding pathways to escape from local minima by swapping up into a more hydrophobic environment, which acts to partially denature it. We propose to develop this promising technique of Hamiltonian Replica Exchange in the context of free energy calculations over a perturbing λ-coordinate.

4.4.2 Replica Exchange over λ

Hamiltonian Replica Exchange forms a generalised ensemble over a collection of Hamiltonians, or forcefields. The λ-coordinate acts by perturbing the forcefield of the system, and thus creates a whole sequence of different Hamiltonians (see section 2.5.1). We can thus form a generalised ensemble over λ, and perform replica exchange moves between different λ-windows (see figure 4.12). This simulation would be set up in an identical manner to an FEP or FDTI simulation, with one trajectory within each λ-window. Periodically, trajectories in neighbouring λ-windows, with coordinates i at $\lambda = A$, and j at $\lambda = B$, would be tested according to,

$$\exp\left[\frac{1}{k_B T}\left(E_B(j) - E_B(i) - E_A(j) + E_A(i)\right)\right] \geq rand(0,1). \qquad (4.14)$$

This equation is derived in appendix C.3. If the test is accepted then the coordinates are swapped between the two λ-windows. At the end of simulation, the complete set of statistics for each λ-window represents a correct ensemble distribution for that λ-value, so normal FEP or FDTI analysis could be used obtain the relative free energies. The simulation is thus *no more expensive or complicated*



1
Get original system. Create replicas of that system (i to m) and distribute over λ. Perform normal NVT or NPT sampling for each replica in parallel.

2
Perform a *Replica Exchange* test between neighbouring replicas. Swap the coordinates of pairs that pass. Do nothing with the pairs that fail. Then run some more NVT or NPT sampling.

3
Keep repeating this test / sampling cycle until the trajectories have converged. At this point, the set of coordinates for each λ-value (e.g. 'i, i, k...' for λ=0.0, or 'j, k, i...' for λ=0.2) form a correct ensemble distribution.

Figure 4.12: The algorithm used for RETI and REFEP.

than normal FEP or FDTI. Indeed, the only extra step is this periodic testing and swapping between neighbouring λ-windows, and the need for additional book-keeping to record which trajectory is at which λ-value. As in the case of Parallel Tempering, the amount of information to exchange for each move is very small, so the method is well suited to application over a Beowulf cluster. The probability of accepting a move is dependent on the amount of overlap between the pair's Hamiltonians. Since this overlap is already high, to allow good convergence of the free energy equations, the acceptance rate of the λ-swap moves is also expected to be high. Since this method is equally applicable to FEP or FDTI, we shall look at both methods, *Replica Exchange Thermodynamic Integration* (RETI) and *Replica Exchange Free Energy Perturbation* (REFEP).

4.4.3 Simulation Conditions

Four copies of REFEP and RETI were performed, using identical starting coordinates and conditions as the corresponding FEP (section 3.3) and FDTI (section

Method	Simulation	Relative free energy at 25 °C / kcal mol^{-1}	
		Forwards	Backwards
REFEP	1	9.7 (0.6)	9.3 (0.6)
REFEP	2	8.7 (0.5)	8.6 (0.5)
REFEP	3	9.0 (0.5)	8.4 (0.5)
REFEP	4	8.7 (0.5)	8.4 (0.5)
RETI	1	8.3 (0.5)	8.4 (0.5)
RETI	2	8.5 (0.5)	8.6 (0.5)
RETI	3	8.3 (0.5)	8.4 (0.5)
RETI	4	8.4 (0.5)	8.4 (0.5)

Table 4.4: The results of four applications each of REFEP and RETI to the calculation of the relative hydration free energy of water and methane. The results are calculated over the last 7 M steps of simulation per λ value, with the standard error calculated shown in parenthesis.

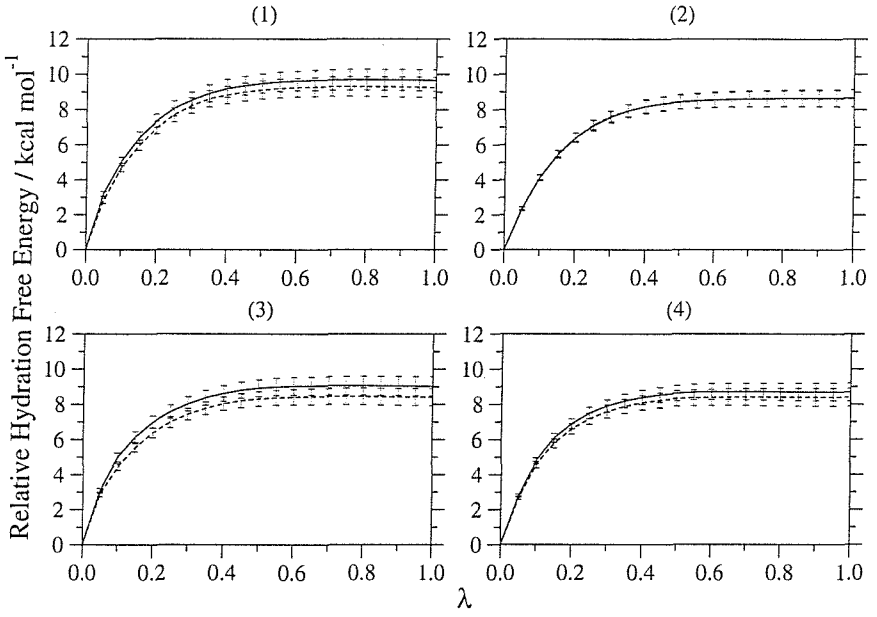
3.4) simulations. The simulations were run for 10 M steps per trajectory, over 200 blocks of 50 K steps. A λ-swap move was attempted at the end of each block. To make sure that the starting coordinates were properly equilibrated for each of their respective λ-values, an extra 900 K steps of equilibration were performed for each trajectory before the start of each simulation.

4.4.4 Results

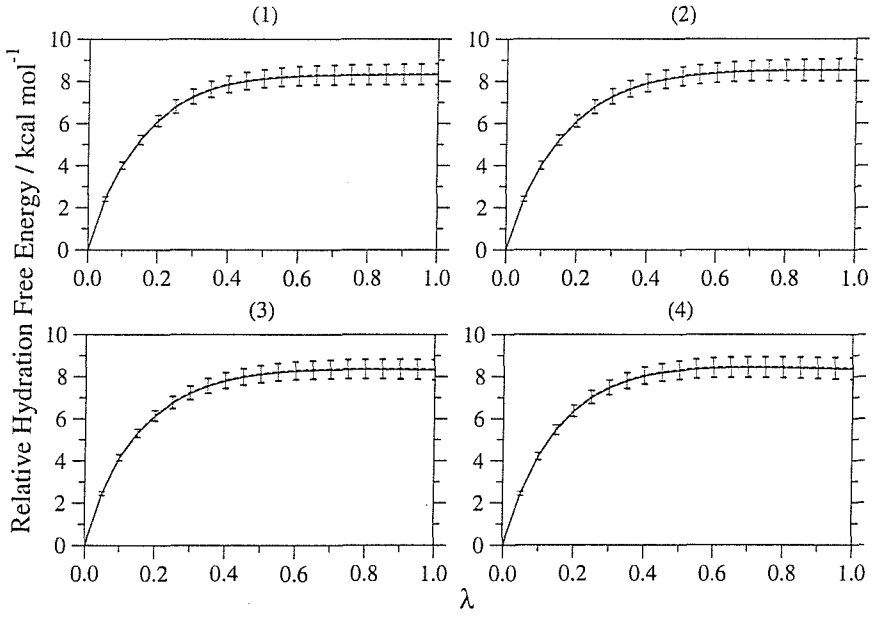
The predicted free energies from the four REFEP and four RETI simulations are shown in table 4.4. The results were calculated over the last 7 M steps, and the standard errors calculated in an identical manner to the FEP and FDTI simulations. The predicted PMFs from each of the eight simulations are shown in figure 4.13.

4.4.5 Analysis

The use of the λ-swap move has improved the quality of these results over standard FEP and FDTI. The REFEP results show a smaller hysteresis (0.1 to 0.4 kcal mol^{-1}) compared to FEP (0.4 to 1.0 kcal mol^{-1}), a smaller spread of results (8.4 kcal mol^{-1} to 9.7 kcal mol^{-1}, compared to 7.3 kcal mol^{-1} to 9.9 kcal mol^{-1}), and a smaller standard deviation (0.5 to 0.6 kcal mol^{-1} compared to 0.6 to 0.7 kcal mol^{-1}).



(a) REFEP



(b) RETI

Figure 4.13: The PMF for the relative hydration free energy of water and methane as calculated by (a) four independent REFEP, and (b) four independent RETI simulations. The results are calculated over the last 7 M steps of simulation, and both the forwards (solid line) and backwards (dashed line) free energies are shown. The standard errors propagated over λ are also shown.

While the λ-move has helped the hysteresis, it has not overcome it, and thus the results still show the same errors that the FEP results show. These could be improved by redistributing the λ-windows, though this requires prior knowledge of the system, and is not valid for a blind application of the method to a new system.

RETI, like FDTI, shows almost no hysteresis, and thus the estimates of the free energy are of a higher quality than the REFEP results. RETI shows a very low spread of results (8.31 kcal mol^{-1} to 8.57 kcal mol^{-1}) compared to FDTI (8.4 kcal mol^{-1} to 9.4 kcal mol^{-1}) and the results have a lower standard deviation (0.5 kcal mol^{-1} compared to 0.6 to 0.7 kcal mol^{-1}). The level of agreement between the four RETI results is graphically shown in the similarity of the four predicted PMFs (figure 4.13). These four PMFs are almost identical in shape and size. What makes the results even more pleasing is that the results are in excellent agreement with the experimental value of 8.31 kcal mol^{-1}. This is a very positive result, especially given the simplicity of the water and methane forcefields used in this study.

4.4.6 λ-exploration

The results from these simulations show that the use of replica exchange has markedly improved the quality of the free energy results. The reason for this can be seen in the λ-trajectories. All eight simulations have very similar λ-trajectories, so only those from the first RETI simulation are shown (figure 4.14). For the sake of clarity, only six of the 21 replicas are shown. The acceptance rate for the λ-moves was very high, at over 80%. This was expected, as the free energy calculation required that there was good overlap between neighbouring λ-values. The trajectories show a high degree of motion, travelling up and down the λ-coordinate multiple times throughout the simulation. The travel of the trajectory can be calculated in a similar manner to that for PTTI (section 4.3.5). All eight simulations show total travel, in other words the statistics at each λ-value include statistics gathered at *all* other λ-values.

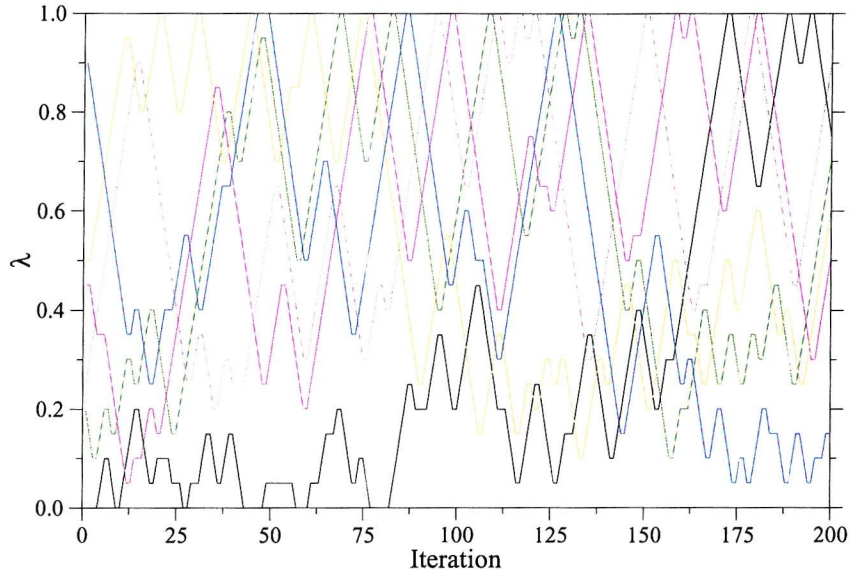
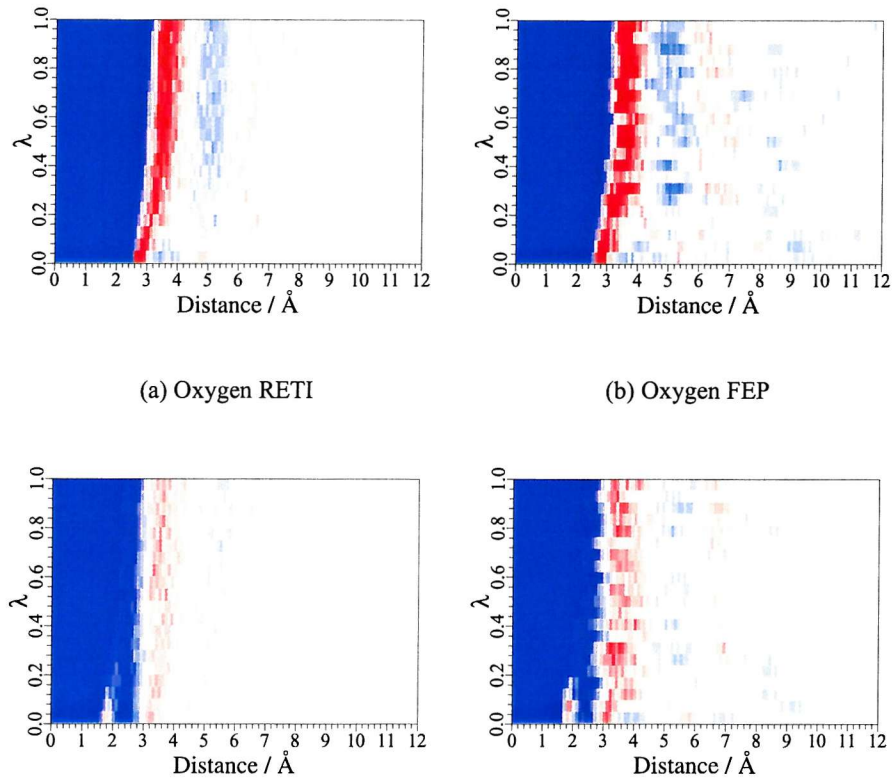


Figure 4.14: The value of λ with respect to iteration for six of the 21 replicas from the first RETI simulation.

4.4.7 Radial Distribution Functions

The high quality of the RETI and REFEP results arises because the individual trajectories are free to move through λ. This can be clearly seen in the plots of the radial distribution functions with respect to λ (figure 4.15). These were composed in an identical manner to those in section 3.3.2, and they are very similar across all eight simulations. Only the RDFs from the first RETI simulation are shown, and are compared with the RDFs from the first FEP simulation. The RETI RDFs show the same features as those from FEP. The RETI RDFs are clearly of a much higher quality than the FEP RDFs, despite having an identical number of sampling points and method of composition. The quality of the RETI RDFs is so high that there is clear evidence of the second, and even third solvation shell. What is also clear is that the RETI RDF is smooth with respect to λ. There is little to no noise in the location of the peaks, and all of the main features of the RDF are consistently well defined across the entire λ-coordinate. This smooth progression across the λ-coordinate arises because the statistics for each λ-value are composed from trajectories that have visited every λ-value, and thus are themselves smoothly moving from water to methane and back again.



(a) Oxygen RETI

(b) Oxygen FEP

(c) Hydrogen RETI

(d) Hydrogen FEP

Figure 4.15: The radial distribution function with respect to λ for the first RETI simulation, and for comparison for the first FEP simulation. Note how much smoother the RETI results are than the FEP. The scale runs from blue (0.0), through white (1.0) up to red (greater than 2.0).

4.4.8 Summary

Two new free energy methods, *Replica Exchange Thermodynamic Integration* (RETI) and *Replica Exchange Free Energy Perturbation* (REFEP) were derived from the equations of Hamiltonian Replica Exchange. The methods are trivial extensions of FDTI and FEP, and merely add a periodic λ-swap move between neighbouring λ-windows. This small and easily implemented extension significantly improves the quality and reliability of the results, resulting in a reduction of hysteresis, improved consistency of results and reduction of standard error. The consistency of the results is enhanced as the λ-swap move allows all trajectories to visit all λ-values, and thus random sampling error is reduced, and properties of the system, like the RDF, progress smoothly and consistently with respect to λ.

RETI and REFEP can be thought of as a replica exchange version of AdUmWHAM over λ. In the previous section, the Parallel Tempering algorithm was discussed as an example of a generalised ensemble over temperature space. This ensemble covered multiple temperatures through the use of multiple independent replicas that could periodically exchange coordinates. Unlike BMW, there was no need to perform an iteration, as the individual replicas were already correctly Boltzmann weighted. An alternative method of forming a generalised ensemble over temperature is *Simulated Tempering*.[98] This method treats temperature as a dynamic coordinate, and tries to sample temperature evenly.[98] This can only be achieved through a tedious iteration that aims to obtain the weighting factor necessary for the MC moves.[98] Parallel tempering and simulated tempering thus represent two different ways of obtaining the same generalised ensemble.[116] This equivalence between Parallel Tempering and Simulated Tempering can be compared to the equivalence between RETI and AdUmWHAM over λ. Just as Simulated Tempering aims to sample temperature evenly, so AdUmWHAM over λ is a method that aims to sample the whole of λ-space evenly. This aim is also achieved through the use of a tedious iteration, which refines the weighting factor (umbrella potential) necessary to encourage even sampling along λ. Conversely, RETI uses multiple replicas over λ, which are allowed to periodically exchange coordinates. Full sampling over λ is achieved without the need for a tedious iteration, as the individual replicas are already correctly Boltzmann weighted. Just as Parallel Tempering and Simulated Tempering represent two different approaches to obtain the same ensemble, so RETI and AdUmWHAM over λ represent two different approaches to achieving full sampling over λ. The advantage of RETI lies not only in the lack of a tedious iteration, but also in the fact that the value of λ is constrained within each replica. In the case of AdUmWHAM, the umbrella potential drives the sampling of the λ coordinate. This can push the system out of equilibrium, and forces the configurational sampling to keep up with the changing λ-value. In contrast, the value of λ for each replica in RETI is constrained to a fixed value. This means that the system is not pushed out of equilibrium. In addition, the exchange of coordinates between replicas is controlled by an MC test.

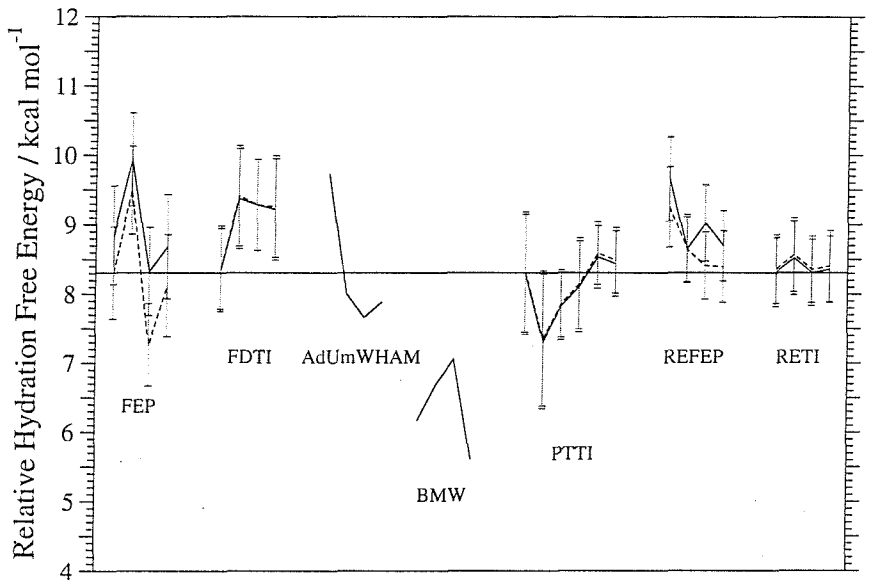


Figure 4.16: A comparison of the predicted relative hydration free energy of water and methane as calculated by all methods. The four simulations from FEP, FDTI and modified AdUmWHAM are shown, as are the four BMW simulations from set 5, as are the six PTTI simulations, and four RETI and REFEP simulations. The experimental result of +8.31 kcal mol^{-1} is also shown as a horizontal line.

This ensures that thermal equilibrium is maintained for the change in λ. This effectively means that configurations will only swap when they are compatible. In this way, the change in λ is driven by the underlying configurational sampling, rather than the changing λ-value driving the configurational sampling.

4.5 Comparison of Methods

A comparison of the predicted relative hydration free energy from all of the methods, and experiment, is shown in figure 4.16. This figure clearly demonstrates that the newly developed method of Replica Exchange Thermodynamic Integration is the method of choice for the water-methane system. This method avoided the problems of hysteresis that reduced the quality of the FEP and REFEP results, as it evaluated only the local gradient of the free energy with respect to λ. The method allowed full exploration of λ by each trajectory, but it avoided driving the sampling as was the case of AdUmWHAM and BMW, as the λ-swap move will only

occur if the trajectories themselves are ready to swap. The configurational sampling within each replica drove the motion along λ, rather than the forced motion along λ driving any configurational change. The method also significantly reduced random sampling error through the use of multiple trajectories at each λ-value, as was the case for PTTI. RETI achieved this without creating any additional overhead over standard FDTI, unlike PTTI which stretches the capacity of even the largest Beowulf cluster. It is useful to note that the predicted free energies of the last two PTTI simulations, which had smaller temperature differences and thus greater temperature exploration, are almost identical to the four free energies from RETI.

4.6 Conclusion

The calculation of the relative hydration free energy of water and methane provided the highly challenging test case on which four new free energy methods were developed. The first, Bivariate Multicanonical WHAM, combined adaptive umbrella sampling with the multicanonical ensemble. Disappointingly, the water-methane system was too challenging for the method, and it failed to converge. The second method developed, Parallel Tempering Thermodynamic Integration, improved standard FDTI by combining it with parallel tempering. The results were a slight improvement over standard FDTI, though this improvement could not justify the sixteen-fold increase in resources necessary. The final two methods, Replica Exchange Free Energy Perturbation and Replica Exchange Thermodynamic Integration combined Hamiltonian replica exchange with FEP or FDTI. The results were a significant improvement over FEP or FDTI. RETI produced especially good results, being reliable to within 0.26 kcal mol^{-1}, and with a standard error of just 0.5 kcal mol^{-1}. RETI achieved this as it combined all that was good about the previous methods, including, paradoxically, the advantages of fixed λ methods, such as FEP and FDTI, with dynamic λ methods, such as AdUmWHAM and BMW. This was achieved through the use of a generalised ensemble over λ, using the techniques of replica exchange investigated in PTTI. RETI truly takes the lessons learned throughout this investigation, and combines their advantages into a highly

effective and reliable technique.

Chapter 5

The Application of Free Energy Methods to a Small Host-Guest System

5.1 Introduction

Over the last two chapters, three established free energy methods, Free Energy Perturbation (FEP), Finite Difference Thermodynamic Integration (FDTI) and Adaptive Umbrella WHAM (AdUmWHAM), and four newly developed methods, Bivariate Multicanonical WHAM (BMW), Parallel Tempering Thermodynamic Integration (PTTI), Replica Exchange Free Energy Perturbation (REFEP) and Replica Exchange Thermodynamic Integration (RETI), have been applied to the calculation of the relative hydration free energy of water and methane. This application has revealed that the new method of RETI was the method of choice for that system, as it overcame the problems of hysteresis in FEP and REFEP, Hamiltonian lag in AdUmWHAM and BMW, and had the benefits of using multiple trajectories at each λ-value, without the computational expense of PTTI. The water-methane system was chosen as a test case as it encapsulated a lot of the common features of protein-ligand binding, namely charge and solvent rearrangement, into a single simulation. The aim of this chapter is to continue the testing of these free energy methods on another small system. The system chosen for this part of the study was the binding of halide ions to a calix[4]pyrrole derivative. This system

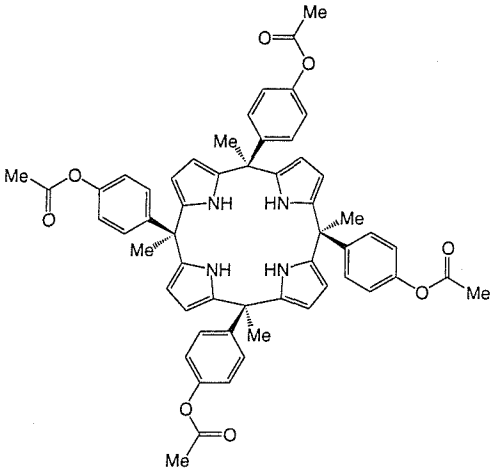
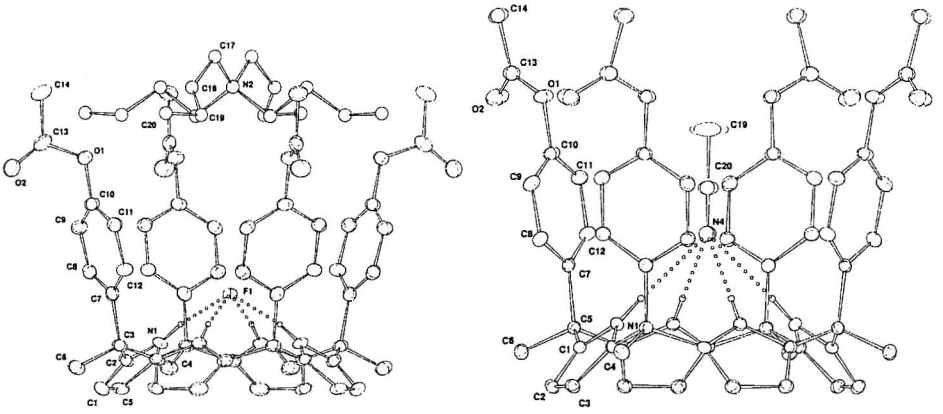


Figure 5.1: Structure of the calix[4]pyrrole derivative.

was chosen as the host is believed to exhibit different configurations depending on which halide is binding.

5.2 Calix[4]pyrrole Complexes

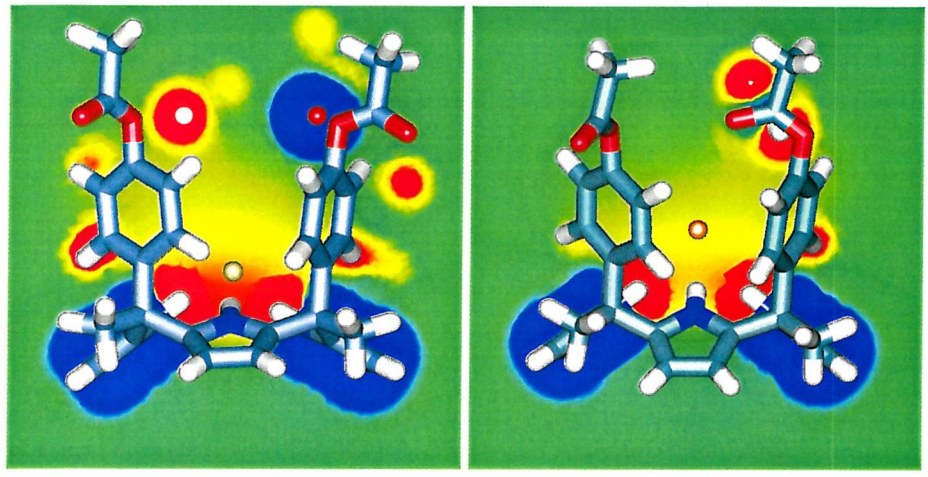
Since the discovery that they are effective anion-binding agents in solution, calix[4]pyrrole derivatives have received much study, leading to their use in anion-binding,[117] sensing,[118,119] and new anion-separation techniques.[120] Recently, Camiolo and Gale[121] synthesised a tetraacetyl-calix[4]pyrrole derivative (figure 5.1), which had the then unique property of selectively binding fluoride ions in dimethylsulfoxide (DMSO). The crystal structure[65] of the fluoride-bound complex was obtained in acetonitrile (figure 5.2), which showed that the complex formed a cup-like configuration, with the fluoride bound deep in bottom of the well. In this configuration, the hydrogens from the four pyrrole groups can hydrogen-bond to the fluoride ion, thus giving stability to the complex. A crystal structure[65] was also obtained for the free calix[4]pyrrole derivative (figure 5.2). This structure showed a single acetonitrile solvent molecule bound in the well, with the calix[4]pyrrole adopting a slightly different configuration. In this configuration, the acetonitrile sits higher up in the cup, and the four pyrrole groups are tilted up more than in the fluoride complex. FEP and AdUmWHAM simulations by Woods *et al.*[65] on the fluoride, chloride and bromide complexes with this derivative, produced



(a) Bound to fluoride (b) Bound to acetonitrile

Figure 5.2: Crystal structures of the calix[4]pyrrole derivative bound to a fluoride ion, and to an acetonitrile solvent molecule. Thermal ellipsoids are drawn at the 30% probability level. The tetrabutylammonium counterion is disordered, so for clarity, is only shown in one configuration in a ball and stick representation.

structures which were in agreement with these two crystal structures. The fluoride was found to bind lower in the complex, with the pyrrole groups tilted towards it. This led to a shorter hydrogen bond (1.9 Å), and a more obtuse hydrogen-fluoride-hydrogen angle (115 °). The chloride and bromide were seen to sit higher in the complex, with longer hydrogen bonds (2.5 Å and 2.6 Å), and a more acute hydrogen-halide-hydrogen angle (85 ° and 79 °). The final simulation structures of the derivative were surrounded by an implicit DMSO solvent, and the electrostatic potential around the molecule was investigated using the Poisson equation.[12,65] A planar slice through the centre of the molecule revealed the presence of a "positive pocket" (figure 5.3),[65] into which only the fluoride was small enough to bind. This pocket only appears when the four pyrrole groups are in the fluoride configuration. The selectivity of the calix[4]pyrrole derivative for fluoride ions was seen as a result of the fluoride inducing slight configurational change in the calix[4]pyrrole, which opened up a positive pocket into which only the fluoride was small enough to bind.



(a) Fluoride (b) Chloride

Figure 5.3: Slices through the electrostatic potential for the calix[4]pyrrole derivative in the configuration adopted when binding fluoride, and when binding chloride. This was calculated via the Poisson equation, with a surrounding DMSO implicit solvent. The location of the fluoride and chloride in the complex is shown as a yellow and orange sphere respectively. The scale ranges linearly from deep red ($\geq$ 10 kcal mol^{-1} e^{-1}), through green (0 kcal mol^{-1} e^{-1}), to deep blue ($\leq$ -10 kcal mol^{-1} e^{-1}).

5.2.1 Literature Simulations

As the calix[4]pyrrole derivative selectively bound fluoride ions, there are no experimental relative binding free energies between fluoride and the other halides. Other workers have investigated related systems. Density functional theory,[122] TI[39] and FEP[123] simulations have been performed on other calix[4]pyrrole derivatives. These studies have concluded that while the host adopts the 1,3-alternate configuration free in solution, the cone configuration is preferred when the receptor binds to an anion. One of these studies[123] predicted a relative binding free energy between the fluoride and chloride of 22 kcal mol^{-1}, while another[39] predicted a relative binding free energy of 17 kcal mol^{-1}. A third study[122] predicted a relative binding energy of approximately 15 kcal mol^{-1}. The relative binding free energy between the fluoride and other halides has been shown to be affected by any contamination by water of the solvent, with increasing water concentration acting to remove the hosts specificity for fluoride.[39,65,124] This is due to water's high affinity for fluoride allowing it to compete with the calix[4]pyrrole derivative.

5.3 Simulation System

The calix[4]pyrrole pyrrole derivative shown in figure 5.1 was used as the test-case for the seven free energy methods. The relative binding free energy of fluoride, chloride and bromide were calculated for the complex solvated by DMSO, with a small amount of water contamination. The system was set up in an almost identical manner to that used by Woods *et al.*[65] The initial z-matrix for the calix[4]pyrrole was constructed from the crystal structure of the complex with fluoride.[65] This was also used to provide starting coordinates for the system. The z-matrix was built using the "fragment" approach of van Hoorn and Jorgensen.[123] They constructed each pyrrole group independently from the dummy atoms, allowing greater flexibility of the macrocycle ring. The OPLS parameters for the macrocycle were also taken from this study,[123] with generic OPLS parameters taken for the pendant arms. The z-matrix and parameters for this system are given in the supplementary material of Woods *et al.*[65] To set up the bound leg of the simulations, the fluoride-bound crystal structure was placed in an orthorhombic box of 1714 DMSO[125] molecules and 30 TIP4P[89] water molecules. The water molecules were positioned randomly, and the box had initial dimensions of 58.8 Å × 59.8 Å × 66.1 Å. A fluoride ion was placed in the host using initial coordinates taken from the crystal structure. This system was then equilibrated for 2 M MC steps, according to the parameters in table 5.1. The free ion leg of the simulation was set up using a fluoride ion, placed in the centre of a box of 640 DMSO molecules and 30 TIP4P molecules, of initial dimension 44.6 Å × 43.7 Å × 45.5 Å. This system was also equilibrated for 2 M steps according to the parameters in table 5.1. These simulations were run using molecule-molecule cutoffs. This meant that if any pair of atoms between two molecules were within the cutoff distance, then the interaction between all pairs of atoms between the molecules were evaluated. All simulations were performed using the same modified version of MCPRO used to perform the water-methane simulations of the previous chapters.

It was decided that the relative binding free energies of three halides would be investigated in this system. The halides chosen were fluoride, chloride and bro-

Parameter	Bound Leg	Free Leg
Ensemble	NPT	NPT
Simulation Temperature	25 °C	25 °C
Simulation Pressure	1 atm	1 atm
Boundary Type	Orthorhombic periodic boundaries	Orthorhombic periodic boundaries
calixpyrrole / halide / solvent move ratio	50 / 1 / 4645	0 / 1 / 1788
Non-bonded cutoff	15 Å	15 Å
Maximum calixpyrrole translation / rotation	0.0 Å / 0.0 °	0.0 Å/ 0.0 °
Maximum halide translation / rotation	0.1 Å/ 5.0 °	0.1 Å/ 5.0 °
Maximum solvent translation / rotation	0.1 Å / 5.0 °	0.1 Å/ 5.0 °
Frequency of volume moves	1 every 10875 MC moves	1 every 4125 MC moves
Maximum volume change	2610 $Å^3$	1005 $Å^3$
Preferential sampling centre	Halide	Halide
Preferential sampling parameter (WKC)	200	200
Number of MC moves per simulation block	250 K	250 K
Number of simulation blocks	8	8

Table 5.1: Parameters used to control the simulations on the calix[4]pyrrole system.

mide, and a closed set of perturbations were run. This closed set, fluoride to chloride, fluoride to bromide and chloride to bromide, enabled the closure on the free energies to be calculated. If the relative binding free energies were determined accurately, then the relative binding free energy between fluoride and bromide should equal the sum of the relative binding free energies of fluoride and chloride, and chloride to bromide. Any deviation from this equality could then be used as a sign of error. The relative binding free energies were calculated from the bound and free legs via the cycle in figure 2.2.

5.4 Application of Established Methods

The first stage of this investigation was the application of the three established free energy methods. These methods, FEP, FDTI and AdUmWHAM, were applied to calculate the relative free energies for the bound and free legs of each of the three perturbations. The application of these methods to the calix[4]pyrrole system closely mirrored their application to the water-methane system. Thus for FEP and FDTI, the λ-coordinate was split into 21 evenly spaced windows, every 0.05 along λ. 10 M steps of simulation were run within each window, over 20 blocks of 500 K steps. FDTI used $\Delta\lambda = 0.001$, and the results were integrated via the trapezium rule. The application of AdUmWHAM used 500 WHAM iterations over two trajectories running asynchronously in parallel. Each iteration consisted of 40 K steps of equilibration and 160 K steps of data collection, and the value of λ was preserved for each trajectory between iterations. In contrast to the simulations on the water-methane system, initial investigations showed little difference in results between simulations which preserved λ between iterations, and simulations that randomised λ. The maximum range of a λ-move was 0.05, and a λ-move was attempted every 100 normal MC steps. This frequency was chosen as previous investigations on this system[65] demonstrated that it did not cause any problems due to Hamiltonian lag.

Each bound and free simulation started from the corresponding final equilibrated coordinates from section 5.3. To investigate how well the free energy methods sampled the locations of the contaminating water molecules, their positions were randomised at the start of each trajectory. The simulations used the same parameters as those used for the equilibration. To allow the reliability of each method to be investigated, each leg of each simulation was repeated four times. Each of these four applications used the same initial starting coordinates, though the positions of the water molecules were randomised for each application. A different random number seed was used for each application.

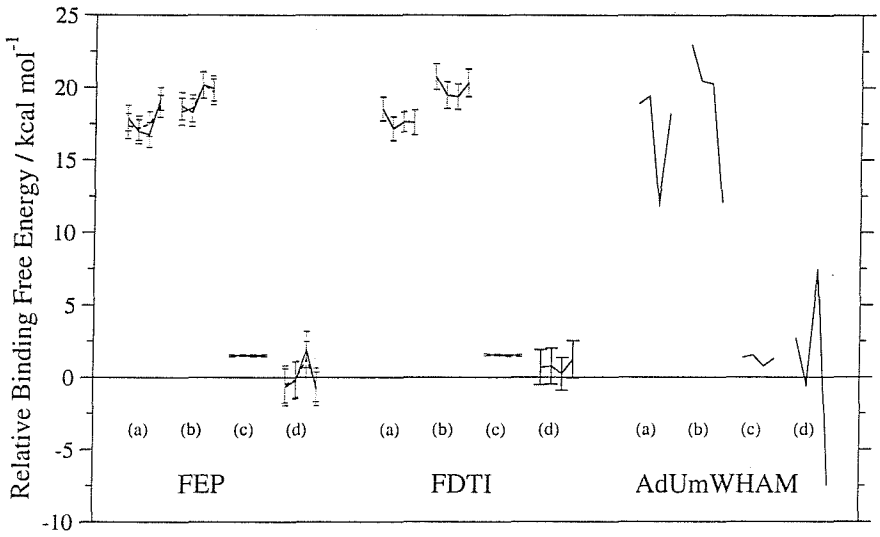


Figure 5.4: Binding free energies and closures for four independent applications of FEP, FDTI and AdUmWHAM. (a) Fluoride to Chloride, (b) Fluoride to Bromide, (c) Chloride to Bromide, (d) Closure. Both the forwards (solid) and backwards (dashed) FEP and FDTI results are shown.

5.4.1 Results

The predicted relative binding free energy for each pair of halides are shown, for four independent applications each of FEP, FDTI and AdUmWHAM. The closure for the cycle is also calculated, as are the combined standard errors (a description of the error analysis is available in appendix E). The results are shown in table 5.2, and for clarity are also plotted in figure 5.4. Analysis of the convergence of the FEP and FDTI results (not shown) resulted in the first 3 M steps per trajectory being discarded as equilibration, with the remaining 7 M steps per trajectory used for data collection.

5.4.2 Analysis

These results all show agreement with the experimental observation that the calix[4]pyrrole derivative selectively binds fluoride ions. Initial analysis of the results suggests that FEP and FDTI have a low spread, and low error, while AdUmWHAM has performed poorly. AdUmWHAM shows large spread and at least one calculation per mutation fails to produce a result in agreement with the either FEP or

	Simulation	Relative Binding Free Energy / kcal mol^{-1}		
		FEP	FDTI	AdUmWHAM
Fluoride to Chloride	1	17.9 (0.9) 17.3 (0.9)	18.5 (0.8) 18.5 (0.8)	18.9
	2	16.9 (0.8) 17.2 (0.8)	17.2 (0.8) 17.2 (0.8)	19.4
	3	16.7 (0.9) 17.5 (0.8)	17.6 (0.7) 17.6 (0.7)	12.0
	4	19.2 (0.8) 18.7 (0.8)	17.6 (0.9) 17.6 (0.9)	18.2
Fluoride to Bromide	1	18.7 (0.9) 18.3 (0.9)	20.7 (0.9) 20.8 (0.9)	23.0
	2	18.3 (1.0) 18.6 (0.9)	19.5 (0.9) 19.5 (0.9)	20.5
	3	20.2 (0.9) 20.2 (0.9)	19.4 (0.9) 19.4 (0.9)	20.3
	4	19.9 (0.9) 19.7 (0.9)	20.3 (1.0) 20.3 (1.0)	12.1
Chloride to Bromide	1	1.51 (0.07) 1.48 (0.07)	1.54 (0.08) 1.54 (0.08)	1.37
	2	1.52 (0.07) 1.51 (0.07)	1.55 (0.07) 1.55 (0.08)	1.53
	3	1.51 (0.08) 1.47 (0.08)	1.50 (0.07) 1.51 (0.07)	0.77
	4	1.52 (0.08) 1.50 (0.08)	1.49 (0.07) 1.49 (0.07)	1.32
Closure	1	-0.7 (1.3) -0.5 (1.3)	0.7 (1.2) 0.7 (1.2)	2.7
	2	-0.2 (1.3) -0.1 (1.3)	0.8 (1.2) 0.8 (1.2)	-0.5
	3	1.9 (1.3) 1.2 (1.3)	0.2 (1.1) 0.2 (1.1)	7.5
	4	-0.8 (1.2) -0.5 (1.2)	1.2 (1.3) 1.2 (1.3)	-7.5

Table 5.2: Binding free energies and closure for four independent applications of FEP, FDTI and AdUmWHAM. Both the forwards and backwards free energies from FEP and FDTI are shown.

FDTI, in spite of the AdUmWHAM simulations appearing to have converged. The predicted free energies for each leg stabilised quickly, by approximately iteration 30, as seen in the convergence plots for the bound leg of the fluoride to chloride perturbation (figure 5.5). This in reinforced by the near even λ-sampling attained during the simulations (not shown).

Visualisation of the structures from the AdUmWHAM simulations revealed

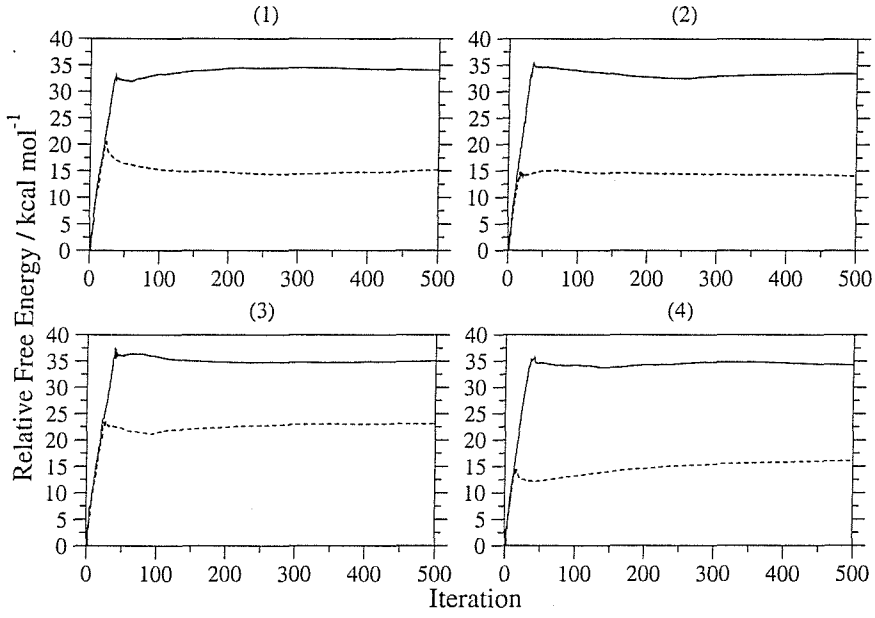


Figure 5.5: Convergence of the bound (solid line) and free (broken line) legs for the four AdUmWHAM simulations on the fluoride to chloride perturbation.

that in some trajectories, a contaminating water molecule had managed to diffuse through solution and hydrogen bond with the halide ion. This can clearly be seen by plotting the distance between the halide ion and the closest water oxygen atom, as a function of AdUmWHAM iteration. These plots are shown for the free leg simulations for the fluoride to chloride perturbation (figure 5.6). Previous studies have shown that hydrogen-bonding between the halide and contaminating water can significantly reduce the binding affinity of the calix[4]pyrrole derivative for fluoride,[39,65,124] as the water is able to compete for the ion. This has the effect of reducing the relative binding free energy between fluoride and the other halides. Examination of the predicted free energies from AdUmWHAM show that the "poor" results are actually the result of the sampling of trajectories that included this hydrogen-bonding. The AdUmWHAM simulations showed no hydrogen bonding with water in the bound legs, but for the free legs, run 3 for fluoride to chloride, run 4 for fluoride to bromide, and run 3 for the chloride to bromide simulations all show that both trajectories exhibited hydrogen bonding for the majority of the simulation. This exactly corresponds to the very low estimates of the relative binding free energy returned by these simulations. This demonstrates the significant effect of water hydrogen bonding on the relative binding free energy.

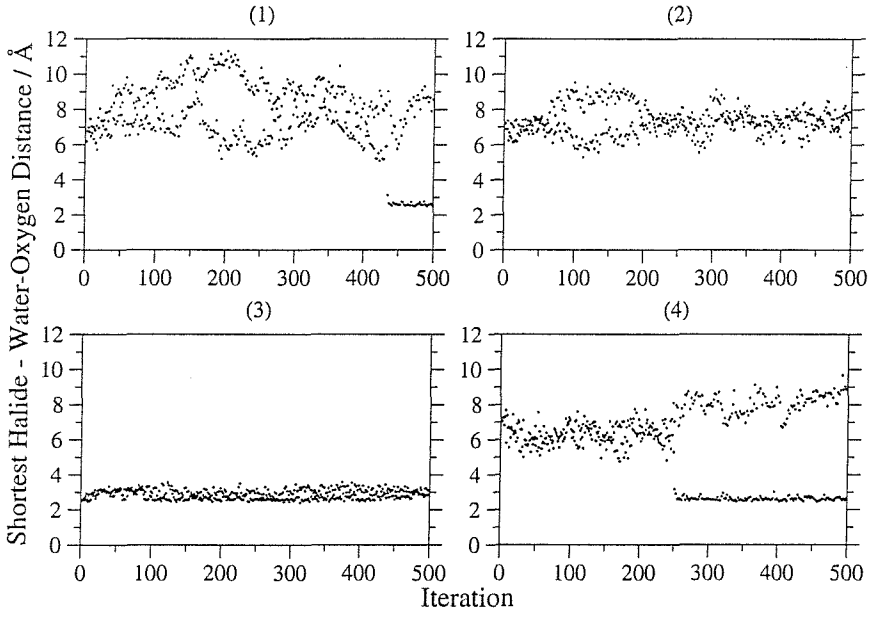
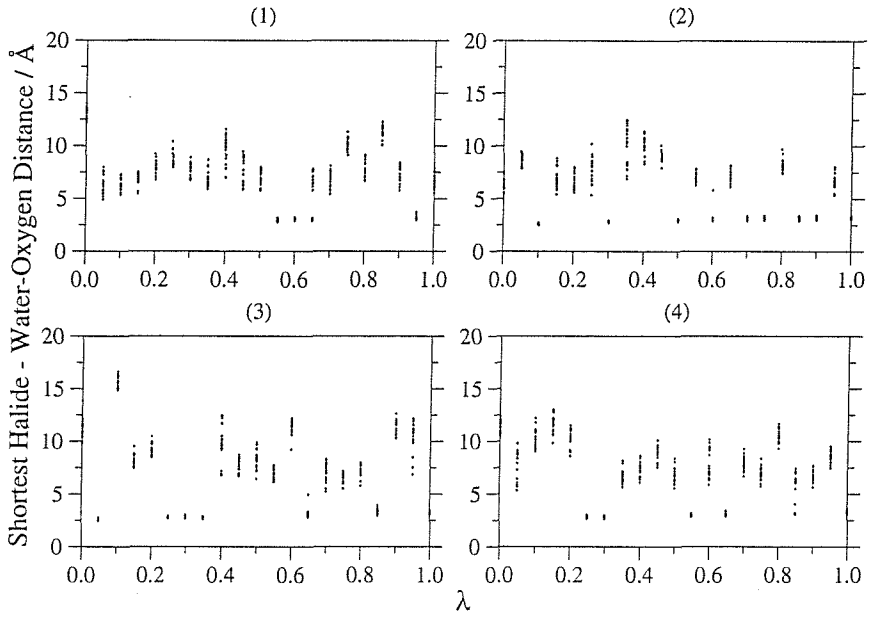
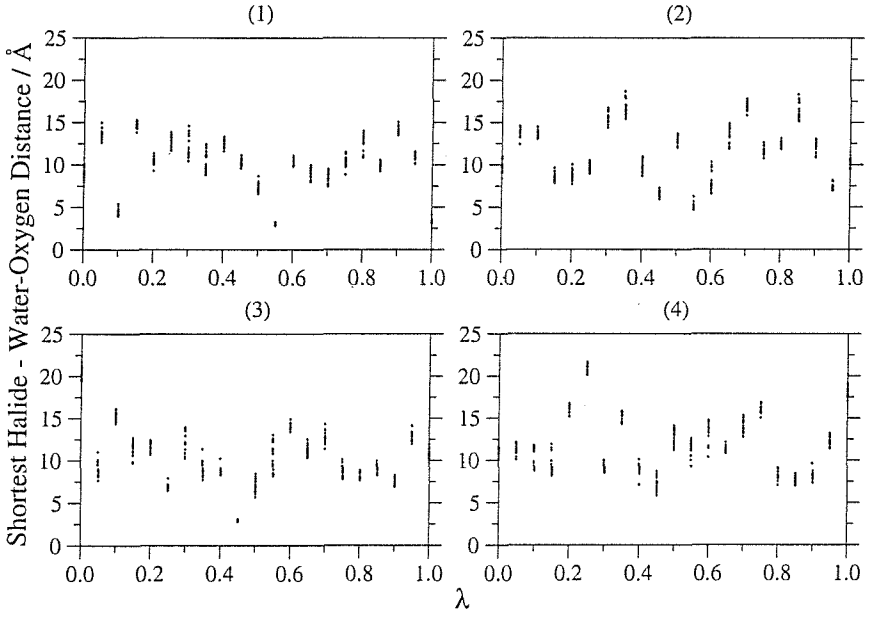


Figure 5.6: The distance between the halide ion and the oxygen of the nearest water molecule for the four AdUmWHAM simulations on the free leg, fluoride to chloride perturbation. The two trajectories within each simulation are clearly visible.

AdUmWHAM fails to properly account for this effect, as its trajectories either do, or do not, exhibit hydrogen bonding. There is no weighting of the hydrogen bond versus non-hydrogen bond states, as there is no free exchange between configurations within a single trajectory. All AdUmWHAM can do, is to ensure that once hydrogen-bonding is obtained, the λ-moves and umbrella allow all values of λ to see this configuration. This is not the case for the FEP or FDTI simulations, which show a mix of hydrogen bond and non-hydrogen bond trajectories across the λ-coordinate. This is seen in the plot of the distance between the halide and nearest water oxygen as a function of λ, for the four free leg, fluoride to chloride simulations (figure 5.7). The 21 different trajectories across the λ-coordinate are either hydrogen bonding, or not hydrogen bonding. This is represented in figure 5.7 by a random zig-zag of the nearest water distance. This pattern is also represented in the bound legs, which can also exhibit hydrogen bonding between the halide and a contaminating water in the host (figure 5.7). The FDTI and FEP results are thus inconsistent across λ, and predict relative binding free energies that are an average over hydrogen bonding *and* non-hydrogen bonding trajectories. This average is in-



(a) FDTI, free leg, fluride to chloride



(b) FEP, bound leg, fluoride to bromide

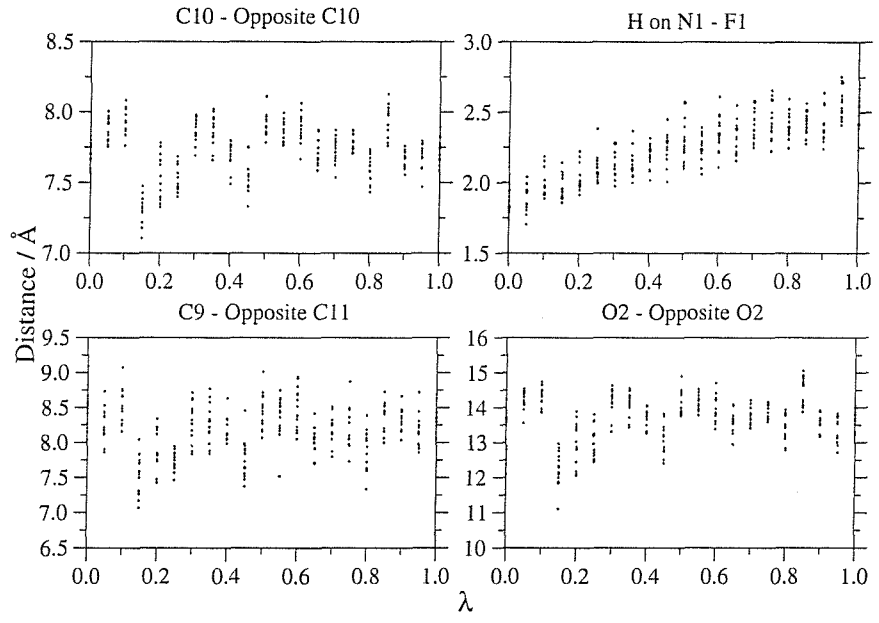
Figure 5.7: The distance between the halide ion and the oxygen of the nearest water molecule for (a) four FDTI simulations on the free leg, fluoride to chloride perturbation, and (b) four FEP simulations on the bound leg, fluoride to bromide perturbation, for each trajectory across λ, and taken every 500 K steps.

correct, as weighting of these states is not performed, and each λ-value sees either one state or the other. The FEP and FDTI results are thus smaller than the pure non-hydrogen bonding results of AdUmWHAM (about 19 kcal mol^{-1} for fluoride to chloride), and larger than for pure hydrogen bonding (about 12 kcal mol^{-1} for fluoride to chloride). The spread in FEP and FDTI results can in part be attributed to the relative ratios of the hydrogen bonding versus non-hydrogen bonding trajectories along λ.

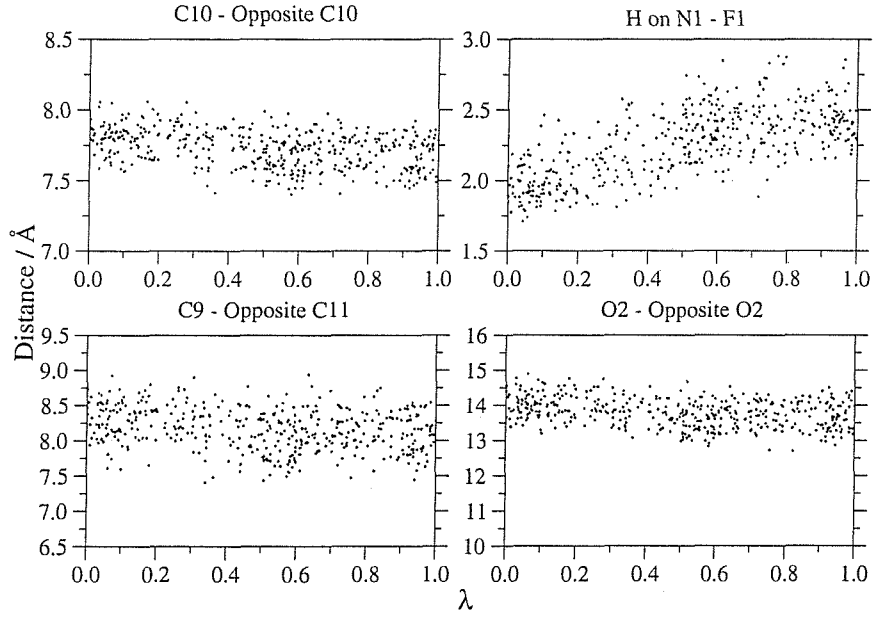
This same effect is also seen in more subtle configurational sampling, for example interatomic distances within the solutes. These distances are important, as they are linked to the configurational change in the host that depends on the bound anion. The distances between selected atoms (using the same naming scheme as in figure 5.2) are shown in figure 5.8. The same trends are seen in all plots, so only those from the first FEP and first AdUmWHAM simulation are shown. These plots show the same patterns as those in figure 5.7. While key configurational changes, like the increasing hydrogen bond length between the host and halide, are sampled smoothly across λ, more subtle configurations are sampled almost randomly. This is a consequence of FEP only running a single trajectory at each λ-value. Each trajectory samples only a particular local minima, from which it is unable to escape. This is in contrast to the AdUmWHAM trajectories, whose motion along λ induces more configurational exploration, and wider sampling of many local minima. All of the distances are seen to change smoothly with λ for the AdUmWHAM simulations, while some distances from the FEP and FDTI simulations exhibit a random zig-zag.

5.4.3 Summary

FEP, FDTI and AdUmWHAM have been applied to calculate the relative binding free energies of three halides to a calix[4]pyrrole derivative. The results show agreement with the experimentally observed binding order, and the simulation structures reproduce the expected configurational change. However, all three sets of free energies contain significant errors. These are a result of the sampling of the positions of the contaminating water molecules. These waters may or may not be



(a) FEP



(b) AdUmWHAM

Figure 5.8: Distances between atoms during the first FEP and first AdUmWHAM simulations, on the bound leg, fluoride to chloride perturbation. The atom names are taken from figure 5.2. Distances calculated every 500 K steps for the last 7 M steps within each λ-window for FEP, and at the end of every iteration from iteration 50 onwards for AdUmWHAM.

positioned to hydrogen bond with the halide ion. None of the three methods correctly samples or weights these different states. AdUmWHAM either has, or does not have, hydrogen bonding, and the use of λ-moves ensures that all points along λ consistently see the same state. This leads to a wide variation in the predicted relative binding free energies, as AdUmWHAM is unable to average the different possibilities correctly. FEP and FDTI use a single trajectory at each λ-value. This trajectory is randomly hydrogen bonding or not hydrogen bonding, and thus there is no consistency across the λ-coordinate. FDTI and FEP average this effect in an incorrect manner, as the relative weights of the two states are not accounted for, and a particular trajectory across λ cannot freely swap between them. This effect is also seen in the subtle configurational sampling of the host, where distances between certain atoms are not seen to change smoothly with λ. FDTI and FEP get stuck in local minima, and while they sample them well, and return highly converged free energies, their results are not consistent across the whole λ-coordinate, and thus the sum of those results contains an unquantifiable random sampling error. Through the use of a dynamic λ-coordinate, AdUmWHAM imparts more configurational motion to its trajectories. These are able to sample more local minima, and distances between atoms are seen to change smoothly with respect to λ.

5.5 Bivariate Multicanonical WHAM

The previous section applied FEP, FDTI and AdUmWHAM to calculate the relative binding free energies of three halides to a calix[4]pyrrole derivative. These methods experienced difficulties with the sampling of the positions of the contaminating water molecules, which in the case of AdUmWHAM led to a wide spread of results. BMW is derived from AdUmWHAM, so is expected to suffer from the same problems. Unlike AdUmWHAM, at the limit of convergence, BMW samples from a multicanonical ensemble. This ensemble samples energy evenly, and removes all energy barriers within a defined range. There is the possibility that the use of the multicanonical ensemble could improve the diffusion of the contaminating water molecules, and in so doing, help to solve the sampling problems. To see if this were the case, four BMW simulations for each bound and free perturbation

were set up in an equivalent manner to AdUmWHAM. These were then run to calculate the relative binding free energies of the three halides for the calix[4]pyrrole host.

5.5.1 Simulation System

To match the AdUmWHAM simulations, a BMW iteration was performed every 200 K MC steps, with each iteration consisting of 40 K steps of equilibration, and 160 K steps of production. A λ-move was made every 100 normal MC steps, with a maximum change of 0.05, and the value of λ was preserved between iterations. Two trajectories were run per simulation, in an asynchronous parallel manner. The energy-λ histogram consisted of 100 evenly sized bins along λ. The previous simulations had shown that the energy range required for these simulations was a lot wider than that used for the water-methane simulations. This energy range, 300 kcal mol^{-1}, required 300 bins of 1 kcal mol^{-1} spacing to be adequately represented. The use of such a large number of bins caused many problems as large amounts of memory were required for their storage, and by iteration 100, the solution of the WHAM equations took more CPU time than the actual running of the iteration. Because of this, only 170 BMW iterations could be performed for each simulation. The bound leg used an energy range of -20 300 kcal mol^{-1} to -20 000 kcal mol^{-1}, while the free leg used a range of -8100 kcal mol^{-1} to -7800 kcal mol^{-1}. To maintain consistency, the same ranges were used for each mutation. In all other respects, the parameters used for the BMW simulations were identical to those used for AdUmWHAM.

5.5.2 Results

The results for each leg, and the combined relative binding free energy, are shown in table 5.3. These were generated at 25 °C from the last umbrella of each iteration, using an identical procedure as that used for the water-methane simulations. Run 2 of the bound leg, fluoride to bromide perturbation only had 121 iterations, due to its memory requirements causing it to crash prematurely. The result from this simulation is slightly lower than the others, so it will not be included in any

	Leg	Relative Free Energy / kcal mol^{-1}			
		1	2	3	4
Fluoride to Chloride	Bound	33.4	33.7	34.1	30.8
	Free	13.4	22.6	16.3	18.9
	Binding	20.0	11.1	17.8	11.9
Fluoride to Bromide	Bound	35.6	32.9	41.7	42.6
	Free	18.6	16.7	16.6	22.5
	Binding	17.0	16.2	25.1	20.1
Chloride to Bromide	Bound	2.45	2.40	2.50	2.47
	Free	1.25	1.13	1.39	1.29
	Binding	1.20	1.27	1.11	1.18
Closure / kcal mol^{-1}		-4.2	3.8	6.2	7.0

Table 5.3: Relative free energies for the bound and free legs, and relative binding free energies for four applications of BMW to the calixpyrrole system.

discussions of spread or quality.

5.5.3 Analysis

These results show much better agreement with AdUmWHAM than was obtained for the water-methane simulations. The quality of the results is poorer than that from AdUmWHAM, and investigation of the λ-energy sampling (not shown) shows that this is due to the method having not fully converged. This is confirmed by the plots of the convergence of the relative binding free energies (figure 5.9) which still shows slight drift in the results. BMW is again seen to be too ambitious, and not viable compared to AdUmWHAM.

5.5.4 Configurational Sampling

To see if the use of the multicanonical ensemble has solved the problem of water sampling, the distance between the halide and nearest water oxygen was plotted as a function of BMW iteration (figure 5.10). These plots showed the same patterns as seen for the AdUmWHAM simulations, with trajectories either exhibiting, or not exhibiting, hydrogen bonding between the halide and contaminating water. While there was some evidence of diffusion of water, there was not the free, back-and-

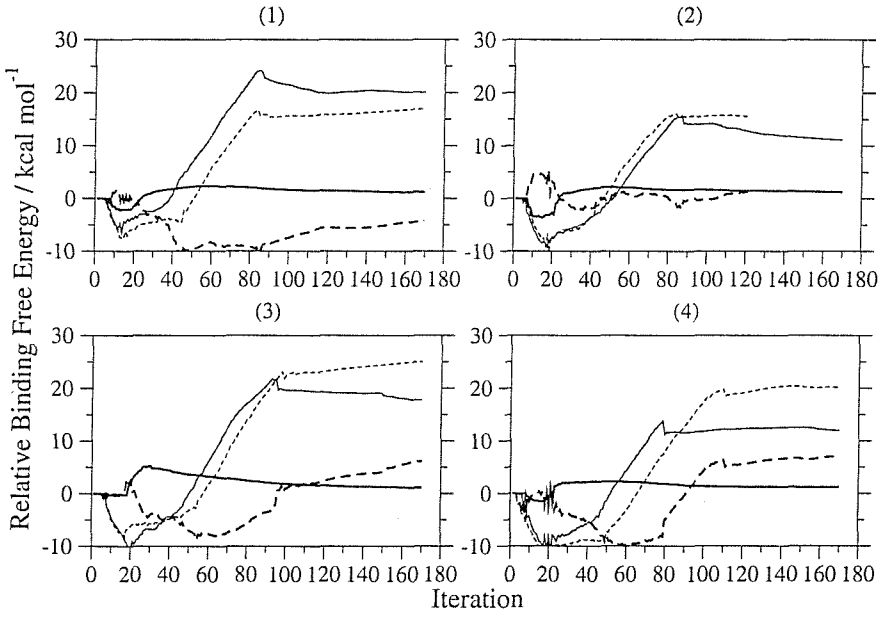


Figure 5.9: Convergence of the fluoride to chloride (solid line), fluoride to bromide (broken line), chloride to bromide (thick solid line) relative binding free energies, and closure (thick broken line), from four BMW simulations of the calixpyrrole system.

forth exchange between hydrogen bonding and non-hydrogen bonding distributions, which would be necessary to properly weight the effects of these two states. Like AdUmWHAM, the spread of the BMW results is seen to almost exactly tally with the presence, or otherwise, of hydrogen bonding with contaminated water.

5.5.5 Summary

The application of BMW to the calculation of the relative binding free energies of halide ions to a calix[4]pyrrole derivative was more successful than that for the relative hydration free energy of water and methane. The results showed good agreement with AdUmWHAM, though still exhibited the same effects of hydrogen-bonding with contaminating water. As in the case of water-methane, the simulations did not converge, and full multicanonical sampling was not obtained. These results confirm that, in its current form, BMW is too ambitious to be applied to calculate relative free energies. This is because the two dimensional λ-energy surface is too large an area to explore within the sampling constraints imposed by the increasing expense of the WHAM equations.

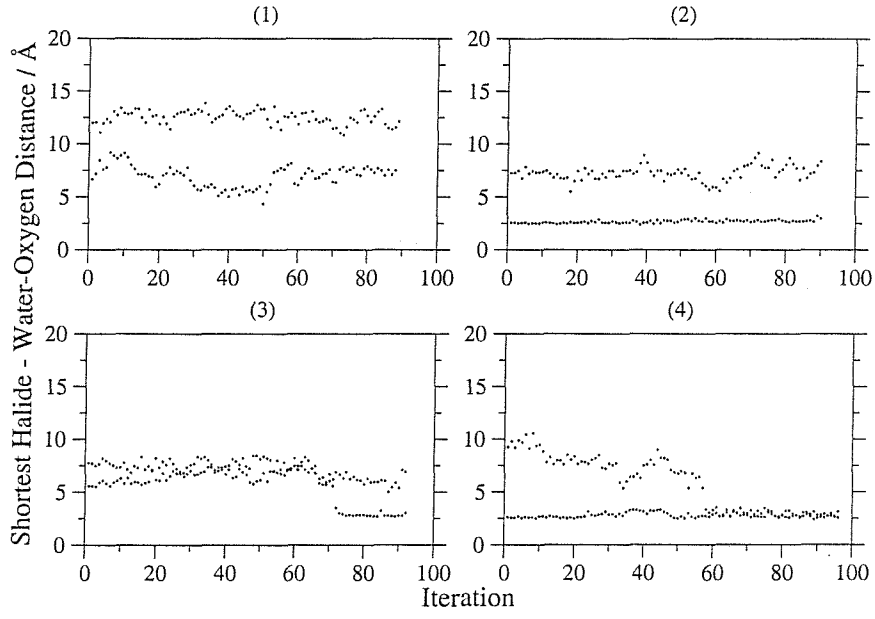


Figure 5.10: The distance between the halide ion and the oxygen of the nearest water molecule for the four BMW simulations on the free leg, fluoride to chloride perturbation.

5.6 Parallel Tempering

PTTI was now applied to calculate the relative binding free energies of the halides to the calix[4]pyrrole derivative. The method is a combination of parallel tempering and FDTI, and uses multiple temperature trajectories at each λ-value to enhance sampling, and reduce random error. The application of PTTI to this system closely mirrored the application of FDTI, with 21 evenly spaced windows along λ, and $\Delta\lambda = 0.001$. Sixteen temperatures were ran at each λ-value, evenly spaced at 5 °C intervals from 10 °C to 85 °C. The same temperatures were used for the bound and free legs of all perturbations to allow the temperature dependence of the relative binding free energies to be determined. Parallel tempering moves were performed every 50 K steps, in an identical manner to the application to the water-methane system. Since a complete set of all mutations requires six free energy simulations (a bound and free leg for each of the three mutations), the costs of PTTI were very substantial. To reduce this cost, only 5 M MC steps were performed for each trajectory. Since even this only reduces the total number of steps to a little over 10 B, the full set of PTTI simulations were only performed once.

	Leg	Relative Free Energy at 25 °C/ kcal mol^{-1}	
		Forwards	Backwards
Fluoride to Chloride	Bound (1)	34.4 (0.9)	34.4 (0.9)
	Bound (2)	34.6 (1.0)	34.7 (1.0)
	Free	16.1 (0.7)	16.1 (0.8)
	Binding (1)	18.3 (1.2)	18.3 (1.2)
	Binding (2)	18.5 (1.3)	18.5 (1.3)
Fluoride to Bromide	Bound	38.9 (1.1)	39.0 (1.1)
	Free	17.4 (0.7)	17.4 (0.7)
	Binding	21.6 (1.3)	21.6 (1.3)
Chloride to Bromide	Bound	3.05 (0.09)	3.05 (0.09)
	Free	1.50 (0.06)	1.50 (0.06)
	Binding	1.55 (0.11)	1.55 (0.11)
Closure (1)		1.7 (1.7)	1.7 (1.7)
Closure (2)		1.5 (1.8)	1.5 (1.8)

Table 5.4: Relative free energies for the bound and free legs, and relative binding free energies as predicted by PTTI.

To allow the consistency of the method to be investigated, the bound fluoride to chloride leg was performed twice.

5.6.1 Results

The results are shown in table 5.4. These are calculated at 25 °C, over the last 3 M steps per trajectory. The standard errors were also calculated, as described in appendix E.

5.6.2 Analysis

These results show reasonable agreement with those from FEP and FDTI. The standard error on the results is slightly larger. This is initially surprising, given that multiple trajectories were run at each λ-value. The spread of results is very good, at only 0.2 kcal mol^{-1}. This is better than standard FDTI, though more duplicate simulations would need to be run to confirm that this is not just a statistical fluke. Plots of the temperature travel for each simulation (figure 5.11) show that

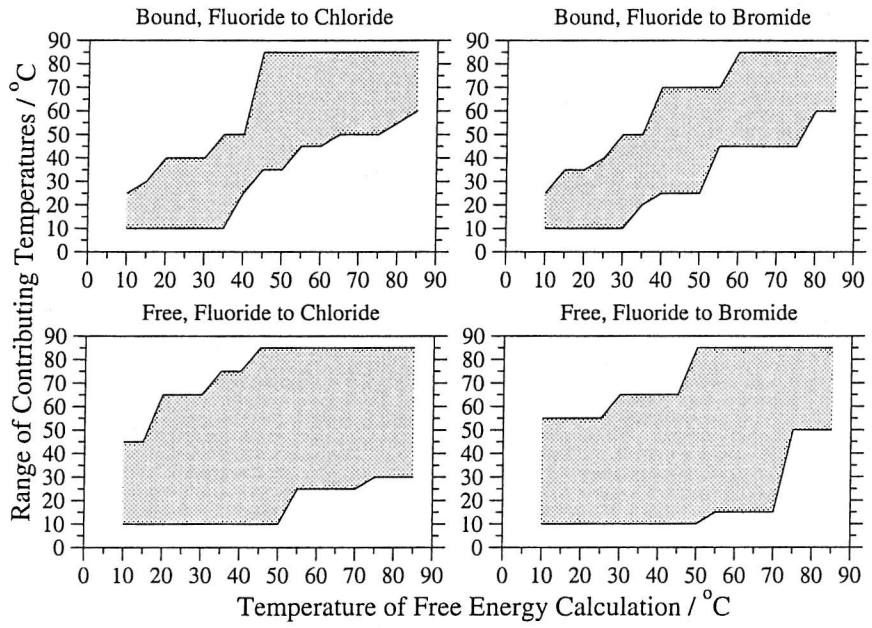


Figure 5.11: The temperature travel for PTTI simulations on the bound and free legs of the fluoride to chloride, and fluoride to bromide perturbations.

the temperature exploration for the bound legs was poorer than that for the free legs. This is due to the increased number of degrees of freedom in the bound legs. Plots of the temperature dependence of the relative binding free energies (figure 5.12) show that the standard error is too large to allow any meaningful trends with respect to temperature to be determined.

5.6.3 Configurational Sampling

The use of PTTI to calculate the relative binding free energies has surprisingly resulted in a larger standard deviation than the applications of FEP or FDTI. This is despite the PTTI simulations using sixteen times the number of Monte Carlo steps. The reason for this observation is made clear in the plots of the distance between the halide and nearest water oxygen as a function of λ (figure 5.13). This shows that hydrogen bonded *and* non-hydrogen bonded trajectories are sampled at each λ-value. Each λ-value has a wide spread of halide-oxygen distances, which results in a wider spread of free energies in the average, and a slightly larger standard error. FEP and FDTI appear to give better converged results as they only sample particular local minima. PTTI samples more water distributions as it allows

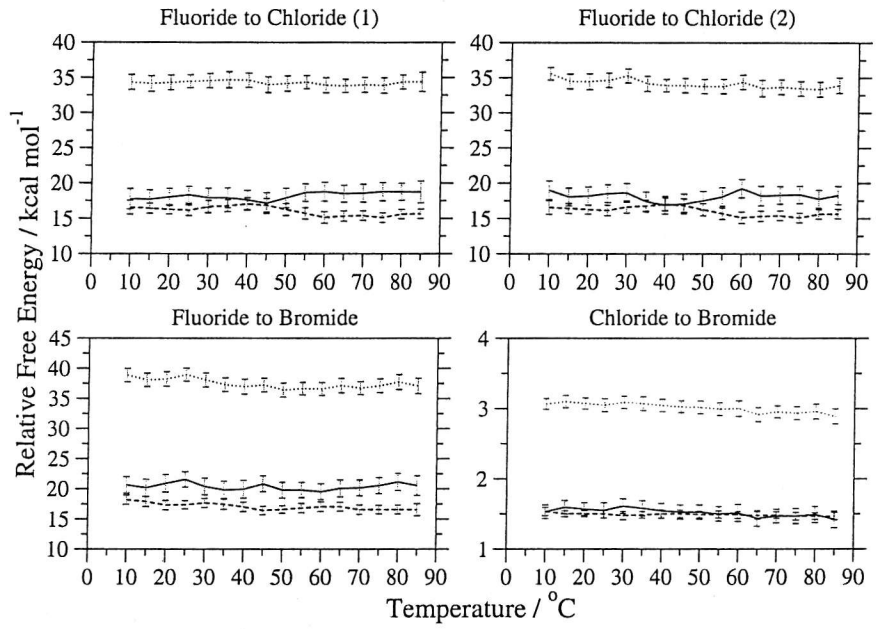


Figure 5.12: The relative free energy with respect to temperature for pairs of halides binding to a calix[4]pyrrole derivative. The relative binding free energy (solid line) is shown, as well as the relative free energies of the bound (dotted line) and free (dashed line) legs.

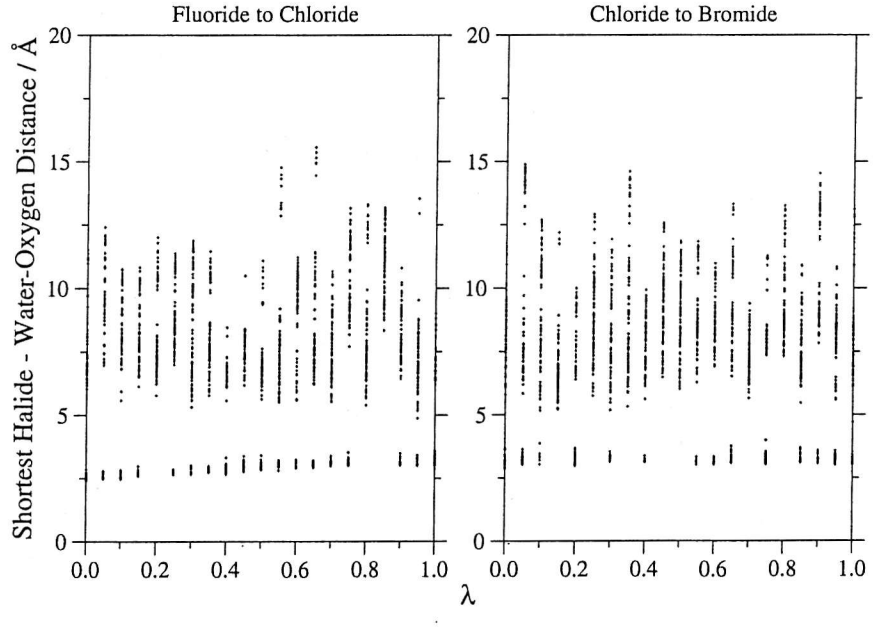


Figure 5.13: The distance between the halide ion and the oxygen of the nearest water, with respect to λ, for the PTTI, free leg simulations on the fluoride to chloride, and chloride to bromide perturbations at 25 °C. The distance is calculated every 50 K steps.

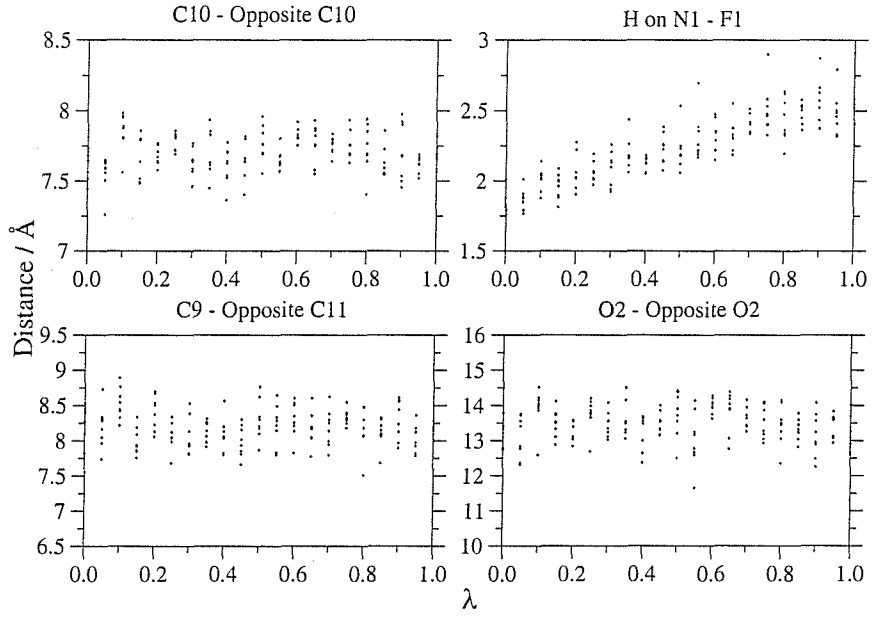


Figure 5.14: Distance between atoms for the first PTTI simulation on the bound leg, fluoride to chloride perturbation. Atom pairs are identical to figure 5.8, and distances are calculated every 500 K steps, for the last 4 M steps per trajectory, at 25 °C.

configurational swapping between temperatures. The set of configurations at each temperature will have a spread of different water configurations, and the temperature swapping allows each λ-value to see most of them. However, the temperature swapping does not seem able to change the initial distribution of water configurations. This means that some λ-values have no sampling from hydrogen bonding trajectories. These are clearly seen as gaps at 2.5 Å in figure 5.13. The sampling across λ was still not consistent, and there was no weighting of the different water configurations.

The use of multiple trajectories at each λ-value is seen to improve the configurational sampling of the host. The plots of interatomic distances versus λ now show smooth progression with respect to λ (figure 5.14).

5.6.4 Summary

PTTI has been applied to calculate the relative binding free energies of three halides to a calix[4]pyrrole derivative. The application of PTTI was computationally demanding, requiring over 10 B MC steps to obtain only one complete set

of relative free energies. The standard errors on the results were larger than either FEP or FDTI. This was a result of the wider configurational sampling brought about through the use of multiple trajectories at each λ-value. While FEP or FDTI tightly converged a particular local minimum, the multiple temperatures of PTTI allowed the exploration of many local minima, leading to a higher quality result. The configurational sampling of the host was seen to be smooth with respect to λ, though the same was not true of the contaminating water distribution. PTTI could not significantly change the contaminating water distribution. This led to holes in the water distance plots, and an inconsistent distribution across λ. While PTTI has improved the results compared to FEP or FDTI, it still has not solved the problem of water distribution, and does not properly weight the relative states.

5.7 Hamiltonian Replica Exchange

The last chapter introduced the combinations of Hamiltonian Replica Exchange with FEP and FDTI. This resulted in two new free energy methods, REFEP and RETI, that improved the predicted relative hydration free energy of water and methane through the simple addition of a periodic λ-swap move. In this move, the coordinates of neighbouring λ-values were periodically tested and swapped. This allowed multiple trajectories to visit each λ-value, and also allowed properties to be determined consistently across the entire λ-coordinate. The results from RETI and REFEP were seen to be very similar, although the use of the λ-swap move did not significantly improve the hysteresis on the REFEP results. The application of REFEP to the calix[4]pyrrole system showed a similar result, namely identical structural sampling to RETI, but similar hysteresis to FEP. Because of this, only the application of RETI to the calix[4]pyrrole system will be discussed in this section.

5.7.1 Simulation Conditions

The application of RETI to the calix[4]pyrrole system was as similar to the application of FDTI as possible. 21 trajectories were evenly spaced every 0.05 along λ, and 10 M steps of simulation were run per trajectory. The value of $\Delta\lambda$ was 0.001,

and a λ-swap move was attempted between neighbouring λ-values, every 50 K steps. The choice of pairs to swap was made in an identical manner to the water-methane simulations. On even iterations, even pairs swapped with the λ neighbour with the larger λ-value, while on odd iterations, odd pairs swapped with the neighbour with the higher λ-value. To enable each λ value to equilibrate before the simulation, an additional 900 K steps of Monte Carlo were run before the start of each simulation. In all other respects, the application of RETI was identical to the application of FDTI. To allow investigation of the reliability of the calculation, four copies of each leg of each perturbation were run, using randomised water positions, and a different random number seed.

5.7.2 Results

The relative binding free energies and relative free energies for each leg of the calculation were calculated, and are shown in table 5.5. These results were calculated over the last 7 M steps per trajectory, and the standard errors calculated as described in appendix E.

5.7.3 Analysis

These results show broad agreement with FEP and FDTI. However, the spread of results is slightly larger, as are the standard errors. In addition, the results are slightly lower than the FEP or FDTI values. The reasons for this are seen in the plots of shortest halide to nearest water oxygen distance, as a function of λ (figure 5.15). These graphs show the features that were expected from the application of RETI, namely multiple water distances for each λ-value, and a consistent set of distances across the entire λ-coordinate. In addition, RETI has been able to weight the different water configurations across λ. The number of hydrogen bonding configurations sampled at low λ is consistently much higher than that sampled at high λ. This corresponds to the higher affinity of fluoride for water than chloride. The plots in figure 5.15 do not show this increasing density clearly, so a histogram of the density of distances with respect to λ was made. The corresponding density plot for figure 5.15 is shown in figure 5.16. This plot shows that not only has RETI

	Leg	Relative Free Energy / kcal mol^{-1}			
		1	2	3	4
Fluoride to Chloride	Bound	34.3 (0.6) 34.3 (0.6)	34.7 (0.7) 34.7 (0.7)	33.9 (0.6) 33.9 (0.6)	34.4 (0.6) 34.4 (0.6)
	Free	18.5 (0.5) 18.5 (0.5)	16.9 (0.4) 16.9 (0.4)	17.5 (0.5) 17.5 (0.5)	17.6 (0.4) 17.6 (0.4)
	Binding	15.8 (0.8) 15.8 (0.8)	17.8 (0.8) 17.8 (0.8)	16.3 (0.8) 16.3 (0.8)	16.8 (0.7) 16.8 (0.7)
Fluoride to Bromide	Bound	37.7 (0.7) 37.8 (0.7)	36.9 (0.7) 37.0 (0.7)	37.1 (0.7) 37.2 (0.7)	38.2 (0.7) 38.2 (0.7)
	Free	20.0 (0.5) 20.0 (0.5)	17.3 (0.5) 17.3 (0.5)	17.5 (0.5) 17.5 (0.5)	16.5 (0.5) 16.5 (0.5)
	Binding	17.7 (0.9) 17.8 (0.9)	19.7 (0.8) 19.7 (0.8)	19.6 (0.8) 19.7 (0.8)	21.7 (0.9) 21.7 (0.9)
Chloride to Bromide	Bound	2.96 (0.04) 2.96 (0.04)	3.03 (0.04) 3.03 (0.04)	3.01 (0.05) 3.01 (0.05)	3.09 (0.05) 3.09 (0.05)
	Free	1.50 (0.03) 1.50 (0.03)	1.49 (0.03) 1.49 (0.03)	1.56 (0.03) 1.57 (0.03)	1.51 (0.03) 1.51 (0.03)
	Binding	1.46 (0.06) 1.46 (0.06)	1.54 (0.05) 1.54 (0.05)	1.44 (0.06) 1.44 (0.06)	1.58 (0.06) 1.58 (0.06)
Closure / kcal mol^{-1}		0.5 (1.2) 0.5 (1.2)	0.4 (1.2) 0.4 (1.2)	1.9 (1.1) 1.9 (1.1)	3.3 (1.1) 3.3 (1.1)

Table 5.5: Relative free energies for the bound and free legs, and relative binding free energies for four applications of RETI to the calixpyrrole system. The results are calculated over the last 7 M steps of simulation. Both the forwards and backwards values are shown, with the standard deviation in parenthesis.

shared the full set of water distributions across λ, but it has also properly weighted those configurations. This weighting is graphically demonstrated by density plots for the bound and free fluoride to bromide, and chloride to bromide perturbations (figure 5.17). Fluoride has a much greater affinity for water than bromide, so the bound and free density plots show that the hydrogen bonding water distributions are weighted towards the fluoride. Chloride and bromide have similar affinities for water, so the bound and free plots show that the hydrogen bonding distributions are shared evenly across the entire λ-coordinate. The effect of this is to reduce the relative binding free energy of fluoride to the other halides, as the fluoride has a greater probability of being hydrogen bonded to water. The RETI simulations have performed much better than any of the other methods at including this significant effect. However, RETI has *not* helped each individual trajectory to sample the water distributions. All it has done is to distribute the original water configurations

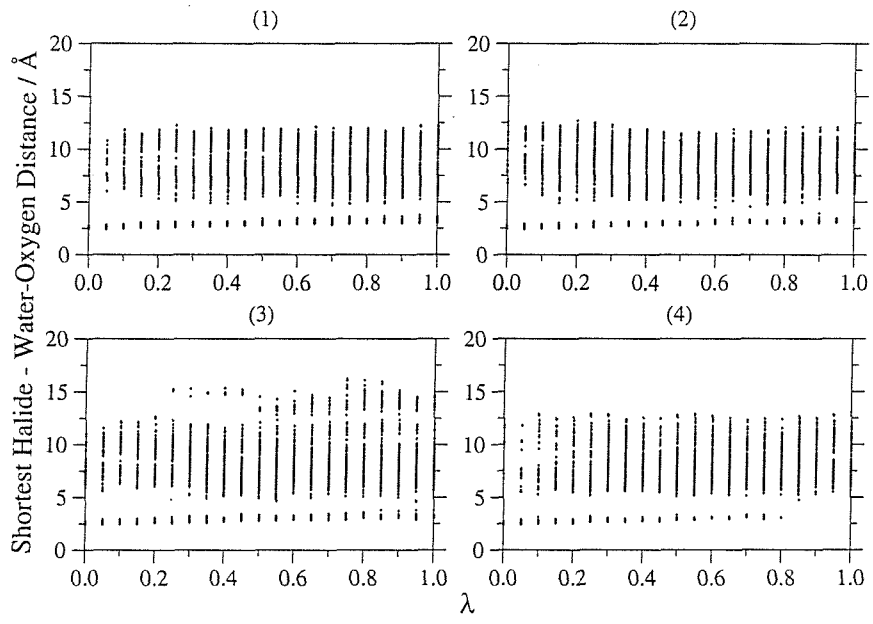
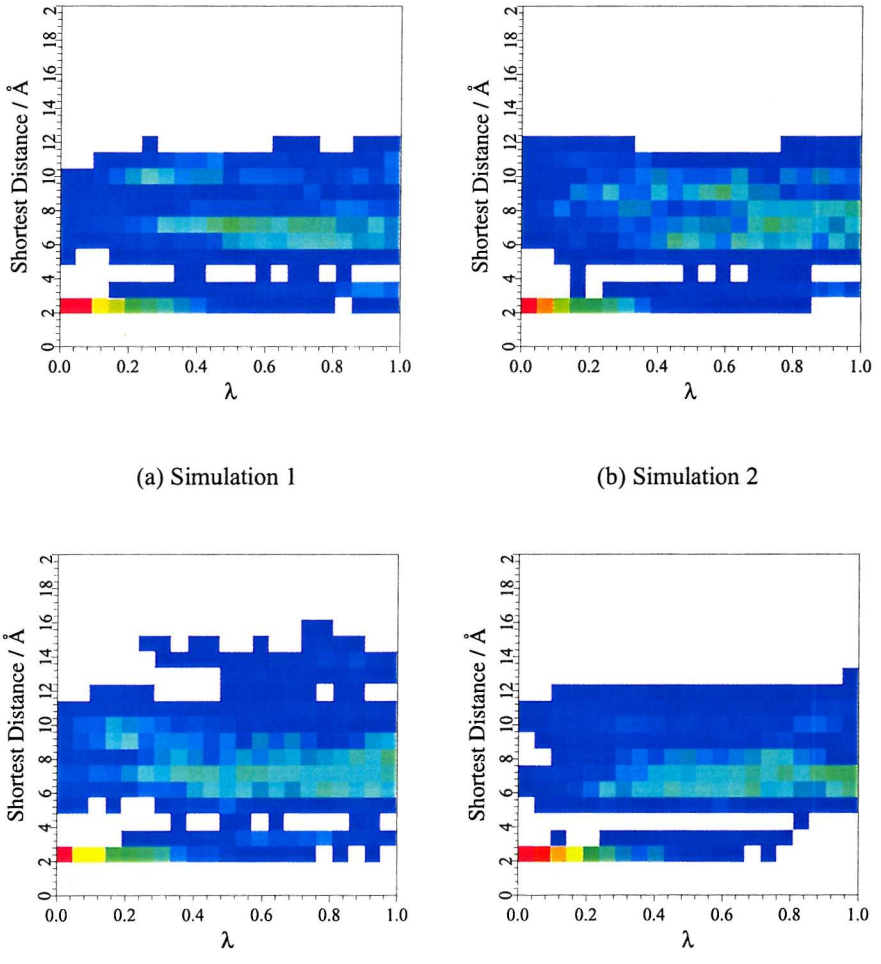


Figure 5.15: The distance between the halide ion and the oxygen atom of the nearest water, as a function of λ, for four applications of RETI to the free leg, fluoride to chloride perturbation. The distances were calculated every 50 K steps.

across λ correctly. RETI has thus handled the water configurations it started with in the best possible manner. There is still a random error in the results associated with the initial water distribution, which is likely to be the source of the spread of results. To make an improvement, new methods must be developed that can effectively sample multiple water distributions. If these methods were combined with RETI, then we believe that the correct relative binding free energies, *including* the effects of contaminating water distribution, could be achieved. Such methods could perhaps use a fourth dimension[126] to soften the water-system interactions, and allow better diffusion of water through the DMSO solvent. Such methods are the subject of future work on this system.

While RETI could not change the initial distribution of water configurations, the use of a dynamic λ-coordinate has given increased configurational flexibility to the calix[4]pyrrole host. Plots of the interatomic distances (figure 5.18) show a wide range of configurational sampling, coupled with smooth and consistent changes in distance across the λ-coordinate.



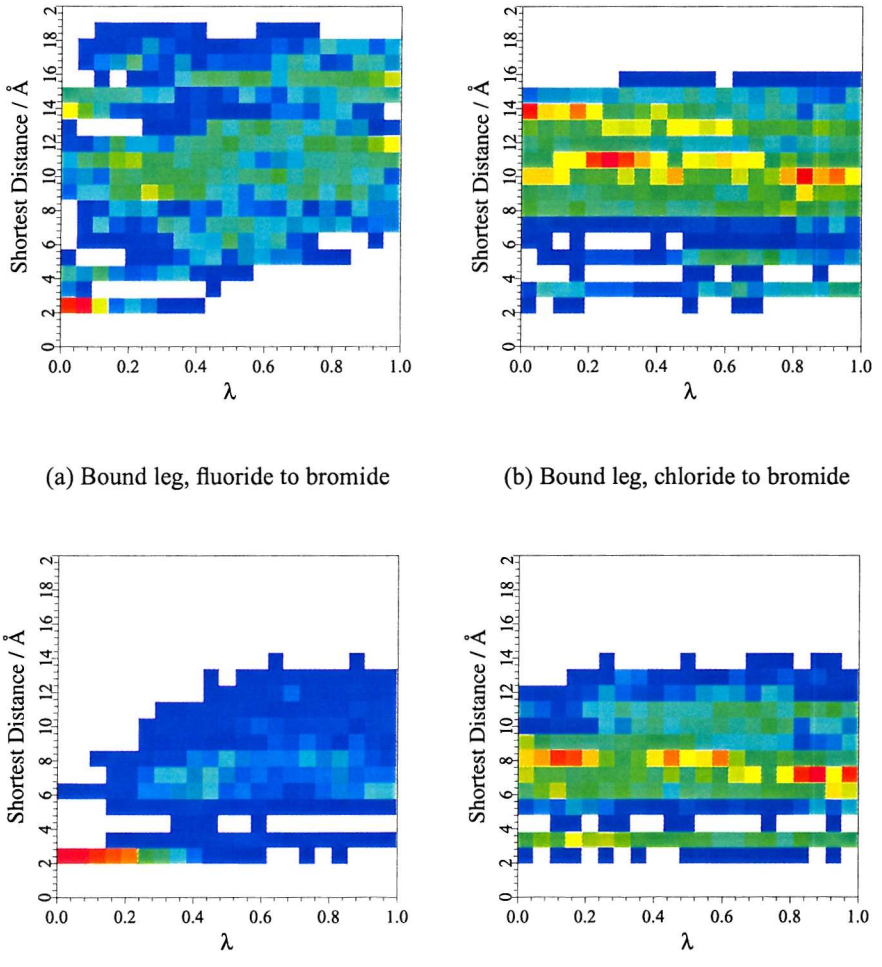
(a) Simulation 1

(b) Simulation 2

(c) Simulation 3

(d) Simulation 4

Figure 5.16: Equivalent plots to figure 5.15, except that the density of points is histogrammed in bins for each λ-value, for every 1.0 Å along halide-oxygen distance. The colour scale goes from no sampling (white), through some sampling (blue) and medium sampling (cyan - green), to high sampling (red).



(a) Bound leg, fluoride to bromide

(b) Bound leg, chloride to bromide

(c) Free leg, fluoride to bromide

(d) Free leg, chloride to bromide

Figure 5.17: Density of halide - nearest water oxygen distances across λ, for RETI simulations on the bound and free legs of the fluoride to bromide, and chloride to bromide perturbations. The colour scale and histogram parameters are identical to figure 5.16.

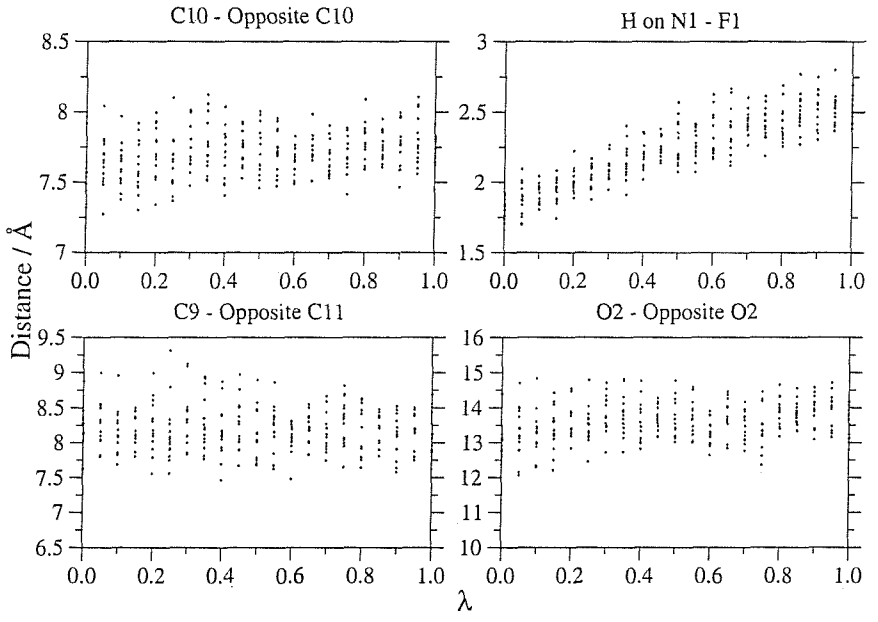


Figure 5.18: Distance between atoms for the first RETI simulation on the bound leg, fluoride to chloride perturbation. Atom pairs are identical to figure 5.8, and distances are calculated every 500 K steps, for the last 7 M steps per trajectory.

5.7.4 Summary

The newly developed technique of RETI was applied to calculate the relative binding free energies of three halides to a calix[4]pyrrole derivative. The results were excellent, with the λ-swap move not only sharing configurational information consistently across the entire λ-coordinate, but also properly weighting the different configurations. RETI was much better able to include the effects of the increased affinity of fluoride for the contaminating waters, compared to the other methods. While RETI was unable to change the initial water distributions, as it does not aid water diffusion, it did ensure that the initial set was distributed correctly to each λ-value. If this method were combined with a technique to aid water diffusion, then we believe that it could reliably obtain the correct relative binding free energies that included the effects of contaminating water.

RETI did an excellent job of sampling calix[4]pyrrole configurations. As in the case of AdUmWHAM, the use of a dynamic λ-coordinate imparted increased configurational flexibility to the host. The increased configurational sampling was then correctly weighted over the entire λ-coordinate via the λ-swap moves.

5.8 Conclusion

Seven free energy techniques were applied to calculate the relative binding free energies of three halide ions to a fluoride-selective calix[4]pyrrole derivative. The results showed that the sampling of contaminating water molecules was critical to predict the correct values. FEP and FDTI failed to sample the water distributions correctly, and used an inconsistent set of water distributions across the λ-coordinate. Each trajectory across λ sampled only a specific configuration. This led to well-converged free energies, which did not include the correct weighting of water, or indeed host conformer, distributions. Since FEP and FDTI used an inconsistent set of distributions across λ, they returned well-converged, yet incorrect, results.

Through the use of a dynamic λ-coordinate, AdUmWHAM and BMW exhibited better configurational sampling than FEP or FDTI. They too suffered from problems sampling the contaminating water distribution, which in their case led to a wide spread of results. This was due to the methods not weighting the relative distributions properly, and instead returning hydrogen bonding, or non-hydrogen bonding free energies.

PTTI used multiple trajectories at multiple temperatures to partially overcome the sampling problem. While the method improved configurational sampling, it failed to enhance water diffusion, and was also not able to obtain a consistent water distribution across λ. This was achieved by the techniques of REFEP and RETI. The use of λ-swap move allowed not only multiple trajectories to sample each λ-value, but also the consistent, and weighted set of water distributions to be obtained across λ. REFEP did not improve the problems of hysteresis inherent in FEP. RETI, on the other hand, had little to no hysteresis, consistent sampling across λ, and weighted water distributions. These results clearly demonstrate that from the seven methods tested, RETI was by far the best choice for the calix[4]pyrrole system. However, improvements could still be made by developing methods that encourage greater diffusion of the contaminating water. Such methods are a subject of current and future work.

Chapter 6

The Application of Free Energy Methods to a Protein-Ligand System

6.1 Introduction

The previous chapters have developed and tested several new free energy methods, and compared them to established techniques. These tests were applied to two systems; the relative hydration free energy of water and methane, and the relative binding free energies of halides to a calix[4]pyrrole derivative. These tests revealed that, of the established free energy methods, Finite Difference Thermodynamic Integration (FDTI) consistently performed better when ran under similar conditions as would be necessary for robust rational drug design. These tests further demonstrated that, out of all tested methods, the newly developed technique of Replica Exchange Thermodynamic Integration (RETI) outperformed even FDTI, and proved itself to be the method of choice for both of the test systems.

The overall aim of this research is to develop free energy methods that may be routinely applied to protein-ligand systems. The new technique of RETI will thus be compared to the established technique of FDTI, on a challenging protein-ligand system. The protein chosen for this study was p38 MAP kinase.

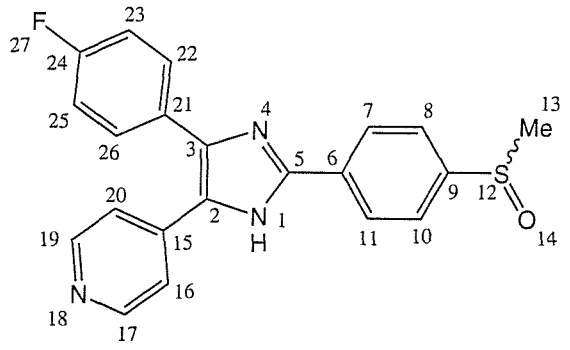


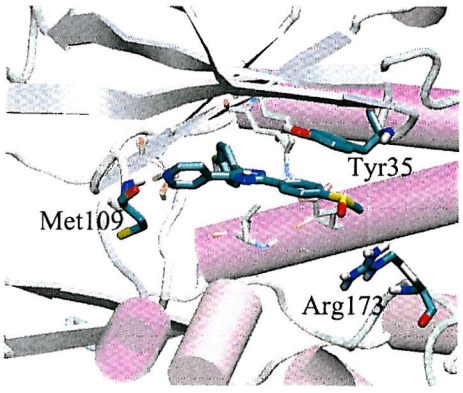
Figure 6.1: The SB1 inhibitor to p38 MAP kinase.

6.2 p38 MAP kinase

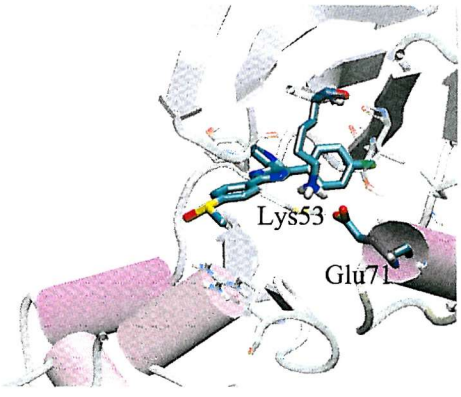
The mitogen-activated protein (MAP) kinases are important potential drug targets for the treatment of inflammation, cancer, leukaemia and other diseases.[127] This is because they play an important part in the regulation of diverse cellular events,[127,128] such as cell transformation, proliferation, differentiation and apoptosis. p38 MAP kinase is known to regulate the release of tumour necrosis factor (TNF)α and interleukin (IL)-1β.[129] These pro-inflammatory cytokines have been implicated as causal agents in the bone and joint destruction characteristic of rheumatoid arthritis.[129] p38 is thus an important therapeutic target for the treatment of rheumatoid arthritis, a condition which effects over 2 M people in the United States alone.[129] There has been a lot of interest in p38,[127–131] and many potential ligands have been developed. A good review of current p38 ligands[132] shows that all of the major pharmaceutical companies are actively pursuing this target. While rheumatoid arthritis was the initial focus of this research, the involvement of p38 in a broader spectrum of disease has become clear.[132] Initial findings suggest that p38 inhibitors could be used to treat a range of conditions, from stroke and Alzheimer's through to osteoarthritis and dermatitis.[132]

SmithKline Beecham have developed an effective and selective group of inhibitors to p38. Of these "SB" inhibitors, SB1 (full name SB203580) (figure 6.1) showed both high affinity and high specificity for the ATP binding site of p38.[127] A range of related ligands have been developed, and their binding affinities to p38 measured.[130] This chapter will investigate the binding of SB1 and related ligands

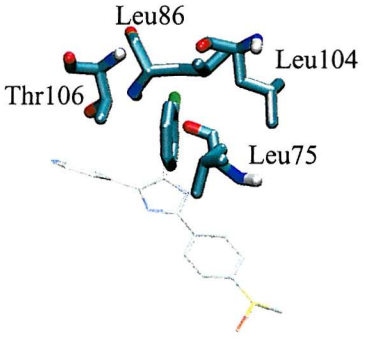
(a) SB1 binds in the central cleft.



(b) H-bonding to Met109, and stacking with Tyr35.



(c) H-bonding to Lys53 induces salt bridge with Glu71.



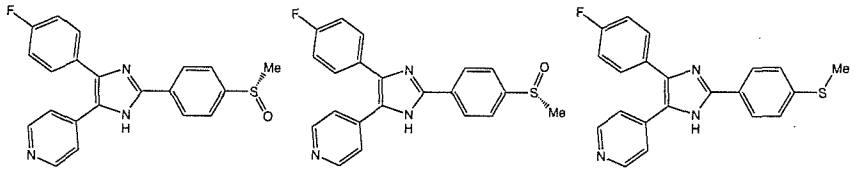
(d) Fluorobenzene ring fits into a hydrophobic pocket.

Figure 6.2: Different views of the crystal structure of p38 with SB1 (pdb code 1A9U, resolution 2.5 Å, collected at pH 7.4). Four different views show the specific interactions important for binding.

to p38, and compare the results to these measured binding affinities.

6.2.1 SB1-based Ligands

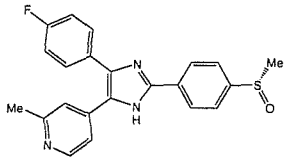
SB1 shows strong affinity and specificity for the ATP binding site of p38.[127] The X-ray crystal structure[127] of the complex between p38 and SB1 is shown in figure 6.2. This structure shows SB1 bound in the ATP binding cleft in the centre of the protein. Like ATP, the pyridine nitrogen atom (N18) forms a hydrogen bond with the backbone NH group of Met109.[127] The formation of a hydrogen bond between N4 on SB1 and Nε of Lys53 suggests that the solvent exposed nitrogen, N1, of the



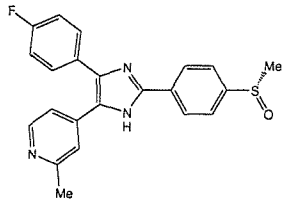
Ligand A, SB1(S) (0.042 μM)

Ligand B, SB1(R) (0.042 μM)

Ligand C, (0.34 μM)



Ligand D, (0.37 μM)



Ligand E, (0.37 μM)

Figure 6.3: The structures of potential ligands to p38. The IC_{50} of each ligand (collected at pH 7.4) is shown in parenthesis.[130]

imidazole ring is protonated.[127] The formation of this hydrogen bond induces a salt bridge between Lys53 and Glu71 which is not present in the uncomplexed form of p38.[127] A stacking interaction is formed between SB1 and Tyr35, and the fluorobenzene ring is buried in a hydrophobic pocket of Thr and Leu groups.[127]

This chapter will investigate the binding of five potential ligands to p38 (figure 6.3). Since the crystal structure was formed from a racemic mixture of SB1,[127] it is unknown whether the binding is enantioselective. While the crystal structure does assign stereochemistry to the S-enantiomer, the accompanying paper does not discuss how this was achieved, nor whether the binding was enantiomerically selective. To investigate this, both the S- and R-enantiomers will be modelled (ligands A and B in figure 6.3). This study will also examine two analogues of SB1. Ligand C, in figure 6.3, is a reduced form of SB1, where the sulfoxide group has been converted into a sulfide. This ligand has a low binding affinity for p38, and has an almost order of magnitude difference in IC_{50} compared to SB1. Ligand D is a methylated version of SB1. This has also resulted in an almost order of magnitude difference in IC_{50}, compared to SB1. The methyl group has been added to the pyridine ring, which is configurationally restrained by the fluorobenzene ring. This

means that there are two distinct conformers of this molecule; ligand D, which has the methyl group pointing inwards, and ligand E which has the group pointing out.

Four perturbations are needed to explore the relative free energies of these five ligands. The first two, A $\rightarrow$ C and B $\rightarrow$ C remove the oxygen atom of the sulfoxide group to form the sulfide. The remaining two perturbations, D $\rightarrow$ A and E $\rightarrow$ A, remove the methyl group from the pyridine ring.

6.3 Parameterisation

The p38-SB1 ligand system was modelled via the OPLS united-atom forcefield. The crystal structure of the SB1-complex (pdb code 1A9U)[127] was used as a basis for all perturbations. The structure was checked via WhatIf[133] and polar hydrogens added based on optimisation of the hydrogen bond network. The titratable residues were also set to the state recommended from this analysis. The residue list, and titration states used are shown in appendix F.1. The PEPZ program, from MCPRO,[13] was used to build a z-matrix of the protein based on the crystal structure. The z-matrix kept most of the protein rigid, with only the side-chains within 15 Å of the centre of SB1 able to move. The flexible residues are also listed in appendix F.1.

The z-matrices of the five ligands were constructed by treating each aromatic ring as a rigid residue. The only flexible degrees of freedom were the dihedrals and angles which connected these rings to each other, and all the non-bond degrees of freedom within the sulfoxide / sulfide functional group. The z-matrices used for the A $\rightarrow$ C and D $\rightarrow$ A perturbations are given in appendix F.2.4. The A $\rightarrow$ C and B $\rightarrow$ C perturbations involved the disappearance of an oxygen atom. This was pulled in as λ increased, such that by $\lambda = 1.0$, its bond length to the sulfur atom was 0.3 Å. As in the case of the water-methane simulations, this was to avoid convergence problems as the atom disappeared. A similar technique was used for the disappearance of the methyl group for perturbations D $\rightarrow$ A and E $\rightarrow$ A. In these cases, the hydrogens were pulled in while being turned into dummy atoms, and the methyl carbon was turned into an aromatic hydrogen.

The parameters for the protein were available from the OPLS united-atom

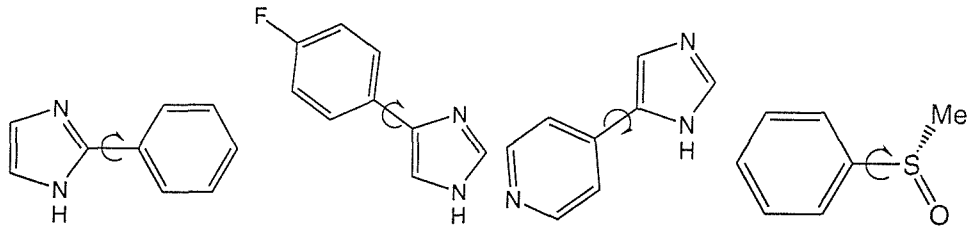
forcefield.[134] The ligand's Lennard Jones, angle and dihedral parameters were taken from the OPLS all-atom forcefield.[6, 10, 135] OPLS lacked appropriate dihedral parameters between the imidazole and phenyl rings, so these had to be individually derived. The partial charges on all of the ligands also had to be derived.

6.3.1 Deriving Partial Charges

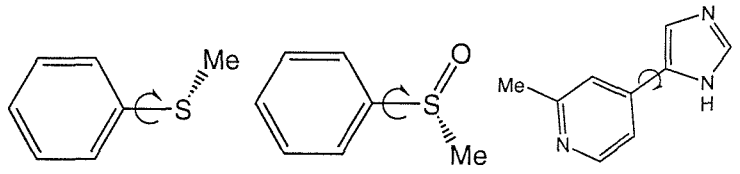
The charge parameterisation scheme for the OPLS forcefield can be very expensive, and requires fitting to known physical properties of the molecule in question.[6] Initial investigations[136] into combining OPLS charges with Quantum Mechanical charges suggested that Hartree-Fock 6-31G*,[4] electrostatic potential surface derived partial charges were appropriate. This removes the need for known physical properties of the system, but this calculation is very expensive, and can encounter problems associated with buried atoms.[137] Recently, the use of semi-empirical Quantum Mechanics methods has been proposed.[138] The AM1 semi-empirical Hamiltonian[139] can efficiently produce atomic partial charges via a Mulliken population analysis.[4] These charges will contain errors, so Storer *et al.* developed empirical correcting functions to produce what they call "AM1-CM1A" charges.[138] The correcting functions were optimised by fitting the AM1 Mulliken charges to the experimental dipole moments of a diverse set of 186 molecules.[138] The resulting charges were compared to high level QM partial charges, and seen to show good agreement.[138]

AM1-CM1A charges were subsequently used by Jorgensen[27,71,125] to obtain OPLS partial charges for new atoms. To make these gas-phase charges compatible with the OPLS condensed phase parameters, Jorgensen scaled them by an α factor. There does not seem to be an optimum value of α, but a similar study to this[71] used $\alpha = 1.08$.

The work presented in this chapter used Amsol 6.6[140] to obtain the AM1-CM1A charges for each of the five ligands. Full geometry optimisation was used, and the charges over symmetrical groups were averaged. The resulting charges were scaled by an α factor of 1.08, then rounded to 3 decimal places (to be con-



Fragment 1 Fragment 2 Fragment 3 Fragment 4



Fragment 5 Fragment 6 Fragment 7

Figure 6.4: Fragments used to obtain the torsional parameters for the five SB1-based ligands.

sistent with OPLS). To correct any rounding error, the sum of charges on each ligand was set to 0.000 by adding or subtracting 0.001 to the first few atoms. The resulting partial charges are given in appendix F.2.2.

6.3.2 Deriving Torsional Parameters

The torsional parameters were derived by constructing representative fragments of each ligand (figure 6.4). Each fragment represented only one of the missing torsions, and was built to ensure that it had as few internal torsions as possible. Gaussian94 was used on each fragment to obtain the quantum mechanical energy as a function of dihedral angle. The dihedral angle of interest was varied in 10 ° increments, from 0 ° to 180 °, and a full, RHF/6-31G* geometry optimisation was performed for each value. A z-matrix was built for each fragment, based on the z-matrices for the full ligands. The Lennard Jones, angle and available dihedral parameters were applied to each fragment from the full ligands, and the AM1-CM1A charges were calculated and scaled as in section 6.3.1. The fragments were missing the parameters for the dihedrals of interest, so these were set to zero. Monte Carlo minimisation was applied for each dihedral angle value between 0 and 180 °, and the difference between the quantum mechanics and molecular mechanics en-

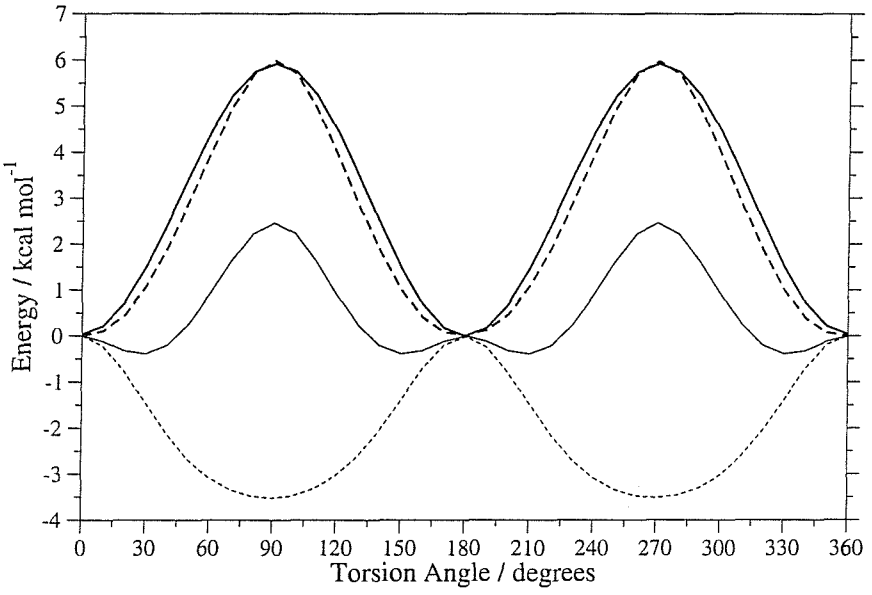


Figure 6.5: The dihedral profiles calculated for fragment 3 in figure 6.4. The thin solid line shows the quantum mechanical profile for this torsion, while the thin dashed line shows the molecular mechanics profile. The difference between the two (thick solid line) is the dihedral angle energy, to which the Fourier terms of equation 2.6 are fit (thick dashed line).

ergies were calculated (figure 6.5). This difference represented the dihedral energy profile of interest. This profile was extrapolated to 360 °, and the Fourier terms of equation 2.6, from chapter 2, were fitted to it via a single value decomposition algorithm[46] (figure 6.5). The quality of each fit was inspected visually, and the dihedral parameters were manually tweaked if necessary. The resulting dihedral parameters are given in appendix F.2.3.

6.4 Simulation Conditions

The simulations of the five SB1-like ligands to p38 were set up in a similar way to that of another simulation study on p38 ligands,[35] and to a similar set of simulations on COX-1 and COX-2 inhibitors.[27] The active site was surrounded by a spherical cap of 1024 TIP4P[89] waters. This was centred on ligand A, and had a radius of 25 Å. The water cap was restrained to the 25 Å barrier through use of a single-sided harmonic restraint, with a 1.5 kcal mol^{-1} force constant. The spherical water cap was modified to include all of the crystallographic waters that were

Parameter	Value
Ensemble	Canonical (NVT)
Simulation Temperature	25 °C
Boundary Type	Solvent cap, centred on ligand, radius of 25 Å, force constant 1.5 kcal mol^{-1}.
Protein / ligand / solvent move ratio	40 / 1 / 40
Non-bonded cutoff	15 Å
Maximum ligand translation / rotation	0.1 Å / 2.0 °
Maximum solvent translation / rotation	0.15 Å / 15 °
Preferential sampling centre	Ligand, imidazole group
Preferential sampling parameter (WKC)	200
Equilibration Scheme	10 M steps of water-only moves, followed by 1 M of protein-only moves, followed by 5 M steps of general equilibration

Table 6.1: Parameters used to control the equilibration of the p38 system.

within the 25 Å radius. Non-bonded cutoffs were set to 15 Å, and residue-molecule cutoffs were used. This meant that if any pair of atoms between a protein residue and ligand or solvent molecule were within the cutoff radius, then all pair-pair interactions between the residue and molecule were included. The non-bonded interactions were feathered over the last 0.5 Å via a harmonic switching function.[12] Preferential sampling[91,92] was used to enhance solvent sampling around the ligand. Initial coordinates for the ligand were taken by mapping to the crystal structure of ligand A bound to p38. This system was equilibrated for 16 M MC steps according to the parameters in table 6.1. This was to ensure that the waters in the solvent sphere were fully relaxed. The final structure from equilibration was then used as the initial structure for all of the bound leg perturbations. The free leg was set up in a similar way, using a 25 Å solvent cap of 2177 TIP4P waters, centred on SB1. To relax the water structure, 10 M steps of general equilibration were run using the parameters from table 6.1, and a solvent / ligand move ratio of 295 to 1. The final structure from this equilibration was then used as the starting point for all of the free leg simulations.

6.5 Free Energy Simulations

The five ligands in table 6.3 were investigated via four perturbations; A $\rightarrow$ C, B $\rightarrow$ C, D $\rightarrow$ A and E $\rightarrow$ A. FDTI and RETI simulations were run on each bound and free leg, and the resulting free energies combined via the cycle in figure 2.2. The applications of FDTI and RETI were as similar as possible, and both used 21 evenly spaced windows across λ. For the bound leg, 15 M steps of simulation were run at each window, over 300 blocks of 50 K steps. The free legs used 10 M steps at each window, over 200 blocks of 50 K steps. A λ-swap move was attempted between neighbouring pairs every 50 K steps during all of the RETI simulations. All simulations used a value of $\Delta\lambda$ of 0.001. To investigate reliability, all simulations were repeated twice using different random number seeds. To ensure that the RETI trajectories were equilibrated for each λ-value, an extra 500 K steps of simulation were run before the start of each trajectory. The remaining simulation parameters were identical to those used during equilibration.

6.5.1 Developing a New Monte Carlo Program

As they stand, using MCPRO, these simulations would have taken about 15 to 20 days each to run. Part of the aim of this research is to make free energy calculations sufficiently fast, such that they may be routinely used in commercial drug discovery. Since the use of MCPRO did not satisfy this aim, a completely new Monte Carlo program was written to perform these free energy simulations. This code was written to have the main features of MCPRO, and was tested to ensure that it could reproduce MCPRO's free energies. Unlike MCPRO, the internal data structure of the code was specifically optimised to run large protein-ligand simulations. This optimisation led to an efficient non-bonded energy routine, which allowed the code to run between 9 to 12 times faster on the p38-SB1 system. This allowed each of the bound leg simulations to be run within 1.5 days, over 21 PIII-1GHz processors. Since the calculation is processor limited, we believe that a similar number of current P4-3GHz processors could accomplish the same simulation in approximately half a day. This is a significant difference to the amount of time these calculations took at the start of this project, when our available computa-

		FDTI / kcal mol^{-1}		RETI / kcal mol^{-1}	
Perturbation	Leg	1	2	1	2
A → C	Bound	24.5 (0.5)	25.1 (0.5)	23.7 (0.6)	24.5 (0.6)
	Free	25.3 (0.6)	26.4 (0.6)	27.1 (0.6)	25.7 (0.6)
	Binding	-0.8 (0.8)	-1.3 (0.7)	-3.4 (0.9)	-1.1 (0.9)
B → C	Bound	27.4 (0.5)	26.9 (0.5)	28.5 (0.6)	27.9 (0.6)
	Free	25.3 (0.6)	26.4 (0.6)	27.1 (0.6)	25.7 (0.6)
	Binding	2.0 (0.8)	0.5 (0.8)	1.4 (0.9)	2.2 (0.9)
D → A	Bound	-5.8 (0.4)	-9.3 (0.4)	-7.1 (0.4)	-7.5 (0.4)
	Free	4.7 (0.3)	5.0 (0.3)	5.6 (0.3)	4.7 (0.3)
	Binding	-10.5 (0.4)	-14.3 (0.5)	-12.6 (0.5)	-12.2 (0.5)
E → A	Bound	8.4 (0.2)	8.1 (0.2)	8.2 (0.2)	8.5 (0.2)
	Free	5.4 (0.3)	4.5 (0.3)	5.5 (0.3)	5.7 (0.3)
	Binding	3.0 (0.3)	3.5 (0.3)	2.7 (0.3)	2.8 (0.3)

Table 6.2: Binding free energies for two applications of FDTI and RETI to the binding of ligands to p38. Only the forwards free energies are shown. The free leg for C → B is taken from the equivalent perturbation of A → B. The free energies are calculated over the last 5 M steps of simulation in each leg, and the standard errors are calculated over blocks of 500 K steps.

tional resources were limited to 8 PII-300MHz processors. Using this equipment, and MCPRO, we estimate that each bound leg would have taken approximately 150 days. In reality, the large Beowulf cluster at Southampton allowed several bound leg simulations to be run simultaneously. At its peak, 8 bound legs were being simulated, with over 168 processors simultaneously in use. One and a half days of simulation today achieved as much as about three and a third years of simulation at the start of this project. It is our belief that the combination of our new code with modern Beowulf clusters means that free energy calculations *are now sufficiently fast* to be used in commercial drug discovery.

6.5.2 Results

The results from the two applications of FDTI and RETI to each perturbation are shown in table 6.2. The relative binding free energies, and standard errors are calculated as in appendix E, and for clarity, are also shown in figure 6.6. Owing to the negligible hysteresis on the results, only the forwards free energies are shown.

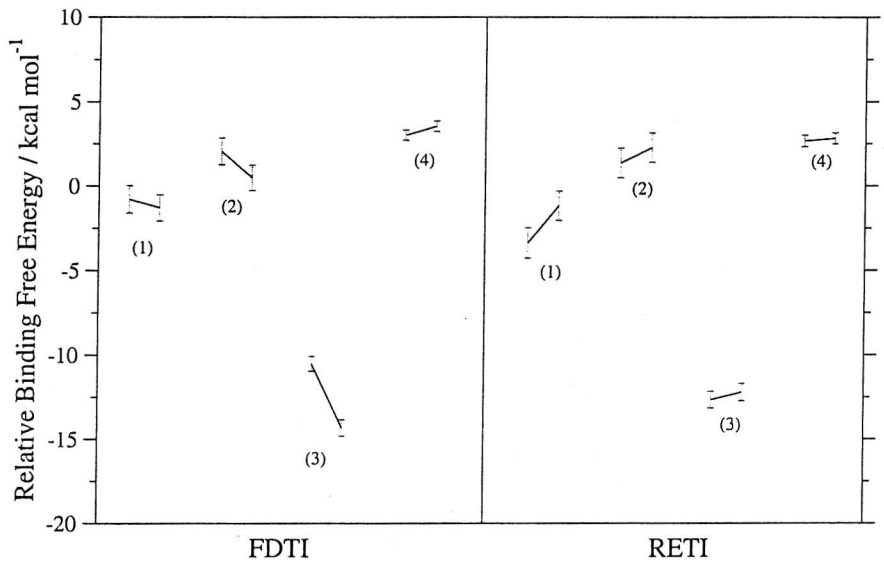
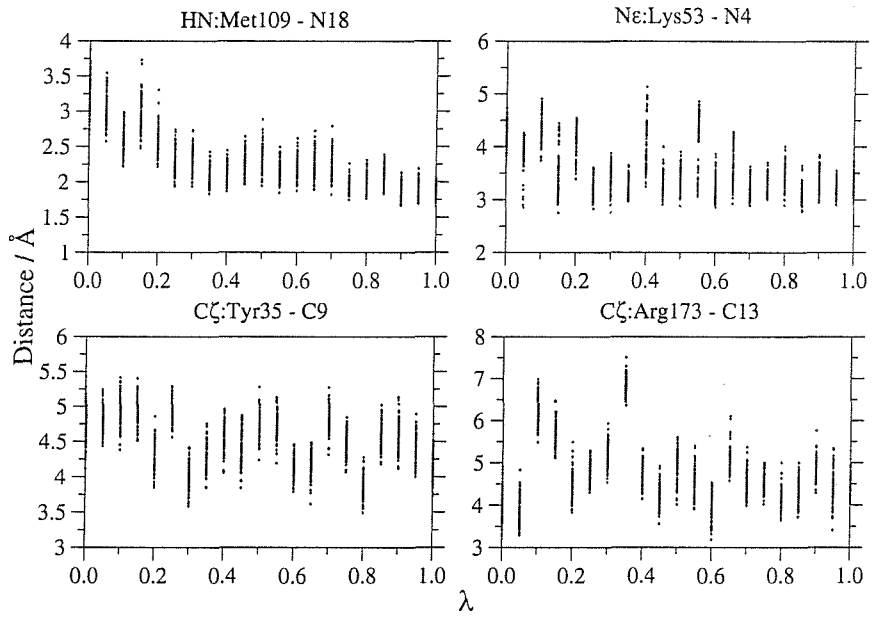


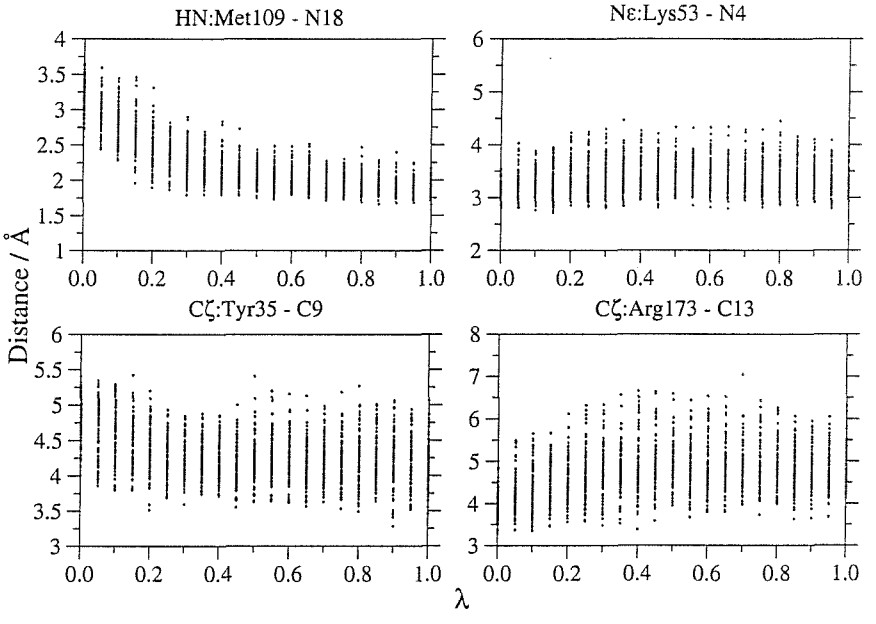
Figure 6.6: Binding free energies for two applications of FDTI and RETI to the binding of ligands to p38. (1) A → C, (2) B → C, (3) D → A, (4) E → A.

6.5.3 Analysis

FDTI and RETI return very similar results, with similar levels of spread, and similar standard errors. The standard errors on FDTI are slightly smaller than those on RETI due to FDTI sampling only single local minima at each λ-value. This is seen in plots of the protein-ligand distances across λ (figure 6.7) and plots of the average interaction energies across λ (figure 6.8). As in all previous simulations, RETI is seen to sample more widely, and more consistently across λ, and again demonstrates its superiority over FDTI. The benefit of wider and more consistent sampling is seen most clearly for the calculation of the relative free energies of ligands D and A. FDTI predicts two, very different free energies for the bound leg; -5.8 kcal mol^{-1} and -9.3 kcal mol^{-1}. In comparison, RETI returns two very similar free energies, that are approximately the average of the two FDTI results; -7.1 kcal mol^{-1} and -7.5 kcal mol^{-1}. Figure 6.8 shows the average electrostatic protein-ligand energy for the bound leg of this perturbation, for both the FDTI and RETI simulations. FDTI is seen to show no clear trend with respect to λ, and does not really return what would be expected, given that the structural sampling (figure 6.7) shows that a hydrogen bond between N18 and HN:Met109 is being formed



(a) FDTI



(b) RETI

Figure 6.7: Distance between selected p38 atoms and selected ligand atoms, for the first FDTI and first RETI simulation on the bound leg, D → A perturbation. The distances were calculated every 50 K steps, for the last 5 M steps of simulation. The ligand atom names are taken from figure 6.1

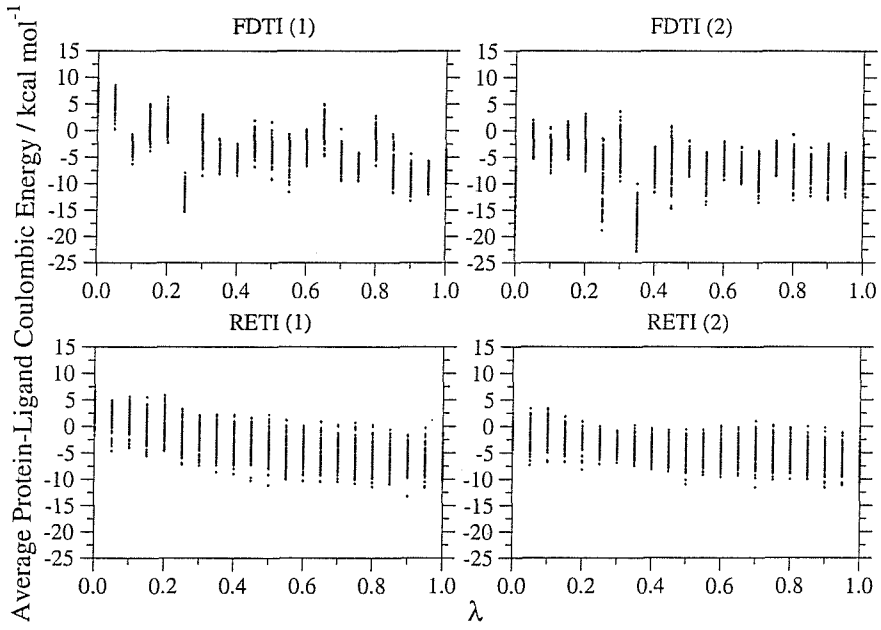


Figure 6.8: The average protein-ligand coulombic energy with respect to λ for both applications of FDTI and RETI to the bound leg, D → A perturbation. The averages are calculated every 50 K steps.

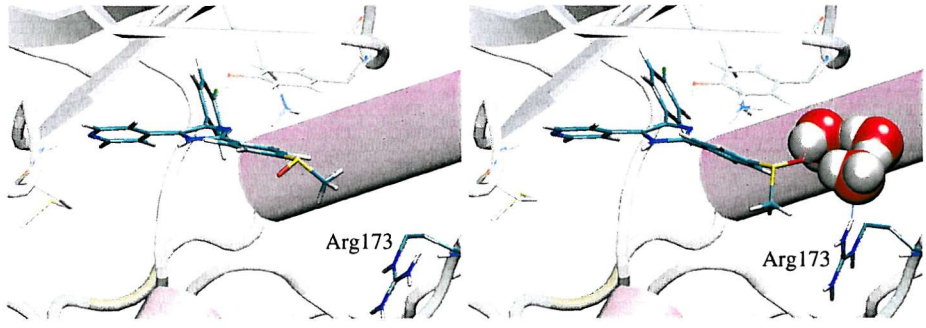
as λ is increased. RETI returns two, near-identical trends that clearly show the decrease in coulombic energy with respect to λ, as expected due to hydrogen bond formation.

6.5.4 Comparison with Experiment

Table 6.3 compares the experimental difference in IC_{50} between pairs of ligands, with the relative binding free energies calculated via the FDTI and RETI simulations. The A → C and B → C perturbations return very different relative binding free energies, suggesting that the binding of SB1 is enantiomerically selective. These results suggest that the SB1(R) ligand binds more strongly to p38 than the reduced form of SB1 (ligand C), which in turn binds more strongly than SB1(S). The combination of the two perturbations suggests that the relative binding free energy of SB1(R) and SB1(S) is about 1.3 to 5.6 kcal mol^{-1}. While this result is initially exciting, there is no experimental evidence to support enantiomeric binding, as the chirality of the binding ligand has not been investigated. Examination of the simulation structures shows little difference in protein or ligand configuration. (figure 6.9) The only significant difference between the two enantiomers is

Perturbation	FDTI / kcal mol^{-1}	RETI / kcal mol^{-1}	ΔIC_{50} / μM	Agree with experiment?
A → C	-0.8 (0.8) -1.3 (0.7)	-3.4 (0.9) -1.1 (0.9)	0.298	No
B → C	2.0 (0.8) 0.5 (0.8)	1.4 (0.9) 2.2 (0.9)	0.298	Yes
D → A	-10.5 (0.4) -14.3 (0.5)	-12.6 (0.5) -12.2 (0.5)	-0.328	Yes, but too big
E → A	3.0 (0.3) 3.5 (0.3)	2.7 (0.3) 2.8 (0.3)	-0.328	No

Table 6.3: Comparison of the experimental difference in IC_{50} between pairs of ligands to p38 with relative binding free energies calculated via two applications of FDTI and RETI.



(a) SB1(S) (b) SB1(R)

Figure 6.9: The structures of the SB1(S) and SB1(R) ligands, in the active site of p38. The structures are taken from the end of corresponding RETI simulations. Note the water-bridged hydrogen bond between SB1(R) and Arg173, which pulls Arg173 up towards the ligand.

that the SB1(S) ligand has the sulfoxide oxygen atom sticking out into the solvent of the active site cleft, while SB1(R) has the oxygen forming a water-bridged hydrogen bond with Arg73 (figure 6.9). This may help SB1(R) gain better electrostatic interactions between itself and the protein, compared to SB1(S), whose oxygen seems only to be interacting with the bulk solvent. The steric barrier to rotation of the sulfoxide group prevented the sampling of configurations where SB1(S) has rotated its oxygen around to interact with the protein. Since this did not occur, the simulations cannot say whether this particular configuration could bind as well as SB1(R), and thus say whether or not the binding is enantiomerically selective. To investigate whether this were the case, the A → C and B → C bound leg, RETI, perturbations were repeated, with the sulfoxide group rotated

Perturbation	Original / kcal mol^{-1}	Rotated / kcal mol^{-1}
A $\rightarrow$ C	23.7 (0.6)	29.0 (0.6)
B $\rightarrow$ C	28.5 (0.6)	26.1 (0.6)

Table 6.4: The relative free energy of two bound leg perturbations, as calculated via RETI. The free energies for the ligands in the original configuration, and for a configuration where the sulfoxide group has been rotated by 180 ° are shown. The results were calculated over the last 5 M steps of MC per λ-window, and the standard errors, shown in parenthesis, were calculated over blocks of 500 K steps.

by 180 °. In these simulations, SB1(S) pointed its oxygen group towards the protein, while SB1(R) pointed it out into bulk solvent. The results are shown in table 6.4, together with the results from the original RETI simulations for comparison. These results show that it is the orientation of the sulfoxide oxygen atom which imparts increased binding affinity to the SB1 ligand. This is because SB1(S), with the rotated sulfoxide group, has an approximately equal relative binding free energy to the sulfide (29.0 kcal mol^{-1}) as SB1(R) in the original crystal configuration (28.5 kcal mol^{-1}). These results suggest that the binding of SB1 is not enantiomerically selective, but that both enantiomers adopt slightly different binding modes. This is not in agreement with the crystal structure, which only shows the SB1(S) ligand bound with the oxygen pointing out into bulk solvent. The paper accompanying the crystal structure gives no information as to how this chirality of the SB1 ligand was determined, and describes how the crystal structure was formed via the soaking in of a racemic mixture of SB1. Without more information, these results suggest that the original assignment of the chirality of the crystal structure may be need to be reevaluated.

The second set of perturbations, involving SB1 and a methylated form, gave disappointing results compared to experiment. The D $\rightarrow$ A perturbation returned a relative binding free energy which agreed with the experimental order, but very much overestimated the amount. We know that this is the case, as the experimental relative binding free energies can be related to the IC_{50} via the Cheng-Prusoff equation.[141] While this requires knowledge of the Michaelis constant of the ligand for the protein, which is not available, it does lead to the general "rule of thumb" that an order of magnitude change in IC_{50} is equivalent to a relative binding free energy of 1 to 2 kcal mol^{-1}. Using this rule, the relative binding free energies of

these ligands should all fall into the range of ± 2 kcal mol^{-1}. An explanation for this overestimate is found in the structures produced from the simulations. These show that the methyl group, in the configuration of ligand D, has a steric clash with the backbone of the protein. To relieve this clash, the important hydrogen bond between N18 and Met109 is broken. This is clearly seen in the distance plots in figure 6.7. In comparison, the methyl group in the ligand E configuration does not clash with the backbone of the protein. The important hydrogen bond is preserved, and the binding mode is not affected. This leads to the conclusion that the D configuration is not the correct binding mode for this ligand, and thus the relative binding free energy to SB1 should be large and positive. Unfortunately, the simulations predict the the ligand in the E configuration is more strongly bound than SB1. This is not supported by the experimental evidence. The reasons for this poor prediction could come from many sources. For example, the simulations used a fixed protein backbone. It is known that the phosphate binding ribbon at the top of the active site is very flexible.[127] Its motion allows the Tyr35 residue to move up and down to interact better with the ligand. Neglect of the motion of this loop could have affected the results. Methods to sample the backbone motion of a protein during an MC simulation, are a subject for future work. Such methods could be based on configurational bias Monte Carlo,[142] or methods that allow the concerted rotation of multiple torsional angles within a single move, e.g. ConRot.[142]

It is likely that the lack of configurational sampling of the ligand dihedrals affected the results. Ligands D and E represent two different binding modes of the same ligand, differing only in dihedral angle. Ideally, both binding modes should be sampled within a single simulation, allowing both configurations to contribute to the same relative binding free energy. Methods to enhance configurational sampling of ligands are a subject of future work.

In summary, the p38-SB1 system presented many challenges to the free energy simulations. The chirality of the binding ligand was only inferred from the crystal structure, the protein had a flexible backbone, and the exact detail of the binding geometry was unknown.

Parameter	Perturbation	Value
Initial size of solvent box	A → C	56.5 Å × 50.1 Å × 43.6 Å
	D → A	57.2 Å × 49.5 Å × 45.4 Å
	E → A	57.3 Å × 51.5 Å × 44.0 Å
Number of TIP4P waters	A → C	3881
	D → A	4059
	E → A	4093

Table 6.5: Initial box sizes and composition for the three free leg perturbations under NPT, periodic boundaries conditions.

Perturbation	Cap / kcal mol^{-1}	NPT / kcal mol^{-1}
A → C	27.1 (0.6)	33.2 (1.1)
D → A	5.6 (0.3)	5.2 (0.4)
E → A	5.5 (0.3)	5.8 (0.3)

Table 6.6: Comparison between the relative free energies for three free leg perturbations, using a spherical solvent cap, and using an orthorhombic box of water under NPT conditions. Standard errors are shown in parenthesis, and the results are averaged over the last 5 M steps of simulation. Only the forwards free energies are shown.

6.5.5 NPT simulations

The results from applications of FDTI and RETI to the five ligands to p38 show poor agreement with experiment. Other workers have suggested that the use of a spherical solvent cap may influence the free energy results.[42,143] To investigate whether this is the case, the free leg perturbations will be performed in an orthorhombic box of water, using standard periodic boundaries. Each free leg perturbation was set up in a orthorhombic box according to table 6.5, and the simulations were equilibrated for 10 M steps under NPT conditions. RETI was applied in an identical manner to section 6.5, using 21 windows across λ, and running for 10 M steps within each window. λ-swap moves were attempted every 50 K steps, and the results were generated over the last 5 M steps within each window. As the newly developed simulation code was then not capable of sampling from the NPT ensemble, these simulations were performed using MCPRO 1.5.[13] The resulting free energies for these free leg simulations are shown in table 6.6, together with the results from the first application of RETI from section 6.5. These results show

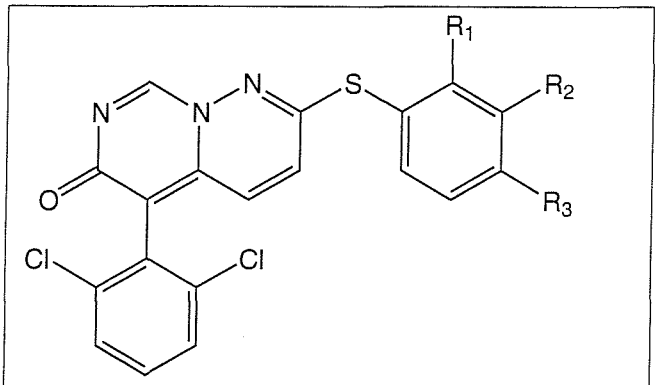
very little difference between the cap and NPT simulation results for the methyl-based perturbations. This suggests that the use of a solvent cap is not the reason why the results for the E → A perturbation returns a result which disagrees with experiment. There is a difference between the free leg A → C perturbations, with the NPT simulation predicting a significantly larger value. When combined with the corresponding bound leg perturbations from section 6.5, this has the effect of changing the binding order, and reporting that the sulfide binds more strongly than SB1. This is a disappointing result, which shows that future work must investigate the effect of solvent boundary conditions on both the bound and free legs of these calculations.

Part of the problem associated with the sulfoxide → sulfide perturbations is that the relative binding free energy is formed as the difference of two large numbers. This is because the relative free energies of the bound and free legs show a large change in the PMF of the order of 30 kcal mol^{-1}. Examination of the energy components across λ (not shown) reveals that this large free energy change is a result of large changes in ligand intramolecular, and ligand-solvent and ligand-protein intermolecular electrostatic energy. These large changes can be traced to the change in partial charge associated with the sulfur atom, as it perturbs from the sulfoxide to the sulfide. This charge changes by 0.85 |e|, as it goes from 0.819 |e| to -0.031 |e|. This large transformation of charge causes the wide variations in electrostatic energy which swamp the bound and free leg free energies. It is likely that the magnitude of the sulfur charges are unrealistic, as the diffuse d-orbitals of the sulfur atom result in an incorrect weighting of electron density from the initial Mulliken analysis. This suggests that future work should investigate the effect of different charge parameterisation schemes on the free energy results, e.g. investigating AM1-CM2 charges,[144] or AM1-BCC.[145]

6.6 Comparison to a Similar Study on p38

While the results presented in this chapter successfully demonstrate the superiority of RETI over FDTI, and reveal that free energy simulations are now sufficiently fast for routine use, the results show poor agreement with experiment. This

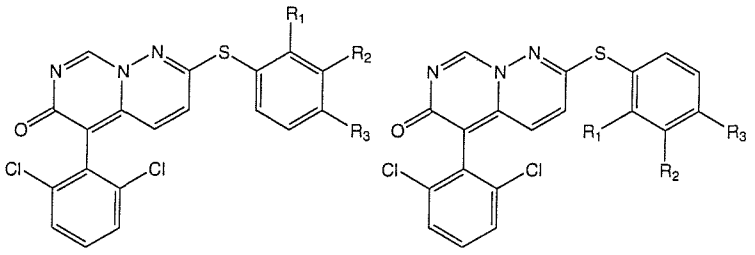
is likely to be a problem with the setup of the system, and should not be interpreted to mean that free energy calculations do not work. Recently, a successful application of thermodynamic integration to a range of ligands bound to p38 has been reported.[35] This study tested TI against less rigorous free energy methods and scoring functions, and concluded that TI was substantially more predictive than the other methods. The study investigated 16 ligands to p38. These ligands formed a congeneric series, and their experimental IC_{50}s showed a wide range from 36 nM to 1.9 μM. The ligands were all based on the same scaffold (figure 6.10). The relative binding free energy for each of these ligands was calculated from a common reference state. This reference state had R_1=F, R_2=H and R_3=F. Since it was possible to map the three R groups onto the phenyl ring in two different ways (representing a 180 ° rotation about the sulfur-phenyl bond, figure 6.11), free energy simulations for both mappings were performed. The parameters for the protein were taken from the Charmm22 forcefield,[9] while the majority of the ligand parameters were taken from AMBER.[7] The ligand partial charges were derived from an electrostatic surface potential (ESP) fit of 3-21g*/6-31g* MK charges,[35] calculated using Gaussian 98. The initial structure for each perturbation was obtained through a flexible dock of each of the ligands into a crystal structure of p38. A spherical cap of TIP3P water, of radius 16 Å, was centred on the ligand, and restrained to the 16 Å barrier by a single-sided harmonic restraint using with a 0.5 kcal mol^{-1} force constant. Most of the protein was kept rigid, with only the protein atoms within 12 Å of the ligand allowed to move during the MD simulations. Positional restraints with a force constant of 0.5 kcal mol^{-1} were applied to the moving protein atoms to keep the protein structure near the initial configuration. These structures were then minimised for 1000 steps, then equilibrated for 500 picoseconds (ps) of MD. Each TI perturbation was performed four times. Each mapping of the ligand (figure 6.11) was run in both the forwards and backwards direction. The final relative binding free energies reported in the paper were the average of these four simulations. Each perturbation was performed over 21 evenly spaced windows over the λ-coordinate, with 10 ps of equilibration and 10 ps of data collection at each λ-value. The simulations at each λ-value were per-



Ligand	R_1	R_2	R_3	pIC_{50}
1	H	H	H	6.602
2	H	H	F	7.000
3	H	H	CH_3	5.854
4	H	Cl	Cl	6.097
5	H	CH_3	H	5.854
6	H	CH_3	CH_3	5.721
7	H	F	H	6.347
8	CH_3	H	H	6.699
9	H	Cl	F	6.301
10	H	Cl	H	6.553
11	CH_3	H	Cl	6.745
12	Br	H	H	6.602
13	CH_3	H	CH_3	6.577
14	OH	H	H	6.444
15	NH_2	H	F	6.658
16	Cl	H	F	7.444

Figure 6.10: The ligands investigated during a literature application of TI on p38.[35]

formed in series, and started from the final structure of the previous λ-value. An extra 100 ps of equilibration was performed at the end of the forwards mutation, to allow the structure to further relax before starting the backwards mutation. The individual free energy gradients were numerically integrated using the trapezium rule. The MD simulations were run with a 2 fs timestep, using SHAKE to constrain bonds.[146] The simulations were performed at 300 K using a Berendsen thermostat, and an 8 Å non-bonded cutoff. The free legs of the binding calculation were per-



(a) Mapping 1 (b) Mapping 2

Figure 6.11: The two primary configurations of the ligand studied.[35]

formed using an NPT periodic box of TIP3P water, of approximately 28 Å on a side. This box was minimised for 1000 steps, then equilibrated for 500 nanoseconds of MD. The perturbations were applied over the same 21 windows across λ, in an equivalent manner to the bound leg, using 20 ps of equilibration and 20 ps of data collection at each λ-value. A further 500 ps of equilibration was performed between the forwards and backwards perturbations.

The results of these TI simulations gave very good agreement with the experimental IC_{50}s. In many ways these simulations are very similar to the simulations presented in this chapter. The main differences are in the sampling method used (MD versus MC), and the method used to obtain the charges (ESP fit of ab initio charges versus corrected Mulliken charges from the AM1 semi-empirical Hamiltonian). This gives weight to the belief that the lack of protein backbone sampling, and the poor charge derivation, may be part of the reason for the disagreement between the results presented in this chapter, and the experimental IC_{50}s. However, the above simulations applied a positional restraint to the protein atoms, which kept them close to their original positions.[35] When the same simulations were performed without this restraint,[35] it was seen that the agreement between the simulation results and the experimental IC_{50}s was markedly reduced. The authors of this study conclude that this was a result of the much greater flexibility of the system, and thus a lack of full configurational sampling.[35] This result highlights the need for configurational sampling methods which can better sample the frustrated energy surfaces of proteins. This result also shows how sensitive the results of free energy simulations can be with regards to the setup of the system, and that perhaps

the p38-ligand system was too challenging a test-case for the development of the free energy methods presented in this chapter.

6.7 Conclusion

FDTI and RETI were applied to calculate the relative binding free energies of ligands to p38 MAP kinase. RETI was again seen to be a superior method to FDTI, as it allowed wide and consistent sampling across the entire λ-coordinate. Given the simplicity of applying RETI, and its clear superiority across three diverse test systems, we would recommend that it is used whenever FEP or FDTI would normally be applied. We have succeeded in our aim of making the free energy calculation reliable, as RETI returns consistent results for each application. We have also succeeded in our aim of making the free energy calculations quick, as each bound leg took only 1.5 days to perform, and many could be run simultaneously. Using the most modern hardware (Pentium 4, 3 GHz) and our newly developed software, we believe that the simulations would take about half a day. This is in comparison to over 150 days on hardware and software available at the start of this project.

The comparison to experimental results was disappointing. While the simulations hinted at the possible enantiomer specific binding of SB1, in general, the simulations failed to agree with experiment. These simulations show how far we have come, but also how much work still needs to be done. The development of RETI allows for the consistent and correct weighting of configurations across the λ-coordinate. However, methods must still be developed that can actually sample significantly different configurations within a single trajectory. These methods should allow the sampling of multiple related binding modes, or multiple related conformers of a ligand. Such methods may help to improve the results of the methylated versions of SB1.

The quality of the results presented in this chapter may have been impaired by the neglect of backbone motion of the protein. This motion is known to be important in the flexible, phosphate binding loop of p38, and methods need to be developed which can include it within the context of a Monte Carlo simulation, e.g. configurational bias Monte Carlo[142] or ConRot.[142] The charge parameterisation

scheme should also be investigated, as problems assigning the charge for sulfur atoms could have affected the results. In addition, the use of a spherical solvent cap was seen to affect the results of the free leg of the sulfoxide $\rightarrow$ sulfide perturbation. Future work will need to investigate the effect of solvent boundaries on both the bound and free legs.

Chapter 7

Concluding Remarks

This research set out with the aim of making free energy calculations on protein-ligand complexes sufficiently reliable, and fast, such that they could be readily used within pharmaceutical drug design. To satisfy this aim, in chapter 3, established free energy methods were tested on a simple, yet challenging test case. This test case required the accurate calculation of the relative hydration free energy of water and methane. This free energy was seen to depend strongly on the exact solvent configurations sampled during each simulation. Two of the three established free energy methods, FEP and FDTI, sampled in an identical manner. They both used a single trajectory at each λ-value, and derived independent free energies for each of these λ-points. The difference between the two methods was that while FEP calculated the difference in free energy between neighboring λ-windows, FDTI only calculated the free energy gradient with respect to λ for each point. Unlike FDTI, the results from the FEP simulations were affected by the window width. This led to a significant hysteresis in the predicted free energies, which was not present in the FDTI results. While this hysteresis could be improved through better positioning of the λ-windows, this would require some *a priori* knowledge of the potential of mean force (PMF) along λ, something which would not be readily available without significant exploratory work.

The sampling in both the FEP and FDTI simulations showed artefacts from the use of a single trajectory at each λ-value. As each trajectory was independent,

different sub-sets of configurations were sampled across the λ-coordinate. This led to inconsistencies across λ, which were highlighted in the radial distribution functions between the perturbing solute and surrounding solvent. These RDFs did not change smoothly with λ, but instead showed significant variation. This was also apparent in the predicted free energies at each λ-value, which varied depending on the exact subset of configurations sampled. This led to a random sampling error in the free energies predicted by FEP and FDTI of about 1 kcal mol^{-1}.

The third established free energy method, AdUmWHAM, was also tested on the water-methane system. Unlike FEP or FDTI, the λ-value in this method was dynamic. This effectively allowed multiple trajectories to sample each λ-value. While this smoothed the RDFs, it also led to the possibility that the motion along λ was being driven through poor solvent configurations. This was again revealed by artefacts in the RDF, and was the result of the solvent configurational change happening too slowly compared to the rate of change of λ.

To solve these problems, in chapter 4, the water-methane system was investigated via four new free energy methods. These methods used ideas from generalised ensembles in an attempt to enhance configurational sampling. The first new method, BMW, combined the multicanonical ensemble with AdUmWHAM, in an attempt to speed up the solvent configurational change. Unfortunately, the method was overly ambitious, and failed to perform as expected. The next new method, PTTI, combined the technique of parallel tempering with FDTI. This used multiple temperature trajectories at each λ-value in an attempt to enhance configurational sampling and to average out the random sampling error. While the method was successful at averaging out the random sampling error, the use of independent simulations across λ meant that the sampling was still inconsistent along λ. In addition, PTTI proved to be too expensive to recommend its use unless there was a real need to use multiple temperatures to overcome large energy barriers. The last two methods combined FEP or FDTI with the technique of replica exchange. These methods, REFEP and RETI, represented the trivial addition of a λ-swap move to standard FEP or FDTI simulations. This move allowed each of the trajectories to visit multiple λ-values. This led to consistent sampling across λ, as seen in smooth RDFs, and a very low random sampling error. While the sampling

was significantly improved by the λ-swap move, it failed to improve the hysteresis problems of FEP. Because of this, RETI was seen to be superior to REFEP, and indeed to all of the other free energy methods tested.

To examine if the superiority of RETI was system-dependent, in chapter 5, all seven methods were applied to calculate the relative binding free energies of halides to a calix[4]pyrrole derivative. The results from these applications showed similar patterns to the results from the water-methane simulations. The simulations on the calix[4]pyrrole system used a DMSO solvent which was contaminated by small amounts of water. The position of the water was randomised for each trajectory, in an attempt to investigate how well each method sampled the effect of the water positions. Of particular significance, was whether the water configuration allowed for any hydrogen bonding with the perturbing halide ion. FEP and FDTI performed poorly in this regard, as the use of a single trajectory at each λ-value led to the use of different water geometries across λ. This was inconsistent, with some λ-values exhibiting hydrogen-bonding, and others not. This led to predicted free energies which were the average of the hydrogen-bonding and non-hydrogen-bonding free energies. In contrast, AdUmWHAM and BMW used a dynamic λ-coordinate. This allowed for consistent sampling across λ, and ensured that if the trajectory contained hydrogen-bonding, then the entire λ-coordinate could experience hydrogen-bonding. However, both methods failed to allow a single trajectory to flip back and forth between the two states, so the free energies produced were either the result of hydrogen bonding trajectories or not. There was little averaging of the results of the two cases, and thus the predicted free energies were larger, or smaller, than those predicted by FEP or FDTI.

PTTI allowed both hydrogen-bonding and non-hydrogen bonding trajectories to be sampled at each λ-value. This was an improvement over FEP or FDTI. Unfortunately, even the use of high temperatures failed to allow the trajectories to flip between the two states. This meant that there was still inconsistency across λ, as some λ-values had randomly either lots of hydrogen-bonding, or very little hydrogen-bonding. This inconsistency again led to the predicted free energies returning the incorrect average of these two states.

Through the use of the λ-swap move, REFEP and RETI allowed hydrogen-

bonding and non-hydrogen-bonding trajectories to move across the entire λ-coordinate. As in the water-methane simulations, this led to improved, consistent sampling across λ. In addition, the λ-swap move allowed the two solvent states to be properly weighted across λ. Thus there was a higher probability of sampling a hydrogen-bonding trajectory for fluoride than there was for chloride or bromide. This meant that REFEP and RETI had gone some way to forming a proper average of the two solvent states. Unfortunately, the λ-swap move only correctly weighted the initial states of the solvent; it did not help the solvent to flip between the two states. This led to free energies which were dependent on the initial number of hydrogen-bonding or non-hydrogen-bonding trajectories. While RETI again produced free energies which were the superior of all of the methods, there is still room for improvement. Such improvement would come from the development of solvent sampling moves which allow the contaminating waters to properly diffuse through the solvent. Such moves are the subject of future work, and could involve the use of a softening parameter to scale down the interactions between the water and the rest of the system. This could perhaps be implemented through use of replica exchange moves over both λ and the softening parameter.

Finally, in chapter 6, the superiority of RETI over FDTI was confirmed. Both methods were applied to calculate the relative binding free energies of SB1-like ligands to p38 MAP kinase. RETI again demonstrated superior sampling compared to FDTI, and returned consistent energetic and structural trends across λ. As on all of the test systems, RETI has demonstrated consistent and reliable sampling. The method shows low sampling error, and thus we believe that the method satisfies the first aim of this research. As part of the work of chapter 6, a new Monte Carlo program was written to perform the free energy simulations. Through use of an optimised data structure, the code allowed these simulations to be performed ten to twelve times more rapidly than via the established Monte Carlo code used for the rest of this research. Combining this improvement with the increasing power of Linux Beowulf clusters, meant that these simulations only took 1.5 days each to run. Using more modern hardware, we predict that they could take less than half a day to run. We believe that this is sufficiently fast for commercial drug design, thus satisfying the second aim of this research.

While we believe that the combination of RETI, our new MC code, and Linux Beowulf clusters satisfies the aims of this research, there are still problems with free energy calculations. These were highlighted by the application to the p38 system, and are related to the setup, and range of sampling within each simulation. These problems led to the disagreement between the predicted free energies and the experimental IC_{50}s. Their solution is the subject of current, and future work, and fall into two distinct groups;

1. Simulation Setup

   (a) There was a problem with the partial charges on the ligand sulfur atoms. This is being investigated through the comparison of partial charges predicted by different parameterisation methods.

   (b) The use of only a ball of solvent, rather than a periodic box, may have affected the results. Future work should investigate this more fully, and investigate the practicality of using a periodic box for the bound leg of the simulation.

   (c) The use of 15 Å, residue-based cutoffs may have affected the results. This may need to be investigated, and longer cutoffs used.

   (d) The binding modes of the ligands were unclear. Higher resolution crystal structures are required to provide data as to the chirality of the binding ligands, and the conformations of important ligand torsional angles.

2. Sampling

   (a) There was a lack of sampling of related ligand binding modes, specifically binding modes which differed only via rotation of a dihedral angle of the ligand. Methods need to be developed which allow multiple related binding modes to be sampled within a single simulation trajectory.

   (b) There was a lack of protein backbone sampling. This may have been important for the flexible loop of p38. Methods are being developed which allow protein backbone moves within an MC trajectory.

(c) There was a lack of sampling of grossly different binding modes. While this is highly ambitious, it should be the ultimate aim of any future developments of this work.

In conclusion, the broad aims of this research have been satisfied. The inclusion of a λ-swap move into a normal FDTI simulation allows for consistent and reliable sampling. This has been demonstrated over three diverse test systems, leading us to recommend that RETI should generally be used in preference to FEP or FDTI. The development of a new simulation code, and the growth of power of Beowulf clusters, means that free energy simulations are now sufficiently fast for general use. We are therefore now at a stage where we can really push the protein-ligand sampling. The speed of the simulations, and the power of RETI, mean that we can now develop methods that implement smart protein or ligand moves. Computers are now so fast that we can use huge, fully solvated protein-ligand systems in NPT periodic solvent boxes. We can also think about incorporating the effect of multiple binding modes into the predicted free energies, or even begin to consider the possibility of calculating absolute binding free energies.

Appendix A

The Boltzmann Equation

Let us imagine a collection of N independent molecules, which can exist in discrete energy states ε_0, ε_1... The lowest energy shall be ε_0. At any particular instant, there may be n_0 molecules in energy state ε_0, n_1 molecules at energy ε_1 etc. This set of populations, n_0, n_1... is the instantaneous configuration of the system.

This configuration will fluctuate with time as the populations change. However, certain configurations will be more probable than others. For example, lets imagine that there are two energy levels, and four molecules (figure A.1). Since the molecules are indistinguishable, there will be a higher probability of the molecules being evenly distributed between both energy levels than all molecules existing in a single energy level. This is because there is only one way to put all molecules in the first energy level, compared to six ways that they may be evenly distributed. For four molecules in two energy states, the most probable configuration appears six times. As the number of molecules or energy levels increases, the number of times the most probable configuration appears also increases. For 10 molecules in 2 energy levels, it will appear 252 times, and for 100 molecules in 2 energy levels, it will appear 184756 times! There will exist one overwhelmingly probable distribution for large numbers of molecules spread over large numbers of energy levels. Since the molecules are most likely to be in this distribution, we can assume that the average properties of the molecules are based solely on this *dominating configuration*. The aim of the Boltzmann equation is to calculate the probabilities within this dominating configuration. To do this, it needs to be able to calculate

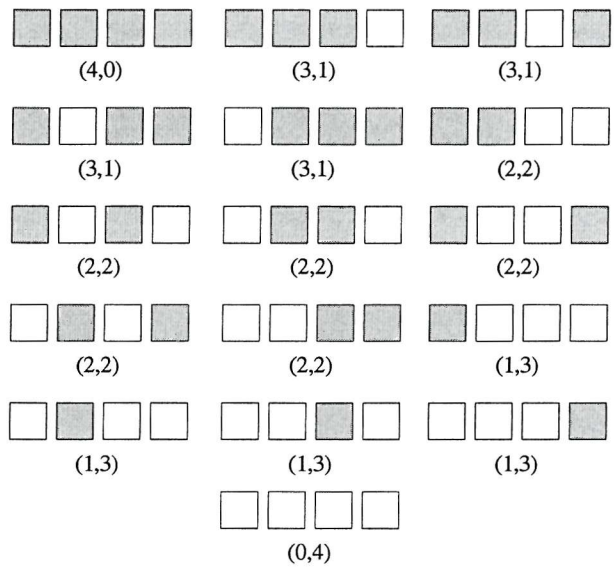


Figure A.1: The sixteen different ways of distributing four molecules between two energy levels. The two energy levels are represented by grey or white boxes. Since the molecules are indistinguishable, there are six ways of obtaining an even distribution of molecules, compared to only one way of obtaining all molecules in the same energy level.

how many ways a particular configuration can be made, subject to the constraint that the configuration is physically reasonable.

In general, the configuration $n_0, n_1 \ldots$ can be achieved in W different ways. W is called the *weight* of the configuration, and is given by the expression,

$$W = \frac{N!}{n_0! n_1! n_2! \ldots}. \tag{A.1}$$

It is more convenient to work with the natural logarithm of W. This enables us to rearrange the equation into the form,

$$\begin{aligned} \ln(W) &= \ln\left(\frac{N!}{n_0! n_1! n_2! \ldots}\right) \\ &= \ln N! - \sum_i \ln n_i!. \end{aligned} \tag{A.2}$$

Stirling's approximation states that, for large x,

$$\ln x! \approx x \ln x - x. \tag{A.3}$$

This allows us to approximate the weight, W, as,

$$\begin{aligned} \ln(W) &= (N\ln N - N) - \sum_i (n_i \ln n_i - n_i) \\ &= N\ln N - \sum_i n_i \ln n_i. \end{aligned} \tag{A.4}$$

The current system can have any number of particles occupying any of the energy levels. This would not be the case for an isolated, or closed system. Such a system will be subject to two constraints; first that the total energy, E, of the system is a constant,

$$\sum_i n_i \varepsilon_i = E, \tag{A.5}$$

and second, that the number of particles, N, in the system is also constant,

$$\sum_i n_i = N. \tag{A.6}$$

To find the dominating configuration, we must maximise $\ln(W)$ while applying these two constraints. This is achieved through examining the rate of change of $\ln(W)$ with respect to changing each of the individual n_i,

$$d\ln(W) = \sum_i \left(\frac{\partial \ln W}{\partial n_i} \right) dn_i. \tag{A.7}$$

At the maximum, $d\ln(W) = 0$. To solve this equation, and take the constraints into account, we must use the technique of *Lagrange Multipliers*. The constraints are multiplied by two constants, α and $-\beta$, and added to equation A.7,

$$\begin{aligned} d\ln W &= \sum_i \left(\frac{\partial \ln W}{\partial n_i} \right) dn_i + \alpha \sum_i dn_i - \beta \sum_i \varepsilon_i dn_i \\ &= \sum_i \left\{ \left(\frac{\partial \ln W}{\partial n_i} \right) + \alpha - \beta\varepsilon_i \right\} dn_i. \end{aligned} \tag{A.8}$$

For $d\ln W = 0$, each of the individual parts of this sum must be equal to zero. Thus,

$$\left(\frac{\partial \ln W}{\partial n_i} \right) + \alpha - \beta\varepsilon_i = 0. \tag{A.9}$$

We can solve this equation by differentiating equation A.4 with respect to n_i,

$$\frac{\partial \ln W}{\partial n_i} = \frac{\partial (N \ln N)}{\partial n_i} - \sum_j \left(\frac{\partial (n_j \ln n_j)}{\partial n_i} \right). \tag{A.10}$$

We shall consider the two parts of this differentiation separately. For the first part, we recognise that N is a function of n_i,

$$N = \ldots + n_{i-1} + n_i + n_{i+1} + \ldots. \tag{A.11}$$

We can thus use the chain rule to differentiate the first term,

$$\begin{aligned} \frac{\partial (N \ln N)}{\partial n_i} &= \left(\frac{\partial (N \ln N)}{\partial N} \right) \times \left(\frac{\partial (\ldots + n_{i-1} + n_i + n_{i+1} + \ldots)}{\partial n_i} \right) \\ &= \left(1 \times \ln N + N \times \frac{1}{N} \right) \times 1 \\ &= \ln N + 1. \end{aligned} \tag{A.12}$$

The second part is slightly more complicated,

$$\sum_j \left(\frac{\partial (n_j \ln n_j)}{\partial n_i} \right) = \sum_j \left\{ \left(\frac{\partial n_j}{\partial n_i} \right) \ln n_j + n_j \left(\frac{\partial \ln n_j}{\partial n_i} \right) \right\}. \tag{A.13}$$

If $i \neq j$, then n_j is independent of n_i and,

$$\frac{\partial n_j}{\partial n_i} = 0, \tag{A.14}$$

while if $i = j$, then

$$\frac{\partial n_j}{\partial n_i} = \frac{\partial n_i}{\partial n_j} = 1. \tag{A.15}$$

Using these two statements, together with the equality (which comes from the chain rule),

$$\frac{\partial \ln n_j}{\partial n_i} = \frac{1}{n_j} \frac{\partial n_j}{\partial n_i}, \tag{A.16}$$

allows us to express equation A.13 as,

$$\sum_j \left\{ \left(\frac{\partial n_j}{\partial n_i} \right) \ln n_j + n_j \left(\frac{\partial \ln n_j}{\partial n_i} \right) \right\} = \sum_j \left\{ \left(\frac{\partial n_j}{\partial n_i} \right) \ln n_j + n_j \left(\frac{1}{n_j} \frac{\partial n_j}{\partial n_i} \right) \right\}$$

$$= \sum_j \left(\frac{\partial n_j}{\partial n_i} \right) (\ln n_j + 1) \qquad \text{(A.17)}$$

$$= \ln n_i + 1, \quad \text{for } i = j, \text{ else all } 0.$$

Combining equations A.17, A.12, A.10 and A.9, gives,

$$\begin{aligned} 0 &= \left(\frac{\partial \ln W}{\partial n_i} \right) + \alpha - \beta \varepsilon_i \\ &= \frac{\partial (N \ln N)}{\partial n_i} - \sum_j \left(\frac{\partial (n_j \ln n_j)}{\partial n_i} \right) + \alpha - \beta \varepsilon_i \\ &= (\ln N + 1) - (\ln n_i + 1) + \alpha - \beta \varepsilon_i \\ &= \ln \left(\frac{N}{n_i} \right) + \alpha - \beta \varepsilon_i \end{aligned} \qquad \text{(A.18)}$$

$$\ln \left(\frac{n_i}{N} \right) = \alpha - \beta \varepsilon_i$$

$$\frac{n_i}{N} = \exp^{\alpha} \exp^{-\beta \varepsilon_i}$$

To obtain the value of α, we note that,

$$N = \sum_j n_j = N \exp^{\alpha} \sum_j \exp^{-\beta \varepsilon_j} \qquad \text{(A.19)}$$

$$\exp^{\alpha} = \frac{1}{\sum_j \exp^{-\beta \varepsilon_j}}.$$

The value of β may be shown[19] to be equal to $1/k_B T$, where k_B is Boltzmann's constant. The combination of all of this is the Boltzmann equation,

$$\frac{n_i}{N} = \frac{exp^{-\varepsilon_i / k_B T}}{\sum_j \exp^{-\varepsilon_j / k_B T}}. \qquad \text{(A.20)}$$

This equation holds for a set of independent, indistinguishable molecules. A similar derivation can be performed using an ensemble of systems. In this derivation, the indistinguishable molecules are replaced by multiple, indistinguishable,

replicas of the system (the ensemble). This collection of systems can adopt many different energy states, and constraints are applied based on the ensemble conditions.

Appendix B

Monte Carlo Sampling

Metropolis Monte Carlo generates random configurations, such that by the end of the simulation, the configurations have all been generated by their correct Boltzmann weight. It achieves this as it builds a *Markov chain* of configurations. A Markov chain is a sequence of trials, set up such that the outcome of each trial belongs to a finite set of outcomes, and that the outcome of each trial only depends on the outcome of the trial that immediately preceded it.[11]

$\rho^{(1)}$ represents the initial guessed probability distribution. π represents the transition matrix. This matrix operates on $\rho^{(1)}$ to produce a new estimate of the probability distribution, $\rho^{(2)}$,

$$\rho^{(2)} = \rho^{(1)}\pi. \tag{B.1}$$

As long as the conditions of the Markov chain are adhered to, then π has the property of making the probability distribution converge onto a limiting value, ρ,

$$\begin{aligned} \rho^{(3)} &= \rho^{(2)}\pi = \rho^{(1)}\pi\pi = \rho^{(1)}\pi^2 \\ \rho &= \lim_{t\to\infty} \rho^{(1)}\pi^t. \end{aligned} \tag{B.2}$$

It is clear from this equation that the limiting probability distribution must satisfy the eigenvalue equation,

$$\begin{aligned} \rho\pi &= \rho \\ \sum_m \rho_m \pi_{mn} &= \rho_n. \end{aligned} \tag{B.3}$$

It is a property of the transition matrix that its rows sum to one,

$$\sum_n \pi_{mn} = 1, \tag{B.4}$$

and that the limiting distribution, ρ, is independent of the initial starting distribution.

The aim of Metropolis Monte Carlo is to generate a transition matrix such that the limiting distribution is equal to the Boltzmann distribution, thus $\rho(i) = p_{Boltzmann}(i)$, for each point, i, in phase space. It is possible to generate the elements of the transition matrix, π_{mn}, and in so doing, generate the correct ensemble. To achieve this, we must solve the eigenvalue equation. Metropolis Monte Carlo achieves this by imposing the unnecessarily strict condition of *microscopic reversibility*,

$$\rho_m \pi_{mn} = \rho_n \pi_{nm}. \tag{B.5}$$

This means that the trajectory must be reversible, namely that it should be equally likely to move from state m to state n as it is to move from state n to state m. We can show that this condition satisfies the eigenvalue equation via,

$$\sum_m \rho_m \pi_{mn} = \sum_m \rho_n \pi_{nm} = \rho_n \sum_m \pi_{nm} = \rho_n. \tag{B.6}$$

The Metropolis solution solves this equation through the use of a symmetrical stochastic matrix,[19] α. This matrix shows that random moves should be attempted between states m and n, and that the probability of attempting the move from m to n should be equal to the probability of attempting to move from state n to m, i.e. $\alpha_{mn} = \alpha_{nm}$. The Metropolis solution recasts π_{mn} in terms of α, for three different cases,

$$\begin{aligned} \pi_{mn} &= \alpha_{mn} \text{ for all } m \neq n \text{ and } \rho_n \geq \rho_m, \\ \pi_{mn} &= \alpha_{mn}(\rho_n/\rho_m) \text{ for all } m \neq n \text{ and } \rho_n < \rho_m, \\ \pi_{mm} &= 1 - \sum_{n \neq m} \pi_{mn}. \end{aligned} \tag{B.7}$$

These can be shown to agree with equation B.5,

$$\begin{aligned} \rho_m \pi_{mn} &= \rho_n \pi_{nm}, \\ \text{using } \rho_n &\geq \rho_m, \\ \rho_m \alpha_{mn} &= \rho_n \alpha_{nm} (\rho_m / \rho_n), \\ \rho_m \alpha_{mn} &= \alpha_{nm} \rho_m, \\ \text{since } \alpha_{mn} &= \alpha_{nm} \\ \rho_m &= \rho_m. \end{aligned} \tag{B.8}$$

These equations show that the Monte Carlo simulation should be performed by attempting random moves from state m to state n with an equal probability to attempting moves from state n to state m. If $\rho_n \geq \rho_m$, then the move should be automatically accepted. If $\rho_n < \rho_m$, then the move should be accepted according to a probability of ρ_n/ρ_m. To do this, this ratio is compared to a random number between 0 and 1. If the ratio is greater than the random number, then the move is accepted. If the ratio is less than or equal to the random number, then the move is rejected, and configuration m is recounted in the average. This ratio will have different forms depending on the desired ensemble for the simulation. For example, in the NVT, canonical ensemble,

$$\begin{aligned} \rho_n/\rho_m &= \frac{\exp(-E_n/k_BT)/Q_{NVT}}{\exp(-E_m/k_BT)/Q_{NVT}} \\ &= \frac{\exp(-E_n/k_BT)}{\exp(-E_m/k_BT)} \\ &= \exp(-(E_n - E_m)/k_BT) \\ &= \exp(-\Delta E/k_BT), \end{aligned} \tag{B.9}$$

while in the NPT ensemble,

$$\begin{aligned} \rho_n/\rho_m &= \frac{\exp\left(-\frac{E_n+PV_n}{k_BT} + N\ln V_n\right)/Q_{NPT}}{\exp\left(-\frac{E_m+PV_m}{k_BT} + N\ln V_m\right)/Q_{NPT}} \\ &= \exp\left(-\frac{\Delta E + P\Delta V}{k_BT} + N\Delta \ln V\right). \end{aligned} \tag{B.10}$$

The Metropolis solution is only one of many solutions to the Monte Carlo equations. In preferential sampling,[11,91] an asymmetric underlying stochastic matrix is used. This allows some moves to be attempted with a higher probability than others, and that the probability of moving in one direction does not have to be equal to the probability of moving back again. For this to satisfy the condition of microscopic reversibility, the values of π_{mn} are given by,

$$\begin{aligned}
&\pi_{mn} = \alpha_{mn} \text{ for all } m \neq n \text{ and } \alpha_{nm}\rho_n \geq \alpha_{mn}\rho_m, \\
&\pi_{mn} = \alpha_{mn}\left(\frac{\alpha_{nm}\rho_n}{\alpha_{mn}\rho_m}\right) \text{ for all } m \neq n \text{ and } \alpha_{nm}\rho_n < \alpha_{mn}\rho_m, \\
&\pi_{mm} = 1 - \sum_{n \neq m} \pi_{mn}.
\end{aligned} \tag{B.11}$$

The agreement with microscopic reversibility is shown by,

$$\begin{aligned}
&\pi_{mn}\rho_m = \pi_{nm}\rho_n \\
&\alpha_{mn}\rho_m = \alpha_{nm}\left(\frac{\alpha_{mn}\rho_m}{\alpha_{nm}\rho_n}\right)\rho_n \\
&\alpha_{mn}\rho_m = \alpha_{mn}\rho_m.
\end{aligned} \tag{B.12}$$

Preferential sampling may be used to increase the probability of sampling solvent molecules that are closest to the central solute. This can increase the sampling of the primary solvation shell, which is likely to have the biggest impact on any relative free energies of the solute.[11] A solvent molecule is picked with a probability, W, which decreases with distance, r, from the central solute,

$$W(r_i) = \frac{1}{r_i^{\nu}} \text{ where } \nu \text{ is a chosen integer parameter.} \tag{B.13}$$

In MCPRO,[13] ν is equal to 1, and the function used is,[92]

$$W(r_i) = \frac{1}{r_i + w_{kc}}. \tag{B.14}$$

A constant, w_{kc} is added to the distance to help prevent the volume from continually expanding during the simulation.[92]

For each configuration, the weight of each solvent molecule is calculated, and all of the weights are normalised,

$$W'(r_i) = \frac{W(r_i)}{\sum_j W(r_j)}. \tag{B.15}$$

A solvent molecule is chosen with a probability of $W'(r_i)$. This is achieved by randomly choosing a solvent molecule, and comparing its weight against a random number between 0 and 1. If the weight is greater than the random number, then that solvent molecule is selected, otherwise a new solvent molecule is chosen and the test repeated. A Monte Carlo move is performed on the chosen solvent molecule. The weight of the solvent before the MC move, W_{old}, and weight after the move, W_{new}, are used in the modified acceptance test,

$$\frac{W_{new}}{W_{old}} \exp(-\Delta E/k_B T) \geq rand(0,1). \tag{B.16}$$

This test is applied in the canonical ensemble, and is derived in a similar manner to equation B.9,

$$\begin{aligned} \alpha_{mn}\left(\frac{\alpha_{nm}\rho_n}{\alpha_{mn}\rho_m}\right) &= \alpha_{nm}\frac{\rho_n}{\rho_m} \\ &= \frac{W_{new}}{W_{old}} \times \frac{\rho_n}{\rho_m} \\ &= \frac{W_{new}}{W_{old}} \exp(-\Delta E/k_B T). \end{aligned} \tag{B.17}$$

Appendix C

Free Energy Methods

C.1 Calculating Free Energy Gradients for TI

Thermodynamic Integration works by calculating the integral of the gradient in the free energy with respect to λ, $(\frac{\partial G(\lambda)}{\partial \lambda})_{\lambda}$. This gradient can be related to the ensemble average of the gradient of the forcefield with respect to λ, $\langle \frac{\partial E(\Gamma,\lambda)}{\partial \lambda} \rangle_{\lambda}$, where the forcefield is evaluated using the coordinates of the system, Γ, and the average is collected over the ensemble of structures generated with the forcefield

perturbed to the value of λ. This equivalence can be shown via,[12]

$$\Delta G = \int_0^1 \left(\frac{\partial G(\lambda)}{\partial \lambda}\right)_\lambda d\lambda$$

since $G(\lambda) = -k_B T \ln Q(\lambda)$

$$\Delta G = -k_B T \int_0^1 \left(\frac{\partial \ln Q(\lambda)}{\partial \lambda}\right)_\lambda d\lambda$$

$$= -k_B T \int_0^1 \frac{1}{Q(\lambda)} \left(\frac{\partial Q(\lambda)}{\partial \lambda}\right)_\lambda d\lambda$$

since $Q(\lambda) = \frac{1}{N!}\frac{1}{h^{3N}} \int \exp\left(-E(\Gamma,\lambda)/k_B T\right) d\Gamma$

then $\left(\frac{\partial Q(\lambda)}{\partial \lambda}\right)_\lambda = \frac{1}{N!}\frac{1}{h^{3N}} \int \left(\frac{\partial \exp\left(-E(\Gamma,\lambda)/k_B T\right)}{\partial \lambda}\right)_\lambda d\Gamma$

differentiating the exponential,

$$\left(\frac{\partial Q(\lambda)}{\partial \lambda}\right)_\lambda = -\frac{1}{N!}\frac{1}{h^{3N}}\frac{1}{k_B T} \int \left(\frac{\partial E(\Gamma,\lambda)}{\partial \lambda}\right)_\lambda \exp\left(-E(\Gamma,\lambda)/k_B T\right) d\Gamma$$

Substituting back into the expression for ΔG gives,

$$\Delta G = \int_0^1 \left(\frac{1}{N!}\frac{1}{h^{3N}} \int \frac{1}{Q(\lambda)} \left(\frac{\partial E(\Gamma,\lambda)}{\partial \lambda}\right)_\lambda \exp\left(-E(\Gamma,\lambda)/k_B T\right) d\Gamma\right) d\lambda$$

$$= \int_0^1 \left(\frac{1}{N!}\frac{1}{h^{3N}} \int \left(\frac{\partial E(\Gamma,\lambda)}{\partial \lambda}\right)_\lambda \frac{\exp\left(-E(\Gamma,\lambda)/k_B T\right)}{Q(\lambda)} d\Gamma\right) d\lambda$$

$$= \int_0^1 \left(\frac{1}{N!}\frac{1}{h^{3N}} \int \left(\frac{\partial E(\Gamma,\lambda)}{\partial \lambda}\right)_\lambda p_\lambda(\Gamma) d\Gamma\right) d\lambda$$

where $p_\lambda(\Gamma)$ is the ensemble probability for configuration Γ for that value of λ,

$$\text{thus } \Delta G = \int_0^1 \left\langle \frac{\partial E(\Gamma,\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda. \tag{C.1}$$

C.2 Derivation of the Multicanonical Move Test

The multicanonical ensemble is one in which all energies are sampled equally. This means that the desired probability distribution is not based on the Boltzmann

equation, but is rather,

$$p_m = 1/n(E_m), \tag{C.2}$$

where $n(E_m)$ is the density of states for the energy of configuration m. The density of states is the number of configurations that have a particular energy. Recalling the results of appendix B, a Monte Carlo simulation can be set up to sample from this distribution. If the move is from state m to state n, then the Monte Carlo test is based on the ratio of p_n and p_m,

$$\begin{aligned} \frac{\rho_n}{\rho_m} &= \frac{p_n}{p_m} \\ &= \frac{n(E_m)}{n(E_n)}. \end{aligned} \tag{C.3}$$

A Monte Carlo simulation can thus be performed. A random move is made, and the density of states of the two states determined. If this ratio is greater than or equal to a random number between 0 and 1, then the move is accepted, else the move is rejected.

The use of this Monte Carlo test is equivalent to an adaptive umbrella sampling simulation, where the umbrella potential acts over potential energy. This can be shown via,

$$\begin{aligned} \frac{\rho_n}{\rho_m} &= \frac{n(E_m)}{n(E_n)}, \text{ using } n(E) = P(E)\exp(E/k_BT), \\ &= \frac{P(E_m)\exp(E_m/k_BT)}{P(E_n)\exp(E_n/k_BT)}, \\ &= \frac{P(E_m)}{P(E_n)}\exp(-(E_n - E_m)/k_BT), \\ &= \exp\left(-\frac{\Delta E}{k_BT} + \ln\frac{P(E_m)}{P(E_n)}\right), \\ &= \exp\left(-\frac{\Delta E + k_BT\Delta\ln P(E)}{k_BT}\right). \end{aligned} \tag{C.4}$$

The Monte Carlo simulation is thus performed using an umbrella of,

$$U(E) = k_BT\ln P(E), \tag{C.5}$$

where $P(E)$ is the Boltzmann probability for energy E.

C.3 Derivation of the Replica Exchange Tests

Let us imagine three different systems, A, B and C. Each system will be in a particular configuration, e.g. system A may adopt configuration i, system B may be in configuration j, and system C make be in configuration k. The probability, $p_A(i)$, that system A is in configuration i, in the NVT ensemble, is given by the Boltzmann equation,

$$p_A(i) = \frac{\exp^{-\beta_A H_A(i)}}{Q_A}, \tag{C.6}$$

where H_A is the Hamiltonian for the system A, β_A is $1/k_B T_A$, where T_A is the temperature of system A, and Q_A is the canonical partition function for the Hamiltonian. Similar equations exist to describe the probability of configuration j in system B, $p_B(j)$, and the probability of configuration k in system C, $p_C(k)$.

Now let us consider this collection of three systems as a whole. We have thus formed a "super-collection" or "super-ensemble" over these three systems. The probability that system A is in configuration i, *and* system B is in configuration j, *and* system C is in configuration k is given by the product of their three individual probabilities. The probability of this set, $P(X)$, is thus given by,

$$\begin{aligned} P(X) &= p_A(i) p_B(j) p_C(k) \\ &= \frac{\exp^{-\beta_A H_A(i)} \exp^{-\beta_B H_B(j)} \exp^{-\beta_C H_C(k)}}{Q_A Q_B Q_C}. \end{aligned} \tag{C.7}$$

In this equation, X is a vector representing the super-configuration set (i, j, k). It is possible that the super-ensemble can exist in configuration set Y, representing (j, i, k), where systems A and B have swapped configurations. Recalling the results of appendix B, a Monte Carlo test could be performed on the move from X to Y,

based on the relative probabilities of X and Y,

$$\begin{aligned}
\frac{\rho_Y}{\rho_X} &= \frac{P(Y)}{P(X)} \\
&= \frac{\exp^{-\beta_A H_A(j)} \exp^{-\beta_B H_B(i)} \exp^{-\beta_C H_C(k)}}{Q_A Q_B Q_C} \\
&\quad \times \frac{Q_A Q_B Q_C}{\exp^{-\beta_A H_A(i)} \exp^{-\beta_B H_B(j)} \exp^{-\beta_C H_C(k)}} \\
&= \frac{\exp^{-\beta_A H_A(j)} \exp^{-\beta_B H_B(i)} \exp^{-\beta_C H_C(k)}}{\exp^{-\beta_A H_A(i)} \exp^{-\beta_B H_B(j)} \exp^{-\beta_C H_C(k)}} \\
&= \frac{\exp^{-\beta_A H_A(j)} \exp^{-\beta_B H_B(i)}}{\exp^{-\beta_A H_A(i)} \exp^{-\beta_B H_B(j)}} \\
&= \exp\left[-\beta_A H_A(j) - \beta_B H_B(i) + \beta_A H_A(i) + \beta_B H_B(j)\right] \\
&= \exp\left[\beta_B\big(H_B(j) - H_B(i)\big) - \beta_A\big(H_A(j) - H_A(i)\big)\right] \\
&= \exp(\Delta)
\end{aligned} \tag{C.8}$$

where $\Delta = \beta_B\big(H_B(j) - H_B(i)\big) - \beta_A\big(H_A(j) - H_A(i)\big)$.

Note how the partition functions for each system cancel out, as do the probabilities of system C. Indeed, this derivation demonstrates that *only* the probabilities of the systems exchanging configurations need to be included. This equation is thus valid for super-collections of *any number of systems*. Also, in this case, Δ was calculated for systems in the NVT canonical ensemble. It is possible to use a similar derivation to obtain Δ for many other ensembles. For example, the derivation of Δ for the NPT ensemble, where the volume, V, of the replicas must be taken into

account, is,

$$p_A(i) = \frac{\exp^{-\beta_A(H_A(i)+P_AV_i)+N\ln V_A}}{Q_A}$$

$$P(X) = p_A(i)p_B(j)p_C(k) \text{ and } P(Y) = p_A(j)p_B(i)p_C(k)$$

$$\frac{P(Y)}{P(X)} = \frac{\exp^{-\beta_A(H_A(j)+P_AV_j)}\exp^{-\beta_B(H_B(i)+P_BV_i)}}{\exp^{-\beta_A(H_A(i)+P_AV_i)}\exp^{-\beta_B(H_B(j)+P_BV_j)}}$$

$$= \exp\left[\beta_B(H_B(j) - H_B(i) + P_B(V_j - V_i)) - \beta_A(H_A(j) - H_A(i) + P_A(V_j - V_i))\right]$$

$$= \exp(\Delta)$$

$$\text{where } \Delta = \beta_B\big(H_B(j) - H_B(i) + P_B(V_j - V_i)\big) - \beta_A\big(H_A(j) - H_A(i) + P_A(V_j - V_i)\big). \tag{C.9}$$

These values of Δ are correct for highly general replica exchange simulations, where the Hamiltonians, temperatures and pressure could vary between replicas. In practise, some of these parameters are kept the same, e.g. parallel tempering keeps the pressure and Hamiltonian the same between replicas. Substituting $H = H_A = H_B$ and $P = P_A = P_B$ into these values of Δ gives the NVT and NPT parallel tempering tests,

$$\text{NVT, } \Delta = \beta_B\big(H(j) - H(i)\big) - \beta_A\big(H(j) - H(i)\big)$$

$$= \big(\beta_B - \beta_A\big)\big(H(j) - H(i)\big)$$

$$\text{NPT, } \Delta = \beta_B\big(H(j) - H(i) + P(V_j - V_i)\big) - \beta_A\big(H(j) - H(i) + P(V_j - V_i)\big)$$

$$= \big(\beta_B - \beta_A\big)\big(H(j) - H(i) + P\big(V_j - V_i\big)\big). \tag{C.10}$$

Hamiltonian replica exchange, which forms the basis of RETI and REFEP, only changes the Hamiltonians between replicas. Thus substituting $\beta = \beta_A = \beta_B$ and $P = P_A = P_B$ into the value of Δ yields,

$$\Delta = \beta\big(H_B(j) - H_B(i) + P(V_j - V_i)\big) - \beta\big(H_A(j) - H_A(i) + P(V_j - V_i)\big)$$

$$= \beta\big(H_B(j) - H_B(i) - H_A(j) + H_A(i)\big) + \beta P\big(V_j - V_i - V_j + V_i\big) \tag{C.11}$$

$$= \beta\big(H_B(j) - H_B(i) - H_A(j) + H_A(i)\big).$$

The volumes are seen to drop out of this test, so both the NVT and NPT tests are equivalent. It is useful to note that this Hamiltonian replica exchange test is just the product of two normal Metropolis Monte Carlo tests; one checking if the change in configuration of j to i is valid for forcefield B, and one checking if the change in configuration of i to j is valid for forcefield A. Hamiltonian replica exchange uses the product of these two tests as both moves happen simultaneously.

Appendix D

Beowulf Clusters

Computers are continually getting faster. At the start of this project, the fastest Intel processors had a clockspeed of 500 MHz. Today, similarly priced processors have a clockspeed of over 3000 MHz. In 1965, Moore[147] speculated that the number of components that could be economically fitted into a single integrated circuit would grow exponentially. Interpretation of this statement by later workers led to *Moore's Law*, which predicts that the processor-power of economically viable computers would double every two years. Despite nearly forty years passing since the original prediction, figure D.1 shows that Moore's law has not only been maintained, but it has even accelerated. The effect of Moore's law is that computer power has become a commodity item. The commodity nature of home PCs can be harnessed via a Beowulf cluster. Lots of standard PCs can be connected together via cheap networking (ethernet or myrinet). The combined power of all of these computers can be harnessed simultaneously, either via the simultaneous running of multiple independent parts of a coarsely parallel job (e.g. multiple windows of a free energy simulation), or through using message passing software to make many individual nodes behave like a single, very fast node. The reasons for the success of Beowulf clusters is that they are now cheap and easy to set up. The processors and networking are commodity items, and are easy to purchase and install. In addition, the growth of the Linux and the Gnu Free Software movement has provided a powerful, enterprise-class operating system, which can be deployed over hundreds of computers without the threat of license fees. This has meant that cash-strapped

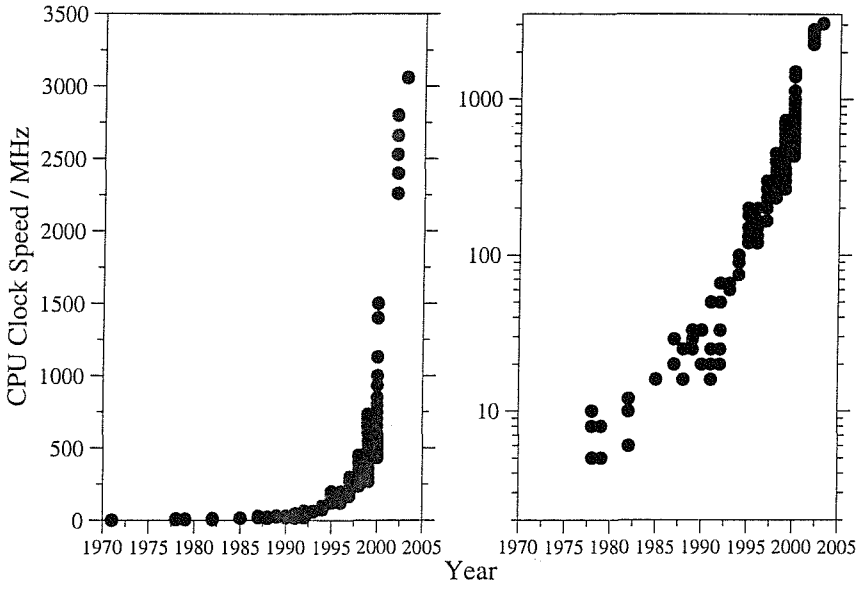


Figure D.1: The processor clock speed for all commodity Intel CPUs from 1974 to 2000. Data taken from 'General Processor Information', *http://bwrc.eecs.berkeley.edu/CIC/summary/*, Tom Burd, 2001. Data was added for the Intel 4004 (1971), and for the latest Intel Pentium 4 processors (*http://www.intel.com*). The data are shown on a linear and a logarithmic scale.

academics are able to purchase and build their own Beowulf clusters, and use existing free software tools to maximise their potential, and ease their support. This initial grass-roots growth of Beowulf clusters has since grown into corporate-, and university-wide adoption of the technology. Such huge purchases are now dealt with via dedicated cluster support companies, who can build Beowulf clusters that contain hundreds, or even thousands of individual nodes.

Appendix E

Error Analysis

To calculate the error on the free energy calculations, the averages for each block of 500 K steps were calculated. The average of these averages was the predicted free energy, and the error on this result was based on the estimate of the standard deviation over these averages, σ. The standard error, Δx, was calculated via,

$$\Delta x = \frac{\sigma}{\sqrt{N}}, \tag{E.1}$$

where N was the number of observations.

The errors were propagated across λ by calculating the maximum, and minimum possible free energies across λ. This was achieved by adding the errors onto the free energy for each λ-value, and then summing or integrating the results.

The errors on the calculated binding free energies were formed as a combination of the errors on the bound and free legs. If Δx_{bound} was the error on the bound leg, and Δx_{free} was the error on the free leg, then the error on the binding free energy, $\Delta x_{binding}$ was calculated via,

$$\Delta x_{binding} = \sqrt{\Delta x_{bound}^2 + \Delta x_{free}^2}. \tag{E.2}$$

Appendix F

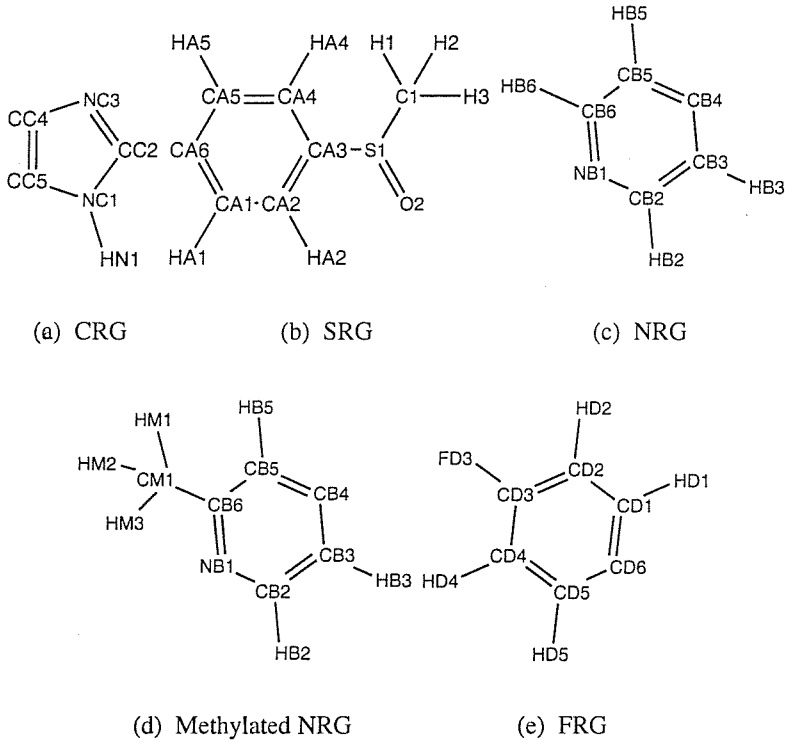
Setup of the p38 Simulations

F.1 Residue List

The residues and titration states used for the simulations on p38 MAP kinase are listed below. The residues with flexible side-chains are shown in bold.

GLU(4)-ARG(5)-PRO(6)-THR(7)-PHE(8)-TYR(9)-ARG(10)-
GLN(11)-GLU(12)-LEU(13)-ASN(14)-LYS(15)-**THR(16)**-ILE(17)-
TRP(18)-GLU(19)-VAL(20)-PRO(21)-GLU(22)-ARG(23)-TYR(24)-
GLN(25)-ASN(26)-LEU(27)-**SER(28)-PRO(29)-VAL(30)-GLY(31)-**
SER(32)-GLY(33)-ALA(34)-TYR(35)-GLY(36)-SER(37)-VAL(38)-
CYS(39)-ALA(40)-ALA(41)-PHE(42)-ASP(43)-THR(44)-LYS(45)-
THR(46)-GLY(47)-LEU(48)-**ARG(49)-VAL(50)-ALA(51)-VAL(52)-**
LYS(53)-LYS(54)-LEU(55)-SER(56)-ARG(57)-PRO(58)-PHE(59)-
GLN(60)-SER(61)-ILE(62)-ILE(63)-HIS(64)-ALA(65)-LYS(66)-
ARG(67)-THR(68)-TYR(69)-**ARG(70)-GLU(71)-LEU(72)**-ARG(73)-
LEU(74)-LEU(75)-LYS(76)-HIS(77)-MET(78)-LYS(79)-HIS(80)-
GLU(81)-ASN(82)-VAL(83)-ILE(84)-GLY(85)-LEU(86)-LEU(87)-
ASP(88)-**VAL(89)**-PHE(90)-THR(91)-PRO(92)-ALA(93)-ARG(94)-
SER(95)-LEU(96)-GLU(97)-GLU(98)-PHE(99)-ASN(100)-ASP(101)-
VAL(102)-TYR(103)-LEU(104)-VAL(105)-THR(106)-HIS(107)-LEU(108)-
MET(109)-GLY(110)-ALA(111)-ASP(112)-LEU(113)-ASN(114)-**ASN(115)**-
ILE(116)-VAL(117)-LYS(118)-CYS(119)-GLN(120)-LYS(121)-LEU(122)-
THR(123)-ASP(124)-ASP(125)-HIS(126)-VAL(127)-GLN(128)-PHE(129)-
LEU(130)-ILE(131)-TYR(132)-GLN(133)-ILE(134)-LEU(135)-ARG(136)-

GLY(137)-LEU(138)-LYS(139)-TYR(140)-ILE(141)-HIS(142)-SER(143)-
ALA(144)-ASP(145)-ILE(146)-ILE(147)-**HIS(148)**-ARG(149)-**ASP(150)**-
LEU(151)-**LYS(152)-PRO(153)-SER(154)-ASN(155)-LEU(156)-ALA(157)-**
VAL(158)-ASN(159)-GLU(160)-ASP(161)-CYS(162)-GLU(163)-LEU(164)-
LYS(165)-ILE(166)-LEU(167)-ASP(168)-PHE(169)-GLY(170)-LEU(171)-
ALA(172)-ARG(173)-HIS(174)-THR(175)-ASP(176)-ASP(177)-**GLU(178)**-
MET(179)-THR(180)-GLY(181)-TYR(182)-VAL(183)-ALA(184)-THR(185)-
ARG(186)-TRP(187)-TYR(188)-ARG(189)-ALA(190)-PRO(191)-GLU(192)-
ILE(193)-MET(194)-LEU(195)-ASN(196)-TRP(197)-MET(198)-HIS(199)-
TYR(200)-ASN(201)-GLN(202)-THR(203)-VAL(204)-ASP(205)-ILE(206)-
TRP(207)-SER(208)-VAL(209)-GLY(210)-CYS(211)-ILE(212)-MET(213)-
ALA(214)-GLU(215)-LEU(216)-LEU(217)-THR(218)-GLY(219)-ARG(220)-
THR(221)-LEU(222)-PHE(223)-PRO(224)-GLY(225)-THR(226)-ASP(227)-
HID(228)-ILE(229)-ASP(230)-GLN(231)-LEU(232)-LYS(233)-LEU(234)-
ILE(235)-LEU(236)-ARG(237)-LEU(238)-VAL(239)-GLY(240)-THR(241)-
PRO(242)-GLY(243)-ALA(244)-GLU(245)-LEU(246)-LEU(247)-LYS(248)-
LYS(249)-ILE(250)-SER(251)-SER(252)-GLU(253)-SER(254)-ALA(255)-
ARG(256)-ASN(257)-TYR(258)-ILE(259)-GLN(260)-SER(261)-LEU(262)-
THR(263)-GLN(264)-MET(265)-PRO(266)-LYS(267)-MET(268)-ASN(269)-
PHE(270)-ALA(271)-ASN(272)-VAL(273)-PHE(274)-ILE(275)-GLY(276)-
ALA(277)-ASN(278)-PRO(279)-LEU(280)-ALA(281)-VAL(282)-ASP(283)-
LEU(284)-LEU(285)-GLU(286)-LYS(287)-MET(288)-LEU(289)-VAL(290)-
LEU(291)-ASP(292)-SER(293)-ASP(294)-LYS(295)-ARG(296)-ILE(297)-
THR(298)-ALA(299)-ALA(300)-GLN(301)-ALA(302)-LEU(303)-ALA(304)-
HIS(305)-ALA(306)-TYR(307)-PHE(308)-ALA(309)-GLN(310)-TYR(311)-
HIS(312)-ASP(313)-PRO(314)-ASP(315)-ASP(316)-GLU(317)-PRO(318)-
VAL(319)-ALA(320)-ASP(321)-PRO(322)-TYR(323)-ASP(324)-GLN(325)-
SER(326)-PHE(327)-GLU(328)-SER(329)-ARG(330)-ASP(331)-LEU(332)-
LEU(333)-ILE(334)-ASP(335)-GLU(336)-TRP(337)-LYS(338)-SER(339)-
LEU(340)-THR(341)-TYR(342)-ASP(343)-GLU(344)-VAL(345)-ILE(346)-
SER(347)-PHE(348)-VAL(349)-PRO(350)-PRO(351)-PRO(352)-LEU(353)-
ASP(354)



(a) CRG (b) SRG (c) NRG

(d) Methylated NRG (e) FRG

Figure F.1: Names of atoms within the SB1-based residues used in these studies.

F.2 Description of the Ligands

F.2.1 Naming Conventions

The ligands were all divided into four residues. These residues were based on the four aromatic rings; the central imidazole ring (CRG), the sulfoxide/sulfide functionalised ring (SRG), the pyrrole group (NRG) and the fluorobenzene group (FRG). Each atom within each group was given a unique name, shown in figure F.1.

F.2.2 Partial Charges

The partial charges for each ligand are shown in table F.1. Ligands A and B were symmetrical, so had the same partial charges, as did ligands D and E.

F.2.3 Dihedral Parameters

The dihedral parameters derived for each of the seven fragments shown in figure 6.4, page 155, are shown in table F.2. This shows the amplitudes of the four cosine

Atom	A & B	C	D + E	Atom	A & B	C	D & E
CC2	0.165	0.188	0.163	CB2	0.004	0.004	0.014
NC3	-0.283	-0.294	-0.282	NB1	-0.309	-0.313	-0.317
CC4	0.021	0.022	0.018	CB6	0.004	0.004	0.074
CC5	0.021	0.012	0.022	CB5	-0.186	-0.187	-0.179
NC1	-0.522	-0.536	-0.521	HB2	0.178	0.177	0.176
HN1	0.423	0.423	0.423	HB3	0.164	0.164	0.153
CA6	0.028	-0.036	0.029	HB5	0.164	0.164	0.172
CA1	-0.134	-0.090	-0.134	HB6	0.178	0.177	N/A
CA2	-0.060	-0.143	-0.060	CM1	N/A	N/A	-0.192
CA3	-0.349	-0.092	-0.350	HM1	N/A	N/A	0.105
CA4	-0.060	-0.143	-0.060	HM2	N/A	N/A	0.105
CA5	-0.134	-0.090	-0.134	HM3	N/A	N/A	0.105
HA1	0.157	0.155	0.157	CD6	-0.026	-0.023	-0.024
HA2	0.163	0.153	0.163	CD1	-0.091	-0.092	-0.091
HA4	0.163	0.153	0.163	CD2	-0.186	-0.187	-0.187
HA5	0.157	0.155	0.157	CD3	0.075	0.074	0.075
S1	0.819	-0.031	0.819	CD4	-0.186	-0.187	-0.187
C1	-0.471	-0.262	-0.470	CD5	-0.091	-0.092	-0.091
O2	-0.615	N/A	-0.615	HD1	0.161	0.162	0.161
H1	0.138	0.122	0.138	HD2	0.162	0.162	0.162
H2	0.138	0.122	0.138	HD4	0.162	0.162	0.162
H3	0.138	0.122	0.138	HD5	0.161	0.162	0.161
CB4	0.028	0.031	0.031	FD3	-0.083	-0.085	-0.084
CB3	-0.186	-0.187	-0.206				

Table F.1: Partial charges used for ligands A to E. The atom names are taken from figure F.1.

functions used in equation 2.6 on page 8. The phases of these cosines were all zero.

Fragment	v_0	v_1	v_2	v_3
1	-0.3324	0.0629	5.9564	0.0180
2	-0.1425	0.0042	5.8251	0.0076
3	-0.3010	0.0263	5.9003	0.0106
4	0.0000	-0.2000	2.4000	0.2000
5	0.2735	-0.0066	2.7049	-0.0824
6	0.0000	-0.2000	2.4000	0.2000
7	0.5372	-0.3253	7.7442	-0.1281

Table F.2: Dihedral parameters derived for fragments 1 to 7 in figure 6.4.

F.2.4 Z-matrices

The z-matrix used for the perturbation of ligand D into ligand A. A similar z-matrix was used for the perturbation of ligand E to ligand A.

```
SB2 methyl
  1 DU1   -1   -1    0    0.000000   0     0.000000   0     0.000000 UNK   -1
  2 DU2   -1   -1    1    1.000000   0     0.000000   0     0.000000 UNK   -1
  3 DU3   -1   -1    2    1.000000   1    90.000000   0     0.000000 UNK   -1
  4 CC2  800  500    2    0.324835   1   117.866681   3    89.790641 CRG    1
  5 NC3  801  501    4    1.364000   2    82.279937   1   304.888886 CRG    1
  6 CC4  802  502    5    1.394696   4   106.537916   2    60.613635 CRG    1
  7 CC5  803  503    6    1.424709   4    73.385011   5   179.832130 CRG    1
  8 NC1  804  504    7    1.397531   4    36.246165   6   180.481055 CRG    1
  9 HN1  805  505    8    0.986714   7   125.415970   6   181.167514 CRG    1
 10 CA6  806  506    4    1.462647   8   123.795579   7   180.164105 SRG    2
 11 CA1  807  507   10    1.401947   4   121.524161   8     0.487547 SRG    2
 12 CA2  808  508   11    1.393279  10   120.145168   4   180.329433 SRG    2
 13 CA3  809  509   12    1.397458  10    90.871347  11   179.832720 SRG    2
 14 CA4  810  510   13    1.397511  11    89.217953  12   179.925054 SRG    2
 15 CA5  811  511   14    1.391886  10    30.132659  13   180.156064 SRG    2
 16 HA1  812  512   11    1.098897  10   121.537112  12   179.694978 SRG    2
 17 HA2  813  513   12    1.105213  11   120.080419  13   180.317551 SRG    2
 18 HA4  814  514   14    1.102066  13   119.766603  15   179.952501 SRG    2
 19 HA5  815  515   15    1.101926  14   119.666105  10   179.901405 SRG    2
 20  S1  816  516   13    1.713986  12   121.137360  11   175.820816 SRG    2
 21  C1  817  517   20    1.739784  13   100.941834  12   115.663042 SRG    2
 22  O2  818  518   20    1.489547  21   105.493544  13   110.464462 SRG    2
 23  H1  819  519   21    1.112736  20   111.131027  13    53.606941 SRG    2
 24  H2  820  520   21    1.113494  23   107.846825  20   125.831942 SRG    2
 25  H3  821  521   21    1.116033  23   108.321649  20   241.350611 SRG    2
 26 CB4  822  522    7    1.450108   6   131.628147   5   181.976367 NRG    3
 27 CB3  823  523   26    1.403352   7   121.063223   6   214.912948 NRG    3
 28 CB2  824  524   27    1.405239  26   118.255113   7   181.019712 NRG    3
 29 NB1  825  525   28    1.343811  26    93.319559  27   180.148989 NRG    3
 30 CB6  826  526   29    1.356729  27    88.901998  28   179.762972 NRG    3
 31 CB5  827  527   30    1.413022  26    30.177406  29   180.064401 NRG    3
 32 HB2  828  528   28    1.106035  27   120.436558  29   180.008651 NRG    3
 33 HB3  829  529   27    1.096614  26   121.626505  28   181.124751 NRG    3
 34 HB5  830  530   31    1.100047  26   120.491557  30   179.644196 NRG    3
 35 CM1  831  531   30    1.494358  31   117.874998  29   180.181157 NRG    3
 36 HM1  832  100   35    1.117826  30   112.301055  31   179.997870 NRG    3
 37 HM2  833  100   35    1.117652  30   109.286162  36   120.598299 NRG    3
 38 HM3  834  100   35    1.118065  30   109.296627  36   239.419292 NRG    3
 39 CD6  835  532    6    1.453783   5   122.423346   4   178.785970 FRG    4
 40 CD1  836  533   39    1.402924   6   120.161186   5    38.467467 FRG    4
 41 CD2  837  534   40    1.390335  39   120.606188   6   180.836376 FRG    4
 42 CD3  838  535   41    1.407432  39    89.759307  40   180.286163 FRG    4
 43 CD4  839  536   42    1.406859  40    90.167447  41   180.023038 FRG    4
 44 CD5  840  537   43    1.390016  42   119.493263  39     0.055165 FRG    4
 45 HD1  841  538   40    1.102009  39   119.876093  41   179.686544 FRG    4
 46 HD2  842  539   41    1.099698  40   120.698341  42   179.793746 FRG    4
 47 HD4  843  540   43    1.099210  42   119.733293  44   179.751446 FRG    4
```

```
  48 HD5   844   541    44     1.101851   43   119.689980   39   180.023779 FRG     4
  49 FD3   845   542    42     1.353693   41   119.922013   43   179.964283 FRG     4
TERZ
  50 CAP    -1    -1     2     0.324835    1   117.866681    3    89.790641 CAP     5
                     Geometry Variations follow    (2I4,F12.6)
  35    1     1.105601
  36    1     0.200000
  37    1     0.200000
  38    1     0.200000
     ***************Variable Bonds follow         (I4)
     ****************Additional Bonds follow       (2I4)
     ****************Harmonic Constraints follow   (2I4,4F10.4)
     ****************Variable Bond Angles follow   (I4)
  23
  21
  20
  11
  10
  27
  26
  40
  39
  49
     ****************Additional Bond Angles follow (3I4)
  24   21   20
  25   21   20
  23   21   24
  23   21   25
  24   21   25
  14   13   20
  15   10    4
  10    4    5
  31   26    7
  26    7    8
   6   39   44
   7    6   39
  43   42   49
       *****************Variable Dihedrals follow     (3I4,F12.6)
  23 606 606     5.000000
  21 602 602     5.000000
  11 601 601     5.000000
  27 614 604     5.000000
  36 612 100     5.000000
  40 603 603     5.000000
       *****************Additional Dihedrals follow   (6I4)
  13   20   21   24 606 606
  13   20   21   25 606 606
  21   20   13   14 602 602
  23   21   20   22 609 609
  24   21   20   22 609 609
  25   21   20   22 609 609
  36   35   30   29 613 100
  31   30   35   37 612 100
  37   35   30   29 613 100
  31   30   35   38 612 100
  38   35   30   29 613 100
                   Domain Definitions follow     (4I4)
                   Excluded Atoms List follows   (10I4)
                   Final blank line
```

The z-matrix used for the perturbation of ligand A into ligand C. A similar z-matrix was used for the perturbation of B to C.

```
SB2 inhibitor
  1 DU1    -1    -1     0     0.000000    0     0.000000    0     0.000000 UNK    -1
  2 DU2    -1    -1     1     1.000000    0     0.000000    0     0.000000 UNK    -1
  3 DU3    -1    -1     2     1.000000    1    90.000000    0     0.000000 UNK    -1
  4 CC2   500   700     2     0.327667    1    36.175083    3   108.952802 CRG     1
  5 NC3   501   701     4     1.364000    2   112.193133    1   302.540728 CRG     1
  6 CC4   502   702     5     1.394981    4   106.577290    2   138.391550 CRG     1
  7 CC5   503   703     6     1.424713    4    73.369384    5   179.790145 CRG     1
  8 NC1   504   704     7     1.398330    4    36.247241    6   180.618341 CRG     1
  9 HN1   505   705     8     0.986467    7   125.492422    6   181.305598 CRG     1
 10 CA6   506   706     4     1.463457    8   123.818787    7   180.376831 SRG     2
```

```
  11 CA1   507   707    10     1.401939    4   121.501854    8     0.242136 SRG     2
  12 CA2   508   708    11     1.392684   10   120.113638    4   180.043247 SRG     2
  13 CA3   509   709    12     1.397431   10    90.880715   11   180.123981 SRG     2
  14 CA4   510   710    13     1.397511   11    89.241684   12   179.752731 SRG     2
  15 CA5   511   711    14     1.391877   10    30.156403   13   180.121739 SRG     2
  16 HA1   512   712    11     1.099770   10   121.512740   12   179.963590 SRG     2
  17 HA2   513   713    12     1.105143   11   120.101024   13   180.180503 SRG     2
  18 HA4   514   714    14     1.101592   13   119.836444   15   180.011804 SRG     2
  19 HA5   515   715    15     1.100833   14   119.714173   10   179.914402 SRG     2
  20  S1   516   716    13     1.713986   12   121.121536   11   175.388576 SRG     2
  21  C1   517   717    20     1.739228   13   101.001762   12   115.823987 SRG     2
  22  O2   518   100    20     1.491105   21   105.426158   13   110.399396 SRG     2
  23  H1   519   719    21     1.113638   20   111.081508   13    52.979012 SRG     2
  24  H2   520   720    21     1.113307   23   107.972360   20   125.761236 SRG     2
  25  H3   521   721    21     1.115991   23   108.357752   20   241.306387 SRG     2
  26 CB4   522   722     7     1.448992    6   131.594616    5   181.902633 NRG     3
  27 CB3   523   723    26     1.404542    7   121.141111    6   214.739842 NRG     3
  28 CB2   524   724    27     1.406165   26   118.537277    7   181.022109 NRG     3
  29 NB1   525   725    28     1.346094   26    93.009871   27   180.139409 NRG     3
  30 CB6   526   726    29     1.347016   27    88.313340   28   179.747867 NRG     3
  31 CB5   527   727    30     1.405453   26    30.712567   29   180.129953 NRG     3
  32 HB2   528   728    28     1.105778   27   120.479450   29   180.021160 NRG     3
  33 HB3   529   729    27     1.097361   26   121.470845   28   181.102897 NRG     3
  34 HB5   530   730    31     1.098789   26   120.669087   30   179.600224 NRG     3
  35 HB6   531   731    30     1.105601   31   120.496874   29   180.188089 NRG     3
  36 CD6   532   732     6     1.453492    5   122.425237    4   178.832412 FRG     4
  37 CD1   533   733    36     1.402924    6   120.142635    5    38.046760 FRG     4
  38 CD2   534   734    37     1.389793   36   120.626477    6   180.861012 FRG     4
  39 CD3   535   735    38     1.406995   36    89.772033   37   180.296239 FRG     4
  40 CD4   536   736    39     1.407872   37    90.169801   38   180.079103 FRG     4
  41 CD5   537   737    40     1.389447   39   119.487976   36     0.055860 FRG     4
  42 HD1   538   738    37     1.102198   36   119.908437   38   179.766529 FRG     4
  43 HD2   539   739    38     1.099684   37   120.729067   39   179.840926 FRG     4
  44 HD4   540   740    40     1.099624   39   119.739260   41   179.747120 FRG     4
  45 HD5   541   741    41     1.101145   40   119.698137   36   179.949296 FRG     4
  46 FD3   542   742    39     1.354393   38   119.977212   40   179.880543 FRG     4
TERZ
  47 CAP    -1    -1     2     0.327667    1    36.175083    3   108.952802 CAP     5
                      Geometry Variations follow     (2I4,F12.6)
  22   1     0.200000
  24   2   108.592580
  25   2   108.915149
  20   3   179.736187
  24   3   125.327637
  25   3   243.778068
     ***************Variable Bonds follow          (I4)
     ***************Additional Bonds follow        (2I4)
     ***************Harmonic Constraints follow    (2I4,4F10.4)
     ***************Variable Bond Angles follow    (I4)
  23
  21
  20
  11
  10
  27
  26
  37
  36
  46
     ***************Additional Bond Angles follow (3I4)
  24  21  20
  25  21  20
  23  21  24
  23  21  25
  24  21  25
  14  13  20
  15  10   4
  10   4   5
  31  26   7
  26   7   8
   6  36  41
   7   6  36
  40  39  46
       ***************Variable Dihedrals follow      (3I4,F12.6)
  23 606 610    5.000000
```

```
21 602 611     5.000000
11 601 601     5.000000
27 604 604     5.000000
37 603 603     5.000000
      ****************Additional Dihedrals follow    (6I4)
13   20   21   24 606 610
13   20   21   25 606 610
21   20   13   14 602 611
23   21   20   22 609 100
24   21   20   22 609 100
25   21   20   22 609 100
                 Domain Definitions follow      (4I4)
                 Excluded Atoms List follows    (10I4)
                 Final blank line
```

References

[1] A. Wlodawer, *Annu. Rev. Med.*, **53**, 595, (2002).

[2] R. D. Taylor, P. J. Jewsbury and J. W. Essex, *J. Comput. Aid. Mol. Des.*, **16**, 151, (2002).

[3] J. G. Morris, *A Biologist's Physical Chemistry*, Edward Arnold, Sevenoaks, Kent, UK, (1991).

[4] C. J. Cramer, *Essentials of Computational Chemistry, Theories and Models*, John Wiley and Sons, Ltd., Chichester, UK, (2002).

[5] W. L. Jorgensen, J. D. Madura and C. J. Swenson, *J. Am. Chem. Soc.*, **106**, 6638, (1984).

[6] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, **118**, 11225, (1996).

[7] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179, (1995).

[8] J. H. Lii and N. L. Allinger, *J. Comput. Chem.*, **12**, 186, (1991).

[9] 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, *J. Phys. Chem. B*, **102**, 3586, (1998).

[10] N. A. McDonald and W. L. Jorgensen, *J. Phys. Chem. B*, **102**, 8049, (1998).

[11] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford, UK, (2001).

[12] A. R. Leach, *Molecular Modelling, Principles and Applications*, Longman, Harlow, UK, (1996).

[13] W. L. Jorgensen, MCPRO 1.5, (1996).

[14] T. A. Halgren and W. Damm, *Curr. Opin. Struc. Biol.*, **11**, 236, (2001).

[15] G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. X. Cao, R. B. Murphy, R. H. Zhou and T. A. Halgren, *J. Comput. Chem.*, **23**, 1515, (2002).

[16] G. G. Ferenczy and C. A. Reynolds, *J. Phys. Chem. A*, **105**, 11470, (2001).

[17] D. A. McQuarrie, *Statistical Mechanics*, Harper and Row, New York, USA, (1976).

[18] P. A. Rejto and S. T. Freer, *Prog. Biophys. Molec. Biol.*, **66**, 167, (1996).

[19] P. W. Atkins, *Physical Chemistry, Fifth Edition*, Oxford University Press, Oxford, UK, (1995).

[20] R. W. Zwanzig, *J. Chem. Phys.*, **22**, 1420, (1954).

[21] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.*, **21**, 1087, (1953).

[22] R. H. Wood, W. C. F. Muhlbauer and P. T. Thompson, *J. Phys. Chem.*, **95**, 6670, (1991).

[23] C. A. Reynolds, P. M. King and W. G. Richards, *Mol. Phys.*, **76**, 251, (1992).

[24] C. Chipot and D. A. Pearlman, *Mol. Simulat.*, **28**, 1, (2002).

[25] C. Jarzynski, *Phys. Rev. E*, **56**, 5018, (1997).

[26] C. Jarzynski, *Phys. Rev. Lett.*, **78**, 2690, (1997).

[27] M. L. P. Price and W. L. Jorgensen, *J. Am. Chem. Soc.*, **122**, 9455, (2000).

[28] D. Q. McDonald and W. C. Still, *J. Am. Chem. Soc.*, **118**, 2073, (1996).

[29] D. J. Price and W. L. Jorgensen, *Bioorg. Med. Chem. Lett.*, **10**, 2067, (2000).

[30] D. J. Price and W. L. Jorgensen, *J. Comput. Aid. Mol. Des.*, **15**, 681, (2001).

[31] M. L. Lamb and W. L. Jorgensen, *J. Med. Chem.*, **41**, 3928, (1998).

[32] J. Vacek and P. A. Kollman, *J. Phys. Chem. A*, **103**, 10015, (1999).

[33] W. L. Jorgensen and C. Ravimohan, *J. Chem. Phys.*, **83**, 3050, (1985).

[34] D. A. Pearlman and P. A. Kollman, *J. Chem. Phys.*, **90**, 2460, (1989).

[35] D. A. Pearlman and P. S. Charifson, *J. Med. Chem.*, **44**, 3417, (2001).

[36] B. C. Oostenbrink, J. W. Pitera, M. M. H. Lipzig, J. H. N. Meerman and W. F. Van Gunsteren, *J. Med. Chem.*, **43**, 4594, (2000).

[37] X. Barril, M. Orozco and F. J. Luque, *J. Med. Chem.*, **42**, 5110, (1999).

[38] J. T. Wescott, L. R. Fisher and S. Hanna, *J. Chem. Phys.*, **116**, 2361, (2002).

[39] J. R. Blas, M. Marquez, J. L. Sessler, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, **124**, 12796, (2002).

[40] A. Melo and M. J. Ramos, *J. Mol. Struc-Theochem.*, **580**, 251, (2002).

[41] M. Mezei, *J. Chem. Phys.*, **86**, 7084, (1987).

[42] C. R. W. Guimaraes and R. B. Alencastro, *Int. J. Quantum Chem.*, **85**, 713, (2001).

[43] C. R. W. Guimaraes and R. B. Alencastro, *J. Med. Chem.*, **45**, 4995, (2002).

[44] C. R. W. Guimaraes and R. B. Alencastro, *J. Phys. Chem. B*, **106**, 466, (2002).

[45] S. Kamath, E. Coutinho and P. Desai, *J. Biomol. Struct. Dyn.*, **16**, 1239, (1999).

[46] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing, Second edition*, Cambridge University Press, Cambridge, UK, (1996).

[47] H. Hu, R. H. Yun and J. Hermans, *Mol. Simulat.*, **28**, 67, (2002).

[48] G. Hummer, *J. Chem. Phys.*, **114**, 7330, (2001).

[49] C. Jarzynski, *Proc. Natl. Acad. Sci.*, **98**, 3636, (2001).

[50] D. A. Hendrix and C. Jarzynski, *J. Chem. Phys.*, **114**, 5974, (2001).

[51] G. Hummer, *Mol. Simulat.*, **28**, 81, (2002).

[52] B. Roux, *Comput. Phys. Commun.*, **91**, 275, (1995).

[53] C. Bartels and M. Karplus, *J. Comput. Chem.*, **18**, 1450, (1997).

[54] S. Kumar, P. W. Payne and M. Vasquez, *J. Comput. Chem.*, **17**, 1269, (1996).

[55] R. W. W. Hooft, B. P. Vaneijck and J. Kroon, *J. Chem. Phys.*, **97**, 6690, (1992).

[56] O. Engkvist and G. Karlstrom, *Chem. Phys.*, **213**, 63, (1996).

[57] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, **13**, 1011, (1992).

[58] X. J. Kong and C. L. Brooks, *J. Chem. Phys.*, **105**, 2414, (1996).

[59] S. Banba, Z. Y. Guo and C. L. Brooks, *J. Phys. Chem. B*, **104**, 6903, (2000).

[60] Z. Guo, C. L. Brooks and X. Kong, *J. Phys. Chem. B*, **102**, 2032, (1998).

[61] G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, **23**, 187, (1977).

[62] T. N. Heinz, W. F. Van Gunsteren and P. H. Hunenberger, *J. Chem. Phys.*, **115**, 1125, (2001).

[63] C. Bartels, M. Schaefer and M. Karplus, *J. Chem. Phys.*, **111**, 8048, (1999).

[64] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, **16**, 1339, (1995).

[65] C. J. Woods, S. Camiolo, M. E. Light, S. J. Coles, M. B. Hursthouse, M. A. King, P. A. Gale and J. W. Essex, *J. Am. Chem. Soc.*, **124**, 8644, (2002).

[66] M. R. Reddy and M. D. Erion, *J. Comput. Chem.*, **20**, 1018, (1999).

[67] J. Åqvist, C. Medina and J. E. Samuelsson, *Protein Eng.*, **7**, 385, (1994).

[68] J. Åqvist and T. Hansson, *J. Phys. Chem.*, **100**, 9512, (1996).

[69] J. Åqvist and S. L. Mowbray, *J. Biol. Chem.*, **270**, 9978, (1995).

[70] T. Hansson and J. Åqvist, *Protein Eng.*, **8**, 1137, (1995).

[71] R. C. Rizzo, M. Udier-Blagovic, D. Wang, E. K. Watkins, M. B. Kroeger Smith, R. H. Smith Jr., J. Tirado-Rives and W. L. Jorgensen, *J. Med. Chem.*, **45**, 2970, (2002).

[72] T. J. Hou, W. Zhang and X. J. Xu, *J. Phys. Chem. B*, **105**, 5304, (2001).

[73] R. H. Zhou, R. A. Friesner, A. Ghosh, R. C. Rizzo, W. L. Jorgensen and R. M. Levy, *J. Phys. Chem. B*, **105**, 10388, (2001).

[74] I. D. Wall, A. R. Leach, D. W. Salt, M. G. Ford and J. W. Essex, *J. Med. Chem.*, **42**, 5142, (1999).

[75] M. E. Davis and J. A. McCammon, *Chem. Rev.*, **90**, 509, (1990).

[76] B. Honig, K. Sharp and A. S. Yang, *J. Phys. Chem.*, **97**, 1101, (1993).

[77] C. J. Woods, M. A. King and J. W. Essex, *J. Comput. Aid. Mol. Des.*, **15**, 129, (2001).

[78] M. A. L. Eriksson, J. Pitera and P. A. Kollman, *J. Med. Chem.*, **42**, 868, (1999).

[79] J. Shen and J. Wendoloski, *J. Comput. Chem.*, **17**, 350, (1996).

[80] L. T. Chong, Y. Duan, L. Wang, I. Massova and P. A. Kollman, *Proc. Natl. Acad. Sci.*, **96**, 14330, (1999).

[81] C. M. Reyes and P. A. Kollman, *J. Mol. Biol.*, **295**, 1, (2000).

[82] J. M. Wang, P. Morin, W. Wang and P. A. Kollman, *J. Am. Chem. Soc.*, **123**, 5221, (2001).

[83] H. Gouda, I. D. Kuntz, D. A. Case and P. A. Kollman, *Biopolymers*, **68**, 16, (2003).

[84] R. Crovetto, R. Fernandez-Prini and M. L. Japas, *J. Chem. Phys.*, **76**, 1077, (1982).

[85] P. H. K. Dejong, J. E. Wilson, G. W. Neilson and A. D. Buckingham, *Mol. Phys.*, **91**, 99, (1997).

[86] J. Hernandez-Cobos, A. D. Mackie and L. F. Vega, *J. Chem. Phys.*, **114**, 7527, (2001).

[87] T. H. Zhu, J. B. Li, G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *J. Chem. Phys.*, **109**, 9117, (1998).

[88] T. Head-Gordon and G. Hura, *Chem. Rev.*, **102**, 2651, (2002).

[89] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, **79**, 926, (1983).

[90] D. A. Pearlman, *J. Phys. Chem.*, **98**, 1487, (1994).

[91] J. C. Owicki and H. A. Scheraga, *Chem. Phys. Lett.*, **47**, 600, (1977).

[92] W. L. Jorgensen, *J. Phys. Chem.*, **87**, 5304, (1983).

[93] J. G. Kirkwood, *J. Chem. Phys.*, **7**, 919, (1939).

[94] C. Chipot, P. A. Kollman and D. A. Pearlman, *J. Comput. Chem.*, **17**, 1112, (1996).

[95] P. H. Axelsen and D. H. Li, *J. Comput. Chem.*, **19**, 1278, (1998).

[96] Y. Iba, *Int. J. Mod. Phys. C*, **12**, 623, (2001).

[97] U. H. E. Hansmann and Y. Okamoto, *Phys. Rev. E*, **54**, 5863, (1996).

[98] U. H. E. Hansmann and Y. Okamoto, *J. Comput. Chem.*, **18**, 920, (1997).

[99] B. A. Berg and T. Neuhaus, *Phys. Lett. B*, **267**, 249, (1991).

[100] J. Higo, N. Nakajima, H. Shirai, A. Kidera and H. Nakamura, *J. Comput. Chem.*, **18**, 2086, (1997).

[101] N. Nakajima, J. Higo, A. Kidera and H. Nakamura, *Chem. Phys. Lett.*, **278**, 297, (1997).

[102] K. Sayano, H. Kono, M. M. Gromiha and A. Sarai, *J. Comput. Chem.*, **21**, 954, (2000).

[103] N. Nakajima, H. Nakamura and A. Kidera, *J. Phys. Chem. B*, **101**, 817, (1997).

[104] H. Shirai, N. Nakajima, J. Higo, A. Kidera and H. Nakamura, *J. Mol. Biol.*, **278**, 481, (1998).

[105] N. Nakajima, J. Higo, A. Kidera and H. Nakamura, *J. Mol. Biol.*, **296**, 197, (2000).

[106] C. Bartels and M. Karplus, *J. Phys. Chem. B*, **102**, 865, (1998).

[107] U. H. E. Hansmann, *Chem. Phys. Lett.*, **281**, 140, (1997).

[108] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, **314**, 141, (1999).

[109] T. Okabe, M. Kawata, Y. Okamoto and M. Mikami, *Chem. Phys. Lett.*, **335**, 435, (2001).

[110] H. Fukunishi, O. Watanabe and S. Takada, *J. Chem. Phys.*, **116**, 9058, (2002).

[111] D. Bedrov and G. D. Smith, *J. Chem. Phys.*, **115**, 1121, (2001).

[112] K. Y. Sanbonmatsu and A. E. Garcia, *Proteins*, **46**, 225, (2002).

[113] R. H. Zhou and B. J. Berne, *Proc. Natl. Acad. Sci.*, **99**, 12777, (2002).

[114] G. M. Verkhivker, P. A. Rejto, D. Bouzida, S. Arthurs, A. B. Colson, S. T. Freer, D. K. Gehlhaar, V. Larson, B. A. Luty, T. Marrone and P. W. Rose, *Chem. Phys. Lett.*, **337**, 181, (2001).

[115] Y. Sugita, A. Kitao and Y. Okamoto, *J. Chem. Phys.*, **113**, 6042, (2000).

[116] Z. R. Zheng, *Stoch. Proc. Appl.*, **104**, 131, (2003).

[117] P. A. Gale, J. L. Sessler, W. E. Allen, N. A. Tvermoes and V. Lynch, *Chem. Commun.*, pp. 665–666, (1997).

[118] P. A. Gale, L. J. Twyman, C. I. Handlin and J. L. Sessler, *Chem. Commun.*, pp. 1851–1852, (1999).

[119] H. Miyaji, P. Anzenbacher, J. L. Sessler, E. R. Bleasdale and P. A. Gale, *Chem. Commun.*, pp. 1723–1724, (1999).

[120] J. L. Sessler, P. A. Gale and J. W. Genge, *Chem. Eur. J.*, **4**, 1095, (1998).

[121] S. Camiolo and P. A. Gale, *Chem. Commun.*, pp. 1129–1130, (2000).

[122] Y. D. Wu, D. F. Wang and J. L. Sessler, *J. Org. Chem.*, **66**, 3739, (2001).

[123] W. P. van Hoorn and W. L. Jorgensen, *J. Org. Chem.*, **64**, 7439, (1999).

[124] N. A. McDonald, E. M. Duffy and W. L. Jorgensen, *J. Am. Chem. Soc.*, **120**, 5104, (1998).

[125] G. A. Kaminski and W. L. Jorgensen, *J. Phys. Chem. B*, **102**, 1787, (1998).

[126] R. Pomes, E. Eisenmesser, C. B. Post and B. Roux, *J. Chem. Phys.*, **111**, 3387, (1999).

[127] Z. L. Wang, B. J. Canagarajah, J. C. Boehm, S. Kassisa, M. H. Cobb, P. R. Young, S. Abdel-Meguid, J. L. Adams and E. J. Goldsmith, *Struct. Fold. Des.*, **6**, 1117, (1998).

[128] R. Newton, L. Cambridge, L. A. Hart, D. A. Stevens, M. A. Lindsay and P. J. Barnes, *Brit. J. Pharmacol.*, **130**, 1353, (2000).

[129] I. M. Mclay, F. Halley, J. E. Souness, J. Mckenna, V. Benning, M. Birrell, B. Burton, M. Belvisi, A. Collis, A. Constan, M. Foster, D. Hele, Z. Jayyosi, M. Kelley, C. Maslen, G. Miller, M. C. Ouldelhkim, K. Page, S. Phipps, K. Pollock, B. Porter, A. I. Ratcliffe, E. J. Redford, S. Webber, B. Slater, V. Thybaud and N. Wilsher, *Bioorgan. Med. Chem.*, **9**, 537, (2001).

[130] J. C. Boehm, J. M. Smietana, M. E. Sorenson, R. S. Garigipati, T. F. Gallagher, P. L. Sheldrake, J. Bradbeer, A. M. Badger, J. T. Laydon, J. C. Lee, L. M. Hillegass, D. E. Griswold, J. J. Breton, M. C. Chabotfletcher and J. L. Adams, *J. Med. Chem.*, **39**, 3929, (1996).

[131] K. U. Birkenkamp, L. M. L. Tuyt, C. Lummen, L. T. J. Wierenga, W. Kruijer and E. Vellenga, *Brit. J. Pharmacol.*, **131**, 99, (2000).

[132] J. C. Boehm and J. L. Adams, *Expert Opin. Ther. Pat.*, **10**, 25, (2000).

[133] G. Vriend, R. W. W. Hooft and D. Van Aalten, WhatIf, (1997).

[134] W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1666, (1988).

[135] W. L. Jorgensen and N. A. McDonald, *Theochem-J. Mol. Struc.*, **424**, 145, (1998).

[136] H. A. Carlson, T. B. Nguyen, M. Orozco and W. L. Jorgensen, *J. Comput. Chem.*, **14**, 1240, (1993).

[137] R. H. Henchman and J. W. Essex, *J. Comput. Chem.*, **20**, 483, (1999).

[138] J. W. Storer, D. J. Giesen, C. J. Cramer and D. G. Truhlar, *J. Comput. Aid. Mol. Des.*, **9**, 87, (1995).

[139] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902, (1985).

[140] G. D. Hawkins, D. J. Giesen, G. C. Lynch, C. C. Chambers, I. Rossi, J. W. Storer, J. Li, T. Zhu, P. Winget, D. Rinaldi, D. A. Liotard, C. J. Cramer and D. G. Truhlar, AMSOL 6.6, (1999).

[141] H. C. Cheng, *J. Pharmacol. Toxicol.*, **46**, 61, (2001).

[142] E. Leontidis, B. M. Forrest, A. H. Widmann and U. W. Suter, *J. Chem. Soc. Faraday T.*, **91**, 2355, (1995).

[143] J. W. Essex and W. L. Jorgensen, *J. Comput. Chem.*, **16**, 951, (1995).

[144] J. B. Li, T. H. Zhu, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. A*, **102**, 1820, (1998).

[145] A. Jakalian, B. L. Bush, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, **21**, 132, (2000).

[146] J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comp. Phys.*, **23**, 327, (1977).

[147] G. E. Moore, *Electronics*, **38**, 114, (1965).