

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI FISICA E ASTRONOMIA "GALILEO GALILEI"

CORSO DI LAUREA IN FISICA

**Fluttuazioni in complessi di proteine:
confronto fra dinamica molecolare e modelli
di rete elastica**

Relatore: Prof. Antonio Trovato

Laureando: Chiara Monni, 1027483

ANNO ACCADEMICO 2014/2015

1	Introduzione: le proteine	1
1.1	Struttura primaria	2
1.2	Struttura secondaria	2
1.2.1	α -elica	2
1.2.2	β -foglietti	2
1.2.3	Ripiegamenti β	3
1.3	Struttura terziaria e quaternaria	3
1.3.1	Proteine globulari	4
2	Affinità di legame	5
2.1	Termodinamica dei processi di associazione	7
2.2	Energia libera vibrazionale	8
2.2.1	Approssimazione armonica classica	9
3	Modelli di rete elastica	13
4	Stima dell'entropia vibrazionale per complessi di proteine	17
4.1	Complesso 1BVK	18
4.1.1	Analisi preliminare	18
4.1.2	Calibrazione delle mobilità	20
4.2	Subunità 1BVL e 3LZT	23
4.3	Contributo vibrazionale all'energia libera di legame	25
5	Conclusioni	29

Bibliografia

31

CAPITOLO 1

Introduzione: le proteine

Le proteine sono le macromolecole biologiche più abbondanti, presenti in tutti i tipi di cellule e in tutte le frazioni subcellulari. Mediano praticamente tutti i processi che hanno luogo nelle cellule e svolgono un numero enorme di funzioni, in particolare sono gli strumenti molecolari con cui si esprime l'informazione genetica. Alla base della struttura delle migliaia di proteine esistenti ci sono unità monomeriche relativamente semplici: le proteine sono formate dalla stessa serie di venti amminoacidi, uniti covalentemente in caratteristiche sequenze lineari. L'enorme varietà di funzioni svolte è profondamente connessa con la rigidità della loro struttura tridimensionale. Per conoscere l'interazione tra molecole proteiche è necessario studiarne la dinamica.

Gli amminoacidi che compongono la proteina sono uniti da legami peptidici, che concorrono a formare una catena. In particolare la sequenza dei residui amminoacidici (ovvero la porzione di amminoacido libero che rimane dopo la polimerizzazione), è unica per ogni proteina e il numero di residui nella catena proteica varia dalla ventina alle centinaia.

Ogni proteina ha una specifica funzione chimica o strutturale, questo porta a pensare che abbia una sola struttura tridimensionale, che sarà determinata dalla sua sequenza amminoacidica, come confermato dall'esperimento di Anfinsen^[1]. La disposizione spaziale degli atomi è detta conformazione. Le conformazioni possibili sono le strutture che la proteina può assumere senza rompere legami covalenti (ad esempio per rotazione intorno a legami singoli). Se la proteina si trova in uno dei suoi stati conformazionali funzionali, è detta nativa. La stabilità viene definita come tendenza a mantenere una conformazione nativa.

1.1 Struttura primaria

La struttura primaria di una proteina è definita come la sequenza lineare degli amminoacidi in una catena polipeptidica. L'ordine con cui compaiono gli amminoacidi lungo la sequenza è fissato ed è determinato geneticamente dai nucleotidi nel gene che codifica per la proteina stessa. La sequenza amminoacidica ha un ruolo fondamentale nel determinare la struttura tridimensionale della proteina, e quindi anche la sua funzione biologica.

1.2 Struttura secondaria

La struttura secondaria descrive l'organizzazione spaziale di un segmento della catena principale, senza tener conto delle catene laterali o della reazione con altri segmenti della proteina. Solo poche strutture secondarie sono regolari e hanno una particolare stabilità, le principali sono l' α elica, la configurazione β e il ripiegamento β . Se non è possibile individuare una struttura regolare, la struttura viene definita casuale (*random coil*).

1.2.1 α -elica

La più semplice organizzazione regolare che una catena polipeptidica può assumere è una struttura elicoidale. Lo scheletro carbonioso polipeptidico si avvolge strettamente attorno ad un asse che attraversa longitudinalmente la parte centrale della spirale, mentre i gruppi dei residui amminoacidici sporgono al di fuori dello scheletro elicoidale. In linea di principio si potrebbero formare sia eliche destrorse che sinistrorse, ma le sinistrorse sono meno stabili e non sono state osservate nelle proteine. Il motivo per cui questa struttura è quella che si forma con maggior facilità è che nell' α elica la disposizione dei legami idrogeno è la migliore possibile: ciascun legame peptidico (eccetto quelli alle estremità) partecipa alla formazione di legami idrogeno. Non tutti i polipeptidi possono formare α eliche stabili.

1.2.2 β -foglietti

Si tratta di una conformazione più estesa della catena polipeptidica, in questo caso lo scheletro si estende in una conformazione a zig-zag, invece che a spirale. Le catene polipeptidiche sono disposte una accanto all'altra, formando una struttura che presenta una serie di pieghettature. In questa disposizione i legami idrogeno si formano tra regioni adiacenti delle catene polipeptidiche. I gruppi R di amminoacidi adiacenti sporgono dalla struttura a zig-zag in direzioni opposte. Le catene adiacenti possono essere parallele o

antiparallele (cioè avere lo stesso orientamento del legame NH–CO, o un orientamento opposto.)

1.2.3 Ripiegamenti β

Nelle proteine globulari, che hanno una struttura ripiegata compatta, quasi un terzo dei residui amminoacidici si trova in ripiegamenti o anse, dove la catena polipeptidica inverte la sua direzione. Questi ripiegamenti collegano tratti successivi in α -eliche o conformazioni β . Comuni sono i ripiegamenti β , che collegano le estremità di due segmenti adiacenti di un foglietto β antiparallelo. In particolare gli anticorpi, complessi proteici che verranno presi in esame in questo lavoro di tesi, sono composti da tre catene, unite da legami non covalenti e ponti disolfuro. Una delle catene interagisce con le altre ad una estremità con le altre due catene, assumendo un forma simile ad una Y. Ogni catena contiene una serie di unità costituite da circa un centinaio di amminoacidi formanti una struttura globulare detta dominio Ig costituito da due foglietti β .

1.3 Struttura terziaria e quaternaria

La disposizione nello spazio di tutti gli atomi di una proteina viene definita struttura terziaria. Questa struttura tiene conto delle relazioni a lungo raggio nella sequenza amminoacidica. Gli amminoacidi che si trovano lontani nella sequenza polipeptidica, e quindi fanno parte di strutture secondarie diverse, possono interagire tra loro nella forma completamente avvolta della proteina. I segmenti della catena polipeptidica vengono mantenuti nelle loro posizioni tipiche dalla struttura terziaria tramite interazioni deboli e talora anche tramite legami covalenti, come ponti disolfuro.

Alcune proteine contengono due o più catene polipeptidiche distinte, o subunità. La disposizione di queste subunità in complessi tridimensionali prende il nome di struttura quaternaria. Considerando questo livello strutturale è possibile classificare le proteine in due classi principali: le proteine fibrose, che hanno catene disposte in lunghi fasci o foglietti, e le proteine globulari, che hanno catene ripiegate e assumono forme globulari o sferiche. I due gruppi sono strutturalmente molto diversi: mentre le proteine fibrose sono costituite in gran parte da un unico tipo di struttura secondaria e la struttura terziaria è relativamente semplice, le proteine globulari contengono più tipi di struttura secondaria.

1.3.1 Proteine globulari

Nelle proteine globulari diversi segmenti della catena polipeptidica si avvolgono gli uni sugli altri, formando una struttura compatta. Questo avvolgimento è responsabile delle molteplicità strutturali che permettono una così vasta gamma di funzioni. Le proteine globulari includono enzimi, proteine di trasporto, proteine regolatrici, immunoglobuline e altre. Ciascuna proteina ha una sua struttura caratteristica, adatta alla sua particolare funzione biologica, tuttavia esse hanno importanti proprietà in comune. Tutte hanno una forma compatta con le catene laterali idrofobiche localizzate all'interno (lontano dall'acqua) e le catene idrofiliche sulla superficie. La struttura tridimensionale delle tipiche proteine globulari può essere considerata come l'insieme di segmenti polipeptidici in α -elica e in conformazione β , uniti da elementi di connessione. La struttura viene descritta definendo il modo in cui i diversi segmenti si dispongono uno rispetto all'altro e come sono uniti dagli elementi di connessione: per studiare una proteina globulare occorre analizzare come le sue parti si avvolgono nello spazio, determinando i processi che regolano il ripiegamento. Come esempio mostriamo in figura 1.1 il complesso 1BVK (formato da tre catene), di cui ci occuperemo in questa tesi.

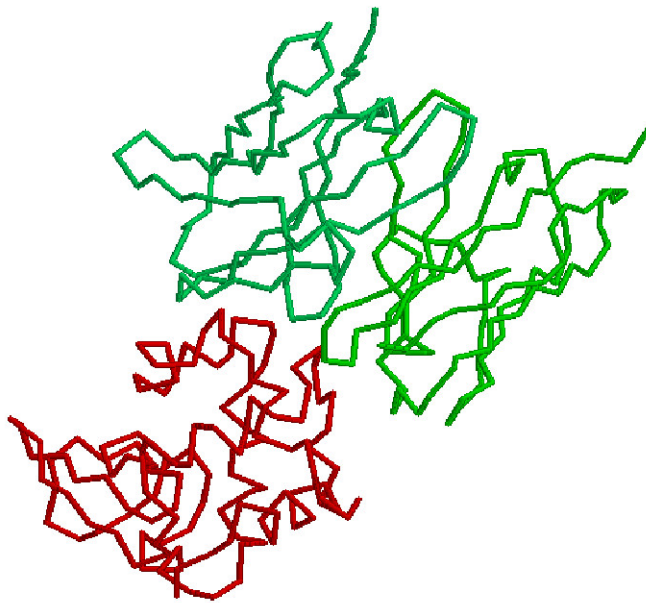


Figura 1.1: Complesso 1BVK, anticorpo monoclonale studiato in questo lavoro di tesi

CAPITOLO 2

Affinità di legame

Nei sistemi biologici le molecole si uniscono e si dissociano continuamente. Le associazioni molecolari sono il primo passo nella maggior parte delle forme di comunicazione biologica.

Per comprendere i processi di associazione tra molecole verranno esaminati i principi termodinamici e meccanico-statistici che sottostanno ai processi chimici di associazione. In particolare ci sono due principi guida:

1. le forze che controllano le associazioni sono solitamente non-covalenti (sono ad esempio forze elettrostatiche, legami idrogeno e interazioni idrofobiche);
2. Le associazioni sono stereospecifiche e dipendono da una precisa disposizione spaziale dei gruppi interagenti

Se due molecole A e B in soluzione formano il complesso C, si avrà un equilibrio chimico del tipo



E' possibile considerare lo stesso processo con A macromolecola e B piccolo ligando, oppure con entrambe macromolecole. Il complesso che si forma tra le due è denotato con C. All'equilibrio le concentrazioni delle tre molecole sono legate dalla relazione

$$\frac{[C]}{[A][B]} = \frac{1}{K_{dis}} \quad (2.2)$$

dove K_{dis} è la costante di dissociazione e ha le unità di misura di una concentrazione. Il suo reciproco è denominato costante di associazione. Un'alta costante di associazione

implica legami più stretti, dunque è possibile utilizzare questa quantità come misura della affinità di legame, o della forza di interazione.

E' possibile collegare la costante di dissociazione all'energia e valutare la forza dei legami non-covalenti che stabilizzano il complesso. L'energia libera per un soluto in una soluzione ideale è

$$G = G^0 + RT \ln[A] \quad (2.3)$$

dove G^0 è l'energia libera dello stato standard. Utilizzando l'equazione 2.3 per ognuna delle tre specie è possibile ottenere la differenza di energia libera per lo stato standard, assumendo ΔG nulla quando le tre specie sono all'equilibrio. Questo porta alla seguente relazione tra ΔG^0 e la costante di dissociazione:

$$\Delta G^0 = -RT \ln \frac{[C]}{[A][B]} = RT \ln K_{dis} \quad (2.4)$$

In generale un complesso si viene a formare tramite legami idrogeno, ponti salini e giustapposizione di superfici idrofobiche. Conoscendo la struttura del complesso è possibile sapere quali interazioni lo tengano insieme e dunque stimare le energie di legame.

Se l' i -esimo contatto contribuisce con un'energia pari a ϵ_i , allora il contributo di tutti i contatti all'energia di legame è, in prima approssimazione:

$$\Delta G_{CT}^0 = \sum_i \epsilon_i \quad (2.5)$$

Le forze attrattive tendono a stabilizzare il complesso, dunque il loro contributo sarà negativo. Possono inoltre essere presenti contributi positivi quando, ad esempio, cariche dello stesso segno sono vicine spazialmente oppure alcuni gruppi sono spinti l'uno contro l'altro nonostante la reciproca repulsione sterica, oppure gruppi idrofobici sono in contatto con gruppi idrofilici. L'approssimazione fatta assume che i vari contatti siano indipendenti, ignora possibili distorsioni della struttura dovute all'associazione delle due molecole in un complesso. Tuttavia quando due molecole si legano c'è un forte decremento di entropia, dovuto alla restrizione di movimento dell'una rispetto all'altra, è quindi necessario tenere conto anche di questi contributi strutturali.

2.1 Termodinamica dei processi di associazione

E' possibile scrivere la funzione di partizione della molecola, denotata con Q , come prodotto di contributi di vari termini:

$$Q = q_{ct}q_tq_rq_vq_{cf}q_s \quad (2.6)$$

dove q_{ct} rappresenta il contributo dei contatti intermolecolari già discusso, q_t , q_r e q_v sono rispettivamente il contributo traslazionale, il contributo rotazionale e quello vibrazionale, q_{cf} rappresenta la flessibilità conformazionale e q_s gli effetti di solvatazione. E' possibile scrivere la funzione di partizione come semplice prodotto dei vari termini assumendo che ogni forma di energia sia indipendente dalle altre e additiva. La funzione di partizione dell'insieme è data dal prodotto su tutte le N molecole, diviso per il fattore combinatorio $N!$, dovuto all'indistinguibilità degli stati. L'energia libera può essere espressa in termini della funzione di partizione:

$$G = -kT \ln \left(\frac{Q^N}{N!} \right) \quad (2.7)$$

Per trattare la generica associazione di due molecole A e B nel complesso C è necessario esprimere la funzione di partizione in termini delle funzioni di partizione delle tre specie. Si assume che le molecole nella soluzione non interagiscano con le altre, ad eccezione dell'interazione che si sta studiando. Dal momento che l'energia libera è una quantità additiva è possibile riscrivere l'equazione 2.7:

$$G = -kT \ln \left(\frac{Q_A^{N_A}}{N_A!} \right) - kT \ln \left(\frac{Q_B^{N_B}}{N_B!} \right) - kT \ln \left(\frac{Q_C^{N_C}}{N_C!} \right) = -kT \ln \left(\frac{Q_A^{N_A} Q_B^{N_B} Q_C^{N_C}}{N_A! N_B! N_C!} \right) \quad (2.8)$$

Se il processo di associazione ha raggiunto l'equilibrio, l'energia libera sarà minima. Convertendo una quantità infinitesima di A e B in C, l'energia libera non varia. Utilizzando questo fatto è possibile derivare la seguente condizione di equilibrio:

$$\frac{Q_A Q_B}{Q_C} = \frac{N_A N_B}{N_C} \quad (2.9)$$

Dal momento che la concentrazione è il numero di moli per unità di volume, è possibile convertire il numero di molecole nella concentrazione. A partire dalla concentrazione possiamo ottenere un'espressione per la costante di dissociazione:

$$\frac{Q_C V N_A}{Q_B Q_A} = \frac{[C]}{[A][B]} = \frac{1}{K_{dis}} \quad (2.10)$$

dove \mathcal{N}_A è il numero di Avogadro, che rientra nell'equazione poiché $[A]$ in moli per unità di volume è uguale a $N_A/V\mathcal{N}_A$. Dividendo la funzione di partizione nei vari contributi è possibile riscrivere l'espressione della costante di dissociazione come segue:

$$\frac{1}{K_{dis}} = \left(\frac{q_{C,ct}}{q_{A,ct}q_{B,ct}} \right) \left(\frac{q_{C,t}V\mathcal{N}_A}{q_{A,t}q_{B,t}} \right) \left(\frac{q_{C,r}}{q_{A,r}q_{B,r}} \right) \left(\frac{q_{C,v}}{q_{A,v}q_{B,v}} \right) \left(\frac{q_{C,cf}}{q_{A,cf}q_{B,cf}} \right) \quad (2.11)$$

Calcolando il logaritmo naturale dell'equazione 2.11 e moltiplicando per RT si torna all'espressione di ΔG , che espresso in questi termini è la somma dei vari contributi all'energia.

E' possibile ora studiare separatamente ogni contributo e, per fare questo, è utile pensare in termini di numero di gradi di libertà. Ogni molecola possiede tre gradi di libertà traslazionali, corrispondenti al movimento nelle direzioni x , y e z . Fatta eccezione per le molecole che presentano particolari simmetrie, possiedono anche tre gradi di libertà rotazionali, che corrispondono alle rotazioni attorno agli assi x , y e z . Due molecole separate possiedono dunque 12 gradi di libertà totali, ma pensando alla molecola che compongono associandosi, questa sembra possedere solo 6 gradi di libertà. In realtà i 6 gradi di libertà apparentemente persi corrispondono a sei nuovi gradi di libertà vibrazionali e librazionali interni.

Nello studio della dinamica del complesso preso in esame in questo lavoro di tesi viene considerato esclusivamente il contributo vibrazionale, che verrà esaminato in dettaglio.

2.2 Energia libera vibrazionale

L'associazione di due molecole in un complesso rimuove alcuni gradi di libertà rotazionali e vibrazionali, dunque questi contributi saranno in generale positivi: la perdita di gradi di libertà si oppone alla formazione del complesso. Tuttavia questi modi vengono convertiti in modi vibrazionali, che ci si aspetta compensino con un contributo negativo all'energia libera di associazione. I nuovi modi vibrazionali derivano da allungamenti e distorsioni dei contatti che tengono assieme le molecole. Il fatto che ci siano proprio sei nuovi modi è peculiare di un'associazione binaria ed è completamente indipendente dal numero di contatti.

Le vibrazioni all'interno del complesso vengono decomposte in modi normali di oscillazione con il comportamento di oscillatori armonici indipendenti. L'energia viene calcolata come somma di termini separati rappresentanti ciascuno un modo normale. La funzione di partizione vibrazionale è il prodotto di ogni termine, che per un oscillatore armonico

ha la seguente forma:

$$q_{vi} = \frac{e^{-h\nu_i/2kT}}{1 - e^{-h\nu_i/kT}} \quad (2.12)$$

dove $\nu_i = (1/2\pi)\sqrt{(\phi_i/\mu_i)}$ è la frequenza di vibrazione, con μ_i massa ridotta e ϕ_i è la costante di forza di richiamo, legata alla matrice Hessiana dell'energia potenziale. E' possibile semplificare l'equazione 2.12 notando che i legami non covalenti sono molto più deboli di quelli covalenti, quindi la costante di forza di legame è minore. Le frequenze di vibrazione per i legami non-covalenti sono piccole e dunque $h\nu \ll kT$. Espandendo l'esponenziale $e^{-h\nu/kT} \sim 1 - h\nu/kT$ l'equazione 2.12 può essere riscritta:

$$q_{vi} = \frac{kT}{h\nu_i} = \frac{2\pi kT}{h} \sqrt{\frac{\mu_i}{\phi_i}} \quad (2.13)$$

Questo è il limite classico per un oscillatore armonico in cui l'energia cinetica media per mole è $\frac{1}{2}RT$, in cui la perdita di energia cinetica traslazionale è completamente recuperata dai moti vibrazionali se i contatti sono abbastanza deboli da produrre vibrazioni classiche. Nella formazione del complesso tutti i modi normali cambiano, poiché sono funzione dell'energia potenziale totale della molecola. Per un'analisi quantitativa è necessario conoscere in dettaglio la funzione energia potenziale.

2.2.1 Approssimazione armonica classica

Studiamo ora in dettaglio come si ottiene l'equazione 2.13 nel caso classico. E' possibile scrivere la lagrangiana:

$$\mathcal{L} = \mathcal{E}_k - \mathcal{E}_p = \frac{1}{2}\dot{\mathbf{q}}^T \mathbf{M} \dot{\mathbf{q}} - \frac{1}{2}\mathbf{q}^T \mathbf{H} \mathbf{q} \quad (2.14)$$

dove \mathbf{q} sono le coordinate generalizzate, $\dot{\mathbf{q}}$ le velocità generalizzate, \mathbf{M} è la matrice delle masse di componenti

$$M_{ij} = \sum_l m_l \frac{\partial \mathbf{r}_l}{\partial q_i} \cdot \frac{\partial \mathbf{r}_l}{\partial q_j}$$

dove la somma corre su tutte le componenti del sistema con massa m_l e coordinate cartesiane \mathbf{r}_l . Infine \mathbf{H} è la matrice Hessiana nelle vicinanze di un punto di equilibrio stabile, per semplicità $\mathbf{q} = 0$, di componenti:

$$H_{ij} = \left. \frac{\partial^2 E_p}{\partial q_i \partial q_j} \right|_{\mathbf{q}=0}$$

Diagonalizzando solo la matrice Hessiana non è possibile ottenere i corretti modi nor-

mali, eccezion fatta per il caso in cui tutte le masse sono uguali, ma   possibile stimare solamente la parte configurazionale della funzione di partizione. La tecnica corretta per ottenere i modi normali di oscillazione   diagonalizzare contemporaneamente entrambe le forme quadratiche \mathbf{M} e \mathbf{H} , cui consegue la ricerca degli autovalori λ_i , collegati alle frequenze normali di oscillazione dalla relazione $\lambda_i = \omega_i^2$ e soluzioni dell'equazione caratteristica:

$$\det|\mathbf{H} - \lambda\mathbf{M}| = 0 \quad (2.15)$$

I modi normali vengono quindi definiti dalle colonne della matrice \mathbf{U} , soluzione dell'equazione agli autovalori:

$$\mathbf{H}\mathbf{M} = \mathbf{M}\mathbf{U}\Lambda \quad (2.16)$$

e sono normalizzati in accordo con la seguente equazione:

$$\mathbf{U}^T\mathbf{M}\mathbf{U} = \mathbb{I} \quad (2.17)$$

Queste relazioni permettono di definire un cambio di coordinate $\mathbf{q} = \mathbf{U}\theta$ che diagonalizza sia l'energia potenziale che quella cinetica. La lagrangiana in queste coordinate  :

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}\dot{\theta}^T\mathbf{U}^T\mathbf{M}\mathbf{U}\dot{\theta} - \frac{1}{2}\theta^T\mathbf{U}^T\mathbf{H}\mathbf{U}\theta \\ &= \frac{1}{2}\dot{\theta}^T\mathbb{I}\dot{\theta} - \frac{1}{2}\theta^T\mathbf{U}^T\mathbf{M}\mathbf{U}\Lambda\theta \\ &= \frac{1}{2}\dot{\theta}^T\mathbb{I}\dot{\theta} - \frac{1}{2}\theta^T\Lambda\theta \end{aligned} \quad (2.18)$$

Questo corrisponde ad un insieme di oscillatori armonici disaccoppiati con frequenza di oscillazione ω_i , che seguono l'equazione del moto $\ddot{\theta}_i = -\lambda_i\theta_i = -\omega_i^2\theta_i$. A questo punto pu  essere calcolata la funzione di partizione, che ora tiene conto del termine cinetico e di quello configurazionale:

$$Q = \prod_i \frac{2\pi kT}{h\omega_i} \quad (2.19)$$

Da questa   possibile ricavare l'energia libera (2.20) e, di conseguenza, energia interna (2.21) ed entropia (2.22) (si   sostituito $\frac{1}{k_B T} = \beta$)

$$\beta G = -\ln Q = N \ln \left[\frac{\beta h}{2\pi} \right] + \sum_i \ln \omega_i \quad (2.20)$$

$$u = \frac{\partial(\beta G)}{\partial \beta} = \frac{N}{\beta} = N k_B T \quad (2.21)$$

$$\frac{s}{k_B} = \beta(u - G) = - \sum_i \ln \omega_i - N \ln \left[\frac{\beta h}{2e\pi} \right] \quad (2.22)$$

Dove $N = \sum_i 1$ è il numero totale di gradi di libertà. Mentre \mathbf{U} e ω_i dipendono dalle masse, Q , G , u ed s non dipendono dalle masse, essendo calcolate all'equilibrio termodinamico.

Modelli di rete elastica

Un metodo efficiente per studiare la dinamica delle proteine è l'analisi dei modi normali (NMA), un metodo per caratterizzare le fluttuazioni molecolari attorno a uno stato di equilibrio. Invece di utilizzare energie potenziali che siano funzioni complicate delle coordinate degli atomi delle proteine, conviene utilizzare un modello semplificato, a grana grossa, basato sulla conoscenza della struttura nativa. Quest'ultima diventa così per definizione lo stato di equilibrio del sistema. Utilizziamo il Gaussian Network Model, che è un modello di rete elastica in cui i nodi corrispondono alle coordinate dei C_α e sono collegati tra loro da molle armoniche. Due residui si considerano legati se la loro distanza è minore di un dato raggio di interazione, chiamato distanza di cutoff. Identificando la posizione di equilibrio del nodo i -esimo attraverso il vettore \mathbf{R}_i^0 e la sua posizione istantanea con il vettore \mathbf{R}_i , le fluttuazioni sono definite dal vettore $\Delta\mathbf{R}_i = \mathbf{R}_i - \mathbf{R}_i^0$. Parimenti le fluttuazioni del vettore distanza tra due residui \mathbf{R}_{ij} possono essere espresse da $\Delta\mathbf{R}_{ij} = \mathbf{R}_{ij} - \mathbf{R}_{ij}^0 = \Delta\mathbf{R}_j - \Delta\mathbf{R}_i$. Assumendo che queste fluttuazioni siano isotropiche e Gaussiane è possibile scrivere il potenziale di una rete con N nodi in termini delle componenti ΔX_i , ΔY_i e ΔZ_i come

$$V_{GNM} = \frac{\gamma}{2} \left[\sum_{i,j}^N \Gamma_{ij} [(\Delta X_i - \Delta X_j)^2 + (\Delta Y_i - \Delta Y_j)^2 + (\Delta Z_i - \Delta Z_j)^2] \right] \quad (3.1)$$

dove Γ_{ij} è l'elemento i, j della matrice di adiacenza che individua quali residui sono in contatto, definita da:

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \text{ e } R_{ij} \leq r_C \\ 0, & \text{if } i \neq j \text{ e } R_{ij} \geq r_C \\ -\sum_{j,j \neq i} \Gamma_{ij}, & \text{if } i = j \end{cases} \quad (3.2)$$

con γ costante elastica delle molle, ipotizzata costante nella rete. Esprimendo le componenti x, y e z dei vettori delle fluttuazioni $\Delta \mathbf{R}_i$ come tre vettori N-dimensionali $\Delta \mathbf{X}$, $\Delta \mathbf{Y}$ e $\Delta \mathbf{Z}$, l'equazione 3.1 può essere riscritta come segue (dove \mathbf{E} è la matrice identità di ordine 3):

$$V_{GNM} = \frac{\gamma}{2} [\Delta \mathbf{X}^T \Gamma \Delta \mathbf{X} + \Delta \mathbf{Y}^T \Gamma \Delta \mathbf{Y} + \Delta \mathbf{Z}^T \Gamma \Delta \mathbf{Z}] = \frac{\gamma}{2} [\Delta \mathbf{R}^T (\Gamma \otimes \mathbf{E}) \Delta \mathbf{R}] \quad (3.3)$$

L'energia potenziale appena ricavata è una approssimazione armonica dell'energia potenziale della proteina, valida in un intorno di uno stato di equilibrio.

Le fluttuazioni quadratiche medie dei residui sono sperimentalmente misurabili tramite diffusione a raggi X di cristalli di proteine^[2] e sono spesso utilizzate per verificare e migliorare modelli e metodi computazionali. In particolare le fluttuazioni predette dal GNM sono in generale in buon accordo con i B-factor misurati sperimentalmente. I B-factor sono collegati alle fluttuazioni dei residui dall'equazione

$$B_i = \frac{8\pi^2}{3} \langle (\Delta \mathbf{R}_i^2) \rangle \quad (3.4)$$

E' di primario interesse la determinazione delle fluttuazioni quadratiche medie di un particolare residuo i data da

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle = \langle \Delta X_i^2 \rangle + \langle \Delta Y_i^2 \rangle + \langle \Delta Z_i^2 \rangle \quad (3.5)$$

Dal momento che nel GNM la distribuzione di probabilità di tutte le fluttuazioni è isotropica Gaussiana, è possibile dimostrare che valgono le relazioni:

$$\langle \Delta \mathbf{X}_i \Delta \mathbf{X}_i^T \rangle = \langle \Delta \mathbf{Y}_i \Delta \mathbf{Y}_i^T \rangle = \langle \Delta \mathbf{Z}_i \Delta \mathbf{Z}_i^T \rangle = \frac{k_B T}{\gamma} [\Gamma^{-1}]_{ij} \quad (3.6)$$

$$\langle \Delta \mathbf{X}_i \Delta \mathbf{X}_i^T \rangle = \langle \Delta \mathbf{Y}_i \Delta \mathbf{Y}_i^T \rangle = \langle \Delta \mathbf{Z}_i \Delta \mathbf{Z}_i^T \rangle = \frac{1}{3} \langle \Delta \mathbf{R}_i \Delta \mathbf{R}_i^T \rangle \quad (3.7)$$

Dove k_B è la costante di Boltzmann, T la temperatura assoluta e γ la costante elastica

delle molle nella rete. Otteniamo quindi

$$\begin{aligned} \langle \Delta \mathbf{R}_i^2 \rangle &= \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ii} \\ \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle &= \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ij} \end{aligned} \quad (3.8)$$

Attraverso l'ipotesi che le fluttuazioni siano isotropiche è dunque possibile ridurre un problema 3N-dimensionale a uno N-dimensionale descritto dall'equazione 3.8.

A questo punto possibile riscrivere la relazione 3.4 in termini della matrice di adiacenza:

$$B_i = \frac{8\pi^2}{3} \langle (\Delta \mathbf{R}_i^2) \rangle = \frac{8\pi^2 k_B T}{\gamma} [\Gamma^{-1}]_{ii} \quad (3.9)$$

A livello pratico per ricavare l'elemento di matrice $[\Gamma^{-1}]_{ii}$ è necessario riscrivere la matrice Γ come prodotto delle matrici degli autovettori \mathbf{u}_i di Γ , della matrice diagonale Λ degli autovalori λ_i e della trasposta di \mathbf{U} :

$$\Gamma = \mathbf{U} \Lambda \mathbf{U}^T \quad (3.10)$$

Gli autovalori λ_i rappresentano le frequenze dei vari modi di oscillazione, mentre gli autovettori definiscono la forma di questi modi. Il primo autovalore λ_1 è identicamente nullo e il corrispettivo autovalore \mathbf{u}_1 è composto di elementi costanti pari a $1/\sqrt{N}$ ed è indicativo di un'assenza di moto interno. Questo riflette il fatto che la macromolecola può traslare rigidamente senza una variazione di energia potenziale. Combinando l'equazione 3.10 con la 3.8 si ottiene la seguente espressione per la correlazione tra le fluttuazioni (dove la somma corre sugli $N - 1$ modi non nulli):

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \frac{3k_B T}{\gamma} \sum_k [\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ij} \quad (3.11)$$

Questo permette di identificare il contributo del k-esimo modo, data da

$$[\Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j] = \frac{3k_B T}{\gamma} \lambda_k^{-1} [\mathbf{u}_k]_i [\mathbf{u}_k]_j \quad (3.12)$$

Dunque gli elementi $[\mathbf{u}_k]_i$ riflettono le mobilità residue del k-esimo modo. Dal momento che λ_k^{-1} funge da peso statistico che riscalda il contributo del k-esimo modo, questo garantisce che il modo più lento porti il contributo maggiore. Inoltre i modi più lenti hanno in generale il più alto grado di collettività: questi modi producono sostanziali cambiamenti conformazionali. Al contrario l'alta frequenza dei modi veloci non porta a cambiamenti

sostanziali. Nel GNM da noi considerato le masse dei residui sono ipotizzate essere uguali fra loro. I risultati ottenuti nel seguito, che coinvolgono quantità di equilibrio, valgono quindi anche nel caso più generale di masse diverse per ogni residuo.

Stima dell'entropia vibrazionale per complessi di proteine

Scopo di questo lavoro di tesi è capire quanto accuratamente un modello semplice come la rete elastica possa riprodurre la fluttuazioni quadratiche medie attorno alla posizione di equilibrio per complessi di proteine.

La procedura utilizzata permette poi di determinare la costante elastica γ della molla introdotta nella rete elastica e di utilizzarne il valore per stimare il contributo entropico vibrazionale all'energia libera di legame. Nel seguito identificheremo i B-factor sperimentali con quelli calcolati a partire da traiettorie di dinamica molecolare disponibili nel gruppo di ricerca in cui è stata svolta questa tesi. E' preferibile usare questi ultimi, invece dei B-factor disponibili da esperimenti di diffrazione di raggi X su cristalli, poiché le fluttuazioni di una proteina in soluzione acquosa possono essere molto diverse da quelle misurate in un cristallo^[3].

Le traiettorie di dinamica molecolare utilizzate corrispondono a 20 *ns* di simulazione effettuata con una energia potenziale classica che è una funzione complicata delle coordinate di tutti gli atomi pesanti delle proteine e delle molecole d'acqua della soluzione^[4].

In particolare è stato preso in esame il complesso 1BVK, un anticorpo monoclonale. Il complesso è composto da due subunità 1BVL (Humanized anti-lysozime) e 3LZT (Lysozime). Inizialmente è stato analizzata l'intera struttura e se ne è studiata la mobilità tramite GNM. Successivamente è stata ripetuta l'analisi per le due subunità separatamente.

4.1 Complesso 1BVK

4.1.1 Analisi preliminare

Attraverso il Gaussian Network Model è possibile calcolare le fluttuazioni teoriche dei vari residui, che risultano, con buona approssimazione, proporzionali ai B-factor sperimentali (la costante di proporzionalità viene determinata tramite il metodo dei minimi quadrati). E' quindi stata definita una GNM a partire dalle coordinate dei C_α delle catene D, E e F del complesso, costituite da 108, 116 e 129 residui, rispettivamente. L'informazione su quali siano le catene interessanti biologicamente e quali le subunità isolate è ricercata da Janin [5].

Per impostare un buon valore di soglia si è cercato di minimizzare l'RMSE (Root Mean Square Error) tra dati teorici e dati sperimentali, calcolandolo per valori di soglia tra 6Å e 16Å (con una griglia di 0.05Å).

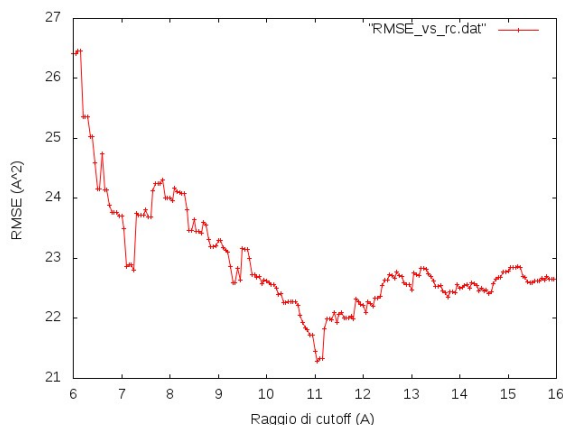


Figura 4.1: RMSE al variare della soglia per il complesso 1BVK

Come si può vedere dalla figura 4.1 l'RMSE decresce fino ad un minimo molto pronunciato per il valore di soglia di 11.05Å superato il quale ritorna ad aumentare e si stabilizza (con qualche oscillazione) per un valore del raggio di cutoff superiore ai 13Å. Andando a studiare parallelamente il coefficiente di correlazione di Pearson, che è un altro indicatore della bontà della calibrazione delle mobilità, si può notare un grafico speculare con il massimo di correlazione sempre a 11.05Å (fig. 4.2).

Si è pertanto deciso di impostare come valore di soglia 11.05 Å e una volta calcolate le mobilità le si è poste a confronto con i valori sperimentali (fig 4.3). Come si può osservare dal grafico l'andamento delle mobilità teoriche si sovrappone piuttosto bene ai dati sperimentali, nonostante ciò si possono notare due importanti incongruenze:

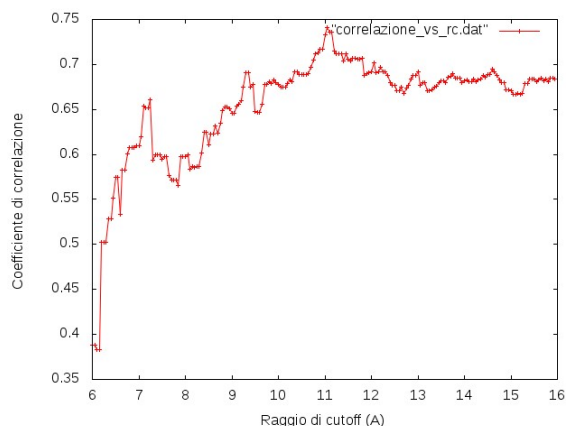


Figura 4.2: Coefficiente di correlazione al variare della soglia per il complesso 1BVK

1. i picchi principali della curva sperimentale sono molto più pronunciati di quelli ottenuti dal modello teorico
2. la baseline del modello teorico è generalmente più alta di quella sperimentale

Queste differenze sono dovute al fatto che il modello utilizzato è lineare, mentre alcuni residui oscillano in maniera non-lineare e quindi le loro fluttuazioni sono più ampie. Nel calcolo della costante di proporzionalità tramite il metodo dei minimi quadrati non si tiene conto di questo fatto e dunque per ottenere picchi più pronunciati si aumenta la costante di proporzionalità ma a questo consegue anche una crescita delle fluttuazioni più piccole, che risultano mediamente più alte di quelle teoriche.

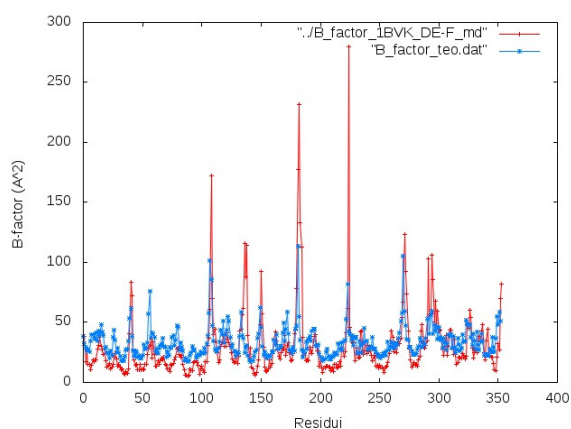


Figura 4.3: Confronto tra i B-factor sperimentali (curva rossa) e B-factor calcolati tramite GNM (curva blu)

4.1.2 Calibrazione delle mobilità

Per ovviare a questi problemi, si è utilizzata una procedura di eliminazione sistematica dei residui a partire da quello che presenta una maggiore differenza tra valore teorico e valore sperimentale.

Durante la procedura si monitora come variano RMSE e coefficiente di correlazione eliminando gradualmente i vari residui.

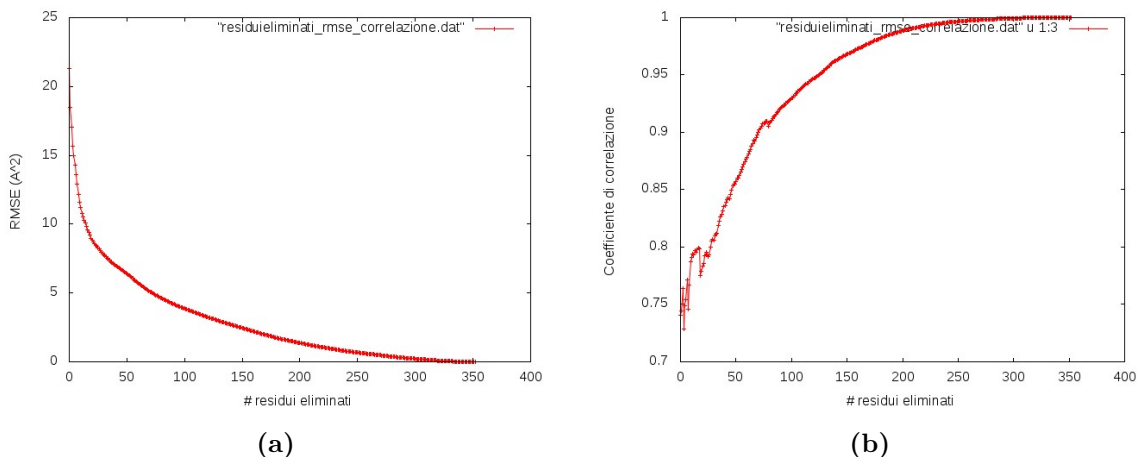


Figura 4.4: RMSE e coefficiente di correlazione al variare del numero di residui esclusi per il complesso 1BVK

Come ci si aspetta l'RMSE diminuisce molto rapidamente eliminando i primi residui, quelli che si discostano maggiormente dai valori sperimentali, e tende a zero, raggiungendolo ovviamente quando rimane un solo residuo.

Parimenti il coefficiente di correlazione cresce all'aumentare dei residui esclusi raggiungendo l'unità quando rimangono solo una decina di residui.

Un'altra quantità importante da monitorare durante questo processo è la costante di proporzionalità tra B-factor teorici e sperimentali ottenuta tramite il metodo dei minimi quadrati. Questa costante vale infatti $\frac{8\pi^2 k_B T}{\gamma}$, dove γ è la costante elastica delle molle nella rete. Se non viene eliminato alcun residuo, per tentare di riprodurre nel modello le fluttuazioni più ampie sarà necessario aumentare la costante di proporzionalità e quindi impostare una costante elastica più piccola, essendo queste ultime inversamente proporzionali. Questo è fisicamente ragionevole: per riprodurre fluttuazioni ampie è necessaria una molla poco rigida.

Quello che ci si aspetta dalla procedura di esclusione dei residui è che γ aumenti insieme al numero di residui esclusi: se non è più necessario riprodurre grosse fluttuazioni è possibile avere molle più rigide. Come si può vedere dalla figura 4.5 l'andamento atteso è rispettato.

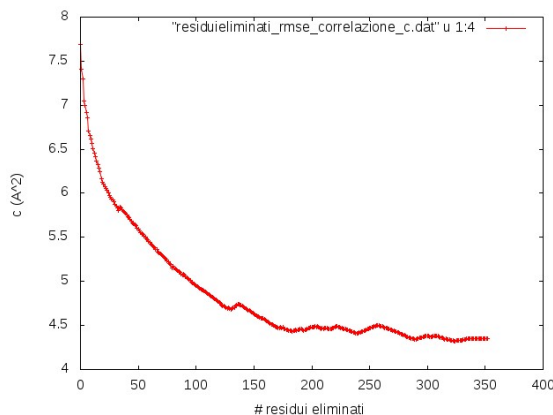


Figura 4.5: Costante di proporzionalità al variare del numero di residui esclusi per il complesso 1BVK

A questo punto si è deciso di eseguire la procedura di esclusione dei residui fino ad avere un valore di RMSE prefissato pari a 4Å^2 . E' possibile utilizzare come parametro per valutare la bontà di questo processo il numero di residui esclusi per ottenere questo valore di RMSE. Osservando come varia il numero di residui esclusi al variare della soglia si può notare che non è più presente il minimo a 11.05Å . E' invece presente un minimo non molto pronunciato a 14Å , mentre per valori di soglia superiori a 15Å il numero di residui esclusi si stabilizza.

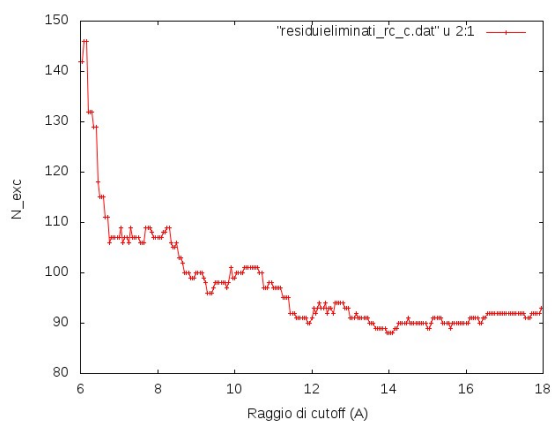


Figura 4.6: Numero di residui esclusi al variare della soglia per il complesso 1BVK

Durante questa procedura sono stati monitorati anche il coefficiente di correlazione e la costante di proporzionalità tra B-factor sperimentali e teorici al variare della soglia. Come ci si poteva aspettare quest'ultima cresce all'aumentare della soglia (fig 4.7 a), infatti aumentando il raggio di cutoff si inseriscono nella rete elastica un numero minore

di molle e dunque se si vogliono riprodurre le stesse mobilità con un numero inferiore di molle è necessario diminuirne la rigidità.

Per quanto riguarda il coefficiente di correlazione (fig. 4.7 b) oscilla fra valori compresi tra 0.9 e 0.94, ma non segue un andamento preciso e in particolare non presenta un massimo al valore di soglia fissato precedentemente (11.05Å). Come ci si aspetta i dati teorici sono in buon accordo con quelli sperimentali (il coefficiente di correlazione è prossimo all'unità). In figura 4.8 sono state poste a confronto le mobilità sperimentali e

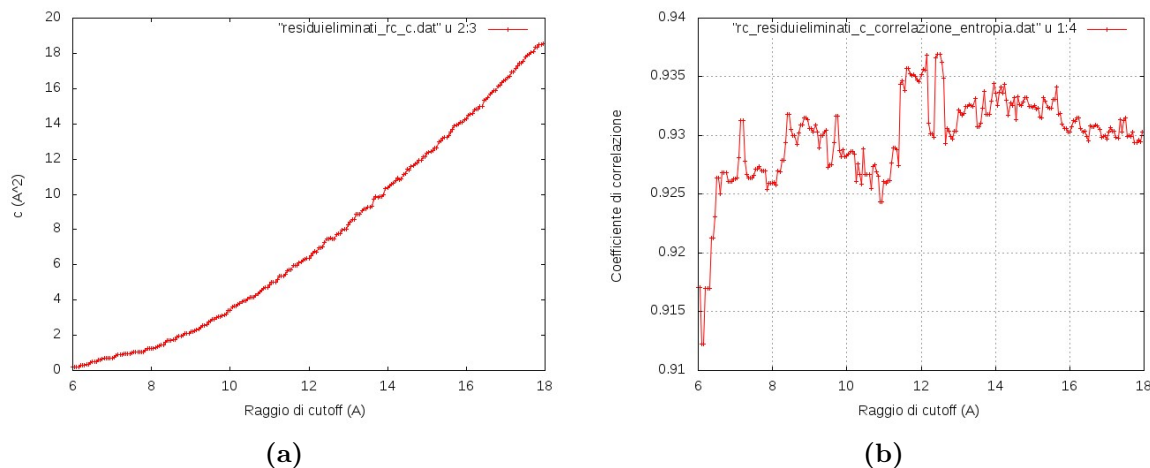


Figura 4.7: Costante di proporzionalità tra B-factor sperimentali e teorici e coefficiente di correlazione al variare del raggio di cutoff

quelle teoriche prima e dopo la procedura di esclusione dei residui che porta ad un valore di RMSE pari a 4Å^2 . Si nota un evidente abbassamento della baseline, che ora combacia con quella attesa sperimentalmente.

Per valutare la regolarità di questa procedura di esclusione si è anche deciso di controllare se i residui eliminati sono sempre gli stessi al variare della soglia. A titolo di esempio riportiamo il grafico dei residui esclusi per un valore del raggio di cutoff pari a 14Å e 16Å, dove l'ordinata rappresenta l'indice di residuo escluso e vari residui sono stati disposti in ordine crescente (fig 4.9).

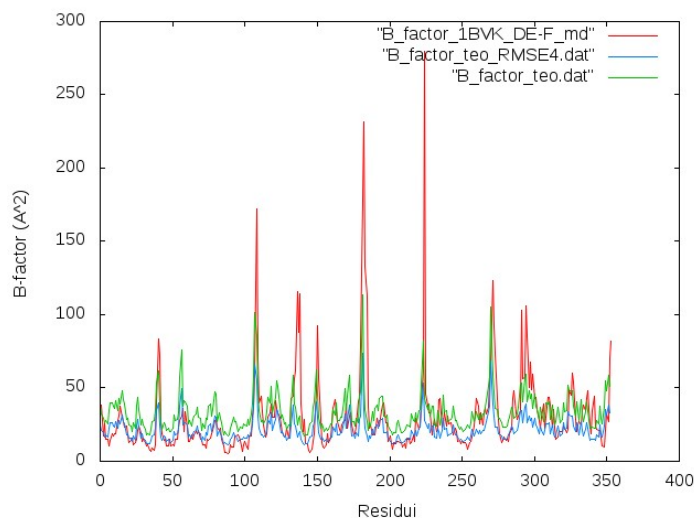


Figura 4.8: Confronto tra B-factor sperimentali (curva rossa) e B-factor calcolati tramite GNM prima della procedura di esclusione (curva verde) e dopo la procedura (curva blu)

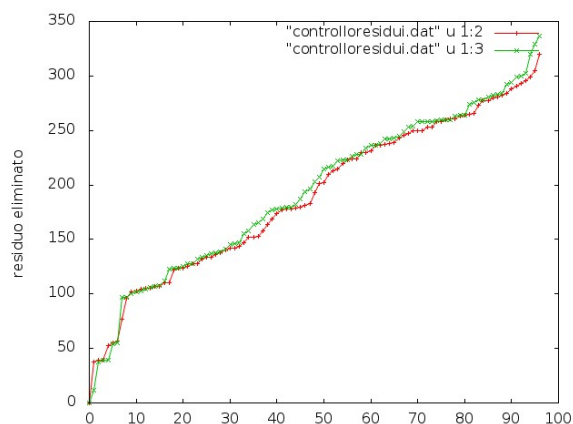


Figura 4.9: Residui eliminati durante la procedura di esclusione per valori della soglia pari a 14\AA (in rosso) e 16\AA (in verde)

4.2 Subunità 1BVL e 3LZT

E' stata eseguita la stessa analisi svolta per il complesso anche per le due subunità separatamente. La subunità 1BVL consta di due catene, di 116 e 108 residui, mentre la subunità 3LZT consta di una catena di 129 residui. Si è quindi eseguito un ciclo sui raggi di cutoff per minimizzare l'RMSE e scegliere il valore di soglia ottimale.

Come si può vedere dai grafici, per la subunità 1BVL il raggio di cutoff presenta un minimo per un valore di 12.7\AA in corrispondenza del quale è presente anche un massimo locale del valore del coefficiente di correlazione. Per quanto riguarda la subunità

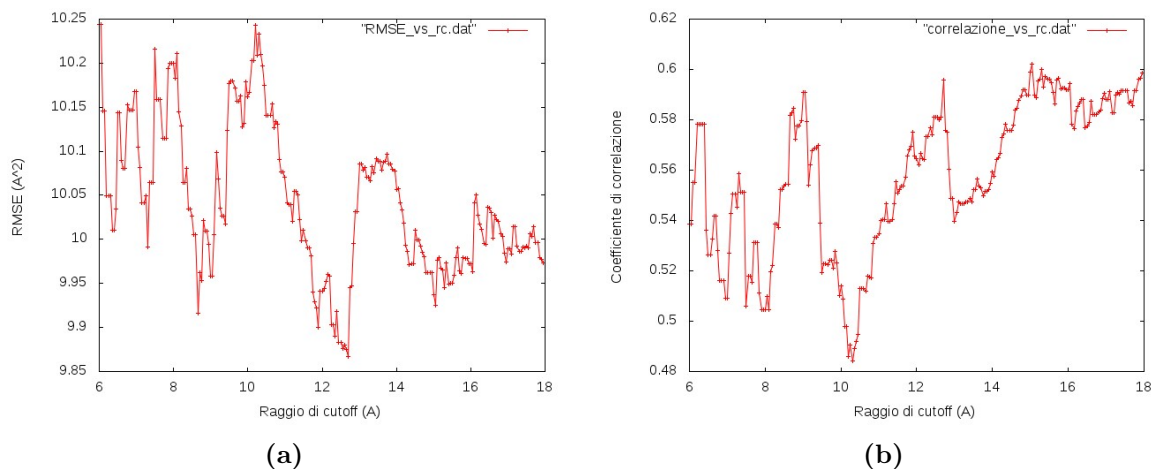


Figura 4.10: RMSE e coefficiente di correlazione al variare della soglia per la subunità 1BVL

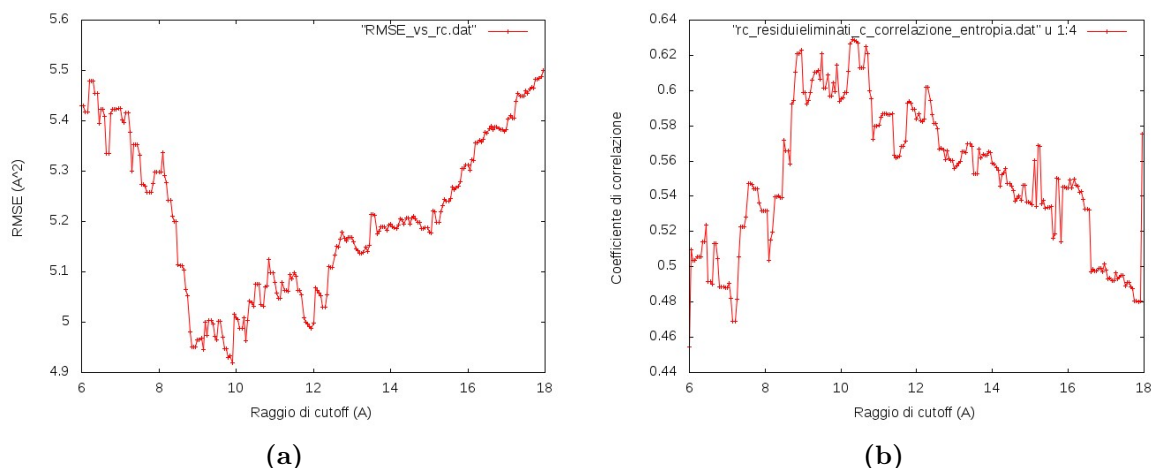


Figura 4.11: RMSE e coefficiente di correlazione al variare della soglia per la subunità 3LZT

3LZT si ha una situazione analoga in corrispondenza del valore di 9.9\AA . Si sono quindi impostati come valori di soglia rispettivamente 12.7\AA e 9.9\AA e sono state calcolate le fluttuazioni teoriche. Ponendole a confronto con quelle sperimentali riscontriamo le stesse problematiche individuate precedentemente.

Si procede dunque con la calibrazione delle mobilità, interrompendo la procedura di esclusione dei residui sempre ad un valore di RMSE pari a 4\AA^2 . Essendo in questo caso le mobilità molto meno ampie rispetto a quelle del complesso, questo valore di RMSE si ottiene eliminando un numero molto minore di residui. In particolare il numero dei residui esclusi rimane praticamente costante al variare del valore di soglia.

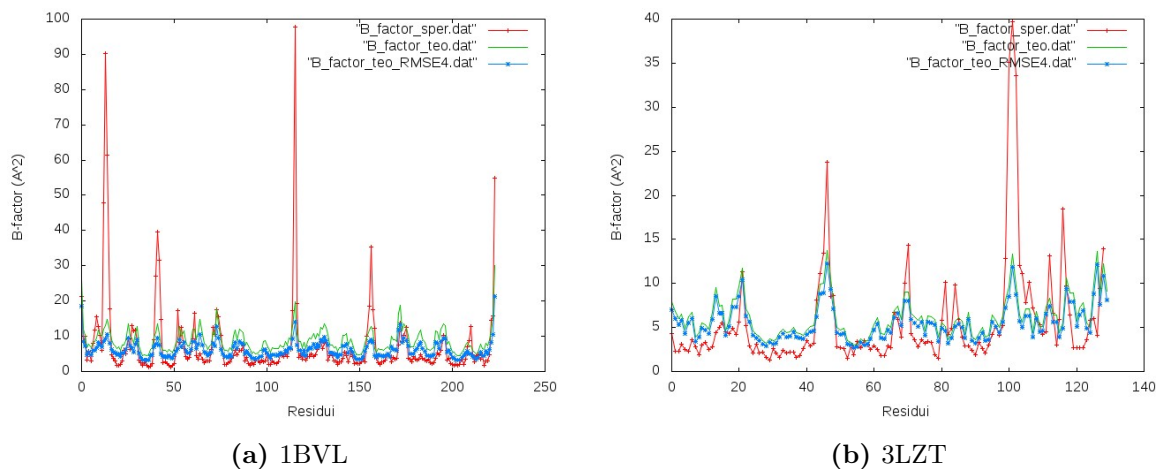


Figura 4.12: Confronto tra B-factor sperimentali (curva rossa) e B-factor calcolati tramite GNM prima della procedura di esclusione (curva verde) e dopo la procedura (curva blu)

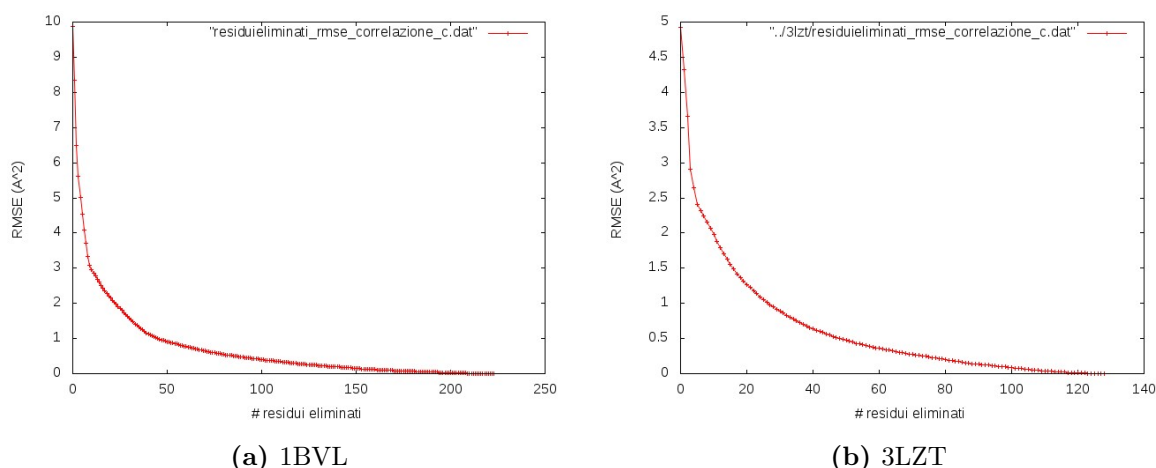


Figura 4.13: RMSE al variare del numero di residui eliminati per le due subunità

4.3 Contributo vibrazionale all'energia libera di legame

La procedura di calibrazione delle mobilità permette di stimare le affinità di legame dei complessi. In particolare è possibile stimare il contributo entropico dovuto alla parte vibrazionale. Nel seguito l'entropia vibrazionale viene stimata in unità adimensionali (s/k_B) e a meno di una costante additiva. Con l'approssimazione di rete elastica l'entropia vibrazionale risulta $S \propto \sum \ln \left(\frac{\lambda}{c} \right)$, con λ autovalori della matrice di adiacenza e c costante di proporzionalità tra i B-factor teorici e quelli sperimentali. E' stata quindi calcolata questa

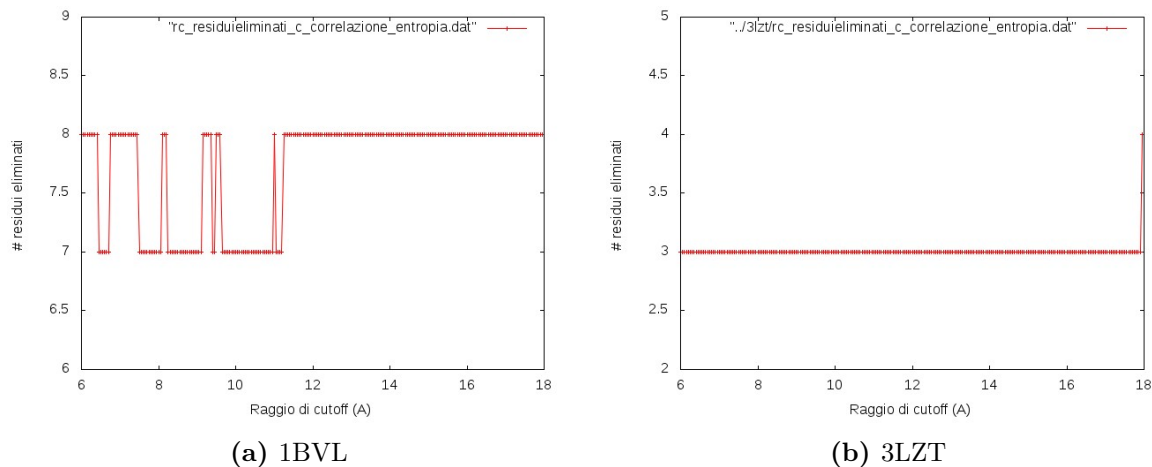


Figura 4.14: Numero di residui eliminati al variare della soglia per le due subunità

quantità al variare della distanza di soglia, se la procedura di esclusione si interrompe ad una RMSE prefissata di 4\AA^2 .

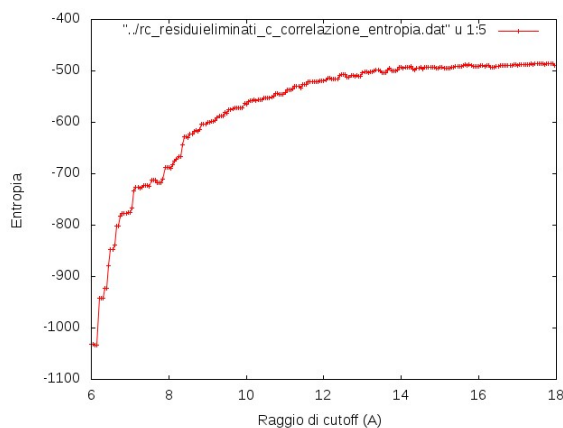


Figura 4.15: Entropia vibrazionale per il complesso 1BVK

Dal momento che ci si aspetta di riprodurre sempre le stesse fluttuazioni, l'entropia dovrebbe rimanere costante durante questo processo. Dai risultati ottenuti per il complesso 1BVK (fig 4.15) si nota invece che si arriva ad un valore costante solamente per valori di soglia sopra i 13\AA .

Questa procedura di esclusione dei residui porta quindi a risultati stabili per valori di soglia superiori ai $13\text{-}14\text{\AA}$. E' stata inoltre ripetuta la procedura di esclusione usando diversi valori di RMSE per interromperla (fig 4.16). Si può notare che per un valore di RMSE più alto la stima dell'entropia è più bassa, infatti i residui sono meno mobili e di conseguenza la costante di proporzionalità c più alta.

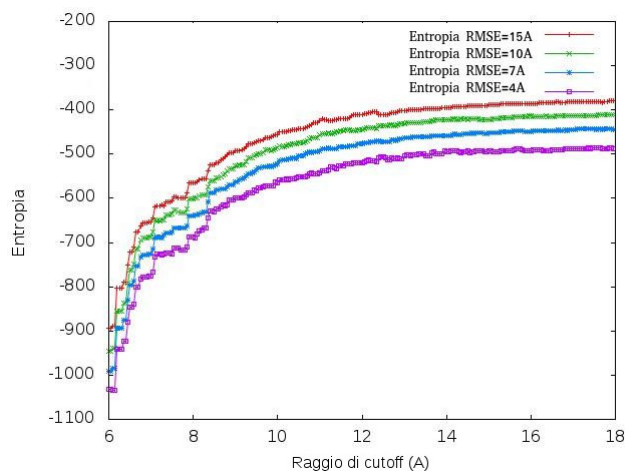


Figura 4.16: Stima dell'entropia vibrazionale del complesso interrompendo la procedura di esclusione per vari valori di RMSE

A questo punto andando a studiare le due subunità separatamente è possibile stimare la differenza di entropia fra il complesso unito AB e le sue subunità isolate: $\Delta S = S_{AB} - (S_A + S_B)$. Anche per le due subunità isolate, 1BVL e 3LZT, è stata stimata la differenza di entropia interrompendo la procedura di esclusione per vari valori di RMSE. In questo caso sono stati inclusi valori di RMSE più bassi, in quanto l'RMSE per la subunità 3LZT non supera mai il valore di 6\AA^2 essendo le sue fluttuazioni piuttosto piccole. Le stime di S_A e S_B (fig. 4.17) confermano la saturazione ad un valore costante per un valore di soglia sopra i 13\AA .

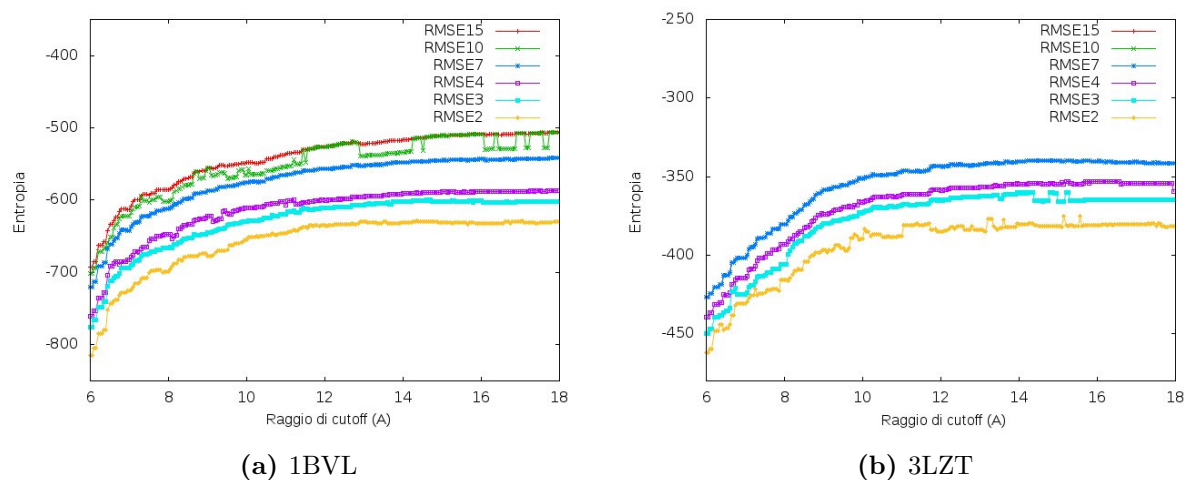


Figura 4.17: Stima dell'entropia vibrazionale delle subunità interrompendo la procedura di esclusione per vari valori di RMSE

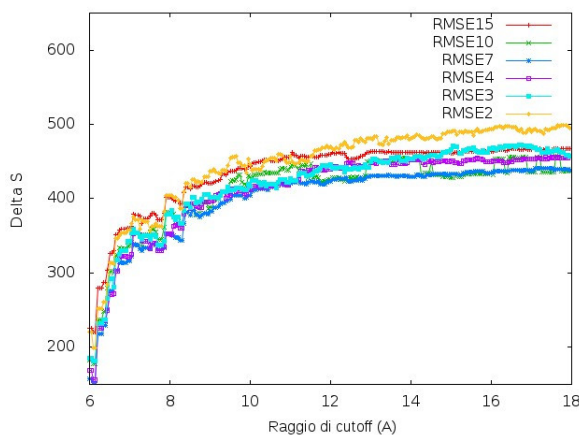


Figura 4.18: $\Delta S = S_{AB} - (S_A + S_B)$

In particolare confrontando le fluttuazioni delle due subunità (fig. 4.12) con quelle del complesso 4.3, si può notare che queste ultime sono molto più ampie. Dunque la formazione del complesso è penalizzata da un aumento di entropia vibrazionale sostanziale, come si può vedere in figura 4.18, dove si mostra la stima di ΔS ottenuta al variare della soglia per diversi valori di RMSE utilizzati nella procedura di esclusione. Da questi grafici emerge inoltre che per le due subunità sono presenti alcuni salti di entropia, che corrispondono alla variazione di un unità nel numero di residui esclusi. Questo porta a pensare che nel calcolo dell'entropia con questa procedura sia più rilevante il numero di residui esclusi rispetto al valore di RMSE in cui interrompere la procedura stessa. Si nota inoltre un andamento non monotono in funzione di RMSE in figura 4.18. La differenza di entropia presenta un minimo per valori intermedi di RMSE ($7-10\text{\AA}^2$). Per evidenziare questo fatto in figura 4.19 è stato graficato il valore di ΔS per un valore arbitrario di soglia (in questo caso 16.25\AA) in funzione del valore di RMSE.

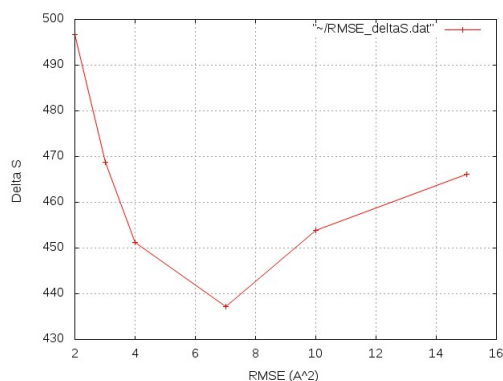


Figura 4.19: ΔS in funzione di RMSE per un valore di soglia pari a 16.25\AA

Conclusioni

Lo studio dei modi normali di oscillazione tramite un modello semplice come il Gaussian Network Model porta a risultati soddisfacenti e che ben riproducono le fluttuazioni ottenute in traiettorie di dinamica molecolare, molto più costose dal punto di vista computazionale. In particolare attraverso la procedura di esclusione dei residui che presentano fluttuazioni ampie e presumibilmente non-lineari è stato possibile ottimizzare il modello ottenendo un notevole abbassamento della baseline come previsto dai dati sperimentali.

Dallo studio delle due subunità separate emerge il fatto che le fluttuazioni per il complesso sono molto più ampie, rispetto a quelle delle due unità separate. Come si può notare dal calcolo dell'entropia, la formazione del complesso è penalizzata da un aumento di entropia sostanziale.

Il calcolo dell'entropia per il complesso e per le due subunità porta inoltre a concludere che il processo di calibrazione delle mobilità per mezzo dell'esclusione dei residui dà risultati stabili solo oltre un determinato valore della soglia di distanza utilizzata per definire le molle della rete elastica, superato il quale l'entropia assume un valore costante, come ci si aspetta dal momento che si devono riprodurre sempre le stesse fluttuazioni. E' stato inoltre caratterizzato come la stima della differenza di entropia dovuta alla formazione del complesso dipenda dal valore di RMSE fra le mobilità per cui si decide di porre fine alla procedura di esclusione dei residui. I risultati presentati in questa tesi potranno essere utili per validare un metodo sistematico per la stima del contributo entropico all'energia libera di legame. In questo contesto, si potrà utilizzare il database di complessi in [6] che riporta anche i valori sperimentali delle affinità di legame. Il contributo dovuto ai contatti

inter-subunità e alla solvatazione si potrà invece stimare utilizzando potenziali empirici sviluppati nel gruppo di ricerca in cui è stato svolto questo lavoro di tesi^[7]

Bibliografia

- [1] Anfinsen, C.B.; *Principles that govern the folding of protein chains*, Science 181:223-230, **1973**.
- [2] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig H.; Shindyalov, I.N.; Bourne, P.E.; *The Protein Data Bank*, Nucleic Acids Res. 28:235-242, **2000**.
- [3] Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J.L. and Orozco, M. *A consensus view of protein dynamics*, Proc Natl Acad Sci USA, 104: 796-801, **2007**.
- [4] Hess, B.; Kutzner, C.; van der Spoel, D. and Lindahl E. *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation*, J Chem Theory Comput 4:435-447, **2008**.
- [5] Kastritis, P. L.; Moal, I.H.; Hwang, H.; Weng, Z.; Bates, P.A.; Bonvin, A.M.J.J. and Janin, J. *A structure-based benchmark for protein-protein binding affinity*, proteinscience.org, **2011**
- [6] Cossio, P.; Granata, D.; Laio, A.; Seno, F. and Trovato, A. *A simple and efficient statistical potential for scoring ensembles of protein structures*. Sci Rep, 2:351, **2012**.
- [7] Sarti, E.; Zamuner, S.; Cossio, P.; Laio, A.; Seno, F. and Trovato, A. *Bachscore. a tool for evaluating efficiently and reliably the quality of large sets of protein structures*. Computer Phys Comm, 184:2860-2865, **2013**.

- [8] Finkelstein, A.; Ptitsyn, O. *Protein Physics, a course of lectures*, Academic Press, **2002**
- [9] Jackson, M.B. *Molecular and Cellular Biophysics*, Cambridge University Press, **2006**
- [10] Cox, M.M; Nelson, D.L *I principi di Biochimica di Lehninger*, Zanichelli, **2010**
- [11] Rader, A.J.; Chennubhotla, C.; Yang, I.; Bahar, *The Gaussian Network Model: Theory and Applications*