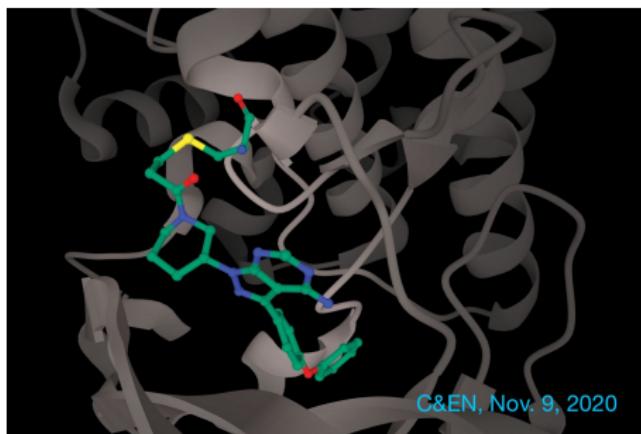


Computational chemistry methods in drug design: Quantum chemical approaches



CCNSB, IIITH, Shampa Raghunathan
MC-610, NIPER Hyderabad: Lectures 1 & 2 (24 & 25 May 2021)
Potential energy surface and minimization

Welcome to MC-610

Objective: To introduce theory and applications of drug design by computational methods (quantum mechanics)

- Mechanism of action of a chemical process
 - ▶ Potential energy surface and minimization
 - ▶ Electronic structure theories
 - Wavefunction based methods
 - Semi-empirical methods
 - Density functional theory
 - ▶ Applications (QM methods in practice)

Optional background reading

Jensen, F. (2017). Introduction to Computational Chemistry. John wiley & Sons.

Cramer, C. J. (2013). Essentials of Computational Chemistry: Theories and Models. John Wiley & Sons.

Leach, A. R. (2001). Molecular Modelling: Principles and Applications. Pearson Education.

Specific references will be provided along with each lecture

Why do we need modelling and simulations?



Jiggling and Wiggling Feynman Lectures on Physics

Certainly no subject or field is making more progress on so many fronts at the present moment than biology, and if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms.

—Richard P. Feynman

Why do we need modelling and simulations?

SPINACH ON THE CEILING: A Theoretical Chemist's Return to Biology

Annual Review of Biophysics and Biomolecular Structure

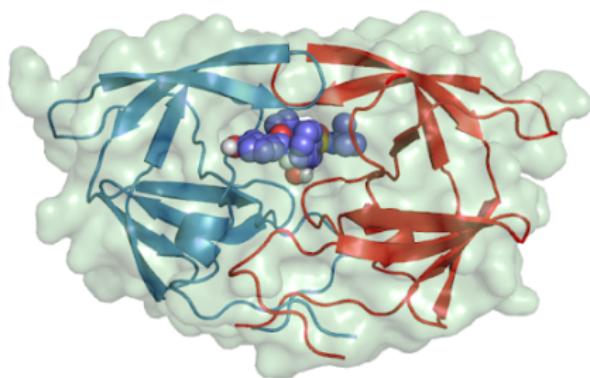


“simulations provide the ultimate detail concerning individual particle motions as a function of time” –Martin Karplus

Why do we need modelling and simulations?

Useful complement, because it can :

- ▶ Explain experiment
- ▶ Aid experiment; in-silico drug discovery, protein engineering
- ▶ Find structure and function relationship at molecular level



de novo design, random screening and combination of these approaches have accelerated the drug discovery process. e.g., HIV protease inhibitors approved by the FDA: saquinavir, ritonavir, fosamprenavir, atazanavir, tipranavir, and darunavir etc.

Illustration:

<http://sbcb.bioch.ox.ac.uk/users/greg/teaching/docking-2012.html>

Definitions: modelling and simulation

A **model** is an idealization of real behavior, i.e., an approximate description based on empirical and/or physical reasoning.

A **simulation** is a study of the dynamical response of a modeled system found by subjecting **models** to inputs and constraints that simulate real events.

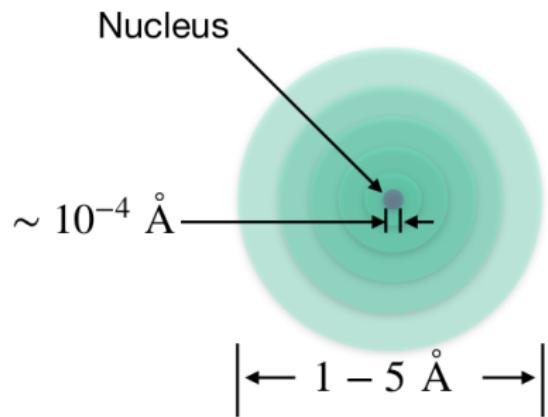
A **simulation** does not mimic reality, rather it mimics a **model** of reality.

Can one separate **simulations** from the underlying **model**? Accuracy & validity?

LeSar, R., Iowa State University

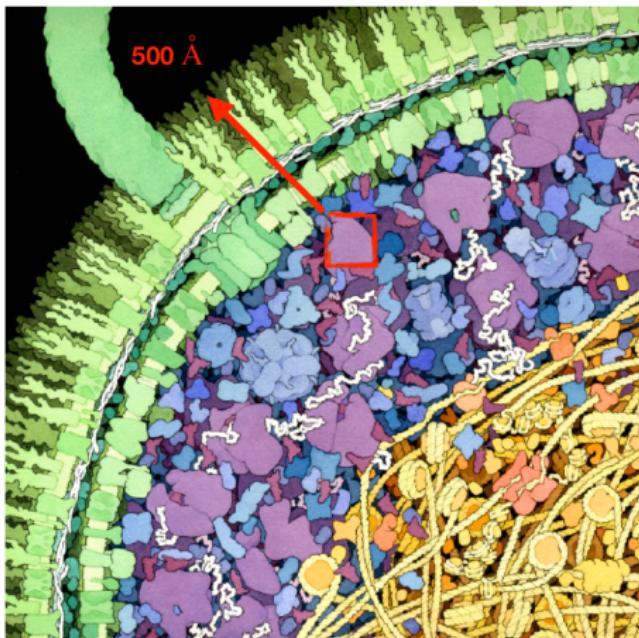
What is an atom?

Protons and neutrons make up the heavy, positive core, the **NUCLEUS** (occupies a small volume of the atom) orbited by electrons.



Length- & time-scales appropriately modelled?

Inconclusive experiments



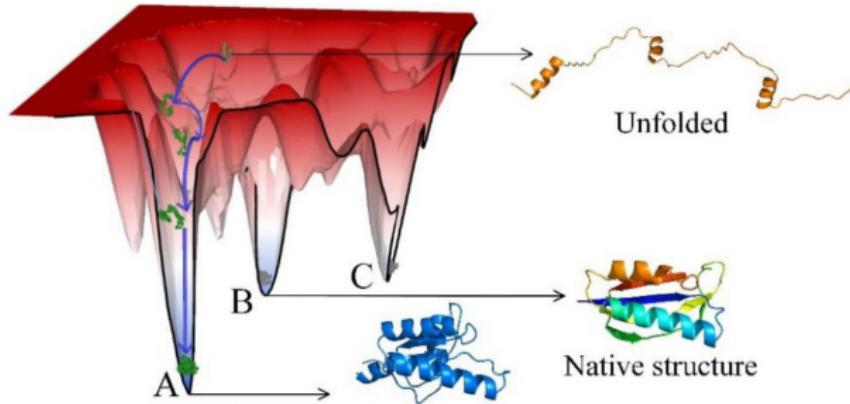
Extreme simplification

Limited model accuracy

Large gaps in time-scales

Illustration: Nilsson, L., Karolinska Institute, Stockholm.

A protein folding landscape



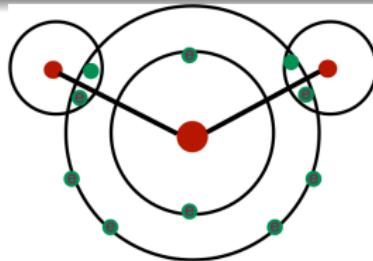
- ▶ Large systems
- ▶ Chemical processes

Illustration: Zhao, K.-L. et al., bioRxiv.

Quantum mechanics vs. Molecular mechanics

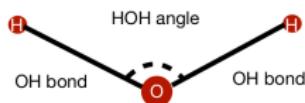
Quantum mechanics (QM)

- ▶ Fundamental entity: electrons
- ▶ Building blocks: atomic orbital basis
- ▶ Approx.: N-electron problem converted to N no. of 1-electron problems



Molecular mechanics (MM)

- ▶ Fundamental entity: atoms
- ▶ Analytic potentials
- ▶ Approx.: Potentials are fitted to exp. and/or QM data

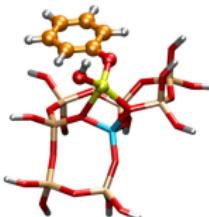


Choosing a computational tool

Quantum Mechanics

Electronic structure, (Schrödinger)

- ▶ More accurate
- ▶ More expensive, small systems

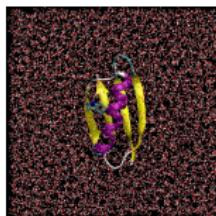


10–100 atoms
10–100 ps

Classical Molecular Mechanics

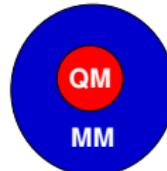
Empirical forces, (Newton)

- ▶ Less accurate
- ▶ Less expensive, large systems



10^4 – 10^5 atoms
10–100 ns

Combined QM/MM



10^4 – 10^5 atoms
10–100 ps

Aim to simulate processes

Polyptide folding
Biomolecular associations
Membrane transportations
Solvent partitioning



thermodynamics, weak
nonbonded interactions
Classical MD
atomic degrees of free-
dom (system+solute)

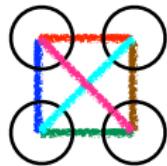
Chemical reactions
Catalytic processes
Photochemical reactions



Strong bonded forces
Quantum MD
electronic, nuclear de-
grees of freedom

Model potential

1 2
3 4



$$V(\mathbf{r}) = \sum_i v_1(\mathbf{r}_i) + \sum_i \sum_{j>i} v_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_i \sum_{j>i} \sum_{k>j>i} v_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots$$

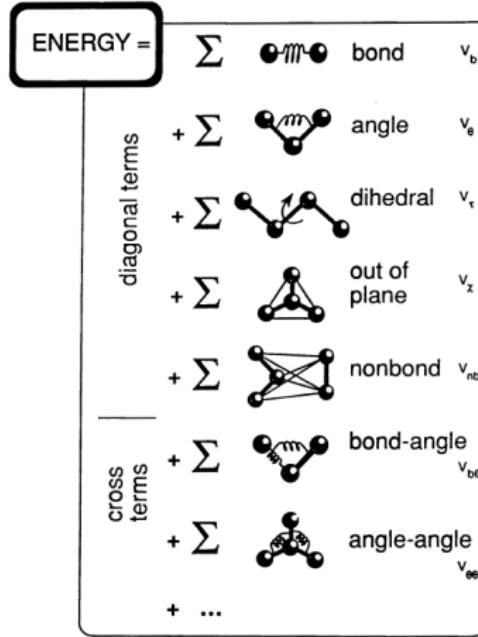


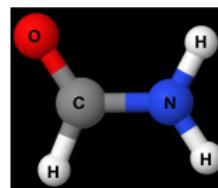
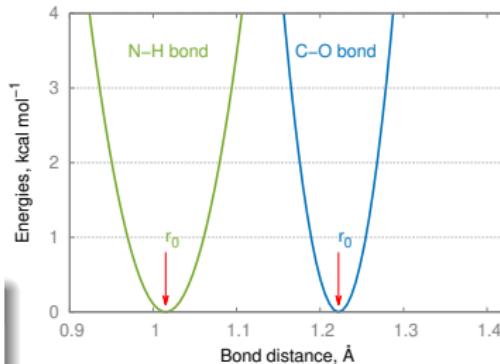
Figure 1 Schematic of molecular force field expression. Diagonal terms refer to interactions that can be expressed as a function of a single internal coordinate, whereas cross terms introduce coupled interactions involving two or more coordinates.

Illustration: Ponder, J., Washington University.

Bonded interactions

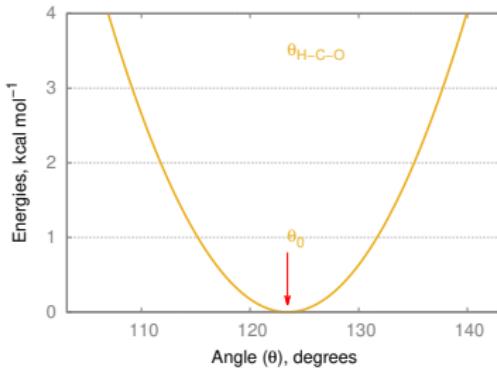
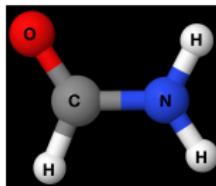
Bond

$$v_b = \frac{k_b}{2}(r - r_0)^2$$

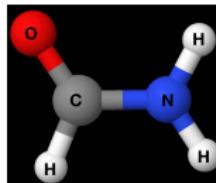


Angles

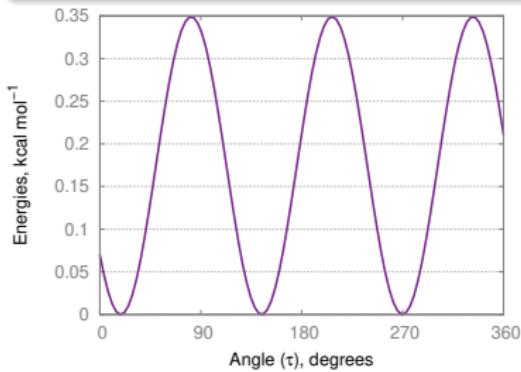
$$v_\theta = \frac{k_\theta}{2}(\theta - \theta_0)^2$$



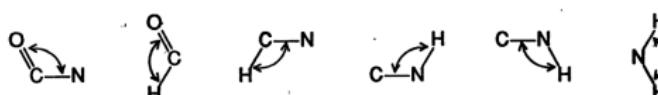
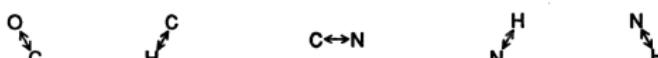
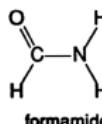
Torsions



$$\nu_{\tau} = \frac{k_{\tau}}{2}(1 + \cos(n\tau - \gamma))$$



Counting the no. of bonded interactions



Nonbonded interactions

$$V_{ij} = \underbrace{\frac{q_i q_j}{r_{ij}}}_{\text{Coulomb}} + 4\epsilon \underbrace{\left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right]}_{\text{Lennard-Jones potential}}$$

vdW interactions, short-range

ϵ : well-depth, σ : collision diameter, r_{\min} : distance at minimum

$$r_{\min} = 2^{1/6} \sigma$$

$$r_{\min,ij} = r_{\min,i} + r_{\min,j}, \quad \epsilon = \sqrt{\epsilon_i \epsilon_j}$$



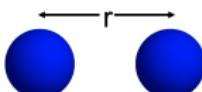
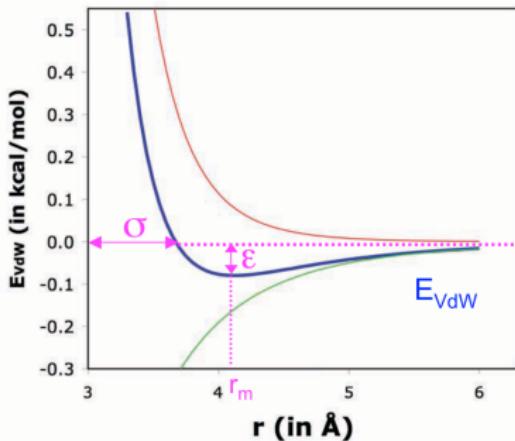
$$\epsilon \left[\left(\frac{r_{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{\min}}{r_{ij}} \right)^6 \right]$$



Repulsive



Attractive

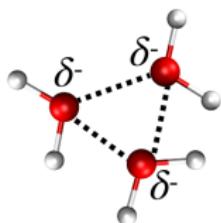


Electrostatic interactions, long-range

$$V_{\text{elec}} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

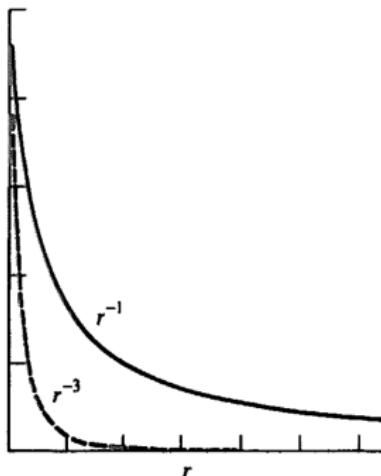
ϵ_0 dielectric constant:

1	for vacuum
4–20	for protein core
80	for water



The Coulomb energy decrease as $1/r$; **long-range interactions**

Special techniques are available to deal with the long-range electrostatic interactions as well as vdW interactions.



A typical molecular mechanics force field

“STERIC” Energy

$$\begin{aligned} V(\mathbf{r}^N) = & \sum_{\text{bonds}} \frac{k_{b_i}}{2} (r_i - r_{i,0})^2 + \sum_{\text{angles}} \frac{k_{\theta_i}}{2} (\theta_i - \theta_{i,0})^2 \\ & + \sum_{\text{torsions}} \sum \frac{k_{\tau_i,n}}{2} (1 + \cos(n\tau_i - \gamma)) \\ & + \sum_{i=1}^N \sum_{j=i+1}^N \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \right) \end{aligned}$$

Examples of force fields: CHARMM, AMBER, GROMOS

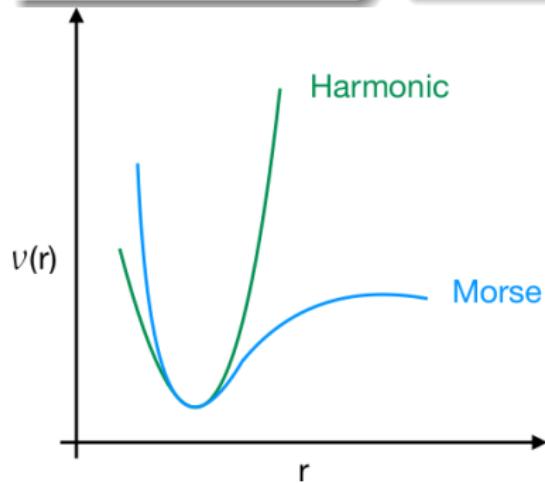
Beyond harmonic approximation

Bond, Harmonic

Morse potential

$$v_b(r) = \frac{k_b}{2}(r - r_0)^2$$

$$v(r) = D_e \{1 - \exp[-a(r - r_0)]\}^2$$



Features of potential energy surface (PES)

Minimum energy path 1: Reactant \rightarrow Transition Structure A \rightarrow Product A
Minimum energy path 2: Reactant \rightarrow Transition Structure B \rightarrow Product B

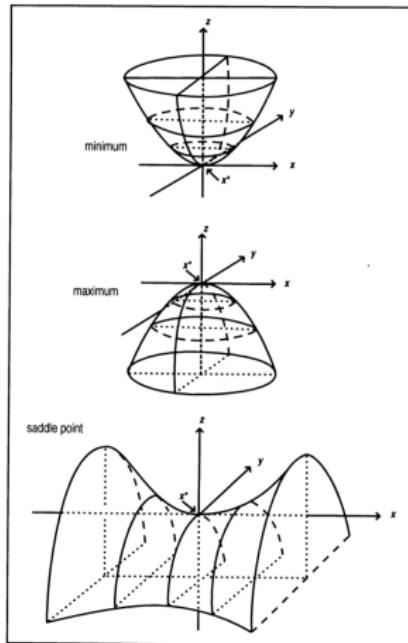
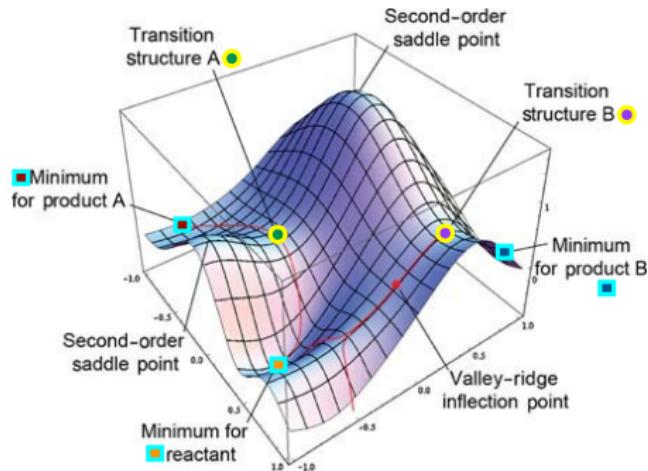
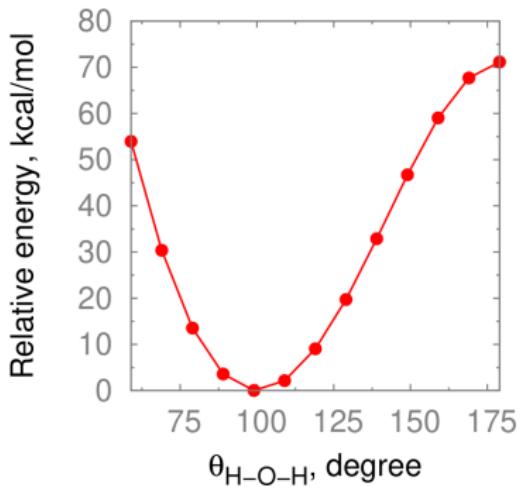
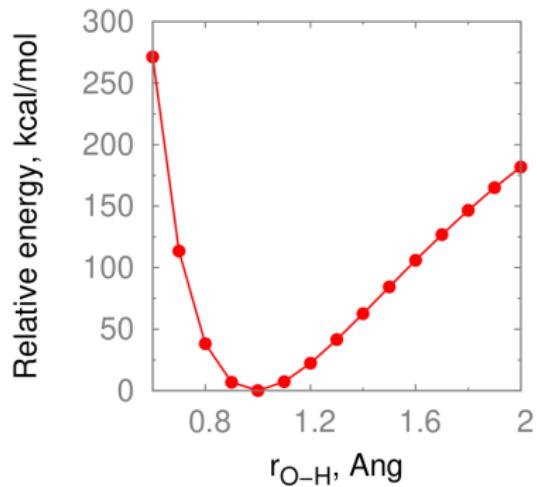
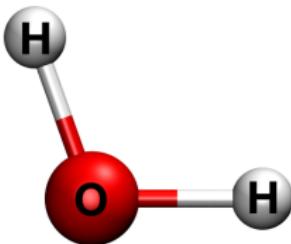


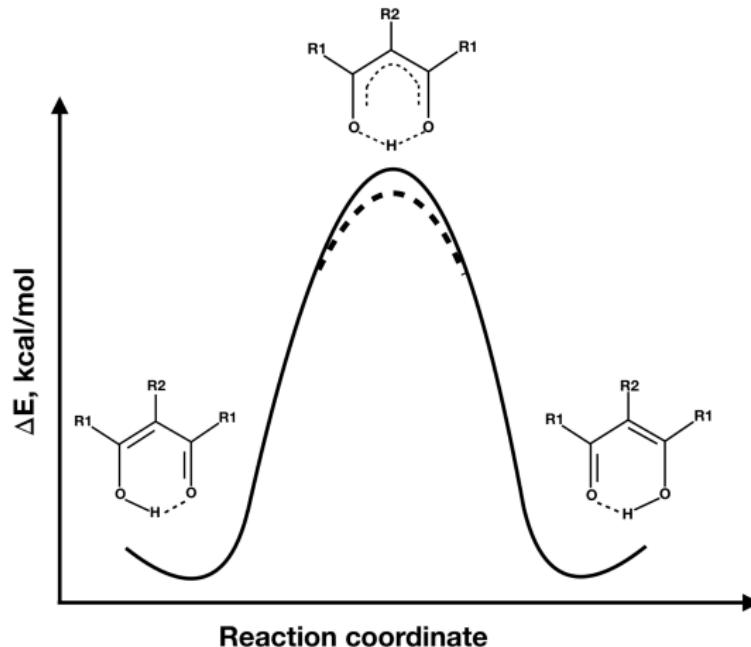
Figure 3 Types of stationary points.

Schlegel, H.B., J.Comput. Chem **24**, 1514 (2003); Schlick T., Rev. Comput. Chem. (1992).

1-dimensional PES of H₂O molecule



Examples of potential energy surfaces for real chemical systems



Hydrogen transfer barrier of 4.1 kcal/mol, calculated at the CCSD(T)/aug-cc-pV5Z level by J. Bowman et al.

Torsional 2-dimensional contour surface of pentane

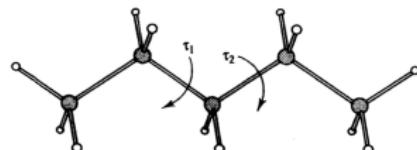
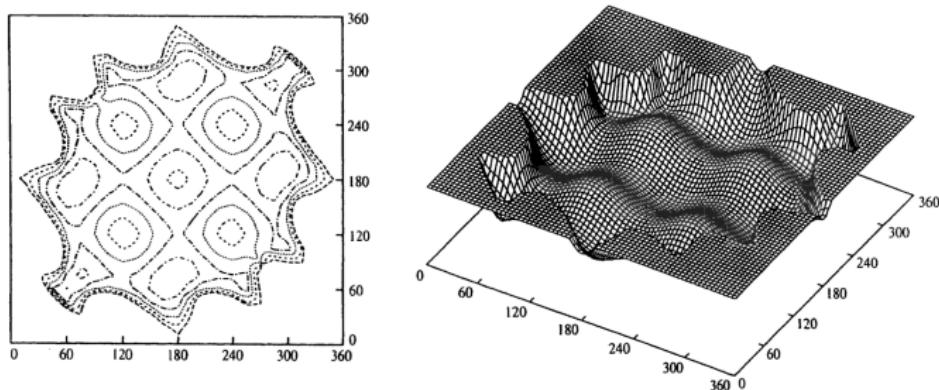


Fig. 5.1: Variation in the energy of pentane with the two torsion angles indicated and represented as a contour diagram and isometric plot. Only the lowest-energy regions are shown.

Leach, A. R. (2001). Molecular Modelling: Principles and Applications. Pearson Education.

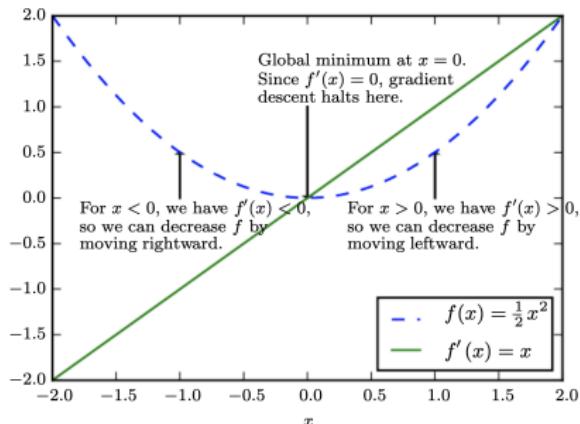
Energy minimization: Statement of the problem

For a given function f which depends on one or more independent variables x_1, x_2, \dots, x_i (Cartesian/internal coordinates), find the values of those variables where f has a minimum value. At a minimum point the first derivative of the function with respect to each of the variables is zero, and the second derivatives are all positive:

$$\frac{\partial f}{\partial x_i} = 0 \text{ and } \frac{\partial^2 f}{\partial x_i^2} > 0$$

- ▶ First-order minimization algorithm
 - Steepest descent (SD)
 - Conjugate gradient (CONJ)
- ▶ Second-order minimization algorithm
 - Newton Raphson

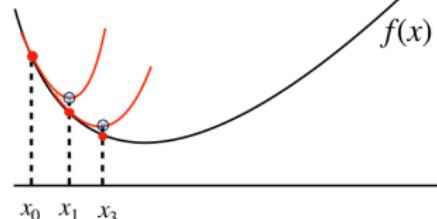
First-order minimization



First derivative of a function $f'(x)$ is used to follow the function downhill to a minimum. The f is decreased by moving in the direction of the negative gradient, in case of the **steepest descent (SD) method.**

Goodfellow, I et al. (2016). Deep Learning. The MIT Press.

Second-order minimization



Second derivative determines the curvature of a function. Moving to a minimum of quadratic fit at each point in case of the **Newton Raphson (NR) method.**

SD	NR
1. Computationally fast	Expensive
2. Good for minimizing initial structures	Unstable far from minimum
3. Does not require initial structure to be near minimum; slow to converge especially near minimum at low gradient values	Quickly converges near minimum

Mapping a complete PES for a chemical reaction

Remember the hypersurface... Lets think about a simple molecule...we have a system of N atoms resulting $3N-6$ variables hypersurface. Do you think we really, need to consider all those variables for a particular reaction to model?

We focus on a single (sometimes two) reaction coordinate, not on full $3N-6$

Finding transition state (TS): For smaller systems

Locating a TS in the PES is more difficult to find than minima.

- ▶ **For extremely small systems:** like, for $\text{H}_2\text{O} + \text{H}^+$, $3N-6$ full dimensional mapping of a PES is possible (limited cases)
- ▶ **For smaller systems:** “Coordinate driving” or “constrained” TS optimization method works well
- ▶ **Choosing a reaction coordinate:** For some cases where, a few of the variables is able to describe the changes occurring between reactant and product. e.g., angle, torsion for conformational changes, and bond involving making and breaking for a chemical reaction

Choose fixed values of the selected **reaction coordinate**, and allow other variables to be optimized. So we are kind of adiabatically mapping the Energy, $E(\text{reaction coordinate})$

Finding TS: “constrained” TS optimization method

- ▶ Reaction coordinate/variable in a “constrained” TS search is good, if it has large coefficient in the Hessian eigenvector with the negative eigenvalue at the TS geometry
- ▶ However, we know of the Hessian of the TS, only after it is found and that was because of the coordinate selection – which is biased
- ▶ If only, one or two variable changes significantly between reactants and products – this method works well

e.g. torsion angle in methyl rotation,

H–C–N angle in HCN to HNC isomerization,

X–C and C–Y distances in SN² reactions, like, X+CH₃Y → Y+CH₃X

More than two reaction coordinates are often used, e.g., double proton transfer process often modelled for RNA base pair modelling

Finding TS: “constrained” TS optimization method

In case of a reaction, if reactants and products are known: Use “constrained” TS optimization method where, one or two geometric parameters are constrained in order to describe the PES of the reaction.

