

# Structural Bioinformatics

Elodie Laine

Master BIM-BMC Semester 3, 2022-2023

Laboratoire de Biologie Computationnelle et Quantitative (LCQB)  
e-documents: <http://www.lcqb.upmc.fr/laine/STRUCT>  
e-mail: [elodie.laine@sorbonne-universite.fr](mailto:elodie.laine@sorbonne-universite.fr)



@LaineElodie

# Tertiary Structure and Dynamics

# The thermodynamic hypothesis (Anfinsen's Dogma)



## The Nobel Prize in Chemistry 1972

"for his work on ribonuclease, especially concerning the connection between the amino acid sequence and the biologically active conformation"

"for their contribution to the understanding of the connection between chemical structure and catalytic activity of the active centre of the ribonuclease molecule"



**Christian B.  
Anfinsen**

1/2 of the prize  
USA



**Stanford Moore**

1/4 of the prize  
USA



**William H. Stein**

1/4 of the prize  
USA

National Institutes  
of Health  
Bethesda, MD, USA

b. 1916  
d. 1995

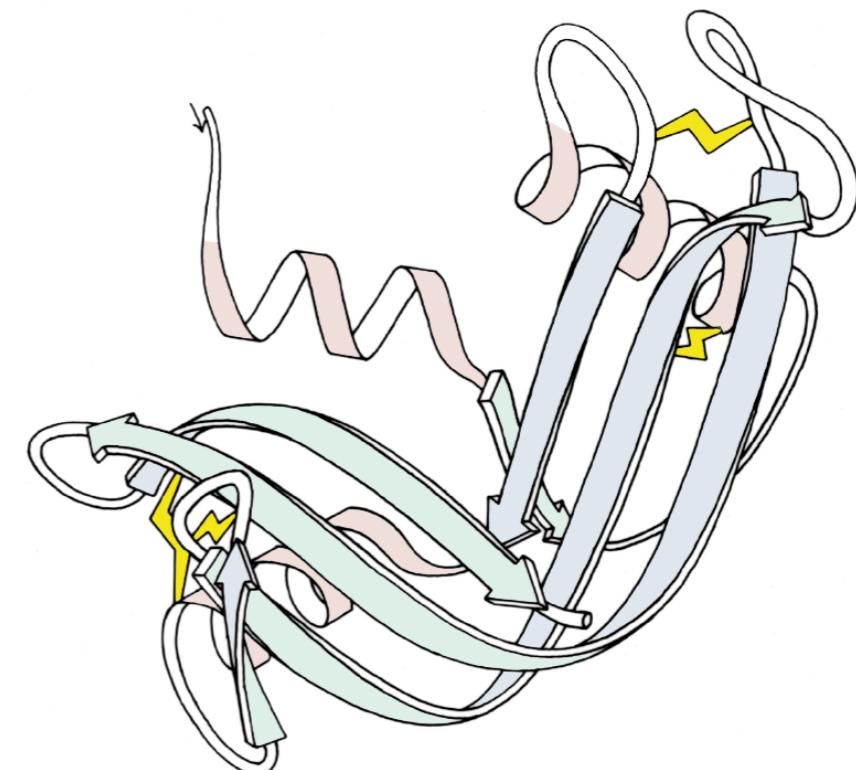
Rockefeller  
University  
New York, NY, USA

b. 1913  
d. 1982

Rockefeller  
University  
New York, NY, USA

b. 1911  
d. 1980

The native structure of a protein corresponds to minimum of free energy

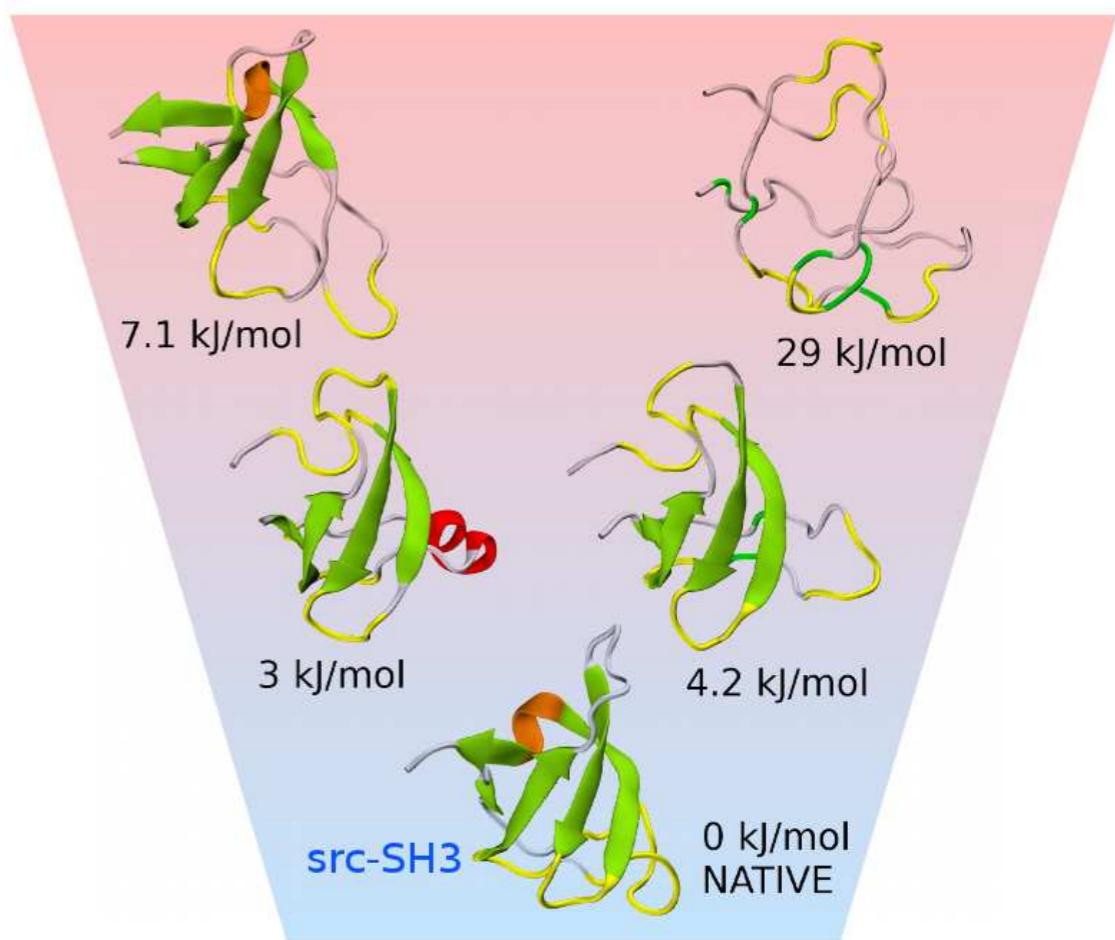


Ribonuclease A

The **free energy difference** between the native state and the ensemble of denatured conformations is 20 à 60 kJ/mol (5-15 kcal/mol).

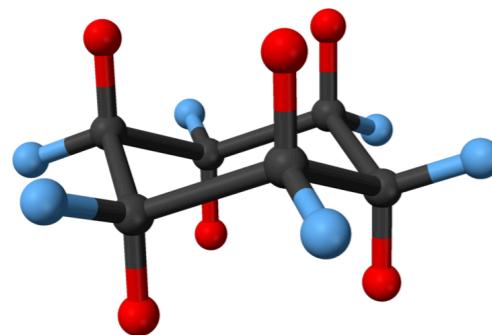
# The thermodynamic hypothesis (Anfinsen's Dogma)

- ◆ **Uniqueness:** the sequence does not have any other configuration with a comparable free energy.
- ◆ **Stability:** small changes in the surrounding environment cannot give rise to changes in the minimum configuration.
- ◆ **Kinetic accessibility:** the folding of the chain must not involve highly complex changes in the shape (like knots or other high order conformations).

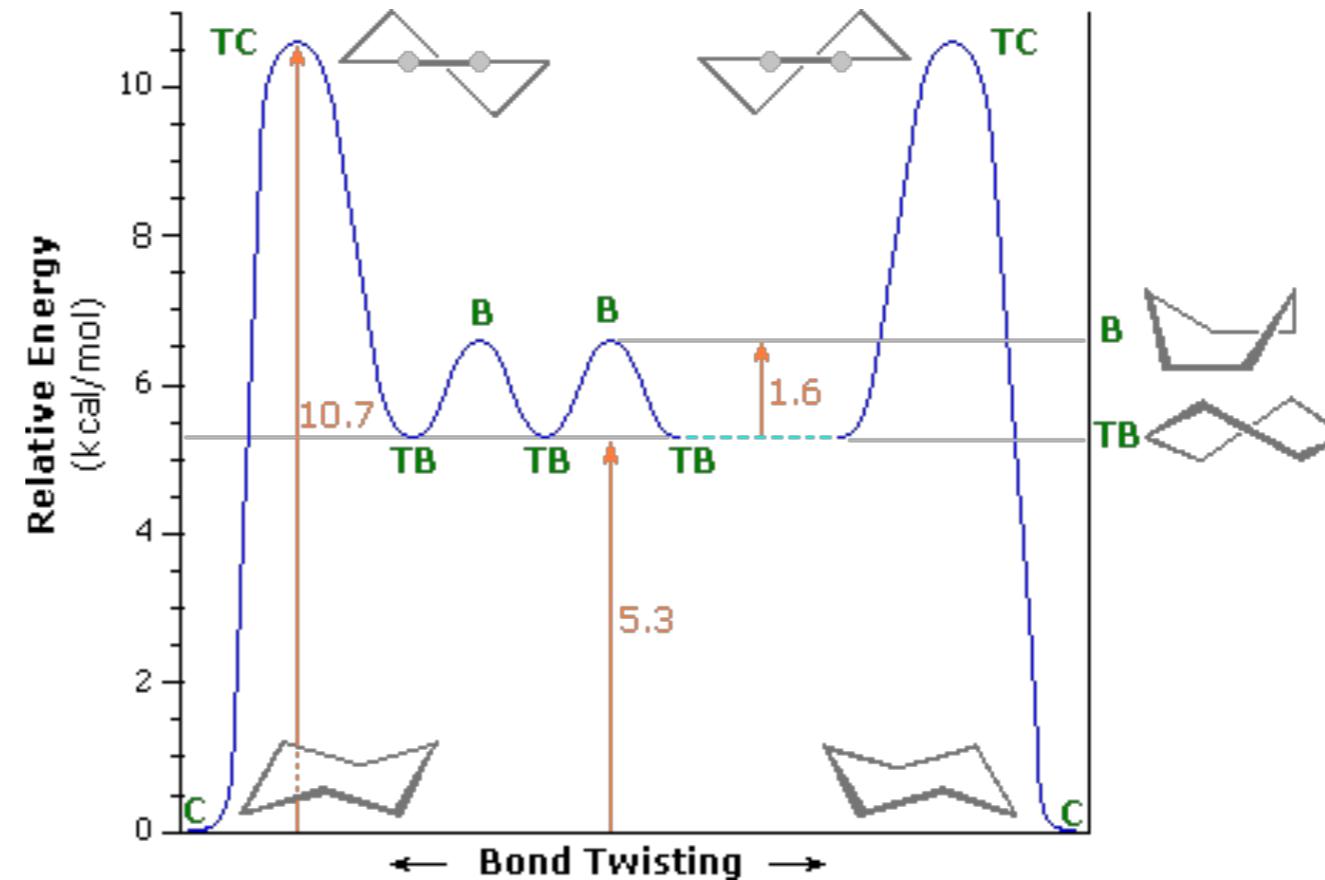


# Energy landscape

The energy landscape is a **multidimensional hypersurface** obtained by mapping all possible conformations of a system & their corresponding (free) energy levels.



cyclohexane

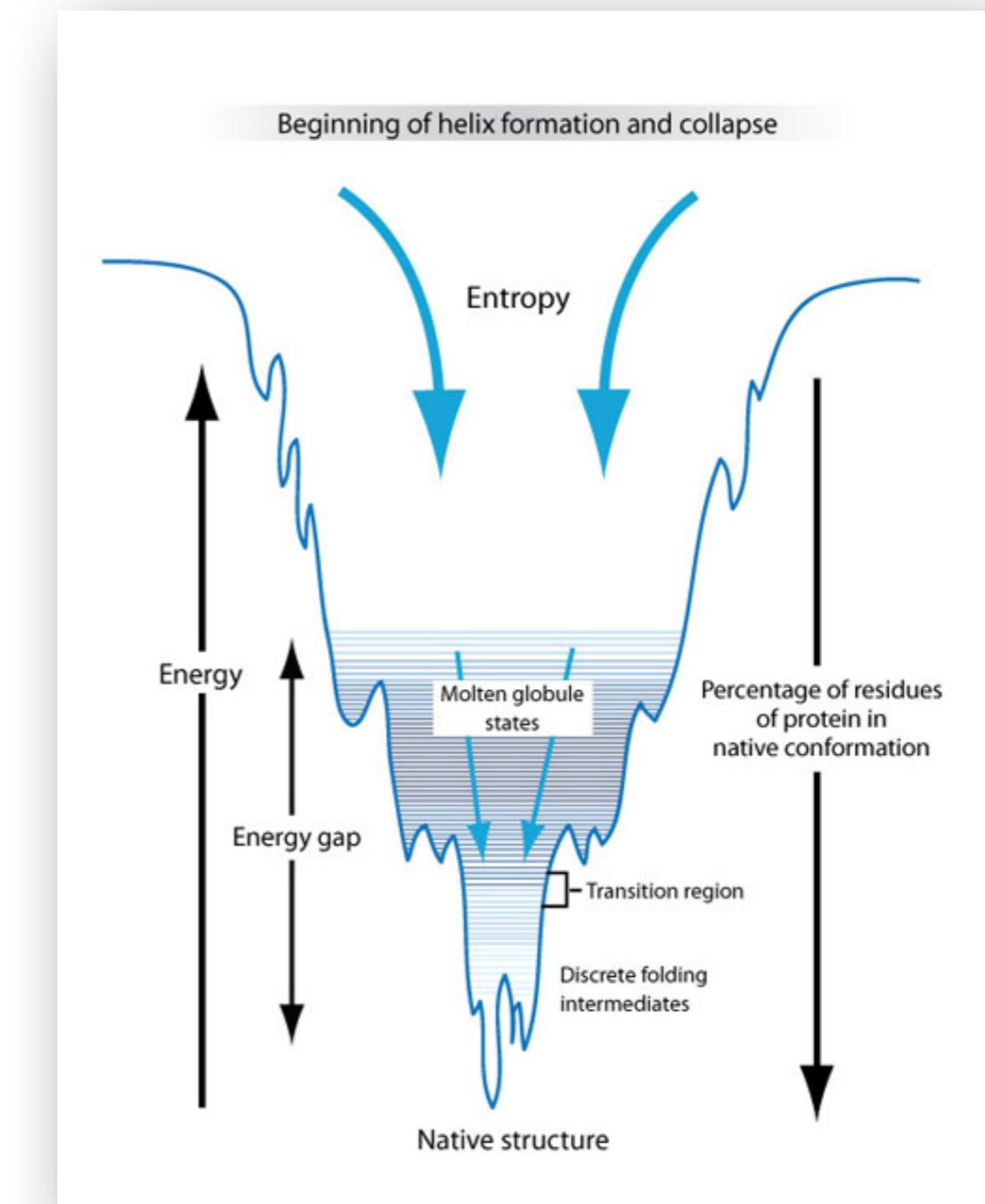


For a system of  $N$  atoms, the number of degrees of freedom is  $3N$ . Generally one represents only a section of the hypersurface along a reaction coordinate.

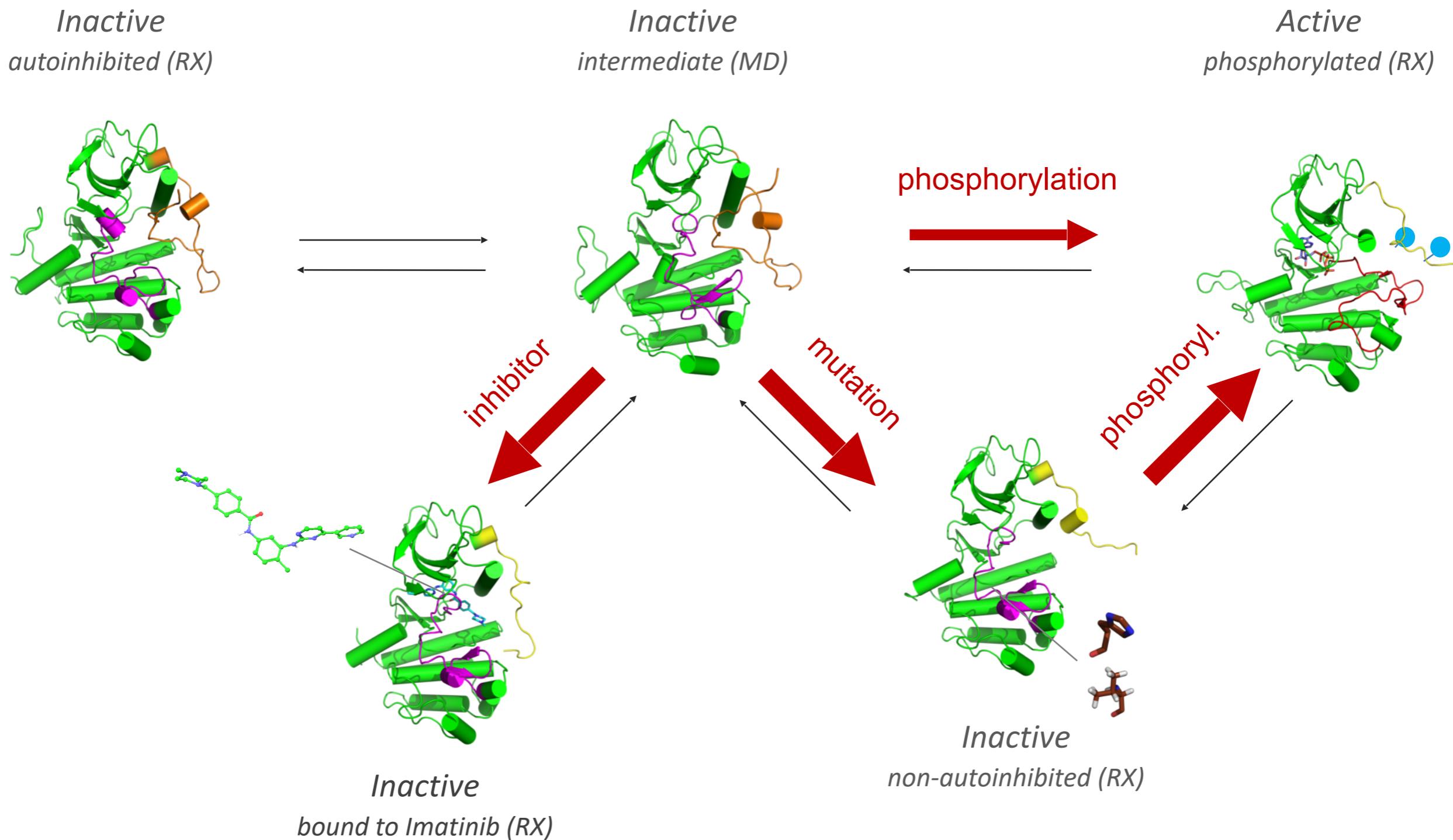
# Energy landscape

The energy landscape is a **funneled rugged hypersurface**:

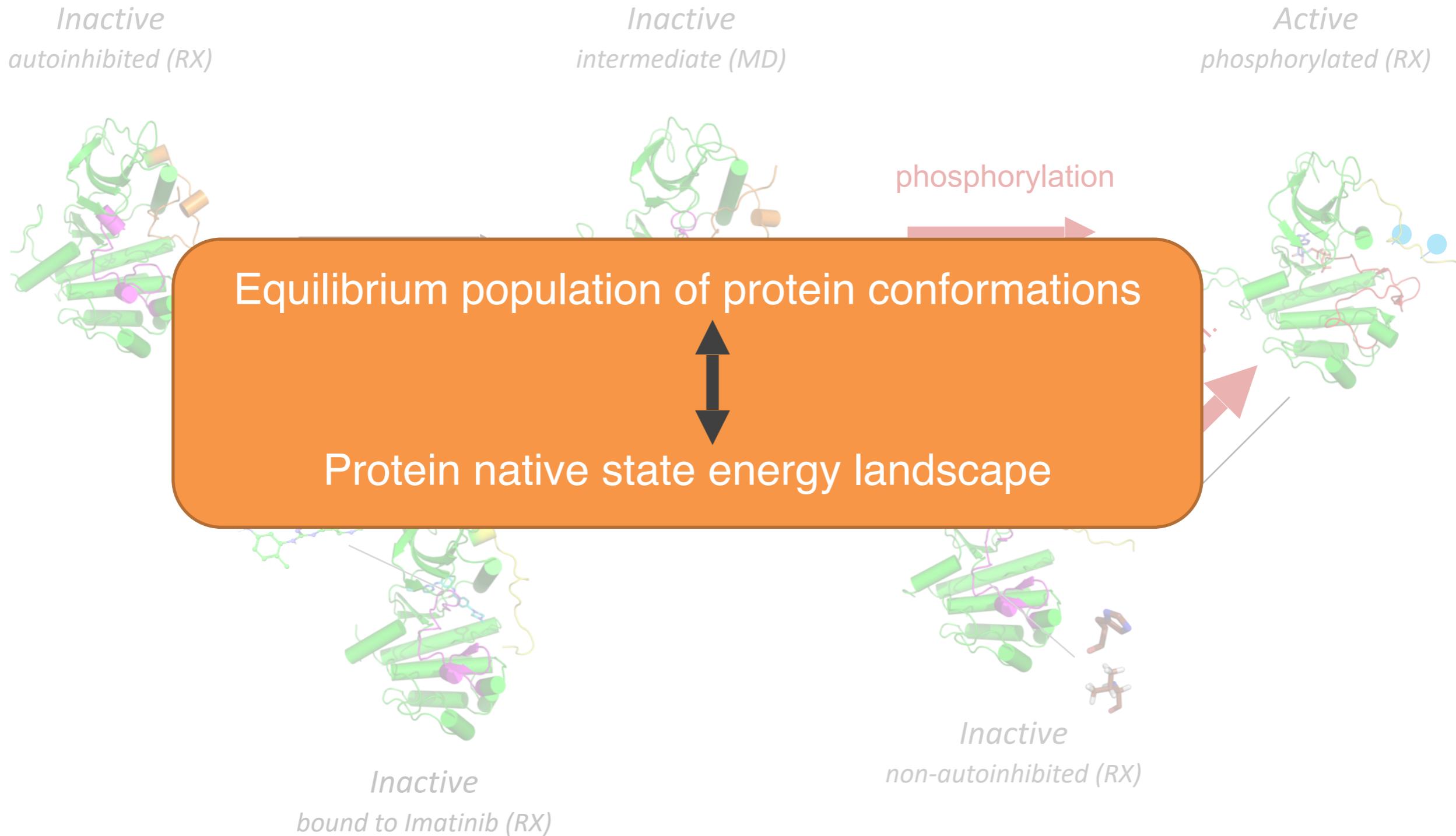
- many non-native local minima
- deep free energy minimum with steep walls.



# Equilibrium of populations



# Equilibrium of populations



# Free energy

The protein in solution is viewed as a **statistical ensemble**.

$$\Delta G = \Delta H - T\Delta S$$

# Free energy

The protein in solution is viewed as a **statistical ensemble**.

$$\Delta G = \Delta H - T\Delta S$$

Gibbs or Helmholtz  
free energy

Energy available  
for  
thermodynamic  
work

# Free energy

The protein in solution is viewed as a **statistical ensemble**.

$$\Delta G = \Delta H - T\Delta S$$

Gibbs or Helmholtz  
free energy

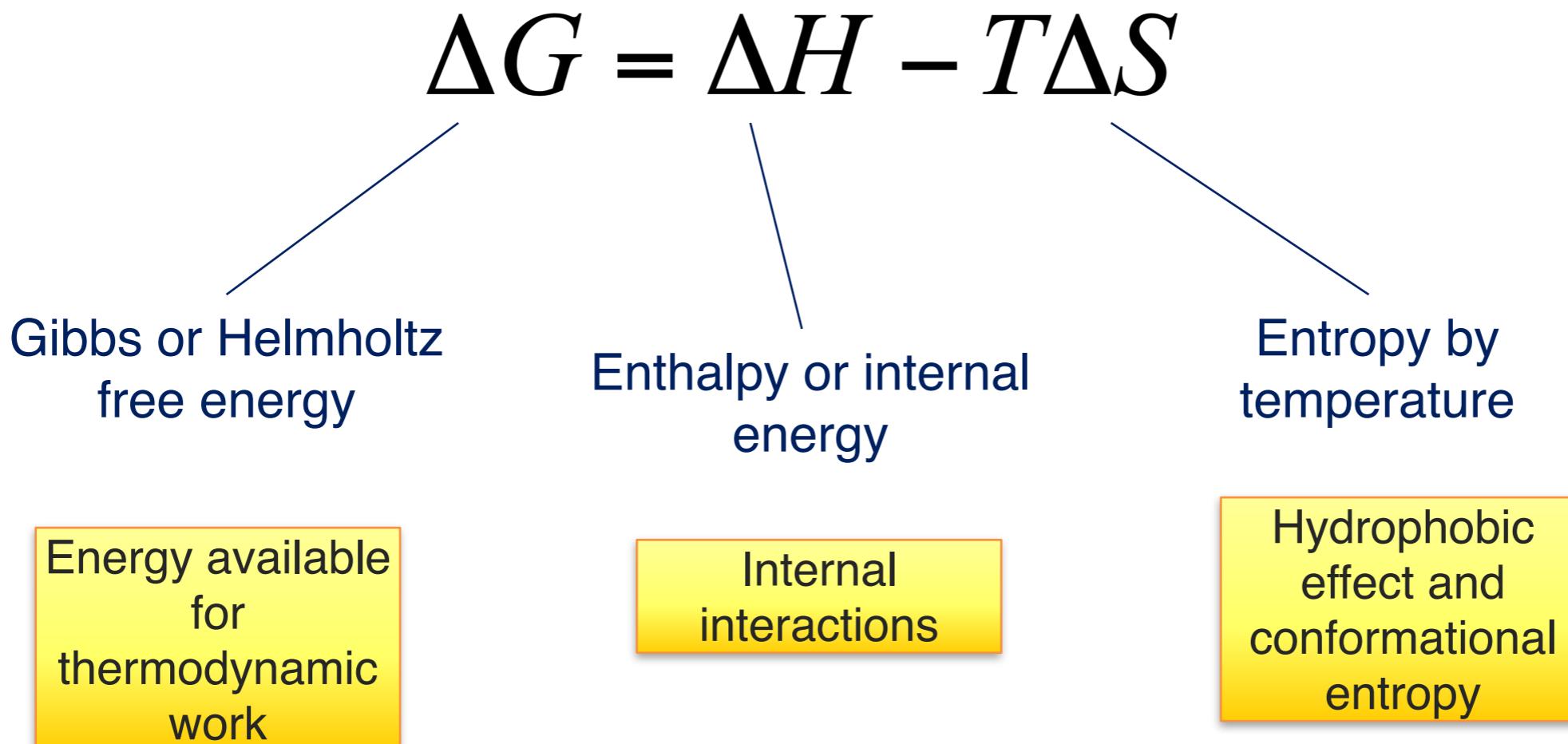
Enthalpy or internal  
energy

Energy available  
for  
thermodynamic  
work

Internal  
interactions

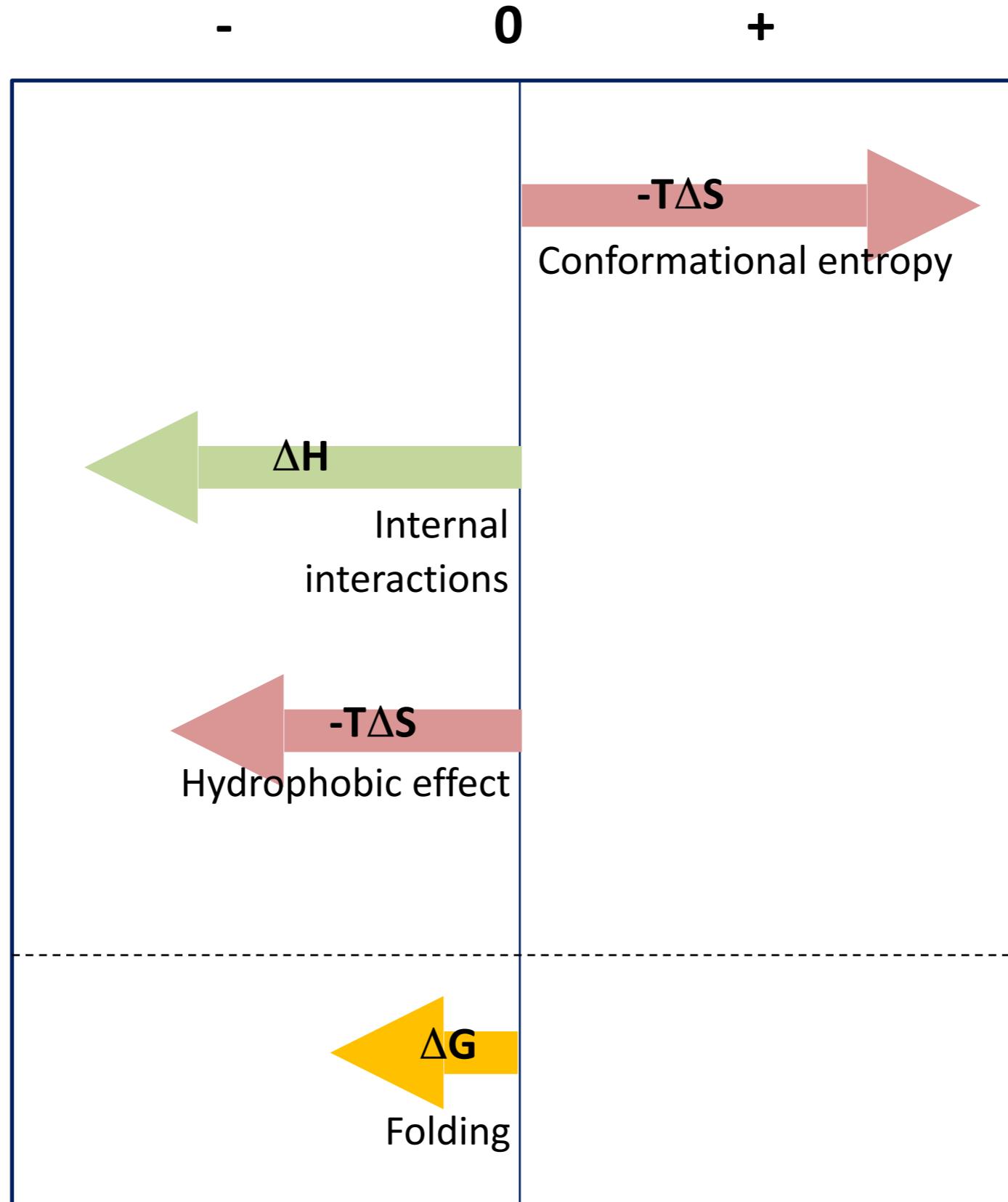
# Free energy

The protein in solution is viewed as a **statistical ensemble**.



# Free energy

Favorable free energy of folding is a net result of thermodynamic forces



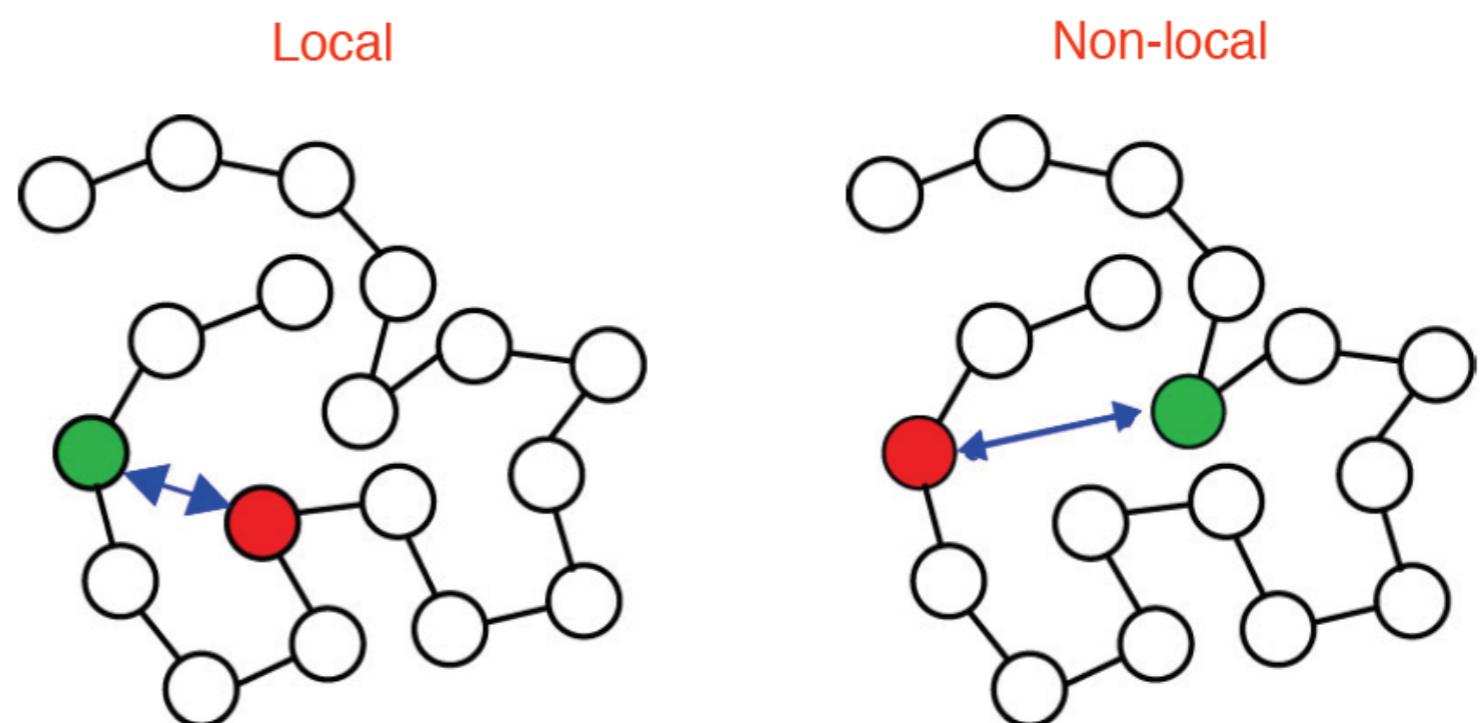
# Interatomic interactions

Amino acids of a protein are joined by **covalent bonding interactions** (primary structure).

The 3D fold is stabilized by **non-bonding interactions** (tertiary structure):

- Electrostatic Interactions (5 kcal/mol)
- Hydrogen-bond Interactions (3-7 kcal/mol)
- Van Der Waals Interactions (1 kcal/mol)
- Hydrophobic Interactions (< 10 kcal/mol)

Interactions are treated differently depending on the number of amino acids along the sequence separating the interacting pair.



# Modeling approach

## Choose a representation of the object

resolution, forces, heterogeneity, extra data

*type of data, purpose, feasibility*

*Atoms or coarser grains, average (rigid) or ensemble (flexible)...*

## Choose a scoring function

quantifies the match between the model and the input data, to be optimized

*Weighted sum of spatial restraints, bayesian posterior...*

## Choose a search strategy

the goal is to explore as much as possible, given computational limitations

*Systematic or biased, stochastic (MC)  
...*

## Filter and/or refine

based on input data, allowing a more expensive representation/search on a subset

*Flexible refinement, 2D projection  
class matching...*

# Modeling approach

Choose a representation of the object  
resolution, forces, heterogeneity, extra data  
*type of data, purpose, feasibility*

*Atoms or coarser grains, average  
(rigid) or ensemble (flexible)...*

Choose a scoring function  
quantifies the match between the model and the  
input data, to be optimized

*Weighted sum of spatial restraints,  
bayesian posterior...*

Choose a search strategy  
the goal is to explore as much as possible, given  
computational limitations

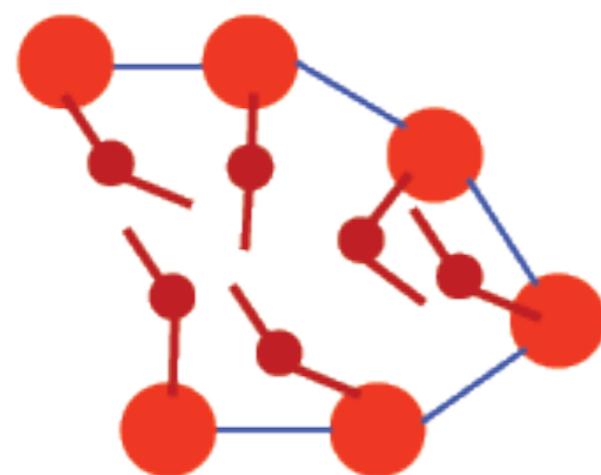
*Systematic or biased, stochastic (MC)  
...*

Filter and/or refine  
based on input data, allowing a more expensive  
representation/search on a subset

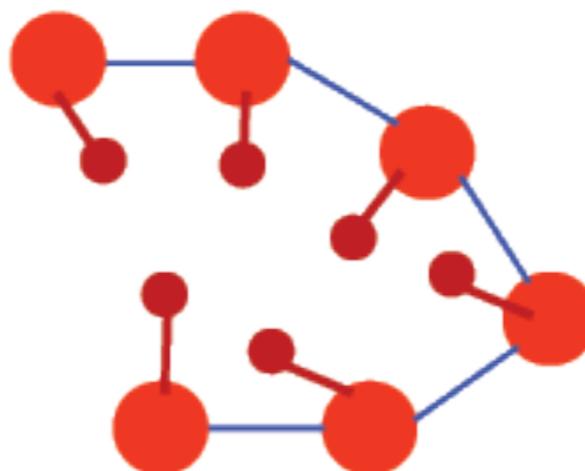
*Flexible refinement, 2D projection  
class matching...*

# Structural representation

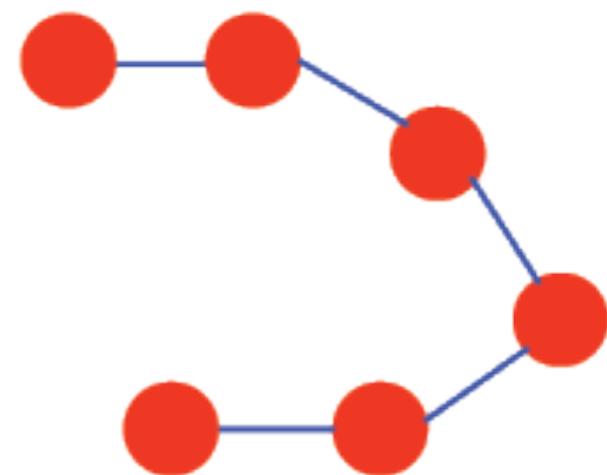
All-atom level



Simplified side-chains



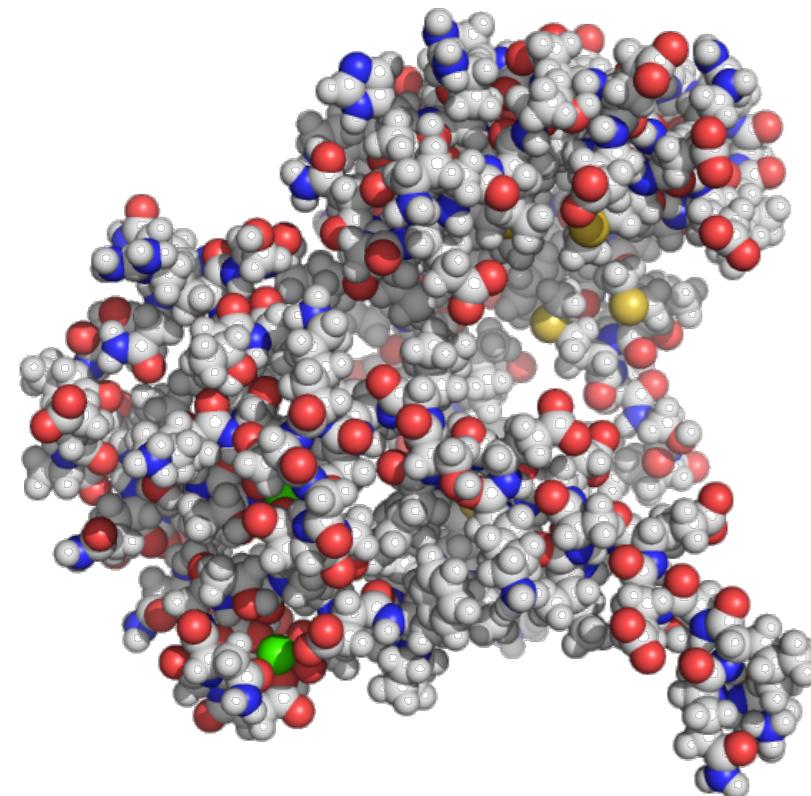
Omitted side-chains



SPEED

PRECISION

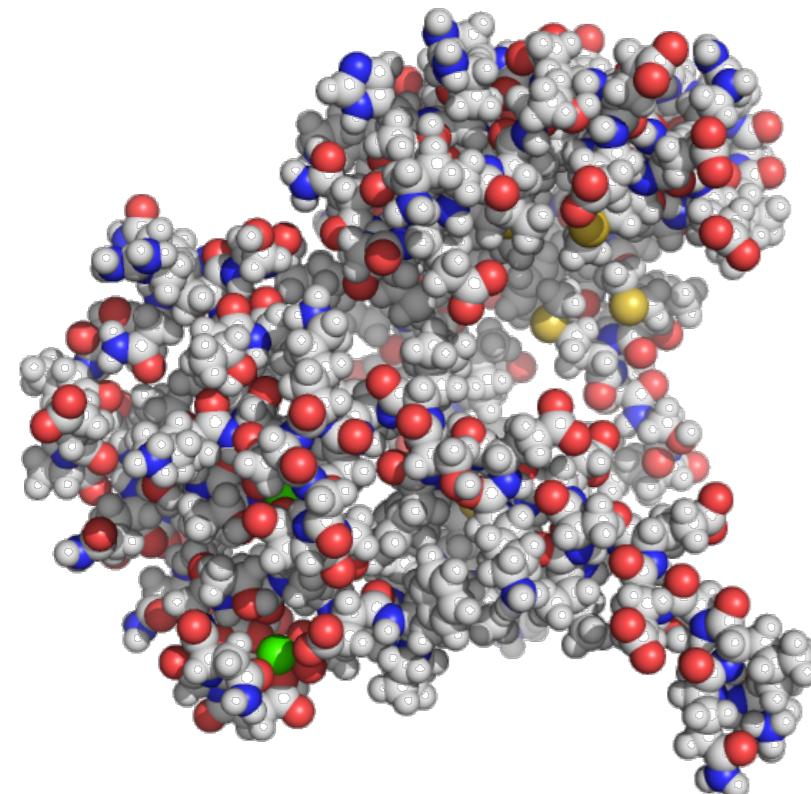
# Structural representation



- oxygen
- nitrogen
- carbon

All-atom model

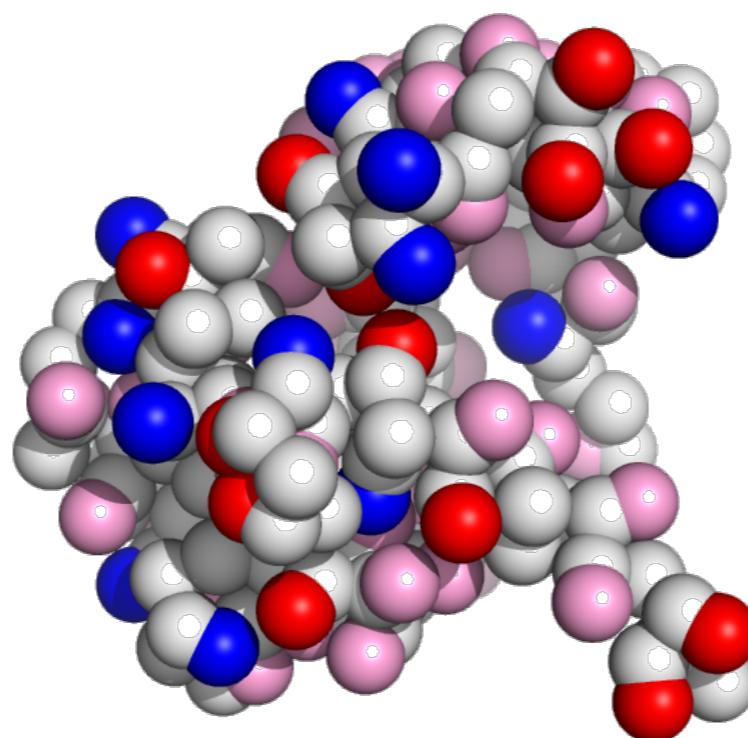
# Structural representation



All-atom model

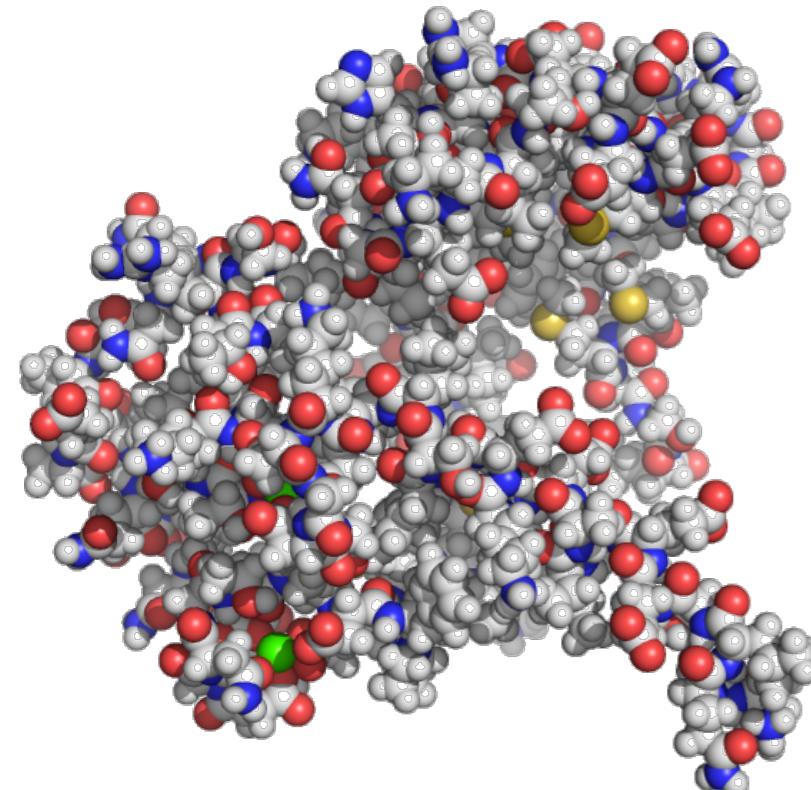
- oxygen
- nitrogen
- carbon

Coarse-grained model  
(pseudo-atoms)



- pseudo -
- pseudo +
- pseudo hydrophobic
- pseudo polar

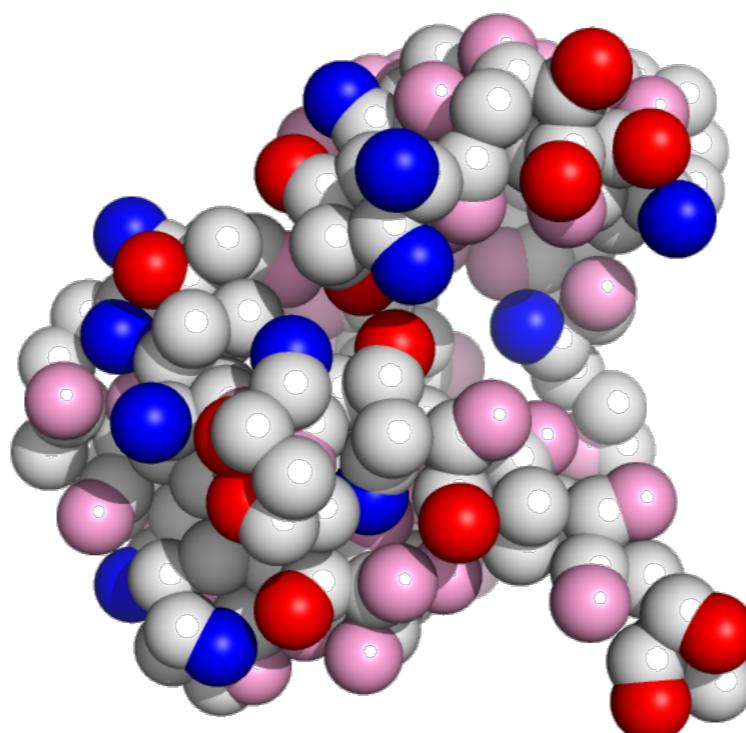
# Structural representation



All-atom model

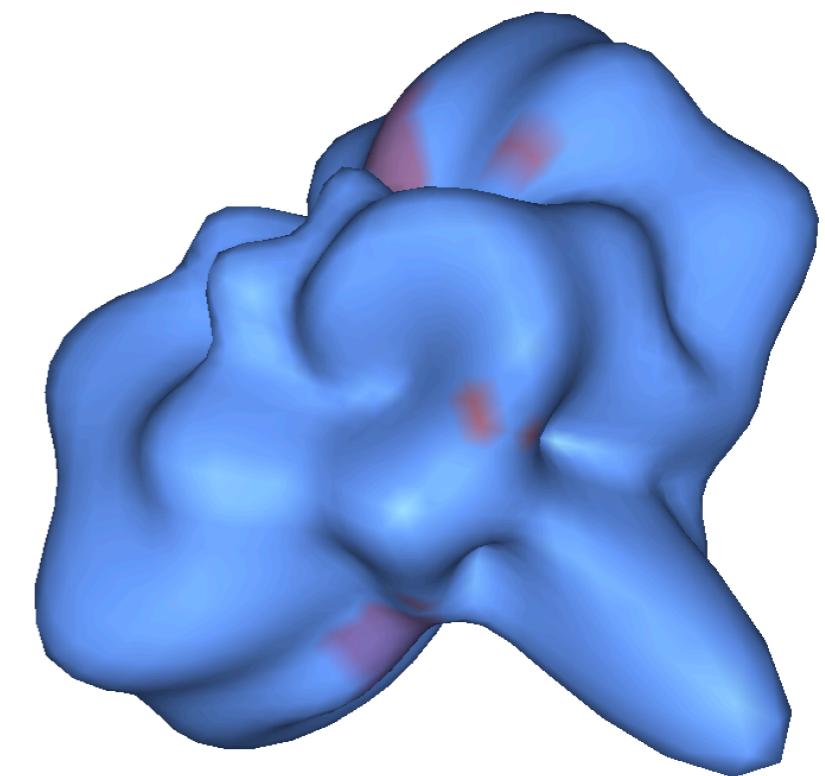
- oxygen
- nitrogen
- carbon

Coarse-grained model  
(pseudo-atoms)



- pseudo -
- pseudo +
- pseudo hydrophobic
- pseudo polar

Spherical harmonics



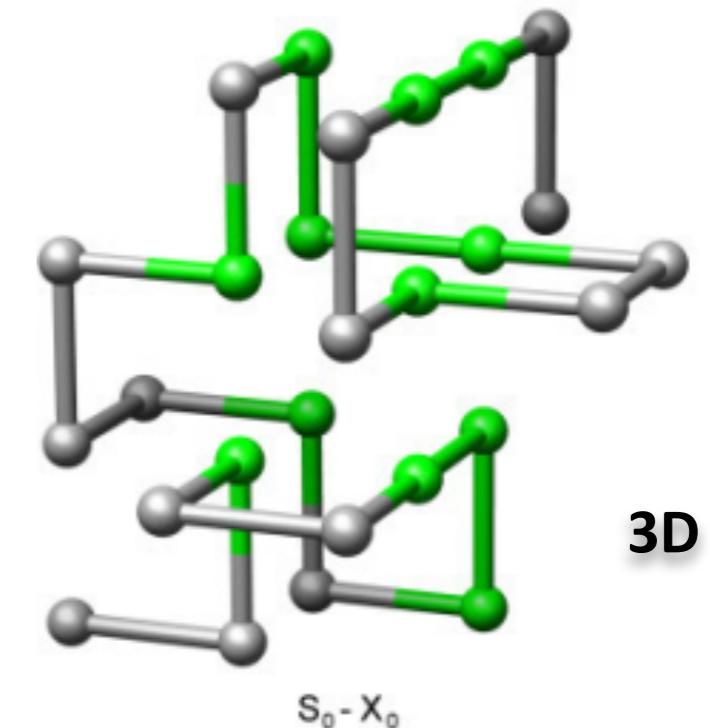
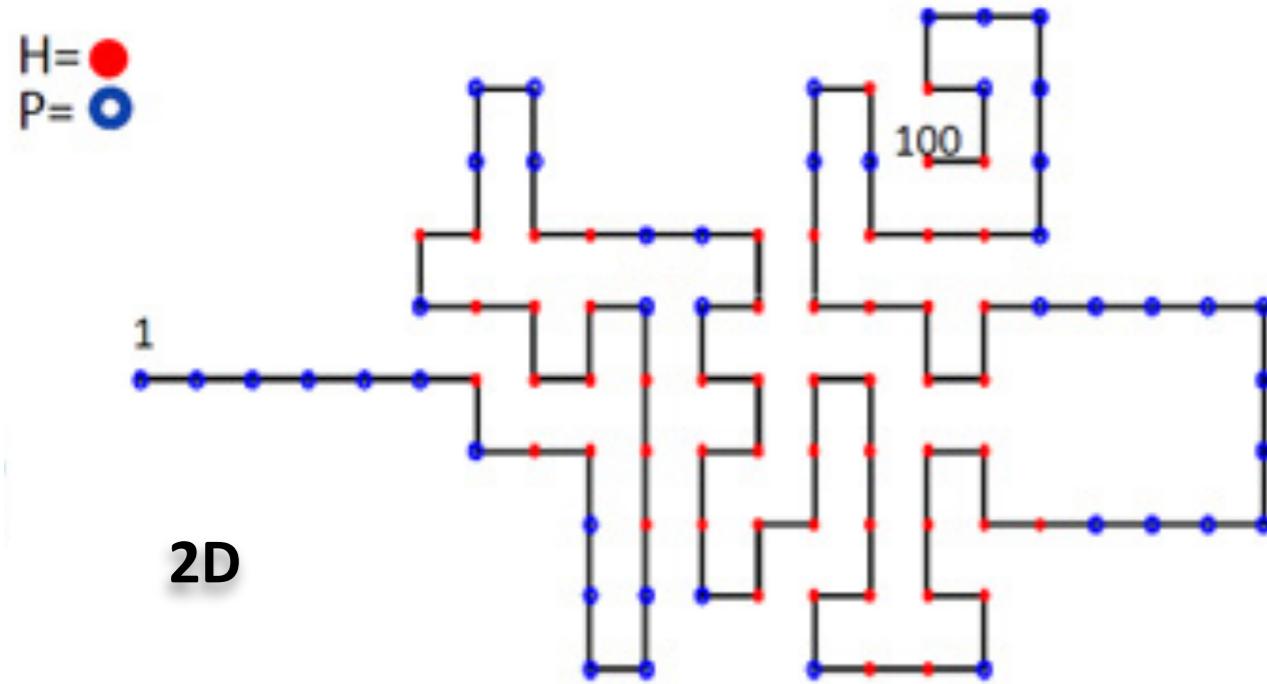
# Structural representation

Conformations can be represented by:

- **Cartesian coordinates**: the constraints of the polypeptide chain must be explicitly addressed (bond length, bond angle, torsion...)
- **Internal coordinates** representing torsion angles  $\phi$ ,  $\psi$ ,  $\omega$ , the bond angles and lengths being fixed: a small change induces a global adjustment of the downstream residues (steric clashes)
- **Interatomic or interresidue distances**: collective variables that reduce the number of dimensions and must be carefully chosen
- **Regular lattices** (cubic, tetrahedral...): almost impossible to represent a real protein structure unless the grid spacing is very small

# Lattices

## Regular lattice



Poor realism, do not enable to represent helices, although compensated motions are permitted.

## Irregular lattice

Enable to represent perfectly (maybe too much) secondary structure elements, but not compensated motions.

# Modeling approach

Choose a representation of the object  
resolution, forces, heterogeneity, extra data  
*type of data, purpose, feasibility*

*Atoms or coarser grains, average  
(rigid) or ensemble (flexible)...*

Choose a scoring function  
quantifies the match between the model and the  
input data, to be optimized

*Weighted sum of spatial restraints,  
bayesian posterior...*

**Choose a search strategy**  
**the goal is to explore as much as possible, given  
computational limitations**

*Systematic or biased, stochastic (MC)  
...*

Filter and/or refine  
based on input data, allowing a more expensive  
representation/search on a subset

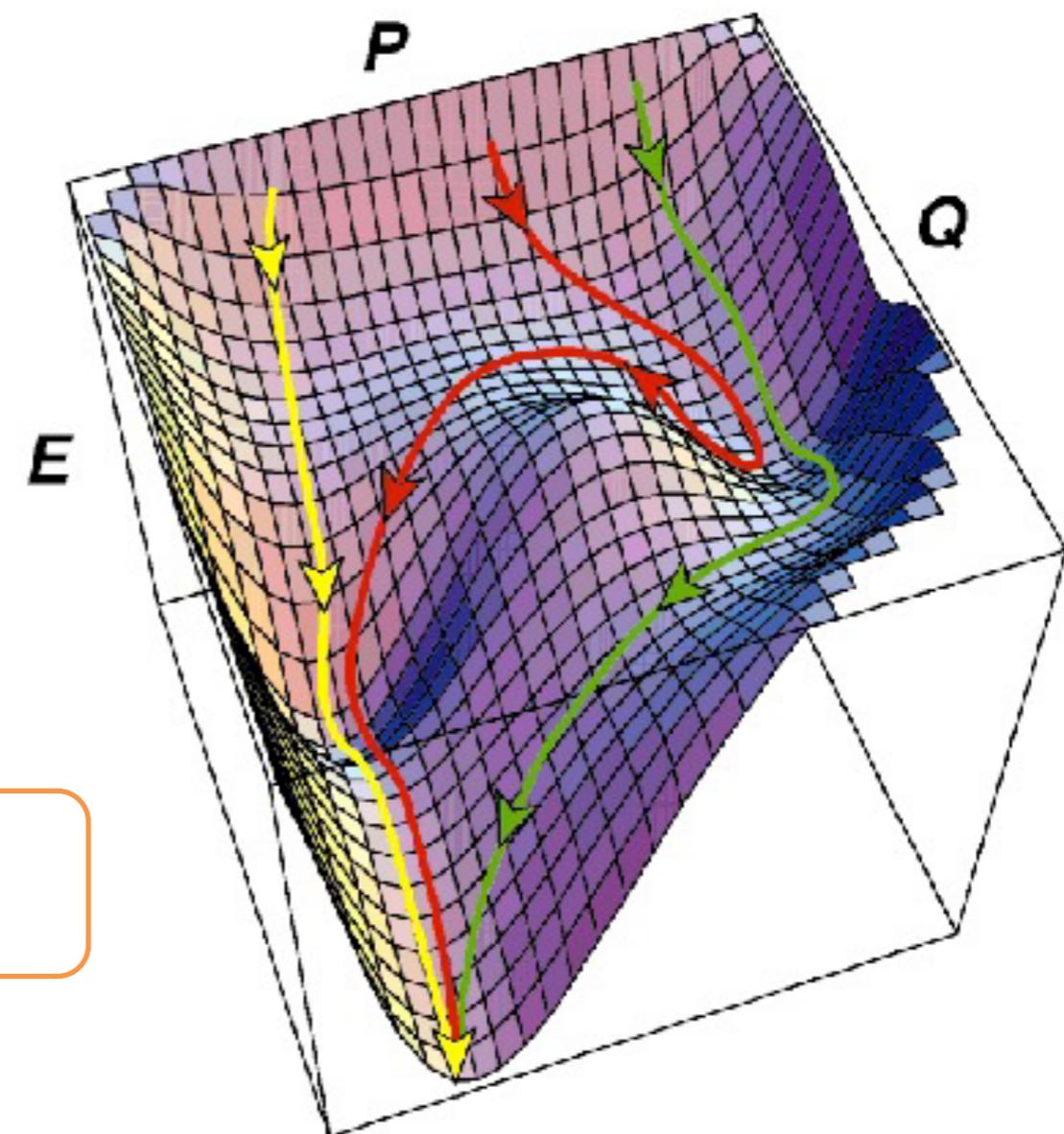
*Flexible refinement, 2D projection  
class matching...*

# *Ab initio* prediction principles

***Ab initio* modelling** consists in searching through the conformational space of the protein. As in the previously described approaches, one assumes that all information necessary to the determination of the protein fold is encoded in its sequence.

- **Explore** the protein conformational space
- Search for the **minimum** of a scoring function

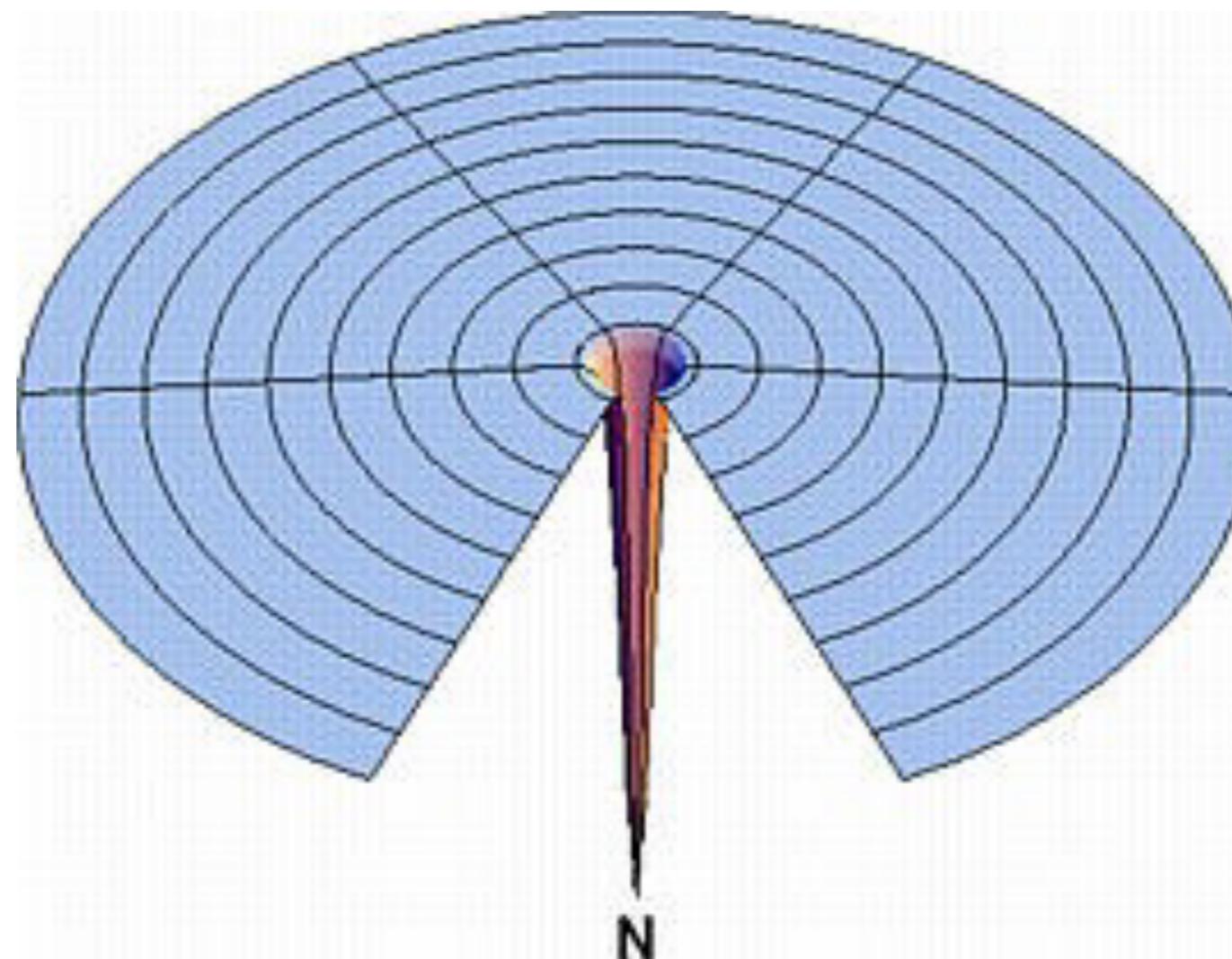
The scoring function corresponds generally to the **interatomic interaction energy**.



# Systematic search

Assumes equal probabilities for non-native states. The conformational space is **systematically explored** by regular changes of the degrees of freedom values.

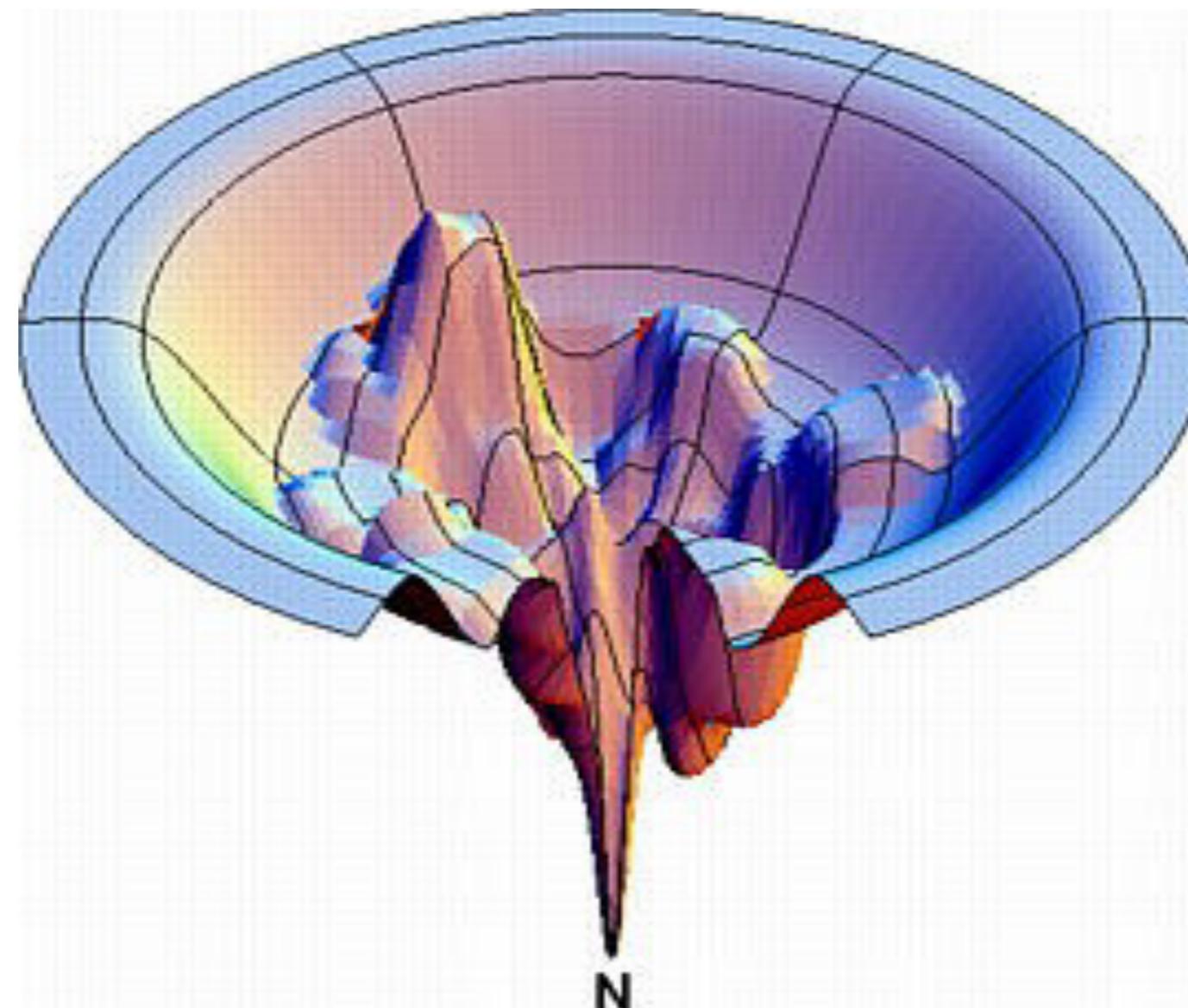
golf-course model  
of the energy  
landscape



# Systematic search

Assumes equal probabilities for non-native states. The conformational space is **systematically explored** by regular changes of the degrees of freedom values.

Rugged funneled view of the energy landscape



# Systematic search

## Example

search for the optimal combination of  $\phi$  et  $\psi$  torsion angle values

- bond lengths and angles remain fixed
- $\omega$  is set to  $180^\circ$  for all residues
- side-chains are omitted

$\phi$  et  $\psi$  are varied from 0 to  $360^\circ$ , with an increment fixed to  $\theta$

**Combinatorial explosion:** the number of conformations being tested is:

$$\prod_{i=1,2N-2} 360/\theta$$

$\theta=30^\circ$

$N=2$ : 144 conformations

$N=3$ : 21,000 conformations

$N=5$ : 430,000,000 conformations

# Systematic search

## Example

search for the optimal combination of  $\phi$  et  $\psi$  torsion angle values

- bond lengths and angles remain fixed
- $\omega$  is set to  $180^\circ$  for all residues

$\phi$  et

Com

### **How to improve search efficiency and reduce computation time?**

- Quality check on partially built structures (structures displaying wrong or high energy substructures are deleted)
- Trade-off between the resolution of the search grid and computational time

$\theta=30^\circ$

N=2: 144 conformations

N=3: 21 000 conformations

N=5: 430 000 000 conformations

# Molecular dynamics

Interatomic forces are represented by an **analytical expression** called **force field**.

It represents **molecular mechanics** models of proteins containing:

- some chosen interactions
- a chosen functional that describes and links them

# Molecular dynamics

Interatomic forces are represented by an **analytical expression** called **force field**.

It represents **molecular mechanics** models of proteins containing:

- some chosen interactions
- a chosen functional that describes and links them

## Schrödinger



$$\frac{\partial}{\partial t} \Psi = \hat{H} \Psi$$

# Molecular dynamics

Interatomic forces are represented by an **analytical expression** called **force field**.

It represents **molecular mechanics** models of proteins containing:

- some chosen interactions
- a chosen functional that describes and links them

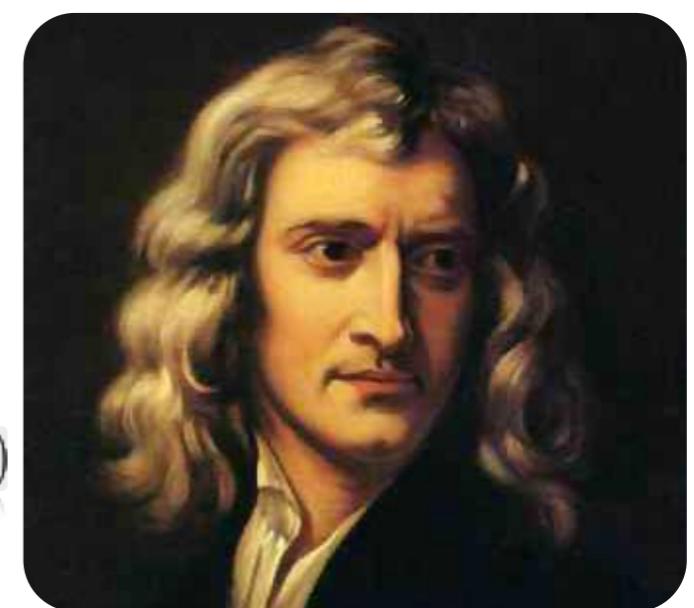
Schrödinger



$$i\hbar \frac{\partial}{\partial t} \Psi = \hat{H} \Psi$$

$$M\ddot{\mathbf{x}}(t) = F(\mathbf{x}(t)) = -\nabla V(\mathbf{x}(t))$$

Newton



# Molecular dynamics

Interatomic forces are represented by an **analytical expression** called **force field**.

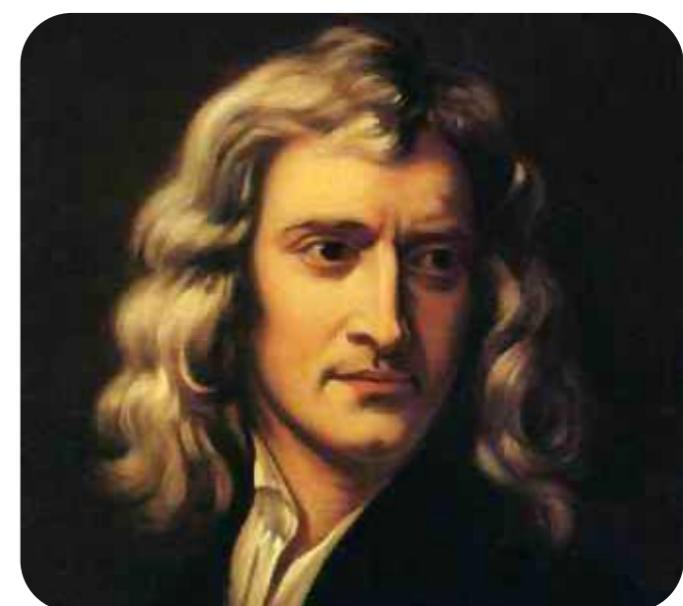
It represents **molecular mechanics** models of proteins containing:

- some chosen interactions
- a chosen functional that describes and links them

Schrödinger



Newton



Born-Oppenheimer

Additivity

Transferability

Empirical

# Molecular dynamics

Interatomic forces are represented by an **analytical expression** called **force field**.

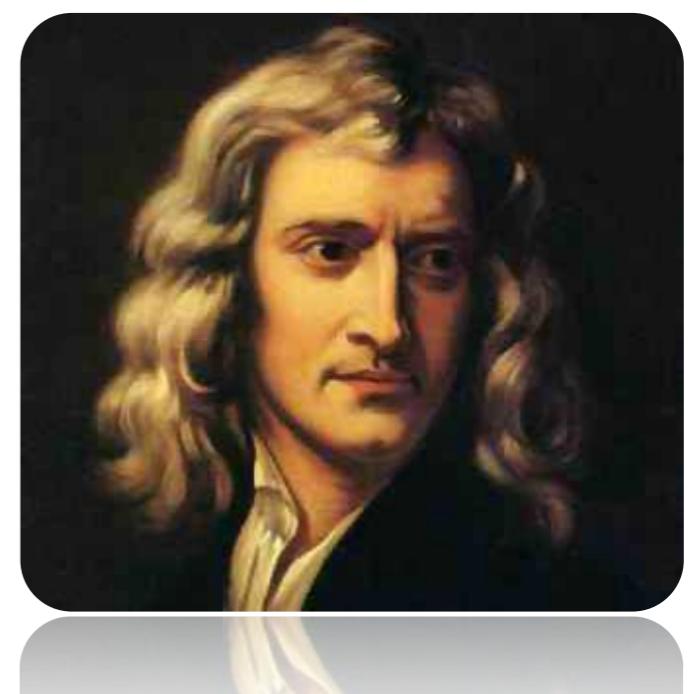
It represents **molecular mechanics** models of proteins containing:

- some chosen interactions
- a chosen functional that describes and links them

Schrödinger



Newton



Born-Oppenheimer

Additivity

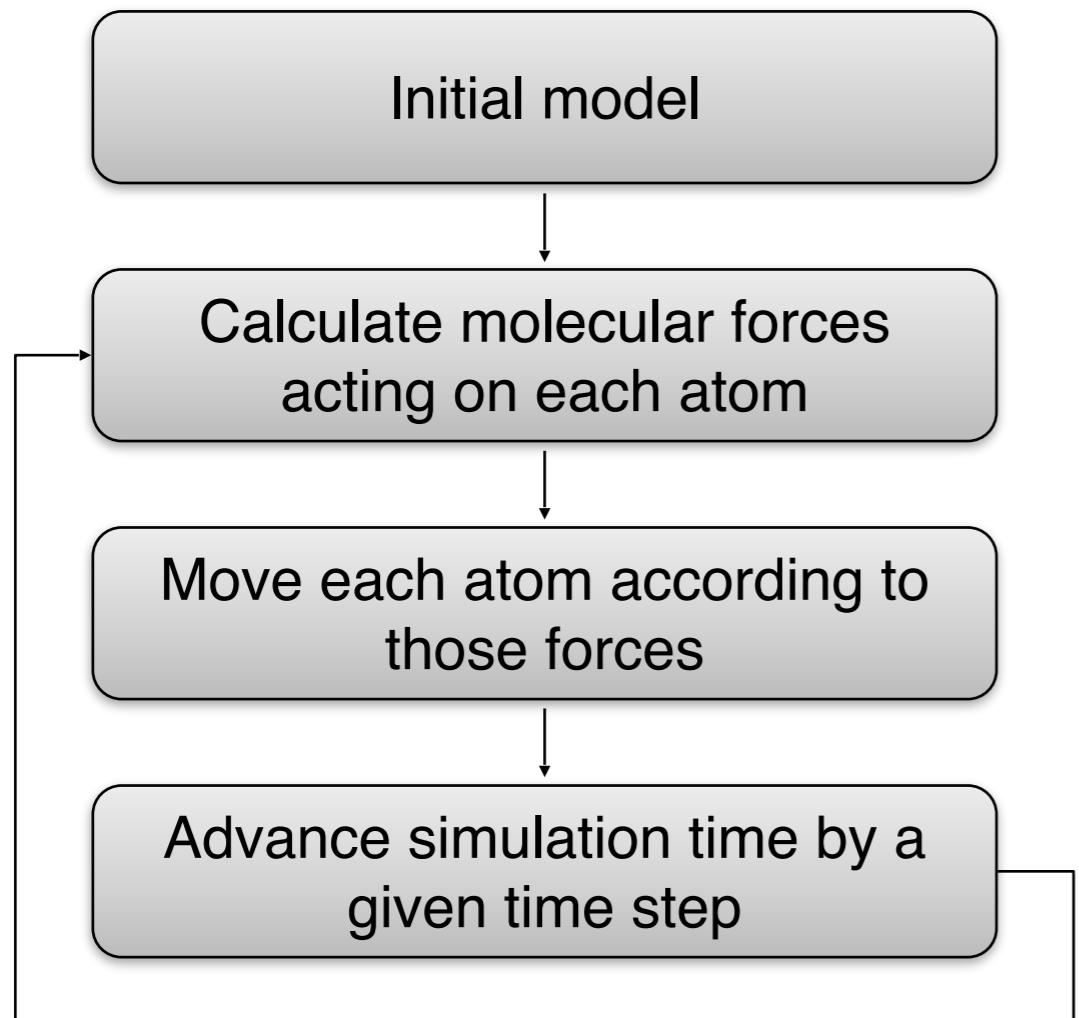
Transferability

Empirical

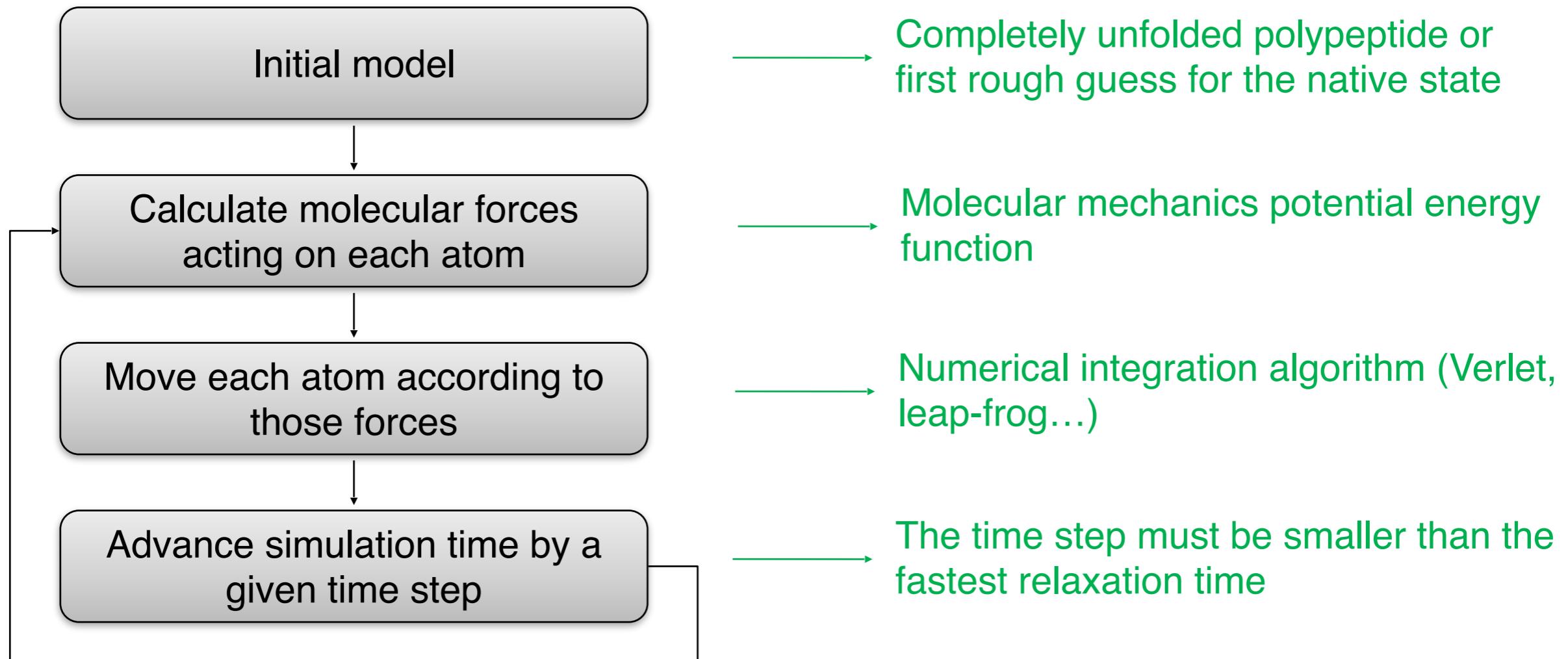
Their general form is:

$$E = E_{bond} + E_{angle} + E_{torsion} + E_{non-bonded} + E_{others}$$

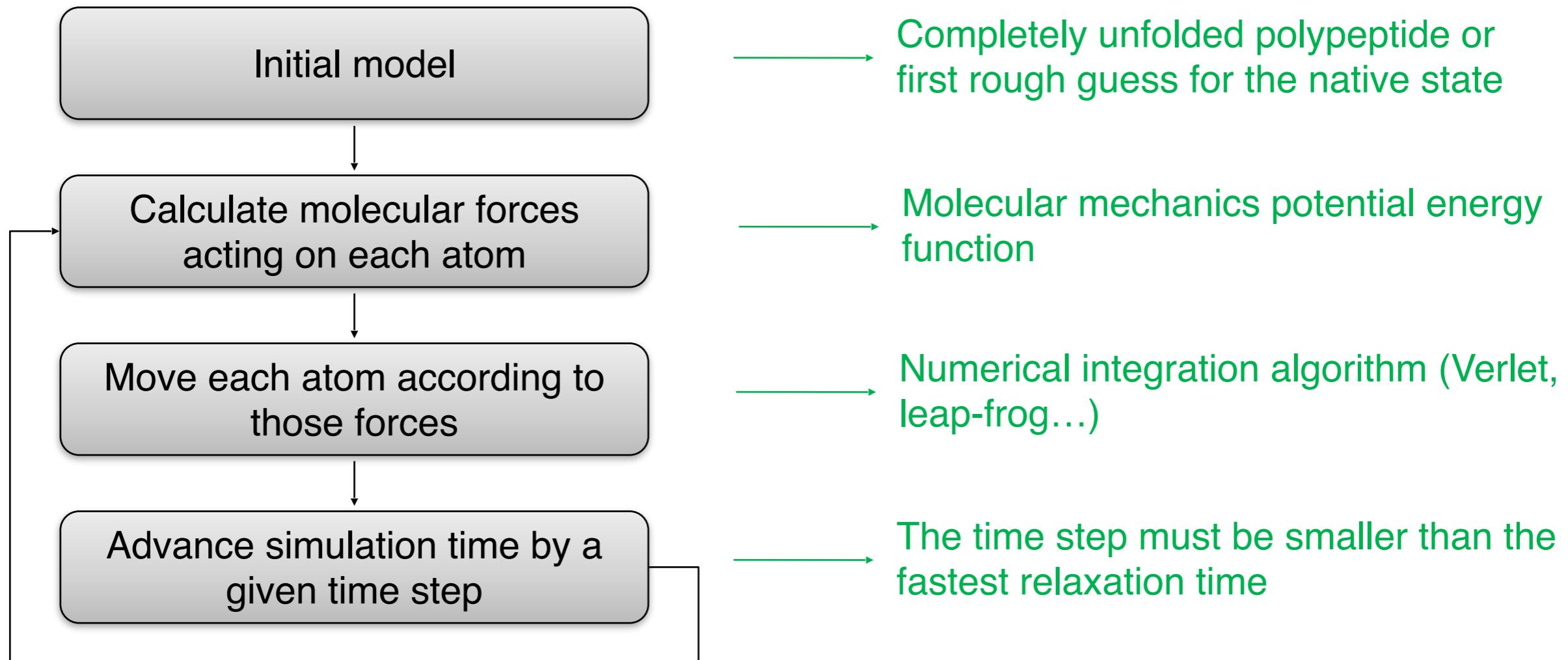
# Molecular dynamics protocol



# Molecular dynamics protocol



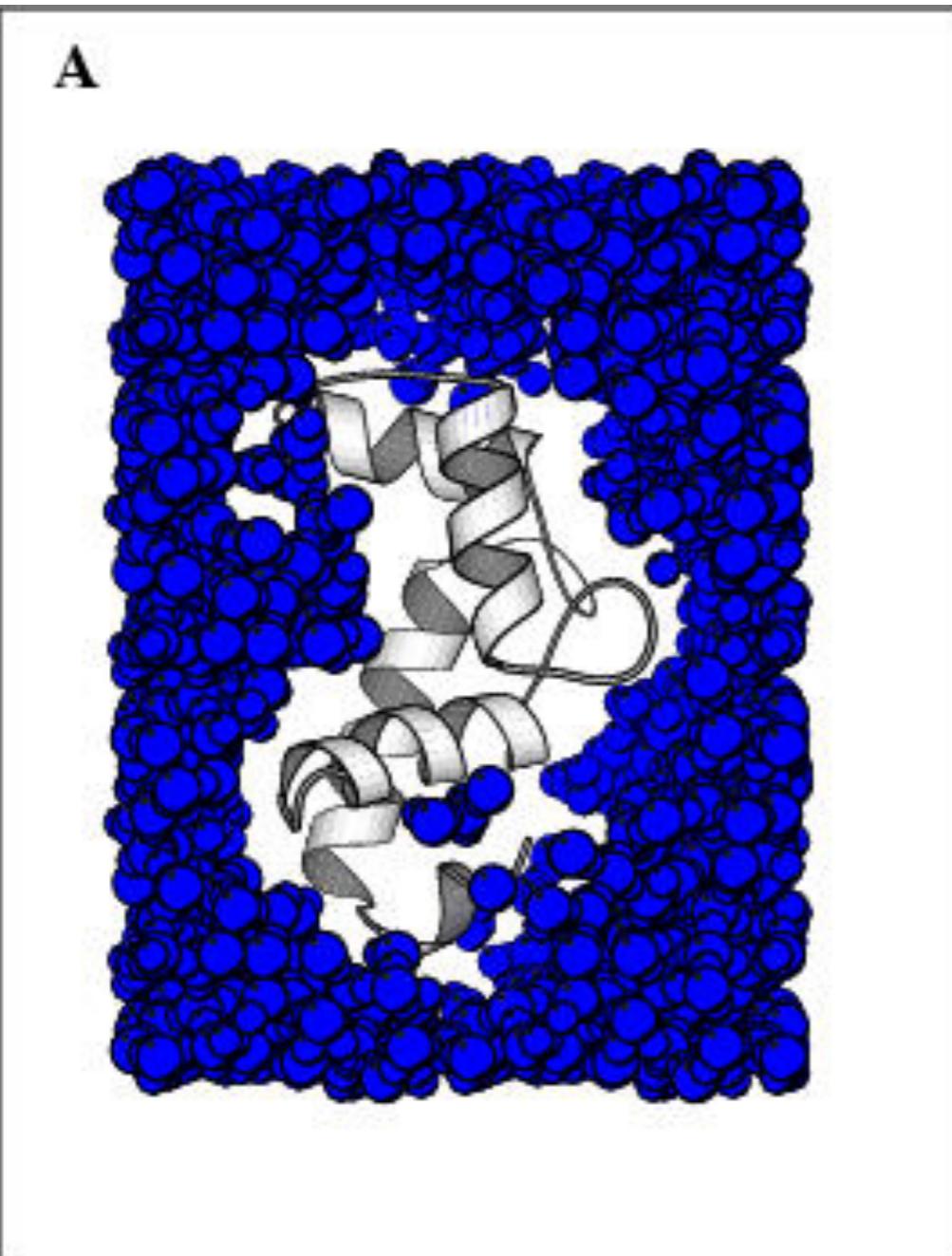
# Molecular dynamics protocol



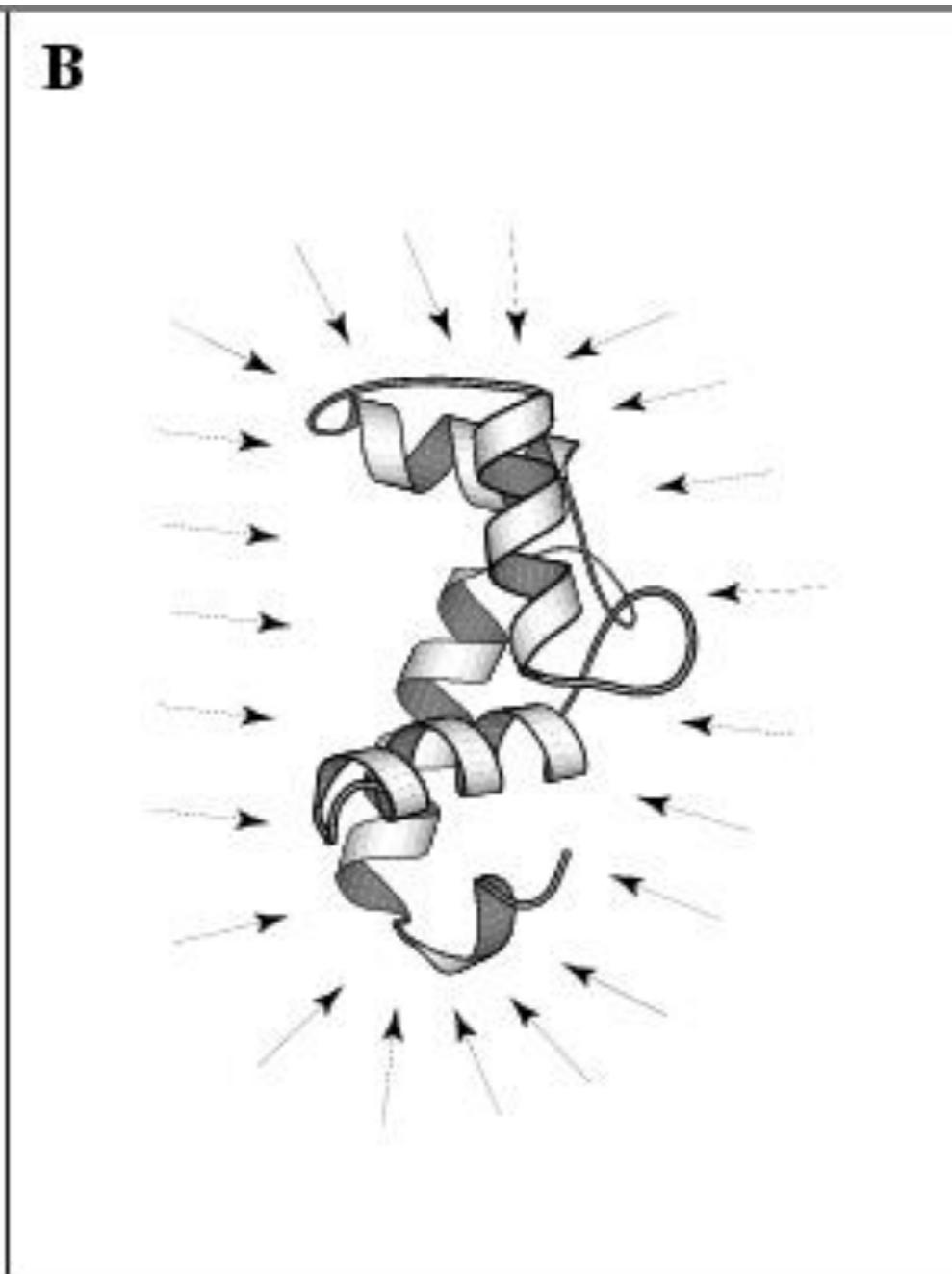
In principle, with long simulations the entire conformational space can be explored (ergodicity)

# Solvent models

Explicit

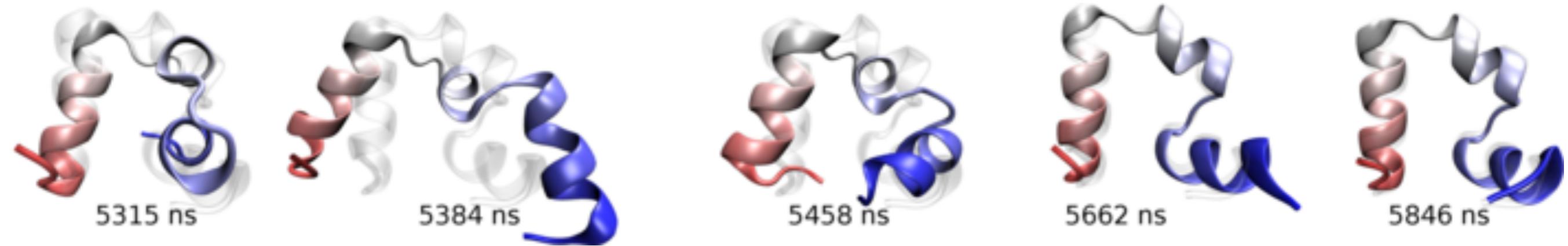
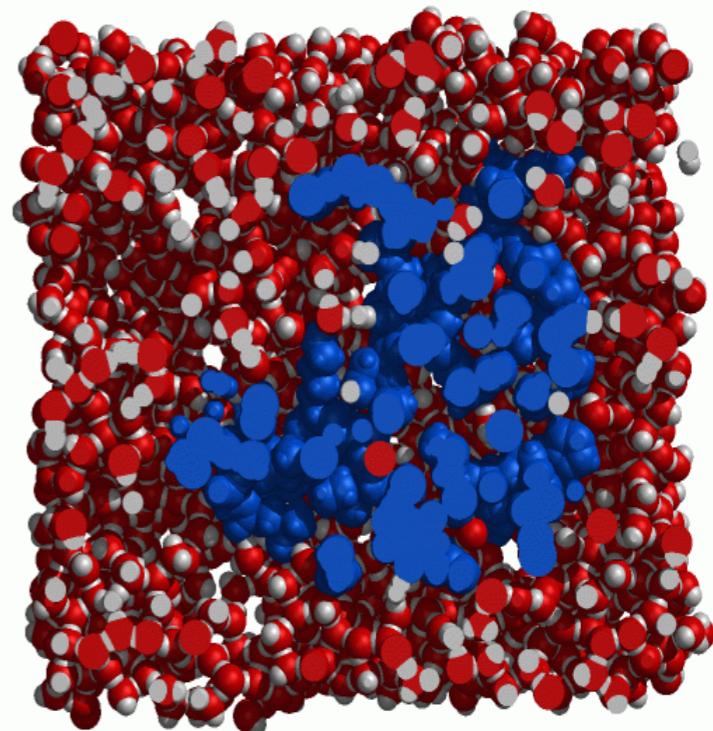


Implicit

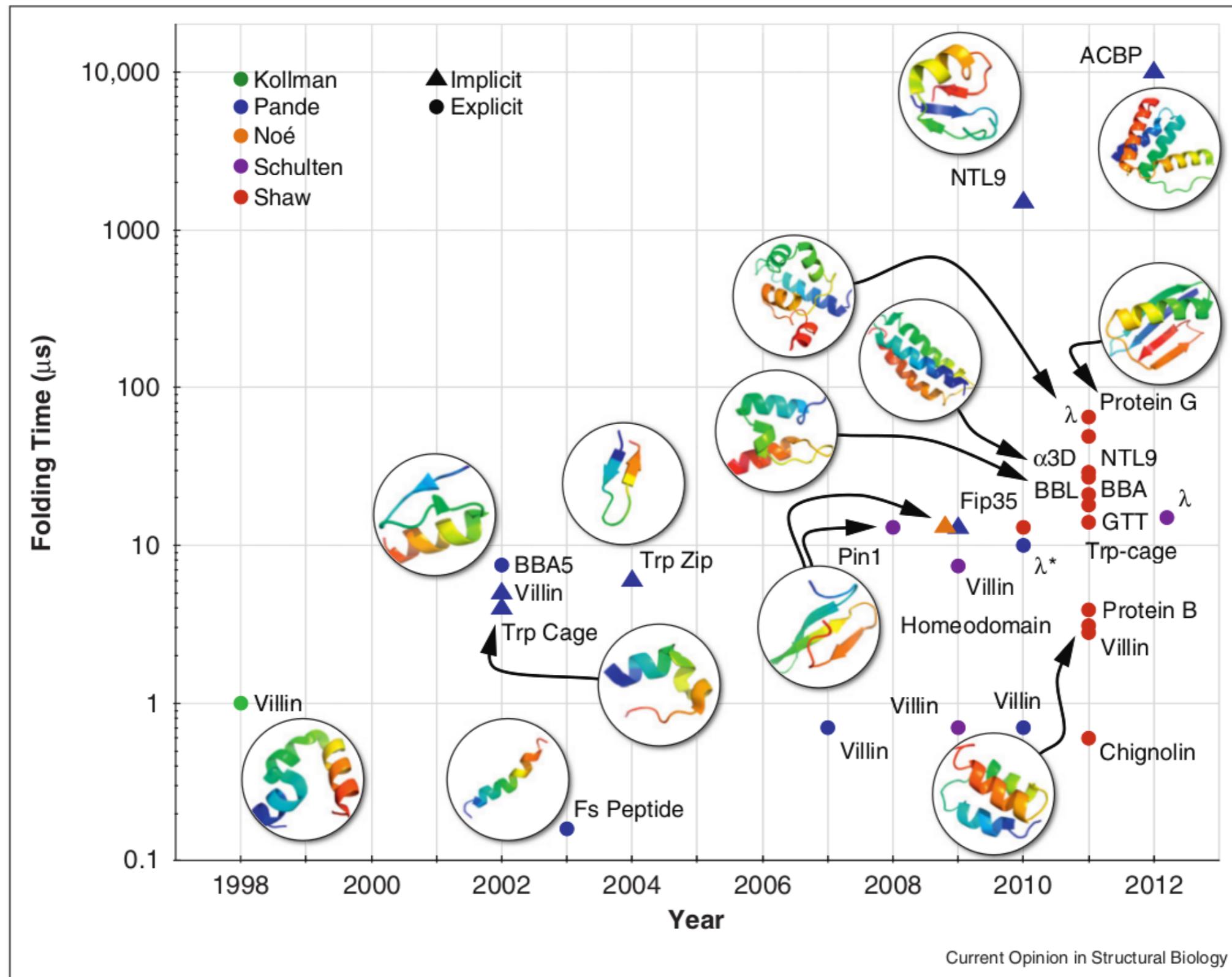


# Simulation set up

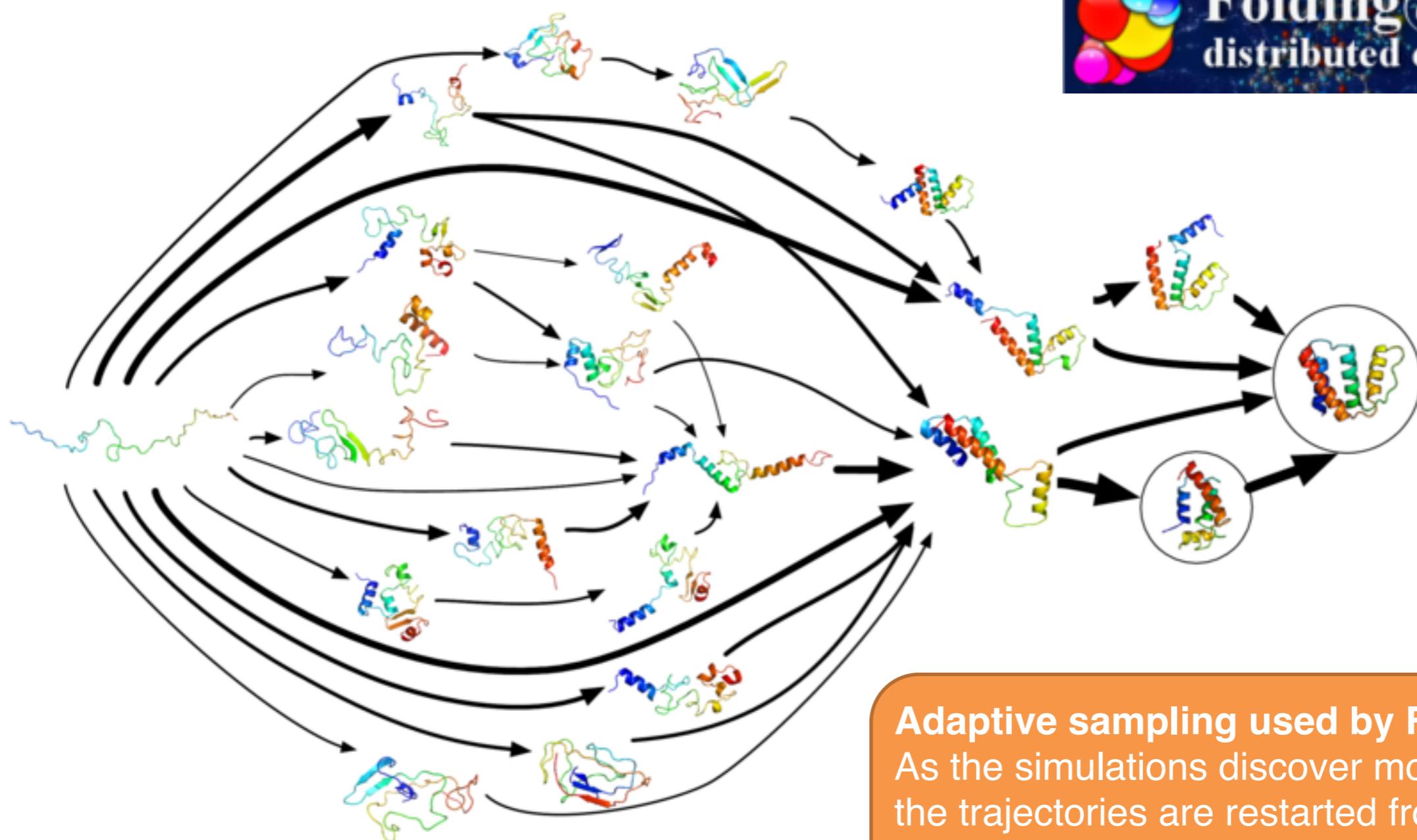
- Solvent representation: explicit or implicit
- High frequency vibrations frozen: longer time step
- Boundary conditions: treating long-range interactions
- Statistical ensemble
  - microcanonical (NVE)
  - canonical (NVT)
  - isothermal-isobaric (NPT)
  - generalized (replica-exchange)



# Folding times accessible by simulation



# Folding@home



**Adaptive sampling used by Folding@home:**  
As the simulations discover more conformations, the trajectories are restarted from them, and a Markov state model (MSM) is gradually created from this cyclic process.

# Modeling approach

Choose a representation of the object  
resolution, forces, heterogeneity, extra data  
*type of data, purpose, feasibility*

*Atoms or coarser grains, average  
(rigid) or ensemble (flexible)...*

**Choose a scoring function**  
**quantifies the match between the model and the**  
**input data, to be optimized**

*Weighted sum of spatial restraints,  
bayesian posterior...*

Choose a search strategy  
the goal is to explore as much as possible, given  
computational limitations

*Systematic or biased, stochastic (MC)*  
...

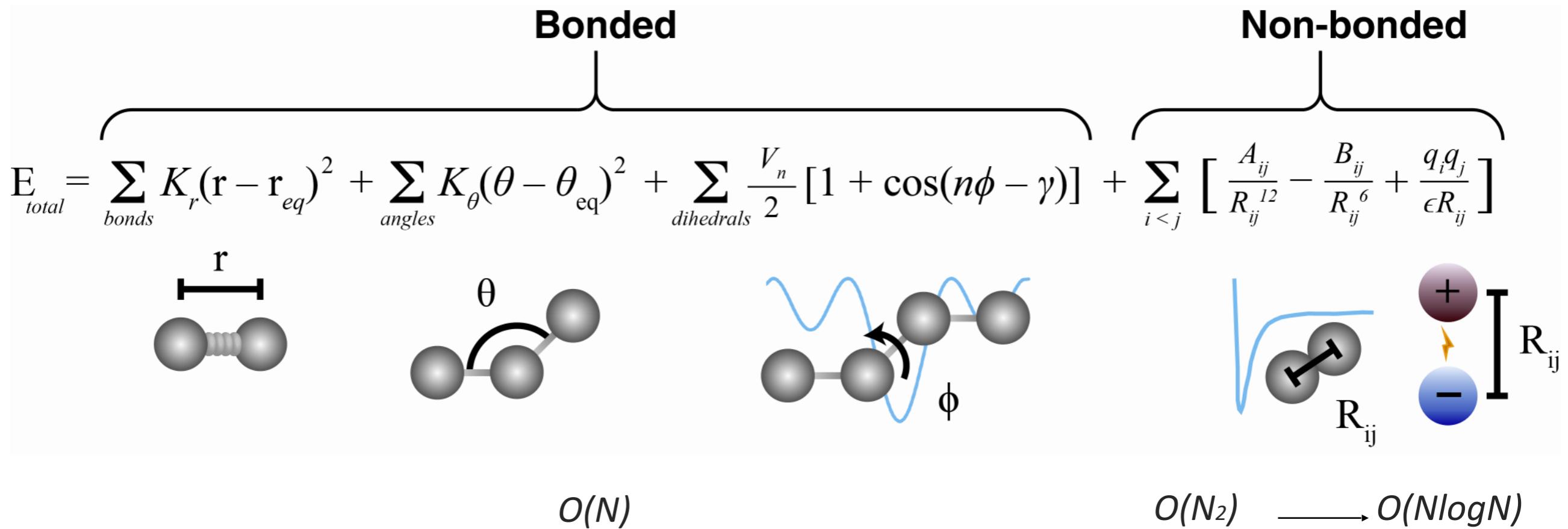
Filter and/or refine  
based on input data, allowing a more expensive  
representation/search on a subset

*Flexible refinement, 2D projection  
class matching...*

# RELYING ON PHYSICS

# Molecular mechanics energy

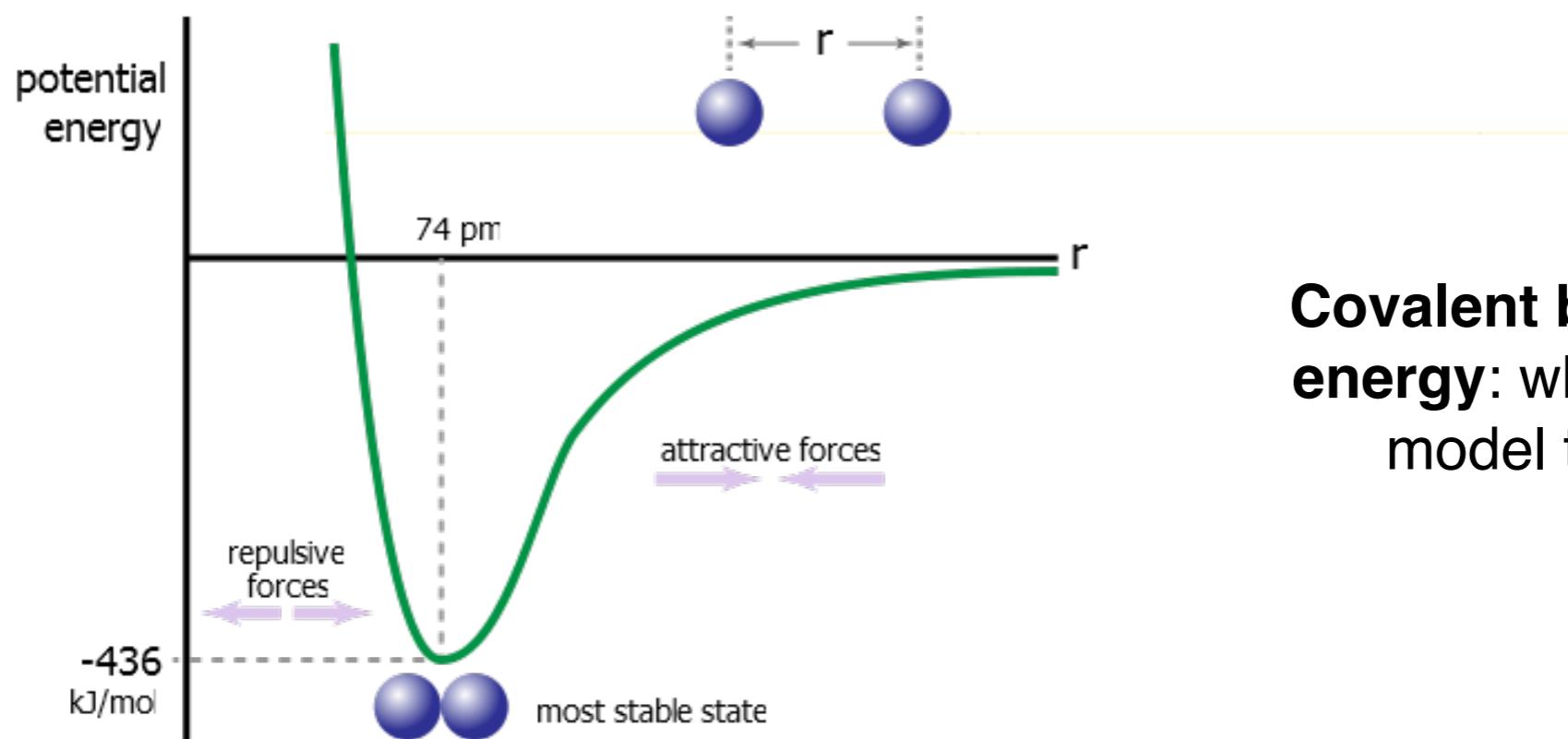
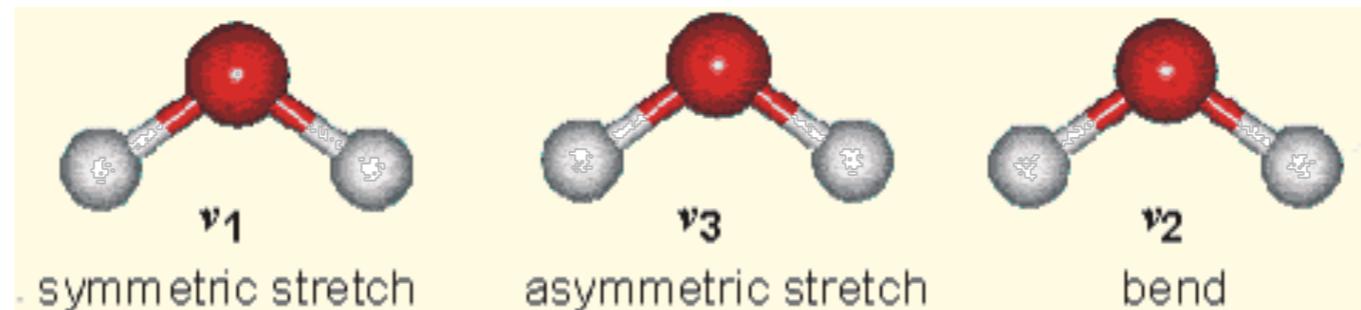
➤ An example: AMBER force field



Many more... CHarMM, OPLS...

# Bonded interactions

How can we represent the variation of the energy corresponding to a covalent bond stretching and bending?



**Covalent bond potential energy:** which function to model this curve ?

# Bonded interactions

## ➤ Morse potential

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

r: interatomic distance

$r_e$ : equilibrium bond distance

$D_e$ : well-depth

a: controls the width of the potential

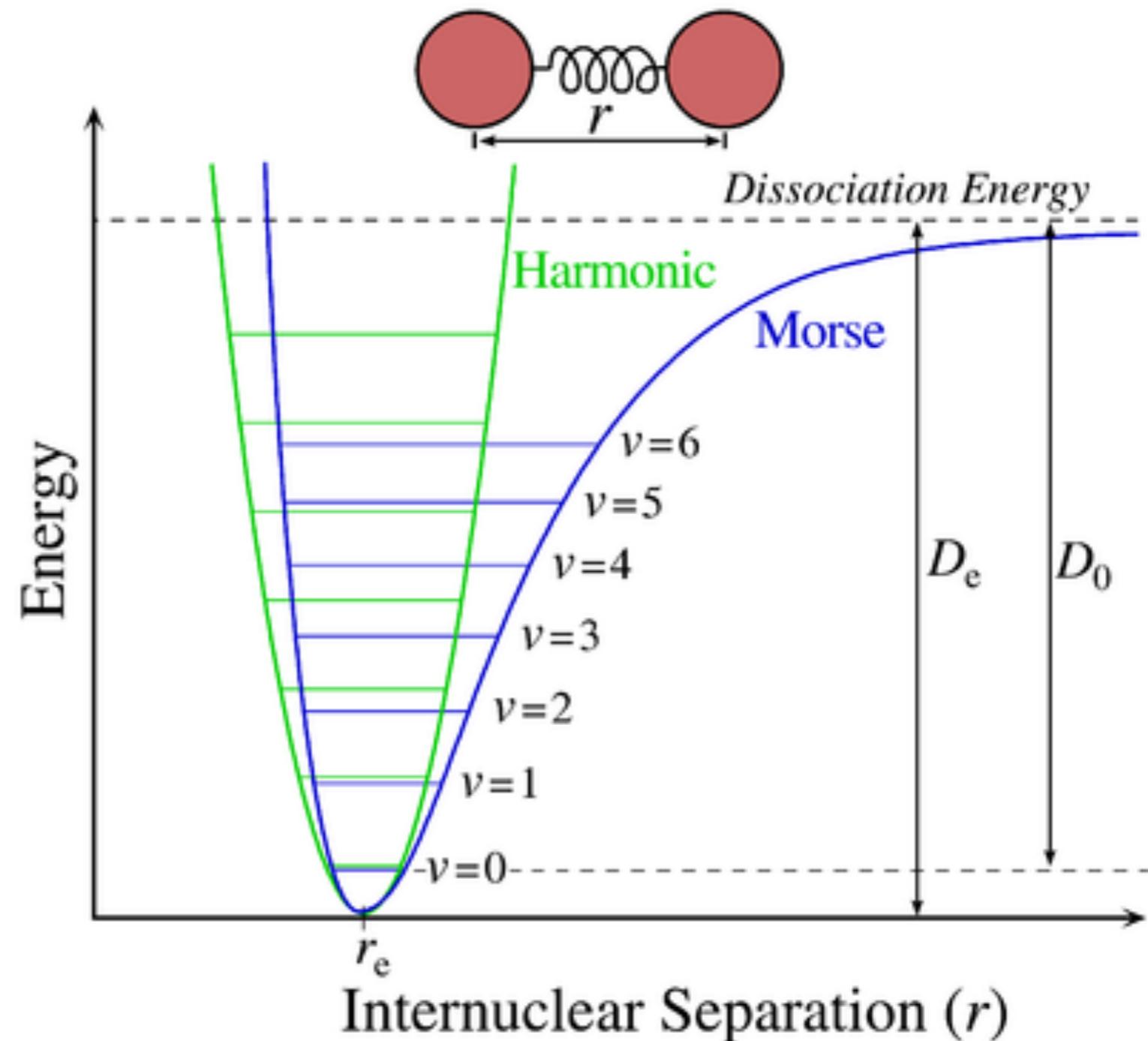
## ➤ Harmonic potential

$$E = \frac{1}{2} k (r - r_e)^2$$

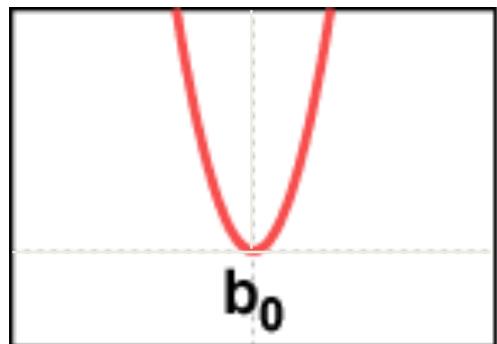
r: interatomic distance

$r_e$ : equilibrium bond distance

k: spring force constant

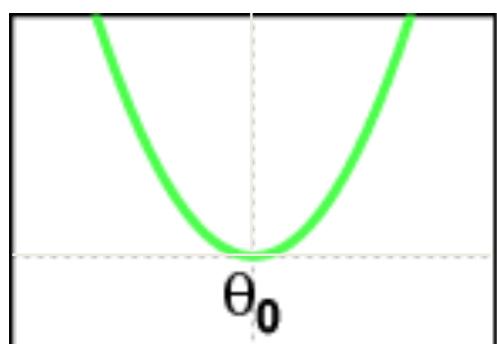


# Bonded interactions



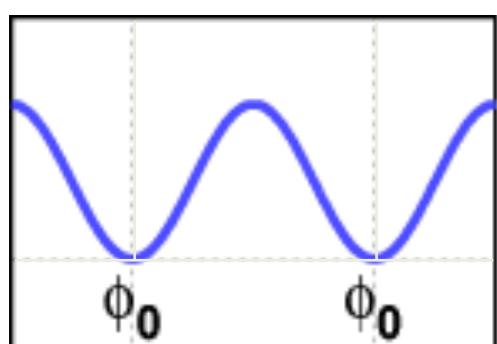
**Bond**

$$\sum_{bonds} K_r (\mathbf{r} - \mathbf{r}_{eq})^2$$



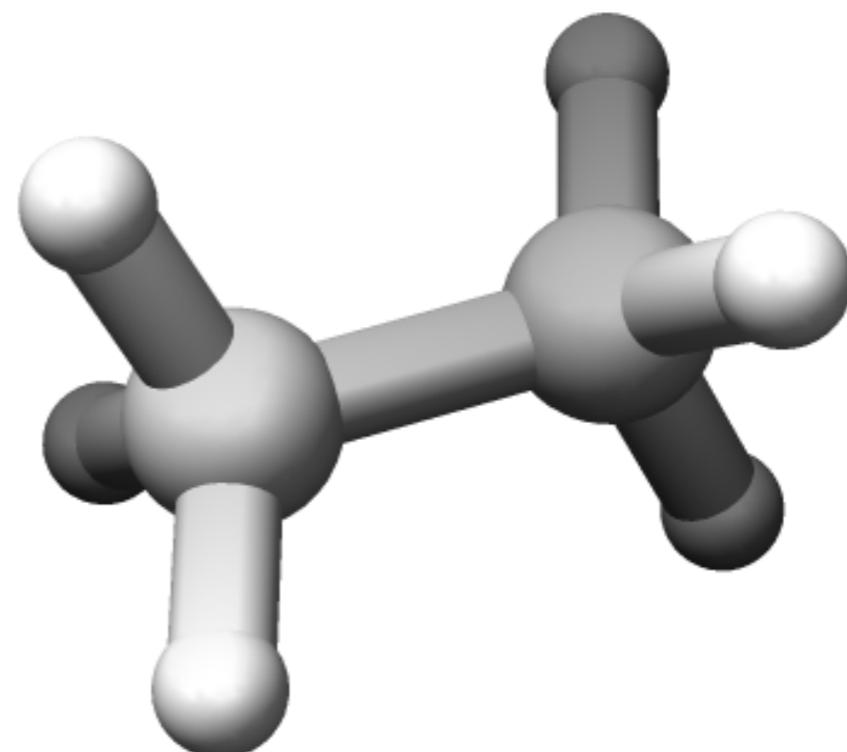
**Angle**

$$\sum_{angles} K_\theta (\theta - \theta_{eq})^2$$

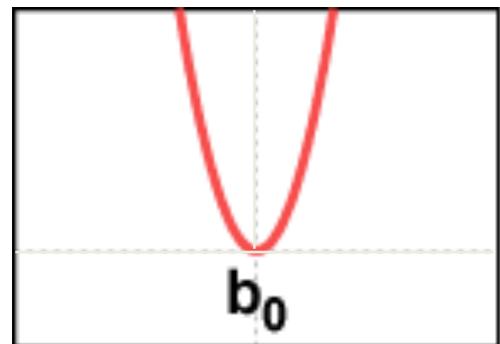


**Dihedral**

$$\sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]$$

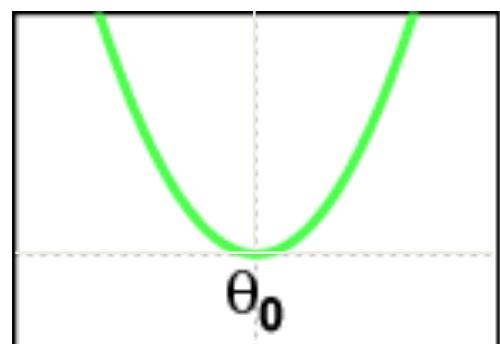


# Bonded interactions



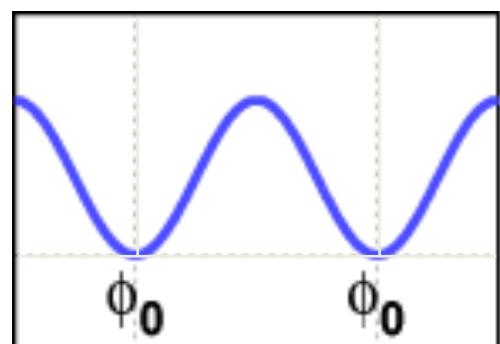
**Bond**

$$\sum_{bonds} K_r (\mathbf{r} - \mathbf{r}_{eq})^2$$



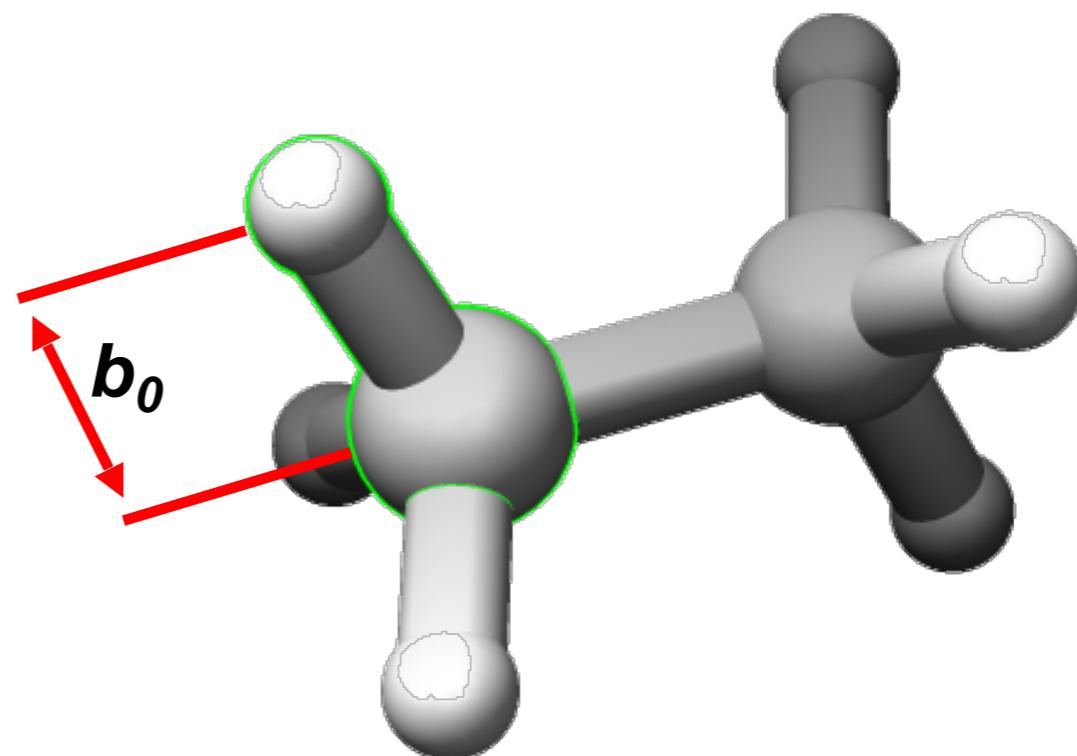
**Angle**

$$\sum_{angles} K_\theta (\theta - \theta_{eq})^2$$

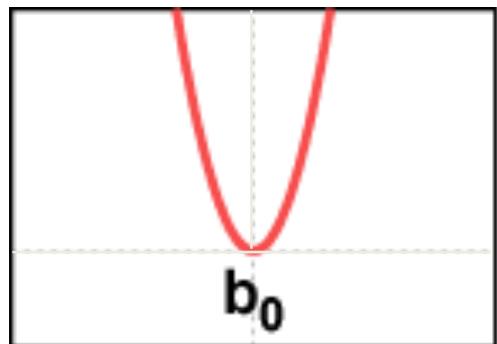


**Dihedral**

$$\sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]$$

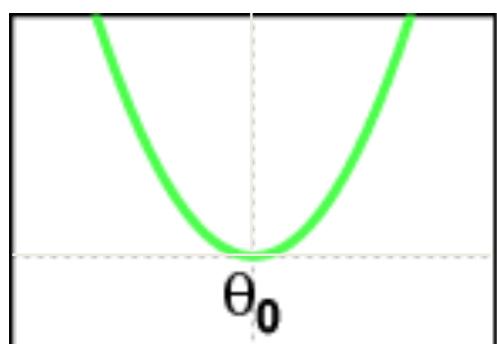


# Bonded interactions



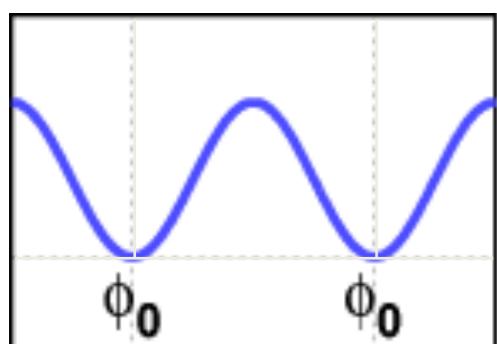
**Bond**

$$\sum_{bonds} K_r (\mathbf{r} - \mathbf{r}_{eq})^2$$



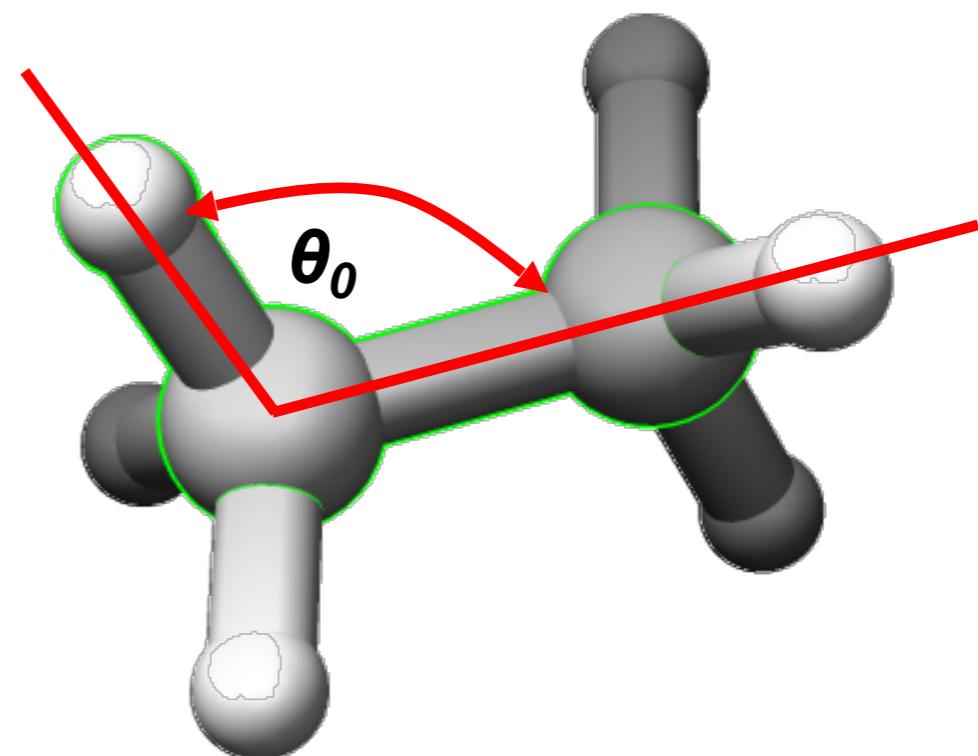
**Angle**

$$\sum_{angles} K_\theta (\theta - \theta_{eq})^2$$

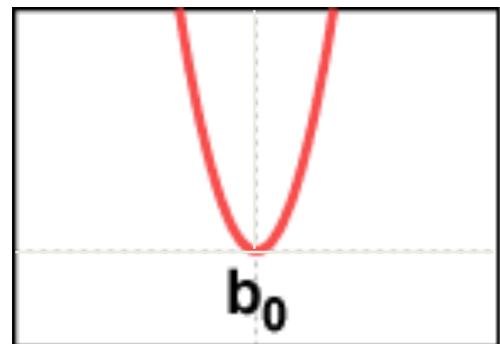


**Dihedral**

$$\sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]$$

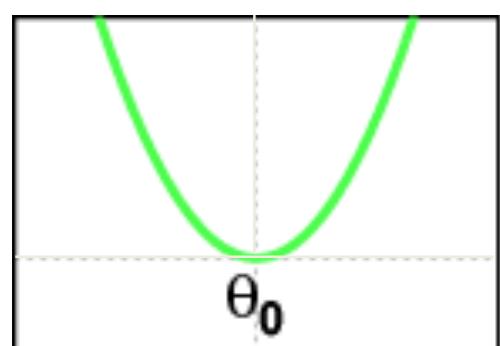


# Bonded interactions



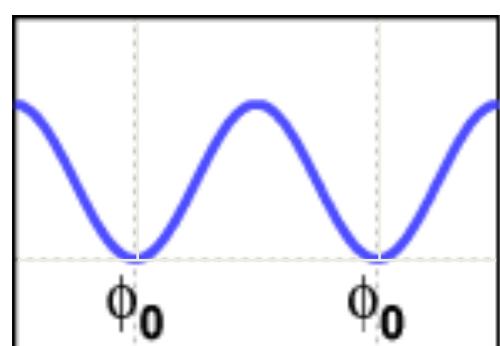
**Bond**

$$\sum_{bonds} K_r (\mathbf{r} - \mathbf{r}_{eq})^2$$



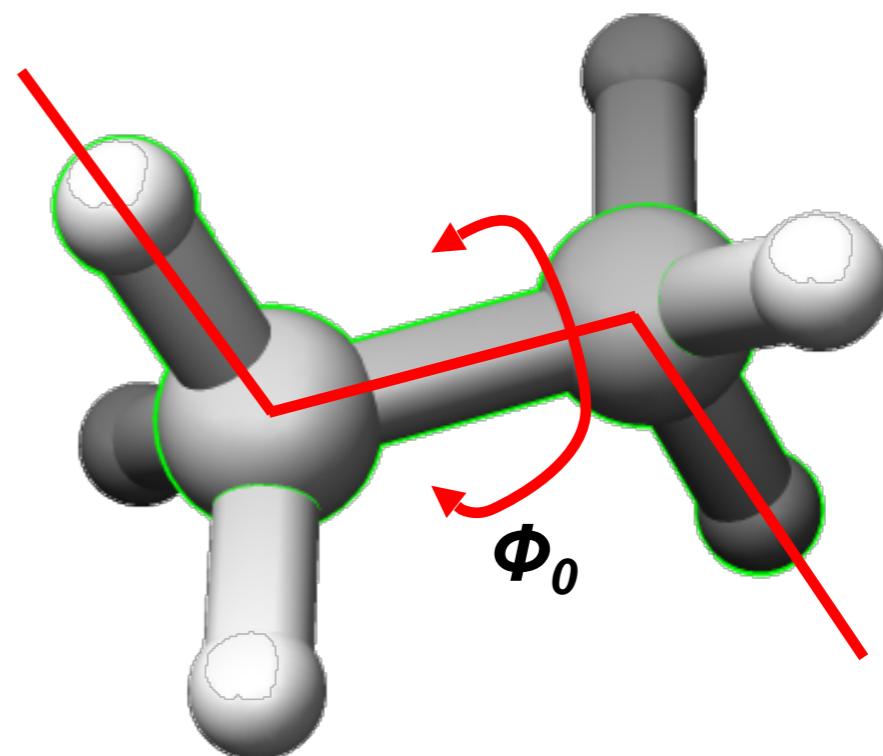
**Angle**

$$\sum_{angles} K_\theta (\theta - \theta_{eq})^2$$

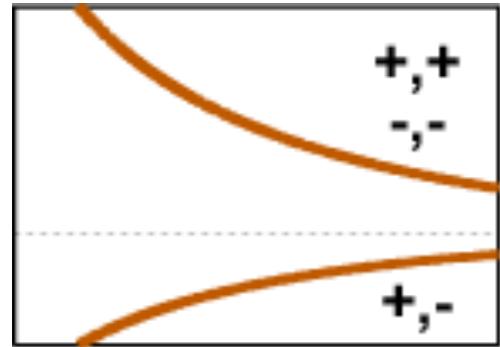


**Dihedral**

$$\sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]$$

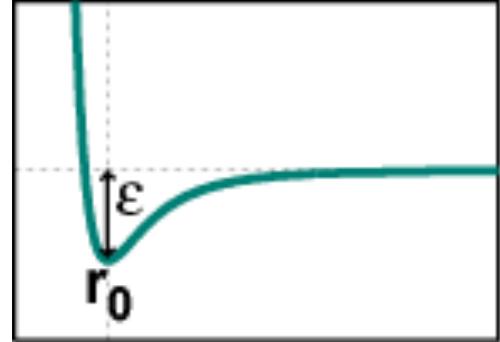


# Non-bonded interactions



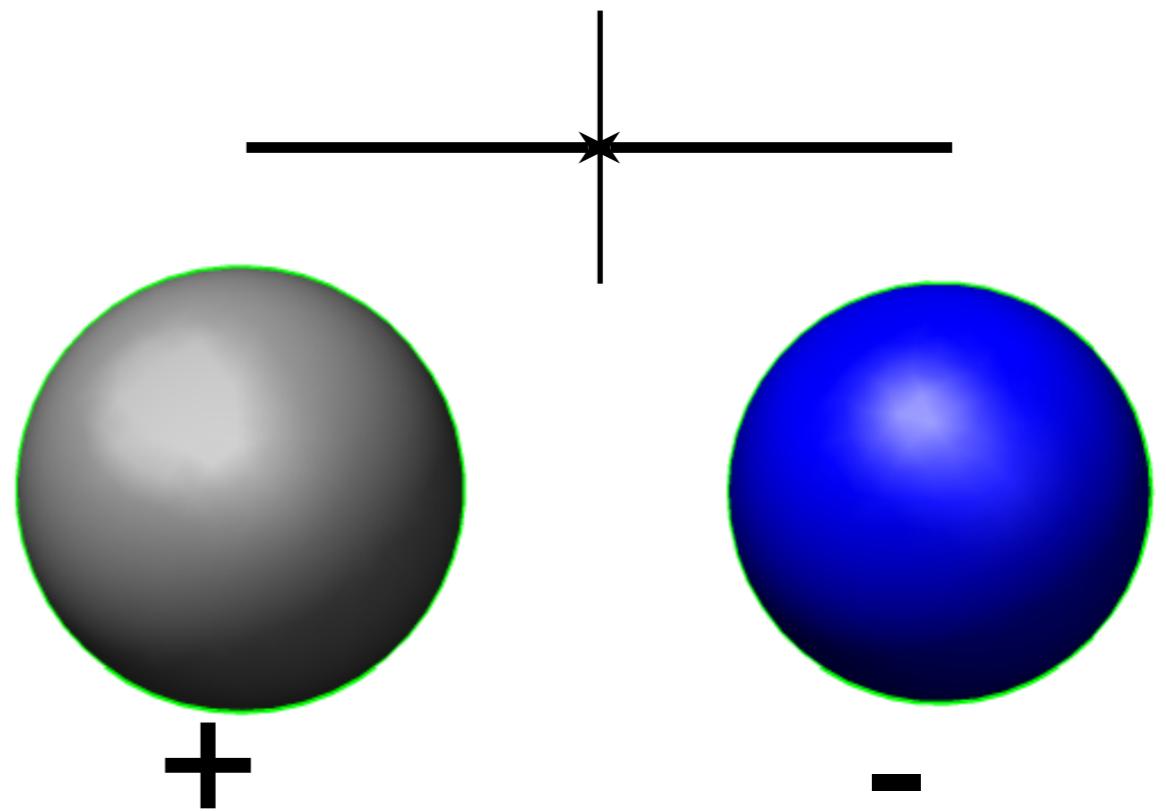
**Electrostatics**

$$\sum_{i < j} \left[ \frac{q_i q_j}{\epsilon R_{ij}} \right]$$



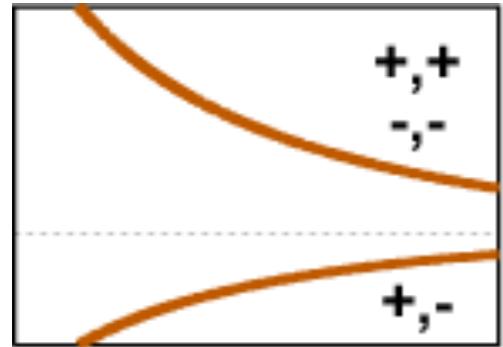
**van der Waals**

$$\sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right]$$



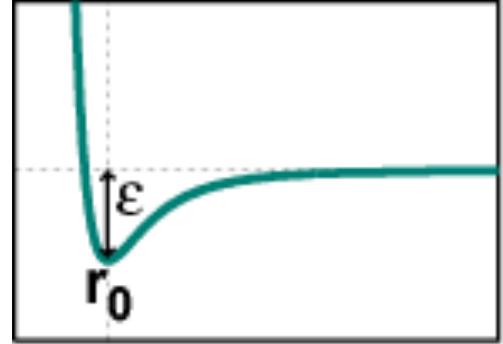
*Coulomb interaction between  
single point charges*

# Non-bonded interactions



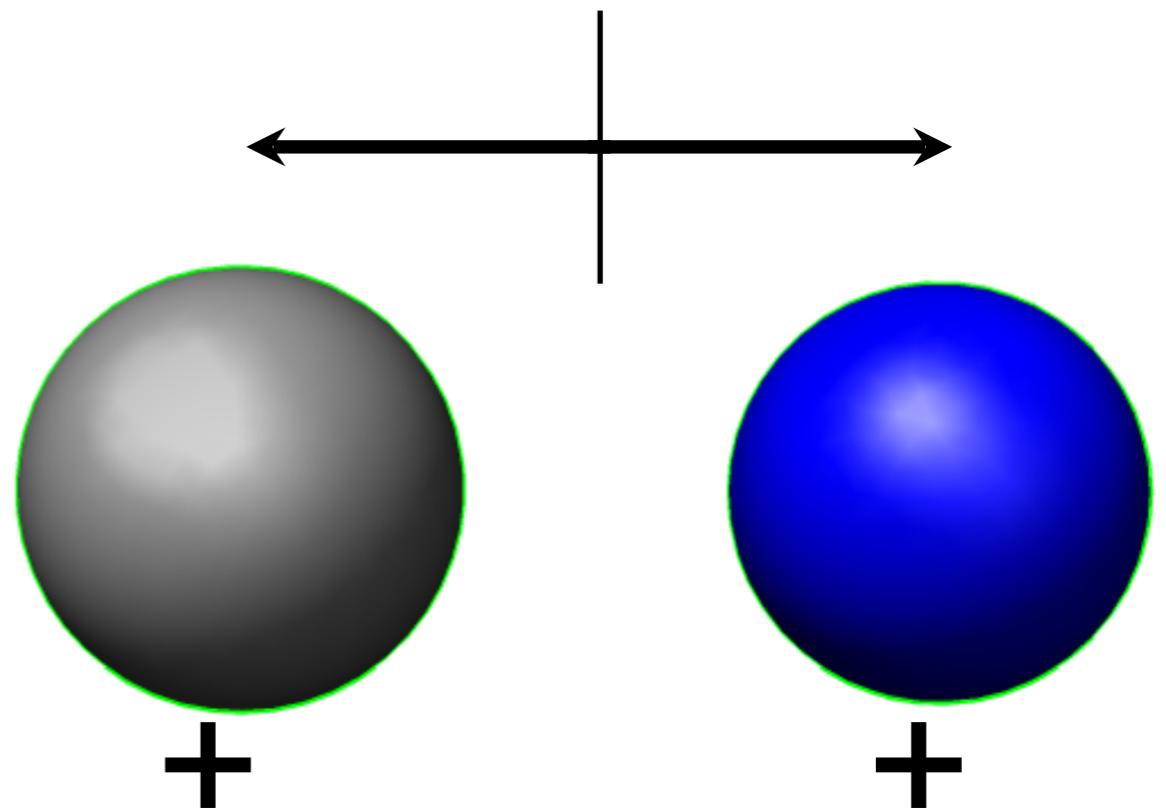
**Electrostatics**

$$\sum_{i < j} \left[ \frac{q_i q_j}{\epsilon R_{ij}} \right]$$



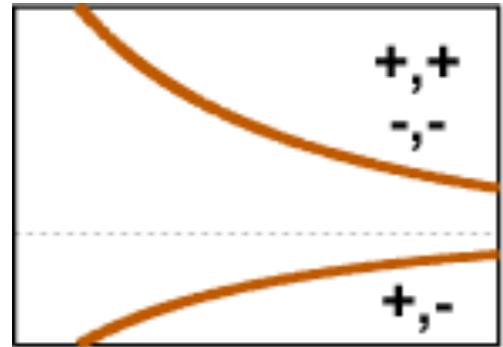
**van der Waals**

$$\sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right]$$



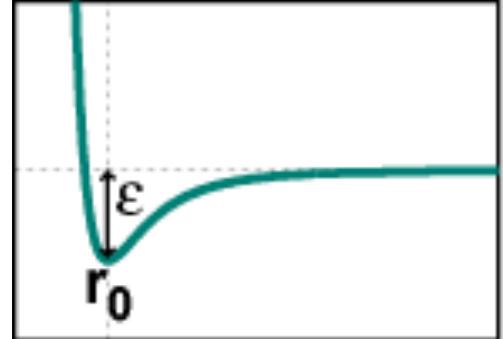
*Coulomb interaction between  
single point charges*

# Non-bonded interactions



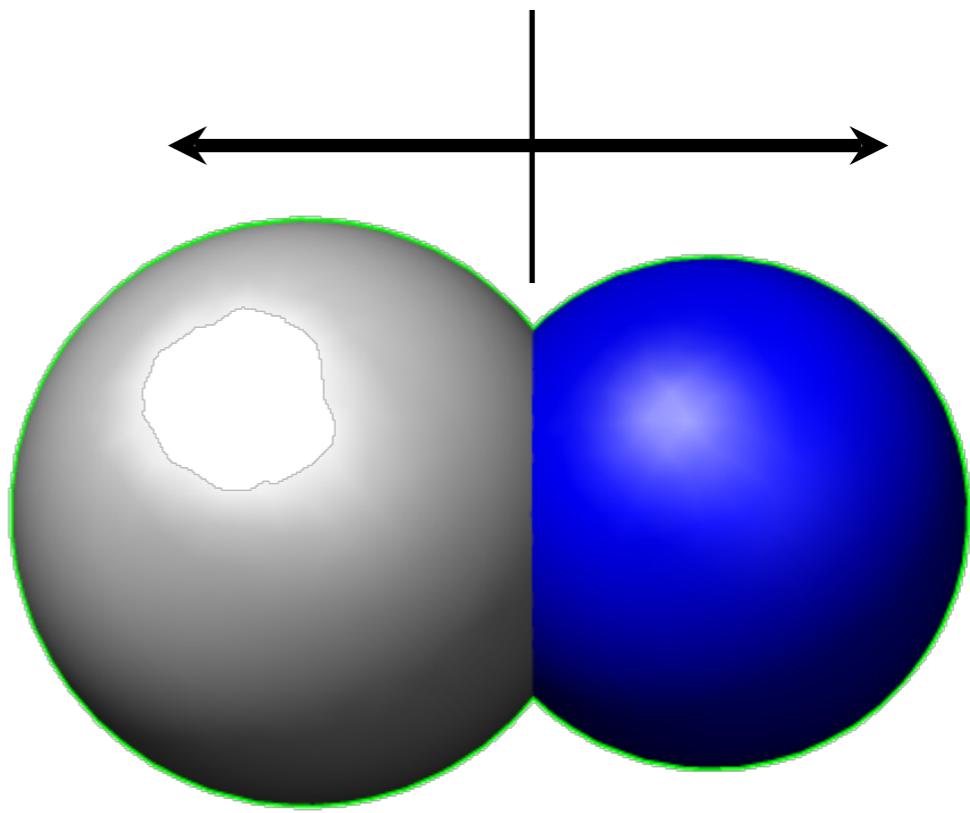
**Electrostatics**

$$\sum_{i < j} \left[ \frac{q_i q_j}{\epsilon R_{ij}} \right]$$



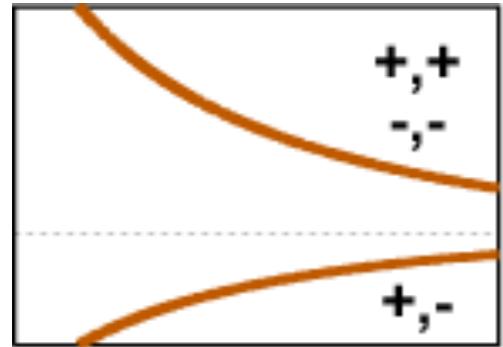
**van der Waals**

$$\sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right]$$



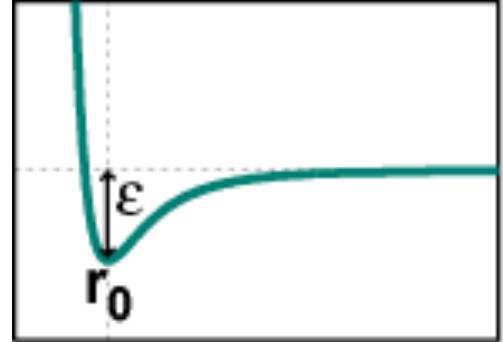
*Hard core repulsion between close atoms*

# Non-bonded interactions



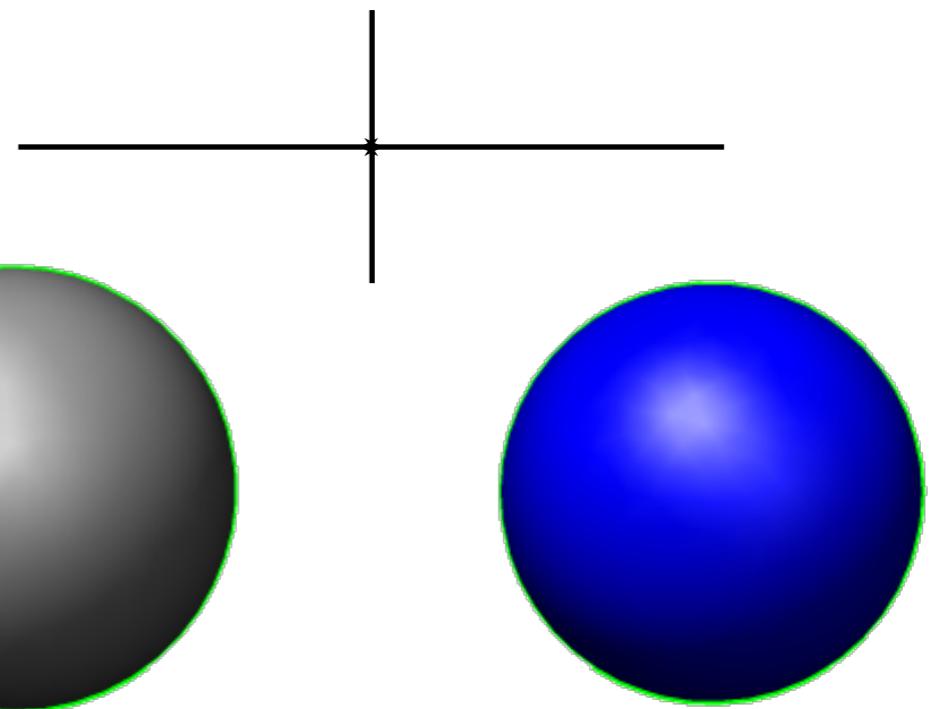
**Electrostatics**

$$\sum_{i < j} \left[ \frac{q_i q_j}{\epsilon R_{ij}} \right]$$



**van der Waals**

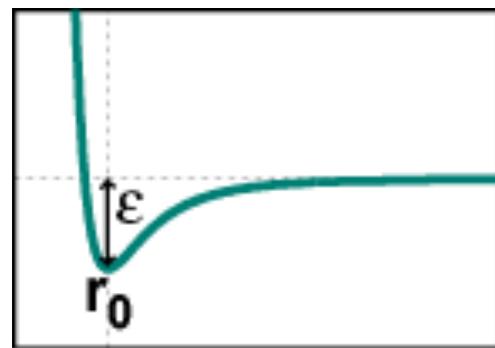
$$\sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right]$$



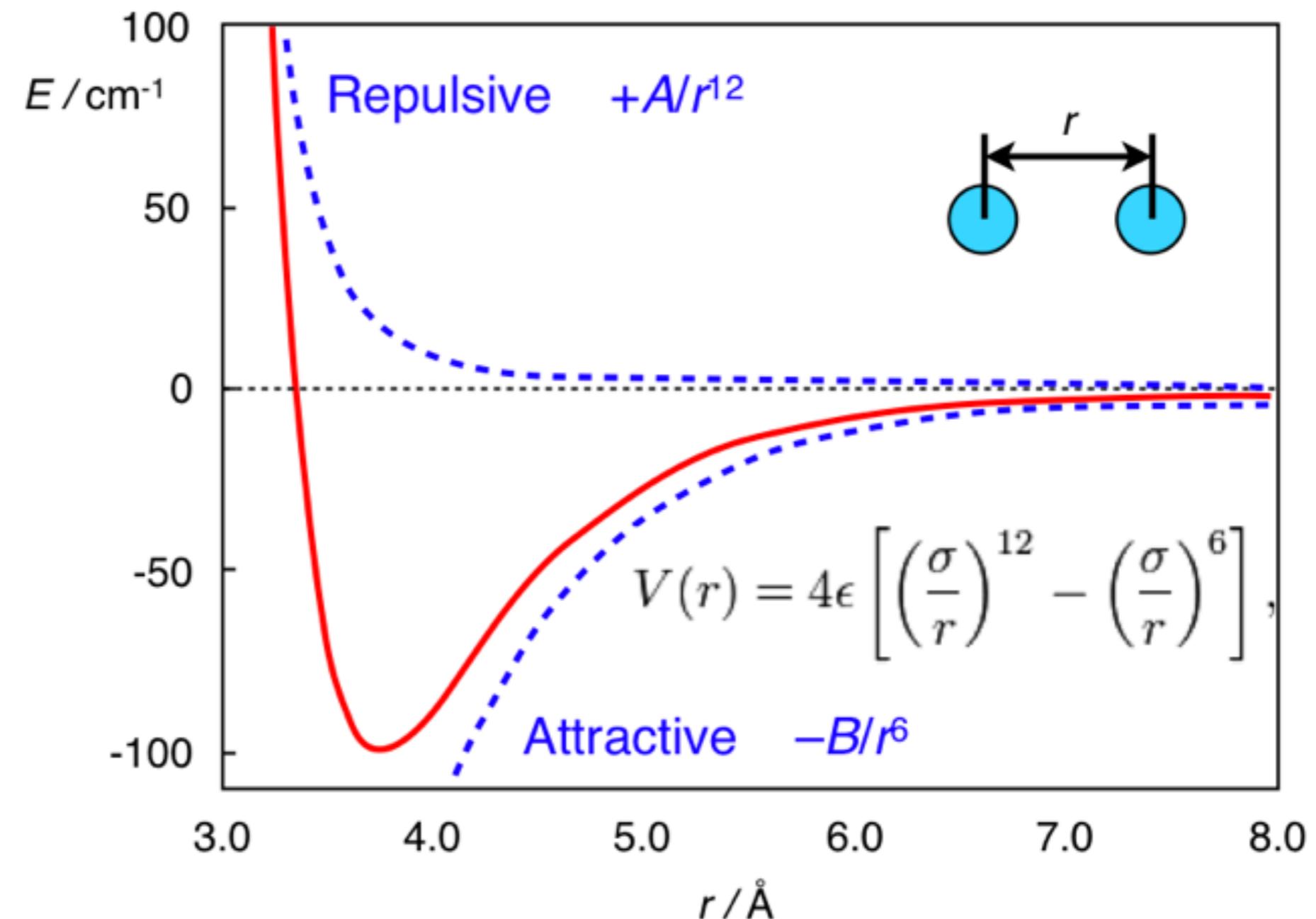
*Weak dipole attraction between distant atoms*

# Lennard-Jones potential

van der Waals

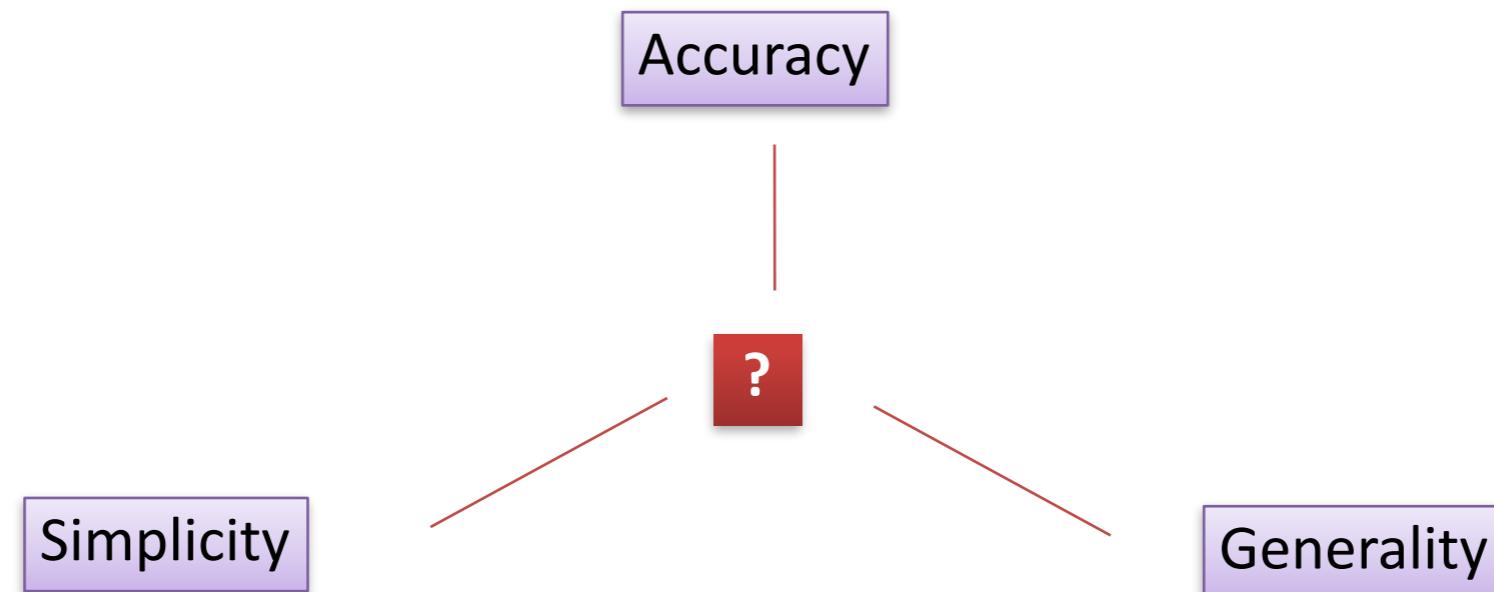


$$\sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right]$$



# Parametrization

Determining parameter values that best fit the force field and lead to the most accurate energy estimates is not trivial.



Parameters are fitted to experimental data (spectroscopy, small molecular crystals...) or quantum mechanics calculations. They are computed for a certain type of molecules (proteins, nucleic acids...) and may not be transferable.

# Pair vs multi-body potentials

Coulombic and van der Waals potentials are summed over pairs of atoms.

How do we account for the influence of all the other particles in the system ?

*Pairs:  $N(N-1)/2$*

*Triplets:  $N(N-1)(N-2)/6$*

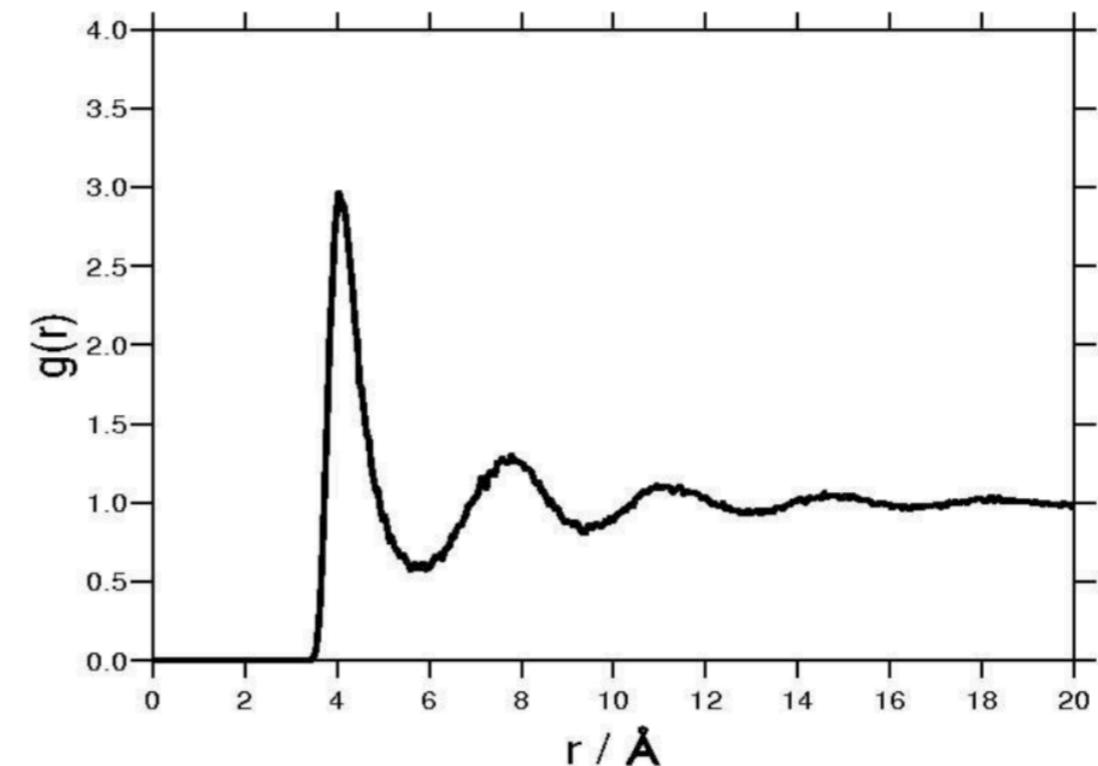
**Effective potentials:** account for the presence of the other entities through parametrization. An effective pair potential does not reflect the « true » interaction energy between two isolated atoms but is parametrized so as to include the effect of the other atoms in the energy of the pair.

# RELYING ON STATISTICS

# Statistical mechanics

- Native structures have distinctive geometric features
- Binding energy can be approximated as a linear superposition of individual interactions
- Protein distribution functions obey Boltzmann statistics

$$\Delta G = -kT \sum_{ij} \ln \left( \frac{g_{ij}(r)}{g(r)} \right)$$



- However, the reference distribution function  $g(r)$  is rarely known!

# Statistical mechanics

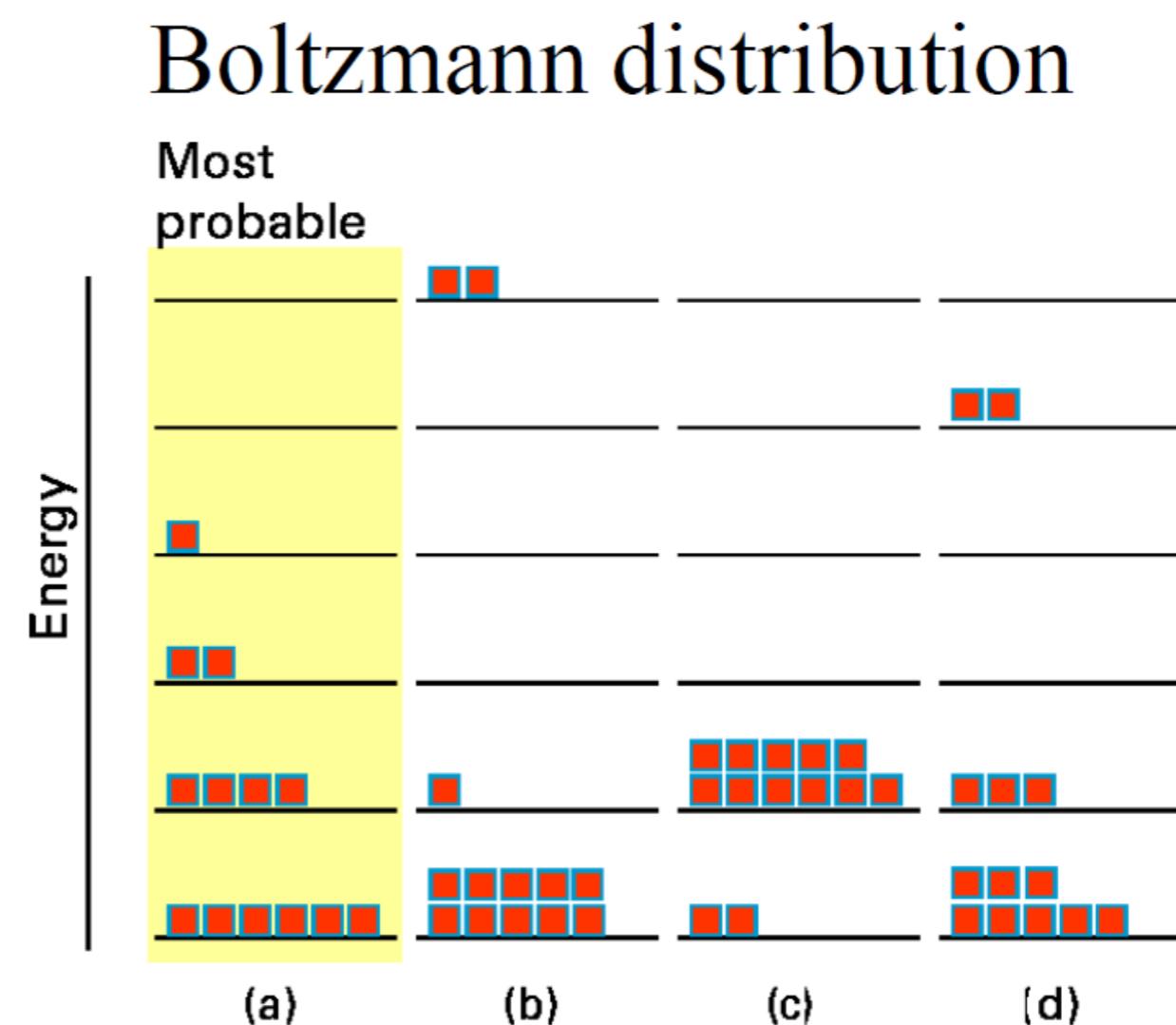
Proteins adopt an ensemble of **conformations** in solution. Not every protein in a large group of them has the lowest energy. The energies are random but they obey certain statistical laws based on the **Boltzmann distribution**.

The probability of observing a given conformation  $C_i$  is:

$$P(C_i) = \frac{\exp(-E(C_i)/kT)}{\sum \exp(-E(C_i)/kT)}$$

Boltzmann coefficient

partition function



# Popular potentials

## ❖ DOPE potential (Modeller)

$$G(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) \approx -k_B T \sum_{i \neq j}^N \ln g_{i,j}^{(2)}(\vec{r}) \\ = \sum_{i \neq j}^N \bar{u}_{i,j}(r)$$

$$\bar{u}_{i,j}(r) = -k_B T \ln\left(\frac{p_{m,n}(r)}{p_{m,n}^{REF}(r)}\right) \approx -k_B T \ln\left(\frac{N_{m,n}^{OBS}(r)}{N_{m,n}^{REF}(r)}\right)$$

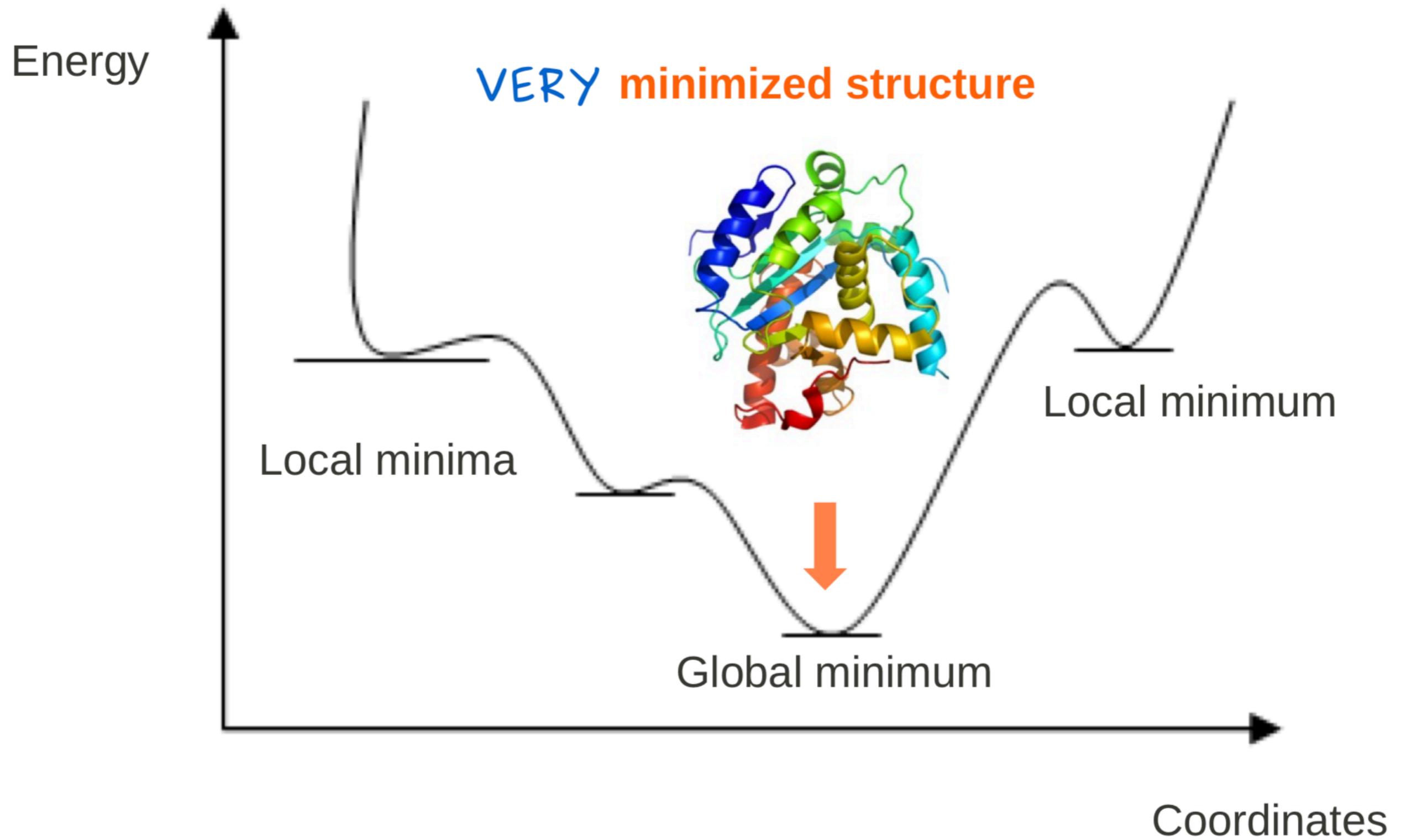
## ❖ ROSETTA all-purpose energy function

It combines a large number of terms, some of them physical, others statistical.

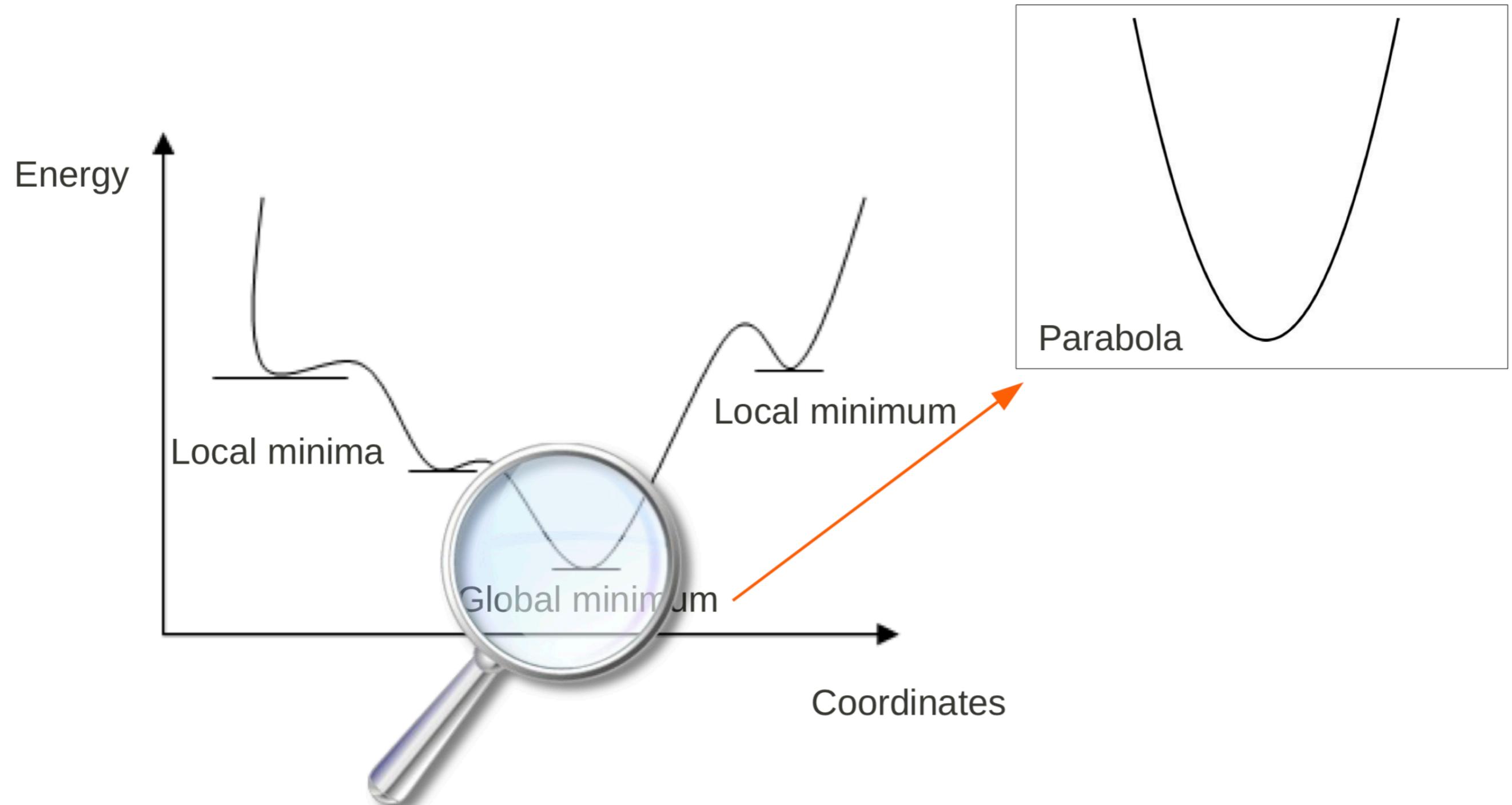
See Alford *et al.* 2017 J. Chem Theory Comput. doi: [10.1021/acs.jctc.7b00125](https://doi.org/10.1021/acs.jctc.7b00125)

# AN ANALYTICAL SOLUTION

# Normal mode analysis



# Normal mode analysis



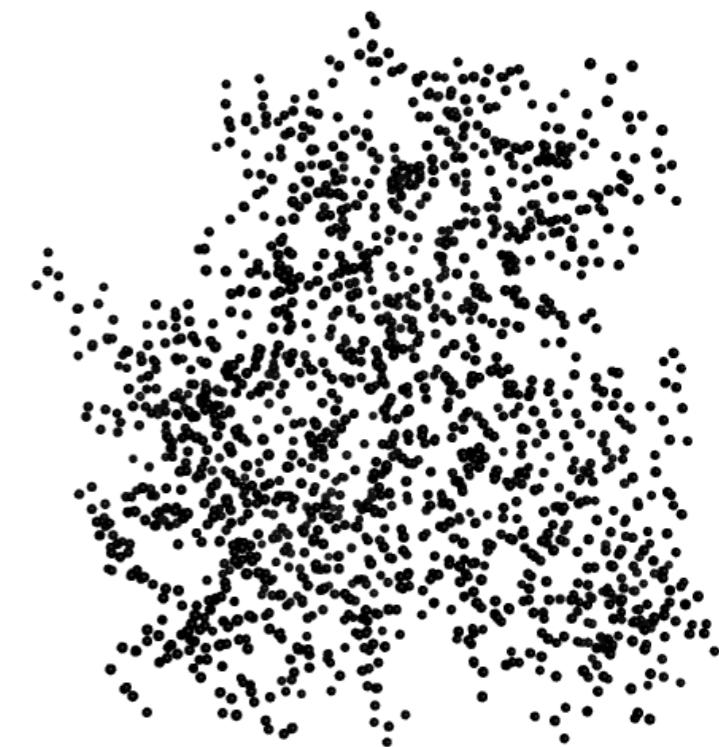
# Normal mode analysis

Let us consider a molecular system with  $N$  atoms at equilibrium position  $q_0 \in \mathbb{R}^{3N}$ .

Let  $V: \mathbb{R}^{3N} \mapsto \mathbb{R}$  be the potential energy of the system.

According to Newton's equation of motion, at any point  $q_0 + q$ ,  
 $F(q_0 + q) = M \times (\ddot{q}_0 + \ddot{q})$ , where  $M$  is the diagonal mass matrix  
and  $F(q_0 + q)$  represents the interatomic forces.

By definition,  $F(q_0 + q) = -\nabla V(q_0 + q)$ ,  
and thus we can write  $M \times (\ddot{q}_0 + \ddot{q}) + \nabla V(q_0 + q) = 0$



# Normal mode analysis

Let us consider a molecular system with  $N$  atoms at equilibrium position  $q_0 \in \mathbb{R}^{3N}$ .

Let  $V: \mathbb{R}^{3N} \mapsto \mathbb{R}$  be the potential energy of the system.

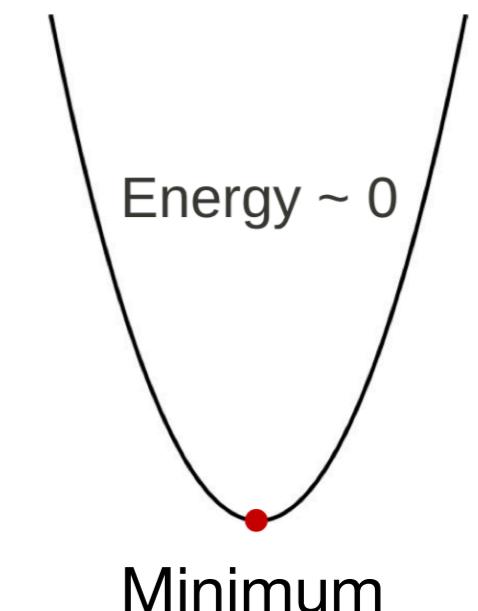
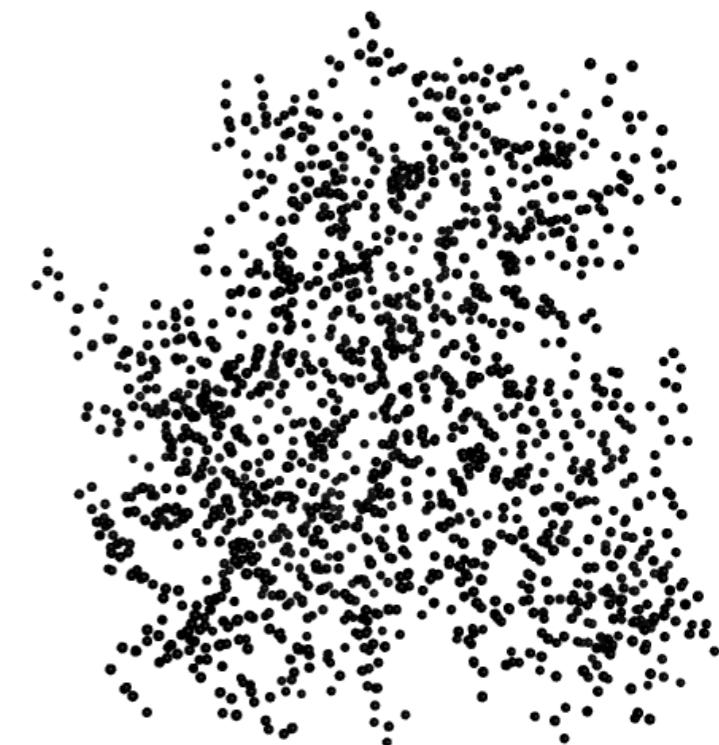
According to Newton's equation of motion, at any point  $q_0 + q$ ,  
 $F(q_0 + q) = M \times (\ddot{q}_0 + \ddot{q})$ , where  $M$  is the diagonal mass matrix  
and  $F(q_0 + q)$  represents the interatomic forces.

By definition,  $F(q_0 + q) = -\nabla V(q_0 + q)$ ,  
and thus we can write  $M \times (\ddot{q}_0 + \ddot{q}) + \nabla V(q_0 + q) = 0$

Around the equilibrium position, if  $q$  is small, then the potential energy  
can be approximated using a Taylor expansion,

$$V(q_0 + q) = V(q_0) + J \times q + \frac{1}{2} H \times q^2 + \dots,$$

where  $J$  and  $H$  are the Jacobian and Hessian matrices evaluated at  $q_0$



# Normal mode analysis

Let us consider a molecular system with  $N$  atoms at equilibrium position  $q_0 \in \mathbb{R}^{3N}$ .

Let  $V: \mathbb{R}^{3N} \mapsto \mathbb{R}$  be the potential energy of the system.

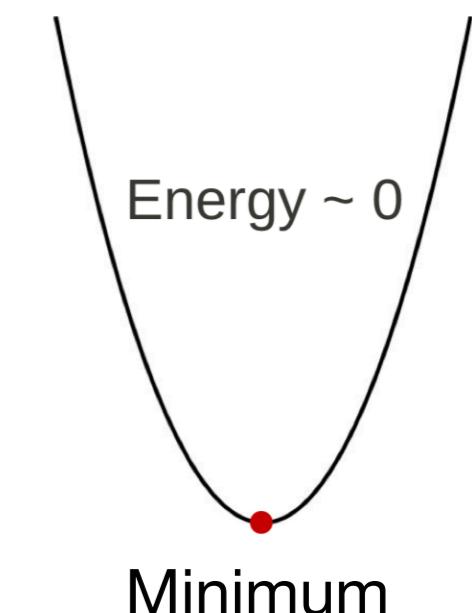
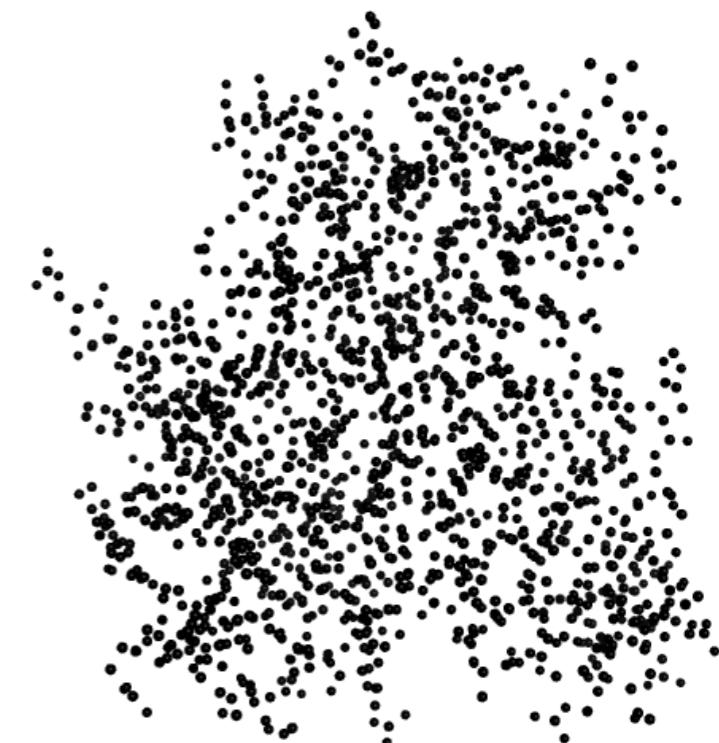
According to Newton's equation of motion, at any point  $q_0 + q$ ,  
 $F(q_0 + q) = M \times (\ddot{q}_0 + \ddot{q})$ , where  $M$  is the diagonal mass matrix  
and  $F(q_0 + q)$  represents the interatomic forces.

By definition,  $F(q_0 + q) = -\nabla V(q_0 + q)$ ,  
and thus we can write  $M \times (\ddot{q}_0 + \ddot{q}) + \nabla V(q_0 + q) = 0$

Around the equilibrium position, if  $q$  is small, then the potential energy  
can be approximated using a Taylor expansion,

$$V(q_0 + q) = \cancel{V(q_0)} + \cancel{J \times q} + \frac{1}{2} H \times q^2 + \dots,$$

where  $J$  and  $H$  are the Jacobian and Hessian matrices evaluated at  $q_0$



# Normal mode analysis

Let us consider a molecular system with  $N$  atoms at equilibrium position  $q_0 \in \mathbb{R}^{3N}$ .

Let  $V: \mathbb{R}^{3N} \mapsto \mathbb{R}$  be the potential energy of the system.

According to Newton's equation of motion, at any point  $q_0 + q$ ,  
 $F(q_0 + q) = M \times (\ddot{q}_0 + \ddot{q})$ , where  $M$  is the diagonal mass matrix  
and  $F(q_0 + q)$  represents the interatomic forces.

By definition,  $F(q_0 + q) = -\nabla V(q_0 + q)$ ,  
and thus we can write  $M \times (\ddot{q}_0 + \ddot{q}) + \nabla V(q_0 + q) = 0$

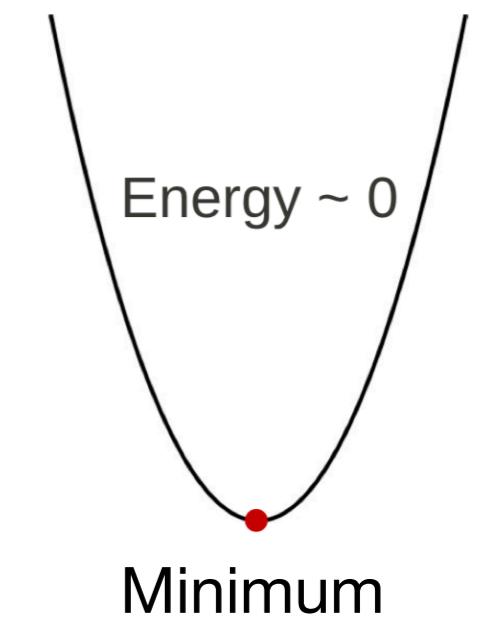
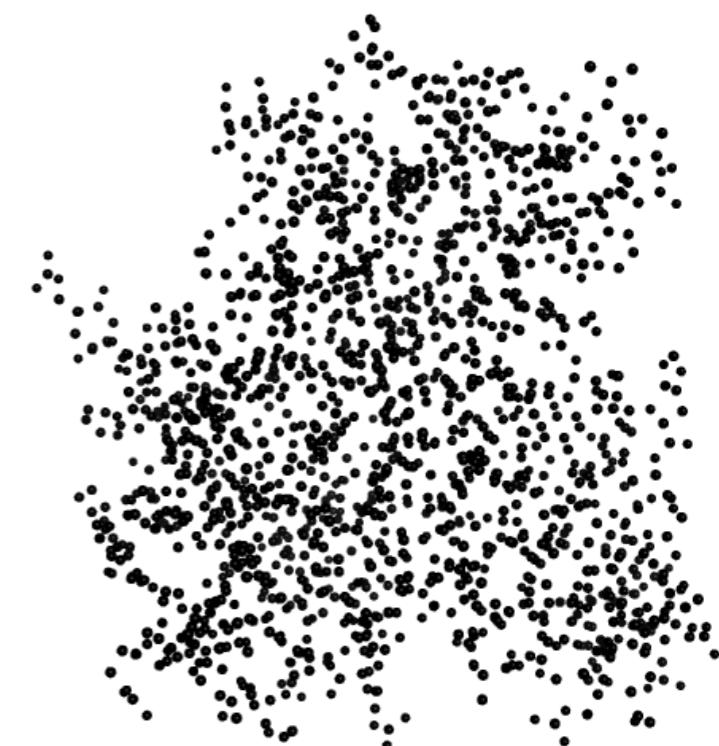
Around the equilibrium position, if  $q$  is small, then the potential energy  
can be approximated using a Taylor expansion,

$$V(q_0 + q) = \cancel{V(q_0)} + \cancel{J \times q} + \frac{1}{2} H \times q^2 + \dots,$$

where  $J$  and  $H$  are the Jacobian and Hessian matrices evaluated at  $q_0$

By replacing  $V(q_0 + q)$  in the equation of motion, we obtain,

$$M \times (\ddot{q}_0 + \ddot{q}) + \frac{1}{2} \nabla (H \times q^2) = 0.$$



# Normal mode analysis

This allows to analytically solve the equation of motion,

$$\begin{aligned} M \times (\ddot{q}_0 + \ddot{q}) + \frac{1}{2} \nabla (H \times q^2) &= 0 \\ \iff M \times (\ddot{q}_0 + \ddot{q}) + \frac{1}{2} \times 2 \times H \times q &= 0 \\ \iff M\ddot{q} + Hq + M\ddot{q}_0 &= 0 \\ \approx M\ddot{q} + Hq &= 0. \end{aligned}$$

# Normal mode analysis

This allows to analytically solve the equation of motion,

$$\begin{aligned} M \times (\ddot{q}_0 + \ddot{q}) + \frac{1}{2} \nabla (H \times q^2) &= 0 \\ \iff M \times (\ddot{q}_0 + \ddot{q}) + \frac{1}{2} \times 2 \times H \times q &= 0 \\ \iff M \ddot{q} + Hq + M \ddot{q}_0 &= 0 \\ \approx M \ddot{q} + Hq &= 0. \end{aligned}$$

We then compute the square matrix of eigenvectors  $L$  and the diagonal matrix of eigenvalues  $\Lambda$  of the mass-weighted Hessian  $H_w = M^{-1/2}HM^{-1/2}$ , such that  $H_w = L\Lambda L^T$ .

# Normal mode analysis

This allows to analytically solve the equation of motion,

$$\begin{aligned} M \times (\ddot{q}_0 + \ddot{q}) + \frac{1}{2} \nabla (H \times q^2) &= 0 \\ \iff M \times (\ddot{q}_0 + \ddot{q}) + \frac{1}{2} \times 2 \times H \times q &= 0 \\ \iff M \ddot{q} + Hq + M \ddot{q}_0 &= 0 \\ \approx M \ddot{q} + Hq &= 0. \end{aligned}$$

We then compute the square matrix of eigenvectors  $L$  and the diagonal matrix of eigenvalues  $\Lambda$  of the mass-weighted Hessian  $H_w = M^{-1/2}HM^{-1/2}$ , such that  $H_w = L\Lambda L^T$ .

By proceeding through a change of variable,  $\eta = L^T M^{1/2} q$ , one can write,

$$\begin{aligned} MM^{-1/2}L\ddot{\eta} + M^{1/2}L\Lambda L^T M^{1/2}M^{-1/2}L\eta &= 0 \\ \iff M^{1/2}L\ddot{\eta} + M^{1/2}L\Lambda\eta &= 0 \\ \iff \ddot{\eta} + \Lambda\eta &= 0 \end{aligned}$$

# Normal mode analysis

This allows to analytically solve the equation of motion,

$$\begin{aligned} M \times (\ddot{q}_0 + \ddot{q}) + \frac{1}{2} \nabla (H \times q^2) &= 0 \\ \iff M \times (\ddot{q}_0 + \ddot{q}) + \frac{1}{2} \times 2 \times H \times q &= 0 \\ \iff M \ddot{q} + Hq + M \ddot{q}_0 &= 0 \\ \approx M \ddot{q} + Hq &= 0. \end{aligned}$$

We then compute the square matrix of eigenvectors  $L$  and the diagonal matrix of eigenvalues  $\Lambda$  of the mass-weighted Hessian  $H_w = M^{-1/2}HM^{-1/2}$ , such that  $H_w = L\Lambda L^T$ .

By proceeding through a change of variable,  $\eta = L^T M^{1/2} q$ , one can write,

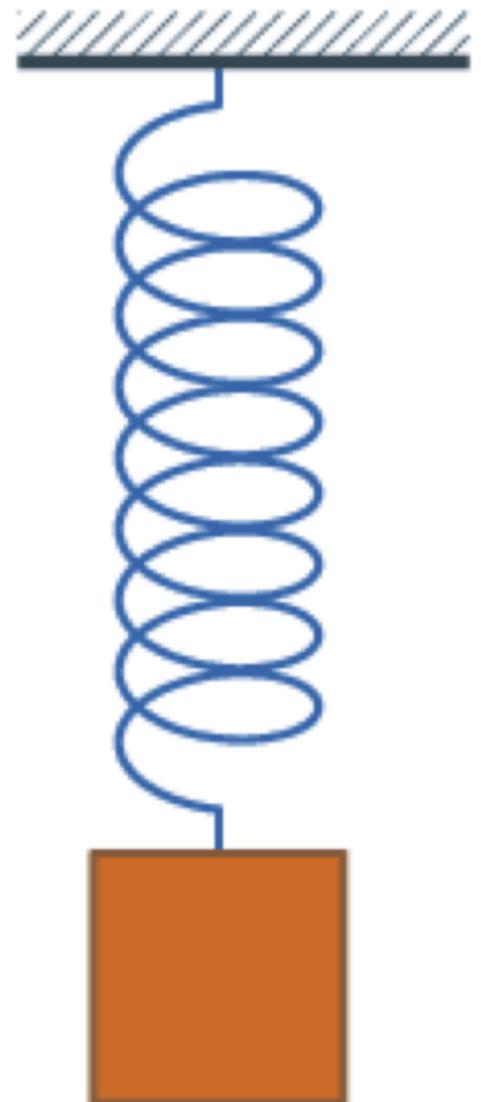
$$\begin{aligned} MM^{-1/2}L\ddot{\eta} + M^{1/2}L\Lambda L^T M^{1/2}M^{-1/2}L\eta &= 0 \\ \iff M^{1/2}L\ddot{\eta} + M^{1/2}L\Lambda\eta &= 0 \\ \iff \ddot{\eta} + \Lambda\eta &= 0 \end{aligned}$$

We thus obtain the following system of uncoupled equations,  
 $\ddot{\eta}_i + \lambda_i \eta_i = 0 \quad i = 1 \dots 3N$ , which can be solved analytically.

# Normal mode analysis

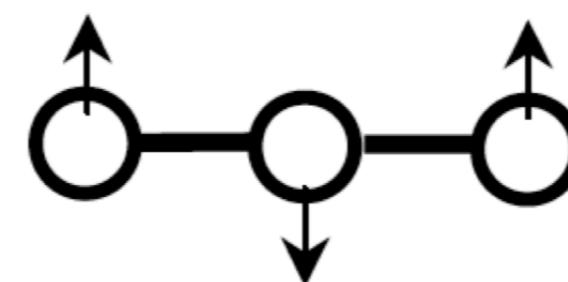
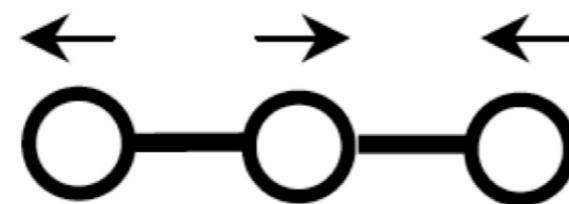
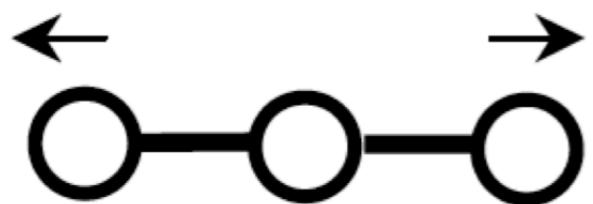
A classical harmonic oscillator is a solution of this system,

$$v(q_0 + q) = \frac{1}{2}k(q_0 + q)^2$$



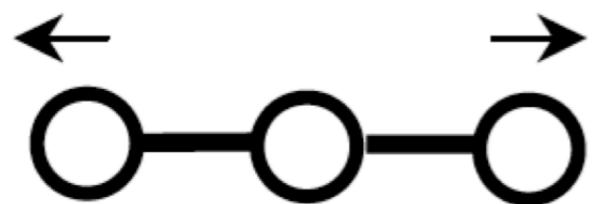
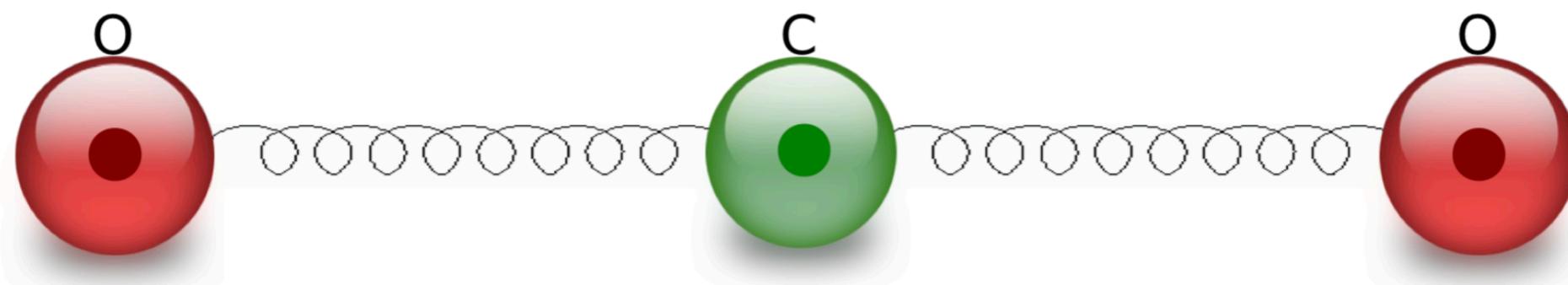
# Normal mode analysis

**Normal modes for a linear 3-atom molecule**

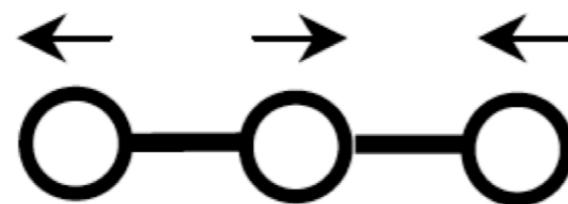


# Normal mode analysis

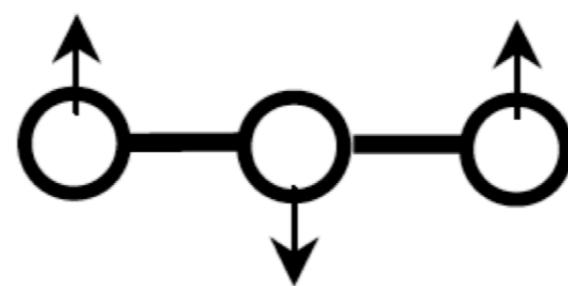
## Normal modes for a linear 3-atom molecule



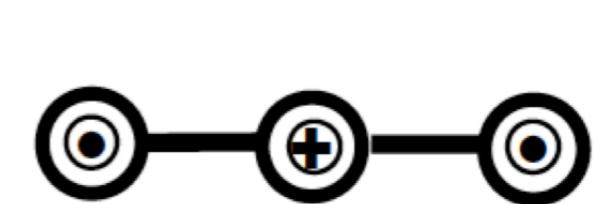
Symmetric stretch



Asymmetric stretch



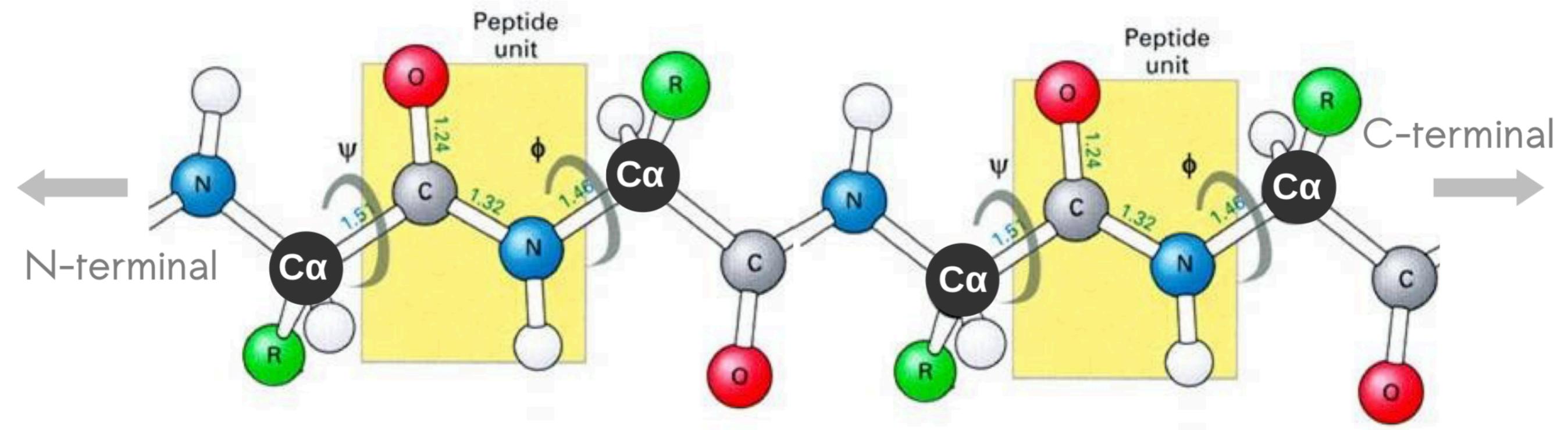
Bend



Bend

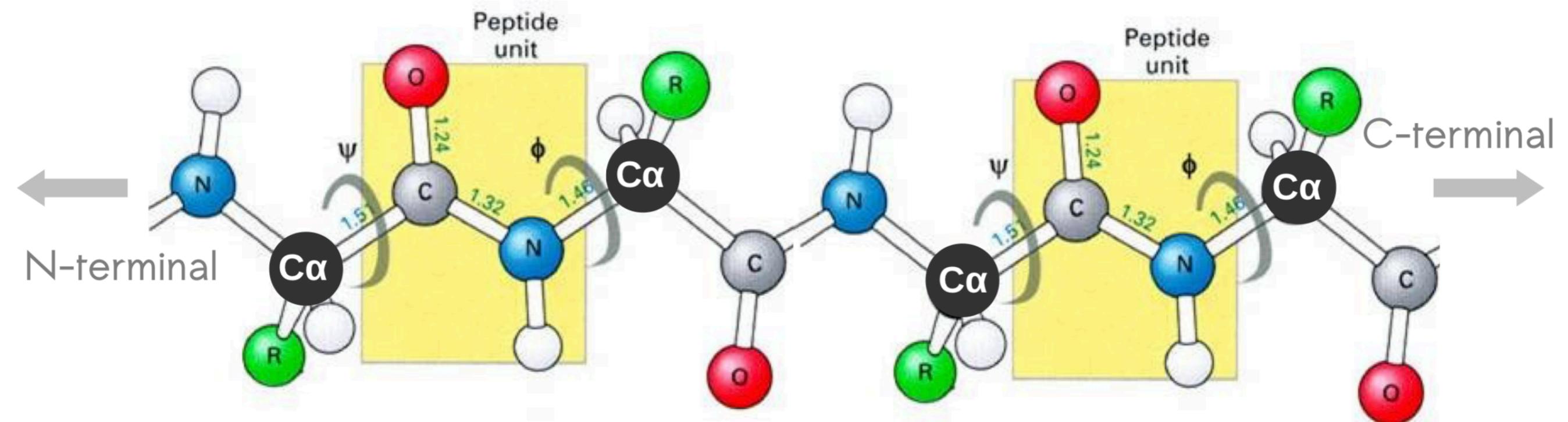
# Normal mode analysis

## All-atom ball-and-spring model



# Normal mode analysis

## All-atom ball-and-spring model



## Only C-alpha ball-and-spring model

