The background of the slide features a complex molecular dynamics simulation. It shows several protein structures in various colors (yellow, green, blue, purple) and a DNA double helix structure in grey. The structures are rendered with a semi-transparent, mesh-like appearance, highlighting their intricate shapes and interactions. The overall scene is set against a dark, teal-colored background.

Ripasso:  
**Meccanica Molecolare**  
**Dinamica Molecolare**

*Prof. R. Urbani*  
*a.a. 2021-2022*

**CALCOLO  
CONFORMAZIONALE**

Individuazione della  
struttura molecolare  
rappresentativa della catena

scelta del metodo:  
- calcolo "ab initio"  
- calcolo semiempirico  
- meccanica molecolare  
- dinamica molecolare

scelta del campo di  
forze per  $E_{tot}$

scelta delle variabili  
descrittive del sistema

**PREDIZIONE DELLE  
STRUTTURE ORDINATE**  
(stato cristallino, struttura  
secondaria dei biopolimeri,...)

scelta dei legami descrittivi  
della catena (reali o virtuali)

scelta del modello probabilistico:  
- catena Bernoulliana  
- catena Markoviana  
- catena di Ising

**STATISTICA  
CONFIGURAZIONALE**  
Stato disordinato (amorfo, in  
soluzione)

scelta del metodo:  
- metodo matriciale di Flory  
- metodo Monte Carlo

## MM

Molecular  
Mechanics



minimizzazione  
dell'energia

$$V(q)$$

$q$  = coordinate degli  
 $N$  atomi



Osservabile  $A$

$$A = f(q)$$

$\langle A \rangle$  = media su  $N$   
stati (insieme)

## MD

Molecular  
Dynamics



integrazione dell'equazione  
di Newton (in vacuo)

$$P_i = F_i \quad (i = 1, \dots, N)$$
$$F_i = \partial V(q) / \partial q_i$$


Osservabile  $A$

$$A = f(q, t)$$

$\langle A \rangle$  = media  
temporale

definizione di un campo di forza:

$$V(q) = k_b(b-b_0)^2 + k_\theta(\theta-\theta_0)^2 + k_\phi[1 + \cos(n\phi)] +$$

$$+ A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^6 + q_i q_j / r_{ij} + \dots$$

**Predizione della struttura:**

**Meccanica Molecolare (MM)**

**Dinamica Molecolare (MD)**

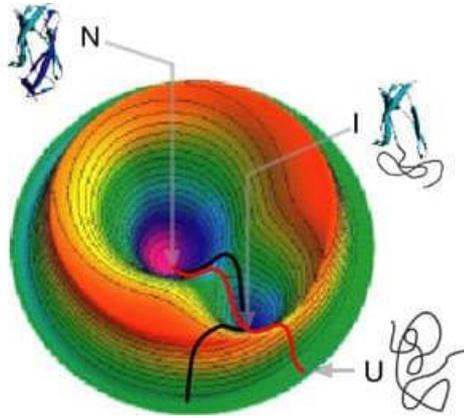
MM ed MD sono approcci computazionali molto usati sia nella scienza dei materiali che nello studio di biomolecole.

Con questi metodi viene seguita passo per passo e simulata in dettaglio la dinamica e l'energetica del sistema.

Mediante l'integrazione delle equazioni del moto, permettono di studiare la dinamica di evoluzione di un sistema fisico e chimico a livello atomico e molecolare.

Di conseguenza, le simulazioni MM/MD possono dare informazioni molto dettagliate sulla stabilità conformazionale e su fenomeni legati ai moti interni delle macromolecole ed alle transizioni conformazionali

Con la **MM/MD** possono essere studiati in particolare:



- stabilità di piccole molecole(oligomeri);
- cambi conformazionali
- polymer folding
- soluzioni polimeriche
- riconoscimento molecolare: interazioni tra segmenti polimerici (proteine, DNA, membrane)
- complessi
- trasporto ionico in sistemi biologici
- .....

La tecnica fornisce anche un valido aiuto nel design di nuovi composti complessi (es. dendrimeri, farmaci, ...) e nella determinazione di strutture di molecole complesse (in combinazione con tecniche sperimentali quali diffrazione di raggi X e NMR).

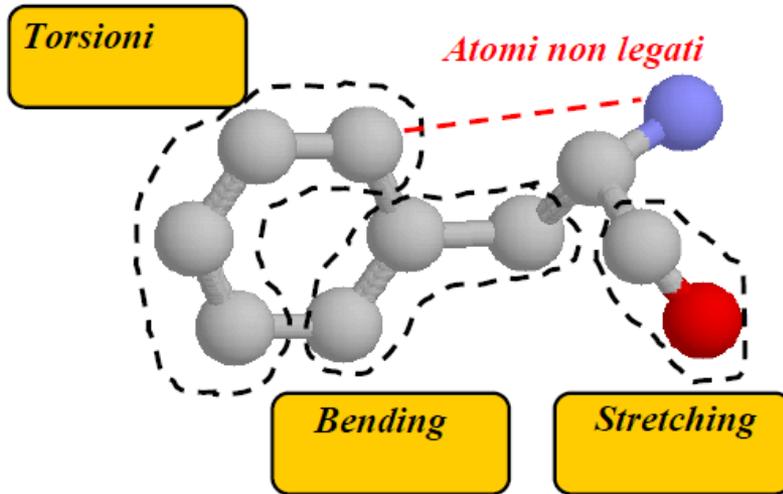
# Metodi 'classici'

- Meccanica Molecolare: ricerca del minimo energetico
- Dinamica Molecolare: simulazione del moto molecolare
- Metodi Monte Carlo: calcolo di proprietà di equilibrio

# Meccanica molecolare

**RICERCA dei MINIMI ASSOLUTI e RELATIVI  
DELL'ENERGIA CONFORMAZIONALE**

# MM



$$U = \sum_{\text{all bonds}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{all angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{\text{all torsions}} K_\phi [1 - \cos(n\phi)] + \sum_{i,j \text{ nonbonded}} \epsilon_{ij} \left[ \left( \frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j \text{ nonbonded}} \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}}$$

U è funzione della conformazione. Si vuole calcolare il minimo di U(C)

# Cutoff

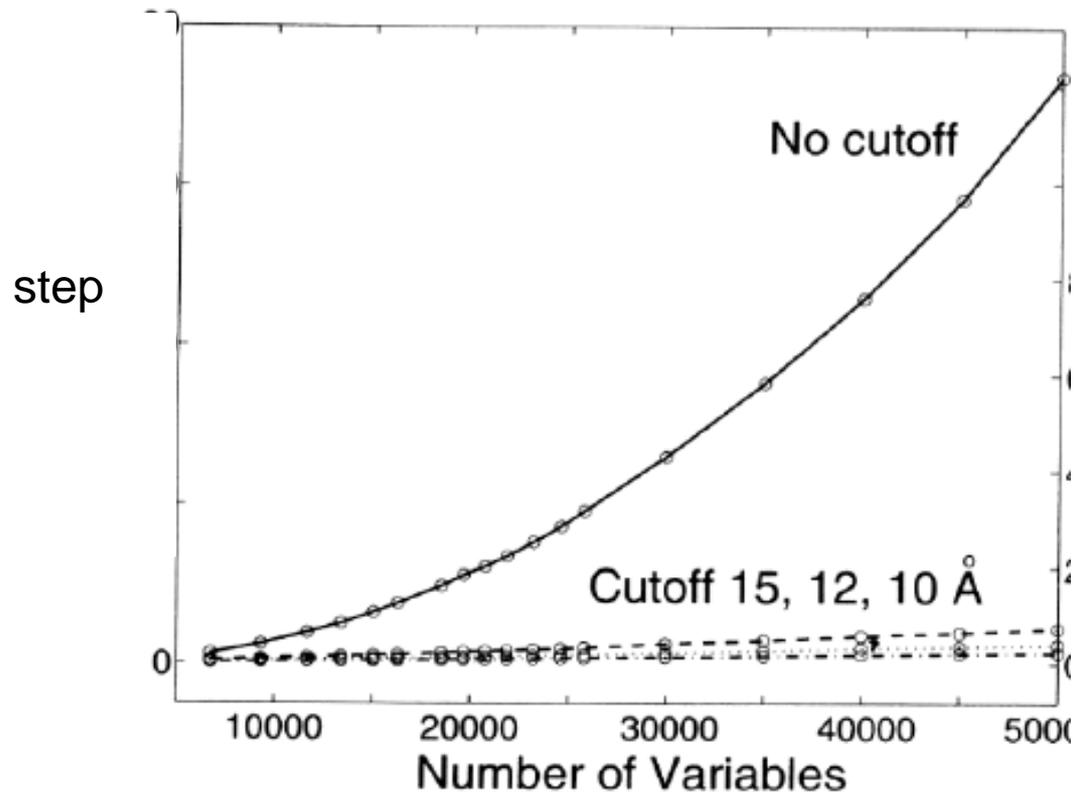
Interazioni locali (intramolecolari) crescono come  $\propto N$  dove N è il numero di atomi.

Interazioni non locali (intermolecolari) crescono come  $\propto N^2$  dove N è il numero di atomi.

$$U_{NB} = \sum_{i,j \text{ nonbonded}} \epsilon_{ij} \left[ \left( \frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j \text{ nonbonded}} \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}}$$

**Cutoff** in  $U_{NB}$  per ridurre il tempo di calcolo

# Cutoff

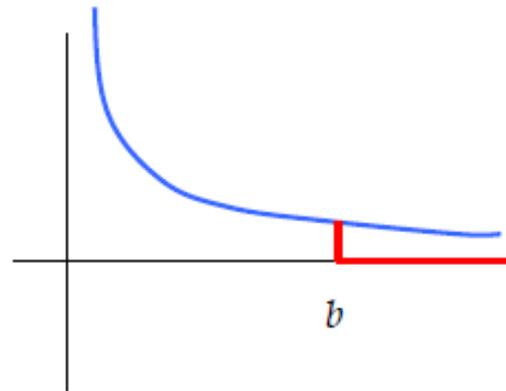


## Cutoff

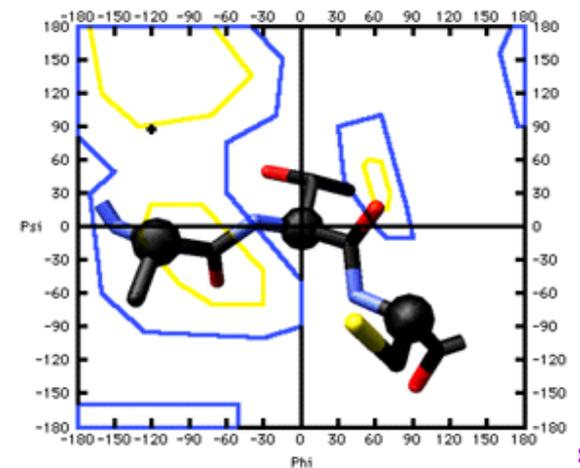
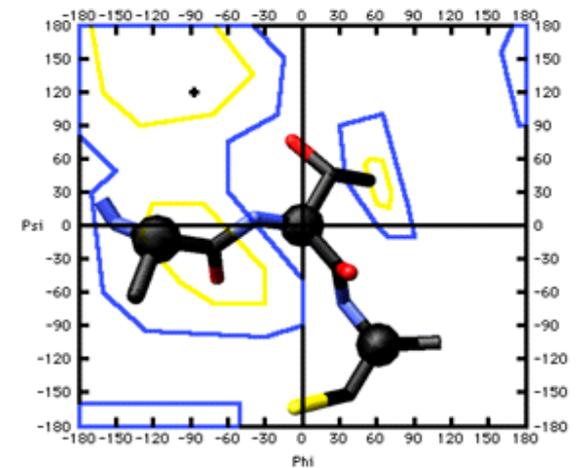
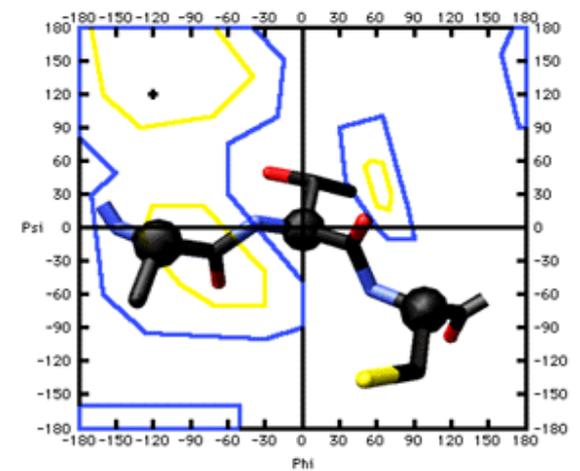
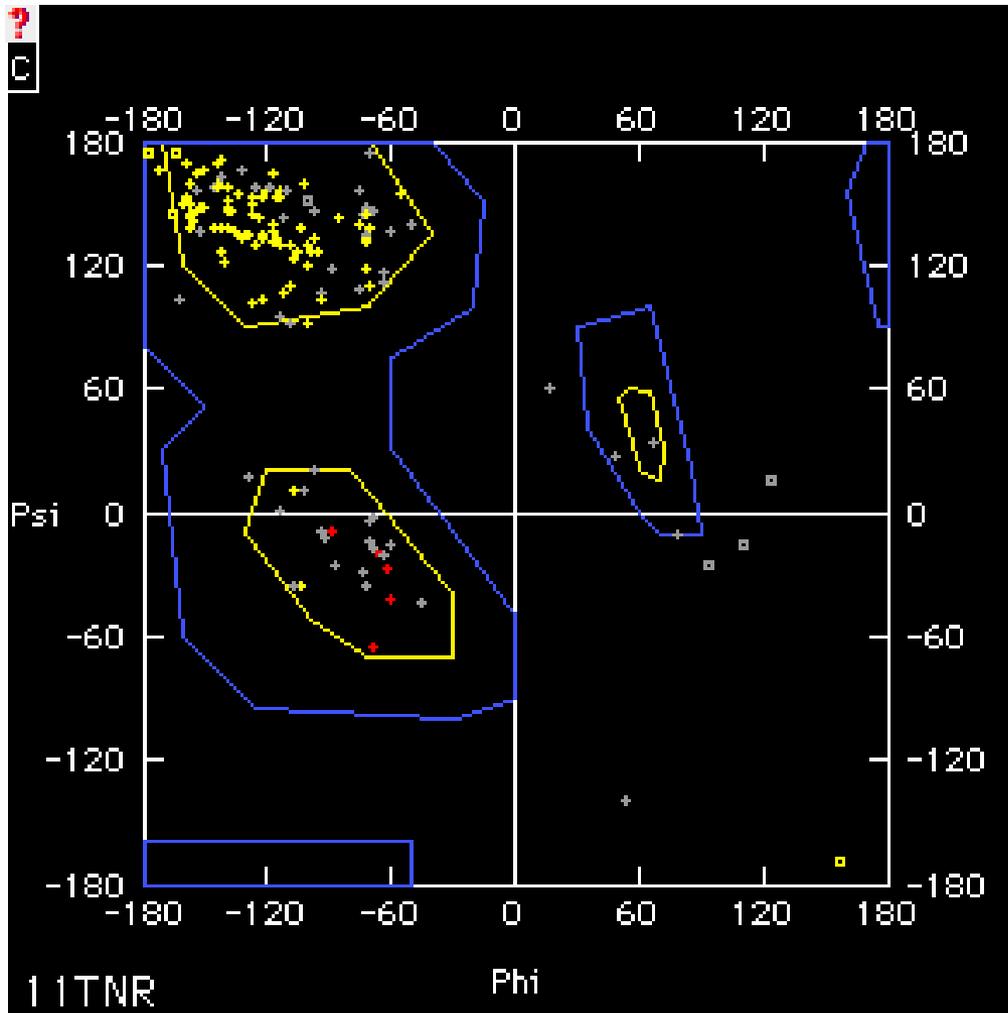
$$U_{NB} = \sum_{i,j} S(r_{ij}) \epsilon_{ij} \left[ \left( \frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j} S(r_{ij}) \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}}$$

$S(r)$  : cutoff function.

$$S(r) = \begin{cases} 1 & r < b \\ 0 & r \geq b \end{cases}$$

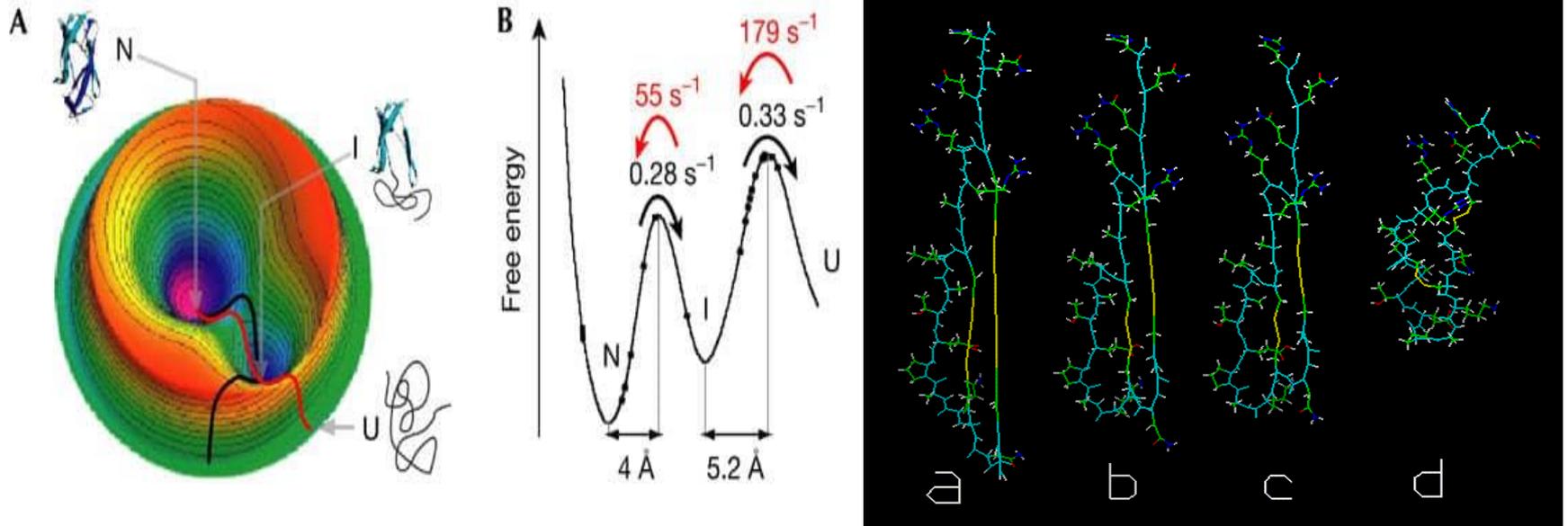


# Controllo della qualità del modello

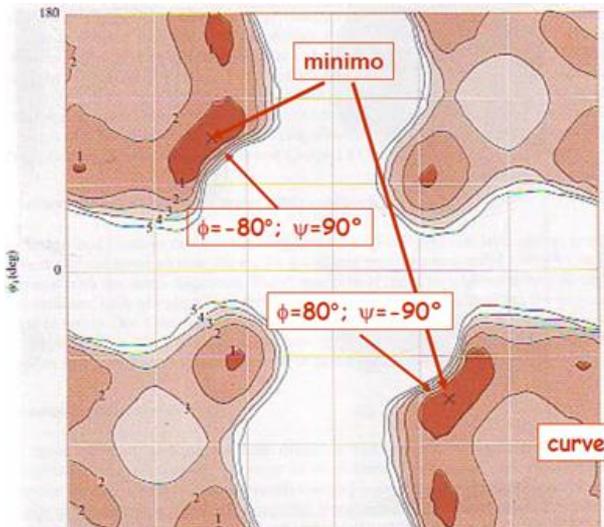


# Minimizzazione dell'energia

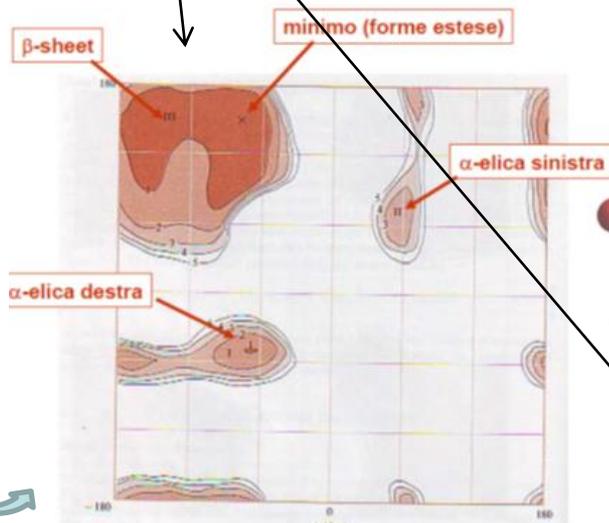
Minimizzare l'energia potenziale di una molecola significa trovare un percorso (costituito dalle variazioni dei gradi di libertà intramolecolari) che conduca da una conformazione iniziale alla conformazione a minima energia più vicina (MINIMO LOCALE), usando il minor numero di calcoli possibile.



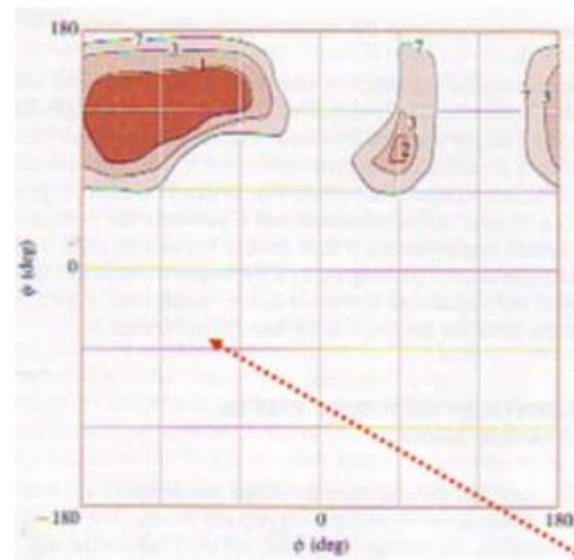
# dalla Meccanica Molecolare



glicina

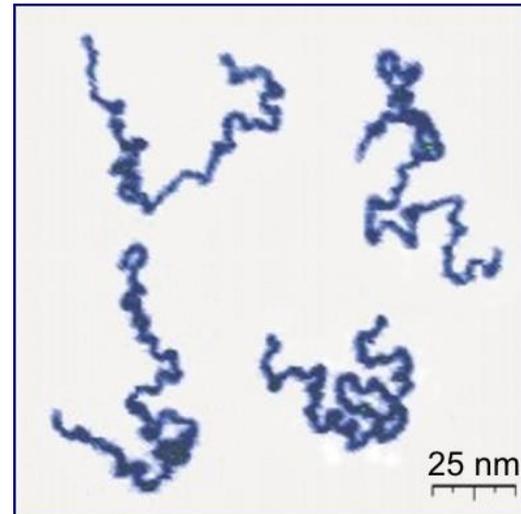
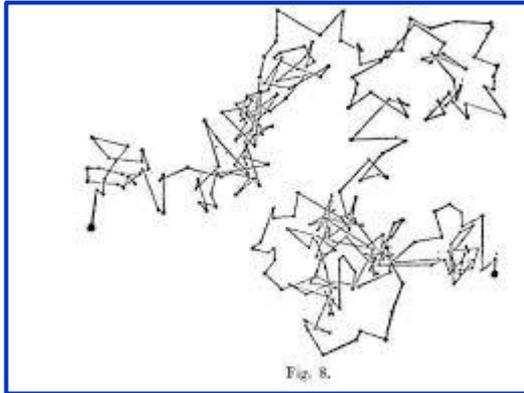


alanina



alanina-prolina

Stato solido amorfo,  
Polimero fuso,  
Polimero in soluzione



$A = f(q)$   
 $\langle A \rangle = \text{media su } N$   
stati (insieme)

## Termodinamica statistica:

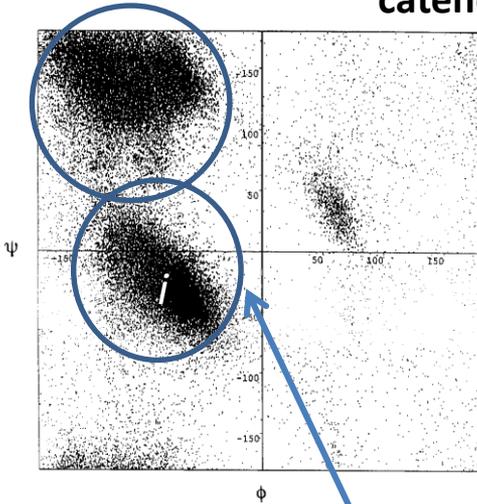
Espressione delle grandezze termodinamiche mediante la funzione di partizione

Derivazione dell'energia interna  $U$  del sistema che è uguale alla media dell'energia degli stati a volume costante:

$$U \equiv \bar{E} = \frac{\sum_j E_j e^{-\frac{E_j}{kT}}}{q} = kT^2 \left( \frac{\partial \ln q}{\partial T} \right)_V$$

# SIMULAZIONE DI CATENE BIOPOLIMERICHE

Dall'insieme di conformazioni accessibili (energie) alla valutazione delle proprietà (medie) delle catene (dimensione e forma, osservabili termodinamiche)



Ad ogni stato viene assegnata una probabilità

$$P_i(\varphi_i, \psi_i) = \frac{e^{-\frac{E_i}{RT}}}{\sum_i e^{-\frac{E_i}{RT}}}$$

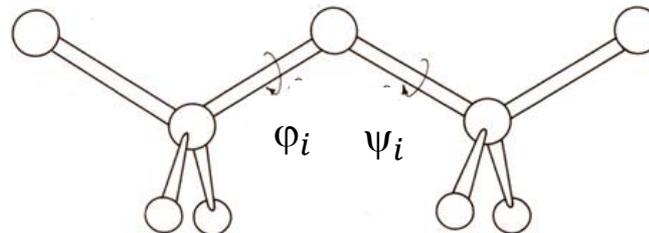
Energie degli stati conformazionali  
 $E_i = E(\varphi_i, \psi_i)$

Popolazione di conformazioni accessibili  
**(più probabili !)**

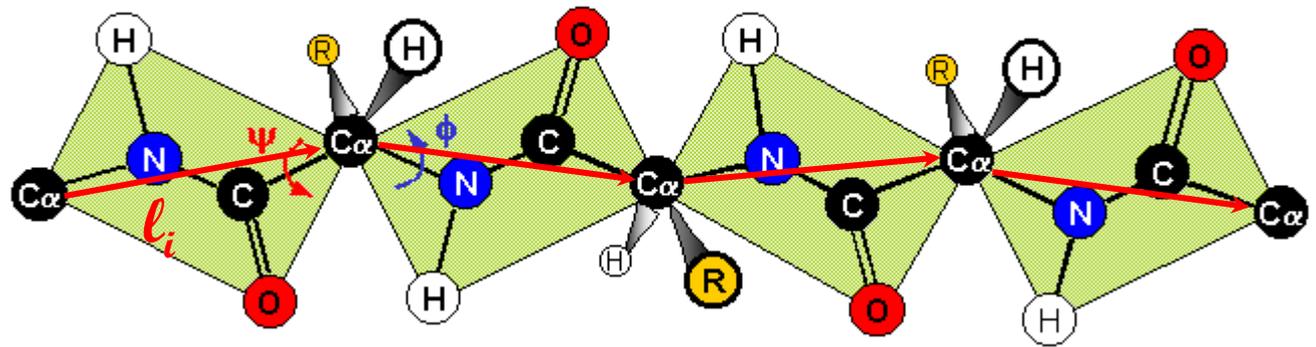
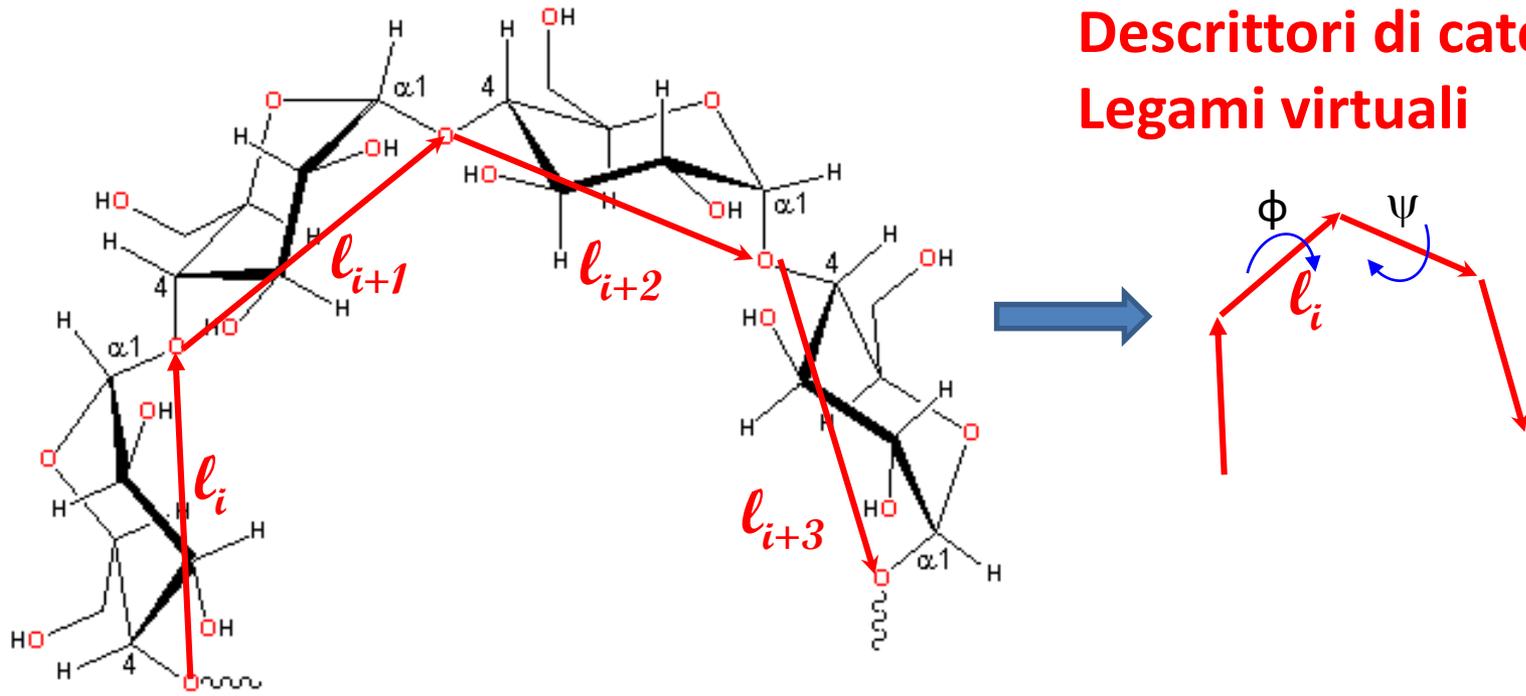
Stato solido amorfo,  
Polimero fuso,  
Polimero in soluzione  
Random coil

Insieme di catene polimeriche **più probabili** ma disordinate

$$\{(\varphi_1, \psi_1)(\varphi_2, \psi_2) \dots (\varphi_n, \psi_n)\}$$

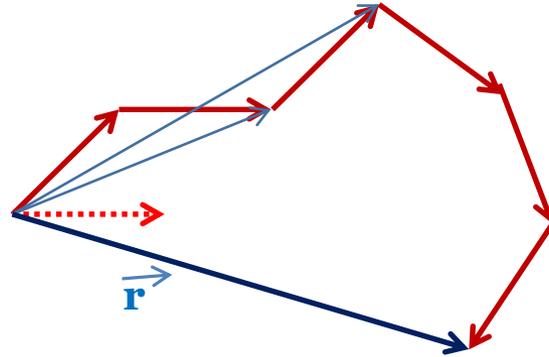
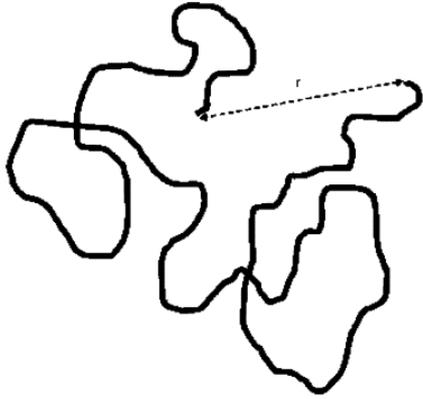


# Descrittori di catena: Legami virtuali



## Distanza quadratica media testa-coda

Distanza testa-coda di un polimero (correlata p.es. alla rigidità del polimero)



Calcolo della distanza testa-coda come somma di vettori ciascuno lungo un legame della catena (tutti i legami hanno stessa lunghezza).

$$\vec{r} = \sum_{i=1}^n \vec{l}_i$$

Il modulo del vettore  $\mathbf{r}$  è:

$$\left( \vec{r} \cdot \vec{r} \right)^{1/2}$$

$$r^2 = \sum_{i=1}^n \vec{l}_i \cdot \sum_{j=1}^n \vec{l}_j = \sum_{i=1}^n \sum_{j=1}^n \vec{l}_i \cdot \vec{l}_j$$

La doppia sommatoria è composta da termini  $i=j$  e da termini  $i \neq j$  uguali a due a due.

$$r^2 = nl^2 + 2 \sum_{j>i}^n \vec{l}_i \cdot \vec{l}_j \quad \text{mediando su diverse catene}$$

$$\langle r^2 \rangle = nl^2 + 2 \sum_{j>i}^n \langle \vec{l}_i \cdot \vec{l}_j \rangle$$

Il calcolo si riduce al prodotto di vettori (proiezione media di un legame su ogni altro legame).

## Rapporto caratteristico:

$$C_n = \frac{\langle r^2 \rangle}{nl^2} = 1 + 2/nl^2 \sum_{j>i}^n \langle \vec{l}_i \cdot \vec{l}_j \rangle$$

$$C_n = \frac{\langle r^2 \rangle}{nl^2} = 1 \quad \text{per una catena liberamente snodata con } n \text{ grande}$$

$$C_\infty = \lim_{n \rightarrow \infty} C_n$$

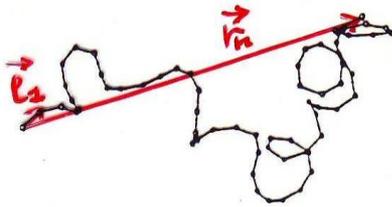
$$C_n \neq 1 \quad \text{per una catena reale}$$

***la velocità di convergenza di  $C_n$  a  $C_\infty$  è una misura della rigidità della catena polimerica.***

Questo concetto viene meglio esplicitato con la definizione della ***lunghezza di persistenza.***

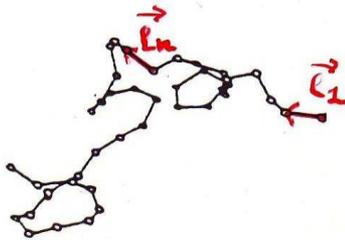
# Proprietà medie di biopolimeri

Lunghezza di persistenza



$$P_n = \left\langle \frac{\bar{l}_1}{l_1} \cdot \bar{r}_n \right\rangle$$

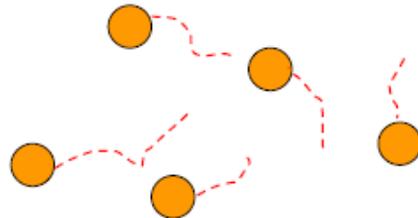
Funzione di correlazione



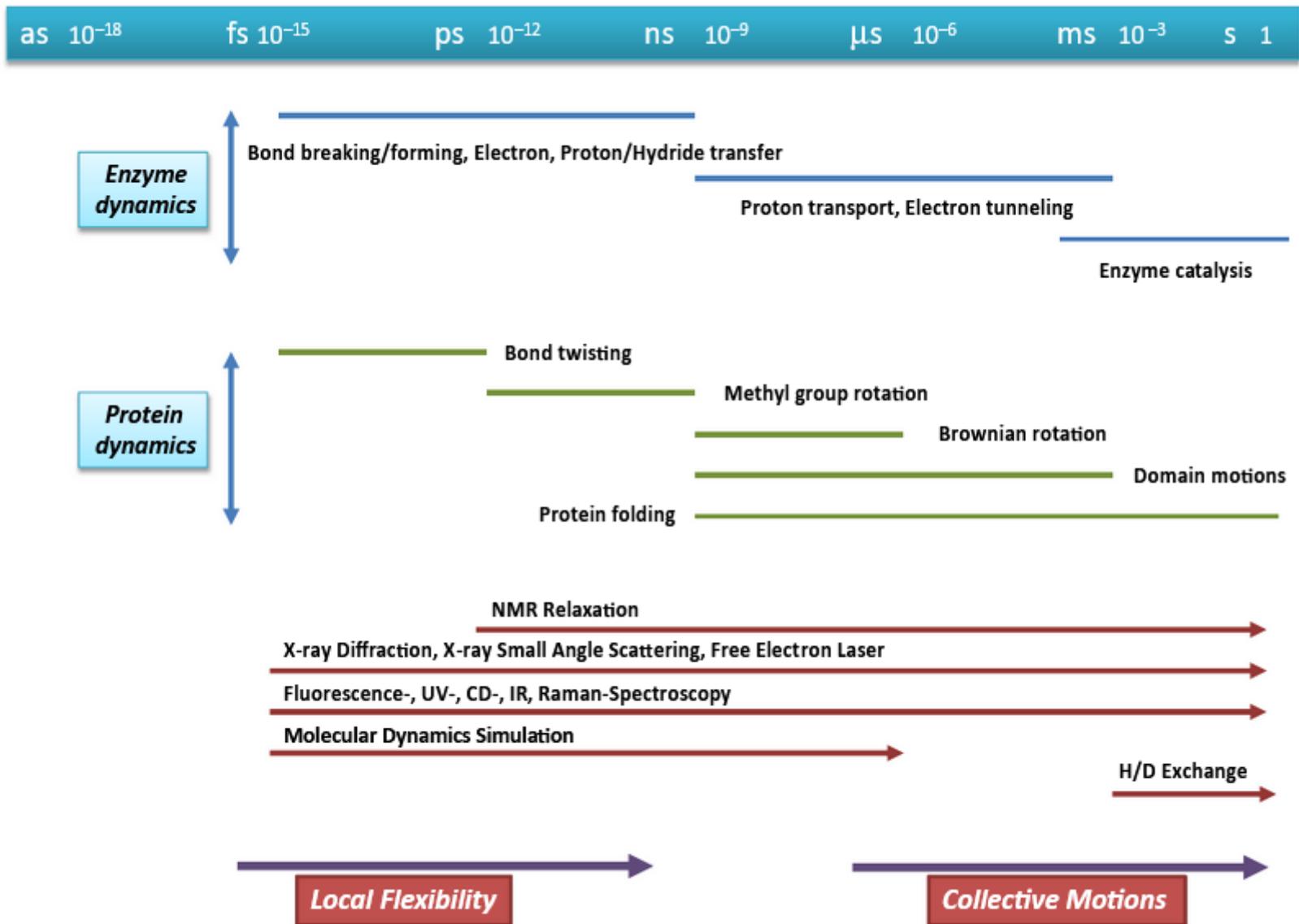
$$F_n = \left\langle \frac{\bar{l}_1}{l_1} \cdot \frac{\bar{l}_n}{l_n} \right\rangle$$

# Dinamica Molecolare

- Simulazione dinamica / moti molecolari
- Scale dei tempi tipiche: nanosecondi
- Atomi → particelle rigide classiche che seguono la II legge di Newton



**Calcola la TRAIETTORIA di un sistema molecolare =**  
la configurazione molecolare in funzione del tempo, ovvero come  
variano nel tempo le posizioni, le velocità e le accelerazioni degli  
atomi della molecola.



**Fig. 2.** Time scales of protein motions and relevant experimental/theoretical techniques for their investigation.

# MD / Es. proteina

Moto	Scala	Ampiezza
<b>Locale:</b>		
Stretching	0.01 ps	
Bending	0.1 ps	< 1 Å
rotazione metile	1 ps	
<b>Raggio medio</b>		
loop motions		
strutture secondarie	ns - μs	1-5 Å
<b>Globale</b>		
tumbling	20 ns	
(water tumbling)	(20 ps)	> 5 Å
protein folding	ms - hrs	

Perciò per calcolare una traiettoria c'è bisogno:

1. delle posizioni iniziali degli atomi (coordinate atomiche)
2. delle velocità iniziali
3. delle accelerazioni

1. le posizioni iniziali  $r_i$  si ricavano da strutture sperimentali (cristallografia raggi X, NMR ecc.) o ottenute con modeling;

2. le velocità iniziali  $v_i$  si ottengono dalla distribuzione di Boltzmann delle velocità ad una data temperatura;

3. le accelerazioni sono determinate dal gradiente dell'energia potenziale.

## MD / Schema

- **Configurazione iniziale**

Es. Coordinate di database (PDB)

- **Velocità iniziali**

Distribuzione statistica (equilibrio termico) :

$$\langle E_{kin} \rangle = \frac{1}{2} \sum_{i=1}^{3N} m_i v_i^2 = \frac{3N}{2} k_B T$$

# Preparing Your System for MD

## Solvation

Biological activity is the result of interactions between molecules and occurs at the interfaces between molecules (protein-protein, protein-DNA, protein-solvent, DNA-solvent, etc).

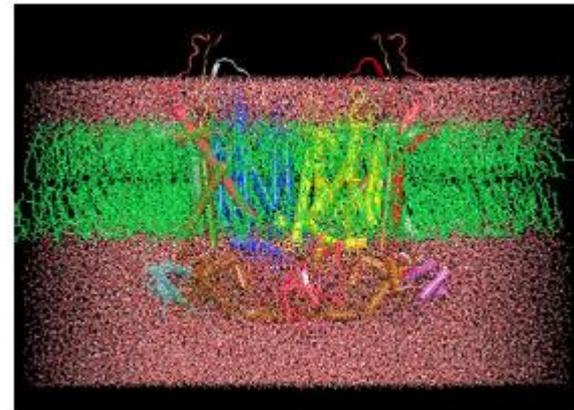
Why model solvation?

- many biological processes occur in aqueous solution
- solvation effects play a crucial role in determining molecular conformation, electronic properties, binding energies, etc

How to model solvation?

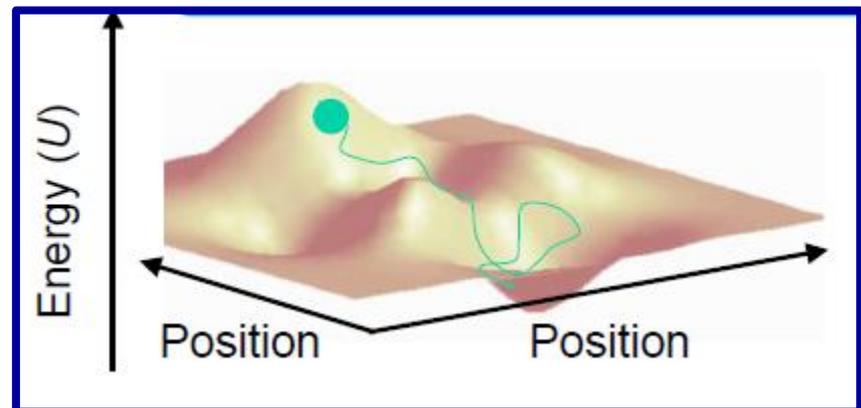
- explicit treatment: solvent molecules are added to the molecular system
- implicit treatment: solvent is modeled as a continuum dielectric or so-called implicit force field

*mitochondrial  
bc1 complex*



Gli atomi sono sempre in moto:

- Nella realtà ed in una simulazione MD, gli atomi sono in costante movimento.
- Non vanno semplicemente da uno stato ad uno di minima energia, fermandosi lì.
- Dato un tempo sufficientemente lungo, la simulazione «campiona» gli stati della distribuzione di Boltzmann.
- Ossia, la probabilità di osservare una particolare disposizione degli atomi è una funzione dell'energia potenziale.
- Spesso il tempo di simulazione non è abbastanza lungo per raggiungere tutte le conformazioni energeticamente favorevoli.



In pratica.....

Ad ogni step (circa 1 fs):

- Si calcolano le forze che agiscono su ogni atomo, usando un campo di forza della meccanica molecolare
- Si muovono gli atomi, aggiornando la posizione e la velocità di ogni atomo usando le leggi del moto di Newton

# Dinamica molecolare

L'integrazione delle equazioni del moto consente di ottenere una **traiettoria** che descrive la variazione nel tempo delle posizioni, velocità ed accelerazioni delle particelle. Le simulazioni di dinamica molecolare sono in genere calcolate sulla scala dei **ns** ( $10^{-9}$  s).

$$\vec{F}_i = m_i \vec{a}_i \quad i = 1, \dots, N \quad (\text{numero atomi del sistema})$$

$$\vec{F}_i = -\nabla_i U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = -\left(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i}\right) \quad \text{dove } U \text{ è energia potenziale}$$

Per moti uniformemente accelerati ( $a = \text{costante}$ ):

$$\vec{a}_i = d\vec{v}_i/dt \quad \Rightarrow \quad \vec{v}_i(t) = \vec{a}_i t + \vec{v}_i(t_0)$$

ma

$$\vec{v}_i = d\vec{r}_i/dt \quad \Rightarrow \quad \vec{r}_i(t) = \vec{v}_i t + \vec{r}_i(t_0)$$

quindi

$$\vec{r}_i(t) = \vec{a}_i t^2 + \vec{v}_i(t_0)t + \vec{r}_i(t_0) \quad \text{dove} \quad \vec{a}_i = - (1/m_i) \, dU/d\vec{r}_i$$

## Soluzione del sistema di equazioni del moto

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}$$

$$\frac{d\mathbf{v}}{dt} = \frac{F(\mathbf{x})}{m}$$

È un sistema di equazioni differenziali ordinario:  
per N atomi abbiamo 3N variabili di posizione e 3N variabili di  
velocità.

La soluzione analitica è impossibile!

# Algoritmo di Verlet

Il potenziale  $U$  è funzione delle posizioni di tutti gli atomi del sistema

E' a causa della complessità di tale funzione che le equazioni del moto devono essere risolte **numericamente**

E' necessario **discretizzare** le equazioni, ovvero passare ad una descrizione in cui il tempo diventa una variabile discreta.

Scegliamo un intervallo di tempo (*time step*)  $\delta t$ , piccolo ed attraverso lo sviluppo in serie di Taylor scriviamo (per una particella):

$$\begin{aligned} \mathbf{r}(t + \delta t) &= \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 + \frac{1}{6}\dot{\mathbf{a}}(t)\delta t^3 \cdot \\ \mathbf{r}(t - \delta t) &= \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 - \frac{1}{6}\dot{\mathbf{a}}(t)\delta t^3 \cdot \end{aligned}$$

Sommando e sottraendo le due relazioni e ricordando  $F = ma$  otteniamo *l'algoritmo di Verlet*:

$$\begin{aligned} \mathbf{r}(t + \delta t) &= \mathbf{r}(t - \delta t) + 2\mathbf{r}(t) + \frac{\mathbf{F}(t)}{m}\delta t^2 \\ \mathbf{v}(t) &= \frac{\mathbf{r}(t + \delta t) - \mathbf{r}(t - \delta t)}{2\delta t} + \mathcal{O}(\delta) \end{aligned}$$

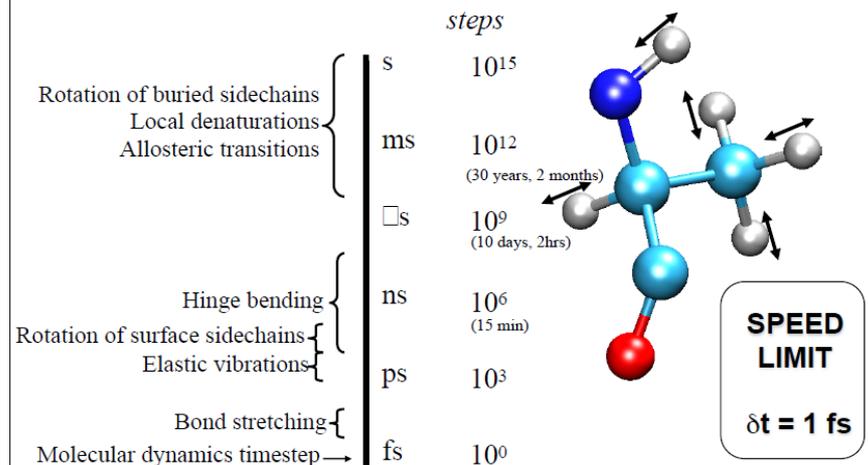
## in pratica:

- Si specificano le posizioni a  $t-dt$  e  $t$ .
- Dalle posizioni si calcolano le forze al tempo  $t$ .
- Si calcolano le posizioni al tempo  $t+ dt$ .
- Si calcolano le velocità

# Scelta del time step $\Delta t$

- ✓ Non troppo corto (ricerca inefficiente nello spazio delle fasi)
- ✓ Non troppo lungo (scarsa risoluzione)
- ✓ Di solito il time step si sceglie un ordine di grandezza più piccolo del moto più veloce
- ✓ Es. stretching C-H  $\sim 10$  fs  $\rightarrow$  time step  $\sim 1$  fs

But long is still a problem!  
*biomolecular timescale and timestep limits*



(NSF center, Shaw Res.)

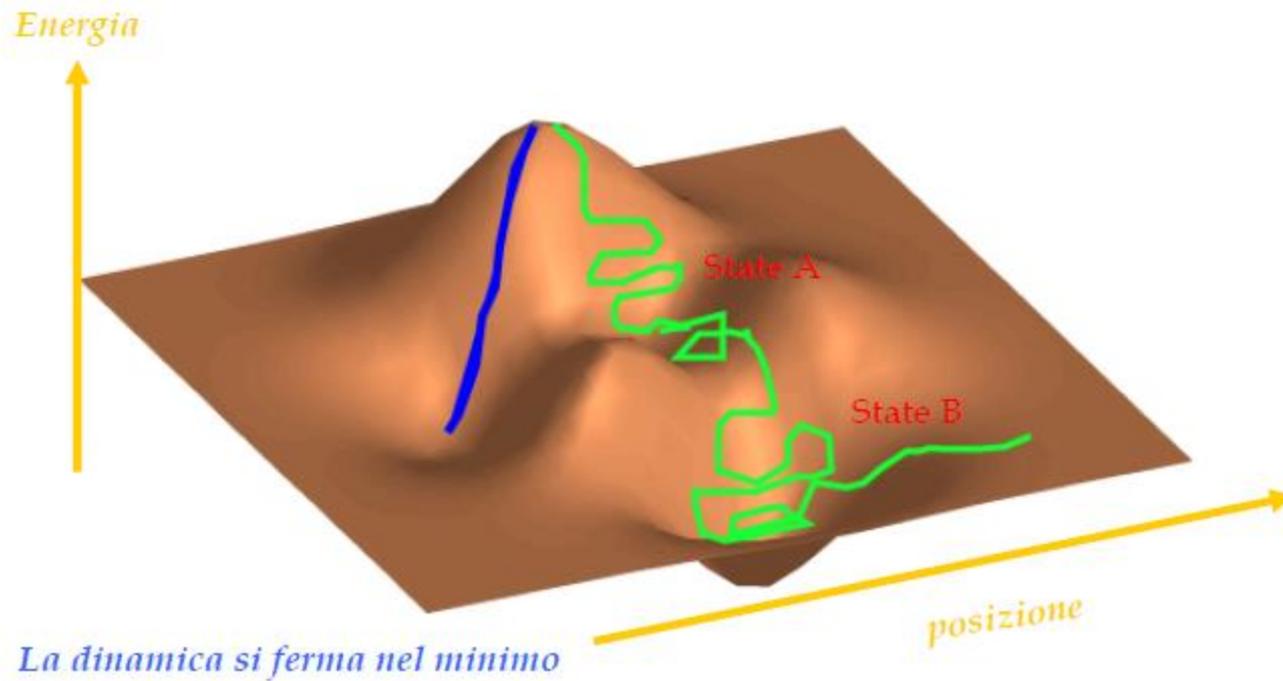
# Insiemi statistici

- Si introducono gli INSIEMI STATISTICI:

$$\frac{3}{2}Nk_bT(t) = \sum_i \frac{1}{2}m_i v_i^2(t).$$

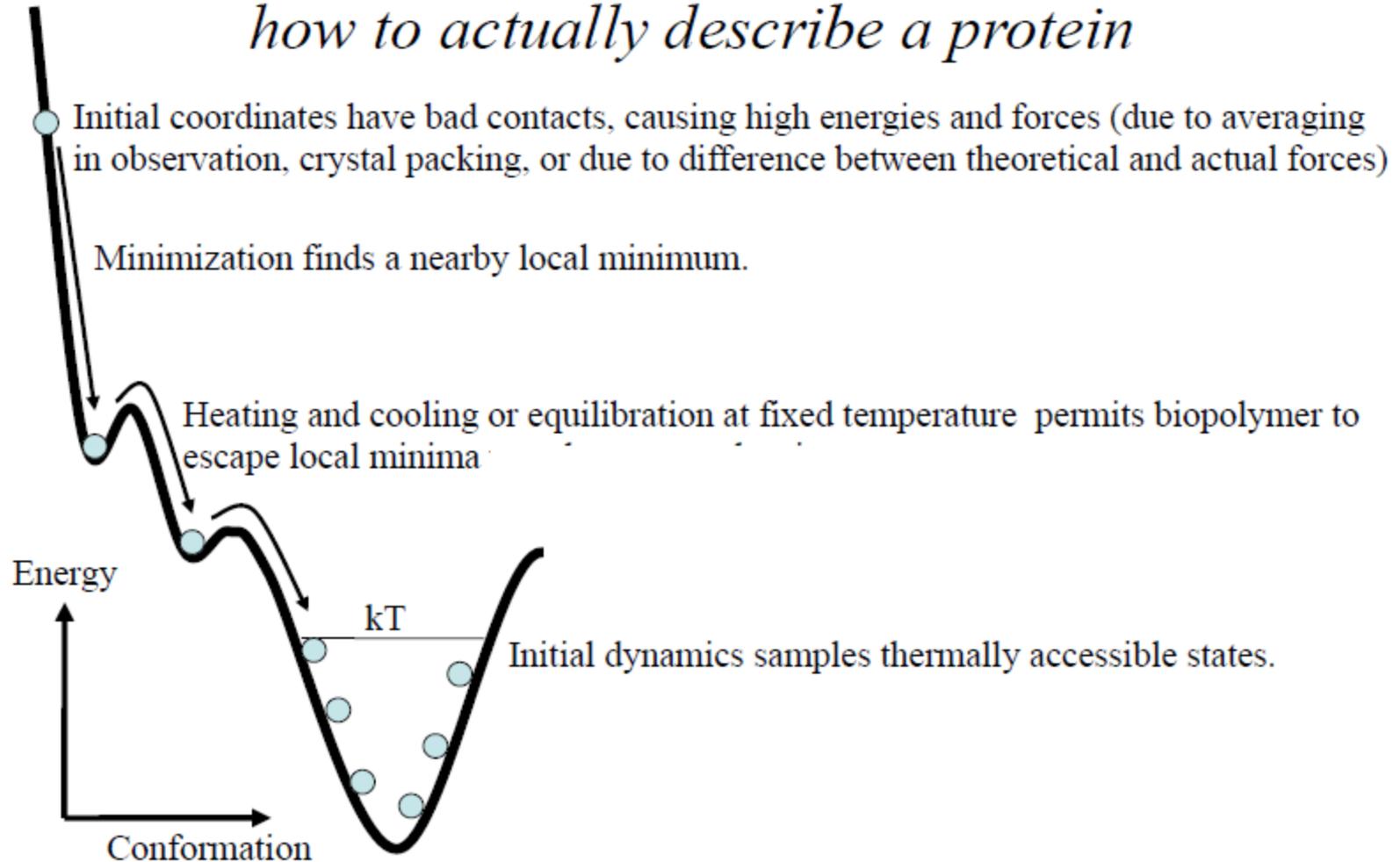
Sistema microcanonico	N, E, V costanti	per mantenere E costante la T fluttua per continua conversione di en.cinetica in potenziale.
Sistema canonico	N, T, V costanti	per mantenere la T costante, le velocità fluttuano secondo una statistica gaussiana
Sistema isotermo-isobarico	N, T, P costanti	per mantenere P costante, vengono lasciati fluttuare il volume e pertanto le distanze

## MD / ricerca di minimi



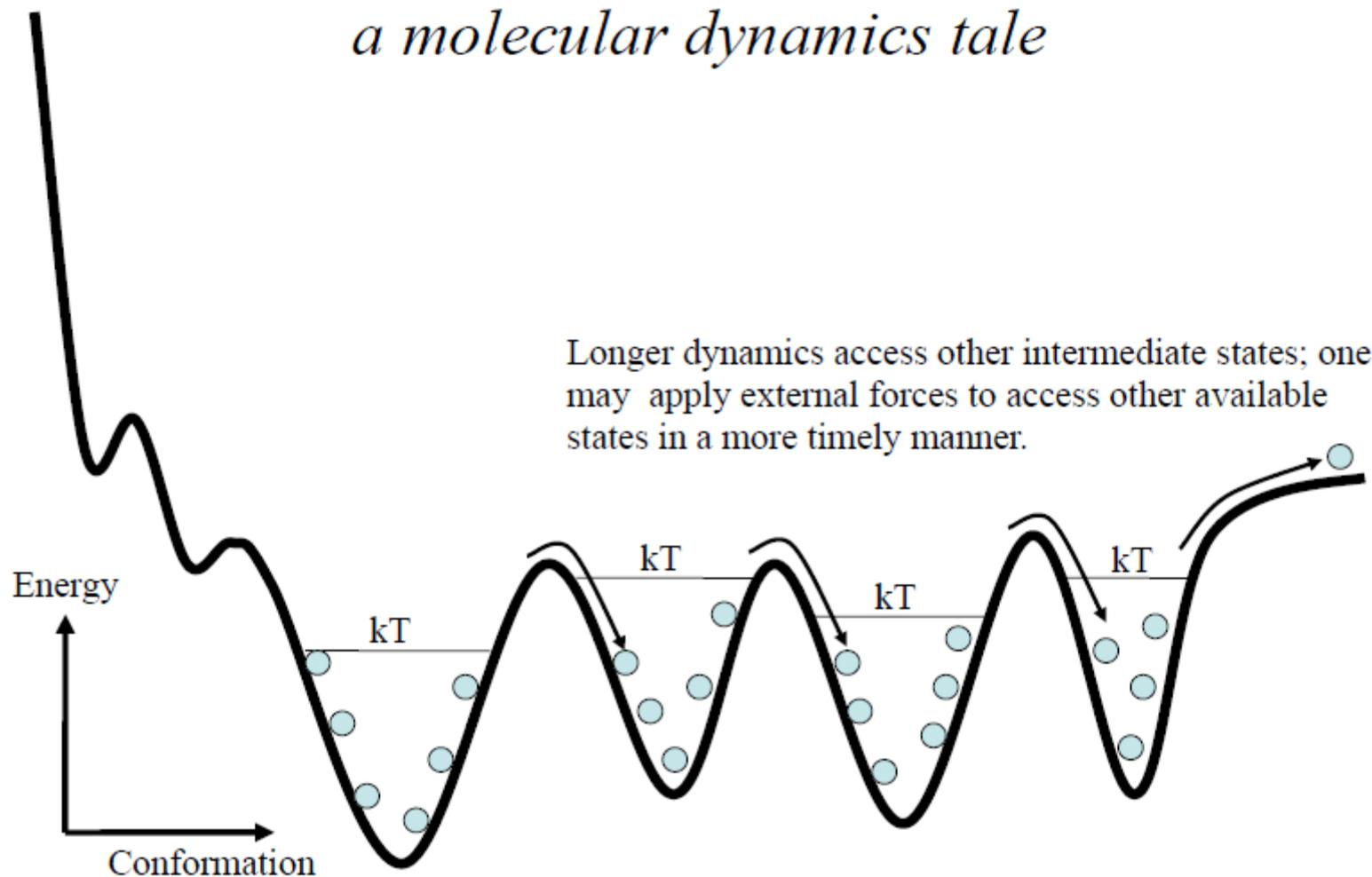
# From the Mountains to the Valleys

*how to actually describe a protein*



# From the Mountains to the Valleys

*a molecular dynamics tale*



# Limiti della MD

## Scale temporali

- Le simulazioni richiedono brevi time step per la stabilità del calcolo (1 time step  $\approx$  2 fs)
- I cambiamenti strutturali nelle proteine possono richiedere nanosecondi, microsecondi, millisecondi, o più lunghi tempi
- Per eventi su scala di ns-ms, sono richiesti milioni o trilioni ( $10^{18}$ ) di time step sequenziali. Fino a poco tempo fa, le simulazioni di 1 microsecondo erano rare
- I progressi nella potenza dei computer hanno permesso negli anni recenti simulazioni di microsecondi. Ottenere simulazioni su scala temporale più lunga è un'area di ricerca molto attiva, che implica:
  - Miglioramento degli algoritmi
  - Calcolo parallelo
  - Sviluppo di hardware; hardware specializzato

## **Accuratezza del campo di forza**

- I campi di forza della meccanica molecolare sono intrinsecamente approssimati
- Sono migliorati sostanzialmente nell'ultimo decennio, ma rimangono molte limitazioni
- In pratica, uno deve acquisire una certa esperienza per sapere come intraprendere e poi interpretare una simulazione.

## **I legami covalenti non possono rompersi o formarsi durante le simulazioni MD standard**

- Una volta creata una proteina, la maggior parte dei suoi legami covalenti non si rompono o si formano altri, durante l'espressione della funzioni biologica.
- Solo alcuni legami covalenti si formano e si rompono più frequentemente:
  - Legami disolfuro tra cisteine
  - I residui di amminoacidi acidi o basici che possono perdere o guadagnare un protone
- Tecniche più avanzate permettono simulazioni di rottura o formazione di almeno alcuni legami covalenti.

The internal motions of macromolecules of biological importance have attracted considerable interest in recent years (since 1970).

Experimental and theoretical studies of the motions have been made and their biological roles have been examined.

MD simulations provide an approach for obtaining information concerning these motions at the atomic level of detail.

By means of such simulations, both the equilibrium properties (e.g. the average structure, fluctuations and thermodynamics) and time-dependent phenomena, such as side-chain rotations, and electronic or nuclear spin dipole correlations, can be examined.

Simulations of proteins are of particular interest because they are complex inhomogeneous systems with properties intermediate between those of liquids and solids; they have dense atom packing, covalent bonding along the polypeptide chain, and weak van der Waals and electrostatic interactions between atoms that are non bonded.

This suggests that their dynamic behavior is likely to involve a multiplicity of time scales

# Energy conservation

- Total energy (potential + kinetic) should be conserved
  - In atomic arrangements with lower potential energy, atoms move faster
  - In practice, total energy tends to grow slowly with time due to numerical errors (rounding errors)
  - In many simulations, one adds a mechanism to keep the temperature roughly constant (a “thermostat”)

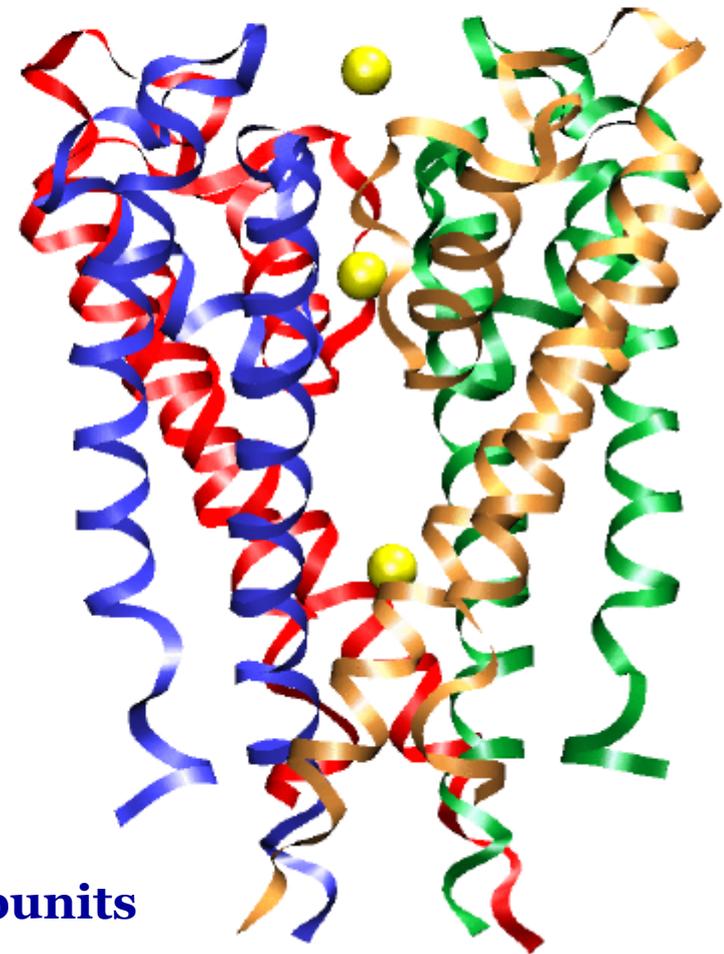
# Water is important

- Ignoring the solvent (the molecules surrounding the molecule of interest) leads to major artifacts
  - Surrounding molecules include: Water, salt ions (e.g., sodium, chloride), lipids of the cell membrane
- Two options for taking solvent into account
  - Explicitly represent solvent molecules
    - High computational expense but more accurate
    - Usually assume periodic boundary conditions (a water molecule that goes off the left side of the simulation box will come back in the right side, like in PacMan)
  - Implicit solvent
    - Mathematical model to approximate average effects of solvent
    - Less accurate but faster

# Example: MD Simulations of the K<sup>+</sup> Channel Protein

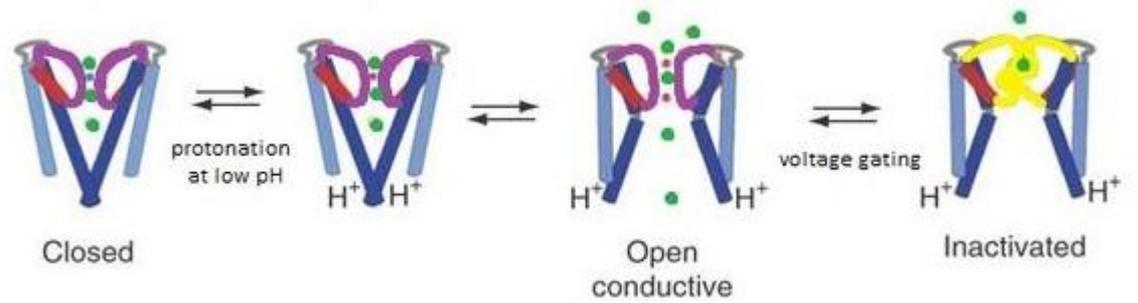
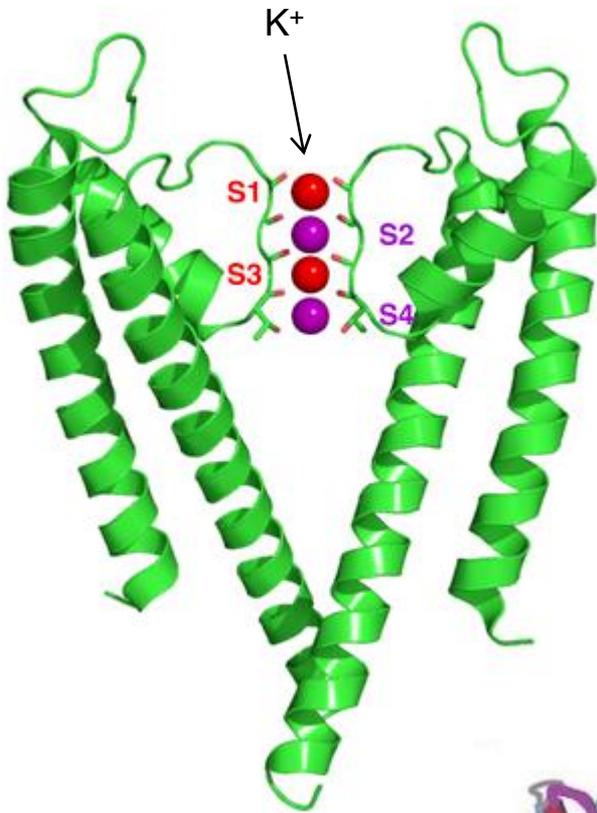
Ion channels are membrane - spanning proteins that form a pathway for the flux of inorganic ions across cell membranes.

Potassium channels are a particularly interesting class of ion channels, managing to distinguish with impressive fidelity between K<sup>+</sup> and Na<sup>+</sup> ions while maintaining a very high throughput of K<sup>+</sup> ions when gated.

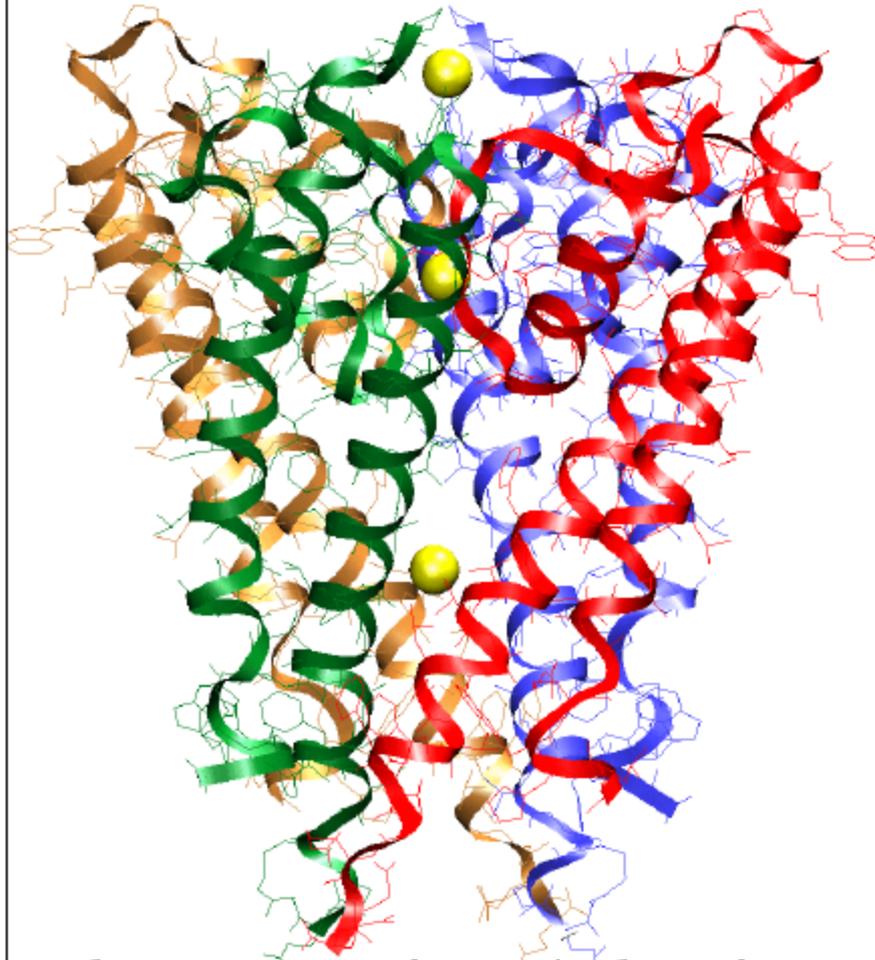


**4 subunits**

2 subunità



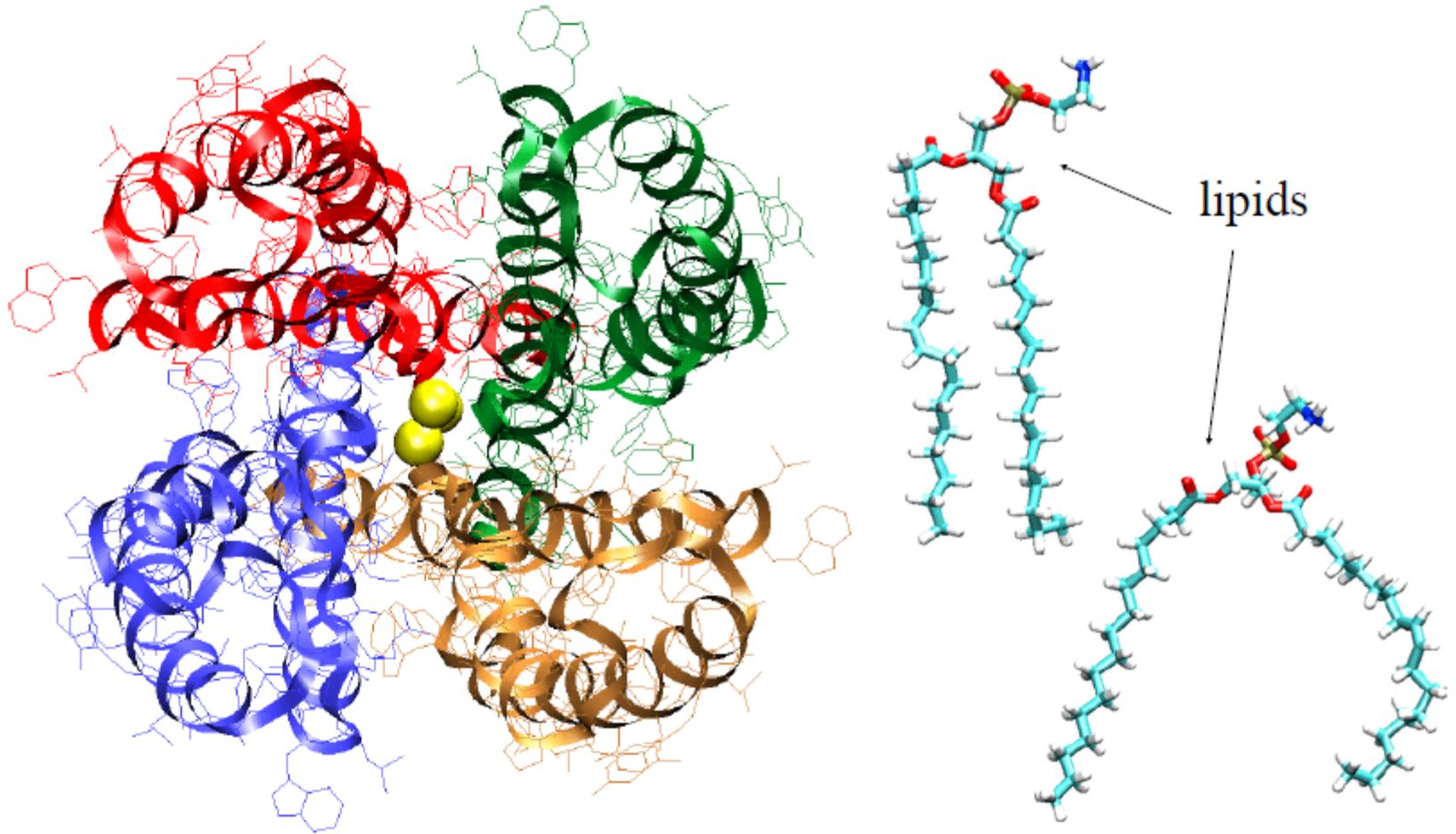
# Setting up the system



- retrieve the PDB (coordinates) file from the Protein Data Bank
- add hydrogen atoms
- use psf and parameter files to set up the structure;
- minimize the protein structure

The structure of KcsA is that of an inverted [cone](#),

# Setting up the system

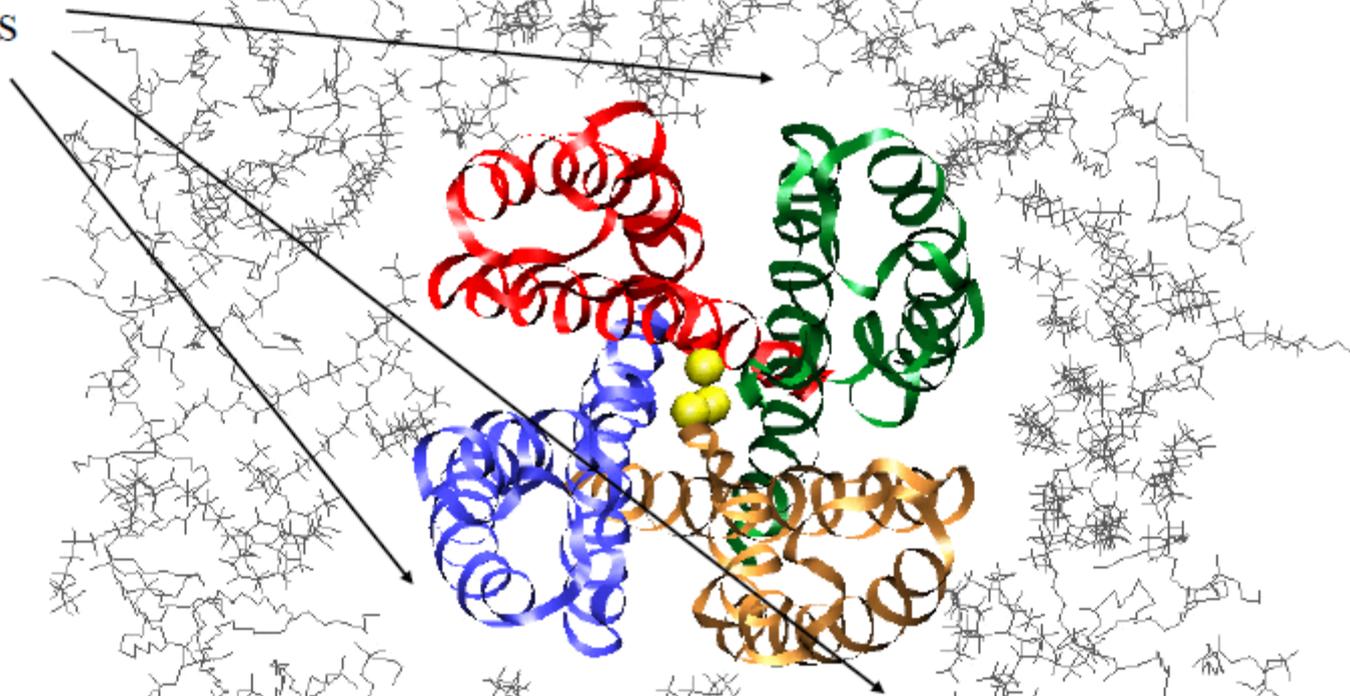


Simulate the protein in its natural environment: solvated lipid bilayer

# Setting up the system

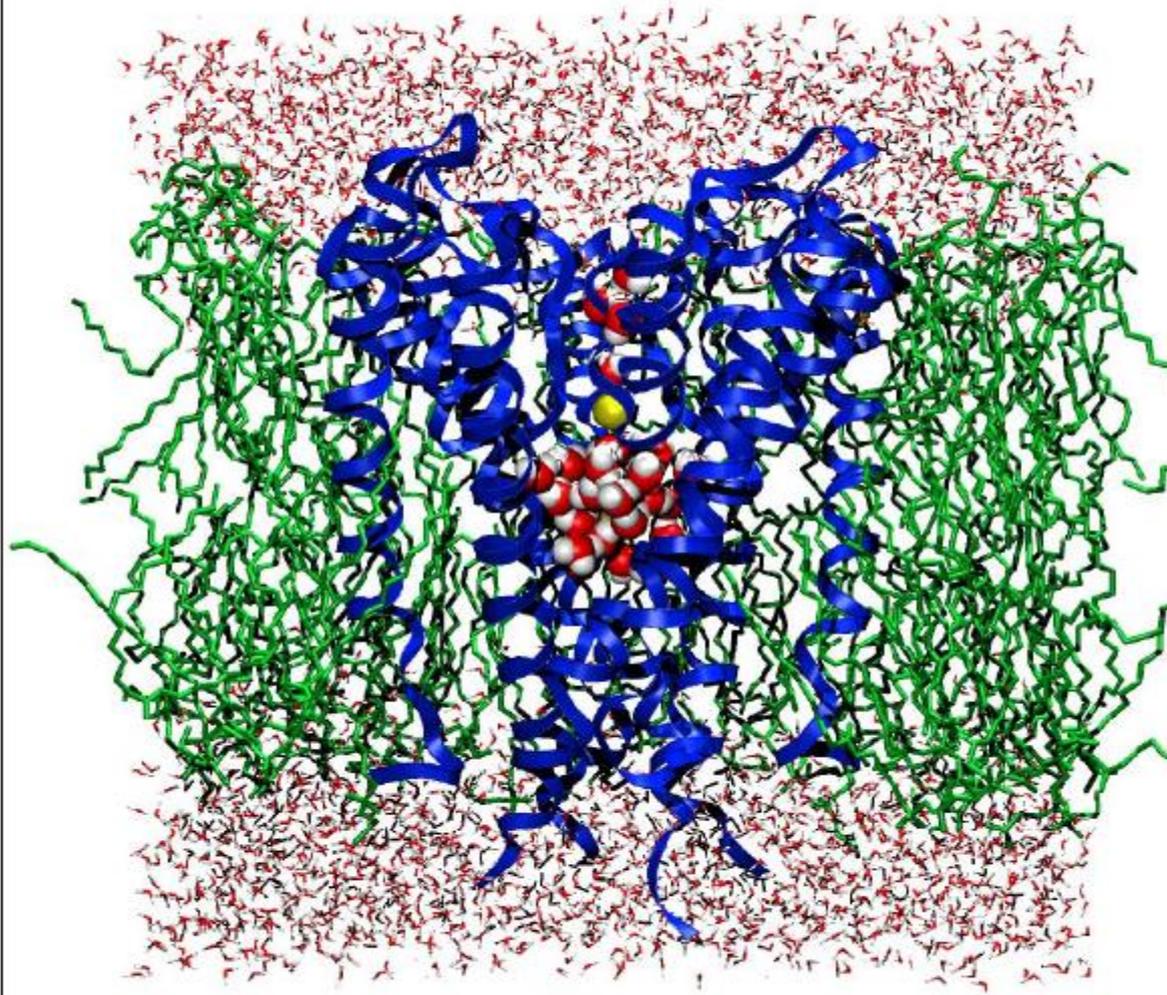
## Inserting the protein in the lipid bilayer

gaps



**Automatic insertion into the lipid bilayer leads to big gaps between the protein and the membrane => long equilibration time required to fill the gaps. Solution: manually adjust the position of lipids around the protein. Employ constant (lateral and normal) pressure control.**

# The system



---

solvent

---

KcsA channel protein (in blue) embedded in a (3:1) POPE/POPG lipid bilayer. Water molecules inside the channel are shown in vdW representation.

---

solvent

---

1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE).

# Simulating the system: Free MD

Summary of simulations:

- protein/membrane system contains 38,112 atoms, including 5117 water molecules, 100 POPE and 34 POPG lipids, plus  $K^+$  counterions
- CHARMM26 forcefield
- periodic boundary conditions, PME electrostatics
- 1 ns equilibration at 310K, NpT
- 2 ns dynamics, NpT

PME properties  
are used to  
specify  
parameters such  
as real space  
cutoff

Program: NAMD2

Platform: Cray T3E (Pittsburgh Supercomputer Center) | or local  
computer cluster; choose ~1000 atoms per processor.

## **B factor (Debye-Waller factor)**

It is a temperature factor, or atomic displacement parameter, used in protein crystallography to describe the attenuation of X-ray or neutron scattering caused by **thermal motion**.

B-factors, available from the Protein Data Bank, can be used to identify the **flexibility** of atoms, side chains, or even whole regions.

The identification and interpretation of rigidity, flexibility, and/or **internal motion** which are crucial in enzymes and in proteins in general.

**High B-factors indicate higher flexibility** as opposed to low B-factors which were believed to occur at more rigid positions.

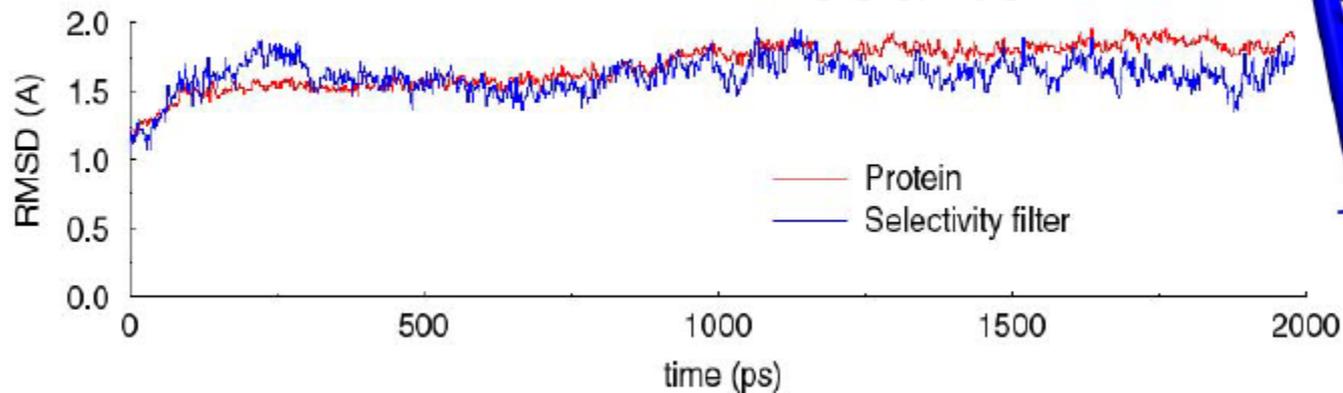
It was found that in all cases the **active site residues occur mainly in regions of low B-factors**, while the residues lining the binding pocket tend to exist in higher B-factor regions.

Residues in the active site were found to have lower B-factors (less flexible) compared to non-active site residues. This seems to be a general rule based on computer programs, which has been applied **to predict active sites** in enzymes.

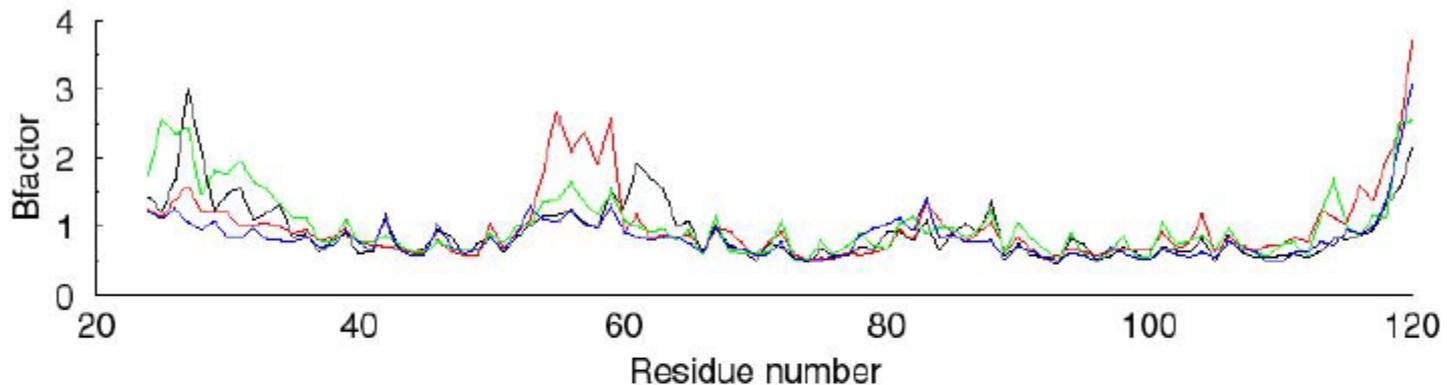
## **B factor**

In summary, numerous studies support the current general view that B-factors are **indicators of the relative vibrational motion of atoms in a protein**, those with low values belonging to a well-ordered site, and those with the highest values being part of the most flexible residues or regions.

# MD Results

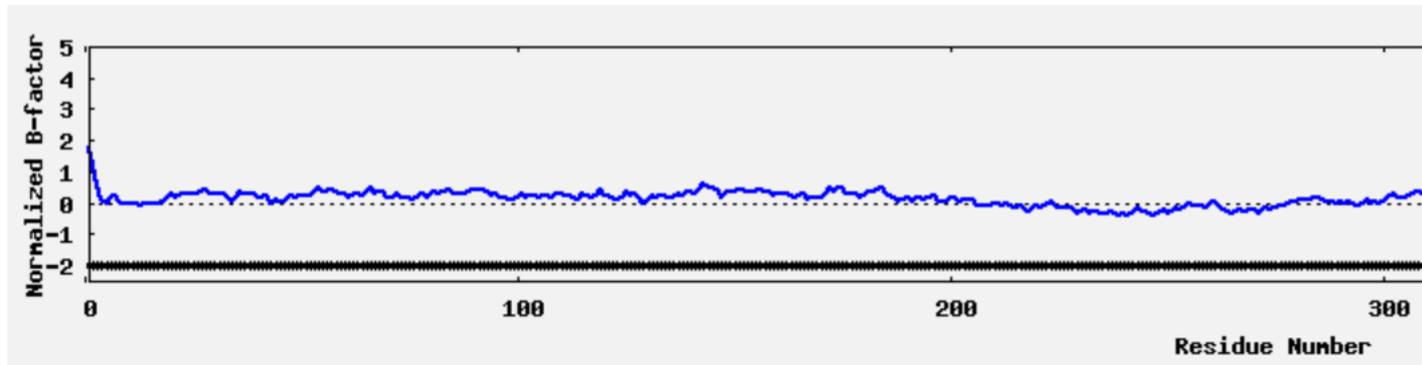


RMS deviations for the KcsA protein and its selectivity filter indicate that the protein is stable during the simulation with the selectivity filter the most stable part of the system.



Temperature factors for individual residues in the four monomers of the KcsA channel protein indicate that the most flexible parts of the protein are the N and C terminal ends, residues 52-60 and residues 84-90. Residues 74-80 in the selectivity filter have low temperature factors and are very stable during the simulation.

## I-TASSER results for protein HELP

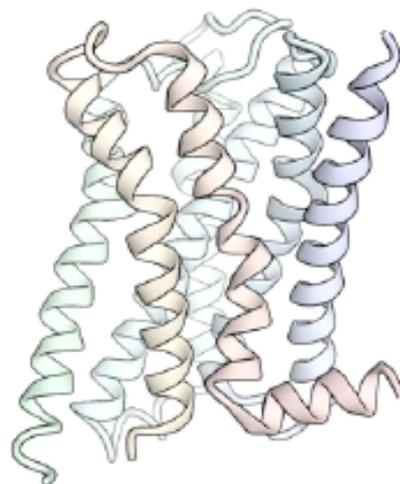
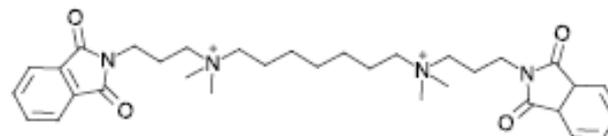
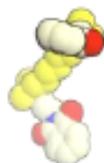


B-factor profile

Altri esempi MD:

## Determining where drug molecules bind, and how they exert their effects

0.00 us



Si sono usate simulazioni per determinare dove questa molecola si lega al suo recettore, e come cambia la forza di legame delle molecole, anche cambiando la struttura.

Dror et al., *Nature* 2013

# Determining functional mechanisms of proteins

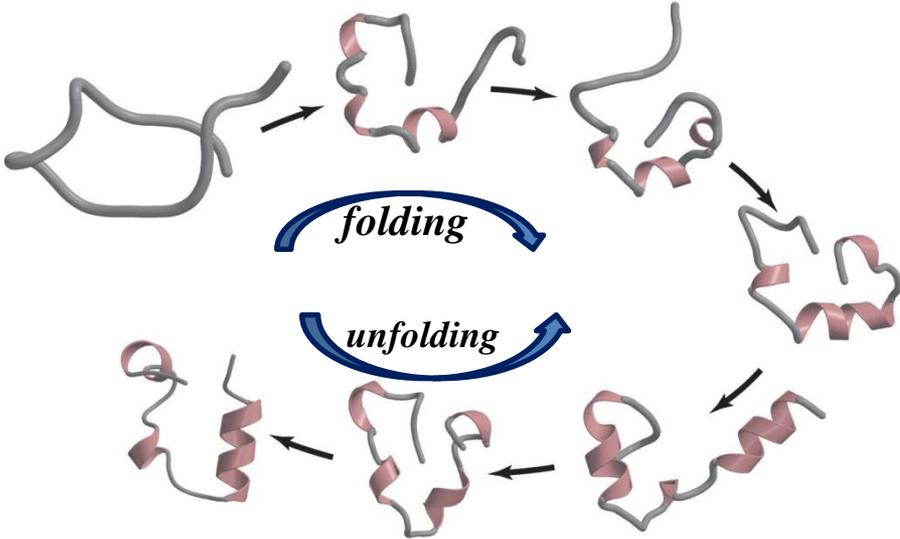
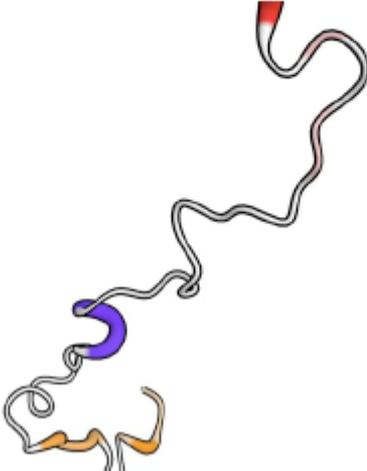


La simulazione partiva dalla struttura attiva verso una struttura inattiva

Si sono eseguite simulazioni in cui un recettore dell'adrenalina transiva spontaneamente dalla sua struttura attiva alla sua struttura inattiva.

Si sono usato queste simulazioni per descrivere il meccanismo con cui i farmaci che si legano ad un'estremità del recettore inducono l'altra estremità del recettore a cambiare forma (attivazione)

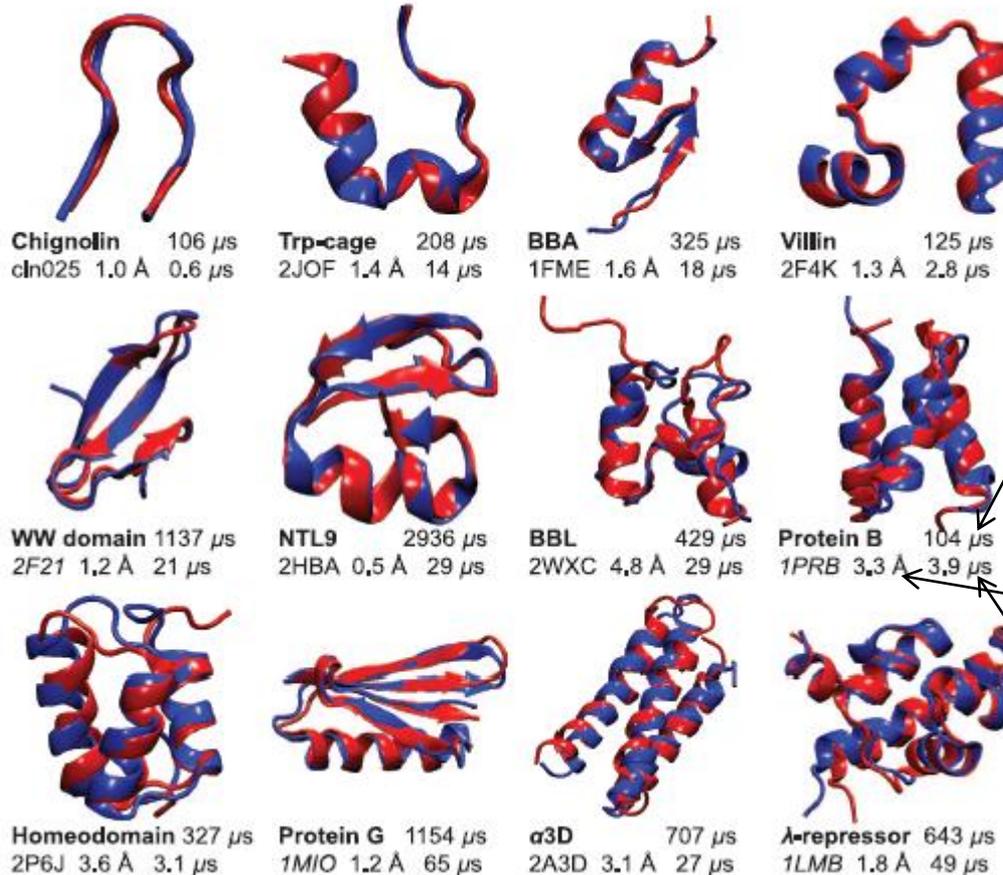
# Understanding the *process* of protein folding



# How Fast-Folding Proteins Fold

Kresten Lindorff-Larsen,<sup>1\*</sup>† Stefano Piana,<sup>1\*</sup>† Ron O. Dror,<sup>1</sup> David E. Shaw<sup>1,2</sup>†

*Science*, 2011



folded structure obtained from **simulation (blue)** superimposed on the **experimentally determined structure (red)**

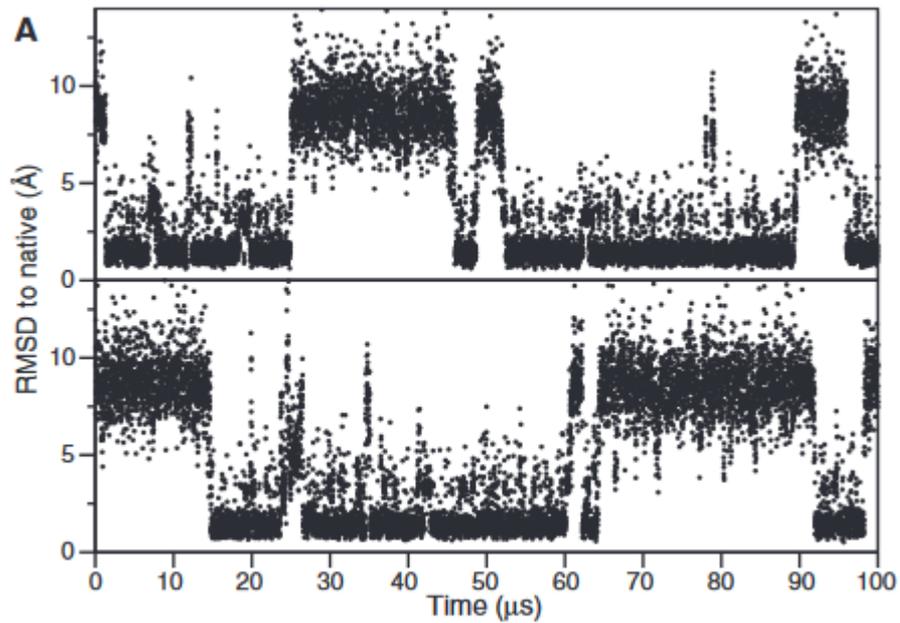
total simulation time

PDB entry of the experimental structure

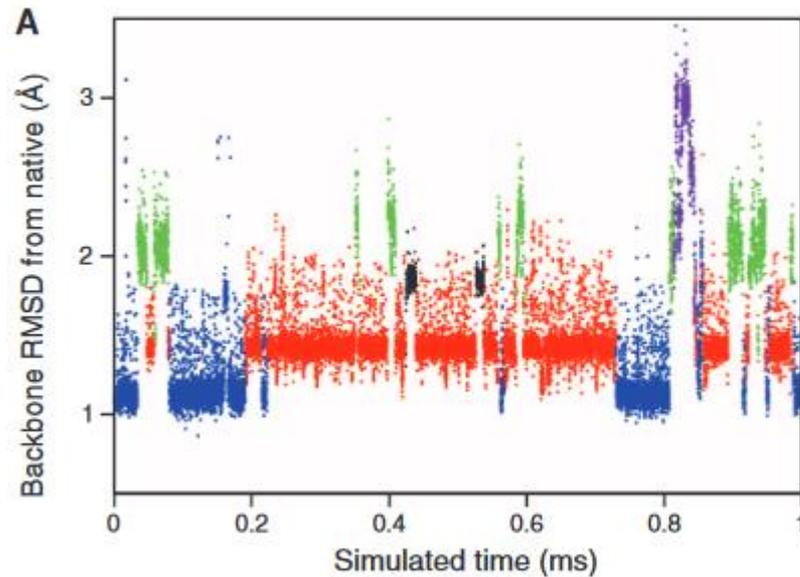
the Ca-RMSD between the two structures

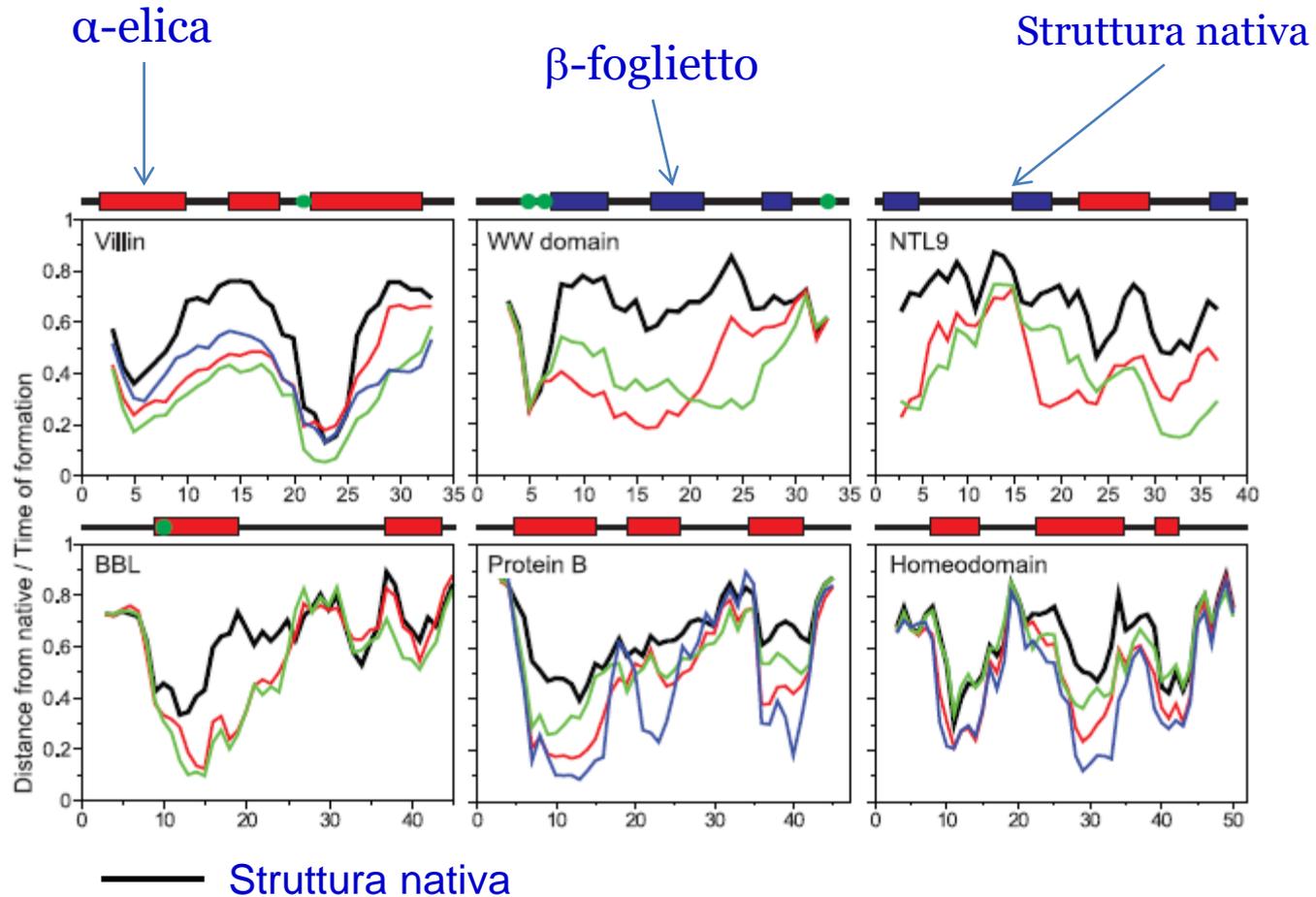
the folding time

$\alpha$ -helical,  $\beta$ -sheet and mixed  $\alpha/\beta$  classes



All-residue backbone RMSD  
from the crystal structure





In gran parte dei casi le differenti traiettorie RMSD ottenute per il folding di ciascuna proteina hanno il profilo simile a quello della struttura terziaria nativa

Per ognuna delle 12 proteine, sono state eseguite da **1 a 4** simulazioni, ciascuna lunga tra **100 ms e 1 ms**, ed osservato un totale di almeno **10 eventi di folding e 10 di unfolding**.

In totale, sono stati raccolti **~8 ms di simulazione**, contenenti più di 400 eventi di folding o unfolding.

Nel complesso, il confronto con i dati sperimentali disponibili indica che il campo di forza usato fornisce una descrizione ragionevole della struttura, della termodinamica e cinetica delle 12 proteine, dando una discreta fiducia all'accuratezza delle simulazioni dei meccanismi di ripiegamento osservati.

Nel processo di ripiegamento, la catena principale della proteina adotta **una topologia simile** a quella nativa, mentre si formano fin dall'inizio alcuni elementi della struttura secondaria.

Nella maggior parte dei casi, il ripiegamento segue un unico percorso in cui gli elementi della struttura nativa appaiono in un ordine altamente correlato alla loro propensione a strutturarsi nello stato nativo

## EXAMPLE: **Carbon monoxy myoglobin (MbCO)**

the system consisted of **1536** atoms:

1217	protein heavy atoms,
270	protein polar hydrogens,
43	heme heavy atoms,
4	heme meso hydrogens
1	CO molecule.

The **potential energy function** contained harmonic terms corresponding to:

- bond stretching,
- angle bending
- improper dihedral deformations,
- a cosine term for dihedral angles,
- Lennard-Jones potential
- electrostatic terms

**Verler algorithm** with an integration step size of **1 fs** was used

**Starting co-ordinates** for the room temperature trajectory are obtained from a MbCO crystal structure determined at 260 K at 1.5 Å resolution

**Disordered side-chains** are positioned in their most probable configuration and then minimized.

The system is then **heated** by 6 °K increments from 260°K to 300°K during 10 ps.

Next a 10 ps equilibration simulation is performed; the velocities of the atoms are randomly reassigned every 0-2 ps, corresponding to a Maxwell distribution at 300 K.

Given the equilibrated system, a 120ps molecular dynamics trajectory was generated. This trajectory, designated as trajectory **R**, was found to correspond to an average temperature of **315 °K** and is referred to as the "room temperature" simulation.

**Three low temperature trajectories**, designated as **L<sub>A</sub>**, **L<sub>A'</sub>**, and **L<sub>B</sub>**, at **80°K** were generated by restarting the dynamics from the **R** trajectory, at 325°K.

**Two X-ray crystal structures** were used in the comparison:

**X<sub>R</sub>** : a 260 °K MbCO structure

**X<sub>L</sub>** : a 80 °K MetMb structure

# Results

## Thermal expansion coefficient of myoglobin

Linear expansion coefficients were calculated from the differences in the myoglobin MD average structures at 80 K and 325 K.

For each pair of atoms ( $i, j$ ) the quantity  $\alpha_{ij}$ , describing the **fractional change in the interatomic distance  $r_{ij}$**  per °K upon a change in temperature from  $T_1$  to  $T_2$  was found from the equation:

$$\alpha_{ij}(T_1, T_2) = \frac{r_{ij}(T_2) - r_{ij}(T_1)}{(T_2 - T_1)\langle r_{ij} \rangle},$$

where  $r_{ij}(T_1)$  and  $r_{ij}(T_2)$  represent the distances between atoms  $i$  and  $j$  at temperatures  $T_1$  and  $T_2$  and  $\langle r_{ij} \rangle$  is the average distance.

The overall expansion coefficient  $\langle \alpha \rangle$  was taken to be the average of  $\alpha_{ij}(T_1, T_2)$  over all atoms  $j$ , with  $i$  equal to the center of mass of the molecule.

*Linear expansion coefficients for MbCO obtained from the molecular dynamics average structures from trajectories R, L<sub>A</sub> and L<sub>B</sub>*

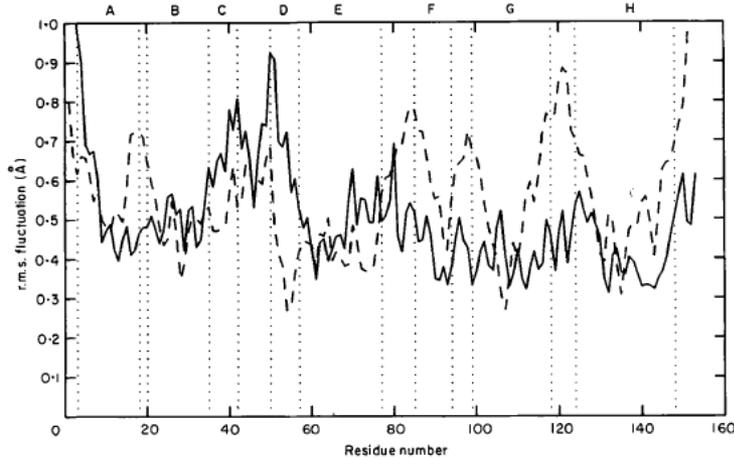
	80 K trajectory			
	$10^{-6} \text{ K}^{-1}$	L <sub>A</sub>	L <sub>B</sub>	Average
All atoms				
$\langle \alpha \rangle$		21	30	25
$\langle \alpha \rangle$ (325 K)		48	69	58

The average value of  $\langle\alpha\rangle$  for the two 80 K trajectories  $L_A$  and  $L_B$  is about  $25 \times 10^{-6} \text{ K}^{-1}$ , which is of the same order as, but smaller by a factor of 2, than the value found from X-ray data.

The average linear expansion coefficient for MbCO found here at 325 °K is  $58 \times 10^{-6} \text{ °K}^{-1}$ , as compared with the value  $115 \times 10^{-6} \text{ °K}^{-1}$  from X-ray data.

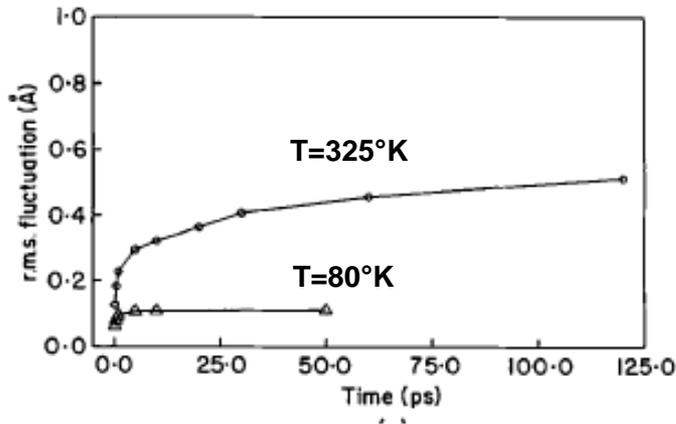
The calculated value is about half that obtained experimentally. It is possible that this difference arises from the crystal interactions in the experimental study, and/or from the neglect of the environment in the simulations.

# Comparison of X-ray and simulated structures



— R: MD at 325°K  
- - - X-ray at 260°K

## Time evolution of atomic fluctuations



At 80 K, the r.m.s, fluctuations are already equal to 50 % of their final values in 0.2 picosecond. The r.m.s, fluctuations then increase slowly with longer averaging intervals, reaching about 80% of their final values at one picosecond and about 90% at ten picoseconds.

The larger differences in fluctuations along the chain at 325°K arise from anharmonic motion (i.e. they are due to transitions among many substates).

At 80°K the protein is in fact sampling only a single minimum.