WHAM

The Weighted Histogram Analysis Method

Alan Grossfield

November 7, 2003

Outline

- Statistical Mechanics
- Non-Boltzmann Sampling
- WHAM equations
- Practical Considerations
- Example: Butane

Statistical Mechanics

- Partition Function

$$Z_{NVT} = \int e^{-U(x_0,\ldots,x_{3N})/k_B T} dx_0 \ldots dx_{3N}$$

- Free Energy

$$A_{NVT} = -k_B T \ln(Z_{NVT})$$

Potential of Mean Force

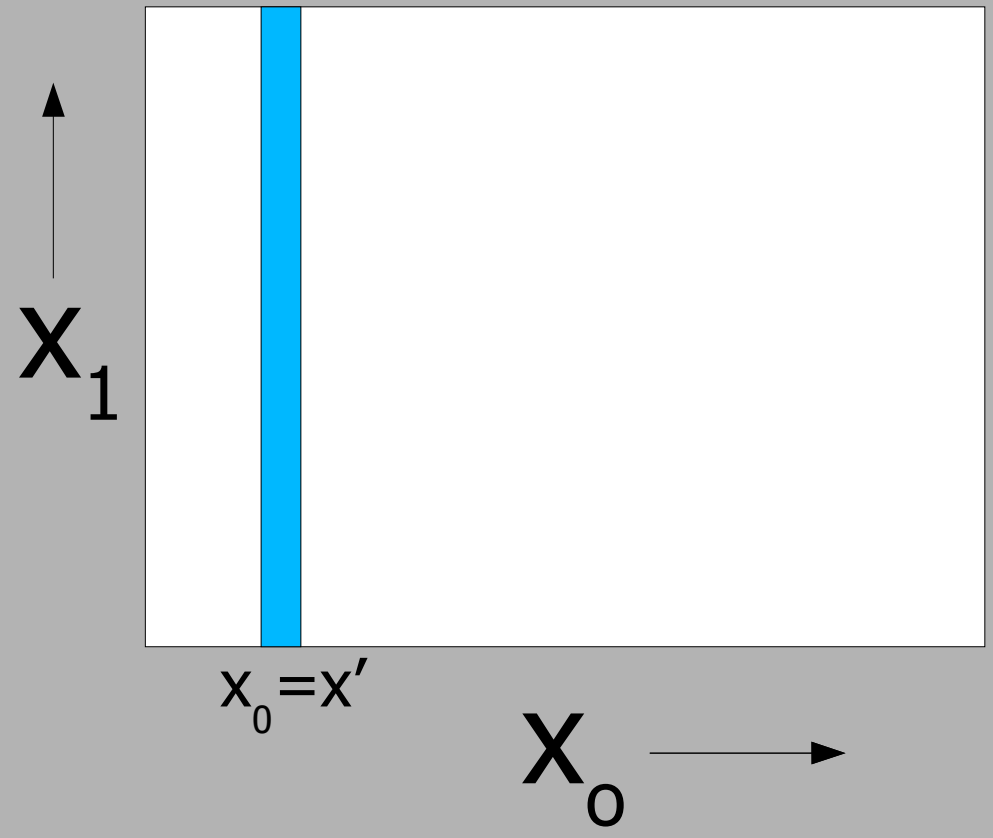
- Extract degree of freedom from partition function

$$Z_{NVT}(x_0') = \int e^{-U(x_0,\ldots,x_{3N})/k_B T} \delta(x_0' - x_0) dx_0 \ldots dx_{3N}$$

- Free energy

$$A_{NVT}(x_0') = -k_B T \ln(Z_{NVT}(x_0'))$$

2-D Example: $U(x_0,x_1)$



$$A(x') = -k_B T \ln \int e^{-U(x', x_1)/k_B T} \, dx_1$$

Dependence on x_1 integrated out

Computing PMFs

- Free energy related to probability

$$P(x') \propto e^{-A(x')/k_B T}$$

- Simple approach
  - Run canonical MD or Monte Carlo
  - Compute probability distribution P(x)
  - P(x) determines A(x) up to a constant

Simple Umbrella Sampling

- Problem: P(x) converges slowly due to barriers along x
- Solution: Add an additional potential term to the energy to encourage barrier crossing

Simple Umbrella Sampling

- Sample with umbrella potential U'(x)
- Compute biased probability P'(x)
- Estimate unbiased free energy

$$A(x) = -k_B T \ln P'(x) - U'(x) + F$$

- F is undetermined but irrelevant

General Umbrella Sampling

- Complex surfaces have multiple barriers
- Need to know the free energy surface to know an efficient bias
- Multicanonical ensemble simulations
  - Iteratively guess appropriate biasing potential to give a flat free energy surface

General Umbrella Sampling

- Harmonic biasing function
- Multiple simulations
- Put the minimum of the bias in a different place for each simulation
- Estimate P'(x) for each simulation
- Combine results from all simulations

Combining Simulations

- From one simulation

$$A(x) = -k_B T \ln P'(x) - U'(x) + F$$

- F depends on U'(x)
- Different simulations have different offsets
- Not obvious how to weight

WHAM

- Weighted Histogram Analysis Method
- Determines optimal F values for combining simulations
- Kumar, et al. J Comput Chem, 13, 1011-1021, 1992
- Generalizations
  - Multidimension reaction coordinates
  - Multiple temperatures

WHAM Equations

$$P(x) = \frac{\sum_{i=1}^{N_{sims}} n_i(x)}{\sum_{i=1}^{N_{sims}} N_i \exp(\left[F_i - U_{bias,i}(x)\right]/k_B T)}$$

$$F_i = -k_B T \ln\left\{\sum_{x_{bins}} P(x) \exp\left[-U_{bias,i}(x)/k_B T\right]\right\}$$

- N_{sims} = number of simulations
- $n_i(x)$= number of counts in histogram bin associated with x
- $U_{bias,i}$, F_i = biasing potential and free energy shift from simulation i
- P(x) = best estimate of unbiased probability distribution
- F_i and P(x) are unknowns
- Solve by iteration to self consistency

Running a Simulation

- Choose the reaction coordinate
- Choose the number of simulations and the biasing potential
- Run the simulations
- Compute time series for the value of the reaction coordinate
- Apply the WHAM equations

Reaction Coordinate

- Sometimes obvious
  - dihedral angle of butane
  - backbone dihedrals of alanine dipeptide
- Some care required
  - PMF depends on choice of coordinate
  - Volume element may not be constant along reaction coordinate

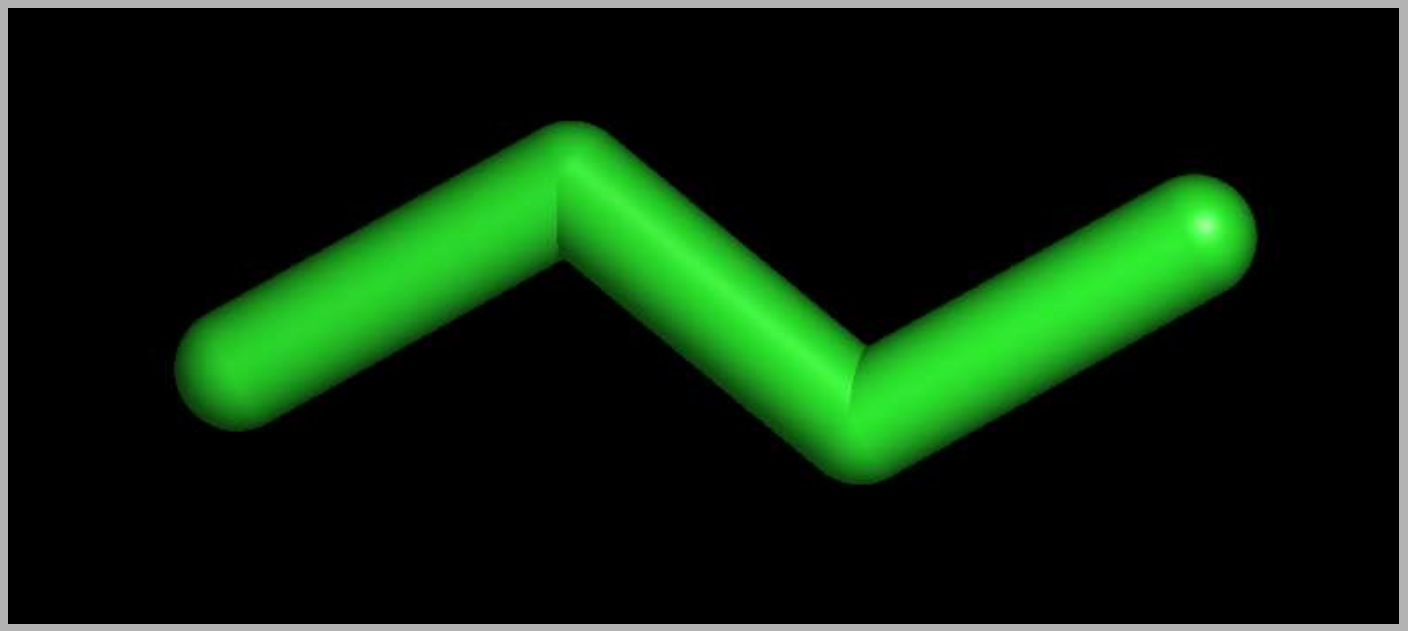
Number of simulations

- Need adequate number of points everywhere on reaction coordinate
- Varies with bias
  - strong bias means more simulations
  - weaker bias allows fewer simulations, but may not ensure sampling in unfavorable regions
  - need to allow each simulation to equilibrate
- Requires knowledge of your system

Running Simulations

- Must generate proper ensemble
  - Derivation assumes NVT or NPT
  - Monte Carlo or Molecular Dynamics
  - Berendsen thermostat is broken
- Must equilibrate
- Must sample "enough"
  - individual histograms should look smooth
  - should have data everywhere

Example: Butane



- Compute PMF for rotating dihedral of united atom butane
- PMF integrates out the effects of flexible bonds and angles

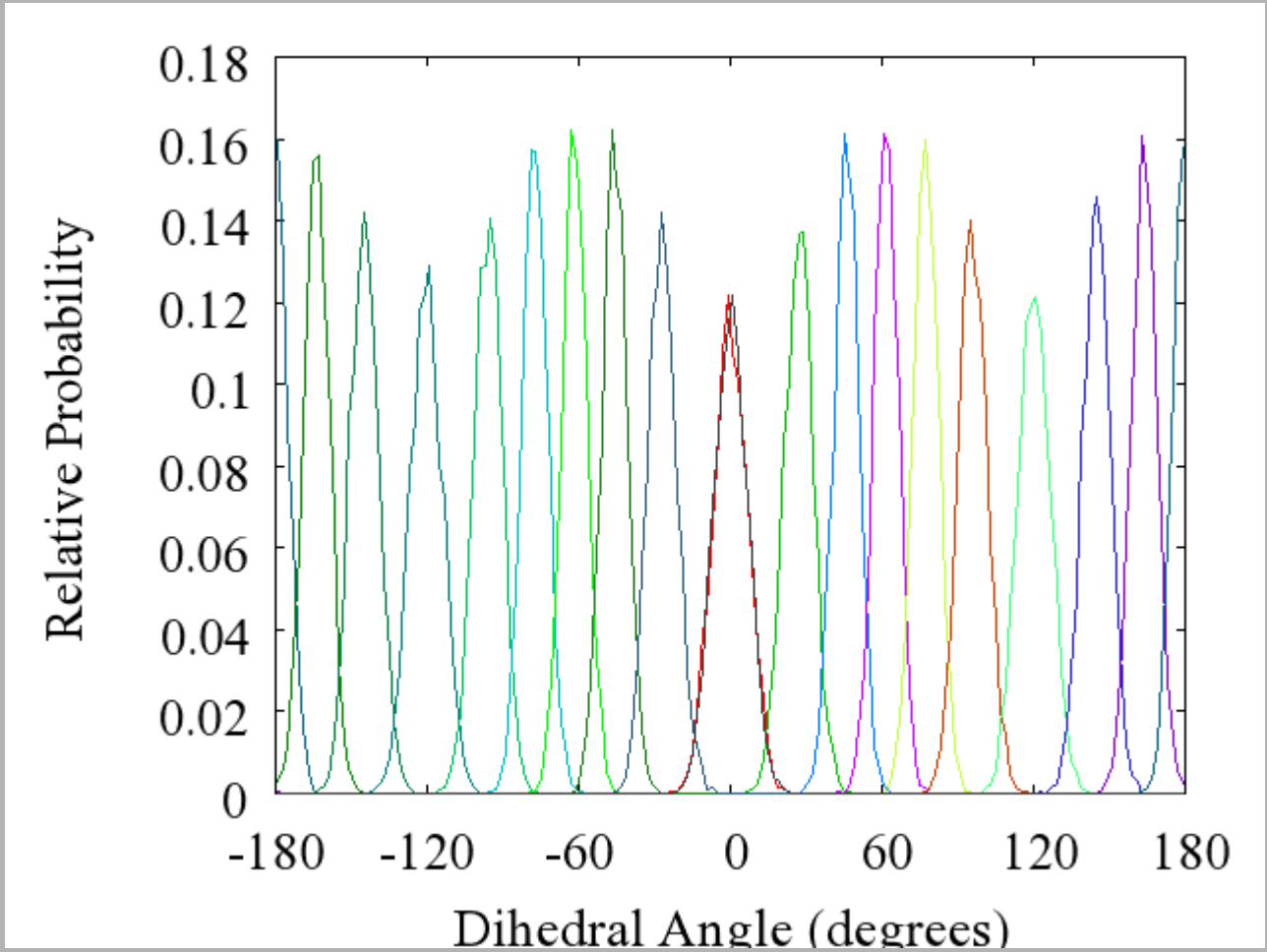
Protocol

- 18 independent simulations
  - 500 ps
  - Restraint spring constant = 0.02 kcal/mol-deg^2
  - T=300K (stochastic dynamics)
- WHAM
  - 90 bins (4^0/bin)
  - Enforced periodicity

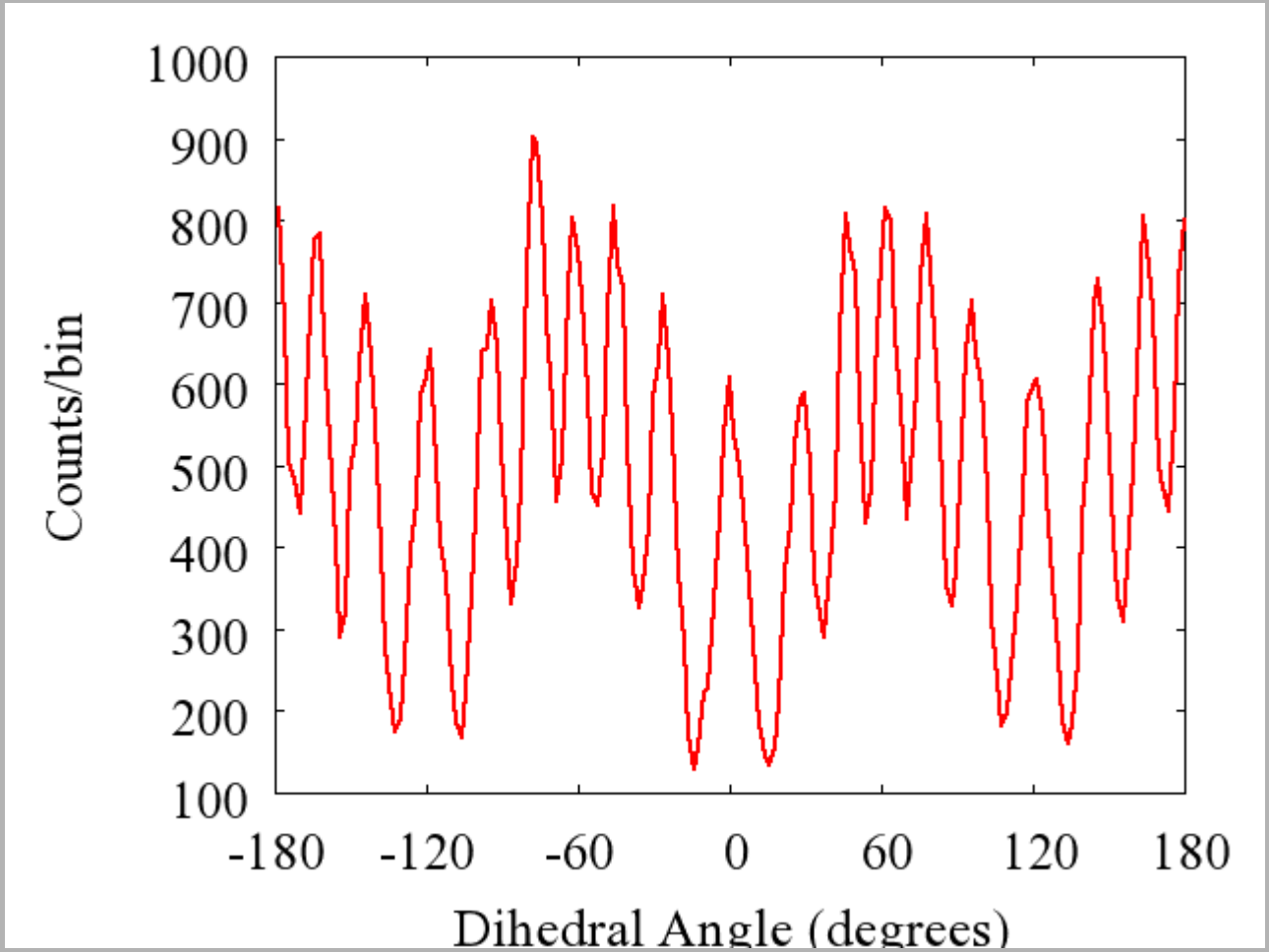
Preliminary Checks

- Examine histograms from individual time series
  - No major shifts from expected value
  - Noisy histograms mean noisy PMF
- Examine histogram of all data
  - Must not have any bare patches
- Figure out appropriate bin size

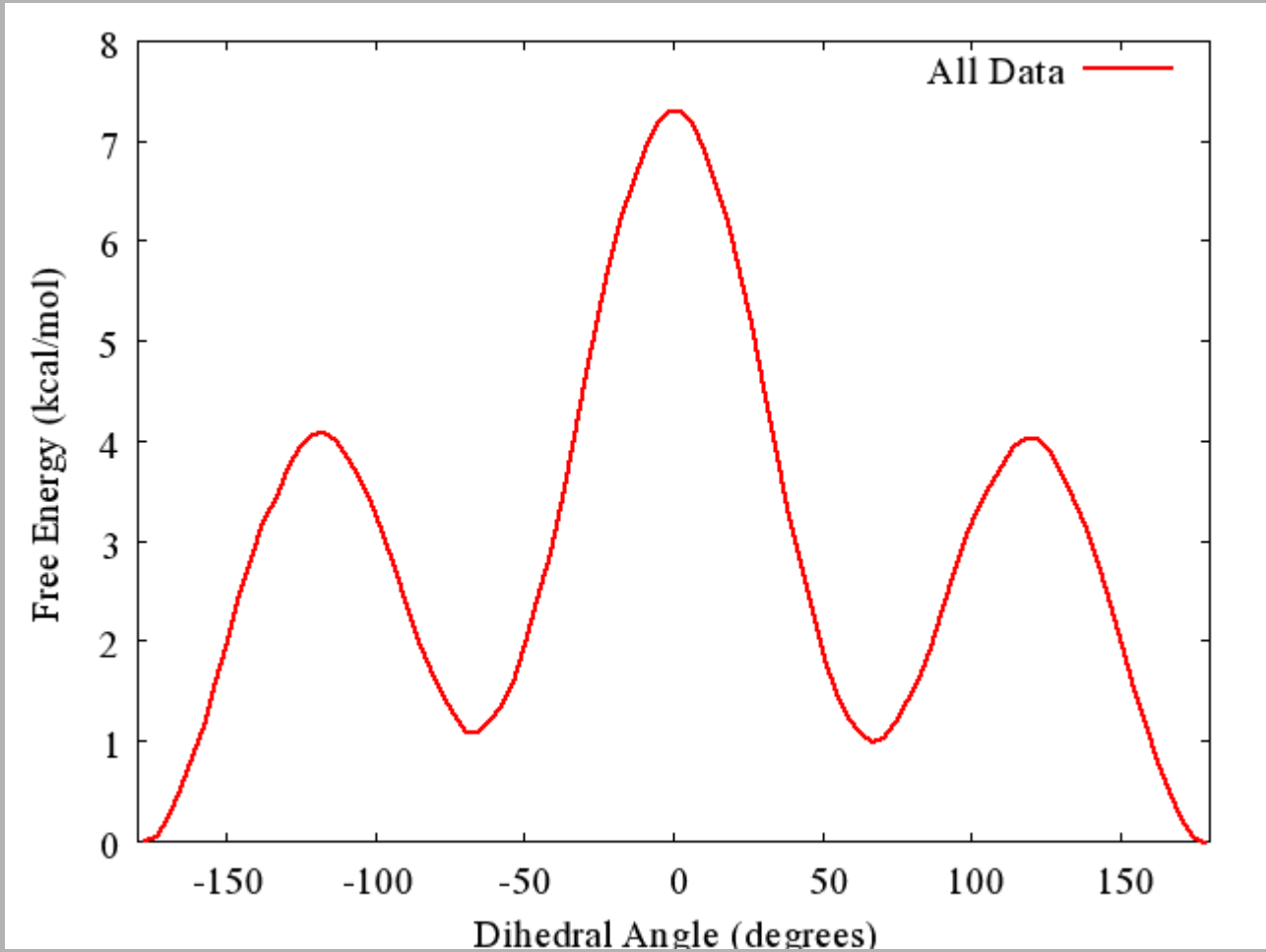
Histograms from Individual Trajectories



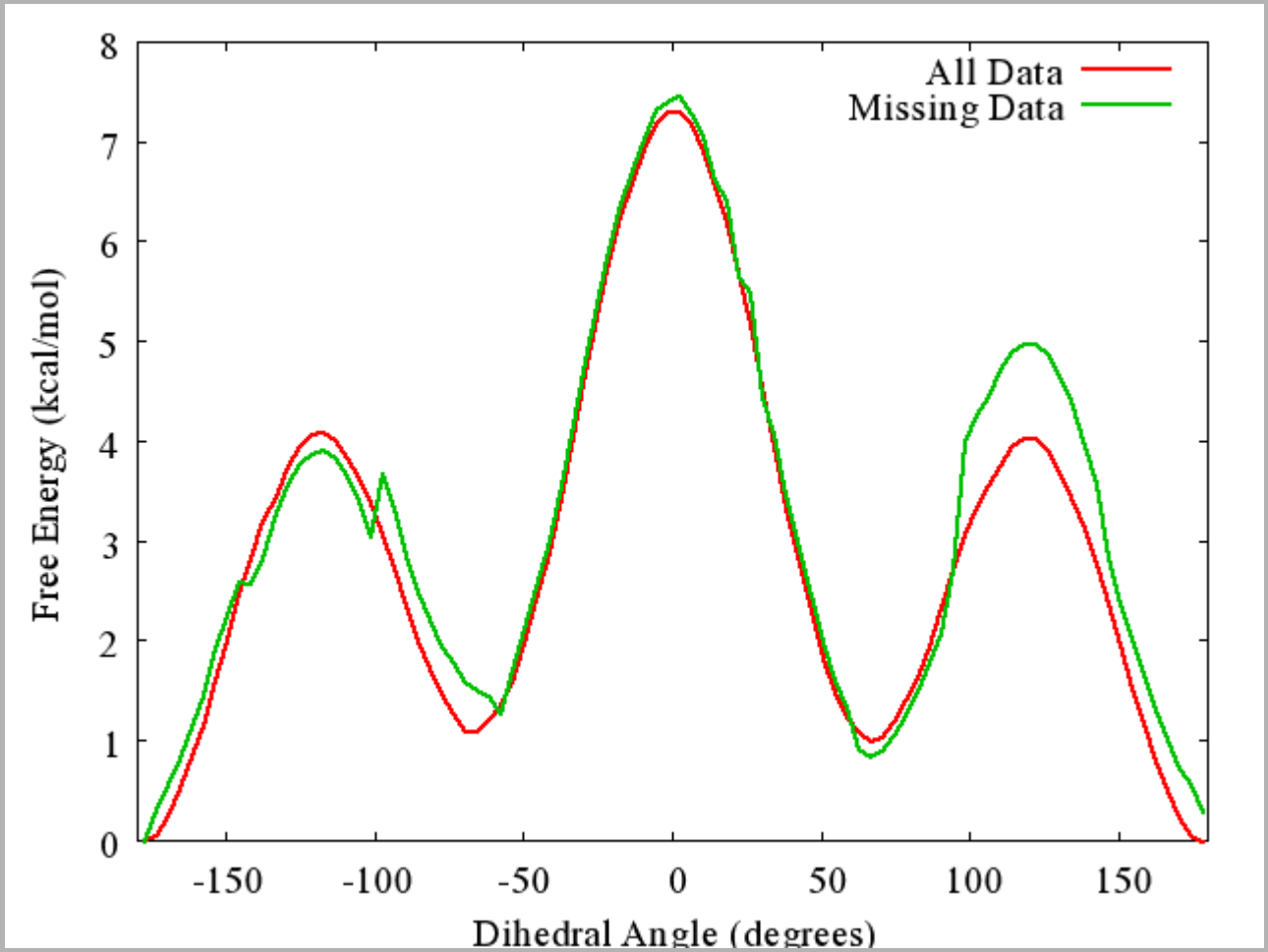
Histogram of Combined Trajectories



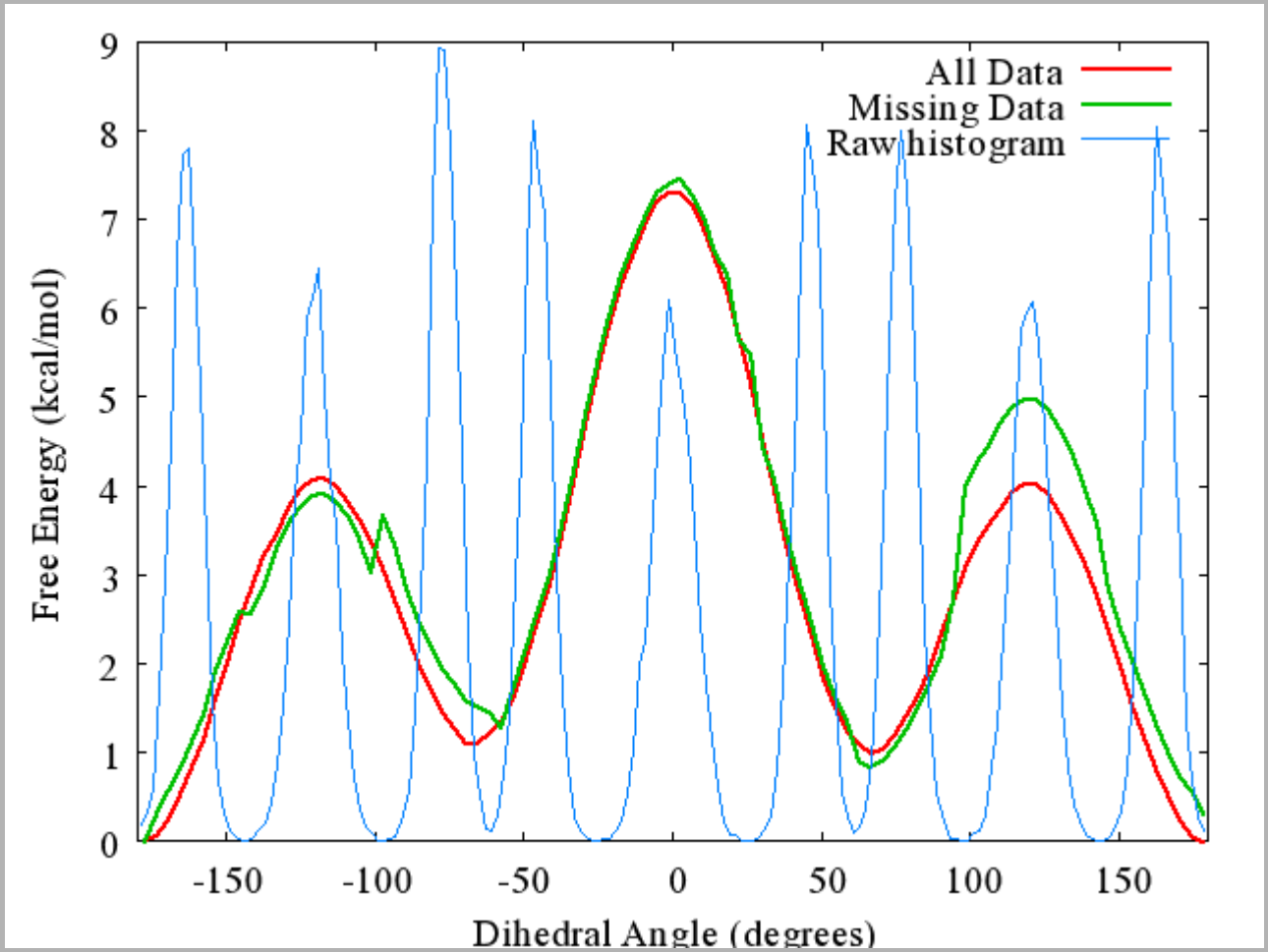
Butane PMF


All Data
Free Energy (kcal/mol)
Dihedral Angle (degrees)
8
7
6
5
4
3
2
1
0
-150
-100
-50
0
50
100
150

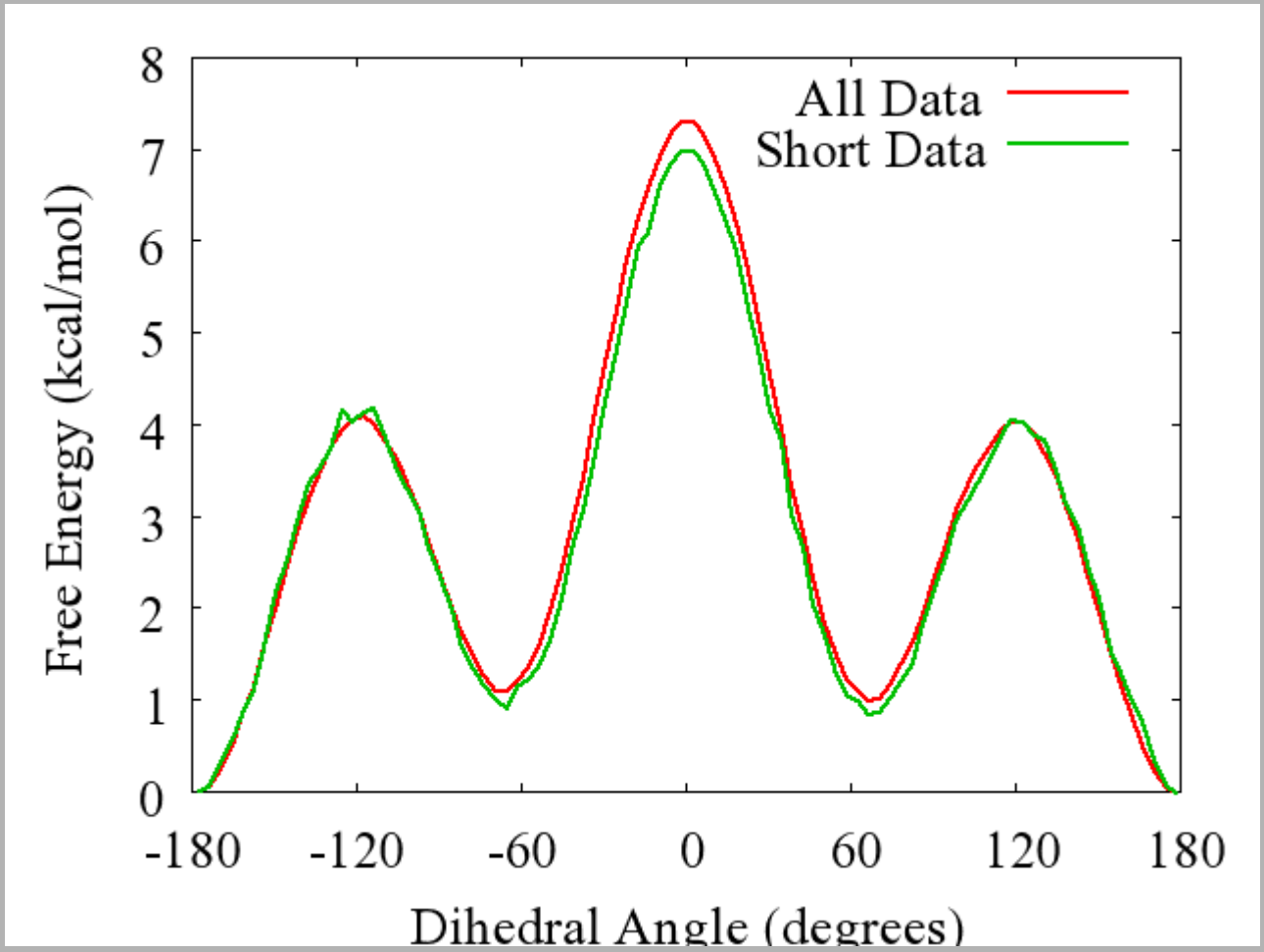
Artifacts of Missing Data



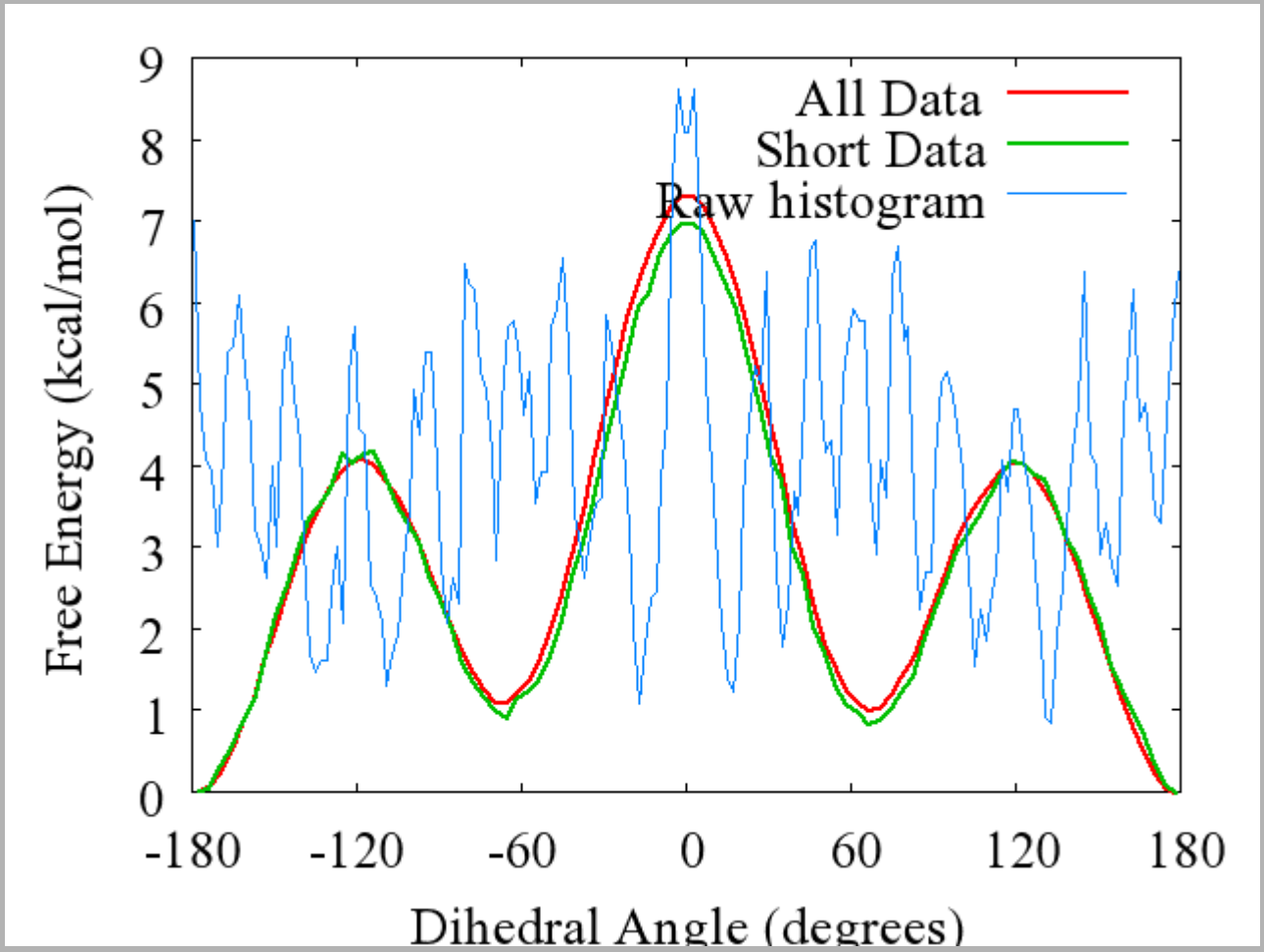
Artifacts of Missing Data



Artifacts of Short Data



Artifacts of Short Data



Computing Other Properties

- Run WHAM to compute F_i
- Compute time series of property T

$$\langle T(x') \rangle = \frac{\sum_{i=1}^{N_{win}} \sum_{j=1}^{N_{points,i}} T_{i,j} \exp[(U_{bias,i}(j) - F_i)/k_B T] \delta x_{i,j}}{\sum_{i=1}^{N_{win}} \sum_{j=1}^{N_{points,i}} \exp[(U_{bias,i}(j) - F_i)/k_B T] \delta x_{i,j}}$$

- subscript i,j means jth point in ith simulation
- $\delta x_{i,j}$ is 1 if $x_{i,j}$ is in the bin corresponding to x', 0 otherwise

Summary

- Calculate PMFs efficiently using umbrella sampling and WHAM
- Choose biasing potential strength and spacing to ensure sampling across entire reaction coordinate
- Can calculate thermodynamic averages of any property using weights calculated from WHAM

Reference

- WHAM Implementation

http://dasher.wustl.edu/alan/wham.tgz

- PDF of this talk

http://dasher.wustl.edu/alan/wham_talk.pdf