University of Southampton

Faculty of Engineering Science and Mathematics

School of Chemistry

The Use of Fast Free Energy Methods in Rational Drug Design

by

Benjamin Philip Cossins

Thesis for the degree of Doctor of Philosophy

August 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

SCHOOL OF CHEMISTRY

Doctor of Philosophy

THE USE OF FAST FREE ENERGY METHODS IN RATIONAL DRUG DESIGN

by Benjamin Cossins

The computationally demanding nature and lack of generality of free energy methods are the main barriers to their common place use in rational drug design. This study investigates the possibility of producing protocols to accurately calculate the binding free energy of protein-ligand complexes more efficiently than presently established methods, using large scale distributed computing. There has been an explosion of useful nonequilibrium work methods recently, mainly due to the discovery of the Jarzynski equilibrium [Jarzynski(1997b)]. After an indepth investigation of these methods a subset, all with the possibility of large scale parallelisation, was chosen for further study. Also, replica exchange fast growth (REFG), was developed, a method which combines replica exchange and fast growth methods in a similar way to replica exchange thermodynamic integration (RETI) [Woods *et al.*(2003a)Woods, Essex & King]. These methods of interest were applied to a large number of harmonic oscillator systems and compared to the established method TI. Those methods deemed to perform best were then applied to some simple solute-solvent test systems and compared to the established method RETI. The best performing method from these studies was then compared to RETI for the calculation of relative binding free energies of two sets of cogeneric inhibitors bound to their receptor proteins. REFG was found to perform as well as RETI and produce constantly predictive results. REFG was able to produce these results in significantly less wall clock time by using large scale distributed computing.

Acknowledgements

First and foremost I would like to thank my family for their support and enthusiasm throughout my education. It is almost certainly the case that my love for science has come directly from my mother and father. An extremely large portion of thanks must also be directed towards my lovely girl friend Vicky who has supported me through this project with which she has sometimes had to compete for my time. Also, thanks to Vicky's family for their support.

Great thanks to my supervisor Jon Essex for all your guidance and teachings. It is a testament to Jon's general quality as a person that he always maintains a jovial and pleasant demeanour with everyone despite the many demands of his position. Also, many congratulations to Jon on having recently been made Professor Jon Essex. Many thanks to the Jon's research group who have all helped me in many ways throughout the last few years especially Julien Michel, Sebastien Foucher and Christopher Woods.

Many thanks must also be reserved for Dr. Colin Edge and GSK who have supported this work. Colin was always very helpful and generally a very nice bloke.

Last but of course by no means least, thank you to all my friends for putting up with my myriad of moans and groans through the time of this PhD and keeping me entertained.

Contents

Chapter 1

Rational drug design

1.1 Introduction

Our ability to find ways to treat the diseases which affect us is one of the major factors in the continual increase in quality and length of human life. Every year billions of dollars are invested in finding and designing drug molecules to help treat disease and health disorders. As with almost all aspects of human endeavour the advent of computers and the rapid increase in computational power has enabled new and highly sophisticated approaches to the problem. The use of computational techniques in drug design/discovery is now common place [Leach(1996)]. This study will be concerned with finding computational methodologies which can efficiently and accurately rank molecules on how well they bind to a target receptor using 3D structures of the candidate molecules bound to the target receptor. These calculations are common in the literature, but they are not, as yet, regularly used in the pharmaceutical industry, because of their computational cost and unreliability.

1.2 Computational Drug design

Most computational drug design techniques fall into four categories: *De novo* molecule design; docking; ligand based screening; binding affinity prediction techniques. *De novo* molecule design, docking and binding affinity prediction generally require that a 3D structure of the receptor understudy is available (receptor based techniques).

De novo design is the creation of novel ligand structures which may bind to particular targets with high affinity. This is a very difficult task and many studies rely on searching large combinatorial libraries of real and/or theoretical molecules. For example Kick *et al.* demon-

strated the coupling of combinatorial chemistry and structure based design resulting in the creation of a number of low nanomolar non-peptide inhibitors of Cathepsin D (Cat-D) [Kick *et al.*(1997)Kick, Roe, Skillman, Liu, Ewing, Sun, Kuntz & Ellman]. This study searched a library of suitable molecules for favourable groups, positioning each of them in turn on a non-peptide backbone, which was a stable mimetic of the tetrahedral intermediate of amide hydrolysis, within the Cat-D X-ray structure, in order to ascertain their suitability.

In the absence of three dimensional receptor information, virtual screening can also be carried out where there are known ligands for a protein (ligand-based virtual screening). Ligand based virtual screening uses descriptors to select molecules with similar properties to known highly active ligands, for further analysis [Leach(1996)]. These empirical binding affinity predictors are generally thought to have problems with transferability as they have been trained on only one set of data. Also, any advantageous interactions not represented in the training data set will not be present in any resulting high scoring binders.

Docking methodology attempts to find solutions to the problem of positioning ligand structures correctly within the binding sites of receptor molecules. Docking is generally used to screen large sets of lead-like molecules to yield a smaller set for further analysis (virtual screening). Often this problem is simplified by treating ligand and receptor structures as rigid and ignoring any solvent effects. In this simplified case docking is a matter of optimising geometries and interactions, in order to fit the ligand to the receptor [Leach(1996)] [Jhoti(2007)]. Even in this simplified case finding solutions is computationally expensive due to the enormous amount of possible ligand and receptor configurations. In the future, docking may be made easier by using knowledge of protein domains common in the proteome to predict binding sites. Current docking algorithms generally make many assumptions, but are still generally able to find real binding poses of drug molecules [Warren *et al.*(2006)Warren, Andrews, Capelli, Clarke, LaLonde, Lambert, Lindvall, Nevins, Semus, Senger, Tedesco, Wall, Woolven, Peishoff & Head]. From the findings of Warren *et al.* it seems that there is a need for more accurate scoring of possible bound structures to pick the correct poses from the possibilities suggested.

Prediction of binding affinity of a given drug candidate generally begins with a ligand-receptor complex structure. Often binding affinity prediction algorithms are bundled in with docking packages; in this case they are called scoring functions. The term scoring function

often refers to the both the potential which drives the optimisation of sampling towards finding possible ligand-receptor positions and the calculations which then rank the possible ligand poses. The latter and more demanding of these two jobs must rank ligand poses in line with experimental free energies (binding affinity scoring functions) [Jhoti(2007)]. Scoring functions are generally one of three types; force-field based, empirical or knowledge based.

Force-field based scoring functions use force-field terms, often with simplified non-bonded terms to describe the interactions between receptor and ligand. To describe solvent screening effects a distance dependent dielectric can be used as it is computationally inexpensive [Jhoti(2007)]. As calculation of force-field terms can be time consuming these functions often approximated to save time. These approximate scoring methods are prone to error as they often ignore entropy and many internal energy effects. Also, unless specific force-field parameters are generated the choice of parameters can be problematic.

Empirical scoring functions are based on the structure-activity relationships developed in the 1960s. Here the binding free energy is decomposed into a number of chemically relevant terms which are calculated individually and summed. Each term is found by multiplying an interaction score by a coefficient found previously through regression fitting to known experimental binding free energies. Common terms include hydrogen bonding, hydrophobic interactions, and internal energy. As with ligand based virtual screening these methods have problems as possibly advantageous interactions not represented in the training set will not be found in the highly ranked ligands [Jhoti(2007)].

Knowledge based methods use the sum of statistically derived potentials of mean force (PMF) scores between protein and ligand atoms from the Protein Data Bank (PDB) as a measure for protein-ligand binding affinity. These statistically derived PMFs are similar but not strictly equal to normal PMFs. Knowledge based methods have shown some improvement over empirical and force field-based methods [Klebe(2000)].

Scoring methods which could possibly discard useful candidates are obviously counter productive. Hence, the success of these docking/scoring methods is generally measured by how many runs it takes to obtain a particular percentage of useful candidates from a large data set. This method of displaying the performance of a docking/scoring method is called enrichment. These less rigorous methods of scoring binding affinity may not be very reliable, but they can be used to obtain an initial set of candidate ligands for higher level analysis by more time

consuming and accurate methods.

Accurate evaluation of drug candidate binding free energies can be achieved through the use of relatively realistic, complex molecular mechanical simulations of ligand-protein complexes and various thermodynamic calculations. These types of calculation are generally termed free energy calculations. Free energy calculations and the simulations which they use are discussed in detail in the next chapter.

The methods we discuss in this study rely on the presence of 3D structures of target receptors which can be time consuming to produce. Problematic assessment of best possible receptor target strategies and slow methods of elucidating 3D protein structures have been a major bottle neck to wide spread use of some of these structure based methods. Integrated use of systems biology and omics (genomics, proteomics, metabolomics, pharmacogenomics and interactomics etc.) techniques and data combined with high-throughput crystallography in the future may offer the ammunition for well directed and efficient structure based computational approaches [Blundell & Patel(2004), Cho *et al.*(2006)Cho, Labow, Reinhardt, van Oostrum & Peitsch].

1.3 Aims of this study

This study will concentrate on the fast, accurate, computational calculation of ligand binding free energies (fast free energy methods) with a view to screening of lead drug candidates (i.e. screening a smaller number of previously identified candidates). Thus, the fields of molecular simulation and free energy calculations will be reviewed with a view to finding new and possibly useful fast free energy methods. These methods will then be investigated and compared to those currently in general use.

Chapter 2

Calculating the Free Energy

2.1 Molecular Simulation

2.1.1 Introduction

In the field of molecular modelling it has become apparent that a single conformation cannot properly represent a molecule. Molecular simulation is the calculation of the prevalence of possible states of a molecular system and generally takes two forms: Molecular Dynamics (MD) and Monte Carlo (MC) simulations. Both forms of simulation use a potential expression to describe the different forces acting on each atom with calculated atom point charges and associated bond, bond angle and dihedral parameters (a molecular mechanics force field) [Atkins(1994)]. Using statistical mechanics it is possible to extract important macroscopic properties from microscopic data which, with accurate simulations, can be used to estimate experimental properties. This is the basis of the ligand binding affinity methods described in this study.

Although the focus of this study is entirely towards classical molecular mechanical simulations it is convenient to consider some of the basics through quantum theory. In quantum theory we are able to consider each individual microstate of a system whereas classical theory presents us with a continuum of states.

2.1.2 Thermodynamics

The word thermodynamics comes from the Greek *thermos* meaning heat and *dynamics* meaning power. Thermodynamics originated through the production of the first vacuum pumps and then steam engines in the 1600s. Thermodynamics has a number of central concepts and

relations [Frenkel & Smit(1996)]. One basic concept is the idea of a system and its surroundings. A system is defined by its particles and their average motions. The system interacts with its surrounding in some defined way. Thermodynamics is the study of how the macroscopic properties such as temperature, pressure and volume of a system and its surroundings relate to each other.

Another basic concept of thermodynamics is energy. Energy can be transferred between systems and particles as heat or work and is attributed as kinetic energy (heat) and potential energy (potential work). If a system has no net transfer of energy between itself and its surroundings and has no change in its macroscopic observables with time it is said to be in equilibrium. When in a state of equilibrium a system's state properties can be described by equations of state such as the classical ideal gas law. In other words we can use the average system properties at equilibrium to calculate other properties. We are now able to think of a model system and how we can study it. In a molecular simulation we make a model of a system in which we are interested and accumulate useful data about the properties of its state, which we can then use to calculate other state properties.

2.1.3 The Boltzmann factor and Statistical Mechanics

A system will always tend towards configurations or microstates with the most ways of distributing the energy amongst its particles. This means that isolated systems tend to occupy lower energy states most of the time, as in these low energy states there are more ways to distribute the energy (for a more in depth explanation see Atkins (1994)).

The Boltzmann factor is a weighting term which can be used to find the probability of individual energy states of a system in thermodynamic equilibrium:

$$P \propto exp\{-E/k_B T\} \tag{2.1}$$

where E is the energy of the system, k_B is the Boltzmann constant and T is the temperature. The Boltzmann factor is not normalised but the ratio of the probabilities of two states is given by the ratio of their Boltzmann factors. The Boltzmann factor can be normalised by dividing by the sum of all the Boltzmann factors of all possible energy states of a system. This is called the partition function. The partition function is very useful as many of the thermodynamic variables of the system can be expressed in its terms. The exact nature of the partition function

of a system relies on how the system is defined, and there are a number of well known system definitions termed ensembles. A statistical mechanical ensemble is a collection of replicas of a system which are effectively different configurations of this system with particular defined constraints.

The Boltzmann constant is generally determined experimentally and is in units of energy per temperature (e.g. JK^{-1}). Thus, the Boltzmann constant relates the energy and temperature of particles in a system. The most basic result associated with the Boltzmann constant is that of the ideal gas equation:

$$PV = N_m k_B T \tag{2.2}$$

where N_m is the number of gas molecules present in the subject system. The ideal gas equation shows us that $k_B T$ is the pressure applied times unit volume for each molecule of an ideal gas.

Statistical mechanics links the microscopic properties of atoms and molecules to the macroscopic properties of thermodynamics. In molecular simulation statistical mechanics can be used to obtain bulk properties of model systems. The main tools of Statistical mechanics are the Boltzmann factor and the partition function.

2.1.4 The Ensembles of Statistical Mechanics

Another important idea in statistical mechanics is the ensemble average. When attempting to study a model system, which is part of a particular ensemble, properties of the whole ensemble (all microstates) are considered, i.e. ensemble averages. This is necessary as in experiment it is not possible to freeze time and evaluate the properties of a system. Instead averages over large systems and amounts of time are taken. We are able to find ensemble averages through sampling and averaging over large numbers of states of a system ensemble. The average does not necessarily have to be over the whole ensemble which is fortunate as systems can have an extremely large number of microstates.

In order that the assertion in the paragraph above may be achieved, statistical mechanics makes some assumptions:-

1. The equal *a priori* probability postulate: this postulate assumes that each microstate has an equal possibility of being occupied and is the fundamental postulate of statistical

mechanics. Justification for the assumption of the equal *a priori* probability postulate can be found with Liouville's theorem [Frenkel & Smit(1996)].

2. The ergodic hypothesis: consider the path of a system across the energy surface made up by plotting all microstates of the systems ensemble. Owing to the equal *a priori* probability postulate the system will, in the long run, visit each area of this energy surface equally and hence each microstate is equally probable over a long period of time. This means that a long time average and an average over the whole statistical ensemble are the same. Equally for systems evolving in a non-time linked fashion, sampled microstates will inevitably be in a localised area, but the assumption of ergodicity allows a large average to equate to the ensemble average. Within this assumption is the idea that ensemble averages are independent of the initial state or configuration of the system.

Thus we can use MD or MC simulations to find properties of the whole ensemble of states of a system by sampling for a long time or sampling a large number of configurations. In practice many systems are not ergodic and it is important to consider this before assuming the ergodicity of an exotic model.

It is important to define a system ensemble when applying statistical mechanics. When creating a model, one of a number of well known system ensembles can be used with differing properties. Each ensemble has a characteristic state function C, a thermodynamic quantity such as the free energy, from which we can directly find the partition function (Z) through,

$$C = K_B T \ln Z \tag{2.3}$$

Next the basic ideas behind some of the well known ensembles will be explained.

Microcanonical ensemble

The microcanonical ensemble refers to a system isolated from the outside environment. As the system is isolated it is unable to exchange energy with the environment and has a fixed total energy. This means that the system can only access those microstates which have the same energy as the system. The entropy of a microcanonical system can only increase and does so

until the equilibrium point is reached (due to the second law of thermodynamics). While in equilibrium the entropy of the microcanonical ensemble can easily be calculated using Boltzmann's famous equation:

$$S = k_B \ln \Omega(E) \tag{2.4}$$

where S is the entropy and $\Omega(E)$ is the number of microstates with the total energy of the system E. The microcanonical ensemble is often referred to as an NVE ensemble as it has fixed numbers of atoms, volume and energy. The number of microstates available to the system in the case of a microcanonical ensemble is found directly from Boltzmann's equation for entropy:

$$\Omega(N, V, E) = exp\{S/k_B\}. \tag{2.5}$$

$\Omega(N, V, E)$ is the microcanonical partition function as in this case the system can only have one energy. As can be seen from eq. 2.5, for the microcanonical ensemble S is the characteristic state function.

Canonical ensemble

The canonical ensemble is an ensemble of states of a system which is able to exchange heat with a heat reservoir. The connected heat reservoir must be very large such that through exchange with the main system both are kept at constant temperature. The main system and the heat reservoir together are isolated from the outside environment and assumed to be in thermal equilibrium. The canonical ensemble is also called an NVT ensemble as it has fixed numbers of atoms, volume and temperature.

The partition function of the canonical ensemble (Z) is,

$$Z = \sum_i exp\{-\beta E_i\}, \tag{2.6}$$

where $\beta = 1/k_B T$. An extra term, g_i is sometimes included to account for the degeneracy of microstates i, i.e. if only macrostates are counted, the number of microstates within each macrostate must be included E_i:

$$Z = \sum_i g_i exp\{-\beta E_i\}. \tag{2.7}$$

The canonical ensemble partition function can be derived in a number of different ways; most rely on the notion that the probability of the system being in a particular state is proportional to the number of states of the heat reservoir. The characteristic state function of the canonical ensemble is the Helmholtz free energy (A) as,

$$Z(N, V, T) = exp\{-\beta A\} \tag{2.8}$$

holds.

In classical mechanics the partition function is a continuous distribution rather than discrete as described above. The number of system states are uncountable as the position and velocity of particles are continuous variables. In this case very similar microstates are grouped together as one and the partition function becomes an integral,

$$Z = \frac{1}{N! h^{3N}} \int_{-\infty}^{\infty} exp\{-\beta H(p_N, q_N)\} dp_N dq_N. \tag{2.9}$$

Here h is Plank's constant, $N!$ is a term which is included to ensure the number of microstates is not over counted, $H(p_N, q_N)$ is the classical Hamiltonian, p is the momentum of particle N and q is the position of particle N. $H(p_N, q_N)$ is made up of the kinetic energy (K) due to the momenta p of the system particles and the potential energy (U) due to the particle coordinates q. In practice, the integration over the momenta can be carried out analytically as K is a quadratic function of the momenta. Thus, the difficulty lies in evaluating functions of the particle positions (p_N). Only in a few simple cases can the potential be evaluated analytically and it is generally found through numerical methods which will be discussed later in this chapter.

Isothermal-isobaric ensemble

The isothermal-isobaric ensemble is similar to the canonical ensemble, except the volume is allowed to change. The volume is controlled to give a specified average pressure. This ensemble can be called the NPT ensemble as pressure is also kept constant. This ensemble is important

as chemical reactions often require constant pressure and simulations often require the volume to be equilibrated. The partition function of the isothermal-isobaric ensemble is,

$$\Delta(N,P,T) = \int \sum_i \exp -\beta(E_i + PV)dV. \tag{2.10}$$

The characteristic state function for the NTP ensemble is the Gibbs free energy (G) as,

$$\Delta(N,P,T) = exp\{-\beta G\}, \tag{2.11}$$

is correct. G is similar to A with an extra term to account for the pressure of the system, $G = U - TS + PV = A + PV$.

The classical, continuous, NPT partition function is derived from the canonical partition function through a various routes, for one example see Frenkel and Smit (1996). The extra part of the equation is due to the work done by the system in going from V_0 to V:

$$\Delta(N,P,T) = \frac{1}{V_0 N! h^{3N}} \int_0^\infty dV \int_{-\infty}^{\infty} e^{-\beta(H(p_N,q_N)+PV)} dp_N dq_N \tag{2.12}$$

where V_0 is the initial volume.

2.1.5 Statistical mechanics and computer simulation

Quantum calculations are intrinsically more exact than the classical equivalent. Without going into details, quantum calculations on a complex system is slow compared to general classical methods. For the purposes of computer simulation of large biologically or chemically relevant systems, a classical mechanical method for evaluating molecular systems has been developed called molecular mechanics.

Molecular mechanics is a formalism where each atom of a system constitutes one particle with an assigned van der Waals (vdW) radius and constant point charge. Atomic point charges are found through quantum mechanics (QM) or experiment. Bond, angle, dihedral and non-bonded forces between atoms are assigned simple mathematical relations to give the behaviour needed for a particular model, this is often called the potential. A set of parameters and the potential needed for molecular mechanics of a particular model is called a force field.

2.1.6 Classical force fields

Classical force fields for bio-molecular simulation have taken up a fairly consistent form over the last few decades. The total energy of the potential is split into:

$$E_{total} = \sum E_{bonded} + \sum E_{non-bonded}, \tag{2.13}$$

$$E_{bonded} = \sum E_{bondlength} + \sum E_{angle} + \sum E_{torsion}, \tag{2.14}$$

$$E_{non-bonded} = \sum E_{electrostatics} + \sum E_{vdW}, \tag{2.15}$$

with bonded referring to covalent bonds [Leach(1996)].

Covalent bonds are most often modelled with harmonic potentials of the form,

$$E_{bond} = \sum_{bonds} k_r(r - r_0)^2, \tag{2.16}$$

$$E_{angle} = \sum_{angles} k_\theta(\theta - \theta_0)^2, \tag{2.17}$$

where k_r and k_θ are the force constants controlling the size of oscillations, r_0 and θ_0 are the reference positions of the bonded particles and r and θ its displacement. The potential of equation 2.16 does not exactly describe the potential of a real covalent bond. However, under the standard thermodynamic conditions for which classical force fields are designed to be used, these harmonic potentials describe the real bond potential well. It is also possible to use the Morse potential to describe covalent bonds. The Morse potential is able to describe real bond potentials more accurately but is rarely used due to the increased computational expense involved.

The torsions of bonds are more complex than bond length or angles as they may have multiple minima and some may have restrictions (e.g. planar rings). Thus, in general torsions are modelled by simple periodic functions with additional harmonics for improper torsions and energies between atoms bonded together through one other atom i.e. coupling angles and dihedrals (1-3 energy or Urey-Bradley energy),

$$E_{torsion} = \sum_{torsions} k_\phi[1 + \cos(n\phi - \delta)] \tag{2.18}$$

$$+ \sum_{impropers} k_\omega(\omega - \omega_0)^2 \tag{2.19}$$

$$+ \sum_{Urey-Bradley} k_u(u - u_0)^2. \tag{2.20}$$

In equation 2.18 k_ϕ is the torsional force constant and dictates the amplitude of the torsional potential, n controls the periodicity of the potential (i.e. how many minima) and δ controls the displacement along the axis of angle ϕ. One notable exception to this rule is the AMBER force field which uses additional periodic functions for improper torsions rather than harmonics. Torsional parameters are generally selected to correct energies and conformations of torsions due to the bond, angle and non-covalent potentials in simple molecules with respect to experiment or quantum calculations.

The non-covalent or non-bonded potential for both electrostatics and vdW is calculated as a pair-wise sum over all non-bonded atoms separated by at least three bonds. The vdW potential between a pair of atoms has a repulsive interaction which increases sharply as the interatomic distance r gets very small, and an attractive interaction which decreases relatively slowly as r is increased (figure 2.1). The vdW force field term is called the Lennard-Jones (LJ) potential and is defined as,

$$E_{vdW} = \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} 4\epsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \tag{2.21}$$

where i and j are each pair of atoms, ϵ is the depth of the potential well and σ is the interatomic distance between the particles when their interaction energy is zero as seen in figure 2.1. The term to the power 12 describes the repulsive interaction (i.e. the hardness of the vdW radii) and the power 6 term describes the attraction. As vdW interactions are relatively short range, the LJ potential is generally truncated (used with a cut off). A cut off will only apply the LJ potential to atoms pairs which are within a certain distance.

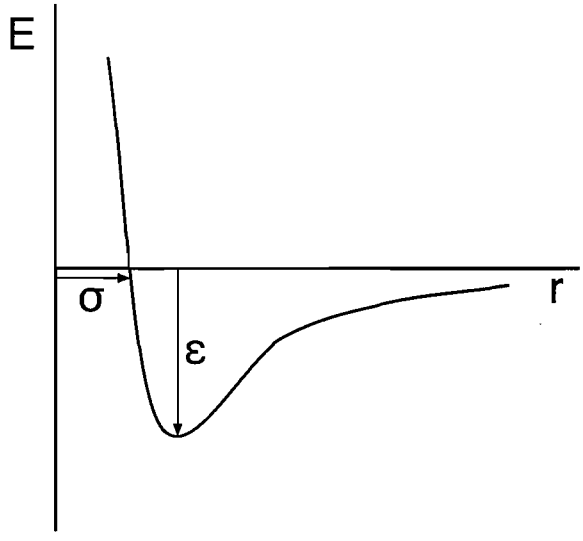


Figure 2.1: Plot of the Lennard-Jones potential with energy (E) on the y-axis and interatomic distance (r) on the x-axis.

The electrostatic potential of a pair of atoms is modeled using the Coulomb potential based on Coulomb's law. The Coulomb potential is proportional to the size of the electric charge on each atom of a pair and inversely proportional to r:

$$E_{elec} = \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \tag{2.22}$$

Here q is the charge on an atom and ϵ_0 is the relative permittivity of free space. The electrostatics of a system can be difficult to model, due to their long-ranged nature.

In the case of bio-systems, long range electrostatics can be important and it is not generally possible to calculate long range electrostatics using the Coulomb sum even if the modelled system is large enough. This is because of the large number of atom pairs, and a cut off is generally used. The size of the electrostatic cut off can make large differences to the calculated properties and care should be taken in their use. In general, the molecular interactions will be scaled down gradually at the cut off boundary in an attempt to reduce the effects of a sudden discontinuity of forces [Frenkel & Smit(1996)]. Also, a faster alternative has been developed called the particle mesh Ewald sum (PME) which is able to closely approximate the long range Coulomb sum [Essmann *et al.*(1995)Essmann, Perera, Berkowitz, Darden, Lee & Pedersen].

The most widely used atomistic force fields for biomolecules are AMBER [Pearlman *et al.*(1995)Pearlman, Case, Caldwell, Ross, Cheatham, Debolt, Ferguson, Seibel & Kollman], CHARMM [Brooks

et al.(1983)Brooks, Bruccoleri, Olafson, States, Swaminathan & Karplus], GROMACS [Spoel *et al.*(2005)Spoel, Lindahl, Hess, Groenhof, Mark & Berendsen] and OPLS [Jorgensen & Tirado-Rives(1988)]. AMBER, CHARMM and OPLS have potential functions exactly of the form described above with options of differing torsional 1-4 terms. GROMACS has the options of different bonding interactions, including the Morse potential and a harmonic approximation. Also, torsional 1-4 interactions can be dealt with by a special LJ interaction or a power series of cos ϕ can be used with out a 1-4 interaction.

The AMBER, CHARMM and GROMACS force fields have been parameterised specifically for MD of biosystems while OPLS has been developed for use with MC, but all can and are used with either style of simulation. The parameters for all of these force fields have been found through a mixture of experimental observations of small organic molecules and high level quantum mechanical (QM) calculations and are thus generally termed empirical. The parameters of almost all force fields are developed self-consistently, and it is never a good idea to mix and match parameters from different parametrisations.

The main deficiencies of these force fields stems from the fact that many of the interactions of molecules (including vdW and electrostatics) are environment dependent [Ponder & Case(2003)]. A new generation of polarizable force fields have recently been developed with polarisable versions of AMBER and CHARMM available. These polarisable force fields are able to account for the electronic polarisation of the environment around each atom. AMBER polarisable force fields can either use point charges as traditional non-polarisable versions, or can add extra off-centre point charges to better model the angular dependence of hydrogen bonds [Cieplak *et al.*(2001)Cieplak, Caldwell & Kollman].

Water models are also extremely important for biosimulations as most biological interactions take place in solutions of water. Solvation in simulations takes two forms, explicit, where water molecules are explicitly present and implicit, where solvent is represented by a continuum.

Implicit solvation models also take two general forms; accessible surface area models such as the generalised born model (GB) [Jeancharles *et al.*(1991)Jeancharles, Nicholls, Sharp, Honig, Tempczyk, Hendrickson & Still] and continuum electrostatics models such as the Poisson-Boltzmann model (PB) [Lu *et al.*(2005)Lu, Zhang & McCammon]. These implicit models improve the speed of molecular simulations dramatically and may be able to reduce errors

found in explicit solvent simulations due to incomplete sampling of solvent conformations. This can have a direct effect on the focus of this study, free energy calculations in protein-ligand systems. A recent study has highlighted the possible advantages of implicit solvation in binding free energy calculations [Michel *et al.*(2006)Michel, Verdonk & Essex].

However, implicit solvent models do have major limitations. Both types of implicit methodology are parameterised especially for the particular type of calculation for example, the parameterisation carried out by Michel *et al.* for free energy calculations with MC [Michel *et al.*(2004)Michel, Taylor & Essex]. Also, both models are unable to properly deal with cases where water molecules are directly involved in a process.

Explicit solvation models can be classified through the number of interaction sites used to define a water molecules and the flexibility of the molecule. The use of flexible water models can add significantly to the time taken for simulation. Thus, the most widely used water models are rigid and use only the non-bonded Coulomb and Lennard-Jones potentials introduced above. Popular 3-site models include TIP3P [Jorgensen *et al.*(1983)Jorgensen, Chandrasekhar, Madura, Impey & Klein] and SPC [Berendsen *et al.*(1981)Berendsen, Postma, van Gunsteren & Hermans] and generally have no Lennard-Jones parameter for the two hydrogen atoms. Other models include up to 6 interaction sites with extra dummy sites used to model the distribution of the electrostatic field around the water molecule [Jorgensen & Tirado-Rives(2005)].

2.1.7 Simulation sampling methodologies

The possibilities of simulation system set-up have now been set out and it only remains to discuss the methods used for sampling these systems. In addition to the MD and MC methods mentioned above, systems can be sampled through minimisation algorithms. Minimisation algorithms attempt to find the minimum of the potential by adjusting a systems degrees of freedom and exploring its energy surface. Minimisation is useful in many ways but is often used only to remove any bad contacts from a starting structure.

Molecular Dynamics

MD uses approximate numerical integration to calculate the positions and velocities of each particle of a system after a small time step (δt). By calculating the state of the system through many successive steps a trajectory through time is produced.

The force acting on an atom can be calculated through the derivative of the energy with respect to the change in the atom's position,

$$\vec{F}_i = \frac{-dU}{dr_i} = -\nabla U. \tag{2.23}$$

where $\vec{F}_i$ is the force on atom i and r_i the position of particle i, and U is the potential energy. By solving Newton's equation, $\vec{F}_i = m_i a_i$, for two particles, it is possible to calculate the trajectory of particle i as a function of time. However, for more than two particles numerical approaches must be sought.

A number of useful MD integration schemes exist, based on a Taylor series expansion approximation. The different integration schemes have differing attributes. The majority of current MD packages use the velocity Verlet or Verlet leapfrog integration schemes due to their good energy conservation at relatively large time steps [Swope *et al.*(1982)Swope, Andersen, Berens & Wilson], [Cramer(2002)]. The velocity Verlet algorithm is generally considered the most complete as it is alone in being able to accurately provide particle positions and velocities at time t. There are 4 steps to the velocity Verlet algorithm, including a step to find the velocity at the mid point of each time step, $t + \delta t/2$ [Swope *et al.*(1982)Swope, Andersen, Berens & Wilson], as follows:

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2, \tag{2.24}$$

$$v(t+\frac{\delta t}{2}) = v(t) + \frac{1}{2}a(t)\delta t, \tag{2.25}$$

$$a(t+\delta t) = -\frac{1}{m}\nabla U\left(r(t+\delta t)\right), \tag{2.26}$$

$$v(t+\delta t) = v(t+\frac{\delta t}{2}) + \frac{1}{2}a(t+\delta t)\delta t. \tag{2.27}$$

All of the MD integrators presently in use are limited in the size of possible time steps. If the time steps are too large the integrator will not be stable, energy will not be conserved and the simulation will uncontrollably increase in energy. Time steps for atomistic simulation are in the region of two femtoseconds (fs). This means that MD simulations are limited in possible timescales. Even with the powerful computers now available MD simulations do not easily go beyond the nano second (ns) timescale. Many biological processes are of a time scale which MD cannot reach with out a large coordinated effort, such as the folding@home effort to understand protein folding [Shirts & Pande(2000)]. Thus, in practice, MD simulations may get stuck in local energy minima, which are very difficult to get out of in the time available. This can hinder the full sampling of phase space and the amassing of accurate ensemble averages for statistical mechanics calculations. It is important to note that MD is time linked, time reversible and deterministic. It is therefore able to model real physical processes such as reaction pathways, unlike MC.

Monte Carlo

MC refers to any method which uses a sequence of random numbers to sample; this means MC is stochastic (i.e. nondeterministic). MC methods can be used for numerical integration problems. They are found to be more efficient than other techniques (quadrature) where the integrand is very large and has a number of dimensions. The most simple representation of an MC integration method for a one dimensional integral I is as follows:

$$I = \int_a^b f(x)dx, \tag{2.28}$$

$$I = (b-a)\langle f(x)\rangle . \tag{2.29}$$

where $\langle f(x)\rangle$ is an average of $f(x)$ in the interval a to b. Using MC this integral is found by evaluating $f(x)$ many times (N_{mc}) with random values of x. As N_{mc} approaches infinity the average would provide I [Frenkel & Smit(1996)].

MC techniques could be applied to evaluate an average of an ensemble of statistical mechanics by randomly sampling states (sampling from all possible arrangements of a system, equivalent to (b - a) above) and weighting them according to the Boltzmann distribution.

However, this would be exceedingly inefficient as most states found this way would not be weighted heavily within the ensemble of interest. The system would become a random jumble of particles which would never occur in reality. Thus, it is important that when using MC with molecular mechanics, we only sample states that have a reasonable probability of occurring according to the Boltzmann distribution.

This is achieved through a special type of importance sampling [Frenkel & Smit(1996)]. With importance sampling random points or configurations are taken from a specific volume of all possibilities, a nonuniform distribution where configurations are more likely to occur according to the Boltzmann distribution. The closer this distribution, which limits our sampling, is to the Boltzmann distribution the more efficient the MC sampling will be. Of course, the optimum choice of limiting distribution is the Boltzmann distribution itself, which is unobtainable due to the limitless nature of the partition function. This problem of defining the limiting distribution is solved through considering the relative probability of sampled configurations as opposed to the absolute probability through the partition function. Using the Boltzmann factor alone the configurations can be sampled with a relative probability proportional to the Boltzmann distribution.

An "MC simulation" in chemistry generally refers to the Metropolis MC method and this shall be the case from now on in this work [Metropolis *et al.*(1953)Metropolis, Rosenbluth, M.N., A.H. & Teller]. Metropolis MC employs a particular Markov process to produce system configurations from a limiting distribution which is the Boltzmann distribution. For an in depth explanation refer to appendix A.

MC in its most simple form perturbs the present configuration O, within preset parameters, to produce a new trial configuration (N) of the system. N is accepted or rejected depending on the relative probability of each state occurring. The criterion for accepting a new configuration is the Metropolis criterion, equation 2.30, where P_{accN} is the probability of accepting a new configuration, $\pi(O \rightarrow N)$ is the probability of the move from O to N being accepted and $\pi(O \rightarrow N)$ visa versa:

$$P_{accN} = min(1, \pi(O \rightarrow N)/\pi(N \rightarrow O)), \tag{2.30}$$

When accepted N becomes O and the properties of N are added to the average being accumulated. If rejected N is discarded, the properties of O are added to the average again and a new trial configuration is generated. In order that the distribution sampled by this process is maintained the probability of moving from any configuration leaving state O must be the same as the probability of any state arriving in state O. An easy way of ensuring this condition is met is to make the probability of leaving O for N the same as the probability of leaving N for O ($\pi(O \rightarrow N) = \pi(O \rightarrow N)$). This condition is termed detailed balance.

$\pi(O \rightarrow N)$ can be split into its component parts,

$$\pi(O \rightarrow N) = \alpha(O \rightarrow N) \times P_{accN}(O \rightarrow N) \tag{2.31}$$

where α is the probability of finding a trial configuration N and acc is the probability of accepting N. By making the perturbation which produces N random the matrix of all possible α is symmetric ($\alpha(O \rightarrow N) = \alpha(N \rightarrow O)$). Thus we can write equation 2.30 in terms of acc only and the distribution sampled by accepted configurations will depend only on the relative probabilities of acc. By using the Boltzmann factors of the O and N configurations the sampled distribution is the Boltzmann distribution:

$$P_{accN} = min\left(1, \frac{exp\{-\beta U_O\}}{exp\{-\beta U_N\}}\right), \tag{2.32}$$

$$= min(1, exp\{-\beta[U_O - U_N]\}), \tag{2.33}$$

where U_O and U_N are the potentials of the old and new configurations respectively. Using this process of sampling, over many cycles, the distribution of accepted configurations has been proven to converge to the Boltzmann distribution regardless of the initial configuration (appendix A). It is possible to use alternative non-symmetric matrices of α and still maintain detailed balance; an outline of such methods is given in appendix B.

Despite the concepts behind this algorthm being difficult, it is quite simple in practice, it is described in figure 2.2.

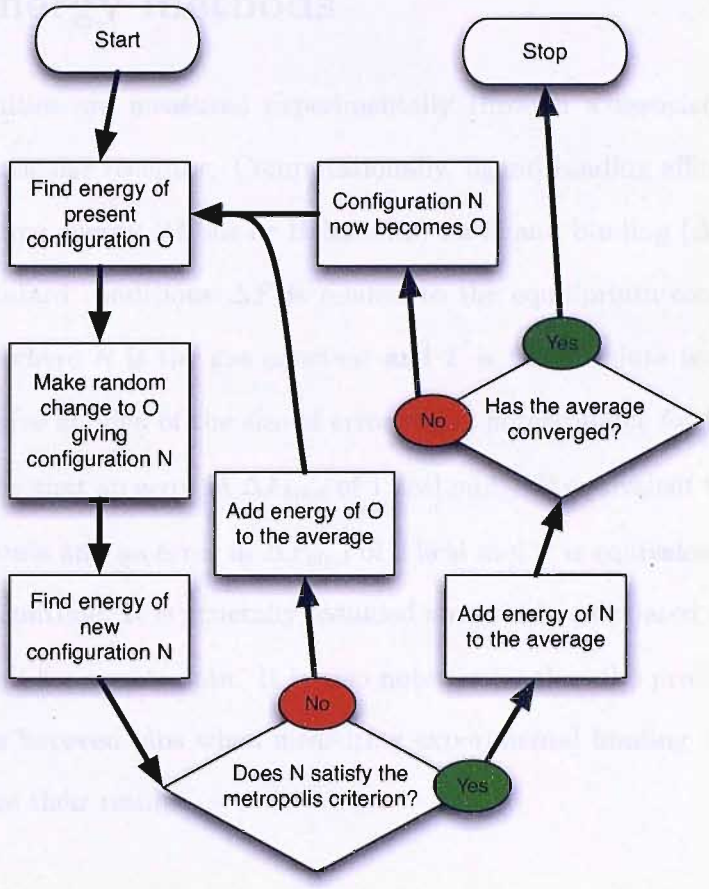


Figure 2.2: Flow diagram of the Metropolis MC algorithm.

In practice N is accepted if the ratio in equation 2.33 is greater than a random number found from 0 to 1.

In the MC algorithm described here there is the possibility of wasting large amounts of time through rejecting many trial configurations. There are some interesting methods for ensuring a high rate of acceptance of trial configurations in difficult cases [Rosenbluth & Rosenbluth(1955)] but they will not be discussed here. The acceptance rate is maintained at a high level through adjustment of the parameters controlling the random perturbation of the present configuration.

In general, the perturbation of O is localised to a random single entity such as a molecule or protein residue. This being the case, when evaluating the potential of N, it is not necessary to calculate the full potential but only the bits of the potential which have been affected by the perturbation. Thus, unlike MD, MC moves only small portions of the configuration in one step, and the time taken to find the new energy is smaller.

2.2 Free energy methods

Ligand binding affinities are measured experimentally through a association constant (K_a) when bound to a particular receptor. Computationally, ligand binding affinities are compared using the change of free energy (Gibbs or Helmholtz) on ligand binding (ΔF_{bind}). At equilibrium and under standard conditions ΔF is related to the equilibrium constant via equation 2.34 [Atkins(1994)], where R is the gas constant and T is the absolute temperature. This is important as it can give an idea of the size of error which is reasonable for free energy calculations. It is noteworthy that an error in ΔF_{bind} of 1 kcal.mol^{-1} is equivalent to an error in K_a of two orders of magnitude and an error in ΔF_{bind} of 2 kcal mol^{-1} is equivalent to an error in K_a of three orders of magnitude. It is generally assumed an error in calculated ΔF_{bind} of around 1 kcal mol^{-1} is the limit for usable data. It is also noteworthy that the problem of maintaining consistent conditions between labs when measuring experimental binding constants can make it difficult to compare their results.

$$\Delta F = -RTlnK_a \tag{2.34}$$

2.2.1 Difficulty of calculating free energy

Through the theory discussed above the route to calculation of system free energies seems clear. In the case of the canonical ensemble the free energy is derived through,

$$exp\{-\beta A\} = \frac{1}{N!h^{3N}} \int_{-\infty}^{\infty} exp\{-\beta H(p_N, q_N)\} dp_N dq_N. \tag{2.35}$$

By calculating the free energy of a ligand bound to a receptor and free in solution the free energy difference could be found. Unfortunately, in practice with large bio-simulations, sufficient sampling to provide an accurate integral with equation 2.35 is beyond the reach of present sampling methods and computational resources. Quantities such as A which depend directly on a system's phase space volume and not on functions of the phase space coordinates are called thermal quantities.

It is possible to attempt to evaluate the free energy as an ensemble average. However, the simulation methods discussed above sample the low energy regions of phase space in the

overwhelming majority. Hence, there are many high energy states which contribute to the free energy which are not encountered with MD or MC. These methods are unable to sample adequately the high energy regions of phase space and converged averages are not possible. As free energies can be either G or A depending on the ensembles under study, in subsequent discussions F will represent general free energy either G or A.

These practical realities enforce the need for special methods for calculating free energy differences. Experimental measurements of free energies are generally carried out in reference to a substance of known free energy, as thermal quantities are also difficult to measure experimentally. For the purposes of drug design the most basic information needed is the relative binding free energy of a set of ligands with one protein receptor. Thus, the rest of the chapter will be concerned with special methods which can be used to calculate free energy differences between two systems, defined by,

$$\Delta F = F_B - F_A \tag{2.36}$$

where F_A is the free energy of system A, F_B is the free energy of system B and ΔF is the free energy difference.

It is important that a free energy method remains within the rules of statistical mechanics to ensure that calculations are accurate. However, even if a method is rigorous with regards to statistical mechanics (from now on referred to as rigorous methods), there are still some assumptions involved particularly related to the choice of potential. On the scale of accurate free energy estimates, statistical mechanically approximate methods (from now on referred to as approximate methods) are in general thought to be less accurate than rigorous ones [Pearlman(2005)] [Kuhn *et al.*(2004)Kuhn, Gerber & Stahl]. This investigation is interested in fast free energy estimates, and thus it may be useful to investigate approximate methods as well as those which are thought to be rigorous.

2.2.2 Approximate binding free energy methods

One major line of investigation is the calculation of the binding free energy difference (ΔF_{bind}) based on partitioning the free energy into separate, individually calculated, components. This methodology assumes that each of the components of ΔF can be calculated separately and

summed to find the total. An example of this is,

$$\Delta F_{bind} = \Delta F_{solv} + \Delta F_{conf} + \Delta F_{int} + \Delta F_{mot}, \tag{2.37}$$

where ΔF_{solv} is the free energy change on ligand-receptor binding due to solvation effects. ΔF_{conf} is the free energy change on binding due to changes in conformational structure of ligand and receptor, ΔF_{int} is the enthalpy change due to interactions such as electrostatic and van der Waals forces between the ligand and receptor, and ΔF_{mot} is the change in entropy due to change in freedom of motion of the atoms of the ligand and receptor. ΔF_{mot} can be split into changes in internal rotations, translational motions and vibrational free energy due to binding [Cramer(2002)].

If it is possible to calculate the total absolute free energy of a system in this way, then ΔF_{bind} can be found by simply calculating the free energy of the receptor-ligand complex, separate receptor and separate ligand and taking the difference as in,

$$\Delta F_{bind} = \Delta F_{complex} - (\Delta F_{receptor} + \Delta F_{ligand}) . \tag{2.38}$$

Calculated energy contributions often take the form of simulation ensemble averages of particular observables.

Assumptions are often made to allow or simplify the calculation of these energy partitioned methods. The basis for these assumptions depends on the types of system under study and the computational time allowed. For example new vibrational modes created on binding are often ignored or approximated in an *ad hoc* manner, as this can be a difficult goal to achieve efficiently.

Linear Interaction Energy (LIE)

The LIE method is based on the partitioning of the total binding energy into polar and non-polar contributions [Aqvist *et al.*(1994)Aqvist, Medina & Samuelsson]. The polar contribution is derived from MD simulation average energies where the solvent is assumed to have a linear response to electrostatic changes due to the solute (i.e. linear response behaviour). The non-polar contributions are found using experimental binding data to create an empirically calibrated function which includes averaged simulation van der Waals energies.

This methodology is based on studies suggesting that solvation energies of non-polar compounds are linearly related to molecular size measures such as surface area or volume [Blokzijl & Engberts(1993)]. Average computer simulation solute-solvent van der Waals energies have also been shown to relate to size measures in a similar way [Aqvist *et al.*(1994)Aqvist, Medina & Samuelsson]. The use of an empirically calibrated parameter is expected to take into account all other contributions to non-polar free energy, such as entropy effects etc.

An LIE calculation uses two MD simulations as depicted in figure 2.3

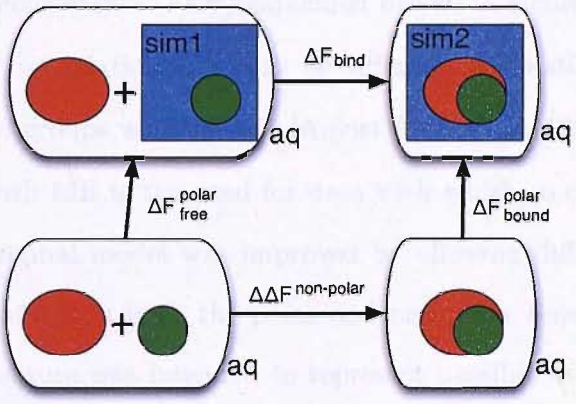


Figure 2.3: Thermodynamic cycle used in the LIE method. The binding free energy ΔF_{bind} is given by the sum of a non-polar contribution $\Delta\Delta G_{nonpolar}$ and a polar contribution $\Delta\Delta F_{polar} = \Delta F_{polarbound} - \Delta F_{polarfree}$. The actual simulation systems are denoted by the blue squares labelled sim1 and sim2.

From these simulations the average electrostatic and van der Waals energies of the ligand with its surroundings are taken and differences calculated. The calculation of polar energies is carried out via equation 2.39 and the non-polar via equation 2.40.

$$\Delta F_{polar} = \beta_{LIE} \left(\langle \Delta E_{polarbound} \rangle - \langle \Delta E_{polarfree} \rangle \right) \tag{2.39}$$

$$\Delta F_{nonpolar} = \alpha_{LIE} \left(\langle \Delta E_{nonpolarbound} \rangle - \langle \Delta E_{nonpolarfree} \rangle \right) \tag{2.40}$$

In equations ?? $\beta_{LIE} = \frac{1}{2}$ according to the linear response assumption, $\langle \Delta E_{polarbound} \rangle$ is the intermolecular electrostatic interaction energy, averaged over an MD trajectory for the bound ligand, $\langle \Delta E_{elecfree} \rangle$ is the same for the unbound ligand. α_{LIE} was originally set at 0.16 using a set of endothiapepsin inhibitors to calibrate this figure, $\langle \Delta V_{vdWbound} \rangle$ is the van

der Waals interaction energy again averaged over an MD trajectory for the bound ligand and $\langle \Delta V_{vdW\,free} \rangle$ is the same for the unbound ligand. Equations 2.39 and 2.40 are combined to calculate ΔF_{bind}. This original model, calibrated using the endothiapepsin inhibitors, has been used reasonably successfully on several systems including HIV protease, trypsin and glucose binding protein [Hansson & Aqvist(1995), Aqvist & Mowbray(1995), Aqvist & Hansson(1996)]. The root mean squared (RMS) error for 18 inhibitors from these studies was 1.57 kcal mol^{-1}, although there were significant errors, particularly for some HIV protease inhibitors.

The validity of the electrostatic linear response approximation was investigated by comparison to Free Energy Perturbation (FEP)(explained below). Calculations were validated for mono-valent ionic solutes but relationships may be different, while still linear, for di-polar solutes especially if hydroxyl groups were present [Aqvist & Hansson(1996)].

The major problem with LIE is the need for data with which to carry out system specific parameterisation. The original model was improved by allowing different scaling factors for the bound and unbound states in both the polar and non-polar components of the LIE free energy equation. This measure was intended to represent possible differences in electrostatic response in the solvent and protein environments. Also a free energy constant term, Γ_{LIE}, was investigated as a non-zero difference between terms in corresponding linear expressions, giving a new free energy as,

$$\begin{aligned} \Delta F_{polar} = \alpha_{LIE,prot} \left(\langle \Delta E_{nonpolarbound} \rangle - \alpha_{LIE,wat} \langle \Delta E_{vdW\,free} \rangle \right) + \\ \beta_{LIE,prot} \left(\langle \Delta E_{polarbound} \rangle - \beta_{LIE,wat} \langle \Delta E_{elecfree} \rangle \right) + \Gamma_{LIE}. \end{aligned} \tag{2.41}$$

This new equation was tested and optimised with the same set of 18 inhibitors as the earlier equations 2.39 and 2.40. Best agreement with experimental data was found when the free energy constant term was disregarded and both α_{LIE} were identical and β_{LIE} values were almost identical. A system of generalised β_{LIE} terms was therefore derived.

$$\begin{aligned}
\beta_{LIE,FEP}(class) &= 0.50 \text{ for "charged class" } (Q \neq 0)\,, \\
&= 0.43 \text{ for class} = 0\,(Q = 0, \text{ No. OH} = 0)\,, \\
&= 0.37 \text{ for class} = 1\,(Q = 0, \text{ No. OH} = 1)\,, \\
&= 0.33 \text{ for class} = 2\,(Q = 0, \text{ No. OH} > 1)\,,
\end{aligned} \tag{2.42}$$

This model divides possible solutes into four classes as in equation 2.42 where Q is the charge and $No.OH$ is the number of hydroxylate groups present. The values of β were found by simulation of typical compounds from each of the classes [Aqvist & Hansson(1996)]. This model has shown an RMS error of 0.84 kcal mol^{-1} using the same set of 18 inhibitors [Marelius & Hansson(1998)]. Whether using the original model or the improved version this methodology has shown it can give results with errors in the order of 1 kcal mol^{-1} in test cases.

A recent study has tested a large number of possible descriptors and coefficients for use in prediction of binding affinities from simulation trajectories [Tominaga & Jorgensen(2004)]. This approach seems to yield limited advantages and descriptors used are similar to the ones used in the LIE methods above.

Molecular Mechanics/Poisson Boltzmann, Surface Area (MM/PBSA)

The MM/PBSA methodology splits free energy into four partitions as in equation 2.43 [Kollman *et al.*(2000)Kollman, Massova, Reyes, Kuhn, Huo, Chong, Lee, Lee, Duan, Wang, Donini, Cieplak, Srinivasan, Case & Cheatham]. Conformational space is sampled by MD simulations with explicit solvent to produce a trajectory of system configurations (snap shots). Each of the simulation snap shots is post-processed to remove solvent atoms. Solvent contributions are calculated through the use of an implicit continuum solvation model.

$$\Delta F = E_{MM} + G_{PB} + G_{SA} - TS_{MM} \tag{2.43}$$

In equation 2.43, E_{MM} is the average molecular mechanical energy of the receptor-ligand complex. E_{MM} can be decomposed into the different terms in the force field equation used (equation 2.15). G_{PB} is the polar solvation free energy which is found by solving the Poisson-Boltzmann

equation. It is possible to replace this solvation free energy term with one based on a GB continuum model [Gohlke & Case(2004)]. G_{SA} represents the non-polar solvation free-energy and is a simple scaled term based on the solvent accessible surface area (SASA):

$$G_{SA} = \gamma_{SA} SASA + b. \tag{2.44}$$

where γ_{SA} and b are 0.00542 kcal.mol^{-1}.Å^2 and 0.92 kcal.mol^{-1} respectively. TS_{MM} is the solute entropy which is estimated by quasi harmonic or normal-mode analyses [Swanson *et al.*(2004)Swanson, Henchman & Mccammon], [Donnini & Juffer(2004)]. It is worth noting that the free energy due to ionic strength effects can be added with a continuum approach. Also, the ΔG_{mot} and ΔG_{conf} terms from equation 2.37 are not properly accounted for within the MM/PBSA method.

When used to find binding free energies of ligand receptor systems there are two possible MM/PBSA forms; one involves analysing separate MD trajectories of receptor, ligand and complex (SEP) while the other uses one trajectory of the receptor-ligand complex which is processed to give data for all four terms in equation 2.43 (SING). Method SING is more efficient, depending on the length of MD trajectories used, although it ignores any free energy due to conformational change on binding (ΔG_{conf}). This deficiency will have varying effects depending on the system under study although promising results have been reported by some (e.g. [Huo *et al.*(2002)Huo, Wang, Cieplak, Kollman & Kuntz], [Kuhn & Kollman(2000)]. A possibly more important point is that the SING method displays a significantly lower level of error than the SEP method. This is due to the fact that the simulations used to find E_{MM} for the complex and its separated constituents are one and the same, meaning that sampling and force field errors cancel.

Three recent studies of the MM/PBSA methodology have introduced some novel methods of estimating energy change due to entropy and translational, rotational freedoms on binding, and also attempted to test the current methodologies on various systems [Swanson *et al.*(2004)Swanson, Henchman & Mccammon], [Donnini & Juffer(2004)], [Gohlke & Case(2004)]. Gohlke and Case utilised structural data of the H-Ras/C-Raf1 complex to investigate MM/PBSA methodology [Gohlke & Case(2004)]. The H-Ras/C-Raf1, protein-protein system has a "small to moderate" level of flexibility on binding and Ras interacts with Raf through the formation of an inter-protein β-sheet. Swanson *et al.* applied their methodologies to the fairly rigid FK506

binding protein (FKBP12), which is a small immunophilin, and its immunosuppressive ligand 4-hydroxy-2-butanone (BUT). Donnini and Juffer used 36 different protein-ligand complexes of various different types including MHC, TIM and SH2 and 3 protein domains. The following is a list of important points to arise from these studies:-

Comparison of PB and GB techniques: The effects of employing PB compared to three different GB implicit, continuum solvent models was assessed using MM/PBSA method SING [Gohlke & Case(2004)]. PB calculations gave over and under estimations of ΔG_{bind} depending on the atomic radii used, with smaller atomic radii giving lower energies. When compared to the PB model, results for the GB models all deviated from PB and over estimated ΔG_{bind} using the same atomic radii. However the same relative trends in scoring snapshots were seen in GB models compared to the PB model. In general the authors comment that the results reflect the way the GB models have been parameterised. Two of the models were parameterised for use with MD i.e. to give constant energy differences between different conformations of one molecule. For use in calculation of ΔG_{bind}, atom radii parameter sets are needed to also give consistent free energies across different molecular species.

Inaccuracy of SASA term: Deficiencies of the SASA non-polar solvation term due to over simplification were addressed, again by Gohlke and Case [Gohlke & Case(2004)]. Non-polar solvation is due to solvent cavity formation (dependent on volume and shape of excluded volume) and solute-solvent van der Waals interactions. Thus the simple SASA dependent term may not describe this completely. A new term was constructed, equation 2.45, where G_{cav} is the cavity solvation free energy which is determined by equation 2.46, and $H_{vdW,solute-solvent}$. $H_{vdW,solute-solvent}$ is the average solute-solvent van der Waals interaction energy from the simulation and is assumed to be approximately equal to the free energy due to solute-solvent van der Waals interactions ($G_{vdW,solute-solvent}$). In equation 2.46 G_{cav} is equal to a surface tension constant, γ_c ,and the molecular surface area, MSA.

$$G_{nonpolar} = G_{cav} + H_{vdW,solute-solvent} \tag{2.45}$$

$$G_{cav} = \gamma_c MSA \tag{2.46}$$

The old SASA dependent term was thought to overlook the van der Waals attraction between the unbound solute and solvent atoms, over estimating the contribution of non-polar solvation to the binding free energy [Gohlke & Case(2004)]. The free energy change due to non-polar solvation found using this new term was more favourable than the old SASA dependent term, which fits well with the view that the old term missed these van der Waals interactions. The $H_{vdW,solute-solvent}$ values taken from snapshots of Ras, Raf and their complex were shown to converge as the size of the water shell used reached 9 and 10 Å, which also backs up the view that this new term is superior. This method of calculating the non-polar solvation energy has previously been used with encouraging results [Vorobjev *et al.*(1998)Vorobjev, Almagro & Hermans].

Comparison of entropic calculations: All three of the papers discussed here used different methods to account for the solute entropic contributions or loss of translational, rotational and vibrational freedom. Gohlke and Case compared quasi-harmonic analyses to the normal mode analyses which would normally be used [Gohlke & Case(2004)]. As would be expected, $T\Delta S_{bind}$ was divided into loss of translational, rotational and vibrational entropy (S_{trans}, S_{rot}, S_{vib} respectively). Almost identical S_{rot} contributions were demonstrated between normal mode and quasi-harmonic analyses. However the S_{vib} contributions calculated with quasi-harmonic analysis were very different from those for normal mode analysis and lead to a very dis-favourable $T\Delta S_{bind}$. The authors highlight the fact that their trajectories do not show convergent, quasi-harmonic, S_{vib} through 10 ns of MD, and the fact that other studies have also failed. This finding is further supported by Swanson *et al.* [Swanson *et al.*(2004)Swanson, Henchman & Mccammon]. This is more pertinent as errors in $T\Delta S_{bind}$ are almost completely attributed to S_{vib}. Thus, quasi-harmonic analysis is seen to be a poor choice due to sampling issues.

Swanson *et al.* attempt to calculate the loss of translational and rotational freedom due to the ligand (which they refer to as "association free energy"). This was done via a conceptual link between the translational configuration integral and the volume of space occupied

by the ligand's centre of mass through the simulation. The volume was measured using the quasi-harmonic model and the covariance matrix to account for coupled motions. The resulting variance was then related to a spring force constant to obtain an average potential energy. A similar method was used for rotational freedom except quaternions were used to represent any motions. The authors then separated the 23.3 kcal mol^{-1} association energy they found into enthalpic and entropic components, assuming that all translational and orientational motions of the ligand can be described by a classical harmonic oscillator. Data for vibrational freedom/conformational flexibility were omitted from this study.

Another interesting method of conformational flexibility estimation has been utilised recently by Donini and Juffer [Donnini & Juffer(2004)]. This method estimates the difference in side chain conformational entropy on binding empirically from the change in side chain accessible surface area. It is difficult to assess how accurate the figures for side chain conformational entropy from this study were, as most calculated binding free energy values were an order of magnitude out from experimental values, with or without these entropy figures.

Inaccuracies due to conformational change on binding (relaxation energy): Both Swanson *et al.* and, Gohlke and Case investigated the use of MM/PBSA SEP, described above [Gohlke & Case(2004), Swanson *et al.*(2004)Swanson, Henchman & Mccammon]. The H-Ras/C-Raf1 system of Gohlke and Case is thought to have more conformational change between bound and unbound states than systems in previous studies. Thus there is little surprise that a comparison between SEP and SING gave differing results for E_{MM}, ΔG_{solv} and $T\Delta S$.

The free energy of the isolated H-Ras/C-Raf1 molecules would be expected to be less favourable while in the bound conformation. An easy rationale for this is that otherwise they would exist in this conformation when not bound. Hence, for this comparison the free energy of the unbound molecules would be expected to be more favourable, with a less favourable ΔG_{bind}, for results from method SEP. However, in this study the opposite was found with ΔG_{bind} being more favourable for method SEP than SING, being -19.3 and -3.1 kcal mol^{-1} respectively. The authors suggest this disparity may arise due to the use of ensemble structures from explicit solvent simulations being used in implicit solvent free energy calculations. In other words the implicit solvent models inadequately model the forces found in explicit solvent simulations.

Swanson *et al.* draw different conclusions from their similar analysis of a more rigid ligand-receptor system. They found that the SEP method gave energy plots which did not converge and were dominated by noise. Hence, they did not use these results to find the relaxation energy of the receptor. However, as the small ligands sampling and therefore convergence was far more complete it was possible to find the ligand relaxation energy (1.7 kcal mol^{-1}).

Statistical Convergence of Average Energies: It is important that the averaged values found using MM/PBSA are converged as otherwise they will be imprecise. Plots of gas-phase and solvation free energies from the 10 ns trajectories used by Gohlke and Case displayed small amounts of absolute drift even after 2 ns of equilibration [Gohlke & Case(2004), Donnini & Juffer(2004)]. As the authors comment, this points to incomplete sampling and the need for longer MD trajectories.

Gohlke and Case reported a standard error of 3.0 kcal mol^{-1} in gas-phase and solvation free energies using separate trajectories (method SEP). This error is too high for useful application in drug design, and thus method SING with an error in gas-phase and solvation free energies of 0.3 kcal mol^{-1} may be more useful. With independent trajectory snapshots, the standard error of the mean is inversely proportional to the square root of the number of snapshots. The authors find the snap shots to be independent as correlation time of fluctuations in the energy is about 1 ps. This ignores large molecular movements in the protein that may be over a much longer period. If snap shots can remain independent, this means that the more snapshots the smaller the error and this may offer a way of minimising the larger errors of method SEP.

Interestingly, despite the problems detailed above, when results taken only from the first and second half of the trajectories were compared with those from the whole trajectory, all were in good agreement in terms of total binding free energy, although errors were higher for the smaller sets of snapshots. The agreement is the result of deviating gas-phase, solvation and entropic energies compensating for each other. Thus, the total free energy is more reliable than the separate energy components especially when using smaller MD trajectories for both methods.

2.2.3 Rigorous relative binding free energy methods

Free energy perturbation methodologies

Zwanzig described a method of calculating free energy differences which was mathematically derived from the partition functions of each system [Zwanzig(1954)]:

$$\Delta F = F_B - F_A, \tag{2.47}$$

$$= -1/\beta \ln\{Z_B/Z_A\}, \tag{2.48}$$

$$= -1/\beta \ln\left\{\frac{\int exp\{-\beta U_B(q_N)\}}{\int exp\{-\beta U_A(q_N)\}}\right\}. \tag{2.49}$$

Let $U_B(q_N) = U_A(q_N) + U_{AB}(q_N)$,

$$= -1/\beta \ln\left\{\frac{\int exp\{-\beta[U_A(q_N) + U_{AB}(q_N)]\}}{\int exp\{-\beta U_A(q_N)\}}\right\}, \tag{2.50}$$

$$= -1/\beta \ln \langle \exp\{-\beta \Delta U_{AB}(q_N)\}\rangle_A . \tag{2.51}$$

Equation 2.51 gives ΔF from the exponential ensemble average of the difference in energy between the two systems A and B, where $\langle ... \rangle_A$ denotes an ensemble average over system A and $\Delta U_{AB}(q_N)$ the difference in the potential energies of A and B due to the present set of coordinates.

The computational implementation of Zwanzig's equation is called free energy perturbation (FEP). FEP in its simplest form entails running a simulation of one system (A or B) which at each simulation step adds to the average in equation 2.51. The $\Delta E_{AB}(q_N)$ term can be found using differing techniques. These techniques are single topology; where the difference between the systems A and B is found by evaluating the configuration of the simulation using two different potentials (A and B) and dual topology; where the difference is found due to the exchange of one or more molecules of system A for those of system B [Cramer(2002)]. Obviously, single topology is only applicable in cases where systems A and B are similar and have the same number of atoms, although, through the use of dummy atoms (atoms which cannot be seen by the rest of the system), differing numbers of atoms can be used. Dual topology can, in theory, be used with any two systems. In the case of a dual topology system which differs in one

molecule, both instances of the differing molecule, A and B, must be present and moved as part of the simulation all of the time. This is possible as the potential of molecules A and B cannot affect each other but only the rest of the system around them. This can cause difficulties as the system may react differently to each molecule. Also, the molecules will behave differently causing increased noise in $\Delta E_{AB}(q_N)$ measurements [Cramer(2002)].

In practice, systems A and B must be very similar if we are to produce a converged ΔF using equation 2.51. This is because FEP simulation techniques use the same system configuration to evaluate both systems A and B. Thus, if the most commonly sampled low energy region of phase space for system A is in a different place from that of system B, it will be difficult to sample the important regions of system B while the simulation is using the potential of system A. The simple solution to this problem is to ensure that any free energy difference evaluated using equation 2.51 has good overlap of important phase space of systems A and B. This often means using a series of intermediate systems between A and B to ensure good phase space overlap. This is achieved by coupling the differences between systems A and B to a simulation parameter, λ, where $\lambda = 0$ gives system A and $\lambda = 1$ gives system B. A simple and widely used example of this is the linear coupling of the system Hamiltonian to λ:

$$H_\lambda = (1 - \lambda)H_A + \lambda H_B. \tag{2.52}$$

This coupling of λ can also be implemented at the level of the parameters which vary between systems A and B. It is then simple to use equation 2.51 on intermediates between A and B and sum the resulting free energy differences to give an overall ΔF. This is the FEP technique, as shown in equation 2.53.

$$\Delta F = F_B - F_A = \sum_{\lambda=0}^{1} -\frac{1}{\beta} \ln \langle \exp\{-\beta \Delta U_{AB}(q_N)\} \rangle_{\lambda i}. \tag{2.53}$$

This technique of coupling the differences in systems A and B to λ is used in the many rigorous free energy methods. In general the change from system A to system B is called a perturbation.

Thermodynamic Integration

Thermodynamic Integration (TI) is a well established rigorous free energy method and is well represented in many texts [Leach(1996)], [Frenkel & Smit(1996)]. TI is based on the perturba-

tion technique, which uses λ, described above for FEP. Simulations are run at values of λ (λ windows) which allow good phase space overlap from systems A to B. The property accumulated by each simulation is the free energy gradient $(\frac{\partial F}{\partial \lambda})_\lambda$. ΔF from A to B is then found by integrating over the measured gradients:

$$\Delta F = F_B - F_A = \int_0^1 \left(\frac{\partial U}{\partial \lambda}\right)_\lambda d\lambda. \tag{2.54}$$

The free energy gradients can be found analytically through the ensemble average of the gradient of the force field. This in turn is found from the gradient of each force field term with respect to λ. Alternatively, the free energy gradients can be approximated numerically by the finite difference as in equation 2.55. TI which uses a finite difference approximation is called Finite Difference Thermodynamic Integration (FDTI) and will be used in this study over other forms of TI due to its simplicity.

$$\left(\frac{\partial F}{\partial \lambda}\right)_\lambda = \left\langle \frac{\Delta F}{\Delta \lambda} \right\rangle_\lambda. \tag{2.55}$$

ΔF in equation 2.55 can be found using the Zwanzig equation (equation 2.51) and potential values at λ and $\Delta\lambda$.

The size of the $\Delta\lambda$ evaluation made to find a gradient measurement, in FDTI, must be small in order that the exact gradient at the correct point is obtained. With FEP the size of $\Delta\lambda$s is dependent on the number of intermediate states between $\lambda = 0$ and $\lambda = 1$ used. Another issue is the presence of a non-linear exponential average in the FEP equation 2.53. This can cause bias in estimates of ΔF and is not present with TI based methods which evaluate a standard linear average. Owing to these necessities the ensemble average gradient may often be converged more easily with FDTI than with FEP. Also, any hysteresis of the same $\Delta\lambda$ evaluation in opposing directions is minimised (i.e. $\lambda = 0 \rightarrow 1$ compared to $\lambda = 1 \rightarrow 0$).

Replica Exchange Thermodynamic Integration

Replica Exchange Thermodynamic Integration (RETI) is a development of TI which incorporates Hamiltonian replica exchange moves between λ window simulations adjacent on the λ

coordinate (λ moves) [Woods *et al.*(2003a)Woods, Essex & King], [Woods *et al.*(2003b)Woods, Essex & King]. λ moves are made periodically and in such a way that one configuration cannot be swapped twice. In order that λ moves adhere to detailed balance, they are accepted or rejected with the equivalent of two metropolis tests, one for each configuration introduced to a new simulation. Thus, moves are accepted if,

$$\exp\{\beta[U_B(j) - U_B(i) - U_A(j) + U_A(i)]\} \geq rand(0,1), \tag{2.56}$$

is true, where i and j are configurations being exchanged, and A and B are the Hamiltonians of the replicas exchanging.

RETI increases sampling especially of the solvent by providing the possibility of ensembles making large jumps in phase space. Also, as simulations are able to move freely across λ, configurations which are more favourable to a particular area of λ may migrate there. These advantages are demonstrated through comparisons to conventional TI using water-methane and calix[4]-pyrrol test systems [Woods *et al.*(2003a)Woods, Essex & King], [Woods *et al.*(2003b)Woods, Essex & King].

The Adaptive Integration Method (AIM) is similar to RETI [Fasnacht *et al.*(2004)Fasnacht, Swendsen & Rosenberg]. AIM calculates ΔF using the same integral as TI and RETI, equation 2.54. An AIM simulation makes λ moves similar to RETI but with a differing Metropolis acceptance test,

$$\exp\{-\beta[E_B(j) - E_A(i)] + \beta[\Delta\hat{F}_n - \Delta\hat{F}_o]\} \geq rand(0,1), \tag{2.57}$$

where, $\Delta\hat{F}_o$ and $\Delta\hat{F}_n$ are the current running free energy estimates after and before the present λ move respectively. The incorporation of current ΔF estimates into the λ move acceptance test smoothes the convergence of the free energy gradients. This may make it possible to use fewer measurements for each ensemble average gradient and consequently more integration points could be used. The λ move acceptance test of AIM, however, only satisfies detailed balance asymptotically and estimates are only rigorously correct when fully converged.

AIM has been compared favourably with other free energy methods using relative hydration free energy test systems as discussed later in the next chapter [Ytreburg *et al.*(2006)Ytreburg, Swendsen & Zuckerman].

2.2.4 Other interesting methods

Methods which calculate free energy differences from computer simulation can be classified neatly into density of states methods and work based methods. An example of work based methods is FEP described above, where the change in energy is measured due to a change made to the Hamiltonian.

Density of states methods calculate ΔF directly through its relation to how the configurations of the systems under study are weighted by the Boltzmann factor [Kofke(2005)]. A simulation able to freely sample systems A and B will reside in each systems configurations in proportion to the Boltzmann factor of their free energies,

$$P(i) \propto exp\{-\beta F_i\} \tag{2.58}$$

where $P(i)$ is the probability of sampling system i with free energy F_i. ΔF is then found as the ratio of these probabilities.

In general, density of states methods have suffered as it is generally difficult to obtain sufficient sampling of all systems of interest. A recent and interesting example is the λ-AFED method of Abrams *et al.* This method uses λ coupling functions such that a large barrier between the end states A and B is created, ensuring that the simulation spends most of its time at the A and B end states and not in less relevant intermediate states [Abrams *et al.*(2006)Abrams, Rosa & Tuckerman]. The introduction of this new barrier can make switches from system A to B rare. To this end the λ variable is given a high temperature so the barrier is crossed easily and a large mass such that λ achieves adiabatic separation from the rest of the system. The free energy is then found easily through the probability distribution of the simulation residing at $\lambda = 0$ and $\lambda = 1$ ($P_{adb}(\lambda)$):

$$F(\lambda) = -k_B T_\lambda \ln P_{adb}(\lambda). \tag{2.59}$$

This method was compared to FEP and TI for three simple test systems including the solvation free energies of methane and methanol and found to be up to 15 times more efficient.

Another recent and relatively successful development is the transition matrix MC method (TMMC) [Errington(2003)]. This methodology uses the information from attempted transitions between configurations of a Markov chain. Using this data estimates of transition probabilities between states along the path of interest which changes the free energy of the system can be

found. However, methods such as these have not been applied to the problem of protein ligand binding free energies.

An interesting method which uses ideas from both density of states and work based view points is the λ dynamics method of Kong and Brooks (1996). The λ variable, which is used to control the mutation from one species to another with FEP is developed, giving λ_i. λ_i represents a set of molecules which simultaneously exist in the same space, whilst being invisible to each other, and vary with λ. One advantage of this type of methodology is that it is possible to evaluate more than two systems at once through the equivalent of a computational competitive binding experiment [Kong & Brooks(1996)], [Zhuyan *et al.*(2003)Zhuyan, Durkin, Fischmann, Ingram, Prongay, Zhang & Madison], although in practice the necessity for several systems existing in the same space can limit the application of this concept.

A λ_i is assigned to each of L ligands and the comparison uses a hybrid potential,

$$V(\lambda, x) = \sum_{i=1}^{L} \lambda_i^2 (V_i(x) - F_i) + V_{env}(x), \tag{2.60}$$

where $V_{env}(x)$ is the interaction involving enviromental atoms (those atoms common to every ligand system) and $V_i(x)$ is the interaction due to those atoms distinct to each (i) ligand system. In equation 2.60 the sum of λ_is is equal to one. F_i is a biasing potential and corresponds to the unbound free energy of the ligand, which has been determined previously. The different λ_i potentials will evolve, through the simulation, to find the lowest free energy regions of λ_i space.

The running averages of each λ_i is the probability distribution of that system and a reflection of the free energy difference between the molecules. The $\Delta\Delta G$ of any two molecules, i and j, can be calculated using equation 2.61, where $P_i(\lambda_i = 1, \{\lambda_k \neq i\} = 0)$ is the probability that the system is dominated by ligand i.

$$\Delta\Delta G_{ij} = -k_b T ln \frac{P_i(\lambda_i = 1, \{\lambda_k \neq i\} = 0)}{P_j(\lambda_j = 1, \{\lambda_k \neq j\} = 0)} \tag{2.61}$$

This method although seemingly complex can be quick to run as it only requires one relatively short simulation to rank a set of ligands, although multiple windows may be used to give increased accuracy with average free energy differences and errors. A recent study used λ-dynamics to find the relative binding free energies of HCV protease (HCVp) inhibitors and

results were compared to similar FEP calculations [Zhuyan *et al.*(2003)Zhuyan, Durkin, Fischmann, Ingram, Prongay, Zhang & Madison]. The suitability of this technique for comparing ligands with many differences and ligands which affect receptor conformation on binding has not been tested. This may prove to be a limiting factor in its use.

2.2.5 Summary and the direction of the present study

Here the possibility of using statistical mechanics and computers to study the interactions of biomolecules has been displayed. Protein-ligand systems can be modelled using molecular mechanics force fields and the behaviour of these systems explored using sampling techniques such as MD and MC. Various free energy methods can then be used to calculate free energy differences between related systems. These free energy differences can then be combined in a thermodynamic cycle to give the relative binding free energies of a set of inhibitors for an enzyme drug target.

Despite the myriad of methods for computational calculation of free energy differences, the methods most widely used for large biosystems are still FEP and TI. These methods are not recent developments having first been mentioned over 50 years ago, although it is only recently that the computational resources have become available with which to use these methods on large protein-ligand systems. Still, rigorous free energy calculations are thought to be computationally costly, limiting there use in the pharmaceutical industry. Hence the use by many of more approximate methods such as LIE and MMPBSA and more simple scoring functions (see chapter 1).

A recent development of work based free energy methods displays relations with both FEP and TI [Jarzynski(1997a)]. This set of methods, called fast growth (FG) in this work, is thought to be extraordinary in its ability to relate nonequilibrium simulation data to equilibrium free energies. Possibly just as extraordinary is the fact that this related set of methods has only recently been discovered so long after its relations FEP and TI and just at the right time to take advantage of the new distributed computational resources to which is it so suited well (see appendix D).

Owing to the potential of FG methods for fast free energy difference calculation these methods will be the subject of this study. The next chapter will review the FG literature with a view

to identifying those methods best suited to protein-ligand calculations and drug design. Useful FG methods will then be compared to the currently established methods FDTI and RETI to gauge there possible utility. This comparison will start with simple analytically tractable toy systems (chapter 4), moving on to simple solute-solvent systems (chapter 5). The best of the remaining methods will then be compared to RETI for two protein-ligand binding free energy studies (chapters 6 and 7). The final chapter will summarise and discuss the issues arising from this work.

Chapter 3

Nonequilibrium Free Energy Methods

3.1 Introduction

In the previous chapters the utility of fast free energy calculations in rational drug design has been described. However, as discussed improvements in efficiency and applicability of these methods are required before wide spread use in the drug design process becomes a reality. Recent advances in nonequilibrium free energy methods have provided the possibility of improving the speed of these calculations through use of large distributed computing resources.

Hence, here nonequilibrium free energy methods will be reviewed in detail and the relative attributes of these methods compared to the presently established methods.

3.2 Slow Growth

Slow Growth or adiabatic switching (SG) uses the same perturbation methodology as FEP and TI. SG was originally presented as a TI-like integration over a continuous sequence of equilibrium states linking systems A and B [Kirkwood(1935)]. A possibly more pleasing description is that SG utilises the basic rule of thermodynamics,

$$\Delta F = W_{\infty}. \tag{3.1}$$

Equation 3.1 states that over the course of a reversible, isothermal process linking two equilibrium states, the work (W) performed on the system is equal to the free energy difference between the two states. For a process linking two states, also known as a switch to be truly

reversible, in principle, it must be infinitely long. For this reason, SG switches cannot be truly reversible. Accuracy of SG calculations relies on how close to the reversible limit the simulated switch is.

In practice an SG switch is a simulation coupled to a λ coordinate, where λ is slowly incremented from 0 to 1. The work performed as a consequence of each λ increment is summed to give a free energy difference estimate, ΔF, of the true free energy difference, ΔF_∞. This is expressed in terms of the potential at the present λ value, $U_{\lambda_i}(q_N)$ and the potential at the next value of λ, $U_{\lambda_{i+1}}(q_N)$, as follows,

$$\Delta F = W = \sum_{i=1}^{n} U_{\lambda_{i+1}}(q_N) - U_{\lambda_i}(q_N). \tag{3.2}$$

An SG calculation invariably produces a systematic error due to its non-equilibrium nature. The simulation lags behind the changing potential, this is often referred to as Hamiltonian lag [Pearlman & Kollman(1989)]. Hamiltonian lag contributes positively to W such that,

$$\Delta F = W - W_{diss}. \tag{3.3}$$

This contribution is called the dissipated work (W_{diss}) and is associated with the increase of entropy during an irreversible process [Jarzynski(1997a)]. Hamiltonian lag is most easily demonstrated through the consideration of switches proceeding in opposite directions i.e. switches from system A to B as already discussed but also starting from system B and ending at system A. A switch from A to B ($\lambda = 0$ to $\lambda = 1$) will from now on be termed forward and B to A ($\lambda = 1$ to $\lambda = 0$) backward, and it will be assumed they have the same internal structure (i.e. numbers of λ increments and MC trials). A non-zero W_{diss} means a hysteresis will exist between switches in the forwards and backwards directions, i.e.

$$\bar{W}_F + \bar{W}_R \neq 0, \tag{3.4}$$

where $\bar{W}_{F/R}$ is the average work in the forwards or backwards directions respectively. The fact that the W_{diss} always contributes positively means that the true ΔF (ΔF_∞) must always be between values for W in switches in opposing directions. Some studies have suggested that

mean W values ($\bar{W}$) of switches in opposing directions can give good estimates of ΔF_∞ as W_{diss} from opposing directions may cancel, i.e.

$$\Delta F = (\bar{W}_F - \bar{W}_R)/2, \tag{3.5}$$

[Wood(1991)], [Hermans(1991)], [Hu *et al.*(2002)Hu, Yun & Hermans], [Hummer(2002)]. Equation 3.5 will provide good ΔF estimates only when the free energy difference of study provides identical barriers to sampling in both forward and backward directions. The systems and changes in potential which display such symmetries are more often simple systems with simple potential changes. However, the size of W_{diss} is proportional to the length of the SG switch, therefore with reasonably slow switch rates the error in an estimate of ΔF_∞ after the use of equation 3.5 may be low even for systems with non-symmetrical forwards and backwards switches [Hu *et al.*(2002)Hu, Yun & Hermans]. The efficiency of the estimation method described by equation 3.5 (symmetric A) will be investigated for various ΔF calculations in this study. Also, unlike FEP and TI, for SG the potential change due to each λ increment is represented by one measurement. These intrinsic problems are undoubtedly the main reasons SG calculations are thought to be relatively inaccurate. With a few exceptions, SG calculations are not presently used to estimate free energy differences [Hu *et al.*(2002)Hu, Yun & Hermans]. However, one advantage of SG and methods which use continuous switching simulations is that the production of a calculation with small levels of hysteresis gives high levels of confidence in the result. In such a case the simulated process is a good approximation of quasi-static process which underlies reversibility and free energy differences.

3.3 Fast Growth

Fast Growth (FG) as the name suggests is similar to SG except the switches made can be made faster. Recently there has been a renewed interest in methods which employ a continuous switching process methodology. This is mainly due to the work of Jarzynski and the equality he derived [Jarzynski(1997b)], [Jarzynski(1997a)].

Jarzynski discovered an equality which relates the distribution of non-equilibrium work values used with equation 3.5 to the equilibrium free energy difference rather than an upper or

lower bounding value.

$$\langle exp\{-\beta W\}\rangle = exp\{-\beta \Delta F\} \tag{3.6}$$

$$\Delta F = -k_B T \ln \langle exp\{-\beta W\}\rangle \tag{3.7}$$

Equation 3.7 reduces to FEP where switches are instantaneous and is related to TI in the case that the work is integrated across λ (as in the previous chapter $\langle ... \rangle$ denotes an ensemble average). The Jarzynski equality is also closely related to other interesting non-equilibrium relations recently derived [Crooks(1999)] [Crooks(2000)].

For equality 3.6 to be true the switches made to produce the distribution of work values must have initial configurations taken from the same equilibrium ensemble (an ensemble of switches). Also, all switches in an ensemble must be of the same internal structure (same numbers of λ increments and simulation steps) as expressed by,

$$exp\{-\beta \Delta F\} = \frac{\int d[\mathbf{Z}_F] \;\; P[\mathbf{Z}_F|\lambda_t] exp\{-\beta W\}}{\int d\mathbf{z}_0 \;\; P(\mathbf{z}_0)}. \tag{3.8}$$

Here, $P[\mathbf{Z}_F|\lambda_t]$ is the probability of the switch, $\mathbf{Z}_F$, being produced when switches are structured with an amount of λ switches and MC trials represented by λ_t and $P(\mathbf{z}_0)$ is the probability of a switch starting configuration $\mathbf{z}_0$ i.e. the equilibrium ensemble probability of a configuration for system A ($P(\mathbf{z}_0) = exp\{-\beta U_A(q_N)\}$). Of course this integral does not need to be considered as switches are produced automatically with the correct frequency. The practical FG protocol is described in figure 3.1 starting with system equilibration and then performing switches from system A (circles) to system B (squares).

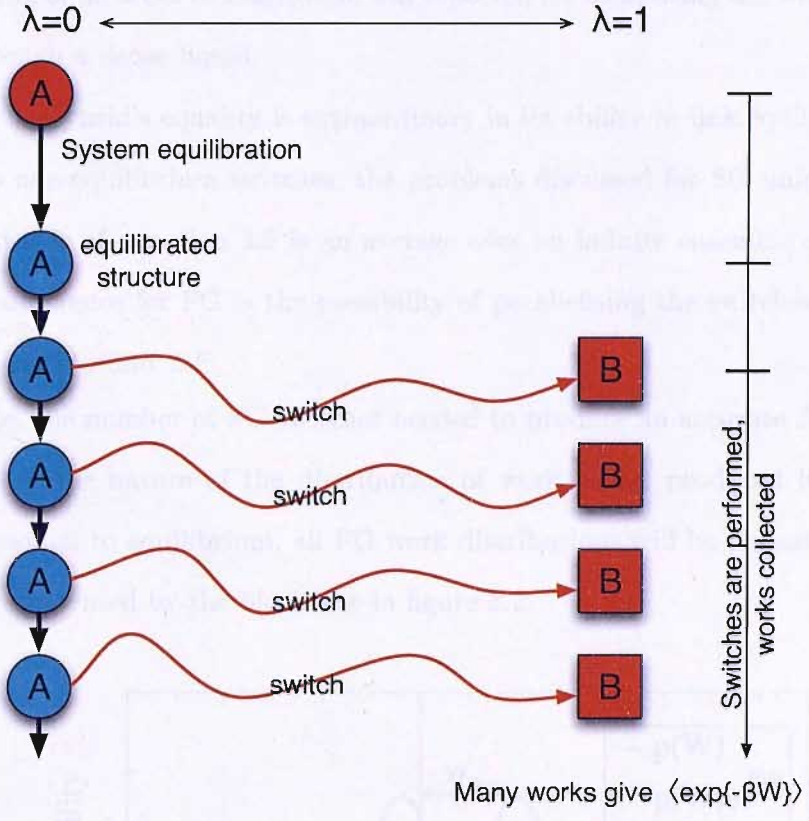


Figure 3.1: Diagram describing the FG protocol. Circles represent the configurations of system A and squares the configurations of system B. Circles/squares filled with cyan are equilibrated while those filled with red are not equilibrated. Full black lines represent equilibrium simulations and thinner red wavy lines represent FG switches.

Equality 3.6 has been explicitly proved for the cases of MC, MD (coupled to various thermostats) and Langevin dynamics [Jarzynski(1997b)], [Jarzynski(1997a)], [Crooks(1998)], the main difference being that MD would produce deterministic switches while MC and MD coupled to stochastic thermostats would produce stochastic forms. In the case of deterministic dynamics, each switch and work value depends only on the equilibrium starting configuration and thus the distributions $P[\mathbf{Z}_F|\lambda_t]$ and $P(\mathbf{z}_0)$ become the same distribution. MC FG switches, although possibly non-equilibrium are assumed to obey the Markov assumption and detailed balance (in this case detailed balance is not assumed through reversibility but through the behaviour of the system while λ is fixed) [Jarzynski(1997b)]. Thus, it is not possible to use MC moves which could not be used for equilibrium MC sampling in FG switches. Interestingly, it has been shown that for stochastic and deterministic MD arbitrarily large time steps can be used and Jarzynski's equality still holds [Lecher *et al.*(2006)Lecher, Oberhofer, Dellago & Geissler]. Using large MD time steps can lower the computational expense of switches and an

efficiency saving of an order of magnitude was reported for computing the work needed to move a particle through a dense liquid.

Although Jarzynski's equality is extraordinary in its ability to link equilibrium free energy differences to non-equilibrium switches, the problems discussed for SG unfortunately remain. The left hand side of equation 3.6 is an average over an infinite ensemble of switching possibilities. The difference for FG is the possibility of parallelising the switches and reducing the clock time needed to find ΔF.

In practice, the number of FG switches needed to produce an accurate ΔF estimate varies and depends on the nature of the distribution of work values produced by the calculation. When close enough to equilibrium, all FG work distributions will be Gaussian, as is the work distribution represented by the black line in figure 3.2.

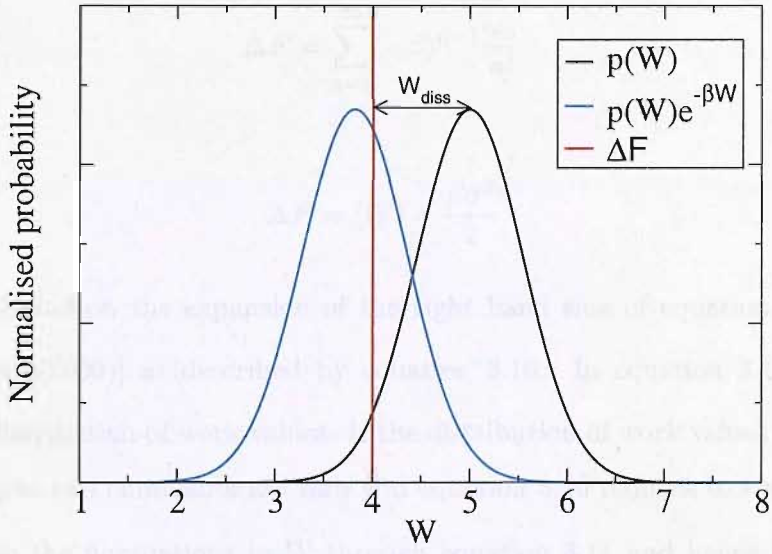


Figure 3.2: Depiction of a Gaussian FG work distribution (in black). The blue line represents $p(W)e^{-\beta W}$ which is the distribution of the exponential weights of the works. Also, the red line denotes the free energy difference.

The exponential average of the Jarzynski equality can be written as an integral over the distribution of work values,

$$\langle exp(-\beta W)\rangle = \int dW p(W) exp(-\beta W). \tag{3.9}$$

The integrand distribution of equation 3.9 labeled $p(W)e^{-\beta W}$ in figure 3.2 is peaked to the left of the average work and ΔF. The work values which contribute most to the right hand side

of the Jarzynski equality (equation 3.6) are those in the peak of $p(W)e^{-\beta W}$ and in the far left hand tail of the work distribution p(W). These important work values (termed dominant) are thus, unfortunately, produced by the rarest switches. This is the major problem of estimating ΔF with equation 3.7.

The probability of producing a switch in the peak of $p(W)e^{-\beta W}$ sharply decreases as a function of W_{diss} [Gore *et al.*(2003)Gore, Ritort & Bustamante]. Therefore, it is important that the $\bar{W}_{diss}$ of the work distribution should always be minimised to produce an efficient FG calculation.

Another useful non-equilibrium free energy relation also introduced by Jarzynski is described by equation 3.11 [Jarzynski(1997a)].

$$\Delta F = \sum_{n=1}^{\infty} (-\beta)^{n-1} \frac{\omega_n}{n!} \tag{3.10}$$

$$\Delta F = \langle W \rangle - \frac{\beta \sigma^2}{2} \tag{3.11}$$

Equation 3.11 is based on the expansion of the right hand side of equation 3.7 as a sum of cumulants [Sornette(2000)] as described by equation 3.10. In equation 3.10 ω_n is the nth cumulant of the distribution of work values. If the distribution of work values used is Gaussian then all but the first two cumulants are zero and equation 3.10 reduces to equation 3.11. The $\bar{W}_{diss}$ is related to the fluctuations in W through equation 3.11 and hence this estimator of the free energy difference is called the fluctuation-dissipation estimator (FD). As 3.11 is only accurate for Gaussian distributions of W it is important to use switches which are close to equilibrium, as although a Gaussian work distribution is possible without switches being close to equilibrium (e.g. a bead being dragged through water by a Hookean spring [Mazonka & Jarzynski(1999)]), being close to equilibrium ensures a Gaussian distribution of W values.

The FD estimator has also been used with some success in a similar arrangement to $\bar{W}$ in equation 3.5 as,

$$\Delta F = \frac{(\bar{W}_F - \bar{W}_R)}{2} - \frac{\beta(\sigma_F^2 - \sigma_R^2)}{12}, \tag{3.12}$$

[Hummer(2002)]. This estimator of ΔF, called symmetric B in this work, will be compared to other interesting methods, with various simple test systems.

Studies of the error and efficiency of FG calculations seem to have come to a number of useful conclusions [Gore *et al.*(2003)Gore, Ritort & Bustamante], [Zuckerman & Woolf(2002b)], [Hummer(2001)], [Shirts & Pande(2005)].

1. As $\bar{W}$ moves away from ΔF the distribution of works gets wider and the probability of a work value close to ΔF becomes less. The probability of a negative $\bar{W}_{diss}$ is related to $\bar{W}_{diss}$, in general the larger the $\bar{W}_{diss}$ the fewer highly weighted switches. Thus, it is important to minimise the $\bar{W}_{diss}$ [Gore *et al.*(2003)Gore, Ritort & Bustamante] [Jarzynski(1997a)].

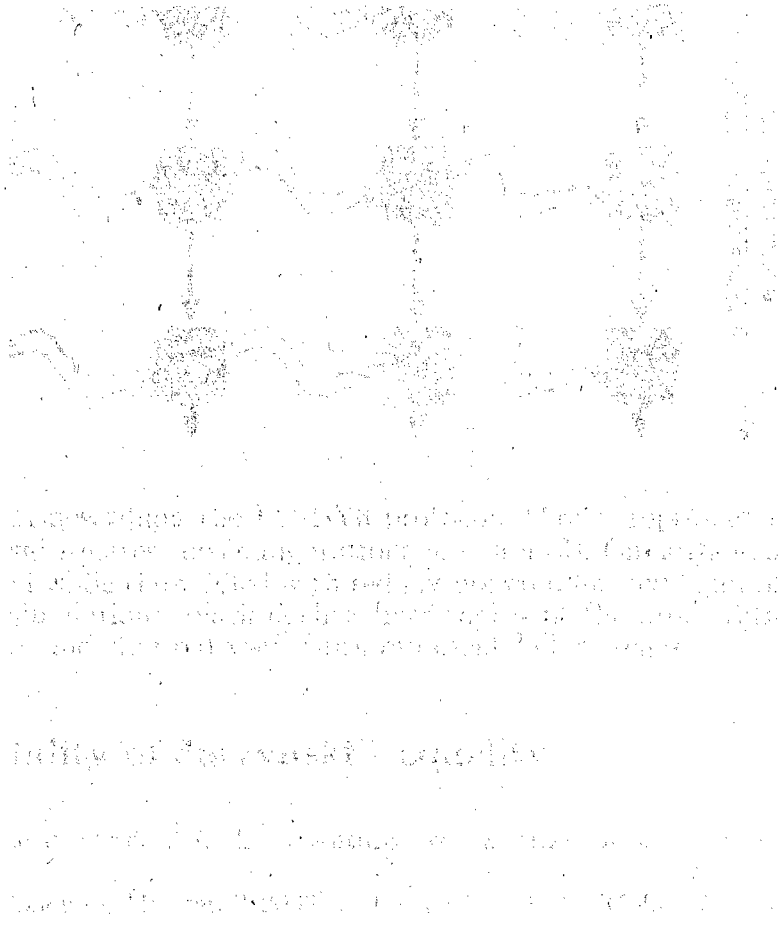
2. The near equilibrium/gaussian work distribution regime should be obtainable from any system if switched slowly enough. This regime is needed for FD to work and for the Jarzinksi estimator to be accurate [Gore *et al.*(2003)Gore, Ritort & Bustamante].

3. The variance of $\bar{W}$ scales approximately as $1/t$ where t = switch time [Hummer(2001)]. Hence, all relevant studies have concluded that it is in general more efficient to use fewer longer switches than more shorter switches [Gore *et al.*(2003)Gore, Ritort & Bustamante], [Zuckerman & Woolf(2002b)], [Shirts & Pande(2005)].

4. For the FD related estimators, if the standard deviation of a distribution of works is of order K_bT statistical error is approximately independent of number of repetitions M for a particular length of switch [Hummer(2001)]. Thus, if the variance of a work distribution is of order K_bT there is no point in adding to the work distribution.

5. It may be optimal to divide the λ co-ordinate into a number of intervals which are evaluated independently. For example, if equilibrium seed simulations are run at $\lambda = 0$, 0.25, 0.75 and 1, switches can be run from $\lambda = 0$ to $\lambda = 0.25$, $\lambda = 0.25$ to $\lambda = 0.75$ and $\lambda = 0.75$ to $\lambda = 1$ with the resultant FG ΔFs added to get the ΔF from $\lambda = 0$ to $\lambda = 1$. This arrangement is described in figure 3.3 and can be called a FG-BY3 protocol as the

λ co-ordinate has been split by three. This organisation of switches increases the overlap of switch end states and therefore could improve convergence in much the same way it does for TI and FEP. However, as the number of intermediate equilibrium starting states across λ is increased, the number of ΔFs, each with a bias and error needing to be added, increases too. Without increasing the number of simulation steps used, the amount of sampling for each independent FG calculation will decrease as the number of λ divisions increases. Thus, for some systems, for which FG converges slowly, extra λ divisions may result in increased bias and error [Hummer(2001)], [Shirts & Pande(2005)].



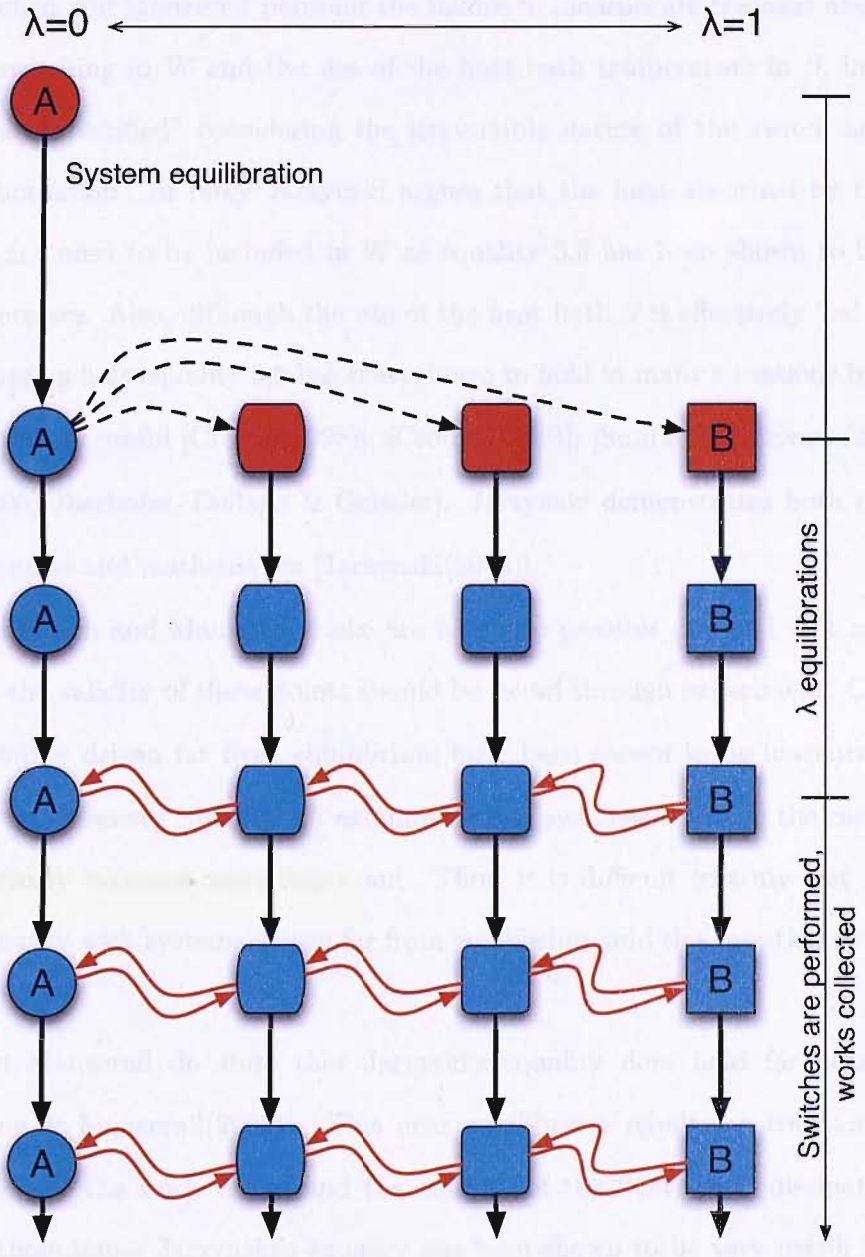


Figure 3.3: Diagram describing the FG-BY3 protocol. Circles represent the configurations of system A and squares the configurations of system B. Circles/squares filled with cyan are equilibrated while those filled with red are not equilibrated. Full black lines represent equilibrium simulations, black dashed lines represent the proliferation of a single system configuration and thin red wavy lines represent FG switches.

3.3.1 The validity of Jarzynski's equality

There has been some discussion in the literature over whether Jarzynski's equality (3.6) is physically correct [Cohen & Mauzerall(2004)], [Jarzynski(2004)], [Cohen & Mauzerall(2005)]. Cohen and Mauzerall claimed that the derivation of equality 3.6 was flawed by the improper handling of heat exchange between the system and the water bath [Cohen & Mauzerall(2004)].

Specifically, Cohen and Mauzerall pinpoint the failure to incorporate the heat absorbed by the system when switching in W and the use of the heat bath temperature in β, in equality 3.6 as "ad hoc and unjustified" considering the irreversible nature of the switch and the possible thermal fluctuation. In reply Jarzynski argues that the heat absorbed by the system in a switch does not need to be included in W as equality 3.6 has been shown to be correct for irreversible processes. Also, although the use of the heat bath β is effectively "ad hoc", it is in no way improper and, as equality 3.6 has been shown to hold in many situations by many other authors, is extremely useful [Crooks(1998)], [Crooks(1999)], [Sun(2003)], [Evans(2003)], [Oberhofer *et al.*(2005)Oberhofer, Dellago & Geissler]. Jarzynski demonstrates both of these ideas with clear examples and mathematics [Jarzynski(2004)].

The points Cohen and Mauzerall make are based on possible physical, not mathematical, errors. Hence, the validity of these points should be found through experiment. Cohen emphasizes that examples driven far from equilibrium have been shown to be inaccurate [Cohen & Mauzerall(2004)]. However equality 3.6 assumes infinite switches which in the case of very fast switching obviously becomes more important. Thus, it is difficult to truly test the ability of Jarzynski's equality with systems driven far from equilibrium and this question requires further study.

Cohen and Mauzerall do state that Jarzyski's equality does hold for near equilibrium switches [Cohen & Mauzerall(2005)]. This near equilibrium requires a constant β, a Gaussian distribution of the work values and the validity of the fluctuation dissipation theorem. Also, despite these issues Jarzynski's equality has been shown to be very useful in many contexts, experimental and simulation [Liphardt *et al.*(2002)Liphardt, Dumont, Smith, Tinoco & Bustamante], [Douarche *et al.*(2005)Douarche, Ciliberto, Petrosyan & Rabbiosi], [Collin *et al.*(2005)Collin, Ritort, Jarzynski, Smith, Tinoco & Bustamante] with and without a Gaussian distribution of work values [Hummer(2002)], [Shirts & Sorin(2005)]. Jarzynski's equality has been shown to give accurate results in comparison to more established, non controversial, equilibrium methods [Shirts & Pande(2005)], [Ytreburg *et al.*(2006)Ytreburg, Swendsen & Zuckerman], [Oostenbrink & Gunsteren(2006)].

3.3.2 Evaluating forward and backward switches together

As discussed above for SG, the Jarzynski estimator applies to switches and work distributions in both forwards and backwards directions, with

$$\langle exp(-\beta W)\rangle_F = exp\{-\beta\Delta F_F\}, \langle exp\{-\beta W\}\rangle_R = exp\{-\beta\Delta F_R\}, \tag{3.13}$$

where ΔF_F is the free energy difference estimate in the forwards direction and ΔF_R in the backwards direction. Hence, $\Delta F_{\infty F} = -\Delta F_{\infty R}$.

The work distributions for FG calculations in the forwards and backwards directions can be related through a generalised version of the entropy fluctuation theorem derived for stochastic microscopically reversible dynamics by Crooks (1999a). The dynamics in question must satisfy a condition of microscopic reversibility as follows,

$$\frac{P[\mathbf{Z}_F|\lambda_t]}{P[\mathbf{Z}_R|\lambda_t]} = exp\{-\beta Q[\mathbf{Z}_F|\lambda_t]\} \tag{3.14}$$

where $P[\mathbf{Z}_F|\lambda_t]$ is the probability of the switch, $\mathbf{Z}_F$, being produced when switches are structured with a number of λ switches and MC trials represented by λ_t and $P[\mathbf{Z}_R|\lambda_t]$ is the probability of the equivalent backward switch. Q is the energy moving from the heat bath to the system, which is odd under time reversal, $Q[\mathbf{Z}_F|\lambda_t] = -Q[\mathbf{Z}_R|\lambda_t]$. This microscopic reversibility is similar but still distinct to the principle of microscopic reversibility at equilibrium originally given by Tolman [Tolman(1924)]. Microscopic reversibility, as in equation 3.14, is generally satisfied with typical stochastic simulation techniques such as Metropolis Monte Carlo and Langevin dynamics as long as individual steps satisfy the condition of detail balance described earlier in chapter 2.

Crooks was able to express the entropy fluctuation theorem, which has been proven for a range of systems [Evans & Searles(1994)], [Kurchan(1998)], in terms of the amount of work performed on a system that starts in equilibrium, i.e. the specific assumptions of the Jarzynski equality, as follows,

$$\frac{P_F(\beta W)}{P_R(-\beta W)} = exp\{-\beta\Delta F\}exp\{\beta W\} = exp\{\beta W_{diss}\}. \tag{3.15}$$

In equation 3.15 $P_F(\beta W)$ is the probability of realising a particular work value for a switch in the forwards direction and $P_R(-\beta W)$ is the same for the backwards switches. Crooks validated

his expression with an analytically tractable one dimensional model system [Crooks(1999)]. For explanations of more general cases where this fluctuation theorem is correct, check Evans and Searles (2002). Crooks and Jarzynski have subsequently shown equation 3.15 to hold for a simple system with non-Gaussian work fluctuations [Crooks & Jarzynski(2007)]. Also, the nonequilibrium fluctuation theorem of Crooks (equation 3.15) has been validated with single molecule force measurements on a real system [Collin *et al.*(2005)Collin, Ritort, Jarzynski, Smith, Tinoco & Bustamante].

Using the Jarzynski equality and equation 3.15 Crooks was subsequently able to show that,

$$exp(-\beta\Delta F) = \frac{\langle f(W)\rangle_F}{\langle f(-W)exp\{-\beta W\}\rangle_R}, \tag{3.16}$$

is true where $f(W)$ is any function of the work with subscript F or R denoting the work is in the forwards or backwards directions [Crooks(2000)]. Equation 3.16 is the same form derived by Bennett for FEP (a limiting case of FG) in his derivation of the Bennetts acceptance ratio method (BAR) [Bennett(1976)].

Bennett found the function $f(W)$ which when used in equation 3.16 gave the ΔF with the lowest variance and therefore the highest statistical accuracy was,

$$f(W) = (1 + \frac{n_F}{n_R} exp\{\beta W - \Delta F\})^{-1} \tag{3.17}$$

where n_F and n_R are the numbers of switches (in the case of FEP these would be instantaneous switches) in the forwards and backwards directions. ΔF can now be found: starting with the lower bound estimate of ΔF, found through the average work, ΔF is increased slowly and iteratively until ΔF satisfies equation 3.16. This ΔF is the BAR estimate.

BAR for nonequilibrium switching has recently been derived using Bayesian theory to show that it is the maximum likelihood estimator for ΔF with a given set of forwards and backwards switches [Shirts *et al.*(2003)Shirts, Blair, Hooker & Pande]. This means that a BAR estimate will have the lowest possible variance of any asymptotic estimator for a given set of forwards and backwards switches. It should be noted that this does not necessarily mean that a BAR estimate is more accurate than others with regards to ΔF_∞.

The ratio in equation 3.15 can be written in Bayesian notation as with $P_F(\beta W)$ being

$P(W|F)$, the conditional probability of a particular work value assuming it comes from a forward switch of a particular structure, and $P_R(\beta W)$ being $P(W|R)$, the conditional probability of a particular work value assuming it comes from a backwards switch of the same structure. As $P(F|W) + P(R|W) = 1$ Bayesian theory allows the following rearrangement:

$$\frac{P(W|F)}{P(W|R)} = \frac{P(F|W)P(R)}{P(R|W)P(F)} = \frac{P(F|W)}{1 - P(F|W)}\frac{P(R)}{P(F)}. \tag{3.18}$$

We can substitute this into equation 3.15 assuming that $P(R)/P(F) = n_R/n_F$, giving,

$$\ln \frac{P(F|W)}{1 - P(F|W)} = \ln \frac{n_F}{n_R}\beta(W - \Delta F) \tag{3.19}$$

Given a particular ΔF value, it is now possible to express $P(F|W_i)$, the probability of a particular forwards work value as,

$$P(F|W_i) = (1 + \frac{n_F}{n_R}exp\{-\beta(W_i - \Delta F)\}^{-1} \tag{3.20}$$

and $P(R|W_i)$, the probability of a particular forwards work value, given a particular estimate of ΔF, as,

$$P(R|W_i) = (1 + \frac{n_F}{n_R}exp\{\beta(W_i - \Delta F)\}^{-1} \tag{3.21}$$

Using these expressions for $P(F|W_i)$ and $P(R|W_i)$ it is now possible to find the likelihood of obtaining a set of forwards and backwards work values for a given ΔF:

$$L(\Delta F) = \prod_{i=1}^{n_F} P(F|W_i) \prod_{j=1}^{n_R} P(R|W_i). \tag{3.22}$$

Thus the best estimate of ΔF is that which maximises $L(\Delta F)$. However, it may be easier to find the maximum log likelihood. Shirts *et al.* (2003a) take the log likelihood and then differentiate this with respect to ΔF, setting it to zero, to give:

$$\frac{\delta \ln L(\Delta F)}{\delta \Delta F} = \sum_{i=1}^{n_F}(1+\frac{n_F}{n_R}exp\{\beta(W_i-\Delta F)\})^{-1} - \sum_{j=1}^{n_R}(1+\frac{n_F}{n_R}exp\{-\beta(W_j-\Delta F)\})^{-1} = 0 \tag{3.23}$$

Equation 3.23 is the same as BAR described by Bennett (1976) and discussed above. Hence, BAR gives the ΔF estimate which, maximises the probability that the observed work values are realised. Shirts *et al.* also discuss and derive a formula for the variance of a BAR estimate although this was not used in this study and will not be discussed here.

It is instructive to consider the efficiency of BAR with different n_F/n_R ratios. In the cases of $n_F \gg n_R$ and $n_R \gg n_F$ BAR (equation 3.23) collapses to the Jarzynski equality (equation 3.6). This is not surprising considering the origins of BAR. Bennett has shown that the optimal ratio of n_F/n_R for BAR is generally close to 1 [Bennett(1976)]. This makes sense as owing to equation 3.15 the forwards and backwards work distributions must be thought of together as a single independent distribution. To estimate this single distribution well, forwards and backwards switches should be used in equal amounts.

As mentioned earlier, the issues discussed here may suggest that BAR should always be used in preference to other FG estimators. In fact the literature does describe cases where BAR is not the most efficient estimator of ΔF. Shirts and Pande describe one such case with simple two dimensional harmonic oscillator systems which give Gaussian work distributions [Shirts & Pande(2005)]. This harmonic oscillator test system was defined by

$$H = \omega(x_i - x_j)^2, \tag{3.24}$$

where H is the Hamiltonian and, x_i and x_j are the particle coordinates. Of the perturbation end points A and B, A had a larger force constant (ω) than B and consequently a smaller range of possible particle positions (phase space). Shirts and Pande found that Jarzynski estimates in the backwards directions where extremely poor with a variance tending to infinity as ω of A becomes twice as big as ω of B. The forwards Jarzynski estimates were found to be more efficient than BAR estimates with slightly lower standard errors. The origin of this result is in the relative behaviour of the forwards and backwards work distributions. The backwards work distribution was so poorly behaved that it causes BAR to be less efficient than the forwards work distribution alone.

While BAR is possibly the best FG estimator in most cases, as it gives the estimate with the lowest variance for a given set of forwards and backwards work values, in some cases a Jarzynski estimate using work values from one direction will be more efficient. With this being the case it becomes important to know when a particular estimator or work distribution should

be used in preference to BAR.

Recently, an interesting extension to BAR has been developed for finding the maximum likelihood estimates of free energy differences of multiple states using parallel tempering simulations [Maragakis *et al.*(2006)Maragakis, Spichty & Karplus]. This method considers a set of systems (N_{sys}) with different Hamiltonians each sampled at a number of replicas with different temperatures. Switches are allowed from any of the replicas of one system to another. Maragakis *et al.* found this method to be many times more efficient than BAR for a simple vacuum test system.

3.3.3 Bias calculation

The FD and Jarzynski estimators are asymptotic and so for finite numbers of switches (N), estimates are biased. This bias is a result of the often inefficient sampling of the important but rare switches discussed above. To be clear, the bias of an FG estimate is,

$$B = \Delta\bar{F} - \Delta F_{\infty} \tag{3.25}$$

where B is the bias, $\Delta\bar{F}$ is the average ΔF, and ΔF_{∞} is the true free energy difference. In the limit of infinite switches this bias will be zero ($\Delta\bar{F} = \Delta F_{\infty}$), and it will increase monotonically with decreasing N, until N=1 and B=$\bar{\mathrm{W}}_{diss}$. Some literature investigations of FG concentrate on attempting to define and correct for the bias of FG estimators [Zuckerman & Woolf(2002b)], [Gore *et al.*(2003)Gore, Ritort & Bustamante], [Wu & Kofke(2004)]. Other studies have attempted to define tests to check for a result that is free of bias [Wu & Kofke(2005a)], [Jarzynski(2006)].

Gore *et al.* and Zuckerman and Woolf have independently derived the same identity for the bias of the Jarzynski estimator (B_J) when N is said to be large (equation 3.29) [Zuckerman & Woolf(2002b)], [Gore *et al.*(2003)Gore, Ritort & Bustamante]. This derivation starts with the reasonable assumption that the variance $Var(exp\{-\beta W\})$ is finite. For large N the central limit theorem guarantees that many realisations of $\langle exp\{-\beta W\}\rangle$ (denoted by Y) gives a normal distribution with mean $exp\{-\beta\Delta F_{\infty}\}$ (denoted by $\bar{Y}$) and variance $Var(exp\{-\beta W\})/N$. A linear expansion of $\ln(Y)$ around $Y = \bar{Y}$ gives,

$$\Delta F = -\frac{1}{\beta}\left[\ln(\bar{Y}) + \frac{(Y-\bar{Y})}{\bar{Y}} - \frac{(Y-\bar{Y})^2}{2!\bar{Y}^2} + \frac{2(Y-\bar{Y})^3}{3!\bar{Y}^3} - ....\right] \tag{3.26}$$

$$\Delta\bar{F} = \Delta F_\infty + \frac{Var(exp\{-\beta W\})}{2\beta exp\{-2\beta\Delta F_\infty\}N} + ... \tag{3.27}$$

Equation 3.27 has ΔF_∞ and a second term equal to the average free energy difference estimate. Hence, Gore *et al.* finds that B_J of the estimate is,

$$B_J = \frac{Var(exp\{-\beta W\})}{2\beta exp\{-2\beta\Delta F_\infty\}N} = \frac{Var(exp\{-\beta W_{diss}\})}{2\beta N}. \tag{3.28}$$

$$B_J = \frac{\exp\{\beta^2\sigma_W^2 - 1\}}{2\beta N} = \frac{\exp\{2\beta\bar{W}_{diss} - 1\}}{2\beta N}. \tag{3.29}$$

where σ_W^2 is the variance of the work values. Gore *et al.* define large N with relation to the $\bar{W}_{diss}$, for switches arbitrarily far from equilibrium as N $\gg$ variance of $exp\{\beta W_{diss}\}$ (Var($exp\{\beta W_{diss}\}$)). Assuming the variance of the work values is finite, with large enough N, the work distribution should be Gaussian as discussed above. However when N is this large the bias is generally not significant and would be dominated by the statistical error. Also, this large N may be difficult to obtain for systems of interest such as biomolecular systems where it is difficult to minimise the $\bar{W}_{diss}$.

It is more useful to concentrate on possibilities where switches are in the near equilibrium regime and the work distribution is Gaussian. Here it may be possible to calculate the bias of FG estimates when the bias is the dominant form of error. Gore *et al.* approximate B_J in the near equilibrium regime with small N as,

$$B_J \approx \frac{\bar{W}_{diss}}{N^{\alpha_G}}. \tag{3.30}$$

Equation 3.30 comes directly from the observation that a log-log plot of B_J against N, where N is small, is approximately linear. The α_G variable is placed to account for the differences in how fast the bias falls away with different amounts of $\bar{W}_{diss}$, as systems with larger $\bar{W}_{diss}$ have a bias which falls away more slowly as N is increased. The α_G parameter must be assigned with consideration of where the small N limit ends. In the near equilibrium regime when N is large,

equation 3.29 is applicable.

Through equations 3.29 and 3.30 it may be possible to find B_J for a Jarzynski estimate and to correct it giving a more accurate estimate of the true ΔF. However, the point at which the calculation is perceived to switch from small to large N must be defined. As equation 3.29 arises by assuming that the first order term of equation 3.27 is dominant, the point $N \gg (e^{2\beta\bar{W}_{diss}-1})$ is a good choice for the small/large N intersect. Thus, Gore *et al.* assume that the intersect of small and large N occurs at,

$$N_C = C(e^{2\beta\bar{W}_{diss}-1}), \tag{3.31}$$

where $C \gg 1$ is a constant that defines how small the bias must be before the large N limit is reached. Hence,

$$B_J(Nc) = \frac{\bar{W}_{diss}}{N_C^{\alpha_G}} = \frac{\exp(2\beta\bar{W}_{diss}-1)}{2\beta N_C} = \frac{1}{2\beta C}. \tag{3.32}$$

The α_G parameter from equation 3.30 can now be defined for small N bias using equation 3.32:

$$\alpha_G = \frac{\ln\left[2\beta C\bar{W}_{diss}\right]}{\ln\left[CVar(e^{2\beta W_{diss}})\right]} = \frac{\ln\left[2\beta C\bar{W}_{diss}\right]}{\ln\left[C(e^{2\beta\bar{W}_{diss}}-1)\right]}. \tag{3.33}$$

It is now possible to calculate the bias of a Jarzynski estimate using the estimated $\bar{W}_{diss}$. However, this estimate of $\bar{W}_{diss}$ is biased due to the bias of the Jarzynski estimate ΔF. Thus, an attempt to correct this estimate of the $\bar{W}_{diss}$ can be made by using $\bar{W}_{diss2}$ as defined in equation 3.34. Figure 3.4 is a flow diagram of the steps taken to calculate the corrected bias of a Jarzynski estimate (Gore bias, B_{J2}).

$$\bar{W}_{diss2} = \bar{W} - \Delta F - B_J \tag{3.34}$$

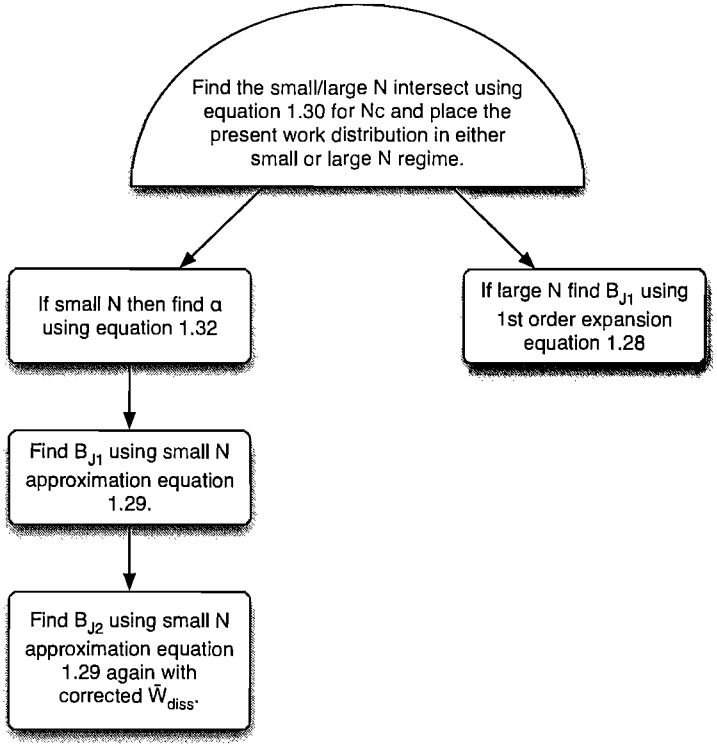


Figure 3.4: Flow diagram of the steps taken to calculate the Gore bias of a Jarzynski estimate.

Gore *et al.* discuss an example using a test system, without sampling noise, where the B_J of Jarzynski estimates with switches displaying a number of $\bar{W}_{diss}$s are calculated using the equations discussed above. These B_Js were compared to the bias found using effective ΔF_∞ (exhaustive numbers of switches) [Gore *et al.*(2003)Gore, Ritort & Bustamante]. The calculated B_Js seem to match with the exhaustive estimates well, although the two start to diverge at the small/large N intersect area as the size of the $\bar{W}_{diss}$ is increased.

In practice, there may be a number of problems with the wider use of the Gore bias correction: The work distributions of large complex systems and perturbations may not be Gaussian at the levels of N that are easily obtainable. The Gore bias correction would possibly be used with faster switches and medium levels of N, which may often be the area where the model works least well. Hence, when a work distribution has a large $\bar{W}_{diss}$ it may be more accurate to use the Gore bias correction with smaller N rather than N which is close to the small/large N intersect. In this study the small N approximation (equations 3.30 and 3.34) will be tested to gauge its suitability for general use. The large N 1st order expansion (equation 3.29) is not used as when this method for estimating the bias is valid the variance is by far the dominant

form of error.

It may be possible to design FG calculation which are well behaved for many systems and display very small levels of bias. However, it is difficult to be sure of the level of bias present in an estimate without having an idea of ΔF_∞. As the Gore bias methodology discussed above shows, the bias of FG calculations behaves in a way which may be statistically tractable and predictable, so it may be possible to predict a well converged estimate.

Wu and Kofke developed a measure of the bias of a Jarzynski estimate based on measuring sampling specifically in the tails of the work distribution called the "neglected tail model" [Lu & Kofke(2001)]. The neglected tail model was originally developed for use with FEP calculations which use many (100,000s) measurements to find the average, $exp(-\beta W)$. This model was extended for general use with FG calculations which use any amount of sampling. The extended model assumes that all of the bias is due to the lack of W contributions below a particular level, W*. The extended neglected tail model was tested on the same Gaussian work model as used by Gore *et al.* and found to have improved agreement with the numerical bias.

Wu and Kofke subsequently defined a measure ($\prod$), for the case of the near equilibrium regime, where $\prod > 0$ when the calculation is fully converged and bias is negligible [Wu & Kofke(2004)].

$$\prod = \sqrt{\mathbf{W}_L\left[\frac{1}{2\pi}(N-1)^2\right]} - \beta\sigma_W. \tag{3.35}$$

Where $\mathbf{W}_L(x)$ is the Lambert W function, defined as the solution for w in $x = wexp\{w\}$. A plot of bias of the Jarzynski estimator against $\prod$ is a curve which meets the point of negligible bias at $\prod = 0$. Again, $\prod$ is susceptible to the generally biased estimate of $\bar{\mathrm{W}}_{diss}$ through σ_W (the standard deviation of the works) of equation 3.35. However, Wu and Kofke attempt to demonstrate that $\prod$ would not take on a value suggesting convergence when convergence had not been achieved and is therefore "fail-safe" [Wu & Kofke(2004)]. It is important to note that this "fail-safe" does not extend to cover the case of non-gaussian work distributions.

Another important idea not considered by equation 3.35 is the fact that the bias is often not symmetric, i.e. it is not same in the forwards and backwards work distributions. Without knowledge of ΔF_∞ or the phase-space relationships of the A and B end states it is difficult to know which direction is more prone to bias. With this in mind Wu and Kofke developed a

method of measuring phase-space relations which they called "relative-entropy measures" [Wu & Kofke(2005a)]. Information theory defines the relative entropy measures S_A and S_B, from each end state as,

$$S_A = \int_\Gamma d\gamma p_A(\gamma) \ln \left[\frac{p_A(\gamma)}{p_B(\gamma)}\right], \tag{3.36}$$

$$S_B = \int_\Gamma d\gamma p_B(\gamma) \ln \left[\frac{p_B(\gamma)}{p_A(\gamma)}\right], \tag{3.37}$$

where Γ is the area of phase space important to the system, i.e. where the system resides almost all of the time, and $p_A(\gamma)$ and $p_B(\gamma)$ are the probability densities for phase space points ($\gamma \epsilon \Gamma$) from the A and B end states. This definition of relative entropy may not be familiar to statistical mechanics and S_A and S_B are not related to the entropies of states A and B, but a connection with the information theory concept has been shown, with a specific example of the non-equilibrium relaxation of a polymer chain system, which is relevant to this work [Qian(2000)].

Equations 3.36 and 3.37 are unfortunately not in an accessible form as the full phase space distributions of systems of interest are not presently tractable. Wu and Kofke expressed equations 3.36 and 3.37 as,

$$S_A = \langle \beta W \rangle_{A \to B} - \beta \Delta F, \tag{3.38}$$

$$S_B = -\langle \beta W \rangle_{B \to A} + \beta \Delta F, \tag{3.39}$$

through the fact that $p_A(\gamma) = e^{-\beta U_A(\gamma)}/Z_A$ and hence γ can be replaced by W in equation 3.37. This new form (equation 3.39) is useful as S_A and S_B are equal to the $\bar{W}_{diss}$ of the respective work distributions. S_A and S_B are equal to zero when the work distributions of switches starting at systems A and B are identical. As the S_A/S_B ratio becomes larger the important phase space sizes of systems A and B become more different, with the system with a lower S having a smaller important phase space. As the size of both S_A and S_B increases, while staying similar relative to each other, the displacement of the two end system phase spaces will also increase. Equations 3.38 and 3.39 have been derived independently by Jarzynski [Jarzynski(2006)].

Using the expressions for $\prod$ and S above a definition for $\prod$ which takes account of asymmetric bias in work distributions in the forwards and backwards directions is,

$$\prod_{A\to B} = \sqrt{\frac{S_A}{S_B}\mathbf{W}_L\left[\frac{1}{2\pi}(N-1)^2\right]} - \sqrt{2S_A}, \tag{3.40}$$

$$\prod_{B\to A} = \sqrt{\frac{S_B}{S_A}\mathbf{W}_L\left[\frac{1}{2\pi}(N-1)^2\right]} - \sqrt{2S_B}. \tag{3.41}$$

Here $\beta\sigma_W$ has been replaced by $\sqrt{2S_A}$ as S_A is equivalent to $\bar{\mathrm{W}}_{diss}$ and a dependency of $\prod$ on $\mathbf{W}_L$ included to account for the phase-space relationship between A and B end states [Wu & Kofke(2005a)].

It may be important to note that only in the case of FEP (instant FG switches) will equations 3.38 and 3.39 connect directly to equations 3.36 and 3.37. However, Wu and Kofke comment that the heuristic in equations 3.41 should generally apply to FG calculations which do not use instant switches [Wu & Kofke(2005b)]. Also, statistical error can affect $\prod$ and to be sure of a bias free estimate, confidence limits of $\prod$ must be checked. For this reason Wu and Kofke suggest using $\prod \geq 0.5$ as a prescription of an accurate estimate.

Again, Wu and Kofke demonstrate the applicability of equation 3.41 with simple one dimensional Gaussian models [Wu & Kofke(2005a)]. The model perturbations with varying phase-space relationships all display the same relation in bias and $\prod$ with negligible bias corresponding to $\prod \geq 0$. Also, logical arguments as to the "fail safe" nature of the $\prod$ measure with regard to the general bias in estimates of S_A and S_B ($\bar{\mathrm{W}}_{diss}$) are presented similar to those discussed above for equation 3.35. An extra measure designed to guard against a possible false positive $\prod \geq 0$, is to swap the ΔF estimates used in producing S_A and S_B as below,

$$S_A = \langle\beta W\rangle_{A\to B} - \beta\Delta F_{B\to A}, \tag{3.42}$$

$$S_B = -\langle\beta W\rangle_{B\to A} + \beta\Delta F_{A\to B}. \tag{3.43}$$

This seemingly arbitrary extra measure gives a larger estimate of S_A and S_B, especially in the case of a large underestimation of S_A or S_B, due to estimator bias, and thus makes it more difficult to obtain a false positive $\prod$. One caveat to the fail safe is noted where N is very low

(less than about 4) and $\prod$ can approach 0. The signal for detection of this error is the rapid decrease of $\prod$ with extra switches.

Central to the idea of microscopic reversibility (equation 3.14) described above and the idea of there being a single work distribution, is the idea that any forward switch has an exact backwards equivalent and *vise versa*, or that switches exist as conjugate pairs related by time reversal. This idea is discussed at length for deterministic dynamics with explanations for most common types of system by Jarzynski [Jarzynski(2006)]. Jarzynski also shows that the dominant switches, important to the exponential average (close to $p(W)e^{-\beta W}$ in figure 3.2) are the conjugate twins of the most common switches, which result in work values close to the $\bar{W}$ peak in the work distribution as shown in equation 3.44 where we use ζ^F_{dom} to denote the dominant switches of the forwards distribution and ζ^F_{com} for the common switches of the backwards distribution.

$$\zeta^F_{dom} = \zeta^R_{com} \tag{3.44}$$

This is nicely illustrated by an ideal gas enclosed within a piston where for the process of pushing the piston into the gas the dominant switches are those where there are no collisions between piston and gas i.e. where the system starts in the type of configuration from which you would expect the process in the opposite direction to start. Further, Jarzynski has shown that the relative contribution of a set of forward switches to the exponential average is equal to the probability of realising the conjugate twin of this switch in the backwards direction and *vise versa.*

These ideas are important when thinking about how many switches may be needed to obtain a converged ΔF estimate and which estimator is most efficient for a particular calculation. To obtain a converged estimate of ΔF from the Jarzynski estimator, dominant switches must be sampled from the region close to $p(W)e^{-\beta W}$. Using equation 3.15 the probability that a sampled forwards switch is the the region ζ^F_{dom} is,

$$P(\mathbf{Z}_{dom}) = \int_{\zeta^F_{dom}} d[\mathbf{Z}_F|\lambda_t]P[\mathbf{Z}_F|\lambda_t] = \int_{\zeta^R_{com}} d[\mathbf{Z}_R|\lambda_t]P[\mathbf{Z}_R|\lambda_t]exp\{-\beta W_{diss}[\mathbf{Z}_R|\lambda_t]\}. \tag{3.45}$$

Then as ζ^R_{com} constitutes almost all of the backwards work distribution, $P[\mathbf{Z}_R|\lambda_t]$, we can say,

$$P(\mathbf{Z}_{dom}) \sim exp\{-\beta \bar{W}_{diss}[\mathbf{Z}_R|\lambda_t]\} \int_{\zeta^R_{com}} d[\mathbf{Z}_R|\lambda_t] P[\mathbf{Z}_R|\lambda_t] \sim exp\{-\beta \bar{W}_{diss}[\mathbf{Z}_R|\lambda_t]\}. \quad (3.46)$$

Thus, a good estimate of the number of switches to obtain an example of ζ^F_{dom} is,

$$N^F_{dom} = P(\mathbf{Z}_{dom})^{-1} \sim exp\{+\beta \bar{W}_{diss}[\mathbf{Z}_R|\lambda_t])\}, \quad (3.47)$$

and similarly in for the backwards calculation,

$$N^R_{dom} \sim exp\{+\beta \bar{W}_{diss}[\mathbf{Z}_F|\lambda_t])\}. \quad (3.48)$$

This suggests that as a rule of thumb the number of switches needed for convergence varies exponentially with the average work dissipated. This is very similar to the relationship discussed above as part of the Gore bias [Gore *et al.*(2003)Gore, Ritort & Bustamante]. However, here it is $\bar{\mathrm{W}}$ in the opposing direction to that of interest which is related to the convergence i.e. the work dissipated in the backward direction determines the convergence of the forwards calculation.

For the case of work distributions where the bias is not symmetric in both directions, as discussed above, this suggests that the direction which has the largest $\bar{W}_{diss}$ will converge most quickly. Thus, this reasoning of Jarzynski agrees with the relative entropy measure approach of Wu and Kofke discussed above which has similar conclusions with regards convergence and $\bar{W}_{diss}$ of the work distribution in the opposite direction. Although this idea is counterintuitive it can be easily rationalised through studying the relative sizes and positions of forwards and backwards work distribution with regard to ΔF. Figure 3.5 shows the forward and backward work distributions for a single free energy difference where the $\bar{W}_{diss}$ is smaller for the backward compared to the forward distribution. The forwards distribution is wider than the backwards, which is much taller. Also, owing to equation 3.44 ζ^F_{dom} is closer to ΔF than ζ^R_{dom}. Consequently the probability that a forward switch is dominant is higher than the probability that a backward switch is dominant, as dominant switches are found deeper in the tail of the backwards distribution. Hence, the switching direction with the highest $\bar{W}_{diss}$ (forwards direction for the case of figure 3.5) will converge to an accurate result with fewer switches than the alternative direction.

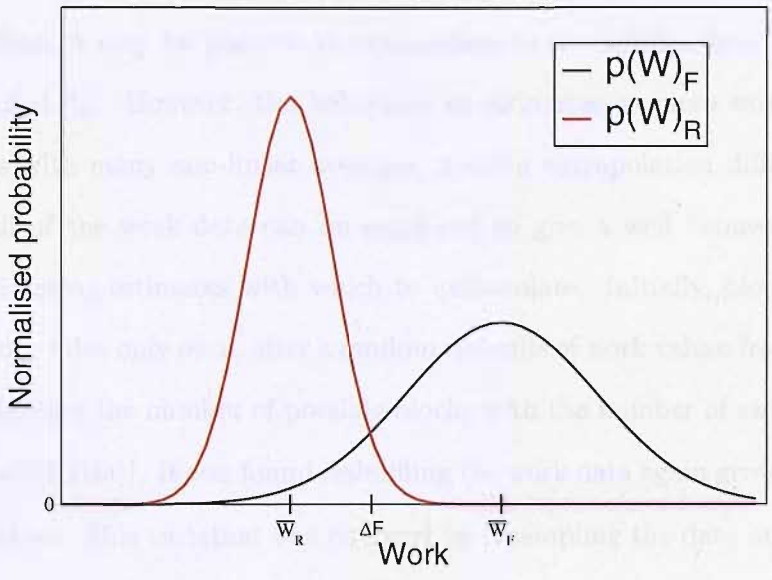


Figure 3.5: Representation of forwards (black) and backwards (red) work distributions where the backwards distribution ($p(W)_R$) has a smaller $\bar{W}_{diss}$ than the forwards distribution than the forwards distribution. The ΔF is found at the point where the two distributions meet.

Importantly, the findings of Jarzynski discussed here and the relative entropy measures of Wu and Kofke discussed above are not limited to Gaussian work distributions as are the Gore bias and Kofke bias measures [Gore *et al.*(2003)Gore, Ritort & Bustamante], [Jarzynski(2006)], [Wu & Kofke(2005a)]. Therefore, it may be that relative entropy measures are more generally applicable to real FG calculations than the Kofke bias measures. By simpily calculating the $\bar{W}_{diss}$ of the forwards and backwards work distributions it may be possible to pick the most accurate estimate in the case that BAR is not the most accurate. However, as discussed for the Kofke bias measures, estimates of $\bar{W}_{diss}$ found through calculations where an exhaustive protocol is not possible are prone to bias. This bias may cause the prediction of the most efficient switching direction to be incorrect when forwards and backwards distributions have relatively similar amounts of bias while still being different.

3.3.4 Extrapolation methods

One of the earliest FG methods attempted to remove the sampling bias (equation 3.25) of standard fast switching calculations through block averaging and extrapolation [Zuckerman & Woolf(2002a)]. In essence this is a simple idea. As more and more switches are added to a work distribution, a Jarzynski estimate will become more and more accurate. Thus, given a certain

amount of work data, it may be possible to extrapolate to the infinite data limit and obtain a good estimate of ΔF_∞. However, the behaviour of estimates as more work data is added is often erratic, as with many non-linear averages, making extrapolation difficult. Therefore, block averaging all of the work data can be employed to give a well behaved monotonically decreasing set of running estimates with which to extrapolate. Initially, block averaging was done using each work value only once, after a random reshuffle of work values from the order they were performed, limiting the number of possible blocks with the number of switches perfromed [Zuckerman & Woolf(2002a)]. It was found reshuffling the work data again gives rise to variation in ΔF estimate curves. This variation was removed by resampling the data and increasing the numbers of blocks used for each ΔF estimate [Ytreburg & Zuckerman(2004)].

Ytreburg and Zuckerman (2004) define their Jarzynski estimate block averaging scheme with N_{tot} work values, N_{blk} work values in each block and m blocks as,

$$\Delta F_{BlkAv} = \frac{1}{m} \sum_{i=1}^{N_{tot}/N_{blk}} -k_B T \ln \langle exp\{-\beta W\} \rangle_i \,. \tag{3.49}$$

Work values are drawn at random and not replaced until a block is complete, so m can be made $\sim 100 \times N_{tot}/N_{blk}$. Then ΔF_{BlkAv} is simply averaged over the number of blocks m to produce an estimate of ΔF for a particular N_{blk} termed here as ΔF_N.

Extrapolation of these block averaged free energy estimates has been carried out with two methods [Ytreburg & Zuckerman(2004)]. Both methods start by plotting ΔF_N as a function of χ where,

$$\chi = \frac{1}{n^\tau}. \tag{3.50}$$

τ must be picked from the range $0.3 < \tau < 0.7$, all of which produce useable results. The choice of τ controls the spread of the data in a plot; decreasing τ causes the data to be confined to a small region near to $\chi = 1$ with a large distance to extrapolate to $\chi = 0$, increasing τ leads to the spread of the data becoming larger and the slope of the tail can become uncertain. Ytreburg and Zuckerman (2004) find a $\tau = 0.5$ is optimum for linear extrapolation with their test systems (discussed below). The limit $N_{tot} = \infty$ is now found at $\chi = 0$.

The first method is a simple linear extrapolation method where after plotting ΔF_N against

χ the data is extended with a straight line using only the small χ tail to $\chi = 0$. Thus, another choice must be made over the extent of the small χ tail. Ytreburg and Zuckerman (2004) use a small χ tail which is 1/5 of the data, as using more introduced bias into any extrapolated estimate and any less resulted in similar estimates as for a χ tail of 1/5, but with greater uncertainty.

The most precise ΔF_N values are always found with small N_{blk} ($\chi \approx 1$) as these estimates are averaged over the most blocks. This being the case it may be a problem that the linear extrapolation method described above uses only the most uncertain ΔF_N estimates at small χ to extrapolate to $\chi = 0$. Therefore, Ytreburg and Zuckerman (2004) devised an alternative extrapolation scheme where all estimates across χ are integrated into an extrapolation.

A cumulative integral estimate (CI) of $\Delta F_N(\chi = 0)$ is given by,

$$CI(\chi) \equiv \int_{\chi}^{1} d\chi' \left(\Delta F_N(\chi') - (1 - \chi') \frac{d\Delta F_N(\chi')}{d\chi'} \right). \tag{3.51}$$

$CI(\chi)$ is started from $CI(\chi = 1)$ and accumulated across a plot to $CI(\chi = 0)$ where $CI(\chi) = \Delta F$. Also the derivative in equation 3.51 is found by numerical methods. τ is chosen differently for this CI extrapolation scheme compared to the previous linear method. Here τ is not fixed but is chosen for each extrapolation process to minimise the slope of the tail of the small χ tail of $CI(\chi)$. Again the small χ tail of $CI(\chi)$ is 1/5 of the data.

Ytreburg and Zuckerman (2004) tested their linear and CI extrapolation methods on five test systems of increasing complexity: a perturbation between harmonic oscillator potentials, a chemical potential calculation with a Lennard-Jones fluid, the growth of a chloride ion in water, a perturbation from methanol to ethanol in water, and a stearic to palmitic acid perturbation in water. Jarzynski estimates and estimates using linear and CI extrapolation are made for each test system with estimates of ΔF, found from subsets of the total data, averaged over 500 repetitions for a range of N_{tot} from 10 to 1000. This range of estimates are compared to Jarzynski estimates found using all work values available and ΔFs found using various equilibrium methods.

For all systems these extrapolation methods are found to give more accurate estimates of ΔF than the Jarzynski estimator for small N_{tot}. For the less complex systems (harmonic oscillator potentials and the Lennard-Jones fluid) Jarzynski and both extrapolation methods converge

to the similar ΔFs which are the same as independent estimates found exhaustively by other methods. For the more complex systems (growth of a chloride ion in water, methanol to ethanol in water and stearic to palmitic acid in water), at $N_{tot} = 1000$, linear extrapolation estimates find similar ΔFs to the exhaustive Jarzynski estimates. However, the independent, exhaustive ΔFs are different due to bias in the exhaustive Jarzynski estimates. CI extrapolation seems to give estimates closer to these independent, exhaustive ΔFs. The authors suggest that the CI extrapolation method is able to see beyond the small amounts of work data it is given. They also compare the amount of work data needed for ΔF estimates to within 2 kcal.mol^{-1} and find that CI extrapolation is 5-40 times more efficient.

Although the CI extrapolation method seems to offer improved accuracy it also displays increased levels of statistical uncertainty compared to using the Jarzynski estimator alone. The standard deviation of estimates from the 500 repeats more than doubled for CI extrapolation compared to the Jarzynski estimator (from $\approx$ 1 to$\approx$ 3 kcal.mol^{-1}) for the palmitic to stearic acid test system, where the difference in accuracy was $\approx$ 2 kcal.mol^{-1}. Also, these 500 repeats were made by drawing small numbers of work values at random from the large number of works available for each system. This means that equilibrium starting configurations used to produce work values in any repeat are spread over a much larger area of phase space than would be the case when using these methods for real calculations where the large number of work values is not available. When using these methods it would make no sense to produce many work values and then only use a small subset for any subsequent calculations. It is not clear how well these extrapolation methods would compare to Jarzynski when used with using small numbers of work values produced from equilibrium starting configurations which are produced in sequence.

Another issue not mentioned in this study of extrapolation methods is the effects of non-symmetric bias in forwards and backwards work distributions [Ytreburg & Zuckerman(2004)]. Especially for the more complex test systems, work distributions may be more efficient in one direction, or together when used with BAR.

3.3.5 Rosenbluth FG sampling

Wu and Kofke recently introduced a set of FG methods which attempt to improve the sampling of the rare, important switches which are close to p(W)e$^{-\beta W}$ in figure 3.2 by applying a bias

to the calculation [Wu & Kofke(2005c)]. The name Rosenbluth FG sampling is taken from the Rosenbluth-sampling methods [Rosenbluth & Rosenbluth(1955)] used in MC simulations of lattice based polymeric chains. The idea of breaking an extremely unlikely event into smaller more likely events is applied to producing an FG switch which has a work value close to the highly weighted $p(W)e^{-\beta W}$ peak.

The structure of a switch within the original FG sampling method has a predefined set of uniformly spaced λ values starting at 0 and ending at 1. A predefined number of MC trials are carried out at each λ value before incrementing to the next and adding the work performed in that λ increment to the total work for the switch (equation 3.2). However, in the course of switching from systems A to B the work performed at each λ increment is often not uniform. Depending on the perturbation, often particular areas of a switch are more or less prone to producing large work values. When a high work is probable, it is possible that the size of the work can be lowered by particular favourable system configurations or by using very small λ increments. However, using very small λ increments throughout an FG calculation may be seen as inefficient and favourable system configurations cannot be relied upon as they must be chosen at random from the Boltzmann distribution.

Wu and Kofke applied Rosenbluth sampling to bias the choice of both the size of individual λ increments and the configuration the system takes up while undergoing a λ increment, in order that the work performed for each λ increment of a switch is low. Three methods were descibed: λ bias FG which biases the size of each λ increment to minimise the work performed, configuration bias FG which biases the configuration the system takes up when undergoing a λ increment, and a hybrid bias FG which does both.

With λ bias FG the λ value to which the system is incremented, is found from a continuum weighted by the potential. This continuum is structured in such a way that λ cannot go backwards with $\lambda_i \in [\lambda_{i-1}, a_i]$ where a_i is a predefined set of incremental constants i.e.

$$0 < a_1 \leq a_2 \cdots \leq a_n - 1 = 1. \tag{3.52}$$

Thus, for each λ increment the new λ (λ_i) is found from a continuum between the current λ (λ_{i-1}) and the next a value (a_i). This allows a variation in the maximum that λ_i can take up.

The selection of λ_i according to the potential of the current configuration ($U_{\lambda_i}(q_{i-1})$ with

probability density

$$P_{q_{i-1}}(\lambda_i) = \frac{1}{R_i(q_{i-1};\lambda_{i-1})} p_i(\lambda_i) e^{-\beta\alpha_K U_{\lambda_i}(q_{i-1})}, \tag{3.53}$$

where $R_i(q_{i-1};\lambda_{i-1})$ is the Rosenbluth weight, which ensures a normalised probability:

$$R_i(q_{i-1};\lambda_{i-1}) = \int_{\lambda_{i-1}}^{a_i} p_i(\lambda) e^{-\beta\alpha_K U_\lambda(q_{i-1})} d\lambda. \tag{3.54}$$

Here α_K is a predefined constant designed to control the influence of the potential of the present configuration ($U_{\lambda_i}(q_{i-1})$) on the weighting of the probability density $P_{q_{i-1}}(\lambda_i)$. Also, $p_i(\lambda)$ is another weighting term designed to control the size of the potential weight from outside the exponential term. It is possible that the form of these two weighting terms rely on i and are different for each λ increment, although this must be predefined and not depend on $U_{\lambda_i}(q_{i-1})$. As there are a predefined number of λ increments (n), the final λ increment must be made such that $\lambda_n = 1$ and hence cannot be weighted.

Each λ bias FG switch will be slightly different depending on the configurations which the system takes up for each λ increment. Because the probability density applied to find each new λ value in each switch is different, the resulting work values do not give an average which relates directly to the free energy difference. The work performed must be modified to account for the specific weighting of each λ bias switch as follows,

$$\beta W(\lambda_{i-1} \rightarrow \lambda_i) = \begin{cases} \beta(1-\alpha_K)U_{\lambda_i}(q_{i-1}) - \beta U_{\lambda_{i-1}}(q_{i-1}) - ln[R_i(q_{i-1};\lambda_{i-1})/I_i(\lambda_{i-1})] & if \quad 1 \leq i < n \\ \beta U_{\lambda_n}(q_{n-1}) - \beta U_{\lambda_{n-1}}(q_{n-1}) & if \quad i = n. \end{cases} \tag{3.55}$$

In equation 3.55 the $I_i(\lambda_{i-1})$ term is defined as,

$$I_i(\lambda_{i-1}) = \int_{\lambda_{i-1}}^{a_i} d\lambda p_i(\lambda) \tag{3.56}$$

to account for the ideal-gas normalisation for switching λ from λ_{i-1} to λ_i. Wu and Kofke also demonstrated that this definition of the work was consistent with the Jarzynski equality (equation 3.6) by defining ΔF in terms of these λ values chosen from a distribution [Wu & Kofke(2005c)].

There are various options in defining the parameters α_K, a_i and $p_i(\lambda)$. The α_K parameter should in essence be varied depending on the general size of the system potential. If the potential is very large it would bias λ increments to be very small, causing a large final λ increment to complete a switch and inevitably large total work values. The α_K parameter should be less than one so that it reduces the effect of the potential on the probability density $P_{q_{i-1}}(\lambda_i)$. Wu and Kofke use,

$$\alpha_K = 1/N_\lambda, \tag{3.57}$$

where N_λ is the "number of atoms or particles involved in the difference between the A and B systems". This came from direct investigation of the use of α_K values from 0 to 1 on four quite different test systems, each of ten independent harmonic oscillators ($N_\lambda = 10$). A minimum of inaccuracy was found in each of the four cases which corresponds to using α_K defined by equation 3.58. Although equation 3.58 is the best definition of α_K for these harmonic oscillator systems, it is probable that a quite different definition may be required for large bio-systems with thousands of atoms, of which only a very small fraction are perturbed in changing from end point systems A to B.

The obvious choice of a_i is to have each a value equal to 1. This would allow each new λ to be any value from the present value to 1. In the case that the potential does not react to a large increase in λ this would allow the switch to proceed to towards $\lambda = 1$ very quickly which would be desirable as it would avoid the situation of have a large forced final λ increment. Unfortunately, this situation can cause large work values as the subsequent small λ increments produce very small Rosenbluth weights (equation 3.54) and the corrected work contributions would not be small. Depending on how it is defined, the effect discussed here can be controlled by the I_i term. Wu and Kofke found the definition,

$$a_i = i/(n-1), \tag{3.58}$$

for a_i preferable to the above for their test systems. Equation 3.58 gives the process an upper bound for each λ increment and prevents initial λ increments from being too large.

Wu and Kofke used $p_i(\lambda) \equiv 1$ in all their presented results [Wu & Kofke(2005c)]. This means that no extra weight, except that involving $U_\lambda(z_{i-1})$, is given to the selection of λ_i.

Wu and Kofke discussed two other possibilities which although interesting were not considered useful and will not be discussed here.

It is worth discussing the practical implementation of λ bias algorithm, carried out as part of this study, as it is not simple and there may be different possibilities not discussed by Wu and Kofke. Figure 3.6 is a flow diagram describing one FG switch of the λ bias algorithm.

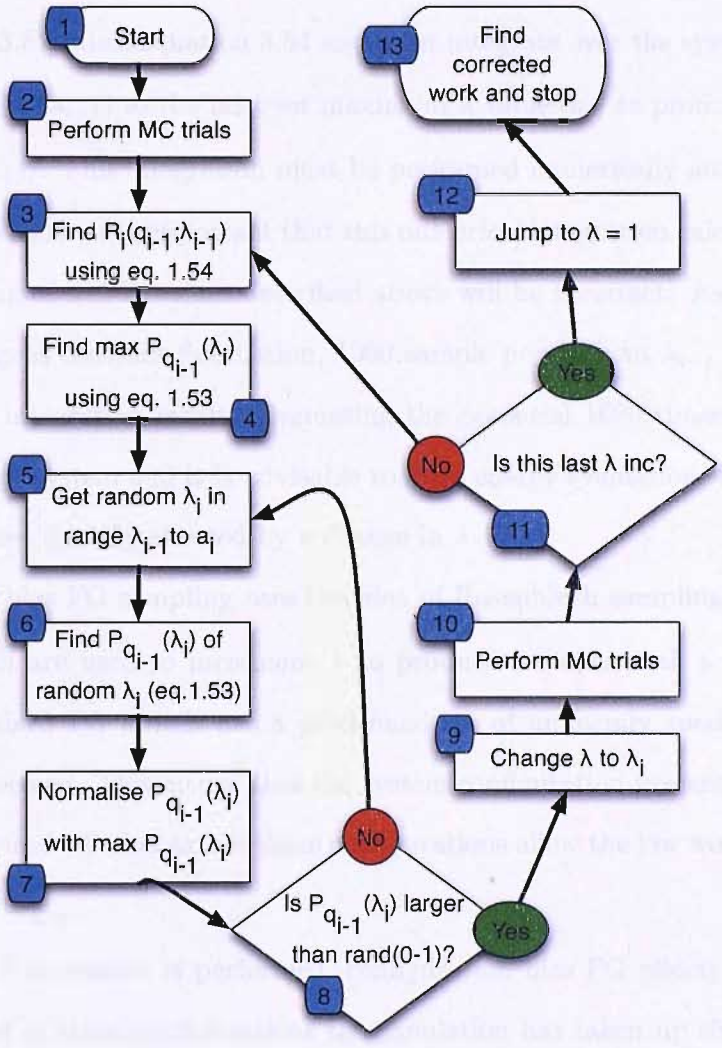


Figure 3.6: Flow diagram describing the λ bias algorithm. Step numbers are marked in light blue.

λ bias FG attempts to find a new λ value from a distribution dictated by the probability density $P_{q_{i-1}}(\lambda_i)$. In practice, sampling from this distribution can be achieved by: Taking a random λ_i from λ_{i-1} to a_i (step 5 figure 3.6) and accepting or rejecting it through comparison of the probability density $P_{q_{i-1}}(\lambda_i)$ to a random value (rand(0-1) in step 8, figure 3.6). If $P_{q_{i-1}}(\lambda_i)$ is less than rand(0-1) it is rejected and a new λ_i is selected. This is repeated until a λ_i

is accepted and this value then becomes λ_{i-1} (step 9). As $P_{q_{i-1}}(\lambda_i)$ is a probability density it must be normalised in some way such that different calculated $P_{q_{i-1}}(\lambda_i)$ values are equivalent. This can be done either by selecting rand(0-1) from the range 0 to $P_{q_{i-1}}(\lambda_i)$ or by finding the maximum value for $P_{q_{i-1}}(\lambda_i)$ through equation 3.53 (step 4 in figure 3.6) and using this to normalise the present $P_{q_{i-1}}(\lambda_i)$ so it will be in the range 0 to 1 and then also finding rand(0-1) in the range 0 to 1.

Step 3 in figure 3.6 utilises equation 3.54 and must integrate over the system potential from the current value of λ (λ_{i-1}) to the next set maximum λ value (a_i) to produce the Rosenbluth weight ($R_i(q_{i-1};\lambda_{i-1})$). This integration must be performed numerically and in this study the trapezoidal rule was used. It is important that this numerical integration calculation is accurate otherwise the biasing of new λ values described above will be incorrect. As the potential of a system often undergoes constant fluctuation, 1000 sample points from λ_{i-1} to a_i were used to ensure an accurate integration result. Evaluating the potential 1000 times can be extremely demanding on a large system and it is advisable to limit energy evaluations for this calculation to the forcefield terms directly affected by a change in λ.

Configurational bias FG sampling uses the idea of Rosenbluth sampling to bias the use of configurations which are used to increment λ to produce switches with a lower $\bar{W}_{diss}$. The structure of a standard FG switch has a predefined set of uniformly spaced points where λ increments are performed. This means that the system configuration present at each of these λ increment points is used whether or not these configurations allow the low work values preferred in producing a low $\bar{W}_{diss}$.

At the point a λ increment is performed, configuration bias FG selects a system configuration from a subset of those configurations the simulation has taken up since the previous λ increment. The choice of configuration is biased to one which produces a low work value when the λ increment is performed. Thus, the switch produced via configuration bias may be more important to the exponential average, $exp(-\beta W)$, in equation 3.6.

The configuration to be used in a λ increment is selected from a set of m taken at uniform intervals from the simulation since the previous λ increment according to,

$$P_{\lambda_i}(q_{i-1}) = \frac{1}{R_i(\lambda_i)} exp\{-\beta f[U_{\lambda_i}(q_{i-1})]\}. \tag{3.59}$$

In equation 3.59 the Rosenbluth weight defined as,

$$R_i(\lambda_i) = \sum_{j=1}^{m} exp\{-\beta f[U_{\lambda_i}(q_{i-1,j})]\}. \tag{3.60}$$

The term $f[U_{\lambda_i}(q_{i-1,j})]$ is a function of the potential for which Wu and Kofke list two possible options: Configuration bias-c, where as with λ bias, $f[H_{\lambda_i}(z_{i-1,j})] = \alpha_K H_{\lambda_i}(z_{i-1})$ and α_K has the same definition as for λ bias (equation 3.58). Configuration bias-d where $f[H_{\lambda_i}(z_{i-1,j})] = H_{\lambda_i}(z_{i-1}) - H_{\lambda_{i-1}}(z_{i-1})$ and is simply the work incurred in performing a λ increment.

Again the definition of the work performed on each switch must be modified to account for the differences in internal structure as follows,

$$\beta W(\lambda_{i-1} \rightarrow \lambda_i) = \beta U_{\lambda_i}(q_{i-1}) - \beta U_{\lambda_{i-1}}(q_{i-1}) - \beta f[U_{\lambda_i}(q_{i-1,j})] - \ln[R_i(\lambda_i)/m]. \tag{3.61}$$

With configuration bias-d equation 3.61 collapses to contain only the term containing the Rosenbluth Weight.

Figure 3.7 is a flow diagram describing a single switch of configurational bias FG.

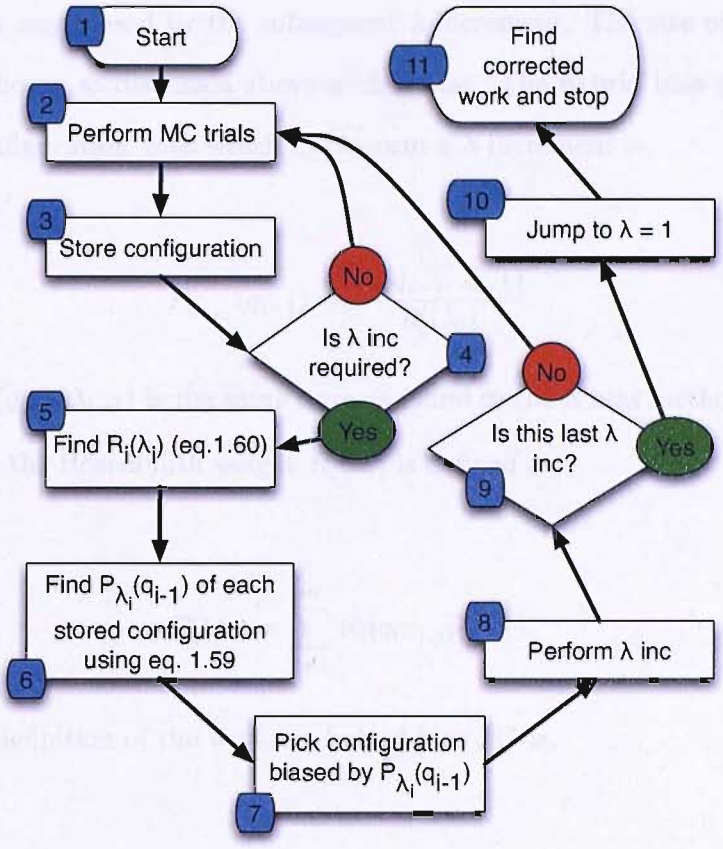


Figure 3.7: Flow diagram describing the configuration bias algorithm. Step numbers are marked in light blue.

The configuration bias algorithm should in theory be less computationally demanding that the λ bias algorithm described above as it does not involve a computationally expensive numerical integration. However, as the whole system configuration must be stored (step 3 in figure 3.7) m times, m must be limited by available memory. A good check on correctness of the algorthm is that the sum of $P_{\lambda_i}(q_{i-1})$s far all stored configurations for a λ increment should be one. For this implementation the biased selection of stored configurations (step 7 in figure 3.7) was performed by summing $P_{\lambda_i}(q_{i-1})$ for each stored configuration and which ever part of this sum a random number from 0 to 1 falls in, is chosen.

As recognised by Wu and Kofke the present λ bias algorithm has limitations for systems with small or no phase space overlap [Wu & Kofke(2005c)]. If a λ bias switch has barriers to sampling after the initial stage's, λ increments will be small and the final forced increment will incur large amounts of work. Wu and Kofke (2005a) attempt to alleviate this problem to some degree through a hybrid of both λ and configuration bias.

Hybrid bias FG is organised so that first a number of system configurations are generated

and one chosen in a way biased by the subsequent λ increment. The size of the subsequent λ increment is then chosen as discussed above with λ bias. The hybrid bias probability density for selecting the configuration with which to perform a λ increment is,

$$P'_{\lambda_{i-1}}(q_{i-1}) = \frac{R_i(q_{i-1};\lambda_{i-1})}{R'_i(\lambda_i)}. \tag{3.62}$$

In equation 3.62 $R_i(q_{i-1};\lambda_{i-1})$ is the same term as found in the λ bias method described above (equation 3.54) and the Rosenbluth weight $R'_i(\lambda_i)$ is defined as,

$$R'_i(\lambda_i) = \sum_{j=1}^{m} R_i(q_{i-1,j};\lambda_{i-1}). \tag{3.63}$$

Then the modified definition of the work for hybrid bias FG is,

$$\beta W(\lambda_{i-1} \rightarrow \lambda_i) = \begin{cases} \beta(1-\alpha_K)U_{\lambda_i}(q_{i-1}) - \beta U_{\lambda_{i-1}}(q_{i-1}) - \ln\{[R'_i(\lambda_{i-1})/mI_i(\lambda_{i-1})]\} & if\ 1 \leq i < n \\ \beta U_{\lambda_n}(q_{n-1}) - \beta U_{\lambda_{n-1}}(q_{n-1}) - \beta f[U_{\lambda_n}(q_{n-1})] - \ln\{R_n(\lambda_n)/m\} & if\ i = n. \end{cases} \tag{3.64}$$

Wu and Kofke only consider the use of hybrid bias with the configuration bias-c definition of $f[U_{\lambda_n}(q_{n-1})]$. This is because the α_K parameter was shown to be important to λ bias, and hybrid bias mainly uses the parameters of λ bias.

Thus, Wu and Kofke have provided three Rosenbluth FG sampling algorithms which, with various parameter options, give a number of possible methods [Wu & Kofke(2005c)]. The six possible methods discussed above and listed in table 3.1 were investigated using four harmonic oscillator test systems (test systems discussed in the next chapter). The Gaussian nature of these harmonic oscillator systems allowed the analytical production of equilibrated configurations at each step. FG switches produced with this method of sampling are still nonequilibrium, due to the discrete steps in λ, but the nonequilibrium effects of Hamiltonian lag are minimised, modelling the effects of large amounts of sampling between λ increments. They found that λ bias has increased efficiency compared to original FG when applied to systems with a reasonable amount of phase space overlap and where the size of phase space important to the starting system is much larger than the destination system. Configurational bias-d FG was found to

give improved accuracy and efficiency compared to original FG when there is less overlap of end point systems A and B.

Wu and Kofke then investigated four of these methods (2, 3, 4, 6) using standard metropolis MC with the same four harmonic oscillator test systems (discussed in the next chapter). In this study, these same methods will be investigated with similar systems and a larger water-methane system, with protocols designed to assess their suitability for calculations with larger bio-systems. These investigations and discussions of these systems start in chapter 3.

Method number	FG sampling method	a_i	$f[U_{\lambda_n}(q_{n-1})]$
1	λ bias	1	
2	λ bias	$i/(n-1)$	
3	configuration bias-c		$\alpha_K U_{\lambda_i}(w_{i-1})$
4	configuration bias-d		$U_{\lambda_i}(q_{i-1}) - \beta U_{\lambda_{i-1}}(q_{i-1})$
5	hybrid bias	1	$\alpha_K U_{\lambda_i}(q_{i-1})$
6	hybrid bias	$i/(n-1)$	$\alpha_K U_{\lambda_i}(q_{i-1})$

Table 3.1: Parameters for six Rosenbluth FG sampling methods investigated by Wu and Kofke [Wu & Kofke(2005c)]. In all methods except method 4 Wu and Kofke used $\alpha_K = 1/N_\lambda$.

3.3.6 FG path sampling

As discussed above and in the literature the major obstacle to efficient use of FG calculations is the bias of ΔF estimates due to the non-linear nature of the exponential work average in equation 3.6 [Gore *et al.*(2003)Gore, Ritort & Bustamante], [Zuckerman & Woolf(2002b)], [Hummer(2001)], [Shirts & Pande(2005)]. In contrast the TI based methods discussed above do not experience this bias as they do not involve any non-linear average (discussed [Shirts & Pande(2005)]). This advantage of TI over FG based methods is the motivation for another set of methods discussed here.

Sun (2003) derived an expression for ΔF which utilises FG switches but is similar to TI based methods in that it does not contain a non-linear average. This derivation starts by defining a function $f(\alpha_S)$ where α_S is a weighting variable, using equation 3.8 above, such that,

$$f(\alpha_S) = \frac{\int d[\mathbf{Z}_F]\ P[\mathbf{Z}_F|\lambda_t] exp\{-\alpha_S \beta W\}}{\int d\mathbf{z}_0\ P(\mathbf{z}_0)}. \tag{3.65}$$

where $\int d[\mathbf{Z}_F]$ is an integration over all possible switches from systems A to B and $P(\mathbf{z}_0)$ is the probability of a switch starting configuration $\mathbf{z}_0$ i.e. the ensemble probability of a configuration for system A ($P(\mathbf{z}_0) = exp\{-\beta U_A(q_N)\}$). According to equation 3.65 the ΔF is found when $\alpha_S = 1$. Thus the derivative of $f()$ with respect to α_S is,

$$\frac{\delta f}{\delta \alpha_S} = -\beta \frac{\int d[\mathbf{Z}_F] \;\; P[\mathbf{Z}_F|\lambda_t] W exp\{-\alpha_S \beta W\}}{\int d[\mathbf{Z}_F] \;\; P[\mathbf{Z}_F|\lambda_t] exp\{-\alpha_S \beta W\}} \times \frac{\int d[\mathbf{Z}_F] \;\; P[\mathbf{Z}_F|\lambda_t] exp\{-\alpha_S \beta W\}}{\int d\mathbf{z}_0 \;\; P(\mathbf{z}_0)}. \tag{3.66}$$

In the more concise notation style of equation 3.6 (Jarzynski equality) we can denote:

$$\langle W \rangle_{\alpha_S} = \frac{\int d[\mathbf{Z}_F] \;\; P[\mathbf{Z}_F|\lambda_t] W exp\{-\alpha_S \beta W\}}{\int d[\mathbf{Z}_F] \;\; P[\mathbf{Z}_F|\lambda_t] exp\{-\alpha_S \beta W\}}. \tag{3.67}$$

The derivative (equation 3.66) above can now be written

$$\frac{\delta f}{\delta \alpha_S} = -\beta \langle W \rangle_{\alpha_S} f(\alpha_S). \tag{3.68}$$

Solving equation 3.68 for $f(\alpha_S)$ gives

$$f(\alpha_S) = exp\left\{-\beta \int_0^{\alpha_S} d\alpha'_S \langle W \rangle_{\alpha'_S}\right\}. \tag{3.69}$$

Now the free energy difference can be expressed as:

$$\Delta F = \int_0^1 d\alpha_S \langle W \rangle_{\alpha_S}. \tag{3.70}$$

Equation 3.70 is a result from which ΔF can be found with a linear average which experiences no internal bias. This linear average is over a new distribution weighted by the work and extra weighting variable α_S ($exp\{-\alpha_S \beta W\}$). Thus the sampling employed to accumulate this average must sample switches such that works of a particular value have the probability $exp\{-\alpha_S \beta W\}$.

Another comparison can help understanding of equation 3.70 and the relevance of the α_S parameter: When switches are made instantly the Jarzynski equality becomes the FEP method described in the previous chapter. In the case of equation 3.70 instant FG switches produce the TI method where α_S replaces λ,

$$\lim_{t=0} \langle W \rangle_{\alpha_S} = \left\langle \frac{\delta H(\alpha_S)}{\delta \alpha_S} \right\rangle_{\alpha_S} \tag{3.71}$$

With the advent of a method of calculating ΔF from a straight average of FG switches, it is important to understand the possible reasons, if any, for using finite switches over instant switches. When evaluating ΔFs of complex bio-systems, which is the ultimate focus of this study, barriers between regions of configurational space, due to hard potentials, can hinder the sampling of the necessary areas. The same is true of switch space, possibly even more so. The longer the switches employed between systems A and B, the easier it becomes to reach all areas of switch space. Also, as shown explicitly by Sun (2003) the variance of $\langle W \rangle_{\alpha_S}$ would be expected to increase with shorter switches.

The actual comparison between FG path sampling (FGPS) and TI as discussed in the previous chapter is not as simple as discussed here and by Sun (2003). It should be made clear that although equation 3.70 collapses to equation 3.71 the α parameter is distinct from λ and hence different parameters are being integrated in FGPS and TI.

A similar FG path sampling approach was developed subsequently by Ytreburg and Zuckerman (2004b) where only a single ensemble average must be converged to obtain ΔF, called single ensemble path sampling (SEPS). Here a distribution is sampled from all possible switches which weights the selection of switches with low work values such that the work average can be related to ΔF. In deriving their methods Ytreburg and Zuckerman (2004b) consider the same ratio seen in equation 3.8 where $P(\mathbf{z}_0)$ is brought into the probability of a switch, $P[\mathbf{Z}_F|\lambda_t]$, giving,

$$exp\{-\beta \Delta F\} = \frac{\int d[\mathbf{Z}_F] \;\; P[\mathbf{Z}_F|\lambda_t] exp\{-\alpha_S \beta W\}}{\int d[\mathbf{Z}_F] \;\; P[\mathbf{Z}_F|\lambda_t]}. \tag{3.72}$$

This can be rearranged by introducing a new, as yet undefined, work weighted distribution $D[\mathbf{Z}_F|\lambda_t]$ giving,

$$exp\{-\beta \Delta F\} = \frac{\int d[\mathbf{Z}_F] \;\; D[\mathbf{Z}_F|\lambda_t] \left(P[\mathbf{Z}_F|\lambda_t]/D[\mathbf{Z}_F|\lambda_t]\right) exp\{-\beta W\}}{\int d[\mathbf{Z}_F] \;\; D[\mathbf{Z}_F|\lambda_t] \left(P[\mathbf{Z}_F|\lambda_t]/D[\mathbf{Z}_F|\lambda_t]\right)}, \tag{3.73}$$

$$= \frac{\sum^{D[\mathbf{Z}_F|\lambda_t]} \;\; P[\mathbf{Z}_F|\lambda_t] exp\{-\beta W\}/D[\mathbf{Z}_F|\lambda_t]}{\sum^{D[\mathbf{Z}_F|\lambda_t]} \;\; P[\mathbf{Z}_F|\lambda_t]/D[\mathbf{Z}_F|\lambda_t]}, \tag{3.74}$$

where $\sum^{D[\mathbf{Z}_F|\lambda_t]}$ signifies a sum over all switches produced according to $D[\mathbf{Z}_F|\lambda_t]$. The work weighted distribution $D[\mathbf{Z}_F|\lambda_t]$ is then defined as,

$$D[\mathbf{Z}_F|\lambda_t] = P[\mathbf{Z}_F|\lambda_t]exp\{-\beta/2W\}. \tag{3.75}$$

Ytreburg and Zuckerman (2004b) found this choice of $D[\mathbf{Z}_F|\lambda_t]$ to be optimal and did try others. With equation 3.75, equation 3.74 collapses to give an expression for ΔF:

$$\Delta F = \frac{1}{\beta} \ln \left\{ \sum^{D[\mathbf{Z}_F|\lambda_t]} exp\{-\beta/2W\} / \sum^{D[\mathbf{Z}_F|\lambda_t]} exp\{+\beta/2W\} \right\}. \tag{3.76}$$

As long as equation 3.76 is correct SEPS should be more efficient than original FGPS as only one average needs to be evaluated.

The implementation of FGPS and SEPS requires that switches are sampled from specific distributions; it is important to discuss the algorithms used as they may present problems which affect how useful these methods are for the sorts of calculations of interest.

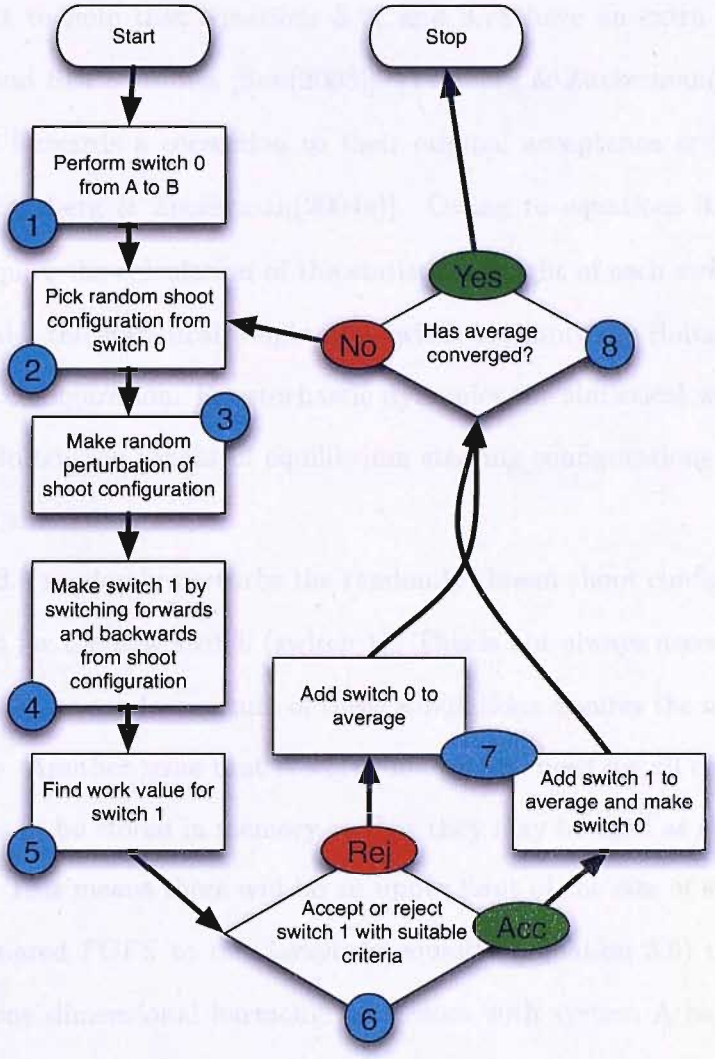


Figure 3.8: Flow diagram describing the path sampling algorithm. Step numbers are marked in light blue.

Figure 3.8 is a flow diagram describing the sampling procedure for both FGPS and SEPS. The major difference between FGPS and SEPS with regard to figure 3.8 is the acceptance test (step 6). For FGPS a new switch $(\mathbf{Z}_{F'})$(switch 1 in figure 3.8) with work W' is accepted according to,

$$\min\left[1, \frac{P[\mathbf{Z}_{F'}|\lambda_t]G[\mathbf{Z}_{F'}]exp\{-\alpha_S\beta W'\}}{P[\mathbf{Z}_F|\lambda_t]G[\mathbf{Z}_F]exp\{-\alpha_S\beta W\}}\right]. \tag{3.77}$$

where $G[\mathbf{Z}_{F'}]$ is the probability of generating switch 1 from switch 0. For SEPS a new switch is accepted according to

$$\min\left[1, \frac{P[\mathbf{Z}_{F'}|\lambda_t]G[\mathbf{Z}_{F'}]exp\{-\beta/2W'\}}{P[\mathbf{Z}_F|\lambda_t]G[\mathbf{Z}_F]exp\{-\beta/2W\}}\right]. \tag{3.78}$$

It may be important to note that equations 3.77 and 3.78 have an extra term compared to the original FGPS and SEPS studies [Sun(2003)], [Ytreberg & Zuckerman(2004b)]. Ytreburg and Zuckerman put forwards a correction to their original acceptance criterion which is the form listed above [Ytreberg & Zuckerman(2004a)]. Owing to equations 3.77 and 3.78, both FGPS and SEPS require the calculation of the statistical weight of each switch ($P[\mathbf{Z}_F|\lambda_t]$. For deterministic dynamics the statistical weight of a switch is simply the Boltzmann weight of its equilibrium starting configuration. For stochastic dynamics the statistical weight of a switch is the product of the Boltzmann weight of equilibrium starting configurations and the transition probabilities of each simulation step.

Step 3 in figure 3.8 randomly perturbs the randomly chosen shoot configuration which will be the starting point for the new switch (switch 1). This is not always necessary in the case of stochastic dynamics as the random nature of these simulations ensures the new switch is different from the old one. Another issue that is worth note is the need for all configurations of the new and old switches to be stored in memory so that they may be used as shoot configurations for the next switch. This means there will be an upper limit of the size of switches.

Sun (2003) compared FGPS to the Jarzynski equality (equation 3.6) using a free energy difference between one dimensional harmonic oscillators with system A having a double well in its potential and system B having a single well. Jarzynski estimates were unable to give an estimate accurate within 15 k_BT using 1x10^6 switches, whereas FGPS used 10,000 switches of the same length to give an accurate estimate. Also, a perturbation of the separation of two methane molecules in a periodic box of water was studied, with the separation being 8 Å at system A and 4 Å at system B. Again with 5 ps deterministic switches FGPS gave accurate estimates where the Jarzynski method did not.

Ytreburg and Zuckerman (2004b) compared the computational time used in Jarzynski estimates using various switch lengths, TI with ten λ windows and SEPS with short switches for two dimensional harmonic oscillators with system A having a single well and system B a double well. They found that SEPS was over 500 times more efficient than either FG or TI [Ytreberg & Zuckerman(2004b)], [Ytreberg & Zuckerman(2004a)].

In a more recent study, Ytreburg *et al.* compared SEPS, TI, AIM, Jarzynski averaging, BAR and SEPS with the BAR estimator (SEPS-BAR) [Ytreburg *et al.*(2006)Ytreburg, Swend-

sen & Zuckerman]. The SEPS-BAR method involves the use of switches in both directions with the SEPS method. BAR generalised to use work values from the SEPS method is,

$$\sum_{n_F}^{D[\mathbf{Z}_F|\lambda_t]} \frac{exp\{+\beta/2W\}}{(1+\frac{n_F}{n_R}exp\{\beta(W_i-\Delta F)\})} = \sum_{n_R}^{D[\mathbf{Z}_F|\lambda_t]} \frac{exp\{+\beta/2W\}}{(1+\frac{n_F}{n_R}exp\{-\beta(W_j-\Delta F)\})}. \tag{3.79}$$

The two test calculations used in this comparison were the free energy of growing a Lennard-Jones sphere and charging of the same Lennard-Jones sphere both in a box of water. For the Lennard-Jones growth calculation it was found that AIM, TI and BAR were able to produce estimates to within 0.5 kcal.mol^{-1} with around 4 times fewer MC steps, while AIM and SEPS-BAR were able to provide the most accurate and precise estimates. For the charging of a Lennard-Jones sphere again AIM, TI and BAR produced estimates to within 0.5 kcal.mol^{-1} faster than the others by around 3-4 times, with BAR using 80,000 MC steps and AIM and TI using 145,000. AIM and BAR were deemed able to give the most precise estimates.

While this most recent study of AIM, TI, FG, BAR, SEPS and SEPS-BAR used a solvated system for calculations, both the growth and charging calculations were not very demanding when compared to many types of calculation that may found with protein-ligand systems. SEPS and SEPS-BAR both suffered from costly equilibrations of their switch sampling processes. These methods may offer advantages for very demanding calculations with large PMF undulations as they fully evaluate across λ and do not experience bias, the two main disadvantages of the other methods. For these systems AIM and BAR proved to be the most useful overall.

3.3.7 Replica Exchange Fast Growth

Combining replica exchange methodology (RE methods) with free energy methods has been shown to provide improvements in accuracy and efficiency [Woods *et al.*(2003a)Woods, Essex & King], [Woods *et al.*(2003b)Woods, Essex & King]. Combining RE methods with FG (REFG) is therefore deemed worthwhile. It is not possible to make the λ moves discussed for RETI in the previous chapter between nonequilibrium switches as apart from the initial ones, configurations are not part of an equilibrium ensemble. Thus, combining RE and FG can only be achieved by performing λ swap moves between configurations of the equilibrium seed simulations. λ swap

moves would again be accepted on the basis of equation 2.56 of chapter 1.

When attempting a λ swap move it is important that there is a large amount of phase space overlap between the two ensembles involved. Attempting λ swaps between equilibrium seed simulations at $\lambda = 0$ and $\lambda = 1$ would result in an extremely low acceptance rate for all but the smallest perturbations. In discussions above of the best practice for FG calculations, the idea of dividing the λ coordinate into a number of smaller intervals which are evaluated by independent FG calculations was found to be more efficient in many situations than performing uninterrupted switches from systems A to B. These differences in protocol are described in figures 3.1 and 3.3 above. By using a protocol with many equilibrium seed simulations across the λ coordinate we can increase the phase space overlap of those equilibrium seed simulations adjacent in λ. In this case of increased phase space overlap RE λ swap moves carried out between equilibrium seed simulations adjacent in λ have the possibility of a high acceptance rate. The use of RE λ swap moves in the generation of FG starting configurations should in theory reduce error associated with incomplete sampling of large systems.

REFG methods will be investigated in this study and compared to original FG and RETI to discern if they have any application in protein-ligand free energy calculations.

3.4 Calculating errors, and inaccuracies of free energy calculations

When producing a computational estimate of ΔF there are often errors. It is extremely useful to be able to measure the error of an estimate and to be able to declare an estimate free of error. The accumulation of averages of simulation configurations and of FG switches can be discussed together as general measurements when error calculation is being discussed.

Errors in simulation averages have their origins in the fact that sampling of systems with infinite phase spaces can never be complete. However the ergodic hypothesis discussed in the previous chapter allows the convergence of simulation averages with a finite amount of sampling. In practice, simulation methods are often unable to negotiate large energy barriers in the phase space of bio-systems, and get stuck sampling only from local minima. Thus two independent simulations can converge to different averages as they have sampled different regions of phase space.

General practice for estimating errors in simulation averages is to use statistical measures of the variance of averages due to K independent blocks of measurements (block variance methods). It is difficult to measure the independence of blocks of measurements. One approximate method for finding independent block sizes for block variance methods is to increase the size of blocks until the calculated error reaches a plateau and does not change significantly. Hummer (2001) plotted error estimates as a function of $\ln K$ to find estimates with independent blocks, as the estimates plateau as K gets smaller (independent blocks method).

The block variance methods discussed here can only give an idea of the variance in the data. If a simulation is unable to overcome barriers in the energy surface of a system and only samples from a subset of phase space these methods will not give a good estimate of the possible range of results. The best way to gauge the possible range of results is to independently repeat calculations a number of times. This can be extremely time consuming for protein-ligand systems as these types of calculation can be slow.

Another method routinely used to gauge the true convergence of free energy calculation is to engineer a cycle of calculations such that the overall ΔF is zero. Owing to the fact that free energy is a function of state a pathway of system perturbations starting and ending with the same system will have a ΔF of zero. If the same cycle of computational estimates of the ΔF is equal to zero it signifies the possibility that the calculations in the pathway are truly converged. The random balancing of unconverged ΔF estimates is of course a real possibility and this method of cycle closure should not be relied upon not to give false positives.

Of course for non-linear averages such as that in the Jarzynski equality (equation 3.6) another source of error is the bias of the average which is not explicitly considered in block variance methods. Various methods of considering the bias of non-linear averages have been discussed above and evaluations of the relative size of statistaical error and bias can be found in the literature [Hummer(2001)], [Gore *et al.*(2003)Gore, Ritort & Bustamante], [Shirts & Pande(2005)], [Wu & Kofke(2005a)]. The Kofke bias measure discussed above is able to give estimates of bias-free results for Jarzynski calculations where the work distribution is Gaussian. The necessary Gaussian nature of Kofke bias estimates may be a major drawback as many systems of interest may not have work distributions which are Gaussian. However the work of Jarzynski (2006a) may give hope of similar measures of the bias of FG calculations which are universally applicable. In this case the bias of Jarzynski averages may become an advantage

giving the ability to predict convergence.

The aim of this study is to investigate the application of free energy calculations to drug design and therefore protein-ligand systems. Thus, it is important not to forget that possibly the most important and difficult area which must be mastered in avoiding errors in ΔF estimates in comparison to experiment is the building of an accurate model system.

3.5 Attributes and limitations of relevant methods

As discussed in the main introduction, this study is concerned with finding the most appropriate methods for fast and accurate computational calculation of ligand binding affinities such that they may be used at the lead optimisation stage of the drug design process. At present the goal of using free energy calculations in drug design is not realised. What is necessary for this to become reality? Some members of the drug design industry feel a successful method would need to be able to calculate the relative binding free energies of a set of lead drug candidates (around 7-15 ligands) for a particular target overnight. Of course it is desirable that this method could deal with a group of ligands which were highly heterogeneous in structure but this is not absolutely necessary. The most important aspect of these calculations is that they are reliable and give results that large pharmaceutical companies will risk money on.

There are examples of calculations which have achieved the benchmark discussed in the paragraph above using FEP the oldest and most basic method discussed here [Price & Jorgensen(2000b)], [Price & Jorgensen(2000a)], [McDonald & Still(1996)]. Also, there are examples of studies using more sophisticated methods and achieving relatively fast and accurate results [Michel *et al.*(2006)Michel, Verdonk & Essex]. However, these examples are still very system specific and generally limited to groups of ligands which have a strictly common structure (congeneric set of ligands). At present, to move beyond these congeneric sets of ligands either a dual topology free energy methodology must be employed or a non-rigourous method such as MMPBSA.

MMPBSA has been shown to be unreliable by some studies [Pearlman(2005)] [Kuhn *et al.*(2004)Kuhn, Gerber & Stahl] and may not be able to reproduce results of the quality needed as discussed above. Dual topology free energy methods suffer from large amounts of sampling error which can result in noise and convergence problems. At present dual topology calculations need many

times the amount of simulation sampling of single topology methods. Thus, faster and more efficient free energy methods are still required.

The speed of a free energy calculation can be measured in a number of ways, normally either computational time or wall clock time. Computational time is the amount of time it would take a single processor to complete the task at hand whereas wall clock time is the amount of time the task takes regardless of the number of processors used. In terms of carrying out the calculations at hand if unlimited funds are available to invest in unlimited computers all that matters is wall clock time.

FG based methods have the advantage that calculations can be parallelised on a large scale, while other methods are able to utilise very limited parallelisation or must be run on a single processor. This is a large advantage often overlooked by using comparisons of computational time. If many processors are available then FG may be able to run calculations at a fraction of the wall clock time of other methods. Thus this study will investigate some of the FG based methods discussed above and compare them to the established equilibrium methods TI and RETI.

All of the FG based methods discussed above have been investigated to some degree. The extrapolation methods discussed above have been investigated. However, any extrapolation results are not discussed here as investigations were not completed due to their unfavourable nature. FGPS and SEPS, although powerful and promising methods, are not easily parallelised; indeed SEPS must be run on a single processor. For that reason FGPS and SEPS will not be investigated here. All of the remaining methods unless otherwise stated are listed in a table 3.2 with any abbreviations, figure legend labels and associated equations or descriptions.

Method name	Abbreviation	Figure legend	Type	Description
Jarzynski	none	Jarz Fwd/Bwd	Estimator	Equation 3.7
Fluctuation dissipation	FD	FD Fwd/Bwd	Estimator	Equation 3.11
Symmetric A	none	Symmetric A	Estimator	Equation 3.5
Symmetric B	none	Symmetric B	Estimator	Equation 3.12
Bennett's acceptance ratio	BAR	BAR	Estimator	Equation 3.23
Gore bias	none	Jarz Fwd/Bwd-GBias	Bias correction	Figure 3.4
Kofke Bias	none	Kofke Bias fwd/bwd	Bias measure	Equations 3.41 and 3.43
λ bias	none	λ bias	Switch sampling	Figure 3.6
Configuration bias-d	none	Confbias-d	Switch sampling	Figure 3.7
Hybrid bias	none	Hybridbias	Switch sampling	Equations 3.62, 3.63 and 3.64
Division of λ	-BY?	-BY?	Switch sampling	Figure 3.3
REFG	none	REFG	Seed sampling	Described above and equation 2.56

Table 3.2: Table of all FG methods to be investigated in this work. The figure legend column contains the labels used for each method in all figures. The type column gives a simple description of the method with a reference to the equation of description in the description column.

The methods of table 3.2 were implemented using the ProtoMS 2.1 MC application [Woods & Michel(2005)] often with large associated scripts [Cossins(2007)]. To be clear, it is possible to combine many of the methods of table 3.2 in the same calculation, and this study will be concerned with finding the most efficient combination for use with protein-ligand systems. These methods have been applied to simple one dimensional harmonic oscillator systems in order that implementations may be validated and their performances measured against results found analytically and compared to the established equilibrium methods. Methods deemed most suitable will then be applied to calculations involving solvent rearrangements, as this is a major factor in protein-ligand calculations and again compared to the established methods. Then the best FG based methods will be applied to two sets of protein-ligand calculations recently presented by Michel (2006) and results compared. The results of this study will ascertain whether any of the FG based methods discussed above have the attributes to be routinely applied to protein-ligand calculations.

Chapter 4

Harmonic Oscillator systems

4.1 Introduction

In all areas of science it is important to be careful and test all your ideas and methods. Fast Growth (FG) is a relatively new method and there are still on-going discussions in the literature as to whether or not the Jarzynski equality (equation 3.6) is exactly thermodynamically and mathematically correct [Jarzynski(1997b)], [Cohen & Mauzerall(2004)], [Jarzynski(2004)], [Cohen & Mauzerall(2005)]. Thus it is important to thoroughly test FG and our implementation of it. Testing can be done on very simple models which can be evaluated quickly and results checked with confidence by comparison with analytical or exact numerical solutions.

4.2 Harmonic Oscillator Models

The independent harmonic oscillator model (IHO) is essentially a simple spring in vacuum. To test our FG implementations we can define two IHO systems and calculate the free energy difference (ΔF) between them. Systems A and B both have the same number of oscillating particles, N, with differing Hamiltonians,

$$H_A = \sum_{i=1}^{N} \omega_A x_i^2 \tag{4.1}$$

$$H_B = \sum_{i=1}^{N} \omega_B (x_i - x_0)^2 \tag{4.2}$$

where x_i is the reference coordinate for particle i and ω_A and ω_B are the force constants of the two systems which control the size of oscillations the particles will undergo. ω_A and ω_B control the size of phase space each system explores. x_0 displaces the reference position of particles in system B which together with ω_A and ω_B gives control over the amount of phase space overlap between the systems.

This IHO model is very simple and we can calculate many of its properties analytically, including ΔF which can be found using equation 4.3. This is very useful as the result from our protocol can be compared to the right answer rather than an answer found using an exhaustive free energy protocol.

$$\Delta F = \frac{1}{2} N k_B T \ln\left(\frac{\omega_B}{\omega_A}\right) \tag{4.3}$$

The IHO systems described here were originally proposed by Wu and Kofke in their study of Rosenbluth FG sampling [Wu & Kofke(2005c)]. Similar IHO systems were used to test FG by Shirts and Pande in a slightly earlier study [Shirts & Pande(2005)]. As we have implemented the FG methods developed by Wu and Kofke for our study it was convenient to validate and test our implementations on the same test systems. The parameters of these systems are listed in table 4.1.

Each of the cases described in table 4.1 is set up in an attempt to model difficulties encountered with free energy calculations on more complex, chemically relevant systems. Those difficulties are: if important phase space of systems A and B is in different places as the system is perturbed, from A to B, it is likely that the system will be in a configuration which is unfavourable for the Hamiltonian of system B and that large energies will result, which can hinder convergence (energetic barriers). Energetic barriers are increased as x_0 is increased, and the amount of important phase space overlap between systems A and B decreased. If the important phase space of one end point (A or B) is much larger than the other as the system is perturbed from A to B it is likely that the system will be in an area of phase space within the larger important phase space but outside of the smaller important phase space again causing large energies and hindering convergence (entropic barriers). Entropic barriers, which may come about both by decreasing phase space overlap and as the relative size of the destination system's important phase space is decreased compared to the starting system i.e. increasing

the ω_B/ω_A ratio.

Case	N	$\omega_B(kcalmol^{-1})$	$\omega_A(kcalmol^{-1})$	x_0	$\beta\Delta F$
A	10	500	1	0	31.07
B	10	20	1	0	14.98
C	10	20	1	1	14.98
D	10	5	1	3	8.05

Table 4.1: Parameters of four test IHO systems used to test our FG implementations

It is instructive to visualise the phase space of the end point systems of cases A-D. As each of the particles in each case is independent of the other, visualisation can take the form of phase space distributions for a single particle. It becomes easy to understand the problems involved in sampling from both distributions in each of the cases when studying figure 4.1.

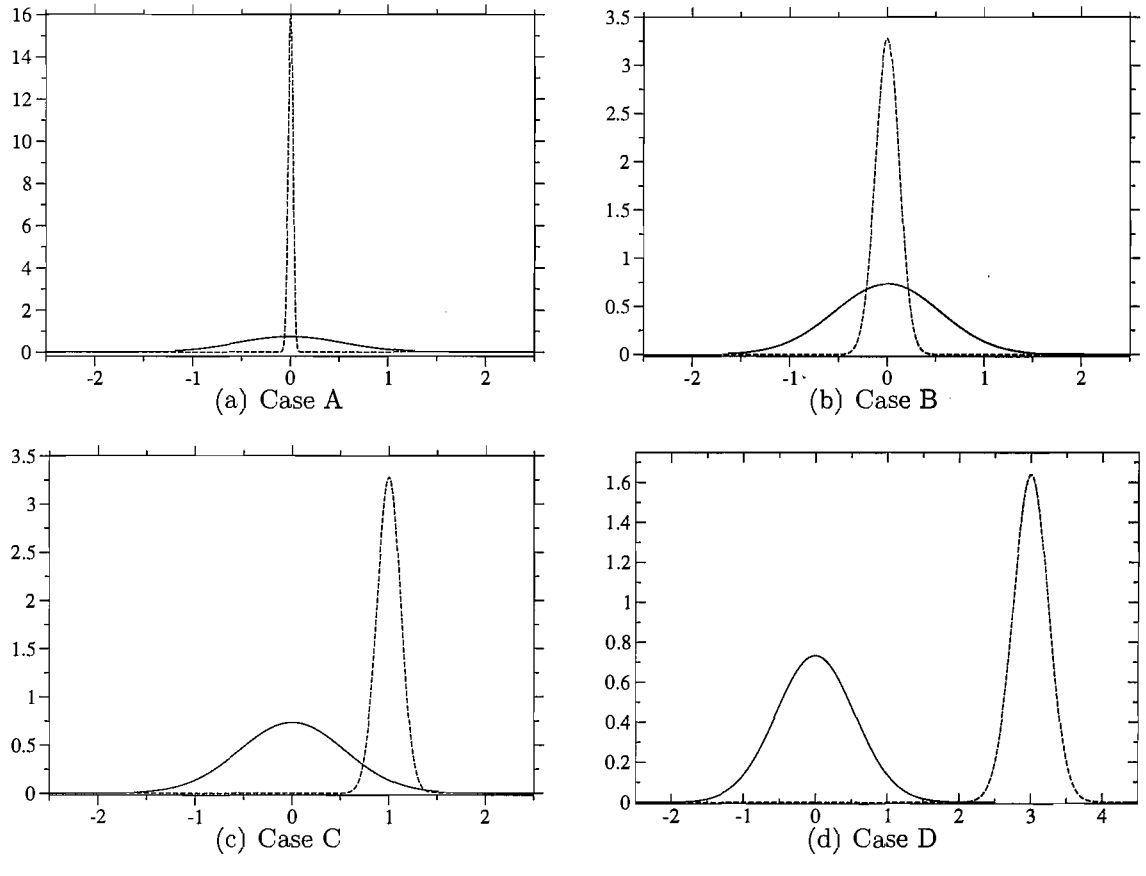


Figure 4.1: Phase space distributions of single particles for IHO cases A-D as defined in table 4.1. Solid lines are phase-space distributions for system A and dashed lines are for system B. The x axis is the dimensionless particle coordinate and all distributions are normalised to unity.

- Case A is an example of an extreme subset relation; ω_B is extremely large in comparison to

ω_A. Thus, it becomes improbable that a simulation will sample configurations important to system B while close to system A.

- Case B is an example of a subset relation; ω_B is larger in comparison to ω_A. This relation is similar to Case A but the entropic barriers associated with the relative size of systems A and B are smaller.

- Case C is an example of a subset, partial overlap relation combination; ω_B is larger in comparison to ω_A and is also moved to the right. In this case it is very improbable that the simulation will sample system B while close to system A as the most probable sampling areas of the two systems are in different places.

- Case D is an example of a slight subset, non-overlap relation combination; ω_B is slightly larger in comparison to ω_A and is moved far enough to the right so the two distributions share no important phase space at all. In this case it is extremely difficult for the simulation to sample phase space important to system B while close to system A.

We would expect to find differing combinations of the problems represented by cases A-D in more complex systems. In the case of protein-ligand perturbations, barriers linked to solute-solvent and solute-protein interactions cause many of the difficulties encountered and energy surfaces are extremely large and rugged. The IHO systems investigated here are unable to represent these problematic interactions and energy surfaces well, and this analysis should not be regarded as a good model specifically for protein-ligand perturbations but instead illustrate the specific issues involved.

4.3 Simulation of Harmonic Oscillator Systems

All IHO systems were set up as molecules with eleven particles and ten bonds. Each of the bonds was setup as an IHO with the force constant ω, all other possible interactions were not used. IHO simulations were carried out using a modified version of our ProtoMS 2.1 Monte Carlo simulation software [Woods & Michel(2005)]. All IHO simulations were coupled to λ through a single topology switching protocol. FG starting configurations were generated every 200 MC trials from equilibrium sampling of the end point systems.

4.4 Harmonic Oscillator Results

4.4.1 Method Validation

This validation must have two levels. First, FG methods must be able to produce ΔFs which are very accurate compared to the analytical value. Second, FG methods should be able to reproduce the efficiency trends which are demonstrated by Wu and Kofke [Wu & Kofke(2005c)]. To achieve this validation we make a direct comparison of our results with Wu and Kofke.

Before making the proposed comparison there are some important points to consider. When publishing their study, Wu and Kofke defined the ratio ω_B/ω_A with out giving the actual values for ω_B and ω_A, for each of the cases A-D. While this information is enough to describe the relative size of the important phase spaces of the two end point systems A and B, it does not define the size of the important phase spaces relative to the displacement x_0 (problem 1). Problem 1 means it is hard to know how big to make the variable x_0 in cases C and D. If we set ω_B and ω_A too high compared to x_0 then the energetic barriers encountered will be bigger than those used by Wu and Kofke and methods will be less efficient.

The range of possible MC particle move sizes are also not mentioned (problem 2). Problem 2 could affect the efficiency of free energy estimates in all of the IHO cases. However as long as Wu and Kofke did not vary the size of MC moves between cases or between MC moves, this is a simple case of finding an optimal maximum move size.

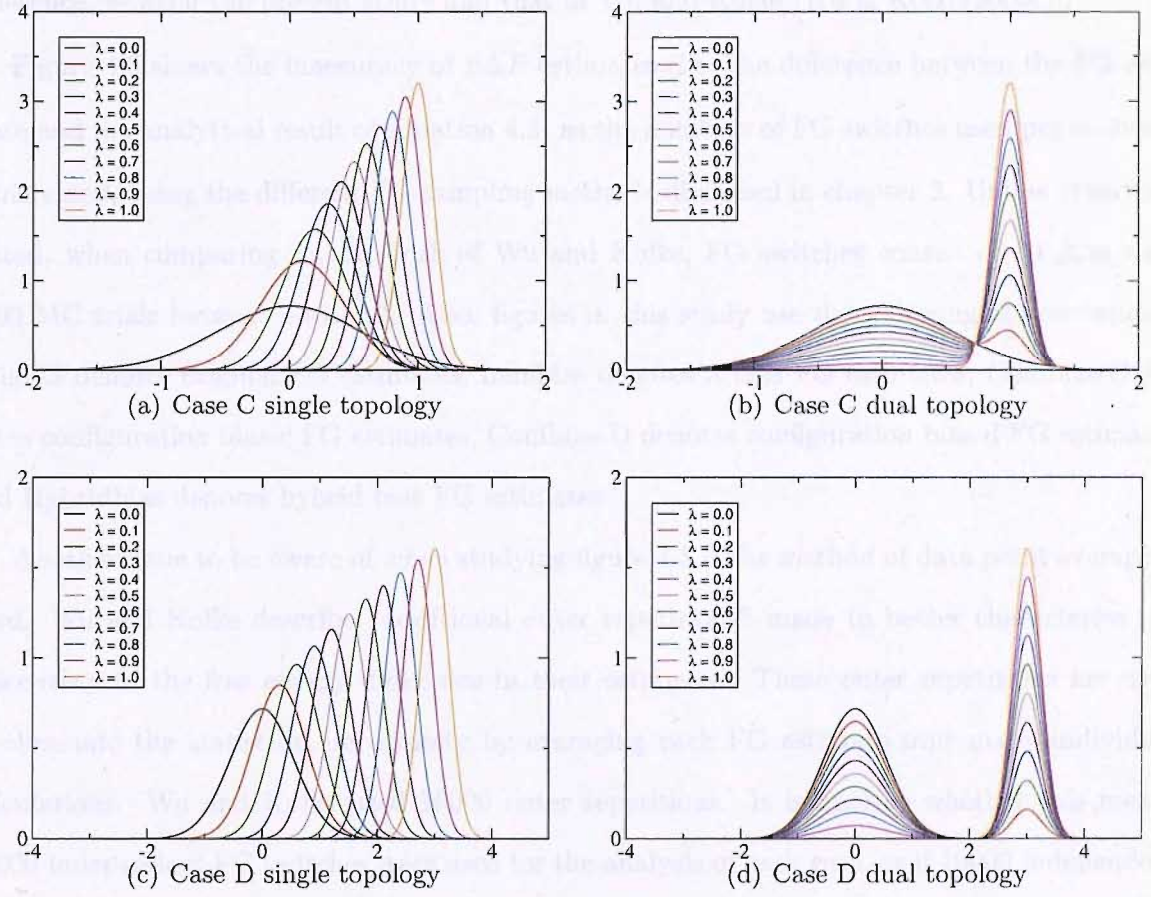
The ten IHOs in each of cases A-D are part of one molecule in this study. This means that each of the bond lengths (i.e. independent oscillators) will be moved for each MC trial. It is not clear whether this is also true for the study carried out by Wu and Kofke (problem 3). This will result in possible differences in MC trial acceptance rates and the efficiency of ΔF calculations.

Wu and Kofke describe their coupling of λ to the Hamiltonian as a linear scaling of the reference and target Hamiltonians (equation 4.4). This method uses two separate systems in a dual topology arrangement. The λ coupling used in the present study is achieved by scaling the parameters of a single system in a single topology arrangement as in equation 4.5.

$$H_\lambda(z) = H_A(z) + \lambda[H_B(z) - H_A(z)] \tag{4.4}$$

$$H_\lambda(z) = \sum_{i=1}^{N}((1-\lambda)\omega_A)(x_i)^2 + \sum_{i=1}^{N}(\lambda\omega_B)(x_i - \lambda x_0)^2 \tag{4.5}$$

H in equations 4.4 and 4.5 is the Hamiltonian and z is the present system configuration. What effect this difference in perturbation method will have is not clear (problem 4). We can attempt to understand the differences between these two methods by plotting the important phase space at the point of each λ increment (figure 4.2).



(a) Case C single topology (b) Case C dual topology

(c) Case D single topology (d) Case D dual topology

Figure 4.2: Phase space distributions of single particles using single and dual topology methods for IHO cases C and D from table 4.1. Each of the coloured distributions represents the system at a particular value of lambda. The x axis is the dimensionless particle coordinate and all distributions are normalised to unity.

It is unclear whether either dual or single topology calculations have an advantage for these IHO cases. In theory it is possible that dual topology sampling has an advantage for cases C and D as the most important regions of phase space (the distribution peaks in figure 4.2) do not shift from their previous position as λ is changed, as they do for single topology sampling.

However, the dual topology method may have higher levels of sampling error compared to single topology as the system does not truly mutate as λ changes and systems A and B could be in different areas of phase space. When applied to large complex systems (such as protein-ligand systems) the high levels of sampling noise generally found with dual topology methods can be a large disadvantage. Therefore single topology methods have been used through out this study as the main focus is towards calculating ligand binding affinities. The issues involved in this comparison of dual and single topology IHO calculations are not clear and their clarification is not within the scope of this study. Thus, it should be noted that this is another source of difference between the present study and that of Wu and Kofke [Wu & Kofke(2005c)].

Figure 4.3 shows the inaccuracy of $\beta\Delta F$ estimates (i.e. the difference between the FG estimate and the analytical result of equation 4.3) as the numbers of FG switches used per estimate is increased, using the different FG sampling methods discussed in chapter 3. Unless otherwise stated, when comparing to the work of Wu and Kofke, FG switches consist of 10 $\Delta\lambda$s with 1000 MC trials between each $\Delta\lambda$. Also, figures in this study use the following abbrieviations: OrigFG denotes original FG estimates, Lambias denotes λ bias FG estimates, Confbias-C denotes configuration bias-c FG estimates, Confbias-D denotes configuration bias-d FG estimates and Hybridbias denotes hybrid bias FG estimates.

Another issue to be aware of when studying figure 4.3 is the method of data point averaging used. Wu and Kofke describe "additional outer repetitions" made to better characterise the inaccuracy in the free energy difference in their estimates. These outer repetitions are used to eliminate the statistical uncertainty by averaging each FG estimate over many individual calculations. Wu and Kofke used 10000 outer repetitions. It is unclear whether this means 10000 independent FG switches were used for the analysis of each case, or if 10000 independent FG estimates were made for each data point (this is problem 5). Our analysis has used the former of these two possibilities. Thus, for each case in figure 4.3 10000 FG switches were made. Hence, the data point for 1 FG switch per FG estimate was averaged over 10000 estimates and the data point for 1000 FG switches per FG estimate was averaged over 10 estimates.

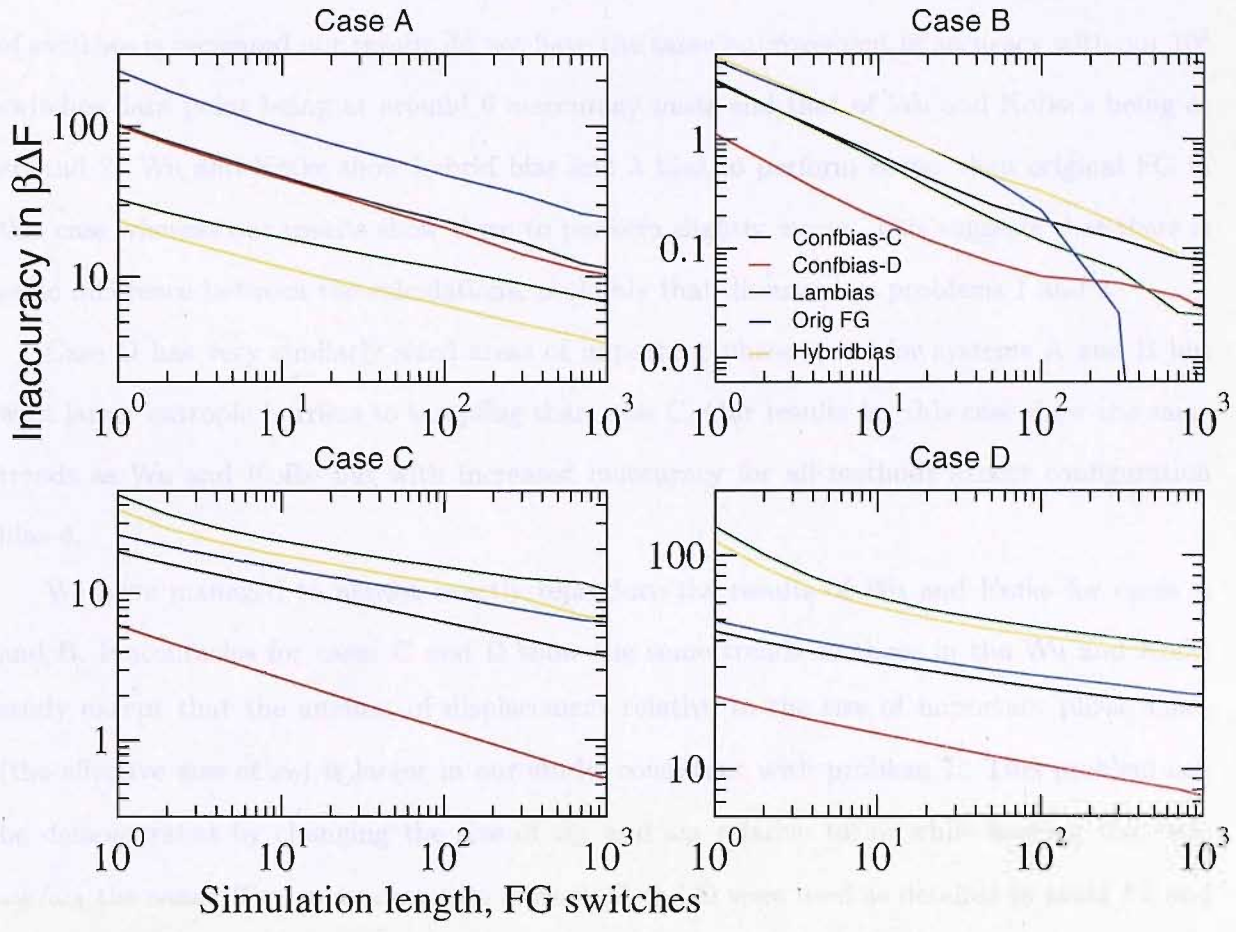


Figure 4.3: IHO results using all Rosenbluth FG methods and original FG for cases A-D showing inaccuracy in $\beta\Delta F$ against numbers of trajectories where a trajectory has 10 λ increments and 1000 MC trials per λ increment. For configuration bias 10 configuration samples were used per λ increment.

Case B is the least demanding of cases A-D, shown in figure 4.3. All the Kofke FG methods can calculate a free energy very close to the analytical value. Our results, for case B, reproduce those seen by Wu and Kofke very well [Wu & Kofke(2005c)]. The calculated free energies diverge from the Wu and Kofke results and give a less clear idea of the inaccuracy as we approach 1000 FG trajectories for Case B, this may be due to problem 5.

Case A is a more demanding free energy difference to evaluate than Case B. Consequently, original FG can only estimate ΔF to within 20-30 units. Hybrid bias and λ bias FG give a performance improvement of almost an order of magnitude. Configuration bias-C and D FG show improved efficiency over original FG. This result is in line with that seen by Wu and Kofke.

Case C is similar to Case B but provides an extra barrier to sampling. Our results for Case C show significant difference from those of Wu and Kofke. Original FG displays similar performance when only one FG switch is used for each $\beta\Delta F$ estimate. However as the numbers

of switches is increased our results do not have the same improvement in accuracy with our 10^3 switches data point being at around 6 inaccuracy units and that of Wu and Kofke's being at around 2. Wu and Kofke show hybrid bias and λ bias to perform better than original FG in this case whereas our results show them to perform slightly worse. This suggests that there is some difference between the calculations, probably that discussed as problems 1 and 3.

Case D has very similarly sized areas of important phase space for systems A and B but with larger entropic barriers to sampling than case C. Our results for this case show the same trends as Wu and Kofke but with increased inaccuracy for all methods except configuration bias-d.

We have managed to almost exactly reproduce the results of Wu and Kofke for cases A and B. Inaccuracies for cases C and D show the same trends as those in the Wu and Kofke study except that the amount of displacement relative to the size of important phase space (the effective size of x_0) is larger in our study, consistant with problem 1. This problem can be demonstrated by changing the size of ω_B and ω_A relative to x_0 while keeping the ratio ω_B/ω_A the same. Two more examples of cases C and D were used as detailed in table 4.2 and figure 4.4. We would expect that our FG methods to be more efficient for new cases Cx0.5 and Dx0.5 and less efficient for cases Cx2 and Cx2. It is also apparent that as the size of the x_0 displacement parameter is increased relative to the size of the A and B phase space i.e. the λ bias methodology gives a larger disadvantage, whereas the confbias methodology gives a larger advantage.

Case	N	$\omega_B(kcalmol^{-1})$	$\omega_A(kcalmol^{-1})$	x_0	$\beta\Delta F$
Cx2	10	40	2	1	14.98
Cx0.5	10	10	0.5	1	14.98
Dx2	10	10	2	3	8.05
Dx0.5	10	2.5	0.5	3	8.05

Table 4.2: Parameters for new case Cx2, Cx0.5 and Dx2 and Dx0.5 systems

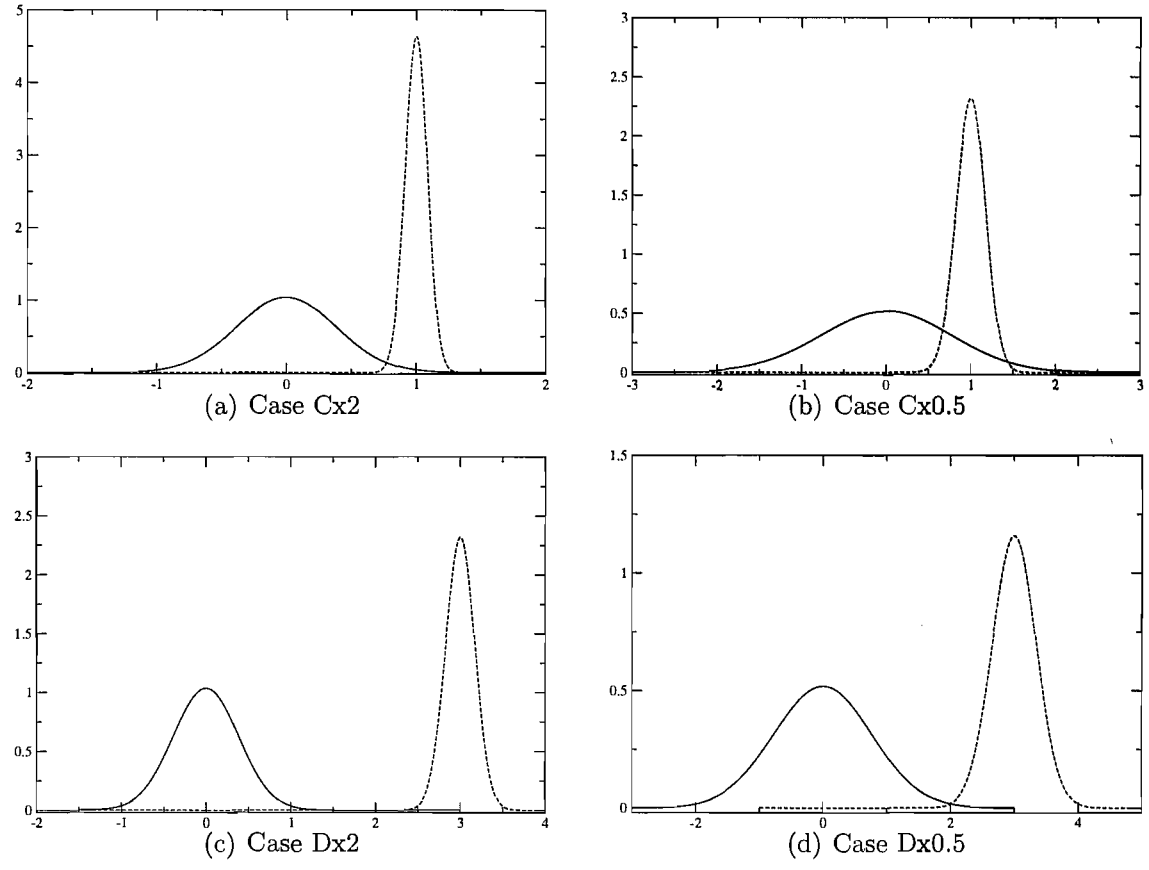


Figure 4.4: Phase space distributions of single particles for IHO new cases C and D as defined in table 4.2. Solid lines are phase-space distributions for system A and dashed lines are for system B. The x axis is the dimensionless particle coordinate and all distributions are normalised to unity.

Figure 4.5 shows that, as predicted, results for cases Cx0.5 and Dx0.5 all FG methods become more effcient compared to cases Cx2 and Dx2. They are less efficient compared to cases C and D. This confirms that problem 1 is an issue for this comparison. Average estimates for case Cx0.5 with configuration bias-d become smaller than the analytical result at around 15 switches, shown in figure 4.5 as the red line curving down to below 0.01 and then immediately increasing again to above 0.1. This behaviour is simply due to statistical error as inaccuracies become very small and data points are averaged over fewer estimates.

The relative performance of FG methods for case Cx0.5 is the same as that seen for case C in figure 4.3. Therefore case Cx0.5 is still different from case C results seen by Wu and Kofke [Wu & Kofke(2005c)] apart from configuration bias-d displaying the least inaccuracy. This suggests there is still a systematic difference in the protocols used. Issues discussed as problem 3 may explain the discrepancy in relative performance of these methods. If the protocol used by Wu and Kofke involves fewer MC trials between each λ increment, configuration bias methods and

original FG would be expected to show an improvement relative to λ bias methods, as configuration bias methods rely on a broad sampling of phase space between λ increments to find the best configuration for a λ increment. This is seen when comparing all case C analyses of this study with case C of Wu and Kokfe.

The relative performance of FG methods for case Dx0.5 is the same as that seen for case D and is therefore similar to case D results of Wu and Kofke. The inaccuracies of all FG methods for case Dx0.5 are much more similar to those seen by Wu and Kofke, although they are not identical.

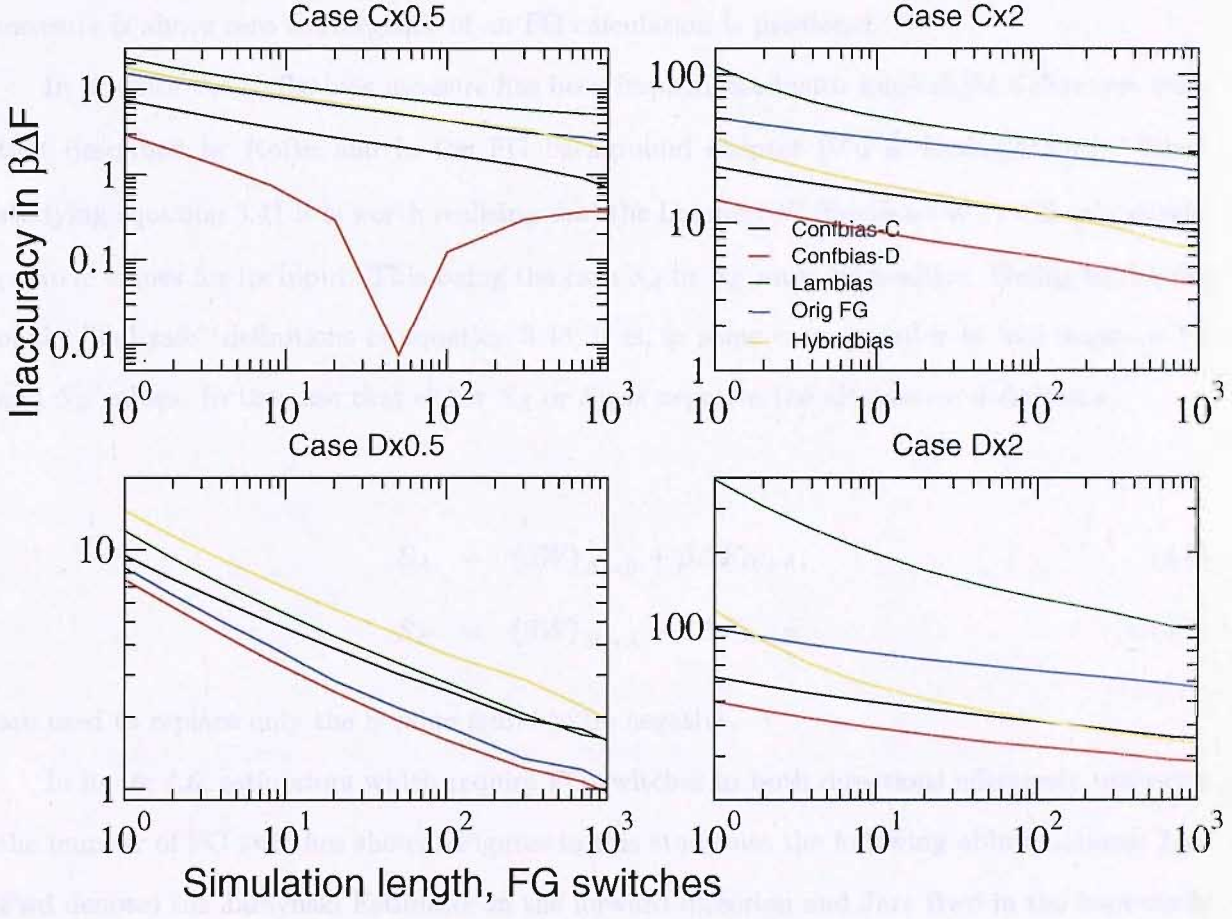


Figure 4.5: inaccuracy in $\beta\Delta F$ against numbers of trajectories for new cases C and D, where a trajectory has 10 λ increments and 1000 MC trials per $\Delta\lambda$. For confbias 10 configuration samples were used per $\Delta\lambda$.

For cases A and B we have quantitatively reproduced the results of Wu and Kofke. The differences found for cases C and D have been rationalised through the differences in protocol from the study of Wu and Kofke. While the difference in protocol may have harmed the comparison with the study of Wu and Kofke, the protocols to be used in this study have been

validated and the possible improved performance of these FG methods demonstrated. For case A, Hybrid bias FG has been shown to be the most efficient, for cases B, C and D Configuration bias-d is the most efficient.

To be sure that all methods used in this study are correct, another analysis was carried out using cases A, B, C and D. The idea in this extra analysis is to produce convergence of ΔF for all our FG estimators. All FG switches are the same as switches in the previous FG method comparison with 10 $\Delta\lambda$s and 1000 MC trials between each $\Delta\lambda$ in both the forward and backward directions. Additionally the convergence heuristic described by Wu and Kofke (Kofke bias) will be tested for each of cases A-D [Wu & Kofke(2005a)]. When the Kofke bias measure is above zero convergence of an FG calculation is predicted.

In practice the Kofke bias measure has been implemented with some slight differences from that described by Kofke and in the FG background chapter [Wu & Kofke(2005a)]. When studying equation 3.41 it is worth realising that the Lambert W function ($\mathbf{W}_L$) will only accept positive values for its input. This being the case S_A or S_B must be positive. Owing to the use of the "fail safe" definitions of equation 3.43, it is, in some cases possible to find negative S_A and S_B values. In the case that either S_A or S_B is negative the alternative definitions,

$$S_A = \langle \beta W \rangle_{A\to B} + \beta \Delta F_{B\to A}, \tag{4.6}$$

$$S_B = \langle \beta W \rangle_{B\to A} + \beta \Delta F_{A\to B}. \tag{4.7}$$

are used to replace only the S value found to be negative.

In figure 4.6, estimators which require FG switches in both directions effectively use twice the number of FG switches shown. Figures in this study use the following abbrieviations; Jarz Fwd denotes the Jarzynski Estimator in the forward direction and Jarz Bwd in the backwards direction, Jarz Fwd-GBias denotes the Jarzynski estimator corrected by the bias as calculated by Gore *et al.* in the forwards direction and Jarz Bwd-GBias in the backwards direction [Gore *et al.*(2003)Gore, Ritort & Bustamante], FD Fwd denotes the fluctuation-dissipation estimator in the forwards direction and FD Bwd in the backwards direction, Symmetric A and B denotes the symmetric A and B estimators, BAR denotes Bennett's acceptance ratio and analytical FE denotes the analytical free energy difference. Results for relatively inaccurate extimators may be omitted from a figure and legend in favour of gaining extra detail for the more accurate

estimators. Errors for all results in this chapter are calculated using the block variance method using 10 blocks, apart from the FD estimators which use the independent blocks method, both described earlier in the main introduction.

For case A, neither forwards only, nor backwards only switching with any estimator has produced a good ΔF estimate. However, BAR converges to a relatively accurate answer at around 20 thousand switches. This is reflected by the Kofke bias measure as both forward and backward predictions are well below zero. Surprisingly, Case B is very different; here only estimates using forward switches and the Jarzynski estimator are accurate and converged. The inaccuracy of estimates using switches in the backwards direction is possibly due to the difficulty of sampling configurations important to system A from a simulation close to system B. This inaccuracy of backwards switches for case B is large enough to inhibit the ability of BAR to give accurate estimates. The measure predicts that backwards Jarzynski estimates are converged with a value well above zero. The forwards value Kofke bias value is just below zero. This Kofke bias convergence measure is obviously wrong in this case.

Cases C and D are similar and show that only BAR and the symmetric B estimators are accurate and easily converged. The relative displacement of the important phase spaces of systems A and B in cases C and D creates a sampling barrier which is seen in estimators using switches in one direction. It is interesting that the symmetric B estimator gives accurate estimates for case C. Case C is a partial overlap relation but the size of the additional subset relation means that the barriers experienced in the forwards and backwards directions are not expected to be similar; consequently the symmetric A estimator is highly inaccurate. The Kofke bias predicts that forwards and backwards Jarzynksi estimates are not converged for cases C and D.

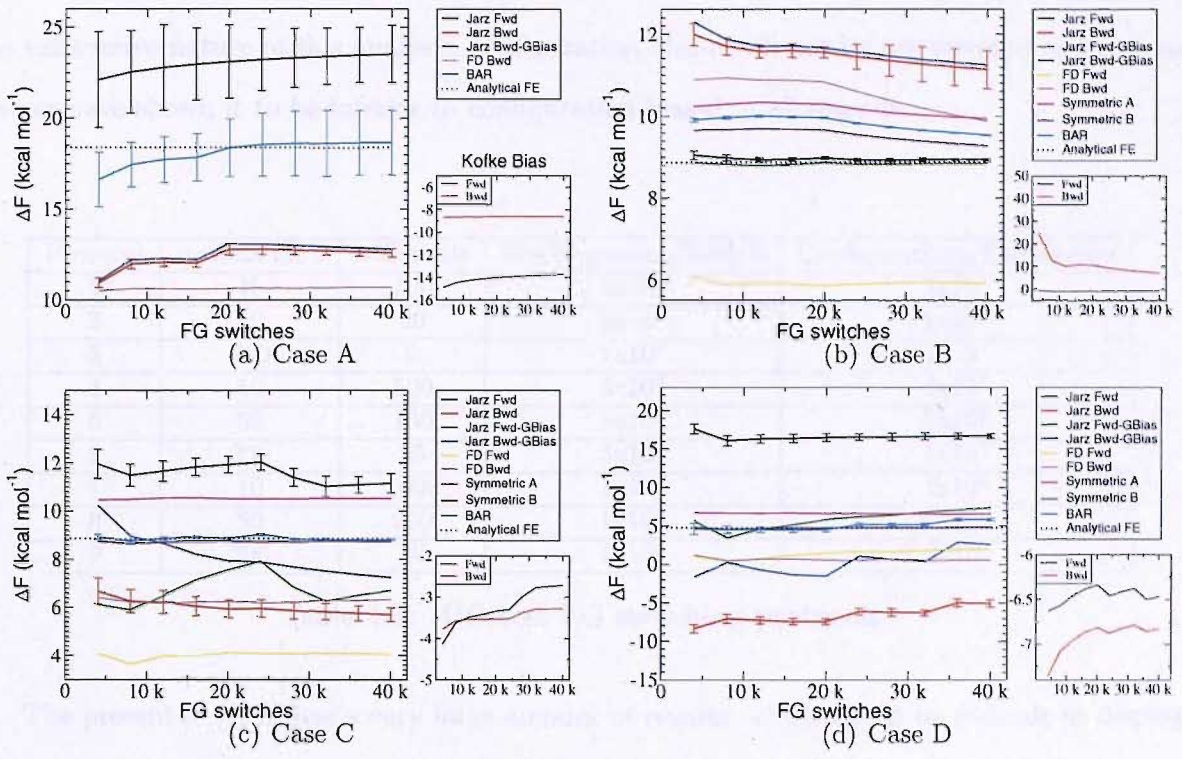


Figure 4.6: ΔF estimates calculated using original FG with all estimators using $2x10^8$ configurations. The plots labelled Kofke Bias display the Kofke bias of estimates plotted in the main windows. If the Kofke bias is above 0-0.5 then the corresponding estimate is predicted to be converged and accurate.

4.4.2 Best practice for accurate ΔF prediction

IHO systems take a very small amount of computer time to sample and simulate. Therefore, it is possible to explore the parameter space of our FG methods more fully than we could with other, larger, more demanding systems.

It is interesting to consider the effect of varying the numbers of λ increment and amount of sampling between each λ increment for the same number of overall configurations used in a switch. It is not immediately obvious whether FG switching protocols differing only in numbers of λ increment and amount of intervening sampling will have substantially different performance characteristics. With this in mind a set of nine switching protocols has been designed to be used with each of the IHO cases A-D and each of our implemented FG sampling methods and estimators (table 4.3). Configuration bias calculations with fewer than 25 MC trials between λ increments will have 5 sampled conformations to pick from for each λ increment rather than 10, as is the case with all other calculations. All FG calculations use the same starting configurations where possible to allow closer comparison of the different protocols. The figures in this section display the number of FG switches for the whole calculation on the x-axis. Owing to

the exhaustive nature of this analysis, configuration bias-c will not be investigated as previous results have shown it to be inferior to configuration bias-d in all regards.

Protocol	lambda inc's	MC trials	Configurations/Switch	Configurations/Calculation
1	10	100	$1x10^3$	$1x10^7$
2	50	20	$1x10^3$	$1x10^7$
3	200	5	$1x10^3$	$1x10^7$
4	10	500	$5x10^3$	$5x10^7$
5	50	100	$5x10^3$	$5x10^7$
6	200	25	$5x10^3$	$5x10^7$
7	10	1000	$1x10^4$	$1x10^8$
8	50	200	$1x10^4$	$1x10^8$
9	200	50	$1x10^4$	$1x10^8$

Table 4.3: IHO test FG switching protocols

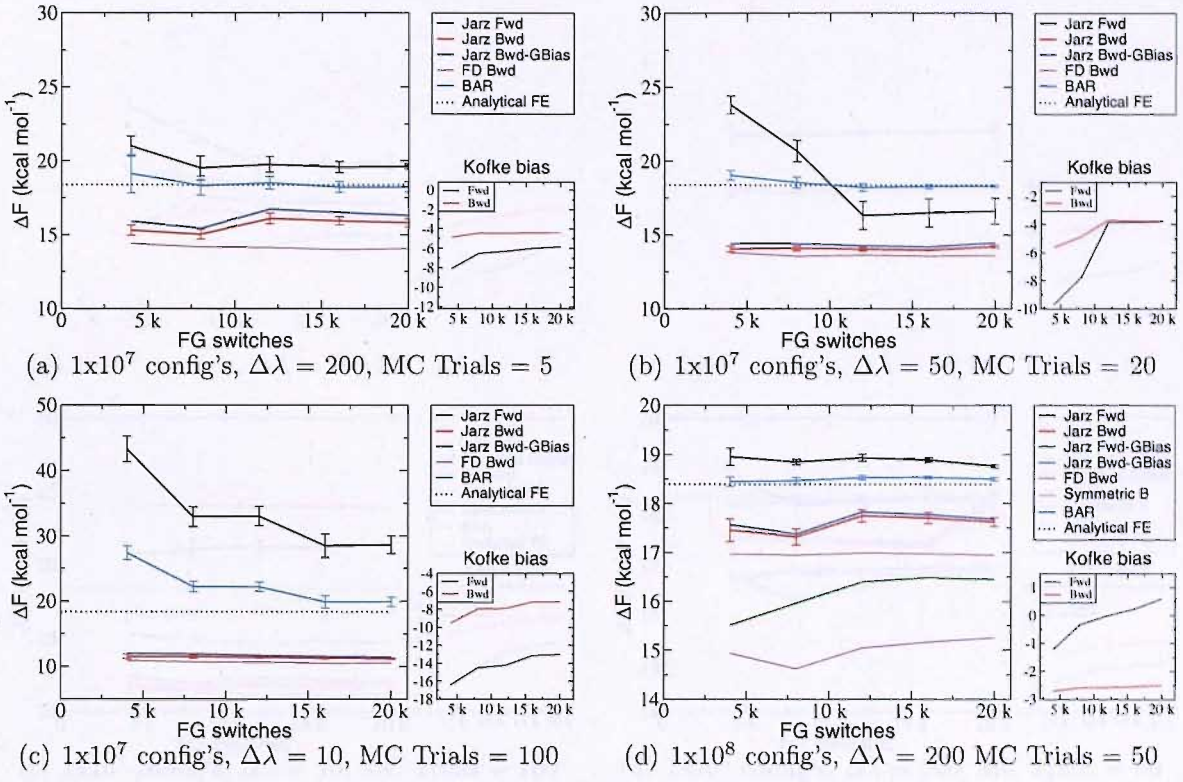
The present analysis has a very large amount of results, which would be difficult to display in full. In each of the following plots only data of interest is displayed. Results of interest are defined as results which show reasonably good performance relative to all other FG methods. In demanding cases, the FD estimator and Gore Bias estimator give wildly incorrect results and large statistical errors. Also in cases with large differences between the forward and backwards calculations the symmetric estimators are inaccurate. These estimators will be referred to in this work as the poorer estimators. In some figures below results for the poorer estimators may be displayed without their statistical errors as these errors are large and can obscure the figure.

Case A: Extreme Subset Relation

Figure 4.7 shows a small yet instructive subset of ΔF estimates using original FG on Case A. Other results from this analysis are displayed in appendix C. The $\Delta\lambda = 200$, MC trials = 50 protocol seems to display the most accurate estimates and lowest statistical uncertainties. The Jarzynski estimator gives the best estimates of the single direction estimators, while BAR is most accurate overall for all of the protocols tested. The FD and symmetric estimators are persistantly very inaccurate due to the challenging nature of the entropic barrier present in case A. Only with the $\Delta\lambda = 200$, MC trials = 50 protocol does the symmetric B estimator produce estimates relatively close to the analytical free energy difference. Also, the Kofke bias measure becomes positive when all 10000 forward switches are used with the $\Delta\lambda = 200$, MC trials = 50 protocol [Wu & Kofke(2005a)].

The forwards Jarzynski calculation gives better estimates than the backwards calculation in each of the nine protocols displayed apart from the $\Delta\lambda = 10$, MC trials = 100 protocol. For the $\Delta\lambda = 10$, MC trials = 100 protocol the backwards Jarzynski estimate is better than the forwards until at least 16000 FG switches are used in estimates when accuracy is similar in either direction.

The behaviour of the forwards and backwards calculations can be explained through the phase space relationship and work distributions. The much larger phase of system A is able to access the whole of the smaller system B phase space, although switches which are very important to the exponential average may be rare due to the entropic problem of finding the much smaller system B phase space from system A. The relatively small phase space of system B is unable to access many of the system A configurations. This confers a forwards work distribution which is much wider than backwards, with a much larger $\bar{W}_{diss}$. Thus, in case A, switches important to the calulation are far more common in the forwards direction than the backwards direction while the work average is further from the analytical answer. This gives the behaviour seen with the $\Delta\lambda = 10$, MC trials = 100 and $\Delta\lambda = 50$, MC trials = 20 protocols with forwards estimates being initially poor and improving quickly to be more accurate than backwards estimates which are inaccurate and showing no change with very low statistical error. This rationale is also demonstrated and discussed by Kofke and Jarzynski [Wu & Kofke(2005a)], [Jarzynski(2006)]. This behaviour with initially poor forwards Jarzynski estimates seems to be eased when using a protocol with many $\Delta\lambda$s, as seen by comparing figures 4.7 a), b) and c). This would be expected as the change from large phase space to small is more gradual. The Kofke bias measure seems to reflect this idea that the backwards Jarzynski estimates are initially closer to the analytical result but converge slowly compared to the forwards estimates which are initially poor but converge fast. The Kofke bias predicts the backwards estimates to be initially more accurate with forwards estimates showing a relative improvement for protocols with $1x10^7$ MC trials but shows the forwards estimates becoming converged for protocols with $1x10^8$ MC trials.



(a) 1x10^7 config's, $\Delta\lambda = 200$, MC Trials = 5

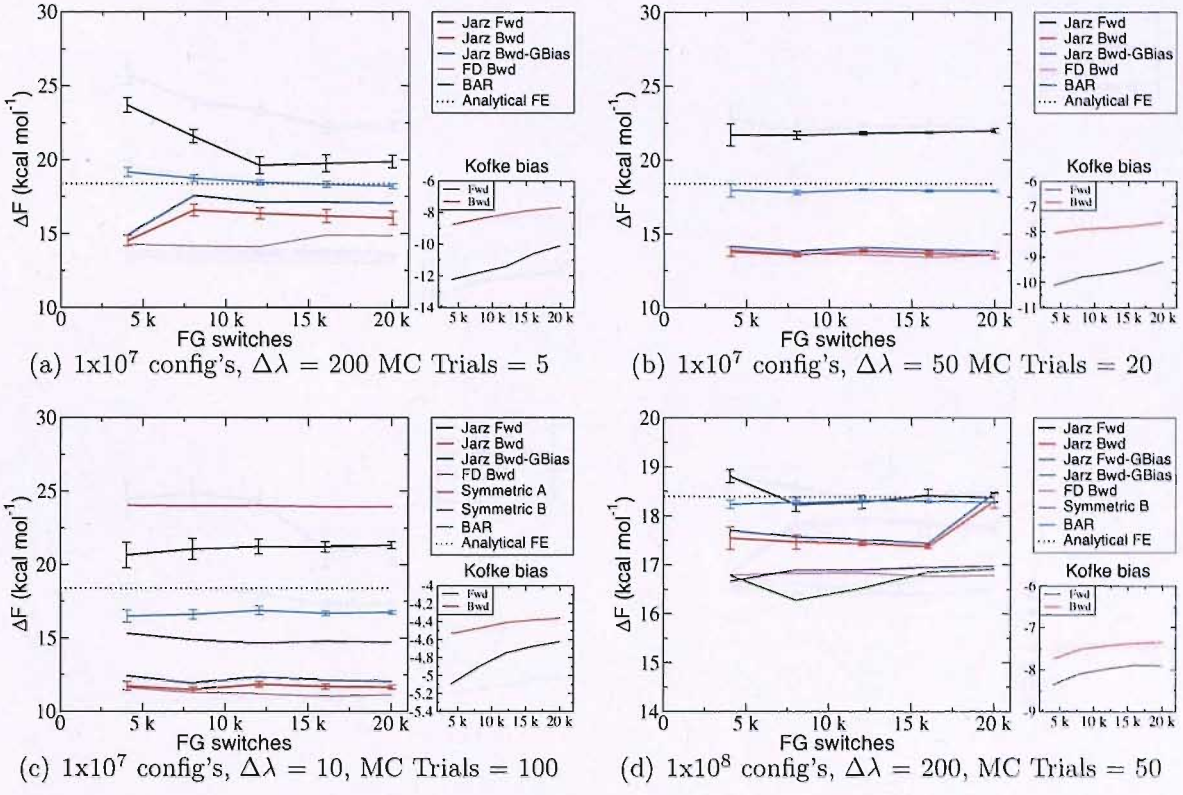
(b) 1x10^7 config's, $\Delta\lambda = 50$, MC Trials = 20

(c) 1x10^7 config's, $\Delta\lambda = 10$, MC Trials = 100

(d) 1x10^8 config's, $\Delta\lambda = 200$ MC Trials = 50

Figure 4.7: ΔF estimates with original FG using 1x10^8 and 1x10^7 configurations for case A.

λ bias results for case A with all estimators seem to converge slightly faster in comparison to original FG for the 1x10^8 protocols apart from BAR which seems almost identical (figure 4.8). However the 1x10^7 λ bias protocols seem to give similar performance to the equivalent original FG results. If the system is not well equilibrated between each λ increment, which may often be the case with larger systems, then λ bias may not offer any improvement. The Kofke bias measure seems to give negative scores to λ bias forward Jarzynski calculations that look like they are converged such as those in figure 4.8 d), while giving positive scores to similar calculations with original FG in figure 4.7 d). This suggests that the Kofke bias measure may be incorrect for these biased FG methods, although it is worth note that the Kofke bias measure does not give any indication of fully converged estimates when there are none (false positives).



(a) $1x10^7$ config's, $\Delta\lambda = 200$ MC Trials = 5

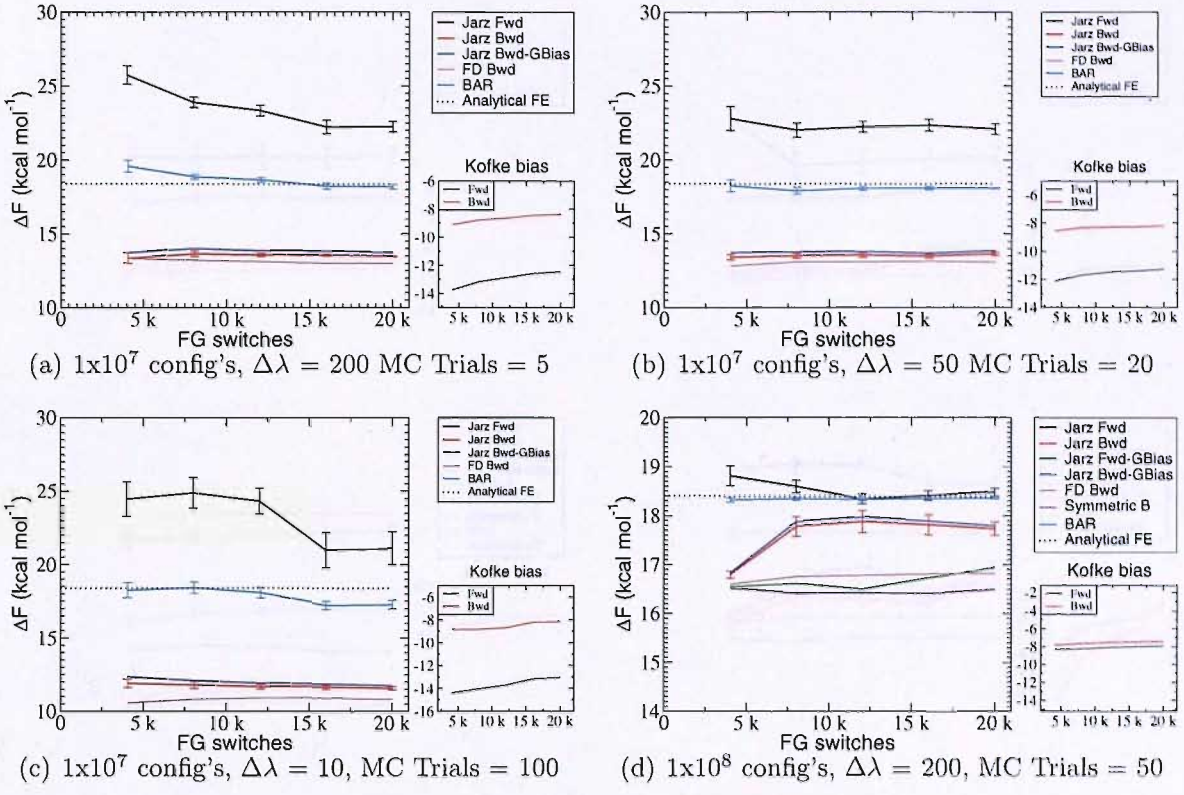
(b) $1x10^7$ config's, $\Delta\lambda = 50$ MC Trials = 20

(c) $1x10^7$ config's, $\Delta\lambda = 10$, MC Trials = 100

(d) $1x10^8$ config's, $\Delta\lambda = 200$, MC Trials = 50

Figure 4.8: ΔF estimates with λ bias FG using $1x10^8$ and $1x10^7$ configurations for case A.

Configuration bias-d shows similar trends to λ bias; there is a slight improvement, of configuration bias-d over original FG, seen with the longer protocols (figure 4.9). Configuration bias-d also performs better than original FG when few λ increments are used but not when many are used. Configuration bias shows better accuracy than original FG for BAR and Fwd Jarzynski with the $\Delta\lambda = 10$, MC trials = 100 protocol but this is reversed for the $\Delta\lambda = 200$, MC trials = 5 protocol. This is almost certainly down to the smaller number of configurational samples taken for the $\Delta\lambda = 200$, MC trials = 5 protocol as previously discussed. For configuration bias FG it may be important to use well separated configuration samples to pick from for each λ increment. Again, the Kofke bias measure fails to give a score reflecting the convergence of the longer protocols, but has no false positives.



(a) $1x10^7$ config's, $\Delta\lambda = 200$ MC Trials = 5

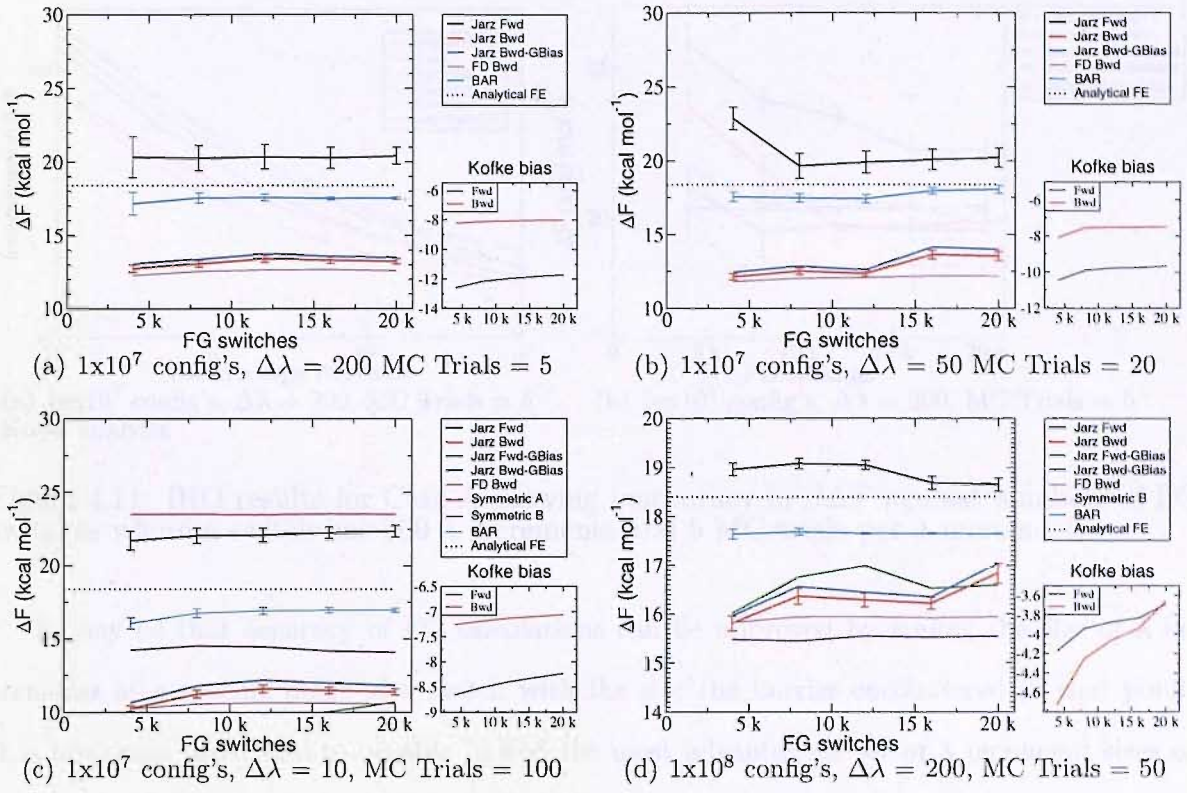
(b) $1x10^7$ config's, $\Delta\lambda = 50$ MC Trials = 20

(c) $1x10^7$ config's, $\Delta\lambda = 10$, MC Trials = 100

(d) $1x10^8$ config's, $\Delta\lambda = 200$, MC Trials = 50

Figure 4.9: ΔF estimates with configuration bias-d FG using $1x10^8$ and $1x10^7$ configurations for case A.

Hybridbias shows similar trends to λ bias. Surprisingly, estimates for the $\Delta\lambda = 200$, MC trials = 50 protocol with Hybrid bias are less accurate than for all the other methods including original FG. The BAR estimates for the $\Delta\lambda = 200$, MC trials = 50 protocol is less accurate than the forwards Jarzynski estimates. This inaccuracy in the BAR estimator is related to the slightly worse performance of estimates using only backwards switches. This may suggest that Hybrid bias is less efficient than other methods when switching from systems with a small important phase space to systems with a large important phase space.



(a) 1x10^7 config's, $\Delta\lambda$ = 200 MC Trials = 5

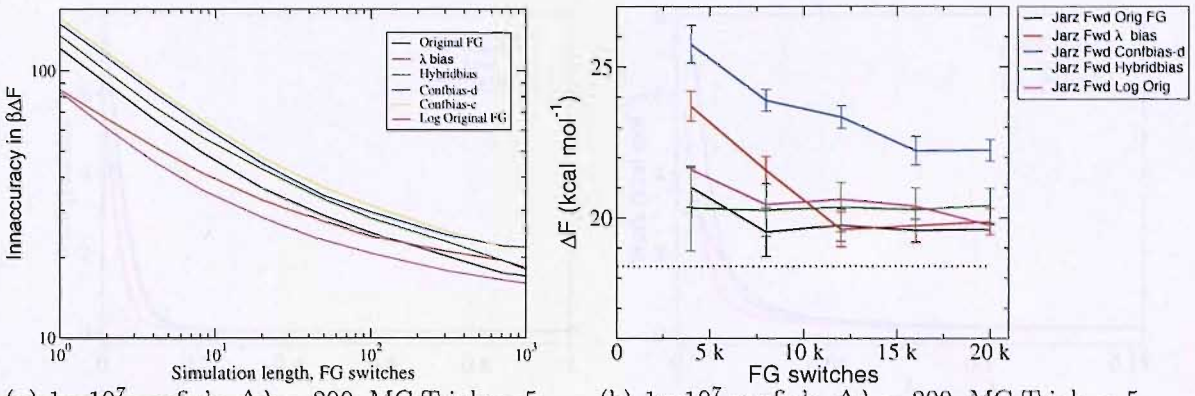
(b) 1x10^7 config's, $\Delta\lambda$ = 50 MC Trials = 20

(c) 1x10^7 config's, $\Delta\lambda$ = 10, MC Trials = 100

(d) 1x10^8 config's, $\Delta\lambda$ = 200, MC Trials = 50

Figure 4.10: ΔF estimates with hybridbias FG using 1x10^8 and 1x10^7 configurations for case A.

The advantages of the Kofke FG methods seem to decline when an FG protocol which does not allow good equilibration between λ increments is used. This is more clearly demonstrated by using the 1x10^7 config's, $\Delta\lambda$ = 200 MC Trials = 5 protocol with the analysis method of Wu and Kofke [Wu & Kofke(2005c)] (figure 4.11 a)). With this protocol all the methods have relatively similar efficiency. λ bias seems more accurate with 1 switch but not with 1000. The same data plotted together in the analysis method of the present study (figure 4.11 b)) shows that although results are very similar, original FG seems to give more accuracy with fewer FG switches.



(a) 1ex10^7 config's, $\Delta\lambda = 200$, MC Trials = 5 Kofke analysis

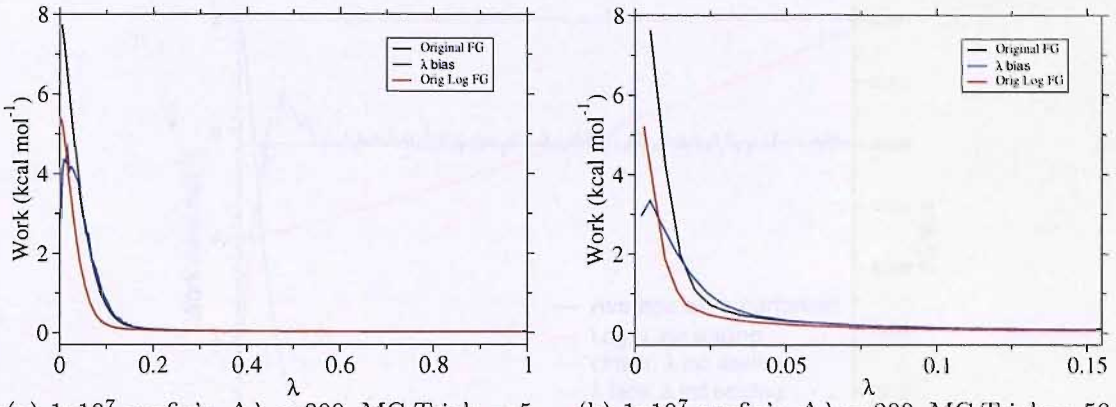
(b) 1ex10^7 config's, $\Delta\lambda = 200$, MC Trials = 5

Figure 4.11: IHO results for Case A showing inaccuracy in $\beta\Delta F$ against numbers of FG switches where a switch has 200 λ increments and 5 MC trials per λ increment.

It may be that accuracy of FG calculations can be improved by scaling the size of λ increments at a specific point of a switch with the size the barrier encountered at that point. It is obviously preferable to be able to find the most advantagous set of λ increment sizes or switching path without prior knowledge of the nature of the switch. Logarithmic paths from system A to B, rather than the usual linear paths as used in previous calculations in this study, have been investigated previously by other authors and used with TI [Shirts & Pande(2005)]. This study described slight advantages using a logarithmic perturbation path in scaling from system A to B for harmonic oscillators with a large subset relation (4.8).

$$H_\lambda(z) = H_A(z)^{1-\lambda} H_B(z)^\lambda \tag{4.8}$$

A logarithmic perturbation path used with original FG also shows a slight improvement in accuracy, compared to original and possibly λ bias FG, when used with the short protocol (figure 4.11). This comparison is further clarified by looking at the work performed at each λ increment, averaged over 1000 switches (figure4.12). Figure 4.12 suggests that λ bias is able to react to the barrier present at the beginning ($\lambda = 0$) of case A switches as the peak is reduced compared to original FG with a linear and logarithmic path. However, after the initial peak λ bias switches are very similar to original FG with a linear path and a logarithmic path shows a slightly lower work.



(a) 1x10^7 config's, $\Delta\lambda = 200$, MC Trials = 5 (b) 1x10^7 config's, $\Delta\lambda = 200$, MC Trials = 50

Figure 4.12: Distribution of average work performed over FG switches for the 200 $\Delta\lambda$, 5 or 50 MC trials per $\Delta\lambda$ protocols.

However, logarithmic scaling of the path from systems A to B may not provide the optimum path for extreme subset relation free energy differences. This can be seen clearly if we compare the size of λ increments used, by a particular method, to the areas of the path which give the largest works (figure 4.13, red, green and blue lines are plotted from the right hand side y-axis while the black line is plotted from the left hand y-axis). Figure 4.13 shows that λ bias attempts to scale its λ increments to the barriers found in a switch with very low λ increments at around $\lambda = 0$ (green line is at a low value when black line is at a high value). λ bias then seems to pick relatively large λ increments at around $\lambda = 0.1$ even though the work values incurred still seem to be relatively large.

This over sizing of λ increments at around $\lambda = 0.1$ is easily explained in terms of the λ bias algorithm: When there are a series of small λ increments at the beginning of a switch, due to heavy Rosenbluth weighting, the range from the current λ (λ_{i-1}) to a_i which the next lambda (λ_i) must be chosen from gets larger. This increases the probability of a λ increment being larger for the same Rosenbluth weight as the calculated probability density is spread over a larger range. Thus, as the average work performed in figure 4.13 (black line) eases off, the possible range of the next λ increment is large and as the Hamiltonian is more favourable than previously to large λ increments, larger λ increments ensue.

In the absence of a large barrier λ bias seems to display a degree of noise in its λ increment choices.

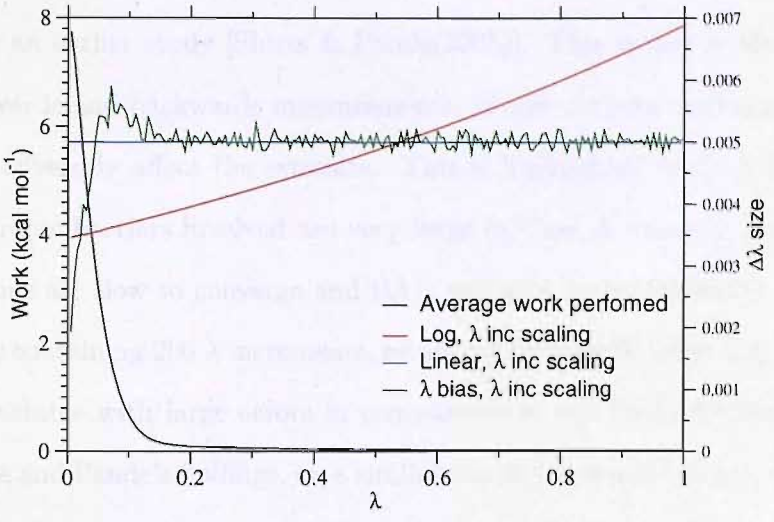


Figure 4.13: Plot of work performed and size of $\Delta\lambda$s against λ for linear and logarithmic perturbation paths using a 200 $\Delta\lambda$, 5 MC trials per $\Delta\lambda$ protocol.

It is clear from comparing FG calculations with 200 and 10 λ increments (and the same number of total configurations used) in the analysis above that increasing the number of λ increments for the same number of configurations can improve accuracy and precision. It is also clear from figure 4.3 case A and figure 4.11 that attempting to scale the size of λ increments to the size of the barriers being experienced in a switch can be more efficient in some cases.

λ bias FG may be ineffective in the case of many barriers which are not near the beginning of a switch as the final λ increment must move the system to $\lambda = 1$. If a switch has large barriers at the beginning and end, then λ bias will perform poorly. A λ bias method could work well if not constrained by the number of λ increments allowed per switch. However, it may be difficult to produce a ensemble of switches of different lengths.

In light of the problem discussed here, the case of many large barriers, original FG with few, long trajectories, many λ increments and BAR could give the best performance. Alternatively configuration bias-d could be used with few, long trajectories and large amounts of sampling between λ increments and BAR.

Case B: Slight Subset Relation

The methods investigated performed unexpectedly with case B, given the case A findings. In general, for original FG sampling with the forward direction, the Jarzynski estimator seems to perform better than BAR. This is demonstrated in figure 4.14 d). This is also true but to a lesser extent with the short protocol as seen in figures 4.14 a), b) and c). These findings are

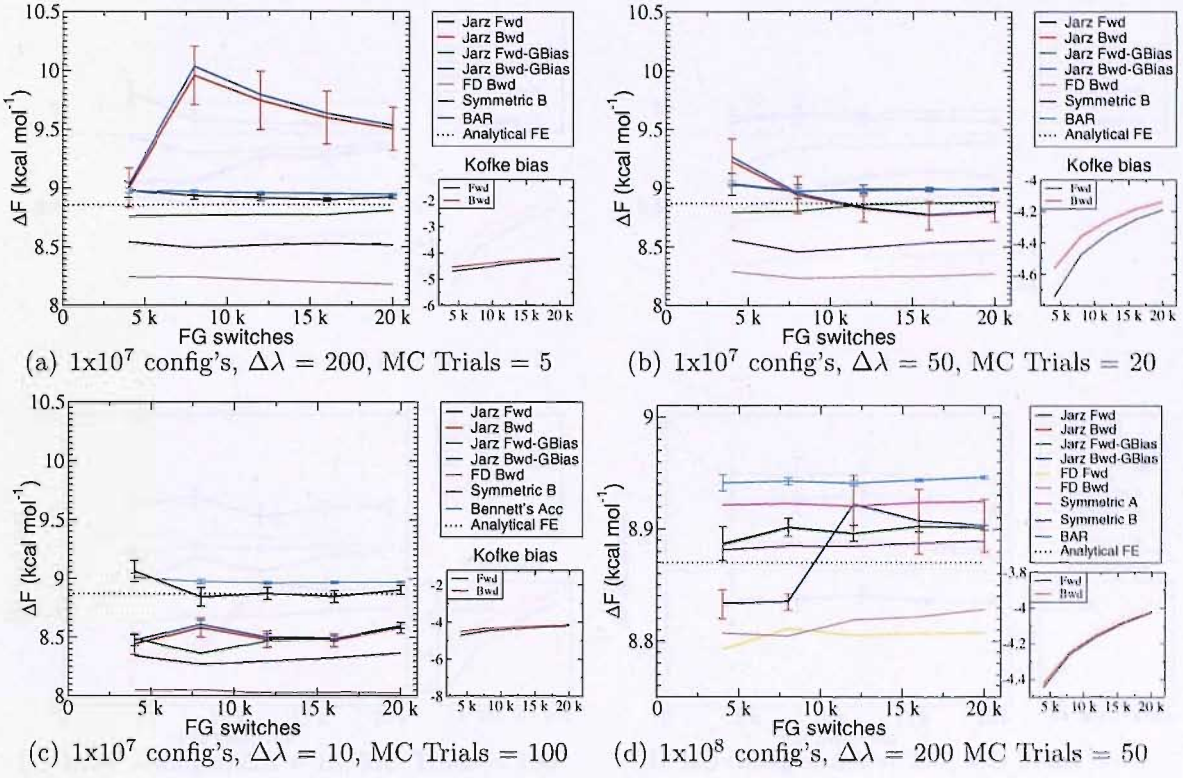
also backed up by an earlier study [Shirts & Pande(2005)]. This is due to the fact that BAR relies on both forwards and backwards measurements. If one of those directions is very poorly behaved this will adversely affect the estimate. This is highlighted in Case B as opposed to Case A as the entropic barriers involved are very large in Case A meaning both forwards and backwards directions are slow to converge and BAR will give better estimates.

With protocols containing 200 λ increments, estimates for case B, made using the backwards directions seem unstable with large errors in comparison to the forwards direction; this is in keeping with Shirts and Pande's findings, on a similar two dimensional system, that estimates in the backwards direction have a variance which tends to infinity as the ratio ω_A/ω_B is increased after $\omega_A > 2\omega_B$ [Shirts & Pande(2005)]. The reason the poor performance of the backwards direction is more marked with case B than case A is because the forwards direction is initially very poor with case A whereas it is very good with case B. A separate study also found the backwards Jarzynski estimator to not converge, where as the forwards direction converges well for systems similar to case B [Wu & Kofke(2005b)]. Thus, it could be assumed that estimates made with the backwards Jarzynski estimator are not reliable.

The different switching protocols in this original FG analysis seem to give comparable estimates with the numbers of configurations, taking into account statistical error when using the forwards Jarzynski estimator. This can be seen through study of figure 4.14 and case B original FG figures in appendix C. However, the error is slightly larger in those protocols with smaller numbers of λ increments as seen by comparing forward estimates in figures 4.14 a), b) and c).

The Kofke bias measures are similar between protocols and switching directions at around -4 to -5. This may reflect the relative ease of convergence for case B. Most of the forwards Jarzynski estimates for case B seem reasonably well converged and this is not reflected by the Kofke bias. The poorer estimators perform better relative to the better estimators in case B compared to case A. The symmetric B estimator gives good ΔF estimates when using $5\text{x}10^7$ configurations or more (figure 4.14 d)).

λ bias FG displays similar trends to original FG with forwards Jarzynski estimates generally being more accurate than BAR (figure 4.15). λ bias, Jarzynski estimates in the backwards direction seem more stable than their original FG counterparts seen by comparing figures 4.14 and 4.15. However, figure 4.15 d) shows set of backwards Jarzynski estimates which are above



(a) 1x10^7 config's, $\Delta\lambda = 200$, MC Trials = 5

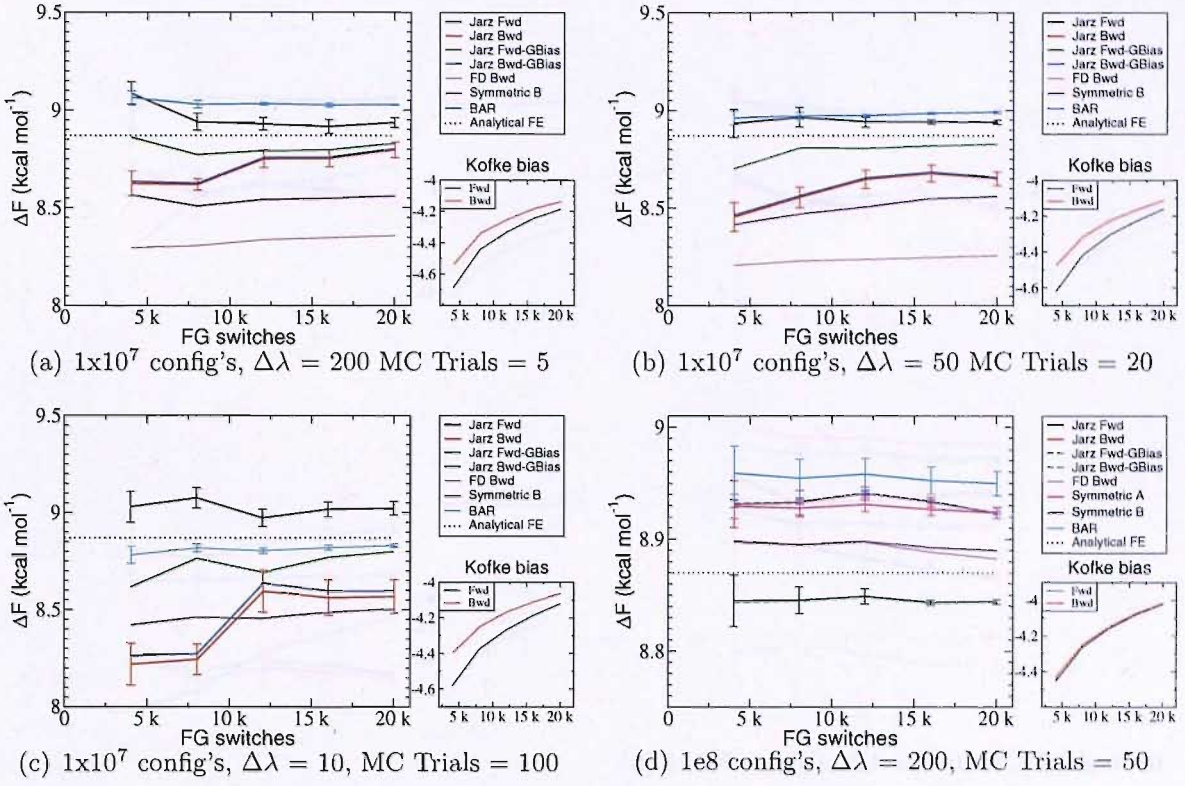
(b) 1x10^7 config's, $\Delta\lambda = 50$, MC Trials = 20

(c) 1x10^7 config's, $\Delta\lambda = 10$, MC Trials = 100

(d) 1x10^8 config's, $\Delta\lambda = 200$ MC Trials = 50

Figure 4.14: ΔF estimates with original FG using 1x10^8 and 1x10^7 configurations for for case B

the forwards estimates rather than below as would be expected. Although slightly worrying this may reflect the fact that the statistical error is the dominant form of estimator bias for case B. Comparing figures 4.14 d) and 4.15 d) it seems that the FD estimator is less accurate for λ bias than original FG.



(a) 1x10^7 config's, $\Delta\lambda$ = 200 MC Trials = 5

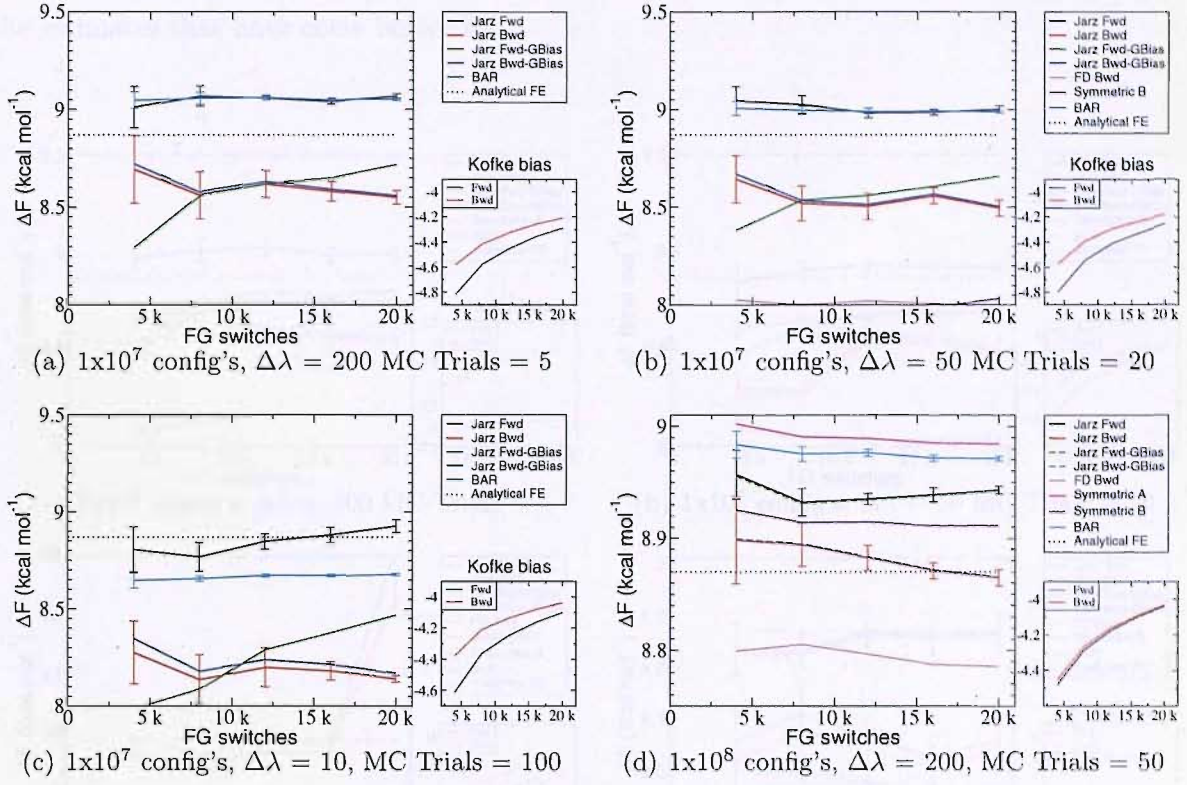
(b) 1x10^7 config's, $\Delta\lambda$ = 50 MC Trials = 20

(c) 1x10^7 config's, $\Delta\lambda$ = 10, MC Trials = 100

(d) 1e8 config's, $\Delta\lambda$ = 200, MC Trials = 50

Figure 4.15: ΔF estimates with λ bias FG using 1x10^8 and 1x10^7 configurations for case B

Figures 4.16 a) b) and c) seem to show that configuration bias-d gives slightly better forwards Jarzynski estimates with the $\Delta\lambda = 10$, MC Trials = 100 protocol than protocols with the same number of configurations but more λ increments. Also, BAR is significatntly below the forwards Jarzynski estimates in figure 4.16 c). Strangely, figure 4.16 d) has backwards Jarzynski more accurate than forwards Jarzynski. This may signify an improvement in backwards estimates with configuration bias-d, yet equally it may simply signify the small differences in these estimates and the presence of statistical errors.



(a) 1x10^7 config's, $\Delta\lambda$ = 200 MC Trials = 5

(b) 1x10^7 config's, $\Delta\lambda$ = 50 MC Trials = 20

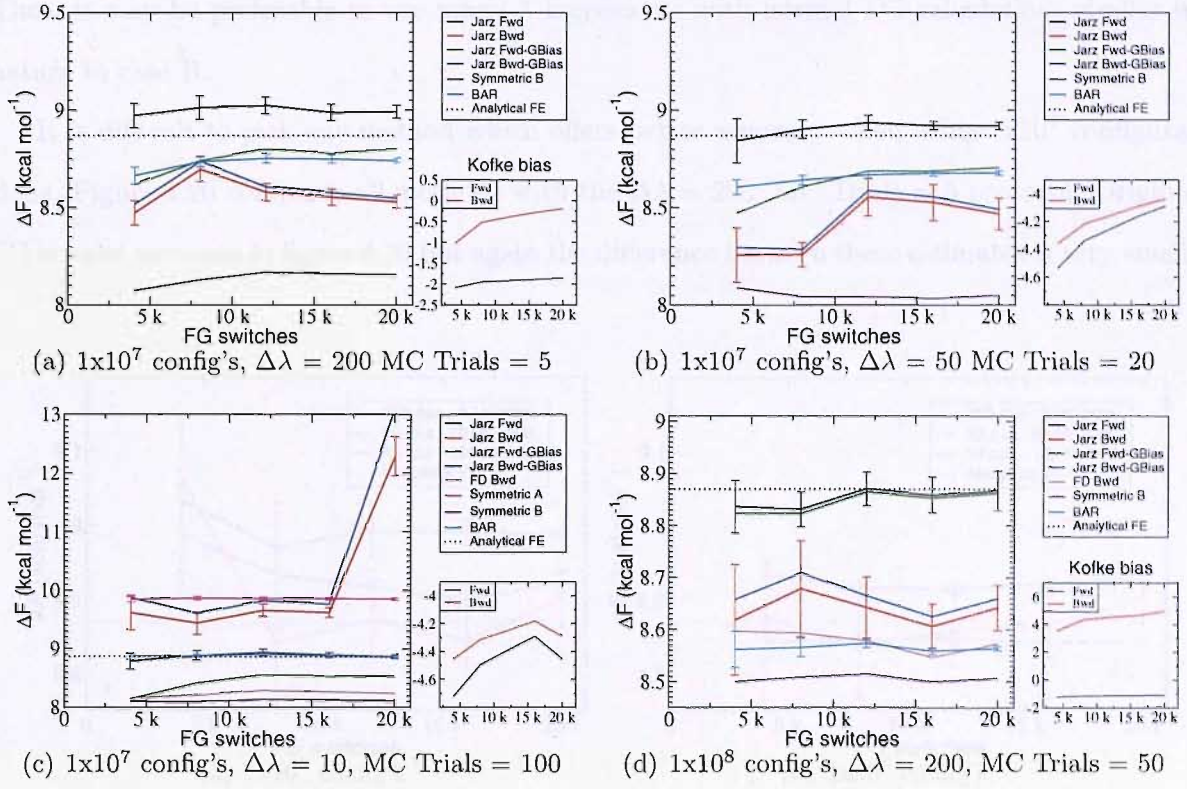
(c) 1x10^7 config's, $\Delta\lambda$ = 10, MC Trials = 100

(d) 1x10^8 config's, $\Delta\lambda$ = 200, MC Trials = 50

Figure 4.16: ΔF estimates with configuration bias-d FG using 1x10^8 and 1x10^7 configurations for case B

Hybrid bias FG produced results for case B similar in their trends to configuration bias FG (figure 4.17). Again, studying figures 4.17 a) b) and c) suggests that estimates with the $\Delta\lambda$ = 10, MC Trials = 100 protocol are slightly improved over protocols with the same number of configurations but more λ increments for forwards Jarzynski estimates. Figures 4.17 a) and b) show a set of BAR estimates between forwards and backwards Jarzyanki estimates while the forwards Jarzynski estimates seem still to be more accurate. This unpredictability in the relative placement of BAR may make it difficult to pick which estimator to use in any particular situation. Figure 4.17 c) has a set of backwards Jarzynski estimates which are above the forwards estimates, which was somthing seen in figure 4.15 d). Also, there is a large jump in value between the final and penultimate estimates of backwards Jarzynski calculation in figure 4.17 c). This again points to the instability of backwards Jarzynski estimates for case B. Figure 4.18 is a scatter plot of all work values for the $\Delta\lambda$ = 10, MC Trials = 100 protocol in the backwards direction. The two data points below -15 in figure 4.18 are responsible for the sudden jump backwards Jarzynski estimates descibed above in figure 4.15 d). If these values are removed the final backwards Jarzynski ΔF estimate then becomes -9.52 $kcalmol^1$ and falls into line with

the estimates that have come before it.



(a) $1x10^7$ config's, $\Delta\lambda = 200$ MC Trials = 5

(b) $1x10^7$ config's, $\Delta\lambda = 50$ MC Trials = 20

(c) $1x10^7$ config's, $\Delta\lambda = 10$, MC Trials = 100

(d) $1x10^8$ config's, $\Delta\lambda = 200$, MC Trials = 50

Figure 4.17: ΔF estimates with hybrid bias FG using $1x10^8$ and $1x10^7$ configurations for case B

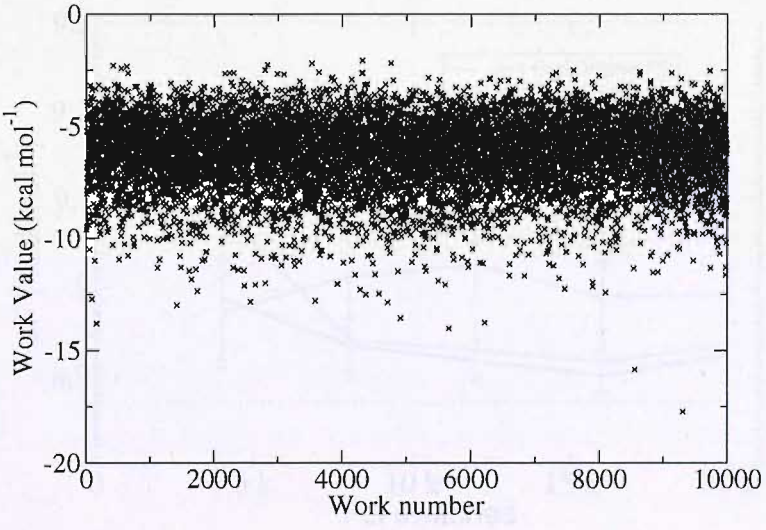
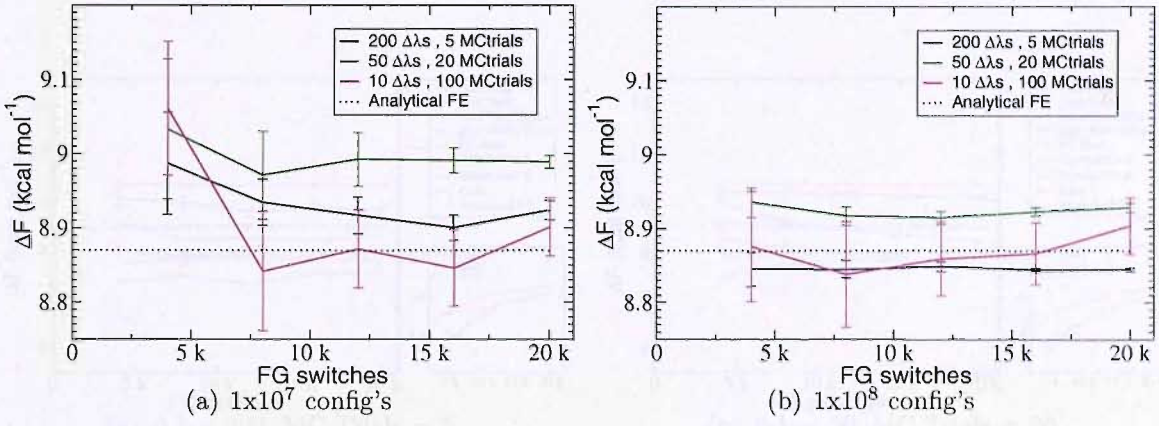


Figure 4.18: Work values for the $\Delta\lambda = 10$, MC Trials = 100 protocol in the backwards direction.

Figure 4.19 a) shows a general trend in $1x10^7$ calculations, that protocols with 10 λ increments are slightly more accurate, seen in all methds but λ bias for case B. However the

difference is small enough to be a random effect and figure 4.19 b) ($1x10^7$ configurations) does not agree. Also, the statistical error gets smaller as you increase the number of λ increments. Thus, it may be preferable to use many λ increments with normal FG calculations similar in nature to case B.

It is difficult to pick any method which offers better accuracy when using $1x10^7$ configurations. Figure 4.20 compares all methods with the $\Delta\lambda = 200$, MC Trials = 5 protocol. Original FG is most accurate in figure 4.20 but again the difference between these estimates is very small.



(a) $1x10^7$ config's (b) $1x10^8$ config's

Figure 4.19: Case B ΔF estimates using original FG with $1x10^7$ or $1x10^8$ configurations and three different switching protocols.

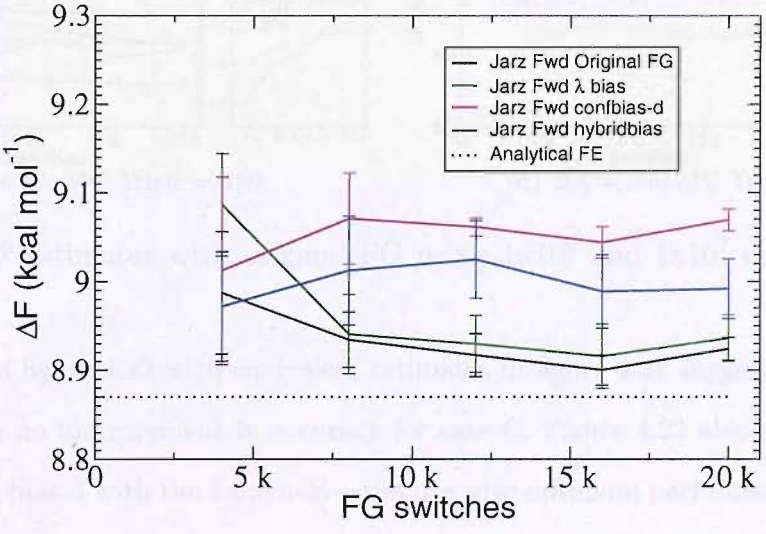
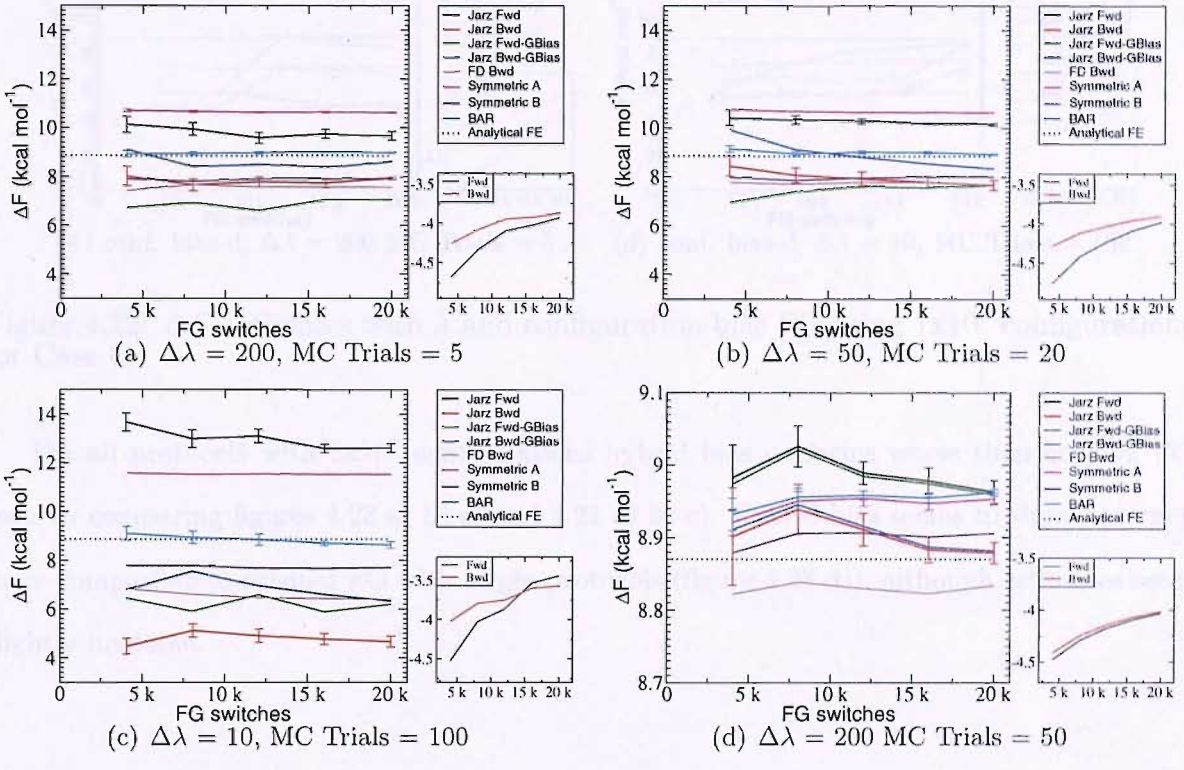


Figure 4.20: ΔF estimates for all methods with the forwards Jarzynski estimator, using $\Delta\lambda = 200$, MC Trials = 5 protocol.

Case C: Subset and Overlap Relation

Owing to the slight displacement of system B in case C, all FG methods using switches from a single direction can be inaccurate. There is an improvement in accuracy of the single direction estimators as the number of $\Delta\lambda$s increase, without extra configurations (figures 4.21 a), b), c)). It is clear that, overall BAR is the most accurate estimator for case C. Single direction Jarzynski with $1x10^8$ configurations and many $\Delta\lambda$s is able to rival BAR (4.21 d)). The symmetric estimators are not as accurate as BAR as forward and backward switches are different due to the slight phase space difference between systems A and B.



(a) $\Delta\lambda$ = 200, MC Trials = 5

(b) $\Delta\lambda$ = 50, MC Trials = 20

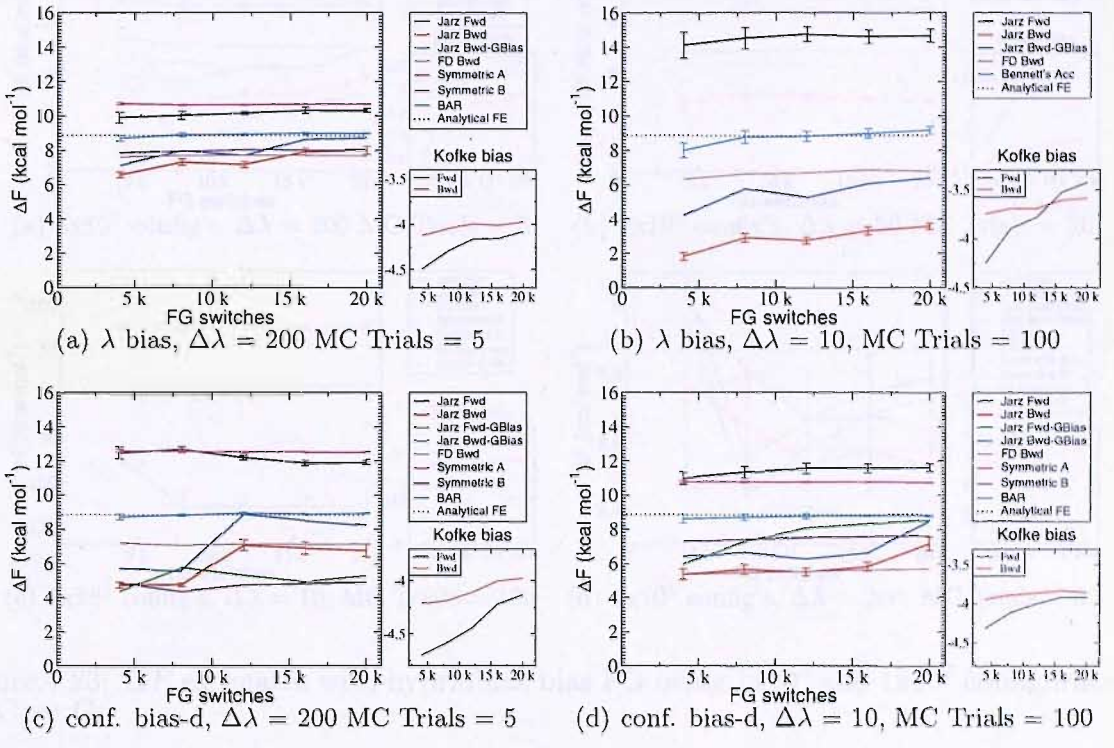
(c) $\Delta\lambda$ = 10, MC Trials = 100

(d) $\Delta\lambda$ = 200 MC Trials = 50

Figure 4.21: ΔF estimates with original FG using $1x10^8$ and $1x10^7$ configurations for case C

Comparison of figure 4.22 with equivalent estimates in figure 4.21 suggests that the biased FG methods offer no improvement in accuracy for case C. Figure 4.22 also shows that λ bias and configuration bias-d with the Jarzynski estimator give optimum performance with different switching protocols. λ bias gives good perfromance similar to original FG with the $\Delta\lambda$ = 200 MC Trials = 5 protocol, while configuration bias is relatively poor with this protocol (figure 4.22). configuration bias-d performs best with the $\Delta\lambda$ = 10 MC Trials = 100 protocol (figure 4.22). This preference in configuration bias-d is probably due to a lack of sampling between

$\Delta\lambda$s meaning all configuration choices are very similar.



(a) λ bias, $\Delta\lambda$ = 200 MC Trials = 5

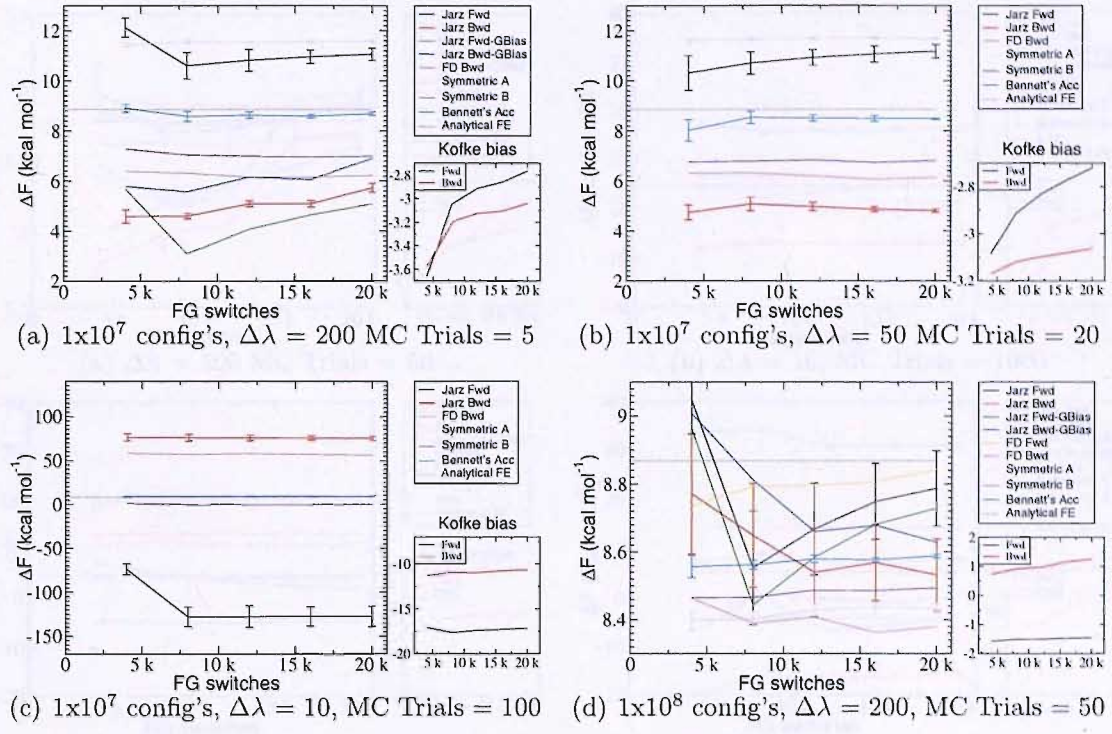
(b) λ bias, $\Delta\lambda$ = 10, MC Trials = 100

(c) conf. bias-d, $\Delta\lambda$ = 200 MC Trials = 5

(d) conf. bias-d, $\Delta\lambda$ = 10, MC Trials = 100

Figure 4.22: ΔF estimates with λ and configuration bias FG using 1x10^7 configurations for Case C

For all protocols with 1x10^7 configurations hybrid bias performs worse than original FG seen by comparing figures 4.23 a) b) c) and 4.21 a) b) c). Hybrid bias seems to show accuracy more comparible to original FG with larger protocols (figure 4.23 d)), although estimates seem slightly unstable.



(a) $1x10^7$ config's, $\Delta\lambda = 200$ MC Trials = 5

(b) $1x10^7$ config's, $\Delta\lambda = 50$ MC Trials = 20

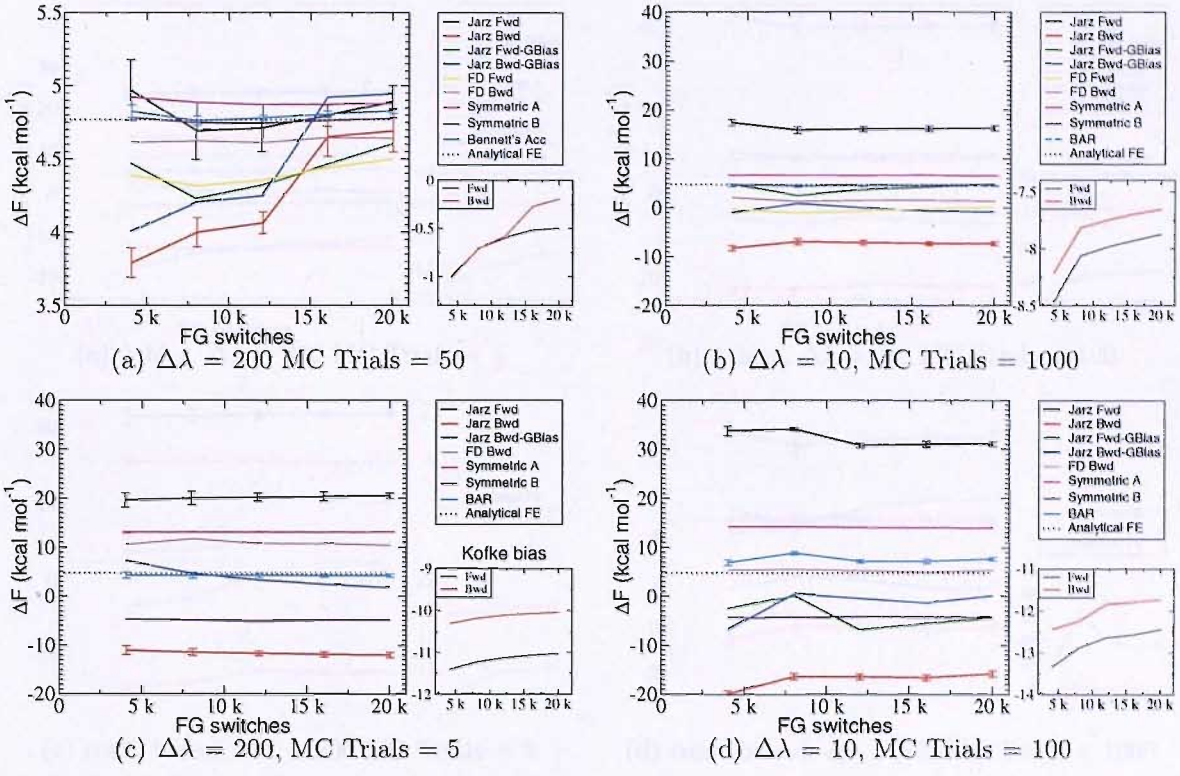
(c) $1x10^7$ config's, $\Delta\lambda = 10$, MC Trials = 100

(d) $1x10^8$ config's, $\Delta\lambda = 200$, MC Trials = 50

Figure 4.23: ΔF estimates with hybridbias bias FG using $1x10^8$ and $1x10^7$ configurations for Case C

Case D: Slight Subset and Non-Overlap Relation

Analysis of case D results show the same trends described for case C. Again, overall, BAR is the most accurate estimator but the symmetric estimators are almost as accurate as BAR for case D (figure 4.24). The improvement in the symmetric estimators is due to the forward and backward directions being more similar in case D, i.e. the important phase spaces of systems A and B are of a more similar size.



(a) $\Delta\lambda = 200$ MC Trials = 50

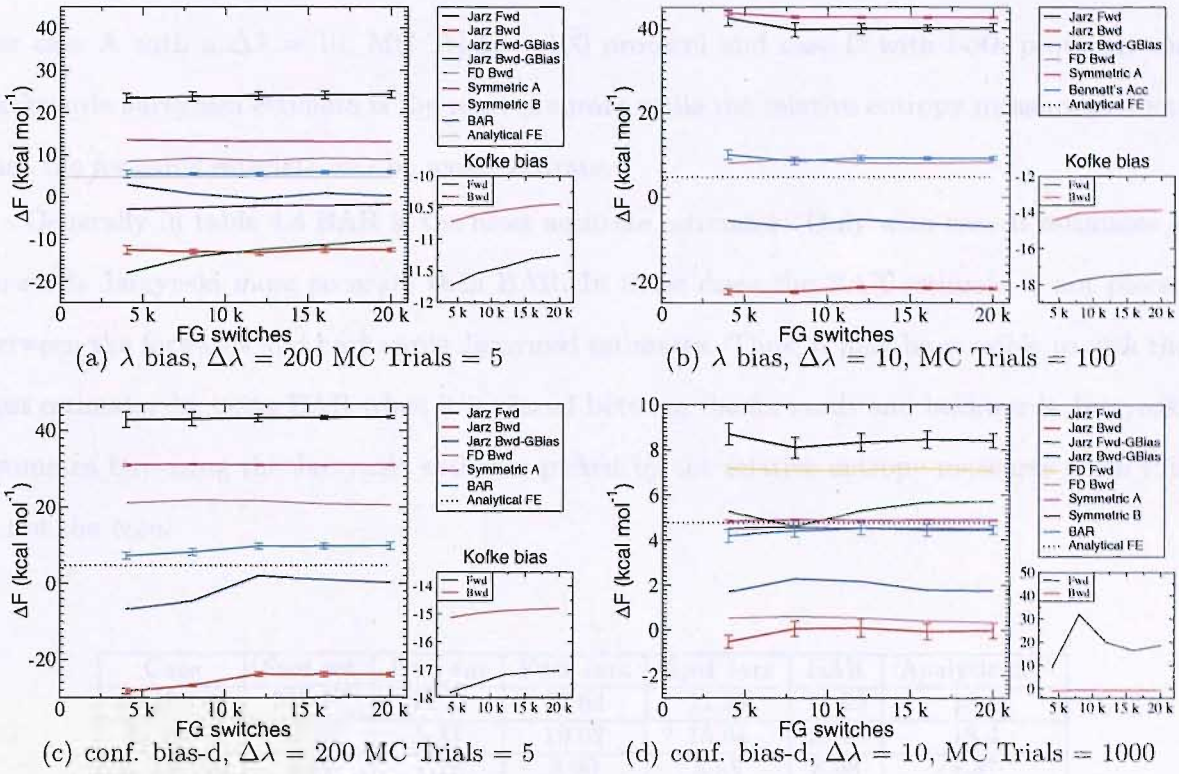
(b) $\Delta\lambda = 10$, MC Trials = 1000

(c) $\Delta\lambda = 200$, MC Trials = 5

(d) $\Delta\lambda = 10$, MC Trials = 100

Figure 4.24: ΔF estimates with original FG using $1x10^8$ and $1x10^7$ configurations for Case D

The biased FG methods perform slightly worse than original FG with single direction estimators (figure 4.25). One exception is configuration bias-d FG with protocols with few $\Delta\lambda$s and many sampling steps in between. Figure 4.25 d) shows improvement of single direction estimates over the equivalent original FG estimates in figure 4.24 b). This is less marked for protocols with fewer configurations (see appendix C). When BAR is used with biased FG methods, accuracy is comparable to original FG (figure 4.25). Hybrid bias FG shows similar behaviour to λ bias (see appendix C).



(a) λ bias, $\Delta\lambda$ = 200 MC Trials = 5

(b) λ bias, $\Delta\lambda$ = 10, MC Trials = 100

(c) conf. bias-d, $\Delta\lambda$ = 200 MC Trials = 5

(d) conf. bias-d, $\Delta\lambda$ = 10, MC Trials = 1000

Figure 4.25: ΔF estimates with λ bias and configuration bias-d bias FG using 1x10^8 and 1x10^7 configurations for Case D

Which estimator should be used when?

It is clear that different estimators often provide the best estimates for different cases and protocols. How do we pick between estimators when we do not know the analytical answer? The Kofke bias measure mentioned earlier attempts to measure the convergence of a calculation [Wu & Kofke(2005a)]. The analysis of figure 4.6 showed that this measure cannot be relied upon in all cases.

Another possible method of picking between estimators is to use the relative entropy measures of Wu and Kofke [Wu & Kofke(2005a)]. As discussed in chapter 3, these measures attempt to give an idea of relative size of important phase space of each end point system. The Jarzynski estimate with the highest relative entropy measure should be the direction starting from the system with the biggest phase space and thus converge faster. Table 4.4 compares these relative entropy measures with the relevant estimates for cases A-D with $\Delta\lambda$ = 10, MC Trials = 100 and $\Delta\lambda$ = 200, MC Trials = 5 switching protocols.

Correlating the highest relative entropy measure with the most accurate Jarzynski estimate shows that this method of picking the most accurate Jarzynski estimate does not work well.

For case A with a $\Delta\lambda = 10$, MC Trials = 100 protocol and case D with both protocols the backwards Jarzynski estimate is the most accurate while the relative entropy measure predicts that the forwards estimate will be most accurate.

Generally in table 4.4 BAR is the most accurate estimator. Only with case B estimates is forwards Jarzynski more accurate than BAR. In those cases the BAR estimate is not placed between the forwards and backwards Jarzynski estimates. Thus, it may be possible to pick the best estimator by using BAR when it is placed between the forwards and backwards Jarzynski estimates but using the Jarzynski estimate picked by the relative entropy measures when this is not the case.

Case	Fwd ent	Bwd ent	Fwd Jarz	Bwd Jarz	BAR	Analytic ΔF
A, 10-100	269.77	36.22	28.64	11.29	19.99	18.4
A, 200-5	63.07	5.17	19.62	15.84	18.25	18.4
B, 10-100	4.44	1.78	8.90	8.58	8.98	8.87
B, 200-5	3.23	2.12	8.92	9.50	8.95	8.87
C, 10-100	13.62	7.87	12.61	4.66	8.63	8.87
C, 200-5	9.33	5.68	9.67	7.96	9.00	8.87
D, 10-100	33.99	21.32	31.10	-15.70	7.73	4.77
D, 200-5	37.55	19.70	20.50	-11.84	4.36	4.77

Table 4.4: Table of forward and backwards relative entropy measures and inaccuracies in Jarzynski and BAR estimates of cases A-D free energy differences with $\Delta\lambda = 10$, MC Trials = 100 and $\Delta\lambda = 200$, MC Trials = 5 switching protocols.

What is the optimum switch length?

Previous studies in the literature tell us, it is more efficient to use fewer longer FG switches [Gore *et al.*(2003)Gore, Ritort & Bustamante], [Ytreburg & Zuckerman(2004)]. Figure 4.26 shows the difference in efficiency using switches of 10,000 configurations (200 λ increments with 50 inter λ increment MC trials) and 1000 configurations (200 λ increments with 5 inter λ increment MC trials). The difference in efficiency between long and short switches can be quite different for the cases A-D. For case B there is no difference in efficiency after 10^6 configurations, while for case D there is a very large difference in efficiency. Therefore when evaluating ΔFs which have barriers only of the form found in case's A and B the advantage of extra parallelisation may be more important for calculation of this type.

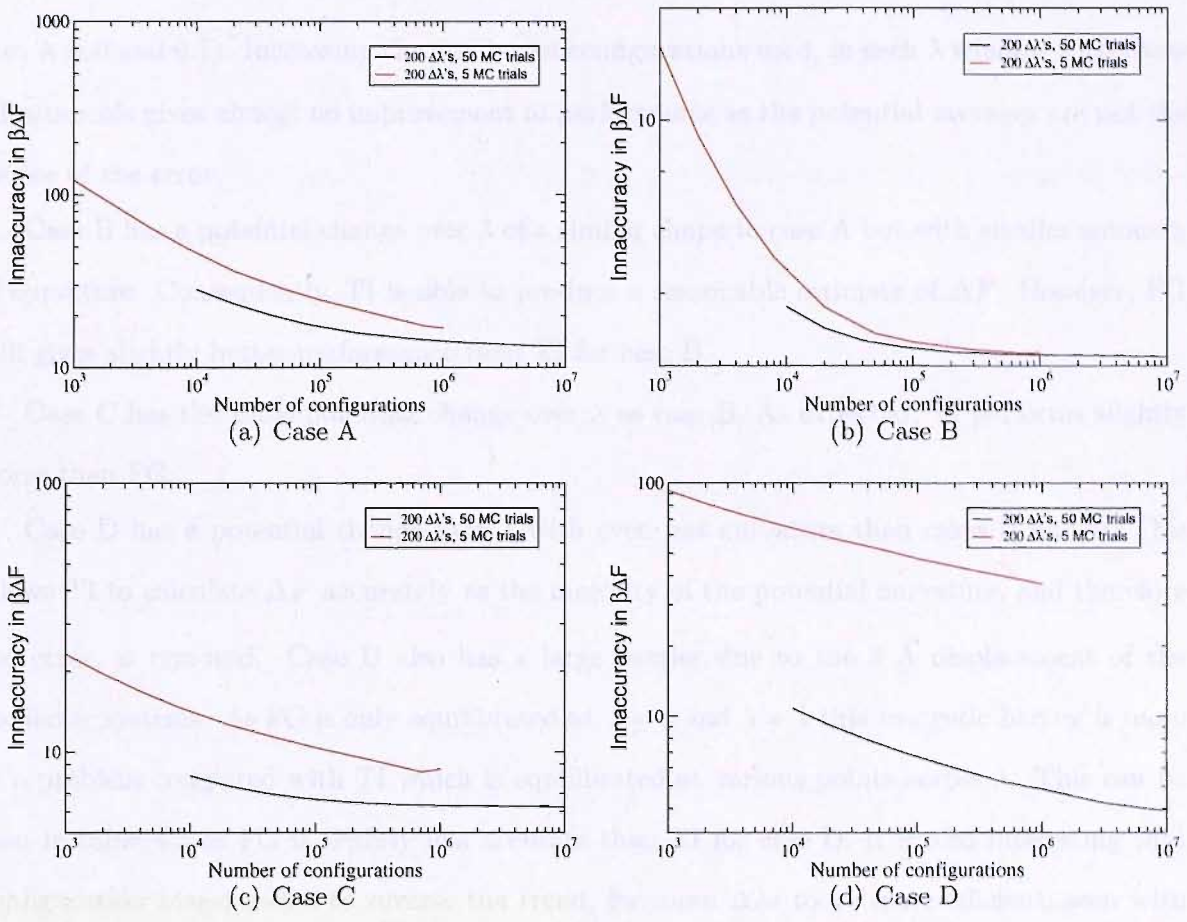


Figure 4.26: IHO results for Original FG showing inaccuracy in $\beta\Delta F$ against numbers of MC configurations where a switch has 200 λ increments and either 5 or 50 MC trials per λ increment. Both calculations use a maximum of 1e7 configurations.

4.4.3 Comparison with TI

Our implementation of FDTI uses the trapezium rule to integrate over the potential gradients found from simulation. If the potential over the λ coordinate undergoes a large amount of localised curvature our TI method will be error prone, as this curvature will not be detected. Accurate TI calculations rely on a smooth free energy gradients across λ. This analysis will investigate TI with 11 and 51 λ increment collection windows using a total of $1x10^7$ MC configurations. In table 4.5, "wins" is the number of windows used by TI and the numbers after the acronym FG relate to the number of $\Delta\lambda$s and MC trials between each λ increment, in that order.

For Case A, TI compares quite poorly with both original and configuration bias-d FG (table 4.5). TI calculations done using 11 windows display very large inaccuracies. This is because of a high level of curvature in the potential between the first and second $\Delta\lambda$ window simulations

(i.e. $\lambda = 0$ and 0.1). Increasing the number of configurations used, in each λ window with these TI protocols gives almost no improvement in performance as the potential averages are not the source of the error.

Case B has a potential change over λ of a similar shape to case A but with smaller amounts of curvature. Consequently, TI is able to produce a reasonable estimate of ΔF. However, FG still gives slightly better performance than TI for case B.

Case C has the same potential change over λ as case B. As expected, TI performs slightly worse than FG.

Case D has a potential change over λ with even less curvature than cases C and B. This allows TI to calculate ΔF accurately as the majority of the potential curvature, and therefore the error, is removed. Case D also has a large barrier due to the 3 Å displacement of the oscillator systems. As FG is only equilibrated at $\lambda = 0$ and $\lambda = 1$ this energetic barrier is more of a problem compared with TI which is equilibrated at various points across λ. This can be seen in table 4.5 as FG is slightly less accurate than TI for case D. It is also interesting that configuration bias-d seems to reverse the trend, for more $\Delta\lambda$s to be more efficient, seen with original FG. This agrees with earlier results suggesting configuration bias-d perfroms best with few λ increments with large amounts of sampling between each λ increment.

Case	TI (11 wins)	TI (51 wins)	FG 10-100	FG 50-20	FG 200-5
Case A, origFG	46.70	6.76	1.59	0.07	0.15
Case A, conf-d	46.70	6.76	1.09	0.30	0.20
Case B, origFG	0.69	0.03	0.11	0.13	0.07
Case B, conf-d	0.69	0.03	0.19	0.12	0.18
Case C, origFG	0.70	0.04	0.25	0.07	0.14
Case C, conf-d	0.70	0.04	0.06	0.05	0.10
Case D, origFG	0.00	0.21	2.96	2.26	0.41
Case D, conf-d	0.00	0.21	0.16	3.50	5.09

Table 4.5: Comparison of inaccuracy in ΔF $(kcalmol^{-1})$ for Cases A-D using TI and BAR

In the case of a demanding system each λ window takes so long to provide a converged potential average that literature studies generally use no more than 21 windows [Price & Jorgensen(2000b)], [Michel *et al.*(2006)Michel, Verdonk & Essex], [Pearlman & Charifson(2001)]. In some cases this may not be enough to capture the true nature of the potential gradient and can introduce error. Methods which are able to sample along the λ coordinate have an advantage as they experience all of the barriers involved in perturbing the system. This view

is backed up by Shirts and Pande [Shirts & Pande(2005)]. Shirts and Pande presented a limiting moment statistical analysis of the variance and bias of FG and TI on simple harmonic oscillator systems. They found that TI had higher standard error for IHO, phase space subset perturbations, which increased at a faster rate than FG as the ratio of ω_B/ω_A was increased. The present study clearly shows that FG should be used in preference to TI when evaluating free energy differences of IHO systems.

This analysis has highlighted a deficiency of FG methods. The system has to react to the changes made to the system by switching. It may increase the performance of FG methods if a λ switch is split up with intermediate starting points and then the ΔF of each part of the switch is summed [Hummer(2001)], [Shirts & Pande(2005)]. This switch splitting methodology should decrease the amount of dissipated work the system must undergo in switching, and bring all simulations closer to equilibrium. Table 4.6 shows results from FG calculations for cases A and D comparing calculations where switches are split into ten equally smaller switches (FG-BY10) and calculations where switches are not split at all. The labels of each column refer to the switching protocol used, with the first number refering to the number of λ increments and the second number after the "-" refers to the number of MC trials between each λ increment. All calculations, in table4.6, used $1x10^7$ configurations. Thus, the 20-5 ($\div 10$) calculations used 200000 switches while all others used 20000.

Case	10-100 (BY10)	10-100	200-5 (BY10)	20-5 (BY10)	200-5
Case A, fwd	0.30	10.24	0.20	0.52	1.22
Case A, bwd	1.73	7.11	0.76	2.59	2.56
Case A, BAR	0.19	1.59	0.37	0.42	0.15
Case D, fwd	0.24	26.33	0.21	0.13	15.73
Case D, bwd	0.07	20.47	0.40	0.24	16.61
Case D, BAR	0.58	2.96	0.73	0.7	0.41

Table 4.6: Comparison of inaccuracy in ΔFs estimated with FG for Cases A and D with the λ switches split into 10, and not split at all.

Splitting FG switches into 10 separate FG calculations has improved the performance for both cases A and D. This is especially marked for the single direction Jarzynski estimates. Case D estimates using 10 $\Delta\lambda$s in the forwards direction improves in accuracy by an order of magnitude when the λ coordinate is split into 10 with switches kept at the same length and out performs BAR using the same protocol. This may be because the variance of BAR does

not reduce as quickly as with Jarzynski as the calculations become better behaved [Shirts & Pande(2005)].

It is interesting to compare the estimates for case A which use the 200-5 BY10 and 20-5 BY10 protocols from table 4.6. These protocols give similar results except the backwards Jarzynski estimate for 20-5 BY10 is far less accurate and similar to the 200-5 estimate. This suggests that for case A backwards Jarzynski estimates, the improvement seen for 200-5 BY10 comes from the effective increase in length of switching rather than the extra equilibrated switching points across λ. If the same protocols are compared for case D, it is clear that to 20-5 BY10 estimates are much improved compared to the 200-5 estimates. Thus, the extra equilibrated points across λ are more important in this case.

Comparing BAR estimates for BY10 and not BY10 protocols seems to suggest that BAR estimates, with a protocol with many λ increments, are slightly more accurate without splitting up the λ coordinate.

4.5 Conclusions: When to use which methods?

The new sampling methods investigated here, originally presented by Kofke *et al.* offer advantages for the test protocols originally presented [Wu & Kofke(2005c)]. λ bias offers large improvements in accuracy for case A using the original test protocol (fig 4.3). When other protocols and estimators are explored these advantages do not necessarily remain. Original FG sampling is more efficient than any of the biased FG methods as long as the size of λ increments is kept very small. The λ bias method must be altered to give λ increments proportional to the barriers throughout a switch, to be effective in all cases. The configuration bias-d method needs large amounts of sampling between λ increments, so that the possible choice of configurations are less correlated. This means that it is often less efficient than original FG.

Of the estimators investigated here, it is clear that for difficult IHO switches BAR is the most efficient. Using Jarzynski in the direction of larger phase space to smaller phase space may be more accurate in the case of a slight subset relation. Jarzynski is also more accurate when the λ coordinate is split into ten and the resultant ΔFs are added together. From these IHO results it maybe possible to set some simple rules to pick between estimators when the correct ΔF is not known. If a BAR estimate is placed between the forwards and backwards

Jarzynski estimates then this is probably the most accurate estimate. If a BAR estimate is not placed between the forwards and backwards Jarzynski estimates then the estimate in the direction of the highest $\bar{W}_{diss}$ should be chosen. There are cases where this method of picking estimates does not work. Specifically, figure 4.16 c) and 4.17 a) although these problem cases are both biased FG examples.

When attempting to think about these results in terms of the best protocol for protein-ligand systems, it is less straight forward as these systems have extra barriers which are not linked to phase space relationships but to interactions between different constituents of the system. Switches between similar ligands in the solvent or protein environments are unlikely to include the large localised amounts of curvature found in case A switches. However, these types of protein switches could involve many barriers due to interactions between the ligand and its environment. This almost certainly means an FG switch would be pushed far from equilibrium. Thus, it seems sensible to use the splitting methodology discussed here.

Splitting an FG switch into smaller parts also makes sense from a computing point of view as it allows more parallelisation. The bottleneck for FG calculations is the production of the starting configurations. Splitting the calculation into ten smaller calculations allows the production of the starting configurations to be parallelised as well. It is also necessary to take a computing point of view when selecting the length of FG switches. The analysis presented here suggests that the longest switches possible should be used. This detracts from the main advantage of FG which is its possible parallelisation. This limits the length of FG switches and a compromise position must be found.

Chapter 5

Solute-Solvent test systems

5.1 Introduction

The IHO systems in the previous chapter allowed the thorough investigation of the accuracy and efficiency of free energy methods for different phase space relationships. The results and concepts taken from the investigation of IHO systems can be applied to the more complex systems which hold more interest and real world application. However, owing to their more complex nature, these systems have extra interactions which must be studied to understand the best approach to calculating their free energy differences.

The protein-ligand systems, which are the ultimate interest of this work, are invariably in a water environment. Thus, one of the extra interactions and possible causes of free energy differences and barriers are the interactions of a solute with a solvent (in this case water). The free energy difference due to rearrangement of water is difficult to capture due to the large number of degrees of freedom and possible long range electrostatics.

Here we have investigated the performance of FG methods in calculating free energy differences with two relatively simple solute-solvent systems. First the free energy of charging of a sodium ion in water and then the relative hydration free energy of water and methane (perturbation of water to methane in water).

5.2 Sodium charging

5.2.1 Description of calculation

Calculating the free energy difference due to ion hydration is a well known test for free energy methods. Of ion hydration calculations, the hydration of a sodium ion in water is possibly the most commonly used test system in the literature [Hummer(2002)], [Ytreburg *et al.*(2006)Ytreburg, Swendsen & Zuckerman], [Han *et al.*(2001)Han, Kim, Mhin & Son]. This is due to its simplicity, and the well behaved nature of the free energy changes involved.

The calculation of sodium ion hydration free energy can be split into two to make it easier. First, the free energy of cavity creation and any solvent dispersive effects for a van der Waals sphere with the same Lennard-Jones parameters as sodium; second, the free energy due to turning on the charge of the sodium ion. Many studies decide to ignore the first calculation (cavity creation and dispersive effects) and simpily study the sodium ion charging. This is due to the relative size of these free energy differences. The free energy due to sodium charging is far larger than that due to growth of a sodium sized Lennard-Jones sphere, and therefore the latter calculation may be ignored without a significant loss of accuracy. This study will concentrate on the free energy due to the charging of a sodium ion. Also, comparison to experimental results for sodium ion hydration is not made, as the simulation software used is unable to reproduce the long range electrostatics as discussed below. This test is used simply as an initial validation and comparison for our FG implementations. Results are compared to TI and similar studies from the literature.

Our study will initially attempt to repeat the calculations of Hummer [Hummer(2002)]. Hummer used the Jarzynski, fluctuation-dissipation (FD), and their symmetric varients to calculate the free energy of charging sodium in water. Studies of ion charging systems have suggested that fluctuations in the electrostatic energy near the solute are approximately gaussian and correspond to a linear response regime [Hummer *et al.*(1996)Hummer, Pratt & Garcia], [Hummer *et al.*(1998)Hummer, Pratt & Garcia], [Ashbaugh & Wood(1997)], [Levy *et al.*(1991)Levy, Belhadj & Kitchen], [Lynden-Bell & Rasaiah(1997)] [Hummer(2002)]. The change in free energy due to charging a sodium ion is purely electrostatic with a large portion of this being the solute-solvent interactions. Thus, the FD estimator and its symmetric varient should perform well, as they rely heavily on the FG work distribution being very close to gaussian.

Our sodium ion is represented as a Lennard-Jones sphere with a point charge and parameters as used by Straatsma and Berendsen in their 1988 study, and subsequantly by Hummer [Straatsma & Berendsen(1988)]. Owing to studies carried out on system size effects of this calculation, Hummer was able to justify using a very small solvent box [Hummer *et al.*(1996)Hummer, Pratt & Garcia], [Hummer(2002)]. This allows very fast sampling of the system which enables exhastive protocols to be used. A periodic box of 64 simple point charge (SPC) water molecules with all dimensions 12.49 Å was used to solvate the ion [Berendsen *et al.*(1981)Berendsen, Postma, van Gunsteren & Hermans]. These parameters are identical to those of Hummer, as is the method of coupling the simulation to λ. Hummer used the Ewald summation for long range electrostatics and corrected the resulting free energies for interactions of the solute with other periodic solute images [Ewald(1921)]. As a modified version of ProtoMS 2.1 was used for these simulations, Ewald was not available and a residue based, non-bonded cutoff was used at 6.2 Å and smoothed for a further 0.3 Å [Woods & Michel(2005)]. The simulation parameters used in this study are summarised in table 5.1.

Parameter description	Setting
Ensemble	NVT
Temperature	25 °C
Pressure	1 atm
Boundary conditions	Orthorhombic periodic
Non-bonded cutoff	6.2 Å
Solute/solvent trials ratio	1/64
Max solute translation/rotation	0.15 Å/ 0.0 °
Max solvent translation/rotation	0.15 Å/ 15.0 °
Preferential sampling centre	Sodium ion
Preferential sampling parameter	200

Table 5.1: Simulation parameters for Sodium charging simulations

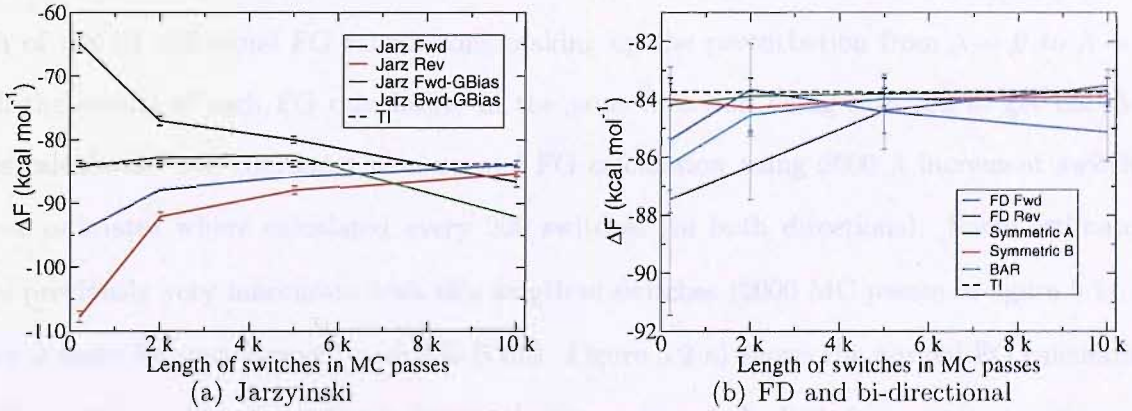
The system set-up used in the present study was expected to give very different free energies to those of Hummer [Hummer *et al.*(1996)Hummer, Pratt & Garcia], due to the difference in the handling of long range electrostatics. The difference between an Ewald and cutoff electrostatic potential was investigated with a very similar sodium charging calculation by Ashbaugh and Wood in a previous study and the resulting free energies were found to differ by 20 % [Ashbaugh & Wood(1997)]. The discrepancy was found to be due to effects at the boundary of the potential cutoff. Therefore, this study will be solely concerned with an efficiency comparsion of

FG estimators and TI when evaluating solvent reorganisation.

The system was equilibrated for 100,000 passes at λ is 0 and 1 where each pass is as many MC trials as there are molecules/residues in the system (in this case 65). Then 500 FG starting configurations where taken, one every 100 passes. Calculations with 500 FG switches in both directions, of 200, 2000, 5000, and 10,000 passes were run. All calculations in this study used switches which contain a λ increment every MC pass unless otherwise stated.

5.2.2 Sodium charging results

Once FG simulations were completed, all the estimators discussed here were applied to the data. These ΔF estimates were compared to FDTI estimates. FDTI was run using the same protocol as Hummer, with 3 FDTI simulations of 680,000 passes resulting in a ΔF of -83.7 $\pm$ 0.5 kcal.mol^{-1} [Hummer(2002)]. This TI result uses similar amounts of configurations to the FG protocol using 2000 λ increments. Figure 5.1 shows estimates from all estimators. The legend of figure 5.1 is the same as used in chapter 4 with "Jarz" refering to the Jarzynski estimator and -GBias refering to the Jarzynski estimator corrected by the bias as calculated by Gore *et al.* (2003)(see table 3.2). All statistical errors in calculations in this section were calculated with the block variance method discussed in the FG background chapter with 10 blocks.



(a) Jarzyinski (b) FD and bi-directional

Figure 5.1: Free energies of sodium charging in water, comparing FG estimators to TI. Figure (a) shows Jarzynski and Jarzynski-Gore bias. Figure (b) shows FD, and all estimators using switches in both forwards and reverse directions.

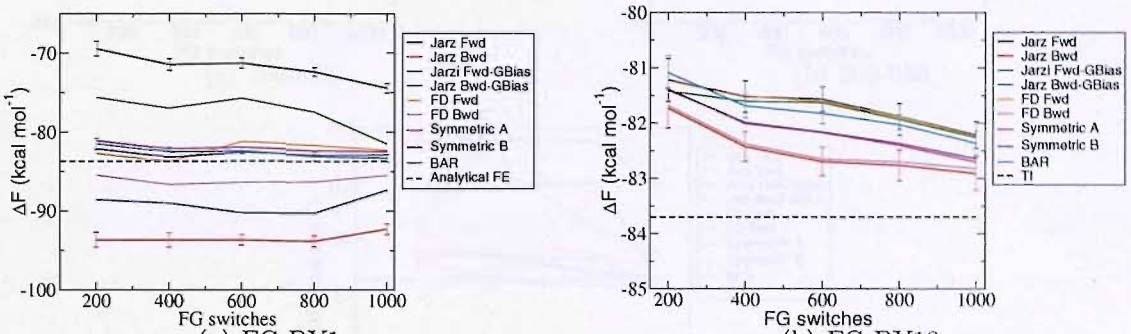
Comparing to FDTI in figure 5.1 it is clear that the use of shorter FG switches with the Jarzynski estimator leads to large inaccuracies. Jarzynski estimates are relatively similar when

using the longest switches with 10,000 λ increments. Jarzynski estimates are also symmetric with forwards and backwards estimates being equally inaccurate. The Jarz-GBias corrected estimates show improvement with forwards estimates giving accurate results with the least expensive protocol. The success of the Jarz-GBias estimator is due to the strictly Gaussian nature of the work distribution. However, this success in bias correction is unreliable as forward Gore bias corrected estimates are less accurate compared to the Jarzynski estimator alone with the longer 10,000 pass switches. This bias over-correction as $\bar{W}_{diss}$ becomes small is a problem noted and discussed by Gore *et al.* in his original study [Gore *et al.*(2003)Gore, Ritort & Bustamante] and here in the FG background chapter.

The FD estimators show much improved accuracy compared to the Jarzynski estimator as predicted by Hummer [Hummer(2002)]. The improvement shown by the FD estimator is undoubtedly specific to cases such as this where the work distribution is almost perfectly gaussian. Possibly the best estimator for this system is the symmetric A estimator closely followed by the symmetric B estimator. The ease of prediction of these symmetric estimators is due to the symmetric nature of switching this system from uncharged to charged. As the change in ΔF is linear with changes in λ these changes are also symmetric.

Considering the results of chapter 4, it was deemed useful to investigate possible improvements found with an FG-BY10 approach. This FG-BY10 calculation used FG switches, with 2000 MC passes and 2000 λ increments, as used by Hummer (2002). 50 switches were used for each of the 10 individual FG calculations making up the perturbation from $\lambda = 0$ to $\lambda = 1$, with the results of each FG calculation in the same direction being summed to get the ΔF. This calculation was compared to a normal FG calculation using 2000 λ increment switches where estimates where calculated every 200 switches (in both directions). Some estimators were previously very inaccurate with this length of switches (2000 MC passes in figure 5.1), so there is room for improvement with FG-BY10. Figure 5.2 a) shows the normal FG calculation results as the number of switches is increased. There seems to be little improvement in the poor Jarzynski estimates through the course of the calculation, although BAR and FD give good accuracy even with only 200 switches. Figure 5.2 b) shows as expected that by splitting the λ coordinate into 10, errors due to hamiltonian lag have been almost removed. Estimates from the FG-BY10 calculation may not have completely converged and seem to be moving towards

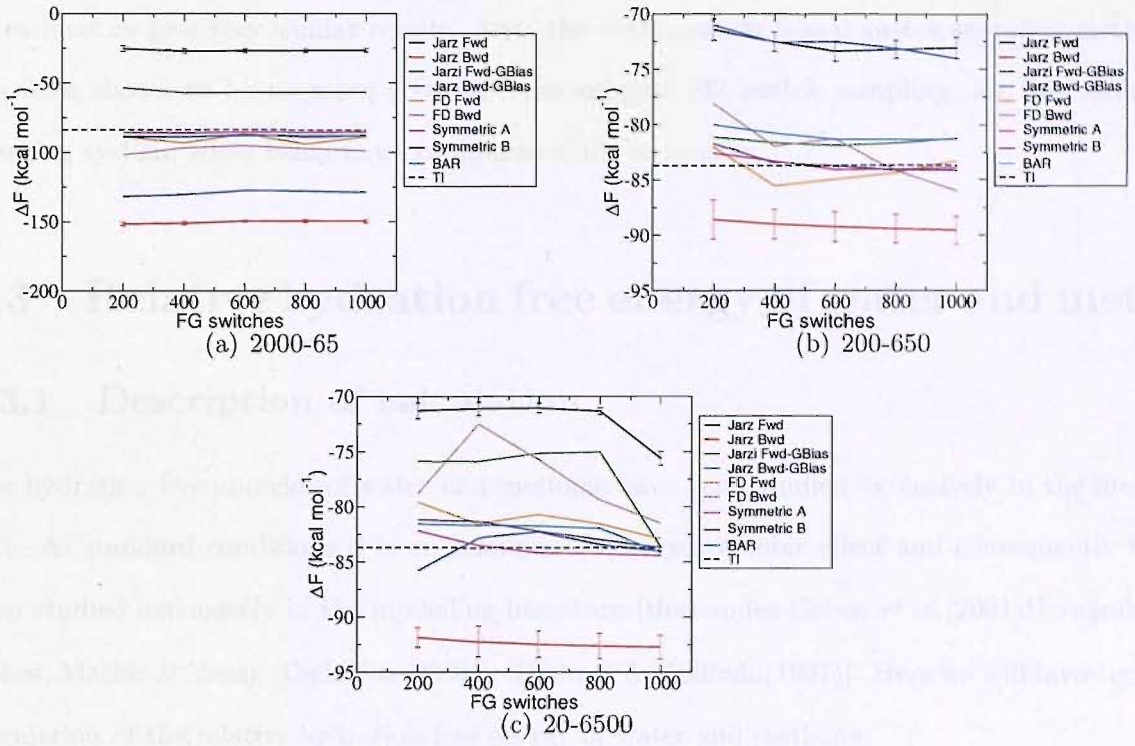
the TI estimate. Here FG-BY10 is able to show similar accuracy to TI. TI is however able to use a small number of λ window simulations as the the change in free energy with λ is linear. Thus, for the charging of sodium, TI may be able to converge more easily than FG methods.



(a) FG-BY1 (b) FG-BY10

Figure 5.2: Estimates of the free energy of sodium charging in water using FG-BY1 and FG-BY10 protocol with all FG estimators.

The configuration bias-d switch sampling method investigated in the previous chapter on harmonic oscillator models was applied to this sodium charging test system. The configuration bias-d results in figure 5.3 used three different switching protocols with differing numbers of λ increments and MC trials. It is clear from these results that using a small number of λ increments with large amounts of MC trials between each is most efficient. However, the most efficient configuration bias-d calculation with 20 λ increments is no more efficient than the FG-BY1 protocol in figure 5.2 above.



(a) 2000-65 (b) 200-650 (c) 20-6500

Figure 5.3: Estimates of the free energy of sodium charging in water using configuration bias-d FG sampling using three switching protocols. 2000-65, 200-650 and 20-6500 denotes switching protocols with 1 MC pass between each of 2000 λ increments, 100 MC pass between each of 200 λ increments and 1000 MC pass between each of 20 λ increments respectively.

5.2.3 Sodium charging conclusions

As previously shown by Hummer (2002) using FG switches and the Jarzynski estimator in a single direction is unable to produce results as accurate as TI using the same number of MC trials. This is mainly due to the possibility of using only three λ simulation windows with TI as the free energy change is linear. However, as previously found by Hummer (2002), the FD estimator and Symmetric estimators are able to offer more accurate estimates of a similar accuracy to TI in the case of sodium charging. This analysis helps validate this implementation of these FG methods as they agree in general with published results [Hummer(2002)].

BAR has been shown to be more accurate than the Jarzynski estimator and as accurate as FD and Symmetric estimators. Also, as discussed in the FG background chapter the Gore bias has been shown to improve Jarzynski estimates when the bias is large (small N) but is unreliable when the bias is very small (large N).

The use of a FG-BY10 protocol has been shown to improve the accuracy of the Jarzynski estimator such that it is as accurate as BAR for the FG-BY1 protocol. Indeed using FG-BY10

all estimators give very similar results. Also, the configuration bias-d switch sampling method has been shown to be no more accurate than original FG switch sampling, for this sodium charging system, while being more computationally expensive.

5.3 Relative hydration free energy of water and methane

5.3.1 Description of calculation

The hydration free energies of water and methane have been studied extensively in the literature. At standard conditions it is an example of the hydrophobic effect and consequently has been studied extensively in the modelling literature [Hernandez-Cobos *et al.*(2001)Hernandez-Cobos, Mackie & Vega], [Delle Site(2001)], [Radmer & Kollman(1997)]. Here we will investigate calculation of the relative hydration free energy of water and methane.

The relative hydration free energy of water and methane can be calculated through the thermodynamic cycle in figure 5.4 in connection to the equivalent experimental steps.

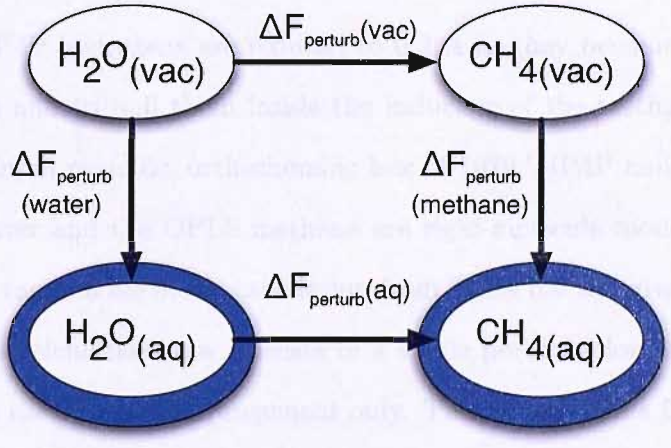


Figure 5.4: Thermodynamic cycle used to calculate the relative hydration free energy of water and methane

Thus, perturbations or switches must be performed from water to methane in vacuum and water environments. The relative free energy difference ($\Delta\Delta F$) is found by taking the ΔF of the water leg from the ΔF of the vacuum leg (equation 5.1).

$$\begin{aligned} \Delta\Delta F_{hyd} &= \Delta F_{pert}(vac) - \Delta F_{pert}(aq) \\ &= \Delta F_{hyd}(methane) - \Delta F_{hyd}(water) \end{aligned}$$

The experimental free energy of hydration of methane is unfavourable at 2.00 kcal.mol^{-1} at 298 K, while the hydration of water has a favourable ΔF of -6.31 kcal.mol^{-1} [Zhou *et al.*(1998)Zhou, Li, Hawkins, Cramer & Truhlar]. This gives an experimental water to methane relative hydration free energy of -8.31 kcal.mol^{-1}.

This calculation has been performed in the literature by Woods *et al.* using many free energy methods in a direct comparison [Woods *et al.*(2003a)Woods, Essex & King]. It is convenient to use the same system setup to give a quick comparison of FG methods with the equilibrium methods investigated by Woods *et al.* The water to methane model consists of a TIP4P water molecule switching to an OPLS united atom methane molecule [Jorgensen *et al.*(1984)Jorgensen, Madura & Swenson], [Jorgensen *et al.*(1983)Jorgensen, Chandrasekhar, Madura, Impey & Klein]. The oxygen atom of the TIP4P water is switched to the OPLS methane while the TIP4P hydrogens and extra "M" atom are switched to dummy atoms. The bond lengths of the TIP4P hydrogens are reduced to 0.2 Å as they become dummy atoms to help smooth the switch and to pull them inside the influence of the methane molecule. The water-methane resides in an periodic, orthorhombic box of 1679 TIP4P molecules.

Both the TIP4P water and the OPLS methane are rigid-molecule models. Consequently, for this calculation, the vacuum leg of the calculation from figure 5.4 will give a ΔF of zero and can be discounted. The calculation now consists of a single perturbation and the free energy difference is now due to the solvent rearrangement only. This is useful as a free energy method can be assessed solely on its ability to evaluate solvent rearrangment and solute-solvent interactions.

The solvated water to methane perturbation may display a decrease in entropy due to excluded volume effects of the hydrophobicity of the methane [Hernandez-Cobos *et al.*(2001)Hernandez-Cobos, Mackie & Vega]. It may be possible that as the water-methane molecule becomes methane the surrounding water forms a rigid shell around it and loses entropy. In terms of the IHO cases discussed in chapter 4, this perturbation may display characteristics of cases C and D, as the number of degrees of freedom explored in the methane system may be less

than the non methane system, so the sizes of phase space when λ is 0 and 1 will differ. Also, the overall structure of the system will change as the solvation shells of the two systems differ. Neutron diffraction results for a methane system show peaks for both hydrogen-methane and oxygen-methane radial distribution function (RDF) at 3.5 Å [Dejong *et al.*(1997)Dejong, Wilson, Neilson & Buckingham]. As the hydrogen peak is broader than the oxygen peak it is thought that the waters orientate themselves with hydrogens towards the methane. The results for pure water show a sharp oxygen-oxygen RDF peak at 2.8 Å with a second shell at 4.5 Å and a third at 6.8 Å. This change in structure is likely to cause FG calculations to display some level of hamiltonian lag where TI equilibrium calculations would not. Also, as the ΔF is mainly due to solvent rearrangement, it is likely that this solvated water to methane perturbation will display a smooth potential energy change. This will allow TI calculations to integrate over the free energy gradient with a good level of accuracy.

Water-methane simulation parameters used by Woods *et al.* for all simulations are set out in table 5.2. The present study used the same parameters, although as ProtoMS 2.1 Monte Carlo simulation software [Woods & Michel(2005)] was used rather than MCPRO 1.5 [Jorgensen(1996)], a small adjustment was required. ProtoMS 2.1 is unable to force volume moves every 10375 MC trials. Instead volume moves are made with a probability relative to solute and solvent moves. The MC trial probability ratios used for water-methane simulation in the present study are volume 2:solute 13:solvent 20800. This is very close to the MC trial probabilities used by Woods *et al.*

Parameter description	Setting
Ensemble	NPT
Temperature	25 °C
Pressure	1 atm
Boundary conditions	Orthorhombic periodic
Non-bonded cutoff	15 Å
Solute/solvent trials ratio	1/1600
frequency of volume trials	1 every 10375 trials
Max solute translation/rotation	0.1 Å/ 5.0 °
Max solute translation/rotation	0.1 Å/ 2.5 °
Maximum volume change	830 $Å^3$
Preferential sampling centre	Oxygen/CH_4 of the solute
Preferential sampling parameter	200

Table 5.2: Simulation parameters for relative hydration free energy of water and methane simulations.

Woods *et al.* equilibrated the water-methane system, at $\lambda = 0$ for 2 million MC trials. For the methods employing fixed λ simulations, FEP, FDTI, PTTI (parallel tempering thermodynamic integration), RETI, and REFEP (replica-exchange free energy perturbation), 21 uniformly spaced windows were used, starting with the final configuration from the equilibration run. 10 million MC trials were allowed in each window, split into 3 million for λ equilibration and 7 million for calculation data. This protocol uses a total of 212 million MC trials.

Woods *et al.* were able to show the improved performance of RETI and REFEP compared to the established non-enhanced methods, FDTI and FEP [Woods *et al.*(2003a)Woods, Essex & King]. The addition of replica-exchange moves over λ reduced the random sampling error and the statistical error. Also, RETI performed consistently better than REFEP and was in very good agreement with experiment.

When designing an FG protocol for this water-methane calculation, it is important to consider the results of chapter 3. Thus, owing to the nature of the water-methane perturbation, a switch splitting protocol is advisable, to attempt to minimise any hamiltonian lag. Also, results from chapter 3 suggest that the longest switching trajectories and smallest λ increments possible should be used to increase efficiency. All FG calculations use as close as possible the same number of MC trials as the RETI protocol of Woods *et al.* (total MC trials used will be noted in brackets the first time each method/protocol is mentioned) [Woods *et al.*(2003a)Woods, Essex & King].

The advantages of a λ swap move in FDTI and FEP protocols is clear [Woods *et al.*(2003a)Woods, Essex & King], [Woods *et al.*(2003b)Woods, Essex & King]. A λ swap move can also be applied across a series of FG calculations which traverse λ (REFG). It makes sense to conduct replica exchanges between the equilibrium simulations used to produce FG starting configurations (seed simulations) which are adjacent on the λ coordinate. As the seed simulations can differ in size depending on the FG protocol (i.e. the number of MC trials between starting configurations), for REFG it may be worthwhile to choose a protocol which maximises seed simulation sampling at the expense of switch length and thus allows more exchange of replicas.

5.3.2 Water-methane results

Figure 5.5 shows water-methane results for non-replica exchange (non-RE) methods, comparing four repeated calculations, which all started with the same equilibrated structure and are joined by a line, using FEP, TI and various protocols of FG. In figure 5.5 the calculation protocols labeled FG-BY10, FG-BY10-Bwd and FG-BY20 are explained below. Black lines denote calculations where λ is being incremented from 0 to 1 and red lines where λ is being incremented from 1 to 0 with cyan lines representing BAR calculations. The dashed black line marks the experimental relative hydration free energy of 8.31 kcal.mol^{-1} $\pm$ 0.5 and the dashed red line marks the recently exhaustively calculated relative hydration free energy of 8.8 kcal.mol^{-1} $\pm$ 0.1 using RETI with randomly chosen λ swap moves, as discussed below [Woods(2007)].

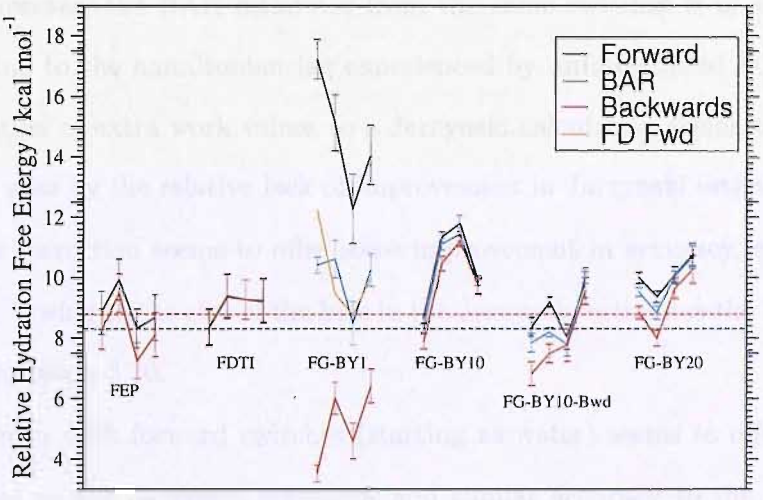


Figure 5.5: Relative hydration free energy of water and methane, estimated by four repetitions for each non-RE method. Each set of four estimates is linked with a line and labeled with the method abbrieviation. These estimates are compared to the experimental value of 8.31 kcal.mol^{-1} which is the black dashed line and the accurate calculated free energy difference of 8.8 kcal.mol^{-1} [Woods(2007)].

FG protocol	Switches	$\Delta\lambda$s	MC trials per $\Delta\lambda$	λ Split	Total MC trials (millions)	Equilibrium MC trials/λ swap
FG-BY1	400	1000	375	1	196	100000
FG-BY10	400	1000	375	10	207	100000
FG-BY20	800	500	375	20	235	100000

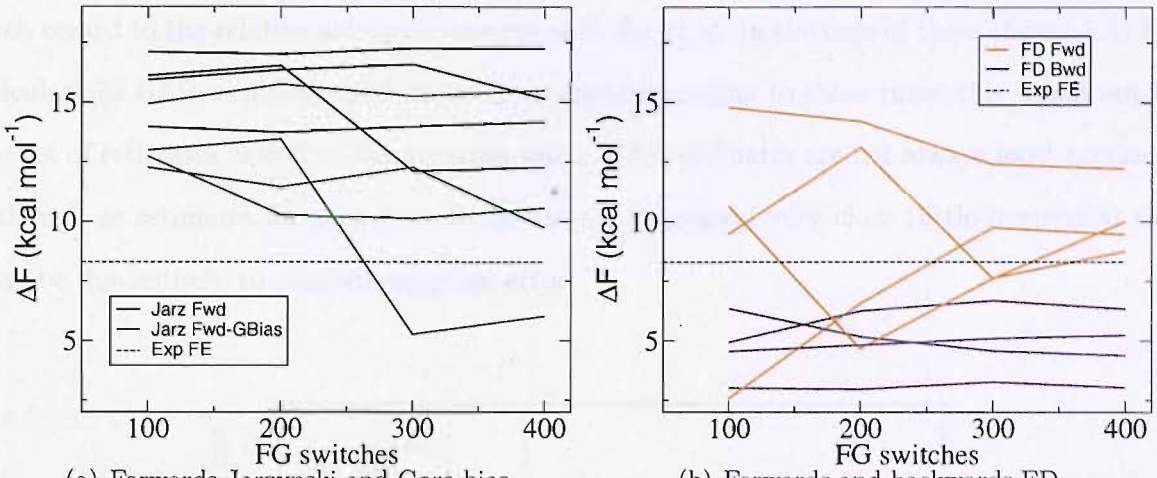
Table 5.3: Table describing non-RE FG protocols used in figure 5.5

FEP estimates display a hysteresis of 0.5 to 1 kcal.mol^{-1}. The problem of choosing a sampling direction may be solved by application of BAR to the FEP results, which would find the

optimum estimate due to the variance of the data. However, more of an issue is the spread of FEP estimates in a single direction which is around 2 kcal.mol^{-1}. This large spread is despite the relatively simple system and the common equilibrated starting structure for each calculation repetition. Each FEP calculation repetition probably samples different areas of configurational phase space giving these different results. This problem can be called the random sampling error, as discussed by Woods *et al.* [Woods *et al.*(2003a)Woods, Essex & King]. FDTI shows very low levels of hysteresis, however this is obviously a poor measure of the level of possible error as the random sampling error is comparable to that of FEP.

The results labelled FG-BY1 use an FG protocol (196 million MC trials) with uninterrupted switches from $\lambda = 0$ to $\lambda = 1$ and numbers of switches, λ increments and MC trials per λ increment listed in table 5.3. The hysteresis of the FG estimates is very large at around 13 kcal.mol^{-1}. However, the BAR estimates from the same switches is of similar quality to FEP and TI. Owing to the hamiltonian lag experienced by uninterrupted FG water-methane switches, the addition of extra work values to a Jarzynski calculation seems to make little difference, as can be seen by the relative lack of improvement in Jarzynski estimates in figure 5.6 a). The Gore bias correction seems to offer some improvement in accuracy, especially in with forwards switches. Owing to the size of the bias in the Jarzynski estimates the bias is calculated accurately using equation 3.30.

The FD estimator with forward switches (starting as water) seems to offer improved performance compared to the Jarzynski estimator and similar accuracy to the estimators using switches in both directions shown clearly in figure 5.5. Forward FD estimates seem to go through large fluctations although they do seem to converge as more switches are used (figure 5.6 b). FD estimates with backwards switches seem to fluctuate less than their forwards counterparts but give less accurate estimates. These results demonstrate that although forwards and backwards jarzynski estimates have similar inaccuracy, forwards and backwards switches contain different levels of bias for this system.



(a) Forwards Jarzynski and Gore bias (b) Forwards and backwards FD

Figure 5.6: ΔF estimates against numbers of FG switches for the FG-BY1 protocol (table 5.3) i.e. where switches are from $\lambda = 0$ to $\lambda = 1$ and various estimators.

FG-BY10 (207 million MC trials) uses the same switching protocol as as FG-BY1. However, as the λ coordinate is split into ten, the effective length of switches is ten times longer (table 5.3). The use of a FG-BY10 protocol massively lowers the hysteresis of ΔF estimates. BAR estimates due to the FG-BY10 protocol are not improved over BAR estimates with uninterrupted switches.

The FG-BY10-Bwd results are where another four FG-BY10 repetitions were carried out, but with the system equilibrated at $\lambda = 1$ (methane) as opposed to $\lambda = 0$ (water). From the eight FG-BY10 and FG-BY10-Bwd estimates reported here it is possible that there is no advantage in equilibrating in one starting system over the other and that both sets of four estimates are fully equilibrated.

The FG-BY20 (235 million MC trials) protocol splits the λ coordinate into 20 with effective length of switches which are the same as FG-BY10 (table 5.3) i.e. twice as many switches are used with each individual switch being half the length of those in the FG-BY10 protocol. This protocol produces quite similar estimates to the FG-BY10 protocol (figure 5.5). FG-BY20 estimates are in general slightly closer to the, thought to be accurate, value of 8.8 kcal.mol^{-1} of Woods *et al.* [Woods(2007)]. However, FG-BY20 seems to produce slightly more hysteresis between forwards and backwards calculations.

For FG, FG-BY10 and FG-BY20 BAR estimates could be picked according to the estimator rules described in the previous chapter; where a BAR estimate is used if it is placed between the forwards and backwards Jarzynski estimates, otherwise the Jarzynski estimates are chosen

with regard to the relative entropy measures of Kofke *et al.* In the case of these (figure 5.5) FG calculations BAR would be used for every estimate according to these rules; this would not be the set of estimates closest to the accurate value. BAR estimates are not always most accurate, although as estimates for each FG calculation are in general very close (little hysteresis) this may be due entirely to random sampling error.

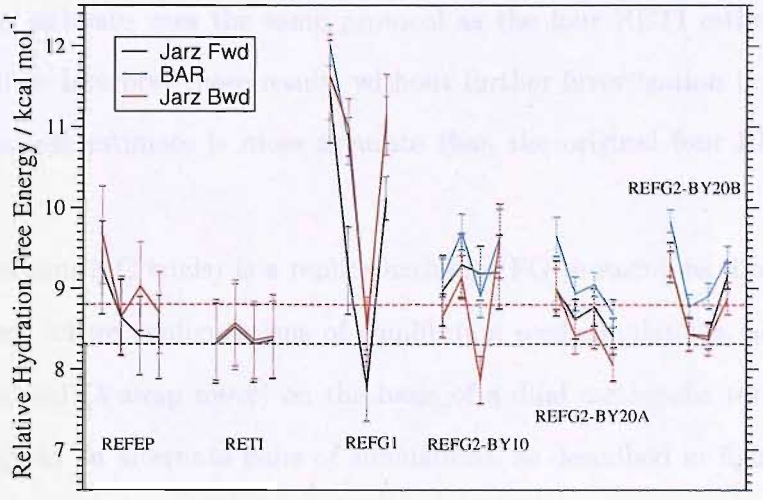


Figure 5.7: Relative hydration free energy of water and methane, estimated by four repetitions for each RE method. Each set of four estimates is linked with a line and labeled with the method abbrieviation. These estimates are compared to the experimental value of 8.31 $kcalmol^{-1}$ which is the dashed line.

FG protocol	Switches	$\Delta\lambda$s	MC trials per $\Delta\lambda$	λ Split	Total MC trials (millions)	Equilibrium MC trials/λ swap
REFG1	800	1000	200	10	217	50000
REFG2-BY10	800	1000	200	10	240	100000
REFG2-BY20A	1600	500	200	20	267	100000
REFG2-BY20B	1600	1000	200	20	427	100000

Table 5.4: Table describing REFG protocols used in figure 5.7

Methods involving λ swap moves display significantly more consistent ΔF estimates. As shown by Woods *et al.* RETI ΔF estimates show very low random sampling error and extremely good agreement with experiment [Woods *et al.*(2003a)Woods, Essex & King]. REFEP has slighty larger levels of hysteresis and random sampling error. Both REFEP and RETI show a flipping of the relative positions of forwards and backwards estimates in comparison to FEP and FDTI, this will be discussed in full later.

The four RETI estimates of figure 5.7 although in very good agreement with experiment

may not be properly converged and accurate. These four RETI calculations were performed with MCPRO [Jorgensen(1996)] and a λ swap scheme where configurations adjacent in λ are swapped in alternate directions as is the case with REFG1 in figure 5.8. The dashed red line marks the more recent RETI estimate performed with ProtoMS 2.1 [Woods & Michel(2005)] and using a λ swap scheme were configurations adjacent in λ are swapped in random directions (RETI-random) unlike all the other RE free energy protocols discussed here [Woods(2007)]. The RETI-random estimate uses the same protocol as the four RETI estimates in all other ways. It is difficult to interpret these results without further investigation but it may be that this new RETI-random estimate is more accurate than the original four RETI estimates of Woods (2003a).

REFG1 (217 million MC trials) is a replica-exchange FG protocol, as discussed in the FG background chapter, where configurations of equilibrium seed simulations, adjacent in λ, are intermittently swapped (λ swap move) on the basis of a dual metropolis test. λ swap moves can only be attempted on alternate pairs of simulations, as described in figure 5.8, any more and configurations could undergo more than one λ swap at once. Switches are performed immediately after a set of λ swap moves have been made.

Figure 5.8 describes REFG1 with a BY3 (λ coordinate is split into three) set of FG calculations, to save space. In figure 5.8 each thick, black, straight, arrowed line represents a section of equilibrium seed simulation. The circles represent system A and squares represent system B, filled with cyan these systems are equilibrated while if filled with red they are not. Each thinner dashed line represents the use of one system configuration to start multiple simulations across λ. Very thick blue lines with arrows at either end represent λ swap moves and thin wavy red lines are FG switches.

As shown in figure 5.8 REFG1 starts with the same system equilibration as other FG methods described in the FG background chapter. The equilibrated structure is then used to start simulations at 11 uniform intervals across λ as with FG-BY10. These simulations are then stopped at prespecified intervals to attempt λ swap moves and then perform FG switches from each equilibrium λ simulation, in both directions to the next equilibrium λ simulation.

REFG1 estimates use a switching protocol, described in table 5.4, which has twice as many switches of almost half the length compared with the protocol used for FG-BY10 estimates. This change is an attempt to allow wider sampling of equilibrium seed simulations. All REFG

estimates use the same switching protocol as REFG1 (table 5.4) except REFG2-BY20.

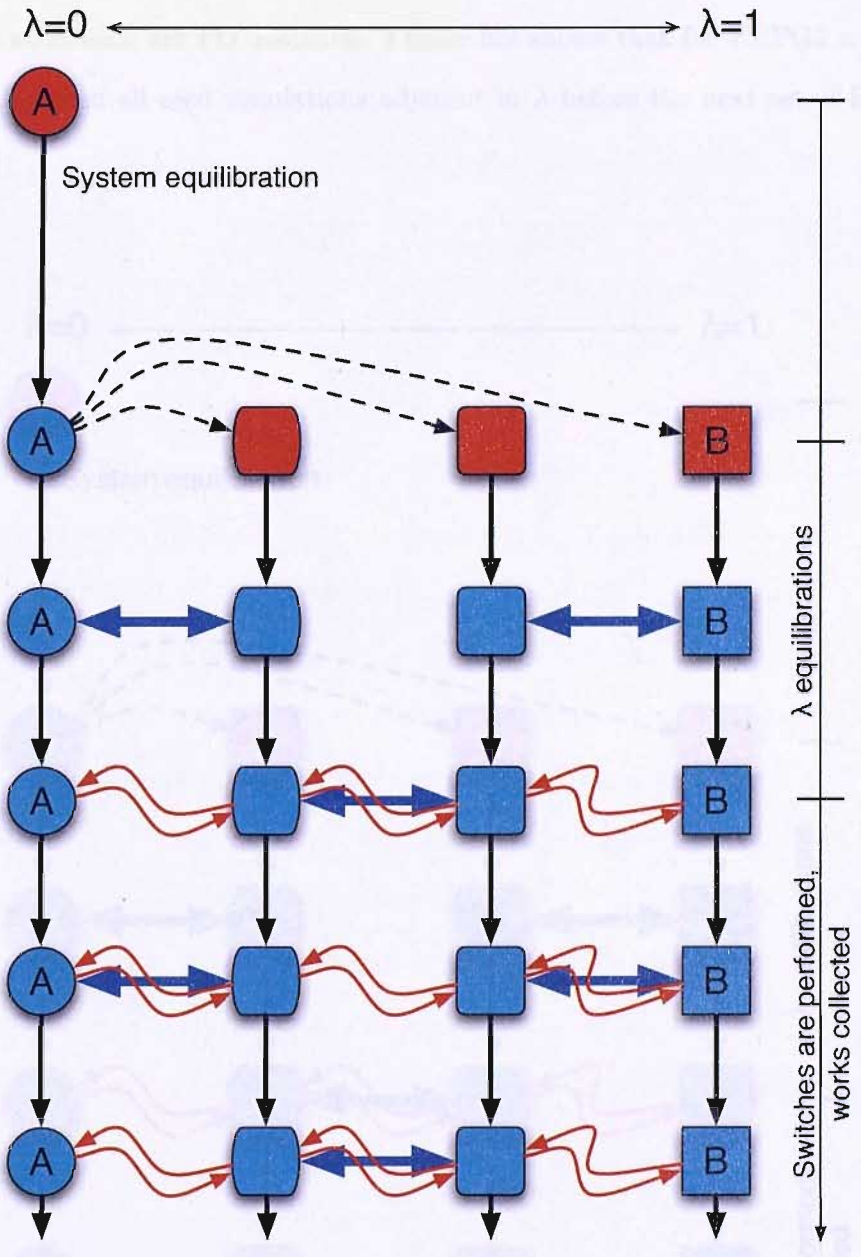


Figure 5.8: Diagram describing the REFG1 protocol. Black lines represent equilibrium simulations, thick blue lines represent λ swap moves, black dashed lines represent the proliferation of a single system configuration and thinner red wavy lines represent FG switches.

REFG1 estimates in figure 5.7 seemed to show higher levels of random sampling error compared to FG-BY10, and larger statistical errors. REFG1 also dislays a flipping of relative placement of forward and backward estimates as seen with FG-BY20, REFEP and RETI. However, unlike FG-BY20 the BAR estimates have moved to be very similar to backward Jarzynski

estimates and consequently forward Jarzynski estimates may be chosen rather than BAR estimates according to the estimator rules previously stated in chapter 4.

REFG2 is similar to REFG1, the difference being that λ swap moves are made in both directions between each set FG switches. Figure 5.9 shows that for REFG2 a λ swap move is attempted between all seed simulations adjacent in λ before the next set of FG switches is started.

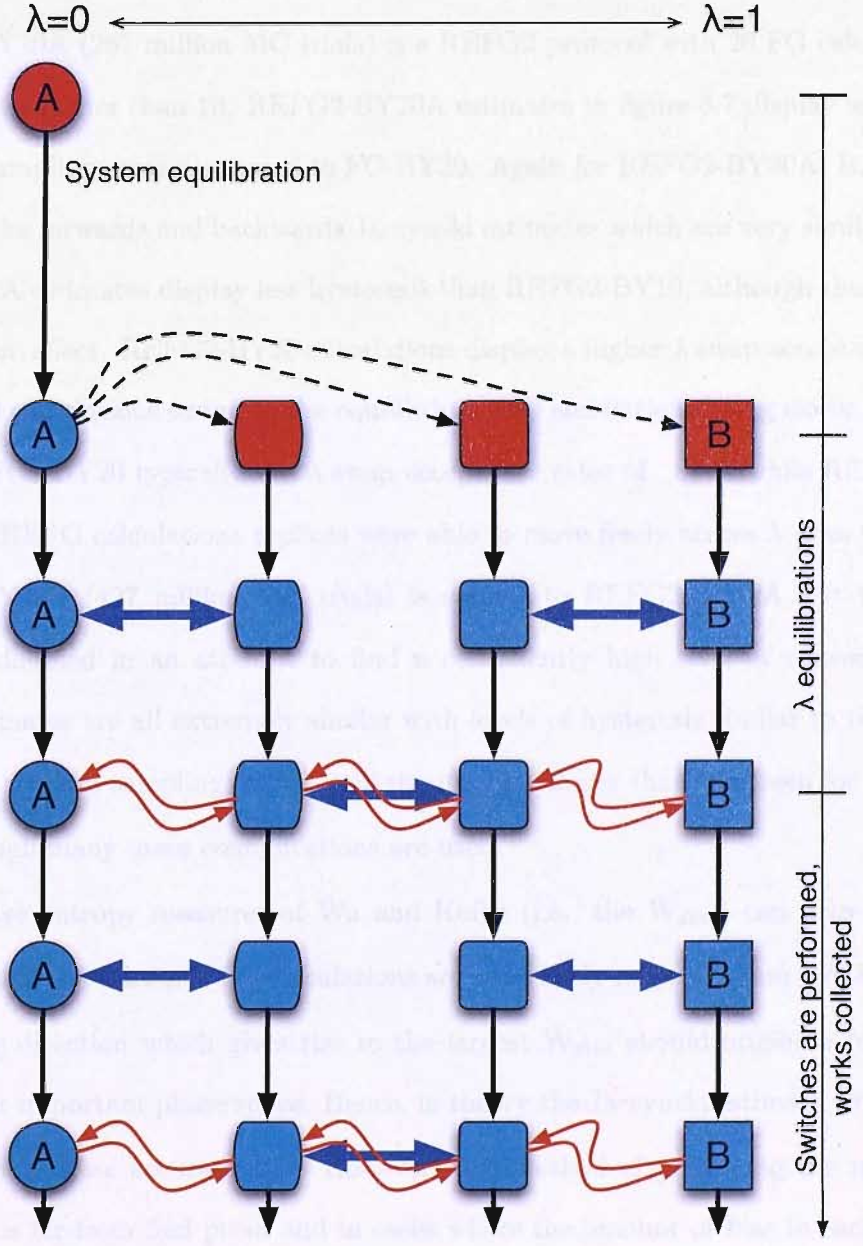


Figure 5.9: Diagram describing the REFG2 protocol. Black lines represent equilibrium simulations, thick blue lines represent λ swap moves, black dashed lines represent the proliferation of a single system configuration and thinner red wavy lines represent FG switches.

REFG2-BY10 (240 million MC trials) estimates in figure 5.7 have less random sampling error than the other FG protocols but slightly more than RETI. It seems that forward Jarzynski REFG2-BY10 estimates are more accurate and consistent than BAR. BAR estimates seem hampered by the less consistent backwards data. REFG2-BY10 forwards Jarzynski estimates seem to have converged slightly above the RETI-random value of 8.8 kcal.mol^{-1} while being as consistent as the four RETI estimates while using approximately half the MC configurations. Although this REFG2-BY10 protocol uses the same alternate direction λ swap scheme as the four original RETI estimates it is closer to the RETI-random estimate.

REFG2-BY20A (267 million MC trials) is a REFG2 protocol with 20 FG calculations over the λ coordinate rather than 10. REFG2-BY20A estimates in figure 5.7 display estimates with less random sampling error compared to FG-BY20. Again for REFG2-BY20A, BAR estimates are all above the forwards and backwards Jarzynski estimates which are very similar. Jarzynski REFG2-BY20A estimates display less hysteresis than REFG2-BY10, although there is a chance this is a random effect. REFG2-BY20 calculations display a higher λ swap acceptance rate than REFG2-BY10 calculations owing to the equilibrium seed simulations being closer to each other across λ. REFG2-BY20 typically has λ swap acceptance rates of 80 % while REFG2-BY10 is 65 %. For all REFG calculations replicas were able to move freely across λ from 0 to 1.

REFG2-BY20B (427 million MC trials) is similar to REFG2-BY20A but the length of switches are doubled in an attempt to find a consistently high level of convergence. Here Jarzynski estimates are all extremely similar with levels of hysteresis similar to the four RETI estimates but random sampling error is still significantly larger than that seen for RETI (figure 5.7) even though many more configurations are used.

The relative entropy measures of Wu and Kofke (i.e. the $\bar{W}_{diss}$) can help us to choose whether forwards or backwards FG calculations are more likely to be accurate [Wu & Kofke(2005a)]. The switching direction which gives rise to the largest $\bar{W}_{diss}$ should originate from a system with the larger important phase space. Hence, in theory the Jarzynski estimate with the largest $\bar{W}_{diss}$ will give a more accurate ΔF. However, this method of predicting the most accurate ΔF estimate is far from fool proof and in cases where the amount of bias in each direction is similar (such as this study of the relative free energy of hydration of water and methane) random factors can result in an incorrect prediction. The Kofke bias measure described in the FG

background chapter may offer a more reliable prediction of the best estimate as this measure attempts to give an idea of the over all bias present.

FG protocols such as REFG2-BY10 discussed above have a number of small FG calculations across λ rather than one calculation relating the A and B systems. This means there are a number of ways estimates of the full free energy difference can be found from the individual FG calculations across λ, here we will test three:

1. **Independently chosen relative entropy measure**: In the case of these multi FG calculation protocols each individual FG calculation can be treated independently and a particular estimator chosen for each, with the result being that different estimators are used for different parts of the perturbation (independently chosen relative entropy measure).

2. **Independently chosen Kofke bias**: Same as above except the Kofke bias is used to pick the most accurate estimator.

3. **Totalled relative entropy measure**: The relative entropy measure can be totalled across the λ coordinate and a direction chosen from these totals, so the same estimator is used for each individual FG calculation across λ (totalled relative entropy measure). This method may help to choose the correct estimator when using few, very long switches and calculations are well behaved as relative entropy measures are likely to be similar in the forwards and backwards direction. Consequently random fluctuations could have an impact.

Table 5.5 shows the Kofke bias measures and free energy estimates for each individual calculation of the first REFG2-BY10 repetition in figure 5.7. For the three sections of the λ coordinate between 0 and 0.3 the forwards Kofke bias measure is negative, predicting that these forwards Jarzynski estimates are not converged. The Kofke bias measure suggests that in general the backwards Jarzynski estimates are more converged than the forwards Jarzynski estimates. These are trends seen in all repetitions of these REFG2-BY10 calculations suggesting the initial portion of the λ coordinate is difficult to converge in the forwards direction and overall the backwards direction may provide converge faster. However, these suggestions do

not fit with observation that the forwards Jarzynski estimates seem more internally consistent than the backwards estimates.

λ	Fwd Kofke	Bwd Kofke	Fwd Jarz	Bwd Jarz	BAR
0 - 0.1	-0.977	1.429	4.600	4.552	4.447
0.1 - 0.2	-0.738	0.550	2.048	1.858	2.182
0.2 - 0.3	-0.216	2.383	1.115	0.943	1.083
0.3 - 0.4	0.195	0.009	0.486	0.608	0.590
0.4 - 0.5	0.074	13.550	0.333	0.372	0.383
0.5 - 0.6	0.572	2.189	0.240	0.143	0.247
0.6 - 0.7	1.624	1.133	0.050	0.079	0.071
0.7 - 0.8	1.622	1.245	0.048	0.061	0.057
0.8 - 0.9	0.510	5.727	0.035	0.040	0.047
0.9 - 1	1.947	1.618	0.017	0.009	0.029

Table 5.5: Table of Kofke bias measure values and free energy differences for each individual calculation of the first repetition of the REFG2-BY10 protocol (all free energy differences in kcal.mol^{-1}). Fwd Kofke and Bwd Kofke denote the Kofke bias measure in the forwards and backwards direction respectively. Also, Fwd Jarz and Bwd Jarz denotes the Jarzynski estimator in the forwards and backwards directions respectively.

FG protocol	Repetition	Fwd Wdiss	Bwd Wdiss	Fwd Jarz	Bwd Jarz	BAR
REFG2	1	1.81	1.63	8.97	8.67	9.13
REFG2	2	1.73	1.38	9.29	9.14	9.7
REFG2	3	1.36	1.23	9.19	7.87	8.92
REFG2	4	1.74	1.86	9.20	9.59	9.62

Table 5.6: Table of relative entropy measure values and relative hydration free energy of water and methane estimates (kcal.mol^{-1}) for FG-BY10 and REFG2 methods.

Relative entropy measures totalled across λ for the REFG2-BY10 calculations given in this study are in table 5.6. Finding estimates using the totalled relative entropy measure for the REFG2-BY10 repetitions we pick the forwards direction in all but repetition 3 where the BAR estimate is placed between the two Jarzynski estimates. While the estimate chosen is not always the closest to the RETI-random estimate of 8.8 kcal.mol^{-1} (red dashed line), the chosen estimates, labelled "Wdiss Pick total" in figure 5.10, are consistently very close to the RETI-random estimate.

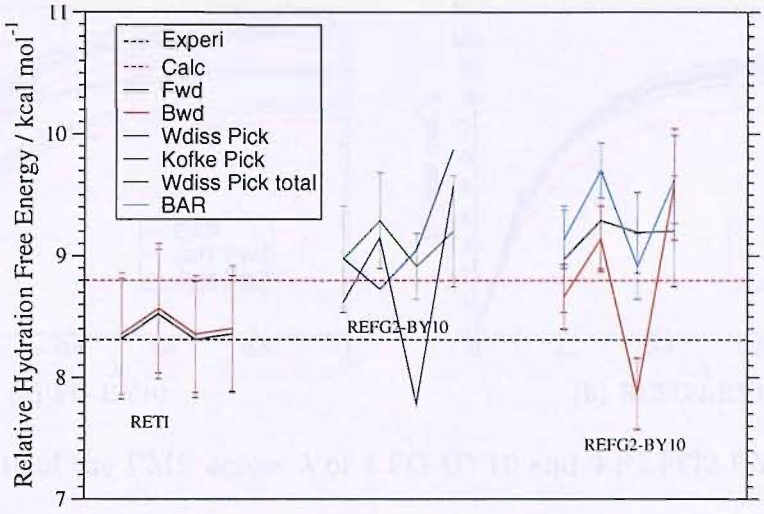


Figure 5.10: Relative hydration free energy of water and methane, These estimates are compared to the experimental value of 8.31 $kcalmol^{-1}$ which is the dashed line.

Finding estimates using the independently chosen relative entropy measure for the REFG2-BY10 repeats (labelled "Wdiss Pick" in figure 5.10) produces estimates of a similar quality to the totalled relative entropy measure but possibly slightly less consistent. Again producing estimates using the independently chosen Kofke bias measure (labelled "Kofke Pick" in figure 5.10) gives similar estimates but even less consistent than the methods using the relative entropy measure.

In theory finding an free energy difference estimate using the independently chosen Kofke bias measure should produce the best results as it should find the estimate with the least bias. In this test it is difficult to see which method is the best, possibly due to the fact that random factors may be the main source of error in these calculations. The totalled relative entropy measure ("Wdiss Pick total" in figure 5.10) seems to produce the most constantly accurate results although it is possibly the most arbitrary.

From the results discussed above it is clear that in general it is advisable to use RE free energy methods as opposed to non-RE methods. REFG2-BY10 is able to produce consistently more accurate results than FG-BY10. The improvement in convergence between similar FG-BY10 and REFG2-BY10 protocols is again demonstrated in figure 5.13. The REFG2-BY10 PMFs are more consistent than the FG-BY10 equivalents.

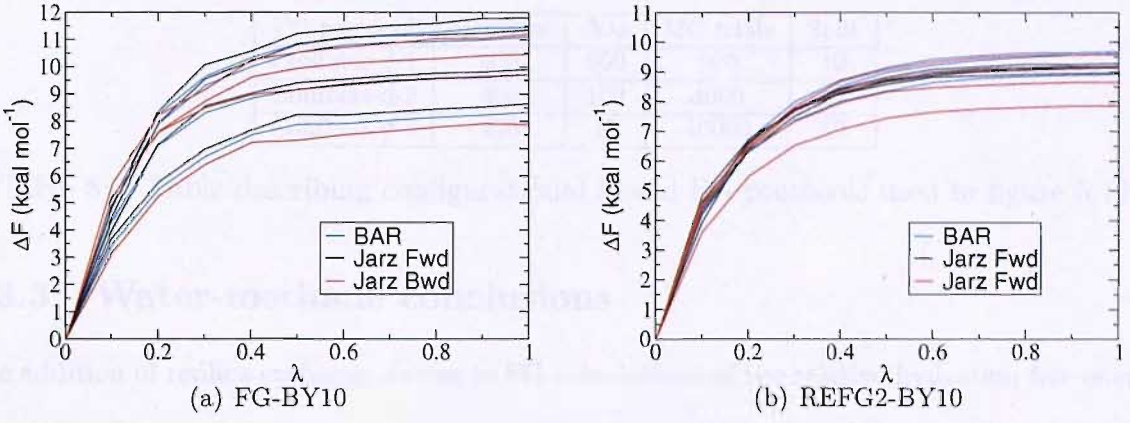


Figure 5.11: Plots of the PMF across λ of 4 FG-BY10 and 4 REFG2-BY10 calculations.

There was evidence, discussed in chapter 4, to suggest that configurational bias-d FG is more efficient than other FG methods with cases C and D. Cases C and D may have similar barriers to sampling as the relative hydration free energy of water and methane calculation in the present study. Thus, it was deemed worthwhile to investigate the performance of configurational bias-d relative to original FG with our water to methane system. Figure 5.12 shows estimates for three configurational bias-d protocols, which are listed in table 5.7, compared to FDTI and FG-BY10 estimates discussed earlier. The configurational bias-d estimates seem to display higher levels of hysteresis between the forwards and backwards Jarzynski estimates. Although, especially for Confbias-d-1 estimates seem more consistent leading to more accurate BAR estimates.

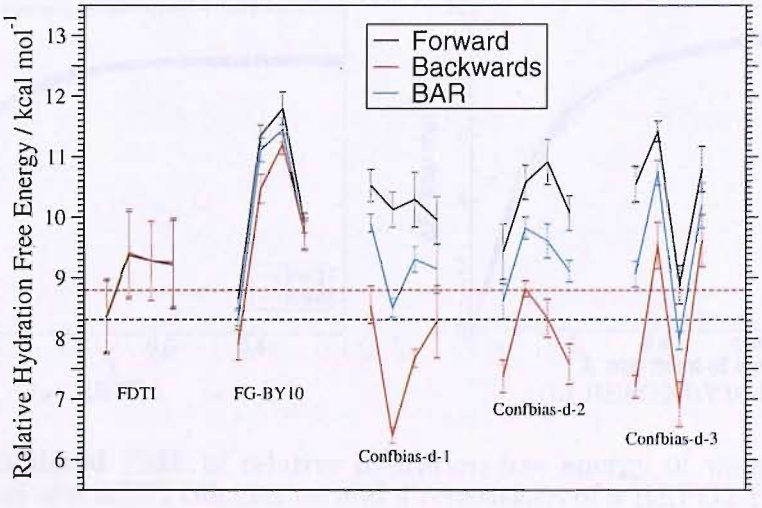


Figure 5.12: Relative hydration free energy of water and methane, estimated by four repetitions for each method. Each set of four estimates is linked with a line and labeled with the method abbrieviation. These estimates are compared to the experimental value of 8.31 kcal.mol^{-1} which is the black dashed line and the RETI-random calculated value of 8.8 kcal.mol^{-1}.

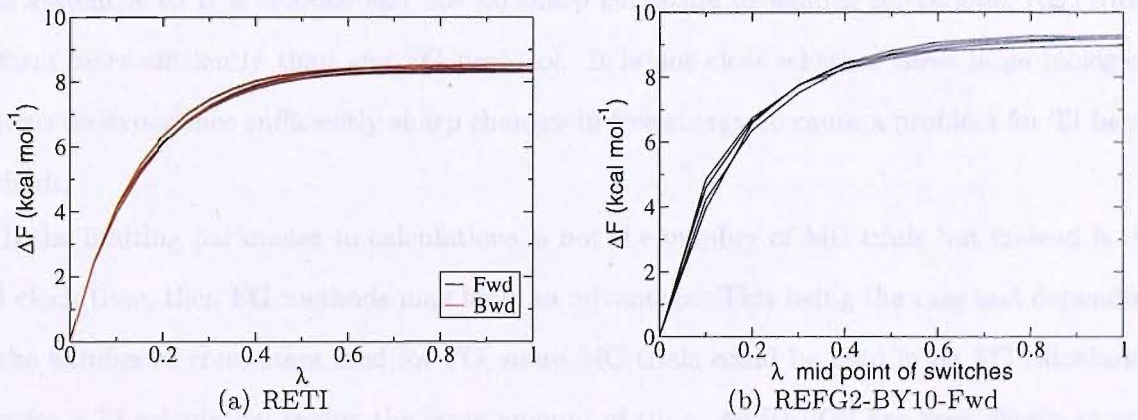
FG protocol	Switches	$\Delta\lambda$s	MC trials	Split
Confbias-d-1	400	500	800	10
Confbias-d-2	400	100	4000	10
Confbias-d-3	400	10	40000	10

Table 5.7: Table describing configurational bias-d FG protocols used in figure 5.12

5.3.3 Water-methane conclusions

The addition of replica exchange moves to FG calculations of the relative hydration free energy of water and methane improves accuracy and precision. REFG2 is able to produce estimates of similar quality to RETI. However, for free energy calculations such as the relative hydration free energy of water and methane RETI should be the method of choice. The PMF of the change in hydration free energy between water and methane is very smooth. This means that a method which makes more evaluations at less discrete points across lambda will produce more easily converged results than a method which makes fewer evaluation at many points across lambda.

Figure 5.13 compares the PMF of four RETI and four REFG2-BY10 forwards Jarzynski calculations. REFG2-BY10 compares well with RETI in this case, especially as half as many configurations are used to produce the REFG2-BY10 forwards estimates as the RETI estimates, although the RETI repeats are slightly more consistent. However, as can be seen from figure 5.7 other REFG protocols and estimators display less reproducible behaviour.



(a) RETI (b) REFG2-BY10-Fwd

Figure 5.13: Calculated PMF of relative hydration free energy of water and methane. From 4 repetitions of a RETI calculation and 4 repetitions of a REFG2-BY10 calcualtion using the forwards Jarzynski estimator. RETI data suplied by C. Woods, taken from [Woods(2003)]

The four RETI calculations of Woods (2003) (figure 5.7) are consistently close to the ex-

perimental value of 8.31 kcal.mol^{-1} $\pm$ 0.5. This may, unfortunately, be due to chance. A more recent RETI-random calculation gives a value with slightly higher precision, of 8.8 kcal.mol^{-1} $\pm$ 0.1 [Woods(2007)]. REFG-BY10 estimates are close to the RETI-random value. It may be that the highly consistent four RETI estimates give a false idea of the level of reproducibility of the RETI protocol, by chance. There is the possibility that there is an issue with the periodicity of λ swap moves in RETI causing differing efficiencies in convergence, although more investigation is required to clear this up. The REFG1 protocol is the same as the REFG2-BY10 protocol except REFG2-BY10 has two λ swaps between each switch starting configuration and REFG1 has only one. Also, REFG1 has 50,000 MC trials between λ swaps whereas REFG2-BY10 has 100,000. The difference in performance between REFG1 and REFG2-BY10 (see figure 5.7) must be attributed to some combination of the extra λ swaps and equilibrium seed simulation MC trials.

When forwards and backwards Jarzynski BAR FG estimators do not give the same value it may be difficult to choose the most accurate estimate. A good method tested here and possibly the most simple is to use the totalled relative entropy measure of Wu and Kofke (2005c) (effectively the $\bar{W}_{diss}$) to find the Jarzynski estimator which is most likely to be accurate, in the case of a BAR estimate not being between forwards and backwards Jarzynski estimates. Alternatively, it may be even simpler to just use BAR on all occasions.

For free energy calculations on chemically and biologically relevant systems, where the PMF from system A to B is smooth and has no sharp curvature or sudden fluctations, RETI may perform more efficiently than any FG protocol. It is not clear whether these large biological systems do experience sufficiently sharp changes in free energy to cause a problem for TI based methods.

If the limiting parameter to calculations is not the number of MC trials but instead is the wall clock time, then FG methods may have an advantage. This being the case and depending on the number of computers used for FG, more MC trials could be used in an FG calculation than for a TI calculation taking the same amount of time. As REFG2 has been shown to give similar performance to RETI with this extra computational advantage both RETI and REFG methods may be equally suitable for calculations similar to the relative hydration free energy of water and methane.

Chapter 6

Protein Ligand Binding Free Energies: Neuraminidase

6.1 Introduction

Neuraminidase is a glycoside hydrolase enzyme and is found on the surface of the influenza virus (EC 3.2.1.18, [Bairoch(2000)]). It is thought that neuraminidase aids in the efficiency of virus release from cells. Thus, neuraminidase has been a major target for drug design programmes and inhibiting compounds in current use as treatments are zanamivir (Relenza) and oseltamivir (Tamiflu).

Neuraminidase inhibitors have been the subject of a recent binding free energy study using RETI with both implicit and explicit solvation [Michel *et al.*(2006)Michel, Verdonk & Essex]. Michel *et al.* used a structure of N9 neuraminidase (PDB code 1BJI) and attempted to predict the binding affinity of 10 ligands originally investigated by Wall *et al.* using LIE [Wall *et al.*(1999)Wall, Leach, Salt, Ford & Essex]. Unfortunately the results of Wall *et al.* and Michel *et al.* cannot be fairly compared as differing receptor structures were used, with Michel *et al.* having the advantage of a newer, possibly more appropriate receptor. Michel *et al.* produced predictions of ligand binding affinities which showed excellent quantitative agreement with experiment using both implicit and explicit methodologies.

Here the reproduction of the explicitly solvated neuraminidase results of Michel *et al.* using FG is investigated. How well do FG methods compare with RETI and experimental analyses?

6.2 The neuraminidase system *in vivo*

Influenza is a wide-spread disease amongst humans, with infections affecting 20 % of the world's population annually [Moscona(2005)]. Influenza can also be extremely virulent. From 1918 to 1920 a pandemic due to the deadly strain H1N1 of influenza A virus killed 50-100 million people worldwide. With the parallels between the recent H5N1 strain and that of the 1918 pandemic, the threat posed by nfluenza is clear [Kamps *et al.*(2006)Kamps, Hoffmann & Preiser]. Through vaccination we can prevent infection and this is the main strategy used to tackle the problem. However, vaccine programs can fail due to antigenic drift, inadequate vaccine supply and the slow nature of present methods of vaccine production. Therefore, anti-viral drugs are important for pandemic strategies and the treatment of individual cases.

The neuraminidase protein has a mushroom shaped head of four co-planar spherical subunits projecting from the virus membrane and a hydrophobic trans-membrane region. Neuraminidase disconnects the influenza viron from cell surface receptors which attach newly released virons to an infected cell (figure 6.1). The neuraminidase active site binds the terminal sialic acid of these cell surface receptors and cleaves the glycosidic linkage between the sialic acid and the adjoining saccharide (EC 3.2.1.18, [Bairoch(2000)]). Without this action the virus is effectively unable to reproduce and the infection is halted.

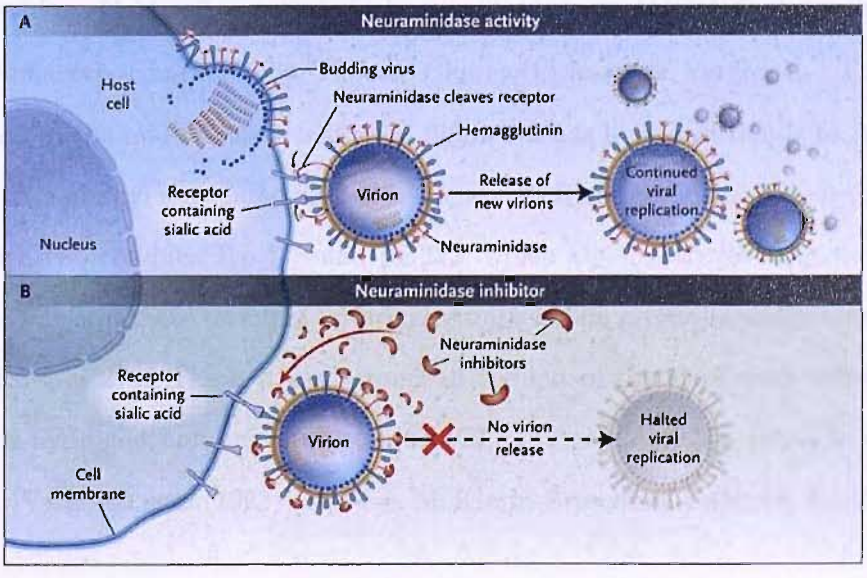


Figure 6.1: The action of neuraminidase in the replication of influenza virons. Taken from [Moscona(2005)]

Effective neuraminidase inhibitors were designed when the three dimensional structure of the enzyme was produced and the position of the active site found [Colman *et al.*(1983)Colman, Varghese & Laver]. Zanamivir and Oseltamivir both closely mimic the natural substrate, sialic acid, but bind more favourably. Current neuraminidase inhibitors have very little toxicity compared to alternatives and are effective against all strains of influenza [Moscona(2005)].

The binding site of neuraminidase is solvent exposed and water molecules play a role in binding of the sialic acid substrate. Figure 6.2 shows the binding site of the x-ray crystal structure of N9 neuraminidase with PDB code 1BJI [Taylor *et al.*(1998)Taylor, Cleasby, Singh, Skarzynski, Wonacott, Smith, Sollis, Howes, Cherry, Colman & Varghese]. The sialic acid substrate is placed correctly in the binding site, although this structure has not undergone any minimisation or equilibration. The slightly enlarged green spheres in figure 6.2 are the oxygen atoms of the crystallographic water molecules present in the binding site. At least one of these bound waters is thought to mediate binding for some inhibitors, including DANA types and amino types [Smith *et al.*(1996)Smith, Sollis, Howes, Cherry, Vobley, Taylor, Whitington, Skarzynski, Cleasby, Singh, Varghese & Colman], [Wall *et al.*(1999)Wall, Leach, Salt, Ford & Essex]. The general protein-ligand interactions seen in figure 6.2 and for all complexes are as follows: the carboxylic acid of the ligand interacts with arginines 118, 292 and 371 through charge-charge based hydrogen bonds, including a possible planar salt bridge to 371 [Varghese *et al.*(1992)Varghese, McKimm-Breschkin, Caldwell, Kortt & Colman], [Taylor *et al.*(1998)Taylor, Cleasby, Singh, Skarzynski, Wonacott, Smith, Sollis, Howes, Cherry, Colman & Varghese]. The acetamido fragment placed in the middle, lower region of figure 6.2 has hydrogen bonds to Arg 152 and a bound water. The methyl associated with this acetamido fragment makes favourable hydrophobic contact with two residues Trp 178 and Ile 222, which are slightly lower in the binding site than Arg 152. The two more terminal hydroxyl groups of the glycerol moiety have a bidentate interaction with Glu 276. The hydroxyl group on the left of the sialic acid molecule in figure 6.2 experiences hydrogen bonding with residues Glu 119 and Asp 151 which is facilitated by a bound water [Varghese *et al.*(1992)Varghese, McKimm-Breschkin, Caldwell, Kortt & Colman].

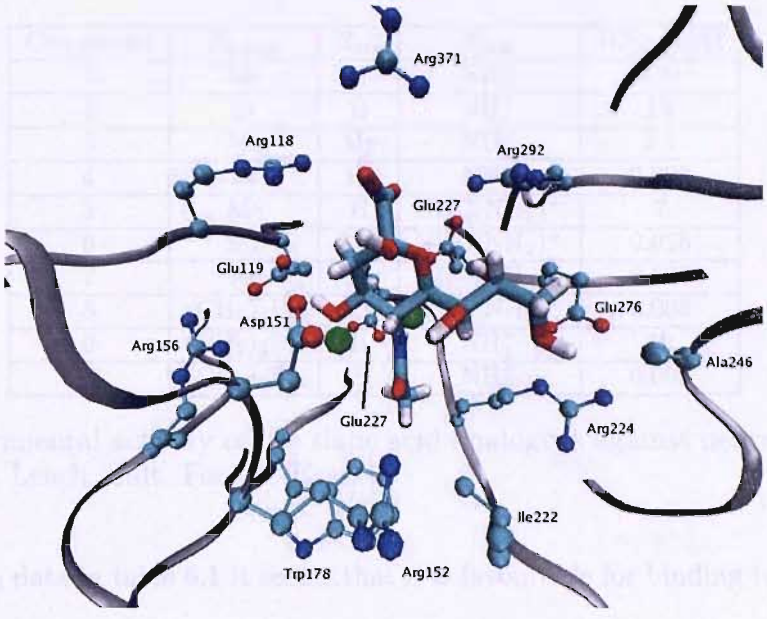


Figure 6.2: The substrate binding site of N9 neuramindase from influenza A (pdb 1BJI [Taylor *et al.*(1998)Taylor, Cleasby, Singh, Skarzynski, Wonacott, Smith, Sollis, Howes, Cherry, Colman & Varghese]) with sialic acid bound. Binding site water oxygen atoms are enlarged and in green. Sidechain hydrogens are removed for clarity. The position of the sialic acid molecule is modelled in the binding site with reference to the sialic acid analogue (ligand 10 in this study) of the 1BJI structure.

Figure 6.3 is a schematic of the basic scaffold for all ligands used in this study, which is based on the substrate sialic acid. The scaffold positions Rcis, Rtrans (referring to the cis and trans pockets described above) and Rpol are filled by different substituents for the different ligands listed in table 6.1.

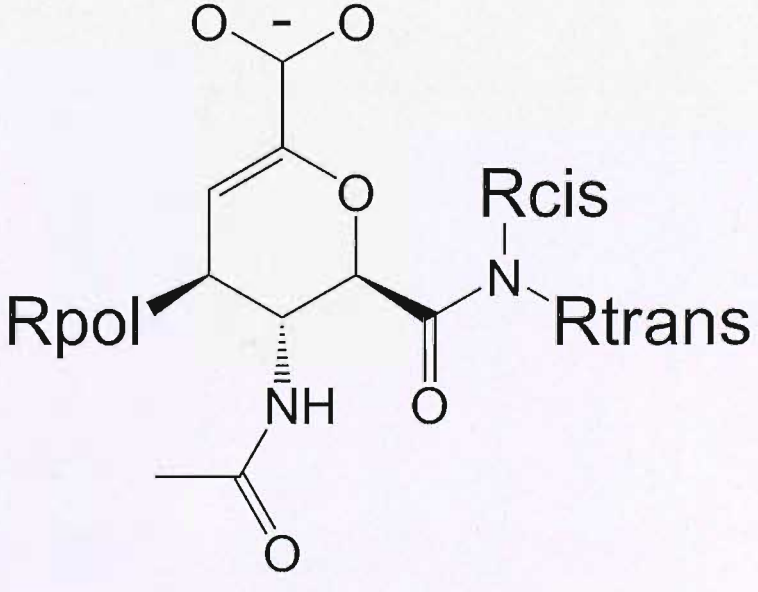


Figure 6.3: 2D chemical structure diagram of the basic scaffold for all neuramindase ligands used in this study.

Compound	R_{trans}	R_{cis}	R_{pol}	IC_{50} (μM)
1	Me	H	NH_3^+	190
2	Et	H	NH_3^+	13
3	Me	Me	NH_3^+	2.4
4	Et	Et	NH_3^+	0.003
5	Me	H	$NHC(NH_2)^+$	7
6	Me	Me	$NHC(NH_2)^+$	0.025
7	Et	Et	$NHC(NH_2)^+$	0.001
8	$(CH_2)_2Ph$	Pr	$NHC(NH_2)^+$	0.005
9	$(CH_2)_2Ph$	H	NH_3^+	12
10	$(CH_2)_2Ph$	Pr	NH_3^+	0.005

Table 6.1: Experimental activity of the sialic acid analogues against neuraminidase [Wall *et al.*(1999)Wall, Leach, Salt, Ford & Essex]

From the IC_{50} data in table 6.1 it seems that it is favourable for binding to fill both the cis and trans pockets. There is little difference between binding affinities of ligands 4, 7, 8 and 10 suggesting there is little profit in bulkier groups than Et for both the trans and cis pockets. The difference between binding affinities of ligands 4, 7, 8 and 10 are within experimental errors of approximately half an order of magnitude [Wall *et al.*(1999)Wall, Leach, Salt, Ford & Essex]. The positioning of the $(CH_2)_2Ph$ and Pr groups in the substrate binding site is shown in figure 6.4. The phenethyl group is placed between Ile222 and Ala246 (the trans pocket), while the propyl is between the Glu276 and Arg224 residues (the cis pocket).

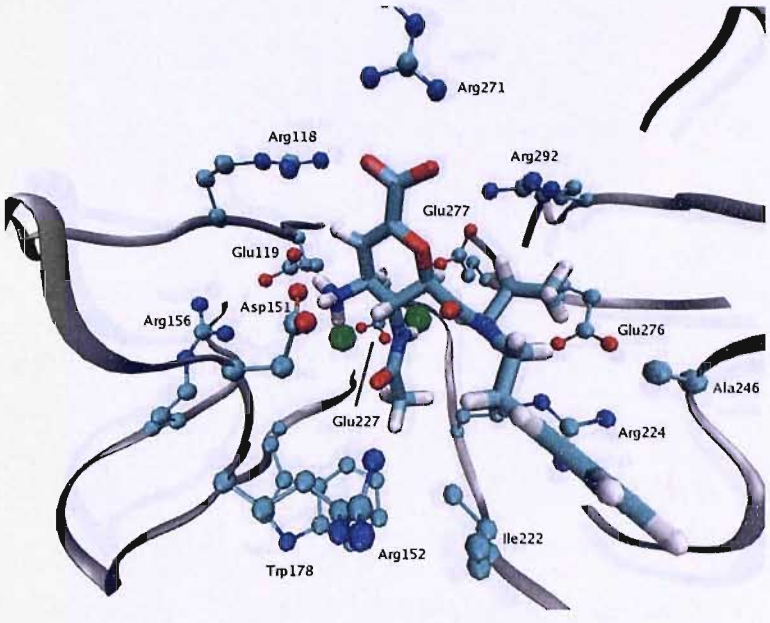


Figure 6.4: The substrate binding site of N9 neuramindase from influenza A (pdb 1BJI [Taylor *et al.*(1998)Taylor, Cleasby, Singh, Skarzynski, Wonacott, Smith, Sollis, Howes, Cherry, Colman & Varghese]) with the amino, sialic acid analogue, ligand 10 bound. Binding site water oxygen atoms are enlarged and in green. Sidechain hydrogens are removed for clarity.

The substitution of an amino group at position Rpol for a guanadino group is also strongly favourable (compare ligand 3 and 6 in table 6.1). This is due to the displacement of a bound water by the guanadino group, which would be present with ligands 1 to 4 ,9 and 10, [Taylor *et al.*(1998)Taylor, Cleasby, Singh, Skarzynski, Wonacott, Smith, Sollis, Howes, Cherry, Colman & Varghese].

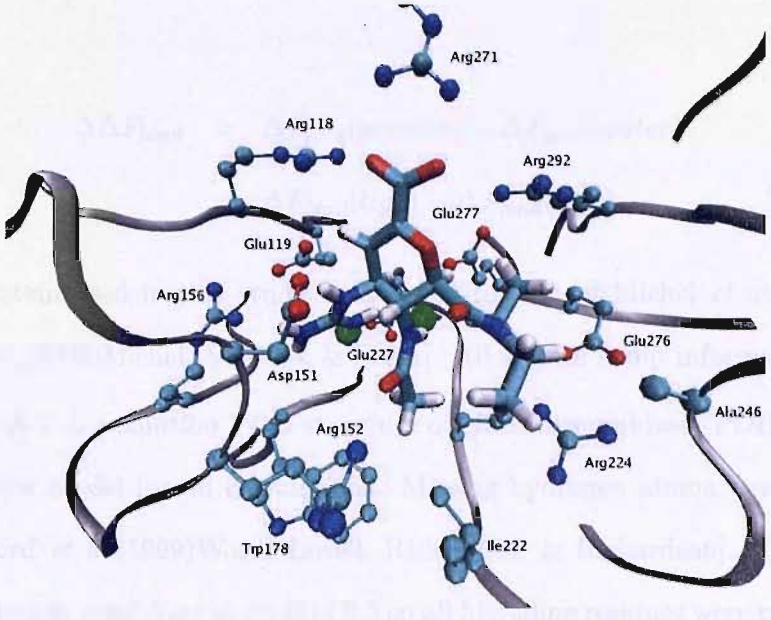


Figure 6.5: The substrate binding site of N9 neuramindase from influenza A (pdb 1BJI [Taylor *et al.*(1998)Taylor, Cleasby, Singh, Skarzynski, Wonacott, Smith, Sollis, Howes, Cherry, Colman & Varghese]) with the guanadino, sialic acid analogue, ligand 7 bound. Binding site water oxygen atoms are enlarged and in green. Sidechain hydrogens are removed for clarity.

6.3 The neuraminidase system *in silico*

This study is concerned with calculating relative binding free energies ($\Delta\Delta F$) of a set of inhibitors for neuraminidase. This allows the ranking of ligands to find those which are the strongest binders. To find $\Delta\Delta F$s for protein-ligand systems it is necessary to run two free energy calculations which are part of a thermodynamic cycle as shown in figure 6.6. $\Delta\Delta F$s are then calculated through equation 6.2.

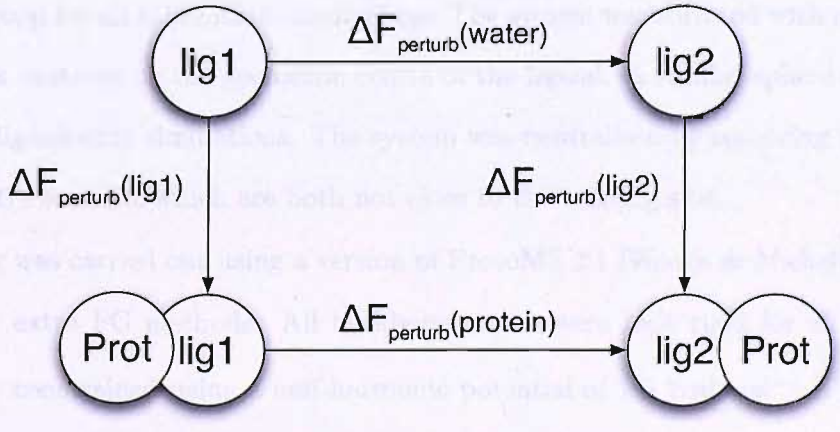


Figure 6.6: Thermodynamic cycle used in the calculation of the relative binding free energy of neuraminidase inhibitors

$$\Delta\Delta F_{bind} = \Delta F_{pert}(protein) - \Delta F_{pert}(water) \quad (6.1)$$

$$= \Delta F_{bind}(lig1) - \Delta F_{bind}(lig2) \quad (6.2)$$

The model system used in this study is identical to that of Michel *et al.* in their recent study [Michel *et al.*(2006)Michel, Verdonk & Essex]. All system setup information is described for completeness. A 2 Å resolution PDB structure of N9 neuraminidase (PDB code 1BJI) was used as the receptor model for all calculations. Missing hydrogen atoms were added to 1BJI using Reduce [Word *et al.*(1999)Word, Lovell, Richardson & Richardson]. The experimental studies on these ligands were done at a pH of 6.5 so all histadine residues were protonated unless evidence of an available hydrogen bonding interaction was found. All ligands were placed in the binding site with the bonding mode of ligand 10 in the crystal structure 1BJI.

The AMBER99 forcefield was used for the protein and GAFF was used to parametrise the ligands [Pearlman *et al.*(1995)Pearlman, Case, Caldwell, Ross, Cheatham, Debolt, Ferguson, Seibel & Kollman]. Partial charges were produced using the AM1/BCC method as implemented in AMBER 8 [Case(2004)] [Jakalian *et al.*(2002)Jakalian, Jack & Bayly].

The complex of protein and ligand 10 was then put through a short minimisation using the Sander simulation program of AMBER 8 with Generalised Born (GB) solvation [Case(2004)]. This minimisation consisted of 100 steps of steepest descent followed by 400 steps of conjugate gradient and was intended only to remove bad contacts.

All residues outside 15 Å from any heavy atom of the ligand (ligand 10) were removed to give a protein scoop for all subsequent simulations. The system was solvated with a 22 Å sphere of TIP4P waters centered on the geometric centre of the ligand. A similar sphere of water was used to solvate ligand-only simulations. The system was neutralised by removing the charge of lysine residues 273 and 432 which are both not close to the binding site.

All sampling was carried out using a version of ProtoMS 2.1 [Woods & Michel(2005)] modified to run any extra FG methods. All backbone atoms were kept rigid for all simulations. The solvent was constrained using a half-harmonic potential of 1.5 kcal.mol^{-1}.Å^{-1} applied to water oxygen atoms which go beyond 22 Å from the geometric centre of the starting position of ligand 10. Residues with all heavy atoms further than 10 Å from all the ligand heavy atoms

were rigid and not sampled. The neutralised lysine residues were in this frozen region. Also, a non-bonded cut off of 10 Å was imposed to increase the efficiency of sampling. This setup gives a belly simulation with a moving core surrounded by a frozen shell. Within the moving core bond angles and torsions of sidechains were sampled except for those which are part of a ring structure. The ligand was sampled in a similar way with the addition of rigid body translations (0.03 Å) and rotations (0.1°).

Simulations were run at a temperature of 37 °C in agreement with experiment [Taylor *et al.*(1998)Taylor, Cleasby, Singh, Skarzynski, Wonacott, Smith, Sollis, Howes, Cherry, Colman & Varghese]. For simulations of protein-ligand complexes, solvent, protein and ligand moves were attempted 85.7 %, 12.8 % and 1.4 % of the time. For ligand-only simulations, solvent and ligand moves were attempted 98.4 % and 1.6 % of the time. In preparation for equilibrium simulations the solvent was equilibrated for 20 million moves, with the protein held rigid, in an attempt to remove any repulsive contacts with the protein or ligand. Then the whole system was equilibrated for another 20 million moves at one extreme of the coupling parameter which corresponds to the largest ligand. The equilibrated system configuration is then used to start all simulations, at different λ values, needed for a particular protocol. The RETI analysis of Michel *et al.* equilibrated each λ window simulation for 10 million MC trials before any data was collected. The FG protocols of this study equilibrated each λ window seed simulation for 2 million MC trials. This reduction in the size of the λ equilibration for FG protocols was made to save time. The smaller size of these equilibrations was found to be sufficient through studying the energy fluctuations of these equilibrations and the fact that test ΔFs were consistent with protocols with longer equilibrations.

The perturbations used were picked to close four thermodynamic cycles shown in figure 6.7. This allowed the calculation of hystereses of different paths to the same $\Delta\Delta F$s giving an idea of the consistency of the calculations. Also, the averaging of $\Delta\Delta F$s from different perturbation paths gives more precise results.

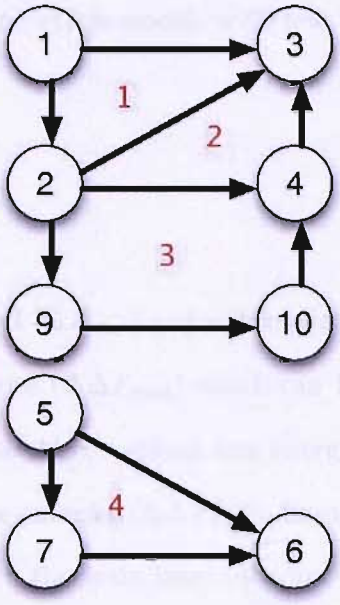


Figure 6.7: Closed perturbation pathways used to find hystereses for consistency checks. Ligand numbers are in black and closed cycle numbers are in red.

Perturbations were performed using a single topology method in a similar style to the water to methane perturbation of chapter 5. The growth of large groups such as benzene rings was made possible by hiding all atoms of the larger group to be grown behind the smaller group from which it is growing. All bonds and atoms in the larger group to be grown were set to 0.2 Å and dummies, respectively, in the end state without the larger group (i.e. not seen by the rest of the system). For example, the perturbation from ligand 2 to 9 grows a benzene ring from a hydrogen. The end state with ligand 2 has a benzene ring of dummy atoms with bonds 0.2 Å long. Even with the benzene ring in its shrunken state, it is unable to hide completely behind the influence of the hydrogen atom from which it grows. However, it is important that as the dummy atoms of the ring are turned on they are not close to any other atoms as this will possibly cause extremely large energies and lead to issues of free energy convergence.

TI calculations carried out by J. Michel used 12 λ windows with 11 spaced uniformly every 0.1 and one at 0.95. Each λ window simulation was equilibrated for 10 million MC trials and then λ measurements were collected for 30 million MC trials. The FDTI $\Delta\lambda$ used in every case was 0.001.

FG protocols for this study were devised with the results and conclusions of chapters 3 and 4 in mind. After experimentation the FG-BY10 protocol was deemed most suitable for this and possibly all protein ligand systems and was therefore tested for original FG, REFG2

and configurational bias FG. Again, FG protocols with few, long switches were employed, as discussed in chapters 3 and 4.

6.4 RETI Results

Table 6.2 shows RETI protein-ligand (ΔF_{prot}) and solvent-ligand ($\Delta\Delta F_{wat}$) free energies which give the relative binding free energies ($\Delta\Delta F_{bind}$) which can be compared to the experimental (Exp) relative binding free energies. Also, vacuum free energies (ΔF_{vac}) are shown, which enable the calculation of solvation free energies ($\Delta\Delta F_{solv}$). Errors were calculated using the block variance method described earlier in the main introduction.

Pert	Exp	$\Delta\Delta G_{bind}$	$\Delta\Delta G_{solv}$	ΔG_{prot}	ΔG_{wat}	ΔG_{vac}
1t2	-1.63	1.44 ± 0.41	1.24 ± 0.32	1.87 ± 0.29	0.43 ± 0.29	-0.81 ± 0.14
1t3	-2.67	-5.25 ± 0.62	-1.58 ± 0.67	21.87 ± 0.34	27.12 ± 0.52	28.70 ± 0.43
2t3	-1.04	-7.19 ± 0.77	-2.52 ± 0.78	19.74 ± 0.43	26.93 ± 0.64	29.45 ± 0.45
3t4	-4.09	-4.14 ± 0.69	1.46 ± 0.54	-8.78 ± 0.51	-4.64 ± 0.47	-6.10 ± 0.27
5t6	-3.45	-5.61 ± 0.58	-1.62 ± 0.62	20.69 ± 0.27	26.30 ± 0.51	27.92 ± 0.35
5t7	-5.15	-7.34 ± 1.18	0.10 ± 1.04	15.12 ± 0.78	22.46 ± 0.88	22.36 ± 0.56
2t4	-5.13	-9.32 ± 1.06	-2.53 ± 1.08	12.04 ± 0.65	21.36 ± 0.84	23.89 ± 0.68
2t9	0.08	-2.56 ± 1.21	0.84 ± 0.93	-7.30 ± 0.80	-4.74 ± 0.91	-5.58 ± 0.20
6t7	-1.70	-1.36 ± 0.66	1.14 ± 0.53	-4.54 ± 0.49	-3.18 ± 0.44	-4.32 ± 0.29
7t8	0.65	-3.97 ± 1.39	3.07 ± 1.04	-4.62 ± 0.99	-0.65 ± 0.97	-3.72 ± 0.37
4t10	0.25	-5.46 ± 1.38	2.34 ± 1.09	-3.12 ± 0.90	2.34 ± 1.04	0.00 ± 0.31
9t10	-4.80	-11.86 ± 1.33	-1.49 ± 1.32	13.33 ± 0.81	25.19 ± 1.06	26.68 ± 0.78

Table 6.2: Experimental and calculated relative binding free energies with the protein-ligand and solvent-ligand free energies used in their calculation. Also, relative solvation free energies calculated with the vacuum-ligand free energies used in their calculation. Experimental free energies are calculated with the formula $\Delta\Delta G = \Delta G_2 - \Delta G_1 = RTln(K_1/K_2)$ assuming that the ratio of the IC_{50}s is equal to the dissociation constants [Cheng & Prusoff(1973)]. All free energies in this table are in kcal.mol^{-1}, were found with RETI and taken from the PhD thesis of J.Michel [Michel(2006)].

The trends in the calculated $\Delta\Delta F_{bind}$s in table 6.2 are similar to those seen from experiment. The calculated $\Delta\Delta F_{bind}$s are in general slightly overestimated.

Table 6.3 shows the changes taking place at the cis and trans positions and compares this with the inaccuracy in $\Delta\Delta F_{bind}$ and the percentage of $\Delta\Delta F_{bind}$ made up by $\Delta\Delta F_{solv}$. In the cis and trans columns the groups listed before and after the "-" is present are at $\lambda = 0$ and $\lambda = 1$, respectively. The same/diff column has "Both diff" and "Both same" if changes at the cis and trans positions are different or the same, respectively. It is interesting that the two

most accurate $\Delta\Delta F_{bind}$ estimates have Me-Et changes at both the cis and trans positions. All $\Delta\Delta F_{bind}$ estimates which grow a $(CH_2)_2Ph$ at the trans position are more inaccurate than the mean unsigned error of 3.39 kcal.mol^{-1}. The Et-$(CH_2)_2Ph$ change of perturbation 2t9 gives a relatively small inaccuracy of 2.64 kcal.mol^{-1}. This suggests that it may be the Et-Pr change which is the source of the relatively high inaccuracies in perturbations 7t8 and 4t10. It is interesting that perturbation 2t3 is very inaccurate as perturbations with similar sized changes in cis and trans groups do not show such large inaccuracies. Hence, the large error for perturbation 2t3 may be due to random sampling error issues.

Pert	trans	cis	same/diff	$\Delta\Delta F_{bind}$ error
1t2	Me-Et	H		-3.07
1t3	Me	H-Me		2.58
2t3	Et-Me	H-Me	Both diff	6.15
3t4	Me-Et	Me-Et	Both same	0.05
5t6	Me	H-Me		2.11
5t7	Me-Et	H-Et	Both diff	2.19
2t4	Et	H-Et		4.19
2t9	Et-$(CH_2)_2Ph$	H		2.64
6t7	Me-Et	Me-Et	Both same	-0.34
7t8	Et-$(CH_2)_2Ph$	Et-Pr	Both diff	4.62
4t10	Et-$(CH_2)_2Ph$	Et-Pr	Both diff	5.71
9t10	$(CH_2)_2Ph$	H-Pr		7.06

Table 6.3: Table showing the perturbations taking place at the cis and trans positions from figure 6.3 and the error of calculated ΔFs compared to experiment. Also, the percentage of the $\Delta\Delta F_{bind}$ value which is contributed by the $\Delta\Delta F_{solv}$ is listed in the last column.

The hystereses of the 4 thermodynamic cycles made by the neuraminidase perturbations are listed in table 6.4. Cycle 2 has a relatively large $\Delta\Delta F_{bind}$ hysteresis of 2.01 kcal.mol^{-1}. However, the total level of error in cycle 1 is larger than cycle 2 as can be seen from table 6.3 and so it could be argued that the large error in perturbation 2t3 has given cycle 1 a high level of consistency, fortuitously. If the possibly random error in perturbation was more in line with errors in other similar perturbations it would be cycle 1 with the high hystereses and not cycle 2.

Cycle pathway	$hyst_{bind}$	$hyst_{solv}$
1 (1,2,3)	0.50 ±1.07	0.30 ± 1.08
2 (2,3,4)	2.01 ± 1.48	1.47 ± 1.43
3 (2,4,9,10)	0.88 ± 2.50	0.46 ± 2.23
4 (5,6,7)	0.37 ± 1.47	0.58 ± 1.32

Table 6.4: Hystereses of 4 thermodynamic cycles for relative binding ($hyst_{bind}$) and solvation ($hyst_{solv}$) free energies of the neuraminidase ligands. All $hyst_{bind}$ and $hyst_{solv}$ values in kcal.mol^{-1}.

Table 6.5 shows the free energy difference of each of the ligands relative to ligand 1. The perturbation 3t6 in table 6.5 is not listed in tables 6.2 and 6.3 above. This is because perturbation 3t6 is problematic due to the necessary expulsion of a water molecule from the binding site as explained earlier [Taylor *et al.*(1998)Taylor, Cleasby, Singh, Skarzynski, Wonacott, Smith, Sollis, Howes, Cherry, Colman & Varghese]. Attempts to tackle this perturbation with relatively simple methods were not successful [Michel(2006)]. The proper treatment of this perturbation would require more complex methodology, outside of the scope of this study. Barillari *et al.* studied this neuraminidase system with the same AMBER99 force field and a more rigourous simulation set up including periodic boundary conditions and protein backbone movement [Barillari *et al.*(2006)Barillari, Taylor, Viner & Essex]. Barillari *et al.* investigated a very similar perturbation with the same water expulsion through the growth of a guanadino group from an amino group. The perturbation consisted of an annihilation of crystallographic bound waters using a double decoupling method [Gilson *et al.*(1997)Gilson, Given, Bush & McCammon] followed by the mutation of ligand 3 to 6. With a non-bonded cut off of 30 Å Barillari *et al.* found a relative binding energy of -3.4 ± 1.1 kcal.mol^{-1} which is close to the experimentally measured value of -2.78 kcal.mol^{-1}. Owing to the complexity of the Barillari *et al.* methodology and the problems of applying this large cutoff (30 Å) with the water droplet methodology of the present study Michel opted to use the relative binding energy value of Barillari *et al.* [Michel(2006)]. Hence, the value of -3.4 ± 1.1 kcal.mol^{-1} will be used for the perturbation of 3t6 through out this study.

Table 6.5 shows whether this RETI analysis has predicted the ranking, from best binder to worst, correctly. Overall most ligands are in the correct order. Ligands 8 and 10 are incorrectly predicted to be better binders than ligands 4 and 7. Also, ligand 4 is predicted to have almost the same $\Delta\Delta F_{bind}$ as ligand 6, whereas experimental $\Delta\Delta F_{bind}$ show that it is a better binder.

Interestingly the large error of perturbation 2t3 discussed above is quenched by the combined error of an opposite sign in perturbations 1t2, giving a $\Delta\Delta F_{bind}$ for ligand 2 which is in line with the general trend. Again this may be fortuitous.

Figure 6.8 again compares the calculated and experimental $\Delta\Delta F_{bind}$ of each ligand relative to ligand 1. Here again, it is clear that the prediction suffers from the large over estimation of ligand 8 (NH_3^+,$(CH_2)_2$Ph,Pr) and 10 ($NHC(NH_2)^+$,$(CH_2)_2$Ph,Pr) $\Delta\Delta F_{bind}$. Figure 6.8 also shows a predictive index (PI) score. PI is a quantitative measure of how useful a predicted set of protein-ligand $\Delta\Delta F_{bind}$s are [Pearlman & Charifson(2001)]. PI is calculated using the formula,

$$PI = \frac{\sum_{j>i}\sum_i w_{ij}C_{ij}}{\sum_{j>i}\sum_i w_{ij}} \tag{6.3}$$

with,

$$w_{ij} = |E(j) - E(i)| \tag{6.4}$$

and,

$$C_{ij} = \begin{cases} +1 & if \quad E(j) - E(i)/P(j) - P(i) < 0 \\ -1 & if \quad E(j) - E(i)/P(j) - P(i) > 0 \\ 0 & if \quad P(j) - P(i) = 0. \end{cases} \tag{6.5}$$

In equations 6.3, 6.4 and 6.5 $P(i)$ is the calculated binding affinity of ligand i and $E(i)$ is the experimental binding affinity of the same ligand. Thus, a PI of 1.0 for a set of ligands means that all ligands are in the correct rank order with respect to the experimental order. A PI of -1.0 means that the predicted rank order of the ligand is the opposite of the experimental order and a PI of 0 means the predicted order is random.

The PI score of the set of estimates shown in figure 6.8 is 0.93. This is a very good score and suggests that the prediction has a near correct rank order and would be very useful in deciding which ligands to investigate further. The mean unsigned error (MUE) of this analysis is 3.37 kcal.mol^{-1}, which is quite high. This high error is obviously largely due to the over estimation of ligands 8 and 10. Figure 6.8 also shows the coefficient of determination which is 0.82.

Compound	Perturbation pathway	Calc $\Delta\Delta F_{bind}$	Expl $\Delta\Delta F_{bind}$
7	[1t3+3t6+6t7]	-10.01 ± 1.42	-7.15
4	[1t3+3t4];[1t2+2t4]	-8.64 ± 1.03	-6.76
8	[1t3+3t6+6t7+7t8]	-13.98 ± 1.99	-6.51
10	[1t2+2t9+9t10];[1t3+3t4+4t10]	-13.92 ± 1.75	-6.51
6	[1t3+3t6]	-8.65 ± 1.26	-5.45
3	[1t3]	-5.25 ± 0.62	-2.67
5	[1t3+3t6+6t5]	-3.04 ± 1.39	-2.00
9	[1t2+2t9]	-1.12 ± 1.28	-1.71
2	[1t2];[1t3+3t2]	1.69 ± 0.70	-1.63
1		0	0

Table 6.5: Experimental and RETI calculated binding free energies with respect to ligand 1. All free energies in this table are in kcal.mol^{-1}.

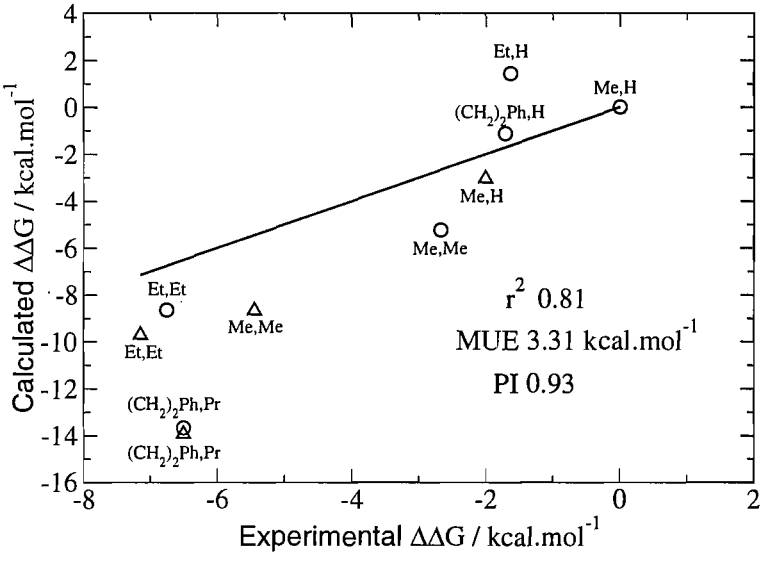


Figure 6.8: Comparison of calculated and experimental $\Delta\Delta F$s for 10 Neuraminidase ligands. Data points are labelled with the chemical groups at the trans and cis positions from table 6.1. Taken from J.Michel *et al.* 2006 [Michel *et al.*(2006)Michel, Verdonk & Essex].

6.5 FG Results

Table 6.6 shows the FG-BY10 calculated and experimental $\Delta\Delta F_{bind}$s, calculated $\Delta\Delta F_{solv}$s and the ΔF_{prot}s, ΔF_{wat}s and ΔF_{vac}s needed for their calculation. This FG-BY10 analysis used 400 switches of 750 MC trials between each of 1000 $\Delta\lambda$s with 71300 MC trials of seed simulation between each starting configuration (400x1000x750). This adds up to 318 million MC trials to obtain a ΔF_{prot} or ΔF_{solv} which is similar to the 360 million MC trials used for the same calculation in the RETI analysis above. Errors in table 6.6 were calculated using the block variance method described earlier in the main introduction with 2 blocks.

Pert	Exp	$\Delta\Delta F_{bind}$	$\Delta\Delta F_{solv}$	ΔF_{prot}	ΔF_{wat}	ΔF_{vac}
1t2	-1.63	1.61 ± 0.21	-0.68 ± 0.20	-0.03 ± 0.13	-1.64 ± 0.17	-0.96 ± 0.11
1t3	-2.67	-4.64 ± 0.19	-1.32 ± 0.15	22.60 ± 0.11	27.24 ± 0.15	28.57 ± 0.01
2t3	-1.04	-2.67 ± 0.32	-3.54 ± 0.17	23.25 ± 0.27	25.92 ± 0.17	29.46 ± 0.01
3t4	-4.09	-4.48 ± 0.23	-0.10 ± 0.12	-11.11 ± 0.20	-6.63 ± 0.11	-6.53 ± 0.03
5t6	-3.45	-4.53 ± 0.15	-2.72 ± 0.13	20.56 ± 0.08	25.09 ± 0.12	27.81 ± 0.01
5t7	-5.15	-5.28 ± 0.26	-3.17 ± 0.20	13.69 ± 0.17	18.98 ± 0.19	22.15 ± 0.06
2t4	-5.13	-7.38 ± 0.34	-4.83 ± 0.31	10.97 ± 0.14	18.35 ± 0.31	23.18 ± 0.01
2t9	0.08	-5.99 ± 0.56	0.58 ± 0.52	-1.17 ± 0.22	4.82 ± 0.51	4.24 ± 0.03
6t7	-1.70	-2.83 ± 0.21	1.64 ± 0.18	-6.00 ± 0.17	-3.17 ± 0.13	-4.81 ± 0.12
7t8	0.65	-5.17 ± 0.53	1.53 ± 0.45	2.24 ± 0.29	7.41 ± 0.65	5.88 ± 0.04
4t10	0.25	-4.41 ± 0.28	-1.23 ± 0.21	4.68 ± 0.21	9.09 ± 0.19	10.32 ± 0.09
9t10	-4.80	-9.83 ± 0.38	-3.93 ± 0.33	13.23 ± 0.19	23.06 ± 0.33	26.98 ± 0.04

Table 6.6: Experimental and calculated relative binding free energies with the protein-ligand and solvent-ligand free energies used in their calculation. Also, relative solvation free energies calculated with the vacuum-ligand free energies used in their calculation. Experimental free energies are calculated with the formula $\Delta\Delta G = \Delta G_2 - \Delta G_1 = RTln(K_1/K_2)$ assuming that the ratio of the IC_{50}s is equal to the dissociation constants [Cheng & Prusoff(1973)]. All calculated free energies in this table were found using FG-BY10 and BAR with 400 switches of 750 MC trials between each of 1000 $\Delta\lambda$s (FG-BY10-BAR-400x1000x750).

Pert	Cis	Trans	same/diff	$\Delta\Delta F_{bind}$ TI error	$\Delta\Delta F_{bind}$ FG error
1t2		Me-Et		-3.07	-3.24
1t3	H-Me			2.58	1.97
2t3	H-Me	Et-Me	Both diff	6.15	1.63
3t4	Me-Et	Me-Et	Both same	0.05	0.39
5t6	H-Me			2.11	1.08
5t7	H-Et	Me-Et	Both diff	2.19	0.13
2t4	H-Et			4.19	2.25
2t9		Et-$(CH_2)_2$Ph		2.64	6.07
6t7	Me-Et	Me-Et	Both same	-0.34	1.13
7t8	Et-Pr	Et-$(CH_2)_2$Ph	Both diff	4.62	5.82
4t10	Et-Pr	Et-$(CH_2)_2$Ph	Both diff	5.71	4.66
9t10	H-Pr			7.06	5.03

Table 6.7: Table showing the perturbations taking place at the cis and trans positions from figure 6.3 and the error of calculated ΔFs compared to experiment for TI and FG-BY10-BAR-400x1000x750.

The FG-BY10 calculated $\Delta\Delta F_{bind}$s are relatively similar to those of the RETI analysis in table 6.2 and show very similar trends. Table 6.7 shows that errors in $\Delta\Delta F_{bind}$s compared with experiment are very similar to those found with the RETI analysis above (table 6.3). The biggest differences are in perturbations 2t3 and 2t9. The large error seen in perturbation 2t3 with RETI is not present with FG-BY10, however, FG-BY10 shows a large error in perturbation 2t9 not seen with RETI. These differences suggest possible occasional large random errors. The average unsigned error of the $\Delta\Delta F_{bind}$s in table 6.6 is 2.30 kcal.mol^{-1} which is lower than the

RETI equivalent of 3.39 kcal.mol^{-1}. However, the consistency of these FG-BY10 results seems much worse than the RETI results as shown in table 6.8. The closed thermodynamic cycles which showed a very low hystereses for the RETI analysis show much larger hystereses with FG-BY10. Only cycle 2, which showed a hysteresis of hyst$_{bind}$ of 2.01 kcal.mol^{-1} for RETI, is relatively accurate with a low hyst$_{bind}$ of 0.23 kcal.mol^{-1}. This relative inaccuracy in hystereses for FG-BY10 may not show any real lack of consistency as the underlying changes causing it are simply due to the differences in perturbations 2t3 and 2t9 and slightly higher accuracy of perturbations 5t6 and 5t7 (table 6.7).

Cycle pathway	hyst$_{bind}$	hyst$_{solv}$
1 (1,2,3)	3.58 ± 0.43	2.9 ± 0.30
2 (2,3,4)	0.23 ± 0.52	1.19 ± 0.37
3 (2,4,9,10)	4.03 ± 0.80	2.71 ± 0.72
4 (5,6,7)	2.08 ± 0.37	2.09 ± 0.30

Table 6.8: Hystereses of 4 thermodynamic cycles for relative binding (hyst$_{bind}$) and solvation (hyst$_{solv}$) free energies of the neuraminidase ligands found using FG-BY10-BAR-400x1000x750 protocol. All hyst$_{bind}$ and hyst$_{solv}$ values in kcal.mol^{-1}.

The underlying ΔF_{prot}s and ΔF_{wat}s of table 6.6 display significant differences from their RETI counterparts. The FG-BY10-BAR-400x1000x750 ΔF_{prot}, ΔF_{wat} and ΔF_{vac} values for perturbations 2t9, 4t10 and 7t8 show significant differences from their RETI counterparts but strangely the differences seem to be much smaller when comparing RETI and FG-BY10-BAR-400x1000x750 $\Delta\Delta F_{bind}$s and $\Delta\Delta F_{solv}$s. It is not coincidence that each of these perturbations showing the largest discrepancies between FG-BY10-BAR-400x1000x750 and RETI involve the growth of a phenyl into the trans pocket. In investigating this discrepancy it was convenient to concentrate on vacuum perturbations as simulations are faster and more accurate than protein and water perturbations. Figure 6.9 shows the PMF across the λ coordinate for perturbation 2t9 with black vertical lines marking the points at which FDTI λ window simulations are run. The large disparity in the PMFs of FG-BY10-BAR-400x1000x750 and RETI for perturbation 2t9 in a vacuum is clear. The FG-BY10-BAR-400x1000x750 PMF undergoing a large peak between $\lambda = 0$ and $\lambda = 0.1$ which is missed by TI.

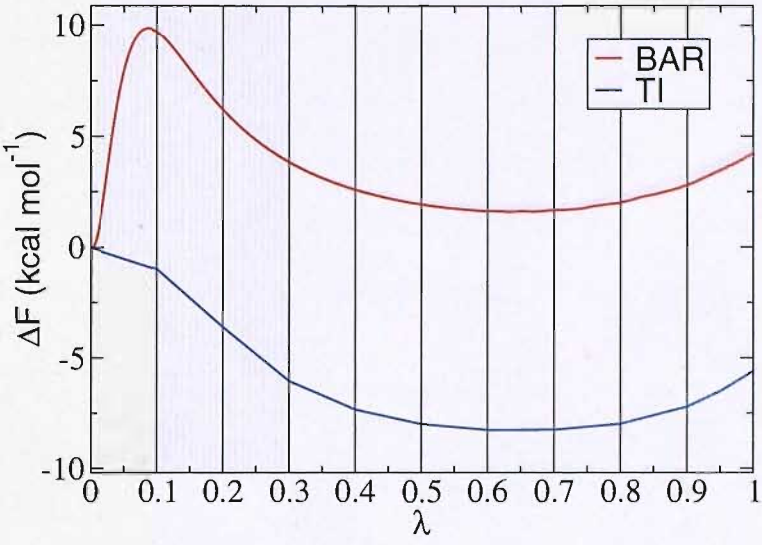


Figure 6.9: Comparison of RETI and FG-BY10 PMFs of perturbation 2t9 in a vacuum.

Figure 6.9 also suggests the reason for the disparity seen in perturbations 2t9, 4t10 and 7t8. The large peak in the BAR PMF between $\lambda = 0$ and $\lambda = 0.1$ is not present in the RETI PMF. Apart from this the two PMFs seem relatively similar. From this it can be deduced that the RETI calculation misses this large peak seen with BAR due to the discrete and non-discrete nature of sampling with RETI and FG-BY10-BAR-400x1000x750 respectively. Figure 6.10 shows the same comparison as figure 6.9 but with 30 RETI $\Delta\lambda$ sampling windows between $\lambda = 0$ and $\lambda = 0.3$ rather than 3. The difference between figures 6.10 and 6.9 is very clear and verifies the idea that the disparity seen between RETI and FG-BY10 for perturbations 2t9, 4t10 and 7t8 is because of a large error in the RETI calculation due to its discrete nature and a lack of necessary $\Delta\lambda$ sampling windows.

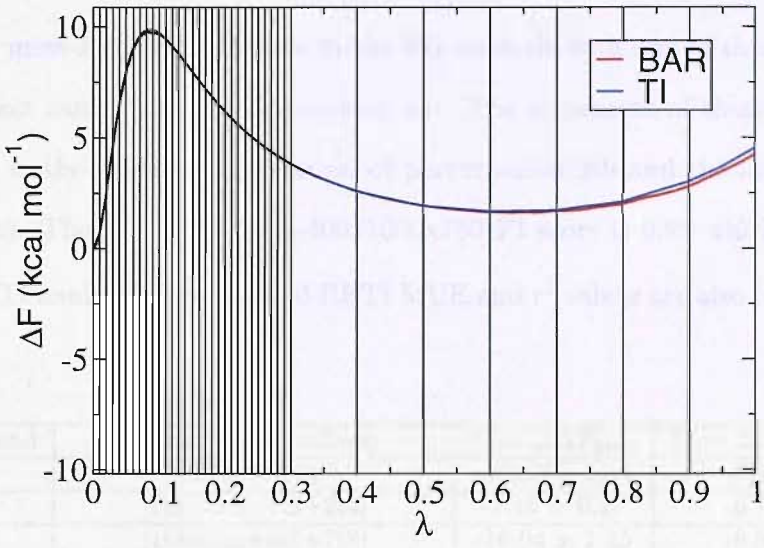


Figure 6.10: Comparison of TI with extra $\Delta\lambda$ sampling windows and FG-BY10 PMFs of perturbation 2t9 in a vacuum. The TI analysis has 30 $\Delta\lambda$ sampling windows between $\lambda = 0$ and $\lambda = 0.3$.

The large peak in the PMF in figure 6.10 is due to very large intra-molecular Lennard-Jones (LJ) and Coulombic forcefield energies. These large LJ and Coulombic energies are produced by the very close proximity of the atoms of the phenyl ring as they are turned on at the beginning of the perturbation (they are 0.2 Å apart at $\lambda = 0$). This means the large peaks in figure 6.10 are an artefact of the perturbation method and would not be expected to be seen commonly in protein-ligand perturbations. The use of larger bond distances for the phenyl group in its shrunken state removes this artefact although increases the risk of clashes with other molecules as the non-bonded terms are turned on. This suggests that the use of a soft core potential [Zacharias *et al.*(1994)Zacharias, Straatsma & McCammon], [Beutler *et al.*(1994)Beutler, Vanschaik, Gerber & van Gunsteren] for perturbation involving the growth a large group may produce a more favourable path. However, the discovery of this error with FG does highlight the importance of sampling across the whole λ coordinate. As the disparity is an internal solute non-bonded effect, it is very well behaved and, as it is present in both protein and water perturbations, is cancelled out in the resulting $\Delta\Delta F_{bind}$ values. The PMFs in figure 6.10 still do not fully converge. There seems to be a small difference between FG-BY10-BAR-400x1000x750 and RETI at the $\lambda = 1$ end. This smaller disparity is again the result of the discrete nature of RETI.

Table 6.9 and figure 6.11 compare the calculated and experimental $\Delta\Delta F_{bind}$ of each ligand relative to ligand 1. The rankings of the ten ligands for FG-BY10-BAR-400x1000x750 and RETI are very similar. The major difference is that the amino, Et, H and amino, $(CH_2)_2Ph$, H

ligands have both moved down the Y axis in the FG analysis with one of these ligand moving back into the correct rank order and one moving out. The movement of these two data points can be attributed to the extra overestimation of perturbation 2t9 and the large improvement in perturbation 2t3. The FG-BY10-BAR-400x1000x750 PI score is 0.95 which is only slightly more than the RETI analysis. The FG and RETI MUE and r^2 values are also both very similar.

Compound	Perturbation pathway	Calc $\Delta\Delta F_{bind}$	Expl $\Delta\Delta F_{bind}$
7	[1t3+3t6+6t7]	-10.87 ± 1.14	-7.15
4	[1t3+3t4];[1t2+2t4]	-7.45 ± 0.25	-6.76
8	[1t3+3t6+6t7+7t8]	-16.04 ± 1.25	-6.51
10	[1t2+2t9+9t10];[1t3+3t4+4t10]	-13.87 ± 0.41	-6.51
6	[1t3+3t6]	-8.04 ± 1.12	-5.45
3	[1t3]	-4.64 ± 0.19	-2.67
5	[1t3+3t6+6t5]	-3.51 ± 0.37	-2.00
9	[1t2+2t9]	-4.38 ± 0.60	-1.71
2	[1t2];[1t3+3t2]	-0.18 ± 0.43	-1.63
1		0	0

Table 6.9: Experimental and FG-BY10-BAR-400x1000x750 calculated binding free energies with respect to ligand 1. All free energies in this table are in kcal.mol^{-1}.

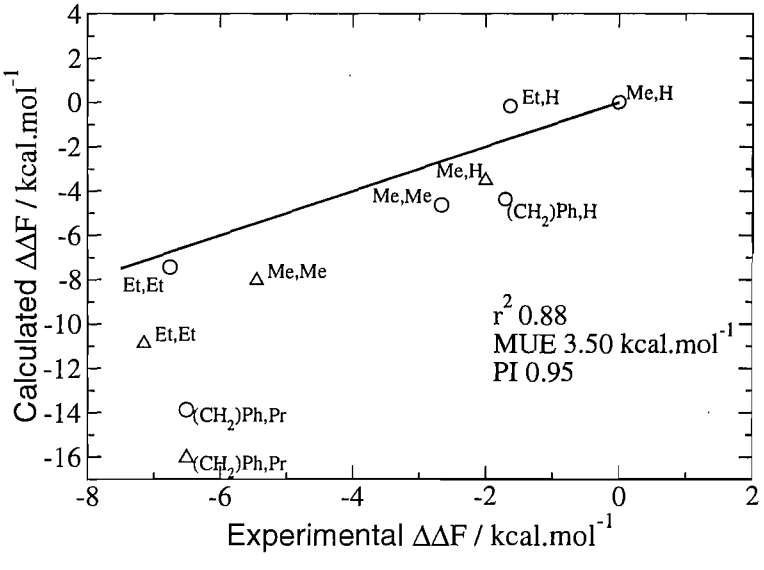


Figure 6.11: Comparison of calculated and experimental $\Delta\Delta F$s for 10 Neuraminidase ligands using FG-BY10-BAR-400x1000x750. Data points are labelled with the chemical groups at the trans and cis positions from table 6.1.

The variation seen between the RETI and FG-BY10-BAR-400x1000x750 analyses seen by comparing figures 6.11 and 6.8 is enough to suggest repeating the FG protocol to check on the possible sensitivity of results. The FG-BY10-BAR-400x1000x750 analysis described here was

therefore repeated another 3 times. These repeats were started with the equilibrated system structure and so all λ equilibrations were included in the repeats. As seen in chapter 5, the statistical errors give a very poor idea of the actual reproducibility of an analysis. Figure 6.12 shows that despite the relatively high r^2 and PI scores of the first FG-BY10-BAR-400x1000x750 analysis discussed above, in general, this protocol may not produce results of high quality. The PI scores of the four repeated analyses range from 0.95 to 0.44.

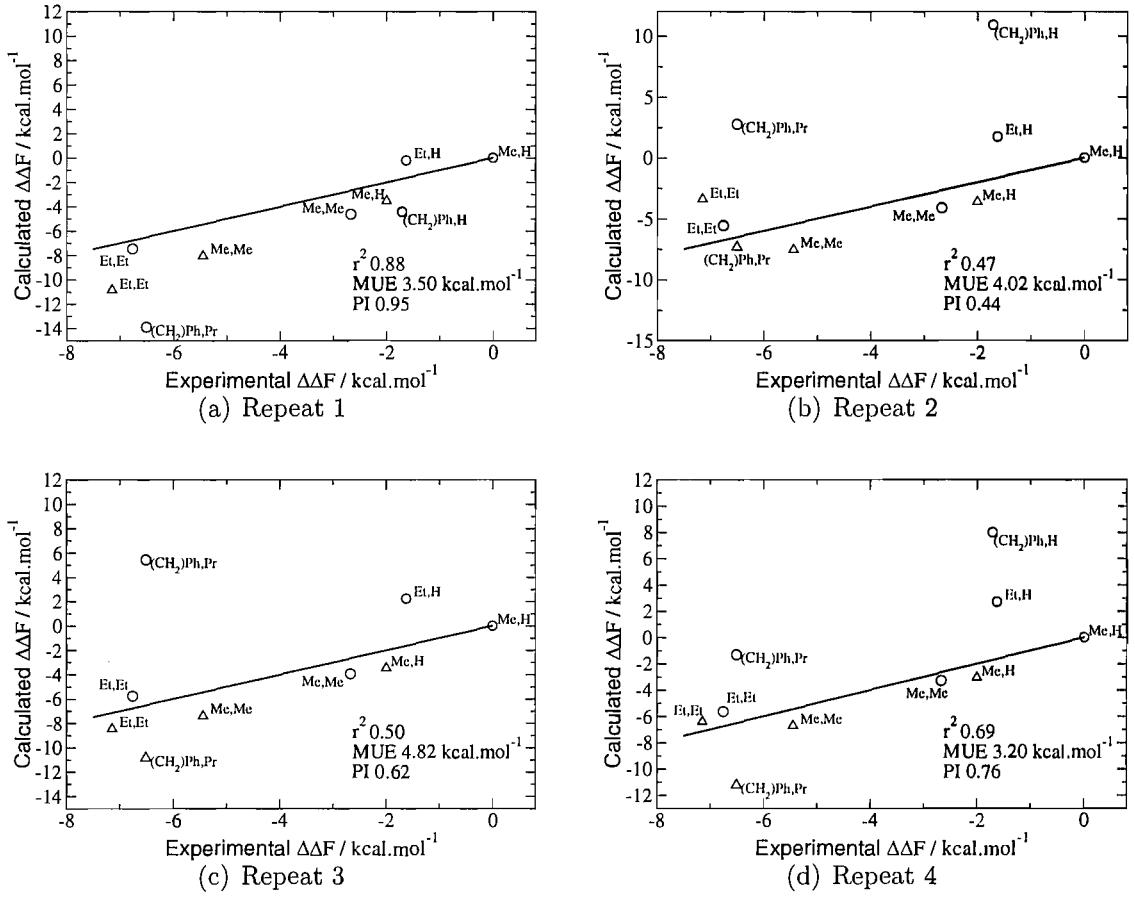


Figure 6.12: Four repeated comparisons of calculated and experimental $\Delta\Delta F$s for 10 Neuraminidase ligands using FG-BY10-BAR-400x1000x750.

Figure 6.13 shows the PI scores of the 4 repeated FG-BY10-BAR-400x1000x750 analyses throughout the calculation. The x-axis shows the number of MC trials for one calculation leg as shown in figure 6.6. Hence figure 6.13 shows the convergence of PI score through the calculation when all simulations are run in parallel. The RETI PI data points are every 100,000 MC trials for each simulation. The FG-BY10-BAR-400x1000x750 PI data points are only calculated every 750,000 MC trials for each simulation, as a new PI can only be found after a set of switches is completed. In the legend of figure 6.13 FG Rep X refers to FG-BY10-BAR-400x1000x750

repeat X.

Figure 6.13 shows that the predictive accuracy of the 4 FG-BY10-BAR-400x1000x750 analyses does not change much through the calculation. Hence the 3 poorer FG repeats are poor through out. The FG-BY10-BAR-400x1000x750 analysis which shows comparable predictive accuracy to RETI (FG Rep 1) seems to converge to a high PI score after a similar number of MC trials have been used.

Figure 6.14 is the same as figure 6.13 but for MUE. The RETI MUE has minimum of about 2.3 kcal.mol^{-1} around a third of the way through the simulation. The final MUE of FG Rep 1, Rep 4 and RETI are all very close at around 3.2 to 3.5 kcal.mol^{-1}. This suggests that MUE is a less sensitive measure of the predictive accuracy of an analysis as FG Rep 4 had a significantly lower PI score than FG Rep 1 and RETI.

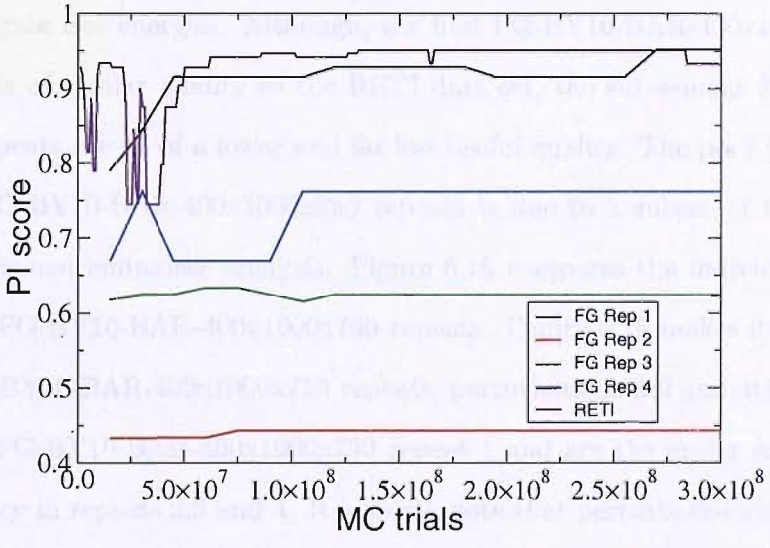


Figure 6.13: Comparison of PI scores for 4 FG-BY10-BAR-400x1000x750 and 1 RETI Neuraminidase analyses.

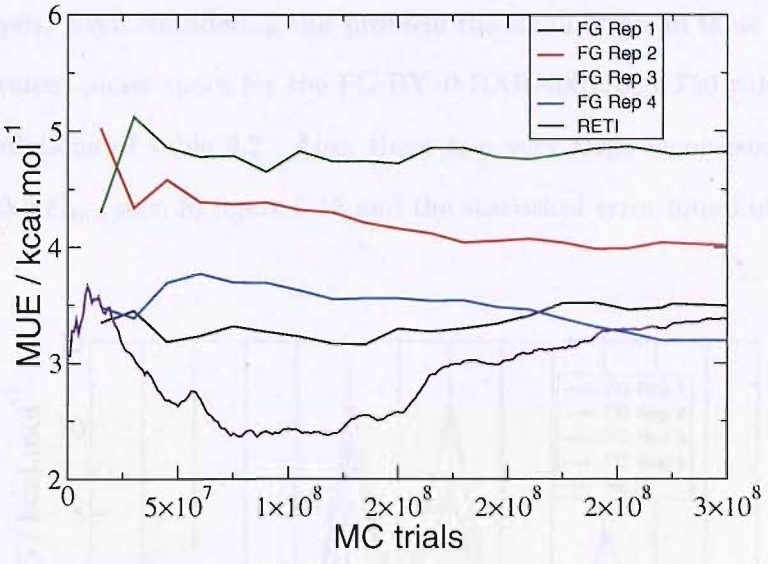


Figure 6.14: Comparison of the MUE for 4 FG-BY10-BAR-400x1000x750 and 1 RETI Neuraminidase analyses.

It is clear that the FG-BY10-BAR-400x1000x750 protocol used here is unable to produce consistently accurate free energies. Although, the first FG-BY10-BAR-400x1000x750 data set discussed above is of similar quality to the RETI data set, the subsequent 3 FG-BY10-BAR-400x1000x750 repeats are all of a lower and far less useful quality. The poor predictive quality of the 3 extra FG-BY10-BAR-400x1000x750 repeats is due to a subset of the perturbations carried out in this neuraminidase analysis. Figure 6.15 compares the individual perturbation $\Delta\Delta F$s for the 4 FG-BY10-BAR-400x1000x750 repeats. Figure 6.15 makes it clear that in the three poorer FG-BY10-BAR-400x1000x750 repeats, perturbations 2t9 and 4t10 differ the most from RETI and FG-BY10-BAR-400x1000x750 repeat 1 and are the major source of the poor predictive accuracy in repeats 2,3 and 4. It is worth note that perturbations 2t9 and 4t10 both involve the growth of a phenyl ring as discussed above.

The repeat of the FG-BY10-BAR-400x1000x750 analysis is also instructive to test the statistical errors found $\Delta\Delta F_{bind}$s in table 6.6, as they are extremely low both in absolute terms and in comparison to their RETI counterparts in table 6.2. These errors have been calculated by finding the standard error between blocks of measurements. As discussed in chapter 3, varying the number of blocks used to calculate statistical errors can affect the size of these errors. In general, the number of blocks is decreased until decreasing them further results in little change in the size of the errors. The small number of measurements used in the FG-BY10-BAR-400x1000x750 protocol allows for only a small amount of adjustment of the error calculation method where as there is much room for adjustment with the many measurements

of the RETI analysis. Even considering this problem the small errors in table 6.6 may signify a low coverage of system phase space for the FG-BY10-BAR-400x1000x750 calculations relative to the RETI simulations of table 6.2. Also, there is a very large inconsistency between the general range in $\Delta\Delta F_{bind}$ seen in figure 6.15 and the statistical error found in table 6.6.

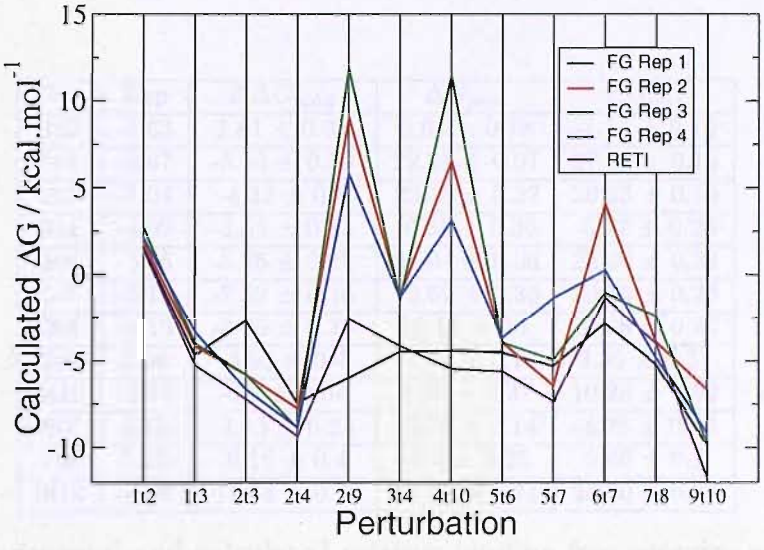


Figure 6.15: Comparison of $\Delta\Delta F_{bind}$s for 4 FG-BY10-BAR-400x1000x750 repeated Neuraminidase analyses.

Considering the improved consistency displayed by the REFG2 protocol in chapter 5, a similar REFG2 protocol was used to see if a similar improvement is found with the neuraminidase system. Table 6.10 shows the results from a neuraminidase analysis using an REFG2 protocol with 800 switches of 375 MC trials between each of 1000 $\Delta\lambda$s with 50000 MC trials of seed simulation between each starting configuration and the BAR estimator (REFG2-BY10-BAR-800x1000x375). This protocol uses 318 million MC trials to obtain a ΔG_{prot} or ΔG_{solv} which is the same as the FG-BY10-BAR-400x1000x750 analysis above and similar to the 360 million MC trials used for the same calculation in the RETI analysis. Errors in table 6.6 were calculated using the block variance method described earlier in the main introduction with 4 blocks.

Comparison of the REFG2-BY10-BAR-800x1000x375 results of table 6.10 to the FG-BY10-BAR-400x1000x750 results of table 6.6 shows no major differences. The MUE of the $\Delta\Delta G_{bind}$s found using the REFG2-BY10-BAR-800x1000x375 protocol is 2.78 kcal.mol^{-1} which is very similar to the equivalent figure for the FG-BY10-BAR-400x1000x750 results of table 6.6 which was 2.30 kcal.mol^{-1}. The highlighted differences, in perturbations 2t3 and 2t9, between the

FG-BY10-BAR-400x1000x750 and RETI analyses of tables 6.6 and 6.2 are not as marked for the REFG2-BY10-BAR-800x1000x375 analysis of table 6.10. The closed thermodynamic cycles which showed relatively large hystereses with FG-BY10-BAR-400x1000x750 are smaller with this REFG2-BY10-BAR-800x1000x375 and of a similar standard to the RETI hystereses in table 6.4.

Pert	Exp	$\Delta\Delta G_{bind}$	ΔG_{prot}	ΔG_{wat}
1t2	-1.63	1.81 ± 0.21	0.67 ± 0.18	-1.14 ± 0.12
1t3	-2.67	-5.53 ± 0.19	22.28 ± 0.07	27.81 ± 0.18
2t3	-1.04	-4.32 ± 0.3	22.21 ± 0.22	26.53 ± 0.19
3t4	-4.09	-3.84 ± 0.43	-9.61 ± 0.36	-5.77 ± 0.23
5t6	-3.45	-5.25 ± 0.22	21.04 ± 0.06	26.29 ± 0.22
5t7	-5.15	-7.52 ± 0.46	12.62 ± 0.36	20.15 ± 0.28
2t4	-5.13	-8.95 ± 0.28	10.13 ± 0.1	19.08 ± 0.26
2t9	0.08	-3.93 ± 0.4	-1.54 ± 0.16	2.39 ± 0.37
4t10	-1.70	-6.2 ± 0.64	4.08 ± 0.37	10.28 ± 0.52
6t7	0.65	-2.53 ± 0.24	-6.78 ± 0.14	-4.25 ± 0.19
7t8	0.25	-6.16 ± 0.4	-0.5 ± 0.25	5.66 ± 0.3
9t10	-4.80	-11.84 ± 0.78	12.86 ± 0.71	24.70 ± 0.33

Table 6.10: Experimental and calculated relative binding free energies with the protein-ligand and solvent-ligand free energies used in their calculation. All free energies in this table are in kcal.mol^{-1} and were found with the REFG2-BY10-BAR-800x1000x375 protocol.

Cycle pathway	$hyst_{bind}$
1 (1,2,3)	3.02 ± 0.41
2 (2,3,4)	0.79 ± 0.59
3 (2,4,9,10)	0.62 ± 1.12
4 (5,6,7)	0.26 ± 0.56

Table 6.11: Hystereses of 4 thermodynamic cycles for relative binding ($hyst_{bind}$) and solvation ($hyst_{solv}$) free energies of the neuraminidase ligands found using REFG2-BY10-BAR-800x1000x375 protocol. All $hyst_{bind}$ and $hyst_{solv}$ values in kcal.mol^{-1}.

Again the REFG2-BY10-BAR-800x1000x375 analysis was repeated four times with the data in tables 6.10 and 6.11 above being for repetition 1. Figure 6.16 shows REFG2-BY10-BAR-800x1000x375 predicted $\Delta\Delta F$s compared to experiment. In general the four REFG2-BY10-BAR-800x1000x375 analyses have a similar PI score to the RETI analysis discussed above. Also, all of the r^2 values of the REFG2 analyses are higher than that of the RETI analysis. This suggests that this REFG2 protocol is at least as predictive as the RETI protocol.

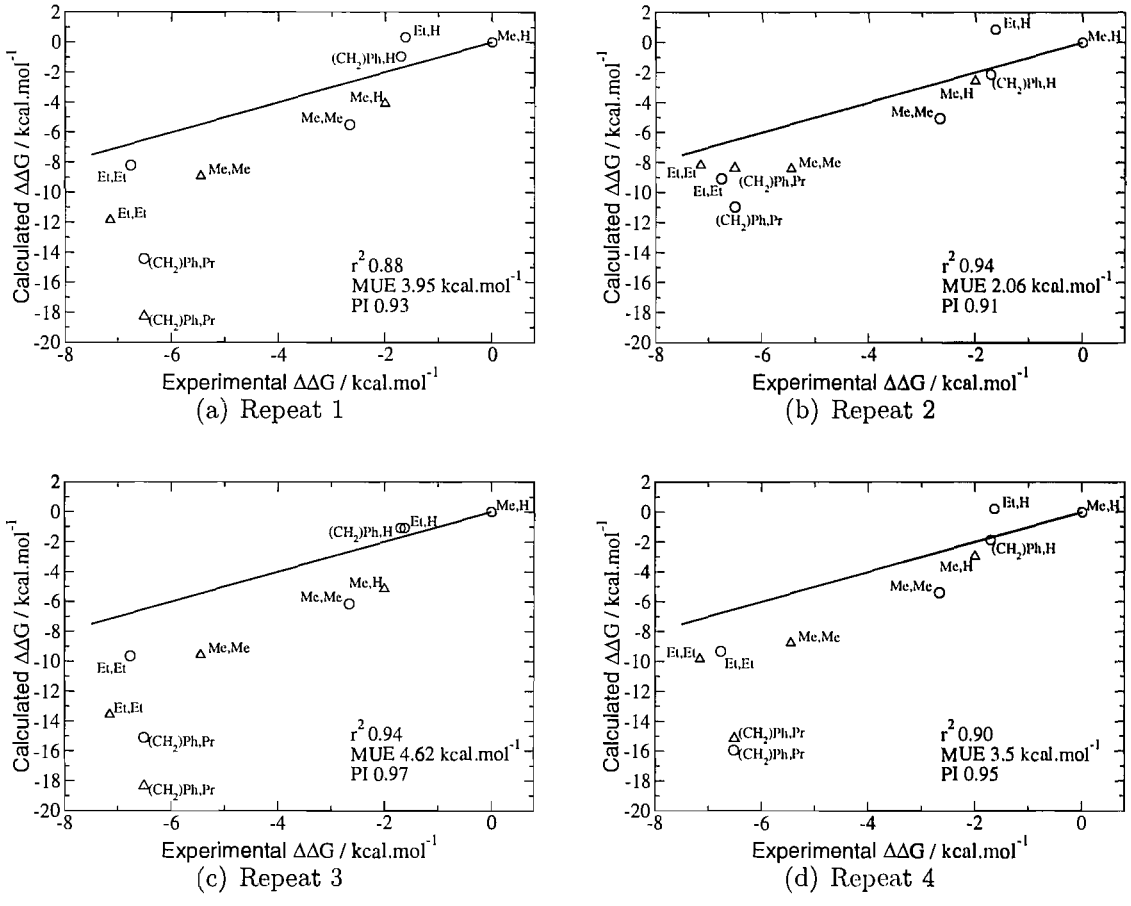


Figure 6.16: Four repeated comparisons of calculated and experimental $\Delta\Delta F$s for 10 Neuraminidase ligands using REFG2-BY10-BAR-800x1000x375.

Interestingly repeat 2 in figure 6.16 has a comparatively low MUE of 2.06 kcal.mol^{-1}. This is due to the fact that, compared with all other neuraminidase analyses, the two $(CH_2)_2$Ph,Pr data points have moved closer to their experimental values. In turn, the difference in the $(CH_2)_2$Ph,Pr data point is due to differing ΔF values for perturbations 4t10, 6t7, 7t8 and 9t10, seen clearly in figure 6.17. This improvement seen only in one repeat is most likely a fortuitous error although the possibility of it being an example of slightly better predictive accuracy with a higher level of convergence remains.

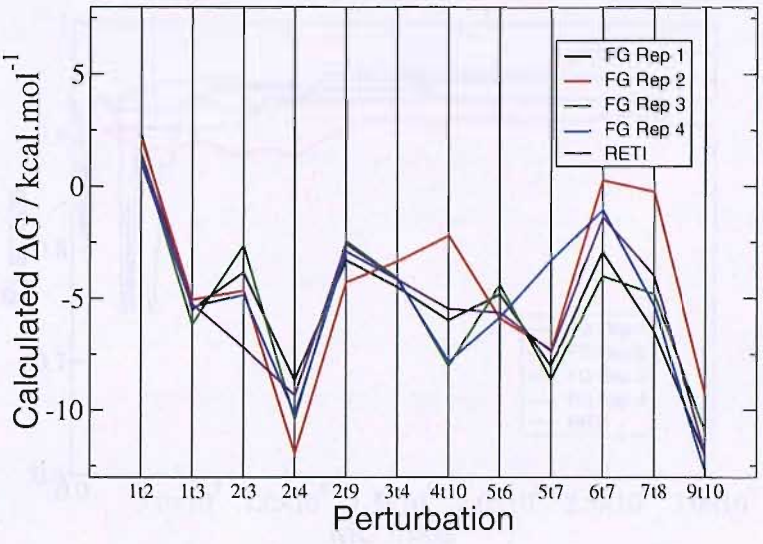


Figure 6.17: Comparison of $\Delta\Delta F$s for 4 REFG2-BY10-BAR-800x1000x375 repeated Neuraminidase analyses.

Figure 6.18 shows the PI score of the four REFG2-BY10-BAR-800x1000x375 analyses throughout the calculations compared to that for the RETI analysis. It is fair to say all are fairly similar in figure 6.18. RETI seems to take slightly longer to converge to a high PI score, this may be due to the fact that the RETI analysis did not include λ swap moves during the equilibration of the λ window simulations. Figure 6.19 shows the MUE of the four REFG2-BY10-BAR-800x1000x375 analyses throughout the calculations compared to that for the RETI analysis. REFG2-BY10-BAR-800x1000x375 repeat 2 starts low and does not fluctuate as much as the other analyses. Again it is interesting how a low MUE does not necessarily translate to a comparatively high PI score, and visa-versa, as seen with repeats 2 and 4.

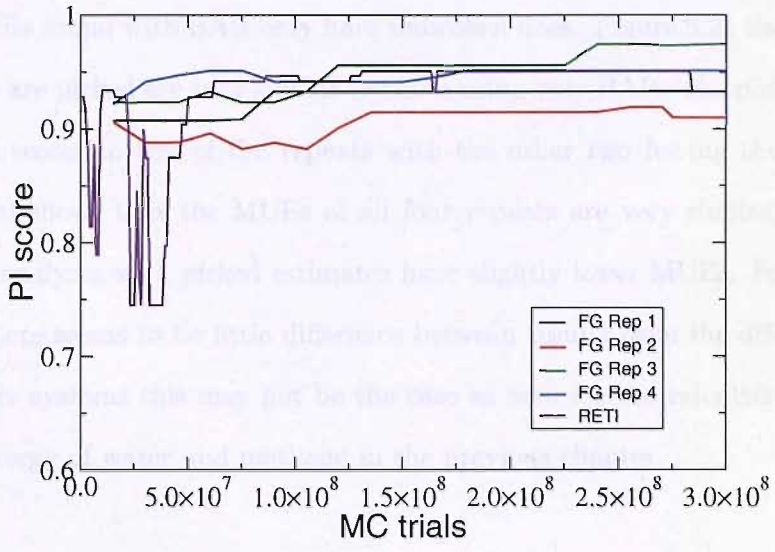


Figure 6.18: Comparison of PI scores for 4 REFG2-BY10-BAR-800x1000x375 and 1 RETI Neuraminidase analyses. The number of MC trials on the x-axis reprents the number of MC trials from one pertyrbation leg i.e. all protein and water simulations are run in parallel.

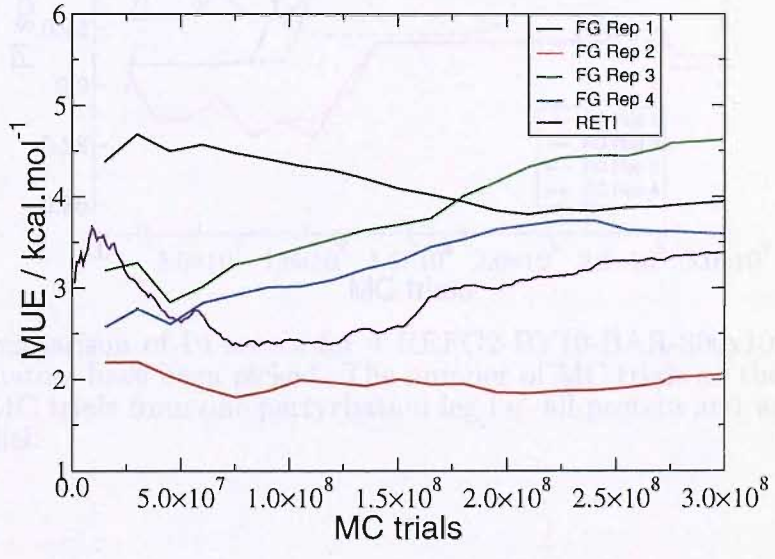


Figure 6.19: Comparison of the MUE for 4 REFG2-BY10-BAR-800x1000x375 and 1 RETI Neuraminidase analyses. The number of MC trials on the x-axis reprents the number of MC trials from one pertyrbation leg i.e. all protein and water simulations are run in parallel.

In previous chapters the idea of picking an FG estimator, for each protein and solvent calculation, from the relative ΔF values has been discussed. If the BAR estimate is between the forwards and backwards Jarzynski estimates, the BAR estimate is chosen. Otherwise the Jarzynski estimate with the largest relative entropy measure as defined by Wu and Kofke [Wu & Kofke(2005a)] is used. Figures 6.20 and 6.21 show the four repeated REFG2-BY10-BAR-800x1000x375 analyses, PIs and MUEs found using these estimator picking rules with a dashed

line, PIs and MUEs found with BAR only have unbroken lines. Figure 6.21 shows that analyses where estimators are picked are very similar to those using only BAR. The picked analyses have slightly lower PI scores in two of the repeats with the other two having the same PI scores. Again, figure 6.20 shows that the MUEs of all four repeats are very similar, although for all four repeats the analyses with picked estimates have slightly lower MUEs. For the system under study here there seems to be little difference between results from the different estimators, although for other systems this may not be the case as seen for the calculation of the relative hydration free energy of water and methane in the previous chapter.

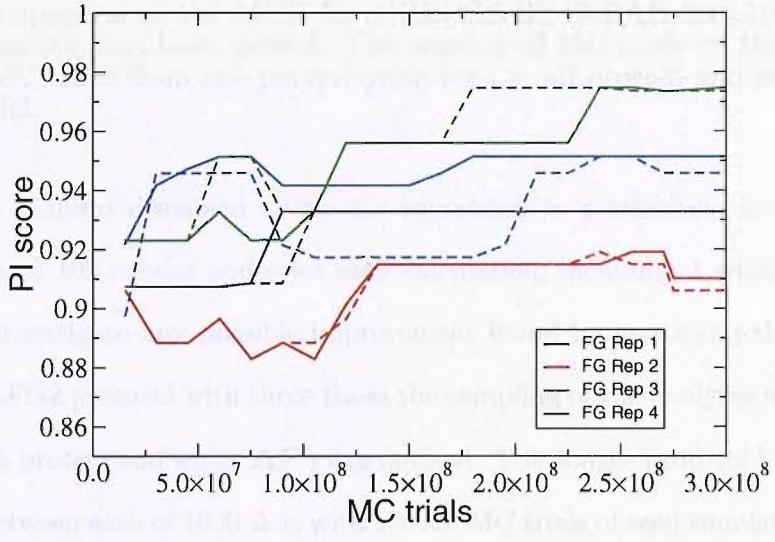


Figure 6.20: Comparison of PI scores for 4 REFG2-BY10-BAR-800x1000x375 analyses where ΔF estimators have been picked. The number of MC trials on the x-axis reprents the number of MC trials from one pertyrbation leg i.e. all protein and water simulations are run in parallel.

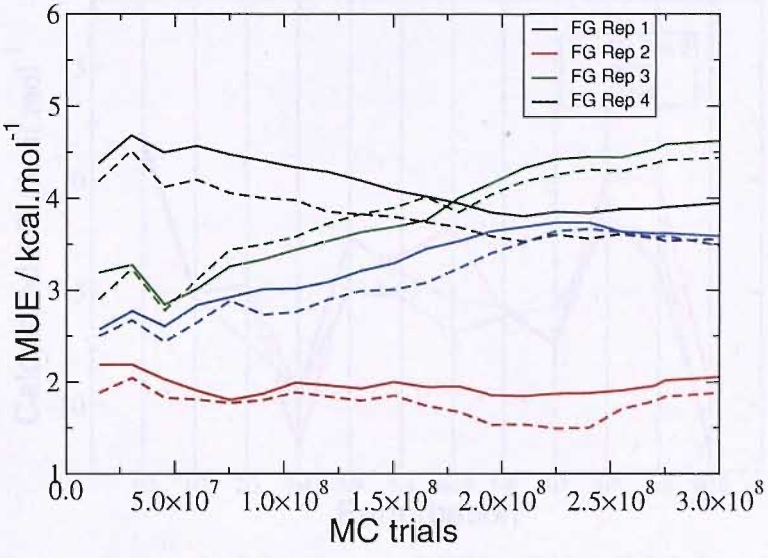


Figure 6.21: Comparison of the MUE for 4 REFG2-BY10-BAR-800x1000x375 analyses where ΔF estimators have been picked. The number of MC trials on the x-axis reprents the number of MC trials from one pertyrbation leg i.e. all protein and water simulations are run in parallel.

As these FG analyses discussed above are completed in a relatively low wall clock time (around 10 hours on 100 condor nodes for each calculation, including λ window equilibration) it is possible to investigate any possible improvement found by increasing the number of MC trials used. A REFG2 protocol with three times the sampling of the analyses above (918 million MC trials for each protein and water ΔF) was devised. This longer protocol has 800 switches of 1125 MC trials between each of 1000 $\Delta\lambda$s with 100000 MC trials of seed simulation between each starting configuration and the BAR estimator (REFG2-BY10-BAR-800x1000x1125). This more expensive protocol can still be completed in approximately 20 hours using 100 condor nodes for each calculation in parallel. In figure 6.22 the new label FG x3 corresponds to the new REFG2-BY10-BAR-800x1000x1125 analysis. Figure 6.22 shows the comparison of calculated $\Delta\Delta F_{bind}$s with experiment, suggesting that the extra sampling gives little advantage.

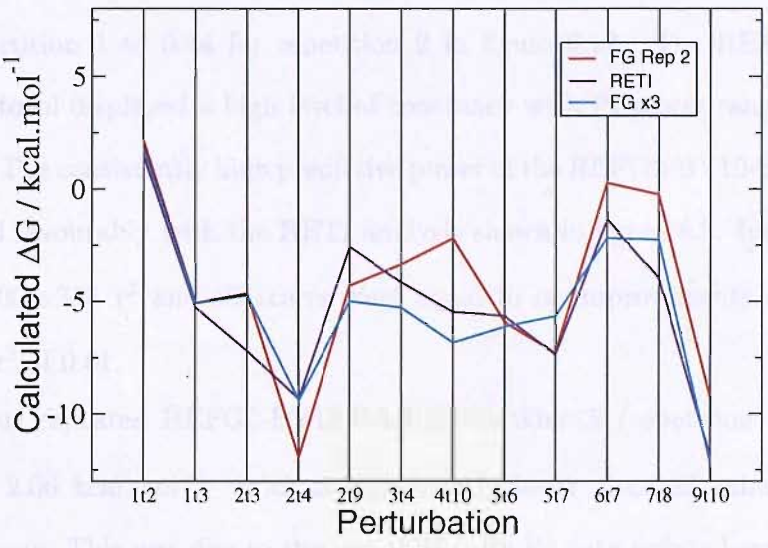


Figure 6.22: Comparison of ΔFs for a REFG2-BY10-BAR-800x1000x1125 Neuraminidase analysis.

Figure 6.23 compares REFG2-BY10-BAR-800x1000x1125 predicted neuraminidase $\Delta\Delta F_{bind}$s to experiment. Again comparing figure 6.23 to figure 6.16 above suggests that the extra sampling in this new analysis offers little predictive improvement.

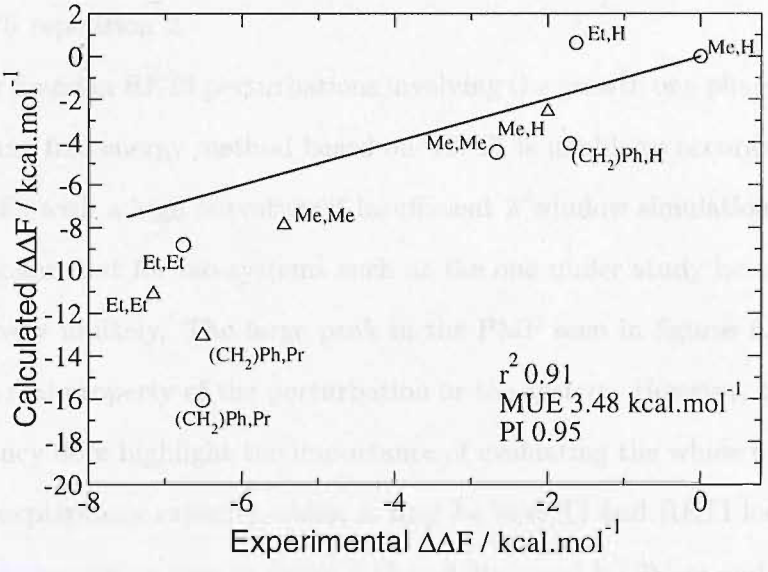


Figure 6.23: Comparison of calculated and experimental $\Delta\Delta F$s for 10 Neuraminidase ligands using REFG2-BY10-BAR-800x1000x1125.

6.6 Conclusions: Does FG offer any thing new?

The FG-BY10-BAR-400x1000x750 analysis was shown to have a large range of predictive performance for this neuraminidase relative binding free energy analysis. The PI scores ranged

from 0.95 for repetition 1 to 0.44 for repetition 2 in figure 6.12. The REFG2-BY10-BAR-800x1000x375 protocol displayed a high level of constancy with PI scores ranging from 0.93 to 0.97 (figure 6.16). The consistently high predictive power of the REFG2-BY10-BAR-800x1000x375 protocol compared favourably with the RETI analysis shown in figure 6.8. Indeed all REFG2-BY10-BAR-800x1000x375 r^2 and PI scores were equal to or improvements on the RETI PI score of 0.93 and r^2 of 0.81.

One of the four repeated REFG2-BY10-BAR-800x1000x375 (repetition 2) analyses produced a MUE of 2.06 kcal.mol^{-1} which is significantly lower than all other neuraminidase analyses in this study. This was due to the two $(CH_2)_2$Ph,Pr data points having moved closer to their experimental values. It is difficult to tell whether this improvement in repetition 2 of figure 6.16 is due to a lack of proper convergence in the other REFG2-BY10-BAR-800x1000x375 repetitions or a lack of convergence in repetition 2. As the REFG2-BY10-BAR-800x1000x375 repetition 2 is the only example of this improvement compared to experiment and the REFG2-BY10-BAR-800x1000x1125 analysis was unable to show this improvement, it may be that this is a chance movement of calculated free energies to give the overall effect in REFG2-BY10-BAR-800x1000x375 repetition 2.

The large error found in RETI perturbations involving the growth of a phenyl ring highlights one deficiency of any free energy method based on TI. TI is unable to accurately calculate the free energy of PMFs with a high curvature if insufficient λ window simulations are performed. It is generally thought that for bio-systems such as the one under study here, a PMF of high curvature is relatively unlikely. The large peak in the PMF seen in figures 6.9 and 6.10 is an artefact and not a real property of the perturbation or the system. However, the unearthing of this large discrepancy does highlight the importance of evaluating the whole perturbation path even if just in an exploratory capacity. Also, it may be that TI and RETI lose accuracy with PMFs of medium curvature as seen in figure 6.10 and discussed by Shirts and Pande [Shirts & Pande(2005)].

The statistical errors for the FG-BY10-BAR-400x1000x750 analysis listed in table 6.6 and the REFG2-BY10-BAR-800x1000x375 analysis listed in table 6.10 are much smaller than the levels of variation of $\Delta\Delta F_{bind}$s seen in figures 6.15 and 6.17. This suggests that statistical errors cannot be used reliably to give an idea of the possible range of variation in binding free energy analyses. Thus, it may be important to repeat calculations especially when testing new

protocols or challenging systems.

When considering which method to use for calculations such as this neuraminidase relative binding free energy analysis, it is clear that FG type methods offer something new and useful. Here the REFG2-BY10-BAR-800x1000x375 protocol has been shown to perform at least as well as RETI. Also, FG discovered some very large errors which were missed with the initial RETI analysis due to its discrete nature. Although, as these errors were present in both protein and water perturbations they were fortuitously cancelled out. Another advantage of FG type methods realised in this study is the possible use of extreme parallelisation. The REFG2-BY10-BAR-800x1000x375 protocol is able to produce a single $\Delta\Delta F_{bind}$ result in around 10 hours on 100 condor nodes while RETI takes around 24 hours on 12 2.2 Ghz Opterons. Owing to the parallelisation of FG the REFG2-BY10-BAR-800x1000x1125 protocol was generally completed in around 20 hours on 100 condor nodes. This is a slightly unfair comparison as the REFG2-BY10-BAR-800x1000x375 protocol involved 2 million MC trials of λ window equilibration while the RETI protocol used 10 million MC trials of λ window equilibration. However, it is clear that FG methods have a large advantage in being able to utilise large clusters of processors. This advantage may become larger as an era of powerful computer processors with many independent processing cores is realised.

Chapter 7

Protein Ligand Binding Free Energies: Cyclooxygenase-2

7.1 Introduction

Cyclooxygenase (COX) is a target for nonsteriodal anti-inflammatory drugs (NSAID) which inhibit the synthesis of prostaglandins by blocking the COX mediated cyclooxygenation of arachidonic acid (AA) to prostaglandin G_2 (PGG_2) (EC 1.14.99.1, [Bairoch(2000)]). PGG_2 is a precursor of many prostaglandins which can possess analgesic, anti-pyretic and anti-inflammatory properties and protection for the gastric mucosa.

The existence of another COX isoform, COX-2 has recently been described and subsequently the crystal structures of both forms elucidated [Picot *et al.*(1994)Picot, Loll & Garavito], [Kurumbail *et al.*(1996)Kurumbail, Stevens, Gierse, McDonald, Stegeman, Pak, Gildehaus, Miyashiro, Penning, Seibert, Isakson & Stallings]. COX-2 is encoded by a different gene to COX-1, is thought to be expressed specifically in inflammatory tissues and is not implicated in production of gastrointestinal tolerability prostaglandins, unlike COX-1 [Xie *et al.*(1991)Xie, Chipman, Robertson, Eriksont & Simmons]. This explains the ulcerogenic side effects of established NSAIDs asprin, ibuprofen and naproxen as these drugs inhibit both COX-1 and COX-2 isoforms [Hawkey(1999)]. A new set of NSAIDs which selectively inhibit COX-2 and therefore display a decreased level of ulcerogenic side effects have been developed. One of the best of these new NSAIDs, celecoxib 1 has been used in treatments for rheumatoid arthritis and osteorarthritis, although is the subject of recent controversy. Celecoxib 1 has also been approved for clinical use against familial adenomatous polyposis in the UK and as the USA.

Celecoxib 1 like, inhibitors have been the subject of two recent binding free energy studies

using RETI and FEP [Michel *et al.*(2006)Michel, Verdonk & Essex], [Price & Jorgensen(2000b)]. Both studies assessed the same set of inhibitors using different free energy methods, force fields and amounts of computation producing binding affinities which showed excellent quantitative agreement with experiment. Here again the results of analysis with FG methods will be compared to more established methods, RETI and FEP.

7.2 The COX-2 system *in vivo*

Prostaglandins are autocrine and paracrine lipid mediators which ligate a subfamily of G-protein-coupled receptors. These receptors are quite varied, and hence prostaglandins have a range of actions which include:

- the constriction or dilation in vascular smooth muscle cells
- the aggregation or disaggregation of blood platelets
- the constriction of smooth muscle
- the regulation of calcium movement
- the regulation of hormone regulation
- the control of cell growth
- the regulation of inflammatory mediation
- the production of pain in spinal neurons

COX facilitates the conversions of AA to PGG_2 using the cyclooxygenase site and PGG_2 to prostaglandin H_2 (PGH_2) through a heme site with peroxidase activity. These COX steps are the rate limiting steps in all prostaglandin biosynthesis. Presently there are three known COX isoenzymes COX-1, 2 and 3. COX-3 is a splice variant of COX-1, while as previously described COX-2 is coded by a separate gene [Chandrasekharan *et al.*(2002)Chandrasekharan, Dai, Lamar Turepu Roos, Evanson, Tomsik, Elton & Simmons]. COX-1 is considered a constitutive enzyme being found in the majority of mammalian cells. COX-2 has not been detected in most tissues and expression is thought to be induced in cells at sites of inflammation.

The non-specific NSAID COX inhibitors have adverse effects such as peptic ulceration and dyspepsia. These adverse effects may be due to direct irritation of the gastric mucosa as many non-specific NSAIDs are acids, as well as reducing production of the protective prostaglandins in the gastrointestinal tract [Price & Fletcher(1990)]. NSAIDs selective for COX-2 are thought to halve the risk of peptic ulceration. However, COX-2 selectivity does not help other possible side effects of NSAIDs such as the increased risk of renal failure [Malhotra *et al.*(2004)Malhotra, Shafiz & Pandhi].

Very recent studies have suggested COX-2 selective NSAIDs increase the risk of myocardial infarction, even with short term use [Kearney *et al.*(2006)Kearney, Baigent, Godwin, Halls, Emberson & Patrono]. Also, courses of large doses of traditional NSAIDs have been associated with a similar increased risk of vascular events. Rofecoxib, a NSAID which is strongly COX-2 selective, was taken of the market in 2004 due to these concerns, while celecoxib with a lower selective strength remains but is the subject of an FDA alert in the US and is prescribed with care [Chan(2006)]. New selective COX-2 inhibitors have been produced and are in use around the world although none have been approved in the US. Also, it has been found that neuroblastomas have abnormal expression levels of COX-2. COX-2 seems to reduce the action of the p53 tumour suppressor. Thus, celecoxib has been shown to help restore p53 function and reduce tumour growth [Johnsen *et al.*(2004)Johnsen, Lindskog, Ponthan, Pettersen, Elfman, Orrego & Sveinbjörnsson].

COX-1 and 2 have a similar size and make up and with 599 and 604 residues respectively and 65% amino acid sequence homology. The COX-1 and 2 active sites are very similar with the most significant difference being the substitution of isoleucine 523 in COX-1 for valine in COX-2. The smaller Val residue in COX-2 opens up a hydrophobic pocket in the active site which the Ile residue obscures.

The COX2 active site is a hydrophobic channel stretching from the membrane bound portion of the protein. There is a relatively small entrance area at one side of the binding site which is gated by Arg120, Glu524, Tyr355, and Arg513 forming hydrogen bonds, which is marked with a dashed orange circle in figure 7.1.

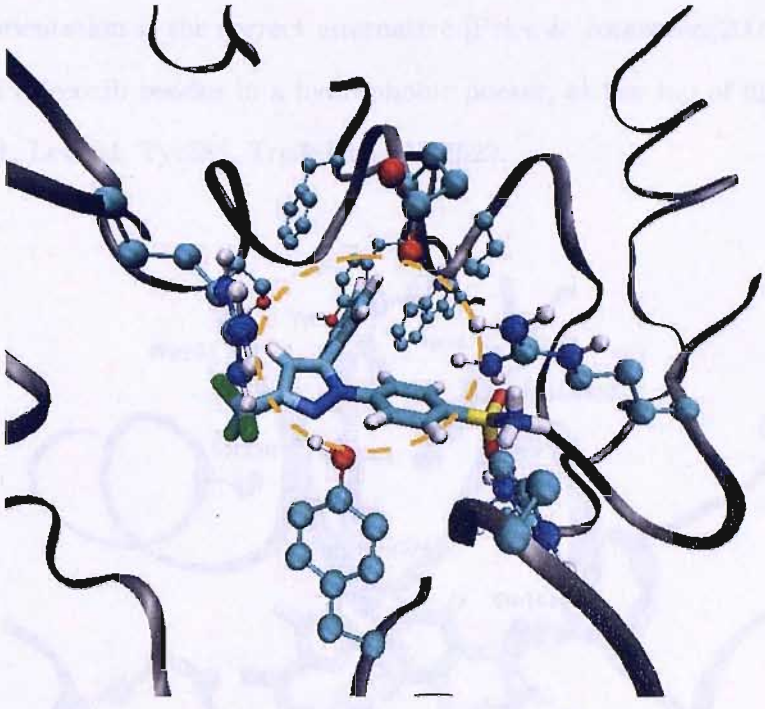


Figure 7.1: The substrate binding site of murine COX-2 (pdb 1CX2 [Kurumbail *et al.*(1996)Kurumbail, Stevens, Gierse, McDonald, Stegeman, Pak, Gildehaus, Miyashiro, Penning, Seibert, Isakson & Stallings]) with compound SC558 (celecoxib) bound. The structure is oriented to show the entrance area of the binding site highlighted in orange. Much of the protein backbone has been removed for clarity.

The residues of COX-2 which have important interactions with binding compounds can be seen in figure 7.2. The trifluromethyl group of celecoxib interacts with the guanidinium group of Arg120 [Kurumbail *et al.*(1996)Kurumbail, Stevens, Gierse, McDonald, Stegeman, Pak, Gildehaus, Miyashiro, Penning, Seibert, Isakson & Stallings]. This is a weakly favourable interaction when compared to the salt bridge formed by many traditional NSAIDs in this area of the binding site and may contribute towards COX-2 selectivity [Grieg *et al.*(1997)Grieg, Francis, Falgueyret, Ouellet, Percival, Roy, Bayly & O'Neill]. In the COX-2 binding site the sulphonamide group is able to reside in the relatively polar pocket made by residues Val523, Arg513, Gln192 and His90. There has, however, been ambiguity over the orientation of the sulphonamide with two crystal structures from the same group (1CX2 and 6COX) showing different sulphonamide positions. Figure 7.2 show the sulphonamide position of structure 1CX2. In 1CX2 an oxygen of the sulphonamide seems to interact unfavourably with the carbonyl oxygen of Gln192. In structure 6COX the sulphonamide is rotated and hydrogen bonds can form between the sulphonamide NHs and the Gln192 carbonyl oxygen. The docking and FEP analyses of the positioning of this sulphonamide group carried out by Price and Jorgensen suggest

that the 6COX orientation is the correct alternative [Price & Jorgensen(2000b)]. The phenyl-bromine group of celecoxib resides in a hydrophobic pocket, at the top of figure 7.2, made up of Ser530, Phe381, Leu384, Tyr385, Trp385 and Met522.

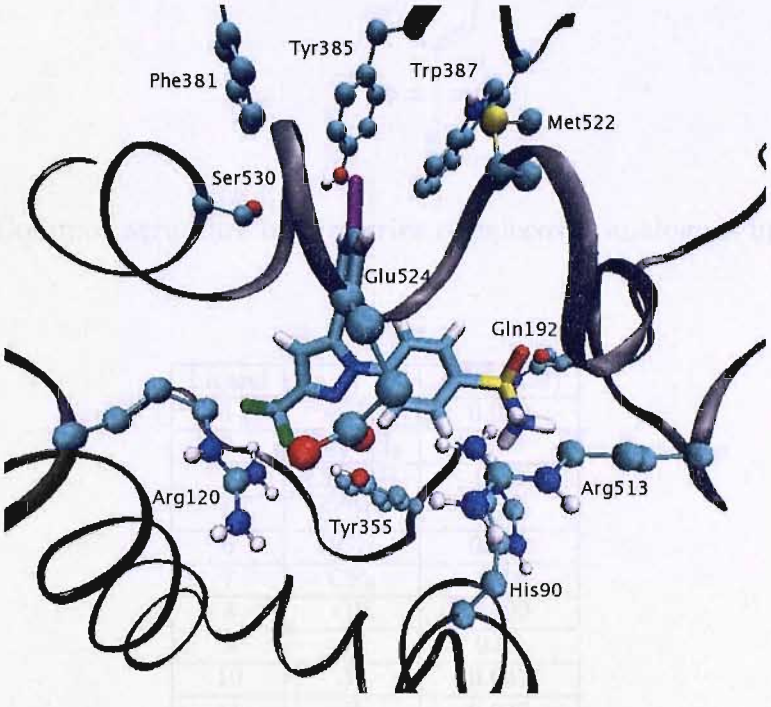


Figure 7.2: The substrate binding site of murine COX-2 (pdb 1CX2 [Kurumbail *et al.*(1996)Kurumbail, Stevens, Gierse, McDonald, Stegeman, Pak, Gildehaus, Miyashiro, Penning, Seibert, Isakson & Stallings]) with compound SC558 (celecoxib) bound. The structure is oriented to display as many of the residue-ligand interactions as possible. Much of the protein backbone has been removed for clarity.

The series of ligands evaluated in this study have the same scaffold as celecoxib in figure 7.2 and vary only at the position of the bromine atom (highlighted in purple). This common scaffold is displayed in figure 7.3 and the ten R group substitutions with corresponding experimental binding affinities is listed in table 7.1. Table 7.1 shows a larger range of binding affinities than the neuraminidase series of the previous chapter, from nanomolar to hundreds of micromolar. It is clear that placing polar hydrogen bonding groups at position R reduces binding affinity (ligands 4, 7 and 8) for COX-2, electron donating groups increase affinity (ligands 5 and 6) and larger groups also decrease affinity (ligand 3). This makes sense considering the nature of the environment in which the R group will reside, discussed above.

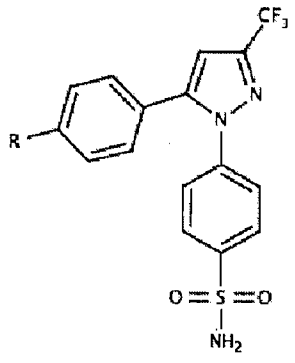


Figure 7.3: Common structure of the series of celecoxib analogues under study.

Ligand	R	IC_{50} (μM)
1	CH_3	0.040
3	CH_2CH_3	0.86
4	CH_2OH	93.3
5	SCH_3	0.009
6	OCH_3	0.008
7	CF_3	8.23
8	OH	>100
9	Cl	0.01
10	F	0.041
11	H	0.032

Table 7.1: Experimental activity of the celecoxib analogues against COX-2

7.3 The COX-2 system *in silico*

The structure of murine COX-2 complexed with celecoxib, 1CX2, was the starting point for the model system [Kurumbail *et al.*(1996)Kurumbail, Stevens, Gierse, McDonald, Stegeman, Pak, Gildehaus, Miyashiro, Penning, Seibert, Isakson & Stallings]. The 1CX2 structure had the advantage of hydrogen atoms being pre-assigned by the crystallographic study authors. The position of the sulphonamide group of each of the ligands in the binding site was changed as detailed by Price and Jorgansen and discussed in the *in vivo* section above [Price & Jorgensen(2000b)]. The heme of the second binding site of COX-2 was removed as it does not interact directly with the COX binding site under study and would have required specific parameterisation. Histidines were visually inspected in order to select an appropriate protonation state and His90, 95, 133, 204, 207, 214, 226, 232, 242, 278, 309, 320, 351, 356, 386, 388 and 417 were assigned δ-tautomer status with others being charged. The ethyl analogue, ligand 3 was

positioned in the COX-2 binding site on the basis of the binding mode of celecoxib in 1CX2. Other ligands with extra degrees of freedom in the substituted R group were modelled using observations from Price and Jorgensen [Price & Jorgensen(2000b)].

The COX-2 system was set up using the same protocol and forcefields as neuraminidase in the previous chapter. Thus, these details will not be reproduced here. Free energy perturbations were designed to close two thermodynamic cycles shown in figure 7.4 to allow the analysis of cycle hystereses and calculation of path averaged $\Delta\Delta F$s. TI and FG protocols used here are also the same as those in the previous chapter.

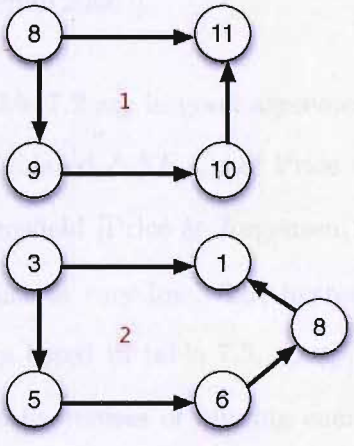


Figure 7.4: Closed perturbation pathways used to find hystereses for consistency checks. Ligand numbers are in black and closed cycle numbers are in red.

7.4 RETI Results

Table 7.2 shows RETI protein-ligand (ΔF_{prot}) and solvent-ligand (ΔF_{wat}) free energies which give the relative binding free energies ($\Delta\Delta F_{bind}$) which can be compared to the experimental (Exp) relative binding free energies. Also, vacuum free energies (ΔF_{vac}) are shown, which enable the calculation of solvation free energies ($\Delta\Delta F_{solv}$). All errors in table 7.2 were calculated using the block variance method described earlier in the main introduction.

Pert	Exp	$\Delta\Delta F_{bind}$	ΔF_{prot}	ΔF_{wat}	$\Delta\Delta F_{solv}$	ΔF_{vac}
1t3	1.82	2.24 ± 0.36	3.56 ± 0.27	1.32 ± 0.24	0.58 ± 0.26	0.74 ± 0.09
1t7	3.16	3.90 ± 0.25	20.71 ± 0.26	16.81 ± 0.23	0.39 ± 0.24	16.42 ± 0.07
3t5	-2.70	-1.75 ± 0.44	-3.19 ± 0.36	-1.49 ± 0.26	-0.46 ± 0.34	-0.50 ± 0.22
4t3	-2.78	-2.18 ± 0.45	0.8 ± 0.26	2.98 ± 0.37	6.39 ± 0.38	-3.91 ± 0.09
5t6	-0.07	-1.75 ± 0.52	-6.56 ± 0.38	-4.81 ± 0.36	-1.18 ± 0.64	-3.66 ± 0.53
8t1	< -4.64	-2.90 ± 0.37	15.03 ± 0.15	17.93 ± 0.34	4.39 ± 0.35	13.53 ± 0.08
8t6	< -5.59	-3.68 ± 0.75	9.35 ± 0.39	13.03 ± 0.64	4.03 ± 0.73	8.91 ± 0.035
8t9	< -5.46	-3.49 ± 0.31	17.12 ± 0.11	20.61 ± 0.29	5.20 ± 0.29	15.40 ± 0.02
10t9	-0.84	-1.33 ± 0.18	-0.19 ± 0.08	1.14 ± 0.16	-0.08 ± 0.16	1.22 ± 0.01
11t8	> 4.77	1.66 ± 0.29	-21.47 ± 0.10	-23.13 ± 0.27	-4.49 ± 0.27	-18.65 ± 0.02
11t10	0.15	0.01 ± 0.17	-3.47 ± 0.05	-3.48 ± 0.18	0.95 ± 0.16	-4.45 ± 0.02

Table 7.2: Experimental and calculated COX-2 relative binding free energies with the protein-ligand and solvent-ligand free energies used in their calculation. Also, relative solvation free energies calculated with the vacuum free energies used in their calculation. All free energies in this table are in kcal.mol^{-1}, were found with RETI and taken from the PhD thesis of J. Michel [Michel(2006)].

The calculated $\Delta\Delta F_{bind}$s of table 7.2 are in good agreement with the experimental values. They are also very close to the calculated $\Delta\Delta F_{bind}$s of Price and Jorgensen which were calculated with FEP and the OPLS forcefield [Price & Jorgensen(2000b)]. The MUE of $\Delta\Delta F_{bind}$s in table 7.2 is 1.25 kcal.mol^{-1} which is very low. The hystereses of the closed cycles of perturbation detailed in figure 7.4 are listed in table 7.3. Despite these two cycles containing 4 and 5 perturbations the respective hystereses of binding energies (hyst$_{bind}$s) of 0.43 and 0.16 kcal.mol^{-1} and hystereses of solvation energies (hyst$_{solv}$s) of 0.70 and 0.16 kcal.mol^{-1} are very low, suggesting all these RETI calculations are consistent and well converged.

Cycle pathway	hyst$_{bind}$	hyst$_{solv}$
1 (8,11,10,9)	0.43 ± 1.14	0.70 ± 1.12
2 (3,1,8,6,5)	0.51 ± 0.49	0.16 ± 0.27

Table 7.3: Hystereses of 2 thermodynamic cycles for relative binding (hyst$_{bind}$) and solvation (hyst$_{solv}$) free energies of the COX-2 ligands. All hyst$_{bind}$ and hyst$_{solv}$ values in kcal.mol^{-1}.

Table 7.4 shows the relative free energies of each ligand with respect to ligand 1. The MUE of $\Delta\Delta F_{bind}$s from table 7.4 is 0.76 kcal.mol^{-1}, which is extremely low. The data from table 7.4 are plotted in figure 7.5 which show the calculated and experimental results to be very similar. The r^2 of 0.85 and PI of 0.96 show this RETI analysis to be highly predictive.

It is interesting to compare these RETI results to those of Price and Jorgensen on the same system using FEP and the OPLS forcefield [Price & Jorgensen(2000b)]. Price and Jorgensen

produced results which agreed with experiment slightly better than these RETI results, with a r^2 of 0.96 and an MUE of 0.40 kcal.mol^{-1}. This improvement was achieved with significantly less computation than used here. Considering the variation in results seen in previous chapters, especially chapter 6, it is possible that this improvement is entirely fortuitous. However, it could also be rationalised by other factors. It is possible that the OPLS forcefield and CM1A atomic partial charges may hold some advantage over the AMER99 forefield and AM1/BCC atomic partial charges used here. Price and Jorgensen solvated there COX-2 system in such a way that water molecules were present in the binding site bridging Ser530 and Tyr385 and for some perturbations hydrogen bonded to Met522 [Price & Jorgensen(2000b)]. These waters were not present in the RETI simulation of Michel and could contribute to any differences [Michel *et al.*(2006)Michel, Verdonk & Essex].

Perturbation of 11t10 is particularly interesting as the water interacting with Met522 in the simulations of Price and Jorgensen also interacted with the substituted R group of the perturbed ligand. This was investigated by Price and Jorgensen, by running an extra simulation without this Met522 interacting water and found that the $\Delta\Delta F$ changed to 1.52 kcal.mol^{-1} from -0.15 kcal.mol^{-1} with the Met522 interacting water [Price & Jorgensen(2000b)]. As can be seen from table 7.2 the result of 0.01 kcal.mol^{-1} for perturbation 11t10 found by Michel is closer to the experimental value than both of the Price and Jorgensen results without the Met522 interacting water being present [Michel *et al.*(2006)Michel, Verdonk & Essex]. It is difficult to come to a conclusion over the probable *in vivo* occupancy of these binding site water molecules and the comparison of results with and without these waters needs the addition of a number of repeated calculations to check their reliability.

Compound	Perturbation pathway	Calc $\Delta\Delta F_{bind}$	Expl $\Delta\Delta F_{bind}$
6	[1t8+8t6];[1t3+3t5+5t6]	-0.99 ± 0.81	-0.95
5	[1t3+3t5]	0.54 ± 0.57	-0.88
9	[1t8+8t9]	-0.58 ± 0.48	-0.82
11	[1t8+8t9+9t10+10t11];[1t8+8t11]	0.99 ± 0.51	-0.13
1		0	0
10	[1t8+8t9+9t10];[1t8+8t11+11t10]	1.00 ± 0.51	0.01
3	1t3	2.25 ± 0.36	1.82
7	1t7	3.90 ± 0.25	3.15
4	[1t3+3t4]	4.42 ± 0.58	4.59
8	1t8	2.90 ± 0.37	4.63

Table 7.4: Experimental and RETI calculated binding free energies with respect to ligand 1. All free energies in this table are in kcal.mol^{-1}.

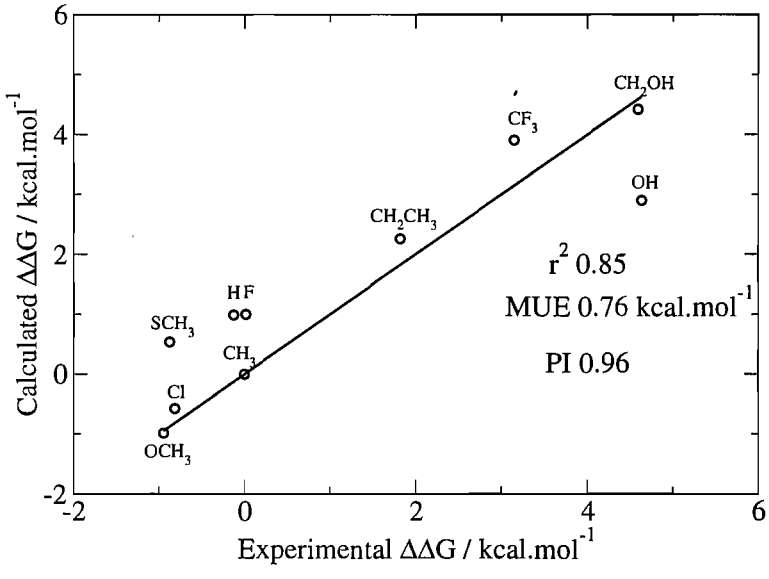


Figure 7.5: Comparison of calculated and experimental $\Delta\Delta F$s for 10 COX-2 ligands. Taken from J.Michel *et al.* 2006 [Michel *et al.*(2006)Michel, Verdonk & Essex].

7.5 FG Results

Table 7.5 shows protein-ligand (ΔF_{prot}) and solvent-ligand (ΔF_{wat}) free energies, found with the REFG2-BY10-BAR-800x1000x375 protocol of the previous chapter, which give the relative binding free energies ($\Delta\Delta F_{bind}$). Also, vacuum free energies (ΔF_{vac}) are shown, which enable the calculation of solvation free energies ($\Delta\Delta F_{solv}$). Errors in table 7.5 were calculated using the block variance method described earlier in the main introduction with 4 blocks.

Pert	Exp	$\Delta\Delta G_{bind}$	ΔG_{prot}	ΔG_{wat}	$\Delta\Delta G_{solv}$	ΔG_{vac}
1t3	1.82	2.48 ± 0.13	3.44 ± 0.08	0.96 ± 0.10	0.24 ± 0.10	0.72 ± 0.002
1t7	3.16	4.15 ± 0.08	20.70 ± 0.03	16.55 ± 0.07	0.20 ± 0.07	16.35 ± 0.003
3t5	-2.70	-3.04 ± 0.14	-3.06 ± 0.13	-0.01 ± 0.06	0.49 ± 0.06	-0.50 ± 0.003
4t3	-2.78	-4.03 ± 0.17	-0.35 ± 0.09	3.68 ± 0.15	7.59 ± 0.15	-3.91 ± 0.003
5t6	-0.07	-1.85 ± 0.14	-5.96 ± 0.08	-4.11 ± 0.12	-0.45 ± 0.12	-3.66 ± 0.008
8t1	< -4.64	-2.71 ± 0.13	14.98 ± 0.04	17.70 ± 0.12	4.17 ± 0.12	13.53 ± 0.003
8t6	< -5.59	-1.66 ± 0.25	9.12 ± 0.11	10.79 ± 0.22	1.88 ± 0.22	8.91 ± 0.005
8t9	< -5.46	-4.01 ± 0.17	16.84 ± 0.04	20.85 ± 0.17	5.45 ± 0.17	15.40 ± 0.001
10t9	-0.84	-2.05 ± 0.06	-0.60 ± 0.02	1.45 ± 0.06	0.23 ± 0.06	1.22 ± 0.0001
11t8	> 4.77	1.78 ± 0.12	-22.11 ± 0.03	-23.88 ± 0.12	-5.23 ± 0.12	-18.65 ± 0.0009
11t10	0.15	0.23 ± 0.07	-3.51 ± 0.01	-3.74 ± 0.07	0.71 ± 0.07	-4.45 ± 0.002

Table 7.5: Experimental and calculated relative binding free energies with the protein-ligand and solvent-ligand free energies used in their calculation. All free energies in this table are in kcal.mol^{-1} and were found with the REFG2-BY10-BAR-800x1000x375 protocol.

Again the calculated $\Delta\Delta F_{binds}$s found using the REFG2-BY10-BAR-800x1000x375 protocol are in good agreement with experiment (table 7.5). The MUE of these $\Delta\Delta F_{binds}$s is 1.51 kcal.mol^{-1} which is slightly more than the equivalent MUE for the RETI results above. The hystereses of the closed cycles of perturbations detailed in figure 7.4 are listed in table 7.6. These hystereses are not as low as those produced from the RETI analysis above since the hyst_{bind} and hyst_{solv} of cycle 1 are significantly larger than the RETI equivalent and the other REFG2-BY10-BAR-800x1000x375 hystereses. This suggests an error or possible lack convergence in the REFG2-BY10-BAR-800x1000x375 $\Delta\Delta F_{solv}$s in table 7.5. This indication of a lack of convergence conflicts with the indication from the forwards and backwards Jarzynski estimates of all of the perturbations in cycle 1, which are all very close, suggesting convergence. The average difference between forwards and backwards Jarzynski estimates from cycle 1 is 0.25 kcal.mol^{-1}.

Cycle pathway	hyst_{bind}	hyst_{solv}
1 (8,11,10,9)	0.53 ± 0.23	0.72 ± 0.16
2 (3,1,8,6,5)	3.46 ± 0.37	2.09 ± 0.30

Table 7.6: Hystereses of 2 thermodynamic cycles for relative binding (hyst_{bind}) and solvation (hyst_{solv}) free energies of the COX-2 ligands. All hyst_{bind} and hyst_{solv} values in kcal.mol^{-1}.

Table 7.5 lists the error, in relation to experimental values, of the calculated $\Delta\Delta F_{bind}$s for the RETI and REFG2-BY10-BAR-800x1000x375 analyses above (columns headed $\Delta\Delta F_{bind}$ RETI/REFG error) and the difference between RETI and REFG2-BY10-BAR-800x1000x375

ΔF_{wat} values (columns headed ΔF_{wat} RETI - REFG). In the column labelled R the groups listed before and after the "-" is present are at $\lambda = 0$ and $\lambda = 1$, respectively. The errors in $\Delta\Delta F_{bind}$ values relative to experiment are very similar for RETI and REFG2-BY10-BAR-800x1000x375, with the only really significant difference found in perturbation 8t6, where the error in REFG $\Delta\Delta F_{bind}$ compared to experiment is 2.02 kcal.mol^{-1} greater than the RETI error. This difference in perturbation 8t6 is almost entirely due to the difference in ΔF_{wat} values and causes the larger $hyst_{bind}$ $hyst_{solv}$ in table 7.6. The REFG2-BY10-BAR-800x1000x375 8t6 perturbation in water does seem to be well converged as forward and backward Jarzynski estimates are very close, being 10.17 and 10.09 kcal.mol^{-1} respectively. Thus, this difference between RETI and REFG calculations may either be caused by random sampling error or a methodological difference, which is not clear.

Pert	R	$\Delta\Delta F_{bind}$ RETI error	$\Delta\Delta F_{bind}$ REFG error	ΔF_{wat} RETI - REFG
1t3	Me-Et	0.42	0.66	-0.36
1t7	Me-CF_3	0.74	0.99	-0.26
3t5	Et-SCH-3	0.95	-0.34	1.48
4t3	CH_2OH-Et	0.60	-1.25	0.70
5t6	SCH_3-OCH_3	-1.68	-1.78	0.70
8t1	OH-Me	1.74	1.93	-0.23
8t6	OH-OCH_3	1.91	3.93	-2.24
8t9	OH-Cl	1.97	1.45	0.24
10t9	F-Cl	-0.49	-1.21	0.31
11t8	H-OH	-3.11	-2.99	-0.75
11t10	H-F	-0.14	0.08	-0.26

Table 7.7: Table showing the perturbations taking place at the R position from figure 7.3 and the error of calculated $\Delta\Delta F$s compared to experiment for RETI and REFG2-BY10-BAR-800x1000x375.

Table 7.8 shows the relative free energies of each ligand with respect to ligand 1 calculated with the REFG2-BY10-BAR-800x1000x375 protocol. The data from table 7.8 is plotted in figure 7.6 which show good agreement between calculated and experimental results. The r^2 of 0.76 and PI of 0.94 show this REFG2-BY10-BAR-800x1000x375 analysis to be slightly less predictive than the RETI analysis presented above. It is possible that this difference in predictive quality, which can be attributed to the overestimation of $\Delta\Delta F_{bind}$s for ligands 10 (F) and 4 (CH_2OH), is a random effect.

Compound	Perturbation pathway	Calc $\Delta\Delta F_{bind}$	Expl $\Delta\Delta F_{bind}$
6	[1t8+8t6];[1t3+3t5+5t6]	-0.68 ± 0.	-0.95
5	[1t3+3t5]	-0.56 ± 0.	-0.88
9	[1t8+8t9]	-1.30 ± 0.	-0.82
11	[1t8+8t9+9t10+10t11];[1t8+8t11]	0.72 ± 0.	-0.13
1		0	0
10	[1t8+8t9+9t10];[1t8+8t11+11t10]	2.50 ± 0.	0.01
3	1t3	2.48 ± 0.	1.82
7	1t7	4.14 ± 0.	3.15
4	[1t3+3t4]	6.50 ± 0.	4.59
8	1t8	2.70 ± 0.	4.63

Table 7.8: Experimental and REFG2-BY10-BAR-800x1000x375 calculated binding free energies with respect to ligand 1. All free energies in this table are in kcal.mol^{-1}.

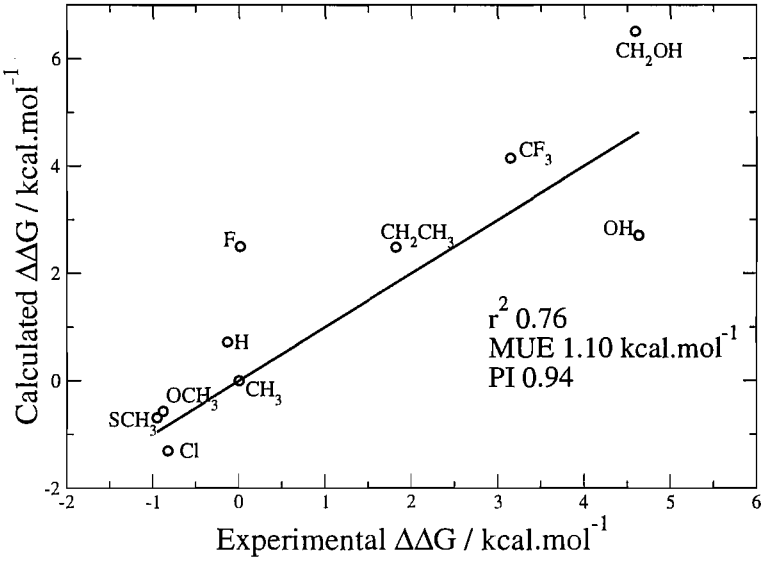


Figure 7.6: Comparison of REFG2-BY10-BAR-800x1000x375 calculated and experimental $\Delta\Delta F$s relative to ligand 1 for 10 COX-2 ligands.

Protocol REFG2-BY10-BAR-800x1000x1125 described in the previous chapter was applied to investigate any predictive improvements possible by increasing the computational time given to the calculation. Table 7.9 shows ΔF_{prot} and ΔF_{wat} free energies, found with the REFG2-BY10-BAR-800x1000x1125 protocol, which give the relative binding free energies ($\Delta\Delta F_{bind}$).

Pert	Exp	$\Delta\Delta G_{bind}$	$\Delta\Delta G_{prot}$	ΔG_{wat}
1t3	1.82	2.13 ± 0.13	3.13 ± 0.09	1.00 ± 0.10
1t7	3.16	4.10 ± 0.25	20.66 ± 0.24	16.56 ± 0.06
3t5	-2.7	-3.15 ± 0.13	-3.64 ± 0.12	-0.49 ± 0.06
4t3	-2.78	-3.28 ± 0.13	0.29 ± 0.09	3.56 ± 0.10
5t6	-0.07	-1.69 ± 0.13	-6.15 ± 0.08	-4.46 ± 0.10
8t1	< -4.64	-3.28 ± 0.12	14.73 ± 0.04	18.01 ± 0.11
8t6	< -5.59	-3.18 ± 0.19	9.54 ± 0.11	12.71 ± 0.16
8t9	< -5.46	-4.14 ± 0.11	16.75 ± 0.03	20.90 ± 0.11
10t9	-0.84	-1.75 ± 0.06	-0.53 ± 0.02	1.22 ± 0.06
11t8	> 4.77	0.95 ± 0.11	-21.98 ± 0.03	-22.94 ± 0.11
11t10	0.15	0.16 ± 0.05	-3.48 ± 0.01	-3.65 ± 0.05

Table 7.9: Experimental and calculated relative binding free energies with the protein-ligand and solvent-ligand free energies used in their calculation. All free energies in this table are in kcal.mol^{-1} and were found with the REFG2-BY10-BAR-800x1000x1125 protocol.

The MUE of the $\Delta\Delta F_{bind}$s in table 7.9 is 1.24 kcal.mol^{-1} which is just lower than the equivalent MUE for the RETI results above. The major difference between the RETI and REFG2-BY10-BAR-800x1000x375 analyses above was perturbation 8t6. Perturbation 8t6 in table 7.9 is more similar to the RETI analysis with the difference being 0.5 kcal.mol^{-1}. The hystereses of the closed cycles of perturbations detailed in figure 7.4 are listed in table 7.10. These hystereses are still higher than the RETI analysis above while cycle 1 is lower than the REFG2-BY10-BAR-800x1000x375 analysis above and cycle 2 is higher.

Cycle pathway	hyst$_{bind}$
1 (3,1,8,6,5)	-2.81
2 (8,11,10,9)	-1.24

Table 7.10: Hystereses of 2 thermodynamic cycles for relative binding (hyst$_{bind}$) and solvation (hyst$_{solv}$) free energies of the COX-2 ligands. All hyst$_{bind}$ and hyst$_{solv}$ values in kcal.mol^{-1}.

Table 7.11 shows the relative free energies of each ligand with respect to ligand 1 calculated with the REFG2-BY10-BAR-800x1000x1125 protocol. Figure 7.11 compares REFG2-BY10-BAR-800x1000x1125 calculated with experimental relative binding free energies. This more costly REFG2-BY10-BAR-800x1000x1125 analysis has a MUE of $\Delta\Delta F_{bind}$s relative to ligand 1 of 0.89 kcal.mol^{-1}, an r^2 of 0.79 and a PI of 0.94 which is a slight improvement over the REFG2-BY10-BAR-800x1000x375 analysis above.

Compound	Perturbation pathway	Calc $\Delta\Delta F_{bind}$	Expl $\Delta\Delta F_{bind}$
6	[1t8+8t6];[1t3+3t5+5t6]	-1.31 ± 0.	-0.95
5	[1t3+3t5]	-1.02 ± 0.	-0.88
9	[1t8+8t9]	-0.86 ± 0.	-0.82
11	[1t8+8t9+9t10+10t11];[1t8+8t11]	1.53 ± 0.	-0.13
1		0	0
10	[1t8+8t9+9t10];[1t8+8t11+11t10]	2.48 ± 0.	0.01
3	1t3	2.12 ± 0.	1.82
7	1t7	4.10 ± 0.	3.15
4	[1t3+3t4]	5.40 ± 0.	4.59
8	1t8	3.23 ± 0.	4.63

Table 7.11: Experimental and REFG2-BY10-BAR-800x1000x375 calculated binding free energies with respect to ligand 1.

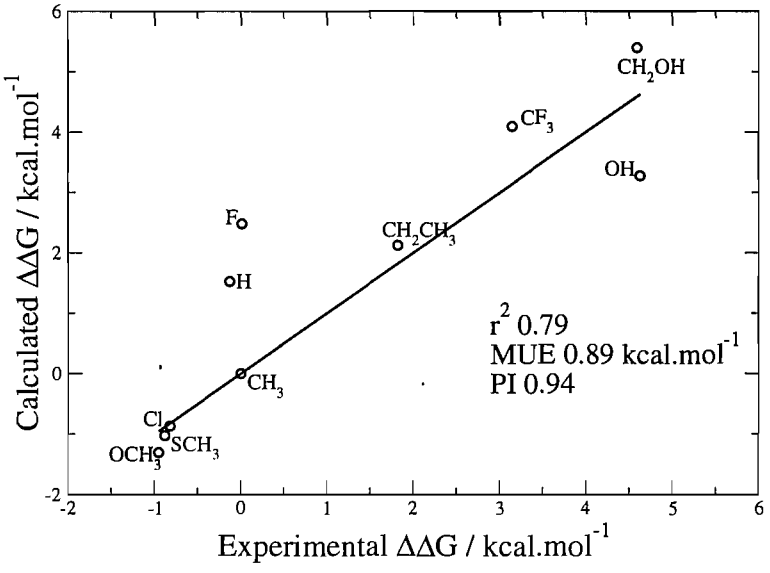


Figure 7.7: Comparison of REFG2-BY10-BAR-800x1000x1125 calculated and experimental $\Delta\Delta F$s relative to ligand 1 for 10 COX-2 ligands.

It is important to study the level of predictability throughout the analyses presented above as this may fluctuate to some extent. Figures 7.8 and 7.9 shows the fluctuations in PI and MUE through the calculations of the COX-2 analyses of this chapter. The number of MC trials of one perturbation leg (protein or water) from figure 7.4 are on the x-axis, as if all perturbation legs are run in parallel. In the legend of figures 7.8 and 7.9 REFG-375 and REFG-1125 refers to the REFG2-BY10-BAR-800x1000x375 and REFG2-BY10-BAR-800x1000x1125 protocols respectively. Also, REFG-375-pick and REFG-1125-pick refers to the same protocols where the FG estimator to be used with each perturbation at each data point is chosen according to the rules stated in previous chapters—: If the BAR estimate is between the forwards and backwards Jarzynski estimates, the BAR estimate is chosen. Otherwise the Jarzynski estimate with the

largest relative entropy measure as defined by Wu and Kofke [Wu & Kofke(2005a)] is used.

Figure 7.8 shows that the RETI analysis converges to a high PI score more quickly than the REFG analyses. Also, the REFG analyses where estimators are picked seem to offer slightly higher PI scores than the BAR-only and RETI analyses, but seem to fluctuate. Figure 7.9 shows that the RETI analysis has a lower MUE than the REFG analyses at all points of the calculations. The REFG analyses with picked estimators offer an slight improvement in MUE for the REFG2-BY10-BAR-800x1000x375 protocol but almost no improvement for the REFG2-BY10-BAR-800x1000x1125 protocol. This lack of improvement in the REFG2-BY10-BAR-800x1000x1125 protocol suggests this set of calculations are more fully converged.

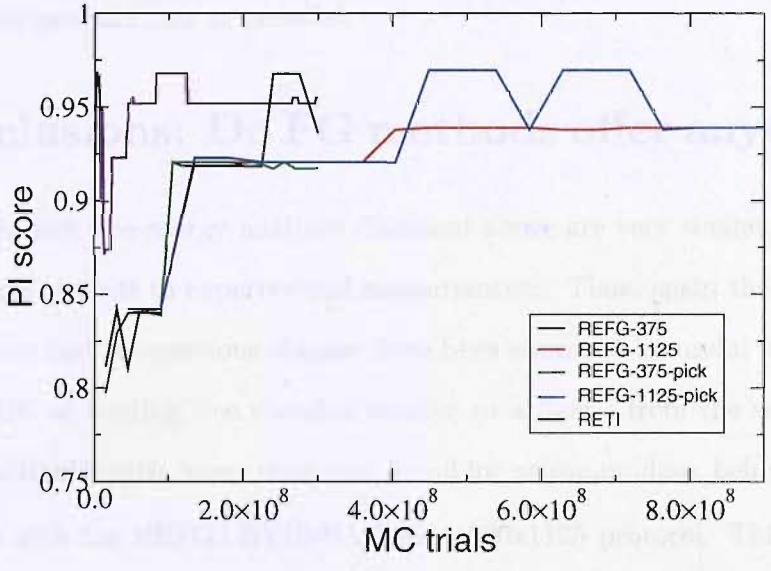


Figure 7.8: Comparison of PI scores for COX-2 analyses. The number of MC trials on the x-axis represents the number of MC trials from one perturbation leg i.e. all protein and water simulations are run in parallel.

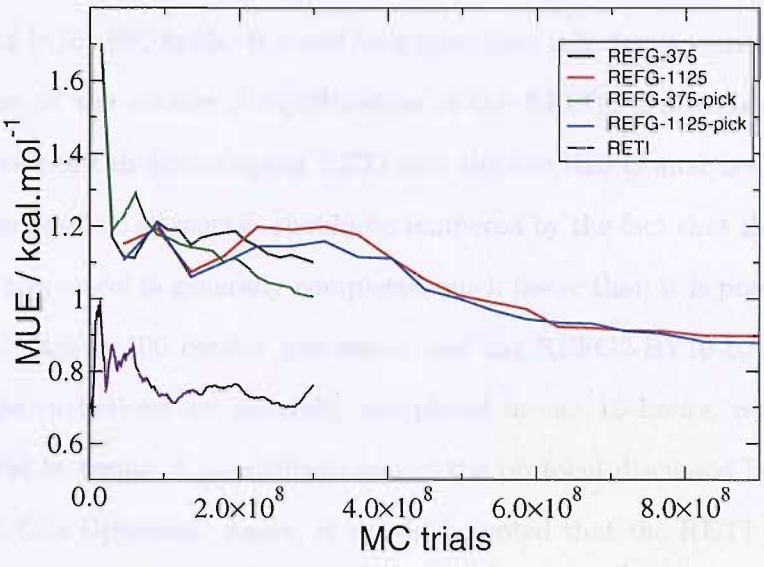


Figure 7.9: Comparison of the MUE for COX-2 analyses. The number of MC trials on the x-axis represents the number of MC trials from one perturbation leg i.e. all protein and water simulations are run in parallel.

7.6 Conclusions: Do FG methods offer anything new?

All the COX-2 binding free energy analyses discussed above are very similar and of good predictive quality with regards to experimental measurements. Thus, again the REFG protocols used in this chapter and the previous chapter have been shown to be useful in the drug design process. The MUE of binding free energies relative to a ligand from the series being evaluated for COX-2 is significantly lower than that found for neuraminidase, being 1.25 versus 3.48 kcal.mol^{-1} found with the REFG2-BY10-BAR-800x1000x1125 protocol. This difference is almost certainly related to the difference in the complexity of the individual perturbations with some neuraminidase perturbations growing large phenyl groups.

The only real disparity between the RETI and REFG analyses was the relatively large difference (2.02 kcal.mol$^-$1) in 8t6 perturbations in the water environment. This disparity was not present in the extra REFG2-BY10-BAR-800x1000x1125 analysis. It is not clear whether this improvement is found because of the extra sampling of the REFG2-BY10-BAR-800x1000x1125 analysis or it is just a random effect. The RETI and REFG COX-2 analyses need to be repeated a number of times to clarify any possible differences.

The RETI protocol discussed here seemed to converge to a high PI score and low MUE using fewer MC trials than the REFG protocols (figure 7.8). RETI converges to a PI score above 0.9 at around 2.46x10^7 MC trials while REFG2-BY10-BAR-800x1000x375 reaches a PI

above 0.9 at around $1x10^8$ MC trials. It could be argued that this slower convergence of REFG2 protocols is because of the smaller λ equilibration of the REFG2 protocols compared to the RETI protocol. However this advantage of RETI over the two REFG analyses could easily be a random effect. This possible advantage should be tempered by the fact that the REFG2-BY10-BAR-800x1000x375 protocol is generally completed much faster than it is possible to complete the RETI protocol. Using 100 condor processors and the REFG2-BY10-BAR-800x1000x375 protocol, COX-2 perturbations are generally completed in ca. 10 hours, whereas RETI calculations are limited in terms of paralellisation and the protocol discussed here takes around 24 hours on 12 2.2 Ghz Opterons. Again, it should be noted that the RETI protocol used 10 million MC trials for λ window equilibration while REFG protocols used 2 million MC trials for the same equilibrations.

Chapter 8

Discussions and Conclusions

8.1 The findings of this study

This study set out with the aim of investigating the use of FG techniques and finding methods which are of use in calculating accurate protein-ligand binding energies for the drug design process. Jarzynski's original FG study triggered the start of an extraordinary torrent of different and powerful FG associated methods [Jarzynski(1997b)]. During the course of the investigation many new methods appeared as work proceeded and the selection of a subset appropriate for our goal of using large scale distributed computing was essential.

Here the results of the investigations into the methods chosen to be investigated are discussed. The implications of these results to the applicability of free energy calculations to rational drug design will then be discussed as a conclusion to this work.

8.1.1 Harmonic oscillator study

The Rosenbluth FG methods of Wu and Kofke were found to offer advantages over original FG sampling only in specific cases and with specific switching protocols. When the possibility of all switching protocols are taken into account there is little efficiency advantage to be found by using these Rosenbluth methods over original FG. In most cases original FG sampling with switching protocols of many small λ increments was found to perform best. However, λ bias may have potential if it can be developed further. This development must be such that the final λ increment is not forced to unity (not constrained by the number of λ increments) and the size of λ increments is made proportional to the suitability of the present configuration for the increment, regardless of the size of previous increments (see figure 4.13). Configuration bias re-

quires large amounts of sampling between λ increments in order that a configuration favourable for the λ increment may be found. This necessity can make it computationally costly.

Overall, testing with harmonic oscillator systems suggested that BAR is the most efficient FG estimator. Certain cases where found where the Jarzynski estimator was found to be more accurate. Specifically, in the case of a slight subset relation, the Jarzynski estimator using switches starting with the system of larger important phase space to smaller phase space will be more accurate (figure 4.14 d)). This deficiency of BAR originates in the relatively poor behaviour of switches starting with the system of smaller important phase space due to the inaccessibility of the destinations phase space from the starting system. This is a view supported by the work of Shirts *et al.* [Shirts & Pande(2005)].

Owing to the possibility of different FG estimators being more accurate in different situations, it is necessary to find a method to pick between estimators without *a priori* knowledge of the free energy. An estimator picking scheme was investigated where BAR is chosen when its result lies between that of the forwards and backwards Jarzynski estimates, otherwise a Jarzynski estimate is chosen using either the relative entropy or bias measures of Wu and Kofke (2005c). Although seemingly arbitrary, these estimator rules are reasonably successful at predicting the most accurate estimate. Alternatively BAR could be used in every situation, as in the cases where BAR is not the most accurate it is generally close, and the loss of accuracy would be negligible within the complex calculations involved in finding protein-ligand binding free energies.

Harmonic oscillator test systems were shown to be more tractable when the λ coordinate is split into a number of smaller sections which are evaluated using FG separately. This technique is used routinely with FEP calculations. Also, as shown in various previous studies, in general fewer longer switches were found to be more efficient than more shorter switches [Gore *et al.*(2003)Gore, Ritort & Bustamante], [Ytreburg & Zuckerman(2004)].

8.1.2 Solute-solvent test systems

The methods found to be most applicable for the study of harmonic oscillators were then applied to two simple solute-solvent test systems.

The study of the hydration free energy due to sodium charging

Here the FG estimators used in this study were further validated successfully. Specifically for this test system, the FD estimator performs with significantly greater efficiency than the Jarzynski estimator due to the linear response nature of the free energy change. Also, the symmetric estimators and BAR perform well for this system.

The configuration bias-d method displayed no efficiency improvement for extra computational expense while the FG-BY10 protocol was shown to improve the efficiency of the Jarzynski estimators such that the performance of all estimators is very similar.

The study of the relative hydration free energy of water and methane

This study was designed to investigate and overcome the problems found in calculating free energy differences for systems with large solvent conformational changes. It can be difficult to obtain consistently accurate free energy differences for these important systems due to random sampling error.

REFG2 was shown to improve the reproducibility of accurate results compared to non-RE FG methods. REFG2 also displays comparable consistent accuracy to RETI [Woods *et al.*(2003a)Woods, Essex & King]. However, when the number of MC trials used for a calculation such as this is limited, RETI should be the method of choice as convergence is generally found more quickly and easily. It may be important with REFG calculations that equilibrium seed sampling is maximised possibly at the expense of switch sampling. This was highlighted by the difference in performance of REFG1 and REFG2-BY10 calculations figure 5.7. Also, there may be need for an investigation of any possible differences between the performance of RETI-random and RETI as some difference in precision has been uncovered by Woods *et al.* (2003), (2007). An interesting study of this issue has found a method of choosing pairs of replicas for a λ swap which are not adjacent which may be of use for methods such as RETI and REFG [Brenner *et al.*(2007)Brenner, Sweet, VonHandorf & Izaguirre].

Another issue with these FG calculations is finding the appropriate estimator for optimum accuracy. A number of methods of picking between forwards and backwards Jarzysnki estimates and BAR were investigated. All of the estimator picking methods gave similar results. However, the method which gave the most consistent and accurate results was the totalled

relative entropy measure. This method of picking uses BAR when a BAR estimate is placed between the forwards and backwards Jarzynski estimates; when this is not the case the Jarzynski estimate with the highest total $\bar{W}_{diss}$ summed over all FG calculations across λ is chosen. The use of BAR alone is also relatively accurate. Hence, these two estimator picking methods were used in the subsequent protein-ligand studies.

Here it has been shown that REFG methods have the possibility of similar levels of accuracy to RETI when a limit is placed on the number of MC trials used in a calculation. It is difficult to compare the speed of RETI and REFG calculations as these RETI calculations were carried out on a small number of dedicated, fast, processors while REFG calculations were carried out on a large number (100) of condor nodes (see appendix D). However, it is clear that REFG methods may be able to run calculations faster than RETI when sufficient computational resources are available.

8.1.3 Study of the binding free energy of a set of inhibitors for Neuraminidase

The neuraminidase analysis of Michel *et al.* (2006) was shown to be relatively demanding and is in some ways, beyond the limits of the present methodology.

The ligands of the group under study which contain an amino group at position Rpol (see figure 6.3) utilise a mediating water to help binding. This mediating water is not used by the ligands with a guanadino group at position Rpol. Thus, in perturbing from an amino ligand to a guanadino ligand the mediating water must be displaced from its buried position in the binding site, something beyond the present methodology. Consequently a free energy from a more complex study of a similar system was used for the perturbation of ligand 3 to ligand 6 [Barillari *et al.*(2006)Barillari, Taylor, Viner & Essex]. This problem highlights one of a number of limitations of using free energy calculations in drug design.

This analysis contained a number of perturbations (namely perturbations of ligands 2 to 9, 4 to 10, 7 to 8 and 9 to 10) which include the growth of large chemical groups where much smaller groups were originally. Owing to the relatively large numbers of atoms affected by these perturbations they can be difficult to converge. It has been shown that the number of switches required to obtain converged free energies is proportional to the number of atoms involved in

the work process [Kofke(2006)].

Owing to the complexity of some of the perturbations in this neuraminidase analysis, it is probable that the converged calculated binding free energies differ significantly from the experimental values. Although the exact nature of these disparities is not clear they are possibly due to deficiencies of the model.

The FG-BY10 protocol was shown to be unreliable in evaluating this set of inhibitors for neuraminidase. PI scores ranged from a very low 0.44 to very high 0.95 and r^2s from 0.47 to 0.88. This suggests that FG-BY10 is unable to produce converged results in within the number of MC configurations used due to the challenging nature of some of the perturbations of this analysis.

REFG2 was shown to compare favourably with RETI in producing results of a repeatedly high predictive standard. PI scores ranged from 0.91 to 0.97 and r^2s from 0.88 to 0.94 with the RETI analysis producing a PI of 0.93 and r^2 of 0.81. These results suggest that RE free energy methods should always be used for protein-ligand binding free energy calculations and that the REFG2 protocol is at least as good as RETI for this purpose.

One advantage of FG methods was highlighted as large inaccuracies were found in some perturbations of the original RETI analysis of Michel *et al.* (2006). A large peak in the PMF of FG perturbations 2 to 9, 4 to 10 and 7 to 8 in both the water and protein environments were not present in the RETI equivalents (figure 6.9). This sudden change in the free energy originates from very large intra-molecular LJ and Coulombic forcefield energies produced by the close proximity of the atoms of the phenyl ring of the cis group at the early stages of these perturbations. These phenyl ring atoms start as dummy atoms but as their non-bonded terms are switched on these atoms are too close and a large energy is produced. This large change in the free energy was not detected by the RETI analysis as it came between the discrete simulation points across λ. Also the large errors in the water and protein leg perturbations was not discovered in the original RETI analysis as this large energy artefact is a very similar size in both water and protein legs resulting in its effective cancelation. This example may serve to highlight the exploratory use of FG methods in analyses such as this or the need for extreme care in the use of TI based methods.

This study also highlighted the limitations of statistical error bars in predicting the reproducibility of free energy calculations on large complex systems. In future it may be important

to routinely reproduce calculations to be sure of the results they give.

The totalled relative entropy measure estimator picking rules were used in this study and compared to using BAR alone. This method of FG estimator picking was shown to give very similar results to using BAR alone, mainly due to the fact that BAR estimates were generally placed between forwards and backwards Jarzynski estimates for these neuraminidase perturbations.

The major result of this neuraminidase study is that REFG has shown significant improvements in calculation wall clock time compared to the established RETI method while maintaining high levels of accuracy. REFG calculations were completed in around 10 hours using 100 condor nodes while RETI calculations took around 24 hours on 12 2.2 Ghz Otperons.

8.1.4 Study of the binding free energy of a set of inhibitors for COX-2

This set of COX-2 perturbations were in general much less demanding than the neuraminidase analysis discussed above. These COX-2 perturbations involve changes to a single group which often consists of one atom only. This means that these COX-2 calculations require a smaller number of MC trials to give a converged free energy estimate. Indeed a similar study of the same system by Price and Jorgensen (2000a) used FEP calculations with around one third of the MC trials of the protocols used in this study and found comparable if not better results.

This COX-2 analysis again demonstrated that REFG2 is able to produce results of a similar predictive quality to RETI. However, for this COX-2 analysis RETI seems able produce converged and accurate results using fewer MC trials than REFG2 (see figures 7.8 and 7.9). The seeming efficiency advantage of RETI may be tempered by the possibility of running MC trials faster using FG methods and many computers in parallel.

Again, the totalled relative entropy measure estimator picking rules were used and gave very similar results to BAR alone. Also a wall clock time advantage of REFG over RETI similar to that found for the neuraminidase study was found for these COX-2 calculations.

8.2 FG Best practice

As can be seen in the FG background chapter (chapter 3) there have been many interesting developments in the area of nonequilibrium work free energy methods over recent years. Here we have picked some particularly interesting methods for investigation with the idea of large scale parallelisation of calculations in mind. Of the methods chosen for investigation some have been found to be more suitable than others for use in rational drug design.

It is clear from the results of this work that overall the BAR estimator is the most efficient. This estimator is able to produce results of very good accuracy in all cases studied here. This is a view shared by other studies in the literature [Shirts & Pande(2005)], [Ytreburg *et al.*(2006)Ytreburg, Swendsen & Zuckerman]. There are cases were the Jarzynski estimator may outperform BAR but differences in accuracy are small. For the case of demanding protein-ligand binding studies BAR is a suitable choice.

Although seeming to have potential, the Rosenbluth sampling FG methods of Wu and Kofke were found, in general, to be less efficient than original FG [Wu & Kofke(2005c)]. REFG protocols were found to perform best for the relative hydration free energy of water and methane and binding free energy studies of neuraminidase and COX-2 systems. REFG was able to show similar levels of accuracy to the established RETI method with the possibility of wall clock time saving through the use of large distributed computational resources.

A series of FG bias calculations methods have been discussed and investigated in this work [Gore *et al.*(2003)Gore, Ritort & Bustamante], [Wu & Kofke(2005a)], [Jarzynski(2006)]. In general these methods rely on a Gaussian distribution of work values, something which is not always possible especially for the large protein-ligand systems of interest here. Hence, these method were found to be less reliable. It may be that in future more generally applicable and reliable methods are developed.

8.3 Free energy calculations in rational drug design

The central purpose of this work was to investigate a relatively new set of free energy methods to test whether they can help further the applications of free energy methods in rational drug design. As discussed earlier the reasons for the as yet low levels of use of these methods by

the pharmaceutical industry are as follows: free energy calculations on protein-ligand systems, although significantly faster than previously, are still relatively slow to carry out and computationally demanding. For use in rational drug design it should be possible to run analyses of around 10 drug targets overnight. These calculations are also limited in applicability as the sets of inhibitors under study must be very similar with low levels of system conformational change (requiring single topology methods). Dual topology methods may be able to speed up the process of system set-up and increase the range of applicable systems. However, dual topology methods are more computationally demanding again. Thus, large increases in efficiency of free energy calculations may be able to increase the usability of these methods in rational drug design.

Here it has been shown that FG based methods are able to produce protein-ligand binding free energy results of a similar quality to the established RETI method in significantly shorter time through the use of a large distributed computer. Here REFG calculations were able to use 3 times more MC trials than RETI in around the same amount of wall clock time. However, this sampling speed improvement did not offer any improvements in predictive accuracy. Further, dual topology calculations may need many 10s of times more sampling than the single topology calculations performed here. Hence extremely large computational resources would be needed to make these sorts of calculations viable in the time frames needed. This study has also highlighted the advantage of FG based methods in being able to find unexpected occurrences in a perturbation that would be missed by TI based methods. In this way FG methods are more reliable.

As more research into the use of free energy calculations in protein-ligand binding studies is carried out, the limitations of these methods will become clearer. This should lead to the gradual increase in use of these methods in drug design. Also, as the power of computational resources continues to increase, the reliability and applicability of these methods will become better.

Appendix A

Markov chains

A Markov chain, named after Andrey Markov, is a sequence of discrete random variables or states where production of the next state (N) is independent of all states except the most recent (O) (known as the Markov property). The changes in state are called transitions, and brought about by application of a particular Markov process or a trial. Hence, the transition probability $\pi(O \rightarrow N)$ can be interpreted as the conditional probability,

$$\pi(O \rightarrow N) = P(X_n = N | X_{n-1} = O), \tag{A.1}$$

where X_n is the present state and X_{n-1} is the immediately previous state. When at O the probabilities of arriving at all possible N states make up a row of a matrix called the transition matrix ($\prod$) the sum of each row must be equal to one. Also, $\pi(O \rightarrow N)$ must be ergodic as discussed later.

This definition of a Markov process has a special property. A clear way to show this property starts with consideration of the probability $P(X_n = N)$ that at step n (the present step) the system is in state N:

$$P(X_n = N) = P(X_n = N | X_{n-1} = O) P(X_{n-1} = O) = \pi(O \rightarrow N) P(X_{n-1} = O), \tag{A.2}$$

Therefore the master equation, which considers the change in this probability over repeated realisations of the Markov process (change in n, where n is continuous rather than discrete ie. $P(X_n = N) = P(N, n)$ can be,

$$\frac{dP(N,n)}{dn} = \sum_{(O \rightarrow N)} \pi(O \rightarrow N) P(N,n) + \sum_{(N \rightarrow O)} \pi(N \rightarrow O) P(O,n). \tag{A.3}$$

This identifies the evolution of the system as a continuous-time Markov process [Landau(2000)]. As long as the two terms on the right hand side of A.3 are kept equal the probability of a state being conserved is,

$$\frac{dP(N,n)}{dn} \equiv 0. \tag{A.4}$$

Also, owing to the Markov property, the state at step n in the process will completely determine the future evolution. Thus, property of such a Markov process is that through many applications of $\prod$ it will converge to a unique limiting distribution regardless of the state in which it begins. The resulting expression of equation A.3 is known as "detailed balance". To satisfy detailed balance the probability of leaving O for N must be the same as the probability of leaving N for O ($\pi(O \rightarrow N) = \pi(O \rightarrow N)$).

It is important to note that not all systems have properties compatible with equation A.3. For this to be the case it is important that the Markov chain produced is ergodic. To be ergodic a Markov chain must be:

aperiodic: A state O has a period J if any return to state O must/may occur in a multiple of J applications of the process. For example, if it is only possible to return to state O in an even number of steps then O is periodic with a period (J) of 2. If J is 1 the state O is aperiodic and the process can return to O without a specific period. A Markov chain is aperiodic if all its states are aperiodic. To be aperiodic a state must also be irreducible, able to reach any point in configuration space from any other in a finite number of transitions.

positive recurrent: A Markov chain is positive recurrent if starting in each state the chain can return to this state in a finite number of applications of the process. If the probability the chain will never return to a given state is non-zero that state is said to be transient.

Another important note is that it is possible to produce a conserved limiting distribution without detailed balance, where equation A.4 is satisfied with the terms on the right hand side of equation A.3 not being equal. However, this is extremely difficult and would need knowledge of $\prod$, which is not easily obtainable.

Appendix B

Monte Carlo moves

B.1 Basic moves

The basic Metropolis MC method entails the construction of a symmetric transition matrix through finding a trial configuration N at random. As discussed before this is generally achieved through the localised perturbation of the present configuration O according to preset parameters. This has the advantages that if configuration O is a member of the Boltzmann limiting distribution then N has a good chance of being a member too and that new total energy evaluations may not be necessary. In the interests of keeping the acceptance rate of new configurations high and the range of possible moves as big as possible, the size of this perturbation is adjusted to an appropriate level which is kept constant, as changing it may violate detailed balance.

With these ideas in mind, in the case of condensed phase simulations, it can be shown that it is generally most efficient to perturb only one particle at a time [Frenkel & Smit(1996)]. On average, a trial move will be rejected if the change in the potential is more than k_BT. Thus, if more than one particle is moved at a time, the time taken to evaluate the new potential is larger and the move is likely to be rejected.

The range in size of any translational move of particles (Δ) is a parameter which is generally set through trial and error. Δ is set such that the acceptance rate of a simulation is a target figure which according to the literature is around 50% [Frenkel & Smit(1996)]. When considering this problem, it may be worth considering that the speed with which a MC code can perform trials has a large effect on possible target acceptance rates. If displacement per CPU time is the term to be optimised, then the faster the code the lower the possible acceptance rate target. For example for MC on hard core systems where a moves can be rejected much faster (as soon as an overlap is detected), optimum acceptance rates may be as low as 20 %.

A size range parameter for rotational movement of particles (θ) is also necessary if particles are molecules. The optimum θ can be found with the same considerations as for translational moves above.

In the case of non-rigid molecules then moves of the internal degrees of freedom must also be undertaken. The range of size parameter for moves of the internal coordinates of each atom can be individual to that atom depending on its environment, with similar considerations to those discussed above for translational moves.

B.2 Biased moves

When choosing a sampling method for molecular simulations, perhaps the majority of investigators employ MD. All other efficiency issues being relatively equal, MD is able to offer information on the time linked dynamic behaviour of a system. However, in specific cases, where MD gets trapped in local energy minima and is too slow to allow convergence of calculations in feasible time scales, specifically biased MC methods can offer important advantages.

Biased MC methods bias the normally random production of new configurations (N) such that they have an increased probability of being accepted into the average being accumulated. This can be an extreme time saving strategy in some cases. Clear examples of this possibility are ensembles where molecules must be inserted as part of a move and the probability of a random move being accepted is very small. Slightly less clear maybe are cases of protein simulations which are directly related to this study. There are however good examples of MC methods which can bias protein back-bone moves and increase the speed of protein sampling, which is notoriously slow and difficult due to large energy barriers between possible conformations [Ulmschneider & Jorgensen(2003)]. Also, a bias may be introduced relatively easily to increase the number of moves performed on molecules close to a point of interest (preferential sampling) [Owicki(1977)]; the most relevant example is biasing moves for water molecules close to a protein binding site.

The basis of biased MC lies in introducing bias into the transition matrix, $\prod$, described earlier and then correcting for this bias in the acceptance test probabilities. If a method of producing state N conducive to a high acceptance rate is found, which uses a function of the potential of the current configuration, the transition probabilities will be:

$$\alpha(O \rightarrow N) = f(U_N), \quad \alpha(N \rightarrow O) = f(U_O). \tag{B.1}$$

In order that detailed balance and the Boltzmann distribution be preserved, these same probabilities must be introduced into the acceptance test:

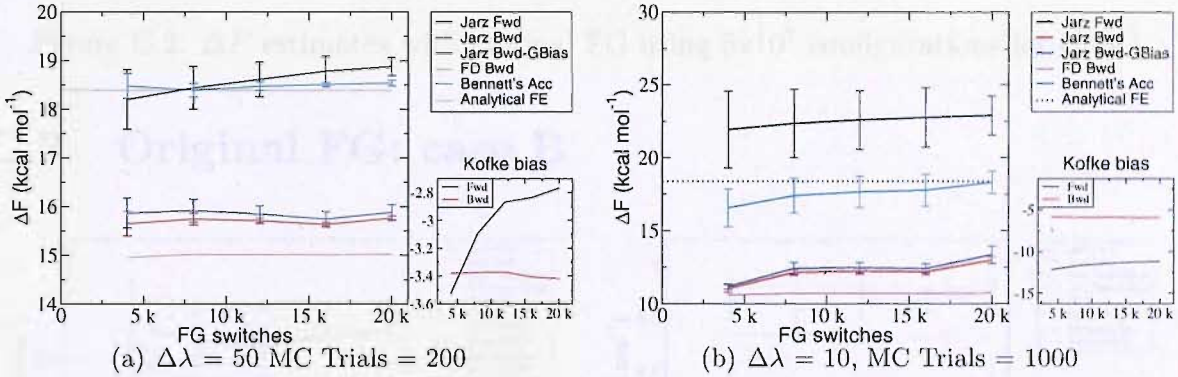
$$P_{accN} = min(1, \frac{f(U_N)}{f(U_O)} exp\{-\beta[U_O - U_N]\}). \tag{B.2}$$

Thus, it is necessary to know the probabilities of $\alpha(O \rightarrow N)$ and $\alpha(N \rightarrow O) = f(U_O)$ which may not always be easily found.

Appendix C

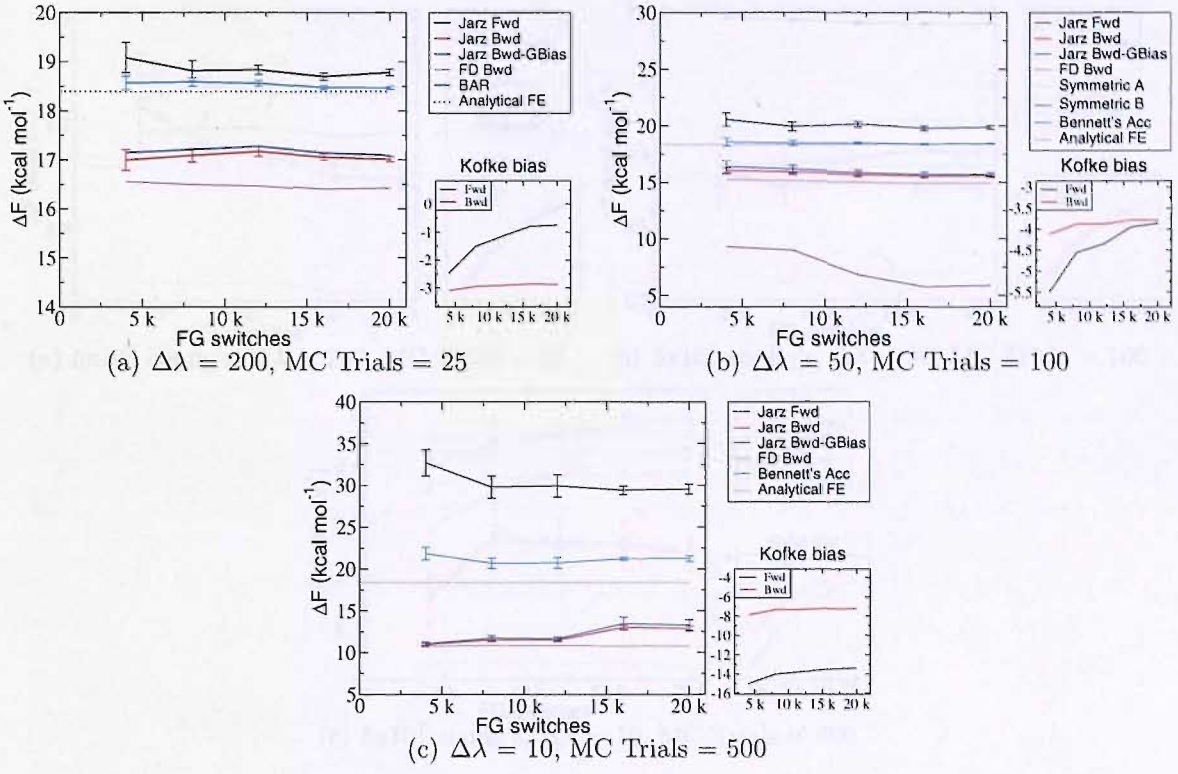
Extra results of harmonic oscillator study

C.1 Original FG: case A



(a) $\Delta\lambda = 50$ MC Trials = 200 (b) $\Delta\lambda = 10$, MC Trials = 1000

Figure C.1: ΔF estimates with original FG using 1x10^8 configurations for CaseA



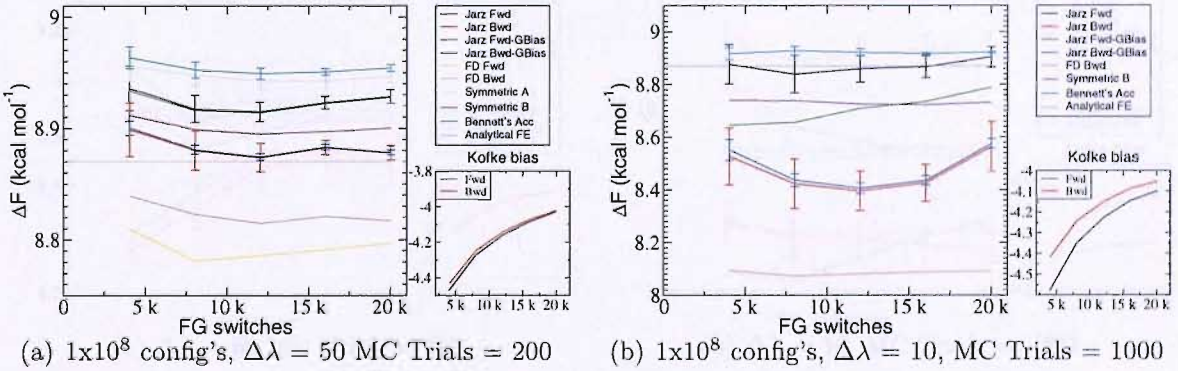
(a) $\Delta\lambda = 200$, MC Trials = 25

(b) $\Delta\lambda = 50$, MC Trials = 100

(c) $\Delta\lambda = 10$, MC Trials = 500

Figure C.2: ΔF estimates with original FG using 5x10^7 configurations for CaseA

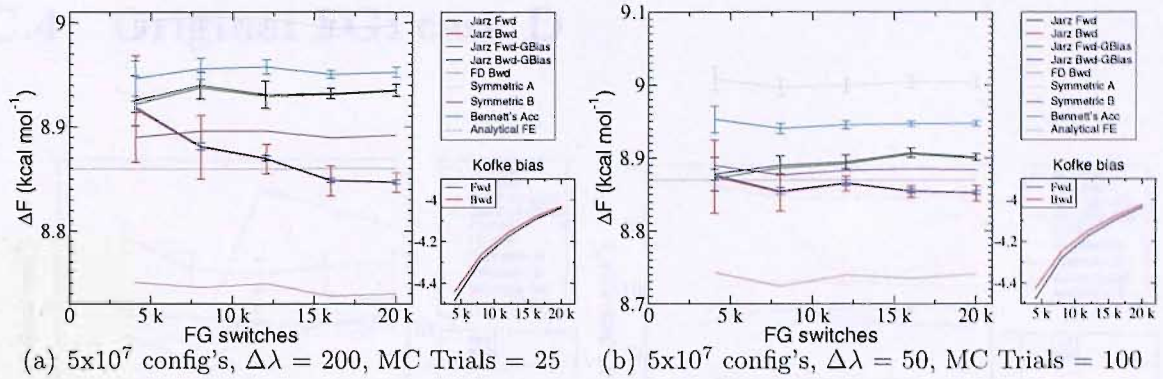
C.2 Original FG: case B



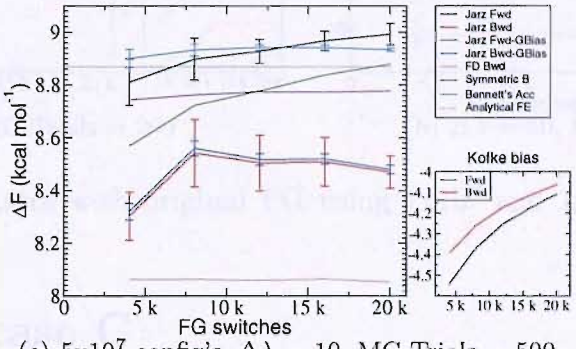
(a) 1x10^8 config's, $\Delta\lambda = 50$ MC Trials = 200

(b) 1x10^8 config's, $\Delta\lambda = 10$, MC Trials = 1000

Figure C.3: ΔF estimates with original FG using 1x10^8 configurations for CaseB



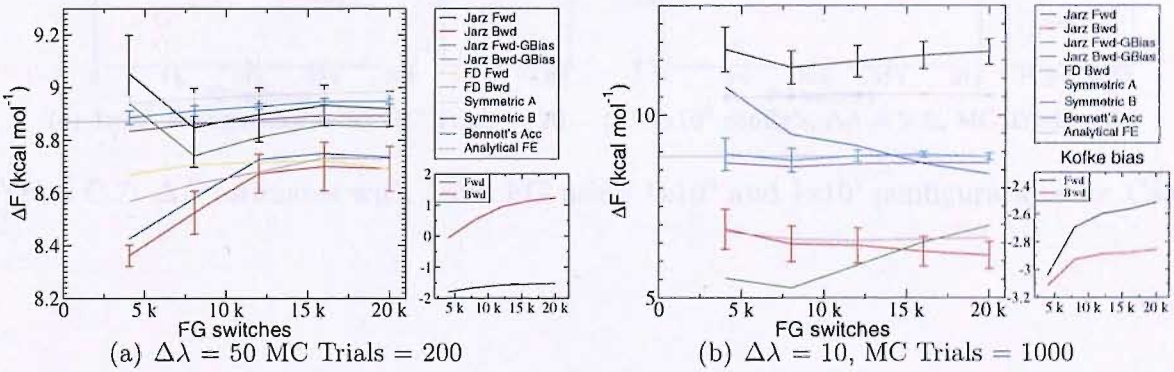
(a) $5x10^7$ config's, $\Delta\lambda = 200$, MC Trials = 25 (b) $5x10^7$ config's, $\Delta\lambda = 50$, MC Trials = 100



(c) $5x10^7$ config's, $\Delta\lambda = 10$, MC Trials = 500

Figure C.4: ΔF estimates with original FG using $5x10^7$ configurations for CaseB

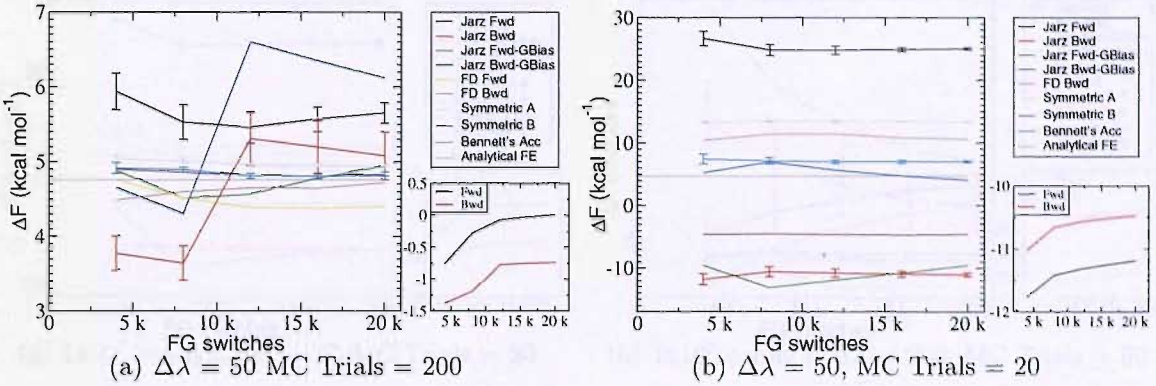
C.3 Original FG: case C



(a) $\Delta\lambda = 50$ MC Trials = 200 (b) $\Delta\lambda = 10$, MC Trials = 1000

Figure C.5: ΔF estimates with original FG using $1x10^8$ configurations for Case C

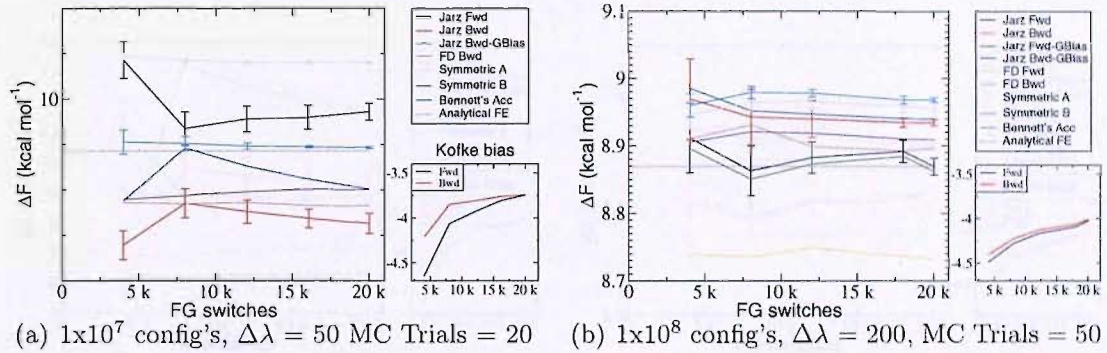
C.4 Original FG: case D



(a) $\Delta\lambda = 50$ MC Trials = 200 (b) $\Delta\lambda = 50$, MC Trials = 20

Figure C.6: ΔF estimates with original FG using $1x10^8$ and $1x10^7$ configurations for Case D

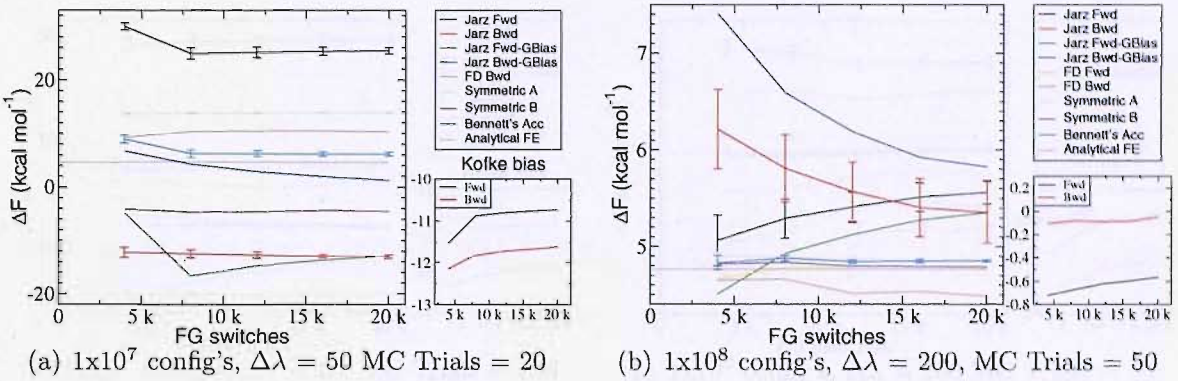
C.5 λ bias: case C



(a) $1x10^7$ config's, $\Delta\lambda = 50$ MC Trials = 20 (b) $1x10^8$ config's, $\Delta\lambda = 200$, MC Trials = 50

Figure C.7: ΔF estimates with λbias FG using $1x10^8$ and $1x10^7$ configurations for Case C

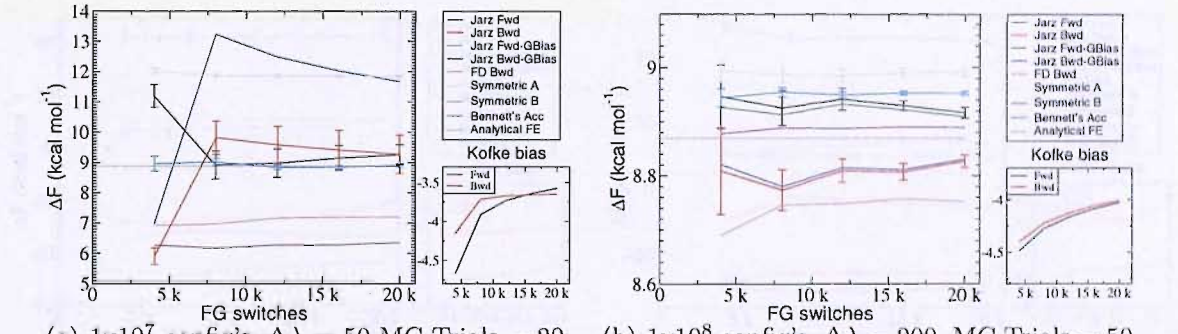
C.6 λ bias: case D



(a) 1x10^7 config's, $\Delta\lambda$ = 50 MC Trials = 20 (b) 1x10^8 config's, $\Delta\lambda$ = 200, MC Trials = 50

Figure C.8: ΔF estimates with λ bias FG using 1x10^8 and 1x10^7 configurations for Case C

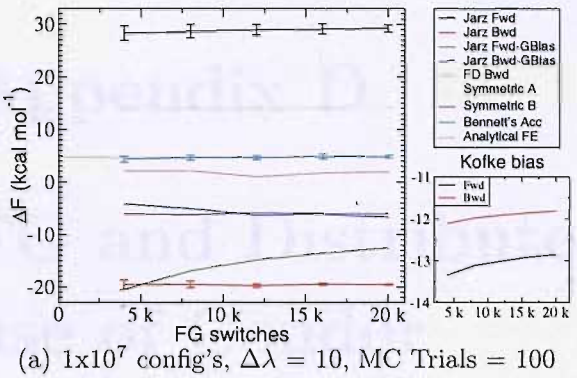
C.7 Configurational bias-c: case C



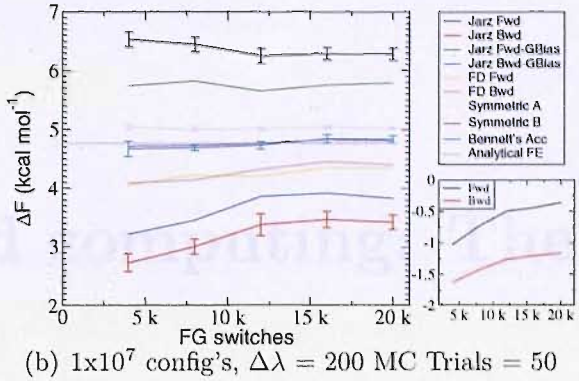
(a) 1x10^7 config's, $\Delta\lambda$ = 50 MC Trials = 20 (b) 1x10^8 config's, $\Delta\lambda$ = 200, MC Trials = 50

Figure C.9: ΔF estimates with confbias-d FG using 1x10^8 and 1x10^7 configurations for Case C

C.8 Configurational bias-d: case D



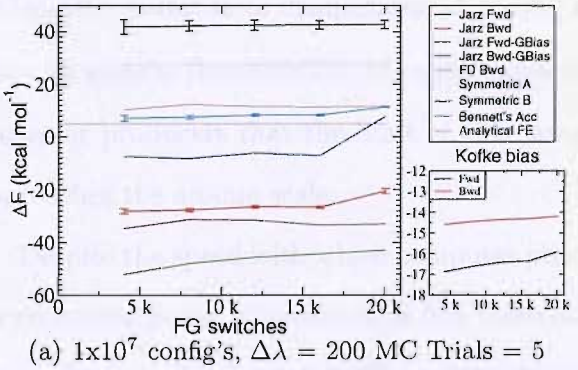
(a) 1x10^7 config's, $\Delta\lambda = 10$, MC Trials = 100



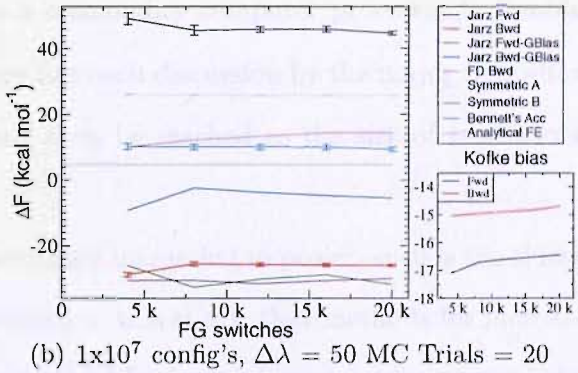
(b) 1x10^7 config's, $\Delta\lambda = 200$ MC Trials = 50

Figure C.10: ΔF estimates with confbias-d FG using 1x10^8 and 1x10^7 configurations for Case D

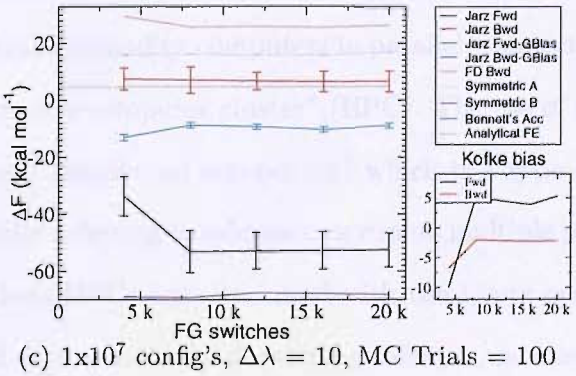
C.9 Hybrid bias: case D



(a) 1x10^7 config's, $\Delta\lambda = 200$ MC Trials = 5



(b) 1x10^7 config's, $\Delta\lambda = 50$ MC Trials = 20



(c) 1x10^7 config's, $\Delta\lambda = 10$, MC Trials = 100

Figure C.11: ΔF estimates with hybrid bias FG using 1x10^8 and 1x10^7 configurations for Case D

Appendix D

FG and Distributed computing: The use of Condor

The power of computer processors has been increasing at an astonishing rate ever since the earliest computers of the 1940s. The rate of increasing power is so steady that the term Moore's Law was coined around 1970 describing the tendency for the number of transistors (which can be loosely converted to computational power) on a commodity computer processor to double every 24 months [Intel(2007)]. In recent times there has been discussion by the major computer processor producers that the limit of this law may soon be reached as the size of transistors approaches the atomic scale.

Despite the speed with which computer processors are increasing in power, such is the thirst for computer power of calculations like those discussed in this study, that methods for increasing power are always sought. The most common method for increasing computer power is to use many cheap yet powerful commodity computers in parallel, connected by a local high speed network in a "high performance computer cluster" (HPC). The use of multiple computers in a single calculation is termed "distributed computing" which is a type of "parallel computing" with the latter term generally referring to calculations run on multiple processors which are part of the same computer. These HPCs are often used with the Linux operating system (OS), as this OS is free and a good environment for developing software, in a set up termed a "Beowulf cluster".

In order that a calculation can be used on a computer cluster, it must be split into smaller calculations which can be distributed across the processors. This limits the efficient use of computer clusters to those calculations which can be parallelised (split up into suitable smaller calculations). Parallel computer program codes (parallel codes) are sometimes able to allocate

jobs to different processors automatically (implicit parallelism). The majority of implicit parallel codes require that all processors and computers included in a calculation are of the same or similar architecture (homogeneous cluster) which can limit the make up of a useful cluster. Also, the number of processors a parallel calculation can use efficiently is in general not unlimited (the scalability of a calculation). This is because the overheads such as the processors communicating with each other over the network slow the calculation [Allan(1999)].

Some calculations are such that they naturally break up into many smaller calculations which can be run independently on single processors of many computers (coarsely parallelised calculations). One example of such a calculation is FG where each switch of a FG calculation can be run on a separate computer, with the work value from each switch combined at the end to give an estimate of the free energy difference.

Coarsely parallelised calculations can be run efficiently with grid computing as there is no necessity for high speed communication between the processors involved. A grid can be a heterogeneous group of computers connected with some level of network to run calculations. The network is often the internet but can be a large non-local network. Grid computing is normally thought to be between computers which are geographically distributed. Some grid clusters are set up to use the unused compute of idle computers in offices and homes. Examples of this are the folding@home project [Shirts & Pande(2000)] which runs molecular simulations on thousands of computers around the world when they are idle.

Condor [Litzkow *et al.*(1988)Litzkow, Livny & Mutka] is workload management software, able to distribute computer jobs within large heterogeneous networks. It is able to identify idle computers (target machines) on the network and distribute jobs to them where the predefined parameters of the computer and jobs fit. Users of Condor submit individual jobs which are then matched to an idle machine which receives the job, completes it and sends the output back to the user. This can effectively produce a large amount of computation that would otherwise be wasted. A good example, which was the case for this study, is a university library which contains many old low specification computers used intermittently by students. Other examples of similar software which is more oriented towards security is Globus [Foster(2006)] and United Devices Virtual Cluster software [UnitedDevices(2007)].

Using Condor with a heterogeneous network of in-use computers can be a delicate and com-

plex task, and thought must be given to producing optimum performance. Here some basic ideas and concepts will be explained from the point of view of the individual user:

First, the code used must be compiled so it can run on the target machines. In the case of this study the majority of the Condor pool used the Microsoft Windows XP operating system (OS), so the code was compiled for Windows.

Each job submission must have a set of parameters defining its target machines on the network. The architecture and OS of the target machines must be given. The memory needs of the individual calculations must be assessed and parameters set such that only machines with enough memory are used. Other possible parameters include, location in the network, disk space and cpu speed.

For each job submitted to a Condor pool the required input files and executables must be transferred to the target machine. Depending on the size of the files to be transferred, this limits the number of jobs a single submit node can service at one time. For the protein-ligand calculations run in this study it was found that 100-150 jobs could be easily serviced from one submit node. This meant that multiple submit nodes were needed to run a whole set of protein-ligand calculations.

Consideration must be given to the length of jobs submitted to a Condor pool. A Condor pool can be set up such that when a user starts using a Condor pool computer which is running a Condor job, this job is either deprioritised or stopped and given back to the Condor pool controller to reassign. Also, sometimes a Condor job can fail and not be returned to the submit node. Either way, the longer a job is, the more chance it has of taking longer than it should or of failing.

However, if jobs are too small the computer from which jobs are submitted to the Condor pool (submit computer) will be unable to finish uploading all the input files for the requested jobs, to the condor pool, before the results of some jobs start to be returned to the submit computer. This issue can effect the efficiency of the calculation as the submit computer is unable to get the requested number of jobs running in the Condor pool. Thus, a balance, must be found allowing the submit computer time to upload all requested jobs, minimising the possibility of jobs slowing or failing, and the demands of the calculations at hand. In this study it was found that FG jobs were at optimum length at around 2-3 hours. Also, it may be advisable to plan for any possible slow or failing jobs by starting a few more than are actually needed.

Using a large Condor pool for free energy calculations has advantages and disadvantages: The speed at which calculations are completed can be unpredictable. If the Condor pool computers are undergoing heavy use by non-Condor users calculations will be very slow. Hence, use overnight and at the weekend can be more efficient. On the other hand, as only coarsely parallelised jobs can be run on Condor and often a large Condor pool is used by only a few projects, extremely large calculations can be run which would be unfeasible on a often oversubscribed dedicated Beowulf clusters. Of course, it would be preferable to run large parallelised FG calculations on large, fast, dedicated clusters as this could produce extremely fast calculations. For large pharmaceutical companies, the use of Condor or Globus pools can be a extremely cost effective solution as they often have large offices full of suitable under-used computers.

Bibliography

[Abrams *et al.*(2006)Abrams, Rosa & Tuckerman] Abrams, J.B., Rosa, L. & Tuckerman, M.E. (2006). Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics. *J Chem Phys*, **125**, 074115.

[Allan(1999)] Allan, R.J. (1999). *High performance computing*. Springer.

[Aqvist & Hansson(1996)] Aqvist, J. & Hansson, T. (1996). On the validity of electrostatic linear response in polar solvents. *J Phys Chem*, **100**, 9512–9521.

[Aqvist & Mowbray(1995)] Aqvist, J. & Mowbray, S.L. (1995). Sugar recognition by a glucose/galactose receptor - evaluation of binding energetics from molecular-dynamics simulations. *J Biol Chem*, **270**, 9978–9981.

[Aqvist *et al.*(1994)Aqvist, Medina & Samuelsson] Aqvist, J., Medina, C. & Samuelsson, J.E. (1994). New method for predicting binding-affinity in computer-aided drug design. *Protein Eng*, **7**, 385–391.

[Ashbaugh & Wood(1997)] Ashbaugh, H. & Wood, R.H. (1997). Effects of long-range electrostatic potential truncation on the free energy of ionic hyrdation. *J Chem Phys*, **106**, 8135–8139.

[Atkins(1994)] Atkins, P.W. (1994). *Physical Chemistry*. Oxford University Press, 5th edn.

[Bairoch(2000)] Bairoch, A. (2000). Enzyme. http://www.expasy.org/enzyme/.

[Barillari *et al.*(2006)Barillari, Taylor, Viner & Essex] Barillari, C., Taylor, J., Viner, R. & Essex, J.W. (2006). Classification of water molecules in protien binding sites. *J Am Chem Soc*, **129**, 2577–2587.

[Bennett(1976)] Bennett, C.H. (1976). Efficient estimation of free energy differences from monte carlo data. *J. Comp. Phys.*, **22**, 245.

[Berendsen *et al.*(1981)Berendsen, Postma, van Gunsteren & Hermans] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. & Hermans, J. (1981). *Intermolecular Forces*. Reidel, Dordrecht, 1st edn.

[Beutler *et al.*(1994)Beutler, Vanschaik, Gerber & van Gunsteren] Beutler, A.E., T. C.; Mark, Vanschaik, R.C., Gerber, P.R. & van Gunsteren, W.F. (1994). A new soft-core potential function for molecular dynamics applied to the prediction of protein loop conformations. *Chem Phys Lett*, **222**, 529–539.

[Blokzijl & Engberts(1993)] Blokzijl, W. & Engberts, J.B.F.N. (1993). Hydrophobic effects - opinions and facts. *Angew Chem Int Edit*, **32**, 1545–1579.

[Blundell & Patel(2004)] Blundell, T.L. & Patel, S. (2004). High-throughput X-ray crystallography for drug discovery. *Current Opinion in Pharmacology*, **4**, 490–496.

[Brenner *et al.*(2007)Brenner, Sweet, VonHandorf & Izaguirre] Brenner, P., Sweet, C.R., VonHandorf, D. & Izaguirre, J.A. (2007). Accelerating the replica exchange method through an efficient all-pairs exchange. *J Chem Phys*, **126**, 074103.

[Brooks *et al.*(1983)Brooks, Bruccoleri, Olafson, States, Swaminathan & Karplus] Brooks, B., Bruccoleri, R.E., Olafson, B., States, D., Swaminathan, S. & Karplus, M. (1983). Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem*, **4**, 187–217.

[Case(2004)] Case, D.A. (2004). Amber 8. *Univeristy of California.*

[Chan(2006)] Chan, F.K.L. (2006). Primer: managing NSAID-induced ulcer complications—balancing gastrointestinal and cardiovascular risks. *Nature Clinical Practice*, **3**, 563–573.

[Chandrasekharan *et al.*(2002)Chandrasekharan, Dai, Lamar Turepu Roos, Evanson, Tomsik, Elton & Simmons] Chandrasekharan, N.V., Dai, H., Lamar Turepu Roos, K., Evanson, N., Tomsik, J., Elton, T.S. & Simmons, D.L. (2002). COX-3, a cyclooxygenase-1 variant inhibited

by acetaminophen and other analgesic/antipyretic drugs: Cloning, structure, and expression. *Proc Nat Acad USA*, **99**, 13926–13931.

[Cheng & Prusoff(1973)] Cheng, Y. & Prusoff, W.H. (1973). Relationship between the inhibition constant (Ki) and the concentration of inhibitor which causes 50 per cent inhibition (IC50) of an enzymatic reaction. *Biochem Pharmacol*, **22**, 3099–3108.

[Cho *et al.*(2006)Cho, Labow, Reinhardt, van Oostrum & Peitsch] Cho, C.R., Labow, M., Reinhardt, M., van Oostrum, J. & Peitsch, M.C. (2006). The application of systems biology to drug design. *Current Opinion in Chemical Biology*, **10**, 294–301.

[Cieplak *et al.*(2001)Cieplak, Caldwell & Kollman] Cieplak, P., Caldwell, J. & Kollman, P. (2001). Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation. *J Comput Chem*, **22**, 1048–1057.

[Cohen & Mauzerall(2004)] Cohen, E.G.D. & Mauzerall, D. (2004). A note on the jarzynski equality. *J Stat Mech: Theory and Experiment*, **P07006**, 1–9.

[Cohen & Mauzerall(2005)] Cohen, E.G.D. & Mauzerall, D. (2005). The jarzynski equaility and the boltzmann factor. *Mol Phys*, **103**, 2923–2926.

[Collin *et al.*(2005)Collin, Ritort, Jarzynski, Smith, Tinoco & Bustamante] Collin, D., Ritort, F., Jarzynski, C., Smith, S.B., Tinoco, I. & Bustamante, C. (2005). Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, **437**, 231–234.

[Colman *et al.*(1983)Colman, Varghese & Laver] Colman, P.M., Varghese, J.N. & Laver, W.G. (1983). Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature*, **303**, 41–44.

[Cossins(2007)] Cossins, B.P. (2007). If copies of extra scripts are required please mail me. bcc@soton.ac.uk.

[Cramer(2002)] Cramer, J. (2002). *Computational Chemistry*. Wiley, 1st edn.

[Crooks(1998)] Crooks, G.E. (1998). Nonequilibrium measurements of free energy differences for microscopically reversable markovian systems. *J Stat Phys*, **90**, 1481–1487.

[Crooks(1999)] Crooks, G.E. (1999). Entropy fluctuation and the nonequilibrium work relation for free energy differences. *Phys rev E*, **60**, 2721–2726.

[Crooks(2000)] Crooks, G.E. (2000). Path-ensemble averages in systems driven far from equilibrium. *Phys Rev E*, **61**, 2361–2366.

[Crooks & Jarzynski(2007)] Crooks, G.E. & Jarzynski, C. (2007). Work distribution for the adiabatic compression of a dilute and interacting classical gas. *Phys Rev*, **75**, 021116.

[Dejong *et al.*(1997)Dejong, Wilson, Neilson & Buckingham] Dejong, P.H.K., Wilson, J.E., Neilson, G.W. & Buckingham, A.D. (1997). Hydrophobic hydration of methane. *Mol Phys*, **91**, 99–103.

[Delle Site(2001)] Delle Site, L. (2001). Methane in water: An ab initio study. *Mol Sim*, **26**, 353.

[Donnini & Juffer(2004)] Donnini, S. & Juffer, A.H. (2004). Calculation of affinities of peptides for proteins. *J Comput Chem*, **25**, 393–411.

[Douarche *et al.*(2005)Douarche, Ciliberto, Petrosyan & Rabbiosi] Douarche, F., Ciliberto, S., Petrosyan, A. & Rabbiosi, I. (2005). An experimental test of the Jarzynski equality in a mechanical experiment. *Europhys Lett*, **70**, 593.

[Errington(2003)] Errington, J.R. (2003). Direct calculation of liquid – vapor phase equilibria from transition matrix monte carlo simulation. *J Chem Phys*, **118**, 9915–9925.

[Essmann *et al.*(1995)Essmann, Perera, Berkowitz, Darden, Lee & Pedersen] Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. & Pedersen, L.G. (1995). A smooth particle mesh Ewald method. *J Chem Phys*, **103**, 8577–8593.

[Evans(2003)] Evans, D.J. (2003). A non-equilibirum free energy theorem for deterministic systems. *Mol Phys*, **101**, 1551–1554.

[Evans & Searles(1994)] Evans, D.J. & Searles, D. (1994). Equilibrium microstates which generate second law violating steady states. *Phys Rev E*, **50**, 1645 – 1648.

[Ewald(1921)] Ewald, P. (1921). Die berechnung optischer und elektrostatischer gitterpotentiale. *Ann. Phys.*, **64**, 253–287.

[Fasnacht *et al.*(2004)Fasnacht, Swendsen & Rosenberg] Fasnacht, M., Swendsen, R.H. & Rosenberg, J.M. (2004). Adaptive integration method for monte carlo simulations. *Phys Rev E*, **69**, 056704.

[Foster(2006)] Foster, I. (2006). Globus toolkit version 4: Software for service-oriented systems. *International Conference on Network and Parallel Computing*, **3779**, 2–13.

[Frenkel & Smit(1996)] Frenkel, D. & Smit, B. (1996). *Understanding molecular simulation*. Academic Press.

[Gilson *et al.*(1997)Gilson, Given, Bush & McCammon] Gilson, M., Given, J., Bush, B. & McCammon, J. (1997). The statistical-thermodynamic basis for computation of binding affinities. a critical review. *Biophys J*, **72**, 1047–1069.

[Gohlke & Case(2004)] Gohlke, H. & Case, D. (2004). Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J Comput Chem*, **25**, 238 250.

[Gore *et al.*(2003)Gore, Ritort & Bustamante] Gore, J., Ritort, F. & Bustamante, C. (2003). Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc Nat Acad Sci USA.*, **100**, 12564–12569.

[Grieg *et al.*(1997)Grieg, Francis, Falgueyret, Ouellet, Percival, Roy, Bayly & O'Neill] Grieg, G.M., Francis, D.A., Falgueyret, J.P., Ouellet, M., Percival, M.D., Roy, P., Bayly, J.A., C. Mancini & O'Neill, G.P. (1997). The interaction of arginine 106 of human prostaglandin g/h synthase-2 with inhibitors is not a universal component of inhibition mediated by nonsteroidal anti-inflammatory drugs. *Mol Pharmacol*, **53**, 829–238.

[Han *et al.*(2001)Han, Kim, Mhin & Son] Han, K.K., Kim, K.H., Mhin, B.J. & Son, H.S. (2001). Application of the multiensemble sampling to the hydration free energy. *J Comp Chem*, **22**, 1004–1009.

[Hansson & Aqvist(1995)] Hansson, T. & Aqvist, J. (1995). Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng*, **8**, 1137 1144.

[Hawkey(1999)] Hawkey, C.J. (1999). COX-2 inhibitors. *Lancet*, **353**, 307–314.

[Hermans(1991)] Hermans, J. (1991). A simple analysis of noise and hysteresis in free energy simulations. *J Phys Chem*, **95**, 9029–9032.

[Hernandez-Cobos *et al.*(2001)Hernandez-Cobos, Mackie & Vega] Hernandez-Cobos, J., Mackie, A.D. & Vega, L.F. (2001). The hydrophobic hydration of methane as a function of temperature from histogram reweighting monte carlo simulations. *J Chem Phys*, **114**, 7527–7535.

[Hu *et al.*(2002)Hu, Yun & Hermans] Hu, H., Yun, R.H. & Hermans, J. (2002). Reversibility of free energy simulations: Slow growth may have a unique advantage. (with a note on the use of ewald summation). *Mol Sim*, **28**, 67–80.

[Hummer(2001)] Hummer, G. (2001). Fast-growth thermodynamic integration: Error and efficiency analysis. *J Chem Phys*, **114**, 7330–7337.

[Hummer(2002)] Hummer, G. (2002). Fast-growth thermodynamic integration: Results for sodium ion hydration. *Mol Sim*, **28**, 81–90.

[Hummer *et al.*(1996)Hummer, Pratt & Garcia] Hummer, G., Pratt, L.R. & Garcia, A.E. (1996). On the free energy of ionic hydration. *J Phys Chem*, **100**, 1206–1215.

[Hummer *et al.*(1998)Hummer, Pratt & Garcia] Hummer, G., Pratt, L.R. & Garcia, A.E. (1998). Molecular theories and simulation of ions and polar molecules in water. *J Phys Chem A*, **102**, 7885–7895.

[Huo *et al.*(2002)Huo, Wang, Cieplak, Kollman & Kuntz] Huo, S.H., Wang, J.M., Cieplak, P., Kollman, P.A. & Kuntz, I.D. (2002). Molecular dynamics and free energy analyses of cathepsin-D inhibitor interactions: Insight into structure-based ligand design. *J Med Chem*, **45**, 1412–1419.

[Intel(2007)] Intel (2007). http://www.intel.com/technology/mooreslaw/index.htm.

[Jakalian *et al.*(2002)Jakalian, Jack & Bayly] Jakalian, A., Jack, D.B. & Bayly, C.I. (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *J Comput Chem*, **23**, 1623–1641.

[Jarzynski(1997a)] Jarzynski, C. (1997a). Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys Rev E*, **56**, 5018–5035.

[Jarzynski(1997b)] Jarzynski, C. (1997b). Nonequilibrium equality for free energy differences. *Phys Rev Lett*, **78**, 2690–2693.

[Jarzynski(2004)] Jarzynski, C. (2004). Nonequilibrium work theorem for a system strongly coupled to a thermal enviroment. *J Stat Mech: Theory and Experiment*, **P09005**, 1-13.

[Jarzynski(2006)] Jarzynski, C. (2006). Rare events and the convergence of exponentially averaged work values. *Phys Rev E*, **73**, 046105.

[Jeancharles *et al.*(1991)Jeancharles, Nicholls, Sharp, Honig, Tempczyk, Hendrickson & Still] Jeancharles, A., Nicholls, A., Sharp, K., Honig, B., Tempczyk, A., Hendrickson, T. & Still, W.C. (1991). Electrostatic contributions to solvation energies - comparison of free-energy perturbation and continuum calculations. *J Am Chem Soc*, **113**, 1454-1455.

[Jhoti(2007)] Jhoti, H. (2007). *Structure based drug discovery*. Springer.

[Johnsen *et al.*(2004)Johnsen, Lindskog, Ponthan, Pettersen, Elfman, Orrego & Sveinbjörnsson] Johnsen, J., Lindskog, M., Ponthan, F., Pettersen, I., Elfman, L., Orrego, A. & Sveinbjörnsson, B. (2004). Cyclooxygenase-2 is expressed in neuroblastoma, and nonsteroidal anti-inflammatory drugs induce apoptosis and inhibit tumor growth in vivo. *Cancer Res*, **64**, 7210–7215.

[Jorgensen(1996)] Jorgensen, W.L. (1996). *MCPRO*.

[Jorgensen & Tirado-Rives(1988)] Jorgensen, W.L. & Tirado-Rives, J. (1988). The OPLS force field for proteins. energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc*, **110**, 1657–1666.

[Jorgensen & Tirado-Rives(2005)] Jorgensen, W.L. & Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc Nat Acad Sci USA.*, **102**, 6665–6670.

[Jorgensen *et al.*(1983)Jorgensen, Chandrasekhar, Madura, Impey & Klein] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J Chem Phys*, **79**, 926–935.

[Jorgensen *et al.*(1984)Jorgensen, Madura & Swenson] Jorgensen, W.L., Madura, J.D. & Swenson, C.J. (1984). Optimized intermolecular potential functions for liquid hydrocarbons. *J Am Chem Soc*, **106**, 6638–6646.

[Kamps *et al.*(2006)Kamps, Hoffmann & Preiser] Kamps, B.S., Hoffmann, C. & Preiser, W. (2006). *Influenza Report*. www.influenzareport.com, 1st edn.

[Kearney *et al.*(2006)Kearney, Baigent, Godwin, Halls, Emberson & Patrono] Kearney, P.M., Baigent, C., Godwin, J., Halls, H., Emberson, J.R. & Patrono, C. (2006). Do selective cyclo-oxygenase-2 inhibitors and traditional non-steroidal anti-inflammatory drugs increase the risk of atherothrombosis? meta-analysis of randomised trials. *Brit Med J*, **332**, 1302–1308.

[Kick *et al.*(1997)Kick, Roe, Skillman, Liu, Ewing, Sun, Kuntz & Ellman] Kick, E.K., Roe, D.C., Skillman, A.G., Liu, G.C., Ewing, T.J.A., Sun, Y.X., Kuntz, I.D. & Ellman, J.A. (1997). Structure-based design and combinatorial chemistry yield low nanomolar inhibitors of cathepsin-D. *Chem Biol*, **4**, 297–307.

[Kirkwood(1935)] Kirkwood, J.G. (1935). Statistical mechanics of fluid mixtures. *J Chem Phys*, **3**, 300–313.

[Klebe(2000)] Klebe, G. (2000). *Virtual Screening*. Springer, 1st edn.

[Kofke(2005)] Kofke, D.A. (2005). Free energy methods in molecular simulation. *Fluid Phase Equilibra*, 41–48.

[Kofke(2006)] Kofke, D.A. (2006). On the sampling requirements for exponential-work free-energy calculations. *Mol Phys*, **104**, 3701–3708.

[Kollman *et al.*(2000)Kollman, Massova, Reyes, Kuhn, Huo, Chong, Lee, Lee, Duan, Wang, Donini, Cieplak Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S.H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D.A. & Cheatham,

T.E. (2000). Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts Chem Res*, **33**, 889–897.

[Kong & Brooks(1996)] Kong, X.J. & Brooks, C.L. (1996). lambda-dynamics: A new approach to free energy calculations. *J Chem Phys*, **105**, 2414–2423.

[Kuhn & Kollman(2000)] Kuhn, B. & Kollman, P.A. (2000). Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J Med Chem*, **43**, 3786–3791.

[Kuhn *et al.*(2004)Kuhn, Gerber & Stahl] Kuhn, B., Gerber, T., P. Schulz-Gasch & Stahl, M. (2004). Validation and use of the MM-PBSA approach for drug discovery. *J Med Chem*.

[Kurchan(1998)] Kurchan, J. (1998). Fluctuation theorem for stochastic dynamics. *J Phys A*, **31**, 3719–3729.

[Kurumbail *et al.*(1996)Kurumbail, Stevens, Gierse, McDonald, Stegeman, Pak, Gildehaus, Miyashiro, Pen
Kurumbail, R.G., Stevens, A.M., Gierse, J.K., McDonald, J.J., Stegeman, R.A., Pak, J.Y., Gildehaus, D., Miyashiro, J.M., Penning, T.D., Seibert, K., Isakson, P.C. & Stallings, W.C. (1996). Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature*, **384**, 644–648.

[Landau(2000)] Landau, D. (2000). *Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press.

[Leach(1996)] Leach, A.R. (1996). *Molecular modelling, principles and applications*. Longman, 1st edn.

[Lecher *et al.*(2006)Lecher, Oberhofer, Dellago & Geissler] Lecher, W., Oberhofer, H., Dellago, C. & Geissler, P.L. (2006). Equilibrium free energies from fast-switching trajectories with large time steps. *J Chem Phys*, **124**, 044113.

[Levy *et al.*(1991)Levy, Belhadj & Kitchen] Levy, R.M., Belhadj, M. & Kitchen, D.B. (1991). Guassian fluctuation formula for electrostatic free energy changes in solution. *J Chem Phys*, **95**, 3627.

[Liphardt *et al.*(2002)Liphardt, Dumont, Smith, Tinoco & Bustamante] Liphardt, J., Dumont, S., Smith, S.B., Tinoco, I. & Bustamante, C. (2002). Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science*, **296**, 1832–1835.

[Litzkow *et al.*(1988)Litzkow, Livny & Mutka] Litzkow, M., Livny, M. & Mutka, M. (1988). Condor - a hunter of idle workstations. *Proceedings of the 8th International Conference of Distributed Computing Systems*, 104–111.

[Lu *et al.*(2005)Lu, Zhang & McCammon] Lu, B., Zhang, D. & McCammon, J.A. (2005). Computation of electrostatic forces between solvated molecules determined by the poisson-boltzmann equation using a boundary element method. *J Chem Phys*, **122**, 214102.

[Lu & Kofke(2001)] Lu, N.D. & Kofke, D.A. (2001). Accuracy of free-energy perturbation calculations in molecular simulation. I. modeling. *J Chem Phys*, **114**, 7303–7311.

[Lynden-Bell & Rasaiah(1997)] Lynden-Bell, R.M. & Rasaiah, J.C. (1997). From hydrophobic to hydrophilic behavoir: A simulation study of solvation entropy and free energy of simple solutes. *J Chem Phys*, **107**, 1981–1991.

[Malhotra *et al.*(2004)Malhotra, Shafiz & Pandhi] Malhotra, S., Shafiz, N. & Pandhi, P. (2004). COX-2 inhibitors: A class act of just vigorously promoted. *Medscape General Medicine*, **6**.

[Maragakis *et al.*(2006)Maragakis, Spichty & Karplus] Maragakis, P., Spichty, M. & Karplus, M. (2006). Optimal estimates of free energies from multistate nonequilibrium work data. *Phys Rev Lett*, **96**, 100602.

[Marelius & Hansson(1998)] Marelius, T. & Hansson, J. (1998). Calculation of ligand binding free energies from molecular dynamics simulations. *Int J Quantum Chem.*

[Mazonka & Jarzynski(1999)] Mazonka, C. & Jarzynski, C. (1999). Exactly solvable model illustrating far-from-equilibrium predictions. *Cond-mat.stat-mech.*

[McDonald & Still(1996)] McDonald, D.Q. & Still, W.C. (1996). Application of free energy perturbation calculations to the enantioselective binding of peptides to C-3-symmetric synthetic receptors. *J Am Chem Soc*, **118**, 207–217.

[Metropolis *et al.*(1953)Metropolis, Rosenbluth, M.N., A.H. & Teller] Metropolis, N., Rosenbluth, A., M.N., R., A.H., T. & Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092.

[Michel(2006)] Michel, J. (2006). The use of free energy simulations as scoring functions. PhD thesis.

[Michel *et al.*(2004)Michel, Taylor & Essex] Michel, J., Taylor, R.D. & Essex, J.W. (2004). The parameterization and validation of generalized born models using the pairwise descreening approximation. *J Comp Chem*, **25**, 1760–1770.

[Michel *et al.*(2006)Michel, Verdonk & Essex] Michel, J., Verdonk, M. & Essex, J. (2006). Protein-ligand binding affinity predictions by implicit solvent simulations: A tool for lead optimization? *J Med Chem*, **49**, 7427–7439.

[Moscona(2005)] Moscona, A. (2005). Neuraminidase inhibitors for influenza. *N Engl J Med*, **353**, 1363–1373.

[Oberhofer *et al.*(2005)Oberhofer, Dellago & Geissler] Oberhofer, H., Dellago, C. & Geissler, P.L. (2005). Biased sampling of nonequilibrium trajectories: Can fast switching simulations outperform conventional free energy calculation methods. *J Phys Chem B*, **109**, 6902.

[Oostenbrink & Gunsteren(2006)] Oostenbrink, C. & Gunsteren, W.F. (2006). Calculating zeros: Non-equilibrium free energy calculations. *Chem Phys*, **323**, 102–108.

[Owicki(1977)] Owicki, H.A., J. C.; Scheraga (1977). Referential sampling near the solute in monte carlo calculations on dilute solutions. *Chem Phys Lett*, **47**, 600–602.

[Pearlman(2005)] Pearlman, D.A. (2005). Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p38 map kinase. *J Med Chem.*

[Pearlman & Charifson(2001)] Pearlman, D.A. & Charifson, P.S. (2001). Are free energy calculations useful in practice? a comparison with rapid scoring functions for the p38 map kinase protein system. *J Med Chem*, **44**, 3417–3423.

[Pearlman & Kollman(1989)] Pearlman, D.A. & Kollman, P.A. (1989). The lag between the hamiltonian and the system configuration in free energy perturbation calculations. *J. Chem. Phys*, **91**, 7831–7839.

[Pearlman *et al.*(1995)Pearlman, Case, Caldwell, Ross, Cheatham, Debolt, Ferguson, Seibel & Kollman] Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E., Debolt, S., Ferguson, D., Seibel, G. & Kollman, P. (1995). AMBER, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun*, **91**, 1–41.

[Picot *et al.*(1994)Picot, Loll & Garavito] Picot, D., Loll, P.J. & Garavito, R.M. (1994). The x-ray crystal-structure of the membrane-protein prostaglandin-h(2) synthase-1. *Nature*, **367**, 243–249.

[Ponder & Case(2003)] Ponder, J.W. & Case, D.A. (2003). Force fields for protein simulations. *Adv Prot Chem*, **66**, 22–85.

[Price & Fletcher(1990)] Price, A.H. & Fletcher, M. (1990). Mechanisms of NSAID-induced gastroenteropathy. *Drugs*, **40**, 1–11.

[Price & Jorgensen(2000a)] Price, D.J. & Jorgensen, W.L. (2000a). Computational binding studies of human pp60(c-src) SH2 domain with a series of nonpeptide, phosphophenyl-containing ligands. *Bioog Med Chem Lett*, **10**, 2067–2070.

[Price & Jorgensen(2000b)] Price, M.P. & Jorgensen, W.L. (2000b). Analysis of binding affinities for celecoxib analogues with COX-1 and COX-2 from combined docking and monte carlo simulations and insight into the COX-2/COX-1 selectivity. *J Am Chem Soc*, **122**, 9455–9466.

[Qian(2000)] Qian, H. (2000). Relative entropy: Free energy associated with equilibrium fluctuations and nonequilibrium deviations. *arXiv:math-ph/0007010v2*.

[Radmer & Kollman(1997)] Radmer, R.J. & Kollman, P.A. (1997). Free energy calculation methods: A theoretical and empirical comparison of numerical errors and a new method for qualitative estimates of free energy changes. *J Comp Chem*, **18**, 902–919.

[Rosenbluth & Rosenbluth(1955)] Rosenbluth, M. & Rosenbluth, A. (1955). Monte carlo calculations of the average extension of molecular chains. *J Chem Phys*, **23**, 356–359.

[Shirts & Pande(2000)] Shirts, M.R. & Pande, V.S. (2000). Screen savers of the world unite. *Science*, **290**, 1903–1904.

[Shirts & Pande(2005)] Shirts, M.R. & Pande, V.S. (2005). Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennetts accentance ratio, and thermodynamic integration. *J Chem Phys*, **122**, 144107.

[Shirts & Sorin(2005)] Shirts, M.R. & Sorin, V.S., E. J. Pande (2005). Direct calulation of the binding free energies of FKBP ligands. *J Chem Phys*, **123**, 084108.

[Shirts *et al.*(2003)Shirts, Blair, Hooker & Pande] Shirts, M.R., Blair, E., Hooker, G. & Pande, V.S. (2003). Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys Rev Let*, **91**, 140601–140604.

[Smith *et al.*(1996)Smith, Sollis, Howes, Cherry, Vobley, Taylor, Whitington, Skarzynski, Cleasby, Singh, V
Smith, P.W., Sollis, S.L., Howes, P.D., Cherry, P.C., Vobley, K.N., Taylor, N.R., Whitington, A.R., Skarzynski, A.J., Cleasby, O., Singh, P.W., T. Wonacott, Varghese, J. & Colman, P. (1996). Novel inhibitors of influenza sialidases related to GC167 - structure - activity, crystalographic and molecular dynamic studies with 4H-pyran-2-carboxylic and molecular dynamic studies with H-pyran-2-carboxylic acid 6-carboxamides. *Bioog Med Chem Lett*, **6**, 2931–2936.

[Sornette(2000)] Sornette, D. (2000). *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization, and Disorder*. 1, Springer.

[Spoel *et al.*(2005)Spoel, Lindahl, Hess, Groenhof, Mark & Berendsen] Spoel, D.D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. & Berendsen, H.J.C. (2005). Gromacs: Fast, flexible, and free. *J Comput Chem*, **26**, 1701–1718.

[Straatsma & Berendsen(1988)] Straatsma, T.P. & Berendsen, H.J. (1988). Free energy of ionic hydration; analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics. *J Chem Phys*, **89**, 5876.

[Sun(2003)] Sun, S.X. (2003). Equilibrium free energies from path samplig of nonequilibrium trajectories. *J Chem Phys*, **118**, 5769.

[Swanson *et al.*(2004)Swanson, Henchman & Mccammon] Swanson, J.M.J., Henchman, R.H. & Mccammon, J.A. (2004). Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys J*, **86**, 67–74.

[Swope *et al.*(1982)Swope, Andersen, Berens & Wilson] Swope, W.C., Andersen, H.C., Berens, P.H. & Wilson, K.R. (1982). The velocity Verlet algorithm. *J Chem Phys*, **76**, 637.

[Taylor *et al.*(1998)Taylor, Cleasby, Singh, Skarzynski, Wonacott, Smith, Sollis, Howes, Cherry, Colman & Taylor, N.R., Cleasby, O., Singh, T., Skarzynski, A.J., Wonacott, P.W., Smith, P.W., Sollis, S.L., Howes, P.D., Cherry, R., P. C. Bethell, Colman, P. & Varghese, J. (1998). Dihydropyrancarboxamides related to zanamivir: A new series of inhibitors of infuenza virus sialidases. 2. crystallographic and molecular modelling study of complexes of 4-amino-4H-pyran-carboxamides and sialidase from influenza virus types A and B. *J Med Chem*, **41**, 798–807.

[Tolman(1924)] Tolman, R.C. (1924). Duration of molecules in upper quantum states. *Phys Rev*, **23**, 693–709.

[Tominaga & Jorgensen(2004)] Tominaga, Y. & Jorgensen, W.L. (2004). General model for estimation of the inhibition of protein kinases using monte carlo simulations. *J Med Chem*, **47**, 2534–2549.

[Ulmschneider & Jorgensen(2003)] Ulmschneider, J.P. & Jorgensen, W.L. (2003). Monte carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a gaussian bias. *J Chem Phys*, **118**, 4261–4271.

[UnitedDevices(2007)] UnitedDevices (2007). http://www.ud.com/solutions/hpc/virtual-cluster.php.

[Varghese *et al.*(1992)Varghese, McKimm-Breschkin, Caldwell, Kortt & Colman] Varghese, J., McKimm-Breschkin, J.L., Caldwell, J.B., Kortt, A.A. & Colman, P. (1992). The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins*, **14**, 327–332.

[Vorobjev *et al.*(1998)Vorobjev, Almagro & Hermans] Vorobjev, Y.N., Almagro, J.C. & Hermans, J. (1998). Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. *Proteins*, **32**, 399–413.

[Wall *et al.*(1999)Wall, Leach, Salt, Ford & Essex] Wall, I.D., Leach, A.R., Salt, D.W., Ford, M.G. & Essex, J.W. (1999). Binding constants of neuraminidase inhibitors: An investigation of the linear interaction energy method. *J Med Chem*, **42**, 5142–5152.

[Warren *et al.*(2006)Warren, Andrews, Capelli, Clarke, LaLonde, Lambert, Lindvall, Nevins, Semus, Senger Warren, G.L., Andrews, C.W., Capelli, A., Clarke, B., LaLonde, J., Lambert, M.H., Lindvall, M., Nevins, N., Semus, S.F., Senger, S., Tedesco, G., Wall, I.D., Woolven, J.M., Peishoff, C.E. & Head, M.S. (2006). A critical assessment of docking programs and scoring functions. *J Med Chem*, **49**, 5912–5931.

[Wood(1991)] Wood, R.H. (1991). Estimation of errors in free energy calculations due to the lag between the hamiltonian and the system configuration. *J Phys Chem*, **95**, 4838- 4842.

[Woods(2003)] Woods, C.J. (2003). The development of free energy methods for protein-ligand complexes. PhD thesis.

[Woods(2007)] Woods, C.J. (2007). The development of a rigorous and efficient QM/MM free energy method. *In preparation*, unpublished results.

[Woods & Michel(2005)] Woods, C.J. & Michel, J. (2005). Protoms2.1. In house Monte Carlo Code.

[Woods *et al.*(2003a)Woods, Essex & King] Woods, C.J., Essex, J.W. & King, M.A. (2003a). The development of replica-exchange-based free-energy methods. *J Phys Chem B*, **107**, 13703–13710.

[Woods *et al.*(2003b)Woods, Essex & King] Woods, C.J., Essex, J.W. & King, M.A. (2003b). Enhanced configurational sampling in binding free-energy calculations. *J Phys Chem B*, **107**, 13711–13718.

[Word *et al.*(1999)Word, Lovell, Richardson & Richardson] Word, J.M., Lovell, J.S., Richardson, J.S. & Richardson, D.C. (1999). Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol*, **285**, 1735–1747.

[Wu & Kofke(2004)] Wu, D. & Kofke, D.A. (2004). Model for small-sample bias of free-energy calculations applied to gaussian-distributed nonequilibrium work measurements. *J Chem Phys*, **121**, 87428747.

[Wu & Kofke(2005a)] Wu, D. & Kofke, D.A. (2005a). Phase-space overlap measures. I. fail-safe bias detaction in free energies calculated by molecular simulation. *J Chem Phys*, **123**, 054103.

[Wu & Kofke(2005b)] Wu, D. & Kofke, D.A. (2005b). Phase-space overlap measures. II. design and implementation of staging methods for free-energy calculations. *J Chem Phys*, **123**, 084109.

[Wu & Kofke(2005c)] Wu, D. & Kofke, D.A. (2005c). Rosenbluth-sampled nonequilibrium work method for calculation of free energies in molecular simulation. *J Chem Phys*, **122**, 204104.

[Xie *et al.*(1991)Xie, Chipman, Robertson, Eriksont & Simmons] Xie, W., Chipman, J.G., Robertson, D.L., Eriksont, R.L. & Simmons, D.L. (1991). Expression of a mitogen-responsive gene encoding prostaglandin synthase is regulated by mRNA splicing. *Proc Nat Acad Sci USA.*, **88**, 2692–2696.

[Ytreberg & Zuckerman(2004a)] Ytreberg, F.M. & Zuckerman, D.M. (2004a). Erratum: "single-ensemble nonequilibrium path-sampling estimates of free energy differences". *J Chem Phys*, **121**.

[Ytreberg & Zuckerman(2004b)] Ytreberg, F.M. & Zuckerman, D.M. (2004b). Single-ensemble nonequilibrium path-sampling estimates of free energy differences. *J Chem Phys*, **120**, 10876–10879.

[Ytreburg & Zuckerman(2004)] Ytreburg, F.M. & Zuckerman, D. (2004). Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems. *J Comput Chem*, **25**, 1749–1759.

[Ytreburg *et al.*(2006)Ytreburg, Swendsen & Zuckerman] Ytreburg, M.F., Swendsen, R.H. & Zuckerman, D. (2006). Comparing free energy methods for molecular systems: Thermodynamic integration, adaptive integration, jarzynski method and path sampling. *Physics*, **2**, 0602088.

[Zacharias *et al.*(1994)Zacharias, Straatsma & McCammon] Zacharias, M., Straatsma, T.P. & McCammon, J.A. (1994). Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J Chem Phys*, **100**, 9025–9031.

[Zhou *et al.*(1998)Zhou, Li, Hawkins, Cramer & Truhlar] Zhou, T.H., Li, J.B., Hawkins, G.D., Cramer, C.J. & Truhlar, D.G. (1998). Density functional solvation model based on CM2 atomic charges. *J Chem Phys*, **109**, 9117–9133.

[Zhuyan *et al.*(2003)Zhuyan, Durkin, Fischmann, Ingram, Prongay, Zhang & Madison] Zhuyan, G., Durkin, J., Fischmann, T., Ingram, R., Prongay, A., Zhang, R. & Madison, V. (2003). Application of the lambda-dynamics method to evaluate the relative binding free energies of inhibitors to HCV protease. *J Med Chem*, **46**, 5360–5364.

[Zuckerman & Woolf(2002a)] Zuckerman, D.M. & Woolf, T.B. (2002a). Overcoming finite-sampling errors in fast-switching free-energy estimates: extrapolative analysis of a molecular system. *Chem Phys Lett*, **351**, 445–453.

[Zuckerman & Woolf(2002b)] Zuckerman, D.M. & Woolf, T.B. (2002b). Theory of a systematic computational error in free energy differences. *Phys Rev Lett*, **89**, 180602.

[Zwanzig(1954)] Zwanzig, R.W. (1954). High temperature equation of state by a perturbation method. I. nonpolar gases,. *J Chem Phys*, **22**, 1420–1426.