

Università di Padova

DIPARTIMENTO DI FISICA E ASTRONOMIA "GALILEO GALILEI"
LAUREA TRIENNALE IN FISICA

**Fluttuazioni critiche
negli stati nativi di proteine**

Candidato:
Ilario Vitali

Relatore:
Antonio Trovato

Anno accademico 2016-2017

Indice

1	Introduzione	1
1.1	Proprietà generali delle proteine	1
1.2	Tecniche sperimentali per ricavare la struttura proteica	2
1.3	Le proteine come sistema critico	3
2	Il modello a rete elastica	6
2.1	Analisi dei modi normali e funzione di correlazione	7
3	Risultati	8
3.1	Calibrazione del modello a rete elastica	8
3.2	Analisi delle strutture cristallografiche	10
4	Conclusioni e problemi aperti	13

Capitolo 1

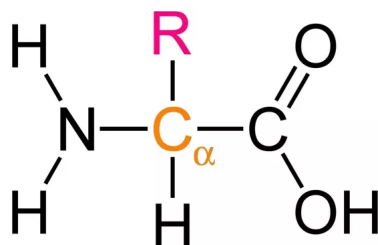
Introduzione

1.1 Proprietà generali delle proteine

Le proteine sono polimeri, ovvero catene di molecole legate da interazioni covalenti forti. Le molecole che, legate insieme, costituiscono le proteine sono chiamate amminoacidi. Ogni amminoacido è composto di alcuni atomi chiamati "backbone", uguali in ogni tipo di questi, e da altri chiamati "sidechain" che variano e ne definiscono il tipo. In **Figura 1.1** si riporta la struttura di un generico amminoacido: gli atomi C, C_α, N sono gli atomi di "backbone" mentre R indica la catena laterale.

Una proteina è caratterizzata da una sequenza di amminoacidi ben definita, chiamata struttura primaria. Questa catena può essere di diversa lunghezza e varia dalle decine alle centinaia di amminoacidi.

Figura 1.1: Struttura di un amminoacido

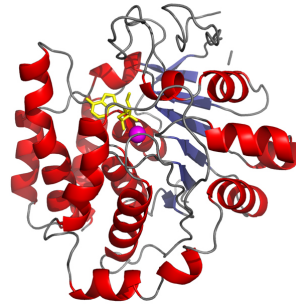


Oltre a una sequenza ben definita la maggior parte delle proteine mostrano una struttura tridimensionale ben definita, detta struttura terziaria, che può essere determinata tramite esperimenti di risonanza magnetica nucleare o di cristallografia a raggi X [3]. Un esempio è visibile in **Figura 1.2**.

Ci sono in aggiunta delle ricorrenze geometriche nella struttura, chiamate strutture secondarie, come alfa eliche e foglietti beta. Sia le alfa eliche che i foglietti beta sono strutture stabilizzate da legami a idrogeno fra atomi di backbone.

Un fatto molto importante è che una proteina con una data struttura primaria (ovvero una data sequenza di amminoacidi) ripiega su una ben definita struttura spaziale. Questa conformazione su cui ripiega è detta solitamente stato nativo della proteina.

Figura 1.2: Esempio di struttura tridimensionale di una proteina



In questa tesi ci occuperemo di proteine che ripiegano su strutture compatte in acqua, chiamate proteine globulari, a causa dell'effetto idrofobico. A causa di ciò le catene laterali idrofobiche tendono a rimanere all'interno della struttura in modo da evitare il contatto con molecole d'acqua.

Lo stato nativo dal punto di vista fisico è un sistema a bassa entropia a un minimo nell'energia libera[3].

Quest'ultimo ha grande importanza anche a livello biologico in quanto la disposizione tridimensionale della proteina permette a questa di svolgere la sua funzione.

1.2 Tecniche sperimentali per ricavare la struttura proteica

Come detto nella **Sezione 1.1** le tecniche sperimentali principali per ottenere la struttura delle proteine sperimentalmente sono due: la cristallografia a raggi X e la risonanza nucleare.

Il primo di questi si basa sulla diffrazione di un fascio elettromagnetico da parte di un cristallo fatto di proteine. Questo agisce come un reticolo di diffrazione formando un pattern di diffrazione che dipende della posizione degli atomi pesanti delle proteine. Nonostante, in linea di principio, basterebbe una singola proteina per creare una mappa di diffrazione, questa sarebbe troppo debole per essere rilevata. Invece un 'cristallo proteico' crea una figura diffrattiva molto nitida. La determinazione della struttura tramite un processo di cristallografia a raggi X si basa sul fatto che il pattern diffrattivo è la trasformata di Fourier della distribuzione elettronica che lo genera. In pratica chiamando con I l'intensità del fascio si ha che:

$$I(\mathbf{q}) = \langle |F(\mathbf{q})|^2 \rangle \quad (1.1)$$

ove F è detto fattore di struttura ed è definito come :

$$F(\mathbf{q}) = \int d^3\mathbf{r} \rho(\mathbf{r}) e^{i\mathbf{r}\cdot\mathbf{q}} \quad (1.2)$$

ove $\rho(\mathbf{r})$ è la densità elettronica, \mathbf{q} è il momento di scattering definito come $\mathbf{q} = \frac{\sin(\frac{\alpha}{2})}{\lambda}$, α è l'angolo di scattering e λ è la lunghezza d'onda del fascio.

Data questa definizione ne consegue che, tramite l'inversa della trasformata di Fourier si può ottenere la mappa della densità elettronica dal fattore di struttura.

Il problema maggiore è che l'intensità misurata non fornisce il fattore di forma ma il suo modulo. Per risolvere questo 'problema di fase' sono state sviluppate molte tecniche che non si tratteranno in questa tesi.

Questo metodo presenta alcuni problemi in quanto non tutte le proteine possono essere cristallizzate ed inoltre la struttura di una proteina cristallizzata può essere diversa da quella

che presenta nel suo ambiente naturale, ovvero in soluzione.

Nonostante questi svantaggi questa procedura permette di ottenere la mappa elettronica, e quindi la posizione, degli elementi più pesanti (non degli idrogeni perchè non contribuiscono in maniera apprezzabile al pattern di diffrazione) anche per grandi proteine[6].

Il secondo metodo utilizzato per determinare la struttura di una data proteina è il metodo NMR.

Questo si basa sull'interazione dei nuclei dispari di una proteina con un forte campo elettromagnetico. Irradiando la proteina con appropriate onde radio è possibile studiare l'energia di risonanza degli spin e da ciò ottenere informazioni su con quali altri atomi essi interagiscono.

In una proteina lunga un centinaio di amminoacidi ogni spin ha una energia di risonanza diversa e conseguentemente può essere identificato fra gli altri. Fondamentalmente per uno spin nucleare isolato la differenza tra due stati di spin $\frac{1}{2}$ in un dato campo magnetico B è di

$$\Delta E = \hbar\gamma B \quad (1.3)$$

ove γ è il rapporto giromagnetico del protone e B il modulo del campo magnetico esterno. Quindi la corrispondente frequenza angolare di risonanza ω sarebbe $\omega = \gamma B$. Nonostante ciò, in una proteina uno spin nucleare non è mai isolato, e un piccolo cambiamento del momento magnetico δB generato da correnti elettroniche nella sua vicinanza cambia la frequenza di risonanza a

$$\omega = \gamma(B + \delta B) = \gamma B(1 + \delta) \quad (1.4)$$

ove $\delta = \frac{\delta B}{B}$ è chiamato "shift chimico" e misura il cambiamento relativo dipendente dagli atomi nella risonanza. Questo cambiamento in tipici esperimenti di NMR con un campo magnetico di all'incirca 1 T è dell'ordine di 10^{-6} .

Un altro concetto chiave per capire la determinazione della struttura proteica tramite esperimenti di tipo NMR è che, quando uno spin nucleare viene eccitato tramite un'onda radio trasversale, questo può decadere eccitando uno spin vicino tramite un accoppiamento dipolo-dipolo. Siccome la forza dipolo dipolo va come d^{-6} (ove d è la distanza tra i due spin) se gli spin sono eccitati correttamente e viene misurato il loro decadimento sullo stato principale è possibile ottenere una densità d'intensità di risonanza proporzionale a una funzione della distanza tra due nuclei. Questo metodo è utilizzabile solo per piccole proteine in quanto con grosse proteine i picchi di assorbimento risultano indistinguibili[6]. Il risultato finale di una misura NMR è un insieme di strutture compatibili con i dati sperimentali descrivente l'ensemble conformazionale caratteristico della proteina in esame.

1.3 Le proteine come sistema critico

In questa sottosezione si vuole riassumere quanto trovato nell'articolo [1] che sarà poi il punto di partenza per il lavoro di questa tesi.

Come abbiamo detto precedentemente la struttura tridimensionale della proteina gioca un ruolo fondamentale nella sua funzione biologica.

Una proteina svolge le sue funzioni passando da una struttura tridimensionale all'altra e per fare ciò deve essere abbastanza sensibile da captare un dato segnale ma deve essere anche abbastanza stabile da fornire funzionalità specifica e robustezza strutturale.

Queste caratteristiche sono solite di un sistema critico. La proprietà basilare di un tale sistema è che la lunghezza di correlazione (lunghezza oltre la quale le correlazioni decadono) è divergente nel limite termodinamico, ovvero quando il numero di componenti del sistema tende all'infinito [7].

Analizzando quindi le fluttuazioni all'equilibrio di una data proteina, partendo dai dati

sperimentali sulla sua struttura, dovremmo trovare delle caratteristiche tipiche di un sistema in questo stato.

Nell'articolo [1] vengono studiate le fluttuazioni all'equilibrio di un vasto numero di proteine partendo dalla loro struttura determinata a partire da esperimenti di risonanza nucleare. Precisamente sono stati utilizzati 4988 ensemble di proteine che avessero non meno di 10 tipi di strutture, allineati attraverso una procedura di minimizzazione della deviazione standard delle posizioni dei C_α atomi.

Dalle deviazioni delle posizioni dei residui di una data struttura (identificate con la posizione dei C_α atomi) dalla loro media definite come in **Equazione 1.5** si definisce una funzione di correlazione come in **Equazione 1.6**.

$$\Delta \mathbf{r}_{i,q} = \mathbf{r}_{i,q} - \langle \mathbf{r}_i \rangle \quad (1.5)$$

($\langle \mathbf{r}_i \rangle = \frac{1}{Q} \sum_{q=1}^Q \mathbf{r}_{i,q}$, q indica il numero di struttura, i il numero di residuo, Q il numero totale di strutture relative a una data proteina)

$$\phi(r) = \frac{\sum_{i<j}^N K_{ij} \delta(r - r_{ij})}{\sum_{i<j}^N \delta(r - r_{ij})} \quad (1.6)$$

con K_{ij} uguale a:

$$K_{ij} = \frac{\bar{C}_{ij}}{(\bar{C}_{ii} \bar{C}_{jj})^{\frac{1}{2}}} \quad (1.7)$$

Ove C_{ij} è uguale a:

$$C_{ij} = \frac{1}{Q} \sum_q \Delta \mathbf{r}_{i,q} \cdot \Delta \mathbf{r}_{j,q} \quad (1.8)$$

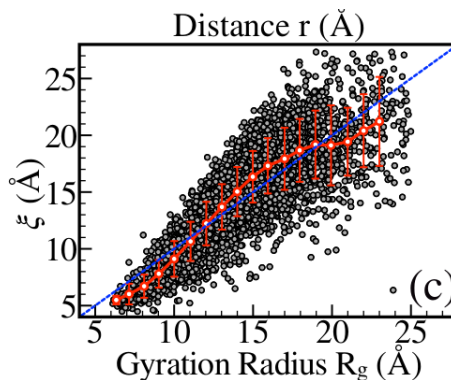
La funzione $\phi(r)$ così definita indica quanto sono correlate in media le direzioni in cui fluttuano due residui a una distanza r . Le funzioni correlazione sono risultate avere delle caratteristiche in comune per ogni proteina, in particolare hanno un minimo negativo all'incirca a -0.2.

Si è visto inoltre che le lunghezze di correlazione delle varie funzioni si distribuivano in modo proporzionale al raggio di girazione, definito come:

$$R_g = \sqrt{\frac{\sum_{i=1}^N (\mathbf{r}_i - \mathbf{r}_C) \cdot (\mathbf{r}_i - \mathbf{r}_C)}{N}} \quad (1.9)$$

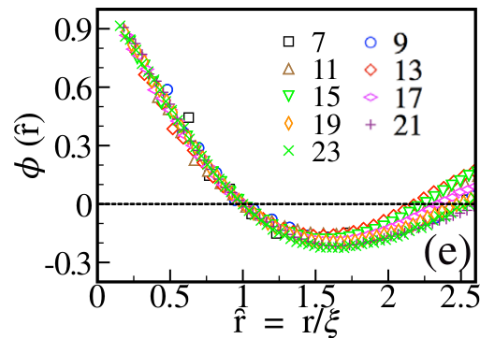
Ove \mathbf{r}_C è la posizione del centro di massa, \mathbf{r}_i la posizione dell' i -esimo residuo e N il numero di residui della data proteina. Questo risultato è visibile nella **Figura 1.3**.

Figura 1.3: Grafico della lunghezza di correlazione in funzione del raggio giratore per le proteine analizzate in [1]



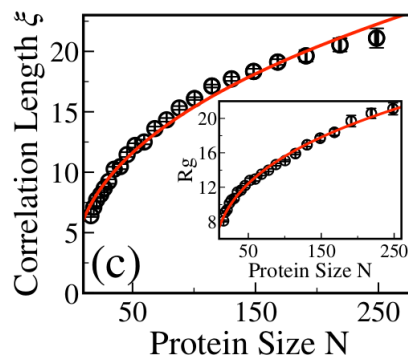
In particolare ricavando funzioni di correlazione per proteine con diversi raggi di girazione e riscalandole per la lunghezza di correlazione si nota che queste si dispongono su una funzione universale come si può vedere in **Figura 1.4**. È stato inoltre dimostrato che il riscaldamento delle correlazioni è indipendente dalla composizione e dalle interazioni dei residui.

Figura 1.4: Funzioni di correlazione riscalate divise per R_g trovate in [1]



Le proprietà di riscaldamento delle funzioni di correlazione descritte sopra sono tipiche di sistemi collettivi ove la correlazione è amplificata in vicinanza di un punto critico e quindi indicano che lo stato nativo è di questo tipo. Si è inoltre trovato che la lunghezza di correlazione e il raggio giratore aumentano con il numero di amminoacidi N secondo la legge a potenza $\xi \propto R_g \propto N^{\frac{1}{3}}$ come si mostra in **Figura 1.5**.

Figura 1.5: Relazione tra N e ξ , e N e R_g



Lo scopo di questa tesi è verificare anche per strutture cristallografiche, più grandi di quelle NMR, che sussistono le proprietà critiche delle fluttuazioni di stati nativi riportate in [1].

Capitolo 2

Il modello a rete elastica

Per calcolare i modi normali di oscillazione della proteina nello stato nativo si è fatto uso di quella che va sotto il nome di analisi dei modi normali (NMA).

L'assunzione implicita di questo metodo è che un dato sistema di equilibrio fluttua attorno a una singola conformazione ben definita e che la natura di queste fluttuazioni termiche può essere calcolata assumendo un semplice potenziale armonico [2]. Questo si è dimostrato nell'articolo [4] in cui si sono confrontati i modi vibrazionali lenti (chiamati anche modi globali) ottenuti tramite un semplice potenziale a rete elastica e tramite un potenziale dettagliato ottenendo risultati quasi identici.

Poichè poi l'energia associata ad un dato modo normale è proporzionale al quadrato della frequenza di oscillazione ω , uno spostamento di egual lunghezza lungo un modo ad alta frequenza sarà energeticamente più dispendioso rispetto ad uno a bassa frequenza. L'energia vibrazionale è distribuita tra i modi normali in modo che l'ampiezza media di oscillazione scali con $\frac{1}{\omega^2}$ e quindi in media gli spostamenti maggiori risultano a bassa frequenza. Concettualmente questi sono i più accessibili in quanto l'energia "sale più lentamente" attraverso i modi globali.[2].

È stato mostrato inoltre che quelle che sembrano fluttuazioni casuali nell'intorno dello stato nativo della proteina nascondono contributi a comportamenti cooperativi direttamente rilevanti per la sua funzione biologica. I moti cooperativi si verificano prevalentemente lungo i modi normali a bassa frequenza[2].

Partendo dall'NMA si è poi definita la funzione di correlazione dipendente dalla distanza da cui si è dedotta poi la criticità del sistema.

Come potenziale per la proteina vicino allo stato nativo si è quindi scelto un potenziale dato da un modello a rete elastica (ENM).

La disposizione della catena è fornita dalla struttura nativa della proteina ed i nodi della rete elastica sono identificati dalla posizione degli atomi C^α (il primo atomo di carbonio che si attacca a un gruppo funzionale) degli amminoacidi.

Nello specifico si è usato il modello a rete elastica che va sotto il nome di *Anisotropic Network Model* (ANM). Il potenziale di interazione usato tra i vari amminoacidi è quindi del tipo:

$$V = \sum_{i,j} \frac{1}{2} \gamma_{ij} (R_{ij} - R_{ij}^0)^2 \quad (2.1)$$

Ove R_{ij}^0 è la distanza fra i residui nello stato nativo.

Poichè i modi globali sono proprietà intrinseche della struttura della proteina questi risultano molto simili con diverse scelte dei parametri nel potenziale definito in **Equazione 2.1**, in particolare con diverse scelte della costante elastica[2].

La scelta di γ_{ij} utilizzata nelle simulazioni è stata:

$$\begin{cases} \gamma_{ij} = 1 & \text{se } R_{ij} < r_e \quad i \neq j \pm 1 \\ \gamma_{ij} = 0 & \text{se } R_{ij} > r_e \\ \gamma_{ij} = 10 & \text{se } i = j \pm 1 \end{cases} \quad (2.2)$$

Ove la scelta delle distanza di cutoff r_e sarà illustrata in **Sezione 3**.

2.1 Analisi dei modi normali e funzione di correlazione

A partire dal potenziale definito in **Equazione 2.1** si è definita la matrice hessiana \mathbf{H} come:

$$H_{ij} = -\frac{\gamma_{ij}}{(R_{ij}^0)^2} \begin{bmatrix} (x_{ij}^0)^2 & x_{ij}^0 y_{ij}^0 & x_{ij}^0 z_{ij}^0 \\ y_{ij}^0 x_{ij}^0 & (y_{ij}^0)^2 & x_{ij}^0 z_{ij}^0 \\ z_{ij}^0 x_{ij}^0 & z_{ij}^0 y_{ij}^0 & (z_{ij}^0)^2 \end{bmatrix} \quad (2.3)$$

Ove con x_{ij}^0 si intende $x_j^0 - x_i^0$ e in maniera del tutto analoga per z_{ij}^0 e y_{ij}^0 .

A questo punto si è definito l'Hessiano mass-weighted $\tilde{\mathbf{H}}$ come

$$\tilde{\mathbf{H}} = \mathbf{M}^{-\frac{1}{2}} \mathbf{H} \mathbf{M}^{-\frac{1}{2}} \quad (2.4)$$

ove \mathbf{M} è la matrice $3N \times 3N$ delle masse degli amminoacidi e N è il numero di residui della proteina considerata.

Calcolando autovettori e autovalori di questa matrice si ricavano le grandezze necessarie per definire i modi normali. In particolare chiamando $\lambda_j = \omega_j^2$ i $3N$ autovalori della matrice $\tilde{\mathbf{H}}$ e con $\tilde{\mathbf{u}}$ gli autovettori della matrice $\tilde{\mathbf{H}}$ (ove omega corrisponde alla velocità angolare di un dato modo normale) possiamo definire la matrice di covarianza \mathbf{C} come in **Equazione 2.5**

$$\mathbf{C}_{ij} = \mathbf{K}_B \mathbf{T} \mathbf{H}^{-1} \quad (2.5)$$

Ove, poichè \mathbf{H}^{-1} non può essere calcolata in quanto ha tipicamente 6 autovalori nulli corrispondenti a 3 gradi di libertà traslazionali e a 3 gradi di libertà rotazionali, viene sostituita dalla sua pseudo-inversa definita come:

$$\tilde{\mathbf{H}}^{-1} = \sum_k \frac{\tilde{\mathbf{u}}_k \tilde{\mathbf{u}}_k^T}{\omega_k^2} \quad (2.6)$$

ove la somma è ristretta agli autovalori, e autovettori corrispondenti, non nulli. La matrice di covarianza $3N \times 3N$ così definita è organizzata in sottomatrici 3×3 del tipo:

$$\mathbf{C}_{ij} = \begin{bmatrix} \langle \Delta x_i \Delta x_j \rangle & \langle \Delta x_i \Delta y_j \rangle & \langle \Delta x_i \Delta z_j \rangle \\ \langle \Delta y_i \Delta x_j \rangle & \langle \Delta y_i \Delta y_j \rangle & \langle \Delta y_i \Delta z_j \rangle \\ \langle \Delta z_i \Delta x_j \rangle & \langle \Delta z_i \Delta y_j \rangle & \langle \Delta z_i \Delta z_j \rangle \end{bmatrix} \quad (2.7)$$

A partire da questa matrice si definisce quindi la matrice di covarianza $N \times N$ delle fluttuazioni dei vettori \mathbf{R}_i , $\bar{\mathbf{C}}$, sostituendo alle sottomatrici 3×3 della matrice di covarianza $3N \times 3N$ la loro traccia.

$$\text{tr}(\mathbf{C}_{ij}) = \langle \Delta x_i \Delta x_j \rangle + \langle \Delta y_i \Delta y_j \rangle + \langle \Delta z_i \Delta z_j \rangle = \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \bar{\mathbf{C}}_{ij} \quad (2.8)$$

Si può quindi definire una funzione di correlazione spaziale come in **Equazione 1.6**.

Capitolo 3

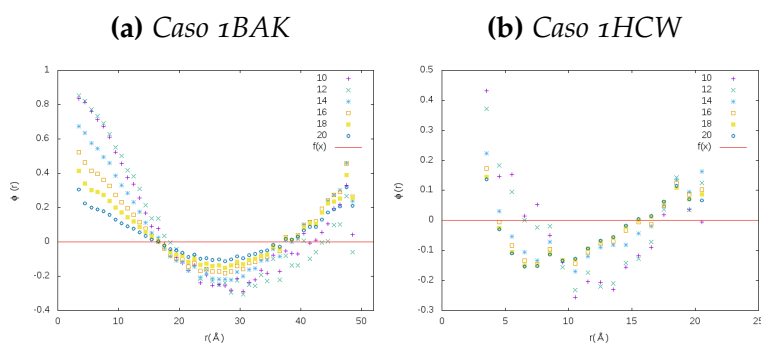
Risultati

3.1 Calibrazione del modello a rete elastica

Si è iniziato verificando che la funzione di correlazione ottenuta tramite il modello spiegato in **Sezione 2.2** fosse la stessa ottenuta dall'analisi svolta nell'articolo [1]. Si è compiuta inizialmente l'analisi sulle strutture NMR di due proteine (1BAK e 1HCW) analizzate anche nell'articolo citato precedentemente in modo da poter confrontare la funzione di correlazione.

Per ognuna delle due proteine erano presenti diverse strutture relative allo stato nativo. Nel caso della proteina 1BAK 20 strutture mentre nel caso della proteina 1HCW 35. Per tarare la distanza di cutoff r_e e per capire quanto questo parametro influenzasse i risultati finali si sono calcolate le funzioni di correlazione per diverse strutture proteiche. Nei casi riportati in **Figura 3.1** sono riportate le diverse funzioni di correlazione ottenute variando la distanza di cutoff r_e da (10 a 20) Å per una struttura di 1HCW e una di 1BAK. Confrontando le funzioni di correlazione ottenute con quelle ottenute in [1] si è deciso di utilizzare come r_e 12 Å nel seguito.

Figura 3.1: Funzioni di correlazione con vario r_e



In **Figura 3.2** sono rappresentate le funzioni di correlazione per ogni struttura relativa alla proteina 1BAK e la relativa media mentre in **Figura 3.3** quelle relative alla proteina 1HCW. In questo caso e anche nei casi presentati successivamente le funzioni di correlazione sono state calcolate come media dei valori di queste in un dato bin; la larghezza del bin utilizzato è di 1 Å in linea con quanto fatto in [1].

Le funzioni di correlazione trovate sono risultate del tutto simili a quanto trovato nell'articolo [1].

Figura 3.2: Funzioni di correlazione per la proteina 1BAK

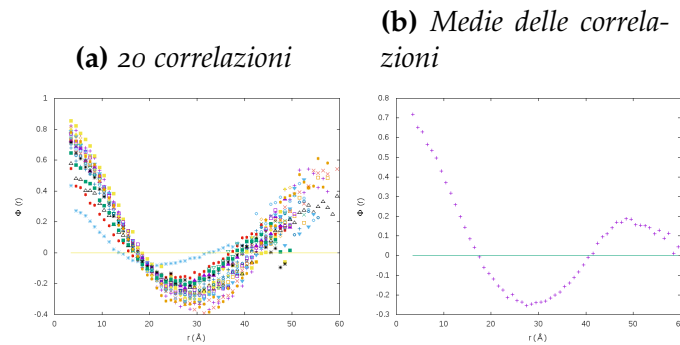
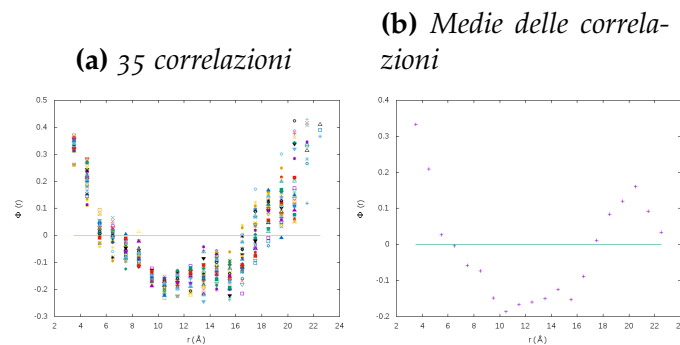


Figura 3.3: Funzioni di correlazione per la proteina 1HCW

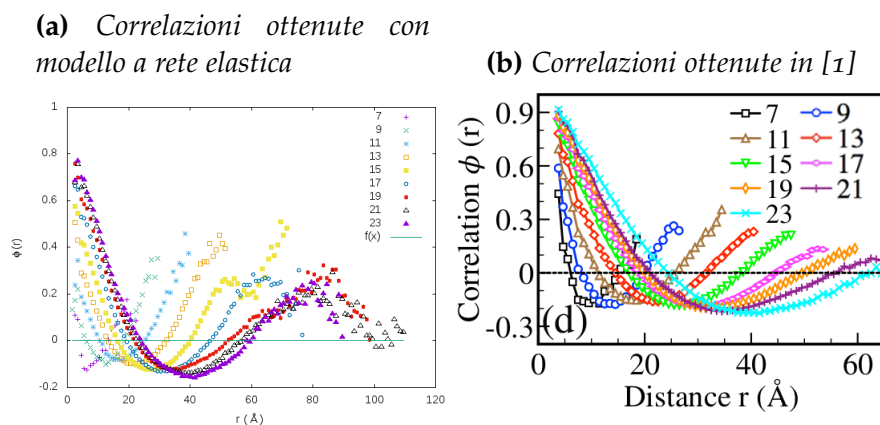


Si sono poi calcolate le funzioni di correlazione delle strutture suddivise per raggio giratore utilizzate in [1] per verificare di ritrovare gli stessi risultati. Si riporta in **Figura 3.4** l'andamento delle funzioni di correlazione suddivise per raggio giratore trovate ripetitivamente con il modello a rete elastica e con il metodo spiegato in **Sezione 1.3** usato nell'articolo sopracitato.

Si nota che l'andamento di quelle trovate attraverso la modellizzazione della proteina come rete elastica non differiscono molto (anche se hanno un minimo leggermente più alto) rispetto a quelle in **Figura 3.4 (b)**.

Questo risultato, oltre a mostrare che il metodo utilizzato in [1] porta agli stessi risultati del metodo che si è utilizzato in questa tesi, è un'ulteriore conferma che la scelta del parametro $r_e = 12 \text{ \AA}$ è ottimale e quindi si è mantenuto tale valore anche per i calcoli che saranno mostrati nelle sezioni successive.

Figura 3.4: Correlazioni suddivise per raggio giratore ricavate da strutture NMR

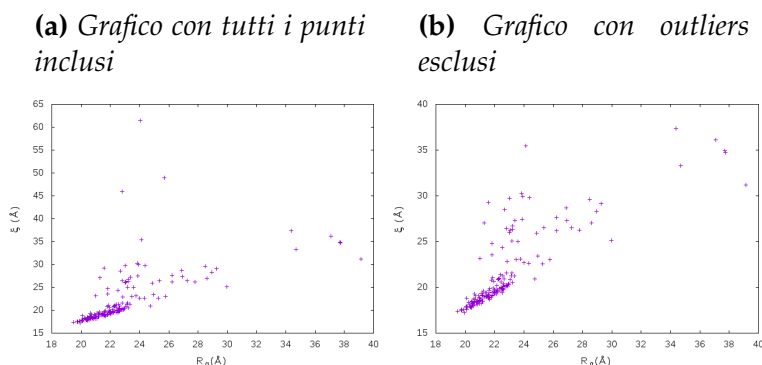


3.2 Analisi delle strutture cristallografiche

Si è quindi passato ad analizzare un set di strutture cristallografiche relative a proteine più grandi (con più di 400 residui) trovate grazie ad esperimenti di diffrazione e poi filtrate in "singoli domini" (ovvero catene o parti di catena che ripiega in maniera autonoma nello stato nativo) grazie alla classificazione CATH ¹.

Delle 176 strutture analizzate si è inizialmente controllata la disposizione del raggio giratore in funzione della lunghezza di correlazione come si riporta in **Figura 3.5** togliendo i tre punti che risultavano come 'outliers'.

Figura 3.5: Disposizione della lunghezza di correlazione in funzione dell raggio giratore



Di queste strutture si è fatta l'analisi precedente suddividendo le proteine per raggio giratore (definito come in **Equazione 1.9**) e si sono riscalate le curve trovate per la loro lunghezza di correlazione trovando, a differenza di quanto trovato in [1], tre curve su cui riscalano i dati.

Le strutture analizzate sono state suddivise per raggio giratore in bin di 1 Å e si è tenuta per la distanza tra residui il bin di 1 Å come in precedenza.

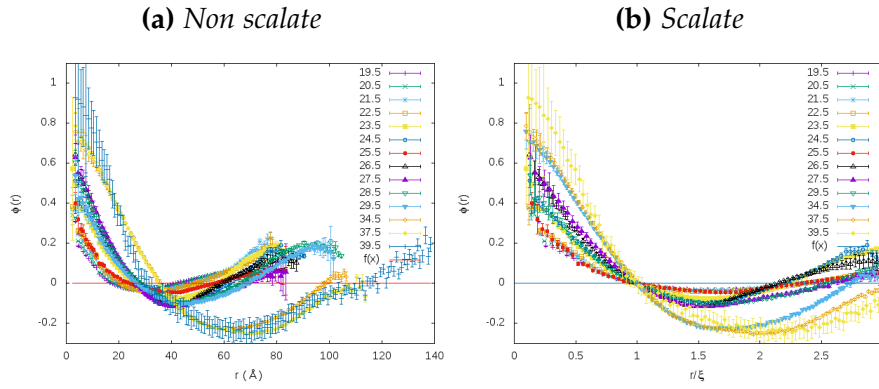
Gli errori associati ai valori della funzione di correlazione nei vari bin sono ricavati come errore sulla media ovvero come:

$$\sigma_{\bar{\phi}_i} = \frac{\sqrt{\langle \phi_i^2 \rangle - \langle \phi_i \rangle^2}}{\sqrt{N_i}} \quad (3.1)$$

Ove i indica il numero del bin. Si riportano le funzioni di correlazione trovate con le relative barre di errore in **Figura 3.6** e in **Figura 3.7** le funzioni di correlazione riscalate.

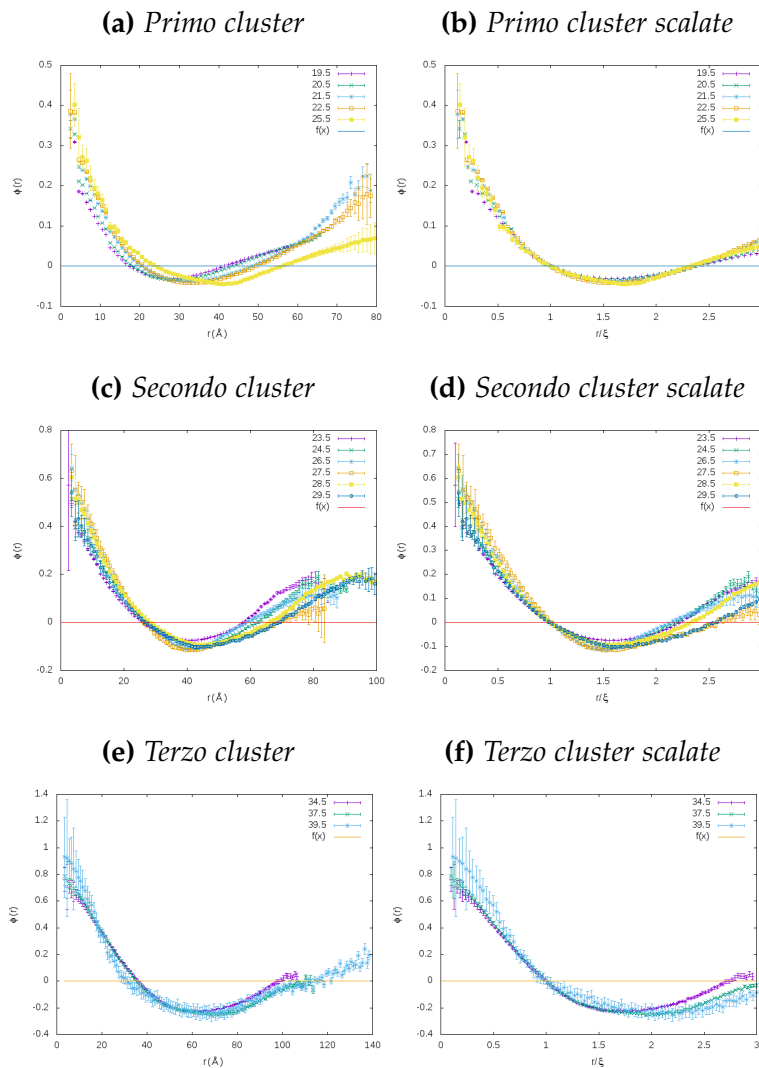
¹I domini delle strutture proteiche date dalle strutture del PDB sono classificati tramite la struttura gerarchica CATH : al livello C i domini sono classificati a seconda della loro struttura secondaria (per esempio α eliche), al livello A si classifica la disposizione delle strutture secondarie nello spazio, al livello T viene classificata la connessione e l'arrangiamento delle strutture secondarie. Sono infine classificate tramite una classe H se queste strutture sono omologhe.

Figura 3.6: Funzioni di correlazione con relativo errore con diversi raggi di girazione



Le funzioni di correlazione riscalano su una data curva a seconda del raggio giratore. Il primo cluster corrisponde a funzioni di correlazione con raggio giratore di $(19.5, 20.5, 21.5, 22.5, 25.5) \pm 0.5 \text{ \AA}$, il secondo a $(23.5, 24.5, 26.5, 27.5, 28.5, 29.5) \pm 0.5 \text{ \AA}$ mentre il terzo a $(34.5, 37.5, 39.5) \pm 0.5 \text{ \AA}$.

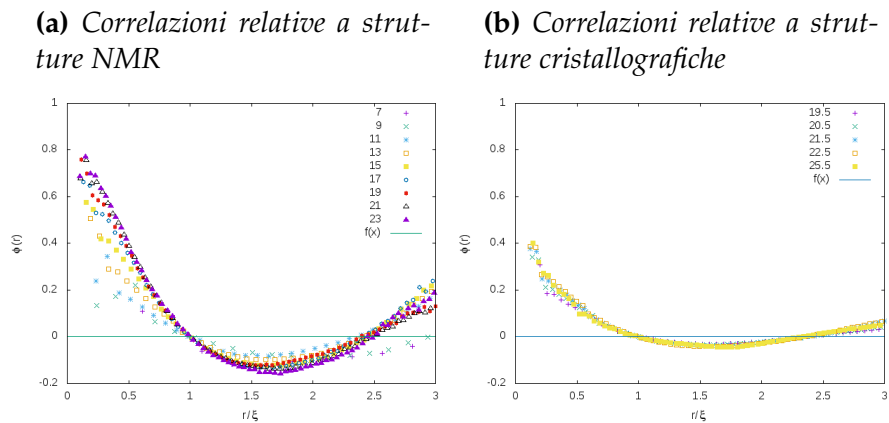
Figura 3.7: Funzioni di correlazione con relativo errore con diversi raggi di girazione riscalate



Analizzando **Figura 3.7** si può notare che solo nel caso del primo cluster migliora la sovrapposizione delle curve dopo il riscaldamento.

Si aggiungono inoltre in **Figura 3.8** tutte le funzioni di correlazione riscalate, trovate con il modello a rete elastica analizzando le strutture usate in [1], a fianco di quelle trovate tramite l'analisi di strutture cristallografiche appartenenti al primo cluster. Da queste ultime due figure risulta evidente un diverso andamento della funzione di correlazione nelle due strutture. Per le strutture cristallografiche con raggio giratore simile a quelli più alti delle strutture NMR (ovvero tra i 18 e i 25 Å) si nota che la funzione di correlazione ottenuta a partire dai due diversi dataset sono molto diverse. In particolare il minimo in quest'ultime è molto più basso rispetto alle prime.

Figura 3.8: Funzioni di correlazione riscalate divise per diversi R_g



Capitolo 4

Conclusioni e problemi aperti

Analizzando quindi con un modello a rete elastica strutture proteiche più grandi di quanto fatto nell'articolo [1] si sono trovati tre diversi tipi di funzioni di correlazione.

Questo potrebbe essere causato da due fattori: un'ipotesi è che questo sia dovuto dalla modellizzazione, forse troppo semplicistica, utilizzata per la proteina mentre l'altra dal fatto che le strutture proteiche ricavate tramite cristallografia a raggi X portino effettivamente a risultati diversi rispetto a quelli ricavati tramite risonanza nucleare (NMR). Si esclude la prima ipotesi in quanto si è notato che, compiendo la stessa analisi utilizzata nel caso delle strutture cristallografiche, per le strutture NMR utilizzate in [1] si sono ottenuti risultati confrontabili con quelli dell'articolo.

I risultati ottenuti, comparando le funzioni scalate del primo cluster ottenute partendo da strutture cristallografiche con quelle relative a strutture NMR, sembrano avvalorare la seconda ipotesi come si può notare in **Figura 3.8**.

Si nota infatti che le due funzioni di correlazione riscalate differiscono nell'andamento; infatti quella ricavata da strutture cristallografiche presenta un minimo molto più schiacciato rispetto a quelle relative a strutture ricavate tramite esperimenti di risonanza nucleare.

Dalla **Figura 3.7** si nota inoltre che il riscaldamento delle funzioni di correlazione relative a strutture cristallografiche non è ottimale nel caso del secondo e del terzo cluster. Questo comportamento è attribuibile alla minor statistica a cui sono affetti questi due cluster rispetto al primo (vedi **Figura 3.5**).

Per completare l'analisi iniziata in questa tesi bisognerebbe anzitutto ripeterla per strutture cristallografiche di qualsiasi dimensione e verificare con certezza che non vi siano artefatti dovuti all'utilizzo del ENM per modellizzare strutture molto grandi. Rimane inoltre da capire appieno il risultato visibile in **Figura 3.8**; infatti nonostante ci si aspetta che la struttura "statica" ricavata tramite NMR e cristallografia a raggi X sia la stessa è anche naturale che le fluttuazioni in un cristallo proteico e in soluzione siano differenti. Questo però non vale per la funzione di correlazione che oltretutto è stata ricavata a partire da un approccio basato sul ENM che dipende unicamente dalla struttura "statica".

Bibliografia

- [1] Qian-Yuan Tang, Yang-Yang Zhang, Jun Wang, Wei Wang, Dante R. Chialvo. *Critical Fluctuations in the Native State of Proteins*, [10.1103/PhysRevLett.118.088102], 2017
- [2] Ivet Bahar, Timothy R. Lezon, Ahmet Bakan, Indira H. Shrivastava *Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins*, 2010
- [3] Alexei V. Finkelstein, Oleg B. Ptitsyn *Protein Physics*, 2002
- [4] Monique M. Tirion *Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis*, 1996
- [5] Chakra Chennubhotla, A J Rader, Lee-Wei Yang, Ivet Bahar *Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies*, 2005
- [6] Guido Tiana *Physics of protein folding*, 2017
- [7] Kerson Hoang *Statistical Mechanics*, 1987, Wiley