

Course Introduction

Atomistic Modeling of Small-Angle Scattering
Data Using SASSIE-web

September 21-23, 2016

Advanced Photon Source

Argonne National Laboratory, Argonne, IL

ccpsas.org

Scatters & Simulators

Develop a community
of users and developers
to advance atomistic
modeling of soft-
condensed matter

Open Source



CCP-SAS Project
About the project

[home](#) | [about CCP-SAS](#) | [meetings/workshops](#) | [resources](#) | [people](#) | [impact](#)

CCP-SAS - Collaborative Computational Project for advanced analyses of structural data in chemical biology and soft condensed matter

an SI2-CHE cyberinfrastructure project addressing Grand Challenges in the Chemical Sciences

A typical bench scientist purifies and characterizes samples, collects the SAXS/SANS/AUC data, and interprets the results using simplistic models. It is rare that the same individual also has the skills to use advanced atomistic simulation software.

The CCP-SAS project is focused on developing an easy-to-use open-source modeling package that enables users to generate physically accurate atomistic models, calculate scattering profiles and compare results to experimental scattering data sets in a single web-based software suite. This enables a broad range of scattering scientists to access often complicated simulation and scattering analysis methods seamlessly thus providing a significant acceleration to the discovery process.

The objectives of the project include providing a web-based GUI front-end with a high-performance back-end to increase the accessibility of advanced atomistic modeling of scattering data by novice users. Advanced analysis modules and new simulation methods, including implicit solvent models, are also being developed to increase the accuracy of scattering calculations and simulation protocols. The project includes both a testing team and a development team and is built around a continuous development and feedback model.

The project is funded through a joint EPSRC/NSF (UK/US) grant with participants from US and UK universities and X-ray and neutron scattering facilities. It aims to leverage a user and developer community that uses high-performance computing resources on a wide-range of cutting-edge chemical problems. The goal is to build a truly international effort that includes facilities and users from around the world which will be self-sustaining long after the end of this grant. To that end we welcome all interested parties to discuss and join with us at whatever level is appropriate.

Components of CCP-SAS

Projects directly involved with this project:

- GenApp** is a modular framework for multiscale science computations being developed at the University of Texas that enables web & gui applications to be generated from same source code.
- SASSIE** is a NIST NCNR project to provide atomistic simulation of SAS data.
- SCT/SCTPL/HYDRO** is a fortran suite out of Prof. Perkins' lab at UCL to compare molecular models of biological systems to SAS data.
- US-SOMO** is part of the University of Texas UltraScan II suite for the analysis of AUC and SAS data.

Related Projects of Interest:

- The **CCP steering panel** brings together the chairs of the active UK CCP projects. Professor Stephen Perkins represents CCP-SAS on that panel.
- SesView** is an open source collaborative SAS analysis software package based primarily on analytical models.
- The **SAS portal** is a project being developed by the **cinSAS** group.

lower barriers



Goals

PRACTICAL introduction to atomistic
methods to model biological systems

with respect to X-ray & neutron
scattering

STRUCTURE

SAXS/SANS

NR

Course Outline

Mixture of “Lectures”
and hands-on
“Lab” exercises

Student Projects
or
Advanced “Labs”

~10:30 AM & ~2:45 PM : Breaks

Noon: Lunch

WED @ 1 PM: Group Photo

Wednesday 9/21/2016

Time	Lead	Activity	File
9:00 - 9:30 AM	JEC	Course Introduction	lecture_0.pdf
9:30 - 10:30 AM	JEC	Lecture 1: Coordinates to Structure	lecture_1.pdf
10:30 - 10:45 AM		Break	
10:45 - Noon	ALL	Lab 0: Software Installation and SASSIE-web Basics	lab_0.pdf ; basic_osx_linux_commands.pdf ; basic_windows_commands.pdf
Noon - 1:00 PM		Lunch	
1:00 - 2:45 PM	SK/ALL	Lab I: VMD	lab_1.pdf
2:45 - 3:00 PM		Break	
3:00 - 5:00 PM	SK/ALL	Lab II: PSFGEN/NAMD	lab_II_osx_linux.pdf ; lab_II_windows.pdf

Files needed for day 1: [aps_2016_day_1_files.zip](#)

https://sassie-web.chem.utk.edu/training/aps_2016/main.html

Course Outline

Wednesday -- Build Structures & Run MD

Thursday -- Build Systems & Run MC, Calc. SAS

Friday -- Advanced MD/MC & Building & Student Projects

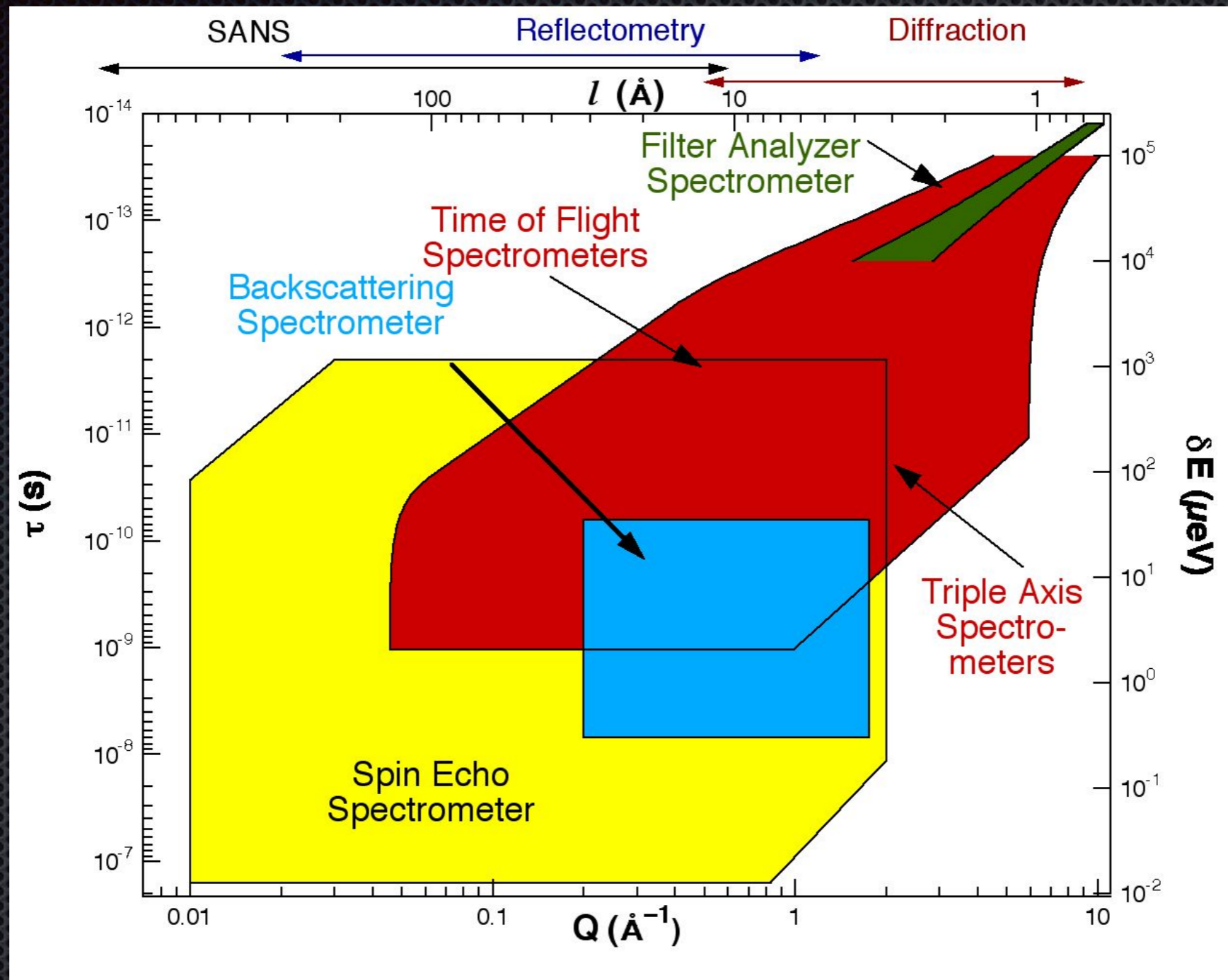
“Student Projects”

Who?

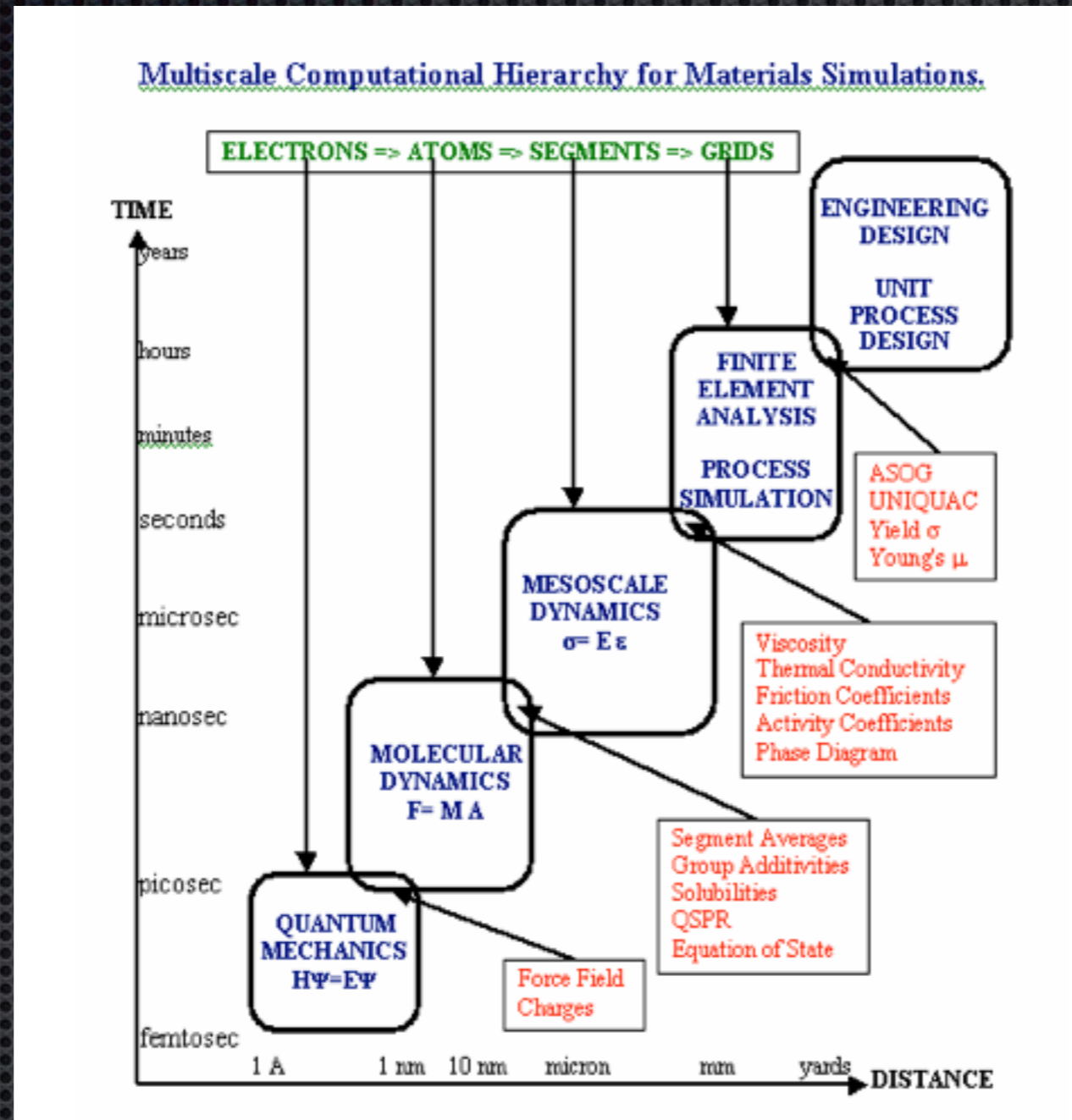
Where?

Project(s)?

Time & Space @ NCNR



Size, Motions, & Theory



Size, Motions, & Theory

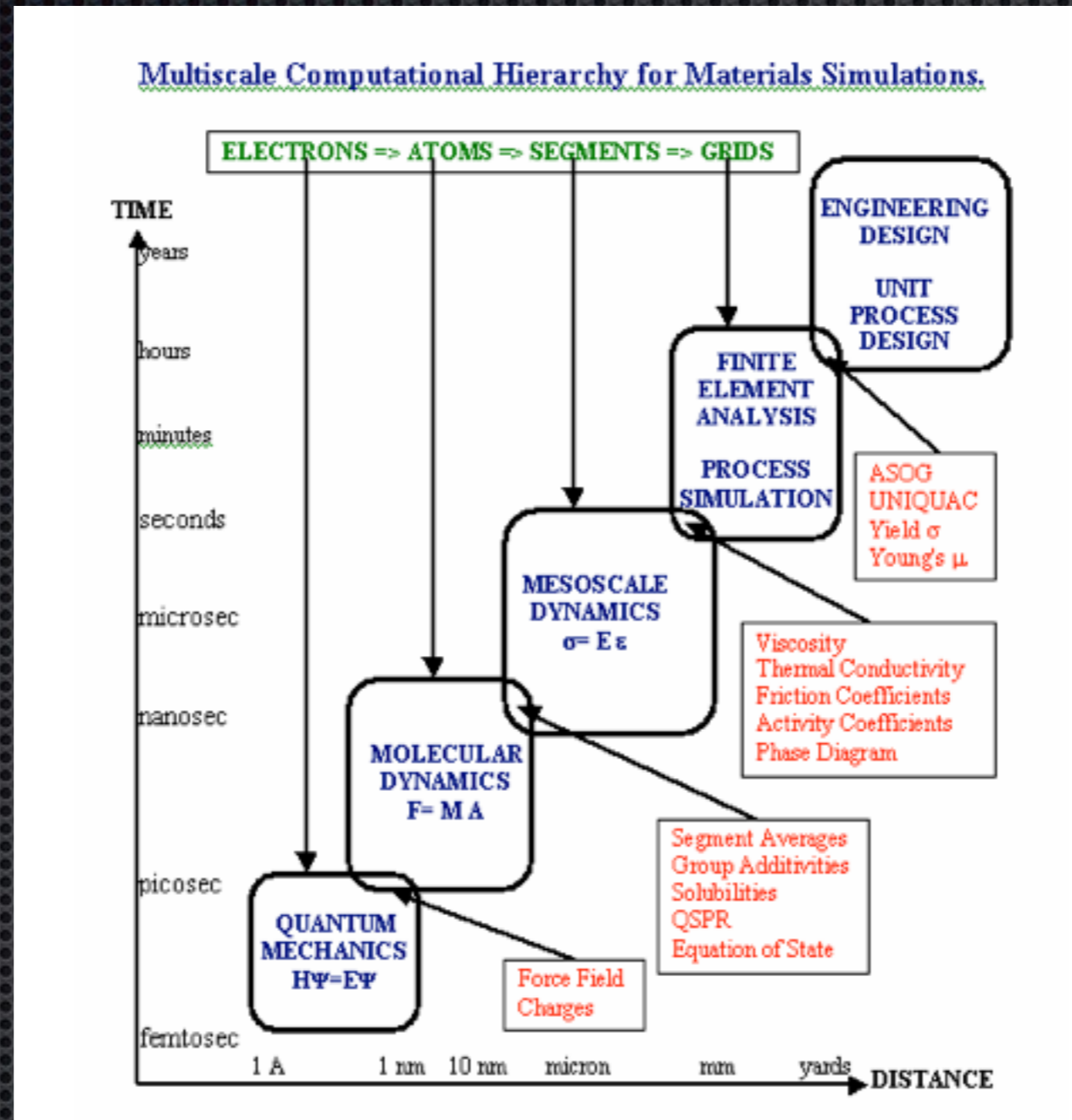
Time scales and amplitudes of protein motions

Presently accessible

Type of Motion	Timescale (s)	Amplitude (Å)	Functionality
Local: <ul style="list-style-type: none">• Atomic fluctuations• Side chain motions	$10^{-15} - 10^{-12}$	< 1	<ul style="list-style-type: none">• Ligand docking flexibility• Ligand diffusion pathways
Medium scale: <ul style="list-style-type: none">• Loop motion• Rigid body motion (e.g. helices)	$10^{-9} - 10^{-6}$	$1 - 5$	<ul style="list-style-type: none">• Active site conformation adaptation• Binding specificity
Large scale: <ul style="list-style-type: none">• Domain motion• Subunit motion	$10^{-6} - 10^{-3}$	$5 - 10$	<ul style="list-style-type: none">• Hinge bending motion• Allosteric transitions
Global: <ul style="list-style-type: none">• Helix-coil transition• Folding/unfolding• Subunit association	$10^{-3} - 10^4$	> 5	<ul style="list-style-type: none">• Hormone activation• Protein functionality

Neutron spectroscopy is unique because it enables the simultaneous measurement of both the time and length scale dependence of dynamics

Size, Motions, & Theory



Goddard: Materials and Process Simulation Center (MSC)

Typical systems

protein: 1745 atoms

protein in solution: 8417 atoms

protein powder:

polycrystalline: 2 proteins

+ 817 waters = 6171 atoms

protein in carbohydrate glass:

~30,068 atoms

protein/dna complex:

25874 atoms

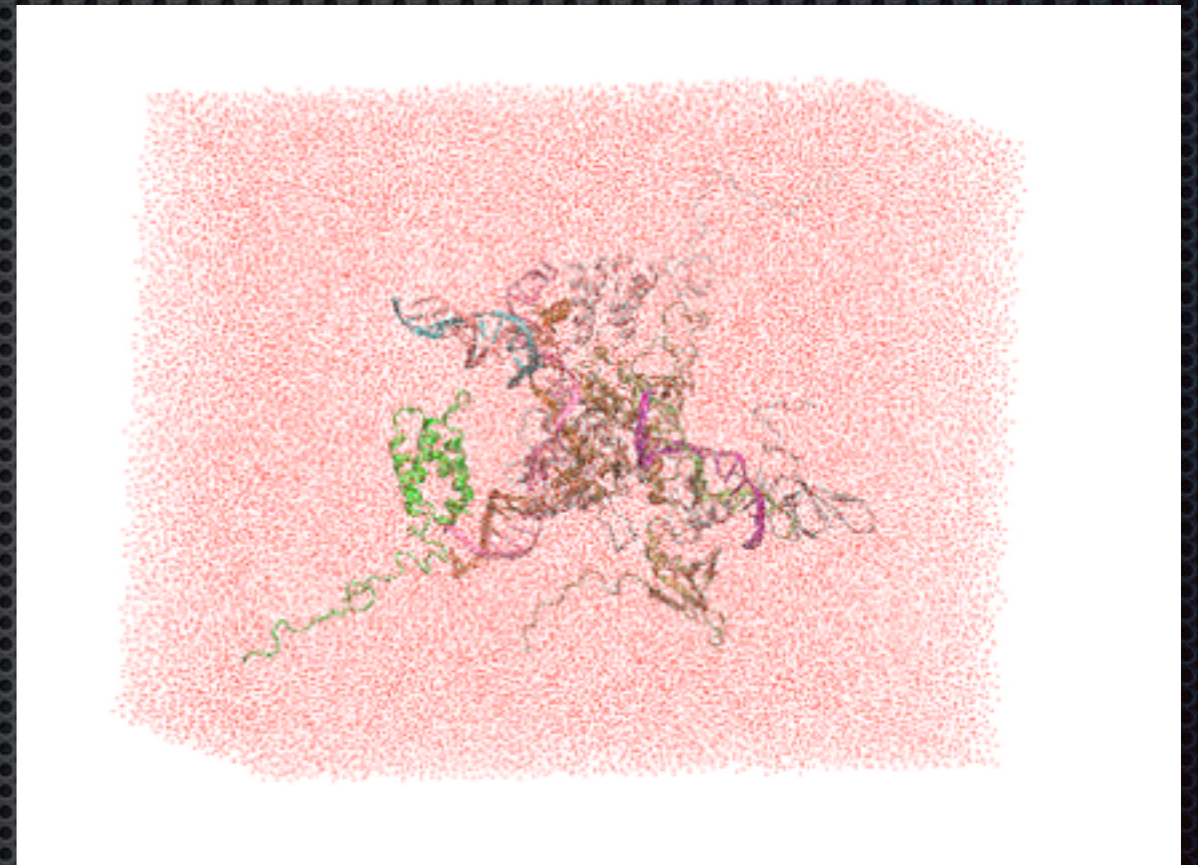
protein/dna complex in solution:

397028 atoms

~19.2 atoms / AA

~136-18 ~ 118 Da / AA

~30 \AA^3 / water



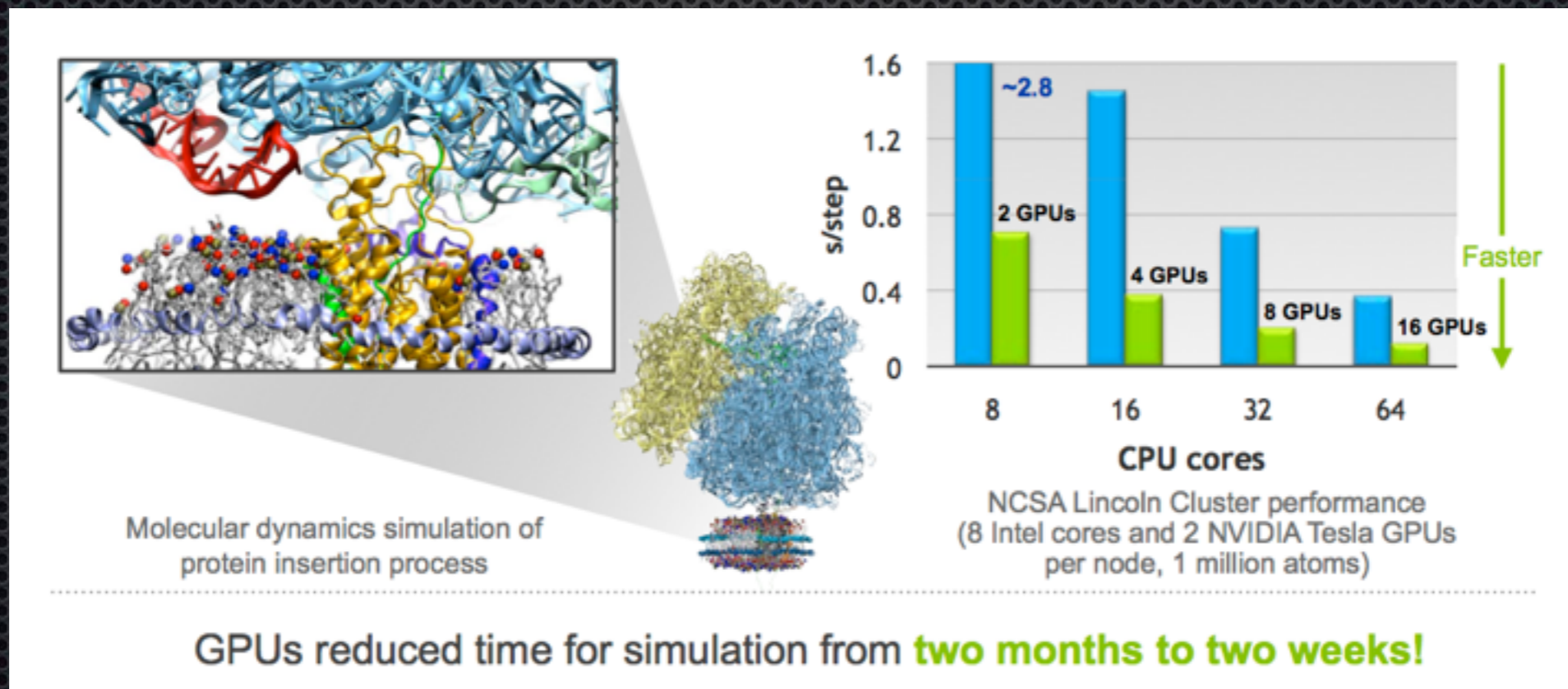
Pushing the limit

2006: STMV (NAMD) 10^6 atoms, 50 ns: “desktop ~ 35 years”, 48 2.4 GHz CPUs, 0.7 ns / day : Structure **14**, 437-449 (2006).

2006: Villin headpiece 2^4 atoms, 500 us: 2^5 CPUs via Folding @ Home: JCP **124**, 164902 (2006). . . . 10,000 trajectories (26 AA + ~5000 waters) . . .

2007: Anton (D.E. Shaw) ~125,000 atoms (limit) ~17 μ s / day for ~24,000 atoms

2011: GPUs “Electrostatics only” calculations: 25-fold faster
MD simulations: 10^6 atoms ~ 4-fold faster . . . up to ~ 3 ns / day



Barriers: BUILD; EQUILIBRATE; PROPAGATE; ANALYZE

What software package(s) and force-fields do I use?

Starting structure?

How do I clean up the structure?

How do I set up a trajectory (time or space)?

How do I calculate scattering observables correctly?

So many options . . .

Bio-simulation (All-atom MD packages):

Amber*
CHARMM*
NAMD*
GROMACS*
GROMOS
LAMPPS*
PINYMD*
HIPPO
GPIUTMD
DL_POLY*
ESPReso
MacroModel*
MACSIMUS
MOLDY
MOSCITO
ProtoMol
TINKER*
MDGrape
Materials Studio (InsightII)*
...

Classical force fields:

Amber*
CHARMM*
CVFF
COSMOS-NMR
GROMACS*
GROMOS
OPLS*
ENZMIX
ECEPP/2
QCFF/PI
UFF
CFF*
MMFF
MM2, MM3, MM4*
XPOL
SIBFA
AMOEBA
VALBOND
DRF90
CG MD*
...

Methods & details*:

Thermostats
Barostats
Electrostatics (PBC)
Polarizability
Implicit Solvent
Langevin Dynamics
Replica Exchange
Parallel Tempering
Steered MD
Free-Energy Calculations
Umbrella Sampling
Normal Mode Analysis
VMD
Pymol
Chimera, O
APBS
Gaussian, Gamess, CPMD
Beowulf, GPU Clusters
Rosetta
PHYRE
Folding @ Home
TIP3, TIP3P, TIP4P, SPC, ST2
...

Let's narrow the options

VMD →

www.ks.uiuc.edu/Research/vmd/

NAMD →

www.ks.uiuc.edu/Research/namd/

CHARMM (MD/MC) →

www.charmm.org/

CHARMM (FF) →

[mackerell.umaryland.edu/
CHARMM_ff_params.html](http://mackerell.umaryland.edu/CHARMM_ff_params.html)

CGMD (Marrink) →

[www.ks.uiuc.edu/Research/vmd/plugins/
cgtools/](http://www.ks.uiuc.edu/Research/vmd/plugins/cgtools/)

MMC (CHARMM & SASSIE) →

[www.smallangles.net/sassie/SASSIE/
SASSIE_HOME.html](http://www.smallangles.net/sassie/SASSIE/SASSIE_HOME.html)

SASSIE-web →

sassie-web.chem.utk.edu/sassie2

COURSE DOCS & FILES →

[sassie-web.chem.utk.edu/training/ccp-
sas_2105](http://sassie-web.chem.utk.edu/training/ccp-sas_2105)

“Student Interests”

“Other MD project interests”:

--> the basics covered here:

topology, parameters --> FF (outline & dev.)

input files for MD runs

T & P control, MTS, NNL, PME, PBC

Analysis of runs (EQ, MTO)

. . . are transferrable

Let's chat during course about setting up your systems as well.

If you wouldn't mind

. . . . **BIO/STRUCTURE** course

Let's not argue about

force-field 1 -vs- force-field 2

water model 1 -vs- water model 2

NAMD -vs- CHARMM -vs- AMBER -vs- ???

classical MD -vs- ab-initio MD

single-chain Nose-Hoover -vs- Nose-Hoover chains

phase space coverage (MD or MC)

etc.

although, you are free to “chat” about your feelings
and questions during labs, breaks

“HOW TO” . . .
Start you on the path

Details ... **there are a lot of them coming!**

At our shoulder in our group . . .

You will go through many of the same problems described in the lectures (and others) in the lab sessions.

WEB: lectures/labs/worked_examples, etc.

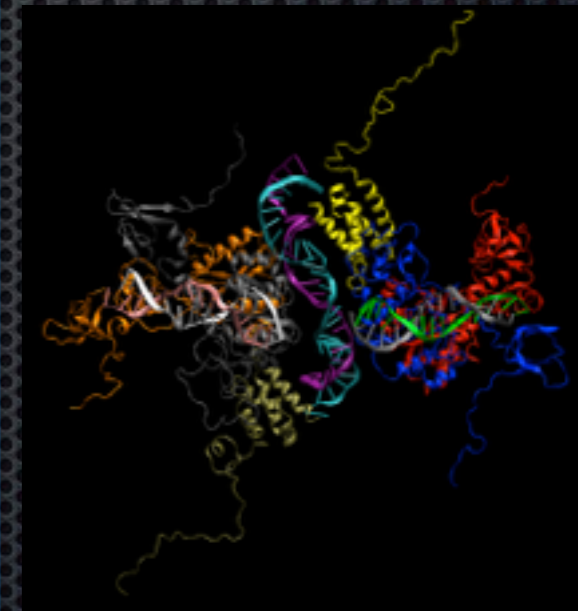
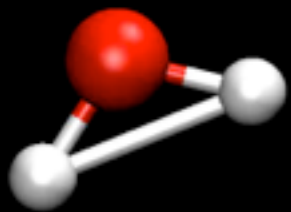
Wednesday

Introduction & background

Oxygen atom to protein-dna complex
in 9 slides

minimize structures in vacuo

*The building scripts you will input are
generic and can be expanded to any
system (FF)*



Thursday

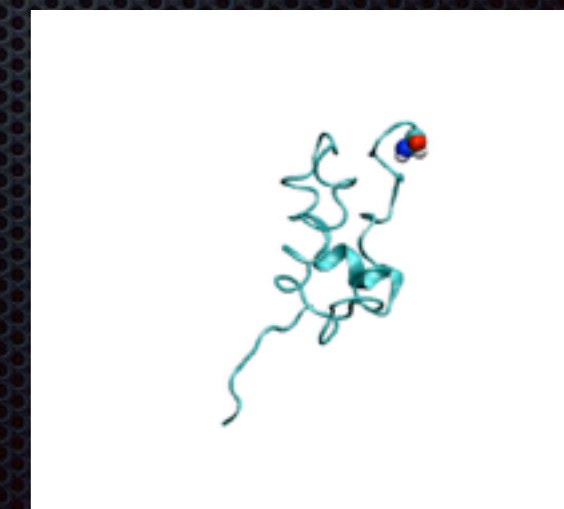
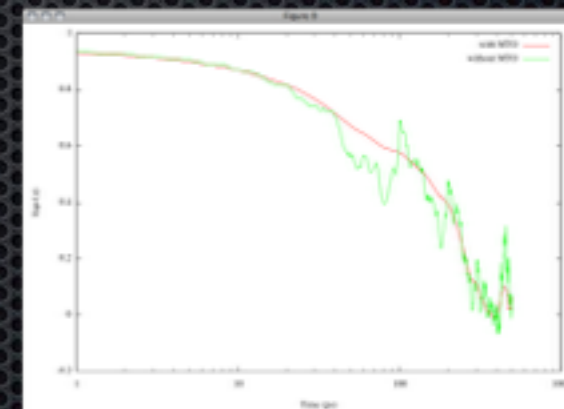
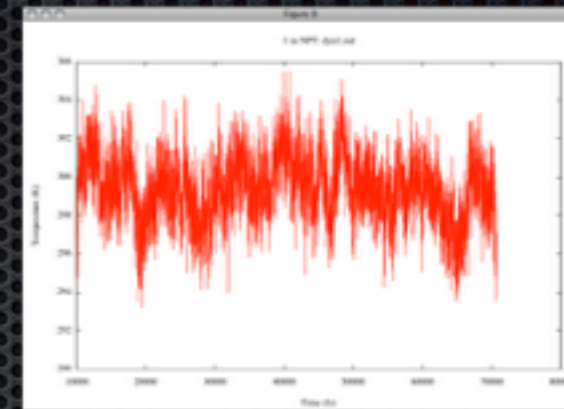
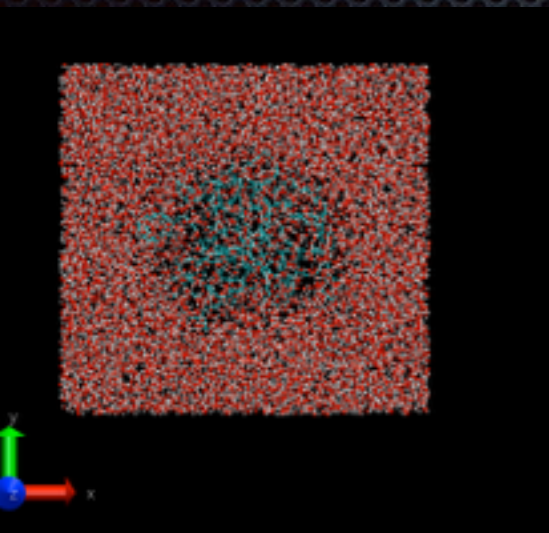
Solvate, PBC, relax, & NVT/NPT dynamics runs

Analyze (check) runs for EQ & analyze trajectory

MMC via SASSIE-web

Calculate SAXS/SANS

Analyze Results



Friday

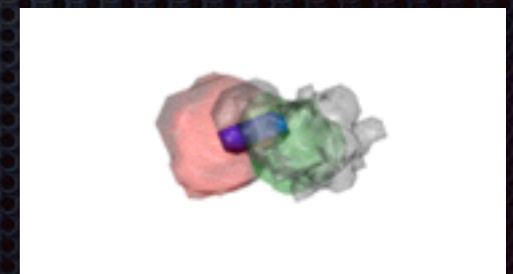
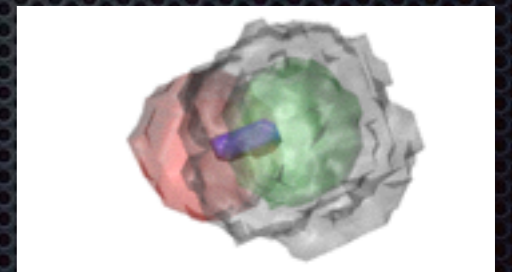
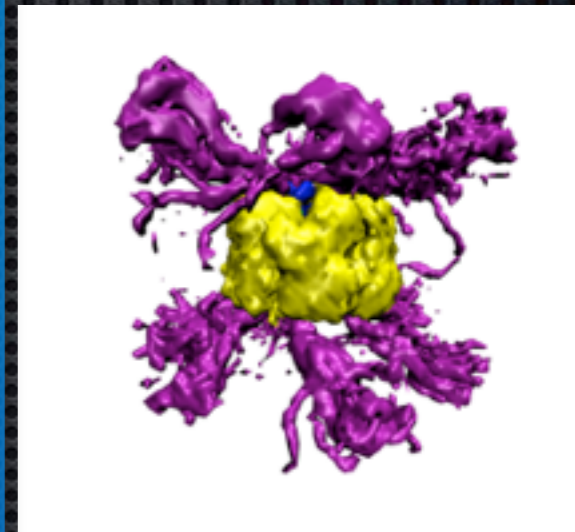
MMC of IDP/ID-NA: Complexes


Calculate SAXS/SANS

Analyze Results

Advanced Building

Student Projects



Tools 

Build 

Interact 

Simulate 

Calculate 

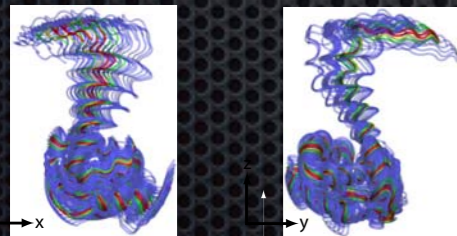
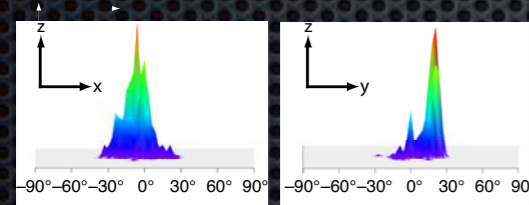
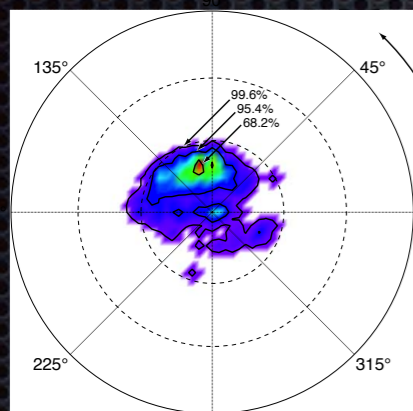
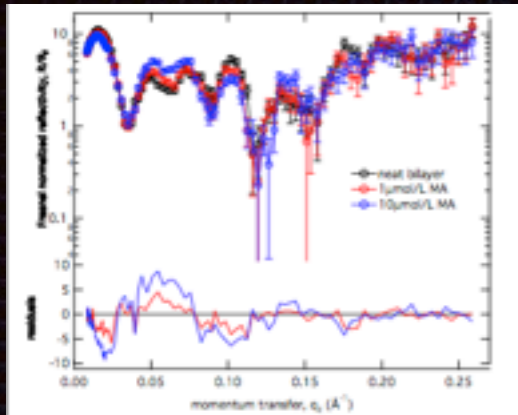
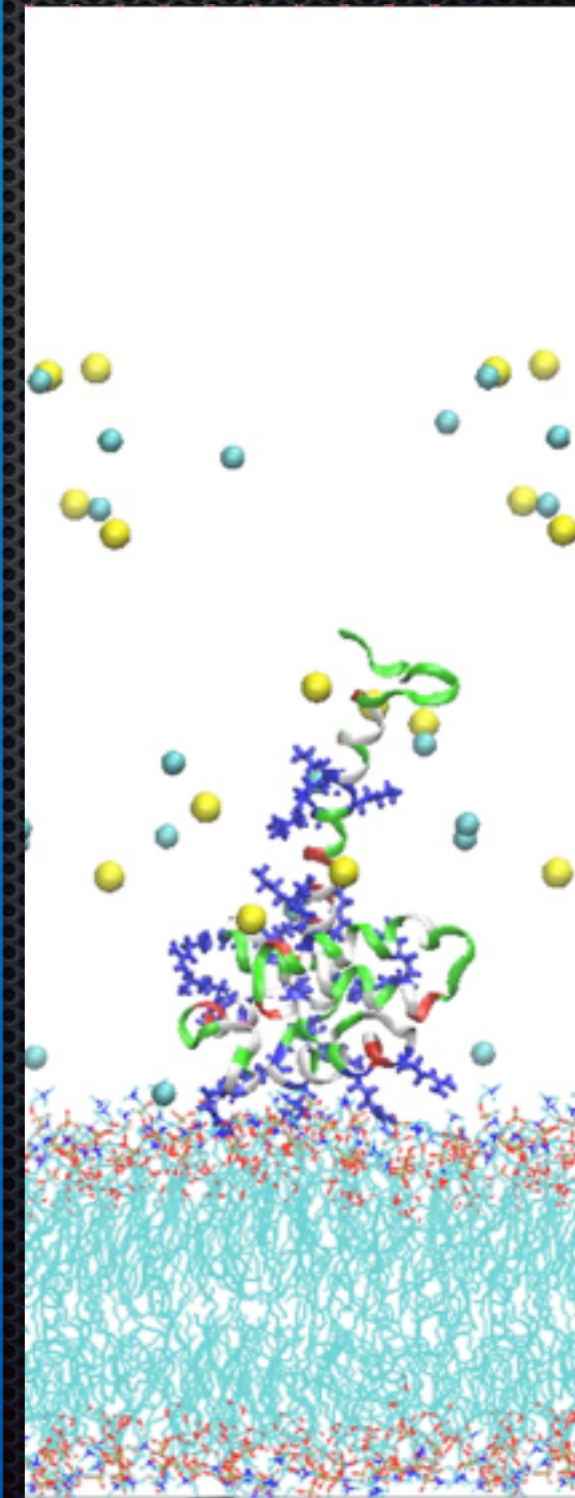
Analyze 

Summary

Steps are Transferrable to almost all MD/MC "Engines"

Complete Structure, Modern FF, Representative Equilibrium

Sample & Converge Observable(s)



Today: “Build Structures & Run”

9:30 AM - 10:30 AM: Coordinates to Structure

~10:30 AM - 10:45 AM: Break

10:45 AM - Noon: Lab 0: Software Installation & SASSIE-web

Noon - 1:00 PM: LUNCH

1:00 PM - 2:45 PM: LAB I: VMD and PDB Scan

2:45 - 3:00 PM: Break

3:00 - 5:00 PM: LAB II: PSFGEN / NAMD