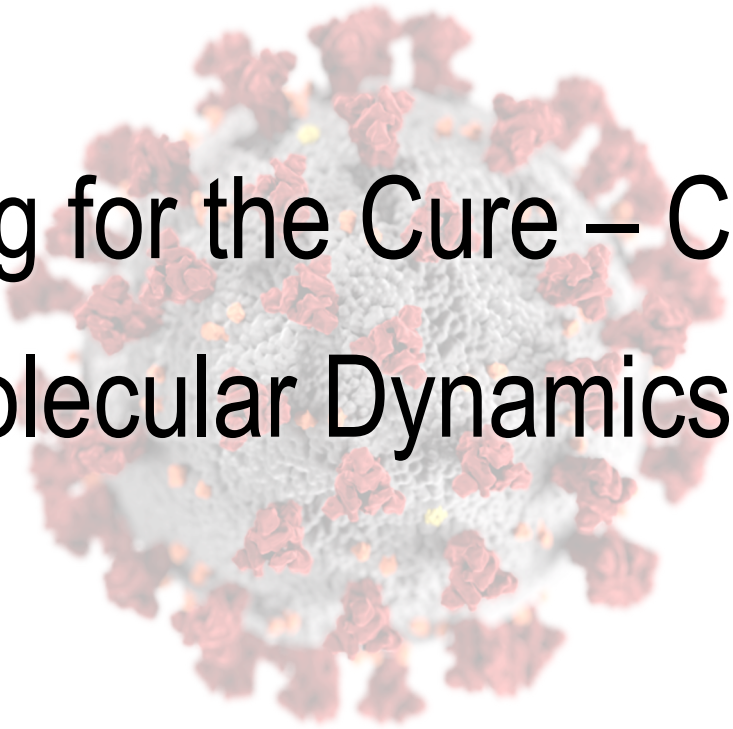


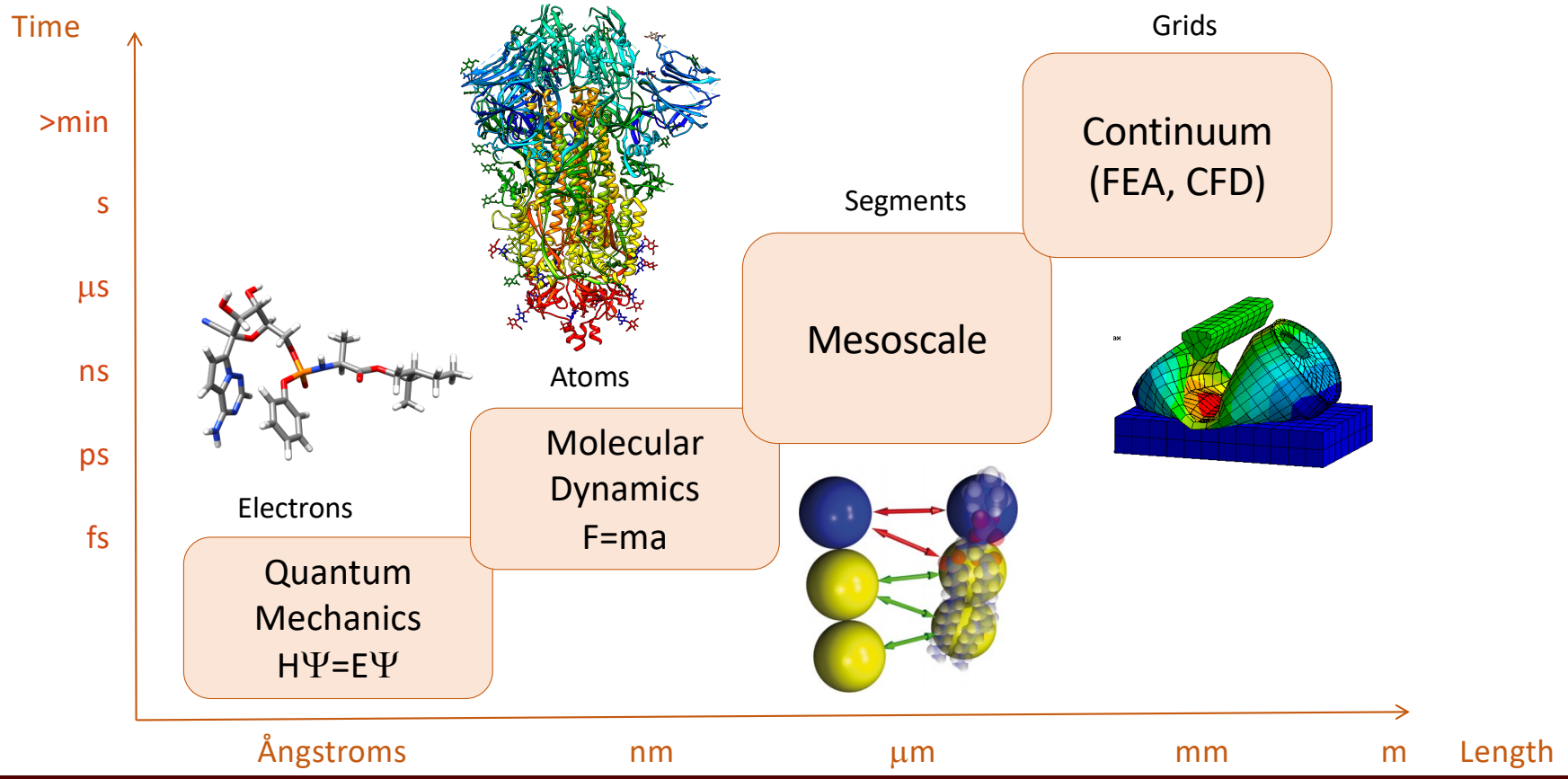
Searching for the Cure – COVID-19

Classical Molecular Dynamics with NAMD

Outline

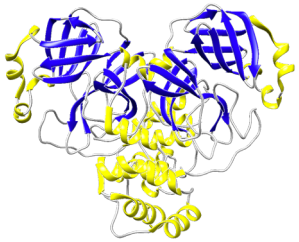
10-10:15	Protein Preparation
10:15-10:45	Hands-on Session 1 – Protein Preparation & Break
10:45-11	Force-Fields & Molecular Mechanics (minimization) lecture
11-11:30	Hand-on Session 2 – Minimization with NAMD & break
11:30-11:45	Molecular Dynamics
11:45-12:15	Hands-on session 3 – Molecular Dynamics with NAMD & Analysis with VMD
12:15-12:30	Wrap-up Lecture

Microscopic ⇔ Macroscopic



Protein Preparation – atomic coordinates

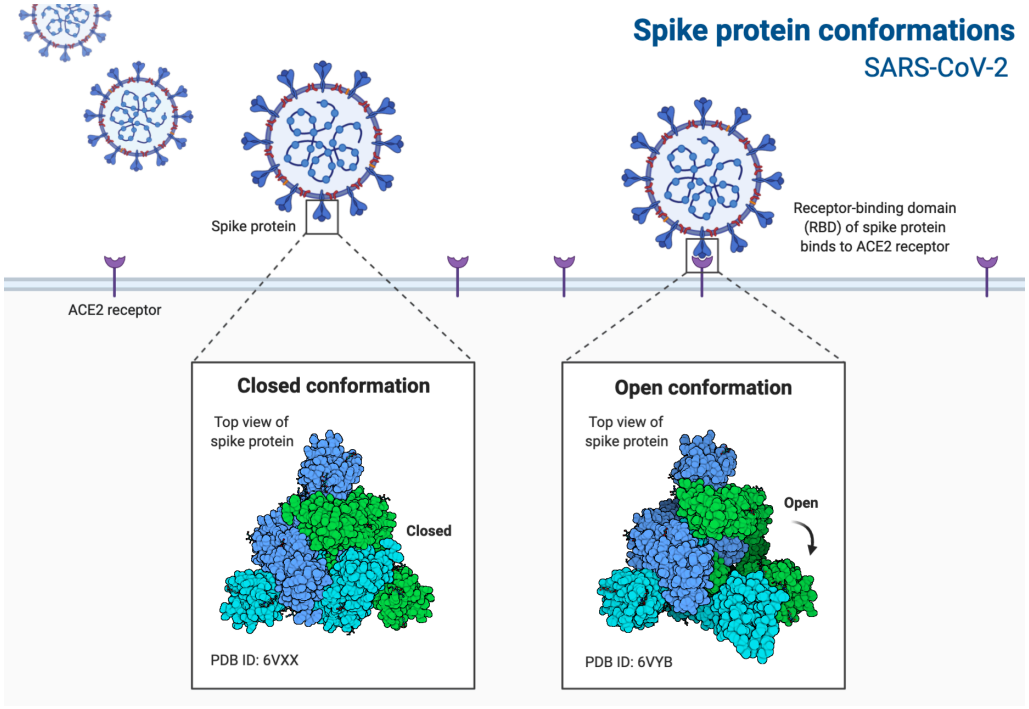
- Crystal Structure
- NMR
- Homology
- cryoelectron microscopy (cryo-EM)



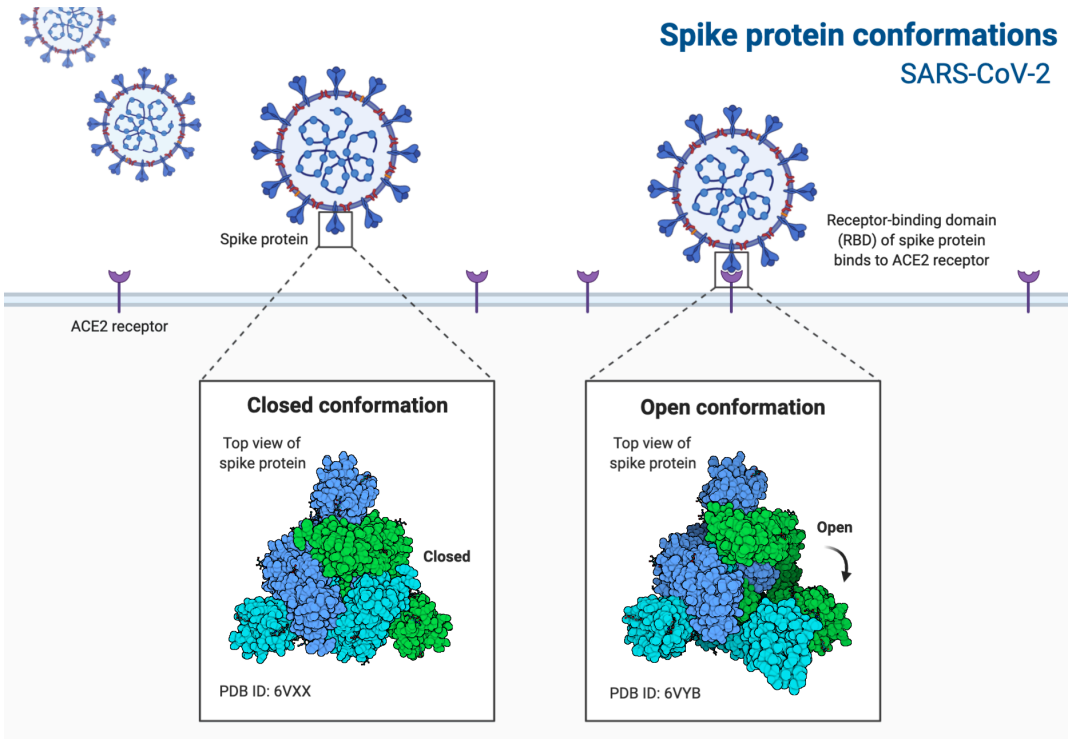
PDB: 6YB7

RCSB Protein Data Bank (PDB)
https://www.rcsb.org/

https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/



Protein Preparation – atomic coordinates



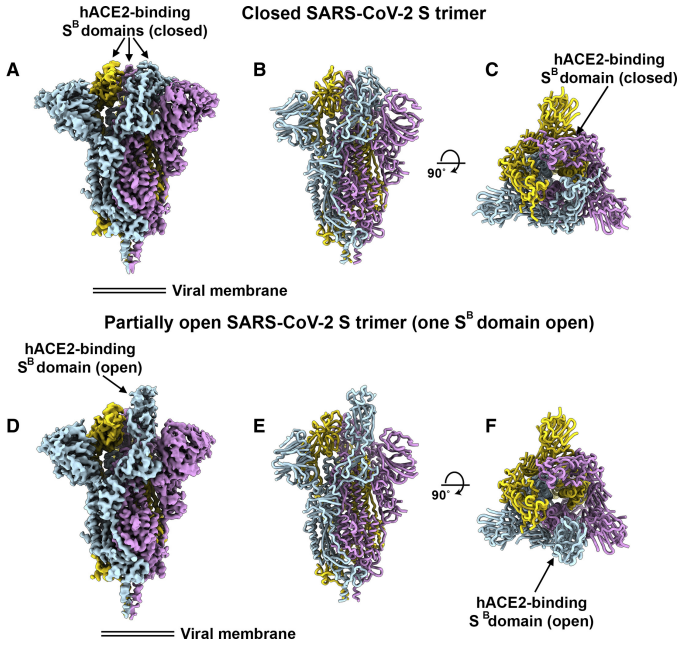
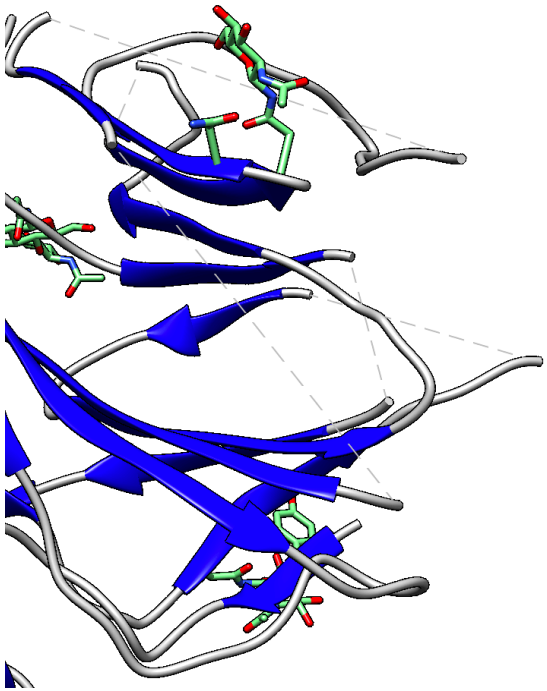


Figure 3. Cryo-EM Structures of the SARS-CoV-2 S Glycoprotein
Walls, A.C., et al https://doi.org/10.1016/j.cell.2020.02.058

6VXX (SARS-CoV-2 spike glycoprotein, closed state)
6VYB (SARS-CoV-2 spike glycoprotein, open state)

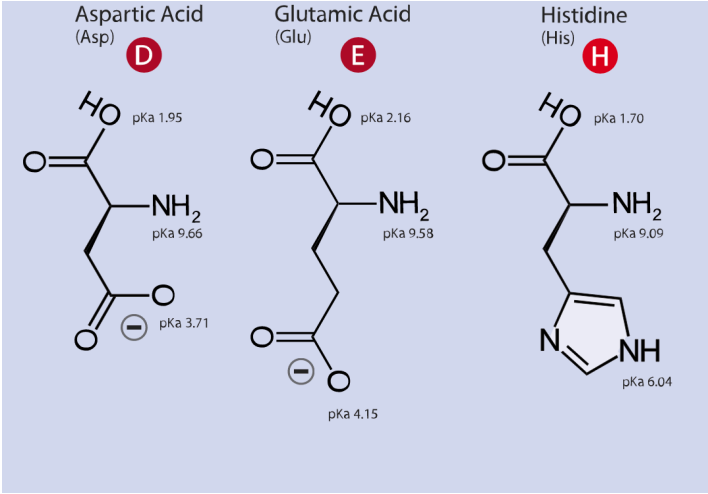
Protein Preparation – Missing atoms

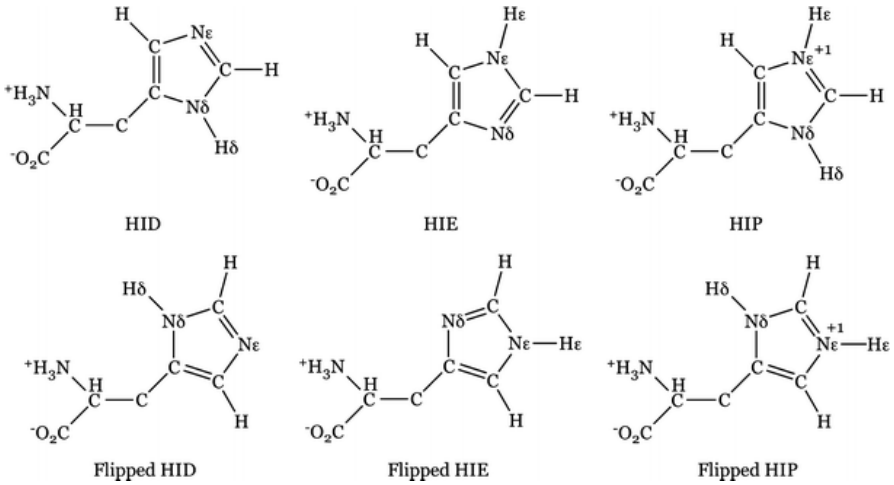
- Missing atoms
  - Hydrogens are not included
  - Entire side chains may be missing
  - There are a number of utilities to fill in missing atoms/sidechain
- Missing segments
  - More complicated to fix
  - Normally requires homology modeling to obtain reasonable results if more than a few residues are missing



Protein Preparation – Protonation states

- ASP, GLU and HIS





Adapted from https://commons.wikimedia.org/wiki/File:Amino_Acids.svg
Dancojocari / CC BY-SA (https://creativecommons.org/licenses/by-sa/3.0)

https://link.springer.com/article/10.1007/s10822-013-9643-9

Hands-on Session 1

Protein Preparation

30 minutes

Molecular Dynamics Software

- AMBER
- CHARMM
- CHARMM Lite (Free)
- Desmond (Free)
- GROMACS (Free)
- LAMMPS (Free)
- NAMD (Free)
- Discovery Studio Suite
- MOE Suite
- Schrödinger Suite

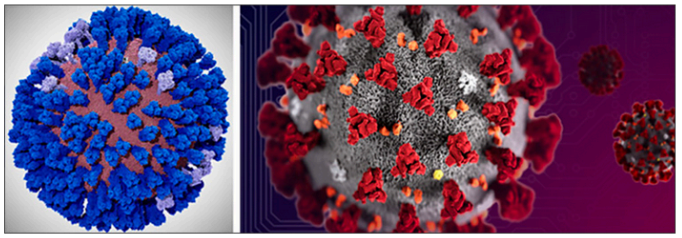
NAMD – Scalable Molecular Dynamics

NAMD, recipient of a **2002 Gordon Bell Award** and a **2012 Sidney Fernbach Award**, is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems. Based on **Charm++ parallel objects**, NAMD **scales** to hundreds of cores for typical simulations and **beyond 500,000 cores** for the largest simulations. NAMD uses the popular molecular graphics program **VMD** for simulation setup and trajectory analysis, but is also file-compatible with AMBER, CHARMM, and X-PLOR. NAMD is distributed **free of charge** with source code. You can **build** NAMD yourself or download **binaries** for a wide variety of platforms. Our **tutorials** show you how to use NAMD and **VMD** for biomolecular modeling.

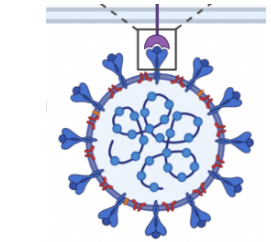
Search all NAMD resources: Search NAMD web site and tutorials Google

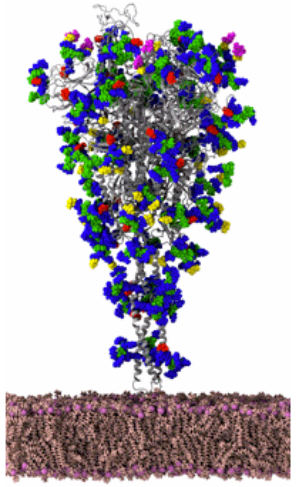
Breaking News

NAMD will be used for groundbreaking coronavirus simulations that will run on the Frontera supercomputer at TACC. The Amaro Lab of UC San Diego is working to build the first complete all-atom model of the SARS-COV-2 coronavirus envelope in order to investigate how the virus interacts with receptors within the host cell membrane. The coronavirus model is anticipated to contain 200 million atoms. The simulations will run on up to 4,000 nodes or about 250,000 processing cores of Frontera.



SARS-CoV-2 spike Simulation





SARS-CoV-2 spike protein of the coronavirus was simulated by the Amaro Lab of UC San Diego on the NSF-funded Frontera supercomputer of TACC at UT Austin. It's the main viral protein involved in host-cell coronavirus infection. Credit: Rommie Amaro, UC San Diego.

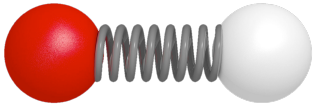
The coronavirus model is anticipated by Amaro to contain roughly 200 million atoms, a daunting undertaking, as the interaction of each atom with one another has to be computed. Her team's workflow takes a hybrid, or integrative modeling approach.

"We're trying to combine data at different resolutions into one cohesive model that can be simulated on leadership-class facilities like Frontera. How we do this is that we basically start with the individual components, where their structures have been resolved at atomic or near atomic resolution, and we have to basically carefully get each of these components up and running and into a state where they are stable. Then we can introduce them into the bigger envelope simulations with neighboring molecules," Amaro said.

The Frontera supercomputer aided efforts of the Amaro Lab on March 12-13, 2020, by running NAMD molecular dynamics simulations on up to 4,000 nodes, or about 250,000 of its processing cores. This is a remarkably large-scale simulation run in itself on Frontera, the #5 top supercomputer in the world and #1 academic supercomputer according to November 2019 rankings of the Top500 organization. Frontera is the leadership-class system in the cyberinfrastructure ecosystem of the National Science Foundation.

From: https://www.tacc.utexas.edu/-/coronavirus-massive-simulations-completed-on-frontera-supercomputer

Molecular Mechanics



- The molecule is considered to be a collection of atoms held together by simple elastic or harmonic forces.

- Force Field - A mathematical expression that describes the dependence of the energy of a molecule on the coordinates of the atoms in the molecule.

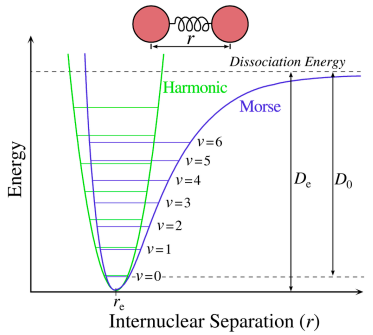
- Force Field Energy Expression:
  - $E = E_{bond} + E_{angle} + E_{torsion} + E_{oop(out\text{-}of\text{-}plane)} + E_{non\text{-}bond} + E_{other}$

Molecular Mechanics – Force Field

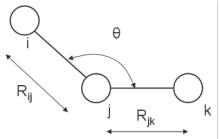
- Choose a Force-Field
  - CHARMM
  - AMBER
  - OPLS
  - etc
- Calibrate
  - against experimental data
  - against benchmark calculations (QM, etc)

CHARMM Force Field

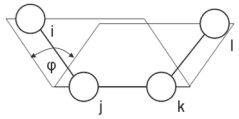
$$U(\vec{R}) = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2$$
$$+ \sum_{\text{Urey-Bradley}} K_{UB}(S - S_0)^2$$
$$+ \sum_{\text{dihedrals}} K_\varphi(1 + \cos(n\varphi - \delta)) + \sum_{\text{impropers}} K_\omega(\omega - \omega_0)^2$$
$$+ \sum_{\substack{\text{non-bonded} \\ \text{pairs}}} \left\{ \varepsilon_{ij}^{\min} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \right\}$$
$$+ \sum_{\text{residues}} U_{\text{CMAP}}(\varphi, \psi)$$



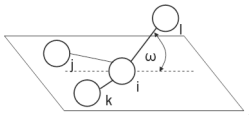
Bond



Angle



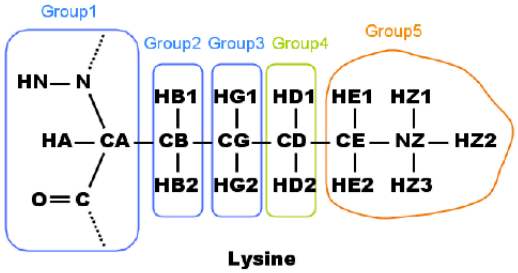
Dihedral



Improper

https://onlinelibrary.wiley.com/doi/10.1002/jcc.21287

Force Field Atom Types



Lysine

```
RESI LYS          1.00
GROUP
ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !  HN-N
ATOM CA   CT1     0.07  !     |   HB1 HG1 HD1 HE1    HZ1
ATOM HA   HB      0.09  !     |   |   |   |   |     /
GROUP                   !  HA-CA--CB--CG--CD--CE--NZ--HZ2
ATOM CB   CT2    -0.18  !     |   |   |   |   |     \
ATOM HB1  HA      0.09  !     |   HB2 HG2 HD2 HE2    HZ3
ATOM HB2  HA      0.09  !   O=C
GROUP                   !     |
```

```
RESI LYS          1.00
GROUP
ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !  HN-N
ATOM CA   CT1     0.07  !     |   HB1 HG1 HD1 HE1    HZ1
ATOM HA   HB      0.09  !     |   |   |   |   |     /
GROUP                   !  HA-CA--CB--CG--CD--CE--NZ--HZ2
ATOM CB   CT2    -0.18  !     |   |   |   |   |     \
ATOM HB1  HA      0.09  !     |   HB2 HG2 HD2 HE2    HZ3
ATOM HB2  HA      0.09  !   O=C
GROUP                   !     |
ATOM CG   CT2    -0.18
ATOM HG1  HA      0.09
ATOM HG2  HA      0.09
GROUP
ATOM CD   CT2    -0.18
ATOM HD1  HA      0.09
ATOM HD2  HA      0.09
GROUP
ATOM CE   CT2     0.21
ATOM HE1  HA      0.05
ATOM HE2  HA      0.05
ATOM NZ   NH3    -0.30
ATOM HZ1  HC      0.33
ATOM HZ2  HC      0.33
ATOM HZ3  HC      0.33
GROUP
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB CA   CG CB   CD CG   CE CD   NZ CE
BOND N  HN   N  CA    C  CA
BOND C  +N   CA HA   CB HB1  CB HB2  CG HG1
BOND CG HG2  CD HD1  CD HD2  CE HE1  CE HE2
DOUBLE   O  C
BOND NZ HZ1  NZ HZ2  NZ HZ3
IMPR N -C CA HN  C CA +N O
DONOR HN N
DONOR HZ1 NZ
DONOR HZ2 NZ
DONOR HZ3 NZ
ACCEPTOR O C
IC -C   CA   *N   HN    1.3482 123.5700  180.0000 115.1100  0.9988
IC -C   N    CA   C     1.3482 123.5700  180.0000 107.2900  1.5187
IC N    CA   C    +N    1.4504 107.2900  180.0000 117.2700  1.3478
IC +N   CA   *C   O     1.3478 117.2700  180.0000 120.7900  1.2277
IC CA   C    +N   +CA   1.5187 117.2700  180.0000 124.9100  1.4487
IC N    C    *CA  CB    1.4504 107.2900  122.2300 111.3600  1.5668
```

http://www.ks.uiuc.edu/Training/Tutorials/science/topology/topology-tutorial.pdf

How to assign the atom types/generate a topology/parameter file

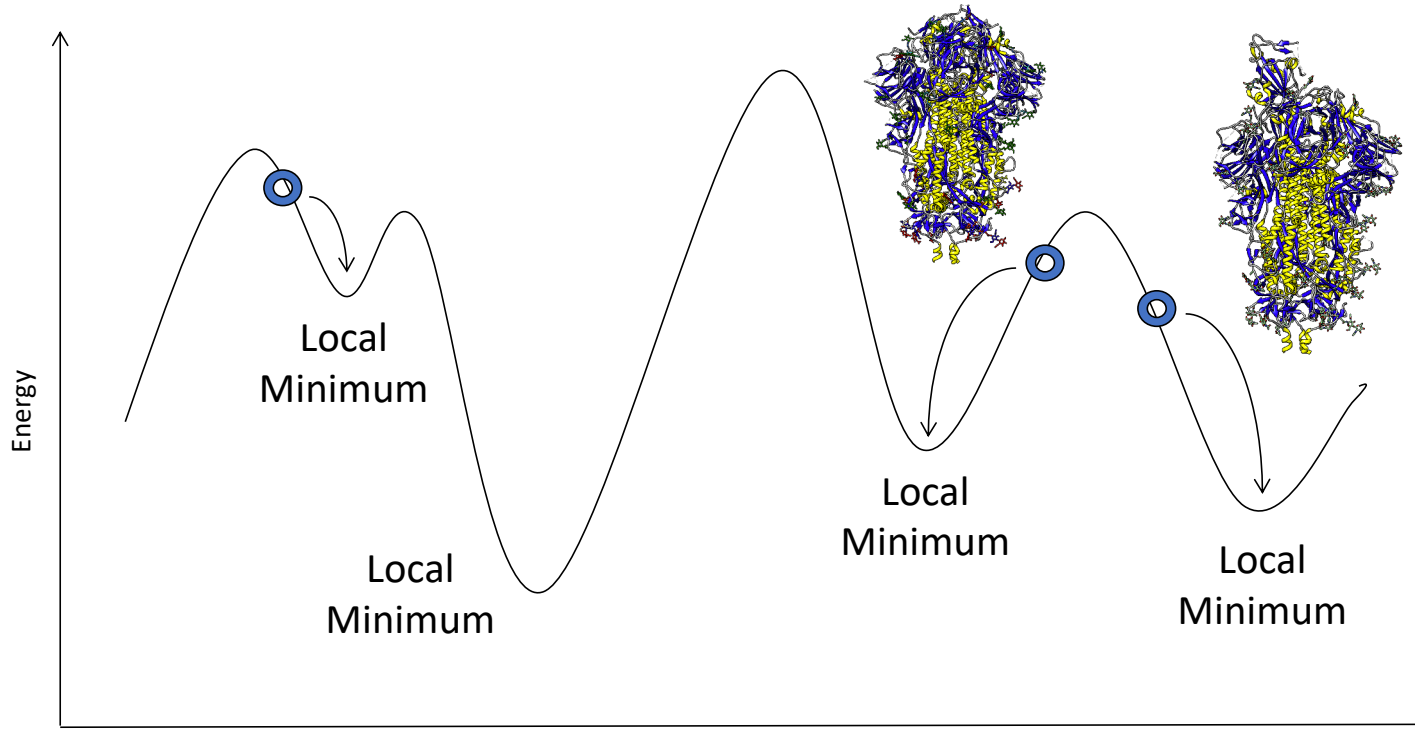
Use a utility

- VMD Plugins
- AmberTools: http://ambermd.org/#AmberTools
- Chimera
- Schrödinger
- MOE
- Discovery Studio
- LAMMPS utilities: http://lammps.sandia.gov/doc/Section_tools.html
- GROMACS: pdb2gmx
- etc

Molecular Mechanics - Minimization

Force Field Energy Expression:

$$E = E_{bond} + E_{angle} + E_{torsion} + E_{oop(out\text{-}of\text{-}plane)} + E_{non\text{-}bond} + E_{other}$$



Solvation & Ions

- Explicit Solvation
  - Include explicit solvent molecules.
  - Computationally expensive
- Implicit Solvation
  - estimate the effect of the solvent molecules using an implicit solvation model.
  - Gives reasonable results for bulk polar solvent effects.
  - Fails if the solvent molecule strongly interacts with the solute (ie strong H-bonding, etc)
- Ions
  - Common ions used: Na^+, K^+, Mg^{2+}, Cl^-
  - Required to balance the charge of the protein (Overall charge of the periodic cell needs to be zero)
  - Ion concentration should be added at reasonable concentrations for the system of interest

Hands-on Session 2

Minimization with NAMD

30 minutes

Molecular Dynamics

$$F = ma = m\frac{d^2r}{dt^2} \; ; \; F = -\frac{\partial V}{\partial r}$$

Newtons equations

- The potential is approximated by an empirical function force field that is fitted to approximately reproduce known interactions
- Applicability is limited by the availability of parameterization
- Generally, the connectivity of atoms cannot change during the simulation
- Generally, not suitable for reaction mechanisms
- Can predict relative energies of different conformational states of material
- And much more

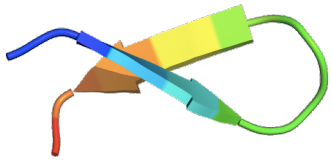
Molecular Dynamics

- Molecular Dynamics Variations
  - Constant Volume - Constant Temperature (NVT)
  - Constant Volume - Constant Energy (NVE)
  - Constant Pressure – Constant Temperature (NPT)
- construct a set of velocities based on the ensemble being used
- Velocities satisfy the Maxwell-Boltzmann distribution

- Each run will start with a different random seed
  - Allow atoms to move for one time-step $r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t)$
  - Calculate the force on the atoms - forcefield
  - Calculate the acceleration F=ma
  - Calculate the new velocity $v(t+\delta t) = v(t) + \frac{1}{2}\delta t(a(t) + a(t+\delta t))$
  - Calculate the new position $r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t)$
  - Repeat for as many time steps as desired

Molecular Dynamics

- Choosing a time step
  - Your time step should be a factor of 10 smaller that the fastest process in your system.
    - Molecular motions such as rotations and vibrations are on the order of 10^{-11} - 10^{-14} s
      - Therefore, a time step of 1 fs (10^{-15} s) or less must be used for most systems.
  - You can increase your time step by restricting the fastest processes
    - SHAKE or RATTLE algorithms restrict the vibrational motion of the molecule of interest
      - Therefore, a time step of 2-3 fs can be used with the SHAKE or RATTLE algorithm
      - There are some modified shake algorithms that claim they are stable up to time steps of 8 fs
- Most simulations are on the order of picoseconds (10^{-12} s) or nanoseconds (10^{-9} s)
- Protein folding tripzip2 (12-residue protein) folds on the order of 2.5 ms (10^{-6} s)



Hands-on Session 3

Molecular Dynamics with NAMD & Analysis with VMD

30 minutes

Illustrative Applications & Training

NAMD: http://www.ks.uiuc.edu/Highlights/
NAMD: http://www.ks.uiuc.edu/Training/

CHARMM https://www.charmm.org/charmm/showcase/

Schrödinger https://www.schrodinger.com/training

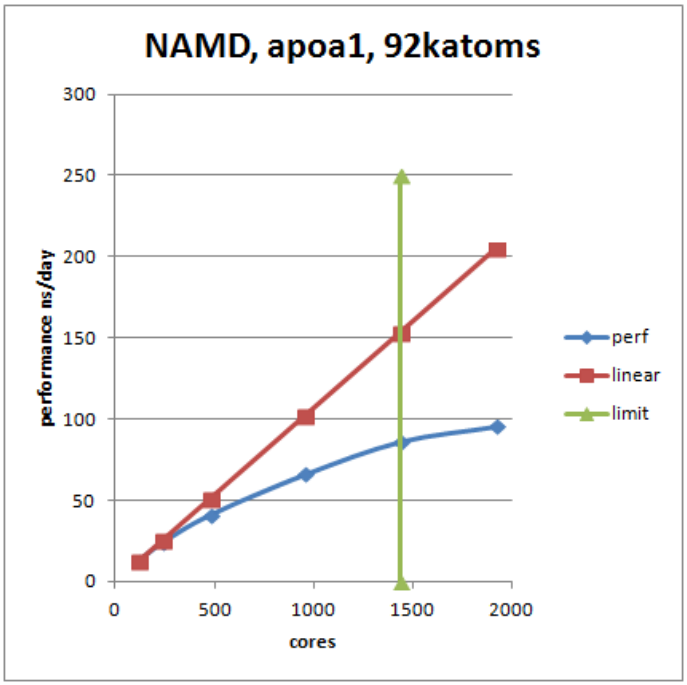
MOE https://www.chemcomp.com/Research-Current_Journals.htm

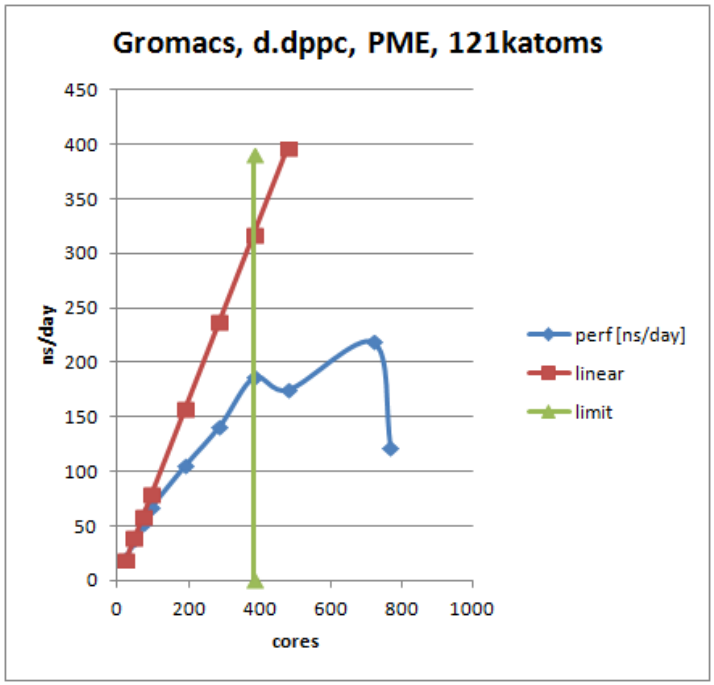
LAMMPS http://lammps.sandia.gov/movies.html

Scalability

Will your calculation run with more than 1 core? How many cores should I use?
Does it require shared memory? How much memory do I need? Disk space?

https://research.csc.fi/





Benchmarking to determine required resources

- Get guidance from group members if possible: Ask group members if someone has tested the efficiency of the code and level of theory on the machine that you are using and hope that they are willing to share their knowledge.
- See if there is guidance in the program manual
- Run your own “benchmarking” calculations
- Run short representative jobs using increasing number of cores and calculate the speed-up using the **WALLTIME**
- Caveats
  - Overhead may dominate the walltime for very short jobs
  - Competition with other users on the system may skew your results
  - Multi-node job timing can be effected by job placement
- IMPORTANT: Just because all of the cpus are busy for your parallel job does NOT mean that you are running an efficiently!
- Seek guidance from HPRC (help@hprc.tamu.edu)

Need Help? Contact the HPRC Helpdesk

Website: hprc.tamu.edu
Email: help@hprc.tamu.edu
Telephone: (979) 845-0219

Help us, help you -- we need more info

- Which Cluster (Terra, Ada, Curie)
- NetID (NOT your UIN)
- Job id(s) if any
- Location of your jobfile, input/output files
- Application used if any
- Module(s) loaded if any
- Error messages
- Steps you have taken, so we can reproduce the problem