

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea Triennale in Fisica

Tesi di Laurea

**Studio di strutture proteiche topologicamente
intrecciate per mezzo di modelli a rete elastica**

Relatore

Prof. Antonio Trovato

Laureando

Michele Puppin

Anno Accademico 2018/2019

Abstract

Lo scopo della tesi è studiare gli effetti delle fluttuazioni delle strutture proteiche sul grado di entanglement delle stesse. Si sono inizialmente presentate le caratteristiche generali delle proteine sottolineando aspetti della loro struttura e composizione di particolare interesse per questo lavoro di tesi. Successivamente si è introdotto l'entanglement topologico di strutture proteiche con particolare attenzione al calcolo dell'entanglement gaussiano, parametro capace di valutarne matematicamente l'entità. Si sono poi trattati i principali metodi di analisi delle strutture proteiche ed i modelli a rete elastica analizzando nel dettaglio il modello anisotropo in seguito utilizzato. Infine si sono presentate le modalità e i risultati dell'analisi svolta su due specifiche proteine.

Indice

1	Introduzione alle proteine	1
1.1	Composizione e struttura delle proteine	1
1.2	Entanglement topologico	3
2	Modelli a rete elastica	5
2.1	Fluttuazioni dello stato nativo	5
2.2	Analisi ai modi normali	6
2.3	Modello anisotropo	7
3	Studio di strutture proteiche	11
3.1	Analisi di 2hqx	11
3.2	Analisi di 3cry	13
3.3	Conclusioni	15
A	Grafici e tabelle	19

Capitolo 1

Introduzione alle proteine

1.1 Composizione e struttura delle proteine

Le proteine sono macromolecole biologiche di natura polimerica costituite da catene di amminoacidi legati tra loro da un legame peptidico. La lunghezza della catena e la sequenza di amminoacidi, i principali sono venti, sono codificate geneticamente. È possibile individuare una catena principale, che consiste nella spina dorsale della molecola, alla quale risultano collegate le catene laterali, ossia le parti dei residui che rimangono libere, come mostrato in Figura 1.1. [1]

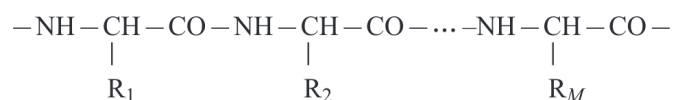


Figura 1.1: Rappresentazione della catena principale. [1]

Le proteine svolgono molteplici funzioni a livello intra e extra cellulare. La maggior parte delle proteine ha funzioni di catalisi enzimatica delle reazioni chimiche svolgendo attività di regolazione nell'espressione genica, sintesi di molecole e ricezione e trasporto di informazioni. Sono inoltre ampiamente utilizzate come proteine di struttura o per il trasporto di altre molecole. Al fine di svolgere tale funzione è di cruciale importanza la struttura tridimensionale che le proteine assumono successivamente al loro assemblaggio. Queste infatti hanno un'elevata specificità in relazione alle molecole con cui interagiscono e questo è spesso dovuto ad una interazione del tipo *chiave-serratura* che richiede una rigida struttura spaziale. Qualora questa risultasse variata anche in minima parte, il funzionamento della proteina risulterebbe compromesso. [1]

Alla base della catena proteica vi è l'unità peptidica rappresentata in Figura 1.2. Questa presenta una configurazione planare a causa della natura del legame tra l'azoto e il carbonio che blocca la rotazione dell'angolo ω rappresentato in Figura 1.3. L'angolo ω può unicamente vibrare a causa dell'agitazione termica ed è quindi fissato in modo tale da riprodurre la configurazione *trans*, che risulta essere marcatamente più frequente, e *cis*. L'angolo χ rappresenta invece l'angolo di rotazione della catena laterale. Questo angolo, assieme alle vibrazioni della lunghezza dei legami, non contribuisce alla flessibilità della proteina e il contributo delle vibrazioni degli angoli è limitato: a conferire elevata flessibilità alla proteina sono quindi le rotazioni attorno agli angoli ψ , ϕ . [1]

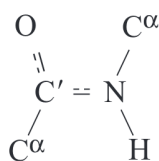


Figura 1.2: Unità peptidica in configurazione *trans*. [1]

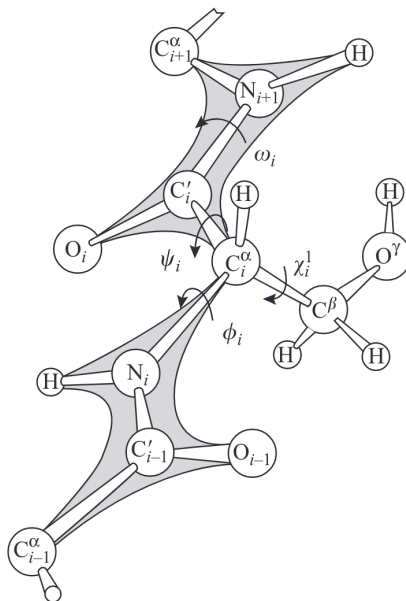


Figura 1.3: Raffigurazione di una catena polipeptidica con il residuo i -esimo. [1]

La struttura proteica può essere studiata nel dettaglio tramite tecniche di cristallografia a raggi X e studi tramite risonanza magnetica nucleare della proteina in soluzione acquosa che permettono di ottenere le coordinate cartesiane dei singoli atomi. Questi approcci tuttavia sono possibili unicamente per lo studio di proteine solubili in acqua, ovvero le proteine globulari, in quanto possono essere facilmente isolate come molecole singole. Le proteine fibrose, così come le proteine di membrana, risultano invece insolubili in acqua ed è quindi difficile ricavarne la struttura. [1]

La macrostruttura di una proteina può essere suddivisa in struttura primaria, secondaria, terziaria e quaternaria. La struttura primaria è data da una determinata sequenza di amminoacidi legati da legame peptidico. Sono invece generalmente legami non covalenti a mantenere la conformazione tridimensionale dando vita a strutture secondarie, terziarie e quaternarie. Le strutture secondarie sono ripiegamenti della catena polipeptidica nelle configurazioni α -elica o β -foglietto raffigurate in Figura 1.4. Le strutture secondarie di una singola catena possono poi ripiegarsi in strutture globulari o fibrose che vanno quindi a costituire gran parte delle strutture terziarie esistenti. Infine l'associazione di più catene che interagiscono generalmente tramite legami deboli dà vita alle strutture quaternarie. Inoltre all'interno della singola catena si possono identificare due o più regioni compatte chiamate domini, generalmente contenenti 100 - 200 residui, che svolgono funzioni specifiche. Ogni dominio ripiega in maniera autonoma. Il ripiegamento della catena polipeptidica per formare strutture secondarie e terziarie avviene, *in vitro*, spontaneamente e sempre nello stesso modo senza la necessità che intervengano processi cellulari specifici, che sono invece presenti *in vivo*. Questo fatto indica chiaramente che la struttura della proteina deriva unicamente dalla specifica sequenza di amminoacidi che la compongono. [1]

A determinare e mantenere le strutture secondarie e terziarie concorrono diversi tipi di interazione. Vi è la possibilità che si formi un ponte disolfuro tra due differenti residui di cisteina contenenti il gruppo $-SH$, ovvero un legame covalente del tipo $-S-S-$. La maggior parte delle interazioni tuttavia è di natura non covalente. Alcuni esempi sono il legame ionico, dovuto essenzialmente all'interazione tra strutture con carica opposta come gruppi carbossilici carichi negativamente e gruppi amminici carichi positivamente, e le forze di Van der Waals, ovvero interazioni deboli ma fondamentali nello studio di agglomerati di grandi dimensioni. Queste sono dovute al fatto che a distanze superiori a 2 - 3 Å gli atomi, invece di respingersi, si attraggono a causa delle vibrazioni coordinate degli elettroni in entrambi gli atomi. Un'interazione molto più intensa delle forze di Van der Waals è invece il legame idrogeno; questo si instaura tra due residui quando un atomo di idrogeno legato ad un atomo elettronegativo si trova in prossimità di un altro atomo elettronegativo

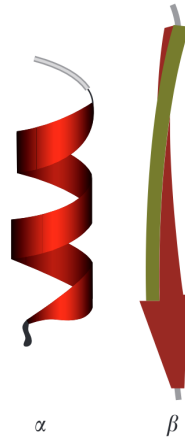


Figura 1.4: Rappresentazione delle configurazioni α -elica e β -foglietto. [1]

come nei casi $O - H :: O$ e $N - H :: N$. In particolare nelle proteine si possono realizzare legami idrogeno del tipo $N - H :: O$ fra l'azoto del gruppo amminico e l'ossigeno del gruppo carbonilico, entrambi situati sulla catena principale, che stabilizzano le strutture secondarie. Infine un ruolo importante nella determinazione della forma globulare della proteina è svolto dall'effetto idrofobico per il quale la proteina tende ad esporre all'acqua i residui polari racchiudendo quindi in una tasca idrofobica quelli apolari. [1]

1.2 Entanglement topologico

L'assemblaggio nei ribosomi della sequenza di amminoacidi a partire dalle informazioni contenute nel DNA, produce una catena lineare. Questa, subito dopo essere stata assemblata, va incontro ad un processo di ripiegamento per mezzo del quale assume la specifica forma tridimensionale, chiamata stato nativo, che risulta biologicamente funzionale. All'interno della cellula il processo viene favorito dall'azione di enzimi specifici che tuttavia hanno il solo scopo di evitare contatti indesiderati con altre molecole all'interno della cellula. Si è visto infatti che le proteine sono capaci di ripiegarsi spontaneamente anche *in vitro*. [1] Il fatto quindi che a partire da una determinata sequenza di amminoacidi si ottenga in modo del tutto spontaneo la medesima configurazione è un chiaro indizio del fatto che lo stato nativo rappresenta uno stato di minimo dell'energia.

Il processo di ripiegamento risulta estremamente complesso e difficile da studiare, sorge dunque naturale domandarsi come questo possa avvenire senza incorrere in errori anche per configurazioni molto complicate. Non è stato del tutto compreso, ad esempio, come tale processo dia vita a proteine che presentano nodi nella topologia della loro struttura, le quali pur presentando una configurazione intricata sono state rilevate in discreto numero fra le proteine del Protein Data Bank (3%). [2] Lo studio di alcuni meccanismi del processo di ripiegamento può essere svolto a partire dall'analisi della configurazione geometrica dello stato nativo. Si è visto, ad esempio, che conoscendo quali coppie di residui risultano in contatto reciproco è possibile ricavare informazioni su quali siano i nuclei di ripiegamento della proteina. [3] Nel caso di proteine dall'elevata complessità, lo studio di proprietà locali non è tuttavia esauriente ed è utile analizzare aspetti più generali della geometria come ad esempio il grado di entanglement topologico di un determinato stato nativo. La proteina può infatti essere pensata come una curva che, deformandosi necessariamente in modo continuo, può essere analizzata con studi di natura topologica. L'entanglement topologico indica l'interconnessione fisica tra due molecole o tra due parti di una stessa molecola. Lo studio di questo aspetto può fornire informazioni utili; si è verificato infatti che la distanza media lungo la catena tra due residui che, entrando in contatto reciproco, danno vita ad un loop della catena proteica risulta fortemente correlata con il tempo di ripiegamento della proteina stessa. [3]

È dunque interessante interrogarsi circa gli effetti che il grado di entanglement della struttura proteica ha a livello energetico e per farlo è necessario valutare matematicamente questa quantità.

A tale scopo si utilizza l'entanglement gaussiano, ovvero una generalizzazione dell'integrale di Gauss utilizzato per calcolare il linking number. [4] Il linking number di due curve chiuse è un parametro che descrive quante volte una curva gira attorno all'altra. In generale se tale parametro è nullo risulta possibile separare le due curve senza spezzarle mentre se è non nullo questa operazione non risulta possibile. In Figura 1.5 un esempio di curve con linking number nullo (sinistra) e pari a uno (destra).

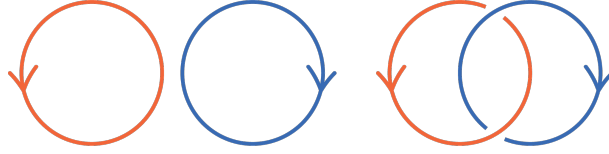


Figura 1.5: Due esempi di entanglement tra curve chiuse

Il linking number tra due curve chiuse $\gamma_i = \{\mathbf{r}^{(i)}\}$ e $\gamma_j = \{\mathbf{r}^{(j)}\}$ in \mathbb{R}^3 è un numero intero e un invariante topologico. Il suo valore può essere calcolato tramite il doppio integrale di Gauss

$$G \equiv \frac{1}{4\pi} \oint_{\gamma_i} \oint_{\gamma_j} \frac{\mathbf{r}^{(i)} - \mathbf{r}^{(j)}}{|\mathbf{r}^{(i)} - \mathbf{r}^{(j)}|^3} \cdot (d\mathbf{r}^{(i)} \times d\mathbf{r}^{(j)}). \quad (1.1)$$

Se si calcola il doppio integrale fra le due curve aperte invece si ottiene un numero reale G' , chiamato entanglement gaussiano, che quantifica quindi proprio l'entanglement fra le due curve. [5]

Anche per le proteine è possibile calcolare l'entanglement gaussiano andando a prendere in considerazione due sottocatene, aperte o chiuse che siano. Un valore del modulo di G' maggiore, o al limite vicino, a uno indica una condizione di entanglement per le due sottocatene. All'interno delle strutture proteiche si possono venire a formare dei loop quando due residui in punti diversi della catena si trovano in contatto. Si è visto che conformazioni nelle quali due loop si intersecano, a sinistra in Figura 1.6, o una parte della catena risulta avvolta da un loop, a destra in Figura 1.6, risultano tutt'altro che rare, essendo presenti nel 30% del data set di proteine a singolo dominio considerate in [4].



Figura 1.6: Esempio di figure di entanglement proteico

Per calcolare l'entanglement gaussiano a partire dalla proteina si prende in considerazione una catena peptidica e si traccia una linea spezzata che colleghi gli atomi di carbonio C^α consecutivi. È quindi possibile individuare le due sottocatene: la prima dal carbonio i_1 al carbonio i_2 corrisponderà alla curva γ_i mentre la seconda da j_1 a j_2 corrisponderà a γ_j . Le due sottocatene devono essere tali che $\gamma_i \cap \gamma_j = \emptyset$. Risulta quindi possibile calcolare l'entanglement gaussiano come

$$G'_{ij} \equiv \frac{1}{4\pi} \sum_{i=i_1}^{i_2-1} \sum_{j=j_1}^{j_2-1} \frac{\mathbf{R}_i - \mathbf{R}_j}{|\mathbf{R}_i - \mathbf{R}_j|^3} \cdot (d\mathbf{R}_i \times d\mathbf{R}_j). \quad (1.2)$$

Nella formula (1.2) una proteina con N residui è pensata come una catena discreta di monomeri ($i = 1, \dots, N$) aventi la posizione \mathbf{r}_i del carbonio C^α e si sono dunque definiti

$$\mathbf{R}_i \equiv \frac{1}{2} (\mathbf{r}_{i+1} + \mathbf{r}_i) \quad d\mathbf{R}_i = (\mathbf{r}_{i+1} - \mathbf{r}_i) \quad (1.3)$$

i vettori utilizzati nella (1.2). [5]

Capitolo 2

Modelli a rete elastica

2.1 Fluttuazioni dello stato nativo

Dopo aver compiuto il processo di ripiegamento, la proteina non risulta statica in quanto gli atomi che la compongono sono soggetti a fluttuazioni termiche. Come già detto, lo stato nativo rappresenta un minimo dell'energia e di conseguenza le fluttuazioni termiche inducono delle variazioni nella struttura proteica che oscillerà attorno a questo. In un intorno del minimo dell'energia si viene quindi a formare un *ensemble* di microstati che si trovano in equilibrio dinamico. Un esempio di quanto accade per una determinata struttura proteica è raffigurato in Figura 2.1 dove si vede come nel minimo dell'energia dello stato nativo, osservando ad una maggiore risoluzione, si possano riconoscere due differenti sottostati, i quali a loro volta sono costituiti da microstati, riscontrabili ad una risoluzione ancora maggiore. Queste variazioni non influiscono sulle strutture secondarie e terziarie ma provocano uno spostamento dei singoli atomi che può essere osservato in termini di alterazioni della lunghezza dei legami e dei loro angoli, della conformazione dei loop e della posizione e orientazione di interi domini o sottounità proteiche. [6]

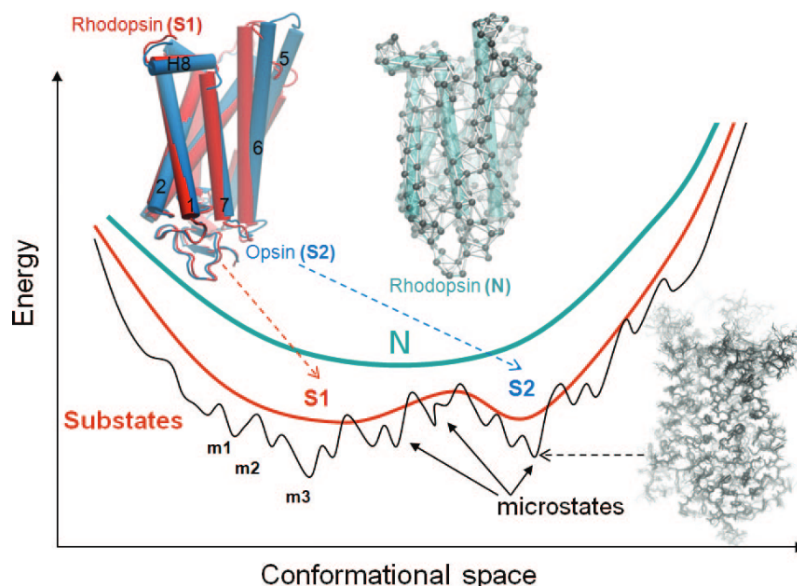


Figura 2.1: Profilo energetico dello stato nativo a differenti risoluzioni [6]

Per lo studio di questi fenomeni si ricorre a metodi di *principal component analysis* (PCA), come ad esempio analisi ai modi normali, dinamica molecolare e simulazioni Monte Carlo. Tramite questi è stato possibile verificare come queste fluttuazioni non siano affatto casuali ma inducano movimenti collettivi e coordinati che sono quindi rilevanti da un punto di vista funzionale. Svolgono importanti funzioni biologiche in quanto non si limitano a rendere possibili i movimenti dei domini

e delle sottounità proteiche ma li conducono e li coordinano attivamente. Risultano fondamentali nello svolgimento delle funzioni di catalisi e in particolare nella regolazione allosterica, rendendo possibile il legame con il substrato. Le funzioni delle proteine di membrana, ad esempio, che svolgono attività di trasmissione dei segnali, apertura dei pori e trasferimento degli ioni, sono rese possibili da questi moti collettivi. La rilevanza delle fluttuazioni evidenzia come la loro presenza sia necessariamente dovuta alla struttura stessa della proteina. La dinamica all'equilibrio risulta influenzata dalla struttura della proteina analogamente alla determinazione dello stato nativo a partire dalla sequenza di amminoacidi. Questi movimenti collettivi sono quindi codificati nella struttura tridimensionale e di conseguenza nella topologia dello stato nativo. [6]

2.2 Analisi ai modi normali

L'analisi ai modi normali (NMA) è tra i metodi PCA più utilizzati. In questo metodo si assume che le fluttuazioni attorno allo stato nativo, dovute all'agitazione termica, siano governate da un potenziale armonico al fine di studiare i modi normali di oscillazione. Data la scelta del potenziale, questo metodo risulta valido in prossimità dell'equilibrio dove l'effettivo potenziale a cui sono soggetti gli atomi della proteina può essere approssimato con quello armonico. Si trascurano quindi altri vincoli che la struttura proteica impone alle deformazioni, come ad esempio la lunghezza dei legami o gli angoli di legame, e questo contribuisce a rendere l'assunzione del potenziale valida unicamente in prossimità dell'equilibrio. L'analisi ai modi normali risulta dunque utile soprattutto per lo studio di variazioni della configurazione che avvengono su larga scala in quanto le specifiche interazioni locali, in particolare quelle di natura elettromagnetica, non sono prese in considerazione. Questo tipo di variazioni, indicate come modi globali, sono di notevole interesse poiché sono quelle che risultano biologicamente funzionali. I modi globali rappresentano le riconfigurazioni lungo le direzioni più facilmente accessibili energeticamente in quanto richiedono la minima quantità di energia per dar luogo ad una data deformazione. [6]

Si considera una proteina con N residui e si assume che per ognuno di questi il centro di interazione del potenziale risieda nel carbonio C^α . Per ognuna delle possibili configurazioni, si associa al carbonio C^α del residuo i -esimo un vettore posizione $\mathbf{R}_i = (x_i, y_i, z_i)^T$. Il vettore associato alla configurazione può quindi essere scritto come

$$\mathbf{q} = ((\mathbf{R}_1)^T, \dots, (\mathbf{R}_N)^T)^T = (x_1, y_1, z_1, \dots, x_N, y_N, z_N)^T. \quad (2.1)$$

Una volta indicata con l'apice 0 la configurazione di equilibrio corrispondente allo stato nativo, è possibile definire il vettore di spostamento $\Delta\mathbf{q} = \mathbf{q} - \mathbf{q}^0$ che descrive la variazione della posizione $\Delta\mathbf{R}_i = \mathbf{R}_i - \mathbf{R}_i^0$ di ognuno degli N siti. Se l'interazione è dominata da un potenziale V , questo può essere espanso come serie di potenze di \mathbf{q} vicino all'equilibrio

$$V(\mathbf{q}) = V(\mathbf{q}^0) + \sum_i \left(\frac{\partial V}{\partial q_i} \right)^0 (q_i - q_i^0) + \frac{1}{2} \sum_{ij} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)^0 (q_i - q_i^0)(q_j - q_j^0) + \dots \quad (2.2)$$

Il primo termine rappresenta il valore minimo dell'energia e può essere arbitrariamente annullato, mentre il secondo termine deve essere nullo in quanto ci si è posti in una condizione di equilibrio. È possibile quindi riscrivere il potenziale nella forma

$$\begin{aligned} V(\mathbf{q}) &= \frac{1}{2} \sum_{ij} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)^0 (q_i - q_i^0)(q_j - q_j^0) \\ &= \frac{1}{2} \sum_{ij} (q_i - q_i^0) H_{ij} (q_j - q_j^0) = \frac{1}{2} \Delta\mathbf{q}^T \mathbf{H} \Delta\mathbf{q} \end{aligned} \quad (2.3)$$

dove si è introdotta la matrice \mathbf{H} di elementi

$$H_{ij} = \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)^0. \quad (2.4)$$

Questa è per costruzione reale e simmetrica. Essendo calcolata a partire da un minimo dell'energia, tutti i suoi autovalori dovranno risultare positivi in quanto rappresentano proprio la curvatura del potenziale e sono positivi per i punti di minimo. È inoltre caratterizzata da sei autovalori nulli che indicano un'assenza di variazione nella configurazione e possono essere quindi associati alle tre traslazioni rigide e alle tre rotazioni rigide.

Al fine di studiare la dinamica della configurazione è necessario introdurre la componente di energia cinetica. Se si assume che i siti siano costituiti da particelle di massa m caratterizzate da un comportamento classico, l'equazione del moto può essere scritta come

$$\mathbf{M} \frac{d^2 \Delta \mathbf{q}}{dt^2} + \mathbf{H} \Delta \mathbf{q} = 0 \quad (2.5)$$

dove \mathbf{M} è una matrice diagonale contenente le masse delle particelle, ognuna ripetuta tre volte. Una soluzione all'equazione (2.5) è fornita dal vettore $\mathbf{u}_k(t) = \mathbf{a}_k \exp(-i\omega_k t)$ dove \mathbf{a}_k è un vettore che contiene l'ampiezza e la fase e ω_k rappresenta la frequenza del modo k -esimo. Sostituendo questa soluzione nell'equazione (2.5), l'equazione del moto diventa

$$\mathbf{H} \mathbf{u}_k = \omega_k^2 \mathbf{M} \mathbf{u}_k. \quad (2.6)$$

Dopo aver assunto che le masse siano tutte uguali e unitarie, si possono riconoscere nei vettori \mathbf{u}_k gli autovettori della matrice \mathbf{H} e in $\lambda_k = \omega_k^2$ i suoi autovalori. Si può vedere come per il singolo modo k -esimo, l'equazione (2.3) può essere riscritta come

$$V(\mathbf{u}_k) = \frac{1}{2} \mathbf{u}_k^T \mathbf{H} \mathbf{u}_k = \frac{\omega_k^2}{2}. \quad (2.7)$$

Osservando l'equazione (2.7) si ricava che l'energia è direttamente proporzionale al quadrato della frequenza. I movimenti lungo modi ad alta frequenza sono quindi energeticamente più dispendiosi rispetto a quelli di uguale entità ma lungo modi a bassa frequenza. Sono proprio questi modi a dare luogo alle deformazioni su larga scala e corrispondono quindi ai modi globali nei quali, come già detto, risiede l'interesse biologico.

2.3 Modello anisotropo

Il punto cruciale dell'analisi ai modi normali consiste nell'individuazione della forma corretta della matrice \mathbf{H} che è direttamente collegata alla scelta del potenziale. Per tener conto della complessità delle interazioni all'interno della proteina è possibile scegliere un potenziale semiempirico che prenda in considerazione i differenti contributi. La complessità di potenziali di questo tipo risulta tuttavia problematica in quanto rende necessaria una minimizzazione preliminare dell'energia per assicurare che la derivata prima del potenziale sia nulla. La minimizzazione infatti richiede un elevato sforzo computazionale e produce una configurazione differente dallo stato nativo. Si è verificato che scegliendo, invece del potenziale semiempirico, un semplice potenziale elastico, dipendente da un unico parametro, è possibile produrre risultati del tutto analoghi. I risultati prodotti con un potenziale elastico risultano inoltre più accurati grazie alla mancata minimizzazione, al minor numero di parametri nel potenziale e all'assenza di altre problematiche dovute ad una differente scelta del potenziale. A rendere efficace questa scelta è il fatto che i modi di interesse sono i modi lenti e coordinati che coinvolgono ampie regioni. Le forze che effettivamente governano questi movimenti sono il risultato della sovrapposizione delle specifiche interazioni che possono essere quindi singolarmente trascurate. [7]

Per l'analisi ai modi normali si introducono quindi dei modelli a rete elastica (ENM). A partire dalla struttura è possibile modellizzare la proteina costruendo una rete elastica di masse accoppiate a delle molle come raffigurato in Figura 2.2. La topologia della rete è definita dallo stato nativo e i nodi della rete vengono collegati tramite delle molle qualora la loro distanza reciproca sia inferiore ad un raggio di cutoff prestabilito. Si è visto inoltre che la scelta della costante elastica delle molle ha effetti molto ridotti sui risultati prodotti dall'analisi. [7] I principali modelli a rete

elastica sono il modello anisotropo (ANM) e il *gaussian network model* (GNM). Quest'ultimo si basa sull'assunzione che le fluttuazioni di ogni residuo siano distribuite in modo gaussiano attorno all'equilibrio e utilizza un potenziale la quale forma fa sì che le fluttuazioni siano isotrope. Di conseguenza tale metodo non fornisce informazioni sulle direzioni del moto nei differenti modi. [8]

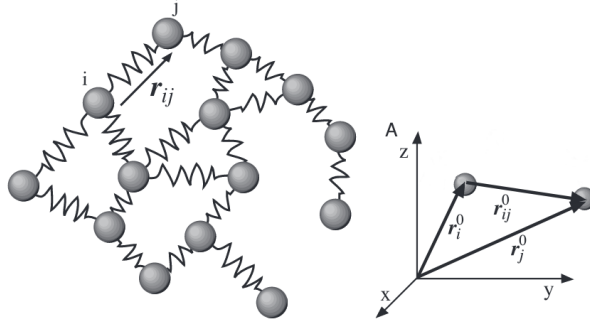


Figura 2.2: Raffigurazione del modello a rete elastica [8]

Per questo lavoro di tesi si è scelto di prendere in considerazione il modello anisotropo nel quale il potenziale scelto ha la forma

$$V = \frac{1}{2} \sum_{ij} \gamma_{ij} (R_{ij} - R_{ij}^0)^2. \quad (2.8)$$

Il metodo è detto anisotropo in quanto osservando l'equazione (2.8) si nota che il potenziale è nullo se $R_{ij} = R_{ij}^0$, indifferentemente dalla direzione dei corrispondenti vettori distanza. Con questo metodo risulta facile scrivere l'espressione della matrice \mathbf{H} utilizzando le equazioni (2.8) e (2.4). In questo caso la derivata seconda del potenziale è semplicemente

$$\frac{\partial^2 V}{\partial x_i \partial y_j} = -\gamma_{ij} \frac{(x_j - x_i)(y_j - y_i)}{R_{ij}^2}. \quad (2.9)$$

Usando la notazione $x_{ij}^0 = (x_j^0 - x_i^0)$ per la prima componente del vettore distanza \mathbf{R}_{ij}^0 e notazioni analoghe per y_{ij}^0 e z_{ij}^0 , si possono scrivere le sottomatrici 3×3 che si trovano fuori dalla diagonale di \mathbf{H} come

$$\mathbf{H}_{ij} = -\frac{\gamma_{ij}}{(R_{ij}^0)^2} \begin{bmatrix} (x_{ij}^0)^2 & x_{ij}^0 y_{ij}^0 & x_{ij}^0 z_{ij}^0 \\ x_{ij}^0 y_{ij}^0 & (y_{ij}^0)^2 & y_{ij}^0 z_{ij}^0 \\ x_{ij}^0 z_{ij}^0 & y_{ij}^0 z_{ij}^0 & (z_{ij}^0)^2 \end{bmatrix} \quad (2.10)$$

mentre quelle sulla diagonale come

$$\mathbf{H}_{ii} = -\sum_{j:j \neq i} \mathbf{H}_{ij}. \quad (2.11)$$

Con questo metodo è possibile generare conformazioni vicine alla struttura data deformandola lungo i modi dominanti. La configurazione deformata lungo il modo k -esimo risulta essere

$$\mathbf{q}^{(k)} = \mathbf{q}^0 \pm s \lambda_k^{-1/2} \mathbf{u}_k. \quad (2.12)$$

Il parametro s dovrebbe scalare come $\sqrt{k_B T}$. La fluttuazione quadratica a temperatura T del residuo i -esimo lungo un dato modo k -esimo è data in termini dell'autovettore \mathbf{u}_k e dell'autovalore λ_k della matrice \mathbf{H} come

$$(\Delta \mathbf{R}_i)^2 \Big|_k = k_B T \operatorname{tr} \{ [\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ii} \} \quad (2.13)$$

e da conto della mobilità di un residuo lungo un determinato modo. Sommando su tutti i modi è possibile ottenere il coefficiente di mobilità che tiene conto della mobilità complessiva del residuo. Inoltre, poiché si è interessati in particolar modo alle deformazioni che riguardano ampie regioni della proteina, è utile andare a valutare quali modi presentino un grado di collettività più elevato.

Il grado di collettività per una deformazione lungo il modo k -esimo può essere valutato tramite il parametro κ_k definito come

$$\kappa_k = \frac{1}{N} \exp \left(- \sum_{i=1}^N \alpha (\Delta \mathbf{R}_i)^2 \Big|_k \log(\alpha (\Delta \mathbf{R}_i)^2) \Big|_k \right) \quad (2.14)$$

dove α è la costante di normalizzazione tale per cui $\sum_i \alpha (\Delta \mathbf{R}_i)^2 \Big|_k = 1$. [6] L'equazione (2.14) rappresenta l'entropia di Shannon definita come

$$S = - \sum_{i=1}^N p_i \ln p_i \quad (2.15)$$

con p_i la probabilità di localizzazione del modo su ogni residuo. L'entropia sarà nulla qualora la deformazione risulti localizzata su un unico residuo, mentre assumerà il valore $S = \ln N$ quando sarà massimamente distribuita su tutti gli N residui. Avendo definito il grado di collettività come $\kappa_k = 1/N \exp(S)$ si vede che questo avrà un valore minimo $\kappa_k = 1/N$ (che per $N \rightarrow \infty$ tenderà ad annullarsi) qualora $S = 0$ e un valore massimo $\kappa_k = 1$ per $S = \ln N$. Si vede che il modo con il più alto grado di collettività avrà la più elevata entropia e sarà quindi distribuito su un maggior numero di residui.

Capitolo 3

Studio di strutture proteiche

Si è interessati a studiare gli effetti delle deformazioni della struttura proteica sull'entanglement della proteina. In particolare si vuole monitorare come varia il valore dell'entanglement gaussiano, calcolato fra due porzioni della catena, al deformarsi della struttura. A tale scopo si è applicata l'analisi ai modi normali con modello anisotropo per lo studio delle proteine 2hqx e 3cry, selezionate fra quelle che presentano un notevole grado di entanglement. Si sono inoltre osservati gli effetti ottenuti aumentando il valore della costante elastica fra i residui terminali di loop selezionati come quelli maggiormente coinvolti nell'entanglement della struttura, in quanto l'analisi di questi contatti risulta interessante per comprendere il processo di ripiegamento. L'intensità di questi e la fase del processo di ripiegamento in cui si instaurano possono infatti essere cruciali; se questi, ad esempio, dovessero venirsi a creare troppo precocemente potrebbero compromettere il corretto ripiegamento della struttura. [4]

3.1 Analisi di 2hqx

Si è analizzata in primo luogo la proteina 2hqx che è una proteina umana composta da due catene polipeptidiche, A e B, e tra queste si è selezionata la catena A, composta da 86 residui. Si sono reperite le coordinate spaziali dell'atomo di carbonio C^α di ogni residuo nel Protein Data Bank, dove sono riportate in Å. Da questo si sono estratti anche i coefficienti B-factor, associati ad ogni atomo e riportati in Å², che indicano il grado di mobilità dei singoli atomi ricavato sperimentalmente dal cristallografo.

Si noti che di seguito per indicare il carbonio C^α di un residuo ci si riferirà per brevità al residuo stesso. Si tenga inoltre in considerazione che nella numerazione dei modi e dei residui al primo elemento è stato associato l'indice zero.

A partire dalle coordinate spaziali degli atomi C^α e dalle equazioni (2.10) e (2.11), si è calcolata la matrice \mathbf{H} . Qualora la distanza tra i due residui risultasse maggiore del valore scelto per il raggio di cutoff, $r_c = 18$ Å, si è assegnato un valore nullo alla costante elastica. A questa si è invece attribuito valore unitario¹ per residui a distanza inferiore del raggio di cutoff con l'eccezione dei residui consecutivi per i quali si è scelto un valore dieci volte maggiore. In questo modo si è tenuta in considerazione la forza del legame peptidico. Una volta ottenuta la matrice si è proceduto con il calcolo degli autovalori e autovettori. Ordinando gli autovalori in modo crescente si vede che i primi sei risultano notevolmente più piccoli degli altri, dell'ordine di 10^{-14} e 10^{-15} , e corrispondono ai sei autovalori nulli attesi.

Al fine di ottenere il profilo di mobilità della proteina, si è calcolato il coefficiente di mobilità m per ogni residuo a partire dall'equazione (2.13) come

$$m_i = \sum_{k=6}^N (\Delta \mathbf{R}_i)^2 \Big|_k \quad (3.1)$$

¹L'effettivo valore della costante elastica risulta in pratica determinato dalla procedura di normalizzazione utilizzata in Figura A.1 per confrontare le mobilità teoriche con i B-factor sperimentali.

dove si è fatta partire la sommatoria dal primo modo con autovalore non nullo. Si sono quindi confrontati graficamente i dati sperimentali con quelli calcolati, sovrapponendo in Figura A.1 i B-factor con i valori di m dopo aver effettuato un'opportuna normalizzazione di quest'ultimi. Osservando il grafico in Figura A.1 si vede che il profilo di mobilità calcolato segue in modo soddisfacente il dato sperimentale. Si è poi verificato che i risultati fossero stabili al variare del raggio di cutoff. In Figura A.2 si sono quindi riportati i profili di mobilità ottenuti impostando $r_c = 12, 14, 16, 18, 20$ Å. È possibile verificare che la variazione del raggio di cutoff attorno al valore $r_c = 18$ Å, indicato in letteratura come ottimale per il modello anisotropo, non comporta variazioni significative dei profili di mobilità. [6] Si è quindi scelto di procedere con il valore del raggio di cutoff precedentemente scelto in quanto produce il profilo di mobilità che più si avvicina ai dati sperimentali. Si sono calcolati infine i gradi di collettività κ_k come indicato nell'equazione (2.14) per i modi con autovalore non nullo e si sono riportati i risultati in Figura A.3.

Si è quindi considerato l'entanglement della proteina. Come si può vedere in Figura 3.1, questa presenta un loop che ha per estremi i residui 10 e 39 all'interno del quale passa un'altra parte di catena della quale si prendono in considerazione la porzione 40-78 che non forma un loop e le porzioni 42-80 e 42-75 che invece costituiscono a loro volta un loop. Si sono scelte queste coppie di sottocatene in quanto (10-39,40-78) è la coppia di sottocatene (di cui solo una è aperta) per cui il valore di $|G'|$ risulta massimo, (10-39,42-80) è invece la coppia di sottocatene chiuse per cui $|G'|$ risulta massimo e (10-39,42-75) è la coppia di sottocatene chiuse contenute in (10-39,40-78) per cui $|G'|$ è massimo. Al fine di misurare l'entanglement fra queste porzioni della catena si è calcolato l'entanglement gaussiano come prescritto dall'equazione (1.2) e si sono ottenuti i seguenti valori: $G' (10-39,40-78) = -1.240$, $G' (10-39,42-80) = -1.010$ e $G' (10-39,42-75) = -0.896$.

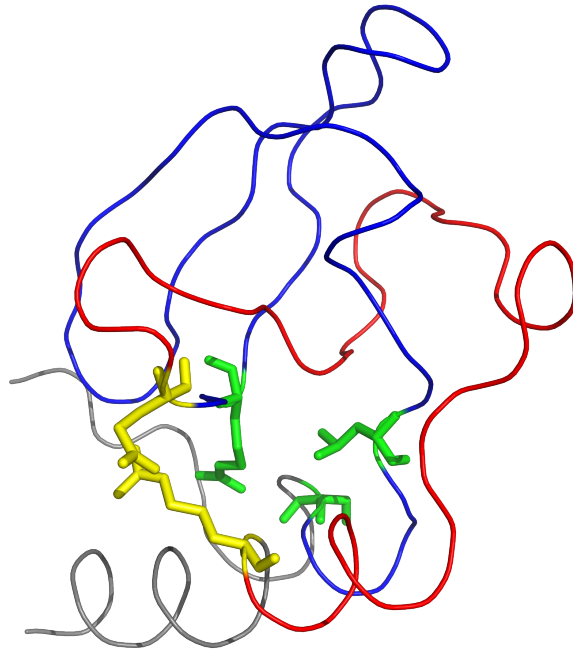


Figura 3.1: Raffigurazione delle sottocatene di interesse della catena A di 2hqx. In rosso la catena 10-39 con i residui terminali a contatto in giallo. In blu la catena 40-80 con i residui a contatto 42, 75 e 80 in verde.

Per scegliere lungo quali modi deformare la proteina si è preso in considerazione il grado di mobilità. Poiché come noto ci si aspetta che i modi lenti siano più collettivi si sono considerati i primi modi non nulli $k = 6, 7, 8, 9, 10, 11$ in quanto presentano autovalori più separati degli altri. Tra questi inoltre i modi 8 e 9 hanno il grado di collettività maggiore come si può vedere in Figura A.3. Il modo con il più alto grado di collettività in assoluto è il modo 121 che, tuttavia, osservando il suo autovettore, presenta una natura casuale e non particolarmente collettiva come atteso per i modi non lenti. Per un'analisi più approfondita si sono andati a considerare i profili di

mobilità separatamente per i primi sei modi non nulli. Si è utilizzata a questo scopo l'equazione (2.13) a k fissato e si sono riportati i risultati in Figura A.4. È possibile osservare come i modi 6 e 7 mostrino fluttuazioni concentrate sugli estremi, risultando quindi scarsamente collettivi, e il modo 11 risulti invece un modo localizzato. Si sono quindi scelti i modi 8 e 9.

Si sono prodotte le deformazioni lungo questi modi utilizzando l'equazione (2.12). Per la scelta di s si è preso un valore s_{max} che determinasse una variazione della distanza del residuo con il più elevato coefficiente di mobilità (il 50-esimo come si vede in Figura A.1) dal centro di massa della proteina di qualche punto percentuale (si è scelto il 10%) in modo da assicurarsi di lavorare intorno alla configurazione non troppo deformata. Si sono quindi scelti 200 valori di s equidistribuiti nell'intervallo $[-s_{max}, s_{max}]$. Per ognuno di questi si è determinato il vettore configurazione deformato e si sono calcolati i valori dell'entanglement gaussiano. Si è poi provato a bloccare separatamente gli estremi dei loop in contatto, 10-39, 42-80 e 42-75, assegnando alla molla che li collega una costante elastica pari a quella assegnata al legame peptidico. I risultati ottenuti sono riportati in Figura A.5, Figura A.6 e Figura A.7 dove è possibile vedere come varia il valore dell'entanglement gaussiano al variare di s nei casi senza molla aggiuntiva (indicati con n.m.) e nei casi con molla aggiuntiva (m.) per i due modi scelti. Si sono inoltre monitorate le distanze fra gli estremi dei loop 10-39, 42-80 e 42-75 i valori delle quali al variare di s nelle differenti configurazioni sono riportati in Figura A.8, Figura A.9 e Figura A.10.

Infine si è osservata la variazione degli autovalori considerati prodotta decuplicando il valore della costante elastica di volta in volta per una differente coppia di residui fra quelle connesse da molle nella rete elastica. Rafforzando la molla si ottiene un irrigidimento della rete elastica che essendo meno mobile presenta autovalori più grandi. La variazione degli autovalori, ottenuta bloccando i contatti terminali degli specifici loop studiati in questo lavoro, risulta in linea con quella prodotta agendo sulle altre coppie di residui.

3.2 Analisi di 3cry

Si è analizzata poi la proteina 3cry che è anch'essa una proteina umana composta da due catene, A e B, delle quali si è selezionata la catena A, composta da 169 residui. Questa presenta un elevato grado di entanglement.

Si è calcolata la matrice \mathbf{H} , scegliendo in questo caso $r_c = 20 \text{ \AA}$ per tener conto del maggior numero di residui di 3cry rispetto a 2hqx, e si è proceduto con l'analisi in modo del tutto analogo a quanto fatto per la proteina precedente. In Figura A.11 e Figura A.12 si possono osservare rispettivamente i profili di mobilità, che anche in questo caso presentano una buona concordanza tra il dato calcolato e quello sperimentale, e i gradi di collettività. Per quanto riguarda l'entanglement questa proteina presenta un loop che ha per estremi i residui 51 e 122 all'interno del quale passa un'altra parte di catena della quale si prendono in considerazione la porzione 1-50 che non forma un loop e la porzione 10-50 che invece costituisce a sua volta un loop, come si può vedere in Figura 3.2. Anche in questo caso si sono scelte queste coppie di sottocatene poiché sono quelle per cui $|G'|$ è massimo, in maniera analoga a quanto descritto per 2hqx. Si sono poi calcolati i valori dell'entanglement gaussiano ottenendo: $G'(51-122,1-50) = -3.067$ e $G'(51-122,10-50) = -2.239$.

Per scegliere lungo quali modi deformare la proteina si è preso in considerazione il grado di mobilità con un procedimento analogo a quello per la proteina 2hqx. In questo caso è stato necessario valutare anche i risultati ottenuti eliminando gli ultimi dieci residui. A partire dai B-factor è infatti possibile osservare che questi residui sono dotati di elevata mobilità. Dalla struttura tridimensionale si vede inoltre che questi compongono una α -elica che risulta abbastanza separata dal resto della struttura, eliminandoli risulta quindi più facile individuare i modi di interesse. Si sono scelti in questo caso i modi $k = 7, 8, 9$ e lungo questi si sono prodotte le deformazioni. Si è inoltre nuovamente proceduto a bloccare gli estremi dei loop in contatto, 51-122 e 10-50. In Figura A.13 e Figura A.14 è possibile osservare l'andamento dei valori dell'entanglement gaussiano per i tre modi scelti e nelle varie configurazioni. L'andamento delle distanze fra gli estremi dei loop può essere invece osservato in Figura A.15 e Figura A.16.

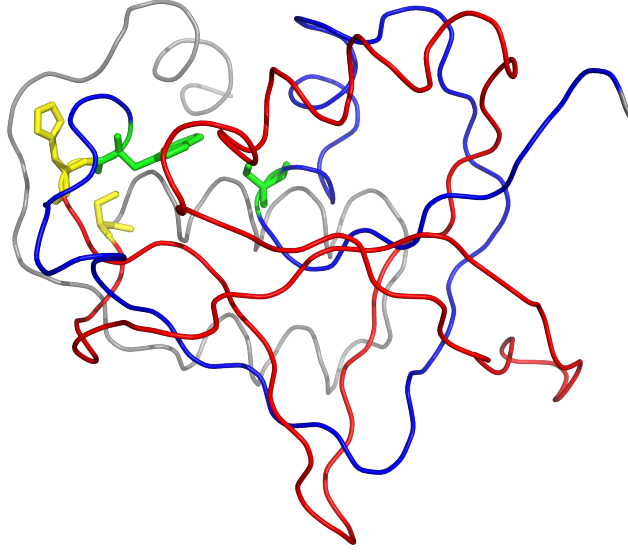


Figura 3.2: Raffigurazione delle sottocatene di interesse della catena A di 3cry. In rosso la catena 51-122 con i residui terminali a contatto in giallo. In blu la catena 1-50 con i residui a contatto 10 e 50 in verde

Al fine di confrontare più agevolmente la variazione dei valori dell'entanglement gaussiano e delle distanze se ne sono calcolate le derivate rispetto al parametro s che controlla le deformazioni lungo i vari modi. Per quanto riguarda G' si è proceduto con un calcolo numerico considerando la derivata centrata

$$\left. \frac{dG'}{ds} \right|_{s=s_0} = \frac{G'(s_0 + \Delta s) - G'(s_0 - \Delta s)}{2\Delta s} \quad (3.2)$$

con $s_0 = 0$ e $\Delta s = 10^{-5}$. Si deve quindi calcolare $G'(\pm\Delta s)$ che è inteso come l'entanglement gaussiano a partire dal vettore configurazione deformato $\mathbf{q}^{(k)} = \mathbf{q}^0 \pm \Delta s \lambda_k^{-1/2} \mathbf{u}_k$. Si è calcolata poi la derivata seconda

$$\left. \frac{d^2G'}{ds^2} \right|_{s=s_0} = \frac{G'(s_0 + \Delta s) + G'(s_0 - \Delta s) - 2G'(s_0)}{\Delta s^2}. \quad (3.3)$$

Per quanto riguarda invece la distanza fra i residui i e j terminali dei un loop si ha che

$$\mathbf{d}(s) = \begin{pmatrix} q_{i,x}^0 \\ q_{i,y}^0 \\ q_{i,z}^0 \end{pmatrix} + s\lambda_k^{-1/2} \begin{pmatrix} u_{k,i,x} \\ u_{k,i,y} \\ u_{k,i,z} \end{pmatrix} - \left[\begin{pmatrix} q_{j,x}^0 \\ q_{j,y}^0 \\ q_{j,z}^0 \end{pmatrix} + s\lambda_k^{-1/2} \begin{pmatrix} u_{k,j,x} \\ u_{k,j,y} \\ u_{k,j,z} \end{pmatrix} \right] \quad (3.4)$$

Si può quindi scrivere $\mathbf{d}(s) = \mathbf{a} + s\mathbf{b}$ con

$$\mathbf{a} = \begin{pmatrix} q_{i,x}^0 - q_{j,x}^0 \\ q_{i,y}^0 - q_{j,y}^0 \\ q_{i,z}^0 - q_{j,z}^0 \end{pmatrix} \quad \text{e} \quad \mathbf{b} = \lambda_k^{-1/2} \begin{pmatrix} u_{k,i,x} - u_{k,j,x} \\ u_{k,i,y} - u_{k,j,y} \\ u_{k,i,z} - u_{k,j,z} \end{pmatrix}. \quad (3.5)$$

Si ha quindi che la distanza risulta $d(s) = \|\mathbf{d}\| = \|\mathbf{a} + s\mathbf{b}\|$ ed è possibile calcolare in modo analitico la derivata prima in zero come

$$d' = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|} \quad (3.6)$$

e la derivata seconda in zero come

$$d'' = \frac{\|\mathbf{b}\|^2}{\|\mathbf{a}\|} - \frac{(\mathbf{a} \cdot \mathbf{b})^2}{\|\mathbf{a}\|^3}. \quad (3.7)$$

In Tabella A.1, Tabella A.2 e Tabella A.3 si sono riportati i moduli delle derivate divise per il valore della funzione corrispondente in $s = 0$ in modo da valutare la variazioni in termini relativi.

3.3 Conclusioni

Dall'analisi dei grafici prodotti e dei valori delle derivate calcolati è possibile osservare come il grado di entanglement rimanga sostanzialmente invariato nelle differenti configurazioni assunte dalla proteina in un intorno dello stato nativo. In tutti i casi infatti i valori di $|G'|$ variano in un range ristretto rispetto al valore di riferimento. Osservando in particolare i valori delle derivate poi è possibile vedere come il valore dell'entanglement gaussiano sia meno dipendente da s rispetto alle distanze che comunque non variano in modo significativo. Si vede quindi che complessivamente le fluttuazioni lungo i modi lenti non inducono variazioni significative sul grado di entanglement della proteina e di conseguenza bloccare i residui a contatto non sortisce particolari effetti. L'analisi di altre proteine e l'utilizzo di modelli più sofisticati di rete elastica [9] saranno necessari per confermare questa conclusione in generale.

Bibliografia

- [1] A. Finkelstein and O. Ptitsyn. *Protein Physics: A Course of Lectures*. Academic Press, 2002.
- [2] E. J. Sophie, A. Suma, and C. Micheletti. How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. *Current opinion in structural biology*, 42:6–14, 2017.
- [3] K. W. Plaxco, K. T. Simons, I. Ruczinski, and D Baker. Topology, stability, sequence and length: Defining the determinants of two-state protein folding kinetics. *Biochemistry*, 39(37):11177–11183, 2000.
- [4] M. Baiesi, E. Orlandini, F. Seno, and A. Trovato. Evolutionary patterns detected in entangled proteins. *arXiv e-prints*, page arXiv:1809.02173, Sep 2018.
- [5] M. Baiesi, E. Orlandini, F. Seno, and A. Trovato. Exploring the correlation between the folding rates of proteins and the entanglement of their native states. *Journal of Physics A: Mathematical and Theoretical*, 50(50):504001, Nov 2017.
- [6] I. Bahar, T. R. Lezon, A. Bakan, and I. H. Shrivastava. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *Chemical reviews*, 110:1463–97, Nov 2009.
- [7] M. M. Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical review letters*, 77(9):1905–1908, Aug 1996.
- [8] C. Chennubhotla, A. J. Rader, L. W. Yang, and I. Bahar. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Physical biology*, 2:S173–80, Nov 2005.
- [9] C. Micheletti, P. Carloni, and A. Maritan. Accurate and efficient description of protein vibrational dynamics: Comparing molecular dynamics and gaussian models. *Proteins: Structure, Function, and Bioinformatics*, 55(3):635–645, 2004.

Appendice A

Grafici e tabelle

Tabella A.1: Derivate prime e seconde per il modo 7 (3cry)

	loop	n.m.	m. (10-50)	m. (51-122)
$(G')'$	(1-50)	0.000156	0.000159	0.000155
$(G')'$	(10-50)	0.0000389	0.0000329	0.0000403
d'	(51-122)	0.000349	0.000345	0.000195
d'	(10-50)	0.000708	0.000485	0.000706
$(G')''$	(1-50)	0.0000347	0.0000550	0.0000521
$(G')''$	(10-50)	0.0000119	0.0000119	0.00000397
d''	(51-122)	0.0514	0.0514	0.0514
d''	(10-50)	0.704	0.704	0.704

Tabella A.2: Derivate prime e seconde per il modo 8 (3cry)

	loop	n.m.	m. (10-50)	m. (51-122)
$(G')'$	(1-50)	0.0000855	0.0000914	0.0000830
$(G')'$	(10-50)	0.0000388	0.0000275	0.0000408
d'	(51-122)	0.000555	0.000553	0.000308
d'	(10-50)	0.00134	0.000926	0.00134
$(G')''$	(1-50)	0.0000347	0.0000231	0.000110
$(G')''$	(10-50)	0.00000793	0.0000357	0.0000516
d''	(51-122)	0.0514	0.0514	0.0514
d''	(10-50)	0.704	0.704	0.704

Tabella A.3: Derivate prime e seconde per il modo 9 (3cry)

	loop	n.m.	m. (10-50)	m. (51-122)
$(G')'$	(1-50)	0.0000333	0.0000348	0.0000343
$(G')'$	(10-50)	0.0000348	0.0000375	0.0000355
d'	(51-122)	0.000199	0.000199	0.000109
d'	(10-50)	0.000321	0.000221	0.000322
$(G')''$	(1-50)	0.0000319	0.0000463	0.000130
$(G')''$	(10-50)	0.0000119	0.0000198	0.0000278
d''	(51-122)	0.0514	0.0514	0.0514
d''	(10-50)	0.704	0.704	0.704

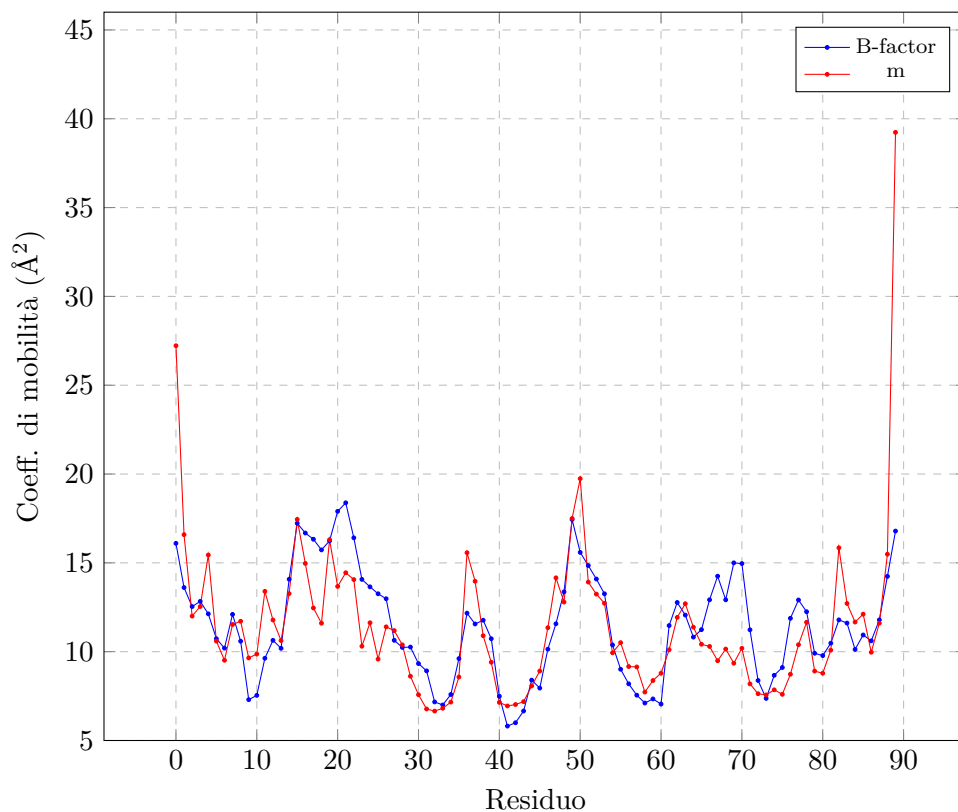


Figura A.1: Confronto tra i profili di mobilità sperimentale e calcolato (2hqx). Sono stati riportati i valori di m normalizzati ottenuti calcolando $(A_B/A_m)m_i$ dove i coefficienti A_B e A_m sono dati dalla somma su tutti i residui dei coefficienti di mobilità rispettivamente sperimentale e calcolato.

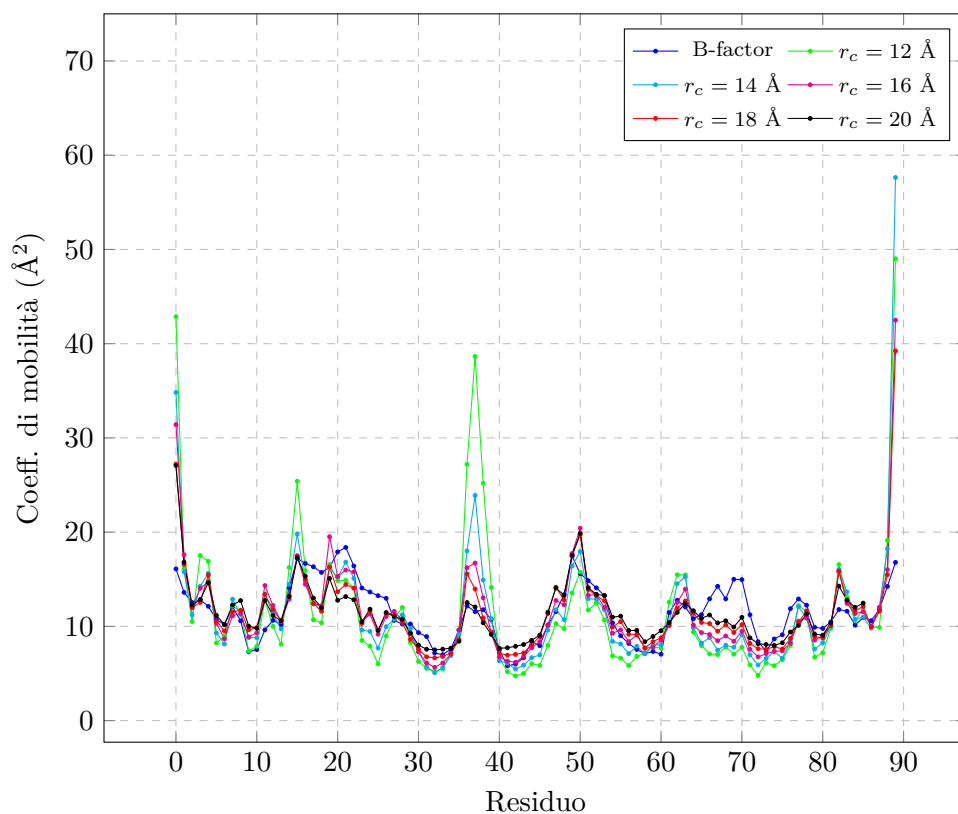


Figura A.2: Profili di mobilità al variare del raggio di cutoff (2hqx)

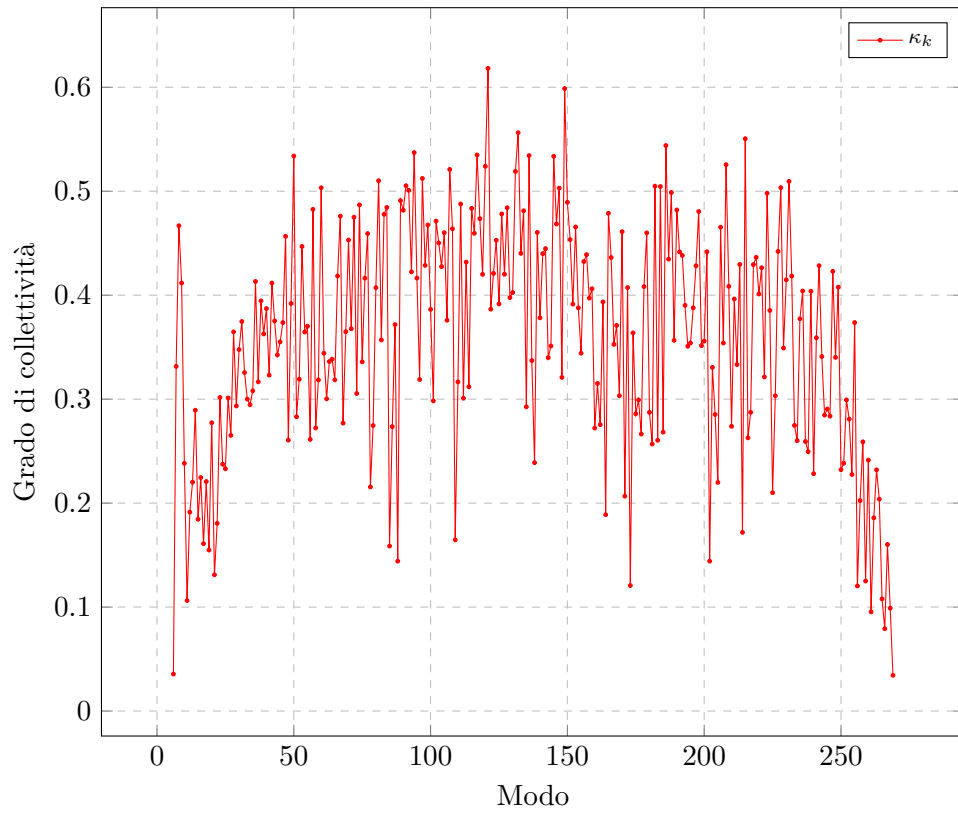


Figura A.3: Gradi di collettività (2hqx)

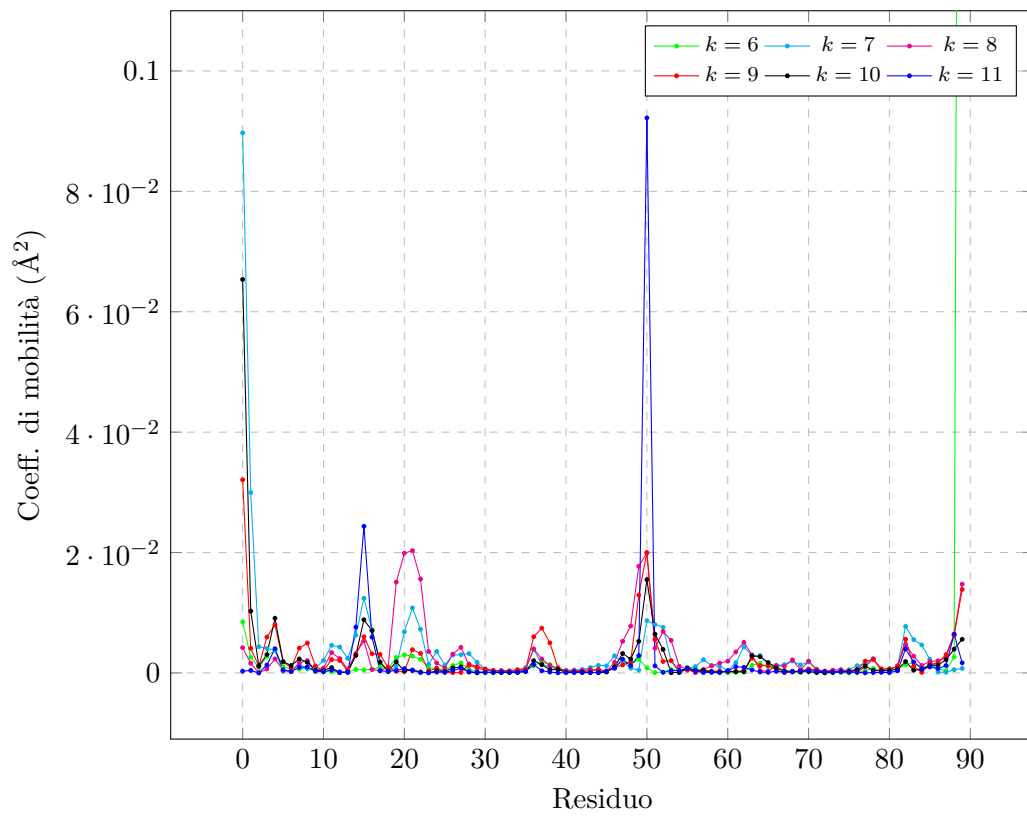


Figura A.4: Profili di mobilità a k fissato (2hqx)

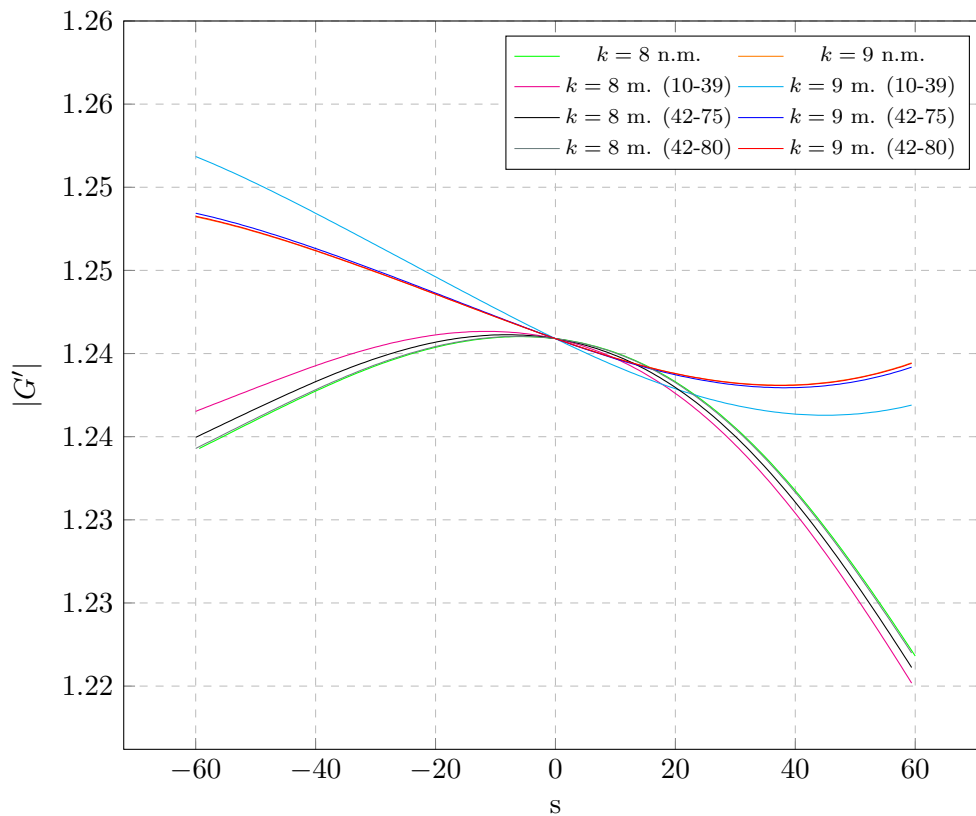


Figura A.5: Valori di $|G'|$ (10-39,40-78) al variare di s (2hqx)

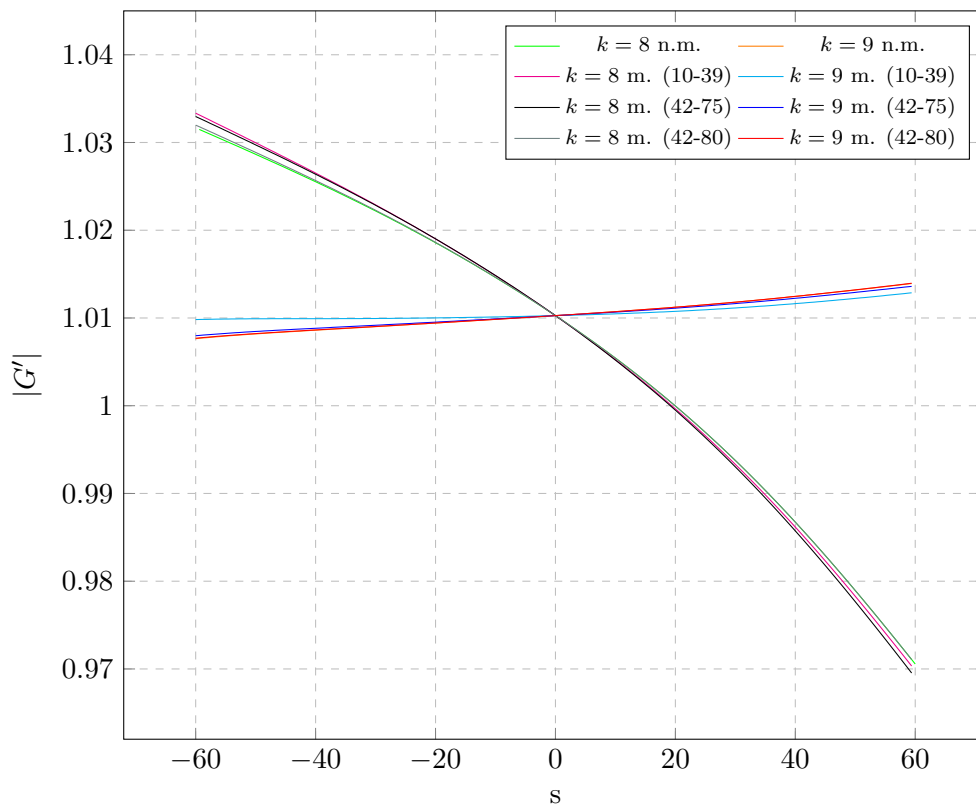


Figura A.6: Valori di $|G'|$ (10-39,42-80) al variare di s (2hqx)

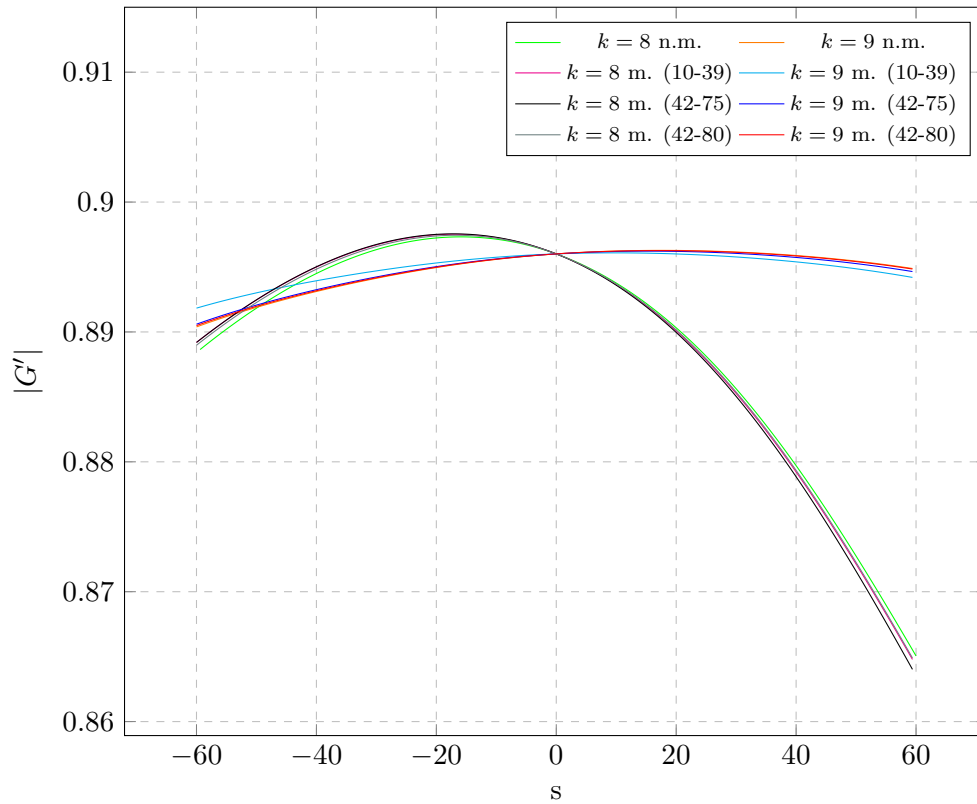


Figura A.7: Valori di $|G'|$ (10-39,42-75) al variare di s (2hqx)

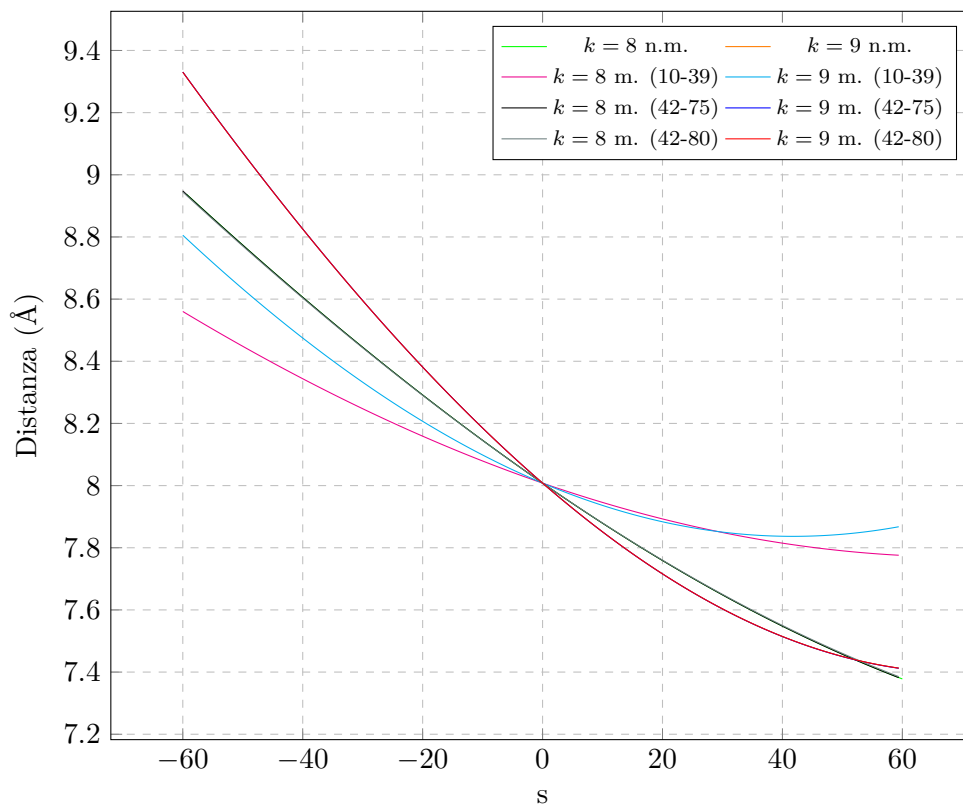


Figura A.8: Distanza tra i residui 10-39 al variare di s (2hqx)

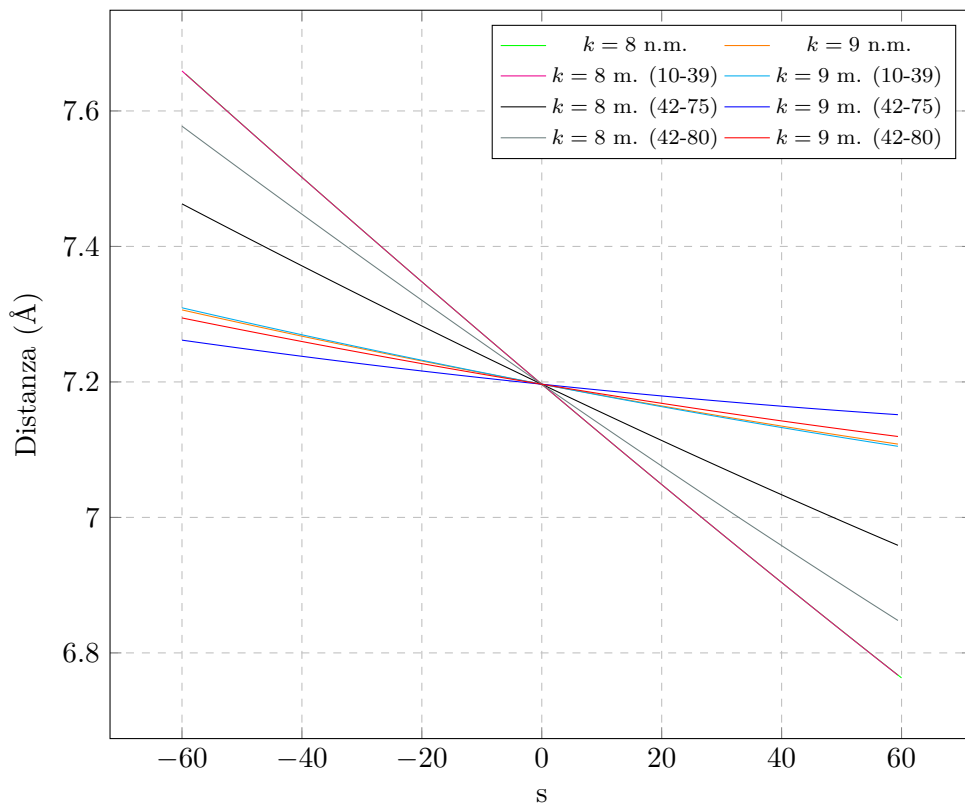


Figura A.9: Distanza tra i residui 42-75 al variare di s (2hqx)

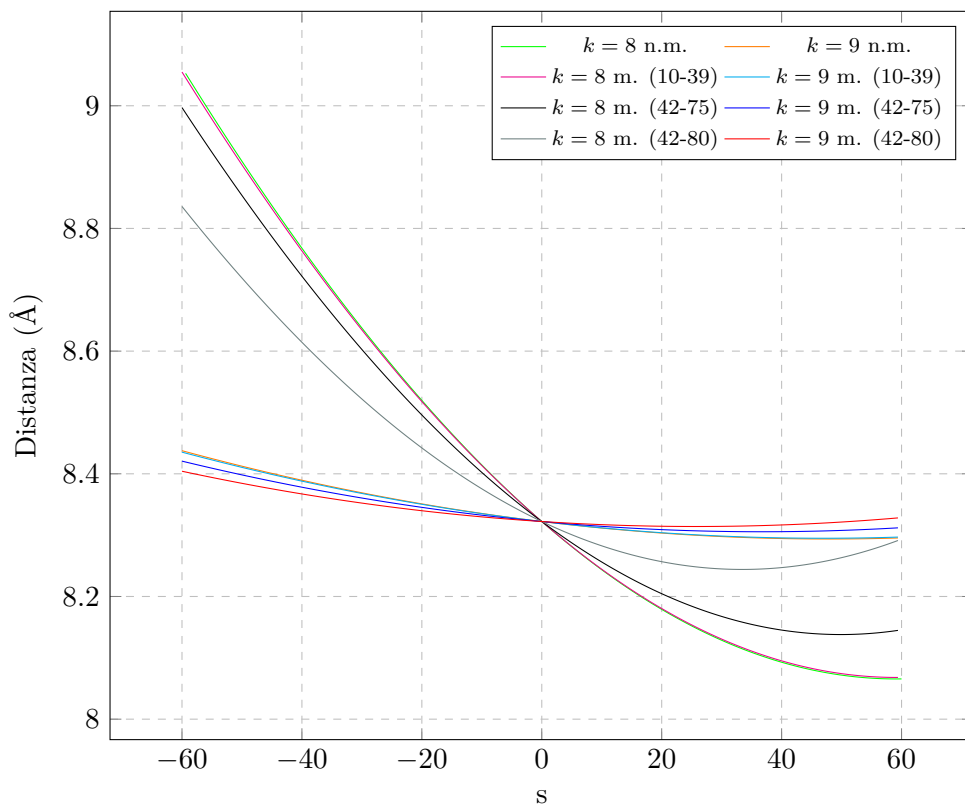


Figura A.10: Distanza tra i residui 42-80 al variare di s (2hqx)

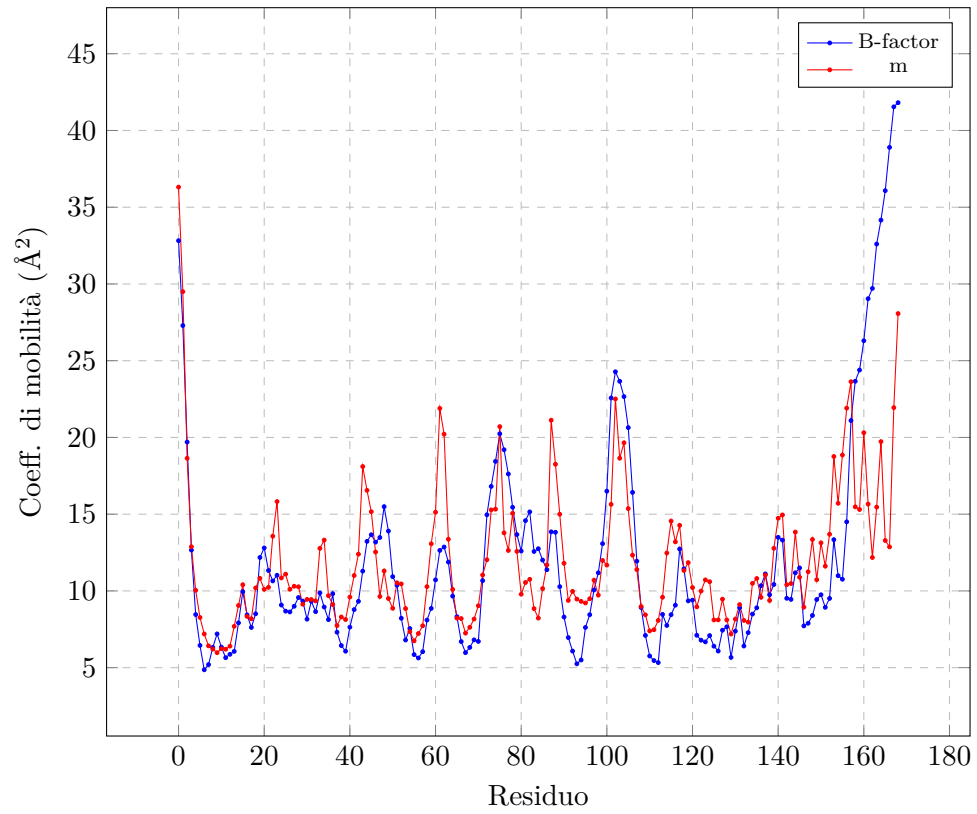


Figura A.11: Confronto tra i profili di mobilità sperimentale e calcolato (3cry)

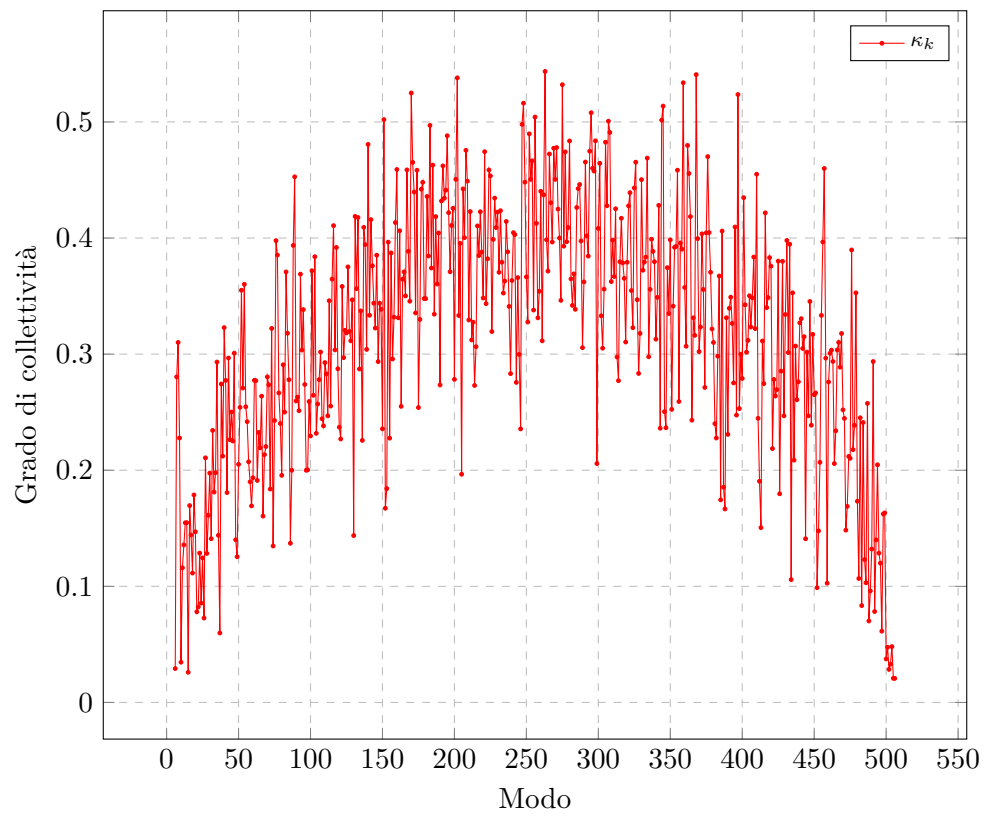
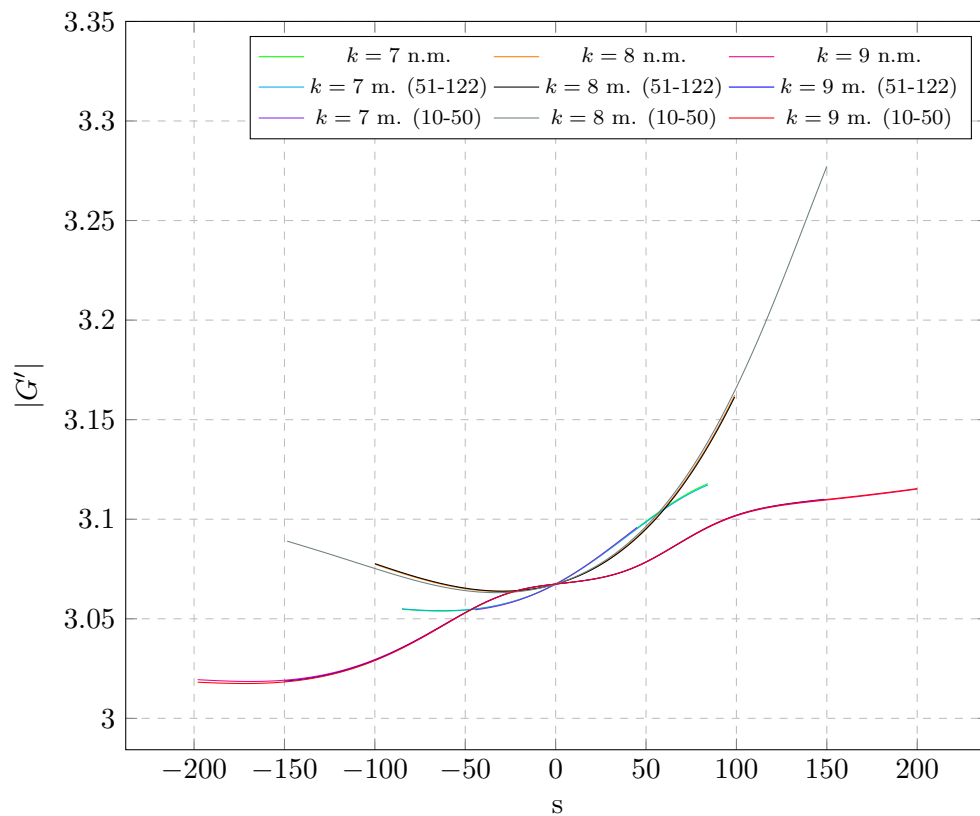
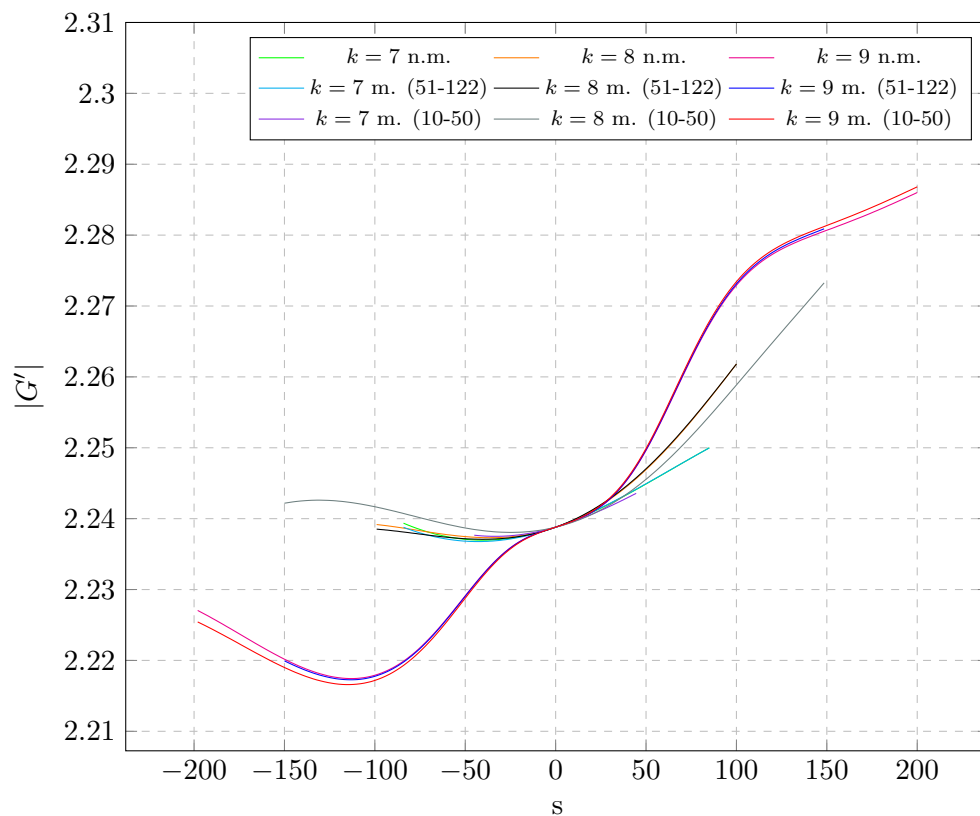


Figura A.12: Gradi di collettività (3cry)

Figura A.13: Valori di $|G'|$ (51-122,1-50) al variare di s (3cry)Figura A.14: Valori di $|G'|$ (51-122,10-50) al variare di s (3cry)

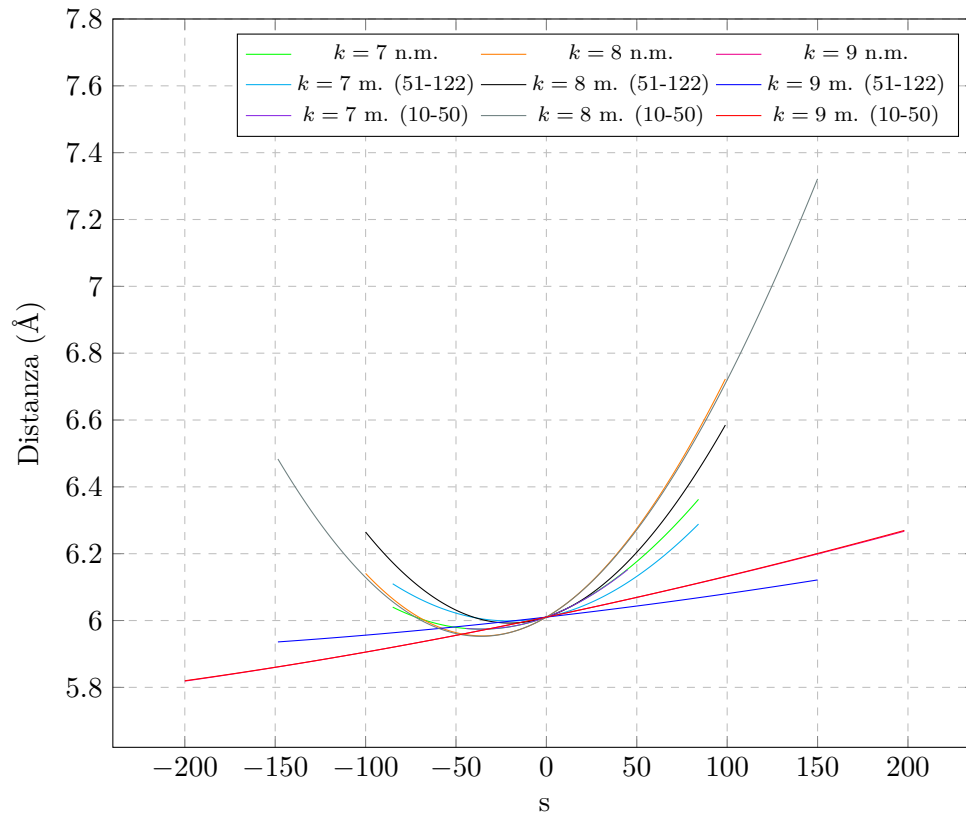


Figura A.15: Distanza tra i residui 51-112 al variare di s (3cry)

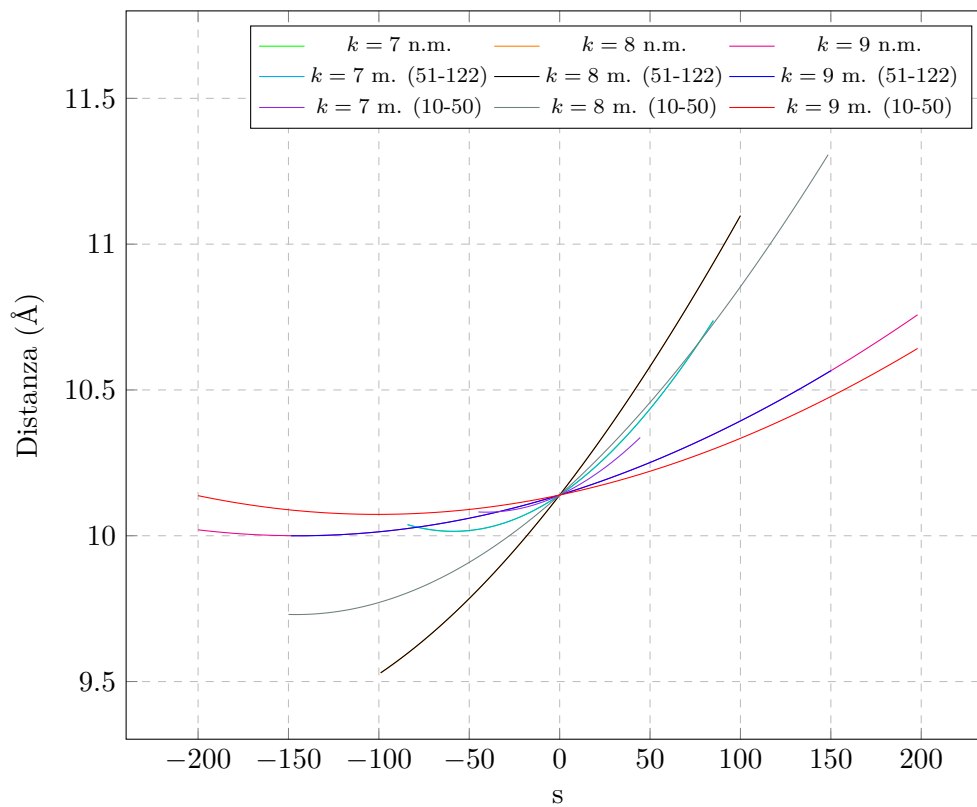


Figura A.16: Distanza tra i residui 10-50 al variare di s (3cry)