

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

**Studio di dimeri proteici concatenati
per mezzo di modelli a rete elastica**

**Relatore
Prof. Antonio Trovato**

**Laureando
Elia Pimazzoni**

Anno Accademico 2017/2018

Indice

1	Struttura delle proteine	2
1.1	Introduzione sulle proteine	2
1.2	Proprietà delle strutture delle proteine	3
1.3	Strutture domain swapped e dimeri concatenati	4
2	Modelli teorici	7
2.1	Modelli a rete elastica per il calcolo di modi normali	7
2.2	Modello anisotropo	9
2.3	Grado di collettività e deformazioni	10
2.4	ENM e funzione biologica	11
3	Risultati	12
3.1	Identificazione dei modi più rilevanti	12
3.2	Fluttuazioni di dimeri concatenati	14
3.3	Deformazioni e Gaussian Entanglement	16
3.4	Confronto con dimeri non concatenati	18
3.5	Dipendenza dal raggio di cutoff	20
4	Conclusioni	22
A	Bibliografia	23

Capitolo 1

Struttura delle proteine

1.1 Introduzione sulle proteine

Le proteine sono macromolecole fondamentali per molti processi biologici. Sono responsabili della catalisi enzimatica e della replicazione del DNA, inoltre le immunoproteine partecipano alle reazioni di anticorpi e le proteine strutturali rafforzano membrane e tessuti. Questo vasto spettro di funzioni è dovuto alla capacità delle proteine di formare legami con specifiche molecole. La configurazione di una proteina è determinata dalla sua composizione chimica, dai legami interni e dall'interazione con l'ambiente. La regione che interagisce con altre molecole è nota come sito di legame ed è strettamente legata alla struttura e alla dinamica della proteina.

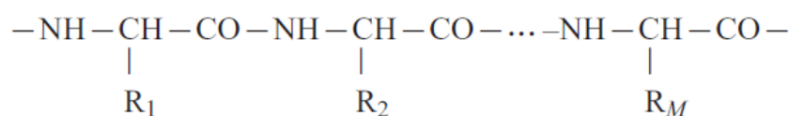


Figura 1.1: Backbone di una proteina e catene laterali (R_1, R_2, \dots, R_M)

Le proteine sono polimeri, ovvero molecole composte da catene di elementi più semplici, i monomeri. Gli elementi costituenti di una proteina sono gli amminoacidi, molecole organiche composte da un gruppo amminico, uno carbossilico e uno specifico gruppo laterale (R). Una catena peptidica viene formata da una sequenza di amminoacidi che interagiscono tramite un legame covalente peptidico tra l'atomo di azoto del gruppo amminico e il carbonio del gruppo

carbossilico. Questo permette di individuare la catena principale o backbone a cui saranno legate le catene laterali, come mostrato in fig. 1.1. Gli atomi di carbonio a cui sono legate le catene laterali vengono identificati come $C\alpha$. I diversi amminoacidi all'interno di una catena vengono chiamati residui e la loro sequenza è unica per ogni proteina.

1.2 Proprietà delle strutture delle proteine

Una proteina può essere descritta secondo diversi livelli di organizzazione. La successione di amminoacidi che compongono una catena polipeptidica viene chiamata struttura primaria ed è uno degli elementi che determinano il ripiegamento della proteina. Osservando la configurazione di diverse proteine si possono osservare alcune strutture che si presentano in segmenti della catena con una certa regolarità. Queste vengono chiamate strutture secondarie. Le più comuni sono α elica e β foglietto. Le prime sono dovute a un ripiegamento regolare della catena dovuto a legami idrogeno tra un residuo e il quarto amminoacido successivo, che causa la formazione di una struttura elicoidale. Nel caso del beta foglietto invece le catene risiedono su una struttura planare e i legami idrogeno uniscono residui di una sezione di catena con quella a fianco, che può essere parallela o antiparallela.

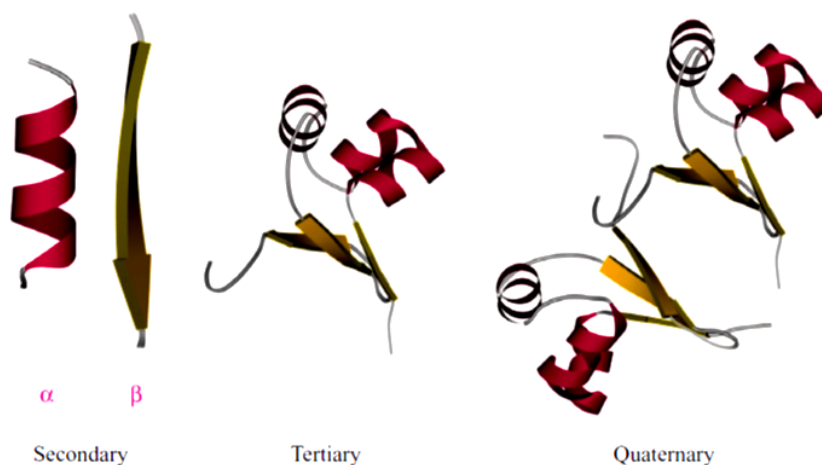


Figura 1.2: Esempio di strutture secondarie, terziarie e quaternarie.

L'intera configurazione tridimensionale di una catena è nota come struttura terziaria, ed esprime come le strutture secondarie si ripiegano. La struttura terziaria può essere governata da diversi tipi di interazione tra cui legami idrogeno, ionici e dipolo-dipolo inoltre può essere anche influenzata dall'ambiente in cui si trova. In alcuni casi, sistemi di più catene possono interagire e formare architetture complesse. Queste vengono descritte come strutture quaternarie.

Le proteine possono essere inoltre divise a seconda della loro struttura generale e del loro ambiente, in tre categorie:

1. Proteine fibrose che si trovano in ambienti poveri d'acqua e hanno una funzione principalmente strutturale.
2. Proteine di membrana che possono avere strutture simili a quelle fibrose in quanto si trovano in ambiente scarso d'acqua. In genere sono tuttavia di dimensioni piuttosto ridotte e svolgono molte funzioni tra cui quelle di recettori, trasportatori o enzimatiche.
3. Proteine globulari che si trovano in ambienti ricchi d'acqua. Queste proteine tendono ad avvolgersi su se stesse in modo che gli amminoacidi apolari idrofobici restino all'interno della struttura.

In questa tesi ci interessiamo di quest'ultima categoria di proteine, per la quale vale il dogma di Anfinsen. Questa ipotesi afferma che per una piccola proteina globulare in condizione fisiologica standard il ripiegamento è completamente determinato dalla sua struttura primaria. Questo postulato è esprimibile anche in termini termodinamici infatti implica che esista una configurazione che sia l'unico minimo di energia libera, che sia stabile e cineticamente accessibile [1]. Questa struttura viene chiamata stato nativo della proteina [2].

1.3 Strutture domain swapped e dimeri concatenati

Alcune proteine presentano strutture che possono esistere ed interagire indipendentemente dal resto della catena peptidica. Questi segmenti, che possono variare da decine a centinaia di

amminoacidi, vengono chiamati domini. In questa tesi vengono studiate un tipo particolare di proteine denominate dimeri domain-swapped. Questo termine è stato introdotto da Bennet, Choe & Eisenberg nel 1994 per descrivere due catene identiche che, scambiandosi lo stesso dominio o elemento strutturale, ricreano una struttura simile a quella che avrebbe il monomero isolato [3]. Il segmento della catena che collega i due monomeri viene chiamato hinge region.

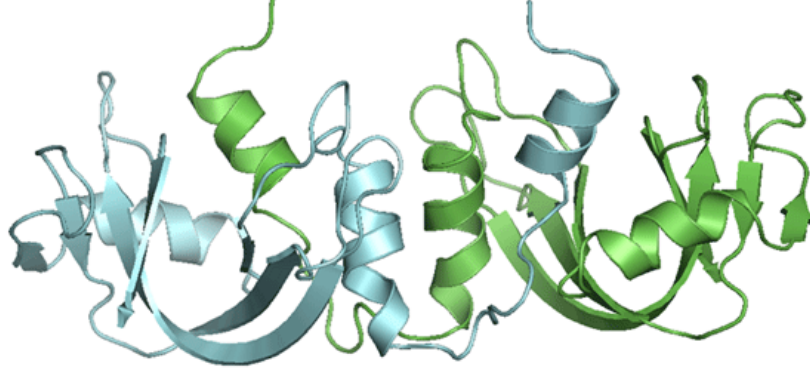


Figura 1.3: Esempio di proteina domain swapped.

Alcune proteine domain-swapped presentano catene concatenate fra loro. Queste possono essere individuate attraverso il calcolo di un parametro topologico, il linking number G . Questa grandezza è definita per due curve chiuse e può essere calcolata utilizzando il doppio integrale gaussiano

$$G \equiv \frac{1}{4\pi} \oint_{\gamma_1} \oint_{\gamma_2} \frac{\vec{r}^{(1)} - \vec{r}^{(2)}}{|\vec{r}^{(1)} - \vec{r}^{(2)}|^3} \cdot (d\vec{r}^{(1)} \times d\vec{r}^{(2)}) \quad (1)$$

Questa definizione può essere utilizzata per calcolare G nel caso di due proteine, scegliendo un'appropriata chiusura delle catene oppure mediando su diverse possibili chiusure. Alternativamente è possibile generalizzare questo parametro a catene aperte composte da N punti, dati dalla posizione degli atomi $C\alpha$. Ponendo $\vec{R}_i^{(1)} \equiv (\vec{r}_{i+1}^{(1)} + \vec{r}_i^{(1)})/2$ e $d\vec{R}_i^{(1)} \equiv \vec{r}_{i+1}^{(1)} - \vec{r}_i^{(1)}$, dove $\vec{r}_i^{(1)}$ sono le coordinate del punto i della prima catena, l'eq. 1 diventa,

$$G' \equiv \frac{1}{4\pi} \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \frac{\vec{R}^{(1)} - \vec{R}^{(2)}}{|\vec{R}^{(1)} - \vec{R}^{(2)}|^3} \cdot (d\vec{R}^{(1)} \times d\vec{R}^{(2)}) \quad (2)$$

E' stato dimostrato che questo parametro, chiamato gaussian entanglement, è utilizzabile come buona alternativa a G nello studio di dimeri concatenati [4].

Capitolo 2

Modelli teorici

2.1 Modelli a rete elastica per il calcolo di modi normali

Come visto nel capitolo 1 le proteine, in condizione fisiologica, tendono ad assumere una precisa conformazione detta stato nativo. In realtà questa struttura non è unica ma è meglio descritta da un ensemble di conformazioni che oscillano attorno ad un minimo dell'energia. Queste oscillazioni possono essere studiate tramite l'analisi dei modi normali, NMA, che permette di individuare i modi più collettivi e di conseguenza più importanti per la funzione di una proteina. L'utilizzo di questa tecnica risale agli anni '80, tuttavia più recentemente sta diventando un metodo molto diffuso grazie all'utilizzo di modelli semplificati e a grana grossa (CG). Infatti, nel suo articolo del 1996, Tirion dimostrò come le predizioni ottenute utilizzando un potenziale semiempirico per la dinamica molecolare (MD), potevano essere raggiunte anche tramite un semplice potenziale con un solo parametro [5].

Utilizzando la NMA una proteina viene considerata come un insieme di N punti di coordinate (x_i, y_i, z_i) in una conformazione $\mathbf{q}^{(k)} = ((\mathbf{R}_1^{(k)})^T, \dots, (\mathbf{R}_N^{(k)})^T)^T = (x_1^{(k)}, y_1^{(k)}, \dots, y_N^{(k)}, z_N^{(k)})$. Attorno a una configurazione di equilibrio \mathbf{q}^0 , il potenziale che lega questi punti può essere sviluppato in serie come

$$V(\mathbf{q}) = V(\mathbf{q}^0) + \sum_i \left(\frac{\partial V}{\partial q_i} \right)^0 (q_i - q_i^0) + \frac{1}{2} \sum_{i,j} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)^0 (q_i - q_i^0)(q_j - q_j^0) + \dots \quad (3)$$

quindi al secondo ordine il potenziale sarà

$$V(\mathbf{q}) = \frac{1}{2} \sum_{i,j} (q_i - q_i^0) H_{ij} (q_j - q_j^0) = \frac{1}{2} \Delta \mathbf{q}^T \mathbf{H} \Delta \mathbf{q} \quad (4)$$

\mathbf{H} rappresenta la matrice Hessiana ottenuta calcolando le derivate seconde del potenziale $H_{ij} = \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)^0$ e sarà quindi una matrice simmetrica di dimensione $3N$. Inoltre \mathbf{H} può essere interpretata come una matrice $N \times N$ i cui elementi sono matrici 3×3 che descrivono il contributo all'energia di ogni coppia di punti. L'Hessiana è quindi diagonalizzabile e i suoi autovettori formano una base ortonormale¹. Gli autovalori sono positivi poiché rappresentano la curvatura locale del potenziale che viene studiato attorno ad un minimo; sei autovalori sono nulli e corrispondono alle rotazioni e traslazioni di corpo rigido.

Scrivendo l'eq. 4 lungo uno specifico modo $V(\mathbf{u}_k) = \frac{1}{2} \mathbf{u}_k^T \mathbf{H} \mathbf{u}_k \propto \frac{\omega_k^2}{2}$ si osserva come l'energia di un modo sia proporzionale al quadrato della frequenza e quindi al suo autovalore, $\lambda_k = \omega_k^2$. Per cui, dato che l'energia vibrazionale sarà divisa ugualmente su tutti i modi, si avranno gli spostamenti maggiori lungo i modi a frequenza minore. Queste saranno perciò le oscillazioni più importanti dal punto di vista funzionale.

Per utilizzare la NMA è quindi necessario individuare un potenziale adeguato. Ad esempio possono essere utilizzati modelli semiempirici, che cercano di considerare i diversi tipi di interazione tra gli atomi, oppure modelli semplificati o a grana grossa. In questa tesi ci interessiamo di modelli a rete elastica (ENM). Le interazioni interne a una proteina vengono quindi semplificate con potenziali armonici tra nodi che possono essere scelti a diversi livelli di coarse graining. In questo modo è possibile tenere conto della topologia di una proteina e ottenere una soluzione esatta per la sua dinamica, anche in caso di molecole di grandi dimensioni. Inoltre con questo modello si ovvia al problema della minimizzazione dell'energia prendendo come configurazione di equilibrio la struttura nativa che può essere ottenuta dal Protein Data Bank (PDB).

¹in realtà per calcolare gli autovettori si dovrebbe utilizzare una matrice Hessiana pesata sulle masse $\tilde{\mathbf{H}} = \mathbf{M}^{-1/2} \mathbf{H} \mathbf{M}^{-1/2}$. Però in questa tesi è stato utilizzato un modello CG ipotizzando che si potessero considerare uguali le masse dei diversi residui e quindi l'Hessiana pesata sarebbe semplicemente $\tilde{\mathbf{H}} = m^{-1} \mathbf{H}$. Gli autovalori di $\tilde{\mathbf{H}}$ sono frequenze al quadrato.

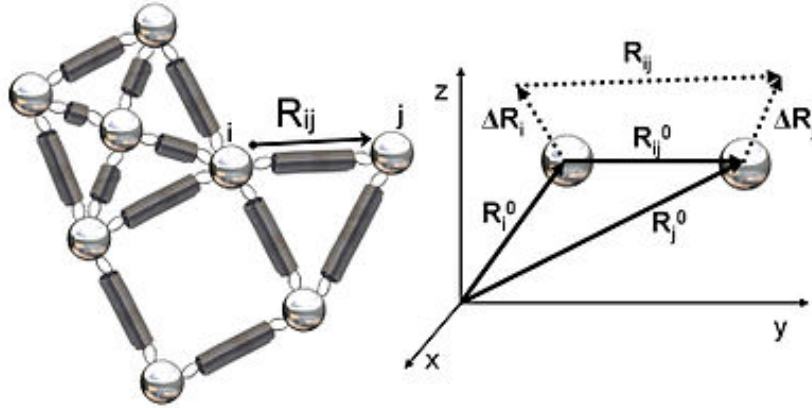


Figura 2.1: Schema di una rete elastica.

Uno dei primi modelli sviluppati in questo campo fu il gaussian network model (GNM) proposto da Bahar et al nel 1997 [6]. Il GNM consiste nell'utilizzo di molle lineari che collegano in maniera isotropa gli atomi di $C\alpha$ entro una certa distanza di cutoff. Questo modello produce risultati riguardo agli spostamenti quadratici medi dei $C\alpha$ in buon accordo con le osservazioni sperimentali. Il GNM tuttavia non fornisce informazioni sulla direzionalità di queste oscillazioni nello spazio tridimensionale.

2.2 Modello anisotropo

Il modello anisotropo (ANM), introdotto nel 2001 [7], è uno dei metodi più semplici e diffusi di ottenere informazioni sulla direzionalità delle oscillazioni. Questo modello a grana grossa utilizza come nodi gli atomi $C\alpha$ soggetti a un potenziale

$$V = \frac{1}{2} \sum_{ij} \gamma_{ij} (R_{ij} - R_{ij}^0)^2 \quad (5)$$

dove R_{ij}^0 rappresenta il modulo del vettore tra due punti della catena nello stato nativo.

In questo modo è possibile, calcolando le derivate seconde del potenziale, risalire all'e-

spressione della matrice Hessiana. In particolare le sottomatrici 3x3 fuori diagonale varranno

$$\mathbf{H}_{ij} = -\frac{\gamma_{ij}}{(R_{ij}^0)^2} \begin{bmatrix} (x_{ij}^0)^2 & x_{ij}^0 y_{ij}^0 & x_{ij}^0 z_{ij}^0 \\ x_{ij}^0 y_{ij}^0 & (y_{ij}^0)^2 & y_{ij}^0 z_{ij}^0 \\ x_{ij}^0 z_{ij}^0 & y_{ij}^0 z_{ij}^0 & (z_{ij}^0)^2 \end{bmatrix} \quad (6)$$

mentre le sottomatrici sulla diagonale saranno

$$\mathbf{H}_{ii} = -\sum_{i;j \neq i} \mathbf{H}_{ij} \quad (7)$$

dove $x_{ij}^0 = (x_j^0 - x_i^0)$ e vale lo stesso per y_{ij} e z_{ij} .

Variando il valore di γ è possibile tener conto della differenza tra i diversi amminoacidi. Per questa tesi γ è stato considerato costante per tutte le coppie di residui a distanza minore di un raggio di cutoff eccetto per $C\alpha$ successivi per i quali è stata considerata una γ dieci volte maggiore. Una volta ottenuta la matrice Hessiana è quindi possibile applicare la NMA per ottenere i modi più lenti.

2.3 Grado di collettività e deformazioni

Il modello anisotropo permette di generare conformazioni vicine a quella di equilibrio considerando le deformazioni lungo i diversi modi. Le fluttuazioni quadratiche di un residuo per un determinato modo possono essere calcolate come

$$(\Delta \mathbf{R}_i)^2 \Big|_k = \text{tr} \{ [\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ii} \} \quad (8)$$

dove λ_k sono gli autovalori di \mathbf{H} , \mathbf{u}_k gli autovettori e la traccia è da considerarsi sulla sottomatrice 3x3 relativa al residuo i .

Le conformazioni deformate lungo un modo k sono

$$\mathbf{q}^{(k)} = ((\mathbf{R}_1^{(k)})^T, \dots, (\mathbf{R}_N^{(k)})^T)^T = ((\mathbf{R}_1^{(0)})^T, \dots, (\mathbf{R}_N^{(0)})^T)^T \pm s \lambda_k^{-1/2} \mathbf{u}_k \quad (9)$$

dove s esprime l'ampiezza della deformazione e può essere scelto in modo da ricreare le oscillazioni osservate sperimentalmente.

Un'altra quantità di particolare interesse è il grado di collettività. Questo parametro è di natura entropica ed esprime quanto il movimento lungo un determinato modo coinvolga un numero elevato di residui. Di norma i modi con un alto grado di collettività corrispondono alle oscillazioni più lente, ovvero con autovalori minori. Questo parametro è definito come

$$k_k = N^{-1} \exp \left\{ - \sum_{i=1}^N \alpha (\Delta \mathbf{R}_i)^2 \Big|_k \log(\alpha \Delta \mathbf{R}_i)^2 \Big|_k \right\} \quad (10)$$

con α costante di normalizzazione calcolata come $\sum_i \alpha (\Delta \mathbf{R}_i)^2 \Big|_k = 1$.

2.4 ENM e funzione biologica

Come discusso nelle sezioni precedenti, i modelli a rete elastica permettono di individuare i modi più significativi per la dinamica globale di una catena peptidica. La conoscenza di queste oscillazioni può essere utilizzata per analizzare come avvengano molti dei processi che determinano la funzione di una proteina. Infatti la dinamica di una catena peptidica può essere utilizzata per studiare come una proteina possa legarsi con altre molecole o come avvenga il trasporto di membrana. Inoltre lo studio di modi normali tramite ENM può essere utilizzato anche per l'analisi della dinamica di proteine motrici [8][9].

Uno studio del 2005 ha analizzato 98 enzimi evidenziando come i siti catalitici si trovino generalmente su residui con basse fluttuazioni, perni delle oscillazioni, mentre siti di legame di ligandi hanno mobilità moderate e anche elevate se devono accettare un ampio spettro di ligandi [10]. In un articolo di Bahar et al del 2010 il ANM è stato applicato a diverse proteine di membrana, in particolare è stato osservato come lo studio dei modi più collettivi aiuti a comprendere diversi fenomeni legati all'apertura e chiusura di trasportatori di membrana. Ad esempio per canali ionici sono stati individuati alcuni modi che mantengono la simmetria della struttura e svolgono un ruolo fondamentale nell'attivazione del processo di gating. Oltre ad evidenziare il funzionamento dei trasportatori di membrana, nell'articolo del 2010, lo studio dei modi normali è stato utilizzato per individuare siti di legame e di importanza allosterica [11].

Capitolo 3

Risultati

In questo capitolo applichiamo i metodi descritti nel capitolo 2 e nella sezione 1.3 a due proteine concatenate domain-swapped, 1m0d e 3die, e a un dimero non concatenato, 2pa7. Si cercherà quindi di osservare come le proprietà di una proteina domain-swapped influenzino la sua dinamica. Inoltre si andrà ad analizzare come le fluttuazioni attorno all'equilibrio influenzino il linking delle proteine concatenate.

3.1 Identificazione dei modi più rilevanti

Applicando il ANM è possibile individuare i modi più significativi ed osservare le fluttuazioni attorno alla posizione di equilibrio. Per determinare lo stato nativo di 1m0d e 3die sono state prese le strutture fornite dal PDB. Per poter applicare il modello a rete elastica si sono quindi isolate le coordinate degli atomi $C\alpha$. Da queste, applicando le formule 6 e 7, è stata calcolata la matrice Hessiana. Il raggio di cutoff è stato scelto come $r_c = 15\text{\AA}$, le motivazioni di questa scelta sono illustrate nella sezione 3.5.

Una volta ottenuta la matrice Hessiana si sono calcolati autovettori e autovalori. In questo modo è stato possibile evidenziare quali fossero i modi di frequenza più bassa ovvero con autovalori minori, esclusi i sei nulli. I primi 10 autovalori non nulli delle due proteine concatenate sono riportati nella tabella 3.1, in unità di γ/m .

1m0d	0.00086	0.00412	0.0134	0.129	0.176	0.374	0.888	0.929	1.034	1.123
3die	0.321	0.398	0.598	1.538	1.860	2.237	2.360	2.453	2.529	2.844

Tabella 3.1

Osservando la tabella si può evidenziare come, per la proteina 1m0d, sia presente un salto tra il sesto e il settimo autovalore. Questo potrebbe essere indicativo del fatto che i modi equivalenti ai primi sei autovalori siano quelli più cooperativi. Invece per 3die è meno evidente quali possano essere gli autovalori più importanti per la dinamica globale della proteina.

Per verificare che effettivamente i modi di frequenza minore corrispondano alle oscillazioni più significative, sono stati calcolati i coefficienti di collettività per i diversi modi, utilizzando la formula 10. I risultati sono mostrati in fig. 3.1 e 3.2.

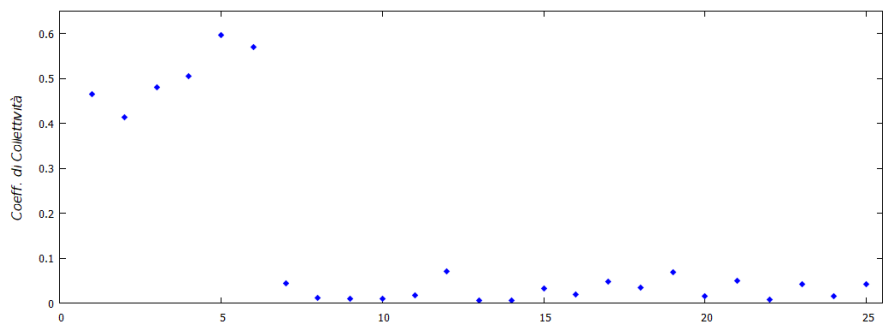


Figura 3.1: Coefficienti di collettività di 1m0d per i primi 25 autovalori

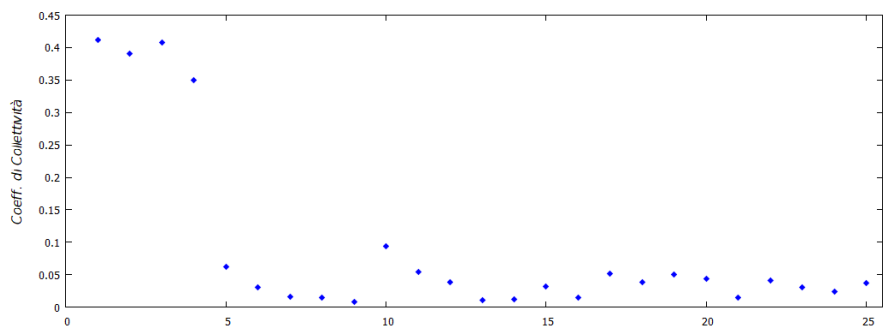


Figura 3.2: Coefficienti di collettività di 3die per i primi 25 autovalori

Il grafico relativo a 1m0d conferma quanto osservato dagli autovalori, i primi sei modi hanno infatti un coefficiente di collettività chiaramente maggiore rispetto a quelli successivi. Nel

caso della proteina 3die invece si può osservare come i modi di maggior interesse siano quelli corrispondenti ai quattro autovalori minori.

3.2 Fluttuazioni di dimeri concatenati

Una volta individuati i modi più importanti per la dinamica è stato possibile studiare il profilo di mobilità dei residui. Sono state calcolate le fluttuazioni, utilizzando l'eq. 8, lungo i primi sei modi di 1m0d e i primi quattro di 3die. I risultati ottenuti sono riportati in unità di $k_B T / \gamma$ nei grafici 3.3 3.4 3.5 3.6, i modi sono ordinati per frequenze crescenti (modo 1, modo 2...).

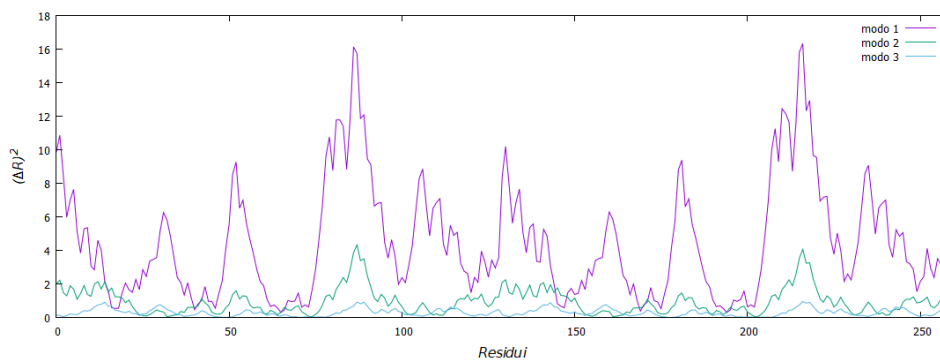


Figura 3.3: Fluttuazioni per i primi tre modi di 1m0d

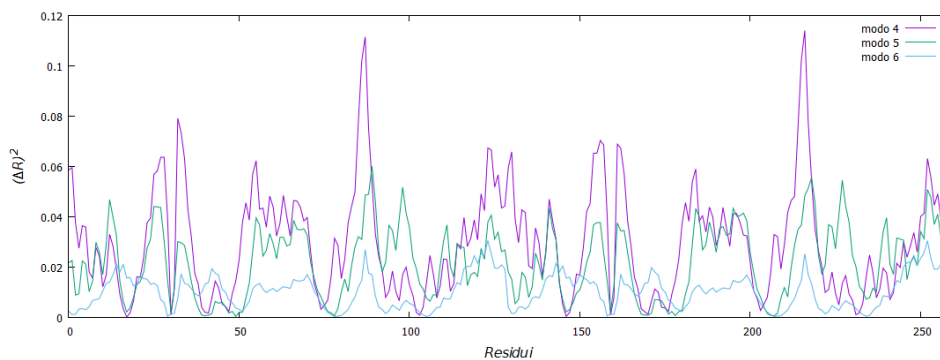


Figura 3.4: Fluttuazioni per i tre modi successivi di 1m0d

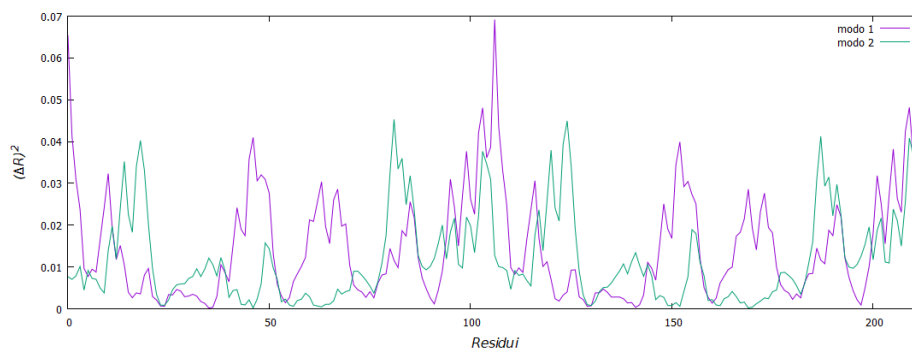


Figura 3.5: Fluttuazioni per i primi due modi di 3die

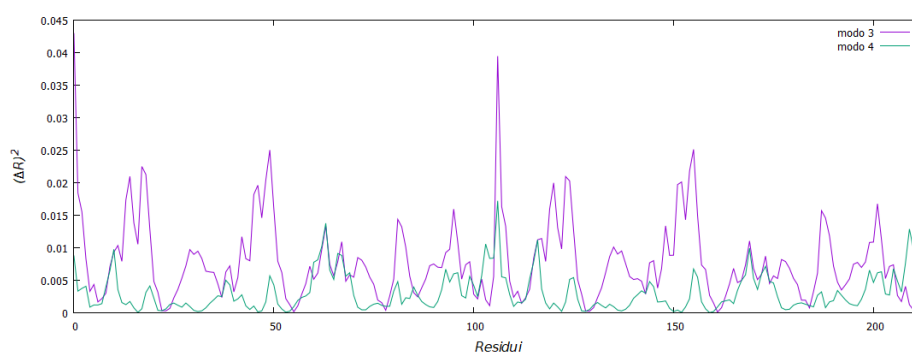


Figura 3.6: Fluttuazioni per i due modi successivi di 3die

Come ci si aspettava dato l'elevato coefficiente di collettività, le fluttuazioni coinvolgono molti residui distribuiti lungo tutta la catena sia per 1m0d che per 3die. Inoltre è possibile osservare come le due catene che compongono la proteina abbiano profili simili ma non identici, infatti dopo la metà del grafico il pattern di massimi e minimi si ripete. Per entrambe le proteine possono essere individuate delle zone in cui sono presenti picchi o minimi della fluttuazione nei diversi modi. In particolare 1m0d presenta delle zone ad elevata mobilità attorno ai residui 53 e 87, invece 3die vicino al quarantottesimo amminoacido. Le regioni in cui la fluttuazione risulta minore sono attorno al settantaquattresimo residuo per 1m0d e ai residui 23, 55 e 77 per 3die. Dato che queste zone di elevata o bassa mobilità si ripresentano lungo diversi modi, sono probabilmente le più importanti per la dinamica globale della proteina.

Può essere interessante studiare le fluttuazioni nella hinge region che per entrambe le proteine si trova attorno al trentesimo residuo, mentre l'elemento swapped include tutti i residui

precedenti. Nel caso di 1m0d la hinge region risulta avere mobilità bassa per tutti i modi eccetto il primo ed il terzo. Invece per 3die presenta delle fluttuazioni piuttosto ridotte per il primo e il quarto modo mentre più elevate per il secondo e il terzo.

3.3 Deformazioni e Gaussian Entanglement

Come osservato nella sezione 2.2 il ANM permette di calcolare le configurazioni deformate lungo i diversi modi. Utilizzando la formula 9 , si è studiato come cambiasse la struttura attorno all'equilibrio per ognuno dei modi evidenziati nella sezione 3.1. Il parametro s è stato scelto in un intervallo $[-s_{max};s_{max}]$ tale che la variazione della distanza tra il centro di massa della proteina e uno dei residui più mobili per ogni modo fosse minore del dieci percento. Sono quindi state calcolate 200 configurazioni nell'intervallo individuato e, per ognuna delle strutture deformate, è stato calcolato il valore del termine di linking utilizzando l'eq. 2. I risultati sono riportati nei seguenti grafici che mostrano le variazioni di G' lungo le deformazioni per ognuno dei modi ordinati per frequenze crescenti.

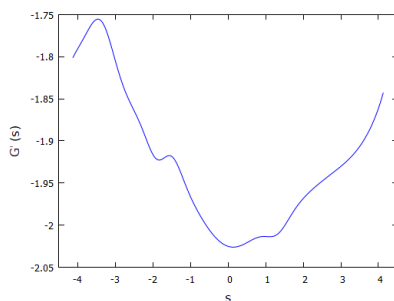


Figura 3.7: 1m0d, modo 1

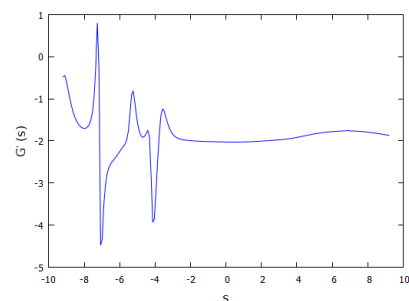


Figura 3.8: 1m0d, modo 2

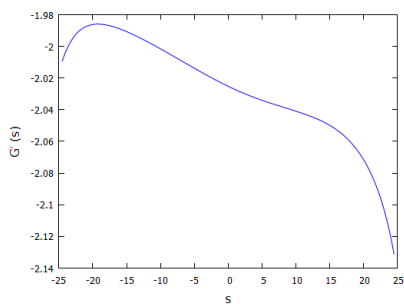


Figura 3.9: 1m0d, modo 3

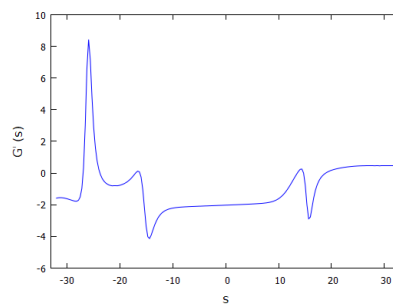


Figura 3.10: 1m0d, modo 4

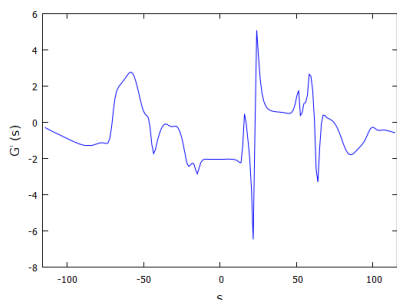


Figura 3.11: 1m0d, modo 5

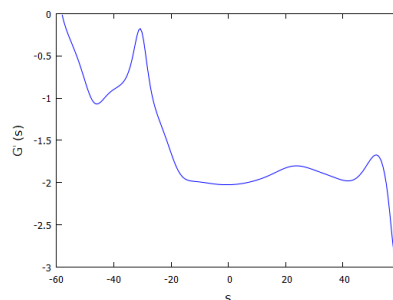


Figura 3.12: 1m0d, modo 6

Si può osservare come, almeno in un intorno della posizione di equilibrio, $s = 0$, il gaussian entanglement risulti piuttosto stabile, variando in un intervallo ridotto di valori. Per le deformazioni relative al sesto modo, fig. 3.12, oltre ad un'ampiezza con $s = -20$ G' tende ad aumentare raggiungendo un valore di -0.2 , questo potrebbe indicare una deformazione che elimina la concatenazione di 1m0d.

Nelle fig. 3.8 3.10 e 3.11 sono presenti delle discontinuità o delle grandi variazioni di G' . Questo potrebbe essere dovuto ad un problema con il calcolo discretizzato del parametro G' . Un'altra possibilità è che l'errore stia nella scelta di s_{max} , infatti le deformazioni studiate dovrebbero essere paragonabili a quelle osservate sperimentalmente. Quindi è possibile che, se la scelta di s_{max} ha portato a deformazioni eccessivamente grandi, l'analisi di modi normali con ENM non sia più valida infatti si basa sull'ipotesi di studiare la dinamica attorno ad un equilibrio.

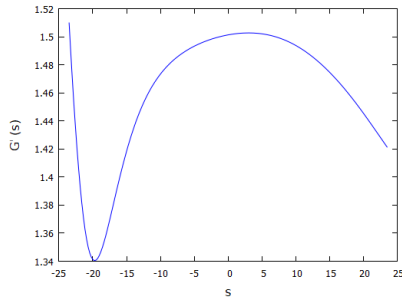


Figura 3.13: 3die, modo 1

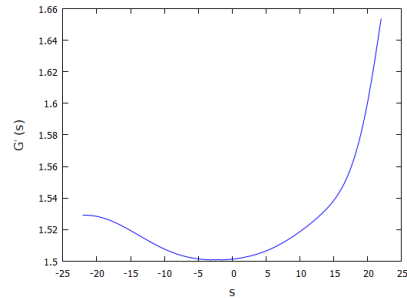


Figura 3.14: 3die, modo 2

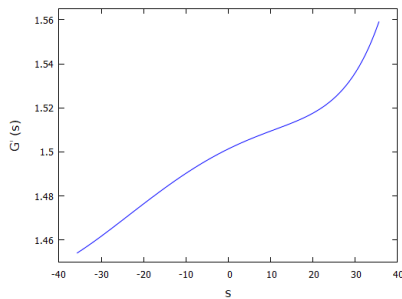


Figura 3.15: 3die, modo 3

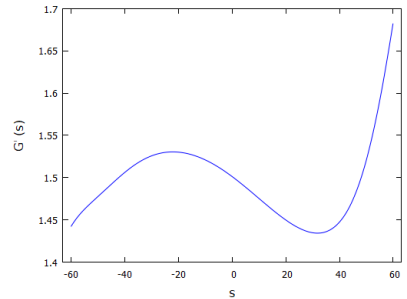


Figura 3.16: 3die, modo 4

Per 3die non si ripresentano le problematiche evidenziate per 1m0d. Le variazioni di G' risultano piuttosto ridotte. Inoltre nessuna delle deformazioni causa variazioni di G' che facciano pensare che possa esserci un'eliminazione della concatenazione.

3.4 Confronto con dimeri non concatenati

Per osservare le differenze nella dinamica di proteine concatenate e non, l'analisi fatta nelle sezioni 3.1 e 3.2 è stata ripetuta con una proteina non concatenata, 2pa7. Quindi, una volta ottenuta la matrice Hessiana si sono calcolati autovalori e autovettori. Come per le proteine concatenate i modi più rilevanti sono quelli relativi agli autovalori minori, i primi dieci sono riportati in tab. 3.2, in unità di γ/m . Sono stati eliminati nel calcolo dei modi normali di oscillazione gli ultimi 15 residui della proteina, perchè appartenenti a una alpha-elica non ben interconnessa al resto della rete elastica.

2pa7	0.777	0.900	1.208	1.385	1.495	1.973	2.033	2.191	2.392	2.708
------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Tabella 3.2

Dalla tab. 3.2 si può osservare come gli autovalori aumentino in maniera piuttosto regolare e non si individuino un gruppo di autovalori chiaramente distaccati da quelli successivi. Quindi per capire se alcuni modi possano essere interessanti per la dinamica globale della proteina sono stati calcolati i coefficienti di collettività, riportati in fig. 3.17.

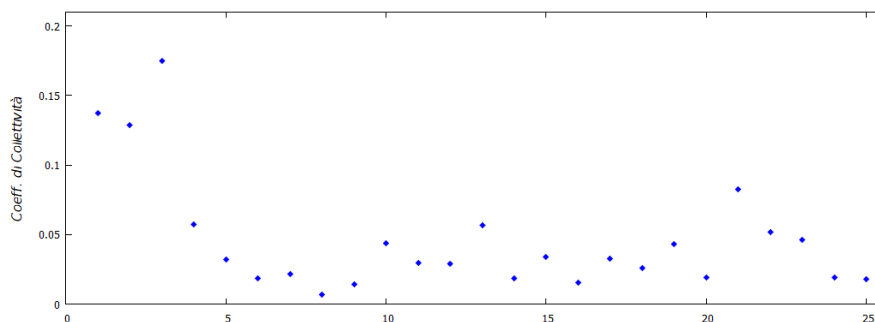


Figura 3.17: Coefficienti di collettività di 2pa7 per i primi 25 autovalori

I primi tre modi risultano quindi avere un coefficiente di collettività superiore a quelli successivi. Tuttavia il loro valore resta piuttosto basso, indicando come le fluttuazioni per 2pa7 siano localizzate in sezioni ridotte della proteina. Per confermare queste osservazioni sono state calcolate le fluttuazioni lungo i tre modi con frequenza minore. I risultati sono mostrati in fig. 3.18, in unità di $k_B T / \gamma$.

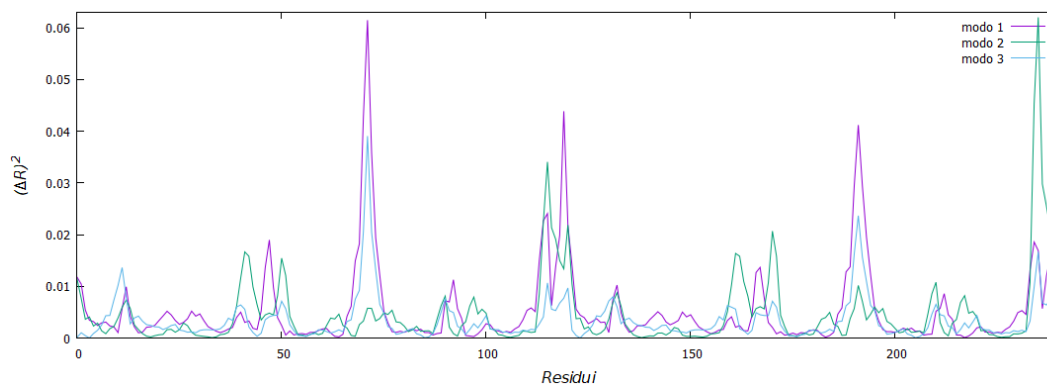


Figura 3.18: Fluttuazioni per i primi tre modi di 2pa7

Dalla fig. 3.18 si ritrova quanto ci si aspettava dati i bassi coefficienti di collettività. Infatti le fluttuazioni sono concentrate su pochi amminoacidi con dei picchi attorno ai residui 71, 191 e nella parte finale delle catene, mentre nel resto della proteina la mobilità è piuttosto bassa. La differenza più evidente tra quanto trovato per 2pa7 rispetto alle proteine concatenate è quindi l'assenza di modi con collettività elevata. È possibile che il linking tra le catene favorisca la presenza di oscillazioni globali. Quest'osservazione potrebbe essere corretta se due catene concatenate tendessero ad interagire maggiormente rispetto a due catene che sono semplicemente vicine spazialmente. Per verificare la veridicità di queste affermazioni sarebbe tuttavia necessario svolgere un'analisi come quella fatta in questo capitolo su molte più proteine.

3.5 Dipendenza dal raggio di cutoff

Per osservare come variassero i risultati cambiando il raggio di cutoff e per decidere quale utilizzare nell'analisi dei risultati sono state calcolate le matrici Hessiane per diversi valori di r_c . In particolare per le proteine 1m0d e 3die è stato applicato il modello anisotropo con $r_c = 12, 15, 18, 21 \text{ \AA}$. Nelle Fig. 3.19 e 3.20 sono mostrati i coefficienti di collettività relativi ai venti autovalori minori non nulli per le due proteine concatenate considerate con i diversi raggi di cutoff.

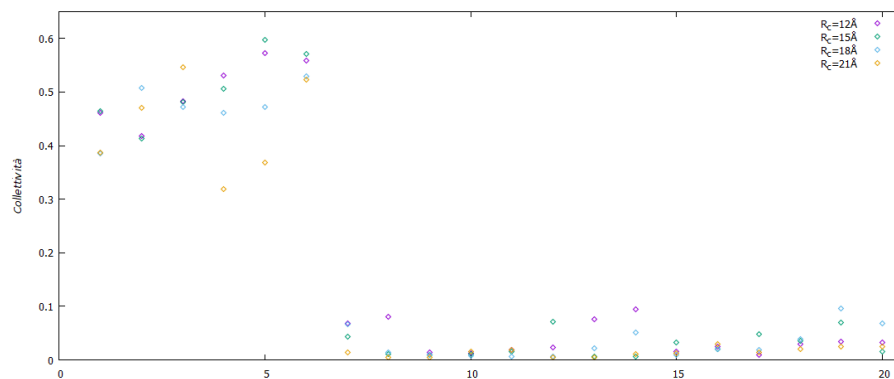


Figura 3.19: Coefficienti di collettività di 1m0d per i primi 20 autovalori

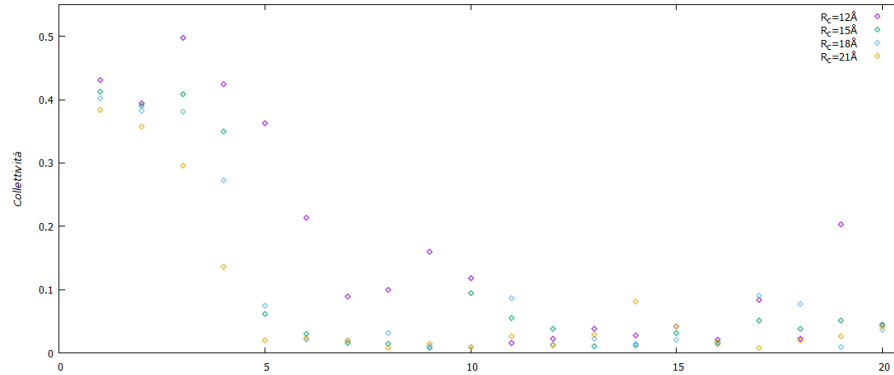


Figura 3.20: Coefficienti di collettività di 3die per i primi 20 autovalori

Si può osservare che, come affermato da Bahar et al [11], il modello anisotropo è piuttosto resistente per cambi di r_c , infatti, in particolare per 1m0d, si ha circa lo stesso numero di modi con alta collettività. Le fig. 3.19 e 3.20 mostrano inoltre come le scelte di r_c che producono i risultati più simili sono 15 e 18 Å per 1m0d e 12 e 15 Å per 3die. Quindi il valore scelto per l'analisi delle oscillazioni è stato $r_c = 15 \text{ \AA}$ per entrambe.

Capitolo 4

Conclusioni

Lo studio di dimeri concatenati tramite un modello a rete elastica anisotropo ha portato a diversi risultati interessanti. Innanzitutto è stata utilizzata l'analisi dei modi normali con un modello anisotropo per schematizzare le interazioni interne delle proteine. Per la scelta del raggio di cutoff è stato studiato come variassero i risultati cambiando R_c ed è stato scelto il calore più appropriato. In questo modo è stato possibile individuare, per 1m0d e 3die, dei modi cooperativi che influenzano la dinamica globale delle catene peptidiche. Inoltre è stato osservato come le fluttuazioni, per alcuni residui, presentassero profili simili lungo i diversi modi. Lo studio del gaussian entanglement ha evidenziato come il parametro G' risulti piuttosto stabile per i modi più lenti attorno all'equilibrio. Inoltre per la proteina 1m0d è emerso un modo che potrebbe portare all'eliminazione della concatenazione tra le catene. Sono state evidenziate delle problematiche legate al gaussian entanglement che potrebbero essere legate alla definizione del parametro o alla scelta di come applicarlo. Per individuare esattamente da dove nasca questa problematica si potrebbero eseguire ulteriori studi sul linking delle deformazioni di proteine concatenate. Grazie all'analisi di 2pa7 sono anche state evidenziate delle interessanti differenze tra dimeri concatenati e dimeri domain swapped non concatenati. In particolare l'assenza di modi con alta collettività per la proteina non concatenata. Questa differenza potrebbe essere dovuta alla maggiore interazione tra due catene concatenate. Anche queste osservazioni necessiterebbero di ulteriori studi per dimostrarne la validità in generale.

Appendice A

Bibliografia

- [1] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 1973.
- [2] Alexei V. Finkelstein and Oleg B. Ptitsyn. *Protein Physics*. Academic Press, 2002.
- [3] M J Bennett, S Choe, and D Eisenberg. Domain swapping: entangling alliances between proteins. *Proceedings of the National Academy of Sciences*, 91(8):3127–3131, 1994.
- [4] Marco Baiesi, Enzo Orlandini, Antonio Trovato, and Flavio Seno. Linking in domain-swapped protein dimers. *Scientific Reports*, 2016.
- [5] Monique M. Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, 77:1905–1908, Aug 1996.
- [6] Ivet Bahar, Ali Rana Atilgan, and Burak Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, 2(3):173 – 181, 1997.
- [7] A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, and I. Bahar. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical Journal*, 80(1):505 – 515, 2001.

- [8] Wenjun Zheng and Sebastian Doniach. A comparative study of motor-protein motions by using a simple elastic-network model. *Proceedings of the National Academy of Sciences*, 100(23):13253–13258, 2003.
- [9] Moon ki Choi, Soojin Jo, Byung Ho Lee, Min Hyeok Kim, Jae Boong Choi, Kyunghoon Kim, and Moon Ki Kim. Dynamic characteristics of a flagellar motor protein analyzed using an elastic network model. *Journal of Molecular Graphics and Modelling*, 78:81 – 87, 2017.
- [10] Chakra Chennubhotla, A J Rader, Lee-Wei Yang, and Ivet Bahar. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Physical Biology*, 2(4):S173, 2005.
- [11] Ivet Bahar, Timothy R. Lezon, Ahmet Bakan, and Indira H. Shrivastava. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chemical reviews*, 2010.