High Performance and High Throughput Using GPU Clusters

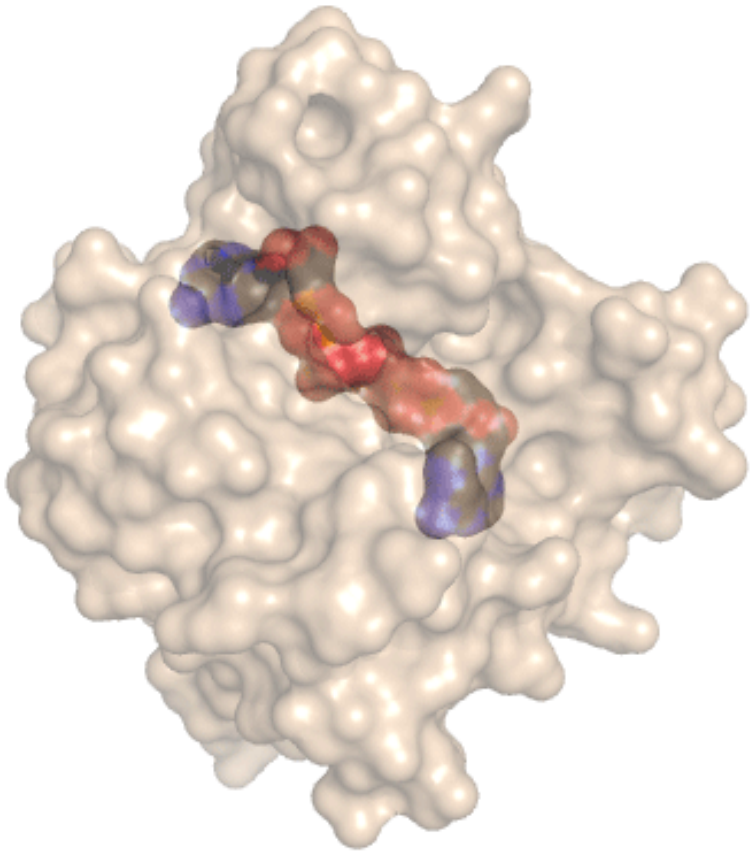
Julie C Mitchell, Spencer Ericksen and Aaron Vassar

University of Wisconsin – Madison

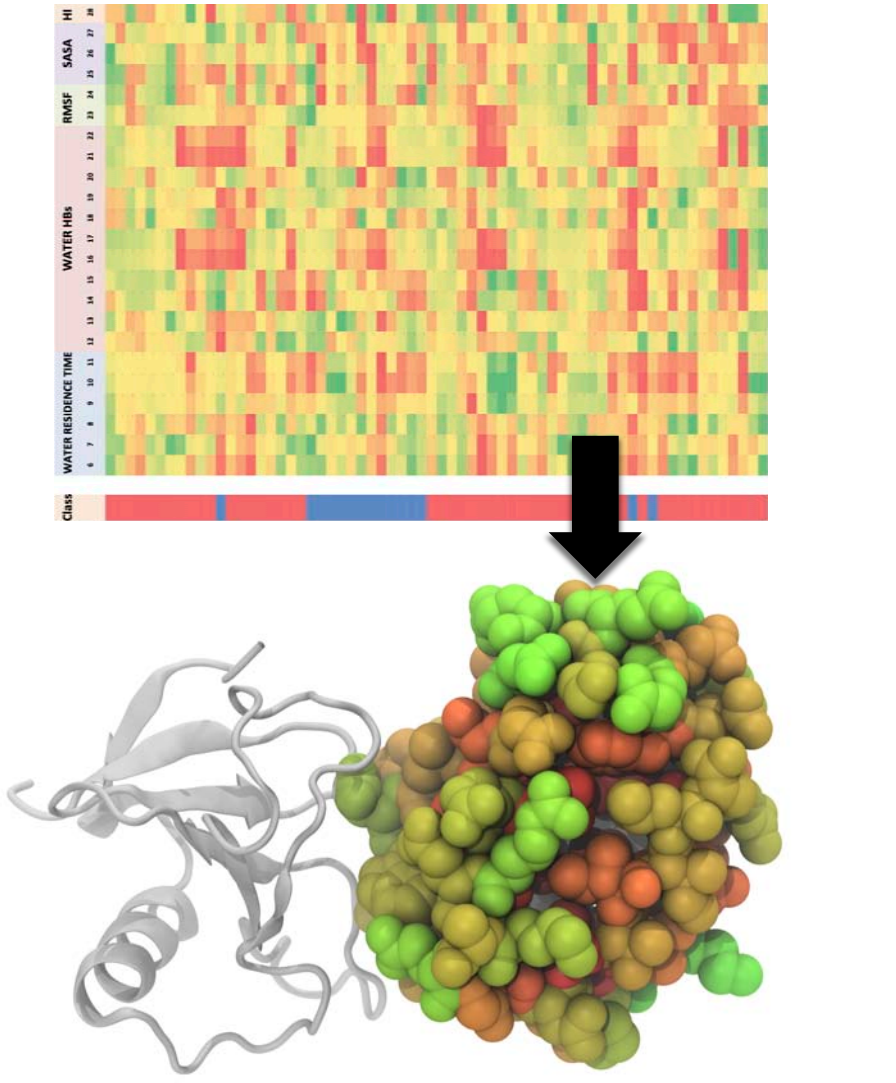
Departments of Biochemistry and Mathematics

Research Goal

- Develop machine learning model to predict protein-protein binding surface residues
- Model uses features generated from molecular dynamics (MD) simulations
- MD is very computationally expensive



Research Procedure Outline



- MD trajectories generated using NAMD
- Feature calculations
  - Trajectory Analysis
    - Number, type of H-bonds
    - Hydration water survival
    - Dynamic SASA
    - Side chain, Backbone RMSF
- Features used to train model
- Model used to predict protein-protein binding surface residues

Molecular Dynamics

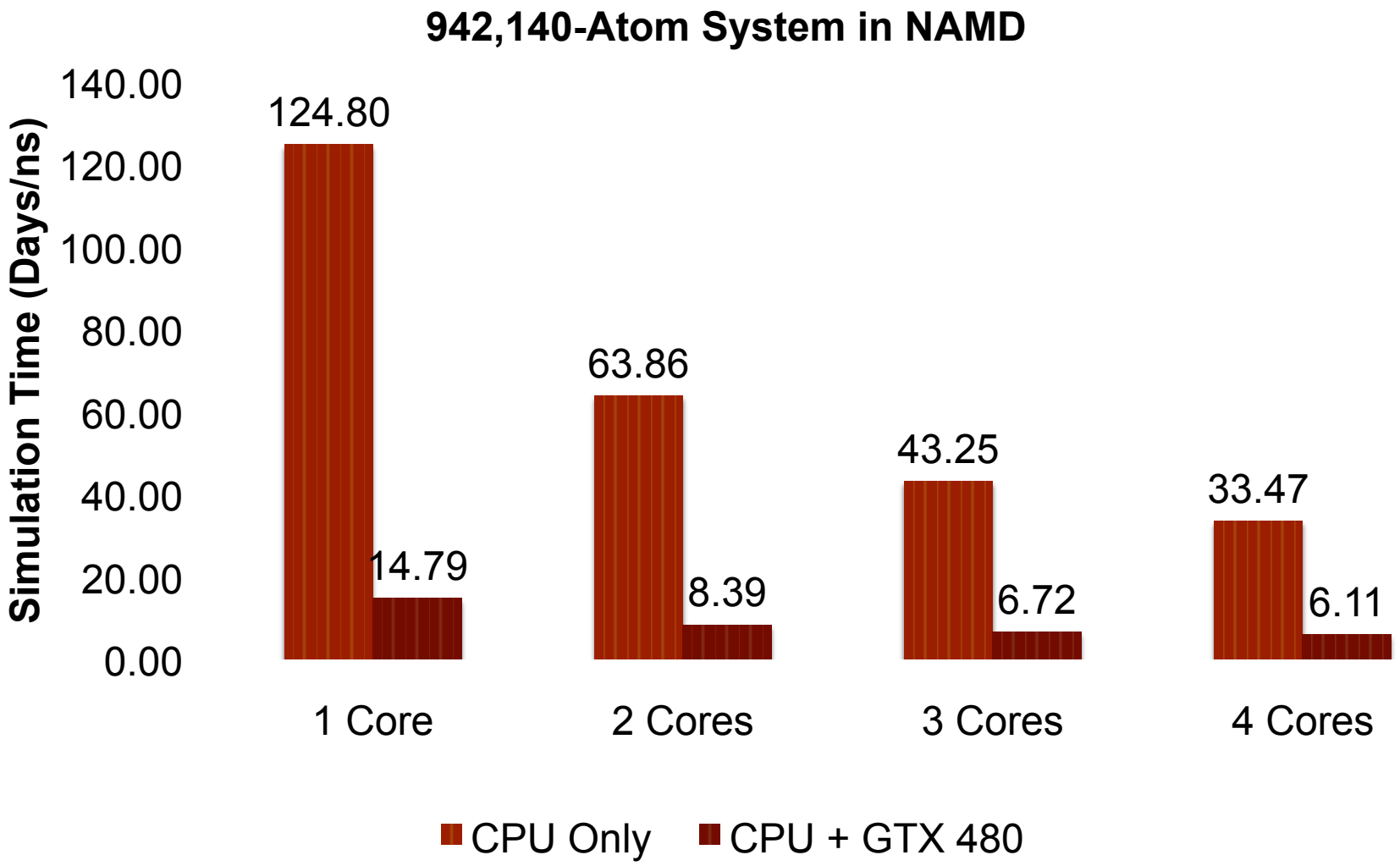
- Computer simulation of all atoms in a system
  - Trajectory tracks movement of atoms through time
- NAMD
  - MD program developed by Univ. of Illinois Theoretical and Computational Biophysics Group
- Certain calculations accelerated with Graphics Processing Unit (GPU)
  - Leads to faster simulations than using CPU alone
- Currently working on 175 systems, 5 simulations per system, for 100ns.

GPU Acceleration of NAMD

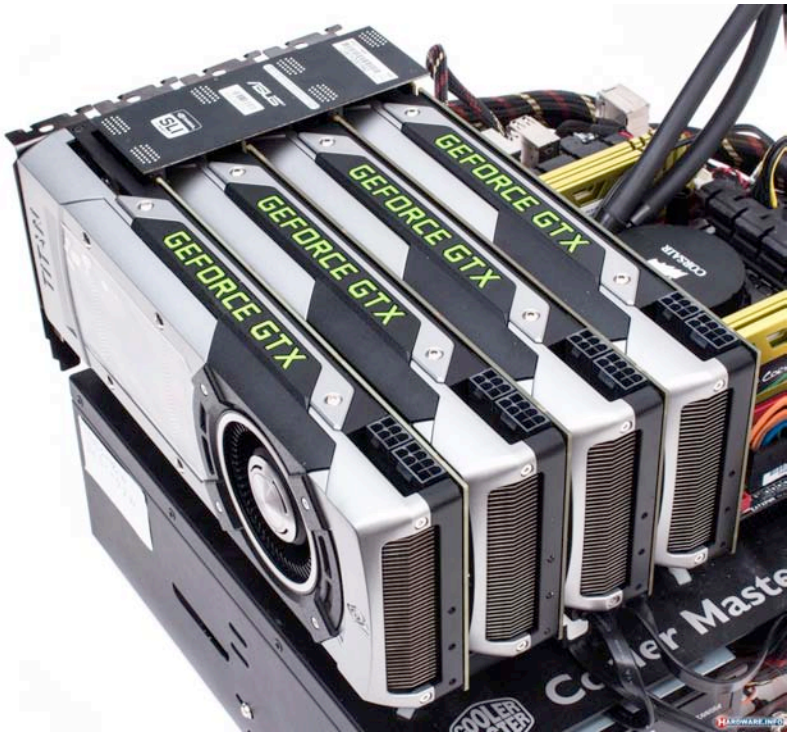
- GPUs contain thousands of specialized cores
- Nvidia CUDA language allows GPU cores to perform highly parallel calculations
- NAMD offloads non-bonded force calculations to GPU
- Massive speedup, as each atom experiences non-bonded forces from

http://www.nvidia.com/content/tesla/images/tesla-k20-series.jpg

GPU Acceleration Increases Simulation Rate 5-10x for Large Systems



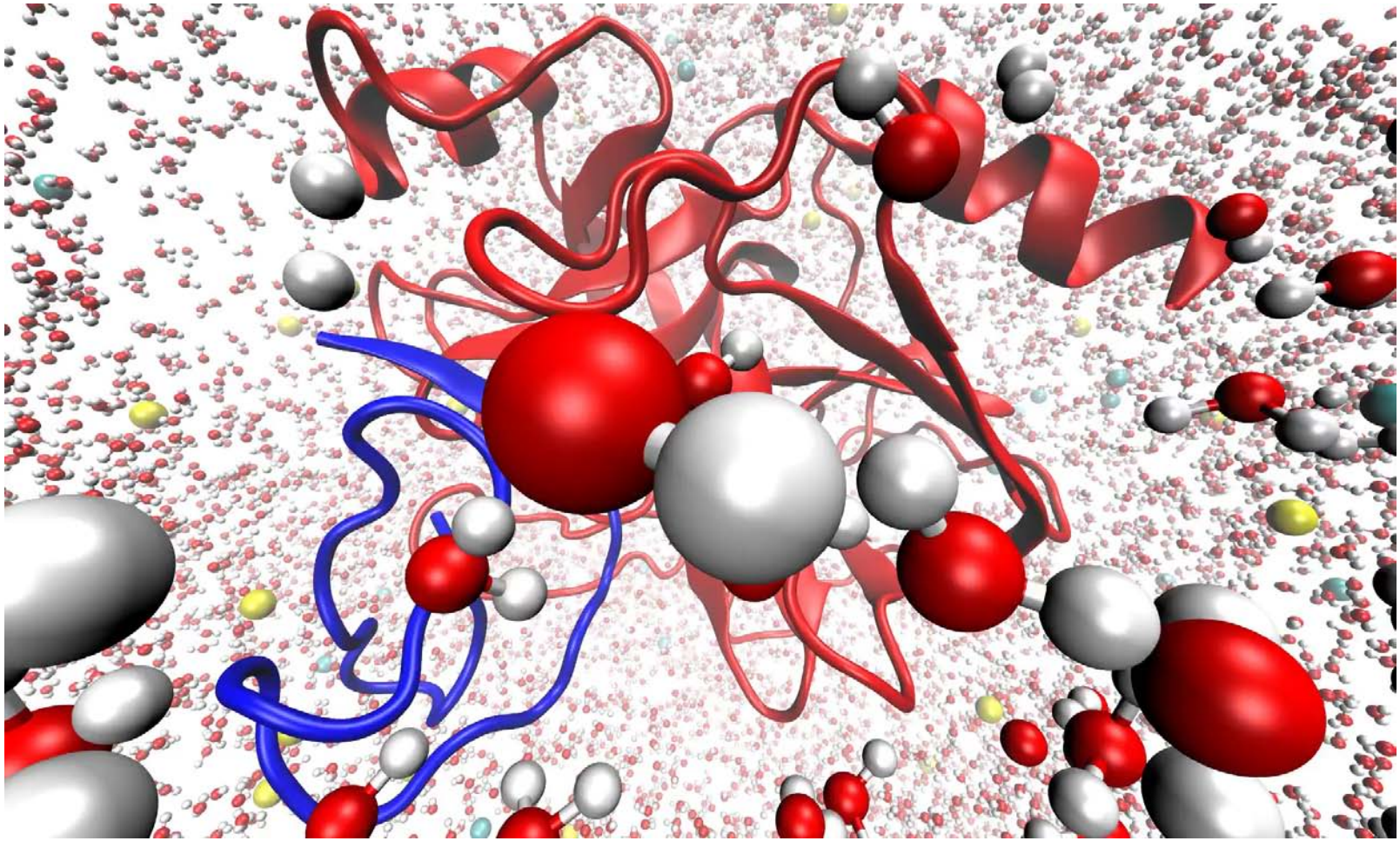
High Throughput MD Using GPU Clusters



- Condor GZK Pool
  - 4 Servers
  - 4 GPUs / Server
  - 16 GPUs Total
- Throughput capacity
  - 16 simultaneous simulations accelerated up to 5x

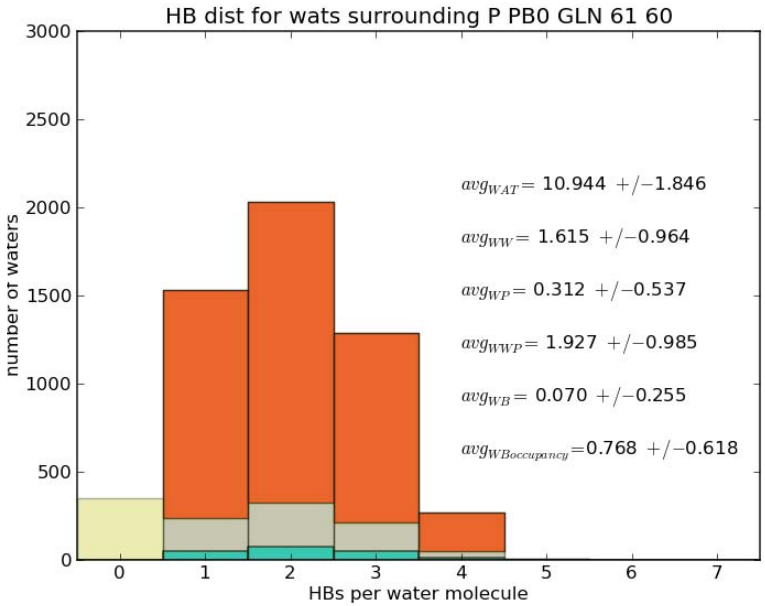
http://content.hwigroup.net/images/products/xl/179054-3.jpg

Tracking Dynamics Down to Individual Water Molecules



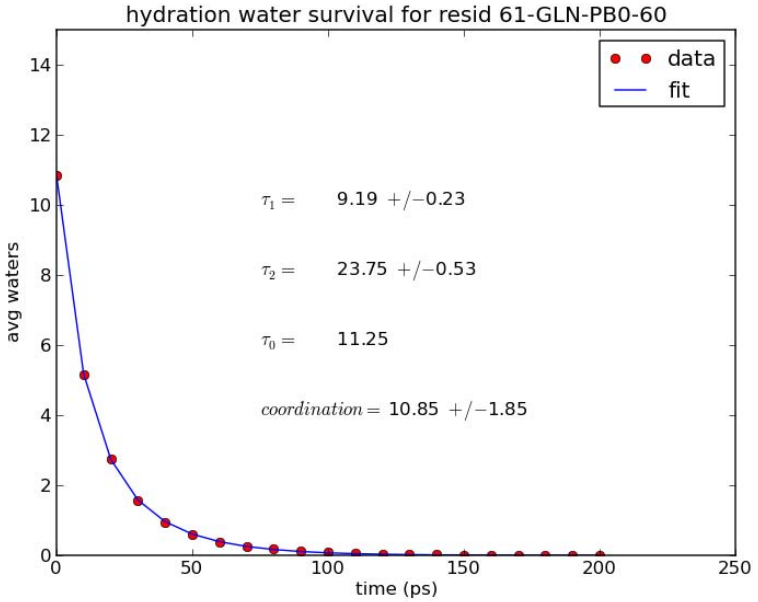
Computing Statistics from MD Trajectories

Computing number and types of HBs formed by hydration water surrounding a particular surface residue.
Hydrophobic patches tend to deprive hydration water of HBs.



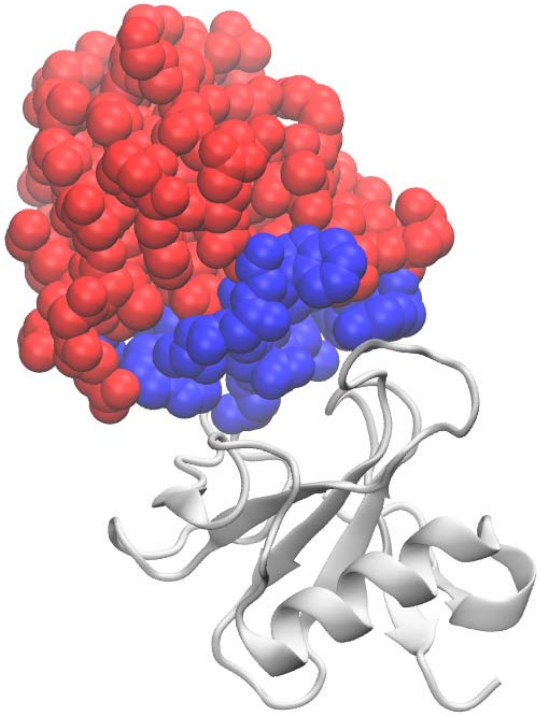
- *WW = wat-wat HB (orange)*
- *WP = wat-prot (brown+cyan)*
- *WWP = wat-wat/prot (orange+brown+cyan)*
- *WB = wat-prot backbone (cyan)*
- *waters with no HBs (yellow)*

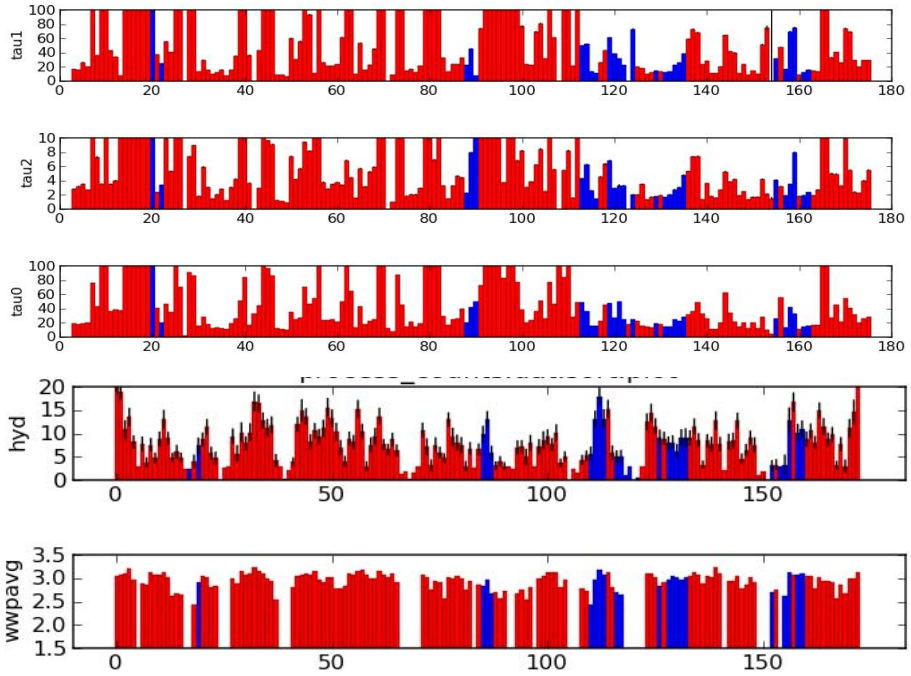
Computing mean residence time for solvation waters around a particular surface residue.



High Throughput Analysis of MD Trajectories

A binary protein complex with one protein rendered as spheres, the other as cartoon. Here the protein rendered in spheres is colored by residue according to whether it is an interface residue (blue) or a non-interface residue (red). Same coloring in the bar plots showing feature values for each of the analyzed residues in the protein.





NOTE: feature calculations are extremely well-suited to Condor. The hydration water features require an analysis of the trajectory that is relatively expensive computationally (requiring a ~1 day per protein residue). However, the analysis is extremely parallelizable since each residue can be analyzed separately.

HTCondor Increases Rate of Data Generation and Analysis

- MD simulations accelerated considerably for our simulations
  - Average ~2 days/ns
- Goal is 100 ns
- HTCondor decreases total time from 1000 days/structure to 200
- Results can be analyzed in parallel
- Features calculated in days, not months