

ANNO ACCADEMICO 1999-2000

Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di Laurea in Fisica

Tesi di Laurea

# **Dinamica e processi di attivazione in modelli di eteropolimeri**

Meccanismi rilevanti per lo studio del ripiegamento delle proteine

Candidata:

Annalisa Tiberio

Relatore:

dott. Alessandro Torcini

Correlatore:

prof. Roberto Livi

10 ottobre 2000

# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Il problema del ripiegamento delle proteine</b>	<b>7</b>
1.1 Cosa è una proteina . . . . .	8
1.2 Attività funzionali delle proteine . . . . .	12
1.3 Struttura primaria, secondaria e terziaria . . . . .	14
1.4 Modello biochimico . . . . .	16
1.5 Il problema della predizione della forma nativa di una proteina . . . . .	19
1.6 Tempi di ripiegamento . . . . .	21
1.7 Proprietà cinetiche del processo di ripiegamento . . . . .	23
1.8 Caratteristiche del processo di denaturazione . . . . .	25
1.9 Cenni sulla teoria dei polimeri lineari . . . . .	28
1.10 Caratterizzazione degli stati denaturati . . . . .	32
<b>2 Modelli fisici e metodi d'indagine</b>	<b>35</b>
2.1 Modelli semplici esatti . . . . .	36
2.1.1 Il modello HP . . . . .	36
2.1.2 Il modello di $G\bar{0}$ . . . . .	38

2.2	Un modello fuori reticolo . . . . .	41
2.3	Il paesaggio energetico . . . . .	44
<b>3</b>	<b>Dinamica di un modello 2D fuori reticolo</b>	<b>48</b>
3.1	Il modello . . . . .	49
3.2	Dettagli sulle simulazioni . . . . .	51
3.3	Le sequenze analizzate . . . . .	54
3.4	Le configurazioni native . . . . .	55
3.5	Caratterizzazione delle valli native . . . . .	56
3.6	Determinazione della temperatura $T_f$ . . . . .	58
3.7	Proprietà di equilibrio . . . . .	60
3.8	Comportamento dinamico alla temperatura $T_f$ . . . . .	64
<b>4</b>	<b>Descrizione a grana grossa della dinamica di ripiegamento</b>	<b>67</b>
4.1	Identificazione dei minimi “vicini” alla configurazione nativa . . . . .	71
4.2	Ricerca dei punti sella . . . . .	74
4.3	Le barriere di energia potenziale . . . . .	78
4.4	Studio della dinamica del processo di salto . . . . .	81
4.5	Probabilità di fuga da uno stato metastabile . . . . .	83
4.5.1	Caso di una particella . . . . .	85
4.5.2	Caso di più particelle . . . . .	89
4.6	Stima delle probabilità di salto tramite la formula teorica . . . . .	92
4.7	Stima numerica delle probabilità di salto . . . . .	96
4.8	Costruzione dei grafi . . . . .	99

<b>Conclusioni</b>	<b>103</b>
<b>A Stima dei tempi di scala reali</b>	<b>108</b>
<b>B Algoritmo per l'identificazione dei cammini principali</b>	<b>109</b>
<b>C Probabilità di fuga nel caso unidimensionale</b>	<b>113</b>
<b>D Probabilità di fuga nel caso multidimensionale</b>	<b>115</b>

# Introduzione

Il presente lavoro di tesi rientra nell'ambito di un'ampia attività di ricerca che ha visto negli ultimi anni molti fisici orientarsi verso problematiche che un tempo erano da ritenersi dominio dei biologi o dei biochimici. A partire dalla proficua collaborazione di Crick, Watson e Wilkins, che consentì l'identificazione della struttura elicoidale del DNA e per cui furono insigniti del premio Nobel nel 1962, sono stati fatti molti passi avanti. In particolare, queste attività di frontiera hanno consentito il trasferimento e l'applicazione di tecniche e metodologie propriamente fisiche a vari campi delle scienze che si occupano degli organismi viventi.

Uno dei temi più studiati, data la sua enorme importanza per capire i processi vitali, è stato fin dagli anni '50 quello riguardante la funzionalità delle proteine. Le proteine sono eteropolimeri costituiti di un numero variabile (da qualche decina a qualche centinaio) di aminoacidi disposti in catena. Dato che il numero di aminoacidi distinti è limitato a venti, e che il numero di proteine sinora individuate è di circa 50'000 è chiaro che è la disposizione degli aminoacidi lungo la catena (cioè la *struttura primaria*) a determinare la conformazione strutturale della proteina (detta *struttura nativa*). La funzionalità di una proteina in condizioni fisiologiche è espressione diretta della sua struttura nativa.

Il problema della determinazione della struttura della proteina nella sua forma nativa a partire dalla sequenza degli aminoacidi che la compongono è, nonostante gli enormi sforzi profusi in questa linea di ricerca, un problema tuttora irrisolto e di grande ed attuale interesse. In particolare non tutte le sequenze di aminoacidi ripiegano in uno stato nativo attivo, ed è di primario interesse capire come mai solo alcune sequenze (dette “buone” sequenze) ripieghino in modo tale da dare luogo ad una proteina, mentre altre (dette “cattive” sequenze) restino in forma di filamento non ripiegato. Un ulteriore problema aperto è legato ai tempi necessari, sotto condizioni fisiologiche, al ripiegamento di una proteina nella sua struttura nativa a partire da una struttura non ripiegata. Tali tempi di ripiegamento sono dell'ordine al più di qualche secondo, e quindi assai rapidi se si considera l'astronomico numero di configurazioni<sup>1</sup> in cui la proteina potrebbe ripiegarsi se tale processo fosse determinato puramente da meccanismi probabilistici. Cioè se la proteina raggiungesse il suo stato nativo dopo aver sistematicamente provato a ripiegarsi in tutte le possibili conformazioni.

La possibilità di disegnare nuove proteine con funzionalità desiderata assegnando semplicemente la sequenza di aminoacidi avrebbe enormi ricadute in vari campi dalla catalisi, ai biosensori, alla progettazione di nuove medicine e di nuovi ormoni. D'altra parte la conoscenza dei dettagli del processo dinamico di ripiegamento potrebbe aiutare enormemente nella progettazione e realizzazione di proteine e nel chiarire quali meccanismi portino alla creazione della struttura nativa tramite l'effetto cooperativo di caratteristiche proprie della catena e vincoli ambientali di natura

---

<sup>1</sup>Tale numero è stato stimato da Levinthal [1] essere dell'ordine di  $10^N$ , ove  $N$  è il numero degli aminoacidi costituenti la proteina.

chimica (il solvente) e termodinamica (la temperatura).

Questi temi sono stati affrontati negli ultimi dieci anni dai fisici come problemi di meccanica statistica d'equilibrio. In particolare, le proteine sono state tipicamente schematizzate come catene unidimensionali costituite da aminoacidi di soli due tipi (polari o idrofobi) che tendono a respingersi o ad attrarsi fra di loro tramite interazioni di tipo classico, in modo da riprodurre in maniera efficace le interazioni col solvente degli aminoacidi reali. Per semplificare ulteriormente il problema viene poi assunto che gli aminoacidi della proteina modello possano assumere solo posizioni fisse su un reticolo bi- o tridimensionale. Per questi modelli assai semplificati è possibile fare una trattazione statistica completa, calcolando ad esempio l'energia libera del sistema e la distribuzione dei livelli energetici del modello (corrispondenti a configurazioni d'equilibrio stabile o metastabile della proteina). Il dibattito scientifico è stato essenzialmente incentrato sulla determinazione di criteri per l'individuazione delle "buone" sequenze. Questi criteri sono stati essenzialmente basati sulla distribuzione dei livelli energetici associati ai minimi metastabili del modello (ad esempio una separazione grande fra il livello più basso e quello subito superiore è stata a lungo considerata propria di una "buona" sequenza [2]). Per determinare questi livelli energetici in modelli appena un pò più complessi è stato necessario affidarsi a metodi numerici, tipicamente ricorrendo a tecniche Monte Carlo di maggiore o minore raffinatezza.

Tuttavia, il processo di ripiegamento di una proteina è un processo tipicamente dinamico, che, partendo da una catena di aminoacidi, termina con una struttura ripiegata globularmente e biologicamente attiva. Quindi non si può sperare di capire

le caratteristiche fondamentali di questo processo limitandoci ad una sua descrizione in termini statistici. Una descrizione dinamica dell'intero processo di ripiegamento non può tuttavia essere affrontata con un approccio *ab-initio*. Infatti, se si considerano modelli abbastanza realistici, che tengano conto di tutte le interazioni presenti in una proteina (a livello semiclassico), si scopre che i tempi associati alla dinamica microscopica sono tipicamente dell'ordine di  $10^{-11}$  -  $10^{-12}$  sec (se si considera l'interazione col solvente in modo efficace, altrimenti si scende a tempi ancora più piccoli), mentre il tempo di ripiegamento è dell'ordine di 0.1 - 1 sec. Quindi una simulazione dinamica che volesse tenere conto con accuratezza di tutte le interazioni richiederebbe di simulare qualche centinaio di atomi per  $10^{11}$ - $10^{12}$  passi di integrazione. Questo è evidentemente un limite irraggiungibile con i mezzi di calcolo oggi disponibili (al più con potenti calcolatori paralleli si arriva a qualche nanosecondo di integrazione di questi modelli *ab-initio*).

È ragionevole quindi ricorrere a modelli efficaci semplificati che consentano di osservare la transizione da uno stato non ripiegato alla struttura nativa o che, quantomeno, permettano di individuare dei segnali precursori di tale ripiegamento tramite l'uso di indicatori appropriati.

Recentemente sono stati messi a punto una serie di semplici modelli meccanici classici in due e tre dimensioni, ove la proteina viene schematizzata come una catena unidimensionale in cui gli aminoacidi sono sostituiti da elementi localizzati con carattere idrofobico o polare che interagiscono tra loro tramite termini di forza efficace sia a corto che a lungo raggio. In questi modelli gli aminoacidi possono muoversi liberamente in un certo volume e le loro posizioni non sono più limitate su un reticolo.



In particolare, tali modelli si basano sull'ipotesi che sia possibile ottenere una descrizione "sufficientemente accurata" della dinamica partendo da una rappresentazione "mesoscopica" del sistema in esame. Questa consiste nel dare una rappresentazione a grana grossa delle forze d'interazione prescindendo dai dettagli "molecolari".

In questo lavoro di tesi abbiamo preso in esame un modello in due dimensioni che contiene un'interazione a primi vicini che ha il compito di simulare un legame approssimativamente rigido (e che per semplicità abbiamo scelto armonica), un'energia di orientazione associata all'angolo che si forma fra monomeri adiacenti ed un'interazione a lungo raggio (tipo Lennard-Jones), che dovrebbe riprodurre in prima approssimazione l'effetto efficace del solvente.

Ci siamo poi limitati a considerare solo due tipi di aminoacidi: polari o idrofobi. Quindi la struttura primaria nel nostro caso può essere rappresentata in termini di una semplice sequenza binaria.

Questo modello riproduce alcune caratteristiche proprie delle proteine reali, ad esempio solo alcune sequenze si ripiegano sempre in un'unica configurazione stabile (la struttura nativa) in tempi "ragionevoli", queste le indicheremo come sequenze "buone". Le altre sequenze saranno definite invece "cattive". Uno degli scopi di questa tesi è stato quello di studiare la dinamica di ripiegamento delle varie sequenze con studi di dinamica molecolare a temperatura costante. Con l'obiettivo di individuare quali siano i comportamenti dinamici collettivi che distinguono sequenze "buone" da sequenze "cattive". Dall'analisi di questo semplice modello mesoscopico si spera di evincere degli indicatori dinamici che poi possano essere applicati utilmente per lo studio di modelli più complessi e sperabilmente per l'analisi di dati sperimentali.

L'altro obiettivo che ci siamo posti è stato quello di mettere a punto delle metodologie per poter descrivere il processo di ripiegamento di una proteina quale un processo di attivazione termica, che tramite transizioni da un minimo metastabile ad un altro porti una proteina da uno stato inizialmente non ripiegato verso la sua struttura nativa.

I contributi originali di questo lavoro di tesi sono descritti in dettaglio nei capitoli 3 e 4 e sono riassunti nelle conclusioni alla tesi. Il primo capitolo è invece dedicato ad una introduzione esaustiva del problema oggetto di questa tesi: il ripiegamento delle proteine nella loro struttura nativa, cioè funzionalmente attiva. Mentre nel secondo capitolo vengono riportate e commentate le schematizzazioni adottate a livello fisico per descrivere il processo di ripiegamento tramite modelli mesoscopici.

Per maggiore chiarezza ogni capitolo è preceduto da un breve preambolo che riassume la struttura del capitolo stesso e i temi in esso trattati.

# Capitolo 1

## Il problema del ripiegamento delle proteine

In questo capitolo ci proponiamo di fornire un quadro essenziale, ma sostanzialmente esauriente, del problema oggetto di questa tesi: il ripiegamento delle proteine nella loro struttura nativa, cioè funzionalmente attiva. Il punto di attacco è necessariamente la definizione di proteina e delle sue proprietà funzionali, che risultano espressione diretta della sua conformazione strutturale. In particolare ci soffermeremo sulla descrizione della natura e dell'ordine di grandezza delle forze, sia interne che dovute alla presenza del solvente, che determinano tale conformazione. Ci addentreremo poi più specificamente nella descrizione di ciò che è noto riguardo al processo di ripiegamento delle proteine a partire da uno stato aperto, inerte, fino a raggiungere la configurazione attiva. Discuteremo anche come il passaggio inverso, dalla struttura attiva a quella aperta, risulti a tutti gli effetti un processo reversibile. Questo mostra in definitiva la specificità del processo di ripiegamento di una

proteina, che risulta ben diverso da un semplice cammino aleatorio in uno spazio di configurazioni. Infine, attraverso brevi cenni alla teoria dei polimeri introdurremo alcune delle proprietà generali dei modelli più semplici di solito utilizzati per schematizzare una proteina.

## 1.1 Cosa è una proteina

Un aminoacido è costituito da un gruppo aminico ( $\text{NH}_2$ ), un gruppo carbossilico ( $\text{C}'\text{OOH}$ ), un atomo di idrogeno (H) e una catena laterale (R), legati ad un atomo di carbonio centrale ( $C_\alpha$ )<sup>1</sup>(vedi figura (1.1a)). In natura esistono 20 tipi di aminoacidi e ciò che li distingue l'uno dall'altro è la composizione chimica della catena laterale [3, 4, 5].

Le proteine sono macromolecole costituite da particolari combinazioni di aminoacidi uniti in catena attraverso legami peptidici (C'- N). Durante la loro biosintesi il gruppo carbossilico del primo aminoacido reagisce con il gruppo aminico del successivo formando il legame peptidico e rilasciando una molecola d'acqua; la ripetizione di questo processo dà luogo alla formazione della catena polipeptidica, cioè della proteina.

La porzione di aminoacido che rimane a far parte della catena è detta residuo; in un residuo si individuano la catena laterale e i restanti atomi formanti la catena principale, identica per ciascun residuo. L'unità ripetitiva elementare lungo la catena principale è  $\text{NH}-C_\alpha\text{H}-\text{C}'\text{O}$  ed è costituita da ciò che rimane della porzione comune

---

<sup>1</sup>Per distinguerlo dall'atomo di carbonio facente parte del gruppo carbossilico (indicato con C'), si indica con  $C_\alpha$  l'atomo di carbonio centrale dell'aminoacido e con  $C_\beta, C_\gamma$  ecc. gli eventuali atomi di carbonio facenti parte della catena laterale R.

di un aminoacido dopo la formazione del legame peptidico (vedi figura (1.1b)).

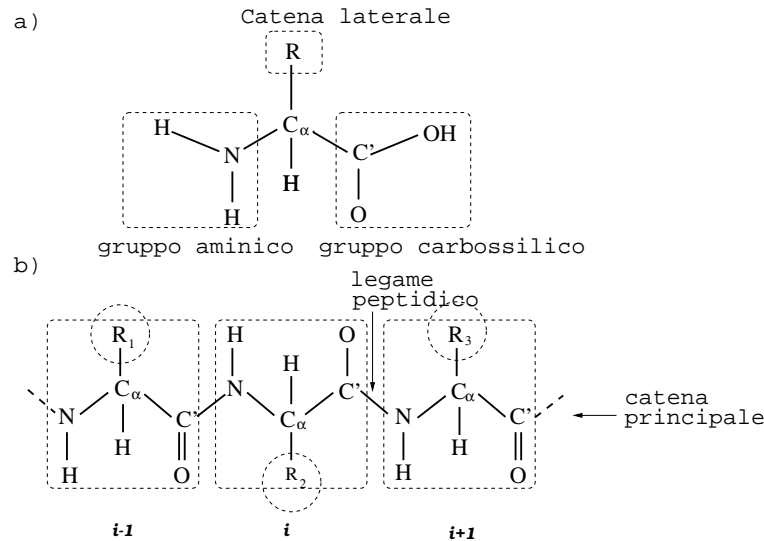


Figura 1.1. a) Rappresentazione di un aminoacido. b) Porzione della catena peptidica.

Si può adottare un altro criterio per scomporre la catena principale in unità ripetitive, che risulta più appropriato per descrivere le proprietà strutturali delle proteine. Ciascuna unità va da un atomo  $C_{\alpha}$  a quello successivo (la distanza media tra questi è circa  $4 \text{ \AA}$ ) e, in questo modo, risulta che tutti gli atomi di un'unità occupano posizioni rigide su un piano, con lunghezze e angoli di legame che sono approssimativamente gli stessi. Questi gruppi rigidi sono uniti attraverso i legami covalenti degli atomi  $C_{\alpha}$ ; in particolare i gradi di libertà per ciascun gruppo si riducono alle rotazioni attorno ai legami  $C_{\alpha} - C'$  e  $N - C_{\alpha}$ , i cui angoli associati vengono detti rispettivamente  $\psi$  e  $\phi$  (vedi fig. (1.2)). La conformazione dell'intera catena principale risulta quindi completamente determinata quando siano dati i valori degli angoli  $\psi_i$  e  $\phi_i$  associati a ciascun residuo  $i$  della proteina; l'indice  $i$  numera ogni carbonio  $C_{\alpha}$  lungo la catena polipeptidica e quindi anche il corrispondente

residuo.

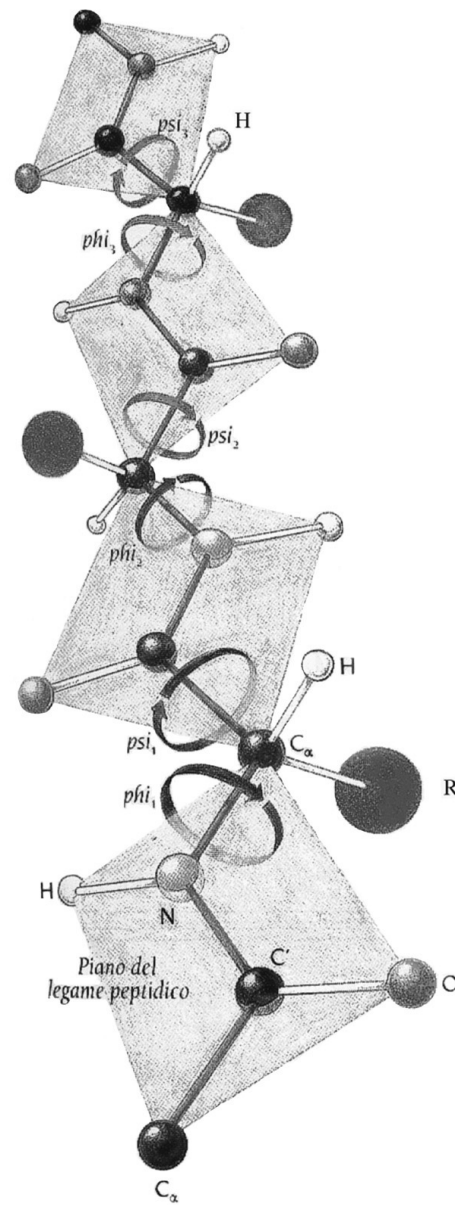


Figura 1.2. Rappresentazione della struttura di una porzione di una catena proteica (la figura è stata presa da [4]).

Generalmente gli aminoacidi vengono distinti in tre classi a seconda della natura chimica della catena laterale. La prima classe comprende aminoacidi con catene

laterali idrofobiche, la seconda classe è formata dagli aminoacidi le cui catene laterali includono gruppi con carica elettrica, gli aminoacidi con catene laterali polari formano infine la terza classe. Considerando tutte le proteine finora identificate, la frequenza media di tali classi è rispettivamente del 48, 22 e 30% [3].

Il numero di aminoacidi di una proteina può variare da poche decine a molte centinaia. Essendo ogni singola proteina costituita da un numero relativamente alto di atomi la si può assimilare ad un sistema mesoscopico e quindi una sua descrizione in termini termodinamici è ragionevole. In particolare, si possono definire potenziali termodinamici (entropia, energia libera, ecc.) che descrivono ogni singola macromolecola. Per chiarezza va fatto notare che negli esperimenti non vengono analizzate le singole proteine, ma soluzioni contenenti un alto numero di queste. Tuttavia questo tipo di analisi fornisce informazioni sulle proprietà della singola proteina, in quanto le interazioni fra i vari eteropolimeri in soluzione sono trascurabili rispetto alle interazioni interne a una proteina.

Inoltre sotto opportune condizioni fisiologiche (soluzione acquosa, 37°C, pH 7, pressione atmosferica), le interazioni fra le catene laterali dello stesso polipeptide prevalgono rispetto a quelle di ciascuna catena laterale con il solvente. In queste condizioni, il polipeptide si ripiega su se stesso assumendo una forma ben definita e termodinamicamente stabile e solo se questo avviene si può parlare effettivamente di proteina. Infatti è proprio questa struttura ripiegata, spesso globulare, a garantire la funzionalità biologica della proteina.

## 1.2 Attività funzionali delle proteine

Le proteine sono presenti solo negli organismi viventi e ne rendono possibile la grande varietà di aspetti morfologici e di adattamenti funzionali [5]. In pratica tutte le proprietà che caratterizzano i processi vitali sono influenzate dalle proteine.

Nel DNA (acido deossiribonucleico) sono codificate le informazioni genetiche (quindi anche la sintesi proteica) e la loro espressione dipende quasi esclusivamente dalle proteine stesse.

Le proteine immagazzinano e trasportano una grande varietà di particelle dai singoli elettroni fino alle macromolecole. Ad esempio esse guidano il flusso di elettroni nel fondamentale processo della fotosintesi; sotto forma di ormoni trasmettono le informazioni tra cellule specifiche e organi negli organismi complessi; come enzimi catalizzano le reazioni chimiche necessarie alla vita; alcune proteine controllano inoltre il passaggio delle molecole attraverso le membrane; gli anticorpi stessi sono proteine. Sono anche costituenti fondamentali del tessuto muscolare e di altri sistemi di conversione da energia chimica in energia meccanica; esse sono inoltre necessarie per il tatto, l'udito e gli altri sensi. Ci sono infine proteine semplicemente strutturali che determinano l'architettura filamentosa dei materiali di cui sono composti i capelli, le unghie, i tendini e le ossa degli animali.

Data una così ampia gamma di funzioni, non ci si attenderebbe che le proteine fossero una classe di molecole relativamente omogenea, che si forma a partire dallo stesso tipo di eteropolimero lineare costruito con varie combinazioni di soli 20 tipi di aminoacidi. Le proteine infatti differiscono tra loro solo per il numero di aminoacidi che le compongono e per la sequenza in cui questi sono disposti lungo la catena



polipeptidica. La loro polifunzionalità è dovuta in parte alla differente composizione chimica dei singoli aminoacidi, ma primariamente alla varietà delle strutture tridimensionali in cui queste catene lineari si dispongono. Le sorprendenti proprietà funzionali delle proteine possono essere comprese solo mettendole in relazione con la loro struttura tridimensionale.

La sequenza di aminoacidi di una data proteina può essere determinata attraverso metodi biochimici, sia osservando direttamente la proteina, sia derivandola dalla sequenza nucleotidica del corrispondente gene in cui è codificata; la banca dati di tutte le sequenze sinora identificate ne contiene già 50'000, numero che va raddoppiando di anno in anno.

Determinare la struttura delle proteine è invece un processo più lungo e complicato e attualmente si conoscono, a livello atomico, le strutture di 10'500 proteine, il più delle quali determinate attraverso la cristallografia a raggi X. Questa tecnica fornisce le informazioni strutturali a una risoluzione abbastanza elevata: la posizione degli atomi è determinata con un errore minore di 2 Å . Anche la risonanza magnetica nucleare (NMR) è un utile mezzo d'indagine per la determinazione delle strutture delle proteine. Essa risulta fornire una risoluzione leggermente peggiore ma ha il vantaggio di poter essere applicata alle proteine in soluzione senza doverle precedentemente cristallizzare.

Non è difficile immaginare che il problema di capire attraverso quali meccanismi queste particolari sequenze si ripieghino e quindi di poter prevedere quale particolare forma assumano sia al centro dell'interesse e dei dibattiti in molti ambiti scientifici.

## 1.3 Struttura primaria, secondaria e terziaria

Nel 1952 il biochimico danese Kai Linderström-Lang [4] coniò il termine struttura primaria, secondaria e terziaria per sottolineare la gerarchia strutturale esistente nelle proteine.

La struttura primaria è semplicemente la sequenza ordinata degli aminoacidi lungo la catena polipeptidica. Per convenzione i residui sono numerati a partire dall'estremo che rimane con il gruppo aminico non impegnato nel legame (N-terminale), procedendo fino all'estremo opposto (C-terminale).

Prima di assumere la forma tridimensionale caratteristica (struttura terziaria) regioni diverse della sequenza formano sottostrutture locali regolari, che identificano le strutture secondarie: le eliche  $\alpha$  e i foglietti  $\beta$ .

Le eliche  $\alpha$  si formano nelle proteine quando gli atomi della catena principale di più residui consecutivi si dispongono a formare un'elica, che in media è composta da 3.6 residui per spira. La struttura elicoidale è stabilizzata da legami idrogeno, che si formano tra il gruppo C'O di un residuo  $i$  e il gruppo NH del residuo  $i+4$ . L'aumento medio di lunghezza dell'elica per residuo lungo il suo asse maggiore è di circa 1.5 Å e la lunghezza media di un'elica è circa 15 Å [5]. Le catene laterali si estendono verso l'esterno dell'elica  $\alpha$  e non interagiscono con la sua struttura.

Il secondo elemento strutturale presente nelle proteine globulari è il foglietto  $\beta$ ; questo motivo strutturale risulta dalla combinazione di più regioni (filamenti  $\beta$ ) della catena polipeptidica, che in genere sono composte da 5-10 residui ed assumono una conformazione quasi completamente distesa. Nella struttura  $\beta$  pieghettata questi filamenti si trovano allineati, giacendo l'uno accanto all'altro in modo da formare

legami idrogeno tra i gruppi C'O di un filamento  $\beta$  e i gruppi NH di un filamento adiacente. Le strutture  $\beta$ , costituite da numerosi di questi filamenti, sono "pieghettate", con gli atomi  $C_\alpha$  e le relative catene laterali che si alternano sopra e sotto il piano del foglietto. Se tutti i filamenti sono allineati nella stessa direzione biochimica, cioè dal gruppo aminico a quello carbossilico ( $N \rightarrow C$ ), il foglietto è detto parallelo; se invece gli aminoacidi presenti in filamenti adiacenti hanno direzioni opposte ( $N \rightarrow C$  e  $C \rightarrow N$ ), il foglietto è detto antiparallelo.

L'interno di una molecola proteica contiene soprattutto catene laterali idrofobiche, che però sono legate alle relative catene principali altamente polari e quindi propense ad esporsi all'interazione con il solvente acquoso. La disposizione in strutture secondarie permette, attraverso i legami idrogeno, di neutralizzare la polarità delle catene principali e di "impacchettare" all'interno della struttura proteica i residui idrofobici.

La struttura terziaria delle proteine è costruita attraverso combinazioni diverse di elementi di struttura secondaria quali eliche  $\alpha$  e foglietti  $\beta$ , disposti su una struttura tridimensionale in scala maggiore. Questi elementi costituiscono le regioni del nucleo interno della molecola (*core*) e sono collegati da regioni a forma di ansa (*loop*) di lunghezza variabile e forma irregolare, presenti sulla superficie della proteina. Questi motivi si combinano a formare strutture globulari compatte, dette domini. Si definisce dominio una catena polipeptidica o parte di essa che può ripiegarsi in maniera indipendente dal resto della molecola e formare una struttura terziaria stabile. Si parla di struttura terziaria per indicare sia il modo in cui i motivi strutturali sono disposti nella configurazione tridimensionale, sia le modalità di ripiegamento

dell'intera catena a formare uno o più domini.

Infine alcune proteine sono costituite da più catene polipeptidiche identiche (molecole multimeriche) che si associano in specifiche strutture quaternarie. Tali subunità possono agire indipendentemente l'una dall'altra, oppure presentare aspetti funzionali dipendenti dall'intera struttura cooperativa.

In questa tesi ci limiteremo a considerare proteine globulari monomeriche, cioè composte da una catena con un solo dominio.

## 1.4 Modello biochimico

In linea di principio per valutare l'energia di una data struttura proteica vanno prese in considerazione tutte le forze non covalenti che intervengono tra gli atomi costituenti e vanno incluse inoltre le interazioni tra questi atomi e le molecole del solvente nel quale la proteina è immersa. Sebbene le forze che determinano il legame peptidico (covalente) siano di tre ordini di grandezza più grandi di quelle non covalenti, il contributo di energia ad esse associato (dell'ordine di  $2.5 \text{ eV}$ ) può essere trascurato, in quanto risulta praticamente costante per ciascuna conformazione che può assumere la proteina e quindi non influisce nel calcolo della differenza di energia tra conformazioni diverse. Le interazioni non covalenti fondamentali sono i legami idrogeno, le forze elettrostatiche e quelle di van der Waals. Esse sono eccitate termicamente a temperatura ambiente e sono responsabili del ripiegamento e delle proprietà termodinamiche della proteina. Queste forze differiscono tra loro per intensità e specificità oltre a essere influenzate in modo diverso dalla presenza del solvente [6].

1. **Legame idrogeno.** Si tratta di un'interazione dipolo-dipolo e si instaura quando due atomi elettronegativi competono per essere legati allo stesso atomo di idrogeno. L'atomo a cui l'idrogeno è legato più strettamente è detto donatore mentre l'altro è detto accettore. È un legame altamente direzionale e risulta più forte quando i tre atomi sono collineari. L'energia di legame associata è dell'ordine di 0.1-0.3 eV e le lunghezze tipiche di legame sono intorno ai 3 Å [7].
2. **Forze elettrostatiche.** Intervengono tra gruppi di atomi parzialmente carichi in accordo con la legge di Coulomb; la loro intensità dipende dalla costante dielettrica del mezzo circostante. Il legame ionico, che si instaura a causa dell'attrazione elettrostatica fra gruppi con carica opposta, ha una lunghezza tipica di 2.8 Å e un'energia dell'ordine dei decimi di eV, analogamente al caso dei legami idrogeno.
3. **Forze di van der Waals.** Tre differenti tipi di forze contribuiscono all'interazione a lungo raggio nota come forza di van der Waals, l'energia di ciascuna delle quali varia come l'inverso della sesta potenza della distanza. I primi due contributi risultano da forze attrattive dipolo-dipolo tra molecole polari e dipolo-dipolo indotto tra molecole polari e non polari. L'ultimo contributo di natura quanto-meccanica si può calcolare considerando il dipolo indotto su una molecola o atomo non polare dal campo elettrico generato dal dipolo istantaneo (a media temporale nulla) di un altro atomo non polare vicino, e la forza attrattiva tra questi due dipoli.

L'attrazione fra coppie di atomi che deriva da queste forze aumenta al diminuire della distanza, ma se gli atomi si avvicinano tanto da sovrapporre i propri orbitali interviene tra loro una forte repulsione come conseguenza del principio di esclusione di Pauli. L'interazione effettiva totale è rappresentabile tramite un potenziale Lennard-Jones 6-12, così definito:

$$V(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]; \quad (1.1)$$

dove  $r$  è la distanza tra le due molecole. L'andamento di  $V(r)$  è mostrato in figura (1.3), i parametri  $\sigma$  e  $\epsilon$  sono rispettivamente la distanza corrispondente allo zero del potenziale e la profondità della buca di potenziale.

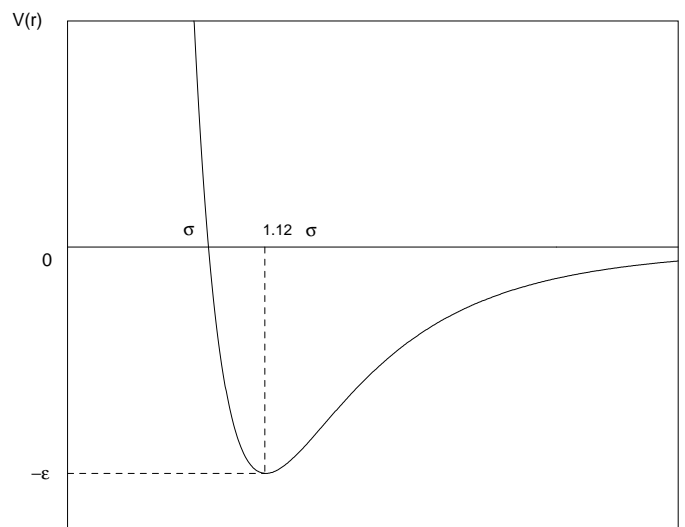


Figura 1.3. Andamento dell'energia potenziale di Lennard-Jones.

L'energia associata a queste interazioni è dell'ordine di  $0.04 \text{ eV}$  e la distanza corrispondente al minimo di  $V(r)$  può variare da  $7$  a  $9 \text{ \AA}$  tenendo in conto le diverse dimensioni dei vari aminoacidi.

Inoltre va tenuta in conto la presenza del solvente, che tipicamente è una soluzione acquosa. Le interazioni tra gli ioni, i dipoli, gli accettori e i donatori dei legami idrogeno vengono ridotte altamente dalla presenza delle molecole di acqua. In più l'acqua aumenta l'attrazione effettiva tra le molecole idrofobiche. Queste ultime infatti non posseggono momento di dipolo permanente (vengono dette anche molecole non polari) e questo non consente loro di partecipare alla formazione di legami idrogeno. Al contrario le molecole d'acqua hanno una forte tendenza a formare tali legami e quindi, quando vengono in contatto con molecole non polari, si orientano in modo tale da instaurare il minor numero possibile di legami con esse. Il risultato è che le molecole d'acqua tendono a circondare le molecole non polari, favorendone l'impacchettamento. Questo processo dà luogo a una attrazione effettiva tra le molecole non polari (o idrofobiche), che viene detta **interazione idrofobica** e la cui scala di energia associata è all'incirca pari a  $0.08 \text{ eV}$ .

Questa interazione è ritenuta essere la maggiore responsabile della stabilità delle proteine e della dinamica del loro ripiegamento [8, 9] durante il quale si tende a formare un nucleo ricco di residui idrofobici.

## 1.5 Il problema della predizione della forma nativa di una proteina

Abbiamo detto che in condizioni fisiologiche standard le proteine sono caratterizzate da proprietà specifiche quali l'attività funzionale, la struttura tridimensionale e altri parametri fenomenologici (carica elettrica globale, solubilità in solventi appropriati,

proprietà spettroscopiche). Questa condizione è denominata *stato nativo* della molecola proteica.

Esistono delle procedure artificiali che portano alla scomparsa di queste proprietà senza che sia distrutta la struttura primaria del polipeptide: innalzamento della temperatura o dei valori del pH, nonché ripetuti congelamenti e scongelamenti. Questi processi sono detti di denaturazione; ad esempio la temperatura di denaturazione “calda” di una proteina è di solito intorno ai 400 °K. La denaturazione può essere realizzata anche mediante l’aggiunta di agenti chimici, che vanno a alterare le interazioni che tengono insieme la struttura tridimensionale della proteina.

In una celebre serie di esperimenti, che gli valsero il premio Nobel per la chimica nel 1972, C. Anfisen e il suo gruppo denaturarono e rinaturarono l’enzima ribonucleasi mediante reazione e rimozione di alcuni agenti denaturanti [10]. Dopo la loro rimozione si otteneva la completa ricomparsa della struttura e delle proprietà dell’enzima nativo.

Questo esperimento ha portato ad alcune importanti conclusioni, che sono state sostenute dai numerosi esperimenti che si sono succeduti e si compiono tuttora per definire quello che in letteratura è denominato il “dogma” fondamentale della determinazione genetica della struttura della proteina:

- tutte le informazioni necessarie al ripiegamento (*fold*ing) di una proteina sono contenute nella sua struttura primaria (cioè nella sequenza degli aminoacidi);
- lo stato nativo corrisponde ad uno stato di minima energia libera che si raggiunge attraverso un processo spontaneo e sequenziale, il cui andamento cinetico è speculare rispetto a quello di denaturazione (*un*folding).



Le questioni di come determinare la struttura tridimensionale che assumerà una data sequenza di aminoacidi (problema del *fold*ing diretto) e di come risalire alle strutture primarie che ammettono una data conformazione tridimensionale (problema del *fold*ing inverso) non sono ancora state risolte dopo trenta anni di intensi studi, che hanno coinvolto sia chimici che biologi, ma anche fisici e matematici. Nel tempo intercorso tra l'esperimento di Anfisen e oggi la ricerca sul problema del ripiegamento delle proteine ha dato comunque risultati innovativi riguardo la dinamica del processo e le interazioni che maggiormente lo determinano. Tutti questi nuovi risultati non hanno contraddetto l'affermazione che le proteine, pur attraverso itinerari diversificati, raggiungano una unica conformazione nativa, che corrisponde alla forma dotata di attività biologica. Essa sembra quindi corrispondere allo stato di massima stabilità termodinamica ed è specifica per ciascuna proteina, in quanto dipende dalla sequenza degli aminoacidi geneticamente determinata.

Risolvere il problema del *fold*ing, sia diretto che inverso, sarebbe di grande importanza nel campo dell'ingegneria genetica per progettare nuove proteine, per disporre di nuovi medicinali e per poter comprendere ed eventualmente controllare la funzionalità delle centinaia di sequenze proteiche che ogni giorno vengono scoperte nei laboratori di biotecnologia.

## 1.6 Tempi di ripiegamento

In linea di principio si potrebbe prevedere la struttura terziaria di una proteina disponendo di un programma di dinamica molecolare, che integri numericamente le leggi di moto di ciascun atomo costituente la proteina considerando un modello il

più realistico possibile, che includa le interazioni osservate sperimentalmente e tutti i dettagli del sistema proteina-solvente (approccio *ab-initio*). I moti vibrazionali degli atomi costituenti le proteine, hanno una frequenza di  $10^{13}$ - $10^{14}$  Hz e quindi un metodo di integrazione numerica *ab-initio* richiede un passo temporale almeno corrispondente ai femtosecondi. Per questa ragione e dato il grande numero di atomi che costituiscono una proteina, i calcolatori più recenti possono simulare attualmente fino a  $10^{-9}$  sec dell'evoluzione di una proteina reale, intervallo di tempo che risulta di fatto insufficiente, essendo molto minore dei tempi tipici di ripiegamento:  $10^{-3}$ - $10$  sec.

In alternativa si potrebbe pensare di individuare lo stato nativo calcolando l'energia libera di tutte le conformazioni possibili in modo da trovare quella con energia libera minore. Questo risulta di nuovo irrealizzabile poiché in una catena ciascun aminoacido può assumere in media 10 orientazioni [1] e quindi il numero di conformazioni possibili di una proteina composta da N aminoacidi è  $10^N$ . Un sistema di calcolatori in parallelo della potenza attuale (fino a  $10^{12}$  operazioni *floating point* al secondo) impiegherebbe un tempo dell'ordine di  $10^{20}$  anni per terminare il calcolo.

Dato il numero enorme di configurazioni è anche da escludere che la proteina si ripieghi vagliando tutte le conformazioni possibili. Una stima teorica del tempo di ripiegamento, assumendo gli stati equiprobabili, porta a tempi maggiori dell'età dell'universo. Questo sottolinea l'importanza degli aspetti cinetici del processo di ripiegamento, dato che, evidentemente, la proteina raggiunge lo stato nativo attraverso itinerari preferenziali, che visitano solo una parte limitata di tutto lo spazio delle fasi accessibile.

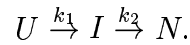
## 1.7 Proprietà cinetiche del processo di ripiegamento

Come abbiamo già notato nel paragrafo precedente, bisogna escludere che le proteine si ripieghino campionando tutte le conformazioni possibili. Su questa considerazione si basa il “paradosso” di Levinthal [1], che evidenzia l'impossibilità della proteina di trovare lo stato nativo attraverso una ricerca casuale tra un numero astronomico di configurazioni possibili. Per realizzarsi nelle scale di tempo osservate sperimentalmente, il processo di ripiegamento deve invece essere diretto lungo cammini specifici nello spazio delle fasi corrispondente. Questo evidenzia l'importanza dei fattori cinetici nel processo di *fold*ing e implica che il processo di ripiegamento non è semplicemente assimilabile alla ricerca di un minimo assoluto di energia libera. Non è detto infatti che uno stato dinamicamente o termodinamicamente stabile sia cineticamente accessibile a partire da qualsiasi condizione iniziale.

Per determinare le caratteristiche dei cammini relativi ai processi di *un*folding e *fold*ing bisogna quindi identificare gli stati intermedi attraverso i quali passa la proteina, anche se questi risultano termodinamicamente instabili. Se la reazione dallo stato denaturato U verso lo stato nativo N avviene senza la presenza di stati intermedi stabili I (transizione a due stati), essa sarà descritta da una sola costante di reazione  $k$ . L'andamento nel tempo della frazione di proteine ripiegate nello stato nativo rispetto al numero totale di proteine presenti in una soluzione avrà la forma:

$$\text{frazione di proteine native} = 1 - \exp(-kt). \quad (1.2)$$

Una reazione che invece avviene attraverso uno stato intermedio I sarà caratterizzata da due costanti di reazione  $k_1$  e  $k_2$ :



La comparsa dello stato N avviene con un ritardo rispetto alla scomparsa dello stato U pari a  $(k_1 + k_2)^{-1}$ , durante il quale lo stato I risulta popolato. Il ritardo aumenta con il numero di stati intermedi sequenziali lungo il cammino.

Osservazioni sperimentali quantitative sulle proteine di singolo dominio indicano che la transizione di *unfolding* avviene tramite una sola costante di reazione, dato che non si osservano tempi di ritardo. La costante  $k$  varia con le condizioni di *unfolding*, in particolare si ha che

$$k \propto \exp\left(-\frac{C}{T}\right).$$

dove  $T$  è la temperatura a cui avviene la transizione e  $C$  è un'opportuna costante.

Nelle transizioni di *folding* si possono invece individuare delle conformazioni denaturate, che si ripiegano più lentamente ( $U_S$ ) e delle conformazioni, sempre denaturate, che hanno tempi di *folding* caratteristici più brevi ( $U_F$ ). In pratica si può schematizzare la transizione di *folding* nel seguente modo:



I processi lenti corrispondono a delle conformazioni che presentano disposizioni dei monomeri particolarmente sfavorevoli a riarrangiarsi nella conformazione nativa. Questi processi a più costanti di reazione avvengono solo in determinate condizioni e se nella proteina è presente un particolare tipo di aminoacido [3]. Se si escludono questi casi, anche la transizione di *folding* appare essere un processo determinato da

una sola fase cinetica. Questo indica che durante la reazione c'è un solo particolare stadio che determina la velocità di reazione; tutti gli stadi precedenti o successivi sono molto più rapidi. Il fatto che una grande varietà di conformazioni eterogenee (le conformazioni denaturate) si ripieghino con la stessa velocità fa pensare a una rapida equilibratura conformazionale precedente al ripiegamento. Che questo accada è indicato dall'osservazione che la velocità di ripiegamento è indipendente dalla procedura di denaturazione, ma dipende unicamente dalle proprietà dello stato nativo finale.

## 1.8 Caratteristiche del processo di denaturazione

Analizziamo adesso le caratteristiche termodinamiche della transizione da stato nativo a stato denaturato (transizione di *unfolding*) di una proteina a un solo dominio. Cambiando gradualmente le condizioni fisiologiche di una proteina allo stato nativo, per esempio aumentando la temperatura o la concentrazione di agente denaturante, abbiamo che inizialmente la conformazione della proteina rimane praticamente invariata. Si potranno verificare aumenti della flessibilità e piccole alterazioni conformazionali localizzate ma la struttura media rimane sostanzialmente la stessa. Procedendo nella variazione di questi parametri si nota che la proteina passa bruscamente dalla fase nativa alla fase denaturata. Se consideriamo una soluzione diluita di proteine, la curva che indica la frazione di proteine denaturate rispetto alla temperatura o alla concentrazione di agente denaturante risulta avere una forte pendenza.

La dipendenza dalla temperatura risulta avere una notevole importanza nello studio

della termodinamica del processo di *unfolding*. Misure calorimetriche [11] sugli effetti dell'innalzamento della temperatura di una soluzione contenente l'enzima lisozima sono mostrate in figura (1.4).

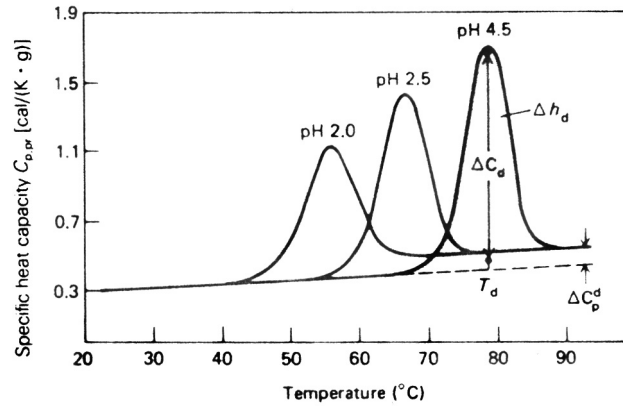


Figura 1.4. Misure calorimetriche di *unfolding* termico di una soluzione di lisozima per tre differenti valori del pH (la figura è stata ripresa da [11]).

La quantità misurata direttamente è il calore specifico a pressione costante dell'intera soluzione: il contributo dovuto alla proteina si ottiene sottraendo i valori ottenuti da misure analoghe eseguite sul solvente puro. Il calore specifico a pressione costante della proteina  $c_p$  risulta avere inizialmente un andamento quasi costante rispetto all'incremento di temperatura. In corrispondenza della transizione di *unfolding* l'andamento del calore specifico mostra un picco evidente, che indica un alto assorbimento di calore. Quando il processo di *unfolding* è terminato il calore specifico  $c_p$  risulta essere di nuovo quasi costante e maggiore del suo valore iniziale. I valori costanti del calore specifico nelle regioni al di fuori della transizione corrispondono allo stato nativo e allo stato completamente denaturato della proteina.

Misure sperimentali [12] hanno infatti evidenziato che il calore specifico dei residui aminoacidi immersi in un solvente povero è direttamente proporzionale all'area della superficie non polare esposta al solvente. Il valore di  $c_p$  ad alte temperature equivale, per esempio, al valore corrispondente al caso in cui tutte le catene laterali siano esposte al solvente. Di conseguenza il valore del calore specifico dopo la transizione è maggiore di quello precedente la transizione; infatti la differenza  $\Delta c_p$  tra questi due valori risulta proporzionale all'area della superficie non polare nascosta nell'interno della proteina ripiegata e successivamente esposta al solvente nella proteina denaturata.

Questi comportamenti fanno supporre che la transizione di *unfolding* sia una transizione a due-stati dove gli stati intermedi sono instabili rispetto agli stati iniziali e finali. Se per esempio pensiamo a una soluzione di proteine, avremo che, nella fase transiente, una parte di proteine si trova nello stato nativo e la restante parte nello stato denaturato, piuttosto che una frazione considerevole di proteine abbia assunto uno stato parzialmente ripiegato. Variazioni dal comportamento a due-stati della transizione si hanno nel caso di proteine a più domini, per le quali può avvenire che i singoli domini si ripieghino indipendentemente in tempi diversi. Nel caso di molte proteine a due domini la curva del calore specifico risulta essere la convoluzione di due picchi, a indicare che la natura del ripiegamento di ogni singolo dominio è simile a quella descritta per una proteina a un singolo dominio.

Inoltre va notato che il processo di denaturazione della maggioranza delle proteine a un solo dominio risulta essere reversibile. Ristabilendo le condizioni ottimali per il ripiegamento la proteina ritorna nello stato nativo (transizione di *refolding*) e non

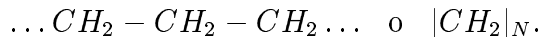
viene osservata alcuna perdita di attività biologica.

## 1.9 Cenni sulla teoria dei polimeri lineari

Poiché gli studi sulle proteine fanno uso di molti concetti introdotti nella teoria dei polimeri, diamo di questa alcuni cenni essenziali, soffermandoci principalmente su quei risultati, a cui in seguito faremo riferimento.

Un polimero lineare (od omopolimero) [13] è costituito da una serie di molecole o gruppi di atomi (monomeri) identici tra loro, legati a formare una catena lineare.

Un esempio semplice è il polietilene, il quale presenta la seguente struttura:



Nella schematizzazione più semplice, spesso utilizzata per rappresentare la struttura dei polimeri, i monomeri vengono ridotti a dei punti materiali <sup>2</sup>, che vengono uniti tramite dei vettori  $\vec{l}$ , come mostrato in figura (1.5).

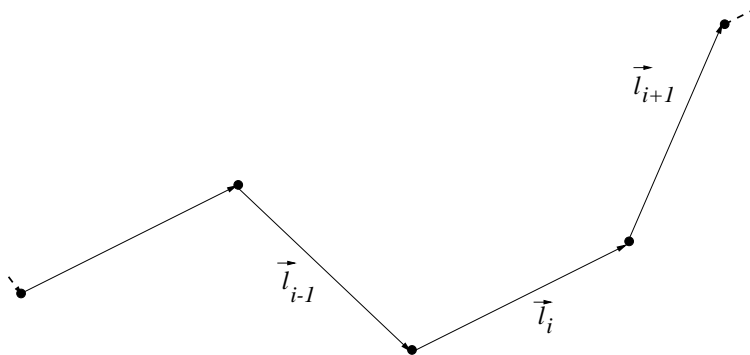


Figura 1.5. Schematizzazione di un polimero

---

<sup>2</sup>Spesso si considerano solo gli atomi che fanno parte dello scheletro della catena (nell'esempio del polietilene, solo gli atomi di carbonio), poichè la posizione degli altri atomi periferici costituenti il monomero è fissa rispetto a questi.



Per individuare i monomeri e i legami lungo la catena si utilizza un indice intero  $i = 1 \dots N$ , dove  $N$  è il numero totale di monomeri, detto anche grado di polimerizzazione del polimero. In certi casi  $N$  può raggiungere l'ordine di grandezza delle migliaia.

Nello studio dei polimeri lineari va tenuto chiaramente in conto l'ambiente con cui interagiscono. Consideriamo il caso di un polimero immerso in un solvente. I solventi vengono classificati in solventi buoni e solventi poveri rispetto al polimero. I buoni solventi sono composti da molecole che hanno polarizzabilità elettronica simile a quella dei monomeri del polimero e sono quindi favorevoli a interagire e a "mescolarsi" con il polimero. I solventi poveri sono invece quei solventi le cui molecole preferiscono legarsi tra loro piuttosto che con il polimero, avendo polarizzabilità elettronica diversa. La differenza tra le due polarizzabilità elettroniche in generale decresce all'aumentare della temperatura, per cui lo stesso solvente potrà essere considerato buono o povero per un polimero a seconda della temperatura a cui si trova il sistema. In ogni caso si suppone di poter trattare il solvente attraverso un termine efficace in modo tale che non ci sia bisogno di una descrizione dettagliata a livello molecolare del solvente per studiarne le interazioni con la catena.

Riprendiamo l'esempio del polietilene e descriviamone la struttura. L'angolo  $\beta$  tra due legami consecutivi  $C - C$  risulta essere sostanzialmente fisso. Questi due legami nel complesso coinvolgono tre atomi di carbonio, per esempio  $C_{i-3}$ ,  $C_{i-2}$ ,  $C_{i-1}$ . Se consideriamo il successivo atomo di carbonio  $C_i$ , esso sarà individuato dall'angolo fisso  $\beta$  tra il vettore  $C_{i-1} - C_i$  e il vettore  $C_{i-2} - C_{i-1}$  e dall'angolo  $\phi_i$  di rotazione del vettore  $C_{i-1} - C_i$  attorno al vettore  $C_{i-2} - C_{i-1}$ . L'andamento dell'energia delle

configurazioni, che si ottengono al variare di  $\phi_i$ , presenta tre minimi [13], due dei quali simmetrici, le cui conformazioni corrispondenti sono dette *cis*; la configurazione corrispondente al terzo minimo rimanente è detta *trans*. Il parametro importante è la differenza di energia  $\Delta\epsilon$  tra le configurazioni *cis* e *trans*. Se l'energia termica  $k_B T$  è maggiore di  $\Delta\epsilon$  si dice che la catena è staticamente flessibile. Infatti se consideriamo la catena per intero, il rapporto tra il numero di conformazioni *cis* e *trans* sarà circa 1 e la catena non si presenterà completamente aperta e stirata ma assumerà una configurazione non ben definita. Questa configurazione può presentare locali aggregazioni ma non risulta comunque compatta: questo stato è detto “spirale” o stato *coil*. Se invece  $\Delta\epsilon > k_B T$  sarà favorito il minimo di energia più basso, che, nel caso del polietilene, corrisponde alla conformazione *trans* e la catena dovrà essere considerata rigida almeno a livello locale.

Un'altra grandezza importante per la caratterizzazione delle configurazioni che può assumere una catena è il raggio di girazione  $R_g$ , definito come segue:

$$R_g^2 = \frac{1}{N} \sum_{i=1}^N |\vec{r}_i - \vec{r}_{cm}|^2; \quad (1.3)$$

dove  $\vec{r}_{cm}$  è il centro di massa del sistema, ed  $\vec{r}_i$  è il vettore posizione del monomero  $i$ .

Un buon modello, che ripropone le caratteristiche principali di un omopolimero immerso in un solvente buono, rappresenta la catena lineare come un cammino su un reticolo periodico in cui ogni sito può essere attraversato una sola volta (*self-avoiding walk* (SAW)). Si identifica, sulla base di questo modello, l'esponente critico  $\nu$  tale che la media statistica del quadrato del raggio di girazione del sistema  $R_g$  scali come  $N^{2\nu}$ :  $\langle R_g^2 \rangle \propto N^{2\nu}$ , con  $\nu = \frac{3}{d+2}$ , essendo  $d$  la dimensione dello spazio in cui il

sistema è definito [13].

In un solvente povero un polimero si comporta piuttosto come un cammino *random* su un reticolo periodico. Per ogni passo verso un sito il passo successivo può avvenire in uno qualunque dei siti primi vicini e il peso statistico associato è lo stesso. Questo modello prevede che la media statistica del quadrato del raggio di girazione  $R_g$  scali ora come:  $\langle R_g^2 \rangle \propto N^{2/d}$  [13]. Il volume occupato da un polimero in un solvente buono è dunque maggiore di quello occupato dallo stesso polimero in un solvente povero.

Consideriamo a questo punto la dinamica di un omopolimero immerso in un solvente al variare della temperatura. Tenere conto in modo esplicito dei gradi di libertà del solvente renderebbe di fatto intrattabile qualunque modello di tipo dinamico. Pertanto si tiene conto del solvente in modo efficace, introducendo un potenziale effettivo d'interazione tra le coppie di monomeri. Al variare della temperatura si osserva che l'omopolimero può presentarsi in due fasi distinte, a seconda che prevalgano le forze di legame o quelle repulsive. Ad alte temperature (nel regime di buon solvente) la parte attrattiva del potenziale è trascurabile mentre sono dominanti le repulsioni steriche; la catena si presenta in fase *coil* e può essere studiata con modelli SAW su reticolo ( $\langle R_g^2 \rangle \sim N^{2\nu}$ ). All'abbassarsi della temperatura (regime di solvente povero) i termini attrattivi diventano più rilevanti e la catena sempre più compatta ( $R_g^2 \sim N^{\frac{2}{d}}$ ).

La temperatura a cui avviene la transizione di fase è detta temperatura  $\theta$  del polimero. In corrispondenza di questa transizione si ha un picco del calore specifico del sistema. Un solvente- $\theta$  per un polimero è un solvente nelle condizioni per cui le

interazioni con la catena bilanciano esattamente le forze attrattive e repulsive tra i differenti monomeri che formano la catena.

Vari lavori sperimentali evidenziano il collasso dei polimeri immersi in solventi poveri (citiamo per esempio [14]); tuttavia la conformazione finale di un polimero che collassa non è unica. Sebbene il numero delle conformazioni massimamente compatte sia una frazione molto piccola del numero totale di conformazioni, il numero assoluto di conformazioni compatte che un polimero può assumere è grande e cresce esponenzialmente con la lunghezza della catena. È evidente quindi che l'unicità della struttura nativa di una proteina non possa provenire dai soli impedimenti sterici, ma che sia codificata nella sequenza degli aminoacidi.

## 1.10 Caratterizzazione degli stati denaturati

Riprendiamo adesso la descrizione della struttura delle proteine considerandone la fase denaturata. Come abbiamo già detto anche le proteine sono catene lineari e possono essere schematizzate come nel paragrafo 1.9. I monomeri rappresentano in questo caso gli aminoacidi e oltre a dover specificare la posizione lungo la catena occorrerà in generale identificarne la specie.

Gli stati denaturati sono caratterizzati dalla perdita dell'attività biologica della proteina e da una crescita evidente delle sue dimensioni. Sotto forti condizioni di denaturazione è stato osservato che le proteine si comportano in modo simile ai polimeri immersi in solventi buoni e possono essere studiate con modelli SAW su reticolo ([3]). Le proteine si presentano quindi nella fase *coil* caratterizzata da una forte entropia configurazionale. Questi modelli valgono però solo parzialmente, infatti le

proteine sono eteropolimeri e ciascun monomero di cui sono composte può interagire in maniera diversa con il solvente o con gli altri monomeri. Di conseguenza, le conformazioni corrispondenti agli stati denaturati possono non avere la stessa energia o probabilità di occorrenza di quelle predette per gli stati *coil* dei polimeri immersi in buoni solventi. Le proteine naturali sono sistemi più complessi dei polimeri, data la diversità delle catene laterali. Un solvente buono per una data catena laterale può risultare povero per altri gruppi idrofobici costituenti la stessa proteina, per cui, nel caso di una proteina, di fatto non è possibile definire il solvente- $\theta$ .

Ci sono ampie evidenze che, sotto condizioni di denaturazione meno estreme, gli stati denaturati di alcune proteine siano diversi dagli stati *coil*. È stata dunque introdotta, oltre allo stato *coil*, un'altra classe di stati denaturati detta stati denaturati compatti o *molten globule*.

Gli stati *coil* e *molten globule* sono indistinguibili termodinamicamente, avendo, ad esempio, lo stesso calore specifico. Le caratteristiche strutturali precise di queste conformazioni non sono ancora ben conosciute; del resto è evidente la difficoltà di caratterizzare le conformazioni non native dato il numero enorme di conformazioni possibili. Lo stato *molten globule* sembra essere lo stato più probabile delle proteine denaturate sotto condizioni di *refolding* e sembra inoltre avere un carattere transiente. Le caratteristiche dello stato *molten globule* sono le seguenti:

- (a) le dimensioni caratteristiche delle conformazioni sono molto minori di quelle dello stato *coil* e lievemente maggiori dello stato nativo; queste conformazioni compatte, possiedono entropia conformazionale finita, e risultano essere non ben definite, come avviene nel caso degli omopolimeri immersi in un solvente

povero;

- (b) lo stato denaturato compatto è caratterizzato dalla presenza di strutture secondarie non necessariamente corrispondenti a quelle dello stato nativo;
- (c) le catene laterali sono completamente esposte al solvente, a differenza dello stato nativo, in cui le catene laterali si trovano in ambienti non omogenei;
- (d) risultano molto più rapide le transizioni dallo stato denaturato compatto verso lo stato *coil* che non quelle verso lo stato nativo.

Le proteine in fase denaturata potrebbero essere caratterizzate da un insieme continuo di conformazioni, che vanno da una conformazione completamente non ripiegata a una conformazione *molten globule* compatta. Bisogna comunque distinguere lo stato *molten globule* dall'insieme delle configurazioni compatte che può assumere una data sequenza sintetizzata artificialmente. Sequenze progettate “scorrettamente” possono ripiegarsi in un insieme di stati compatti corrispondenti a stati multipli di energia minima, detti stati *gemisch* [15]. La distinzione sperimentale tra gli stati *molten globule* e gli stati *gemisch* è che questi ultimi, al variare delle condizioni sperimentali, non subiscono transizioni di fase verso stati più ordinati (corrispondenti a una diminuzione dell'entropia). In altre parole, nonostante la similitudine tra le strutture *gemisch* e *molten globule*, la differenza sostanziale sta nella capacità o meno di ripiegarsi in maniera univoca sotto opportune condizioni sperimentali. Ad esempio le sequenze composte da un alta concentrazione di monomeri idrofobici spesso si ripiegano in stati *gemisch*, risultando di fatto cattivi ripiegatori.

# Capitolo 2

## Modelli fisici e metodi d'indagine

Avendo mostrato nel precedente capitolo l'effettiva impraticabilità di un approccio " *ab initio*" al problema del ripiegamento delle proteine, qui ci proponiamo di descrivere alcuni dei modelli semplificati che hanno riscosso maggior consenso e interesse in letteratura e che sono assai prossimi a quello preso in considerazione in questa tesi. Gli obiettivi ai quali miriamo in questo breve capitolo sono i seguenti. Discutere gli aspetti rilevanti e anche i limiti di queste schematizzazioni spesso drastiche di una struttura biologica complessa, come risulta essere in realtà una proteina. Introdurre una serie di concetti, che risulteranno particolarmente utili in seguito per la descrizione del modello studiato in questa tesi e introdotto nel capitolo successivo. Nell'ultimo paragrafo sono invece presentate alcune considerazioni generali sulle caratteristiche del paesaggio energetico tipico di una proteina, che, è bene sottolinearlo, si comporta come un sistema termodinamico, a dispetto del numero relativamente piccolo dei suoi gradi di libertà.

## 2.1 Modelli semplici esatti

In una vasta parte di letteratura scientifica le proteine vengono schematizzate attraverso modelli di catene di eteropolimeri su reticoli in due e tre dimensioni spaziali. In questi modelli ciascun sito del reticolo può essere libero, oppure occupato da un monomero. Gli angoli di legame possono assumere solo alcuni valori discreti, fissati dalla struttura reticolare. I limiti dei modelli su reticolo sono evidenti: si perde risoluzione nella rappresentazione dei dettagli strutturali ed energetici delle proteine. D'altra parte i modelli semplici risolubili esattamente hanno anche notevoli vantaggi. Come abbiamo detto nel paragrafo 1.6 le simulazioni a livello atomico possono esplorare solamente piccoli cambiamenti nello spazio delle configurazioni, che avvengono nei brevi tempi raggiungibili (tipicamente dell'ordine dei nanosecondi). Con questi modelli semplici si può seguire l'evoluzione del sistema fino a tempi molto più lunghi, che includono anche i tempi di ripiegamento. Inoltre, nel caso di catene sufficientemente corte, è possibile l'individuazione dell'intero spazio delle configurazioni e quindi il calcolo della funzione di partizione, attraverso la quale, grandezze fondamentali come l'entropia, l'energia libera, il calore specifico ecc. sono facilmente calcolabili senza ulteriori approssimazioni oltre a quelle intrinseche alla definizione del modello stesso.

### 2.1.1 Il modello HP

Uno dei modelli più utilizzati è il modello HP su reticolo quadrato o cubico semplice (vedi per esempio [16, 17]), nel quale vengono considerate solo due classi di monomeri a rappresentare la tendenza idrofobica o polare degli aminoacidi. L'Hamiltoniana



$\mathcal{H}$ , associata a una catena composta da  $N$  monomeri in una data sequenza  $S = (s_1, s_2, \dots, s_N)$  che si trova in una configurazione spaziale  $\Gamma$ , è data da:

$$\mathcal{H}_S(\Gamma) = \sum_{ij} \epsilon(s_i, s_j) \Delta(\vec{r}_i - \vec{r}_j); \quad (2.1)$$

dove  $\vec{r}_i$  è la posizione sul reticolo del monomero  $i$ -esimo nella configurazione  $\Gamma$  e  $\Delta(\vec{r}_i - \vec{r}_j)$  è 1 se  $\vec{r}_i$  e  $\vec{r}_j$  sono primi vicini nel reticolo senza essere monomeri contigui nella catena (in questo caso  $\vec{r}_i - \vec{r}_j$  viene detto contatto non di legame) e 0 altrimenti. Il monomero  $s_i$  può essere idrofobico, ( $H$ ), o polare, ( $P$ ), in modo che  $\epsilon(H, H)$ ,  $\epsilon(H, P)$ ,  $\epsilon(P, P)$  rappresentino le possibili energie d'interazione. Per favorire la tendenza dei monomeri idrofobici ad agglomerarsi in un nucleo interno non esposto al solvente, viene scelta tra questi un' interazione attrattiva ( $\epsilon(H, H) < 0$ ), energeticamente favorita rispetto alle altre interazioni ( $\epsilon(H, H) < \epsilon(H, P), \epsilon(P, P)$ ). Le applicazioni di tale modello in due o tre dimensioni hanno dato risultati qualitativi simili. Peraltro è stato osservato che per riprodurre il rapporto corretto tra il volume e la superficie esterna di una proteina, come la mioglobina, è necessario considerare una catena di 154 monomeri per il modello in 3D, mentre sono sufficienti appena 16-20 monomeri in un modello 2D ([18]). Questo indica comunque che sono da attendersi delle sostanziali differenze quantitative in funzione delle dimensioni spaziali.

Il modello HP bidimensionale riproduce molte proprietà generali delle proteine globulari. Ad esempio, sotto condizioni che favoriscono la denaturazione (per  $\epsilon(H, H) \lesssim 0$ ), le catene visitano un insieme relativamente grande dello spazio delle configurazioni, corrispondente a stati denaturati della proteina. Con l'aumentare della forza attrattiva tra i monomeri idrofobici le catene iniziano a visitare un insieme sempre

più piccolo dello spazio delle configurazioni (per alcune sequenze tale insieme è composto da una sola configurazione o due), che risultano compatte e hanno un nucleo idrofobico al proprio interno. Nella figura (2.1) riportiamo la configurazione nativa di una data sequenza per il modello HP in 3 dimensioni dove, in particolare, i parametri d'interazione  $\epsilon(H,H)$ ,  $\epsilon(P,P)$  e  $\epsilon(H,P)$  sono stati presi rispettivamente pari a -1, 0, 0.

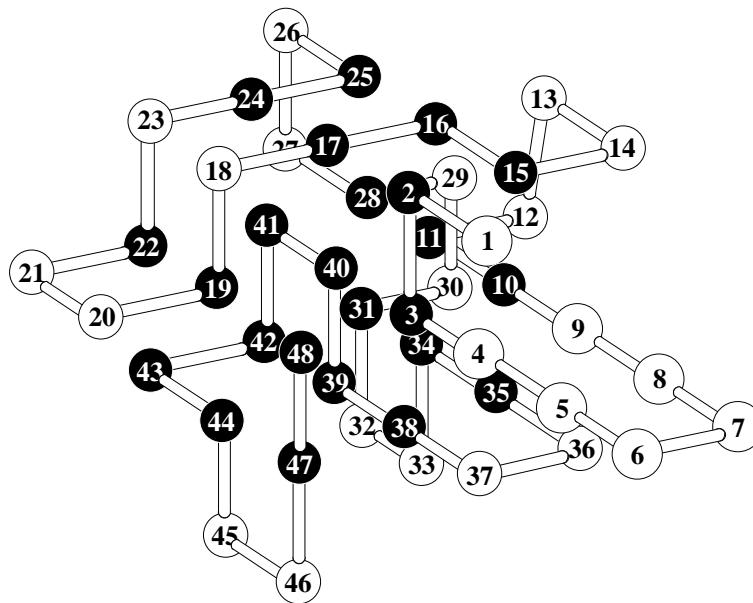


Figura 2.1. Configurazione nativa del modello HP in 3D (la figura è stata ripresa da [17]).

### 2.1.2 Il modello di $G\bar{o}$

Un altro modello che è stato ampiamente utilizzato e rivisitato in tempi recenti è il cosiddetto modello di  $G\bar{o}$ , che fu introdotto a metà degli anni '70 ([19],[20]). Nel lavoro originale  $G\bar{o}$  schematizzò una proteina con una catena lineare costituita da 49 monomeri che potevano occupare i siti di un reticolo quadrato. Una volta fissata

a priori una configurazione nativa arbitraria, ma massimamente compatta (cioè che occupi tutti i siti di un quadrato  $7 \times 7$ ), Gō introdusse delle specifiche interazioni *ad-hoc* a “corto raggio” e a “lungo raggio”, in modo tale da favorire il ripiegamento della catena sulla configurazione nativa prefissata. Si intende per interazioni a “lungo raggio” quelle che si instaurano fra le coppie di monomeri non contigui lungo la catena, che si trovino però ad essere primi vicini sul reticolo. Per interazioni a “corto raggio” si intendono, invece, quelle fra monomeri che siano contigui lungo la catena. Le interazioni a “corto raggio” vengono rappresentate tramite un termine di energia funzione dell’angolo di legame  $\alpha_i$  relativo a ciascun monomero  $i$  della catena, cioè l’angolo formato dai due vettori congiungenti rispettivamente tale monomero al precedente,  $i-1$ , e al successivo,  $i+1$  (secondo il verso prefissato di percorrenza della catena). L’energia associata a queste interazioni è determinata in modo tale da favorire la formazione della struttura nativa. Infatti, se  $\alpha_i$  è uguale al valore che assume nella conformazione nativa, all’energia si attribuisce un valore pari a  $-\epsilon'$ , mentre negli altri casi tale valore è pari a zero.

La trattazione delle interazioni a “lungo raggio” è leggermente più complicata; in questo caso i “contatti nativi” (cioè quelli corrispondenti alla configurazione nativa), non sono favoriti in maniera esclusiva ma è favorita anche la formazione di altri contatti (non nativi). Questo per rendere conto degli intrappolamenti in stati metastabili, che la proteina subisce durante il reale processo di ripiegamento. In pratica, fra tutti i possibili contatti realizzabili fra monomeri non contigui della catena, se ne seleziona un terzo (detti contatti favoriti) e ad ognuno di questi si associa un contributo di energia  $-\epsilon$ , mentre la formazione di uno qualsiasi degli altri contatti

non altera l'energia della catena. Fra i contatti favoriti sono inclusi ovviamente tutti i contatti nativi e una certa quota di contatti non nativi scelti in maniera casuale. In questo modo, per quanto riguarda il contributo delle interazioni a lungo raggio, alla configurazione nativa è associata l'energia minore possibile (cioè  $-36\epsilon$ ), mentre una conformazione arbitraria compatta ha in media un'energia "a lungo raggio" pari a  $-12\epsilon$ , essendo i contatti favoriti un terzo di tutti quelli possibili.

Al fine di studiare le proprietà meccanico-statistiche di questo modello, Gō ha applicato il metodo Monte Carlo per diversi valori di  $(\epsilon, \epsilon')$  e della temperatura <sup>1</sup>. Egli ha effettuato sia simulazioni di *folding* (partendo da una conformazione aperta) che di *unfolding* (partendo dalla configurazione nativa). Da questo modello si possono trarre informazioni interessanti sulla dinamica del processo di ripiegamento, seguendo lo stato del sistema durante l'evoluzione Monte Carlo. Gō sottolineò che per riproporre un comportamento simile a quello di una proteina reale è fondamentale la presenza di entrambe le interazioni. Infatti, se ci si limita a tener in conto solo delle interazioni a "lungo raggio", si osserva una corretta descrizione del fenomeno dell'*unfolding*. Per temperature sufficientemente basse, la proteina fluttua attorno allo stato nativo con ampiezza delle fluttuazioni crescente con la temperatura. Quando la temperatura si avvicina a un valore critico (temperatura di transizione) il sistema dapprima evolve intorno allo stato nativo per poi passare bruscamente allo stato denaturato, mostrando un picco nel calore specifico. Tuttavia nelle simulazioni di *folding*, alla temperatura di transizione, il sistema non raggiunge lo stato nativo, ma viene intrappolato in una conformazione corrispondente a un minimo locale dell'energia.

---

<sup>1</sup>Nel metodo Monte Carlo si definisce temperatura il parametro presente nel peso di Boltzmann da cui dipende la probabilità di accettare o meno i cambiamenti delle conformazioni che aumentano l'energia del sistema.

Aggiungendo al modello le interazioni a corto raggio si può infine osservare anche il processo di ripiegamento verso la configurazione nativa.

## 2.2 Un modello fuori reticolo

Il modello che studieremo in questa tesi è molto simile al modello introdotto alcuni anni fa da Stillinger *et al.* ([21],[22]). Nel modello di Stillinger la catena lineare, che rappresenta la proteina, è definita su un piano e i monomeri non sono costretti a occupare i siti di un reticolo.

La lunghezza di legame è fissata pari a 1<sup>2</sup> e l'angolo compreso tra due vettori di legame  $\vec{l}_{i-1}$  e  $\vec{l}_i$  è detto  $\theta_i$  (vedi figura (2.2)). Come nel modello HP i monomeri

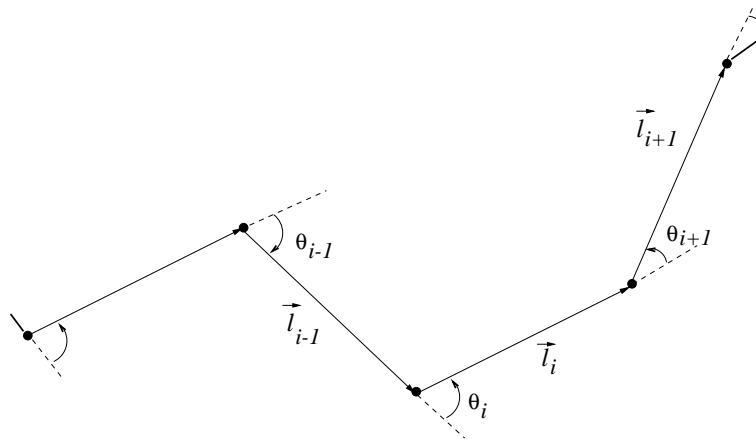


Figura 2.2. Catena lineare in due dimensioni

possono essere di soli due tipi: idrofobici (H) e polari (P) e vengono identificati dalle variabili binarie  $\xi_i$ , che assumono i valori +1 e -1 a seconda che il monomero  $i$  sia rispettivamente di tipo H o P.

---

<sup>2</sup>In questo modello tutte le grandezze in gioco sono rese adimensionali.

Si assume che per ciascun monomero  $i$  si debbano considerare due contributi all'energia potenziale: il primo,  $V_c$ , dipende solo dall'angolo  $\theta_i$  e quindi è legato alla "curvatura" della catena intorno al monomero considerato (interazione a corto raggio); il secondo,  $V_l$ , è legato alle interazioni con tutti gli altri monomeri non direttamente legati a  $i$  (interazione a lungo raggio). L'energia potenziale totale  $V$  relativa a una catena composta da  $N$  monomeri in sequenza  $\{\xi_i\}$ , con  $i = 1, \dots, N$ , è data da:

$$V = \sum_{i=2}^{N-1} V_c(\theta_i) + \sum_{i=1}^{N-2} \sum_{j=i+2}^N V_l(r_{ij}, \xi_i, \xi_j), \quad (2.2)$$

dove  $r_{ij}$  è la distanza euclidea tra i monomeri  $i$  e  $j$  che può essere espressa in funzione degli angoli  $\theta_i$ . Il primo termine della (2.2) tende a favorire l'allineamento della catena;

$$V_c(\theta_i) = \frac{1}{4}(1 - \cos(\theta_i)). \quad (2.3)$$

Il termine  $V_l$ , legato alle interazioni non di legame, ha la forma tipo Lennard-Jones 12-6:

$$V_l = \left[ \frac{1}{r_{i,j}^{12}} - \frac{C(\xi_i, \xi_j)}{r_{i,j}^6} \right];$$

dove la dipendenza dalla natura dei monomeri è contenuta nel coefficiente:

$$C(\xi_i, \xi_j) = \frac{1}{8}(1 + \xi_i + \xi_j + 5\xi_i\xi_j). \quad (2.4)$$

In questo modo l'interazione tra due monomeri idrofobici risulta fortemente attrattiva ( $C(+1, +1) = 1$ ), mentre l'interazione tra due monomeri polari e tra un monomero idrofobo e uno polare risulta rispettivamente debolmente attrattiva ( $C(-1, -1) = \frac{1}{2}$ ) e debolmente repulsiva ( $C(+1, -1) = -\frac{1}{2}$ ).

L'energia  $V$  è funzione solo degli angoli  $\theta_i$  ed è quindi invariante per traslazioni rigide dell'intera catena e per rotazioni dei monomeri attorno a un'asse perpendicolare al

piano. Essa è inoltre invariante per riflessioni ( $\theta_i \rightarrow -\theta_i$  per  $\forall i$ ).

Combinando il metodo Monte Carlo con un'ulteriore procedura di minimizzazione dell'energia, Stillinger *et al.* determinarono la configurazione dello stato fondamentale per tutte le sequenze composte da  $N \leq 7$  monomeri <sup>3</sup>. Per stabilire la conformazione fondamentale i termini del potenziale competono tra loro; per esempio il termine di “curvatura” tende a far allineare la catena, mentre i termini attrattivi di Lennard-Jones tendono a ripiegarla, avvicinando i monomeri. Per le varie sequenze cambia il bilanciamento di questi termini, dando luogo a un'ampia varietà nella conformazione degli stati fondamentali. Nel caso  $N = 6$ , per esempio, la configurazione dello stato fondamentale per la sequenza PPHHP è lineare mentre quella per la sequenza HHPHH ricorda la struttura del foglietto  $\beta$ . Nel caso della sequenza HHPHPP la forma della struttura dello stato fondamentale ricorda quella di un'elica, e inoltre la sequenza HHPHP fornisce un esempio di struttura globulare con nucleo idrofobico.

Lo stesso modello è stato ripreso successivamente da Irbäck *et al.* ([23], [24]). Questi autori, con metodi di Monte Carlo, hanno studiato tutte le possibili catene contenenti 8 e 10 monomeri a una temperatura fissata, tale che sia la sequenza composta da soli monomeri H che quella composta da soli monomeri P fossero in fase compatta. Essi trovarono che solo poche sequenze (dette buoni ripiegatori) sono caratterizzate, a quella temperatura, da un unico stato ripiegato. Determinando inoltre i livelli più bassi dello spettro di energia per alcune sequenze, notarono che la differenza di energia tra il minimo fondamentale e i minimi ad esso conformazionalmente più

---

<sup>3</sup>Data una catena composta da  $N$  monomeri il numero di sequenze distinguibili per il modello di Stillinger è uguale a  $2^{m-1}(2^m + 1)$ , se  $N$  è un numero pari con  $N = 2m$ , mentre è uguale a  $2^m(2^m + 1)$ , se  $N$  è un numero dispari con  $N = 2m + 1$ .

simili era molto maggiore per i buoni ripiegatori; questo è analogo a ciò che accade in natura infatti non tutte le sequenze dan luogo a delle proteine.

In un articolo successivo gli stessi autori hanno studiato 300 sequenze scelte in maniera casuale tra l'insieme di tutte le catene distinguibili contenenti 14 monomeri H e 6 monomeri P <sup>4</sup>. Di queste sequenze solo una piccola frazione (il 12 % ) sono state identificate come buoni ripiegatori.

## 2.3 Il paesaggio energetico

La cinetica del ripiegamento può essere descritta attraverso lo studio delle caratteristiche del paesaggio di energia libera del sistema. La traiettoria di un punto rappresentativo del sistema nello spazio delle coordinate può essere vista come un cammino sulla superficie multidimensionale dell'energia. La dinamica è fortemente influenzata dalla topografia del paesaggio: minimi locali, altezza delle barriere, bacini d'attrazione ecc.. Ciascuna valle nell'ipersuperficie descrive una particolare conformazione della proteina.

L'analisi di Levinthal ([1]), per esempio, assume che il paesaggio energetico sia piatto, con un solo minimo principale corrispondente allo stato nativo; la stima del tempo che occorre per raggiungere lo stato di minima energia in sistemi di questo tipo assume, in letteratura, la denominazione pittoresca di “problema del giocatore di golf bendato”.

L'energia di una proteina è data dalla somma di vari termini che competono tra loro come, per esempio, le interazioni tra gli aminoacidi e i vincoli topologici della catena.

---

<sup>4</sup>Tale insieme contiene 19440 sequenze.



Quando, durante il ripiegamento, differenti segmenti di un eteropolimero vengono in contatto, niente garantisce che le varie interazioni risultanti non si contrastino a vicenda nel tentativo di minimizzarsi. La proteina è dunque un sistema caratterizzato dalla presenza di frustrazione. Per questo gli stati corrispondenti ai minimi dell'energia sono connessi tra loro in maniera complessa e il paesaggio energetico delle proteine risulta frastagliato.

In recenti studi ([25], [26],[27]) si è supposto che il paesaggio energetico delle proteine, nonostante sia frastagliato e quindi presenti delle piccole “trappole” dove la proteina può risiedere a livello transiente, abbia una pendenza globale verso la configurazione nativa. In questa visione il paesaggio di una proteina naturale consiste in una collezione di cammini cineticamente convergenti verso lo stato nativo, corrispondente a uno stato termodinamicamente stabile. Una tale struttura del paesaggio è denominata a imbuto (*funnel*) poiché focalizza la varietà di strutture non ripiegate verso la struttura nativa.

Un paesaggio energetico frastagliato è caratterizzato da più valli corrispondenti ai minimi locali. L'intrappolamento transiente in queste valli rallenta il cammino verso la configurazione nativa più stabile. Se comunque la temperatura è minore di un certo valore (che indichiamo con  $T_g$ ), può avvenire che il sistema rimanga definitivamente intrappolato in una di queste valli non native, similmente a ciò che avviene nella transizione da stati liquidi a stati vetrosi, dove il sistema non raggiunge lo stato cristallino corrispondente a un minimo assoluto dell'energia. Quindi il minimo assoluto di energia per una proteina deve risultare termodinamicamente stabile alla temperatura di transizione vetrosa, al di sotto della quale non è più possibile

“scappare” dai minimi locali.

Nella figura (2.3) è mostrato “l’imbuto energetico” di un modello su reticolo di una catena di 27 monomeri con tre specie possibili di monomeri (la figura è stata ripresa da [26]). La figura mostra che il ripiegamento verso lo stato nativo è caratterizzato

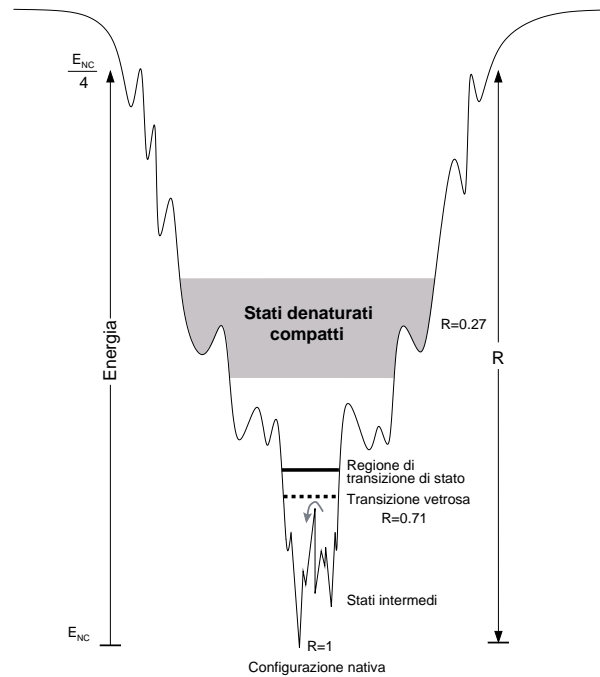


Figura 2.3. Forma schematizzata dell’imbuto energetico nel modello di Onuchic *et al.* ([26]).

da un collasso iniziale verso un ampio insieme di stati denaturati compatti, tali che la frazione  $R$  di contatti tra primi vicini nel reticolo, corrispondenti a quelli della configurazione nativa, è pari approssimativamente a un quarto. In seguito si ha un processo di riconfigurazione molto più lento, durante il quale il sistema si trova nella regione di transizione caratterizzata da stati cinetici intermedi geometricamente simili alla configurazione nativa ( $R \sim \frac{3}{4}$ ).

L'esistenza di questo imbuto per il ripiegamento dipende dalla sequenza della catena. Una sequenza il cui paesaggio di energia presenti un unico imbuto, che porta a un'unica conformazione stabile, corrisponde a una sequenza ripiegabile, anche detta "buon ripiegatore". Una sequenza che invece sia caratterizzata da più imbuti verso stati multipli di minima energia, corrispondenti a quelle configurazioni che abbiamo precedentemente denominato stati *gemisch*, corrisponde a un cattivo ripiegatore.

# Capitolo 3

## Dinamica di un modello 2D fuori reticolo

In questo capitolo introdurremo innanzitutto il modello semplificato di proteina oggetto degli studi originali condotti in questa tesi. In particolare specificheremo i tre tipi di sequenze scelte allo scopo di identificare le proprietà che consentono di distinguere un “buon ripiegatore” (cioè un buon candidato come modello di proteina) da una “cattivo ripiegatore”. A differenza di quanto viene fatto nella maggior parte degli studi riportati in letteratura, dove tali proprietà vengono identificate sulla base di metodi Monte Carlo (MC), qui ci siamo proposti di utilizzare un approccio dinamico canonico (in ogni caso, tutti i gradi di libertà dell’eteropolimero sono termalizzati tramite accoppiamento con un bagno termico, che fissa la temperatura del sistema). Questo approccio ci ha consentito di ottenere una serie di risultati interessanti. Abbiamo verificato che è possibile individuare in modo rapido ed efficace le medesime configurazioni native ottenute con i metodi MC. Si è anche potuto ottenere una

rappresentazione del paesaggio energetico delle sequenze, nonché la temperatura di *folding*, cioè la temperatura al di sotto della quale un buon ripiegatore mantiene la sua conformazione strutturale sostanzialmente inalterata rispetto a quella nativa. L'analisi di grandezze termodinamiche come energia interna e calore specifico, ci ha inoltre consentito di individuare ulteriori elementi distintivi di un “buon ripiegatore”. Infine, lo studio di opportune osservabili dinamiche ci ha permesso di caratterizzare l'evoluzione nel tempo di un buon ripiegatore verso la forma nativa.

### 3.1 Il modello

Il modello che consideriamo in questa tesi è una generalizzazione del modello di Stillinger *et al.* [21] introdotto nel paragrafo 2.2. Analogamente a questo modello si schematizza la proteina come una catena lineare nel piano composta da  $N$  monomeri che possono essere di tipo idrofobico (H) o polare (P) e che sono identificati dalle variabili binarie  $\xi_i$  con  $i = 1, \dots, N$ <sup>1</sup>. Invece di considerare legami rigidi fissando a 1 la lunghezza di legame, si è aggiunto nell'espressione dell'energia potenziale  $V$  dell'equazione (2.2) un terzo contributo  $V_a$  di tipo armonico tra primi vicini nella catena, scegliendo una forte costante di richiamo armonica  $\alpha$  e prendendo pari a 1 la lunghezza di riposo  $r_0$ . Questa modifica, oltre ad avvicinarsi maggiormente ad una descrizione realistica, ha il vantaggio di rendere più semplice lo studio della dinamica del modello, evitando di dover introdurre quei vincoli meccanici, dovuti alla rigidità dei legami assunta nel modello di Stillinger. Veniamo adesso alla definizione esplicita

---

<sup>1</sup>Come nel modello di Stillinger  $\xi_i = 1$  o  $-1$  a seconda che il monomero  $i$  sia di tipo idrofobico o polare.

del nostro modello.

Il grado di polimerizzazione  $N$  è stato preso uguale a 20. L'Hamiltoniana  $\mathcal{H}$  associata a una catena composta da  $N$  monomeri caratterizzati dalla sequenza  $\{\xi_i\}$  con  $i = 1, \dots, N$  è data da:

$$\mathcal{H} = \sum_{i=1}^N \frac{p_{x,i}^2 + p_{y,i}^2}{2} + \sum_{i=2}^{N-1} V_c(\theta_i) + \sum_{i=1}^{N-2} \sum_{j=i+2}^N V_l(r_{ij}, \xi_i, \xi_j) + \sum_{i=1}^{N-1} V_a(r_{i,i+1}); \quad (3.1)$$

dove la massa di ciascun monomero è assunta essere unitaria,  $(p_{x,i}, p_{y,i}) = (\dot{x}_i, \dot{y}_i)$  e  $r_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ , con  $(x_i, y_i)$  coordinate di posizione del monomero  $i$ . L'angolo  $\theta_i$  è l'angolo compreso tra i vettori di legame  $\vec{l}_{i-1}$  e  $\vec{l}_i$  (vedi figura (2.2)).

Il termine  $V_c(\theta_i)$  che compare nella (3.1) è dato da:

$$V_c(\theta_i) = \frac{1}{16} \times (1 - \cos(\theta_i)) \quad (3.2)$$

dove

$$\cos(\theta_i) = \frac{(x_i - x_{i-1})(x_{i+1} - x_i) + (y_i - y_{i-1})(y_{i+1} - y_i)}{r_{i,i-1}r_{i+1,i}}; \quad (3.3)$$

il termine  $V_l$ , legato alle interazioni tra monomeri non primi vicini nella catena, è pari a:

$$V_l = \left[ \frac{1}{r_{i,j}^{12}} - \frac{C(\xi_i, \xi_j)}{r_{i,j}^6} \right] \quad (3.4)$$

dove  $C(\xi_i, \xi_j)$  è dato dalla (2.4)<sup>2</sup>. Infine il terzo contributo all'energia potenziale  $V_a$  tende a tenere i monomeri contigui nella catena a una distanza pari a  $r_0$ ;

$$V_a(r_{i,i+1}) = \alpha(r_{i,i+1} - r_0)^2 \quad (3.5)$$

con  $\alpha = 20$  e  $r_0 = 1$ . Il valore di  $\alpha$  è tale da rendere il legame quasi rigido infatti

al tutte le temperature esaminate il valore di equilibrio del rapporto  $r_{i,i+1}/r_0$  oscilla

<sup>2</sup>I termini  $V_c$  e  $V_l$  sono pari a un quarto di quelli corrispondenti nel modello di Stillinger, questo risulta irrilevante perché ciò che conta ai fini dei risultati sono le intensità relative tra i termini dell'energia potenziale.

intorno a 1 al massimo dell'1–2%. Volendo fare una stima dei tempi caratteristici del sistema riportiamo le frequenze associate ai vari contributi dell'energia potenziale. I termini di energia che contribuiscono alle interazioni tra un singolo monomero e i propri monomeri contigui nella catena sono il termine armonico  $V_a$  e quello di “curvatura”  $V_c$ . Le frequenze  $\omega_a$  e  $\omega_c$  associate a tali termini intorno ai relativi punti di minimo risultano  $\omega_a = 12.65$  e  $\omega_c = 0.25$ . La frequenza associata al termine Lennard-Jones (nel caso  $C(\xi_i, \xi_j) = 1$ ) intorno alla distanza di equilibrio è invece dell'ordine  $\omega_l = 3.77$ .

L'energia potenziale totale  $V$  è dunque data da:

$$V = \frac{1}{16} \sum_{i=2}^{N-1} (1 - \cos(\theta_i)) + \sum_{i=1}^{N-2} \sum_{j=i+2}^N \left[ \frac{1}{r_{i,j}^{12}} - \frac{C(\xi_i, \xi_j)}{r_{i,j}^6} \right] + \alpha \sum_{i=1}^{N-1} (r_{i,i+1} - r_0)^2. \quad (3.6)$$

Essendo  $V$  dipendente solo dalla distanza reciproca dei monomeri della catena  $r_{ij}$  essa risulta invariante per traslazioni rigide della catena nel piano  $(x,y)$ , per rotazioni attorno a un asse perpendicolare al piano  $(x,y)$  e per riflessioni rispetto a un piano perpendicolare al piano  $(x,y)$ .

## 3.2 Dettagli sulle simulazioni

La dinamica a temperatura costante del modello presentato nel paragrafo 3.1 è stata studiata integrando numericamente le corrispondenti equazioni di Hamilton-Jacobi in presenza di un bagno termico. Questo è stato realizzato con uno schema standard di Nosè-Hoover [28] per ciascun monomero della catena (abbiamo verificato che tale algoritmo conserva la temperatura costante con una precisione di una parte

su 10'000). Più precisamente le equazioni per l'evoluzione delle componenti  $x_i$  e  $p_{x,i}$  di ciascun monomero  $i$  si esprimono come:

$$\begin{aligned} \dot{x}_i &= p_{x,i} \\ \dot{p}_{x,i} &= -\frac{\partial H}{\partial x_i} - \zeta_i p_{x,i} \\ \dot{\zeta}_i &= \frac{1}{\tau^2} \left( \frac{p_{x,i}^2 + p_{y,i}^2}{2T} - 1 \right) \end{aligned} \quad (3.7)$$

dove  $\zeta_i$  indica l'accoppiamento con il bagno mantenuto a una temperatura costante  $T$  e  $\tau$  è il tempo di reazione del bagno. Analoghe equazioni valgono anche per le componenti  $y_i$  e  $p_{y,i}$  con la stessa variabile di accoppiamento con il bagno  $\zeta_i$ . L'integrazione numerica delle equazioni del moto è stata eseguita adottando l'algoritmo d'integrazione di Runge-Kutta del secondo ordine [29] dove si è scelto come passo temporale d'integrazione  $\delta t = 0.025$ ; in questo modo  $\delta t$  risulta essere molto minore dei tempi caratteristici associati al nostro modello. Se infatti consideriamo il potenziale armonico dell'equazione (3.5) la frequenza caratteristica  $\omega$  sarà data da:  $\omega = \sqrt{\frac{8\alpha}{m}}$  il cui periodo associato,  $T_\omega$ , è pari a 0.5.

Sono state eseguite sia simulazioni di *unfolding* (SU) che simulazioni di *folding* (SF). Le SU sono state realizzate considerando come condizione iniziale la configurazione corrispondente al minimo assoluto dell'energia potenziale, che abbiamo assunto coincidere con la configurazione nativa (CN). Le medie delle varie quantità termodinamiche sono state eseguite, a temperatura fissata, durante un tempo  $\Delta t \sim 5 \cdot 10^5$ . Le SF sono state realizzate considerando come condizione iniziale una configurazione ottenuta mettendo i monomeri a una distanza  $r_0$  e scegliendo gli angoli  $\theta_i$  in maniera casuale nell'intervallo  $[-\pi/4; \pi/4]$ . Abbiamo inoltre controllato che la configurazione della catena, così ottenuta, non fosse cinematicamente inaccessibile (per esempio che



non ci fossero autointersezioni della catena). Il sistema è stato dunque lasciato rilassare per un certo tempo  $t_r$  ( $5 \cdot 10^5 \leq t_r \leq 10^7$ ) a seconda delle temperature esaminate. Le medie sono state eseguite, dopo il processo di rilassamento, durante un tempo  $\Delta t$  ( $5 \cdot 10^5 \leq \Delta t \leq 1.9 \cdot 10^6$ ). Per quanto riguarda le simulazioni di *folding* la media è stata inoltre eseguita su 10 condizioni iniziali diverse.

Durante le simulazioni sono state calcolate varie grandezze: l'energia totale  $E$ , l'energia potenziale totale  $V$ , l'energia interna  $U = \langle E \rangle$ , la temperatura media  $T$ :

$$T = \frac{1}{N} \left\langle \sum_i^N \frac{p_{x,i}^2 + p_{y,i}^2}{2} \right\rangle, \quad (3.8)$$

e il calore specifico a volume costante:

$$c_v = \frac{\langle E^2 \rangle - \langle E \rangle^2}{NT^2}. \quad (3.9)$$

La costante di Boltzmann è assunta uguale a 1 così come la massa dei singoli monomeri e il simbolo  $\langle \cdot \rangle$  indica la media nel tempo lungo una traiettoria nello spazio delle fasi.

Dovuto alle invarianze di  $V$  ogni configurazione è individuata a meno di traslazioni e rotazioni nel piano  $(x,y)$  nonché a meno di riflessioni rispetto a un piano perpendicolare al piano  $(x,y)$ . Perciò al fine di valutare la “distanza” tra due configurazioni  $h$  e  $k$  di una data catena è stata definita la seguente quantità che abbiamo denominato distanza angolare:

$$\delta_{h,k}(t) = \min \frac{1}{N-2} \sum_{i=2}^{N-1} |\theta_i^{(h)}(t) - \theta_i^{(k)}(t)|; \quad (3.10)$$

dove il minimo è calcolato considerando anche la configurazione ottenuta riflettendo la catena. Infatti non è necessario minimizzare  $\delta_{h,k}$  rispetto a traslazioni e rotazioni

della configurazione, poiché la distanza  $\delta_{h,k}$  così definita è invariante per tali trasformazioni. Generalmente abbiamo considerato la distanza angolare di una certa configurazione  $k$  rispetto alla configurazione nativa e l'abbiamo indicata con  $\delta_k^0$ .

Per distanza euclidea  $\Delta_{h,k}$  tra due configurazioni  $h$  e  $k$  si intende invece di aver ruotato, traslato e riflesso una delle due configurazioni in modo tale da averne minimizzato la distanza euclidea dall'altra;

$$\Delta_{h,k} = \min \sqrt{\sum_{i=1}^N (x_i^h - x_i^k)^2 + (y_i^h - y_i^k)^2}.$$

### 3.3 Le sequenze analizzate

Al fine di caratterizzare le proprietà relative al ripiegamento di questo modello semplificato, abbiamo analizzato in dettaglio le seguenti sequenze:

- $[S_0]$  Omopolimero; ovvero la sequenza dove tutti i residui sono idrofobici ( $\xi_i = -1, \forall i = 1, \dots, N$ );

- $[S_1]$  La sequenza composta da 14 monomeri H e 6 monomeri P e identificata in [23] con metodi Monte Carlo come un buon ripiegatore per il modello di Stillinger et al. [21] (in [23] la sequenza è classificata con il numero 81).

In dettaglio  $S_1 = \text{HHHPHHHPHHHPHPPHHPPHHH}$ .

- $[S_2]$  Un'altra sequenza, studiata in [23], sempre composta da 14 monomeri H e 6 monomeri P ma identificata come un cattivo ripiegatore (in [23] questa sequenza è stata classificata con il numero 50).

In dettaglio  $S_2 = \text{PHPHHHHHHHHPHHHHPPHPP}$ .

### 3.4 Le configurazioni native

La configurazione nativa di ciascuna sequenza è stata individuata nel seguente modo. Una volta fatto equilibrare il sistema ad una temperatura fissata abbastanza alta (tipicamente  $T=0.08$  in unità naturali del modello) in modo da evitare intrappolamenti in minimi locali dell'energia potenziale, abbiamo seguito la dinamica del sistema. A intervalli di tempo  $\Delta t = 5$  si identifica il minimo locale dell'energia, corrispondente al bacino d'attrazione che il sistema sta "sorvolando" in quel dato istante di tempo  $t$ . A tale scopo abbiamo preso le configurazioni dinamiche  $(x_i, y_i)$  al tempo  $t$  come condizioni iniziali (ottenute azzerando tutte le velocità  $(\dot{x}_i, \dot{y}_i)$ ) e le abbiamo fatte evolvere attraverso una dinamica sovrasmorzata regolata dalle seguenti equazioni del moto:

$$\dot{x}_i = -\frac{1}{\tilde{\gamma}} \frac{\partial H}{\partial x_i} \quad (3.11)$$

$$\dot{y}_i = -\frac{1}{\tilde{\gamma}} \frac{\partial H}{\partial y_i} \quad (3.12)$$

dove  $\tilde{\gamma}$  è la costante di smorzamento. Tale dinamica fa convergere il sistema verso il minimo locale corrispondente al bacino d'attrazione in cui si trova il sistema al tempo  $t$ . In questo modo possiamo associare a un qualunque stato dinamico del sistema un minimo locale dell'energia, che corrisponde al minimo che il sistema raggiunge una volta che gli viene applicata la dinamica sovrasmorzata dell'equazione (3.12). In questo senso dunque diremo che il sistema dinamico sta "visitando" un dato minimo. Se si segue la dinamica del sistema per tempi sufficientemente lunghi, si individua un gran numero di minimi locali del sistema e tra tutti i minimi si considera come configurazione nativa quello corrispondente allo stato di minima

energia <sup>3</sup>. Abbiamo verificato che, per le sequenze  $S_1$  e  $S_2$ , questo metodo è in grado di identificare non solo la configurazione nativa ma anche i 10 minimi di minore energia precedentemente determinati in [23] con procedure Monte Carlo assai raffinate.

Tipicamente, una simulazione della durata  $t = 2.5 \cdot 10^5$  risulta essere sufficiente a identificare la struttura nativa e i minimi corrispondenti alle energie più basse. Nella tabella 3.1 sono riportati il valore dell'energia potenziale  $V_0$  corrispondente allo stato nativo per ciascuna sequenza, insieme con quello del secondo minimo osservato con energia più bassa  $V_1$ . È abbastanza ovvio che la configurazione nativa dell'omopolimero abbia l'energia più bassa di tutte, essendo composta da un numero più elevato di monomeri H. Infine nella figura (3.1) riportiamo la struttura della CN

Tabella 3.1. L'energia potenziale minima  $V_0$  (corrispondente alla CN) è qui riportata insieme a quella del minimo successivo  $V_1$ , per le sequenze  $S_0$ ,  $S_1$ ,  $S_2$ .

	$S_0$	$S_1$	$S_2$
$V_0$	-7.04226	-4.66668	-4.62830
$V_1$	-7.01674	-4.57445	-4.62582

per ciascuna sequenza. Notiamo che per la sequenza  $S_1$  è più evidente la formazione del “nucleo idrofobico” e il suo “schermaggio” da parte delle molecole polari.

### 3.5 Caratterizzazione delle valli native

Per studiare le caratteristiche del paesaggio energetico nella regione intorno al minimo assoluto dell'energia potenziale (valle nativa), abbiamo eseguito, per ciascuna

<sup>3</sup>I livelli energetici sono stati determinati con una risoluzione pari a  $10^{-6}$ .

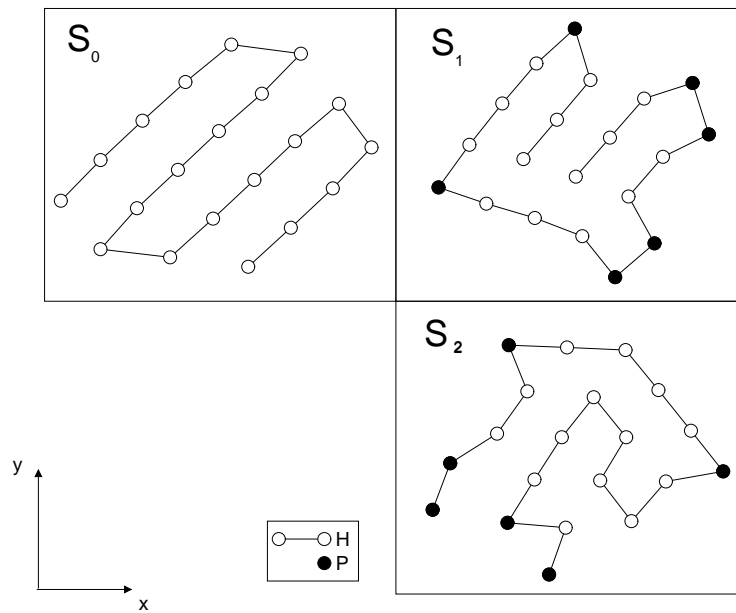


Figura 3.1. Struttura delle configurazioni native per le sequenze  $S_0$ ,  $S_1$  e  $S_2$

sequenza, simulazioni di *unfolding* a varie temperature  $T$ . Per ogni temperatura  $T$  abbiamo considerato una singola simulazione della durata di  $t = 2.5 \cdot 10^5$  e per ogni  $\Delta t = 5$  abbiamo proceduto alla identificazione dei minimi “visitati” come descritto nel paragrafo 3.2. In questo modo il numero di minimi distinti che è possibile identificare è al più pari a 50'000.

Una caratteristica osservata per tutte le sequenze è che al decrescere di  $T$  il numero di minimi “visitati” decresce notevolmente. Questo è dovuto al fatto che, abbassando la temperatura, la dinamica del sistema ha sempre meno tendenza ad allontanarsi dai minimi di energia più bassa e in particolare dalla CN. Si osserva comunque una certa differenza di comportamento tra le tre sequenze considerate: per  $S_0$  sono stati identificati 42'249 minimi differenti a  $T = 0.11$ , mentre a  $T = 0.04$  solo 118; per  $S_1$  abbiamo che a  $T = 0.11$  vengono visitati 20'290 minimi distinti, mentre per  $T = 0.04$  solo 48. In pratica il numero dei minimi visitati dall'omopolimero è molto maggiore

e alle temperature più alte considerate ( $T \sim 0.15$ ) raggiunge il valore massimo di 50'000, mentre per le sequenze eteropolimeriche è dell'ordine di 30'000. Questo indica che nemmeno ad alte temperature gli eteropolimeri visitano in maniera casuale lo spazio delle configurazioni (vedi figura (3.2)).

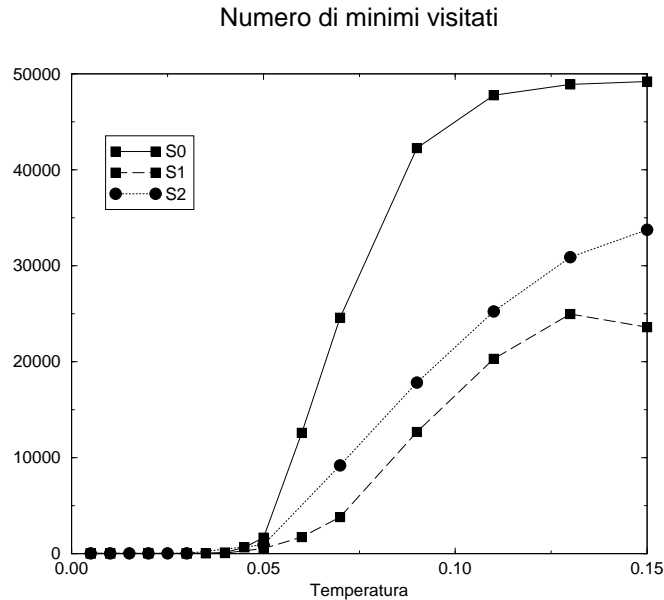


Figura 3.2. Numero di minimi visitati nelle SU a varie temperature  $T$

### 3.6 Determinazione della temperatura $T_f$

La definizione sperimentale della temperatura di *folding*  $T_f$  di una proteina è la seguente [3]: considerando un insieme di proteine immerse in un solvente, si definisce temperatura di *folding* quella per la quale il numero di proteine che si trovano nello stato nativo è pari alla metà del numero totale. In una simulazione dinamica di un modello su reticolo, la definizione più simile per  $T_f$  è la seguente:  $T_f$  è la temperatura alla quale la probabilità che il sistema sia nella configurazione nativa è pari a  $\frac{1}{2}$ . In un modello definito nello spazio continuo occorre dare una definizione diversa

di tale probabilità; infatti la CN è, nello spazio continuo, un punto a dimensione nulla. In questa tesi abbiamo proceduto nel seguente modo: abbiamo introdotto la probabilità  $P(T)$  che il sistema, alla temperatura  $T$  stia “visitando” (nel senso spiegato nel paragrafo 3.4), la CN o uno dei minimi che ne risultano “direttamente connessi”<sup>4</sup> alla CN. L’andamento di  $P(T)$  è mostrato in figura (3.3). La definizione

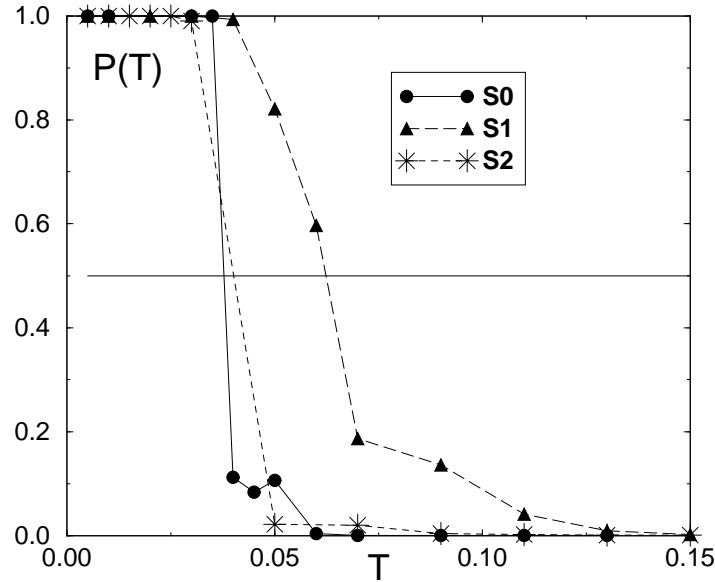


Figura 3.3. Probabilità  $P(T)$  in funzione della temperatura  $T$ . Le misure sono state eseguite durante SU della durata di  $t = 2.5 \cdot 10^5$ , dove la procedura di rilassamento sovrasmorzato è applicata ogni  $\Delta t = 5$

di  $T_f$  è la seguente:

$$P(T_f) = \frac{1}{2};$$

e i corrispondenti valori per ciascuna sequenza sono riportati nella tabella 3.2. Si osserva che per  $S_1$  il valore di  $T_f$  è molto più alto che per  $S_0$  e  $S_2$ .

<sup>4</sup>Nel cap.4 verrà data una definizione rigorosa di tali minimi

### 3.7 Proprietà di equilibrio

Durante le simulazioni dinamiche, una volta raggiunto l'equilibrio, abbiamo calcolato la media temporale di varie grandezze fisiche. Citiamo per esempio l'energia totale  $E$  ( $\langle E \rangle = U$ ), il raggio di girazione  $R_g$  e la distanza angolare rispetto alla configurazione nativa  $\delta^0$ .

Abbiamo eseguito sia SU che SF e un aspetto comune a tutte le medie di equilibrio è che i risultati ottenuti tramite le SU e le SF sono coincidenti solo al di sopra di una certa temperatura, che definiamo temperatura di transizione vetrosa  $T_g$ . Alle temperature al di sotto di  $T_g$  si ha un arresto dinamico e il sistema viene intrappolato in minimi locali, non raggiungendo mai la configurazione di minimo assoluto dell'energia (entro i nostri limiti massimi di tempo d'integrazione  $t \simeq 1.5 \cdot 10^7$ ). Nella tabella 3.2 sono riportati i valori di  $T_g$  per ciascuna sequenza, ottenuti confrontando le curve  $T(U)$  (vedi per esempio la figura (3.4)) ricavate durante le SU e le SF a varie temperature  $T$ . Vediamo che per  $S_1$  è maggiore la differenza tra  $T_f$  e  $T_g$ ; questa è un'indicazione che  $S_1$  si può ripiegare più facilmente. Infatti alla temperatura  $T_f$  il sistema è ancora lontano dalla fase di rallentamento dinamico.

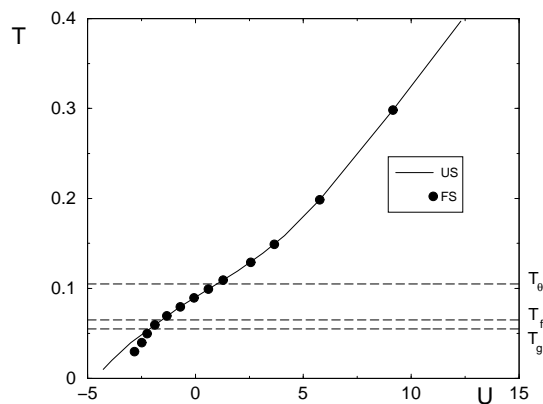


Figura 3.4. Dati di equilibrio per  $T = T(U)$  relativi alla sequenza  $S_1$ . Le barre di errore sono minori della dimensione dei simboli.



Abbiamo studiato inoltre l'andamento rispetto alla temperatura, del calore specifico a volume costante  $c_v$  (vedi figura (3.5)). Tutte le sequenze osservate mostrano un massimo a una certa temperatura che definiamo temperatura  $T_\theta$  in analogia con i risultati sperimentali (vedi per esempio la figura (1.4)) che mostrano un picco nell'andamento del calore specifico in corrispondenza di una temperatura che viene chiamata temperatura  $\theta$ . Riportiamo nella tabella 3.2 il valore di  $T_\theta$  per ciascuna sequenza. La temperatura  $\theta$  è la temperatura di transizione da stato *coil* a stato compatto che si osserva sia negli omopolimeri che negli eteropolimeri.

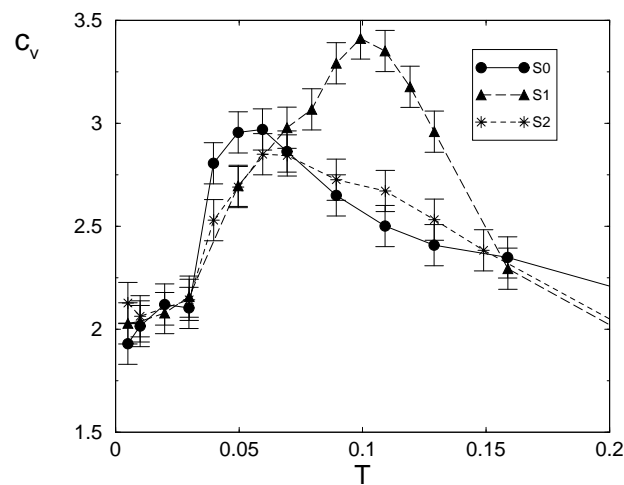


Figura 3.5. Calore specifico  $c_v$  in funzione di  $T$ ; i dati si riferiscono alle simulazioni SU.

La sequenza  $S_1$  presenta il valore più alto del massimo di  $c_v$  e della temperatura corrispondente  $T_\theta$  e questo vuol dire che la sequenza  $S_1$  raggiunge la fase compatta già a valori più alti della temperatura.

Consideriamo adesso la media della distanza angolare  $\langle \delta^0 \rangle$  rispetto alla CN. I risultati relativi alle SU e alle SF per le tre sequenze analizzate sono riportati nelle figure (3.6) e (3.7). Per temperature sufficientemente alte ( $T > T_g$ ) i risultati essenzialmente coincidono. Come caratteristica tipica osserviamo un incremento

Tabella 3.2. Temperatura di *folding*  $T_f$ , temperatura  $T_g$ , temperatura  $T_\theta$  e fattore di ripiegamento  $Q$ .

	$S_0$	$S_1$	$S_2$
$T_f$	0.036	0.065	0.036
$T_g$	0.030	0.055	0.035
$T_\theta$	0.050	0.105	0.060
$Q$	0.270	0.620	0.270

di  $\langle \delta^0 \rangle$  con la temperatura, fino a un valore asintotico di saturazione  $\delta_a$  (per  $T > 0.20$ ) dipendente dalla sequenza. La dipendenza di  $\langle \delta^0 \rangle$  dalla temperatura, ottenuta nelle SU può dare un'idea della forma della valle nativa. Si osserva che  $\langle \delta^0 \rangle$  mostra, nel caso della sequenza  $S_0$ , un incremento molto più brusco che nel caso delle sequenze  $S_1$  e  $S_2$ . Se durante la dinamica di ripiegamento il sistema non è capace di entrare nella valle nativa a nessuna delle temperature esaminate questa è una chiara indicazione che la sequenza esaminata è un cattivo ripiegatore. Un'indicazione dell'accessibilità della valle nativa può essere il valore di  $\langle \delta^0 \rangle$  più basso per cui i risultati delle SU e nelle SF siano ancora coincidenti.

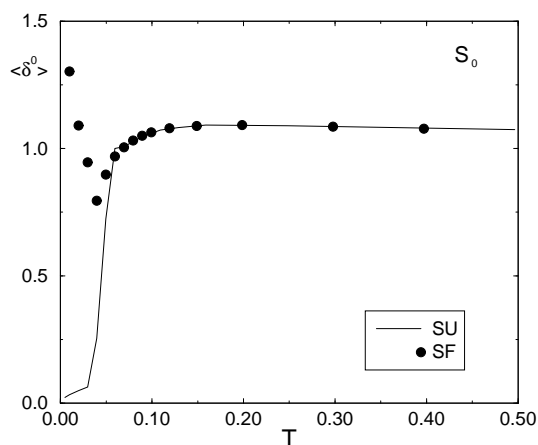


Figura 3.6. Distanza angolare media  $\langle \delta^0 \rangle$  per  $S_0$ .

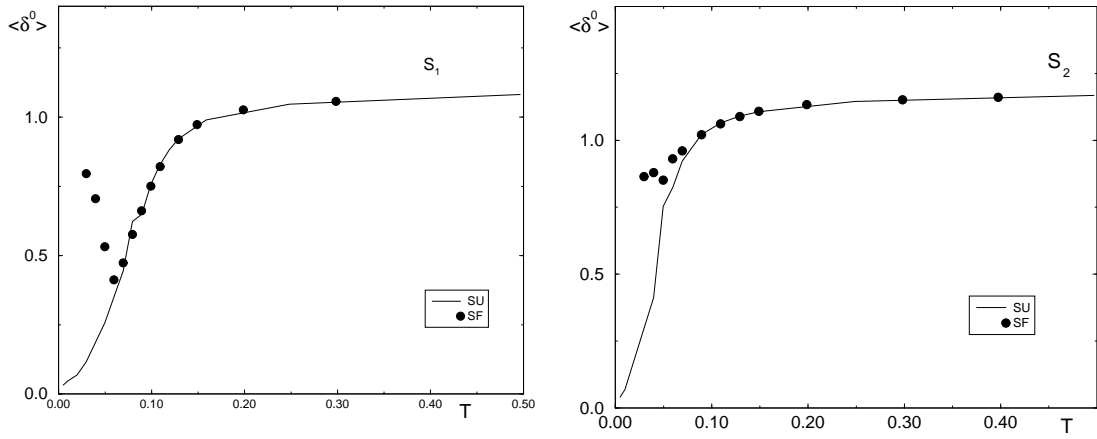


Figura 3.7. Distanza angolare media  $\langle \delta^0 \rangle$  per  $S_1$  e  $S_2$

Per confrontare le differenti sequenze a livello più quantitativo abbiamo introdotto il seguente fattore di ripiegamento  $Q$ :

$$Q = \frac{\delta_a - \delta_m}{\delta_a}$$

dove  $\delta_m$  è il minimo valore raggiunto da  $\langle \delta^0 \rangle$  tra le varie SF eseguite a diverse temperature. La temperatura della SF a cui viene raggiunto il valore minimo  $\delta_m$  è in generale differente per ciascuna sequenza. Un alto valore del fattore  $Q$  indica che il sistema è capace di avvicinarsi alla configurazione nativa prima che il sistema si arresti dinamicamente a  $T = T_g$ ; se  $Q$  è relativamente basso questo suggerisce che il sistema non raggiunge la valle nativa prima che la dinamica sia notevolmente rallentata dalla transizione vetrosa. Come riportato nella tabella 3.2 il valore di  $Q$  per  $S_1$  è molto maggiore di quelli di  $S_0$  e  $S_2$ , quindi, in accordo con i risultati in [23],  $S_1$  può essere identificato come un buon ripiegatore mentre  $S_0$  e  $S_2$  come cattivi ripiegatori.

### 3.8 Comportamento dinamico alla temperatura $T_f$

Partendo da condizioni iniziali casuali, come descritto nel paragrafo 3.2, abbiamo seguito la dinamica di *folding* a temperatura costante, per una data sequenza e per vari valori della temperatura.

Per la sequenza  $S_1$ , che abbiamo identificato come un buon ripiegatore, osserviamo tipicamente che (per  $T_g < T < T_f$ ) la catena raggiunge sempre un certo stato di equilibrio, caratterizzato da una distanza angolare  $\bar{\delta}^0$ . Abbassando ulteriormente la temperatura da  $T_f$  a  $T_g$ , dopo questa fase di rilassamento, si ha che  $\delta^0$  decresce gradualmente fino a 0. Inoltre, la catena raggiunge questo stato di equilibrio attraverso salti ben definiti e quasi istantanei, che connettono una serie di stati metastabili. (vedi figura (3.8)).

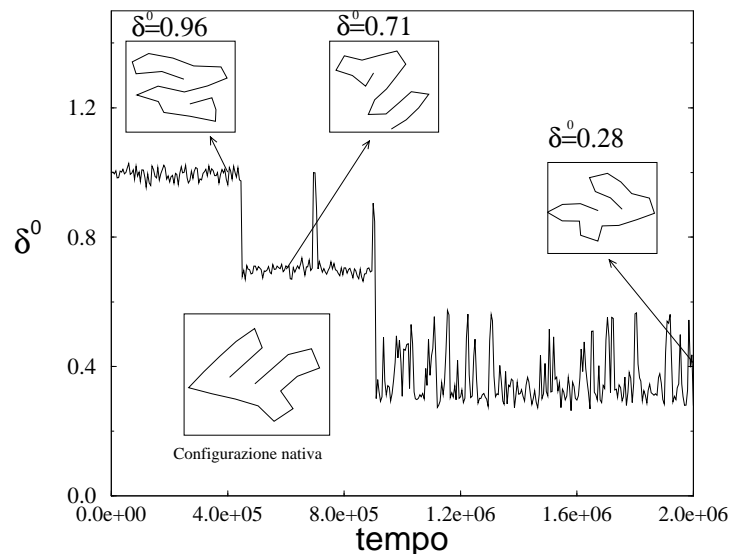


Figura 3.8. Distanza angolare in funzione del tempo. Con condizioni iniziali casuali la dinamica è stata integrata a temperatura costante  $T = 0.06$ . La dinamica di ripiegamento è chiaramente visibile.

Si osserva che la decrescita della distanza angolare non è sempre associata a un corrispondente abbassamento dell'energia potenziale (vedi per esempio il primo salto

nelle figure (3.8) e (3.9)). Solo quei salti che portano verso la configurazione nativa sono associati a una notevole decrescita dell'energia potenziale (in particolare del contributo Lennard-Jones totale  $V_{LJ}$ ) come mostrato in figura (3.9). Notiamo inoltre

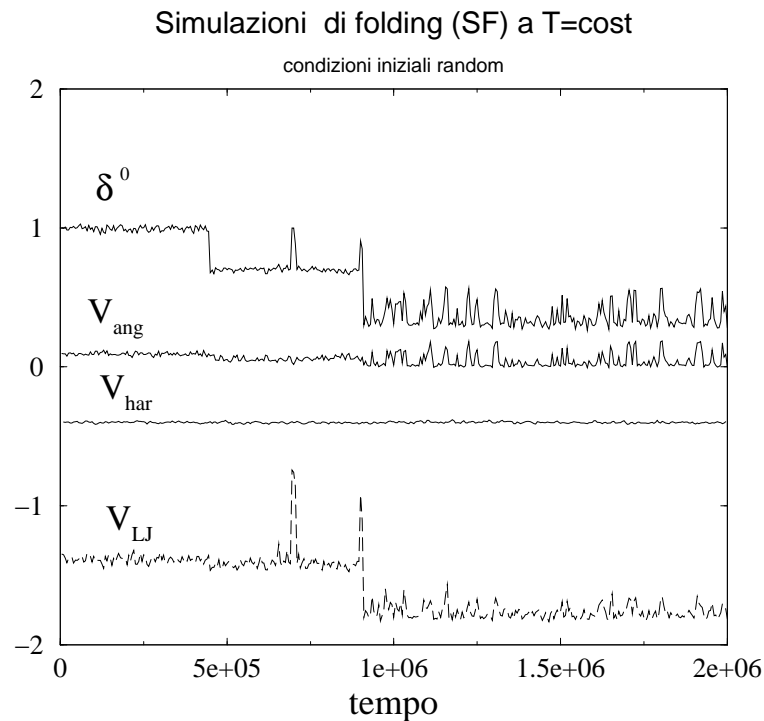


Figura 3.9. Andamento della distanza angolare e dei termini di energia potenziale in funzione del tempo durante la fase di ripiegamento del sistema. I vari termini di energia sono stati traslati di una quantità arbitraria per rendere maggiormente chiara la figura.

che, una volta che viene raggiunta la valle nativa, la sequenza  $S_1$  vi rimane in maniera stabile; infatti non vengono più registrati, entro i tempi d'integrazione eseguiti, salti verso stati conformazionali caratterizzati da valori di  $\delta^0$  molto più alti ( $\delta^0 > 0.7$ ). Il tempo di ripiegamento totale è risultato essere dell'ordine di  $10^6 - 10^7$  unità naturali che in unità reali corrispondono a circa  $10^{-5}$  sec (per la conversione tra le unità di naturali e quelle reali si rimanda all'appendice A). Il tempo di ripiegamento è dunque risultato essere molto minore dei tempi tipici osservati sperimentalmente,

ma vi sono risultati recenti riguardo a tempi di ripiegamento di moduli di proteine dell'ordine dei  $\mu$ -sec [30]. La differenza può essere dovuta anche alla dimensionalità del sistema studiato: nel piano si hanno vincoli geometrici più forti che in 3D e ciò da sì che l'eteropolimero converga in modo più rapido verso la CN.

Per i cattivi ripiegatori, dove tipicamente  $T_f \sim T_g$ , non abbiamo osservato, entro i tempi d'integrazione considerati, un ripiegamento definitivo verso la configurazione nativa. Inoltre non si osserva un andamento sostanzialmente decrescente della distanza angolare  $\delta^0$  e dei termini di energia potenziale. Avviene che il sistema si avvicina alla configurazione nativa, ma ci rimane sempre per un breve tempo. Risultano sufficienti piccole fluttuazioni termiche per riportare il sistema fuori dalla valle nativa; di conseguenza l'energia  $E$  subisce notevoli oscillazioni. Gli andamenti della distanza angolare  $\delta^0$  e dell'energia totale  $E$  per la sequenza  $S_2$  sono mostrate in figura (3.10).

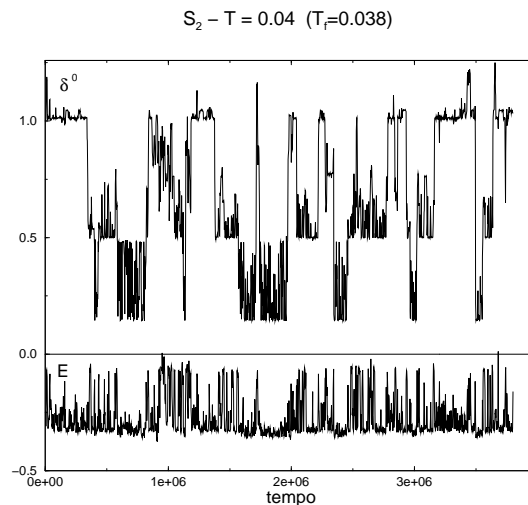


Figura 3.10. Andamento dell'energia totale  $E$  e della distanza angolare rispetto alla CN in funzione del tempo  $t$  per la sequenza  $S_1$

# Capitolo 4

## Descrizione a grana grossa della dinamica di ripiegamento

Per esporre con maggiore chiarezza i contenuti di questo capitolo il preambolo sarà un pò più articolato e conterà di qualche pagina.

L'energia potenziale associata ad un modello di eteropolimero è tipicamente caratterizzata da un altissimo numero di minimi. Sembra quindi una buona idea rappresentare a *grana grossa* la dinamica di *fold*ing e *un*fold*ing* di una proteina come una successione di salti da un minimo del sistema ad un'altro. Per completare la descrizione in termini statistici di questa dinamica sarà altresì necessario associare una probabilità di transizione da un minimo ad un altro direttamente connesso <sup>1</sup>.

Tale probabilità è in definitiva legata al profilo di energia libera del sistema, in quanto dipende non solo dall'altezza della barriera energetica che separa i due minimi

---

<sup>1</sup>Il concetto di minimi direttamente connessi in uno spazio a  $N > 3$  dimensioni non è di immediata comprensione in termini geometrici. Consideriamo comunque due minimi che siano separati da un'unica barriera, rimandando nel corso della tesi a una definizione più rigorosa.

direttamente connessi ma anche da proprietà entropiche del sistema, quali la sua configurazione prima e durante il salto. In particolare, i minimi fra loro contigui saranno separati da barriere di energia potenziale la cui altezza sarà uno degli elementi che determinano la probabilità di transizione da un minimo ad un altro.

Ad ogni istante la proteina può essere rappresentata come un punto nello spazio delle fasi che può essere associato ad un *bacino di attrazione* di un dato minimo. Il bacino di attrazione soggiacente la dinamica della proteina ad un dato istante può essere identificato congelando il sistema a quell'istante (ponendo quindi la parte cinetica dell'energia a zero) e lasciando successivamente rilassare il sistema lungo la direzione di massima pendenza del potenziale. Questa tecnica è stata ampiamente usata per lo studio di vetri, *cluster* atomici e liquidi molecolari [31]. Una rappresentazione di questo tipo ci consente di ottenere una partizione dello spazio delle fasi, i cui elementi saranno i minimi del potenziale (o equivalentemente i rispettivi bacini di attrazione).

Per meglio chiarire questi concetti, consideriamo come esempio un semplice ma non banale profilo bidimensionale di energia potenziale: il potenziale a “scatola d'uova” (riportato in figura (4.1) in una sua rappresentazione tridimensionale). Questo potenziale mostra una ripetizione periodica di una data cella elementare. La cella elementare presenta un minimo, quattro punti sella e quattro massimi, data la conformazione di questo potenziale la cella elementare coincide col bacino di attrazione del minimo considerato. Se ad ogni punto estrema del potenziale associamo



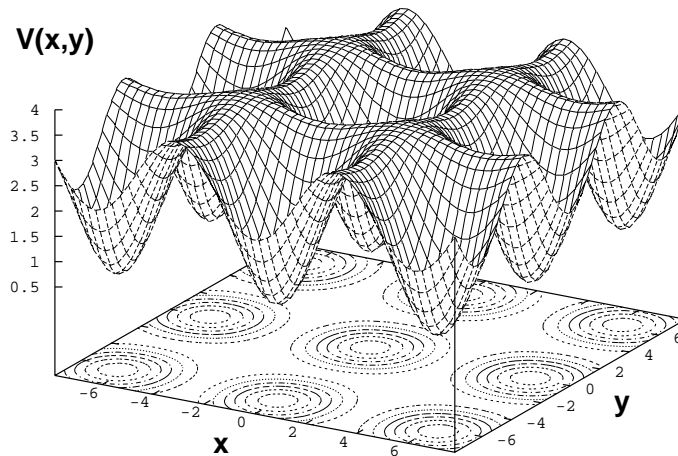


Figura 4.1. Rappresentazione tridimensionale del potenziale a scatola d'uova.  
 $V(x,y) = 3 - \cos x - \cos y - \cos x \cos y.$

le corrispondenti *varietà* stabili e instabili <sup>2</sup> è evidente dalla figura (4.1) che il bacino di attrazione di un dato minimo è delimitato dalle varietà stabili dei punti sella che lo circondano; si noti che questo risultato non dipende dalla dimensionalità del problema in esame.

Come si è accennato in precedenza, se vogliamo dare una descrizione in termini statistici della dinamica di una particella che si muove in un campo di potenziale come quello a “scatola d'uova”, è necessario poter associare a ogni transizione da un minimo ad un altro la corrispondente probabilità di salto. Questa dipenderà

---

<sup>2</sup>Per varietà di un dato punto estremale si intende il luogo dei punti invarianti sotto la dinamica del sistema. Le varietà instabili sono il luogo di accumulazione dei punti del piano a cui si applichino iterazioni successive della dinamica, le varietà stabili altresì sono il luogo ove si accumulano i punti del piano sottoposti a iterazioni all'indietro nel tempo della dinamica. Per una trattazione dettagliata vedi [32]

certamente dalla temperatura  $T$  a cui è posto il sistema, ma in larga parte dipenderà anche dalla forma del potenziale in prossimità dei minimi coinvolti nel salto e in prossimità del punto sella che divide i due minimi, nonché dall'altezza  $W$  della barriera che li separa, come vedremo nel paragrafo 4.5. In prima approssimazione, per un processo di attivazione termica, tale probabilità di transizione sarà proporzionale a  $e^{-W/T}$  (assumendo unitaria la costante di Boltzmann). Dato il carattere esponenziale della legge riportata, è evidente che i salti più probabili sono quelli per cui la particella passa per il punto sella associato alla minima barriera di potenziale che divide i due minimi. Sarà quindi essenziale mettere a punto una tecnica per identificare il punto (o i punti) sella che separano due minimi ed in particolare quello associato alla barriera di potenziale più bassa. Facendo ancora riferimento al potenziale a “scatola d'uova”, rappresentato in due dimensioni nella figura (4.2), descriviamo come procedere per mettere a punto gli elementi di base di tale tecnica che ci saranno utili in seguito.

È chiaro che se congiungiamo due minimi  $\mathbf{a}_1$  e  $\mathbf{a}_2$  con una linea generica adagiata lungo il profilo di energia potenziale questa in generale passerà da punti con energia maggiore di quella del punto sella  $\mathbf{s}$  (vedi figura (4.2)). Se poi, con una opportuna dinamica di rilassamento, facciamo evolvere tutta la linea di punti in modo che segua la direzione di massima pendenza del potenziale, è abbastanza pacifico che questa linea coinciderà asintoticamente con la varietà instabile associata al punto sella  $\mathbf{s}$  fra i due minimi.

Riassumendo, nel seguito di questo capitolo prima identificheremo i minimi che risultano più prossimi alla CN con un approccio di tipo dinamico e poi, con la tecnica

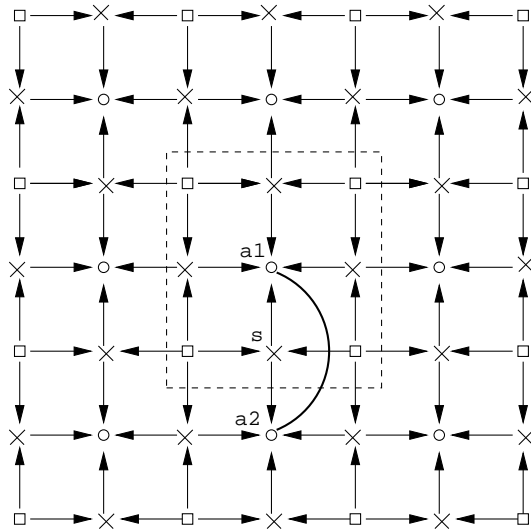


Figura 4.2. Schematizzazione dei minimi  $\circ$ , dei massimi  $\square$  e delle selle  $\times$  di un potenziale a scatola d'uova. Le linee delimitate dalle frecce indicano le varietà stabili ( $\rightarrow\leftarrow$ ) e instabili ( $\leftarrow\rightarrow$ ) della funzione energia potenziale.

appena descritta, identificheremo le barriere di potenziale che li separano nonché i punti sella interposti tra loro. Queste informazioni ci consentiranno, nel paragrafo 4.6, di dare una stima teorica della probabilità di transizione da un minimo all'altro e, in seguito, di sintetizzare i principali cammini che la dinamica di *unfolding* di una proteina può seguire in termini di un grafo che connette fra loro i minimi accessibili a partire dalla CN.

## 4.1 Identificazione dei minimi “vicini” alla configurazione nativa

Per studiare le caratteristiche dell'energia potenziale intorno alla configurazione nativa abbiamo individuato innanzitutto i minimi del sistema che risultano essere “vicini” alla CN, cioè quei minimi che vengono visitati per primi, una volta che il

sistema si allontani dal bacino di attrazione associato alla CN. La ricerca di tali minimi è stata realizzata, per ogni sequenza, seguendo varie evoluzioni dinamiche del sistema, quando questo lascia il bacino corrispondente alla CN e raggiunge quello corrispondente a un altro minimo. Abbiamo dunque proceduto nella seguente maniera: si considera come configurazione iniziale quella nativa e si riscalda lentamente il sistema portandolo a una data temperatura  $T > T_g$ . Successivamente si fa evolvere il sistema alla temperatura  $T$  e ad intervalli di tempo  $\Delta t$  fissati si controlla, tramite la tecnica descritta nel paragrafo 3.4, se il sistema sta ancora “visitando” la CN o se è passato a “visitare” un altro minimo. Se il sistema sta “visitando” ancora la configurazione nativa, si prosegue l’evoluzione dinamica del sistema per un altro intervallo di tempo  $\Delta t$ , altrimenti, se risulta che sta “visitando” un altro minimo, si arresta l’evoluzione e si identifica il nuovo minimo come minimo “vicino” della CN. Si registra inoltre ogni volta il tempo di permanenza  $t_p$  durante il quale il sistema “visita” la configurazione nativa, prima di passare nel bacino d’attrazione di un altro minimo:  $t_p$  sarà dato da  $n \cdot \Delta t$ , dove  $n$  è il numero di controlli successivi in cui il sistema si viene a trovare sempre nella CN.

L’intervallo di tempo  $\Delta t$  deve essere sufficientemente piccolo da poter individuare efficacemente quando il sistema passa nel bacino di un altro minimo<sup>3</sup>; quindi  $\Delta t$  dovrà risultare abbastanza minore di  $t_p$ . Nello stesso tempo applicare la dinamica sovrasmorzata troppo frequentemente rallenta inutilmente la durata del calcolo.

Abbiamo stimato che  $\Delta t = 1$  fosse un buon compromesso. Abbiamo poi ripetuto

---

<sup>3</sup>Con un passo  $\Delta t$  troppo grande c’è il rischio che il sistema non solo esca dal bacino nativo ma con una serie di salti successivi giunga nel bacino di un minimo non direttamente connesso alla CN. Perciò in generale non è detto che i minimi “vicini” siano sempre direttamente connessi alla CN.

questa procedura per  $N_v$  volte, in modo tale da avere una statistica sufficiente a dare una buona stima della frequenza con cui ciascun minimo “vicino” viene visitato.  $N_v$  dovrà essere molto maggiore del numero totale di minimi distinti che possono essere individuati (abbiamo tipicamente considerato  $N_v = 4000$ ). Abbiamo infine calcolato il tempo di permanenza medio  $\langle t_p \rangle$  nel minimo nativo prima di un generico salto, dove la media è eseguita sulle 4000 prove.

Per le tre sequenze introdotte nel paragrafo 3.3, sono stati considerati più valori della temperatura intorno alle temperature  $T_g$  e  $T_f$ . Come mostrato nella tabella 4.1 il tempo medio di permanenza  $\langle t_p \rangle$  nella configurazione nativa diminuisce al crescere della temperatura, mentre il numero  $n_t$  di minimi distinti visitati aumenta. Questo si spiega col fatto che a temperature molto basse il sistema può visitare

Tabella 4.1. In questa tabella sono riportate per ciascuna temperatura  $T$  il numero di minimi distinti  $n_t$  in cui passa per primo il sistema partendo dalla CN. Inoltre è riportato il tempo di permanenza medio nella CN insieme con la propria deviazione standard. (\*)Per il minimo  $S_0$  a temperatura  $T=0.03$  si sono fatte solo  $N_v = 500$  prove.

$S_0$			$S_1$			$S_2$		
$T$	$\langle t_p \rangle$	$n_t$	$T$	$\langle t_p \rangle$	$n_t$	$T$	$\langle t_p \rangle$	$n_t$
0.03*	$(32 \pm 1) \cdot 10^3$	10	0.04	$1600 \pm 20$	44	0.03	$9.6 \pm 0.1$	2
0.04	$1960 \pm 30$	27	0.05	$188 \pm 3$	57	0.04	$8.0 \pm 0.1$	5
0.05	$300 \pm 5$	47	0.06	$48 \pm 1$	65	0.05	$7.1 \pm 0.1$	20
0.06	$79 \pm 1$	82	0.07	$20.7 \pm 0.3$	89	0.06	$6.3 \pm 0.1$	38

un numero limitato di minimi; in quanto la probabilità di visitare i minimi corrispondenti alle barriere energetiche più alte è praticamente nulla. Man mano che la temperatura aumenta, il sistema visita più minimi diversi tra loro ma è sempre più probabile che i bacini di alcuni di tali minimi non siano direttamente connessi a quello della CN. Comunque dalla tabella 4.1 è evidente che a  $T \sim T_f$  la sequenza

$S_1$  è quella la cui CN è connessa al massimo numero di vicini mentre  $S_2$  è connessa a un numero assai limitato di vicini. Si nota inoltre come a tutte le temperature esaminate,  $\langle t_p \rangle$  per  $S_2$  resti quasi costante. Come spiegheremo in seguito questo è dovuto al fatto che la CN della sequenza  $S_2$  è connessa tramite una barriera assai bassa a un minimo, con il quale forma quasi un unico bacino d'attrazione.

## 4.2 Ricerca dei punti sella

Per una descrizione a grana grossa della dinamica del nostro sistema come una successione di salti tra minimi è necessario conoscere la posizione relativa nello spazio delle coordinate dei vari bacini di attrazione dell'energia potenziale e l'altezza delle barriere energetiche che li separano. Consideriamo per esempio due minimi dell'energia potenziale e i possibili cammini sull'ipersuperficie dell'energia che uniscono questi minimi. In generale, in un sistema multidimensionale, esistono infiniti cammini che connettono un minimo a un altro; il nostro scopo è di individuare tra tutti i cammini quello *principale*, cioè quello che corrisponde al valore minimo della barriera di energia che è necessario superare. Tale cammino coinciderà con una o più varietà instabili del sistema dinamico associate ai punti sella dell'energia che intercorrono tra questi minimi. Se il cammino principale passa da un unico punto sella e quindi non è necessario passare da ulteriori minimi per raggiungere un minimo dall'altro, i due minimi sono connessi direttamente, altrimenti, se il cammino principale passa per più minimi, sarà necessario passare per più punti sella per connettere i due minimi.

In questa tesi abbiamo messo a punto una nuova tecnica attraverso la quale, dati due

minimi, si è in grado di individuare il cammino principale che connette tali minimi. Illustriamo adesso quali sono le idee principali su cui si basa questa tecnica, rimanendo all'appendice B per una descrizione dettagliata dell'algoritmo.

L'energia potenziale  $V$  del nostro modello è funzione delle coordinate spaziali  $q_x(i)$  e  $q_y(i)$  (con  $i = 1 \dots N$ ) degli  $N = 20$  monomeri costituenti il sistema. In pratica un cammino tra due minimi è rappresentabile come una successione di punti in uno spazio 40-dimensionale, dove ogni punto corrisponde ad una configurazione del sistema. Se immaginiamo di poter modificare un qualsiasi cammino nello spazio delle coordinate, in modo tale che non perda le proprietà di continuità e che le configurazioni che lo costituiscono assumano valori dell'energia via via più bassi, si può pensare che tale cammino converga verso il cammino principale, in quanto corrisponde a varietà instabili del sistema. Un modo di modificare una configurazione riducendone l'energia è chiaramente applicare la dinamica sovrasmorzata. A livello pratico insorgono però delle difficoltà; infatti nei calcoli numerici non è possibile rappresentare l'intero cammino con continuità, ma è necessario rappresentarlo tramite un numero finito di configurazioni  $M$ . Se consideriamo queste  $M$  configurazioni e le sottoponiamo a una dinamica sovrasmorzata per un tempo breve (molto minore del tempo di convergenza verso i minimi), avremo che quelle configurazioni che inizialmente si trovavano vicino a punti di minimo o a punti sella dell'energia avranno subito cambiamenti dell'energia molto meno pronunciati di quelle configurazioni che si trovavano in zone corrispondenti a forze maggiori o, se si vuole, a pendenze locali del potenziale molto più elevate in modulo. Si avrà quindi che la dinamica sovrasmorzata in generale

produrrà un'accumulazione verso i punti estremali dell'energia e un conseguente diradamento nelle altre zone. L'insieme discreto delle nuove configurazioni, ottenute evolvendo quelle iniziali, rischia, anche solo dopo pochi passi, di non essere più idoneo a campionare il cammino con sufficiente risoluzione.

Dato, invece, che lo scopo è individuare il cammino tra i due minimi che attraversa la barriera di energia nel punto più basso, siamo interessati a mantenere un campionamento a risoluzione costante dei cammini nell'ipersuperficie dell'energia. È allora necessario far seguire alla fase di evoluzione una fase di interpolazione, dove vengono aggiunte delle nuove configurazioni per campionare più fittamente il cammino ove necessario, dopo di che si potrà far rilassare nuovamente (sempre per un tempo breve) l'insieme delle nuove configurazioni. In questo modo, alternando la fase di rilassamento a quella di interpolazione per un qualsiasi cammino iniziale, si ha che esso converge con continuità verso il cammino principale. Possiamo quindi calcolare numericamente le configurazioni corrispondenti ai punti sella di tale cammino, che saranno quelle corrispondenti ai punti di massima energia. Riportiamo nelle figure (4.3) l'andamento del profilo dell'energia e della somma dei valori assoluti delle forze agenti sui singoli monomeri delle configurazioni considerate, durante alcune fasi intermedie del processo di rilassamento fino al profilo di convergenza finale. Poiché il numero totale di configurazioni  $N_{tot}$  può cambiare dopo ogni fase d'interpolazione, riportiamo in ascissa il rapporto tra l'indice della configurazione  $m$ <sup>4</sup> e il numero totale di configurazioni della fase considerata. Le figure (4.3) si riferiscono al caso in cui i due minimi sono la configurazione nativa e un suo minimo direttamente

---

<sup>4</sup>L'indice indica la posizione in successione delle configurazioni lungo il profilo energetico a partire dalla CN ( $m = 0$ ) fino al minimo vicino considerato ( $m = N_{tot}$ ).



connesso.

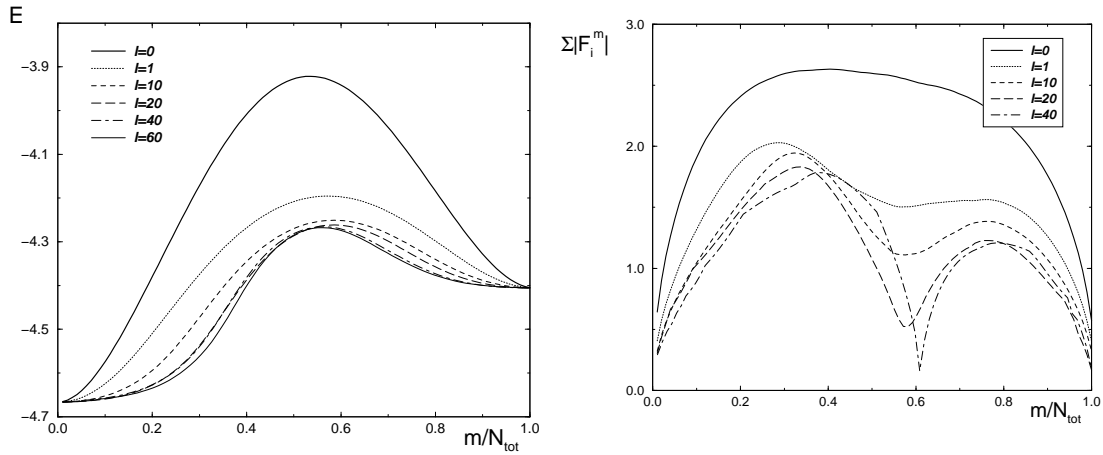


Figura 4.3. a) Profilo di energia di un cammino durante alcune fasi di rilassamento. L'indice  $l$  indica il numero di cicli di interpolazione e evoluzione che sono stati eseguiti. b) Somma del valore assoluto delle forze agenti sui singoli monomeri.

Può accadere che l'identificazione dinamica dei minimi “vicini” della CN, con la tecnica descritta nel paragrafo 4.1 porti all'individuazione di minimi che in realtà non risultano contigui alla CN. Attraverso questa nuova tecnica siamo però in grado di raffinare l'analisi dinamica e di individuare quando il minimo nativo è connesso ai minimi “vicini” non direttamente ma passando da altri minimi; possiamo inoltre individuare la configurazione di tali minimi intermedi e dei relativi punti sella che li separano. Nella figura (4.4) mostriamo per esempio il profilo di energia e la somma dei valori assoluti delle forze agenti sui monomeri delle configurazioni costituenti il cammino principale tra la CN e un minimo  $\mathbf{b}_1$ , non direttamente connesso alla CN. In questo caso abbiamo che tra i bacini d'attrazione corrispondenti a questi due minimi è presente un terzo bacino intermedio.

Per completezza va notato che, in linea di principio, non è detto che un qualunque cammino iniziale converga al cammino principale. Per esaminare questa eventualità

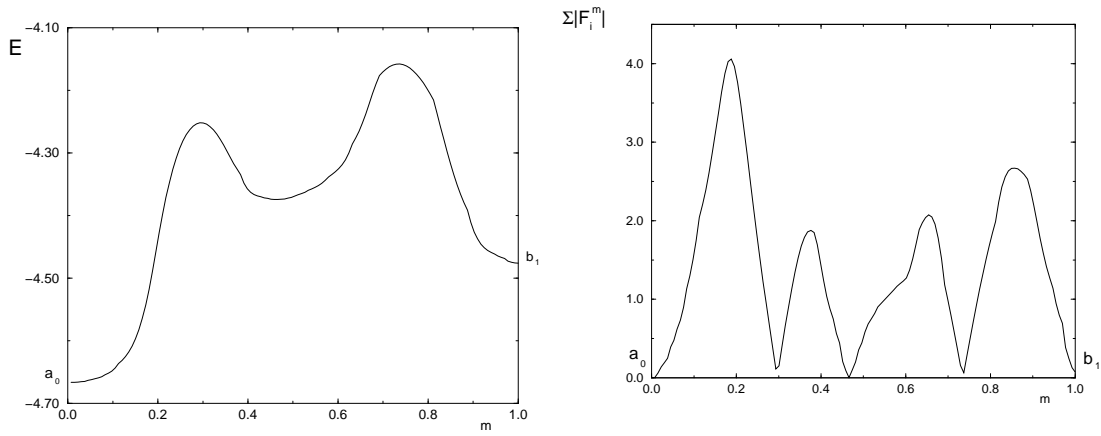


Figura 4.4. a) Energia delle configurazioni finali costituenti il cammino che unisce la CN con  $\mathbf{b}_1$ . b) Somma dei moduli delle forze agenti sui singoli monomeri per ciascuna configurazione finale.

abbiamo dunque considerato più cammini iniziali diversi, con l'intento di prendere tra le barriere di energia ottenute dal profilo asintotico di ciascun cammino iniziale quella di altezza minore. In pratica si è osservato che ciascun cammino iniziale è rilassato sempre verso lo stesso cammino, che è stato perciò individuato come il cammino principale. Per comodità si è perciò in seguito scelto sempre come cammino iniziale quello rettilineo, cioè quello ottenuto come l'interpolazione lineare tra le configurazioni corrispondenti ai due minimi in esame.

### 4.3 Le barriere di energia potenziale

Per prima cosa abbiamo considerato i minimi dinamicamente accessibili dalla CN, che abbiamo individuato attraverso la procedura illustrata nel paragrafo 4.1. Applicando la tecnica descritta nel paragrafo 4.2 all'insieme di tali  $n_t$  minimi, abbiamo potuto distinguere quali di questi minimi siano direttamente connessi alla CN e abbiamo calcolato l'altezza  $W$  della barriera che li separa dalla CN. Facciamo notare

che alcuni degli  $n_t$  minimi non sono risultati essere direttamente connessi alla CN, cioè il loro profilo di energia finale mostra la presenza di almeno un minimo ulteriore tra tali minimi e la CN. In ogni caso si tratta di minimi poco visitati rispetto al numero totale di prove eseguite  $N_p$ . Se infatti introduciamo una soglia sulle probabilità e consideriamo solo i minimi che sono stati visitati un numero di volte pari almeno al 2% del numero totale di prove, abbiamo che tali minimi risultano tutti direttamente connessi alla configurazione nativa.

Abbiamo a questo punto a disposizione tutte le configurazioni dei punti sella corrispondenti ai vari cammini, e le altezze delle barriere di energia.

Per ciascuna temperatura  $T$  consideriamo tra tutti i minimi “vicini” ( $n_t$ ) quei minimi direttamente connessi a CN ( $n_{dc}$ ). Mentre  $n_t$  aumenta considerevolmente con la temperatura, il sottoinsieme di  $n_t$  composto da quei minimi che risultano direttamente connessi alla CN ( $n_{dc}$ ) subisce un incremento molto più lento. Infatti il numero totale di minimi direttamente connessi a un dato minimo è una quantità che prescinde dalla temperatura a cui si eseguono le dinamiche dei salti. A temperature basse il sistema visita solo alcuni dei minimi direttamente connessi al sistema e solo aumentando la temperatura inizia a visitarne di ulteriori. Oltre certi valori della temperatura però il sistema non troverà più nuovi minimi direttamente connessi alla CN mentre il numero di minimi “vicini” che il sistema visita ( $n_t$ ) continuerà ad aumentare. Riportiamo, nella tabella 4.2 il valore di  $n_{dc}$  e di  $n_{co}$  per ciascuna sequenza e per le varie temperature esaminate.

I minimi in generale più visitati sono risultati essere effettivamente quelli corrispondenti alle barriere di energia più basse. Ordiniamo dunque i minimi direttamente

Tabella 4.2. Numero di minimi visitati alla temperatura  $T$  che risultano direttamente connessi ( $n_{dc}$ ) alla CN. Numero di minimi visitati almeno il 2% delle  $N_{tot} = 4000$  prove ( $n_{co}$ ).(\*)Per la sequenza  $S_0$  a  $T = 0.03$ ,  $N_{tot} = 500$ .

$S_0$			$S_1$			$S_2$		
$T$	$n_{dc}$	$n_{co}$	$T$	$n_{dc}$	$n_{co}$	$T$	$n_{dc}$	$n_{co}$
0.03*	10	1	0.04	29	11	0.03	2	1
0.04	23	4	0.05	28	13	0.04	4	1
0.05	31	9	0.06	34	13	0.05	8	1
0.06	43	11	0.07	37	13	0.06	11	2

connessi in ordine crescente rispetto all'altezza della barriera  $W$ . Riportiamo dunque, nella tabella 4.3, l'energia potenziale  $V$ , la distanza angolare  $\delta^0$  e l'altezza della barriera di potenziale  $W$  dei primi dieci minimi per ognuna delle sequenze esaminate. Notiamo che le indicazioni sulle barriere di energia sono più attendibili di quelle della semplice differenza di energia tra i minimi. Il quarto minimo della sequenza  $S_1$ , per esempio, pur avendo un'energia abbastanza alta, è uno dei minimi più visitati, e difatti risulta avere una delle barriere di energia più basse.

Tabella 4.3. Valori dell'energia potenziale  $V$ , della distanza angolare  $\delta^0$  e dell'altezza  $W$  della barriera rispetto alla CN dei primi 10 minimi ordinati secondo il valore di  $W$  per le tre sequenze  $S_0$ ,  $S_1$  e  $S_2$ . Le cifre riportate nella tabella sono tutte significative.

$S_0$			$S_1$			$S_2$		
$V$	$\delta^0$	$W$	$V$	$\delta^0$	$W$	$V$	$\delta^0$	$W$
-6.78307	0.19667	0.31290	-4.57445	0.30158	0.38569	-4.61073	0.04211	0.01760
-6.92689	0.39618	0.43785	-4.40565	0.20675	0.39633	-4.59036	0.39361	0.34677
-6.91915	0.39808	0.44267	-4.44757	0.20055	0.40032	-4.37062	0.35321	0.43905
-6.72890	0.24557	0.44746	-4.29895	0.22598	0.40967	-4.33501	0.22470	0.44660
-6.89044	0.35064	0.46194	-4.42167	0.33760	0.40978	-4.39782	0.33069	0.44911
-6.89680	0.35459	0.47079	-4.37439	0.36744	0.41476	-4.36965	0.30871	0.47014
-6.71883	0.12130	0.50322	-4.41957	0.30925	0.41720	-4.34052	0.34673	0.53599
-6.70549	0.30451	0.51233	-4.57444	0.12309	0.43291	-4.24869	0.36136	0.56353
-6.68994	0.11564	0.52755	-4.30789	0.28734	0.46028	-4.06183	0.27290	0.57089
-6.54018	0.20023	0.531456	-4.44403	0.30836	0.49616	-4.26547	0.48305	0.58147

## 4.4 Studio della dinamica del processo di salto

Riconsideriamo adesso più in dettaglio la dinamica di “salto” dalla CN a un suo minimo direttamente connesso. Come osservato nel paragrafo 4.1, l’evoluzione dinamica tipica che si ha in un tale processo consiste in una fase iniziale, in cui il sistema “visita” la CN, seguita da un brusco salto, ad un dato tempo  $t_f$ , nel bacino di un minimo vicino. Questo fa pensare che al tempo  $t_f$  il sistema lascia il bacino corrispondente alla CN e, passando dal relativo punto sella, inizia a “sorvolare” un altro minimo. Pertanto abbiamo cercato un indicatore fisico che caratterizzi al meglio questa transizione. Se si considera, per esempio, l’andamento dell’energia o della distanza angolare sia dalla configurazione nativa che dal minimo di arrivo, si hanno fluttuazioni troppo grandi di queste quantità per poter individuare alcun cambiamento significativo in corrispondenza di  $t_f$ . Dato che ci aspettiamo che il sistema arrivi al bacino direttamente connesso passando dal punto sella corrispondente alla barriera energetica più bassa è abbastanza ovvio considerare come indicatore la proiezione del moto del sistema lungo il vettore congiungente la CN al punto sella in questione. Oltre alla dinamica di salto dalla CN a un altro minimo, siamo in grado di conoscere anche la configurazione corrispondente al punto sella relativo, che chiaramente è uno dei candidati a darci informazioni utili. Prendiamo per esempio la sequenza  $S_1$  e consideriamo una particolare dinamica di salto in cui dal minimo  $\mathbf{a}_0$  viene raggiunto il minimo  $\mathbf{a}_1$  ( $\mathbf{a}_1$  sia il primo minimo in ordine dell’altezza della barriera  $W$ ); consideriamo i minimi come vettori dello spazio 40-dimensionale, per esempio  $\mathbf{a}_0 = (q_x^0(1), \dots, q_x^0(20), q_y^0(1), \dots, q_y^0(20))$ . Indichiamo con  $\mathbf{A}_1$  il vettore della configurazione corrispondente al punto sella tra il minimo  $\mathbf{a}_0$  e il minimo  $\mathbf{a}_1$ .

Definiamo quindi il vettore nello spazio 40-dimensionale delle coordinate, dato dalla differenza tra  $\mathbf{A}_1$  e il vettore  $\mathbf{a}_0$  e ne consideriamo il corrispondente versore  $\mathbf{n}$ . Durante l'evoluzione dinamica possiamo inoltre definire, a ogni tempo  $t$ , il vettore delle configurazioni  $\mathbf{a}(t) = (q_x(1), \dots, q_x(20), q_y(1), \dots, q_y(20))$ , dove la posizione dei monomeri è valutata al tempo  $t$ , e quindi il vettore differenza  $\mathbf{d} = \mathbf{a} - \mathbf{a}_0$ . Nella figura (4.5) mostriamo l'andamento rispetto al tempo del prodotto scalare tra la proiezione del vettore differenza  $\mathbf{d}$  sul versore  $\mathbf{n}$ , cioè  $(\mathbf{a}(t) - \mathbf{a}_0) \cdot \mathbf{n}$ . Notiamo che

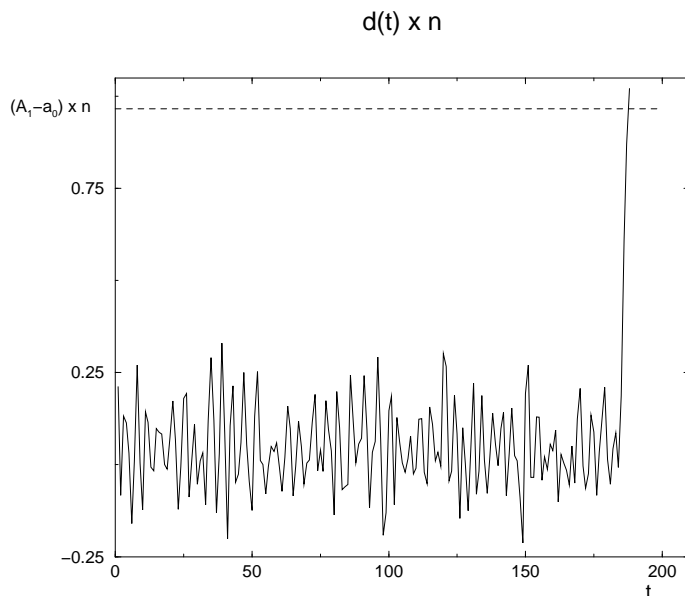


Figura 4.5. Andamento rispetto al tempo del prodotto scalare  $\mathbf{d}(t) \cdot \mathbf{n}$ . La linea tratteggiata è il valore del prodotto scalare tra  $\mathbf{A}_1 - \mathbf{a}_0$  e  $\mathbf{n}$ , dove  $\mathbf{A}_1$  è il vettore corrispondente al punto sella

nell'istante  $t_f$  in cui il sistema finalmente passa dalla CN al nuovo minimo  $\mathbf{a}_1$ ,  $\mathbf{d}(t_f)$  si orienta lungo la direzione  $(\mathbf{A}_1 - \mathbf{a}_0)$ . Per ogni minimo direttamente connesso ad  $\mathbf{a}_0$  si ha un andamento simile a quello mostrato in figura (4.5) per il minimo  $\mathbf{a}_1$ . Questi risultati ci hanno confermato l'intuizione che il sistema salti il bacino d'attrazione del minimo di partenza passando dal punto sella corrispondente alla barriera

energetica più bassa. Inoltre se consideriamo per esempio il “salto” riportato nella figura (4.5), la temperatura a cui abbiamo effettuato la simulazione è dell’ordine di  $T \sim 0.05$ , mentre la barriera energetica che viene superata è dell’ordine di  $W \sim 0.4$ , cioè quasi dieci volte superiore. Perciò, se il processo di salto fosse regolato soltanto dall’altezza della barriera avremmo che la probabilità di salto sarebbe assai piccola. In realtà, come mostreremo nel paragrafo seguente, i contributi entropici alla probabilità di salto sono assai rilevanti tanto da rendere tale probabilità non trascurabile anche alla temperatura in questione. Nel prossimo paragrafo riporteremo delle stime teoriche raffinate riguardanti la probabilità di fuga da uno stato metastabile. Tali analisi teoriche riguardano i processi di fuga, dovuti a processi di attivazione, nel caso di un sistema unidimensionale [33] e multidimensionale [34]. Tramite una discussione critica delle ipotesi con cui sono state derivate tali espressioni teoriche e dal confronto con i dati numerici deriveremo i loro limiti di applicabilità.

## 4.5 Probabilità di fuga da uno stato metastabile

Il problema della fuga da uno stato metastabile ha un ruolo importante in molti processi fisici come la diffusione nei solidi, il trasporto degli elettroni nei semiconduttori, la cinetica delle reazioni chimiche, ecc.. La descrizione del processo di fuga da uno stato di equilibrio, attivata dalle fluttuazioni termiche, si basa principalmente sulla teoria del moto Browniano. In pratica si assume che ci sia una netta separazione tra le scale di tempo del sistema e quelle relative alla dinamica interna al bagno termico; si assume inoltre che i gradi di libertà relativi al bagno termico siano solo debolmente accoppiati a quelli del sistema e che quindi le loro interazioni possano

non essere esplicitamente considerate nel termine di energia potenziale, ma tenute in conto tramite un termine di smorzamento del moto e forze casuali (o *random*) responsabili delle proprietà Browniane del sistema. Pertanto le equazioni del moto, che governano la dinamica del processo di fuga, contengono delle forze *random*, che permettono al sistema di uscire da stati localmente stabili. Generalmente queste forze sono abbastanza deboli, per cui le traiettorie del sistema sono localizzate per la maggior parte del tempo all'interno della buca di potenziale intorno al minimo iniziale fino a quando intervengono delle fluttuazioni casuali che consentono al sistema di attraversare la barriera di potenziale e localizzare nuovamente la traiettoria in prossimità di un nuovo minimo locale. Consideriamo un sistema classico composto da  $N$  particelle di massa  $m$  in uno spazio a dimensione  $d \leq 3$ , assimilabili a punti materiali, con  $\vec{r}_i$  e  $\vec{v}_i$  ( $i = 1, \dots, N$ ) vettori posizione e velocità delle singole particelle. Se questo sistema è immerso in un bagno termico a temperatura  $T$ , avremo che le corrispondenti equazioni del moto, per ciascuna particella di massa  $m$ , assumono la seguente forma alla Langevin:

$$\begin{aligned}\dot{\vec{r}}_i &= \vec{v}_i \\ \dot{\vec{v}}_i &= -\frac{1}{m}\nabla_{\vec{r}_i}V - \gamma\vec{v}_i + \vec{\xi}(t).\end{aligned}\tag{4.1}$$

L'accelerazione *random*  $\vec{\xi}(t)$  è causata dalle interazioni con i gradi di libertà del bagno termico, per esempio dalle collisioni molto rapide con le molecole del fluido circostante. Questo termine è considerato equivalente ad un processo stocastico a media nulla, la cui funzione di autocorrelazione è legata al fattore di smorzamento  $\gamma$  tramite il teorema di fluttuazione-dissipazione, per cui considerando ciascuna



componente  $\xi^j$ , con  $j = 1, \dots, d$ , si avrà che:

$$\langle \xi^h(t) \cdot \xi^k(s) \rangle = \left( \frac{2k_B T \gamma}{m} \right) \delta_{hk}(t - s).$$

La media è calcolata su un insieme composto da varie rappresentazioni del nostro sistema, o anche su successive osservazioni, richiedendo che queste ultime siano sufficientemente distanziate nel tempo in modo da poter essere considerate non correlate e quindi statisticamente indipendenti.

### 4.5.1 Caso di una particella

Consideriamo dapprima per semplicità il caso di una particella di massa  $m$  soggetta a un campo di forze  $F(x)$  unidimensionale e immersa in un bagno termico a temperatura costante  $T$ , supponiamo inoltre che l'energia potenziale  $V(x)$  abbia la forma mostrata in figura (4.6). A noi interessa l'espressione per la probabilità di fuga della particella per unità di tempo dal punto  $x = a$ , corrispondente a un minimo dell'energia potenziale, per l'attraversamento del punto  $x = s$ , corrispondente ad un massimo dell'energia potenziale. La particella eccitata termicamente deve dunque superare la barriera  $W = V(s) - V(a)$  (ci limitiamo al caso in cui la temperatura del bagno sia tale che  $kT \ll W$ ).

Definiamo quindi la funzione di distribuzione  $P(x, v, t)$  in modo tale che  $P(x, v, t) dx dv$  sia la probabilità che il sistema si trovi nella posizione compresa tra  $x$  e  $x + dx$ , con una velocità compresa tra  $v$  e  $v + dv$  al tempo  $t$ ; sia inoltre  $P(x, v, t)$  normalizzata nello spazio  $(x, v)$ :

$$\int P(x, v, t) dx dv = 1. \quad (4.2)$$

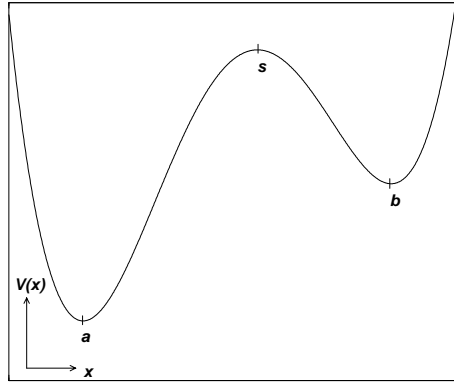


Figura 4.6. Andamento qualitativo dell'energia potenziale  $V(x)$  unidimensionale.

Si dimostra (vedi [35]) che la dinamica di Langevin (4.1) nel caso di una sola particella in un campo di forze unidimensionale, può essere rappresentata dal processo Markoviano  $x(t)$  la cui corrispondente densità di probabilità  $P(x,v,t)$  verifichi l'equazione di Fokker-Planck unidimensionale:

$$\partial_t P(x,v,t) = -v \frac{\partial P}{\partial x} - \frac{F(x)}{m} \frac{\partial P}{\partial v} + \gamma \frac{\partial}{\partial v} (PV) + \frac{1}{2} \frac{\partial^2}{\partial v^2} \left( \gamma \frac{k_B T}{m} \right) P. \quad (4.3)$$

Nelle ipotesi che tutto il sistema sia all'equilibrio termodinamico e che siano trascurabili salti all'indietro una volta che la particella abbia attraversato la sella, possiamo considerare la soluzione stazionaria  $P^0(x,v)$  dell'equazione (4.3) che avrà la forma:

$$P^0(x,v) = \begin{cases} C \exp[-(V(x) + \frac{mv^2}{2})/k_B T] & \text{se } x \leq s \\ \approx 0 & \text{se } x > s. \end{cases} \quad (4.4)$$

La costante  $C$  si ottiene imponendo la condizione di normalizzazione (4.8) (sviluppando  $V(x)$  in un intorno di  $a$ ). La probabilità di fuga  $\Gamma$  è data quindi dal flusso in avanti di  $P^0(x,v)$  attraverso la cima  $x = s$  per cui si ottiene la seguente espressione

per  $\Gamma$ :

$$\Gamma = \int_0^{+\infty} v P^0(s, v) dv = \frac{w_a}{2\pi} \exp[-W/kT]. \quad (4.5)$$

con

$$w_a^2 = \left. \frac{1}{m} \frac{d^2 V}{dx^2} \right|_{x=a}$$

Soffermiamoci sulle ipotesi utilizzate per ricavare l'espressione (4.5) per  $\Gamma$ . Se non si considerano temperature troppo alte è ragionevole assumere che siano trascurabili i salti all'indietro della particella, mentre sembra meno plausibile assumere che gli effetti di non equilibrio siano trascurabili durante l'intero processo, dato che questo, per essere attivato, ha bisogno che la particella esca da uno stato di equilibrio. La precedente assunzione potrà comunque ritenersi un'approssimazione tanto più valida quanto più il tempo di permanenza medio all'interno di una buca sarà maggiore del tempo di termalizzazione.

Procedendo quindi con un approccio perturbativo rispetto alla teoria di equilibrio, riprendiamo l'equazione (4.3) e consideriamo il caso di alta viscosità, cioè si assume che l'effetto dovuto alle forze Browniane sulla velocità della particella sia maggiore di quello dovuto al campo  $F(x)$ .

Si suddivide l'intervallo  $(a, s)$  in due intervalli  $(a, c)$  e  $(c, s)$  scegliendo un punto  $c$  tale che:

- (i) per  $x < c$  la soluzione  $P(x, v)$  coincida con la soluzione stazionaria  $P^0(x, v)$
- (ii) per  $x > c$  si possa approssimare  $V(x)$  con la parabola

$$V(x) = V(s) - \frac{1}{2} m \omega_s^2 (x - s)^2;$$

con

$$w_s^2 = -\frac{1}{m} \frac{d^2V}{dx^2} \Big|_{x=s}$$

Il punto  $c$  esiste se  $\gamma$  non è troppo piccolo ( $\gamma \gtrsim \omega_s$ ); in questo regime la probabilità di fuga è limitata dalle sole collisioni che avvengono in prossimità del punto sella e un tipico cammino non va direttamente da una parte all'altra della sella ma attraversa avanti e indietro la cima prima di scappare definitivamente. Questi attraversamenti ripetuti della barriera modificheranno la distribuzione di equilibrio nella regione prossima al punto sella in particolare si avrà una diminuzione di  $P(x,v,t)$  rispetto al caso di equilibrio. Inoltre se  $F(x)$  non varia molto sulla distanza  $\sqrt{\frac{kT}{m\gamma^2}}$  ci aspettiamo che, partendo da una distribuzione iniziale arbitraria, si instauri presto (dopo un tempo dell'ordine di  $1/\gamma$ ), una distribuzione Maxwelliana delle velocità per ciascun valore delle  $x$ .

Per  $x \sim s$ , si può riscrivere la soluzione dell'equazione (4.3)  $P(x,v)$  come una perturbazione rispetto alla soluzione stazionaria:

$$P(x,v) = Q(x,v)P^0(x,v).$$

In queste ipotesi si ottiene dunque la seguente espressione per  $\Gamma$ :

$$\Gamma = \frac{\omega_s \omega_a}{\alpha} \frac{1}{2\pi} e^{-\frac{W}{kT}}; \quad (4.6)$$

con  $\alpha = \frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{2} + \omega_s^2}$ .

Nell'appendice C è riportata la derivazione dettagliata dell'espressione (4.6).

Nell'limite  $\gamma \gg \omega_s$  (caso sovrasmorzato) l'espressione per la probabilità di fuga diviene:

$$\Gamma = \frac{\omega_a \omega_s}{2\pi\gamma} e^{-\frac{W}{kT}}. \quad (4.7)$$

La (4.7) può essere ricavata con un procedimento analogo partendo da un'equazione di diffusione tipo Fokker-Planck per le sole variabili spaziali detta equazione di Smoluchowski [35].

Notiamo infine che nel limite  $\gamma \ll \omega_s$  (caso sottosmorzato) la (4.6) viene a coincidere con la (4.5), ma tale espressione per  $\Gamma$  in realtà sovrastima l'effettiva probabilità di salto (del resto una delle ipotesi che si utilizzano per ricavare la (4.6) è che  $\gamma$  sia maggiore di  $\omega_s$ <sup>5</sup>).

Il valore del coefficiente di viscosità  $\gamma$  (che in generale può dipendere dalla temperatura) è una misura dell'intensità con la quale il sistema interagisce con il mezzo circostante. Se  $\gamma$  è molto piccolo la particella una volta che ha attraversato la barriera può rimbalzare indietro senza aver raggiunto l'equilibrio all'interno della buca in cui era arrivata e la probabilità di transizione è minore rispetto al valore fornito dalla (4.5); ciò accade anche, come abbiamo appena visto, per alti valori di  $\gamma$  essendo  $\alpha > \omega_s$ . Nella regione intermedia ( $\omega_a \frac{kT}{W} \lesssim \gamma \lesssim \omega_s$ ) la (4.5) può essere ritenuta valida per il calcolo della probabilità di fuga.

## 4.5.2 Caso di più particelle

Consideriamo adesso il caso più attinente al nostro modello di un sistema classico composto da  $N$  particelle di massa  $m$  in uno spazio  $d$ -dimensionale ( $d \leq 3$ ). Lo stato dinamico del sistema è rappresentato dal vettore  $\vec{\eta} = (\vec{x}, \vec{p}) = (\vec{r}_i, \vec{p}_i, i = 1, \dots, N)$  definito nello spazio  $2dN$ -dimensionale, dove  $\vec{r}_i$  e  $\vec{p}_i$  sono i vettori che individuano

---

<sup>5</sup>Rimandiamo all'articolo originale di Kramers [36] per una trattazione rigorosa del caso di smorzamento debole ( $\gamma \ll \omega_s$ ).

le coordinate e i momenti del punto  $i$ -esimo <sup>6</sup>. Definiamo inoltre la funzione di distribuzione  $P(\vec{\eta}, t)$  in modo tale che  $P(\vec{\eta}, t)\delta\vec{\eta}$  sia la probabilità che il sistema si trovi nello stato compreso tra  $\vec{\eta}$  e  $\vec{\eta} + \delta\vec{\eta}$  al tempo  $t$ , normalizzata nello spazio  $2dN$  dimensionale:

$$\int \prod_{i=1}^{2dN} d\eta_i P(\vec{\eta}, t) = 1. \quad (4.8)$$

Sia  $H(\vec{x}, \vec{p}) = K + V$  l'Hamiltoniana di questo sistema e siano inoltre  $\vec{x}_a$  e  $\vec{x}_b$  due punti nello spazio delle coordinate corrispondenti a due minimi dell'energia potenziale  $V(\vec{x})$ . Consideriamo il caso che i minimi  $\vec{x}_a$  e  $\vec{x}_b$  siano direttamente connessi tra loro, cioè che il cammino principale che congiunge  $\vec{x}_a$  a  $\vec{x}_b$  passi da un unico punto sella  $\vec{x}_s$  <sup>7</sup>. Supponiamo inoltre che il sistema sia in contatto con un bagno termico a una temperatura costante  $T$ .

Vogliamo calcolare la probabilità, per unità di tempo, che il sistema, inizialmente vicino allo stato  $\vec{x}_a$ , passi per attivazione termica allo stato  $\vec{x}_b$ , superando la barriera di energia di altezza  $W = V(\vec{x}_s) - V(\vec{x}_a)$ . Indicheremo tale probabilità con  $\Gamma$ .

Si dimostra (vedi [35]) che la dinamica di Langevin (4.1) può essere rappresentata dal processo Markoviano  $\vec{\eta}(t)$  con probabilità  $P(\vec{\eta}, t)$  che verifichi l'equazione di Fokker-Planck multidimensionale.

Nel caso sovrasmorzato possiamo considerare inoltre la funzione  $P(\vec{x}, t)$  ottenuta integrando nello spazio delle velocità  $P(\vec{\eta}, t)$ ; essa dovrà soddisfare l'equazione (vedi

---

<sup>6</sup>I vettori  $\vec{r}_i$  e  $\vec{p}_i$  sono definiti nello spazio  $d$ -dimensionale mentre  $\vec{x}$  e  $\vec{p}$  nello spazio  $dN$ -dimensionale

<sup>7</sup>Si definisce punto sella  $\vec{x}_s$  di una funzione  $V$  quel punto tale che lo Jacobiano di  $V$  calcolato in  $S$  è nullo e gli autovalori della matrice Hessiana di  $V$  calcolata in  $\vec{x}_s$  sono tutti non negativi tranne uno che è negativo.

[35]) nello spazio delle coordinate:

$$\frac{\partial P(\vec{x},t)}{\partial t} = \sum_{i=1}^{dN} \frac{1}{m\gamma} \frac{\partial}{\partial x_i} \left( \frac{\partial V}{\partial x_i} P(\vec{x},t) + kT \frac{\partial P(\vec{x},t)}{\partial x_i} \right). \quad (4.9)$$

Partendo da questa equazione e limitandosi al caso sovrasmorzato si ricava la seguente espressione per  $\Gamma$  (vedi appendice D):

$$\Gamma = \frac{2\pi}{m\gamma} \sqrt{|\lambda_1|} e^{-W/kT} \frac{\prod_{n=1}^M (\lambda_n^a)^{1/2}}{\prod_{n'=2}^M (\lambda_{n'})^{1/2}}. \quad (4.10)$$

Le quantità  $\{\lambda_n\}$  e  $\{\lambda_n^a\}$  sono gli autovalori non nulli della matrice Hessiana dell'energia potenziale  $V$ , calcolata rispettivamente nel punto di sella  $\vec{x}_s$  e nel punto di minimo  $\vec{x}_a$  e  $M$  è il numero di gradi di libertà del sistema. Se infatti  $M$  è minore della dimensione dello spazio delle coordinate  $dN$ , in cui si esprime  $V$ , si avranno in generale  $dN - M$  autovalori nulli, che chiaramente non compaiono nel termine di produttoria. I restanti autovalori non nulli sono tutti positivi nel caso del minimo  $\vec{x}_a$  (indichiamo tali autovalori con  $\lambda_n^a$  con  $n = 1, \dots, M$ ) mentre sono tutti positivi tranne uno ( $\lambda_1$ ) nel caso del punto di sella  $\vec{x}_s$ .

Volendo riscrivere la (4.10) in termini delle frequenze  $\omega_i = \sqrt{|\lambda_i|/m}$ , avremo finalmente l'espressione:

$$\Gamma = \frac{\omega_1}{\gamma} \left[ \frac{1}{2\pi} e^{-W/kT} \frac{\prod_{n=1}^M (\omega_n^a)}{\prod_{n'=2}^M (\omega_{n'})} \right]. \quad (4.11)$$

Nell'appendice D è riportata la derivazione della (4.11). La quantità all'interno delle parentesi quadre è la probabilità di transizione che si ottiene con la teoria di equilibrio, il fattore di riduzione  $\omega_1/\gamma$  è lo stesso che si ottiene nel caso unidimensionale (equazione (4.7)); più piatta è la sella lungo la direzione instabile (cioè più piccolo è  $\omega_1$ ) maggiore è l'opposizione al moto diffusivo.

## 4.6 Stima delle probabilità di salto tramite la formula teorica

Nel paragrafo 4.1 abbiamo individuato i minimi “vicini” alla CN. Tramite il programma descritto nel paragrafo 4.2 abbiamo quindi individuato quali di questi minimi sono direttamente connessi alla CN. Abbiamo dunque calcolato l’altezza della barriera per ciascun minimo e inoltre conosciamo la configurazione del corrispondente punto sella. Per questi minimi possiamo pensare di poter applicare la formula dell’equazione (4.11), volendo calcolare qual’è la probabilità per unità di tempo che il sistema, inizialmente nella CN, passi, attraversando la barriera di energia, in uno di questi minimi. Abbiamo dunque bisogno di conoscere gli autovalori della matrice Hessiana dell’energia potenziale  $V$  calcolata sia nella configurazione nativa che in tutti i punti sella corrispondenti ai minimi direttamente connessi a CN. Infine ci occorre conoscere il valore della costante di smorzamento  $\gamma$ , che misura l’interazione effettiva tra il nostro sistema e i gradi di libertà del bagno termico. Descriviamo adesso come abbiamo calcolato tali quantità.

- **Autovalori della matrice Hessiana** - La matrice Hessiana dell’energia potenziale  $V$  è una matrice simmetrica  $40 \times 40$  che, espressa in blocchi, ha la forma ( $\forall i, j = 1, \dots, N$ ):

$$H = \begin{pmatrix} \frac{\partial^2 V}{\partial r_x(i) \partial r_x(j)} & \frac{\partial^2 V}{\partial r_x(i) \partial r_y(j)} \\ \frac{\partial^2 V}{\partial r_y(i) \partial r_x(j)} & \frac{\partial^2 V}{\partial r_y(i) \partial r_y(j)} \end{pmatrix}.$$



Gli autovalori della matrice Hessiana si calcolano tramite un programma che diagonalizza le matrici (preso da [29]), una volta scritti esplicitamente gli elementi della matrice rispetto a un generico punto dello spazio delle coordinate  $(r_x(i), r_y(i))$ . Come potevamo aspettarci dalle invarianze di  $V$  per traslazioni e riflessioni, tre degli autovalori della matrice Hessiana sono risultati nulli nei limiti dell'errore numerico ( $< 10^{-13}$ ). Come ci si aspettava, nel caso della CN (e di tutti gli altri minimi di cui abbiamo calcolato gli autovalori) gli altri autovalori sono risultati essere tutti positivi e, nel caso di tutti i punti sella considerati, gli altri autovalori, oltre ai tre nulli, sono risultati essere tutti positivi tranne uno negativo, che corrisponde alla frequenza espandente  $\omega_1$ , che compare nell'equazione (4.11). Nella figura (4.7) sono riportati i valori dei 40 autovalori della matrice  $H$ , calcolata nella CN delle tre sequenze.

- **Calcolo della costante di smorzamento** - In tutte le misure che abbiamo eseguito, il sistema si trova in equilibrio con un bagno termico a una temperatura costante  $T$ . Se pensiamo di perturbare il sistema portandolo leggermente fuori dall'equilibrio, esso verrà riportato nuovamente allo stato stazionario dall'azione del bagno termico. La velocità con cui il sistema ritorna all'equilibrio dipenderà dalle interazioni tra il bagno termico e il sistema e quindi sarà legata alla costante di smorzamento  $\gamma$ . In particolare abbiamo supposto che il moto di ciascun monomero del sistema, immerso nel bagno termico, possa essere descritto dall'equazione di Langevin (4.1), dove  $\gamma$  è la costante di smorzamento che vogliamo appunto calcolare. Dalle equazioni (4.1) otteniamo l'equazione di moto per il centro di massa del sistema,  $\vec{r}_{cm}$ ,

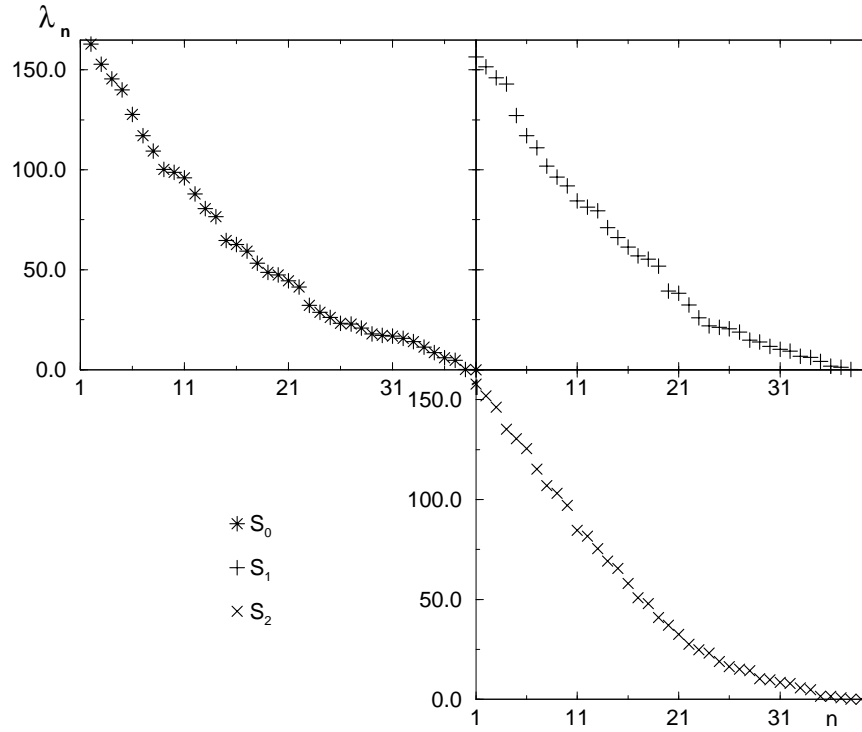


Figura 4.7. Autovalori della matrice Hessiana  $H$  calcolata nel punto dello spazio delle coordinate corrispondente alla CN delle sequenze  $S_0$ ,  $S_1$  e  $S_2$ .

che avranno la forma:

$$\dot{\vec{r}}_{cm} = \vec{v}_{cm} \quad (4.12)$$

$$\dot{\vec{v}}_{cm} = -\gamma \vec{v}_{cm} + \vec{\xi}(t).$$

dove  $\vec{\xi}(t)$  è la risultante delle forze *random*, che agiscono sul sistema. Notiamo quindi che, se si considera la media (eseguita su varie rappresentazioni del sistema) della velocità del centro di massa, nell'equazione (4.12) si annullerà anche il termine *random* cosicché una qualunque componente della media di  $\vec{v}_{cm}$ , per esempio  $\langle (v_{cm})_x \rangle$ , avrà un andamento esponenziale  $\langle (v_{cm})_x \rangle = v_0 e^{-\gamma t}$  con costante di decadimento  $\gamma$  e dove  $v_0$  dipende delle condizioni iniziali del moto.

Per calcolare il valore di  $\gamma$  dinamicamente abbiamo dunque agito nel seguente modo. Si prende come condizione iniziale del moto la CN e si riscalda progressivamente il sistema fino a che raggiunge una temperatura fissata  $T$ . Una volta che il sistema si trovi alla temperatura  $T$  lo si fa evolvere a temperatura costante per un certo tempo, in modo tale da raggiungere una situazione stazionaria.

Alla componente  $x$  della velocità di ciascun monomero si aggiunge poi una perturbazione costante  $v_0$  che si sceglie pari a 10 volte il valore di  $T$ . Si continua a far evolvere il sistema registrando, ad ogni passo temporale ( $\delta t = 0.025$ ), la componente  $x$  della velocità del centro di massa  $v_x(t)$ . Per eliminare le fluttuazioni nei valori di  $v_x(t)$  abbiamo dunque mediato questa quantità su 400 realizzazioni distinte dello stesso processo (data la natura stocastica del termostato, ogni fase di riscaldamento produce uno stato del sistema leggermente diverso). La quantità  $\langle v_x(t) \rangle$  avrà un andamento esponenziale decrescente  $v_0 e^{-\gamma t}$ ; stimando la rapidità del decadimento si può ottenere il coefficiente  $\gamma$ , che rappresenta la costante di smorzamento del moto. Il valore di  $\gamma$  per le varie temperature considerate e per le tre sequenze è risultato essere praticamente costante, come mostrato nella tabella 4.4. Nella figura (4.8) è mostrato, come esempio, l'andamento della quantità  $\frac{\langle v_x \rangle}{v_0}$  rispetto al tempo  $t$ , alla temperatura  $T = 0.06$ , nel caso della sequenza  $S_1$ .

Siamo quindi in grado di calcolare, avvalendoci della (4.11), i valori della probabilità di fuga dalla CN, per unità di tempo, verso tutti i minimi direttamente connessi

Tabella 4.4. Valori della costante di smorzamento  $\gamma$  per le sequenze  $S_1$ ,  $S_2$  e  $S_3$ , a vari valori della temperatura  $T$ , l'incertezza su  $\gamma$  è pari a 0.1.

$S_0$		$S_1$		$S_2$	
$T$	$\gamma$	$T$	$\gamma$	$T$	$\gamma$
0.03	$6.9 \pm 0.1$	0.04	$7.1 \pm 0.1$	0.03	$6.9 \pm 0.1$
0.04	$6.8 \pm 0.1$	0.05	$6.8 \pm 0.1$	0.04	$6.9 \pm 0.1$
0.05	$6.9 \pm 0.1$	0.06	$7.0 \pm 0.1$	0.05	$6.9 \pm 0.1$
0.06	$7.0 \pm 0.1$	0.07	$6.9 \pm 0.1$	0.06	$6.9 \pm 0.1$

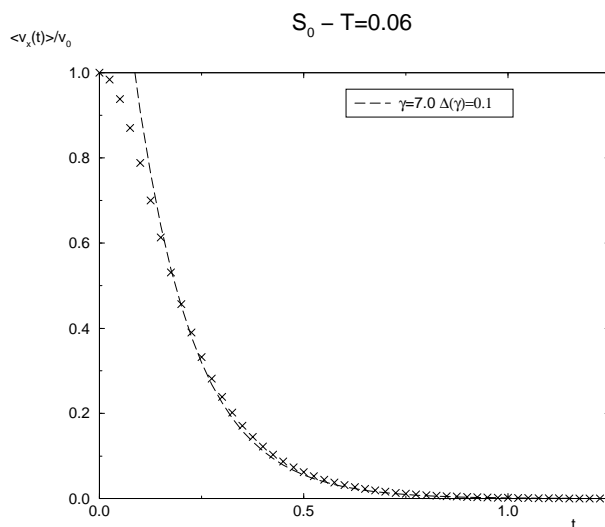


Figura 4.8. Andamento della componente della velocità del centro di massa mediata su 400 condizioni iniziali per la sequenza  $S_1$  alla temperatura  $T = 0.06$

alla CN di ciascuna sequenza. Sottolineiamo che siamo effettivamente nel caso sovrasmorzato ( $\gamma > \omega_1$ ); infatti i valori di  $\omega_1$  nei vari punti sella considerati non sono mai superiori a 1.5. Nel prossimo paragrafo confronteremo tale stima con il corrispondente valore ottenuto dai dati numerici.

## 4.7 Stima numerica delle probabilità di salto

Avendo a disposizione una buona statistica relativa alle dinamiche di salto, possiamo dare una stima diretta della probabilità di salto, per unità di tempo, dalla CN ai suoi

minimi “vicini”. Analogamente a quanto riportato nel paragrafo 4.1, consideriamo la serie di  $N_v$  salti dalla CN ad altri bacini di attrazione. Sia  $N_m$  il numero di volte che il sistema è saltato in un certo minimo  $\mathbf{a}_m$  su  $N_v$  salti totali, sia  $\langle t_p \rangle$  il tempo di permanenza medio nella CN prima di un qualsiasi salto a temperatura  $T$ . La probabilità di transizione dalla configurazione nativa al minimo  $\mathbf{a}_m$ , per unità di tempo, ( $\Gamma_{\mathbf{a}_0 \rightarrow \mathbf{a}_m}$ ) è data da:

$$\Gamma_{\mathbf{a}_0 \rightarrow \mathbf{a}_m} = \frac{N_m}{N_v \langle t_p \rangle}. \quad (4.13)$$

Se tra i minimi “vicini” ci si limita al sottoinsieme dei minimi direttamente connessi, possiamo confrontare i dati ottenuti dinamicamente con quelli della espressione teorica (4.11) a varie temperature. Nelle figura (4.9) sono mostrati i dati per i primi 15 minimi (ordinati in funzione dell’altezza della barriera energetica  $W$  che li separa dalla CN) direttamente connessi alla configurazione nativa relativi alle sequenze  $S_0$  e  $S_1$  per due temperature diverse <sup>8</sup>. Nel caso della sequenza  $S_1$  l’accordo tra le due curve è molto buono per entrambe le temperature soprattutto considerando le numerose approssimazioni che sono state eseguite per derivare la (4.11). Nel caso della sequenza  $S_0$  l’andamento delle due curve è simile ma vi è uno scostamento sistematico abbastanza evidente. Al crescere della temperatura la stima di Langer [34] è sempre meno valida in quanto viene meno l’ipotesi fondamentale  $W > T$ , cioè che il salto sia un processo di attivazione termica.

Alla luce di questi risultati possiamo affermare che la stima di Langer è una buona approssimazione per le probabilità di salto (dovute ad attivazione termica) tra i

---

<sup>8</sup>La sequenza  $S_2$  ha sostanzialmente un unico minimo connesso alla CN e la barriera di energia associata è molto bassa. Risulta quindi privo d’interesse confrontare i dati dinamici per la probabilità di fuga con quelli teorici derivati nelle ipotesi che  $W \gg T$ .

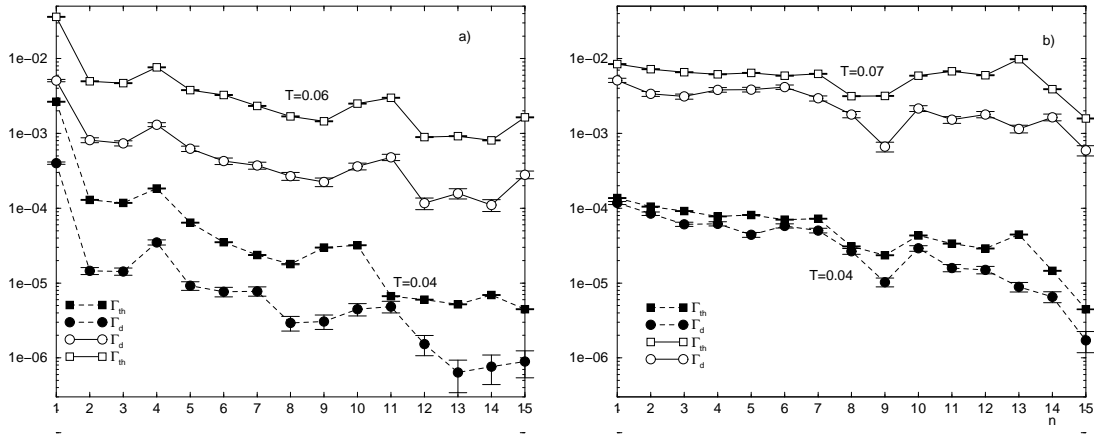


Figura 4.9. Probabilità di salto per unità di tempo, in scala logaritmica, dalla CN al minimo  $\mathbf{a}_n$  valutato dinamicamente ( $\Gamma_d$ ) e attraverso la formula teorica (4.11) ( $\Gamma_{th}$ ). a) Sequenza  $S_0$ . b) Sequenza  $S_1$ .

minimi del potenziale almeno in prossimità della temperatura di *foldng*.

Sebbene l'espressione (4.11) sia stata ampiamente usata in letteratura nello studio di vetri [37] e *cluster* [38], non ne è mai stata verificata l'applicabilità. Il lavoro riportato in questa tesi rappresenta il primo confronto accurato dell'espressione (4.11) con dati numerici per un sistema multidimensionale complesso. Si può altresì affermare che le differenze fra la (4.11) e le probabilità di transizione osservate dinamicamente sono probabilmente dovute al grado di approssimazione con cui Langer tiene conto dei contributi entropici. Tali contributi sono legati alla struttura geometrica del minimo di partenza e del punto sella valicato, ma Langer si limita a tener conto delle correzioni al secondo ordine all'energia potenziale in corrispondenza di tali punti estremali. Per migliorare la (4.11) si può pensare di riscriverla sostituendo alle frequenze "nude"  $\{\omega_n^a\}$  e  $\{\omega_n\}$  delle frequenze efficaci, che tengano conto anche dei termini di ordine superiore al secondo nello sviluppo dell'energia.

## 4.8 Costruzione dei grafi

Per avere una visione dettagliata delle caratteristiche del paesaggio di energia intorno alla CN abbiamo voluto dare una rappresentazione grafica dei minimi e delle connessioni fra loro esistenti. Per prima cosa abbiamo considerato la CN e l'insieme dei minimi direttamente connessi ad essa, limitandoci a quei minimi che nelle nostre prove dinamiche sono stati visitati a partire dalla CN almeno nel 5% delle prove totali. Li abbiamo dunque rappresentati in un grafo ponendo in ordinata il valore dell'energia potenziale e in ascissa una quantità arbitraria <sup>9</sup>. Di ognuno di questi minimi sono stati a loro volta studiati i propri minimi vicini e le relative connessioni reciproche, su questi ultimi minimi abbiamo agito analogamente ottenendo un terza generazione di minimi. Le informazioni sono state sintetizzate in un grafo generale che rappresenta i minimi connessi alla CN, i loro primi secondi e terzi vicini e le loro reciproche connessioni. I minimi direttamente connessi alla CN sono indicati con la lettera a, quelli connessi alla CN tramite un minimo intermedio con la lettera b e quelli connessi tramite due minimi intermedi con la lettera c. Abbiamo riportato le connessioni tra un minimo e un altro solo se la percentuale di visite è pari almeno al 5% delle prove eseguite. Per indicare la connessione tra un minimo e un altro con energia maggiore si è usata una linea continua, la connessione con un minimo di energia inferiore è indicata con una linea tratteggiata. Per individuare i minimi connessi a un dato minimo sono stati esaminati 500 salti a una data temperatura  $T$ , scelta in modo da consentire l'analisi in tempi accettabili; la procedura è quella riportata nel paragrafo 4.1. Successivamente per ogni coppia di minimi si è verificato

---

<sup>9</sup>Dato che a noi interessa principalmente dare l'idea delle connessioni tra i minimi, la disposizione lungo l'asse delle  $x$  serve solo ad aumentarne la chiarezza.

che la connessione fosse diretta o meno attraverso la tecnica descritta in appendice B. Nelle figure (4.10), (4.11) e (4.12) sono riportati i grafi corrispondenti alle tre sequenze esaminate:  $S_0$ ,  $S_1$  e  $S_2$ . È immediatamente evidente dalle figure la dif-

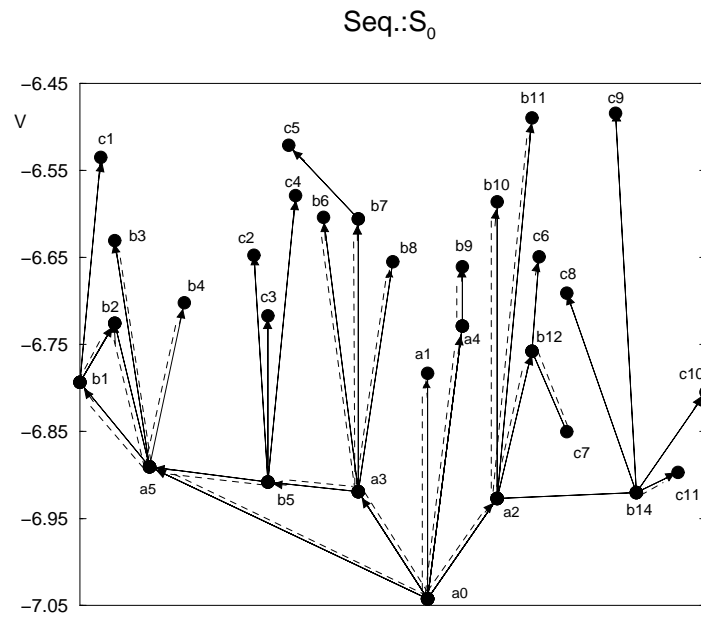


Figura 4.10.

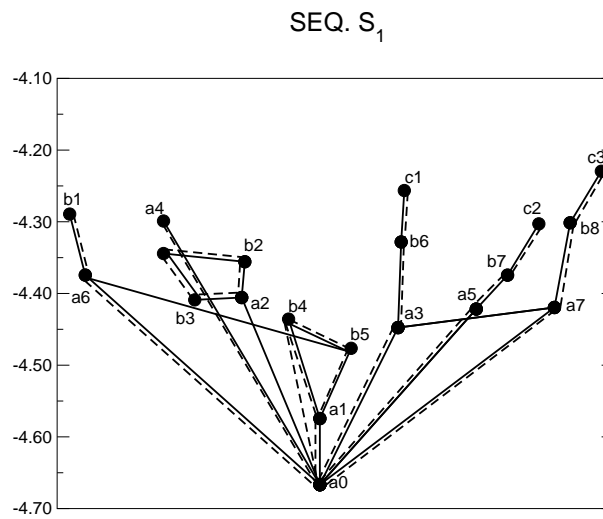


Figura 4.11.



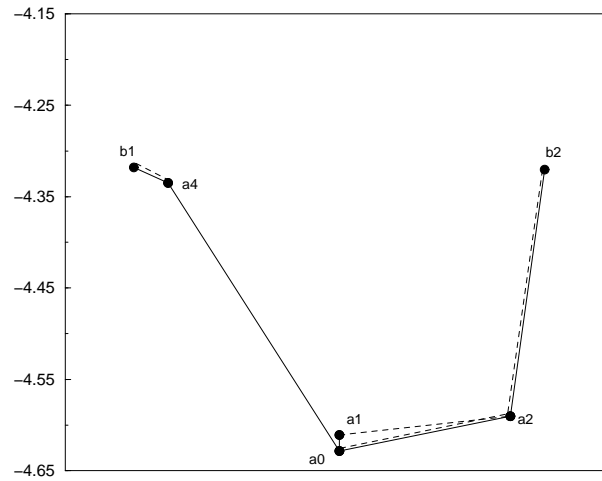
Seq.: $S_2$ 

Figura 4.12.

ferenza tra il grafo della sequenza  $S_2$  e gli altri grafi. Nel caso della  $S_2$  si osserva che il minimo nativo è connesso a pochissimi altri minimi i quali a loro volta sono connessi a pochissimi altri. Questa è una chiara indicazione che il sistema, a partire da una configurazione iniziale troverà con estrema difficoltà il “cammino” dinamico che attraverso successivi ripiegamenti la porterebbe ad assumere la forma della CN. Questo perché esisterà un numero molto limitato di cammini nello spazio delle fasi che connettono una configurazione aperta (*coil*) alla CN e la probabilità che questo avvenga su tempi non astronomici è molto bassa. Ciò conferma quanto osservato nel capitolo 3 dove la sequenza  $S_2$  è stata classificata come un cattivo ripiegatore. Per quanto riguarda le sequenze  $S_0$  e  $S_1$  la distinzione fra di esse, sulla base dei grafi realizzati, è meno evidente. Per una completa caratterizzazione dello spazio delle fasi di queste sequenze, data la ricchezza di minimi fra loro direttamente connessi,

sarebbe necessario indagare un numero molto più ampio di minimi e le loro reciproche connessioni. Questo studio va al di là degli scopi di questa tesi ma è già stato programmato per un prossimo futuro.

Vediamo comunque di trarre alcune conclusioni sulla base dei dati a nostra disposizione. Una prima osservazione è che nel caso della sequenza  $S_0$  si osservano molti più minimi, a parità di variazione energetica, che nel caso della  $S_1$ . Inoltre la distanza angolare media tra i minimi del grafo della  $S_0$  è pari a  $\langle \delta^0 \rangle \sim 0.45$ , valore che risulta maggiore di quello ottenuto per la  $S_1$  ( $\langle \delta^0 \rangle \sim 0.30$ ). Questo sta a indicare come, anche una volta entrato nella valle nativa, il sistema possa assumere configurazioni che differiscono maggiormente dalla CN nel caso della sequenza  $S_0$  rispetto al caso della  $S_1$ . Questo risultato conferma l'analisi del fattore di ripiegamento  $Q$  riportata nel paragrafo 3.7. Si è osservato inoltre che i minimi a energia più bassa nel caso della  $S_1$  sono direttamente connessi alla CN; ciò non accade nel caso delle altre due sequenze  $S_0$  e  $S_2$ . Possiamo perciò supporre che la  $S_0$  si comporti come un cattivo ripiegatore poiché la sua valle nativa è connessa facilmente a molte altre valli ove il sistema può restare intrappolato. Pensiamo che ciò non accada per la  $S_1$  ma, come già detto, per chiarire definitivamente questo punto saranno necessarie nuove analisi.

# Conclusioni

L'attività di ricerca originale descritta in questa tesi ha avuto come scopo quello di mettere in luce e caratterizzare, attraverso lo studio della dinamica di modelli “mesoscopici”, i meccanismi che presiedono al ripiegamento delle proteine.

A tal scopo abbiamo considerato un modello “fuori reticolo” in due dimensioni spaziali, ove la proteina viene schematizzata come una catena unidimensionale in cui i “residui” sono sostituiti da elementi localizzati con carattere idrofobico o polare che interagiscono tra loro tramite termini di forza efficace sia a corto che a lungo raggio. Un dato eteropolimero è quindi identificato univocamente dalla sequenza dei residui idrofobici o polari lungo la catena.

In particolare, ci siamo concentrati sulla caratterizzazione delle proprietà strutturali, termodinamiche e dinamiche che distinguono le sequenze “buone”, che cioè in tempi relativamente brevi si ripiegano sempre nella stessa struttura nativa, da tutte le altre sequenze che non mostrano tali caratteristiche (indicate come “cattive”).

L'analisi con tecniche di tipo Monte Carlo (MC) di questo modello [24] aveva messo in luce già in precedenza un aspetto assai interessante: solo poche sequenze si ripiegano su scale di tempi “ragionevoli” mentre la maggioranza necessita di tempi che divergono esponenzialmente col numero dei residui considerati (analogamente a

quanto avviene per le proteine reali).

Come primo punto abbiamo trovato che le tecniche di dinamica molecolare a temperatura costante rappresentano una valida alternativa alle tecniche MC per lo studio di questi sistemi complessi, in quanto siamo stati in grado di ottenere con estrema precisione sia le strutture native, che la distribuzione dei minimi locali dell'energia potenziale ricavati in precedenza con simulazioni MC. Il ricorso alla dinamica molecolare è però preferibile dato che non permette solo di ricavare la struttura del panorama energetico ma anche di seguire nel tempo il percorso della proteina verso lo stato nativo. Si è proceduto quindi a studiare l'evoluzione del modello in presenza di un bagno termico (insieme canonico). Gli studi di dinamica sono stati effettuati sia osservando il ripiegamento del sistema a partire da condizioni iniziali casuali, sia studiando la dinamica di *unfolding*, cioè le modalità e le scale dei tempi su cui la proteina esce dalla "valle nativa" a causa delle fluttuazioni di temperatura. Questi studi hanno riguardato catene composte da 20 residui, esaminate per 3 sequenze distinte a varie temperature.

Dall'andamento dell'energia e del calore specifico in funzione della temperatura è stato poi possibile individuare una prima temperatura (equivalente alla transizione  $\theta$  per gli omopolimeri) che corrisponde alla "transizione" da uno stato in cui prevalgono le configurazioni aperte ad uno stato ove la proteina per la maggior parte del tempo sta in forma compatta. Queste due fasi sono state individuate in tutte le sequenze esaminate. Alla temperatura  $T_\theta$  che le separa tutte presentano un massimo (più o meno pronunciato) del calore specifico. Abbiamo poi dato una definizione dinamica della temperatura di *folding*  $T_f < T_\theta$ , cioè della temperatura a cui la proteina

inizia a passare la maggior parte del tempo nella sua “valle nativa”. Dal confronto dei risultati di *folding* ed *unfolding* è poi stato possibile definire la temperatura di transizione vetrosa  $T_g$ , al di sotto della quale si osserva un arresto strutturale del sistema. Nel caso di una “buona” sequenza il rapporto  $T_f/T_g$  è maggiore che per le “cattive” sequenze; questo indica che esiste un intervallo di temperature nel quale le proteine possono ripiegarsi verso la conformazione nativa senza essere rallentate significativamente dal rallentamento dinamico tipico della fase vetrosa.

Inoltre, per una “buona” sequenza alla temperatura di *folding* si osserva un comportamento caratteristico : la proteina si avvicina attraverso una successione di salti bruschi alla configurazione nativa. Ad ogni salto è associata una netta riduzione sia della distanza media dalla struttura nativa che dell’energia potenziale, in particolare dell’energia associata all’interazione di tipo Lennard-Jones.

D’altra parte, per una “cattiva” sequenza si osservano variazioni brusche nelle osservabili misurate, ma non si nota una decrescita né della energia potenziale né della distanza dallo stato a minima energia; le osservabili fluttuano semplicemente nel tempo attorno ai loro valori medi.

Dalle nostre stime, le scale di tempo tipiche per il ripiegamento sono dell’ordine di  $10^6 - 10^7$  unità temporali  $\tau$  (per  $T \sim T_f$ ), mentre nelle proteine reali tali scale di tempo sono a temperatura ambiente dell’ordine di  $10^{11} - 10^{12} \times \tau$  (essendo tipicamente  $\tau \sim 10^{-12}$  sec). Questa differenza in parte è ascrivibile al numero limitato di residui considerati (le proteine sono tipicamente più lunghe), ed in parte è sicuramente dovuto alla dimensionalità: nel piano esistono vincoli geometrici più forti che in 3D e questo permette all’eteropolimero di convergere più rapidamente verso

la configurazione nativa.

Ci siamo poi mossi nell'ottica di descrivere i processi di *folding* e *unfolding* di una proteina come una successione di transizioni dovute ad attivazioni termiche da un bacino di attrazione, associato ad un dato minimo dell'energia potenziale, ad un bacino adiacente. Per dare una descrizione completa della dinamica delle proteine in questi termini non è sufficiente individuare tutti i minimi, ma è anche necessario associare ad ogni "salto" la corrispondente probabilità di transizione. Per la stima di queste quantità ci siamo rifatti ad una teoria assai raffinata dovuta a Langer [34] che tiene conto per il calcolo di queste probabilità non solo dell'altezza della barriera che separa due minimi (come nella usuale teoria di Kramers [36]) ma include, anche se in modo approssimato, i termini entropici, termini che dipendono dalla forma assunta dalla proteina prima e durante il salto, mentre viene valicato il punto sella che separa i due minimi coinvolti nella transizione. Al riguardo abbiamo messo a punto una tecnica numerica per il calcolo delle altezze delle barriere energetiche che separano i vari minimi visitati dalla proteina durante la sua dinamica. Vogliamo notare come questa tecnica sia abbastanza generale e possa essere applicata proficuamente a tutti i modelli a molti corpi in cui siano presenti paesaggi energetici complessi (come ad esempio nei vetri).

Ottenute le informazioni strutturali riguardanti il paesaggio energetico abbiamo potuto stimare, per varie transizioni fra minimi contigui, le corrispondenti probabilità secondo la formula di Langer. Abbiamo poi verificato che, almeno in prossimità della temperatura di transizione, tale stima teorica è in buon accordo con i dati derivati direttamente dalla dinamica del modello. Questa stima rappresenta la prima

verifica accurata mai fatta della teoria di Langer per un sistema assai complesso, quale è una proteina. Infatti, in letteratura [37, 38], abbiamo trovato solo un utilizzo *tout-court* della formula.

Con lo scopo di dare una rappresentazione il più fedele possibile del paesaggio di energia in prossimità della configurazione nativa, abbiamo costruito un “grafo” che mostra le connessioni della struttura nativa con tutti i minimi suoi primi, secondi e terzi vicini. Cioè quei minimi che sono raggiungibili dallo stato nativo valicando un punto sella, due punti sella o tre punti sella consecutivamente. Nel caso di sequenze “cattive” la struttura del paesaggio energetico è tale che la valle nativa è di fatto connessa con pochissimi altri minimi. Ciò rende assai poco probabile che partendo da una struttura aperta qualsiasi la proteina riesca a trovare un cammino (rappresentato da una successione di minimi fra loro connessi) che la porti verso la CN. Invece, nel caso di sequenze “buone”, la CN è connessa a molti più cammini che possono portare una configurazione casuale a ripiegarsi nella struttura nativa. Nel prossimo futuro e con lo scopo di giungere ad una pubblicazione scientifica di questi risultati, ci proponiamo di integrare l’analisi arrivando a considerare qualche centinaio di minimi connessi alla CN calcolando le corrispondenti probabilità di transizione. Ciò permetterà di completare la descrizione della dinamica di ripiegamento di una proteina in termini di processi Markoviani.

Le metodologie e gli indicatori messi a punto in questo lavoro di tesi per discriminare una “buona” da una “cattiva” sequenza tramite un’approccio dinamico rimangono valide anche per modelli più realistici, a cui possono essere parimenti applicate.

# Appendice A

## Stima dei tempi di scala reali

In questa appendice introdurremo, nel nostro modello adimensionale, le unità di misura che caratterizzano le interazioni delle proteine reali, con lo scopo di confrontare i tempi di ripiegamento osservati nelle nostre simulazioni con quelli sperimentali. Il termine d'interazione di Lennard-Jones nel nostro modello vuole riproporre per esempio l'attrazione idrofobica tra i monomeri. Usando l'espressione (1.1) di tale energia il tempo caratteristico associato  $\tau$  avremo che:

$$\tau \sim \left( \frac{m\sigma^2}{\epsilon} \right)^{\frac{1}{2}} ; \quad (\text{A.1})$$

dove  $m$  è la massa dell'aminoacido,  $\epsilon$  la profondità della buca di energia e  $\sigma$  è circa la distanza di equilibrio. Nel nostro modello si considerano  $m$ ,  $\sigma$  e  $\epsilon$  pari a uno mentre se prendiamo le grandezze reali dovremo considerare che la massa di un aminoacido è circa pari a  $2-3 \cdot 10^{-25} kg$ , la distanza di equilibrio è  $7-9 \text{ \AA}$  e l'energia d'interazione idrofobica è circa  $0.04 \cdot eV$ . Sostituendo nell'equazione (A.1) si ha che  $\tau \sim 6 \cdot 10^{-12}$  sec e quindi i tempi di ripiegamento da noi osservati ( $10^6 - 10^7$  unità di tempo naturali) corrispondono a un tempo dell'ordine di  $10^{-5}$  sec.



# Appendice B

## Algoritmo per l'identificazione dei cammini principali

In questa appendice riportiamo una schematica descrizione dell'algoritmo numerico impiegato per l'identificazione del cammino principale che connette due minimi dell'energia potenziale. Questo algoritmo può trovare utile applicazione per lo studio dei profili energetici di sistemi fisici a molti corpi.

Prendiamo l'esempio concreto di due configurazioni corrispondenti a due minimi  $\mathbf{a}_0$  e  $\mathbf{a}_1$  dell'energia potenziale.

Date le invarianze dell'energia potenziale del modello che consideriamo, ogni configurazione è individuata a meno di una rotazione attorno a un'asse perpendicolare al piano  $(xy)$ , di una traslazione rigida nel piano  $(xy)$  e della riflessione rispetto a un piano perpendicolare al piano  $(xy)$ . Una volta fissata dalle coordinate spaziali  $(q_x^0(i), q_y^0(i))$  (con  $i = 1 \dots N$ ) una data orientazione arbitraria di  $\mathbf{a}_0$ , si considera

quella particolare configurazione  $(q_x^1(i), q_y^1(i))$  che si ottiene ruotando, traslando e riflettendo  $\mathbf{a}_1$  in modo tale da minimizzare la distanza euclidea da  $\mathbf{a}_0$ . Il programma che individua il cammino principale tra questi due minimi può essere diviso in tre fasi che mostriamo nelle seguenti sezioni.

**1. Scelta delle configurazioni iniziali.** Nella prima fase iniziale si considera il cammino  $\{\mathcal{C}_0\}$  che unisce i due minimi  $\mathbf{a}_0$  e  $\mathbf{a}_1$ . Il cammino viene scelto nel seguente modo: si considera il cammino rettilineo  $\mathcal{C}_0$  cioè quel cammino costituito, al variare di  $r$ , dalle configurazioni i cui monomeri hanno coordinate  $(q_x(i), q_y(i))$  date da:

$$\begin{aligned} q_x(i) &= q_x^0(i) + r(q_x^1(i) - q_x^0(i)) \\ q_y(i) &= q_y^0(i) + r(q_y^1(i) - q_y^0(i)) \end{aligned} \tag{B.1}$$

dove  $r$  è una variabile continua compresa tra 0 e 1. Infatti per  $r = 0$  la configurazione  $(q_x(i), q_y(i))$  coincide con  $\mathbf{a}_0$  e per  $r = 1$  con  $\mathbf{a}_1$ . Lungo tale cammino si considera l'insieme finito delle configurazioni  $(q_x^m(i), q_y^m(i))$  ottenute al variare di  $m$  ( $0 \leq m \leq M$ ):

$$\begin{aligned} q_x^m(i) &= q_x^0(i) + \frac{m}{M}(q_x^1(i) - q_x^0(i)) \\ q_y^m(i) &= q_y^0(i) + \frac{m}{M}(q_y^1(i) - q_y^0(i)). \end{aligned}$$

Il numero  $M$  di configurazioni che campionano ogni cammino è stato preso pari a 100.

2. **Fase di evoluzione.** Una volta fissate le configurazioni del cammino iniziale, si inizia la seconda fase detta di evoluzione dove si fa evolvere ciascuna configurazione mediante una dinamica sovrasmorzata:

$$\begin{aligned}\dot{q}_x(i) &= -\frac{1}{\tilde{\gamma}} \frac{\partial H}{\partial q_x(i)} \\ \dot{q}_y(i) &= -\frac{1}{\tilde{\gamma}} \frac{\partial H}{\partial q_y(i)}.\end{aligned}$$

La costante d'attrito  $\tilde{\gamma}$  in realtà può anche non comparire esplicitamente, infatti aumentare lo smorzamento sul sistema è equivalente a diminuire l'intervallo di tempo d'integrazione  $\delta t$ . Inizialmente si esegue il calcolo di tutte le forze che agiscono sui monomeri di ogni configurazione appartenenti al cammino iniziale e se ne considera il valore massimo  $F_{max}$ . La scelta di  $\delta t$  viene eseguita in modo tale che il prodotto  $F_{max} \cdot \delta t$  sia pari a una costante. Se infatti un monomero di una configurazione è soggetto a una forza molto grande un tempo d'integrazione  $\delta t$  non sufficientemente piccolo potrebbe far esplodere l'energia in conseguenza, per esempio, a un avvicinamento eccessivo di due monomeri. Al solito un tempo d'integrazione piccolo rallenta la durata del calcolo ed è per evitare di utilizzare un  $\delta t$  molto piccolo anche quando non ce ne sia bisogno che si esegue il controllo all'inizio di ogni fase evolutiva e si sceglie  $\delta t$  (o analogamente  $\tilde{\gamma}$ ) in base alle forze in gioco. Una volta fatto evolvere il sistema per un numero limitato di passi d'integrazione si passa alla fase d'interpolazione.

3. **Fase di interpolazione.** Nella terza fase, quella di interpolazione, si considerano le nuove configurazioni ottenute dopo la fase di evoluzione e si calcola quindi la distanza euclidea tra configurazioni successive lungo il cammino

$\Delta_{m,m+1}$ . Se  $\Delta_{m,m+1}$  è maggiore di una certa soglia che prendiamo pari a  $2\Delta$ , ( $\Delta$  è la distanza che intercorre tra due configurazioni successive lungo il cammino iniziale) si inserisce una nuova configurazione ottenuta per interpolazione lineare dalle due configurazioni. Se invece  $\Delta_{m,m+1}$  è minore di  $\frac{\Delta}{2}$  si elimina la configurazione  $(m+1)$ -esima. In questo modo si genera un nuovo insieme di configurazioni spaziate in modo approssimativamente uniforme.

A ogni fase di evoluzione ne segue una d'interpolazione e così via fino a quando la somma del valore assoluto di tutte le forze agenti sui monomeri della configurazione di massima energia è minore di una certa soglia che abbiamo fissato pari a  $10^{-3}$ . Tale configurazione viene dunque memorizzata come quella del punto sella.

# Appendice C

## Probabilità di fuga nel caso unidimensionale

In questa appendice deriveremo in dettaglio l'espressione per la probabilità per unità di tempo  $\Gamma$  nel caso unidimensionale. Seguendo [36] la soluzione  $P(x,y)$  della versione stazionaria dell'equazione:

$$\partial_t P(x,v,t) = -v \frac{\partial P}{\partial x} - \frac{F(x)}{m} \frac{\partial P}{\partial v} + \gamma \frac{\partial}{\partial v} (PV) + \frac{1}{2} \frac{\partial^2}{\partial v^2} \left( \gamma \frac{KT}{m} \right) P; \quad (\text{C.1})$$

ottenuta come perturbazione rispetto alla soluzione stazionaria avrà la forma:

$$P(x,v) = Q(x,v) \exp\left[ \frac{mw_s^2(x-s)^2 - mv^2}{2kT} \right].$$

Sostituendo nella (C.1) avremo che  $Q(x,v)$  dovrà soddisfare l'equazione (ponendo  $s=0$ ):

$$-v \frac{\partial Q}{\partial x} - w_s^2 x \frac{\partial Q}{\partial v} - \gamma v \frac{\partial Q}{\partial v} + \frac{\gamma kT}{m} \frac{\partial^2 Q}{\partial v^2} = 0; \quad (\text{C.2})$$

assumiamo che  $Q(x,v)$  sia funzione di una sola variabile;  $Q(x,v) = f(z) = f(v - \alpha x)$ ,

sostituendo nella (C.2) si ottiene:

$$(\alpha v - \omega_s^2 x - \gamma v) \frac{df}{dz} = -\frac{\gamma kT}{m} \frac{d^2 f}{dz^2}.$$

Questa equazione può essere soddisfatta solo se il coefficiente che moltiplica  $\frac{df}{dz}$  è funzione solo di  $z$ , per cui imponendo  $\alpha v - \omega_s^2 x - \gamma v = \lambda(v - \alpha x)$  si ha che:

$$\begin{cases} \alpha - \gamma = \lambda \\ \omega_s^2 = \alpha \lambda \end{cases} \rightarrow \begin{cases} \alpha_{\pm} = \frac{1}{2}(\gamma \pm \sqrt{\gamma^2 + 4\omega_s^2}) \\ \lambda_{\pm} = -\alpha_{\mp} \end{cases} \quad (C.3)$$

con questa scelta abbiamo:

$$\lambda z \frac{df}{dz} = -\frac{\gamma kT}{m} \frac{d^2 f}{dz^2}$$

la cui soluzione generale è definita a meno di due costanti d'integrazione:

$$f(z) = B' + C' \int^z \exp\left[-\frac{\lambda m z'^2}{2\gamma kT}\right] dz'.$$

Imponendo che per  $x < 0$ ,  $P(x,v)$  coincida con la soluzione stazionaria e svanisca per  $x \rightarrow +\infty$  e  $v < 0$  si deve prendere  $\alpha$  positiva e sostituire il limite d'integrazione inferiore con  $-\infty$ . Si determinano così le costanti  $B'$  e  $C'$  e si ottiene un'espressione approssimata di  $P(x,v)$  valida per qualsiasi  $x$ .

$$P(x,v) = \frac{\omega_a \omega_s}{\sqrt{\alpha_+ \gamma}} \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} e^{-\frac{W}{kT}} \exp\left[\frac{m\omega_s x^2 - mv^2}{2kT}\right] \int_{-\infty}^{v-\alpha_+ x} \exp\left[-\frac{\lambda m z'^2}{2\gamma kT}\right] dz'$$

La probabilità di fuga è data dal flusso netto di  $P(x,v)$  attraverso la cima  $x = s = 0$ :

$$\Gamma = \int_{-\infty}^{+\infty} v P(0,v) dv = \frac{\omega_a \omega_s}{2\pi \alpha_+} e^{-\frac{W}{kT}}; \quad (C.4)$$

con  $\alpha_+$  dato dalla (C.3).

# Appendice D

## Probabilità di fuga nel caso multidimensionale

In questa appendice deriveremo in dettaglio l'espressione per la probabilità per unità di tempo  $\Gamma$  nel caso multidimensionale. L'equazione di Fokker-Planck multidimensionale per le sole variabili spaziali ha la forma:

$$\frac{\partial P(\vec{x},t)}{\partial t} = \sum_{i=1}^{dN} I_i \frac{\partial}{\partial x_i} \left( \frac{\partial V}{\partial x_i} P(\vec{x},t) + kT \frac{\partial P(\vec{x},t)}{\partial x_i} \right) \quad (\text{D.1})$$

che, definendo il vettore densità di corrente  $\vec{J}$  come:

$$J_i = -I_i \left( \frac{\partial V}{\partial x_i} P + kT \frac{\partial P}{\partial x_i} \right), \quad (\text{D.2})$$

può essere scritta nel seguente modo:

$$\frac{\partial P(\vec{x},t)}{\partial t} = -\vec{\nabla} \cdot \vec{J}.$$

Notiamo che la soluzione all'equilibrio della (D.1)

$$P_0(\vec{x}) = C \exp(-V(\vec{x})/kT) \quad (\text{D.3})$$

corrisponde a  $J_i = 0$  per qualunque  $i$ . Noi invece cerchiamo una soluzione stazionaria che descriva una corrente di probabilità finita nelle immediate vicinanze del punto sella.

Risulterà conveniente lavorare con un nuovo sistema di coordinate  $\xi_n$  con  $n = 1, \dots, dN$ , che ha origine in  $\vec{x}_s$  e tale che nello sviluppo al secondo ordine di  $V(\vec{\xi})$  in un intorno di  $\vec{\xi} = 0$  i termini misti delle derivate seconde siano nulli;

$$V(\vec{\xi}) = V_a + 1/2 \sum_{n=1}^{dN} \lambda_n \xi_n^2 + \dots$$

$\lambda_n$  autovalori della matrice Hessiana di  $V$  calcolata nel punto di sella  $\vec{\xi} = \vec{0}$ .

Per definizione di punto sella uno dei  $\lambda_n$  (ad esempio  $\lambda_1$ ) è negativo e il corrispondente autovettore  $\vec{\xi}_1$  individua la direzione espandente. Nel caso in cui il numero gradi di libertà del sistema sia minore della dimensione dello spazio  $\Omega$  avremo anche un certo numero di  $\lambda_n$  nulli.

Il vettore densità di corrente nel nuovo sistema di coordinate avrà componenti:

$$J'_n = - \sum_m I'_{nm} \left( \frac{\partial V}{\partial \xi_m} P + kT \frac{\partial P}{\partial \xi_m} \right)$$

con  $I'_{nm} = \sum_i D_{ni} I_i D_{mi}$  e  $D$  matrice associata al cambiamento di base. Se consideriamo  $I_i$  costante, come possiamo aspettarci supponendo che il bagno termico interagisca uniformemente con il nostro sistema, e considerando che le masse  $m_i$  delle particelle siano uguali tra loro, avremo  $I'_{nm} = I \delta_{nm}$  e l'equazione per il vettore densità di corrente diviene:

$$J'_n = -I \left( \frac{\partial V}{\partial \xi_n} P + kT \frac{\partial P}{\partial \xi_n} \right).$$

La versione stazionaria dell'equazione (D.1) avrà di nuovo la forma:

$$\sum_n \frac{\partial J'_n}{\partial \xi_n} = 0$$



che, ponendo  $P(\vec{\xi}) = Q(\vec{\xi})\exp(-V/kT)$ , diviene:

$$IkT \sum_n \frac{\partial}{\partial \xi_n} \left( \frac{\partial Q}{\partial \xi_n} e^{-V/kT} \right) = 0 \rightarrow \vec{J}' = -IkT(\vec{\nabla}Q)e^{-V/kT}. \quad (\text{D.4})$$

La funzione  $Q(\vec{\xi})$  indica la perturbazione rispetto allo stato di equilibrio del sistema. Per ottenere una densità di corrente  $\vec{J}'$  non trascurabile, dovremo avere che nelle vicinanze del punto sella, dove  $e^{-V/kT}$  è piccolo,  $\vec{\nabla}Q$  assuma valori relativamente grandi mentre  $Q$  può essere ritenuta praticamente costante vicino ai punti di minimo. Possiamo quindi aspettarci che cambiamenti rilevanti dell'andamento tipo Boltzmann per  $P(\vec{\xi})$  avvengano solo in prossimità della sella.

In questa zona si utilizza l'approssimazione quadratica per  $V$  e si assume che  $\vec{J}'$  abbia direzione costante parallela alla direzione espandente  $\xi_1$ ; quest'ultima forte assunzione fa dipendere  $Q$  solo da  $\xi_1$  e fa sì che la (D.4) diventi un'equazione nella sola variabile  $\xi_1$ ;

$$IkT \left( \frac{\partial^2 Q}{\partial \xi_1^2} - \frac{\lambda_1 \xi_1}{kT} \frac{\partial Q}{\partial \xi_1} \right) = 0,$$

la cui soluzione non costante ha la forma:

$$Q(\xi_1) = \int d\xi_1 \exp\left(\frac{\lambda_1 \xi_1^2}{2kT}\right),$$

che è accettabile fisicamente solo se  $\lambda_1$  è negativa, cosa che accade avendo preso come direzione di  $\vec{J}'$  quella espandente.

La funzione  $P(\vec{\xi})$  dovrà avere le seguenti proprietà:

- per  $\xi_1 < 0$ ,  $P$  deve essere molto simile alla distribuzione di equilibrio (D.3) centrata nel minimo iniziale;

- per  $\xi_1 > 0$ ,  $P$  deve tendere velocemente a zero;
- deve passare in maniera continua dal valore di equilibrio  $P_0$  a zero dentro una regione limitata intorno alla sella di larghezza  $\Delta\xi_1 \simeq \sqrt{\frac{kT}{|\lambda_1|}}$ , nella quale si concentrerà il flusso di  $\vec{J}$ .

Sulla base di queste considerazioni scegliamo

$$Q(\xi_1) = \frac{1}{Z_0} \sqrt{\frac{|\lambda_1|}{2\pi kT}} \int_{\xi_1}^{+\infty} \exp\left(\frac{\lambda_1 u^2}{2kT}\right) du$$

con  $Z_0$  la costante di normalizzazione che si ottiene imponendo che nella regione del minimo di partenza  $P$  sia normalizzata a 1;

$$Z_0 = \int \exp\left(-\frac{V(\vec{x})}{kT}\right) d\vec{x} \simeq e^{-V_a/kT} \prod_{n=1}^M \left(\frac{2\pi kT}{\lambda_n^a}\right)^{1/2}, \quad (\text{D.5})$$

$M$  è il numero di  $\lambda$  non nulli cioè il numero di gradi di libertà effettivo del sistema. La probabilità che il sistema esegua la transizione dal minimo  $\vec{x}_a$  al minimo  $\vec{x}_b$  è data dal flusso di  $\vec{J}$  attraverso una qualunque superficie non parallela a  $\vec{J}$  che tagli il punto sella, (prendiamo ad esempio la superficie perpendicolare  $\xi_1 = 0$ )

$$\Gamma = \int J'_1|_{\xi_1=0} d\xi_2 d\xi_3 \dots;$$

svolvendo i calcoli troviamo l'espressione

$$\Gamma = \frac{IkT}{Z_0} \sqrt{\frac{|\lambda_1|}{2\pi kT}} e^{-V_s/kT} \prod_{n'=2}^M \left(\frac{2\pi kT}{\lambda_{n'}}\right)^{1/2}$$

da cui, esprimendo  $Z_0$  tramite la (D.5) si ha che:

$$\Gamma = \frac{I}{2\pi} \sqrt{|\lambda_1|} e^{-W/kT} \frac{\prod_{n=1}^M (\lambda_n^a)^{1/2}}{\prod_{n'=2}^M (\lambda_{n'})^{1/2}}. \quad (\text{D.6})$$

Volendo riscrivere la (D.6) in termini delle frequenze  $\omega_i = \sqrt{|\lambda_i|/m}$ , e considerando che  $I = (m\gamma)^{-1}$  avremo finalmente l'espressione per la probabilità di fuga:

$$\Gamma = \frac{\omega_1}{\gamma} \left[ \frac{1}{2\pi} e^{-W/kT} \frac{\prod_{n=1}^M (\omega_n^a)}{\prod_{n'=2}^M (\omega_{n'})} \right]. \quad (\text{D.7})$$

# Bibliografia

- [1] C. Levinthal, *Are there pathways for protein folding?*, J. Chim. Phys. 65, 44-45 (1968)
- [2] E. Shakhnovich, *Proteins with Selected Sequences Fold into Unique Native Conformation*, Phys. Rev. Lett. 72, 3907-3910 (1991)
- [3] T.E. Creighton, **Proteins: Structure and Molecular Properties**, W.H. Freeman, New York (1983)
- [4] C. Branden e J. Tooze, **Introduzione alla Struttura delle Proteine**, Zanichelli, Bologna (1993)
- [5] G. Rotilio, **Struttura e Funzione delle Proteine**, Nis, Bologna (1993)
- [6] J.N. Israelachvili, **Intermolecular and surface forces**, Academic Press, London (1985)
- [7] L. Stryer, **Biochimica**, Zanichelli, Bologna (1994)
- [8] J.T. Kellis Jr, K. Nyberg, Daša Šali e A.R. Fersht, *Contribution of hydrophobic interactions to protein stability*, Nature 333, 784-786 (1988)
- [9] P.L. Privalov e S.J. Gill, *Stability of protein structure and hydrophobic interaction*, Adv. Prot. Chem. 39, 191-234 (1988)
- [10] C. Anfisen, *Principles that govern the folding of protein chains* Science 181, 223 (1973)
- [11] P.L. Privalov e N.N. Khechinashvili, *A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study*, J. Mol. Biol. 86, 665-684 (1974)

- [12] G.I. Makhatadze e P.L. Privalov, *Heat capacity of proteins. 1. Partial molar heat capacity of individual amino acid residues in aqueous solution: Hydration effect*, J. Mol. Biol. 213, 375-384 (1990)
- [13] P.G. de Gennes, **Scaling Concept in Polymer Physics**, Cornell University Press, London (1979)
- [14] S. Fujishige, K. Kubota e I. Anto, *Phase transitions of aqueous solutions of poly(N-isopropylacrylamide) and poly(N-isopropylmethacrylamide)*, J. Phys. Chem. 93, 3311-3313 (1989)
- [15] K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas e H.S. Chan, *Principles of protein folding- A perspective from simple exact models*, Prot. Science 4, 561-602 (1996)
- [16] H.S. Chan e K.A. Dill, *Transition states and folding dynamics of proteins and heteropolymers*, J. Chem. Phys. 100, 9238-9257 (1994)
- [17] K. Yue e K.A. Dill, *Forces of tertiary structural organization of globular proteins*, Proc. Natl. Acad. Sci. USA 92, 146-150 (1995)
- [18] H.S. Chan e K.A. Dill, *The Protein Folding Problem*, Phys. Today 46, 24-32 (1993)
- [19] H. Taketomi, Y. Hueda e N. Gō, *Studies on protein folding, unfolding and fluctuation by computer simulation*, Int. J. Peptide Protein Res. 7, 445-459 (1975)
- [20] N. Gō e H. Taketomi, *Respective roles of short- and long-range interaction in protein folding*, Proc. Natl. Acad. Sci. USA 75, 559-563 (1978)
- [21] F.H. Stillinger, T.H. Head-Gordon e C. Hirshfeld, *Toy model for protein folding*, Phys. Rev. E 48, 1469-1477 (1993)
- [22] T.H. Head-Gordon e F.H. Stillinger, *Optimal neural networks for protein-structure prediction*, Phys. Rev. E 48, 1502-1515 (1993)
- [23] A. Irbäck, C. Peterson e F. Potthast, *Identification of Amino Acid in an Off-Lattice Model*, Phys. Rev. E 55, 860-874 (1997)
- [24] A. Irbäck e F. Potthast, *Studies of an off-lattice model for protein folding: Sequence dependence and improved sampling at finite temperature*, J. Chem. Phys. 103, 10298-10305 (1995)

- [25] P.E. Leopold, M. Montal e J.N. Onuchic, *Protein folding funnels: A kinetic approach to the sequence-structure relationship*, Proc. Natl. Acad. Sci. USA 89, 8721-8725 (1992)
- [26] J.N. Onuchic, P.G. Wolynes, Z. Luthey-Schulten e N.D. Socci, *Toward an outline of the topography of a realistic protein-folding funnel*, Proc. Natl. Acad. Sci. 92, 3626-3629 (1995)
- [27] P.G. Wolynes, J.N. Onuchic e D. Thirumalai, *Navigating the Folding Routes*, Science 267, 1619-1620 (1995)
- [28] D.J. Evans e G.P. Morriss, **Statistical Mechanics of Nonequilibriums Liquids**, Academic Press, London (1990)
- [29] H. Press, S.A. Teukolsky, W.T. Vetterling e B.P. Flannery, **Numerical Recipes in Fortran**, Cambridge University Press, New York (1992)
- [30] C.M. Johnson, U. Mayor, H. Oschkinat e A.R. Fersht, *Protein folding on a microsecond time scale*, contributo presentato alla conferenza: “*Protein Folding: Simple models and experimenti*” Torino, maggio 2000.
- [31] F.H. Stillinger e T.A. Weber, *Packing Structure and transitions in liquids and solids*, Science 225, 983 (1984)
- [32] V.I. Arnold e A. Avez, **Ergodic problems of classical mechanics**, Benjamin, New York (1968)
- [33] P. Hanggi, *Escape from a Metastable State*, J. Stat. Phys. 42, 105-148 (1986)
- [34] J.S. Langer, *Statistical Theory of the Decay of Metastable States*, Ann. Phys. 54, 258-275 (1969)
- [35] N.G. van Kampen, **Stochastic processes in physics and chemistry**, North-Holland personal library, Amsterdam (1997)
- [36] H.A. Kramers, *Brownian motion in a field of force and the diffusion model of chemical reaction*, Physica 7, 284- (1940)
- [37] L. Angelani, G. Parisi, G. Ruocco, G. Vilianni, *Potential energy landscape and long-time dynamics in a simple model glass*, Phys. Rev. E 61, 1681-1691 (2000)
- [38] G. Ruocco, G. Daldoss, O.Pilla e C. Brangian, *Energy landscape, two-levels systems, and entropy barriers in Lennard-Jones clusters*, Phys. Rev. B 50, 3200-3205 (1999)