



Università degli Studi “*La Sapienza*”
Facoltà di Scienze Matematiche Fisiche e Naturali
Corso di Laurea in Chimica

STUDIO DEL FOLDING DI ANALOGHI DELLA GRAMICIDINA S MEDIANTE SIMULAZIONI DI DINAMICA MOLECOLARE

Laurea Specialistica

Relatore	Relatore esterno	Laureando
Alfredo Di Nola	Andrea Amadei	Giuliano Carchini

matricola 692574

Anno Accademico 2009/2010

Indice

1	Introduzione	3
1.1	Gli antibiotici	3
1.1.1	Un'importante classe di antibiotici: I peptidi	4
1.1.2	Un caso particolare: i D,L- α -peptidi ciclici	4
1.1.3	La Gramicidina S e i suoi analoghi	5
2	Metodi	10
2.1	Dinamica molecolare classica	10
2.1.1	Modelli di force fields	12
2.1.2	Le condizioni al contorno	14
2.1.3	Integrazione delle equazioni del moto	15
2.1.4	Metodi per migliorare l'efficienza	18
2.1.5	Calcolo delle interazioni a lungo raggio	19
2.1.6	Metodi per il controllo della temperatura e della pressione	23
2.1.7	Dinamica essenziale	27
2.2	Il Perturbed Matrix Method	31
2.2.1	La teoria di base	31
2.2.2	Il calcolo degli spettri Infrarosso	37
3	Risultati	39
3.1	Simulazioni di dinamica molecolare	39
3.1.1	Analisi delle traiettorie di dinamica molecolare	40
3.1.2	Calcolo della costante cinetica di folding	59
3.1.3	Confronto con il GS6	63
3.2	Calcoli degli spettri infrarosso	64
3.2.1	Spettri infrarosso del trans-NMA	65
3.2.2	Spettri infrarosso del β -hairpin	67
4	Conclusioni	76
	Bibliografia	79

Capitolo 1

Introduzione

1.1 Gli antibiotici

Gli antibiotici sono metaboliti microbici o sostanze la cui sintesi si è ispirata ad essi: a piccole dosi, inibiscono la crescita e la sopravvivenza dei microorganismi senza causare eccessiva tossicità nell'organismo ospite; il concetto chiave della loro azione consiste quindi nella loro selettività.

Gli antibiotici sono inoltre tra i farmaci più prescritti al giorno d'oggi, sebbene la loro efficacia sia messa in forse dalla resistenza batterica, sviluppatasi a causa della pressione evolutiva e di un uso scorretto.

In molti casi l'utilità degli antibiotici naturali è stata esaltata da modifiche chimiche della struttura che ne hanno allargato lo spettro d'azione antimicrobica, aumentato la potenza, ridotto gli effetti tossici, migliorato la somministrazione, ecc.

Sotto il nome di antimicrobici sono classificati invece i molti composti sintetici che inibiscono la crescita o la vita dei microor-

ganismi ma che non sono strutturalmente correlati a prodotti di origine naturale. [1]

1.1.1 Un'importante classe di antibiotici: I peptidi

Un gran numero di peptidi atipici biologicamente attivi sono stati isolati da batteri, attinomiceti e funghi.

Strutturalmente rappresentano un gruppo estremamente differente dagli antibiotici non peptidici, che comprende quei metaboliti contenenti due o più amminoacidi legati da un legame peptidico, ma che possiedono altre caratteristiche non proprie delle proteine; queste possono includere residui amminoacidici inusuali con configurazione D o portati ad un livello di ossidazione superiore, oppure ancora legami non peptidici tra i residui (per esempio esteri, lattoni o un γ -glutammi-ammide). In aggiunta, le molecole possono essere lineari o cicliche, contenere una o più delle caratteristiche sopracitate, essere modificate da ulteriori interazioni tra le catene laterali degli amminoacidi o coniugati con lipidi o zuccheri. [2]

1.1.2 Un caso particolare: i D,L- α -peptidi ciclici

I D,L- α -peptidi ciclici possiedono caratteristiche strutturali uniche, non rilevate in alcun'altra classe naturale di antibiotici peptidici o nei loro derivati.

I peptidi ciclici con un numero pari di D e L- α -amminoacidi alternati possono adottare conformazioni piatte ad anello nelle quali le funzionalità ammidiche della catena principale sono orientate

perpendicolarmente alle catene laterali e al piano della struttura ad anello.

In condizioni particolari, i peptidi ciclici si possono impilare per formare strutture a cavità tubolari simili a foglietti β , aperte all'estremità, presentando le catene laterali degli amminoacidi verso la superficie esterna dell'insieme. Questo porta il peptide ad avere proprietà antibatteriche poiché, una volta assemblatosi nella membrana batterica, ne aumenta la permeabilità. [3]

1.1.3 La Gramicidina S e i suoi analoghi

Un importante rappresentante della categoria succitata è la Gramicidina S (GS), un peptide solubile in acqua, composto da una serie di 10 amminoacidi D ed L alternati e ciclizzati mediante due β -turns, un tipico ripiegamento secondario presente in molte proteine.

Questo peptide presenta un'interessante attività antibiotica, ma anche caratteristiche emolitiche, il che ne preclude purtroppo il possibile uso come farmaco; da qui l'esigenza di produrre analoghi sintetici che accentuino la sua efficacia come farmaco, ma che riducano o eliminino la tossicità.

Sono stati sintetizzati diversi analoghi della GS, in particolare un analogo con sei residui (GS6) ed un analogo a 10 residui (GS10), le cui strutture sono mostrate in Figura 1.1; di questi il GS10 in particolare possiede la richiesta attività antibiotica, mentre la caratteristica emolitica è assente.

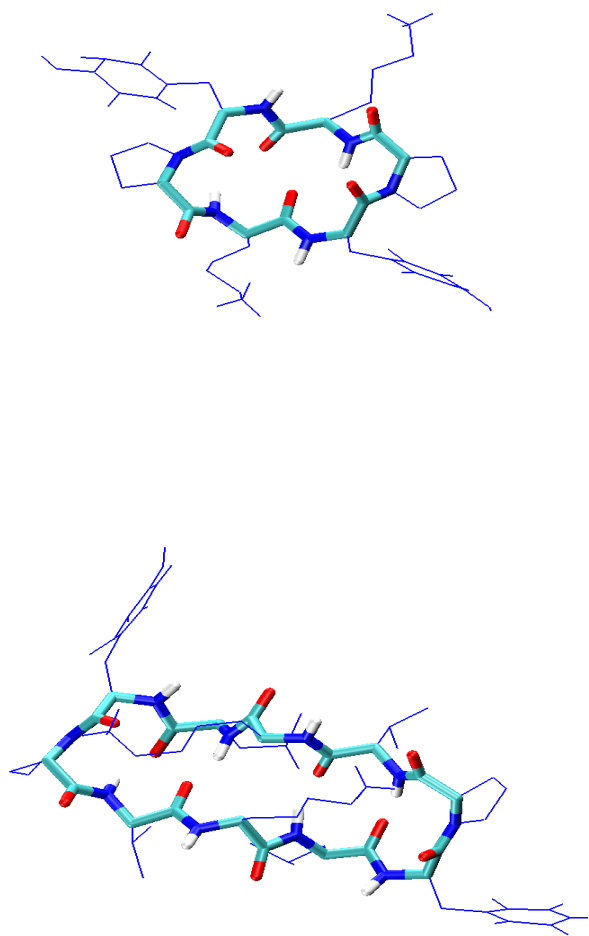


Figura 1.1: Strutture degli analoghi della Gramicidina S GS6 e GS10, ricavate da simulazioni MD.

Lo studio di questi peptidi presenta inoltre un altro aspetto interessante. Infatti, essendo composti da pochi atomi, sono particolarmente indicati per simulazione al calcolatore al fine di studiarne il meccanismo di folding/unfolding e la relativa cinetica e termodinamica.

Questi peptidi sono poi costretti stericamente nella loro ricerca dello spazio conformazionale e, di conseguenza, la formazione del ripiegamento e la propagazione dei legami idrogeno possono essere misurate separatamente dai processi diffusivi della catena (collasso).

Le loro strutture sono state risolte mediante spettroscopia NMR, che ha rilevato la presenza di strutture β -hairpin antiparallele, collegate da due β -turns di tipo II'. Da un punto di vista sperimentale, questi peptidi sono stati largamente studiati con varie tecniche spettroscopiche.

Mediante dicroismo circolare è stata dimostrata la maggiore tendenza a formare foglietti β rispetto agli analoghi a 8, 12 e 16 residui [6]. La dinamica della formazione di β -turn e le frequenze di folding/unfolding sono state poi studiate usando la spettroscopia infrarossa in trasformata di fourier e la sua variante time-resolved [7, 8], particolarmente adatta per studiare strutture secondarie di proteine e peptidi. Per questo scopo, grazie alla loro sensibilità alle strutture secondarie e alle variazioni di solvatazione, vengono spesso presi in esame i modi vibrazionali dei gruppi ammidici della catena principale; in particolare il I modo di vibrazione del-

l'ammide, che corrisponde essenzialmente allo stretching del C=O ($1600\text{-}1700\text{ cm}^{-1}$) è il più studiato. Questo metodo è tuttavia limitato dal fatto che le forme delle bande per diverse strutture possono risultare simili ed altamente sovrapposte [9], portando a spettri di difficile interpretazione; inoltre l'interpretazione di spettri IR è tipicamente basata su regole empiriche e assunzioni, la cui validità generale non può essere sempre dimostrata (per esempio, l'ipotesi che maggiore la temperatura, maggiore la popolazione dello stato unfoldato). Il comportamento di rilassamento è stato infine caratterizzato usando un modello a due stati, dal quale ricavare le cinetiche di folding e i parametri di attivazione. Costanti di folding/unfolding dell'ordine di 10^7 s^{-1} sono state ottenute, a dimostrazione della cinetica molto rapida [10].

L'introduzione di nuove tecniche sperimentali basate sull'uso di un impulso laser, che permettono il verificarsi di processi di folding in tempi di nanosecondi [4], ha avuto un notevole impatto sullo studio del folding proteico da un punto di vista sia sperimentale che teorico-computazionale. Queste nuove tecniche hanno permesso di studiare i meccanismi di formazione degli elementi di struttura secondaria - α eliche, β hairpin e ripiegamenti - così come il meccanismo del folding per proteine con un ripiegamento sufficientemente rapido.

Nel nostro laboratorio abbiamo già affrontato lo studio cinetico e termodinamico sul GS6 mediante dinamica molecolare. I risultati sono stati confrontati con i dati sperimentali ottenuti me-

dante spettroscopia IR tempo/risolta. Ciò ha permesso quindi di ottenere una maggiore comprensione dei meccanismi cinetici e termodinamici coinvolti nel processo di folding/unfolding di questo, e più in generale, di piccoli peptidi [11].

Lo scopo di questa tesi è quello di affrontare lo studio del GS10 mediante simulazioni di dinamica molecolare. In particolare è stato analizzato il meccanismo di folding/unfolding ed è stato possibile stimare l'ordine di grandezza di alcune costanti cinetiche. Infine abbiamo utilizzato le simulazioni ottenute in modo da calcolare gli spettri IR mediante un metodo misto QM/MD, e li abbiamo confrontati con quelli sperimentali.

Capitolo 2

Metodi

2.1 Dinamica molecolare classica

In questa sezione sono introdotti i concetti e le metodologie alla base delle simulazioni di dinamica molecolare (MD) classica; viene qui riportata una descrizione dei principi e delle tecniche utilizzate in questo studio.

Lo scopo delle simulazioni di sistemi molecolari è quello di studiare il loro comportamento macroscopico a partire dalle interazioni microscopiche. Si costruisce un modello del mondo reale, se ne elaborano le proprietà misurabili e quelle non misurabili, operando poi un confronto tra le prime ed i dati sperimentali. Se il modello è validato dal confronto, può poi essere utilizzato per predire quantità sconosciute o non misurabili. Un trattamento teorico dei moti e delle interazioni delle molecole dovrebbe essere fondato, a rigore, sui principi della quantomeccanica, a causa della loro natura microscopica. Sfortunatamente, un approccio del genere è spesso impraticabile, poiché vengono richiesti calcolatori

di grandi dimensioni, risultando del tutto proibitivo per sistemi contenenti migliaia di atomi. Diventa necessario quindi un certo livello di approssimazione, tale che possano essere sufficientemente campionati quei gradi di libertà essenziali per una giusta valutazione della quantità o proprietà di interesse. Possiamo assumere perciò con una certa sicurezza che, escludendo reazioni chimiche, basse temperature o il moto dettagliato degli atomi di idrogeno, il sistema sia governato dalle leggi della meccanica classica.

Nella MD classica una traiettoria (configurazioni in funzione del tempo) di un sistema, è generata dall'integrazione simultanea delle equazioni del moto di Newton per tutti gli atomi del sistema, che per sistemi conservativi sono:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = m_i^{-1} \mathbf{F}_i \quad (2.1)$$

$$\mathbf{F}_i = - \frac{\partial V(\mathbf{r}_i \dots \mathbf{r}_N)}{\partial \mathbf{r}_i} \quad (2.2)$$

Con $\mathbf{F}_i$ si intende la forza agente sull'atomo i -esimo, con m_i la sua massa e con t il tempo. Simulazioni MD richiedono il calcolo del gradiente dell'energia potenziale $V(\mathbf{r}_i \dots \mathbf{r}_N)$, che deve quindi essere una funzione differenziabile delle coordinate atomiche $\mathbf{r}_i$. Questa funzione di energia potenziale, o *force field*, è chiamata anche funzione di *interazione effettiva*, essendo l'effetto medio dei

gradi di libertà omessi (elettronici) incorporato nell'interazione tra i gradi di libertà (atomici) esplicitamente presenti nel modello.

La scelta di un modello molecolare e del force field è essenziale per una predizione adatta delle proprietà del sistema. E' quindi di grande importanza essere al corrente delle fondamentali assunzioni, semplificazioni e approssimazioni impliciti nei vari tipi di modelli presenti in letteratura.

2.1.1 Modelli di force fields

Attualmente viene utilizzata una grandissima varietà di force fields (CHARMM [12], AMBER [13], GROMOS [14]...), i quali differiscono a volte solo per piccole variazioni. Un tipico force field molecolare, o potenziale effettivo ha questa forma:

$$\begin{aligned}
V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = & \sum_{bonds} \frac{1}{2} K_b [b - b_{eq}]^2 + \sum_{angles} \frac{1}{2} K_\theta [\theta - \theta_{eq}]^2 \\
& + \sum_{\substack{improper \\ dihedrals}} \frac{1}{2} K_\xi [\xi - \xi_{eq}]^2 \\
& + \sum_{dihedrals} K_\phi [1 + \cos(n\phi - \delta)] \\
& + \sum_{pairs(i,j)} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
& + \sum_{pairs(i,j)} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.3}
\end{aligned}$$

Il primo termine rappresenta, mediante un potenziale armonico, l'interazione di stretching di legame covalente tra due atomi, dove b_{eq} è la lunghezza di legame di minima energia e K_b è la costante di forza, diversa a seconda del tipo di legame. Il secondo termine è un'interazione a tre corpi corrispondente alla deformazione dell'angolo di valenza, θ , espressa sempre da un potenziale armonico, dove θ_{eq} è l'angolo di valenza all'equilibrio e K_θ la costante di forza. Il terzo e il quarto termine sono invece usati per le interazioni di angolo diedro a quattro corpi: un termine armonico per gli angoli diedri impropri, e un termine sinusoidale per tutti gli altri angoli diedri, ϕ . Gli ultimi due termini sono le somme su tutte le coppie di atomi non legati e rappresentano le interazioni effettive di non legame espresse in termini di interazioni di Van der Waals e Coulumbiane tra gli atomi i e j alla distanza r_{ij} . I parametri ϵ_{ij} e σ_{ij} sono le costanti tipiche che definiscono il potenziale Lennard-Jones, mentre q_i e q_j sono le cariche degli atomi e ϵ_0 la costante dielettrica nel vuoto.

I parametri usati nel force field possono essere determinati in diversi modi. In generale due approcci principali vengono seguiti. Un primo metodo fa uso di calcoli *ab initio* su piccoli cluster molecolari. In alternativa è possibile ricavare i parametri da dati sperimentali, come strutture cristalline, dinamica energetica e di reticolo, dati infrarossi o Raman su piccole molecole, proprietà liquide

come densità ed entalpia di vaporizzazione, energia libera di solvatazione, dati NMR, etc. Qualunque sia il metodo utilizzato, il modello risultante sarà tuttavia lontano dall'essere universale. Vale la pena far notare che ogni force field è di solito ben adatto a specifiche condizioni termodinamiche (temperatura, densità, pressione, etc.) ma anche a condizioni al contorno. Inoltre, sono ottimizzati per specifiche classi di molecole, come molecole inorganiche, organiche, biomolecole (DNA, proteine, lipidi), etc.

2.1.2 Le condizioni al contorno

Una caratteristica importante delle simulazioni MD è il modo in cui viene trattato il contorno. A causa dei limiti computazionali, un tipico sistema simulato contiene 10^4 - 10^5 atomi, abbastanza piccolo se confrontato con la materia macroscopica. Questo significa che, se le molecole sono sistemate in un box cubico, una gran parte di esse si troveranno sulla superficie, e sperimenteranno perciò forze diverse dalle molecole al suo interno. La conseguenza di una dimensione finita del sistema è che le condizioni al contorno possono influenzare seriamente il risultato delle simulazioni, specie quando il sistema di interesse è un liquido omogeneo o una soluzione. Per ridurre gli effetti di superficie, tipicamente si adottano condizioni periodiche al contorno (PBC) [15]: questa tecnica consiste nel simulare il sistema in un box centrale circondato da un infinito numero di copie di se stesso. Durante la simulazione, le molecole nel box originale e le loro immagini periodiche si muovo-

no nello stesso modo; quando una molecola lascerà il box centrale quindi, una delle sue immagini vi entrerà dalla parte opposta. Come risultato, non esistono più barriere fisiche, né molecole in superficie. Da notare che è possibile utilizzare altre forme per il box, come l'ottaedro troncato o il dodecaedro rombico.

2.1.3 Integrazione delle equazioni del moto

In una simulazione MD le traiettorie degli N atomi del sistema si ottengono dall'integrazione delle equazioni del moto di Newton. Vista la complessità dell'espressione del campo di forze (2.1.1), l'integrazione di queste equazioni non ha soluzione analitica. Per questo sono stati sviluppati nel corso degli anni diversi metodi approssimati che estraggono le traiettorie da una simulazione MD per via numerica.

Le caratteristiche che deve avere un buon algoritmo di integrazione possono essere così riassunte:

- deve permettere l'uso di un time step lungo. E' evidente che maggiore risulta il passo di integrazione, minore è il numero di integrazioni necessarie in una simulazione. D'altro canto, se il Δt è troppo grande, si rischia di non riuscire a seguire certe componenti del moto del sistema, come ad esempio le vibrazioni, portando così a risultati a volte catastrofici. E' pertanto opportuno tarare il passo di integrazione sui moti a frequenza più alta;

- deve essere reversibile temporalmente e conservare le grandezze costanti nel sistema. Le equazioni di Newton sono reversibili temporalmente ed è pertanto auspicabile che l'algoritmo utilizzato per integrarle riproduca con fedeltà questo dato.

Gli algoritmi utilizzati si differenziano nella precisione di calcolo (ordine dell'algoritmo), nella occupazione di memoria e nel tipo di rappresentazione. La natura degli algoritmi usati deriva dallo sviluppo in serie di Taylor delle grandezze $r_i(t+\Delta t)$ e $v_i(t+\Delta t)$. In MD si adoperano generalmente algoritmi che utilizzano espansioni di Taylor fino al terzo ordine in quanto si è osservato che algoritmi di ordine inferiore sono poco accurati mentre quelli di ordine superiore sono troppo dispendiosi in termini di risorse computazionali. Gli algoritmi comunemente usati sono:

- l'algoritmo di Verlet;
- l'algoritmo leap-frog.

Il primo algoritmo è veloce, richiede poca memoria, presenta un piccolo errore per tempi lunghi ed una buona conservazione dell'energia per tempi brevi. Si ottiene da addizioni e sottrazioni di espansioni di Taylor delle coordinate $\mathbf{r}_i(t + \Delta t)$ e $\mathbf{r}_i(t - \Delta t)$ di particelle di massa m_i :

$$\mathbf{r}_i(t \pm \Delta t) = \mathbf{r}_i(t) \pm \Delta t \dot{\mathbf{r}}_i(t) \pm \frac{(\Delta t)^2}{2!} \ddot{\mathbf{r}}_i \pm \frac{(\Delta t)^3}{3!} \dddot{\mathbf{r}}_i + O(\Delta t^4) \quad (2.4)$$

da cui si ricava:

$$\mathbf{r}_i(t + \Delta t) = -\mathbf{r}_i(t - \Delta t) + 2\mathbf{r}_i(t) + \frac{(\Delta t^2)}{m_i}\mathbf{F}_i(t) \quad (2.5)$$

dove $\mathbf{F}_i(t)$ è la forza esercitata sulla particella i al tempo t e dove i termini di quarto ordine e superiori sono stati trascurati. Questo schema di integrazione è esatto fino al quarto ordine in Δt . Le velocità in questo schema sono date da:

$$\mathbf{v}_i(t) = \dot{\mathbf{r}}_i(t) = \frac{1}{2\Delta t}[\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t)] \quad (2.6)$$

in cui l'errore sulle velocità è dell'ordine di $(\Delta t)^3$.

Il leap-frog è una variante dello schema d'integrazione appena visto. Viene utilizzato in GROMACS, ed in particolare è stato impiegato in tutte le simulazioni di questo lavoro. Questo schema usa per le coordinate intervalli interi e per le velocità intervalli semi-interi.

$$\mathbf{v}_i(t + \Delta t/2) = \mathbf{v}_i(t - \frac{1}{2}\Delta t) + \frac{\mathbf{F}_i(t)\Delta t}{m_i} \quad (2.7)$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \Delta t/2)\Delta t \quad (2.8)$$

L'algoritmo leap-frog è molto efficiente da un punto di vista computazionale, essendo semplice e richiedendo meno tempo di CPU e memoria rispetto ad altri più avanzati.

2.1.4 Metodi per migliorare l'efficienza

Dalla pubblicazione della prima applicazione a sistemi biomolecolari [20], poco più di 30 anni fa, sono stati sviluppati diversi metodi per aumentare i tempi delle simulazioni MD. Quando le equazioni del moto di Newton vengono integrate, il fattore limitante che determina il time step è la frequenza più elevata presente nel sistema; nelle macromolecole biologiche solvate, corrisponde a quella associata alle vibrazioni dei legami comprendenti atomi d'idrogeno. Prendendo in considerazione il legame O-H, la sua frequenza di stretching si trova tipicamente intorno a 10^4 Hz, così che il periodo medio sarà dell'ordine di 10 fs [21]: questo limiterà quindi la scelta del time step a circa 0.5 fs (infatti, affinché sia effettuato un buon campionamento di una funzione periodica, si è stabilito un campionamento pari ad almeno venti volte in un periodo). L'introduzione di un constraint per questi legami (o nei fatti per tutti i legami covalenti) ha permesso quindi di incrementare il time step ad un valore di 2 fs (mediante SHAKE o LINCS) [22]. Inoltre, dato che queste vibrazioni sono praticamente disaccoppiate dalle altre vibrazioni del sistema, la dinamica del sistema non

risulta alterata in modo significativo. Questo non è vero invece per le fluttuazioni dell'angolo-legame, che costituiscono la seconda più alta frequenza di vibrazione. Il blocco di queste interazioni ha infatti un notevole impatto su molte altre fluttuazioni, anche su quelle di carattere globale e collettivo, limitandone perciò l'uso a pochi casi specifici [21].

E' possibile definire algoritmi più efficienti grazie al fatto che un certo numero di classi discrete di frequenze di fluttuazioni possono essere distinte nelle simulazioni. Le forze che fluttuano rapidamente devono essere ricalcolate ad una frequenza più alta rispetto alle altre. Anche se di non facile implementazione, alcune applicazioni di algoritmi chiamati *multiple time step* permettono un time step 4-5 volte superiore.

2.1.5 Calcolo delle interazioni a lungo raggio

In una simulazione di MD il 95% del tempo di calcolo è occupato dal calcolo delle interazione di non legame. E' quindi opportuno cercare un metodo che permetta di aumentare l'efficienza computazionale della simulazione agendo sulle componenti di van der Waals e coulombiana. In questo senso, sono state sviluppate varie tecniche per ridurre lo sforzo computazionale. Tra quelle più utilizzate:

- il raggio di cut-off;
- le somme di Ewald.

Il primo metodo consiste nel troncamento delle interazioni di non legame quando la distanza supera un certo valore, detto raggio di cut-off (r_c), pertanto ogni particella del sistema potrà interagire solo con le altre che sono all'interno di una sfera di raggio r_c . Ovviamente l'uso di questo metodo introduce delle approssimazioni, soprattutto nel calcolo della parte elettrostatica, che possono influenzare notevolmente le proprietà dinamiche e strutturali del sistema. Per superare questo problema, oltre ad aumentare il valore del raggio di cut-off, si può introdurre una funzione di switch che assicura la continuità delle forze alla distanza di cut-off. Anche se l'uso del raggio di cut-off soffre di queste limitazioni, rimane uno tra i metodi più diffusi per il calcolo delle forze intermolecolari. Il secondo metodo, più recente, propone l'uso del reticolo periodico, in cui si considerano tutte le interazioni tra il box centrale e le sue infinite repliche. L'energia coulombiana in un sistema periodico formato da N particelle è ottenuta sommando su tutte le coppie di atomi tra i quali uno è nel box centrale e l'altro in una delle repliche:

$$E = \frac{1}{8\pi\epsilon_0} \sum_{|n|=0}^{\star\infty} \left[\sum_{i=0}^N \sum_{j=0}^N \frac{q_i q_j}{|r_{ij} + n|} \right] \quad (2.9)$$

dove n sono le immagini periodiche, i, j le particelle e il simbolo $\star$ indica che la sommatoria non include il termine con $i = j$ se

$n = 0$. Si può dimostrare che questa sommatoria, per potenziali che vanno come $\frac{1}{r}$, può non essere totalmente convergente, cioè il suo limite può variare o divergere se l'ordine dei termini viene cambiato; pertanto questa espressione non può essere utilizzata nel calcolo dell'energia elettrostatica. Per risolvere questo problema si può considerare ogni carica circondata da una distribuzione di carica ρ^G di uguale intensità e segno opposto. Questa distribuzione, che supponiamo essere gaussiana, avrà l'effetto di schermare le interazioni tra le cariche vicine e potremmo quindi considerare l'energia di interazione di tipo *short-range*. La distribuzione di carica totale è

$$\rho_i(r) = \rho_i^q(r) + \rho_i^G(r) \quad (2.10)$$

dove $\rho_i^q(r)$ è la distribuzione della carica della particella i -esima e $\rho_i^G(r)$ è la distribuzione gaussiana corrispondente:

$$\rho_i^q(r) = q_i \delta(r - r_i) \quad (2.11)$$

Dopo aver calcolato l'energia di interazione in (2.10) nello spazio reale, per tornare alla funzione di distribuzione originaria $\rho_i(r)$, si sottrae una funzione uguale a $\rho_i^G(r)$ nello spazio reciproco, usan-

do la trasformata di Fourier. L'espressione finale dell'energia di interazione totale sarà quindi:

$$\begin{aligned}
E = & \frac{1}{8\pi\epsilon_0} \sum_{i=0}^N \sum_{j=0}^N \left[\sum_{|n|=0}^{\infty} \frac{q_i q_j \operatorname{erfc}(\alpha |r_{ij} + n|)}{|r_{ij} + n|} \right] + \\
& + \frac{1}{8\pi L^3} \sum_{k=0} \frac{4\pi q_i q_j}{k^2} \exp(-k^2/4\alpha^2) \cos(k \cdot r_{ij}) + \\
& - \frac{\alpha}{4\pi^{3/2}\epsilon_0} \sum_{i=1}^{\infty} q_i^2 + \frac{|\sum_{i=1}^N q_i r_i|^2}{2\epsilon_0 L^3 (2\epsilon' + 1)} \quad (2.12)
\end{aligned}$$

dove $\operatorname{erfc}(x)$ è la funzione errore complementare, che tende a zero per $x \rightarrow \infty$. Per quanto riguarda il primo termine, per α abbastanza grande, la somma si riduce ad un unico termine con $n = 0$; il secondo termine è invece una somma sui vettori reciproci $k = 2\pi n/L$. Gli ultimi due termini sono, rispettivamente, la funzione di correzione, dovuta alla presenza di un termine di auto-interazione in $\rho_i^G(r)$, e il contributo del campo depolarizzante all'energia, che è compensato dall'effetto del dielettrico esterno. Questo ultimo punto necessita di un ulteriore commento, infatti nelle somme di Ewald, i box virtuali sono visti come strati sferici concentrici ordinati dall'interno verso l'esterno; la sommatoria è ovviamente troncata dopo un certo numero di strati e questo sistema sferico è immerso in un dielettrico continuo con costante dielettrica uguale a ϵ' . L'ultimo termine nell'equazione è la somma dei contributi del campo depolarizzante e il campo esterno dovuto

al dielettrico. Se la sfera è immersa in un mezzo con una costante dielettrica infinita, questo termine scompare, perché è compensato esattamente da questi due effetti. Recentemente sono stati sviluppati altri metodi della stessa famiglia, come il *Particle-Particle, Particle-Mesh (PPPM)* o il *Particle Mesh Ewald (PME)* [25], che basandosi sugli stessi principi sui quali sono state sviluppate le somme di Ewald, sono computazionalmente più efficienti e stabili. In particolare, in tutte le simulazioni di questo lavoro, è stato usato il PME come metodo di calcolo per le forze elettrostatiche. Il metodo del PME sostanzialmente è identico alle somme di Ewald, con la sola differenza che per calcolare l'energia di una certa distribuzione di carica si considerano le cariche disposte su una griglia (mesh); questo rende notevolmente più efficiente il calcolo della parte in cui è coinvolta la trasformata di Fourier, che costituisce la parte computazionalmente più pesante.

2.1.6 Metodi per il controllo della temperatura e della pressione

Per poter confrontare i risultati di una simulazione MD con i dati sperimentali occorre mettersi nelle stesse condizioni in cui questi sono stati ricavati. Questo spesso costringe ad eseguire la simulazione a pressione e temperatura costanti. Nel corso degli anni sono stati sviluppati, a questo proposito, vari algoritmi. I metodi generalmente impiegati per generare insiemi statistici isotermini e isobari si possono dividere in due gruppi. Del primo fanno parte

quella classe di procedure definite con il nome di *extended system methods* [28, 29]. Questi metodi fanno uso di meccanismi di retroazione per correggere le deviazioni di parametri controllati (temperatura e pressione) da valori medi prestabiliti, questo implica che il valore può fluttuare, ma l'ampiezza della fluttuazione può essere controllata. L'altro gruppo è rappresentato dai cosiddetti *constraint methods*, in cui le equazioni del moto vengono accoppiate ad un opportuno vincolo (per esempio, nel caso della temperatura questo vincolo è l'energia cinetica costante). Nel caso della temperatura esistono due ulteriori gruppi costituiti dai cosiddetti *stochastic methods* e dai *weak coupling methods* [26]. I primi consistono nell'aggiustare la temperatura in modo da riprodurre l'effetto di un bagno termico esterno o, in modo da generare ad ogni passo una distribuzione maxwelliana di velocità in modo da riportare il valore dell'energia cinetica totale a quello stabilito; nei secondi le equazioni del moto vengono modificate in modo che ci sia un rilassamento del primo ordine in T verso un valore di riferimento T_0 secondo la classica equazione:

$$\frac{dT(t)}{dt} = \frac{(T_0 - T(t))}{\tau} \quad (2.13)$$

Nelle simulazioni eseguite nel corso di questa tesi si è fatto uso del metodo di Berendsen o dell'accoppiamento debole con un bagno termico che fa parte del gruppo dei *weak coupling methods*.

Non si è utilizzato alcun metodo per il controllo della pressione poiché tutte le simulazioni sono state effettuate in un box a dimensioni fisse. Di seguito verrà illustrato brevemente il metodo di Berendsen. Uno dei pregi di questo metodo è la semplicità. L'algoritmo simula l'accoppiamento del sistema con un bagno termico esterno alla temperatura T_0 . La velocità dell'interazione tra questo bagno termostato ed il sistema è determinata da una costante di tempo, τ , che può essere liberamente scelta. L'accoppiamento viene ottenuto modificando ad ogni passo la velocità degli atomi. Questo può essere fatto moltiplicando per una opportuna costante λ le velocità delle singole particelle. Il livellamento (*rescaling*) della temperatura T ad un valore limite T_0 seguendo una legge esponenziale può essere ottenuto come segue, tenendo conto che $T(t)$ è la temperatura ottenuta dal valore delle velocità dopo l'integrazione e $T'(t)$ è il valore di temperatura ottenuto dopo il *rescaling*:

$$\frac{dT}{dt} = \frac{1}{\tau}[T_0 - T(t)] \quad (2.14)$$

$$\Delta T = \frac{\Delta t}{\tau}[T_0 - T(t)] \quad (2.15)$$

$$T'(t) = \lambda T(t) \quad (2.16)$$

$$\Delta T = T'(t) - T(t) = (\lambda - 1)T(t) \quad (2.17)$$

da (2.15) e (2.17) si ottiene :

$$(\lambda - 1)T(t) = \frac{\Delta t}{\tau}[T_0 - T(t)] \quad (2.18)$$

$$\lambda = 1 + \frac{\Delta t}{\tau}\left(\frac{T_0}{T(t)} - 1\right) \quad (2.19)$$

$$T'(t) = \left[1 + \frac{\Delta t}{\tau}\left(\frac{T_0}{T(t)} - 1\right)\right]T(t) \quad (2.20)$$

Il metodo che è stato impiegato in questo lavoro di tesi per il controllo della temperatura è l'*isotermo gaussiano* [27]. Tale metodo appartiene ai cosiddetti *constraint methods* e utilizza l'energia cinetica totale del sistema come una variabile vincolata; le equazioni del moto, utilizzando l'accoppiamento isotermo gaussiano divengono

$$\dot{x}_i = \frac{\partial H(\mathbf{x}, \mathbf{p})}{\partial p_i} \quad (2.21)$$

$$\dot{p}_i = -\frac{\partial H(\mathbf{r}, \mathbf{p})}{\partial x_i} + \gamma p_i \quad (2.22)$$

dove γ è equivalente a un coefficiente di attrito e varia implicita-

mente nel tempo così da garantire che il valore dell'energia cinetica sia costante, ovvero mantiene il vincolo

$$\dot{K} = \sum_{i=1} \frac{p_i \dot{p}_i}{m_i} = 0 \quad (2.23)$$

Si può dimostrare che questo metodo genera una distribuzione nello spazio delle configurazioni in accordo con la meccanica statistica.

2.1.7 Dinamica essenziale

La dinamica essenziale (ED) viene utilizzata per la ricerca dell'insieme di gradi di libertà che meglio approssimano le fluttuazioni totali di un sistema dinamico; è basata su un'analisi delle componenti principali (PCA). Una PCA è una procedura di fitting lineare multidimensionale. Passando alla dinamica proteica, l'usuale spazio cartesiano tridimensionale (3D) dove si rappresentano le coordinate proteiche (dove vengono rappresentate anche le strutture delle proteine nel Brookhaven Protein Data Bank o PDB), viene sostituito da uno spazio multidimensionale nel seguente modo: una molecola di N particelle può essere rappresentata da N punti in uno spazio 3D e con 3 coordinate per punto, questo porta ad averne $3N$; in uno spazio $3N$ -dimensionale una tale struttura viene invece rappresentata da un singolo punto, caratterizzato da $3N$ coordinate. Questa rappresentazione è conveniente dato che ci

permette di rappresentare una collezione o traiettoria di strutture come un insieme di punti. In due dimensioni, esisterà sempre una retta che meglio approssimerà questi punti: se tale retta riproduce bene i dati, i dati stessi potranno perciò essere approssimati dalla sola posizione lungo la retta trascurando le fluttuazioni ortogonali; se prendiamo questa retta come nuovo asse del sistema di riferimento, allora la posizione di un punto potrà essere visualizzata da una singola coordinata. In modo simile la procedura funziona per più dimensioni, dove di solito un numero limitato di questi assi (le componenti principali) riproducono adeguatamente i dati. Tali assi definiscono il sottospazio responsabile della gran parte delle fluttuazioni che quindi viene chiamato *sottospazio essenziale*. Applicazioni di questa procedura alle configurazioni di alcune proteine hanno dimostrato come le prime dieci o venti componenti principali sono responsabili del 90% delle fluttuazioni totali [31, 32, 33]. Queste componenti principali corrispondono a coordinate collettive, essendo costruite a partire da ogni atomo della molecola. Riassumendo, un numero limitato di moti collettivi è responsabile per una larga percentuale delle fluttuazioni strutturali di una molecola.

Se tutti gli atomi in una proteina potessero muoversi indipendentemente l'uno dall'altro, un'approssimazione della fluttuazione totale basata su poche coordinate collettive non sarebbe possibile. Il fatto che tale approssimazione abbia successo è quindi il risultato della presenza di un gran numero di correlazioni e veri e

propri vincoli meccanici ('near-constraint') formati dalle interazioni presenti in una data struttura. Interazioni atomiche, a partire dai legami covalenti (l'interazione più forte) fino alle interazioni di non legame, insieme con il denso impaccamento di atomi nella struttura nativa sono all'origine di queste correlazioni e vincoli.

Nello studio della dinamica proteica, siamo di solito interessati alle sole fluttuazioni interne; il primo passo nella dinamica essenziale sarà quindi quello di rimuovere la rotazione e la traslazione complessiva. A questo scopo viene traslato il centro di massa di ogni configurazione nell'origine, avendo effettuato un fitting rotazionale degli atomi ai minimi quadrati su di una struttura di riferimento. Si passa poi all'analisi delle componenti principali, basata sulla costruzione e la diagonalizzazione della matrice di covarianza delle fluttuazioni delle posizioni atomiche. Definendo con $\mathbf{X}(t)$ il vettore $3N$ dimensionale che rappresenta le coordinate atomiche al tempo t , la matrice di covarianza è costruita in questo modo:

$$\mathbf{C} = \langle \Delta \mathbf{X} \Delta \mathbf{X}^T \rangle$$

dove $\Delta \mathbf{X} = \mathbf{X}(t) - \langle \mathbf{X} \rangle$ e le parentesi angolari rappresentano la media sul tempo o sull'ensemble. Le particelle che si muovono in modo correlato corrispondono ad elementi positivi della matrice (correlazione positiva) o ad elementi negativi (correlazione negativa); elementi indipendenti avranno invece nella matrice valori

molto piccoli. La trasformazione ortogonale $\mathbf{T}$ che diagonalizza questa matrice (simmetrica) conterrà gli autovettori di $\mathbf{C}$ mentre la matrice diagonale risultante $\mathbf{\Lambda}$ sarà costituita dai corrispondenti autovalori:

$$\mathbf{\Lambda} = \mathbf{T}^T \mathbf{C} \mathbf{T}$$

Gli autovalori corrispondono alle fluttuazioni quadratiche medie lungo i corrispondenti autovettori. Quando gli autovettori sono disposti secondo il decrescere degli autovalori, i primi rappresenteranno quei moti collettivi che meglio approssimano la totalità delle fluttuazioni, mentre gli ultimi corrisponderanno a quei gradi di libertà più rigidi (vincoli approssimati). Le caratteristiche di queste fluttuazioni collettive possono quindi essere studiate proiettando l'ensemble delle strutture sui singoli autovettori, traslando poi queste proiezioni nello spazio 3D, per visualizzare gli spostamenti atomici connessi con quell'autovettore. Come già detto, le analisi di traiettorie provenienti da diverse proteine hanno mostrato che la dinamica della struttura nativa è dominata da poche coordinate collettive ed in un certo numero di casi si è visto che proprio questi modi principali sono coinvolti nella dinamica funzionale delle proteine stesse [31, 34, 35]. Le analisi di ED possono essere applicate a qualsiasi sottoinsieme di atomi dell'ensemble di strutture: sono state riportate in letteratura applicazioni a col-

lezioni di strutture ottenute dalla diffusione di raggi X [35, 36], dalla risonanza magnetica nucleare (NMR) [37], e strutture derivate sulle distanze [38]. Dato che le fluttuazioni collettive (della catena principale) dominano la dinamica delle proteine, di solito viene utilizzata solo la catena principale o le coordinate dei $C\alpha$, in modo da semplificare la procedura computazionale ed ottenere una buona convergenza degli autovettori ed autovalori della matrice di covarianza, che richiederebbe un campionamento statistico molto più esteso qualora anche i gradi di libertà delle catene laterali fossero inclusi nell'analisi.

2.2 Il Perturbed Matrix Method

La trattazione principale del PMM è ampiamente descritta nei seguenti articoli [39, 40, 41, 42]. In questa sezione verrà descritta la teoria e le implicazioni fisiche del PMM. Infine, si mostrerà come utilizzare le analisi statistiche su i risultati del PMM, con lo scopo di caratterizzare gli stati quantistici perturbati del *centro quantistico* (QC).

2.2.1 La teoria di base

Definendo con r_n le coordinate del QC e con x le coordinate degli atomi che generano il campo perturbante (classico), si può scrivere il seguente operatore Hamiltoniano elettronico complessivo (con Hamiltoniano elettronico complessivo si intende l'operatore dell'e-

nergia elettronica dove le interazioni nucleo-nucleo sono incluse)
 $\hat{H}$ come

$$\hat{H}(r_n, x) = \hat{H}^0(r_n) + \hat{V}_{QC,e}(r_n, x) + \hat{H}_e^0(x) \quad (2.24)$$

dove $\hat{H}^0$ è l'operatore Hamiltoniano del QC isolato, $\hat{V}_{QC,e}$ rappresenta l'operatore di interazione del QC con l'ambiente e $\hat{H}_e^0$ è l'operatore Hamiltoniano dell'ambiente (escludendo le interazioni QC-ambiente). Per ogni configurazione atomica definita da r_n e x si assume sempre che lo stato elettronico complessivo Ψ_T possa essere espresso dal prodotto di una funzione d'onda Φ che descriva il QC e una funzione d'onda che descriva l'ambiente Ψ_e , ossia $\Psi_T \cong \Phi\Psi_e$. Tale ipotesi può essere considerata come una approssimazione eccellente fintanto che il QC e l'ambiente possano essere ben distinti e trattati come appartenenti a diversi sistemi fisici, condizione che viene rispettata nel caso in cui non si considerino reazioni chimiche o interazioni molto strette tra il QC e l'ambiente. Se si assume inoltre che per ogni stato eccitato del QC lo stato fondamentale dell'ambiente rimanga praticamente invariato, per ogni stato quantico complessivo in cui solamente il QC è in uno stato eccitato si ha che

$$\hat{H}\Psi_{T,n} = H_n\Psi_{T,n} \quad (2.25)$$

$$\Psi_{T,n} = \Phi_n\Psi_{e,0} \quad (2.26)$$

il quale implica

$$\hat{H}_{\Psi_{e,0}}\Phi_n = H_n\Phi_n \quad (2.27)$$

$$\hat{H}_{\Psi_{e,0}} = \langle \Psi_{e,0} | \hat{H} | \Psi_{e,0} \rangle \quad (2.28)$$

in cui H_n è l'autovalore dell'Hamiltoniano elettronico complessivo. Inoltre, dalla definizione dell'operatore Hamiltoniano si ha

$$\hat{H}_{\Psi_{e,0}} = \hat{H}^0 + \hat{V}_{\Psi_{e,0}} + H_{e,0} \quad (2.29)$$

$$\hat{V}_{\Psi_{e,0}} = \langle \Psi_{e,0} | \hat{V}_{QC,e} | \Psi_{e,0} \rangle \quad (2.30)$$

$$H_{e,0} = \langle \Psi_{e,0} | \hat{H}_e^0 | \Psi_{e,0} \rangle \quad (2.31)$$

Si noti che $\hat{V}_{\Psi_{e,0}}$ corrisponde alla perturbazione del QC ad opera della distribuzione di carica dell'ambiente descritta dalla funzione d'onda $\Psi_{e,0}$ (cioè lo stato fondamentale perturbato dell'ambiente). Quindi, separando il termine dovuto alle cariche atomiche dell'ambiente $\hat{V}_{\Psi_{e,0}}^{ac}$ da quelli di ordine superiore, espressi da $\hat{V}_{\Psi_{e,0}}^{ac}$, si ottiene

$$\hat{H}_{\Psi_{e,0}} = \hat{H}^0 + \hat{V}_{\Psi_{e,0}}^{ac} + \Delta\hat{V}_{\Psi_{e,0}} + H_{e,0} \quad (2.32)$$

Fornendolo invece in notazione matriciale otteniamo

$$\widetilde{H}_{\Psi_{e,0}} = \widetilde{H}^0 + \widetilde{V}_{\Psi_{e,0}}^{ac} + \Delta\widetilde{V}_{\Psi_{e,0}} + H_{e,0}\widetilde{I} \quad (2.33)$$

in cui si usa il *basis-set* imperturbato del QC per esprimere tutte le matrici (cioè la matrice Hamiltoniana $\widetilde{H}^0$ del QC è diagonale). L'espressione dell'operatore ottenuta si può sviluppare ulteriormente considerando che per un QC interagente con un set di cariche puntiformi, le cariche atomiche, le interazioni di dispersione a coppia e le repulsioni a corto raggio non sono presenti e quindi, assumendo $\Delta\widetilde{V}_{\Psi_{e,0}} \cong \Delta V_{\Psi_{e,0}}\widetilde{I}$ (ossia che gli autovettori di $\widetilde{H}_{\Psi_{e,0}}$ coincidano con gli autovettori di $\widetilde{H}^0 + \widetilde{V}_{\Psi_{e,0}}^{ac}$), si può scrivere

$$\begin{aligned}
\widetilde{H}_{\Psi_{e,0}} &\cong \widetilde{H}^0 + \widetilde{V}_{\Psi_{e,0}}^{ac} + (\Delta V_{\Psi_{e,0}} + H_{e,0}) \widetilde{I} \\
&\cong \widetilde{H}^0 + \widetilde{V}_{\Psi_{e,0}}^{ac} + (U'_{env} + U'_{D,R}) \widetilde{I} \quad (2.34)
\end{aligned}$$

dove, all'interno dell'approssimazione utilizzata

$\Delta V_{\Psi_{e,0}} + H_{e,0} \cong U'_{env} + U'_{D,R}$ il termine U'_{env} è l'energia potenziale dell'ambiente (escludendo tutte le interazioni semi-classiche con il QC) e il termine $U'_{D,R}$ rappresenta le interazioni a coppia di repulsione e dispersione tra il QC e l'ambiente generate dal campo di forze utilizzato nella simulazione di MD (si noti che il potenziale di interazione $\Delta V_{\Psi_{e,0}}$ non è in generale equivalente al termine di interazione di dispersione e repulsione come per $U'_{D,R}$ perché in questo caso possono essere coinvolte energie di rilassamento intramolecolari). Infine espandendo $\widetilde{V}_{\Psi_{e,0}}^{ac}$ in multipoli del QC e approssimando tutti i termini da i quadrupoli in poi tramite un potenziale a corto raggio (cioè $\Delta V_{\Psi_{e,0}}^{ac} \widetilde{I}$ con $\widetilde{I}$ la matrice identità), si ottiene [42, 39, 43]

$$\begin{aligned}
\widetilde{H} &= \widetilde{H}_{\Psi_{e,0}}(r_n, x) - U'_{env} \widetilde{I} = \widetilde{H}^0(r_n) + \widetilde{V}(r_0, x) \\
&\cong \widetilde{H}^0(r_n) + q_T \mathcal{V}(r_0, x) \widetilde{I} + \widetilde{Z}_1(E(r_0, x), r_n) + \Delta V(r_n, x) \widetilde{I} \quad (2.35)
\end{aligned}$$

in cui q_T è la carica totale del QC, $\mathcal{V}(r_0, x)$ e $E(r_0, x)$ sono rispettivamente il potenziale elettrico perturbante e il campo elettrico prodotto dalle cariche atomiche dell'ambiente, per una data geometria del QC r_0 (tipicamente il centro di massa o geometrico del QC), $\tilde{Z}_1(E, r_n)$ è la matrice di perturbazione dell'energia data esplicitamente da $[\tilde{Z}_1]_{l,l'} = -E \cdot \langle \Phi_l^0 | \hat{\mu} | \Phi_{l'}^0 \rangle$ in cui $\hat{\mu}$ è l'operatore di dipolo, e $\Delta V(r_n, x) = \Delta V_{\Psi_{e,0}}^{ac} + U'_{DR}$ approssima tutti i termini di ordine superiore come un potenziale a corto raggio (nel caso in cui il QC fosse solo una parte di una molecola più grande, ΔV potrebbe includere una costante additiva corrispondente alla possibile variazione dell'energia di riferimento). L'equazione (2.35) mostra la matrice Hamiltoniana perturbata del QC $\tilde{H}$ e fornisce l'espressione del PMM per un QC interagente in maniera semiclassica con l'ambiente. Esprimendo tale equazione in notazione matriciale

$$\begin{aligned} \hat{H} &= \hat{H}^0 + \hat{V} \\ &\cong \hat{H}^0 + q_T \mathcal{V} - E \cdot \hat{\mu} + \Delta V \end{aligned} \quad (2.36)$$

con l'operatore di perturbazione $\hat{V}$ fisicamente coincidente alla perturbazione dovuta al campo generato dallo stato fondamentale dell'ambiente agente sul QC. Gli autovettori di $\tilde{H}$, cioè gli

autostati di $\hat{H}$, possono essere quindi utilizzati per ottenere le proprietà elettroniche perturbate del QC. Deve essere sottolineato tuttavia che le equazioni (2.35) e (2.36), basandosi sull'approssimazione che l'operatore di perturbazione possa essere troncato al termine dipolare includendo esplicitamente i termini di ordine superiore come un potenziale a corto raggio indipendente dallo stato elettronico, possono non essere adeguate per descrivere il tipo di interazioni QC-ambiente in cui sia necessario esplicitare termini di ordine superiore o operatori multipolari distribuiti. D'altronde, quando è possibile ignorare, stati elettronici eccitati ad alta energia l'approssimazione utilizzata nelle equazioni (2.35) e (2.36) è tipicamente piuttosto accurata. L'introduzione di termini multipolari, benché possibile in teoria [42, 43], richiederebbe l'utilizzo di multipoli di transizione che allo stato attuale dei software sono meno accurati dei corrispondenti dipoli, il che si tradurrebbe in un PMM di peggior qualità.

2.2.2 Il calcolo degli spettri Infrarosso

La procedura sopra descritta è stata applicata per studiare il I modo dell'ammide del trans-N-metilacetammide (trans-NMA) (Figura 2.1), un sistema modello per il gruppo peptidico, in una soluzione di D_2O ; a partire da questo è stato ricostruito lo spettro IR dei peptidi solvatati. Questo protocollo è stato adottato per valutare le frequenze del I modo dell'ammide perturbato per ogni gruppo peptidico (e quindi per ogni residuo) appartenente al GS6 e al GS10

in soluzione. Come QC è stato usato il trans-NMA di cui è stato eseguito un fitting su ogni residuo: gli autovettori corrispondenti (imperturbati) hanno prodotto quindi i modi vibrazionali di ogni gruppo peptidico, che tramite PMM, hanno prodotto lo spettro di frequenza perturbato. La catena laterale del gruppo considerato, gli altri N-1 residui ed il solvente definiscono quindi l'ambiente perturbante ad ogni configurazione generata e le distribuzioni delle frequenze perturbate di ogni residuo formano la banda complessiva caratteristica del I modo dell'ammide.

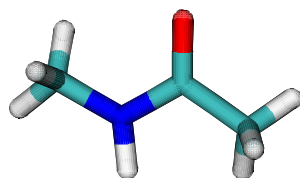


Figura 2.1: Il trans-N-metilacetammide, la molecola usata come modello per studiare il I modo ammidico nei peptidi.

Capitolo 3

Risultati

3.1 Simulazioni di dinamica molecolare

La struttura iniziale del ciclo [(Val-Lys-Leu-dTyr-Pro)₂], è stata creata usando Pymol versione 0.99 (DeLano Scientific) sulla base della struttura chimica sperimentale [5, 6]. E' stata poi eseguita l'ottimizzazione dell'energia usando il potenziale empirico del force field GROMOS96 43a1 [14]. Il peptide, nella sua conformazione iniziale, è stato solvatato con acqua e posto in un box cubico periodico grande abbastanza da contenerlo insieme a 0.5 nm di solvente su tutti i lati. Successivamente sono state protonate le catene laterali delle lisine in modo da riprodurre un pH di circa 7 e sono stati aggiunti due controioni negativi (Cl⁻) in sostituzione di due molecole di acqua, in modo da raggiungere la condizione di neutralità. Le simulazioni MD, nell'ensemble NVT, con lunghezze di legame fisse [49] sono state effettuate con il pacchetto software GROMACS [50] e con il force field GROMOS 43a1, usando per l'acqua il modello single point charge (SPC) [51]. Per le interazioni

di Van der Waals è stato usato un raggio di cut-off di 9.0 Å. Le elettrostatiche sono state trattate con il metodo particle mesh Ewald [52]. La temperatura è stata mantenuta costante al valore desiderato mediante un accoppiamento isocinetico [53]. Dopo vari run MD di equilibramento, sono state effettuate simulazioni di tutti gli atomi in acqua esplicita a temperature e tempi differenti: 200 ns a 280 K, 300 ns a 310 K, 900 ns a 330 K, 300 ns a 335 K, 360 ns a 340 K, 980 ns a 360 K e 500 ns a 400 K.

3.1.1 Analisi delle traiettorie di dinamica molecolare

In questa sezione vengono analizzate le traiettorie ottenute dalle simulazioni alle diverse temperature. Le diverse conformazioni assunte dal peptide sono caratterizzabili strutturalmente attraverso le distanze di quattro legami idrogeno intramolecolari:

1. quello formato tra l'HN della valina 1 e l'O della leucina 8 (Hb1);
2. quello formato tra l'O della valina 1 e l'HN della leucina 8 (Hb2);
3. quello formato tra l'HN della leucina 3 e l'O della Valina 6 (Hb3);
4. quello formato tra l'O della leucina 3 e l'HN della Valina 6 (Hb4).

dove Hb1 e Hb4 sono i legami idrogeno terminali che vanno a costituire i due hairpin, mentre H2 e H3 sono quelli più interni. Naturalmente, essendo la molecola simmetrica, la numerazione utilizzata per i residui e per gli H-bonds è assolutamente arbitraria e non ha alcun significato chimico. Questo ci permette quindi di ridurre le nostre coordinate da $3N$ a 4 facilitando notevolmente l'analisi. Successivamente abbiamo utilizzato queste stesse distanze come base per costruire e diagonalizzare la matrice di covarianza, in modo da ottenere 4 autovettori (le nuove coordinate) su cui proiettare le diverse traiettorie, come descritto nella sezione 2.1.7. Gli autovettori utilizzati in tutte le analisi sono stati quelli ricavati a 330K, che risulta essere la temperatura di equilibrio tra lo stato foldato e quello unfoldato secondo i dati ottenuti da Maness[10]; in Figura 3.1 è riportato lo spettro dei 4 autovettori. La figura mostra che il primo autovalore è molto più grande degli altri e descrive il 75 % della fluttuazione totale del sistema, permettendo da solo una buona descrizione del comportamento del peptide.

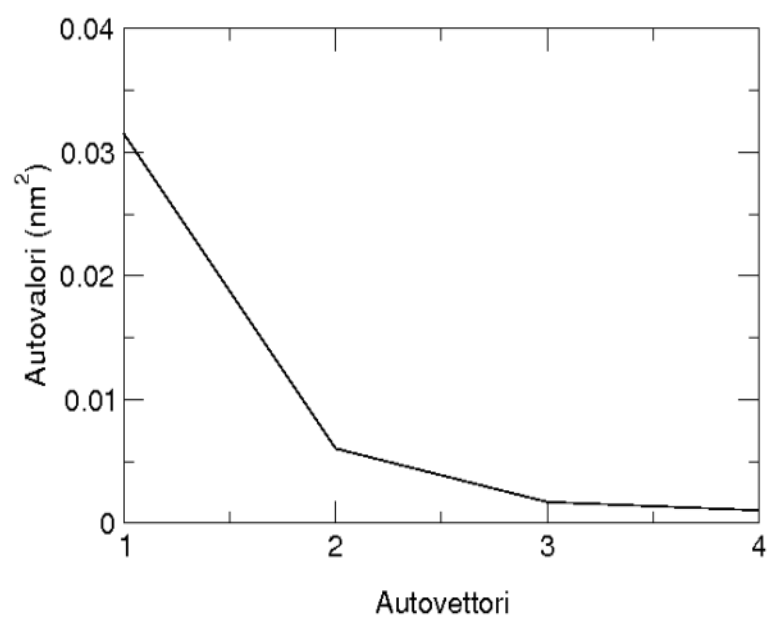


Figura 3.1: Spettro degli autovettori ottenuti a 330 K.

Dalla simulazione eseguita a 280 K vengono riportate le traiettorie delle distanze N-O dei quattro legami idrogeno in Figura 3.2 da cui si vede che tutti i legami si mantengono stabili nel tempo e quindi la molecola rimane nel suo stato foldato. Questo è confermato dalla traiettoria del primo autovettore della matrice di covarianza, che rimane stabile nel tempo (Figura 3.3).

In Figura 3.4 viene riportata la proiezione della traiettoria ottenuta a 280 K nello spazio formato dai primi due autovettori ottenuti a 330 K: la figura mostra che viene esplorata una regione molto ristretta.

Infine, l'istogramma normalizzato delle occorrenze mostra la presenza di un unico massimo intorno a -0.8, relativo allo stato foldato (Figura 3.5).

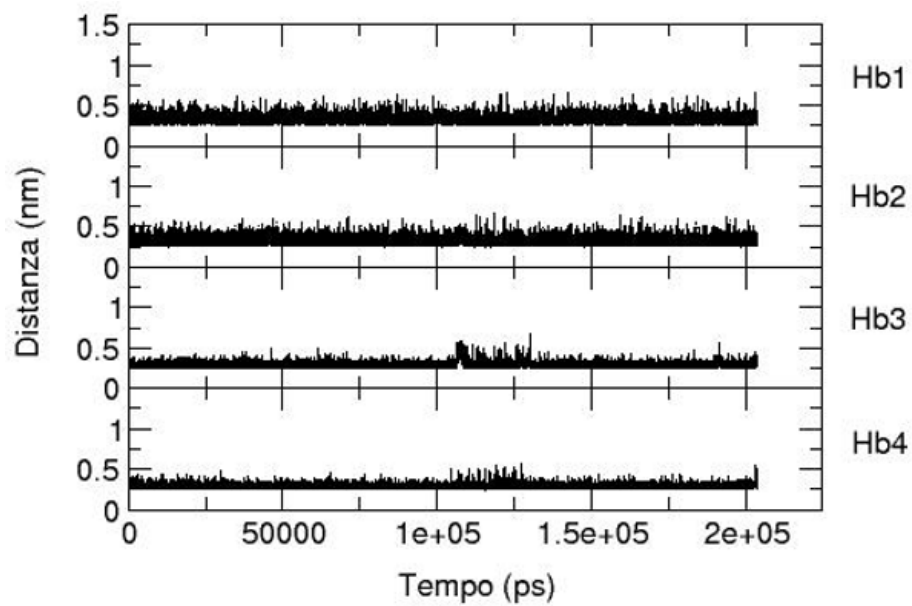


Figura 3.2: Andamento delle distanze dei 4 legami idrogeno nel tempo a 280 K.

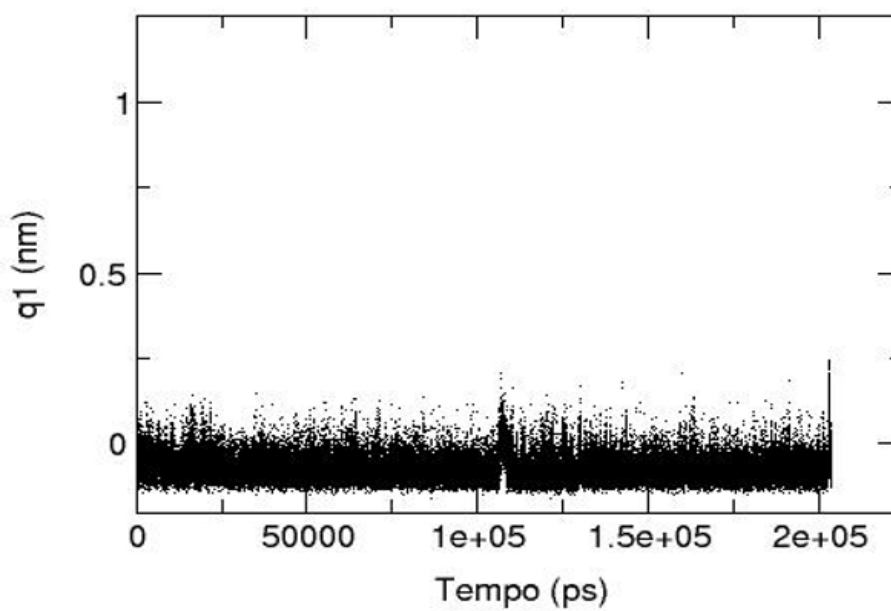


Figura 3.3: Traiettoria ottenuta a 280 K del primo autovettore calcolato a 330 K.

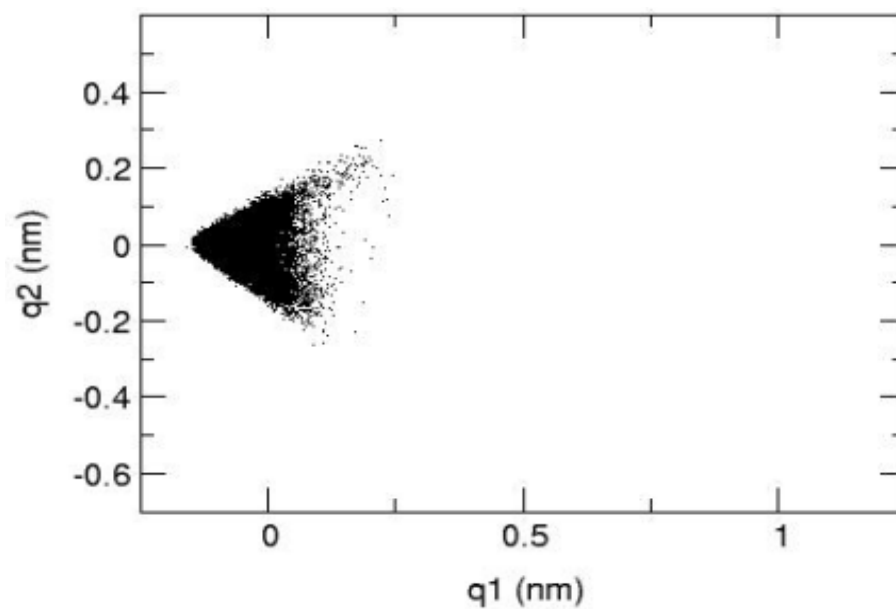


Figura 3.4: Proiezione della traiettoria ottenuta a 280 K sul piano definito dai primi due autovettori calcolati a 330 K.

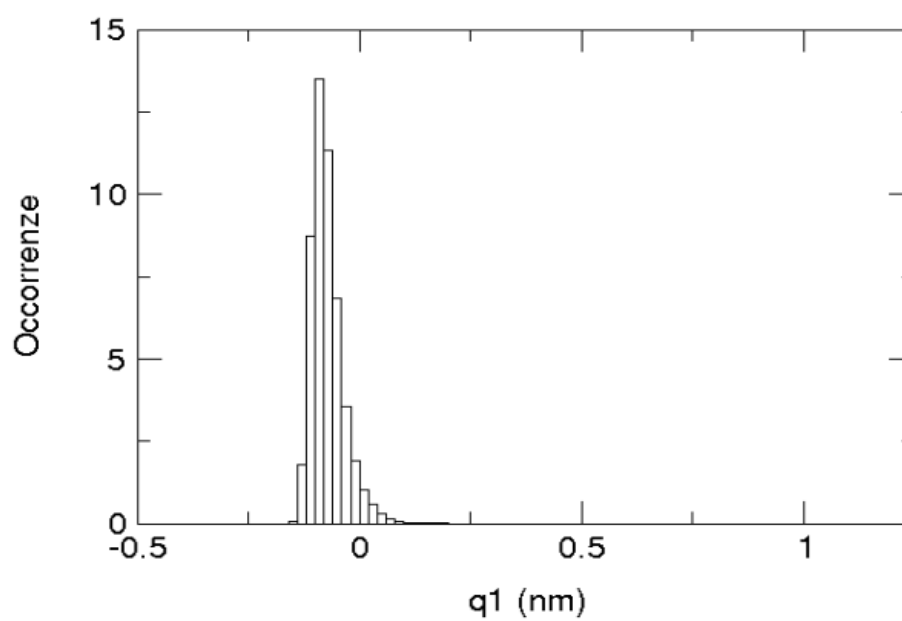


Figura 3.5: Istogramma normalizzato delle occorrenze a 280 K.

La simulazione a 310 K mostra una maggiore fluttuazione rispetto a quella a 280 K, come si può vedere dall'andamento delle distanze dei legami idrogeno nel tempo, ma nessuna transizione significativa (Figura 3.6); ciò è evidente maggiormente dalla traiettoria del primo autovettore calcolato a 330 K in Figura 3.7 in cui si nota la presenza di alcuni picchi isolati, con valori di q_1 fino a 0.5. Dalla proiezione sul piano dei primi due autovettori q_1 e q_2 della matrice di covarianza a 330 K (Figura 3.8) si vede come sia stata esplorata una maggiore porzione dello spazio delle fasi, ma che comprende ancora il solo stato foldato. L'istogramma normalizzato delle occorrenze (Figura 3.9) mostra un andamento molto simile a quello a 280 K anche se più esteso.

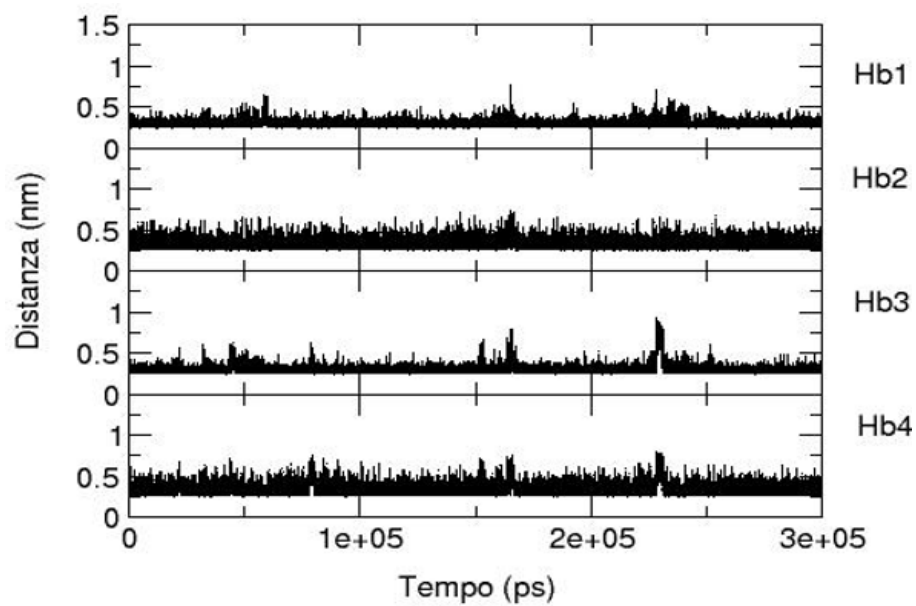


Figura 3.6: Andamento delle distanze dei 4 legami idrogeno nel tempo a 310 K.

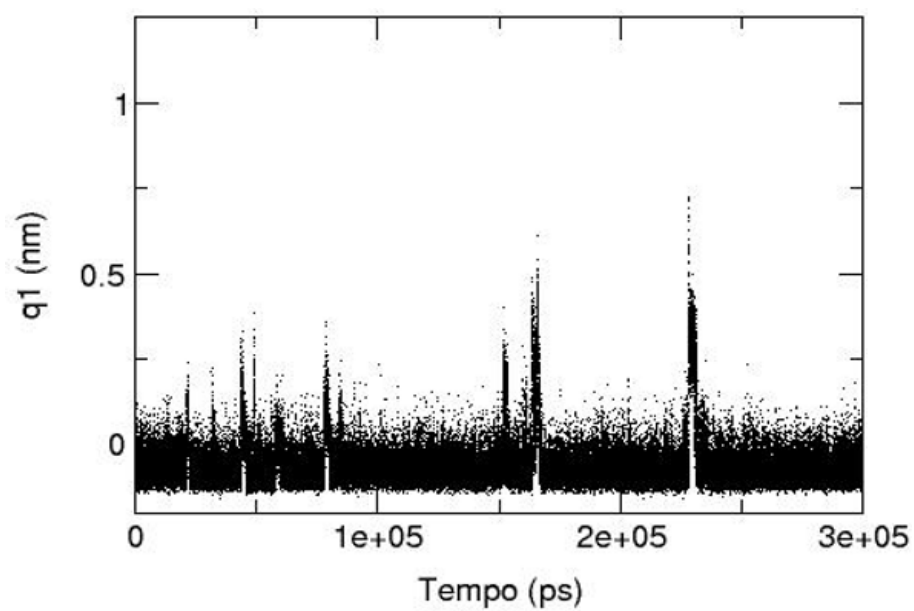


Figura 3.7: Traiettoria ottenuta a 310 K del primo autovettore calcolato a 330 K.

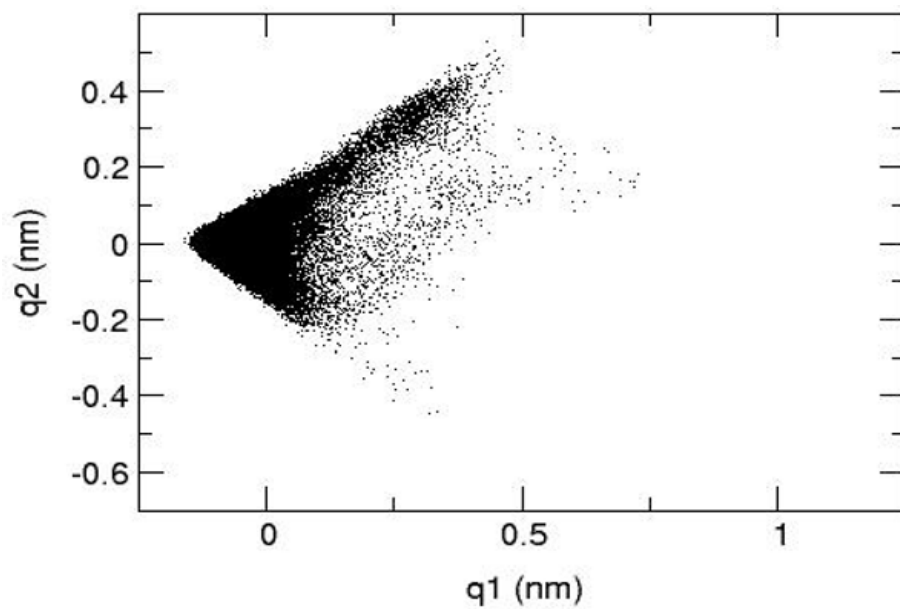


Figura 3.8: Proiezione della traiettoria ottenuta a 310 K sul piano definito dai primi due autovettori calcolati a 330 K.

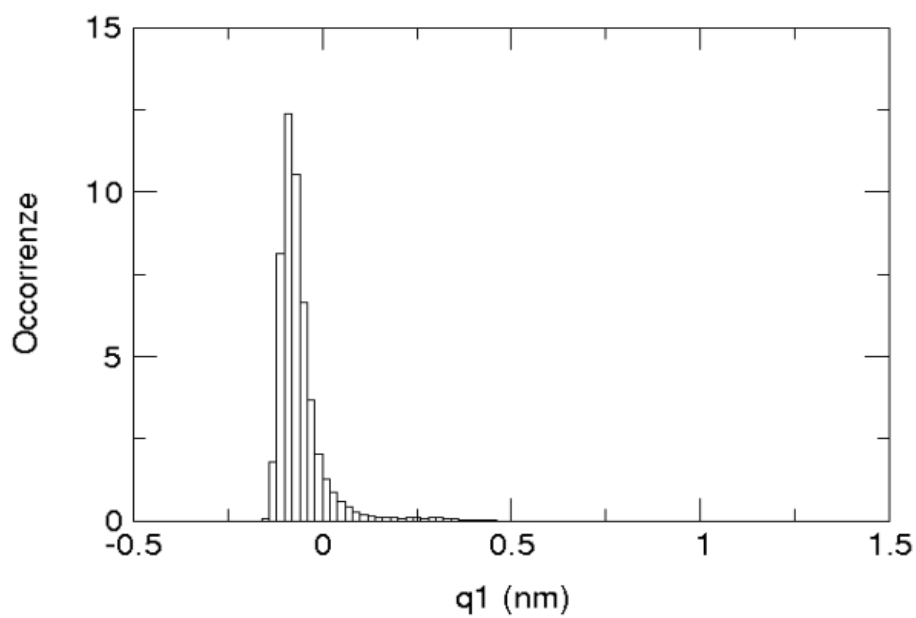


Figura 3.9: Istogramma normalizzato delle occorrenze a 310 K.

Le simulazioni a 330 K e 335 K, temperature prossime a quella di equilibrio, mostrano la presenza di alcune transizioni tra i due stati, come evidenziato dalle Figure 3.10 A e 3.10 B; ciò risulta maggiormente evidente dalle traiettorie del primo autovettore della matrice di covarianza calcolato a 330 K, caratterizzate da valori di q_1 intorno a 0.5, ad identificare lo stato unfoldato (Figure 3.11 A e 3.11 B). Lo spazio campionato è maggiore che nelle simulazioni precedenti, come ci si può aspettare date le temperature più elevate e come si vede bene dalle proiezioni sul piano definito dai primi due autovettori q_1 e q_2 calcolati a 330 K (Figure 3.12 A e 3.12 B), in cui si identificano bene stato foldato e stato unfoldato. Sono riportati anche gli istogrammi normalizzati delle occorrenze (Figure 3.10 A e 3.10 B) a confronto dove si possono osservare due massimi: uno a -0.8 relativo allo stato foldato e l'altro a 0.6 relativo invece allo stato unfoldato.

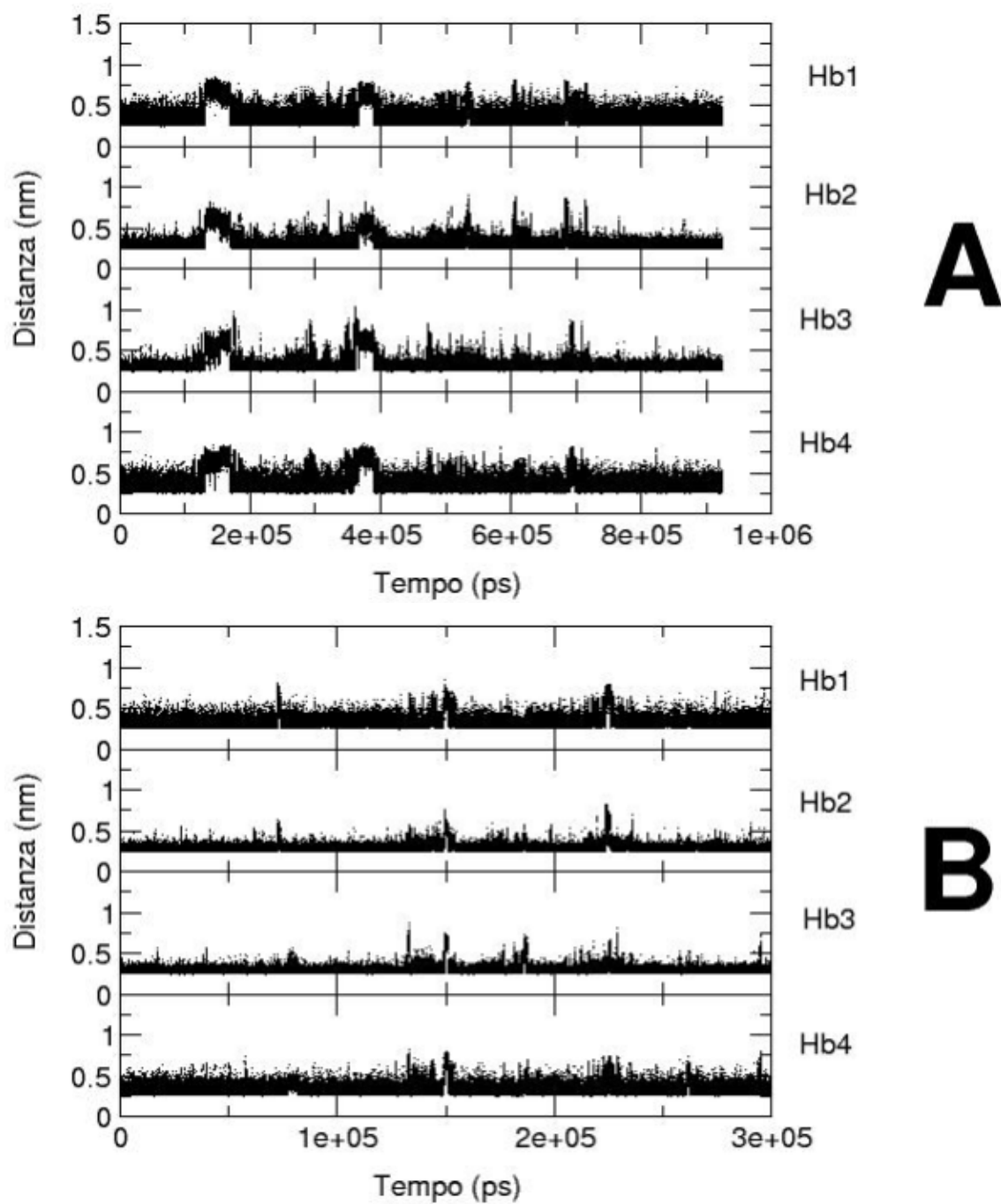


Figura 3.10: (A) Andamento delle distanze dei 4 legami idrogeno nel tempo a 330 K e (B) a 335 K.

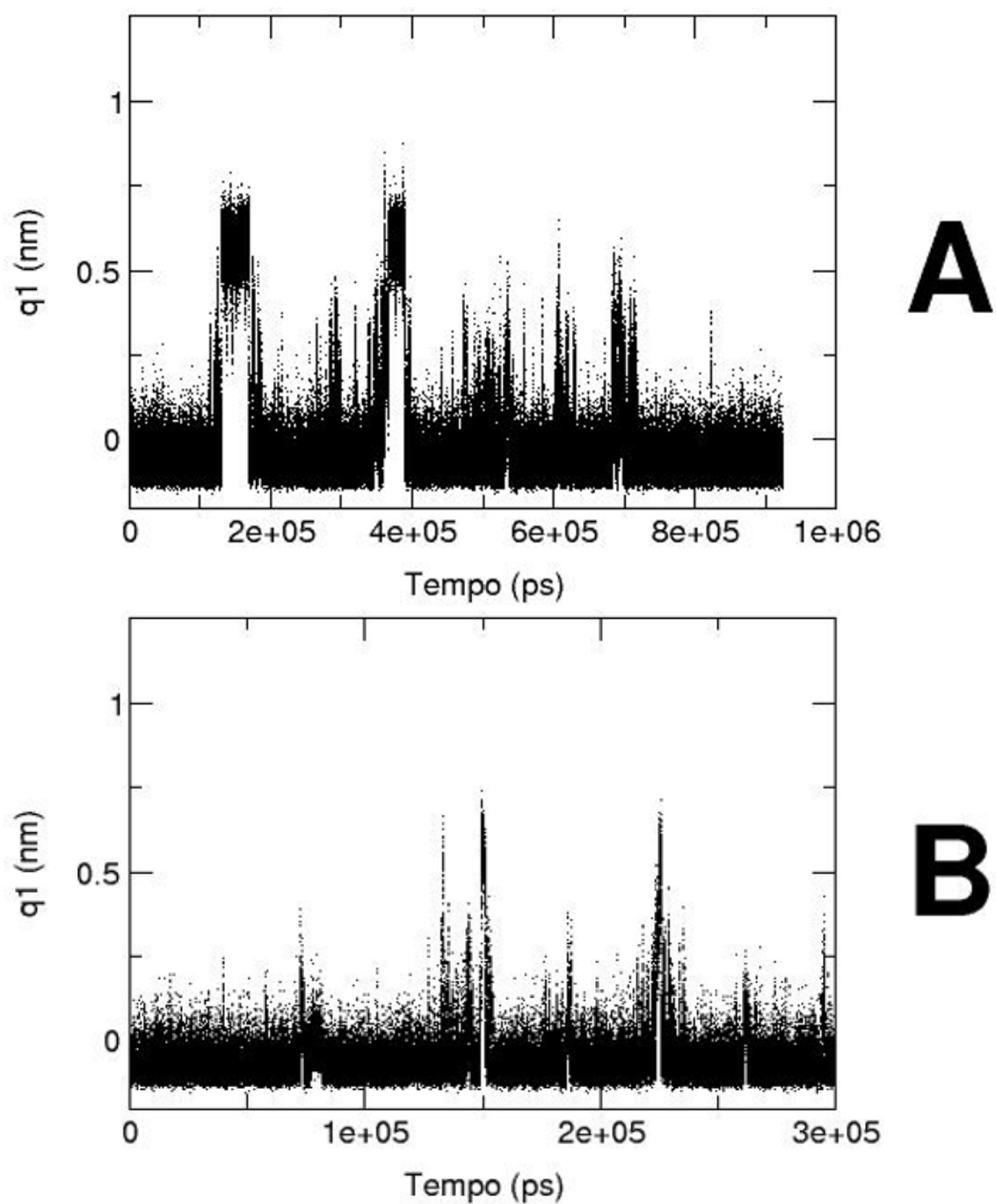


Figura 3.11: (A) Traiettoria del primo autovettore calcolato a 330 K ottenuta dalla simulazione a 330 K e (B) a 335 K.

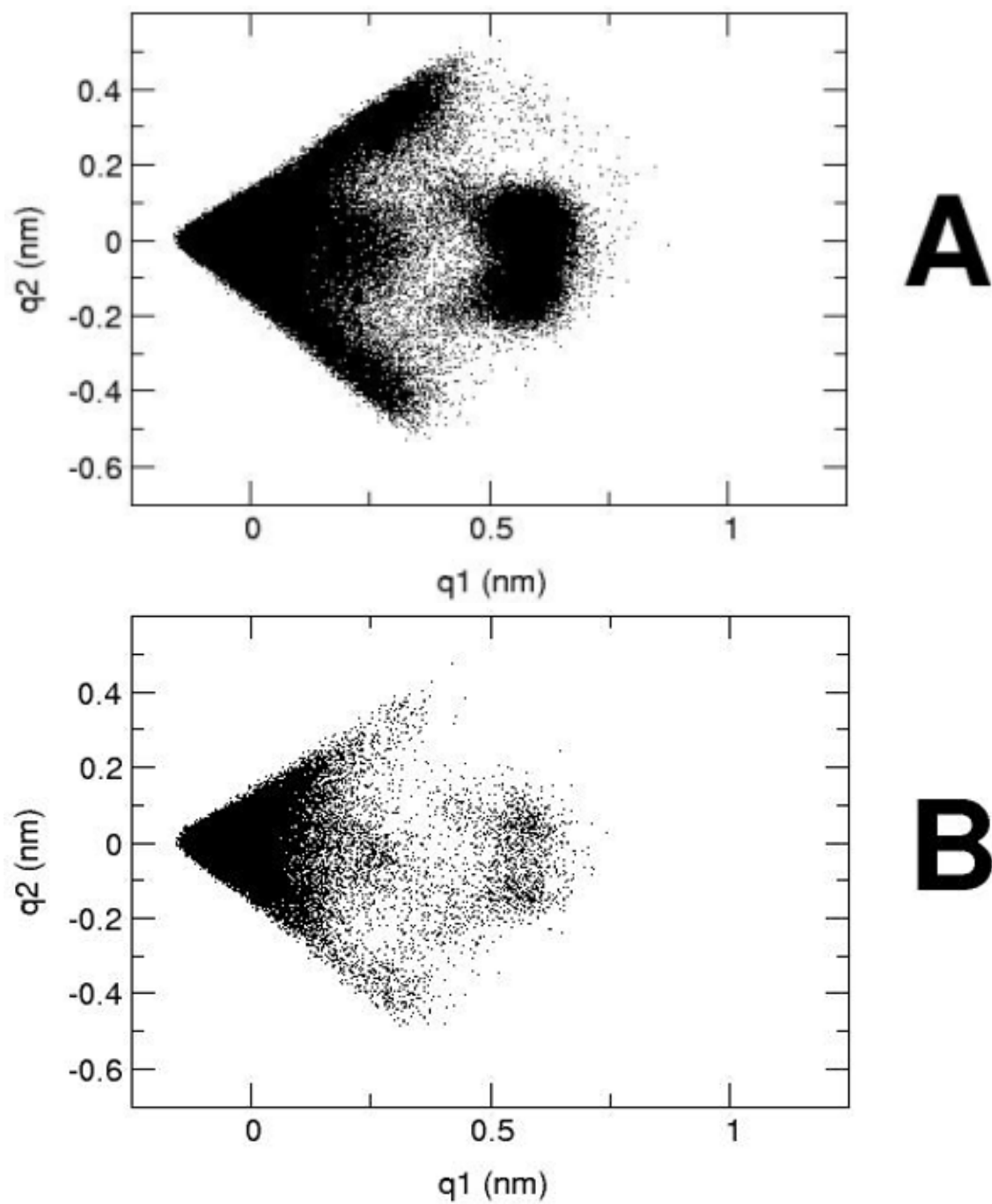


Figura 3.12: (A) Proiezione sul piano definito dai primi due autovettori calcolati a 330 della traiettoria a 330 K e (B) a 335 K.

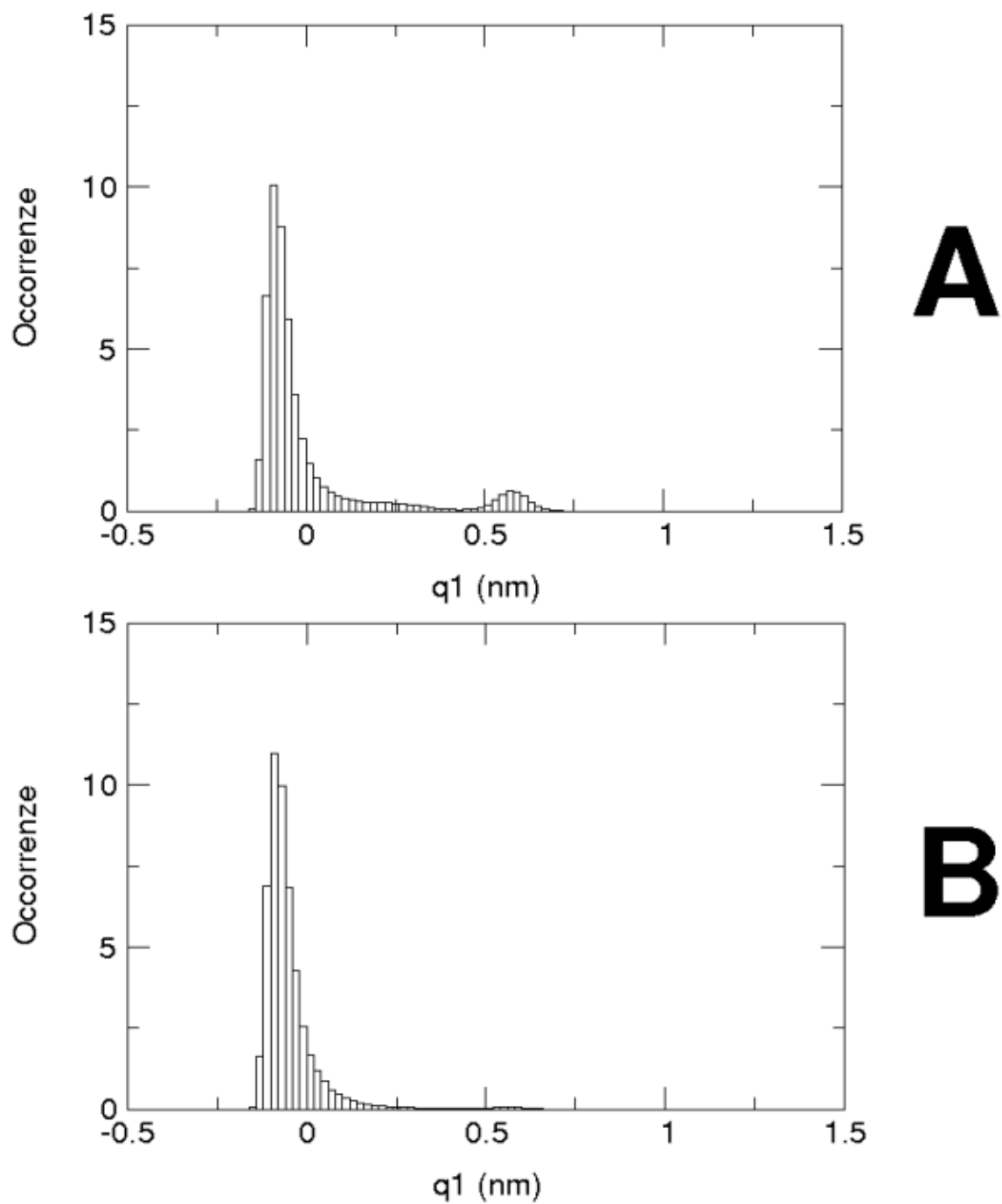


Figura 3.13: (A) Istogramma normalizzato delle occorrenze a 330 K e (B) a 335 K.

A 340 K e 360 K, le simulazioni sono dominate principalmente da una transizione dallo stato foldato a quello unfoldato nel quale il peptide permane per gran parte del tempo di simulazione. Ciò si vede bene dall'andamento delle distanze dei legami idrogeno (Figure 3.14 A e 3.14 B), che passano da 0.3 nm, valore caratteristico dello stato foldato, a 0.7 per quelli terminali e ≈ 1 nm per gli H-bonds interni, valori che identificano invece lo stato unfoldato. Le traiettorie del primo autovettore calcolato a 330 K, oltre a sottolineare quanto già detto per le distanze dei legami idrogeno, ci permettono di identificare la presenza di alcuni stati intermedi, caratterizzati da valori di q_1 intorno a 0.5, compresi tra 0 (stato foldato) e 0.7 (stato unfoldato) (Figure 3.15 A e 3.15 B). Infine la proiezione delle traiettorie sul piano dei primi due autovettori calcolati dalla matrice di covarianza a 330 K, ci mostra come sia stato esplorato in buona parte l'intero spazio accessibile al peptide (Figure 3.16 A e 3.16 B). Infine in Figure 3.17 A e 3.17 B vengono mostrati gli istogrammi normalizzati delle occorrenze dove si notano i due massimi a -0.8 (stato foldato) e 0.8 (stato unfoldato).

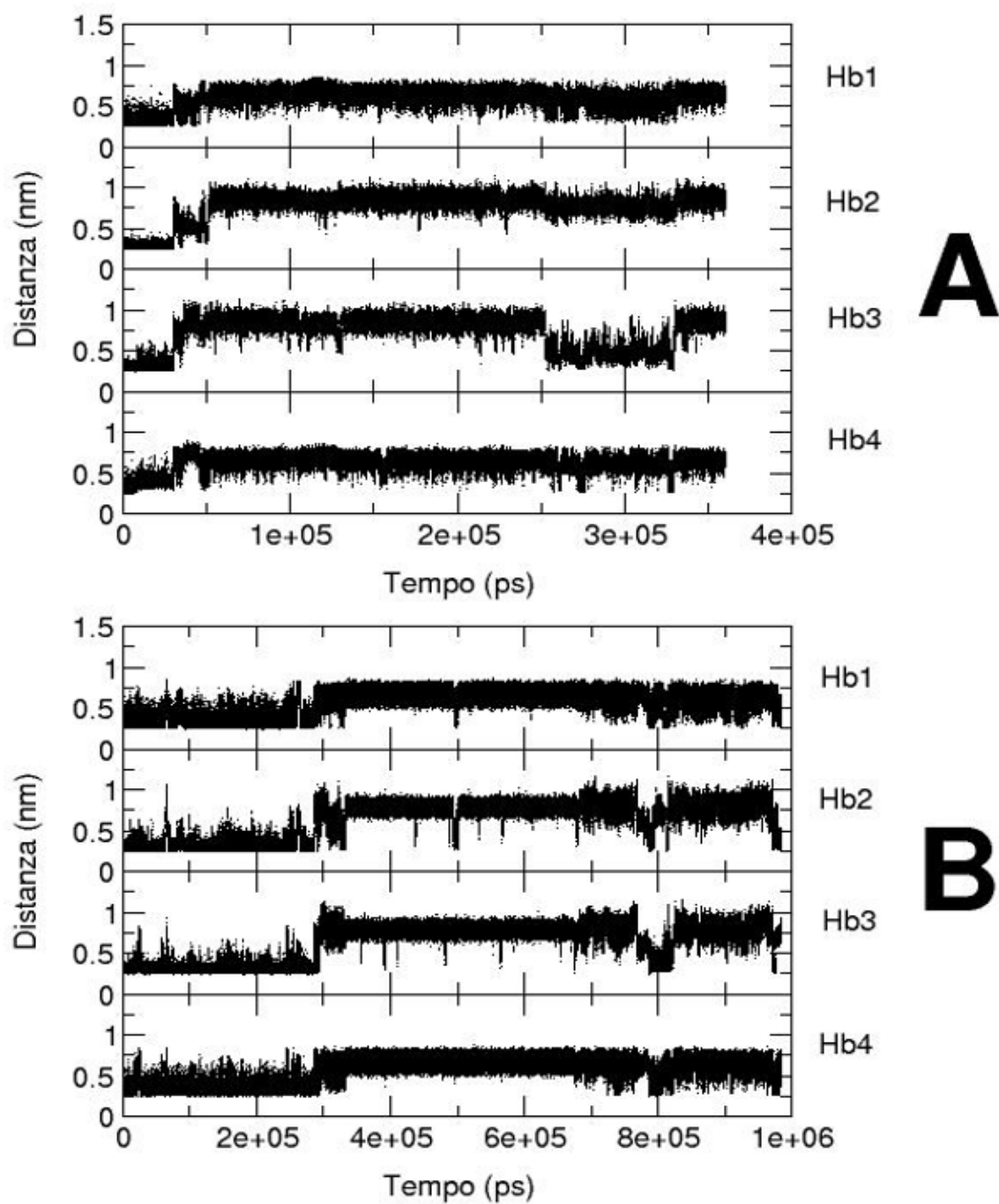


Figura 3.14: (A) Andamento delle distanze dei 4 legami idrogeno nel tempo a 340 K e (B) a 360 K.

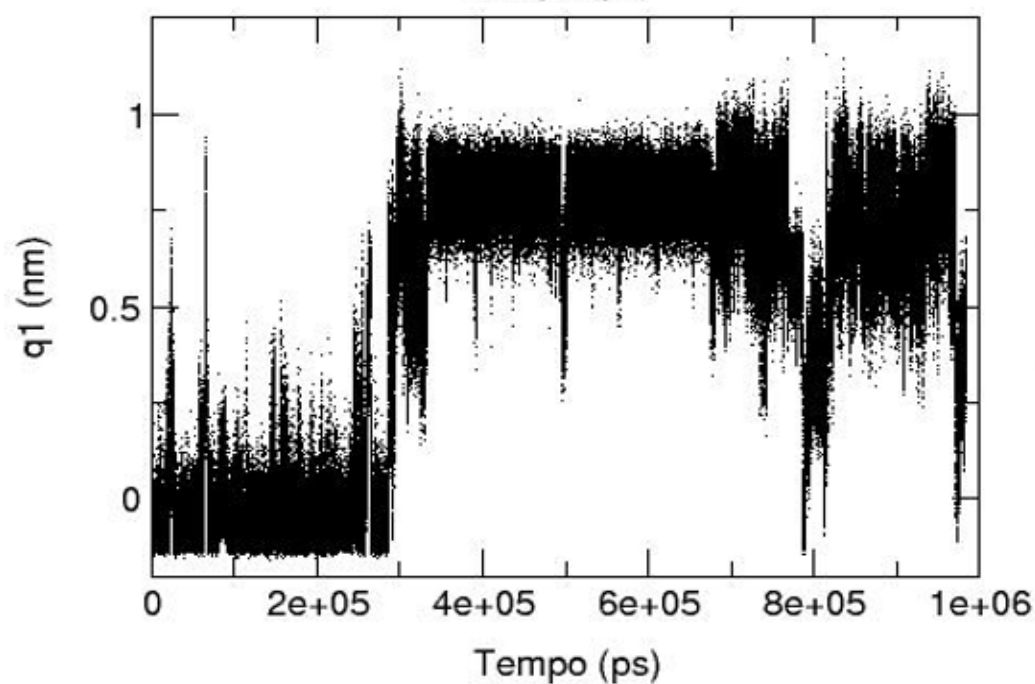
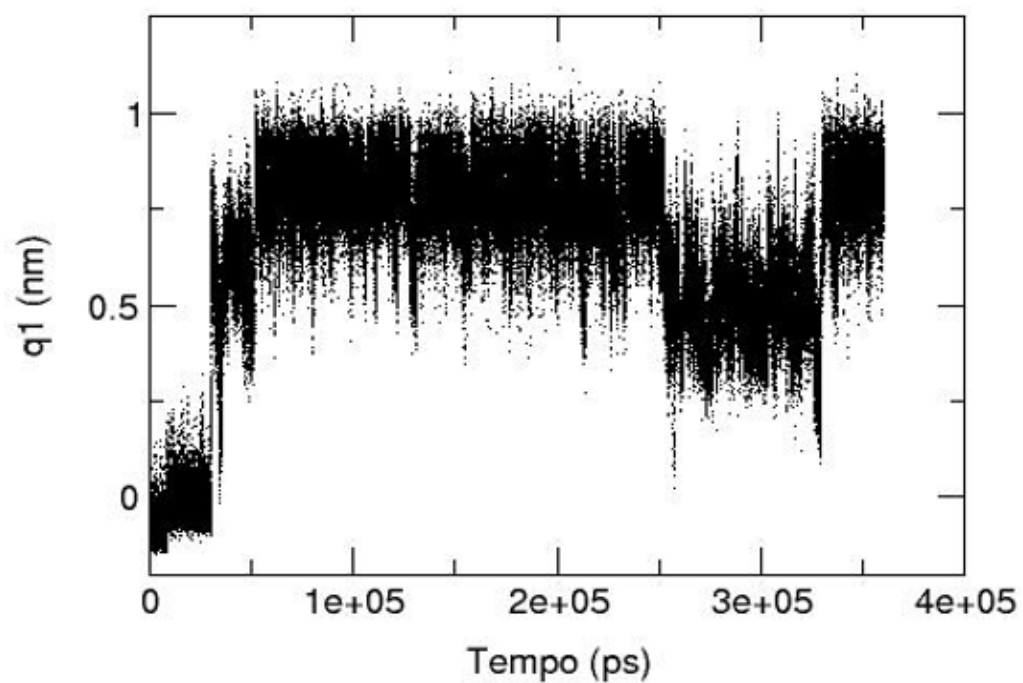


Figura 3.15: (A) Traiettoria del primo autovettore calcolato a 330 K ottenuta dalla simulazione a 340 K e (B) a 360 K.

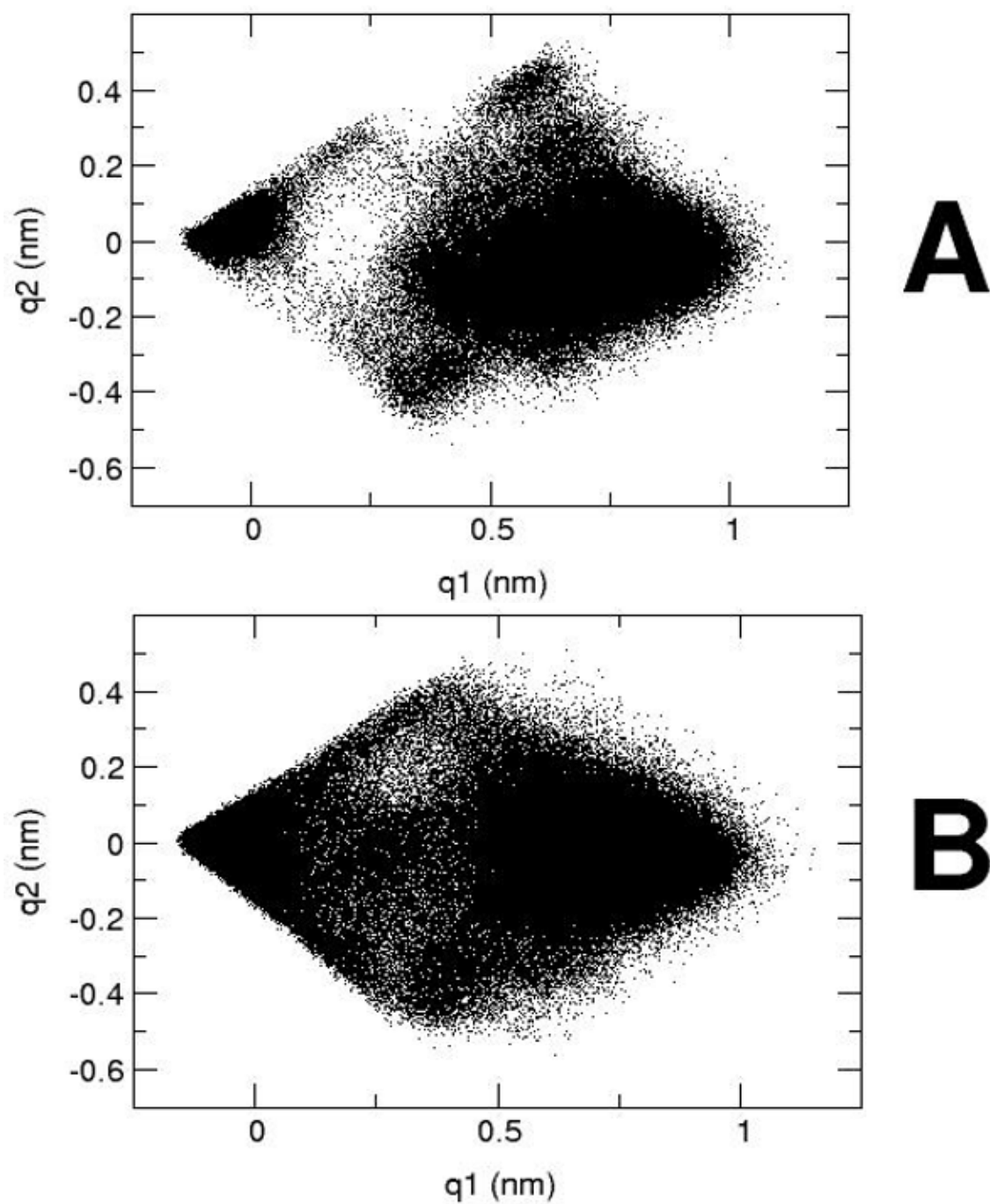


Figura 3.16: (A) Proiezione sul piano definito dai primi due autovettori calcolati a 330 della traiettoria a 340 K e (B) a 360 K.

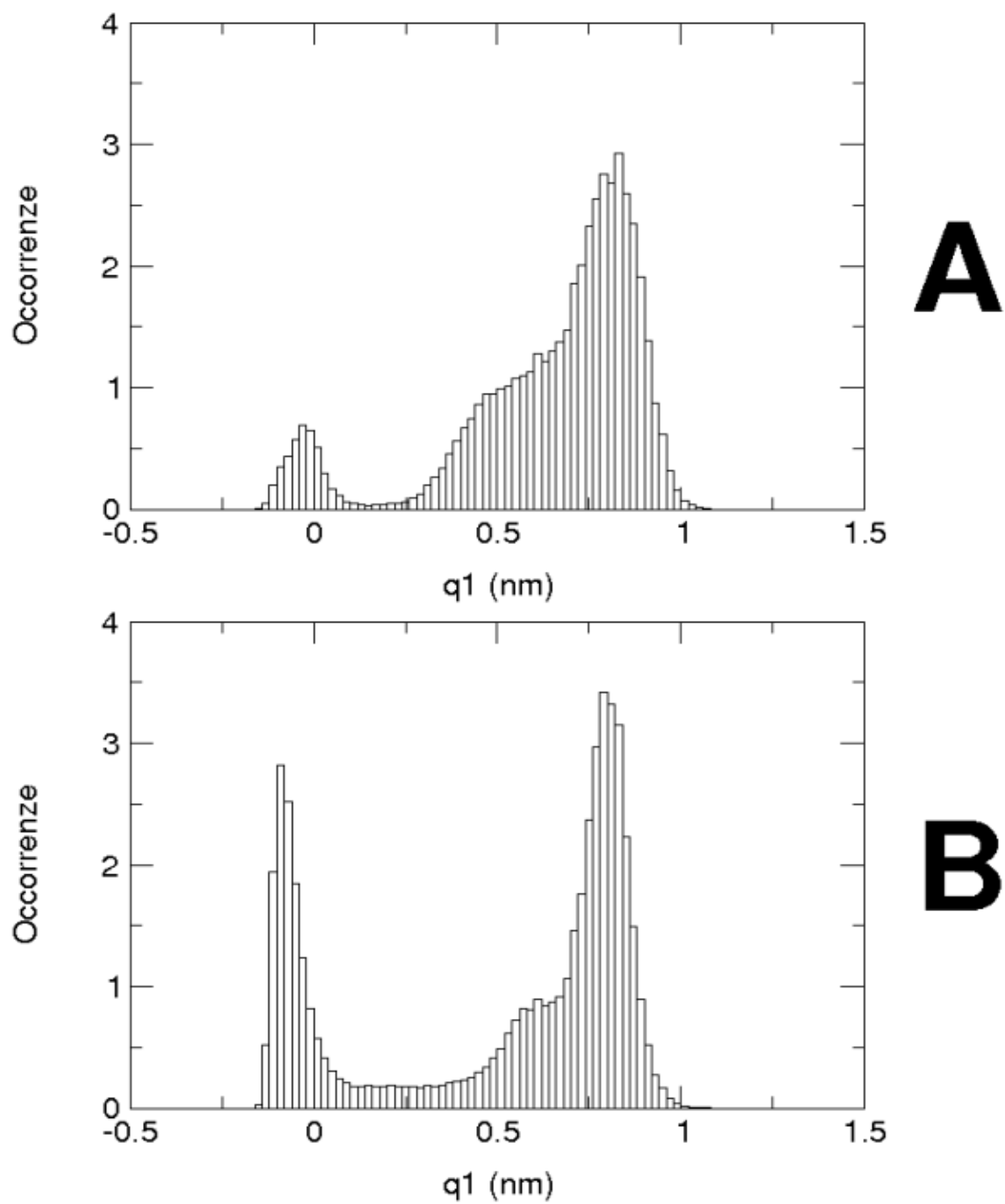


Figura 3.17: (A) Istogramma normalizzato delle occorrenze a 340 K e (B) a 360 K.

3.1.2 Calcolo della costante cinetica di folding

Pur non potendo calcolare in modo rigoroso la costante cinetica di folding, è stato tuttavia possibile stimarne l'ordine di grandezza da confrontare con il dato sperimentale, ottenuto tramite spettroscopia infrarossa mediante trasformata di Fourier all'equilibrio e rilassamento dopo T-jump [10].

A questo scopo abbiamo utilizzato l'analisi essenziale effettuata sulla traiettoria a 330 K perché, in quanto temperatura intermedia, presentava un maggior numero di transizioni. Per poter procedere è stato necessario prima dare una definizione operativa di transizione:

1. dall'istogramma (Figura 3.18) calcolato sulla proiezione della traiettoria sul primo autovettore, abbiamo ricavato una soglia di riferimento che separa i due stati e i loro massimi di probabilità;
2. viene considerata come transizione non solo il semplice passaggio oltre la soglia di riferimento, ma anche il raggiungimento del massimo di probabilità dello stato verso cui il sistema transisce.

Seguendo questa definizione, abbiamo analizzato la traiettoria punto a punto, riportando poi gli inizi e le durate dei due stati tra una transizione e l'altra (Tabella 3.1). Abbiamo poi calcolato i tempi di vita media dello stato foldato e di quello unfoldato; i reci-

proci di questi valori corrispondono rispettivamente alla costante cinetica di folding e di unfolding (Tabella 3.2).

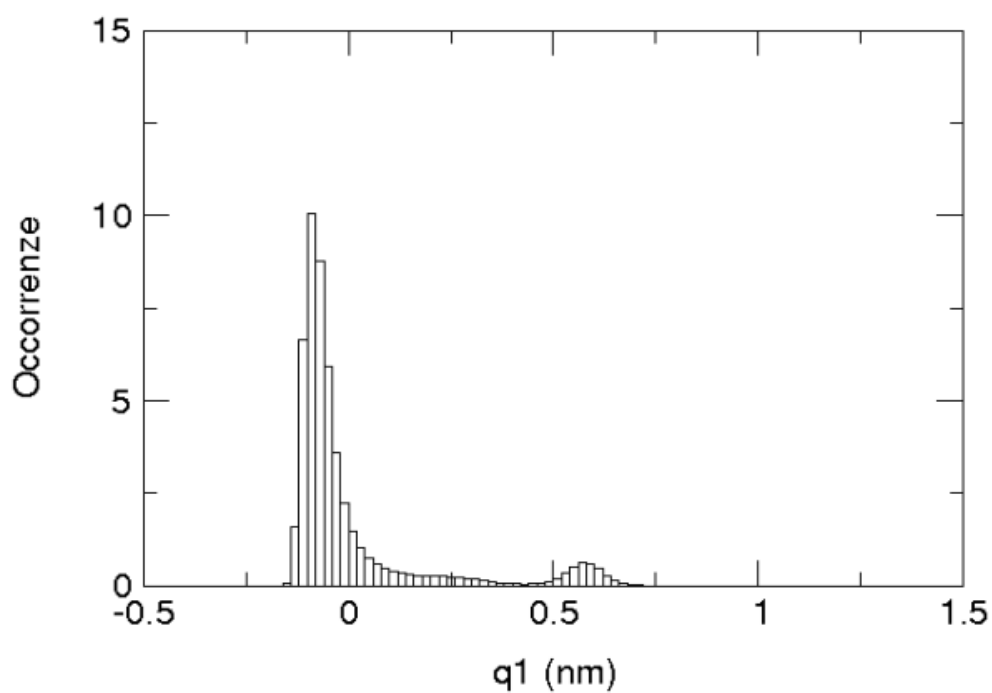


Figura 3.18: Istogramma normalizzato delle occorrenze calcolato dalla traiettoria a 330 K.

Tabella 3.1:

Inizio (ns)	Durata (ns)	Stato
0	130	F
130	40	U
170	192	F
362	28	U
390	216	F
606	1	U
607	87	F
694	2	U
696	227	F

Inizio e tempo di vita degli stati foldato (F) e unfoldato (U) tra una transizione e l'altra.

Tabella 3.2:

Stato	Tempo di vita medio (ns)	costante cinetica (s)	costante cinetica (s) sperimentale
F	170.4	$0.59 \cdot 10^7$	$0.7 \cdot 10^7$
U	18.0	$5.6 \cdot 10^7$	$0.7 \cdot 10^7$

Tempo di vita media dello stato foldato (F) e unfoldato (U) e relative costanti cinetiche (foldato = K_f , unfoldato = K_u), confrontate con quelle ottenute sperimentalmente [10].

La costante cinetica della transizione tra lo stato foldato e quello unfoldato è stata ottenuta sperimentalmente mediante studio della cinetica di rilassamento in seguito al T-jump indotto tramite laser [10]. Da un confronto con i dati sperimentali [10] si può vedere come, specie per la costante di folding, sia stato stimato correttamente l'ordine di grandezza, a confermare l'affidabilità delle simulazioni; tuttavia, a causa della condizione di equilibrio fortemente spostata verso lo stato foldato, lo stato unfoldato ha un tempo di vita molto ridotto e ciò porta ad ottenere una costante cinetica di unfolding affetta da un non drammatico errore.

Dal confronto tra gli istogrammi ottenuti alle diverse temperature, ci si potrebbe chiedere perché come traiettoria di riferimento per le varie analisi, sia stata scelta quella ottenuta dalla simulazione a 330 K (Figura 3.13 A), in cui l'equilibrio è molto spostato verso lo stato foldato, piuttosto che quelle ottenute dalla simulazioni a temperature superiori (Figure 3.17 A e 3.17 B), in cui gli stati foldato e unfoldato mostrano praticamente la stessa probabilità. Questa scelta è dovuta al fatto che a 330 K, pur con una situazione molto sbilanciata, sono presenti diverse transizioni complete: il sistema infatti transisce dallo stato foldato a quello unfoldato, per poi ritornare allo stato iniziale. A 340 K come a 360 K invece, pur avendo un miglior rapporto tra le due conformazioni, siamo in presenza di una sola transizione foldato/unfoldato, senza un successivo ritorno allo stato iniziale; per questo motivo pur con le già citate problematiche abbiamo scelto 330 K come

temperatura di riferimento, perché maggiormente rappresentativa dell'intero sistema e della sua evoluzione nel tempo.

3.1.3 Confronto con il GS6

Per quanto riguarda il GS6 invece, un precedente studio svolto in questo laboratorio [11], ha prodotto risultati particolarmente accurati, in accordo con quelli sperimentali, ottenuti sempre tramite spettroscopia IR in trasformata di Fourier all'equilibrio e rilassamento dopo T-jump [10]. Per far ciò sono state eseguite una serie di simulazioni di dinamica molecolare a diverse temperature, in modo analogo a quanto effettuato in questa sede per il GS10; le proprietà osservate sperimentalmente sono state riprodotte quantitativamente mediante l'uso di un modello meccanico statistico basato sulla QGE (*Quasi Gaussian Entropy Theory*) per descrivere la termodinamica di folding/unfolding del peptide.

E' stato possibile utilizzare tale procedura perché, a differenza del GS10, il peptide mostra un gran numero di transizioni complete (foldato-unfoldato-foldato) alle varie temperature: ciò grazie ad una minore libertà conformazionale e alle minori dimensioni rispetto al GS10, che hanno permesso quindi una maggiore esplorazione dello spazio delle fasi in tempi più brevi. Questa fluttuazione è essenziale per poter eseguire un'analisi di tipo statistico, quale quella basata sulla QGE.

3.2 Calcoli degli spettri infrarosso

Spettri IR di questi peptidi, sono stati ottenuti sperimentalmente dallo studio del I modo ammidico (corrispondente essenzialmente allo stretching del C=O) mediante l'uso della spettroscopia infrarossa in trasformata di Fourier e della sua variante time-resolved [7, 8].

In questa sezione viene descritta la procedura per calcolare gli spettri relativi al GS6 ed al GS10, vengono descritti i risultati ottenuti e confrontati con quelli sperimentali.

Come modello del gruppo peptidico, cioè il centro quantistico trattato esplicitamente a livello elettronico, è stato scelto il trans-N-metilacetammide (trans-NMA). Calcoli quantomeccanici sono stati svolti sulla molecola isolata mediante la teoria del funzionale di densità dipendente dal tempo (TDDFT) con il basis set 6-31+G(d); questo livello di teoria rappresenta infatti un buon compromesso tra costo computazionale e accuratezza. La matrice Hessiana massa-pesata è stata calcolata sulla geometria ottimizzata a livello di teoria B3LYP/6-31+G(d) e successivamente diagonalizzata, ottenendo autovettori imperturbati e relativi autovalori.

L'autovettore corrispondente al modo I dell'ammide nel vuoto è stato utilizzato quindi per generare una griglia di punti (cioè configurazioni): è stato adottato uno step di 0.05 a.u. ed è stato determinato un numero di punti tale da esplorare un intervallo di

energia di 20 kJ/mol (nel caso presente 31 punti). Per ogni punto, sei stati elettronici imperturbati sono stati poi valutati allo stesso livello di teoria che fornisce il basis set per i calcoli PMM.

3.2.1 Spettri infrarosso del trans-NMA

La frequenza sperimentale di vibrazione del I modo dell'amide del trans-NMA isolato va da 1707 cm^{-1} (in una matrice di argon a 20 K [54]) fino a $1714\text{-}1731\text{ cm}^{-1}$ (in fase gassosa a $\approx 100\text{ K}$ [55]) a seconda delle condizioni sperimentali. La frequenza valutata mediante ottimizzazioni della geometria standard e calcolata tramite l'approssimazione armonica (sezione 2.2.1), è risultata pari a 1754 cm^{-1} , 47 cm^{-1} più alta rispetto a quella sperimentale (prendendo come valore di riferimento quello registrato nella matrice d'argon a 20 K [54]); questa differenza è abbastanza comune per questo tipo di calcoli [54, 56] e può essere attribuita ad incertezze del calcolo quantistico, tra cui la stessa approssimazione armonica [57].

Lo spettro sperimentale del trans-NMA nella regione del I modo ammidico in D_2O a temperatura ambiente [58] viene mostrato in Figura 3.19. Il picco massimo corrisponde a circa $1622\text{ - }1623\text{ cm}^{-1}$, cioè $\approx 80\text{ - }85\text{ cm}^{-1}$ più bassa rispetto a quella del composto isolato (frequenza imperturbata) [55, 59], con una massima ampiezza a mezza altezza (fwhm) pari a $\approx 30\text{ cm}^{-1}$. La banda calcolata mediante la procedura PMM/MD è caratterizzata da un massimo a 1698 cm^{-1} , spostato cioè di 56 cm^{-1} , e comprende quindi la maggior parte dello shift negativo osservato speri-

talmente. Inoltre, la forma dello spettro e l'ampiezza sono ben riprodotte dai calcoli PMM/MD ($\text{fwhm} = 30 \text{ cm}^{-1}$). La differenza di $25\text{-}30 \text{ cm}^{-1}$ tra i due spettri è dovuta ad una combinazione di effetti di ordine superiore, incluse piccole inesattezze dei dipoli calcolati nella matrice $\tilde{Z}_1$ (vedere sezione 2.2.1), ed il fatto che non viene considerato alcun accoppiamento eccitonico e/o comportamento anarmonico.

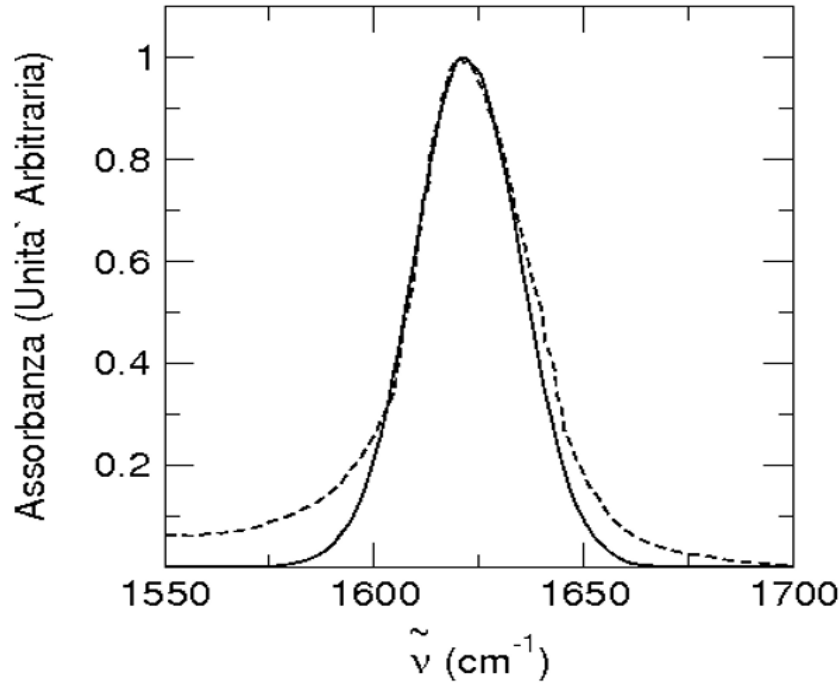


Figura 3.19: Confronto tra lo spettro IR calcolato e quello sperimentale [64] della regione I ammidica del trans-NMA in una soluzione di D_2O . Per un migliore confronto, le frequenze calcolate sono state traslate uniformemente ad un valore più basso di 76 cm^{-1} , in modo da allinearle al picco sperimentale centrato a 1622 cm^{-1} .

3.2.2 Spettri infrarosso del β -hairpin

La termodinamica e la cinetica del GS6 sono state in precedenza [11] caratterizzate quantitativamente, tramite simulazioni MD e l'uso di un modello meccanico-statistico basato sulla *Quasi Gaussian Entropy Theory* (QGE) [60] - [63], ottenendone i seguenti risultati:

1. lo stato unfoldato è risultato termodinamicamente più stabile nell'intero intervallo di temperatura considerato, presentando una maggiore stabilità per tutte le temperature;
2. lo stato unfoldato è caratterizzato da un più basso valore di energia ed entropia;
3. i tempi di vita media dei due stati sono dell'ordine di nano-secondi.

I risultati concernenti le caratteristiche strutturali della conformazione foldata, l'ordine di grandezza della variazione di potenziale chimico tra i due stati e le costanti cinetiche della transizione sono risultate inoltre in buon accordo con i dati sperimentali disponibili [10, 6]; le variazioni di energia ed entropia hanno mostrato invece segno opposto rispetto ai valori sperimentali: ciò si traduce in un incremento della popolazione dello stato foldato all'aumentare della temperatura.

In base a queste considerazioni, sono stati ricostruiti gli spettri in funzione della temperatura a partire dalle traiettorie simulate,

da confrontare con quelli sperimentali, per offrire una giustificazione agli stessi ed ai dati termodinamici. Lo spettro IR ottenuto da Maness et al. [10] mostra un picco centrato ad $\approx 1630 \text{ cm}^{-1}$, corrispondente alla banda I ammidica; questo picco è stato studiato a diverse temperature (da 2°C a 85°C) tramite spettroscopia d'equilibrio FTIR in modo da effettuare una stima della termodinamica dei diversi stati, assumendo che alle temperature estreme la popolazione si trovi interamente in un unico stato (foldato a 2°C , unfoldato a 85°C). All'aumentare della temperatura, il picco si sposta a frequenze maggiori e la sua intensità diminuisce: di conseguenza, gli spettri ottenuti per sottrazione di quello ottenuto alla temperatura più bassa, mostrano un andamento negativo-positivo lungo l'asse delle frequenze. Il corrispondente spettro calcolato a 6 temperature diverse (280 K, 310 K, 360 K, 400 K, 500 K, e 600 K) viene mostrato in Figura 3.20. Nella stessa figura vengono mostrati gli spettri ottenuti sottraendo lo spettro a 280 K da quelli a 360 K e 310 K. Dalla figura si nota ancora come la procedura PMM/MD ben approssimi i dati sperimentali [10], riproducendo la forma dello spettro ($\text{fwhm} \approx 35 \text{ cm}^{-1}$ contro $\approx 40 \text{ cm}^{-1}$ del valore sperimentale) e l'andamento con la temperatura. Per una migliore comprensione, gli spettri calcolati sono stati traslati di $\approx 95 \text{ cm}^{-1}$ a frequenze più basse, allineando in tal modo la banda del I modo ammidico con il picco massimo ottenuto sperimentalmente.

Tale shift, dovuto in gran parte alle inesattezze dei calcoli *ab-initio*, è stato poi applicato a tutti i segnali IR del GS6 qui riportati.

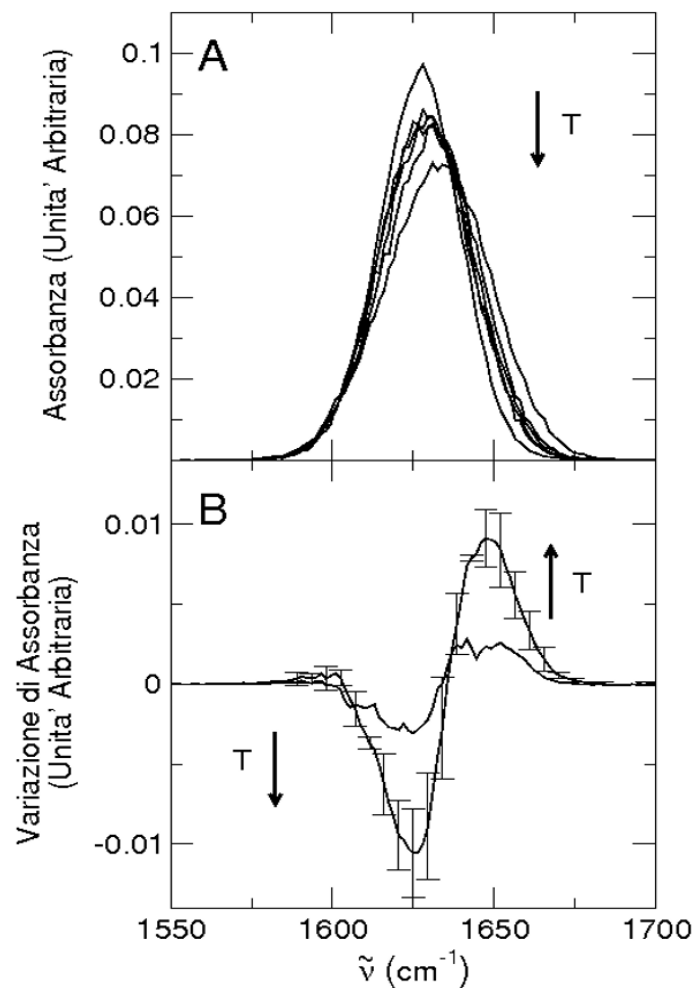


Figura 3.20: A: Spettri IR nella regione I ammidica del GS6 in D_2O elaborati a 280 K, 310 K, 360 K, 400 K, 500 K, e 600 K. - B: Spettri ottenuti sottraendo quello calcolato a 280K dagli spettri elaborati a 310 K e 360 K. Nella figura viene riportata l'incertezza corrispondente all'errore standard valutato su tre traiettorie indipendenti.

Gli spettri alle diverse temperature sono stati utilizzati per monitorare l'andamento delle popolazioni dei diversi stati, in modo da ricavare informazioni sulla cinetica della transizione e la termodinamica del β -hairpin, assumendo che all'aumentare della temperatura corrisponda una popolazione dello stato unfoldato maggiore.

In seguito abbiamo applicato la procedura mista PMM/MD agli ensemble foldato e unfoldato derivanti dall'analisi delle traiettorie MD (vedi sezione 2.2.2), ricostruendo poi gli spettri dei due stati. In Figura 3.21 sono riportati gli spettri elaborati del GS6, ottenuti dalla differenza foldato-unfoldato, a 280 K, 310 K e 360 K (da notare che risultano praticamente indistinguibili, nei limiti del rumore di fondo). Questi spettri mostrano un andamento positivo-negativo all'aumentare della frequenza, in contrasto a quanto accade in quelli sperimentali e negli spettri ricostruiti al variare della temperatura: questo risulta in accordo con i risultati termodinamici [11] a dimostrazione della maggiore stabilità energetica dello stato foldato ad una temperatura più alta (in accordo con quanto ottenuto in un precedente lavoro svolto in questo laboratorio [11]); di conseguenza è giustificato l'andamento opposto tra i due tipi di spettri, quello foldato/unfoldato e quello sulla temperatura. Si può notare infine l'aumento di popolazione dello stato foldato all'aumentare della temperatura grazie all'effetto spettroscopico, insieme al contributo dello stato foldato e unfoldato alle temperature considerate.

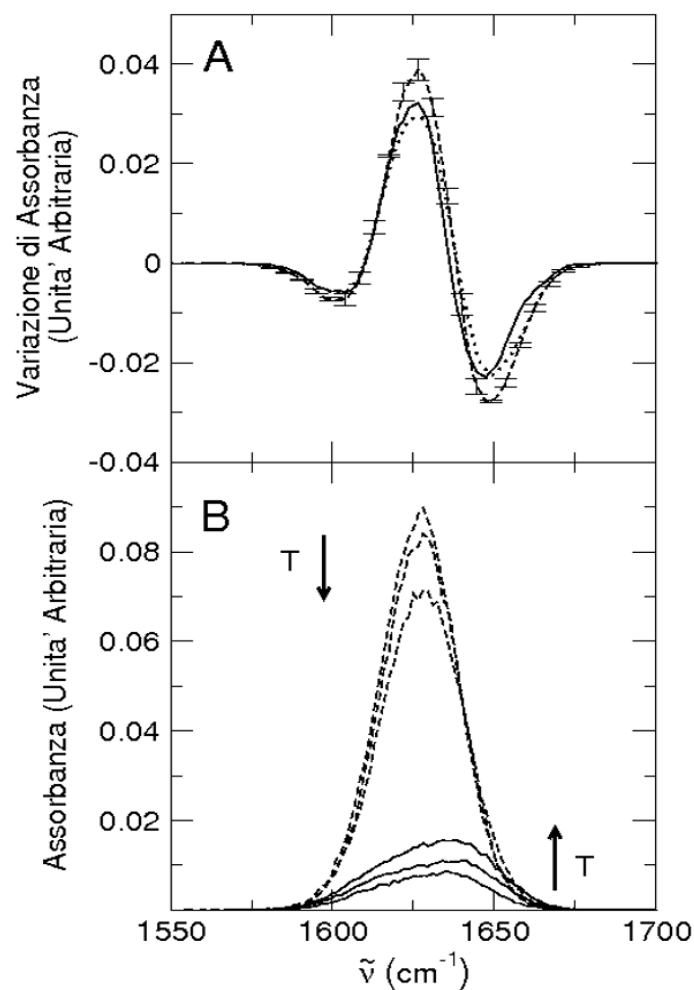


Figura 3.21: A: Spettri relativi al GS6, ottenuti sottraendo lo stato foldato da quello unfoldato a 280 K (linea solida), 310 K (linea tratteggiata) e 360K (linea punteggiata). Viene anche riportata l'incertezza in base all'errore standard valutato su tre traiettorie indipendenti per gli spettri a 310 K. - B: Contributo dello stato foldato (linea solida) e di quello unfoldato (linea tratteggiata) agli spettri IR a 280 K, 310 K e 360 K.

Per quanto riguarda il GS10, sono stati ricostruiti gli spettri IR in funzione della temperatura alle temperature di 280 K, 330, 360 K, 400 K (Figura 3.22 A). C'è un buon accordo tra gli spettri calcolati e quelli sperimentali, sia nell'andamento in funzione della temperatura (Figura 3.22 B ed i corrispondenti dati sperimentali [10]), sia nella massima ampiezza a mezza altezza ($\approx 40 \text{ cm}^{-1}$ nello spettro sperimentale e $\approx 44 \text{ cm}^{-1}$ in quello elaborato). Anche in questo caso, gli spettri elaborati sono stati traslati a frequenze minori di $\approx 87 \text{ cm}^{-1}$ per allineare la banda I ammidica con il massimo sperimentale.

E' stato esaminato il contributo di tutti i residui amminoacidici allo spettro, in modo da comprendere meglio la relazione tra caratteristiche strutturali e spettroscopiche. L'andamento complessivo è il risultato del contributo di entrambi i segnali, negativo-positivo e positivo-negativo; nel GS10 il trend negativo-positivo per lo spettro, ottenuto dalla differenza foldato/unfoldato, è dovuto per la maggior parte ai residui idrofobici (Leu e Val), come risulta evidente in Figura 3.23, dove sono riportati i contributi dei diversi residui allo spettro totale.

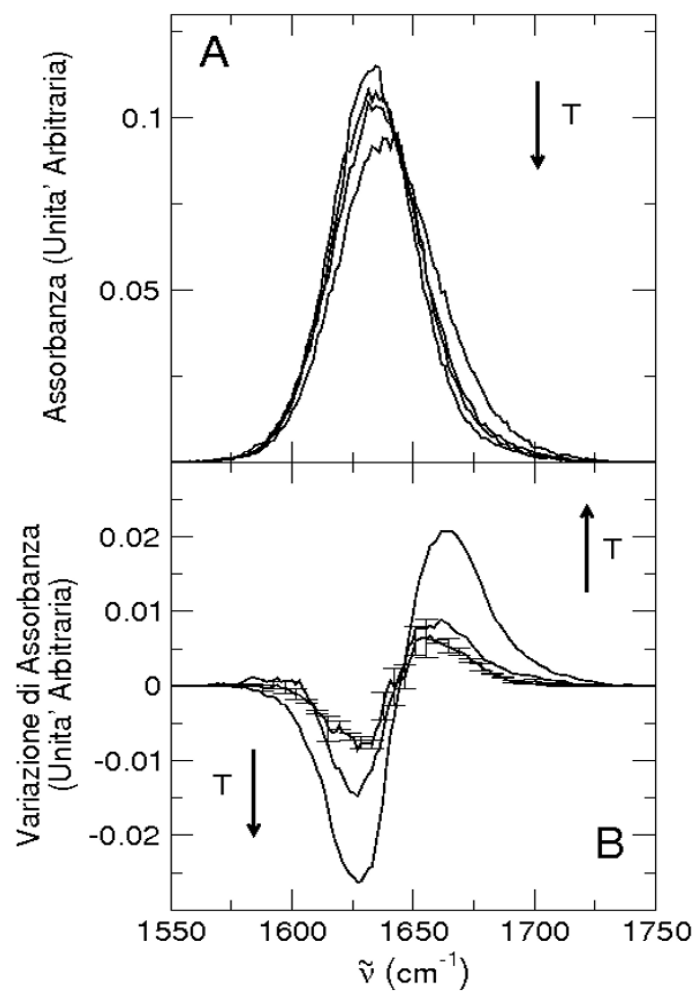


Figura 3.22: A: Spettri IR del GS10 nella regione del I modo ammidico in una soluzione di D_2O a 280 K, 330K, 360 K e 400 K. - B: Spettri ottenuti sottraendo lo spettro calcolato a 280 K da quelli a temperatura maggiore. Per il primo è stata anche riportata l'incertezza, corrispondente all'errore standard calcolato su tre distinte traiettorie eseguite a 330 K.

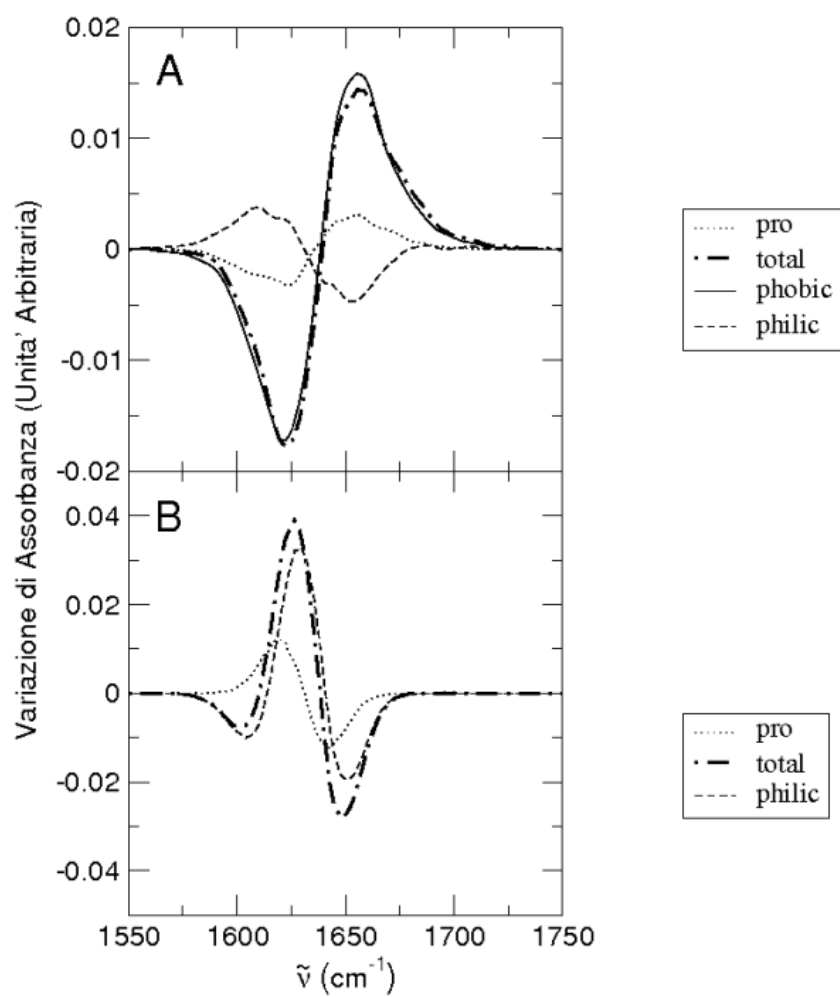


Figura 3.23: Contributo dei residui idrofilici (linea tratteggiata), idrofobici (linea solida) e della prolina (linea punteggiata) allo spettro unfoldato/foldato (linea più spessa) del GS10 (A) e del GS6 (B)

Nel caso del GS6 invece, non sono presenti residui idrofobici e lo spettro presenta un trend positivo-negativo: l'assenza di residui idrofobici risulta quindi alla base di questa inversione di tendenza. Inoltre, il comportamento termodinamico inusuale del GS6 [11], guidato dall'esposizione a gruppi carichi o polari del solvente durante l'unfolding che determina un riarrangiamento dello stesso, è determinato dalla presenza nel peptide delle sole catene laterali idrofiliche.

In tal modo viene spiegata l'apparente discrepanza tra dati ottenuti da calcoli e quelli sperimentali per il GS6, che risiede nella particolare termodinamica; ciò impedisce sia l'applicazione dell'ipotesi comunemente accettata *maggiore la temperatura, maggiore la popolazione dello stato unfoldato*, sia l'uso di regole empiriche, che assegnano un contributo spettrale ad un particolare elemento strutturale.

Capitolo 4

Conclusioni

L'analisi delle traiettorie relative al GS10 ha mostrato come il passaggio del peptide dallo stato foldato a quello unfoldato avvenga in un intervallo di temperatura molto ristretto, ciò a causa dei quattro legami idrogeno la cui formazione e rottura avviene in modo concertato. Non è stato perciò possibile effettuare una simulazione ad una determinata temperatura, che possedesse un numero di transizioni tale da ricavare proprietà termodinamiche e cinetiche con sufficiente accuratezza, così come svolto in un precedente lavoro per il GS6 [11]. Tuttavia uno studio più approfondito della traiettoria a 330 K ha permesso di ottenere una stima qualitativa della costante di folding risultata dello stesso ordine di grandezza di quella sperimentale [10], ciò a dimostrazione della affidabilità delle simulazioni.

Dalle analisi è inoltre emersa una differenza notevole nella cinetica dei due peptidi: infatti se il GS6 è perfettamente interpretabile con un meccanismo a due stati (foldato e unfoldato), per quanto riguarda il GS10 si è visto che tale meccanismo è molto

più complesso, e comprende, oltre allo stato foldato e quello unfoldato, diversi stati intermedi. Ciò è dovuto al maggior numero di residui che compongono il GS10 e che gli donano una maggiore libertà conformazionale.

L'analisi degli spettri calcolati in funzione della temperatura sono in buon accordo con quelli sperimentali, mostrando l'usuale andamento negativo-positivo, mentre quelli ottenuti dalla differenza foldato/unfoldato risultano differenti per i due peptidi. Per quanto riguarda il GS10, lo spettro foldato/unfoldato mostra lo stesso andamento di quello in funzione della temperatura, mostrando all'aumentare della temperatura una maggiore popolazione nello stato foldato; nel caso del GS6 abbiamo invece un andamento opposto, cioè un segnale positivo-negativo, per lo spettro della differenza foldato/unfoldato. Abbiamo attribuito questo fatto all'assenza di catene laterali idrofobiche che costituiscono il contributo principale allo spettro. L'inversione tra i due spettri indica che nell'intervallo di temperatura considerato la popolazione dello stato foldato aumenta all'aumentare della temperatura; ciò a conferma dei dati precedenti sulla termodinamica del GS6 [11]. Di conseguenza, per il peptide GS6, non è possibile considerare gli spettri IR funzione della temperatura come una rappresentazione del processo di folding.

L'uso di procedure PMM/MD per ricostruire gli spettri IR, ha permesso di identificare infine il diverso contributo allo spettro delle catene laterali, sottolineando come la diversa natura (idrofilica od idrofobica) definisca il segnale IR finale dei peptidi.

Bibliografia

- [1] Mitscher, L.A. *Foye's principi di chimica farmaceutica*, Piccin.
- [2] Gross, E. *The Peptides*, Academic Press, NY, 1983.
- [3] Fernandez-Lopez, S.; Kim H.S.; Choi, E.C.; Delgado, M.; Granja, J.R.; Khasanov, A.; Kraehenbuehl, K.; Long, G.; Weinberger, D.A.; Wilcoxon, K.M.; Ghadiri, M.R. *Antibacterial agents based on the cyclic D,L-alpha-peptide architecture*. Nature, 2001, 412, 452-455.
- [4] Jones, C.M.; Henry, E.R.; Hu, Y.; Chan, C.K.; Luck, S.D.; Bhuyan, A.; Roder, H.; Hofrichter, J.; Eaton, W.A. *Fast Events in Protein-Folding Initiated by Nanoseconds Laser Photolysis*. Proc. Natl. Acad. Sci. USA, 1993, 90, 11860-11864.
- [5] Gibbs, A.C.; Bjorndahl, T.C.; Hodges, R.S.; Wishart, D.S. *Probing the Structural Determinants of type II' β -Turn Formation in Peptides and Proteins*. J. Am. Chem. Soc., 2002, 124, 1203-1213.

- [6] Gibbs, A.C.; Kondejewski, L.H.; Gronwald, W.; Nip, A.M.; Hodges, R.S.; Sykes, B.D.; Wishart, D.S. *Unusual β -sheet periodicity in small cyclic peptides*. Nat. Struct. Biol., 1998, 5, 284-288.
- [7] Barh, A. *Infrared spectroscopy of proteins*. Biochem. Biophys. Acta, 2007, 1767, 1073-1101
- [8] Chen, E.; Goldbeck, R.A.; Kliger, D.S. *Nanosecond time-resolved spectroscopy of biomolecular processes*. Annu. Rev. Biophys. Biomol. Struct., 1997, 26, 327-355.
- [9] Hahn, S.; Ham, S.; Cho, M. *Simulation studies of Amide I IR absorption and two-dimensional IR spectra of beta-hairpins in liquid water*. J. Phys. Chem. B, 2005, 109, 11789-11801.
- [10] Maness, S.J.; Franzen, S.; Gibbs, A.C.; Causgrove, T.P.; Dyer, R.B. *Nanosecond Temperature Jump Relaxation Dynamics of Cyclic β -Hairpin Peptides*. Biophysical Journal, 2003, 84, 3874-3882.
- [11] Zanetti-Polzi, L.; Anselmi, M.; D'Alessandro M.; Amadei, A.; Di Nola, A. *Structural, Thermodynamic, and Kinetic Properties of Gramicidin Analogue GS6 Studied by Molecular Dynamics Simulations and Statistical Mechanics*. Biopolymers, 2009, 91, 1154-1160.
- [12] Brooks, B.R.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *CHARMM: a program for macromo-*

- lecular energy, minimization, and dynamics calculations. J. Comput. Chem.*, 1983, 4, 187.
- [13] Weiner, S.; Kollman, P.; Nguyen, D.; Case, D.; *An all atom force field for simulations of proteins and nucleic acids. J. Comput. Chem.*, 1986, 7, 230.
- [14] van Gunsteren, W.F.; Billeter, S.R.; Eising, A.A.; Hünenberger, P.H.; Krüger, P.; Mark, A.E.; Scott, W.R.P.; Tironi, I.G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Hochschulverlag AG an der ETH Zürich, Zürich, 1996.
- [15] Allen, M.P.; Tildesly, D.J. *Computer simulation of liquids*. Oxford University Press, Oxford, 1989.
- [16] Verlet, L. *Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys. Rev.*, 1967, 159, 98.
- [17] Verlet, L. *Computer experiments on classical fluids. II. Equilibrium correlation functions. Phys. Rev.*, 1968, 165, 201-223.
- [18] Beeman, D. *Some multistep methods for use in molecular for use in molecular dynamics calculations. J. Comput. Phys.*, 1976, 20, 130.

- [19] Gear, C. *Numerical Initial Value Problems in Ordinary Differential Equations*. Englewood Cliffs, NJ, USA, 1971.
- [20] McCammon, J.; Gelin, B; Karplus, M. *Dynamics of folded proteins*. Nature, 1977, 265, 585.
- [21] Schlick, T.; Barth, E.; Mandziuk, M. *BIOMOLECULAR DYNAMICS AT LONG TIMESTEPS: Bridging the Timescale Gap Between Simulation and Experimentation*. Annu. Rev. Biomol. Struct., 1997, 26, 181-222.
- [22] Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. *Numerical integration of the cartesian equations of motion in a system with constraints: molecular dynamics of n-alkanes*. J. Comp. Phys., 1977, 23, 327-341.
- [23] Amadei, A.; Chillemi, G.; Ceruso, M.A.; Grottesi, A.; Di Nola, A. *Molecular dynamics simulations with 0 rotational motions: Theoretical basis and statistical mechanical consistency*. J. Chem. Phys., 2000, 112, 9-23.
- [24] Bekker, H.; van den Berg, J.P.; Wassenaar, T.A. *A method to obtain a near-minimal-volume molecular simulation of a macromolecule, using periodic boundary conditions and rotational constraints*. J. Comput. Chem., 2004, 25, 1037-1046.

- [25] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L.G. *A smooth particle mesh Ewald method*. J. Chem. Phys., 1995, 103, 8577-8593.
- [26] Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; Di Nola, A.; Haak, J.R. *Molecular dynamics with coupling to an external bath*. J. Chem. Phys., 1984, 81, 3684-3690.
- [27] Brown, D.; Clarke, J. *Molecular dynamics simulations of polymer fiber microstructure*. J. Chem. Phys., 1986, 84, 2858.
- [28] Nosé, S. *Constant temperature molecular dynamics methods*. Prog. Theoret. Phys. Supplement, 1991, 103, 1-46.
- [29] Hoover, W. *Canonical dynamics: Equilibrium phase-space distributions*. Phys. Rev. A31, 1985, 1695-1697.
- [30] Martyna, G.J.; Klein, M.L.; Tuckerman, M. *Nosé Hoover chains: The canonical ensemble via continuous dynamics*. J. Chem. Phys., 1992, 97, 2635-2645.
- [31] Amadei, A.; Linssen, A.B.M.; Berendsen, H.J.C. *Essential Dynamics of Proteins*. Proteins: Struct. Funct. Genet., 1993, 17, 412-425.
- [32] García, A.E. *Large-amplitude nonlinear motions in proteins*. Phys. Rev. Lett., 1992, 68, 2696-2699.
- [33] Kitao, A.; Hirata, F.; Gō, N. *The effects of solvent on the conformation and the collective motions of protein: normal*

- mode analysis and molecular dynamics simulations of melittin in water and in vacuum.* J. Chem. Phys., 1991, 158, 447-472.
- [34] van Aalten, D.M.; Amadei, A.; Linssen, A.B.; Eijssink, V.G.; Vriend, G.; Berendsen, H.J.C. *The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water.* Proteins: Struct. Funct. Genet., 1995, 22, 45-54.
- [35] de Groot, B.L.; van Aalten, D.M.; Amadei, A.; Berendsen, H.J.C. *Domain motions in Bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data.* Proteins: Struct. Funct. Genet., 1998, 31, 116-127.
- [36] van Aalten, D.M.; Conn, D.A.; de Groot, B.L.; Berendsen, H.J.C.; Findlay, J.B.; Amadei, A. *Protein dynamics derived from clusters of crystal structures.* Biophys. J., 1997, 73, 2891-2896.
- [37] Abseher, R.; Horstink, L.; Hilbers, C.W.; Nilges M. *Essential spaces defined by NMR structure ensembles and molecular dynamics simulations show significant overlap.* Proteins: Struct. Funct. Genet., 1998, 31, 370-382.
- [38] de Groot, B.L.; van Aalten, D.M.; Scheek, R.M.; Amadei, A.; Vriend, G.; Berendsen, H.J.C. *Prediction of protein confor-*

- mational freedom from distance constraints*. Proteins: Struct. Funct. Genet., 1997, 29, 240-251.
- [39] Amadei, A.; D'Abramo, M.; Zazza, C.; Aschi, M. *Electronic properties of formaldehyde: a theoretical study*. Chem. Phys. Lett., 2003, 381, 187-193.
- [40] Amadei, A.; D'Alessandro, M.; D'Abramo, M.; Aschi, M. *Theoretical characterization of electronic states in interacting chemical systems*. J. Chem. Phys., 2009, 130(8), 084109.
- [41] Aschi, M.; Spezia, R.; Di Nola, A.; Amadei, A. *A first principles method to model perturbed electronic wave-functions: the effect of an external electric field*. Chem. Phys. Lett., 2001, 344, 374-380.
- [42] Spezia, R.; Aschi, M.; Di Nola, A.; Amadei, A. *Extension of the perturbed matrix method: application to a water molecule*. Chem. Phys. Lett., 2002, 365, 450-456.
- [43] Amadei, A.; Marinelli, F.; D'Abramo, M.; D'Alessandro, M.; Anselmi, M.; Di Nola, A.; Aschi, M. *Theoretical modeling of vibroelectronic quantum states in complex molecular system: Solvated carbon monoxide: a test case*. J. Chem. Phys., 2005, 122, 124506.
- [44] Anselmi, M.; Aschi, M.; Di Nola, A.; Amadei, A. *Theoretical Characterization of Carbon Monoxide Vibrational Spectrum*

- in Sperm Whale Myoglobin Distal Pocket*. Biophys. J., 2007, 92.
- [45] Aschi, M.; Amadei, A. *Theoretical-computational modeling of quantum vibrational states of polyatomic systems in complex enviroment*. IJLSS In Press, 2009.
 - [46] Gao, J.; Truhlar, D.G. *Quantum mechanical methods for enzyme kinetics*. Ann. Rev. Phys. Chem., 2002, 53, 467-505.
 - [47] Senn, H.M.; Thiel, W. *QM/MM studies of enzymes*. Curr. Opin. Chem. Biol., 2007, 11, 182-187.
 - [48] Vreven, T.; Morokuma, K. *Hybrid methods: ONIOMQM:MM and QM/MM*. Ann. Rep. Comput. Chem., 2006, 2, 35-52.
 - [49] Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. *LINCS: A linear constraint solver for molecular simulations*. J. Comput. Chem., 1997, 18, 1463-1472.
 - [50] Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. *GROMACS: A message-passing parallel molecular dynamics implementation*. Comput. Phys. Commun., 1995, 91, 43-56.
 - [51] Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; Hermans, J. *Interaction models for water in relation to protein hydration in Intermolecular Forces*. D. Reidel Publishing Company, Dordrecht, The Netherlands, 1981.

- [52] Darden T.; York, D.; Pedersen, L. *Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems*. J. Chem. Phys., 1993, 98, 10089-10092.
- [53] Brown, D.; Clarke, J.H.R. *A comparison of constant energy, constant temperature and constant pressure ensembles in molecular dynamics simulations of atomic liquids*. Mol. Phys., 1984, 51, 1243-1252.
- [54] Ataka, S.; Takeuchi, H.; Tasumi, M. *Infrared studies of the less stable cis form of N-methylformamide and N-methylacetamide in low-temperature nitrogen matrices and vibrational analysis of the trans and cis forms of these molecules*. J. Mol. Struct., 1984, 113, 147-160.
- [55] Kubelka, J.; Keiderling, T.A. *Ab Initio Calculation of Amide Carbonyl Stretch Vibrational Frequencies in Solution with Modified Basis Sets. 1. N-Methyl Acetamide*. J. Phys. Chem. A, 2001, 105, 10922-10928.
- [56] Mirkin, N.G.; Krimm, S. *Ab initio vibrational analysis of hydrogen-bonded trans- and cis-N-methylacetamide*. J. Am. Chem. Soc., 1991, 113, 9742-9747.
- [57] Gregurick, S.K.; Chaban, G.M.; Gerber, R.B. *Ab Initio and Improved Empirical Potentials for the Calculation of the Anharmonic Vibrational States and Intramolecular Mode Cou-*

- pling of N-Methylacetamide.*, J. Phys. Chem. A, 2001, 106, 8696-8707.
- [58] Song, S.; Asher, S.A.; Krimm, S.; Shaw, K.D. *Ultraviolet resonance Raman studies of trans and cis peptides: photochemical consequences of the twisted π^* excited state.* J. Am. Chem. Soc., 1991, 113, 1155-1163.
- [59] Kwac, K.; Cho, M. *Molecular dynamics simulation study of N-methylacetamide in water. I. Amide I mode frequency fluctuation.* J. Chem. Phys., 2003, 119, 2247-2255.
- [60] Amadei, A.; Apol, M.E.F.; Berendsen, H.J.C. *Extensions of the quasi-Gaussian entropy theory.* J. Chem. Phys., 1997, 106, 1893-1912.
- [61] Amadei, A.; Apol, M.E.F.; Berendsen, H.J.C. *On the use of the quasi-Gaussian entropy theory in noncanonical ensembles. I. Prediction of temperature dependence of thermodynamic properties.* J. Chem. Phys., 1998, 109, 3004-3016.
- [62] Amadei, A.; Apol, M.E.F.; Di Nola, A.; Berendsen, H.J.C. *The quasi-Gaussian entropy theory: Free energy calculations based on the potential energy distribution function.* J. Chem. Phys., 1996, 104, 1560-1574.
- [63] Apol, M.E.F.; Amadei, A.; Berendsen, H.J.C. *On the use of the quasi-Gaussian entropy theory in noncanonical ensem-*

- bles. II. Prediction of density dependence of thermodynamic properties.* J. Chem. Phys., 1998, 109, 3017-3027.
- [64] Hayashi, T.; Mukamel, S. *Vibrational-exciton couplings for the amide I, II, III and A. Modes of peptides.* J. Phys. Chem. B, 2007, 111, 11032-11046.