Analysis of MD Trajectories

Justin Gullingsrud

Beckman Institute
University of Illinois at Urbana-Champaign

Analysis of MD Data

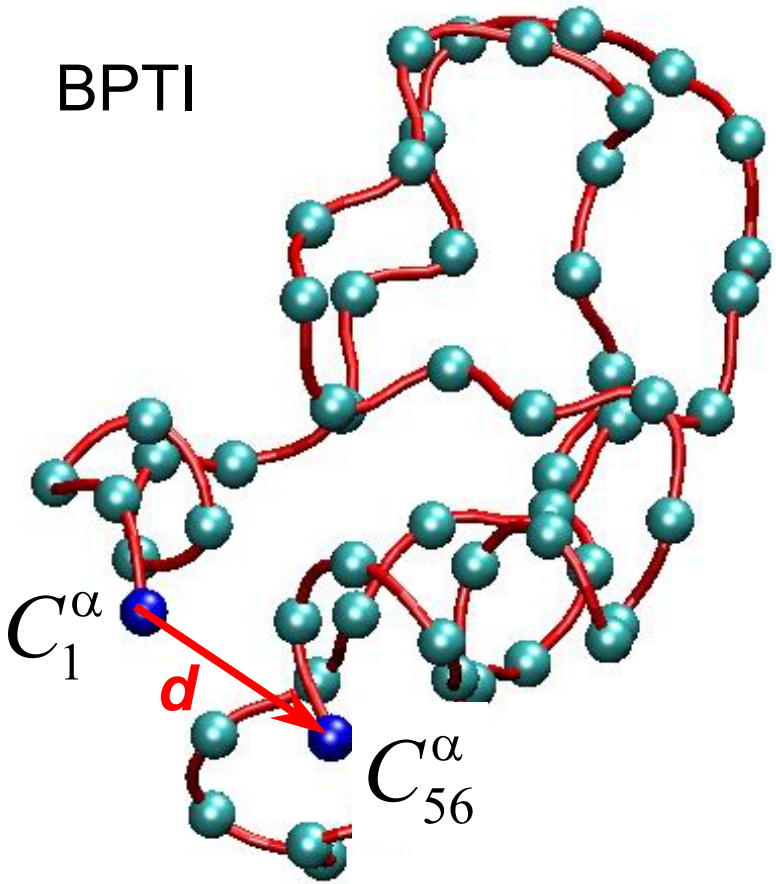
1. Structural properties
2. Equilibrium properties
3. Non-equilibrium properties

Equilibrium properties can be studied via both **equilibrium** and/or **non-equilibrium** MD simulations

Structural Properties of Biopolymers

1. End-to-end Distance
2. Radius of Gyration
3. Mean Square Displacement (MSD)
4. Root Mean Square Deviation (RMSD)
5. Debye-Waller Factor

1. End-to End Distance



Represents the average distance ***d*** between the first and last segment of a (bio)polymer

Suitable to describe linear polymers

2. Radius of Gyration

$$R_G = \sqrt{\sum_{a=1}^{N} m_a \left(\boldsymbol{r}_a - \boldsymbol{r}_{COM}\right)^2 \Big/ \sum_{a=1}^{N} m_a}$$

Mass weighted RMS average distance of the selected atoms from their center of mass (COM)

Suitable to describe branched chains with large number of ends

3. Mean Square Displacement

$$MSD = \left\langle \sum_{a} (\boldsymbol{r}_a - \boldsymbol{r}_{a0})^2 \right\rangle$$

Describes the “distance” between two conformations of a (bio)polymer (or group of selected atoms)

First the two conformations must be aligned

4. Root Mean Square Deviation

$$RMSD = \left[\left\langle \sum_{a} (\boldsymbol{r}_a - \boldsymbol{r}_{a0})^2 \right\rangle \right]^{1/2}$$

Describes the “distance” between two conformations of a (bio)polymer (or group of selected atoms)

First the two conformations must be aligned

5. Debye-Waller Factor

(B- or Temperature Factor)

$$B_a = \frac{8}{3}\pi^2 \langle \delta \boldsymbol{r}_a^2 \rangle, \quad \delta \boldsymbol{r}_a = \boldsymbol{r}_a - \langle \boldsymbol{r}_a \rangle_t$$

Describes the reduction of the intensity of Bragg scattering due to motion of atoms about their equilibrium position

Atomic scattering factor:

$$f = f_0 \cdot \exp[-B(\sin\phi/\lambda)^2]$$

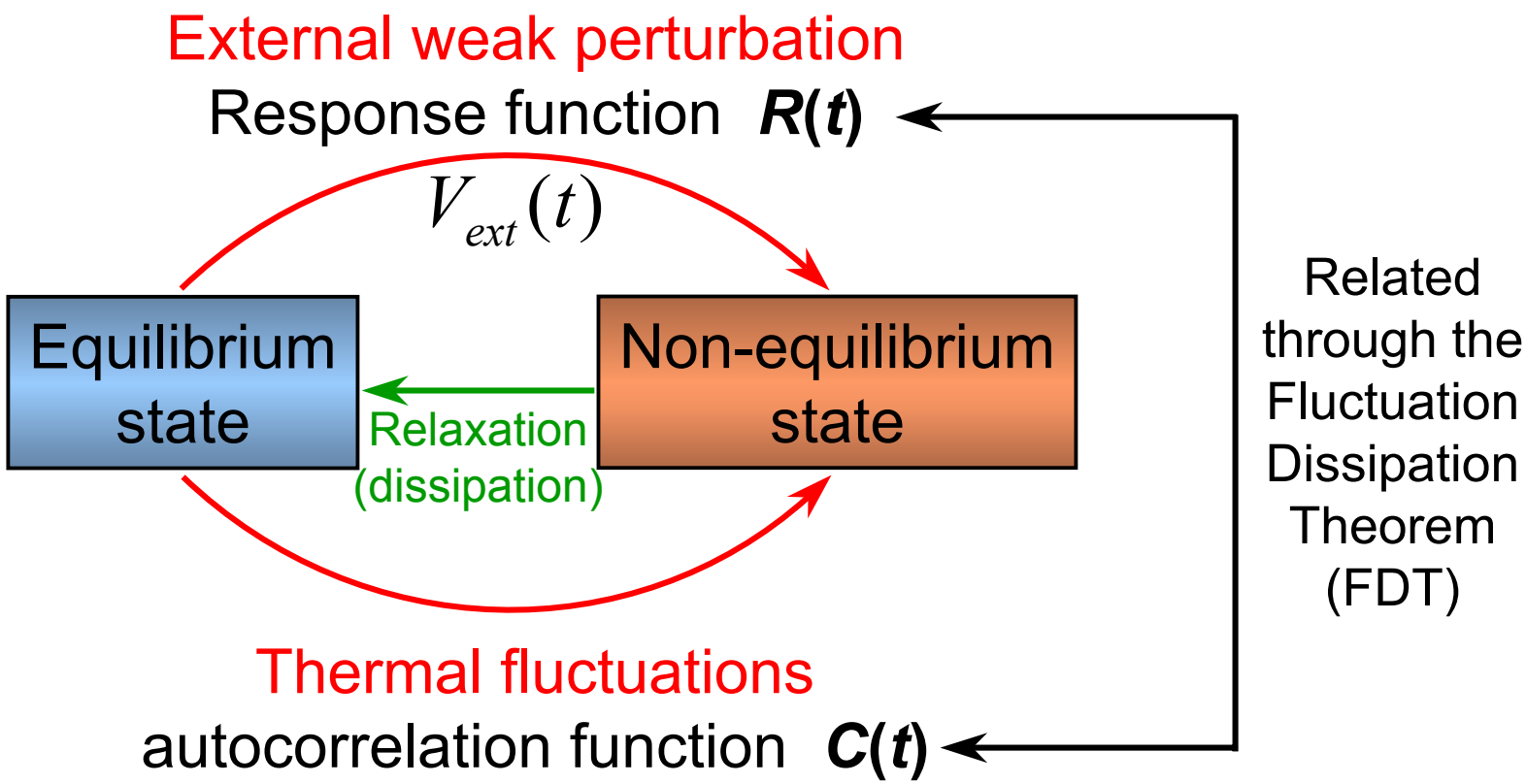
Does not vanish even at T=0 because of the zero point motion of the atoms !

Non-equilibrium Properties

1. Transport properties
2. Spectral properties

Can be obtained from *equilibrium* MD simulations by employing *linear response theory*

Linear Response Theory



Time Correlation Functions

$$C_{AB}(t-t') = \underbrace{\langle A(t)\,B(t')\rangle = \langle A(t-t')\,B(0)\rangle}_{\text{since } \rho_{eq} \text{ is } t \text{ independent !}}$$

$A \neq B$ cross-
$A = B$ auto- } correlation function

Correlation time: $\tau_c = \int_0^\infty dt\, C_{AA}(t) / C_{AA}(0)$

Estimates how long the “memory” of the system lasts

In many cases (but not always): $C(t) = C(0)\exp(-t/\tau_c)$

Response Function

or *generalized susceptibility*

External perturbation: $V_{ext}(t) = -A \cdot f_{ext}(t)$

Response of the system: $\langle A(t) \rangle = \int_0^t dt' \, R(t-t') \, f_{ext}(t')$

Response function: $R(t) = \langle \{A(t), A\}_{PB} \rangle = -\beta \langle \partial_t A(t) \, A \rangle$

with $\beta = 1/k_B T$

Generalized susceptibility: $\chi(\omega) \equiv R(\omega) = \int_0^\infty dt \, e^{i\omega t} \, R(t)$

Rate of energy dissipation/absorption:

$$Q_\omega \equiv \overline{\langle A(t) \rangle \frac{df}{dt}} = \frac{1}{2} \omega \chi''(\omega) \, | f_0 |^2, \quad f(t) = \mathrm{Re} \, f_0 \, e^{-i\omega t}$$

Fluctuation-Dissipation Theorem

Relates $R(t)$ and $C(t)$, namely:

$$\chi''(\omega) = (\beta\omega/2)\,C(\omega)$$

In the static limit ($t \to \infty$): $C(0) = \langle A^2 \rangle = k_B T\, R(0)$

Note: quantum corrections are important when $k_B T \leq \eta\omega$

$$\chi''(\omega) = \eta^{-1} \tanh(\beta\eta\omega/2)\,C(\omega)$$

Example: Absorption of Radiation by Electric Dipoles

Perturbation: $V_{ext}(t) = -\boldsymbol{P} \cdot \boldsymbol{E}(t), \quad E(t) = E_0 \, \hat{\boldsymbol{e}} \cos \omega t$

Correlation function: $C(t) = 1/3 \langle \boldsymbol{P}(t) \, \boldsymbol{P} \rangle$

Absorption coefficient: $\alpha(\omega) = (4\pi\omega / c)[\chi''(\omega) / \varepsilon'(\omega)]$

Applying the FDT:

$$\alpha(\omega) = (2\pi\omega^2\beta / c)[C(\omega) / \varepsilon'(\omega)]$$

P(t), and *C*(*t*) can be computed from a suitable MD trajectory

Diffusion Coefficient

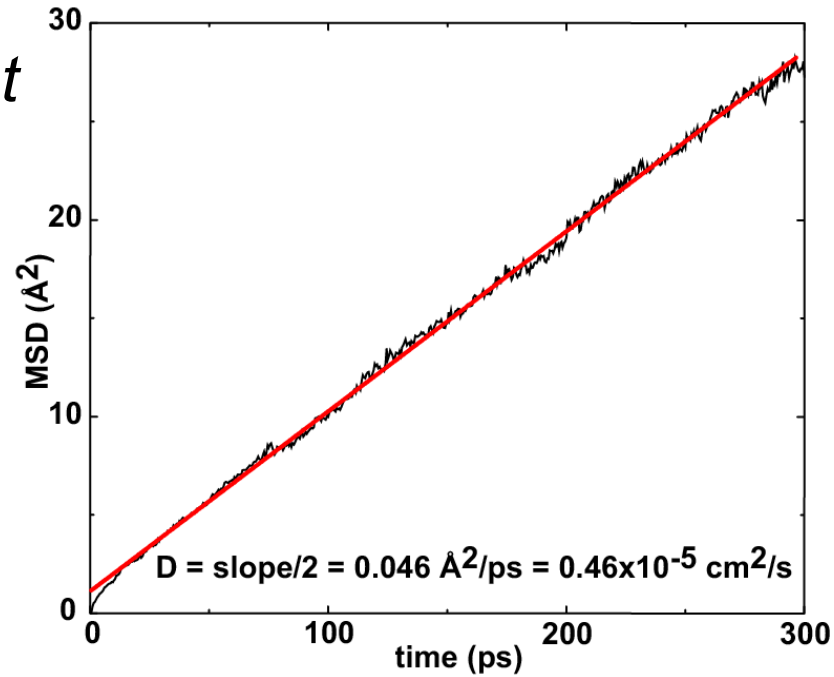
Generic transport coefficient: $\gamma = \int_0^\infty dt \langle \partial_t A(t) \partial_t A(0) \rangle$

Einstein relation: $2\gamma t = \langle [A(t) - A(0)]^2 \rangle$

Example: *self-diffusion coefficient*

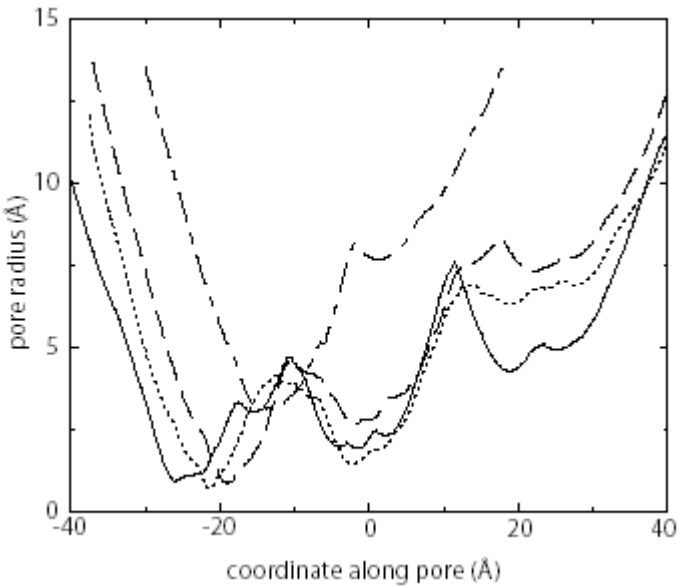
$$D = \frac{1}{3} \int_0^\infty dt \langle \boldsymbol{v}(t) \boldsymbol{v}(0) \rangle$$

$$6Dt = \langle | \boldsymbol{r}(t) - \boldsymbol{r}(0)]^2 \rangle$$



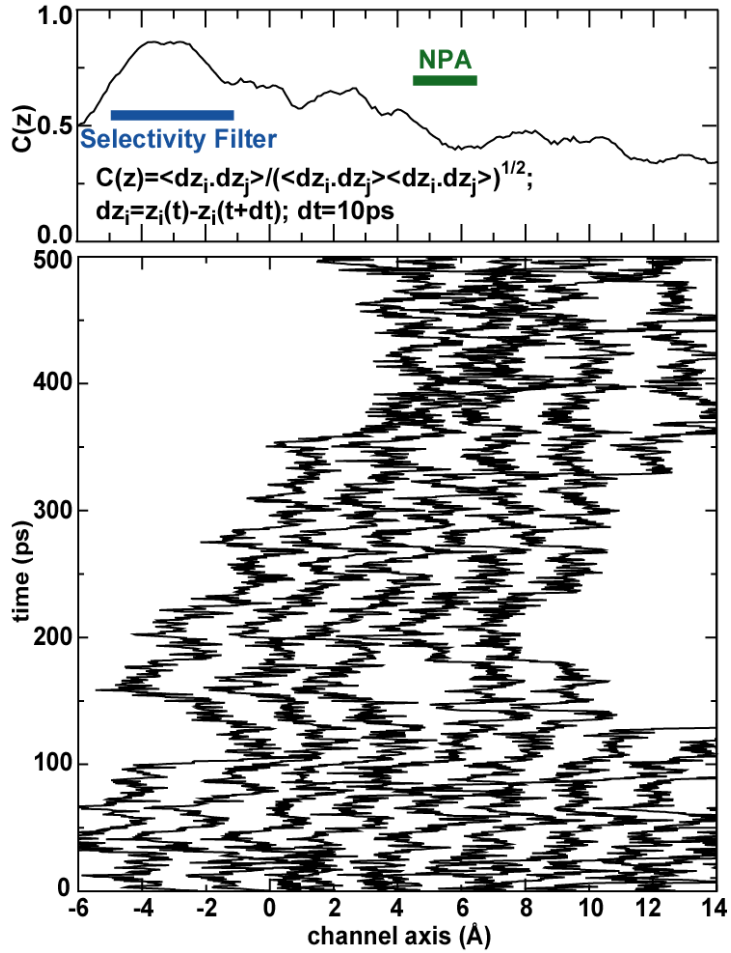
Analyzing Ion Channels

- The program HOLE can be used to obtain an estimate of pore radii along an axial coordinate.
- Other properties: diffusion coefficient, DELPHI analysis…



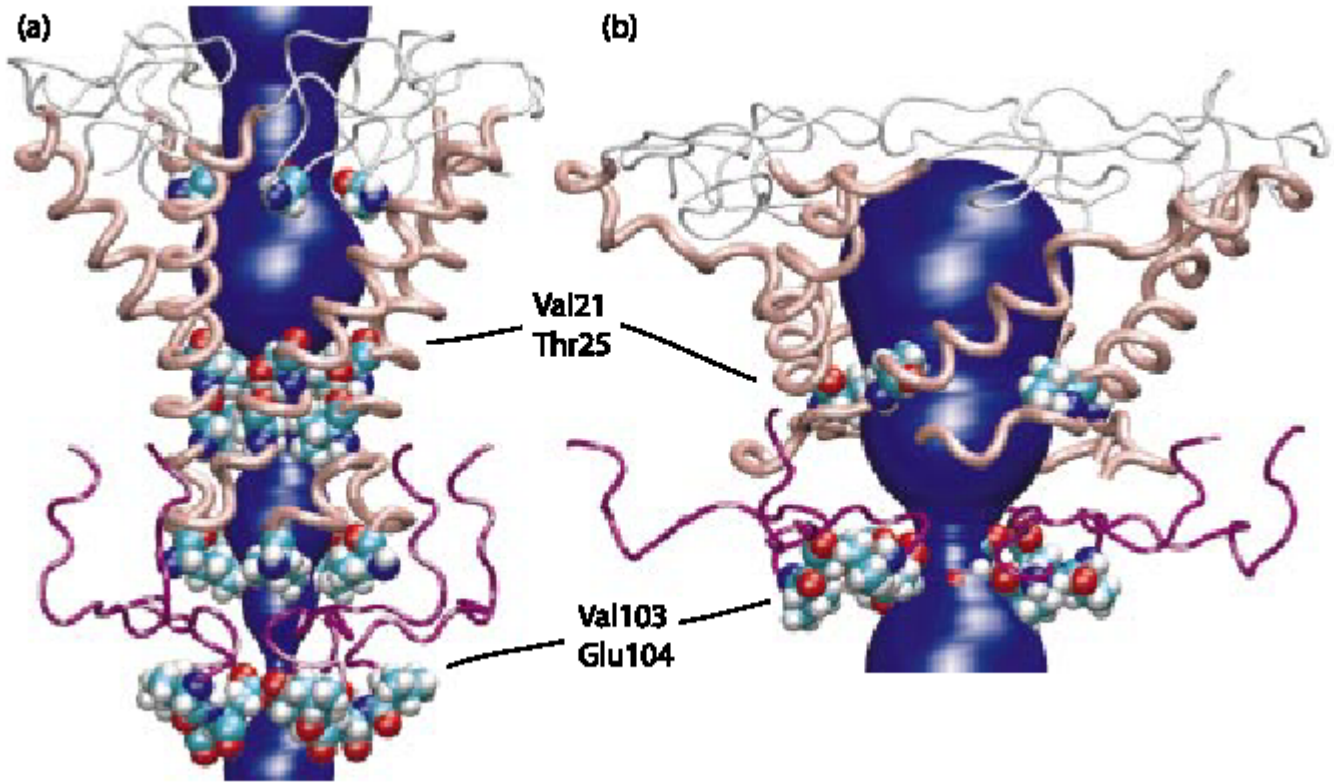
Analysis of water in channels

- Analysis of single water molecules demonstrates single-file motion.
- Correlation functions can be obtained directly from the trajectories.



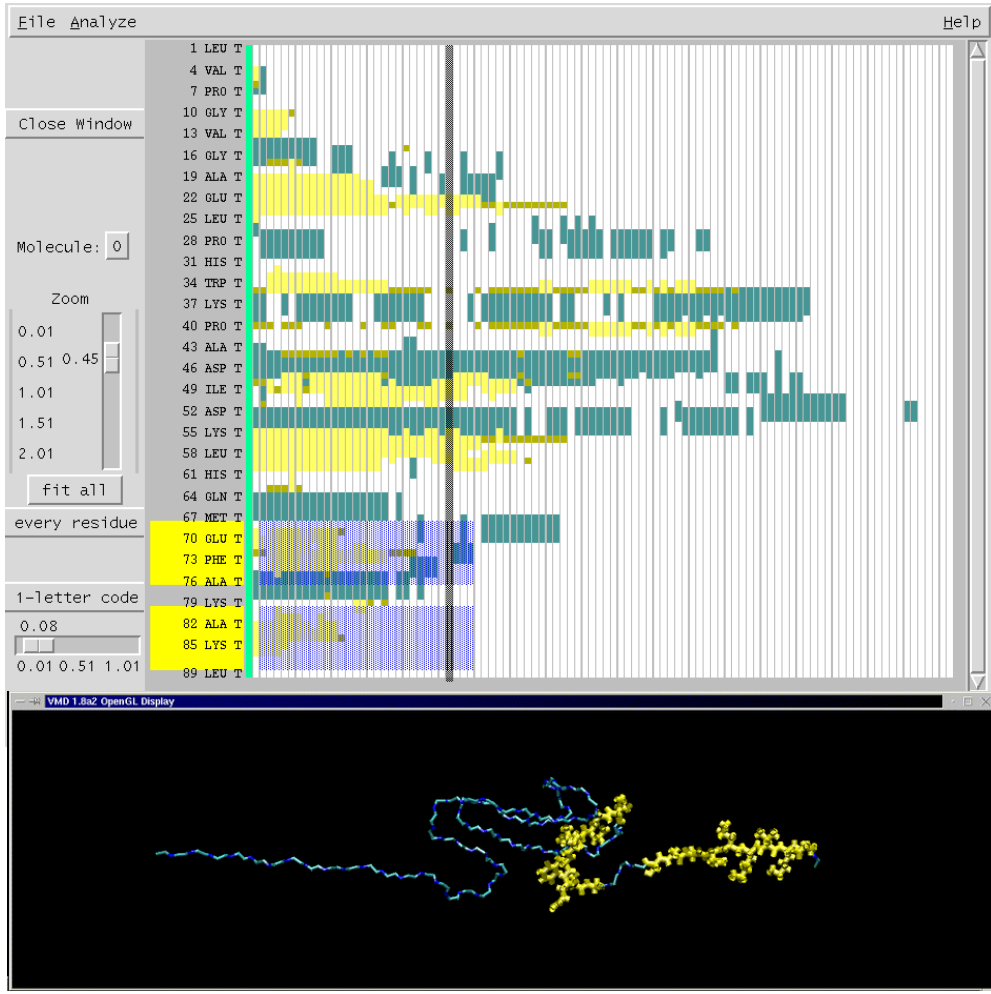
Combining HOLE with VMD

Data from HOLE can easily be imported into VMD for structural analysis.



Secondary Structure Analysis

- Secondary structure trajectories can be displayed along with coordinate data.



Trajectory Analysis in Practice

- Trajectory files are often very large, requiring special tools to work with them.
- Quite a bit of trajectory analysis can be done within VMD.
  - Use fast atom selections
  - Take advantage of fast methods (think vector code)
- Environments like Matlab are also well suited for trajectory analysis.

Collection of MD Data

- DCD trajectory file
  - coordinates for each atom
  - velocities for each atom
- Output files
  - global energies
  - temperature, pressure, …
  - unit cell dimensions

CatDCD

- Merge or split DCD files from NAMD:

```
catdcd -o simA.dcd simA-01.dcd simA-
  02.dcd simA-03.dcd
```

- Create a new DCD file containing only selected atoms
  - Saves memory
  - Makes atom selections go faster

```
catdcd -I protein.ind -o protein.dcd
  simA.dcd
```

Other uses for CatDCD

- Count how many frames are in a DCD file:

```
catdcd -num min_all.dcd
catdcd -num *.dcd
```

- Grab the last 5 frames out of a DCD file:

```
catdcd -first 196 min_all.dcd
```

Analysis Features of VMD

- RMSD analysis, RMS best fits, mass-weighted RMSD, etc. can all be done easily in VMD.
- Phi/psi angles are available for a given atom selection:

```
set ca [atomselect top "name CA"]
set philist [$ca get phi]
```

- Bond, angle and dihedrals values can be quickly determined for an entire trajectory:

```
label add Bonds 0/10 0/20
set bondval [label graph Bonds 0]
```

Matrices and vectors in Tcl

- In VMD, vectors are Tcl lists and matrices are nested lists.
- Vecadd, vecsub, transvec, etc. are mostly implemented in C and are reasonably fast.

VMD/Tcl Performance Tips

- Use a few large vectors (like x, y, z), rather than separate vectors for each atom, for your analysis scripts.
- Use CatDCD to whittle down a DCD file to just what you need.
- Atom selections are fast and scale well, but use them sparingly.

VMD and Python

- Many scientific packages for Python exist which can be useful for MD analysis
  - Numeric Python
  - Scientific Python
  - MMTK
- You can use these packages from within VMD by switching to the built-in Python interpreter (`gopython`).

MatDCD

- NAMD DCD files can be loaded to/from Matlab for analysis.
- Easy to compute correlation functions, perform principal component analysis, SVD, etc.
- No atom selection language, though VMD could be used to generate the Matlab script…

Mindy

- "Minimal Molecular Dynamics" program based on NAMD source code.
- Provides a framework for "hacking" one's own analysis tools without having to understand all of NAMD.
- Has been used to find hydrogen bonds, compute interaction energies between subsets of atoms.