THE UNIVERSITY of TEXAS

HEALTH SCIENCE CENTER AT HOUSTON

SCHOOL of HEALTH INFORMATION SCIENCES

Molecular Dynamics Simulation: Analysis

For students of HI 6001-100 “Biomolecular Modeling”

Willy Wriggers, Ph.D.

School of Health Information Sciences

http://biomachina.org/courses/modeling/06.html

Practical Tips for the Production Run

1. *Dynamics Restarts*
2. *Trajectory Output*
3. *Pre-Processing of Trajectory Files*
4. *X-PLOR Trajectory Analysis*

1. Dynamics Restarts

Break up dynamics into smaller steps to guard against a system crash

At end of preceding dynamics run:

```
write coordinates
          output = restart.pdb
end
vector do (X = VX) (all)
vector do (Y = VY) (all)
vector do (Z = VZ) (all)
write coordinates
          output = restart.vlo
end
stop
```

At beginning of new run:

```
coordinates
          @restart.pdb

coordinates
          disposition = comparison
          @restart.vlo

vector do (VX = XCOMP) (all)
vector do (VY = YCOMP) (all)
vector do (VZ = ZCOMP) (all)
```

Don’t forget to set
`iasvel = current`
In dynamics statement!

2. Trajectory Output

Binary trajectory files save disk space
Both velocity and coordinate trajectories may be written

Implicit declaration:

```
evaluate ($trajname = “tra.dcd")
evaluate ($veloname = "vel.dcd")
dynamics verlet
          ascii=false          ! binary coordinate trajectory file
          vascii=false         ! binary velocity trajectory file
          nstep=1000
          nprint=10
          timestep=0.001
          iasvel=current       ! assumes velocities initialized
          traj=$trajname       ! coordinates trajectory file name
          velo=$veloname       ! velocity trajectory file name
          nsavc=10             ! frequency of coord. trajectory frames
          nsavv=10             ! frequency of velo. trajectory frames
end
```

2. Trajectory Output

Explicit declaration (coords shown only):

```
evaluate ($dcdname = "out.dcd")
evaluate ($counter = 1)                 ! main loop counter
while ($counter LE 100) loop main
          dynamics verlet
                    nstep=1000
                    nprint=10
                    timestep=0.001
                    iasvel=current      ! assumes velocities initialized
          end
          if ($counter = 1) then        ! must initialize trajectory file
                    write trajectory
                    ascii = false
                    selection = (all)
                    output = $dcdname
                    end
          else
                    write trajectory next end
          end if
          evaluate ($counter = $counter + 1)
end loop main
```

3. Pre-Processing of Trajectory Files

- Some processing of trajectory files may be required to allow analysis or speed up analysis.

3. Pre-Processing of Trajectory Files

Reading Trajectories

The example below reads frames from a fictitious molecular dynamics trajectory of two files until the last frame is reached:

```
read trajectory
  asci=true
  input=pti_00_50.crd
  input=pti_50_100.crd
  begin=1000
  skip=1000
  stop=100000
end

while ($status # "COMPLETE") loop traj
  read trajectory next end
end loop traj
```

(for details see online X-PLOR manual, chapter 11)

3. Pre-Processing of Trajectory Files

Writing Trajectories

The following example reads a set of PDB coordinate files and merges them into a single trajectory file.

```
evaluate ($count=0)
for $1 in ( a.pdb b.pdb c.pdb d.pdb e.pdb ) loop main
  evaluate ($count=$count+1)
  if ($count=1) then
    write trajectory
       output=trajectory.dcd
       ascii=false
    end
  else
    write trajectory
      next
    end
  end
  write trajectory
    reset
  end
end loop main
```

(for details see online X-PLOR manual, chapter 11)

3. Pre-Processing of Trajectory Files

Merging Trajectories

The trajectory output data were originally stored in two files. The first file contains the first 50 psec, the second file the second 50 psec. These two files are then combined into one file.

```
dynamics merge
  ascii=false
  input=pti_00_50.crd
  input=pti_50_100.crd
  begin=1000
  skip=1000
  stop=100000

  oasci=false
  output=pti_00_100.crd
end
```

(for details see online X-PLOR manual, chapter 11)

4. X-PLOR Trajectory Analysis

Statistical Tools

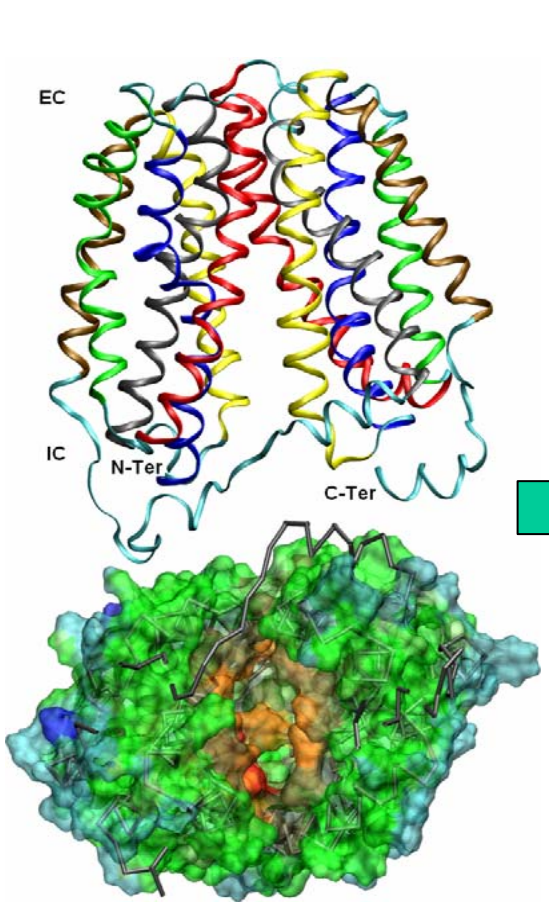
In addition to the standard geometric display and analysis tools described in the last session (which also can be used on trajectory files in combination with ‘read trajectory’), X-PLOR offers certain statistical tools for trajectory analysis:

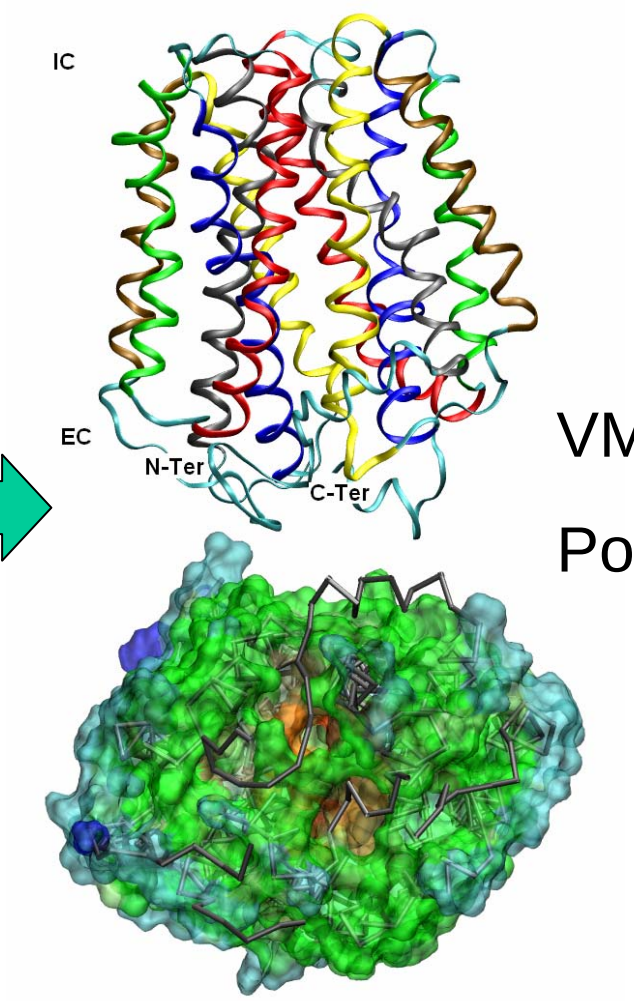
- Average coordinates and fluctuations
- Density analysis
- Covariance Analysis
- Time Correlation Analysis
- Radial Distribution Functions
- Angular Distribution Functions
- Power Spectrum Analysis

(for details see online X-PLOR manual, chapter 11)

General Analysis

Pretty Pictures





VMD

Pov-Ray

Start

20ns

RMS Deviation from Start Structure

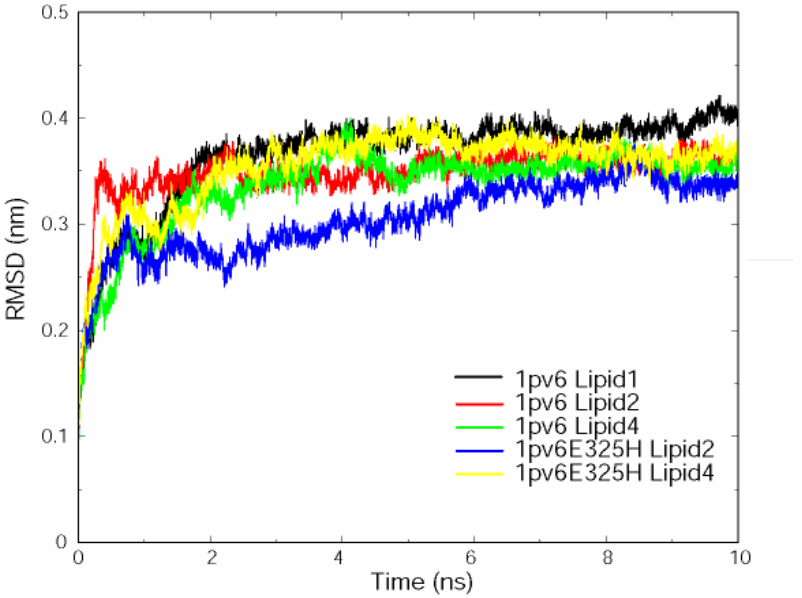
```
coordinates
        disposition = comparison
        @start.pdb
end
```

```
set display = "rmsd.dat" end
evaluate ($frame = 1)
while ($frame LE 60) loop frameloop

  read trajectory
          ascii=false
          input=input.dcd
  end

  coordinates fit
          selection = (all)
          lsq = true
  end

  !calculate rms deviation
  coordinates
          rms
          selection = (all)
  end
  evaluate ($deviation = $RESULT)

  !output results
  evaluate ($runtime = 10 * $frame)
  evaluate ($frame = $frame + 1)
  display $runtime $deviation

end loop frameloop
```

Root Mean Square Deviation (RMSD)

- Cα atom RMSD from start structure is a good indication of structure stability and simulation integrity.

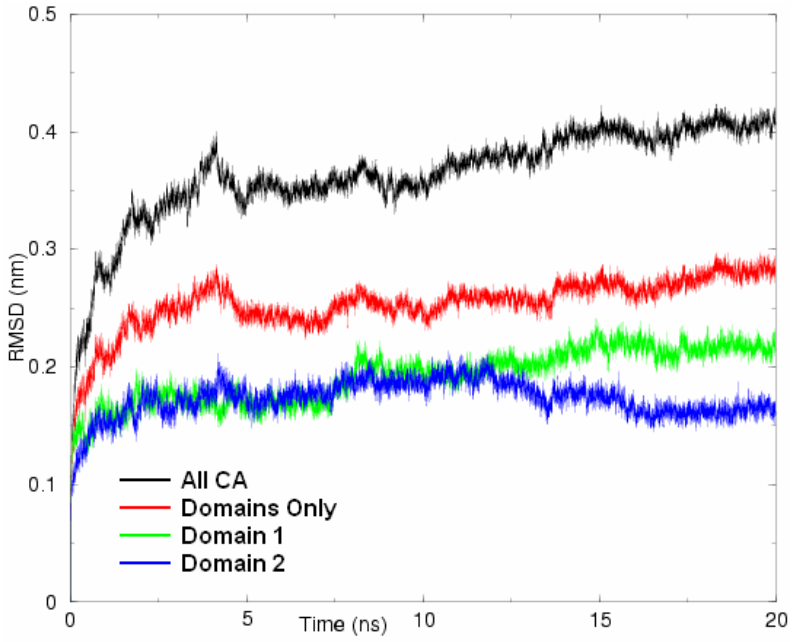
  => Continuous increase indicates sustained changes in structure.



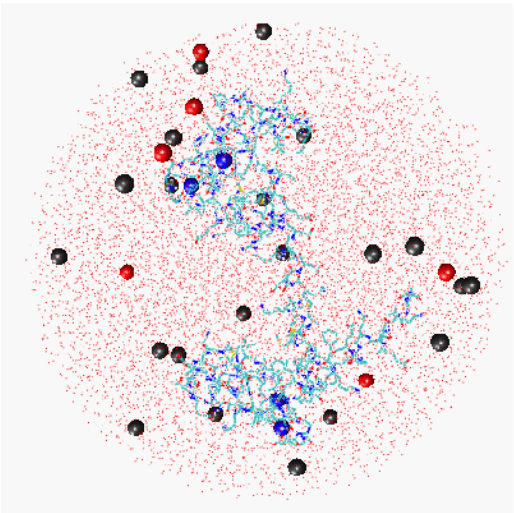
RMSD's Can Be Misleading

- Flexible regions e.g. Large loops, unwound termini can cause large contributions to RMSD.

  => Look at sub selections of Cα atoms e.g. core secondary structure regions.



Calmodulin Example

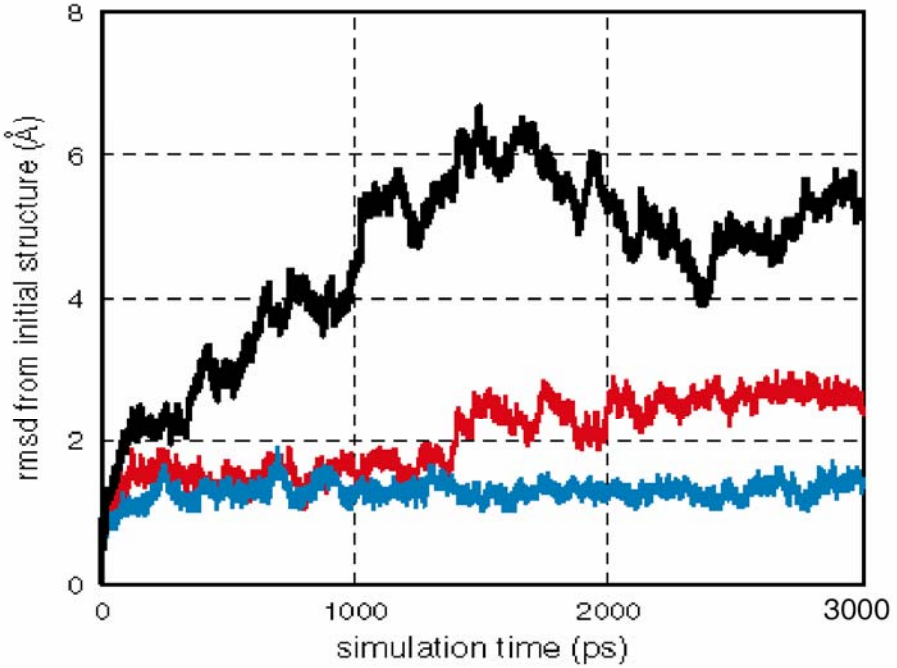


RMS Deviation from Crystal Structure

Total RMSD

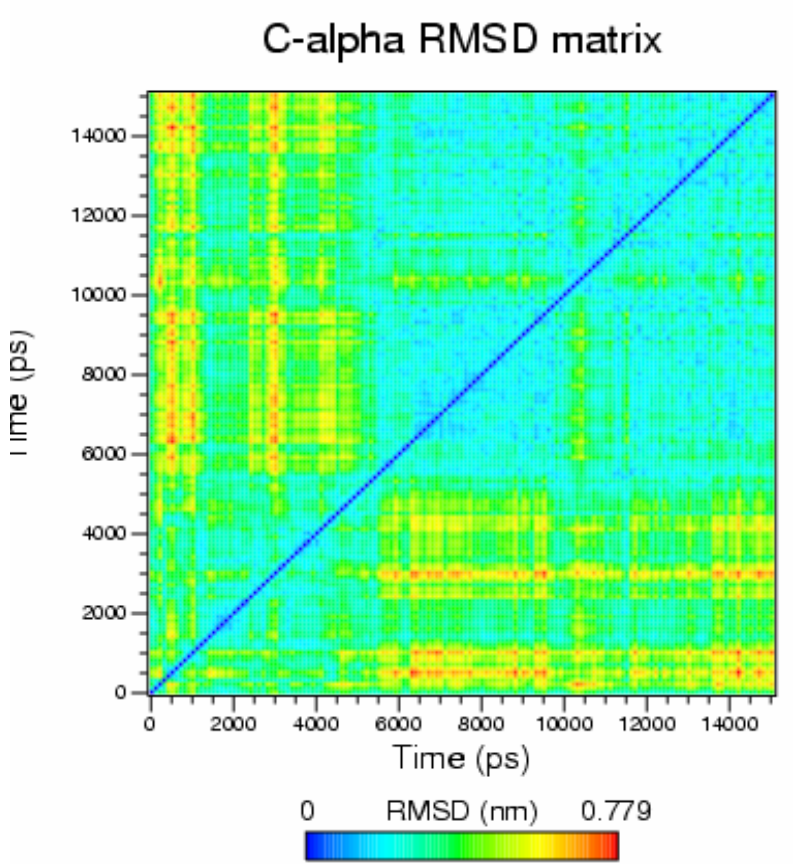
N-terminus domain (1-77)

C-terminus domain (78-148)



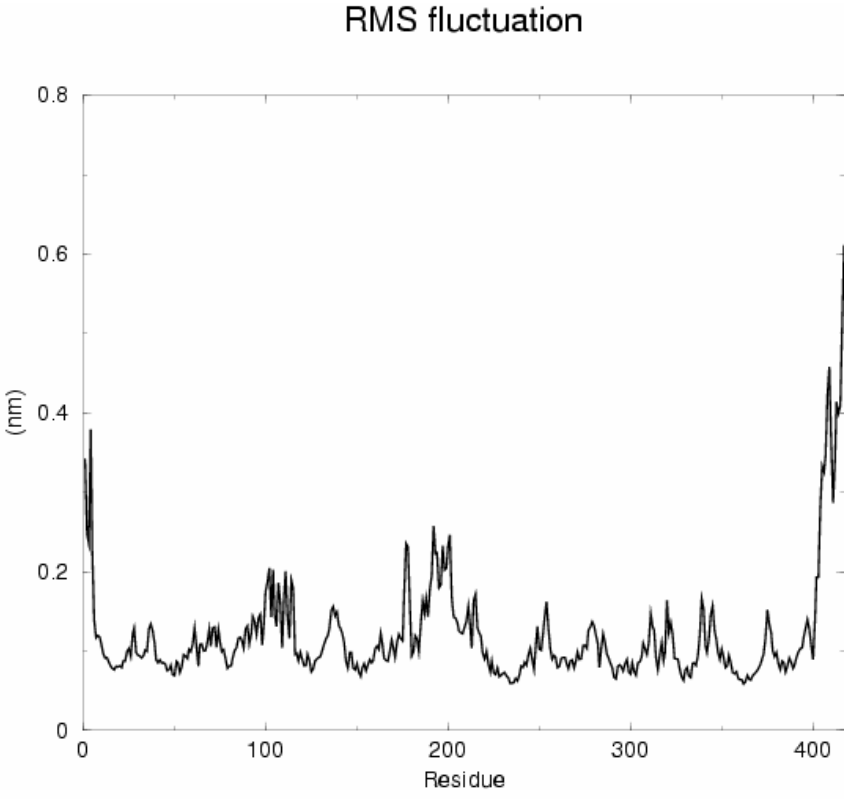
RMSD Matrix

- Allows detection of periodic changes in structure.



Root Mean Square Fluctuations

- Cα RMSF is a measure of the local chain flexibility.
- It is the standard deviation of the atom position calculated from the average structure.



RMS Fluctuations and B-values

```
dynamics merge  ! need to merge trajectories because ...
          ensemble = true
          ascii = false
          input = file1.dcd
          input = file2.dcd
          oasci = false
          output = file3.dcd
end

dynamics analyze average  ! ... this only works for single trajectory file
          ascii = false
          input = file3.dcd
end
! now have average in X,Y,Z and rms fluctuations in B

vector identify (store1) (tag) ! norm fluct over residue
evaluate ($residnr = 0)
for $atom_id in id (store1) loop normrmsf
          vector show norm (B) ((byres((id $atom_id)))and(not(hydrogen)))
          evaluate ($residnr = $residnr + 1)
          display $residnr $RESULT
end loop normrmsf
```

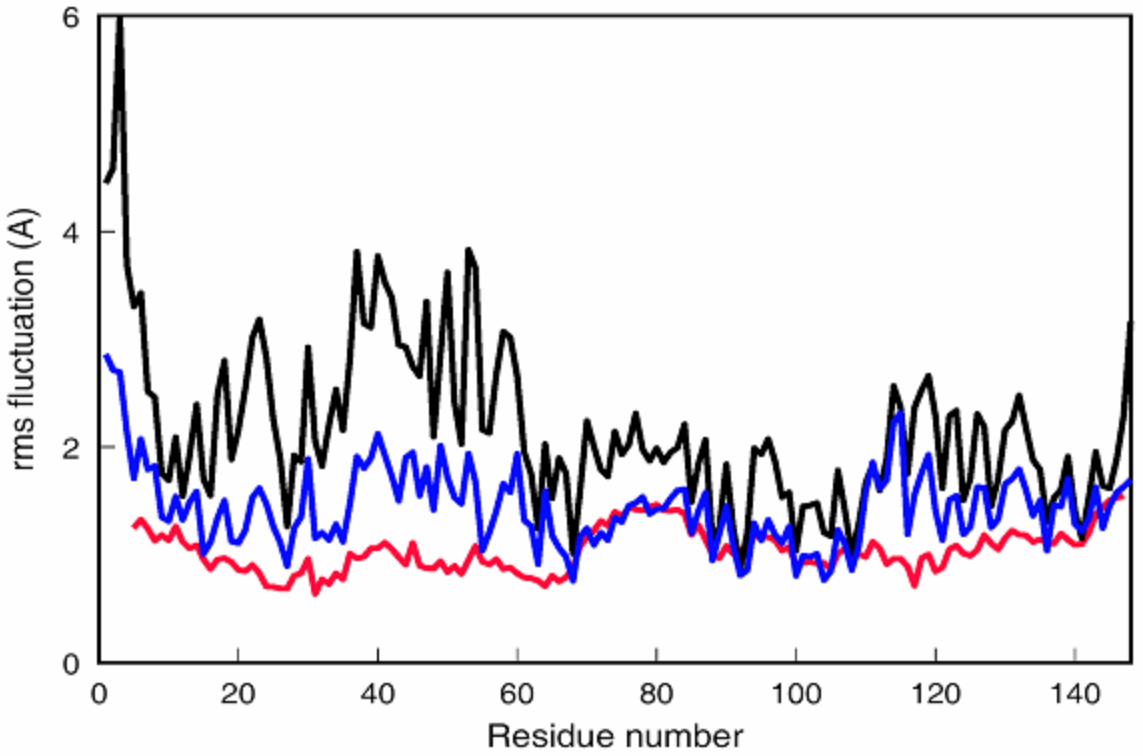
RMS Fluctuations and B-values

Crystal packing forces constrain calmodulin's flexibility:

Fluctuations 0-3ns

Fluctuations 2-3ns

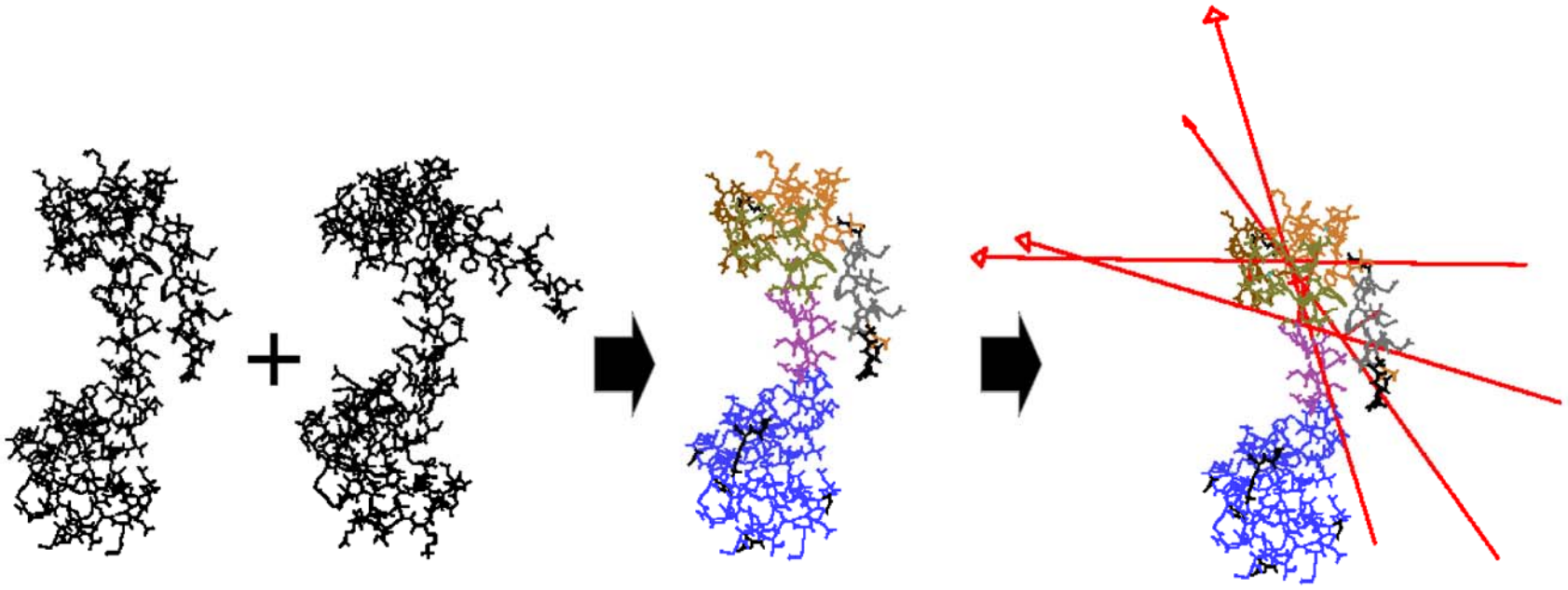
Fluctuations from crystal B-factors:
$B = (8\pi^2/3) \langle(\Delta \mathbf{r})^2\rangle$



Conformational Analysis

Hingefind:

Willy Wriggers and Klaus Schulten. Proteins: Structure, Function, and Genetics 1997, 29:1-14.



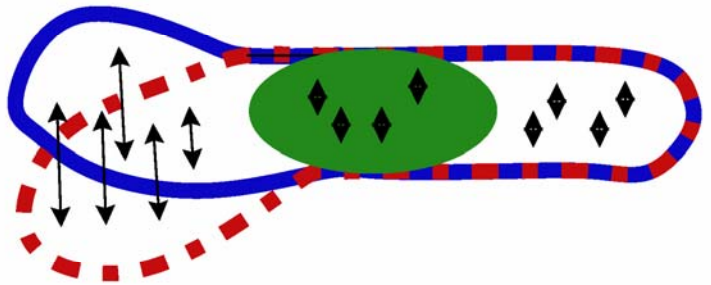
Compare two structures.

Extract rigid domains (regions of preserved packing). Choose resolution to filter out noise from imprecision of coordinates.

Visualize relative movements of rigid domains by effective rotation axes (hinges).

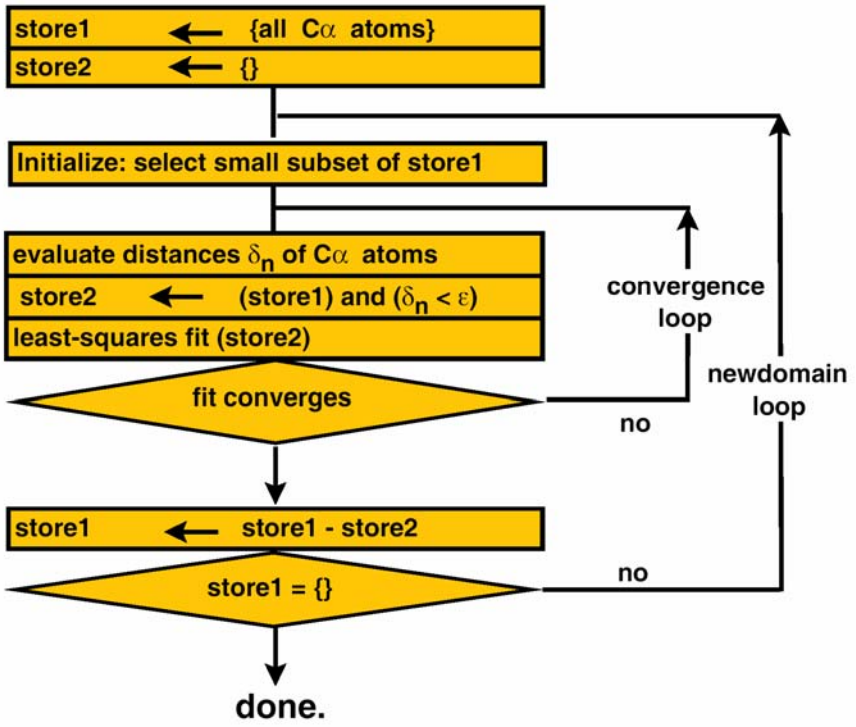
Extracting Rigid Domains

A measure of topological conformance with a subset of atoms:
↔ distances δ_n between pairs of corresponding residues.



Conformation 1
Conformation 2

Iterative *adaptive selection* routine:

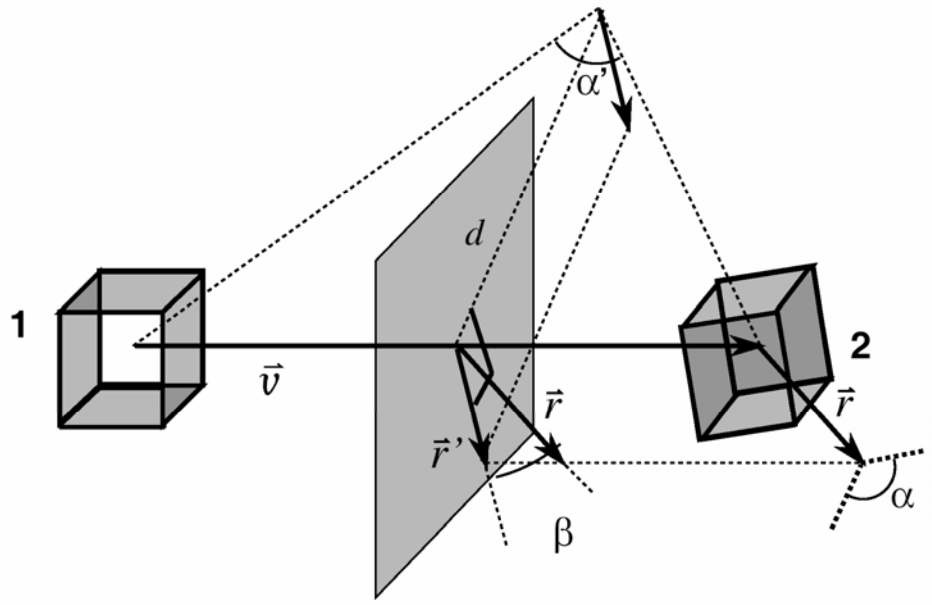


Locating Hinge Axes

Express rigid-body movement (6 degrees of freedom) as a rotation about an effective rotation axis (5 degrees of freedom)

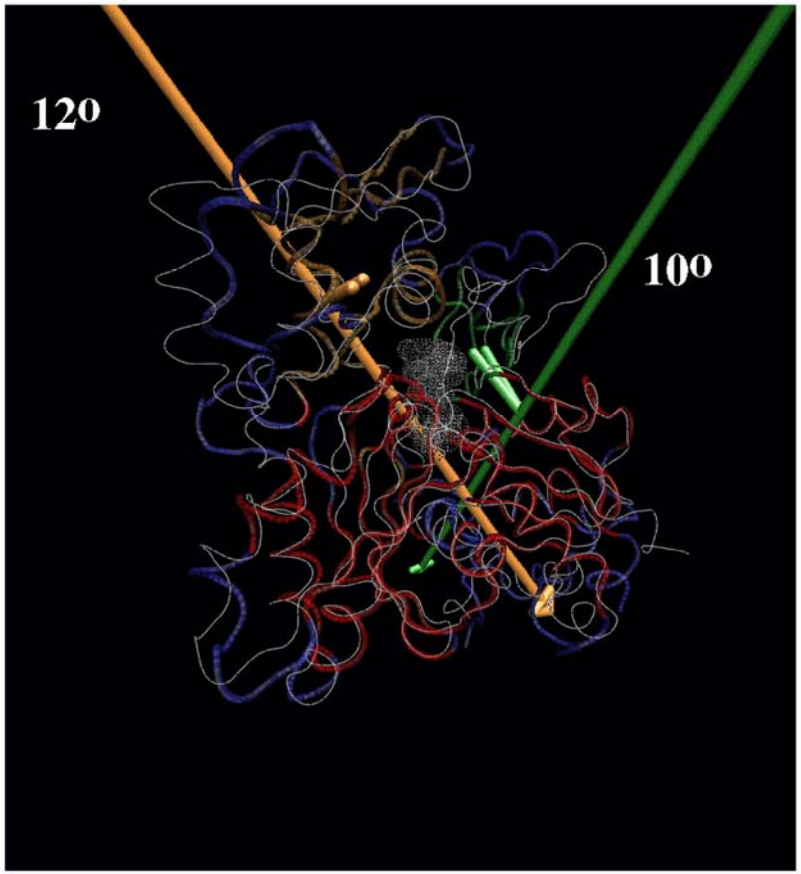
Premise: Domains are connected by flexible joints, which constrain their movement.

Solution: Keep removal of COM translation, but approximate rotation.

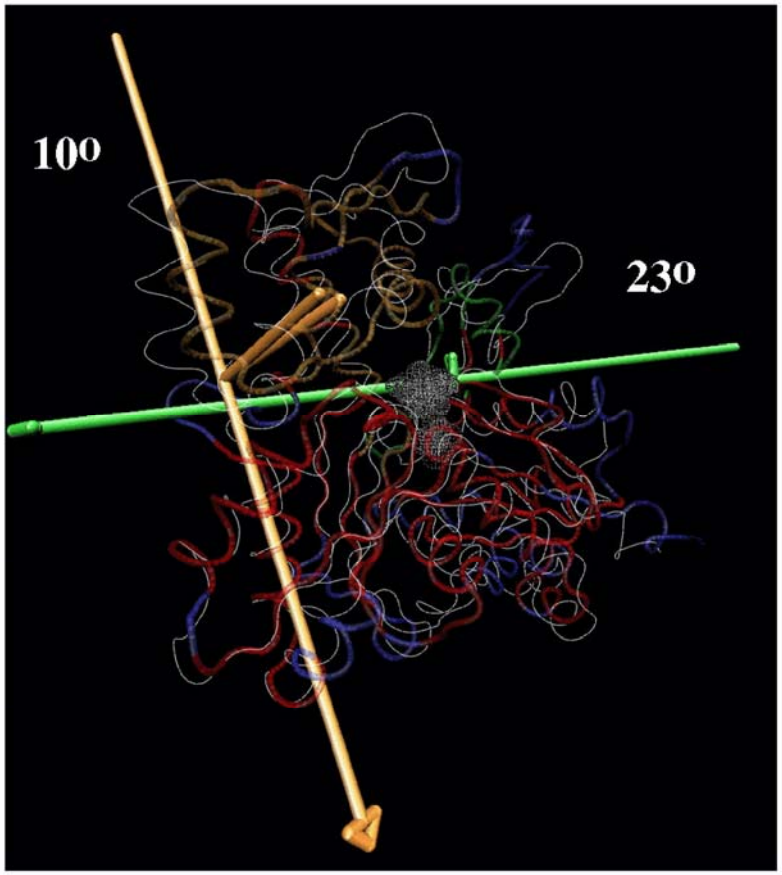


Example

Actin Cleft Closure: MD-Simulation vs. Fiber Diffraction Model of the Filament.



500ps MD simulation vs. Kabsch crystal structure.



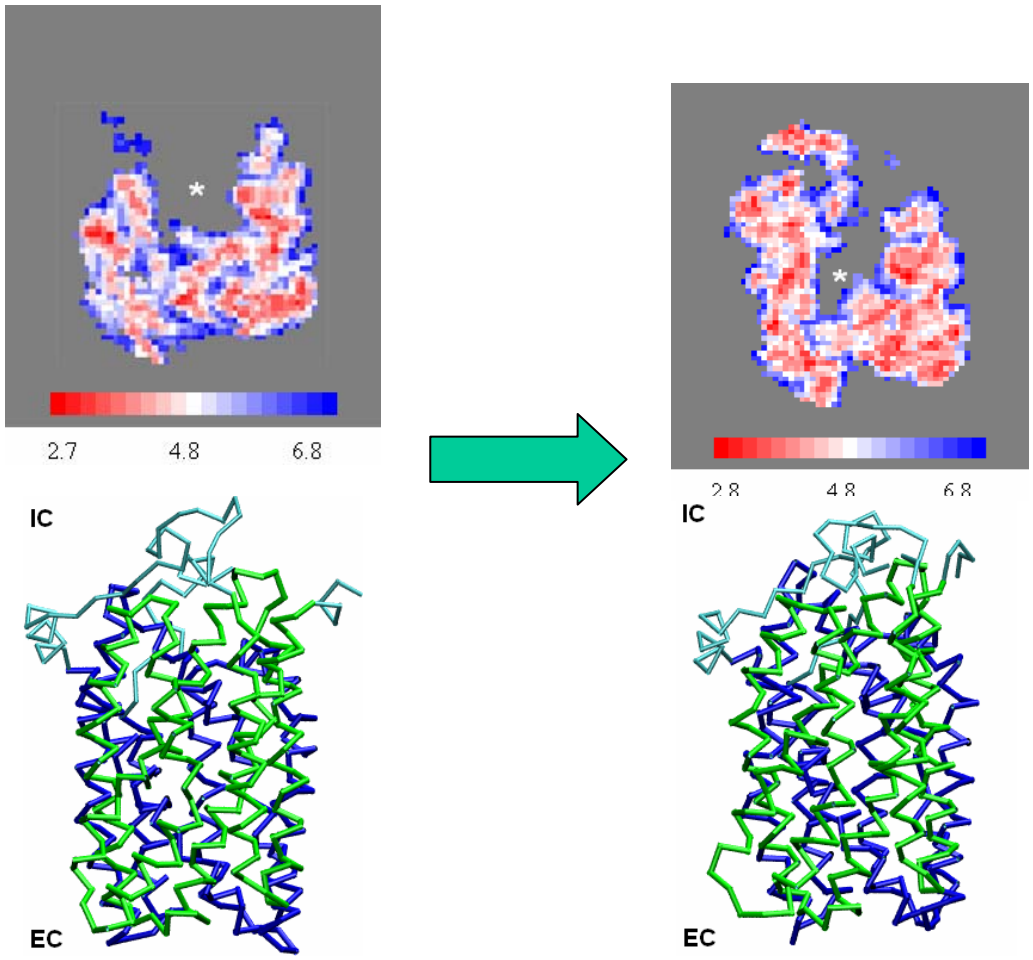
Lorenz filament model vs. crystal structure.

Hingefind Availability

http://www.biomachina.org/disseminate/hingefind/hingefind.html

Tcl (VMD plugin) and X-PLOR (standalone) scripts

Interaction Surfaces



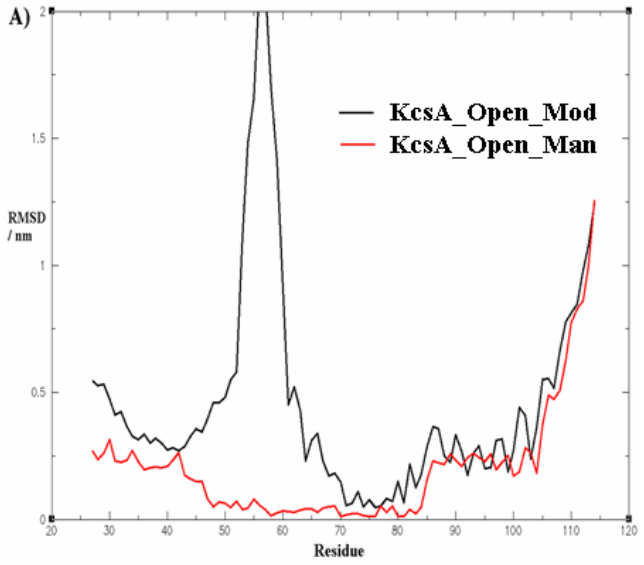
MOLSURFER: characterises interaction surface between domains or two proteins.

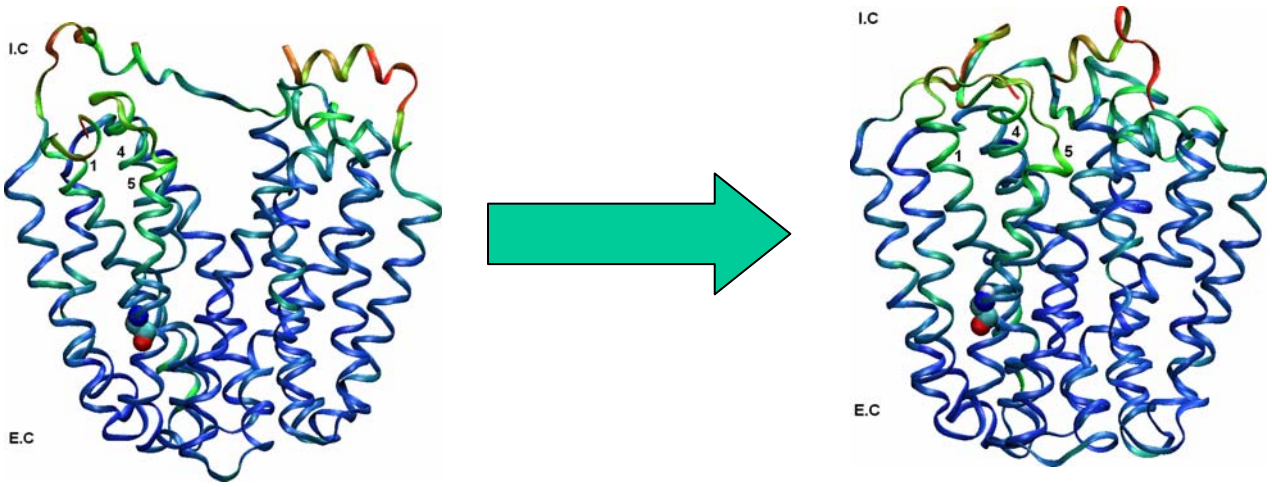
Residue Displacement

Characterizes changes in structure:

Gives an indication which residues are changing position from the start structure.

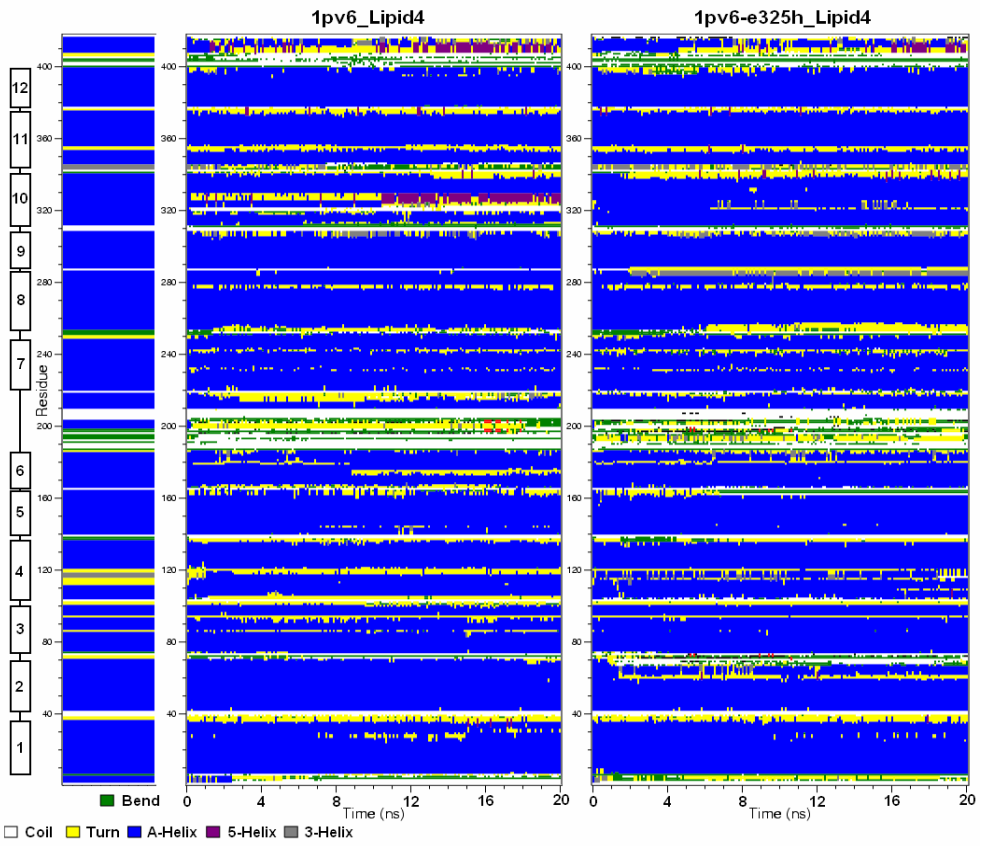
Graph, or display on the structure.





Secondary Structure

- Secondary structure assignment along a trajectory can indicate unstable regions – regions undergoing structural changes.
- Highlights helix breaking effects of prolines.



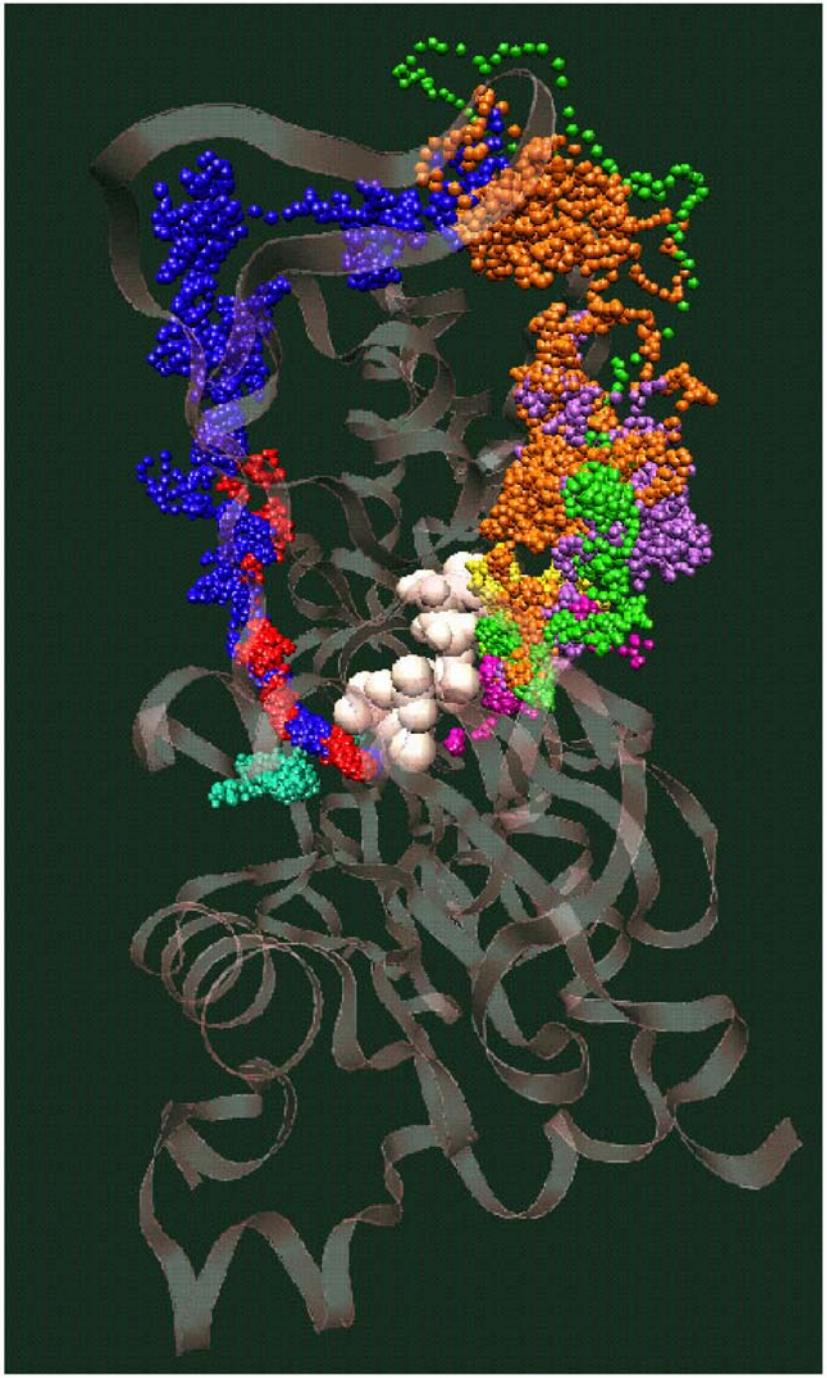
Time-Dependent Quantities

Example:

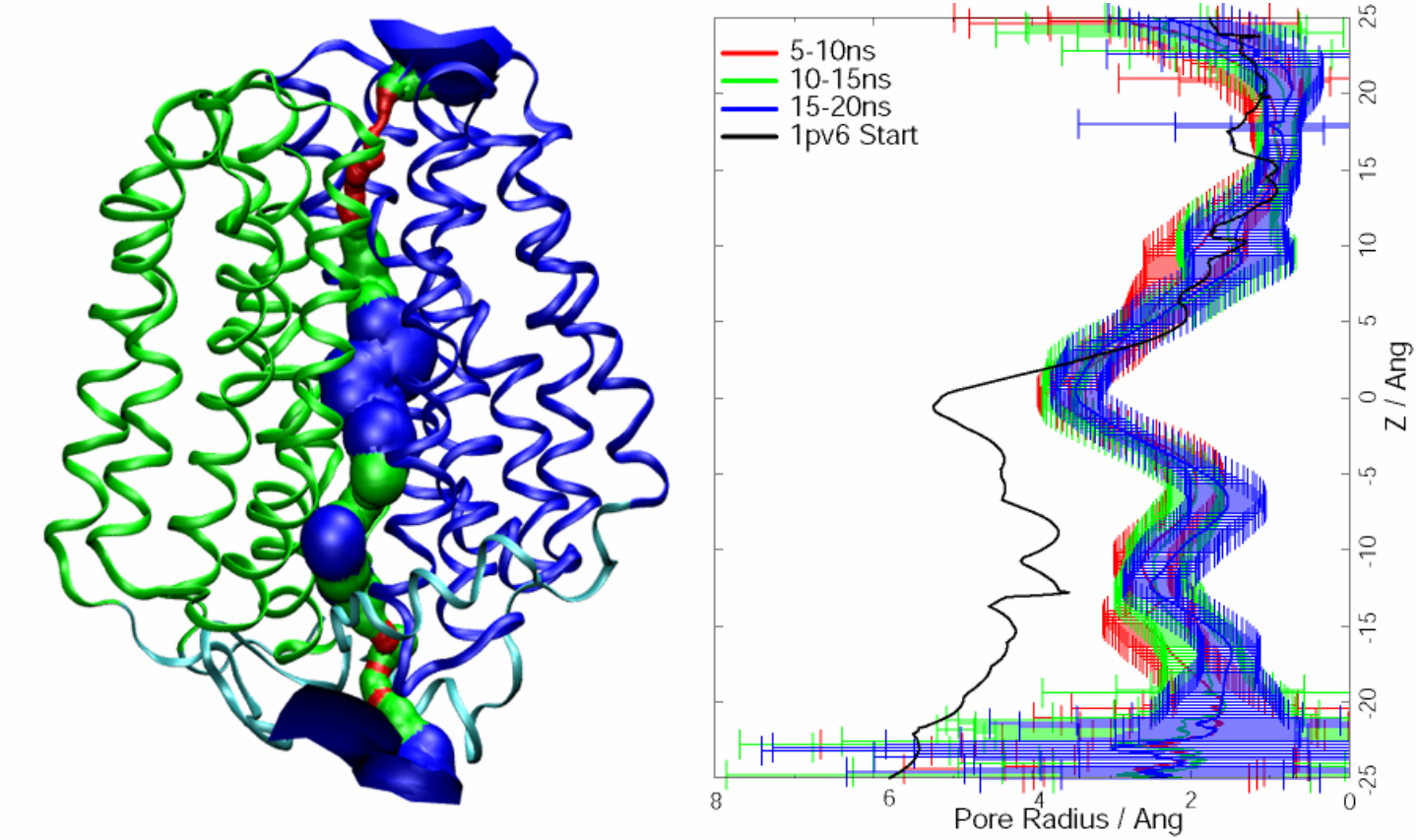
Diffusion of water molecules into actin's enzymatic site

Strategy:

- fit trajectory frames to reference
- extract water oxygen positions



Pore Profiles

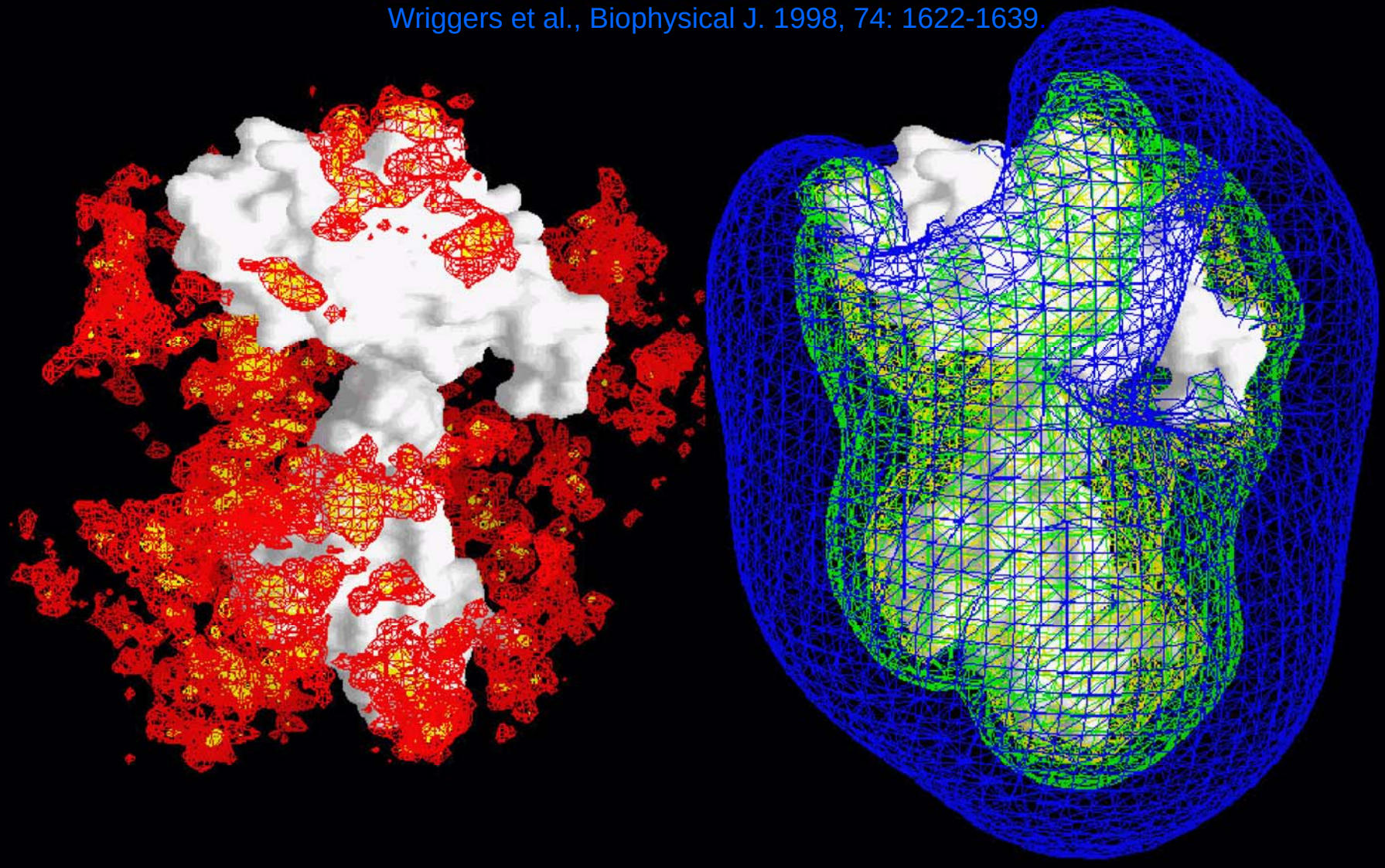


- Hole can calculate the pore radius profile through a protein.
- Surface of the pore can also be visualised
- Useful to look at the average pore profile with standard deviation.

  => Flexibility of the pore

Counterion Distribution

Wriggers et al., Biophysical J. 1998, 74: 1622-1639.



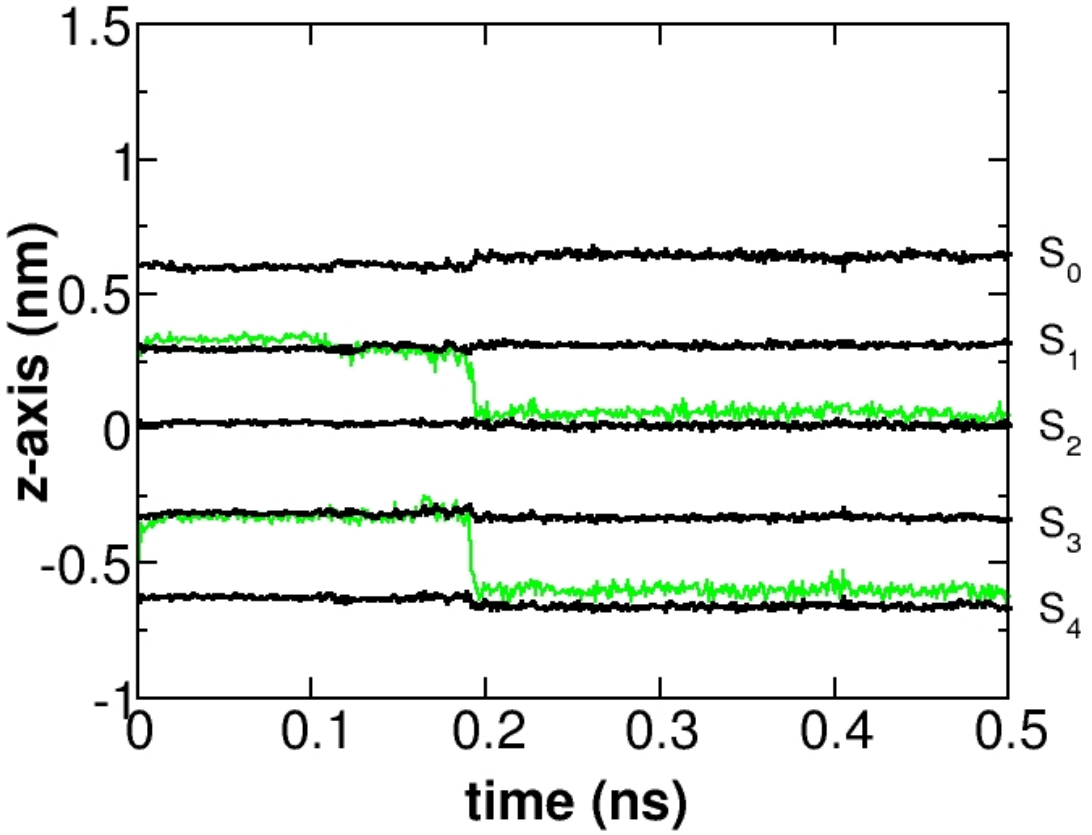
3ns trajectory

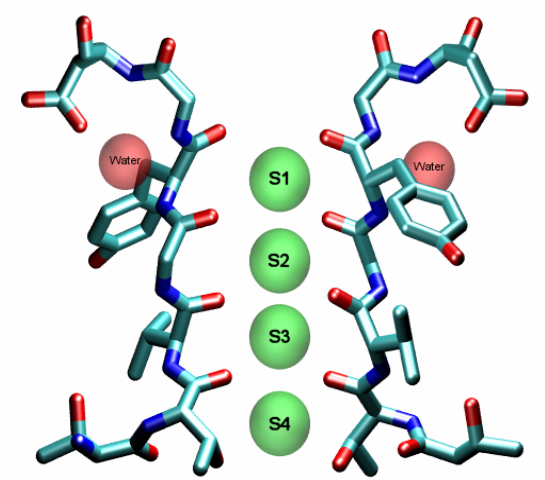
[Na] = 3.0 M, [Na] = 1.0 M

continuum electrostatic theory

[Na] = 0.5 M, [Na] = 0.2 M, [Na] = 0.1 M

Channels: Ion Trajectory





Solvent Diffusion

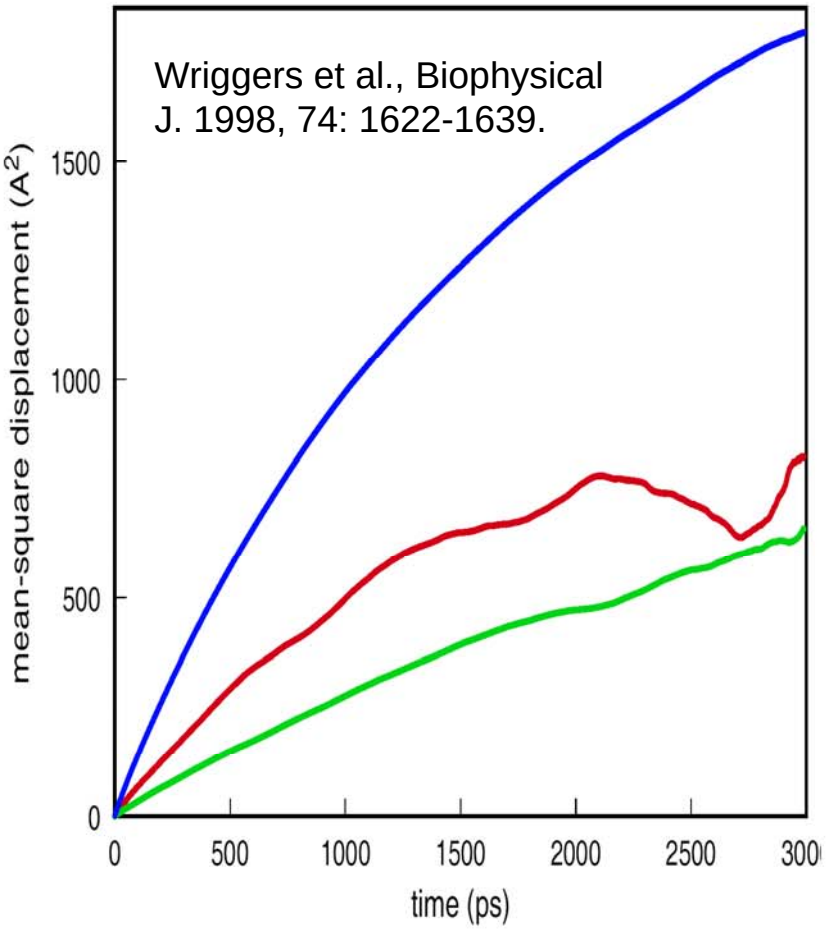
Mean-square displacement:

$$\langle\langle[\mathbf{r}(t+\tau)-\mathbf{r}(t)]^2\rangle_t\rangle_{ens}$$

H_2O

Na^+

Cl^-



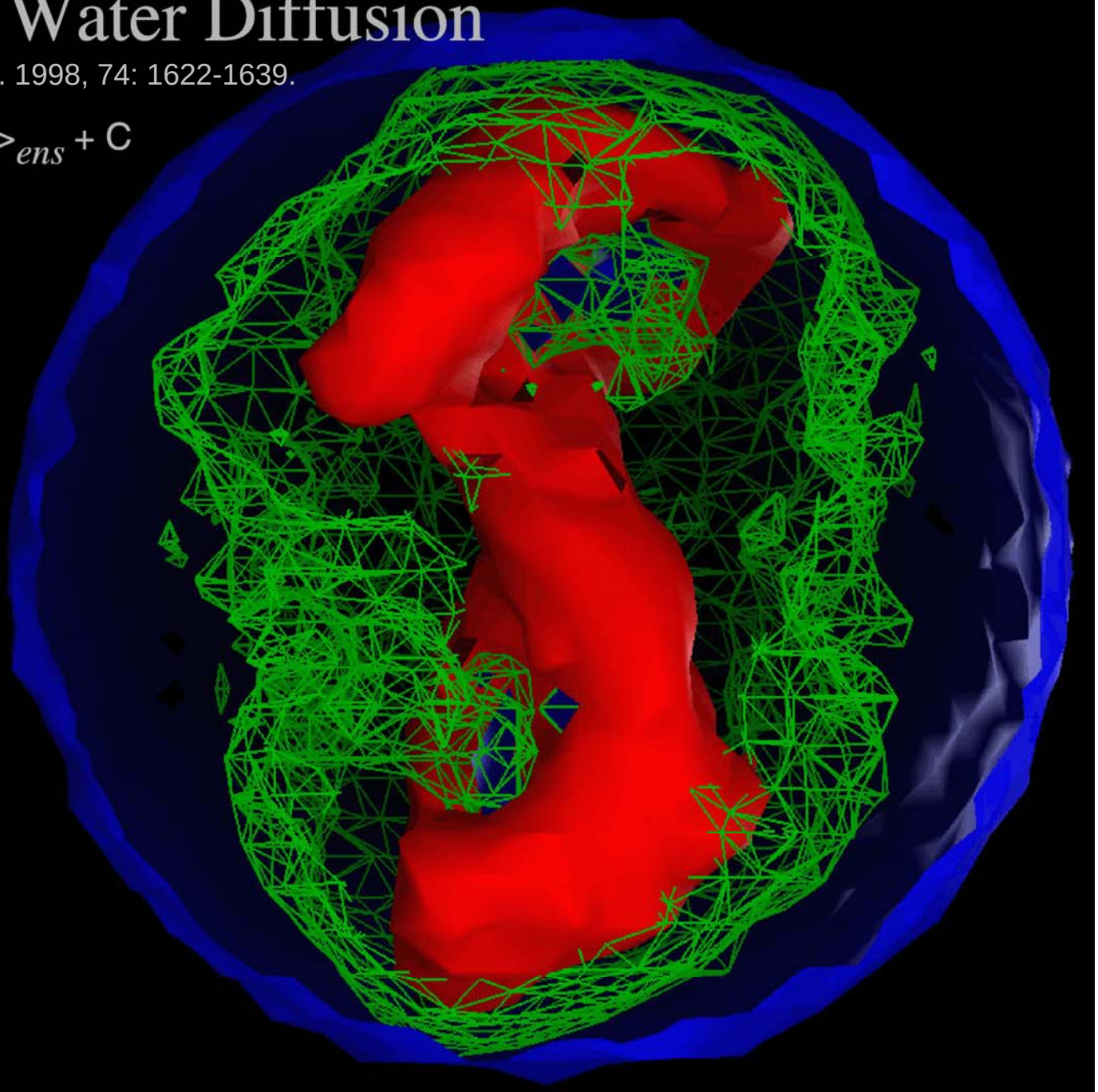
Diffusion constants from Einstein relation: $\lim_{t\to 0} \langle\langle[\mathbf{r}(t+\tau)-\mathbf{r}(t)]^2\rangle_t\rangle_{ens} = 6Dt$

H_2O:	$D = 2.5$ $[10^{-9}\ m^2/s]$	Experiment: $D = 2.3$ [~]	Simulation: $D = 1.3\text{-}4.2$ [~]
Na^+:	$D = 0.53$ [~]	$D = 1.4$ [~]	$D = 0.5\text{-}5$ [~]
Cl^-:	$D = 1.3$ [~]	$D = 2.1$ [~]	---

Translational Water Diffusion

Wriggers et al., Biophysical J. 1998, 74: 1622-1639.

$$6Dt = \langle\langle [r(t+\tau)-r(\tau)]^2 \rangle_\tau \rangle_{ens} + C$$



$D(r) = 2.0 \cdot 10^{-9}\ m^2/s$

$D(r) = 3.0 \cdot 10^{-9}\ m^2/s$

$D(r) = 4.0 \cdot 10^{-9}\ m^2/s$

$D_{exp} = 2.7 \cdot 10^{-9}\ m^2/s$

Protein-Specific Analysis

The analysis carried out will depend on the structure you are looking at, and the features you are exploring.

Requires writing of your own analysis scripts/programs.

Resources and Further Reading

Papers:
http://www.biomachina.org/publications_web/WRIG98B.pdf
http://www.biomachina.org/publications_web/WRIG97.pdf

WWW:
http://cmm.info.nih.gov/intro_simulation
http://xplor.csb.yale.edu/

Books:
Schlick, Chapters 8, 9, 12, 13
Brunger, X-PLOR Version 3.1, Chapters 1-11
online free at http://alpha2.bmc.uu.se/local_html/xplor_mirror.html

Acknowledgements/Credits

Mark Samson, U. Oxford