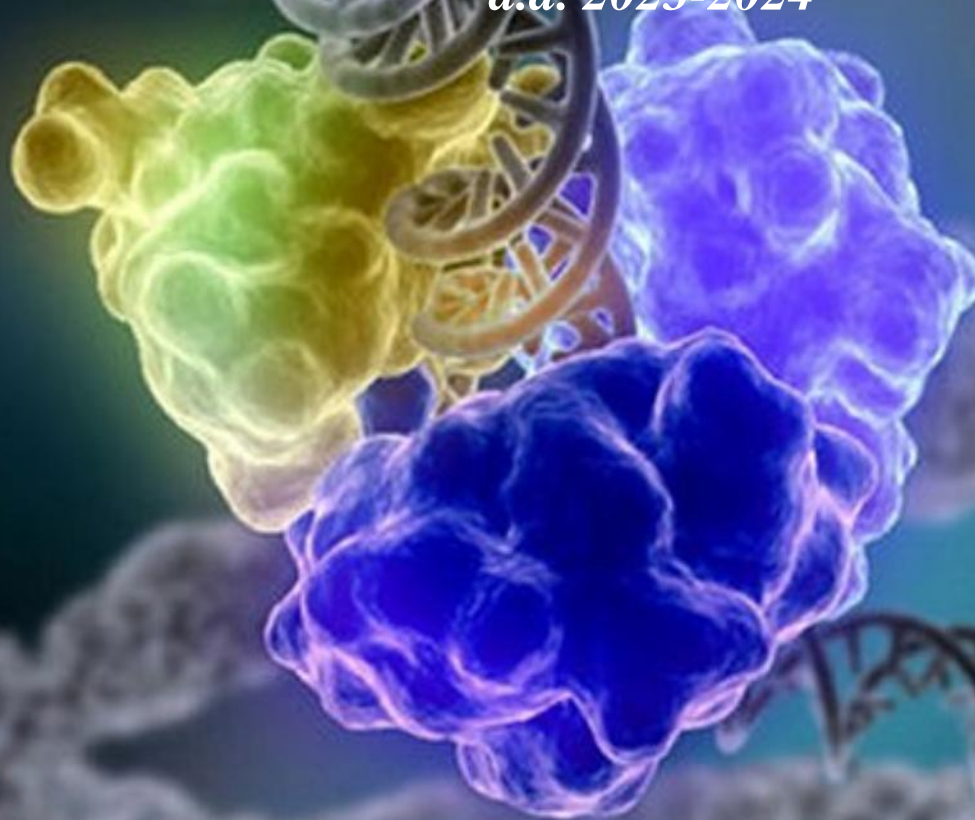


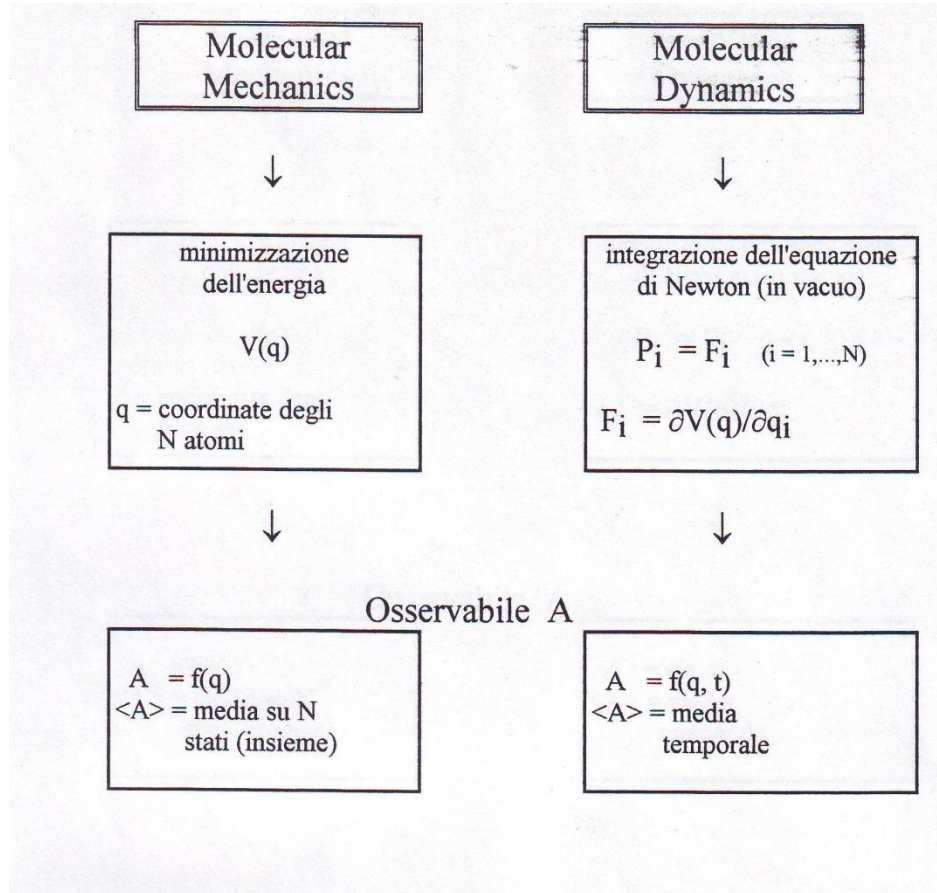
Dinamica Molecolare

Prof. R. Urbani
a.a. 2023-2024



MM

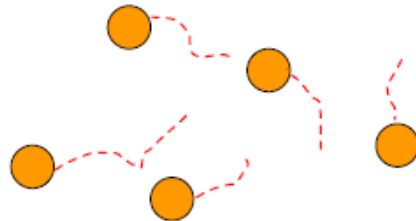
MD



$$V(q) = \sum_i^{\text{bonds}} k_i (r_i - r_{i,e})^2 + \sum_i^{\text{angles}} k_i (\theta_i - \theta_{i,e})^2 + \sum_i^{\text{some dihedrals}} v_i [1 \pm \cos(n_i \phi_i)]^2 + \sum_i^{\text{atoms}} \sum_{j>1}^{\text{atoms}} \left(-\frac{A_i A_j}{r_{ij}^6} + \frac{B_i B_j}{r_{ij}^{12}} + \frac{q_i q_j}{r_{ij}} \right)$$

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.

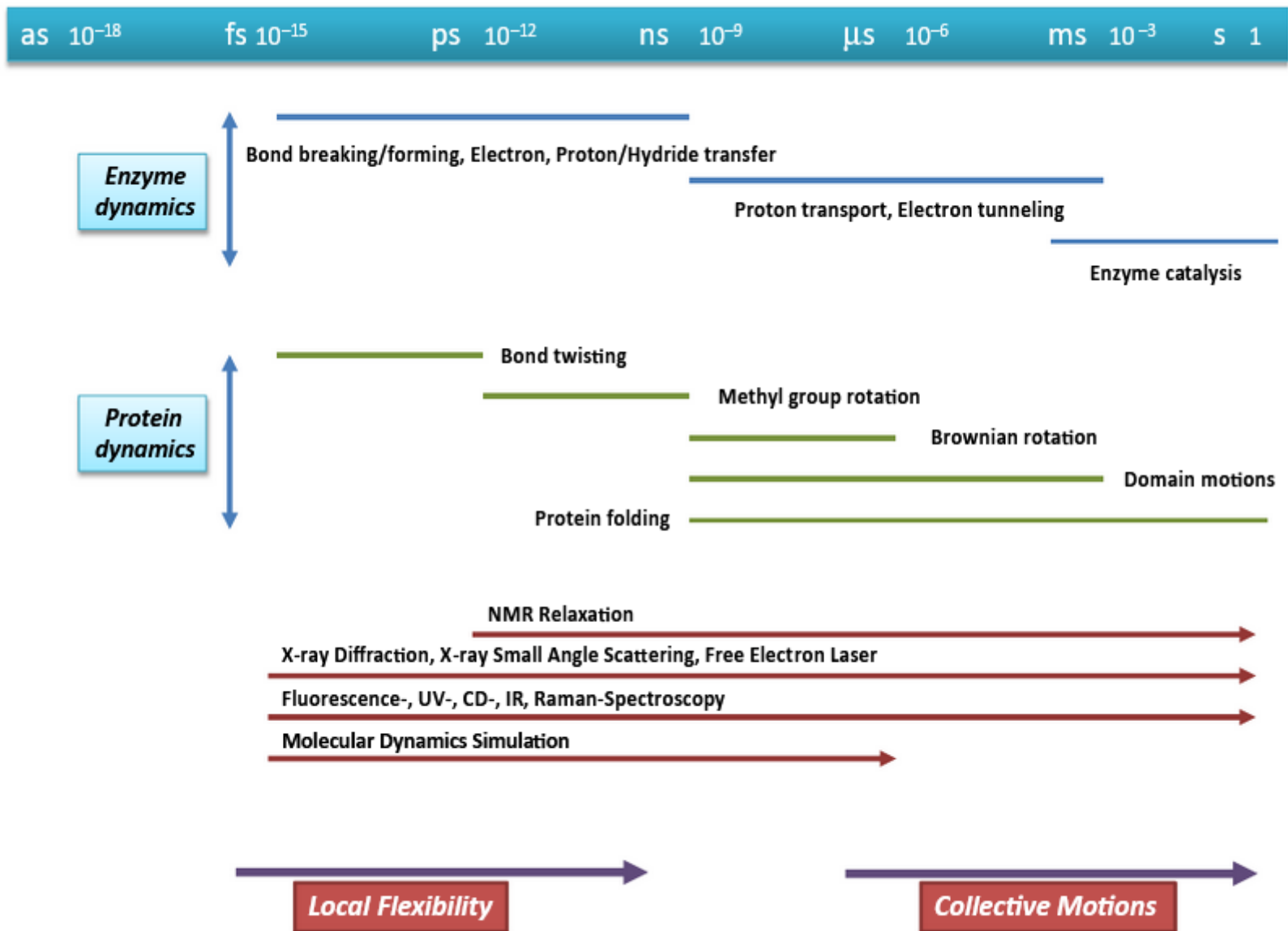


Fig. 2. Time scales of protein motions and relevant experimental/theoretical techniques for their investigation.

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 \mathbf{r}_i si ricavano da strutture sperimentali (cristallografia raggi X, NMR ecc.) o ottenute con modeling;

2. le velocità iniziali \mathbf{v}_i si ottengono dalla distribuzione di Boltzmann delle velocità ad una data temperatura;

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

$$\mathbf{F}_i = m_i \mathbf{a}_i = - dV/dr_i$$

La traiettoria è generata da integrazioni simultanee dell' equazione del moto di Newton

$$\mathbf{F}_i = m_i \cdot \mathbf{a}_i$$

per tutti gli atomi del sistema molecolare

Tenendo presente che la forza si può esprimere come **gradiente dell'energia potenziale**:

$$\mathbf{F}_i = - dV/dr_i$$

si combinano le due equazioni e si ottiene:

$$- dV/dr_i = m_i d^2r_i/dt^2$$

che collega la derivata dell'energia potenziale alle variazioni di posizione in funzione del tempo. E' questa che viene **integrata**.

MD / Schema

- Configurazione iniziale

Es. Coordinate di database (PDB)

- Velocità iniziali e controllo della temperatura

- la simulazione dinamica coinvolge tecniche di valutazione e di controllo della T che si realizzano in genere utilizzando dei **termostati (virtuali)** opportuni.
- La temperatura è una grandezza che traduce a livello macroscopico la situazione a livello atomico mediante il termine dell'energia cinetica
- Tale relazione passa attraverso il valore della velocità degli atomi (**distribuzione di Maxwell-Boltzmann**)

$$p(\mathbf{v})d\mathbf{v} = \left(\frac{m}{2\pi kT} \right)^{3/2} e^{-\frac{m\mathbf{v}^2}{2kT}} 4\pi v^2 d\mathbf{v}$$

La regolazione della T

- Le velocità assegnate come valori iniziali del processo di integrazione vengono determinate in modo tale che **sia soddisfatta la distribuzione di Maxwell-Boltzmann una volta scelto il valore T**
- Durante l'esperimento virtuale tale **distribuzione muta nel tempo** mano a mano che il calcolo procede.
- Dovendo T essere **costante**, il sistema preposto alla termostatazione deve configurarsi come una tecnica che permetta di generare la corretta configurazione statistica.

La regolazione della T

- **Metodi di Termostatazione:**
- Si basano sul concetto di “bagno di termostatazione virtuale”(Berendsen , Langevin)



- Valutano la **differenza tra T impostata e T calcolata** dall'energia cinetica. Questa differenza riscalata in funzione di un parametro di rilassamento permette di calcolare un fattore correttivo delle velocità (intesa come distribuzione di probabilità)

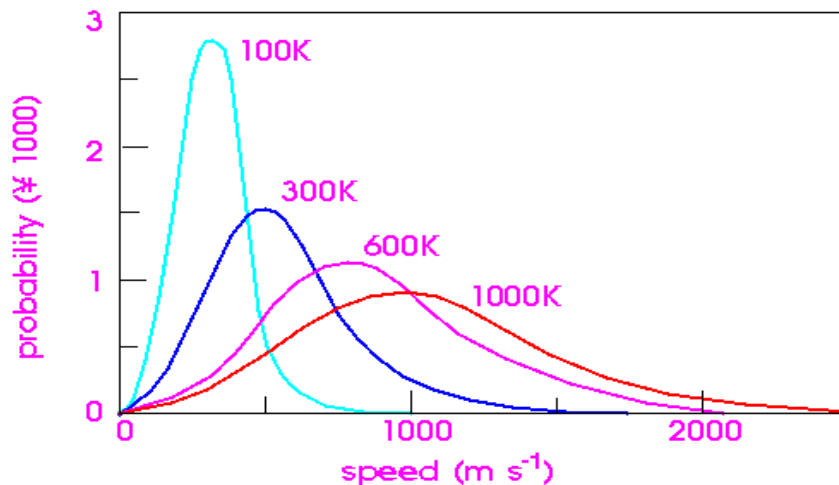
Quindi la distribuzione delle velocità iniziali sono derivate da una distribuzione Maxwelliana random corrispondente alla T desiderata:

$$p(\vec{v}_i) = \sqrt{\frac{m_i}{2\pi K_B T}} \exp[-m_i \vec{v}_i^2 / 2K_B T]$$

$p(\vec{v}_i)$ = probabilità che l'atomo i -esimo abbia velocità \vec{v}_i alla temperatura T .

L'iniziale distribuzione random delle velocità viene scelta in modo tale che non ci sia momento totale della quantità di moto, cioè:

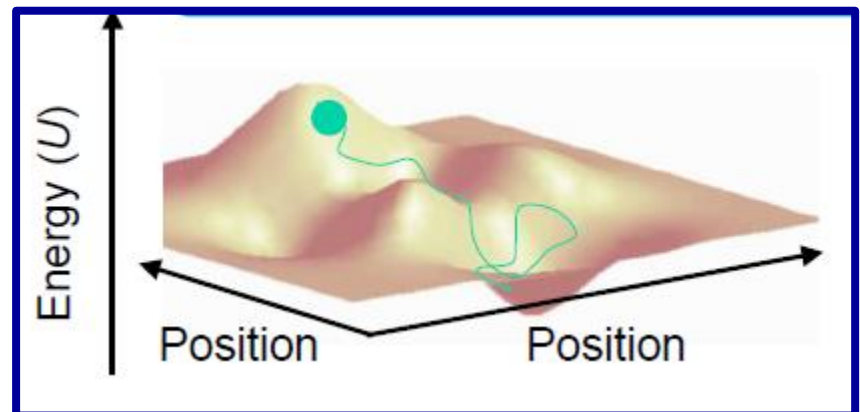
$$\vec{P} = \sum_{i=1}^N m_i \vec{v}_i = 0$$



$$\frac{3}{2} N k_b T(t) = \sum_i \frac{1}{2} m_i v_i^2(t)$$

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.



La **simulazione MD** è in molti aspetti simile ad un vero e proprio **esperimento di laboratorio**, perciò gli *step* da seguire per ottenere una simulazione realistica sono correlati alla sua controparte reale:

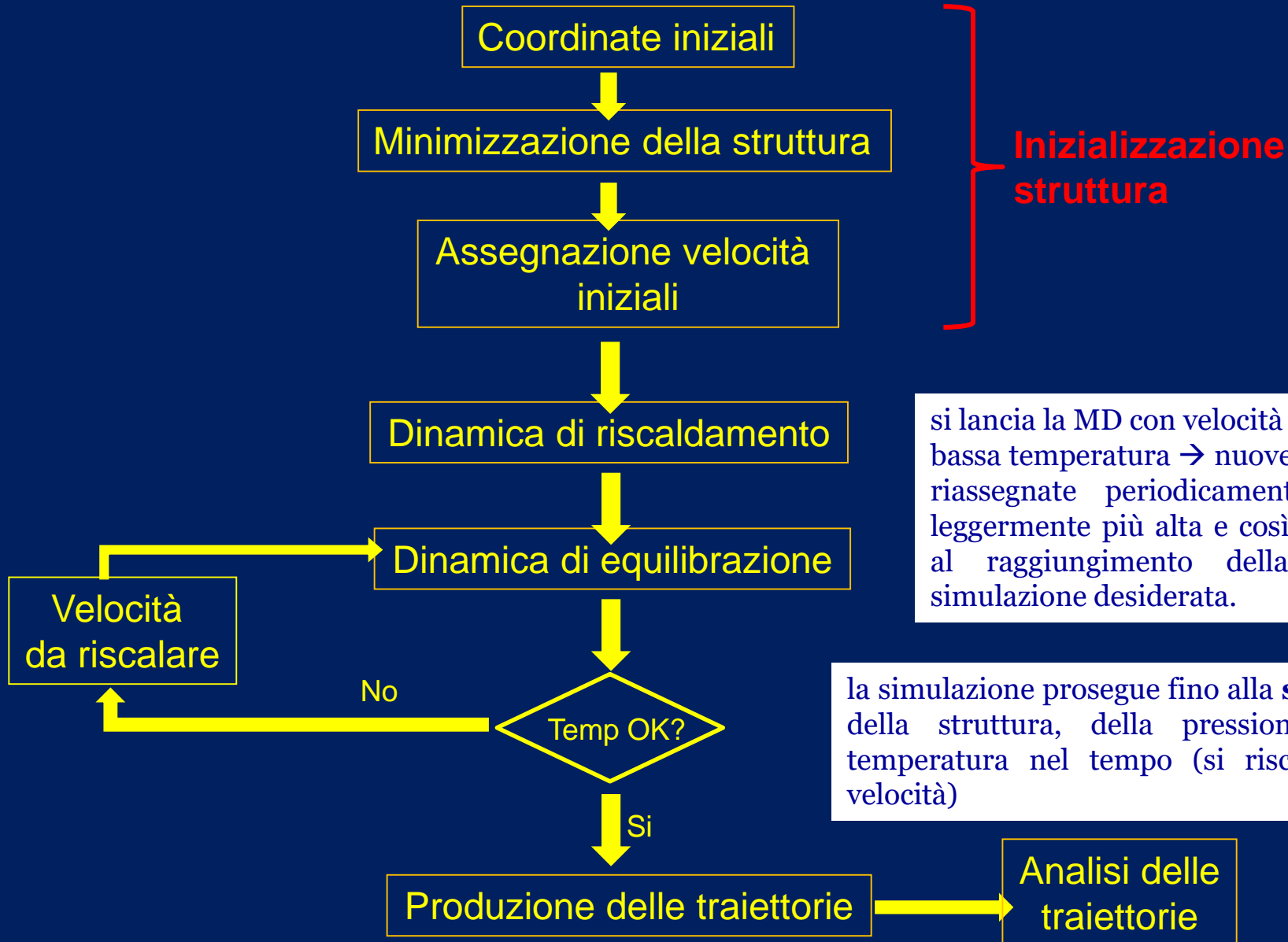
1. Per prima cosa occorre preparare il campione da simulare, costituito da N particelle, risolvendo le equazioni di Newton affinché le proprietà temporali siano quelle di equilibrio (**processo di equilibratura**);

2. **Simulazione dell'evoluzione del sistema** monitorando le grandezze termodinamiche importanti: **Energia, Temperatura, Volume, Pressione e Numero di particelle devono essere costanti**; questa è la vera parte di simulazione di dinamica molecolare in cui avviene la produzione dei dati da analizzare;

3. A simulazione conclusa, dopo aver definito delle opportune osservabili, **si misurano le grandezze fisiche di interesse al variare del tempo**.

Questa quindi è la parte di analisi dati che ci permetterà di trarre delle conclusioni alla fine del nostro esperimento computazionale.

Set-up e running di una Simulazione MD



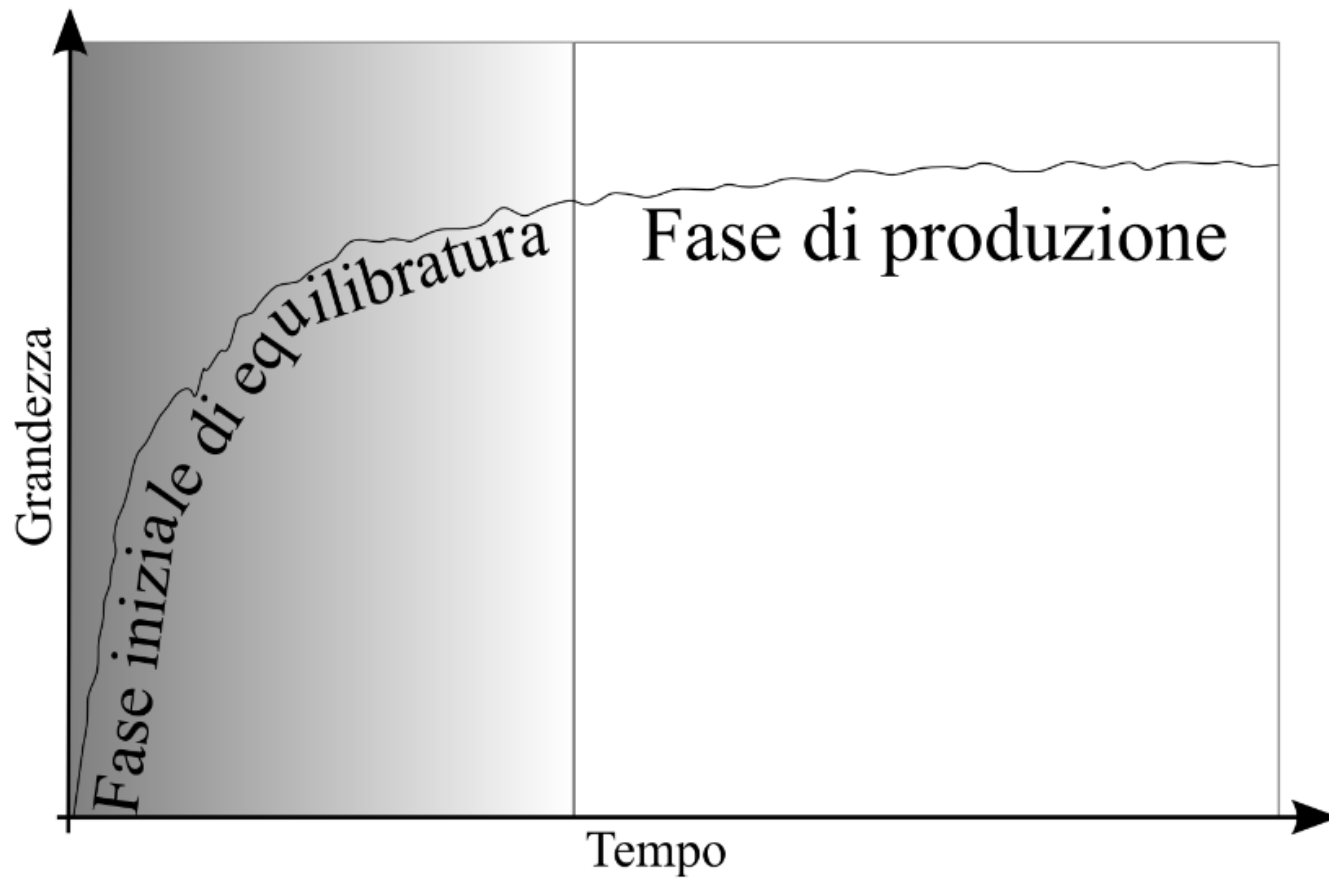


Figura 2 Fase di equilibratura e fase di produzione per una generica grandezza durante una simulazione di dinamica molecolare.

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*) δt , 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 Δt

- ✓ 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 ~ 10 fs \rightarrow time step ~ 1 fs

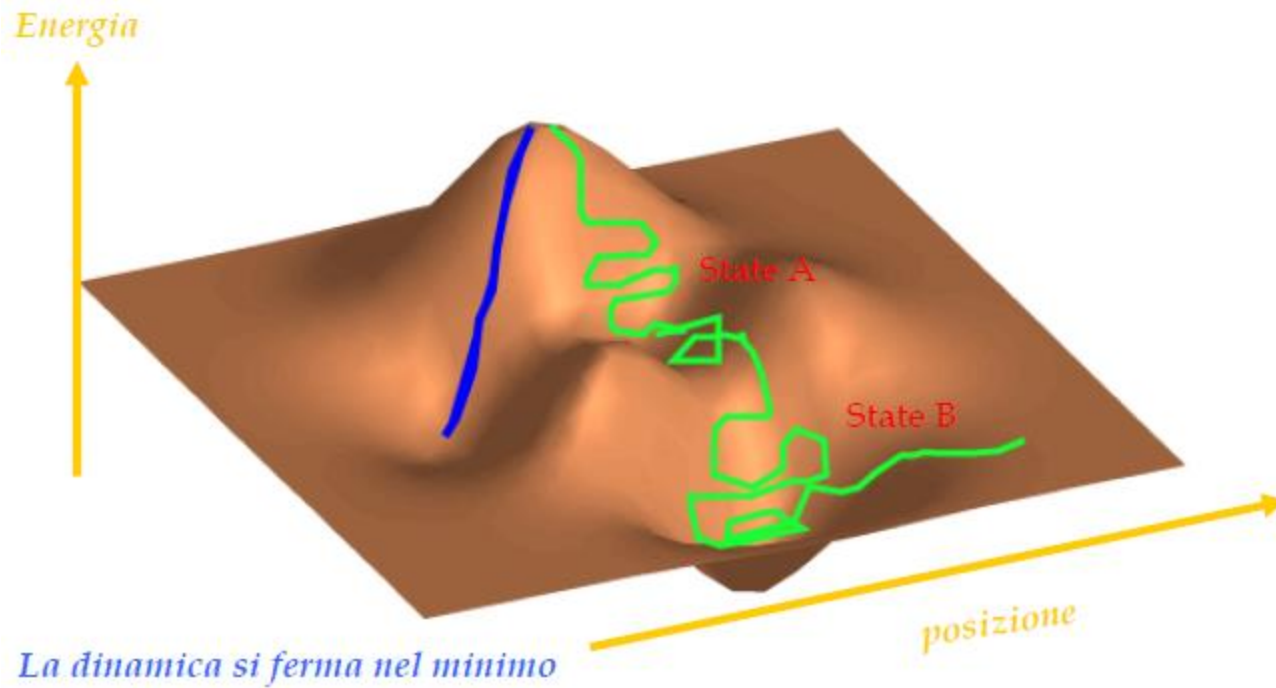
Insiemi statistici

- Si introducono gli INSIEMI STATISTICI:

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

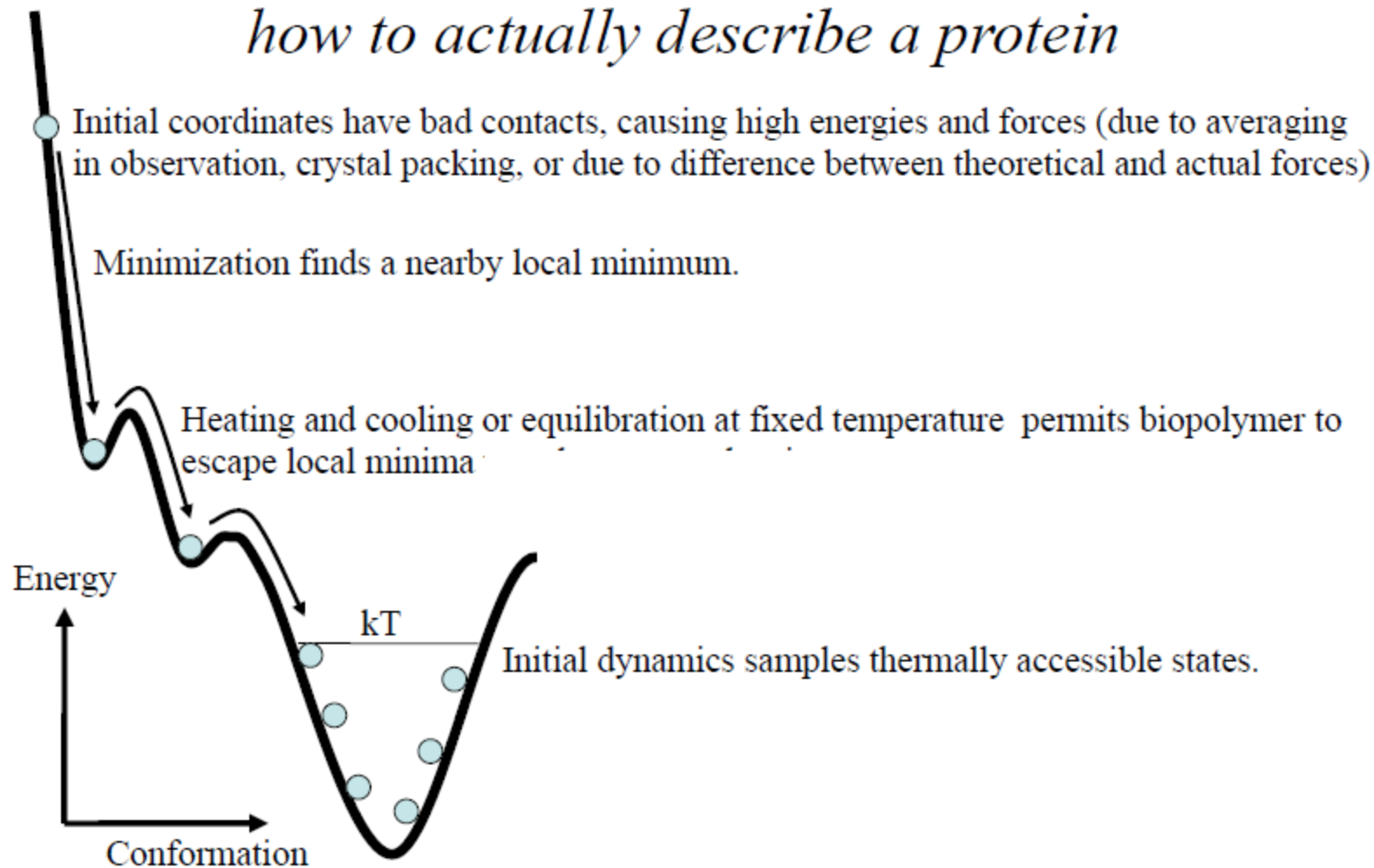
Sistema microcanonico	N, E, V costanti	per mantenere E costante la T fluttua per continua conversione di en.cinetica in potenziale.
Sistema canonico	N, T, V costanti	per mantenere la T costante, le velocità fluttuano secondo una statistica gaussiana
Sistema isotermo-isobarico	N, T, P costanti	per mantenere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

MD / ricerca di minimi



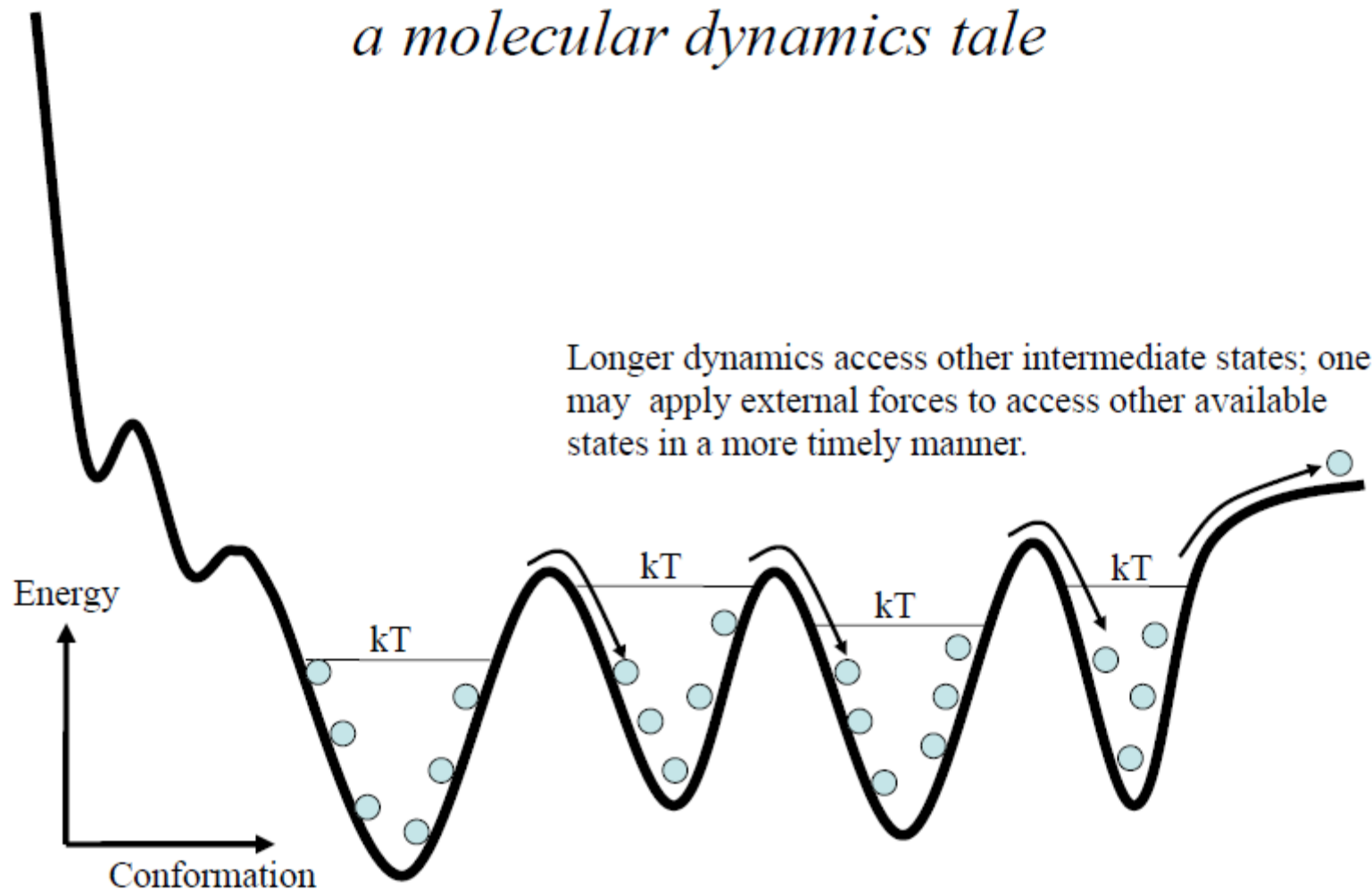
From the Mountains to the Valleys

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** nei biopolimeri 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 creato un biopolimero, la maggior parte dei suoi legami covalenti **non si rompono o si formano altri**, durante l'espressione delle funzioni biologiche.
- Nelle proteine 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.

I moti interni delle macromolecole di importanza biologica hanno suscitato un notevole interesse negli ultimi anni (dal 1970).

Sono stati effettuati **studi sperimentali e teorici** sui moti e sono stati esaminati i loro ruoli biologici.

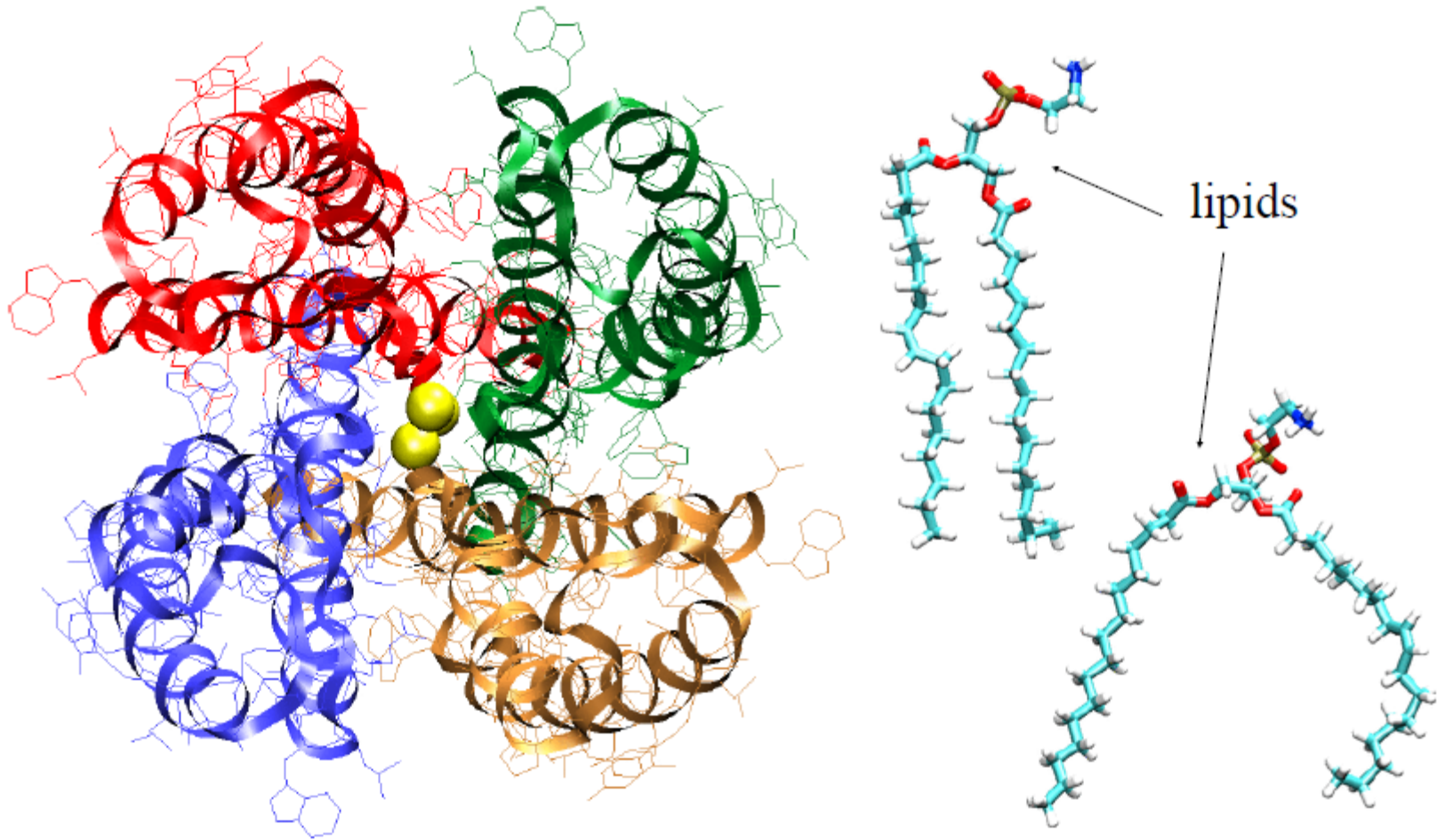
Le simulazioni MD forniscono un approccio per ottenere informazioni in dettaglio su questi moti **a livello atomico**.

Con tali simulazioni è possibile esaminare sia le **proprietà di equilibrio** (ad esempio la struttura media, le fluttuazioni e la termodinamica) sia i **fenomeni dipendenti dal tempo**, come le rotazioni delle catene laterali e le correlazioni di dipolo elettronico o nucleare.

Le simulazioni delle proteine sono di particolare interesse perché sono sistemi complessi e disomogenei con proprietà intermedie tra quelle dei liquidi e dei solidi; presentano **un denso impacchettamento di atomi, legami covalenti lungo la catena polipeptidica e deboli interazioni van der Waals ed elettrostatiche tra gli atomi non legati**.

Ciò suggerisce (come abbiamo visto) che il loro comportamento dinamico coinvolge probabilmente una molteplicità di scale temporali.

Setting up the system

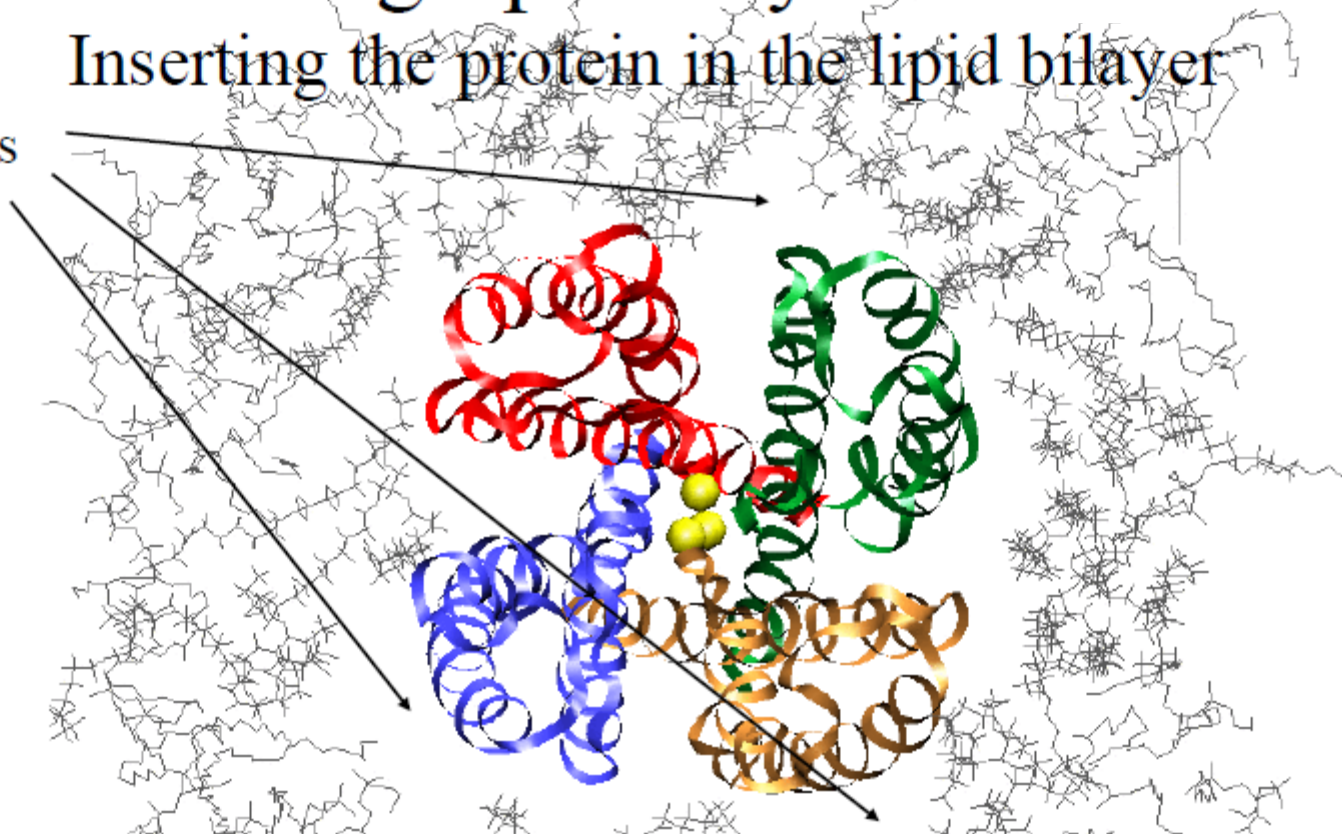


Simulate the protein in its natural environment: solvated lipid bilayer

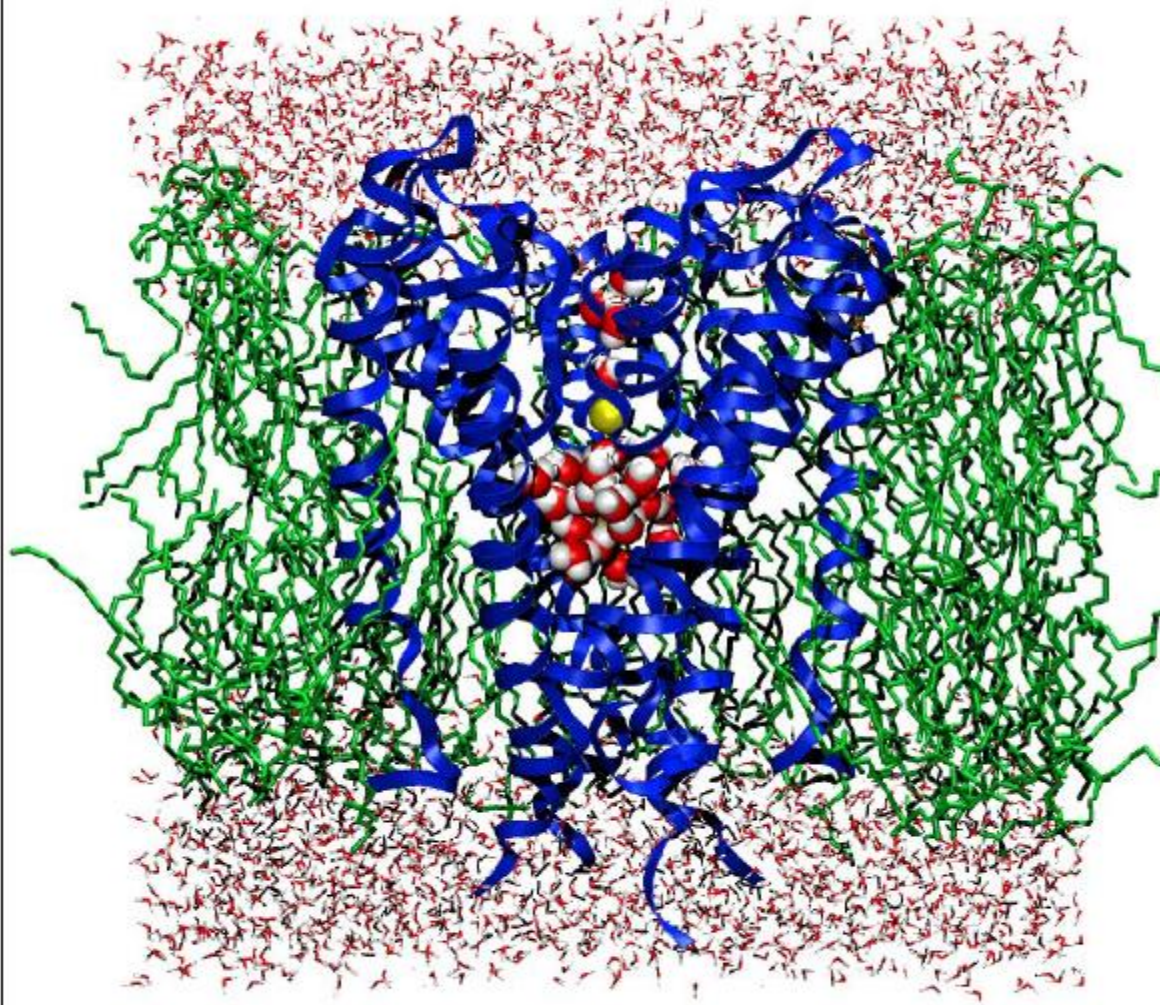
Setting up the system

Inserting the protein in the lipid bilayer

gaps



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).

Macromolecular properties of cepacian in water and in dimethylsulfoxide

Yury Herasimenka,^{a,†} Paola Cescutti,^a Carlos E. Sampaio Noguera,^b José R. Ruggiero,^b Ranieri Urbani,^a Giuseppe Impallomeni,^c Flavio Zanetti,^d Stéphane Campidelli,^{e,‡} Maurizio Prato^e and Roberto Rizzo^{a,*}

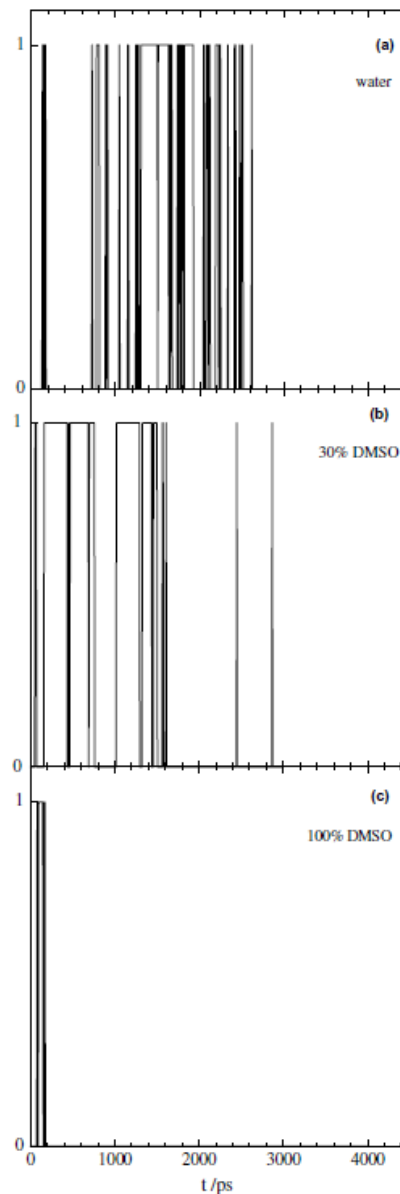
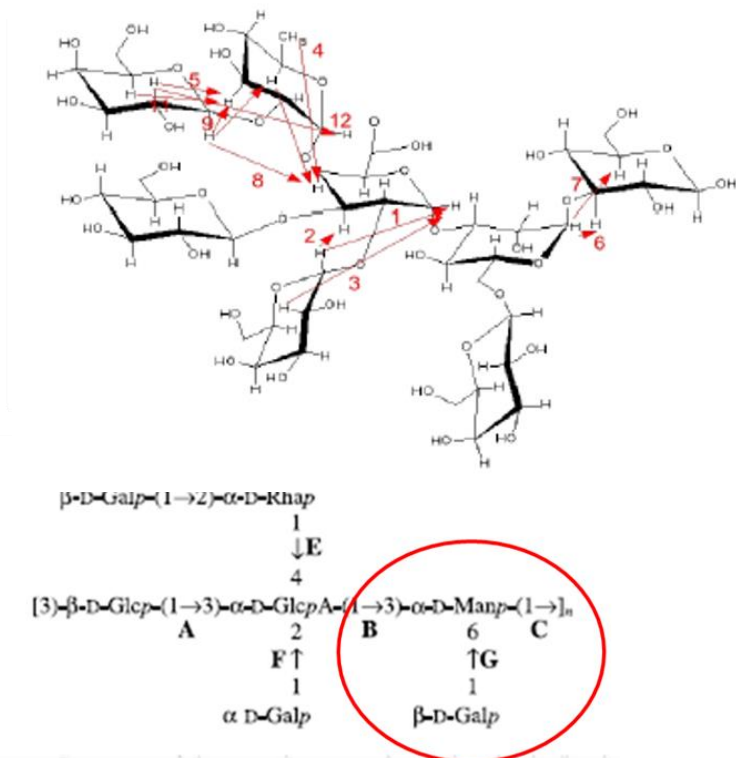
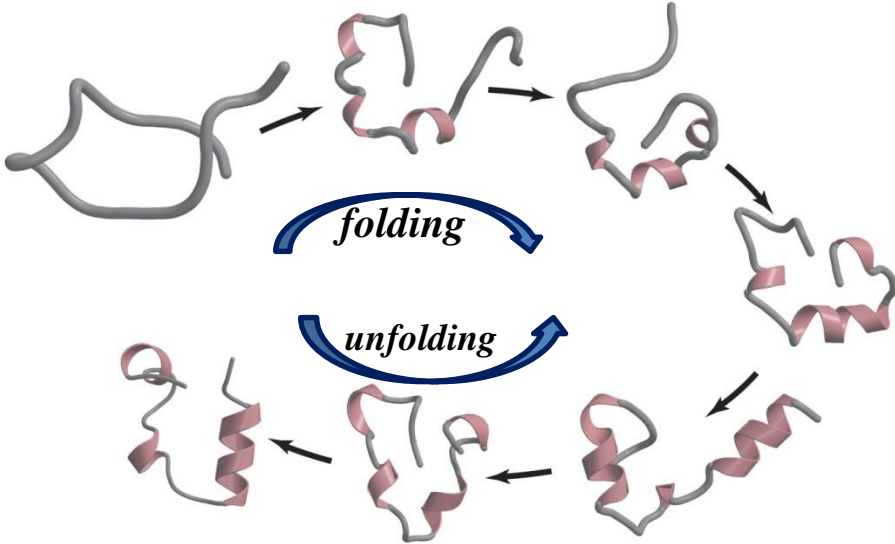
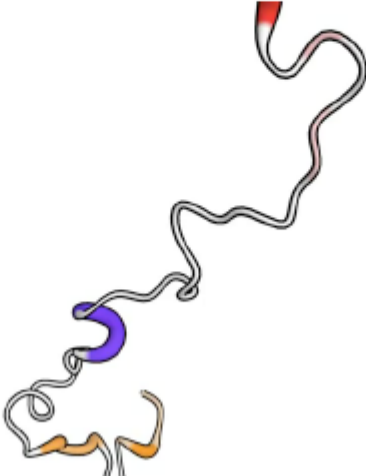


Figure 10. Occurrence of hydrogen bonding during the 3 ns simulation time for the 3-OH ($\alpha\text{-D-Rhap}$)-O-6 ($\beta\text{-D-Glcp}$) interaction in (a) water, (b) 30% Me_2SO -water, and (c) 100% Me_2SO solutions.

Understanding the *process* of protein folding



Solvatazione

E' importante poiché molte molecole sono in soluzione

Le molecole di solvente polare si dispongono attorno allo ione o alla molecola con il lato di polarità opposta rispetto alla carica (parziale) ed esercitano un campo elettrico

Questo campo elettrico da luogo ad un **potenziale di interazione** che deve essere aggiunto all'Hamiltoniana dell'energia

Nei metodi di Meccanica e Dinamica Molecolare si aggiunge semplicemente un contributo elettrostatico nell'espressione empirica dell'energia e si considera poi una costante dielettrica nell'interazione elettrostatica tra i vari atomi



- Effetto solvente** {
1. Trattamento **IMPLICITO** (modello CONTINUO)
 2. Trattamento **ESPLICITO** (modello ALL-ATOMS)

Termodinamica di solvatazione

Energia libera di solvatazione: *variazione di energia libera nel trasferimento di una molecola dal vuoto al solvente.*

$$\Delta G_{solv} = \Delta G_{elett} + \Delta G_{VdW} + \Delta G_{cav} + \Delta G_{HB}$$

Diagram illustrating the components of the free energy of solvation (ΔG_{solv}):

- elettrostatico** (electrostatic) points to ΔG_{elett}
- Van der Waals** points to ΔG_{VdW}
- cavitazione** (cavitation) points to ΔG_{cav}
- Legame idrogeno** (hydrogen bond) points to ΔG_{HB}

Modello del solvente continuo

Il solvente agisce come una perturbazione del nostro sistema in fase gassosa e questo concetto è alla base dei modelli del solvente “continuo”.

L'energia libera di solvatazione (ΔG_{sol}) include tre componenti:

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{cav} + \Delta G_{HB}$$

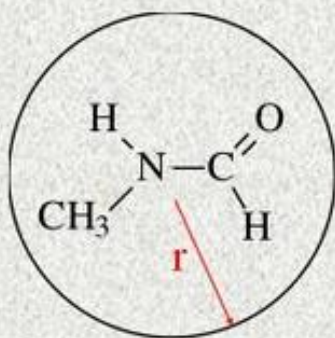
dove ΔG_{elec} è la componente elettrostatica (la polarizzazione del solvente con costante dielettrica ϵ).

ΔG_{vdw} rappresenta le interazioni di van der Waals tra il soluto ed il solvente; questa può essere ulteriormente suddivisa in un termine repulsivo, ΔG_{rep} , ed in un termine di dispersione attrattiva, ΔG_{disp} .

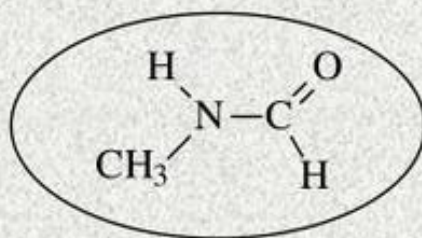
ΔG_{cav} è l'energia libera richiesta per formare la **cavità** del soluto all'interno del solvente. Questa componente è positiva e comprende la penalizzazione entropica associata alla riorganizzazione delle molecole di solvente attorno al soluto ed al lavoro fatto contro la pressione del solvente nel creare la cavità.

1a. Solute Cavity Size and Shape

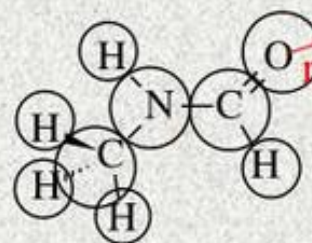
Spherical
(Born)



Ellipsoidal
(Onsager)



van der Waals
(Kirkwood)



SOLVATAZIONE

Trattamento **IMPLICITO**

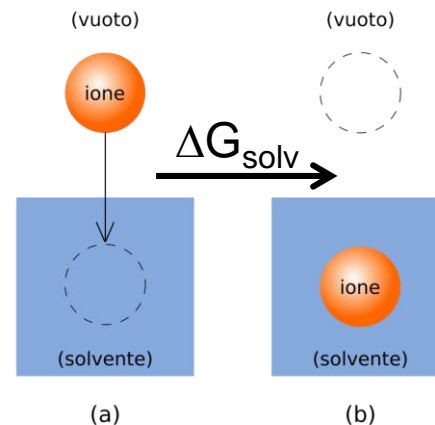
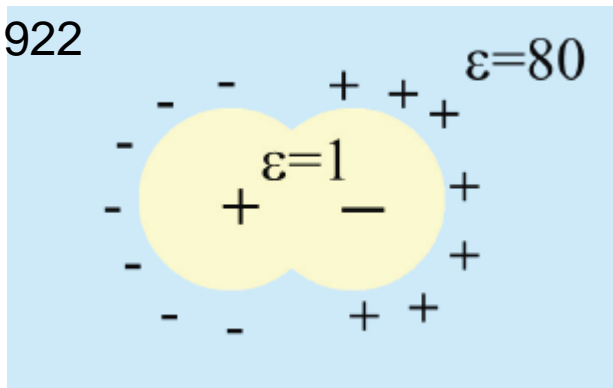
Il modo implicito più semplice è quello di includere una costante elettrostatica nel termine elettrostatico dell'energia potenziale, usando una costante dielettrica effettiva dipendente dalla distanza:

$$\epsilon_{\text{eff}} = \epsilon(\mathbf{r}_{ij}).$$

Le molecole di solvente sono implicitamente considerate modificando opportunamente le interazioni tra i nuclei come se essi subissero l'effetto "medio" cioè statistico della presenza di un gran numero di molecole di solvente. In questo caso il costo computazionale è notevolmente ridotto rispetto alla trattazione con solvente esplicito.

Modello di Born

1922



$$\Delta G_{\text{sol}} = \Delta G_{\text{elec}} + \Delta G_{\text{vdw}} + \Delta G_{\text{cav}}$$

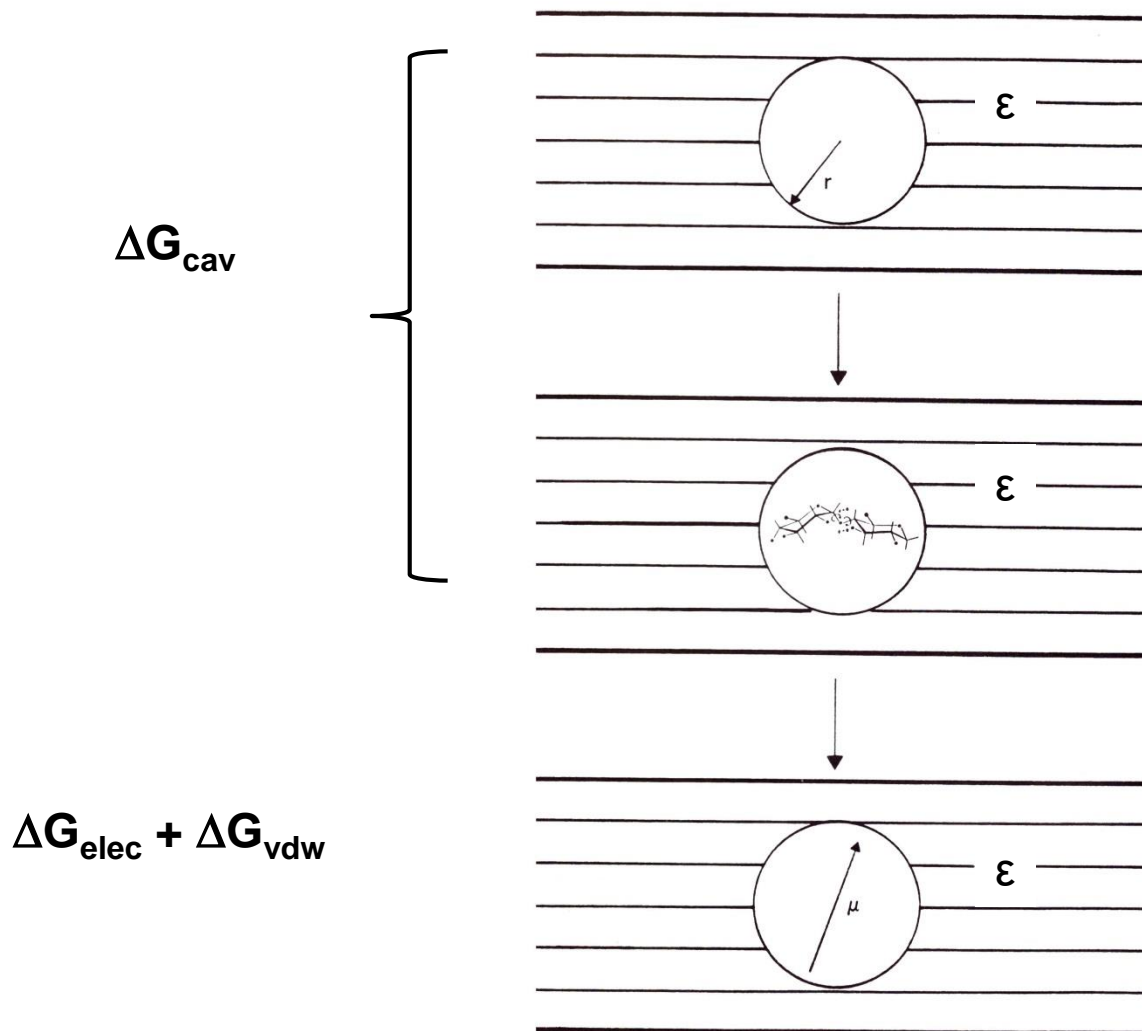


Figure 1 Scheme of the theoretical approach for the calculation of the free energy of solvation. First, a cavity of proper size is created in the solvent. Secondly, the real molecule, with a given conformational energy, is inserted in the cavity. Thirdly, the surrounding molecular layer of the solvent interacts with the solute molecules with given charge displacement.

Trattamento ESPLICITO (All-atoms)

Consiste nell'includere molecole di solvente nel sistema. Le molecole forniscono direttamente l'effetto di schermo (o almeno la parte legata al riorientamento delle stesse).

Condizioni al contorno

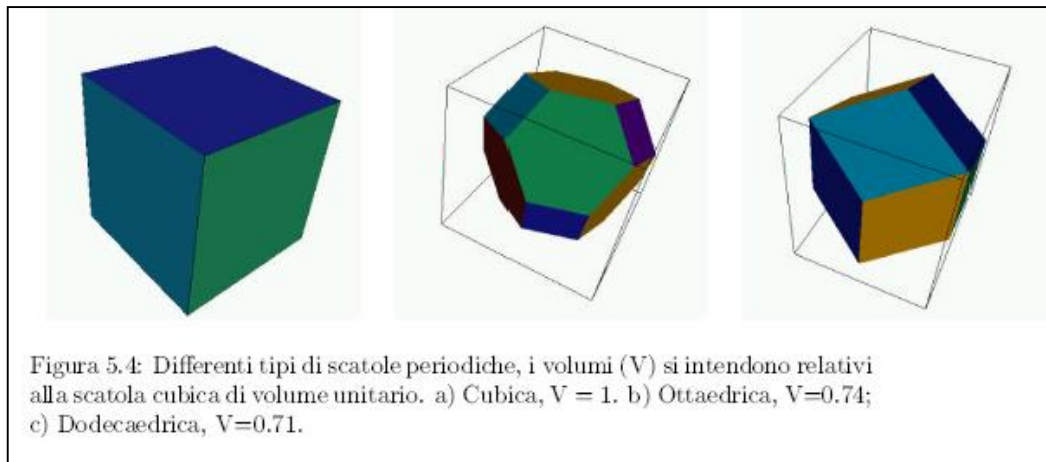


Figura 5.4: Differenti tipi di scatole periodiche, i volumi (V) si intendono relativi alla scatola cubica di volume unitario. a) Cubica, $V = 1$. b) Ottaedrica, $V=0.74$; c) Dodecaedrica, $V=0.71$.

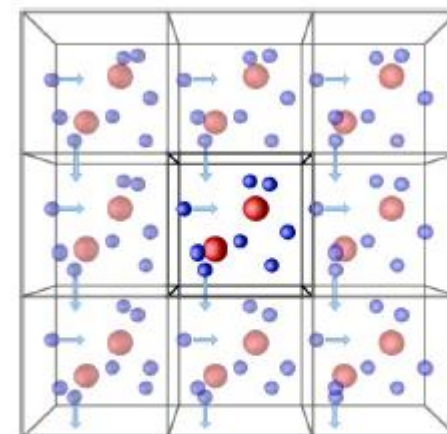
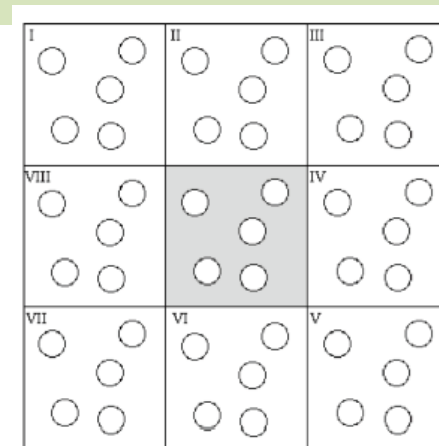


Figura 2.1. Rappresentazione delle condizioni periodiche al contorno: l'unico sistema realmente simulato è quello centrale ma grazie a questa condizione è come se fosse circondato da sistemi identici. Fonte: isaacs.sourceforge.net

Per realizzare **la scatola si deve conoscere la densità del liquido ad una certa temperatura.**

Questa ci permette di calcolare il volume medio occupato da una singola molecola.

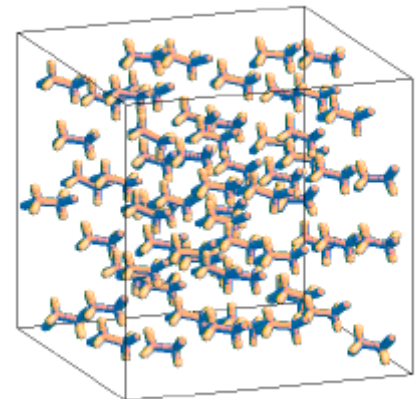
Se **d** è la densità in kg m^{-3} , allora il **volume in nm** occupato da una molecola e' uguale a:

$$V_m = \frac{1000M}{dN_a} 10^{21}$$

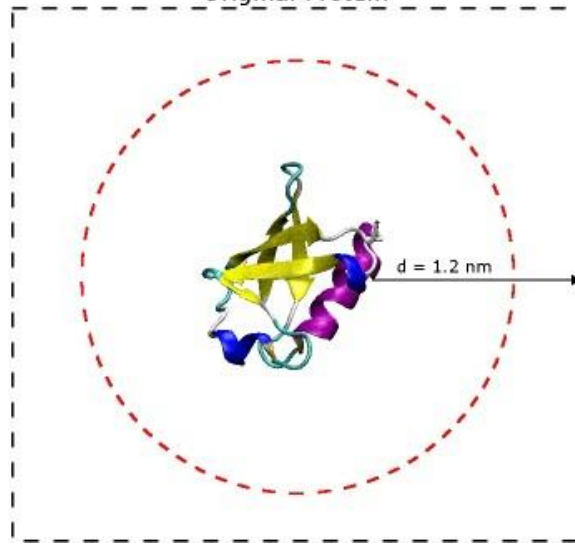
Dal volume (V_m) si ottiene il lato (l) del volumetto che contiene la molecola.

Per esempio, nel caso del 1,2-dicloroetano abbiamo che $M = 98.96\text{g}$ e $d = 1246.3 \text{ kg m}^{-3}$, da cui si ottiene $l = \sqrt[3]{0.132} = 0.509 \text{ nm}$.

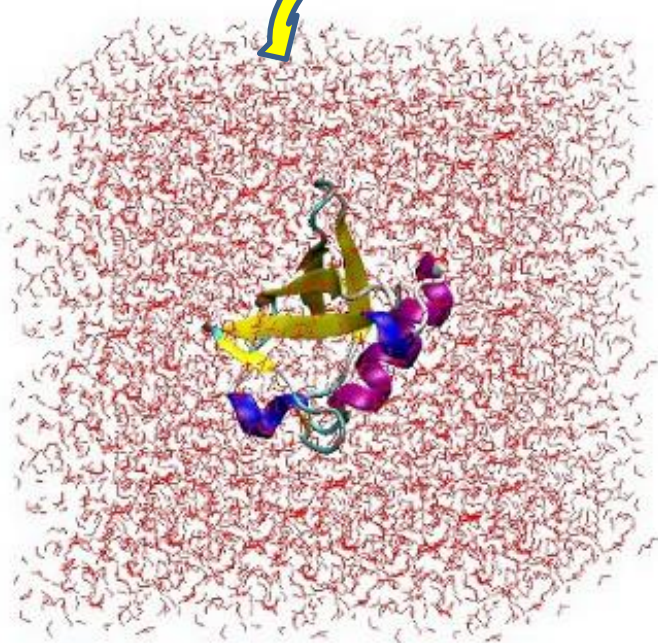
Il metodo replica il volumetto di dimensione (0.509 x 0.509 x 0.509) avente la densità del liquido (ad una data T) e la molecola in essa contenuta su un reticolo di dimensione voluta. Le molecole vengono distribuita ed orientate in modo *random*.



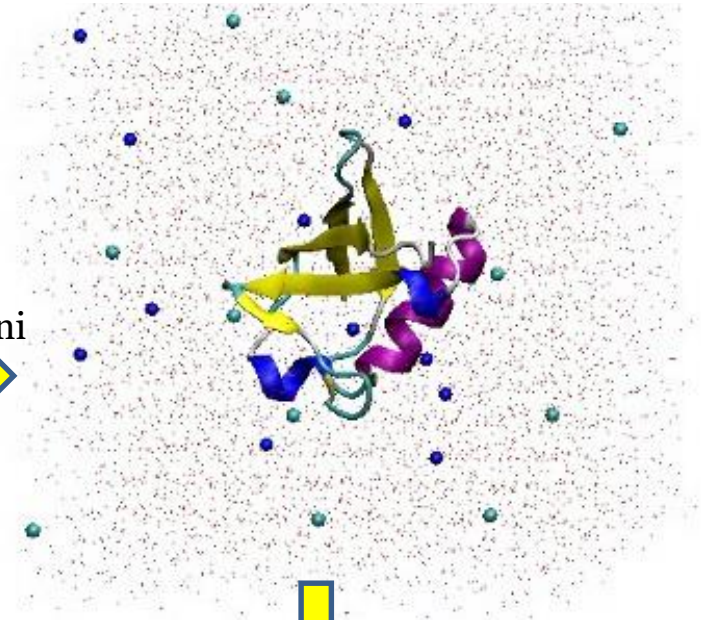
Original Protein



Addizione di solvente



Addizione di ioni



Ottimizzazione della geometria del sistema (soluto + solvente + controioni)

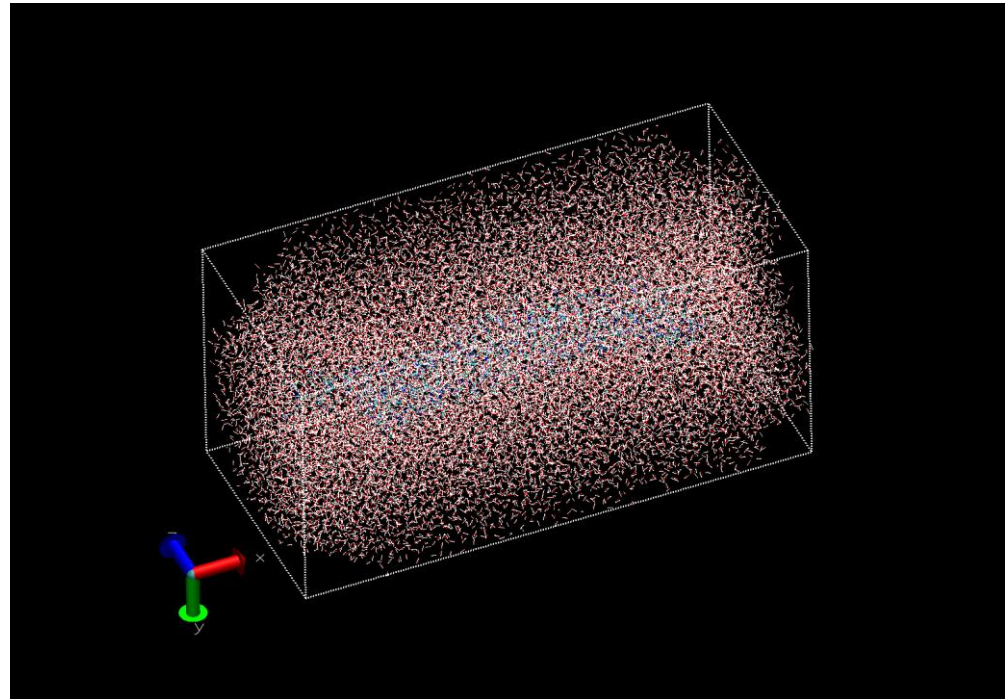
Esempio:
scatola periodica di molecole d'acqua

$P = 1 \text{ atm}$

$T = 300 \text{ K}$

Dimensione della scatola periodica: 60 \AA

N di molecole = **22487**



l l'amiloide β 42 ($A\beta$ 42). Questo peptide, lungo 42 aminoacidi, originato dall'*Amyloid Precursor Protein* (APP), è il maggior costituente delle placche amiloidee ritrovate dall'analisi *post mortem* del cervello di pazienti affetti da morbo di Alzheimer.

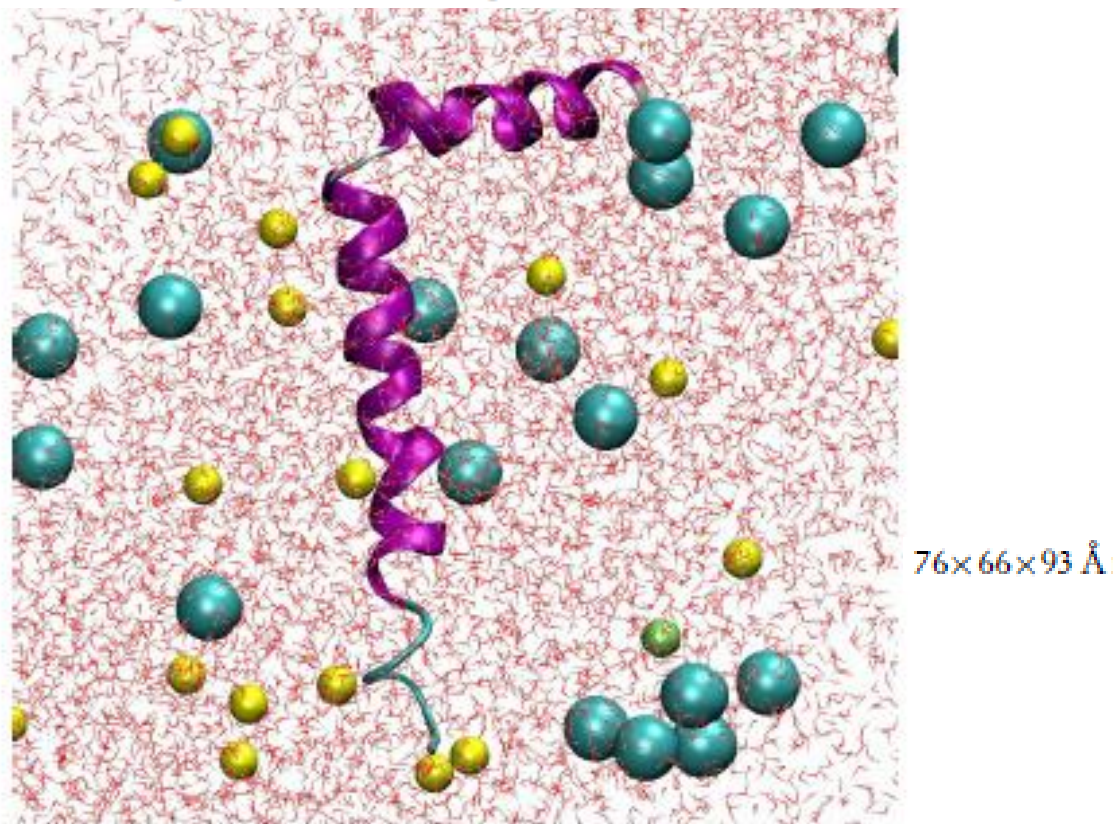


Figura 2.3: Set-up molecolare di partenza dell' $A\beta$ 42 in soluzione acquosa che mima l'ambiente extra-cellulare. Impostando un valore di Forza Ionica di 140mM in NaCl, il box d'acqua era sufficientemente grande da contemplare anche la presenza di uno ione Ca^{2+} (sfera di Van der Waals color verdino). In giallo sono rappresentate come sfere di Van der Waals i cationi Na^+ ; in celeste gli anioni Cl^- . In viola sono rappresentate le due α -eliche del motivo *Helix-Kink-Helix* della struttura 1IYT così come ottenuta da NMR. Quest'immagine è stata ottenuta con VMD [31]. La struttura secondaria viene assegnata con il plugin STRIDE [32].