Bioinformatics

Modeling of biological systems

Adrian E. Roitberg

University of Florida

ECAR2012 -
Escuela de Computacion de Alto Rendimiento

Despues de un largo viaje de un mes y medio….

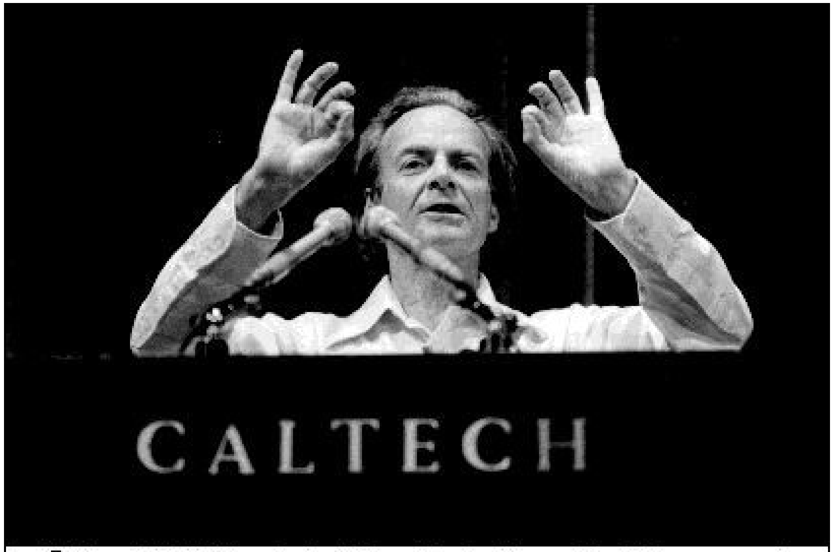
- Basic Principles

- Modern Simulations of Biomolecules

- Advanced Techniques

"everything that living things do can be understood in terms of the jiggling and wiggling of atoms."

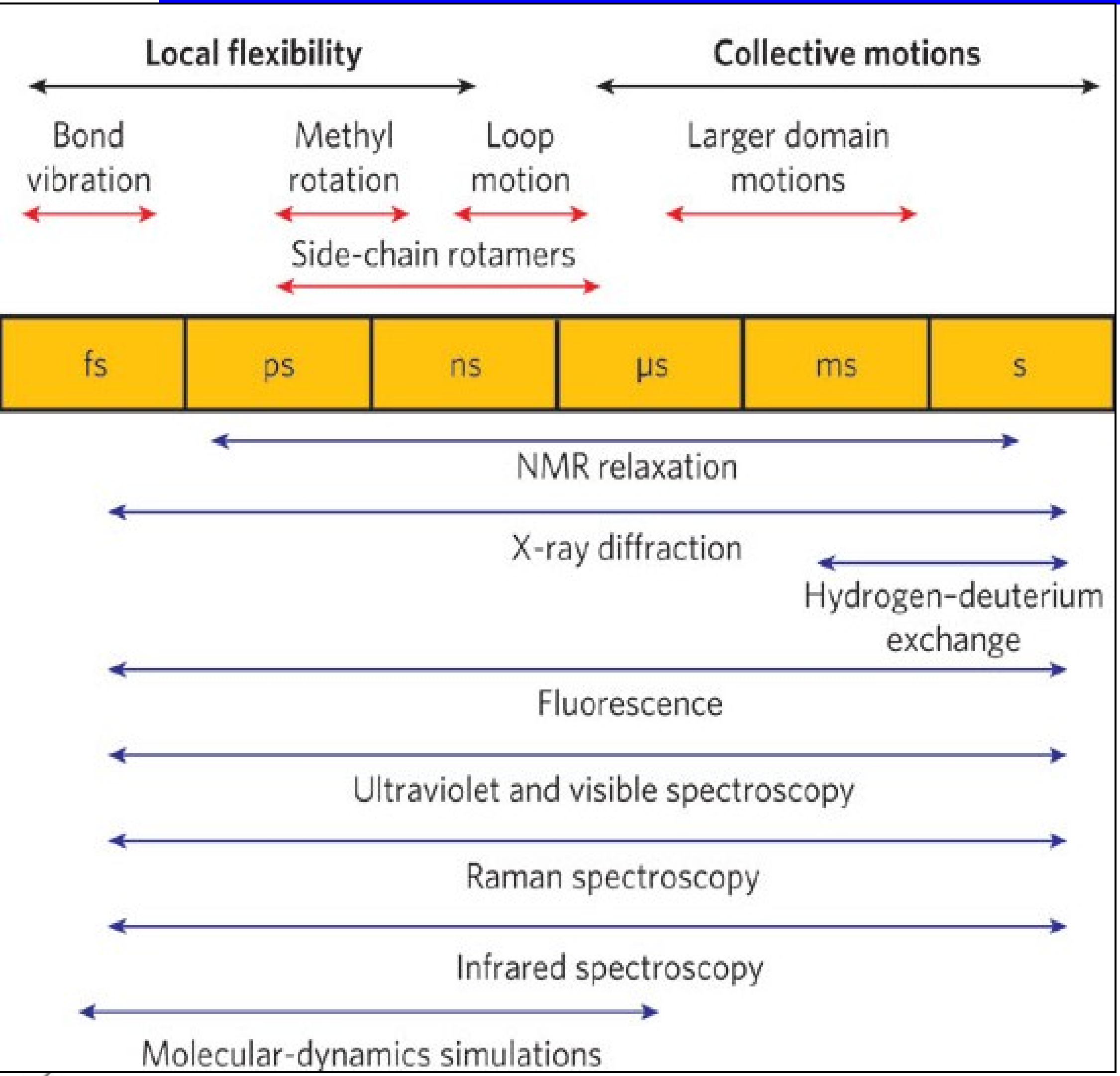
Richard Feynman, 1963



Biology involves motion

- Biomolecules undergo conformational changes
- These can be difficult to probe with experiments
- Function arises from dynamics on a wide range of time scales

Biomolecular timescales



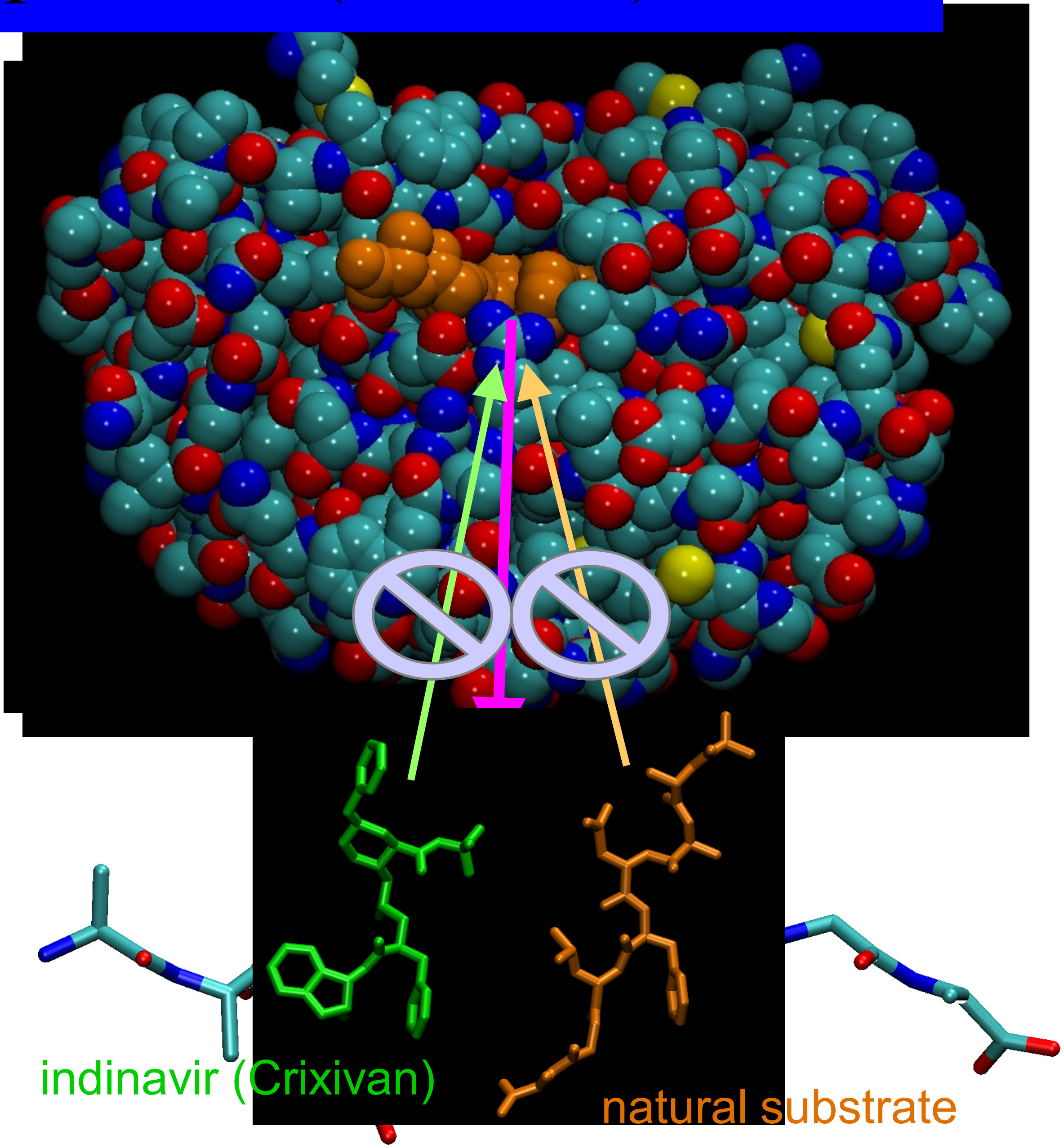
- Dynamics occur on timescales that vary over many orders of magnitude

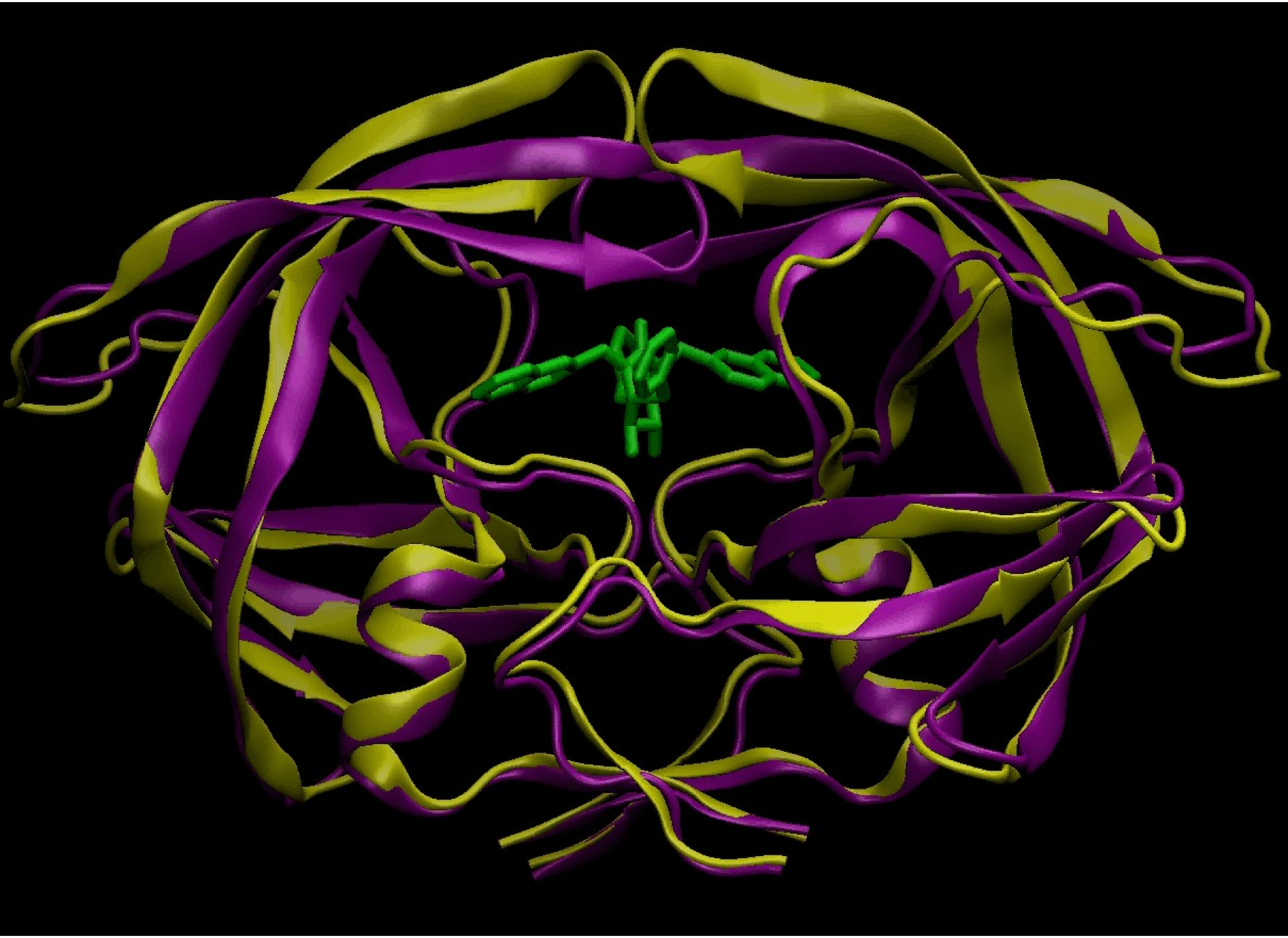
- Most experimental data represents an average over time – know the limitations of your data!

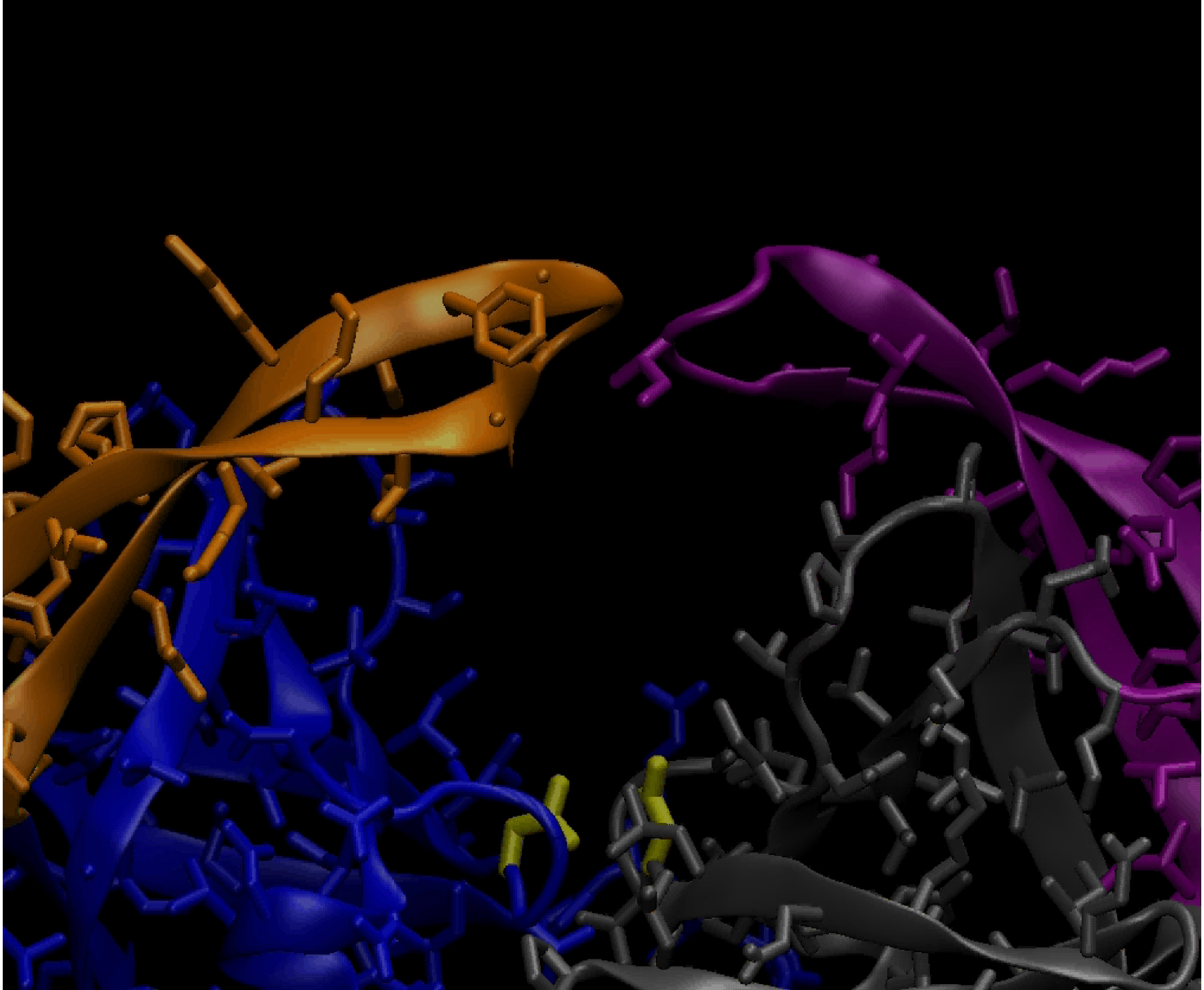
Katherine Henzler-Wildman & Dorothee Kern. *Dynamic personalities of proteins*. Nature, December 2007

HIV-1 protease (HIV-PR)

- Homodimer, 198aa
- 2 gly-rich flaps
- Cleaves gag, gag-pol polyproteins during maturation
- AIDS drug target (saquinavir, indinavir, etc.)
- Drug resistance can arise from remote mutations







Why run simulations?

- Models for time- or ensemble-averaged experimental observations
- Predictions: positive or negative
- Physical insight- why does it act this way?
- Connect structure and energy
- Nonphysical or impractical states: arbitrary chemistry, conditions, substrate-enzyme complex, etc.
- Test experimental assumptions

Why MD?

- Atoms move!
  - We may be interested in studying time dependent phenomena, such as molecular vibrations, phonons, diffusion, etc.
  - We may be interested in studying temperature dependant phenomena, such as free energies, anharmonic effects,
  - etc.
- Ergodic Hypothesis
  - Time average over trajectory is equivalent to an ensemble average.
  - Allows the use of MD for statistical mechanics studies.

Alternatives

- Monte Carlo
  - Can do thermal averages.
  - Hard to do time dependant things.

- Hybrid MD/MC
  - Bad MD as good MC.
  - Generate configurations using poor/cheap/fast MD but then evaluate contribution to ensemble average using MC.

Energy

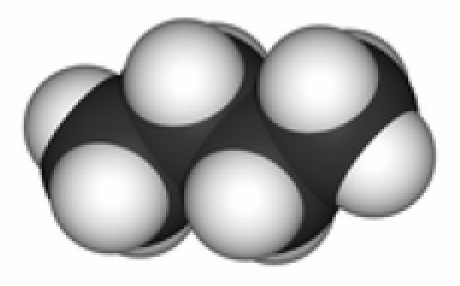
Energy: any arrangement of atoms and molecules in the system has a particular energy and the energy varies as the positions of the atoms (electrons) and molecules change.

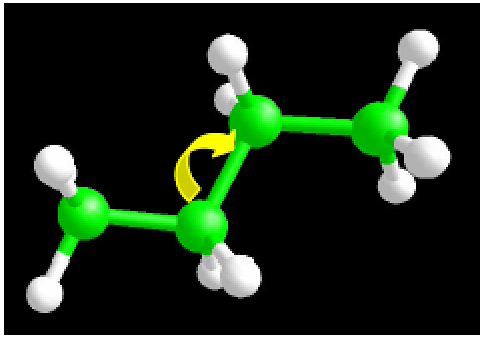
$$E \sim f\,(\text{atomic positions})$$

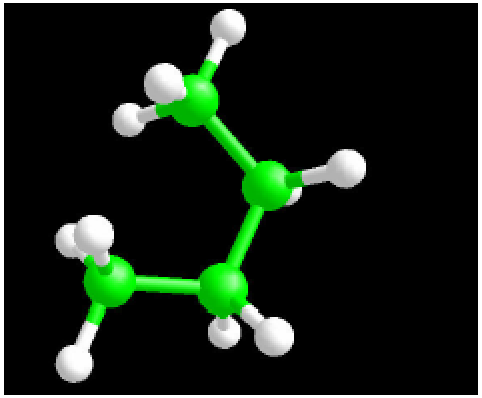
The most stable conformation of a molecule is the one with the lowest energy, but it is not the only one we are interested in.

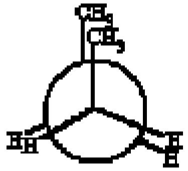
Examples: n-butane

$CH_3CH_2CH_2CH_3$



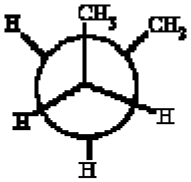






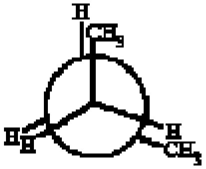
Syn (Eclipsed)
Dihedral Angle = 0, 360 degrees

0



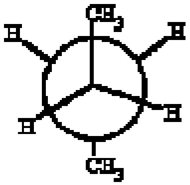
Gauche (Staggered)
Dihedral Angle = 60 degrees

-4.1



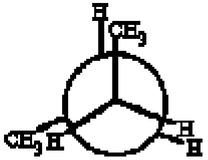
Eclipsed
Dihedral Angle = 120 degrees

-1.5



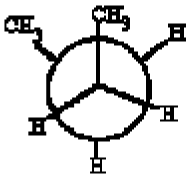
Anti (Staggered)
Dihedral Angle = 180 degrees

-5



Eclipsed
Dihedral Angle = 240 degrees

-1.5



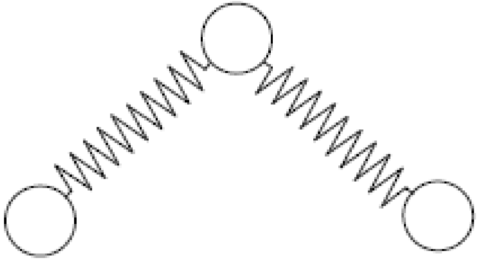
Gauche (Staggered)
Dihedral Angle = 300 degrees

-4.1

(ΔE in kcal/mol)

How do we Calculate the Energy?

- Quantum mechanics
  - Electrons are the smallest particle represented.
  - Solves the Schrödinger equation.
- Molecular mechanics
  - Atoms are the smallest particles.
  - Atoms are represented by balls.
  - Bonds are represented by strings.
  - “balls and springs” model



Force Fields

- Equations and parameters that relate the chemical structure and conformation to energy.

  E ~ **f** (atomic positions)
- FF used in molecular modeling are primarily designed to reproduce structural properties.
- A ff is usually designed to reproduce a given type of data (parameterized accordingly).
- Force fields are empirical, there is not a ’correct’ energy function or parameters.

Types of Force Field

- Compromise between accuracy and computational efficiency.
- Transferability is necessary to ensure predictability.
    - Class I FF: simple potential energy function which limits transferability.
    - Class II FF: extended potential energy function, including cross terms. Increases transferability.

⚠ **Increasing transferability can limit accuracy.**

Common Force Fields

Class I

- CHARMM
- CHARMm (Accelrys)
- AMBER
- GROMOS
- OPLS
- …

Class II

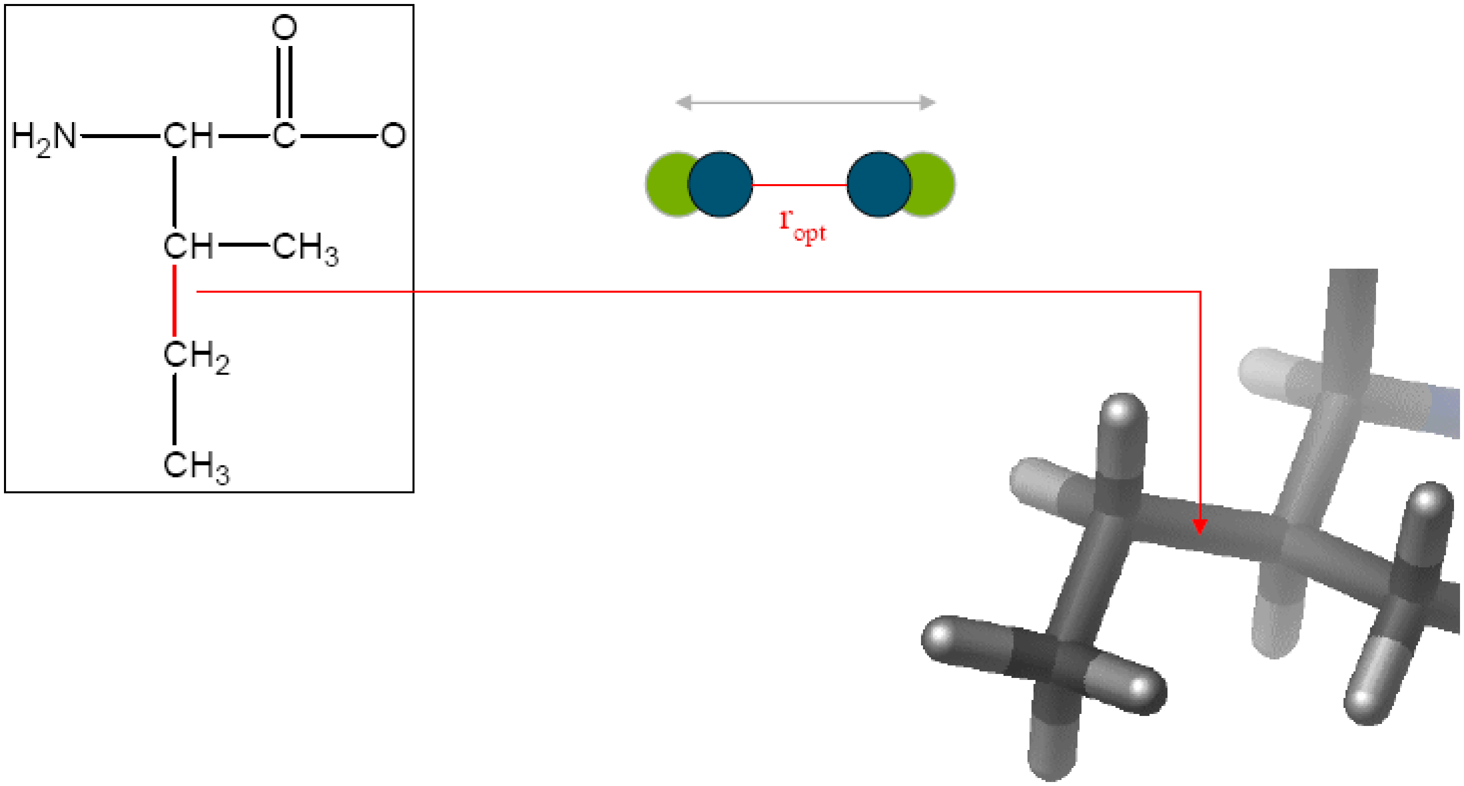
- CFF95 (Accelrys)
- MM3
- MMFF (Charmm, Macromodel, Moe,…)
- UFF
- Dreiding
- …

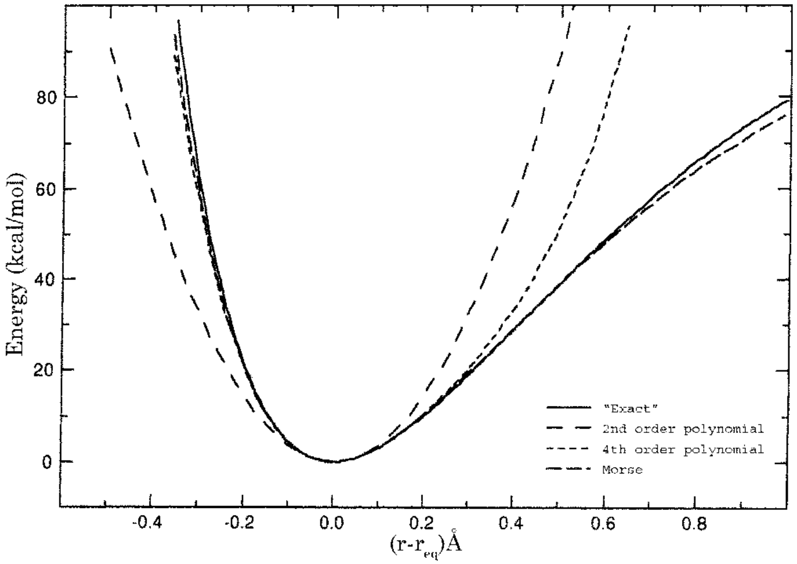
NOTE: There are often multiple versions of each of these force fields.

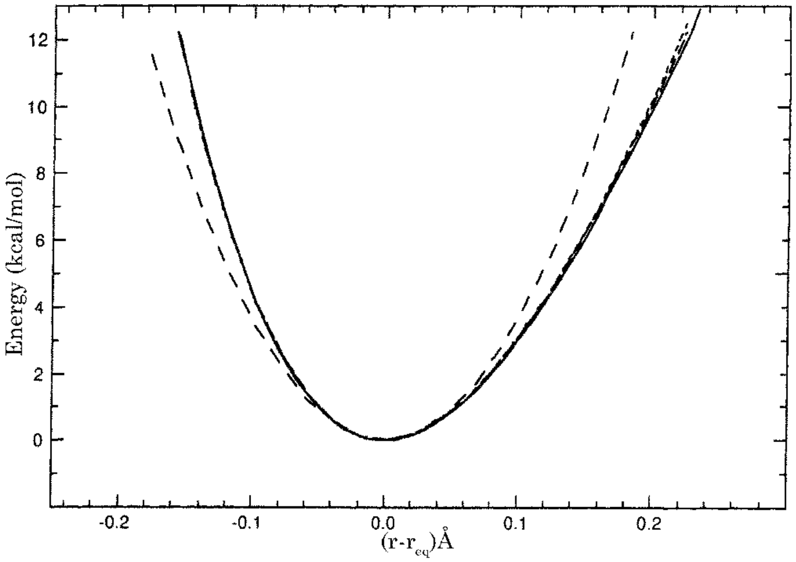
A Function for Energy?

$$E = \overbrace{E_{stretch} + E_{bend} + E_{torsion} + E_{impropers}}^{E_{\text{bonded}}} + \underbrace{E_{electrostatic} + E_{vanderWaals}}_{E_{\text{non-bonded}}}$$

Bonds







Morse potential

$$E_{stretching} = D_e.\{1-\exp[-a(r-r_o)]\}^2$$

Harmonic potential (Hooke)

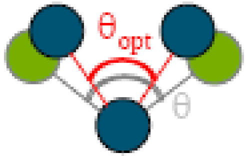
$$E_{stretching} = \sum_{1,2\,pairs} K_r (r-r_0)^2$$

ca. 100 to 500 kCal/mol/A²

- Morse potential requires 3 parameters so computationally expensive and difficult to parameterize.
- Harmonic potential is usually good enough since most bonds remain within +/- 0.1 Angstroms of “optimum”.

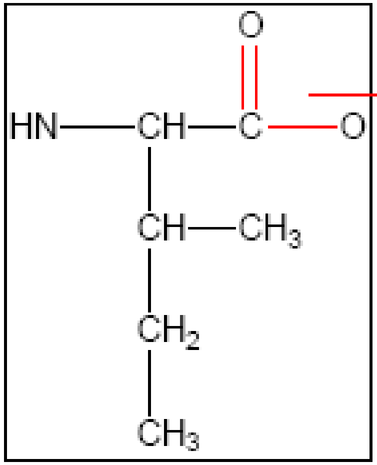
Angles

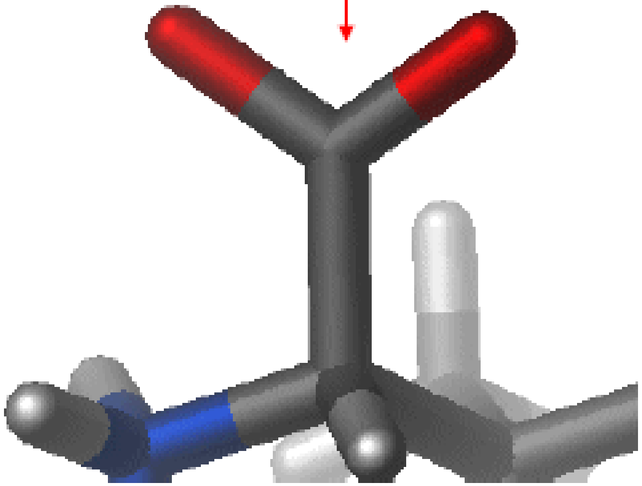
$$E_{bending} = \sum_{1,2\,pairs} K_\theta (\theta - \theta_0)^2$$



less energy is required to distort an angle than to stretch or compress a bond;

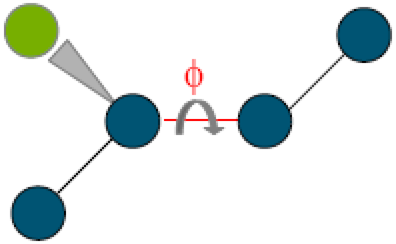
lower K_θ values: e.g. 10^{-3} kcal.mol^{-1}.deg^{-2}

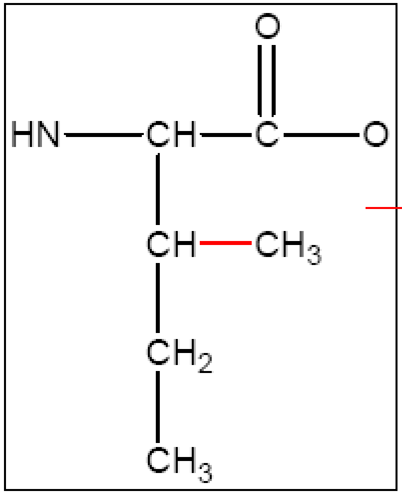


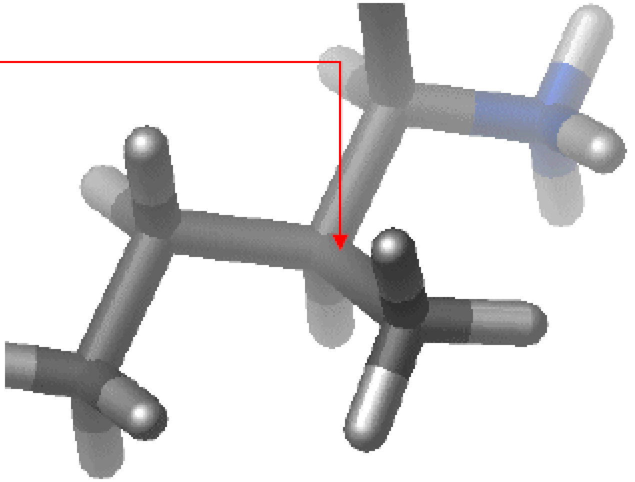


Dihedrals

$$E_{torsion} = \sum_{1,4pairs} K_{\phi} (1 - \cos(n\phi - \delta))$$







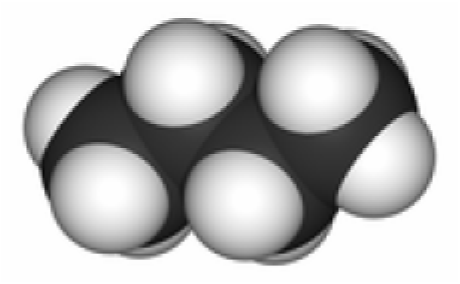
Dihedrals Contd.

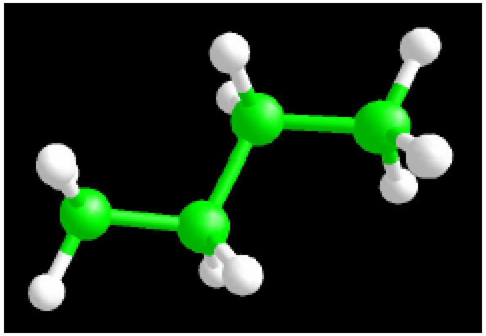
- Most of the variation in structure and energies is due to the interplay between torsional and non-bonded contributions.
- Almost always expressed as a cosine series expansion.

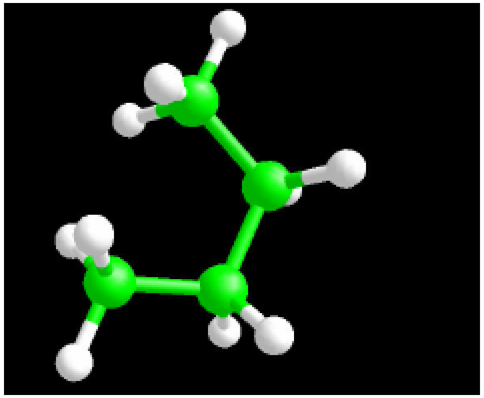
$$E_{torsion} = \sum_{1,4\,pairs} K_{\phi}\,(1 - \cos(n\phi - \delta\,))$$

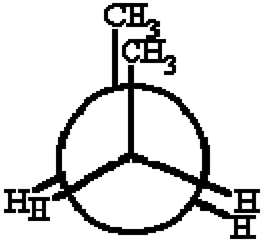
- K is often referred to as the barrier height (misleading since other terms contribute to the barrier: non-bonded, etc…)
- n: multiplicity, number of minimum points.
- d: phase factor, determines when the torsion potential has its minimum value.

Example: Butane



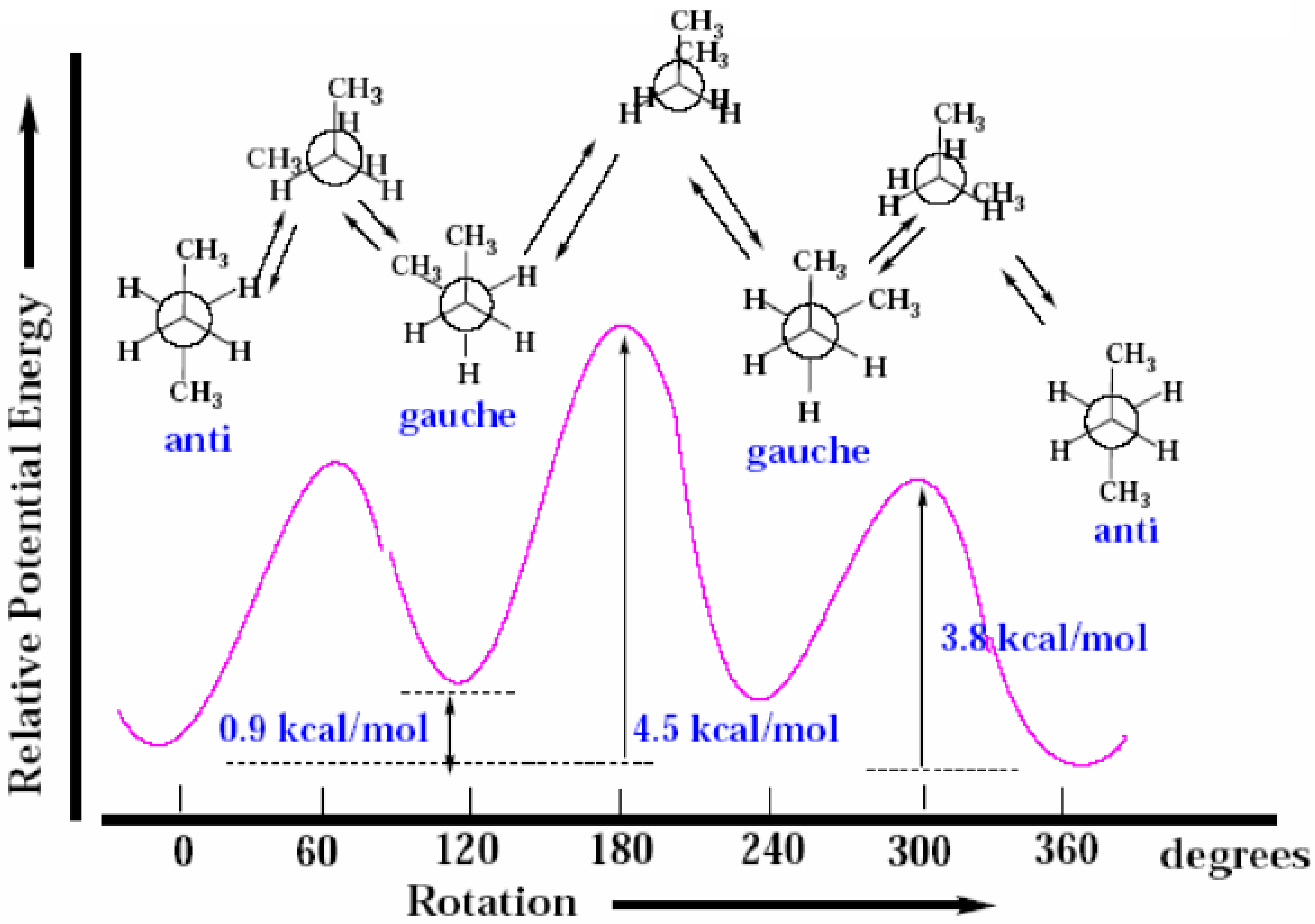




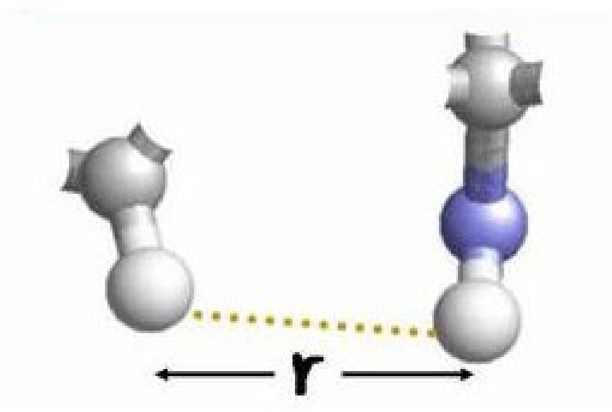


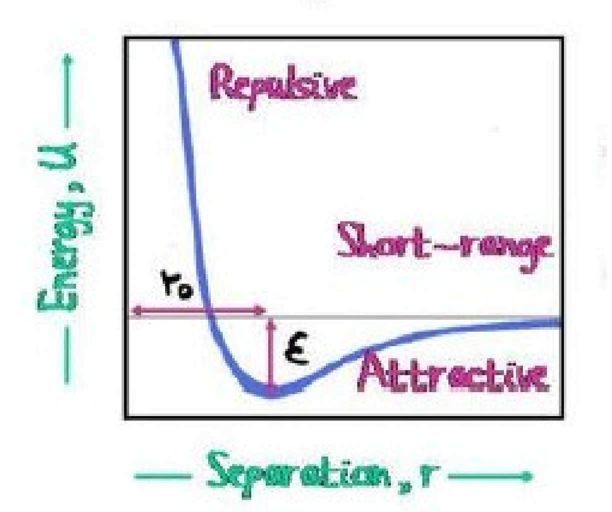
Syn (Eclipsed)
Dihedral Angle = 0, 360 degrees

Example: Butane



Non-Bonded Interactions: VDW





Lennard-Jones potential

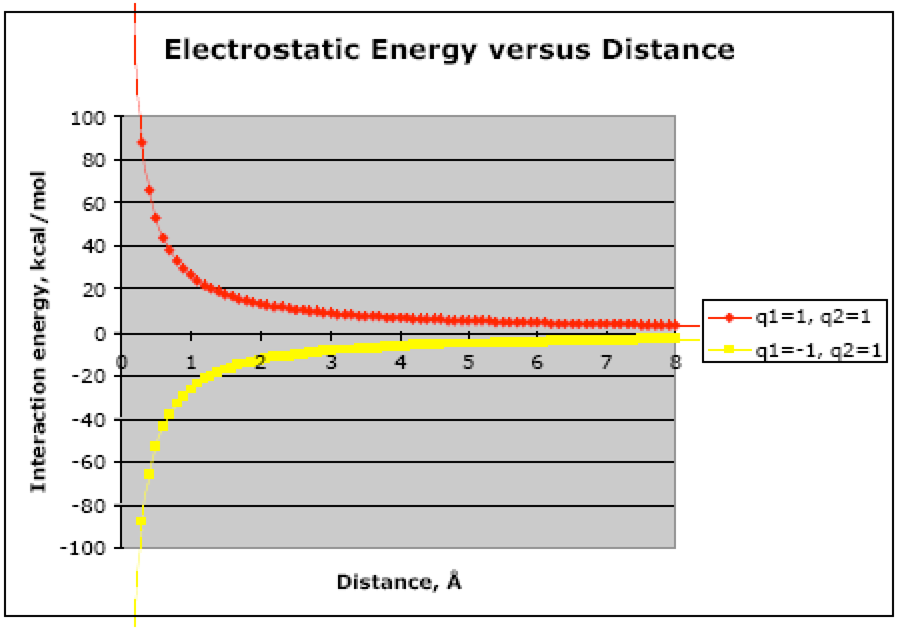
$$E_{vdW} = \sum_{nonbondedpairs} \varepsilon . \left[\left(\frac{r_0}{r} \right)^{12} - 2 . \left(\frac{r_0}{r} \right)^{6} \right]$$

$$E_{vdW} = \sum_{nonbondedpairs} \left[\left(\frac{A}{r^{12}} \right) - 2 . \left(\frac{C}{r^{6}} \right) \right]$$

Electrostatics

Coulomb potential

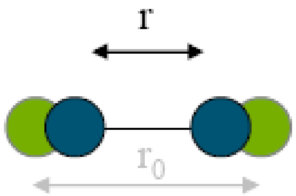
$$E_{vdW} = \sum_{nonbondedpairs} \frac{q_i . q_j}{\varepsilon . r}$$

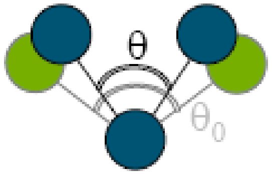


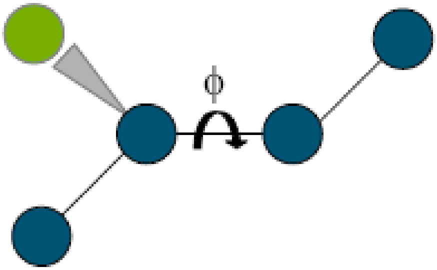
- point charge model

- no dipole-induced effects

Summary







$$E_{stretching} = \sum_{1,2\,pairs} K_r (r - r_0)^2$$

$$E_{bending} = \sum_{1,2\,pairs} K_\theta (\theta - \theta_0)^2$$

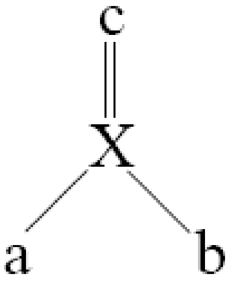
$$E_{torsion} = \sum_{1,4\,pairs} K_\phi (1 - \cos(n\phi - \delta))$$

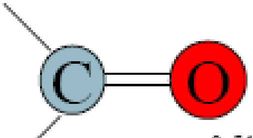
$$\boxed{E = E_{stretch} + E_{bend} + E_{torsion} + E_{impropers} + E_{electrostatic} + E_{vanderWaals}}$$

$$E_{impropers} = \sum_{impr.} K_\varphi (\varphi - \varphi_0)^2$$

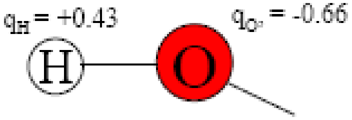
$$E_{electrostatic} = \sum_{\substack{nonbonded\\ i\text{-}k\ pairs}} q_i \cdot q_k / D.r_{ik}$$

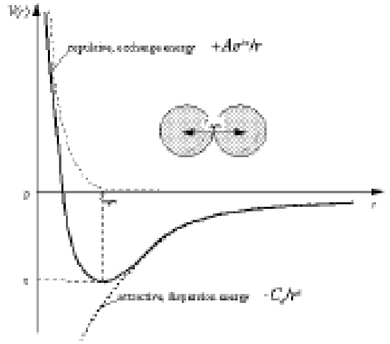
$$E_{van-der-Waals} = \sum_{\substack{nonbonded\\ pairs}} \left(\frac{\boxed{A_{ik}}}{r_{ik}^{12}} - \frac{\boxed{C_{ik}}}{r_{ik}^{6}} \right)$$





$q_O = -0.51$





The AMBER Force Field Equation

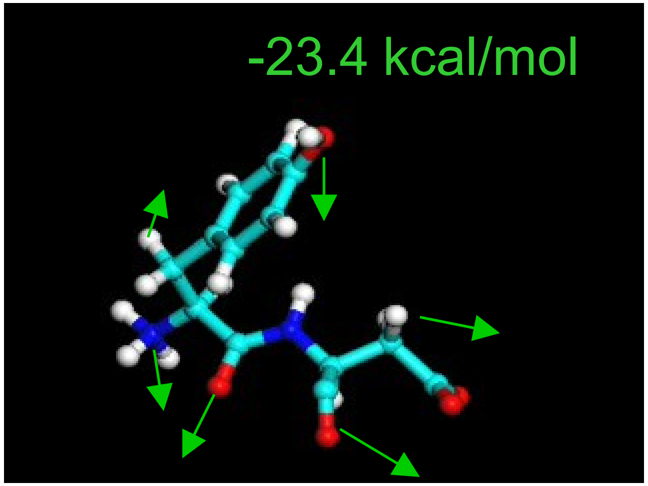
$$V\left(r^n\right)=\sum_{bonds} K_r\left(r-r_{eq}\right)^2+\sum_{angles} K_\theta\left(\theta-\theta_{eq}\right)^2$$

$$+\sum_{dihedrals}\frac{V_n}{2}\left[1+\cos\left(n\phi-\gamma\right)\right]+\sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}}-\frac{B_{ij}}{R_{ij}^{6}}+\frac{q_i q_j}{\varepsilon_r R_{ij}}\right]$$

Note: 1-2 and 1-3 non-bond interactions are parameterized into the bond and angle terms. Dihedral term also includes some of the non-bond interaction.

1-4 EEL scaled by 1.2
1-4 VDW scaled by 2.0

From structure/parameters to Energy/Forces



Molecular Dynamics

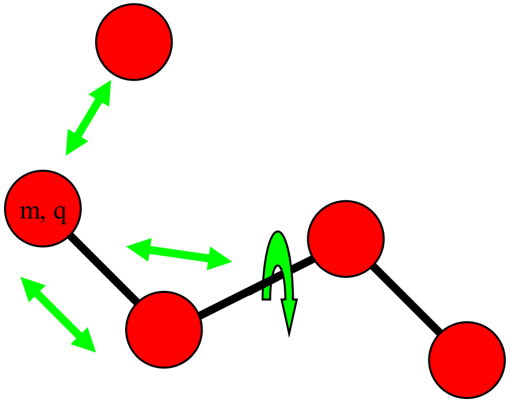
Molecular Dynamics Simulations

…are simulations of the time evolution of a chemical system at atomic details described by an empirically derived classical potential function

need:
- a cartesian representation of the chemical system
- a potential function appropriate for the studied subject
- a propagator to simulate system time evolution
- efficient numerical algorithms

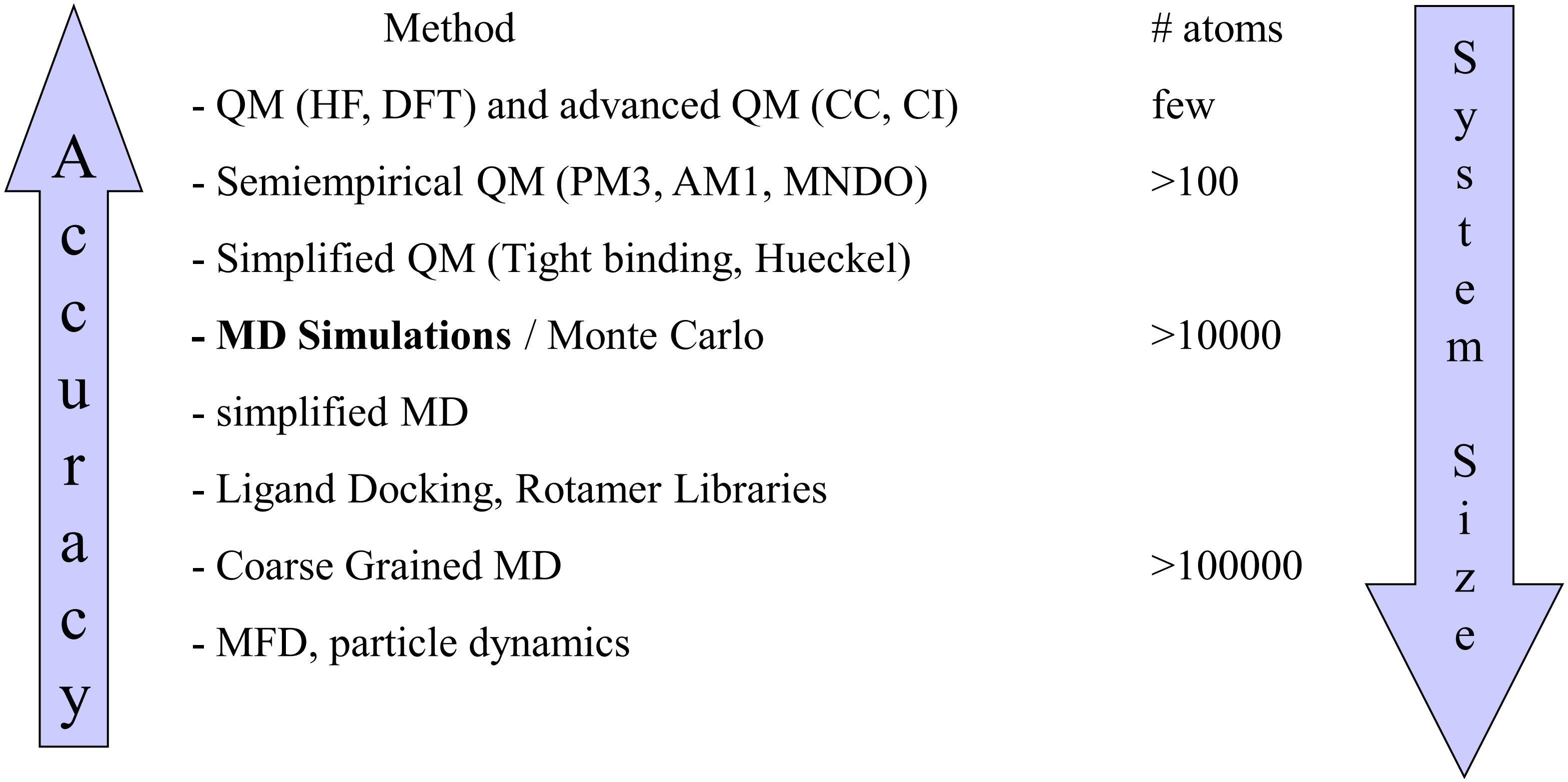
provide:
- a dynamic representation at the given temperature
- relative energies of conformers
- small scale conformational changes
- time dependent properties
- a description of noncovalent association

do NOT provide:
- chemical reactions
- quantum effects
- large barrier crossings
- precise energies and geometries
- unusual compounds



System: f (m, q, x, v, topology, V)

MD compared to other Modelling Techniques

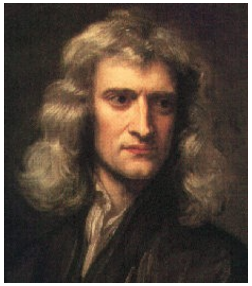


Accuracy

Method	# atoms
- QM (HF, DFT) and advanced QM (CC, CI)	few
- Semiempirical QM (PM3, AM1, MNDO)	>100
- Simplified QM (Tight binding, Hueckel)	
- MD Simulations / Monte Carlo	>10000
- simplified MD	
- Ligand Docking, Rotamer Libraries	
- Coarse Grained MD	>100000
- MFD, particle dynamics	

System Size

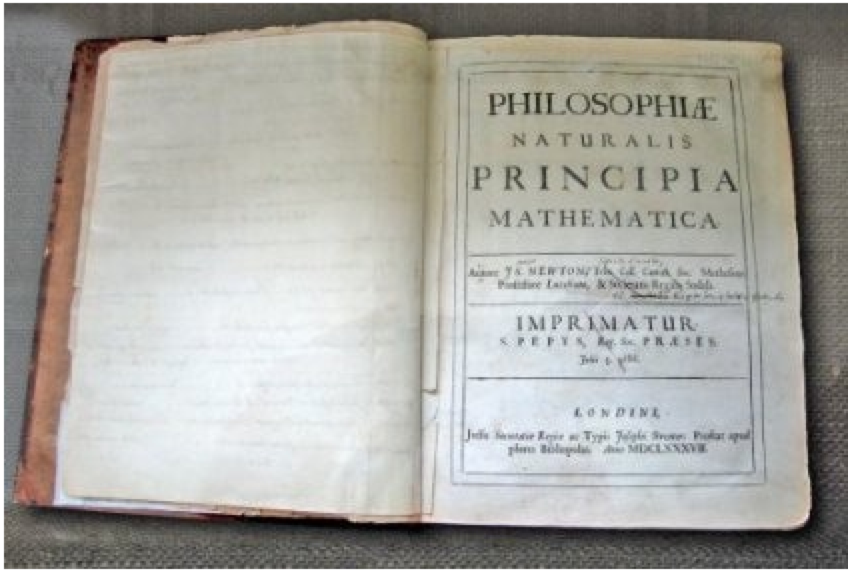
Newtons Equations of Motion



Molecular Mechanics, 19th century style:

$$\vec{F}_i = -\frac{\partial V}{\partial \vec{r}_i} \qquad \vec{F}_i = m_i \vec{a}_i$$

$$x(t+\Delta t) \approx x(t) + \Delta t \left[\frac{dx}{dt}\right]_t + \frac{\Delta t^2}{2}\left[\frac{d^2x}{dt^2}\right]_t$$



PHILOSOPHIÆ

NATURALIS

PRINCIPIA

MATHEMATICA

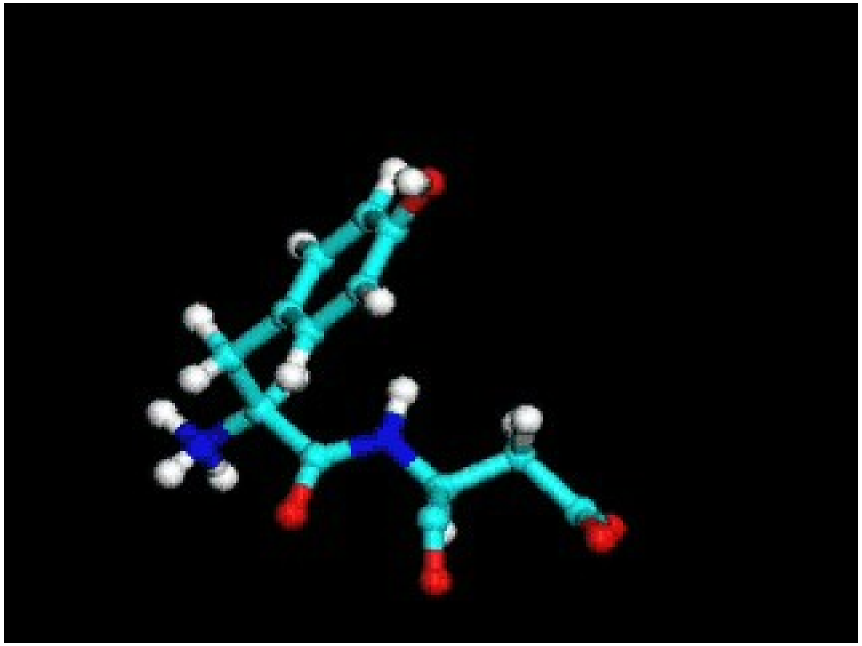
IMPRIMATUR

S. PEPYS, Reg. Soc. PRÆSES.

LONDINI

Not really true at atomic scale, but often a good model with added empirical parameters

From structure/parameters to forces to dynamics



Atoms are Quantum no?

- It is difficult to accept that the behavior of atomistic systems, which behave according to quantum rather than classical laws, could be accurately described by the application of classical Newtonian mechanics.
- The justification for this can be made by considering the de Broglie expression for the thermal wavelength Λ

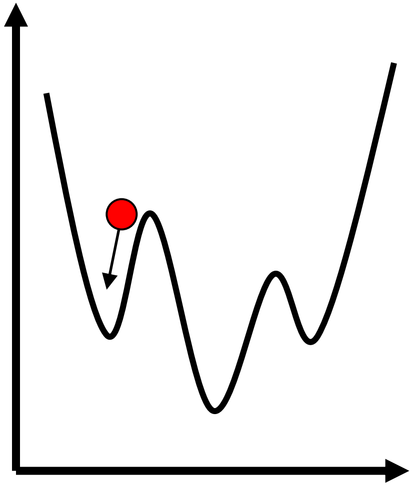
$$\Lambda = \sqrt{\frac{2\pi\hbar^2}{Mk_B T}}$$

- where T is the temperature and M is the atomic mass.
- The approximation of classical behavior holds if $\Lambda \ll \alpha$, where α is the mean nearest neighbor separation.
- This holds for ‘heavy’ liquid systems at all but the lowest temperatures, at which quantum effects become important.

Dynamics, Minimizations, Monte Carlo

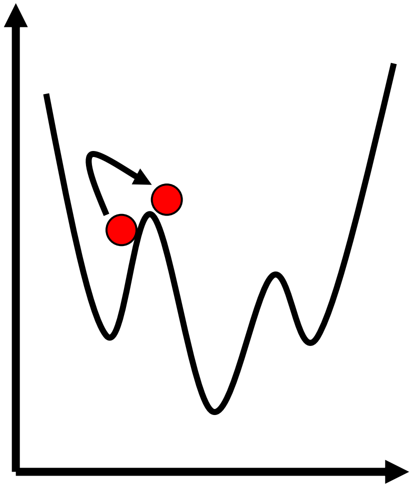
No analytical solutions for interesting systems

Minimizations



$$\Delta x = -\nabla V(x)$$

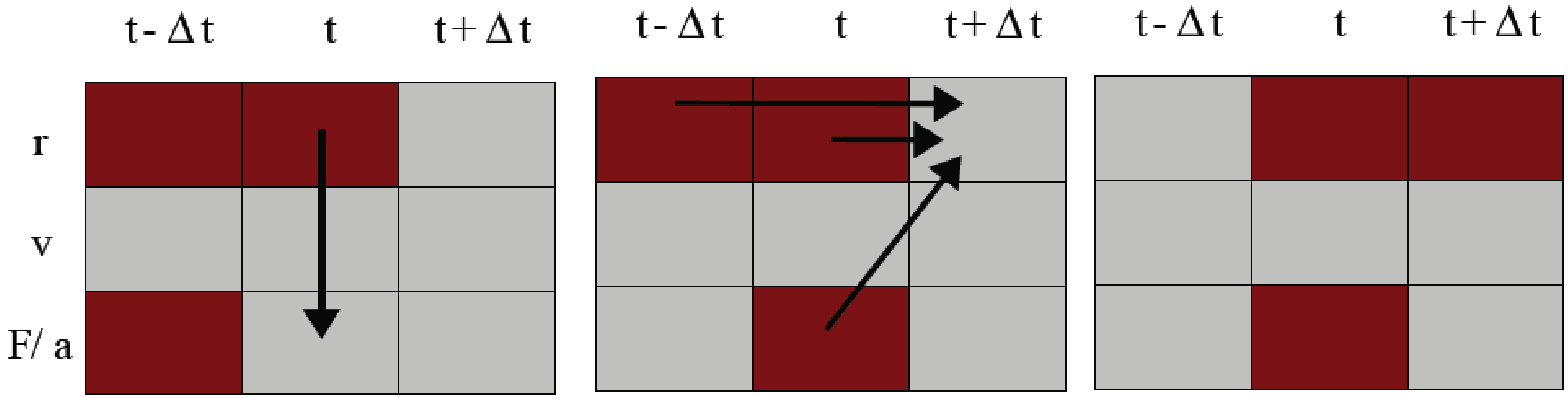
Monte Carlo



$$P(\Delta x) = \min(1, e^{(-\Delta E/kT)})$$

Molecular Dynamics

Time Evolution Propagator



Numerical algorithms introduce a timestep and truncate the Taylor expansion at the second derivative of the positions, i.e. the accelerations/forces

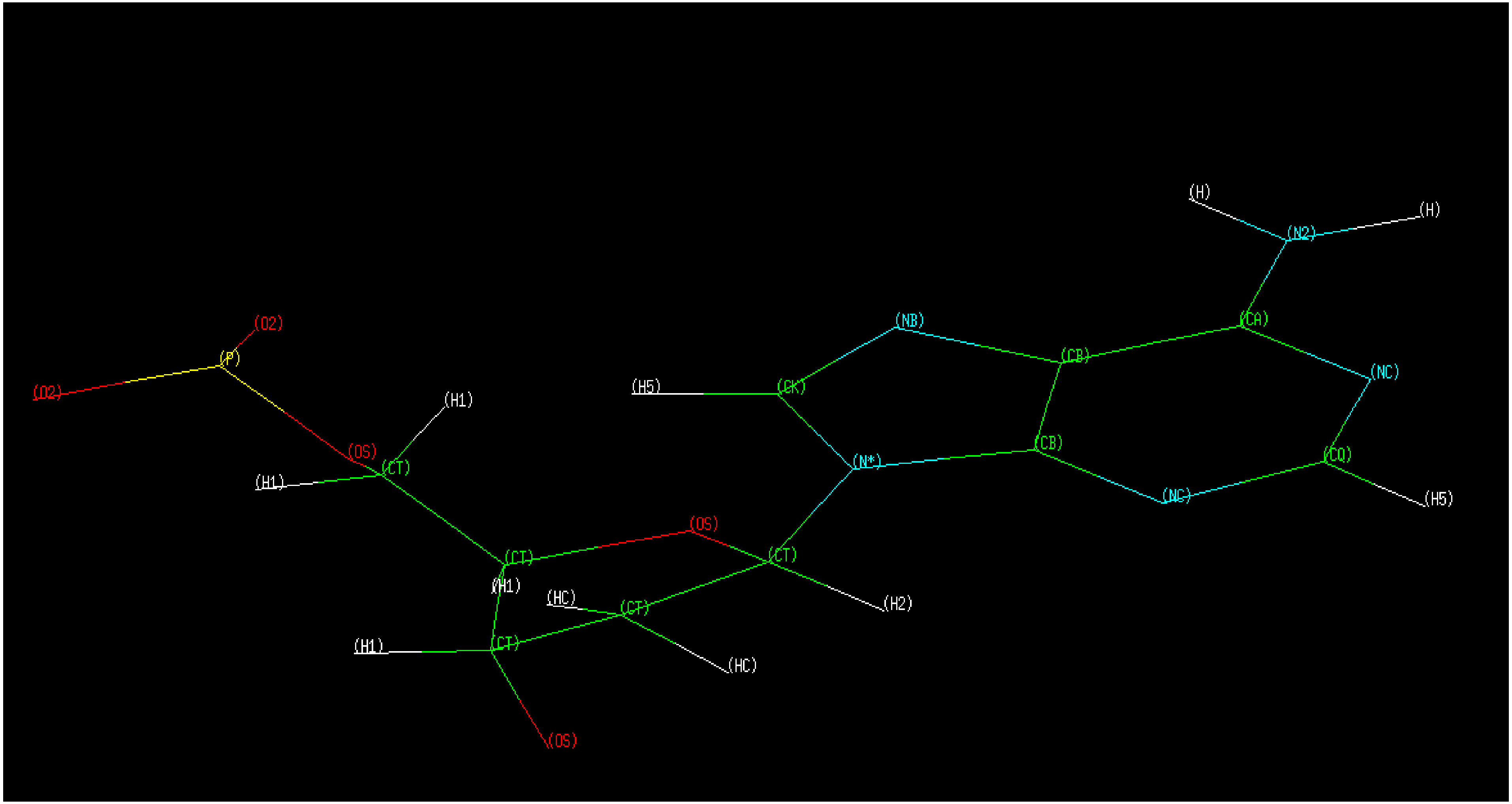
Parameters

$$V(x) = \sum_{bonds} K_b (r - r_0)^2 + \sum_{angles} K_a (\phi - \phi_0)^2 + \sum_{dihedrals} \frac{K_d}{2} \left[1 + \cos(n\vartheta + \gamma)\right] + \sum_{i,j<i} \left(\frac{A}{r^{12}} - \frac{C}{r^6} + \frac{q_i q_j}{r} \right)$$

From the amber gaff.dat general atom forcefield:

```
AMBER General Force Field for organic mol., add. info. at the end (June, 2003)
c  12.01         0.616               Sp2 C carbonyl group
c1 12.01         0.360               Sp C
c2 12.01         0.360               Sp2 C
c3 12.01         0.878               Sp3 C
ca 12.01         0.360               Sp2 C in pure aromatic systems
cp 12.01         0.360               Head Sp2 C that connect two rings in biphenyl sys.
cq 12.01         0.360               Head Sp2 C that connect two rings in biphenyl sys. identical to cp
[…]
ha 1.008         0.135               H bonded to aromatic carbon
hc 1.008         0.135               H bonded to aliphatic carbon without electrwd. group
hn 1.008         0.161               H bonded to nitrogen atoms
ho 1.008         0.135               Hydroxyl group
hp 1.008         0.135               H bonded to phosphate
hs 1.008         0.135               Hydrogen bonded to sulphur
hw 1.008         0.135               Hydrogen in water
hx 1.008         0.135               H bonded to C next to positively charged group
f  19.00         0.320               Fluorine
cl 35.45         1.910               Chlorine
br 79.90         2.880               Bromine
i  126.9         4.690               Iodine
n  14.01         0.530               Sp2 nitrogen in amide groups
[…]
c2-no  327.6    1.463        SOURCE3     4    0.0013    0.0013
c2-o   546.2    1.261        SOURCE3     4    0.0144    0.0144
c2-oh  425.4    1.333        SOURCE1    53    0.0000    0.0000
c2-os  392.6    1.357        SOURCE1   315    0.0088    0.0097
c2-p2  375.9    1.670        SOURCE3    62    0.0078    0.0147
[…]
<angles>
<dihedrals>
```

Atom typing


(H)
(H)
(N2)
(O2)
(NB)
(CA)
(P)
(CB)
(NC)
(O2)
(H5)
(CK)
(H1)
(OS)
(CT)
(CB)
(CQ)
(N*)
(H1)
(NC)
(H5)
(OS)
(CT)
(CT)
(H1)
(HC)
(CT)
(H2)
(H1)
(CT)
(HC)
(OS)

The Amber forcefield family

Amber contains a variety of continously improving force fields:

General Biomolecules:

leaprc.ff86	Weiner *et al.* 1986	parm91X.dat
leaprc.ff94	Cornell*et al.* 1994	parm94.dat
leaprc.ff96	"	parm96.dat
leaprc.ff98	"	parm98.dat
leaprc.ff99	"	parm99.dat
leaprc.ff03	Duan*et al.* 2003	parm99.dat+frcmod.ff03
leaprc.ff03ua	Yang *et al.* 2003	parm99.dat+frcmod.ff03+frcmod.ff03ua
leaprc.ff02	reduced (polarizable) charges	parm99.dat+frcmod.ff02pol.r1
leaprc.ff02EP	" + extra points	parm99EP.dat
leaprc.ff99SB	"	parm99.dat+frcmod.ff99SB
leaprc.ff99bsc0	BSC	parm99.dat+frcmod.ff99SB+frcmod.parmbsc0
leaprc.ff10	BSC0+ff99SB+Ions08	parm10.dat

General organic molecules

leaprc.gaff	none	gaff.dat

Carbohydrates

leaprc.glycam04	Woods *et al.*	glycam04.dat
leaprc.glycam04EP	"	glycam04EP.dat

Infinite Systems

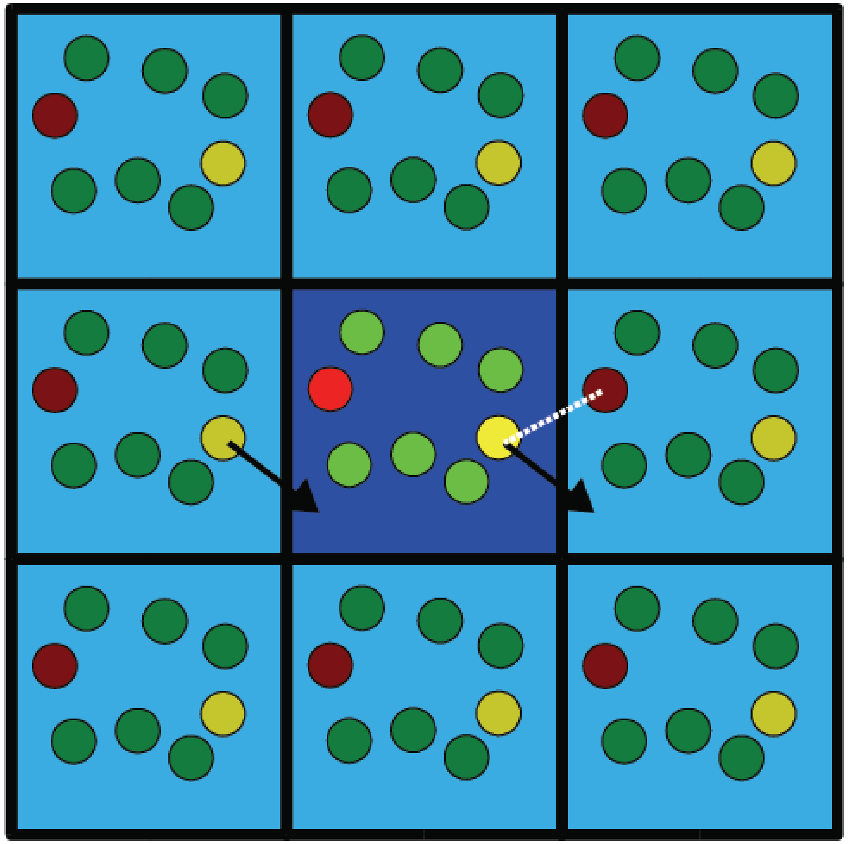
- Real systems are almost infinite in size

- In vacuo simulation are rarely a good idea

Periodic boundary conditions: Simulation boxes and nearest image conventions

Infinite electrostatics: Ewald summation

Problem:
physical properties depend on box size



System Sizes and Time Scales

The extremes:

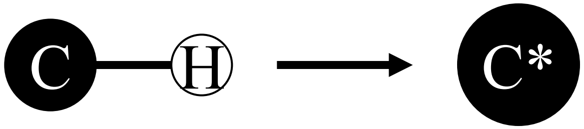
1977:	BPTI, 500 atoms, 10 ps, in vacuo
2002:	F_0F_1-ATPase, 300 kDa Protein, 1 ns, large scale conformational changes
2006:	TMV, 1 Mio atoms, 50 ns
	Villin Headpiece, 20k atoms, 500 μs

State of the art simulations on moderate computational ressources:

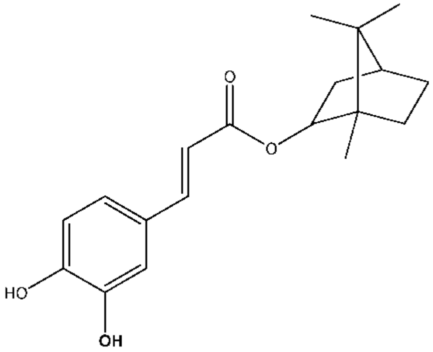
10k-100k atoms,	GOOD	-> most proteins are accesible
Full solvation, PBC	GOOD	-> physically meaningful parameters
1-10 ns simulation time	BAD	-> many biochemical system operate on longer timescales

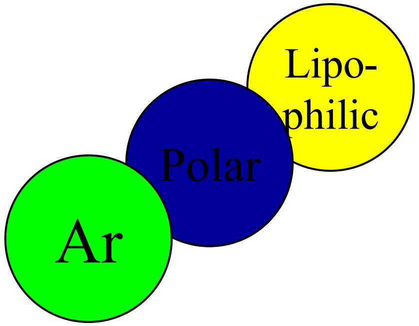
Coarse Graining

- United Atom Model:

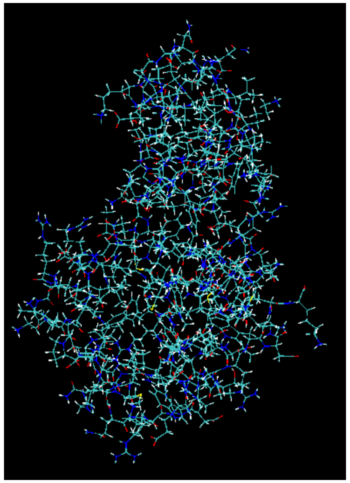


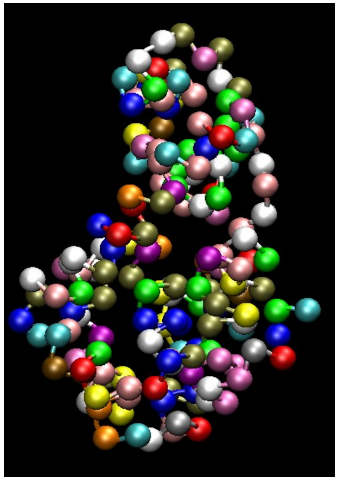
Molecular Fragment Dynamics:





Simplified proteins:

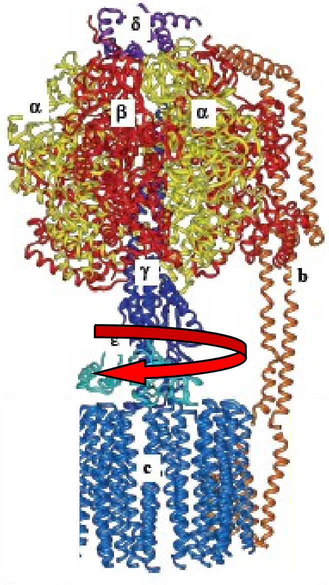




Guided Dynamics

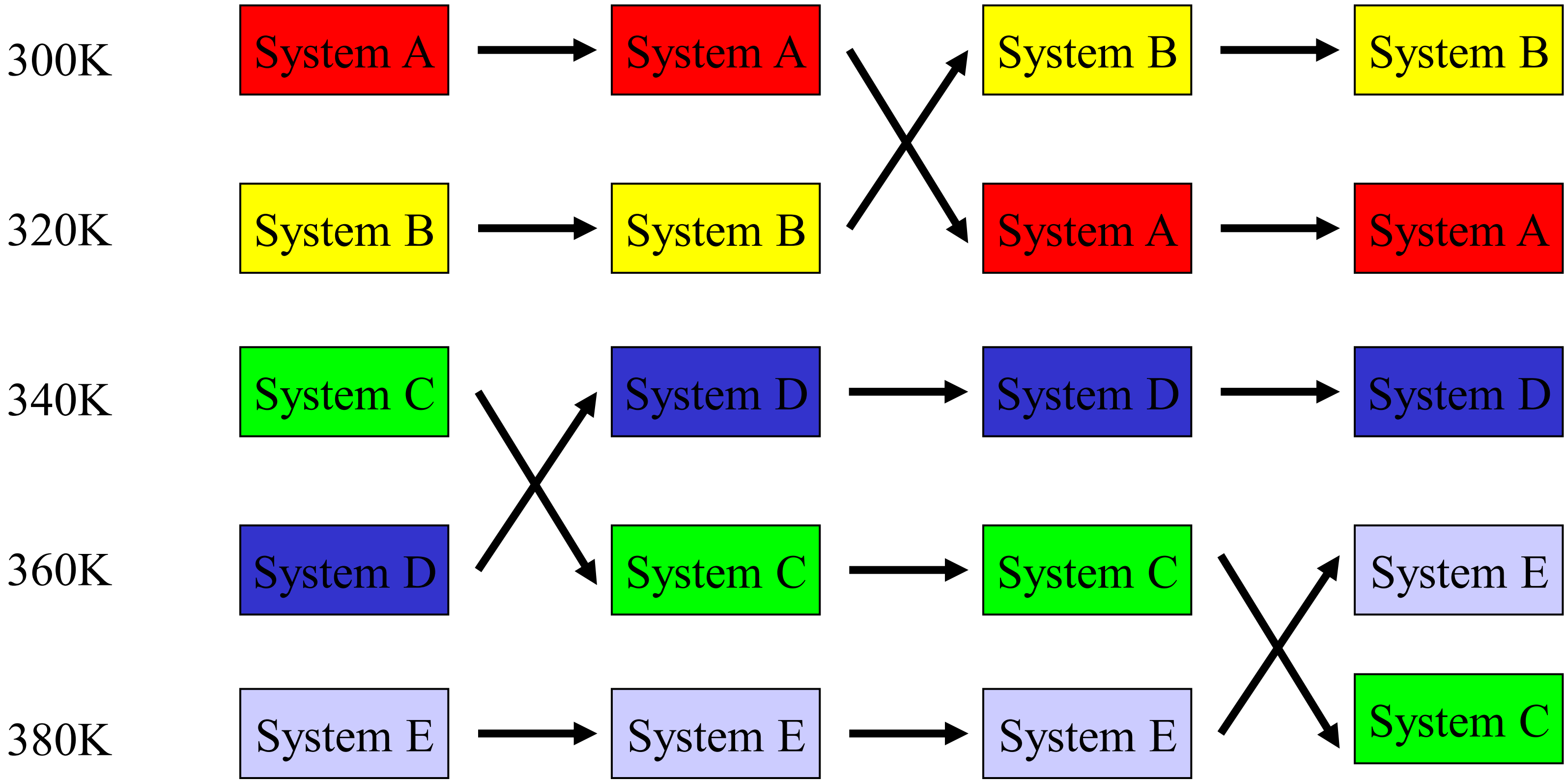
Add an additional biasing potential to sample regions of interest

- Steered MD
- Targetted MD



Replica Exchange

Run several Replica of your system at different temperatures and swap them occasionally

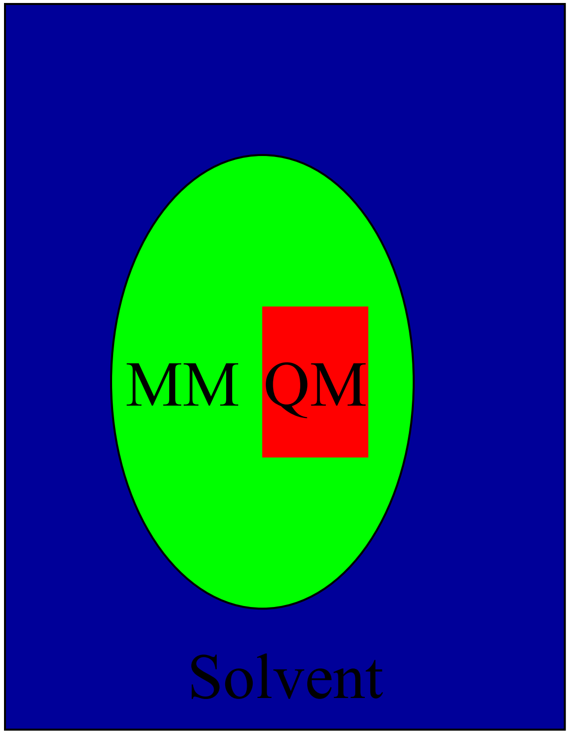


300K	System A	System A	System B	System B
320K	System B	System B	System A	System A
340K	System C	System D	System D	System D
360K	System D	System C	System C	System E
380K	System E	System E	System E	System C

QM/MM Hybrid Models

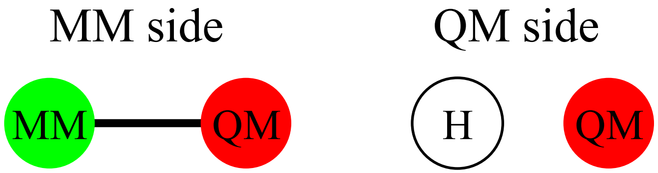
Treat the most interesting part of your system quantum mechanically, the rest by the forcefield.

Many applications in biochemical reactions mechanisms



Problems:

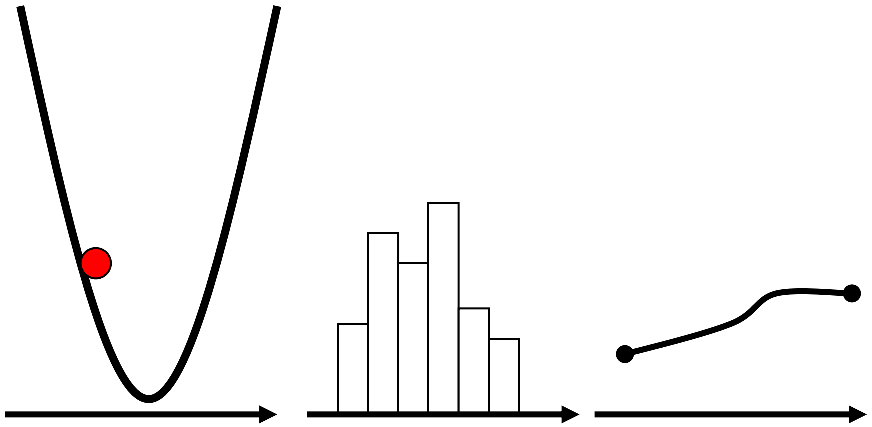
- one sided polarization
- vdW interactions between QM and MM
- boundary crossings and link atoms:

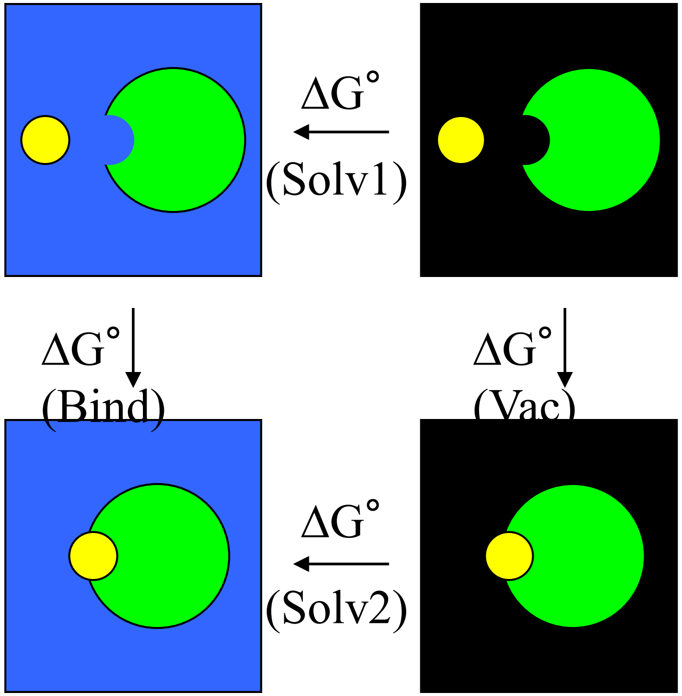


Free Energy Methods

- MM-PBSA
- Umbrella Sampling & WHAM
- Linear Interaction Energy
- FEP & Thermodynamic Integration

(-> see afternoon session…)





The MM-PBSA thermodynamic cycle