

POLITECNICO DI TORINO

Collegio di Ingegneria Chimica e dei Materiali

Corso di Laurea Magistrale
in Ingegneria Chimica e dei Processi Sostenibili

Tesi di Alessio Lombardo Pontillo

Modellazione molecolare dell'assorbimento di una proteina del tuorlo d'uovo su un'interfaccia acqua-olio



**Politecnico
di Torino**

Relatori

Prof. Daniele Marchisio
Prof. Antonio Buffo
Prof. Gianluca Boccardo
Prof. Francesca Bosco

Candidato

Alessio Lombardo Pontillo

Marzo, 2022

Indice

Introduzione	1
1 Introduzione	1
1.1 Le emulsioni	2
1.2 Il tuorlo d'uovo	3
2 Dinamica Molecolare e Metadinamica	7
2.1 La meccanica statistica	7
2.1.1 Interazione tra molecole	9
2.1.2 Dallo stato molecolare allo stato termodinamico	10
2.2 La dinamica molecolare	11
2.2.1 Il campo di forze o force field	11
2.2.2 Simulazione di dinamica molecolare	14
2.3 La Metadinamica	15
2.3.1 Sampling nello spazio delle variabili collettive	15
2.3.2 Well-Tempered Metadynamics	17
2.3.3 Variabili Collettive	18
2.3.4 Varianti della metadinamica	19
2.3.5 Confronto metadinamica e dinamica molecolare	20
3 Dettagli computazionali	22
3.1 Le condizioni periodiche e il cut-off	22
3.2 Preparazione del sistema	24
3.2.1 Neutralizzazione della carica elettrica	24
3.2.2 Minimizzazione energetica	25
3.2.3 Equilibratura	25
3.3 Ipotesi semplificative	26
3.3.1 Fase non-acquosa: olio alimentare	27
3.3.2 Fase acquosa	27
3.3.3 La proteina	28
3.4 Parametri della dinamica molecolare	31
3.5 Parametri della Metadinamica	31
3.6 Dettagli delle simulazioni	33

3.6.1	Caso 1	33
3.6.2	Caso 2	34
3.6.3	Caso 3	34
3.6.4	Caso 4	34
3.6.5	Caso 5	35
3.6.6	Caso 6	36
3.7	Tabella riassuntiva	36
3.8	Post-processing	37
3.8.1	Il raggio di giro	37
3.8.2	RMSD	37
3.8.3	Distanza testa-coda	38
3.8.4	MSD e diffusività	38
3.8.5	Area superficiale accessibile al solvente	38
4	Analisi dei risultati	39
4.1	Caso 1	39
4.2	Caso 2	44
4.3	Caso 3	48
4.4	Caso 4	55
4.5	Caso 5	57
4.6	Caso 6	60
4.7	Confronto: Casi 1 e 2	62
4.8	Confronto: Casi 3 e 4	62
4.9	Confronto: Casi 5 e 6	63
5	Conclusioni	64
	Bibliografia	66

Elenco delle figure

1.1	Schematizzazione di una emulsione con i principali elementi che la caratterizzano (elaborazione immagine estratta da https://sacrachimicablog.wordpress.com/2017/03/23/emulsioni/).	2
1.2	Rappresentazione dell'assorbimento di un granulo all'interfaccia acqua-olio (elaborazione immagine estratta da Ferrari et al. (2022)).	4
2.1	Tempi caratteristici di alcuni eventi tipicamente studiati con la dinamica molecolare (elaborazione immagine estratta da (Lindahl 2008)).	12
2.2	Effetto del bias sulla probabilità. Dopo l'applicazione del potenziale (grafico rosso) il picco A risulta più basso e per il sistema sarà più semplice raggiungere la configurazione B (elaborazione immagine estratta da Ref. Valsson, Tiwary e Michele Parrinello (2016)).	17
2.3	(a) FES nello spazio delle CV $\mathbf{S} = (S_1, S_2)$ con lo stato iniziale A, lo stato finale B e il percorso di riferimento \mathbf{S}_0 . (b) il valore di s che esprime il completamento del percorso. (c) il valore di z che esprime quanto un percorso generico si discosta dal percorso di riferimento S_0 (elaborazione immagine estratta da Ref. Valsson, Tiwary e Michele Parrinello (2016)).	19
3.1	Rappresentazione delle PBC. La box centrale rappresenta il sistema che si vuole simulare, le box laterali rappresentano delle ideali copie del sistema.	23

3.2	Utilizzo delle PME: nel grafico superiore si nota la differenza tra il potenziale reale (linea continua) e il potenziale calcolato tenendo conto solo della distanza di cut-off (linea tratteggiata). Nel grafico inferiore si notano i due contributi del potenziale calcolati con le PME: il potenziale a lungo raggio (linea tratto punto) e il potenziale a corto raggio (linea tratteggiata), confrontati con il potenziale reale (linea continua) (elaborazione immagine estratta da Lindahl (2008)).	24
3.3	Struttura della proteina estratta dal software AlphaFold: il colore indica l'affidabilità di predizione.	30
3.4	Affidabilità di predizione per ogni residuo.	30
3.5	Distribuzione normalizzata rispetto al numero totale di atomi dell'olio lungo l'asse Y della box.	32
4.1	Caso 1: raggio di giro.	40
4.2	Caso 1: RMSD.	40
4.3	Caso 1: SASA.	41
4.4	Caso 1: distanza EtE.	41
4.5	Caso 1: MSD della traiettoria di 100 ns (linea blu continua) e regressione lineare (linea rossa tratteggiata).	42
4.6	Caso 1: singoli MSD delle dieci simulazioni di 10 ns.	42
4.7	Caso 1: MSD medio delle ultime dieci simulazioni (linea blu continua) e regressione lineare (linea rossa tratteggiata).	43
4.8	Caso 2: raggio di giro.	44
4.9	Caso 2: RMSD.	45
4.10	Caso 2: SASA.	45
4.11	Caso 2: distanza EtE.	46
4.12	Caso 2: MSD della traiettoria di 100 ns (linea blu continua) e regressione lineare (linea rossa tratteggiata).	46
4.13	Caso 2: singoli MSD delle dieci simulazioni di 10 ns.	47
4.14	Caso 2: MSD medio delle ultime dieci simulazioni (linea blu continua) e regressione lineare (linea rossa tratteggiata).	47
4.15	Caso 3: raggio di giro.	49
4.16	Caso 3: RMSD.	49
4.17	Caso 3: SASA.	50
4.18	Caso 3: distanza EtE.	50
4.19	Caso 3: distanza minima tra la fase oleosa e la proteina.	51
4.20	Caso 3: profilo di energia al variare della posizione della proteina.	51
4.21	Caso 3: probabilità al variare della posizione della proteina.	52
4.22	Caso 3: profilo di energia al variare del raggio di giro della proteina.	52

4.23	Caso 3: probabilità al variare del raggio di giro della proteina (per questioni di visualizzazione, l'asse della probabilità è stato tagliato e non raggiunge l'unità).	53
4.24	Caso 3: probabilità e posizione dei singoli amminoacidi, con il punto rosso sono stati evidenziati i punti di assorbimento.	54
4.25	Caso 3: probabilità e posizione degli amminoacidi divisi per tipo, con il punto rosso sono stati evidenziati i punti di assorbimento.	54
4.26	Caso 4: profilo di energia al variare della posizione della proteina.	55
4.27	Caso 4: probabilità al variare della posizione della proteina.	56
4.28	Caso 4: profilo di energia al variare del raggio di giro della proteina.	56
4.29	Caso 4: probabilità al variare del raggio di giro della proteina.	57
4.30	Caso 5: profilo di energia al variare della posizione della proteina.	58
4.31	Caso 5: probabilità al variare della posizione della proteina.	58
4.32	Caso 5: profilo di energia al variare del raggio di giro della proteina.	59
4.33	Caso 5: probabilità al variare del raggio di giro della proteina.	59
4.34	Caso 6: profilo di energia al variare della posizione della proteina.	60
4.35	Caso 6: probabilità al variare della posizione della proteina.	61
4.36	Caso 6: profilo di energia al variare del raggio di giro della proteina.	61
4.37	Caso 6: probabilità al variare del raggio di giro della proteina.	62

Elenco delle tabelle

1.1	Composizione del tuorlo d'uovo.	3
3.1	Composizione dell'olio di semi di soia. Tipo = numero di carboni:numero di insaturazioni, m_i = massa di una molecola, χ_{mass} = frazione massica sperimentale (Pryde 1980), χ_{num} = frazione numerica di molecole calcolata a partire dalla frazione massica.	27
3.2	Carica elettrica degli amminoacidi di un singolo monomero. Ciascun residuo è individuato da: sigla del nome-posizione nella catena.	29
3.3	Parametri Plumed delle simulazioni di metadinamica.	32
3.4	Descrizione delle simulazioni.	33
3.5	Contenuto numerico di molecole di ogni caso simulato.	36
3.6	Composizione in percentuale massica [%] dell'olio in ogni caso simulato.	36
4.1	Indice delle analisi dei risultati.	39
4.2	Caso 1: riassunto dei risultati.	43
4.3	Diffusività in acqua.	44
4.4	Coefficiente di autodiffusione e viscosità dell'acqua, simulato e sperimentale.	44
4.5	Caso 2: riassunto dei risultati.	48
4.6	Diffusività in olio.	48

Capitolo 1

Introduzione

In campo alimentare è frequente l'utilizzo di emulsioni per unire in un unico composto due fasi che normalmente sarebbero immiscibili. Tra i più classici esempi si hanno la panna montata, il burro, la vinaigrette e la maionese. L'emulsione si forma quando, a seguito di un'agitazione vigorosa, una fase si riduce in goccioline che si disperdono nella seconda fase che sarà continua. Le emulsioni sono instabili ma possono mantenere la loro struttura per un certo periodo di tempo a seconda di determinate condizioni. La panna montata, per esempio, tenderà a tornare liquida, liberando l'aria che aveva incorporato con l'agitazione, dopo un certo periodo di tempo e questo può essere influenzato dalla temperatura. Conservando la panna montata a temperature intorno ai 4 °C l'emulsione rimarrà stabile per più tempo rispetto che a conservarla a temperatura ambiente. In alcuni casi però non sono solo le condizioni ambientali a determinare la stabilità di un'emulsione ma anche la presenza di fattori emulsionanti o tensioattivi. Nel caso della maionese, in cui si ha un'emulsione di una fase oleosa in una fase acquosa, la stabilità è fortemente influenzata dalla presenza di proteine e fosfolipidi che consentono di mantenere la consistenza tipica della maionese. Questo lavoro di tesi si propone di studiare il comportamento dell'Apovitellenina-1, una delle principali proteine del tuorlo d'uovo che presenta proprietà emulsionanti. Inoltre, sono stati studiati la cooperazione e la competizione di questa proteina con il fosfolipide maggiormente presente nel tuorlo d'uovo. Lo strumento adottato per questo studio è la dinamica molecolare che consente di studiare le interazioni a livello atomistico e lo sviluppo nel tempo di un sistema. Poiché la dinamica molecolare richiede elevati costi computazionali è stato necessario l'utilizzo di risorse di calcolo adeguate (high performance computing, HPC) per poter ottenere i risultati delle simulazioni in tempi brevi.

1.1 Le emulsioni

Con il termine emulsione si intende la miscela di due sostanze normalmente immiscibili. Questo è possibile perché, anche se le due fasi sono immiscibili l'una nell'altra, grazie all'agitazione si crea una dispersione di goccioline di una fase, che sarà chiamata fase dispersa o interna, all'interno dell'altra fase, che sarà chiamata fase continua o esterna. Si possono avere emulsioni A/O con dispersione di acqua in olio, come nel caso del burro, oppure emulsioni O/A con dispersione di olio in acqua, come nel caso della maionese. Questa

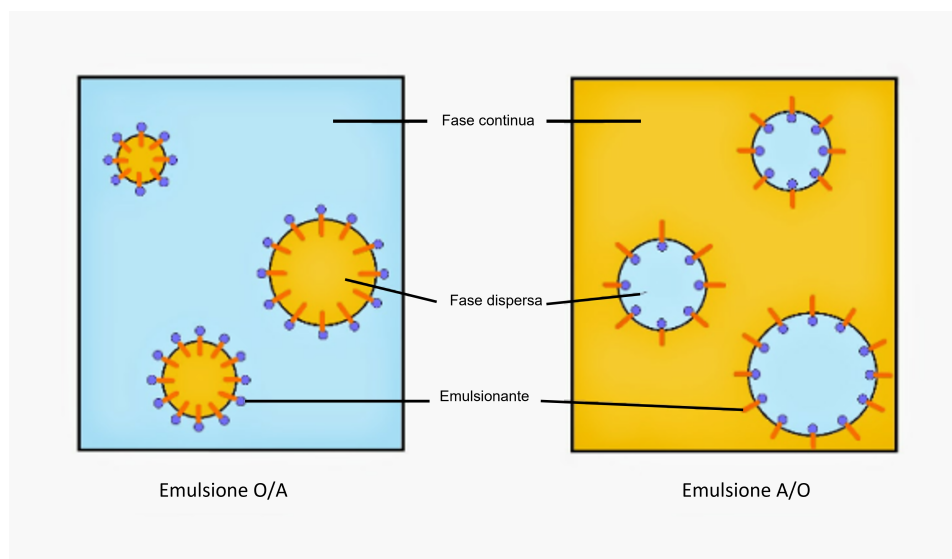


Figura 1.1: Schematizzazione di una emulsione con i principali elementi che la caratterizzano (elaborazione immagine estratta da <https://sacrachimicablog.wordpress.com/2017/03/23/emulsioni/>).

non è comunque una divisione rigorosa perché esistono emulsioni di altri tipi di cui l'esempio più famoso è la panna montata dove si emulsionano bolle di aria all'interno di una fase liquida che è a sua volta un'emulsione, ovvero la panna. Per avere una emulsione non è però sufficiente agitare e miscelare le due fasi, acquosa e oleosa ma è anche necessario che la miscela sia stabile per un certo periodo di tempo che può deve essere più o meno lungo a seconda dell'utilizzo che se ne vuole fare. La vinaigrette, cioè la famosa miscela di aceto, olio e sale tipica delle insalate, ha un tempo di stabilità di pochi minuti e perciò viene preparata solo poco prima di essere servita. Il latte ha invece un tempo di stabilità molto più lungo, anche di diversi giorni. Passato un certo periodo di tempo le gocce della fase dispersa tendono a coalescere fino ad ottenere di nuovo due fasi distinte e separate per densità. Le emulsioni si classificano in base alla dimensione delle gocce della fase dispersa, in tre gruppi:

- Macroemulsioni: dimensioni delle gocce di $0,2\div 50\ \mu\text{m}$, termodinamicamente instabili.
- Miniemulsioni: dimensione delle gocce di $0,1\div 0,4\ \mu\text{m}$, termodinamicamente instabili ma con una stabilità maggiore rispetto alle macroemulsioni.
- Microemulsioni: dimensione delle gocce di $10\div 100\ \text{nm}$, termodinamicamente stabili.

Più le gocce sono grandi, più facilmente tenderanno ad avvicinarsi l'una all'altra fino alla separazione di fase. Le emulsioni alimentari ricadono in grossa parte nella classe delle macroemulsioni e sono perciò termodinamicamente instabili. La stabilità può comunque essere migliorata variando le condizioni ambientali, in particolare pH e temperatura, o grazie alla presenza di tensioattivi quali fosfolipidi e proteine. Queste due macromolecole sono particolarmente importanti nel caso della maionese, un'emulsione di olio in acqua, con l'aggiunta di sale e aceto che contribuisce ad abbassare il pH, in cui i tensioattivi derivano dal tuorlo d'uovo (McClements 2007; Das e Kinsella 1990).

1.2 Il tuorlo d'uovo

Nel caso della maionese, il tuorlo d'uovo fornisce tre componenti per la produzione e la stabilità dell'emulsione: l'acqua, i fosfolipidi e le proteine. La composizione del tuorlo d'uovo si può trovare in Tabella 1.1 (Anton 2007). I

Tabella 1.1: Composizione del tuorlo d'uovo.

Composizione [%]	
acqua	51,1
lipidi	30,6
proteine	16,0
carboidrati	0,6
minerali	1,7

lipidi a loro volta si dividono per il 62% in trigliceridi, per il 33% in fosfolipidi, per una frazione minore al 5% in colesterolo e meno dell'1% è rappresentata dai carotenoidi che donano il tipico colore arancione. Le proteine invece sono per la maggior parte lipoproteine e in particolare si ha il 68% di lipoproteine a bassa densità (LDL), il 16% di lipoproteine ad alta densità (HDL), il 10% di livetine e il 4% di fosfoviti. Tra le varie LDL è di particolare importanza l'Apopitellenina-1 (Apo-VLDL-II), un'apoproteina trasferita dal sangue direttamente nel tuorlo. Questa proteina è un omodimero di dimensioni

contenute (circa 9 kDa per monomero) che presenta un ponte disolfuro tra le due cisteine in posizione 75, le uniche presenti all'interno della molecola. Inoltre, è stato dimostrato sperimentalmente che svolge un ruolo importante nella stabilità della maionese grazie alle sue proprietà emulsionanti (Zambrowicz et al. 2015). Queste proteine non si trovano libere nel plasma del tuorlo d'uovo poiché non sono solubili in acqua ma si presentano in micelle composte da fosfolipidi, trigliceridi e colesterolo. Quando queste strutture raggiungono l'interfaccia che separa la fase acquosa da quella oleosa, nella fase di produzione della maionese, si rompono e liberano i fosfolipidi e le apoproteine che sono libere di diffondere all'interfaccia, denaturandosi e formando un film che stabilizza le emulsioni, come mostrato in Figura 1.2. Purtroppo, le informazioni sulla struttura dell'Apovitellenina-1, così come per altre proteine del tuorlo, sono limitate alla conoscenza della struttura primaria e a poco altro poiché risulta molto difficile purificarle dal tuorlo d'uovo. Altri componenti che influenzano la stabilità dell'emulsione, come

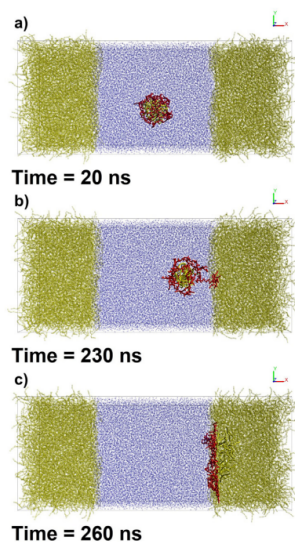


Figura 1.2: Rappresentazione dell'assorbimento di un granulo all'interfaccia acqua-olio (elaborazione immagine estratta da Ferrari et al. (2022)).

già detto, sono i fosfolipidi, molecole di glicerolo a cui sono legati due acidi grassi e un gruppo fosfato carico negativamente. La presenza di una parte polare e una apolare favorisce la stabilità tra le due fasi immiscibili: le code apolari si legheranno alla fase oleosa mentre la testa polare si legherà alla fase acquosa. È proprio grazie alla cooperazione di questi due elementi che si ottiene la stabilità della maionese e, in generale, di molte emulsioni.

Questo lavoro di tesi si propone di studiare, con tecniche di dinamica molecolare e di metadinamica, la stabilità della proteina Apovitellenina-1 all'interfaccia acqua-olio e la sua diffusività all'interno delle due fasi oltre

che l'interazione che si ha tra la proteina e i fosfolipidi analizzando sempre il comportamento all'interfaccia e come questo sia influenzato dalla presenza di altri tensioattivi. Il capitolo due conterrà i principali aspetti teorici della meccanica statistica, della dinamica molecolare classica e della metadynamica. Il capitolo tre descriverà invece i dettagli computazionali di tutte le simulazioni con la descrizione dei parametri e la spiegazione di eventuali ipotesi semplificative. Nel capitolo quattro saranno presentati i risultati delle simulazioni con tutte le operazioni di post-processing che sono state eseguite, per poi arrivare al capitolo cinque in cui verranno discussi i risultati ottenuti.

Capitolo 2

Dinamica Molecolare e Metadinamica

2.1 La meccanica statistica

La seguente sezione è tratta dal libro *Physical Chemistry* di Ira Levine (2009). La meccanica statistica rappresenta la base teorica delle simulazioni atomiche e molecolari e fornisce il contesto in cui, tramite analisi analitiche o computazionali, è possibile estrarre le proprietà macroscopiche di un sistema chimico a partire dal comportamento delle singole molecole che compongono il sistema. L'approccio statistico è necessario per definire il comportamento complessivo di un sistema chimico a partire dal comportamento delle singole molecole. È quindi necessario distinguere tra macrostato, che rappresenta lo stato termodinamico del sistema, e microstato, che rappresenta lo stato molecolare del sistema, definito a sua volta dallo stato delle singole molecole che lo compongono. A causa della natura delle interazioni fra le molecole e delle dimensioni delle molecole stesse lo stato delle molecole è definito dalla meccanica quantistica. Gli atomi e le molecole evolvono infatti obbedendo alle leggi della meccanica quantistica, al contrario degli oggetti macroscopici che evolvono secondo le leggi della meccanica classica, quali le leggi del moto di Newton. Molto spesso si trascurano comunque gli effetti quantistici e si considera il comportamento degli atomi e delle molecole classicamente, soprattutto quando si fa riferimento ai gradi di libertà traslazionali, rotazionali e vibrazionali. Questo è quello che viene fatto, ad esempio, nella dinamica molecolare classica, come verrà spiegato più nel dettaglio nel prosieguo di questa tesi.

Per parlare di meccanica statistica è necessario introdurre il concetto di *ensemble*. Con questo termine si intende l'insieme di infiniti sistemi non interagenti che si trovano tutti nello stesso macrostato, rispetto al sistema di interesse, ma che possono avere microstati differenti. I sistemi che compon-

gono l'ensamble possono essere visti come delle scatole incompressibili immerse in un bagno termico, con pareti rigide che garantiscono il passaggio di energia ma non quello di materia. Un insieme di sistemi con composizione, temperatura e volume costanti è detto *Canonical ensemble* e viene indicato con la sigla NVT. Se invece si ha la pressione costante al posto del volume si avrà il *Isothermal-Isobaric ensemble* indicato semplicemente con NpT. Una volta chiarito il concetto di ensemble si può parare dei due principi su cui si basa la meccanica statistica (Tolman 1979):

- Postulato *equal a priori*: dato un sistema termodinamico nell'ensemble canonico (composizione, volume e temperatura costanti), tutti i microstati con egual energia hanno la stessa probabilità di verificarsi.
- Teorema Ergodico: dato un sistema di interesse, la media temporale di una proprietà sarà uguale alla media d'insieme nell'ensemble canonico.

In particolare, l'equazione che lega la probabilità all'energia è la seguente:

$$p_j = \frac{e^{-\frac{E_j}{k_b T}}}{\sum_j e^{-\frac{E_j}{k_b T}}} = \frac{e^{-\frac{E_j}{k_b T}}}{Z}. \quad (2.1)$$

Il denominatore dell'equazione precedente, indicato con Z , in riferimento al termine tedesco *Zustandsumme* che tradotto significa *somma sugli stati*, è detto *Canonical partition function* e rappresenta la somma su tutti i possibili stati quantici ad un volume ed una temperatura fissati, T è la temperatura del sistema e k_b è la costante di Boltzman. Tale definizione è valida nel caso di un sistema che obbedisce alle leggi della meccanica quantistica. In questo caso infatti il sistema può trovarsi solo in determinati stati quantici, che risultano quantizzati e discreti. La stessa definizione è applicabile anche a sistemi che obbediscono le leggi della meccanica classica, in cui lo stato delle molecole varia in modo continuo fra un valore minimo e massimo. In questo caso le sommatorie sono sostituite da degli integrali. Dall'Eq. (2.1) si possono poi ottenere tutte le proprietà macroscopiche come l'energia interna che sarà la somma su tutte le possibili energie, come riportato nell'Eq. (2.2):

$$U = \sum_j p_j E_j = \frac{\sum_j E_j e^{-\frac{E_j}{k_b T}}}{Z} = k_b T^2 \left(\frac{\partial \ln Z}{\partial T} \right)_{V,N}. \quad (2.2)$$

La pressione può essere calcolata sapendo che per un sistema adiabatico reversibile la variazione di energia è uguale alla variazione di lavoro che a sua volta è funzione della variazione di volume secondo l'Eq. (2.3):

$$dU = -dw_{rev} = -P_j dV = dE_j. \quad (2.3)$$

Quindi la pressione può essere calcolata secondo l'Eq. (2.4):

$$P = \sum_j p_j P_j = - \frac{\sum_j e^{-\frac{E_j}{k_b T}} \left(\frac{\partial E_j}{\partial V} \right)}{Z} = k_b T \left(\frac{\partial \ln Z}{\partial V} \right)_{T,N}. \quad (2.4)$$

In generale, dalla *Canonical partition function* è possibile calcolare tutte le proprietà termodinamiche come l'entropia (Eq. 2.5), l'energia libera di Helmholtz (Eq. 2.6), il potenziale chimico (Eq. 2.7) e l'energia libera di Gibbs (Eq. 2.8):

$$S = \frac{U}{T} + k_b \ln Z, \quad (2.5)$$

$$A = -k_b T \ln Z, \quad (2.6)$$

$$\mu_B = -RT \left(\frac{\partial \ln Z}{\partial N_B} \right)_{T,V,N_{C \neq B}}, \quad (2.7)$$

$$G = -k_b T \ln Z + k_b T \left(\frac{\partial \ln Z}{\partial \ln V} \right)_{T,N}. \quad (2.8)$$

2.1.1 Interazione tra molecole

L'interazione tra due molecole all'interno del sistema è definita dall'interazione elettrostatica (Eq. 2.9), nel caso in cui si abbia a che fare con molecole cariche elettricamente, e dall'energia potenziale (Eq. 2.10) da cui poi è possibile determinare la forza agente sulle due molecole e per un sistema di N molecole si ha:

$$V_e = \sum_{i=1}^N \sum_{j<i} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}, \quad (2.9)$$

$$V_p \approx \sum_{i=1}^N \sum_{j<i} v_{ij}(r_{ij}). \quad (2.10)$$

Il calcolo del potenziale v_{ij} agente su due molecole a distanza r_{ij} può risultare estremamente complesso perciò si adottando delle approssimazioni per calcolarlo, in particolare si può avere il potenziale dipolo-dipolo (Eq. 2.11), il potenziale dipolo-dipolo-indotto (Eq. 2.12) e il potenziale di London (Eq. 2.13):

$$v_{d-d}(r) = -\frac{2}{3k_b T} \frac{\mu_1^2 \mu_2^2}{(4\pi\epsilon_0)^2 r^6}, \quad (2.11)$$

$$v_{d-di}(r) = -\frac{\mu_1^2 \alpha_2 + \mu_2^2 \alpha_1}{(4\pi\epsilon_0)^2 r^6}, \quad (2.12)$$

$$v_L(r) \approx -\frac{3I_1 I_2}{2(I_1 + I_2)} \frac{\alpha_1 \alpha_2}{(4\pi\epsilon_0)^2 r^6}, \quad (2.13)$$

dove μ_1 e μ_2 sono i momenti di dipolo delle due molecole che si stanno considerando, r è la distanza tra le molecole, k_b è la costante di Boltzman, T è la temperatura, ε_0 è la costante dielettrica nel vuoto, α_1 e α_2 sono le polarizzabilità delle due molecole e I_1 e I_2 sono le energie di ionizzazione. Dalla somma dei tre potenziali attrattivi a lungo raggio appena citati si può calcolare una forza (si ricorda che $F = -\nabla V$) attrattiva che prende il nome di forza di van der Waals. Oltre al contributo attrattivo a lungo raggio, il potenziale totale conterrà anche un contributo repulsivo a corto raggio che solitamente viene approssimato con l'Eq. (2.14):

$$v_{rep} \approx \frac{A}{r^n}, \quad (2.14)$$

dove A è una costante positiva, r è la distanza tra le molecole e n è un numero intero solitamente compreso tra 8 e 18. Il potenziale totale più utilizzato, somma del contributo repulsivo e di quello attrattivo, è detto *potenziale 6-12 di Lennard-Jones* mostrato nell'Eq. (2.15):

$$v = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (2.15)$$

dove il primo termine rappresenta il contributo repulsivo e il secondo quello attrattivo, ε è il valore di energia minima e σ è la distanza tra le molecole a cui il potenziale è nullo.

2.1.2 Dallo stato molecolare allo stato termodinamico

Quindi, per calcolare le variabili termodinamiche è necessario definire l'operatore Hamiltoniano per ogni molecola del sistema termodinamico (Eq. 2.16) e risolvere l'equazione di Schrödinger (Eq. 2.17) per ottenere i valori di energia E_j da cui poi è possibile calcolare la funzione di partizione canonica e da questa è possibile ricondursi alle proprietà termodinamiche citate sopra (Eq. da 2.5 a 2.8).

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2m_j} \frac{\partial}{\partial \mathbf{r}^2} + V(\mathbf{r}), \quad (2.16)$$

$$\hat{H}\psi_j = E_j\psi_j, \quad (2.17)$$

dove \hbar è la costante di Plank divisa per 2π , \mathbf{r} è il vettore posizione, $V(\mathbf{r})$ è l'energia potenziale, m_j è la massa di una molecola e ψ_j è la relativa funzione d'onda. Con questo approccio però è possibile ricavare soluzioni analitiche solo per sistemi ideali con potenziali di interazione nulli o molto semplici, come ad esempio il potenziale a sfere rigide o di van der Waals. Per tutti gli altri casi, in cui il potenziale è più complesso, è necessario utilizzare modelli computazionali in cui le equazioni vengono risolte numericamente con algoritmi implementati al calcolatore e questo è quello che avviene, ad

esempio, nella dinamica molecolare classica. È da ricordare inoltre, che quest'ultima trascura gli effetti quantici e le interazioni atomiche sono descritte con un insieme di parametri, detto campo di forze, che derivano da calcoli quanto-meccanici.

2.2 La dinamica molecolare

Le tecniche di dinamica molecolare (in inglese Molecular Dynamics, MD) sono diventate, negli anni, complementari alle tecniche di studio sperimentali, soprattutto nel caso in cui si vogliano studiare biomolecole o proteine. Secondo il teorema ergodico si può infatti studiare l'andamento nel tempo di una proprietà di un certo sistema e confrontarla con le medie d'insieme della stessa proprietà misurata sperimentalmente o sono usate per esplorare le varie configurazioni assunte da una proteina nel tempo (Hansson, Oostenbrink e Gunsteren 2002). Le basi su cui si pone la dinamica molecolare sono sostanzialmente due: il teorema ergodico e la risoluzione delle equazioni di Newton per un sistema di molecole, costituite a loro volta da atomi. Ogni sistema viene quindi trattato a livello atomistico, andando a considerare le interazioni tra ciascun atomo e tutti gli atomi che lo circondano. Ovviamente questo già presenta una limitazione poiché è la meccanica quantistica a descrivere correttamente il comportamento fisico di un sistema e non la meccanica classica ma ad oggi sono pochi gli strumenti computazionali abbastanza potenti da poter risolvere le equazioni di Schrödinger per un sistema che contiene un numero elevato di molecole, dell'ordine delle migliaia (Lindahl 2008). Dall'altro lato, simulare un sistema più piccolo può essere impossibile visto che la dinamica molecolare, come già detto, è utilizzata per macromolecole e proteine che già intrinsecamente contengono un numero elevato di atomi e che devono essere poste in un sistema a sua volta composto da altre molecole. A partire dalle tecniche MD sono state sviluppate anche altre tecniche più accurate, come la Metadinamica (MetaD), ma ovviamente questo ha un costo, in termini di tempo computazionale, molto maggiore. La risoluzione delle equazioni del moto di Newton risulta quindi un buon compromesso finale dato che riesce spesso a simulare correttamente le proprietà misurate sperimentalmente. In Figura 2.1 sono rappresentati alcuni dei possibili fenomeni studiati dalla dinamica molecolare con le relative scale temporali coinvolte. In particolare, si può osservare che eventi che richiedono tempi relativamente lunghi, come la sintesi dei ribosomi o il ripiegamento delle proteine non siano studiabili atomisticamente a causa dell'elevato costo computazionale.

2.2.1 Il campo di forze o force field

Il primo passo per effettuare una simulazione MD è la scelta di un campo di forze, ovvero l'insieme di approssimazioni ed equazioni empiriche e non per il calcolo delle interazioni atomiche e dell'energia potenziale del sistema. Una

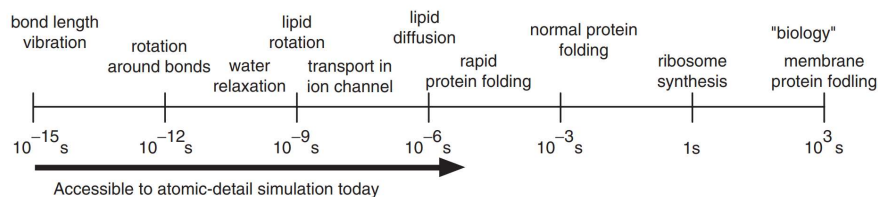


Figura 2.1: Tempi caratteristici di alcuni eventi tipicamente studiati con la dinamica molecolare (elaborazione immagine estratta da (Lindahl 2008)).

caratteristica generale che accomuna quasi tutti i campi di forze è che non riescono a predire la formazione e la rottura di legami (esistono comunque alcuni campi di forze che contengono parametri che sono in grado di predire, con molta approssimazione, le reazioni chimiche), non arrivano quindi a riprodurre gli effetti quantistici del sistema in esame messi in gioco ad esempio durante le reazioni chimiche. Un'altra caratteristica tipica di ogni campo di forze è la suddivisione delle funzioni per il calcolo dell'energia potenziale in interazioni di legame, che riguardano le vibrazioni, rotazioni e torsioni di un legame tra due o più atomi di una molecola e in interazioni di non legame, che riguardano invece in interazioni di Coulomb e di Lennard-Jones (si ricorda che queste interazioni non si hanno solo tra atomi di molecole diverse ma anche tra atomi di una stessa molecola se questa è molto grande, come nel caso delle proteine o dei polimeri). In generale, i parametri contenuti nel campo di forze riguardano un particolare tipo di atomo all'interno di una particolare molecola. Ad esempio, l'ossigeno legato all'idrogeno nella molecola dell'acqua sarà diverso dall'ossigeno legato ad un idrogeno in un alcol. Una pratica usata di frequente riguarda quella di inserire, all'interno di un campo di forze, i parametri relativi ad un singolo atomo di un legame e poi estrapolare gli altri parametri da medie geometriche e aritmetiche. Ad esempio, i parametri del legame C-N saranno dati da una media tra i parametri del legame C-C e quelli del legame N-N (Israelachvili 2011). Oltre tutto, i campi di forza trattano con lo stesso approccio appena descritto le interazioni di van der Waals e di Coulomb. In particolare, per queste ultime vengono definite delle cariche parziali. All'interno delle molecole, gli elettroni tenderanno ad orbitare vicino agli atomi più elettronegativi ma è comunque possibile che, nel moto di rotazione, si trovino vicino ad un atomo poco elettronegativo. Con le cariche parziali si esprime quindi la possibile polarità di una molecola anche se questa è scarica. Nel caso dell'acqua ad esempio, in uno dei modelli più utilizzati si pone una carica di $-0,834 e^-$ per l'ossigeno e una carica di $+0,417 e^-$ per ciascun idrogeno (Jorgensen, Chandrasekhar et al. 1983). Di seguito sono riportate le principali equazioni utilizzate per il calcolo dell'energia potenziale e delle corrispondenti forze che

le coppie di atomi si scambiano:

$$E_{bonds} = \sum_{bonds} K_r (r - r_{eq})^2, \quad (2.18)$$

$$E_{angles} = \sum_{angles} K_\theta (\theta - \theta_{eq})^2, \quad (2.19)$$

$$E_{torsion} = \sum_i \frac{V_1^i}{2} [1 + \cos(\phi_i + f_1^i)] + \frac{V_2^i}{2} [1 + \cos(2\phi_i + f_2^i)] + \frac{V_3^i}{2} [1 + \cos(3\phi_i + f_3^i)], \quad (2.20)$$

$$E_{ab} = \sum_i^{on} \sum_j^{on} \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \right] f_{ij}, \quad (2.21)$$

$$E = E_{ab} + E_{torsion} + E_{angles} + E_{bonds}, \quad (2.22)$$

$$\mathbf{F}_i = -\nabla E_i = m_i \frac{d^2 \mathbf{r}}{dt^2}, \quad (2.23)$$

dove K_r e K_θ delle Eq. (2.18) e (2.19) sono rispettivamente le costanti elastiche per la vibrazione di legame e angolare e $(r - r_{eq})$ e $(\theta - \theta_{eq})$, in riferimento alle stesse equazioni, sono rispettivamente gli spostamenti spaziale e angolare rispetto alla condizione di equilibrio. Nel contributo energetico dovuto alla torsione dell'Eq. (2.20) i termini V_1 , V_2 , V_3 , corrispondono ai coefficienti di Fourier della serie mentre ϕ rappresenta l'angolo torsionale e f_1 , f_2 , f_3 rappresentano gli angoli di fase. Infine, il termine energetico per le interazioni di non legame dell'Eq. (2.21) è dato dal contributo di Coulomb sommato a quello di Lennard-Jones. È da tenere in considerazione che quest'ultima formula è applicata anche ad atomi facenti parte di una stessa molecola ma separati da tre o più legami. Inoltre, $f_{ij} = 1$ a parte nel caso in cui si ha un legame intermolecolare di tipo 1,4 in cui $f_{ij} = 0,5$. Infine, le Eq. (2.22) e (2.23) rappresentano la risoluzione delle equazioni di Newton. La differenza tra i diversi campi di forze esistenti in letteratura è proprio nei parametri con cui viene calcolata l'energia. Non esiste un campo di forza che racchiuda tutti gli atomi e le molecole esistenti, però è possibile scegliere un campo di forze come base per poi andare a variare alcuni parametri, o per fonderlo con un altro campo di forze se, ad esempio, il secondo contiene informazioni su molecole che non erano presenti nel campo di forze originariamente scelto. Attualmente, tra i vari campi di forza, quello che risulta più completo è l'OPLS-AA (Optimized Potentials for Liquid Simulations – All Atom) sviluppato da Jorgensen, Madura e Swenson (1984). Questo campo di forze tiene conto di tutte le interazioni che ci sono tra tutti gli atomi che compongono il sistema. Altri tipi di campi di forza contengono delle semplificazioni: ad esempio non presentano i parametri per le interazioni tra

gli idrogeni di una proteina andando così a ridurre il costo computazionale (Jorgensen, Maxwell e Tirado-Rives 1996). Ciò significa che due simulazioni simili, con stesso numero e posizione iniziale degli atomi, che si differenziano solo per il tipo di campo di forze scelto, possono dare risultati molto diversi.

2.2.2 Simulazione di dinamica molecolare

La simulazione di dinamica molecolare è caratterizzata dalla risoluzione delle equazioni del moto (Eq. da 2.18 a 2.23) con metodi numerici. È quindi necessario definire un intervallo di tempo per l'integrazione delle equazioni del moto di pochi femtosecondi, di solito 1 o 2 fs. Inoltre, per poter regolare temperatura e pressione, vengono introdotti un termostato e un barostato, illustrati meglio nel Paragrafo 3.2. Il metodo per la risoluzione delle equazioni del moto è l'algoritmo di Verlet (Swope et al. 1982):

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{\Delta t^2}{2m_i}\mathbf{F}_i(t), \quad (2.24)$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{\Delta t}{m_i}[\mathbf{F}_i(t) + \mathbf{F}_i(t + \Delta t)]. \quad (2.25)$$

Questo algoritmo garantisce che se due sistemi identici vengono simulati partendo da identiche condizioni iniziali giungeranno allo stesso risultato finale. In molte situazioni non è dato sapere quanto tempo sarà necessario al sistema per raggiungere la situazione desiderata o uno stato di equilibrio particolare ma in ogni caso si può decidere di iniziare la simulazione con un certo numero di passi temporali e poi eventualmente ripartire dalla disposizione finale del sistema per lanciare altri passi temporali senza dover ripetere tutte le fasi di minimizzazione energetica ed equilibratura. Solitamente, è comunque consigliabile scegliere come durata della simulazione un tempo almeno dello stesso ordine di grandezza del tempo caratteristico dell'evento che si vuole studiare.

2.3 La Metadinamica

Discutiamo ora molto sinteticamente i principi alla base della metadinamica. Maggiori dettagli sono disponibili in letteratura come ad esempio nell'articolo di Valsson, Tiwary e Michele Parrinello (2016).

Nonostante l'ottimizzazione delle risorse e dei sistemi per svolgere simulazioni di dinamica molecolare, ad oggi il problema principale restano i lunghi tempi di simulazione che alcuni processi presentano. Questo perché lo strumento principale per lo studio di sistemi con simulazioni MD, ovvero la parallelizzazione, non permette di intervenire in maniera significativa sulla scala temporale, che è intrinsecamente seriale, ma piuttosto aiuta a riprodurre sistemi più grandi e complessi. Oltretutto c'è da considerare che sistemi proteici possono necessitare di simulazioni lunghe decine di microsecondi e che spesso sono necessarie più simulazione per ottenere un dato statistico, soprattutto se questo riguarda eventi di transizione. Per superare le limitazioni della scala temporale si possono usare metodi di campionamento delle variabili introducendo un bias, un potenziale energetico, che dipende dalle variabili scelte (metodo di campionamento a ombrello). Poiché costruire il bias non è un'operazione semplice si può scegliere come soluzione quella di far variare il bias stesso durante la simulazione. Uno dei metodi che segue questa strategia è proprio la metadinamica (MetaD).

2.3.1 Sampling nello spazio delle variabili collettive

Per descrivere un sistema complesso si deve cercare un numero di variabili ristretto che possano descrivere al meglio l'evoluzione del sistema stesso. Da un punto di vista matematico, queste variabili collettive (CV), sono funzioni generalmente non lineari delle coordinate atomiche. Una volta definite le CV si può definire una distribuzione di equilibrio secondo l'Eq. (2.26):

$$P(\mathbf{s}) = \int \delta[\mathbf{s} - \mathbf{s}(\mathbf{r})] P(\mathbf{r}) d\mathbf{r} = \langle [\mathbf{s} - \mathbf{s}(\mathbf{r})] \rangle, \quad (2.26)$$

dove il termine $P(\mathbf{r})$, espresso dall'Eq. (2.27),

$$P(\mathbf{r}) = \frac{e^{-\beta U(\mathbf{r})}}{Z} \quad (2.27)$$

è la distribuzione di Boltzmann corrispondente all'energia potenziale U e \mathbf{s} è lo spazio delle variabili collettive. Si può inoltre definire la superficie di energia libera (FES) con l'Eq. (2.28):

$$F(\mathbf{s}) = -\frac{1}{\beta} \log P(\mathbf{s}). \quad (2.28)$$

Questa funzione consente di distinguere chiaramente le zone metastabili e le eventuali barriere energetiche del sistema al variare delle CV, la cui scelta

risulta quindi importante ai fini della determinazione degli stati metastabili del sistema. La barriera energetica tra due stati metastabili può essere calcolata dall'Eq. (2.29):

$$\Delta F_{A,B} = -\frac{1}{\beta} \log \frac{\int_A e^{-\beta F(\mathbf{s})} d\mathbf{s}}{\int_B e^{-\beta F(\mathbf{s})} d\mathbf{s}}. \quad (2.29)$$

Per un sistema ergodico la FES si può anche calcolare dalle Eq. (2.30) e (2.31):

$$F(\mathbf{s}) = -\frac{1}{\beta} \lim_{t \rightarrow +\infty} (\log (N(\mathbf{s}, t))), \quad (2.30)$$

$$N(\mathbf{s}, t) = \left(\frac{\int_0^t \delta[\mathbf{s} - \mathbf{s}(\mathbf{r}(t'))] dt'}{\int_0^t dt'} \right), \quad (2.31)$$

dove N corrisponde ad un istogramma normalizzato valutato a partire da una simulazione di dinamica molecolare classica. Per stimare la FES sembrerebbe quindi sufficiente “contare”. Il problema è che nella maggior parte dei casi un sistema non può essere considerato ergodico per il tempo di simulazione, durante il quale il sistema tenderà solo ad oscillare attorno ad un particolare stato locale di equilibrio. Per superare questo problema si può considerare un potenziale di polarizzazione $V(\mathbf{s}(\mathbf{R}))$ che è progettato per migliorare il campionamento nello spazio delle CV. Si ottiene quindi la distribuzione di equilibrio secondo l'Eq (2.32):

$$P_V(\mathbf{s}) = \int \delta[\mathbf{s} - \mathbf{s}(\mathbf{r})] P_V(\mathbf{r}) d\mathbf{r} = \frac{e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{\int e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}_{\mathbf{s}\mathbf{s}}} = \frac{e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{Z}, \quad (2.32)$$

Questa è la distribuzione di Boltzmann del sistema a cui è stato applicato un potenziale. Da questa equazione si può ricavare l'Eq. (2.33) che rappresenta la FES della simulazione priva di bias:

$$F(\mathbf{s}) = -\frac{1}{\beta} \log N_V(\mathbf{s}) - V(\mathbf{s}), \quad (2.33)$$

In linea di principio, dall'Eq. (2.34) si può stimare qualsiasi funzione dipendente da \mathbf{r} pesando ciascuna configurazione ottenuta nell'*ensemble* a cui è applicato il bias con il valore stesso del bias che agisce sul sistema:

$$\langle O(\mathbf{r}) \rangle = \frac{\langle O(\mathbf{r}) e^{\beta V(\mathbf{s}(\mathbf{r}))} \rangle_V}{\langle e^{\beta V(\mathbf{s}(\mathbf{r}))} \rangle_V}, \quad (2.34)$$

dove $O(\mathbf{r})$ è una generica funzione dipendente dalle coordinate atomiche. È da tenere in considerazione che questa relazione è valida solo se il potenziale applicato al sistema è statico. Un esempio di potenziale che si può applicare è quello presentato nell'Eq. (2.35):

$$V(\mathbf{s}) = -\left(1 - \frac{1}{\gamma}\right) F(\mathbf{s}), \quad (2.35)$$

in cui $\gamma > 1$ è una costante e il limite $\gamma \rightarrow 1$ corrisponde all'*ensemble* senza bias. L'effetto di questo bias è quello di amplificare i due picchi in modo che la transizione tra uno stato metastabile e l'altro sia più favorita come mostrato nella rappresentazione di Figura 2.2. Per $\gamma \rightarrow +\infty$ il valore della probabilità tende ad una costante e si perdono i picchi, è quindi importante la scelta di γ , oltre alla scelta delle CV, per poter esplorare le diverse configurazioni di energia minima. Il suo valore non deve essere troppo piccolo, altrimenti le barriere energetiche resterebbero difficili da superare, ma nemmeno troppo grande, poiché si rischierebbe di andare ad esplorare valori di CV improbabili e quindi aumenterebbe il costo computazionale. Con questo bias, la probabilità di distribuzione è legata al valore senza bias dall'Eq.(2.36):

$$P_V(\mathbf{s}) = \frac{[P(\mathbf{s})]^\frac{1}{\gamma}}{\int [P(\mathbf{s})]^\frac{1}{\gamma} d\mathbf{s}} \quad (2.36)$$

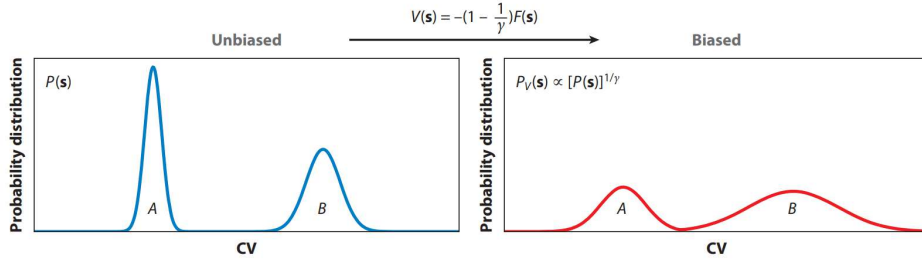


Figura 2.2: Effetto del bias sulla probabilità. Dopo l'applicazione del potenziale (grafico rosso) il picco A risulta più basso e per il sistema sarà più semplice raggiungere la configurazione B (elaborazione immagine estratta da Ref. Valsson, Tiwary e Michele Parrinello (2016)).

2.3.2 Well-Tempered Metadynamics

La Well-Tempered Metadynamics (WTMetaD) è una procedura che costruisce iterativamente un bias $V(\mathbf{s})$ che tende asintoticamente all'Eq. (2.35) grazie all'aggiunta successiva di piccoli kernel del bias repulsivi $G(\mathbf{s}, \mathbf{s}')$. Questi kernel (Eq. 2.37 e 2.38) assumono la forma di una Gaussiana localizzata nello spazio delle variabili collettive \mathbf{s} :

$$G(\mathbf{s}, \mathbf{s}') = W e^{\|\mathbf{s} - \mathbf{s}'\|^2}, \quad (2.37)$$

$$\|\mathbf{s} - \mathbf{s}'\|^2 = \frac{1}{2} \sum_{i,j} (\mathbf{s}_i - \mathbf{s}'_i) \sum_{i,j}^{-1} (\mathbf{s}_j - \mathbf{s}'_j), \quad (2.38)$$

dove W rappresenta l'altezza della Gaussiana, $\|\mathbf{s} - \mathbf{s}'\|$ indica la distanza tra \mathbf{s} e \mathbf{s}' , $\sum_{i,j}$ rappresenta la matrice di covarianza e $\sum_{i,j}^{-1}$ la sua inversa.

Dall'Eq. (2.39) si può trovare l'espressione del bias:

$$V_n(\mathbf{s}) = V_{n-1}(\mathbf{s}) + G(\mathbf{s}, \mathbf{s}_n) e^{-\frac{1}{\gamma-1} \beta V_{n-1}(\mathbf{s}_n)}, \quad (2.39)$$

in cui $V_0(\mathbf{s}) = 0$ e il termine correttivo diminuisce man a mano che la simulazione procede poiché l'argomento dell'esponenziale decresce come $1/n$. Questa equazione viene applicata ogni N_G steps. Si può quindi ottenere la probabilità di distribuzione del bias al tempo t dall'Eq. (2.40):

$$P_V(\mathbf{r}, t) = \frac{e^{-\beta[U(\mathbf{r})+V(\mathbf{s}(\mathbf{r},t))]}{\int e^{-\beta[U(\mathbf{r})+V(\mathbf{s}(\mathbf{r},t))]} d\mathbf{r}} = P(\mathbf{r}) e^{-\beta[V(\mathbf{s}(\mathbf{r},t))-c(t)]}, \quad (2.40)$$

con il termine $c(t)$ espresso nell'Eq. (2.41):

$$c(t) = \frac{1}{\beta} \log \left(\frac{\int e^{-\beta F(\mathbf{s})} d\mathbf{s}}{\int e^{-\beta[F(\mathbf{s})+V(\mathbf{s},t)]} d\mathbf{s}} \right). \quad (2.41)$$

In conclusione, si ottiene che, data una funzione dipendente dalle coordinate atomiche $O(\mathbf{r})$, si può calcolare la media sul sistema privo di bias dalla traiettoria della simulazione di metadinamica dall'Eq. (2.42):

$$\langle O(\mathbf{r}) \rangle = \langle O(\mathbf{r}) e^{-\beta[V(\mathbf{s}(\mathbf{r},t))-c(t)]} \rangle_V \quad (2.42)$$

.

2.3.3 Variabili Collettive

La metadinamica è, in linea di principio, indipendente dalla scelta delle CV. In pratica però la scelta delle CV influenza il costo computazionale. Per poter scegliere con cura le CV ci si può basare su simulazioni precedenti di sistemi noti o da dati di letteratura ma in generale, se il sistema finale non è noto, è buona norma scegliere delle CV che possano fornire una moltitudine di risultati. Una volta selezionato un set di CV preliminare è necessario eseguire una simulazione MetaD ed analizzare l'andamento del sistema. Ci si può trovare in tre situazioni:

- Il sistema è fisso nello stesso stato di partenza, la scelta dei CV è errata.
- Il sistema esplora correttamente diversi metastati e si ottiene una stima dell'energia convergente in un breve lasso di tempo, la scelta dei CV è corretta.
- Il sistema è intrappolato per lungo tempo in diversi bacini energetici e il potenziale di bias non cresce in modo uniforme, la scelta dei CV è incompleta.

La terza situazione è la più probabile e per far fronte a ciò bisognerebbe identificare parametri che variano lentamente e che potrebbero essere responsabili dell'intrappolamento occasionale del sistema in un certo metastato. Assumendo di avere un sistema con due stati noti A e B e che il cammino tra questi due stati possa essere espresso attraverso un insieme di CV. Si scrive una traiettoria $\mathbf{S}(t)$ che collega i due stati e si sceglie, tra gli infiniti cammini possibili, un cammino di riferimento $\mathbf{S}_0(t)$. È poi necessario definire una metrica che esprime le distanze nello spazio di \mathbf{S} . Si possono quindi costruire due funzioni (Eq. 2.43 e 2.44):

$$s(\mathbf{r}) = \lim_{\lambda \rightarrow +\infty} \left(\frac{\int_0^1 t e^{-\lambda \|\mathbf{S}(\mathbf{r}) - \mathbf{S}_0(t)\|^2} dt}{\int_0^1 e^{-\lambda \|\mathbf{S}(\mathbf{r}) - \mathbf{S}_0(t)\|^2} dt} \right), \quad (2.43)$$

$$z(\mathbf{r}) = \lim_{\lambda \rightarrow +\infty} \left(-\frac{1}{\lambda} \log \int_0^1 e^{-\lambda \|\mathbf{S}(\mathbf{r}) - \mathbf{S}_0(t)\|^2} dt \right). \quad (2.44)$$

La variabile s induce una foliazione nello spazio \mathbf{S} e misura l'avanzamento lungo il percorso e il suo valore è compreso tra 0, che corrisponde allo stato iniziale, e 1, che corrisponde al raggiungimento dello stato finale. La variabile z esprime la distanza dal percorso di riferimento \mathbf{S}_0 , il suo valore sarà tanto maggiore quanto più il percorso si discosta dal percorso di riferimento. In Figura 2.3 si possono vedere graficamente gli andamenti delle due variabili.

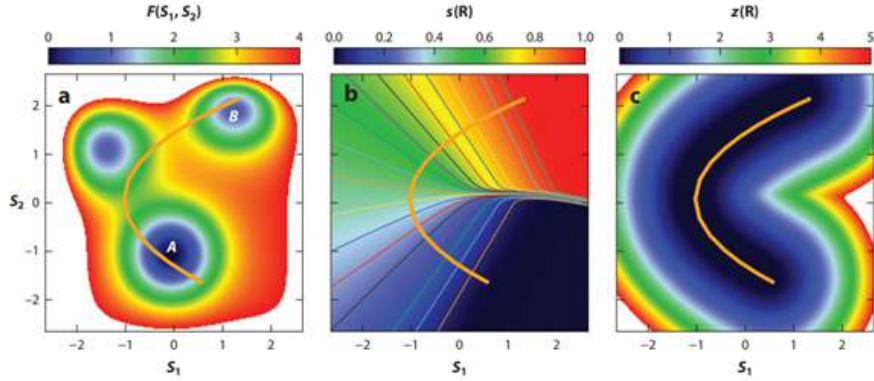


Figura 2.3: (a) FES nello spazio delle CV $\mathbf{S} = (S_1, S_2)$ con lo stato iniziale A, lo stato finale B e il percorso di riferimento \mathbf{S}_0 . (b) il valore di s che esprime il completamento del percorso. (c) il valore di z che esprime quanto un percorso generico si discosta dal percorso di riferimento \mathbf{S}_0 (elaborazione immagine estratta da Ref. Valsson, Tiwary e Michele Parrinello (2016)).

2.3.4 Varianti della metadinamica

Di seguito sono riportati alcuni altri esempi e varianti della metadinamica rispetto alla WTMetaD.

- Metadinamica dei camminatori multipli: in questo tipo di simulazioni si ha sempre la creazione iterativa di un bias depositando kernel successivi ma questo viene fatto in contemporanea su più simulazioni in parallelo. Il bias totale sarà dato dalla somma di tutti i kernel depositati in ogni simulazione.
- Metadinamica con Gaussiana adattativa: nello schema Gaussiano adattativo la matrice di covarianza viene calcolata durante la simulazione effettuando una media temporale sui CV e questo porta ad un adattamento continuo dei kernel per la creazione iterativa del bias. Scegliendo questo metodo si migliorano le prestazioni ma le equazioni presentate precedentemente non sono più valide.
- Metadinamica Flux-Tempered: in questo caso il potenziale di bias porta alla distribuzione ottimizzata per diffusione secondo l'Eq. (2.45):

$$P_V(s) \propto \frac{1}{\sqrt{D(s)}}, \quad (2.45)$$

dove D è la diffusività locale dipendente da s che è compresa nell'intervallo $\mathbf{s}_- \leq s \leq \mathbf{s}_+$. La costruzione del bias procede in due fasi: nella prima parte viene eseguita una normale simulazione MetaD e la deposizione delle gaussiane per la creazione dei kernel viene stoppata dopo un certo tempo, fissando quindi il bias; nella seconda parte il bias viene nuovamente aggiornato attraverso un diverso schema iterativo che porta poi all'Eq. (2.45).

- Metadinamica Transition-Tempered: questa può essere vista come la combinazione della MetaD classica e della WTMetaD. In questo caso i kernel vengono mantenuti costanti fino a quando il sistema non esplora tutti i bacini di stabilità. Successivamente la forza dei kernel viene gradualmente ridotta. Viene quindi combinata la convergenza della WTMetaD con l'esplorazione più rapida di ciascun FES offerta dalla MetaD. È una procedura molto utile ma ha l'inconveniente che è necessario conoscere tutti i bacini che si vogliono esplorare per poter definire quando il contributo dei kernel deve venire meno.

2.3.5 Confronto metadinamica e dinamica molecolare

La metadinamica risulta essere uno strumento di notevole aiuto nello studio di un sistema. Sono presenti però delle limitazioni. Innanzi tutto, è necessario tenere in considerazione che, a parità di numero di passi temporali tra una simulazione di dinamica classica e una simulazione di metadinamica, l'implementazione di una simulazione di MetaD ha un costo computazionale notevolmente maggiore. Però una singola simulazione di metadinamica fornisce molte più informazioni, per le variabili che riguardano le CV, rispetto

ad una simulazione di dinamica molecolare classica. Dall'altro lato, con la metadinamica, si perde qualsiasi informazione temporale. Si può solo sapere se, partendo da una configurazione A, il sistema evolverà verso una configurazione B e con che probabilità ma non si può sapere in quanto tempo questo evento si concretizzerà. È comunque importante sottolineare che la metadinamica viene scelta come strumento di studio per sistemi che impiegherebbero molto tempo ad evolvere verso un nuovo stato stabile e che per valutare la probabilità di accadimento di un evento (per esempio il passaggio dalla configurazione A del sistema alla configurazione B) sarebbero necessarie numerose simulazioni di dinamica molecolare classica su cui poter poi fare una media d'insieme per applicare il teorema ergodico.

Capitolo 3

Dettagli computazionali

Tutte le simulazioni di dinamica molecolare descritte in questo capitolo sono state svolte con GROMACS versione 20.6, le simulazioni di metadinamica sono invece state eseguite con GROMACS 20.6 e Plumed 2.7.2.

3.1 Le condizioni periodiche e il cut-off

Se immaginiamo il sistema da simulare come un parallelepipedo, la dimensione di ciascun lato avrà una dimensione variabile circa tra 5 nm e 30 nm, a seconda del tipo e del numero di molecole che conterrà, perciò, ci si pone il problema di cosa ci sia oltre il sistema in esame, ovvero in quella porzione dello spazio che non è stata definita. Poiché considerare il vuoto all'esterno del sistema o delle pareti solide non avrebbe avuto un senso fisico sono state adottate le Periodic Boundary Conditions (PBC). Una volta creato il sistema di interesse, di questo ne verranno create delle copie ideali in tutte le direzioni, in questo modo, se immaginiamo che la box contenga acqua, una molecola che esce da una faccia della box rientrerà dalla faccia opposta con la stessa velocità in modulo, direzione e verso che aveva quando si trovava nel momento dell'uscita dalla box, come raffigurato nell'immagine rappresentativa di Figura 3.1. Ovviamente con questa soluzione si rischia che le molecole di interesse della simulazione, tipicamente una proteina, risentano della propria immagine periodica e per risolvere questo problema si inserisce una distanza di cut-off, ovvero si considera che ogni singolo atomo interagisca solo con gli atomi presenti all'interno di una sfera che ha come centro l'atomo di riferimento e come raggio la distanza di cut-off. In questo modo, per evitare l'interazione tra una molecola di interesse e la sua immagine periodica è sufficiente creare una box che abbia una lunghezza per ciascun lato pari ad almeno la massima estensione raggiungibile dalla molecola sommata al doppio della distanza di cut-off. Per tener conto di quali atomi si trovano ad una distanza minore del raggio di cut-off rispetto ad una certa particella si usa la lista di Verlet (Chialvo e Debenedetti 1990), una particolare algo-

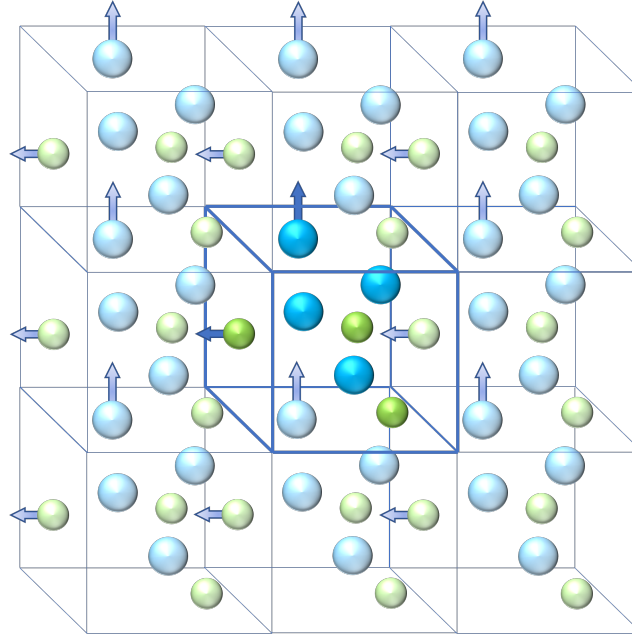


Figura 3.1: Rappresentazione delle PBC. La box centrale rappresenta il sistema che si vuole simulare, le box laterali rappresentano delle ideali copie del sistema.

ritmo a basso costo computazionale che permette di tener conto facilmente dell'elenco di particelle contenute all'interno di una sfera di raggio pari alla distanza di cut-off. Questo algoritmo prevede che il sistema sia diviso in celle cubiche con l'atomo di riferimento all'interno della cella 0. A ciascun altro atomo, che si troverà in una cella i , viene assegnato un valore che dipende dalla distanza tra la cella i e la cella 0 e dalla dimensione del lato della cella. Maggiore sarà il valore assegnato e maggiore sarà la distanza dall'atomo di riferimento. Oltretutto, limitando il calcolo delle interazioni a lungo raggio solo a quelle comprese all'interno del cut-off, si riduce notevolmente il costo computazionale. Inoltre, grazie alla presenza della periodicità si può calcolare il potenziale di Coulomb con le Particle Mesh Ewald summation (PME). Una condizione necessaria per poter applicare queste formule è proprio il fatto che si stia lavorando con celle periodiche (viene tipicamente usata nello studio dei cristalli) (Lindahl 2008). Come si può vedere nell'immagine sottostante, Figura 3.2, considerando solo la distanza di cut-off, il potenziale di Coulomb avrebbe un andamento privo di significato fisico. Grazie alle PME il contributo di Coulomb viene scomposto nella somma di due contributi, il primo a corto raggio, che tipicamente raggiunge lo zero asintotico ad una distanza inferiore al cut-off nello spazio reale, e il secondo a lungo raggio che converge rapidamente nello spazio di Fourier (Di Pierro, Elber e Leimkuhler 2015). La somma di questi due contributi garantisce che il potenziale sia

una funzione che tende a 0 quando r , la distanza tra i centri di due atomi, tende ad infinito ma allo stesso tempo consente di avere un andamento più realistico del potenziale.

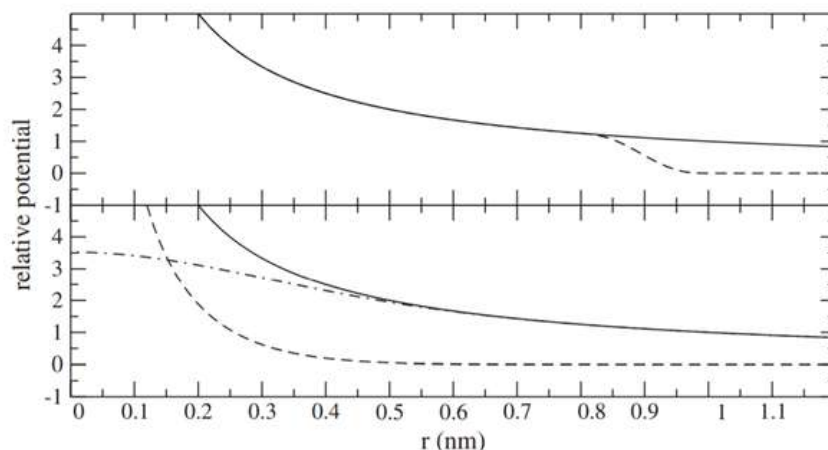


Figura 3.2: Utilizzo delle PME: nel grafico superiore si nota la differenza tra il potenziale reale (linea continua) e il potenziale calcolato tenendo conto solo della distanza di cut-off (linea tratteggiata). Nel grafico inferiore si notano i due contributi del potenziale calcolati con le PME: il potenziale a lungo raggio (linea tratto punto) e il potenziale a corto raggio (linea tratteggiata), confrontati con il potenziale reale (linea continua) (elaborazione immagine estratta da Lindahl (2008)).

3.2 Preparazione del sistema

Una volta creata la box che si vuole simulare sono necessari ulteriori tre passi prima di poter effettuare la simulazione di dinamica molecolare vera e propria. Questi tre passaggi sono in ordine:

1. Neutralizzazione della carica elettrica
2. Minimizzazione energetica
3. Equilibratura di temperatura e densità

3.2.1 Neutralizzazione della carica elettrica

Con l'aggiunta di ioni quello che si vuole ottenere è l'annullamento della carica del sistema. Infatti, le simulazioni MD sono tipicamente simulazioni all'equilibrio e la presenza di una carica elettrica potrebbe influenzare negativamente l'andamento di tutta la simulazione dato che in natura non esiste

una carica netta ma tutti i sistemi sono elettricamente neutri. Oltretutto, la scelta di alcuni parametri o modelli, come per esempio il modello di Ewald per il calcolo del potenziale elettrostatico, funzionano bene solo se il sistema ha carica nulla (Kolafa e Perram 1992). Solitamente, per annullare la carica si aggiungono al sistema ioni Cl^- per contrastare la carica negativa e ioni Na^+ per contrastare quella positiva. Tipicamente questi ioni si sostituiscono al solvente e la loro presenza non cambia significativamente la composizione della box.

3.2.2 Minimizzazione energetica

Una volta aggiunti gli ioni la box può dirsi a tutti gli effetti completa per quanto riguarda il numero e i tipi di atomi presenti. Il secondo passaggio è quindi la minimizzazione energetica. Si effettua una simulazione di pochi passi temporali con lo scopo di andare a muovere gli atomi attorno alle loro posizioni iniziali per ottenere un minimo di energia locale del sistema. Al momento dell'inserimento delle molecole, queste potrebbero avere un'orientazione tale per cui l'energia avrebbe un valore elevato che poi porterebbe ad un valore della forza elevato. Questo potrebbe essere dovuto al semplice inserimento randomico delle molecole all'interno della box che potrebbe portare ad avere due atomi inizialmente molto vicini oppure la configurazione di una molecola inserita nella box potrebbe non essere quella ad energia minore. Questa pratica non è obbligatoria ma è fortemente consigliata dal momento che lanciare una simulazione senza la minimizzazione energetica potrebbe portare al collasso del sistema a causa delle elevate forze che potrebbero agire sugli atomi (Chatzieftheriou, Adendorff e Lagaros 2016). Inoltre, la minimizzazione energetica è un primo controllo della buona disposizione degli atomi nella box, questo perché, in caso di atomi troppo vicini o addirittura sovrapposti, si avrebbe un valore di energia pari ad infinito.

3.2.3 Equilibratura

Terminata la fase di minimizzazione energetica non si può ancora lanciare la simulazione vera e propria poiché è necessario portare all'equilibrio temperatura e densità. Teoricamente si vorrebbe portare all'equilibrio anche la pressione ma una delle grosse limitazioni della dinamica molecolare è proprio legato alle enormi oscillazioni del valore di pressione rispetto al valore di riferimento impostato (Nosé e Klein 1983). Come conseguenza si ha che anche tutte le proprietà calcolate partendo dalla pressione, come per esempio la tensione interfacciale, risentiranno di questo errore e risulteranno molto difficili da stimare. Una delle pratiche più comuni consiste nell'eseguire l'equilibratura in due fasi: nella prima si esegue una simulazione di tipo NVT (Canonical ensemble: numero di molecole, volume e temperatura costanti) per portare la temperatura al valore desiderato e per equilibrare, per

quanto possibile, la pressione; nella seconda si esegue una simulazione NpT (Isothermal-Isobaric ensemble: numero di molecole, pressione e temperatura costanti) per equilibrare la densità del sistema. Durante l'equilibratura le molecole di soluto (es. proteine o altre macromolecole di interesse) vengono fissate nelle loro posizioni inserendo dei vincoli energetici che non consentono il movimento. In questo modo tutte le molecole di solvente adattano la loro posizione per portare all'equilibrio le grandezze elencate precedentemente. Per ottenere il valore di pressione e temperatura desiderati si inseriscono rispettivamente un barostato e un termostato, ovvero un insieme di equazioni che mirano a portare a convergenza le due variabili fino al valore desiderato. In particolare, tra i metodi più noti, ed utilizzati anche in questo lavoro di tesi, si ha il termostato di Berendsen (Berendsen 1991), mostrato nell'Eq. (3.1) e il barostato di Parrinello-Rhman (M. Parrinello e Rahman 1981), a sua volta descritto nelle Eq. da (3.2) a (3.5) :

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}, \quad (3.1)$$

dove T_0 e T sono rispettivamente la temperatura di riferimento e la temperatura attuale del sistema, mentre τ è una costante di tempo che dipende dalla capacità termica totale del sistema e dal numero totale di gradi di libertà.

$$\frac{d^2 \mathbf{b}}{dt^2} = V \mathbf{W}^{-1} \mathbf{b}'^{-1} (\mathbf{P} - \mathbf{P}_{ref}), \quad (3.2)$$

$$\frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - \mathbf{M} \frac{d\mathbf{r}_i}{dt}, \quad (3.3)$$

$$\mathbf{M} = \mathbf{b}^{-1} \left[\mathbf{b} \frac{d\mathbf{b}'}{dt} + \frac{d\mathbf{b}}{dt} \mathbf{b}' \right] \mathbf{b}'^{-1}, \quad (3.4)$$

$$\mathbf{W}_{i,j}^{-1} = \frac{4\pi^2 \beta_{i,j}}{3\tau_p^2 L}, \quad (3.5)$$

dove \mathbf{P} e \mathbf{P}_{ref} sono rispettivamente la pressione nel sistema in un certo istante di tempo e la pressione di riferimento. Il termine \mathbf{W} è una matrice di parametri legati alla forza calcolata a partire dalla compressibilità isoterma β (generalmente una matrice diagonale con elementi tutti uguali), dalla massima dimensione della box L e da una costante di tempo τ_p che generalmente è molto più grande rispetto a quella usata nel calcolo della temperatura, il termine \mathbf{b} rappresenta la matrice dei vettori dimensione della box. Inoltre compare un termine nel calcolo dell'equazione di Newton che corrisponde ad una forza di attrito.

3.3 Ipotesi semplificative

In tutte le simulazioni eseguite sono state assunte delle semplificazioni, alcune dovute ad una mancanza di informazioni, altre legate al fatto che le

simulazioni di dinamica molecolare e di metadinamica hanno già intrinsecamente un costo computazionale elevato e talvolta inserire certe semplificazioni può ridurre questo costo e, conseguentemente, abbassare il tempo totale di simulazione.

3.3.1 Fase non-acquosa: olio alimentare

Gli oli alimentari sono composti da una miscela di diversi trigliceridi, molecole formate da un glicerolo con tre acidi grassi sostituiti. L'olio scelto per le simulazioni è un olio di semi di soia di cui è nota la composizione degli acidi grassi in Tabella 3.1 (Pryde 1980). Poiché non è nota la composizione effettiva dei trigliceridi una prima ipotesi semplificativa è stata assumere che tutti i trigliceridi fossero composti da un unico tipo di acido grasso. Per semplicità, da questo punto in avanti i trigliceridi verranno indicati con il nome dell'acido grasso che li compone. Dai valori di composizione mas-

Tabella 3.1: Composizione dell'olio di semi di soia. Tipo = numero di carboni:numero di insaturazioni, m_i = massa di una molecola, χ_{mass} = frazione massica sperimentale (Pryde 1980), χ_{num} = frazione numerica di molecole calcolata a partire dalla frazione massica.

Trigliceride	Massa molare [g/mol]	Tipo	m_i [g]	χ_{mass} [%]	χ_{num} [%]
Palmitico	807,31	16:0	$1,34 \times 10^{-21}$	$10,75 \pm 2,75$	11,71
Stearico	891,47	18:0	$1,48 \times 10^{-21}$	$3,70 \pm 1,70$	3,65
Oleico	885,42	18:1	$1,47 \times 10^{-21}$	$23,50 \pm 6,50$	23,34
Linoleico	879,37	18:2	$1,46 \times 10^{-21}$	$53,50 \pm 5,50$	53,50
Linolenico	873,33	18:3	$1,45 \times 10^{-21}$	$7,75 \pm 3,25$	7,80

sa su massa è stata calcolata la composizione moli su moli e da questa la composizione numerica di ciascun trigliceride. Una volta ottenuta questa, è stato necessario creare i file di struttura e di topologia di ciascun trigliceride. GROMACS, infatti, non contiene informazioni sui lipidi nel suo database di molecole; quindi, i file sono stati ottenuti da un software esterno (LigPar-Gen <http://zarbi.chem.yale.edu/ligpargen/>) (Jorgensen e Tirado-Rives 2005; Dodda, Vilseck et al. 2017; Dodda, Cabeza de Vaca et al. 2017).

3.3.2 Fase acquosa

Per l'acqua è stato scelto il modello TIP3P (Transferable Intermolecular Potential with 3 Points) in cui ciascuna molecola viene rappresentata da tre punti dotati di massa, uno per ogni atomo. Il problema principale di questo

metodo riguarda il fatto che alcune proprietà, come la temperatura di fusione o la diffusività, non vengono calcolate correttamente. Esistono anche altri modelli successivi, come il modello TIP5P, in cui una molecola d'acqua viene rappresentata da cinque punti: tre dotati di massa, che rappresentano gli atomi, e due privi di massa ma dotati di carica elettrica, che vanno a rappresentare la coppia di elettroni di valenza dell'ossigeno. L'inconveniente di questo modello è che l'utilizzo di due punti in più per ogni molecola andrebbe ad appesantire incredibilmente la simulazione in termini di costo computazionale. Maggiori informazioni riguardo ai modelli per l'acqua possono essere trovati nel testo di Allen e Tildesley (2017).

3.3.3 La proteina

Anche per la proteina è stato necessario assumere diverse semplificazioni dettate dal fatto che, come già detto precedentemente, le informazioni sulla struttura sono molto scarse. La prima cosa da determinare è stata la carica dei vari amminoacidi. Ciascuna catena della proteina presenta infatti sei lisine, sei arginine, quattro asparagine e cinque acidi glutammici, oltre ai due terminali di catena, e ciascuno di questi residui può essere protonato o non protonato. Per verificare lo stato di protonazione si è ricorso ad un altro software (H++ <http://biophysics.cs.vt.edu/H++>) (Anandakrishnan, Aguilar e Onufriev 2012; Myers et al. 2006; Gordon et al. 2005) che partendo dal file pdb della proteina e dal pH, che in questo caso è pari a 3,8, restituisce la pKa di ciascun amminoacido in catena con cui poi è possibile verificare se il residuo in esame è protonato oppure no. La pKa degli amminoacidi liberi può infatti essere diversa se questi sono legati in una catena; perciò, per determinare la carica dei residui non si può semplicemente cercare la pKa di ciascun amminoacido in letteratura per confrontarlo con il pH. Nella Tabella 3.2 è rappresentato il risultato di questa analisi, per una singola catena del dipeptide, con il pKa di ciascun amminoacido e la carica totale. Si ricorda che se il pH è inferiore al pKa allora il gruppo funzionale del residuo sarà protonato, viceversa nella situazione opposta. Al termine di tutto, la carica netta è risultata essere pari a +14 per tutta la proteina.

Una volta ottenuta la carica esatta della proteina è stato necessario individuare una struttura quaternaria per non eseguire le simulazioni su una proteina non ripiegata. Per fare ciò è stato usato un software di DeepMind, AlphaFold (<https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>) (Jumper et al. 2021), in grado di predire la struttura ripiegata delle proteine. Per questo lavoro di tesi è stata utilizzata una versione base del software che non permette di simulare la struttura di un dimero, ma è possibile cercare la struttura dei due monomeri contemporaneamente. Nonostante i risultati ottenuti non fornissero una buona accuratezza, i due monomeri simulati venivano posti in una posizione tale per cui le due cisteine fossero vicine, come mostrato nelle Figure 3.3

Tabella 3.2: Carica elettrica degli amminoacidi di un singolo monomero. Ciascun residuo è individuato da: sigla del nome-posizione nella catena.

Residuo	pKa	Gruppo funzionale	Carica netta
LYS-1 (TER)	10,610	-NH ₃ ⁺	+1
LYS-1	6,849	-NH ₃ ⁺	+1
ASP-5	2,349	-COO ⁻	-1
ARG-6	>12,000	-NH ₃ ⁺	+1
GLU-7	4,013	-COOH	0
ARG-8	>12,000	-NH ₃ ⁺	+1
ARG-9	>12,000	-NH ₃ ⁺	+1
ASP-10	3,699	-COO ⁻	-1
ASP-16	3,392	-COO ⁻	-1
GLU-24	3,755	-COO ⁻	-1
LYS-28	11,824	-NH ₃ ⁺	+1
ARG-32	>12,000	-NH ₃ ⁺	+1
ASP-39	4,291	-COOH	0
ARG-50	>12,000	-NH ₃ ⁺	+1
GLU-56	3,873	-COOH	0
ARG-59	>12,000	-NH ₃ ⁺	+1
LYS-62	10,404	-NH ₃ ⁺	+1
GLU-65	3,857	-COOH	0
GLU-69	3,785	-COO ⁻	-1
LYS-70	10,595	-NH ₃ ⁺	+1
LYS-72	11,017	-NH ₃ ⁺	+1
LYS-78	10,580	-NH ₃ ⁺	+1
TYR-82 (TER)	2,200	-COO ⁻	-1
TOTALE	—	—	+7

3.4. Dal file di struttura ottenuto da AlphaFold è stato necessario avvicinare manualmente i due monomeri con un'operazione di traslazione per far sì che i due zolfi delle due cisteine si trovassero ad una distanza compresa tra 0,18 nm e 0,22 nm. Questo perché GROMACS individua la presenza di un legame solo se la distanza tra gli atomi è pari alla distanza di equilibrio (che per il ponte disolfuro vale 0,2 nm) più o meno il 10 % del valore della distanza stessa. Con la struttura finale ottenuta è stato poi possibile modificare manualmente la carica di ciascun amminoacido andando a specificare per ognuno lo stato di protonazione, seguendo i risultati ottenuti dal software H++. Con la struttura finale è poi stata individuata la dimensione minima per il lato di una box cubica per evitare che durante le simulazioni la proteina “vedesse” la sua immagine periodica.

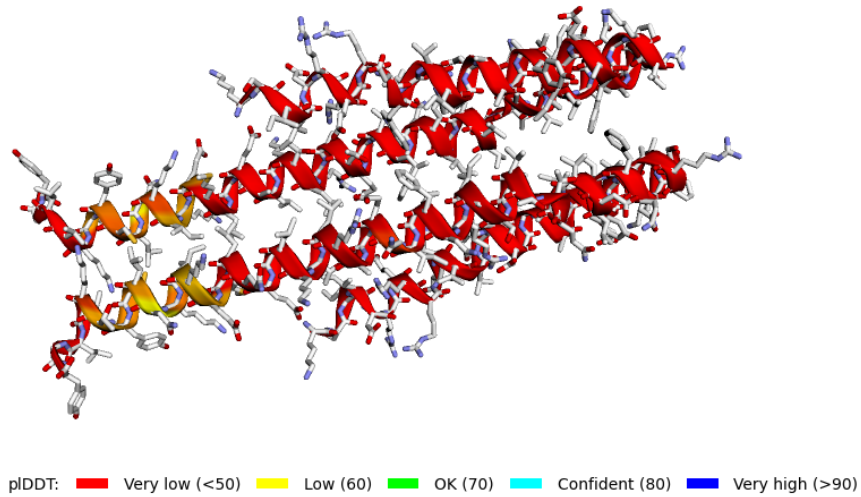


Figura 3.3: Struttura della proteina estratta dal software AlphaFold: il colore indica l'affidabilità di predizione.

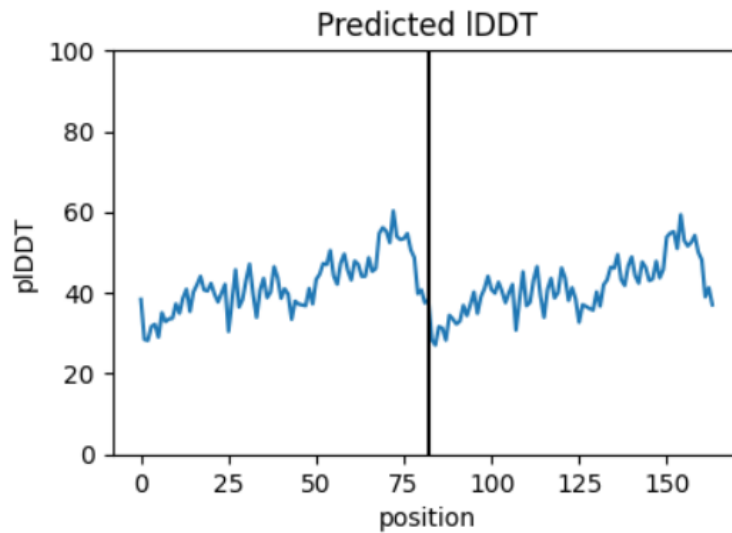


Figura 3.4: Affidabilità di predizione per ogni residuo.

3.4 Parametri della dinamica molecolare

In questa sezione verranno esposti i parametri comuni a tutte le simulazioni. Tutte le simulazioni sono state svolte alla temperatura di 298 K, ad una pressione di 1 bar. Il termostato scelto sia per le fasi di equilibratura che per le simulazioni di produzione è il Berendsen (Eq. 3.1) con temperatura di riferimento $T_0 = 298$ K e $\tau_T = 0,1$ ps e come barostato è stato usato il Parrinello-Rhman (Eq. da 3.2 a 3.5) con compressibilità isoterma uguale a quella dell'acqua e pari a $\beta = 4,5 \times 10^{-5}$ bar $^{-1}$, costante di tempo $\tau_p = 2$ ps e pressione di riferimento $P_{ref} = 1$ bar. Le PBC sono state considerate in tutte e tre le direzioni e la distanza di cut-off è stata scelta pari ad 1 nm sia per le interazioni di Coulomb che per il potenziale di Lennard-Jones e per entrambi è stata utilizzata la lista di Verlet per tenere conto di quali particelle rientrassero nella distanza di cut-off. Le interazioni elettrostatiche sono state calcolate tramite le PME. La fase di neutralizzazione della carica è stata eseguita andando a sostituire quattordici molecole di solvente con altrettanti ioni Cl $^-$ per compensare la carica della proteina. Successivamente è stata eseguita prima la minimizzazione energetica e poi l'equilibratura divisa in due fasi: prima 100 ps ($\Delta t = 2$ fs, $N_{steps} = 50000$) con un *ensemble* NVT per portare la temperatura a 298 K e poi 1 ns ($\Delta t = 2$ fs, $N_{steps} = 500000$) con un *ensemble* NpT per stabilizzare la densità. La scelta di questi due tempi è stata giustificata a posteriori andando a visualizzare graficamente l'andamento nel tempo della temperatura e della densità per controllare che le variazioni fossero contenute e che il valore delle due grandezze fosse circa costante.

3.5 Parametri della Metadinamica

Tutte le simulazioni di metadinamica sono state svolte in seguito alla fase di equilibratura con la tecnica della WTMetaD. Le variabili collettive scelte sono state:

- Minima distanza della proteina dalla fase oleosa.
- Raggio di giro della proteina.

Per ridurre il costo computazionale, la distanza minima non è stata calcolata considerando tutti gli atomi della proteina rispetto a tutti gli atomi di olio ma per la proteina è stato considerato il COM, mentre la fase olio è stata individuata scegliendo come punto di riferimento un atomo di carbonio che appartenesse a tutte le molecole di trigliceridi ma non alla proteina o ai fosfolipidi. GROMACS individua ciascun atomo con un codice identificativo, quindi è stato necessario verificare che l'atomo di carbonio scelto come riferimento per rappresentare la fase oleosa, identificato col codice C01, avesse un codice identificativo uguale in tutti i trigliceridi ma diverso da tutti gli

altri carboni nel sistema. Nel grafico di Figura 3.5 si può vedere la distribuzione dei carboni C01 rispetto alla distribuzione di tutti gli atomi dell'olio per giustificare l'approssimazione fatta. Per la scelta delle CV fatte, è stato

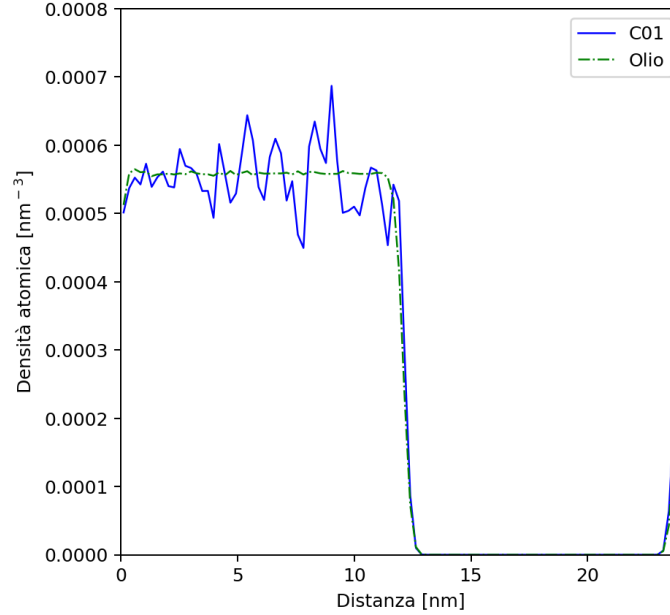


Figura 3.5: Distribuzione normalizzata rispetto al numero totale di atomi dell'olio lungo l'asse Y della box.

inserito un bias bidimensionale dinamico. La bidimensionalità è dovuta al fatto che è applicato un valore di bias diverso per ogni CV, la dinamicità è dovuta al fatto che, secondo la tecnica della MetaD, il bias è applicato con delle gaussiane successive e perciò non è costante. I parametri da specificare sono quindi: l'altezza iniziale delle gaussiane, l'ampiezza (scelta solitamente come il 35-40% della variazione della CV nella simulazione MD), il fattore di bias (già visto nell'Eq. 2.35)) e il numero di passi tra la deposizione di una gaussiana e la successiva. Nella tabella 3.3 sono presenti i parametri computazionali scelti per ogni CV.

Tabella 3.3: Parametri Plumed delle simulazioni di metadinamica.

	Raggio di giro	Distanza minima
Altezza gaussiana	1	1
Ampiezza gaussiana	0,04	0,5
Numero di passi	500	500
Fattore di bias	10	10

3.6 Dettagli delle simulazioni

Di seguito, è riportata la Tabella 3.4 riassuntiva di tutti i casi simulati che verranno descritti e dettagliati successivamente:

Tabella 3.4: Descrizione delle simulazioni.

Caso	Descrizione	MD	MetaD
1	Proteina nel bulk dell'acqua.	✓	
2	Proteina nel bulk dell'olio.	✓	
3	Sistema bifasico con proteina nel bulk dell'acqua.	✓	✓
4	Sistema bifasico con proteina e fosfolipidi nel bulk dell'acqua con concentrazione pari a quello nel tuorlo d'uovo.		✓
5	Sistema bifasico con proteina nel bulk dell'acqua e fosfolipidi all'interfaccia con concentrazione pari a metà della saturazione.		✓
6	Sistema bifasico con proteina nel bulk dell'acqua e fosfolipidi all'interfaccia con concentrazione di saturazione.		✓

3.6.1 Caso 1

La box della proteina in acqua è sicuramente il caso più semplice da simulare. Una volta ottenuta la struttura corretta della proteina questa è stata posta al centro di una box di dimensioni $X = Y = Z = 12,00$ nm a cui poi è stata aggiunta l'acqua (modello TIP3P). Quattordici molecole d'acqua casuali sono state sostituite da altrettanti ioni Cl^- prima di eseguire la minimizzazione energetica e l'equilibratura. Successivamente è stata effettuata una simulazione di 100 ns. Al termine dei 100 ns sono state condotte altre dieci simulazioni di 10 ns ciascuna, che avevano tutte come posizione iniziale degli atomi la posizione ottenuta dopo 100 ns ma con la differenza che la velocità iniziale è stata scelta casualmente dalla distribuzione di Maxwell-Boltzman a 298 K. In questo modo è stato poi possibile eseguire una media d'insieme per il calcolo di alcune proprietà come la diffusività, sfruttando il teorema ergodico.

3.6.2 Caso 2

Per il caso della proteina in olio la proteina è stata inserita in una box di dimensioni $X = Y = Z = 25,00$ nm. Successivamente è stato inserito un numero di trigliceridi che rispettasse le concentrazioni di Tabella 3.1. La dimensione maggiore della box rispetto al caso 1 è stata scelta per poter garantire lo spazio per l'inserimento delle molecole di olio senza che vi fosse sovrapposizione di atomi, lasciando quindi dello spazio vuoto tra le varie molecole. In seguito all'annullamento della carica, in cui sono state rimosse quattordici molecole di trilinoleato di glicerile per far posto agli ioni Cl^- , è stata eseguita la minimizzazione energetica e l'equilibratura. In particolare, durante l'equilibratura della densità, è stato rimosso lo spazio vuoto all'interno della box e questo ha ridotto le dimensioni a $X = Y = Z = 12,18$ nm, molto simili al caso precedente. Infine, è stata condotta la simulazione di produzione con la stessa procedura del caso 1: una simulazione di 100 ns e dieci simulazioni di 10 ns ciascuna con velocità scelta casualmente dalla distribuzione di Maxwell-Boltzman a 298 K.

3.6.3 Caso 3

Per il terzo caso la costruzione della box è avvenuta in due passaggi: nel primo è stata costruita una box di olio nello stesso modo descritto nel Caso 2 ma senza la proteina al centro. Questa è poi stata equilibrata (non è stato necessario inserire ioni dato che i trigliceridi non sono carichi elettricamente) per portare le dimensioni a $X = Y = Z = 12,15$ nm. Nel secondo passaggio, la dimensione Y della box è stata raddoppiata, senza modificare le posizioni dei trigliceridi, ed è stata inserita la proteina al centro dello spazio vuoto che si è creato. Infine, tutta la box è stata solvatata con acqua. Gli ioni Cl^- in questo caso sono andati a sostituirsi con le molecole d'acqua. Dopo la minimizzazione energetica è stata svolta l'equilibratura che ha avuto come scopo, oltre a stabilizzare temperatura e densità, quello di separare correttamente la fase oleosa dalla fase acquosa, soprattutto in prossimità dell'interfaccia. Infine, sono state lanciate due diverse simulazioni:

- Caso 3a: simulazione MD di 300 ns.
- Caso 3b: simulazione MetaD di 50000000 steps.

3.6.4 Caso 4

Per la creazione della box del Caso 4 è stato sufficiente partire dalla box del Caso 3 e sostituire parte delle molecole d'acqua con dieci molecole di fosfolipide POPC. Questo numero è stato scelto sulla base della composizione del tuorlo d'uovo. La composizione del POPC è nota da letteratura (Ali et al. 2017), per la proteina si è assunto che tutte le proteina facenti parte delle LDL fossero Apovitellenina-1 (Anton 2007). Sapendo che la massa

della proteina è di circa 9331 Da per monomero e che quella del POPC è di circa 760 Da si è potuto passare dalla frazione massica al rapporto numerico. Nell'Eq. (3.6) sono riportati i calcoli svolti per trovare il numero di molecole di POPC per molecola di proteina:

- Frazione di LDL nel tuorlo d'uovo secco: $\chi_{LDL} = 0,66$
- Frazione di proteine nel LDL: $\chi_P = 0,10$
- Frazione di lipidi nel tuorlo d'uovo secco: $\chi_L = 0,63$
- Frazione di fosfolipidi nei lipidi: $\chi_{FL} = 0,33$
- Frazione di fosfoditilcoline nei fosfolipidi: $\chi_{PC} = 0,76$
- Frazione di POPC nelle fosfoditilcoline: $\chi_{POPC} = 0,154$
- Massa proteina: $m_P = 18662$ Da
- Massa POPC: $m_{POPC} = 760$ Da

$$\frac{N_{POPC}}{N_P} = \frac{\chi_L \times \chi_{FL} \times \chi_{PC} \times \chi_{POPC}}{\chi_{LDL} \times \chi_P} \times \frac{m_P}{m_{POPC}} = 9,05 \approx 10 \% \quad (3.6)$$

Poiché i fosfolipidi hanno complessivamente carica nulla non è stato necessario aggiungere o rimuovere ioni Cl^- . Una volta completata la box, e dopo aver minimizzato l'energia ed equilibrato temperatura e densità, sono state eseguite due simulazioni:

- Caso 4a: simulazione MD di 100 ns.
- Caso 4b: simulazione MetaD di 50000000 steps.

3.6.5 Caso 5

Anche per il Caso 5 la box di partenza è stata quella già equilibrata utilizzata nel Caso 3. Questa box presenta l'interfaccia XZ ricoperta da fosfolipidi in numero tale per cui la concentrazione fosse il 50 % della concentrazione di saturazione. Il valore di quest'ultima è pari a $0,635 \text{ nm}^2/\text{Lipide}$ (Jämbeck e Lyubartsev 2012) e, note le dimensioni della box pari a $X = 0,5Y = Z = 11,70 \text{ nm}$ si è ottenuto un numero di circa 108 molecole di fosfolipide per ogni interfaccia. Questi sono stati sostituiti in parte sia all'acqua che all'olio. Successivamente sono stati eseguiti la minimizzazione energetica, l'equilibratura e due simulazioni di produzione:

- Caso 5a: simulazione MD di 100 ns.
- Caso 5b: simulazione MetaD di 50000000 steps.

3.6.6 Caso 6

Per quest'ultimo caso è stata simulata una box con l'interfaccia completamente ricoperta di fosfolipidi. Partendo sempre dalla box del Caso 3, sono stati inserite 216 molecole di fosfolipide per interfaccia, in modo da riprodurre le condizioni di saturazione citate nel Caso 5. Anche in questo caso, dopo le fasi di minimizzazione energetica ed equilibratura, sono state lanciate due simulazioni:

- Caso 6a: simulazione MD di 100 ns.
- Caso 6b: simulazione MetaD di 50000000 steps.

3.7 Tabella riassuntiva

Nelle tabelle seguenti viene riassunto il contenuto molecolare di ogni simulazione (Tabella 3.5) e la composizione dell'olio in ogni box confrontata con quella sperimentale (Tabella 3.6):

Tabella 3.5: Contenuto numerico di molecole di ogni caso simulato.

Caso	1	2	3	4	5	6
Acqua	55690	0	55845	55065	42619	42619
Palmitico	0	132	132	132	90	90
Stearico	0	42	42	42	30	30
Oleico	0	263	263	263	172	172
Linoleico	0	588	602	602	452	452
Linolenico	0	88	88	88	61	61
POPC	0	0	0	10	216	432
Cl ⁻	14	14	14	14	14	14

Tabella 3.6: Composizione in percentuale massica [%] dell'olio in ogni caso simulato.

Trigliceride	Caso					Sperimentale
	2	3	4	5	6	
Palmitico	10,98	10,84	10,84	10,34	10,34	10,75±2,75
Oleico	3,86	3,81	3,81	3,81	3,81	3,70±1,70
Stearico	23,99	23,69	23,69	21,68	21,68	23,50±6,50
Linoleico	53,26	53,85	53,85	56,58	56,58	53,50±5,50
Linolenico	7,92	7,82	7,82	7,58	7,58	7,75±3,25

3.8 Post-processing

In questa sezione verranno descritti tutti i parametri analizzati al termine delle varie simulazioni per poi presentare, nel prossimo capitolo, i risultati ottenuti. Si ricorda che, per le simulazioni di metadinamica è necessario considerare il fatto che sia stato applicato un bias, applicando quindi l'Eq. (2.42).

3.8.1 Il raggio di giro

Il raggio di giro di un corpo attorno ad un asse di rotazione è definito come la distanza radiale da un punto che avrebbe un momento di inerzia uguale alla distribuzione effettiva della massa del corpo, se la massa totale del corpo fosse concentrata in quel singolo punto. Avendo un sistema di N molecole, ciascuna con massa m_i , il raggio di giro può essere calcolato con la formula in Eq. (3.7) e si può considerare come un indice di quanto la massa è distribuita attorno ad un asse di rotazione. Nel caso di una proteina, il raggio di giro si potrà intendere come una stima di quanto la proteina è denaturata, maggiore sarà il raggio di giro e maggiore sarà la denaturazione.

$$R_g(t) = \sqrt{\frac{\sum_{i=1}^N m_i ||\mathbf{r}_i(t)||^2}{\sum_{i=1}^N m_i}}. \quad (3.7)$$

3.8.2 RMSD

La radice della deviazione quadratica media (dall'inglese *Root Mean Square Deviation*, o RMSD) è una misura della distanza media tra gli atomi di una proteina rispetto ad una struttura di riferimento. Minore è il valore dell'RMSD e minore sarà la variazione di configurazione strutturale rispetto alla proteina di riferimento. A differenza di altre proprietà, non viene calcolata su tutti gli atomi che compongono la proteina ma solo sullo scheletro (atomi C, N, O, S, etc.), vengono cioè esclusi gli idrogeni. Può essere calcolata dall'Eq. (3.8) e anche questa può essere interpretata come un indice di stabilità della struttura.

$$RMSD(t) = \sqrt{\frac{1}{M} \sum_{i=1}^N m_i ||\mathbf{r}_i(t) - \mathbf{r}_i(t_0)||^2}. \quad (3.8)$$

Nell'equazione precedente $M = \sum_{i=1}^N m_i$ rappresenta la massa totale della molecola di cui si vuole calcolare l'RMSD e N rappresenta il numero di atomi che compongono la molecola. il termine $\mathbf{r}_i(t_0)$ indica le posizioni di una struttura di riferimento che può essere rappresentata o dalla struttura all'inizio della simulazione ($t_0 = 0$) o, come in questo caso, dalla struttura ottenuta dopo la minimizzazione energetica.

3.8.3 Distanza testa-coda

La distanza testa-coda, o *end-to-end* (EtE), rappresenta la distanza tra due atomi o gruppi specifici al variare del tempo. Nelle simulazioni eseguite è stata calcolata tra il C terminale della lisina e il N terminale della tirosina di uno stesso monomero, come riportato dall'Eq. 3.9:

$$d_{EtE}(t) = \|\mathbf{r}_{LYS}(t) - \mathbf{r}_{TYR}(t)\|. \quad (3.9)$$

3.8.4 MSD e diffusività

Lo spostamento quadratico medio (in inglese Mean Square Displacement, MSD) può essere inteso come la porzione di spazio esplorata da una molecola nel tempo. Esiste una relazione, detta relazione di Einstein che lega lo spostamento quadratico medio alla diffusività di una molecola secondo l'Eq. 3.10.

$$\lim_{t \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \|\mathbf{r}_i(t) - \mathbf{r}_i(t_0)\|^2 = 6D_A t \quad (3.10)$$

L'argomento del limite è il MSD per una molecola composta da N atomi. Per ottenere il valore della diffusività si può quindi dividere per sei la pendenza della funzione MSD, che ad infinito è una retta passante per l'origine.

3.8.5 Area superficiale accessibile al solvente

L'area superficiale accessibile al solvente o SASA (dall'inglese *Solvent Accessible Surface Area*) è un indice di quanto una data molecola sia affine al solvente. Nel caso di una proteina con un carattere maggiormente apolare, se viene immersa in un solvente polare, la SASA avrà un valore minore poiché la proteina tenderà ad agglomerarsi per poter esporre la minor superficie possibile verso un solvente a cui non è affine, viceversa se la stessa proteina viene immersa in un solvente apolare. Per calcolare la SASA si utilizza un particolare algoritmo detto DCLM (*Double Cubic Lattice Method*) in cui si crea una rete cubica nelle tre direzioni in modo tale che ogni cubo abbia lato di $2r_{i,max}$, dove $r_{i,max}$ rappresenta la somma dei raggi di van der Waals del solvente e dell'atomo i di interesse. In questo modo, fissata la cella dell'atomo i , esso sarà considerato "vicino" solo agli atomi di solvente che si trovano nelle celle adiacenti a quella fissata. Per il calcolo effettivo dell'area si crea poi una maglia di punti ad una distanza fissata dagli atomi della molecola di riferimento che si trovano vicini al solvente (solitamente la distanza è scelta come il raggio di van der Waals di un atomo più il raggio di una molecola d'acqua, cioè circa 1,4 Å). La SASA sarà quindi l'area occupata dalla maglia di punti. Maggiori informazioni sull'algoritmo e sul metodo DCLM possono essere trovati nell'articolo di Eisenhaber et al. (1995).

Capitolo 4

Analisi dei risultati

In questo capitolo verranno presentati i risultati delle simulazioni. Per comodità ci si riferirà a ciascuna simulazione con il riferimento indicato in Tabella 3.4. Nella Tabella 4.1 è possibile vedere un indice dei grafici che saranno presentati nei paragrafi successivi.

Tabella 4.1: Indice delle analisi dei risultati.

	Raggio di Giro	RMSD	SASA	EtE	MSD	Distanza Minima
Caso 1	✓	✓	✓	✓	✓	
Caso 2	✓	✓	✓	✓	✓	
Caso 3	✓	✓	✓	✓	✓	✓
Caso 4	✓					✓
Caso 5	✓					✓
Caso 6	✓					✓

4.1 Caso 1

Per il Caso 1 sono stati analizzati il raggio di giro (Figura 4.1), il RMSD (Figura 4.2), la SASA (Figura 4.3), la distanza EtE (Figura 4.4) e il MSD. In particolare, per quest'ultimo è riportato sia il calcolo su tutta la traiettoria della simulazione (Figura 4.5), sia una media d'insieme dei risultati ottenuti dalle ultime dieci simulazioni (Figure 4.6 e 4.7).

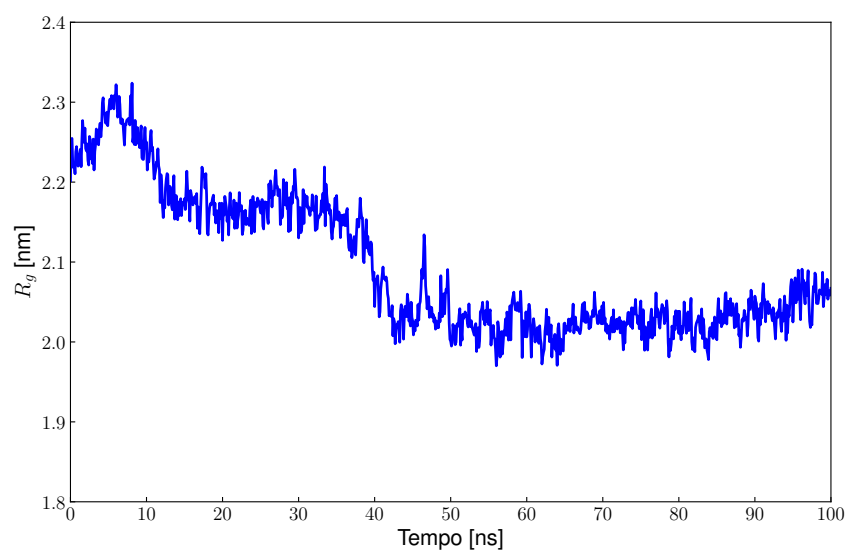


Figura 4.1: Caso 1: raggio di giro.

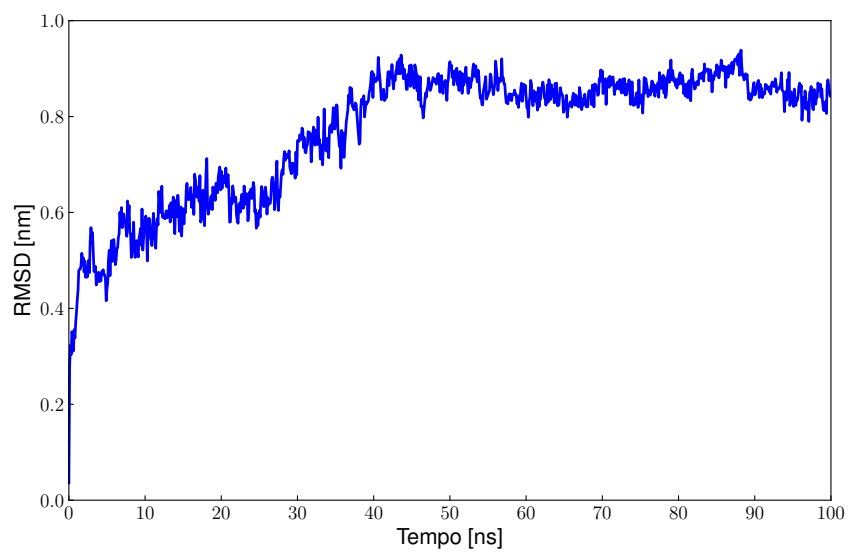


Figura 4.2: Caso 1: RMSD.

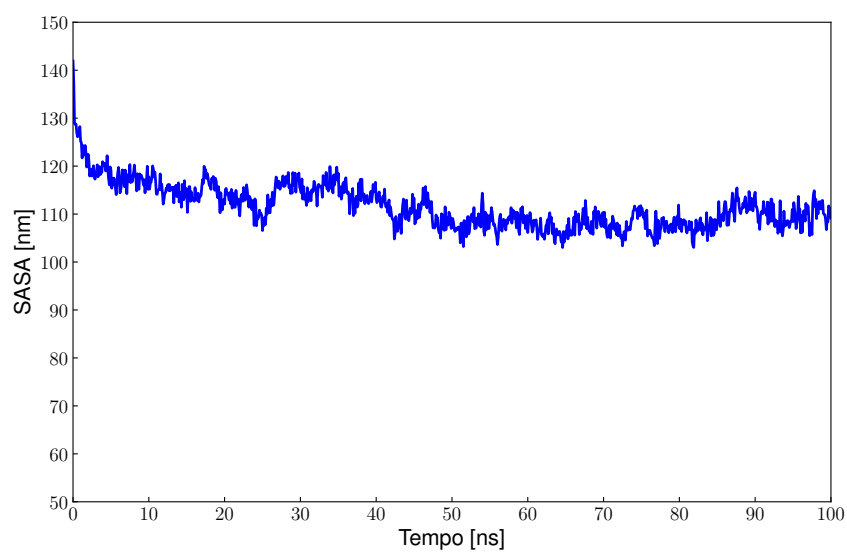


Figura 4.3: Caso 1: SASA.

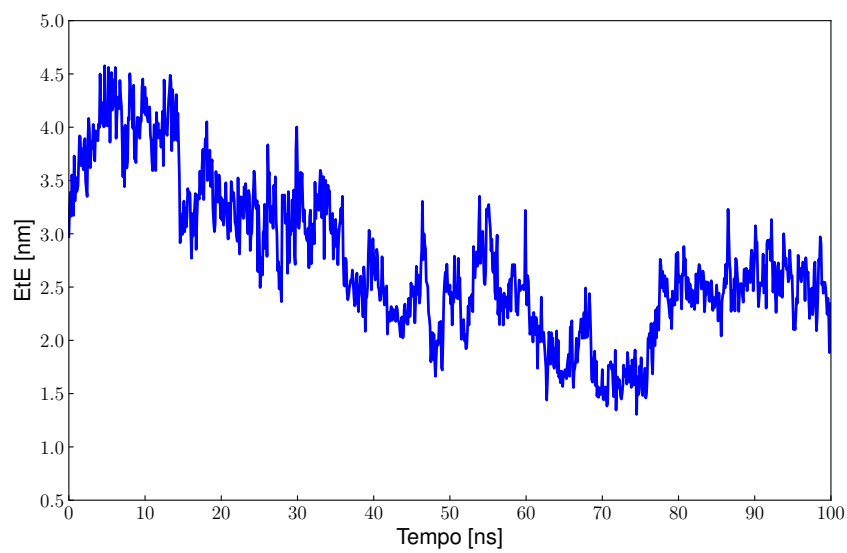


Figura 4.4: Caso 1: distanza EtE.

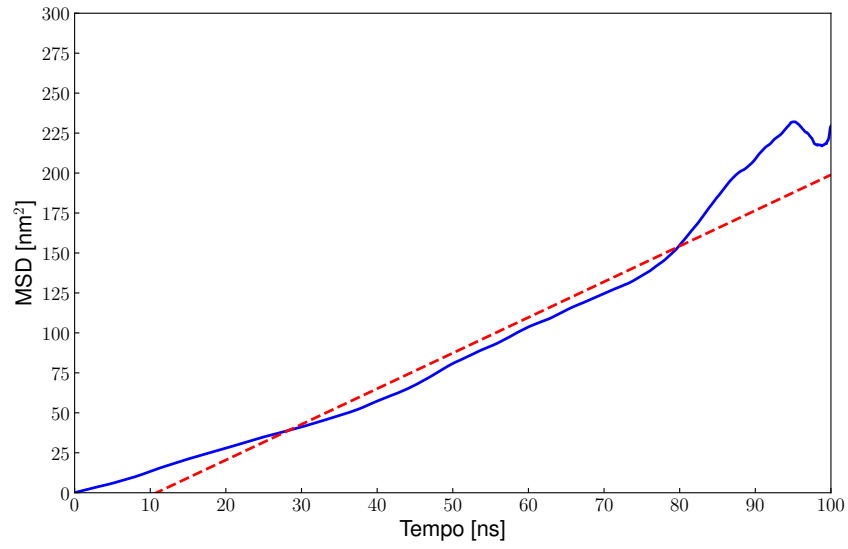


Figura 4.5: Caso 1: MSD della traiettoria di 100 ns (linea blu continua) e regressione lineare (linea rossa tratteggiata).

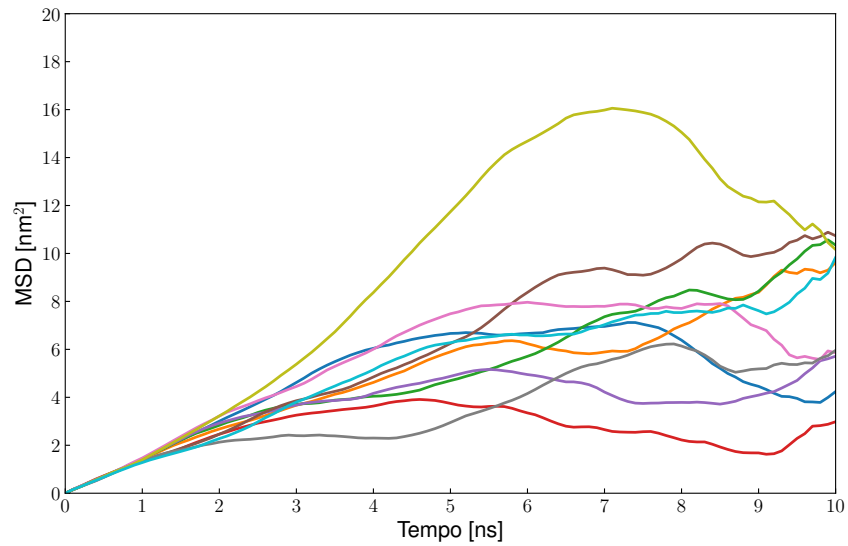


Figura 4.6: Caso 1: singoli MSD delle dieci simulazioni di 10 ns.

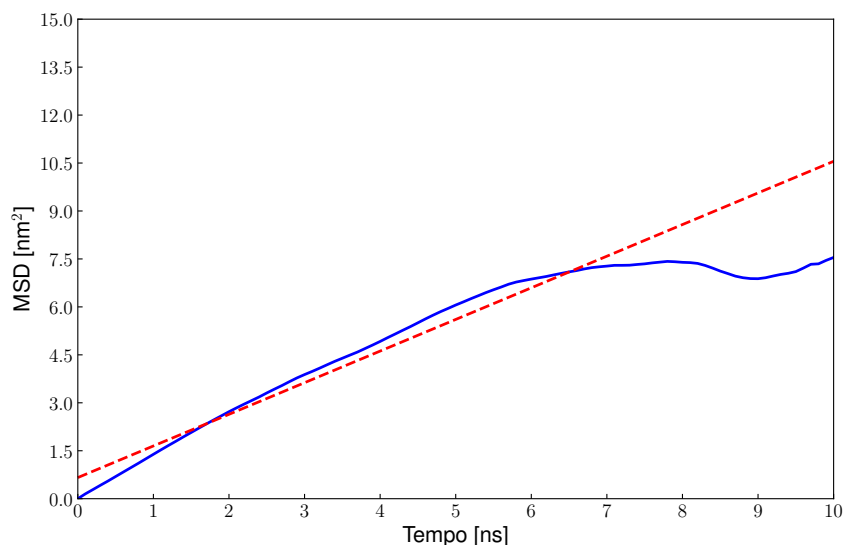


Figura 4.7: Caso 1: MSD medio delle ultime dieci simulazioni (linea blu continua) e regressione lineare (linea rossa tratteggiata).

Dai grafici precedenti si può notare l'instabilità della proteina in acqua. Poiché raggio di giro, distanza EtE e SASA diminuiscono mentre il RMSD, facendo intendere che la proteina cerca di adattare la sua forma ad un solvente a cui non è affine. Nella Tabella 4.2 è presente un riassunto dei risultati con il valore medio che questi raggiungono, calcolato solo nell'ultimo 30 % della simulazione con la relativa deviazione standard ed errore. Oltretutto, dai grafici del MSD è stato possibile calcolare la diffusività dell'Apovitellenina-1 in acqua secondo la formula di Einstein dell'Eq. 3.10 (Tabella 4.3). Anche se questo valore non è confrontabile con dati sperimentali è presumibilmente errato dal momento che GROMACS calcola in modo errato il coefficiente di autodiffusione e la viscosità dell'acqua, come riportato in Tabella 4.4.

Tabella 4.2: Caso 1: riassunto dei risultati.

Caso 1	Valore Medio	Deviazione Standard	Errore [%]
Raggio di Giro [nm]	2,03	$2,19 \times 10^{-2}$	1,08
RMSD [nm]	$8,64 \times 10^{-1}$	$2,56 \times 10^{-2}$	2,97
SASA [nm ²]	108,88	2,49	2,28
EtE [nm]	2,33	$4,16 \times 10^{-1}$	17,86

Tabella 4.3: Diffusività in acqua.

Caso 1	D [m ² /s]	R ²
100 ns	$3,72 \times 10^{-10}$	0,95
Media d'insieme	$1,65 \times 10^{-10}$	0,98

Tabella 4.4: Coefficiente di autodiffusione e viscosità dell'acqua, simulato e sperimentale.

	GROMACS	Sperimentale
Diffusività [m ² /s]	$4,41 \times 10^{-9}$	$2,57 \times 10^{-9}$
Viscosità [Pa s]	$3,14 \times 10^{-4}$	$8,94 \times 10^{-4}$

4.2 Caso 2

Per il Caso 2 sono stati analizzate le stesse proprietà già viste per il Caso 1 in modo da poter avere un confronto tra la proteina in solvente polare e apolare. Quindi sono stati calcolati: il raggio di giro (Figura 4.8), il RMSD (Figura 4.9), la SASA (Figura 4.10), la distanza EtE (Figura 4.11) e il MSD. Anche in questo caso, per il MSD, è riportato sia il calcolo su tutta la traiettoria della simulazione (Figura 4.12), sia una media d'insieme dei risultati ottenuti dalle ultime dieci simulazioni (Figure 4.13 e 4.14). Si può notare che, applicando il teorema ergodico, il calcolo risulta più accurato.

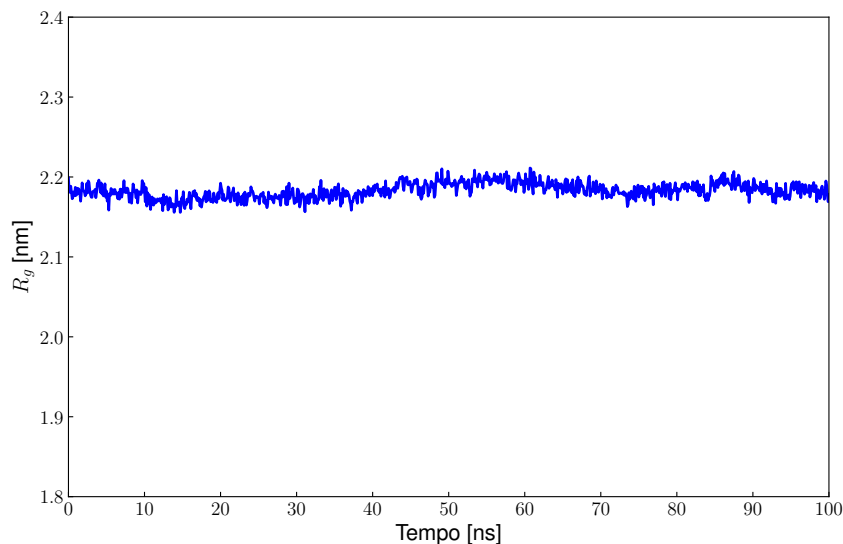


Figura 4.8: Caso 2: raggio di giro.

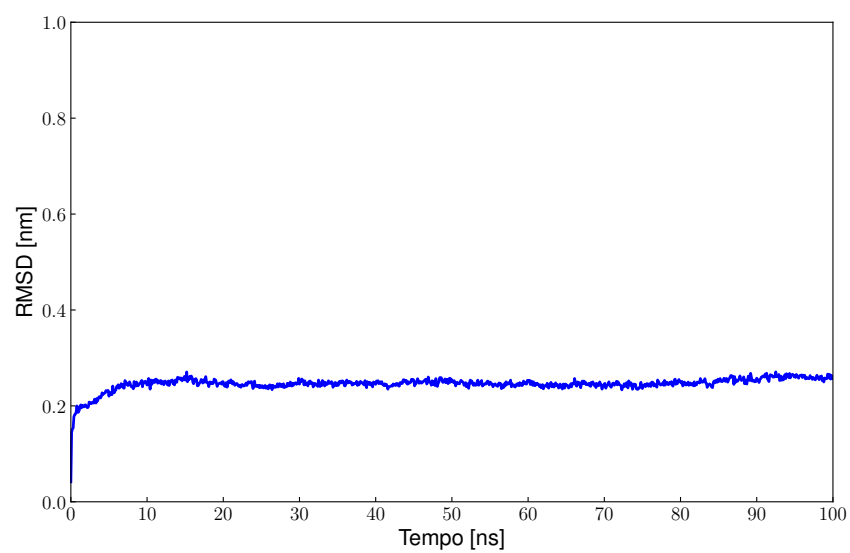


Figura 4.9: Caso 2: RMSD.

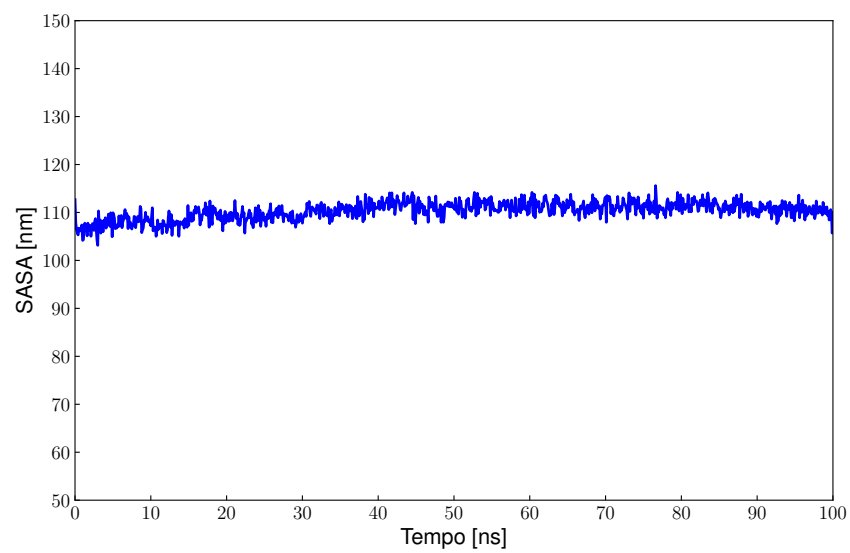


Figura 4.10: Caso 2: SASA.

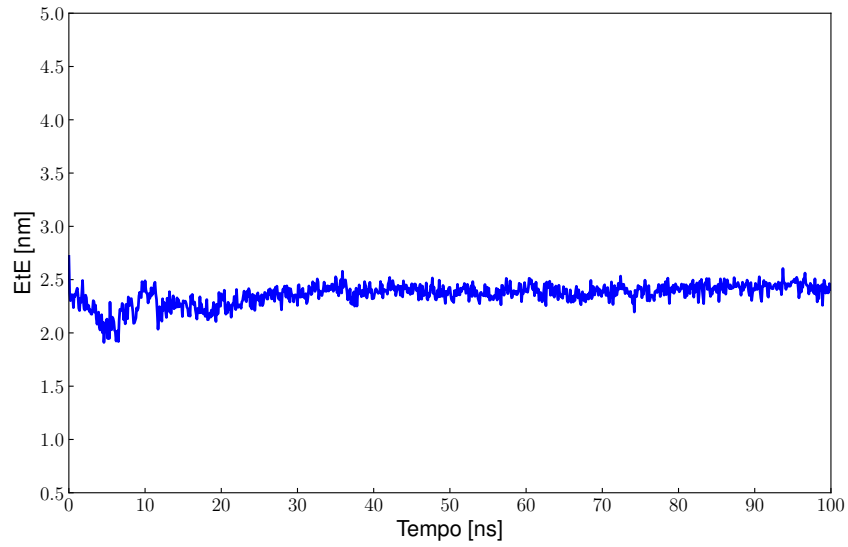


Figura 4.11: Caso 2: distanza EtE.

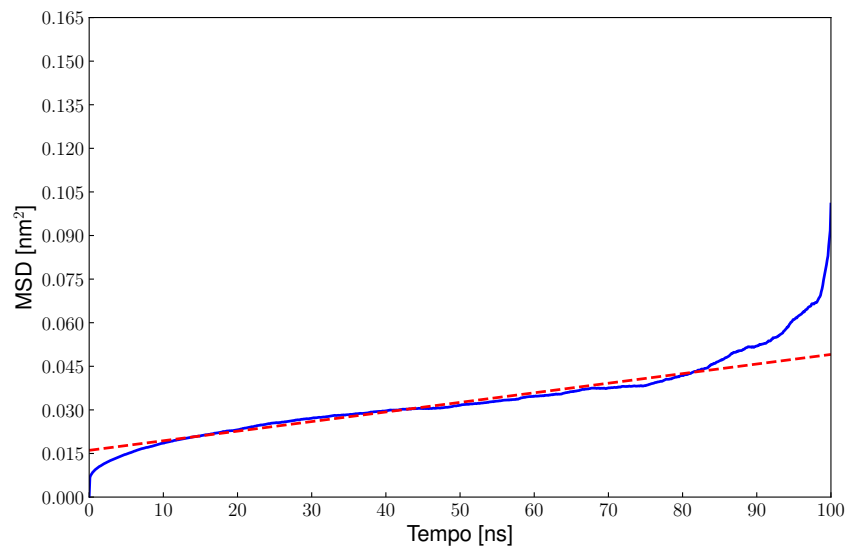


Figura 4.12: Caso 2: MSD della traiettoria di 100 ns (linea blu continua) e regressione lineare (linea rossa tratteggiata).

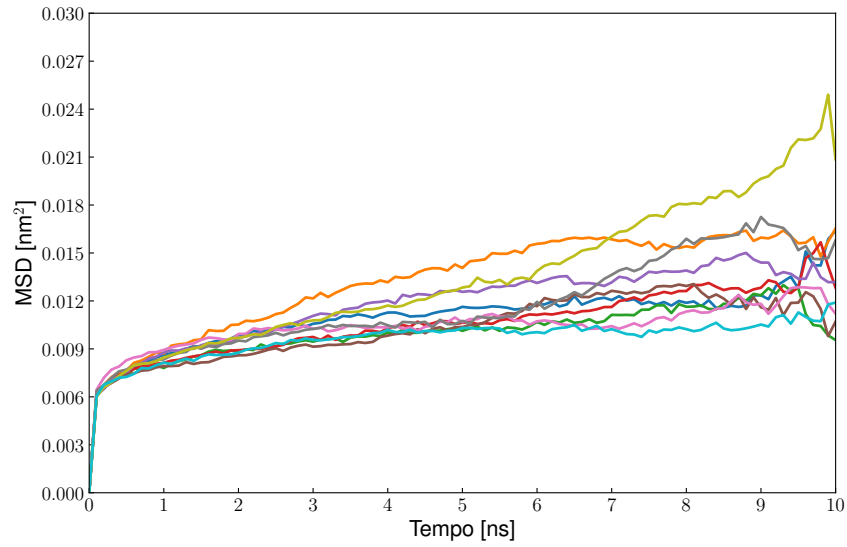


Figura 4.13: Caso 2: singoli MSD delle dieci simulazioni di 10 ns.

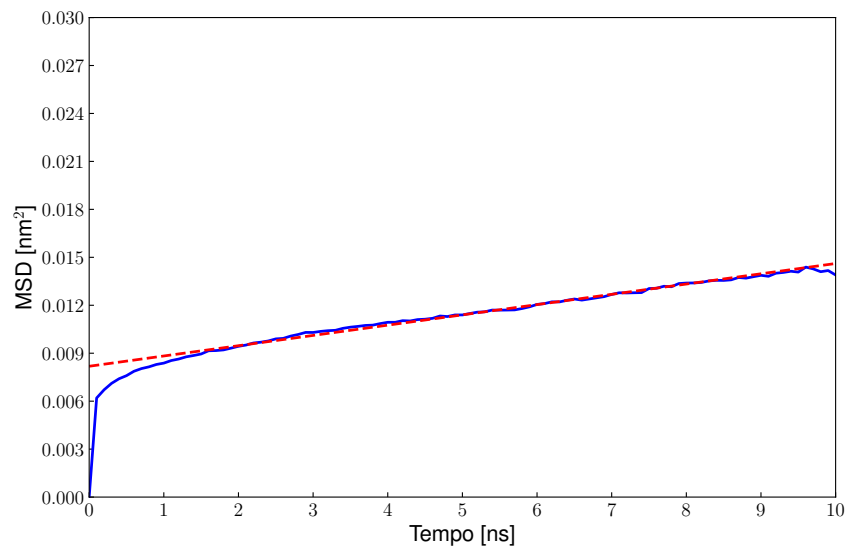


Figura 4.14: Caso 2: MSD medio delle ultime dieci simulazioni (linea blu continua) e regressione lineare (linea rossa tratteggiata).

In questo secondo caso la proteina appare più stabile nel solvente apolare. Tutte le proprietà analizzate risultano all'incirca costanti per tutta la durata della simulazione, l'Apovitellenina-1 risulta quindi più affine alla fase oleosa piuttosto che a quella acquosa. Anche in questo caso, sono riportati nelle Tabelle 4.5 e 4.6 i risultati ottenuti.

Tabella 4.5: Caso 2: riassunto dei risultati.

Caso 2	Valore Medio	Deviazione Standard	Errore
Raggio di Giro [nm]	2,18	$7,72 \times 10^{-3}$	0,35
RMSD [nm]	$2,52 \times 10^{-1}$	$8,9 \times 10^{-3}$	3,21
SASA [nm ²]	111,08	1,34	1,21
EtE [nm]	2,42	$5,68 \times 10^{-2}$	2,35

Tabella 4.6: Diffusività in olio.

Caso 1	D [m ² /s]	R ²
100 ns	$5,5 \times 10^{-14}$	0,95
Media d'insieme	$1,07 \times 10^{-13}$	0,99

4.3 Caso 3

Per il terzo caso sono stati analizzate le stesse proprietà già viste per i due casi precedenti ad esclusione del MSD. In particolare, sono stati calcolati: il raggio di giro (Figura 4.15), il RMSD (Figura 4.16), la SASA (Figura 4.17) e la distanza EtE (Figura 4.18). In aggiunta è stata analizzata la distanza minima tra la fase oleosa e la proteina (Figura 4.19) per individuare il tempo a cui questa si è assorbita. Inoltre, dalle simulazioni di metadinamica, sono stati ottenuti i profili di energia in funzione della distanza dall'interfaccia (Figura 4.20) e del raggio di giro (Figura 4.22) con le relative probabilità calcolate secondo l'Eq. (2.1) (Figure 4.21 e 4.23).

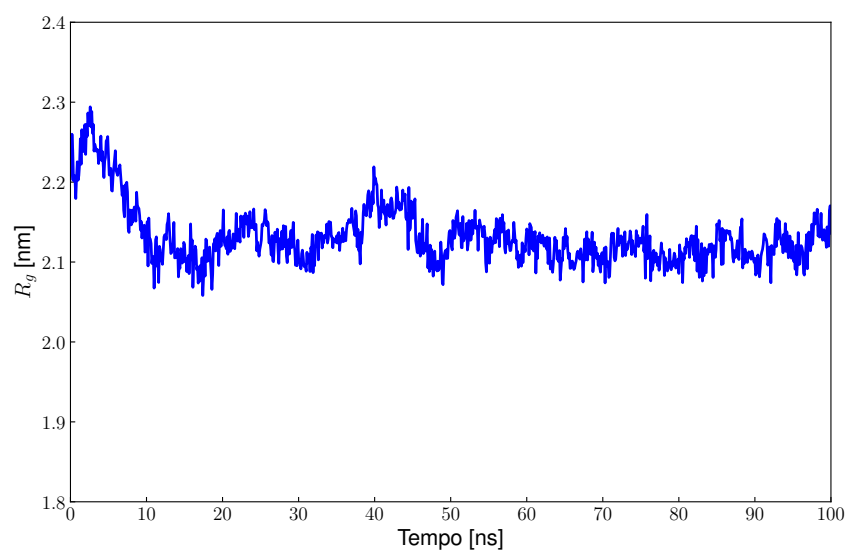


Figura 4.15: Caso 3: raggio di giro.

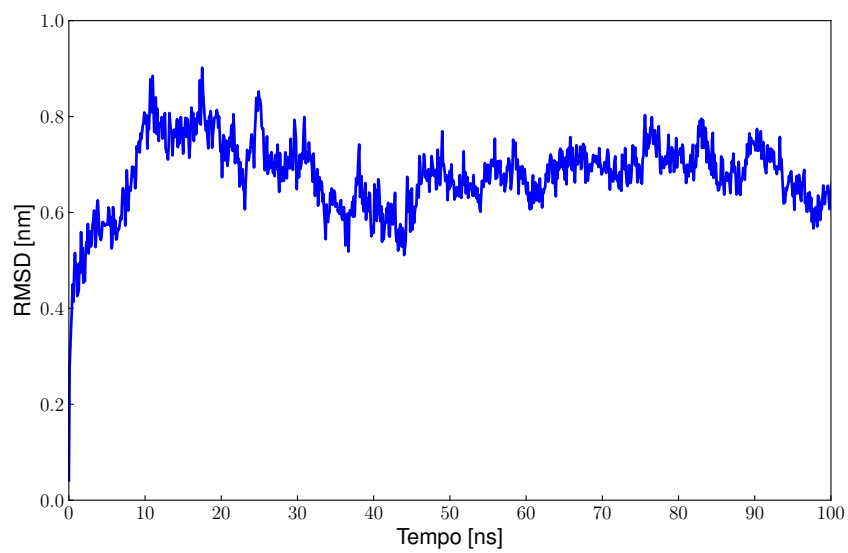


Figura 4.16: Caso 3: RMSD.

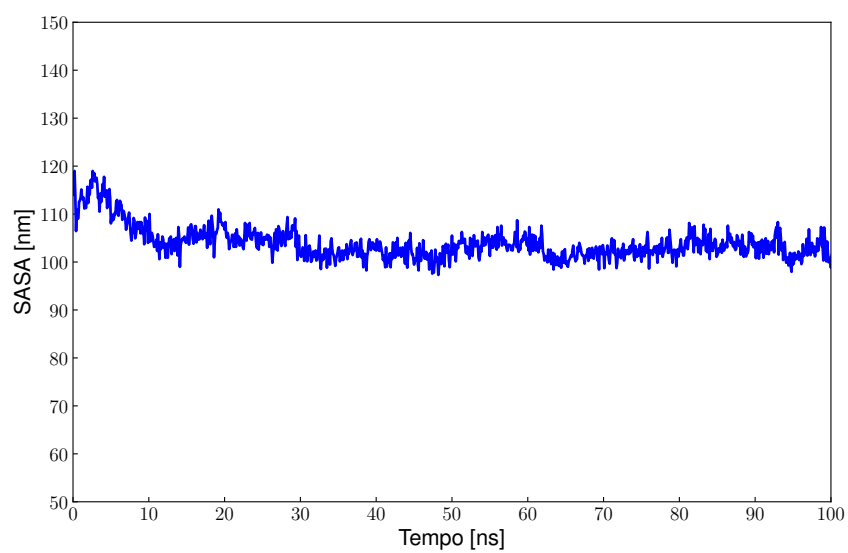


Figura 4.17: Caso 3: SASA.

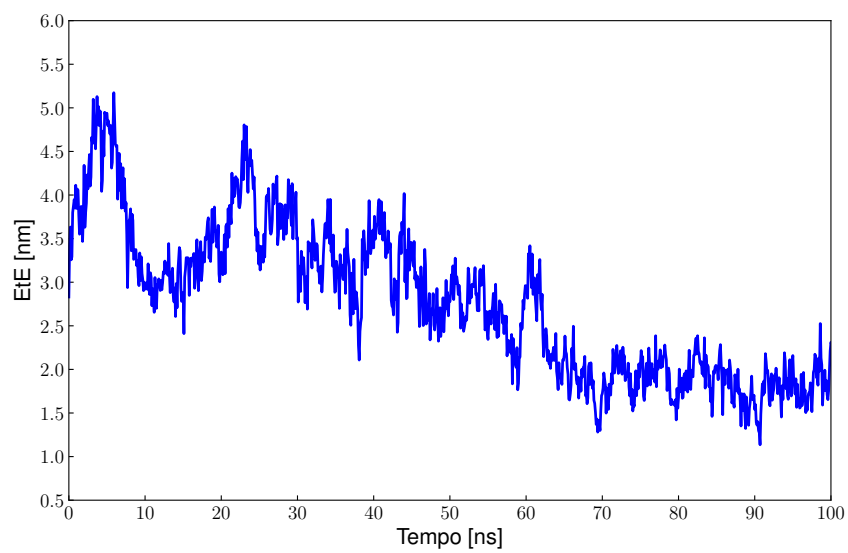


Figura 4.18: Caso 3: distanza EtE.

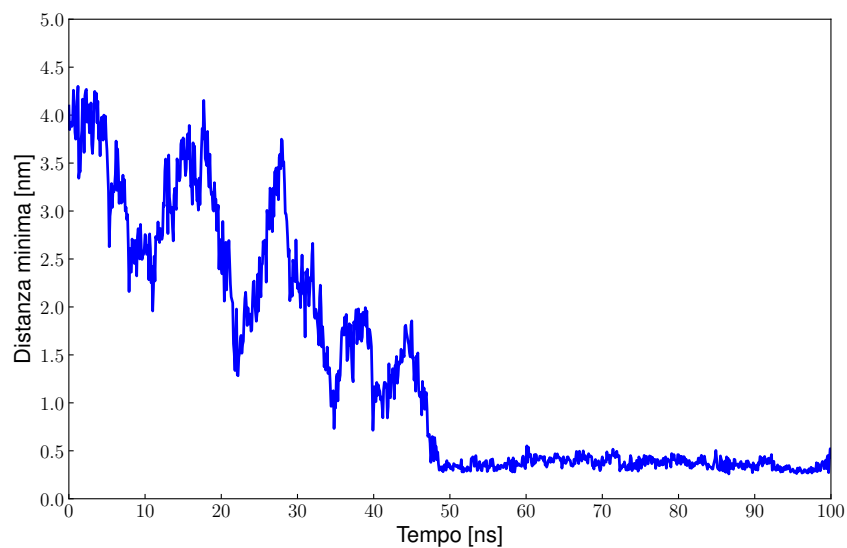


Figura 4.19: Caso 3: distanza minima tra la fase oleosa e la proteina.

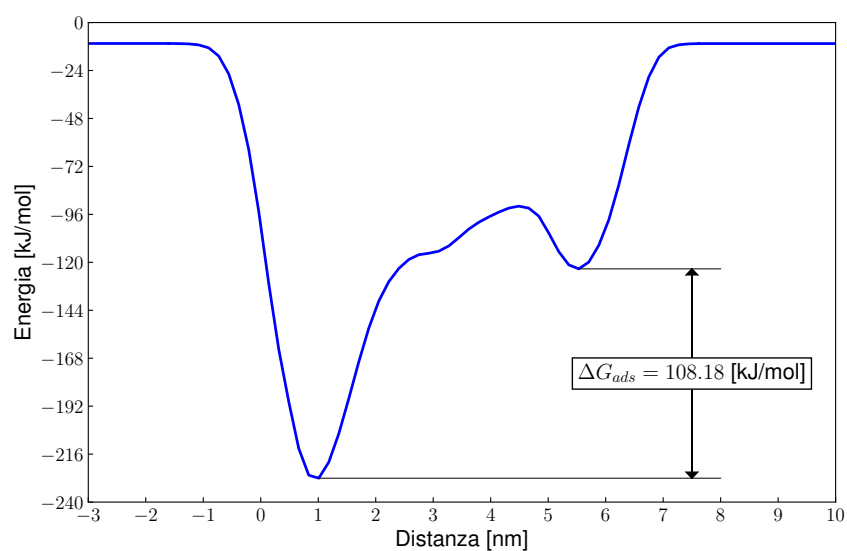


Figura 4.20: Caso 3: profilo di energia al variare della posizione della proteina.

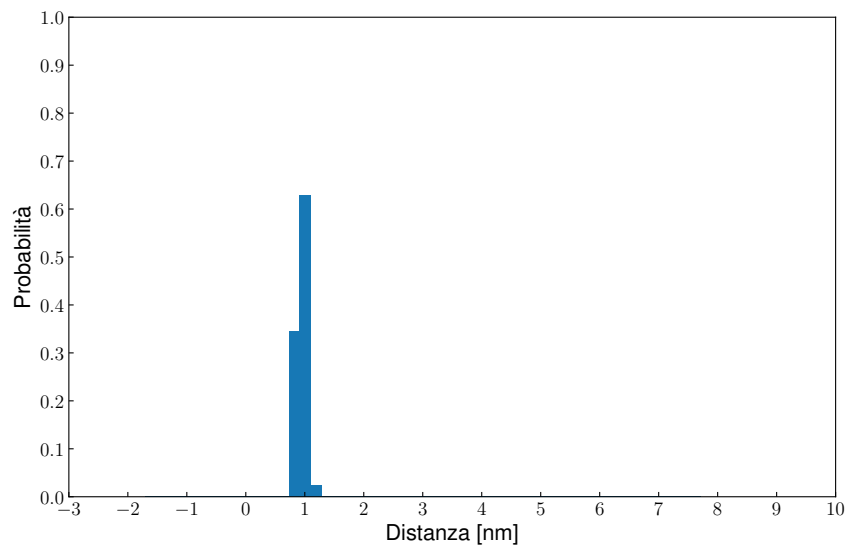


Figura 4.21: Caso 3: probabilità al variare della posizione della proteina.

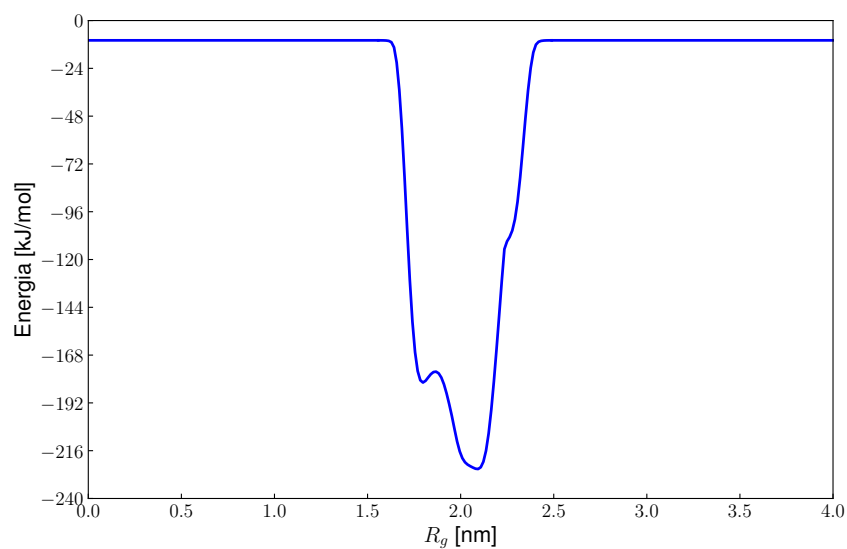


Figura 4.22: Caso 3: profilo di energia al variare del raggio di giro della proteina.

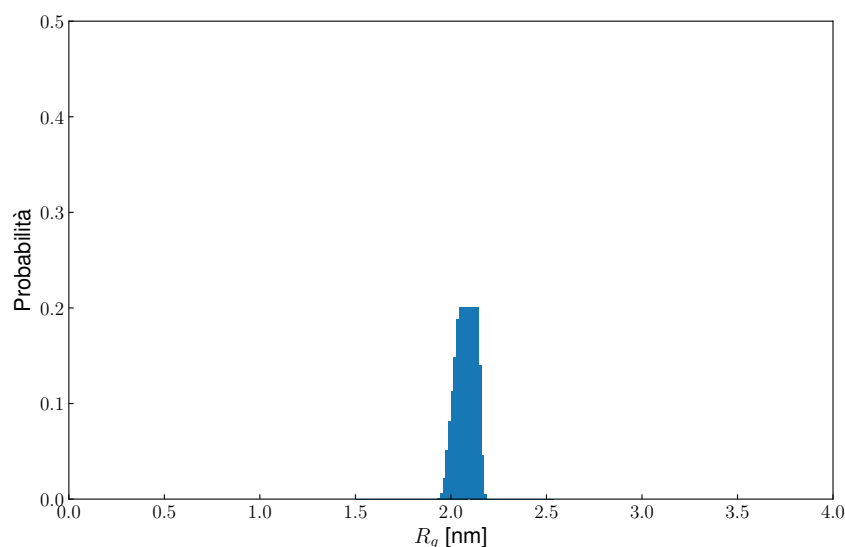


Figura 4.23: Caso 3: probabilità al variare del raggio di giro della proteina (per questioni di visualizzazione, l'asse della probabilità è stato tagliato e non raggiunge l'unità).

Nel Caso 3 si può notare come le proprietà esplorate (Figure da 4.15 a 4.18) abbiano una variazione nella prima metà della simulazione, ovvero quando la proteina si muove nell'acqua, per poi essere più stabili dopo l'assorbimento. Oltretutto, dai grafici dell'energia e della probabilità si può osservare come la proteina sia effettivamente più stabile all'interfaccia piuttosto che nel bulk dell'acqua. Inoltre, è possibile individuare un delta di energia dal grafico di Figura 4.20 che corrisponde all'energia di assorbimento. Oltre a questo, è stata eseguita un'analisi più dettagliata per individuare esattamente quali fossero gli amminoacidi assorbiti, come si può osservare nelle Figure 4.24 e 4.25. In particolare, si può notare come la proteina sia composta maggiormente da amminoacidi apolari e come siano proprio questi quelli che si assorbono maggiormente. Oltretutto, si può osservare la concordanza tra i dati della MetaD e della MD. Il raggio di giro assume lo stesso valore medio anche se calcolato con due tecniche differenti (Figure 4.15, 4.22 e 4.23).

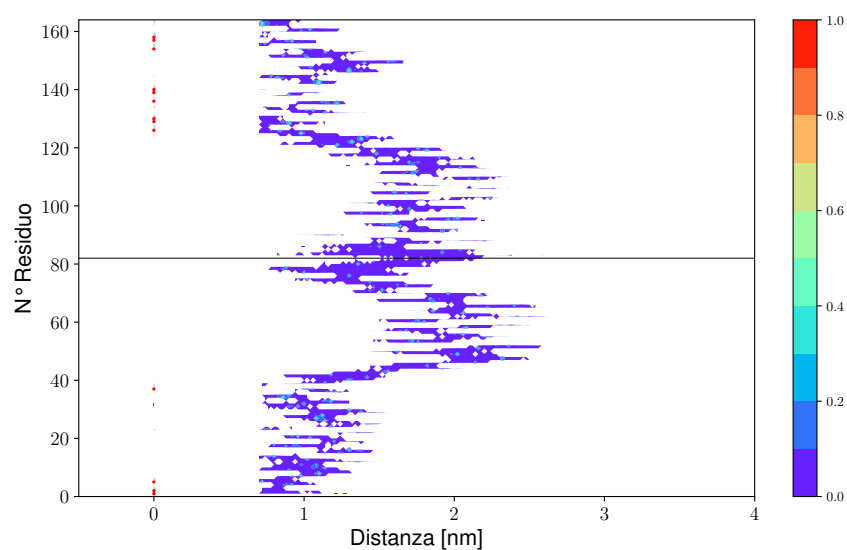


Figura 4.24: Caso 3: probabilità e posizione dei singoli amminoacidi, con il punto rosso sono stati evidenziati i punti di assorbimento.

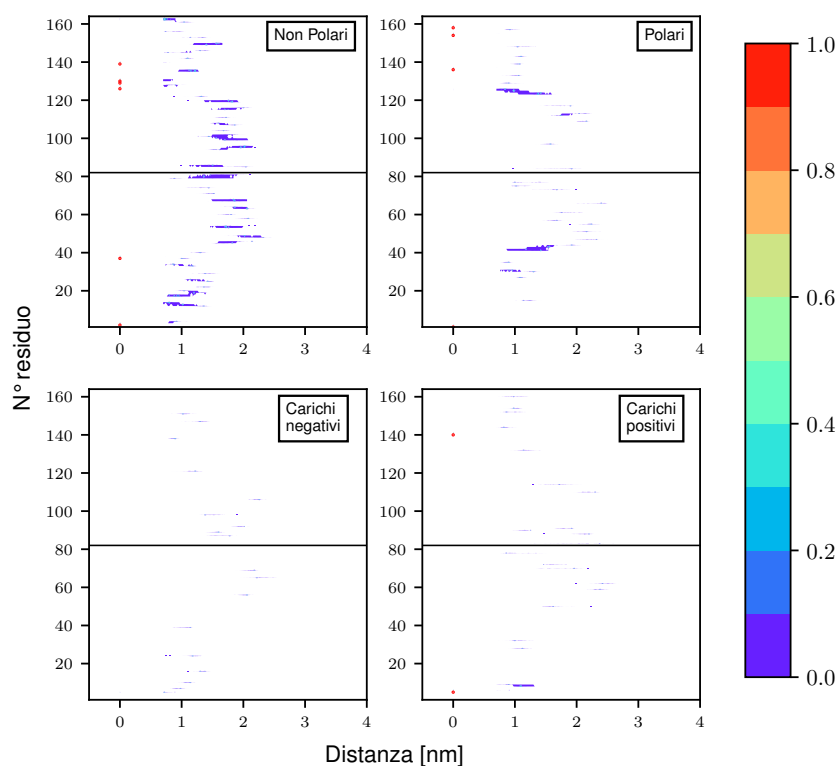


Figura 4.25: Caso 3: probabilità e posizione degli amminoacidi divisi per tipo, con il punto rosso sono stati evidenziati i punti di assorbimento.

4.4 Caso 4

Nel Caso 4, in cui è stata eseguita solo la simulazione di metadinamica, sono stati analizzati i profili di energia e le probabilità rispetto alla distanza dall'interfaccia (Figure 4.26 e 4.27) e rispetto al raggio di giro (Figure 4.28 e 4.29). Anche in questo caso, in presenza dei fosfolipidi, si nota ancora una forte stabilità della proteina all'interfaccia piuttosto che nel bulk dell'acqua.

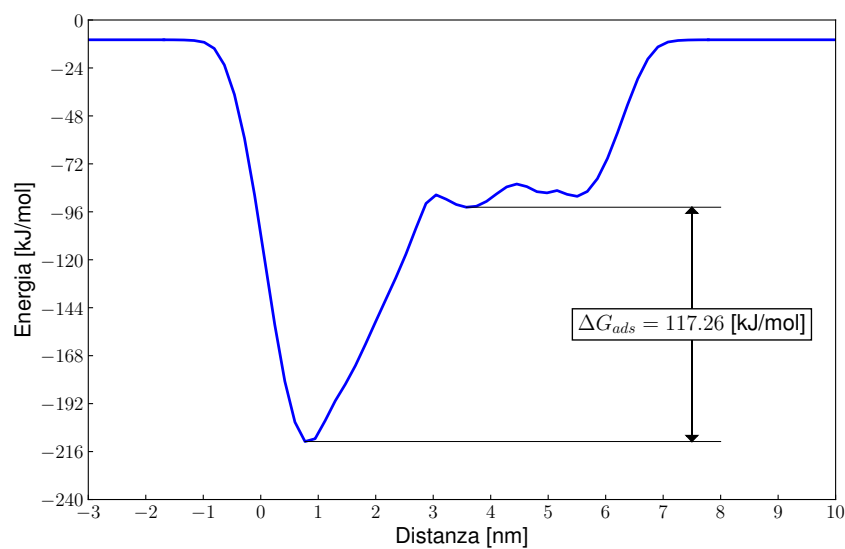


Figura 4.26: Caso 4: profilo di energia al variare della posizione della proteina.

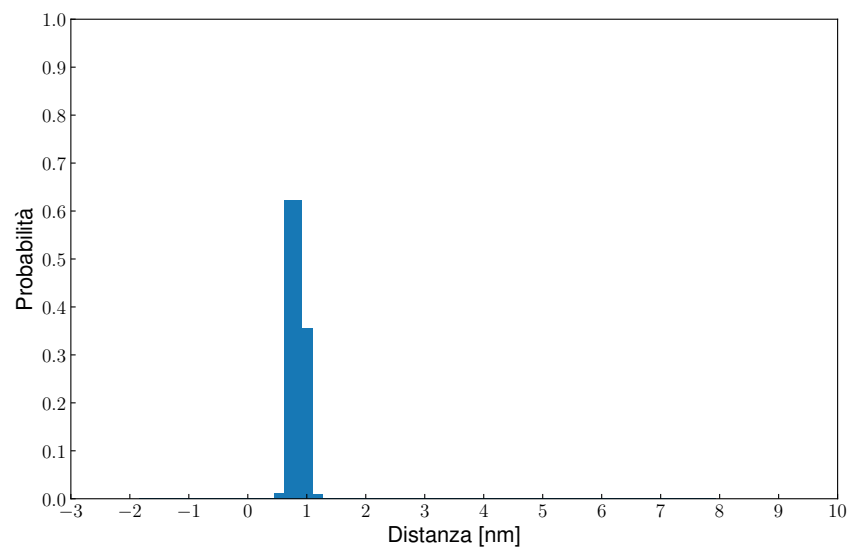


Figura 4.27: Caso 4: probabilità al variare della posizione della proteina.

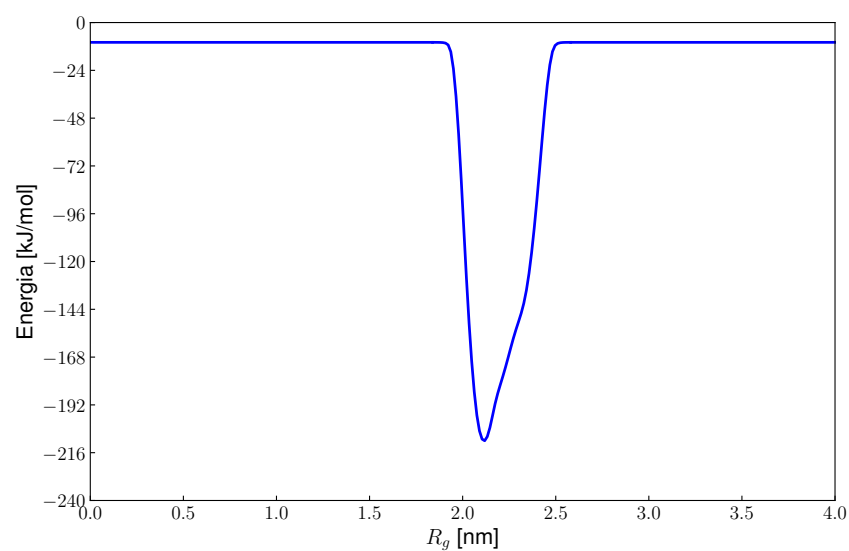


Figura 4.28: Caso 4: profilo di energia al variare del raggio di giro della proteina.

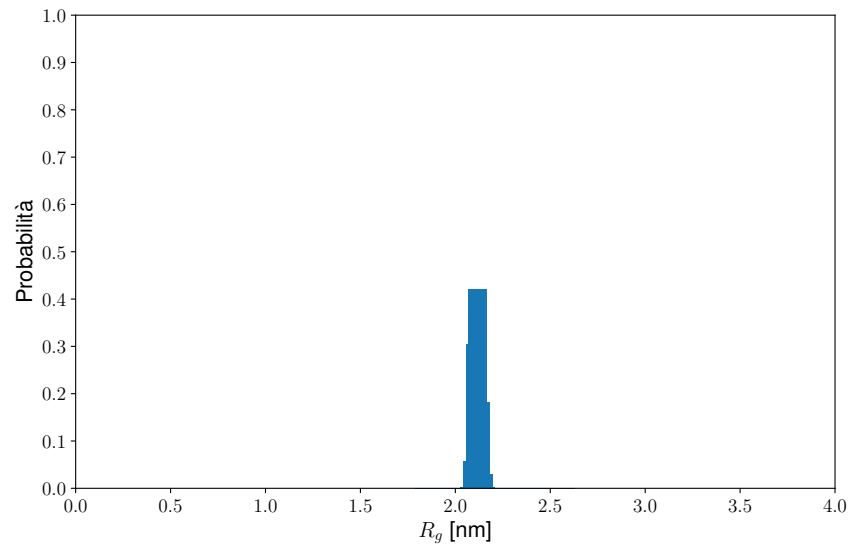


Figura 4.29: Caso 4: probabilità al variare del raggio di giro della proteina.

4.5 Caso 5

Anche nel Caso 5 sono stati analizzati i profili di energia e le probabilità rispetto alla distanza dall'interfaccia (Figure 4.30 e 4.31) e rispetto al raggio di giro (Figure 4.32 e 4.33). In questo caso si può notare che la proteina, per quanto sia più stabile vicino all'interfaccia, non riesce ad assorbirsi a causa dell'ingombro sterico dato dallo strato di fosfolipidi. Si nota anche una variazione nel raggio di giro, che tende verso valori inferiori rispetto ai casi precedenti.

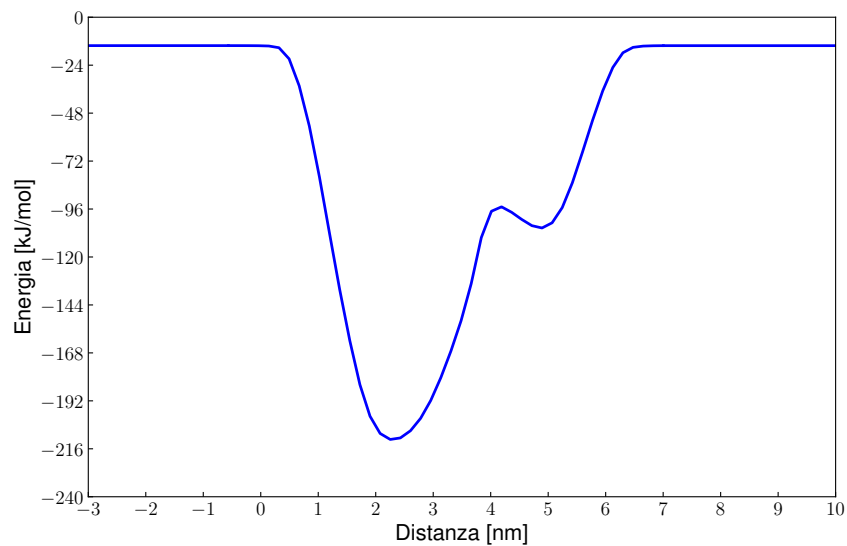


Figura 4.30: Caso 5: profilo di energia al variare della posizione della proteina.

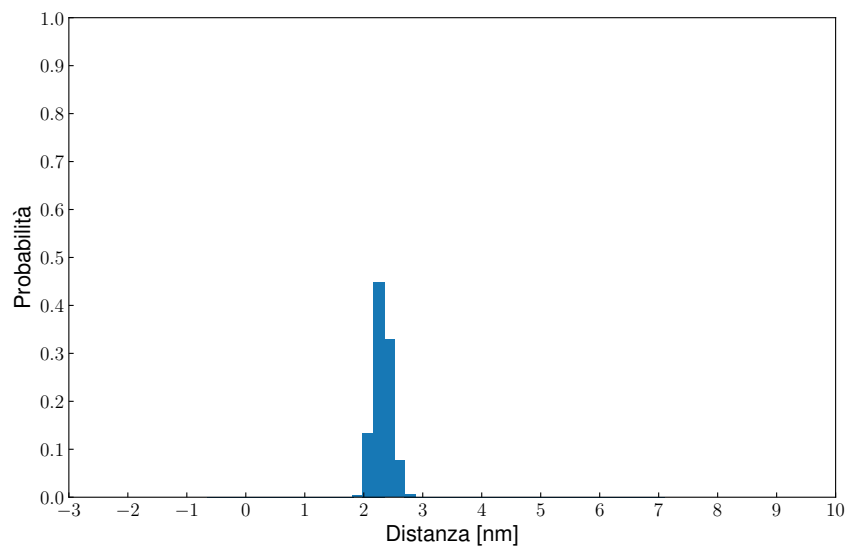


Figura 4.31: Caso 5: probabilità al variare della posizione della proteina.

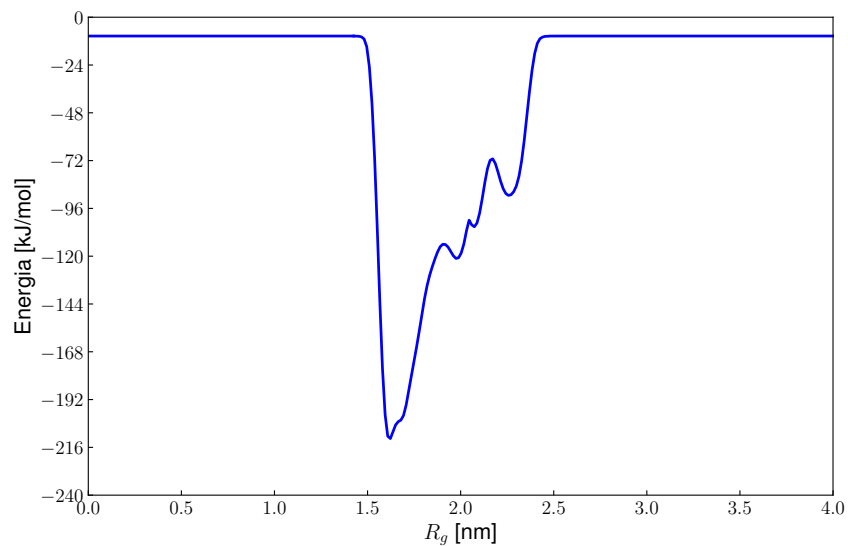


Figura 4.32: Caso 5: profilo di energia al variare del raggio di giro della proteina.

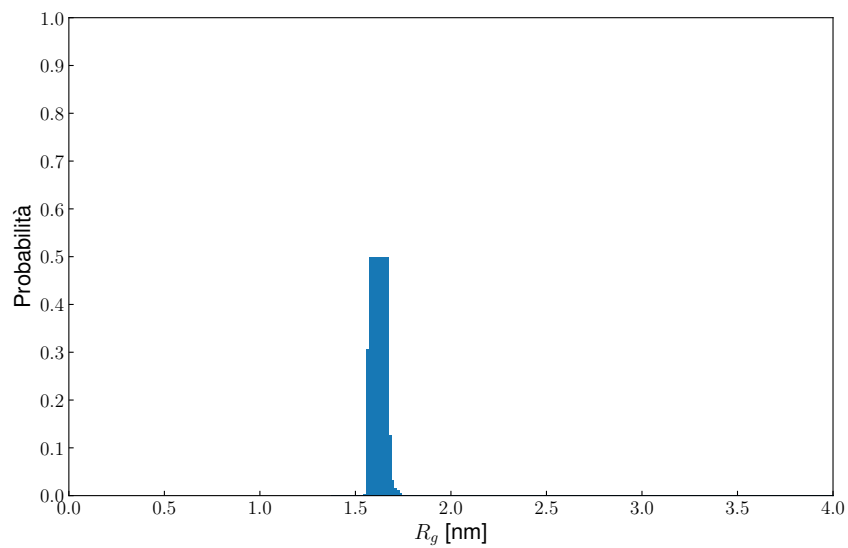


Figura 4.33: Caso 5: probabilità al variare del raggio di giro della proteina.

4.6 Caso 6

Per il sesto ed ultimo caso, come già visto per i precedenti due, sono riportati i profili di energia e le probabilità rispetto alla distanza dall'interfaccia (Figure 4.34 e 4.35) e rispetto al raggio di giro (Figure 4.36 e 4.37). Anche in questo caso si può notare il mancato assorbimento della proteina, dovuto all'ingombro sterico dato dallo strato di fosfolipidi ancora più fitto rispetto al Caso 5, e la riduzione del raggio di giro, già osservata nel caso precedente.

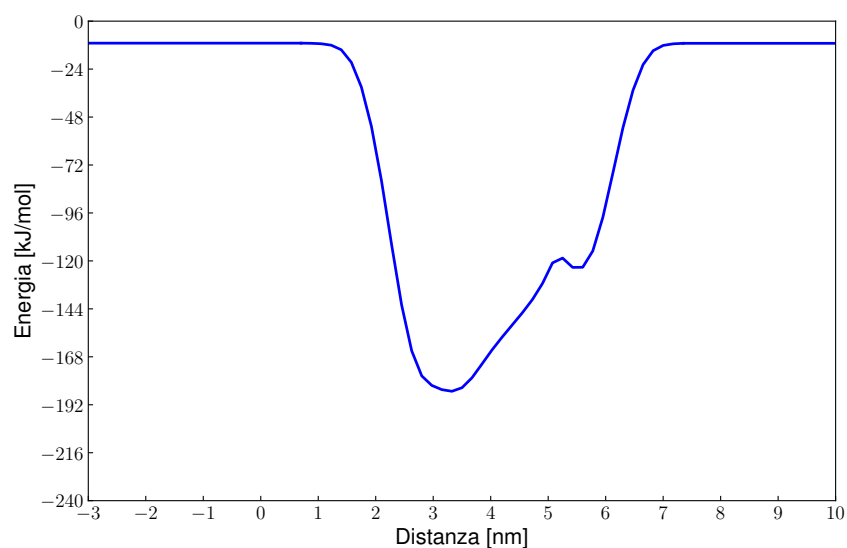


Figura 4.34: Caso 6: profilo di energia al variare della posizione della proteina.

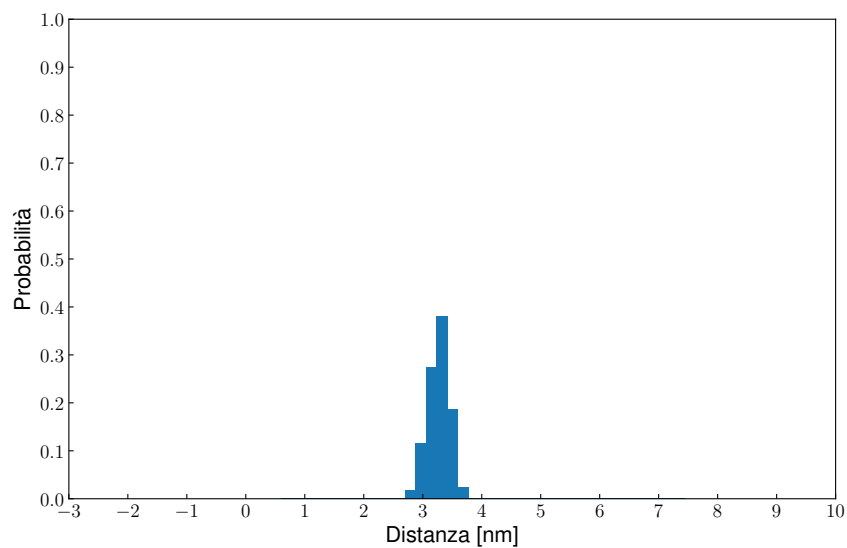


Figura 4.35: Caso 6: probabilità al variare della posizione della proteina.

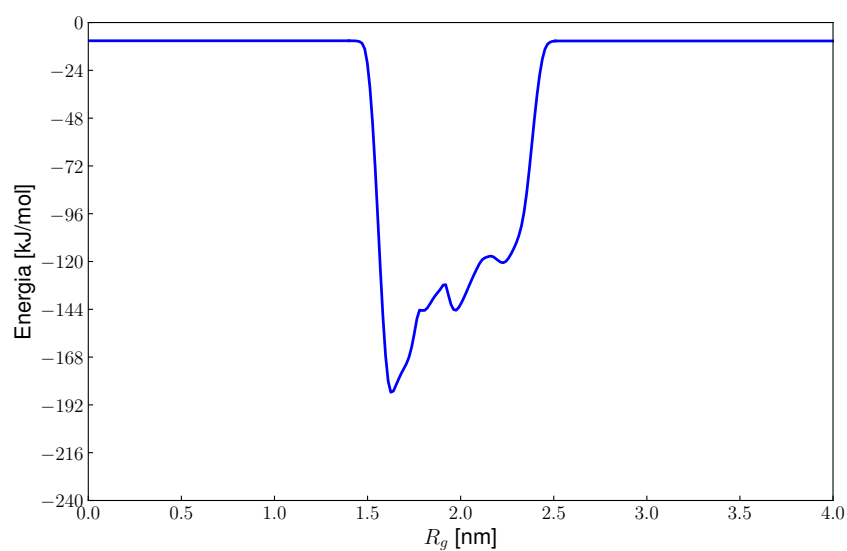


Figura 4.36: Caso 6: profilo di energia al variare del raggio di giro della proteina.

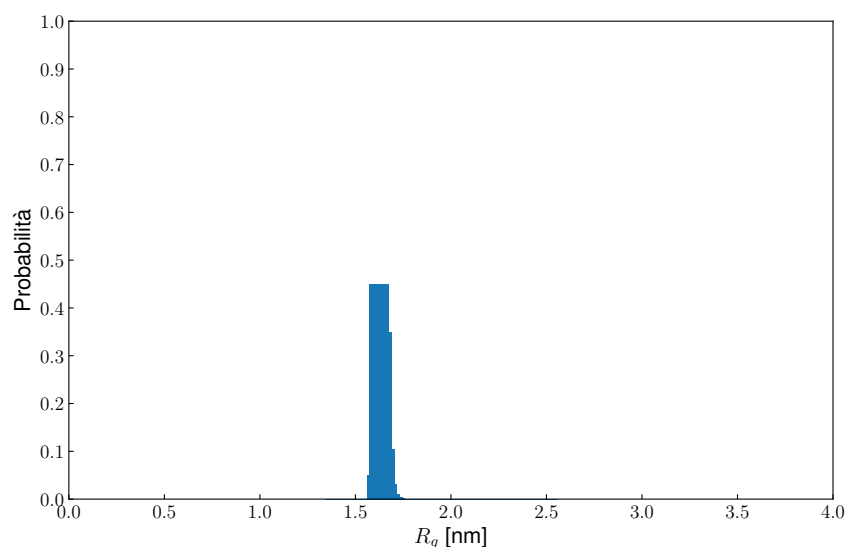


Figura 4.37: Caso 6: probabilità al variare del raggio di giro della proteina.

4.7 Confronto: Casi 1 e 2

Dalle simulazioni della proteina in acqua (Caso 1) e in olio (Caso 2) si può osservare come la proteina sia più stabile in un solvente apolare piuttosto che in uno polare. Confrontando tra di loro i grafici (Caso 1, Figure da 4.1 a 4.4; Caso 2, Figure da 4.8 a 4.11) dei due casi si nota come le proprietà calcolate abbiano oscillazioni maggiori in acqua piuttosto che in olio, con la distanza EtE che in acqua raggiunge valori di $\pm 17,86\%$ rispetto al valore medio. Infatti, come osservato nella Figura 4.25, la proteina è composta principalmente da amminoacidi apolari, più affini all'olio che all'acqua. Oltretutto, l'influenza del solvente oleoso si nota anche nella mobilità della proteina. A causa della dimensione delle molecole di olio, molto più grandi di quelle di acqua, il MSD in olio raggiunge valori di quattro ordini di grandezza inferiori rispetto al caso in acqua.

4.8 Confronto: Casi 3 e 4

La differenza tra i Casi 3 e 4 è che quest'ultimo presenta dei fosfolipidi in acqua insieme alla proteina (si ricorda che il rapporto del numero di molecole di fosfolipidi e proteina corrisponde a quello nel tuorlo d'uovo). Dai grafici ottenuti si può vedere che i fosfolipidi influenzano l'assorbimento, infatti la differenza di energia tra la proteina in bulk e all'interfaccia è maggiore nel Caso 4 (Caso 3, Figura 4.20, $\Delta G_{ads} = 108,08$ [kJ/mol]; Caso 4, Figura 4.26, $\Delta G_{ads} = 117,26$ [kJ/mol]). Dal momento che nel Caso 4 è quasi assente il minimo di energia nel bulk si può affermare che i fosfolipidi influenzano

indirettamente l'assorbimento, andando a rendere ancora meno stabile la proteina in acqua di quanto già non fosse naturalmente. Invece il raggio di giro (Caso 3, Figure 4.22 e 4.28) resta quasi invariato, non è quindi presente alcun effetto dei fosfolipidi sulle dimensioni della proteina.

4.9 Confronto: Casi 5 e 6

Gli ultimi due casi analizzati presentano entrambi uno strato di fosfolipidi all'interfaccia in diverse concentrazioni, nel Caso 5 la concentrazione di fosfolipidi è pari alla metà della concentrazione di saturazione e nel Caso 6 la concentrazione è esattamente quella di saturazione. In entrambi i casi si osserva che la proteina non raggiunge l'interfaccia a causa dell'ingombro sterico dei fosfolipidi ma tenta comunque di rimanere più vicina alla fase oleosa a cui è più affine (Caso 5, Figure 4.30 e 4.31; Caso 6, Figure 4.34 e 4.35). Inoltre, si nota una differenza nel raggio di giro che diminuisce rispetto a tutti i casi precedenti (Caso 5, Figure 4.32 e 4.33; Caso 6, Figure 4.36 e 4.37). La proteina tenta di farsi spazio nello strato fosfolipidico andando a ridurre le proprie dimensioni per cercare di avvicinarsi il più possibile all'interfaccia.

Capitolo 5

Conclusioni

Nel lavoro di tesi appena presentato sono state utilizzate tecniche di dinamica molecolare e metadinamica per lo studio e il comportamento della proteina Apovitellenina-1 del tuorlo d'uovo. Lo studio è stato caratterizzato da una crescente complessità del sistema. Prima di eseguire le simulazioni si è cercato di ottenere una struttura terziaria sufficientemente affidabile per la proteina dato che in origine era nota solo la struttura primaria. Inizialmente è stata studiata la proteina in acqua, poi si è cercato di riprodurre, nonostante le semplificazioni adottate, una soluzione che potesse rappresentare l'olio di semi di soia all'interno del quale è stata simulata la proteina per poter avere un confronto tra le proprietà in fase acquosa e in fase oleosa. Successivamente è stato creato il sistema composto da un volume equamente diviso tra acqua e olio, con la proteina posta nella fase acquosa. È stato verificato l'assorbimento della proteina sia con la tecnica della dinamica molecolare, sia con la tecnica della metadinamica per avere una conferma dell'effettivo assorbimento e della stabilità dell'Apovitellenina-1 all'interfaccia piuttosto che nel bulk della fase acquosa. Dall'analisi dell'assorbimento degli amminoacidi si è verificato che fossero effettivamente quelli apolari ad avvicinarsi maggiormente e ad ancorarsi all'interfaccia. Infine, il sistema è stato ulteriormente complicato andando ad inserire i fosfolipidi, per ottenere una condizione ancora più simile a quella effettivamente presente nel tuorlo d'uovo. Uno dei risultati più importanti ottenuti riguarda il valore del ΔG tra la posizione della proteina nel bulk dell'acqua e all'interfaccia. Questo può essere considerato come un'energia di assorbimento di una singola molecola proteica e tramite questa è possibile ricondursi ad un modello termodinamico e da questo ad un modello cinetico di assorbimento. In questo modo, si può ottenere l'abbassamento della tensione superficiale nel tempo e da quest'ultimo si può studiare la coalescenza delle gocce di olio all'interno della matrice acquosa. Minore è la tensione superficiale e maggiore sarà l'energia necessaria alle gocce per coalescere e quindi maggiore sarà la stabilità dell'emulsione. Oltretutto, la presenza di fosfolipidi e proteine all'interfaccia

crea un rivestimento che stericamente impedisce e sfavorisce ancora di più il fenomeno di coalescenza. Quindi dal modello cinetico è possibile studiare anche il tempo al quale si raggiunge la saturazione della superficie.

Acronimi

COM Center Of Mass. 31

CV Collective variables. iii, 15–20, 32

DCLM Double Cubic Lattice Method. 38

EtE End-to-End distance. 38, 39, 43, 44, 48, 62

FES Free Energy Surface. iii, 15, 16, 19, 20

MD Molecular Dynamics. 11, 15, 24, 32, 34–36, 53

MetaD Metadynamics. 11, 15, 18, 20, 32, 34–36, 53

MSD Mean Square Displacement. 38, 39, 43, 44, 48, 62

NpT Isothermal-Isobaric ensemble. 8, 26, 31

NVT Canonical ensemble. 8, 25, 31

PBC Periodic Boundary Conditions. iii, 22, 23, 31

PME Particle Mesh Ewald summation. iv, 23, 24, 31

RMSD Root Mean Square Deviation. 37, 39, 43, 44, 48

SASA Solvent Accessible Surface Area. 38, 39, 43, 44, 48

TIP3P Transferable Intermolecular Potential with 3 Points. 27, 33

WTMetaD Well Tempered Metadynamics. 17, 19, 20, 31

Bibliografia

- Ali, Abdelmoneim H. et al. (Apr. 2017). «Identification of phospholipids classes and molecular species in different types of egg yolk by using UPLC-Q-TOF-MS». In: *Food Chemistry* 221, pp. 58–66. ISSN: 03088146. DOI: 10.1016/j.foodchem.2016.10.043. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0308814616316600> (visited on 02/04/2022).
- Allen, Michael P and Dominic J Tildesley (2017). *Computer simulation of liquids*. Oxford university press.
- Anandakrishnan, Ramu, Boris Aguilar, and Alexey Onufriev (May 2012). «H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations». In: *Nucleic acids research* 40, W537–41. DOI: 10.1093/nar/gks375.
- Anton, Marc (2007). «Chapter 1 Composition and Structure of Hen Egg Yolk». In: p. 6.
- Berendsen, HJC (1991). «Transport properties computed by linear response through weak coupling to a bath». In: *Computer Simulation in Materials Science*. Springer, pp. 139–155.
- Chatzieftheriou, Stavros, Matthew R. Adendorff, and Nikos D. Lagaros (Oct. 2016). «Generalized Potential Energy Finite Elements for Modeling Molecular Nanostructures». In: *Journal of Chemical Information and Modeling* 56.10, pp. 1963–1978. ISSN: 1549-9596, 1549-960X. DOI: 10.1021/acs.jcim.6b00356. URL: <https://pubs.acs.org/doi/10.1021/acs.jcim.6b00356> (visited on 02/04/2022).
- Chialvo, Ariel A. and Pablo G. Debenedetti (1990). «On the use of the Verlet neighbor list in molecular dynamics». In: *Computer Physics Communications* 60.2, pp. 215–224. ISSN: 0010-4655. DOI: [https://doi.org/10.1016/0010-4655\(90\)90007-N](https://doi.org/10.1016/0010-4655(90)90007-N). URL: <https://www.sciencedirect.com/science/article/pii/001046559090007N>.
- Das, K.P. and J.E. Kinsella (1990). «Stability Of Food Emulsions: Physicochemical Role Of Protein And Nonprotein Emulsifiers». In: *Advances in Food and Nutrition Research*. Vol. 34. Elsevier, pp. 81–201. ISBN: 978-0-12-016434-9. DOI: 10.1016/S1043-4526(08)60007-8. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1043452608600078> (visited on 02/04/2022).

- Di Pierro, Michele, Ron Elber, and Benedict Leimkuhler (Dec. 2015). «A Stochastic Algorithm for the Isobaric–Isothermal Ensemble with Ewald Summations for All Long Range Forces». In: *Journal of Chemical Theory and Computation* 11.12, pp. 5624–5637. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.5b00648. URL: <https://pubs.acs.org/doi/10.1021/acs.jctc.5b00648> (visited on 02/04/2022).
- Dodda, Leela S., Israel Cabeza de Vaca, et al. (Apr. 2017). «LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands». In: *Nucleic Acids Research* 45.W1, W331–W336. ISSN: 0305-1048. DOI: 10.1093/nar/gkx312. eprint: <https://academic.oup.com/nar/article-pdf/45/W1/W331/23740908/gkx312.pdf>. URL: <https://doi.org/10.1093/nar/gkx312>.
- Dodda, Leela S., Jonah Z. Vilseck, et al. (2017). «1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations». In: *The Journal of Physical Chemistry B* 121.15. PMID: 28224794, pp. 3864–3870. DOI: 10.1021/acs.jpcb.7b00272. eprint: <https://doi.org/10.1021/acs.jpcb.7b00272>. URL: <https://doi.org/10.1021/acs.jpcb.7b00272>.
- Eisenhaber, Frank et al. (Mar. 1995). «The Double Cubic Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies». In: *Journal of Computational Chemistry* 16, pp. 273–284. DOI: 10.1002/jcc.540160303.
- Ferrari, Marco et al. (2022). «Molecular modeling of the interface of an egg yolk protein-based emulsion». In: *Physics of Fluids* 34.2, p. 021903. DOI: 10.1063/5.0079883. eprint: <https://doi.org/10.1063/5.0079883>. URL: <https://doi.org/10.1063/5.0079883>.
- Gordon, John et al. (Aug. 2005). «H++: A server for estimating pKas and adding missing hydrogens to macromolecules». In: *Nucleic acids research* 33, W368–71. DOI: 10.1093/nar/gki464.
- Hansson, Tomas, Chris Oostenbrink, and Wilfred F van Gunsteren (2002). «Molecular dynamics simulations». In: p. 7.
- (2011). In: *Intermolecular and Surface Forces (Third Edition)*. Ed. by Jacob N. Israelachvili. Third Edition. San Diego: Academic Press, pp. 635–660. ISBN: 978-0-12-375182-9. DOI: <https://doi.org/10.1016/B978-0-12-375182-9.10023-5>. URL: <https://www.sciencedirect.com/science/article/pii/B9780123751829100235>.
- Jämbeck, Joakim P. M. and Alexander P. Lyubartsev (2012). «An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes». In: *Journal of Chemical Theory and Computation* 8.8. PMID: 26592132, pp. 2938–2948. DOI: 10.1021/ct300342n. eprint: <https://doi.org/10.1021/ct300342n>. URL: <https://doi.org/10.1021/ct300342n>.

- Jorgensen, William L., Jayaraman Chandrasekhar, et al. (1983). «Comparison of simple potential functions for simulating liquid water». In: *The Journal of Chemical Physics* 79.2, pp. 926–935. DOI: 10.1063/1.445869. eprint: <https://doi.org/10.1063/1.445869>. URL: <https://doi.org/10.1063/1.445869>.
- Jorgensen, William L., Jeffrey D. Madura, and Carol J. Swenson (Oct. 1984). «Optimized intermolecular potential functions for liquid hydrocarbons». In: *Journal of the American Chemical Society* 106.22, pp. 6638–6646. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja00334a030. URL: <https://pubs.acs.org/doi/abs/10.1021/ja00334a030> (visited on 02/04/2022).
- Jorgensen, William L., David S. Maxwell, and Julian Tirado-Rives (Nov. 1996). «Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids». In: *Journal of the American Chemical Society* 118.45, pp. 11225–11236. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja9621760. URL: <https://pubs.acs.org/doi/10.1021/ja9621760> (visited on 02/04/2022).
- Jorgensen, William L. and Julian Tirado-Rives (2005). «Potential energy functions for atomic-level simulations of water and organic and biomolecular systems». In: *Proceedings of the National Academy of Sciences* 102.19, pp. 6665–6670. ISSN: 0027-8424. DOI: 10.1073/pnas.0408037102. eprint: <https://www.pnas.org/content/102/19/6665.full.pdf>. URL: <https://www.pnas.org/content/102/19/6665>.
- Jumper, John et al. (2021). «Highly accurate protein structure prediction with AlphaFold». In: *Nature* 596.7873, pp. 583–589. DOI: 10.1038/s41586-021-03819-2.
- Kolafa, Jiri and John W. Perram (Jan. 1992). «Cutoff Errors in the Ewald Summation Formulae for Point Charge Systems». In: *Molecular Simulation* 9.5, pp. 351–368. ISSN: 0892-7022, 1029-0435. DOI: 10.1080/08927029208049126. URL: <http://www.tandfonline.com/doi/abs/10.1080/08927029208049126> (visited on 02/04/2022).
- Levine, I. (2009). *Physical Chemistry*. McGraw-Hill international edition. McGraw-Hill Education. ISBN: 9780072538632. URL: <https://books.google.it/books?id=L5juAAAAMAAJ>.
- Lindahl, Erik R (2008). «Molecular Dynamics Simulations». In: *Molecular Dynamics Simulations*, p. 21.
- McClements, David Julian (Sept. 2007). «Critical Review of Techniques and Methodologies for Characterization of Emulsion Stability». In: *Critical Reviews in Food Science and Nutrition* 47.7, pp. 611–649. ISSN: 1040-8398, 1549-7852. DOI: 10.1080/10408390701289292. URL: <https://www.tandfonline.com/doi/full/10.1080/10408390701289292> (visited on 02/04/2022).
- Myers, Jonathan et al. (June 2006). «A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules». In: *Proteins* 63, pp. 928–38. DOI: 10.1002/prot.20922.

- Nosé, Shuichi and M.L. Klein (Dec. 1983). «Constant pressure molecular dynamics for molecular systems». In: *Molecular Physics* 50.5, pp. 1055–1076. ISSN: 0026-8976, 1362-3028. DOI: 10.1080/00268978300102851. URL: <http://www.tandfonline.com/doi/abs/10.1080/00268978300102851> (visited on 02/04/2022).
- Parrinello, M. and A. Rahman (1981). «Polymorphic transitions in single crystals: A new molecular dynamics method». In: *Journal of Applied Physics* 52.12, pp. 7182–7190. DOI: 10.1063/1.328693. eprint: <https://doi.org/10.1063/1.328693>. URL: <https://doi.org/10.1063/1.328693>.
- Pryde, E H (1980). «COMPOSITION OF SOYBEAN OIL». In: p. 19.
- Swope, William C. et al. (1982). «A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters». In: *The Journal of Chemical Physics* 76.1, pp. 637–649. DOI: 10.1063/1.442716. eprint: <https://doi.org/10.1063/1.442716>. URL: <https://doi.org/10.1063/1.442716>.
- Tolman, R.C. (1979). *The Principles of Statistical Mechanics*. Dover Books on Physics. Dover Publications. ISBN: 9780486638966. URL: <https://books.google.it/books?id=4TqQZo962s0C>.
- Valsson, Omar, Pratyush Tiwary, and Michele Parrinello (2016). «Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint». In: *Annual Review of Physical Chemistry* 67.1. PMID: 26980304, pp. 159–184. DOI: 10.1146/annurev-physchem-040215-112229. eprint: <https://doi.org/10.1146/annurev-physchem-040215-112229>. URL: <https://doi.org/10.1146/annurev-physchem-040215-112229>.
- Zambrowicz, Aleksandra et al. (2015). «Multifunctional peptides derived from an egg yolk protein hydrolysate: isolation and characterization». In: *Amino Acids* 47.2, pp. 369–380. DOI: 10.1007/s00726-014-1869-x. URL: <http://link.springer.com/10.1007/s00726-014-1869-x>.

Ringraziamenti

A conclusione di questo lavoro di tesi, si ringraziano HPC@POLITO (<http://www.hpc.polito.it>) e CINECA (<http://www.cineca.it/hpc>) per le risorse computazionali fornite.

Desidero inoltre ringraziare i professori Daniele Marchisio, Antonio Buffo, Gianluca Boccardo e la professoressa Francesca Bosco per avermi dato l'opportunità di svolgere questo lavoro di tesi senza farmi mancare il supporto, i consigli e l'aiuto di cui potevo aver bisogno, seguendomi passo dopo passo fino alla conclusione. Ringrazio inoltre tutte le persone del gruppo di ricerca che ho conosciuto e che mi hanno aiutato nella parte finale di questo percorso facendomi sentire parte del gruppo. Un ringraziamento speciale va a Marco, che fin dall'inizio mi ha seguito e mi ha insegnato tutto quello che dovevo sapere sia per lo svolgimento di questo lavoro, sia per il proseguimento degli studi. Ringrazio i miei genitori, mia sorella e mio fratello che mi hanno sempre sostenuto nonostante la distanza da casa. Infine ringrazio tutti gli amici che ho incontrato durante il percorso universitario, che mi hanno accompagnato e sostenuto fino ad ora, anche nei momenti più difficili.

Grazie a tutti.