CDFD

Introduction to Molecular Mechanics

for

M.Tech. (Bioinformatics)

Hampapathalu A. Nagarajaram
Laboratory of Computational Biology
Centre for DNA Fingerprinting & Diagnostics (CDFD)
Gandipet Campus
Hyderabad

www.cdfd.org.in
han@cdfd.org.in

CDFD

Computational Biology Approaches to Protein Structure Prediction

- Simulations and *Ab initio* methods

- Knowledge based
  - Threading/Fold-recognition
  - Comparative/homology modelling

What determines fold?

- Anfinsen's experiments in 1957 demonstrated that proteins can fold spontaneously into their native conformations under physiological conditions. This implied that primary structure does indeed determine folding or 3-D stucture.

Protein folding problem!

Given a protein sequence how to predict its structure?

The Factors

- Physical properties of protein that influence stability & therefore, determine its fold:
  - Rigidity of backbone
  - Amino acid interaction with water
    - Hydropathic nature of side chains
  - Interactions among amino acids
    - Electrostatic interactions
    - Hydrogen, disulphide bonds
    - Volume constraints

Pioneering Contributions

Pauling

- “Nature of Chemical Bonding”

Semi-empirical formulations:

- Scheraga

- Ramachandran

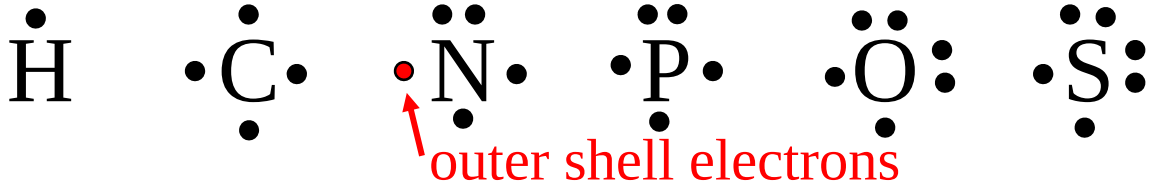
- Karplus, Mc Cammon, Gunstern

Quantum mechanical formulations:

- Pullman and Pullman

Covalent Interactions (bonds)

The most stable bond type: involves electron sharing



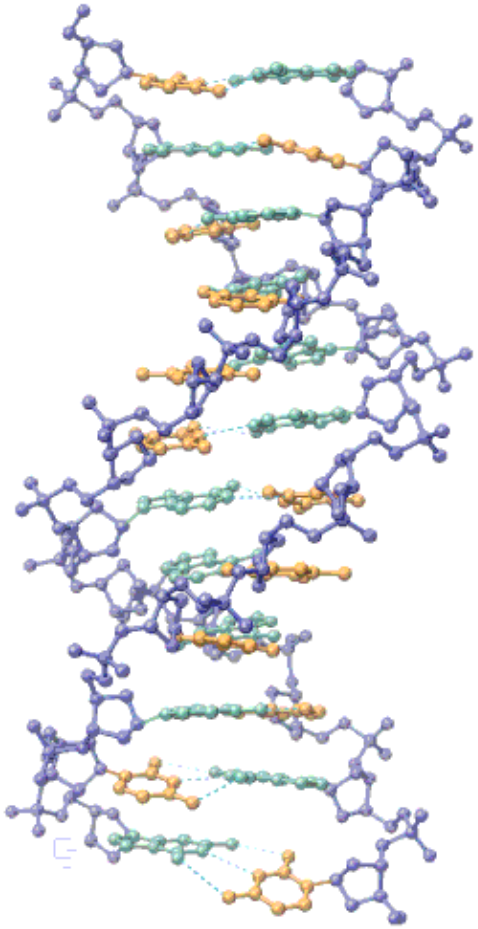
Each type of atom forms a characteristic number of covalent bonds with other atoms; shortest bond lengths

Covalent bonds are strong and require a lot of energy to break

thermal energy at 25^{o} (i.e. in cell) is 1 kcal/mol

C-C bond requires 83 kcal/mol to break

Non-covalent Interactions (Bonds)



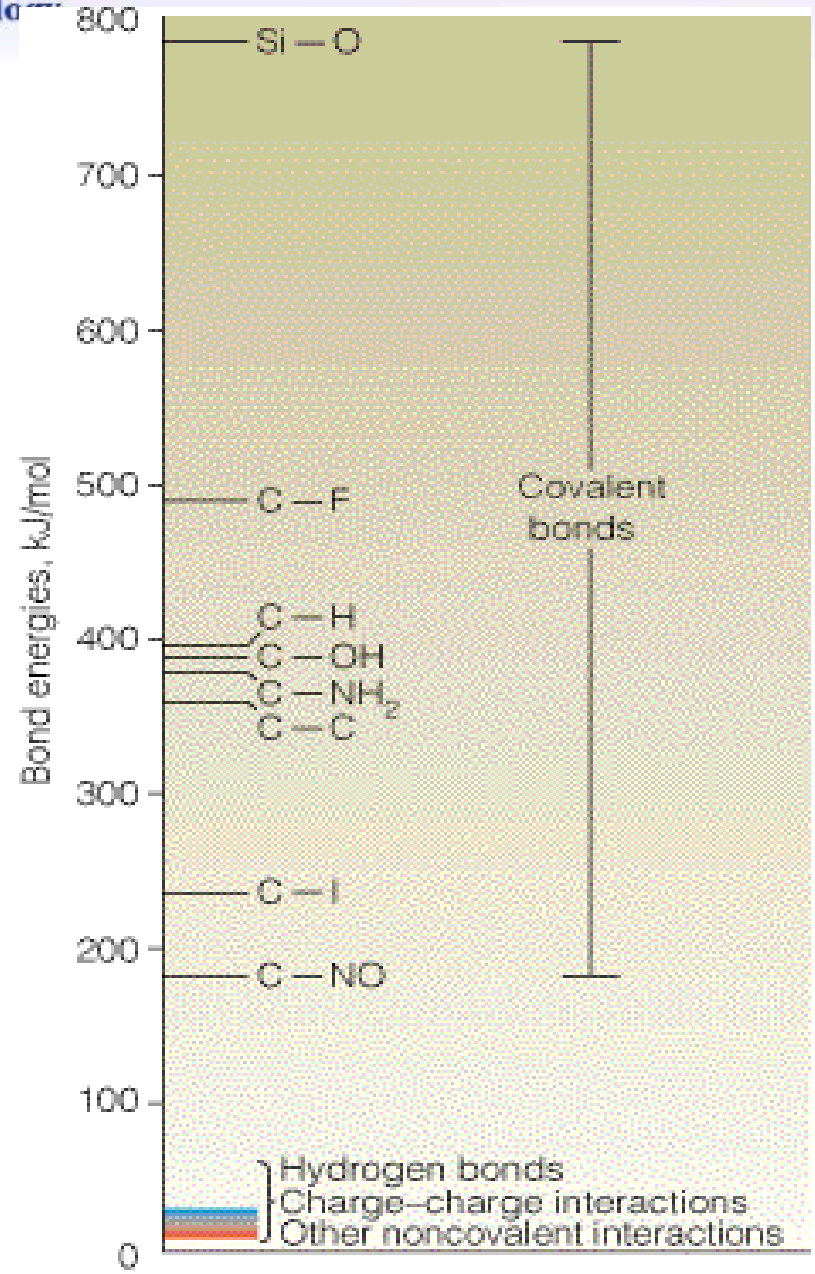
Much weaker than covalent bonds

- these bonds break and reform at Room Temperature (RT)

'Transient Bonds'

- however, <u>cumulatively</u> they are very effective e.g. α helix for proteins and double helix for DNA

Bond energies



These bonds are strong (stable) but not so strong as to not being possible to break

most biologically relevant covalent bonds

These bonds are continuously broken and reformed

all biologically relevant noncovalent bonds

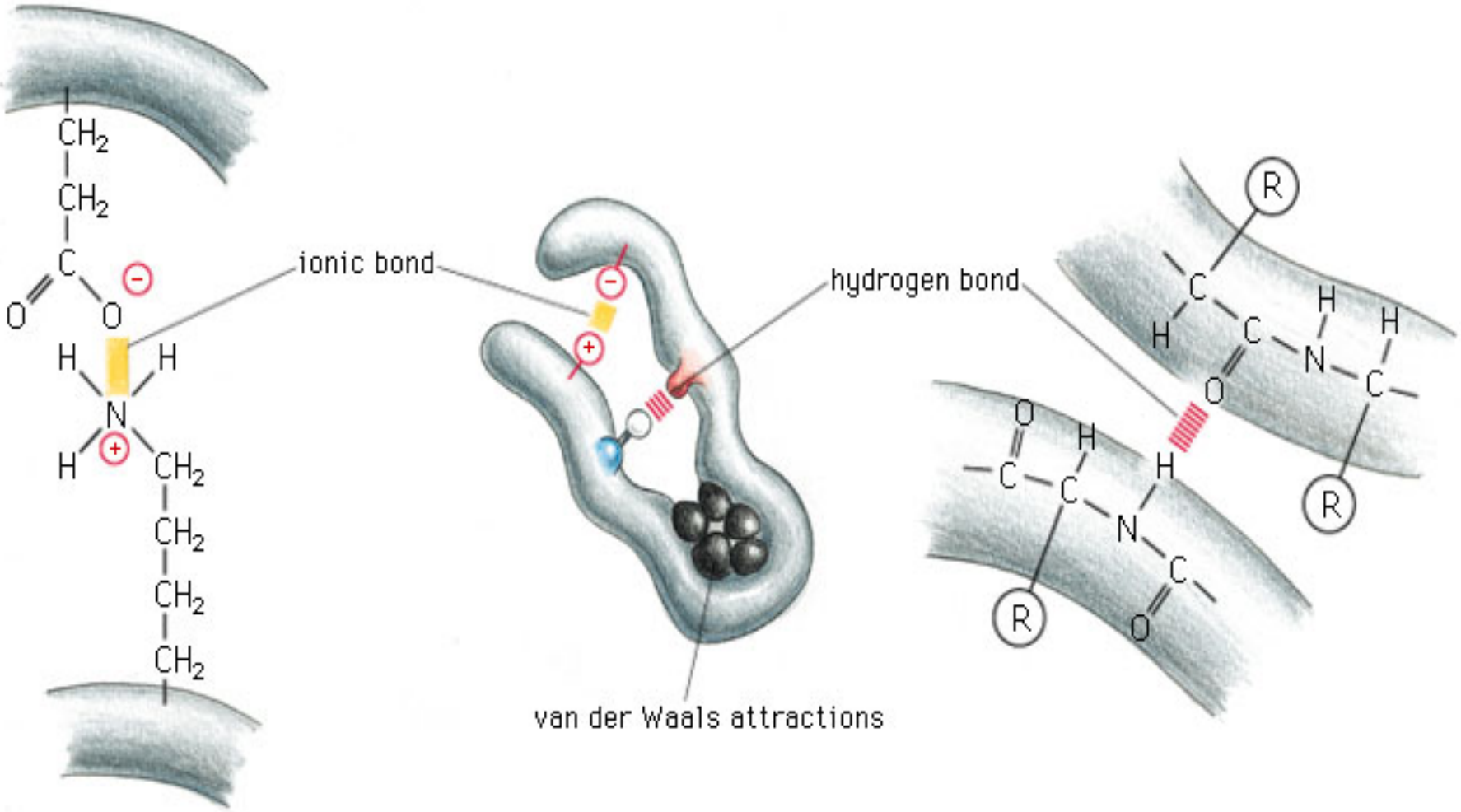
Non-covalent interactions

- Ionic bond (electrostatic interaction)
- Van der Waals interactions
- Hydrogen bond
- Hydrophobic interactions

Computational Biology

CDFD

Non-covalent interactions



Charge-charge interactions: energy of interaction

Energy *of Interaction (U)*

-energy required to separate two charged particles from distance r to infinite distance

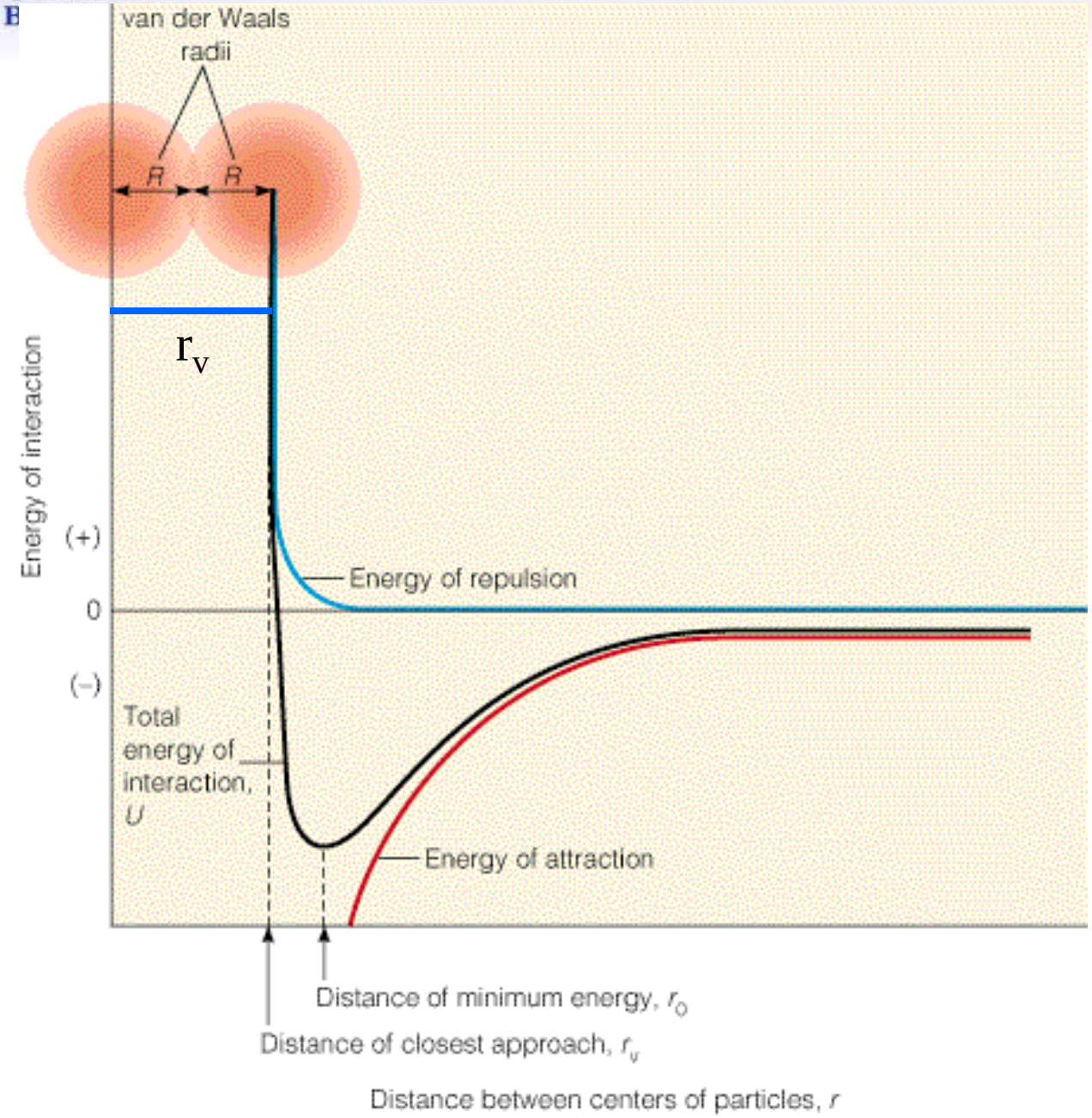
$$U = k\frac{q_1 \times q_2}{\varepsilon r}$$

Interaction/repulsion between charges:

- *is nondirectional*
- *varies linearly with distance (as opposed to some of the other interactions)*

(as *r* becomes very large, U approaches *zero*)

van der Waals interaction



van der Waals radii

	R (nm)
Atoms	
H	0.12
O	0.14
N	0.15
C	0.17
S	0.18
P	0.19
Groups	
—OH	0.14
—NH_2	0.15
—CH_2—	0.20
—CH_3	0.20
Half-thickness of aromatic ring	0.17

Note:

Covalently – bonded atoms are <u>closer</u> than the van der waals distance

“6-12 potential”

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

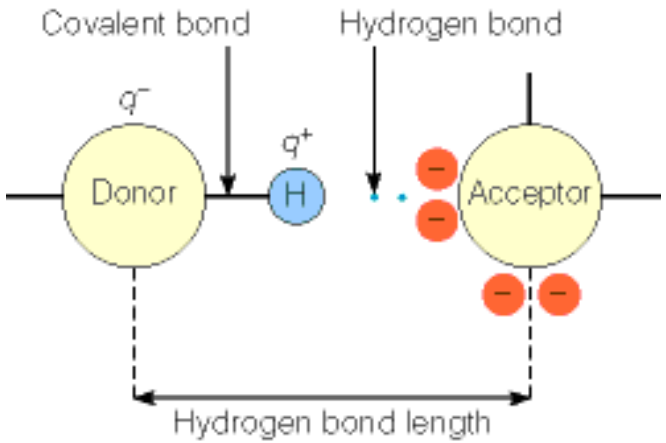
“Buckingham potential” :

Repulsive term is represented by an exponential of separation distance

Hydrogen Bonds

VERY important.

An interaction between a covalently bonded hydrogen atom on a *donor group* and a pair of non-bonded electrons on an *acceptor group*.

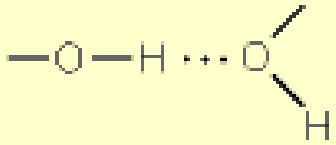
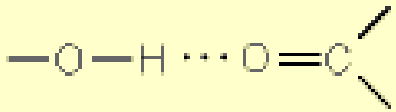
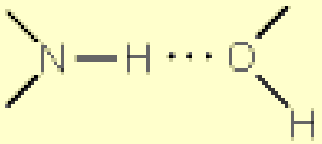
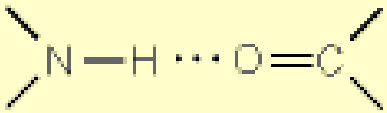
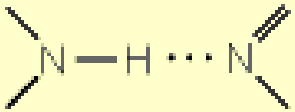
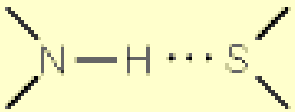


sufficiently electronegative

↓

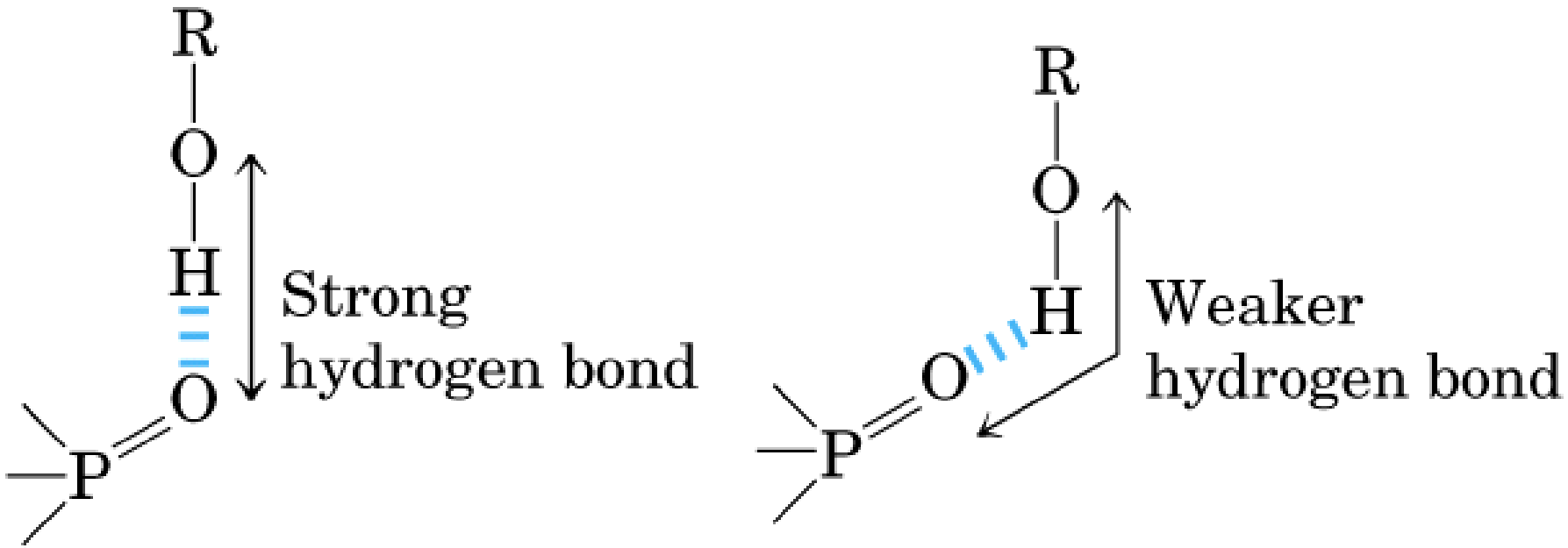
donor groups O-H, N-H

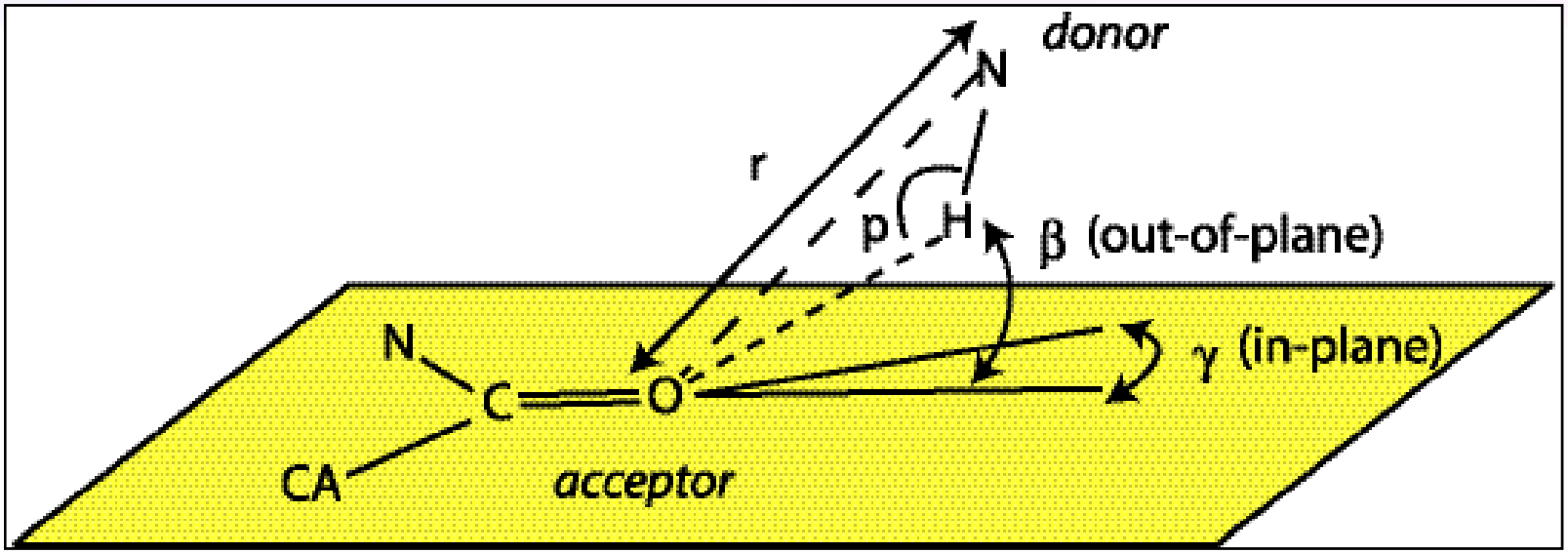
acceptor groups N, O, (S) (nonbonded electron pair)

Donor···Acceptor	Bond Length[a] (nm)	Comment
	0.28 ± 0.01	H bond formed in water
	0.28 ± 0.01	Bonding of water to other molecules often involves these
	0.29 ± 0.01	
	0.29 ± 0.01	Very important in protein and nucleic acid structures
	0.31 ± 0.02	
	0.37	Relatively rare; weaker than above

[a]Defined as distance from center of donor atom to center of acceptor atom. For example, in the N—H···O=C< bond it is the N—O distance.

Directionality of hydrogen bond





- Hydrogen bond not really a covalent ”bond”--not much orbital overlap.
- Model as an **electrostatic** interaction between two dipoles consisting of the H-N bond and the O sp^2 lone pair. In electrostatic theory, the optimal orientation of two such dipoles is ***head-to-tail***. The energy of such an arrangement should decrease as the head and tail are brought together as long as atomic van der Waals radii are not violated (then repulsive forces quickly take over).
- “Ideal” hydrogen bond in this model would have r~3.0 Å, p=180°, β=0° and γ=±60°. Convince yourself of this.
- In small molecule crystals, this is *approximately* what is observed, though there is a lot of variation in the angles β and γ. Thus the precise C=O…H angle parameters are not critical.
- Main chain-main chain hydrogen bonds found in proteins will show various deviations from this geometry, partly due to the topological constraints imposed by forming secondary structures.

Criteria for identifying hydrogen bonds in protein structures

- **What is a reasonable hydrogen bond? Criteria for identifying hydrogen bonds are somewhat arbitrary and many have been used. Here are a couple of examples.**

- **Geometric criteria:** Often H-bonds are just identified by two parameters, the O…N (acceptor-donor) *distance* r, and a O…H-N *angle* p. The angles describing the C=O…H geometry are ignored. Typical cutoffs: p > 120° and r < 3.5 Å. (Baker & Hubbard, 1984)

- 10-12 potential is used by to calculate the hydrogen bond energy

- **Electrostatic criteria:** One of the most commonly used criteria is a potential function based on a pure electrostatic model (Kabsch & Sander, 1983). Place partial positive and negative charges on the C,O ($+q_1,-q_1$) and N,H ($+q_2,-q_2$) atoms and compute a binding energy as the sum of repulsive and attractive interactions between these four atoms:

  $E=q_1q_2(1/r(ON)+1/r(CH)-1/r(OH)-1/r(CN))*f$

  where q_1=0.42e and q_2=0.20e, f is a dimensional factor (=332) to convert E to kcal/mol, and r(AB) is the interatomic distance between atoms A and B.

  A hydrogen bond is then identified by a binding energy less than some arbitrary cutoff, e.g. E< -0.5 kcal/mol.

- Note that the criteria defined above are only applicable *when hydrogen atom positions are available*. Crystal structures do not have hydrogens--however, their positions can be computed in many cases.

‘Hydrophobic’ Interaction

Hydrophobic, water-fearing
Hydrophilic, water-loving

Amphipathic, both H-philic and H-phobic, *e.g.*, lipids

NOT a true force.

The consequence of the energy needed to insert a non-polar molecule into water.

- water clatherate structure formed
- hydrogen bonds broken
- non-polar molecules form van der Waals interactions between each other

Various covalent and non-covalent interactions

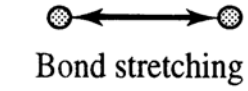
Protein 3D Structure

Total potential energy of the system

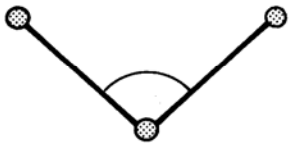
$$V(R) = E_{bonded} + E_{non-bonded}$$

$$E_{bonded} = E_{bond-stretch} + E_{angle-bend} + E_{rotate-along-bond}$$

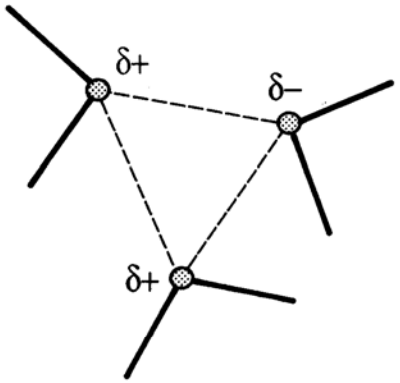
Empirical Force Fields



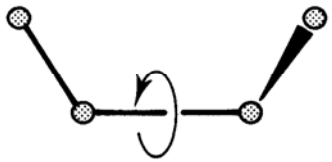
Bond stretching



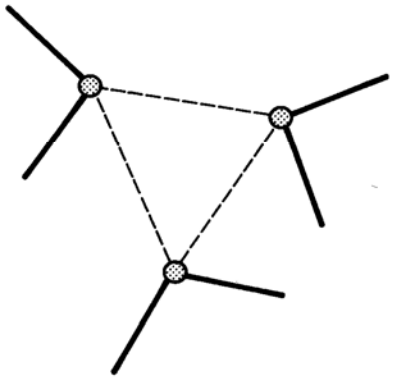
Angle bending



Non-bonded interactions
(electrostatic)



Bond rotation
(torsion)



Non-bonded interactions
(van der Waals)

- describe interaction of atoms or groups
- the parameters are "empirical", i.e. they are dependent on others and have no direct intrinsic meaning
- Examples:
  GROMOS96 (van Gusteren)
  CHARMM (M. Karplus)
  AMBER (Kollman)

Force constant

$$E_{bond-stretch} = \sum_{1,2\,pairs} K_b \left(b - b_0\right)^2$$

Bond length

Canonical average

Hormonic potential

$$E_{bond-bend} = \sum_{angles} K_\theta (\theta - \theta_0)^2$$

$$E_{rotate-along-bond} = \sum_{1,4\,pairs} K_{\phi}\left(1-\cos\left(n\phi\right)\right)$$

$$E_{non-bonded} = E_{van-der-Waals} + E_{electrostatic}$$

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

"6-12 potential"

$$E_{electrostatic} = \sum_{\substack{nonbonded \\ pairs}} \frac{q_i q_k}{D r_{ik}}$$

q_i and q_k are the charges
r_{ik} the distance between them
D dielectric constant

Force constants:

Bonds	K
Cα-H	472.0
N-H	810.0
N-C	806.0
C=O	1090.0
Cα-C	274.0
Cα-N	522.0
Cα-Cβ	220.0

Angles

C-N-H	53.2
C-N-Cα	109.0
Cα-N-H	62.8
N-C=O97.0	
Cα-C-N	66.2
O=C-Cα	81.8
C-Cα-N	66.2

Rotations

N-Cα	0.6	-1	3
Cα-C	0.2	+1	3
C-N	20.0	-1	2

Force field

The various potential energy functions along with all the force constants and canonical average values (standard values) together define a "Force-field".

Different force-fields are available which give empirical description of atomic interactions in proteins, nucleic acids and lipids, sugars, organic and inorganic molecules.

For proteins and nucleic acids:

Ex: AMBER (Assisted Model Building with Energy Refinement)

V(R) describes an n-dimensional energy surface where n= 3(#of atoms)

Conformational Space

How to sample conformational space?
(What are the probable conformations?)
The Most asked question in Molecular mechanics

It is assumed that biologically active conformation of a biomolecule corresponds to either global minimum or one of the local minima close to the global minimum.

So.. given a biomolecule search is made in the conformational space (energy space) to look for various energy minima.

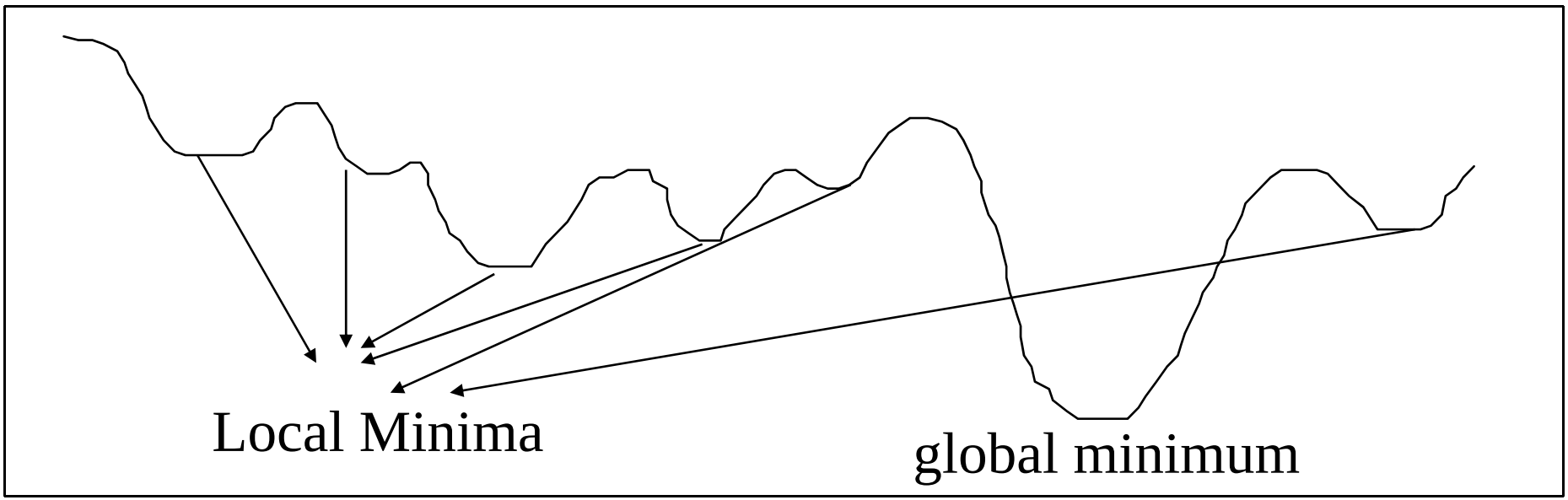
Conformational search

a) Systematic search

b) Random search

Energy Minimization

- Local minimum vs global minimum
- Many local minima; only ONE global minimum
- Methods: Newton-Raphson (block diagonal), steepest descent, conjugate gradient, others.



CDFD

Molecular Dynamics: Introduction

- At physiological conditions, the biomolecules undergo several movements and changes
- The time-scales of the motions are diverse, ranging from few femtoseconds to few seconds
- These motions are crucial for the function of the biomolecules

An overview of various motions in proteins (

Computational Biology

CDFD

Motion	Spatial extent (nm)	Log_{10} of characteristic time (s)
Relative vibration of bonded atoms	0.2 to 0.5	-14 to -13
Elastic vibration of globular region	1 to 2	-12 to –11
Rotation of side chains at surface	0.5 to 1	-11 to –10
Torsional libration of buried groups	0.5 to 1	-14 to –13

An overview of various motions in proteins (2)

Computational Biology

CDFD

Motion	Spatial Extent (nm)	Log_{10} of characteristic time (s)
Relative motion of different globular regions (hinge bending)	1 to 2	-11 to –7
Rotation of medium-sized side chains in interior	0.5	-4 to 0
Allosteric transitions	0.5 to 4	-5 to 0
Local denaturation	0.5 to 1	-5 to 1
Protein folding	???	-5 to 2

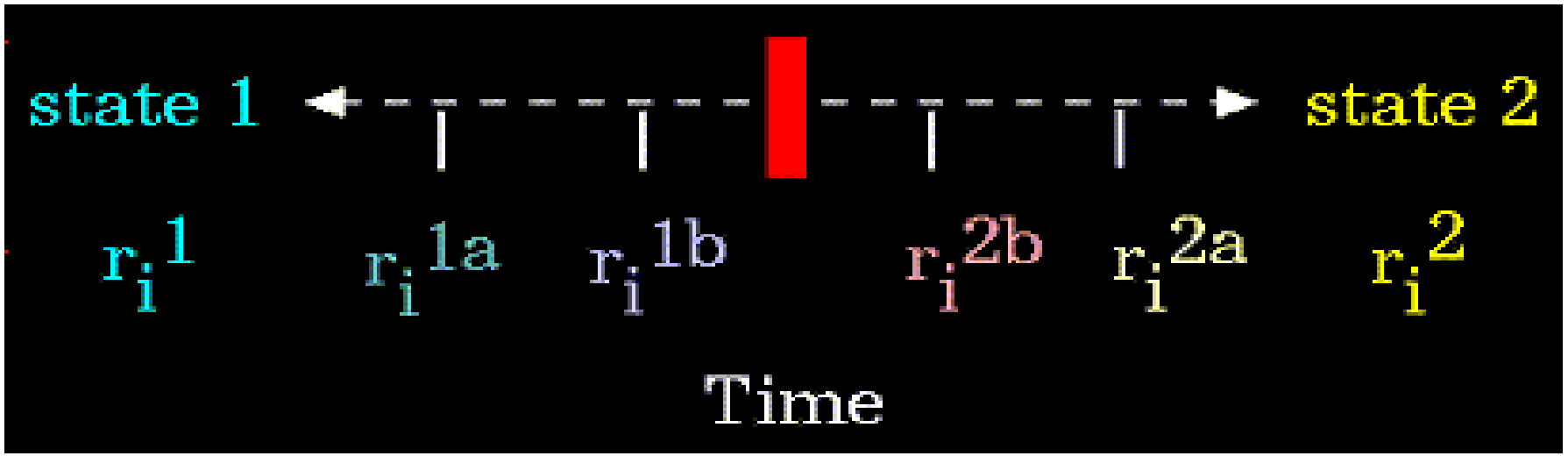
Thermodynamics state 1 state 2 state n

Kinetics state 1 ⟷ state 2 ⟷ state n

Thermodynamics → what states are possible??

Kinetics → how (and how fast) do states interconvert??

Conformational transitions and local vibrations are the usual subjects of molecular dynamics studies. Molecular dynamics alters the intramolecular degrees of freedom in a step-wise fashion, analogous to energy minimization. The individual steps in energy minimization are merely directed at establishing a down-hill direction to a minimum. The steps in molecular dynamics, on the other hand, meaningfully represent the changes in atomic position, r_i, over time (i.e. velocity).



Calculating of physical properties (which are macroscopic) of a system

Ergodic Hypothesis:

Ensemble average = Time average

If you simulate a system sufficiently a long time
you would have visited all the conformational states!

But how long we have to simulate?

The molecular dynamics simulation method is based on Newton's second law or the equation of motion,

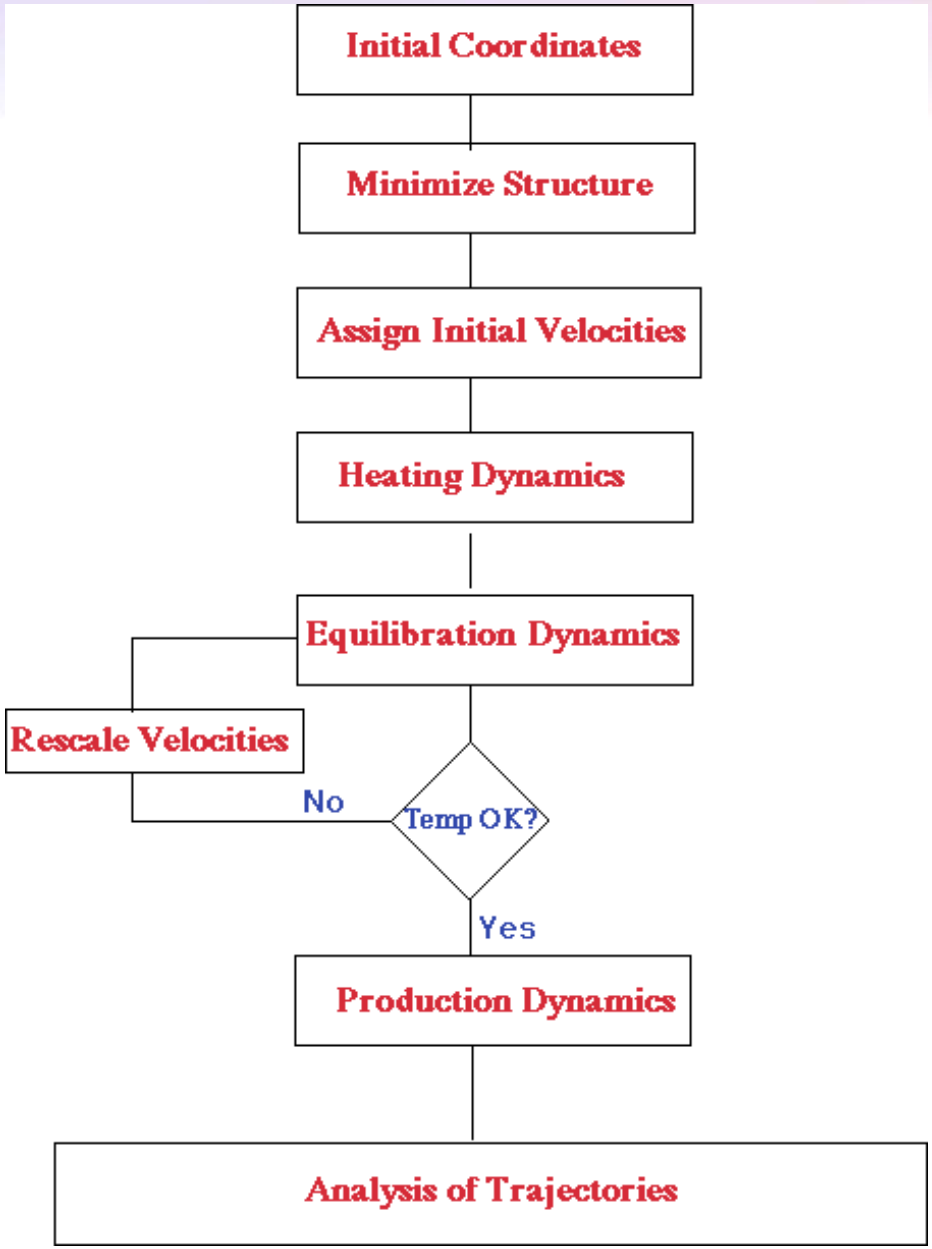
F=ma, where **F** is the force exerted on the particle, **m** is its mass and **a** is its acceleration.

From a knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory (time evolution of a given property) that describes the positions, velocities and accelerations of the particles as they vary with time.

From this trajectory, the average values of properties can be determined.

The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past.

Molecular dynamics simulations can be time consuming and computationally expensive. However, computers are getting faster and cheaper. Simulations of solvated proteins are calculated up to the nanosecond time scale, however, simulations into the millisecond regime have been reported.


Initial Coordinates
Minimize Structure
Assign Initial Velocities
Heating Dynamics
Equilibration Dynamics
Rescale Velocities
No
Temp OK?
Yes
Production Dynamics
Analysis of Trajectories

Newton's equation of motion is give by where *Fi* is the force exerted on particle *i*, *mi* is the mass of particle *i* and *ai* is the acceleration of particle *i*.

$$F_i = m_i a_i$$

The force can also be expressed as the gradient of the potential energy,

$$F_i = -\nabla_i V$$

Combining these two equations yields

where *V* is the potential energy of the system. **Newton's** equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

To do this, we should know

at given time t,

- initial position of the atom

$$x_1$$

- its velocity

$$v_1 = dx_1/dt$$

- and the acceleration

$$a_1 = d^2x_1/dt^2 = m^{-1}F(x_1)$$

Equations of Motion from classical mechanics

The position x_2 $$x_2 = x_1 + v_1 \Delta t$$ d be,

an

$$v_2 = v_1 + a_1 \Delta t = v_1 + m^{-1} F(x_1) \Delta t = v_1 - m^{-1} \left. \frac{dV}{dx} \right|_{x_1} \Delta t$$

In general, given the values x_1, v_1 and the potential energy V(x), the molecular trajectory x(t) can be calculated, using,

$$x_i = x_{i-1} + v_{i-1}\Delta t$$

$$v_i = v_{i-1} - m^{-1}\left.\frac{dV(x)}{dx}\right|_{x_{i-1}}\Delta t$$

The velocities, *vi,* are often chosen randomly from a Maxwell-Boltzmann or Gaussian distribution at a given temperature, which gives the probability that an atom i has a velocity v_x in the x direction at a temperature T.

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right]$$

The initial distribution of velocities are usually determined from a random distribution with the magnitudes conforming to the required temperature and corrected so there is no overall momentum, i.e.,

$$P = \sum_{i=1}^{N} m_i v_i = 0$$

The temperature can be calculated from the velocities using the relation where N is the number of atoms in the system.

$$T = \frac{1}{(3N)} \sum_{i=1}^{N} \frac{|p_i|}{2m_i}$$

Pressure – computed from Virial Theorem as due to positions of particles and forces acting on them

$$P = \frac{1}{V}\left[Nk_B T - \frac{1}{3} \sum_{i=1}^{N} \sum_{j=i+1}^{N} x_i f_{ij} \right]$$

Integration Algorithms (Dynamics of the system)

The potential energy is a function of the atomic positions (3N) of all the atoms in the system. Due to the complicated nature of this function, there is *no analytical solution* to the equations of motion; they must be solved numerically.

Numerous numerical algorithms have been developed for integrating the equations of motion. They are:

- Verlet algorithm
- Leap-frog algorithm
- Velocity Verlet
- Beeman's algorithm

Important: In choosing which algorithm to use, one should consider the following criteria:

- The algorithm should conserve energy and momentum.
- It should be computationally efficient
- It should permit a long time step for integration.

Integration Algorithms

All the integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion:

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \ldots$$

$$v(t+\delta t) = v(t) + a(t)\delta t + \frac{1}{2}b(t)\delta t^2 + \ldots$$

$$a(t+\delta t) = a(t) + b(t)\delta t + \ldots$$

Where *r* is the position, *v* is the velocity (the first derivative with respect to time), *a* is the acceleration (the second derivative with respect to time), etc.

Verlet algorithm

To derive one can write

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$

$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$

Summing these two equations, one obtains

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\delta t^2$$

The Verlet algorithm uses positions and accelerations at time *t* and the positions from time *t-dt* to calculate new positions at time *t+dt.* The Verlet algorithm uses no explicit velocities. The advantages of the Verlet algorithm are, *i)* it is straightforward, and *ii)* the storage requirements are modest . The disadvantage is that the algorithm is of moderate precision.

The Leap-frog algorithm

$$r(t+\delta t) = r(t) + v\left(t+\frac{1}{2}\delta t\right)\delta t$$

$$v\left(t+\frac{1}{2}\delta t\right) = v\left(t-\frac{1}{2}\delta t\right) + a(t)\delta t$$

In this algorithm, the velocities are first calculated at time *t+1/2dt*; these are used to calculate the positions, *r*, at time *t+dt*. In this way, the velocities *leap* over the positions, then the positions *leap* over the velocities. The advantage of this algorithm is that the velocities are explicitly calculated, however, the disadvantage is that they are not calculated at the same time as the positions. The velocities at time *t* can be approximated by the relationship:

$$v(t) = \frac{1}{2}\left[v\left(t-\frac{1}{2}\delta t\right) + v\left(t+\frac{1}{2}\delta t\right)\right]$$

The Velocity Verlet algorithm

This algorithm yields positions, velocities and accelerations at time *t*. There is no compromise on precision.

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$

$$v(t+\delta t) = v(t) + \frac{1}{2}[a(t) + a(t+\delta t)]\delta t$$

Beeman's algorithm

This algorithm is closely related to the Verlet algorithm

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{2}{3}a(t)\delta t^2 - \frac{1}{6}a(t-\delta t)\delta t^2$$

$$v(t+\delta t) = v(t) + v(t)\delta t + \frac{1}{3}a(t)\delta t + \frac{5}{6}a(t)\delta t - \frac{1}{6}a(t-\delta t)\delta t$$

The advantage of this algorithm is that it provides a more accurate expression for the velocities and better energy conservation. The disadvantage is that the more complex expressions make the calculation more expensive.

Improper Dihedral Angles

An improper is a "virtual" torsion angle defined so as to keep an atom and its bonded neighbours in a certain configuration

ex. 1 - a chiral carbon with four different neighbours

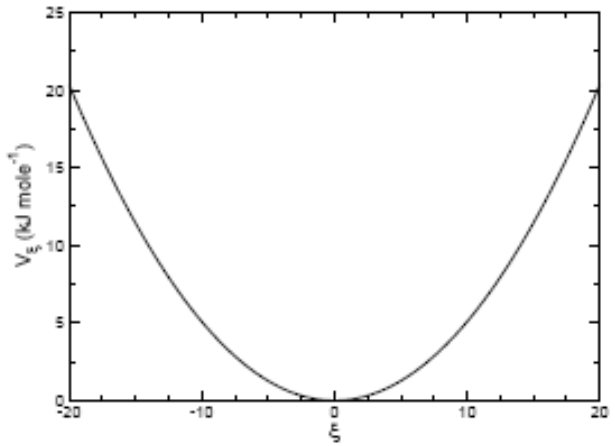
CA in all amino acids except Gly) CA-N-C-CB is an improper torsion (about the non-existent N-C bond);

it's value is ~+34 degrees for an L amino acid, -34 for a D amino acid

ex. 2 - a carboxylate carbon has to be flat (e.g., in aspartate); the improper torsion CG-CB-OD1-OD2 should be zero

Potential for improper dihedral angle is simply a harmonic potential as given below

$$V_{id}(\xi_{ijkl}) = \frac{1}{2} k_\xi (\xi_{ijkl} - \xi_0)^2$$



Solvent treatment in Molecular Dynamics

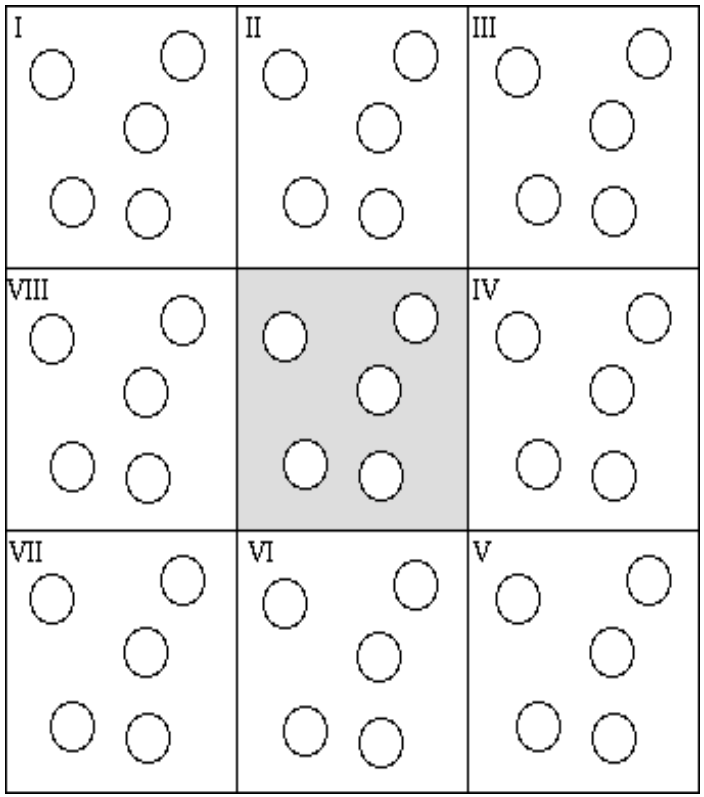
The most important effects of the solvent is the screening of electrostatic interactions. The electrostatic interaction between two charges is given by Coulomb's law,

$$V_{elec} = \frac{q_i q_j}{\varepsilon_{eff} r_{ij}}$$

where qi, qj are the partial atomic charges, ε_{eff} is the effective dielectric constant and rij is the relative distance between the two particles. ε_{eff} can be calcuated as the distant dependent dielectric constant as = $\varepsilon * r_{ij}$

Periodic Boundary conditions

Periodic boundary conditions enable a simulation to be performed using a relatively small number of particles in such a way that the particles experience forces as though they were in a bulk solution.



Furthermore, non-bonded cut-offs also used to reduce the list of interactions to be computed during each integration step

However, this affects long range electrostatic interactions. Therefore, Ewald summation or Particle Mesh Ewald (PME) is used for calculation of long range electrostatic interactions

Assignment: Write a note on PME.

Limit high-frequency motion

- SHAKE/LINCS algorithm:
  - restrain motion
  - most common use: hydrogen - heavy atom bonds
  - correct positions of H-atoms in such a way that r_{XH} remains constant
- allows somewhat bigger time step (1-2 fs)
- enhances stability of integration

Analysis

1. Snap shots saved during dynamics are used
2. Translation and rotations are removed by superposing to a reference structure which is usually the starting structure
3. Various properties are calculated.

$$RMS = \left\langle \left(r_i^{\alpha} - r_i^{\beta} \right)^2 \right\rangle^{\frac{1}{2}} = \sqrt{\frac{1}{N_i} \sum_i \left(r_i^{\alpha} - r_i^{\beta} \right)^2}$$

$$RMS_i^{fluct} = \sqrt{\frac{1}{N_f} \sum_f \left(r_i^f - r_i^{ave} \right)^2}$$

$$RadiusGyration = \sqrt{\frac{1}{Ni} \sum_i \left(r_i - r_{cm} \right)^2}$$

Termination of MD:

Molecular dynamics has no defined point of termination other than the amount of time that can be practically covered. Unfortunately, the current nano second order of magnitude limit is often not long enough to follow many kinds of state to state transformations, such as large conformational transitions in proteins.

- MD in the context of proteins is used for:
  - Protein Folding/Unfolding
  - NMR, X-ray structure determination
  - Homology modelling (MODELLER)
  - Conformational sampling
    - Start MD at temp T1
    - Simulate for long time and collect snap shots
    - EM snap shots; compare; collect all those that are conformationally distinct
    - Change temp and redo the whole exercise

Higher sampling using MD

- Higher the temperature the greater and faster the motion & more of conformational space sampled.
- Use
  - (a) to overcome energy barriers to find better structure
  - (b) explore motion

The Langevin or brownian dynamics:

Computational Biology

CDFD

In the Langevin equation, the force on the particle is divided into three components (see the equation below).

The first component is the interatomic force, *Fi{xi(t)}*, due to the interaction between the atoms in the system, for example, between two atoms of a protein. This is the same force used in Newton's equation of motion.

The second component is a frictional force which describes the drag on the particle due to the solvent. The magnitude of the drag is related to the friction coefficient, γi.

The third component, *Ri(t)*, is the random or stochastic force due to thermal fluctuations of the solvent. The solvent is not explicitly represented but its effects on the explicit atoms comes from the frictional and random forces. When the frictional and random forces are zero, the Langevin equation reduces to Newton's equation of motion.

$$m_i a = F_i\{x_i(t)\} - \gamma_i v_i m_i + R_i(t)$$

The Langevin equation generates classical Brownian dynamics which describes the motion of a particle under the influence of random collisions with the surrounding solvent.

Monte Carlo (MC) Simulation

Instead of evaluating forces to determine incremental atomic motions, Monte Carlo simulation simply imposes relatively large motions on the system and determines whether or not the altered structure is energetically feasible at the temperature simulated. The system jumps abruptly from conformation to conformation, rather than evolving smoothly through time. It can traverse barriers without feeling them; all that matters is the relative energy of the conformations before and after the jump. Because MC simulation samples conformation space without a true `time' variable or a realistic dynamics trajectory, it cannot provide time-dependent quantities. However, it may be much better than MD in estimating average thermodynamic properties for which the sampling of many system configurations is important.

Monte Carlo Algorithm

1. Generate initial structure R. Calculate V(R).

2. Modify structure to R' . Calculate V'(R').

3. Check: Is V' < V then accept R'

Else
Generate a random number RAND
if exp (-(V'-V)/kT) > RAND then accept R' and go to step 2
Else
Generate new R' randomly and go to step 2
Endif

4. Stop after generating a large number of Rs.

For simple systems, the structural modifications are often tuned so that about 50% of the conformations are accepted. For macromolecular systems, this acceptance ratio can be much smaller, e.g. when dihedral angles are modified by large amounts. It is then generally expedient to bias the random moves in favor of known structural preferences such as side chain rotamers (`biased probability Monte Carlo') and to do some energy minimization before evaluating the energy . Also, explicit water molecules hinder the acceptance of new conformations. So Monte Carlo simulations of macro-molecules generally use an implicit model of solvation, e.g., a term in the empirical potential energy function that mimics the effects of water. See, for example, reference.