

**POLITECNICO DI TORINO**

Collegio di Ingegneria Chimica e dei Materiali

**Corso di Laurea Magistrale  
in Ingegneria Chimica e dei Processi Sostenibili**

Tesi di Laurea Magistrale

**Dinamica Molecolare applicata alla  
formulazione di biofarmaci: le interazioni  
amminoacido-eccipiente sono additive?**



**Relatore**

prof. Roberto Pisano

**Candidato**

Lorenzo La Cortiglia

S252569

Ottobre 2019



# Indice

<b>RINGRAZIAMENTI.....</b>	<b>7</b>
<b>1. INTRODUZIONE.....</b>	<b>9</b>
1.1 Introduzione al processo di liofilizzazione.....	9
1.1.1 Tipologie di liofilizzazione.....	9
1.2 Le fasi della liofilizzazione .....	10
1.2.1 Congelamento: parametri di processo.....	11
1.2.2 Prima fase di essicamento: la sublimazione del ghiaccio.....	14
1.2.3 Essicamento secondario.....	15
1.3 Liofilizzazione di biofarmaci .....	16
1.3.1 Struttura delle proteine.....	18
1.4 Denaturazione nelle proteine .....	21
1.4.1 Stabilizzazione delle proteine tramite eccipienti .....	23
1.4.2 Meccanismi di protezione durante la liofilizzazione .....	24
1.4.3 Termodinamica della crioprotezione: teoria dell'esclusione preferenziale.....	25
1.4.4 Interazione proteina-eccipiente nella fase congelata.....	26
1.4.5 Protezione durante l'essicamento e lo stoccaggio.....	27
1.5 Instabilità allo stato amorfo .....	27
1.5.1 Instabilità durante lo stoccaggio .....	28
1.6 Reazioni di denaturazione .....	29
1.7 Obiettivi della tesi.....	31
<b>2. METODI: APPROCCIO COMPUTAZIONALE E DESIGN DELLE SIMULAZIONI .....</b>	<b>33</b>
2.1 Introduzione alla Dinamica Molecolare .....	33
2.1.1 Periodic boundary conditions – condizioni periodiche al contorno .....	36
2.1.2 Neighbour list – lista dei 'vicini' .....	37
2.2 Procedura di simulazione .....	38
2.2.1 Creazione del file di struttura.....	39
2.2.2 Creazione della box .....	44
2.2.3 Inserimento dello zucchero.....	44
2.2.4 Inserimento dell'acqua.....	46
2.2.5 Neutralizzazione della box e minimizzazione energetica del sistema .....	47
2.2.6 Equilibratura del sistema.....	48
2.2.7 Simulazione della traiettoria .....	49
2.3 Analisi dei risultati .....	50
2.3.1 Funzione di distribuzione radiale ( <i>rdf</i> ) .....	50

2.3.2	Valutazione dei legami a idrogeno .....	52
2.3.3	Valutazione superficie esposta e analisi.....	54
2.3.4	Valutazione del clustering .....	55
2.3.5	Analisi dei KBI .....	55
2.4	Additività delle interazioni proteina-cosoluto .....	57
2.4.1	Backbone .....	59
2.4.2	Sidechain .....	60
2.4.3	Metodi di calcolo per l'additività utilizzati nell'elaborazione dei dati .....	62
2.5	Design delle simulazioni .....	62
<b>3.</b>	<b>RISULTATI.....</b>	<b>69</b>
3.1	Introduzione .....	69
3.2	Comportamento degli amminoacidi in presenza di saccarosio o trealosio .....	69
3.2.1	Valutazione del <i>KBI</i> proteina – zucchero ( $G_{23}$ ) .....	69
3.2.2	Valutazione del <i>KBI</i> proteina – acqua ( $G_{12}$ ) .....	72
3.2.3	Valutazione del coefficiente di esclusione preferenziale .....	74
3.2.4	Valutazione del <i>water replacement</i> (fattore $\chi$ ) .....	76
3.2.5	Valutazione delle aree esposte con <i>gmx sasa</i> .....	78
3.2.6	Valutazione del clustering .....	80
3.3	Verifica dell'additività.....	82
3.3.1	Descrizione delle analisi .....	82
3.3.2	Valutazione del coefficiente di esclusione preferenziale per il backbone .....	87
3.3.3	Valutazione del coefficiente di esclusione preferenziale della sidechain .....	88
3.3.4	Valutazione delle aree esposte nei tripeptidi Gly – X – Gly .....	89
3.3.5	Valutazione dell'additività per il GB1 – hairpin.....	90
3.3.6	Valutazione dell'additività per le catene proteiche .....	93
3.4	Commenti sui risultati .....	97
<b>4.</b>	<b>COMMENTI FINALI E PROSPETTIVE FUTURE.....</b>	<b>101</b>
	<b>LISTA DEI SIMBOLI.....</b>	<b>103</b>
	LETTERE GRECHE .....	104
	APICI E PEDICI .....	105
	<b>APPENDICE.....</b>	<b>107</b>
A.1	File di minimizzazione energetica ( <i>enermin.mdp</i> ) .....	107
A.2	File di equilibratura ( <i>npt.mdp</i> ).....	107
A.3	File di simulazione della traiettoria ( <i>md.mdp</i> ).....	108
A.4	File di calcolo dei <i>KBI</i> ( <i>KBI.py</i> ) .....	110
A.5.1	File di analisi delle aree ponderate per il backbone ( <i>sasa-back.py</i> ) .....	110

A.5.2 File di analisi delle aree ponderate per la sidechain (sasa-side.py).....	111
<b>BIBLIOGRAFIA .....</b>	<b>113</b>



## **RINGRAZIAMENTI**

Ringrazio i miei genitori per aver sempre creduto in me. A loro dedico questo traguardo.

Ringrazio il relatore Roberto Pisano per la grande professionalità e la grande disponibilità dimostrate nei miei confronti durante questo periodo.

In particolar modo, ringrazio anche il dottorando Andrea Arsiccio per avermi messo nelle condizioni migliori per svolgere questo lavoro, con disponibilità e competenze che difficilmente ho avuto modo di trovare in altre persone.

Ringrazio inoltre la mia ragazza Francesca e il mio amico Ray per avermi sempre supportato durante la mia carriera scolastica.



# 1. INTRODUZIONE

## 1.1 Introduzione al processo di liofilizzazione

Con il termine liofilizzazione (freeze-drying) ci si riferisce a un processo in cui si rimuovono, attraverso sublimazione, le molecole di solvente, spesso acqua, presenti in un dato materiale. Il processo di essiccamento è condotto in condizioni di bassa temperatura e alto vuoto, le quali consentono la conservazione di tutti quei prodotti che sono termosensibili. La temperatura del prodotto durante il processo di essiccamento è inferiore agli 0° C (-20/-40 °C), mentre la pressione operativa è inferiore alla tensione di vapore del solvente a quella determinata temperatura (5-100 Pa). A causa di queste condizioni estremamente cautelative, i processi chimici e biologici che possono portare alla denaturazione del prodotto si fermano, mantenendo inalterate le proprietà biologiche e terapeutiche del prodotto. Ovviamente questo risultato è raggiunto solo se si utilizzano degli eccipienti che possono proteggere le molecole più sensibili dagli stress indotti dal processo di liofilizzazione stesso (problema della *formulazione*) e se il processo è condotto in modo adeguato (Rey, 2010).

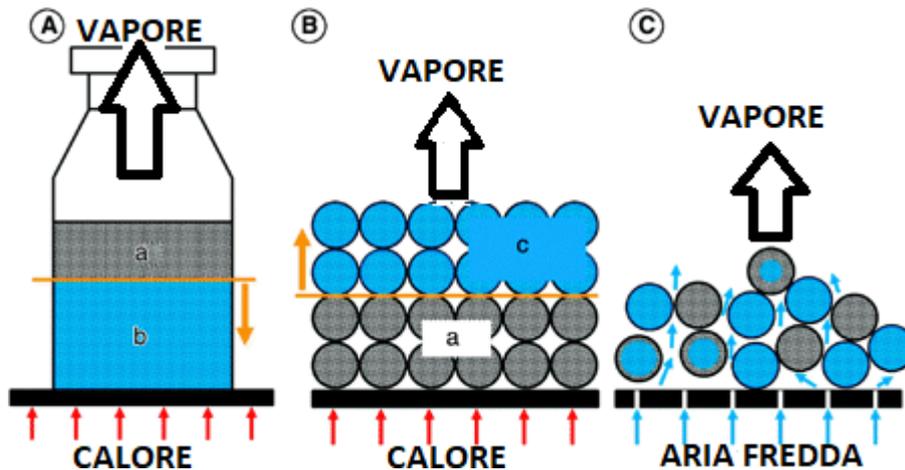
Concentrandosi sulla liofilizzazione applicata in campo farmaceutico i vantaggi che si possono ottenere sono molteplici. Questi includono la stabilità dei principi attivi, la facilità di stoccaggio, anche a temperatura ambiente, e di somministrazione del liofilizzato, la possibilità di condurre il processo in ambienti privi di contaminanti (ambienti sterilizzati, a temperature molto basse e in condizioni di vuoto spinto), e la possibilità di reidratare il componente essiccato in tempi brevi (Wang, 2000). La liofilizzazione è il processo più comunemente utilizzato oggi per produrre farmaci allo stato solido (Cleland *et al.*, 1993) e, quindi, tutti quei principi attivi che non sono stabili in fase liquida.

### 1.1.1 Tipologie di liofilizzazione

La liofilizzazione può essere in genere classificata in due gruppi di processo (Figura 1.1):

- Liofilizzazione sotto vuoto per contatto
- Liofilizzazione atmosferica

Nel primo processo si opera in una atmosfera rarefatta e quasi interamente occupata dai vapori di sublimazione; non è presente ossigeno a contatto con il prodotto, ed è così inibita qualsiasi reazione di inattivazione causata dal contatto con esso. Inoltre il calore può essere fornito principalmente per conduzione (contatto) e irraggiamento. Nel caso della liofilizzazione atmosferica, il vapore viene rimosso per sublimazione attraverso un flusso di gas secco a bassa temperatura. In questo caso è possibile lavorare a pressione atmosferica, purché la pressione parziale del solvente nel gas di processo sia inferiore alla tensione di vapore a quella determinata temperatura. Il calore può pertanto essere trasmesso efficacemente, e il contributo dominante è quello per convezione (Kharaghani *et al.*, 2017).

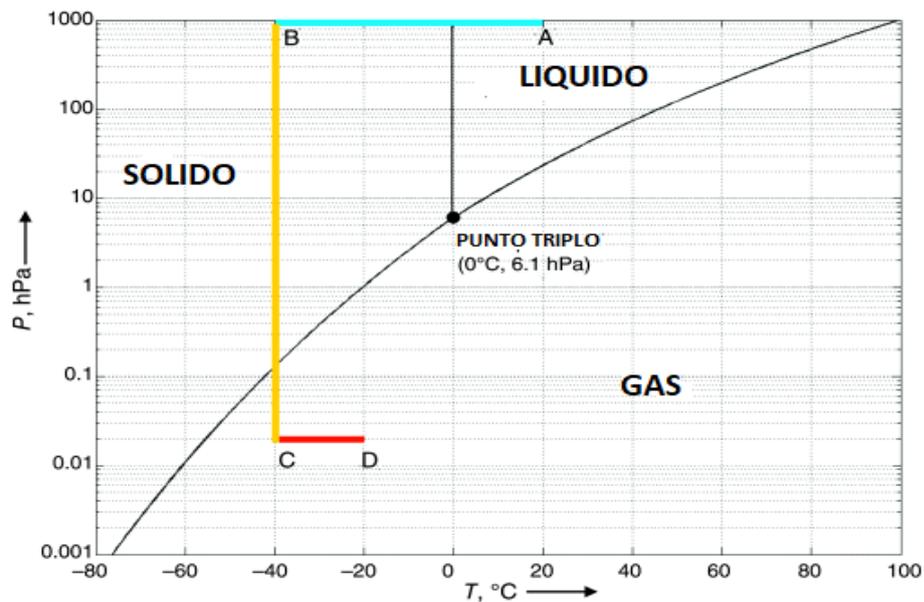


**Figura 1.1:** Figura schematica dei processi di liofilizzazione in fiala (A), sotto vuoto per contatto (B), entrambi rientranti nella liofilizzazione sotto vuoto, e del processo di liofilizzazione atmosferica (C). Le regioni essiccate (a) sono distinte da quelle congelate (b) da una linea orizzontale arancione, chiamata fronte di sublimazione, la freccia verticale indica l'andamento del fronte. La lettera 'c' indica un congelamento solo tra le particelle. Immagine tratta, con modifiche, da Ullman's Encyclopedia of Chemical Engineering (Kharaghani et al., 2017).

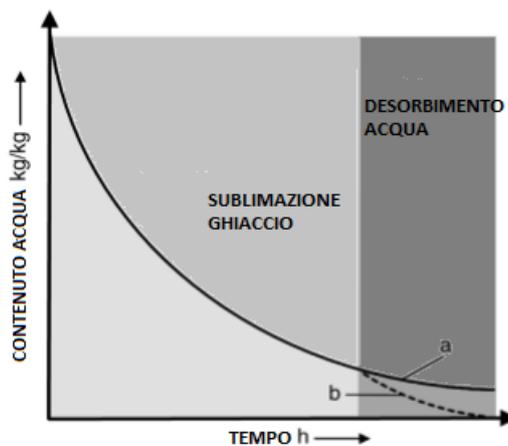
## 1.2 Le fasi della liofilizzazione

Il processo di liofilizzazione consta di tre fasi: il congelamento (solidificazione), essiccamento primario (sublimazione del ghiaccio), essiccamento secondario (rimozione per desorbimento delle molecole di solvente non solidificate poiché legate alle molecole di soluto (Arakawa *et al.*, 1993)). Durante il congelamento (tratto A-B, Figura 1.2) si abbassa la temperatura del prodotto sino a  $-40 / -50$  °C, permettendo così la formazione di ghiaccio, che inizia sottoforma di nucleazione, continuando poi con la crescita dei cristalli. È importante che tutto il prodotto sia completamente congelato, così da evitare collasso per l'aumento di temperatura richiesto per le fasi successive. Durante l'essiccamento primario (tratti BC e CD, Figura 1.2) il ghiaccio viene rimosso per sublimazione riducendo la pressione in camera al di sotto della tensione di saturazione del solvente, e aumentando la temperatura così da sopperire al calore rimosso dalla sublimazione del ghiaccio. Alla fine di questa fase il contenuto di acqua all'interno del liofilizzato è tipicamente attorno al 5-10%. Questa umidità residua, dovuta alla presenza di acqua non congelata e ancora legata al soluto, viene ridotta durante la fase di essiccamento secondario, in cui, rispetto all'essiccamento primario, la temperatura viene ulteriormente aumentata. Quest'ultima parte è relativamente breve rispetto alle prime due e, alla fine di essa, si avrà un contenuto di acqua pari a 1-2% rispetto alla massa di materiale secco.

Per un farmaco i tempi di liofilizzazione possono variare da alcuni giorni ad alcune settimane e dipendono dalla temperatura critica del prodotto, che determina la temperatura del ripiano riscaldante, dalla pressione, dal coefficiente di scambio di calore tra il contenitore e il ripiano, dalla resistenza al trasferimento di materia del prodotto e infine dal volume di riempimento. Durante la parte finale il comportamento del prodotto in relazione al suo contenuto di acqua può variare a seconda che esso sia



**Figura 1.2:** Diagramma di fase dell'acqua pura. Immagine tratta, con modifiche, da Ullman's Encyclopedia of Chemical Engineering (Kharaghani et al., 2017).



**Figura 1.3:** Variazione del contenuto di acqua durante la liofilizzazione per un prodotto igroscopico (a) e non igroscopico (b). Immagine tratta, con modifiche, da Ullman's Encyclopedia of Chemical Engineering (Kharaghani et al., 2017).

igroscopico o non igroscopico. Nel primo caso, durante l'ultima fase di processo, il contenuto di acqua si porterà a un valore prossimo a quello di equilibrio, che per i materiali non igroscopici è prossimo a zero (Figura 1.3).

### 1.2.1 Congelamento: parametri di processo

Il primo step della liofilizzazione, ossia il congelamento è la parte più delicata da sviluppare, poiché essa determina la struttura e la morfologia del prodotto. In particolare questa parte del processo influisce su:

- Tempi di essiccamento

- Massima temperatura raggiunta dal prodotto
- Pressione cui l'acqua può esser desorbita
- Contenuto di umidità residua a fine processo
- Tempi di reidratazione del prodotto

In genere grandi cristalli di ghiaccio, quindi grossi pori nel liofilizzato sono preferibili, in quanto si traducono in una ridotta resistenza al trasferimento di materia, e quindi in un essiccamento più veloce (Willemer, 1992). Inoltre, la temperatura massima raggiunta dal prodotto è più bassa, e ciò consente di minimizzare il rischio di collasso. Infine, anche l'interfaccia acqua-ghiaccio è meno estesa se i cristalli sono di grosse dimensioni, e ciò riduce il rischio di adsorbimento e denaturazione interfacciale (Strambini e Gabellieri, 1996).

Per comprendere nel dettaglio la liofilizzazione di una generica miscela acqua-soluto, è opportuno partire dalle proprietà dell'acqua pura. Dal punto di vista termodinamico, l'acqua pura congela a 0 °C a pressione atmosferica, condizioni in cui si ha un equilibrio termodinamico solido-liquido; tuttavia, a causa di fenomeni di metastabilità dettati da ragioni cinetiche, la spontanea formazione di ghiaccio per nucleazione omogenea, ossia la formazione di nuclei in assenza di impurità o superfici esterne che possano fungere da centro di nucleazione, si ha a - 40°C (Moore e Bolinero, 2011). Durante il processo di liofilizzazione si osserva tipicamente nucleazione eterogenea, che si verifica a temperature inferiori a 0 °C. In un processo di congelamento di una miscela soluto-acqua può accadere che alcune molecole di acqua si trovino a non potersi muovere verso i cristalli in accrescimento o i centri di nucleazione, a causa della crescente viscosità attorno a esse, causata dalla transizione liquido-solido. Queste molecole, cui ci riferisce con il nome di *acqua legata*, rimangono non congelate e, possono portare al collasso del prodotto nella fase di essiccamento secondario, poiché espletano un'azione plasticizzante.

Il comportamento dell'acqua cambia però leggermente in presenza di soluti diversi. A seconda della frazione massica del soluto ci si trova in diverse zone del diagramma di fase (Figura 1.4). Per aver il *congelamento* della miscela si deve raffreddare al di sotto della temperatura di eutettico ( $T_e$ ), se si sta trattando un soluto cristallino, oppure al di sotto della temperatura di transizione vetrosa ( $T_g$ ), se si sta trattando un soluto amorfo. In generale possono essere presenti entrambe le tipologie di soluto, e in questo caso si dovrà tener in considerazione che gli amorfi *congeleranno* solamente al di sotto della  $T_g$ .

La  $T_g$  può esser calcolata con l'equazione di Gordon – Taylor, che tiene conto della presenza di acqua nel prodotto (Octjen e Haseley, 2003):

$$T_g = \frac{(1 - x_w)T_{g,s} + x_w k T_{g,w}}{(1 - x_w) + x_w k} \quad (1.1)$$

$x_w$  è la frazione massica di acqua,  $T_{g,w} = - 135$  °C, è la temperatura di transizione vetrosa dell'acqua,  $k$  è una costante che dipende dal soluto e  $T_{g,s}$  la temperatura di transizione vetrosa del soluto allo stato puro.

La struttura del prodotto liofilizzato dipende fortemente anche da altri parametri, come il numero di nuclei formati durante la cristallizzazione, il grado di sotto raffreddamento, la temperatura a fine processo, la mobilità dell'acqua all'interno della formulazione e la velocità di raffreddamento. Questi parametri sono interdipendenti, ma si possono evidenziare due situazioni distinte che in termini generali descrivono come avverrà il processo di liofilizzazione (Kharaghani *et al.*, 2017):

- Raffreddamento lento ( $< 0.5 \frac{^{\circ}\text{C}}{\text{min}}$ ): bassa velocità di congelamento, formazione di grandi cristalli di ghiaccio, concentrazione del soluto costante in tutto il volume di prodotto.
- Raffreddamento veloce ( $5-300 \frac{^{\circ}\text{C}}{\text{min}}$ ): piccoli cristalli di ghiaccio, che impediscono la fuoriuscita dell'acqua non congelata e allungano il tempo di essiccazione. Il sottoraffreddamento può essere di 10-20 °C, il soluto si concentrerà solo in alcune regioni della soluzione. Alcune molecole di acqua non riescono a muoversi verso i nuclei e si comporteranno come un solido, rimanendo intrappolate tra i cristalli, seppur ancora allo stato liquido.

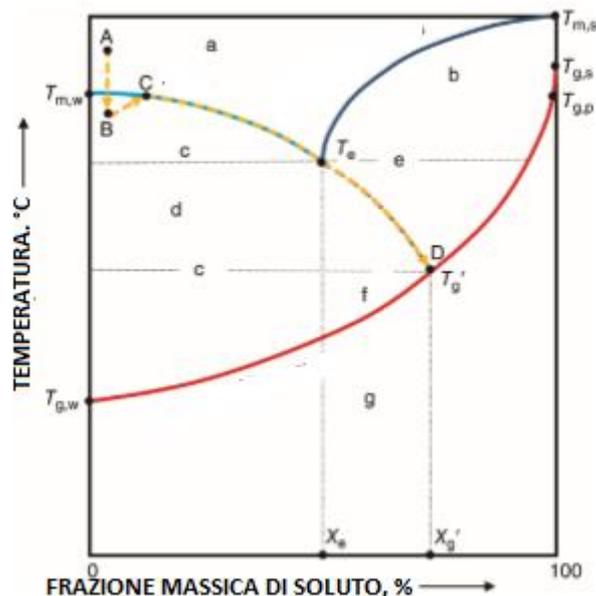
Il tempo di raffreddamento di una miscela può essere stimato con la equazione di Steinbach (Octjen e Haseley, 2003):

$$t_r = \frac{\Delta h}{\Delta T} \rho_f \left( \frac{L^2}{2\lambda_{eff}} + \frac{L}{\alpha_{surf}} \right) \quad (1.2)$$

dove  $\Delta h$  è la differenza di entalpia tra il punto di inizio e il punto di fine congelamento,  $\Delta T$  è la differenza di temperatura tra il fronte di congelamento e il fluido refrigerante,  $L$  è la dimensione del lato di prodotto parallela al flusso di calore prevalente,  $\rho_f$  è la densità del prodotto congelato,  $\lambda_{eff}$  indica la conducibilità termica del prodotto,  $\alpha_{surf}$  indica la superficie media di scambio di calore tra prodotto e mezzo refrigerante, e va misurata per ogni tipo di fiala o ripiano utilizzato.

Il congelamento può essere condotto in camere separate o nelle stesse camere di liofilizzazione, a pressione atmosferica. I farmaci sono in genere congelati direttamente all'interno di fiale poste su ripiani raffreddati della camera di liofilizzazione (Rey, 2010).

Se il prodotto non viene completamente congelato con il primo raffreddamento sui ripiani, è necessario congelare l'acqua rimasta allo stato liquido riscaldando il prodotto sino a una temperatura superiore a quella di transizione vetrosa che consenta all'acqua di potersi muovere e raggiungere i centri di nucleazione o i cristalli già formati. In seguito si procede con un secondo raffreddamento per completare il processo. Al fine di ottenere un completo congelamento, questo ciclo, che prende il nome di *annealing*, può essere eseguito più volte (Octjen e Haseley, 2003).

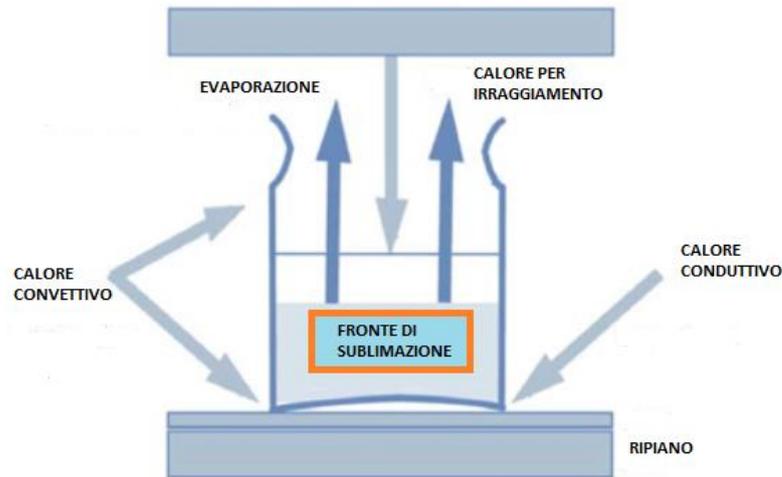


**Figura 1.4:** Diagramma di fase per una miscela acqua (w) – soluto (s) in funzione della frazione massica di soluto.  $T_{g,w}$  indica la temperatura di transizione vetrosa dell’acqua pura,  $T_{g,s}$  del soluto puro,  $T_{m,w}$  indica la temperatura di fusione del ghiaccio,  $T_{m,s}$  quella di fusione del soluto,  $T_e$  indica la temperatura di eutettico.  $T_g'$  indica la temperatura di transizione vetrosa per la miscela in cui è massima la frazione di soluto. La frazione massica  $x_e$  indica la frazione massica di soluto dell’eutettico, ossia la composizione tale per cui la miscela si comporta come un elemento monocomponente,  $x_{g'}$  indica la frazione massica di soluto massima per cui si ha ancora comportamento amorfo della miscela. La linea gialla tratteggiata indica il percorso di congelamento di una miscela contenente un amaro (es. zucchero). Le lettere indicano: a) soluzione sottosatura; b) soluzione + solido; c) ghiaccio + soluzione; d) ghiaccio + solido; e) soluzione sovrasatura; f) zona allo stato gommoso; g) zona vetrosa. (Immagine tratta, con modifiche, da Liu, 2016).

### 1.2.2 Prima fase di essiccamento: la sublimazione del ghiaccio

In questa fase vengono rimossi, per sublimazione, tutti i cristalli di ghiaccio e una parte dell’acqua rimasta in fase liquida. Per far avvenire ciò è necessario fornire calore ed essere in condizioni di vuoto spinto. Il calore viene trasmesso dai ripiani al prodotto tramite conduzione o irraggiamento (Figura 1.5), attraversando tutti gli strati (gap tra parete e contenitore, parete del contenitore e parti di prodotto già essiccate) presenti prima del fronte di sublimazione (Pikal *et al.*, 1984).

Durante il processo la temperatura del liofilizzato deve rimanere al di sotto di quella di fusione all’eutettico ( $T_e$ ) oppure al di sotto della temperatura di transizione vetrosa ( $T_g$ ). Nel caso in cui, per un solido cristallino o per un solido amorfo, questa condizione non venga rispettata, si instaura un flusso di materia che provoca un cedimento strutturale dei pori formati nelle regioni già essiccate adiacenti al fronte di sublimazione. Questo fenomeno viene chiamato *collasso*, e si verifica tipicamente a temperature prossime alla  $T_g$  o alla  $T_e$ ; un ciclo di liofilizzazione ottimale lavora in condizioni di temperatura prossime a quella di collasso, così da portare alla formazione di una struttura porosa omogenea con buone proprietà di reidratazione.



**Figura 1.5:** Rappresentazione schematica di una fiala cui il calore viene fornito per irraggiamento, convezione e conduzione. Immagine tratta, con modifiche, da [www.spscientific.com/freeze-drying-lyophilization-basics](http://www.spscientific.com/freeze-drying-lyophilization-basics).

Per ciò che concerne il calore apportato tramite irraggiamento si deve considerare che si può creare eterogeneità nel prodotto. Bisogna considerare, ad esempio, che le fiale al centro della camera saranno molto più schermate rispetto a quelle ai lati (Hottot *et al.*, 2005). Nel percorso che il flusso di calore deve compiere per arrivare al fronte di sublimazione possono esservi numerose resistenze, come spazi riempiti da vapore o strati di prodotto congelato

Oltre alla pressione della camera, anche la forma della fiala e la sua grandezza influenzano il flusso di calore (Pikal *et al.*, 1984).

Il processo di sublimazione comporta un flusso irreversibile di massa, quantità di moto ed energia, che occorre in due differenti regioni: all'interno del prodotto e nello spazio di vuoto della camera. Nella regione del prodotto si individuano due sezioni: quella contenente prodotto ancora congelato e quella porosa in cui l'acqua è già sublimata; a separarle c'è ciò che viene chiamato fronte di sublimazione, che nel tempo si ritrae progressivamente (Figura 1.5).

### 1.2.3 Essiccamento secondario

A seguito della sublimazione del ghiaccio, rimane ancora una piccola quantità di acqua da eliminare dal prodotto. L'essiccamento secondario è la parte di processo in cui, per desorbimento, si riesce a ridurre alla quantità desiderata il contenuto di solvente all'interno del liofilizzato. Generalmente l'essiccamento secondario dura molto meno del primario, e inizia quando tutto il ghiaccio presente dopo la fine della fase di *congelamento*, risulta sublimato. Alla fine di questa fase il contenuto residuo di acqua rispetto alla massa secca di prodotto è tra il 3% e lo 0.3%. Tuttavia non si può stabilire un confine netto tra le due fasi, poiché entrambe coesistono durante tutto il ciclo di essiccamento.

A livello fisico questa parte di processo inizia a svilupparsi quando il ghiaccio è sublimato e, nelle regioni porose che si sono formate nel prodotto, l'acqua rimasta disciolta nel liofilizzato inizia a desorbirsi dando origine a un flusso di materia all'interno di queste regioni. Per un solvente che ha dato origine a un solido cristallino, il contenuto di umidità residuo al termine della sublimazione si trova allo stato liquido ed è costituito da molecole che sono rimaste legate al soluto tramite forze di tipo chimico o elettrostatico. Il desorbimento, con conseguente deflusso dell'acqua residua, è regolato dalla concentrazione di equilibrio, funzione della natura del solvente e del materiale solido considerato, che essa possiede a una determinata temperatura e, oltre a questo, è regolato anche dalla superficie porosa disponibile: più sono grandi i cristalli di ghiaccio formati e sublimati, più lento sarà l'essiccamento secondario.

Nel caso in cui la miscela congelata sia un amorfo, i tempi di essiccamento secondario si allungheranno, poiché l'acqua non congelata può essere presente fino al 40% in peso su base secca (Pikal *et al.*, 1990), e il trasporto molecolare deve avvenire attraverso la fase amorfa stessa. Per solidi amorfi l'essiccamento secondario può durare fino a metà del tempo richiesto per il primario (Kharaghani *et al.*, 2017).

Per fornire l'energia necessaria al desorbimento, fenomeno endotermico, si alza la temperatura dei ripiani su cui sono poste le fiale di liofilizzato. Bisogna prestar attenzione per gli amorfi a non superare la  $T_g$  del prodotto, per non correre il rischio di transizione da una matrice solida a una viscosa, in cui ci potrebbe essere collasso della struttura porosa; per i solidi cristallini invece questo rischio è minore.

In campo farmaceutico, al fine di avere un prodotto il più stabile possibile, il contenuto di acqua alla fine dell'essiccamento secondario dovrà essere inferiore all' 1% (Hagemann, 1988), ma ciò va bilanciato con l'esigenza che la temperatura nella zona più calda della fiala non superi la temperatura di combustione spontanea (*scorch temperature*) del liofilizzato, sopra la quale possono verificarsi reazioni di inattivazione del farmaco. Il processo ideale riesce quindi a combinare condizioni di basso stress termico con condizioni di elevata rimozione di acqua.

### 1.3 Liofilizzazione di biofarmaci

Il termine *biofarmaci* è stato coniato negli anni Ottanta per descrivere prodotti generati dalle moderne biotecnologie e per distinguerli dai farmaci tradizionali. Rappresentano ormai una fetta importante del mercato e sono costituiti principalmente dai vaccini, costituiti da proteine come immunoglobine, o altre biomolecole estratte direttamente da materiale biologico (Walsh, 2014).

Fino a dicembre 2018 sono stati autorizzati e immessi sul mercato 316 biofarmaci diversi (Walsh, 2018) in Unione Europea e Stati Uniti e nel solo 2017 le vendite globali hanno raggiunto i 155 miliardi di dollari di ricavi.

Nel periodo 2014-2018 hanno ricevuto licenza per l'immissione in commercio 155 nuovi prodotti biofarmaceutici. Di questi 33 destinati a cure contro il cancro, 24 contro malattie legate a infiammazioni, 16 contro l'emofilia e 15 contro il diabete, altri utilizzati contro asma, HIV ed emicrania.

Come si può vedere in Figura 1.6 l'utilizzo dei biofarmaci è in continua crescita, e per questo motivo si stanno ottimizzando sempre più le tecniche di ricerca e produzione per immettere sul mercato prodotti sempre più efficaci.

L'importanza e la novità dei biofarmaci sono legate al fatto che utilizzano come agente curante le proteine, che sono le molecole più dinamiche nel nostro organismo, e si adattano a svolgere diversi ruoli; ad esempio trasportano altre molecole da un organo ad un altro, oppure catalizzano reazioni biologiche, assumendo così una azione altamente specifica nel combattere una determinata malattia.

Cercando di suddividere i biofarmaci in macrocategorie si può dapprima individuare una distinzione tra quelli che contengono proteine con conformazione simile a quella nativa, e quelli contenenti proteine prodotte tramite tecnologia del DNA ricombinante.

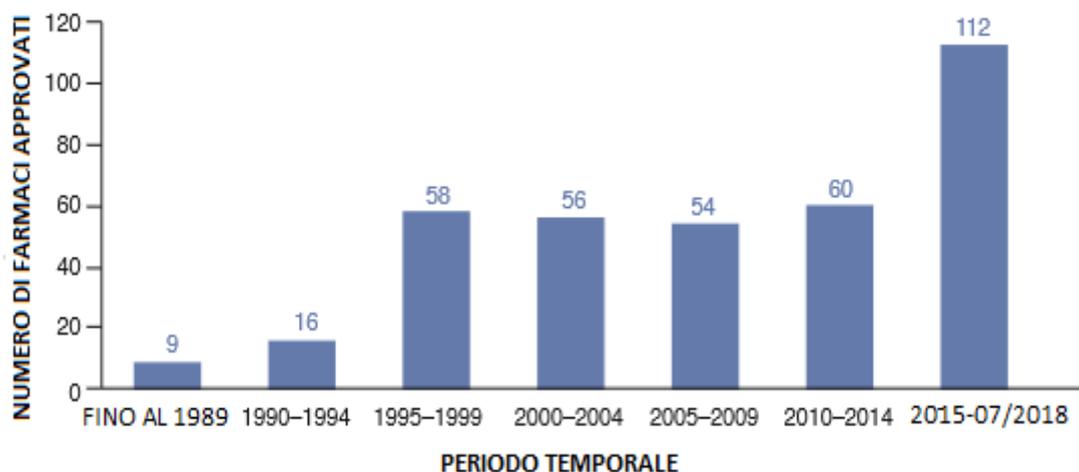
Suddividendo invece in base alla funzione e applicazione terapeutica, i biofarmaci possono esser così suddivisi (Leader, 2008):

- Proteine terapeutiche ad attività regolatoria o enzimatica (*gruppo I*)
- Proteine terapeutiche a target specifico (*gruppo II*)
- Vaccini contenenti proteine (*gruppo III*)
- Proteine per attività diagnostiche (*gruppo IV*)

Appartenenti al *gruppo I* vi sono i farmaci che contribuiscono a sostituire una proteina non funzionante, provvedono a una nuova funzione per l'organismo, migliorano lo svolgimento di una determinata attività biologica. I biofarmaci del secondo gruppo servono da carrier per determinate molecole o proteine o interferiscono con molecole già presenti nell'organismo. Appartenenti al *gruppo III* sono i vaccini che vengono utilizzati per proteggere contro patogeni, curare malattie autoimmuni, svolgere funzione antitumorale. Il quarto gruppo comprende infine proteine sviluppate per effettuare attività diagnostica e aiutare il processo decisionale di cura.

Il principale problema dei biofarmaci è la delicatezza dei loro componenti e la conseguente difficoltà nella conservazione e anche nella somministrazione. L'instabilità proteica infatti è il motivo principale per cui i biofarmaci sono somministrati attraverso iniezione e non per via orale (Wang, 2000). Inoltre la forma solida renderebbe molto più competitive le vie di conservazione e trasporto degli stessi, ma a causa degli stress cui un processo come la liofilizzazione, utilizzato per essiccare un farmaco a partire dalla forma liquida, può sottoporre le proteine, spesso si preferisce lo stoccaggio in forma liquida, nel caso in cui, ovviamente, il farmaco sia stabile per un tempo sufficientemente lungo (Wang, 2000).

Fondamentale quindi diventa comprendere i meccanismi di denaturazione delle proteine durante i processi mirati a rendere i biofarmaci sempre più economicamente competitivi, a partire dalle reazioni che portano a denaturazione delle stesse, correlate agli stress e alle condizioni cui si trovano sottoposte.



**Figura 1.6** Numero di biofarmaci approvati per l'immissione nel mercato UE e USA. Immagine tratta, con modifiche, da Walsh, 2018

### 1.3.1 Struttura delle proteine

Le proteine sono dei polimeri di amminoacidi in cui ogni monomero è legato a quello vicino tramite un legame covalente. Per descrivere il monomero spesso si usa anche il termine *residuo* (Lehninger *et al.*, 2005).

Gli amminoacidi presenti in natura sono 20, e sono accomunati dalla presenza di una struttura, chiamata *backbone*, composta da un atomo di carbonio al centro, legato a un gruppo carbossilico (-COOH), a un gruppo amminico (-NH<sub>2</sub>) e a un atomo di idrogeno (-H). Il quarto legame del carbonio è con un gruppo caratteristico, diverso per ogni amminoacido. Questo gruppo, chiamato anche *catena laterale* o *sidechain* conferisce specifiche proprietà al residuo, che lo differenziano dagli altri. Vista la formazione di legami con quattro sostituenti diversi, il carbonio centrale è un centro chirale, e per questo ogni amminoacido presenta due configurazioni speculari, chiamate enantiomeri, della stessa molecola, definite con le lettere *L-* o *D-* (Lehninger *et al.*, 2005).

Come criterio di classificazione degli amminoacidi, in letteratura e conseguentemente nel presente lavoro, si adotta una suddivisione basata sulla tipologia di catena laterale di ognuno di essi. Le categorie sono le seguenti (Figura 1.7):

- *amminoacidi non polari*: alanina, valina, leucina, isoleucina, glicina, prolina
- *amminoacidi polari*: serina, treonina, cisteina, asparagina, glutammina, metionina
- *amminoacidi aromatici*: fenilalanina, tirosina, triptofano
- *amminoacidi carichi positivamente*: lisina, arginina, istidina
- *amminoacidi carichi negativamente*: aspartato e glutammato

La categoria degli *amminoacidi carichi* comprende quelli aventi carica netta non nulla a pH pari a 7. Infatti, al variare del pH, i residui possono avere sia una complessiva carica negativa, che una complessiva carica positiva, a causa della presenza sia del gruppo carbossilico che del gruppo amminico (Figura 1.8). Il punto in cui l'amminoacido ha

carica complessiva uguale a zero, e si comporta come *zwitterione* o *anfotero*, si chiama *punto isoelettrico*, ed è caratteristico per ogni amminoacido (Lehninger *et al.*, 2005).

Il legame covalente che tiene uniti i residui in una catena polipeptidica è il legame *peptidico*, che si forma tra il gruppo carbossilico di un amminoacido e il gruppo amminico dell'amminoacido seguente, con eliminazione di acqua. Queste catene polipeptidiche possono avere diversa lunghezza e sequenza di amminoacidi, ma tutte sono accumulate dalla presenza di una certa serie di residui interni, una estremità con solo un gruppo amminico ed una con solo un gruppo carbossilico.

La struttura formata da una sequenza generica di amminoacidi viene definita *struttura primaria* delle proteine, e tiene conto solamente del tipo di residui e del loro susseguirsi in una catena polipeptidica. Tenendo conto anche della configurazione spaziale della catena principale degli amminoacidi (*backbone*) si parla di *struttura secondaria* della proteina. Il contributo a questa struttura non è solamente dato dai legami peptidici tra i residui, ma anche da legami di tipo Van Der Waals, legami idrogeno e ponti disolfuro, che si possono formare tra residui non consecutivi, e che possono conferire alla catena amminoacidica differenti disposizioni spaziali. Le più comuni strutture secondarie sono l' $\alpha$ -elica e il  $\beta$ -foglietto (Figura 1.9). In natura è comune trovare delle strutture  $\beta$ -foglietto parallele le une alle altre e legate da un legame peptidico ad una delle estremità, così da formare la struttura secondaria denominata  $\beta$ -hairpin. Esempi di strutture  $\beta$ -foglietto sono le proteine globulari (Lehninger *et al.*, 2005).

Se si considerano non solo la disposizione spaziale in una catena e le loro interazioni intra-molecolari, ma anche le interazioni tra gli atomi di una catena con quelli di una catena adiacente, si definisce la cosiddetta *struttura terziaria* delle proteine. Un insieme di strutture terziarie in complessi tridimensionali dà origine alla *struttura quaternaria* delle proteine, l'ultima tipologia di struttura e anche la più complessa. Spesso all'interno di una struttura quaternaria possono essere presenti eteroatomi, come a esempio il ferro nel caso dell'emoglobina.

Principalmente le proteine possono essere distinte in *proteine fibrose* e *proteine globulari*. Le prime sono caratteristicamente disposte in lunghi fasci o foglietti, le seconde hanno invece catene polipeptidiche ripiegate in forma globulare o sferica (Lehninger *et al.*, 2005).

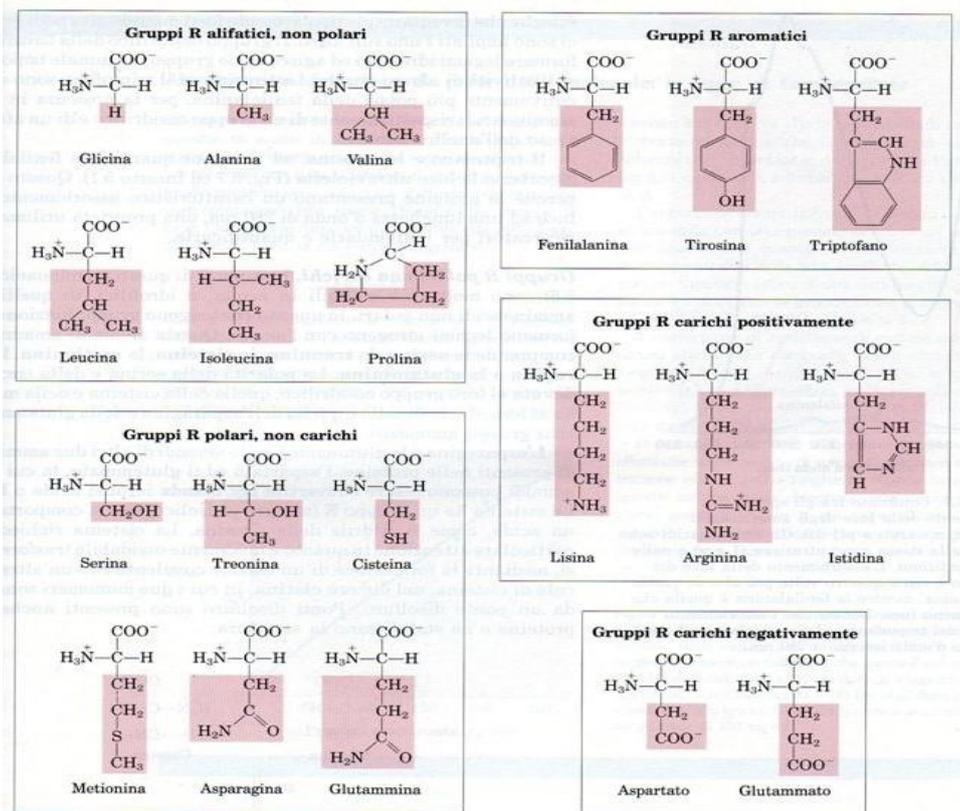


Figura 1.7 Classificazione degli amminoacidi. Immagine tratta con modifiche da Lehninger et al., 2005.

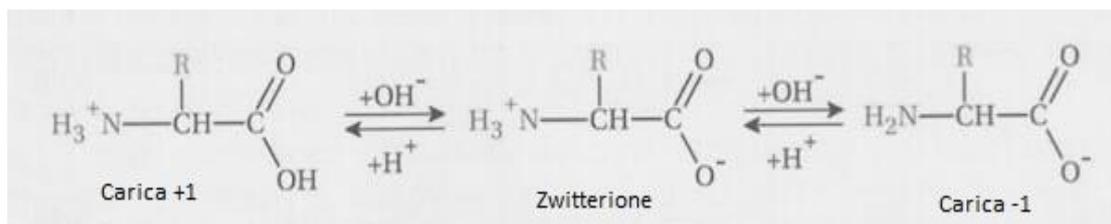
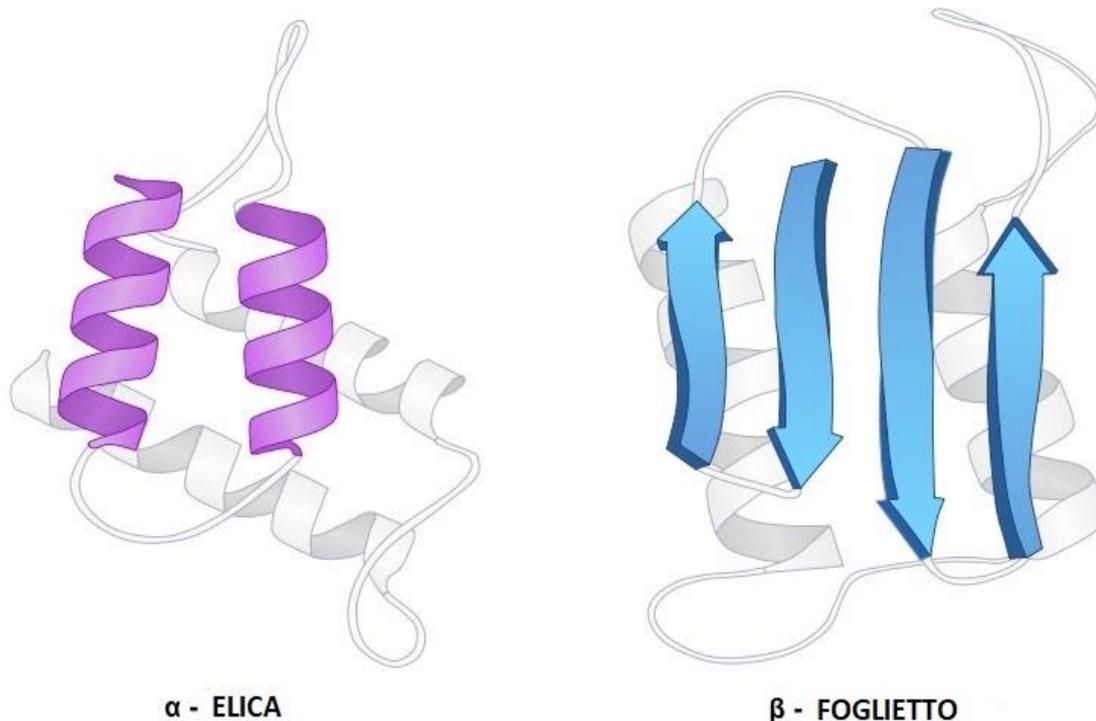


Figura 1.8 Forma acida, neutra e basica di un generico amminoacido. Immagine tratta con modifiche da Lehninger et al., 2005.



**Figura 1.9** Struttura  $\alpha$ -elica e  $\beta$ -foglietto. Immagine tratta con modifiche da <https://ib.bioninja.com.au/standard-level/topic-2-molecular-biology/24-proteins/protein-structure.html>

## 1.4 Denaturazione nelle proteine

La difficoltà principale nel formulare un prodotto farmaceutico è rappresentata dal fatto che, per portarlo alla forma solida, il processo di liofilizzazione implica numerosi stress che rischiano di modificarne o annullarne l'attività. Per quanto la forma in soluzione acquosa di un farmaco proteico sia la più facile da preparare, spesso, per rendere il prodotto competitivo in termini di stoccaggio, shelf life, trasporto e somministrazione è necessario ricondurlo allo stato solido, cercando di superare le difficoltà che si possono verificare a livello di progettazione del processo e della formulazione (Wang, 2000).

Uno dei fenomeni biologici che si verificano più frequentemente durante il processo di liofilizzazione è l'*unfolding* delle proteine nella formulazione a seguito degli stress cui esse sono sottoposte durante il processo, che può condurre alla denaturazione irreversibile del farmaco (Prestelski *et al.*, 1993).

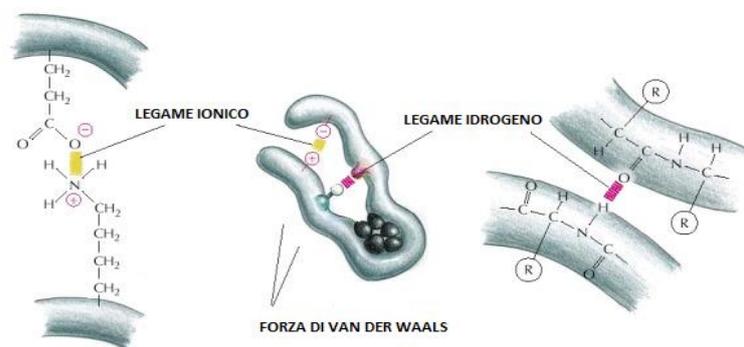
L'*unfolding* di una proteina corrisponde alla perdita di legami intermolecolari (figura 1.10) tra le catene polipeptidiche che le conferiscono una struttura ripiegata nello spazio (*struttura secondaria*). Oltre a questi legami, un'altra forza esercita un ruolo centrale nell'indurre le proteine ad assumere una certa conformazione spaziale: la presenza di acqua. L'acqua infatti tende a creare legami idrogeno con le sidechain degli amminoacidi (figura 1.11), qualora queste siano polari, mentre il backbone, apolare, tende a essere escluso dal solvente.

La *denaturazione* di una proteina invece consiste nella perdita della struttura tridimensionale tale per cui essa perda le proprie funzioni biologiche (Lehninger *et al.*,

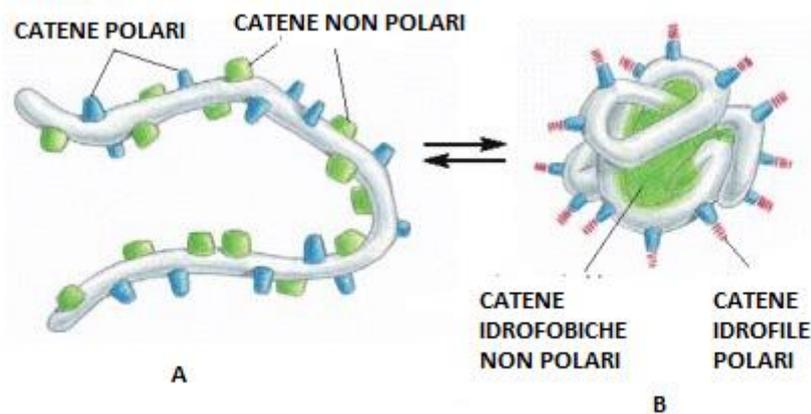
2005). La denaturazione non sempre corrisponde a un completo unfolding della proteina, ma spesso questo può esserne il motivo principale. La denaturazione avviene in genere a temperature molto alte, a valori estremi di pH, in presenza di cambiamenti di fase/rimozione del solvente attorno alla proteina (Wang, 2000).

Al fine di evitare l'unfolding è necessario studiare nel dettaglio le condizioni, come ad esempio il pH, in cui condurre la liofilizzazione e introdurre nella formulazione del farmaco alcune molecole, chiamate *eccipienti*, che possano stabilizzare la soluzione e proteggere le proteine (Wang, 2000). Oltre a ciò vanno anche ottimizzate la  $T_g$  della soluzione e il contenuto di acqua residuo a fine processo (Carpenter e Chang., 1996).

Altri fenomeni che portano alla denaturazione delle proteine possono essere l'aggregazione o la denaturazione all'interfaccia. L'aggregazione può essere dovuta a eccessivi tempi di stoccaggio o stress indotti dal processo di liofilizzazione, e porta alla formazione di aggregati di proteine che, nella nuova conformazione, perdono le loro proprietà denaturandosi. I fenomeni di denaturazione all'interfaccia invece sono dovuti alla concentrazione delle proteine in zone in cui la tensione superficiale comporta stress maggiori rispetto al bulk come, ad esempio, le superfici dei cristalli di ghiaccio durante congelamento.



**Figura 1.10** Tipologie di legami secondari: Legami polari, forze di Van Der Waals, legami a idrogeno. Immagine tratta, con modifiche, da [oregonstate.edu/instruction/bi314/fall11/cellchemistrycontd.html](http://oregonstate.edu/instruction/bi314/fall11/cellchemistrycontd.html)



**Figura 1.11** *Struttura proteica unfolded (sinistra) e ripiegata (destra). Immagine tratta, con modifiche, da [biologicalphysics.iop.org/cws/article/lectures/46262.html](http://biologicalphysics.iop.org/cws/article/lectures/46262.html)*

#### 1.4.1 Stabilizzazione delle proteine tramite eccipienti

Il primo step per valutare l'efficacia di una certa formulazione durante il ciclo di liofilizzazione consiste nel valutare se in essa siano presenti elementi che alterino l'energia libera di *unfolding* della proteina (Chang *et al.*, 1996).

Generalmente a livello di protezione durante la liofilizzazione si possono individuare diverse categorie di eccipienti non specifici che possono stabilizzare la maggior parte delle proteine, come zuccheri e polimeri. Gli zuccheri sono molto utilizzati quindi come eccipienti in ambito farmaceutico. I più utilizzati sono saccarosio, trealosio, lattosio e maltosio (Wang, 2000), con gli ultimi due che possono però causare imbrunimento e danneggiamento della proteina in fase di stoccaggio a seguito della reazione di Maillard, che può essere accelerata dal contenuto residuo di acqua non evaporata in fase di essiccamento secondario (Rey e May, 2010).

In generale una formulazione farmaceutica contiene proteine, agenti di bulk (sostanze aggiunte alla soluzione per rendere più resistente meccanicamente la struttura essiccata a seguito della sublimazione del ghiaccio), *tamponi per la correzione del pH* ed eccipienti. Un eccipiente parzialmente amorfo è più adatto per la stabilizzazione durante l'essiccamento poiché la proteina, a seguito di congelamento, tende ad accumularsi e ad interagire con gli altri composti in fase amorfa, ed è fondamentale che questa interazione venga mantenuta durante l'intero processo di liofilizzazione (Prestelski *et al.*, 1993). Per quanto sia ormai accertato che sia necessario avere un eccipiente amorfo nella formulazione, diverse sono le teorie su quale sia il meccanismo attraverso cui questo riesca a proteggere la proteina dal denaturarsi in fase di essiccamento.

## 1.4.2 Meccanismi di protezione durante la liofilizzazione

Principalmente sono due le proposte per spiegare come un eccipiente amorfo stabilizzi una formulazione durante il processo di liofilizzazione: la prima afferma che le proteine sono semplicemente immobilizzate nella fase amorfa durante tutto l'essiccamento a causa della presenza di altre macromolecole attorno a esse che non ne permettono la denaturazione (Franks *et al.*, 1991); la seconda teoria, spesso chiamata del *water replacement*, afferma che alcuni zuccheri proteggono le proteine dalla denaturazione formando legami a idrogeno con i gruppi polari degli amminoacidi mentre l'acqua viene rimossa. È stato dimostrato, nel caso della seconda teoria, che più uno zucchero è piccolo, maggiore sarà il suo effetto stabilizzante, poiché l'ingombro sterico inferiore renderà più facile formare legami a idrogeno (Tanaka *et al.*, 1991). È stato dimostrato da Tanaka (Tanaka *et al.*, 1991) che il grado di protezione è direttamente proporzionale alla concentrazione del protettore in soluzione, e ciò conferma la teoria del *water replacement* poiché più sono le molecole di zucchero presenti, maggiore è il numero di legami idrogeno che esse possono formare in fase di essiccamento. Inoltre conferma di ciò si trova negli studi effettuati su un polipeptide modello, la poli – L – lisina, in cui è dimostrato che, a prescindere da quali siano le conformazioni iniziali che questa assuma in soluzione, alla fine del processo di liofilizzazione si trova sempre in configurazione  $\beta$ -foglietto, poiché questa permette una maggiore compensazione dei legami idrogeno persi in fase di essiccamento. In presenza di saccarosio, però, lo stesso polipeptide mantiene la configurazione iniziale che aveva in soluzione, poiché il numero di legami idrogeno persi in fase di essiccamento vengono compensati da quelli che si formano con il protettore, così che la proteina non necessita più di un cambiamento di conformazione per compensare il numero di legami scomparsi (Prestelski *et al.*, 1993). La teoria del *water replacement* trova anche fondamento nel fatto che le proteine, allo stato nativo, assumono una conformazione tale per cui le catene laterali (spesso idrofobiche) siano rivolte parzialmente all'interno della struttura, formando legami a idrogeno con altre catene della proteina stessa. Queste interazioni portano a un effetto termodinamicamente destabilizzante nei confronti dello stato nativo nel momento in cui i legami interni non vengono bilanciati all'esterno da simili interazioni della proteina con altri elementi presenti nella soluzione. Si dimostra che la differenza di energia libera favorevole che viene portata dall'interazione delle catene laterali con gli altri elementi presenti nella soluzione, quali possono essere protettori come gli zuccheri, è spesso superiore alla differenza di energia libera tra lo stato nativo e quello denaturato, e da ciò deriva quindi l'effetto stabilizzante (Lehninger *et al.*, 2005).

È necessario considerare gli stress cui una proteina è sottoposta non solo durante l'essiccamento, ma anche durante il congelamento, poiché la denaturazione che potrebbe manifestarsi in questa fase non è reversibile nelle successive fasi del processo (Rey e May, 2010). Per proteggere il prodotto finale è necessario aggiungere degli agenti che in bulk permettano alla proteina di non perdere la propria conformazione; gli zuccheri possono esser considerati ottimi *crioprotettori*, in particolar modo i disaccaridi come trealosio e saccarosio (Carpenter *et al.*, 1987).

I disaccaridi sono di fatto gli unici composti in grado di funzionare sia da *crioprotettori* che da *lioprotettori*, poiché riescono a proteggere le proteine sia in fase di congelamento sia in fase di essiccamento, rimanendo in fase amorfa insieme a esse e sostituendo con efficacia i legami idrogeno durante l'essiccamento (Rey e May, 2010).

In fase di congelamento è inoltre necessario considerare la presenza di stress per la proteina legati al fatto che essa andrà a trovarsi in una frazione di prodotto caratterizzata da una concentrazione molto maggiore di soluti rispetto al bulk. La denaturazione potrebbe essere pertanto causata proprio dalla variazione di forza ionica, pH o composizione indotta dalla crioconcentrazione (Rey e May, 2010). Altro fattore di stress molto importante è quello causato dall'adsorbimento all'interfaccia con il solvente ghiacciato, che può portare anch'esso a denaturazione (Strambini e Gabelleri, 1996).

### 1.4.3 Termodinamica della crioprotezione: teoria dell'esclusione preferenziale

A livello termodinamico la denaturazione di una proteina può essere vista come conseguenza di uno stato in cui l'*unfolding* porti ad avere una minore energia libera di Gibbs. Timasheff e Arakawa hanno proposto una teoria termodinamica che spiega come si possa passare a una conformazione denaturata delle proteine (Arakawa *et al.*, 1993), ma prima di addentrarsi nel dettaglio del suo studio possono esser fatte delle considerazioni più generali.

Una proteina, in soluzione, allo stato nativo, si trova, dal punto di vista termodinamico, in un minimo globale di energia libera. I legami con il solvente sono formati nel momento in cui termodinamicamente questa sia la situazione favorita, mentre in caso contrario non avviene interazione. Per valutare l'effetto dell'eccipiente sulla proteina, si deve tener presente che questo favorisce la conservazione dello stato nativo solo se questa è la situazione termodinamicamente favorevole. Allo stesso modo, l'eccipiente può ricondurre spontaneamente allo stato nativo una proteina denaturata solo se la transizione comporta una variazione negativa del  $\Delta G$  di energia libera. Nel momento in cui una proteina formi comunque dei legami con il solvente, la situazione favorita termodinamicamente è quella in cui si forma il maggior numero di legami, poiché è quella in cui l'energia libera ha valore minore (Rey e May, 2010).

In base alla teoria di Timasheff (Arakawa *et al.*, 1993), sono quasi sempre richieste grandi concentrazioni di crioprotettore ( $> 0.2$  M), poiché le interazioni proteina-soluto sono relativamente deboli. Queste interazioni vengono sperimentalmente determinate tramite l'utilizzo di membrane semipermeabili e misure di potenziale chimico, e quantitativamente esprimono se una proteina interagisca *preferenzialmente* con l'eccipiente o con l'acqua. I soluti come gli zuccheri, che tendono a stabilizzare la proteina, sono preferenzialmente esclusi dalla superficie della proteina, in modo tale da rendere lo stato nativo termodinamicamente favorito, quelli come urea o GdmCl, che tendono a denaturarla, invece interagiscono preferenzialmente con essa e con la sua superficie in modo tale da rendere lo stato denaturato più favorito termodinamicamente rispetto a quello nativo.

La teoria di Timasheff può essere espressa dall'equazione:

$$\left(\frac{\partial \mu_2}{\partial m_3}\right) = -\left(\frac{\partial m_2}{\partial m_3}\right)\left(\frac{\partial \mu_3}{\partial m_3}\right) \quad (1.3)$$

Dove il pedice 2 la proteina, il pedice 3 il soluto,  $m$  è la concentrazione molale del componente,  $\mu$  il potenziale chimico. Il termine  $(\partial m_2/\delta m_3)$  quantifica l'interazione preferenziale del soluto con la proteina a potenziale chimico costante del solvente e del soluto, e un valore positivo di questo parametro indica un eccesso di soluto nei dintorni della proteina. Il termine  $(\partial \mu_2/\delta m_3)$  è legato alla variazione di potenziale chimico della proteina in presenza del soluto. L'equazione esprime dunque che i composti esclusi dalla superficie ( $\partial m_2/\delta m_3 < 0$ ) avranno valori positivi di  $(\partial \mu_2/\delta m_3)$  e aumentano l'energia libera di Gibbs della proteina nel caso in cui si trovino nel suo intorno, aumentando l'instabilità termodinamica. Inoltre, in presenza di soluti esclusi, l'esclusione sarà maggiore in presenza di proteine in forma denaturata, poiché queste espongono una maggiore area superficiale, così che  $(\partial m_2/\delta m_3)_{\text{Denaturata}} < (\partial m_2/\delta m_3)_{\text{Nativa}} < 0$ . Di conseguenza è più stabile la forma nativa, poiché l'instabilità termodinamica è maggiore per le proteine in forma denaturata:  $(\partial \mu_2/\delta m_3)_{\text{Denaturata}} > (\partial \mu_2/\delta m_3)_{\text{Nativa}} > 0$ . Il discorso opposto vale per i soluti che interagiscono favorevolmente con la superficie della proteina (es. urea). Nella notazione utilizzata non compare il pedice 1, che viene utilizzato per indicare il solvente.

In generale, finché  $(\partial m_2/\delta m_3)_{\text{Denaturata}} < (\partial m_2/\delta m_3)_{\text{Nativa}}$ , la forma nativa viene stabilizzata, implicando così che la stabilizzazione può avvenire anche quando il soluto interagisce favorevolmente con la forma nativa, ma risulta escluso dalla forma denaturata, o quando il soluto interagisce favorevolmente con entrambe le forme, ma in maniera meno pronunciata con la forma denaturata (Rey e May, 2010).

Il termine  $(\frac{\partial \mu_3}{\delta m_3})$  può esser così espresso (Arakawa *et al.*, 1993):

$$\left(\frac{\partial \mu_3}{\delta m_3}\right) = \left[\left(\frac{RT}{m_3}\right) + RT\left(\frac{\partial \ln \gamma_3}{\delta m_3}\right)\right] \quad (1.4)$$

Dove  $\gamma_3$  è il coefficiente di attività del soluto, R la costante universale dei gas perfetti.

#### 1.4.4 Interazione proteina-eccipiente nella fase congelata

A oggi non si può misurare direttamente l'interazione preferenziale tra proteina e soluti in fase solida, non potendo quindi stabilire se i crioprotettori siano esclusi o meno dalle proteine. Tuttavia è stato dimostrato con l'equazione 1.3 che l'influenza del soluto sul potenziale chimico della proteina in soluzione fornisce informazioni su quale sia il suo effetto stabilizzante anche in fase congelata. In particolare, maggiore è l'innalzamento del potenziale, maggiore è l'effetto stabilizzante del crioprotettore. Si potrebbe quindi ipotizzare che sia possibile aggiungere concentrazioni sempre maggiori di soluto così da stabilizzare ulteriormente la proteina. In realtà ciò non è possibile per il fatto che il potenziale può incrementare sino a un certo livello, una volta superato il quale sono termodinamicamente favoriti fenomeni di precipitazione delle proteine (Rey e May, 2010).

### 1.4.5 Protezione durante l'essiccamento e lo stoccaggio

A causa della loro complessità, le proteine spesso sono affette da instabilità. Nel momento in cui si voglia creare un farmaco con shelf life sufficiente per essere competitivo sul mercato è necessario rimuovere dalla formulazione la fase acquosa, poiché in tempi elevati può essere essa stessa causa di denaturazione (Rey e May, 2010). La liofilizzazione è dunque un processo mirato a incrementare la stabilità di un determinato farmaco, ma anch'essa presenta problemi legati al fatto che in fase di essiccamento gli stress cui le proteine sono sottoposte possono danneggiarle irreversibilmente (Wang, 2000).

In ambito farmaceutico il termine *stabilità* si riferisce alla capacità di una proteina di essere processata e utilizzata senza cambiamenti nella propria struttura primaria, conformazione e stato di aggregazione. La locuzione *stabilità della proteina* indica invece la posizione di equilibrio tra la forma nativa e quella denaturata. Se una proteina necessita di alte concentrazioni di denaturante o alte temperature per passar dalla forma nativa a quella denaturata viene definita *stabile*, e questa accezione di stabilità viene definita *stabilità termodinamica* (Rey e May, 2010). I tempi della destabilizzazione termodinamica sono nell'ordine di ore e coinvolgono cambiamenti a livello fisico, quelli della destabilizzazione farmaceutica riguardano periodi di durata superiore e possono anche portare a degradazione chimica. I due tipi di stabilità non sono necessariamente correlati, dal momento che una proteina può rimanere in forma nativa per lunghi tempi, tuttavia perdere la propria attività a seguito di reazioni chimiche; se una proteina invece è destabilizzata dal punto di vista termodinamico, può essere favorevolmente portata a situazioni di destabilizzazione durante il periodo di stoccaggio, in quanto possono insorgere fenomeni di aggregazione tipici della degradazione farmaceutica (Rey e May, 2010).

Pur non essendo ancora stata trovata una teoria condivisa riguardo i meccanismi di stabilizzazione durante l'essiccamento, si possono individuare due categorie principali: la *stabilizzazione termodinamica* e la *stabilizzazione cinetica*. Nella prima rientrano le teorie del *water replacement* e del *soluto escluso* elencate in precedenza, che prevedono una interazione eccipiente – proteina tale per cui lo stato nativo sia sempre favorito, e tempi di occorrenza molto minori (minuti/ore) rispetto a quelli dell'essiccamento (ore); nella seconda rientrano le teorie come quella della *vitrificazione*, che prevede l'immobilizzazione delle proteine allo stato vetroso, così che queste non diano origine a fenomeni di aggregazione o denaturazione, se non in tempi molto lunghi, spesso superiori alla durata dell'essiccamento (Franks, 1990). Si può affermare inoltre che la stabilizzazione tramite *water replacement* possa essere considerata anche una stabilizzazione di tipo cinetico, poiché molti protettori, formando legami a idrogeno, rendono molto lento l'*unfolding*, anche se favorito in modo termodinamico, portando il tempo di denaturazione a una durata superiore a quello di essiccamento (Rey e May, 2010). Non essendoci quindi una teoria univoca su quali fenomeni stabilizzino prevalentemente una proteina durante l'essiccamento, per ogni soluzione è necessario valutare le proprietà allo stato amorfo e le interazioni tra proteina e soluto, così da comprendere quali meccanismi di denaturazione siano prevalenti e come strutturare la formulazione.

### 1.5 Instabilità allo stato amorfo

Il processo di formazione del ghiaccio comporta una maggior concentrazione di soluto e un aumento della viscosità della formulazione. Durante il congelamento, la maggior parte dell'acqua viene allontanata dalla frazione proteica della soluzione, che rimane amorfa, e

si possono instaurare delle tensioni che portano alla denaturazione. In prima battuta ci si può quindi aspettare che, diminuendo la presenza di acqua, la proteina sia maggiormente esposta a fenomeni di *unfolding*, o possa entrare più facilmente in contatto con altri soluti che portino a reazioni di degradazione. In realtà è stato dimostrato che un elevato abbassamento della temperatura influenza favorevolmente la stabilità della formulazione poiché, più la temperatura diminuisce, più aumenta la viscosità del sistema e i processi di degradazione sono fortemente rallentati, poiché controllati dal *mass transfer* (Franks, 1990). Si deve poi considerare l'effetto della temperatura sul *tempo di rilassamento*, ossia il tempo che la struttura amorfa impiega per portarsi alla configurazione di equilibrio. Esistono due *tempi di rilassamento*; uno, più lungo, legato alla velocità con cui si muovono le componenti della matrice amorfa fino a che non si raggiunga una situazione di equilibrio a seguito di cambiamenti fisici del sistema (disidratazione); uno invece, più breve, legato al tempo di reazione, ossia al tempo in cui si verificano alcune reazioni di degradazione, o alcune variazioni conformazionali. In entrambi i casi un eccipiente che garantisca un *tempo di rilassamento* maggiore è migliore per la stabilità della proteina, poiché il tempo di rilassamento è inversamente proporzionale alla costante cinetica per le reazioni di degradazione, e ne rallenta quindi la cinetica (Rey e May, 2010).

I due tempi di rilassamento sono in genere relazionati con la viscosità e la temperatura del sistema dalla relazione di Stokes – Einstein (Rey e May, 2010):

$$\frac{z^2}{2\tau} = \frac{k_b T}{6\pi\eta a} \quad (1.5)$$

Dove  $z$  indica la distanza spaziale dei salti diffusionali considerati,  $k_b$  è la costante di Boltzmann,  $T$  è la temperatura del sistema,  $\tau$  è il tempo di rilassamento della struttura,  $\eta$  è la viscosità del sistema,  $a$  è il raggio idrodinamico delle specie diffusive. I salti diffusionali sono le modifiche strutturali che possono avvenire in fase amorfa a seguito della diffusione e della mobilità delle molecole. Possono portare a configurazioni tali per cui avvengano reazioni chimiche o cambiamenti fisici all'interno dell'amorfo considerato (Rey e May, 2010).

A seguito del congelamento inoltre vanno valutati gli stress generati all'interfaccia con la superficie dei cristalli di ghiaccio in formazione, che possono portare a denaturazione della proteina. Maggiore è la superficie acqua-ghiaccio esposta, maggiore è il rischio di denaturazione; per questo nelle formulazioni si aggiungono spesso dei tensioattivi, che riducono l'assorbimento delle proteine all'interfaccia e così facendo le rendono meno esposte al rischio di denaturazione (Wang, 2000).

### 1.5.1 Instabilità durante lo stoccaggio

Gli stress presenti in fase di stoccaggio sono i medesimi presenti in fase di essiccamento, ma vanno valutati su tempi ovviamente molto più lunghi. In questo caso il tempo di rilassamento della struttura risulta molto maggiore, e i tempi di denaturazione diventano molto lunghi.

In realtà si deve tener conto che le proteine dopo il processo di liofilizzazione hanno subito delle perturbazioni che ne hanno alterato la struttura e per questo, pur a temperatura molto più bassa della  $T_g$ , possono incorrere nella denaturazione, poiché il numero di salti diffusionali necessari è molto minore rispetto a quello di una proteina allo stato nativo non sottoposta a liofilizzazione (Carpenter *et al.*, 1997).

Ai fini del controllo del processo in fase di stoccaggio vanno quindi valutati gli eccipienti migliori non per una stabilizzazione di tipo termodinamico, bensì per una stabilizzazione di tipo cinetico (Rey e May, 2010).

## 1.6 Reazioni di denaturazione

A livello chimico sono diverse le reazioni che possono coinvolgere i gruppi funzionali degli amminoacidi costituenti le proteine, modificandone la struttura molecolare. Le reazioni più comuni per le proteine sono quelle di idrolisi, ossidazione, deamidificazione,  $\beta$  – scissione, aggregazione e, in presenza di zuccheri riducenti, la reazione di Maillard (Wang, 2000).

- La reazione di idrolisi, chiamata anche *proteolisi*, consiste in una serie di passi che portano a una rottura del legame peptidico in presenza di acqua e in ambiente con pH molto alto o molto basso (Pham *et al.*, 2014).
- La reazione di ossidazione invece coinvolge le proteine in presenza di ossigeno e porta a frammentazione delle proteine sia a livello di backbone che di sidechain degli amminoacidi e inter-ricombinamento tra esse (Zhang *et al.*, 2013).
- La deamidificazione invece è una reazione in cui un gruppo funzionale ammidico ( $-NH_2$ ) nella sidechain viene rimosso e sostituito con un altro gruppo funzionale ed è influenzata principalmente dal pH e dalla temperatura. Coinvolge gli amminoacidi Asparagina e Glutamina (Robinson, 2002).
- La  $\beta$  – scissione è una reazione che coinvolge i gruppi alchilici delle sidechain e porta anch'essa a frammentazione dei polipeptidi. Avviene in presenza di ossigeno e vengono rilasciati dei gruppi carbonilici ( $R - C = O - R'$ ) derivanti dall'ossidazione dei gruppi alchilici (Headlam e Davies, 2002).
- La reazione di aggregazione bimolecolare avviene quando due proteine diverse, a seguito di movimenti all'interno della formulazione causati da temperature più alte del necessario oppure da fenomeni di dinamica molecolare, si aggregano formando un nuovo composto (Lapidus, 2003).
- La reazione di Maillard coinvolge invece zuccheri riducenti, come il glucosio, e amminoacidi liberi nella formulazione, in presenza di alte temperature e specifiche condizioni di pH. Porta a un imbrunimento della formulazione. Nella realtà la reazione consiste in una complessa serie di reazioni consecutive (Rey e May, 2010).

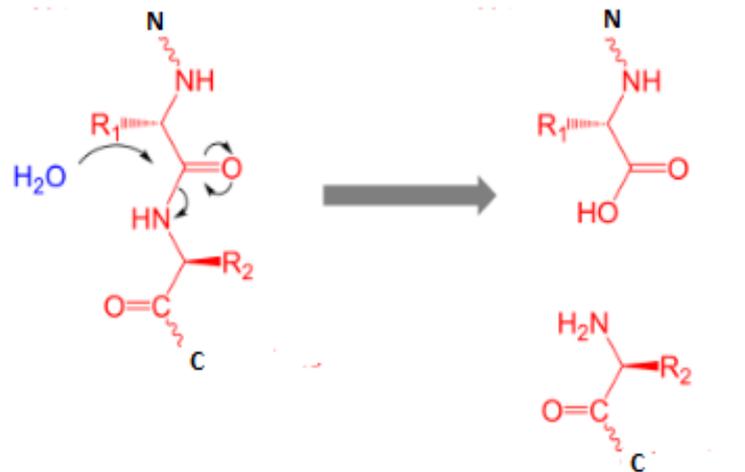


Figura 1.12 Reazione di proteolisi. Immagine tratta con modifiche da wikipedia.org

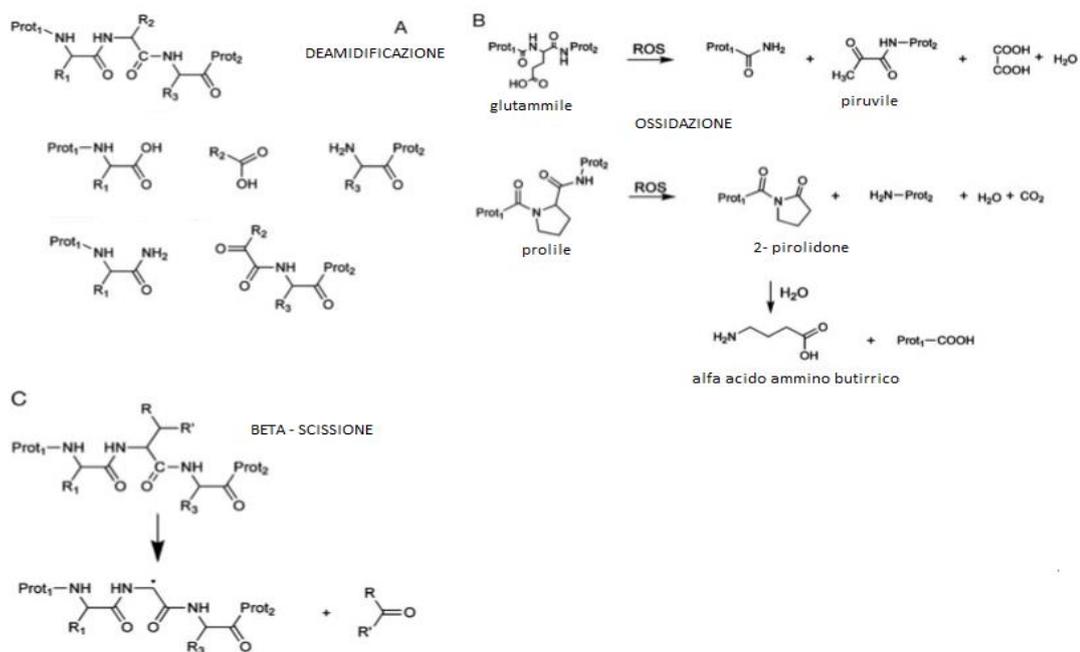


Figura 1.13 Meccanismo di deamidificazione (A), meccanismo di frammentazione per ossidazione (B), meccanismo di  $\beta$  - scissione (C). Immagine tratta con modifiche da Barelli, 2008.

## 1.7 Obiettivi della tesi

Il presente lavoro di tesi, tramite l'utilizzo del software di dinamica molecolare GROMACS, si occupa di valutare alcuni aspetti qualitativi dei meccanismi di lioprotezione dei farmaci, valutati a oggi in modo quasi esclusivamente sperimentale. L'obiettivo principale è quello di verificare se sia possibile predire l'efficacia di uno zucchero nella protezione di un principio attivo polipeptidico, e di valutare se l'interazione tra un generico polipeptide e protettore sia additiva rispetto all'interazione dei residui che lo compongono con il protettore medesimo. Si struttura il lavoro simulando le interazioni di ognuno dei venti amminoacidi con saccarosio o trealosio in soluzione acquosa, per poi valutare nelle medesime soluzioni le proprietà di alcuni polipeptidi presi come campione. I risultati, confrontati con quelli sperimentali, possono essere immediatamente verificati per dimostrare l'effettivo raggiungimento dell'obiettivo.

Il lavoro propone quindi una parte di introduzione alla dinamica molecolare e descrizione dei metodi computazionali e teorici utilizzati per valutare le proprietà degli amminoacidi e dei polipeptidi. L'ultima parte riguarda invece la presentazione dei risultati ottenuti, con commenti riguardo il confronto con i dati teorici e i possibili sviluppi futuri sulla dinamica molecolare applicata all'additività nei polipeptidi.



## 2. METODI: APPROCCIO COMPUTAZIONALE E DESIGN DELLE SIMULAZIONI

In questo capitolo si spiegano alcuni concetti fondamentali della dinamica molecolare, utilizzata per il presente lavoro di tesi, e si presentano i dettagli computazionali delle simulazioni effettuate.

### 2.1 Introduzione alla Dinamica Molecolare

La dinamica molecolare (spesso abbreviata in *MD*) è una branca della chimica fisica che si occupa di analizzare le interazioni tra le molecole a livello atomico, studiando le forze che si sviluppano tra esse e le traiettorie da loro assunte. Per quanto le dimensioni dei sistemi analizzati siano molto ridotte (si simulano *box* di dimensioni nanometriche) l'approccio utilizzato dai programmi di *MD* è quello della meccanica classica, poiché i calcolatori non riuscirebbero a risolvere in tempi ragionevoli le equazioni legate alla meccanica quantistica. In ogni caso, considerazioni di tipo quantistico sono implicitamente incluse nel calcolo delle forze di interazione presenti tra gli atomi simulati.

Una simulazione di dinamica molecolare consiste nell'integrazione, *step by step*, della equazione del moto elaborata da Newton:

$$F_i = m_i a_i \quad (2.1)$$

Con:

$$F_i = - \frac{d}{dr_i} V \quad (2.2)$$

Dove  $F_i$  è la forza agente su ogni atomo  $i$ -esimo presente nel sistema e viene derivata dal potenziale  $V(r^N)$ , dove  $r^N = (r_1, r_2, \dots, r_N)$  rappresenta le tre coordinate spaziali per ogni molecola presente nel sistema.

La funzione potenziale è legata a diversi tipi di interazione. Per quanto riguarda le interazioni tra molecole non legate si prevede un potenziale nella seguente forma:

$$V_{non\ legata}(r^N) = \sum_i u(r_i) + \sum_i \sum_{j>i} u(r_i, r_j) + \dots \quad (2.3)$$

Il termine  $u(r_i)$  rappresenta i potenziali esterni presenti nella *box* oppure l'effetto sulle molecole delle pareti del contenitore. Il secondo termine rappresenta il potenziale di interazione tra le coppie di molecole; per semplicità si trascurano le interazioni tra tre o

più molecole contemporaneamente. I potenziali possono essere determinati sperimentalmente o modellati in modo teorico (Maitland *et al.*, 1981; Stone, 1996).

Il potenziale *continuo* più utilizzato è il potenziale di Lennard – Jones (Allen, 2004), teorizzato inizialmente per spiegare il comportamento delle molecole di Argon (Verlet, 1967):

$$V_{LJ}(r) = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad (2.4)$$

Dove  $\sigma$  e  $\varepsilon$  sono due parametri determinati empiricamente. Se invece si sta studiando un caso in cui siano presenti interazioni di tipo elettrostatico si utilizza il potenziale coulombiano:

$$V_{Coulomb}(r) = \frac{Q_A Q_B}{4\pi\varepsilon_0 r} \quad (2.5)$$

Per ciò che riguarda le interazioni intramolecolari il modello più semplice che può essere proposto contiene i seguenti termini (Allen, 2004):

$$\begin{aligned} V_{intramolecolare} = & \frac{1}{2} \sum_{legami} k_{ij}^r (r_{ij} - r_{eq})^2 + \frac{1}{2} \sum_{angoli\ di\ bending} k_{ijk}^\vartheta (\vartheta_{ijk} \\ & - \vartheta_{eq})^2 \\ & + \frac{1}{2} \sum_{angoli\ di\ torsione} \sum_m k_{ijkl}^{\varphi, m} (1 + \cos(m\varphi_{ijkl} - \gamma_m)) \end{aligned} \quad (2.6)$$

La configurazione spaziale corrispondente è rappresentata in Figura 2.1. I *legami* riguardano due molecole adiacenti, a distanza  $r_{ij} = |r_i - r_j|$  e si assume che tra esse il potenziale assuma la forma del potenziale armonico, con uno specifico valore di distanza di equilibrio ( $r_{eq}$ ).

Gli *angoli di bending* ( $\vartheta_{ijk}$ ) riguardano legami consecutivi, come  $r_i - r_j$  e  $r_k - r_j$  e coinvolgono quindi tre coordinate atomiche (Allen, 2004):

$$\cos \vartheta_{ijk} = \widehat{r}_{ij} \times \widehat{r}_{jk} = (r_{ij} \times r_{ij})^{-\frac{1}{2}} (r_{jk} \times r_{jk})^{-\frac{1}{2}} (r_{ij} \times r_{jk}) \quad (2.7)$$

Anche per gli *angoli di bending* si valuta la differenza rispetto a un valore di equilibrio ( $\vartheta_{eq}$ ).

Gli *angoli di torsione* ( $\varphi_{ijkl}$ ) riguardano tre legami, e coinvolgono quindi quattro coordinate atomiche (Allen, 2004):

$$\cos \varphi_{ijkl} = \widehat{n}_{ljk} \times \widehat{n}_{jkl} \quad \text{dove} \quad \widehat{n}_{ljk} = \frac{r_{ij} \times r_{jk}}{r_{jk} \times r_{kl}}; \quad \widehat{n}_{jkl} = \frac{r_{jk} \times r_{kl}}{r_{jk} \times r_{kl}} \quad (2.8)$$

e  $\widehat{n} = \frac{\mathbf{n}}{n}$  è la normale al piano definito da ogni coppia di legami.

Il modello di potenziale riportato è solamente una semplificazione di quello che viene implementato da un programma di dinamica molecolare, che avrà un modello specifico detto *force field*. Per quanto riguarda molecole più complesse come polimeri e proteine i *force field* più utilizzati sono AMBER (Weiner *et al.*, 1984), CHARMM (Brooks *et al.*, 1983), GROMOS (Van Gunsteren *et al.*, 1996) e OPLS (Jorgensen *et al.*, 1996) e sono stati studiati principalmente per la fase condensata e modificati a seguito di considerazioni legate alla meccanica quantistica.

Una volta specificato il potenziale di interazione tra le molecole, il programma può proseguire nel calcolo delle forze atomiche con l'equazione 2.2.

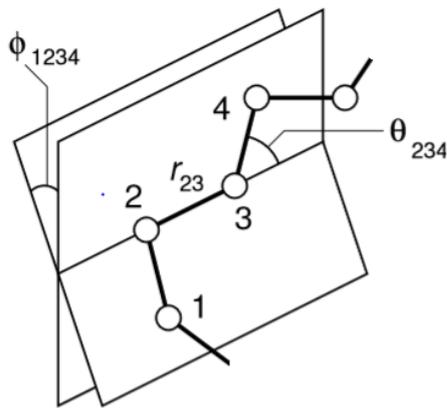
Per determinare le forze di interazione i programmi di calcolo fanno ricorso a metodi alle differenze finite, che permettono di approssimare l'operatore derivativo tramite operazioni di somma e prodotto. Per minimizzare possibili errori di tipo numerico, il software GROMACS utilizza l'algoritmo di Verlet (Swope *et al.*, 1982):

$$v \left( t + \frac{1}{2} \Delta t \right) = v(t) + \frac{\Delta t}{2m} * F(t) \quad (2.9)$$

$$r(t + \Delta t) = r(t) + \Delta t * v \left( t + \frac{1}{2} \Delta t \right) \quad (2.10)$$

$$v(t + \Delta t) = v \left( t + \frac{1}{2} \Delta t \right) + \frac{\Delta t}{2m} * F(t + \Delta t) \quad (2.11)$$

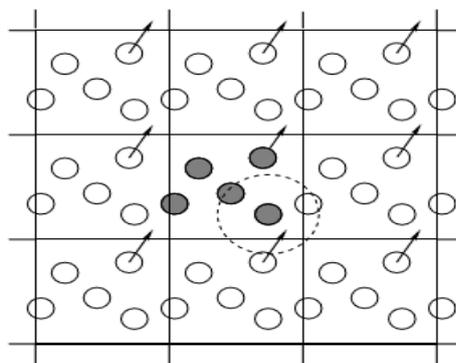
Come si può vedere, ad ogni passo temporale  $\Delta t$ , si calcolano la coordinata spaziale e la velocità di ogni particella del sistema utilizzando i valori ottenuti nel passo precedente. È importante scegliere il  $\Delta t$  corretto, poiché un passo troppo breve risulterebbe eccessivamente dispendioso dal punto di vista computazionale, mentre un passo eccessivamente lungo produrrebbe risultati inaccurati (Allen, 2004).



**Figura 2.1.** Rappresentazione di legami, angoli di bending e angoli di torsione. Immagine tratta da Allen, 2004 con modifiche.

### 2.1.1 Periodic boundary conditions – condizioni periodiche al contorno

Quando si simula un sistema di particelle in dinamica molecolare si deve inoltre tener conto che, per limitare il costo computazionale, i sistemi simulati hanno una dimensione ridotta. A causa di questa condizione c'è il rischio, ad esempio nel caso in cui si crei uno spazio cubico in cui si calcolano traiettorie e velocità, che le particelle collocate a ridosso delle facce siano trattate in modo diverso da quelle al centro del cubo, poiché soggette ad effetti di bordo. Per evitare questo problema ogni programma di MD adotta le *periodic boundary conditions*, ossia simula il sistema in questione come confinante con altri sistemi identici. In questa maniera le molecole a ridosso della superficie possono essere trattate come se fossero all'interno di una cella, risultando in un sistema più simile a quello reale (Allen, 2004). GROMACS adotta inoltre la *minimum image convention*, un artificio che prevede che ogni molecola interagisca solamente con quella che si trova più vicina a essa nella cella confinante. Inoltre, come mostrato in figura 2.2, si prevede che nel momento in cui una molecola esca da una cella, la stessa venga rimpiazzata da una identica proveniente da una cella confinante (che, come scritto prima, è identica ad essa). Le *periodic boundary conditions* si possono utilizzare sia per celle cubiche che per celle rombiche dodecaedriche e ottaedriche troncate (Figura 2.2).



**Figura 2.2** Periodic boundary condition in un programma di MD. Immagine tratta da Allen, 2004 con modifiche.

### 2.1.2 Neighbour list – lista dei ‘vicini’

Nel calcolo delle interazioni tra molecole non legate, ogni programma dovrebbe in principio considerare una grandissima quantità di coppie che, nel momento in cui si trovassero a una certa distanza, avrebbero una interazione di fatto trascurabile. Per questo sono presenti, a livello computazionale, i cosiddetti *cut-off*,  $r_{cut}$ , ossia dei limiti spaziali tali per cui se la distanza  $r_{ij}$  tra una molecola  $i$  e una molecola  $j$  è superiore a  $r_{cut}$ , il potenziale  $V_{ij}$  viene posto uguale a 0. Nonostante ciò in ogni ambiente di calcolo contenente  $N$  molecole vanno esaminate circa  $\frac{1}{2}N(N - 1)$  interazioni, di cui per ognuna va calcolato almeno  $r_{ij}^2$ . Poiché questa procedura risulta eccessivamente dispendiosa per il programma di calcolo, viene utilizzata la tecnica proposta da Verlet (Verlet, 1967), che prevede di considerare per ogni molecola un guscio sferico di raggio  $r_{list}$ , di dimensione superiore a  $r_{cut}$ , avente origine nel proprio centro di massa e che racchiuda un certo numero di molecole al suo interno. Queste verranno inserite in una lista di ‘vicini’ e nei passi successivi verranno considerate solamente le interazioni tra la molecola e i suoi ‘vicini’. Periodicamente (ogni 10 step su GROMACS) questa lista viene aggiornata, facendo attenzione a stabilire un passo temporale tale per cui nel guscio non entrino molecole prima non racchiuse (Figura 2.3). La scelta del valore di  $r_{list}$  deve essere un adeguato compromesso tra precisione e costo computazionale. Per sistemi con numero di particelle molto elevato (es. box con più di 1000 particelle) si preferisce utilizzare celle di lato superiore a  $r_{cut}$ . In questi sistemi la lista dei ‘vicini’ viene comparata con la lista delle molecole presenti nella cella in cui si trova la molecola e in quelle adiacenti (Hockney et Eastwood, 1988), in cui potrebbero essere inclusi dei ‘vicini’ (Figura 2.4). Questo processo risulta molto rapido ed efficiente per sistemi di grandi dimensioni (Allen, 2004).

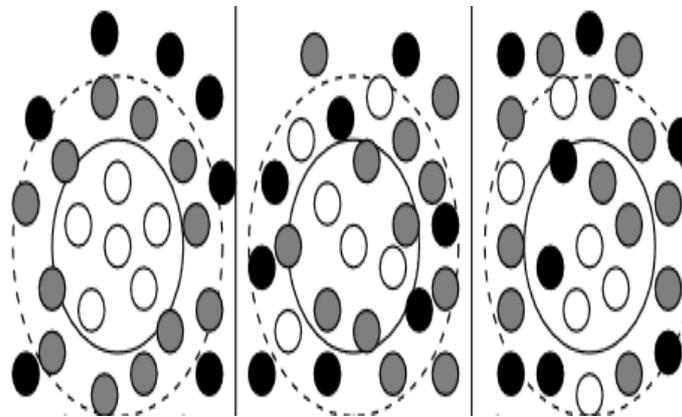
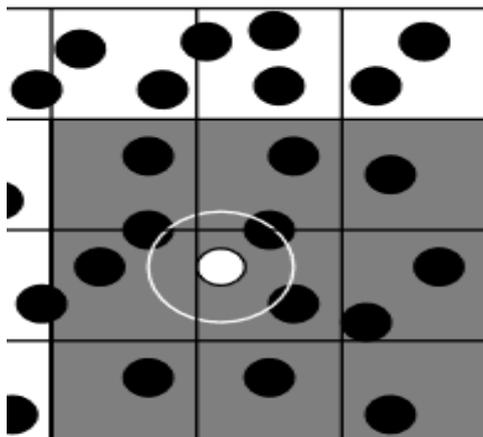


Figura 2.3 Neighbour list in un programma di MD. Immagine tratta da Allen, 2004 con modifiche



**Figura 2.4** Celle e  $r_{list}$  in un programma di MD. Immagine tratta da Allen, 2004 con modifiche

## 2.2 Procedura di simulazione

L'obiettivo principale del presente lavoro è quello di valutare in ambiente computazionale le interazioni tra ognuno dei venti amminoacidi e due eccipienti, comunemente utilizzati in ambito farmaceutico come lio- e crio-protettori, ovvero saccarosio e trealosio (Wang, 2000). Per ottenere questo risultato è stato utilizzato GROMACS, uno dei più comuni programmi di Dinamica Molecolare. Le condizioni in cui si vuole valutare il comportamento del nostro sistema si rifanno a quelle già presenti in letteratura (Ganguly *et al.*, 2015) in modo da aver dei risultati confrontabili con altre trattazioni e simulazioni che rispecchino in modo adeguato la realtà del processo.

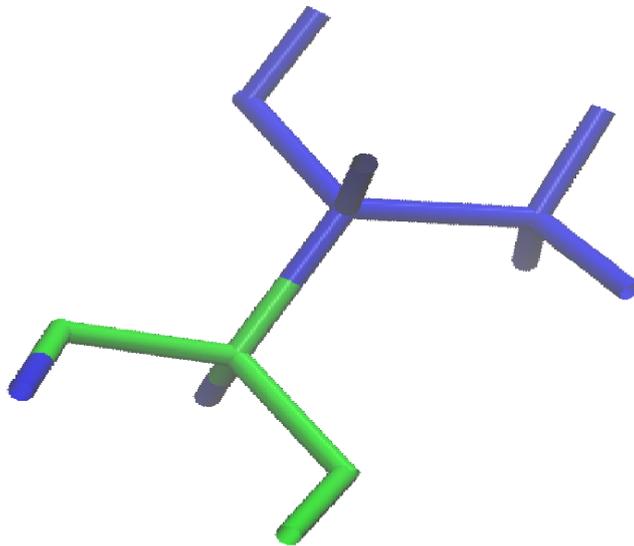
Per ogni simulazione si è creata una box cubica di lato 8 nm, in cui sono state poste 25 molecole di un singolo amminoacido e 308 molecole di saccarosio o trealosio, pari a una concentrazione 1 M di eccipiente. Questa concentrazione viene comunemente utilizzata in ambito farmaceutico (Wang, 2000) poiché sufficiente a garantire un'adeguata protezione per le proteine e viene presa anche come riferimento per gli studi sull'additività delle interazioni proteine-cosoluto (Auton e Bolen, 2005) che vengono presi come modello nell'ultima parte di questo lavoro di tesi. La box viene infine saturata con acqua e contiene al suo interno quindi una soluzione ideale di un farmaco in condizioni precedenti la liofilizzazione. Le condizioni di temperatura e pressione su cui si basa la simulazione sono di 300 K e 1 bar. Ciò sembrerebbe differire da quello che avviene in un ambiente reale di liofilizzazione. In realtà utilizzare tali condizioni permette di arrivare all'equilibrio in tempi brevi rispetto a condizioni di bassa pressione e bassa temperatura. Questo si traduce, a livello computazionale, in un numero inferiore di passi richiesti al calcolatore per completare la simulazione, arrivando al medesimo risultato finale. Il tempo per cui si simula l'evoluzione del sistema è pari a 60 ns, considerato sufficiente per raggiungere l'equilibrio.

Una volta dato il via alla simulazione, GROMACS calcola la traiettoria delle molecole presenti nella box per il periodo desiderato, in step temporali che corrispondono a una durata di 2 femsecondi ognuno, per un totale di 30 milioni di passi.

Alla fine della simulazione si ottiene un file di testo contenente tutte le informazioni spaziali e temporali riguardanti la traiettoria delle molecole, e, dalla sua analisi si riesce a

risalire a tutte le informazioni desiderate riguardo il sistema esaminato. La presente ricerca in particolar modo si concentra sulla valutazione all'equilibrio dei legami idrogeno che si formano tra gli amminoacidi e le molecole di solvente o protettore, dell'area esposta dai singoli amminoacidi alla soluzione in cui sono immersi, del cosiddetto *Kirchwood – Buff Integral* (KBI) tra gli amminoacidi e le molecole di solvente o protettore. Il significato del KBI viene spiegato in seguito nel paragrafo 2.3.5. Per ogni amminoacido si valutano questi parametri sia per l'intera molecola, che per le sue due componenti singole, la *backbone* e la *sidechain*, rispettivamente la struttura principale di ogni amminoacido, identica per tutti e 20 gli amminoacidi, e la catena laterale attaccata alla *backbone*, differente per ogni amminoacido (Figura 2.5).

In questa sezione viene descritto nel dettaglio il codice utilizzato per le simulazioni realizzate su GROMACS.



**Figura 2.5** Immagine generata con software VMD dell'amminoacido Treonina. In verde si può vedere il backbone, parte uguale per tutti gli amminoacidi, in blu la sidechain, parte caratterizzante ogni singolo residuo.

### 2.2.1 Creazione del file di struttura

```
gmx pdb2gmx -f amminoacido.pdb -o amminoacido.gro -water tip3p
```

Il primo comando inserito su GROMACS prevede di partire (-f) dal file di struttura dell'amminoacido, con estensione *pdb*, e trasformarlo in output (-o) in un altro tipo di file di struttura, con estensione *gro*, così da permettere la creazione di molecole di amminoacido da inserire in un ambiente simulativo. Il file di struttura viene modificato aggiungendo atomi di idrogeno agli amminoacidi. L'aggiunta di questi atomi viene fatta seguendo quanto prescritto dal campo di forza selezionato. Per le presenti simulazioni è stato utilizzato il CHARMM36 modificato da Trout (Cloutier *et al.*, 2018). Le molecole di acqua hanno topologia di tipo *tip3p*. Con questa linea di comando viene anche generato il

file di topologia *topol.top*, che contiene le informazioni riguardo il campo di forza utilizzato dal sistema e la topologia dell'amminoacido (file di estensione *itp*, Figura 2.6c).

### 2.2.1.1 File di struttura (.pdb e .gro)

I file di struttura con estensione *pdb* (Figura 2.6a) e *gro* (Figura 2.6b) di un generico amminoacido descrivono la posizione degli atomi in una struttura molecolare (Abraham *et al.*, 2005).

```

|TITLE          S  C  A  M  O  R  G
|MODEL         1
|ATOM          1  H  LIG      1      0.300  1.270  1.530  1.00  0.00
|ATOM          2  C  LIG      1      0.870  0.810  0.720  1.00  0.00
|ATOM          3  O  LIG      1      1.890  0.070  1.360  1.00  0.00
|ATOM          4  C1 LIG      1      2.830 -0.590  0.480  1.00  0.00
|ATOM          5  H1 LIG      1      2.290 -1.310 -0.150  1.00  0.00
|ATOM          6  C2 LIG      1      3.780 -1.360  1.400  1.00  0.00
|ATOM          7  H2 LIG      1      4.510 -1.900  0.790  1.00  0.00
|ATOM          8  H3 LIG      1      4.330 -0.640  2.030  1.00  0.00
|ATOM          9  O1 LIG      1      3.090 -2.320  2.180  1.00  0.00
|ATOM         10  H4 LIG      1      2.430 -1.830  2.700  1.00  0.00
|ATOM         11  C3 LIG      1      3.530  0.440 -0.400  1.00  0.00
|ATOM         12  H5 LIG      1      4.100  1.130  0.240  1.00  0.00
|ATOM         13  O2 LIG      1      4.400 -0.250 -1.290  1.00  0.00
|ATOM         14  H6 LIG      1      4.710  0.410 -1.930  1.00  0.00
|ATOM         15  C4 LIG      1      2.490  1.250 -1.170  1.00  0.00
|ATOM         16  H7 LIG      1      1.970  0.570 -1.860  1.00  0.00
|ATOM         17  O3 LIG      1      3.190  2.250 -1.910  1.00  0.00
|ATOM         18  H8 LIG      1      2.520  2.760 -2.390  1.00  0.00
|ATOM         19  C5 LIG      1      1.460  1.870 -0.220  1.00  0.00
|ATOM         20  H9 LIG      1      1.960  2.620  0.390  1.00  0.00
|ATOM         21  O4 LIG      1      0.450  2.560 -0.940  1.00  0.00
|ATOM         22  H10 LIG     1     -0.100  1.890 -1.380  1.00  0.00
|ATOM         23  O5 LIG      1      0.010 -0.020 -0.060  1.00  0.00
|ATOM         24  C6 LIG      1     -0.850 -0.850  0.730  1.00  0.00
|ATOM         25  H11 LIG     1     -0.260 -1.340  1.510  1.00  0.00
|ATOM         26  O6 LIG      1     -1.840 -0.120  1.400  1.00  0.00
|ATOM         27  C7 LIG      1     -2.790  0.570  0.560  1.00  0.00
|ATOM         28  H12 LIG     1     -2.270  1.300 -0.060  1.00  0.00
|ATOM         29  C8 LIG      1     -3.750  1.310  1.490  1.00  0.00
|ATOM         30  H13 LIG     1     -4.580  1.680  0.880  1.00  0.00
|ATOM         31  H14 LIG     1     -4.150  0.600  2.230  1.00  0.00
|ATOM         32  O7 LIG      1     -3.040  2.370  2.110  1.00  0.00
|ATOM         33  H15 LIG     1     -3.670  2.850  2.670  1.00  0.00
|ATOM         34  C9 LIG      1     -3.520 -0.450 -0.330  1.00  0.00
|ATOM         35  H16 LIG     1     -4.080 -1.150  0.300  1.00  0.00
|ATOM         36  O8 LIG      1     -4.400  0.260 -1.190  1.00  0.00
|ATOM         37  H17 LIG     1     -4.720 -0.380 -1.840  1.00  0.00
|ATOM         38  C10 LIG     1     -2.500 -1.240 -1.140  1.00  0.00
|ATOM         39  H18 LIG     1     -1.990 -0.550 -1.830  1.00  0.00
|ATOM         40  O9 LIG      1     -3.220 -2.220 -1.880  1.00  0.00
|ATOM         41  H19 LIG     1     -2.570 -2.740 -2.380  1.00  0.00
|ATOM         42  C11 LIG     1     -1.460 -1.890 -0.230  1.00  0.00
|ATOM         43  H20 LIG     1     -1.960 -2.650  0.380  1.00  0.00
|ATOM         44  O10 LIG     1     -0.470 -2.590 -0.970  1.00  0.00
|ATOM         45  H21 LIG     1      0.080 -1.910 -1.410  1.00  0.00
|TER
|ENDMDL

```

**Figura 2.6a** File di struttura *pdb* del saccarosio. Si possono vedere le coordinate relative alla posizione e alla velocità di ogni specifico atomo.

Metionina in water  
49946

1MET	N	1	6.051	4.998	0.620	-0.5929	0.6612	0.3670
1MET	H1	2	6.094	4.914	0.576	-0.7400	0.2070	1.0744
1MET	H2	3	6.039	4.979	0.721	-0.1908	0.4702	0.3802
1MET	H3	4	5.956	5.009	0.580	-1.5165	0.9642	2.5067
1MET	CA	5	6.135	5.116	0.589	0.4322	-0.1858	-0.1440
1MET	HA	6	6.116	5.146	0.487	-0.5233	-1.7464	-0.4559
1MET	CB	7	6.067	5.224	0.674	0.0407	-0.5543	0.0146
1MET	HB1	8	6.076	5.216	0.785	-1.2029	-1.1158	0.0883
1MET	HB2	9	6.122	5.320	0.658	-1.0798	0.1591	0.3233
1MET	CG	10	5.920	5.253	0.641	0.1600	0.1586	0.0963
1MET	HG1	11	5.916	5.241	0.531	1.4740	-1.4043	0.1723
1MET	HG2	12	5.864	5.169	0.681	-1.2688	0.2591	1.7952
1MET	SD	13	4.864	5.419	0.696	-0.9488	0.0660	0.1775
1MET	CE	14	5.701	5.411	0.625	-0.1440	0.7572	0.3208
1MET	HE1	15	5.655	5.509	0.649	-1.5904	-0.1475	1.4020
1MET	HE2	16	5.700	5.408	0.514	-0.0797	-0.6138	0.3443
1MET	HE3	17	5.638	5.329	0.667	0.6461	-0.2785	-0.4738
1MET	C	18	6.286	5.103	0.608	0.4230	0.0988	0.1301
1MET	OT1	19	6.343	4.994	0.633	0.0003	-0.0568	0.4402
1MET	OT2	20	6.358	5.205	0.591	0.2110	0.1600	-0.4073
2MET	N	21	7.704	6.806	0.029	0.3919	-0.2357	-0.4979
2MET	H1	22	7.613	6.848	0.055	0.1437	-0.4001	-1.0770
2MET	H2	23	7.765	6.889	0.017	0.0853	0.2519	1.1057
2MET	H3	24	7.695	6.774	-0.070	0.5454	-1.3769	-0.1590
2MET	CA	25	7.749	6.695	0.116	0.1193	0.4108	0.4822
2MET	HA	26	7.729	6.720	0.219	0.3889	0.1429	0.6024
2MET	CB	27	7.698	6.555	0.076	0.2162	0.3379	0.6116
2MET	HB1	28	7.745	6.538	-0.023	-2.3009	0.4066	-0.6777
2MET	HB2	29	7.754	6.487	0.144	0.4599	-0.1557	-0.0751
2MET	CG	30	7.549	6.530	0.098	0.3338	-0.1928	0.8022
2MET	HG1	31	7.525	6.544	0.206	0.1974	-2.0500	1.0594
2MET	HG2	32	7.505	6.610	0.035	-0.5979	-2.4012	-1.5209
2MET	SD	33	7.489	6.370	0.035	0.0762	0.1559	0.1574
2MET	CE	34	7.332	6.361	0.127	0.2050	-0.7556	0.2969
2MET	HE1	35	7.260	6.441	0.098	-1.0666	-2.8077	-2.6563
2MET	HE2	36	7.278	6.265	0.124	1.4299	-1.4603	0.0423
2MET	HE3	37	7.351	6.379	0.235	1.3137	2.0424	-0.2782
2MET	C	38	7.901	6.084	0.107	0.0907	0.1334	0.3183
2MET	OT1	39	7.965	6.747	0.019	0.1744	-0.5718	-0.1391
2MET	OT2	40	7.964	6.605	0.183	0.4243	0.0857	-0.0061
3MET	N	41	4.924	5.902	4.535	-0.1291	0.1340	0.0907
3MET	H1	42	5.910	5.951	4.504	-0.7874	0.8146	-0.7294
3MET	H2	43	4.949	5.826	4.600	-0.7015	-0.3950	-0.2979
3MET	H3	44	4.888	5.853	4.451	2.8459	0.3976	-1.4486
3MET	CA	45	4.843	6.008	4.600	-0.1534	0.0259	0.2363
3MET	HA	46	4.841	6.089	4.529	-1.2990	-2.8876	-3.3536
3MET	CB	47	4.694	5.969	4.611	-0.0585	-0.4545	-0.1140
3MET	HB1	48	4.681	5.895	4.693	-1.0871	-0.2032	-0.0409
3MET	HB2	49	4.634	6.060	4.635	-0.1759	0.3129	-0.9206
3MET	CG	50	4.636	5.915	4.481	-0.0321	0.0138	-0.3221
3MET	HG1	51	4.648	5.986	4.396	2.5518	-0.0861	-0.1132
3MET	HG2	52	4.601	5.820	4.462	-0.4635	0.1058	0.1334
3MET	SD	53	4.464	5.859	4.502	0.0182	-0.0051	0.0446
3MET	CE	54	4.446	5.812	4.327	-0.2747	-0.1773	0.1209
3MET	HE1	55	4.482	5.889	4.257	0.5862	-0.6128	-0.5245
3MET	HE2	56	4.530	5.739	4.321	-0.9225	-0.8151	-1.7335
3MET	HE3	57	4.346	5.779	4.291	-0.5540	-0.3569	1.0134
3MET	C	58	4.896	6.059	4.733	0.0947	0.2650	0.0453
3MET	OT1	59	5.008	6.024	4.781	0.4307	0.7926	-0.3459
3MET	OT2	60	4.824	6.141	4.796	-0.4678	0.3036	-0.6405
4MET	N	61	5.993	5.208	6.836	0.1408	0.6182	-0.6182

Figura 2.6b File di struttura gro per alcuni residui di metionina in soluzione acquosa. In questo file viene riportata la posizione di tutti gli atomi per ogni residuo.

```
[ moleculetype ]
; Name          nrexcl
Prolina        3

[ atoms ]
; nr      type  resnr residue  atom  cgnr      charge      mass  typeB  chargeB
; residue 2 PRO  rtp  PRO   q  0.0
1         NP    2     PRO   N    1     -0.07      14.007  ; qtot -0.07
2         HC    2     PRO  HN1   2     0.24       1.008  ; qtot 0.17
3         HC    2     PRO  HN2   3     0.24       1.008  ; qtot 0.41
4         CP3   2     PRO   CD    4     0.16       12.011 ; qtot 0.57
5         HA2   2     PRO  HD1   5     0.09       1.008  ; qtot 0.66
6         HA2   2     PRO  HD2   6     0.09       1.008  ; qtot 0.75
7         CP1   2     PRO   CA    7     0.16       12.011 ; qtot 0.91
8         HB1   2     PRO   HA    8     0.09       1.008  ; qtot 1
9         CP2   2     PRO   CB    9     -0.18      12.011 ; qtot 0.82
10        HA2   2     PRO  HB1   10    0.09       1.008  ; qtot 0.91
11        HA2   2     PRO  HB2   11    0.09       1.008  ; qtot 1
12        CP2   2     PRO   CG   12    -0.18      12.011 ; qtot 0.82
13        HA2   2     PRO  HG1   13    0.09       1.008  ; qtot 0.91
14        HA2   2     PRO  HG2   14    0.09       1.008  ; qtot 1
15        CC    2     PRO   C    15    0.34       12.011 ; qtot 1.34
16        OC    2     PRO  OT1   16    -0.67      15.9994 ; qtot 0.67
17        OC    2     PRO  OT2   17    -0.67      15.9994 ; qtot 0

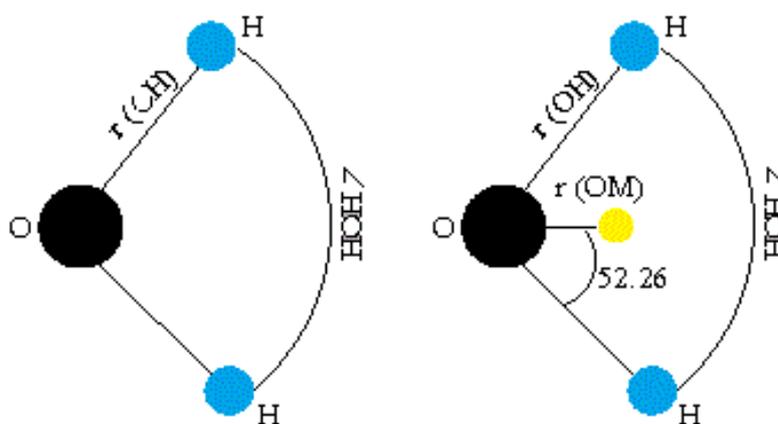
[ bonds ]
; al      aj      funct      c0      c1      c2      c3
1         2         1
1         3         1
1         4         1
1         7         1
4         5         1
4         6         1
4         12        1
7         8         1
7         9         1
7         15        1
9         10        1
9         11        1
9         12        1
12        13        1
12        14        1
15        16        1
15        17        1

[ pairs ]
; al      aj      funct      c0      c1      c2      c3
1         10        1
1         11        1
1         13        1
1         14        1
1         16        1
1         17        1
2         5         1
```

Figura 2.6c Esempio tratto dal file di topologia della Prolina. Il file di topologia (estensione itp) è sempre presente insieme al file di struttura di una molecola, e indica i legami e gli angoli tra i vari atomi, oltre che le relative cariche (Abraham et al., 2005)

### 2.2.1.2 Topologia dell'acqua

La topologia scelta per l'acqua nelle simulazioni è la tipologia water tip3p, e viene utilizzata in accordo con quanto riportato in letteratura (Cloutier *et al.*, 2018). Quando in dinamica molecolare si tratta di soluzioni acquose non è banale scegliere un force field ottimale. Le molecole d'acqua possono essere descritte con molecole a tre, quattro o cinque siti (Jorgensen *et al.*, 1983), come si può vedere in Figura 2.7. Un modello a tre siti, come il tip3p, è particolarmente adatto per i calcoli in cui è presente acqua allo stato liquido e prevede due siti con carica positiva, corrispondenti nella realtà agli atomi di idrogeno, e un sito con carica negativa, corrispondente nella realtà all'atomo di ossigeno, con  $q_O = -2q_H$  (Jorgensen *et al.*, 1983). Successivamente la topologia a tre siti è stata ottimizzata per lo stato liquido in modo tale che i risultati computazionali riproducessero quelli sperimentali per ciò che concerne le funzioni di distribuzione radiale degli atomi di ossigeno e idrogeno (Kusalik e Svishchev, 1994) e per avere densità della soluzione uguale a quella reale. Le topologie a quattro e cinque siti utilizzano dei modelli in cui si aggiungono altre cariche fittizie per rendere ancora migliore il risultato computazionale sia per lo stato liquido che per gli altri stati di aggregazione della materia, ma comportano un maggior costo computazionale (Jorgensen *et al.*, 1983).



**Figura 2.7** Rappresentazione grafica del modello per l'acqua a 3 siti e 4 siti. Immagine tratta, con modifiche, da <http://www.mdpi.org/cji/cji/2004/069061ne.html>

### 2.2.1.3 Campo di forza

La scelta del campo di forza è molto importante in una simulazione di dinamica molecolare, poiché da esso dipendono il calcolo della traiettoria seguita dalle molecole e, come già descritto nella sezione precedente, la struttura spaziale delle singole molecole nel file di estensione *gro*. Il fine principale di ogni simulazione è infatti quello di produrre dei risultati in tutto e per tutto affini a quelli sperimentali.

Il nostro campo di forza di partenza è il CHARMM36, abitualmente utilizzato per simulare proteine, carboidrati e acqua. Si è visto però che, utilizzando questo campo, in presenza di proteine e zuccheri in soluzione acquosa, a concentrazioni di zuccheri simili a quelle utilizzate nei processi di liofilizzazione, le simulazioni danno come risultato un andamento opposto rispetto ai risultati sperimentali, indicando inclusione degli zuccheri rispetto alla proteina, invece che esclusione (Cloutier *et al.*, 2018; Ha *et al.*, 1988).

Per ovviare a questo problema il CHARMM36 è stato modificato da Trout e altri ricercatori, avendo come obiettivo quello di ottenere per diverse simulazioni contenenti carboidrati (tra cui saccarosio e trealosio), proteine e acqua (con topologia *tip3p*) dei valori di KBI corrispondenti alla realtà sperimentale, oltre che congrui valori di *clustering*, ossia di dimensione degli aggregati di carboidrati ottenuti al raggiungimento dell'equilibrio. A livello di simulazione questi valori si possono ottenere calcolando la funzione di distribuzione radiale, mentre sperimentalmente si ottengono valutando parametri della soluzione tra cui compressibilità, coefficiente di attività e volumi molari parziali.

La modifica del CHARMM36 è avvenuta sui cosiddetti *non bonded parameters* ossia i valori che nel campo di forza riguardano le interazioni tra molecole di tipo Van Der Walls. Si sono quindi modificate le cariche parziali degli atomi di ossigeno, idrogeno e carbonio, e i parametri del potenziale di Lennard-Jones.

Le modifiche apportate al CHARMM36 ne hanno cambiato i valori di carica parziale per O e H dei gruppi alcolici, con conseguente aumento della carica parziale del C ad esso legato per mantenere il sistema neutro. Il parametro  $\sigma$  dell'equazione di Lennard – Jones è rimasto il medesimo, mentre per il parametro  $\epsilon$  sono stati definiti due diversi valori per ogni tipo di atomo (come mostrato in Tabella 2.1): uno per le interazioni con l'acqua, uno diverso per le altre interazioni.

A seguito di queste modifiche, rispetto al campo di forza di partenza, si è riscontrata una migliore aderenza ai dati sperimentali per quanto riguarda le interazioni tra acqua e carboidrati e tra proteina e carboidrati, mentre le interazioni tra carboidrati e carboidrati si sono dimostrate leggermente differenti da quelle ricavate sperimentalmente. A differenza però delle simulazioni effettuate con il CHARMM36, gli andamenti con l'aumentare della concentrazione di carboidrati si sono dimostrati uguali a quelli sperimentali, cioè è stato visto che all'aumentare della concentrazione aumenta l'interazione tra carboidrati e carboidrati (Cloutier *et al.*, 2018). Oltre ad una corretta predizione di queste interazioni, è stato anche dimostrato che le modifiche apportate al CHARMM36 permettono di ottenere corretti valori di densità della soluzione, dei coefficienti di attività degli elementi presenti in soluzione e dei loro volumi molari parziali (Cloutier *et al.*, 2018). Tutti i risultati sono corretti per concentrazioni molari di carboidrati pari o inferiori a 2M.

Per ciò che riguarda le interazioni tra carboidrati e proteine i risultati con le modifiche apportate al campo di forza permettono di ottenere andamenti pressoché uguali a quelli determinati sperimentalmente, avendo come punto più importante la corrispondenza nell'esclusione dei carboidrati dalla superficie delle proteine all'aumentare della concentrazione dei primi, postulato fondamentale nel considerare i carboidrati come dei protettori (si veda il Capitolo 1). È stato visto che in presenza di amminoacidi con elevata superficie esposta ed idrofobicità (ALA, ILE, PHE), i carboidrati sono maggiormente inclusi poiché tendono a sostituire le sfavorevoli interazioni dei residui con l'acqua, contribuendo quindi alla stabilità proteica (Cloutier *et al.*, 2018).

In conclusione, il CHARMM36 modificato nei suoi parametri *non bonded* da Trout permette di ottenere valori molto simili a quelli sperimentali per le interazioni delle proteine in soluzione acquosa con zuccheri quali trealosio e saccarosio, e risulta il più adatto per le simulazioni svolte in questo lavoro.

**Tabella 2.1** Parametri utilizzati per il potenziale di Lennard-Jones per i gruppi -OH (Cloutier et al., 2018).

Atomo	Interazione	$\epsilon$ (kJ/mol)	$\sigma$ (nm)
O	Generale	0.45	0.314
	Acqua	0.9	
H	Generale	0.12	0.04
	Acqua	0.3	

### 2.2.2 Creazione della box

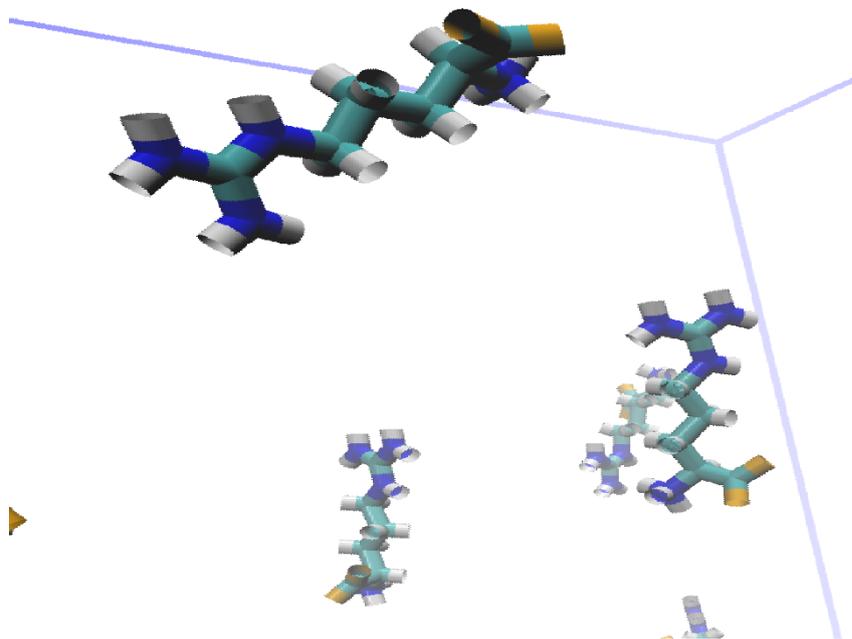
```
gmx insert-molecules -ci ala.gro -nmol 25 -o pbox.gro -box 8 8 8
```

Questo comando permette di creare su GROMACS una parte tridimensionale di spazio denominata *box* in cui vengono collocate, in modo casuale, 25 molecole di amminoacido (negli esempi riportati c'è sempre l'amminoacido *alanina*) tramite i comandi *insert-molecules* e *nmol*. Tramite il comando *box* si specificano le dimensioni del nostro spazio tridimensionale, che nel presente lavoro è un cubo di lato 8 nanometri (Figura 2.8). Il file contenente la box con al suo interno le 25 molecole di amminoacido è denominato *pbox.gro*, e riporta le coordinate spaziali di ogni atomo all'interno della *box* (Abraham et al., 2005).

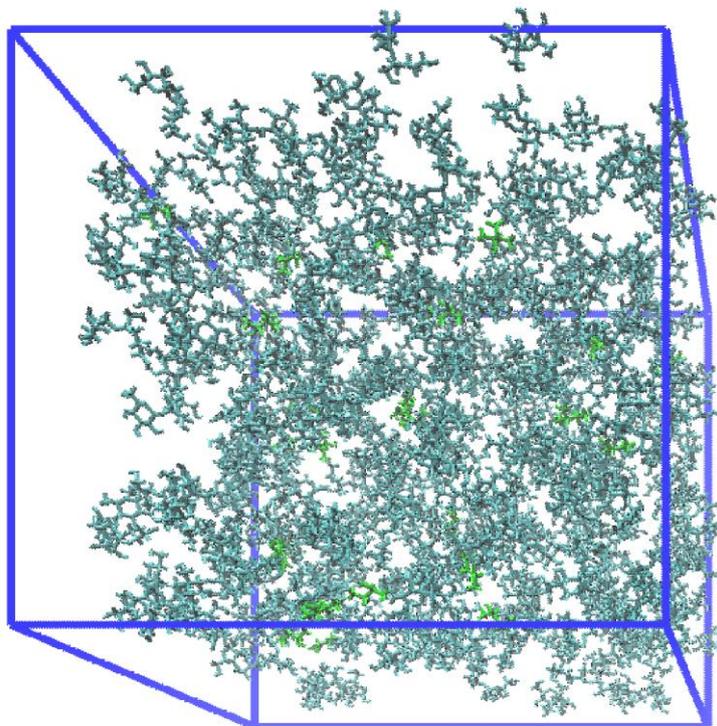
### 2.2.3 Inserimento dello zucchero

```
gmx insert-molecules -fpbox.gro -ci suc.pdb -nmol 308 -o psbox.gro
```

Per inserire le molecole di zucchero si usa nuovamente il comando *insert-molecules*, specificando tramite il comando *nmol* che le molecole di zucchero da aggiungere nelle simulazioni sono 308, pari a concentrazione 1 M per la nostra soluzione. Il comando è il medesimo per trealosio e saccarosio, considerando che, essendo isomeri, le molecole da aggiungere sono le stesse per entrambi. Gli zuccheri vengono rappresentati con un file di struttura avente estensione *pdb* e anche in questo caso il loro inserimento è casuale. In output si ha il file *psbox.gro* (Figura 2.9) che contiene le informazioni sulle coordinate spaziali di zuccheri e amminoacidi all'interno della box (Abraham et al., 2005).



**Figura 2.8** *Rappresentazione con VMD di alcuni residui di arginina in una box (delimitata dalle linee blu). Il colore bianco rappresenta gli atomi di idrogeno, il verde acqua gli atomi di carbonio, il blu gli atomi di azoto, l'oro gli atomi di ossigeno.*

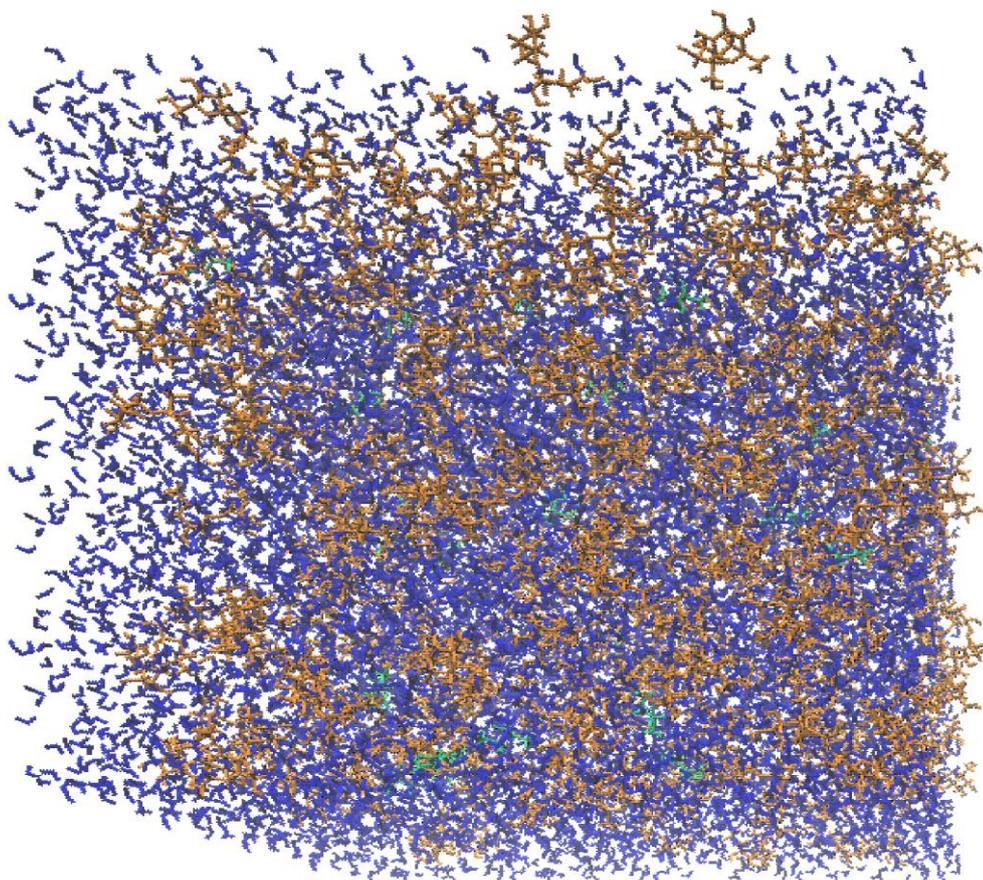


**Figura 2.9** *Rappresentazione tramite VMD di una box contenente 308 molecole di saccarosio (colore azzurro) e 25 di amminoacido (colore verde).*

## 2.2.4 Inserimento dell'acqua

```
gmx solvate -cp psbox.gro -cs spc216.gro -o pswbox.gro -p topol.top
```

L'acqua viene aggiunta nella *box* tramite il comando *solvate* che satura la *box* con molecole d'acqua. Il comando riceve in input (*-cp*) la struttura della *box* creata negli step precedenti e vi inserisce come solvente ciò che viene specificato dall'utente (*-cs*). Nel presente lavoro viene inserita una specifica topologia di acqua che segue il modello *tip3p*, ossia la topologia *spc216.gro*, che consiste in un file di testo che specifica la posizione di 72 molecole di acqua (216 atomi) nello spazio e viene ripetuto nella nostra *box* finché tutto lo spazio libero viene saturato. In uscita è prodotto il file di struttura *pswbox.gro* (Figura 2.10) e viene aggiornato (*-p*) il file di topologia del sistema (un esempio è riportato in Figura 2.11) in questione *topol.top* (Abraham *et al.*, 2005).



**Figura 2.10** Immagine generata tramite VMD di un sistema contenente acqua (colore blu), 308 molecole di saccarosio (colore arancione) e 25 molecole di amminoacido (colore verde). In questo caso non sono state riportate le linee di contorno della *box*.

```

!
; Include forcefield parameters
#include "./charmm36-modified-trout.ff/forcefield.itp"

#include "gln.itp"

#include "Sucrose.itp"

; Include water topology
#include "./charmm36-modified-trout.ff/tip3p.itp"

#ifdef POSRES_WATER
; Position restraint for each water oxygen
[ position_restraints ]
; i funct      fcx      fcy      fcz
  1  1      1000      1000      1000
#endif

; Include topology for ions
#include "./charmm36-modified-trout.ff/ions.itp"

[ system ]
; Name
Glutamm in water

[ molecules ]
; Compound      #mols
Glutamm          25
Sucrose          308

SOL              11898

```

**Figura 2.11.** Esempio di file di topologia con estensione top per 25 molecole di Glutammato in acqua. Si possono vedere le topologie dei singoli composti, riportate con estensione itp e tutte le indicazioni riguardanti le molecole presenti nel sistema e il campo di forza.

## 2.2.5 Neutralizzazione della box e minimizzazione energetica del sistema

```

gmx grompp -f ions.mdp -c pswbox.gro -p topol.top -o ions.tpr
gmx genion -s ions.tpr -o pswbox.gro -p topol.top -pname NA -nname CL- nn/np 25
gmx grompp -f enermin.mdp -c pswbox.gro -p topol.top -o em.tpr
gmx mdrun -v -deffnm em

```

Nel momento in cui l'amminoacido presente nella box sia uno tra Aspartato, Glutammato, Arginina, Istidina, risulta che il sistema, dopo l'aggiunta di solvente, non è in equilibrio elettrostatico, essendo presente una carica per ogni amminoacido, positiva nel caso di Arginina e Istidina, negativa nel caso di Glutammato e Aspartato. In totale si hanno quindi 25 cariche in eccesso, che vanno equilibrate prima di iniziare la simulazione di minimizzazione energetica. Per far questo si utilizza il comando *grompp* che prende le istruzioni presenti nel file *ions.mdp* e permette di stimare la carica presente nella *box pswbox.gro*, stima contenuta nel file *ions.tpr*.

Si utilizza in seguito il comando *genion* che prende in input il file *ions.tpr* e restituisce in output il file di struttura *pswbox.gro* in cui 25 molecole di solvente sono state sostituite (-*np* 25 per gli ioni positivi oppure -*nn* 25 per gli ioni negativi), a seconda della carica totale del sistema, con 25 ioni Na<sup>+</sup> (in presenza di Glutammato o Aspartato) o con 25 ioni Cl<sup>-</sup> (in

presenza di Arginina e Istidina). Una volta effettuata questa operazione si aggiorna la topologia *topol.top*, aggiungendo il nome e la quantità di ioni presenti.

Dopo aver reso neutro l'ambiente di simulazione si procede con la minimizzazione energetica del sistema, cioè si portano le molecole della *box* ad una situazione di partenza tale per cui l'energia totale del sistema sia inferiore al valore scelto di  $1000 \text{ kJ/mol}$ . Per questo si utilizzano due righe di comando in sequenza. Nella prima si utilizza il comando *grompp* che prendendo il file di struttura *pswbox.gro* e il file di testo *enermin.mdp* produce un file binario con estensione *tpr* che contiene tutte le informazioni necessarie per l'inizio della simulazione di minimizzazione energetica. Le informazioni fornite dal file *pswbox.gro* riguardano la *box* e la disposizione spaziale delle molecole al suo interno, quelle fornite dal file *enermin.mdp* riguardano i parametri per la simulazione. Viene anche aggiornato il file *topol.top* con le informazioni fornite dal file *enermin.mdp* (Abraham *et al.*, 2005).

#### 2.2.5.1 File di minimizzazione energetica (*enermin.mdp*)

Il file di minimizzazione energetica *enermin.mdp* (v. Appendice A.1) stabilisce la durata massima della simulazione di minimizzazione energetica (*nsteps*), il passo temporale da utilizzare (*dt*), il tipo di algoritmo da utilizzare per la minimizzazione (*integrator*) e la sua frequenza di applicazione (*ntcgsteep*), la frequenza di aggiornamento della *neighbour list* (*nstlist*) e il suo modo di compilazione (*ns-type*), l'algoritmo con cui vengono calcolate le forze di interazione elettrostatica coulombiane (*coulombtype*) e il suo ordine di interpolazione (*pme-order*), la distanza spaziale del primo step di integrazione (*emstep*) e il valore di convergenza sotto il quale si considera conclusa la minimizzazione energetica (*emtol*), (Abraham *et al.*, 2005).

#### 2.2.6 Equilibratura del sistema

```
gmx grompp -f npt.mdp -c em.gro -p topol.top -o npt.tpr
gmx mdrun -v -deffnm npt
```

La linea di comando per l'equilibratura del sistema serve a portare la nostra *box*, una volta effettuata la minimizzazione energetica, nelle condizioni di temperatura e pressione desiderate per la nostra simulazione. Come già descritto in precedenza, si valuta il comportamento della soluzione a pressione di 1 bar e temperatura di 300 K. Il nostro sistema, date le dimensioni e le condizioni operative che decidiamo di tener costanti, corrisponde a un *NpT-ensemble*, ossia un sistema composto da una ipotetica quantità infinita di stati non interagenti, aventi la stessa temperatura, pressione e numero totale di molecole (Levine, 2009).

Anche in questo comando è presente un file di testo con estensione *mdp*, *npt.mdp*, contenente tutti i parametri e le informazioni sull'equilibratura e gli algoritmi utilizzati per ottenerla. Questo file viene letto, tramite il comando *grompp*, insieme al file di struttura generato dalla minimizzazione energetica, *em.gro*, per creare il file *npt.tpr*, che serve per far iniziare la vera e propria simulazione di dinamica molecolare nelle condizioni desiderate. Anche in questa linea di comando viene aggiornato il file di topologia *topol.top*.

Dopo aver creato il file *npt.tpr* si procede con la vera e propria simulazione per l'equilibratura del sistema. Il comando *mdrun* in questo caso permette di generare file

denominati *npt* (tramite il comando *-defnm - npt*) e produce un file partendo dal file di input *npt.tpr*. In questo caso quindi verranno svolte diverse iterazioni sino a che il sistema non raggiunga le condizioni di temperatura e pressione desiderate (Abraham *et al.*, 2005).

#### 2.2.6.1 File di equilibratura (*npt.mdp*)

Il file di equilibratura (*npt.mdp*) (v. Appendice A.2) include il file *posre.itp*, contenente i parametri di limitazione sulla posizione delle molecole, la durata della simulazione di equilibratura (*nsteps*), il passo temporale da utilizzare (*dt*), il tipo di algoritmo da utilizzare per la minimizzazione (*integrator*), il tempo di inizio della simulazione (*tinit*), il numero di steps che trascorrono nel trascrivere la traiettoria delle molecole (*nstxout*), la velocità (*nstvout*) e le energie (*nstenergy*), i gruppi per cui trascrivere la traiettoria (*xtc\_grps*) e quelli per cui trascrivere le energie (*energygrps*), la frequenza di aggiornamento della *neighbour list* (*nstlist*), il suo modo di compilazione (*ns-type*) e la distanza massima a cui vengono valutati (*rlist*), l'algoritmo con cui vengono calcolate le forze di interazione elettrostatica coulombiane (*coulombtype*) e il suo ordine di interpolazione (*pme-order*), l'algoritmo con cui vengono calcolate le forze di interazione elettrostatica di tipo Van Der Waals, la distanza fino a cui vengono considerate (*rvdw*) e le correzioni da effettuare in caso di distanze superiori (*DispCorr; ewald\_rtoll*).

Per quanto riguarda l'equilibratura della temperatura, nel file sono contenuti i comandi che descrivono quale algoritmo utilizzare per portare il sistema alla temperatura desiderata (*tcoupl*) e i relativi comandi riguardanti parametri per l'algoritmo di Berendsen (Berendsen *et al.*, 1984) (*tau; tc-grps; ref\_t*) che permettono di portare tutti i gruppi presenti nel nostro sistema (*tc-grps*) alla temperatura di 300 K (*ref\_t*).

Per quanto riguarda l'equilibratura della pressione, nel file sono contenuti i comandi che descrivono quale algoritmo utilizzare per portare il sistema alla pressione desiderata (*pcoupl*) e i relativi comandi riguardanti i parametri per l'algoritmo di Berendsen (Berendsen *et al.*, 1984) (*taup; ref\_p; compressibility; refcoord-scaling*) che permettono di portare il nostro sistema alla pressione di 1 bar (*ref\_p*). Il file prevede inoltre le istruzioni per generare in modo casuale le velocità del sistema (*gen\_vel; gen\_seed*) così da avere il valore iniziale per l'algoritmo di Berendsen, che accoppia velocità e pressione (Abraham *et al.*, 2005).

Infine sono riportate le istruzioni per l'algoritmo con cui applicare eventuali vincoli rigidi su molecole di cui si desidera limitare la mobilità (*constraints-algorithm*) e i relativi parametri (*lincs-order; lincs-iter*), (Abraham *et al.*, 2005).

### 2.2.7 Simulazione della traiettoria

```
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md.tpr
gmx mdrun -v -deffnm md
```

Una volta effettuata l'equilibratura del sistema si può procedere con la simulazione della traiettoria delle particelle nel sistema per un periodo pari a 60 ns. Al termine di questa si ottengono diversi file, il più importante dei quali, *md.xtc*, contiene tutte le traiettorie delle molecole del sistema per tutta la durata della simulazione; su di esso verranno effettuate tutte le considerazioni e i calcoli relativi al presente lavoro.

La linea di comando per il calcolo della traiettoria, tramite il comando *grompp*, prende il file di struttura prodotto dall'equilibratura, *npt.gro*, lo unisce al file di testo con estensione *mdp*, *md.mdp*, prende i dati dell'ultimo frame della traiettoria prodotta dall'equilibratura (-

*t npt.cpt*) e crea il file *md.tpr* tramite cui può essere iniziata la simulazione. Anche in questo caso viene aggiornato il file di topologia *topol.top*.

Dopo aver creato il file *md.tpr* si procede con il calcolo della traiettoria. Il comando *mdrun* in questo caso permette di generare file denominati *md* (tramite il comando *-defnm - md*) e produce un file partendo dal file di input *md.tpr*. In questo caso quindi verranno calcolati 30 000 000 di step, ognuno pari a 2 femsecondi, in cui si implementano gli algoritmi di calcolo per le equazioni del moto della meccanica newtoniana classica. Alla fine del calcolo si ha un sistema in cui la posizione e la velocità spaziale delle particelle corrisponde idealmente a quella di un sistema reale nelle medesime condizioni (Abraham *et al.*, 2005).

### 2.2.7.1 File di calcolo della traiettoria (*md.mdp*)

Il file con le istruzioni per la simulazione (*md.mdp*) (v. Appendice A.3) include informazioni simili al file *npt.mdp* usato per l'equilibratura.

Per quanto riguarda la temperatura, in questo caso si utilizza un altro algoritmo, ovvero il Nosè-Hoover (Nosè, 1984), che, a differenza del Berendsen, consente di riprodurre esattamente la distribuzione canonica delle velocità alla temperatura desiderata.

Anche per la pressione, l'algoritmo usato, noto come Parrinello-Rahman (Parrinello et Rahman, 1981), è diverso rispetto a quello utilizzato nel file *npt.mdp*. Questo algoritmo risponde più rapidamente a variazioni di pressione rispetto al Berendsen, ma è più aggressivo, e per questo motivo si preferisce non usarlo in fase di equilibratura. Il file *md.mdp* prevede inoltre di partire dalle velocità del sistema in output dall'equilibratura per avere il valore iniziale da usare nell'algoritmo Parrinello-Rahman, che, così come il Berendsen, accoppia velocità e pressione, (Abraham *et al.*, 2005).

## 2.3 Analisi dei risultati

A seguito della simulazione del nostro sistema, GROMACS restituisce come output almeno due file (Abraham *et al.*, 2005):

- Un file contenente la traiettoria nel tempo di tutte le molecole appartenenti al nostro sistema (*md.xtc*).
- Un file contenente la struttura della box alla fine della simulazione (*md.gro*).

Il file sicuramente più importante per i risultati che si vogliono valutare è *md.xtc*, che contiene la traiettoria del nostro sistema e sul quale sono state effettuate tutte le analisi in seguito descritte.

### 2.3.1 Funzione di distribuzione radiale (*rdf*)

```
gmx rdf -f md.xtc -s md.tpr -rdf mol_com -o rdf_ps.svg -b 15000
```

Il primo parametro che si valuta per ogni *box* in esame è la funzione di distribuzione radiale di alcune molecole del nostro sistema rispetto ad altre:

- Distribuzione del solvente rispetto agli amminoacidi
- Distribuzione dello zucchero rispetto agli amminoacidi

- Distribuzione del solvente rispetto alla backbone degli amminoacidi
- Distribuzione dello zucchero rispetto alla backbone degli amminoacidi
- Distribuzione del solvente rispetto alla sidechain degli amminoacidi
- Distribuzione dello zucchero rispetto alla sidechain degli amminoacidi
- Distribuzione dello zucchero rispetto alle altre molecole di zucchero
- Distribuzione dello zucchero rispetto alle molecole di solvente

Dalla funzione di distribuzione radiale riusciamo a ottenere importanti informazioni riguardo i meccanismi di stabilizzazione degli zuccheri in una soluzione contenente dei residui polipeptidici, e la sua analisi, non solo sull'amminoacido in sé, ma anche sulle sue componenti, permette di avere una visione più completa dei fenomeni in gioco.

Il comando per l'analisi della funzione di distribuzione radiale è unico per tutti i casi e, una volta lanciato, permette di scegliere di quale coppia di componenti valutare la distribuzione. Nell'esempio riportato a inizio paragrafo si genera in output (-o) un file contenente la *rdf* della coppia amminoacido-zucchero (indicata con *ps*) che può essere successivamente analizzata. Il comando per la funzione di distribuzione radiale prende in input il file della traiettoria (*md.xtc*) e il file *tpr* della simulazione (*md.tpr*), e restituisce in output un file di estensione *xvg*, scelta poiché utilizzabile sia per la generazione di immagini, tramite il software *Grace*, che per l'analisi numerica tramite uno script Python. Il comando *-rdf mol\_com* indica che la funzione di distribuzione radiale viene valutata considerando i centri di massa delle molecole poste attorno al riferimento. Infine è presente il comando *-b 15000* che indica di iniziare la valutazione della *rdf* dal 15 ns di simulazione. La scelta del tempo di inizio presume che da quel momento in poi il sistema sia già arrivato in una situazione di equilibrio cinetico e termodinamico e la valutazione della distribuzione radiale possa esser quindi valutata e mediata temporalmente nel periodo che intercorre dal 15esimo ns alla fine della simulazione.

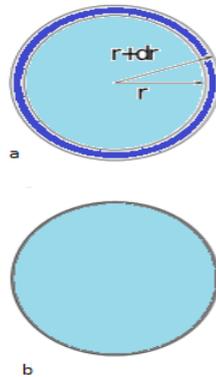
La funzione di distribuzione radiale  $g_{AB}$  di una molecola B rispetto a una molecola A è definita su GROMACS (Abraham *et al.*, 2005) nel seguente modo:

$$g_{AB} = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{local}} = \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij} - r)}{4\pi r^2} \quad (2.12)$$

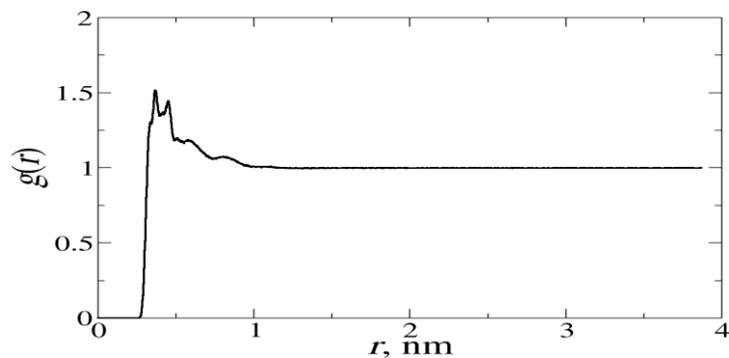
Con  $\langle \rho_B(r) \rangle$  che indica la densità delle particelle B a distanza  $r$  da A, mentre  $\langle \rho_B \rangle_{local}$  è il termine di normalizzazione rispetto alla *box* ed indica la media della densità di particelle B, che si trovano entro la distanza  $r_{max}$ , attorno a tutte le particelle di A. In questo caso  $r_{max}$  è uguale a metà del lato della *box* (impostazione di default utilizzata da GROMACS), ossia 4 nm. Come è stato scritto in precedenza, la media viene fatta anche nel tempo oltre che nello spazio, e viene ulteriormente mediata rispetto a tutte e 25 le molecole di amminoacido, *backbone* o *sidechain*.

In Figura 2.12a si può vedere come il sistema venga diviso in gusci sferici di raggio compreso tra  $r$  e  $r+dr$  in cui viene valutata la *rdf* rispetto alla particella di riferimento, idealmente posta al centro della sfera. In Figura 2.12b si può invece vedere il valore calcolato per la normalizzazione, che prende la media di tutte le particelle presenti nella sfera attorno alla particella di riferimento. In Figura 2.13 è invece riportata una generica funzione di distribuzione radiale calcolate nelle simulazioni. Come si può vedere è una

funzione resa continua partendo da valori discreti, e, allontanandosi dalla molecola, il valore della *rdf* tende a 1, indicando quindi che la densità di molecole a questa distanza corrisponde alla densità media della *box*.



**Figura 2.12 a e b.** In figura a è possibile vedere il guscio sferico in cui si valuta  $\langle \rho_B(r) \rangle$ , mentre in figura b è possibile invece vedere la normalizzazione  $\langle \rho_B \rangle_{local}$ . Immagine tratta da [manual.gromacs.org/current/reference-manual/analysis/radial-distribution-function.html](http://manual.gromacs.org/current/reference-manual/analysis/radial-distribution-function.html) con modifiche.



**Figura 2.13** Grafico che rappresenta la funzione di distribuzione radiale per il backbone della Treonina e l'acqua. Si può vedere sull'asse delle ordinate la distribuzione delle molecole rispetto alla media della *box*, mentre sull'asse delle ordinate la distanza dal centro di massa del backbone.

### 2.3.2 Valutazione dei legami a idrogeno

```
gmx hbond -f md.xtc -s md.tpr -o hbonds_bw.svg -b 15000
gmx analyze -f hbonds.svg
```

Fondamentale per confermare la teoria della stabilizzazione tramite *water replacement* (Paragrafo 1.4.2) è la valutazione dei legami idrogeno che si sono formati tra:

- Solvente e amminoacidi
- Zucchero e amminoacidi
- Solvente e backbone degli amminoacidi
- Zucchero e backbone degli amminoacidi

- Solvente e sidechain degli amminoacidi
- Zucchero e sidechain degli amminoacidi

Questa operazione viene fatta così da analizzare quanti legami siano stati sostituiti dagli zuccheri una volta sublimata l'acqua e per verificare che effettivamente gli zuccheri formino legami a idrogeno, fondamentali nel rallentare la denaturazione della proteina durante l'essiccamento (Paragrafo 1.4.2). Anche in questo caso si valuta sia l'amminoacido nella sua interezza, che i contributi delle sue singole parti; si può anticipare che i legami idrogeno formati da un amminoacido corrispondono alla somma dei legami formati dalla backbone e dalla sidechain (Abraham *et al.*, 2005).

Il comando per l'analisi dei legami a idrogeno è unico per tutti i casi e, una volta avviato, permette di scegliere di quale coppia di componenti valutare la distribuzione. Nell'esempio riportato a inizio paragrafo si genera in output (-o) un file di testo contenente il numero di legami a idrogeno della coppia backbone-acqua (indicata con *bw*) che può essere successivamente analizzata e mediata sul numero totale di amminoacidi. Il comando *hbond* prende in input il file della traiettoria (*md.xtc*) e il file *tpr* della simulazione (*md.tpr*), e restituisce in output (-o) un file di estensione *xvg*, scelta poiché analizzabile con il comando *analyze*. Il comando *-b 15000*, come nel caso della *rdf*, indica di iniziare la valutazione *hbond* dal 15 ns di simulazione.

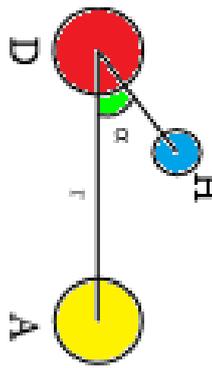
Il comando *analyze* viene inserito una volta generato il file *xvg* e permette di effettuare la lettura e mediare i valori riportati al suo interno. In output (-o) viene generato il file *hbonds.xvg*, che mostra all'utente il valore medio dei legami idrogeno formati nella *box* per la coppia selezionata e la sua deviazione standard.

Dal punto di vista computazionale il comando *hbond* analizza i legami che si formano tra un gruppo idrogeno (H) – accettore (A) – donatore (D) (Figura 2.14). L'ossigeno è sempre l'accettore, OH e NH possono essere i donatori. Per determinare dalla traiettoria se effettivamente esista un legame a idrogeno, si applica il seguente criterio geometrico (Abraham *et al.*, 2005):

**Tabella 2.2** Parametri utilizzati per la valutazione dei legami a idrogeno (Abraham *et al.*, 2005).

Misura	Criterio	Valore
Distanza	$r \leq r_{HB}$	0.35 nm
Angolo	$\alpha \leq \alpha_{HB}$	30°

Il valore di  $r_{HB}$  è stato scelto poiché corrisponde al primo punto di minimo della *rdf* dell'acqua con topologia SPC (Abraham *et al.*, 2005; Pisano e Arsiccio, 2017).



**Figura 2.14** Rappresentazione del criterio geometrico della valutazione dei legami a idrogeno su GROMACS. Immagine tratta da (Abraham *et al.*, 2005) con modifiche.

Il comando *hbond* quindi analizza i valori di  $r$  tra donatore e accettore e i valori di  $\alpha$  tra idrogeno, donatore e accettore in ogni frame di tempo indicato dall'utente (Abraham *et al.*, 2005).

### 2.3.3 Valutazione superficie esposta e analisi

```
gmx sasa -f md.xtc -s md.tpr -o scsasa.xvg
gmx analyze -f scsasa.xvg
```

Un altro parametro di interesse da valutare nelle simulazioni è la superficie esposta alla soluzione acqua-zucchero da ogni singolo amminoacido. La valutazione viene fatta sia per il backbone che per la sidechain del residuo.

Il comando *sasa* prende in input il file della traiettoria della simulazione e il file *tpr* da essa generato, producendo in output (-o) un file *xvg*, con elencate le aree esposte di ogni amminoacido della *box*. Il comando permette di far scegliere all'utente quale molecola e quale sua parte analizzare nella sua area esposta alla soluzione, permettendo così di effettuare le analisi richieste.

Dal punto di vista computazionale, l'area superficiale esposta viene in genere calcolata da GROMACS (Abraham *et al.*, 2005) immaginando di far 'rotolare' ('rolling-ball algorithm') al di sopra della molecola di interesse una sfera di raggio pari a quello delle molecole di solvente, ovvero 1.4 Å nel caso delle molecole di acqua (Shrake e Rupley, 1973). Questo è solo uno dei diversi metodi possibili proposti in letteratura. Considerando che le pubblicazioni di Auton e Bolen, cui l'ultima parte del presente lavoro fa riferimento per le verifiche sull'additività delle interazioni proteina-cosoluto, utilizzano un algoritmo diverso, teorizzato da Lee e Richards (Lee e Richards, 1971), si è deciso di utilizzare il loro stesso metodo di calcolo per le aree esposte nei polipeptidi (programma *freesasa 2.0.3*).

### 2.3.4 Valutazione del clustering

```
gmx clustsize -f md.xtc -s md.tpr -o nclust.xvg -b 15000
gmx analyze -f nclust.xvg
```

Un altro parametro di interesse riguarda il *clustering*, ossia la formazione di aggregati tra i singoli residui all'interno della *box*. L'aggregazione infatti può essere un meccanismo di denaturazione (Wang e Roberts, 2010, paragrafo 1.4) e un buon eccipiente dovrebbe prevenirne la formazione. GROMACS permette la valutazione del *clustering* tramite il comando *clustsize*, che prende in input il file della traiettoria (*md.xtc*) e il file *tpr* della simulazione (*md.tpr*) e restituisce in output (-o) il file *nclust.xvg*, in cui è riportato il numero medio dei cluster rilevati a partire dal nanosecondo 15 della nostra simulazione (-b 15000). In seguito tramite il comando *analyze* viene restituito il numero medio di cluster presenti. Il numero, nel caso delle presenti simulazioni, è compreso tra 1 e 25: 1 indica che i residui si sono aggregati in unico cluster, 25 indica invece che ogni residuo è rimasto separato dagli altri (Abraham *et al.*, 2015).

### 2.3.5 Analisi dei KBI

Il parametro probabilmente più importante per questo lavoro è l'integrale di Kirkwood-Buff, *Kirkwood-Buff Integral (KBI)*, che viene calcolato tramite uno script Python (v. Appendice A.4) a partire dalla funzione di distribuzione radiale nel seguente modo (Ganguly *et al.*, 2015):

$$KBI = G_{ij} = 4\pi \int_0^{\infty} [g_{ij}^{nPT}(r) - 1] r^2 dr \quad (2.13)$$

La funzione di distribuzione radiale tra due diverse molecole *i* e *j* viene indicata qui con *g<sub>ij</sub>*, mentre gli estremi di integrazione, nel nostro caso, vanno dalla particella di riferimento sino alla lunghezza convenzionale di metà della *box*, pari a 4 nm, e pari anche alla distanza massima in cui è stata valutata la *rdf*.

L'integrale di Kirkwood-Buff è un parametro utilizzato per coniugare le correlazioni tra le molecole, come ad esempio la funzione di distribuzione radiale, con proprietà della soluzione, come la variazione del potenziale chimico, la compressibilità, il volume parziale molare (Ganguly *et al.*, 2015; Kirkwood e Buff, 1951) e l'energia libera di solvatazione (Ganguly *et al.*, 2015). Come si può vedere nell'equazione 2.13 il KBI è un integrale di volume applicabile su un insieme canonico a numero di molecole, pressione e temperatura costanti. Il valore dell'integrale può essere utilizzato sia per considerazioni immediate sulla esclusione/inclusione di una molecola rispetto alla superficie di un'altra, sia essere combinato con altre costanti per determinare la variazione di energia libera della soluzione al variare della concentrazione di un determinato soluto.

Nel primo caso si può semplicemente dire che se  $KBI > 0$ , la particella *j* tende ad essere accumulata nei dintorni della particella *i*, mentre se  $KBI < 0$  si può parlare di esclusione dalla superficie (Wang e Roberts, 2010; Shulgin e Ruckenstein, 2008).

I KBI possono anche essere correlati con la teoria del *soluto escluso*, presentata nel paragrafo 1.4.3. Un parametro che può essere calcolato partendo dai KBI è il *coefficiente di esclusione preferenziale*, così definito (Wang e Roberts, 2010):

$$\Gamma_{32} = \rho_3 (G_{23} - G_{21}) = M_1 N_{Av} c_1 m_3 (G_{23} - G_{21}) \quad (2.14)$$

La notazione è la medesima dell'equazione 1.9. In questo caso si valuta il *coefficiente di esclusione preferenziale* tra la proteina (2) e lo zucchero (3). Questo è uguale al prodotto tra la densità molecolare dello zucchero nella soluzione ( $\rho_3$ ) e la differenza dei KBI per le coppie proteina-zucchero ( $G_{23}$ ) e proteina-solvente ( $G_{21}$ ). Chiaramente  $\rho_3$  può anche essere espressa come prodotto della massa molare dell'acqua ( $M_1$ ), il numero di Avogadro ( $N_{Av}$ ), la concentrazione molare dell'acqua ( $c_1$ ) e la molalità dello zucchero ( $m_3$ ).

Il *coefficiente di esclusione preferenziale* misura quindi la concentrazione locale di soluto nell'intorno di una proteina rispetto a quella del bulk, e può altrimenti essere così espresso (Cloutier *et al.*, 2018):

$$\Gamma_{32}(r, t) = n_3(r, t) - n_1(r, t) \left( \frac{n_3 - n_3(r, t)}{n_1 - n_1(r, t)} \right) \quad (2.15)$$

Dove  $n_i$  indica il numero totale di particelle presenti nella simulazione. Il valore che si ottiene nelle simulazioni effettuate è una media temporale e spaziale del valore rilevato per ogni distanza  $r$  e tempo  $t$ .

La correlazione tra il *coefficiente di esclusione preferenziale* e la teoria del *soluto escluso* di Timasheff e Arakawa (paragrafo 1.4.3) permette di valutare il grado di stabilizzazione della proteina da parte del cosoluto, come dimostrato con le seguenti equazioni, valide per un insieme canonico  $nPT$  (Wang e Roberts, 2010; Anderson *et al.*, 2002):

$$\left( \frac{\partial \mu_2}{\partial m_3} \right) = - \left( \frac{\partial m_2}{\partial m_3} \right) \left( \frac{\partial \mu_3}{\partial m_3} \right) \quad (1.9)$$

$$\left( \frac{\partial \mu_2}{\partial \mu_3} \right) = \left( \frac{\partial \mu_2 / \partial m_3}{\partial \mu_3 / \partial m_3} \right) = - \Gamma_{32} \quad (2.16)$$

Si può qui vedere come nel caso in cui il valore di  $\Gamma_{32}$  sia minore di 0, la variazione di potenziale chimico della proteina al variare della concentrazione di soluto sia positiva. Questo indica che nel momento in cui attorno a una proteina la concentrazione locale di molecole di soluto è inferiore a quella di molecole di solvente (Wang e Roberts, 2010), la configurazione stabilizzata è quella nativa (paragrafo 1.4.3).

Si può anche calcolare la variazione di energia libera di Gibbs, da stato *nativo* ad *unfolded*, al variare della concentrazione di protettore, tramite l'utilizzo dei coefficienti di

interazione preferenziale, valutati in condizioni di proteina allo stato nativo ( $\Gamma_{32}^{\text{Nat}}$ ) e allo stato denaturato ( $\Gamma_{32}^{\text{Den}}$ ). L'equazione è la seguente (Wang e Roberts, 2010):

$$\frac{d \Delta G}{d m_3} = [\Gamma_{32}^{\text{Nat}} - \Gamma_{32}^{\text{Den}}] \left( \frac{\partial \mu_3}{\delta m_3} \right) \quad (2.17)$$

## 2.4 Additività delle interazioni proteina-cosoluto

L'ultima parte del presente lavoro di tesi si occupa di valutare l'additività delle interazioni tra eccipienti e polipeptidi, ovvero di verificare se l'interazione eccipiente-proteina possa essere ottenuta come somma delle interazioni tra lo stesso eccipiente ed i singoli amminoacidi presenti nella struttura primaria del polipeptide. Per fare ciò si prende spunto dai lavori di Auton e Bolen, ricercatori che per primi hanno proposto questo approccio. Nel loro lavoro, essi hanno verificato sperimentalmente per un ampio set di proteine modello l'esistenza di additività tra le energie libere di trasferimento ( $\Delta G_{\text{trasf}}$ ) dei singoli amminoacidi componenti la proteina.  $\Delta G_{\text{trasf}}$  è definita come la variazione di energia libera conseguente al trasferimento del composto di riferimento da soluzioni acquose a soluzioni 1 M di comuni osmoliti (Auton e Bolen, 2005).

È importante stabilire una teoria che, in qualsiasi condizione di concentrazione del protettore, o di scala chimica utilizzata, permetta di verificare l'additività. Per ottenere questo risultato è necessario prima esaminare quali equazioni permettano di esprimere l'energia libera di trasferimento. Trascurando la variazione di attività della proteina in presenza di soluzioni con protettore, per la scala molare e molale, i risultati sono i seguenti (Auton e Bolen, 2004):

$$\Delta G_{app,c} = RT \ln \left( \frac{n_{i,w}}{n_{i,os}} \right) + RT \ln \left( \frac{V_{s,os}}{V_{s,w}} \right) \quad (2.18)$$

$$\Delta G_{app,m} = RT \ln \left( \frac{n_{i,w}}{n_{i,os}} \right) + RT \ln \left( \frac{wt_{os}}{wt_w} \right) \quad (2.19)$$

Dove  $n_{i,w}$  e  $n_{i,os}$  rappresentano le moli di soluto solubili in 100 g di soluzione acquosa e in 100 g di soluzione 1 M di protettore.  $V_{s,os}$  e  $V_{s,w}$  sono i volumi totali della soluzione acquosa e di quella contenente il protettore,  $wt_{os}$  e  $wt_w$  sono la massa totale di acqua e di soluzione 1 M di protettore.

Come si può osservare, i valori di energia libera di trasferimento sono qui indicati come *apparenti*, ossia legati alla scala di misurazione utilizzata. È dunque necessario normalizzare i risultati ottenuti per avere un risultato univoco. Per far ciò l'approccio indicato da Auton e Bolen prevede di sottrarre dalle equazioni 2.18 e 2.19 le medesime, valutate però in condizioni di  $n_i \rightarrow 0$ .

Il valore di correzione viene chiamato  $\Delta G_{SC,\infty}$  e per la scala molare e molale corrisponde ai seguenti valori (Auton e Bolen, 2004):

$$\Delta G_{SC,\infty, c} = \Delta G_{int} + RT \ln (d_w/d_{os}) \quad (2.20)$$

$$\Delta G_{SC,\infty, m} = \Delta G_{int} \quad (2.21)$$

Dove  $d_w$  e  $d_{os}$  sono rispettivamente la densità dell'acqua pura e quella della soluzione 1 M di soluto.

Il valore di energia libera di trasferimento intrinseca per una soluzione è definita come  $\Delta G_{int}$ , che, a prescindere dal sistema di riferimento, può esser così calcolata (Auton e Bolen, 2004):

$$\Delta G_{int} = \Delta G_{app} - \Delta G_{SC,\infty} = \Delta \mu^\circ + RT \ln (\gamma_{i,w}/\gamma_{i,os}) \quad (2.22)$$

Dove  $\gamma_{i,w}$  e  $\gamma_{i,os}$  sono i valori di solubilità del soluto in acqua pura e in soluzione 1M di protettore. Per solubilità basse la variazione di energia interna corrisponde alla variazione di potenziale chimico standard.

Utilizzando valori di energie di trasferimento così calcolati, Auton e Bolen hanno dimostrato che l'energia libera di trasferimento di una catena polipeptidica può essere ottenuta come media pesata delle energie libere di trasferimento dei singoli amminoacidi costituenti la catena.

L'additività si può ottenere però solo pesando i  $\Delta G_{trasf}$  dei singoli amminoacidi tramite le aree esposte da ogni singolo residuo nel polipeptide, e comporta l'adozione di alcune convenzioni per la valutazione dei parametri riguardanti il contributo di *backbone* e *sidechain* per migliorare i risultati ottenuti, come qui riportato:

$$\begin{aligned} \left( \frac{\partial \mu_2}{\delta m_3} \right) &= \Delta G_{trasf} = \\ &= \sum_i n_i \alpha_i^{sc} \Delta G_{trasf}^{sc} + \Delta G_{trasf}^{bb} \sum_i n_i \alpha_i^{bb} \end{aligned} \quad (2.23)$$

Nell'equazione 2.23 compaiono per la prima volta i termini  $\Delta G_{trasf}^{sc}$  e  $\Delta G_{trasf}^{bb}$ , che rappresentano l'energia libera di trasferimento di *backbone* e *sidechain* del residuo *i-esimo*, e i termini  $\alpha_i^{sc}$  e  $\alpha_i^{bb}$ , che rappresentano le aree frazionali esposte di *backbone* e *sidechain* del residuo *i-esimo*. Nel seguito si forniscono informazioni più dettagliate riguardo la valutazione di questi termini.

L'obiettivo del presente lavoro è quello di valutare se l'additività sia confermata anche per i  $\Gamma_{32}$  dei polipeptidi rispetto agli amminoacidi che li compongono, replicando nei calcoli

l'approccio usato da Auton e Bolen per i  $\Delta G_{trasf}$  e utilizzando i risultati delle simulazioni effettuate su GROMACS. L'approccio proposto è ammissibile per il fatto che il  $\Delta G_{trasf}$  di *backbone* e *sidechain* sono legati a  $\Gamma_{32}$  dalle seguenti equazioni (Wang e Roberts, 2010):

$$\Gamma_{32}^{bb} = \frac{-\Delta G_{trasf}^{bb}}{\frac{d\mu_3}{dc_3}} \quad (2.24)$$

$$\Gamma_{32}^{sc} = \frac{-\Delta G_{trasf}^{sc}}{\frac{d\mu_3}{dc_3}} \quad (2.25)$$

Nel seguito si spiega nel dettaglio l'approccio utilizzato sia in letteratura che nel nostro lavoro per le valutazioni sull'additività.

### 2.4.1 Backbone

Volendo replicare l'approccio utilizzato da Auton e Bolen si vuole iniziare riportando che nelle loro pubblicazioni, per le catene polipeptidiche, si valutano, in modo separato e con metodi diversi, i contributi di *backbone* e *sidechain* (Auton e Bolen, 2005), poichè solamente in questo modo si ottengono i risultati di additività desiderati. Alla luce delle equazioni 2.24 e 2.25, è necessario replicare questo approccio anche per i valori del *coefficiente di esclusione preferenziale*.

La descrizione dei possibili approcci utilizzati per valutare il  $\Delta G_{trasf}$  del *backbone* viene riportata in una pubblicazione del 2004 (Auton e Bolen, 2004).

Innanzitutto, per ciò che riguarda le interazioni del *backbone*, la forza spingente per il meccanismo di protezione risiede nella interazione solvofobica di questa struttura con il solvente (Bolen e Baskakov, 2001). Poiché il *backbone* è lo stesso per ogni residuo polipeptidico, si può ipotizzare che eventuali variazioni di contributo all'energia libera di trasferimento tra i diversi amminoacidi siano dovute esclusivamente alla *sidechain* (Auton e Bolen, 2004). A seguito di questa considerazione si è quindi provveduto a cercare dei modelli peptidici con caratteristiche simili a quelle del *backbone*, attraverso cui valutare il valore di energia libera di trasferimento comune a tutti i residui (Tanford, 1970). Per le simulazioni effettuate è stata utilizzata solamente la serie di *N-acetilglicinammide* (Figura 2.15).

Per costruire la serie di *N-acetilglicinammide* è opportuno creare una serie contenente numero variabile di residui interni di glicina (nel nostro caso compresi tra 3 e 6), ed utilizzare poi il *metodo degli incrementi costanti* (Auton e Bolen, 2004):

$$\Delta G_{tr,polip} = \Delta G_{EG} + N_i \left( \frac{\partial \Delta G_{tr}}{\partial N_{IU}} \right) \quad (2.26)$$

dove  $\Delta G_{tr,polip}$  è l'energia intrinseca di trasferimento del polipeptide considerato,  $\Delta G_{EG}$  è l'energia libera di trasferimento dei residui alle estremità del polipeptide,  $N_i$  è il numero di

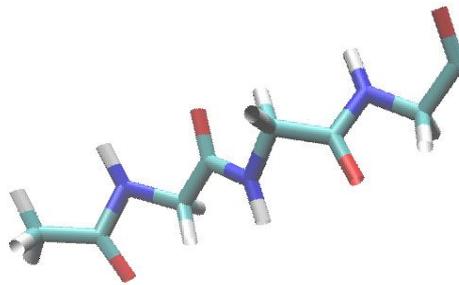
residui interni al polipeptide (variabile tra 3 e 6),  $\frac{\partial \Delta G_{tr}}{\partial N_{IU}}$  è l'energia libera di trasferimento di ogni residuo interno al polipeptide, considerata uguale per ognuno e che rappresenta il valore modello per il backbone di un generico amminoacido.

Conseguentemente, il valore di  $\Gamma_{32}^{bb}$ , viene così calcolato per analogia:

$$\Gamma_{32}^{bb} = \Gamma_{32,EG} + N_i (\Gamma_{32}^{Nacet}) \quad (2.27)$$

Dove  $\Gamma_{32,EG}$  corrisponde al valore del *coefficiente di interazione preferenziale* per i gruppi terminali e  $\Gamma_{32}^{Nacet}$  corrisponde al *coefficiente di interazione preferenziale* per la molecola modello *N-acetilglicinamide*, avente  $N_i$  unità di glicina.

Dato che la presenza di gruppi terminali carichi può essere causa di erronee misure dei valori di  $\Delta G_{EG}$  o  $\Gamma_{32,EG}$  (Auton e Bolen, 2004), nel presente lavoro si è scelto di utilizzare dei terminali acetilici e amminici alle estremità del polipeptide analizzato, così da avere valori più precisi per i dati estrapolati.



**Figura 2.15** Immagine generata con VMD della *N-acetiltryglicinamide*. In bianco gli atomi di idrogeno, in blu quelli di azoto, in rosso quelli di ossigeno, in verde acqua quelli di carbonio.

## 2.4.2 Sidechain

Per ciò che concerne la *sidechain*, ogni residuo possiede un valore di  $\Delta G_{trasf}$  diverso, poiché è proprio questa parte strutturale che distingue gli amminoacidi e le loro proprietà. Si cerca quindi di trovare un metodo di calcolo univoco per risalire al valore di  $\Delta G_{trasf}$  di ogni residuo.

Partendo da considerazioni termodinamiche, per ogni residuo si ha un valore apparente,  $\Delta G_{app}$ , di energia libera di trasferimento, determinabile sperimentalmente (Auton *et al.*, 2007):

$$\Delta G_{app} = RT \ln \left( \frac{S_{i,w}}{S_{i,os}} \right) \quad (2.28)$$

Dove  $s_{i,w}$  e  $s_{i,os}$  indicano le concentrazioni molari di soluto in acqua pura e nella soluzione in cui è presente il protettore.

Per aver il valore intrinseco di energia libera di trasferimento per la *sidechain* di un amminoacido, si prende il valore ottenuto con l'equazione 2.28 e si effettua la differenza tra questo valore e il valore di energia libera di trasferimento calcolato per la glicina. Infatti la glicina è una molecola assimilabile al *backbone* e, ipotizzando additività anche nelle proprietà dei singoli residui (Auton e Bolen, 2004), le proprietà della *sidechain* si possono calcolare per differenza tra quelle del residuo intero e quelle del suo *backbone*. L'equazione per il calcolo dell'energia libera di trasferimento intrinseca della *sidechain* di un amminoacido  $\Delta G_{int}^{sc}$  è pertanto la seguente (Auton *et al.*, 2007):

$$\Delta G_{int}^{sc} = RT \ln (s_{i,w}/s_{i,os}) - [ RT \ln (s_{gly,w}/s_{gly,os}) + RT \ln (\gamma_{gly,w}/\gamma_{gly,os}) ] \quad (2.29)$$

Questa formula è valida solamente nel momento in cui si decida di trascurare la variazione di attività del residuo in presenza o meno di protettore. Questa approssimazione non è sempre realistica, dal momento che non tutti gli amminoacidi hanno la stessa solubilità in soluzioni diverse, ma in assenza di dati sperimentali e vista la difficoltà sperimentale nella determinazione del coefficiente di attività, è la miglior scelta possibile per verificare i risultati attesi. Utilizzando però i veri valori di energia libera di trasferimento per la glicina (di cui si è riusciti a determinare la variazione di attività) come termine di confronto nell'equazione 2.29, si ha una notevole corrispondenza con i dati sperimentali.

Allo stesso modo, per analogia, il valore di  $\Gamma_{i,32}^{sc}$  per ogni residuo è stato calcolato come:

$$\Gamma_{i,32}^{sc} = \Gamma_{32}^i - \Gamma_{32}^{gly} \quad (2.30)$$

Dove  $\Gamma_{32}^{gly}$  è il coefficiente di interazione preferenziale per la glicina,  $\Gamma_{32}^i$  quello di un generico residuo  $i$ , e  $\Gamma_{i,32}^{sc}$  il coefficiente di interazione preferenziale per la *sidechain*.

Si è detto che, oltre ai valori di energia libera di trasferimento, fondamentale importanza nel calcolo dell'additività è riservata al calcolo delle aree esposte alla soluzione da parte di ogni amminoacido costituente il polipeptide, per quanto riguarda sia il *backbone* che la *sidechain*. Il metodo utilizzato da Auton e Bolen (Auton e Bolen, 2005) prevede, per questa valutazione, l'utilizzo dell'algoritmo di Lee e Richards (Lee e Richards, 1971). L'obiettivo è quello di valutare la differenza nell'area esposta dal polipeptide e dai suoi residui tra lo stato nativo e quello denaturato, e di normalizzarla rispetto al valore massimo ottenibile.

$$\Delta\alpha_i^{bb\ o\ sc} = \frac{\sum_i (Area_{i,Den} - Area_{i,Nat})}{n_i Area_{i,Gly-X-Gly}} \quad (2.31)$$

Il valore massimo si ottiene considerando un polipeptide modello, Gly – X – Gly, in cui i due terminali sono molecole di glicina, mentre al centro è presente l'amminoacido X, di cui si desidera valutare l'area esposta. Questa operazione viene fatta sia per il *backbone* che per la *sidechain*. Nel seguito si riportano le equazioni per il calcolo dell'area utilizzate in letteratura (Auton *et al.*, 2007).

### 2.4.3 Metodi di calcolo per l'additività utilizzati nell'elaborazione dei dati

Dopo avere descritto nei paragrafi 2.4.1 e 2.4.2 le scelte effettuate per la verifica dell'additività, si riportano qui le equazioni che relazionano l'approccio additivo di Auton e Bolen con quello del presente lavoro (Wang e Roberts, 2010).

Per quanto riguarda i  $\Delta G_{trasf}$  vale:

$$\left(\frac{\partial \mu_2}{\partial m_3}\right) = \sum_i n_i \alpha_i^{sc} \Delta G_{trasf}^{sc} + \Delta G_{trasf}^{bb} \sum_i n_i \alpha_i^{bb} \quad (2.23)$$

Ma il  $\Delta G_{trasf}$  è legato al coefficiente di interazione preferenziale  $\Gamma_{32}$  da:

$$- \Gamma_{32} * \frac{d\mu_3}{dc_3} = \frac{d\mu_2}{dm_3} \quad (2.32)$$

Considerando le equazioni 2.17 e 2.18, si conclude che:

$$\Gamma_{32} = \Gamma_{32}^{bb} * \sum n_i \alpha_i^{bb} + \sum n_i \alpha_i^{sc} \Gamma_{i,32}^{sc} \quad (2.33)$$

## 2.5 Design delle simulazioni

In Tabella 2.3 si riporta un riassunto delle simulazioni effettuate, con relative condizioni di simulazione.

Le simulazioni effettuate prevedono (Paragrafo 2.2) la valutazione del comportamento di ognuno dei venti amminoacidi in presenza di saccarosio (simulazioni 1-20) e trealosio (simulazioni 21-40). Gli amminoacidi sono stati presi nella loro forma zwitterionica, ossia senza terminale acetilico e amminico ad annullarne la carica totale, e la metodologia utilizzata per le simulazioni è stata riportata ampiamente nel sottocapitolo 2.3.

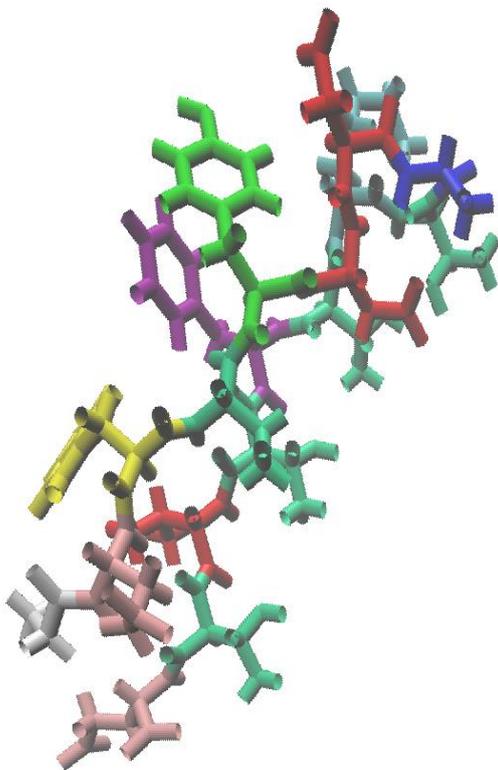
Per la valutazione dell'additività degli amminoacidi in strutture polipeptidiche si è fatto ricorso ad ulteriori simulazioni.

Sei simulazioni (simulazioni 41-46) sono state necessarie per avere il valore modello del coefficiente di esclusione preferenziale per il *backbone* (Paragrafo 2.4.3) in presenza di trealosio e saccarosio. Per far questo, con l'algoritmo riportato nel capitolo 2.3, si è creata una *box* con una molecola campione di *N-acetilglicinammide* in presenza di acqua e saccarosio o trealosio in concentrazione 1 M. Riproponendo l'approccio di Auton e Bolen (Auton e Bolen, 2004) si è valutato il coefficiente di esclusione preferenziale,  $\Gamma_{32}$  per la molecola campione composta da tre, quattro, cinque e sei residui di glicina, così da risalire

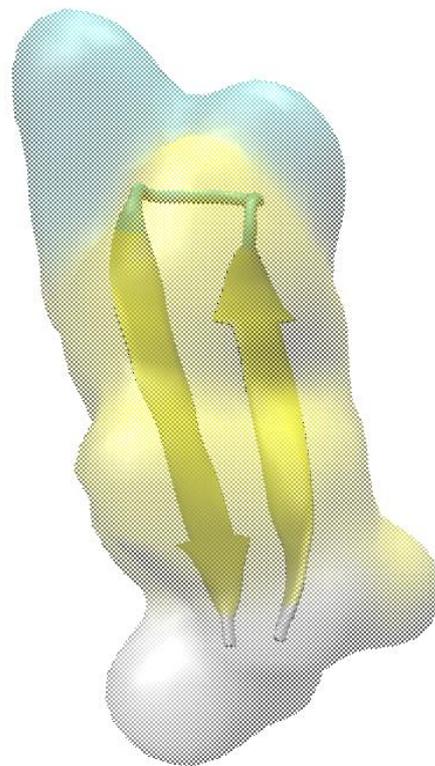
tramite interpolazione lineare e *metodo degli incrementi finiti* al valore da utilizzare come valore modello per il *backbone*.

Una simulazione (simulazione 47) è stata necessaria per avere un file di struttura e un file di topologia per la forma denaturata del GB1 – Hairpin (Figura 2.16 a e b), polipeptide composto da 16 residui e utilizzato come molecola modello per i calcoli sull'additività. Volendo infatti valutare non solo la forma nativa del polipeptide, ma anche la forma denaturata, si è reso necessario ottenere quest'ultima tramite una apposita simulazione. Per far questo si è scelto di mettere una molecola di GB1 in forma nativa in una *box* in cui fossero presenti urea in concentrazione 8 M e acqua. Ampia è la letteratura che dimostra la capacità denaturante dell'urea (Ganguly *et al.*, 2015; Auton *et al.*, 2007), e per questo si è scelta per ottenere il GB1 in forma denaturata (Figura 2.16c). La simulazione, svolta seguendo l'algoritmo presentato nel capitolo 2.3, è stata effettuata a una temperatura di 500 K, così da aumentare la velocità di denaturazione, ed è stata monitorata per un tempo (non impostato quindi a 60 ns) tale per cui il *raggio di giro* del GB1, ossia il raggio formato dalla struttura  $\beta$ -hairpin del polipeptide, raggiungesse valori molto alti, segno di una perdita totale della struttura ripiegata a favore di una lineare, tipica della forma denaturata.

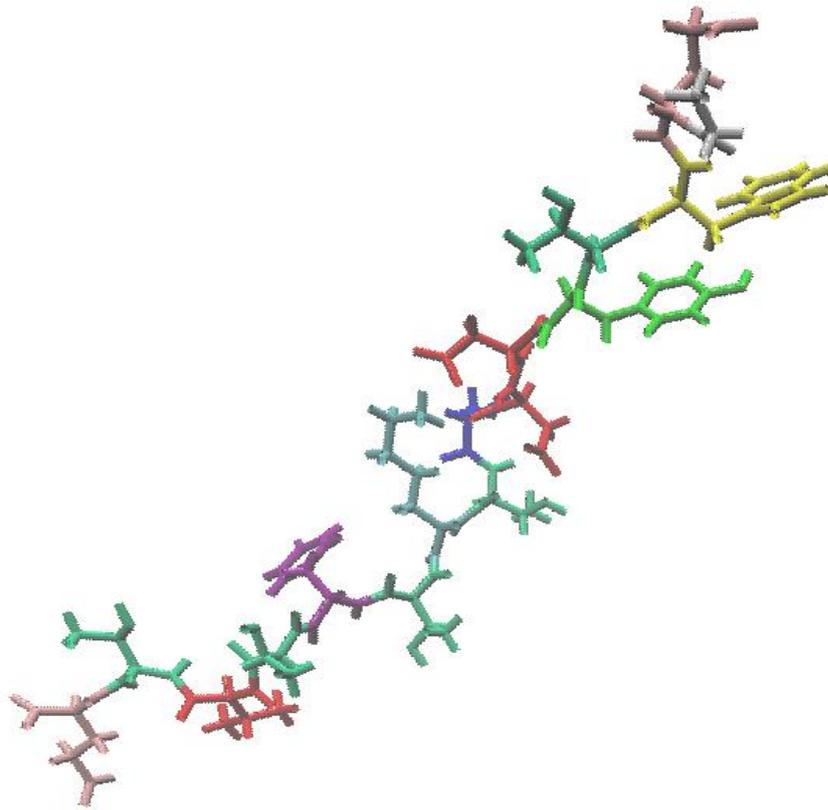
Quattro simulazioni invece, seguendo l'algoritmo descritto nel capitolo 2.3, sono state realizzate per valutare le interazioni tra GB1, in forma nativa e denaturata, con trealosio e saccarosio. È stata sempre posta in una *box* contenente concentrazione 1 M di protettore e acqua, una molecola di GB1, e alla fine della simulazione, sempre della durata di 60 ns, si sono raccolti i risultati.



**Figura 2.16a** Immagine generata con VMD del GB1 in forma  $\beta$ -hairpin. Ogni residuo ha un colore diverso



**Figura 2.16b** Immagine generata con VMD della struttura  $\beta$ -hairpin.



**Figura 2.16c** Immagine generata con VMD del GB1 in forma denaturata. Ogni residuo ha un colore diverso.

**Tabella 2.3** *Riassunto delle simulazioni effettuate.*

<b>N°</b>	<b>Amminoacido</b>	<b>Numero di Amminoacidi</b>	<b>Eccipiente</b>	<b>Concentrazione Eccipiente</b>	<b>Box [nm]</b>	<b>T [K]</b>	<b>P [bar]</b>
1	Glicina (Gly)	25	Saccarosio	1 M	8x8x8	300	1
2	Alanina (Ala)	25	Saccarosio	1 M	8x8x8	300	1
3	Prolina (Pro)	25	Saccarosio	1 M	8x8x8	300	1
4	Valina (Val)	25	Saccarosio	1 M	8x8x8	300	1
5	Leucina (Leu)	25	Saccarosio	1 M	8x8x8	300	1
6	Isoleucina (Ile)	25	Saccarosio	1 M	8x8x8	300	1
7	Serina (Ser)	25	Saccarosio	1 M	8x8x8	300	1
8	Treonina (Thr)	25	Saccarosio	1 M	8x8x8	300	1
9	Cisteina (Cys)	25	Saccarosio	1 M	8x8x8	300	1
10	Asparagina (Asn)	25	Saccarosio	1 M	8x8x8	300	1
11	Glutammina (Gln)	25	Saccarosio	1 M	8x8x8	300	1
12	Metionina (Met)	25	Saccarosio	1 M	8x8x8	300	1
13	Fenilalanina (Phe)	25	Saccarosio	1 M	8x8x8	300	1
14	Tirosina (Tyr)	25	Saccarosio	1 M	8x8x8	300	1
15	Triptofano (Trp)	25	Saccarosio	1 M	8x8x8	300	1
16	Lisina (Lys)	25	Saccarosio	1 M	8x8x8	300	1
17	Istidina (Hys)	25	Saccarosio	1 M	8x8x8	300	1
18	Arginina (Arg)	25	Saccarosio	1 M	8x8x8	300	1

<b>19</b>	Aspartato (Asp)	25	Saccarosio	1 M	8x8x8	300	1
<b>20</b>	Glutammato (Glu)	25	Saccarosio	1 M	8x8x8	300	1
<b>21</b>	Glicina (Gly)	25	Trealosio	1 M	8x8x8	300	1
<b>22</b>	Alanina (Ala)	25	Trealosio	1 M	8x8x8	300	1
<b>23</b>	Prolina (Pro)	25	Trealosio	1 M	8x8x8	300	1
<b>24</b>	Valina (Val)	25	Trealosio	1 M	8x8x8	300	1
<b>25</b>	Leucina (Leu)	25	Trealosio	1 M	8x8x8	300	1
<b>26</b>	Isoleucina (Ile)	25	Trealosio	1 M	8x8x8	300	1
<b>27</b>	Serina (Ser)	25	Trealosio	1 M	8x8x8	300	1
<b>28</b>	Treonina (Thr)	25	Trealosio	1 M	8x8x8	300	1
<b>29</b>	Cisteina (Cys)	25	Trealosio	1 M	8x8x8	300	1
<b>30</b>	Asparagina (Asn)	25	Trealosio	1 M	8x8x8	300	1
<b>31</b>	Glutammina (Gln)	25	Trealosio	1 M	8x8x8	300	1
<b>32</b>	Metionina (Met)	25	Trealosio	1 M	8x8x8	300	1
<b>33</b>	Fenilalanina (Phe)	25	Trealosio	1 M	8x8x8	300	1
<b>34</b>	Tirosina (Tyr)	25	Trealosio	1 M	8x8x8	300	1
<b>35</b>	Triptofano (Trp)	25	Trealosio	1 M	8x8x8	300	1
<b>36</b>	Lisina (Lys)	25	Trealosio	1 M	8x8x8	300	1
<b>37</b>	Istidina (Hys)	25	Trealosio	1 M	8x8x8	300	1
<b>38</b>	Arginina (Arg)	25	Trealosio	1 M	8x8x8	300	1

39	Aspartato (Asp)	25	Trealosio	1 M	8x8x8	300	1
40	Glutammato (Glu)	25	Trealosio	1 M	8x8x8	300	1
41	N-acetil triglicinammide	1	Saccarosio	1 M	8x8x8	300	1
42	N-acetil triglicinammide	1	Trealosio	1 M	8x8x8	300	1
43	N-acetil tetraglicinammide	1	Saccarosio	1 M	8x8x8	300	1
44	N-acetil tetraglicinammide	1	Trealosio	1 M	8x8x8	300	1
45	N-acetil pentaglicinammid e	1	Saccarosio	1 M	8x8x8	300	1
46	N-acetil pentaglicinammid e	1	Trealosio	1 M	8x8x8	300	1
47	N-acetil esaglicinammide	1	Saccarosio	1 M	8x8x8	300	1
48	N-acetil esaglicinammide	1	Trealosio	1 M	8x8x8	300	1
49	GB1 – Hairpin Nativo	1	Urea	1 M	8x8x8	500	1
50	GB1 – Hairpin Nativo	1	Saccarosio	1 M	8x8x8	300	1
51	GB1 – Hairpin Denaturato	1	Saccarosio	1 M	8x8x8	300	1
52	GB1 – Hairpin Nativo	1	Trealosio	1 M	8x8x8	300	1
53	GB1 – Hairpin Denaturato	1	Trealosio	1 M	8x8x8	300	1



## 3. RISULTATI

### 3.1 Introduzione

In quest'ultima parte dell'elaborato si riportano i risultati ottenuti dalle simulazioni e le verifiche sull'additività delle proprietà degli amminoacidi in polipeptidi complessi. Il capitolo si compone di due parti. La prima sarà dedicata all'analisi dei singoli amminoacidi in forma zwitterionica, messi in soluzione acquosa di saccarosio o trealosio in concentrazione 1 M. Nella seconda parte invece si prendono i dati raccolti nella prima parte e li si elabora secondo il metodo di Auton e Bolen descritto nel paragrafo 2.4. Si riportano quindi i risultati delle simulazioni effettuate per valutare il coefficiente di esclusione preferenziale di saccarosio e trealosio dal backbone proteico. L'obiettivo principale di quest'ultima parte è quello di verificare se l'interazione tra gli eccipienti e polipeptidi modello risponda ad una relazione di additività, e sia pertanto ottenibile come somma delle interazioni eccipiente-amminoacido (Auton *et al.*, 2007).

### 3.2 Comportamento degli amminoacidi in presenza di saccarosio o trealosio

In questa prima parte, come criterio di distinzione tra le classi di amminoacidi, si è scelto di utilizzare la divisione canonica presente in letteratura, ossia la divisione a seconda delle proprietà della catena laterale (Paragrafo 1.3.1). Le classi sono le seguenti:

- Amminoacidi non polari: glicina, alanina, valina, prolina, isoleucina, leucina
- Amminoacidi polari: serina, treonina, cisteina, asparagina, glutammina, metionina
- Amminoacidi aromatici: fenilalanina, tirosina, triptofano
- Amminoacidi carichi positivamente: lisina, istidina, arginina
- Amminoacidi carichi negativamente: aspartato e glutammato.

All'interno delle varie classi, come criterio di distinzione, si è scelto di ordinare i residui in ordine di peso molecolare crescente, poiché risulta criterio omogeneo per tutte le classi elencate.

Oltre al valutare le proprietà in relazione alla classificazione canonica, si è scelto di evidenziare in ogni grafico il confronto tra le proprietà dello stesso residuo in presenza di saccarosio o trealosio. L'analisi è stata fatta per verificare, alla luce delle teorie di esclusione preferenziale e water replacement, differenze nel comportamento dei due protettori in presenza di un determinato amminoacido. Si fornisce un'ampia panoramica delle proprietà degli amminoacidi in forma zwitterionica in presenza di saccarosio e trealosio, valutate tramite il campo di forza CHARMM36 modificato da Trout (si veda paragrafo 2.2.1.3).

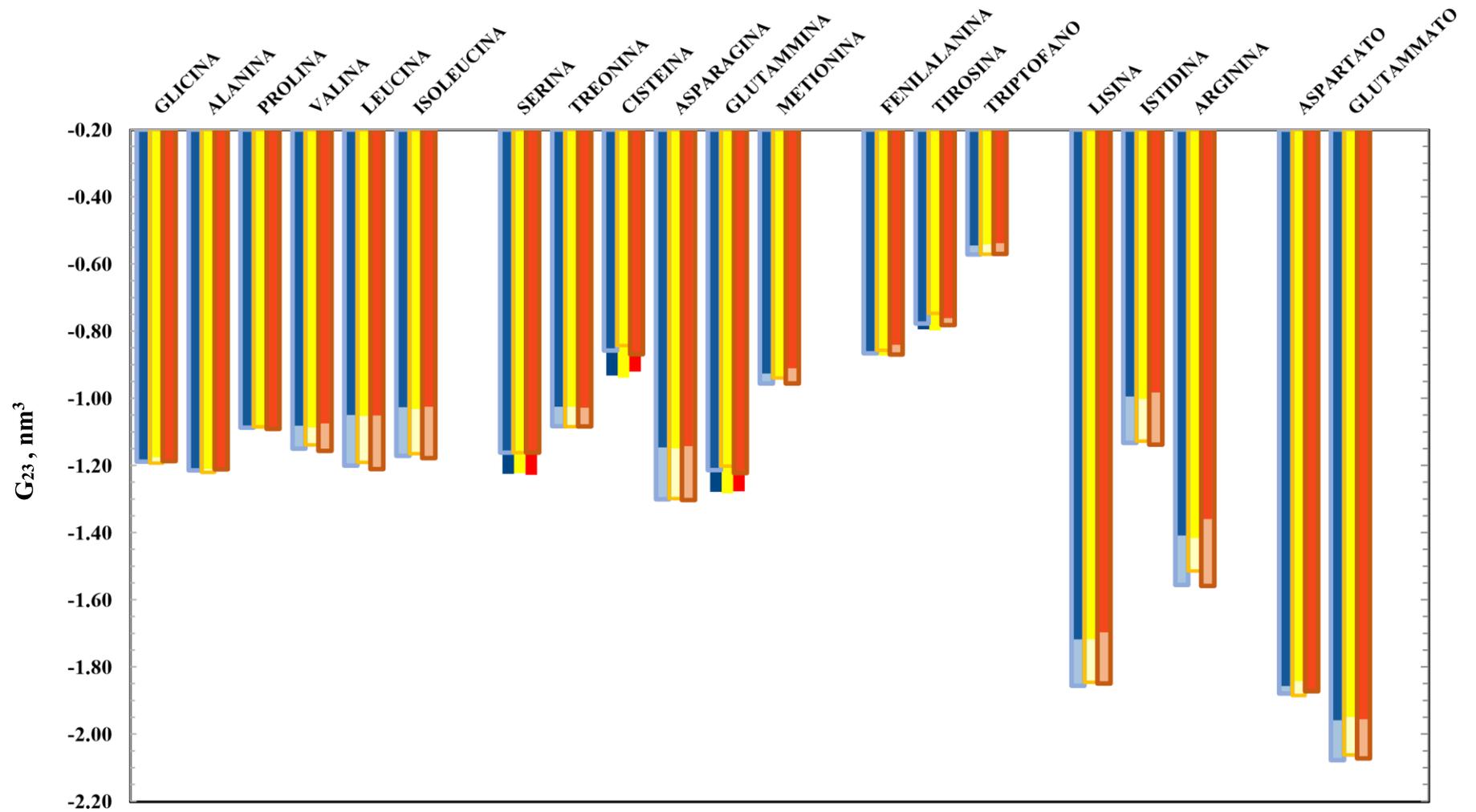
#### 3.2.1 Valutazione del *KBI* proteina – zucchero ( $G_{23}$ )

Come si può vedere in Figura 3.1, i risultati delle simulazioni confermano quanto previsto dalla teoria dell'esclusione preferenziale di Timasheff e Arakawa (paragrafo 1.4.3). Per tutti gli amminoacidi, lo zucchero risulta, all'equilibrio, escluso dalla superficie dell'amminoacido e delle sue parti (il valore del *KBI* per la proteina è infatti uguale alla media tra il valore per il backbone e il valore per la sidechain). Questo risultato indica che saccarosio e trealosio sono utilizzabili come protettori della struttura proteica durante la liofilizzazione. Si può vedere che

l'esclusione maggiore avviene per i residui carichi negativamente e positivamente, mentre quella minore avviene per i residui con catena laterale in cui siano presenti degli aromatici, con valore più basso in assoluto per il triptofano; questo può essere spiegato con il fatto che sia saccarosio che trealosio, isomeri, sono molecole non cariche, e per questo non presentano particolari interazioni con gruppi invece carichi.

Non si evidenziano particolari relazioni tra l'aumento del peso molecolare e il valore di KBI, mentre proprietà simili possono esser legate a omogeneità degli eteroatomi presenti nelle catene laterali: cisteina e metionina, ad esempio, hanno valori simili poiché entrambe possiedono un atomo di zolfo (S) nella sidechain. Stesso discorso può essere fatto per asparagina e glutammina, che possiedono entrambe un gruppo amminico e uno carbonilico, e per serina e treonina, che possiedono entrambe un gruppo alcolico. Tra gli amminoacidi carichi positivamente, invece, l'istidina interagisce maggiormente con saccarosio e trealosio, questo perché presenta sulla catena laterale una struttura ciclica che ne riduce la carica totale.

Come si può vedere in Figura 3.1, la differenza di valori ottenuti tra trealosio e saccarosio è modesta. In generale si può affermare che entrambi siano due ottimi stabilizzatori e che per entrambi l'esclusione maggiore sia con le molecole cariche, con cui non interagiscono vista la loro neutralità. Per tutti gli amminoacidi non polari, l'interazione con il saccarosio è minore rispetto a quella con il trealosio, così come per tutti gli amminoacidi carichi negativamente e positivamente. Per quanto riguarda i gruppi polari e aromatici non si può tracciare un comportamento generale, e le differenze tra i due zuccheri sono minime



**Figura 3.1** Rappresentazione del KBI zucchero-amminoacido. In blu il valore per l'intero residuo, in rosso il valore per il backbone, in giallo il valore per la catena laterale. In trasparenza e con bordo spesso i valori per l'interazione con il saccarosio, senza bordo e in colore pieno i valori per l'interazione con trealosio.

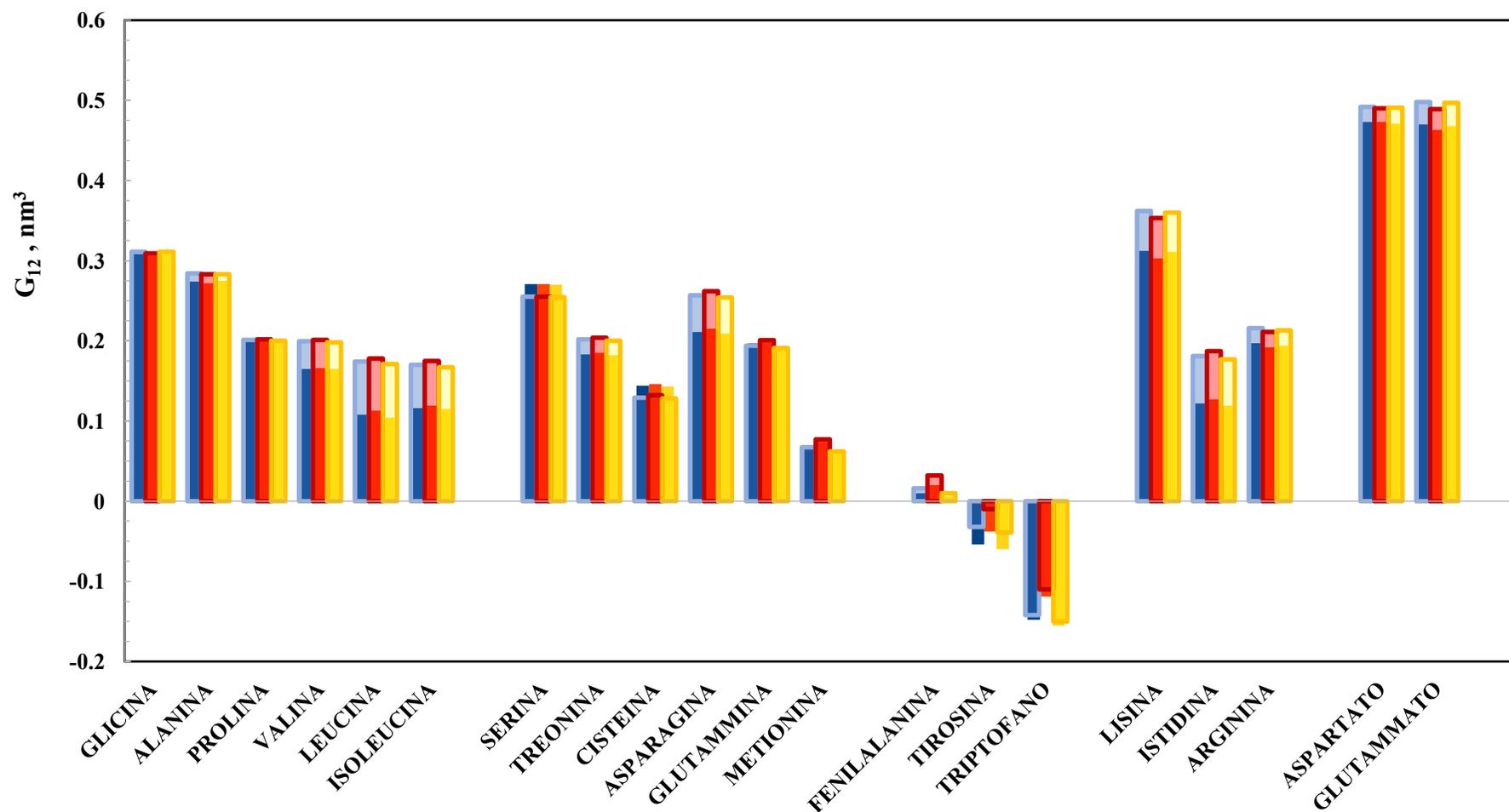
(tranne nel caso dell'asparagina). In generale si può quindi affermare, alla luce dei dati rilevati, che il saccarosio sia leggermente più escluso dalla superficie degli amminoacidi.

### 3.2.2 Valutazione del *KBI* proteina – acqua ( $G_{12}$ )

Come si può vedere in Figura 3.2, i risultati delle simulazioni indicano che per tutti gli amminoacidi, ad esclusione di tirosina e triptofano, l'acqua risulta inclusa rispetto alla superficie dell'amminoacido e delle sue componenti. Anche in questo caso il *KBI* della proteina risulta uguale alla media dei *KBI* di *backbone* e *sidechain*. Per ciò che riguarda tirosina e triptofano invece valgono le considerazioni fatte nel paragrafo precedente sulle peculiarità della presenza di aromatici nella *sidechain*. In questo caso l'acqua è esclusa dalla superficie degli amminoacidi ( $KBI < 0$ ).

Per le interazioni con solvente, nel caso degli amminoacidi *non polari*, si vede una progressiva diminuzione del valore di *KBI* al crescere del peso molecolare, così come accade per gli *aromatici* e per le coppie metionina e cisteina, glutammina e asparagina, serina e treonina, negli amminoacidi *polari*; questo può indicare il fatto che aumentando la superficie esposta della particella, a causa della presenza di catene laterali più lunghe o atomi più pesanti, è maggiore la possibilità che la molecola non si trovi circondata solamente da molecole di solvente, ma anche da altre molecole presenti nella soluzione. Le interazioni maggiori con l'acqua si possono osservare con i gruppi carichi *negativamente* e con la lisina.

Come si può vedere in Figura 3.2 inoltre, la differenza di valori ottenuti tra trealosio e saccarosio è di modesta entità. L'andamento è simile sia per il trealosio che per il saccarosio, anche se il solvente è maggiormente incluso in presenza di saccarosio per i gruppi *non polari*, *aromatici*, *carichi positivamente* e *negativamente*, al contrario di quanto accade per i gruppi *polari*, per i quali il solvente è maggiormente incluso in presenza di *trealosio* (tranne che per lisina e treonina). Partendo da queste osservazioni si può passare alla valutazione del *coefficiente di interazione preferenziale*, che nel confronto delle interazioni residuo-zucchero può restituire una migliore analisi qualitativa di quale protettore tra trealosio e saccarosio possa avere un miglior effetto nella stabilizzazione di una formulazione proteica.



**Figura 3.2** Rappresentazione del KBI con solvente per i singoli amminoacidi. In blu il valore per l'intero residuo, in rosso il valore per il backbone, in giallo il valore per la catena laterale. In trasparenza e con bordo spesso i valori per l'interazione con il saccarosio, senza bordo e in contrasto i valori per l'interazione con trealosio.

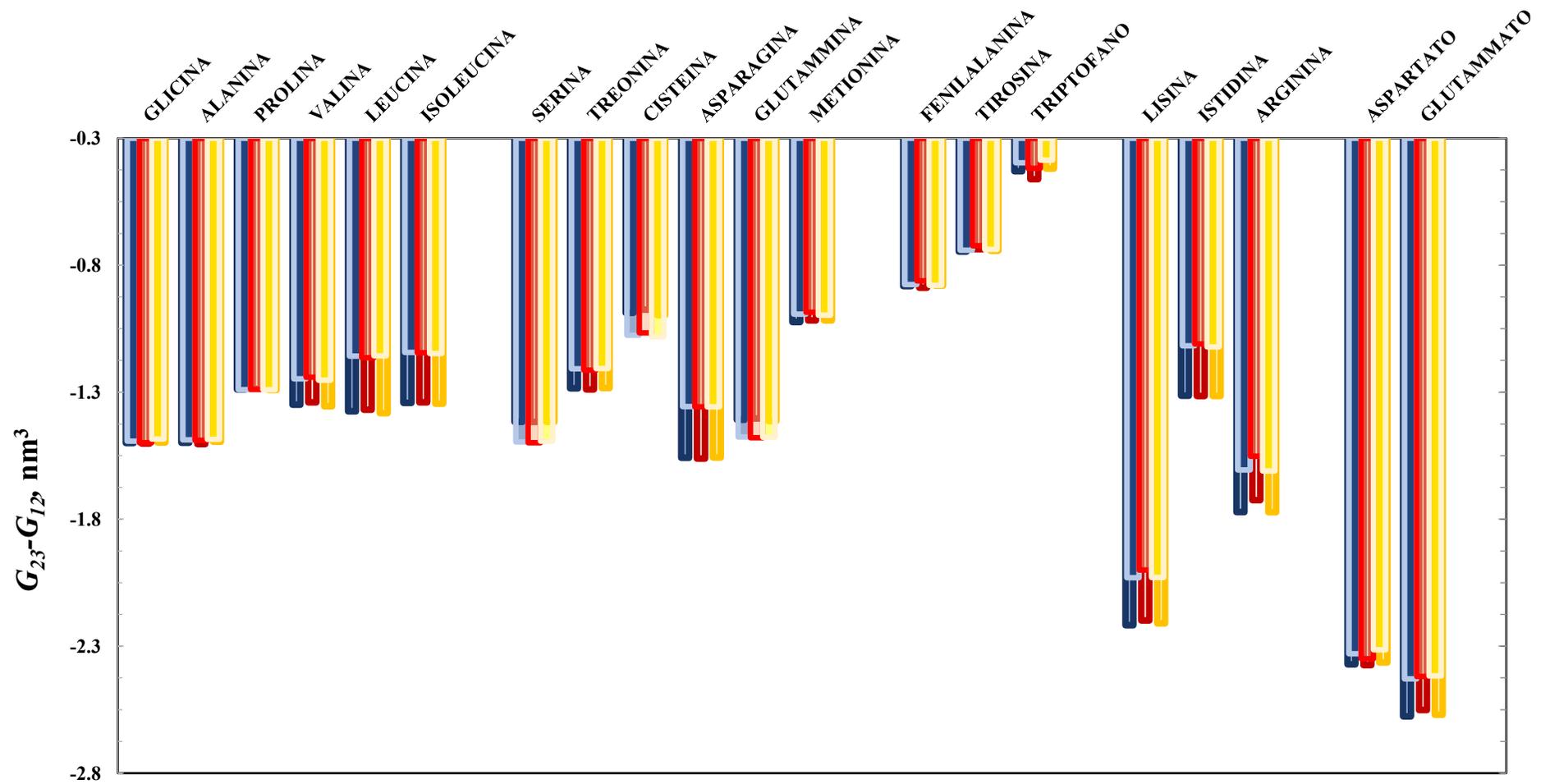
### 3.2.3 Valutazione del coefficiente di esclusione preferenziale

Il *coefficiente di interazione preferenziale*, come espresso nel paragrafo 2.3.5, è dato dalla differenza tra il *KBI* proteina-zucchero e il *KBI* proteina-solvente ( $G_{23}-G_{12}$ ), moltiplicato per la densità molecolare dello zucchero. È un parametro molto importante sia per la presente ricerca di additività nelle interazioni eccipiente-polipeptide che per considerazioni di tipo termodinamico sulla stabilizzazione garantita dai protettori nelle formulazioni proteiche. Il valore di questo coefficiente permette di stabilire, in casi come quelli del triptofano e della tirosina, per cui entrambi i *KBI* sono negativi, se sia più escluso il soluto o il solvente dalla superficie del residuo.

In Figura 3.3 è stato riportato il valore  $G_{23}-G_{12}$ , dal momento che questo rappresenta la parte variabile del *coefficiente di esclusione preferenziale*. Si può osservare che per gli amminoacidi *non polari e aromatici* il coefficiente di interazione preferenziale aumenta con l'aumentare del peso molecolare, risentendo dell'andamento decrescente del *KBI* tra proteina e solvente. Negli amminoacidi *carichi negativamente* invece, con l'aumentare del peso molecolare diminuisce il *coefficiente di interazione preferenziale*. Per gli amminoacidi *polari* si hanno valori simili per le coppie con catena laterale simile, come metionina e cisteina, asparagina e glutammina, serina e treonina, anche se l'andamento con il peso molecolare è diverso a seconda della coppia e dello zucchero. Per gli amminoacidi *carichi positivamente* non si può stabilire un andamento caratteristico, poiché troppo diversa è la conformazione della *sidechain*, che caratterizza in modo diverso le relazioni di ognuno dei tre residui.

Per tutti gli amminoacidi il *coefficiente di esclusione preferenziale* è negativo, sia per il trealosio che per il saccarosio, confermando quindi il fatto che in ogni soluzione lo zucchero utilizzato sia un buon protettore secondo la teoria del *soluto escluso*.

Come si può vedere in Figura 3.3 inoltre, la differenza dei valori di *coefficiente di esclusione preferenziale* tra trealosio e saccarosio non è molto accentuata. In generale l'interazione con il soluto in presenza di trealosio è maggiore rispetto a quella in presenza di saccarosio (tranne che per serina, cisteina e glutammina), e per tutti i gruppi tranne quelli *polari*, l'andamento con il peso molecolare è il medesimo. Alla luce di queste osservazioni si può quindi affermare che il saccarosio fornisca una protezione termodinamica leggermente superiore al trealosio nelle formulazioni contenenti amminoacidi.



**Figura 3.3** Rappresentazione del coefficiente di esclusione preferenziale per i singoli amminoacidi. In blu il valore per l'intero residuo, in rosso il valore per il backbone, in giallo il valore per la catena laterale. In trasparenza e con bordo chiaro i valori per l'interazione con il trealosio, con bordo scuro e in colore pieno i valori per l'interazione con il saccarosio.

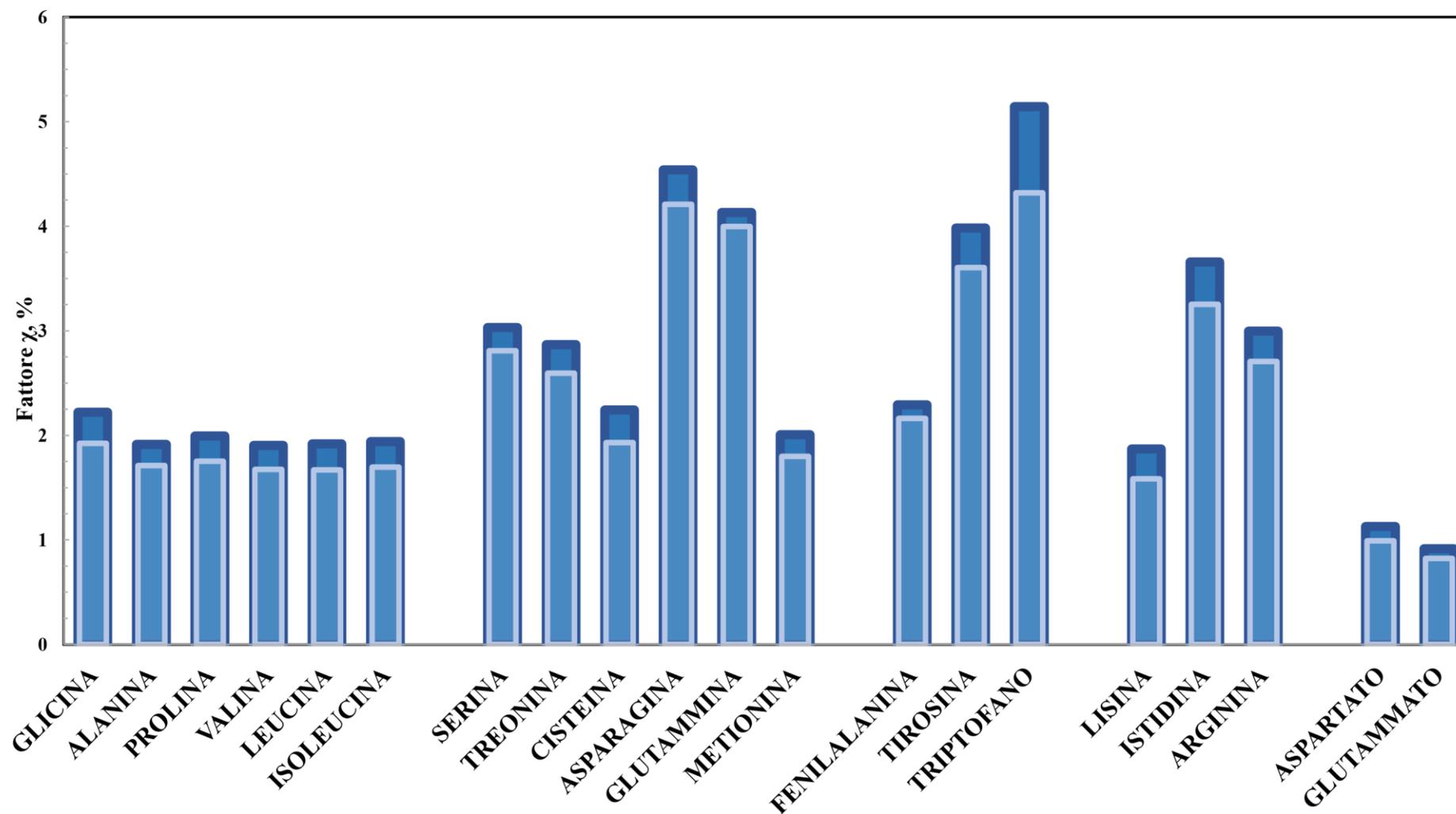
### 3.2.4 Valutazione del *water replacement* (fattore $\chi$ )

Come descritto nei paragrafi precedenti, la valutazione dei legami a idrogeno tra residuo e protettore e tra acqua e residuo è fondamentale nella verifica della funzione del protettore di anti-denaturante durante il processo di liofilizzazione. Infatti la perdita di acqua può comportare instabilità nella proteina, ed è fondamentale che il protettore riesca a sostituire le molecole d'acqua nel soddisfare la richiesta di legami idrogeno della catena peptidica. Questa teoria si chiama del *water replacement* (v. 1.4.2), e può essere studiata a livello simulativo con prove a diversa concentrazione del solvente, valutando i legami idrogeno tra la proteina e l'eccipiente e tra la proteina e il solvente (Pisano e Arsiccio, 2017). Seguendo la metodologia riportate in letteratura (Pisano e Arsiccio, 2017), nel presente lavoro si sono calcolati sia i legami a idrogeno tra protettore e residuo che quelli tra residuo e solvente, per poi normalizzare i valori ottenuti per ogni residuo ed avere un parametro confrontabile per tutti gli amminoacidi. Il parametro cui si fa riferimento viene indicato con la lettera  $\chi$ , ed è così espresso:

$$\chi = \frac{(n^{\circ} \text{ legami zucchero} - \text{residuo}) * 100}{n^{\circ} \text{ legami zucchero} - \text{residuo} + n^{\circ} \text{ legami solvente} - \text{residuo}} \quad (3.1)$$

In Figura 3.4 si può vedere che non ci sono grandi differenze quantitative all'interno dei vari gruppi. L'andamento a coppie per affinità di *sidechain* è confermato per il *gruppo polare*, gli amminoacidi *non polari* hanno valori molto simili, con andamenti che, come per tutti i gruppi analizzati, non hanno relazioni particolari con il peso molecolare, tranne che per gli *aromatici*, per cui il numero di legami con lo zucchero in proporzione è molto maggiore all'aumentare della lunghezza della *sidechain*. Si può quindi ipotizzare per quest'ultimo caso che maggiore sia l'estensione della catena laterale, maggiore sia la possibilità che l'amminoacido entri in contatto con il protettore. Si nota anche che gli amminoacidi che possiedono atomi di azoto nella catena laterale, come asparagina e glutammina, formano un numero elevato di legami a idrogeno. Lo stesso accade anche per gli amminoacidi *aromatici* che possiedono atomi di azoto nella *sidechain*, quali triptofano e tirosina. Tra gli amminoacidi *carichi* si osserva interazione maggiore per quelli *carichi positivamente*.

Come si può vedere in Figura 3.4, la differenza dei valori di  $\chi$  tra trealosio e saccarosio è modesta. Entrambi seguono lo stesso andamento per tutti gli amminoacidi, ma il trealosio forma sempre un numero leggermente più alto di legami rispetto al saccarosio. Si può quindi ipotizzare che il trealosio sia un protettore leggermente migliore per quanto riguarda la teoria del *water replacement*. Ciò è legato alla conformazione della molecola del trealosio, che mostra una più elevata esposizione superficiale dei siti per la formazione di legami a idrogeno rispetto a quella del saccarosio. Queste ipotesi sono confermate anche in letteratura (Pisano e Arsiccio, 2017).



**Figura 3.4** Rappresentazione del fattore  $\chi$  per i singoli amminoacidi. In blu scuro e contrasto i valori per i residui in presenza di trealosio, in azzurro e trasparenza i valori per i residui in presenza di saccarosio.

Analizzando il parametro  $\chi$  si evince anche che il numero di legami formati tra i residui e il solvente è circa due ordini di grandezza superiore rispetto al numero di legami formati con lo zucchero, in accordo con la differenza tra il numero di molecole di acqua e di protettore presenti nella box (più di 100000 molecole di acqua, rapportate a 308 molecole di trealosio o saccarosio).

Per quanto quindi dalla superficie delle proteine i protettori risultino esclusi, si forma già in soluzione acquosa un certo numero di legami a idrogeno. Nel momento in cui si voglia fare una analisi più dettagliata sulla verifica della teoria del *water replacement*, si dovrebbe gradualmente aumentare la concentrazione di eccipiente. È stato mostrato che quando questo accade il numero di legami inizialmente formati con il solvente è sostituito parzialmente da quelli formati con il protettore, e ciò permette che il farmaco proteico non si denaturi durante il passaggio allo stato solido (Pisano e Arsiccio, 2017).

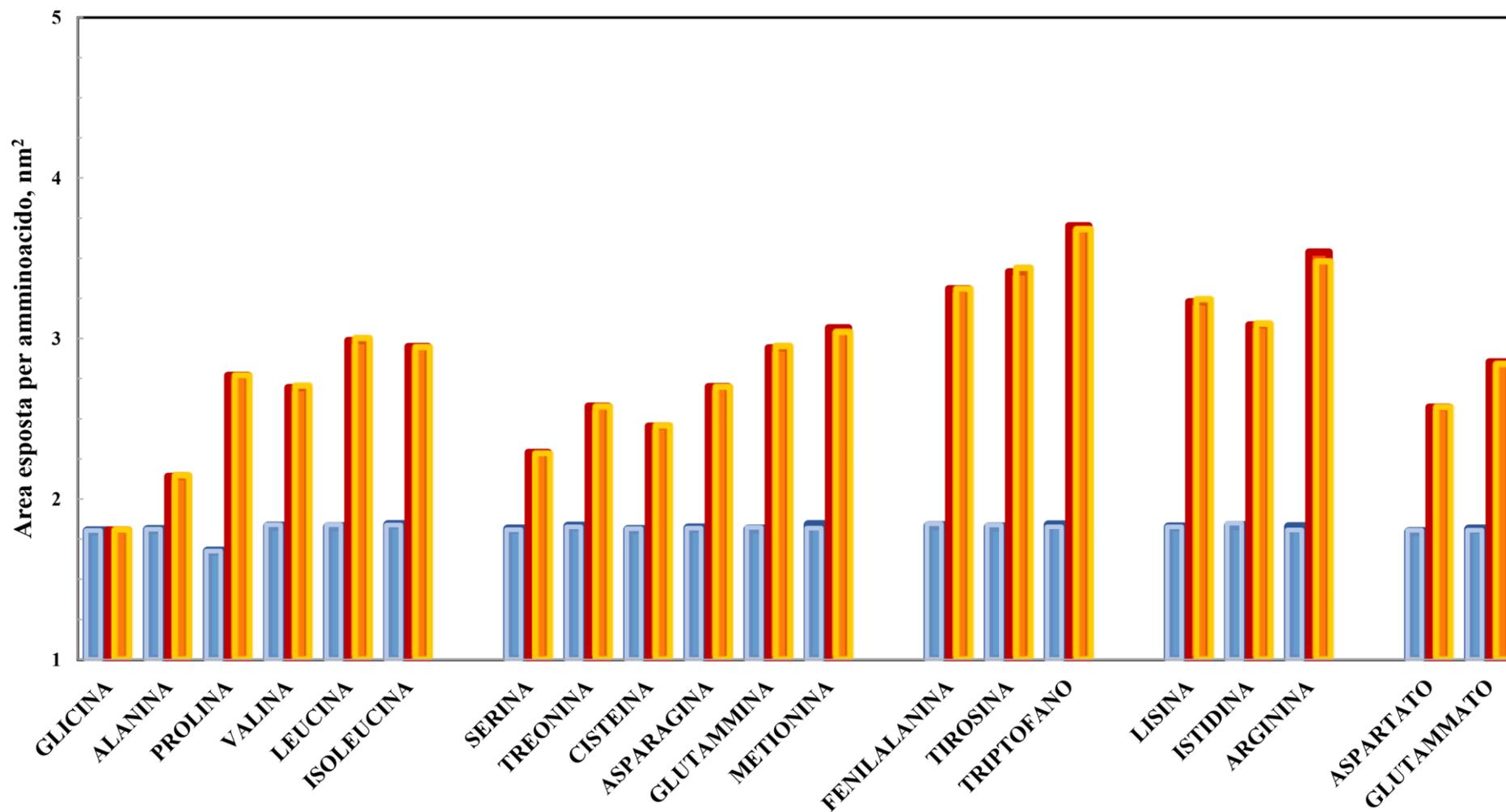
### 3.2.5 Valutazione delle aree esposte con *gmx sasa*

La valutazione delle aree esposte tramite il tool di GROMACS *sasa* permette di conoscere per ogni residuo la superficie esposta al solvente all'interno della soluzione. Questo tool, che segue il metodo di calcolo teorizzato da Eisenhaber (Eisenhaber *et al.*, 1995) è stato utilizzato solo nella presente analisi, mentre per la valutazione dell'additività si è preferito utilizzare l'algoritmo di Lee e Richards (v. paragrafo 2.4.3).

Come si può vedere in Figura 3.5, il valore di area esposta per il backbone è molto simile per ogni residuo, e ciò indica che, pur nella diversità della composizione delle sidechain, il solvente entra in contatto con il corpo centrale dell'amminoacido circa nella stessa misura. Per le sidechain invece l'area esposta varia per ogni residuo. All'interno di ogni gruppo si può vedere un progressivo aumento di superficie esposta all'aumentare del peso molecolare, dovuto alla maggior ampiezza della sequenza di atomi che compone la sidechain. Il valore più ampio di tutti è per il triptofano, poichè questo presenta due strutture cicliche nella sua sidechain.

In generale si può quindi affermare che, nel caso delle aree esposte, l'andamento dei risultati è influenzato principalmente dall'ampiezza della catena laterale, mentre la distribuzione spaziale delle stesse non è così importante da indurre grandi differenze a livello di esposizione del residuo.

Come si può vedere in Figura 3.5, la differenza dei valori di superficie esposta in presenza di trealosio o saccarosio è molto modesta. Si osserva infatti lo stesso andamento per tutti gli amminoacidi. Tuttavia in presenza di saccarosio, tranne che per lisina, istidina, fenilalanina, tirosina e glutammina, le aree esposte per la sidechain e il backbone sono sempre leggermente inferiori. Questo può indicare che in presenza di saccarosio, gli amminoacidi assumono in genere una forma leggermente più compatta. Ciò è in linea con la maggiore esclusione del saccarosio dalla superficie degli amminoacidi precedentemente discussa.



**Figura 3.5** Rappresentazione delle aree esposte per i singoli amminoacidi. In blu il valore per il backbone in presenza di trealosio, in arancione il valore per la sidechain in presenza di trealosio, in azzurro con bordo spesso il valore per il backbone in presenza di saccarosio, in arancione con bordo spesso il valore della sidechain in presenza di saccarosio.

### 3.2.6 Valutazione del clustering

La valutazione del numero di cluster formati è stata fatta a seguito delle altre valutazioni, per verificare se, laddove ci fosse poca interazione sia tra residui e zucchero che tra residui e solvente, non si fossero formati aggregati di peptidi. Ricordando (v. paragrafo 1.4) che l'aggregazione proteica è un fenomeno di denaturazione, questo parametro ha una importanza rilevante nel caso della valutazione della stabilizzazione proteica da parte di crio/lioprotettori.

Si può vedere in Figura 3.6 che il numero di cluster è sempre superiore a venti. Ricordando che gli amminoacidi all'interno di ogni box di simulazione sono 25, si può quindi concludere che i residui non hanno dato origine a macro-aggregati, ma sono di fatto rimasti perlopiù in forma monomeric. Non ci sono relazioni tra l'aumento del peso molecolare e il numero di cluster, mentre in generale si può dire che gli amminoacidi carichi tendano a formare aggregati in misura leggermente maggiore, a causa probabilmente delle interazioni elettrostatiche tra le sidechain ed i terminali zwitterionici.

Come si può vedere in Figura 3.6, la differenza dei valori di clustering in presenza di trealosio o saccarosio è minima. Solo per il gruppo non polare l'aggregazione è minore in presenza di saccarosio rispetto al trealosio, mentre per tutti gli altri gruppi non si può stabilire un comportamento caratteristico in presenza di un protettore rispetto ad un altro. Per entrambi gli zuccheri sono gli amminoacidi carichi a presentare i valori di clustering maggiori.

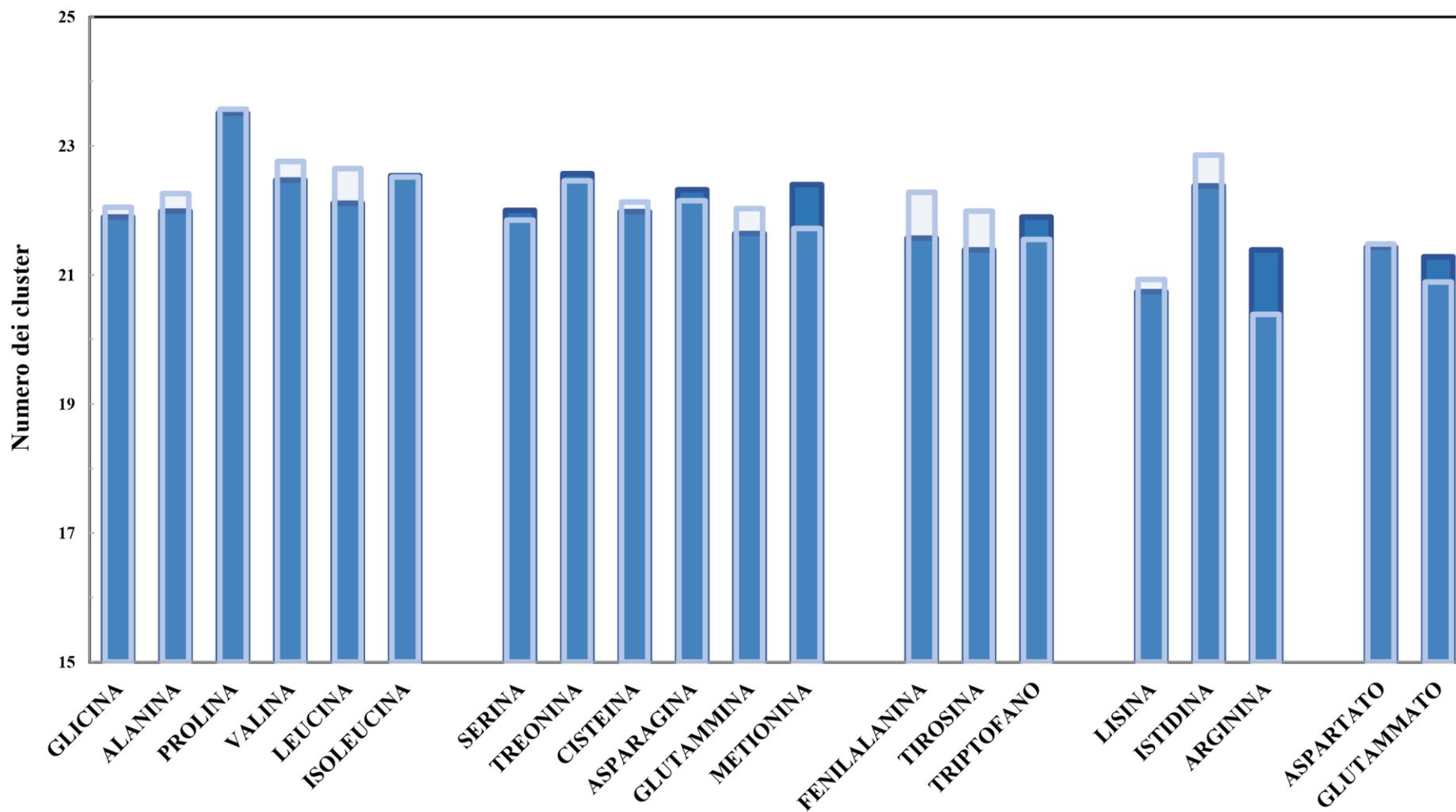


Figura 3.6 Rappresentazione del clustering per i singoli amminoacidi. In blu scuro e contrasto i valori per i residui in presenza di trealosio, in azzurro e trasparenza i valori per i residui in presenza di saccarosio.

### 3.3 Verifica dell'additività

In questa sezione si descrivono i risultati ottenuti nella verifica dell'additività per il *coefficiente di esclusione preferenziale*  $\Gamma_{23}^I$  (**inteso come la differenza  $G_{23} - G_{12}$ , ed espresso in  $\text{nm}^3$** ) secondo i metodi descritti nei paragrafi 2.4 e seguenti. Questa ipotesi, se verificata, permetterebbe di venire immediatamente a conoscenza del tipo di interazione tra un polipeptide e un eccipiente semplicemente conoscendone la struttura e la sequenza primaria. Per passare da  $\Gamma_{23}^I$  a  $\Gamma_{32}$  è necessario moltiplicare  $\Gamma_{23}^I$  per la densità molecolare dell'eccipiente.

#### 3.3.1 Descrizione delle analisi

Si è scelto di effettuare l'analisi di additività per i seguenti otto sistemi:

1. GB1-hairpin nativo in saccarosio 1M
2. GB1-hairpin nativo in trealosio 1M
3. GB1-hairpin denaturato in saccarosio 1M
4. GB1-hairpin denaturato in trealosio 1M
5. Lisozima (codice PDB: 1e81) nativo in trealosio 1M
6.  $\alpha$  – chimotripsinogeno ( $\alpha$  – CGN, codice PDB: 2cga) nativo in saccarosio 1M
7. Ribonuclease A (RNase A, codice PDB: 1kf5) nativo in trealosio 1M
8. RNase A (codice PDB: 1kf5) nativo in saccarosio 1M

Per i sistemi 5-8 si sono usati i valori di  $\Gamma_{32}$  presenti in letteratura (Cloutier *et al.*, 2018) e convertiti in  $\text{nm}^3$  (v. paragrafo 3.3).

I dati utilizzati per questi sistemi sono pertanto frutto di simulazioni leggermente diverse da quelle effettuate in questo lavoro per i singoli amminoacidi e per il GB1-hairpin. Infatti sono stati ottenuti con una versione di GROMACS antecedente (v. 5.0.5) e sono l'esito di simulazioni leggermente più brevi (55 ns). Queste differenze, però, non sono tali da portare a risultati non confrontabili con quelli ottenuti con il metodo di simulazione utilizzato in questo lavoro. Ipotizzando che in presenza di protettori come trealosio e saccarosio una proteina mantenga configurazione uguale a quella del proprio stato nativo (v. paragrafo 1.4.3), l'area esposta è stata valutata direttamente sui file *pdb* delle proteine coinvolte nei sistemi 5-8. Ciò è stato ottenuto tramite l'algoritmo di Lee e Richards (Lee e Richards, 1971), lo stesso utilizzato per valutare le aree esposte dai residui componenti il GB1 e per l'area esposta dall'amminoacido X nel tripeptide Gly – X – Gly (si veda paragrafo 2.4.2). Per tutte e otto le simulazioni si è dunque utilizzato nella sua interezza il metodo di valutazione proposto da Auton e Bolen (Auton e Bolen, 2005), così da valutare se anche in ambito simulativo i risultati fossero concordi con quelli sperimentali.

Prima di esporre i risultati è necessario verificare dal punto di vista termodinamico i passaggi che hanno portato a ottenere l'equazione 2.33:

$$\Gamma_{32} = \Gamma_{32}^{bb} * \sum n_i \alpha_i^{bb} + \sum n_i \alpha_i^{sc} \Gamma_{i,32}^{sc} \quad (2.33)$$

Questa equazione è valida se e solo se la differenza tra il *KBI* eccipiente-eccipiente ( $G_{33}$ ) e il *KBI* eccipiente-acqua ( $G_{13}$ ) si mantiene pressochè costante ( $G_{33}-G_{13} \approx \text{cost}$ ). Infatti vale, per i  $\Delta G_{trasf}$ :

$$\Delta G_{trasf} = \sum_i n_i \alpha_i^{sc} \Delta G_{trasf}^{sc} + \Delta G_{trasf}^{bb} \sum_i n_i \alpha_i^{bb} \quad (2.23)$$

Dove il  $\Delta G_{trasf}$  è pari a:

$$\Delta G_{trasf} = \frac{k_b T M_1 N_{Av} c_1 (G_{21} - G_{23})}{1 - M_1 N_{Av} c_1 m_3 (G_{13} - G_{33})} \quad (3.2)$$

Considerato che il coefficiente di interazione preferenziale  $\Gamma_{32}$  può essere espresso come:

$$\Gamma_{32} = \rho_3 (G_{23} - G_{21}) = M_1 N_{Av} c_1 m_3 (G_{23} - G_{21}) \quad (2.14)$$

è evidente che la sostituzione di  $\Gamma_{32}$  al posto del  $\Delta G_{trasf}$  nell'equazione 2.23 per ottenere l'equazione 2.33 è possibile solo se  $G_{33}-G_{13} \approx \text{cost.}$ , poiché per tutti i membri di questa equazione si otterrebbe:

$$\Delta G_{trasf} = \frac{C_1 * (G_{23} - G_{21})}{C_2} \quad (3.3)$$

con  $C_1$  e  $C_2$  che indicano due valori costanti, e che, essendo uguali per tutti i termini dell'equazione 2.23, permettono di passare da questa equazione all'equazione 2.33. Si ritiene pertanto necessario verificare che il termine  $G_{13}-G_{33}$  sia uguale per ambo i membri dell'equazione 2.23, ossia sia per il polipeptide che per i singoli amminoacidi. La verifica si traduce quindi nel confrontare le differenze tra  $G_{13}$  e  $G_{33}$  per tutti i singoli amminoacidi e per i polipeptidi analizzati. Nel momento in cui non si verificano grosse differenze tra i valori misurati, si può considerare valida l'equazione 2.33. Si riportano in Figure 3.7 e 3.8 i valori ottenuti per trealosio e saccarosio per ognuno dei venti amminoacidi, per il GB1 hairpin e per i vari composti di N-acetilglicinammide. Come si può vedere le differenze sono pressochè costanti e le variazioni minime rispetto, a esempio, a quelle rilevate per  $G_{23}$  e  $G_{12}$ . Si può quindi considerare valida l'ipotesi di additività anche per i *coefficienti di interazione preferenziale*  $\Gamma_{23}$  o  $\Gamma_{23}^I$ . Si può inoltre notare che sia  $G_{13}$  che  $G_{33}$  sono negativi, dimostrando quindi che non vi sono fenomeni di clustering tra molecole di zucchero né fenomeni di interazione tra lo zucchero e il solvente. In ogni caso, sia per il trealosio che per il saccarosio, l'esclusione tra zuccheri è molto maggiore rispetto all'esclusione zucchero-solvente.

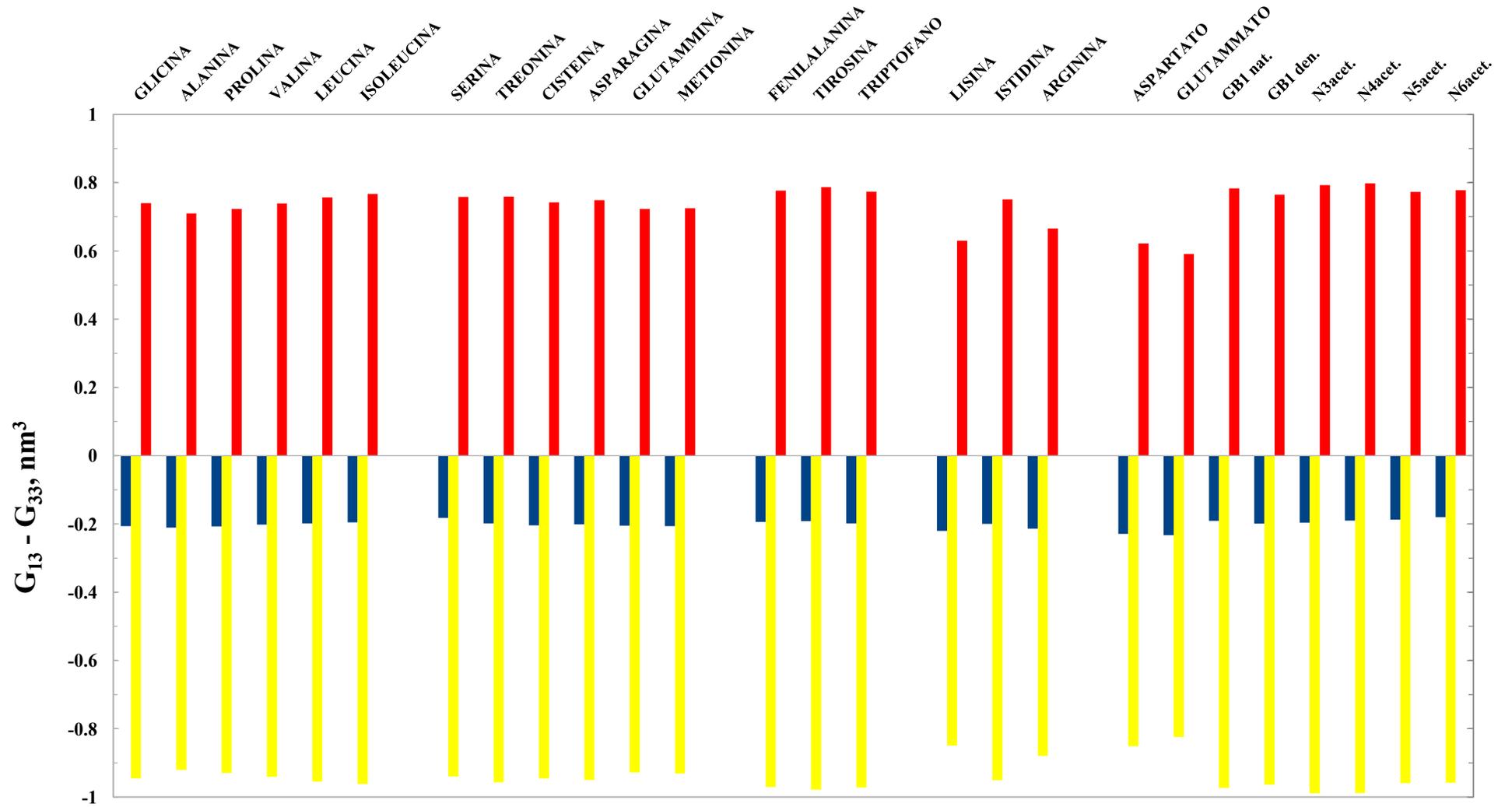
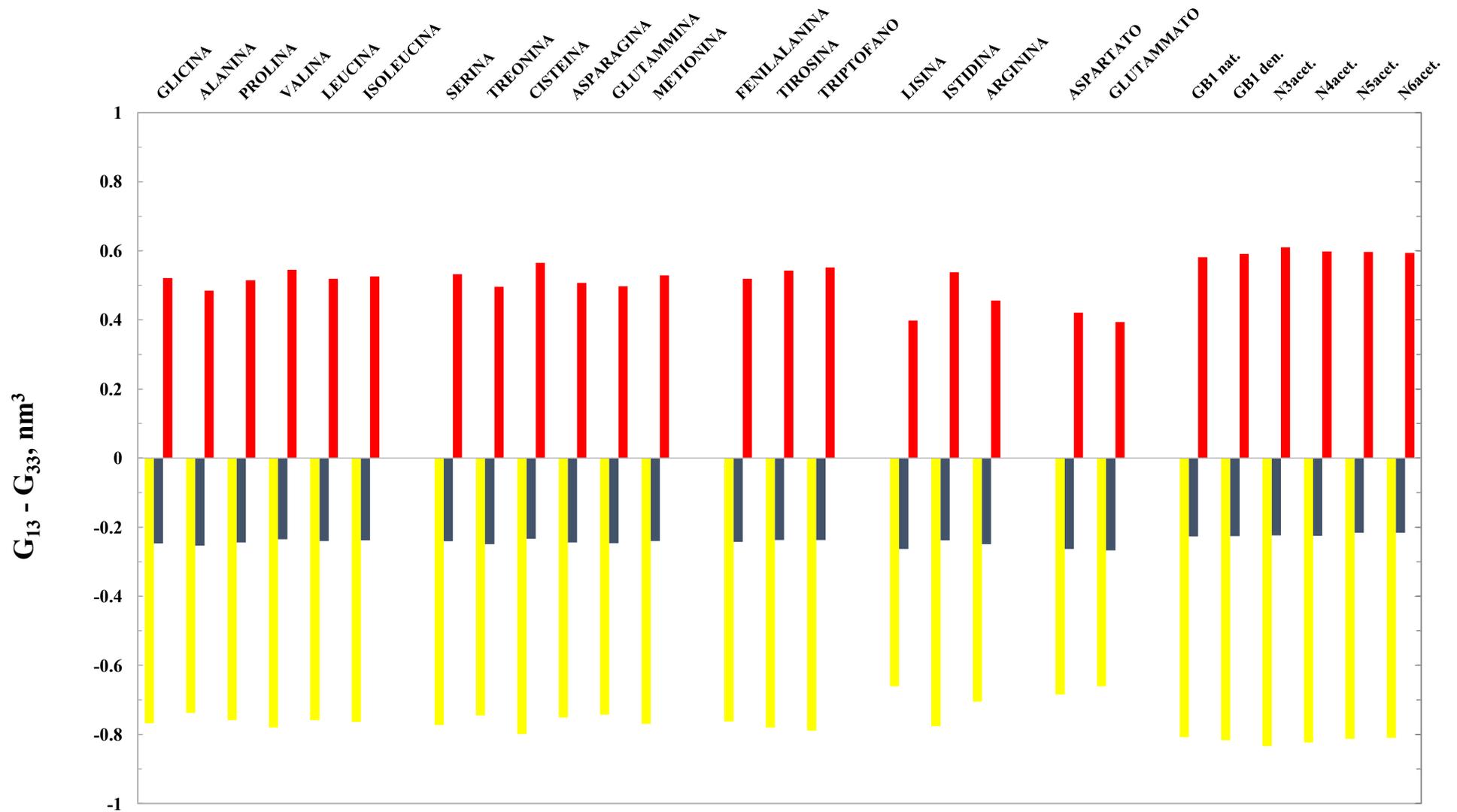


Figura 3.7 Differenza  $G_{13}-G_{33}$  in presenza di trealosio. In blu scuro  $G_{13}$ , in giallo  $G_{33}$ , in rosso la differenza tra i due valori.



**Figura 3.8** Differenza  $G_{13}-G_{33}$  in presenza di saccarosio. In blu scuro  $G_{13}$ , in giallo  $G_{33}$ , in rosso la differenza tra i due valori.

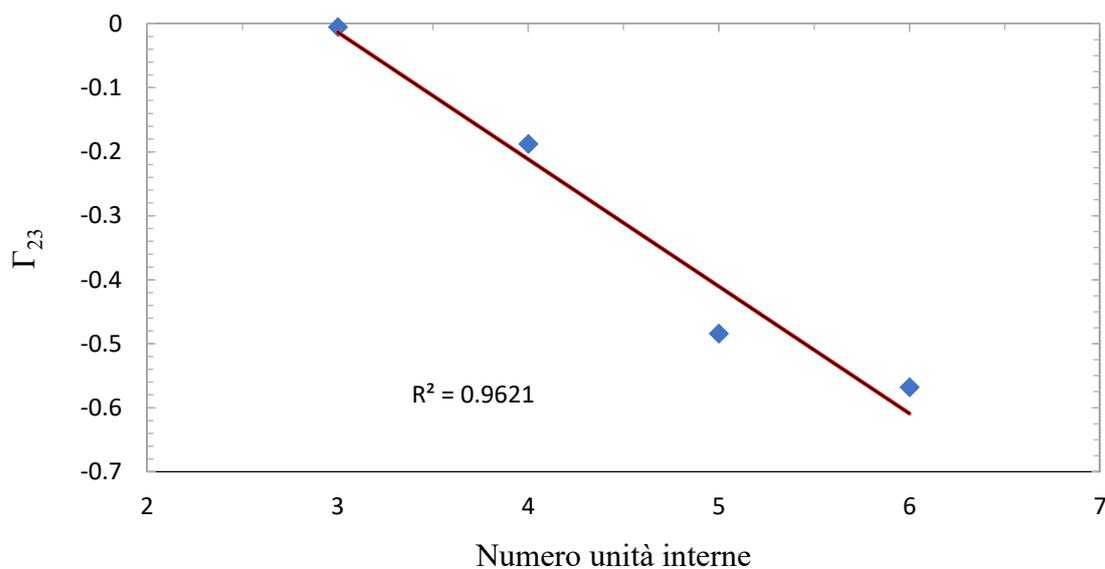
Alla luce dei risultati illustrati in questo capitolo, si riporta qui un raffronto tra i risultati ottenuti sperimentalmente per il  $\Delta G_{trasf}$  delle sidechain (Auton *et al.*, 2008) e quelli ottenuti partendo dalle simulazioni effettuate. Il confronto può essere fatto solamente per 17 amminoacidi, poiché solo per questi sono disponibili i risultati sperimentali di  $\Delta G_{trasf}$  (non sono disponibili per la cisteina, per cui si usa lo stesso valore della metionina, mentre per tirosina e triptofano sono disponibili solo in presenza di saccarosio 1M, e si utilizzano i medesimi per le interazioni con trealosio). I risultati qui presentati mostrano che solamente per gli amminoacidi serina, istidina e prolina è presente una buona approssimazione computazionale rispetto a quanto calcolato sperimentalmente, mentre gli altri amminoacidi presentano valori molto distanti, anche qualitativamente. Quanto qui riportato verrà utilizzato nel paragrafo 3.4 per considerazioni finali sulla bontà del campo di forza attualmente utilizzato nelle simulazioni di dinamica molecolare.

**Tabella 3.1** Valori del  $\Delta G_{trasf}$  delle sidechain in presenza di trealosio e saccarosio

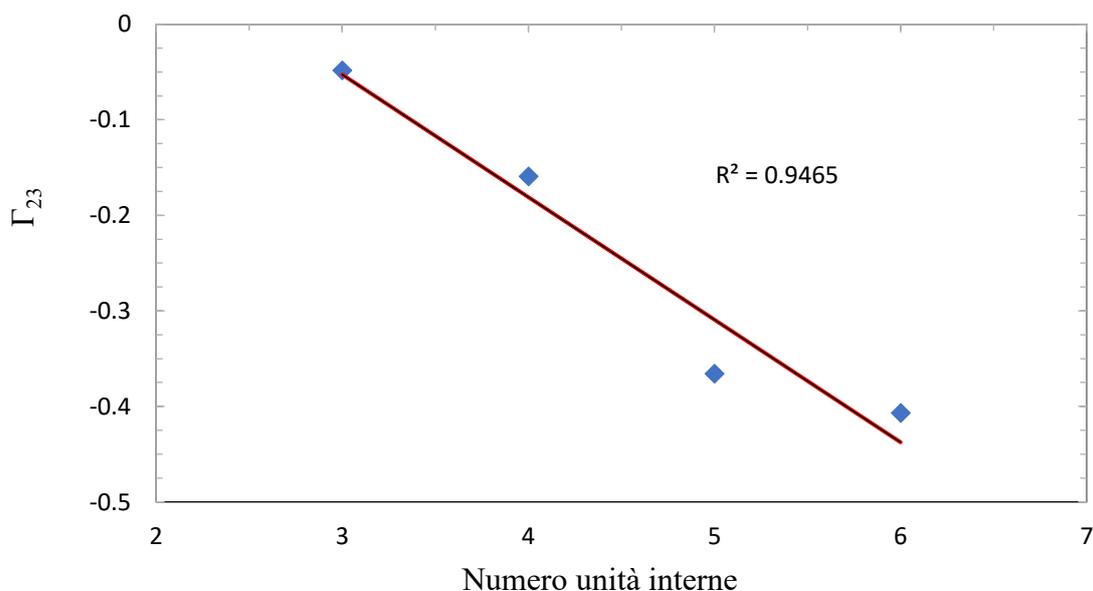
	SACCAROSIO		TREALOSIO	
	$\Delta G_{trasf}$ , <i>calc.</i> , cal/mol	$\Delta G_{trasf}$ , <i>sperim.</i> , cal/mol	$\Delta G_{trasf}$ , <i>calc.</i> , cal/mol	$\Delta G_{trasf}$ , <i>sperim.</i> , cal/mol
<b>ALA</b>	-24.37	21.96	3.54	33.25
<b>ARG</b>	91.65	-79.41	40.60	-50.33
<b>ASN</b>	20.31	-28.37	-34.67	48.67
<b>ASP</b>	353.20	-41.75	236.30	-85.92
<b>GLN</b>	-62.90	-40.97	-1.47	-36.34
<b>GLU</b>	424.90	-37.27	270.66	-96.54
<b>GLY</b>	0.00	0.00	0.00	0.00
<b>HYS</b>	-86.31	-118.76	-93.88	-98.75
<b>ILE</b>	-79.03	28.03	-89.35	79.66
<b>LEU</b>	-66.20	37.01	-84.52	96.18
<b>LYS</b>	260.81	-39.70	156.63	-50.07
<b>MET</b>	-243.95	-6.76	-121.67	29.19
<b>PHE</b>	-321.93	-96.45	-155.35	-17.88
<b>PRO</b>	-113.15	-73.12	-47.68	-94.67
<b>SER</b>	-35.95	-2.88	-1.88	-0.98
<b>THR</b>	-126.05	20.72	-72.52	26.32
<b>VAL</b>	-63.44	33.82	-60.43	96.79

### 3.3.2 Valutazione del coefficiente di esclusione preferenziale per il backbone

Come si può osservare in Figure 3.9 e 3.10, In entrambi i casi i valori sono risultati negativi, in accordo con gli esperimenti di Auton e Bolen (Auton e Bolen, 2005), in cui il  $\Delta G_{trasf}$  per il backbone, avente segno opposto al  $\Gamma_{23}$ , risulta positivo. Quantitativamente i valori di  $\Delta G_{trasf}$  del backbone ottenuti tramite GROMACS, calcolati come descritto nel paragrafo precedente, sono in discreto accordo con quelli sperimentali. Per il saccarosio si ha un valore sperimentale pari a 50 cal/mol, mentre nel presente lavoro risulta pari a 72 cal/mol; per il trealosio invece il valore sperimentale è pari a 47 cal/mol, mentre nel presente lavoro risulta pari a 48 cal/mol.



**Figura 3.9** Retta interpolante (in rosso) per la N-acetilglicinamide con trealosio. Equazione della retta:  $y = -0.198x + 0.1853$ . Valore di coefficiente di esclusione preferenziale  $\Gamma_{23}^I$  per il backbone =  $-0.198 \text{ nm}^3$ . In azzurro i valori di gamma per la serie di N-acetilglicinamide.



**Figura 3.10** Retta interpolante (in rosso) per la N-acetilglicinamide con saccarosio. Equazione della retta:  $y = -0.128x + 0.0757$ . Valore di coefficiente di esclusione preferenziale  $\Gamma_{23}^I$  per il backbone =  $-0.128 \text{ nm}^3$ . In azzurro i valori di gamma per la serie di N-acetilglicinamide.

### 3.3.3 Valutazione del coefficiente di esclusione preferenziale della sidechain

Nella valutazione del coefficiente di esclusione preferenziale della *sidechain*, come descritto nel paragrafo 2.4.3, si sono presi i valori di *coefficiente di esclusione preferenziale* per gli interi residui e ad ognuno di essi è stato sottratto il valore di *coefficiente di esclusione preferenziale* per la glicina. Come si è già visto per i risultati sperimentali riportati in letteratura, utilizzando questo metodo, i contributi per la *sidechain* dei singoli amminoacidi possono essere sia positivi che negativi. In Tabella 3.1 si riportano i risultati ottenuti per saccarosio e trealosio. È stato dimostrato che i valori ottenuti per gli amminoacidi in forma zwitterionica, utilizzati nel presente lavoro, sono molto simili a quelli ottenuti per gli amminoacidi in forma capped, ossia dotati di terminale acetilico e amminico, e non aventi cariche sul backbone. Si può quindi escludere fin da ora che eventuali differenze dei valori ottenuti per il *coefficiente di esclusione preferenziale* per la sidechain rispetto ai dati sperimentali siano dovute all'utilizzo della forma *carica* degli amminoacidi nelle nostre simulazioni.  $\Gamma_{32}^{RES}$  indica il *coefficiente di esclusione preferenziale*  $\Gamma_{23}^I$  per il residuo,  $\Gamma_{32}^{GLY}$  quello per la glicina,  $\Gamma_{32}^{SC}$  quello finale ottenuto per la sidechain.

Si può vedere, innanzitutto, che  $\Gamma_{32}^{GLY}$  presenta valori molto simili sia in presenza di saccarosio che di trealosio. I valori di  $\Gamma_{32}^{SC}$  più negativi si hanno per catene laterali cariche, confermando quindi che gli amminoacidi carichi, a causa di fenomeni di natura elettrostatica, interagiscono maggiormente con l'acqua (Figura 3.2) rispetto alla glicina. Gli amminoacidi non carichi presentano valori di  $\Gamma_{32}^{RES}$  molto vicini a quelli della glicina, risultando in  $\Gamma_{32}^{SC}$  tutti di segno positivo, e di maggior valore assoluto in presenza di trealosio. Gli amminoacidi polari presentano valori di  $\Gamma_{32}^{SC}$  positivi e seguono l'andamento a coppie descritto nei paragrafi 3.2.1 e seguenti, con valori molto simili in presenza di saccarosio o trealosio; uniche eccezioni sono i valori negativi di asparagina in presenza di saccarosio e serina in presenza di trealosio. Valori positivi di  $\Gamma_{32}^{SC}$  si riscontrano anche per gli aromatici, con simili risultati sia in presenza di saccarosio che di trealosio. Tutti gli amminoacidi con catena non carica quindi, sia in presenza di saccarosio che di trealosio, confrontati con la glicina, esprimono, a livello di additività, un contributo di catena laterale che tende a interagire preferenzialmente con lo zucchero ( $\Gamma_{32}^{SC}$  positivo).

**Tabella 3.2** Valori del coefficiente di esclusione preferenziale per la sidechain

	SACCAROSIO			TREALOSIO		
	$\Gamma_{32}^{RES}, nm^3$	$\Gamma_{32}^{GLY}, nm^3$	$\Gamma_{32}^{SC}, nm^3$	$\Gamma_{32}^{RES}, nm^3$	$\Gamma_{32}^{GLY}, nm^3$	$\Gamma_{32}^{SC}, nm^3$
ALANINA	-1.499		0.001	-1.488		0.004
ARGININA	-1.772		-0.272	-1.606		-0.114
ASPARAGINA	-1.558		-0.058	-1.357		0.135
ASPARTATO	-2.371		-0.871	-2.33		-0.838
CISTEINA	-0.987		0.513	-1.076		0.416
GLUTAMMINA	-1.408		0.092	-1.476		0.016
GLUTAMMATO	-2.576		-1.076	-2.429		-0.937
GLICINA	-1.5		0	-1.492		0
ITSTIDINA	-1.314		0.186	-1.117		0.375
ISOLEUCINA	-1.342	-1.5	0.158	-1.143	-1.492	0.349
FENILALANINA	-0.882		0.618	-0.876		0.616
LEUCINA	-1.375		0.125	-1.158		0.334
LISINA	-2.218		-0.718	-2.03		-0.538
METIONINA	-1.023		0.477	-0.933		0.599
PROLINA	-1.289		0.211	-1.29		0.202
SERINA	-1.417		0.083	-1.496		-0.004
TIROSINA	-0.746		0.754	-0.741		0.751
TREONINA	-1.285		0.215	-1.208		0.284
TRIPTOFANO	-0.43		1.07	-0.397		1.095
VALINA	-1.349		0.151	-1.247		0.245

### 3.3.4 Valutazione delle aree esposte nei tripeptidi Gly – X – Gly

Come già descritto nel paragrafo 2.4.2, per normalizzare il contributo dei singoli residui nei polipeptidi si rende necessario, sia per il backbone che per la sidechain, dividere l'area esposta dalla sidechain o dal backbone del residuo per i valori che lo stesso residuo X espone nel tripeptide *Gly – X – Gly*, contenente alle estremità due molecole di glicina. I residui sono stati creati con pymol e le aree computate con l'algoritmo di Lee e Richards, lo stesso utilizzato per la valutazione delle aree esposte nei polipeptidi. Si riportano in Tabella 3.3 i risultati ottenuti. Lo stesso valore è stato riscontrato per il *backbone* di ogni residuo, mentre i risultati per le *sidechain* sono diversi per ogni amminoacido. L'andamento è simile a quello calcolato con l'algoritmo di Eisenhaber et al (Eisenhaber *et al.*, 1995) nel paragrafo 3.2.5, ma

**Tabella 3.3** Valori di area esposta per i tripeptidi Gly – X - Gly

AREA ESPOSTA Gly - X - Gly, nm <sup>2</sup>		
	BACKBONE	SIDCHAIN
ALANINA		1.32
ARGININA		2.759
ASPARAGINA		1.853
ASPARTATO		1.798
CISTEINA		1.615
FENILALANINA		2.52
GLUTAMMINA		2.138
GLUTAMMATO		2.085
GLICINA		0
ITSTIDINA	0.786	2.316
ISOLEUCINA		2.088
LEUCINA		2.103
LISINA		2.506
METIONINA		2.295
PROLINA		1.8
SERINA		1.456
TREONINA		1.73
TRIPTOFANO		2.995
TIROSINA		2.685
VALINA		1.836

i valori sono molto più bassi. Le differenze sono principalmente dovute al fatto che l'analisi fatta utilizzando il tripeptide riduce la superficie esposta a causa dei due terminali di glicina. Nel momento in cui si voglia valutare l'additività secondo il metodo di Auton e Bolen, l'utilizzo del tripeptide come parametro di normalizzazione per le superfici esposte permette una migliore congruità con i risultati ottenuti per un intero polipeptide, dove i singoli residui non sono isolati, ma inclusi nella sequenza amminoacidica.

### 3.3.5 Valutazione dell'additività per il GB1 – hairpin

Per il GB1 – hairpin è stata valutata l'additività sia nella forma denaturata che nella forma nativa, utilizzando la procedura mostrata in Tabelle 3.4 e 3.4. La forma denaturata è derivata dalla simulazione 49 (GB1 in urea). Il GB1 si compone di sedici residui, e per le nostre simulazioni ne è stata utilizzata la forma *capped*, ossia con un terminale acetilico ed uno amminico alle estremità. Questa scelta rende non carichi i residui di testa, permettendo di trattarne il *backbone* allo stesso modo dei residui interni (v. paragrafo 2.4.2) e di restituire dei

valori di *coefficiente di interazione preferenziale* non influenzati da fenomeni legati alle cariche (Auton e Bolen, 2004). Si sono quindi confrontati i *coefficienti di interazione preferenziale*  $\Gamma_{23}^I$  ottenuti dalle simulazioni in presenza di trealosio e saccarosio (prima riga in Tabella 3.6) con quelli ottenuti considerando il metodo additivo, simulato e sperimentale, teorizzato da Auton e Bolen (seconda e terza riga in Tabella 3.6). Come si può osservare dalla Tabella 3.6, l'approssimazione per le forme denaturate è discreta (-1.718 vs -3.582 per il trealosio, -1.330 vs -2.840 per il saccarosio), mentre non altrettanto valida è la previsione per la forma nativa (-0.674 vs -3.052 per il trealosio, -0.375 vs -2.67 per il saccarosio). Si può notare inoltre che, mentre nelle simulazioni su GROMACS l'esclusione dell'eccipiente aumenta in maniera considerevole all'aumentare della frazione di backbone esposta a seguito di denaturazione, i valori di esclusione per la forma nativa non sono invece molto diversi da quelli per la forma denaturata quando si utilizza l'approccio additivo. Molto migliori, seppur non comunque corretti e con simile andamento, sono stati i risultati ottenuti utilizzando i dati sperimentali. Nel seguito si ipotizzeranno le possibili cause di queste discrepanze nei risultati ottenuti con l'approccio additivo.

**Tabella 3.4** Valori di area e contributo additivo per i residui del  $\beta$ -hairpin in presenza di trealosio

	BACKBONE				SIDECHAIN			
	AREA, nm <sup>2</sup>		$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>		AREA, nm <sup>2</sup>		$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>	
	NAT.	DEN.	NAT.	DEN.	NAT.	DEN.	NAT.	DEN.
<b>GLY</b>	0.597	0.801	-0.118	-0.159	0	0	0	0
<b>GLU</b>	0.524	0.749	-0.104	-0.148	1.808	1.971	-0.812	-0.886
<b>TRP</b>	0.692	0.804	-0.137	-0.159	1.881	2.143	0.688	0.783
<b>THR</b>	0.530	0.744	-0.105	-0.147	0.979	1.327	0.161	0.218
<b>TYR</b>	0.606	0.776	-0.120	-0.154	1.408	1.946	0.394	0.544
<b>ASP</b>	0.440	0.785	-0.087	-0.155	0.756	1.643	-0.352	-0.766
<b>ASP</b>	0.733	0.770	-0.145	-0.153	1.197	1.463	-0.558	-0.682
<b>ALA</b>	0.803	0.704	-0.159	-0.139	0.836	1.008	0.003	0.003
<b>THR</b>	0.721	0.751	-0.143	-0.149	0.935	1.198	0.154	0.197
<b>LYS</b>	0.612	0.795	-0.121	-0.157	1.554	2.013	-0.334	-0.432
<b>THR</b>	0.502	0.632	-0.099	-0.125	0.958	1.359	0.157	0.223
<b>PHE</b>	0.626	0.734	-0.124	-0.145	1.311	2.253	0.321	0.551
<b>THR</b>	0.570	0.777	-0.113	-0.154	1.051	1.522	0.173	0.250
<b>VAL</b>	0.620	0.723	-0.123	-0.143	0.915	1.330	0.122	0.177
<b>THR</b>	0.510	0.811	-0.101	-0.161	1.054	1.607	0.173	0.264
<b>GLU</b>	1.108	1.229	-0.219	-0.243	1.718	1.903	-0.772	-0.855

**Tabella 3.5** Valori di area e contributo additivo per i residui del  $\beta$ -hairpin in presenza di saccarosio

	BACKBONE				SIDECHAIN			
	AREA, nm <sup>2</sup>		$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>		AREA, nm <sup>2</sup>		$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>	
	NAT.	DEN.	NAT.	DEN.	NAT.	DEN.	NAT.	DEN.
<b>GLY</b>	0.542	0.861	-0.069	-0.110	0.000	0.000	0.000	0.000
<b>GLU</b>	0.529	0.748	-0.068	-0.096	1.601	1.636	-0.826	-0.844
<b>TRP</b>	0.644	0.682	-0.082	-0.087	1.766	2.557	0.631	0.913
<b>THR</b>	0.539	0.714	-0.069	-0.091	0.877	1.242	0.109	0.154
<b>TYR</b>	0.676	0.829	-0.087	-0.106	1.269	2.401	0.356	0.674
<b>ASP</b>	0.420	0.665	-0.054	-0.085	0.701	1.324	-0.340	-0.641
<b>ASP</b>	0.680	0.784	-0.087	-0.100	0.900	1.448	-0.436	-0.702
<b>ALA</b>	0.796	0.742	-0.102	-0.095	0.890	0.978	0.001	0.001
<b>THR</b>	0.740	0.786	-0.095	-0.101	0.939	1.539	0.117	0.191
<b>LYS</b>	0.602	0.790	-0.077	-0.101	1.689	2.112	-0.484	-0.605
<b>THR</b>	0.516	0.765	-0.066	-0.098	0.981	1.579	0.122	0.196
<b>PHE</b>	0.623	0.782	-0.080	-0.100	1.203	1.906	0.295	0.467
<b>THR</b>	0.513	0.754	-0.066	-0.097	0.925	1.369	0.115	0.170
<b>VAL</b>	0.592	0.754	-0.076	-0.096	1.038	1.321	0.085	0.109
<b>THR</b>	0.508	0.766	-0.065	-0.098	0.861	1.366	0.107	0.170
<b>GLU</b>	1.070	1.257	-0.137	-0.161	1.733	1.992	-0.894	-1.028

**Tabella 3.6** Confronto dei valori di gamma ottenuti durante le simulazioni su GROMACS, o utilizzando l'approccio additivo

	TREALOSIO		SACCAROSIO		
	$\Gamma_{23}^I$ , nm <sup>3</sup>	NATIVO	DENATURATO	NATIVO	DENATURATO
<b>FORMA ADDITIVA</b>		-3.052	-3.582	-2.670	-2.84
<b>FORMA ADDITIVA SPERIMENTALE</b>		-1.1	-1.4	-1.46	-1.66
<b>GROMACS</b>		-0.674	-1.718	-0.375	-1.330

### 3.3.6 Valutazione dell'additività per le catene proteiche

La stessa analisi effettuata per il  $\beta$ -hairpin è stata ripetuta per intere catene proteiche (lisozima,  $\alpha$ -chimotripsinogeno e ribonuclease A), sfruttando i dati riportati in letteratura (Cloutier *et al.*, 2018). La procedura di calcolo è schematicamente illustrata in Tabelle 3.7-3.10, mentre i risultati del confronto tra esperimenti, simulazioni su GROMACS e approccio additivo è riportato in Tabella 3.11. Anche in questo caso, così come già osservato per il  $\beta$ -hairpin, i valori ottenuti tramite approccio additivo non riescono a riprodurre adeguatamente i risultati attesi.

**Tabella 3.7** Valori di area e contributo additivo per i residui del lisozima del bianco d'uovo in presenza di trealosio

	BACKBONE			SIDECHAIN	
	N° RES	AREA, nm <sup>2</sup>	$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>	AREA, nm <sup>2</sup>	$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>
<b>ALA</b>	28	1444.63	-3.639	1352.1	0.041
<b>ARG</b>	1	65.21	-0.164	133.77	-0.055
<b>ASN</b>	6	230.16	-0.580	292.18	0.213
<b>ASP</b>	21	1001.47	-2.523	1370.24	-6.386
<b>CYS</b>	2	69.3	-0.175	76.41	0.197
<b>GLN</b>	9	473.59	-1.193	736.46	0.055
<b>GLU</b>	5	238.74	-0.601	319.27	-1.435
<b>GLY</b>	37	1819.03	-4.582	0	0
<b>HIS</b>	1	75.06	-0.189	103.24	0.167
<b>ILE</b>	18	912.44	-2.299	810.2	1.354
<b>LEU</b>	19	843.7	-2.125	853.15	1.355
<b>LYS</b>	10	606.33	-1.527	1539.62	-3.305
<b>MET</b>	0	0	0	0	0
<b>PHE</b>	19	934.87	-2.355	793.33	1.939
<b>PRO</b>	13	737.05	-1.857	1033.37	1.160
<b>SER</b>	48	2632.87	-6.632	3589.98	-0.099
<b>THR</b>	47	2501.62	-6.302	3878.82	6.368
<b>TRP</b>	5	220.29	-0.555	164.83	0.603
<b>TYR</b>	13	618.26	-1.557	729.72	2.041
<b>VAL</b>	23	1034.84	-2.607	724.73	0.967

**Tabella 3.8** Valori di area e contributo additivo per  $\alpha$ -chimotripsinogeno in presenza di saccarosio

	<b>BACKBONE</b>			<b>SIDECCHAIN</b>	
	N° RES	AREA, nm <sup>2</sup>	$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>	AREA, nm <sup>2</sup>	$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>
<b>ALA</b>	44	2132.48	-3.473	1913.07	0.014
<b>ARG</b>	8	396.51	-0.646	1061.96	-1.047
<b>ASN</b>	28	1756.05	-2.860	2199.23	-0.688
<b>ASP</b>	18	902.04	-1.469	1273.07	-6.167
<b>CYS</b>	20	958.45	-1.561	624.58	1.984
<b>GLN</b>	20	1039.49	-1.693	1703.67	0.733
<b>GLU</b>	10	581.53	-0.947	652.38	-3.367
<b>GLY</b>	46	1936.58	-3.154	0	0.000
<b>HIS</b>	4	287.73	-0.469	274.24	0.220
<b>ILE</b>	20	952.54	-1.551	1031.74	0.781
<b>LEU</b>	38	1802.25	-2.935	1659.2	0.986
<b>LYS</b>	28	1705.93	-2.778	4006.88	-11.480
<b>MET</b>	4	153.75	-0.250	172.67	0.359
<b>PHE</b>	12	719.23	-1.171	786.08	1.928
<b>PRO</b>	18	1002.31	-1.632	898.39	1.053
<b>SER</b>	56	3216.67	-5.238	3588.03	2.045
<b>THR</b>	46	2443.72	-3.980	3445.33	4.282
<b>TRP</b>	16	810.14	-1.319	845.79	3.022
<b>TYR</b>	8	370.61	-0.604	677.98	1.904
<b>VAL</b>	46	2076	-3.381	1971.96	1.622

**Tabella 3.9** Valori di area e contributo additivo per ribonuclease A in presenza di trealosio

	BACKBONE			SIDECHAIN	
	N° RES	AREA, nm <sup>2</sup>	$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>	AREA, nm <sup>2</sup>	$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>
<b>ARG</b>	4	214.26	-0.540	517.26	-0.214
<b>ASN</b>	10	621.57	-1.566	996.58	0.726
<b>ASP</b>	5	291.33	-0.734	292.19	-1.362
<b>CYS</b>	8	314.31	-0.792	175.11	0.451
<b>GLN</b>	7	366.31	-0.923	546.31	0.041
<b>GLU</b>	5	282.22	-0.711	440.68	-1.980
<b>GLY</b>	3	208.82	-0.526	0	0.000
<b>HIS</b>	4	136.43	-0.344	263.62	0.427
<b>ILE</b>	3	112.79	-0.284	75.26	0.126
<b>LEU</b>	2	123.31	-0.311	93.32	0.148
<b>LYS</b>	10	628.18	-1.582	1525.79	-3.276
<b>MET</b>	4	194.44	-0.490	135.92	0.884
<b>PHE</b>	3	146.48	-0.369	101.09	0.247
<b>PRO</b>	4	270.58	-0.682	320	0.359
<b>SER</b>	15	920.77	-2.319	1130.69	-0.031
<b>THR</b>	10	490.98	-1.237	724.84	1.190
<b>TRP</b>	0	0	0.000	0	0.000
<b>TYR</b>	6	304.55	-0.767	553.11	1.547
<b>VAL</b>	9	433.74	-1.093	399.51	0.533

**Tabella 3.10** Valori di area e contributo additivo per ribonuclease A in presenza di saccarosio

BACKBONE			SIDECHAIN		
N° RES	AREA, nm <sup>2</sup>	$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>	AREA, nm <sup>2</sup>	$\Gamma_{23}^I$ * AREA NORMALIZZATA, nm <sup>3</sup>	
<b>ALA</b>	12	656.99	-1.070	584.44	0.004
<b>ARG</b>	4	214.26	-0.349	517.26	-0.510
<b>ASN</b>	10	621.57	-1.012	996.58	-0.312
<b>ASP</b>	5	291.33	-0.474	292.19	-1.415
<b>CYS</b>	8	314.31	-0.512	175.11	0.556
<b>GLN</b>	7	366.31	-0.597	546.31	0.235
<b>GLU</b>	5	282.22	-0.460	440.68	-2.274
<b>GLY</b>	3	208.82	-0.340	0	0.000
<b>HIS</b>	4	136.43	-0.222	263.62	0.212
<b>ILE</b>	3	112.79	-0.184	75.26	0.057
<b>LEU</b>	2	123.31	-0.201	93.32	0.055
<b>LYS</b>	10	628.18	-1.023	1525.79	-4.372
<b>MET</b>	4	194.44	-0.317	135.92	0.283
<b>PHE</b>	3	146.48	-0.239	101.09	0.248
<b>PRO</b>	4	270.58	-0.441	320	0.375
<b>SER</b>	15	920.77	-1.499	1130.69	0.645
<b>THR</b>	10	490.98	-0.800	724.84	0.901
<b>TRP</b>	0	0	0.000	0	0.000
<b>TYR</b>	6	304.55	-0.496	553.11	1.553
<b>VAL</b>	9	433.74	-0.706	399.51	0.329

**Tabella 3.11** Confronto dei valori di  $\Gamma_{32}$  ottenuti sperimentalmente, tramite simulazione su GROMACS (Cloutier et al., 2018), oppure sfruttando l'approccio additivo.

	<b>Lisozima</b>	<b><math>\alpha</math> – CGN A</b>	<b>Rnase A</b>	
$\Gamma_{23}^I$ , nm <sup>3</sup>	TREALOSIO	SACCAROSIO	SACCAROSIO	TREALOSIO
<b>FORMA ADDITIVA</b>	-36.28	-42.92	-14.37	-17.64
<b>FORMA ADDITIVA SPERIMENTALE</b>	-39	-50.5	-11.9	-14.4
<b>GROMACS</b>	-4.5	-11.1	-10.8	-10.6
<b>SPERIMENTALE</b>	-6.1	-12.5	-10.3	-10.3

### 3.4 Commenti sui risultati

Come si può vedere dai risultati presentati sia per il  $\beta$ -hairpin (Tabella 3.6) che per intere catene proteiche (Tabella 3.11), i risultati ottenuti con il metodo additivo sono molto differenti da quelli ottenuti sia sperimentalmente, sia tramite simulazione su GROMACS. Sia per il trealosio che per il saccarosio tutti i risultati sono qualitativamente corretti, essendo infatti il *coefficiente di esclusione preferenziale*  $\Gamma_{23}^I$  negativo per tutti i polipeptidi studiati, ma con il metodo di Auton e Bolen si tende a sovrastimare quantitativamente il valore di questo coefficiente. Si può notare in generale che maggiore è il numero di residui che compongono la proteina (16 residui compongono il  $\beta$ -hairpin, 124 Rnase -A, 245  $\alpha$  – CGN A, 330 il lisozima), maggiore è la differenza tra il valore calcolato e quello ricavato con approccio additivo. Si può inoltre osservare (Tabella 3.6) che, per il  $\beta$ -hairpin, la differenza è molto più accentuata per la forma nativa rispetto a quella denaturata.

Alla luce delle considerazioni effettuate, si possono proporre alcune ipotesi su quali motivi abbiano portato ad ottenere dei risultati quantitativamente non accurati.

È importante notare che una ottima additività è stata osservata per la serie di N-acetilglicinammide (v. paragrafo 3.3.2). Infatti il *coefficiente di esclusione preferenziale* diminuisce in maniera lineare (come dimostrato dall'elevato valore di  $R^2$ ) con l'aumentare dei residui di glicina nella catena, rispecchiando con ottima approssimazione la realtà sperimentale (Auton e Bolen, 2004). Tuttavia l'accordo non è stato altrettanto buono per polipeptidi di grosse dimensioni, specialmente quando nella forma nativa. Si può quindi ipotizzare che la discrepanza osservata applicando il metodo additivo scaturisca principalmente dal fatto che, in grossi polipeptidi, i residui interni sono esposti a condizioni di solvatazione molto diverse rispetto a quelli sulla superficie.

In particolare, in un polipeptide con molti residui, la maggior parte di essi è esposta ad altri residui, e non al solvente (v. paragrafo 1.4), e perciò non interagisce con esso. Questa

ipotesi è stata confermata dal fatto che, anche valutando l'additività con le energie di trasferimento sperimentali fornite da Auton e Bolen (Auton e Bolen, 2005), si sono ottenute discrepanze simili a quelle riscontrate usando i valori forniti da GROMACS (Tabella 3.11). Inoltre, anche in questo caso la qualità dell'accordo è risultata migliore per i polipeptidi più piccoli. Ciò indica che la carenza di accordo con i dati sperimentali è dovuta non, o non solo, ad imprecisioni del campo di forza usato, ma principalmente ad inaccuratezze intrinseche all'approccio additivo stesso. Ulteriore conferma è stata fornita dalla valutazione di additività sui polipeptidi effettuata con gli stessi metodi elencati nei paragrafi 2.4 e seguenti, ma corretta, tramite uno script Python (v. appendice A.5), includendo solamente i contributi di *coefficiente di esclusione preferenziale* (o energia libera di trasferimento) dei residui che esponessero al solvente almeno l'80% della superficie di sidechain o backbone esposta in condizioni massime (tripeptide GLY-X-GLY). Applicando questa correzione, i risultati ottenuti migliorano di molto la corrispondenza coi dati sperimentali (Tabella 3.12).

Anche per il GB1-hairpin i risultati forniti dall'approccio additivo usato in questo lavoro sono in accordo con quelli ottenibili usando direttamente i valori presenti nei lavori di Auton e Bolen (Tabella 3.6). Inoltre, si può di nuovo osservare che l'accordo dell'approccio additivo con il valore ottenuto per l'intero polipeptide migliora qualora si considerino solo i residui effettivamente esposti al solvente esterno (Tabella 3.13). Ciò è vero però solo per la forma nativa, e non per quella denaturata, dove gli amminoacidi sono già sufficientemente esposti all'ambiente acquoso circostante.

**Tabella 3.12** Confronto dei valori ponderati di  $\Gamma_{32}^I$  ottenuti sperimentalmente, tramite simulazione su GROMACS (Cloutier et al., 2018), oppure sfruttando l'approccio additivo. Solo i residui che esponessero al solvente almeno l'80% della superficie di sidechain o backbone sono stati considerati.

	<b>Liozima</b>	<b><math>\alpha</math> – CGN A</b>	<b>Rnase A</b>	
$\Gamma_{32}^I, \text{nm}^3$	TREALOSIO	SACCAROSIO	SACCAROSIO	TREALOSIO
<b>FORMA ADDITIVA</b>	-14	-17.2	-4.8	-6.3
<b>FORMA ADDITIVA SPERIMENTALE</b>	-14.6	-22.5	-4.7	-5.1
<b>GROMACS</b>	-4.5	-11.1	-10.8	-10.6
<b>SPERIMENTALE</b>	-6.1	-12.5	-10.3	-10.3

**Tabella 3.13** Confronto dei valori ponderati di  $\Gamma_{32}^I$  per il GB1 -hairpin ottenuti sperimentalmente, tramite simulazione su GROMACS oppure sfruttando l'approccio additivo. Solo i residui che esponessero al solvente almeno l'80% della superficie di sidechain o backbone sono stati considerati.

$\Gamma_{23}^I, \text{nm}^3$	TREALOSIO		SACCAROSIO	
	NATIVO	DENATURATO	NATIVO	DENATURATO
<b>FORMA ADDITIVA</b>	-2.99	-5.63	-1.84	-2.41
<b>FORMA ADDITIVA SPERIMENTALE</b>	-0.55	-1.72	-1.27	-1.83
<b>GROMACS</b>	-0.674	-1.718	-0.375	-1.330

Si può quindi concludere che il metodo additivo funziona in maniera adeguata se si considerano solo i residui effettivamente esposti al solvente, come si evince confrontando le Tabelle 3.6 e 3.11 con le Tabelle 3.13 e 3.12.

I dati riportati in questa tesi dimostrano che l'approccio additivo, così come formulato da Auton e Bolen, si basa sul presupposto che tutti i residui si trovino nelle stesse condizioni di solvatazione, e non funziona quindi adeguatamente per grossi polipeptidi nella forma nativa. Infatti, nei lavori di Auton e Bolen (Auton e Bolen, 2005) si sono perlopiù studiati cosoluto non lioprotettivi, ma denaturanti, come ad esempio l'urea, e si sono utilizzati polipeptidi con catene più corte (dai 98 ai 145 residui) rispetto a quelli investigati in questo lavoro. Molto probabilmente quindi, l'additività osservata sperimentalmente è relativa principalmente a molecole polipeptidiche che espongono gran parte della loro superficie alla soluzione, come possono essere proteine di piccola dimensione o nella forma denaturata.



#### 4. COMMENTI FINALI E PROSPETTIVE FUTURE

Il presente lavoro di tesi può essere considerato come un approccio nuovo al problema della formulazione dei farmaci. Utilizzando a pieno le attuali possibilità computazionali, si è provato a trattare in modo diverso un problema usualmente affrontato a livello sperimentale.

I risultati ottenuti permettono di considerare la dinamica molecolare come un buon elemento di aiuto per la formulazione di farmaci a base proteica. Le formule di struttura dei componenti chimici, le topologie dell'acqua disponibili e l'elaborazione di un campo di forza adeguato hanno permesso di riprodurre in modo dettagliato lo stato di equilibrio di un ambiente molto simile a quello di una formulazione farmaceutica (Wang, 2000).

Prima ancora di valutare la bontà qualitativa dei risultati ottenuti, si sono utilizzate le simulazioni su scala molecolare per confermare i meccanismi proposti in letteratura per descrivere come un eccipiente zuccherino possa stabilizzare una formulazione proteica, permettendo, oltre a ciò, di confermare come la dinamica molecolare possa aiutare nella valutazione di dati non sempre facilmente accessibili sperimentalmente. È stato dunque confermato che gli eccipienti, se buoni lioprotettori, risultano esclusi all'equilibrio dalla superficie del principio attivo rispetto al solvente, ma che con lo stesso principio attivo tendono comunque a formare un certo numero di legami a idrogeno che, aumentando dopo la sublimazione del ghiaccio, risultano fondamentali in una stabilizzazione non più termodinamica, bensì cinetica, della formulazione. È stato inoltre osservato come le differenze qualitative, a livello di amminoacidi, siano esclusivamente legate alle proprietà delle catene laterali sia per le interazioni con gli zuccheri che per le interazioni con il solvente. Queste corrispondenze permettono di affermare che anche in presenza di singoli amminoacidi e non solo di grandi polipeptidi, il campo di forza è qualitativamente corretto.

A livello quantitativo è stato importante effettuare un confronto con i dati sperimentali. A questo scopo si è tentato di correlare il *coefficiente di esclusione preferenziale* valutato per i singoli amminoacidi con le energie libere di trasferimento ottenute da Auton e Bolen (Auton *et al.*, 2008). La mancata corrispondenza in questo caso indica ancora la necessità di miglioramento del campo di forza CHARMM36 modificato da Trout (Cloutier *et al.*, 2018), che porta a ottimi risultati per grandi polipeptidi, ma non altrettanto per i residui osservati in queste simulazioni.

Per quanto infatti il campo di forza porti a risultati sempre qualitativamente corretti, e quantitativamente molto vicini a quelli sperimentali (Cloutier *et al.*, 2018), piccoli errori relativi ai singoli amminoacidi possono portare a grandi differenze in una valutazione di tipo additivo come quella effettuata nel presente lavoro. Prendendo ad esempio il metodo di calcolo della *sidechain*, un valore errato di *coefficiente di esclusione preferenziale* per la glicina può portare a una valutazione errata dei contributi per la catena laterale di tutti i residui componenti un polipeptide. Si è visto infatti in letteratura (Auton *et al.*, 2008) che il valore di  $\Delta G_{trasf}$  per la glicina calcolato sperimentalmente risulta molto distante rispetto a quello ricavato nei presenti calcoli sia per trealosio che per il saccarosio (217 vs. 368 cal/mol e 218 vs. 779 cal/mol) e ciò indica una importante discrepanza considerando la notevole precisione di calcolo disponibile per questo amminoacido (Auton *et al.*, 2007). Per il *backbone* invece una valutazione imprecisa di anche solo una delle componenti della serie di *N-acetilglicinammide* può portare a una valutazione errata del suo *coefficiente di esclusione preferenziale*, per quanto, in questo caso, i risultati ottenuti sono in ottimo accordo con quelli sperimentali (si veda il paragrafo 3.3.2). Si può inoltre far

riferimento ai dati riportati in Tabella 3.1, dove è chiaramente mostrato che sia per il saccarosio che per il trealosio, i dati calcolati per il  $\Delta G_{trasf}$  delle catene laterali divergono significativamente da quelli ottenuti sperimentalmente. In quest'ultimo caso si può però aggiungere che, come descritto in letteratura (Auton *et al.*, 2007), anche i risultati sperimentali, eccetto che per la glicina, possono risultare imprecisi a causa delle incertezze di misura.

Anche il calcolo delle aree esposte potrebbe tuttavia essere una fonte di errore. La valutazione delle aree può infatti rivelarsi diversa a seconda dell'algoritmo utilizzato, e, nel caso dei polipeptidi trattati nella pubblicazione di Cloutier *et al.* del 2018, è stata fatta l'ipotesi aggiuntiva che la conformazione assunta dal polipeptide all'equilibrio in una soluzione con protettore e solvente sia la struttura cristallina riportata nell'archivio pdb. Queste differenze, anche se di modesta entità, possono portare a grandi variazioni a livello di risultato finale, soprattutto per polipeptidi di grandi dimensioni.

In conclusione, e come prospettiva per il futuro, si può quindi affermare che, sia dal punto di vista sperimentale che computazionale, sono possibili miglioramenti per ciò che riguarda la teoria dell'additività nelle formulazioni contenenti biofarmaci. Il campo di forza può essere migliorato cercando di rendere concordi i dati ottenuti a livello simulativo con quelli ottenuti sperimentalmente anche per residui più piccoli, concentrandosi maggiormente sulla modellazione dei singoli amminoacidi o di piccoli polipeptidi, e considerando che probabilmente le interazioni meglio modellate sono quelle dei residui polari, poiché sono i più esposti nella formulazione allo stato nativo dal momento che interagiscono favorevolmente con l'acqua e sfavorevolmente con lo zucchero, e rendono quindi quest'ultimo escluso dalla superficie, portando a distribuzioni in una *box* coerenti con quelle della realtà sperimentale. A livello di additività sperimentale sarebbe invece opportuno verificare quanto l'esposizione al solvente possa alterare i dati ottenuti anche in polipeptidi di maggior dimensione, tenendo conto che una proteina allo stato nativo è principalmente esposta a sé stessa, più che al solvente esterno. Non si può quindi escludere che la teoria dell'additività, a oggi, debba essere parzialmente riformulata.

Sicuramente, alla luce dell'attuale stato dell'arte, la dinamica molecolare, insieme a un approccio additivo, potranno in futuro permettere di accelerare e semplificare la procedura per la selezione di formulazioni proteiche, risultando di grande aiuto al crescente mercato dei biofarmaci, sempre più utilizzati per le loro più complesse e varie proprietà curative.

## LISTA DEI SIMBOLI

SIMBOLO	UNITÀ DI MISURA	DEFINIZIONE
$a$	m	Raggio idrodinamico
$a$	$\text{m s}^{-2}$	accelerazione
$c$	$\text{Mol l}^{-1}$	Concentrazione molare
$C_1$	-	Generica costante
$C_2$	-	Generica costante
$d$	$\text{kg m}^{-3}$	densità
$F$	N	Forza
$g$	adimensionato	Funzione di distribuzione radiale
$G$	$\text{nm}^3$	Integrale di Kirchwood - Buff
$\Delta G$	J	Energia libera di trasferimento
$\Delta h$	J	Differenza di entalpia
$k$	-	Generica costante
$k_b$	$\text{J K}^{-1}$	Costante di Boltzmann
$L$	m	Lunghezza
$m$	$\text{mol (kg di solvente)}^{-1}$	Molalità
$m$	kg	Massa componente
$M$	$\text{mol (l di solvente)}^{-1}$	molarità
$M$	$\text{kg mol}^{-1}$	Peso molecolare
$N$ o $n$	unità	Numero di unità
$N_{av}$	$\text{mol}^{-1}$	Numero di Avogadro
$Q$	Coulomb	Carica elettrostatica
$r$	m	Distanza spaziale
$R$	$\text{J mol}^{-1} \text{K}^{-1}$	Costante universale dei gas perfetti

$s$	adimensionato	solubilità
$t$	s	tempo
$t_r$	s	Tempo di raffreddamento
$\Delta t$	s	Differenza temporale
$T$	K	Temperatura
$T_g$	K	Temperatura di transizione vetrosa
$\Delta T$	K	Differenza di temperatura
$u$	$\text{N m}^{-1}$	Potenziale elettrostatico esterno
$v$	$\text{m s}^{-1}$	Velocità
$V$	$\text{N m}^{-1}$	Potenziale elettrostatico
$wt$	kg	Massa totale soluzione
$x$	adimensionata	Frazione massica
$z$	m	Distanza spaziale salti diffusionali

### LETTERE GRECHE

$\alpha$	adimensionato	Area normalizzata
$\alpha_{surf}$	$\text{W m}^{-1} \text{K}^{-1}$	Superficie media di scambio
$\gamma$	-	Coefficiente di attività
$\Gamma$	-	Coefficiente di esclusione preferenziale
$\Gamma^I$	$\text{nm}^3$	Coefficiente di esclusione preferenziale
$\Theta$	-	Angolo di bending
$\Phi$	-	Angolo di torsione
$\varepsilon$	$\text{J mol}^{-1}$	Coefficiente Potenziale LJ
$\varepsilon_0$	$\text{F m}^{-1}$	Costante dielettrica nel vuoto
$\eta$	cP	Viscosità
$\lambda_{eff}$	$\text{W m}^{-2} \text{K}^{-1}$	Conducibilità termica
$\mu$	$\text{J mol}^{-1}$	Potenziale chimico
$\rho$	$\text{kg m}^{-3}$	Densità
$\sigma$	m	Coefficiente Potenziale LJ

$\tau$	s	Tempo di rilassamento
$\chi$	-	Fattore water replacement

## APICI E PEDICI

PEDICE	SIGNIFICATO	APICE	SIGNIFICATO
$\infty$	Valore di correzione	$^{\circ}$	Condizioni standard
1	solvente	bb	backbone
2	proteina	Den	Forma Denaturata
3	cosoluto	Gly	Glicina
A	Generico componente	Nacet	N-acetilglicinammide
app	apparente	Nat	Forma nativa
B	Generico componente	nPT	Ensamble nPT
c	Concentrazione molare	res	residuo
Coulomb	-	sc	sidechain
EG	Gruppi terminali	$\vartheta$	Angolo di bending
f	Prodotto congelato	$\Phi$	Angolo di torsione
i	Generico componente		
int	intrinseco		
Intramolecolare	-		
IU	Gruppi interni		
j	Generico componente		
k	Generico componente		
l	Generico componente		
LJ	Lennard - Jones		
m	Concentrazione molale		
r	Distanza spaziale		
s	solido		
os	Presenza di cosoluto		



## APPENDICE

### A.1 File di minimizzazione energetica (*enermin.mdp*)

```
integrator = steep
dt = 0.001
nsteps = 3000
nstcgsteep = 2
nstlist = 10
ns_type = grid
rlist = 1.0
```

```
coulombtype = PME
pme_order = 4
optimize_fft = yes
```

```
emtol = 1000.0
emstep = 0.01
```

### A.2 File di equilibratura (*npt.mdp*)

```
define                = -DPOSRES
integrator             = md
tinit                 = 0
dt                    = 0.002
;emstep               = 0.01
comm_grps             = system
nsteps                = 500000
nstxout               = 10000
nstvout               = 0
nstlog                = 10000
nstenergy             = 1000
;nstxtcout            = 10000
;xtc_grps             = Protein
energygrps            = System
nstlist               = 5
ns_type               = grid
pbc                   = xyz
```

```
=====
; non-bonded
=====
rlist                 = 1.2
coulombtype           = PME
rcoulomb              = 1.2
pme_order             = 4
optimize_fft          = yes
vdwtype               = cut-off
rvdw                  = 1.2
DispCorr              = enerPres
cutoff-scheme         = group
```

```

;ewald_rtol          = 1e-5
;=====
;          T-coupling:
;=====
tcoupl              = berendsen
tc-grps             = Protein Non-Protein
tau_t               = 1.0 1.0
ref_t               = 300 300
;=====
;          P-coupling:
;=====
Pcoupl              = berendsen
tau_p               = 1.0
ref_p               = 1.0
compressibility     = 4.5e-5
refcoord_scaling   = com
;=====
;          initial velocities
;=====
gen_vel             = yes
gen_temp            = 300
gen_seed            = 182756
;=====
;          constraints
;=====
constraints         = all-bonds
constraint-algorithm = Lincs

;lincs_order        = 8
;lincs_iter         = 100

```

### A.3 File di simulazione della traiettoria (md.mdp)

```

integrator          = md
tinit              = 0
dt                 = 0.002
emstep             = 0.01
comm_grps          = system
nsteps             = 30000000
nstxout            = 0
nstvout            = 0
nstfout            = 0
nstlog             = 10000
nstenergy          = 1000
nstxtcout          = 1000
;xtc_grps          = Protein
energygrps         = System

```

```

nstlist          = 5
ns_type          = grid
pbc              = xyz

;=====
;          non-bonded
;=====
rlist            = 1.2
coulombtype      = PME
rcoulomb         = 1.2
pme_order        = 4
optimize_fft     = yes
vdwtype          = cut-off
rvdw             = 1.2
DispCorr         = enerPres
cutoff-scheme    = group
;ewald_rtol      = 1e-5

;=====
;          T-coupling:
;=====
tcoupl           = nose-hoover
tc-grps          = Protein Non-Protein
tau_t            = 0.5 0.5
ref_t            = 300 300

;=====
;          P-coupling:
;=====
Pcoupl           = Parrinello-Rahman
pcoupltype       = isotropic
tau_p            = 3.0
ref_p            = 1.0
compressibility  = 4.5e-5

;=====
;          initial velocities
;=====
gen_vel          = no
gen_temp         = 300
gen_seed         = 182756

;=====
Constraints
;=====
constraints      = all-bonds
constraint-algorithm = Lincs

;lines_order     = 8
;lines_iter      = 100

```

#### A.4 File di calcolo dei *KBI* (KBI.py)

```
#!/usr/bin/python
import sys
import math
import numpy as np
import matplotlib.pyplot as plt
#check command lines arguments
if len(sys.argv)!=2 :
print("I need the name of the rdf file"); exit(1)
with open(sys.argv[1], 'r') as f:
data = f.readlines()
distanza = []
integrando = []
for row in data:
values = row.split()
if values[0]!='#' and values[0]!='@' and values[0]!='@TYPE':
d = float(values[0])
distanza.append(d)
i = float(values[1])
integrando.append((i-1)*d*d)
integral = []
for linenr, radius in enumerate(distanza):
if radius >= 1 and radius <= 1.4:
integral.append(np.trapz(integrando[0:linenr], distanza[0:linenr]))
KBI = np.mean(integral)*4*math.pi
print("KBI is %f" %KBI)
```

#### A.5.1 File di analisi delle aree ponderate per il backbone (sasa-back.py)

```
#!/usr/bin/python
import sys
import math
import numpy as np
import pandas as pd
#check command lines arguments
if len(sys.argv)!=3 :
    print("I need the name of the input file and the fraction to be considered"); exit(1)
with open(sys.argv[1], 'r') as f:
    sasa = f.readlines()
sasadict={}
resname=[]
ressasa=[]
```

```

for row in sasa:
    resname.append(row.split()[3])
    ressasa.append(float(row.split()[5]))
sasadict={'name':resname, 'sasa':ressasa}
df_real=pd.DataFrame(sasadict)
with open('max_sasa_back', 'r') as f:
    maxsasa = f.readlines()
max_sasadict={}
resname=[]
ressasa=[]
for row in maxsasa[:-1]:
    resname.append(row.split()[0])
    ressasa.append(float(row.split()[1]))
max_sasadict={'name':resname, 'sasa':ressasa}
df_max=pd.DataFrame(max_sasadict)
#print(df.groupby('name').sum())
for numres, restype in df_max.iterrows():
    rescount=0
    sasasum=0
    for numseq, res in df_real.iterrows():
        if res['name']==restype['name'] and res['sasa']>float(sys.argv[2])*restype['sasa']:
            rescount+=1
            sasasum+=res['sasa']
    #print("Ci sono %d residui %s con area totale %f" %(rescount, restype['name'], sasasum))
    print("%d %s %f " %(rescount, restype['name'], sasasum))

```

#### A.5.2 File di analisi delle aree ponderate per la sidechain (sasa-side.py)

```

#!/usr/bin/python
import sys
import math
import numpy as np
import pandas as pd
#check command lines arguments

```

```

if len(sys.argv)!=3 :
    print("I need the name of the input file and the fraction to be considered"); exit(1)
with open(sys.argv[1], 'r') as f:
    sasa = f.readlines()
sasadict={}
resname=[]
ressasa=[]
for row in sasa:
    resname.append(row.split()[3])
    ressasa.append(float(row.split()[5]))
sasadict={'name':resname, 'sasa':ressasa}
df_real=pd.DataFrame(sasadict)
with open('max_sasa_side', 'r') as f:
    maxsasa = f.readlines()
max_sasadict={}
resname=[]
ressasa=[]
for row in maxsasa[:-1]:
    resname.append(row.split()[0])
    ressasa.append(float(row.split()[1]))
max_sasadict={'name':resname, 'sasa':ressasa}
df_max=pd.DataFrame(max_sasadict)
#print(df.groupby('name').sum())
for numres, restype in df_max.iterrows():
    rescount=0
    sasasum=0
    for numseq, res in df_real.iterrows():
        if res['name']==restype['name'] and res['sasa']>float(sys.argv[2])*restype['sasa']:
            rescount+=1
            sasasum+=res['sasa']
#print("Ci sono %d residui %s con area totale %f" %(rescount, restype['name'], sasasum))
print("%d %s %f " %(rescount, restype['name'], sasasum))

```

## BIBLIOGRAFIA

- Abraham, M.J., Hess, B., Van der Spoel, D., Lindahl, E. e GROMACS team, *Gromacs User Manual version 5.0.7.*, www.gromacs.org, Department of Biophysical Chemistry, University of Groningen, Paesi Bassi: Groningen, 2005.
- Allen, M.P., Introduction to Molecular Dynamics Simulation, Computational Soft Matter: From Synthetic Polymers to Proteins, 2004. *Lecture Notes, NIC Series*, **Vol. 23**: 1-28.
- Anderson, C.F., Courtenay, E.S., Record, M.T.J., Thermodynamic expressions relating different types of preferential interaction coefficients in solutions containing two solute components, 2002. *J. Phys. Chem. B.*, **106**: 418 – 433.
- Arakawa, T., Prestrelski, S.J., Kenney, W.C., Carpenter, J.F., Factors affecting short-term and long-term stabilities of protein, 1993. *Adv. Drug Deliv. Rev.*, **10**: 1–28.
- Auton, M., Bolen, D.W., Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale, 2004. *Biochem.*, **43**: 1329-1342.
- Auton, M., Bolen, D.W., Holthausen, L.M.F., Anatomy of energetic changes accompanying urea-induced protein denaturation, 2007. *PNAS*, **104**: 15317-15322.
- Auton, M., Bolen, D.W., Predicting the energetics of osmolyte-induced protein folding/unfolding, 2005. *PNAS*, **102**: 15065-15068.
- Barelli, S., Cannellini G., Thadikkaran, L., Crettaz, D., Quadroni, M., Rossier, J.S., Tissot, J.D., Lion, N., Oxidation of proteins: Basic principles and perspectives for blood proteomics, 2008. *Proteomics Clin. Appl.*, **2**: 142–157.
- Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R., Molecular dynamics with coupling to an external bath, 1984. *J. Chem. Phys.*, **81**: 3684.
- Bolen, D.W., Baskakov, I. V., The osmophobic effect: natural selection of a thermodynamic force in protein folding, 2001. *J. Mol. Biol.*, **310**: 955-963.
- Brooks, B.R., Bruccoleri, R.E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M., CHARMM - A program for macromolecular energy, minimization, and dynamics calculations, 1983. *J. Comput. Chem.*, **4**: 187–217.
- Carpenter, J.F., Chang, B.S., *Lyophilization of protein pharmaceuticals*. In: Avis, K.E., Wu, V.L., *Biotechnology and Biopharmaceutical Manufacturing, Processing and Preservation*, Interpharm Press: Buffalo Grove, Stati Uniti, 1996.
- Carpenter, J.F., Crowe, L.M., Crowe, J.H., Stabilization of phosphofructokinase with sugars during freeze-drying: characterization of enhanced protection in the presence of divalent cations, 1987. *Biochim. Biophys. Acta*, **923**: 109–115.
- Carpenter, J.F., Pikal, M.J., Chang, B.S., Randolph, T.W., Rational design of stable lyophilized protein formulations: Some practical advice, 1997. *Pharm. Res.*, **14**: 969–975.
- Chang, B.S., Beauvais, R.M., Dong, A., Carpenter, J.F., Physical factors affecting the storage stability of freeze-dried interleukin-I receptor antagonist: Glass transition and protein conformation, 1996. *Arch. Biochem. Biophys.*, **331**: 249–258.

Cleland, J.L., Powell, M.F., Shire, S.J., The development of stable protein formulations: a close look at protein aggregation, deamidation, and oxidation, 1993. *Crit. Rev. Ther. Drug Carrier Syst.*, **10**: 307–377.

Cloutier, T., Sudrik, C., Sathish, H.A., Trout, B.L., Kirkwood-Buff Derived Alcohol Parameters for Aqueous Carbohydrates and Their Application to Preferential Interaction Coefficient Calculations of Proteins, 2018. *J. Phys. Chem. B.*, **122**: 9350-9360.

Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C., Scharf, M., The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies, 1995. *J. Comput. Chem.*, **16**: 273-284.

Franks, F., Hatley, R.H.M., Mathias, S.F., Materials science and the production of shelf stable biologicals, 1991. *Bio. Pharm.*, **4**: 38–55.

Franks, F., Improved freeze-drying: from empiricism to predictability, 1990. *Cryo-Letters*, **11**: 93–110.

Ganguly, P., Hajari, T., Shea, J.E., van der Vegt, N.F., Mutual exclusion of urea and trimethylamine N-oxide from amino acids in mixed Solvent Environment., 2015. *J Phys Chem Lett.*, **6**: 581-585.

Ha, S.N., Giammona, A., Field, M., Brady, J.W., A Revised Potential-Energy Surface for Molecular Mechanics Studies of Carbohydrates, 1988. *Carbohydr. Res.*, **180**: 207-221.

Hageman, M., The role of moisture in protein stability, 1988. *Drug Dev. Ind. Pharm.*, **14**: 2047.

Headlam, H.A., Davies, M.J., Beta-scission of side-chain alkoxy radicals on peptides and proteins results in the loss of side-chains as aldehydes and ketones, 2002. *Free Radic. Biol. Med.*, **32**: 1171-1184.

Hess, B., Bekker, H., Berendsen, H. J. C., Fraaije, J.G.E. M., LINCS: A linear constraint solver for molecular simulations, 1997. *J. Comp. Chem.*, **18**: 1463-1472.

Hockney, R.W., Eastwood, J.W., *Computer simulations using particles*. Adam Hilger, Bristol: Gran Bretagna, 1988.

Hottot, A., Vessot, S., Andrieu, J., Determination of mass and heat transfer parameters during freeze-drying cycle of pharmaceutical products, 2005. *PDA J. Pharm. Sci. Technol.*, **59**: 138.

Ira Levine, *Physical Chemistry*, McGraw-Hill, Singapore, 1997.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D. Impey, R. W., Klein, M.L., Comparison of Simple Potential Functions for Simulating Liquid Water, 1983. *J. Chem. Phys.*, **79**: 926- 935.

Jorgensen, W.L., Maxwell, D.S., Tirado Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, 1996. *J. Am. Chem. Soc.*, **118**: 11225–11236.

Kharaghani, A., Metzger, T., Tsotsas, E., A proposal for discrete modeling of mechanical effects during drying, combining pore networks with DEM, 2011. *AIChE J.*, **57**: 872.

Kharaghani, A., Tsotsas, E., Wolf, C., Beutler, T., Guttzeit, M., Oetjen, G.-W., *Freeze-Drying*, In *Ullmann's Encyclopedia of Industrial Chemistry*; Wiley-VCH Verlag GmbH & Co.: Weinheim, Germania, 2017.

Kirkwood, J. G., Buff, F.P.J., The Statistical Mechanical Theory of Solutions. I, *Chem. Phys.*, **19**: 774.

Kusalik, P. G., Svishchev, I.M., The spatial structure in liquid water, 1994. *Science*, **265**: 1219-1221.

Lapidus, L.J., Understanding protein aggregation from the view of monomer dynamics, 2013. *Mol. Biosyst.*, **9**: 29-35.

Leader, B., Baca, Q.J., Golan, D.E., Protein therapeutics: A summary and pharmacological classification, 2008. *Nat. Rev. Drug Discov.*, **7**: 21-39.

Lee, B., Richards F.M., The interpretation of protein structures: Estimation of static accessibility, 1971. *J. Mol. Biol.*, **55**: 379–400

Lehninger, A., Nelson, D.L., Cox, M.M, *Lehninger Principles of Biochemistry*, IV, W.H. Freeman & Co., New York: Stati Uniti, 2005.

Maitland, G.C., Rigby, M., Smith, E.B., Wakeham, W.A., *Intermolecular forces: their origin and determination*; Clarendon Press, Oxford: Gran Bretagna, 1981.

Moore, E.B., Molinero, V., Structural transformation in supercooled water controls the crystallization rate of ice, 2011. *Nature*, **479**: 506.

Nosé, S., A unified formulation of the constant temperature molecular-dynamics methods, 1984. *J. Chem. Phys.*, **81**: 511–519.

Oetjen, W., Haseley, P., *Freeze-Drying*, 2<sup>nd</sup> ed., Wiley-VCH Verlag GmbH & Co.: Weinheim, Germania, 2003.

Parinello, M., Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method, 1981. *J. Appl. Phys.*, **52**: 7182.

Pham, V.C., Anania, V.G., Qui, T., Phung, J., Lill, R., Chapter Fourteen - Complementary Methods for the Identification of Substrates of Proteolysis, 2014. *Methods Enzymol.*, **544**: 359 – 380.

Pikal, M.J., Roy, M.L., Shah, S., Mass and heat transfer in vial freeze-drying of pharmaceuticals: Role of the vial, 1984. *J. Pharm. Sci.*, **73**: 1224.

Pikal, M.J., Shah, S., Roy, M.L., Putman, R., The secondary drying stage of freeze-drying: Drying kinetics as a function of temperature and chamber pressure, 1990. *Int J. Pharm.*, **60**: 203.

Pisano, R., Arsiccio, A., Stability of Proteins in Carbohydrates and Other Additives during Freezing: The Human Growth Hormone as a Case Study, 2017. *J. Phys. Chem. B.*, **121**: 8652 – 8660.

Prestrelski, S.J., Tedeschi, N., Arakawa T., Carpenter, J.F., Dehydration-induced conformational changes in proteins and their inhibition by stabilizers, 1993. *Biophys. J.*, **65**: 661–671.

Rey, L., May, C.J., *Freeze Drying/Lyophilization of Pharmaceutical and Biological Products*, III, Informa Healthcare: New York, Stati Uniti, 2010.

Robinson, N.E., Protein deamidation, 2002. *PNAS*, **99**: 5283-5288.

Shrake, A., Rupley, J.A., Environment and exposure to solvent of protein atoms. Lysozyme and insulin, 1973. *J Mol Biol.*, **79**: 351–371.

Shulgin, I.L., Ruckenstein, E., Local composition in solvent + polymer or biopolymer systems, 2008. *J. Phys. Chem. B.*, **112**: 3005 – 3012.

Stone, A.J., *The Theory of Intermolecular Forces*; Clarendon Press, Oxford: Gran Bretagna, 1996.

Strambini, G.B., Gabellieri, E., Proteins in frozen solutions: evidence of ice-induced partial unfolding, 1996. *Biophys. J.*, **70**: 971–976.

Swope, W.C., Andersen, H.C., Berens, P.H., Wilson, K.R., A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters, 1982. *J. Chem. Phys.*, **76**: 637–649.

Tanaka, T., Takeda, T., Miyajama, R., Cryoprotective effect of sachharides on denaturation of catalase during freeze-drying, 1991. *Chem. Pharm. Bull. (Tokyo)*, **39**: 1091–1094.

Tanford, C., Protein denaturation. C. Theoretical models for the mechanism of denaturation, 1970. *Adv. Protein. Chem.*, **24**: 1-95.

Van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hunenberger, P. H., Kruger, P., Mark, A. E., Scott, W. R. P., Tironi, I. G., *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Vdf Hochschulverlag AG an der ETH Zurich, Zurigo, Svizzera: 1996.

Verlet, L., Computer experiments on classical fluids. Thermodynamical properties of Lennard-Jones molecules, 1967. *Phys. Rev.*, **159**: 98–103.

Walsh, G., Biopharmaceutical benchmarks 2014, 2014. *Nat. Biotechnol.*, **32**: 992 – 1000.

Walsh, G., Biopharmaceutical benchmarks 2018, 2018. *Nat. Biotechnol.*, **36**: 1136 – 1145.

Wang, W., Lyophilization and development of solid protein pharmaceuticals, 2000. *Int. J. Pharm.*, **203**: 1-60.

Wang, W., Roberts, C.J., *Aggregation of therapeutic proteins*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2010.

Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., Weiner, P., A new force-field for molecular mechanical simulation of nucleic acids and proteins, 1984. *J. Am. Chem. Soc.*, **106**: 765–784.

Willemer, H., Measurements of temperatures, ice evaporation rates and residual moisture contents in freeze-drying, 1992. *Dev. Biol. Stand.*, **74**: 123–134.

Zarekar, S., Bück, A., Jacob, M., Tsotsas, E., Reconsideration of the hydrodynamic behavior of fluidized beds operated under reduced pressure, 2016. *Powder Technol.*, **287**: 169.

Zhang, W.; Xiao, S., Ahn, D.U., Protein oxidation: Basic principles and implications for meat quality, 2013. *Crit. Rev. Food Sci. Nutr.*, **53**: 1191-1201.

