

Università degli studi di Padova

DIPARTIMENTO DI FISICA E ASTRONOMIA "GALILEO GALILEI"

Corso di Laurea in Fisica

Separazione di fase per proteine intrinsecamente disordinate

Laureando:

Benedetta Mariani

Relatore:

Prof. Antonio Trovato

Anno accademico 2016/2017

Indice

1	Introduzione alle proteine e alle proteine intrinsecamente disordinate.	1
1.1	Proteine globulari	1
1.2	Proteine intrinsecamente disordinate	2
1.3	Separazione di fase di IDP	3
1.4	Obiettivo del lavoro di tesi	4
2	Teoria di Flory Huggins e RPA	4
2.1	Entropia di mescolamento di Flory Huggins	4
2.2	Energia interna di Flory Huggins	6
2.3	Energia elettrostatica con Random Phase Approximation (RPA)	7
2.4	Determinazione dei diagrammi di fase	10
3	Separazione di fase per poliamfoliti	10
4	Dipendenza del diagramma di fase dalle sequenze di IDP	14
5	Conclusioni e prospettive	16
	Riferimenti bibliografici	20

1 Introduzione alle proteine e alle proteine intrinsecamente disordinate.

1.1 Proteine globulari

Le proteine sono le maggiori componenti della cellula vivente. Esse giocano ruoli cruciali nel mantenimento della vita, e una delle loro funzioni meglio capite è la catalisi enzimatica delle reazioni chimiche che avvengono fuori e dentro la cellula. Alla base di questo compito vi è la loro stessa struttura: essa presenta una particolare regione, chiamata sito attivo, che è modellata specificatamente sulla base della molecola con cui la proteina deve interagire. La stessa struttura tridimensionale della proteina è disposta in modo da consentire la stabilità di questo sito attivo, e, di conseguenza, la conformazione 3D di una proteina determina con grande specificità la funzione.

Le proteine sono polimeri, costituite da amminoacidi legati insieme a formare una catena peptidica: essa si forma attraverso il gruppo carbossilico e il gruppo amminico di due amminoacidi differenti, che legano i loro atomi C e N tramite la liberazione una molecola d'acqua (come mostrato in Figura 1).

Nei primi anni '50 Sanger mostrò che le sequenze dei residui degli amminoacidi (ovvero la porzione di amminoacido libero che persiste dopo la polimerizzazione) sono uniche per ogni proteina ([1]); la catena della proteina dunque è costituita da una *backbone* chimicamente regolare unita dal legame peptidico, da cui si dipartono diverse catene laterali costituite dai residui.

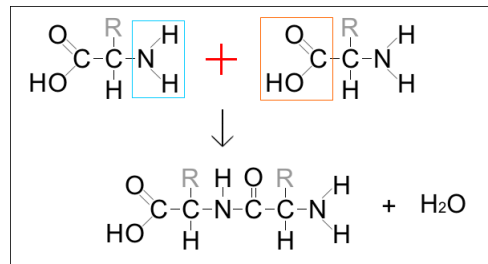


Figura 1: Formazione del legame peptidico tra due amminoacidi. Come si nota dalla figura, degli originari amminoacidi sopravvivono il gruppo carbonilico e il gruppo NH (tra cui a causa delle loro cariche parziali si potranno stabilire interazioni di legame a idrogeno) e i residui R_i , la cui specifica sequenza caratterizza la proteina.

La classe di proteine di cui si possiedono più conoscenze sono le proteine globulari solubili in acqua, essendo queste facili da studiare tramite la cristallografia a raggi X e la spettroscopia NMR. La configurazione 3D di queste proteine si viene a costituire in seguito a una gerarchia di strutture. La prima, la struttura primaria, è costituita dalla sequenza stessa di amminoacidi, ed è tenuta insieme da legami chimici covalenti molto più forti di quelli che mantengono la struttura 3D.

In seguito vi è una struttura secondaria, che è dovuta al ripiegamento della struttura primaria su se stessa e alla formazione di legami a idrogeno tra il gruppo carbonilico e il gruppo NH dei legami peptidici, e si può manifestare in una conformazione a α

eliche o a pieghe (β) detta anche a foglietto ripiegato. Un successivo ripiegamento di questa struttura su se stessa e interazioni tra gruppi di molecole della catena vanno a costituire la struttura terziaria finale.

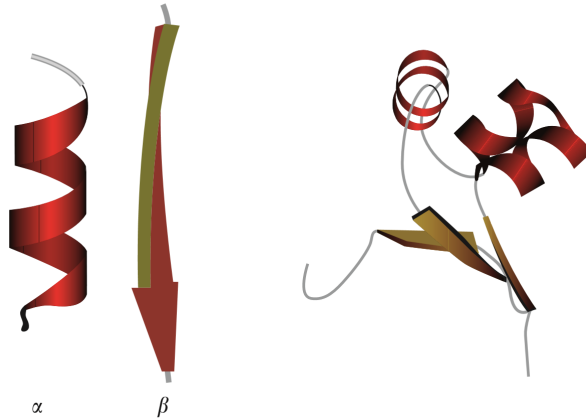


Figura 2: Struttura secondaria di una catena polipeptidica (α elica e a foglietto β), e struttura terziaria di una proteina globulare. Solitamente, considerate insieme, α elica e foglietti β costituiscono circa la metà della catena in una proteina globulare.

Il meccanismo con cui una proteina nel momento della biosintesi, o persino in vitro, si ripiega su se stessa, è chiamato folding. Esso è un processo "tutto o nulla", ovvero macroscopicamente corrisponde a una transizione del primo ordine, e da uno stato inizialmente disordinato chiamato *random coil* evolve al suo stato nativo di struttura terziaria. E' spesso però individuabile uno stato intermedio metastabile chiamato *molten globule*, che presenta meno compattezza rispetto allo stato nativo e più mobilità delle catene laterali.

1.2 Proteine intrinsecamente disordinate

Tuttavia, dopo che nel corso del ventesimo secolo molte strutture tridimensionali di proteine sono state comprese e inserite in una piattaforma di raccolta dati chiamata Protein Data Bank, sono state scoperte nuove proteine fundamentalmente distinte dalle solite proteine globulari. Queste particolari proteine non presentano una struttura ben definita, terziaria e/o secondaria e vengono chiamate intrinsecamente disordinate. Consistono infatti di strutture dinamiche che si modificano nel tempo e, disciolte in soluzione acquosa, possono assumere un ensemble di conformazioni diverse (dell'ordine delle centinaia). Tuttavia, è bene tenere a mente che la maggior parte di esse ha qualche regione di struttura secondaria transiente e che, in generale, la maggior parte delle proteine contiene regioni ordinate e disordinate in rapporti differenti. Solo il 32% delle strutture cristallizzate del Protein Data Bank sono completamente prive di disordine ([2]). A seconda del contenuto di struttura residua

e a seconda dell'ambiente, possono trovarsi per lo più in tre situazioni: Molten globule (MG), dunque simile allo stato intermedio del folding proteico, premolten globule e random coil. Recenti analisi hanno mostrato come le sequenze polari e di residui carichi nelle IDPs favoriscano il loro collasso in una struttura compatta di MG in soluzione acquosa.

Tra le caratteristiche fondamentale delle IDPs vi sono delle strutture primarie ricche di amminoacidi polari e carichi, dunque idrofili (come Tyr, Ser, Thr), che dunque non permettono loro di formare il core idrofobico che caratterizza le proteine ordinate e non possono stabilizzarsi in una struttura ordinata e globulare. Sono comunque in grado, interagendo con altri partner, di ripiegarsi su se stesse e di svolgere la loro funzione biologica, spesso importante nei casi di controllo cellulare o regolazione trascrizionale e traduzionale.

1.3 Separazione di fase di IDP

Interazioni tra IDPs multivalenti possono dare avvio a una separazione di fase di tipo liquido-liquido, da uno stato solubile in soluzione acquosa a uno insolubile, caratterizzato dalla formazione di gocce compatte della dimensione di micrometri. La formazione di queste gocce, quattro ordini di grandezza più viscoso dell'acqua, è importante per la vita della cellula in quanto si è recentemente scoperto che costituiscono degli organelli privi di membrana, sede di molte attività. Hanno funzioni importanti tra cui la conservazione e l'elaborazione dell'RNA. Queste transizioni avvengono in seguito a una supersaturazione molecolare, raggiunta attraverso cambiamenti regolati dall'espressione genica nella concentrazione delle proteine o dell'RNA, cambiamenti nello stato delle cariche (post-traslational modifications), cambiamenti nella concentrazione del sale/protoni e variazioni di temperatura ([3]). Stanno diventando però importanti anche per quanto riguarda lo studio di aggregati patologici di proteine: questa separazione di fase può rendere le proteine vulnerabili all'aggregazione e portare alla formazione di aggregati fibrillari insolubili. Questi aggregati patologici sono associati a inclusioni di elementi composti di fibre amiloidi semi-cristalline, come mostrato nel caso della proteina actin-regolatrice responsabile della sindrome neurale Wiskott-Aldrich. IDPs che danno avvio a questa transizione di fase sono spesso caratterizzate da sequenze con una bassa complessiva diversità di amminoacidi, e sequenze di interazione che si ripetono lungo la sequenza: risultano ricche di catene laterali polari, includono piccoli motivi multipli, e blocchi di cariche alternati. In questi sistemi le interazioni elettrostatiche tra gruppi di residui di cariche opposte promuovono la separazione di fase attraverso attrazioni a lungo range nella catena. Per studiare i polielettroliti è necessario tenere presente il patterning delle cariche e studi recenti hanno iniziato a trattare l'impatto del pattern di cariche nel comportamento delle IDPs.

1.4 Obiettivo del lavoro di tesi

L'obiettivo di questa tesi è riprodurre i diagrammi di fase di tale transizione per diverse sequenze proteiche, basandosi su una trattazione della termodinamica del sistema che fa affidamento alla teoria della Random Phase Approximation, con alcune riprese della teoria di Flory Huggins. Prima sono oggetto di analisi dei poliamfoliti caratterizzati dalla ripetizione di blocchi dalla carica alternata, poi la particolare proteina $Ddx4^{N1}$ e la sua mutazione tramite uno scambio di cariche $Ddx4^{N1}CS$, in modo da mostrare come le differenze tra queste sequenze influenzino le temperature di transizione.

Vengono infine mostrate le medesime curve dei diagrammi di fase per delle sequenze per le quali in un articolo su *Nature Materials* del 2015 ([4]) sono stati presentati dei valori sperimentali delle temperature di transizione, discutendo quanto il modello utilizzato in questa sede possa prevedere i risultati sperimentali e come potrebbe essere migliorato.

2 Teoria di Flory Huggins e RPA

Il comportamento delle IDPs è governato dalla loro sequenza di amminoacidi. Per compiere progressi nella comprensione di questa relazione tra sequenza e fase, è possibile utilizzare un'applicazione della RPA per eteropolimeri per considerare al meglio la elettrostatica specifica per ogni sequenza. Serve infatti una teoria che vada oltre la teoria classica di Flory Huggins per la separazione di fase, che è consona se le forze in gioco sono solo a corto range e di contatto come quelle che nascono tra sequenze idrofobiche. Rispetto ad altre teorie, l'approccio qui presentato ha il vantaggio di trattare la connettività della catena piuttosto esplicitamente e permette un inserimento diretto e chiaro del pattern di cariche. La teoria è concepita per soluzioni acquose di poliamfoliti neutri o quasi neutri come $Ddx4^{N1}$, con piccoli controioni monovalenti e sale.

2.1 Entropia di mescolamento di Flory Huggins

Un'introduzione alla teoria di Flory Huggins è comunque necessaria dato che la maggior parte della entropia configurazionale del sistema qui presentato è basata appunto sul modello a reticolo di FH: il volume spaziale è ripartito in siti del reticolo ognuno con volume a^3 , che è dell'ordine di una singola molecola di liquido. Questa teoria assume che vi siano due specie mescolate A e B, che riempiono l'intero reticolo formando un miscuglio di volume $V_A + V_B$. Possono essere dunque definite delle frazioni di volume ϕ_A e ϕ_B , dove $\phi_A = \frac{V_A}{V_A + V_B}$ e $\phi_B = 1 - \phi_A$. Una delle due specie può essere un polimero, nel caso più generale comunque il volume di un sito del reticolo v_0 è definito dalle unità più piccole (quindi dalle molecole di solvente o da monomeri). L'entropia è definita dal prodotto della costante di Boltzmann k_B col logaritmo naturale del numero di modi possibile di organizzare le molecole $S = k_B \ln \Omega$. Si hanno due contributi all'entropia di rimescolamento: il numero di

stati traslazionali possibile per una molecola sarà semplicemente pari al numero di posizioni indipendenti che può avere sul reticolo; in un miscuglio omogeneo di specie A e B sarà dunque pari a $\Omega_{AB} = M$, con M , numero dei siti di reticolo occupati, corrispondente a $\frac{V_A+V_B}{v_0}$. Andrà sottratto un termine dovuto agli stati occupati dalla specie A prima del rimescolamento ($\Omega_A = M\phi_A$); di conseguenza, per una singola molecola di specie A, la differenza di entropia dovuta al rimescolamento sarà data da $\Delta S_A = k_B \ln \frac{\Omega_{AB}}{\Omega_A} = k_B \ln \frac{1}{\phi_A}$, che evidenzia come il cambiamento di entropia in seguito a un rimescolamento sia sempre positivo. Per calcolare la entropia totale di rimescolamento, i due contributi relativi alle due specie vengono sommati entrambi con il loro peso:

$$\Delta S_{mix} = n_A \Delta S_A + n_B \Delta S_B = -k_B(n_A \ln \phi_A + n_B \ln \phi_B) \quad (1)$$

In cui n_A e n_B sono il numero di polimeri delle due specie, tra i quali ognuno ha rispettivamente N_A e N_B monomeri. Sostituendo nella formula i valori di n_A e n_B , ovvero $n_A = \frac{\phi_A M}{N_A}$ e $n_B = \frac{\phi_B M}{N_B}$, e dividendo tutto per M , il numero di siti di reticolo occupati, in modo da ottenere una grandezza intensiva, si ricava il seguente valore di entropia di rimescolamento:

$$\Delta s_{mix} = \frac{\Delta S_{mix}}{M} = -k_B \left(\frac{\phi_A}{N_A} \ln \phi_A + \frac{\phi_B}{N_B} \ln \phi_B \right) \quad (2)$$

in cui per soluzioni di polimeri $N_A = N$ e $N_B = 1$ [5].

Applicando questa teoria al caso di interesse si assume per semplicità che la molecola di un solvente sia della stessa dimensione di un monomero del polimero, e che ognuna di queste specie occupi un singolo sito di reticolo. Tramite la teoria FH si può definire una entropia configurazionale generalizzata per una collezione di m differenti tipi di polimeri ognuno con N_i monomeri di dimensione $r_i a^3$. In un tale sistema, il numero di siti di reticolo occupati sarà pari a $M = \frac{V}{a^3} = \sum_{i=1}^m r_i n_i N_i$, dove n_i è il numero di monomeri di tipo i nella soluzione. Per semplicità in seguito le dimensioni r_i saranno poste pari a 1. E' inoltre dunque possibile definire la frazione di monomeri di tipo i in unità di a^3 come $\phi_i = \rho_i a^3 = \frac{n_i N_i}{M}$. Seguendo le argomentazioni della originale derivazione della FH, si separa l'entropia configurazionale $S = k_B \ln \Omega$ in un termine che deriva dal grado di libertà traslazionale del primo monomero (o del centro di massa) di ogni polimero (Ω_{CM}), e in un termine che tiene conto del grado di libertà conformazionale per ogni polimero, fissando il centro di massa (Ω_{conf}) in modo tale che $\Omega = \Omega_{CM} \Omega_{conf}$. Il termine traslazionale corrisponderà al numero di modi di riarrangiare $\sum_i n_i$ monomeri, quando ognuno di essi è il primo monomero di un polimero.

$$\Omega_{CM} = \frac{M!}{(M - \sum_{i=1}^m n_i)! \prod_{i=1}^m n_i!} \quad (3)$$

Per quanto riguarda il termine conformazionale, si deriva una formula per ogni monomero i con z monomeri vicini sul sito di reticolo:

$$\Omega_{conf} = (z-1)^{M_1} \prod_{l=0}^{M_2-1} \left(1 - \frac{\sum_{i=1}^m n_i + l}{M} \right) = \frac{(z-1)^{M_1}}{M^{M_2}} (M - \sum_{i=1}^m n_i)! \quad (4)$$

dove $M_1 = M_2 = \sum_i n_i(N_i - 1)$ è il numero totale di siti di reticolo che possono essere riempiti dopo il posizionamento del primo monomero. Combinando le due formule e inoltre modificandole per tenere conto di una possibile violazione del volume escluso si arriva alla seguente espressione:

$$\Omega = \Omega_{CM} \Omega_{conf} = \frac{M!(z-1)^{\sum_{i=1}^m n_i(N_i-1)}}{\prod_{i=1}^m n_i! M^{\sum_{i=1}^m n_i(r_i N_i - 1)}} \quad (5)$$

Così, la entropia negativa configurazionale per sito di reticolo, in unità di $k_B T$, risulta

$$-s = -\frac{1}{M} \ln \Omega = \sum_{i=1}^m \frac{n_i}{M} \ln n_i + \{\text{termini lineari in } n_i\} = \sum_{i=1}^m \frac{\phi_i}{N_i} \ln \phi_i + \{\text{termini lineari in } \phi_i\}$$

trascurando termini lineari in ϕ_i che possono essere omessi dato che è dimostrabile che non hanno effetto sulla separazione di fase. Di conseguenza, definendo con ϕ_m , ϕ_c , ϕ_s e ϕ_w le frazioni di volume rispettivamente dei monomeri, dei controioni, del sale e dell'acqua, l'entropia di mescolamento risulta essere

$$-s = \frac{\phi_m}{N} \ln \phi_m + \phi_c \ln \phi_c + 2\phi_s \ln \phi_s + \phi_w \ln \phi_w. \quad (6)$$

In essa si nota come il sale sia moltiplicato per un fattore 2, dato che disciolto in soluzione esso presenta sia ioni negativi che positivi.

2.2 Energia interna di Flory Huggins

Essendo sistemi a volume costante, la grandezza estensiva che conviene analizzare per definire lo stato della soluzione è la energia libera di Helmholtz. L'altra grandezza che insieme all'entropia definisce l'energia libera di Helmholtz è la energia interna. Ricordando che M , il numero dei siti del reticolo occupati, corrisponde a $\frac{V}{a^3}$, la energia libera totale del sistema considerata in unità di $k_B T$ per sito di reticolo è

$$f = \frac{F a^3}{V k_B T} = -s + f_{int} \quad (7)$$

$-s$ è il contributo entropico alla energia derivato dalla teoria di Flory Huggins. f_{int} è modellata tenendo conto che in questo caso siamo interessati a poliamfoliti in cui quindi l'elettrostatica è il contributo entalpico dominante. Di fatto tuttavia f_{int} contiene anche contributi entropici dovuti alla connettività della catena.

In questo caso si andrà oltre l'approccio della teoria di Flory Huggins: questa infatti tratta una energia interna di mixing in termine di tre contributi di energia di accoppiamento (u_{AA} , u_{AB} e u_{BB}) tra siti di reticolo adiacenti occupati dalle due specie. Essi vanno a costituire la energia media di interazione per ogni specie con uno dei monomeri vicini tramite una somma pesata sulle frazioni di volume; ad esempio, per la specie A, l'energia media di interazione risulta $U_A = u_{AA}\phi_A + u_{AB}\phi_B$, e in simil modo per la specie B. Ogni sito di reticolo ha z elementi più vicini, in cui z

è il numero di coordinazione. Tenendo conto di questo numero e del fatto che ogni interazione di coppia è contata due volte, si arriva alla seguente formula:

$$U = \frac{zU_A M \phi_A}{2} + \frac{zU_B M \phi_B}{2} = \frac{zM}{2}(u_{AA}\phi^2 + 2u_{AB}\phi(1-\phi) + u_{BB}(1-\phi)^2) \quad (8)$$

In cui nell'ultimo membro è stato posto $\phi_A = \phi$ e di conseguenza $\phi_B = 1 - \phi$. E' possibile ricavare l'energia totale prima del rimescolamento:

$$U = \frac{zM}{2}(u_{AA}\phi + u_{BB}(1-\phi)) \quad (9)$$

e di conseguenza si conosce il cambiamento di energia in seguito al mescolamento per sito di reticolo (quantità intensiva). E' possibile definire il parametro di interazione della teoria di Flory, $\chi = \frac{z}{2} \frac{2u_{AB} - u_{AA} - u_{BB}}{kT}$, e si può riscrivere la variazione di energia per sito di reticolo in funzione del parametro di interazione.

$$\Delta u_{mix} = \chi \phi(1-\phi)kT \quad (10)$$

In queste formule u_{AA} , u_{AB} e u_{BB} sono le energie di campo medio per sito di reticolo che quantificano la forza delle interazioni tra unità di polimero e unità di solvente, coppie di unità di polimero e coppie di unità di solvente rispettivamente. La grandezza di χ quantifica il costo energetico di avere siti adiacenti al sito del polimero occupati da particelle di solvente. Infatti, per $\chi > 0$, l'interazione tra solvente e polimero è bassa, il polimero si trova in un solvente povero che favorisce la separazione di fase e inibisce il rimescolamento delle due fasi. Dunque vi sarà un valore critico di χ , sopra cui il termine energetico della energia libera supererà il termine entropico così che la energia libera di mescolamento avrà una regione in cui sarà concava. La instabilità risultante verrà risolta tramite una separazione di fase, una fase povera di polimeri che coesisterà con una ricca di polimeri. E' stato stabilito che soluzioni acquose sono solventi poveri, poco efficaci, e quindi stimolanti la separazione di fase ($\chi > 0$) per IDPs che sono ricche di amminoacidi polari come G, Q, N e S. Questa trattazione tuttavia non è adatta per descrivere soluzioni in cui i polimeri siano caratterizzati da forti interazioni elettrostatiche, quindi a lungo range e non solo a corto range. Overbeek e Vorn hanno esteso il formalismo di Flory Huggins per descrivere la separazione di fase in presenza di interazioni elettrostatiche.

2.3 Energia elettrostatica con Random Phase Approximation (RPA)

La trattazione dell'interazione elettrostatica a cui si fa riferimento è quella della RPA, e dunque la densità della particella è assunta essere omogenea per permettere un calcolo basato solamente su correlazioni a due corpi di fluttuazioni di densità. In maniera simile l'energia libera elettrostatica è approssimata da un integrale gaussiano sulle fluttuazioni di cariche. Si terrà conto di poliamfoliti quasi neutri in soluzione acquosa, con piccoli controioni e sale. Ogni poliamfolite consiste di N monomeri (residui di amminoacidi), con cariche $\sigma_i = \sigma_1, \sigma_2, \dots, \sigma_N$, dove σ_i è in unità della

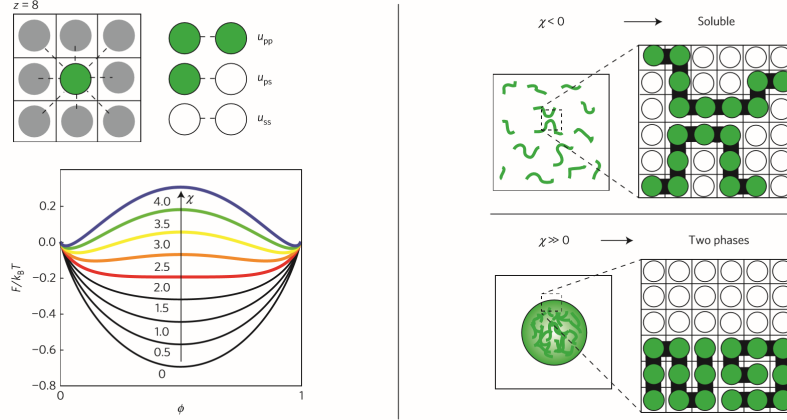


Figura 3: Separazione di fase dal punto di vista della Flory Huggins Theory: al crescere di χ , l'interazione tra le due specie considerate, il solvente e il polimero, diminuisce e il mescolamento viene inibito. Le due specie si disporranno dunque sul reticolo evidenziando una fase insolubile in soluzione acquosa, e l'energia libera in funzione della concentrazione esibirà una concavità.

carica elettronica e e può quindi assumere valori $+1$, -1 , 0 . I controioni e il sale sono monovalenti in questa trattazione. Le densità dei monomeri, dei controioni e del sale sono denotate con ρ_m , ρ_c e ρ_s e il numero dei controioni è considerato tale da bilanciare la carica netta dei monomeri, se diversa da zero. Tramite la RPA si arriva a un termine di energia interna elettrostatica:

$$fel = \frac{1}{2} \int_0^\infty \frac{d^3(ka)^3}{(2\pi)^3} \{ \ln[\det(1 + \hat{G}_k \hat{U}_k)] - Tr(\hat{\rho} \hat{U}_k) \} \quad (11)$$

dove \hat{G}_k e \hat{U}_k e $\hat{\rho}$ sono matrici $(N + 2) \times (N + 2)$. \hat{G}_k esprime una correlazione di densità senza effetti elettrostatici e \hat{U}_k è la correlazione per interazioni Coulombiane, entrambe definite nello spazio reciproco k . Il termine logaritmico che contiene queste correlazioni è il risultato di integrali gaussiani standard nella RPA. Il secondo termine, la traccia, sottrae la energia elettrostatica propria per tutte le densità di cariche per eliminare la divergenza non fisica del primo termine per $k \rightarrow \infty$. La matrice densità è una matrice diagonale, con I_N la identità N dimensionale e $\hat{\rho}_I$ una matrice diagonale 2×2 per gli ioni positivi e negativi monovalenti.

$$\hat{\rho} = \begin{bmatrix} (\frac{\rho_m}{N}) \hat{I}_N & 0 \\ 0 & \hat{\rho}_I \end{bmatrix} \quad (12)$$

\hat{G}_k , la matrice di correlazione, combina la correlazione monomero-monomero per una catena gaussiana e la matrice di densità per gli ioni monovalenti.

$$\hat{G}_k = \begin{bmatrix} (\frac{\rho_m}{N}) \hat{G}_M(k) & 0 \\ 0 & \hat{\rho}_I \end{bmatrix} \quad (13)$$

Dove $\hat{G}_M(k)$ è la matrice $N \times N$ per le catene gaussiane, con elementi $\hat{G}_M(k)_{ij} = e^{-\frac{(kb)^2 |i-j|}{6}}$, che corrispondono alla correlazione tra l' i -esimo e il j -esimo monomero

in una catena gaussiana. La matrice di interazione \hat{U}_k è la trasformata di Fourier del potenziale di Coulomb che si considera, ovvero con un cutoff fisico a corto range, sulla scala della dimensione di un monomero:

$$U(r) = \frac{l_B(1 - e^{-\frac{r}{b}})}{r} \quad (14)$$

Dunque nello spazio k diventa:

$$\hat{U}_k = \frac{4\pi l_B}{k^2[1 + (kb)^2]} |q\rangle \langle q| = \lambda(k) |q\rangle \langle q| \quad (15)$$

Dove $l_B = \frac{e^2}{4\pi\epsilon_0\epsilon_r k_B T}$ è la lunghezza di Bjerrum, e rappresenta la distanza tra due cariche interagenti elementari alla quale l'interazione elettrostatica risulta confrontabile con l'energia termica $k_B T$; dentro a l_B compaiono ϵ_0 , la permittività del vuoto, e ϵ_r la costante dielettrica relativa. $|q\rangle$ in equazione (15) invece è il vettore colonna per le cariche dei monomeri e degli ioni monovalenti, e $\langle q| = |q\rangle^T$ è il vettore riga corrispondente, con componenti $q_i = \sigma_i$ per $1 \leq i \leq N$, $q_{N+1} = 1$ e $q_{N+2} = -1$. Il determinante in equazione (11) può essere semplificato in :

$$\det(1 + \hat{G}_k \hat{U}_k) = 1 + \lambda(k) \langle q| \hat{G}_k |q\rangle = 1 + \lambda(k)(2\rho_s + \rho_c + \frac{\rho_M}{N} \langle \sigma| \hat{G}_M(k) |\sigma\rangle)$$

tramite l'identità di Sylvester, in cui $\lambda(k) = \frac{4\pi l_B}{k^2[1+(kb)^2]}$. Inoltre nel limite di per k tendente all'infinito vale $\ln[\det(1 + \hat{G}_k \hat{U}_k)] \rightarrow \lambda(k) \langle q| \hat{G}_k |q\rangle$. Questa proprietà permette di sostituire la traccia in equazione (11) con $\lambda(k) \langle q| \hat{G}_k |q\rangle$ per facilitare l'integrazione numerica per la *fel*.

Per le analisi che verranno fatte ci si riconduce a delle variabili adimensionali, quindi a una temperatura $T^* = \frac{b}{l_B} = \frac{4\pi\epsilon_0\epsilon_r k_B T b}{e^2}$. Inoltre, l'integrale di *fel* è riscritto in termini del numero d'onda ridotto $\tilde{k} = kb$. Infine, tramite le densità riscalate $\phi_i = \rho_i a^3$ si hanno delle espressioni puramente in termini di variabili adimensionali. Tramite queste semplificazioni si arriva alla formula definitiva per la *fel* :

$$fel = \int_0^\infty \frac{d\tilde{k} \tilde{k}^2}{4\pi^2} \left\{ \frac{1}{\eta} \ln[1 + \eta G(\tilde{k})] - G(\tilde{k}) \right\} \quad (16)$$

dove $\eta = (\frac{b}{a})^3$ è il rapporto tra il cubo della lunghezza del legame del polimero e la dimensione di una molecola d'acqua, e in cui

$$G(\tilde{k}) = \frac{4\pi}{\tilde{k}^2(1 + \tilde{k}^2)T^*} (2\phi_s + \phi_c + \frac{\phi_m}{N} \langle \sigma| \hat{G}_M(\tilde{k}) |\sigma\rangle). \quad (17)$$

2.4 Determinazione dei diagrammi di fase

L'equilibrio termodinamico è lo stato con un minimo nell'energia libera di Helmholtz. E' necessario studiare dunque $F_{mix}(\phi)$ e la sua relazione con $F_{\alpha\beta}(\phi)$, energia dello stato separato in due fasi, α e β . Al di sotto di una temperatura critica T_c compare un range di composizione in cui la energia libera in funzione della concentrazione è concava. Questo avviene tra due punti di flesso, all'interno dei quali $\frac{\partial^2 F}{\partial \phi^2} < 0$, il che rende le composizioni intermedie instabili: si tratta della regione spinodale, in cui anche la minima fluttuazione in concentrazione porta a una separazione di fase. Lo stato di equilibrio sarà definito da quello con la minore energia libera tra tutti gli stati in fase separata. Le fasi presenti sono determinate dalla regola della tangente comune: il sistema per minimizzare l'energia libera sceglie le composizioni che hanno la linea più bassa possibile che connette le due composizioni di fase. E' pertanto necessario eguagliare le derivate prime dell'energia libera rispetto alle concentrazioni nelle due composizioni di equilibrio. Per la determinazione dei confini di fase sarà dunque sufficiente risolvere il sistema di equazioni (18) e (19) nel range di concentrazioni desiderato e a temperatura fissata.

$$F(\phi_\alpha) - \phi_\alpha F'(\phi_\alpha) = F(\phi_\beta) - \phi_\beta F'(\phi_\beta) \quad (18)$$

$$F'(\phi_\alpha) = F'(\phi_\beta) \quad (19)$$

Le due fasi che coesistono all'equilibrio necessitano infatti di una uguaglianza dei potenziali: la F risulta nella pratica un Grand Potential di Gibbs, l'equazione (19) imporrà dunque la necessaria uguaglianza dei potenziali chimici all'equilibrio, e di conseguenza la (18) ridarà l'uguaglianza tra le energie libere in quanto tali delle due specie. La *miscibility gap* sottende la regione tra le due composizioni di minimo, e comprende regioni metastabili e instabili (regione binodale e regione spinodale). Per ogni temperatura fissata, sono quindi ricavabili le concentrazioni in cui coesistono due fasi separate, e, ponendole in un grafico sulle cui ordinate compare la temperatura, si ottiene la curva binodale nel diagramma di fase (che coincide con la curva di coesistenza all'equilibrio). Essa separa il diagramma in una regione a una fase e in una regione a due fasi, seppure comunque al di sotto della temperatura di transizione l'equilibrio stabile sia rappresentato dalle due fasi separate.

3 Separazione di fase per poliamfoliti

Questo metodo di determinazione del diagramma di fase verrà ora applicato a diverse sequenze proteiche in soluzione acquosa che danno avvio a una transizione liquido-liquido. Tramite una integrazione numerica è ottenibile infatti la $f = -s + f_{el}$ caratteristica di ogni sequenza a una data temperatura; definendo poi la $f(\phi)$ stessa e $\frac{\partial f}{\partial \phi}(\phi)$ come funzioni della concentrazione, e risolvendo il sistema di equazioni (18) e (19) per ogni temperatura in termini delle concentrazioni ϕ_α e ϕ_β , è possibile costruire il diagramma di fase.

In primo luogo sono oggetto di questa trattazione delle semplici soluzioni acquose prive

di sale ($\rho_s = 0$) di poliamfoliti, i quali consistono di n blocchi dalla carica alternata, di lunghezza $\frac{N}{L}$. Per semplicità verrà assunto che la lunghezza del legame del polimero b equivalga alla scala di lunghezza del solvente a , e che dunque ($b = a, \eta = 1$). Questi poliamfoliti presentano un n pari e dunque sono neutri globalmente, il che impone $\rho_c = 0$. L'obiettivo è riprodurre le curve binodali presentate per tali sequenze in un articolo pubblicato nel 2017 su "*Journal of molecular liquids*" [6]. Vengono considerate le sequenze in Figura 4. Per sequenze così regolari per quanto riguarda il

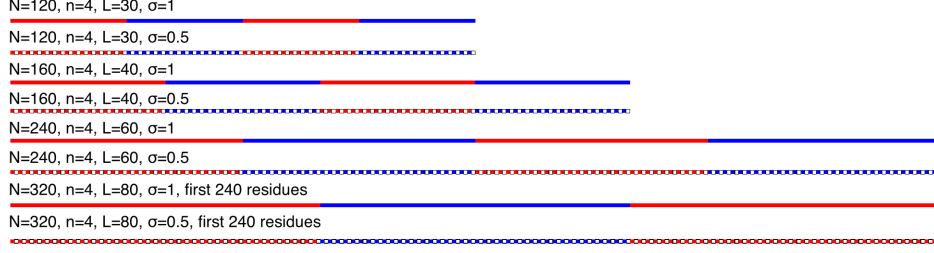


Figura 4: Poliamfoliti a blocchi di carica alternata, globalmente neutri. Ogni monomero è rappresentato da un quadratino, di colore rosso se di carica positiva, blu se negativa e bianco se neutro.

pattern di cariche, è possibile introdurre la densità di carica σ in un blocco e servirsi di questa grandezza nelle formule: infatti il termine $\langle \sigma | \hat{G}_M(k) | \sigma \rangle$, in equazione (17) è ottenibile focalizzandosi sui singoli blocchi, calcolando prima gli elementi diagonali della matrice $\hat{G}_M^{block}(k)$, che tengono conto delle correlazioni interne al blocco, poi gli elementi fuori dalla diagonale, che tengono conto delle correlazioni tra i blocchi. Il contributo dell'RPA per \hat{G}_M si calcola poi sommando tutti gli elementi $\hat{G}_M^{block}(k)$ con fattori di peso ± 1 alternati, ricavando eq. (20).

$$\langle \sigma | \hat{G}_M(k) | \sigma \rangle = N \left[\sigma \frac{1 + \zeta}{1 - \zeta} - \frac{1}{L} \frac{4\zeta}{(1 - \zeta)^2} \left(\frac{1 - \zeta^{L\sigma}}{1 + \zeta^{L\sigma}} \right) \right] + \frac{2\zeta}{(1 - \zeta)^2} \left(\frac{1 - \zeta^{L\sigma}}{1 + \zeta^{L\sigma}} \right)^2 (1 - (-1)^n \zeta^{N\sigma}) \quad (20)$$

In cui $\zeta = e^{-\frac{(\vec{k})^2}{6\sigma}}$.

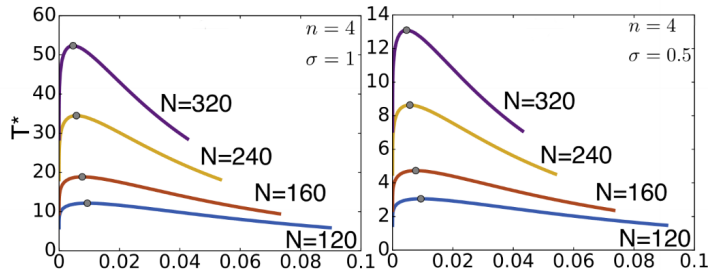


Figura 5: Diagramma di fase teorico, ottenuto tramite eq.(16) per poliamfoliti con $n = 4$, σ che assume valori 1 e 0.5 e N variabile come in legenda. Grafico tratto dall'articolo di Yi-Hsuan Lin, Julie D. Forman-Kay e Hue Sun Chan comparso nel 2017 su "*Journal of molecular liquids*" ([6]).

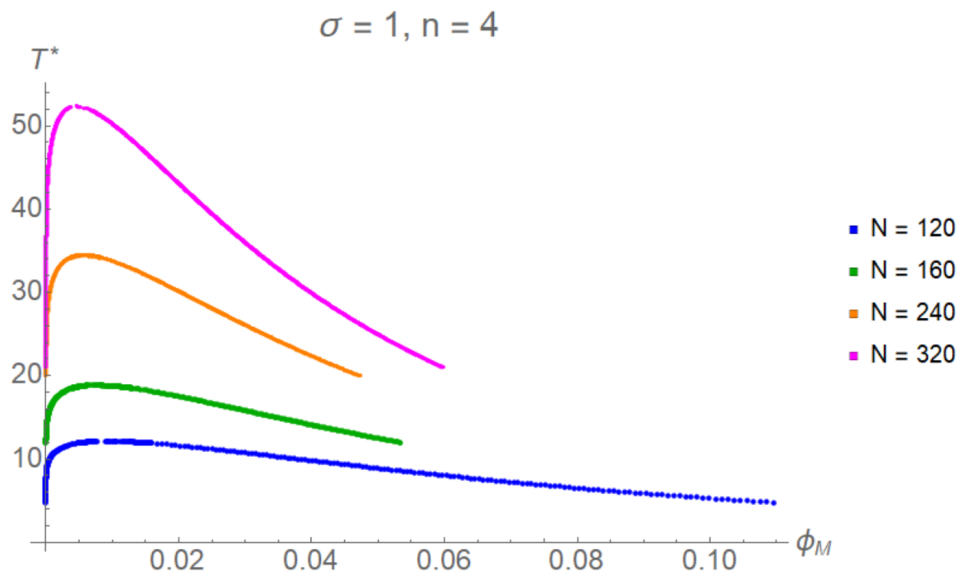


Figura 6: Riproduzione del diagramma di fase di fig.5, considerando tra le sequenze di fig.4 quelle con $\sigma = 1$.

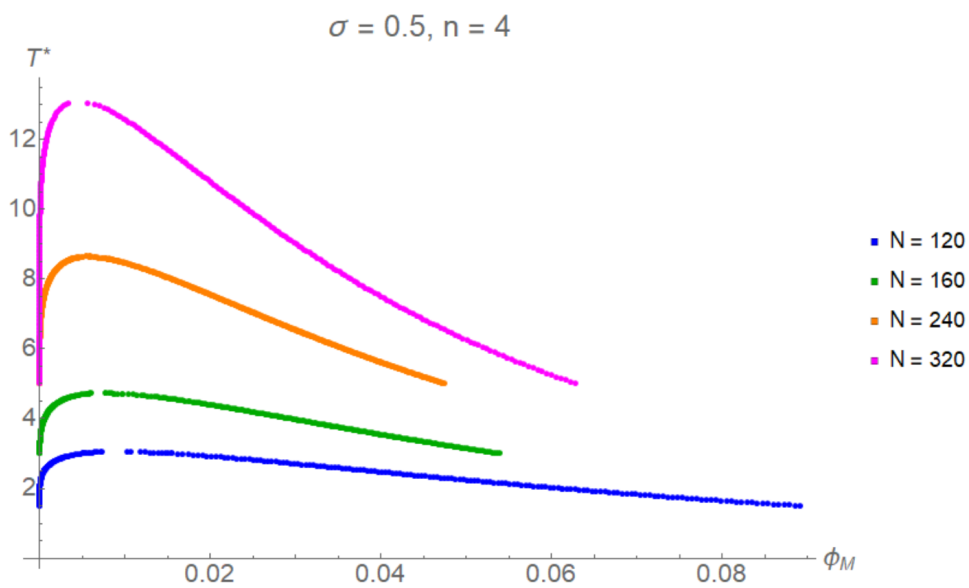


Figura 7: Riproduzione del diagramma di fase di fig.5, considerando tra le sequenze di fig.4 quelle con $\sigma = 0.5$

In secondo luogo invece viene applicata questa teoria alle precise sequenze di amminoacidi delle proteine $Ddx4^{N1}$ e alla sua sequenza modificata tramite uno scambio di cariche $Ddx4^{N1}CS(charged\ scrambled)$. A differenza di prima, non si tratta di poliamfoliti con blocchi dalla carica alternata, bensì sono sequenze caratterizzate da 236 amminoacidi, tra i quali 32 residui sono carichi positivamente e 36 negativamente; sono dunque sequenze quasi neutre ma non del tutto: in questa trattazione viene considerata anche la presenza dei controioni, la cui frazione è $\phi_c = \frac{|\sum_{i=1}^N \sigma_i \phi_m|}{N}$. Il sale verrà considerato assente anche in questa trattazione. Si assegna $\sigma_i = 1$ a ognuno dei 28 residui R e dei 4 residui K, $\sigma_i = -1$ a ognuno dei 18 residui D e dei 18 residui E, e $\sigma_i = 0$ a tutti gli altri residui. Si inseriscono questi valori di σ_i dentro a $|\sigma\rangle$ in equazione (17) (ora un vettore di lunghezza 236) per calcolare le f_{el} delle due sequenze e i corrispondenti diagrammi di fase. In questo caso infatti non si utilizzerà più eq. (20) che teneva conto di una ripetizione regolare di blocchi con la stessa densità di carica, bensì $\langle \sigma | \hat{G}_M(k) | \sigma \rangle$ corrisponderà a $\sum_{i,j=1}^N \sigma_i \sigma_j e^{-\frac{k^2|i-j|}{6}}$, in cui i e j definiscono la posizione lungo la sequenza dell'amminoacido, a cui corrisponde la carica i -esima o j -esima. Anche in questo caso l'obiettivo è riprodurre i diagrammi di fase per queste due sequenze proteiche così come mostrati in un articolo del 2016 pubblicato su *Physical review letters* [7].

```

Ddx4N1:
MGDEEDWEAII NPHMSSVVPFI FEKDRYSGEN GDNFNTPAS SSEMDDGPSR RDHFMSGFA SGNFNGNRDA GECNKRDNNTS
TMGGFGVGRS FGNRCFSNS FEDGDSSGFW EESSNDCEDN PTNNGFSKR GGYFDGNNSE ASGPYRRGGR GSFRCGRGGE
GLGSPNNDLD PDECMQRTGG LFGSRPVLVLS GTGNGDTSQS RSGSGSERGG YKGLNEEVIT GSGKNSWRSE ARGGES

Ddx4N1CS:
MGDEEDWEAII NPHMSSVVPFI FEKDRYSGEN GRNFNDTPAS SSEMDDGPSR RDHFMSGFA SGNFNGNRDA GECNKRDNNTS
TMGGFGVGRS FGNRCFSNS FEDGDSSGFW EESSNDCEDN PTNNGFSKR GGYFDGNNSE ASGPYRRGGR GSFRCGRGGE
GLGSPNNDLD PRECMQRTGG LFGSRPVLVLS GTGNGDTSQS RSGSGSERGG YKGLNEEVIT GSGKNSWRSE ARGGES

```

Figura 8: Sequenze di amminoacidi delle proteine $Ddx4^{N1}$ e $Ddx4^{N1}CS$. Il colore rosso evidenzia i residui carichi positivamente, l'azzurro i residui carichi negativamente e il giallo i residui aromatici.

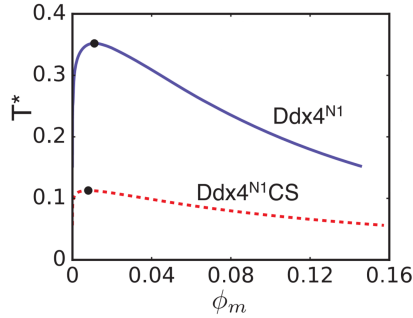


Figura 9: Diagrammi di fase teorici per le due proteine considerate, in condizioni di assenza di sale. Grafico tratto dall'articolo Yi-Hsuan Lin, Julie D. Forman-Kay e Hue Sun Chan, pubblicato nel 2016 su *Physical Review Letters* [7].

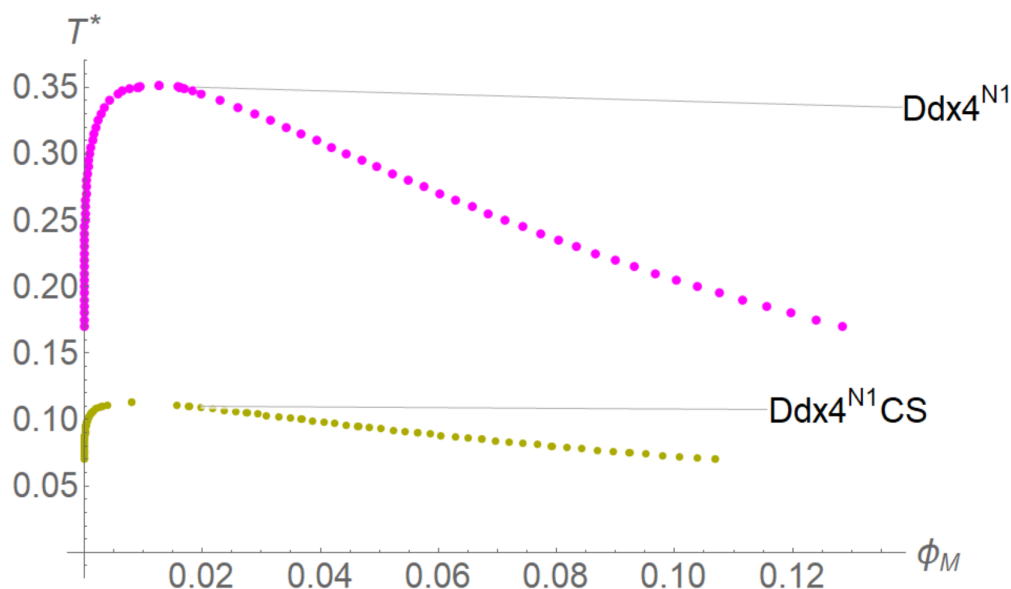


Figura 10: Riproduzione del diagramma di fase in Figura 9.

4 Dipendenza del diagramma di fase dalle sequenze di IDP

Le analisi seguenti si riferiscono invece a un articolo uscito nel 2015 su *Nature Materials* ([4]). In questo articolo sono riportati dei risultati sperimentali per la transizione di fase di un set di IDPs, le quali sono state progettate ad hoc tramite un approccio euristico: notando i pattern ricorrenti in alcune proteine di cui era conosciuto il comportamento di transizione liquido-liquido, sono state sintetizzate sequenze tramite tali pattern, ipotizzando che costruite in tale modo esibissero comportamenti simili alle proteine note. L'ambizioso obiettivo ultimo era infatti predire il comportamento di fase e ricostruire il modo in cui è codificato nella sequenza stessa.

Questo set di IDPs è caratterizzato in generale dal motivo PX_nG ricorrente (dove X è un amminoacido e n varia da 0 a 4), e da coppie di residui di carica opposta: sono state notate infatti queste proprietà in IDPs che sono conosciute per dare avvio a una transizione di fase che si manifesta in una UCST (Upper Critical Solution Temperature) [8], a differenza di altre categorie di IDPs che invece evidenziano una LCST (Lower Critical Solution Temperature: in questo caso, opposto a quello qui affrontato, la separazione di fase si ottiene aumentando la temperatura). In questo articolo viene analizzata la loro transizione di fase sperimentale per notare come evolva al variare dei parametri. E' stato verificato come il livello di idrofobia della unità ripetuta moduli anch'essa la UCST: per esempio, sostituire Val con il forte residuo idrofobico Tyr nel pattern $RGDSPYG$, ha come conseguenza un polimero

con una UCST non misurabile. Anche la concentrazione modula il *cloudpoint*, dato che è osservabile come la temperatura di transizione vari linearmente con il logaritmo della concentrazione dell'IDP.

In questa tesi vengono analizzati i risultati ottenuti nel caso delle temperature di transizione a concentrazione fissata ($50 \mu M$) di sequenze ricche di P (proline) e G (glicine), in funzione del numero di ripetizione dei blocchi.

A questo proposito viene utilizzato il modello teorico introdotto nel capitolo 2 per produrre le curve binodali delle ripetizioni di tre delle unità proteiche in Figura 11, con diversi valori di n fissati. Sono prese in considerazione le sequenze $(GRDGSPYG)_n$ (in Figura 11 la sequenza verde), con n pari a 14 e 20, $(GRGDSPYG)_n$ (in Figura 11 la sequenza magenta) con n pari a 18, 22, 32, e $(GRGDSAY)_n$ (in Figura 11 la sequenza nera) con n pari a 15 e 18.

Per poter istituire un paragone con i dati sperimentali in questo caso vengono abbandonate alcune approssimazioni, quali $\eta = 1$, al cui interno sono inseriti i valori sperimentali di $b = 3.8 \text{ \AA}$ (lunghezza del legame proteico), e $a = 2.8 \text{ \AA}$ (distanza fra due molecole d'acqua), e dunque risulta $\eta \simeq 2.5$. Viene inoltre considerato un ambiente cellulare che contempli una presenza di sale, in particolare una concentrazione di 150 mM di NaCl, che, diviso la concentrazione di H_2O , assumendo che vi sia una molecola d'acqua per sito di reticolo ($[H_2O] = \frac{1}{a^3 6.022 \times 10^{23} 10^3} = 75.9 \text{ M}$), viene convertito in una frazione $\phi_s = 0.002$ adimensionale, che è usata nel compiere i calcoli. Con un ragionamento analogo di conversione, i grafici sono infine espressi in unità di misura realistiche quanto alla concentrazione: le frazioni adimensionali di monomeri sono trasformate in concentrazioni molari di proteine ($[M]$) tramite la formula $\Phi_M[M] = \frac{[H_2O]\phi_m}{N}$.

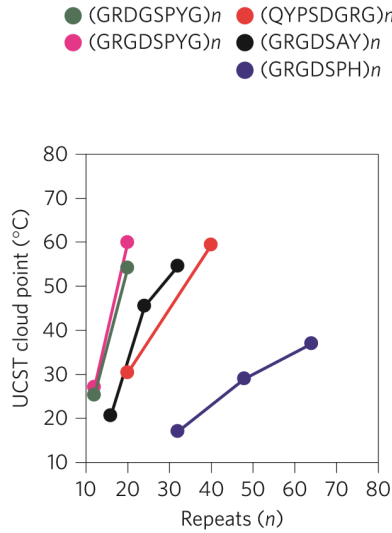


Figura 11: Temperature di transizione di un set di IDPs, estratte da dati sperimentali a una concentrazione fissata di $50 \mu M$, in funzione del numero di ripetizione dei blocchi.

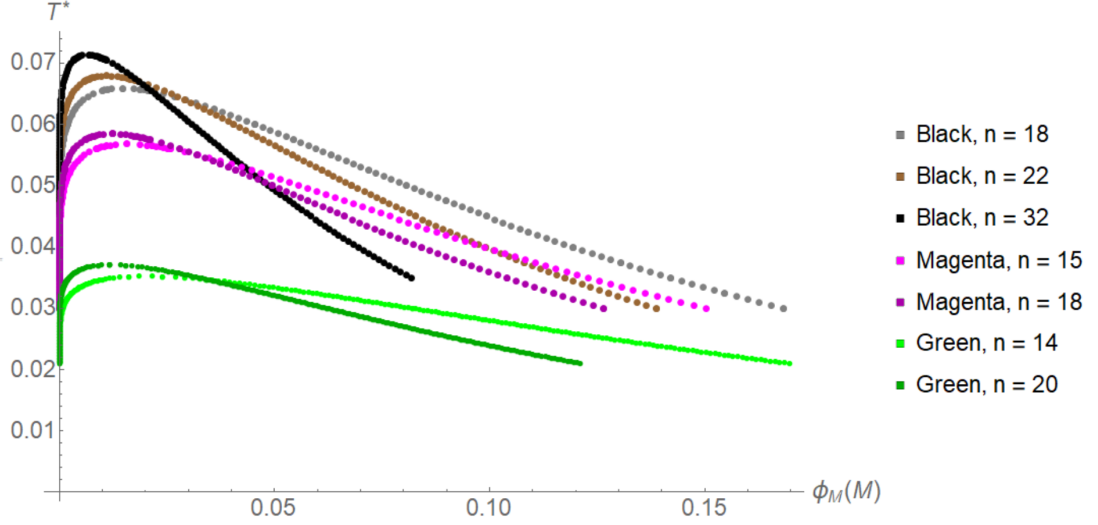


Figura 12: Diagrammi di fase di diverse ripetizioni delle unità proteiche indicate in Figura 11 coi colori magenta, verde e nero.

5 Conclusioni e prospettive

Dai risultati ottenuti dalla prima analisi condotta sui poliamfoliti neutri a blocchi dalla carica alternata si possono trarre alcune osservazioni. Il trend mostrato nei grafici delle figure 6 e 7 mostra che i poliamfoliti, a parità di n , caratterizzati da una maggiore N e da una maggiore frazione di cariche σ , danno avvio alla separazione di fase a una maggiore T^* . In particolare, la temperatura critica T_{cr}^* aumenta mentre la frazione di volume critica (concentrazione) $\phi_{cr}(= \phi_m \text{ alla } T_{cr}^*)$ diminuisce con l'aumentare di L (la lunghezza dei blocchi). Questo trend di $T_{cr}^* \rightarrow \infty$, $\phi_{cr} \rightarrow 0$ per $N \rightarrow \infty$ quando n è fissato è consistente con precedenti reports per piccoli poliamfoliti a due blocchi. Nel limite per $N \rightarrow \infty$ il primo termine in eq.(20) domina nell'integrazione per f_{el} , e in questa situazione il numero di blocchi n diventerebbe irrilevante e il sistema si comporterebbe come una soluzione di poliamfoliti a due blocchi con $N \rightarrow \infty$, in cui l'interazione elettrostatica è così forte da rompere l'RPA.

Per quanto riguarda l'analisi delle sequenze $Ddx4^{N1}$ e $Ddx4^{N1}CS$ è facile osservare come quest'ultima, con il suo pattern di cariche scambiate, abbia una T_{cr}^* pari a circa $\frac{1}{3}$ quella della sequenza non mutata, nel modello dell'RPA: viene evidenziata nuovamente l'importanza del pattern di cariche nel comportamento di separazione di fase, in accordo con i dati sperimentali secondo cui $Ddx4^{N1}$ dà avvio a una separazione di fase mentre $Ddx4^{N1}CS$ no. Tuttavia la concentrazione critica ϕ_{cr} di $Ddx4^{N1}$ è inferiore rispetto a quella della specie non mutata, questo può suggerire che $Ddx4^{N1}CS$ abbia una lunghezza maggiore di blocchi carichi ma una densità di carica

σ inferiore. Questa variazione rispetto al caso dei poliamfoliti a blocchi nel trend di (T_{cr}^*, ϕ_{cr}) rimanda al trend ricavabile invece per dei poliamfoliti strettamente alternati al diminuire di σ . Infatti, la prossimità delle cariche in $Ddx4^{N1}CS$ si manifesta in una σ effettiva molto più debole: la separazione di fase per $Ddx4^{N1}$ richiede dunque che sia mantenuta una densità di carica media ragionevolmente alta in una scala di lunghezza che includa diversi residui di amminoacidi.

Per quanto riguarda la riproduzione del diagramma di fase per IDPs che sono state analizzate sperimentalmente, il grafico in questione è in forma diversa rispetto a quello dell'articolo per cui non è subito evidente un confronto o un riscontro. E' da sottolineare che figura 11 faccia riferimento a una concentrazione fissata di $50\mu M$, molto più piccola della scala del range di concentrazioni di figura 12. In questo grafico, con l'allontanarsi delle concentrazioni dalla concentrazione critica, le sequenze nere sembrano via via avere delle temperature che scendono al di sotto delle altre sequenze. Si auspicherebbe un comportamento del genere anche nel range desiderato, ovvero quello dei micrometri, e per analizzarlo conviene esprimere l'asse delle ascisse in scala logaritmica, come mostrato in figura 13, in cui la linea rossa segnala il valore di concentrazione corrispondente ai dati sperimentali di figura 12. Si nota prima di tutto come nella regione a basse concentrazioni sia rispettato il trend al variare delle ripetizioni della stessa sequenza: a un n maggiore corrisponde una temperatura di transizione maggiore. Sembra non essere rispettato invece il trend secondo cui le sequenze etichettate col colore nero avrebbero temperature di transizione in generale minori (o tutt'al più a volte uguali, per precisi valori di n) rispetto alle altre sequenze.

Qualitativamente sono dunque riscontrabili o meno delle analogie coi dati sperimentali, mentre quantitativamente, cercando di convertire oltre alle concentrazioni anche i valori delle temperature in unità realistiche, si incontrano alcuni problemi. Per avere una conversione rigorosa dalla T^* adimensionale alla T sperimentale in gradi centigradi è utilizzabile la formula $T = \frac{T^*}{\frac{4\pi\epsilon_0\epsilon_r k_B b}{e^2}}$; bisognerebbe tuttavia essere in possesso della particolare ϵ_r della proteina, che in generale dipende dalla sequenza stessa, e inoltre dipende dalla concentrazione stessa delle IDPs, tanto che un ulteriore campo di analisi potrebbe consistere nel vedere come variano i diagrammi di fase considerando una ϵ_r non più costante lungo tutta la trattazione bensì variabile. In generale si può osservare come per ottenere delle temperature consistenti in gradi K in questo caso servirebbe una ϵ_r dell'ordine delle unità.

Un'altra osservazione che si può apportare è che la sequenza possa essere dipendente dalle cariche parziali che gli atomi dello scheletro dell'amminoacido portano su di sé. E' infatti possibile pensare a tre atomi di *backbone* per ogni amminoacido (C, N, C'), ognuno dei quali è portatore di una carica parziale (non intera) dovuta anche al gruppo a lui collegato, seppure la carica totale dell'amminoacido resti sempre +1, 0 o -1. Si tratta comunque di una approssimazione: in realtà la carica parziale è un

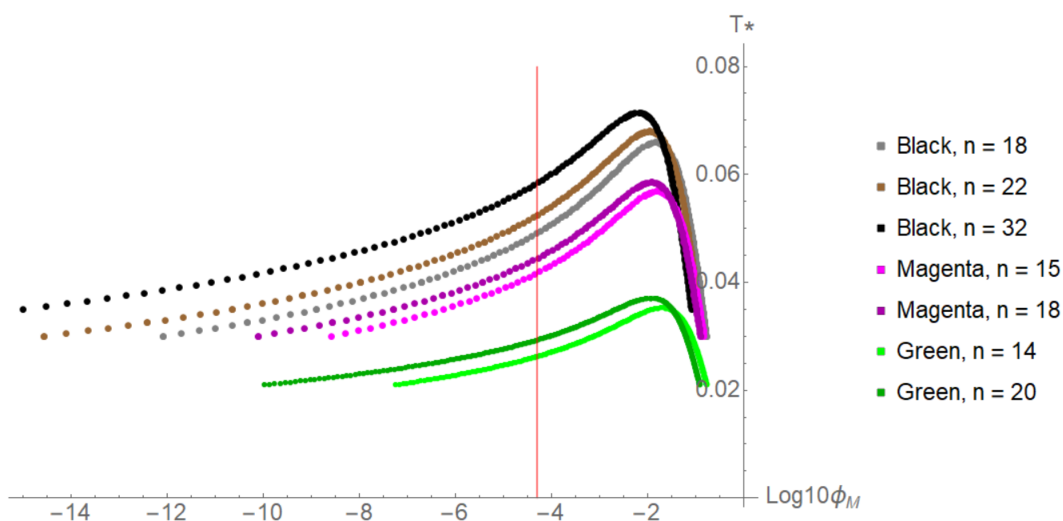


Figura 13: Diagrammi di fase di figura 12 esprimendo l'asse delle ascisse in scala logaritmica.

effetto quantistico ed esistono liste di cariche parziali "empiriche" a seconda degli usi. Si possono includere tali cariche parziali empiriche in questa trattazione, in quanto la presenza di proline o meno (sono assenti nelle sequenze nere) potrebbe essere determinante, dato che P, insieme a D e R presentano delle cariche parziali differenti da tutti gli altri amminoacidi caratteristici di queste sequenze. E' stata implementata dunque la presenza delle cariche parziali nel modello precedentemente utilizzato, facendo riferimento a dei valori di cariche parziali per amminoacidi presenti in letteratura [9], ottenendo sequenze di lunghezza $3N$, e utilizzando per sequenze così costruite un $b = 1.5 \text{ \AA}$; costruiti i diagrammi di fase relativi a tali sequenze, non si è però evidenziato un miglioramento tra la vicinanza delle temperature T^* per le sequenze nere e le altre sequenze. Si può dunque asserire che il modello utilizzato sia in questo caso troppo semplicistico per rendere conto dell'andamento preciso di questi dati sperimentali.

Senz'altro un'analisi più da vicino della sequenza degli amminoacidi potrebbe essere la strada per costruire un modello più realistico e meno riduttivo. Infatti la catena polipeptidica è essenzialmente una catena di dipoli e si potrebbe formalizzare una trattazione dell'RPA che tratti i componenti della catena come dipoli. In aggiunta a ciò, sono stati recentemente rivelati anche i contributi alla separazione di fase dovuti a interazioni a corto range, direzionali tra residui di arginine carichi positivamente e residui aromatici entro e tra le molecole di $Ddx4^{N1}$ [3]. Infatti, esplorando gli effetti della dipendenza dal sale nella transizione di $Ddx4^{N1}$, come avviene nell'articolo del 2016 su *Physical Review Letters* [7], i diagrammi di fase sotto differenti concentrazioni di sale non corrispondono ai valori sperimentali. Questa mancata corrispondenza significa che certe interazioni di fondo nella transizione del $Ddx4^{N1}$ che sono meno dipendenti dal sale non sono state prese in considerazione. Infatti, ci si aspetta che

le interazioni aromatiche contribuiscano significativamente alle proprietà di fase del $Ddx4^{N1}$, dato che il suo mutante con lo stesso pattern di cariche ma con meno residui aromatici non induce una separazione di fase. Queste osservazioni mostrano come interazioni *catione* – π e π – π (interazione aromatica, dovuta alla sovrapposizione degli orbitali p) possano giocare ruoli centrali nella separazione di fase. E' possibile dunque inserire nella teoria RPA un contributo di campo medio che tenga conto delle interazioni π , che sono note per essere a corto range e possono dunque essere formulate in prima approssimazione con un termine, dipendente dal sale, nel formalismo della teoria di Flory Huggins, così come avviene nell'articolo sopracitato. Sarebbe utile applicare questo metodo ad altri poliamfoliti e generalizzare la formulazione presente per incorporare interazioni non elettrostatiche specifiche della sequenza.

In generale dunque, sembra esserci un'interferenza gerarchica tra differenti tipi di interazione: un'interazione elettrostatica tra motivi carichi promuove interazioni a lungo range, con una specificità addizionale raggiunta tramite interazioni a corto range e direzionali tra dipoli (G, Q, N, S) o tra cariche positive e gruppi aromatici. Chiarire questo gioco di ruoli porterebbe a una comprensione biofisica molecolare delle sequenze che inducono la formazione di liquidi intracellulari.

Riferimenti bibliografici

- [1] Alexi V. Finkelstein and Oleg Ptitsyn. *Protein Physics, a course of lectures*. Academic press, 2002.
- [2] Johnny Habchi, Peter Tompa, Sonia Longhi, and Vladimir N. Uversky. Introducing protein intrinsic disorder. *Chemical Reviews*, 114:6561–6588, 2014.
- [3] Clifford P. Brangwynne, Peter Tompa, and Rohit V. Pappu. Polymer physics of intracellular phase transitions. *Nature Physics*, 11:899–904, 2015.
- [4] Felipe Garcia Quiroz and Ashutosh Chilkoti. Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nature materials*, 14:1164–1171, 2015.
- [5] Michael Rubinstein and Ralph H. Colby. *Polymer Physics*. Oxford University Press.
- [6] Yi-Hsuan Lin, Julie D. Forman-Kay, and Hue Sun Chan. Random-phase approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation for intrinsically disordered proteins. *Journal of Molecular Liquids*, 228:179–193, 2017.
- [7] Yi-Hsuan Lin, Julie D. Forman-Kay, and Hue Sun Chan. Sequence-Specific Polyampholyte Phase Separation in Membranellar Organelles. *Physical review letters*, 117:178101,1–178101,6, 2016.
- [8] Alex S. Holehouse and Rohit V. Pappu. Encoding phase transitions. *Nature materials*, 14:1083–1084, 2015.
- [9] D. S. Cerutti, J. E. Rice, W. C. Swope, and D. A. Case. Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model and Implicit Polarization. *The Journal of Physical Chemistry B*, 117:2328–2338, 2013.