

## Chapter 3

# Statistica di catene molecolari

Consideriamo qui l'analisi della statistica delle configurazioni di una catena molecolare libera, dove la valutazione statistica delle proprietà della catena viene fatta mediando su tutte le conformazioni. Le proprietà così valutate sono il riflesso della geometria della catena (angoli e lunghezze di legame) e delle limitazioni delle rotazioni del *backbone*. La proprietà medie di catena libera, che spesso denomineremo *coil*, così ottenute, devono essere nettamente distinte da quelle dello stato nativo in cui molte molecole biologiche vengono normalmente trovate.

Nel trattare la statistica delle catene porremo l'attenzione principale nel calcolare le distanze medie intra-catena, cioè le distanze medie tra elementi della catena. Le informazioni così ottenute possono essere utili sia per il protein folding che per altre proprietà delle macromolecole. In particolare, relativamente al folding, tali modelli risulteranno adatti a descrivere le proprietà statistiche di una proteina nello stato *unfolded*. D'altra parte, c'è una stretta connessione tra la distanza end-to-end e le proprietà idrodinamiche e altre proprietà fisiche di un coil, come la viscosità intrinseca ed i coefficienti di sedimentazione. In effetti, è proprio da queste proprietà che in genere si può cominciare a capire se la catena può essere trattata come un semplice coil, o invece si debbono introdurre altri tipi di modelli.

### 3.0.3 Calcolo delle proprietà dipendenti dalle conformazioni

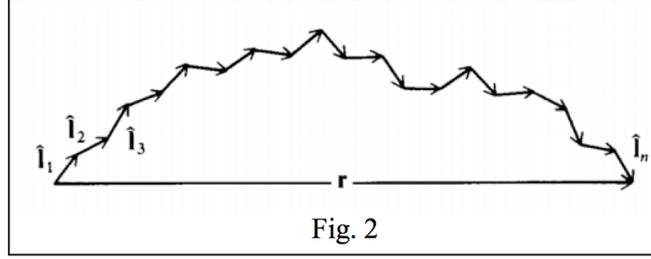
Nella figura illustriamo schematicamente la conformazione di una ipotetica catena polimerica, in cui ogni *pallina* indica un atomo o un gruppo di atomi del suo *backbone*. I bonds sono indicati con gli indici da 1 a  $n$  (il bond  $i$  unisce gli elementi  $i - 1$  e  $i$  della catena). Gli angoli *diedrici* sono quindi numerate da  $\phi_2$  a  $\phi_{n-1}$ , e quindi gli angoli 1 e  $n$  (perdendo di significato nel caso presente) non verranno definiti.

Nel trattare le conformazioni di catene reali, assumeremo che sia le lunghezze che gli angoli dei bonds siano fisse. Ovviamente questa è un'approssimazione, ma dato che in genere le deviazioni (per esempio quelle tra gli angoli) tendono a compensarsi, essa non genera grandi inaccuratezze. Perciò le conformazioni della catena saranno determinate sulla base dei valori degli angoli di rotazione  $\phi_i$ .

### 3.0.4 Distanza end-to-end e raggio di girazione

La distanza end-to-end  $r$  può essere calcolata come la somma dei vettori  $\hat{\mathbf{I}}_i$  associati ai bonds, dove intenderemo che il vettore  $\hat{\mathbf{I}}_i$  è parallelo al bond  $i$  e punta dall'elemento  $i - 1$  all'elemento  $i$ :

$$\mathbf{r} = \sum_{i=1}^n \hat{\mathbf{I}}_i. \quad (3.1)$$



Il modulo quadro  $r^2$  di  $\mathbf{r}$  è dato da  $r^2 = (\mathbf{r} \cdot \mathbf{r})$ , pari a

$$r^2 = \left( \sum_{i=1}^n \hat{\mathbf{I}}_i \right) \cdot \left( \sum_{j=1}^n \hat{\mathbf{I}}_j \right) = \sum_{i=1}^n \sum_{j=1}^n \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j. \quad (3.2)$$

Portando fuori dalla sommatoria gli  $n$  termini diagonali ( $i = j$ ), e assumendo, per semplicità, che tutti i bonds abbiano la stessa lunghezza  $l$ , abbiamo

$$r^2 = nl^2 + 2 \sum_{j>i}^n \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j. \quad (3.3)$$

Il valor medio di tale quantità è

$$\langle r^2 \rangle = nl^2 + 2 \sum_{j>i}^n \langle \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j \rangle \quad (3.4)$$

dove le parentesi angolari denotano i valori medi. Quindi, il problema del calcolo del valor medio del quadrato della distanza end-to-end, si riduce al calcolo della proiezione media che ogni bond ha sugli altri.

Tale media sarà calcolata attraverso le leggi della meccanica statistica, le quali affermano che la probabilità di una particolare conformazione è  $p(\phi_2, \phi_3, \dots, \phi_{n-1})d\phi_2d\phi_3\dots d\phi_{n-1}$  (conformazione determinata quando gli angoli cadono negli intervalli  $[\phi_2, \phi_2 + d\phi_2]$ ,  $[\phi_3, \phi_3 + d\phi_3]$ , etc...) è determinata dalla distribuzione di Boltzmann. Ciò è

$$p(\phi_2, \phi_3, \dots, \phi_{n-1})d\phi_2d\phi_3\dots d\phi_{n-1} = \frac{1}{P} \exp[-E(\phi_2, \phi_3, \dots, \phi_{n-1})/k_B T] \quad (3.5)$$

dove il fattore di normalizzazione  $P$  è dato da

$$P = \int_0^{2\pi} \dots \int_0^{2\pi} \exp[-E(\phi_2, \phi_3, \dots, \phi_{n-1})/k_B T] d\phi_2d\phi_3\dots d\phi_{n-1}. \quad (3.6)$$

dove  $E(\phi_2, \phi_3, \dots, \phi_{n-1})$  è l'energia associata alla conformazione specificata da  $\phi_2, \phi_3, \dots, \phi_{n-1}$ . Tuttavia, nei casi che vedremo, le distribuzioni di probabilità che incontreremo saranno per lo più determinate dagli eventuali vincoli associati alle rotazioni dei legami; in generale, invece, essi possono includere altri fattori

che tengono conto delle interazioni con il solvente e/o con gli altri elementi (residui) della catena.

Quindi, la valutazione dei prodotti  $\hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j$  dipende dai valori permessi agli angoli di rotazione tra i bonds  $i$  e  $j$ ; cioè l'orientazione di  $\hat{\mathbf{I}}_j$  rispetto a quella di  $\hat{\mathbf{I}}_i$  dipende da  $\phi_{j-1}, \phi_{j-2}, \dots, \phi_{i+1}$ . Quindi, la somma (per  $j > i$ )  $\sum_{j>i} (\hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j)$  dipende dall'intero set di angoli di rotazione, e il suo valor medio è

$$\sum_{j>i} \langle \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j \rangle = \int_0^{2\pi} \dots \int_0^{2\pi} \sum_{j>i} (\hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j) p(\phi_2, \phi_3, \dots, \phi_{n-1}) d\phi_2 d\phi_3 \dots d\phi_{n-1}. \quad (3.7)$$

Sebbene l'eq.(3.7) possa apparire molto complessa, le situazioni che analizzeremo ci permetteranno di renderla molto semplice. La sostituzione dell'eq.(3.7) nell'eq.(3.4) equazione nella precedente ci permette di ottenere  $\langle r^2 \rangle$ .

La relazione tra  $\langle r^2 \rangle$  e le conformazioni della catena è insita nelle equazioni (3.5) e (3.7) precedenti. Se  $E(\phi_2, \phi_3, \dots, \phi_{n-1})$  ha molti dei suoi minimi di energia per configurazioni che corrispondono a valori grandi di  $r^2$ , il valore risultante di  $\langle r^2 \rangle$  sarà molto grande. Un unico minimo di energia molto stretto a  $\phi'_2, \phi'_3, \dots, \phi'_{n-1}$  risulterà invece in un valore di  $\langle r^2 \rangle$  molto vicino a quello corrispondente alla conformazione specificata da  $\phi'_2, \phi'_3, \dots, \phi'_{n-1}$ . Infine, nel limite di  $E(\phi_2, \phi_3, \dots, \phi_{n-1})$  costante per tutti i valori di  $\phi_2, \phi_3, \dots, \phi_{n-1}$  (*freely rotating chain*),  $\langle r^2 \rangle$  corrisponderà alla semplice media aritmetica di  $r^2$  su tutte le conformazioni della catena. È chiaro, quindi, che  $\langle r^2 \rangle$  è il riflesso della distribuzione delle conformazioni e conseguentemente di quelle prevalentemente adottate dalla catena.

## Il raggio di girazione

Il raggio di girazione  $R_G$  è definito come la distanza quadratica media di un gruppo di atomi dal loro centro di massa. Per una catena di  $n + 1$  elementi identici (uniti da  $n$  bonds, vedi Figura), se  $R_{G,i}$  è la distanza dell'elemento  $i$  dal centro di massa della catena per una particolare conformazione, allora

$$R_G^2 = \frac{1}{n+1} \sum_{i=0}^n R_{G,i}^2. \quad (3.8)$$

Tuttavia, è possibile mostrare che

$$R_G^2 = \frac{1}{(n+1)^2} \sum_{j>i}^n r_{i,j}^2. \quad (3.9)$$

dove  $r_{i,j}^2$  è la distanza fra gli elementi  $i$  e  $j$ . Perciò il calcolo del raggio di girazione si riduce al calcolo delle distanze intra-catena, come già visto per la end-to-end distance, Eq.(3.4).

**Dimostrazione** Si consideri un oggetto composto da  $n + 1$  elementi. L' $i$ -esimo elemento sia posizionato nel punto  $(x_i, y_i, z_i)$  e abbia massa  $m_i$ . Il centro

di massa dell'oggetto è un punto che ha la seguente proprietà: se  $\mathbf{s}_i$  è il vettore che va dal centro di massa all'elemento  $i$ , allora

$$\sum_{i=0}^n m_i \mathbf{s}_i = 0. \quad (3.10)$$

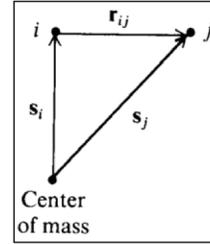
Se l'oggetto è composto da elemento di massa identica, allora la precedente si semplifica:

$$\sum_{i=0}^n \mathbf{s}_i = 0. \quad (3.11)$$

Dato il centro di massa, il raggio di girazione può essere calcolato attraverso la (3.8). Spesso, però, è conveniente utilizzare un'espressione del raggio di girazione dipendente solo dalle coordinate dei singoli elementi componenti l'oggetto, analoga alla (3.9). Dimostriamo l'equivalenza tra queste due equazioni. Prima di tutto notiamo che

$$\sum_{j>i}^n r_{ij}^2 = \frac{1}{2} \sum_{j=0}^n \sum_{i=0}^n \mathbf{r}_{ij} \cdot \mathbf{r}_{ij}. \quad (3.12)$$

Consideriamo ogni coppia di elementi dell'oggetto  $i$  e  $j$  come mostrato in figura. Se  $\mathbf{s}_i$  e  $\mathbf{s}_j$  sono i vettori che vanno dal centro di massa a tali elementi, allora  $\mathbf{r}_{ij} = \mathbf{s}_j - \mathbf{s}_i$ . Quindi, la doppia somma può essere scritta nella forma:



$$\frac{1}{2} \sum_{j=0}^n \sum_{i=0}^n (\mathbf{s}_j - \mathbf{s}_i) \cdot (\mathbf{s}_j - \mathbf{s}_i) = \frac{1}{2} \sum_{j=0}^n \sum_{i=0}^n (\mathbf{s}_i^2 + \mathbf{s}_j^2 - 2\mathbf{s}_i \cdot \mathbf{s}_j). \quad (3.13)$$

Ogni termine può essere sommato facilmente. Dall'eq. (3.8), sappiamo che  $\sum_i \mathbf{s}_i^2 = (n+1)R_G^2$  mentre l'effetto della somma su  $j$  è semplicemente quello di moltiplicare tale termine per  $n+1$ . Analogamente si ha per la somma di  $\mathbf{s}_j^2$ . Per il terzo termine, possiamo scrivere

$$\sum_{j=0}^n \sum_{i=0}^n \mathbf{s}_i \cdot \mathbf{s}_j = \left( \sum_{j=0}^n \mathbf{s}_j \right) \cdot \left( \sum_{i=0}^n \mathbf{s}_i \right) = 0. \quad (3.14)$$

Perciò, alla fine possiamo scrivere

$$\sum_{j>i}^n r_{ij}^2 = (n+1)^2 R_G^2. \quad (3.15)$$

### 3.1 Modelli di catene senza volume escluso

Nella seguente trattazione della statistica delle conformazioni non viene tenuto conto degli effetti del volume escluso. Il volume escluso è semplicemente il

vincolo fisico legato al fatto che due elementi della catena non possono simultaneamente occupare la stessa posizione. Nella trattazione di una catena reale, le conformazioni che includono le interazioni di volume escluso tra elementi di catena vicini nella sequenza sono implicitamente considerate, assegnando ad esse valori molto grandi dell'energia. Tuttavia, le possibili intersezioni tra elementi lontani nella sequenza sono molto difficili da tenere in conto. Quindi, la presente statistica delle conformazioni si applica a catene non perturbate da effetti di volume escluso.

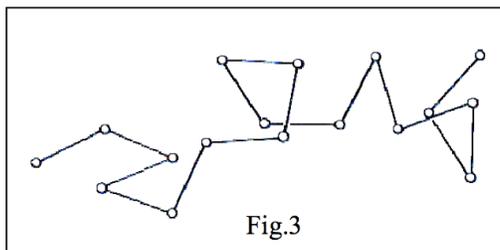
L'eq. (3.4) può essere usata, ovviamente, per calcolare ogni distanza intracatena  $r_{ij}$ . Osservando l'eq. (3.9), può non sorprendere, che  $\langle R_G^2 \rangle$  e  $\langle r^2 \rangle$  siano, sotto certe condizioni, direttamente correlati. In particolare nel limite  $n \rightarrow \infty$  si trova

$$\langle R_G^2 \rangle = \frac{1}{6} \langle r^2 \rangle \quad \text{per } n \rightarrow \infty \quad (3.16)$$

(Flory, 1969). Quindi nel limite di catene infinitamente lunghe, la distanza end-to-end quadratica media e il raggio di girazione quadratico medio sono direttamente proporzionali. Nella seguente discussione focalizzeremo l'attenzione sul calcolo di  $\langle r^2 \rangle$  piuttosto che di  $\langle R_G^2 \rangle$ , ma il metodo utilizzato può essere facilmente esteso anche al calcolo di quest'ultima quantità.

### 3.1.1 Il modello di freely jointed chain (o random walk chain)

La catena *freely jointed* o *random walk* è un modello astratto in cui la direzione di ogni bond è completamente scorrelata da ogni altro bond della catena. Quindi, le conformazioni della catena sono descritte da un random walk (cammino casuale) in cui ogni passo, di lunghezza  $l$  è fatto in una direzione casuale indipendente dal precedente.



Nella Fig.3 mostriamo un esempio di una configurazione di una catena freely jointed. Analogamente a quanto visto prima, la sua end-to-end distance sarà ottenuta mediante la seguente

$$\langle r^2 \rangle = nl^2 + 2 \sum_{j>i}^n \langle \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j \rangle \quad (3.17)$$

ma ora, essendo ogni passo della catena completamente scorrelato da tutti gli altri, deve essere

$$\langle \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j \rangle = 0. \quad (3.18)$$

In tali condizioni l'end-to-end distance segue la legge

$$\langle r^2 \rangle = nl^2 \quad \Rightarrow \quad \langle r^2 \rangle^{1/2} = l\sqrt{n}. \quad (3.19)$$

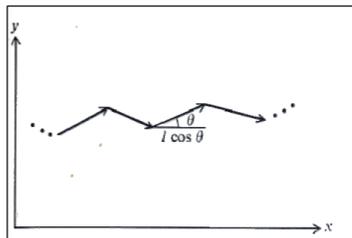
Si noti che vedendo la catena come il cammino casuale seguito da una particella, il tempo  $t$  intercorso tra il primo e l'ultimo step sarà essenzialmente

proporzionale al numero di step fatti,  $n$ , e quindi il risultato precedente si traduce in

$$\langle r^2 \rangle^{1/2} \propto \sqrt{t}. \quad (3.20)$$

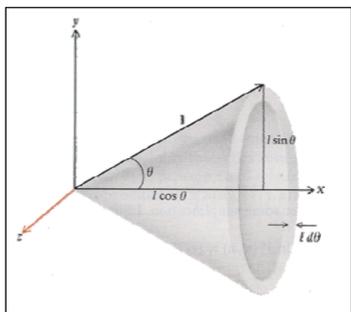
Questa relazione costituisce una delle proprietà caratteristiche di una particella che segue un moto browniano (per esempio, particella di soluto in un solvente).

Ricaveremo la distribuzione di  $r$  calcolando, prima la lunghezza media della proiezione di un passo qualsiasi della catena, e quindi, tenendo conto del loro orientamento, si introdurranno tutte le possibili configurazioni.



Considerando un passo (bond) qualsiasi della catena, la sua proiezione sull'asse  $x$  sarà data da  $l_x = l \cos \theta$  dove  $\theta$  è l'angolo che esso forma con l'asse stesso. Perciò il suo valore quadratico medio sarà dato da

$$\overline{l_x^2} = l^2 \overline{\cos^2 \theta}. \quad (3.21)$$



Il calcolo di questi valori medi deve tenere conto della distribuzione degli angoli  $\theta$ . In effetti, osservando la figura qui a fianco, si può notare che la probabilità di avere un valore compreso fra  $\theta$  e  $\theta + d\theta$  è proporzionale all'area  $2\pi l \sin \theta \cdot l d\theta$ .

Quindi, il valor medio precedente sarà dato da

$$\overline{l_x^2} = \frac{\int_0^\pi l^2 \cos^2 \theta \cdot 2\pi l^2 \sin \theta d\theta}{\int_0^\pi 2\pi l^2 \sin \theta d\theta} = -l^2 \frac{\int_0^\pi \cos^2 \theta d(\cos \theta)}{\int_0^\pi \sin \theta d\theta} \quad (3.22)$$

dalla quale otteniamo

$$\overline{l_x^2} = \frac{l^2}{3} \quad \rightarrow \quad (\overline{l_x^2})^{1/2} = \frac{l}{\sqrt{3}}. \quad (3.23)$$

Di conseguenza possiamo dire che la lunghezza media della proiezione di un qualsiasi bond della catena sui tre assi è

$$l_{x,y,z} = \frac{l}{\sqrt{3}}. \quad (3.24)$$

Ora, sempre considerando le proiezioni lungo l'asse  $x$ , supponiamo di averne  $n_+$  orientate in verso positivo e  $n_-$  in verso negativo; questo comporta un'estensione della catena lungo l'asse  $x$  pari a

$$x = \frac{l}{\sqrt{3}}(n_+ - n_-) = \frac{l}{\sqrt{3}}(2n_+ - n). \quad (3.25)$$

D'altra parte, il numero di configurazioni possibili con  $n_+$  proiezioni positive è dato da

$$W(n, n_+) = \frac{n!}{n_+!(n - n_+)!}, \quad (3.26)$$

e il numero di tutte le configurazioni possibili da  $2^n$ . La corrispondente probabilità è

$$P(n, n_+) = \frac{W(n, n_+)}{2^n} = \frac{1}{2^n} \frac{n!}{n_+!(n - n_+)!} = \frac{1}{2^n} \frac{n!}{\left(\frac{n+m}{2}\right)! \left(\frac{n-m}{2}\right)!} \quad (3.27)$$

dove abbiamo introdotto la quantità  $m = n_+ - n_- = 2n_+ - n$ . Si noti che  $m$  varia a passi di 2 e quindi  $m = -n, -n + 2, -n + 4, \dots, n - 4, n - 2, n$ .

Per grandi  $n$  possiamo utilizzare la formula di Stirling, nella sua versione più accurata:

$$\ln N! \simeq N \ln N - N + \frac{1}{2} \ln 2\pi + \frac{1}{2} \ln N, \quad (3.28)$$

tramite la quale otteniamo

$$\begin{aligned} \ln P(n, m) &= n \ln \frac{1}{2} + \ln n! - \ln \left(\frac{n+m}{2}\right)! - \ln \left(\frac{n-m}{2}\right)! \\ &\simeq \ln \sqrt{\frac{2}{n\pi}} - \frac{n+1}{2} \ln \left[1 - \left(\frac{m}{n}\right)^2\right] - \frac{m}{2} \ln \left[\frac{1 + \frac{m}{n}}{1 - \frac{m}{n}}\right] \end{aligned} \quad (3.29)$$

Osservando che per  $x \ll 1$  si ha  $\ln(1+x) \simeq x$  e  $\ln[(1+x)/(1-x)] \simeq 2x$ , allora per  $m \ll n$  la precedente può essere scritta nella forma

$$\ln P(n, m) \simeq \ln \sqrt{\frac{2}{n\pi}} - \frac{n+1}{2} \left(-\frac{m}{n}\right)^2 - \frac{m}{2} 2 \frac{m}{2} \simeq \ln \sqrt{\frac{2}{n\pi}} - \frac{m^2}{2n} \quad (3.30)$$

dalla quale ricaviamo che la probabilità di avere un dato valore di  $m$  è data da

$$P(n, m) \simeq \sqrt{\frac{2}{n\pi}} e^{-\frac{m^2}{2n}}. \quad (3.31)$$

Ma dato che  $x = (l/\sqrt{3})m$ , e ricordando che in generale  $x$  è continua, possiamo dire che un dato valore di  $m$  corrisponde ad avere  $x$  nell'intervallo  $[x(m) - l/\sqrt{3}, x(m) + l/\sqrt{3}]$  di larghezza  $\Delta x = 2l/\sqrt{3}$  (ricordarsi che  $m$  varia a passi di 2). Quindi, se con  $P(n, x)$  indichiamo la densità di probabilità (probabilità per unità di lunghezza) di avere una catena di lunghezza  $n$  con estensione lungo l'asse  $x$  tra  $x - \Delta x/2$  e  $x + \Delta x/2$ , abbiamo

$$P(n, x) = \frac{P(n, m)}{\Delta x} \Rightarrow P(n, x) = \sqrt{\frac{3}{2\pi n l^2}} e^{-3x^2/2nl^2}. \quad (3.32)$$

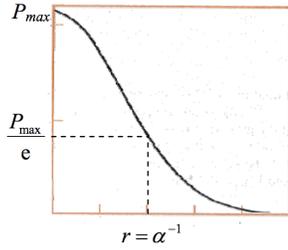
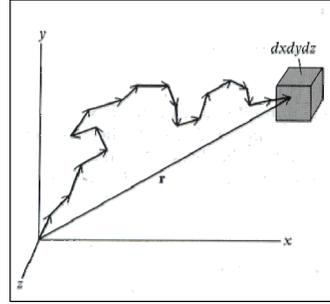
Infine, dato che si dovrà avere lo stesso comportamento anche a riguardo delle estensioni lungo gli assi  $y$  e  $z$ , possiamo concludere che la densità di probabilità totale sarà data da:

$$P(n, x, y, z) = P(n, x)P(n, y)P(n, z) \quad (3.33)$$

$$= \left(\frac{3}{2\pi n l^2}\right)^{3/2} e^{-3r^2/2nl^2} = \left(\frac{\alpha}{\pi}\right)^3 e^{-\alpha^2 r^2}, \quad (3.34)$$

dove  $r^2 = x^2 + y^2 + z^2$  e  $\alpha = \sqrt{3/2nl^2}$ .

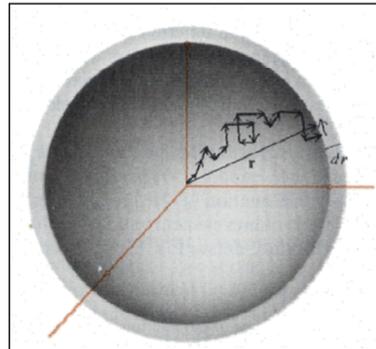
Ricordiamo che (ponendo il primo estremo della catena nell'origine) il significato di tale funzione è il seguente:  $P(n, x, y, z)dxdydz$  corrisponde alla probabilità che il secondo estremo cada esattamente nel volumetto di lati  $x \div x + dx, y \div y + dy, z \div z + dz$  come schematizzato nella figura qui a fianco.



La forma Gaussiana della funzione comporta un andamento della probabilità suddetta del tipo riportato nel grafico a lato nel quale si mostra che la larghezza della distribuzione è dell'ordine di  $\alpha^{-1} \propto \sqrt{n}$ . Si noti anche che la densità di probabilità  $P(n, x, y, z)$  è normalizzata. Infatti:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(n, x, y, z)dxdydz = \left[ \frac{\alpha}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\alpha^2 x^2} dx \right]^3 = 1. \quad (3.35)$$

D'altra parte, osservando la figura qui a destra, si capisce facilmente che la probabilità di avere un valore di  $r$  compreso tra  $r$  e  $r + dr$  (indipendentemente dalla direzione) corrisponde alla quantità

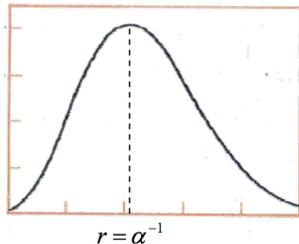


$$\begin{aligned} P(n, r)dr &= \int_{dV} P(n, x, y, z)dxdydz \\ &= \left( \frac{\alpha}{\sqrt{\pi}} \right)^3 e^{-\alpha^2 r^2} 4\pi r^2 dr \end{aligned} \quad (3.36)$$

dove si è integrato nella regione corrispondente al guscio sferico di raggio  $r$  e spessore  $dr$  (di volume pari a  $4\pi r^2 dr$ ). La funzione

$$P(n, r) \left( \frac{\alpha}{\sqrt{\pi}} \right)^3 4\pi r^2 e^{-\alpha^2 r^2} \quad (3.37)$$

è detta *funzione di distribuzione radiale* (o *densità di probabilità radiale*).



Il suo andamento in funzione di  $r$  è mostrato qui a fianco: essa presenta un massimo in corrispondenza di  $r = \alpha^{-1}$ . Tale distribuzione permette anche di semplificare il calcolo di  $\langle r^2 \rangle$ . Infatti otteniamo

$$\langle r^2 \rangle = \int_0^{\infty} r^2 P(r, n)dr = 4\pi \left( \frac{\alpha}{\sqrt{\pi}} \right)^3 \int_0^{\infty} r^4 e^{-\alpha^2 r^2} dr. \quad (3.38)$$

Integrando per parti si vede facilmente che

$$\int_0^\infty r^4 e^{-\alpha^2 r^2} dr = \frac{1}{\alpha^5} \int_0^\infty y^4 e^{-y^2} dy = \frac{1}{\alpha^5} \left[ -\frac{1}{2} y^3 e^{-y^2} \Big|_0^\infty + \frac{3}{2} \int_0^\infty y^2 e^{-y^2} dy \right] \quad (3.39)$$

dalla quale otteniamo

$$\int_0^\infty r^4 e^{-\alpha^2 r^2} dr = \frac{1}{\alpha^5} \frac{3\sqrt{\pi}}{8} \quad (3.40)$$

e infine

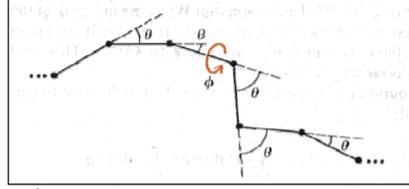
$$\langle r^2 \rangle = 4\pi \left( \frac{\alpha}{\sqrt{\pi}} \right)^3 \frac{1}{\alpha^5} \frac{3\sqrt{\pi}}{8} = \frac{3}{2} \frac{1}{\alpha^2} = \frac{3}{2} \frac{2nl^2}{3} = nl^2, \quad (3.41)$$

che ovviamente ha il valore già ottenuto all'inizio!

### 3.1.2 Catena con bonds ad angoli fissi

Tale modello, anche se non include effetti di volume escluso, è un passo in avanti rispetto al modello precedente.

Infatti, supporremo che i bonds della catena formino angoli fissi (pari a  $\theta$ ) mantenendo però la libertà di rotazioni intorno all'asse dei bonds stessi (vedi figura qui a fianco).



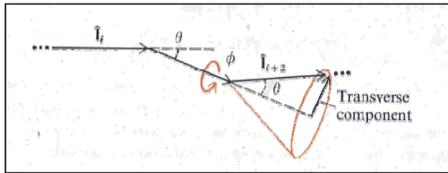
La maggior realistica del modello risiede nel fatto che, per esempio nelle proteine i bonds che compongono i peptidi che costituiscono il backbone della proteina stessa formano angoli essenzialmente fissi tra loro. In tali condizioni il calcolo di  $\langle r^2 \rangle$  comporta l'effettivo calcolo della sommatoria

$$\sum_{j>i}^n \langle \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j \rangle, \quad (3.42)$$

ma per il calcolo dei valori medi  $\langle \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j \rangle$  possiamo avvalerci del fatto che gli angoli  $\phi$  sono tutti equiprobabili. In effetti, è immediato ricavare che

$$\langle \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_{i+1} \rangle = l^2 \cos \theta. \quad (3.43)$$

Per quanto riguarda  $\langle \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_{i+2} \rangle$ , osservando la figura qui sotto, possiamo dire:



al variare del  $\phi$  relativo ad  $\hat{\mathbf{I}}_{i+1}$  vengono generate tutte le orientazioni di  $\hat{\mathbf{I}}_{i+2}$  rispetto ad  $\hat{\mathbf{I}}_i$  e questo comporta che, facendo la media, tutte le componenti di  $\hat{\mathbf{I}}_{i+2}$  perpendicolari a  $\hat{\mathbf{I}}_{i+1}$  si cancellano.

Al contrario, la componente parallela ad  $\hat{\mathbf{I}}_{i+1}$ , pari a  $l \cos \theta$ , si conserva e quindi la proiezione di tale lunghezza su  $\hat{\mathbf{I}}_i$  è pari a  $l \cos \theta \cdot \cos \theta = l \cos^2 \theta$ . Perciò abbiamo

$$\langle \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_{i+2} \rangle = l^2 \cos^2 \theta. \quad (3.44)$$

Procedendo in questo modo, per induzione, si può ricavare che

$$\langle \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j \rangle = l^2 (\cos \theta)^{j-i} \quad (3.45)$$

e che

$$\langle r^2 \rangle = nl^2 + 2l^2 \sum_{j>i} (\cos \theta)^{j-i} = nl^2 + 2l^2 \sum_{j>i} \lambda^{j-i}, \quad (3.46)$$

dove si è posto  $\lambda = \cos \theta$ . Inoltre, osservando che nell'ultima sommatoria esistono  $n - k$  termini del tipo  $\lambda^k$ , allora possiamo scrivere

$$\sum_{j>i} \lambda^{j-i} = \sum_{k=1}^{n-1} (n-k) \lambda^k. \quad (3.47)$$

Tale somma può essere ricavata utilizzando il seguente metodo. Poniamo

$$S_m = \sum_{k=1}^m \lambda^k \quad (3.48)$$

e osserviamo che valgono le seguenti relazioni:

$$S_{m+1} = S_m + \lambda^{m+1}; \quad S_{m+1} = (1 + S_m)\lambda. \quad (3.49)$$

Da notare che  $\lambda \leq 1$ . Mettendo a sistema si ricava

$$S_m = \frac{\lambda - \lambda^{m+1}}{1 - \lambda}. \quad (3.50)$$

D'altra parte

$$\sum_{k=1}^m k \lambda^k = \lambda \frac{dS_m}{d\lambda}, \quad \frac{dS_m}{d\lambda} = \frac{1 - \lambda^m}{(1 - \lambda)^2} - m \frac{\lambda^m}{1 - \lambda}, \quad (3.51)$$

da cui

$$\langle r^2 \rangle = nl^2 + 2l^2 \sum_{k=1}^{n-1} (n-k) \lambda^k = nl^2 + 2l^2 \left( nS_{n-1} - \lambda \frac{dS_{n-1}}{d\lambda} \right) \quad (3.52)$$

Quindi, con un pò di passaggi si ottiene:

$$\langle r^2 \rangle = nl^2 \left[ \frac{1 + \lambda}{1 - \lambda} - \frac{2\lambda(1 - \lambda^n)}{n(1 - \lambda)^2} \right]. \quad (3.53)$$

Tale espressione mostra che, contrariamente al caso della catena freely jointed (random walk),  $\langle r^2 \rangle$  dipende da  $n$ . Tuttavia, nel limite  $n \rightarrow \infty$  (per  $\lambda \neq 1$ ) abbiamo

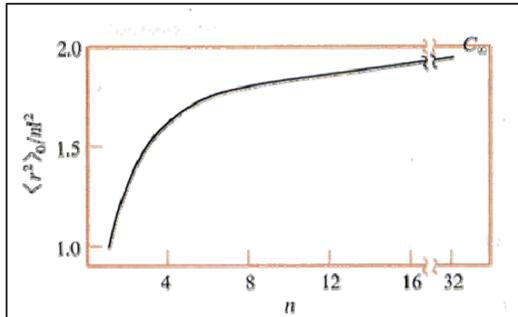
$$\langle r^2 \rangle \simeq nl^2 \left( \frac{1 + \lambda}{1 - \lambda} \right) \propto nl^2. \quad (3.54)$$

che ci riporta verso un andamento simile a quello del random walk puro.

A volte viene definita la quantità

$$C_n = \frac{\langle r^2 \rangle}{nl^2}. \quad (3.55)$$

Per esempio, per  $\lambda = 1/3$  ( $\theta \simeq 70.5^\circ$ ),  $C_n$  ha l'andamento mostrato nel grafico qui a lato; si noti che  $C_\infty = 2$ .

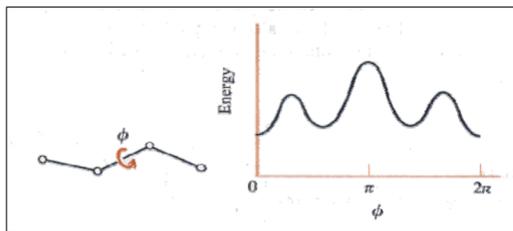


E' interessante notare che nel caso di  $\lambda = 1$ ,  $C_n$  cresce indefinitamente. Ciò è dovuto al fatto che i bond della catena sono allineati. In effetti, tanto più grande è il valore limite di  $C_n$ , cioè  $C_\infty$ , tanto più i bond della catena tendono ad essere allineati.  $C_\infty$  può quindi essere visto come una misura di questa tendenza.

### 3.1.3 Catene reali e rotazioni intorno ai bond

Un passo ulteriore verso un modello più realistico si ottiene introducendo una distribuzione degli angoli  $\phi$  che tiene conto delle energie di interazione associate ai suoi diversi valori.

In effetti, se da una parte il legame peptidico permette la rotazione completa del legame, ai diversi valori dell'angolo  $\phi$  corrisponde un'energia che segue l'andamento riportato nella figura qui a fianco.



Quindi, seguendo la statistica di Boltzmann, potremmo ricavarci la distribuzione di probabilità associata ai diversi  $\phi$ , attraverso la seguente espressione:

$$p(\phi)d\phi = \frac{e^{-E(\phi)/k}BT d\phi}{\int_0^{2\pi} e^{-E(\phi)/k}BT d\phi}. \quad (3.56)$$

Tale distribuzione avrà, ovviamente, i suoi massimi in corrispondenza dei minimi della funzione  $E(\phi)$ , ed essi, in caso di temperature non troppo elevate, potranno essere particolarmente stretti. Ciò significa che i corrispondenti valori di  $\phi$  saranno molto più probabili degli altri possibili valori. In considerazione di ciò, è possibile ridurre la complessità di tale modello, considerando soltanto gli angoli  $\phi_i$  corrispondenti ai minimi dell'energia  $E(\phi)$ , associando ad essi le probabilità

$$p(\phi_i) = \frac{e^{-E(\phi_i)/k}BT}{\sum_{i=1}^3 e^{-E(\phi_i)/k}BT}. \quad (3.57)$$

Tuttavia, anche se tali idee possono aiutare nell'adattamento del modello a specifiche caratteristiche del sistema considerato, questo tipo di approccio non porta a sostanziali differenze rispetto a quanto visto in precedenza. Per osservare differenze sostanziali si debbono introdurre effetti di volume escluso.

## 3.2 Effetti di volume escluso: metodo di Flory

In questo paragrafo ricaviamo gli effetti di volume escluso per una catena sulla base di argomenti qualitativi (alla Flory) che però risultano essere di grande portata. In effetti, molte delle valutazioni relative a catene che non si intersecano (*self avoiding walks*) si basano su enumerazioni di configurazioni di catene fatte al computer ed estrapolazioni numeriche che risultano essere in notevole

accordo con le conclusioni che trarremo con tale metodo. Le considerazioni che faremo qui risulteranno utili anche in relazione al fenomeno del *protein folding*.

Per prima cosa cerchiamo di capire quale nuovo ingrediente entra in gioco quando consideriamo una catena che non può intersecarsi. Sappiamo che gli effetti di volume escluso sono legati al fatto che quando due unità della catena si trovano a piccole distanze, nascono delle interazioni repulsive che si oppongono al loro ulteriore avvicinamento. Perciò, per una catena formata da  $n$  unità, ognuna delle quali ha un volume escluso effettivo pari a  $v_{ex}$ , con un diametro medio dell'ordine di  $r$  (essenzialmente pari a  $\langle r^2 \rangle^{1/2}$ ), si può pensare che l'energia repulsiva associata alle interazioni di volume escluso sia proporzionale alla densità delle unità (rapporto tra il volume escluso complessivo,  $nv_{ex}$ , e il volume occupato dalla catena, dell'ordine di  $r^3$ ) e ad  $n$ , cioè:

$$E_{ex} \sim k_b T \cdot \frac{nv_{ex}}{r^3} \cdot n = k_b T v_{ex} \frac{n^2}{r^3}. \quad (3.58)$$

Al contrario, quando la catena è molto estesa (dominano gli effetti entropici) sappiamo che si comporta come un random walk e quindi la probabilità  $P(r)$  che essa abbia una estensione pari ad  $r$  dovrà essere proporzionale a  $\exp(-3r^2/2nl^2)$ . In tali condizioni, l'entropia della catena è proporzionale al logaritmo della probabilità appena menzionata, e quindi

$$\frac{S}{k_B T} \sim -\frac{3r^2}{2nl^2}. \quad (3.59)$$

Volendo ricavare un'espressione che esprimesse l'estensione media  $r$  di una catena in funzione di  $n$  che tenesse anche conto degli effetti di volume escluso, Flory (nel 1953) espresse l'energia libera di Helmholtz della catena,  $A$ , per mezzo delle quantità appena stimate. Scrisse quindi l'energia libera nella forma

$$A = E_{ex} - TS \sim k_B T v_{ex} \frac{n^2}{r^3} + k_B T \frac{3r^2}{2nl^2} = k_B T \left( v_{ex} \frac{n^2}{r^3} + \frac{3r^2}{2nl^2} \right). \quad (3.60)$$

All'equilibrio, tale funzione deve essere minima e pertanto

$$\left( \frac{\partial A}{\partial r} \right)_{T,n} = 0 \Rightarrow k_B T \left( v_{ex} \frac{3n^2}{r^4} + \frac{3r}{nl^2} \right) = 0 \rightarrow r^5 = v_{ex} l^2 n^3, \quad (3.61)$$

dalla quale abbiamo

$$r \simeq \langle r^2 \rangle^{1/2} \propto n^{3/5}. \quad (3.62)$$

Sorprendentemente, tale andamento è in ottimo accordo con calcoli effettuati su enumerazioni esatte di self-avoiding walks tramite computer.

D'altra parte, supponendo di considerare catene in spazi a  $d$  dimensioni ( $d = 1, 2, 3, \dots$ ), il metodo precedente porta alla seguente generalizzazione:

$$A_d = E_{ex} - TS \sim k_B T \left( v_{ex} \frac{n^2}{r^d} + \frac{3r^2}{2nl^2} \right). \quad (3.63)$$

da cui

$$\left( \frac{\partial A_d}{\partial r} \right)_{T,n} = 0 \Rightarrow r \simeq \langle r^2 \rangle^{1/2} \propto n^{3/(2+d)}. \quad (3.64)$$

Gli andamenti che si ottengono alle diverse dimensioni spaziali sono in ottimo accordo con le determinazioni computazionali e sperimentali.