



Corso di
Proprietà di Biopolimeri

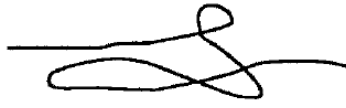
Prof. R.URBANI

Dipartimento di Scienze Chimiche e Farmaceutiche
a.a. 2023-2024

PREDIZIONE DELLA CONFORMAZIONE:
Meccanica Molecolare (MM)
Statistica conformazionale



Compact



Extended



Random



Rod



Sphere



**Quali sono i fattori stabilizzanti di queste strutture?
Come posso determinare/predire la topologia, le proprietà, delle catene?**

Metodologie sperimentali

Quali sono le tecniche
che mi danno informazioni sulla
dimensione e forma
delle macromolecole?

Approccio teorico

Quali sono i modelli descrittivi
che permettono di predire le
dimensioni e forma delle
macromolecole?

Metodi 'classici'

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

Le conformazioni di una catena polimerica sono stabilizzate da legami non covalenti (**interazioni deboli**) intramolecolari

Interazioni che stabilizzano le conformazioni:

forze di van de Waals :

1. forza dipolo permanente-dipolo permanente o **forza di Keesom** (energia dell'ordine di 4 kJ/mole);
2. forza dipolo permanente-dipolo indotto (una molecola apolare) (o **forza di Debye**);
3. forza dipolo indotto istantaneo-dipolo indotto istantaneo (o **forza di dispersione di London**)

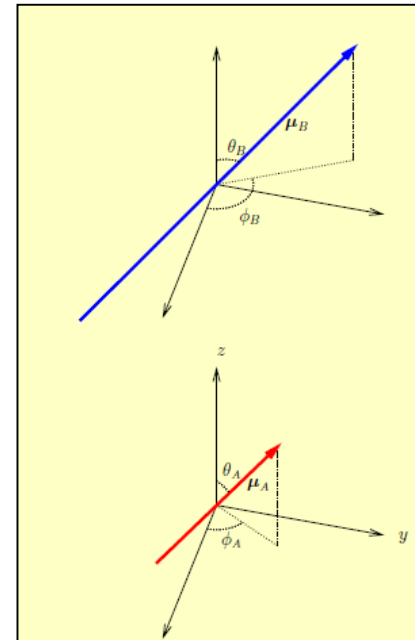
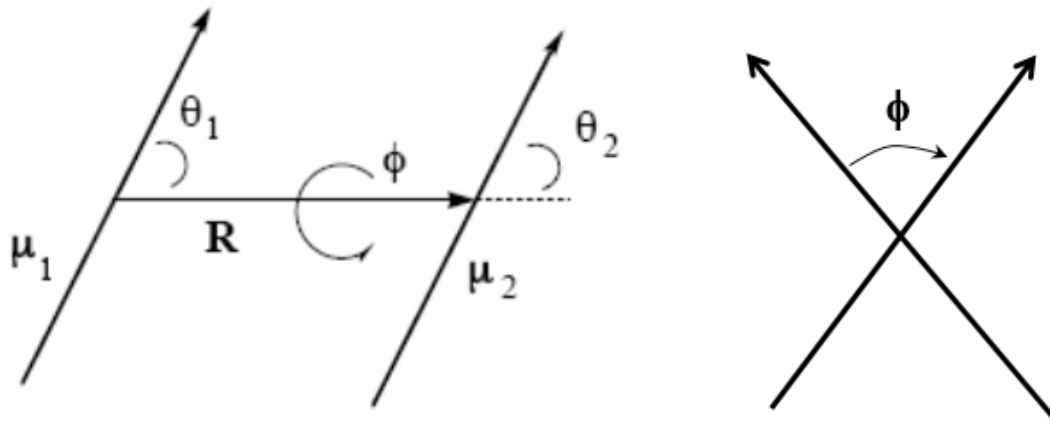
forza di Coulomb :

$$F = \frac{1}{4\pi\epsilon} \frac{q_1 q_2}{r^2}$$

legame idrogeno:

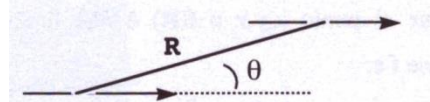
1. forza dipolo permanente-dipolo permanente o **forza di Keesom**

In questo caso le due molecole hanno momento di dipolo permanente, μ_1 e μ_2



$$V_{12} = \frac{\mu_1 \mu_2}{4\pi\epsilon_0 R_{12}^3} (\cos \phi - 3 \cos \theta_1 \cos \theta_2)$$

I) Per $0 < \theta_1 = \theta_2 = \theta < \pi/2$ (e quindi $\phi = 0$):

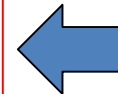
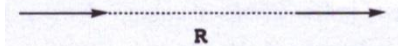


$$V_{12} = \frac{\mu_1 \mu_2}{4\pi\epsilon_0 R^3} (1 - 3\cos^2\theta)$$

II) per $\theta_1 = \theta_2 = 0$ (a), $\cos^2\theta = 1$, l'eq (6-4) diventa:

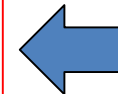
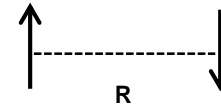
$$V_{12} = -\frac{2 \mu_1 \mu_2}{4\pi\epsilon_0 R^3}$$

in questo caso, $V_{12} = V_{\min}$: essa corrisponde all'interazione attrattiva *testa-coda*.



III) per $\theta_1 = \theta_2 = \pi/2$ e $\phi = 0$

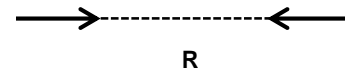
$$V_{12} = -\frac{\mu_1 \mu_2}{4\pi\epsilon_0 R^3}$$



IV) per $\theta_1 = 0, \theta_2 = \pi$ (a), e quindi $\phi = \pi$, $\cos\theta_1 = 1$ e $\cos\theta_2 = -1$:

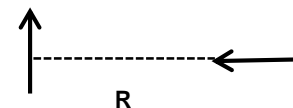
$$V_{12} = +\frac{2 \mu_1 \mu_2}{4\pi\epsilon_0 R^3}$$

In questo caso, V_{12} corrisponde all'interazione repulsiva *testa-testa*.



V) per $\theta_1=0, \theta_2=\pi/2$ e $\phi = \pi/2$

$$V_{12}=0$$



2. forza dipolo permanente-dipolo indotto (una molecola apolare) (o forza di Debye)

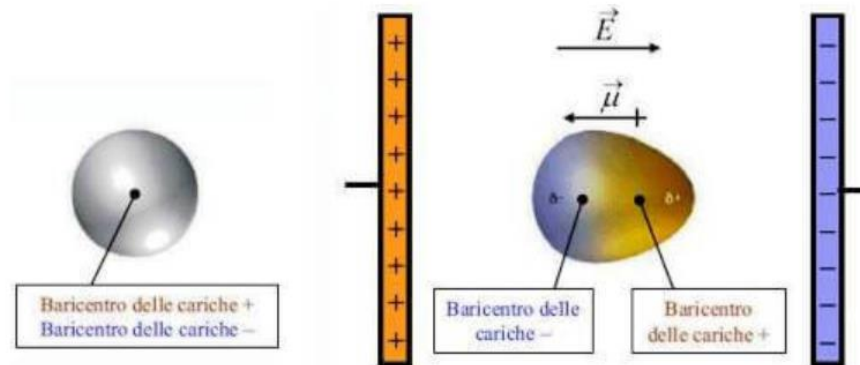
La forza elettrostatica attrattiva o repulsiva di un dipolo permanente influenza il moto degli elettroni delle molecole vicine (polari o apolari) trasformandole a loro volta in dipoli indotti.

Per comprendere tale interazione bisogna considerare cosa succede ad una **molecola apolare** (o a un atomo) quando viene immersa in un **campo elettrico**.

Nella molecola non polarizzata il baricentro delle cariche positive coincide con il baricentro delle cariche negative.

Quando la molecola viene immersa in un campo elettrico, succede che la nuvola elettronica carica negativamente viene deformata ed attratta dal polo positivo; viceversa il **nucleo dell'atomo** carico positivamente viene attratto dal polo negativo.

Il campo elettrico esterno pertanto induce una separazione di carica nella molecola e la formazione di un **dipolo temporaneo (dipolo indotto)**.



Tanto maggiore è intensità del campo elettrico E applicato tanto maggiore sarà l'intensità del dipolo indotto μ ;

- Si genera quindi un momento di dipolo elettrico indotto μ^{ind}
- La relazione tra \mathbf{E} e μ^{ind} definisce la matrice di polarizzabilità α

$$\mu^{\text{ind}} = \alpha \cdot \mathbf{E}$$

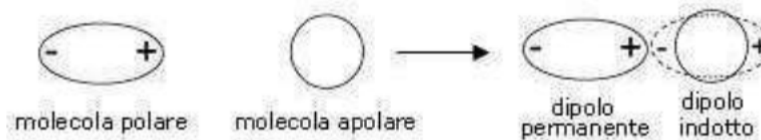
«Tensore polarizzabilità»

$$\begin{bmatrix} \mu_x^{\text{ind}} \\ \mu_y^{\text{ind}} \\ \mu_z^{\text{ind}} \end{bmatrix} = \begin{bmatrix} \alpha_{xx} & \alpha_{yx} & \alpha_{zx} \\ \alpha_{xy} & \alpha_{yy} & \alpha_{zy} \\ \alpha_{xz} & \alpha_{yz} & \alpha_{zz} \end{bmatrix} \cdot \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}$$

Quando una **molecola polare** si avvicina ad una molecola non polare, si ha lo stesso effetto visto in precedenza: la molecola polare induce un dipolo (**dipolo indotto**) sulla molecola non polare.

Tale dipolo perdura sino a quando le due molecole non si allontanano.

Si genera in questo modo un'**attrazione dipolo permanente-dipolo indotto**.



L'intensità della forza attrattiva dipende chiaramente dall'intensità del momento dipolare μ della molecola polare e dalla polarizzabilità della molecola non polare.

L'intensità dell'interazione (**interazione di Debye**) che si esercita tra una molecola polare di momento μ e una molecola non polare di polarizzabilità a poste ad una distanza r è:

$$E_{Debye} = -\frac{\mu^2 \cdot \alpha'}{4\pi\epsilon_0} \cdot \frac{1}{r^6}$$

in cui:

μ = momenti di dipolo della molecola polare (C · m)

ϵ_0 = **costante dielettrica del vuoto** = $8,854 \cdot 10^{-12}$ (C² · m⁻² · N⁻¹)

r = distanza tra le due molecole (m)

α' = volume di polarizzabilità (m³)

Il **volume di polarizzabilità** è una grandezza che ha le dimensioni di un volume e vale:

$$\alpha' = \frac{\alpha}{4\pi\epsilon_0}$$

3. forza dipolo indotto istantaneo-dipolo indotto istantaneo (**forza di dispersione di London**)



- ◇ Deriva dalle fluttuazioni della carica (e quindi della densità elettronica) di una molecola A . Queste fluttuazioni creano un campo elettromagnetico che interagisce con la densità di carica della seconda molecola B . Il risultato netto è un termine di energia attrattivo. ■
- L'espressione per l'energia potenziale si ricava attraverso un trattamento quantomeccanico dell'oscillatore armonico, in cui le frequenze di oscillazione per A e B sono ν_A e ν_B (modello di Drude),

$$U_{\mu_A^{\text{ind}} \mu_B^{\text{ind}}}(R) = -\frac{3h\alpha_A\alpha_B}{2(\epsilon_0\epsilon)^2} \frac{\nu_A\nu_B}{(\nu_A + \nu_B)} \frac{1}{R^6}$$

h è la costante di Planck.

L'energia media è sempre attrattiva e varia come R^{-6}

Tabella riassuntiva

Tipo di interazione	$U(R, \Omega)$
Carica-carica	$\frac{Q_A Q_B}{4\pi\epsilon_0\epsilon} \frac{1}{R}$
Carica-dipolo	$-\frac{1}{4\pi\epsilon_0\epsilon} \cos\theta \frac{Q_A \mu_B}{R^2}$
Dipolo-dipolo	$\frac{1}{4\pi\epsilon_0\epsilon} [\sin\theta_A \sin\theta_B \cos(\phi_A - \phi_B) - 2 \cos\theta_A \cos\theta_B] \frac{\mu_A \mu_B}{R^3}$
Carica-dipolo indotto	$-\frac{Q_A^2 \alpha_B}{2(4\pi\epsilon_0\epsilon)^2} \frac{1}{R^4}$
Dipolo-dipolo indotto	$-\frac{\mu_A^2 \alpha_B}{2(4\pi\epsilon_0\epsilon)^2} (3 \cos^2\theta + 1) \frac{1}{R^6}$
Dipolo-quadrupolo	...
Quadrupolo-quadrupolo	...
Dipolo indotto-Dipolo indotto	$-\frac{3h\alpha_A\alpha_B}{2(\epsilon_0\epsilon)^2} \frac{\nu_A\nu_B}{(\nu_A+\nu_B)} \frac{1}{R^6}$

- Tutti i termini di dipolo R^{-6} sono compresi nel cosiddetto termine "attrattivo" dell'interazione di van der Waals, e rappresentano la parte sostanziale dell'interazione a corto raggio.

Forze di Van der Waals

Attrazione di van der Waals

In generale, l'interazione attrattiva totale fra molecole *neutre*, polari o apolari, denominata attrazione di van der Waals, dipende dal contributo delle interazioni descritte precedentemente, dipolare, di induzione e di dispersione. Dato che tutte e tre dipendono dall'inverso della sesta potenza della distanza si può esprimere unitariamente l'energia di attrazione come

$$U = -\frac{A}{r^6}$$

dove la costante di proporzionalità A dipende dall'identità delle molecole interagenti.

Repulsione di van der Waals

Tutti gli atomi e le molecole si respingono a brevi distanze, a causa della repulsione dei nuclei e degli elettroni, chiamata repulsione sterica o di van der Waals. La repulsione avviene sia fra due atomi di una stessa molecola (*intramolecolare*), sia fra due atomi di molecole diverse (*intermolecolare*). L'energia di interazione della repulsione è di solito rappresentata dalla relazione

$$U = \frac{B}{r^{12}}$$

dove la costante di proporzionalità B dipende dalla natura degli atomi. La repulsione cresce inversamente a r^{12} , cioè molto bruscamente.

Potenziale di Lennard-Jones

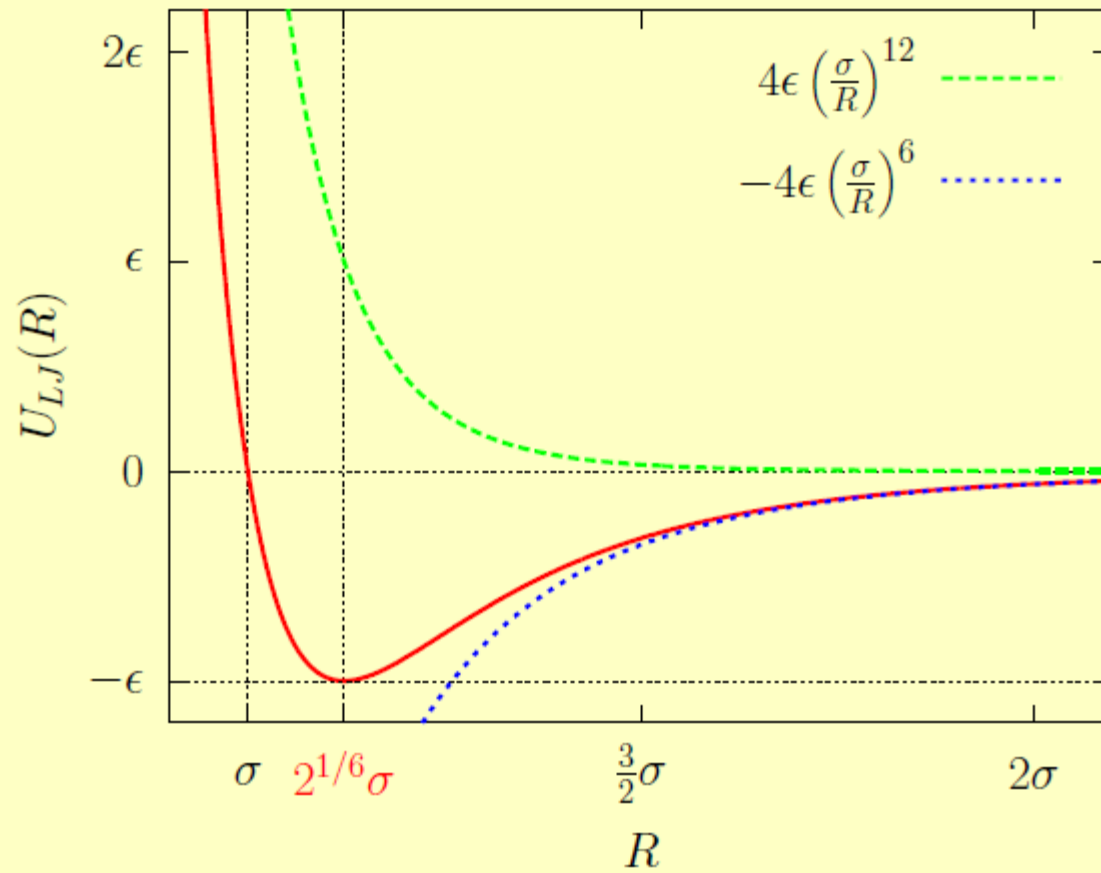
- ◇ E' uno dei potenziali più frequentemente utilizzati per descrivere l'interazione tra due molecole prive di carica.

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

- σ è il raggio della sfera che approssima l'atomo o la molecola in un modello a sfera rigida.
- E' costituito da una parte repulsiva di tipo R^{-12} e una parte attrattiva di tipo van der Waals R^{-6} . I due parametri che lo caratterizzano sono la distanza σ che annulla U_{LJ} e la profondità della buca ϵ . La distanza d'equilibrio risulta $2^{1/6}\sigma \approx 1.122\sigma$.

Esempio: per H_2O si ha $\sigma = 3.1655\text{\AA}$ e $\epsilon = 0.6502 \text{ kJ/mol}$

Plot di $U_{LJ}(R)$



Legame idrogeno

- ◇ Il legame idrogeno è un caso particolare di interazione fra dipolo-dipolo fra dipoli permanenti.



- X è un atomo molto elettronegativo come N, O, o F, che attrae gli elettroni di valenza, acquisendo una parziale carica negativa δ^- lasciando l'idrogeno con una parziale carica positiva δ^+ . Il gruppo $X - H$ viene detto donatore D.
- Il legame idrogeno si forma quando la carica positiva su H viene in contatto con un doppietto elettronico di un gruppo funzionale di un'altra molecola, il quale lega l'H e viene definito accettore A.

Legame idrogeno

D	A	R_{AD} (Å)
N – H	:: O =	2.90
N – H	: OH	2.90
N – H	: N =	3.10
N – H	: S	3.70
O – H	:: O =	2.75
O – H	: OH	2.75

- ◇ L'energia del legame idrogeno $U_{HB} \simeq 40$ kJ/mol è inferiore al legame covalente ($\simeq 500$ kJ/mol), ma superiore alla energia di Van der Waals ($\simeq 1$ kJ/mol).

Legame idrogeno

- Una descrizione quantitativa di U_{HB} si ha con un trattamento quantomeccanico

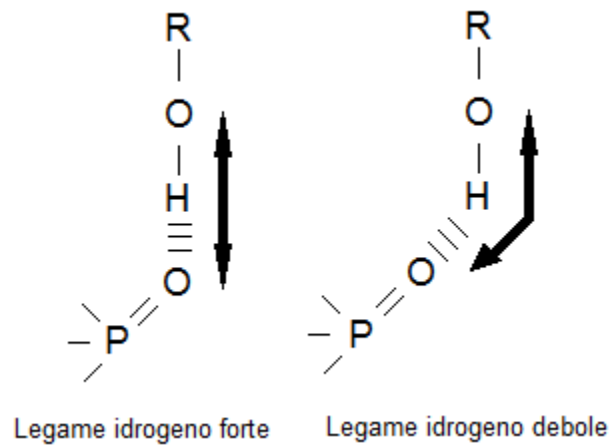
$$\psi_{HB} = a\psi_{A,D}^0 + b\psi_{A-D}^1$$

$\psi_{A,D}^0$ è la funzione d'onda di A e D separati, ψ_{A-D}^1 quella di una coppia ionica in cui un elettrone è passato da D a A. a e b sono due parametri "aggiustabili".

- ◇ Una recente approssimazione dell'energia di interazione risulta

$$U_{HB} = \cos \theta \left(\frac{B'_{ij}}{R_{ij}^{12}} - \frac{A'_{ij}}{R_{ij}^6} \right) + (1 - \cos \theta) \left(\frac{B_{ij}}{R_{ij}^{12}} - \frac{A_{ij}}{R_{ij}^6} \right)$$

θ è l'angolo DHA, le costanti A'_{ij} , B'_{ij} , A_{ij} e B_{ij} dipendono dalla coppia ij accettore/donatore.



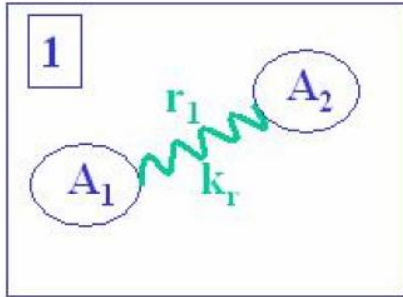
L'attrazione i dipoli è massima quando gli atomi coinvolti (in questo caso O, H e O) sono disposti in **linea retta** sullo stesso piano.

Energia e conformazione di una macromolecola

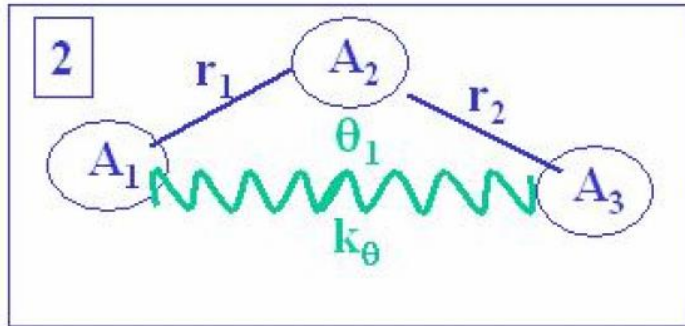
- ◇ Si studia nel dettaglio l'energia potenziale conformazionale di una macromolecola. Si distingue tra energia potenziale di legame (b) e di non-legame (nb)

$$U = U^b + U^{nb}$$

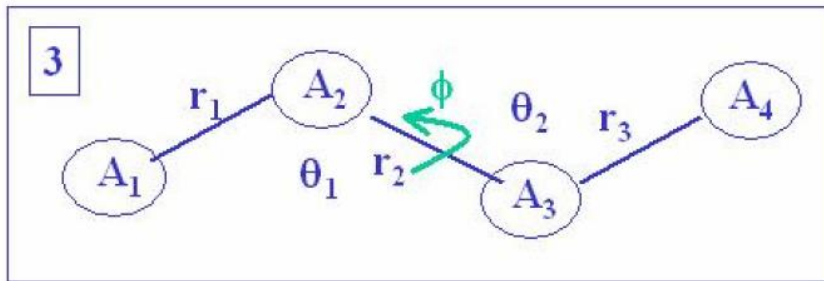
- Il termine U^b rappresenta l'energia conformazionale dovuta ai legami chimici. Dipende:
 1. Dalla distanza di legame r di ciascun legame
 2. Dall'angolo di legame θ tra 3 atomi adiacenti
 3. Dall'angolo diedro ϕ tra 4 atomi adiacenti



$$U(r) = \frac{1}{2} K_r (r - r_0)^2$$



$$U(\vartheta) = \frac{1}{2} K_\vartheta (\vartheta - \vartheta_0)^2$$



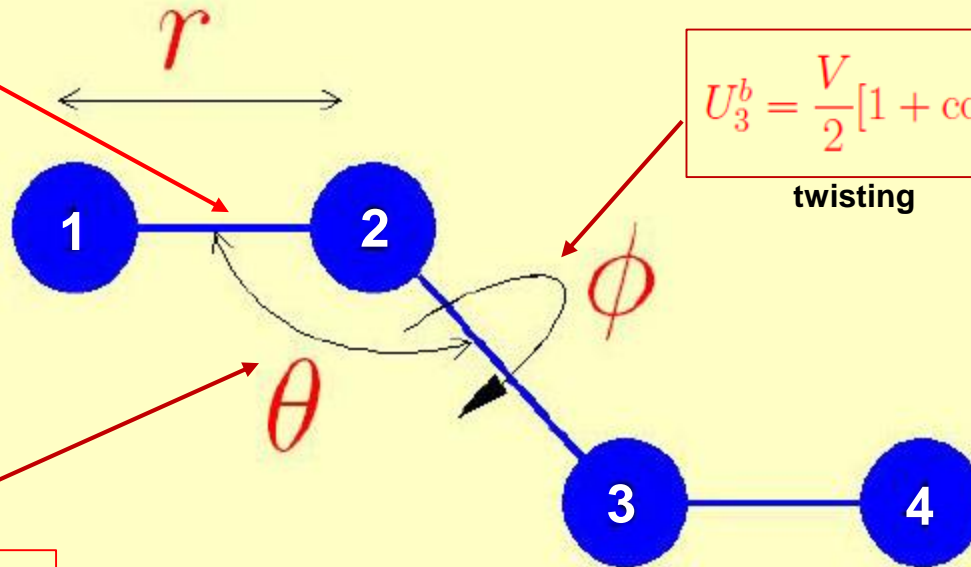
$$U(\varphi) = \sum_k c_k \cos^k \varphi \approx \frac{1}{2} [1 + \cos(3\varphi)]$$

Energia potenziale di legame

Quattro atomi adiacenti

$$U_1^b = K_1(r - r_0)^2$$

stretching

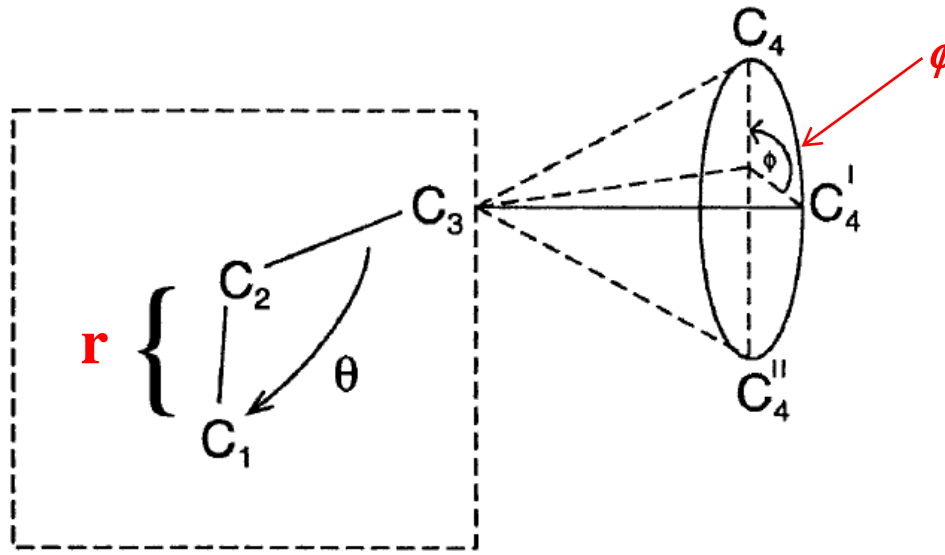


$$U_3^b = \frac{V}{2}[1 + \cos(n\phi - \gamma)]$$

twisting

$$U_2^b = K_2(\theta - \theta_0)^2$$

bending



La configurazione della catene di 4 atomi è determinata da 3 parametri: la lunghezza di legame r ; l'angolo di legame θ e l'angolo diedro ϕ .

Energia potenziale di non-legame

- ◇ E' dovuta a tutti gli altri effetti ad esclusione dei legami chimici. Distinguiamo almeno tre termini

1. Interazione di Coulomb

$$U_1^{nb} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon} \frac{1}{r_{ij}}$$

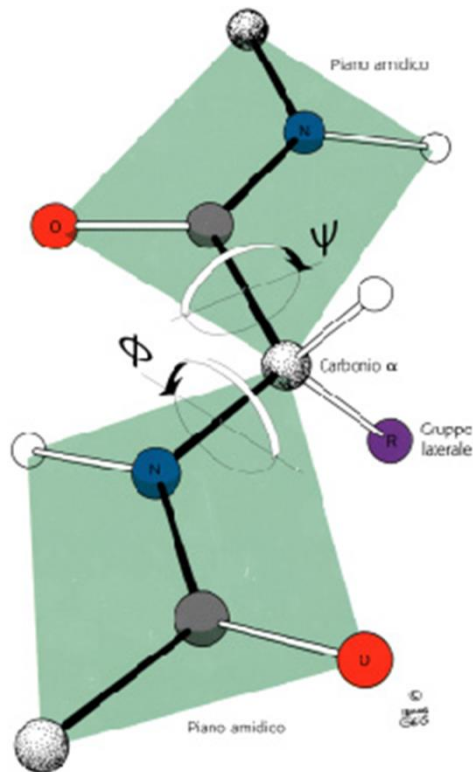
2. Interazione di Lennard-Jones

$$U_2^{nb} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

3. Legami idrogeno

$$U_3^{nb} = \cos \theta \left(\frac{B'_{ij}}{r_{ij}^{12}} - \frac{A'_{ij}}{r_{ij}^6} \right)$$

La conformazione della macromolecola definisce la disposizione degli atomi e dei gruppi atomici nello spazio, quale risulta a seguito della rotazione che avviene attorno ai legami che costituiscono lo scheletro della catena polimerica.



**Gli angoli ϕ e ψ
possono ruotare**

No-bonded interactions - Interazione tra **atomi non-legati**

$$U_C = \frac{1}{4\pi\epsilon_0\epsilon_r} \sum_{i>j} \frac{q_i q_j}{r_{ij}}$$

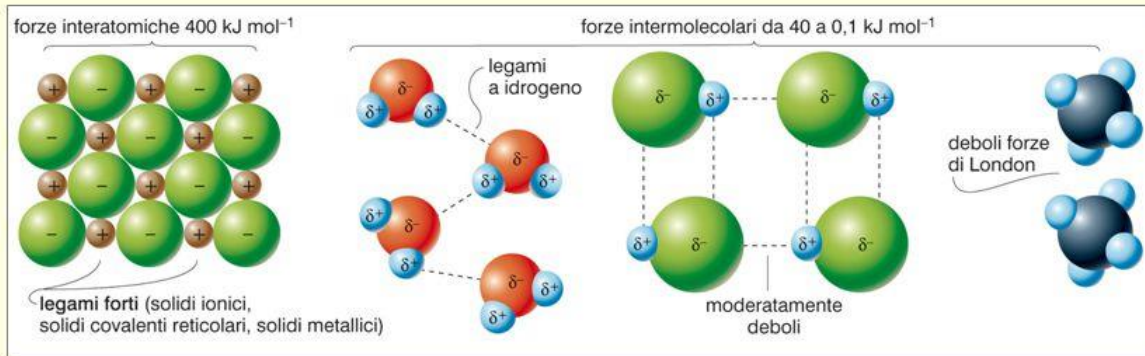
$$U_{HB} \text{ (legame idrogeno)}$$

$$U_{LJ} = \sum_{i>j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}$$

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

Legami a confronto

■ La scala di energia dei legami atomici è assai diversa da quella delle interazioni di *non-bonding* (van der Waals)



Interazione di van der Waals	0.1-10 (kJ/mole)
Legame H	2-50
Legame covalente	400

Scala dei tempi dei moti molecolari

stretching	1	→ 20 fs
modi elastici di domini	100 fs	→ 100 ps
riorientazione dell'acqua	4 ps	
bending di domini	10 ps	→ 100 ns
tumbling di proteine globulari	1	→ 10 ns
flipping dell'anello aromatico	100 μ s	→ 100 s

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

$E_{\text{stretch}} + E_{\text{bend}} = \text{costante}$



Rigid geometry approximation

Considerazioni statistiche

T << **273 K**: (al di sotto della temperatura di cristallizzazione) tutte le molecole avranno la stessa conformazione descritta dalla posizione più stabile .

T ambiente: in un sistema costituito da un numero N molto grande di molecole, in ciascun istante sono presenti tutte le conformazioni possibili, da quelle più stabili a quelle meno stabili.

Ma con **diversa probabilità**: per esempio, le conformazioni sfalsate sono più probabili di quelle eclissate.

La distribuzione delle diverse situazioni nel sistema è descritta dall'**equazione di Boltzman**, che fornisce la probabilità p_i di ciascun stato conformazionale i (definito dall'angolo di rotazione) in funzione dell'energia conformazionale $E(\varphi)$ e della temperatura T :

La **probabilità di occupazione** del j -esimo stato è:

$$p_j = \frac{e^{-\frac{E_j}{kT}}}{q}$$

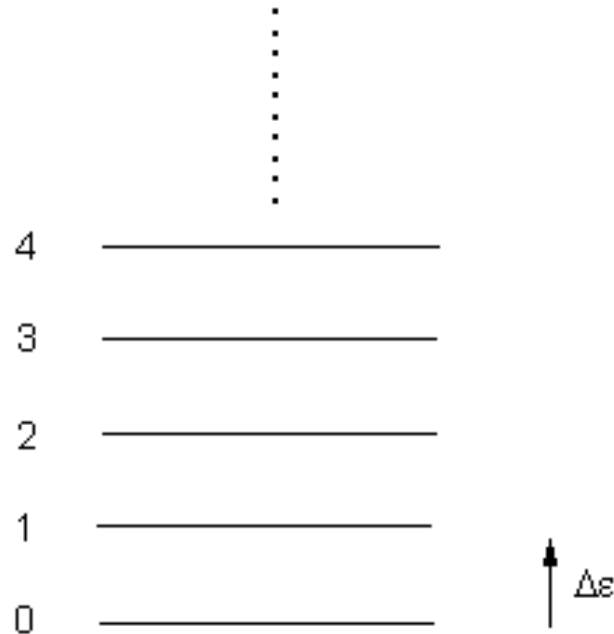
dove k è la costante di Boltzman e q la **funzione di partizione**:

$$q = \sum_j e^{-\frac{E_j}{kT}}$$

ma prima un ripasso di termodinamica statistica

FUNZIONE DI PARTIZIONE

Consideriamo un equilibrio semplice tra più livelli energetici non degeneri equispaziati



Il rapporto tra le probabilità di occupazione tra due stati è:

$$\frac{P_i}{P_{i-1}} = \exp\left(-\frac{\Delta\varepsilon}{RT}\right)$$

$$\frac{P_1}{P_0} = \exp\left(-\frac{\Delta\varepsilon}{RT}\right)$$

$$\frac{P_2}{P_1} = \exp\left(-\frac{\Delta\varepsilon}{RT}\right)$$

Quindi

$$\frac{P_2}{P_1} \times \frac{P_1}{P_o} = \frac{P_2}{P_o} = \exp\left(-2 \frac{\Delta\varepsilon}{RT}\right)$$

generalizzando

$$\frac{P_n}{P_o} = \exp\left(-n \frac{\Delta\varepsilon}{RT}\right)$$

Conoscendo P_o e $\Delta\varepsilon$ conosciamo anche tutti gli altri P_n

Poichè

$$\sum_{n=0}^{\infty} P_n = 1$$

e

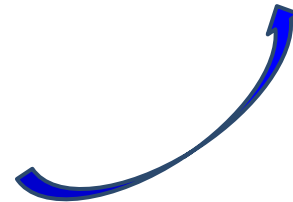
$$P_n = P_o \exp\left(-n \frac{\Delta\varepsilon}{RT}\right)$$

si ha

$$P_o \sum \exp\left(-\frac{n\Delta\varepsilon}{RT}\right) = 1$$

e

$$P_o = \frac{1}{\sum_{n=0}^{\infty} \exp\left(-\frac{n\Delta\varepsilon}{RT}\right)}$$



In generale

$$P_n = \frac{\exp\left(-\frac{n\Delta\varepsilon}{RT}\right)}{q}$$

dove

$$q = \sum_{n=0}^{\infty} \exp\left(-\frac{n\Delta\varepsilon}{RT}\right)$$

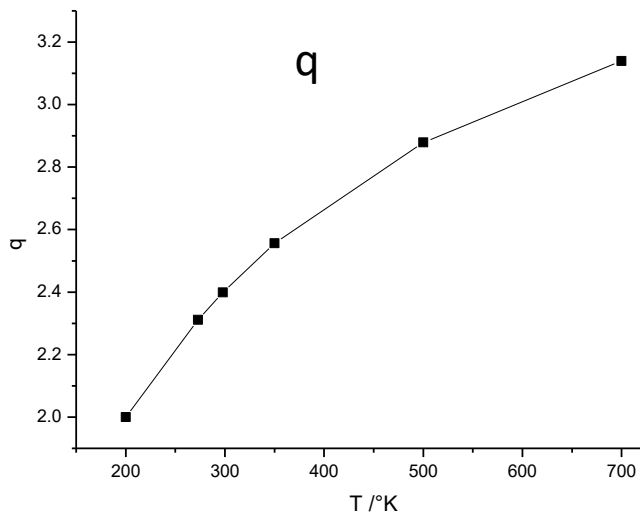
viene chiamata **funzione di partizione**

T bassa (stato solido):

sono privilegiati gli stati conformazionali ad energia più bassa,

T alta:

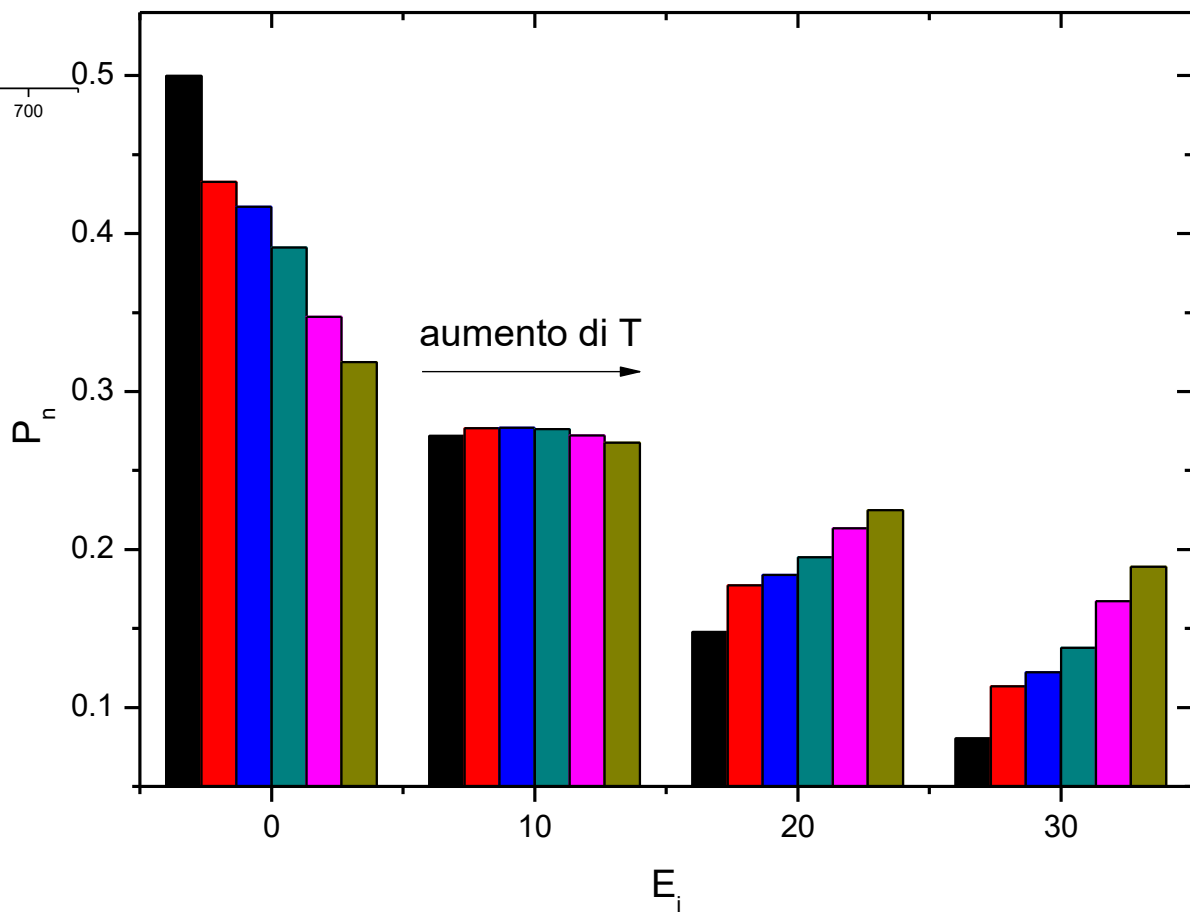
in assenza di vincoli di simmetria o di impacchettamento (stato fuso o soluzione), troveremo una **distribuzione di tutti gli stati conformazionali** in sequenza statistica e intercambiabili gli uni negli altri (struttura dinamica) con una frequenza che dipende dall'esponente della probabilità di Boltzman.



$$p_j = \frac{e^{-\frac{E_j}{kT}}}{q}$$

$$q = \sum_j e^{-\frac{E_j}{kT}}$$

$E_4=30$ _____
 $E_3=20$ _____
 $E_2=10$ _____
 $E_1=0$ _____



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:

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



n. 1844

† Duino, 5 settembre 1906



$$U \equiv \bar{E} = \frac{\sum_j E_j e^{-\frac{E_j}{kT}}}{q}$$

$$\begin{aligned} \frac{\partial}{\partial T} \exp\left(-\frac{E_i}{kT}\right) &= \exp\left(-\frac{E_i}{kT}\right) \frac{\partial}{\partial T} \left(-\frac{E_i}{kT}\right) \\ &= -\frac{E_i}{kT^2} \exp\left(-\frac{E_i}{kT}\right) \end{aligned}$$

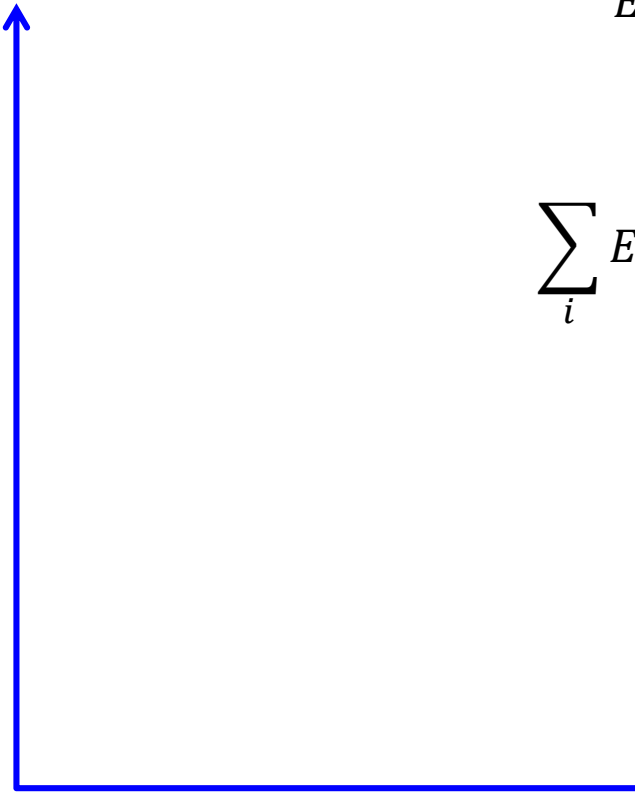
$$E_i \exp\left(-\frac{E_i}{kT}\right) = kT^2 \frac{\partial}{\partial T} \exp\left(-\frac{E_i}{kT}\right)$$

$$\sum_i E_i \exp\left(-\frac{E_i}{kT}\right) = kT^2 \frac{\partial}{\partial T} \sum_i \exp\left(-\frac{E_i}{kT}\right)$$

$$\sum_i \exp\left(-\frac{E_i}{kT}\right) = kT^2 \frac{\partial q}{\partial T}$$

$$U = \frac{kT^2}{q} \frac{\partial q}{\partial T}$$

$$U = kT^2 \frac{\partial \ln q}{\partial T}$$



La **varianza** nell'energia (o *fluttuazione dell'energia*) è

$$\langle(\delta E)^2\rangle \equiv \langle(E - \langle E \rangle)^2\rangle = \frac{\partial^2 \ln q}{\partial \beta^2}. \quad \beta = 1/kT$$

L'**entalpia** è

$$H = -\frac{\partial \ln q}{\partial \beta} + k_B T V \frac{\partial \ln q}{\partial V}.$$

La **capacità termica** è

$$C_v = \frac{\partial \langle E \rangle}{\partial T} = \frac{1}{k_B T^2} \langle \delta E^2 \rangle.$$

L'**entropia** è

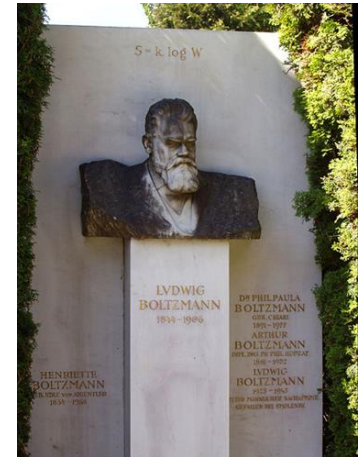
$$S \equiv -k_B \sum_j P_j \ln P_j = k_B (\ln q + \beta \langle E \rangle) = \frac{\partial}{\partial T} (k_B T \ln q) = -\frac{\partial A}{\partial T}$$

dove A è l'**energia libera di Helmholtz** definita come $A = U - TS$, ove $U = \langle E \rangle$ è l'energia totale e S è l'**entropia**, cosicché

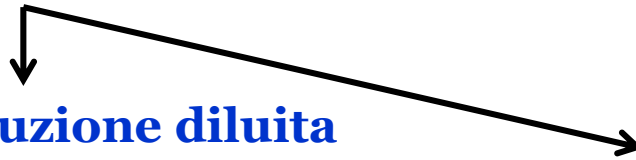
$$A = \langle E \rangle - TS = -k_B T \ln Z.$$

L'**energia libera di Gibbs** è

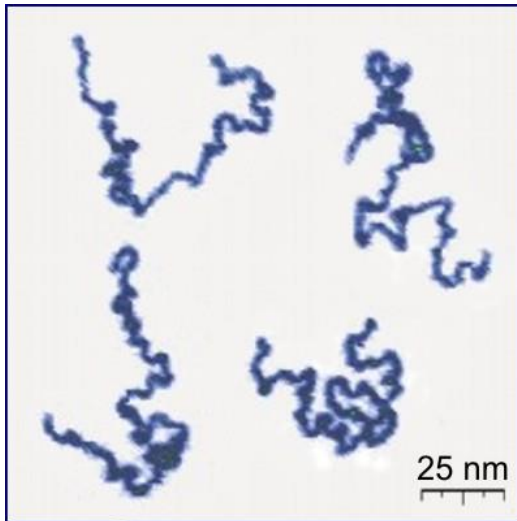
$$G = -k_B T \ln Z + k_B T V \frac{\partial \ln Z}{\partial V}$$



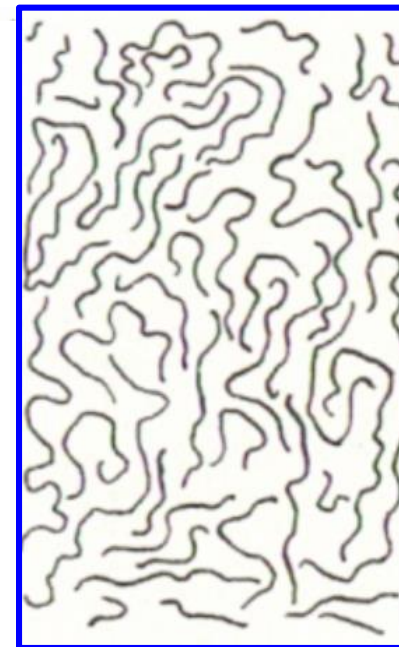
il sistema macroscopico è costituito da un numero molto grande di catene aventi tutte **conformazioni diverse** (*catene disordinate, stato amorfo*).



Polimeri in soluzione diluita



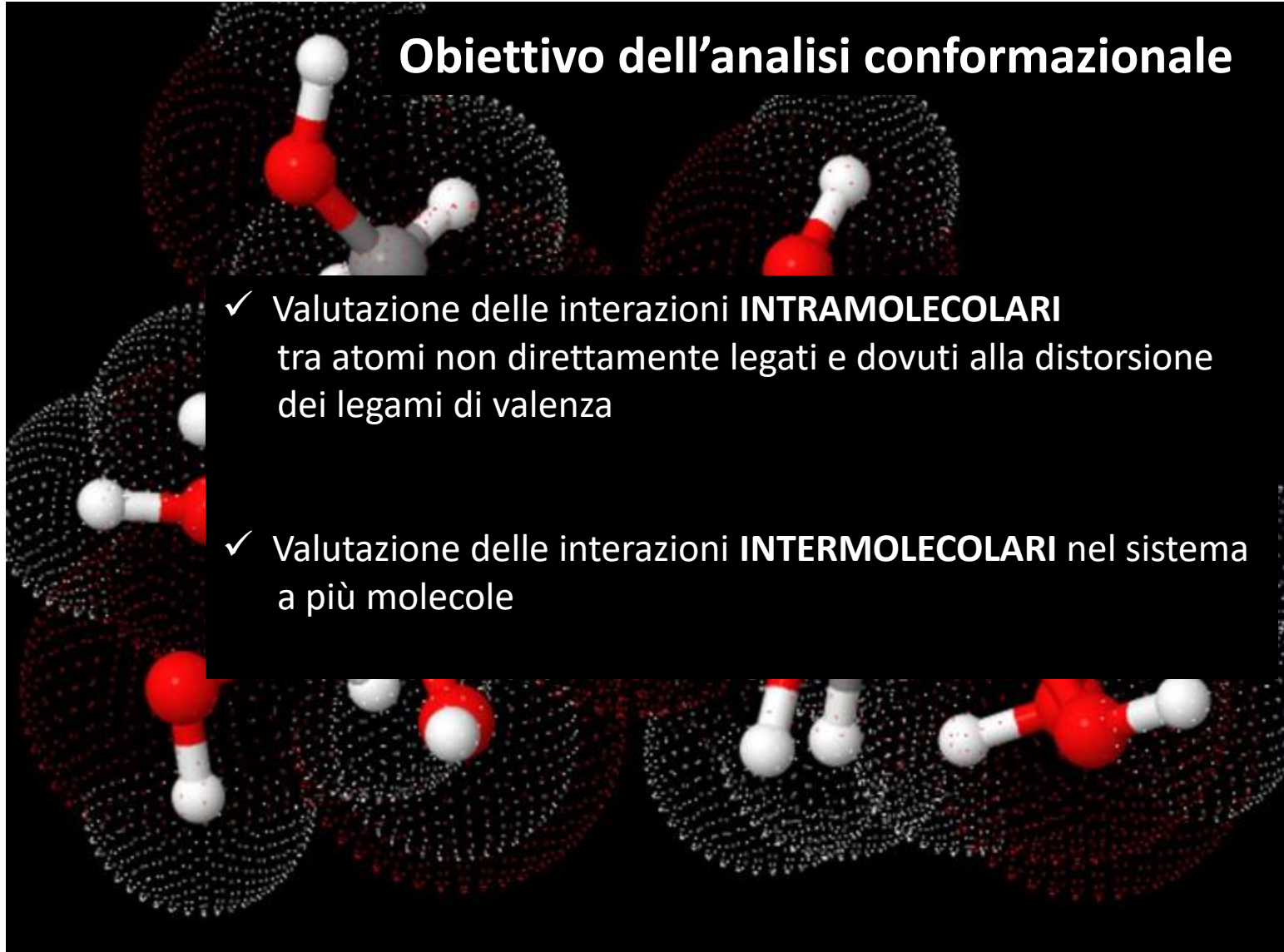
**Polimeri in soluzione concentrata
o allo stato fuso**



LE PROPRIETA' CONFORMAZIONALI

Obiettivo dell'analisi conformazionale

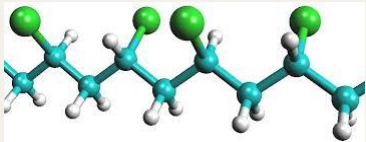
- ✓ Valutazione delle interazioni **INTRAMOLECOLARI** tra atomi non direttamente legati e dovuti alla distorsione dei legami di valenza
- ✓ Valutazione delle interazioni **INTERMOLECOLARI** nel sistema a più molecole



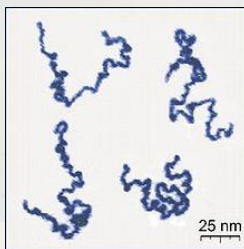
APPROCCIO TEORICO

CALCOLO CONFORMAZIONALE

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



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



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

Energia
conformazionale

scelta dei legami descrittivi
della catena (reali o virtuali)

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

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

$$A = f(q)$$

<A> = 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

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

<A> = media
temporale

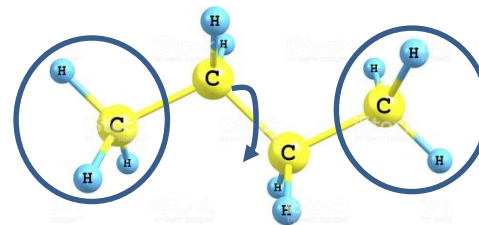
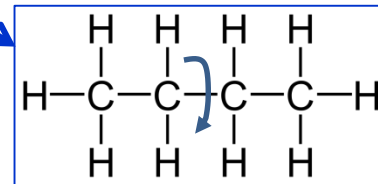
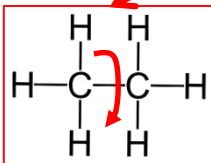
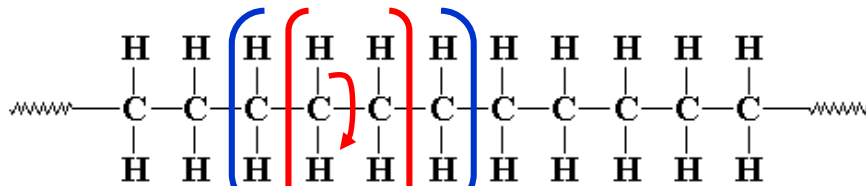
definizione di un campo di forza:

$$V(q) = k_b(b-b_0)^2 + k_\theta(\theta-\theta_0)^2 + k_\phi[1 + \cos(n\phi)] +$$
$$+ A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^6 + q_i q_j / r_{ij} + \dots$$

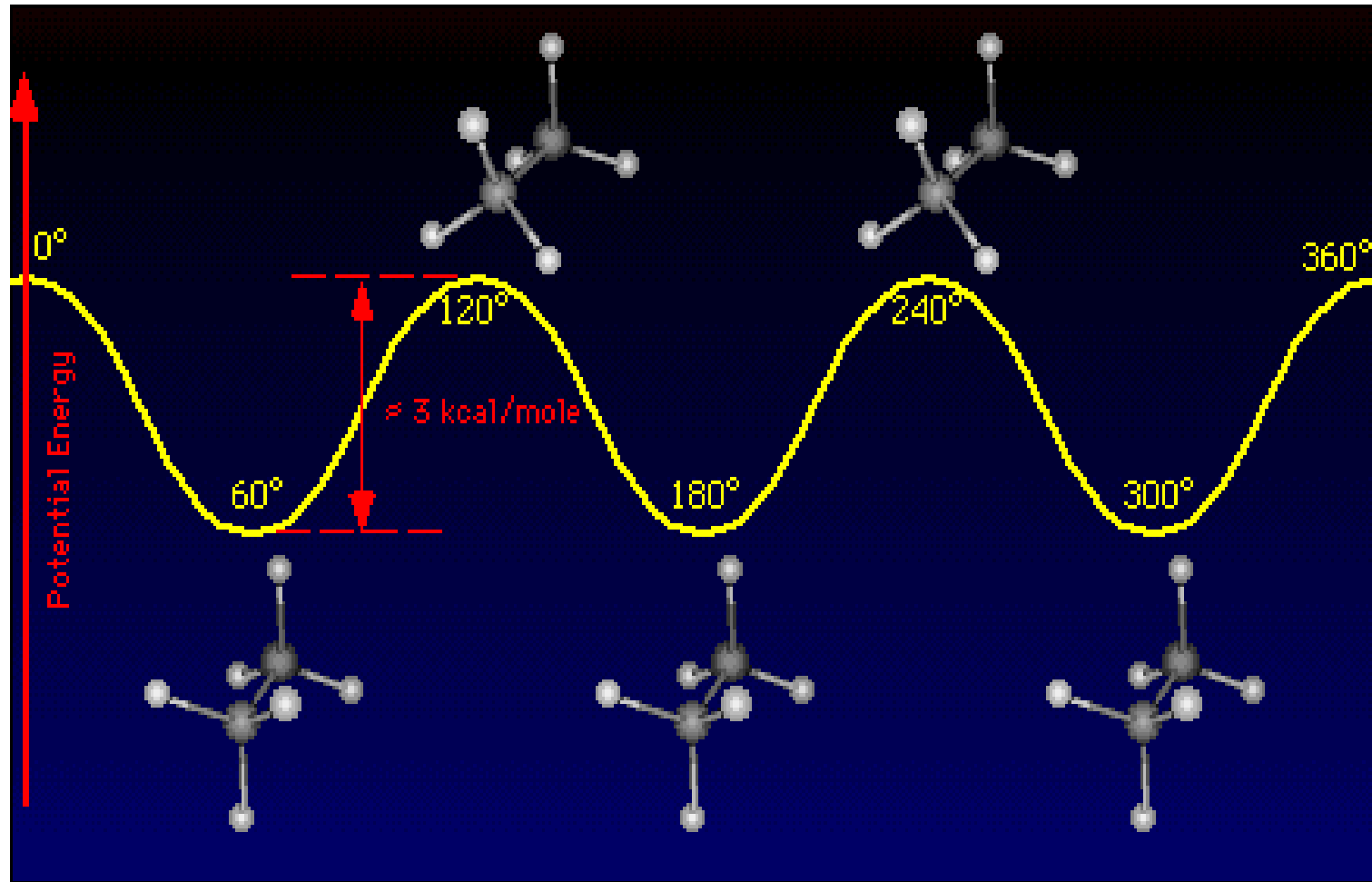
Scelta della unità ripetitiva o residuo che rappresenta il dato polimero



Minimo numero di atomi sufficienti alla valutazione dell'energia potenziale (conformazionale) del residuo che rappresenta la data macromolecola



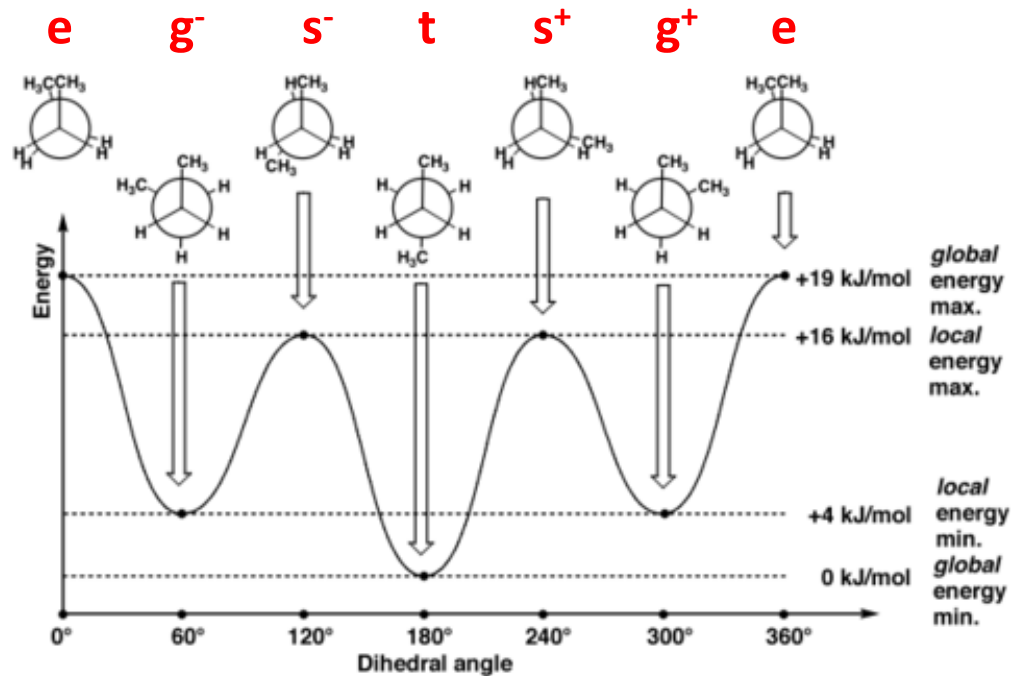
Energia potenziale



Le diverse conformazioni sono quindi regolate dai legami chimici che costituiscono la catena.

- il legame doppio **C=C** ha una libertà rotazionale molto limitata;
- il legame **C-C** ha maggior libertà rotazionale, ma limitata da vicoli geometrici e sterici.

**Esempio:
butano**



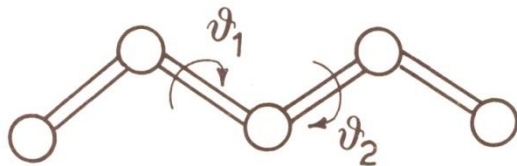
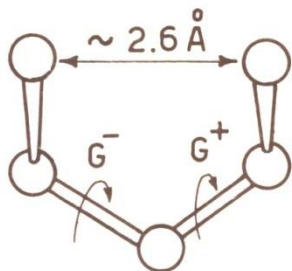


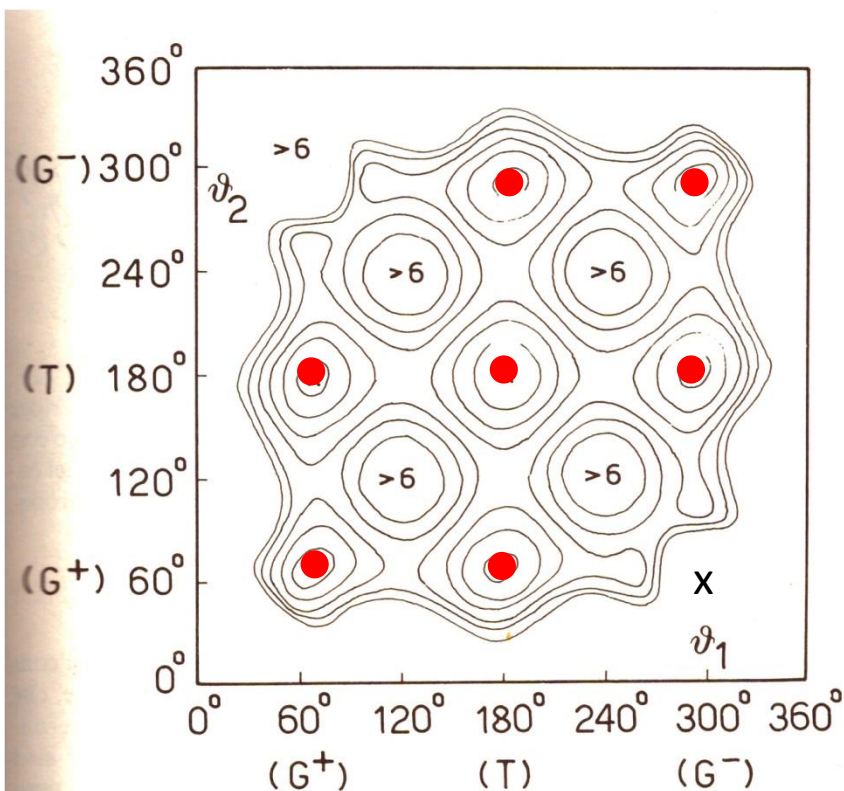
Fig. 1.3.8 - a) Ignorando gli atomi di idrogeno, la conformazione del *n*-pentano è definita da due angoli di rotazione interna θ_1 e θ_2 .



b) Se $\theta_1 = 300^\circ$ (G^-) e $\theta_2 = 60^\circ$ (G^+), o viceversa, i gruppi metilici terminali si trovano molto più vicini della distanza di van der Waals ($\sim 4 \text{ \AA}$).

$$E_{conf} = E(\vartheta_1, \vartheta_2)$$

n-pentano



c) Energia conformazionale del *n*-pentano calcolata secondo la Eq. 1.3.9) in funzione di θ_1 e θ_2 ($E(180^\circ, 180^\circ) = 0$, curve di livello separate da 1 kcal/mol, minimi di energia indicati con crocette).

2,2,4,4-tetrametil--pentano

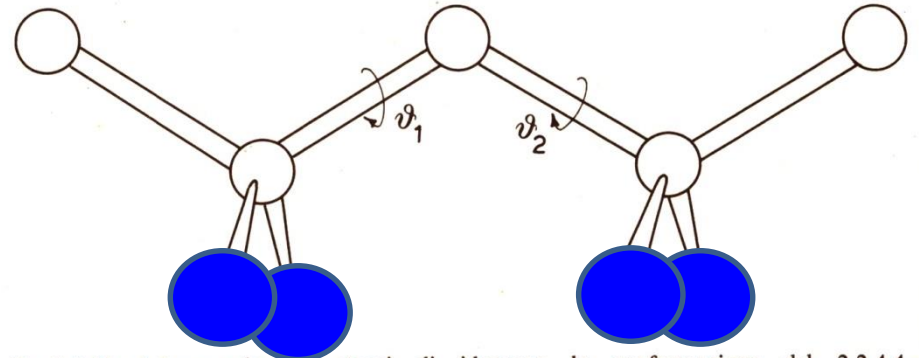
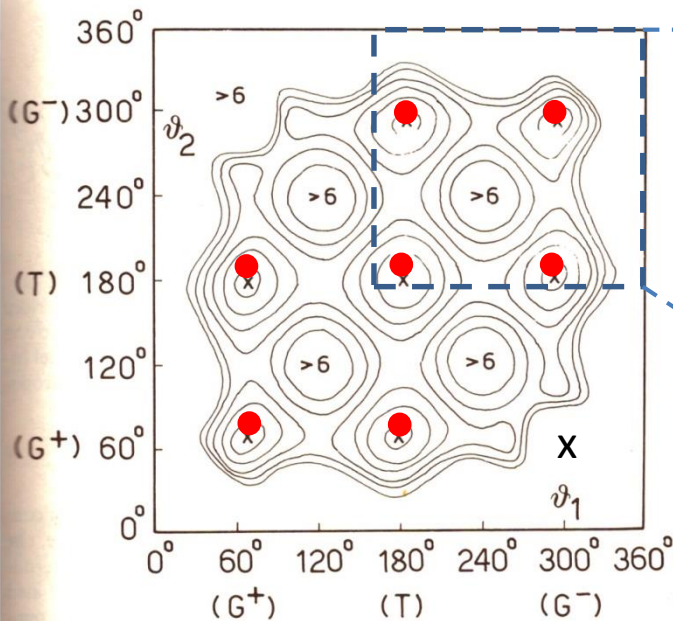
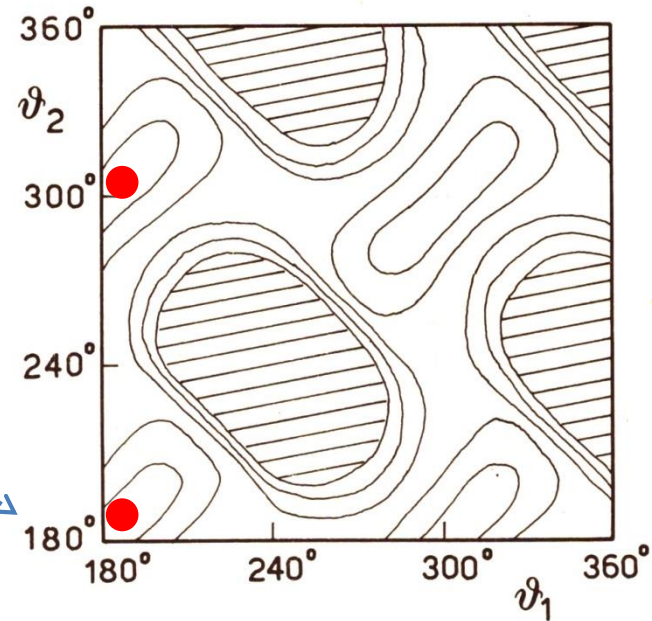


Fig. 1.3.11 - a) Ignorando gli atomi di idrogeno la conformazione del 2,2,4,4-tetrametilpentano è definita da due angoli di rotazione interna (θ_1 e θ_2).

n-pentano

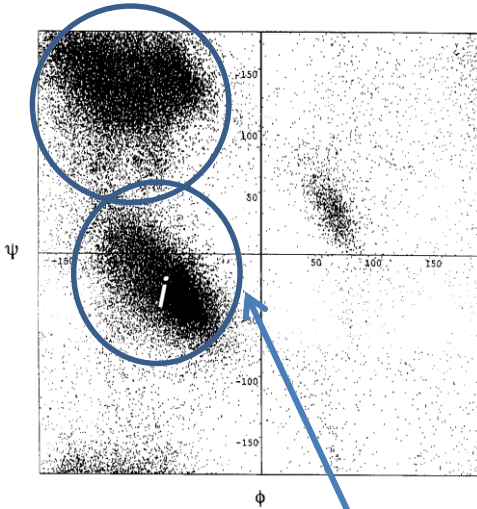


c) Energia conformazionale del n-pentano calcolata secondo la Eq. 1.3.9) in funzione di θ_1 e θ_2 ($E(180^\circ, 180^\circ) = 0$, curve di livello separate da 1 kcal/mol, minimi di energia indicati con crocette).



b) Energia conformazionale del 2,2,4,4-tetrametilpentano, calcolata secondo la Eq. 1.3.8) in funzione di θ_1 e θ_2 (vedi Fig. 1.3.11a). Curve di livello a 1,2...5 kcal/mol sopra i minimi; le zone tratteggiate corrispondono a $E > 5$ kcal/mol.

SIMULAZIONE DI CATENE POLIMERICHE



Ad ogni stato i viene assegnata una probabilità

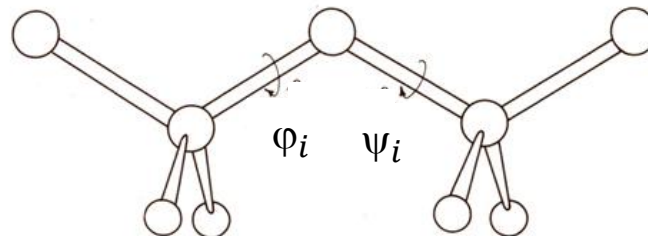
$$P_i(\varphi_i, \psi_i) = \frac{e^{-\frac{E_i}{RT}}}{\sum_i e^{-\frac{E_i}{RT}}}$$

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

Popolazione di conformazioni accessibili
(più probabili !)

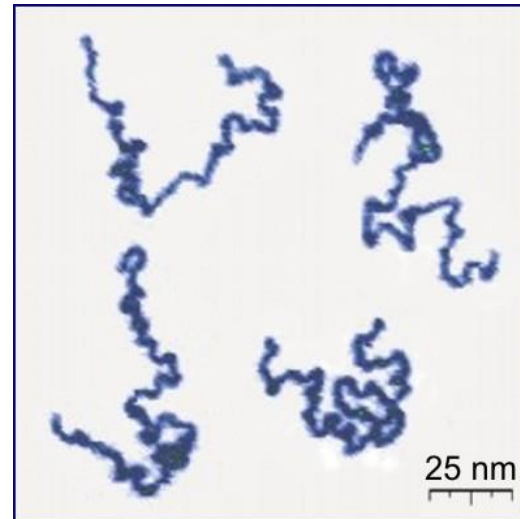
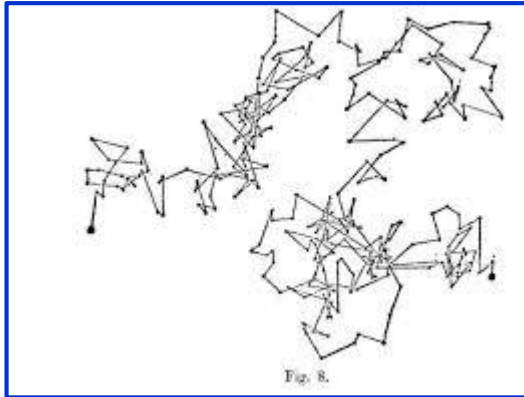
Insieme di catene polimeriche più probabili ma disordinate

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

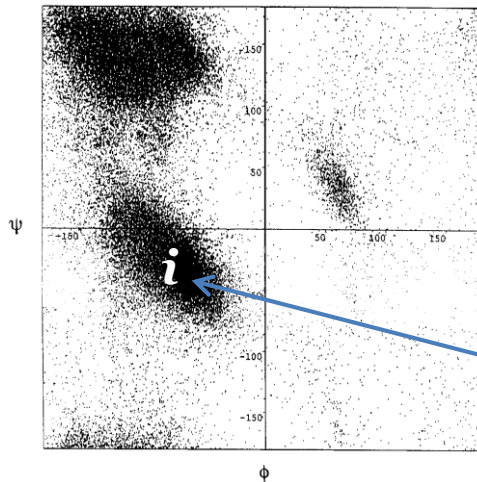


Stato solido amorfo,
Polimero fuso,
Polimero in soluzione

Stato solido amorfo,
Polimero fuso,
Polimero in soluzione



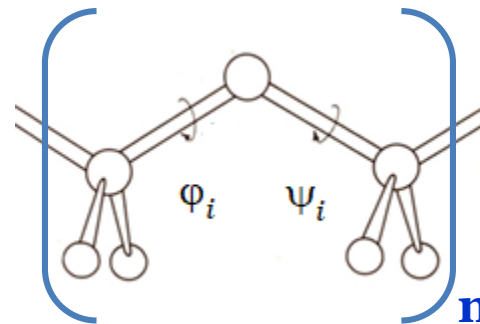
SIMULAZIONE DI CATENE POLIMERICHE



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

Una catena polimerica **ordinata (elicoideale)** è descritta da **una sola coppia** di angoli torsionali lungo tutta la catena di **n** unità ripetitive

$$\{(\phi_i, \psi_i)(\phi_i, \psi_i) \dots (\phi_i, \psi_i)\}_n$$



Predizione della struttura: Meccanica Molecolare (MM) Dinamica Molecolare (MD)

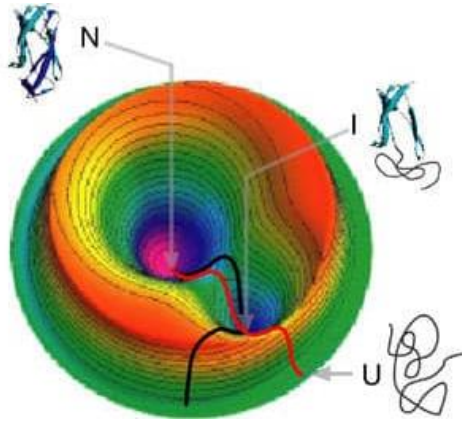
MM ed MD sono approcci computazionali molto usati sia nella scienza dei materiali che nello studio di biomolecole.

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

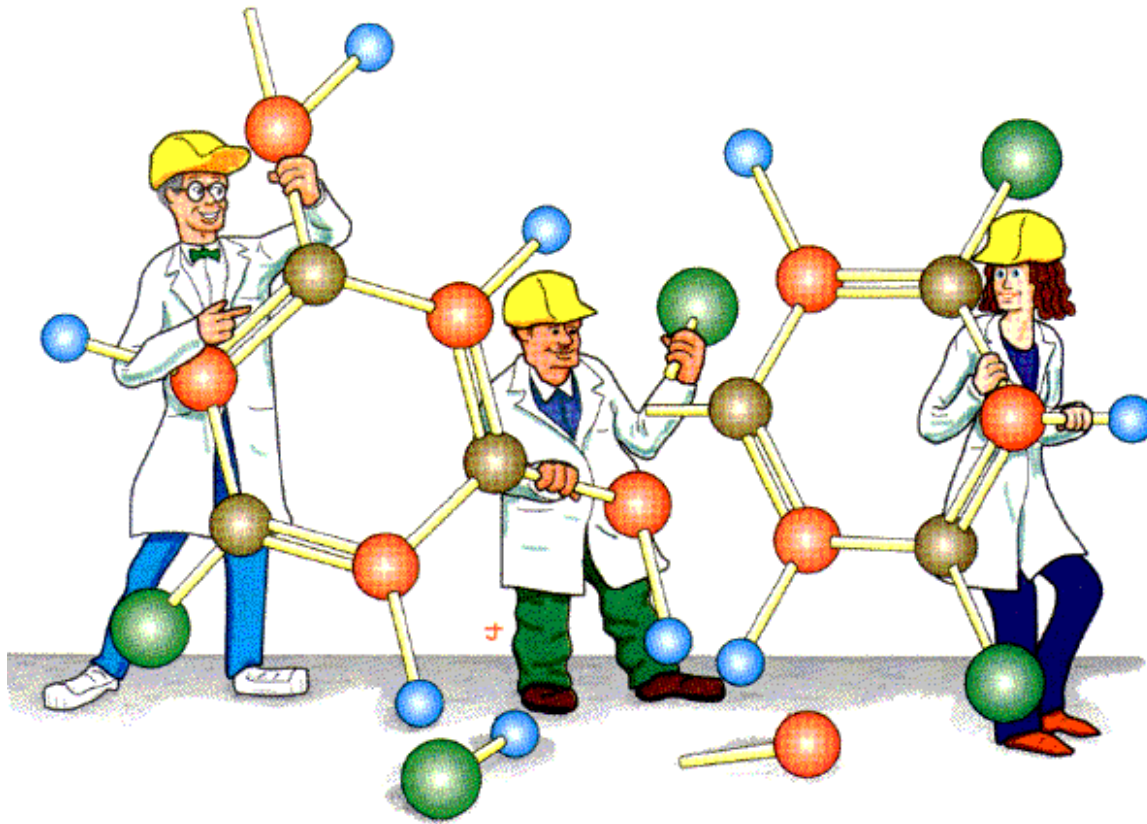
Con la **MM/MD** possono essere studiati in particolare:



- stabilità di piccole molecole(oligomeri);
- cambi conformazionali
- polymer folding
- soluzioni polimeriche
- interazioni tra segmenti polimerici
- 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.

Modellizzazione Molecolare



Limiti delle tecniche sperimentali

risoluzione spaziale



misure ad alta risoluzione di strutture molecolari sono possibili solo per sistemi relativamente rigidi (cristalli, fibre,...)

risoluzione energetica



analisi delle energie di interazione atomica difficoltosa

risoluzione temporale



i primissimi eventi nelle reazioni chimiche e nei processi biologici sono spesso di difficile misurazione

Limiti delle tecniche computazionali

Sistemi molecolari troppo complessi

Si ricorre alla meccanica classica con funzioni di interazione **semi-empiriche** per descrivere le interazioni tra gli atomi del sistema molecolare

Simulazione del comportamento di un sistema molecolare

Solo per un **numero limitato** di atomi o di gradi di libertà (di solito 10^2 - 10^5 atomi), per un limitato periodo di tempo (10^2 - 10^4 picosecondi)

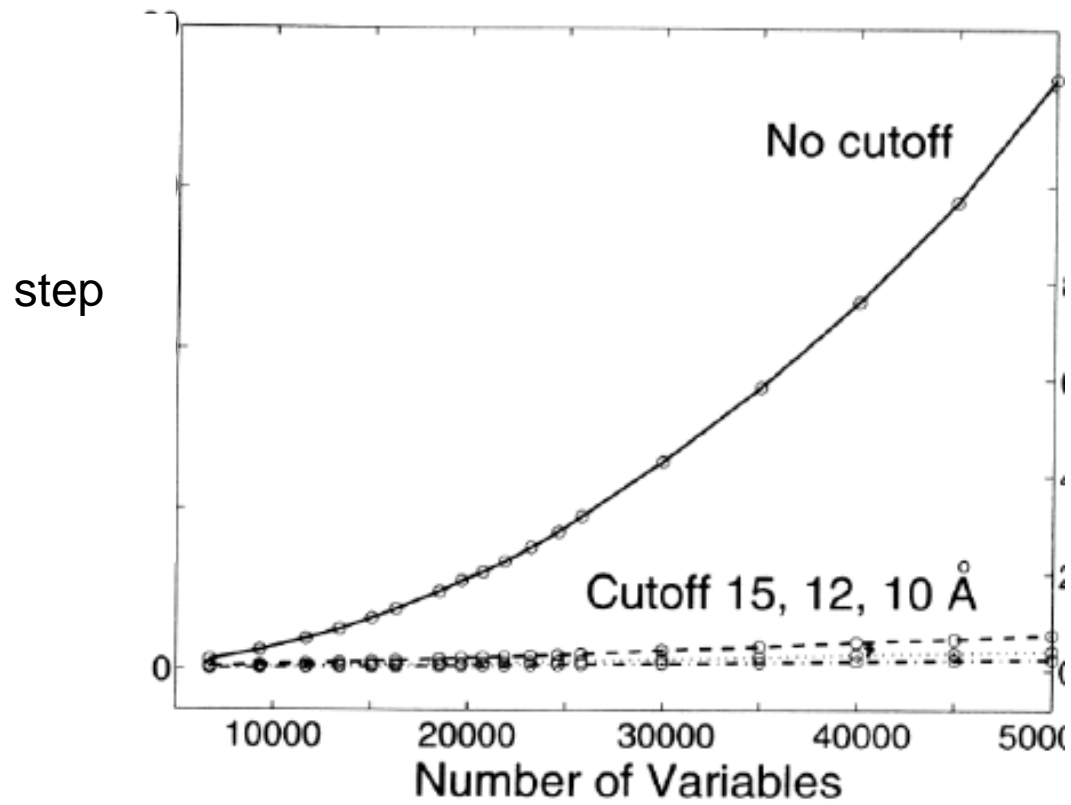
Campionatura limitata dello spazio delle conformazioni di una macromolecola

Si utilizzano dati sperimentali per restringerlo

Meccanica molecolare

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

Cutoff

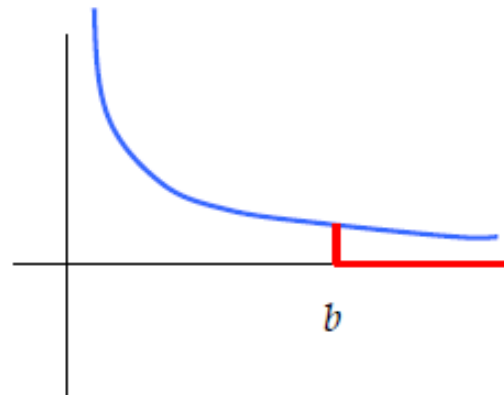


Cutoff

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

$S(r)$: cutoff function.

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



Conformazione energeticamente preferita



minimo globale dell'energia potenziale



$$F_i = -\partial V / \partial r_i$$



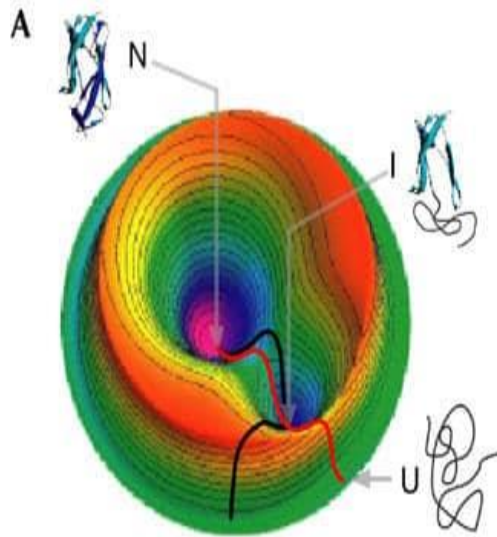
equilibrio



stato più popolato

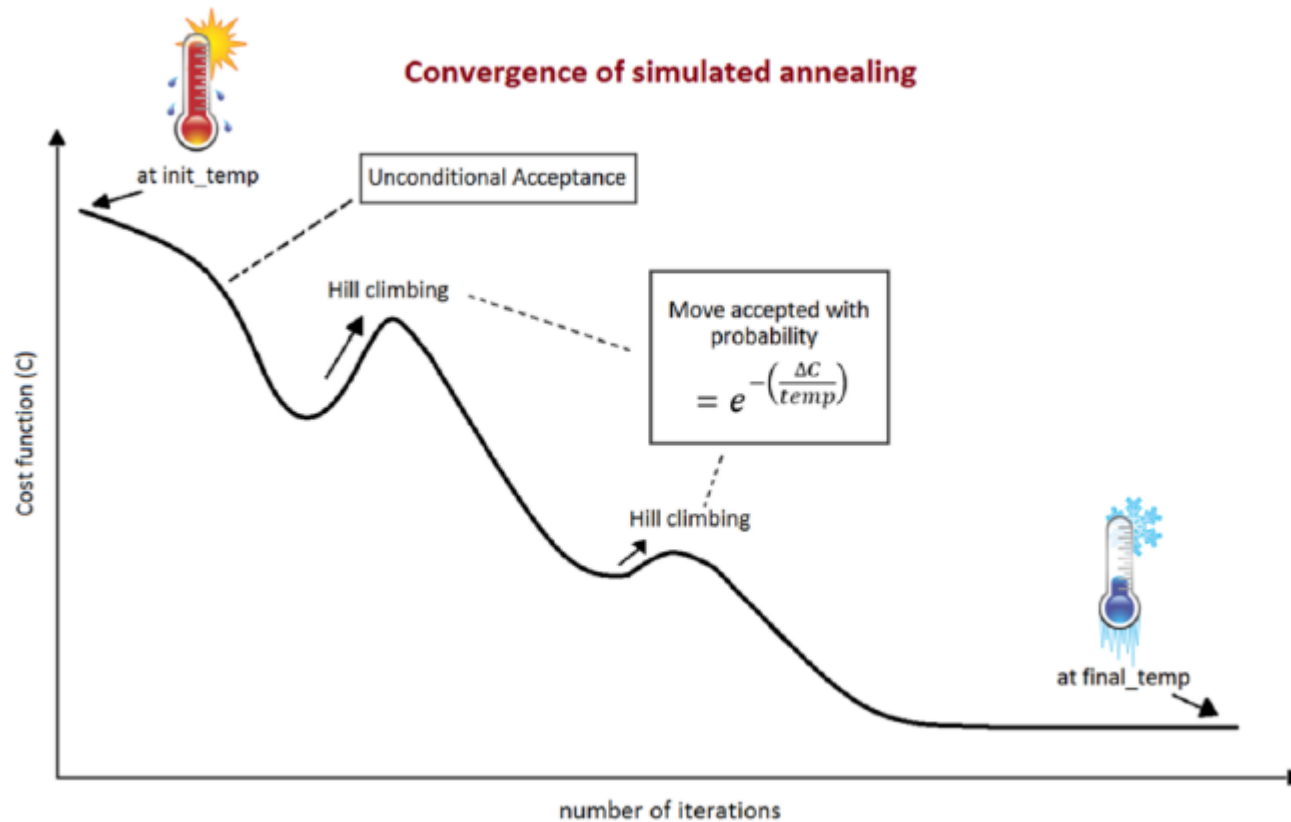
Minimizzazione dell'energia

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.



➤ **Annealing simulato:**

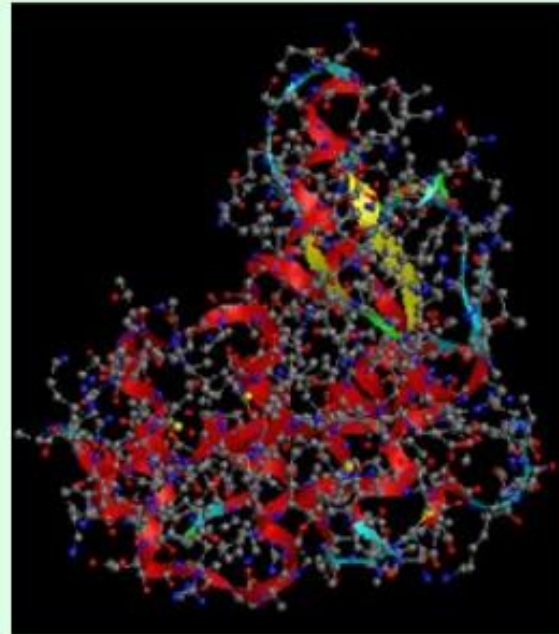
riscaldamento ad alta temperatura che modifica la struttura superando le barriere energetiche, seguito da un raffreddamento lento.



E' impossibile visualizzare l'intera superficie di energia potenziale, tranne che in casi molto semplici in cui l'energia è funzione di solo una o due coordinate.

Lisozima T4, 1328 atomi circa
Superficie di energia molto complicata.

Ipersuperficie 3984-dimensionale!!



Ricerca del minimo (minimizzazione o ottimizzazione)

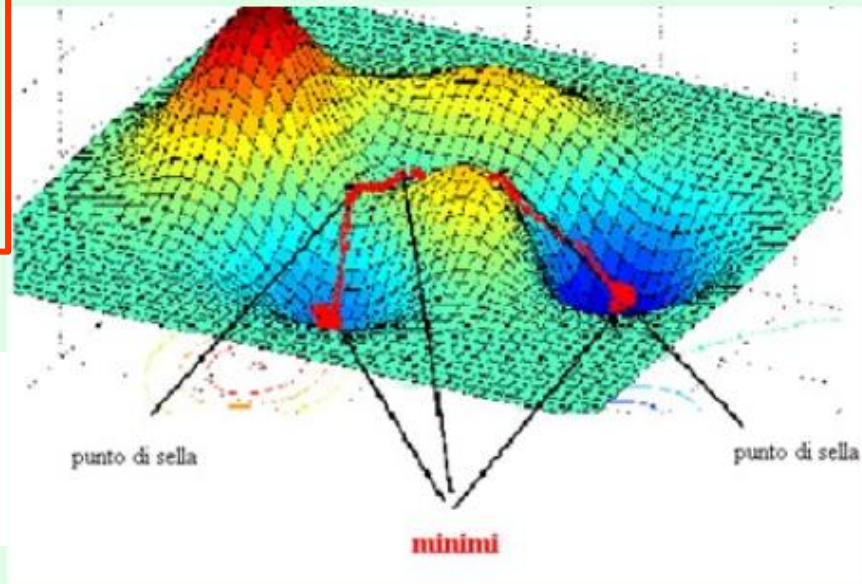
Quello della ricerca di un algoritmo di minimizzazione efficace è un problema molto noto in matematica. Nei casi complessi per determinare i minimi sulla superficie di energia potenziale in genere si ricorre ad algoritmi matematici basati su metodi numerici. C'è una vasta letteratura su questi metodi, qui ci soffermeremo su quegli approcci più comunemente usati in meccanica molecolare.

Punti di minimo

Un punto di minimo oltre ad avere le derivate prime rispetto a tutte le coordinate uguali a zero deve avere tutte le derivate seconde positive.

$$\frac{\partial V}{\partial r_i} = 0; \quad \frac{\partial^2 V}{\partial r_i^2} > 0$$

Se la derivata seconda è >0 allora è un punto di minimo




Ricerca dei punti di minimo

In termini matematici il problema è dunque quello di minimizzare una funzione -appunto l'energia- di più variabili -le coordinate interne degli atomi.

Qualora l'energia di un sistema di particelle possa essere rappresentata da una funzione analitica, il minimo può essere trovato con metodi di calcolo standard. Questo non è però generalmente possibile per sistemi molecolari complicati nei quali l'energia varia con le coordinate in maniera complessa. In questi casi i minimi sono localizzati usando metodi numerici nei quali le coordinate vengono variate gradualmente producendo configurazioni con energie sempre più basse fino al raggiungimento del minimo.

Scelta dell'algorithm.

Ci sono molti fattori che determinano la scelta dell'algorithm più appropriato per un dato problema:

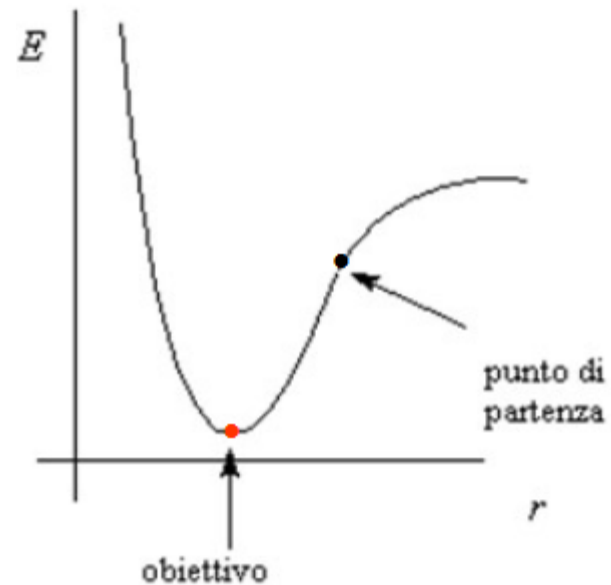
- veloce
- robusto 
- che richiede poca memoria

La **robustezza** indica la capacità di un metodo di non risentire degli effetti delle variazioni operative e della configurazione del sistema. Quindi può considerarsi come la capacità del metodo di non essere significativamente influenzato da variazioni delle condizioni analitiche.

Algoritmi di minimizzazione

Si possono classificare gli algoritmi di minimizzazione in due gruppi:

- 1) quelli che non usano le derivate dell'energia rispetto alle coordinate
- 2) quelli che usano le derivate dell'energia rispetto alle coordinate.



Criteri di Convergenza

A differenza di quello che succede con le funzioni analitiche nelle applicazioni "reali" di molecular modelling è possibile identificare l'esatta posizione dei minimi solo raramente, possiamo solo sperare di raggiungere una buona approssimazione. E' necessario quindi avere dei criteri per decidere quando il calcolo di minimizzazione è sufficientemente vicino ad un minimo in modo da poter essere terminato. Ogni calcolo è ovviamente limitato dalla precisione con cui vengono immagazzinati i numeri, ma nella maggior parte dei casi ci si ferma ben prima di questo limite.

Ci sono vari criteri tra cui scegliere (o applicare quello che si verifica per primo). Ci si ferma quando:

- si raggiunge un massimo numero di iterazioni
- quando la differenza di energia tra due passi successivi è minore di una certa tolleranza
- quando il passo (cioè la differenza tra due configurazioni successive) è piccola.

Scelta del punto di partenza

Dipende da quello che vogliamo ottenere:

Se vogliamo ottenere la struttura di minima energia (minimo globale) conviene partire da una struttura che probabilmente è già vicina a quel minimo:

- es. struttura a raggi X (dai database)

- se vogliamo utilizzare i metodi della meccanica quantistica (più pesanti) ci conviene partire da una struttura già minimizzata con i metodi della meccanica molecolare.

Se vogliamo la geometria di tutte le strutture a bassa energia (più minimi locali) allora possiamo partire da diverse strutture casuali ("random"). (

Applicazioni della minimizzazione

- per trovare il minimo globale
- per preparare il sistema per altre simulazioni:
 - rilassare le conformazioni ad energia più alta
 - avere una geometria di partenza buona
- ricerca conformazionale (conformational search)

Potenziale atomi legati

Potenziale atomi non legati

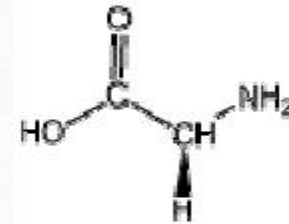
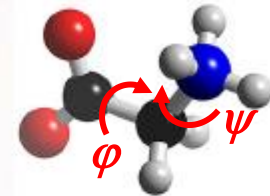
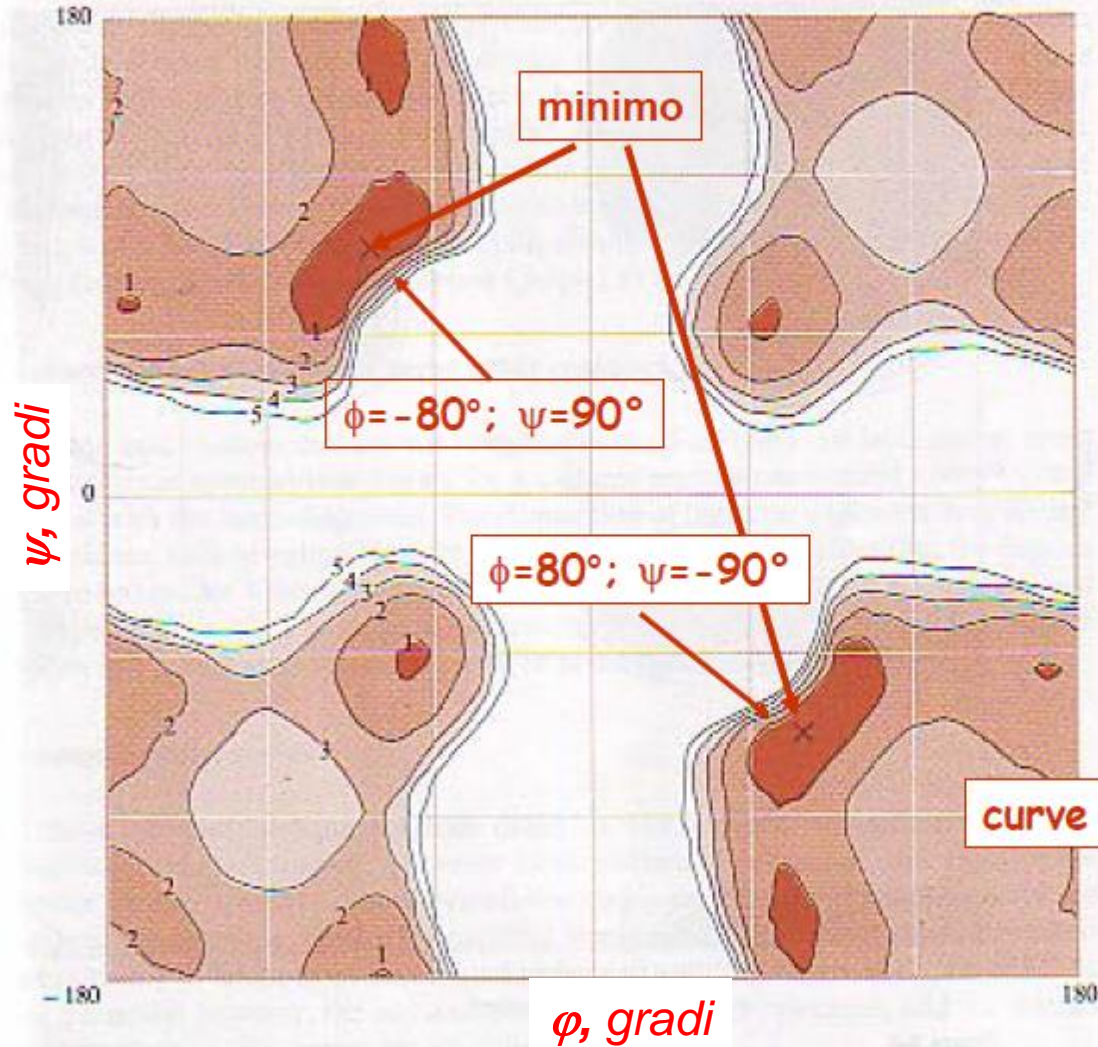
$$E = \sum_i^{\text{bonds}} k_i (r_i - r_{i,e})^2 + \sum_i^{\text{angles}} k_i (\theta_i - \theta_{i,e})^2 + \sum_i^{\text{some dimerals}} v_i [1 \pm \cos(n_i \phi_i)]^2 + \sum_i \sum_{j>1}^{\text{atomsatoms}} \left(-\frac{A_i A_j}{r_{ij}^6} + \frac{B_i B_j}{r_{ij}^{12}} + \frac{q_i q_j}{r_{ij}} \right)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

$$E_{\text{bond}} + E_{\text{ang}} = \text{costante}$$

Rigid geometry approximation

Mappe di Ramachandran

$$E(\varphi_i, \psi_i) = E_{LJ}(\varphi_i, \psi_i) + E_C(\varphi_i, \psi_i) + E_{tor}(\varphi_i, \psi_i)$$



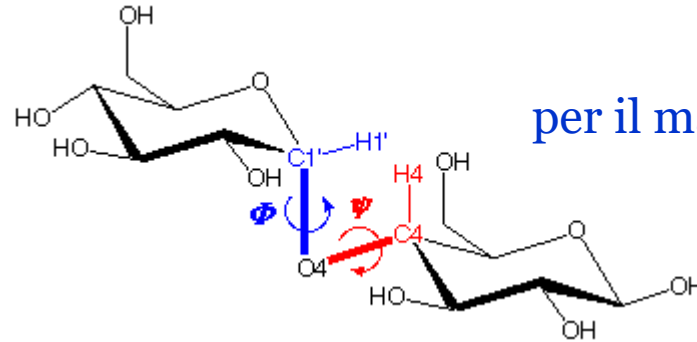
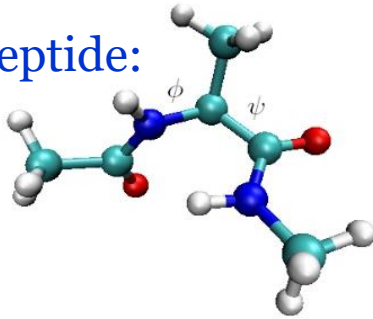
Glicina

curve di livello ogni 1 Kcal

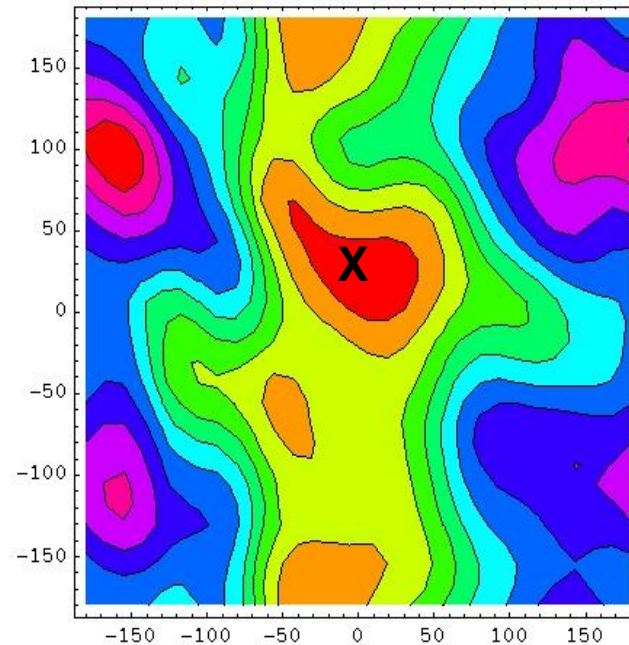
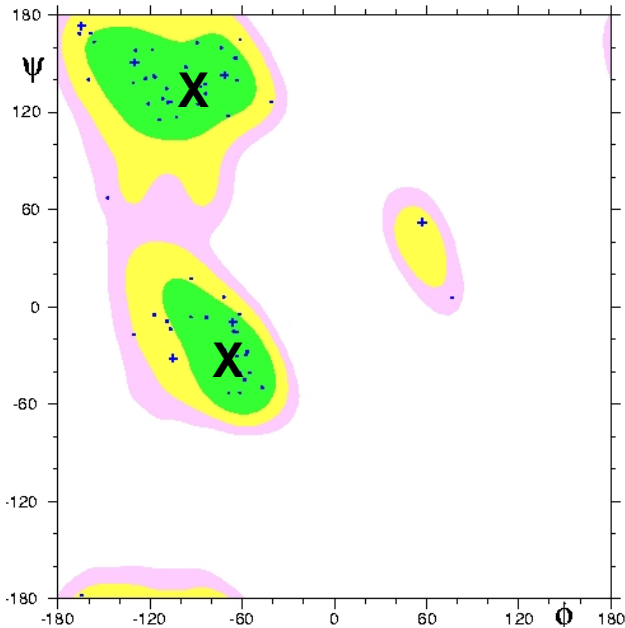
La libertà conformazionale è data dalle regione a bassa energia delle **mappe di Ramachandran**.

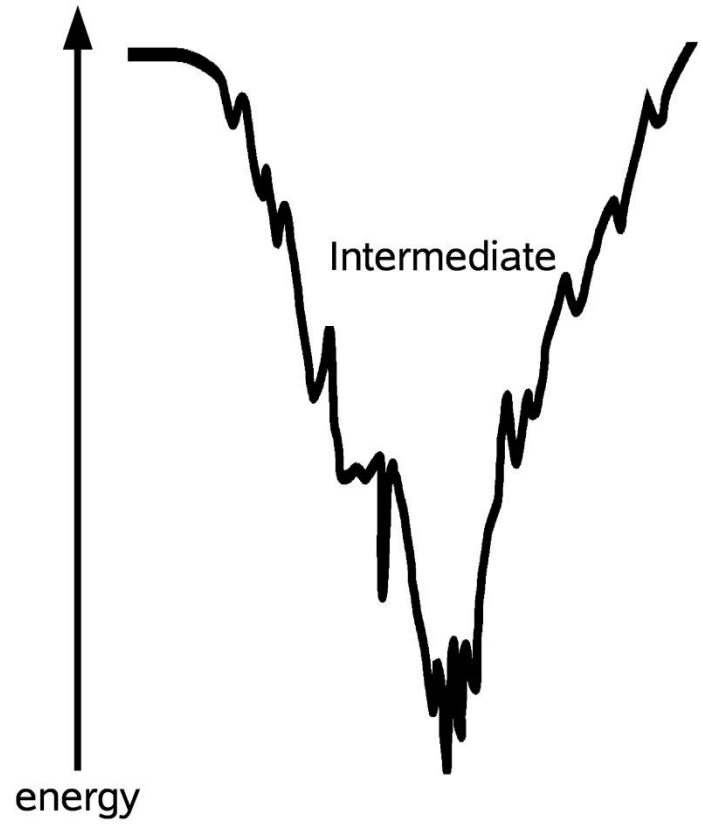
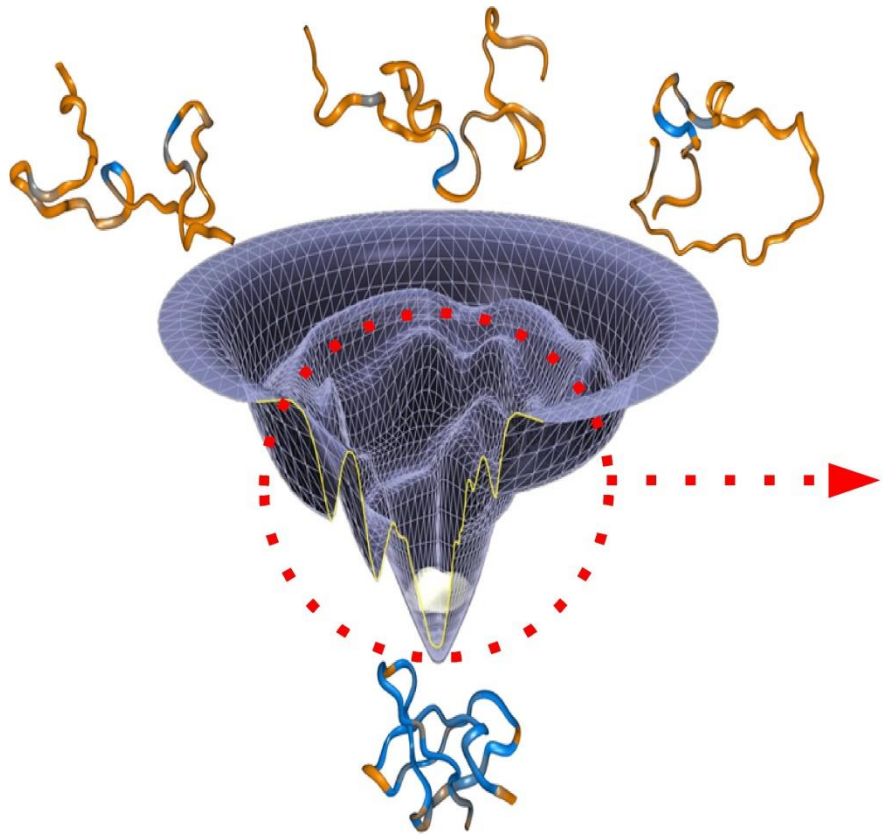
Esempio: le mappe relative a disaccaridi sono molto diverse da quelle relative a dipeptidi

per un dipeptide:



per il maltosio:





**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,...)

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

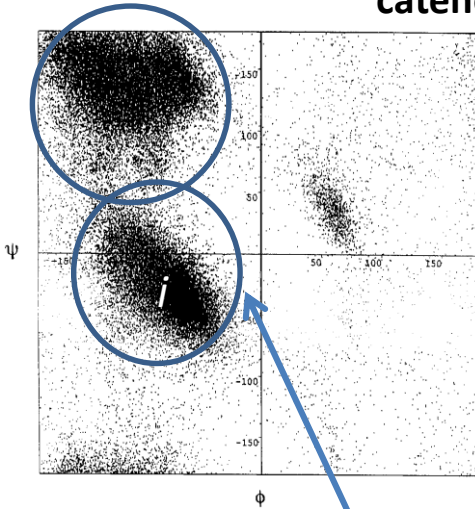
scelta dei legami descrittivi
della catena (reali o virtuali)

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

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

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

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

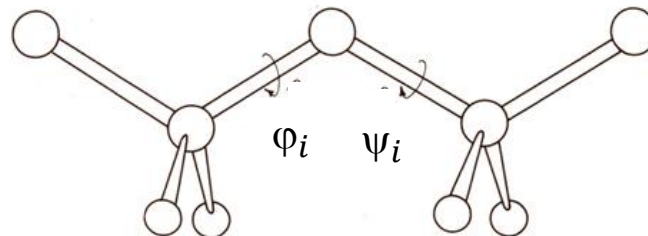
Energie degli stati conformazionali
 $E_i = E(\varphi_i, \psi_i)$

Popolazione di conformazioni accessibili
(più probabili !)

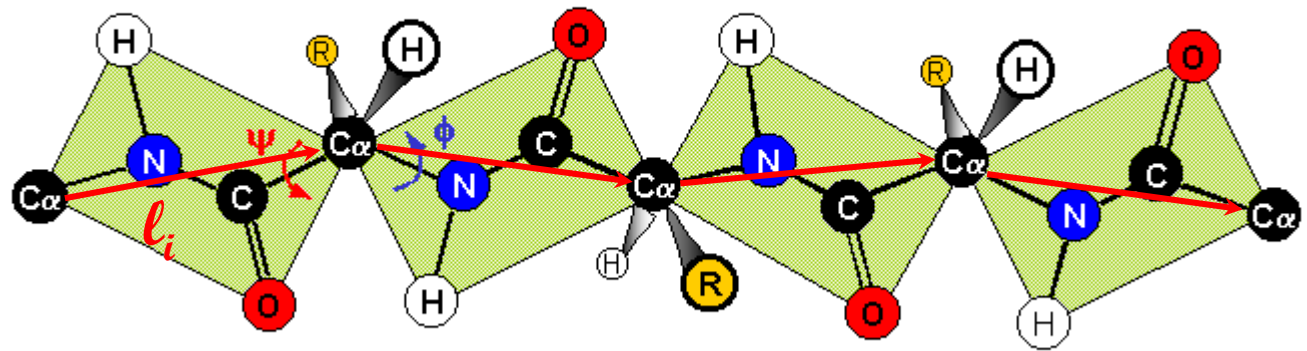
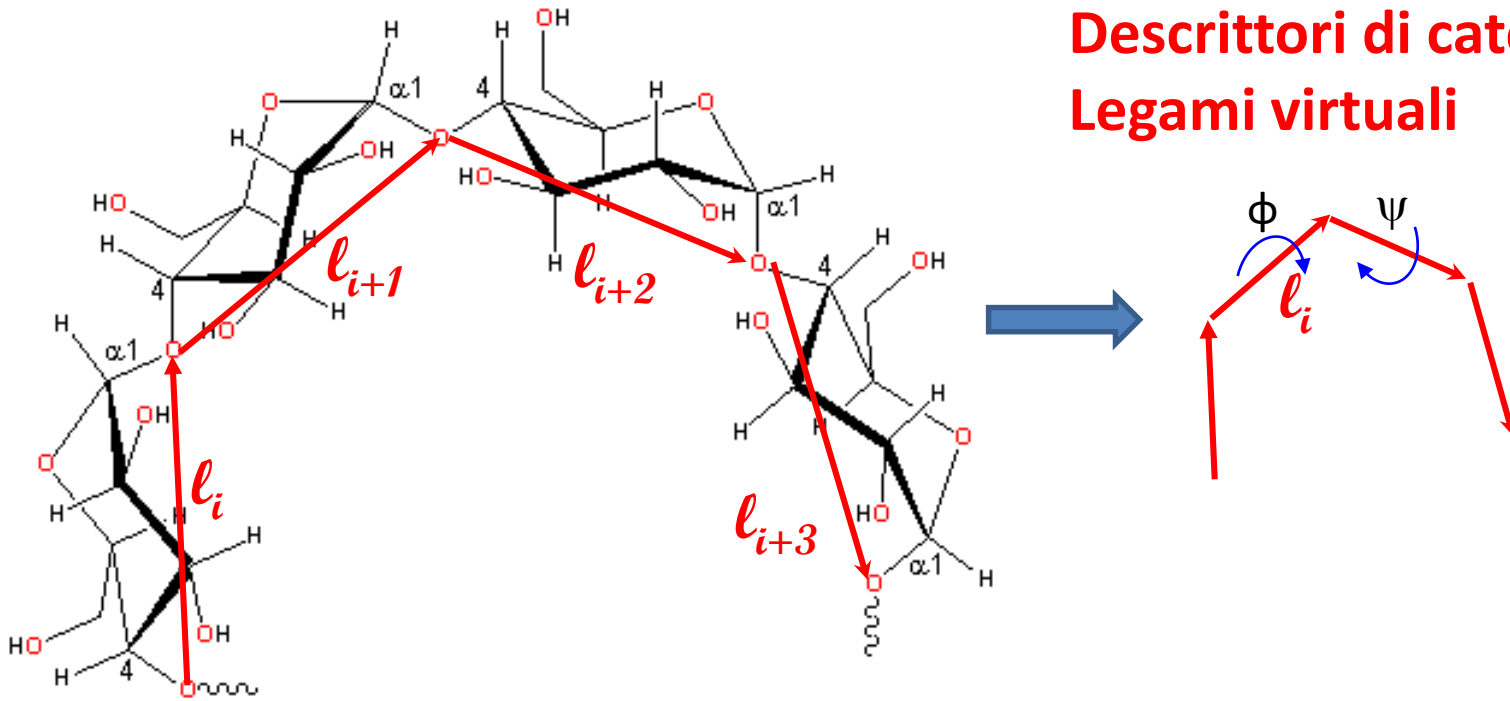
Stato solido amorfo,
Polimero fuso,
Polimero in soluzione
Random coil

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



**Dagli stati conformazionali $E(\varphi_i, \psi_i)$
alla predizione della topologia di catena**



**Metodi stocastici per la predizione
della topologia di macromolecole
nello stato disordinato**

I **modelli deterministici** sono i più semplici; in essi, le variabili di **input assumono valori fissi**.

E' vero che i risultati (*output*) generati da questi modelli possono tener conto, entro certi limiti, della *variabilità* e dell'effetto del caso; tuttavia, si tratta sempre di elaborazioni di tipo deterministico, in quanto non si tiene in considerazione l'incertezza associata alle variabili di *input*.

i **modelli stocastici** (stocastico = dovuto al caso, aleatorio, dal greco *stochastikòs*=congetturale) **tengono in considerazione le variazioni (causali e non) delle variabili di input**, e quindi forniscono risultati in termini di "**probabilità**".

È importante sottolineare che ciò che differenzia i modelli deterministici da quelli stocastici è che in questi ultimi si tiene conto della variabilità dei dati di input.

In genere i modelli stocastici hanno una struttura più complessa di quelli deterministici.

I metodi stocastici hanno trovato larga diffusione in vari settori, (scientifici e non):

Fisica (metodi quantistici, reattori nucleari, aerodinamica, ottica...),

Astrofisica (modelli di evoluzione di galassie),

Biologia (sistemi biologici),

Finanza (calcolo del valore di una compagnia, per prezzare i derivati e le opzioni non standard),

Ingegneria (variazioni in circuiti integrati, fluidodinamica,...),
altri.....

altri settori:

Traffico stradale,

Videogiochi, metodi scacchistici, ...

Meteorologia,

Progettazione grafica, ecc...

Disordine :

- Disordine nelle velocità'
- Disordine nelle posizioni:
 - Liquidi
 - Polimeri
 - Solidi disordinati (vetri, amorfi)

la distribuzione di Maxwell-Boltzmann

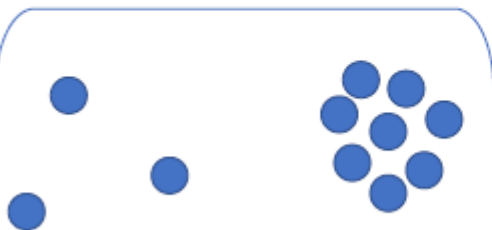
$$f_{MB}(v) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[-\frac{mv^2}{2kT}\right]$$



f_{MB} predetta in sistemi in equilibrio classico, potenziali dipendenti dalla sola posizione e velocità non relativistiche. Confermata in:

Fluidi (alta mobilità')

Solidi (mobilità' trascurabile)



Gas ideali

Liquidi

Cristalli

Vetri

NB:

Le leggi dei sistemi disordinati sono di natura statistica

Liquidi: il moto browniano e l'esistenza degli atomi



J. Perrin

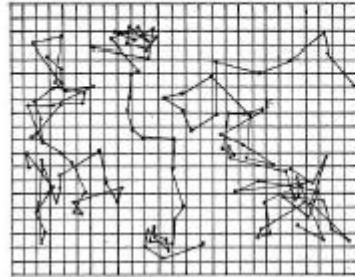


Fig. 7.

Jean Baptiste Perrin, The Nobel Prize in Physics 1926

"for his work on the **discontinuous structure of matter...**"



A. Einstein

Articoli 1905 (Annus Mirabilis), in ordine di invio:

- Effetto fotoelettrico (Nobel 1921)
- **Moto Browniano**
- Relatività speciale
- $E=mc^2$

$$\langle \mathbf{R}^2(t) \rangle = 6 D t$$

$$k_B = 6\pi a \frac{D \eta}{T}$$

D: coefficiente di diffusione
 η : viscosità

Legge di Stokes-Einstein

*Esistono diversi **modelli statistici** che vengono usati nella descrizione dei processi stocastici relativi a sistemi polimerici*

- La teoria dei processi stocastici riguarda lo studio di sistemi che evolvono nel tempo (ma anche più in generale nello spazio) **secondo leggi probabilistiche.**

- Tali sistemi o modelli descrivono **fenomeni complessi** del mondo reale che hanno la possibilità di essere **aleatori.**

- Tali fenomeni sono più frequenti di quanto si possa credere e si incontrano tutte quelle volte in cui le quantità alle quali siamo interessati **non sono prevedibili con certezza.**

Polimeri: analogia con il moto browniano

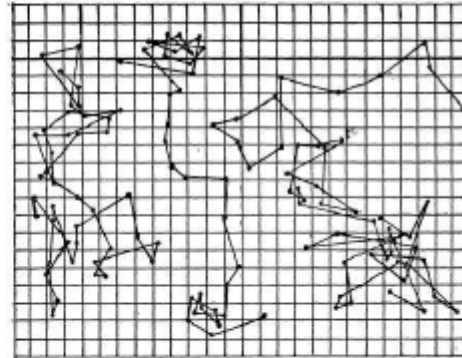
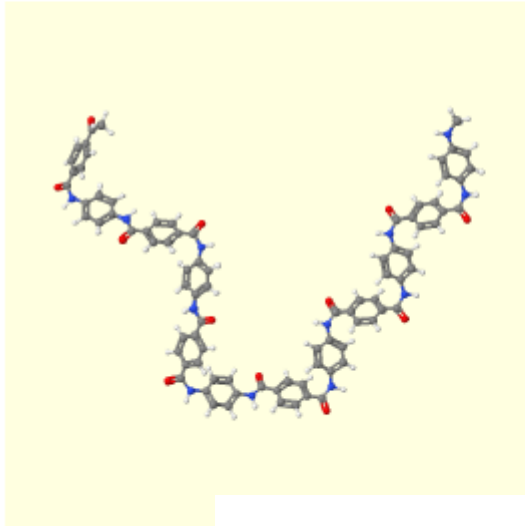
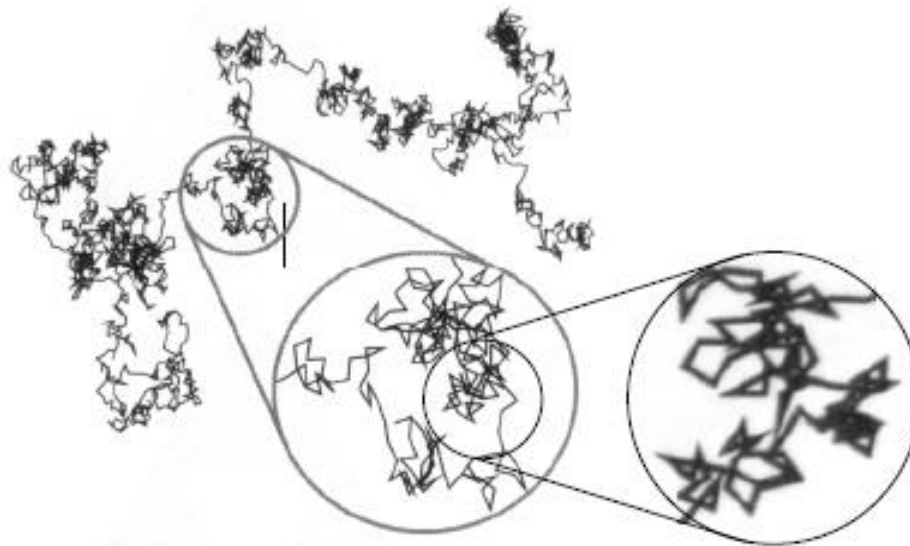


Fig. 7.



Alcuni esempi di modelli statistici sono:

- 1) **catena Bernoulliana:** in questo modello la probabilità di avere una determinata conformazione in catena è indipendente dallo stato conformazionale precedente e tale evento non ha alcun effetto sulla conformazione del residuo successivo in catena;
- 2) **catena Markoviana:** in una catena Markoviana di ordine n la probabilità di avere uno stato conformazionale ad un certo DP in catena è dipendente dagli n stati conformazionali che lo precedono. Quando $n=0$ la catena Markoviana coincide con quella Bernoulliana;
- 3) **catena di Ising:** la probabilità di avere un dato stato ad un certo punto della catena dipende dallo stato conformazionale dei residui primi vicini.

Metodo Monte Carlo

(Nicholas Constantine Metropolis)

Sono metodi probabilistici utilizzati per ottenere stime di grandezze d'interesse attraverso estrazioni di numeri casuali.

Il metodo è usato per trarre stime attraverso simulazioni.



Con metodo (di) Monte Carlo si intende una classe di algoritmi computazionali che, a partire da un insieme di **numeri generati casualmente** o pseudo-casualmente, permetta di ottenere un risultato medio molto vicino al comportamento teorico.

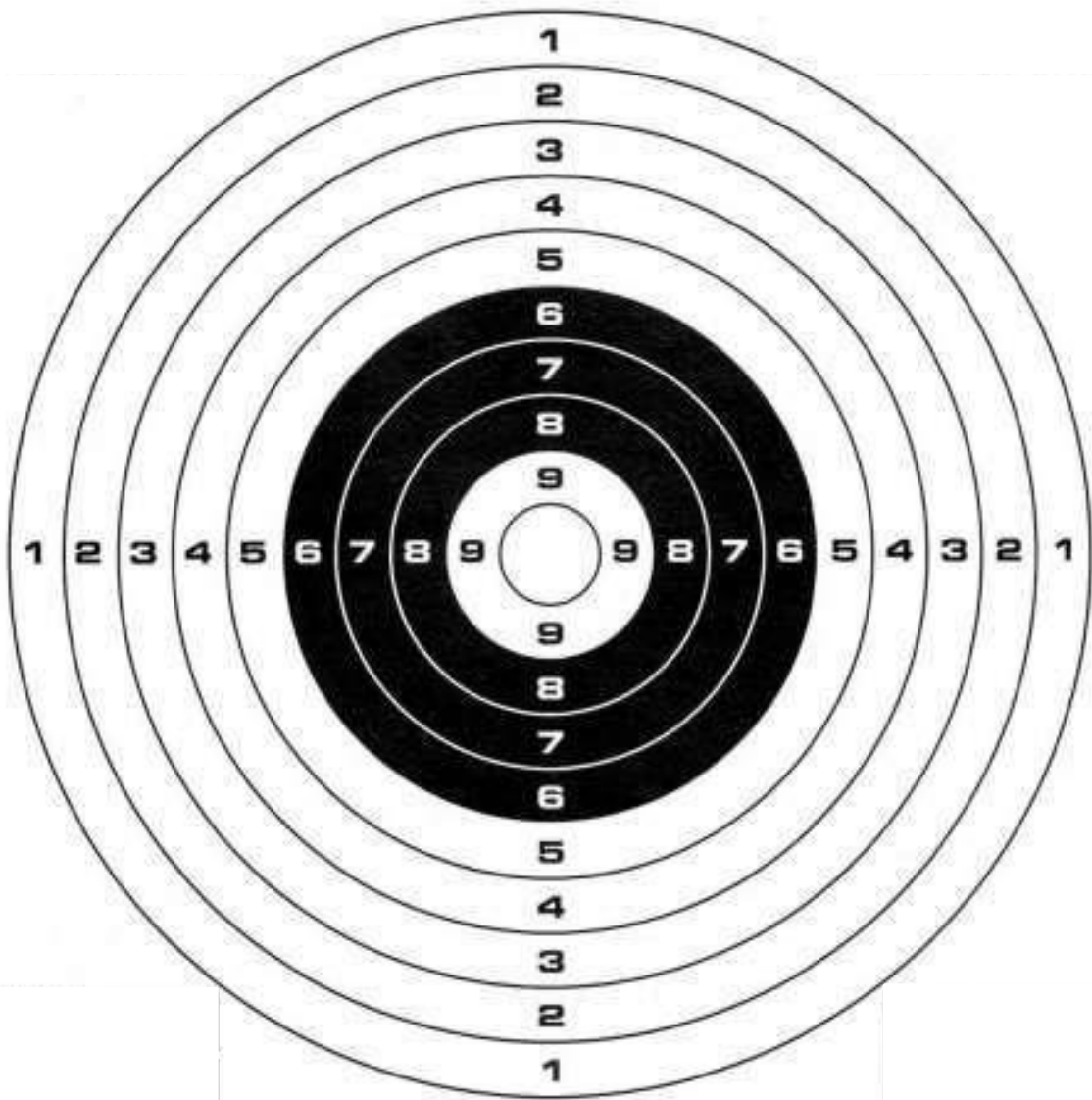
Prevede dunque la risoluzione di complicati problemi deterministici con metodi probabilistici.

Prende particolarmente piede, in campo scientifico, laddove l'osservazione diretta del fenomeno non sia più sufficiente per arrivare alla formulazione di un modello teorico (per es. in fisica subatomica).

Con il metodo Monte Carlo (MC) **si simula un processo stocastico associando agli stadi che si susseguono nel processo, una probabilità *a priori*.**

Per fare questo vengono generati dei **numeri casuali (random)** e confrontati con le **probabilità assegnate *a priori*** ai vari possibili eventi, in modo da determinare quale di questi avviene a quel particolare stadio del processo.

La simulazione Monte Carlo calcola una **serie di realizzazioni possibili** del fenomeno in esame, con il **peso** della probabilità di tale evenienza, cercando di esplorare in modo «denso» tutto lo spazio dei parametri del fenomeno



In generale viene generata una serie di **N** numeri random che determina un campione di popolazione di tutte le possibili conformazioni della catena ognuna **pesata** mediante l'opportuno **fattore probabilistico di Boltzmann** chiamato **peso statistico** :

$$q = \sum_{k=1}^M \exp\left[-\frac{E_k}{kT}\right]$$

Funzione di partizione

$$p_j = \frac{e^{-\frac{E_j}{kT}}}{q}$$

Probabilità di evento dello stato j-esimo

Generatori Lineari Congruenziali (LCG)

Uno dei metodi matematici più noti per generare numeri casuali distribuiti uniformemente è rappresentato dai generatori lineari **congruenziali** (**numeri pseudo-casuali**)

Tali generatori vennero presentati per la prima volta da D.H. Lemer negli anni '40.

Il metodo LCG, come tutti i generatori di numeri casuali, ha bisogno di un **seed** (seme) per inizializzare la sequenza di numeri secondo la regola

$$x_{n+1} = (a x_n + c) \bmod m, \quad n \geq 0,$$

dove **a**, **c** ed **m** sono opportuni numeri interi costanti. La notazione “mod *m*”, ossia **modulo m**, significa che ax_n+c viene diviso per **m** e x_{n+1} **posto uguale al resto intero della divisione**. Quindi x_{n+1} assume valori interi tra 0, 1, 2, ..., m-1.

Algoritmo di generazione di numeri random

$$x_i = (N_1 x_{i-1} + N_2) \bmod [(N_1 x_{i-1} + N_2), N]$$

$$\bmod [(N_1 x_{i-1} + N_2), N] = (N_1 x_{i-1} + N_2) - N \cdot \text{int} \left[\frac{(N_1 x_{i-1} + N_2)}{N} \right]$$

$$0 \leq N_1, N_2 < N$$

dove N_1 , N_2 ed N sono numeri interi ed $\text{int}[\dots]$ è una funzione che prende la parte intera del rapporto espresso in parentesi.

ESEMPIO

Se come esempio consideriamo $x_0=4$, $N_1=5$, $N_2=3$ ed $N=8$, si ottiene la seguente serie di numeri:

$\{4, 7, 6, 1, 0, 3, 2, 5, 4, 7, 6, 1, \dots\}$

che si ripete ciclicamente dopo 8 numeri. La periodicità (numeri pseudo-random) è in genere uguale ad N , cosicchè prendendo N molto grande, la lunghezza del ciclo diventa più lunga della dimensione della serie di numeri random che usualmente si richiede nella maggioranza delle simulazioni Monte Carlo.

In effetti nessun calcolatore è in grado di generare **numeri puramente casuali**, ma solo numeri *pseudo-casuali*. Si tratta di numeri generati da algoritmi deterministici in grado di superare una serie di test statistici che conferiscono a tali numeri una **apparente casualità**.

Il problema della scelta dei migliori valori per a , c ed m è quindi il punto cruciale del metodo.

Un aspetto importante è la **lunghezza del periodo** della successione da cui segue che m dovrà essere molto grande.

Inoltre per un dato m , i valori di a e c dovranno essere tali che la successione abbia periodo massimo.

Una delle scelte più *popolari* è

$$\begin{aligned} m &= 2^{31}-1, \\ a &= 75, \\ c &= 0. \end{aligned}$$

Questo garantisce **un periodo di oltre due miliardi di numeri casuali.**

Metodo middle-square

- Supponiamo di voler generare un numero casuale di 4 cifre
- Il metodo richiede come tutti i generatori di numeri casuali un valore iniziale, detto **seme**, dal quale vengono generati i successivi valori
- Ad es. a partire da 1234 avente 4 cifre
 - eleviamo al quadrato e otteniamo le 8 cifre 01522756
 - consideriamo solo le 4 cifre di mezzo 5227
 - ripetiamo il procedimento ottenendo 27321529 e 3215 e così via
- Ogni nuovo numero è determinato univocamente dal predecessore. Ogni successione di numeri generata da questo algoritmo si ripeterà prima o poi. Il numero di numeri della sequenza prima che intervenga una ripetizione è detta **periodo** della sequenza

N.B. Suggesto da John von Neumann nei primi anni dell'era dei computer (1946) per generare numeri pseudo-casuali distribuiti in modo uniforme.

Generatori più recenti: il Cristoforo Colombo

I generatori di numeri casuali più recenti, come quello incluso nelle versioni di MATLAB sono una **combinazione di operazioni di spostamento di registri** e manipolazione sui bit che non richiedono nessuna operazione di moltiplicazione o divisione.

Questo nuovo approccio risulta estremamente veloce ed inoltre garantisce periodi di sequenza incredibilmente lunghi.

Nelle ultime versioni MATLAB, **il periodo è 2^{1492}** , che ad un milione di numeri casuali al secondo richiederebbe 10^{435} anni (circa 10^{425} volte l'età dell'universo) prima di ripetersi!

Data la coincidenza dell'esponente con la data della scoperta dell'America (12 ott. 1492), questo generatore viene comunemente chiamato **il generatore di "Cristoforo Colombo"**.

Controlli statistici

REQUISITI DEI GENERATORI DI NUMERI RANDOM

1) uniformità: gli n numeri random devono essere uniformemente distribuiti nell'intervallo $[0,1]$ e tale proprietà viene stimata attraverso i momenti della distribuzione, ossia:

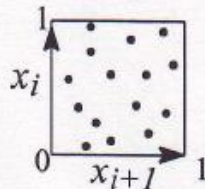
$$m_1 = \frac{1}{n} \sum_i^n x_i = 0.500$$

$$m_2 = \frac{1}{n} \sum_i^n x_i^2 = 0.333$$

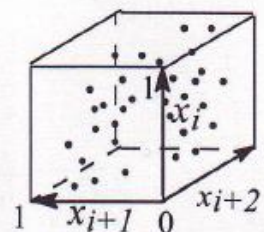
$$m_3 = \frac{1}{n} \sum_i^n x_i^3 = 0.250$$

2) assenza di correlazione: un test molto semplice di correlazione dei numeri della serie è quello di analizzare copie (x_i, x_{i+1}) e terne (x_i, x_{i+1}, x_{i+2}) nell'insieme di numeri random e vedere con quale frequenza si distribuiscono, rispettivamente, nel piano e nello spazio. Per un insieme non-correlato i punti devono risultare uniformemente distribuiti.

(x_i, x_{i+1})



(x_i, x_{i+1}, x_{i+2})



Considerazioni

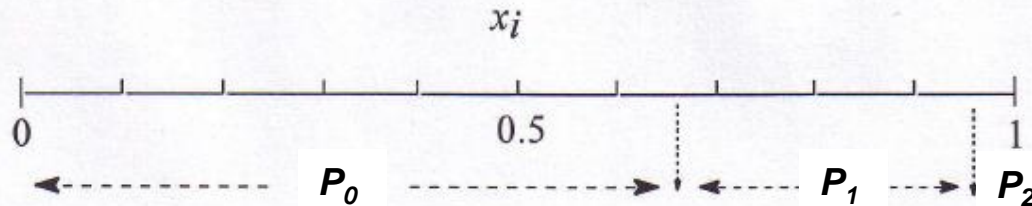
- Le caratteristiche che deve avere un generatore di sequenze pseudocasuali:
 - Ripetibilità, affinché sia possibile ripetere gli stessi esperimenti più volte
 - Soddisfacimento dei test: servono per verificare che il generatore sia quanto più simile ad un generatore di numeri casuali ideale.
 - Semplicità e rapidità: un simulatore che usa il generatore risulta quindi efficiente.
 - Periodicità lunga, in modo da poter effettuare simulazioni lunghe senza riutilizzare più volte la stessa sequenza.
 - Portabilità, cioè rendere l'implementazione del generatore indipendente dalla piattaforma.

ESEMPIO: Generazione di N numeri random x_i

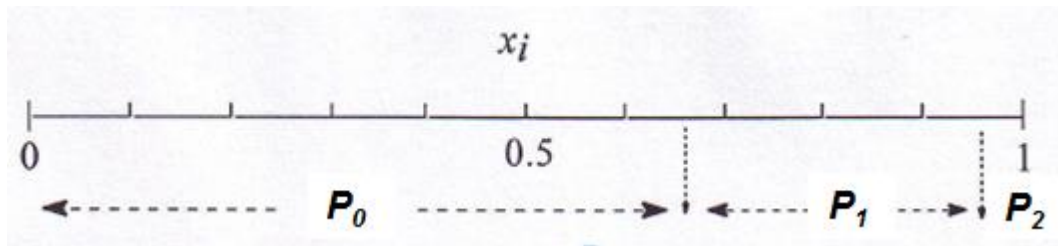
Stati energetici	$E_0=0$	➔	Peso statistico	$P_0=0.66$
	$E_1=0.5$		$P_1=0.28$	
	$E_2=1.5$		$P_2=0.06$	

- la molecola adotterà la conformazione $E_0 = 0$ se $0 \leq x_i \leq P_0$;
- la molecola adotterà la conformazione $E_1 = 0.5$ se $P_0 \leq x_i \leq P_0 + P_1$;
- la molecola adotterà la conformazione $E_2 = 1.5$ se $P_0 + P_1 \leq x_i \leq P_0 + P_1 + P_2 = 1$

E' chiaro che i numeri random cadranno con maggiore frequenza in quella parte del segmento $[0, 1]$ più ampia.



Per N=20



n_i	x_i	conform.
1	.00002	E_0
2	.08503	E_0
3	.60135	E_0
4	.89161	E_1
5	.96796	E_2
6	.18969	E_0
7	.51498	E_0
8	.39801	E_0
9	.26291	E_0
10	.74351	E_1
11	.08955	E_0
12	.56039	E_0
13	.58223	E_0
14	.80957	E_1
15	.59192	E_0
16	.51171	E_0
17	.87663	E_1
18	.99508	E_2
19	.72621	E_1
20	.96661	E_2

$$\mathbf{R_g, \langle r \rangle^{1/2}}$$



- **Diffusione della luce**
(Laser Light Scattering)
- **Viscometria**