Practical considerations in running simulations in NAMD

James C. (JC) Gumbart
Georgia Institute of Technology

Computational Biophysics Workshop | DICP | July 11 2018

PDB files *provide structure and starting positions of atoms*

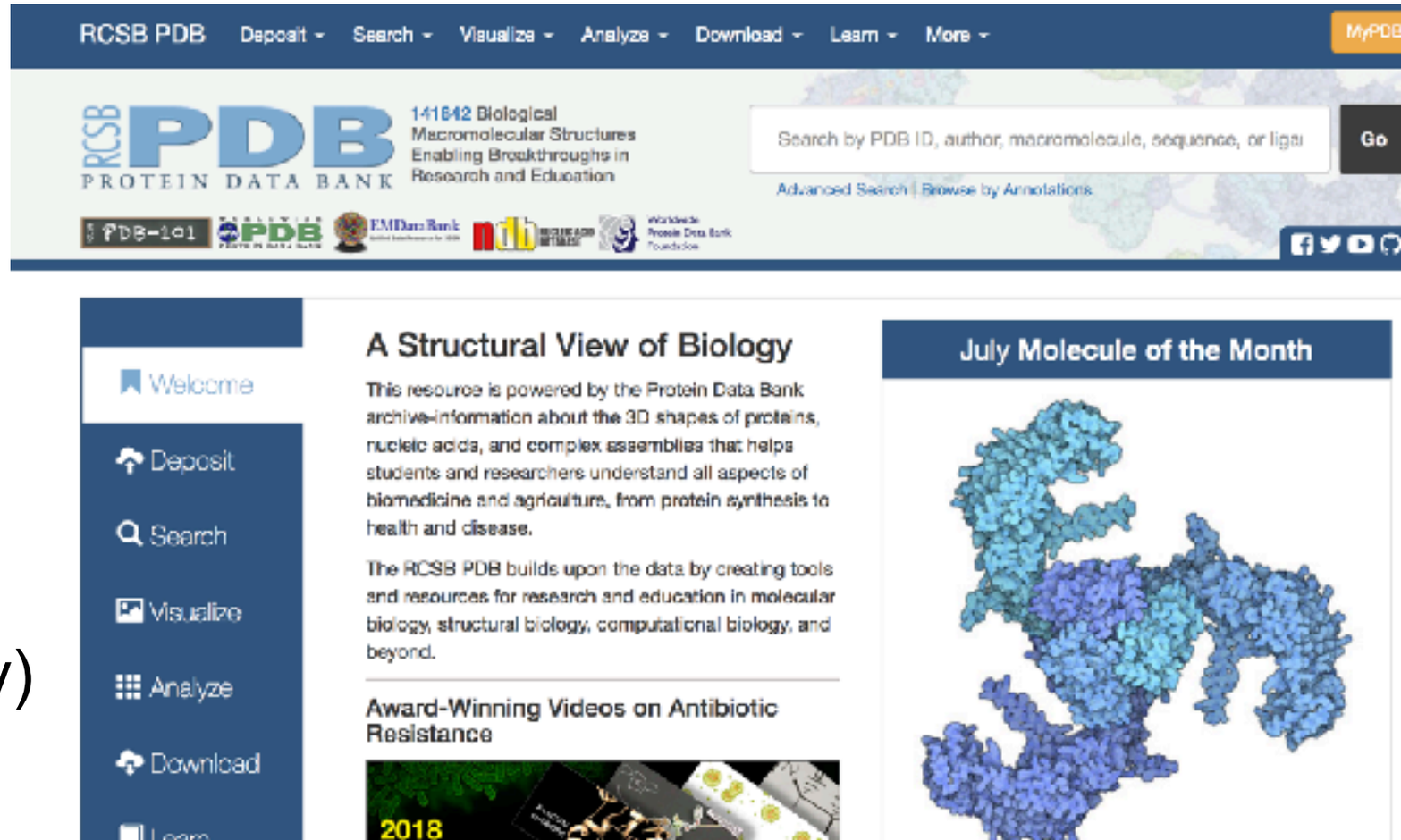
Simulations start with a crystal structure from the Protein Data Bank, in the standard PDB file format.

• PDB files contain standard records for species, tissue, authorship, citations, sequence, secondary structure, etc.

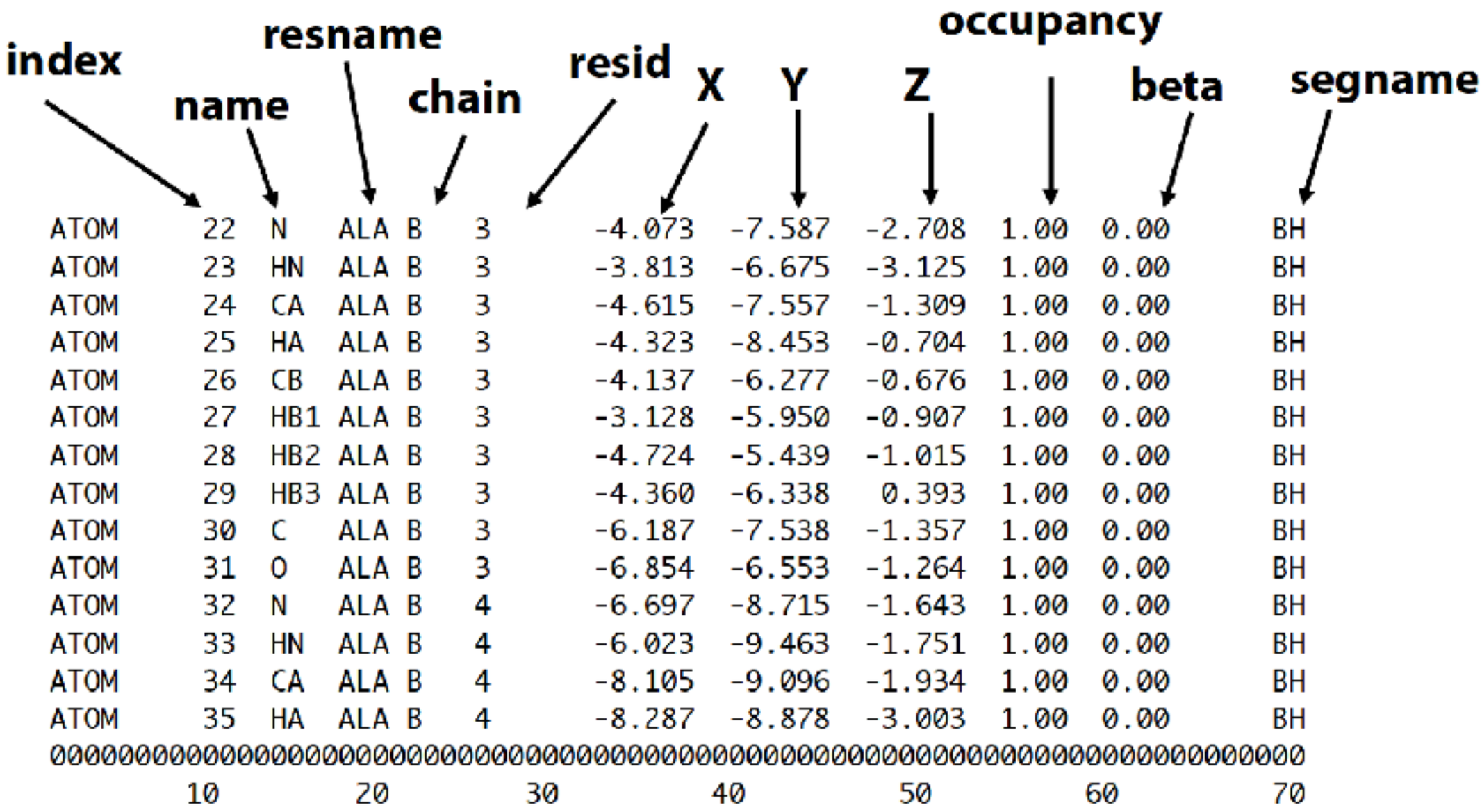
• We only care about the atom records…
  – atom name (N, C, CA)
  – residue name (ALA, HIS)
  – residue id (integer)
  – coordinates (x, y, z)
  – occupancy (0.0 to 1.0)
  – temp. factor (a.k.a. beta)
  – segment id (6PTI)

• No hydrogen atoms! (usually)
We must add them ourselves.

http://www.rcsb.org/



Structure of a PDB File

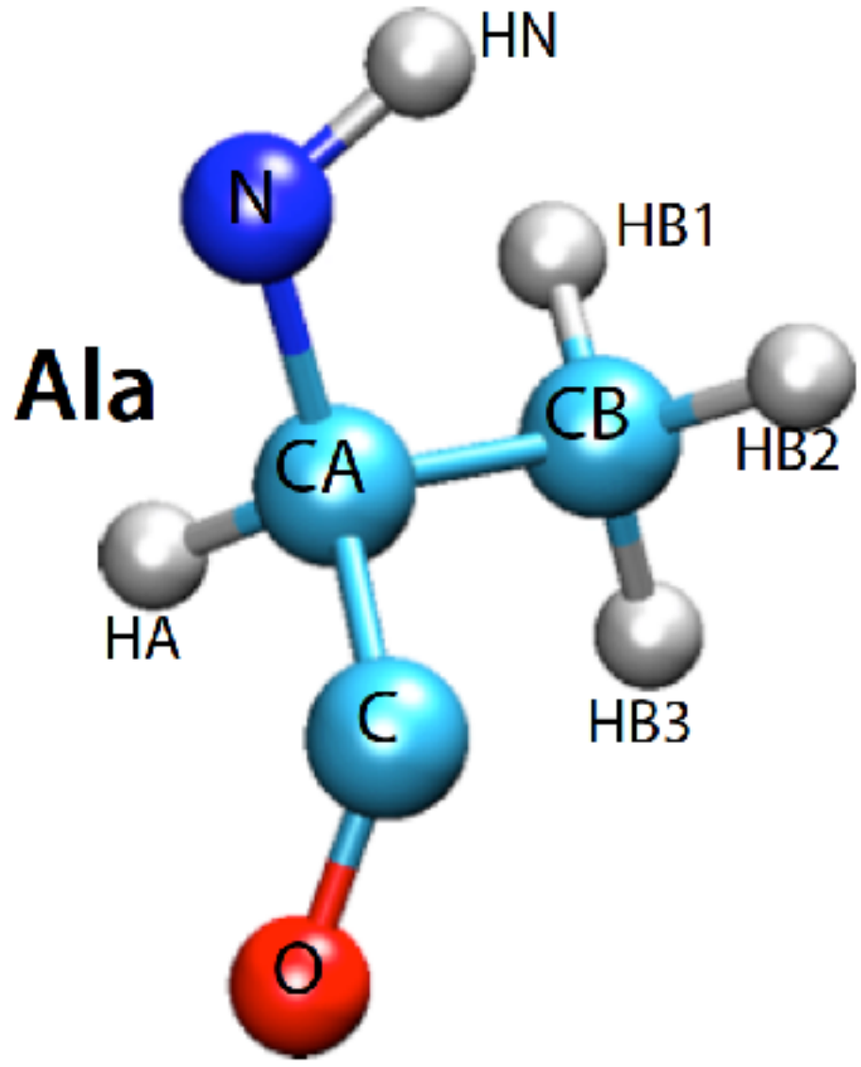


It is an ASCII, **fixed-width** file, which generally does not contain any connectivity information

PSF files *provide the topology and charges*

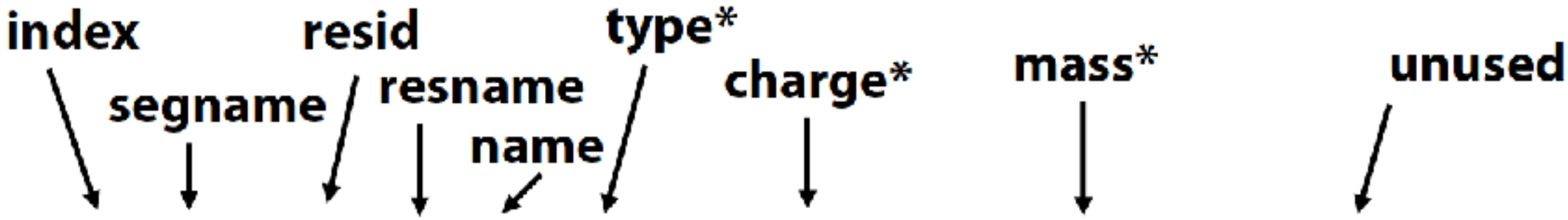
Every atom in the simulation is listed.

- Provides all static atom-specific values:
  - atom name (N, C, CA)
  - atom type (NH1, C, CT1)
  - residue name (ALA, TRP)
  - residue id (integer)
  - segment id (6PTI)
  - atomic mass (in atomic mass units)
  - partial charge (in electronic charge units)



- What is not in the PSF file?
  - coordinates (dynamic data, initially read from PDB file)
  - velocities (dynamic data, initially from Boltzmann distribution)
  - force field parameters (non-specific, used for many molecules)

Structure of a PSF file




```
 160 C     10    GLY  C     C      0.510000       12.0110           0
 161 C     10    GLY  O     O     -0.510000       15.9990           0
 162 C     11    ALA  N     NH1   -0.470000       14.0070           0
 163 C     11    ALA  HN    H      0.310000        1.0080           0
 164 C     11    ALA  CA    CT1    0.070000       12.0110           0
 165 C     11    ALA  HA    HB1    0.090000        1.0080           0
 166 C     11    ALA  CB    CT3   -0.270000       12.0110           0
 167 C     11    ALA  HB1   HA3    0.090000        1.0080           0
 168 C     11    ALA  HB2   HA3    0.090000        1.0080           0
 169 C     11    ALA  HB3   HA3    0.090000        1.0080           0
 170 C     11    ALA  C     C      0.510000       12.0110           0
 171 C     11    ALA  O     O     -0.510000       15.9990           0
 172 C     12    THR  N     NH1   -0.470000       14.0070           0
 173 C     12    THR  HN    H      0.310000        1.0080           0
 174 C     12    THR  CA    CT1    0.070000       12.0110           0
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
    10        20        30        40        50        60        70
```

Also an ASCII, **fixed-width** file, which does **not** contain coordinate information

Structure of a PSF file (extended)

index segname resid resname name type* charge* mass* unused

```
      82 A          212        THR        O          O          -0.510000       15.9990           0
      83 A          213        ALA        N          NH1        -0.470000       14.0070           0
      84 A          213        ALA        HN         H           0.310000        1.0080           0
      85 A          213        ALA        CA         CT1         0.070000       12.0110           0
      86 A          213        ALA        HA         HB1         0.090000        1.0080           0
      87 A          213        ALA        CB         CT3        -0.270000       12.0110           0
      88 A          213        ALA        HB1        HA3         0.090000        1.0080           0
      89 A          213        ALA        HB2        HA3         0.090000        1.0080           0
      90 A          213        ALA        HB3        HA3         0.090000        1.0080           0
      91 A          213        ALA        C          C           0.510000       12.0110           0
      92 A          213        ALA        O          O          -0.510000       15.9990           0
      93 A          214        GLU        N          NH1        -0.470000       14.0070           0
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
        10         20         30         40         50         60         70         80
```

"Extended" format supports **long** atom types and names (> 4 characters)
Current versions of NAMD and VMD handle this automatically

```
PSF EXT CMAP
```

"**EXT**" at the beginning of the file indicates extended format

Structure of a topology file

Topology files contain instructions for building different types of biomolecules, used to build the PSF

```
RESI ALA          0.00
GROUP
ATOM N    NH1    -0.47   !
ATOM HN   H       0.31   !   HN-N
ATOM CA   CT1     0.07   !       |      HB1
ATOM HA   HB1     0.09   !       |     /
GROUP                    !   HA-CA--CB-HB2
ATOM CB   CT3    -0.27   !       |     \
ATOM HB1  HA3     0.09   !       |      HB3
ATOM HB2  HA3     0.09   !     O=C
ATOM HB3  HA3     0.09   !       |
GROUP                    !
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB CA  N  HN  N  CA
BOND C  CA  C  +N  CA HA  CB HB1  CB HB2  CB HB3
DOUBLE O  C
IMPR N -C CA HN  C CA +N O
CMAP -C  N  CA  C   N  CA  C  +N
DONOR HN N
ACCEPTOR O C
IC -C   CA   *N   HN    1.3551 126.4900  180.0000 115.4200  0.9996
IC -C   N    CA   C     1.3551 126.4900  180.0000 114.4400  1.5390
IC N    CA   C    +N    1.4592 114.4400  180.0000 116.8400  1.3558
IC +N   CA   *C   O     1.3558 116.8400  180.0000 122.5200  1.2297
IC CA   C    +N   +CA   1.5390 116.8400  180.0000 126.7700  1.4613
```

residue name, total charge

atom name, type, charge (after ! is a comment)

groups not required, but generally indicate sets of atoms with integer charges

bonds explicitly listed, but not angles, dihedrals

impropers maintain planarity

adjustment to dihedral terms from QM for proteins

ignored

Internal coords. help build missing atoms

Structure of a parameter file

Parameter files (used during simulation) tell NAMD what the force constants, etc. are - organized by **atom type**, *not name*

```
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb          b0
!
```

entries for bonds, angles, dihedrals, impropers, and LJ

! comments

type 1 **type 2** **force constant** **eq. bond length**

```
NH2   CT1    240.000     1.4550 ! From LSN NH2-CT2
!
!Indole/Tryptophan
CA    CAI    305.000     1.3750 ! from CA CA
CAI   CAI    305.000     1.3750 ! atm, methylindole, fit CCDSS
CPT   CA     300.000     1.3600 ! atm, methylindole, fit CCDSS
```

NOTE: always check for the latest force-field files (**multiple** topology and parameters) at their source - we currently use CHARMM36

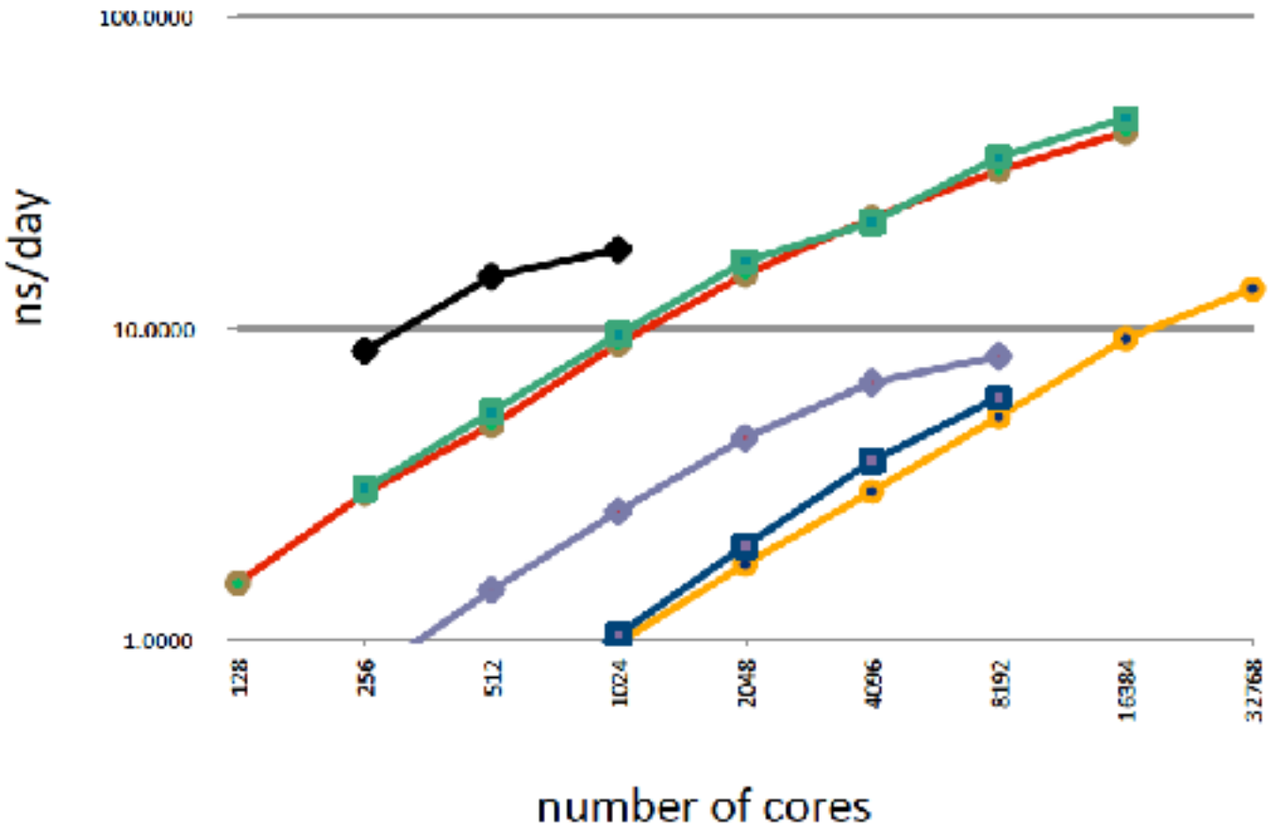
http://mackerell.umaryland.edu/charmm_ff.shtml

Measuring performance

Check your scaling!!!

grep "Benchmark" *log

```
Info: Benchmark time: 42 CPUs 0.0879267 s/step 1.01767 days/ns 87.665 MB
memory
```



$$\text{Efficiency} = \frac{\text{\# s/step (1 cpu)}}{n * \text{\# s/step } (n \text{ cpus})}$$

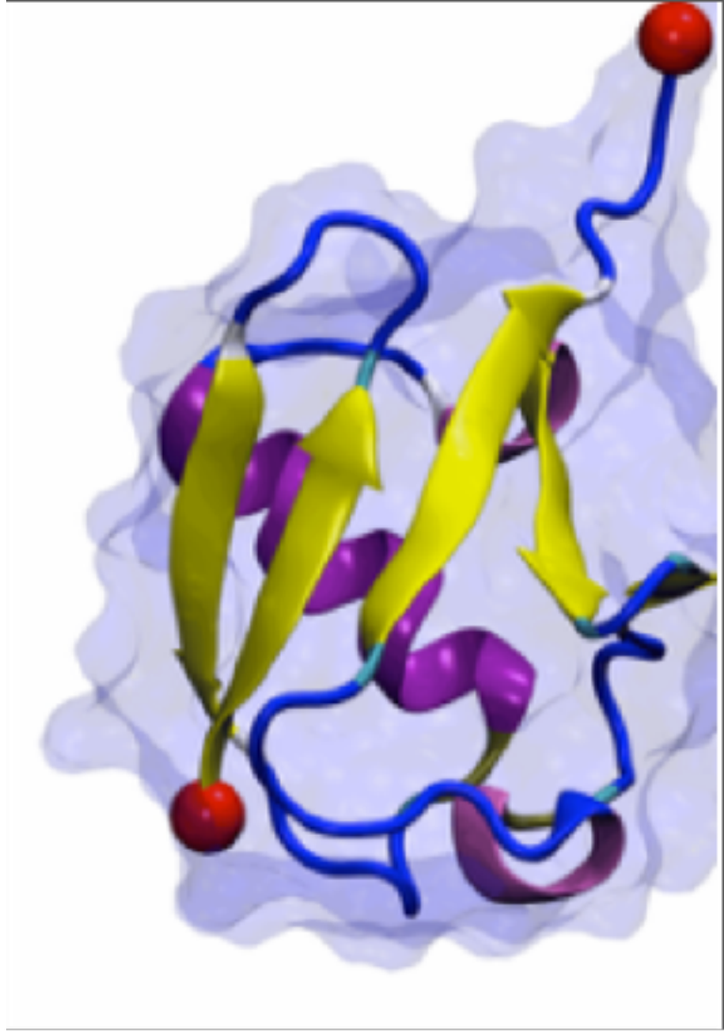
TYPICAL RANGE:
250-1000 atoms/core

grep “TIMING" *log

```
TIMING: 3000 CPU: 346.34, 0.07938/step Wall: 466.648, 0.0879514/step,
6.08331 hours remaining, 88.341812 MB of memory in use.
```

Dealing with crashes

Some errors are obvious…



"Cannot specify both an initial temperature and a velocity file"

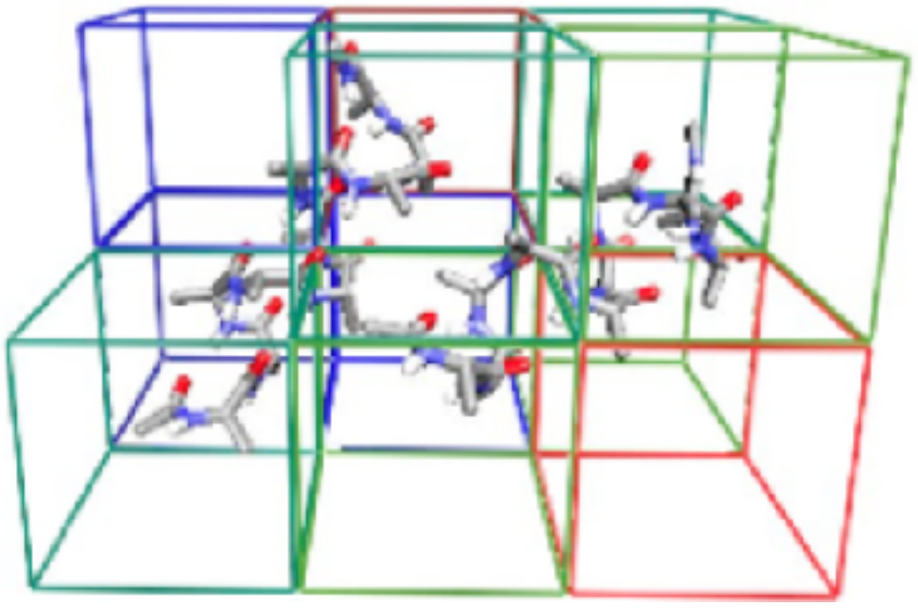
"stepsPerCycle must be a multiple of fullElectFrequency"

etc…

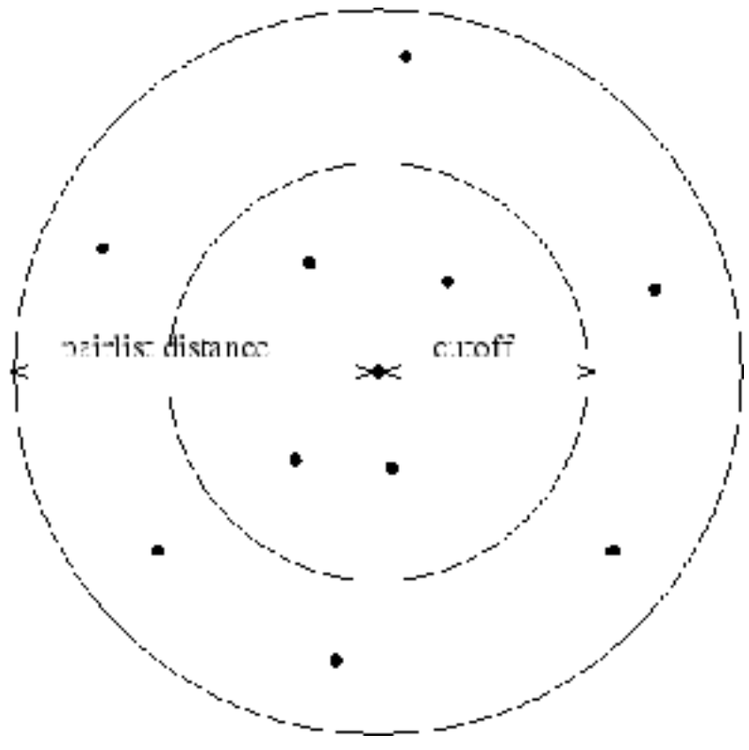
Others not so much…

```
FATAL ERROR: Periodic cell has become too small for original patch grid!
Possible solutions are to restart from a recent checkpoint,
increase margin, or disable useFlexibleCell for liquid simulation.
```

-relates to how NAMD parallelizes the simulation



Atoms that move close enough to interact (defined by cutoff) but are not on neighboring patches causes a crash



-typically happens because of large volume fluctuations (normal during initial equilibriation in NpT ensemble), but **CHECK OUTPUT TO BE SURE**

**can set “margin 2” (force bigger patches) in configuration file
**lower “stepsPerCycle” in configuration file
**just restart

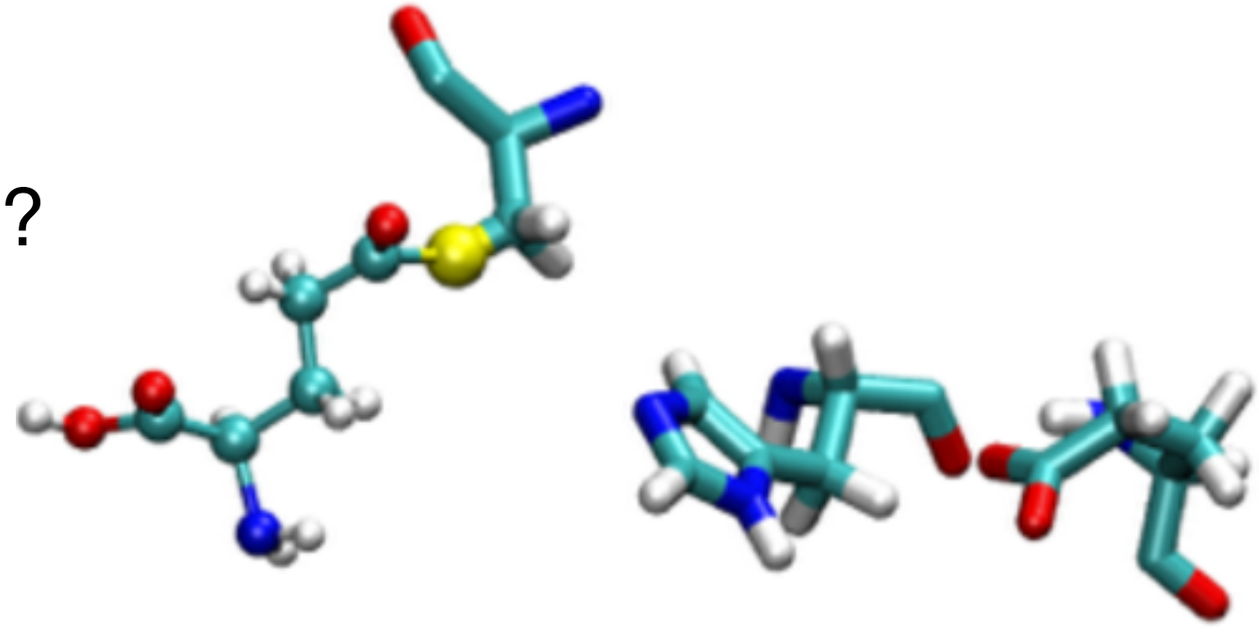
Missing parameters?

```
DIDN'T FIND vdW PARAMETER FOR ATOM TYPE CT3
```

**Did you specify all the needed parameter files?

**Was your system (PSF/PDB) constructed correctly? (Check for errors/warnings from PSFGen or AutoPSF!)

**Do you have an unusual ligand?
(need to either remove or
develop parameters for it)

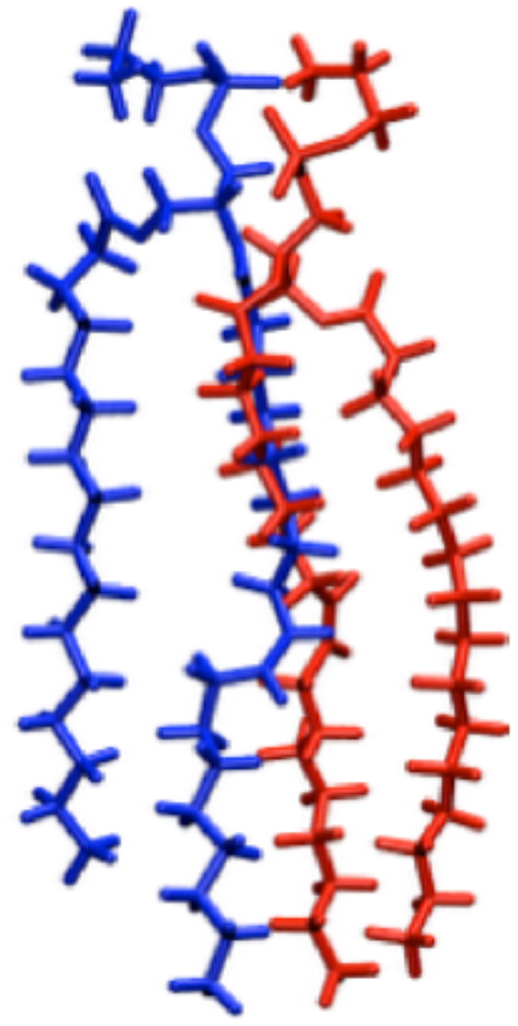


Simulation instability

```
ERROR: Atoms moving too fast; simulation has become
unstable.
ERROR: Constraint failure in RATTLE algorithm for atom
1897!
```

Both errors almost always derive from bad system configurations!

**Check your system in VMD near the noted atoms
**Use the “measure contacts” command to check for atoms that are very close (say, within 0.1 Å)
**Look for atoms at (0,0,0) whose positions didn’t get initialized when building PSF/PDB
**Check that the periodic box dimensions are big enough
**Minimize for longer, or set margin higher
If all else fails, change your DCDFreq to 1 and watch the simulation up to the point of the crash **very carefully**

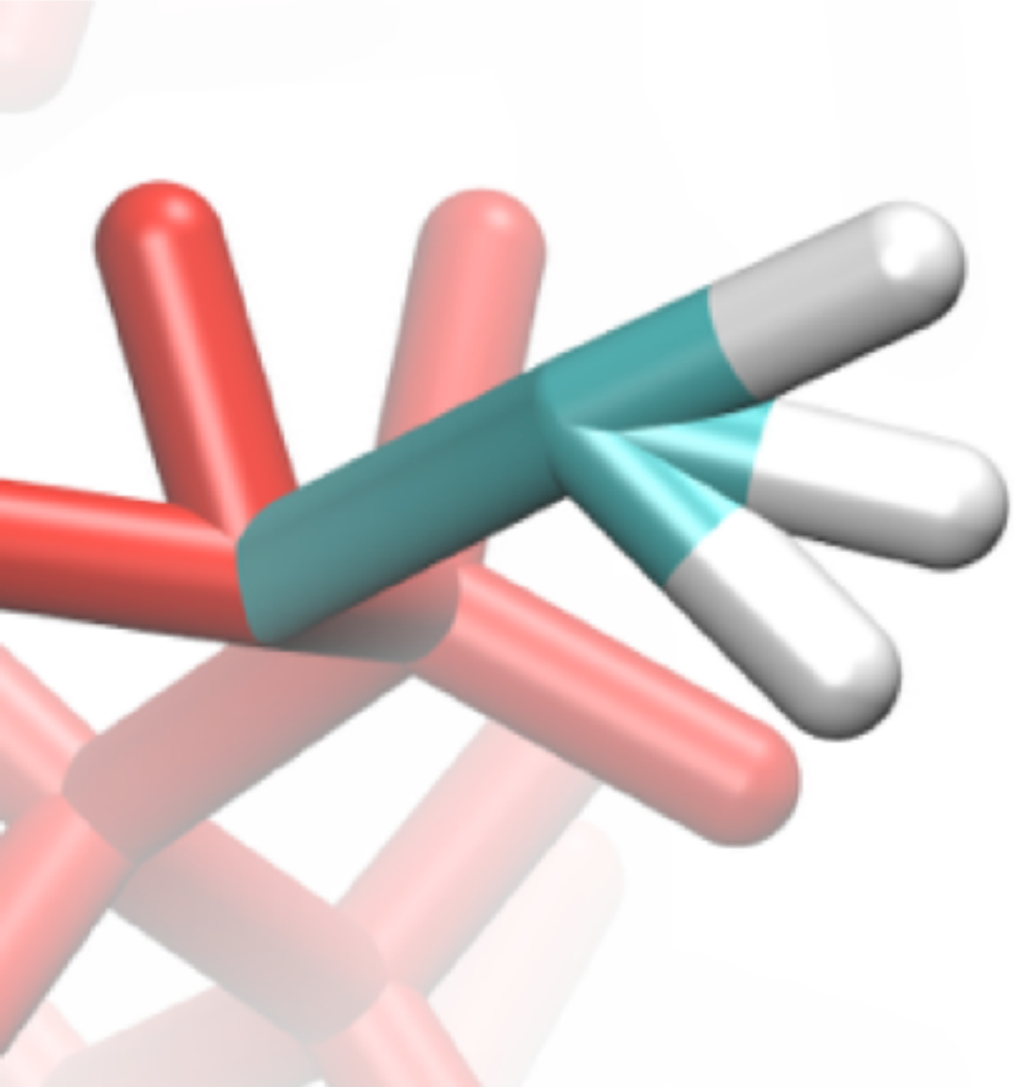


NOTE: NAMD uses 1-based numbers whereas VMD typically uses 0-based
In VMD - use “serial ____” instead of “index”

Simulation instability (cont.)

```
FATAL ERROR: Bad global exclusion count.
```

typically results from bad starting configuration, similar to previous errors



-besides previous solutions, consider the possibility of missing angle or dihedral entries from the PSF file

-when a patch is applied by PSFGen, the command

“regenerate angles dihedrals”

may need to be issued before guessing coordinates and writing the PSF (when in doubt, add it!)

any questions?