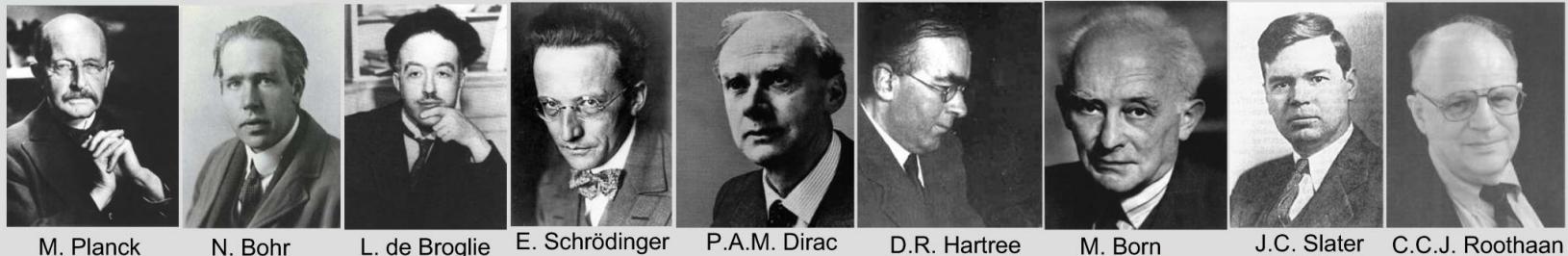


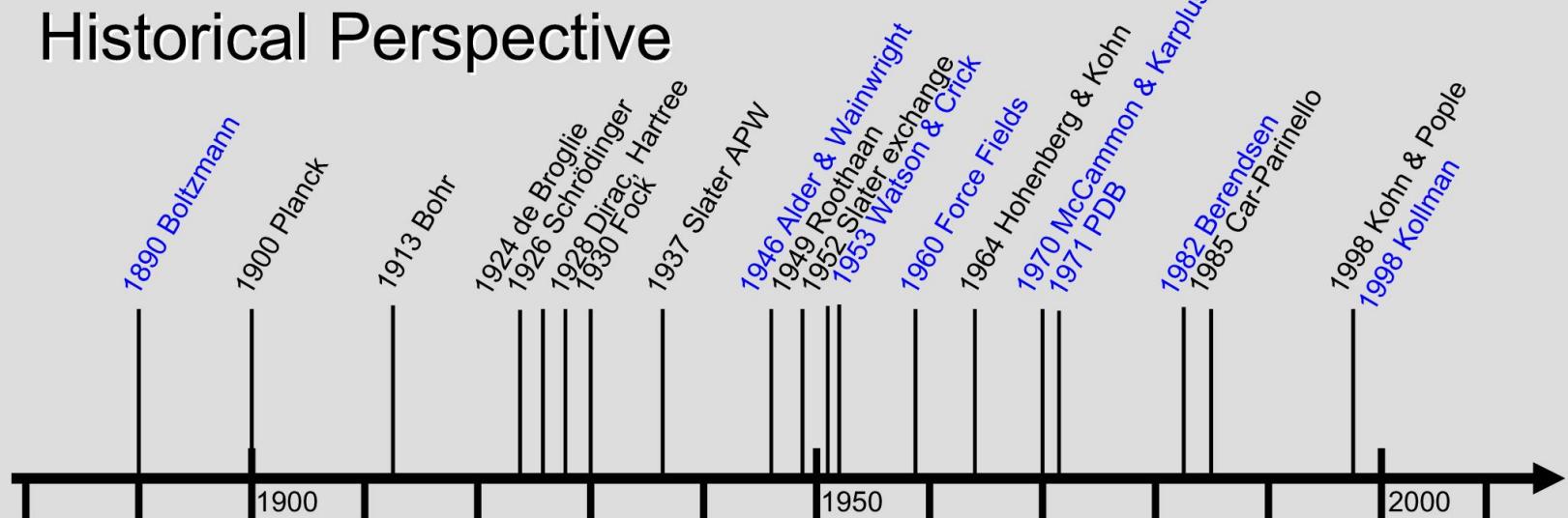
Structural Modeling Basics

Yazdan Asgari
2020

Historical Perspective



Historical Perspective



Wavefunction

Ψ

Each “particle” is represented by a wavefunction Ψ (position, time) such that $\Psi^*\Psi$ = the probability of finding the particle at that position at that time.

The wavefunction is used in the Schrodinger equation. The Schrodinger equation plays the role of Newton’s laws and conservation of energy in classical mechanics - i.e., it predicts the future behavior of a dynamic system. It predicts analytically and precisely the probability of events or outcome. The detailed outcome depends on chance, but given a large number of events, the Schrodinger equation will predict the distribution of results.

Schrodinger Equation

Kinetic Energy + Potential Energy = E

Classical Conservation of Energy
Newton's Laws

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = E$$
$$F = ma = -kx$$

Harmonic oscillator example.

Quantum Conservation of Energy
Schrodinger Equation

In making the transition to a wave equation, physical variables take the form of "operators".

$\frac{p^2}{2m} + \frac{1}{2}kx^2$ The energy becomes the Hamiltonian operator

$p \rightarrow \frac{\hbar}{i} \frac{\partial}{\partial x}$ $x \rightarrow x$ Wavefunction

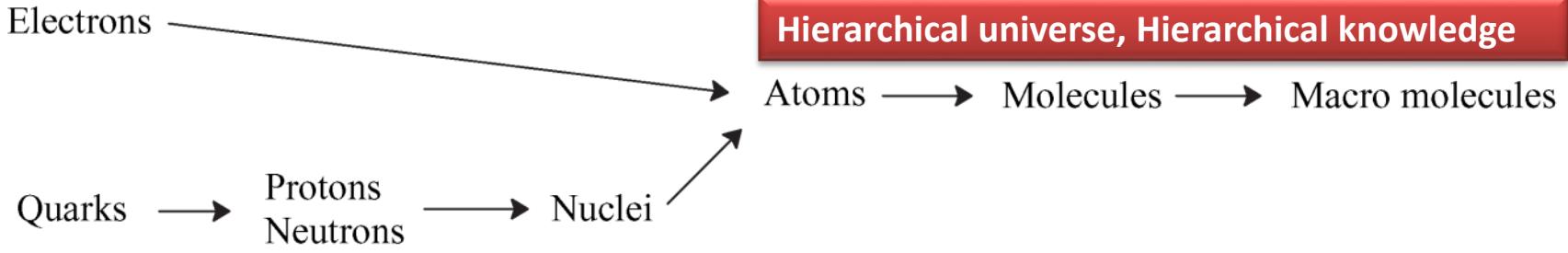
$H \rightarrow -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + \frac{1}{2}kx^2$ Energy "eigenvalue" for the system.

$H\Psi = E\Psi$ The form of the Hamiltonian operator for a quantum harmonic oscillator.



Biomolecular Modeling :

Application of computational chemistry to understand the structure, dynamics, and thermodynamics of biological molecules.



1) System description

What are the fundamental structural units or “particles”, and how many are there?

2) Starting condition

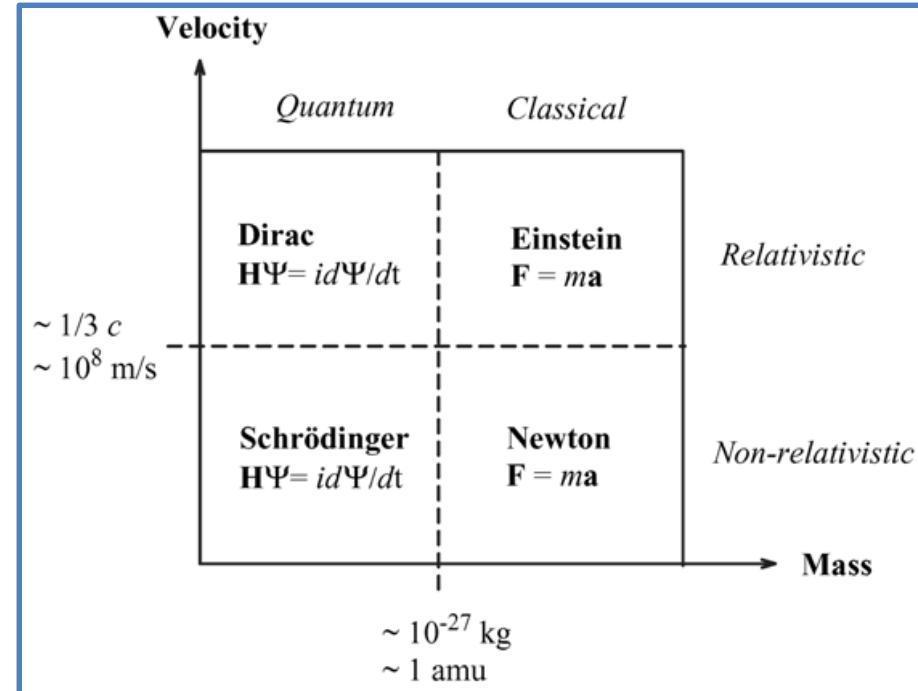
Where are the particles and what are their velocities?

3) Interaction

What are the forces acting between particles?

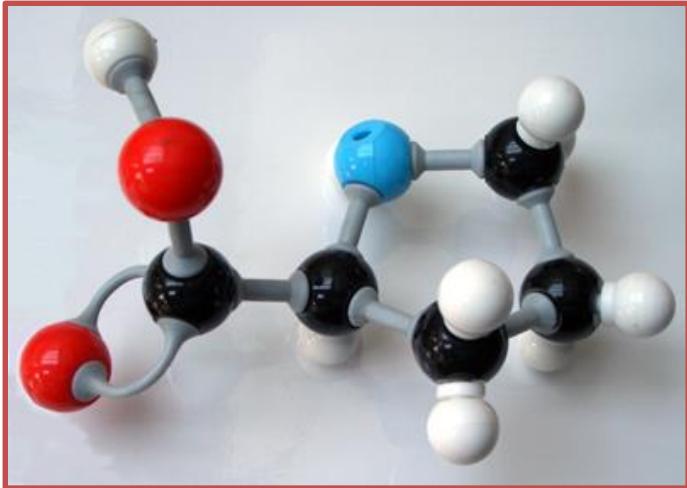
4) Dynamical equation

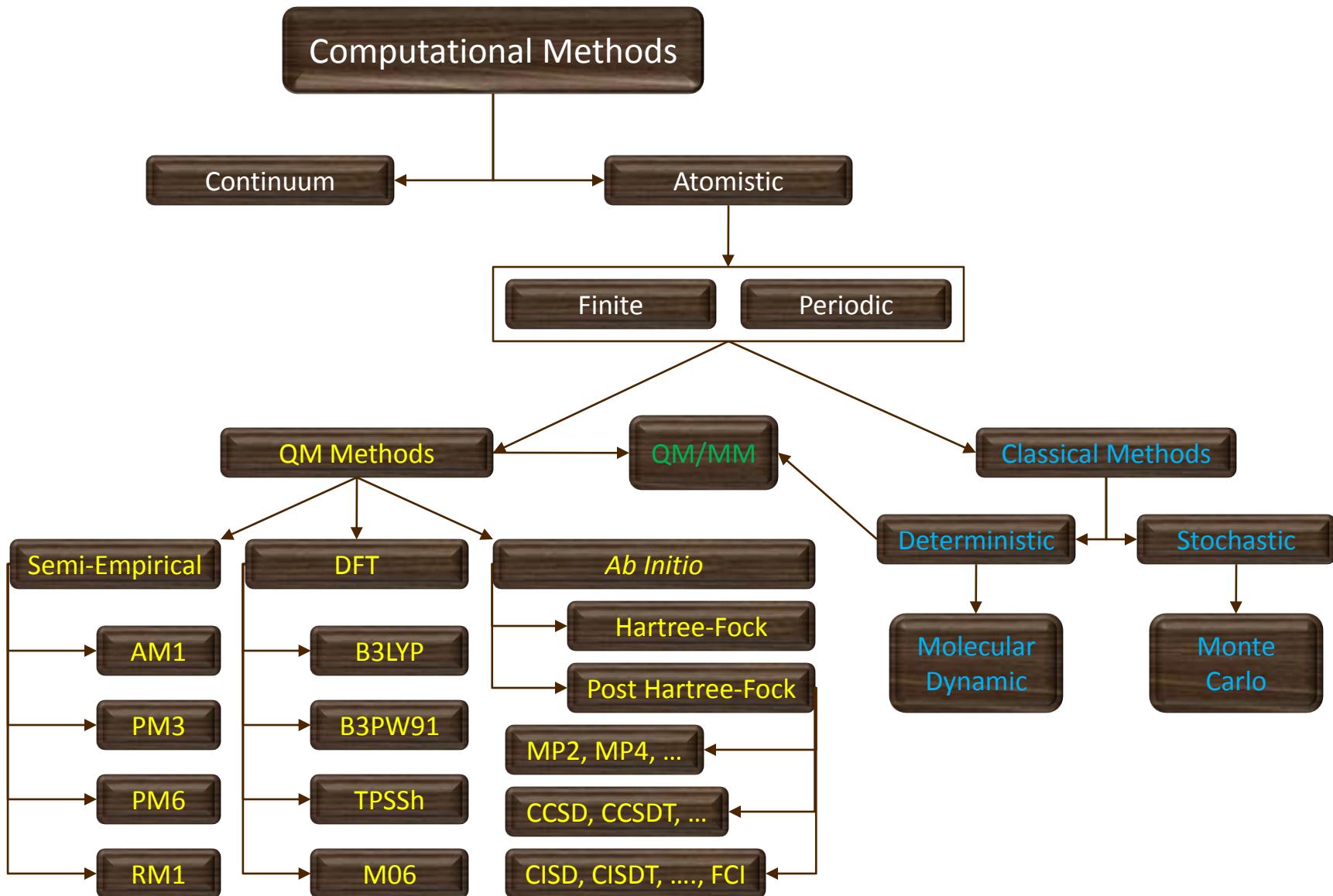
How the system evolves in time?



Plastic molecular models

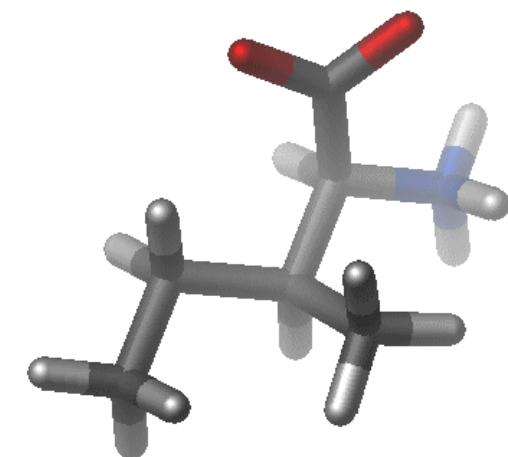
- Assemble from standard parts
- Fixed bond lengths and coordination geometries
- Good enough for qualitative modeling of the structure of some molecules
- Easy and cheap to use
- Provide a good feeling for the 3 dimensional structure of molecules
- No information on properties, energetics or reactivity





Molecular mechanics

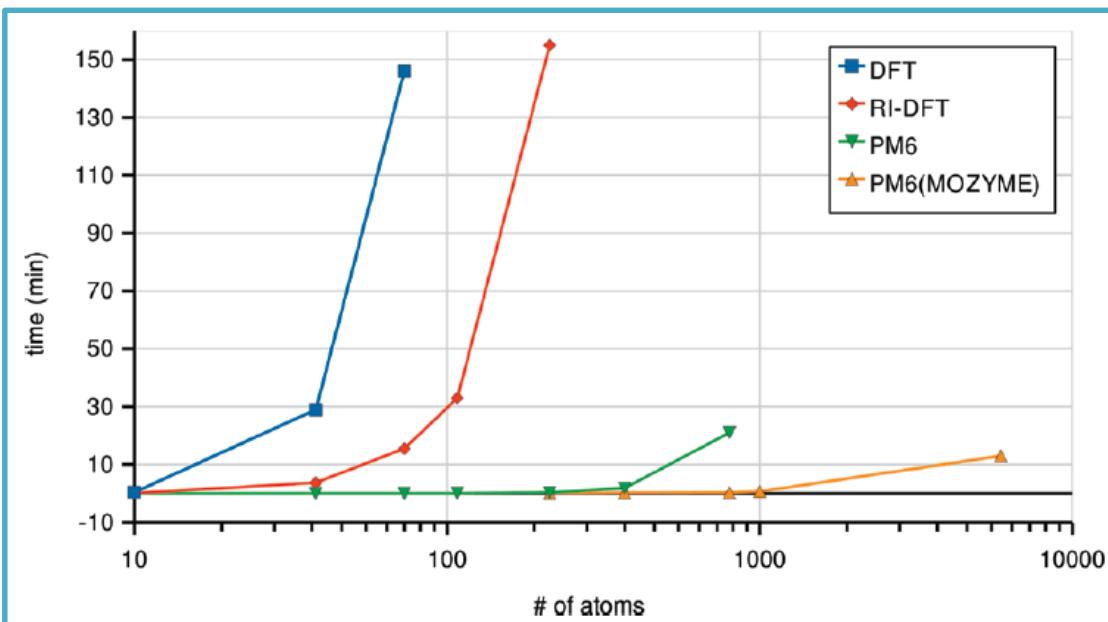
- Ball and spring description of molecules
- Better representation of equilibrium geometries than plastic models
- Able to compute relative strain energies
- Cheap to compute
- Lots of empirical parameters that have to be carefully tested and calibrated
- Limited to equilibrium geometries
- Does not take electronic interactions into account
- No information on properties or reactivity
- Cannot readily handle reactions involving the making and breaking of bonds

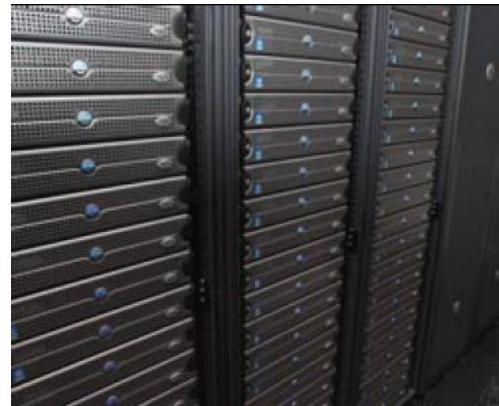
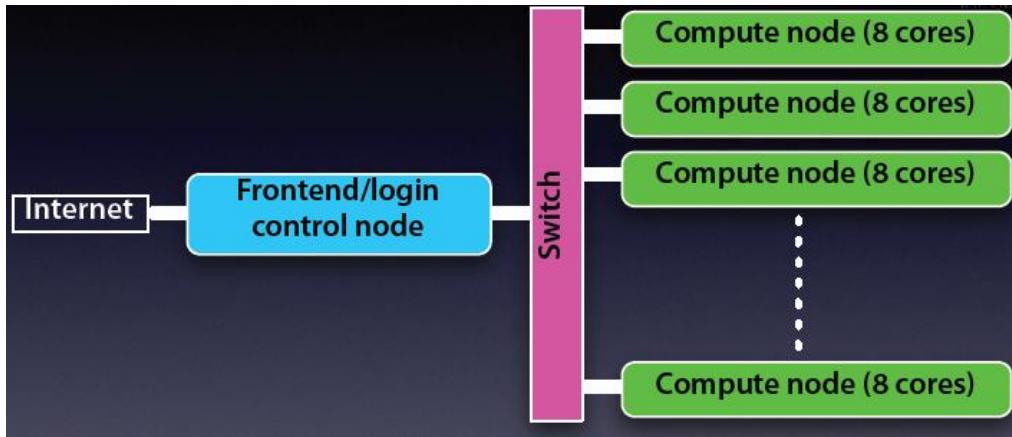


Semi-empirical molecular orbital methods

- ❖ Approximate description of valence electrons
- ❖ Obtained by solving a simplified form of the Schrödinger equation
- ❖ Many integrals approximated using empirical expressions with various parameters
- ❖ Semi-quantitative description of electronic distribution, molecular structure, properties and relative energies
- ❖ Cheaper than *ab initio* electronic structure methods, but not as accurate

Comparison of computational resources required for one SCF cycle on a single protein molecule.





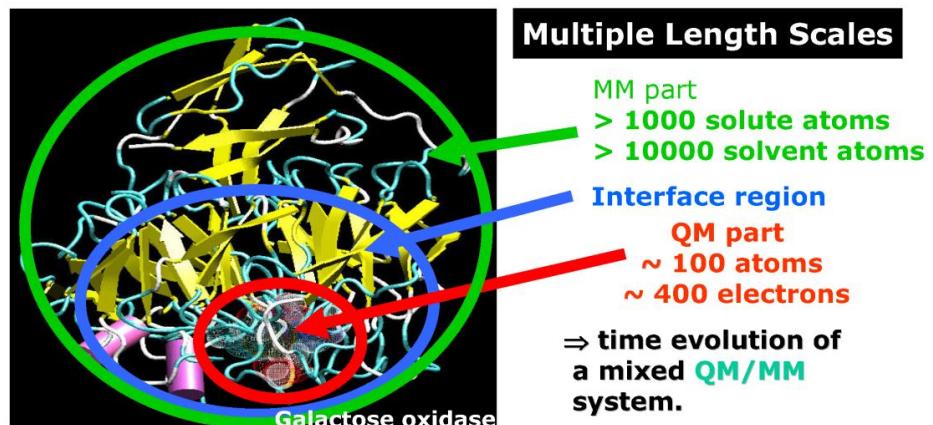
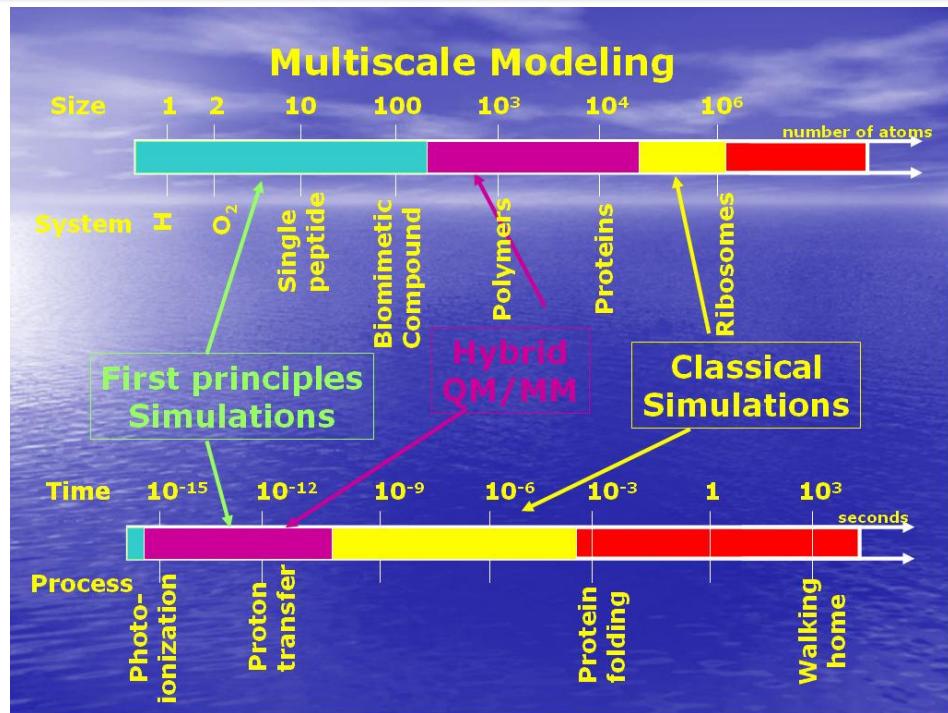
Dealing with different length/size/time scales

Multiscale Modeling:

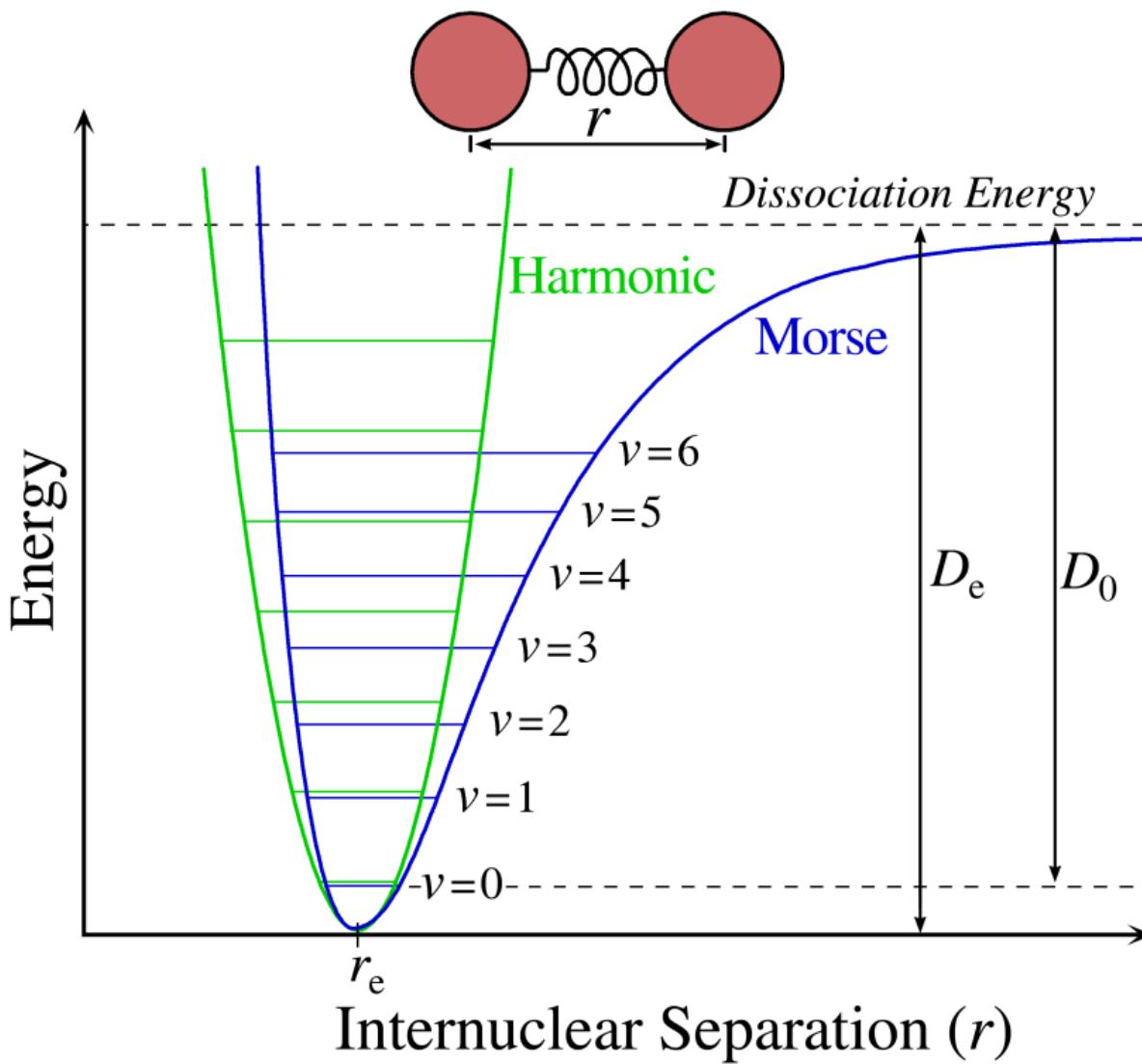
A coherent application of quantum mechanics, molecular dynamics and coarse grained modeling.

Especial methods Targeted on Biomolecular Problems :

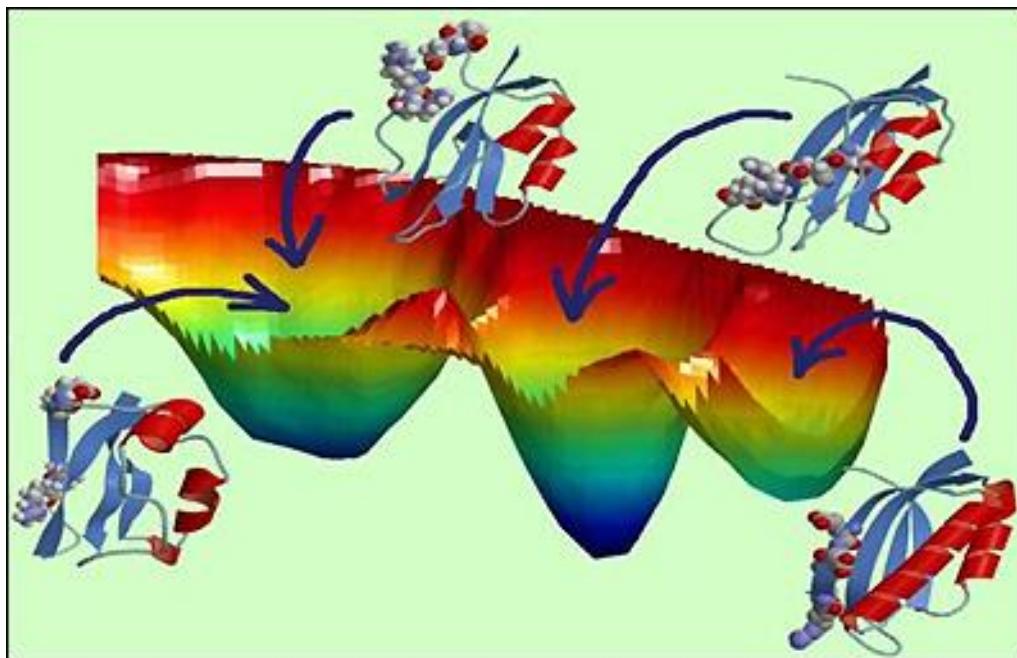
- 1) MD Simulations
- 2) QM/MM methods
- 3) Effective Fragment Potential (EFP)
- 4) Fragment Molecular Orbital (FMO)



Bonds in MM & QM



Exploring The Energy Landscape



Starting Point: Schrödinger Equation

For particles with small masses, primarily electrons, quantum mechanics must be employed.

At low velocities, the relevant equation is the time-dependent Schrödinger equation.

Hamiltonian operator is given as a sum of kinetic and potential energy operators.

Solving the Schrödinger equation gives the wave function as a function of time.

Probability of observing a particle at a position \mathbf{r} and time t is given as the square of the wave function.

Separating space and time variables results in time-independent Schrödinger equation.

$$\mathbf{H}\Psi = i \frac{\partial\Psi}{\partial t}$$

$$\mathbf{H}_{\text{Schrödinger}} = \mathbf{T} + \mathbf{V}$$

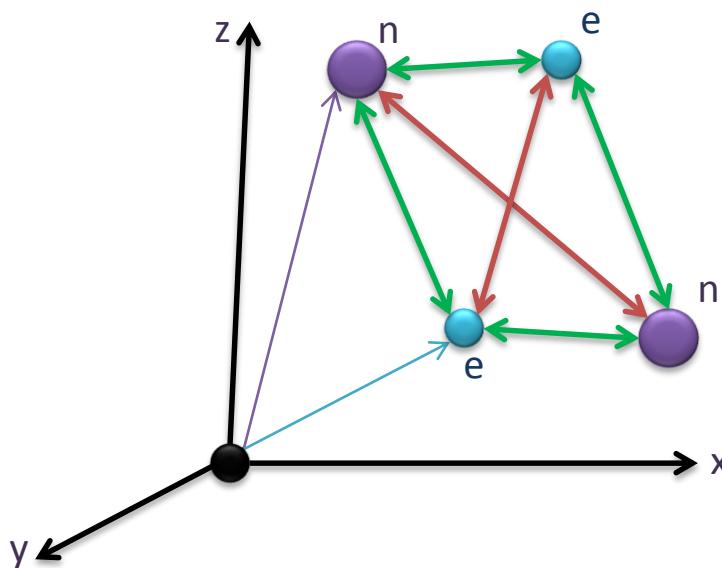
$$\mathbf{T} = \frac{\mathbf{p}^2}{2m} = -\frac{1}{2m}\nabla^2$$

$$P(\mathbf{r}, t) = \Psi^2(\mathbf{r}, t)$$

$$\mathbf{H}(\mathbf{r})\Psi(\mathbf{r}) = E(\mathbf{r})\Psi(\mathbf{r})$$

$$\mathcal{H}|\Phi\rangle = E|\Phi\rangle$$

Separating nuclear and electronic problems



Neglecting the coupling between the nuclear and electronic velocities.
The nuclei are stationary from the electronic point of view.

$$H = T_e + T_n + V_{ee} + V_{ne} + V_{nn}$$

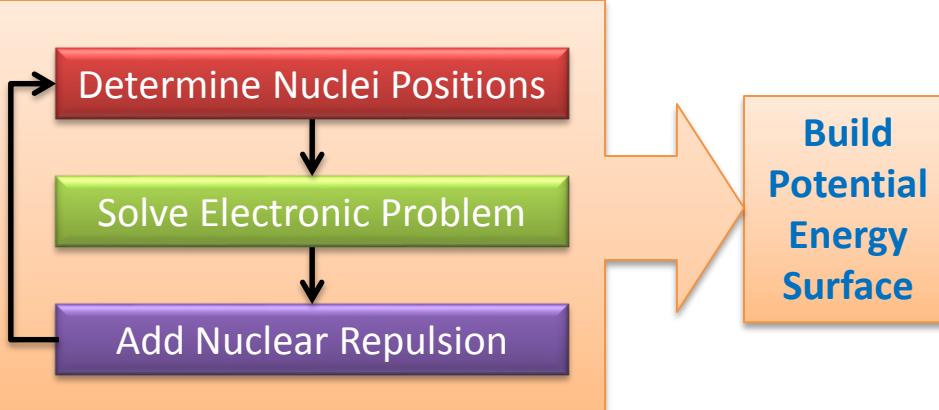
$$H\Psi = E\Psi$$

$$H_e = T_e + V_{ee} + V_{ne}$$

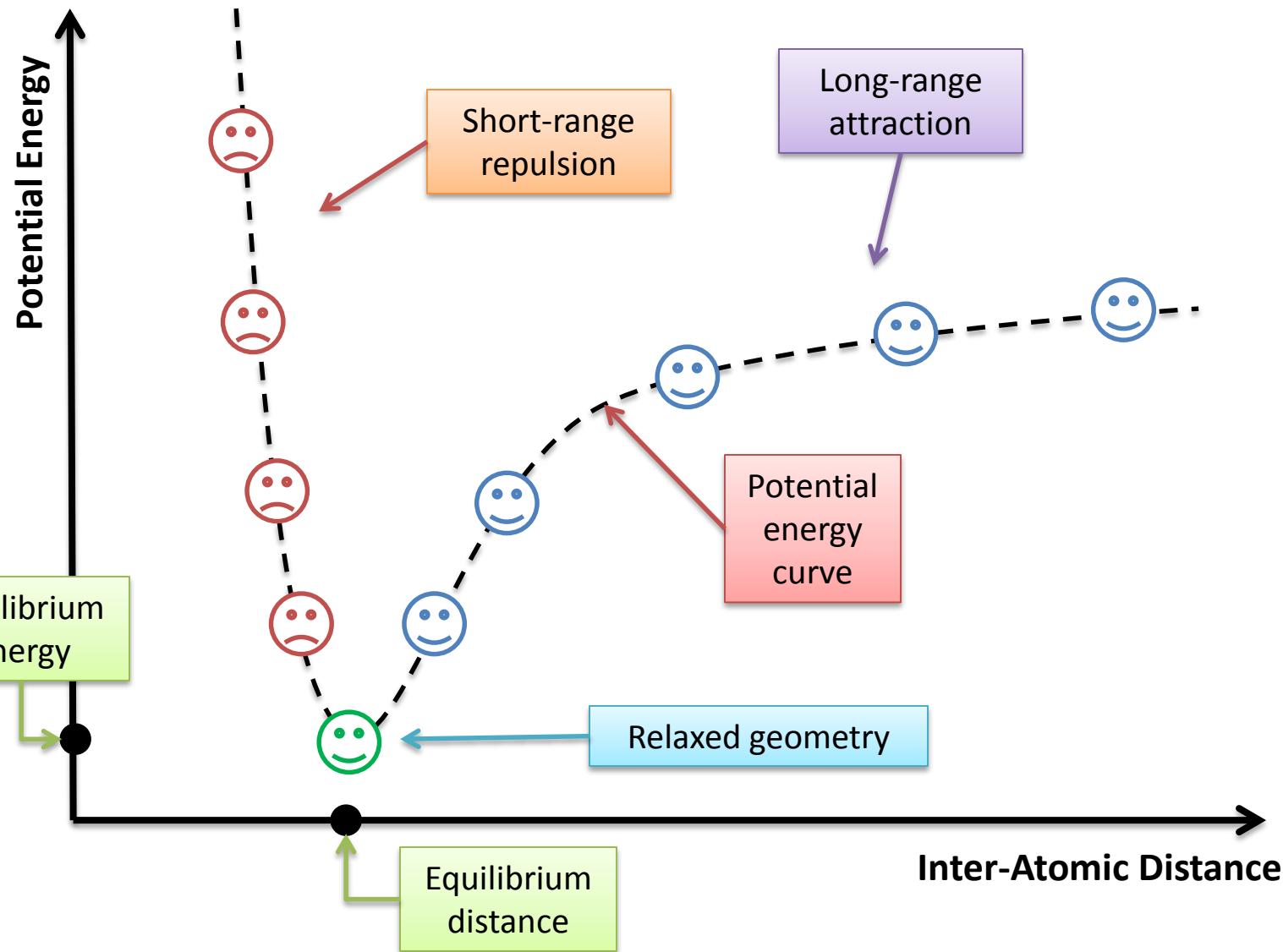
$$H_e \Psi_e = E_e \Psi_e$$

$$E_{tot} = E_e + V_{nn}$$

This separation is known as the
Born–Oppenheimer approximation



Building a PES: Diatomic System



Mathematical Models: Morse potential

$$V(r) = D_e (1 - e^{-a(r-r_e)})^2$$

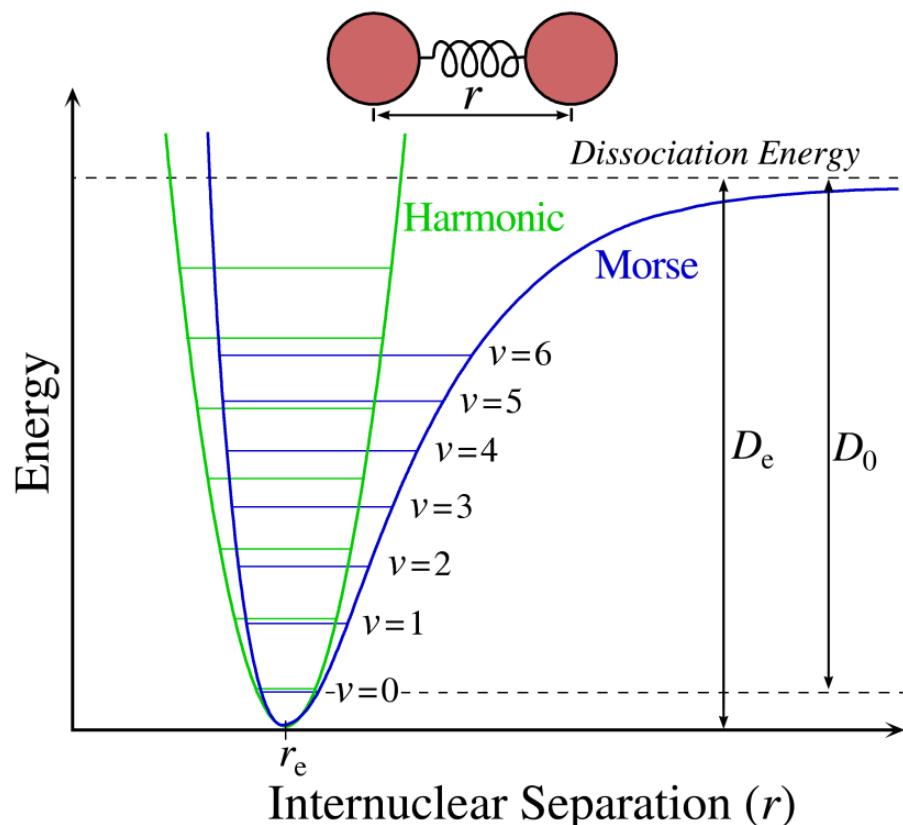
Named after physicist Philip M. Morse.

a convenient model for the potential energy of a diatomic molecule.

better approximation for the vibrational structure of the molecule than the quantum harmonic oscillator.

Explicitly includes the effects of bond breaking.

It also accounts for the anharmonicity of real bonds.



Mathematical Models: Lennard-Jones potential

parameters can be fitted to reproduce experimental data or accurate quantum chemistry calculations.

Due to its computational simplicity, Lennard-Jones potential is used extensively in computer simulations even though more accurate potentials exist.

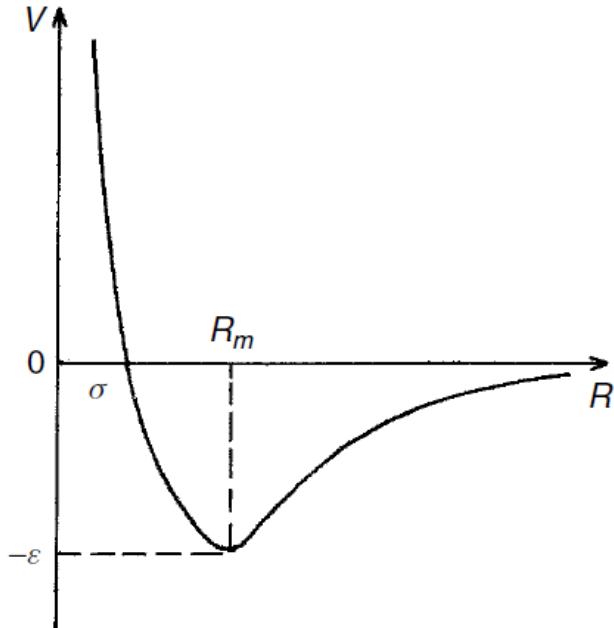
The general form

$$V(R) = \frac{\lambda_n}{R^n} - \frac{\lambda_m}{R^m}$$

The most popular form

$$V(R) = 4\epsilon \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right]$$

$$R_m = 2^{1/6}\sigma$$



	He–He	Ne–Ne	Ar–Ar
σ (Å)	2.556	2.749	3.405
ϵ/k (K)	10.22	35.60	119.8

Modifications: (12–6–4) potential

$$V(R) = 2\epsilon \left[(1 + \gamma) \left(\frac{\sigma}{R} \right)^{12} - 2\gamma \left(\frac{\sigma}{R} \right)^6 - 3(1 - \gamma) \left(\frac{\sigma}{R} \right)^4 \right]$$

A Two-Dimensional PES: Symmetric Water Molecule

The water geometry is defined in general by two bond lengths and a bond angle.

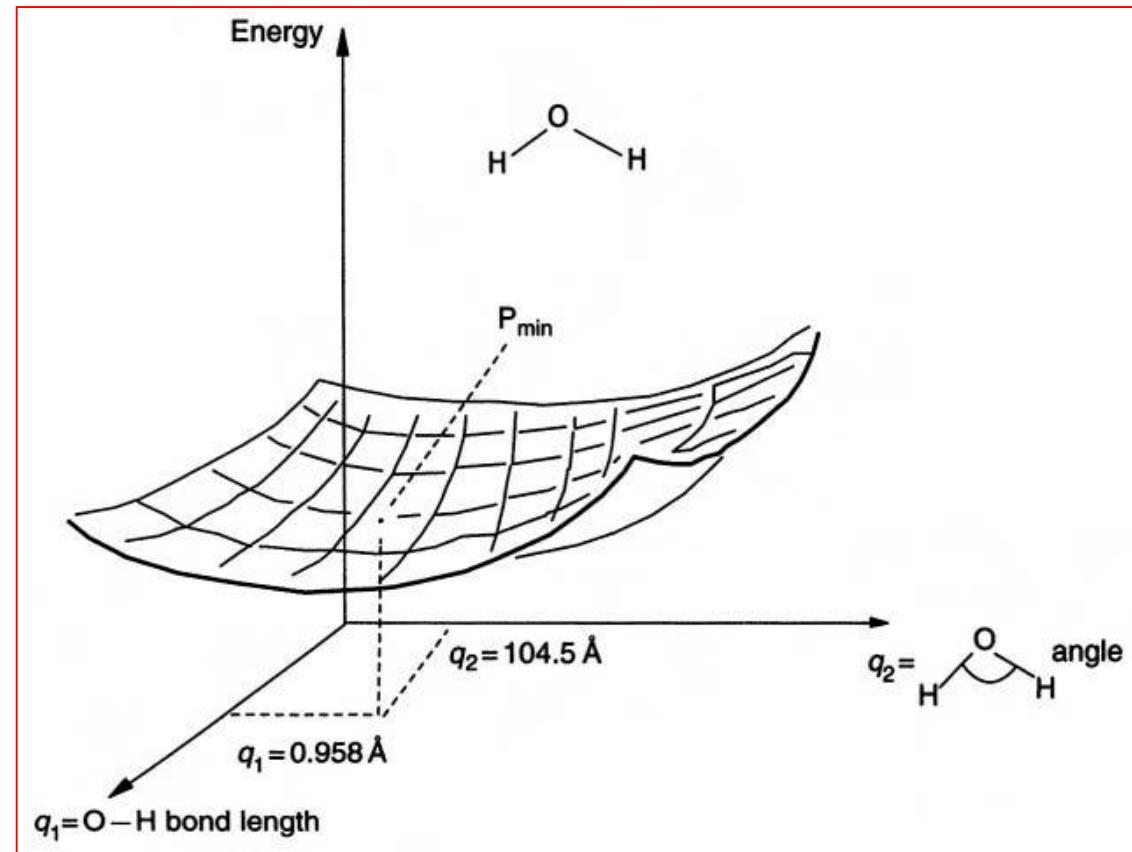
C_{2v} symmetry:

two planes of symmetry and a two-fold symmetry axis.

Symmetry Constraint:
keep the two bond lengths to the same.

Geometric parameters:
 q_1 is the O-H bond length and
 q_2 is the H-O-H bond angle.

Equilibrium Geometry:
The minimum potential energy geometry is the point at which
 $dE/dq_1=dE/dq_2=0$



In higher dimensions:
PESs are called potential energy **hypersurfaces**. Despite the problem of drawing a hypersurface, we can define the **equation** $E=f(q_1,q_2,q_3)$

Dealing with Dimensionality: Using a slice of a Complex PES

Slice parallel to an axis:

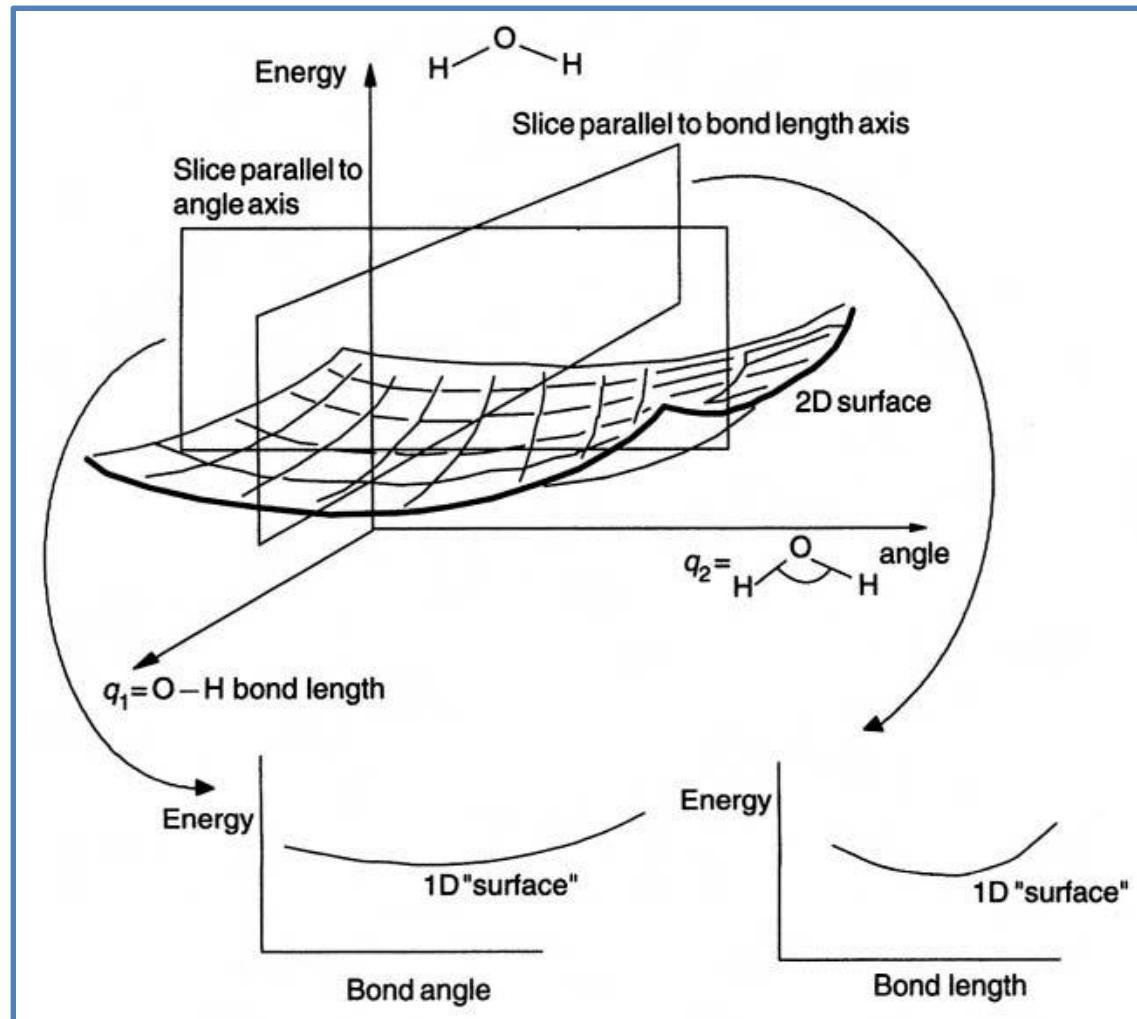
The slice could be made holding one or the other of the two geometric parameters constant.

Slice parallel to no axis:

The slice could involve both of them, giving a diagram in which the geometry axis is a composite of more than one geometric parameter.

Relaxed or Frozen PES slice:

Other geometrical parameters might be **fixed** or **relaxed**.



Stationary Points on the PES: Minima and Saddle Points

At all Stationary Points:

$$\frac{\partial E}{\partial q_1} = \frac{\partial E}{\partial q_2} = \cdots = 0$$

At a minimum for all coordinates q :

$$\frac{\partial^2 E}{\partial q^2} > 0$$

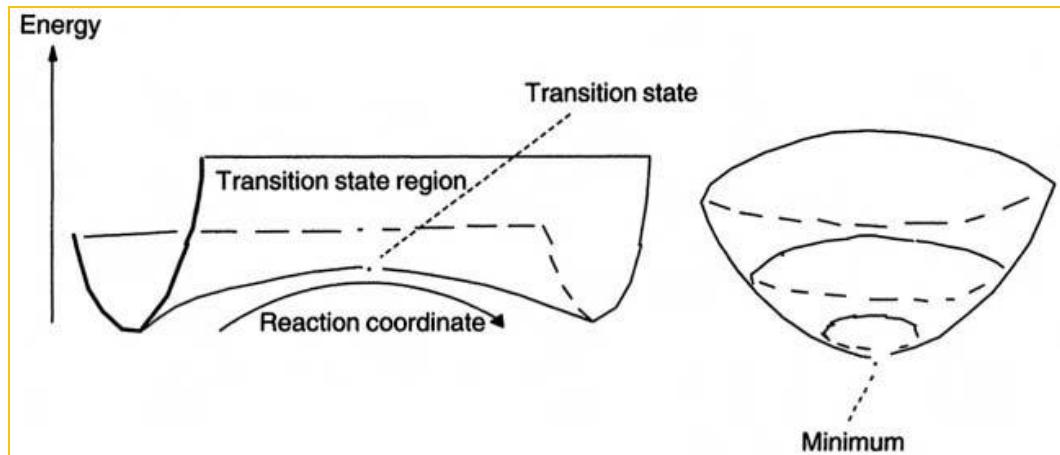
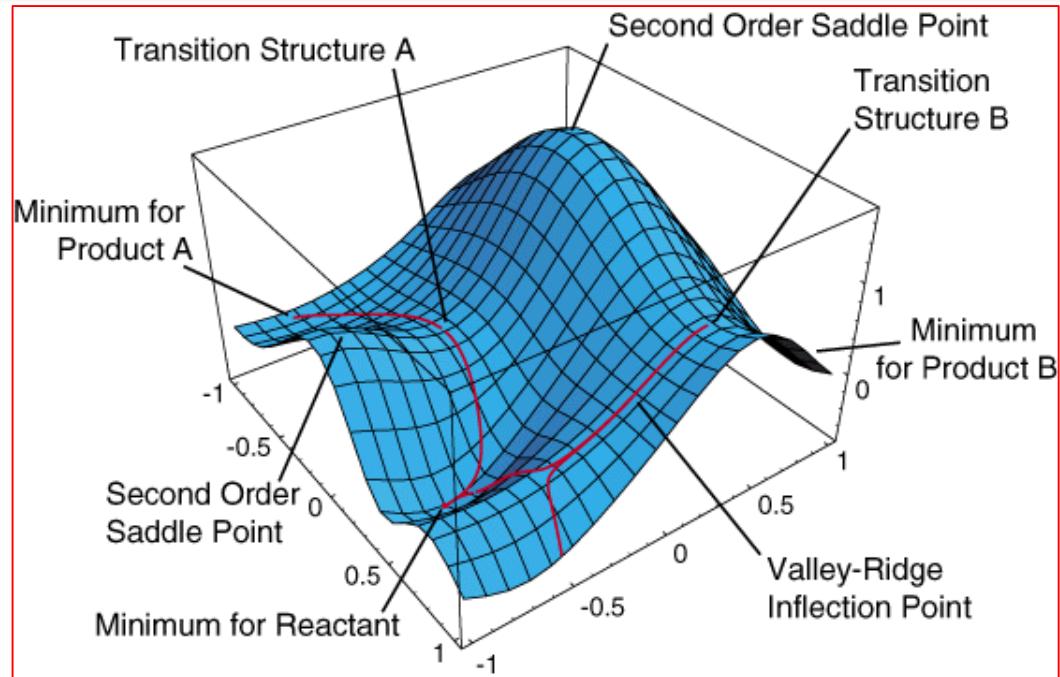
At a transition state:

for all q , except along reaction coordinate:

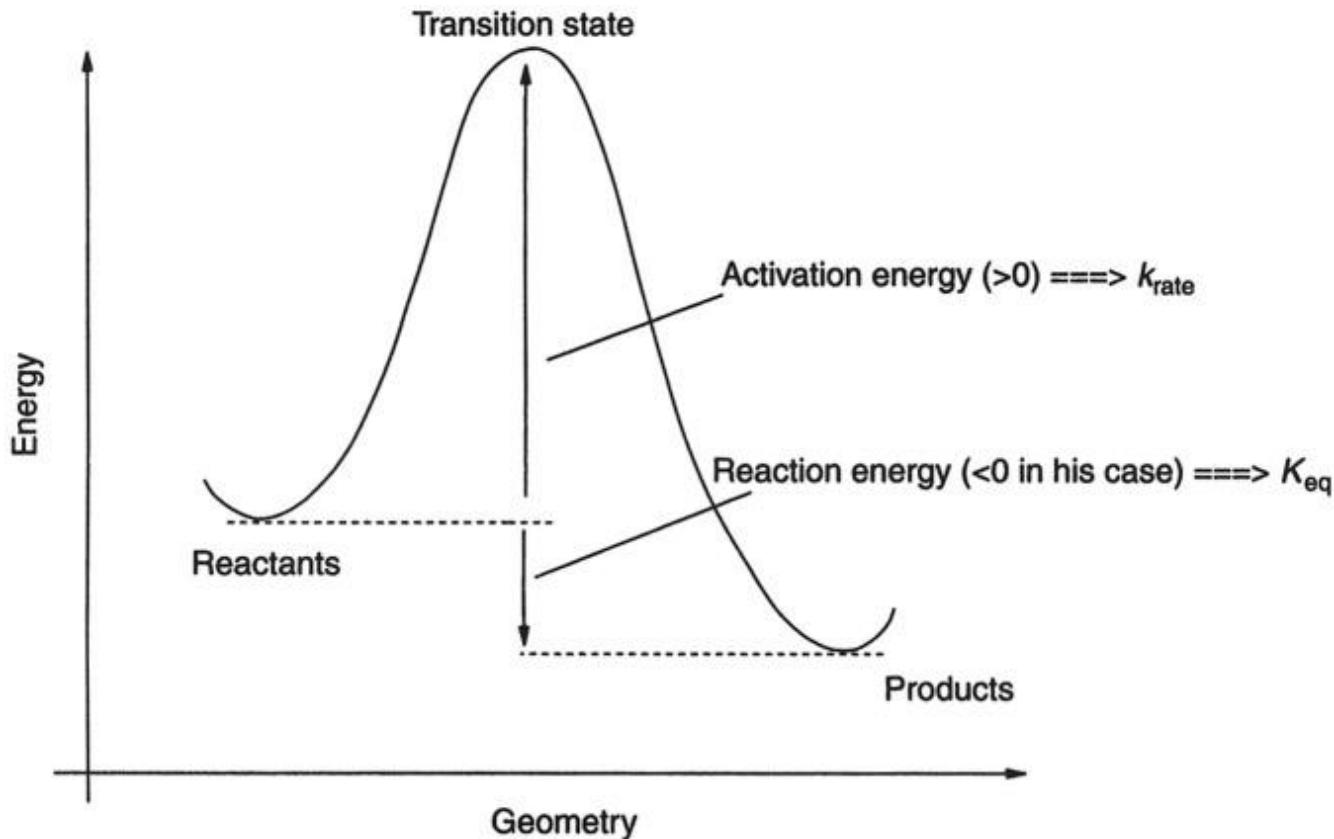
$$\frac{\partial^2 E}{\partial q^2} > 0$$

And along the reaction coordinate:

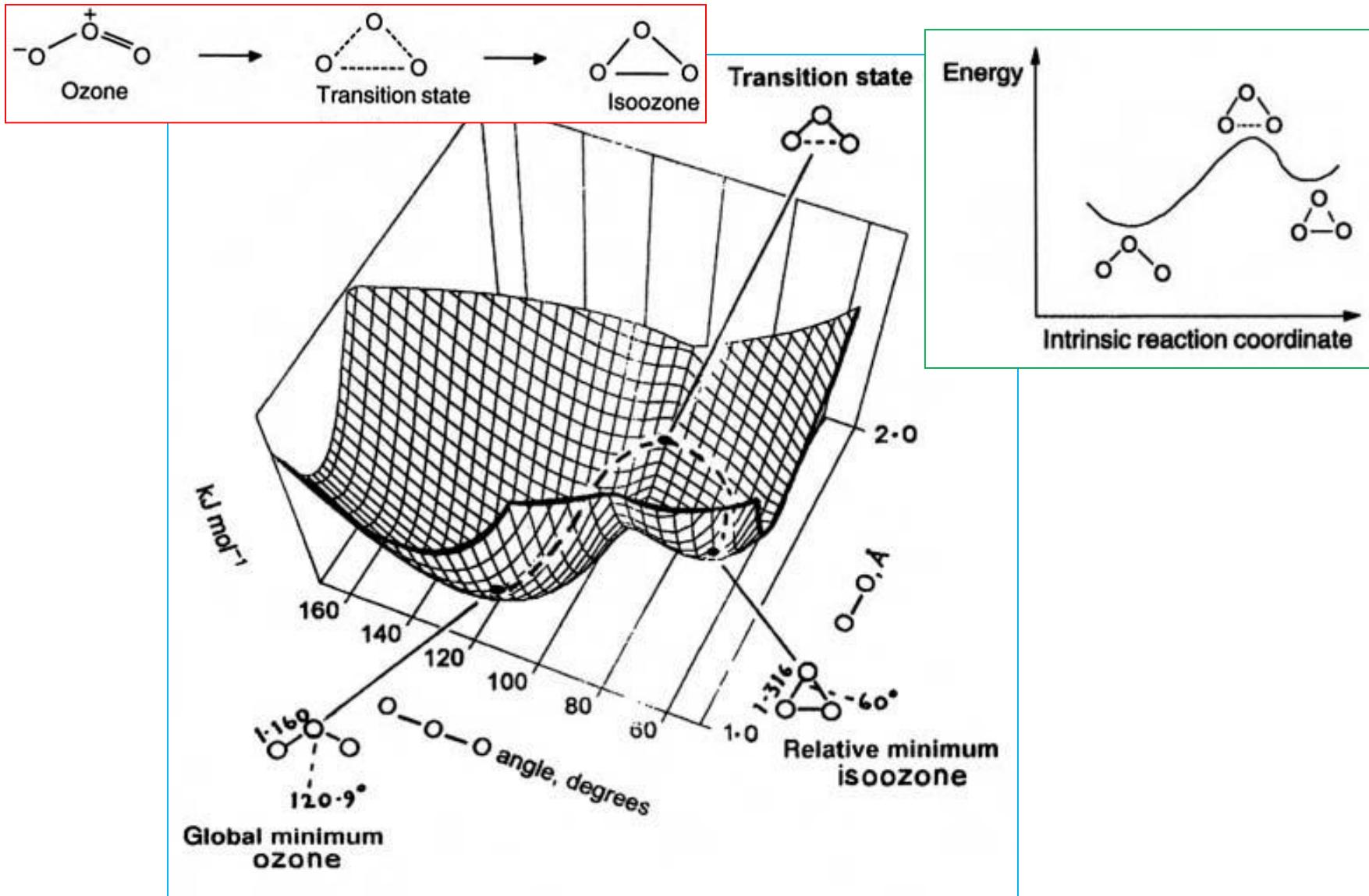
$$\frac{\partial^2 E}{\partial q^2} < 0$$



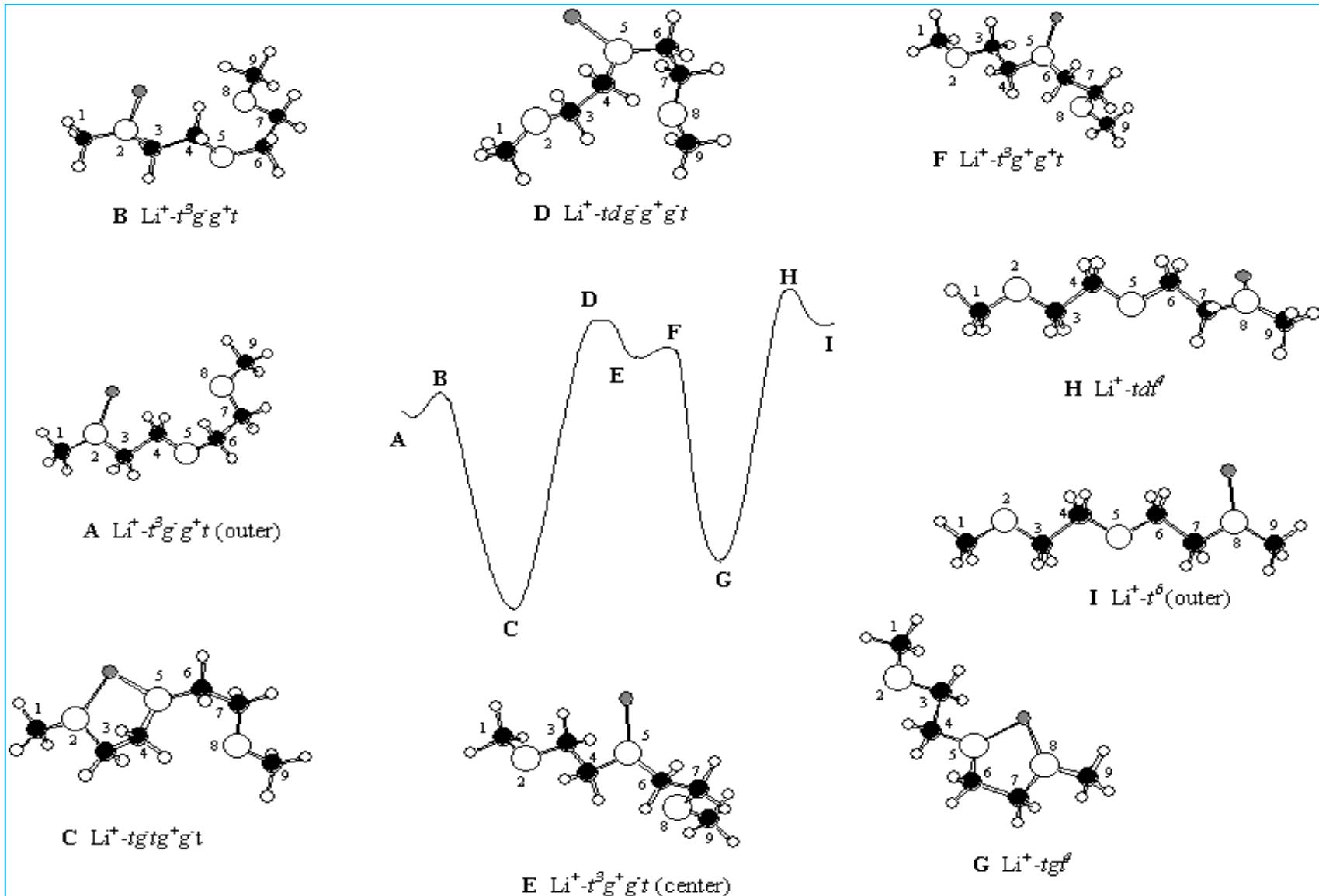
Thermodynamic and Kinetic Properties



Example: conversion of ozone to isoozone



Example: migration of Li cation along a polymer



Gradient Vector and Hessian Matrix

Gradient Vector:

vector that points in the direction of greatest rate of increase of a function, and its magnitude is the greatest rate of change.

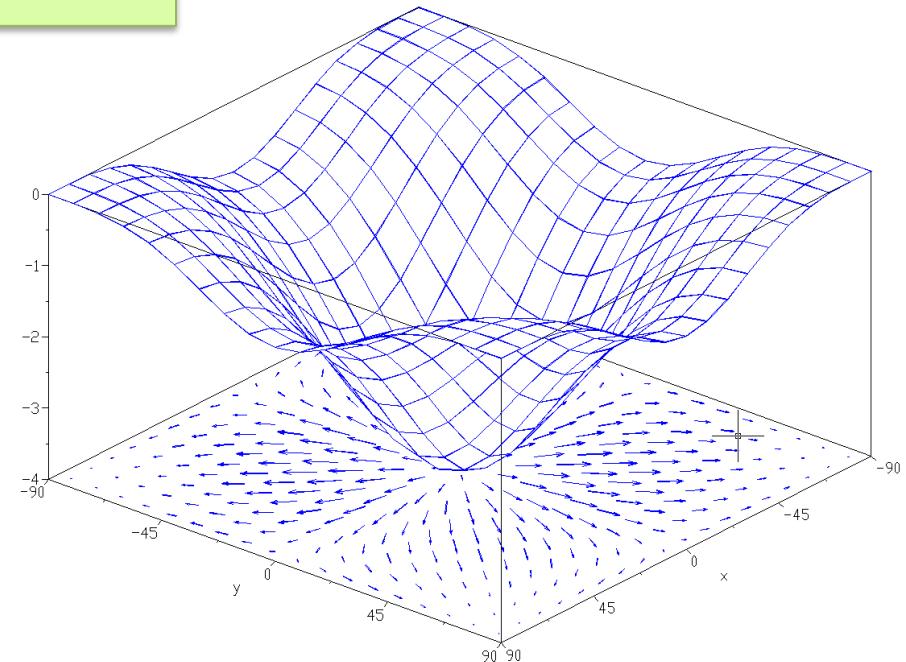
Hessian matrix:

square matrix of second-order partial derivatives of a function.

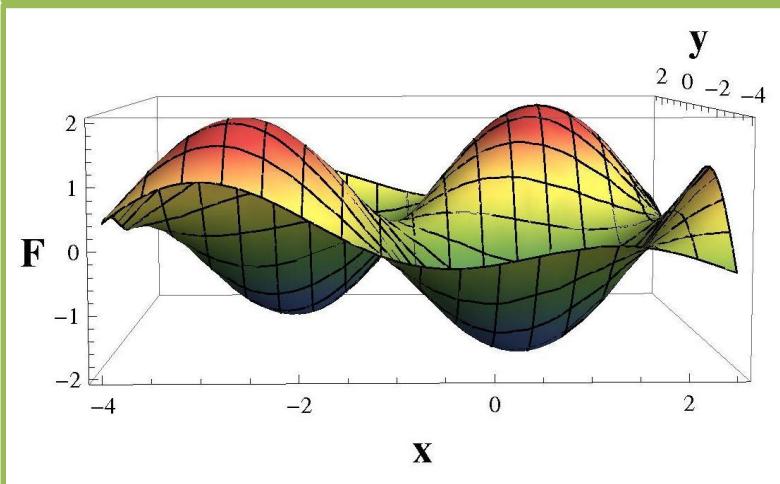
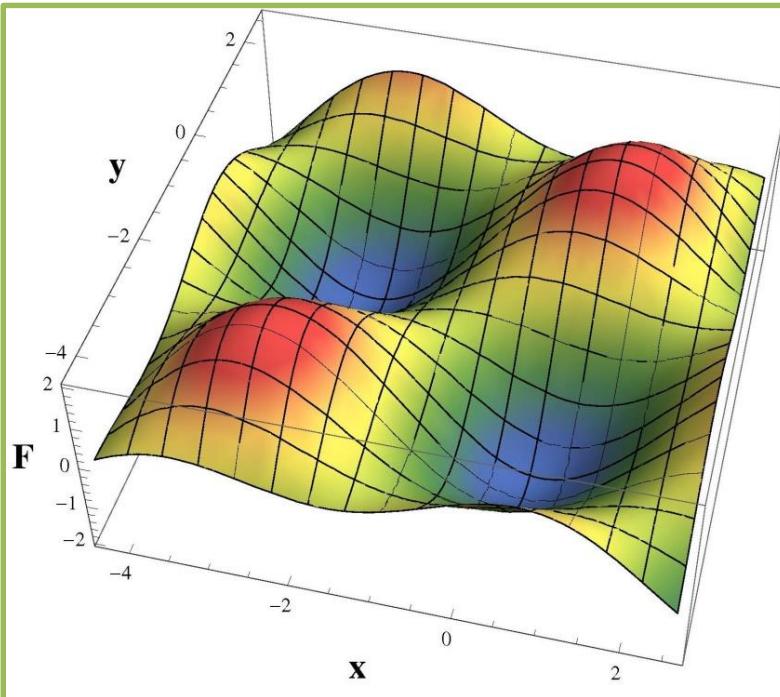
$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

$$f(x_1, x_2, \dots, x_n),$$

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right).$$



Gradient Vector and Hessian Matrix



$$F(x, y) = \sin(ax + by) + \cos(cx - dy)$$

Gradient Vector

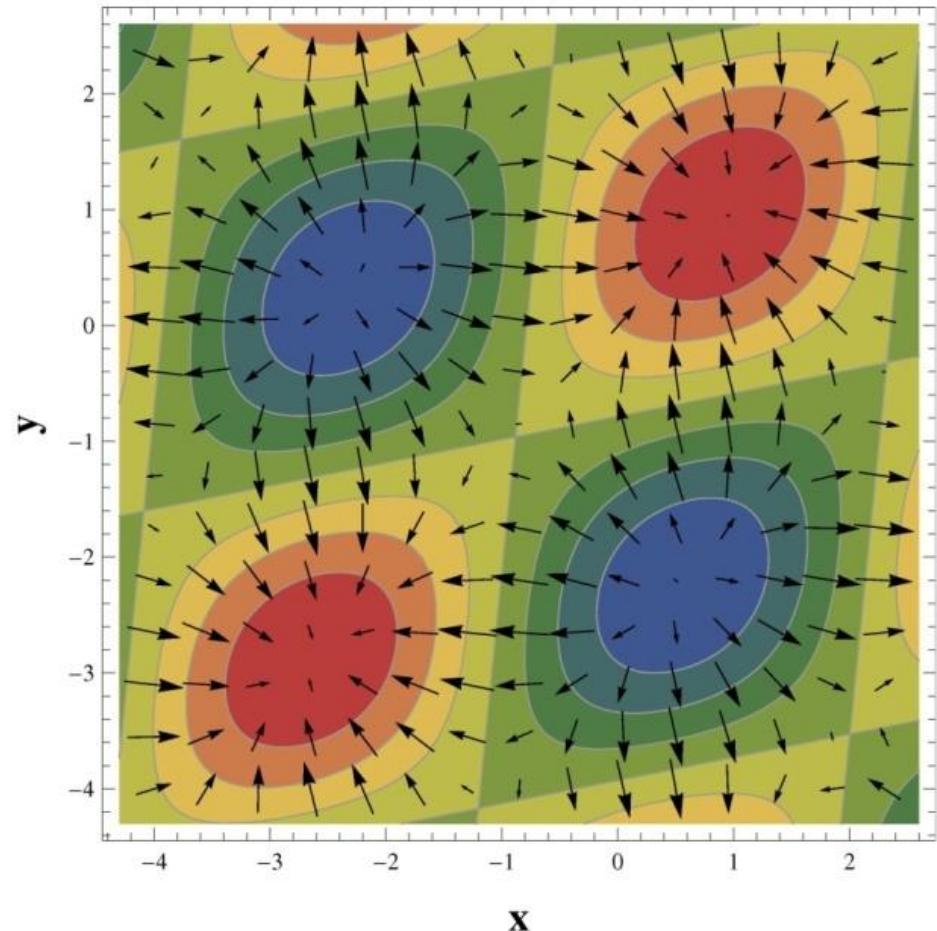
$$\bar{G}(x, y) = \begin{bmatrix} \frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial y} \end{bmatrix}$$

Hessian Matrix

$$\bar{H}(x, y) = \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} \end{bmatrix}$$

Gradient Vector and Hessian Matrix

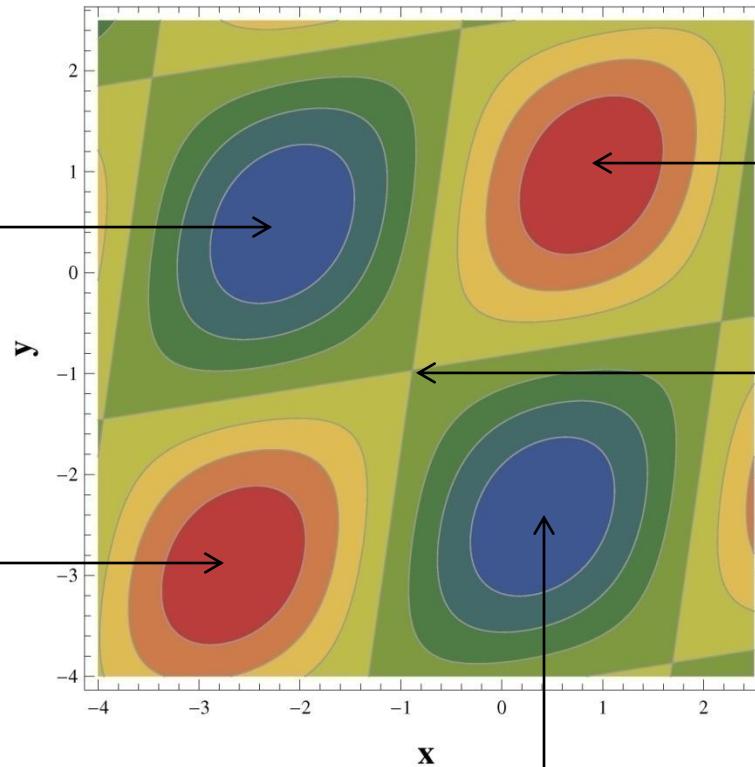
Gradient vector points in the direction where the function increases most.



$$\bar{G}(\min1) = \bar{G}(\min2) = \bar{G}(\max1) = \bar{G}(\max2) = \bar{G}(\text{sadl}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Gradient Vector and Hessian Matrix

$$min2 = \begin{bmatrix} -2.17 \\ 0.48 \end{bmatrix}$$



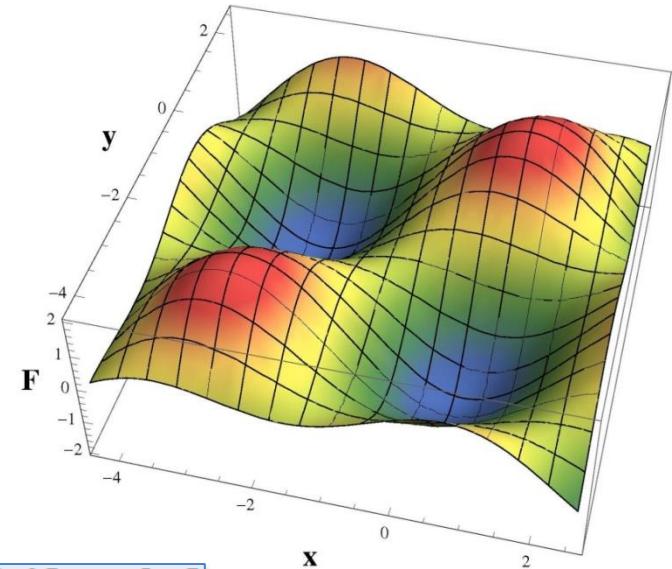
$$max2 = \begin{bmatrix} 0.89 \\ 0.97 \end{bmatrix}$$

$$max1 = \begin{bmatrix} -2.66 \\ -2.9 \end{bmatrix}$$

$$sadl = \begin{bmatrix} -0.886 \\ -0.967 \end{bmatrix}$$

$$min1 = \begin{bmatrix} 0.40 \\ -2.42 \end{bmatrix}$$

Gradient Vector and Hessian Matrix



$$\bar{H}(\text{min1}) = \bar{H}(\text{min2}) = \begin{bmatrix} 2.25 & -0.6 \\ -0.6 & 1.85 \end{bmatrix} \xrightarrow{\text{diagonalize}} \begin{bmatrix} 2.68 & 0 \\ 0 & 1.42 \end{bmatrix}$$

$$\bar{H}(\text{sadl}) = \begin{bmatrix} -0.63 & 2.04 \\ 2.04 & -0.57 \end{bmatrix} \xrightarrow{\text{diagonalize}} \begin{bmatrix} -2.64 & 0 \\ 0 & 1.44 \end{bmatrix}$$

$$\begin{aligned} \bar{H}(\text{max1}) &= \bar{H}(\text{max2}) \\ &= \begin{bmatrix} -2.25 & 0.6 \\ 0.6 & -1.85 \end{bmatrix} \xrightarrow{\text{diagonalize}} \begin{bmatrix} -2.68 & 0 \\ 0 & -1.42 \end{bmatrix} \end{aligned}$$

Optimization and Minimization Methods

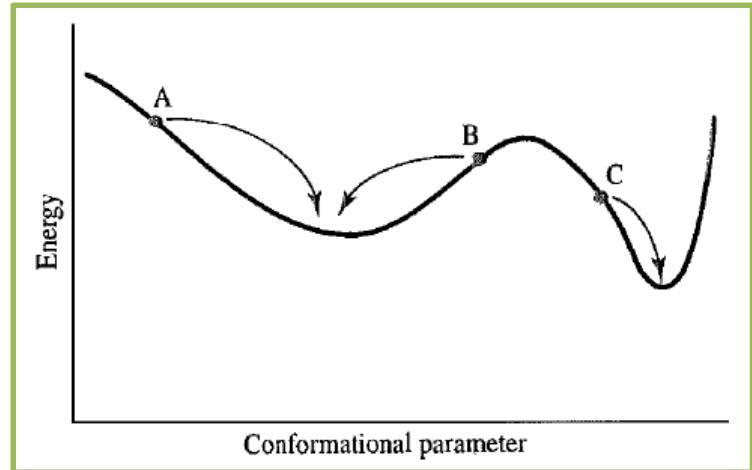
Optimization

finding *stationary* points of a function

Mostly,
desired stationary point is a minimum.

Minimization Algorithms

- 1) only go downhill on PES.
- 2) only locate the nearest minimum.



Conformational Search

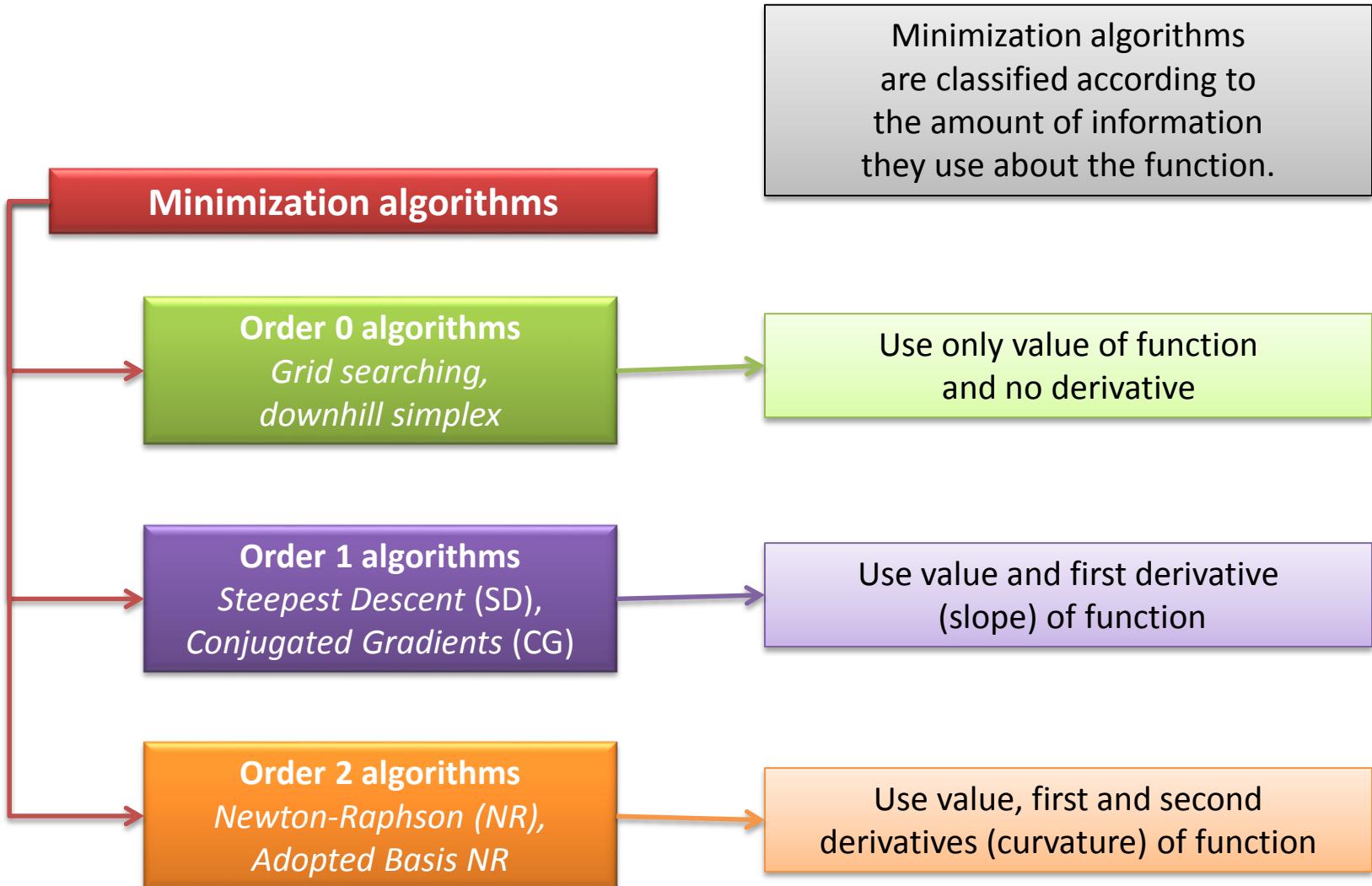
Minimizing each starting point

Generate different starting points

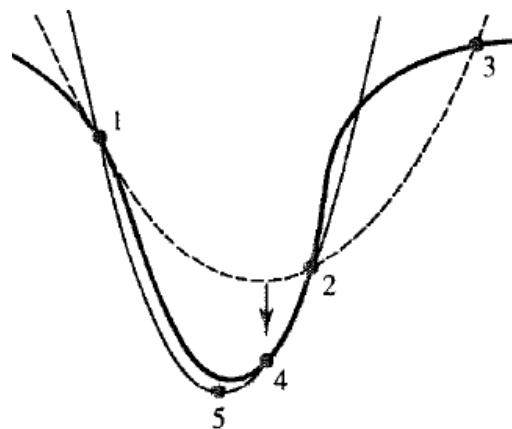
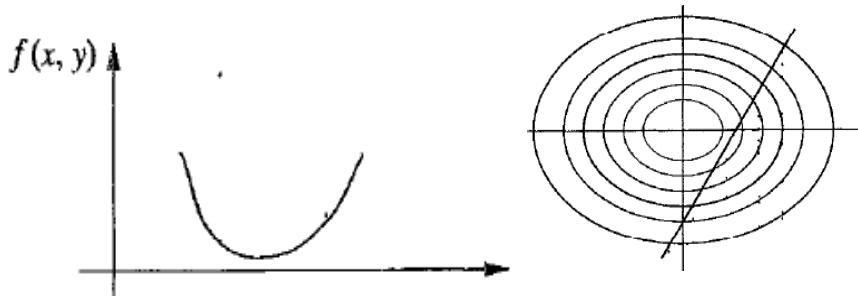
locate more than one minimum

Locate global energy minimum

Minimization algorithms



Order 1 algorithms: Steepest Descent



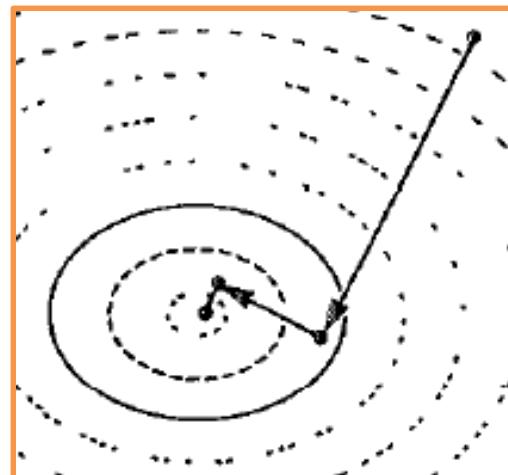
A sure-fire algorithm
If accurate line minimization, always lowering energy,
guarantee to approach a minimum

Step direction
negative of gradient direction

$$\vec{S}_k = -\vec{g}_k = -\nabla U(\mathbf{r})$$

At each step

- 1) Line search along step direction.
- 2) Scale step size according to energy change.



Order 1 algorithms: Conjugated Gradients

Step direction

First Step: negative of gradient direction

$$\vec{S}_1 = -\vec{g}_1$$

Other Steps: weighted average of current gradient and previous direction(s)

$$\vec{S}_k = -\vec{g}_k + b_k \vec{S}_{k-1} \quad b_k = |g_k|^2 / |g_{k-1}|^2$$

At each step

- 1) Line search along step direction.
- 2) Use of memory of gradients calculated in previous steps.
- 3) Each direction is conjugate to previous direction(s).
- 4) Scale step size according to energy change.

For quadratic surfaces of dimension **N**, converging on the **N**th step,

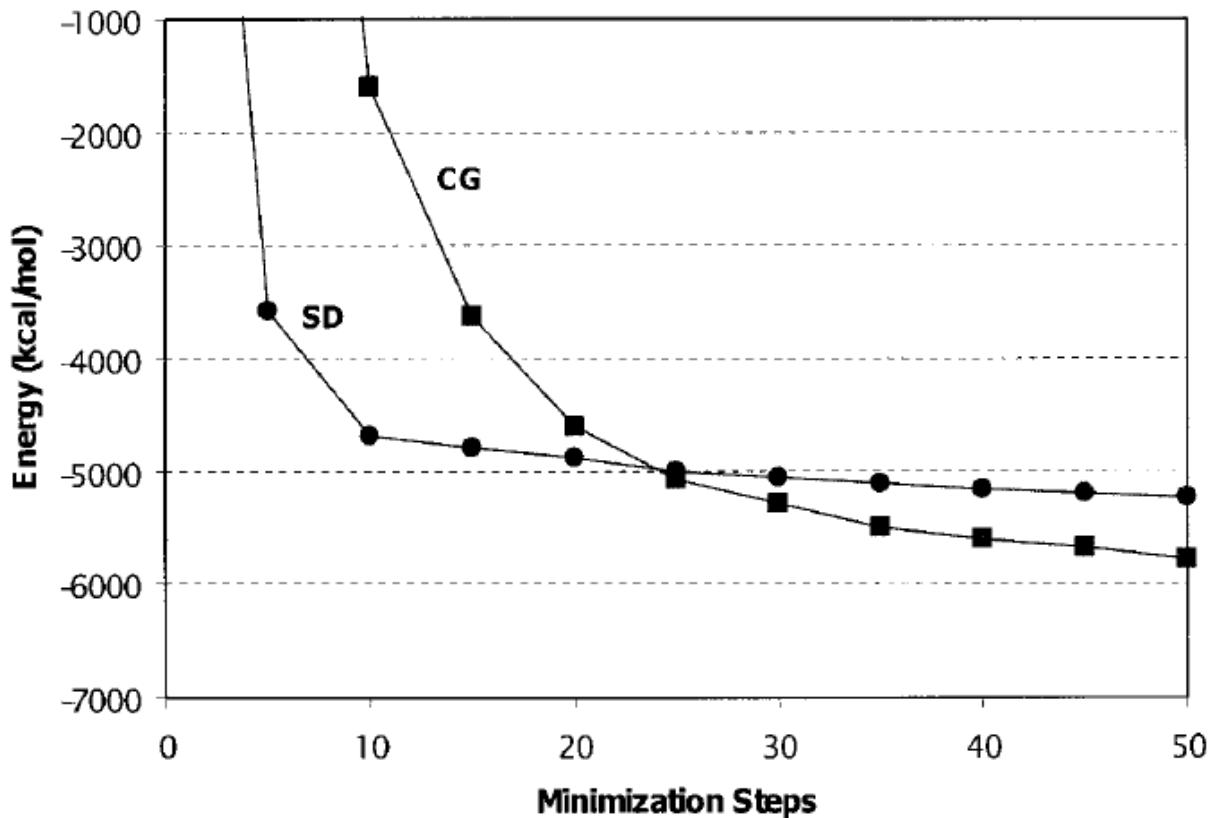
BUT

Molecular energy surfaces are non-quadratic.

Nonetheless,

CG converges much better than SD.

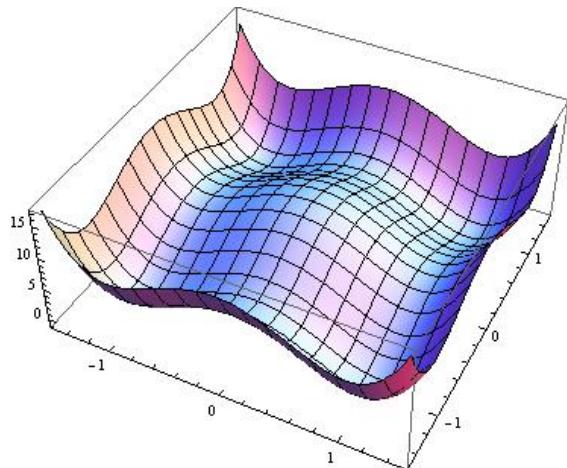
SD versus CG: Performance Comparison



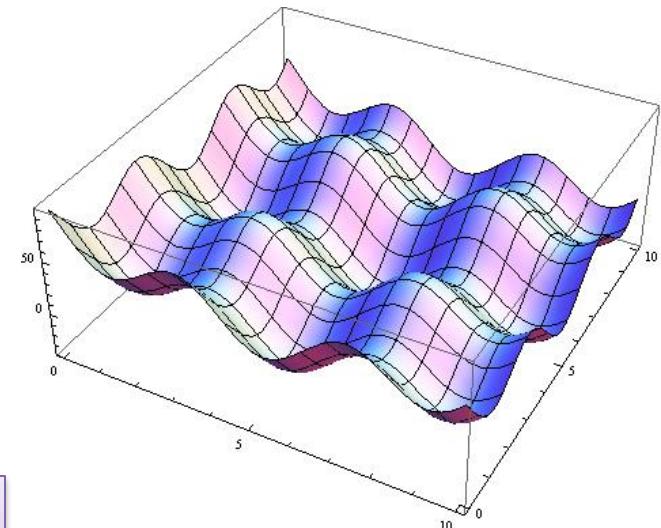
Termination criterion
is usually defined in
terms of the gradient
RMS (GRMS).

$$\text{RMS} = \sqrt{\frac{\mathbf{g}^T \mathbf{g}}{3N}}$$

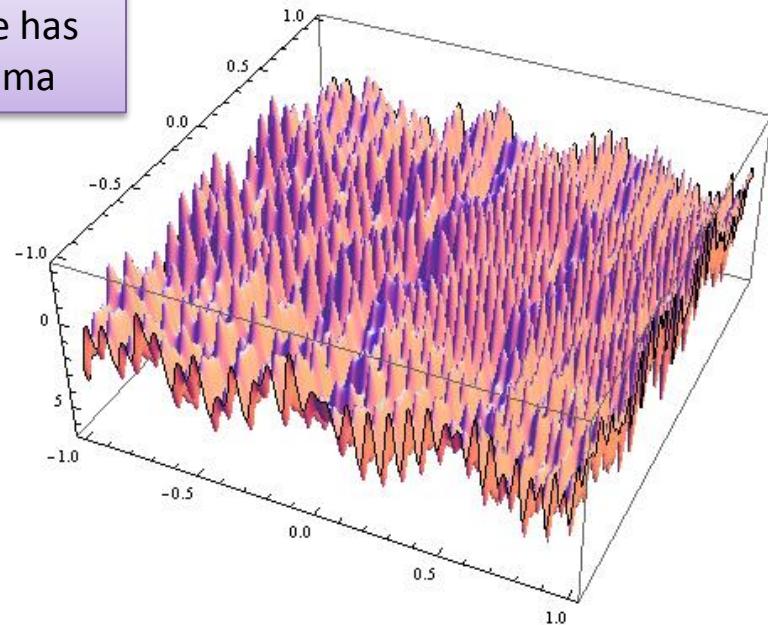
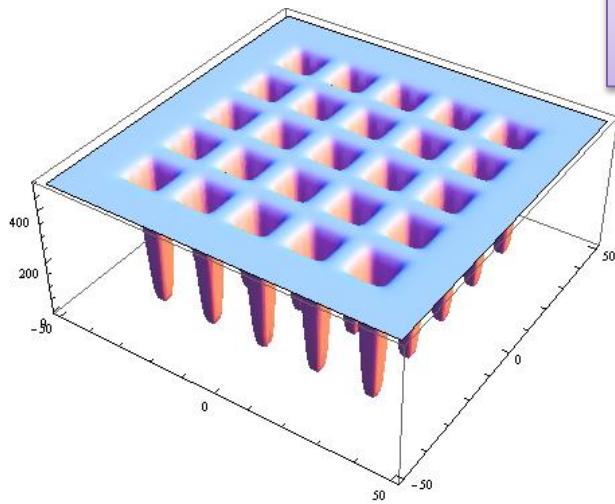
Energy Landscape of Complex Systems



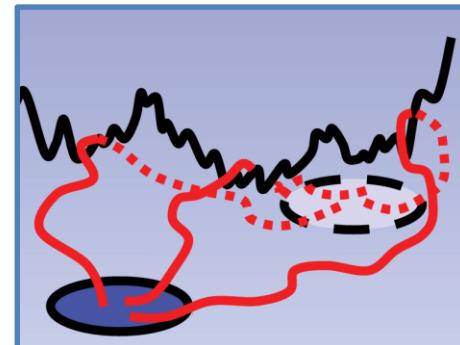
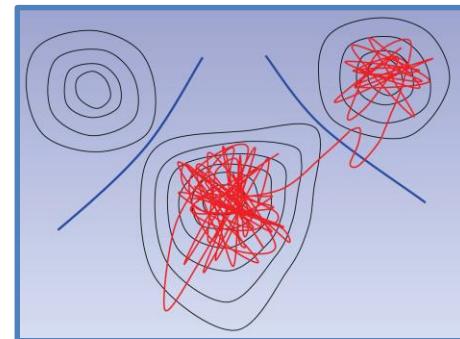
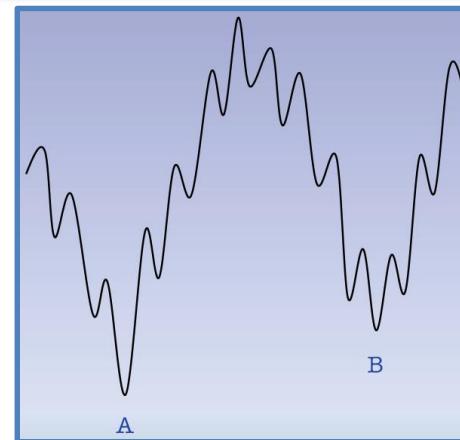
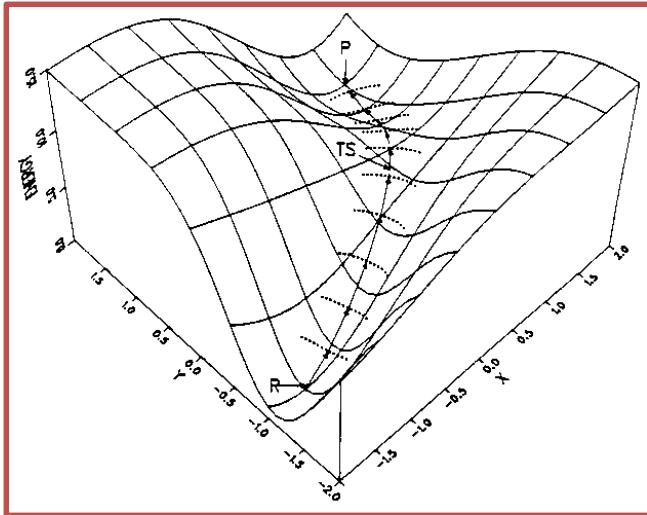
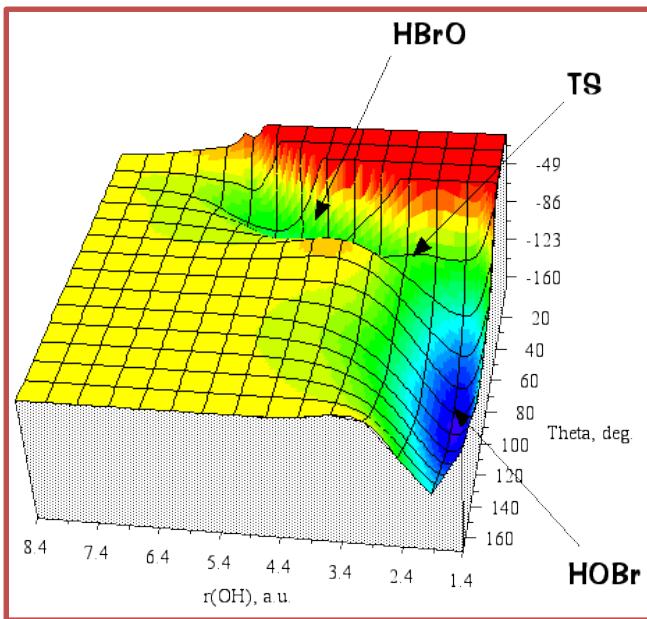
In a complex system it
should not be a simple
smooth surface



In a complex system the
energy landscape has
many local minima



Smooth Potentials versus Rugged Landscapes

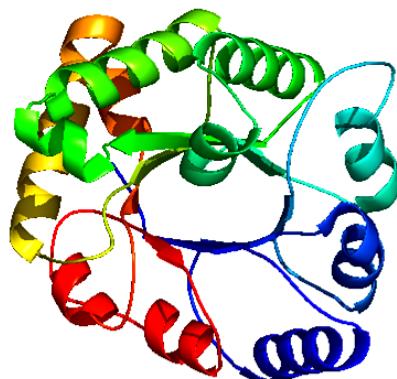


Protein Structure Diversity

1CTF



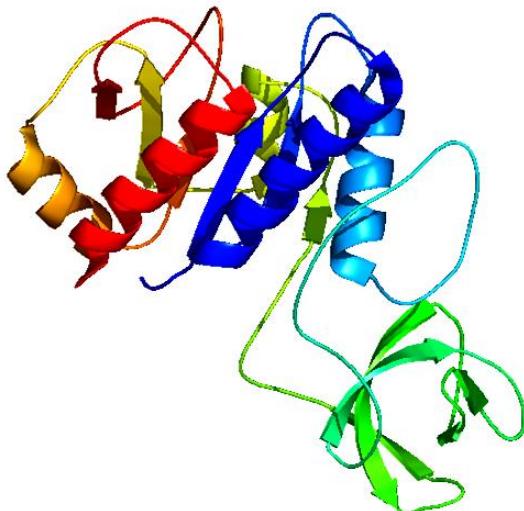
1TIM



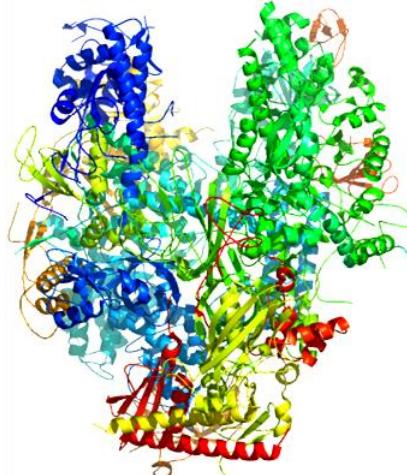
1A1O



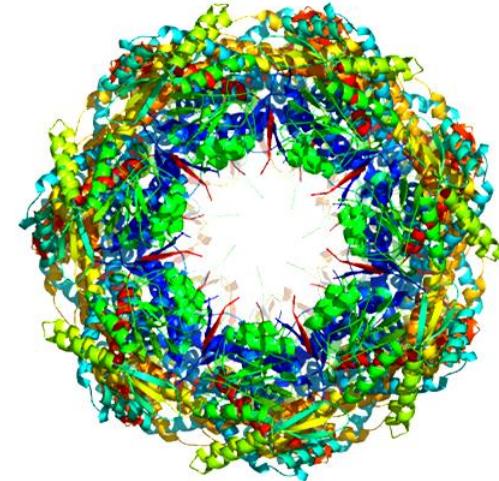
1K3R



1NIK



1AON



Anfinsen Experiment

A remarkable discovery (In about 1960):

A globular protein is capable of spontaneous folding in vitro.

renaturation requires:

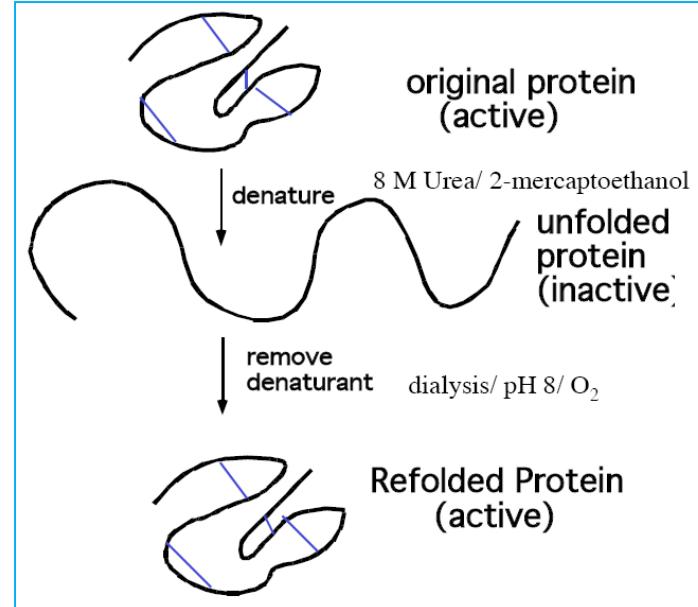
careful selection of experimental conditions

otherwise,

aggregation can prevent protein chains from folding

chemically synthesized protein chain, (without any cell or ribosome), placed in proper conditions, folds into a biologically active protein.

Protein folding is the most simple case of pure self-organization:
Nothing “biological” (but for the sequence!) helps protein chain to fold.



All of the information necessary for folding the peptide chain into its “native” structure is contained in amino acid sequence of peptide.

Protein Folding Puzzle: Levinthal's Paradox

On the one hand,

the same native state is achieved by various folding processes:

- 1) in vivo on the ribosome
- 2) in vivo after translocation through the membrane
- 3) in vitro after denaturation with various agents
- 4) in vitro after chemically synthesizing the protein chains

This suggests that:

the native state is thermodynamically
the most stable state under
the “biological” conditions.

Protein Folding Puzzle: Levinthal's Paradox

On the other hand,

For a small protein with 100 amino acids:

2 conformations for each amino acid

$2^{100} \approx 10^{30}$ conformations for protein

10^{-12} s for each conformational change

The time required to sample all conformations:

$$(10^{30} \text{ conformations}) \times (10^{-12} \text{ s/conformation})$$

$$= 10^{18} \text{ s}$$

$$\approx 10^{10} \text{ years}$$

Age of the universe $\approx 1.37 \times 10^{10}$ years

Protein Folding Puzzle: Levinthal's Paradox

The paradox is that,

on the one hand,

achievement of the same (native) state
by a variety of processes is
a clear-cut evidence of its stability.

On the other hand,

protein simply does not have enough time
to prove that the native structure is the most stable
among all possible structures!

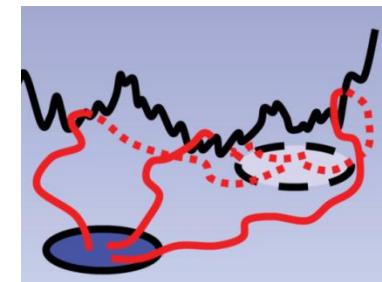
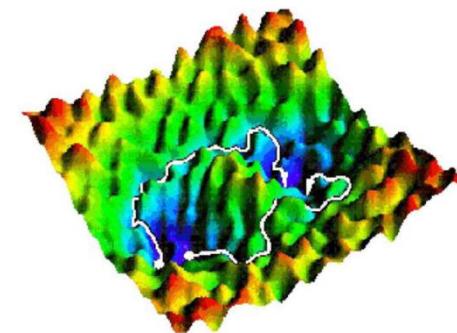
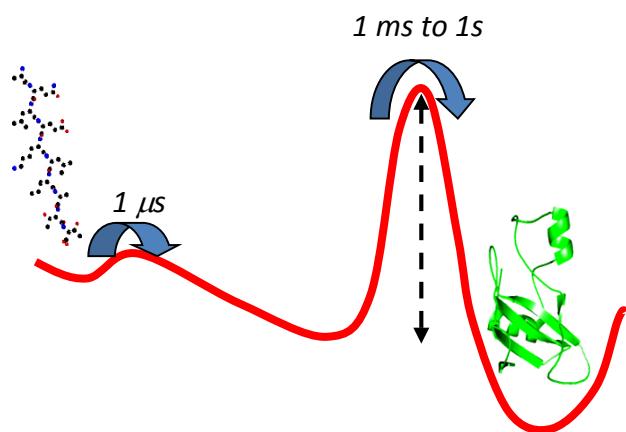
Then,

how does the protein choose
its native structure among zillions of others?

Protein Folding Puzzle: Levinthal's Paradox

How Do Proteins Know How to Fold?

- ⇒ proteins do not fold by a random search of conformation space.
- ⇒ there exists a specific folding pathway.
- ⇒ folding pathway is narrow and a small part of conformational space is sampled.
- ⇒ native structure is under kinetic rather than thermodynamic control.



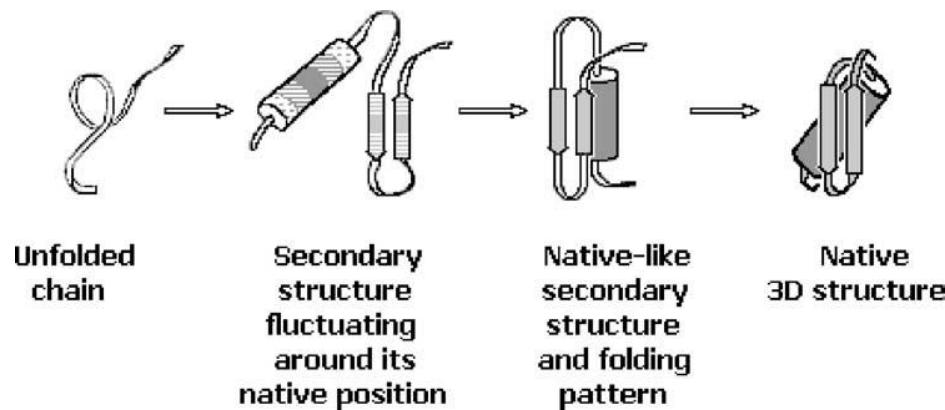
- ⇒ native structure is the end of folding pathway rather than the most stable chain fold.
- ⇒ native structure is not global but easily accessible energy minimum.

folding pathways and intermediates

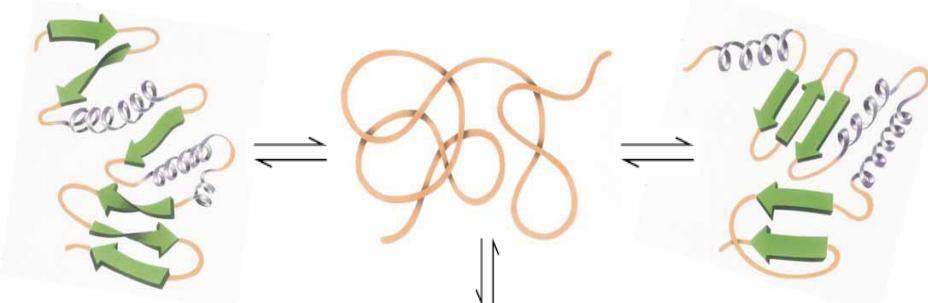
Framework model of stepwise folding:

Stepwise involvement of different interactions:

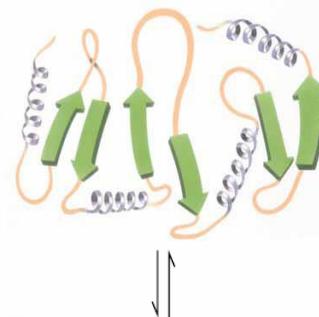
- 1) Rapidly folded α -helices and β -hairpins at the initial folding steps.
- 2) Gluing of these helices and hairpins into a native-like globule.
- 3) Final crystallization of the structure within this globule.



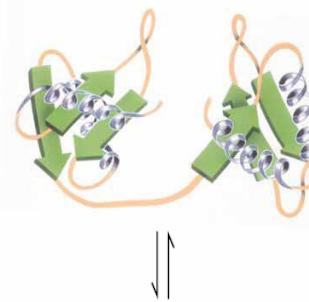
(1) Rapid and reversible formation of local secondary structures.



(2) Formation of domains by cooperative aggregation of folding nuclei.



(3) Molten globule formation of the assembled domains.



(4) Adjustment in the conformation of the domains.

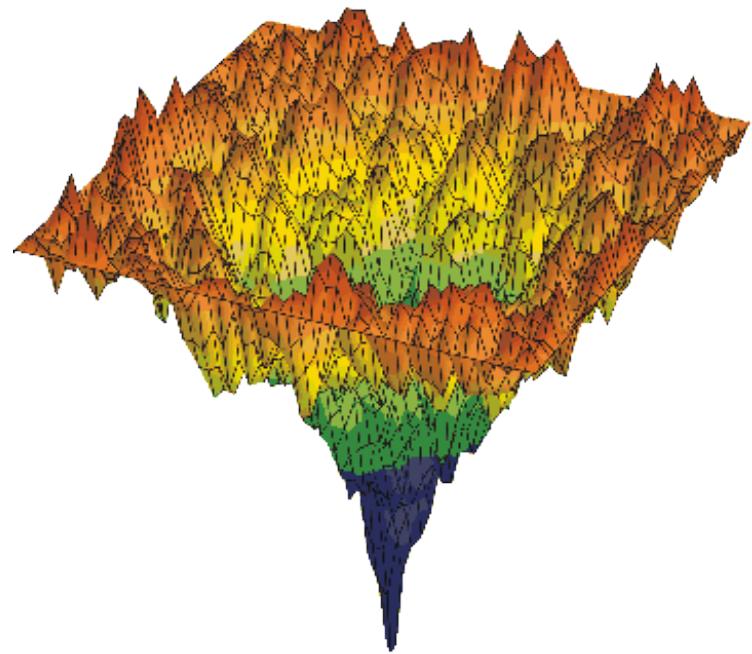
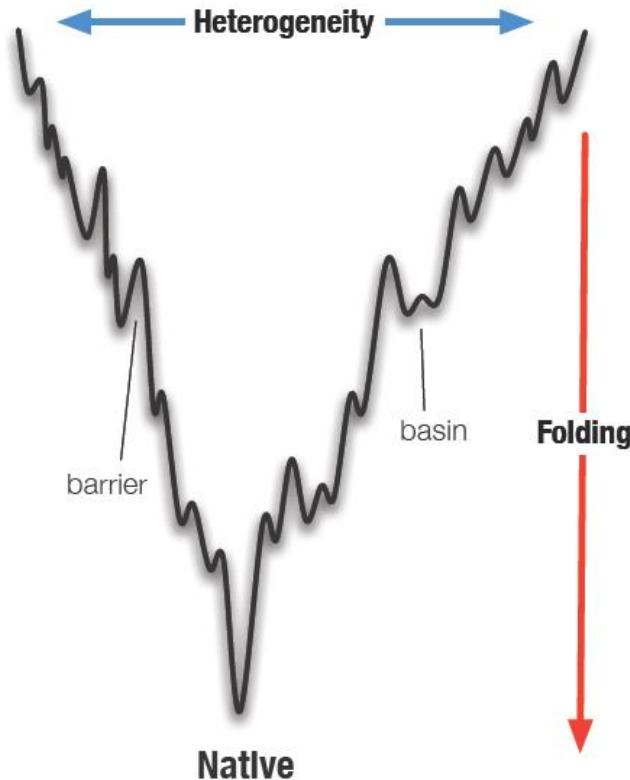


(5) Final protein monomer.

Funnel Like Landscapes of Proteins

The landscape appears to be a funnel.

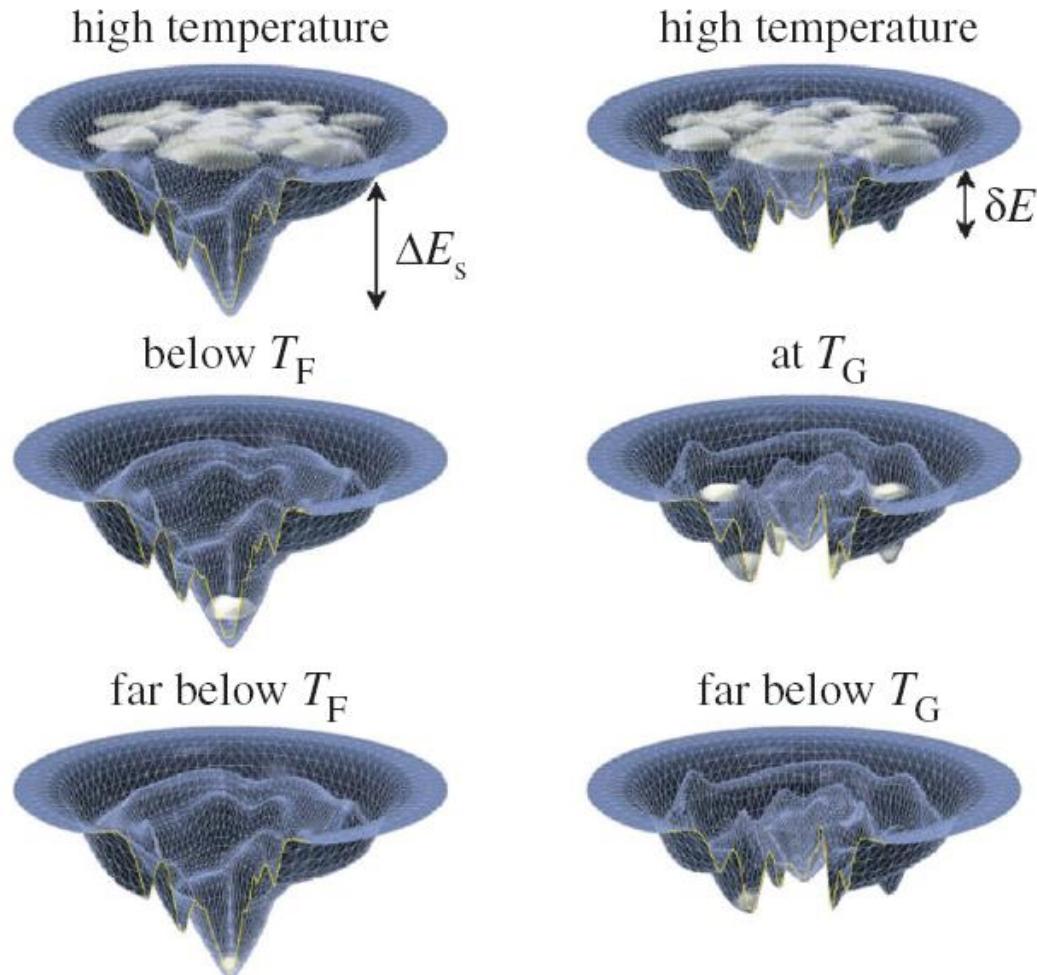
Collective dominance of native-like interactions over other alternate combinations of interactions organizes the energy landscape of a foldable protein.



The mouth of the funnel represents high entropy of denatured states.

Heterogeneity or entropy becomes smaller as more contacts are formed and at the same time as stabilization energy decreases.

Comparison of the Landscapes of Proteins and Honey

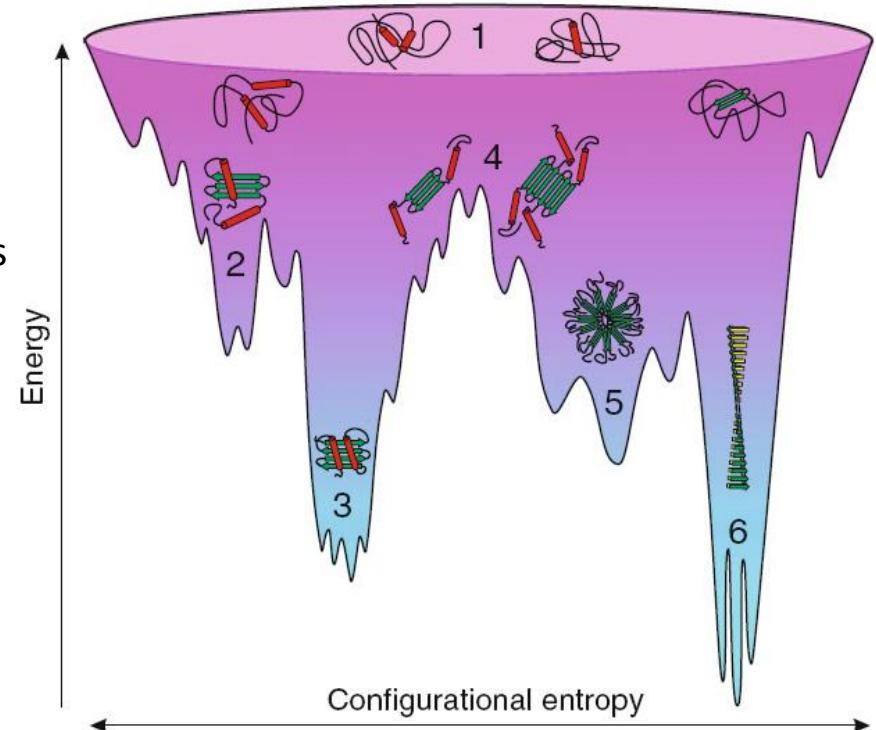


Linked Landscapes and Conformational Conversions

how the folding, misfolding and aggregation of proteins are linked?

The ruggedness of the landscape corresponds to intermediates and other metastable states:

- (1) Unfolded states showing varied conformations and residual structures.
- (2) Folding intermediates.
- (3) Native state, showing functional conformational changes.
- (4) Misfolded intermediates and small oligomers.
- (5) Toxic oligomers and protofibrils.
- (6) Amyloid fibrils, may have several distinct morphologies



proteins fold and misfold on linked, funnelled energy landscapes, where rarely populated states can provide access to misfolded and sometimes aggregated ensembles.

How Nature has encoded the target structures for systems that are good ‘structure seekers’ in the potential energy landscape

This question links
protein folding, crystallization,
'magic number' clusters in molecular beams and self-assembly processes.

In each case
the probability of finding the right structure in a random search is extremely small.

In fact, the PES can
effectively guide the system through the configuration space
if the local minima are connected in the right way.

The resulting topology can be identified directly
by constructing a connectivity graph.

The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics

Oren M. Becker

School of Chemistry, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel and Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

Martin Karplus

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138 and Laboratoire de Chimie Biophysique, Institut Le Bel, Université Louis Pasteur, 67000 Strasbourg, France

(Received 15 April 1996; accepted 18 September 1996)

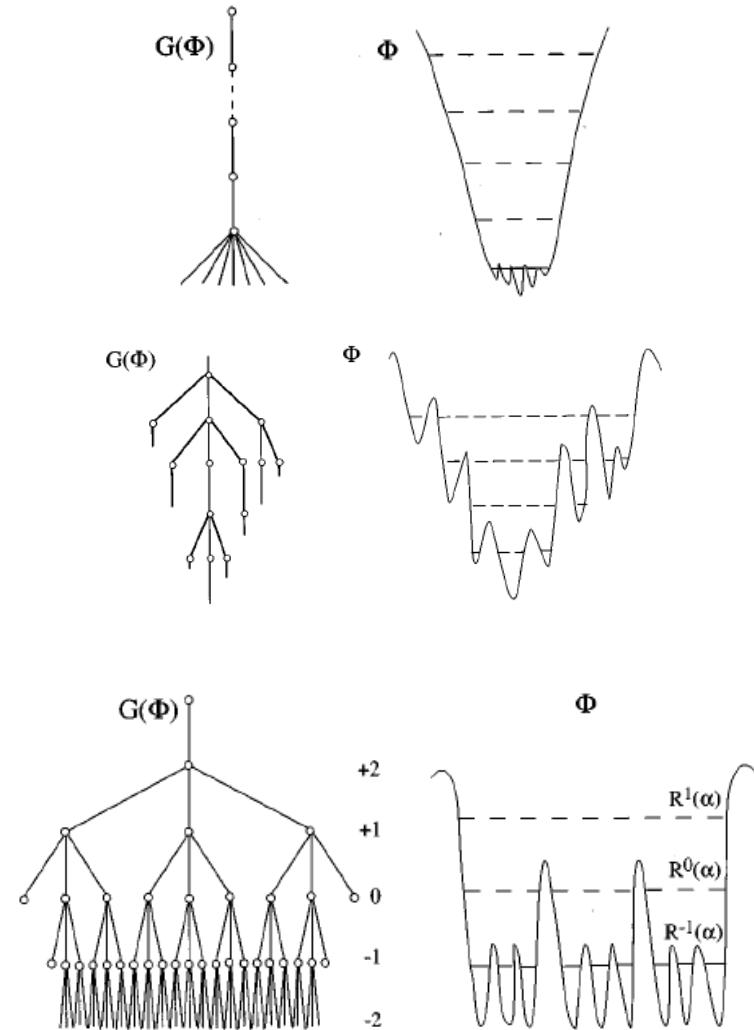
Visualizing the PES using disconnectivity graphs

A connected database of local minima can be partitioned into disjoint sets, or ‘superbasins’, at any given total energy, E , where the members of each set can all be interconverted by one or more rearrangements without exceeding E .

This basin analysis is performed at a series of energies, $E_1 < E_2 < E_3 < \dots$, and each distinct superbasin is represented as a point, or node, on the horizontal axis.

The vertical axis corresponds to the potential energy, and lines are drawn upwards starting from the potential energy of each local minimum, which is joined to the node for the superbasin in which it lies at the next energy level.

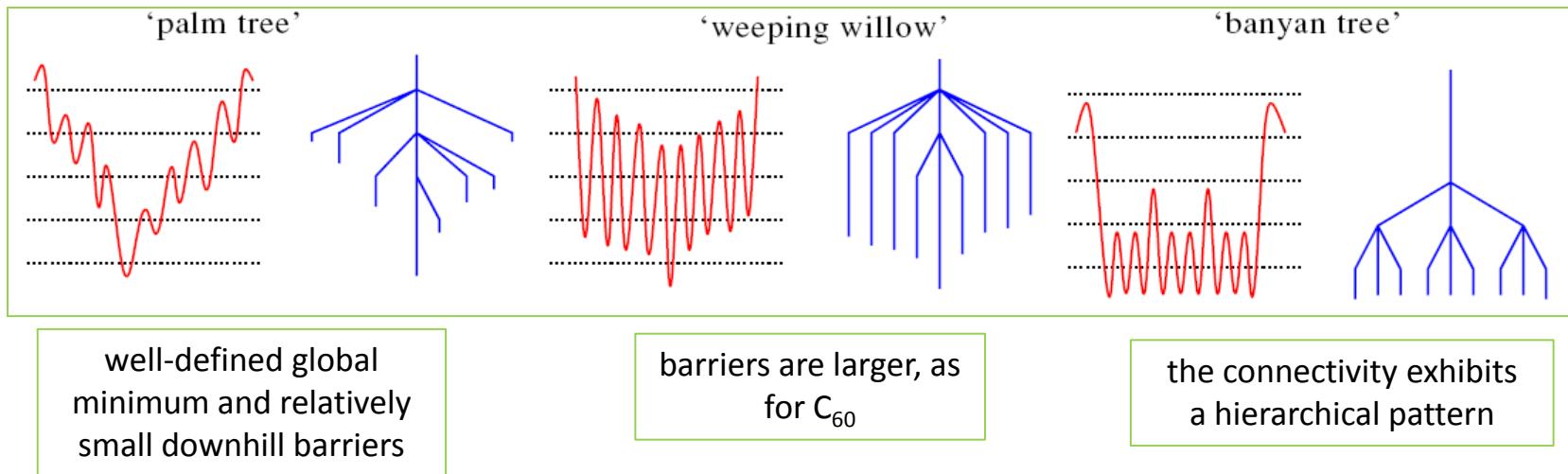
Lines are drawn between nodes at adjacent levels if they correspond to the same superbasin or to superbasins that merge together.



Visualizing the PES using disconnectivity graphs

Disconnectivity graphs are connected, but contain no cycles.

They are classified as tree graphs.



Hierarchy arises because local minima with similar energies are separated by barriers on two different energy scales.

Qualitative behaviour of important observables may often be deduced from the form of the graph.

Palm tree motif show efficient relaxation to global minimum over a wide range of temperature.

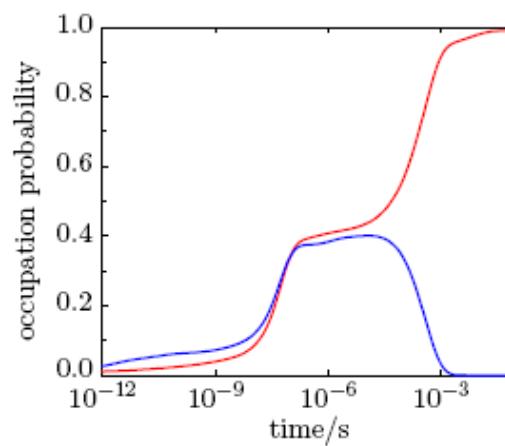
Such graphs can be viewed as a set of kinetically convergent pathways (a potential energy funnel).

The disconnectivity graph for ala₈

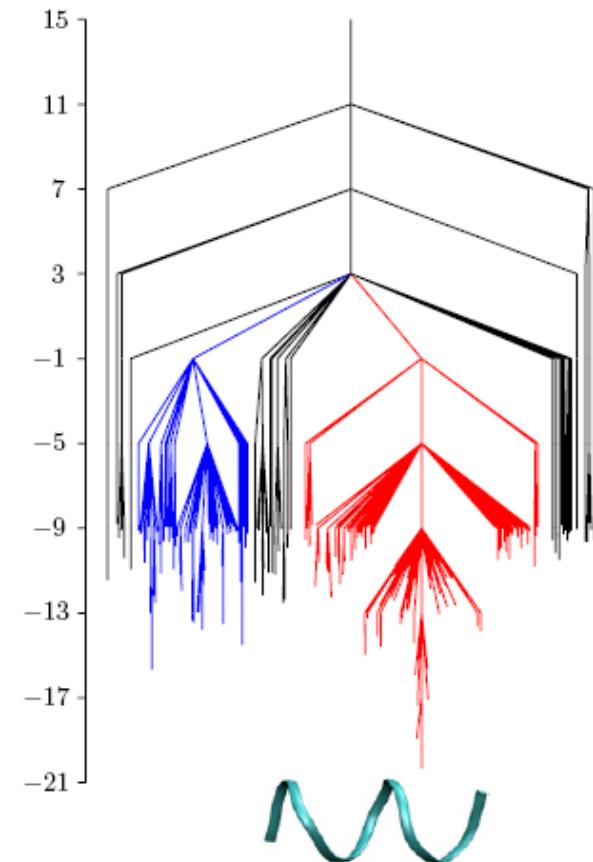
Two competing funnels result in a separation of relaxation time scales

The energy is in kcal mol⁻¹.

Branches leading to minima with α and β character are coloured red and blue.

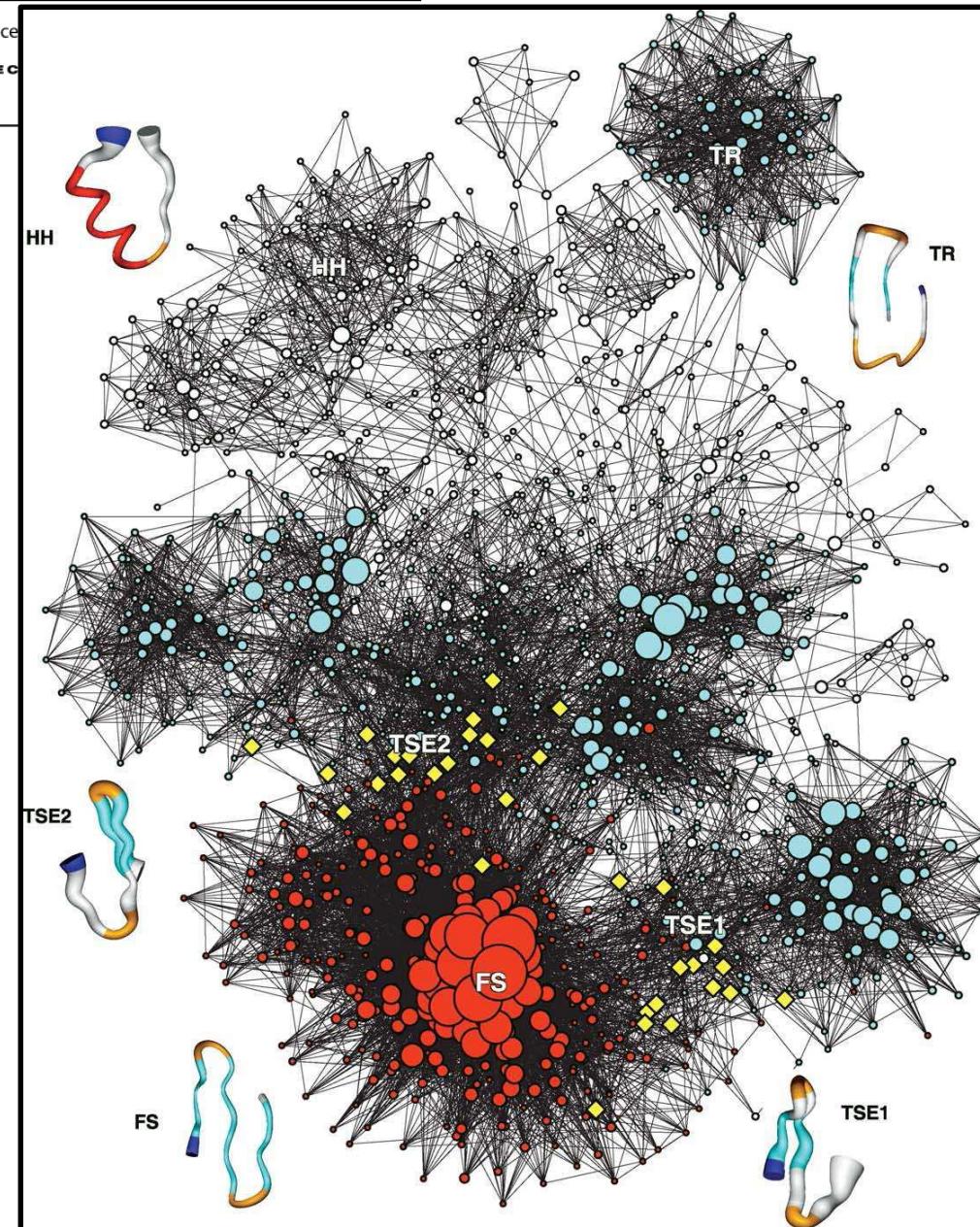


Occupation probability is plotted as a function of time starting from a high temperature equilibrium distribution.



The Protein Folding Network

Francesco Rao and Amedeo Caflisch*



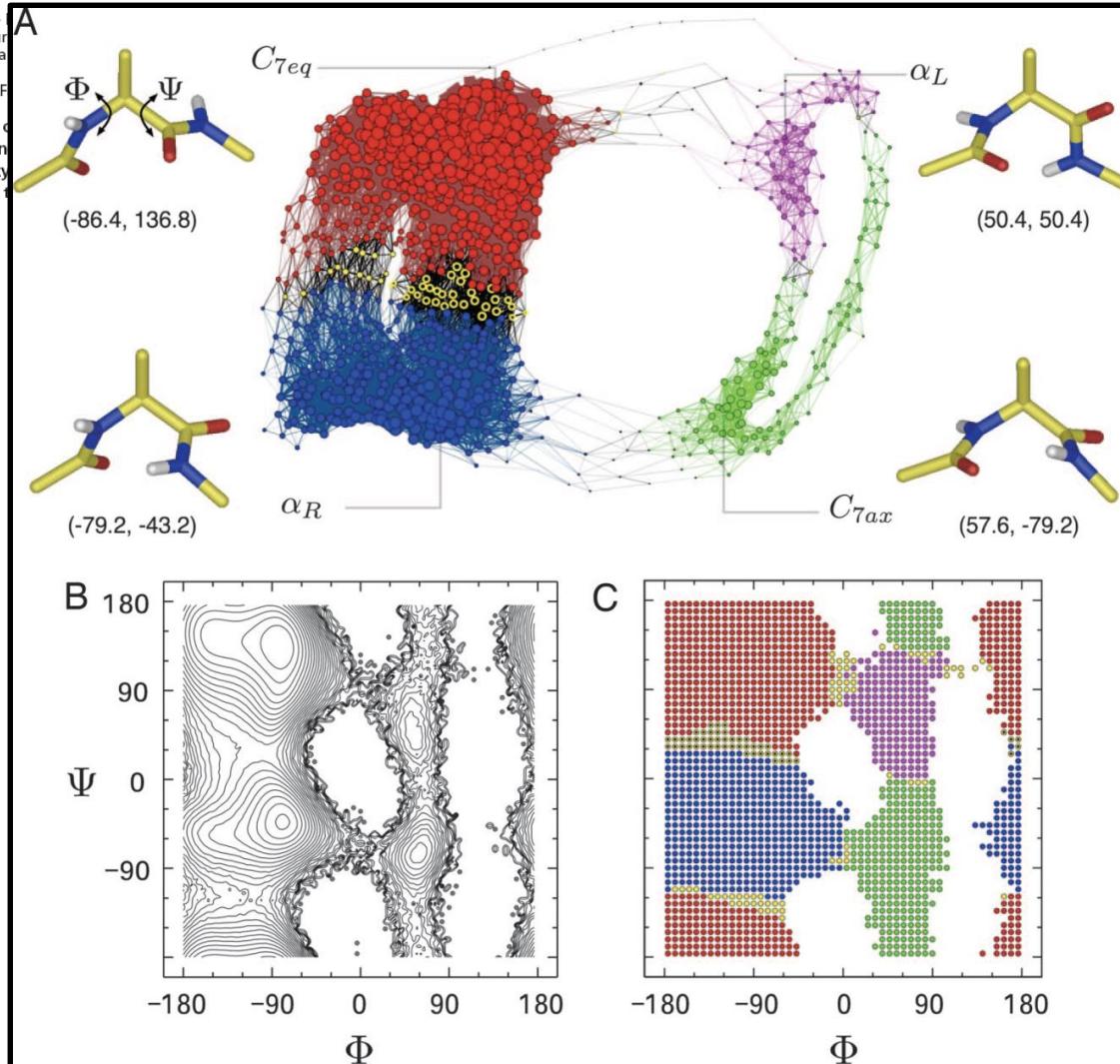
Complex network analysis of free-energy landscapes

D. Gfeller[†], P. De Los Rios[†], A. Caflisch^{§¶||††}, and F. Rao^{§¶||††}

[†]Laboratoire de
Université de Zurich
Fisica, Universita

Edited by Hans P.

The kinetics of protein folding dimensionality into one or



References

