- **De novo prediction / Ab initio**

  - Secondary structure prediction; conformation of short fragments (Rosetta); molecular dynamics; Monte Carlo; quantum mechanics (unfeasible)

  - Tough computation

- **Fold recognition**

  - Try to fit with known folds

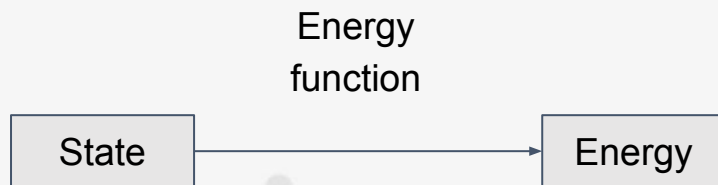 - The fold space is not completely known (50% success)

- **Homology modelling**

  - Similar sequences have similar structures (+50% sequence identity)

  - 40% of genes are not homologous to known structures

Energy functions

A mathematical algorithm that inputs the state of a system and outputs the energy

Energy
function

State → Energy

System state may include
- Coordinates
- Bonded structure
- External elements
- Empirical parameters
- Charge, spin

Energy function may include
- Logic
- Arithmetic
- Algebra
- Calculus
- Numerical procedures

Energy function can be
- Simple or complex
- Analytic or numerical
- QM or MM or other
- Ab initio or empirical
- Atomistic of Coarse-grained

Molecular Mechanics (MM)

A set of models which use an **empirical** (lot of parameters), **algebraic** (formulas), **atomistic** (individual atoms) energy function for chemical systems

$$E_{total} = E_{bonded} + E_{non-bonded}$$

$$E_{bonded} = E_{bonds} + E_{angles} + E_{torsions}$$

$$E_{non-bonded} = E_{elst} + E_{vdw}$$

Force field models

- Atoms: 3D point particles

- Structure: bonds, angles, torsions

Molecular Mechanics (MM)

Force fields vary in

- Energy function form
- Empirical parameters
- Simulation targets

Limitations

- Up to 1M of atoms
- Up to 10^{15} configurations
- Accuracy of energy
- Scope of parameters

Examples

- **AMBER**, assisted model building energy refinement (orginally for DNA)

- **CHARM**, chemistry at harvard MM (Martin Karplus nobel 2013)

- **OPLS**, optimized potential for liquid simulations (small molecules)

- **GROMOS**, used by Gromacs software

- **MMFT**, Merk Molecular force field (pharmaceutical)

Force field parameters

An arbitrary constant whose value characterize an element of a system

$$E_{total} = \sum_{bonds} K_b(r - r_{eq})^2 + ...$$

$E \rightarrow$ [kcal/mol] $r \rightarrow$ [Å]

$R_{eq} \rightarrow$ [Å] $K_b \rightarrow$ [kcal/mol*Å²]

Sources
- Structures → crystal, density
- Spectra → IR, NMR
- Simulation → PES (potential energy surface, quantum mechanics)

Desired properties
- General
- Transferable
- Accurate

Example

O-H in H_2O in AMBER95

R_{eq} = 0.960 Å k_b = 553.0 kcal/mol x Å²

Bond stretch terms

$$E_{bonds} = \sum_{bonds} K_b(r - r_{eq})^2$$

- Parameters
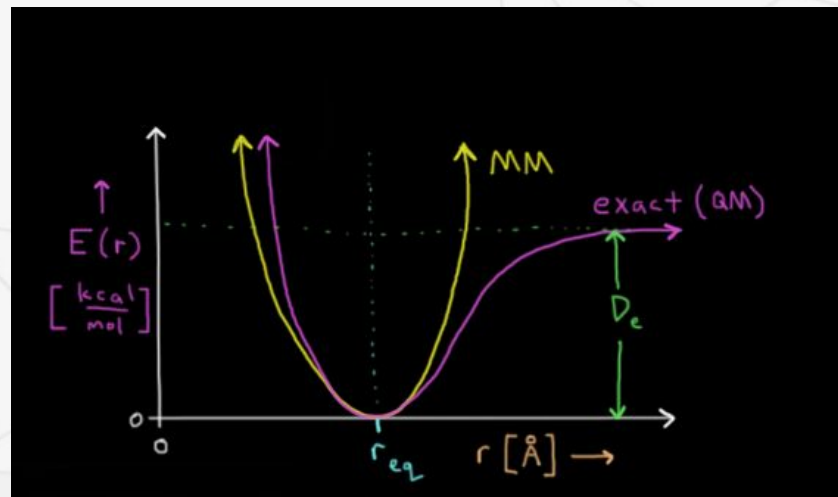
  $r_{eq} \rightarrow$ equilibrium bond length

  $K_b \rightarrow$ bond spring constant

- Variable

  $r \rightarrow$ bond length

- $D_e \rightarrow$ dissociation energy
- Typical $K_b \approx 300\text{-}600$ kcal/mol*Å^2

Angle bend terms

$$E_{angles} = \sum_{angles} K_a(\theta - \theta_{eq})^2$$
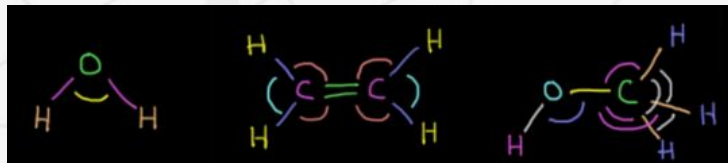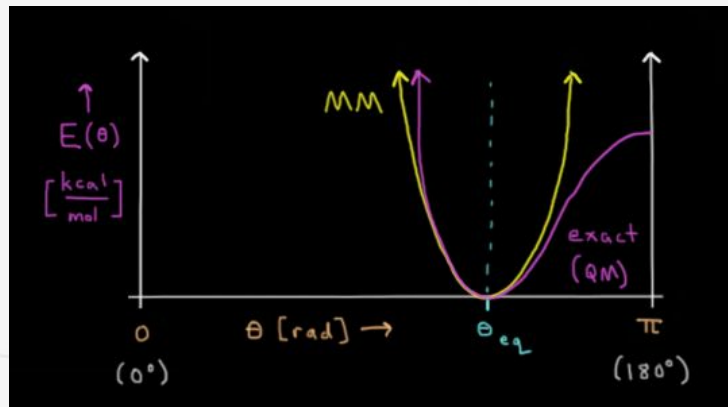
Parameters

$\theta_{eq} \rightarrow$ equilibrium bond angle [rad]

$\theta_b \rightarrow$ angle spring constant

Variable

$\theta \rightarrow$ bond angle [rad], $0 \leq \theta \leq \pi$

Typical $K_a \approx$ 100-150 kcal/mol*rad^2

Van der Waals terms

$$E_{vdw} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \varepsilon_{ij} \left[\left(\frac{r_0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_0}{r_{ij}} \right)^6 \right]$$
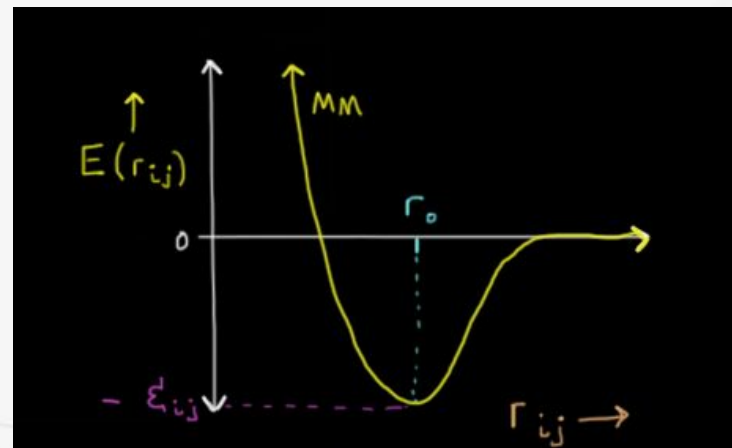
- Parameters

  $\varepsilon_{ij} \rightarrow$ interaction strength [kcal/mol]

  $r_0 \rightarrow$ van der Waals radius (Å)

- Variable

  $r_{ij} \rightarrow$ interatomic distance (Å)

- Short range → exchange / steric repulsion (12)
- Long range → London dispersion attraction (6)

How to define parameters for all possible pairs? → "Combining rules"

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$$ "geometric mean"

$R_0 = \frac{1}{2}(r_i + r_j)$ "arithmetic mean"

$\varepsilon_i, \varepsilon_j \rightarrow$ atomic parameters

Electrostatics terms

$$E_{elst} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$
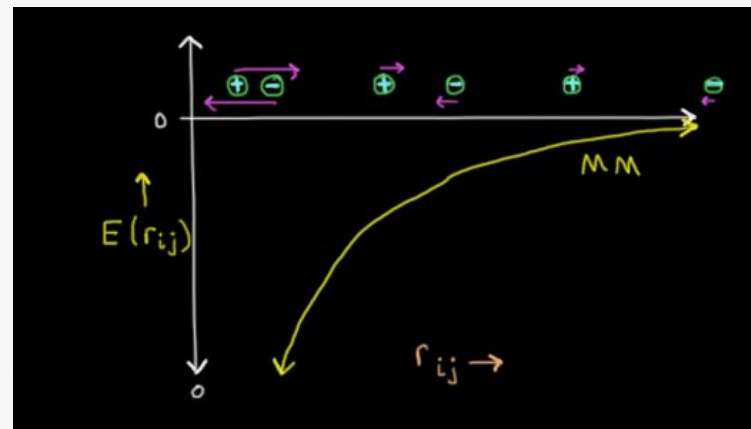

- Parameters

  q_i, $q_j \rightarrow$ atomic partial charges (e)

- Variable

  $r_{ij} \rightarrow$ interatomic distance (Å)

- Same charge → repulsion
- Opposite charge → attraction
- Q_{tot} usually is zero

vdw $\propto \left(\frac{1}{r}\right)^6$ → quick decay

elst $\propto \left(\frac{1}{r}\right)^1$ → slow decay

Elst is rate limiting step of MM

- O(N) bonded terms
- $O(N^2)$ non-bonded terms

Energy gradient
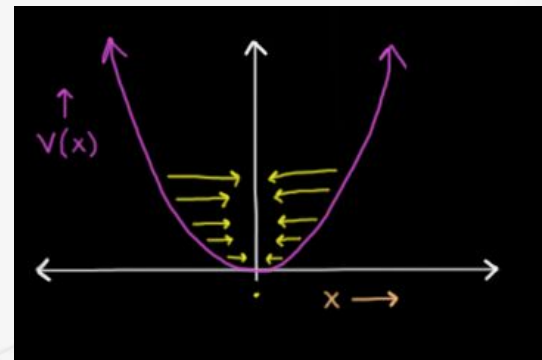
$$V(x) \rightarrow \text{potential energy} \quad (\text{example: } \tfrac{1}{2} kx^2)$$

$$F(x) = -\frac{dV(x)}{dx} \rightarrow \text{force} \quad (\text{example: } -kx)$$

In 3D

$$V(x) \rightarrow V(x,y,z)$$

$$-\frac{dV(x)}{dx} \rightarrow -\left(\frac{\partial V}{\partial x}\hat{x} + \frac{\partial V}{\partial y}\hat{y} + \frac{\partial V}{\partial z}\hat{z} \right)$$

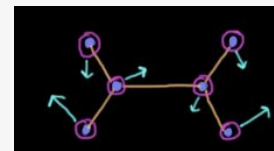"Gradient" $-\nabla V(x,y,z)$

N atoms → 3N coordinates

$$V = V(x_1, y_1, z_1, x_2, y_2, z_2, ..., x_N, y_N, z_N)$$

$$= V(\bar{r}_1, \bar{r}_1, ..., \bar{r}_N) \qquad \bar{r}_1 = (x_i, y_i, z_i)$$

$$= V(\bar{r}^{3N})$$

$$V(\bar{r}^{3N}) \rightarrow \text{energy}$$

$$-\nabla V(\bar{r}^{3N}) \rightarrow \text{force}$$

Energy minimization

$$\frac{dV(x)}{dx} = 0 \qquad \rightarrow \text{ "stationary point"}$$

$$\frac{d^2V(x)}{dx^2} \begin{cases} > 0 \text{ local minimum} \\ < 0 \text{ local maximum} \\ = 0 \text{ unknown ("saddle" point)} \end{cases}$$

1D $\rightarrow$ **3D**

$$\frac{dV(x)}{dx} = 0 \qquad \left| \nabla V(\bar{r}^{3N}) \right| = 0$$

$$\frac{d^2V(x)}{dx^2} > 0 \qquad \{\lambda_H\} \geq 0, \ H_{ij} = \frac{\partial^2 V(\bar{r}^{3N})}{\partial q_i \partial q_j}$$

- Lowest $V(\bar{r}^{3N})$ among local minima $\rightarrow$ global minimum

- "Geometry Optimization" $\rightarrow$ algorithm to find $\{\bar{r}^{3N}\}$ with minimum P.E.

$H \rightarrow$ 3N x 3N real, symmetric "Hessian matrix"

Steepest descent

Method for obtaining local stationary points of a multi-dimensional function

Simple method for geometry optimization

1. Guess X_0 "initial geometry"
2. Compute $V(x_i)$ and $\nabla V(x_i)$ energy and gradient
3. Update $x_{i+1} = x_i - \gamma_i \nabla V(x_i)$ displacement
4. Check for convergence
   a. If yes: done $x_f = x_i$
   b. If no: return to step 2

Various methods for choosing γ_n
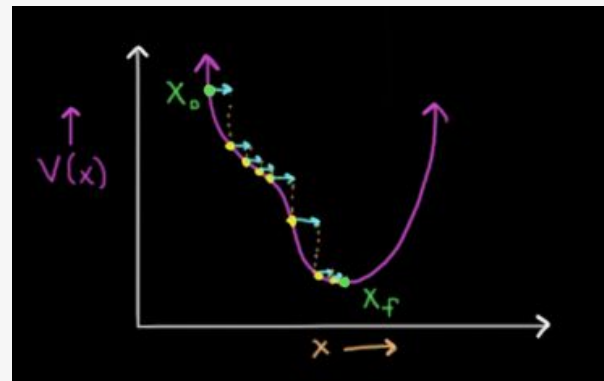
Convergence checks

- $\|\nabla V(x_i)\| \approx 0$ gradient
- $|V(x_i)-V(x_{i-1})| \approx 0$ energy
- $\|x_i - x_{i-1}\| \approx 0$ displacement

Molecular dynamics (MD)

Calculate system properties by propagating through time and averaging

- Classical Mechanics → Newton's Laws (F=ma)

- MM → conservative (total energy is conserved, kinetic + potential energy), time-independent

$$F = ma = m\frac{d^2(x)}{dt^2} = \frac{d}{dt}\left(m\frac{d(x)}{dt}\right) = \frac{d}{dt}(mv) = \frac{dp}{dt}$$

$$F = -\frac{dV(x)}{dx} \qquad p = mv = m\frac{dx(t)}{dt}$$

$$\frac{\partial p_{x_i}}{\partial t} = -\frac{\partial V}{\partial x_i}$$

Solve for all 3N coordinates

Can't solve exactly → approximate numerically

Molecular dynamics

1. Assign initial x_0, q_0, $t = 0$

2. Compute $V(x_i)$ and $\nabla V(x_i)$

3. Update $x_{i+1} = x_i + 1/m\ p_i \Delta t$

4. Update $p_{i+1} = p_i - \nabla V(x_i) \Delta t$

   $\left.\right\}$ $t = t + \Delta t$

5. Repeat (steps 2-4) until $t \geq t_f$

6. Compute desidered properties

$\Delta t \rightarrow$ timestep (1fs = 10^{-15}s)

$T \rightarrow$ depends on goal

- Vibrations $\rightarrow$ 10 - 1,000 fs

- Conformation changes $\rightarrow$ ps - ns

- Diffusion / interaction $\rightarrow$ ns - μs

- Protein folding $\rightarrow$ μs - s

References & Links

TMP Chem (Trent Parker's YouTube channel)

https://www.youtube.com/user/TMPChem

Computational Chemistry

- 2.1, 2.2, 2.3
- 2.5, 2.6
- 2.9, 2.10
- 3.1, 3.2, 3.3
- 3.8