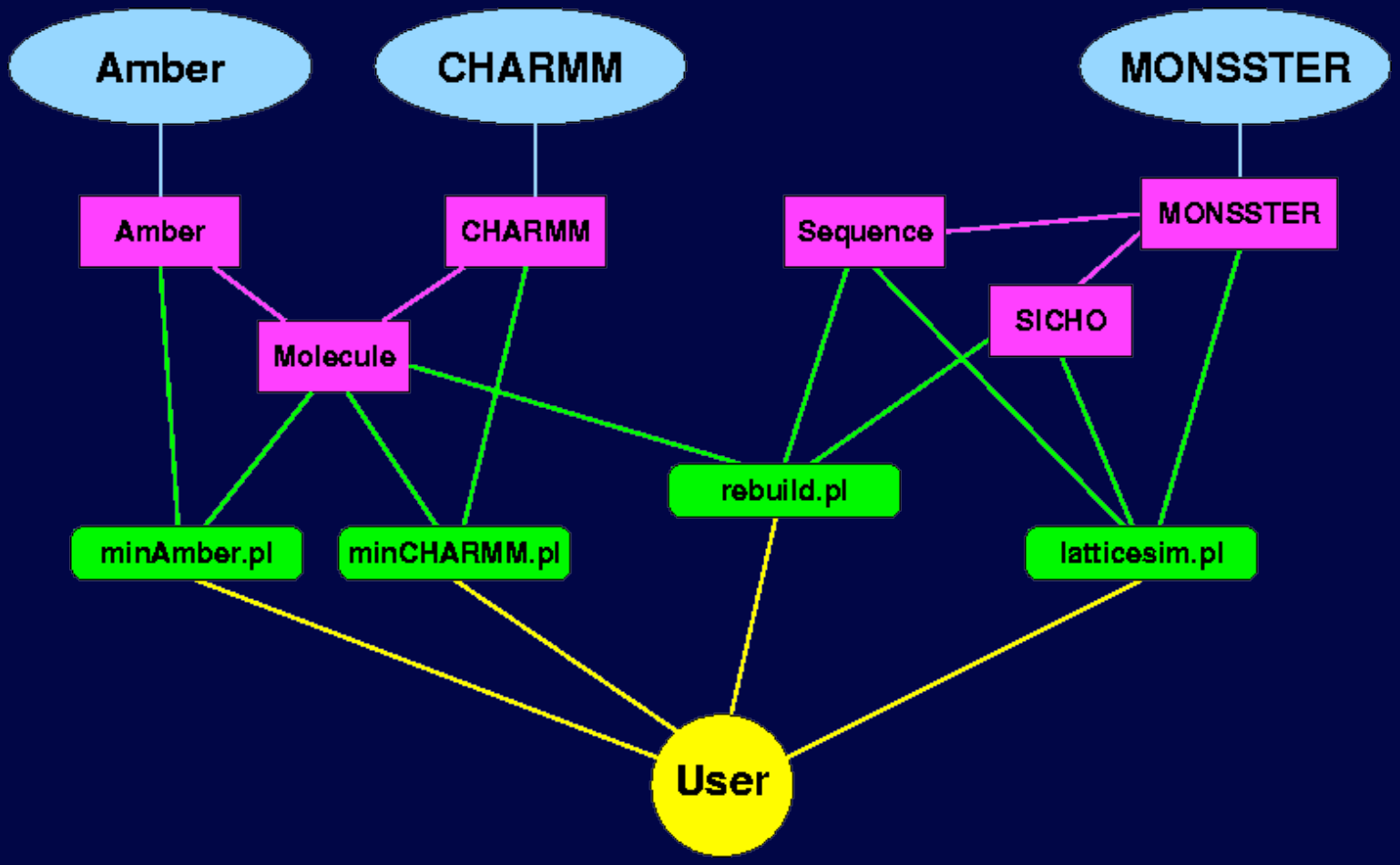
MMTSB Toolset

Michael Feig
MMTSB/CTBP
2006 Summer Workshop

MMTSB Tool Set Overview



application programs

Perl libraries

Perl utilities

MMTSB Tool Set Functionality

- Structure manipulation and analysis
- Classical modeling of biomolecules
- Enhanced sampling
- Multi-scale modeling
- Ensemble computing
- Protein structure prediction

MMTSB Tool Set Functionality

- **Structure manipulation and analysis**
  - > PDB file manipulation, structure subselection
  - > Rotation/translation, residue mutation
  - > RMSD, ϕ/ψ dihedrals, native contacts, clustering
  - > Trajectory analysis
- Classical modeling of biomolecules
- Enhanced sampling
- Multi-scale modeling
- Ensemble computing
- Protein structure prediction

MMTSB Tool Set Functionality

- Structure manipulation and analysis
- **Classical modeling of biomolecules**
  - Molecular dynamics simulations, energy evaluation, minimization
  - **CHARMM**/Amber interfaces
- Enhanced sampling
- Multi-scale modeling
- Ensemble computing
- Protein structure prediction

MMTSB Tool Set Functionality

- Structure manipulation and analysis
- Classical modeling of biomolecules
- **Enhanced sampling**
  - > Replica exchange sampling: 1D, multidimensional
  - > Weighted histogram analysis (WHAM)
- Multi-scale modeling
- Ensemble computing
- Protein structure prediction

MMTSB Tool Set Functionality

- Structure manipulation and analysis
- Classical modeling of biomolecules
- Enhanced sampling
- **Multi-scale modeling**
  - All-atom sampling with CHARMM and Amber
  - Low-resolution sampling with MONSSTER
- Ensemble computing
- Protein structure prediction

MMTSB Tool Set Functionality

- Structure manipulation and analysis
- Classical modeling of biomolecules
- Enhanced sampling
- Multi-scale modeling
- **Ensemble computing**
  - > Minimization, energy evaluation, clustering for ensembles
  - > Arbitrary operations on ensemble level
- Protein structure prediction

MMTSB Tool Set Functionality

- Structure manipulation and analysis
- Classical modeling of biomolecules
- Multi-scale modeling
- Enhanced sampling
- Ensemble computing
- **Protein structure prediction**
  - > sequence alignment
  - > secondary structure prediction
  - > template-based modeling
  - > abinitio folding

Using the MMTSB Tool Set

- **Command-line utilities**
  - for common applications
  - easy to learn and use

```
> minCHARMM.pl -par minsteps=100 1vii.pdb
```

- Programming with MMTSB libraries

Using the MMTSB Tool Set

- Command-line utilities
- **Programming with MMTSB libraries**
  - > for complex and/or specialized tasks
  - > requires programming experience

```
#!/usr/bin/perl
use Molecule;
use Analyze;
my $mol = &Molecule::new("1vii.pdb");
my $rgyr=&Analyze::radiusOfGyration($mol);
for ($i=0; $i<10; $i++) {
 $mol->{chain}->[0]->{atom}->[$i]->{xcoor}+=$rgyr;
 …
}
$mol->writePDB("1vii.moved.pdb");
```

Documentation/Help

- -help option

```
> dihed.pl -help
usage:   dihed.pl [options] [refPDB [cmpPDB]]
options: [-l min:max[...]]
         [-list phi|psi|chi1|omega]
```

- Online manual pages with examples

  http://mmtsb.scripps.edu/cgi-bin/tooldoc

- Email support: mmtsb@scripps.edu
- MMTSB Forum: http://www.charmm.org

Command Line Utilities: Input/Output

- Input usually expected from file name or standard input

```
> minCHARMM.pl 1vii.pdb
> cat 1vii.pdb | minCHARMM.pl
```

- Output usually written to standard output

```
> minCHARMM.pl 1vii.pdb > 1vii.min.pdb
```

- "-" can be used for standard input/output

```
> cat 1vii.pdb | lsqfit.pl - 1vii.min.pdb
> minCHARMM.pl -log - 1vii.pdb
```

Command Line Utilities: Pipes

```
> convpdb.pl -solvate -cubic 1vii.pdb |
  minCHARMM.pl -par minsteps=500 1vii.pdb |
  mdCHARMM.pl -par dynsteps=5000,dyntemp=100 -final - |
  mdCHARMM.pl -par dynsteps=5000,dyntemp=200 -final - |
  mdCHARMM.pl -par dynsteps=5000,dyntemp=250 -final - |
  mdCHARMM.pl -par dynsteps=100000 -trajout traj.dcd
```

Solvate protein -> Minimize -> Equilibrate -> Run production

Structure Preparation:
PDB Conversion

- Generate specific format

```
> convpdb.pl -segnames -out charmm22 1vii.pdb
ATOM      1  N   MET     1       1.177 -10.035  -3.493  1.00  0.00      PRO0
ATOM      2  HT1 MET     1       1.833 -10.061  -2.687  1.00  0.00      PRO0
ATOM      3  HT2 MET     1       1.719  -9.981  -4.379  1.00  0.00      PRO0
ATOM      4  HT3 MET     1       0.596 -10.897  -3.495  1.00  0.00      PRO0
ATOM      5  CA  MET     1       0.292  -8.839  -3.377  1.00  0.00      PRO0
…
```

```
> convpdb.pl -setchain A -out amber 1vii.pdb
ATOM      1  N   MET A   1       1.177 -10.035  -3.493  1.00  0.00
ATOM      2  H1  MET A   1       1.833 -10.061  -2.687  1.00  0.00
ATOM      3  H2  MET A   1       1.719  -9.981  -4.379  1.00  0.00
ATOM      4  H3  MET A   1       0.596 -10.897  -3.495  1.00  0.00
ATOM      5  CA  MET A   1       0.292  -8.839  -3.377  1.00  0.00
…
```

- Renumber residues

```
> convpdb.pl -addres 5 1vii.pdb
```

Structure Preparation:
Subselection

- Basic selection options

```
> convpdb.pl -sel 20:36 1vii.pdb
> convpdb.pl -chain A 1i3h.pdb
> convpdb.pl -selseq KAVFG 1vii.pdb
```

- Expanded selection syntax

```
> convpdb.pl -nsel 20-36.CA 1vii.pdb
> convpdb.pl -nsel LEU/20-36.CA 1vii.pdb
> convpdb.pl -nsel A+B:polar 1i3h.pdb
> convpdb.pl -nsel @KAVFG.CB+CA 1vii.pdb
> convpdb.pl -nsel nucleic 1i3h.pdb
```

Structure Preparation:
Orientation

- Center

```
> convpdb.pl -center 1vii.pdb
```

- Translate

```
> convpdb.pl -translate -1.2 1.0 5.0 1vii.pdb
```

- Rotate

```
> convpdb.pl -rotatex 90.0 1vii.pdb
> convpdb.pl -rotate 1 0 0 0 -1 0 0 0 1 1vii.pdb
```

Structure Preparation
Residue Mutation

- Proteins/peptides

```
> mutate.pl -seq 10:AGT 1vii.pdb
```

```
   MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF
-> MLSDEDFKAAGTMTRSAFANLPLWKQQNLKKEKGLF
```

- Nucleic acids

```
> mutateNA.pl -seq A3:B23:AT dna.pdb
```

```
   5′-GCGCAATTGCGC-3′      5′-GATCAATTGCGC-3′
      ||||||||||||     ->     ||||||||||||
   5′-CGCGTTAACGCG-3′      3′-CTAGTTAACGCG-5′
```

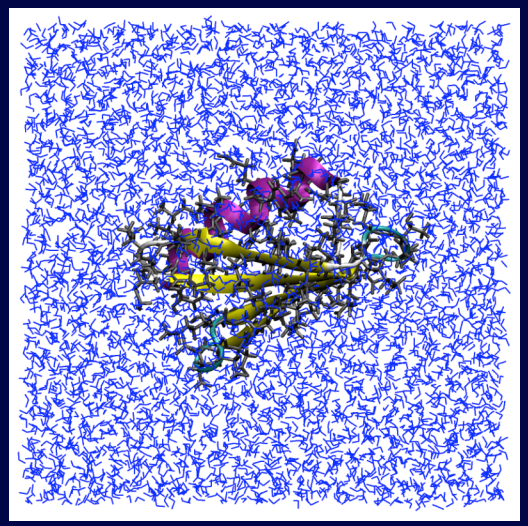
Structure Preparation
Solvation

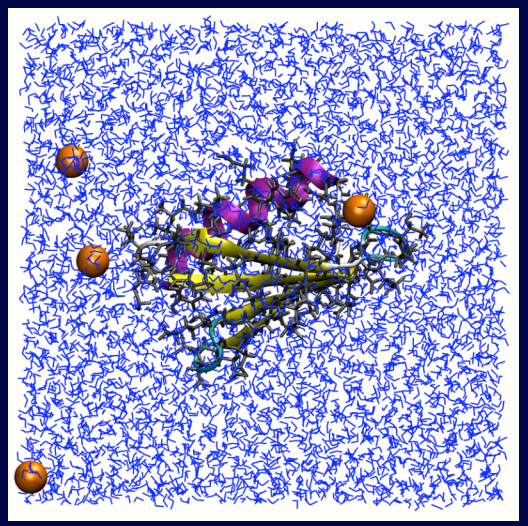
- Add explicit water box

```
> convpdb.pl -solvate -cubic 1vii.pdb > 1vii.water.pdb
```

- Add counterions/additional salt

```
> convpdb.pl -ions SOD:10 1vii.water.pdb > 1vii.ions.pdb
```





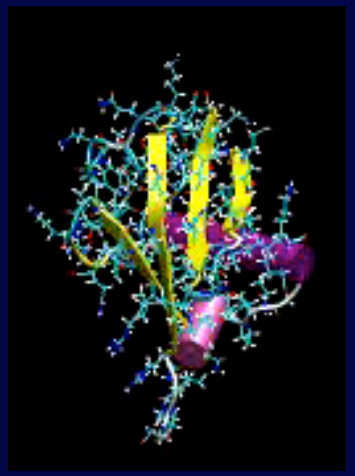
All-atom Modeling of Biomolecules

- User interface to CHARMM/Amber

  minCHARMM.pl, minAmber.pl

  mdCHARMM.pl, mdAmber.pl

  pbCHARMM.pl



```
> mdCHARMM.pl
  -par dynsteps=10000,gb,cutoff=16,lang,langfbeta=50
  -cons CA self 10:20_10
  -custom extra.commands
  -trajout traj.dcd -final last.pdb -restout restartfile
  -log charmm.log -cmd cmdfile
  init.pdb
```

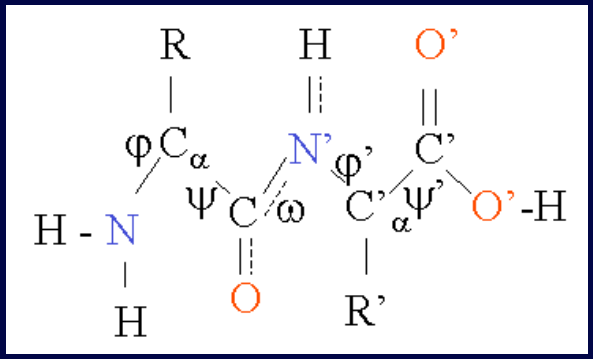
Replica Exchange Sampling


```
> aarex.pl -n 100 -temp 4:300:500
  -mdpar dynsteps=500,gb
  -dir rex.1 -log rex.log
  init.pdb
```

Structural Analysis

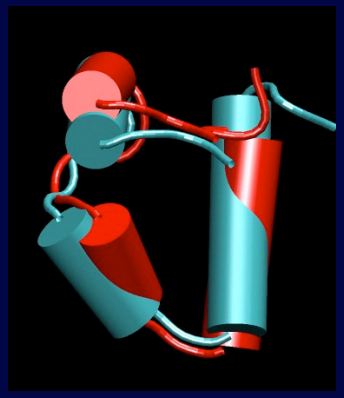
- Protein backbone torsions

```
> dihed.pl -list phi,psi 1vii.pdb
MET1:     0.000   132.907
LEU2:   -64.308   100.191
SER3:   -63.139   154.230
…
```



- RMSD calculation

```
> rms.pl -out CA -fit 1vii.pdb 1vii.sample.pdb
  2.6944 CA
```

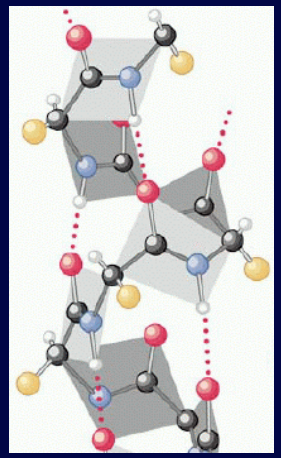


Secondary Structure Analysis

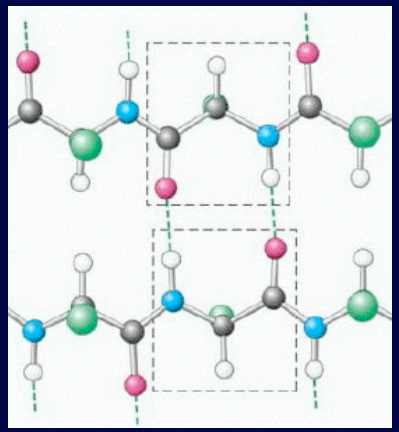
```
> genseq.pl -out onesec -dssp 1vii.pdb
MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF
UUUHHHHHUUUUUUHHHHUUUUHHHHHHHHHHUUUU
```

Secondary structure identified from hydrogen bonding:

H – helical (α)



E – extended (β)

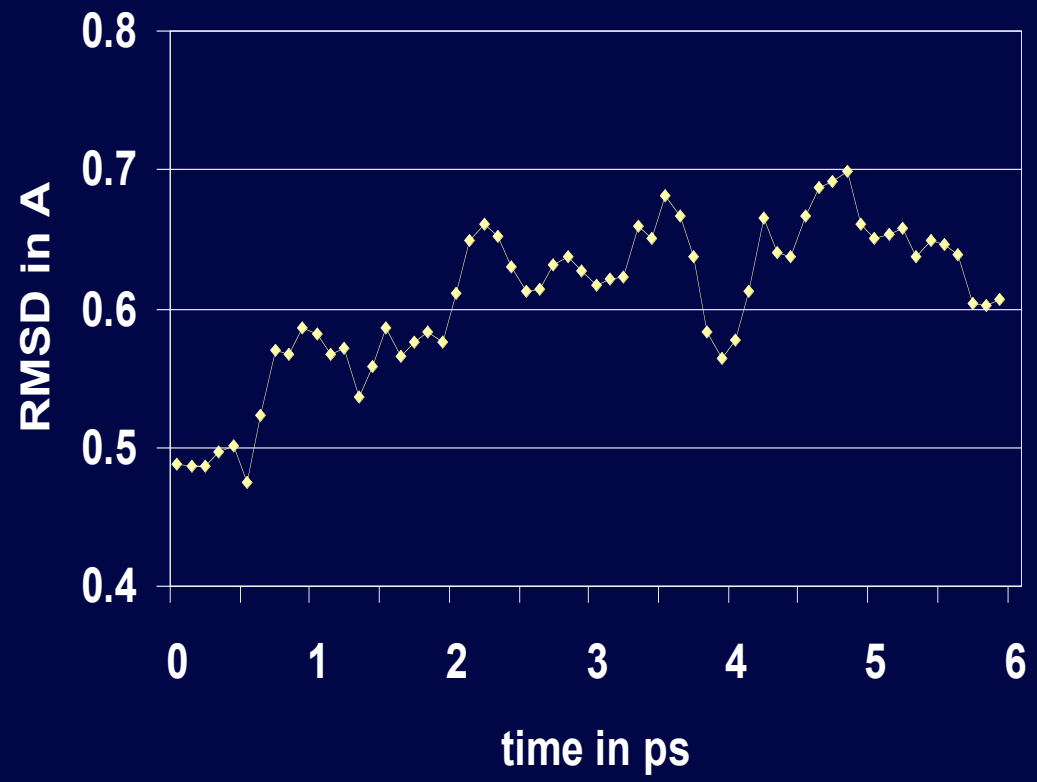


U – random coil/undefined

Trajectory Analysis

- RMSD from reference vs. time

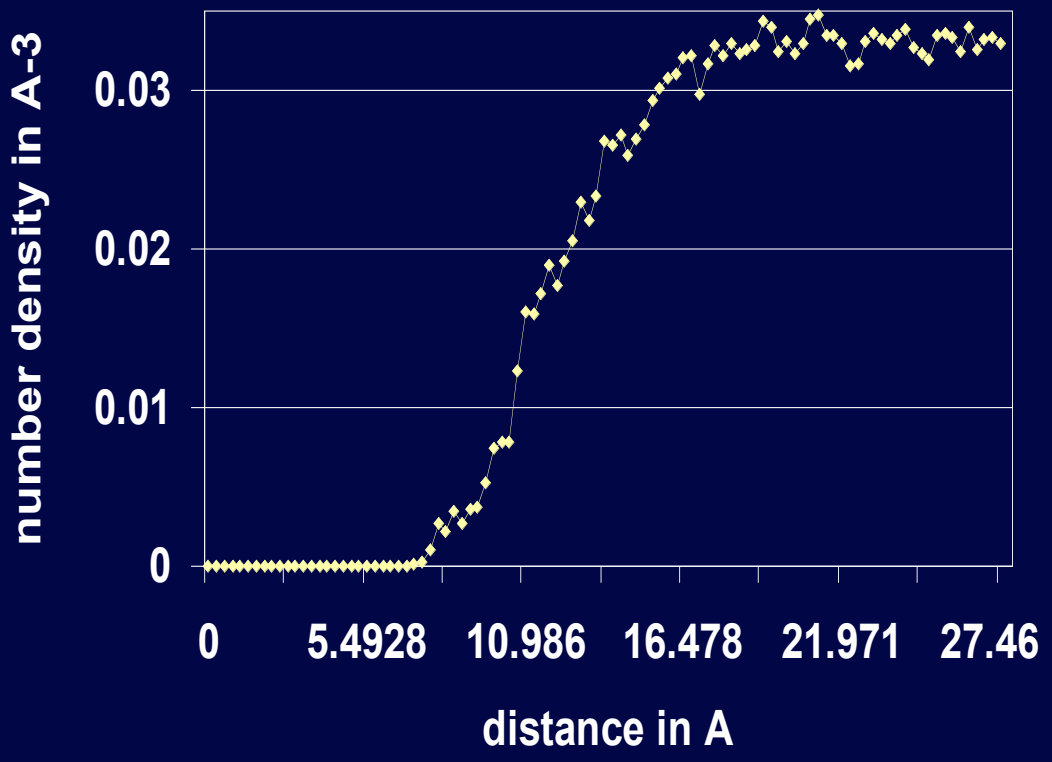
```
> analyzeCHARMM.pl -rms -sel CA -comp 1vii.pdb
                   -pdb md.1.pdb md.1.dcd md.2.dcd
```



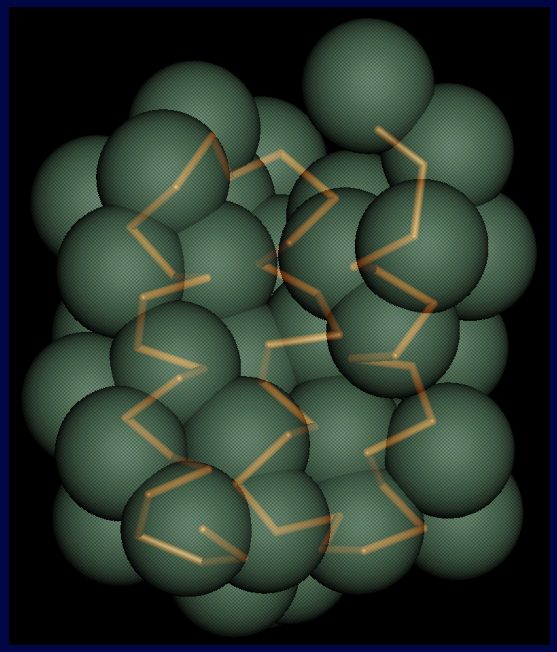
Trajectory Analysis

- Radial distribution function

```
> analyzeCHARMM.pl -rdist -dsel water.OH2 solute.nitrogen
                   -pdb md.1.pdb md.1.dcd md.2.dcd
```

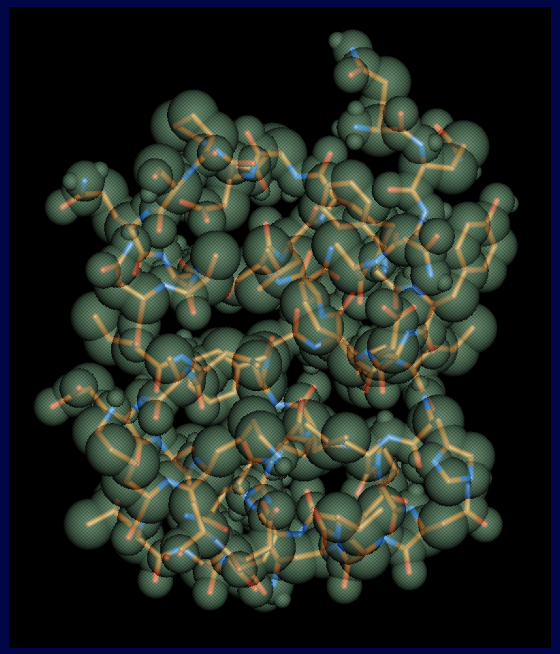


Low Resolution Modeling of Proteins



Reduction

Reconstruction



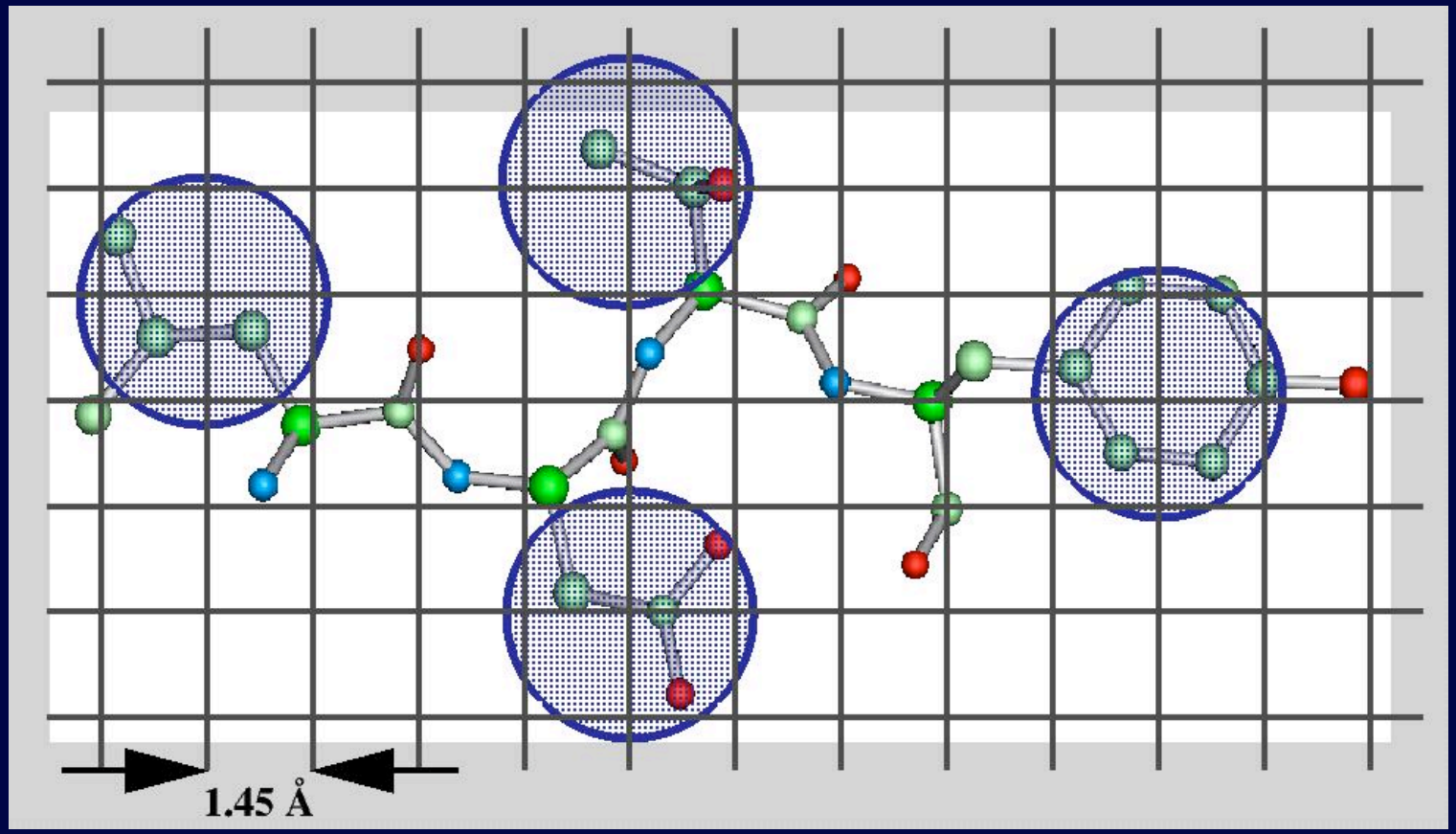
residue-level

atom-level

Sampling

Accuracy

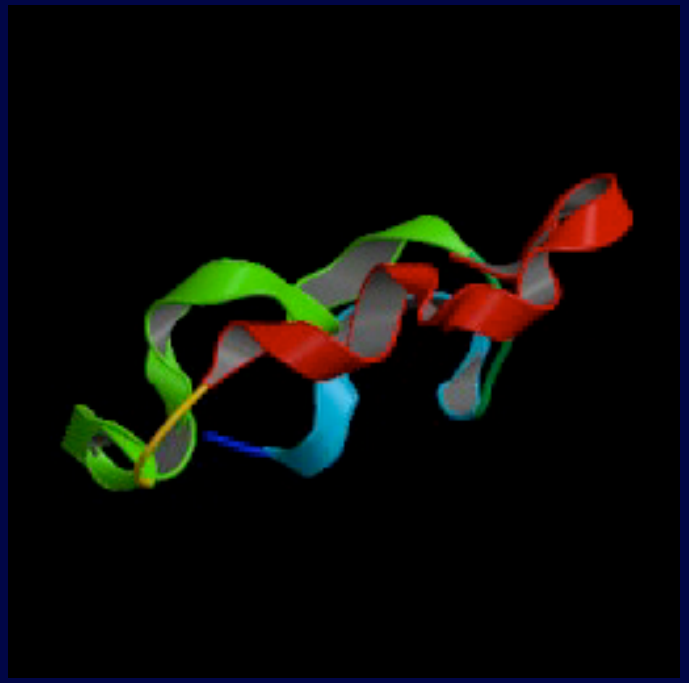
SICHO Lattice Model



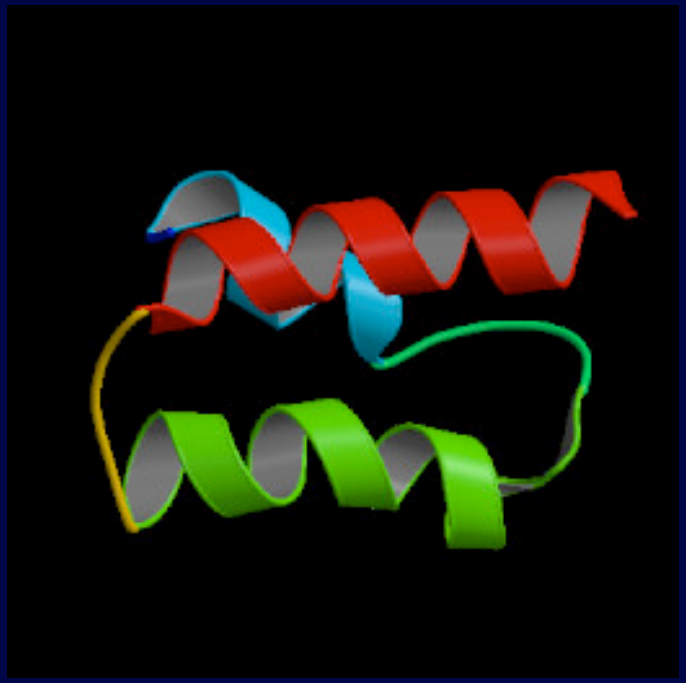
Kolinski & Skolnick: Proteins *32*, 475 (1998)

Protein Folding with SICHO Model

Protein A



Lattice simulation
(Simulated Annealing)



Native

MMTSB and MONSSTER

MOdeling of New Structures from Secondary and TErtiary Restraints

```
> latticesim.pl -sa 2.5 -pdb 1vii.pdb 1vii.seq
```

```
> genseq.pl -dssp 1vii.pdb
    1    MET     1     1
    2    LEU     1     1
    3    SER     1     1
    4    ASP     2     1
    5    GLU     2     1
    …
```

```
> psipred.pl seq.file | genseq.pl -2ndone - 1vii.pdb
    1    MET     1     1
    2    LEU     1     1
    3    SER     1     1
    4    ASP     2     1
    5    GLU     2     1
    …
```

Model Translation

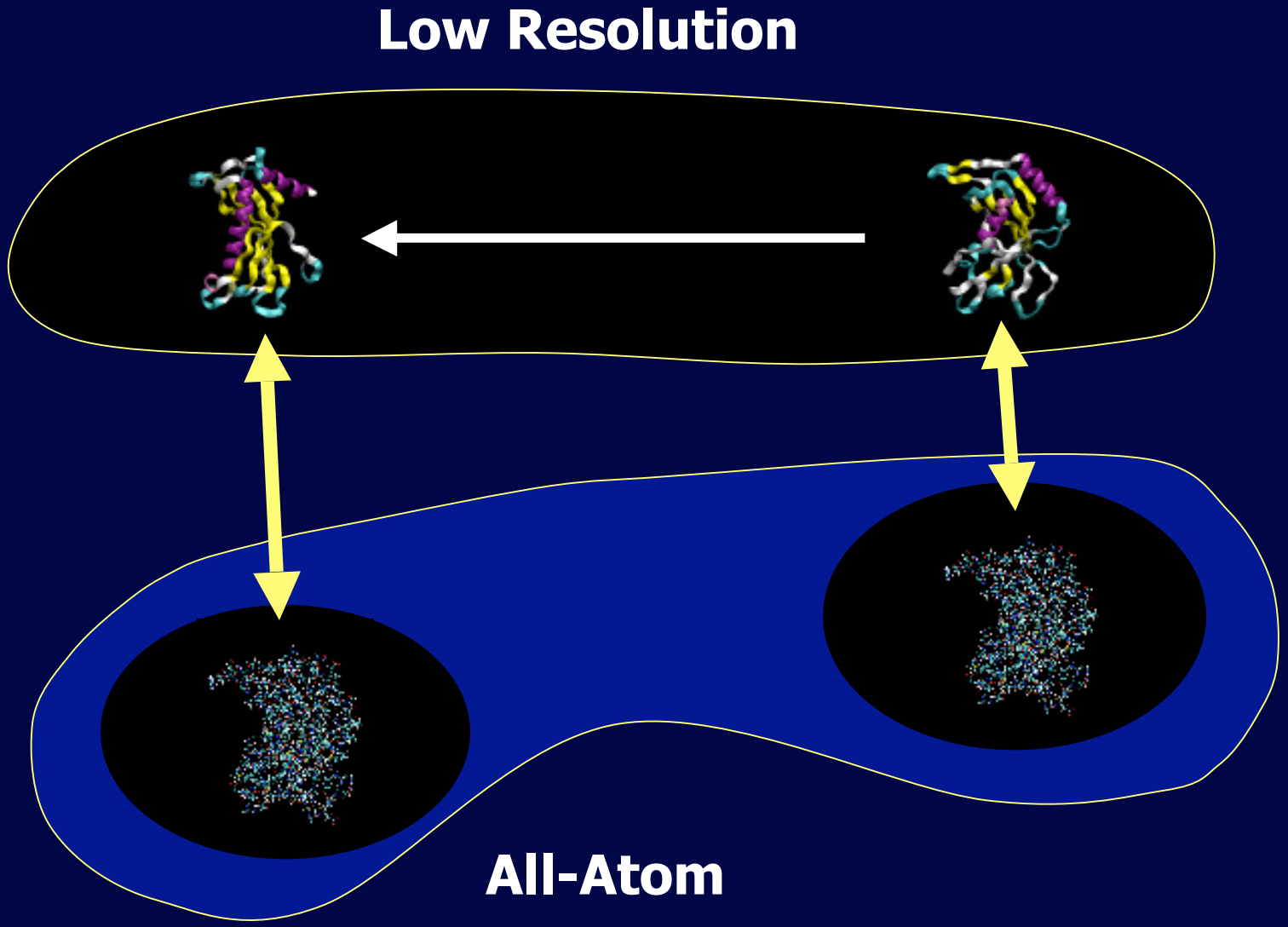
- Reduction: All-atom -> Lattice

```
> genchain.pl 1vii.pdb
38
   48   45   43
   49   43   46
   50   45   51
   …
```

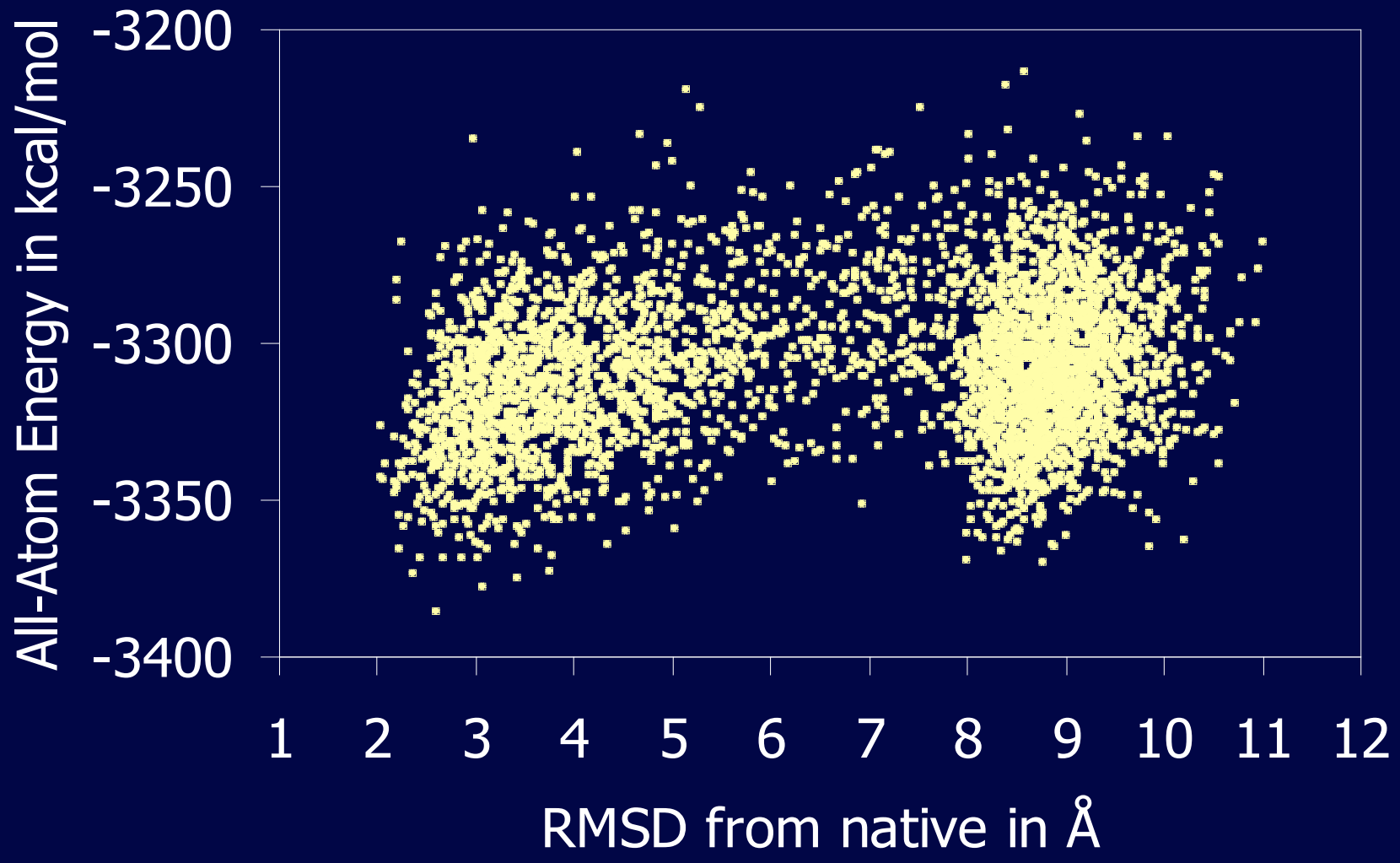
- Reconstruction: Lattice -> All-atom

```
> rebuild.pl 1vii.seq 1vii.chain
ATOM   1 N   MET  1   -0.194  -8.104  -7.197  1.00 0.00
ATOM   2 CA  MET  1    0.003  -8.457  -5.787  1.00 0.00
ATOM   3 C   MET  1   -1.063  -8.084  -5.005  1.00 0.00
…
```

Multi-Scale Sampling



Conformational Sampling Data



Ensemble Computing

Common operations on conformational ensembles:

- Structural derivatives

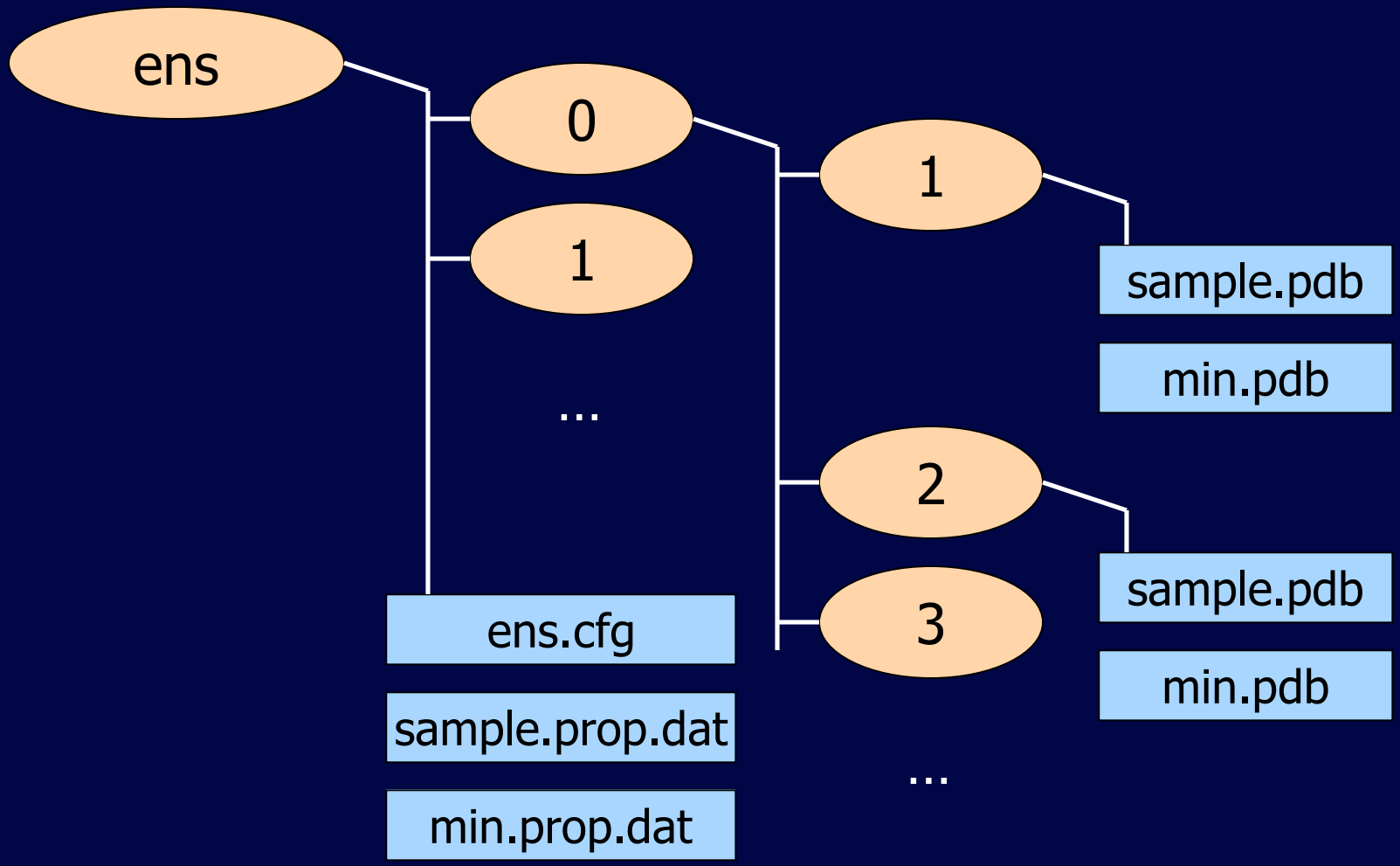
```
> ensmin.pl -par minsteps=100 sample min
```

sample -> min

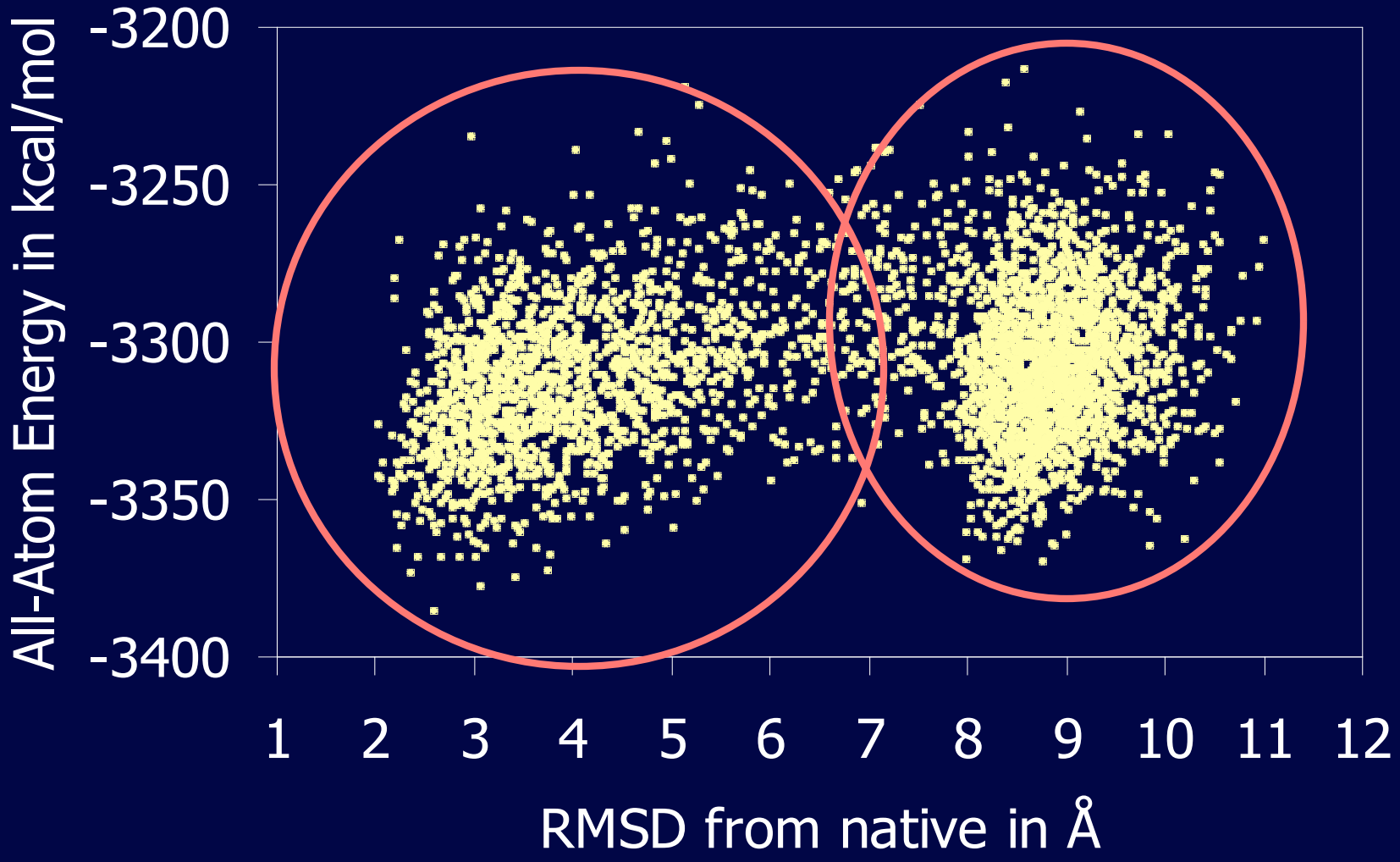
- Collective analysis

```
> ensrun.pl -set score:1 min enerCHARMM.pl -par gb
> getprop.pl -prop score min
1 -1989.4
2 -2020.5
3 -1801.9
4 -1923.0
...
```

Ensemble Data Structure



Clustering



Data Clustering with MMTSB

- Ensemble clustering

```
> enscluster.pl -radius 4 -kclust min
```

- Analyze clusters

```
> bestcluster.pl -prop score -crit avglow min
t.3      30  27   -2242.7933  33.7171  6.4889
t.4      12  10   -2162.9970  21.8099  6.8969
t.2      53  50   -2160.8799  44.4020  6.2794
…

> ensfiles.pl -cluster t.3 -sort score min
./0/18/sample.pdb -2299.6500
./2/44/sample.pdb -2285.5200
./1/66/sample.pdb -2275.8300
…
```

Parallel Execution

- **Parallelism within CHARMM/Amber:**

  parallel molecular dynamics / minimization

  tightly coupled, based on MPI

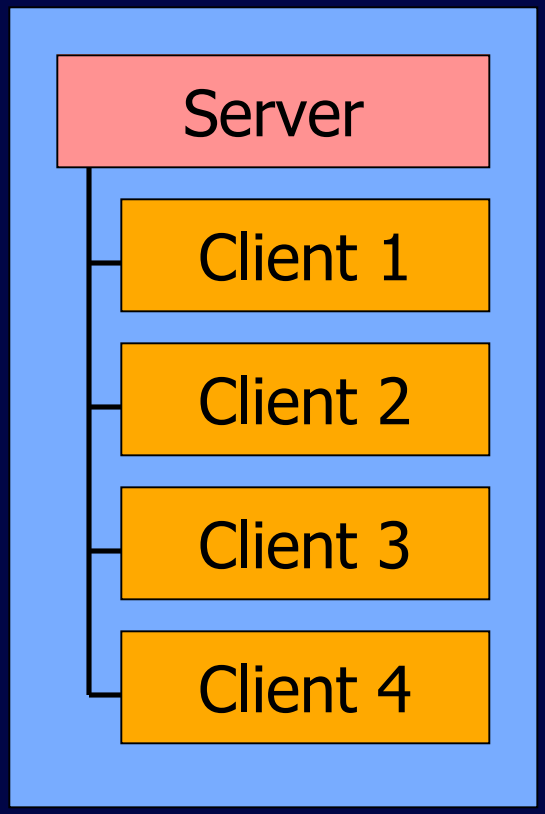
```
setenv CHARMMEXEC "mpirun -n 8 /usr/local/bin/charmm.exe"
setenv SANDEREXEC "mpirun -n 8 /usr/local/bin/sander.exe"
```

- **Parallelism within MMTSB Tool Set**

  ensemble computing, replica exchange simulations

  TCP/IP socket-based server/client architecture

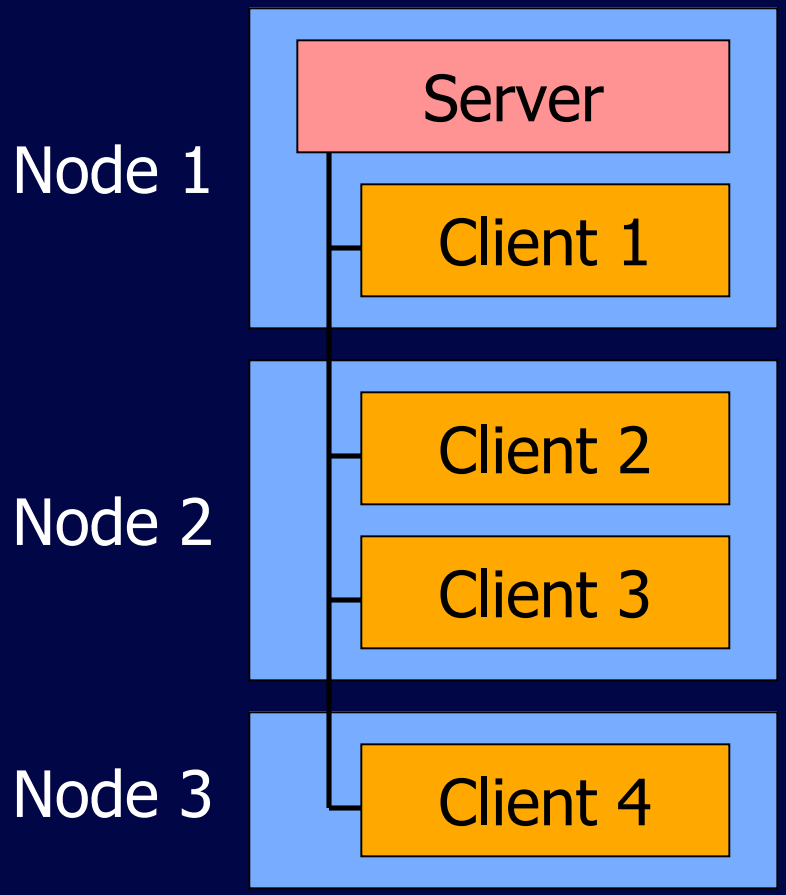
MMTSB Parallel Options: Shared Memory Architecture




```
> ensrun.pl -cpus 4 …
> aarex.pl -temp 4:300:350 …
```

use 4 CPUs for 4 clients + server on SMP architecture

MMTSB Parallel Options: Distributed Environment




```
> ensrun.pl -cpus 4 -hosts hlist …
> ensrun.pl -cpus 4 -mp -hosts hlist …
```

Host file:

```
node1 1 /tmp/mmtsb
node2 2 /tmp/mmtsb
node3 1 /tmp/mmtsb
```

uses 4 CPUs on 3 nodes

requires rsh/ssh without password