Mathematical and Numerical Methods for Free Energy Calculations in Molecular Systems

Workshop held from June 15 to 20, 2008 at the Banff International Research Station for Mathematical Innovation and Discovery

Organized by Eric Darve (Stanford University) and Christophe Chipot (Nancy Université)

Banff International Research Station
for Mathematical Innovation and Discovery

Program of the workshop

Monday: Symplectic integrators for improved molecular dynamics simulations

8:30 am	9:00 am	*beginning of workshop.* **Welcome** **Introduction to BIRS**
9:00 am	10:00 am	**plenary talk: Andrew Pohorille** *Towards efficient calculations of free energies: Modeling probability distributions and adaptive methods*
10:00 am	10:30 am	**coffee break**
10:30 am	11:00 am	**25-minute talk 1: Jesús Izaguirre** *Molecular dynamics using dynamically updated normal modes*
11:00 am	11:30 am	**25-minute talk 2: Tony Lelièvre** *Calculation of free energy differences using parallel adaptive dynamics*
11:30 am	12:00 am	**25-minute talk 3: Alan Mark** *Single point perturbation approaches for estimating relative binding free energies*
12:00 pm	1:00 pm	**lunch break**
1:00 pm	2:00 pm	*Guided tour of the campus*
2:30 pm	3:30 pm	**plenary talk: Ben Leimkuhler** *Equilibrium methods for nonequilibrium modelling*
3:30 pm	4:00 pm	**coffee break**
4:00 pm	4:30 pm	**25-minute talk 4: Emad Noorizadeh** *Hypoelliptic thermostatting methods*
4:30 pm	5:00 pm	**25-minute talk 5: Adrian Lew** *A brief of asynchronous variational integrators*
5:00 pm	5:30 pm	**25-minute talk 6: Frédéric Legoll** *Effective dynamics using conditional expectations*
5:30 pm	6:30 pm	*poster session*
		- He Huang **- Carol Post** **- Juanfang Shen** **- Suriyanarayanan** **- Vaikuntanathan** **- William Fong** **- John Chodera** **- Eric Cyr**
6:30 pm	7:30 pm	**dinner**
7:30 pm		*wine and cheese*

Tuesday: Non-equilibrium methods for equilibrium free energy

9:00am	10:00 am	**plenary talk: Christopher Jarzynski** *Nonequilibrium free energy estimates: improving convergence by reducing dissipation*
10:00 am	10:30 am	**coffee break**
10:30 am	11:00 pm	**25-minute talk 1: Eric Darve** *Coarse graining molecular dynamics models using stochastic equations*
11:00 am	11:30 am	**25-minute talk 2: Peter Bolhuis** *Kinetics and free energy by replica exchange transition interface sampling*
11:30 am	12:00 am	**25-minute talk 3: Sanghyun Park** *Computing conformational free energy by deactivated morphing*
12:00 pm	2:30 pm	**lunch break** *photo session*
2:30 pm	3:30 pm	**plenary talk: Vijay Pande** *Predicting protein-ligand free energies*
3:30 pm	4:00 pm	**coffee break**
4:00 pm	5:30 pm	**25-minute talk 4: Marty Ytreberg** *Computing absolute binding affinities via non-equilibrium unbinding simulations*
4:30 pm	5:00 pm	**25-minute talk 5: Robert Skeel** *Application of the restraining potential method for binding energies*
5:00 pm	5:30 pm	**25-minute talk 6: Angel García** *Calculation of the pressure-temperature folding/unfolding free energy diagrams from molecular dynamics simulations*
6:30 pm		**dinner**

Wednesday: Error analysis, estimation of accuracy and modeling of the density of states

9:00 am	10:00 am	**plenary talk: Benoît Roux** *Building Markov state models along pathways to determine free energies and rates of transitions*
10:00 am	10:30 am	**coffee break**
10:30 am	11:00 pm	**25-minute talk 1: Thomas Beck** *Modeling molecular and ionic solvation free energies with quasi-chemical theory bounds*
11:00 am	11:30 am	**25-minute talk 2: Gabriel Stoltz** *Nonequilibrium computation of free energy differences: Some new algorithms*
11:30 am	12:00 am	**25-minute talk 3: David Kofke** *Fail-safe bias detection and design of stages through consideration of phase-space overlap*
12:00 pm	2:30 pm	**lunch break**
2:30 pm	6:30 pm	***free afternoon***
6:30 pm		**dinner**

Thursday: Methods for enhanced ergodic sampling

9:00 am	10:00 am	**plenary talk: Thomas Woolf** *Improving sampling of rare events: Use of non-Gaussian probability distributions*
10:00 am	10:30 am	**coffee break**
10:30 am	11:00 pm	**25-minute talk 1: Florent Calvo** *Extensions and applications of the Wang-Landau method*
11:00 am	11:30 am	**25-minute talk 2: Jaegil Kim** *Optimal replica exchange method combined with statistical temperature sampling and Tsallis weight sampling*
11:30 am	12:00 am	**25-minute talk 3: Stephen Bond** *Estimating algorithmic bias in molcular dynamics averages*
12:00 pm	2:30 pm	**lunch break**
2:30 pm	3:30 pm	**plenary talk: Mark Tuckerman** *Enhanced conformational sampling and free energies via novel spatial-warping transformations and adiabatic dynamics*
3:30 pm	4:00 pm	**coffee break**
4:00 pm	4:30 pm	**25-minute talk 4: Thanos Panagiotopoulos** *A computational investigation of the solubility and conformations of long n-alkanes in water*
4:30 pm	5:00 pm	**25-minute talk 5: Celeste Sagui** *Adaptively biased molecular dynamics for free energy calculations*
5:00 pm	5:30 pm	**25-minute talk 6: Daniel Zuckerman** *What's your N? Ideas for statistical quality control*
6:30 pm		**dinner**

**Friday: Transition path sampling, ordering parame-
meters *versus* reaction coordinates**

9:00 am	10:00 am	**plenary talk:Eric Vanden-Eijnden**
		Single-sweep methods for free energy calculations
10:00 am	10:30 am	**coffee break**
10:30 am	11:00 pm	**25-minute talk 1: Phil Geissler**
		How fast is best? Optimizing rapid transformations for the purpose of computing reversible work
11:00 am	11:30 am	**25-minute talk 2: Wei Cai**
		Comparison of thermal properties predicted by interatomic potential models
11:30 am	12:00 am	**25-minute talk 3: Christophe Chipot**
		Reaction coordinates versus order parameters. Exploring free-energy landscapes using an adaptive biasing force
12:00 pm	2:30 pm	**lunch break**
2:30 pm		***end of workshop***

Monday, June 17, 2008

Symplectic integrators for improved molecular dynamics simulations

Towards efficient calculations of free energies: Modeling probability distributions and adaptive methods

Andrew Pohorille
NASA Ames Research Center and University of California, San Francisco, U.S.A.

A common objective of molecular simulations in chemistry and biology is to calculate the free energy difference, ΔA, between states of a system of interest [1]. Important examples are protein-drug interactions, protein folding and ionization states of chemical groups. Considerable complexity of these systems dictates the need for efficient determination of ΔA. This is not a simple task, as can be seen from the free energy perturbation (FEP) formalism. In this formalism, ΔA can be represented in terms of a one-dimensional integral of $\exp(-\Delta E / k_B T) \times P(\Delta E)$ over ΔE [1]. In this expression, ΔE is the energy difference between two states of the system, $P(\Delta E)$ is the probability distribution of ΔA, k_B is the Boltzmann constant and T is temperature. For finite systems, $P(\Delta E)$ is a distorted Gaussian. The exponential factor weights heavily the low ΔE tail of $P(\Delta E)$, which is usually known with low statistical precision.

One way to improve estimates of ΔA is to model $P(\Delta E)$. Generally, this approach is rarely successful. Here, however, one can take advantage of the "Gaussian-like" shape of $P(\Delta E)$ and represent it as a linear combination of Gram-Charlier polynomials [3]. The number of terms, N, in this expansion supported by the data is determined using a Bayesian approach. This is done by calculating the posterior probability, $P(N/\Delta E)$, where ΔE stands for all sampled values of ΔE. The dependence of the likelihood function on the coefficients of the expansion, C_N is marginalized by determining their optimal values and then expanding $P(\Delta E)/C_N$, N) around the optimal solution. As expected, the maximum likelihood solution consists of two terms. One is related to the optimal values of C_N and always increases with N. The second term is an "Ockham's Razor" penalty and involves a multivariate Gaussian integral. The method offers the largest improve- ments over conventional approaches when $P(\Delta E)$ is broad and deviates from a Gaussian, and sample size is relatively small.

One alternative to FEP is thermodynamic integration (TI). It appears that in its conventional version, based on constrained simulations, it offers no advantages over FEP. An alternative approach relies on calculating thermodynamic force along a selected order parameter, ξ in unconstrained simulations coupled with an adaptive algorithm for subtracting the position-dependent average force from the instantaneous force, which yields a uniform sampling along ξ [3]. High efficiency of this approach, compared to other, adaptive or non-adaptive algorithms, is due to the fact that forces, in contrast to probabilities, are local properties and, therefore, can be efficiently updated without exploring the reminder of the configurational

space. In its new formulation we have shown how the same information can be obtained from first derivatives with respect to the order parameters and time only [4], which greatly simplifies implementation of this method for ξ that are complex functions of Cartesian coordinates.

The results presented here offer the basis for general statements about efficiencies of different methods for calculating ΔA.

[1] C. Chipot and A. Pohorille (Eds.) Free Energy Calculations: Theory and Applications in Chemistry and Biology, Springer, 2006.
[2] Pohorille, A. and Darve E. (2006) A Bayesian approach to calculating free energies in chemical and biological systems, in Bayesian Inference and Maximum Entropy Methods in Science and Engineering. American Institute of Physics, Melville, NY, pp 23-30.
[3] Darve, E. and Pohorille, A. (2001) Calculating free energies using average force, J. Chem. Phys., 115:9169-9183.
[4] Darve, E. Rodriguez-Gomez, D and Pohorille, A (2008) Adaptive Biasing Force method for scalar and vector free energy calculations, J. Chem. Phys., 128.

Molecular dynamics using dynamically updated normal modes

Jesús A. Izaguirre
University of Notre Dame, U.S.A.

Normal Mode Langevin dynamics (NML) is a method to split Langevin dynamics into slow and fast motions according to normal mode analysis of a protein using an implicit solvent model [Sweet *et al.,* J. Chem. Phys. 128, 145101 (2008)]. The dynamics in the subspace of slow modes are propagated using Langevin dynamics, whereas the fast modes are enslaved to the slow modes. This allows long time steps in the slow dynamics. The method in its original form needs that the slow subspace span with sufficient accuracy the conformational space of interest. To extend the applicability of NML, we use an adaptive partial rediagonalization technique based on filtering spurious fast frequencies introduced to the slow space, and on an iterative method to compute the updated solution of the slow space. This method allows more accurate computation of kinetics or sampling with very few modes (10-12), and is capable of describing processes with substantial conformational change. We also introduce a new Langevin integrator which is more effective for very long time steps. Simulation results on poly-alanine peptides and BPTI will be presented.

Calculation of free energy differences using parallel adaptive dynamics

Tony Lelièvre
Ecole Nationale des Ponts et Chaussées, France

We will present a mathematical viewpoint on adaptive methods for free energy computations, with an emphasis on the Adaptive Biasing Force (ABF) techniques. A new implementation in

parallel of the ABF method is proposed, with an interacting particle system with a selection mechnism associated to a birth death process. Preliminary computations on toy models are presented.

Single point perturbation approaches for estimating rela-tive binding free energies

Alan E. Mark
University of Queensland, Brisbane, Australia

Despite the fact that free energy calculations have been repeatedly demonstrated to provide highly accurate predictions of relative binding free energies for a wide range of drug targets they are still rarely used in the pharmaceutical industry owing to their high cost in terms of both computational and human resources. To facilitate the use of free energy calculations in a drug design context a number of single point approaches have been developed which attempt to estimate the change in free energy from an examination of the end states. One particularly promising approach is the use of the perturbation formula in conjunction with non-physical reference states. In this approach a biasing potential is added to the system such that the phase space sampled in the reference state encompasses a wide range of alternative physical states (e.g. a set of ligands) [1, 2]. The trajectory can then be stored and the binding affinities for a range of alternative end states (compounds) can be estimated by rescanning the trajectory. As the only interactions that that change between the reference state and the chosen end state this approach can be highly efficient. The convergence of the answer is however, heavily dependent on the choice of reference state [3]. The talk will discuss the potential of this approach and current approach to designing alternative reference states.

[1] Schafer, H., van Gunsteren, W.F., Mark, A.E. 1999 Estimating relative free energies from a single ensemble: Hydration free energies J. Comp. Chem. 20, 1604-1617.
[2] Oostenbrink, C, van Gunsteren, W.F. 2005 Free energies of ligand binding for structurally diverse compounds. P.N.A.S., 102, 6750-6754.
[3] Christ, C.D., van Gunsteren, W.F. 2007 Enveloping distribution sampling: A method to calculate free energy differences from a single simulation J. Chem. Phys., 126, 18,184110.

Equilibrium methods for nonequilibrium modelling

Benedict Leimkuhler
University of Edinburgh, Scotland, United Kingdom

Molecular dynamics (MD), if properly used, is an extraordinary tool for computing equilibrium properties and structures, but the portion of phase space accessible to standard MD is often inadequate due to the presence of energetic and/or entropic barriers. For this reason molecular dynamics is increasingly used as part of sophisticated schemes which force the

system into rarely visited regions of phase space, placing increasing demands on the MD apparatus. In particular, thermal equilibration (relaxation of trajectories to equidistribution of energy) is often difficult, and modern techniques require us to compute many well-equilibrated trajectories, in regions of high dynamical activity, with minimal disturbance of dynamics. In this talk, I will describe ongoing work on the development of flexible families of thermostatting methods for molecular dynamics. Hamiltonian based thermostatting methods allow for the use of symplectic integrators which in turn enable estimation of the error in averages as a function of the integration stepsize. Alternatively, stepsize-dependent reweighting formulas allow thermodynamic computations to performed with higher accuracy. While ergodicity-enhancing devices can be incorporated into the Hamiltonian framework, it appears that some type of stochastic perturbation is helpful for obtaining a rapid convergence to canonical sampling in many practical situations. I will describe alternative formulations for stochastic dynamics, generalizing Langevin dynamics but including kinetic energy control, each of which can be analyzed using standard methods. I will also discuss the challenge of obtaining adaptive thermostatting methods.

This talk discusses recent work with a number of collaborators: S. Bond (Illinois), E. Noorizadeh (Edinburgh), F. Legoll (ENPC), F. Theil (Warwick), C. Sweet (Notre Dame), E. Barth (Kalamazoo).

Hypoelliptic thermostatting methods

Emad Noorizadeh
University of Edinburgh, Scotland, United Kingdom

Molecular dynamics requires an artificial device (a thermostat) to sample from the canonical ensemble. Typical thermostats are either stochastic (e.g. Langevin dynamics) or deterministic (e.g. Nose dynamics). Rigorous proofs of ergodicity have been provided for the stochastic methods, but these schemes have several drawbacks (poor control of kinetic energy, loss of smoothness of trajectories, large dynamical perturbation). Deterministic thermostats can correct some of these flaws, but are generally not ergodic. In this talk we describe new methods which introduce stochastic perturbation in a restricted way, retaining some of the advantages of the deterministic approaches. We use the concept of hypoellipticity of the forward Kolmogorov equation, and a Hörmander's condition, to investigate the existence of a unique invariant measure, which implies ergodicity. We compare different schemes in numerical experiments, investigating in particular the smoothness of sampling trajectories obtained in each case.

This talk is based on a recent joint work with Ben Leimkuhler (School of Mathematics, University of Edinburgh) and Florian Theil (Mathematics Institute, University of Warwick).

A brief of asynchronous variational integrators

Adrian Lew
Stanford University, U.S.A.

In this talk I will review the formulation and properties of asynchronous variational integrators. These are explicit multi-time-step integrators in which no integer time step ratio among different time steps are required. I will discuss our recent studies on their stability properties. We demonstrated that if arbitrary rational time steps ratios are permitted, then the set of unstable time steps may be dense. Additionally, I will comment on a recently created asynchronous contact algorithm, with applications in solid dynamics. I will illustrate the performance of these integrators with numerical examples in solid and molecular dynamics.

Effective dynamics using conditional expectations

Frédéric Legoll
Ecole Nationale des Ponts et Chaussées, France

This work aims at getting some insight in the dynamical content of the free energy. More precisely, consider a molecular system evolving according to the overdamped Langevin dynamics, and a given reaction coordinate $\xi(q)$ on this system. The free energy completely determines the statistics of the reaction coordinate. In this work, we try and design an effective dynamics on ξ, that is a low-dimensional dynamics which is a good approximation of $\xi[q(t)]$. Using conditional expectations, we build an original dynamics, and discuss how it is related to the free energy itself. Numerical simulations will illustrate the accuracy of our approach.

Tuesday, June 17, 2008

Non-equilibrium methods for equilibrium free energy

Nonequilibrium free energy estimates: improving convergence by reducing dissipation

Christopher Jarzynski
University of Maryland, College Park, U.S.A.

Nonequilibrium estimates of equilibrium free energy differences are often hampered by convergence difficulties similar to those that plague traditional perturbation methods. In the nonequilibrium context, dissipation is the culprit behind poor convergence. I will discuss a general strategy for addressing these difficulties, in which dissipation is reduced by adding non-physical terms to the equations of motion. When these terms are constructed so as to escort the system along a near-equilibrium path, the efficiency of the free energy estimate can improve dramatically. After sketching the general features of this strategy, I will illustrate its application using specific examples, and will discuss its relation to other non-equilibrium and equilibrium strategies that are similar in spirit.

Coarse graining molecular dynamics models using stochastic equations

Eric Darve
Stanford University, U.S.A.

One application of free energy computation is the calculation of coarse grained models, which can be used to make predictions at time scales not accessible to classical molecular dynamics. There are many formalisms to coarse grain the equations of motion. In this talk we will focus on the Mori-Zwanzig projector formalism. We will present a new algorithm, called Discrete Mori Projector (DMP), to calculate Generalized Langevin Equations based on the Mori-Zwanzig projector. The main advantage of DMP is that the calculation is tractable while remaining generally applicable. We will present applications to computing the generalized friction in implicit water models. We will also present results of the application of DMP to the normal mode dynamics of Izaguirre and Sweet. Finally we will compare stochastic models with fine grained molecular dynamics for a model of insertion of a peptide in a membrane.

Kinetics and free energy by replica exchange transition inter-face sampling

Peter Bolhuis
University of Amsterdam, The Netherlands

The computation of the experimentally accessible rate constants by molecular simulation is

difficult in the case of rare events. Rare events are caused by the presence of high free energy barriers, and special techniques have to be applied to obtain the free energy barriers, such as umbrella sampling or adaptive biasing force. Other techniques focus on the rate constant computation, e.g. path sampling, and its related technique transition interface sampling (TIS). I present a method based on a combination of replica exchange with TIS. Sampling a distribution of dynamical pathways, this methods allows, in a single simulation, the computation the complete kinetics and free energy. Moreover, the path sampling efficiency improves dramatically. I will demonstrate the method for a simple 2D potential.

Computing conformational free energy by deactivated morphing

Sanghyun Park
Argonne National Laboratories, U.S.A.

Despite the significant advances in free-energy computations for biomolecules, there exists no general method to evaluate the free-energy difference between two conformations of a macromolecule that differ significantly from each other. A crucial ingredient of such a method is the ability to find a path between different conformations that allows an efficient computation of the free energy. The free-energy difference is the same no matter what path is taken, but not all paths are equally practical. In this talk I will present a new method named `deactivated morphing'. An important feature of this method is the use of nonphysical paths, which makes the method robustly applicable to conformational changes of arbitrary complexity.

Predicting protein-ligand free energies

Vijay Pande
Stanford University, U.S.A.

Predicting protein-ligand binding free energies is a major challenge for free energy methods, both due to the importance of this problem in areas such as computational drug design as well as the challenging free energy problem this presents. I will discuss our recent advances in free energy methodology in this area, with applications to specific protein-ligand binding systems.

Computing absolute binding affinities via non-equilibrium unbin-ding simulations

F. Marty Ytreberg
University of Idaho, Moscow, U.S.A.

We discuss the advantages and disadvantages of estimating absolute binding affinities via

non-equilibrium unbinding simulations. The study utilizes the FKBP protein bound to two different ligands as a model system. The non-equilibrium methodology utilized is straight-forward, requiring little or no modification to modern molecular simulation packages, and is trivially parallelizable. The approach makes use of a physical pathway, eliminating the need for complicated alchemical decoupling schemes.

Application of the restraining potential method for binding energies

Robert Skeel
Purdue University, West Lafayette, U.S.A.

Calculation of the absolute free energy of binding is a difficult task due to the complexity of sampling the transition state. The restraining potential method introduced in a 2005 PNAS article by Woo and Roux finesses this difficulty by means of a thermodynamic cycle involving a restrained version of the original system. The basic ideas of this method are presented here, from a slightly different perspective. The method was originally tested on a protein-ligand complex for a ligand of 4 amino acid residues. Presented here is a calculation for the association of the kinase inducible domain of the CREB protein with the 28-residue KIX domain of the CREB-binding protein. Results are obtained for both the unphosphorylated and phosphorylated KID. These are compared to experimental results. The calculations uncover a dual role for the contribution of the phosphorlyation to a reduction in the binding free energy.

Calculation of the pressure-temperature folding/unfolding free energy diagrams from molecular dynamics simulations

Angel E. García
Rensselaer Polytechnic Institute, New York, U.S.A.

The folded state of a protein is described as a highly packed conformation. However, upon an increase of the hydrostatic pressure, proteins will unfold. This seems to contradict physical intuition where a low volume state is preferred at high pressures. The solution of the clue is that the unfolded state of the protein reduces the overall volume of the protein in solution by packing water molecules in the protein interior. An analysis of the potential of mean force of small non polar molecules in water supports this idea. However, the balance between energy and volume in biomolecular systems may be more complex than for simple hydrophobic solutes in water.

In my talk I will describe atomic simulations of the folding/unfolding equilibrium of a small protein (trp-cage) and of an RNA tetraloop that exhibit pressure induced unfolding and cold denaturation. We will show that the structure and hydration of the unfolded state at low T and high P is different from the unfolded state at high T and low P. The equilibrium pressure-

temperature free energy of folding, $\Delta G(P,T),$ is calculated from replica exchange molecular dynamics simulations. This free energy diagram has an elliptical shape, similar to what has been observed in globular proteins.

Wednesday, June 18, 2008

Error analysis, estimation of accuracy and modeling of the density of states

Building Markov state models along pathways to determine free energies and rates of transitions

Benoît Roux
University of Chicago, U.S.A.

An efficient method is proposed for building Markov models with discrete states able to accurately describe the slow relaxation of a complex system with two stable conformations. First, the reaction pathway described by a set of collective variables between the two stable states is determined using the string method with swarms-of-trajectories. Then, short trajectories are initiated at different point along this pathway to build the state-to-state transition probability matrix. It is shown, using a model system, how this strategy makes it possible to use trajectories that are significantly shorter than the slowest relaxation time to efficiently build a reliable and accurate Markov model. Extensions of the method to multiple pathways, as well as some common pitfalls arising from poorly relaxed paths or an inappropriate choice of collective variables are illustrated and discussed.

Modeling molecular and ionic solvation free energies with quasi-chemical theory bounds

Thomas L. Beck
University of Cincinnati, U.S.A.

A recently developed statistical mechanical Quasi-Chemical Theory (QCT) has led to significant insights into solvation phenomena for both hydrophobic and hydrophilic solutes. The QCT exactly partitions solvation free energies into three components:
1) inner-shell chemical, 2) outer-shell packing, and 3) outer-shell long-ranged contributions. Here we discuss efficient methods for computing each of the three parts of the free energy. A Bayesian estimation approach is developed to compute the inner-shell chemical and outer-shell packing contributions. The outer-shell long-ranged component can be expressed in two equivalent ways, resulting in upper and lower bounds on that part of the free energy. Also, the spatial conditioning in this free energy piece leads to near-Gaussian distributions of solute-solvent interaction energies. Thus, the average of the two mean-field bounds yields an accurate and efficient free energy estimate. Aqueous solvation free energy results are presented for several solutes, including methane, perfluoromethane, water, and the sodium and chloride ions. The results demonstrate the accuracy and efficiency of the methods developed here. The approach should prove useful in computing solvation free energies in inhomogeneous, restricted environments.

Nonequilibrium computation of free energy differences: Some new algorithms

Gabriel Stolz
Ecole Nationale des Ponts et Chaussées, France

The Jarzynski equality $\mathscr{E}(e^{-\beta W}) = e^{-\beta F}$ is of great theoretical interest in the context of nonequilibrium statistical mechanics. However, it is unclear at the moment whether numerical strategies based on this equality are competitive (as compared to more traditional approaches such as thermodynamic integration, or recent adaptive strategies). I present some recent developments contributing to more accurate simulations based on the Jarzynski equality, in particular:

- an extension to the case of reaction coordinates (when transitions cannot be indexed by some alchemical parameter);
- selection strategies to avoid the degeneracies of the exponential weights;
- refined path sampling stratgies when stochastic dynamics are used.

Fail-safe bias detection and design of stages through conside-ration of phase-space overlap

David A. Kofke
University at Buffalo, U.S.A.

Work-based free-energy calculation methods are highly prone to inaccuracy, and in some cases it is easy to proceed correctly and yet obtain very precise numbers that are wrong. The quality of the result depends very much on the nature of the systems being studied together with the details of how the work-based calculation is performed. The key to understanding bias is the relation between the parts of phase space important to each system (labelled *A* and *B*), a relation which may range from full overlap (system *B* a phase-space subset of *A*), to partial overlap, to no overlap. Intuitively appealing constructs can be developed from this picture, and their application can guide the use of molecular simulation so that it provides results that are both precise and accurate.

Phase-space relations can be quantified by the relative entropies, s_A and s_B. These quantities are related to the dissipated work for instantaneous processes $A{\to}B$ and $B{\to}A$, respectively. The relative entropy can be used to scale the amount of sampling to obtain an apparently universal bias curve. This result leads to the development of a simple heuristic which can be applied to determine whether a work-based free-energy measurement is accurate. We demonstrate that this bias-detection measure is effective in identifying inaccuracy in single- as well as multi-stage calculations.

Thursday, June 19, 2008

Methods for enhanced ergodic sampling

Improving sampling of rare events: Use of non-Gaussian probability distributions

Thomas Woolf
John Hopkins University, Baltimore, U.S.A.

Rare events frequently dominate the convergence properties of equilibrium and non-equilibrium sampling for free energy differences. Getting faster and more confident answers from a simulation thus depends on improvements in reaching to and finding the rare events behind a process. This talk will discuss ideas for improved sampling of rare events and present ideas based on non-Gaussian density functionals and importance sampling. The ideas will be framed in both the equilibrium and non-equilibrium contexts and for both alchemical and conformational relative free energy changes.

Extensions and applications of the Wang-Landau method

Florent Calvo
Université Paul Sabatier Toulouse III, France

The Wang-Landau algorithm was proposed in 2001 as a recursive procedure to compute the density of states of classical systems. Even though its theoretical foundations still need to be clarified, the method has been increasingly used since its introduction. In the present work, we describe possible improvements for continuous
classical systems, and a simple application to the calculation of free energies along a reaction coordinate. We also discuss another extension of the method to the case of quantum densities of states for systems whose energy levels can be cast by a Dunham expansion.

Optimal replica exchange method combined with statistical temperature sampling and Tsallis weight sampling

Jaegil Kim
Boston University, U.S.A.

Abstract: Replica exchange method (REM) has been combined with the statistical temperature sampling (RESTMC) and Tsallis weight sampling (Tsallis-REM) to accelerate the convergence of the conventional temperature replica exchange method (t-REM) and alleviate its slowing down of the convergence with the increase of the system size. Instead of

Gibbs-Boltzmann sampling at a specific temperature in t-REM each replica samples a range of temperature using the replica-dependent statistical temperature in RESTMC and utilizes a generalized ensemble sampling associated with an optimally parameterized Tsallis weight in Tsallis-REM, which enable a considerable acceleration of the convergence of simulations with much fewer replicas by systematically modulating energy overlaps between neighboring replicas. The performance of our methods has been explicitly demonstrated in various simulation conditions using tunneling events in replica and energy space in Lennard-Jones atomic clusters, exhibiting single- and double-funneled energy landscapes, respectively.

Estimating algorithmic bias in molecular dynamics averages

Stephen Bond
University of Illinois at Urbana-Champaign, U.S.A.

In the context of molecular dynamics, accuracy and efficiency of a numerical method should be measured with respect to statistical averages, rather than deviations from an exact trajectory. In this talk I will survey some results from backward error analysis and show how (under certain assumptions) these results can be applied to compute estimates of the truncation error in averages from molecular dynamics simulations. Results from several test problems will be explored, including examples from canonical and microcanonical molecular dynamics.

Enhanced conformational sampling and free energies via novel spatial-warping transformations and adiabatic dynamics

Mark E. Tuckerman
Courant Institute of Mathematical Sciences, New-York University, U.S.A.

One of the computational grand challenge problems is to develop methodology capable of sampling conformational equilibria in systems with rough energy landscapes. If met, many important problems, most notably protein structure prediction, could be significantly impacted. In this talk, I will present a new approach [1,2] in which molecular dynamics is combined with a novel variable transformation designed to warp configuration space in such a way that barriers are reduced and attractive basins stretched. The new method rigorously preserves equilibrium properties while leading to very large enhancements in sampling efficiency. The performance of the method is demonstrated on long polymer chains and simple protein models and is shown to significantly outperform replica-exchange Monte Carlo with only one trajectory [2]. If time permits, I will discuss the use of adiabatic dynamics to generate multi-dimensional free-energy surfaces [3,4], our recent development that circumvents the need for explicit coordinate transformations based on recent work of Maragliano and Van den Eijnden [5], and the possibility of adapting this approach to help ease the implementation of spatial-warping transformations.

[1] Z. W. Zhu, M. E. Tuckerman, S. O. Samuelson and G. J. Martyna, Phys. Rev. Lett. 88 100201 (2002).

[2] P. Minary, M. E. Tuckerman and G. J. Martyna, SIAM J. Sci. Comput. (in press).

[3] L. Rosso, P. Minary, Z.W. Zhu and M. E. Tuckerman, J. Chem. Phys. 116, 4389 (2002).

[4] L. Rosso, J. B. Abrams and M. E. Tuckerman, J. Phys. Chem B 109, 4162 (2005).

[5] L. Maragliano and E. Vanden-Eijnden, Chem. Phys. Lett. 426, 168 (2006).

A computational investigation of the solubility and conforma-tions of long n-alkanes in water

Athanassios Z. Panagiotopoulos
Princeton University, U.S.A.

The aqueous solubility of long n-alkanes is so slight that equilibrium concentrations become extremely low, leading to significant scatter in experimental data for chains heavier than tetradecane and essentially no reliable data beyond hexadecane. Computer simulations provide a method to determine the long-chain solubilities, but due in part to the severe sampling difficulties associated with the Widom insertion of long chains into the solvent, have only been conducted for the lighter chains. We present a heteropolymer generalization of the incremental Widom insertion technique combined with replica exchange molecular dynamics simulations to obviate these sampling problems and provide a reliable determination of n-alkane solubilities up to docosane. Agreement of our results with experimental data for the lighter chains is excellent and the results for the heavier chains are, to our knowledge, the first ever reported. Henry's constant exhibits a maximum as a function of chain length around hexadecane, which is attributable to an increasing proportion of compact conformers that possess more favorable free energies of solvation. The solvated chains show remarkable structural similarity to the ideal gas phase, with quite subtle, yet important, differences.

Adaptively biased molecular dynamics for free energy calculations

Celeste Sagui
North Carolina State University, Raleigh, U.S.A.

We present an Adaptively Biased Molecular Dynamics (ABMD) method for the computation of the free energy surface of a reaction coordinate using non-equilibrium dynamics. The ABMD method belongs to the general category of umbrella sampling methods with an evolving biasing potential. The ABMD method has several useful features, including a small number of control parameters, and an $O(t)$ numerical cost with molecular dynamics time t. The ABMD method naturally allows for extensions based on multiple walkers and replica exchange, where different replicas can have different temperatures and/or collective variables. This is beneficial not only in terms of the speed and accuracy of a calculation, but also in terms of the amount of useful information that may be obtained from a given simulation. The workings of the ABMD method are illustrated via a study of the folding of U-turn peptide in a gaseous and solvated environment.

What's your *N*? Ideas for statistical quality control

Daniel Zuckerman
University of Pittsburgh, U.S.A.

The statistical quality of equilibrium sampling is a limiting factor for many free energy calculations. Furthermore, the assessment of force fields and sampling methodologies cannot proceed systematically without statistical quality control. We therefore propose as a standard N_{phys}, the effective sample size governing the statistical uncertainty in physical-state populations. The use of state populations can greatly decrease the role of subjectivity in sample size assessment, and surprisingly, it is possible to estimate N_{phys} "blindly" without any a priori information about a system.

Friday, June 20, 2008

Transition path sampling, ordering parameters *versus* reaction coordinates

Single-sweep methods for free energy calculations

Eric Vanden-Eijnden
Courant Institute of Mathematical Sciences, New-York University, U.S.A.

Free energy (or potential of mean force) calculations are a central issue in biophysics. Molecular dynamics (MD) simulations provide a tool for performing such calculations on a computer in a way which is potentially both precise and inexpensive. Since a free energy associated with some collective variables is proportional to the logarithm of the probability density function of these variables, it can in principle be calculated by histogram methods based on the binning of an MD trajectory. This direct approach, however, turns out to be unpractical in general because the time scale required for the trajectory to explore all the relevant regions of configuration space is prohibitively long. The standard way around this difficulty is to use umbrella sampling methods such as the weighted histogram analysis method (WHAM), but this technique is restricted to calculations of free energies in low dimension, i.e. when the number of collective variables is small (typically, 1 or 2). An alternative approach, however, is to calculate the gradient of the free energy (i.e. the mean force) locally, and use these data to reconstruct the free energy globally. In this talk, I will present a new method based on this idea which can be used to map free energy landscapes in multidimensions. The method uses radial-basis functions to represent the free energy and a variational approach to reconstruct the free energy globally from the local data on the mean force. The method will be illustrated on the several examples and compared with existing techniques based on histogram methods such as WHAM or metadynamics.

Comparison of thermal properties predicted by interatomic potential models

Wei Cai
Stanford University, U.S.A.

We report melting points and other thermal properties of several semiconducting and metallic elements as they are modelled by different empirical interatomic potential models, including Stillinger-Weber (SW), embedded-atom-method (EAM), Finnis-Sinclar (FS) and modified-embedded-atom-method (MEAM). The state-of-the-art free energy methods are used to determine the melting points of these models within a very small error bar, so that they can be cross-compared with each other. The comparison reveals several systematic trends

among elements with the same crystal structure. It identifies areas that require caution in the applications of these models and suggests directions for their future improvements.

How fast is best? Optimizing rapid transformations for the purpose of computing reversible work

Philip Geissler
University of California, Berkeley, U.S.A.

The discovery of profound connections between the statistics of irreversible processes and those of equilibrium ensembles opened intriguing new routes for calculating free energies. Although many advances have been made for exploiting these connections, e.g., biased sampling of irreversible trajectories according to the work they require, it is not immediately clear whether they in fact improve upon the efficiency of more conventional approaches. We have performed a systematic comparison of the computational resources required for biased free energy calculations via nonequilibrium transformations and the alternative of standard importance sampling. We find for model high-dimensional systems that the optimal irreversible transformation proceeds in a single step, corresponding to Zwanzig's method of free energy perturbation. Using an analogy between non-equilibrium dynamics and mappings between equilibrium ensembles, we examine the generality of this result and offer a novel improvement based on integrating equations of motion with large time steps. Finally, we describe an advance in the methodology of path sampling, appropriate for the harvesting of driven trajectories, that extends its applicability to sluggish dynamics in highly chaotic systems.

Reaction coordinates *versus* order parameters. Exploring free- energy landscapes using an adaptive biasing force

Chris Chipot
Université Henri Poincaré, Nancy Université, France

The adaptive biasing force algorithm constitutes a robust, converging method for measuring free-energy changes along a given order parameter, ξ, or combination thereof. It is aimed at providing a uniform sampling of ξ without any a priori know-ledge of the actual free-energy landscape. As will be shown through a variety of case examples, convergence of the calculation is rapidly reached if ξ is fully decoupled from other, slowly relaxing degrees of freedom. Unfortunately, under most circum-stances, reaction coordinates are modeled by surrogate, usually one-dimensional order parameters. By and large, order parameters that depart from the *true* reaction coordinate yield convergence issues that hamper uniform sampling along ξ.

List of participants

Ioan Andricioaei	University of Michigan, U.S.A.	andricio@umich.edu
Thomas Beck	University of Cincinatti, U.S.A.	becktl@email.uc.edu
Peter Bolhuis	University of Amsterdam, The Netherlands	bolhuis@science.uva.nl
Stephen Bond	University of Illinois, U.S.A.	sdbond@illinois.edu
Wei Cai	Stanford University, U.S.A.	caiwei@stanford.edu
Florent Calvo	Université Paul Sabatier Toulouse III, France	fcalvo@lasim.univ-lyon1.fr
John Chodera	University of California, San Francisco, U.S.A.	jchodera@gmail.com
Christophe Chipot	Nancy Université, Nancy, France	Christophe.Chipot@edam.uhp-nancy.fr
Eric Cyr	University of Illinois, Urbana-Champaign, U.S.A.	ericcyr@uiuc.edu
Eric Darve	Stanford University, U.S.A.	darve@stanford.edu
William Fong	Stanford University, U.S.A.	willfong@stanford.edu
Angel García	Rensselaer Polytechnic Institute, New York, U.S.A.	angel@rpi.edu
Philip Geissler	University of California, Berkeley, U.S.A.	geissler@cchem.berkeley.edu
He Huang	Purdue University, U.S.A.	huang23@purdue.edu
Jesús Izaguirre	University of Notre-Dame, U.S.A.	izaguirr@nd.edu
Christopher Jarzynski	University of Maryland, College Park, U.S.A.	cjarzyns@umd.edu
Jaegil Kim	Boston University, U.S.A.	jaegil@bu.edu
David Kofke	University at Buffalo, U.S.A.	kofke@buffalo.edu
Frédéric Legoll	Ecole Nationale des Ponts et Chaussées, France	legoll@lami.enpc.fr
Tony Lelièvre	Ecole Nationale des Ponts et Chaussées, France	lelievre@cermics.enpc.fr
Ben Leimkuhler	University of Leicester, U.K.	B.Leimkuhler@ed.ac.uk

Adrian Lew	Stanford University, U.S.A.	lewa@stanford.edu
Alan Mark	University of Queensland, Brisbane, U.S.A.	a.mark@uq.edu.au
Emad Noorizadeh	Edinburgh University, U.K.	e.noorizadeh@sms.ed.ac.uk
Athanassios Panagiotopoulos	Princeton University, U.S.A.	azp@princeton.edu
Vijay Pande	Stanford University, U.S.A.	pande@stanford.edu
Sanghyun Park	Argonne National Laboratories, U.S.A.	sp.spark@gmail.com
Andrew Pohorille	NASA Ames Research Center, U.S.A.	pohorill@raphael.arc.nasa.gov
Carol Post	Purdue University, U.S.A.	cbp@purdue.edu
Benoît Roux	University of Chicago, U.S.A.	roux@uchicago.edu
Celeste Sagui	North Carolina State University, U.S.A.	sagui@unity.ncsu.edu
Juanfang Shen	Purdue University, U.S.A.	shen12@purdue.edu
Robert Skeel	Purdue University, U.S.A.	skeel@cs.purdue.edu
Gabriel Stoltz	Ecole Nationale des Ponts et Chaussées, France	stoltz@cermics.enpc.fr
Mark Tuckerman	New-York University, U.S.A.	mark.tuckerman@nyu.edu
Suri Vaikuntanathan	University of Maryland, U.S.A.	svaikunt@umd.edu
Eric Vanden-Eijnden	New-York University, U.S.A.	eve2@cims.nyu.edu
Thomas Woolf	Johns Hopkins University, U.S.A.	woolf@groucho.med.jhmi.edu
F. Marty Ytreberg	University of Idaho, U.S.A.	ytreberg@uidaho.edu
Daniel Zuckerman	University of Pittsburgh, U.S.A.	dmz@ccbb.pitt.edu