



Università Ca'Foscari Venezia

Corso di laurea magistrale in chimica e tecnologie sostenibili

Dinamica Molecolare di Biomolecole: validazione mediante dati sperimentali ed analisi delle componenti entropiche ed entalpiche del folding

Relatore:

Prof. Achille Giacometti

Correlatore:

Dott.ssa Tatjana Škrbić

Laureando:

Marco Mondin

Matricola 826624

Anno Accademico 2015-2016

Indice

1	Introduzione	1
2	Le proteine	4
2.1	Amminoacidi e legami	4
2.2	Strutture secondarie e superiori	10
2.3	Ripiegamento	18
2.3.1	Driving force	19
2.3.2	Problemi relativi al folding	22
2.4	Metodi sperimentali per la caratterizzazione della proteina	23
2.4.1	Cristallografia a raggi X	23
2.4.2	Spettroscopia NMR per le proteine	25
2.4.3	Calorimetria	29
3	Dinamica Molecolare	33
3.1	<i>Force field</i>	33
3.2	Modelli per l'acqua	34
3.3	Condizioni periodiche	36
3.4	Schema generale di una simulazione	36
3.5	Dettagli dell'implementazione	37
4	Validazione della dinamica	39
5	Componenti entropiche ed entalpiche del folding	46
5.1	Ciclo termodinamico	46
5.2	Dati sperimentali per la proteina 1UBQ	49
5.3	Risultati	50
5.3.1	Legami Idrogeno	50
5.3.2	Scorporamento dei vari termini entalpici aggiuntivi	51
6	Conclusioni e prospettive	61
	Appendices	

INDICE

Appendice A	Integrazione termodinamica	65
Appendice B	Tabella di conversione delle unità di misura per l'energia	67
Appendice C	Parametro di ordine nematico	68

Elenco delle figure

2.1	Concentrazione delle tre forme di de/protonate in funzione del pH. Immagine tratta da [7].	5
2.2	I 9 amminoacidi apolari. Immagine tratta da [7].	7
2.3	I 6 amminoacidi polari. Immagine tratta da [7].	8
2.4	I 5 amminoacidi polari e non neutri. Immagine tratta da [7].	8
2.5	La sequenza dell'ubiquitina, in blu i residui identici. Nell'ultima colonna è riportato il codice per il database NCBI. Immagine tratta da [8].	9
2.6	Rappresentazione planare del gruppo peptidico. Immagine tratta da [9].	9
2.7	In rosso gli angoli accessibili a tutti gli amminoacidi tranne la prolina. In giallo le coppie addizionali accessibili alla glicina. Immagine tratta da [10].	10
2.8	Rappresentazione schematica della formazione del ponte disolfuro.	10
2.9	α -elica destrogira. Immagine tratta da [9].	12
2.10	Un fogliettino- β che presenta una parte parallela in alto ed una anti-parallela in basso. Immagine tratta da [7].	13
2.11	Fogliettino- α . Immagine tratta da [11].	13
2.12	Amminoacido generico e la prolina nei duei isomeri cis e trans. Immagine tratta da [12].	14
2.13	β -turn di tipo I e II. Immagine tratta da [9].	15
2.14	Un esempio di Ω -loop, 1cvl:262-267. Immagine elaborata a partire da [15].	15
2.15	Alcuni motivi strutturali delle proteine. Immagine tratta da [9].	17
2.16	a) Collagene, una proteina fibrosa. L'unione delle tre eliche forma una struttura quaternaria. b) Mioglobina, una proteina globulare. In giallo i residui idrofobici, in blu i residui carichi, e i restanti in bianco. Immagine elaborata a partire da [7]. . .	18

ELENCO DELLE FIGURE

2.17	a) Denaturazione termica per la apomioglobina del cavallo e ribonucleasi A. b) Denaturazione della ribonucleasi ad opera di cloruro di guanidinio. Immagine tratta da [9].	19
2.18	Una rappresentazione ad imbuto del ripiegamento della proteina. Ad alte energia sono accessibili un numero maggiore di strutture, che diminuiscono via via che la proteina inizia a ripiegarsi. Lungo questo ripiegamento esistono degli intermedi metastabili con minimi locali di energia. Immagine tratta da [7].	20
2.19	A: legame idrogeno interno, B: legame ionico, C: contributo di vdW. immagine elaborata a partire da [17].	21
2.20	Spettro COSY ottenuto a 600MHz per la maiganina. Nel box, la fingerprint region. Immagine tratta da [25].	26
2.21	Popolazione dei livelli energetici dopo aver saturato la transizione X. Immagine tratta da [24].	27
2.22	Spettro NOESY per la stricnina. In alto è riportata la struttura della molecola con la legenda dei protoni. Immagine tratta da [26].	28
2.23	Rappresentazione schematica di un calorimetro a scansione differenziale. I triangoli sono amplificatori. Sono inoltre riportate la dipendenza della temperatura dal tempo e della flusso di calore dalla temperatura. Immagine modificata a partire da [27].	29
2.24	Calorimetro a titolazione isotermitica. Immagine tratta da [28].	31
2.25	Titolazione isotermitica di un pentamero di una sotto-unità della tossina del colera. Immagine tratta da [28].	31
3.1	Le tre componenti bonded del potenziale.	34
3.2	Due strutture per 1ubq. A sinistra la struttura nativa, a destra una delle possibile conformazioni della proteina denaturata. . .	38
4.1	RMSD per i primi 72 residui. In nero AMBER99, in rosso AMBER99sb, in verde AMBER99sb-ildn.	42
4.2	RMSF dei C $_{\alpha}$ della proteina. In nero AMBER99, in rosso AMBER99sb, in verde AMBER99sb-ildn.	42
4.3	Ubiquitina ripiegata con le zone mobili marcate.	43
4.4	RMSF della proteina mostrata insieme alle zone mobili della proteina.	43
4.5	Criteri geometrici per il legame idrogeno. D è un generico atomo donatore, A un generico atomo accettore. Immagine tratta dal manuale di GROMACS [40].	44

ELENCO DELLE FIGURE

4.6	Legami idrogeni intramolecolari e con l'acqua nelle 3 simulazioni. In nero AMBER99, in rosso AMBER99sb, in verde AMBER99sb-ildn.	45
5.1	Il ciclo termodinamico usato per studiare il sistema.	47
5.2	Dipendenza dalla temperatura della capacità termica. La linea tratteggiata rappresenta la capacità termica della proteina nella stato nativo, mentre la linea con i pallini rappresenta la capacità termica della proteina denaturata. Immagine tratta da [44].	49
5.3	Energia libera di Gibbs di <i>unfolding</i> in funzione del pH per l'ubiquitina. I simboli in bianco sono dati ottenuti per la denaturazione indotta da guanidina, in nero i dati ottenuti in assenza di denaturante. Nel primo caso i quadrati si riferiscono a dati ottenuti con dicroismo circolare mentre i pallini sono ottenuti da fluorescenza; nel secondo caso i quadrati si riferiscono a dati ottenuti da Ibarra-Molero, mentre i pallini sono i dati riportati da Wintrode. Immagine tratta da [45]. . .	50
5.4	Termine LJ fra la proteina nello stato nativo e l'acqua.	52
5.5	Termine coulombiano fra la proteina nello stato nativo e l'acqua.	52
5.6	Termine LJ fra la proteina denaturata e l'acqua.	53
5.7	Termine coulombiano fra la proteina denaturata e l'acqua.	53
5.8	Termine LJ per la proteina nella struttura nativa e denaturata, con $\Delta\lambda=0.01$	53
5.9	$dG/d\lambda$ per la proteina nello stato nativo con $\Delta\lambda=0.05$	54
5.10	$dG/d\lambda$ per la proteina nello stato nativo con $\Delta\lambda=0.02$	54
5.11	$dG/d\lambda$ per la proteina nello stato nativo con $\Delta\lambda=0.01$	54
5.12	$dG/d\lambda$ per la proteina nello stato denaturato con $\Delta\lambda=0.05$	55
5.13	$dG/d\lambda$ per la proteina nello stato denaturato con $\Delta\lambda=0.02$	55
5.14	$dG/d\lambda$ per la proteina nello stato denaturato con $\Delta\lambda=0.01$	55
5.15	I punti rossi rappresentano i valori di λ simulati, da cui si ottiene una curva ideale (in rosso). In nero la curva effettivamente ottenuta. Lungo l'asse y è riportato $\langle \partial H / \partial \lambda \rangle_\lambda$. Immagine tratta da [46].	56
5.16	$\int_0^\lambda \left(\frac{\delta G}{\delta \lambda'} \right) d\lambda'$ per la proteina nello stato nativo con $\Delta\lambda=0.05$	57
5.17	$\int_0^\lambda \left(\frac{dG}{d\lambda'} \right) d\lambda'$ per la proteina nello stato nativo con $\Delta\lambda=0.02$	57
5.18	$\int_0^\lambda \left(\frac{dG}{d\lambda'} \right) d\lambda'$ per la proteina nello stato nativo con $\Delta\lambda=0.01$	57
5.19	$\int_0^\lambda \left(\frac{dG}{d\lambda'} \right) d\lambda'$ per la proteina nello stato denaturato con $\Delta\lambda=0.05$	58

ELENCO DELLE FIGURE

5.20 $\int_0^\lambda \left(\frac{dG}{d\lambda'}\right) d\lambda'$ per la proteina nello stato denaturato con $\Delta\lambda=0.02$. 58

5.21 $\int_0^\lambda \left(\frac{dG}{d\lambda'}\right) d\lambda'$ per la proteina nello stato denaturato con $\Delta\lambda=0.01$. 58

5.22 La densità dell'acqua nel box in funzione della posizione. In rosso nel caso della proteina denaturata, mentre in verde è riportata la densità dell'acqua in presenza della proteina nello stato nativo. 59

5.23 Ubiquitina. In verde sono rappresentati il backbone e le strutture secondarie, mentre in rosso sono rappresentate la catene laterali. 60

Elenco delle tabelle

2.1	Angoli diedri ideali fra i piani in quattro tipi di β -turn [13].	14
2.2	Dettagli della proteina visibili in base alla risoluzione.	25
3.1	Parametri dei modelli SPC e TIPI3P per l'acqua.	35
4.1	Risultati statistici delle simulazioni.	40
4.2	Media dei legami idrogeno intramolecolari P-P, fra proteina ed acqua P-W, e totali.	44
4.3	Energia cinetica, entalpia, ed energia potenziale per le tre corse di 20 ns.	44
4.4	Volumi e densità per le tre corse di 20 ns.	45
5.1	Entropie ed entalpie sperimentali [44]. T_f è la temperatura a cui avviene la denaturazione della proteina.	49
5.2	Media dei legami idrogeno intramolecolari, P-P, fra proteina ed acqua, P-W, e totali.	50
5.3	Energia libera dei processi di deidratazione per le proteine nello stato nativo N e denaturato U, e di unfolding quando il sistema è disaccoppiato. Tutti i valori sono in kJ mol^{-1}	59
5.4	Confronto fra i valori da noi ottenuti ed i valori sperimentali per $T = 300 \text{ K}$. La Tutti i valori sono in kJ mol^{-1}	60
B.1	Tabella di conversione delle unità di misura dell'energia.	67

Capitolo 1

Introduzione

Il ruolo delle simulazioni numeriche è notevolmente cresciuto negli ultimi anni, tanto da diventare un elemento di fondamentale importanza in qualsiasi studio dei sistemi biologici. Ciò è dovuto sia al fatto che esse rappresentano un utile complemento agli studi sperimentali sia (e soprattutto) al fatto che esse possono essere fatte anche in condizioni in cui l'esperimento non sia facilmente eseguibile. In questa seconda accezione, la simulazione numerica riveste spesso l'importante ruolo di guida dell'esperimento stesso verso condizioni più favorevoli. All'interno del campionario di metodi di simulazione numerica, la dinamica molecolare occupa senz'altro un posto di primo piano nei sistemi biologici per le sue capacità predittive della dinamica del sistema. Un esempio importante del vantaggio dello studio a livello molecolare viene dallo studio del processo di solvatazione di un amminoacido. La metodologia tradizionale usa le scale idrofobiche, in cui ad ogni amminoacido è assegnato un valore che descrive la sua idrofobicità in rapporto agli altri amminoacidi. Fra gli altri problemi, con queste scale non si è in grado di misurare direttamente i contributi entropici al processo di idratazione e, di conseguenza, non si riescono ad avere informazioni sulla dinamica dell'acqua. Su scala maggiore, lo studio del processo del ripiegamento delle proteine presenta ulteriori difficoltà. La caratteristica fondamentale delle proteine è l'unicità della loro struttura nativa, che determina le funzioni delle proteine, che possono essere, ad esempio, di trasporto o strutturali. A causa dell'enorme quantità di interazioni e conformazioni, e conseguente alto costo computazionale, la simulazione diretta del ripiegamento di una proteina è possibile solo per piccole proteine e usando supercomputer dedicati [1]. Una soluzione, consapevoli di perdere la dinamica specifica del ripiegamento, è studiare il processo inverso, la denaturazione. In questo secondo caso, si parte dallo stato nativo e si ottiene lo stato denaturato (aperto), o mediante l'aumento della temperatura, oppure variando le proprietà di solvatazione del solvente. Tale processo

è fattibile in quanto si segue un solo percorso, evitando quindi l'immensa complessità della scelta del corretto percorso di *folding*. Il prezzo da pagare è naturalmente quello di un numero di informazioni molto inferiore. In questo quadro semplificato, è però possibile disaccoppiare alcune quantità che di solito non sono separatamente calcolabili: in particolare, è in linea di principio possibile calcolare separatamente i contributi entropici ed entalpici al processo di denaturazione, e quindi anche di *folding*, essendo entrambe funzioni di stato, e quindi indipendenti dal percorso. La separazione dei due contributi è una questione di interesse attuale, anche per sistemi piccoli come gli amminoacidi, come dimostrato da un importante articolo uscito nel corso del lavoro di questa tesi [2]. Partendo da questa idea, in questa tesi ci siamo proposti di compiere dei passi in direzione della completa caratterizzazione delle quantità termodinamiche del processo di ripiegamento. Questo metodo consiste nella sostituzione del passaggio diretto fra stato nativo e denaturato con due processi: il primo consiste in una desolvatazione della proteina nello stato nativo, il secondo in una solvatazione della proteina denaturata. I due processi sono uniti da una denaturazione della proteina desolvata. Questa metodologia ci consente di poter separare quindi l'energia libera in entalpia, misurabile a partire dai due stati iniziale e finale, in entropia associata al solvente, misurabile nella variazione del volume escluso, ed in una entropia della proteina. Data l'elevata importanza della quantificazione dei vari contributi, risulta cruciale essere certi di aver eseguito delle simulazioni di dinamica accurate. A tale scopo, per dare una validazione sperimentale allo strumento usato, nella prima parte della tesi si sono eseguiti dei calcoli di parametri d'ordine NMR del legame azoto-idrogeno per i quali esistono risultati sperimentali noti. Per questo si è scelto l'ubiquitina IUBQ [3], sia per la numerosità dei dati sperimentali disponibili sia per il suo interesse specifico, ancora intenso come dimostrato da un recente articolo sul tema [4]. La scoperta della proteolisi mediata dall'ubiquitina è stata riconosciuta con il premio Nobel per la chimica per l'anno 2004 [5]. Questa proteina ha assunto un'importanza notevole nella medicina degli ultimi 20 anni, anche in virtù di interessanti prospettive future nello studio dei meccanismi ubiquitina-dipendenti tumorali [6]. La validazione dello strumento numerico è stata poi sfruttata nella seconda parte della tesi per il calcolo delle parti entropiche ed entalliche del processo di *folding* della proteina stessa.

Piano della tesi

Nel secondo Capitolo presentiamo le proteine e gli amminoacidi. In particolare sono descritte le varie strutture secondarie derivate dalla sequenza degli amminoacidi. Sono presenti anche tre metodi sperimentali caratteristici del-

lo studio delle proteine. Nel Capitolo 3 sono presenti una descrizione della metodologia generale delle simulazioni di dinamica molecolare ed il metodo specifico impiegato in questa tesi. Il quarto Capitolo riporta i primi risultati di validazione ottenuti mediante un confronto fra i parametri d'ordine NMR sperimentali e simulati. Sono presenti inoltre dei risultati di dinamica. Il quinto Capitolo contiene una descrizione del ciclo termodinamico ideato per lo studio dei termini di ripiegamento della proteina, i risultati sperimentali che abbiamo usato per validare il nostro studio, e i risultati ottenuti dalle simulazioni. La tesi si chiude con le conclusioni e con le prospettive future.

Capitolo 2

Le proteine

Le proteine sono macromolecole di origine biologica, formate principalmente da azoto, ossigeno, carbonio ed idrogeno. Sono sintetizzate per svolgere una enorme varietà di funzioni, quali quella enzimatica, di trasporto, o come parete a livello cellulare. A livello di struttura le proteine sono polimeri lineari formati da una sequenza di monomeri, gli amminoacidi, che si ripiegano spontaneamente in una struttura tridimensionale unica per ogni sequenza amminoacidica. Nella forma ripiegata è possibile distinguere la formazione delle strutture secondarie ad opera di legami idrogeno e di strutture terziarie dovute ad interazioni fra amminoacidi vicini nello spazio. Le proteine sono in grado di formare strutture quaternarie per interazioni con altre proteine.

2.1 Amminoacidi e legami

Tutte le proteine sono prodotte a partire da 20 amminoacidi¹, che possono essere considerati come una sorta di alfabeto. Chimicamente gli amminoacidi sono costituiti da un carbonio centrale, denominato C_α , a cui sono legati un gruppo amminico, un gruppo carbossilico, un idrogeno ed un gruppo R detto catena laterale. Il carbonio centrale, essendo tetrasostituito, è chirale e potrebbe presentarsi in entrambi gli enantiomeri, ma (quasi) solo quello L è usato a livello biologico. Caso particolare è la glicina in quanto la catena laterale è costituita da un solo idrogeno e quindi non si ha chiralità. La struttura dell'amminoacido consente di trovarlo in un forma cationica protonata, zwitterionica, complessivamente neutra ma con cariche separate, oppure anionica deprotonata, a seconda del pH. In un pH fisiologico di 7,

¹In realtà altri 2 amminoacidi, la pirrolisina e la selenocisteina, sono usati. La prima è stata trovata in un enzima di un archabatterio, la seconda non è universale in quanto è assente in piante e funghi.

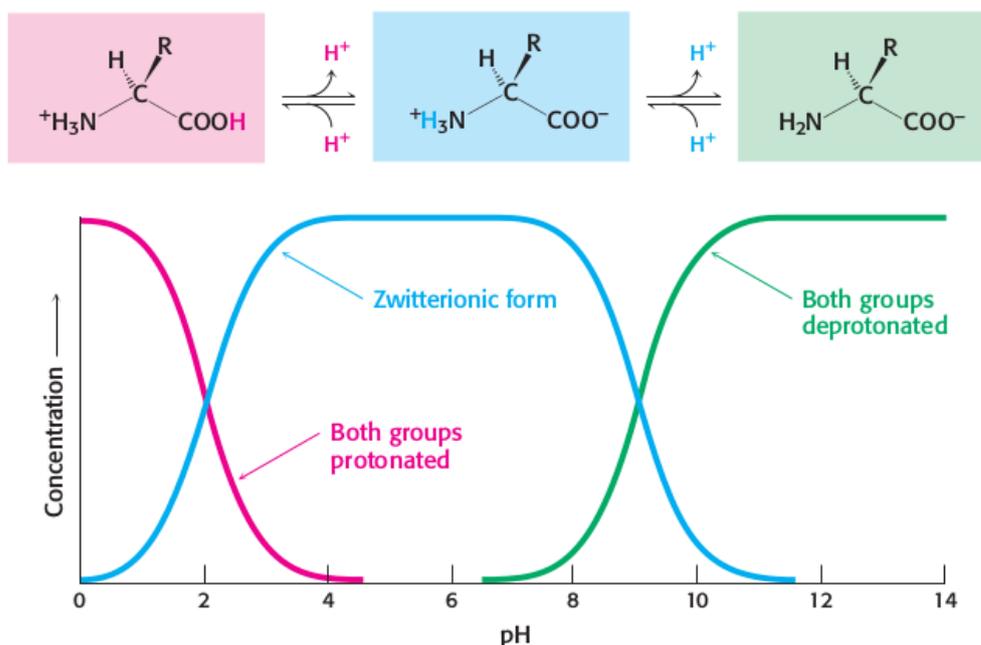


Figura 2.1: Concentrazione delle tre forme di de/protonate in funzione del pH. Immagine tratta da [7]

la forma predominante è quella zwitterionica, Figura 2.1. Gli amminoacidi possono essere separati in vari modi, quella presentata qui è legata alla polarità e carica delle catene laterali. Possiamo quindi distinguere le catene in 4 gruppi: apolari, polari neutre, polari cationiche, polari anioniche. I gruppi sono rappresentati nelle Figure 2.2, 2.3 e 2.4.

Apolari

Catene apolari sono alanina, valina, leucina, isoleucina, glicina, metionina, prolina, fenilalanina, e triptofano. La glicina è già stata menzionata in precedenza. Altri amminoacidi notevoli sono la prolina, in quanto è l'unico amminoacido ciclico, e di conseguenza è particolarmente rigido, e la metionina, uno dei due amminoacidi con lo zolfo. Queste catene, per quanto classificate come apolari, presentano un range di polarità in cui, ad esempio, la fenilalanina è decisamente più apolare degli altri amminoacidi aromatici. L'isoleucina contiene un secondo centro stereogenico, si trova solo il diastereomero in cui l'idrogeno di questo secondo centro è trans rispetto all'idrogeno del C_{α} .

Polari neutre

Sono la serina, la treonina, la tirosina, l'asparagina, la glu-

tammina, e la cisteina. Le prime 3 terminano con un gruppo idrossilico, tirosina ed asparagina terminano con un gruppo ammidico e la cisteina con un tiolo. La cisteina gioca un ruolo importante nella strutturazione delle proteine mediante la formazione di un legame disulfuro fra due residui cisteinici. Anche la treonina, come l'isoleucina, contiene un secondo centro stereogenico, anche qui i due idrogeni sono disposti trans.

Polari cationiche

Lisina, arginina ed istidina. Terminano tutte con un gruppo azotato, rispettivamente amminico, guanidinico, ed imidazolico. Come le catene anioniche, sono molto più idrofiliche delle polari neutre. L'anello imidazolico ha un pKa di 6, e in un pH fisiologico di 7 esso si troverà protonato o meno a secondo dell'intorno chimico.

Polari anioniche

Aspartato e glutammato, entrambi contenenti un gruppo carbossilico.

La struttura primaria di una proteina è la sequenza di amminoacidi uniti da un legame peptidico fra il carbonio carbonilico e l'azoto di due amminoacidi diversi lungo il *backbone*. Questa struttura primaria presenta un ordine ben preciso: la testa della catena è il residuo con il gruppo amminico mentre la coda è il residuo con il gruppo carbossilico. La formazione del legame peptidico è energeticamente sfavorevole e comporta la perdita di una molecola di acqua. L'equilibrio è spostato verso l'idrolisi, però il legame formato è molto stabile. La lunghezza del legame carbonio azoto è 1.32Å, intermedio fra C-N (1.49Å) e C=N (1.27Å). Questo, unito al fatto che gli atomi del gruppo peptidico sono coplanari, indica che la rotazione attorno al legame C-N è impedita dalla sua natura di quasi doppio legame. Si può vedere il centro C_α come un punto in cui due piani si toccano. Il legame peptidico preferisce l'isomeria trans in quanto la disposizione cis è sfavorita da interazioni steriche fra le catene laterali. Un caso particolare è la prolina che non mostra particolare preferenza fra trans e cis, (Figura 2.6). La rotazione è consentita nel legame N- C_α e C_α -C. Definiamo due angoli diedri, ϕ per la rotazione di N- C_α e ψ per quella attorno C_α -C, ed assumiamo che valgano 180° quando tutti i gruppi peptidici sono nello stesso piano. Non tutti gli angoli sono permessi, come evidenziato dal grafico di Ramachandran, Figura 2.7. Gli angoli ammessi corrispondono a strutture secondarie caratteristiche quali le α -eliche e i fogliettini- β .

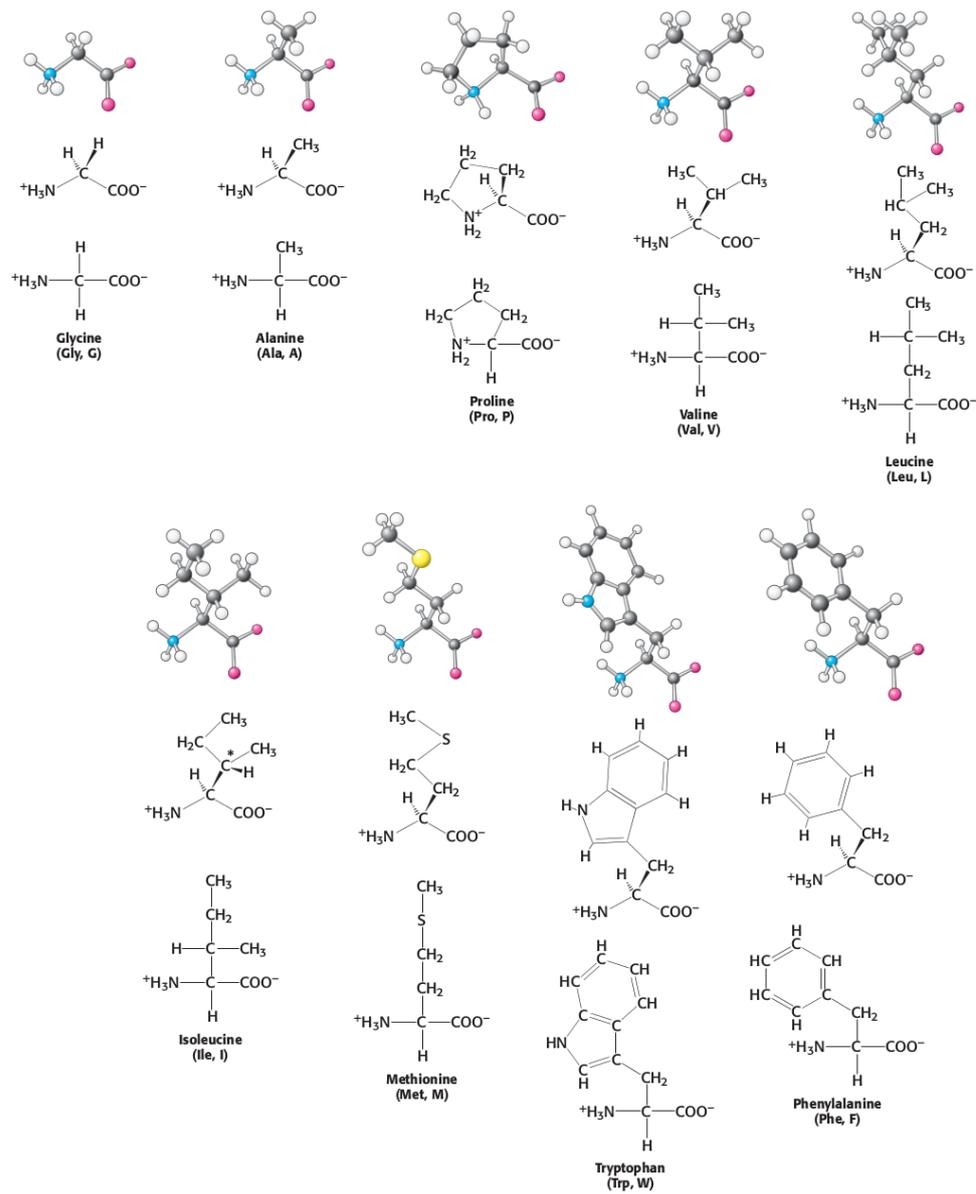


Figura 2.2: I 9 amminoacidi apolari. Immagine tratta da [7].

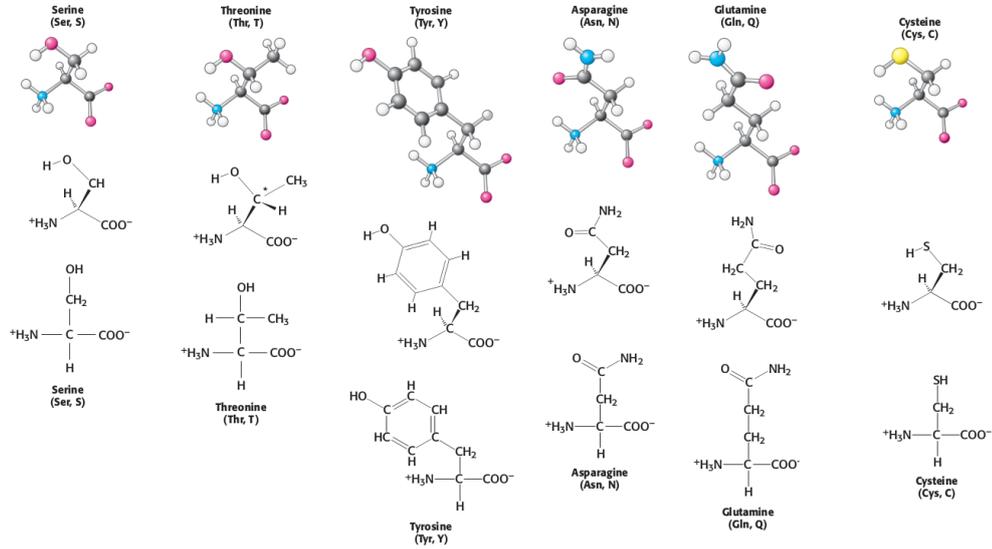


Figura 2.3: I 6 amminoacidi polari. Immagine tratta da [7].

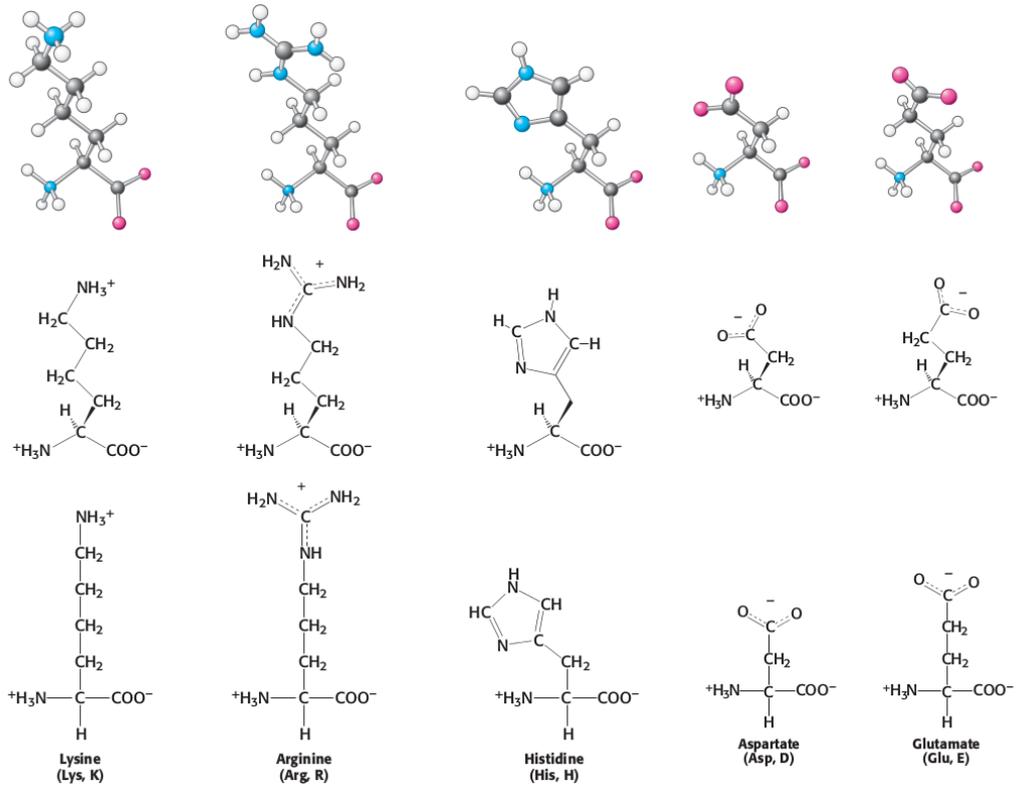


Figura 2.4: I 5 amminoacidi polari e non neutri. Immagine tratta da [7].

Organism	Sequence Alignment	Swiss-P
Amoeba	MQIFVKTLTGKTIITLVESSDTIENVKQKIQDKEGIPDQQRLLFAGKLEGGRTLADYNIQKESTLHLVLRRLGG	P49634
Green alga	MQIFVKTLTGKTIITLVESSDTVENVKSKIQQKEGIPDQQRLLFAGKLEGGRTLADYNIQKESTLHLVLRRLGG	P42739
Chlamy. reinhardtii	MQIFVKTLTGKTIITLVESSDTIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLADYNIQKESTLHLVLRRLGG	P14624
Mouse	MQIFVKTLTGKTIITLVEPSDTIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P62991
Human (*)	MQIFVKTLTGKTIITLVEPSDTIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P62988
Slime mold	MQIFVKTLTGKTIITLVEGSDNIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P08618
Purple sea urchin	MQIFVKTLTGKTIITLVEPSDSIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P23398
Eimeria bovis	MQIFVKTLTGKTIITLVEPSDTIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P46574
T. pyriformis	MQIFVKTLTGKTIITLVEASDTIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P20685
C. elegans	MQIFVKTLTGKTIITLVEASDTIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P14792
Red alga	MQIFVKTLTGKTIITLVESSDTIENVKTKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P42740
Neurospora crassa	MQIFVKTLTGKTIITLVESSDTIDNVKQKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P13117
Baker's yeast	MQIFVKTLTGKTIITLVESSDTIDNVKSKIQQKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P61864
Inky cap fungus	MQIFVKTLTGKTIITLVESSDTIDNVKAKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P19848
Garden pea (**)	MQIFVKTLTGKTIITLVESSDTIDNVKAKIQDKEGIPDQQRLLFAGKLEGGRTLADYNIQKESTLHLVLRRLGG	P03993
Euplotes eurytomus	MQIFVKTLTGKTIITLVEQSDTIIDNVKTKIQDKEGIPDQQRLLFAGKLEGGRTLADYNIQKESTLHLVLRRLGG	P23324
Potato late blight fungus	MQIFVKTLTGKTIITLVEPSDSIDNVKQKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	P22589
Leishmania major	MQIFVKTLTGKTIITLVEPSDTIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLSDYNIQKESTLHLVLRRLGG	Q05550
Sauroleish. tarentolae	MQIFVKTLTGKTIITLVEPSDTIENVKAKIQDKEGIPDQQRLLFADKLEGGRTLSDYNIQKESTLHLVLRRLGG	P49635
T. brucei brucei	MQIFVKTLTGKTIITLVEASDTIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLADYNIQKESTLHLVLRRLGG	P15174
Trypanosoma cruzi	MQIFVKTLTGKTIITLVESSDTIENVKAKIQDKEGIPDQQRLLFAGKLEGGRTLADYNIQKESTLHLVLRRLGG	P08565

Figura 2.5: La sequenza dell'ubiquitina, in blu i residui identici. Nell'ultima colonna è riportato il codice per il database NCBI. Immagine tratta da [8].

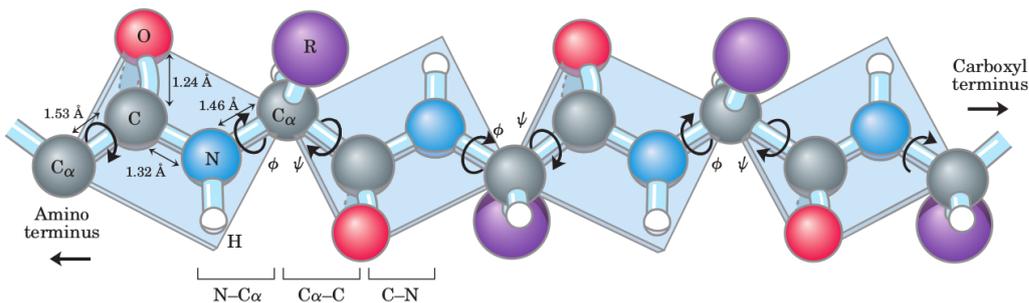


Figura 2.6: Rappresentazione planare del gruppo peptidico. Immagine tratta da [9].

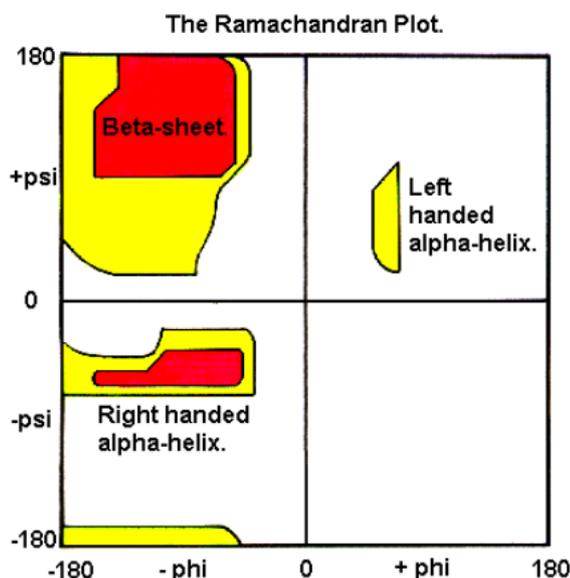


Figura 2.7: In rosso gli angoli accessibili a tutti gli amminoacidi tranne la prolina. In giallo le coppie aggiuntive accessibili alla glicina. Immagine tratta da [10].

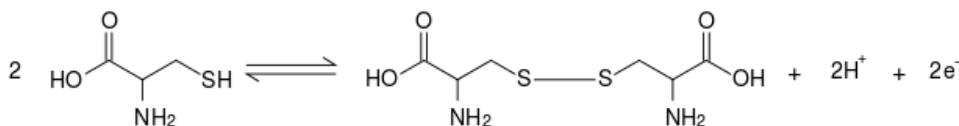


Figura 2.8: Rappresentazione schematica della formazione del ponte disolfuro.

Il legame peptidico, limitato ad amminoacidi sequenziali, non è l'unico legame covalente possibile. Due cisteine possono dare una reazione di ossidazione per formare una cistina, contenente un ponte disolfuro, Figura 2.8. Questo legame è tipico della struttura di proteine extra-cellulari e di proteine resistenti alle alte temperature.

2.2 Strutture secondarie e superiori

Il ripiegamento della struttura primaria porta alla formazione di motivi ricorrenti chiamati strutture secondarie e terziarie. Le strutture secondarie sono formate a livello locale mediante legami idrogeno. Possiamo distinguere quattro tipi di struttura secondaria: α -elica, fogliettino- β , β -turn ed Ω -ansa. L' α -elica, Figura 2.9, è una struttura a bastoncino in cui il *backbone* si trova all'interno e le catene laterali all'esterno. Nell'elica ogni residuo si

trova alzato di 1.5\AA e ruotato di 100° , questo comporta una distanza di 5.4\AA per ogni rotazione completa dell'elica, mentre gli angoli fra i residui sono $\phi = (-180^\circ, -60^\circ)$ e $\psi = -45^\circ$, come mostrato in Figura 2.7. Le eliche possono essere sia levogire che destrogire, ma per questioni di ingombro sterico la vasta maggioranza sarà destrogira. Possono essere formate da amminoacidi L o R ma non da una combinazione. Le interazioni stabilizzanti dell'elica sono i legami idrogeno fra il carbonile e l'ammina di due residui diversi e distanti 3 o 4 residui. Tutti i gruppi CO e NH sono coinvolti nella formazione di legami idrogeno, tranne i CO del C terminale e i gruppi NH del residuo N terminale, si ha una sorta di separazione di carica che induce la presenza di gruppi carichi negativamente nella testa e carichi positivamente nella coda. Questo tipo di stabilizzazione implica che certi gruppi possano portare ad una diminuzione di stabilità della catena. Amminoacidi con gruppi ramificati come valina, isoleucina e treonina tendono a destabilizzare l'elica per interazioni steriche, amminoacidi con catene laterali capaci di formare legami idrogeno (serina, asparagina, e l'acido aspartico) portano ad una diminuzione di stabilità per concorrenza nella formazione dei legami idrogeno fra *backbone* e catena laterale, e la prolina, sia per la mancanza di un gruppo NH capace di partecipare al legame sia per l'incapacità di formare l'angolo ϕ necessario per l'elica, è anch'essa destabilizzante. Un ulteriore effetto destabilizzante è la vicinanza fra gruppi carichi positivamente (o negativamente) al pH in cui si opera, ad esempio una sequenza di lisine e/o arginine a pH 7 tenderà a sfavorire la formazione dell'elica per repulsione elettronica. Oltre alla struttura α , esistono altre tipi di eliche. Nell' α -cheratina si ha la formazione di una super-elica levogira formata da due eliche destrogire che presentano 3.5 residui per rotazione completa, facendo sì che le interazioni fra le catene laterali si ripetano una volta ogni sette residui. Il collagene è composto da tre eliche in cui si ha assenza di legame idrogeno intra-elicale. La struttura è mantenuta dalle interazioni steriche fra proline, e si ha la formazione di una macrostruttura a cavo fra le tre eliche mediante legami idrogeno fra residui appartenenti a diverse eliche.

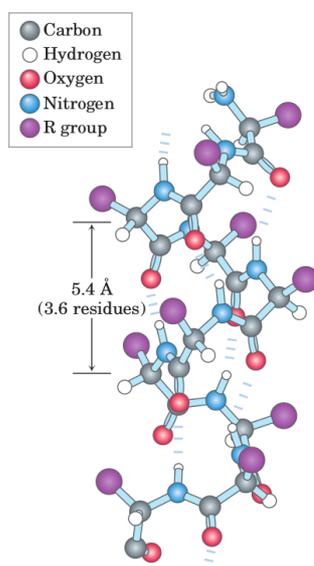


Figura 2.9: α -elica destrorsa. Immagine tratta da [9].

Un secondo tipo di struttura secondaria è il foglietto- β , Figura 2.10, che può essere presente nelle varianti parallela ed anti-parallela. In questo tipo di struttura il *backbone* è disposto a zig-zag e non in modo elicoidale. Rispetto all' α -elica la struttura è molto meno compatta con una distanza fra residui di 3.5\AA (un po' inferiore nel caso del foglietto parallelo), oltre il doppio rispetto all'elica, formando un filamento β . Le interazioni fra due o più filamenti portano alla formazione del foglietto, che viene distinto sulla base di come si formano i legami idrogeno fra i filamenti, come mostrato in Figura [?]. Nel caso in cui il legame NH-CO avvenga fra gli stessi due amminoacidi legati dall'interazione CO-NH, si parla di fogliettino antiparallelo, se invece ogni amminoacido forma legami idrogeno con due amminoacidi diversi si ha un fogliettino parallelo. Un foglietto β può anche essere mescolato e contenere sezioni parallele ed anti-parallele. In ogni caso, a prescindere dalla struttura, un filamento β avrà i gruppi laterali disposti in posizione sfalsata mentre nel caso del foglietto essi saranno disposti in posizione eclissata, comportando una limitazione a quali amminoacidi possono formare il foglietto. Anche se i foglietti possono essere planari, la maggioranza tende a ruotare, a volte fino a 90° , nel corso della struttura. Esiste anche una variante α , Figura 2.11, più rara in quanto energeticamente meno stabile. In questo tipo di struttura i gruppi amminici e carbonilici sono orientati in due direzioni diverse, il gruppo carbonilico forma legami idrogeno con un filamento diverso rispetto al gruppo amminico, questa causa una separazione di carica sulla superficie del foglietto, rendendo la struttura instabile.

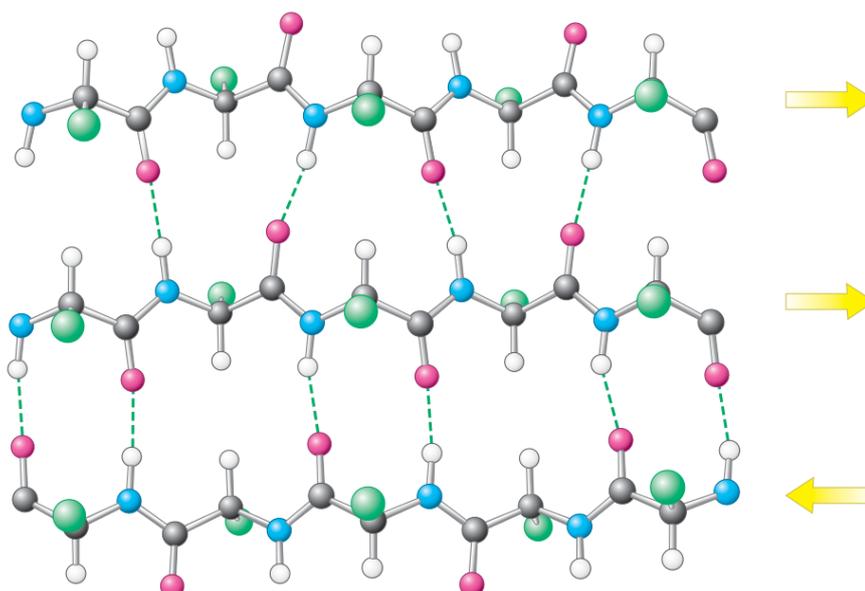


Figura 2.10: Un fogliettino- β che presenta una parte parallela in alto ed una anti-parallela in basso. Immagine tratta da [7].

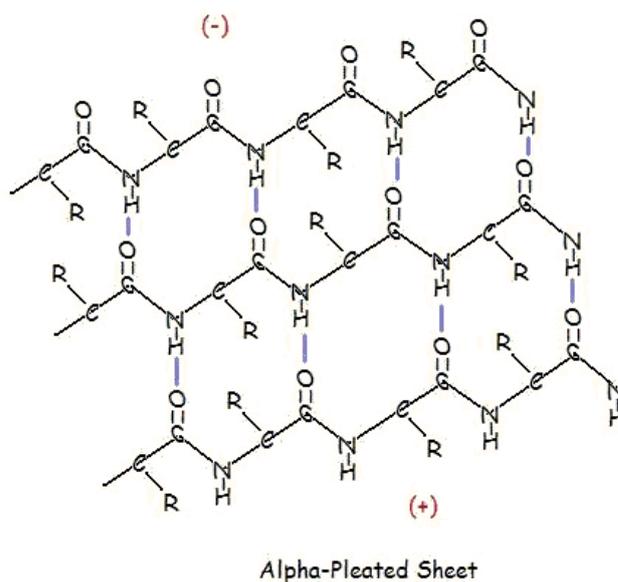


Figura 2.11: Fogliettino- α . Immagine tratta da [11].

I β -turn sono abbondanti nelle proteine globulari. Sono strutture secondarie composte da quattro residui che compiono un curva di 180° , con un legame idrogeno fra l'ossigeno del primo residuo e l'azoto del quarto, mentre i due residui centrali non partecipano. Particolarmente frequenti in queste

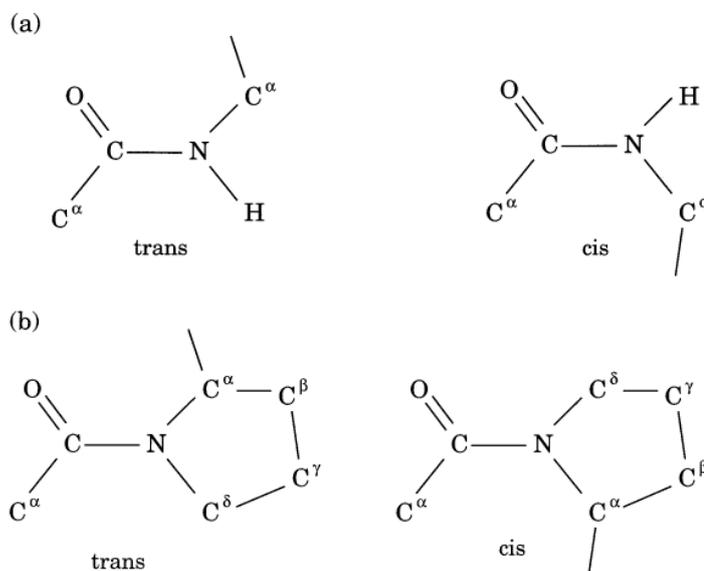


Figura 2.12: Amminoacido generico e la prolina nei duei isomeri cis e trans. Immagine tratta da [12].

strutture sono le glicine, per la loro flessibilità, e le proline, data la relativa tendenza ad esistere come isomero cis, dato che i due stereoisomeri sono molto simili, Figura 2.12. Esistono vari tipi di β -turn, classificati in base agli angoli diedri fra i residui, Tabella 2.1. I più importanti sono il tipo I e II, Figura 2.13, ma sono presenti anche i tipi I' e II', rispettivamente speculari a I e II. I tipi I e II si differenziano per la disposizione spaziale dell'ossigeno del secondo residuo e dell'azoto del terzo residuo, inoltre il tipo II presenterà sempre un residuo glicinico.

	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
Tipo I	-60°	-30°	-90°	0°
Tipo I'	60°	30°	90°	0°
Tipo II	-60°	120°	80°	0°
Tipo II'	60°	-120°	-80°	0°

Tabella 2.1: Angoli diedri ideali fra i piani in quattro tipi di β -turn [13].

Un ulteriore tipo di turn, molto meno comune, è il γ -turn, costituito da 3 residui con un legame idrogeno fra il residuo n ed il residuo $n+1$, con angoli medi $\phi_{i+1} = 75^\circ$ e $\psi_{i+1} = -64^\circ$ [14].

Gli Ω -loop presentano una struttura non regolare ma comunque rigida e definita, Figura 2.14. Si tratta di parti di sequenza che compiono una curva

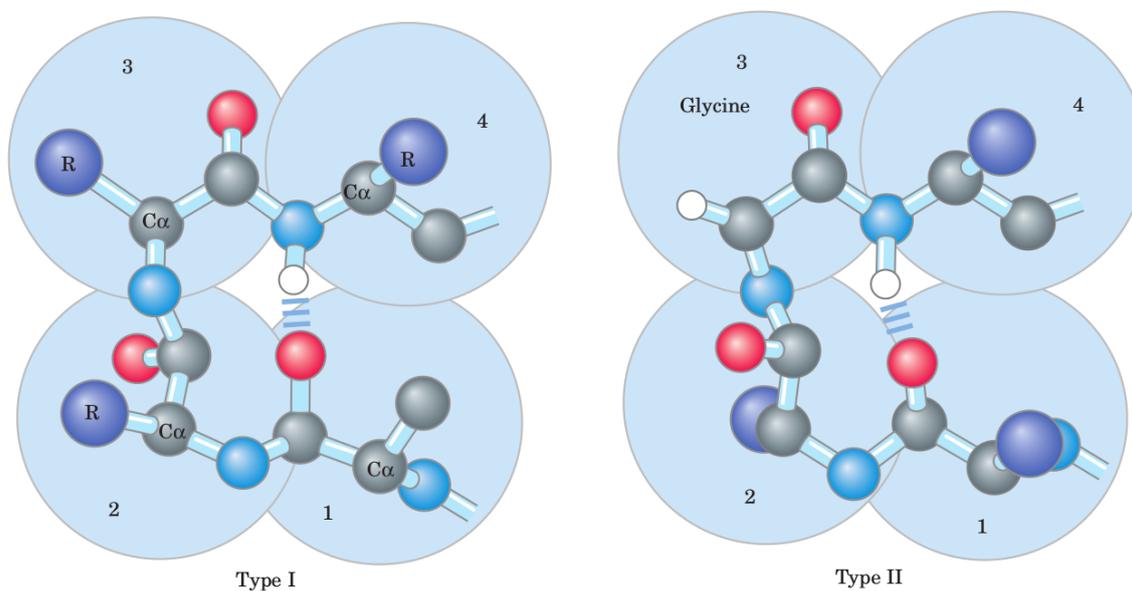


Figura 2.13: β -turn di tipo I e II. Immagine tratta da [9].

di 180° , ricchi in legami idrogeno e più lunghi dei turn. Solitamente queste strutture si trovano sulla superficie della proteina.

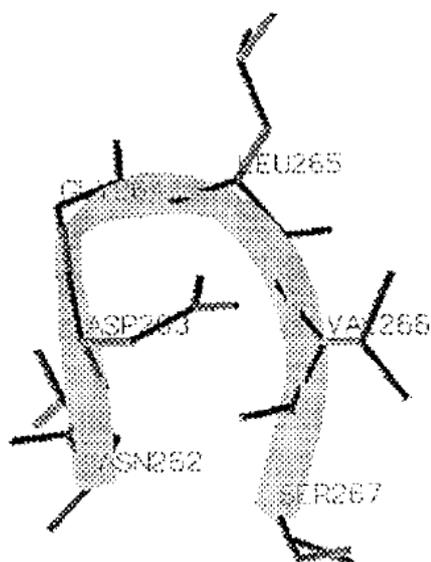


Figura 2.14: Un esempio di Ω -loop, 1cvl:262-267. Immagine elaborata a partire da [15].

Alcune combinazioni strutture secondarie sono presenti in molte proteine e spesso presentano funzioni simili. Queste combinazioni, che vengono

chiamate domini, possono portare alla formazione di strutture complesse a partire da elementi semplici come le eliche e i foglietti. Un esempio può essere il motivo elica-turn-elica composto da due α -eliche unite da un turn e con funzioni di riconoscimento e legame. La presenza di questi motivi viene usata per separare le proteine in famiglie. A livello complessivo la proteina è però disposta in una struttura terziaria, che risulta unica per ogni proteina e, quindi, per ogni sequenza di amminoacidi. Queste strutture terziarie sono formate da interazioni di van der Waals (vdW), legami idrogeno, e, nel caso di proteine rigide, da ponti disolfuro. Sulla base della loro struttura terziaria le proteine possono essere separate in due famiglie, le fibrose e le globulari. Le proteine fibrose, formate da catene disposte come fogli, presentano delle superfici idrofobiche che le rendono insolubili. Queste superfici vengono nascoste in seguito all'impacchettamento delle proteine. Una proteina globulare, invece, presenta un ripiegamento più compatto. Rispetto alle proteine lineari, che essenzialmente avevano un ruolo strutturale, le proteine globulari possono svolgere più funzioni, da enzimatica ad una di trasporto. Nelle proteine globulari si assiste alla formazione di una parte interna contenente principalmente residui idrofobici, mentre l'esterno della proteina avrà anche residui idrofilici, facendo diminuire le interazioni sfavorevoli fra un solvente polare e le catene idrofobiche. Esistono anche proteine, come le porine, che operano in un ambiente polare e hanno una struttura rovesciata, con i gruppi idrofilici all'interno. Proteine globulari di piccole dimensioni tenderanno ad essere stabilizzate da legami covalenti, in quanto avranno sia un minor numero di interazioni interne stabilizzanti sia una minor capacità di isolare i residui idrofobici rispetto a proteine di dimensioni maggiori. Catene amminoacidiche molto lunghe possono ripiegarsi in un numero di regioni compatte globulari, dette domini, composte da 300-400 residui ed unite da una parte flessibile della catena. Le proteine che sono composte da diverse catene di polipeptidi posseggono un quarto livello strutturale, denominato struttura quaternaria e composto dalle disposizione spaziale relativa delle catene. Un esempio è il già citato collagene, [Figura 2.16](#). Le sub-unità che compongono una proteina possono essere non identiche, portando a strutture complesse. Sub-unità identiche appartenenti ad una stessa proteina tenderanno ad essere disposte con simmetria rotazionale o elicoidale. L'uso di questo tipo di struttura, rispetto ad una singola sequenza, porta a dei vantaggi sia di risparmio di spazio richiesto dal gene incaricato di sintetizzare la proteina, sia ad un più facile controllo degli errori che avvengono durante la sintesi della proteina.

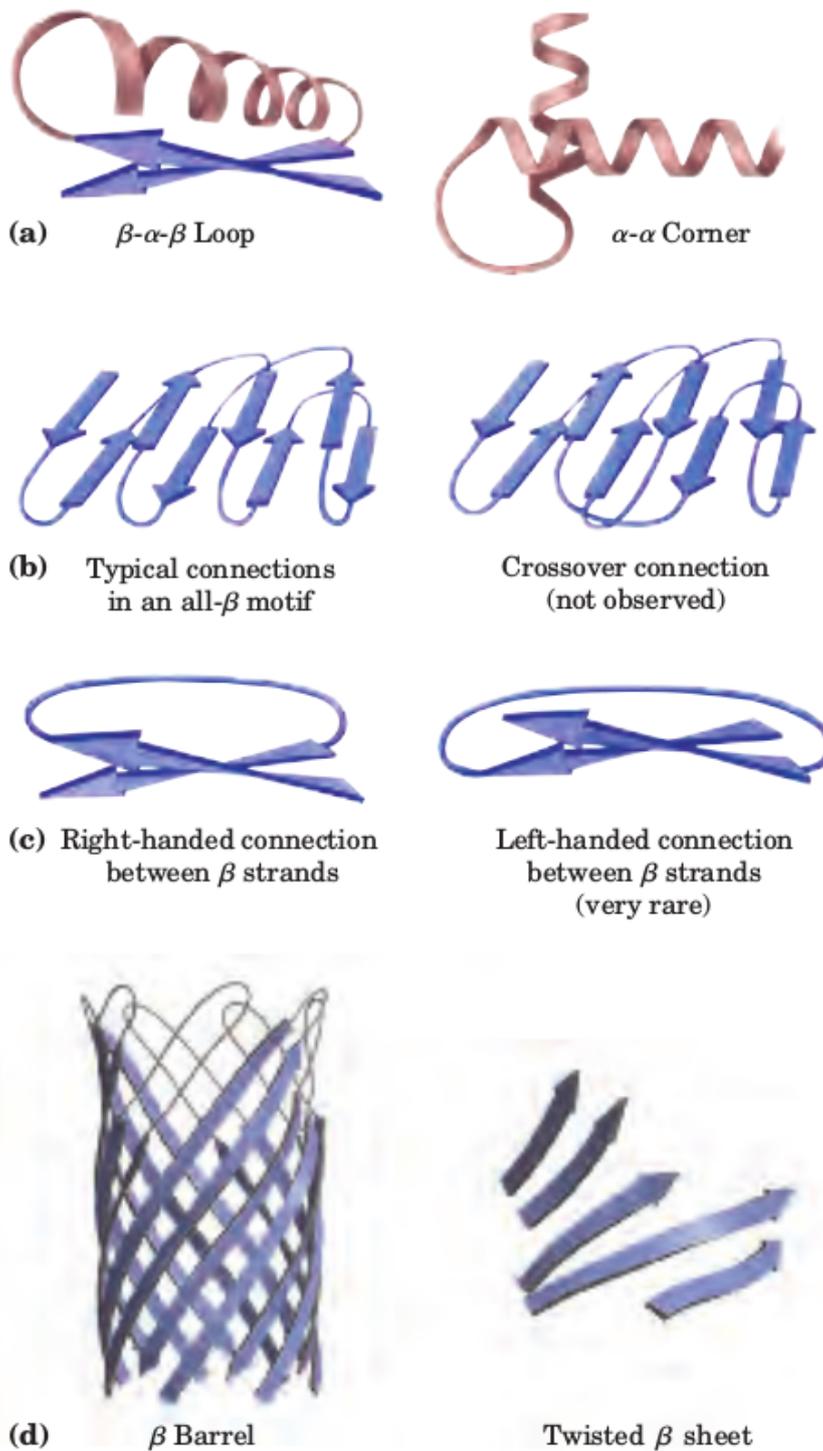


Figura 2.15: Alcuni motivi strutturali delle proteine. Immagine tratta da [9].

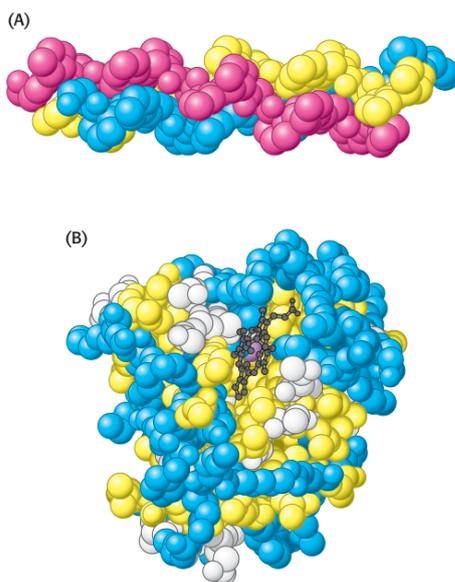


Figura 2.16: a) Collagene, una proteina fibrosa. L'unione delle tre eliche forma una struttura quaternaria. b) Mioglobina, una proteina globulare. In giallo i residui idrofobici, in blu i residui carichi, e i restanti in bianco. Immagine elaborata a partire da [7].

2.3 Ripiegamento

La prima evidenza che tutte le informazioni sulla struttura 3d della proteina fossero contenute nella sequenza fu l'esperimento di Anfinsen [16]. Anche partendo da un ribonucleasi in una struttura stabile ma diversa da quella nativa ed inattiva, aggiungendo del 2-mercaptoetanololo, era possibile tornare alla struttura attiva mediante una rottura ed una formazione di legami disolfuro dovuta dalla diminuzione di energia libera nel passaggio dalla struttura errata a quella ripiegata in maniera corretta. Il meccanismo esatto del ripiegamento è un fenomeno complesso e non totalmente compreso. Dalle curve di denaturazione, Figura 2.17, sia in funzione della temperatura che della concentrazione un agente denaturante, si vede come la perdita di struttura sia una transizione molto *sharp*, simile ad un fenomeno critico del tipo transizione ferromagnetica, suggerendo che il ripiegamento sia un fenomeno cooperativo e che la perdita di struttura in una parte della proteina induca ad un collasso globale. Anche prendendo una soluzione in condizioni tali da essere esattamente nel punto della transizione, quello che si vede è semplicemente una miscela di proteine completamente ripiegate e proteine completamente inattive, e non si trovano mai proteine parzialmente ripiegate. Un'altra evidenza della natura cooperativa del ripiegamento si ha dall'analisi della tendenza dei vari amminoacidi a trovarsi in specifiche strutture secondarie.

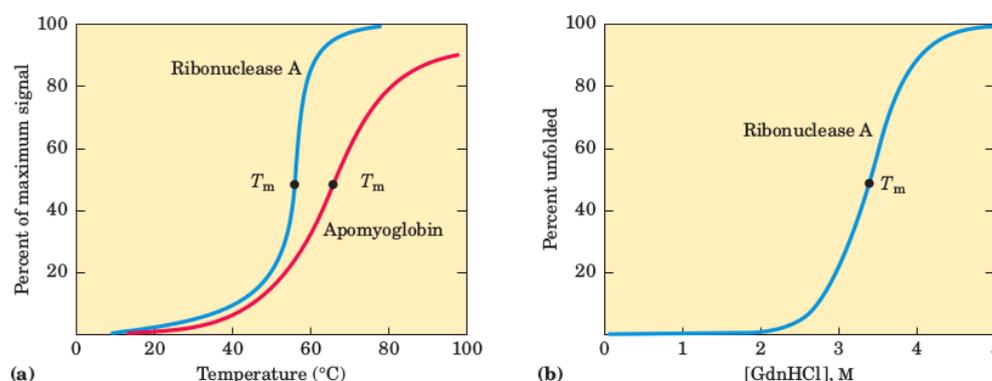


Figura 2.17: a) Denaturazione termica per la apomioglobina del cavallo e ribonucleasi A. b) Denaturazione della ribonucleasi ad opera di cloruro di guanidinio. Immagine tratta da [9].

Nonostante amminoacidi diversi abbiano tendenza diverse a presentarsi in strutture secondarie diverse, queste non sono mai delle preferenze estremamente forti ed esclusive. Inoltre, è possibile trovare una stessa breve sequenza di amminoacidi in strutture diverse in proteine diverse. Per il ripiegamento sono quindi importanti le interazioni fra residui avvicinati da una prima strutturazione della proteina che abbandona lo stato di *random coil*. Sono state proposte due teorie per spiegare come avviene il ripiegamento di una proteina. Un primo modello, di tipo gerarchico, prevede la formazione di strutture secondarie a livello locale che avvicinano nello spazio amminoacidi lontani, le interazioni a lunga distanza formano una struttura stabile, con un processo che prosegue fino al raggiungimento della struttura nativa ed energeticamente più stabile. Un secondo modello prevede la formazione di uno stato intermedio, detto globulo fuso, dovuto ad un compattamento spontaneo guidato dalle interazioni idrofobiche. Il globulo fuso presenta un alto contenuto di strutture secondarie, simili a quelle della struttura nativa, ma le catene non sono ancora fissate e il globulo è estremamente flessibile e dinamico. In ogni caso il ripiegamento non segue un percorso preciso ma avviene attraverso una serie di percorsi diversi, con strutture diverse, generando una sorta di imbuto energetico in cui si ha convergenza verso una singola struttura, come mostrato in Figura 2.18.

2.3.1 Driving force

Il processo di folding è un processo spontaneo, quindi, per definizione, porterà ad una diminuzione di energia libera di Gibbs del sistema:

$$\Delta G = \Delta H - T\Delta S \quad (2.1)$$

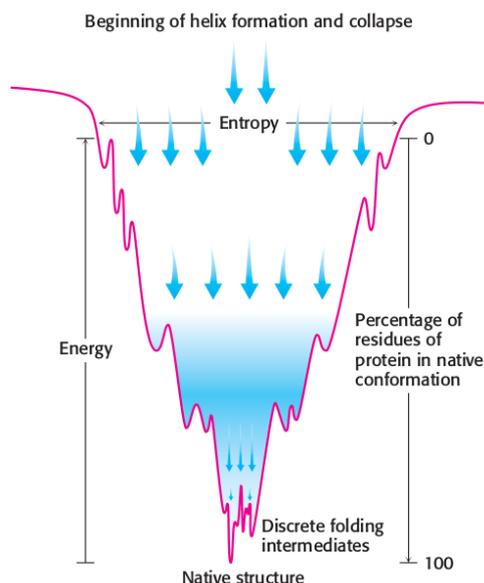


Figura 2.18: Una rappresentazione ad imbuto del ripiegamento della proteina. Ad alte energie sono accessibili un numero maggiore di strutture, che diminuiscono via via che la proteina inizia a ripiegarsi. Lungo questo ripiegamento esistono degli intermedi metastabili con minimi locali di energia. Immagine tratta da [7].

La discriminazione fra i contributi entalpici ed entropici costituisce un problema complesso ed è uno dei punti che questa tesi si propone di affrontare. Nel ripiegarsi la proteina perde entropia ($\Delta S_{prot} < 0$) mentre il solvente la guadagna ($\Delta S_{solv} > 0$). In parallelo, la proteina rompe dei legami idrogeno con l'acqua e ne forma altri interni fra i singoli amminoacidi. Analogamente, il solvente (acqua) rompe alcuni legami idrogeno con la proteina ma ne forma altri con sé stesso per riformare la caratteristica struttura tetraedrica che è piuttosto stabile. In questo caso, il bilancio relativo delle componenti è molto meno chiaro, e il segno non può essere stabilito a priori. L'intensità di un legame idrogeno varia a seconda dell'ambiente e degli atomi coinvolti, si tratta di un legame energeticamente intermedio fra un legame covalente ed un legame van der Waals. Esclusi casi eccezionali, come lo ione idrogenodifluoruro, si tratta di decine di kJ mol^{-1} . Un legame idrogeno è lungo circa 2.5-3 Å. Nel contesto proteico in acqua, gli atomi coinvolti in un legame idrogeno saranno l'azoto amminoacidico (donatore), l'ossigeno amminoacidico (accettore), e l'ossigeno dell'acqua. Il legame idrogeno è un legame direzionale, idealmente l'angolo fra i 3 atomi coinvolti dovrebbe essere di 180° . Questa rigidità è alla base della formazione delle strutture secondarie. Il legame idrogeno tende anche ad avvicinare due strutture lontane nella sequenza amminoacidica, formando una struttura secondaria compatta. Un secondo contributo

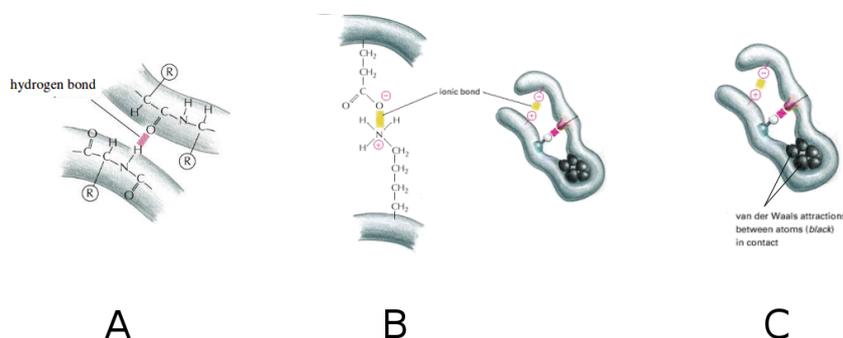


Figura 2.19: A: legame idrogeno interno, B: legame ionico, C: contributo di vdW. immagine elaborata a partire da [17].

energetico rilevante è un termine ionico dovuto all'attrazione fra cariche. Nel pH fisiologico di 7 si avranno catene laterali completamente protonate (equivalenti ad una carica positiva netta) e catene laterali deprotonate (ossia una carica negativa). Solitamente i residui coinvolti sono acido aspartico o glutammico come anione e lisina o arginina come catione. In una struttura ripiegata le catene saranno sia più vicine sia in un ambiente più apolare di quello in cui si trovano nella proteina aperta. Si avrà un vantaggio energetico nel ripiegamento dato che l'energia nella **forza di Coulomb** è inversamente proporzionale alla distanza e alla permittività elettrica:

$$F = \frac{1}{4\pi\epsilon} \frac{q_1 q_2}{r^2} \quad (2.2)$$

Il terzo termine è dovuto a interazioni a breve distanza fra dipoli indotti, che vengono identificati come termini di van der Waals. In una struttura aperta le interazioni di questo tipo saranno trascurabili, dato che le catene sono troppo lontane per poter distorcere la nube elettronica delle altre catene, mentre in una struttura ripiegata e compatta esse saranno un ulteriore contributo alla stabilità della struttura chiusa. Il vero motore alla base del ripiegamento, però, è l'effetto idrofobico. Si tratta della tendenza alla separazione fra una fase polare ed una apolare. In una proteina, questa separazione conduce alla formazione di una struttura compatta in cui le catene apolari sono nascoste all'acqua. Questa driving force è caratteristica dell'acqua, in quanto le interazioni dipolo-dipolo sono molto più forti delle interazioni dipolo-dipolo indotto che l'acqua farebbe con una struttura apolare. La perdita energetica dovuta al fatto che, dove c'è la catena apolare, l'acqua non può formare legami idrogeno, viene compensata dalla formazione di legami

idrogeno acqua-acqua molto corti e lineari. Questi legami idrogeno comportano però un riarrangiamento locale dell'acqua intorno alla proteina in un ordine simile a quello del ghiaccio, una situazione sfavorevole dal punto di vista entropico. Il ripiegamento delle catene apolari consente un rilassamento della struttura cristallina con aumento dell'entropia e quindi calo dell'energia libera del sistema. Nell'acqua bulk ogni molecola d'acqua si troverebbe al centro di un tetraedro, con l'atomo di ossigeno interno distante 0.28 nm dagli ossigeni angolari. L'acqua forma solo due dei 4 possibili legami idrogeno, dando origine ad una disposizione fra le sei energeticamente equivalenti possibili. Una struttura idealmente cristallina avrebbe invece quattro legami idrogeno formati su un massimo di quattro. Dalla definizione di entropia come funzione del logaritmo dei microstati possibili è facile comprendere come ci sia un guadagno entropico nel passaggio dalla struttura ordinata quasi-cristallina a quella disordinata del liquido, anche se il grado di ordine nella struttura reale è inferiore a quello del ghiaccio [18]. Va inoltre notato come la strutturazione cristallina aumenta il volume parziale dell'acqua mentre l'aggiunta di un solvente apolare lo fa calare. Un'altra conseguenza dell'effetto idrofobico e di questa cristallizzazione dell'acqua è l'aumento di capacità termica in seguito all'unfolding della proteina [19]. Dalle osservazioni precedenti, ne seguirebbe che le catene polari dovrebbero stare all'esterno, per massimizzare il numero di legami idrogeno con l'acqua. In realtà è possibile trovarle anche all'interno, dove contribuiscono alla definizione della struttura tridimensionale della proteina mediante interazioni reciproche.

2.3.2 Problemi relativi al folding

Il ripiegamento presenta notevoli difficoltà dal punto di vista dello studio mediante simulazione. Un primo problema è il numero di possibili conformazioni che devono essere esplorate. Per una proteina composta da 100 amminoacidi sono possibili 3^{198} strutture, derivate dai 99 legami peptidici che formano 198 angoli ψ e ϕ , ognuno dei quali si può trovare in tre configurazioni stabili. Assumendo che servano $10^{-13}s$ per passare da una struttura all'altra, servirebbero circa 10^{77} anni per il ripiegamento, quindi molto maggiore dell'età dell'universo che risulta nell'ordine dei 10^{10} anni. Queste considerazioni, dette paradosso di Levinthal in onore di chi le ha suggerite per primo [20], suggeriscono che la proteina non esplori tutte le possibili strutture, ma tenda a mantenere le parti favorevoli, ovvero a minor energia, delle strutture esplorate. Questa soluzione non è applicabile da un punto di vista computazionale in quanto la differenza energetica fra lo stato nativo e denaturato di una proteina di 100 residui è nell'ordine di 4 kJ mol^{-1} per residuo [21]. Questo valore è confrontabile con l'energia termica a temperatura ambiente, 2.4

kJ mol^{-1} , il che significa che la proteina può essere destabilizzata nelle sue fasi di ripiegamento intermedie. Un secondo problema deriva dal fatto che, come visto, la conoscenza della sequenza locale è insufficiente a questo livello per determinare la struttura secondaria e sono richieste le interazioni a lunga distanza. Questa situazione limita l'utilizzo di metodi basati sul confronto fra sequenze simili in proteine diverse. Un problema specifico delle simulazioni è la necessità di non poter fare una integrazione troppo grossolana per non perdere informazioni sulla dinamica dell'acqua, rendendo simulazioni nell'ordine dei secondi troppo complesse. Un ulteriore problema deriva dal fatto che certe proteine non avranno necessariamente, in toto o in una parte, una singola struttura. Una prima classe, definita IUP (*intrinsically unstructured proteins*), è costituita da proteine che in condizioni fisiologiche non hanno una struttura specifica che viene invece assunta quando interagiscono con una seconda proteina. Queste proteine sono ricche in gruppi polari e carichi. Una seconda classe, le proteine metamorfiche, esistono in un equilibrio di strutture ad energia simile. Alcune proteine non sono in grado di ripiegarsi da sole e richiedono l'intervento di proteine specializzate dette Chaperoni, che facilitano i corretti percorsi di ripiegamento.

2.4 Metodi sperimentali per la caratterizzazione della proteina

In questo capitolo passeremo in rassegna i metodi ai fini della validazione degli strumenti numerici usati in questa tesi.

2.4.1 Cristallografia a raggi X

Questa tecnica si basa sull'analisi della diffrazione da parte di un cristallo. I raggi X, fotoni con una lunghezza d'onda λ compresa fra 0.1 e 100 Å, sono formati bombardando con elettroni ad alta energia un bersaglio. Introduciamo il vettore di scattering \mathbf{K} , il cui modulo risulta legato all'angolo θ di scattering dalla nota relazione

$$|\mathbf{K}| = \frac{1}{\pi} \frac{\sin \theta}{\lambda} \quad (2.3)$$

e risulta quindi compreso fra 0 e $\frac{1}{\pi\lambda}$. Il fattore di struttura risulta legato alla densità di carica $\rho(\mathbf{r})$ dalla trasformata di Fourier [22]

$$F(\mathbf{K}) = \int \rho(\mathbf{r}) e^{i\mathbf{K}\cdot\mathbf{r}} d^3r \quad (2.4)$$

Possiamo ottenere la densità a carica $\rho(\mathbf{r})$ a partire da F come la trasformata inversa di Fourier

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{K}} F(\mathbf{K}) e^{i\mathbf{K}\cdot\mathbf{r}} \quad (2.5)$$

Il fattore di struttura è un numero complesso, composto da un termine reale ed un termine immaginario F_i

$$F = |F|^2 e^{i\phi} \quad (2.6)$$

$$F = F_R + iF_i \quad (2.7)$$

Quello che viene misurato in maniera sperimentale è l'intensità, ossia il prodotto del fattore di struttura per la sua complessa coniugata

$$I(\mathbf{K}) = F(\mathbf{K})F^*(\mathbf{K}) = |F|^2 \quad (2.8)$$

La cui trasformata di Fourier si chiama funzione di Patterson. Nelle applicazioni pratiche, la $\rho(\mathbf{r})$ è in realtà formata da cariche localizzate nelle determinate posizioni degli atomi che compongono la proteina. Del fattore di struttura è non misurabile direttamente il termine di fase $e^{i\phi}$. Nel caso dello studio di macromolecole di grandi dimensioni, come le proteine, l'analisi è complessa. Dopo aver preparato dei cristalli con la proteina, vengono preparati dei cristalli isomorfi contenenti atomi pesanti. Solitamente questi cristalli modificati vengono preparati a partire dal cristallo della proteina in cui viene fatto diffondere un reagente contenente il metallo pesante. Un esempio di questi reagenti è può essere AgNO_3 , capace di legarsi a siti SH della proteina. La posizione degli atomi viene individuata con una funzione di Patterson ottenuta per differenza. Il fattore di struttura dell'isomorfo contenente gli atomi pesanti, F_{PH} , è la somma del fattore di struttura della proteina, F_P e degli atomi pesanti, F_H :

$$F_{PH}(\mathbf{K}) = F_P(\mathbf{K}) + F_H(\mathbf{K}) \quad (2.9)$$

Possiamo approssimare l'ampiezza $|F_H(\mathbf{K})|$ come la differenza delle ampiezze del isomorfo e della proteina. Possiamo quindi calcolare una funzione di Patterson come

$$\Delta P = \frac{1}{V} \sum_{\mathbf{K}} ||F_{PH}| - |F_P||^2 e^{i\mathbf{K}\mathbf{R}} \quad (2.10)$$

dove \mathbf{R} è la posizione degli atomi. Una volta nota la posizione e la natura degli atomi pesanti, F_H è interamente noto. I valori di F_P sono un cerchio

di raggio $|F_P|$ centrato nell'origine. F_{PH} è un cerchio di raggio F_{PH} col centro spostato di F_H rispetto all'origine. Questi due cerchi si intersecano in due punti che soddisfano l'equazione 2.26. Per scegliere quale fra i due punti sia quello esatto, ossia per scegliere fra due fasi ϕ_a e ϕ_b , si ricorre a più isomorfi, contenenti atomi pesanti diversi. Possiamo quindi ottenere una F_P che useremo per calcolare la densità elettronica dalla ampiezza misurata e la fase calcolata. La posizione degli atomi pesanti viene raffinata calcolando la densità elettronica degli atomi pesanti come la differenza fra le densità elettroniche della proteina e dell'isomorfo calcolate usando la fase della proteina sia per la proteina che per l'isomorfo. Una volta ottenuta una F_P accettabile, si procede con l'assegnazione di una struttura alla mappa per determinare la struttura. La capacità di attribuire una struttura accurata dipende dalla risoluzione della mappa, che dipende a sua volta da quando gli isomorfi sono simili alla proteina di partenza.

Risoluzione	Dettagli visibili
6Å	Le α -eliche appaiono come bastoncini, si possono individuare sub-unità.
3Å	Si può tracciare il percorso dello scheletro. Catene laterali voluminose diventano visibili.
2.5Å	Quasi tutte le catene laterali sono visibili, e il gruppo carbonile di ogni peptide consente di stabilire l'orientazione del piano.
1.5Å	Sono individuabili gli atomi, e si possono identificare direttamente molte catene laterali

Tabella 2.2: Dettagli della proteina visibili in base alla risoluzione.

Di fatto a partire da una risoluzione di 2.5Å diventa possibile assegnare una struttura alla proteina, ma più la risoluzione è bassa più il processo è accurato.

2.4.2 Spettroscopia NMR per le proteine

Lo studio delle proteine via NMR passa attraverso due tecniche. La prima è la COSY, o spettroscopia di correlazione, che permette di risalire alla distanza fra gli atomi di idrogeno mediante la costante di coupling J . È limitata a protoni connessi da 1 o 2 legami. La seconda è la NOESY, basata sull'effetto nucleare Overhauser, e da informazioni sugli atomi di idrogeno vicini nello spazio. In entrambi i casi si tratta di nmr bidimensionale, un tipo di

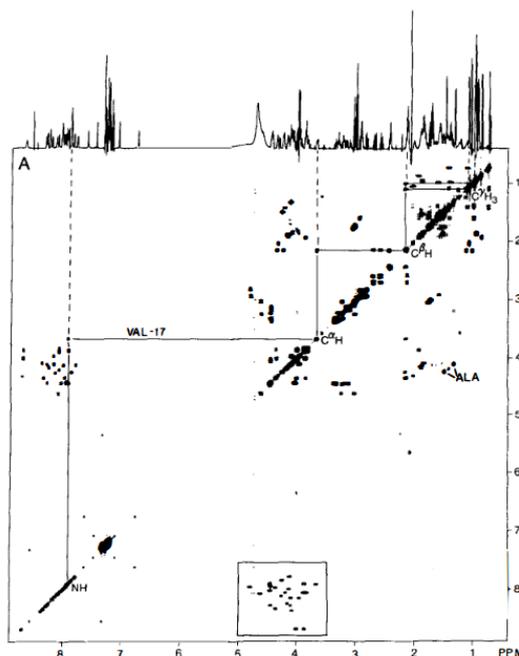


Figura 2.20: Spettro COSY ottenuto a 600MHz per la maiganina. Nel box, la fingerprint region. Immagine tratta da [25].

NMR in cui vengono studiate due frequenze e l'intensità in loro funzione. In entrambe le tecniche menzionate i dati ottenuti sono presentati come grafici bi-dimensionali con il chemical shift di due atomi di idrogeno diversi lungo i due assi. Nella tecnica COSY lo studio è limitato al singolo residuo, in quanto la presenza di un legame peptidico impedisce il coupling fra due protoni appartenenti a residui diversi. Uno spettro COSY presenta una linea diagonale, composta dai punti per cui $\delta_1 = \delta_2$, ed una serie di punti, detti cross-peak, quando i due chemical shift δ_1 e δ_2 sono diversi. Questi punti sono simmetrici rispetto alla diagonale. Una regione particolare, detta fingerprint region, mostra le interazioni fra NH e C_α H del residuo e permette di calcolare l'angolo diedro ϕ secondo l'equazione di Karplus [23]

$$J = A + B \cos \phi + C \cos(2\phi) \quad (2.11)$$

dove A, B, e C sono parametri empirici che dipendono dall'atomo. Un esempio di COSY è la Figura 2.20. Dopo aver identificato il residuo, attraverso l'accoppiamento di NH, C_α H e C_β H, esso viene inserito nella sequenza attraverso la NOESY. L'effetto Overhauser può essere spiegato a partire da un sistema bi-nucleare $\alpha_A\alpha_X$, mostrato in Figura 2.21. Se noi saturiamo la transizione X, applicando un impulso di lunga durata, la popolazione dei livelli di X diventerà uguale e non si vedranno più le transizioni fra α_X e β_X .

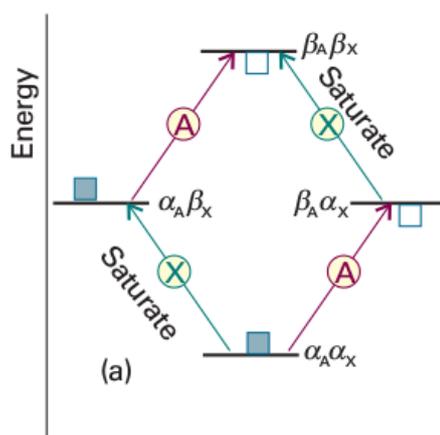


Figura 2.21: Popolazione dei livelli energetici dopo aver saturato la transizione X. Immagine tratta da [24].

La transizione A non viene modificata fino a quando non viene considerato l'effetto dei rilassamenti. Sono possibili due rilassamenti che modificano la popolazione dei livelli. Una prima transizione è dovuta all'azione del campo magnetico e comporta la contemporanea transizione $\alpha \leftrightarrow \beta$, portando ad un aumento della popolazione dello stato $\alpha_A \alpha_X$, una diminuzione della popolazione dello stato $\beta_A \beta_X$, e la non modifica della popolazione degli stati $\alpha_A \beta_X$ e $\beta_A \alpha_X$. Questo comporta un rafforzamento della transizione A, ovvero $\alpha_A \alpha_X \rightarrow \beta_A \alpha_X$ e $\alpha_A \beta_X \rightarrow \beta_A \beta_X$. Un secondo effetto è invece il rilassamento causato dall'interazione dipolare fra gli spin, portando ad una transizione $\alpha_A \beta_X \leftrightarrow \beta_A \alpha_X$ che equilibra le popolazioni dei due stati, ne consegue un indebolimento della transizione A. Quale dei due fenomeni prevale dipende dalla velocità di rotazione della molecola. Minore è la rotazione e più prevale la diminuzione, più vicina la frequenza di rotazione alla frequenza di transizione e più prevale l'effetto rafforzante l'intensità. Definiamo il valore di rafforzamento come η

$$\eta = \frac{I_A - I_A^\circ}{I_A} \quad (2.12)$$

dove I_A° e I_A sono le intensità prima e dopo aver saturato la transizione X. Quando i due nuclei sono uguali η è compreso fra -1 e $\frac{1}{2}$, mentre per due nuclei diversi il massimo di incremento è

$$\eta_{max} = \frac{\gamma_A}{2\gamma_X} \quad (2.13)$$

Un esempio di spettro NOESY è riportato in Figura 2.22. L'effetto Overhauser può essere usato per determinare le distanze fra protoni in quanto dipende

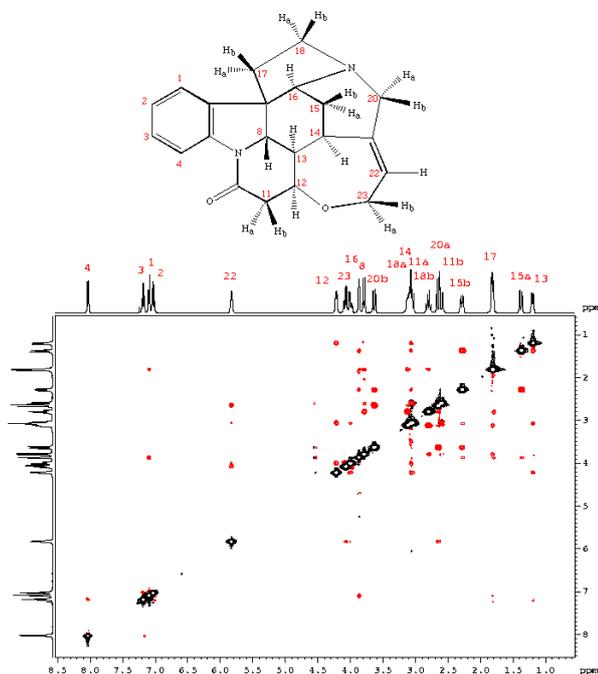


Figura 2.22: Spettro NOESY per la stricnina. In alto è riportata la struttura della molecola con la legenda dei protoni. Immagine tratta da [26].

dalla frazione del rilassamento di reticolo di A che è dovuto alle interazioni dipolari con X. Il campo dipolare ha un andamento come r^{-3} , dove r è la distanza fra protoni, e il rilassamento è proporzionale al quadrato del campo. La distanza fra due protoni i e j è legata alla matrice di rilassamento R_{ij} come $r_{ij} \propto R_{ij}^{-\frac{1}{6}}$. Questa equazione può essere sostituita, secondo la *Isolated Spin Pair Approximation*, con

$$r_{ij} = \alpha V_{ij}^{-\frac{1}{6}} \quad (2.14)$$

dove V_{ij} è il volume di cross-peak e α un fattore di calibrazione determinato da una distanza nota. Gli spettri NMR possono essere eseguiti con nuclei diversi dal protone. Il carbonio ^{13}C , nonostante sia meno sensibile a parità di numero di nuclei a causa di un basso rapporto giromagnetico e della scarsa abbondanza naturale, trova impiego. I principali vantaggi sono l'allargamento della finestra del chemical shift rispetto agli spettri protonici, in cui i chemical shift cadevano all'interno di 8 ppm, e l'assenza dello splitting ^{13}C - ^{13}C . Oltre agli svantaggi già elencati, gli spettri ^{13}C delle proteine denaturate e native sono meno diversi rispetto agli spettri protonici, rendendo più difficile lo studio di cambiamenti strutturali.

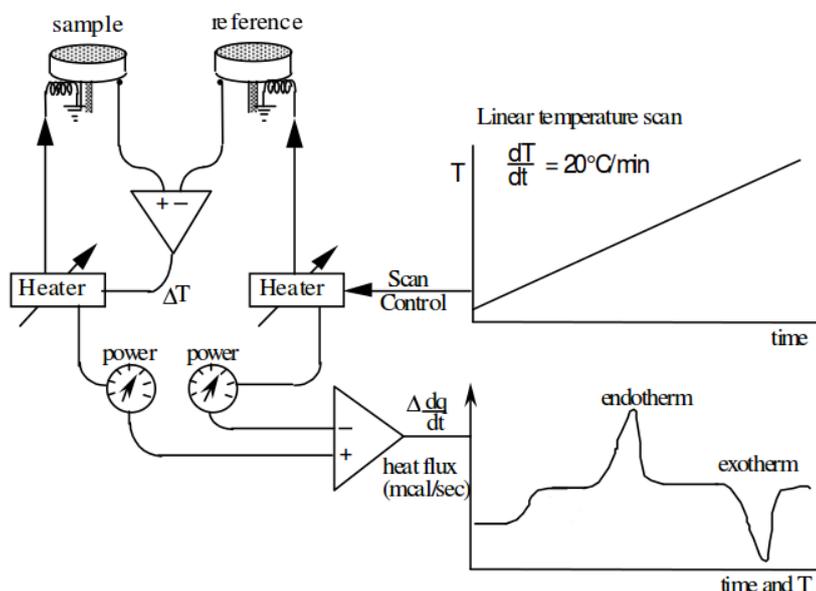


Figura 2.23: Rappresentazione schematica di un calorimetro a scansione differenziale. I triangoli sono amplificatori. Sono inoltre riportate la dipendenza della temperatura dal tempo e della flusso di calore dalla temperatura. Immagine modificata a partire da [27].

2.4.3 Calorimetria

Il ripiegamento e quindi la formazione della struttura tridimensionale di una proteina è un processo reversibile e termodinamico, come dimostrato dall'esperimento di Anfinsen. Il suo studio ha richiesto lo sviluppo di due tecniche ultra-sensibili, denominate calorimetria differenziale a scansione (DSC) e calorimetria isoterma di titolazione (ITC). La DSC prevede il confronto fra il campione da analizzare e un riferimento, scaldando i due elementi in modo tale che la loro temperatura resti uguale. Dato che campione e riferimento sono diversi, quello che si misura è la differenza in calore fornito al campione e al riferimento, ottenendo un grafico che misura questa differenza in funzione della temperatura. Lavorando a pressione costante, la variazione di calore è pari alla variazione d'entalpia

$$\left(\frac{dQ}{dT}\right)_P = \frac{dH}{dT} \quad (2.15)$$

Possiamo quindi ottenere l'entalpia di un processo dall'integrazione del picco:

$$\int \frac{dH_{sample}}{dT} dT = \Delta H_{sample} \quad (2.16)$$

La capacità termica viene quindi determinata a partire dal confronto con la capacità termica di un campione noto

$$C_p = C_{p,ref} \frac{\left(\frac{dq}{dT}\right)_P m}{\left(\frac{dq_{ref}}{dT}\right)_P m_{ref}} \quad (2.17)$$

con m e m_{ref} massa del campione e del campione noto, rispettivamente. In Figura 2.23 sono riportati un generico DSC, una transizione endotermica, che assorbe calore e quindi $\Delta \frac{dQ}{dT}$ ha un picco, ed una esotermica, che emette calore e quindi $\Delta \frac{dQ}{dT}$ ha una valle. La tecnica ITC, invece, misura la variazione di energia di un sistema a temperatura costante e la variabile decisa dall'operatore è la concentrazione di un reagente. L'esperimento è condotto titolando una soluzione con un titolante, e monitorando il calore assorbito o rilasciato dalla reazione fra titolante e reagenti nella soluzione. In Figura 2.24 è riportato un esempio di ITC, in cui le termopile rappresentano l'unica connessione fra le celle di titolazione ed un dissipatore ed agiscono da detector, e generano un potenziale proporzionale alla differenza di temperature fra dissipatore e cella, a sua volta proporzionale alla velocità di trasferimento di calore fra la cella ed il dissipatore. La quantità monitorata è quindi la differenza nella produzione di calore fra la cella con il campione ed una cella di riferimento. Un nanovoltmetro misura le differenze di potenziale fra le due termopile, differenze funzione delle differenze di temperature fra le celle, ed invia il dato ad un computer, che corregge la corrente applicata agli elementi delle celle per bilanciare le differenze di temperatura. La quantità misurata sperimentalmente è quindi la velocità di trasferimento di calore, proporzionale al calore sviluppato dalla reazione. Facendo una serie di misure a temperature diverse è possibile calcolare la capacità termica del campione. Da un grafico della capacità termica C_p in funzione della temperatura di una proteina, si possono ricavare molte informazioni. L'area del picco rappresenta la variazione di entalpia dovuta al processo di unfolding, come definito in precedenza. Dalla forma del picco, possiamo invece calcolare l'entalpia di van't Hoff, definita come

$$\Delta H_{vH} = \frac{4RT_t^2 C_{p,max}}{\Delta H_{cal}} \quad (2.18)$$

dove T_t è la temperatura di transizione, R è la costante universale dei gas, $C_{p,max}$ la capacità termica massima, e ΔH_{cal} è l'area del picco della transizione. Nel caso in cui le due entalpie misurate abbiano lo stesso valore, si può assumere che il processo studiato sia senza stati intermedi. In questo caso la variazione di energia libera alla temperatura di unfolding è zero e possiamo

2.4. METODI SPERIMENTALI PER LA CARATTERIZZAZIONE DELLA PROTEINA31

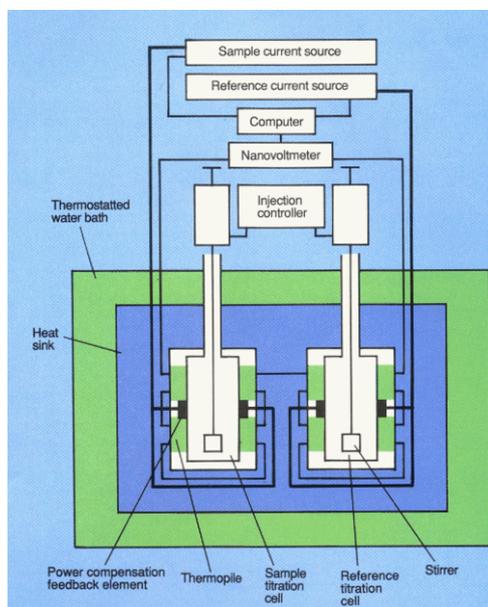


Figura 2.24: Calorimetro a titolazione isoterica. Immagine tratta da [28].

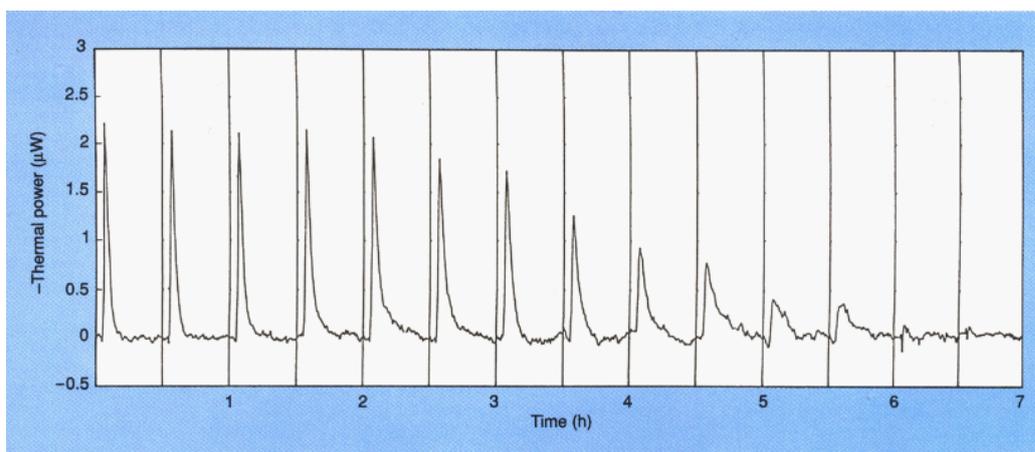


Figura 2.25: Ttitolazione isoterica di un pentamero di una sotto-unità della tossina del colera. Immagine tratta da [28].

calcolare l'entropia di folding. Dalle relazioni termodinamiche

$$\Delta H = T\Delta S + V\Delta P \quad (2.19)$$

e

$$\Delta Q = T\Delta S \quad (2.20)$$

definendo l'entropia e l'entalpia dovute all'*unfolding* come $\Delta_N^U S(T_t)$ e $\Delta_N^U H(T_t)$, rispettivamente, ossia come le differenze fra entropia ed entalpia nello proteina denatura e nello stato nativo si ottiene che

$$\Delta_N^U S(T_t) = \frac{\Delta_N^U H(T_t)}{T_t} \quad (2.21)$$

Una volta note le capacità termiche a varie temperature, possiamo calcolare l'entalpia e l'entropia di *folding* in funzione della temperatura:

$$\Delta_N^U H(T) = \Delta_N^U H(T_t) + \int_{T_t}^T \Delta_N^U C_p(T') dT' \quad (2.22)$$

e

$$\Delta_N^U S(T) = \frac{\Delta_N^U H(T_t)}{T_t} + \int_{T_t}^T \Delta_N^U C_p(T') d \ln T' \quad (2.23)$$

Solitamente gli esperimenti per la misura delle energie di folding sono condotti in condizioni sperimentali per cui si ottengono temperature di folding diverse. Questa situazione si può ottenere o lavorando a pH diversi oppure, oltre alla variazione del pH, si possono usare agenti denaturanti. La dipendenza della temperatura di transizione dal pH può essere usata per calcolare il numero di idrogeni rilasciati durante l'*unfolding*:

$$\Delta_t v = - \frac{\Delta H_{cal}(T_t)}{2.303RT_t^2} \frac{dT_t}{d(pH)}. \quad (2.24)$$

Capitolo 3

Dinamica Molecolare

La dinamica molecolare usa la meccanica classica per simulare l'evoluzione temporale di un sistema usando la seconda legge di Newton, che lega l'accelerazione del sistema al gradiente del potenziale. Per sistemi conservativi si trova che

$$m \frac{d^2 \mathbf{r}_l}{dt^2} = -\nabla_l \mathcal{V}(\mathbf{r}_1 \dots \mathbf{r}_n) \quad (3.1)$$

dove $\mathcal{V}(\mathbf{r}_1 \dots \mathbf{r}_n)$ è il potenziale di un sistema di N molecole di massa m aventi posizioni $(\mathbf{r}_1 \dots \mathbf{r}_n)$. Questa accelerazione viene calcolata per ogni componente del sistema, le posizioni delle molecole vengono aggiornate, e si ricalcola l'accelerazione. La variazione nel tempo delle posizioni genera la traiettoria del sistema, da cui sono calcolabili le proprietà macroscopiche. L'uso della meccanica classica implica la perdita di informazioni a livello quantistico ma permette di studiare sistemi più grandi. I potenziali classici (*force field*) sono opportunamente parametrizzati per rappresentare i sistemi in modo da ottenere osservabili macroscopici quanto più vicini possibile alla realtà.

3.1 *Force field*

Le interazioni che compongono il sistema sono studiate considerando gli atomi come delle sfere rigide legate da molle. A partire da questa descrizione l'energia del sistema è poi calcolata come due macro-contributi, uno di legame dovuto a variazioni nella lunghezza l del legame fra due corpi, nell'angolo θ formato da tre corpi, e nell'angolo diedro ϕ che descrive l'orientamento reciproco di due piani formati da quattro corpi, ed uno di non legame dovuto ad un termine coulombiano ed un termine di Lennard-Jones.

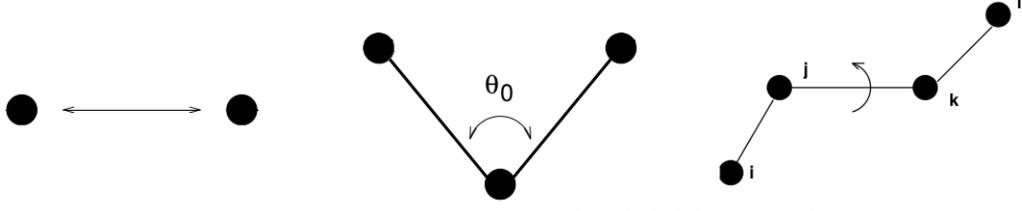


Figura 3.1: Le tre componenti bonded del potenziale.

$$\begin{aligned}
 \mathcal{V}(\mathbf{r}_1 \dots \mathbf{r}_N) = & \sum_i \frac{k_i}{2} (l_i - l_{i,0})^2 && \{\text{Termine armonico lunghezza legame}\} \\
 & + \sum_i \frac{h_i}{2} (\theta_i - \theta_{i,0})^2 && \{\text{Termine armonico angolo legame}\} \\
 & + \sum_{i,n} \frac{V_{i,n}}{2} (1 + \cos(n\omega_i - \gamma_i)) && \{\text{Termine armonico angoli diedri}\} \\
 & + \sum_{i < j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] && \{\text{Termine Lennard - Jones}\} \\
 & + \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon r_{ij}} && \{\text{Termine coulombiano}\}
 \end{aligned} \tag{3.2}$$

dove k_i , h_i , e $V_{i,n}$ sono delle costanti elastiche, e $\sum_{i < j} = \sum_{i=1}^N \sum_{i=1; j > i}^N$. I potenziali legati alle lunghezze e agli angoli sono calcolati in modo simile, ossia con un potenziale di tipo armonico rispetto ad una posizione di equilibrio. Il termine legato ai diedri è calcolato come una combinazione lineare di funzioni periodiche, dove ω_i è l'angolo e γ è la fase. Il termine di Lennard-Jones è composto da un termine a corto raggio repulsivo ed un termine a più lungo raggio attrattivo che rappresenta i contributi di tipo vdW. Questa equazione rappresenta un modello generale, i termini sono dipendenti da una serie di parametri che costituiscono i diversi *force field* e li rendono quindi degli strumenti utili per sistemi specifici. Va notato che le interazioni di non legame fra atomi più vicini di 3 legami, sono calcolate considerando un fattore di smorzamento il cui valore dipende dal *force field*.

3.2 Modelli per l'acqua

Tutte le simulazioni eseguite nel corso di questa tesi sono state condotte usando acqua pura come solvente. Una modellazione accurata dell'acqua è quindi

un requisito fondamentale per ottenere dei risultati affidabili. I vari modelli per l'acqua si possono dividere in grosse famiglie basate sul numero di siti che descrivono il modello. Un maggior numero di siti aumenta la precisione del modello, però porta anche ad un rapido aumento delle coordinate richieste per calcolare le energie di interazione. L'energia potenziale è calcolata come somma di termini coulombiani e di Lennard-Jones [29]:

$$\mathcal{V}_w(\mathbf{S}_i \dots \mathbf{S}_n) = \sum_{l < m} f \frac{q_l q_m}{S_{lm}} + \sum_{l < m} 4\epsilon \left[\left(\frac{\sigma_{lm}}{S_{lm}^{OO}} \right)^{12} - \left(\frac{\sigma_{lm}}{S_{lm}^{OO}} \right)^6 \right] \quad (3.3)$$

Con $f = 138.935 \text{ 485 kJ mol}^{-1} \text{ nm e}^{-2}$, $S_{lm} = |\mathbf{S}_l - \mathbf{S}_m|$ indica la distanza fra due atomi, e $S_{lm}^{OO} = |\mathbf{S}_l^{OO} - \mathbf{S}_m^{OO}|$ indica la distanza fra due atomi di ossigeno. Nel corso di questa tesi abbiamo usato due modelli per l'acqua, il modello TIP3P ed il modello SPC. Sono entrambi modelli a 3 siti, dove ogni sito rappresenta un atomo della molecola. Ogni sito è dotato di una carica ed il sito dell'ossigeno ha dei parametri LJ. La tabella 3.1 riporta i parametri usati per descrivere i due modelli di acqua usati nelle nostre simulazioni.

	SPC	TIP3P
r_{OH} (nm)	0.1	0.09572
K_r (kJ mol ⁻¹ nm ⁻¹)	345000	502416.0
θ_{HOH} (deg)	109.47	104.52
K_θ (kJ mol ⁻¹ deg ⁻¹)	383	628
q_O (e)	0.82	0.834
q_H (e)	0.41	0.417
σ_O (nm)	0.316557	0.315061
ϵ_O (kJ mol ⁻¹)	0.650629	0.636386
σ_H (nm)	0	0
ϵ_H (kJ mol ⁻¹)	0	0

Tabella 3.1: Parametri dei modelli SPC e TIP3P per l'acqua.

Si vede come solo le interazioni fra ossigeni contengono dei termini di LJ. La differenza più evidente fra i due modelli è l'angolo fra gli idrogeni, nel modello SPC la molecola viene rappresentata con l'angolo ideale tetraedrico di 109.47° mentre il modello TIP3P usa il valore osservato dell'angolo. L'uso di un modello rispetto all'altro porta ad ottenere delle proprietà (densità, capacità termica, etc.) diverse, ed una diversa struttura dell'acqua. Uno studio sulle proprietà e strutture dell'acqua si può trovare in [29], dove si vede come le proprietà calcolate dai due modelli siano simili ed in buon accordo con valori sperimentali. L'eccezione è rappresentata dalla compressibilità isoterma ed il coefficiente di espansione termico, entrambi molto lontani

dal valor vero. Aumentare il numero di siti, spostando dall'ossigeno verso gli idrogeni, non migliora le cose. L'unica differenza fra i due modelli in questo contesto è l'elevata sovrastima di C_p da parte di SPC. Il modello SPC, d'altra parte, si dimostra migliore nel rappresentare la struttura dell'acqua bulk, in particolare nel secondo picco della funzione di distribuzione radiale OO.

3.3 Condizioni periodiche

Per evitare artefatti dovuti alla presenza di una superficie e del nulla oltre i confini del sistema, si usano condizioni al contorno periodiche dove il sistema viene considerato come una cella all'interno di un cristallo formato da una infinita serie di identiche celle in tutte le direzioni. Questo consente ad esempio l'uscita di una particella da un lato ed il suo ingresso nella direzione opposta. Questa soluzione introduce però un nuovo tipo di problema, cioè la possibilità per un atomo in una delle celle di interagire con la sua copia in un'altra cella. Questo problema è risolto usando la distanza di cut-off, ossia nel calcolo dei termini di non legame solo gli atomi all'interno di una sfera con un raggio arbitrario verranno considerati. La sfera può includere più celle.

3.4 Schema generale di una simulazione

Una simulazione di dinamica molecolare generica può essere scomposta in una serie di passaggi.

Solvatazione La struttura di partenza viene posta in un box di dimensioni specificate e viene circondata con un solvente, solitamente acqua. Oltre alle molecole di acqua, possono essere introdotte atomi che costituiscono un sale, e può essere stabilito il pH del sistema.

Minimizzazione Prima di iniziare la simulazione è necessario essere certi che non ci siano errori dovuti a sovrapposizioni non fisiche fra atomi. Questo rilassamento del sistema avviene attraverso il processo di minimizzazione del sistema, in cui il sistema viene portato verso un minimo energetico.

Heating O simulazione NVT. Il sistema viene portato ad una temperatura stabilita, mentre la proteina è costretta ed è soggetta ad una minor dinamica.

Equilibrio O simulazione NPT. Il sistema viene portato ad una pressione stabilita, mantenendo la temperatura costante e con limitata mobilità.

Production run

Il sistema viene lasciato libero di muoversi, e si ottengono informazioni sulla traiettoria e sulle proprietà del sistema. Può essere condotta in condizioni NVT o NPT.

3.5 Dettagli dell'implementazione

Scopo di questa prima parte è quella della scelta del modello più adatto, mediante validazione sperimentale. Lo studio dei termini NMR è stato condotto mediante simulazioni di dinamica molecolare con Gromacs 4.6.5. La proteina usata per lo studio è l'ubiquitina umana, discussa nel capitolo 3. La struttura di partenza è quella contenuta nel protein data bank [3], codice 1UBQ. Come modelli per l'acqua sono stati usati il modello SPC ed il modello TIP3P, abbinati a potenziali AMBER99[30], AMBER99sb[31], AMBER99sb-ildn [32], CHARMM27 [33] [34] e OPLSAA [35] [36]. Una descrizione dei modelli per l'acqua e dei *force field* si trova nei capitoli 4 e 5. La proteina è stata solvata in un box dodecaedrico di raggio 1.2 nm e con 7165 molecole di acqua e si è minimizzata con algoritmo steep, per un massimo di 50000 step, e con *emtol*, un termine che indica la tolleranza della variazione di energia del sistema entro cui si considera conclusa la procedura di minimizzazione, $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. Dopo la minimizzazione la proteina è stata portata ai valori desiderati di temperatura (300K), prima, e di pressione (1 bar), poi, mediante due simulazioni della durata ciascuna di 50 ps; 25000 step di integrazione per 0.002 ps come intervallo d'integrazione per la simulazione NVT atta a portare il sistema a 300K passando attraverso 0, 50 e 150K, e 50000 step per 0.001 ps d'intervallo per la dinamica NPT. Per il coupling della temperatura e della pressione sono stati usati gli algoritmi di Berendsen.

Le dieci strutture ottenute sono state poi studiate con due production run, distinte e non sequenziali, una in condizioni NPT ed una in condizioni NVT. Sono state ottenute in tutto quindi venti traiettorie di production run. Le run NPT sono state condotte con integratore md, 5 ns (5000000 step per 0.001 ps di intervallo), con algoritmo v-rescale per la temperatura e Berendsen per la pressione. Le corse AMBER99/SPC/NPT, AMBER99sb/SPC/NPT, e AMBER99sb-ildn/SPC/NPT sono state prolungate di altri 15 ns per un totale di 20 ns di production run. Per tutte le simulazioni, incluse quelle di preparazione esclusa la minimizzazione, è stato usato l'algoritmo SHAKE

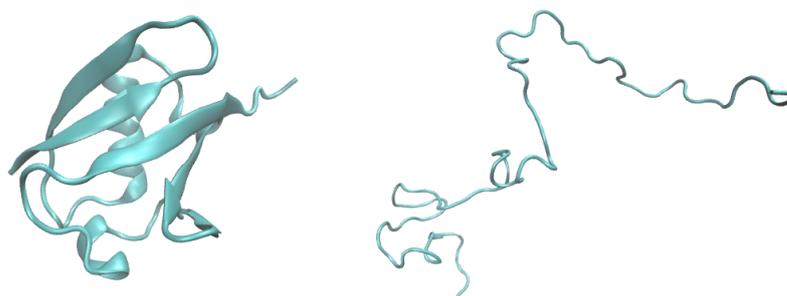


Figura 3.2: Due strutture per 1ubq. A sinistra la struttura nativa, a destra una delle possibili conformazioni della proteina denaturata.

per i vincoli. Per la minimizzazione è stato necessario usare quello LINCS in quanto specificatamente richiesto dall'integratore steep.

Lo studio dei termini calorimetrici è stato eseguito su due strutture distinte, usando il modello TIP3P per l'acqua ed il FF AMBER99sb-ildn. Un primo studio è stato condotto su una struttura aperta della proteina Ubiquitina umana. Questa struttura è stata ottenuta partendo dalla struttura nativa della proteina, sottoposta ad una preparazione costituita da minimizzazione con algoritmo steep, per un massimo di 50000 step, e con *emtol* $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. Si è poi fatto un heating a 1000K per 2ns, condotto senza restrizioni per le posizioni degli atomi in modo da aprire la proteina. Da questa traiettoria è stato estratto un frame caratterizzato dall'assenza di strutture secondarie e legami idrogeno interni. Questa struttura è proposta nella Figura 3.2 a destra, mentre a sinistra è riportata la struttura nativa. La struttura così ottenuta è stata minimizzata due volte. La prima è stata una minimizzazione usando l'integratore steep, la seconda l'integratore l-bfgs. Entrambe le minimizzazioni sono state condotte per un massimo di 50000 step con *emtol* 10.0. La proteina è stata scaldata a 300K in 500 ps (500000 step per 0.001 ps d'intervallo) con integratore md e successivamente portata a pressione 1 bar con impostazioni identiche ed usando l'algoritmo di Parrinello-Rahman. La production run è stata condotta per 1 ns (500000 step per 0.002 ps d'intervallo con integratore sd). L'interazione soft-core è stata descritta con *sc-alpha* = 0.5, *sc-power* = 1 e *sc-sigma* = 0.3. Le simulazioni sono state eseguite con queste impostazioni per diversi $\Delta\lambda$, ossia per diverse distanze in senso di accoppiamento fra la proteina ed il solvente.

Capitolo 4

Validazione della dinamica

Il confronto fra i vari *force field* nelle varie condizioni di simulazione è basato sul confronto dei parametri d'ordine nematico S^2 per il legame N-H, quadrato del parametro d'ordine generalizzato definito da Lipari e Szabo [37], ottenuti dalle nostre simulazioni e quelli ottenuti in maniera sperimentali dal gruppo di Tjandra [38]. Il termine S^2 è stato ricavato originalmente come segue. In soluzione il rilassamento dovuto alle interazione dipolo-dipolo fra due nuclei può essere descritto dalla funzione di correlazione

$$C(t) = \frac{1}{5} \langle P_2(\boldsymbol{\mu}_{LF}(0) \cdot \boldsymbol{\mu}_{LF}(t)) \rangle \quad (4.1)$$

dove $\boldsymbol{\mu}_{LF}$ è il versore che unisce i due nuclei nel sistema di coordinate del laboratorio e P_2 è il secondo polinomio di Legendre $P_2(x) = \frac{1}{2}(3x^2 - 1)$. La media è su tutte le possibili configurazioni. La funzione di correlazione descrive come varia nel tempo la correlazione fra due quantità. Nel caso isotropico la funzione di correlazione può essere fattorizzata esattamente come

$$C(t) = C_0(t)C_1(t) \quad (4.2)$$

dove

$$C_0(t) = \frac{1}{5} e^{-t/\tau_M} \quad (4.3)$$

con τ_M tempo di correlazione della macromolecola, è la funzione di correlazione del moto complessivo, e

$$C_1(t) = \langle P_2(\boldsymbol{\mu}(0) \cdot \boldsymbol{\mu}(t)) \rangle \quad (4.4)$$

con $\boldsymbol{\mu}$ versore che descrive l'orientazione del vettore d'interazione dipolo-dipolo in un nuovo riferimento, è la funzione di correlazione del moto interno.

$C_1(t)$ può essere a sua volta espresso con una serie di esponenziali

$$C_1(t) = \sum_{i=0} a_i e^{-\frac{t}{\tau_i}} \quad (4.5)$$

dove a_i sono i coefficienti per il peso del tempo caratteristico, mentre τ_i è il tempo di correlazione. Si può dimostrare [37] che

$$C_1(\infty) = \sum_{m=-2}^2 |\langle C_{2m}(\Omega) \rangle|^2 \quad (4.6)$$

dove C_{2m} sono le armoniche sferiche modificate di Brink e Satchler, [37] e Ω_i sono gli angoli polari di $\boldsymbol{\mu}_i$. Leghiamo la funzione di correlazione ad un tempo molto lungo con il parametro d'ordine generalizzato come

$$C_1(\infty) = S^2 \quad (4.7)$$

S^2 è compreso fra 1 e 0. Se il moto interno è isotropico e tutte le orientazioni di $\boldsymbol{\mu}$ sono equiprobabili, $S^2 = 0$. Se invece il moto è completamente bloccato, $S^2 = 1$. Si vede Appendice C per dettagli. Il parametro d'ordine consente di approssimare la funzione di correlazione per il moto interno come

$$C_1(t) = S^2 + (1 - S^2)e^{-t/\tau_e} \quad (4.8)$$

ossia C_1 decade con un tempo di correlazione effettivo τ_e ad un valore di plateau S^2 . In questa tesi S^2 è stato calcolato a partire dalla funzione di correlazione dei singoli N-H.

		R_p	R_s	RMSD	$\langle S_{md}/S_{exp} \rangle$
AMBER99	spc-npt	0.889	0.669	0.095	0.943 ± 0.168
	spc-nvt	0.836	0.634	0.074	1.015 ± 0.155
	tip3p-npt	0.870	0.722	0.081	0.965 ± 0.141
	tip3p-nvt	0.889	0.724	0.077	0.971 ± 0.187
AMBER99sb	spc-npt	0.937	0.653	0.051	0.995 ± 0.088
	spc-nvt	0.905	0.666	0.060	1.008 ± 0.109
	tip3p-npt	0.948	0.659	0.059	0.970 ± 0.143
	tip3p-nvt	0.930	0.678	0.081	0.943 ± 0.184
AMBER99sbildn	spc-npt	0.949	0.659	0.066	0.959 ± 0.165
	spc-nvt	0.906	0.730	0.066	1.000 ± 0.190
	tip3p-npt	0.908	0.698	0.072	0.975 ± 0.163
	tip3p-nvt	0.931	0.695	0.055	0.991 ± 0.141

Tabella 4.1: Risultati statistici delle simulazioni.

I valori riportati sono R_p , il coefficiente di Pearson, R_s , coefficiente di Spearman, RMSD, e la media del rapporto fra i valori di S^2 ottenuti e quelli sperimentali ottenuti da Tjandra et al. [38] R_p è stato calcolato come

$$R_p = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (4.9)$$

Dove x ed y sono rispettivamente il valore teorico e quello sperimentale, e \bar{x} ed \bar{y} indicano la media dei valori.

R_s è stato invece calcolato come

$$R_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \quad (4.10)$$

dove d^2 è definito come il quadrato della differenza fra il rank dei valori teorici e sperimentali, ed n è il numero di misure.

RMSD è invece definito come

$$RMSD = \sqrt{\frac{1}{n} \sum_i (x_i - y_i)^2} \quad (4.11)$$

con n numero dei residui. Dopo le analisi sui termini NMR, la prima analisi che abbiamo fatto è il calcolo della root mean square deviation della backbone della proteina. Nel contesto della dinamica molecolare con RMSD s'intende la differenza fra la struttura in un determinato tempo ed una struttura di partenza:

$$RMSD(t_1, t_2) = \sqrt{\frac{\sum_{i=1}^N m_i [(r_i(t_1) - r_i(t_2))]^2}{\sum_{i=1}^N m_i}} \quad (4.12)$$

L'RMSD è stato calcolato solo per le run AMBER99, AMBER99sb e AMBER-99sb-ildn, in condizioni NPT con acqua SPC, e solo per i residui 1-72. Gli ultimi quattro residui compongono una coda estremamente mobile. I risultati sono essenzialmente comparabili. Come risulta dalla Figura 4.1 che riporta la RMSD dei C_α , l'unica differenza notevole è un picco elevato intorno a 1420 ps per AMBER99sb (in rosso). Successivamente abbiamo calcolato la *root mean square fluctuations*, RMSF, che rappresenta il movimento medio di uno specifico residuo, ossia RMSD calcolato per un residuo mediato nel tempo. In accordo con [39] vediamo come il forcefield AMBER99sb si dimostri più rigido dell'originale AMBER99. Il più nuovo AMBER99sb-ildn, invece, offre risultati contrastanti rispetto a AMBER99sb, più rigido a volte più lasco altre. RMSD e RMSF sono riportati nelle Figure 4.1 e 4.2.

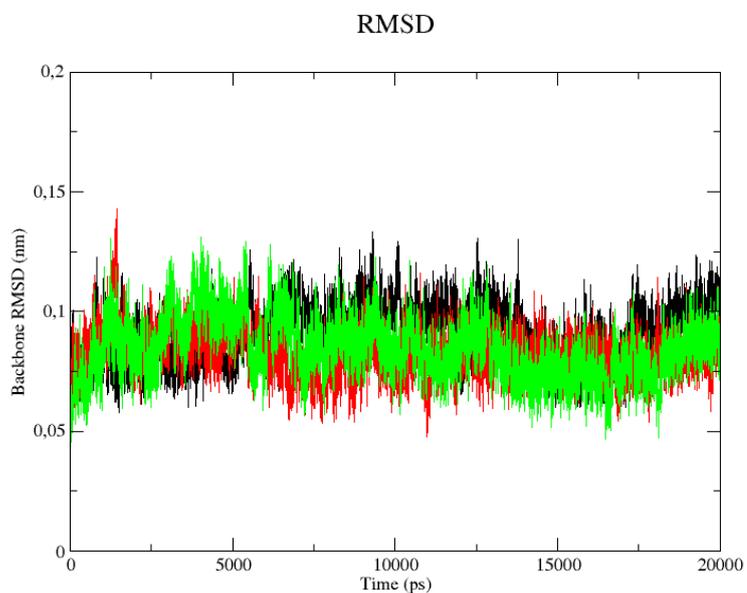


Figura 4.1: RMSD per i primi 72 residui. In nero AMBER99, in rosso AMBER99sb, in verde AMBER99sb-ildn.

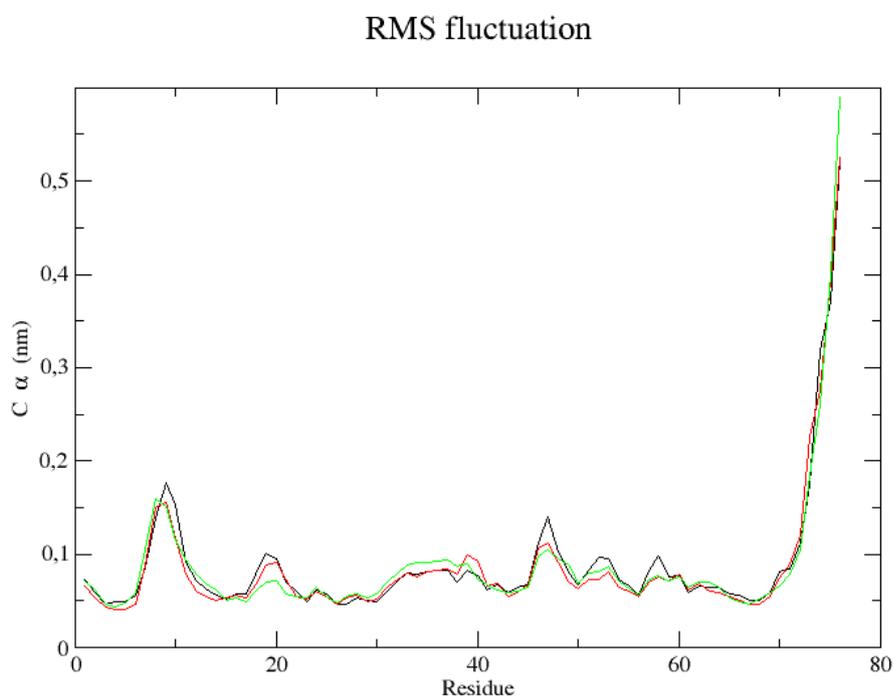


Figura 4.2: RMSF dei C_{α} della proteina. In nero AMBER99, in rosso AMBER99sb, in verde AMBER99sb-ildn.

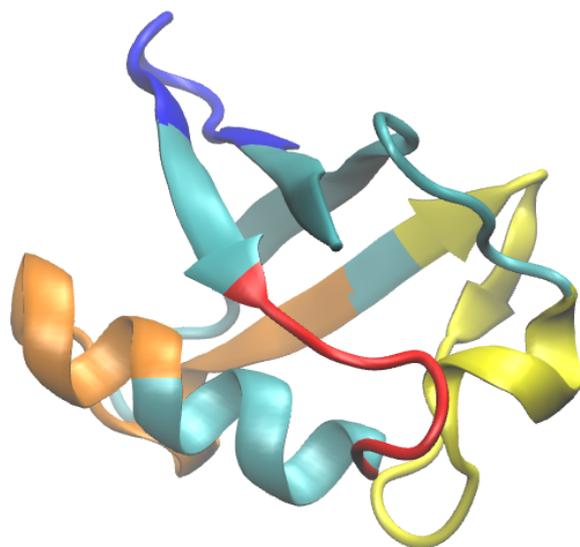


Figura 4.3: Ubiquitina ripiegata con le zone mobili marcate.

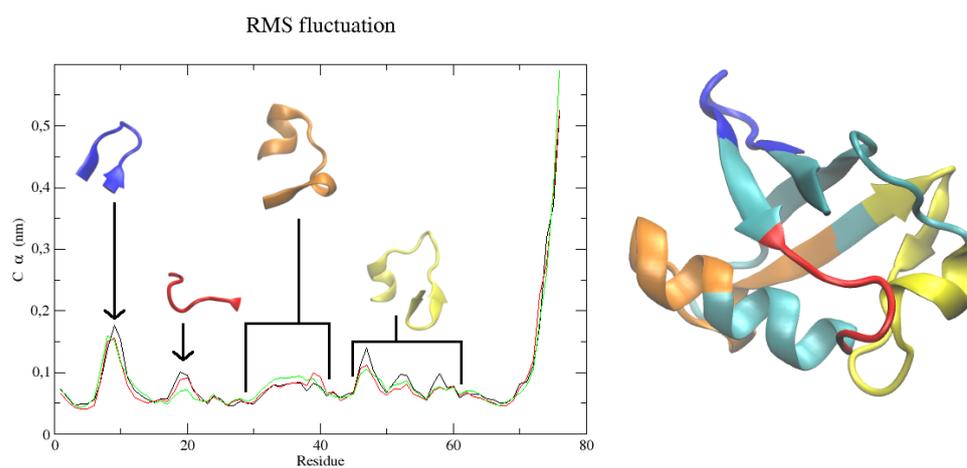


Figura 4.4: RMSF della proteina mostrata insieme alle zone mobili della proteina.

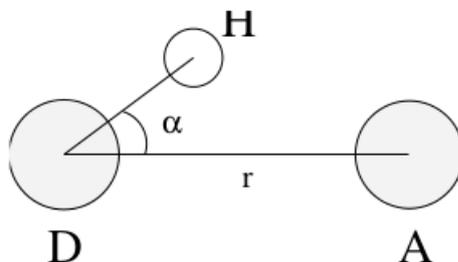


Figura 4.5: Criteri geometrici per il legame idrogeno. D è un generico atomo donatore, A un generico atomo accettore. Immagine tratta dal manuale di GROMACS [40].

L'analisi successiva è stata di fatto il conteggio dei legami idrogeno intra-proteina e fra la proteina e l'acqua. I legami idrogeno sono stati definiti sulla base di criteri geometrici di distanza ed orientazione, mostrati in Figura 4.5. Per essere un legame idrogeno $r \leq 0.35$ nm e $\alpha \leq 30^\circ$. I risultati si possono trovare nella Tabella 4.2 e nella Figura 4.6. I risultati sono sostanzialmente identici, anche se la dinamica di queste simulazioni specifiche non è la stessa, inclusa una sottostima del numero di legami idrogeno interni rispetto a quanto riportato in letteratura, [41], dove sperimentalmente sono stati misurati 76 legami intramolecolari.

	P-P	P-W	Totale
AMBER99	58	188	246
AMBER99-sb	58	188	246
AMBER99-sbildn	58	189	248

Tabella 4.2: Media dei legami idrogeno intramolecolari P-P, fra proteina ed acqua P-W, e totali.

Dalla Figura 4.6 si vede come la fluttuazione per i legami idrogeno proteina-proteina sia all'interno del 10 %.

Nelle Tabelle 4.3 e 4.4 sono riportati alcuni risultati delle nostre simulazioni. I tre *force field* risultano praticamente identici, in particolare i volumi misurati rientrano nei rispettivi errori.

	K (kJ mol ⁻¹)	H (kJ mol ⁻¹)	V (kJ mol ⁻¹)
AMBER99	57423.0 ± 1.1	-251646.0 ± 7.5	-309092.0 ± 6.8
AMBER99-sb	57443.1 ± 1.5	-251525.0 ± 8.7	-308972.0 ± 8.9
AMBER99-sbildn	57432.2 ± 1.1	-251647.0 ± 5.1	-309093.0 ± 5.1

Tabella 4.3: Energia cinetica, entalpia, ed energia potenziale per le tre corse di 20 ns.

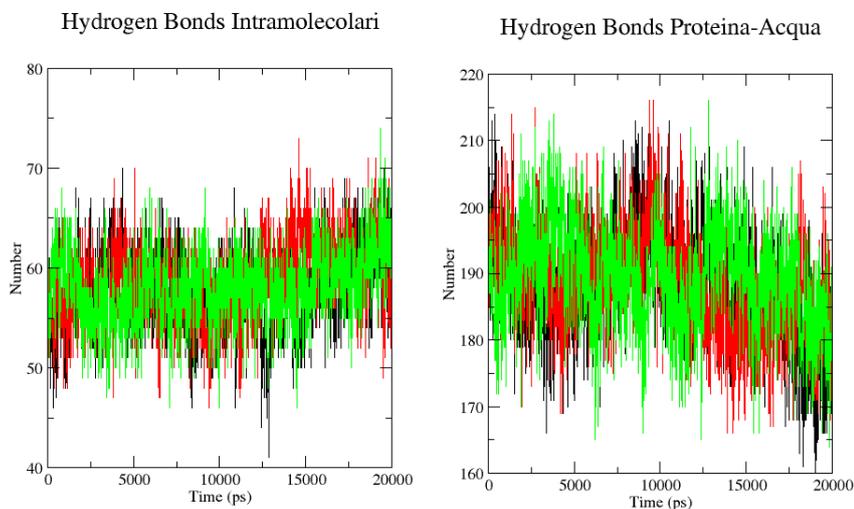


Figura 4.6: Legami idrogeni intramolecolari e con l'acqua nelle 3 simulazioni. In nero AMBER99, in rosso AMBER99sb, in verde AMBER99sb-ildn.

	Volume (nm ³)	Densità (kg nm ⁻³)
AMBER99	231.30 ±0.09	988.19 ±0.04
AMBER99-sb	231.30 ±0.04	988.19 ±0.02
AMBER99-sbildn	231.33 ±0.05	988.07 ±0.02

Tabella 4.4: Volumi e densità per le tre corse di 20 ns.

Dalla Tabella 4.1 e dalla Figura 4.4 si riscontra quanto già riportato in studi precedenti [39], ossia un miglior accordo fra il set sperimentale e i valori calcolati usando AMBER99sb rispetto a AMBER99. Al contrario di quanto riportato da Showalter [39] però, in queste simulazioni la dinamica molecolare sovrastima la dinamica delle proteine, in quando otteniamo mediamente un S_{md} più basso di S_{exp} in ogni condizione. La combinazione SPC-nvt invece sottostima la dinamica nei *force field* AMBER99 e AMBER99sb. Le dinamiche misurate usando AMBER99sb e AMBER99sb-ildn sono invece simili, l'unica differenza notevole sembra essere l'RMSF nella sequenza 17-22, un β -turn.

Capitolo 5

Componenti entropiche ed entalpiche del folding

5.1 Ciclo termodinamico

A causa della complessità computazionale e del percorso di simulazione che presenta molteplici minimi locali, non è solitamente possibile simulare il ripiegamento di una proteina. Un caso eccezionale è rappresentato dal supercomputer Anton [1]. In questa tesi abbiamo deciso di studiare i vari termini che compongono l'energia libera del processo di ripiegamento (*folding*) di una proteina, l'ubiquitina umana, senza doverlo simulare direttamente. Il processo di *folding* è stato scomposto in un ciclo termodinamico composto da tre passaggi: la desolvatazione della proteina, la sua apertura, e la sua solvatazione. Il processo è rappresentato schematicamente in Figura 5.1.

La variazione di energia libera dovuta al *folding* è definita come

$$\Delta G_{folding} = \Delta H_{folding} - T\Delta S_{folding} = -\Delta G_{unfolding} \quad (5.1)$$

dove ΔH e ΔS sono la variazione di entalpia ed entropia, rispettivamente. La variazione di energia libera dovuta all'unfolding può essere calcolata come

$$\Delta G_{unfolding}(\lambda = 0) = \Delta G^N(0 \rightarrow 1) + \Delta G_{unfolding}(\lambda = 1) - \Delta G^U(0 \rightarrow 1) \quad (5.2)$$

dove $(\lambda = 0)$ e $(\lambda = 1)$ indicano rispettivamente una condizione di accoppiamento e disaccoppiamento totale fra la proteina e l'acqua, e $\Delta G^N(0 \rightarrow 1)$ e $\Delta G^U(0 \rightarrow 1)$ sono la variazione di energia libera dovuta al passaggio dalla situazione di accoppiamento a quella di disaccoppiamento, per la proteina nello stato nativo e denaturata, in cui le uniche interazioni fra proteina ed acqua sono di tipo sterico. Questo disaccoppiamento rappresenta il processo di

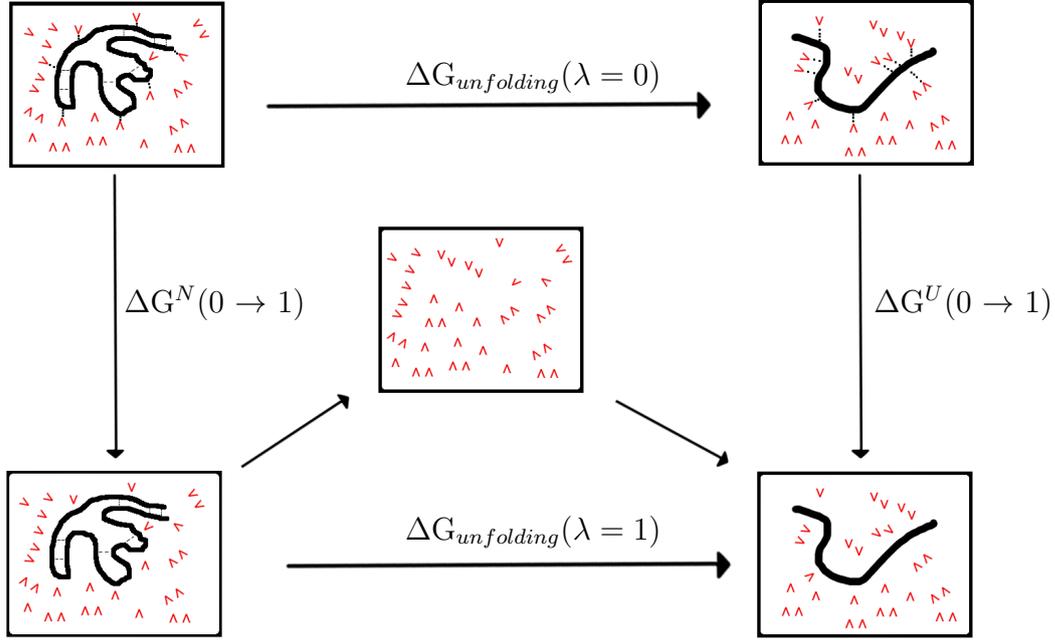


Figura 5.1: Il ciclo termodinamico usato per studiare il sistema.

desolvatazione della proteina. Dei processi $\Delta G^N(0 \rightarrow 1)$ e $\Delta G^U(0 \rightarrow 1)$ noi conosciamo la variazione entropica ed entalpica. Lo studio di questi processi è stato condotto mediante integrazione termodinamica, in cui la variazione di energia libera è calcolata con il seguente integrale

$$\Delta G = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (5.3)$$

con

$$\frac{dG}{d\lambda} = \left\langle \frac{dH}{d\lambda} \right\rangle_{\lambda} \quad (5.4)$$

dove H è l'Hamiltoniano che rappresenta l'energia del sistema. Per una trattazione sull'integrazione termodinamica si veda l'Appendice A. Il processo di desolvatazione della proteina nella struttura nativa è stato eseguito con una dinamica in cui la struttura della proteina è stata artificialmente irrigidita in modo da preservare la struttura durante tutto il processo di desolvatazione. In assenza di questo irrigidimento, si assiste ad una parziale denaturazione della proteina. Questo irrigidimento varia l'energia della proteina di un termine ΔH_{res} .

$$H^N(\lambda = 0) = H_{sim}^N(\lambda = 0) + \Delta H_{res}(\lambda = 0) \quad (5.5)$$

e

$$H^N(\lambda = 1) = H_{sim}^N(\lambda = 1) + \Delta H_{res}(\lambda = 1) \quad (5.6)$$

Il processo di *unfolding* quando la proteina è disaccoppiata può essere trasformato nel processo di riempimento di un buco e formazione di un secondo buco di forma e dimensioni diverse. La variazione di energia libera dovuta a questo processo ha origine puramente entropica e può essere calcolata come [42]:

$$\Delta G_{unfolding}(\lambda = 1) = -k_B T \log\left(\frac{V^*}{V_N}\right) + k_B T \log\left(\frac{V^*}{V_U}\right) \quad (5.7)$$

dove V^* è il volume esclusivamente delle molecole di acqua usate nella simulazione, e V_N e V_U sono i volumi delle molecole di acqua e delle proteina nei due stati nativa ed denaturata quando le cariche sono accoppiate. Possiamo riscrivere questo termine come

$$\Delta G_{unfolding}(\lambda = 1) = k_B T \log\left(\frac{V_N}{V_U}\right) \quad (5.8)$$

e visto che questa variazione di energia libera è dovuta solo all'entropia

$$\Delta S_{unfolding}(\lambda = 1) = -k_B \log\left(\frac{V_N}{V_U}\right) \quad (5.9)$$

Dal calcolo dell'energia libera possiamo risalire alla variazione di entropia nel folding:

$$\Delta S_{unfolding} = \frac{\Delta H_{unfolding} - \Delta G_{unfolding}}{T} \quad (5.10)$$

Abbiamo quindi caratterizzato termodinamicamente il processo di folding di una proteina. Il termine $\Delta S_{unfolding}$ è la somma di due termini, uno dovuto alla proteina ed uno dovuto all'acqua:

$$\Delta S_{unfolding}(\lambda = 0) = \Delta S_p + \Delta S_w \quad (5.11)$$

Il termine $\Delta S_{unfolding}(\lambda = 1)$ contiene esclusivamente il contributo dell'acqua. Inoltre, dato che le interazioni fra solvente e soluto non hanno influenza sul termine entropico, esso è identico al termine ΔS_w . Possiamo quindi risalire al termine entropico dovuto esclusivamente alla proteina come

$$\Delta S_p = \Delta S_{unfolding}(\lambda = 0) - \Delta S_{unfolding}(\lambda = 1) \quad (5.12)$$

5.2 Dati sperimentali per la proteina 1UBQ

La proteina ubiquitina è stata soggetto di diversi studi. Principalmente questi studi sono stati condotti in presenza di un agente denaturante, in quanto l'ubiquitina è una proteina molto stabile con una temperatura di unfolding superiore ai 100°C in pH neutro [43]. Wintrode et al. nel 1994 [44] fece uno studio in cui calcolava sperimentalmente la variazione di entropia ed entalpia di *unfolding* in funzione della temperatura di *unfolding*. Questo studio calcolava queste quantità a partire dalle equazioni 2.22 e 2.23 ottenute mediante DSC. La capacità termica della proteina nello stato nativo oltre i 65 °C è stata stimata per estrapolazione lineare, come mostrato in Figura 5.2 mentre la capacità termica della proteina denaturata è stata ottenuta a partire dalla capacità termica dei singoli amminoacidi. In Tabella 5.1 sono riportate le entalpie ed entropie di unfolding quando la denaturazione termica avviene a 100 e 125 °C.

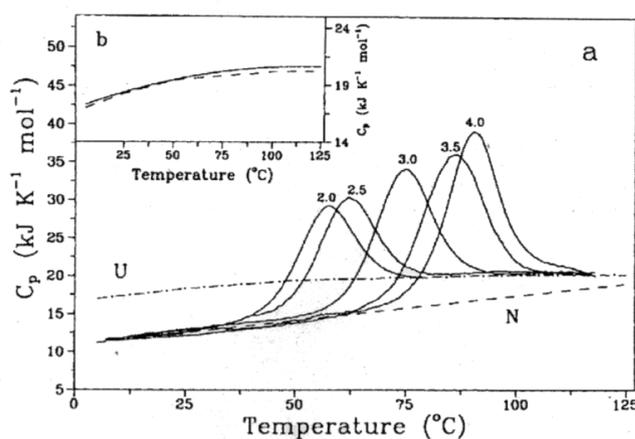


Figura 5.2: Dipendenza dalla temperatura della capacità termica. La linea tratteggiata rappresenta la capacità termica della proteina nella stato nativo, mentre la linea con i pallini rappresenta la capacità termica della proteina denaturata. Immagine tratta da [44].

T_f	100 °C	125 °C
ΔH (kJ mol ⁻¹)	351	393
$T_f \Delta S$ (kJ mol ⁻¹)	390	459

Tabella 5.1: Entropie ed entalpie sperimentali [44]. T_f è la temperatura a cui avviene la denaturazione della proteina.

Non sono disponibili dati sull'unfolding a pH superiori a 4 in assenza di agente denaturante. Dal lavoro di Wintrode e da altra letteratura [45] si può

stimare il raggiungimento di un plateau per pH superiori a 4 per entalpia, energia libera di Gibbs e temperatura di unfolding. Un esempio è riportato in Figura 5.3.

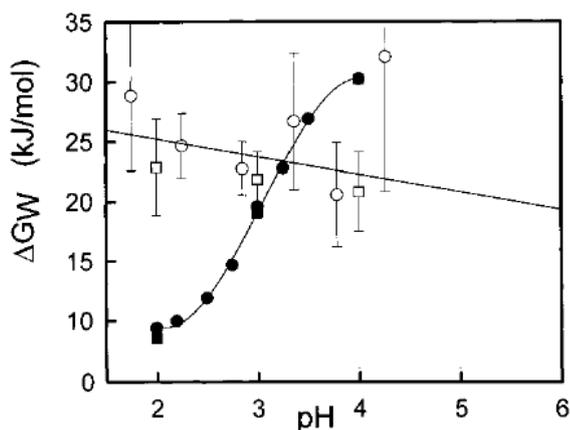


Figura 5.3: Energia libera di Gibbs di *unfolding* in funzione del pH per l'ubiquitina. I simboli in bianco sono dati ottenuti per la denaturazione indotta da guanidina, in nero i dati ottenuti in assenza di denaturante. Nel primo caso i quadrati si riferiscono a dati ottenuti con dicromismo circolare mentre i pallini sono ottenuti da fluorescenza; nel secondo caso i quadrati si riferiscono a dati ottenuti da Ibarra-Molero, mentre i pallini sono i dati riportati da Wintrodte. Immagine tratta da [45].

5.3 Risultati

5.3.1 Legami Idrogeno

Un primo risultato nel confronto fra la proteina nello stato nativo e denaturata è il numero di legami idrogeno, riportato in Tabella 5.2. I numeri riportati hanno un'incertezza approssimativamente del 10% dovuta alle fluttuazioni termiche lungo la traiettoria.

	P-P	P-W	Totale
Nativa	59	189	247
Unfolded	29	266	295

Tabella 5.2: Media dei legami idrogeno intramolecolari, P-P, fra proteina ed acqua, P-W, e totali.

La perdita di legami idrogeno interni dovuta alla denaturazione della proteina viene compensata dall'aumento di legami idrogeno con l'acqua. I due valori ottenuti, all'interno delle fluttuazioni, possono essere considerati simili, indice

del fatto che ci sia un bilanciamento fra i legami idrogeno intramolecolari persi durante la denaturazione e quelli guadagnati con il solvente.

5.3.2 Scorporamento dei vari termini entalpici aggiuntivi

Abbiamo condotto le simulazioni di integrazione termodinamica con diversi $\Delta\lambda$, ossia con una diversa distanza in termini di accoppiamento fra acqua e proteina fra una simulazione e la successiva. Questo termine sembra non avere effetto sul valore e sul comportamento dei due termini, indicando il valore più grande di $\Delta\lambda = 0.05$ come sufficiente per la precisione richiesta. I risultati per la proteina nello stato nativo sono riportati nelle Figure 5.4 e 5.5, quelli per la proteina nello stato di random coil in 5.6 e 5.7. In Figura 5.4 si vedono due risultati inattesi: per λ molto bassi il termine di LJ diventa più attrattivo rispetto all'accoppiamento totale, mentre per λ intorno a 0.8 il termine di LJ diventa repulsivo, mentre ci si attendeva un andamento monotono con un plateau intorno a 0. Nel caso della proteina denaturata si vede come un $\Delta\lambda$ più piccolo porti ad una simulazione più sensibile che individua una serie di fluttuazioni, dovute a fluttuazioni all'interno dell'ensemble denaturato, mentre il termine coulombiano non sembra sentire particolari effetti, Figura 5.6 e 5.7. I picchi positivi intorno a $\Delta\lambda = 0.8$ sono identici per entrambe le forme della proteina, Figura 5.8. La proteina denaturata presenta un termine di LJ più basso della proteina nello stato nativo in quanto ci sono più contatti fra i suoi amminoacidi e l'acqua. Nelle Figure 5.9 - 5.14 sono riportate le variazioni di energia libera fra λ e $\lambda + \Delta\lambda$ durante il processo di disaccoppiamento per vari $\Delta\lambda$ e per gli stati nativo e denaturato della proteina.

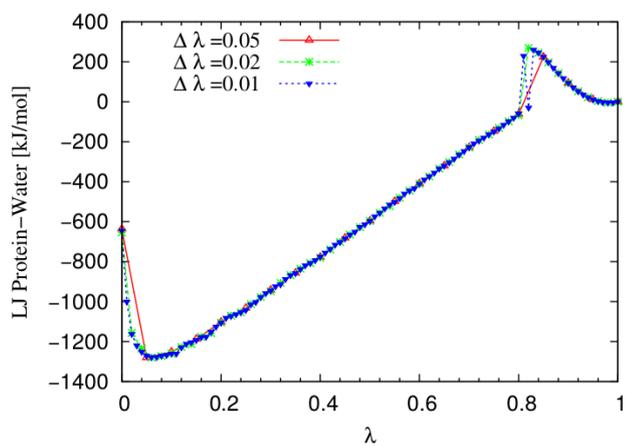


Figura 5.4: Termine LJ fra la proteina nello stato nativo e l'acqua.

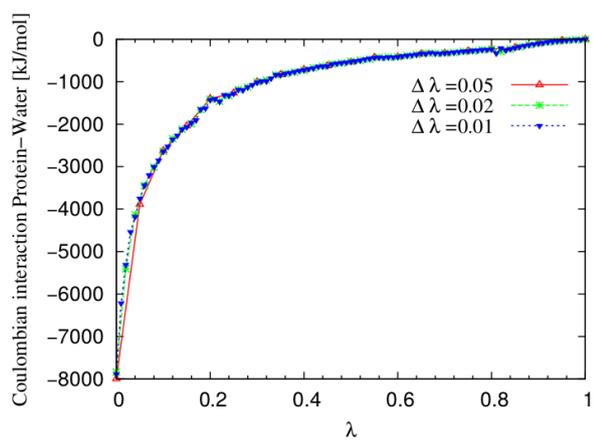


Figura 5.5: Termine coulombiano fra la proteina nello stato nativo e l'acqua.

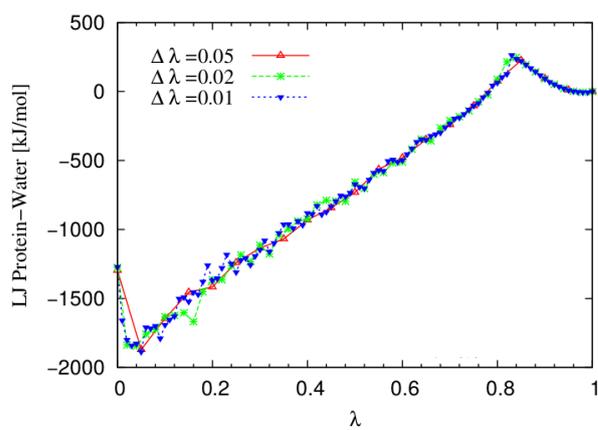


Figura 5.6: Termine LJ fra la proteina denaturata e l'acqua.

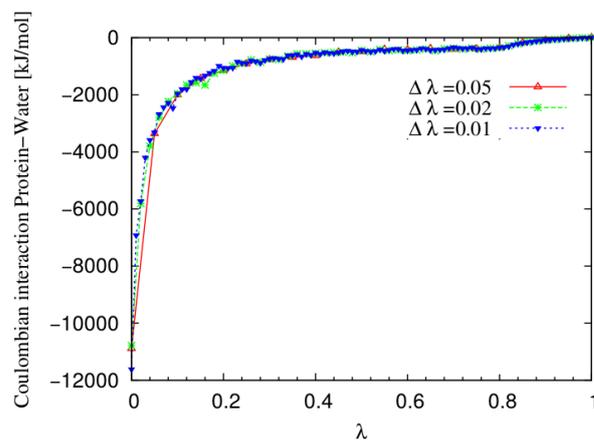
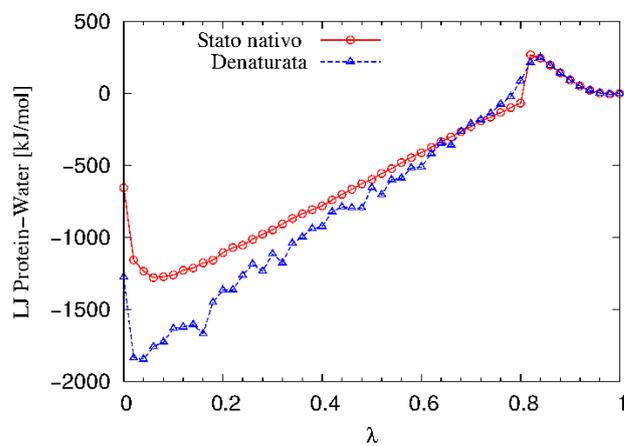


Figura 5.7: Termine coulombiano fra la proteina denaturata e l'acqua.

Figura 5.8: Termine LJ per la proteina nella struttura nativa e denaturata, con $\Delta\lambda=0.01$.

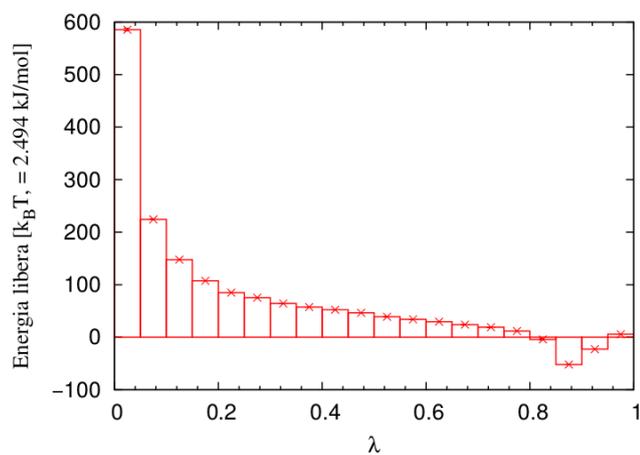


Figura 5.9: $dG/d\lambda$ per la proteina nello stato nativo con $\Delta\lambda=0.05$.

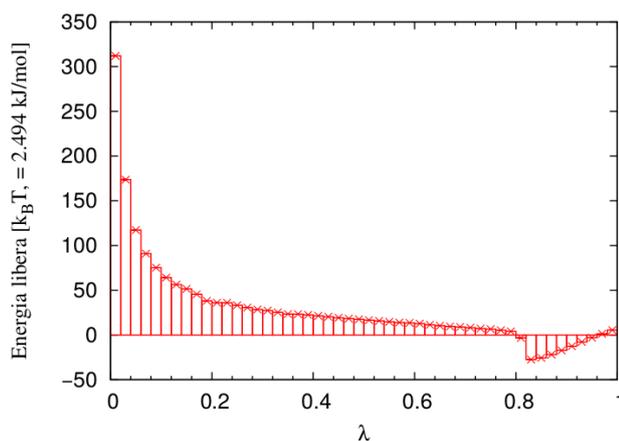


Figura 5.10: $dG/d\lambda$ per la proteina nello stato nativo con $\Delta\lambda=0.02$.

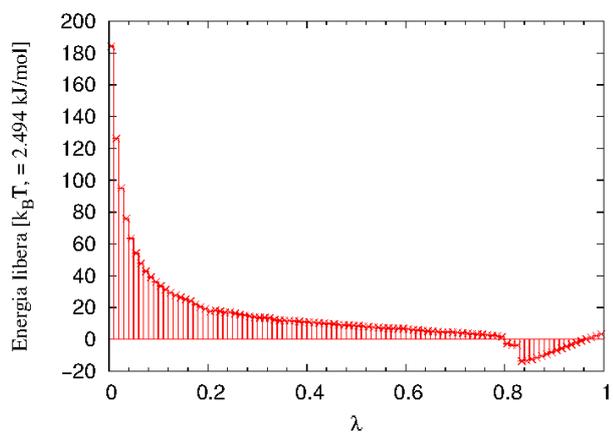
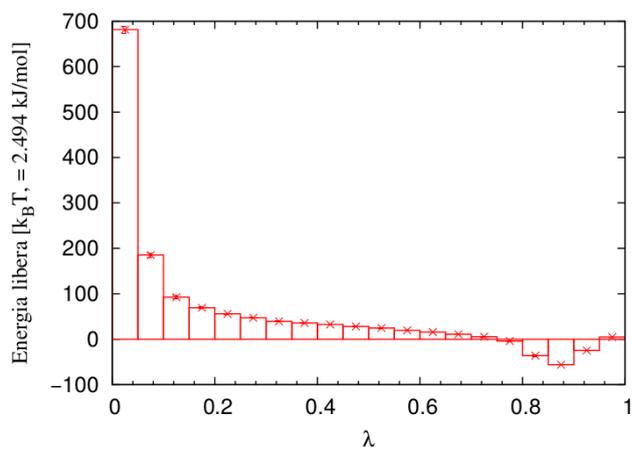
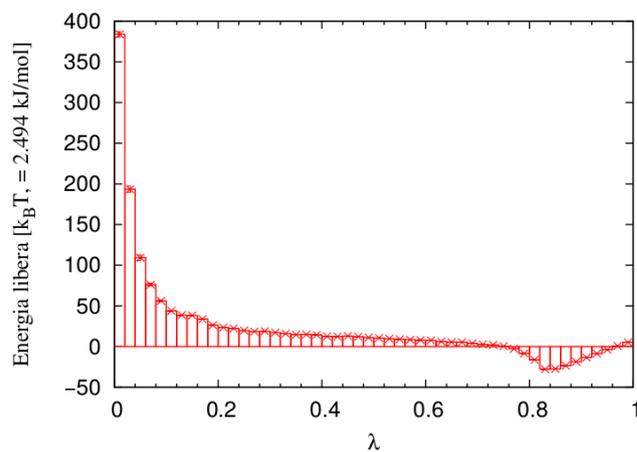
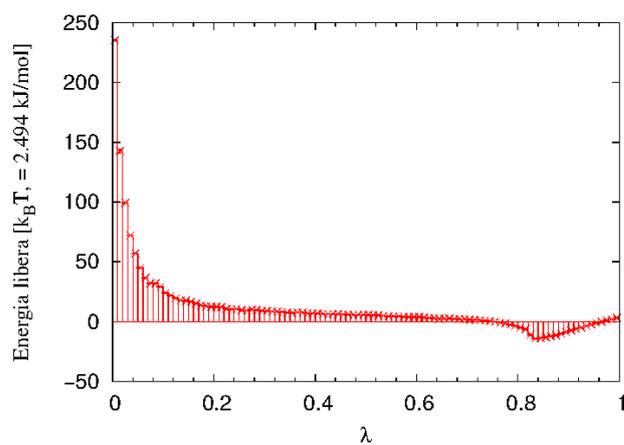


Figura 5.11: $dG/d\lambda$ per la proteina nello stato nativo con $\Delta\lambda=0.01$.

Figura 5.12: $dG/d\lambda$ per la proteina nello stato denaturato con $\Delta\lambda=0.05$.Figura 5.13: $dG/d\lambda$ per la proteina nello stato denaturato con $\Delta\lambda=0.02$.Figura 5.14: $dG/d\lambda$ per la proteina nello stato denaturato con $\Delta\lambda=0.01$.

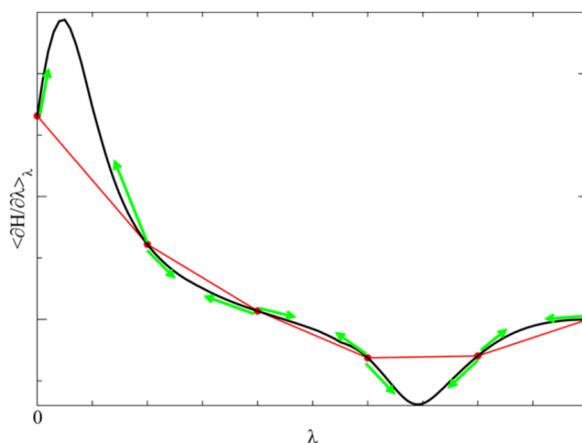


Figura 5.15: I punti rossi rappresentano i valori di λ simulati, da cui si ottiene una curva ideale (in rosso). In nero la curva effettivamente ottenuta. Lungo l'asse y è riportato $\langle \partial H / \partial \lambda \rangle_\lambda$. Immagine tratta da [46].

La presenza di una zona in cui la differenza di energia libera diventa negativa sembra essere una situazione sistematica dell'integrazione termodinamica, come evidenziato in Figura 5.15. L'allontanamento dall'attesa nei punti iniziali e finali può essere attribuito esclusivamente al termine LJ, in quanto dalle nostre simulazioni si vede come il termine Coulombiano segua l'idealità. Nel nostro caso il contributo di Lennard-Jones nei primi punti viene coperto dall'elevata variazione del termine coulombiano, spiegando l'assenza del primo picco presente invece nella Figura 5.15 mentre per λ elevati i due termini sono comparabili e quindi la non idealità del termine LJ si ripercuote sulla variazione di energia libera di Gibbs, evidenziato dalla variazione del segno.

Nelle Figure 5.16 - 5.21 sono riportate invece le sommatorie correnti delle variazioni di energia libera fra 0 e λ durante il processo di disaccoppiamento, definite come

$$\int_0^\lambda \left(\frac{dG}{d\lambda'} \right) d\lambda' \quad (5.13)$$

per vari $\Delta\lambda$ e per gli stati nativo e denaturato della proteina.

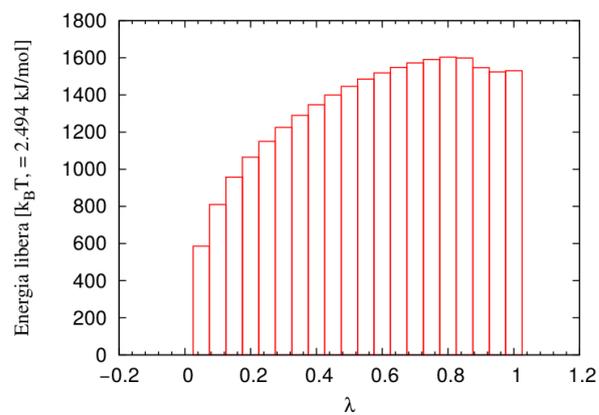


Figura 5.16: $\int_0^\lambda \left(\frac{\delta G}{\delta \lambda'} \right) d\lambda'$ per la proteina nello stato nativo con $\Delta\lambda=0.05$.

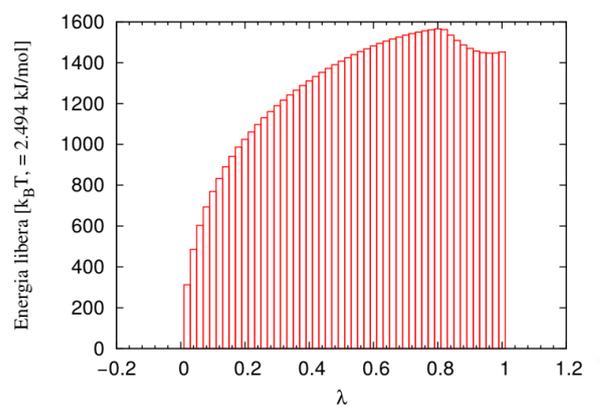


Figura 5.17: $\int_0^\lambda \left(\frac{dG}{d\lambda'} \right) d\lambda'$ per la proteina nello stato nativo con $\Delta\lambda=0.02$.

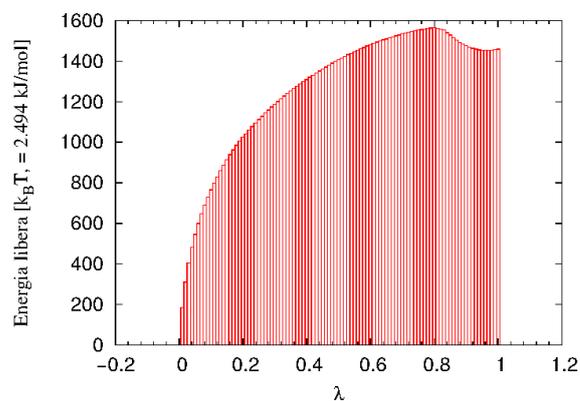


Figura 5.18: $\int_0^\lambda \left(\frac{dG}{d\lambda'} \right) d\lambda'$ per la proteina nello stato nativo con $\Delta\lambda=0.01$.

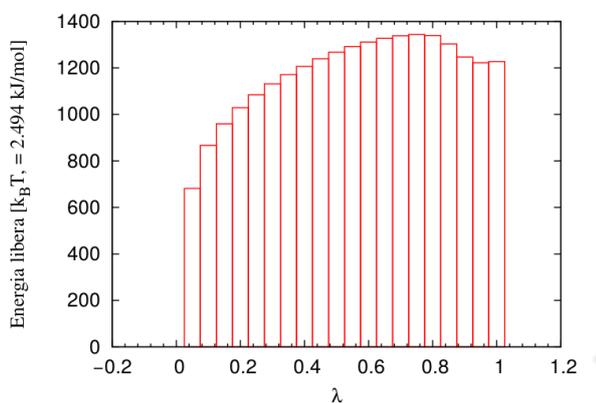


Figura 5.19: $\int_0^\lambda \left(\frac{dG}{d\lambda'} \right) d\lambda'$ per la proteina nello stato denaturato con $\Delta\lambda=0.05$.

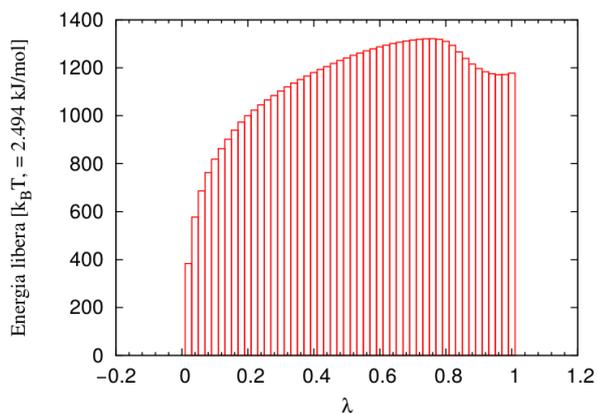


Figura 5.20: $\int_0^\lambda \left(\frac{dG}{d\lambda'} \right) d\lambda'$ per la proteina nello stato denaturato con $\Delta\lambda=0.02$.

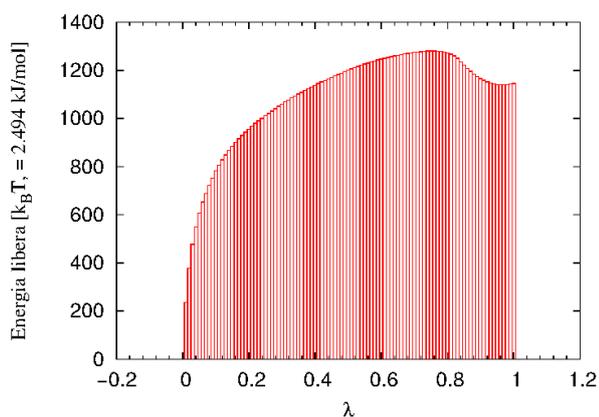


Figura 5.21: $\int_0^\lambda \left(\frac{dG}{d\lambda'} \right) d\lambda'$ per la proteina nello stato denaturato con $\Delta\lambda=0.01$.

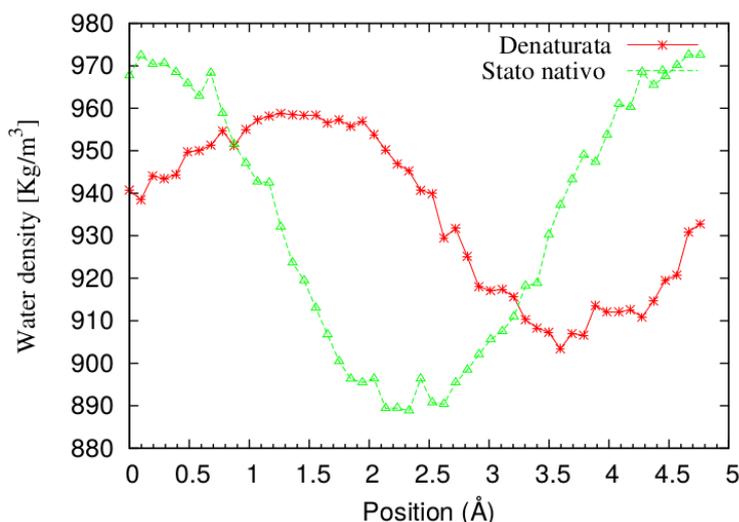


Figura 5.22: La densità dell'acqua nel box in funzione della posizione. In rosso nel caso della proteina denaturata, mentre in verde è riportata la densità dell'acqua in presenza della proteina nello stato nativo.

Abbiamo poi misurato le quantità calorimetriche associate ai processi che compongono il ciclo, i risultati sono riportati in Tabella 5.3. L'energia libera di Gibbs per l'unfolding quando il sistema è disaccoppiato è stata calcolata secondo l'equazione 5.7, ed è risultata estremamente piccola rispetto alle energie di idratazione. Benché piccola, essa è risultata positiva in ogni simulazione, in quanto, contrariamente alle attese, il volume della proteina ripiegata è risultato più grande del volume della proteina denaturata, probabilmente a causa della minor densità dell'acqua all'interno della proteina. Il profilo di densità dell'acqua è mostrata in Figura 5.22, mentre in Figura 5.23 è rappresentata l'ubiquitina in modo da mettere in evidenza il buco formato durante il ripiegamento.

	N(0 → 1)	U(0 → 1)	Unfolding($\lambda = 1$)
$\Delta G_{\Delta\lambda=0.05}$	3816 ± 10	3060 ± 29	$4.7 \cdot 10^{-3}$
$\Delta G_{\Delta\lambda=0.02}$	3624 ± 2	2936 ± 31	$3.5 \cdot 10^{-3}$
$\Delta G_{\Delta\lambda=0.01}$	3641 ± 3	2859 ± 17	$10.0 \cdot 10^{-3}$

Tabella 5.3: Energia libera dei processi di deidratazione per le proteine nello stato nativo N e denaturato U, e di unfolding quando il sistema è disaccoppiato. Tutti i valori sono in kJ mol^{-1} .

Dal ciclo possiamo misurare la variazione di energia libera del processo di unfolding secondo l'equazione 5.2, mentre la variazione di entalpia può essere



Figura 5.23: Ubiquitina. In verde sono rappresentati il backbone e le strutture secondarie, mentre in rosso sono rappresentate la catene laterali.

calcolata confrontando i due stadi finale ed iniziale, ed è quindi non dipendente da $\Delta\lambda$. La variazione di entalpia è stata calcolata mediando diverse simulazioni. I risultati sono mostrati nella Tabella 5.4.

	$\Delta\lambda = 0.01$	$\Delta\lambda = 0.02$	$\Delta\lambda = 0.05$	[44]
$\Delta G_{unfolding}(\lambda = 0)$	780 ± 20	688 ± 32	756 ± 39	37.8
$\langle \Delta H \rangle$	485	485	485	351
$T\Delta S$	-295	-202	-290	313

Tabella 5.4: Confronto fra i valori da noi ottenuti ed i valori sperimentali per $T = 300$ K. La Tutti i valori sono in kJ mol^{-1} .

La differenza di energia libera dovuta all'unfolding non sembra essere direttamente legata al $\Delta\lambda$. Il valore di ΔH da noi misurato è paragonabile al dato sperimentale, ed è in accordo con il valore energetico per il singolo amminoacido riportato in [21], $1.5 \text{ kcal mol}^{-1}$. Esiste una elevata differenza di entalpia fra le singole simulazioni, dovuta al fatto che non esiste una unica struttura denaturata ma un insieme di strutture con legami idrogeno intramolecolari che portano a grandi fluttuazioni entalpiche.

Capitolo 6

Conclusioni e prospettive

In questa tesi, si sono applicati metodi numerici per studiare alcune proprietà del processo di folding di una proteina, l'ubiquitina.

La tesi è divisa in due parti.

La prima parte è caratterizzata dall'analisi comparativa di diversi possibili *force field*, caratteristici della Dinamica Molecolare da noi utilizzata come strumento numerico, e la loro validazione mediante confronto con dati sperimentali NMR. Ciò ha permesso di selezionare il *force field* che meglio descriveva le principali caratteristiche chimico-fisiche delle proteine.

Nella seconda parte, tale *force field* è stato utilizzato per lo studio del numero dei legami idrogeno e dell'energia libera di solvatazione per una particolare proteina, l'ubiquitina, con un'enfasi particolare data alla discriminazione tra effetti entropici ed entalpici. L'interesse per tale problema è sia di natura metodologica generale, sia specifico per la particolare proteina, per la quale un tale discrimine non è ancora stato definito, nonostante esistano un certo numero di studi, sia teorici che sperimentali.

L'analisi dei risultati per i *force field* AMBER99sb e AMBER99sbildn ed il confronto con i risultati sperimentali NMR, suggeriscono il fatto che entrambi riproducano abbastanza bene le principali proprietà, e che le loro differenze si limitino a zone specifiche, come evidenziato in Figura 4.2 e Tabella 4.1. Tale risultato non supporta la conclusione di un netto miglioramento nell'uso del *force field* più moderno, come suggerito nella referenza [39]. Queste differenze potrebbero essere dovute all'uso di diversi set per il confronto dei termini S^2 . Dai risultati ottenuti non siamo riusciti ad individuare, in maniera oggettiva, quale fosse il *force field* migliore. Abbiamo quindi scelto di partire dal modello dell'acqua, prendendo il modello TIP3P in quanto strutturalmente più realistico, condizioni NPT in quanto sono lo standard, ed il *force field* AMBER99sbildn in quanto presentava un miglior accordo con i dati sperimentali per i residui intorno al ventesimo.

Fatto tale passaggio, si è quindi passato alla seconda parte della tesi, che riguarda lo studio dei termini energetici nel processo di *fold*ing. L'analisi dei termini energetici è iniziata con il conteggio dei legami idrogeno intramolecolari e con l'acqua. Da tale analisi, emerge una sostanziale invarianza tra stato nativo e denaturato nel numero dei legami idrogeno formati dalla proteina, come evidenziato nella Tabella 4.2. Ciò si spiega assumendo che sostanzialmente tutti i legami idrogeno originariamente formati con l'acqua nella fase denaturata vengono riformati, in parte con l'acqua ed in parte all'interno della proteina stessa, anche nello stato nativo. Anche se le nostre simulazioni sembrano sottostimare in maniera sistematica il numero di legami idrogeno nella proteina ripiegata, va comunque tenuta in conto la possibilità di una saturazione nei legami idrogeno causata dal ripiegamento e dalla presenza di una zona interna di acqua con minor densità del *bulk*. Successivamente abbiamo considerato tutte le interazioni fra proteina e solvente e la loro variazione durante il disaccoppiamento che simulava il processo di deidratazione e la corrispondente variazione di energia libera ΔG_{solv} . Il processo di transizione dallo stato nativo allo stato denaturato (*un*folding) è stato trasformato in una sequenza di due deidratazioni per la fase nativa e denaturata, più un processo di pura trasformazione entropica, mediante un opportuno ciclo termodinamico, rappresentato in Figura 5.1. Per ognuno dei due processi di deidratazione, è stata eseguita una integrazione termodinamica che forniva la curva $\frac{dG}{d\lambda}$ il cui integrale rispetto a λ compreso fra 0 ed 1 fornisce l'energia libera di solvatazione ΔG_{solv} . La presenza di alcune punti del grafico $dG/d\lambda$ che si allontanano dall'idealità è una caratteristica dell'integrazione termodinamica, si vedano Figure 5.9 - 5.14 e 5.15. Confrontando Figura 5.4 e 5.5 si vede come il termine problematico sia quello di Lennard-Jones. Un miglioramento della griglia di integrazione mediante l'uso di $\Delta\lambda$ più piccoli non sembra migliorare il problema. Simulazioni future dovrebbero quindi concentrarsi su altre possibilità, quali il modo in cui vengono calcolate le interazioni di non legame. Una caratteristica interessante del picco del limite di disaccoppiamento del termine di Lennard-Jones è che non sembra dipendere dalla forma della proteina. Il termine coulombiano invece non mostra punti di particolare interesse e si comporta secondo quanto ci si aspettava. Possiamo concludere quindi che il termine di Coulomb sia meno critico di quello di LJ, e che sia possibile simularlo in maniera accurata impiegando relativamente pochi punti. Al contrario il termine di LJ è il termine critico, su cui bisogna spendere particolare attenzione. In particolare quando λ tende a zero, e i due contributi sono piccoli e comparabili, il comportamento del termine di LJ si riflette sull'energia libera. Le energie libere di deidratazione sono state usate per studiare il processo di folding/unfolding della proteina, secondo lo schema in Figura 5.1. Dal confronto dei termini in Tabella 5.3 si vede come

l'energia libera dovuta all'acqua risulta estremamente piccola e trascurabile nel calcolo dell'energia libera di *unfolding*. Un valore positivo di energia libera di *unfolding* per la proteina disaccoppiata, che dal punto di vista del solvente è semplicemente la variazione nel volume escluso dato dal *fold*ing della proteina, potrebbe essere una caratteristica dell'ubiquitina dovuta alla forma della stessa. La già citata zona interna con una minore densità nella proteina nella struttura nativa probabilmente compensa la diminuzione di volume escluso dovuta esclusivamente agli atomi della proteina. I valori assoluti di energia libera calcolati risultano essere significativamente diversi da quelli sperimentali e l'uso di una entalpia mediata su varie simulazioni, che si avvicina al risultato sperimentale, porta alla stima, sicuramente incorretta, di una variazione di entropia negativa.

I risultati di questa tesi possono costituire il punto di partenza di una serie di miglioramenti sistematici della procedura di calcolo. In prima battuta, in questa tesi ci siamo limitati a confrontare *force field* appartenenti alla stessa famiglia. Uno studio più completo sul parametro d'ordine confrontando *force field* diversi su proteine diverse sarebbe il naturale prosieguo del lavoro di questa tesi. La struttura denaturata risulta in realtà non ben definita. La nostra aspettativa era che differenti strutture denaturate, usate come possibili condizioni iniziali del processo di ripiegamento della proteina, fossero tutte statisticamente equivalenti, e quindi caratterizzate da entalpie molto simili. I nostri risultati indicano invece viceversa una notevole fluttuazione in entalpia tra le varie strutture denaturate, che si ripercuote negativamente sul risultato generato dall'intero ciclo termodinamico. Questo non è un problema derivante dall'uso di simulazioni ma è una caratteristica delle catene polipeptidiche, le quali, anche in assenza di gruppi idrofobici, presentano legami idrogeno intramolecolari che inducono ad un rapido collasso della proteina [47]. Studi futuri sull'ubiquitina, a prescindere da variazioni nella metodologia, necessitano di un numero maggiore di simulazioni partendo da strutture denaturate diverse, in modo da poter ottenere dei dati mediati. Un aumento del numero delle simulazioni potrebbe inoltre determinare se la differenza nel numero di legami idrogeno fra proteina nativa e denaturata sia effettivamente da imputare ad una semplice fluttuazione. Nello specifico delle simulazioni di deidratazione, il termine di LJ è quello che necessita di un numero maggiore di punti. Data la possibilità di separare i due disaccoppiamenti, e data la robustezza del termine di Coulomb, il termine di Lennard-Jones può essere studiato sia aumentando il numero di punti λ sia cambiando le condizioni della simulazione. In particolare, un termine che merita attenzione è la distanza di *cut-off* oltre cui le interazioni di non legame vengono trascurate. Questo termine potrebbe essere aumentato per meglio descrivere quello che succede, con la consapevolezza questo potrebbe favorire un relativo aumento

del rapporto rumore/segnale.

Un punto dolente del nostro sistema per lo studio dell'apertura della proteina riguarda l'entropia di unfolding. Anche ignorando il fatto di aver ottenuto un valore negativo, l'errore risultante è evidentemente troppo grande. Un possibile sviluppo è l'uso di tecniche diverse, quali la *Grid Inhomogeneous Solvation Theory* (GIST) [2]. Queste tecniche calcolano l'entalpia a partire dalla distribuzione g_{sw} , funzione della posizione \mathbf{r} e dell'orientazione ω , e non si limitano a fare una stima a partire da energia libera ed entalpia. Una ulteriore possibilità è lo studio di diverse simulazioni a temperature diverse e il calcolo dell'entropia a partire dalla dipendenza dell'energia libera di Gibbs dalla temperatura. I metodi per stimare l'entropia a partire da simulazioni di dinamica molecolare sono analizzati in [48].

Una soluzione più drastica sarebbe quella di passare ad una proteina più piccola. Il soggetto di questo studio è la proteina ubiquitina umana, una proteina che per quanto si possa considerare piccola relativamente alla media delle proteine, resta comunque una molecola di oltre 1200 atomi. Lo studio della solvatazione per sistemi circa 200 volte più piccoli è risultato estremamente accurato, sarebbe quindi d'interesse uno studio sistematico a dimensioni crescenti. L'analisi e la quantificazione dei singoli termini entropici ed entalpici che appaiono nell'energia libera di *unfolding*, è un problema di grande interesse attuale [2] perché aprirebbe la strada ad una più corretta interpretazione del processo di *foldig*.

Una volta ottenuto un sistema funzionante ed accurato per lo studio dei vari termini di unfolding, lo sviluppo successivo sarebbe quello di variare il solvente, e le condizioni termodinamiche. È stato riportato [49] che un solvente meno polare dell'acqua induce la proteina a preferire la formazione di α -eliche rispetto all'acqua. Studi in solventi diversi, quali ad esempio gli idrocarburi, sarebbero utili nell'ottica dello studio della vita extra terrestre e più in generale in solventi diversi dall'acqua.

Appendice A

Integrazione termodinamica

L'integrazione termodinamica è basata sull'assunto che ci sia una variazione lineare di \mathcal{V} rispetto ad un parametro λ tale che

$$\mathcal{V}(\lambda) = \mathcal{V}_A + \lambda(\mathcal{V}_B - \mathcal{V}_A) \quad (\text{A.1})$$

Possiamo quindi calcolare come varia l'energia libera fra due stati, sfruttando il fatto che, essendo una funzione di stato e non di processo, la differenza fra le energie libere non dipende dal percorso ma solo dagli stati iniziale e finale. L'uso di processi non fisicamente possibili, quali lo spegnimento delle interazioni di non legame, è chiamato *alchemical simulation*. Nel caso di una simulazione in cui si spengono le cariche, esiste il rischio che due atomi si sovrappongano. Per evitare questa situazione, si usa un potenziale di soft-core. La funzione 5.3 può essere derivata nel modo seguente. La funzione di partizione, che descrive la costante di normalizzazione usata per il calcolo della probabilità di trovare un sistema in un determinato stato ha la forma

$$Q = \frac{1}{h^{3N} N!} \int d\mathbf{p} d\mathbf{r} \exp \left[-\beta \left(\sum_{i=1}^N \frac{p_i^2}{2m_i} + \mathcal{V}(\mathbf{r}^N) \right) \right] \quad (\text{A.2})$$

Dove \mathbf{p} è il momento di moto e \mathcal{V} il potenziale. Le componenti cinetiche e potenziali sono indipendenti, in quanto l'energia cinetica dipende dai momenti e quella potenziale dalla posizione. La componente cinetica può essere integrata, dando come risultato $(\frac{2\pi m}{\beta})^{\frac{3}{2}N}$. Definendo Z la componente configurazionale

$$Z = \int d\mathbf{r} \exp \left[-\beta \mathcal{V}(\mathbf{r}^N) \right] \quad (\text{A.3})$$

Q diventa

$$Q(N, P, T) = \frac{Z(N, P, T)}{\Lambda(T)^{3N} N!} \quad (\text{A.4})$$

dove $\Lambda(T)$ è la lunghezza d'onda termica di De Broglie. La funzione di Helmholtz, definita come $F = U - TS$, che caratterizza un sistema con numero di componenti, volume e temperatura costante, è legato alla funzione di partizione Q essendone il logaritmo:

$$G = -k_B T \log Q_{NPT} \quad (\text{A.5})$$

Possiamo calcolare la variazione di energia libera rispetto ad un parametro di coupling λ

$$\begin{aligned} \left(\frac{dG(\lambda)}{d\lambda} \right)_{NPT} &= -k_B T \frac{\log Q_{N,P,T}(\lambda)}{d\lambda} & (\text{A.6}) \\ &= -\frac{k_B T}{Q_{N,P,T}(\lambda)} \frac{Q_{N,P,T}(\lambda)}{d\lambda} \\ &= \frac{\int d\mathbf{r}^N \left(\frac{d\mathcal{V}(\lambda)}{d\lambda} \right)_\lambda \exp[-\beta\mathcal{V}(\lambda)]}{\int d\mathbf{r}^N \exp[-\beta\mathcal{V}(\lambda)]} \\ &= \left\langle \frac{d\mathcal{V}(\lambda)}{d\lambda} \right\rangle_\lambda & (\text{A.7}) \end{aligned}$$

Integrando l'equazione A.6 è possibile trovare la differenza di energia libera fra lo stato A e B.

$$\Delta G_{AB} = G(\lambda = 1) - G(\lambda = 0) = \int_{\lambda=0}^{\lambda=1} d\lambda \left\langle \frac{d\mathcal{V}(\lambda)}{d\lambda} \right\rangle_\lambda \quad (\text{A.8})$$

Appendice B

Tabella di conversione delle unità di misura per l'energia

Energia

	kcal/mol	kJ/mol	eV
kcal/mol	1	4.184	$2.611E^{22}$
kJ/mol	0.239	1	$6.242E^{21}$
eV	$3.829E^{-23}$	$1.602E^{-21}$	1

Tabella B.1: Tabella di conversione delle unità di misura dell'energia.

Appendice C

Parametro di ordine nematico

La funzione di correlazione $C(t)$ ha la forma

$$C(t) = \frac{1}{5} \langle P_2(\boldsymbol{\mu}_L(0) \cdot \boldsymbol{\mu}_{LF}(t)) \rangle \quad (\text{C.1})$$

dove P_2 è il secondo polinomio di Legendre

$$P_2(x) = \frac{1}{2}(3x^2 - 1) \quad (\text{C.2})$$

P_0 e P_1 invece valgono 1 e X , rispettivamente. Il secondo polinomio di Legendre viene usato per rappresentare l'ordinamento di una serie di vettori. Il prodotto scalare fra due vettori vale

$$\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}} = xy \cos(\theta) \quad (\text{C.3})$$

dove θ è l'angolo fra i vettori. la media del secondo polinomio del prodotto fra vettori quindi vale

$$S = \langle P_2(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}) \rangle = \langle P_2(\cos \theta) \rangle \quad (\text{C.4})$$

che, dato che i vettori usati sono versori, diventa

$$S = \langle P_2(\cos \theta) \rangle = \frac{1}{2} \langle 3 \cos^2 \theta - 1 \rangle = \frac{1}{2} (3 \langle \cos^2 \theta \rangle - 1) \quad (\text{C.5})$$

dove $\langle \rangle$ rappresenta la media su tutte le configurazioni. S è compreso fra un massimo di 1, indice che tutti i vettori sono orientati nella stessa direzione, $\theta = 0$ per tutti gli angoli, dando $3 \langle \cos^2 \theta \rangle = 3$, ed un minimo di 0. Il minimo si ha quando si ha disordine completo, in questo caso la media del quadrato dei coseni vale $\frac{1}{3}$ e $3 \langle \cos^2 \theta \rangle - 1 = 0$. La media di P_2 si può calcolare come

$$\langle P_2 \rangle = \int P(\Omega) P_2(\cos \theta) d\Omega \quad (\text{C.6})$$

dove Ω è l'angolo solido e $P(\Omega)$ la probabilità di trovare il vettore nell'angolo Ω . Se siamo in una situazione completamente casuale, $P(\Omega)$ è una costante e vale $\frac{1}{4\pi}$. In coordinate sferiche $d\Omega = d\phi \sin \theta d\theta$ dove ϕ e θ sono rispettivamente l'angolo fra l'asse x e la proiezione di \mathbf{r} nel piano xy, e l'angolo fra l'asse z e \mathbf{r} . L'integrale diventa quindi

$$\langle P_2 \rangle = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_0^\pi P_2(\cos \theta) \sin \theta d\theta \quad (\text{C.7})$$

Dato che $\sin \theta d\theta = -d(\cos \theta)$ si ottiene

$$\langle P_2 \rangle = \frac{1}{4\pi} \int_0^{2\pi} d\phi \int_{-1}^1 P_2(\cos \theta) d \cos \theta \quad (\text{C.8})$$

dove l'integrazione è ora fra -1 ed 1. $\int_0^{2\pi} d\phi = 2\pi$ e si ottiene quindi

$$\langle P_2 \rangle = \frac{1}{2} \int_{-1}^1 P_2(\cos \theta) d \cos \theta \quad (\text{C.9})$$

Separando l'integrale si ottiene

$$\langle P_2 \rangle = \frac{1}{2} \left(3 \int_{-1}^1 \cos^2 \theta d(\cos \theta) - \int_{-1}^1 1 d(\cos \theta) \right) \quad (\text{C.10})$$

L'integrale $\int_{-1}^1 X^2 dX$ vale $\frac{2}{3}$, mentre $\int_{-1}^1 1 d(\cos \theta) = 2$. $\langle P_2 \rangle$ quando l'orientamento è completamente casuale diventa quindi

$$\langle P_2 \rangle = \frac{1}{2} \left[3 \frac{2}{3} - 2 \right] = 0 \quad (\text{C.11})$$

Quando i vettori sono orientati nella stessa direzione, la probabilità $P(\Omega)$ diventa

$$P(\Omega) = \frac{1}{2\pi} \delta(\cos \theta - 1) \quad (\text{C.12})$$

dove $\delta(\cos \theta - 1)$ è la delta di Dirac, e l'integrale diventa

$$\langle P_2 \rangle = \int \frac{1}{2\pi} \delta(\cos \theta - 1) P_2(\cos \theta) d\Omega \quad (\text{C.13})$$

che, con la stessa trattazione di prima, equivale a

$$\langle P_2 \rangle = \int_{-1}^1 \delta(\cos \theta - 1) P_2(\cos \theta) d \cos \theta \quad (\text{C.14})$$

La delta di Dirac ha la proprietà

$$\int_a^b \delta(x - c) f(x) dx = f(c) \quad (\text{C.15})$$

quando x è compreso nell'intervallo d'integrazione. Si ottiene quindi

$$\langle P_2 \rangle = P_2(1) = \frac{1}{2} (3 - 1) = 1 \quad (\text{C.16})$$

Bibliografia

- [1] S. Piana, K. Lindorff-Larsen, D. E. Shaw, Atomic-level description of ubiquitin folding, *Proceedings of the National Academy of Sciences of the United States of America*, vol 110 no 15:5915, (2013).
- [2] M. Schauperl, M. Podewitz, B. J. Waldner, K. R. Liedl, Enthalpic and Entropic Contributions to Hydrophobicity, *J. Chem Theory, Comput.*, 12:4600, (2016).
- [3] PDB ID: 1UBQ, Vijay-Kumar, S., Bugg, C.E., Cook, W.J., Structure of ubiquitin refined at 1.8 Å resolution. *J.Mol.Biol.* 194:531 (1987).
- [4] M. Senske, L. Törk, B. Born, M. Havenith, C. Hermann, S. Ebbinghaus, Protein Stabilization by Macromolecular Crowding through Enthalpy Rather Than Entropy, *J. Am. Chem. Soc.* 136:9036, (2014).
- [5] www.nobelprize.org
- [6] J. A. Diehl, S. Y. Fuchs, D. S. Haines, Ubiquitin and Cancer: New Discussions for a New Journal, *Genes & Cancer*, vol 1 no 7:679, (2010).
- [7] J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry*, printed in the United States of America, W. H. Freeman, 7th edition.
- [8] <http://www.ks.uiuc.edu/Training/CaseStudies/pdfs/ubq.pdf>
- [9] D. L. Nelson, M. M. Cox, *Principles of biochemistry*, printed in the United States of America, 2005, W. H. Freeman, 4th edition.
- [10] C. Poletto, *Solvent Induced Interactions in Biopolymers: Origin of Secondary Motifs* (tesi di dottorato, Università degli studi di Padova), (2009).
- [11] www.chemwiki.ucdavis.edu/Core/Biological_Chemistry/Proteins/Protein-Structure

- [12] W. J. Wedmeyer, E. Welker, H. A. Scheraga, Proline Cis-Trans Isomerization and Protein Folding, *Biochemistry*, 41:14637, (2002).
- [13] P. Kountoris, J. D. Hyrst, Predicting b-turns and their types using predicted backbone dihedral angles and secondary structures *BMC Bioinformatics*, 11:407, (2010).
- [14] J. Milner-White, B. M. Ross, et al., One type of γ -turn, rather than the other, gives rise to chain reversal in proteins, *J. Mol. Biol.*, 204:777, (1988).
- [15] M. Pal, S. Dasgupta, The Nature of the Turn in Omega Loops of Proteins, *Proteins: Structure, Functions, and Genetics*, 51:591, (2003).
- [16] C. B. Anfinsen, Principles that govern the folding of protein chains., *Science*, 181:223, (1973).
- [17] <http://sites.fas.harvard.edu/lsci1a/10-12notes.pdf>
- [18] N. T. Southall, K. A. Dill, A. D.J. Haymet, A View of the Hydrophobic Effect, *J. Phys. Chem.*, 106:521, (2002).
- [19] N. V. Prabhu, K. A. Sharp, Heat capacity in proteins, *Annu. Rev. Phys. Chem.*, 56:521, (2005).
- [20] C. Levinthal, How to Fold Graciously, *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois: 22–24*, (1969).
- [21] J. M. Scholtz, S. Marqusee, R. L. Baldwin, E. J. York, J. M. Stewart, M. Santoro, D. Wayne. Bolen, Calorimetric determination of the enthalpy change for the α -helix to coil transformation of an alanine peptide in water, *Proc. Natl. Acad. Sci. USA*, 88:2854, (1990).
- [22] C. R. Cantor, P. R. Schimmel, *Biophysical Chemistry Part II*, Printed in the United States of America, W. H. Freeman, (1980).
- [23] M. Karplus, Vicinal Proton Coupling in Nuclear Magnetic Resonance, *J. Am. Chem. Soc.*, 85:2870, (1963)
- [24] P. Atkins, J. de Paula, *Atkins' Physical Chemistry*, New York, NY, W. H. Freeman, 8th edition, (2006).
- [25] A. Bax, Two-Dimensional NMR and Protein Structure, *Annu. Rev. Biochem.*, 58:223, (1989).

- [26] http://triton.iqfr.csic.es/guide/tutorials/specdata/spectra/str_noesy.html
- [27] <http://www.colby.edu/chemistry/PChem/lab/DiffScanningCal.pdf>
- [28] E. Freire, O. L. Mayorga, M. Straume, Isothermal Titration, *Analytical Chemistry*, vol 62 no 18:950, (1990).
- [29] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water , *J. Chem. Phys.* 79:926, (1983).
- [30] AMBER99 protein, nucleic AMBER94 (Wang et al., *J. Comp. Chem.* 21, 1049-1074, 2000).
- [31] AMBER99SB protein, nucleic AMBER94 (Hornak et al., *Proteins* 65, 712-725, 2006).
- [32] AMBER99SB-ILDN protein, nucleic AMBER94 (Lindorff-Larsen et al., *Proteins* 78, 1950-58, 2010).
- [33] MacKerell, Jr., A. D., Feig, M., Brooks, C.L., III, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, *Journal of Computational Chemistry*, 25: 1400-1415, 2004.
- [34] MacKerell, Jr., A. D., et al. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *Journal of Physical Chemistry B*, 1998, 102, 3586-3616.
- [35] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. (1996) *J. Am. Chem. Soc.* 118 (45): 11225-11236.
- [36] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen (2001) *J. Phys. Chem. B* 105 (28): 6474-6487 .
- [37] G. Lipari, A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1.Theory and Range of Validity, *J. Am. Chem. Soc.*, 104, 4546, (1982).
- [38] N. Tjandra, A. Szabo, A. Bax, Protein Backbone Dynamics and ^{15}N Chemical Shift Anisotropy from Quantitative Measurement of Relaxation Interference Effects, *J. Am. Chem. Soc.*, 118, 6986, (1996).

- [39] S. A. Showalter, R. Brüschweiler, *J. Chem. Theory Comput.*, 3, 961, (2007).
- [40] D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS User Manual version 4.6.5*, www.gromacs.org, (2013).
- [41] D. F. Stickle, L. G. Presta, K. A. Dill, G. D. Rose, Hydrogen Bonding in Globular Proteins, *Journal of molecular biology*, vol 226 no 4:1143, (1992).
- [42] M. R. Shirts, J. W. Pitner, W. C. Swope, V. S. Pande, Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins, *J. Chem. Phys.* 119, 5740, (2003).
- [43] G. I. Makhatadze, M. M. Lopez, J. M. Richardson III, S. T. Thomas, Anion binding to the ubiquitin molecule, *Protein Science*, 7:689, (1998).
- [44] P. L. Wintrode, G. I. Makhatadze, P. L. Privalov, Thermodynamics of Ubiquitin Unfolding, *Proteins: structure, functions, and Genetics* 18:246 (1994).
- [45] B. Ibarra-Molero, V. V. Loladze, G. I. Makhatadze, J. M. Sanchez-Ruiz, Thermal versus Guanidine-Induced Unfolding of Ubiquitin. An Analysis in Terms of the Contributions from Charge-Charge Interactions to Protein Stability, *Biochemistry*, 38:8138, (1999).
- [46] A. De Ruiter, C. Oostenbrink, Extended Thermodynamic Integration: Efficient Prediction of Lambda Derivatives at Nonsimulated Points, *J. Chem Theory, Comput.*, 12:4476, (2016).
- [47] A. Möglich, K. Joder, T. Kiefhaber, End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation, *Proc. Natl. Acad. Sci. USA*, vol 103 no 33:12394, (2006).
- [48] C. Peter, C. Oostenbrink, A. van Dorp, W. F. van Gunsteren, Estimating entropies from molecular dynamics simulations, *The Journal of Chemical Physics* 120:2652, (2004).
- [49] S. Yasuda, H. Oshima, M. Kinoshita, Structural Stability of Proteins in Aqueous and Nonpolar Environments *J. Chem. Phys.* 137, 135103, (2012).