Recent Advances in Transition Path Sampling: Accurate Reaction Coordinates and Diffusive Transition Paths
Baron Peters, Gregg Beckham, Bernhardt Trout
Department of Chemical Engineering and Department of Chemistry, University of California, Santa Barbara, CA, 93106

Background
An accurate reaction coordinate is tantamount to a complete mechanistic understanding. With an accurate reaction coordinate we can:
- efficiently compute reaction rates and - kinetically meaningful free energy barriers: $\Delta G^\circ, \Delta H^\circ, \Delta S^\circ$, etc.
- characterize the transition state ensemble and - use properties of the transition states to predict kinetic trends
- identify the degrees of freedom for which large fluctuations lead to reactions vs. fluctuations that led nowhere

Computational chemistry has long excelled at using saddle search algorithms, transition state Theory (TST), and intrinsic reaction coordinates to understand the mechanisms of bond breaking-making reactions. By comparison we have been utterly unable to identify accurate reaction coordinates (and therefore mechanisms) for processes like nucleation, protein folding, and reactions that involve solvent molecules.

There is a fundamental reason these two groups of activated processes pose very different challenges. In bond breaking-making processes, key bonds stretch (or shorten) by just -1Å as the system moves from deep in the reactant basin to deep in the product basin. On that length scale the landscape is smooth and therefore theories like harmonic TST work fairly well. Now consider crystal nucleation from a supersaturated solution. Solutes must migrate through a sea of solvents to join the nucleus. After many solute attachments and detachments the nucleus may become much larger than the critical size. Although the free energy as a function of nucleus size is smooth, the underlying potential energy surface for barrier crossing is rugged. Initial momenta become irrelevant, and we must instead think about mean first passage times for motion along collective variable reaction coordinates.

Nucleation in the LJ fluid

Transition path sampling (TPS), in principle provided a framework for studying activated processes on rugged landscapes. However, there were two key problems with it:
1. "TPS is inefficient for transitions that take longer than a few picoseconds."
2. "Identifying reaction coordinates is a trial and error procedure."

TPS and Aimless Shooting

Transition Path Sampling
- Monte Carlo in (reactive) trajectory space
- dynamics not biased along pre-chosen $q$
- true ensemble of reactive trajectories
- original TPS by shooting and shifting

Aimless Shooting = TPS +
- shoots are indep. realizations of $p_b(x)$
- shooting points are auto-distributed as $p(T|p(x))(T|x)$ (near $p(x)$ of 1 and in the transition pathway)
- highly efficient for long diffusive paths
- trajectories deconvoluted more rapidly than in original TPS, one-sided shooting, or precision shooting
- very easily implemented
 www.engineering.ucsb.edu/baronp

Likelihood Maximization

Weekly curvature $\kappa$ in separate TPS trajectories from $q_1$ to $q_2$ for reaction coordinate $q_1$, $q_2$
- $\kappa > 0$ indicates potential energy surface is smoother near $q_2$ than near $q_1$
- $\kappa < 0$ indicates potential energy surface is rugged near $q_2$ than near $q_1$
- $\kappa = 0$ indicates potential energy surface is flat near $q_2$ than near $q_1$

Conclusions

Aimless shooting efficiently samples long transition paths
Likelihood maximization identifies accurate reaction coordinates from Aimless Shooting data
Sampling error deconvolution makes coordinate validation fast and quantitative
Identified a structure-corrected nucleus size metric that accurately predicts the committor probability – an assumption of classical nucleation theory

Improved Coordinate Validation

Original trial and error approaches used cummouted analysis to validate reaction coordinates. Committor analysis remains the most stringent test for coordinate accuracy. The procedure to validate a coordinate $q$ in
1. estimate the dividing surface $q = q^*$ (was previously done based on a free energy calculation $q^* = \text{argmax} \langle F(q), \text{now likelihood maximization gives estimate for } q^* \rangle$
2. harvest a putative transition state ensemble (TSE) from $p_b(x|q<q^*)$ (with restrained molecular dynamics or MC)
3. for each $x$ in the putative TSE, compute the committor probability and create a histogram $H(q_b(x))$. Contrast to earlier prescriptions, accurate $p_b(x)$ estimates not needed for the histogram – see next step!
4. deconvolute physically meaningful coordinate error from the binomial $p_b(x)$ estimation error.

More Likely Max Examples

peptide hydrolysis in water, Pan and Trout (2009)
solid-state polymorph transition, Biesinger et al. (2007, 2008)
NaCl dissociation in water, Mullen et al. in preparation

committor probability

Each configuration $x$ has a committor (or splitting) probability $p_b(x)$ – the fraction trajectories that commit to the product state (8) when initiated at $x$ with Boltzmann momenta $p(x)$ is the reaction coordinate for rugged landscapes (also a good reaction coordinate for smooth landscapes) However, physical coordinates like nucleus size or contact maps for protein folding provide more mechanistic insight. Therefore we seek models of the committor in terms of simple physical collective variables.

References