Molecular Dynamics Simulations

Amedeo Caflisch

Department of Biochemistry

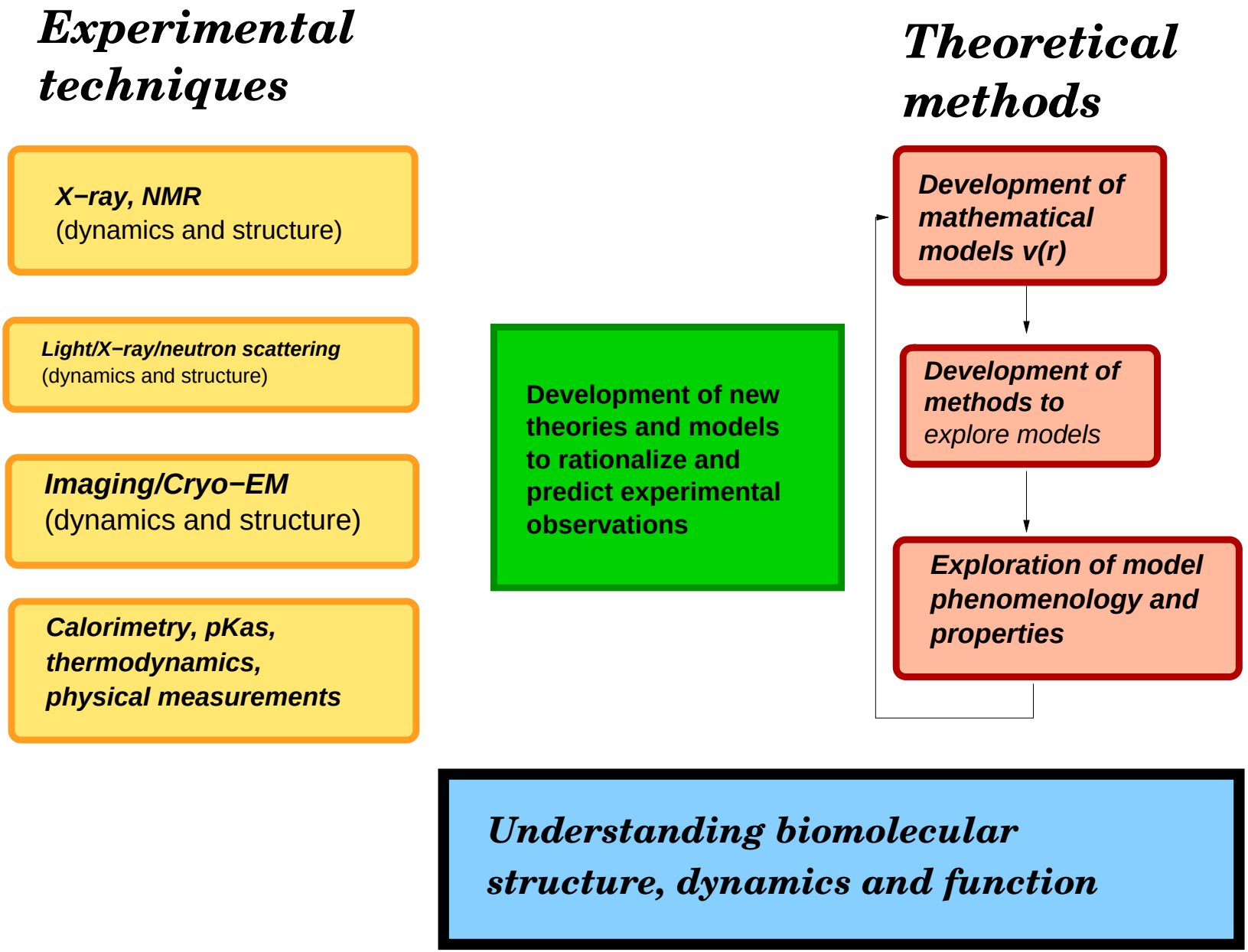
University of Zürich

caflisch@bioc.unizh.ch
http://www.biochem-caflisch.unizh.ch

April 2006

Contents

1 Introduction



2 Force field

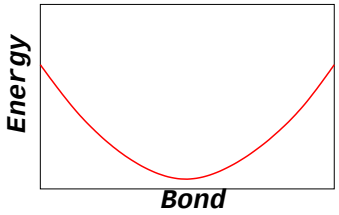
- A force field is an empirical approximation for expressing structure-energy relationships in biopolymers
- It is a compromise between speed and accuracy
- Common form (CHARMM):

$$
\begin{aligned}
E(\mathbf{r_1}, \mathbf{r_2}, \ldots, \mathbf{r_N}) &= \sum_{bonds} \frac{1}{2} k_i^b (d_i - d_i^0)^2 \\
&+ \sum_{angles} \frac{1}{2} k_i^\theta (\theta_i - \theta_i^0)^2 \\
&+ \sum_{torsions} k_i^\phi [1 + \cos(n_i \phi_i - \delta_i)] \\
&+ \frac{1}{2} \sum_{nonbond} \left(\epsilon_{ij}^{min} \left[\left(\frac{d_{ij}^{min}}{d_{ij}} \right)^{12} - 2 \left(\frac{d_{ij}^{min}}{d_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon d_{ij}} \right)
\end{aligned}
$$

2.1 Energy terms

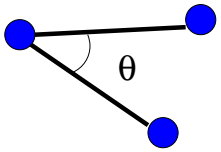
From specctroscopy, IR, NMR

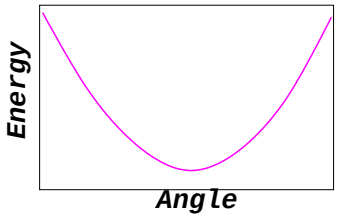
Bonds



$$v_{bond} = \frac{1}{2}k_i^b(d_i - d_i^0)^2$$

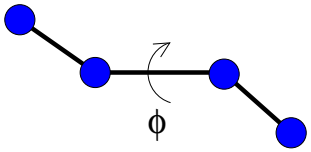
Angles

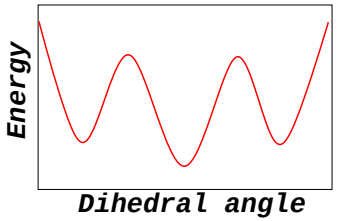




$$v_{angle} = \frac{1}{2}k_i^\theta(\theta_i - \theta_i^0)^2$$

Dihedrals

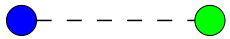


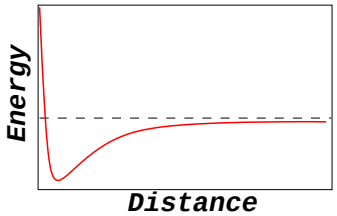


$$v_{dihedral} = k_i^\phi[1 + \cos(3\phi_i)] + k_i^{\phi'}[1 - \cos(\phi_i - \pi)]$$

From quantum chemistry, thermodynamics

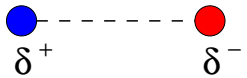
Van der Waals

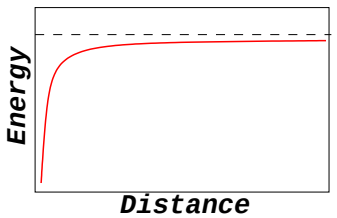




$$v_{vdW} = \epsilon_{ij}^{min}\left[\left(\frac{d_{ij}^{min}}{d_{ij}}\right)^{12} - 2\left(\frac{d_{ij}^{min}}{d_{ij}}\right)^{6}\right]$$

Electrostatics

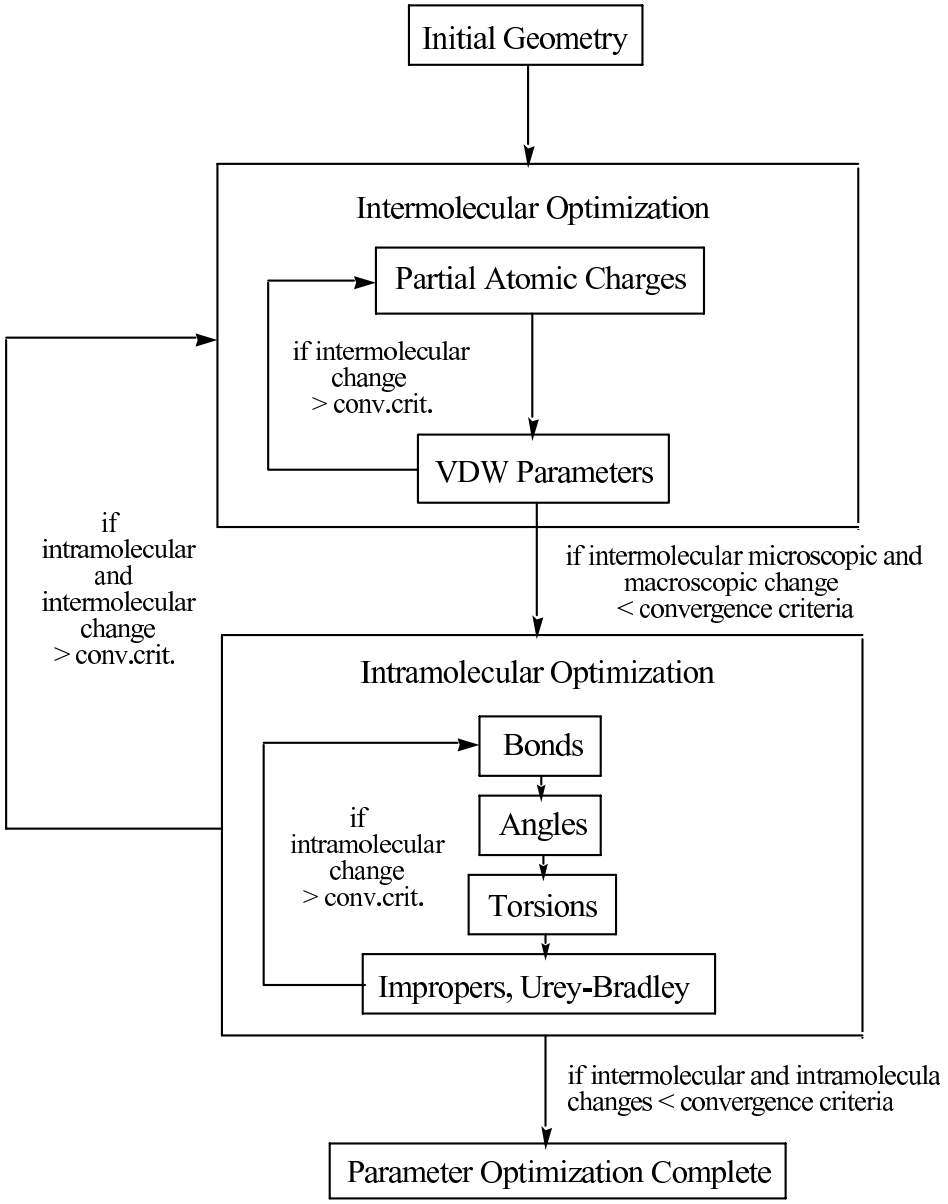




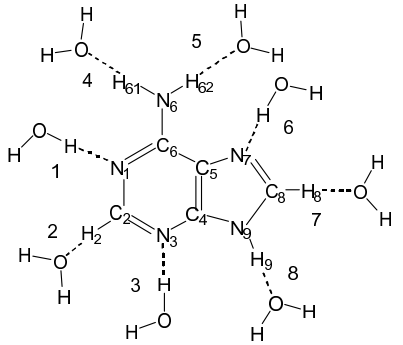
$$v_{Coulomb} = \frac{q_i q_j}{\epsilon d_{ij}}$$

2.2 Parametrization

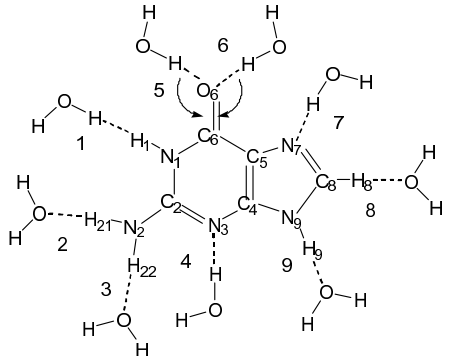
MacKerell, Parameter Notes

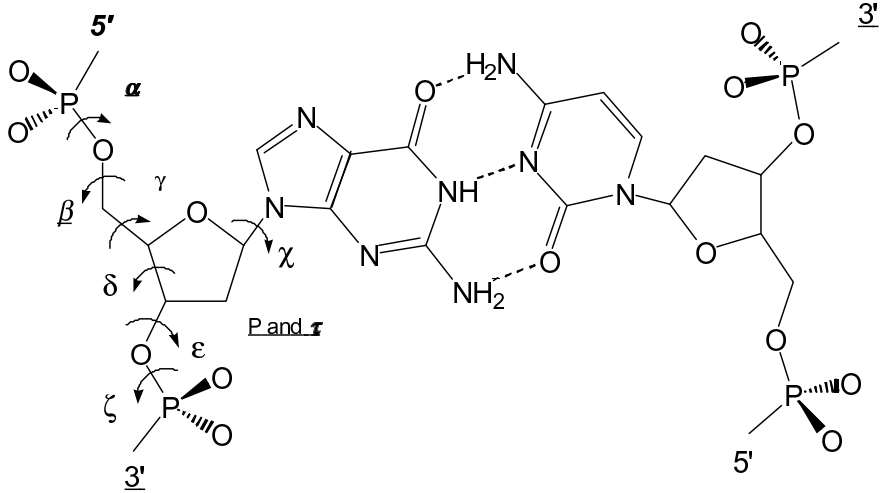


D

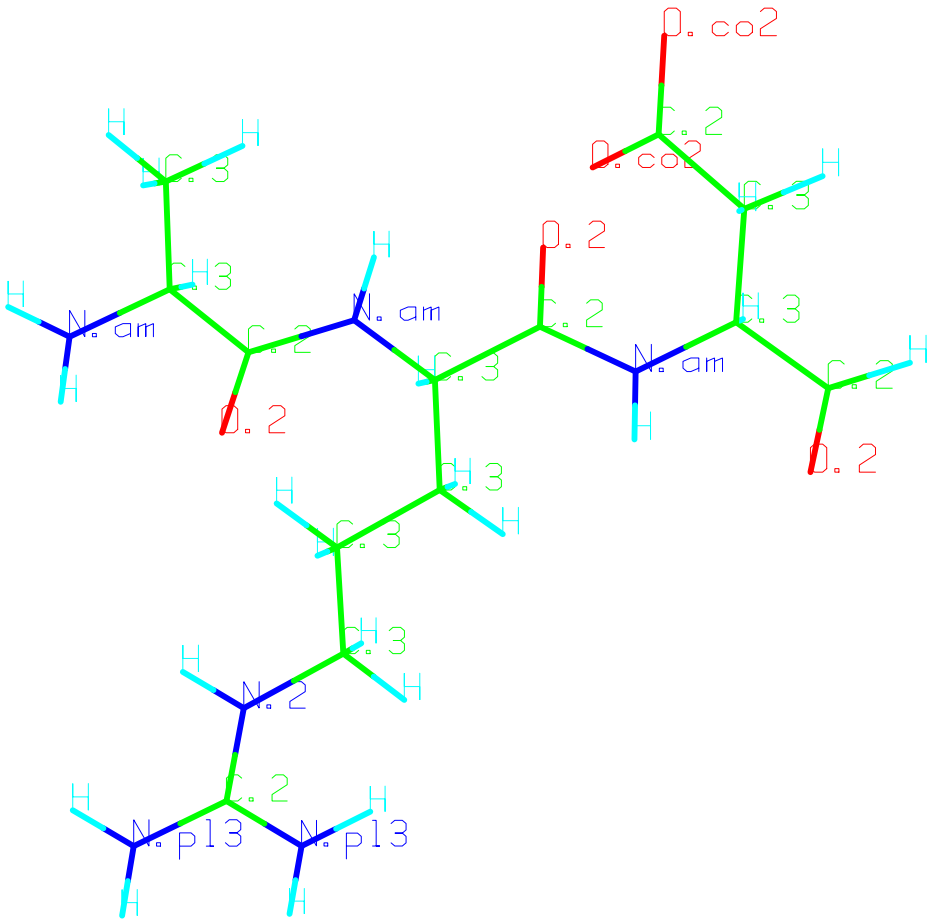


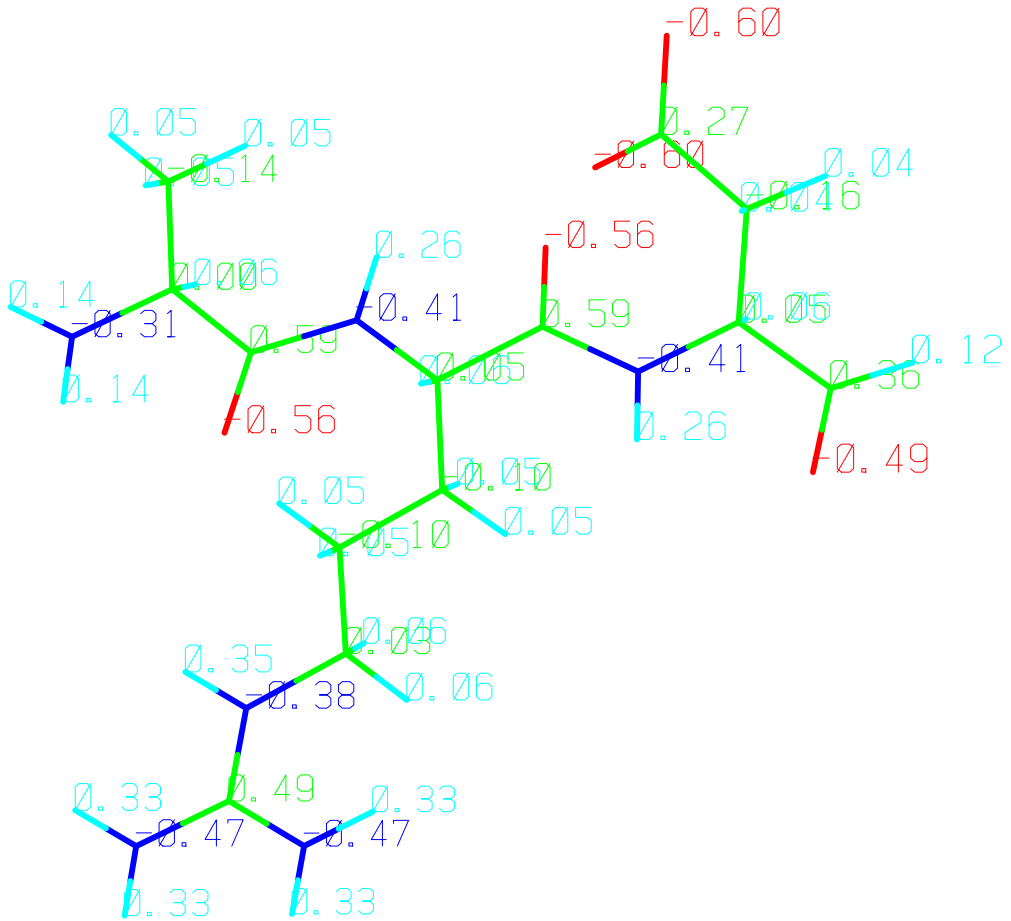
E





2.3 Atom types and partial charges

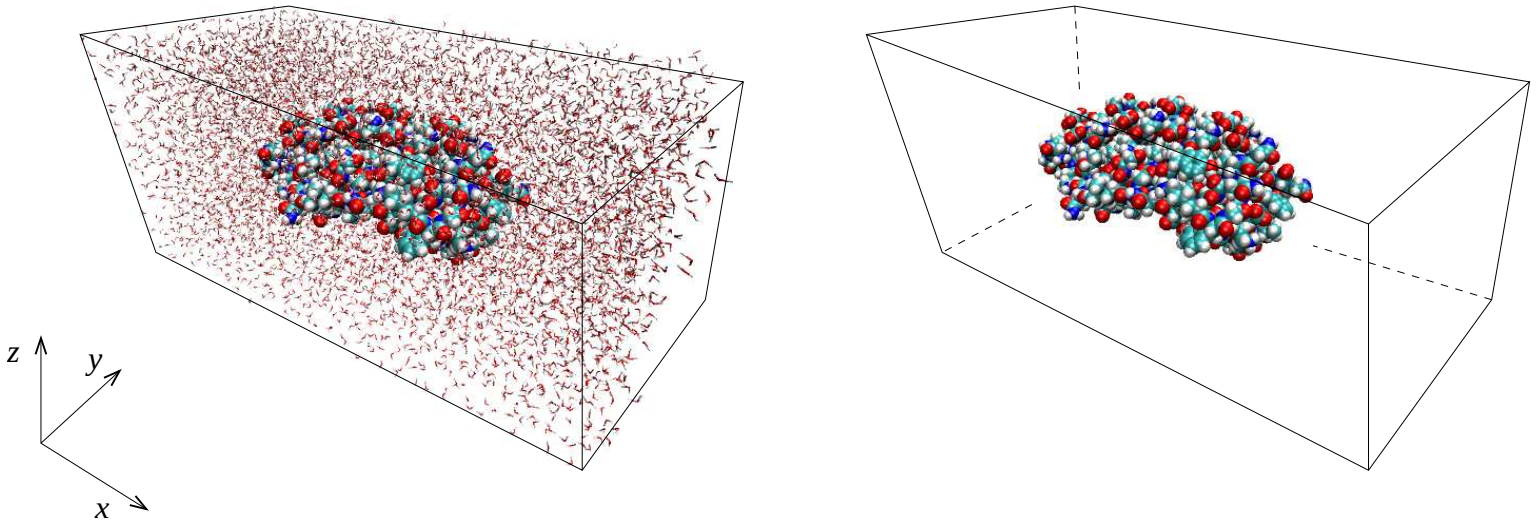




2.4 Models for simulation studies

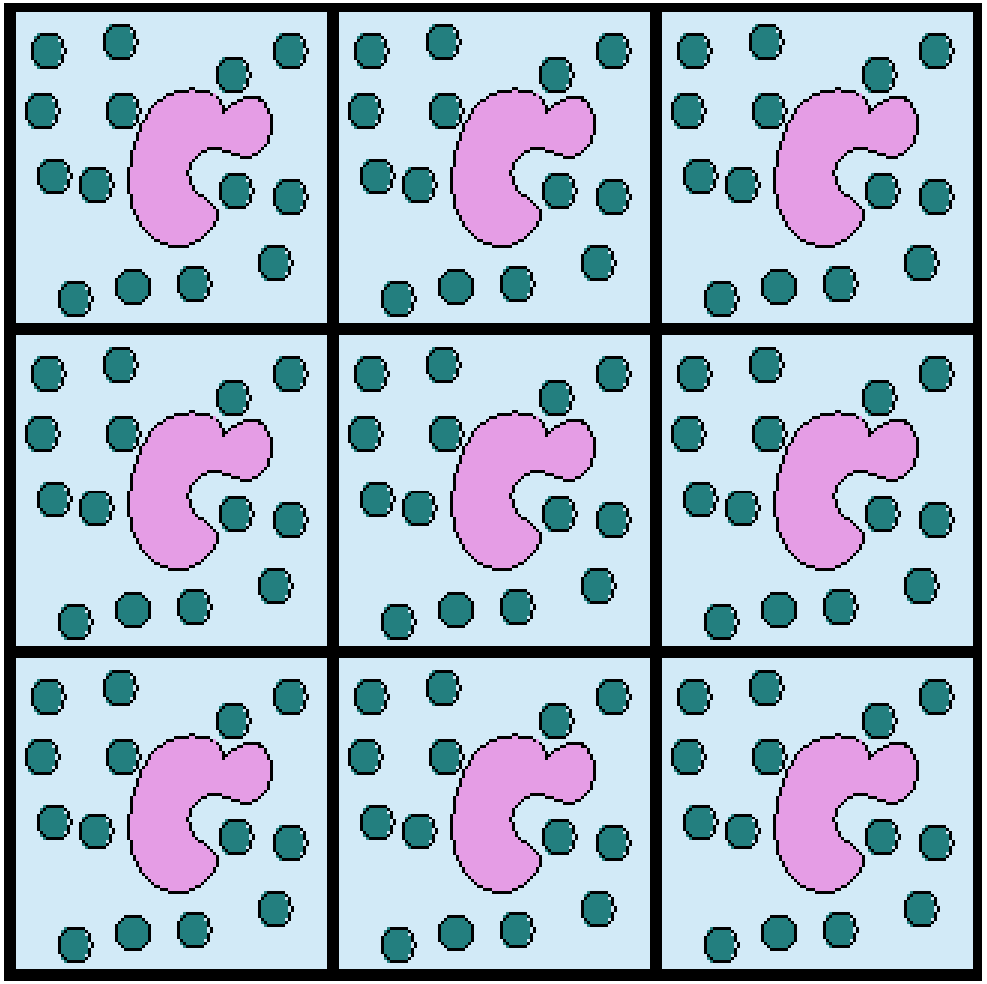
- How similar is the (macro)molecular system being studied to systems that were employed in the development of the model?
- The generality (i.e., range of possible applications) of a given model can only be established by comparison to experimental data for a wider and wider variety of (macro)molecular systems.
- Example: protein force field parameters are developed using small compounds which represent the peptide bond (e.g., N-methyl-acetamide) and the functional groups of the side chains of the natural amino acids (e.g., propane for Val, benzene for Phe, imidazole for His, indole for Trp, etc). Hence, it is straigthforward to determine the parameters for a norleucine side chain (–C_β-C_γ-C_δ-C_ϵ) but not for a phosphotyrosine because a protein force field does not usually have parameters for phospate.
- Some properties are more amenable to accurate computation than others. Example: Energy minimization to obtain refined structures is much simpler than calculating free energy differences.

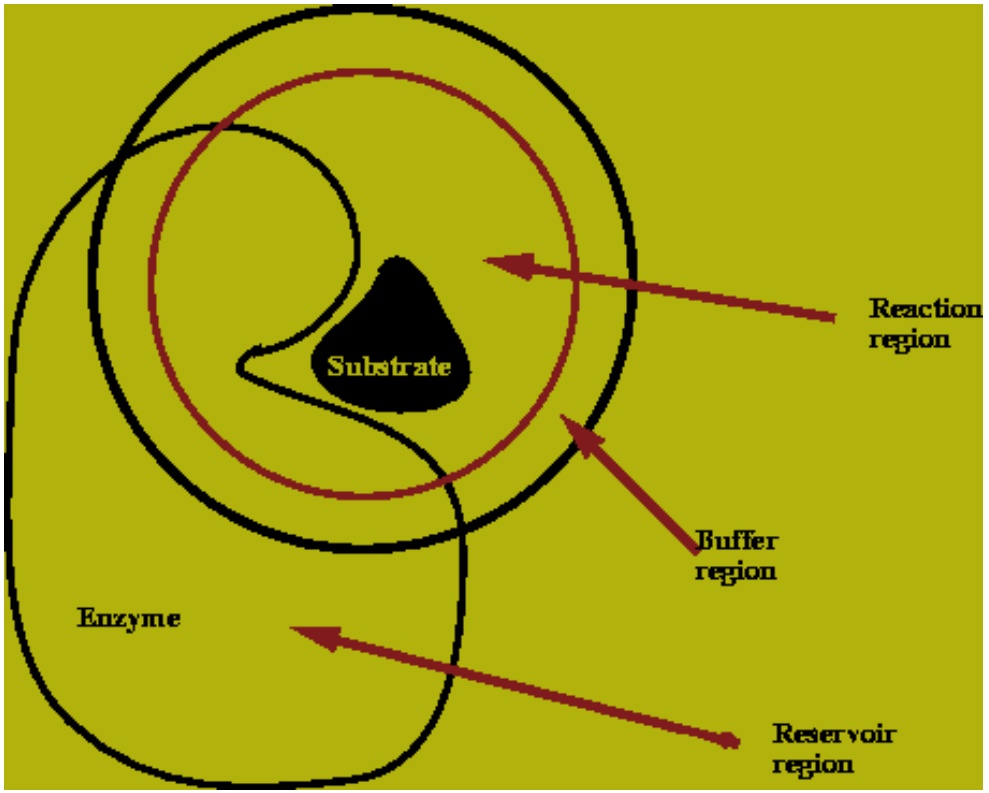
3 Solvation models



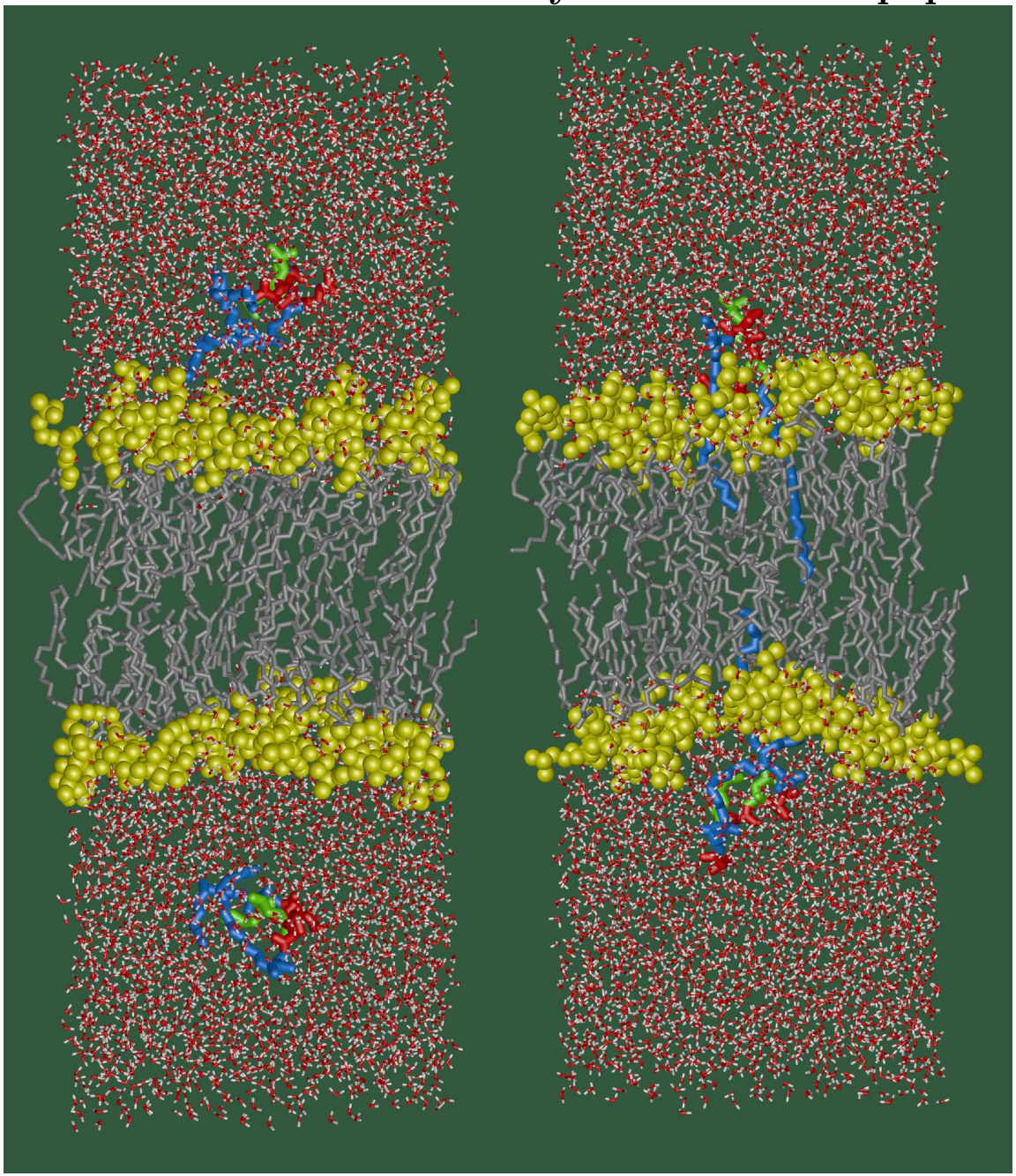
- Simulations *in vacuo* suffer from serious artifacts.
- Implicit solvation models reduce computational costs by removal of H_2O interaction centers and degrees of freedom.
- Solvation term based on the solvent accessible surface area (Ferrara et al., *Proteins* **46**, 24, 2002).

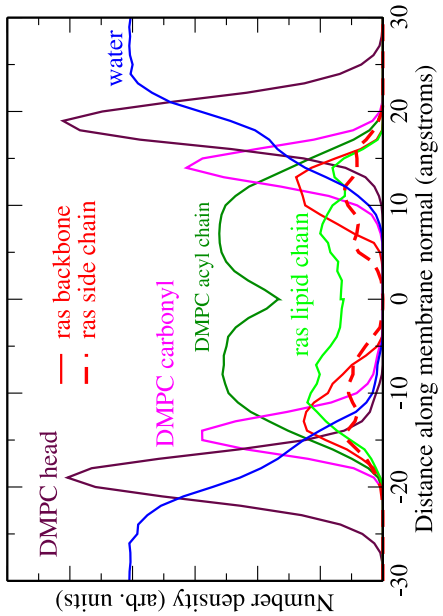
3.1 Periodic boundary conditions (PBC) vs. spherical boundary





3.2 Periodic boundary conditions for peptide/membrane system



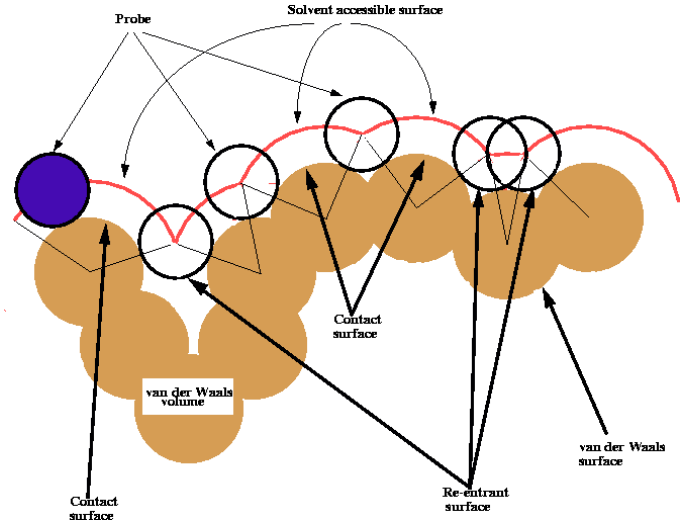


Gorfe et al., Membrane localization and flexibility of a lipidated ras peptide studied by molecular dynamics simulations. J. Amer. Chem. Soc. 126, 15277-15286, 2004.

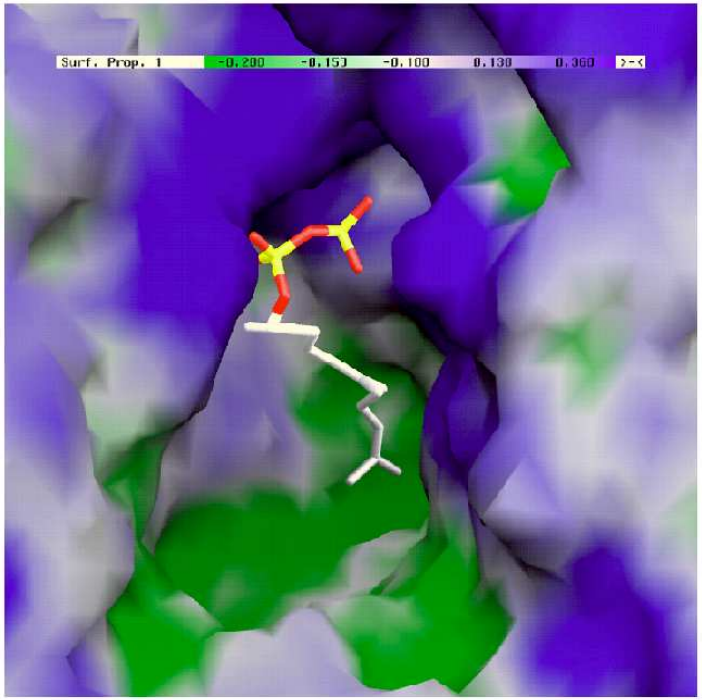
3.3 Solvent accessible surface area

$$W(\mathbf{r}) = \mathbf{E}(\mathbf{r}) + \mathbf{V_{solvation}}(\mathbf{r})$$

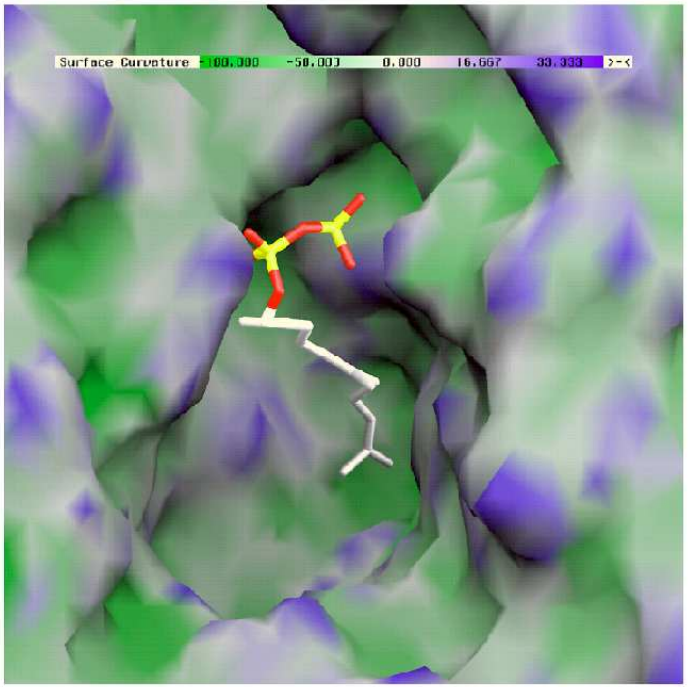
$$V_{solvation}(\mathbf{r}) = \sum_{\mathbf{i=1}}^{\mathbf{N_{polar}}} \sigma_{\mathbf{N,O}} \mathbf{A_i} + \sum_{\mathbf{i=1}}^{\mathbf{N_{apolar}}} \sigma_{\mathbf{C,S}} \mathbf{A_i}$$



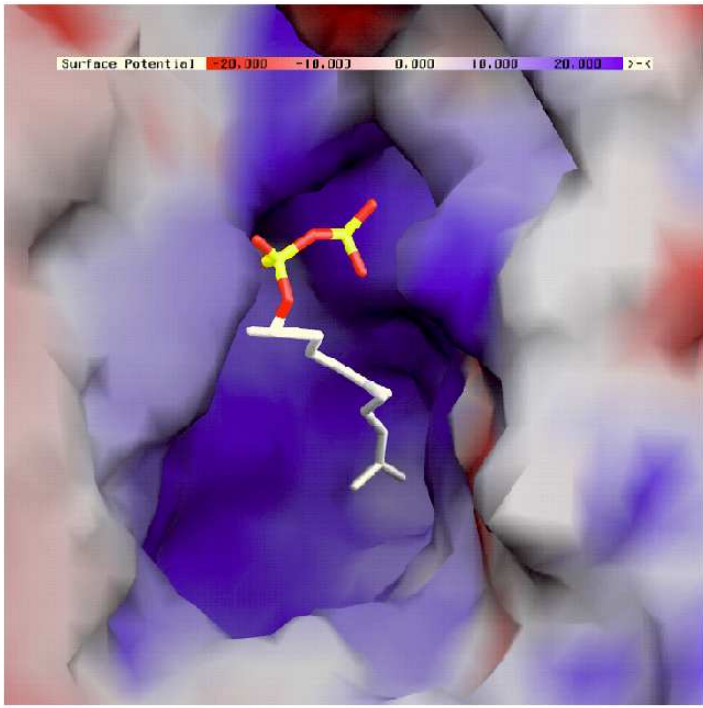
Farnesyltransferase-farnesylpyrophosphate complex



hydrophilic/hydrophobic



concave/convex

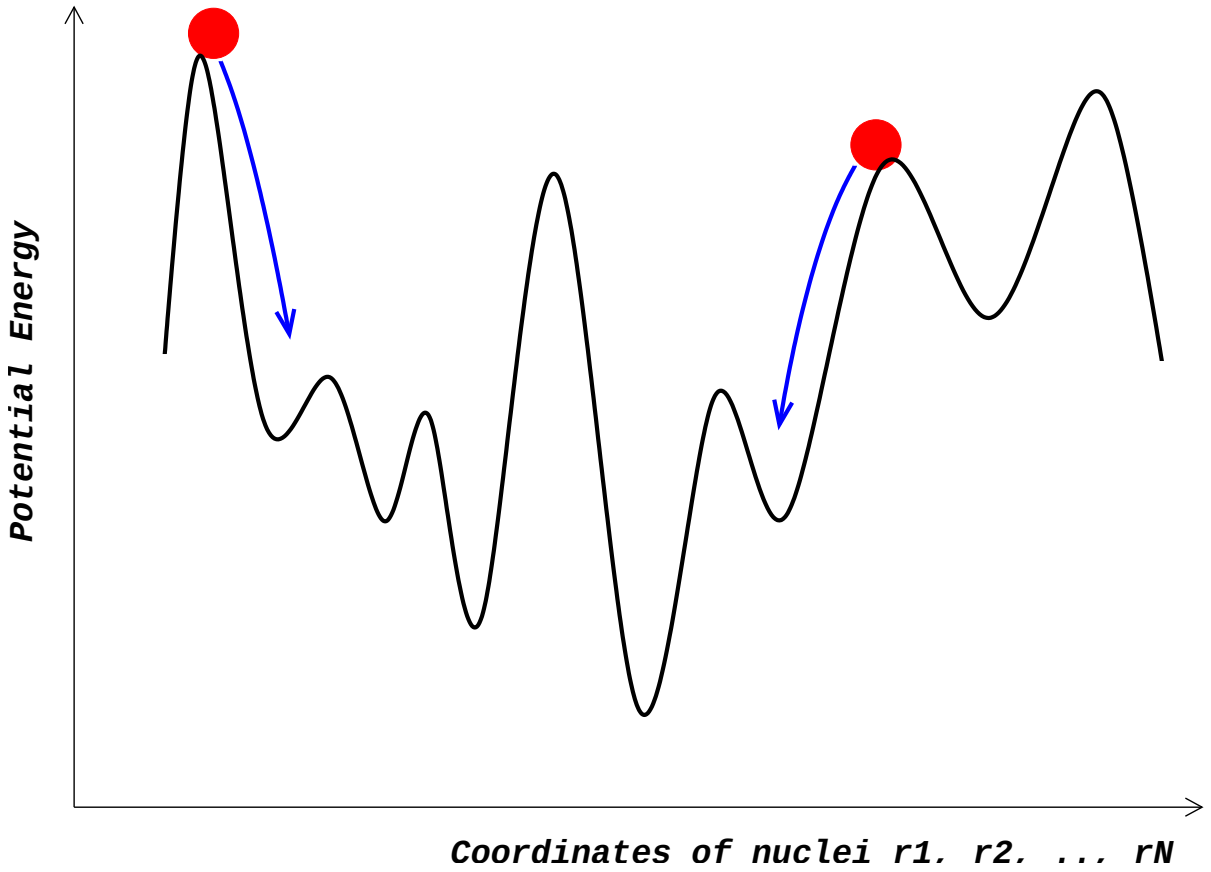


electrostatic potential

Scarsi et al., *Proteins* **37**, 565, 1999

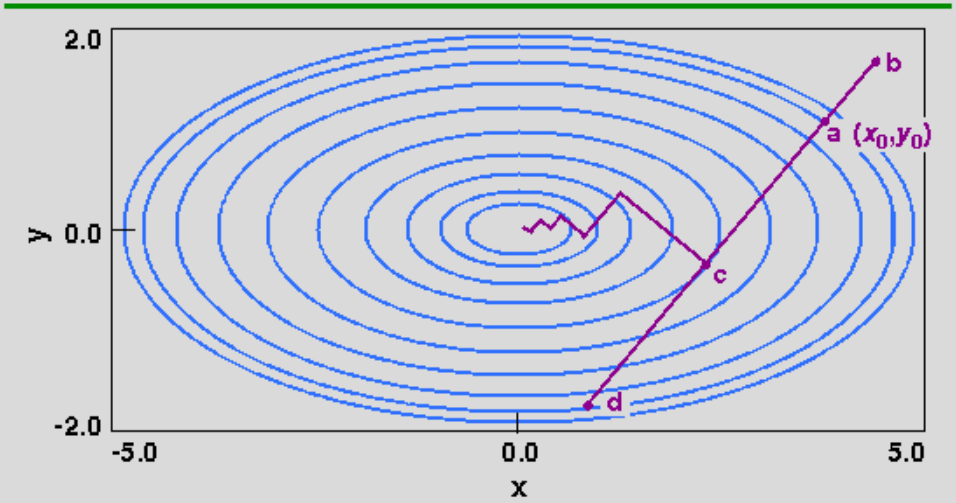
4 Sampling techniques

4.1 Energy minimization

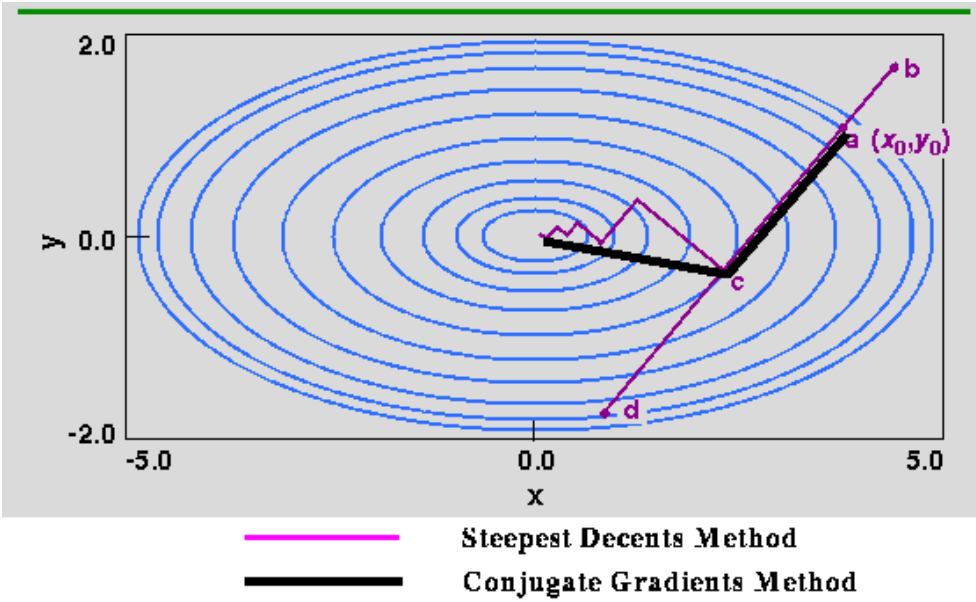


- Minimization follows gradient
- Reaches the nearest local minimum
- Steepest descent, conjugate gradient

4.2 Steepest descent vs. conjugate gradient

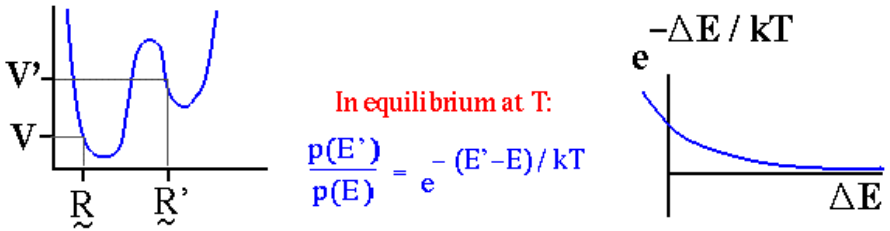


The derivative vector from the initial point **a** (x_0, y_0) defines the line search direction. Note that the derivative vector does not point directly toward the minimum. Compare this representation with that in Figure "right", where the line (**b**–**a**–**c**–**d**) is searched in one dimension for the minimum. Note that the minimum (point **c**) occurs precisely at the point where the derivative vector is tangent to the energy contours, which implies that the subsequent derivative vectors are orthogonal to the previous derivatives.



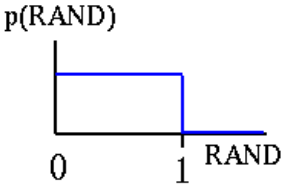
4.3 Metropolis Monte Carlo (Boltzmann statistics)

MC compares energies. No forces calculated.



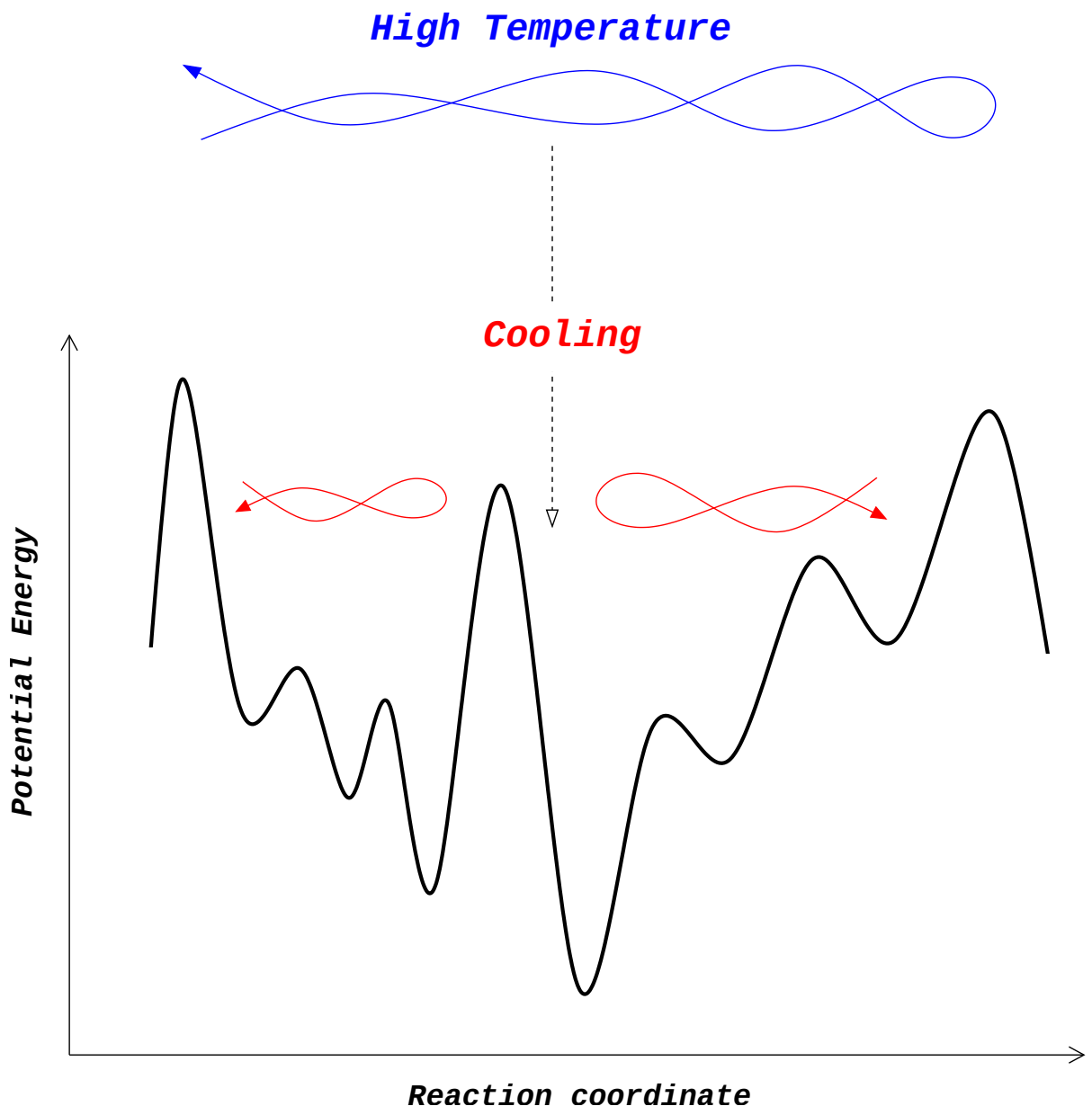
Monte Carlo Algorithm

Generate initial structure $\underset{\sim}{R}$. Calculate $V(\underset{\sim}{R})$.
Modify structure to $\underset{\sim}{R}$'. Calculate $V'=V(\underset{\sim}{R}')$.
If $V' < V$ then
$\underset{\sim}{R} \leftarrow \underset{\sim}{R}'$, $V \leftarrow V'$
else
Generate random number RAND:
If $e^{-(V'-V)/kT} >$ RAND
$\underset{\sim}{R} \leftarrow \underset{\sim}{R}'$, $V \leftarrow V'$
End if
End if
Repeat for N steps.



- Metropolis Monte Carlo yields an ensemble (Boltzmann statistics).
- Ergodicity: every accessible point in configuration space should be reached in a finite number of Monte Carlo steps from any other point.
- Kinetics are usually not meaningful.

4.4 Simulated annealing (good for sampling but no ensemble)



4.5 Parallel tempering (equilibrium Monte Carlo scheme)

- M *non-interacting* copies of the system at different T_m

- A *state* is defined by

$$X = \overbrace{\left(x^1_{m(1)}, \ldots, x^M_{m(M)}\right)}^{T_1,\ T_2,\ldots,\ T_M}, \qquad x^i_m \equiv (q^i, p^i)_m$$

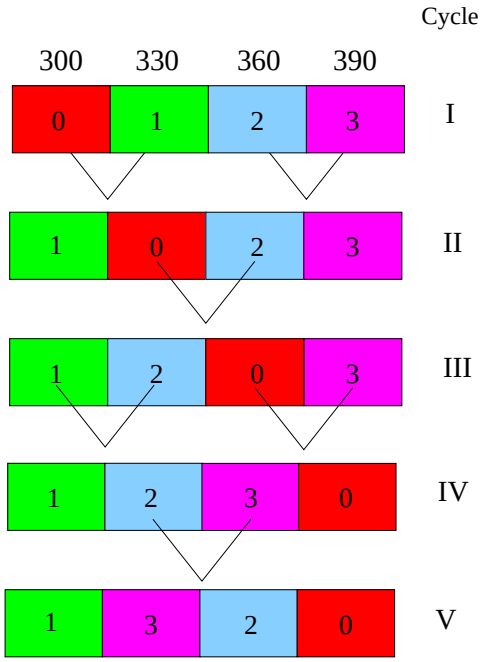
- In order to converge toward equilibrium the *detailed balance* should be satisfied. Therefore:

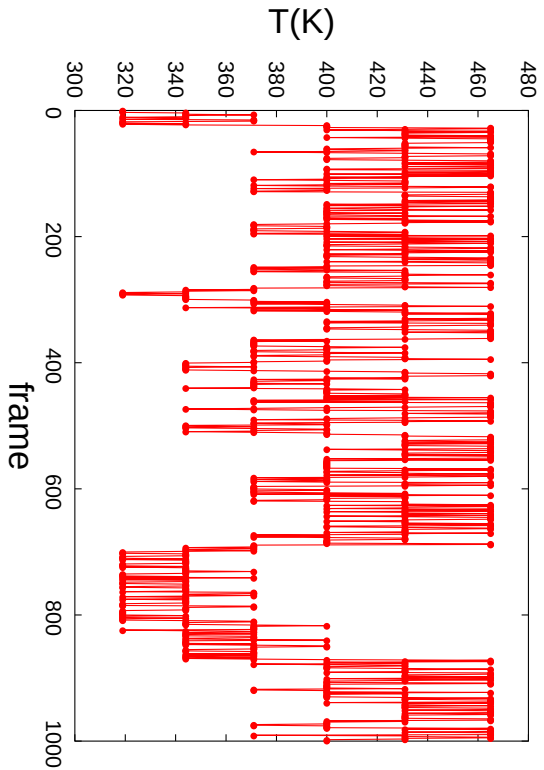
$$w(X \to X') = \begin{cases} 1, & \Delta \leq 0, \\ \exp(-\Delta), & \Delta > 0. \end{cases}$$

where $\Delta \equiv [\beta_n - \beta_m]\left(\mathcal{E}(x^i_m) - \mathcal{E}(x^j_n)\right)$.

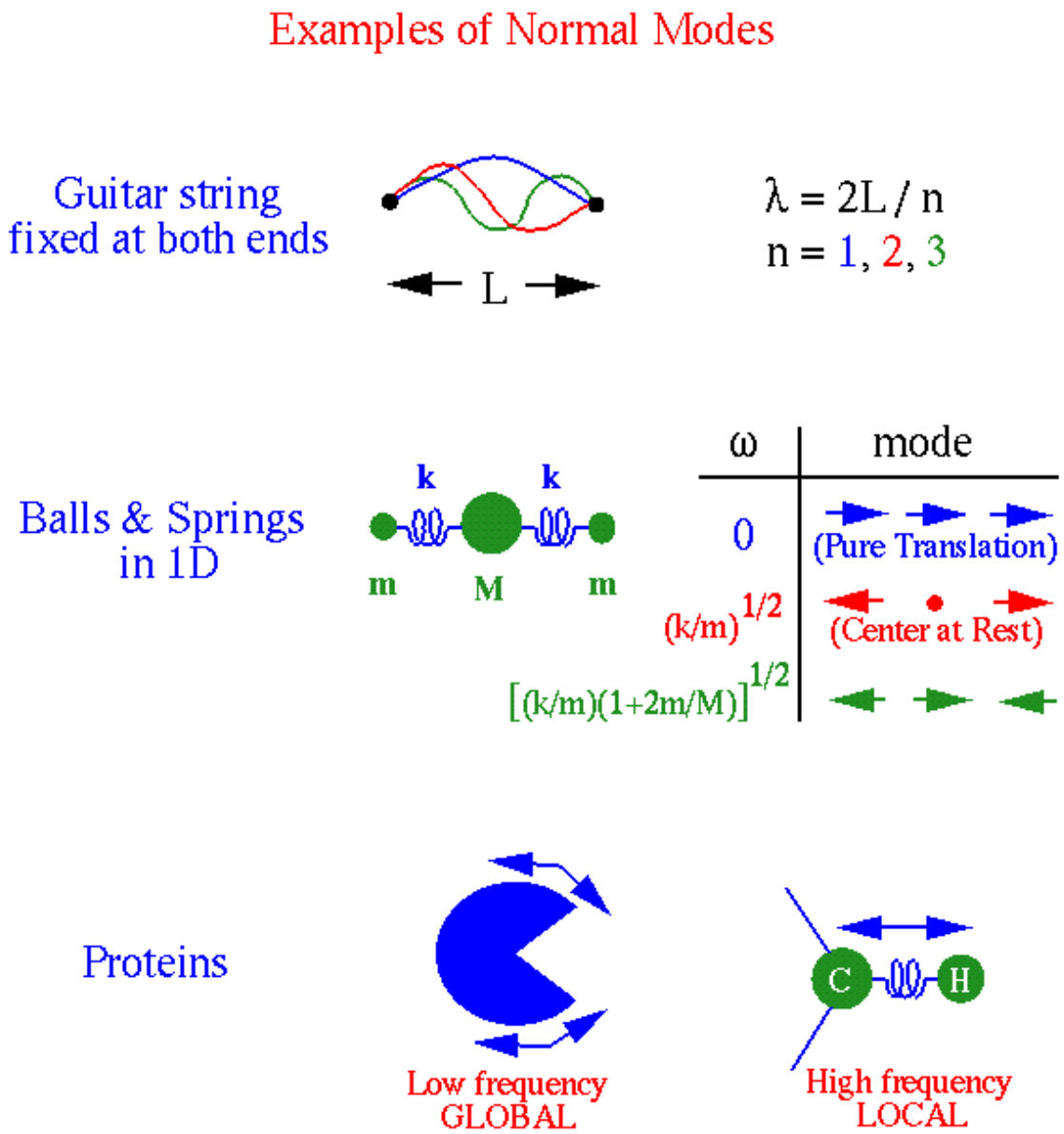
High T	replicas jump from basin to basin	(*inter-basin*)
Low T	replicas explore a single valley	(*intra-basin*)

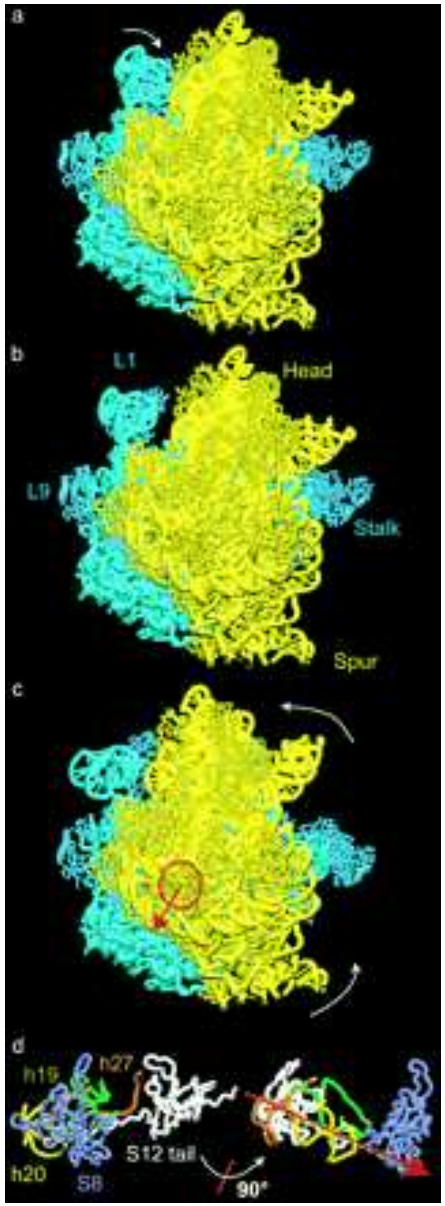
Rao and Caflisch, *J. Chem. Phys.* **119**, 4035, 2003





5 Normal mode analysis





Tama et al., PNAS 100, 9319, 2003

6 Molecular dynamics

6.1 Basics

- Atomic positions (coordinate file) ⟶
- Covalent structure (topology file) ⟶
- Potential energy function (parameter file) ⟶
- Additional atoms (solvent, counterions) ⟶
- Special features (PBC, constant T and/or P) ⟶
- Atomic velocities ⟶

- Effective temperature (through kinetic energy)
- Forces on each atom

6.2 Equations

- $\mathbf{F}_i = \mathrm{m}_i \mathbf{a}_i \qquad \mathbf{F}_i = -\mathbf{grad}_i E \qquad E = E_{bonding} + E_{non-bonding}$

1. solve for a_i at t $\qquad -\frac{dE}{dr_i} = F_i = m_i a_i(t)$
2. update v_i at $t + \Delta t/2$ $\qquad v_i(t + \Delta t/2) = v_i(t - \Delta t/2) + a_i(t)\Delta t$
3. update r_i at $t + \Delta t$ $\qquad r_i(t + \Delta t) = r_i(t) + v_i(t + \Delta t/2)\Delta t$
4. go to 1.

- Timestep controls accuracy of numerical solution.
- Fundamental timestep is determined by high frequency vibrations (covalent bonds $\longrightarrow \Delta t = 10^{-15}$ sec).
- Highest frequency motions, i.e., hydrogen atom vibrations, can be removed with holonomic constraints.

6.3 Thermodynamic variables T and P

- Statistical ensembles connect *microscopic* to *macroscopic*

  Microcanonical (NVE, entropy)

  Canonical (NVT, Helmhotz free-energy)
  · $T = \sum m \left\langle v^2 \right\rangle / (3k_b)$

  Isothermal-isobaric (NPT, Gibbs free-energy)
  · $P = kinetic + virial\ contributions$

- Thermostats, barostats allow to choose the appropriate ensemble.
  · Andersen, Nosé, Hoover.

7 Applications of molecular dynamics

1. Sampling configuration space,
   e.g., simulated annealing to determine or refine structures.

2. Obtaining a description of the system at equilibrium,
   i.e., sampling with appropriate Boltzmann factor:

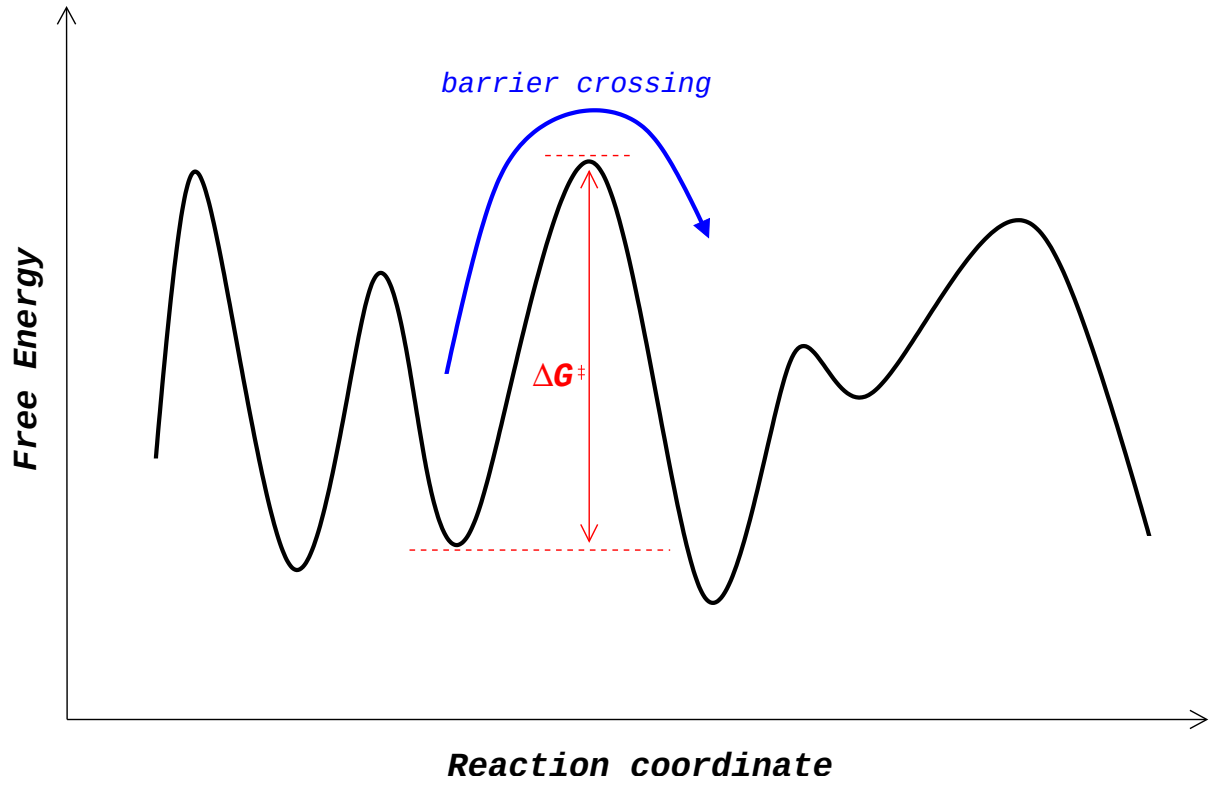
   - structural and motional properties
   - values of thermodynamic parameters

3. Obtaining actual dynamics and kinetics,
   i.e., sampling with appropriate Boltzmann factor and
   correct representation of the development of the system over time.

For (1) and (2) Monte Carlo simulations can be also used.

From Karplus and McCammon, Nature Structural Biology **9**, 646, 2002.

8 Free-energy barriers and timescales



- To cross a free-energy barrier $\boxed{\tau = \tau_0 \exp(\Delta G^{\ddagger}/k_B T)}$ with $\tau_0 \sim 10^{-12}$ s:

  $1\ kcal/mol : \sim ps$, $5\ kcal/mol : \sim ns$, $10\ kcal/mol :\ \mu s$ or longer

- Sampling should exceed timescales of interest by $\sim$ 10-fold.
- System size and complexity increase required timescales (equilibration of ions, complex landscapes, multiple minima)

9 Approximations in molecular dynamics

- Approximations inherent to the force field (E) $\longrightarrow$
  Systematic error:
  Calculations of free energy differences is still very difficult.

- Time scale and sampling problem $\longrightarrow$
  Statistical error:
  Conformational transitions that require $> 0.1 - 1\mu$s cannot be simulated (yet) by conventional molecular dynamics techniques.

- Other simulation approaches:
  - MD with implicit solvent (approximate)
  - Brownian dynamics
  - Monte Carlo (move definitions are difficult for macromolecules)

10 References

- Y. Gao, W. Yang and M. Karplus, *A structure-based model for the synthesis and hydrolysis of ATP by F_1-ATPase.* Cell 123, 195-205, 2005.
- M. Karplus and J. A. McCammon, *Molecular dynamics simulations of biomolecules.* Nature Structural Biology 9, 646-652, 2002.
- V. Daggett and A. Fersht, *The present view of the mechanism of protein folding.* Nature Reviews Molec. Cell. Biology 4, 497-502, 2003.
- J. Gsponer, U. Haberthür and A. Caflisch, *The role of side chain interactions in the early steps of aggregation: Molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup35.* Proc. Natl. Acad. Sci. USA 100, 5154-5159, 2003.
- J. Gsponer and A. Caflisch, *Molecular dynamics simulations of folding from the transition state.* Proc. Natl. Acad. Sci. USA 99, 6719-6724, 2002.
- P. Ferrara and A. Caflisch *Folding simulations of a three-stranded antiparallel β-sheet peptide.* Proc. Natl. Acad. Sci. USA 97, 10780-10785, 2000.
- M. J. Field, *A practical introduction to the simulation of molecular systems.* Cambridge University Press, 1999.
- *Workshop on Methods and Applications of Molecular Dynamics to Biopolymers (2000)* `http://www.psc.edu/general/software/packages/charmm/tutorial/index.html`
- P. J. Steinbach, *Introduction to macromolecular simulation.* `http://cmm.cit.nih.gov/intro_simulation`
- D. Frenkel and B. Smit *Understanding molecular simulations.* Academic Press, 2002.
- M. P. Allen and D. J. Tildesley, *Computer simulation of liquids.* Oxford University Press.