

# UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea Triennale in Fisica

Tesi di Laurea

Il ruolo della sequenza amminoacidica nella  
separazione di fase di proteine

Relatore

Prof. Antonio Trovato

Laureando

Corrado Pancotti

Anno Accademico 2017/2018



# Indice

<b>1</b>	<b>Introduzione alle proteine</b>	<b>3</b>
1.1	Classificazione delle strutture proteiche . . . . .	4
1.2	Separazione di fase di proteine . . . . .	5
1.3	Le IDPs . . . . .	8
<b>2</b>	<b>Teoria di Flory-Huggins</b>	<b>9</b>
2.1	Entropia di mescolamento . . . . .	9
2.2	Energia interna di mescolamento . . . . .	11
2.3	Energia libera e parametro di interazione . . . . .	13
2.4	Determinazione dei diagrammi di fase . . . . .	13
<b>3</b>	<b>RPA (Random Phase Approximation)</b>	<b>17</b>
3.1	Entropia di configurazione . . . . .	17
3.2	Energia elettrostatica . . . . .	19
<b>4</b>	<b>Risultati per la proteina <math>Ddx4</math></b>	<b>21</b>
4.1	$Ddx4^{N1}$ e $Ddx4_{CS}$ . . . . .	21
4.2	$Ddx4$ in presenza di NaCl . . . . .	23
4.3	Permittività dipendente dalla concentrazione . . . . .	24
4.4	RPA + Flory-Huggins . . . . .	26
<b>5</b>	<b>Confronto con i dati sperimentali e conclusioni</b>	<b>28</b>

# Presentazione del lavoro di tesi

Argomento centrale del seguente elaborato è la fisica delle proteine. Nello specifico andremo a presentare alcuni modelli atti alla descrizione dei meccanismi e delle interazioni presenti tra le catene amminoacidiche.

Dopo una breve introduzione alle proteine in generale, seguendo le orme di un articolo pubblicato nel 2017 sul *Journal Molecular of Liquids* [4], tratteremo un particolare tipo di proteina, la *Ddx4<sup>N1</sup>*. Questa rientra nella più generale categoria delle proteine intrinsecamente disordinate (IDPs), cioè proteine che sono costituite principalmente da amminoacidi polari e carichi in cui non vi è una struttura tridimensionale ben definita e stabile.

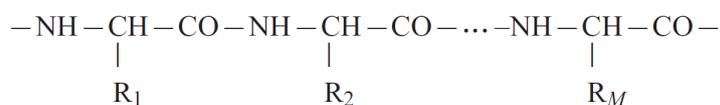
Vedremo come la teoria classica di Flory-Huggins non è sufficiente a descrivere il comportamento della nostra proteina; infatti, sebbene questo modello descriva molto bene il comportamento di proteine standard, sarà necessario ampliarlo per computare le interazioni coulombiane presenti tra i residui carichi.

Tramite la teoria RPA studieremo la proteina *Ddx4<sup>N1</sup>* originale e la sua proteina mutante, *Ddx4<sub>CS</sub>* (charge scrambled), mettendo in evidenza come la separazione di fase in soluzione acquosa dipenda fortemente dalla disposizione degli amminoacidi all'interno della sequenza. La *Ddx4*, come altre IDPs, è in grado di effettuare una separazione di fase liquido-liquido in cui coesistono due differenti concentrazioni in ambiente acquoso, processo attraverso il quale si formano delle goccioline che possono funzionare come componenti di organelli privi di membrana (tra cui nucleoli, corpi di Cajal e granuli di stress). Questi organelli svolgono ruoli importanti nella regolazione genica, nel processo di omeostasi e nel ciclo cellulare. Pertanto comprendere la biofisica della separazione di fase di IDPs ha rilevanza, oltre che prettamente biologica, anche di tipo medico. Obiettivo principale del lavoro di tesi sarà la riproduzione dei diagrammi di fase presenti nell' articolo di riferimento.

# Capitolo 1

## Introduzione alle proteine

Con il termine proteina si definisce, in generale, una macromolecola biologica costituita da catene di amminoacidi tenuti assieme da legami peptidici. Gli amminoacidi sono molecole organiche che nella loro struttura recano il gruppo funzionale amminico (-NH<sub>2</sub>) chiamato N-terminus e quello carbossilico (degli acidi carbossilici) (-COOH) denominato C-terminus. Si può pensare alla proteina come ad un insieme di unità di base, gli amminoacidi appunto, tenute insieme da un legame covalente che permette la formazione di un corpo unico. Le proteine sono delle vere e proprie macchine molecolari che svolgono molteplici ruoli negli organismi viventi. Tra i principali elenchiamo la catalisi di reazioni metaboliche, la comunicazione cellulare, il trasporto di altre molecole e le funzioni di sintesi, come ad esempio la replicazione del DNA. La grande varietà delle funzioni che queste possono svolgere si riflette in una elevata specificità di forma. Si può immaginare, per agevolare la comprensione, che esista una relazione chiave-serratura tra le proteine e le molecole con le quali esse interagiscono. Questo richiede necessariamente una definita struttura e una configurazione spaziale che identifica la proteina. Una conoscenza della struttura tridimensionale della proteina permette la comprensione delle funzioni della stessa. Cerchiamo quindi di capire più nel dettaglio come è fatta. Negli anni 50' il chimico britannico Frederick Sanger mostrò che la sequenza di residui amminoacidici è unica per ogni proteina <sup>1</sup>. Questa è costituita da una catena principale regolare alla quale sono allacciate altre catene laterali diverse ( $R_1, R_2, \dots, R_M$ ).



Il numero M di residui in una catena proteica può variare da poco più di una decina a molte migliaia. In natura esistono venti specie principali di residui amminoacidici, pertanto le possibili combinazioni di queste sono in numero elevatissimo. Tuttavia la posizione dei residui all'interno della catena è dettata dalla sequenza nucleotidica conservata nei geni.

E' bene dire che la proteina "operativa", cioè la proteina nel momento in cui svolge la sua funzione, si trova in una struttura strettamente specifica, struttura influenzata necessariamente dall'ambiente in cui questa vive. Infatti in base alle diverse condizioni ambientali è possibile dividere le proteine in tre differenti classi:

---

<sup>1</sup> un residuo è ciò che rimane dell'amminoacido libero dopo la polimerizzazione, cioè dopo il processo che unisce gli amminoacidi in un polimero, la proteina

1. Proteine fibrose. Sono proteine lunghe, dalla struttura semplice a forma di barra o filamento i cui gruppi laterali sono costituiti da elementi idrofobici. Si trovano principalmente sotto forma di aggregati in ambienti in cui vi è una scarsa presenza di acqua. La loro struttura è spesso caratterizzata da un elevato numero di legami idrogeno. E' molto regolare e mantenuta assieme soprattutto dalle forti interazioni che vi sono tra le varie catene.
2. Proteine di membrana. Sono proteine che risiedono principalmente in ambienti di membrana in cui è presente un basso contenuto d'acqua (sebbene una piccola porzione di queste è a contatto con l'acqua). La parte intra-membranosa è costituita da un gran numero di legami idrogeno che conferisce regolarità alla struttura, sebbene questa sia di dimensione ridotta rispetto alle proteine fibrose data la sottigliezza della membrana. Queste assieme alle fibrose svolgono funzioni di tipo strutturale.
3. Proteine globulari. Sono proteine caratterizzate da una forma sferica e sono generalmente solubili in acqua. A differenza delle proteine fibrose e di membrana, che hanno un ruolo prettamente strutturale, queste agiscono anche come enzimi, catalizzando reazioni chimiche, o come fattori di trascrizione. Svolgono poi anche azioni di trasporto di altre molecole.

La classificazione appena fatta è certamente approssimativa.

Esistono infatti proteine costituite da una "coda" fibrosa e da una "testa" globulare.

Siamo a conoscenza di più di centinaia di migliaia di sequenze proteiche e di circa una decina di migliaia di configurazioni spaziali (salvate all'interno del PDB, Protein Data Banks [2]).

Ciò che conosciamo delle strutture tridimensionali concerne solitamente proteine globulari. Le configurazioni spaziali note di proteine fibrose e membranose sono in numero assai ridotto. Questa differenza è da ricercare nel fatto che le proteine globulari sono solubili in ambiente acquoso, pertanto risulta relativamente semplice isolarle in molecole separate e osservare la loro struttura mediante moderne tecniche di microscopia ottica quali la cristallografia a raggi X o l'NMR (nuclear magnetic resonance).

## 1.1 Classificazione delle strutture proteiche

Le architetture proteiche, specialmente quelle che riguardano le proteine globulari, sono complesse e molto diverse tra loro. Per quanto concerne la loro struttura è bene dire che i principali responsabili di questa sono i legami che entrano in gioco. La forma tridimensionale della proteina è costituita principalmente da legami di tipo covalente, molto più robusti di quelli non covalenti, che tengono insieme gli amminoacidi costitutivi. La sequenza amminoacidica nella sua forma grezza, cioè come alternanza di residui molecolari, prende il nome di **struttura primaria**. La struttura primaria della proteina è tenuta insieme da legami peptidici che vengono formati durante il processo di biosintesi. La struttura primaria è determinata dal gene corrispondente alla proteina. Una specifica sequenza dei nucleotidi nel DNA è trascritta nell'mRNA che viene poi letto dai ribosomi nel processo di trascrizione.

La **struttura secondaria** di una proteina è la forma tridimensionale di segmenti locali della proteina. I due principali elementi della struttura secondaria sono la  $\alpha$ -elica e il foglietto  $\beta$ . Queste forme si costruiscono spontaneamente come fase intermedia prima del ripiegamento della proteina nella struttura terziaria. Le  $\alpha$ -eliche e i foglietti  $\beta$  vanno poi a formare una struttura globulare compatta durante quello che viene definito *fold*ing proteico.

Al termine del processo di ripiegamento, che avviene spontaneamente, si ottiene la **struttura terziaria** della proteina. La formazione *in vivo* della struttura nativa terziaria avviene durante

la biosintesi o immediatamente dopo. Tuttavia è possibile *in vitro* ottenere il ripiegamento di qualche catena proteica partendo da una situazione unfolded tramite appunto un processo di "rinaturazione" in cui si soddisfano le condizioni ambientali affinché la proteina riesca a raggiungere spontaneamente il suo stato nativo. Questa possibilità implica che la struttura spaziale della proteina sia determinata esclusivamente dalla sua sequenza amminoacidica e che perciò la proteina sia in grado di auto-organizzarsi [2]. Molte proteine, soprattutto di origine enzimatica, nella loro forma funzionale sono costituite dall'associazione di due o più monomeri che, dopo aver raggiunto la struttura terziaria, si uniscono in strutture super-globulari tenute assieme da legami di natura idrofobica. Questa configurazione prende il nome di **struttura quaternaria**. In figura mostriamo le varie strutture proteiche.

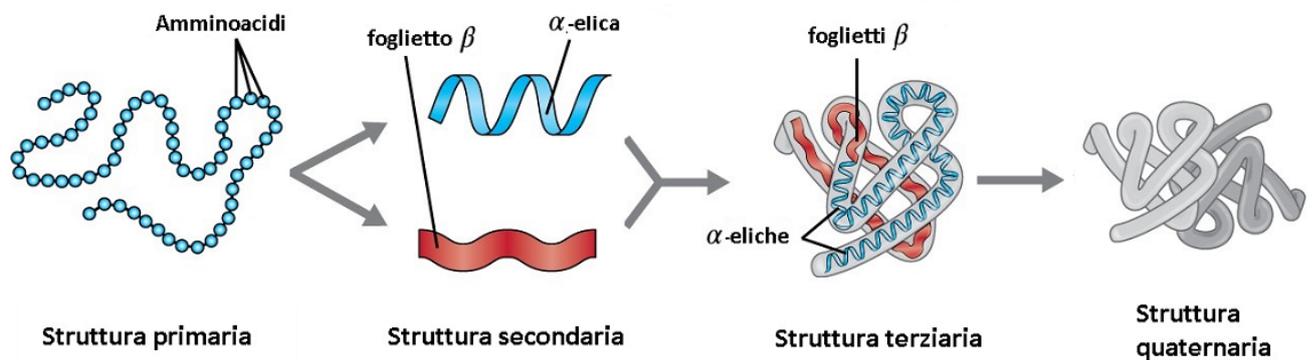


Figura 1. Livelli di organizzazione strutturale: struttura primaria (sequenza amminoacidica), struttura secondaria (conformazioni  $\alpha$ -elica e foglietto beta), struttura terziaria ripiegata di una catena e struttura quaternaria.

Abbiamo detto che la struttura di una proteina determina il suo comportamento e le sue specifiche funzioni. Tuttavia il contrario, cioè che le funzioni svolte da una proteina determinino la sua forma tridimensionale, non è in generale vero. Infatti, nonostante in alcuni casi si siano osservate delle correlazioni tra il comportamento della proteina e la sua struttura 3D, è possibile affermare che questa correlazione esista soltanto per una classificazione strutturale superficiale, cioè quella connessa alle condizioni ambientali. Ad esempio le proteine che svolgono funzioni di tipo strutturale, quelle che costituiscono capelli o fibre sono proteine fibrose, i recettori sono proteine membranose.

A conferma di quanto sopra affermato esistono però siti attivi strutturalmente diversi che svolgono le stesse funzioni. Quindi non è possibile affermare che il comportamento di una proteina determini la sua complessa architettura spaziale [2].

## 1.2 Separazione di fase di proteine

Nella prima sezione abbiamo effettuato una generica classificazione delle proteine in base alle condizioni ambientali in cui vivono. Ciò significa che queste sono presenti in soluzione a una particolare concentrazione e si trovano sotto determinate condizioni di pressione e a date temperature. Al fine di comprendere i meccanismi che stanno alla base del loro comportamento è possibile pertanto utilizzare un approccio di tipo termodinamico.

Come vedremo più nel dettaglio nei capitoli successivi, un modello che ben descrive il comportamento delle proteine è quello di Flory-Huggins in cui si va a computare l'energia libera di Helmholtz del sistema studiando le interazioni che avvengono tra gli amminoacidi costituenti la proteina.

Inoltre si tiene conto anche dei contributi che derivano dall'entropia di configurazione, cioè dall'entropia derivante dal mescolamento di specie chimiche differenti. Tuttavia esistono anche amminoacidi carichi positivamente e altri carichi negativamente (che vedremo essere presenti soprattutto in un particolare tipo di proteine), dunque tratteremo anche della più moderna teoria della RPA (Random Phase Approximation) che cerca di computare l'energia interna dovuta alle interazioni elettrostatiche presenti nelle catene amminoacidiche.

Immaginiamo ora di avere a che fare con una soluzione acquosa di volume  $V$  all'interno della quale vi sono due specie chimiche diverse tra loro, che chiameremo A e B, aventi una differente concentrazione che possiamo chiamare  $\phi_A$  e  $\phi_B$ . Le concentrazioni altro non sono che frazioni di volume di una determinata sostanza:

$$\phi_A = \frac{V_A}{V_A + V_B} \quad (1.1)$$

$$\phi_B = \frac{V_B}{V_A + V_B} = 1 - \phi_A \quad (1.2)$$

Indipendentemente dal metodo che si utilizza per lo studio del problema ciò che interessa capire è, a grandi linee, quali siano le condizioni di stabilità di un siffatto sistema.

Lo stato miscelato risulta stabile o tende a separarsi in due fasi diverse?

L'equilibrio in termodinamica è associato allo stato del sistema con la minima energia libera. Quando si ha a che fare con sistemi a volume costante, la grandezza estensiva che conviene considerare è l'energia libera di Helmholtz. Questa è data da un termine di energia interna e da un termine entropico

$F = U_{mix} - T\Delta S_{mix}$ . In generale sarà funzione delle concentrazioni  $\phi_A$  e  $\phi_B$ .

Supponiamo ora che il risultato del mescolamento di queste differenti sostanze formi un liquido omogeneo di concentrazione complessiva  $\phi_0$  ( $\phi_A < \phi_0 < \phi_B$ ) e consideriamo la sua energia libera  $F_{mix}(\phi_0)$ .

In generale possiamo esprimere la concentrazione complessiva come una combinazione lineare delle due

$$\phi_0 = f_A\phi_A + f_B\phi_B \quad (1.3)$$

La stabilità di questo sistema dipende dal fatto che l'energia libera dello stato miscelato  $F_{mix}(\phi_0)$  sia maggiore o minore di quella dello stato costituito dalle due fasi separate,  $F_{AB}(\phi_0)$ .

I coefficienti che compaiono nell'equazione 1.3 rappresentano le frazioni di concentrazione dei due materiali. In formule:

$$f_A = \frac{\phi_B - \phi_0}{\phi_B - \phi_A} \quad (1.4)$$

$$f_B = \frac{\phi_0 - \phi_A}{\phi_B - \phi_A} \quad (1.5)$$

L'energia libera dello stato costituito dalle due fasi separate, che è la media pesata delle energie libere di ogni materiale nelle rispettive fasi, è invece data da:

$$F_{AB}(\phi_0) = f_AF_A + f_BF_B \quad (1.6)$$

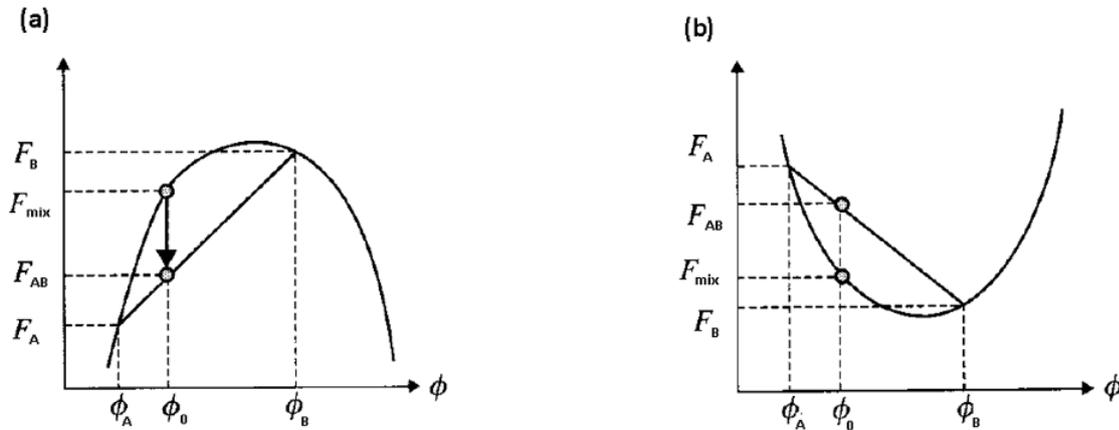


Figura 2. Dipendenza dell'energia libera dalla concentrazione. Esempi di instabilità a) e di stabilità locale b). La stabilità è determinata dalla derivata seconda dell'energia rispetto alla composizione.

In figura mostriamo le configurazioni possibili per l'energia libera.

Come è facile osservare la curvatura determina la stabilità locale. Se la dipendenza dell'energia libera dalla composizione è concava (figura (a)) il sistema può spontaneamente abbassare la propria energia separandosi in due fasi diverse, poiché  $F_{AB}(\phi_0) < F_{mix}(\phi_0)$ . Al contrario, se è convessa (figura (b)), ogni stato mescolato ha un'energia libera minore di ogni possibile stato separato, rendendo lo stato di mixing localmente stabile.

Possiamo quindi scrivere le condizioni di stabilità in termini della derivata seconda dell'energia libera rispetto alla concentrazione:

$$\frac{\partial^2 F}{\partial \phi^2} < 0 \quad \text{instabile} \quad (1.7)$$

$$\frac{\partial^2 F}{\partial \phi^2} > 0 \quad \text{localmente stabile} \quad (1.8)$$

Approfondiremo la separazione di fase di proteine nei prossimi capitoli quando focalizzeremo la nostra attenzione su una proteina specifica.

Per comprenderne il comportamento sarà necessaria un'analisi più sofisticata e accurata.

Ciò che ci interessa sottolineare qui è l'importanza che riveste la separazione di fase nello studio delle funzioni che la proteina può svolgere. Vi sono ad esempio alcune particolari proteine che possono dar luogo a separazioni di fase liquido-liquido, a seguito di interazioni con l'ambiente o con altre sostanze, in cui coesistono fra loro due fasi solubili in acqua, caratterizzate da diverse concentrazioni; Questo significa che a separazione avvenuta la proteina subisce un importante cambiamento nelle proprietà e nelle funzioni che può svolgere. La fase risultante è molto importante per la vita di alcune cellule; Si è recentemente osservato infatti che la formazione di gocce derivanti da una separazione di fase liquido-liquido è necessaria per la costituzione di alcuni organelli che svolgono importanti funzioni biologiche.

### 1.3 Le IDPs

Nella prima sezione si è cercato di mettere in evidenza l'importanza della struttura spaziale delle proteine e di come questa dia importanti informazioni sulle funzioni biologiche coinvolte. Per lungo tempo si è pensato che una struttura terziaria ben definita e specifica fosse il requisito essenziale per il funzionamento di una proteina, fino a quando negli anni 90' alcuni esperimenti cominciarono a mettere in luce l'esistenza di proteine differenti, che per le loro caratteristiche furono chiamate IDP, acronimo che sta per *intrinsically disordered protein*. Una proteina intrinsecamente disordinata è una proteina che manca di una specifica struttura tridimensionale. Tra queste sono incluse proteine che vanno dalle completamente prive di struttura alle parzialmente strutturate [3]. Solo una piccola percentuale delle proteine conosciute possiede una struttura completamente disordinata. Nonostante la mancanza di una struttura stabile, le IDPs sono in elevato numero e si differenziano dalle normali proteine per funzionalità, sequenze, interazioni e evoluzione. E' ormai ampiamente accettato dalla comunità scientifica che queste proteine esistano sotto forma di insiemi di strutture simili. Possiamo quindi pensarle come un insieme di differenti configurazioni che la proteina si trova ad assumere in base alle interazioni in cui è coinvolta e alle condizioni ambientali in cui vive. Le IDPs sono generalmente caratterizzate da una maggior presenza di amminoacidi polari e carichi piuttosto che di amminoacidi idrofobici. Queste caratteristiche stanno alla base della struttura disordinata, infatti la scarsità di elementi idrofobici le rende solubili in acqua e la grande presenza di elementi polari e carichi, nel caso in cui vi sia una carica totale netta, favorisce le interazioni elettrostatiche repulsive che risultano da residui ugualmente carichi. Dunque un modo per distinguere le proteine disordinate da quelle ordinate è quello di andare a studiare la composizione amminoacidica. Vengono infatti considerati come promotori di disordine i seguenti amminoacidi, idrofili e carichi, A, R, G, Q, S, P, E e K, mentre sono caratterizzati come promotori di ordine, W, C, F, I, Y, V, L, and N che sono ovviamente idrofobici e privi di carica. I restanti amminoacidi si possono invece trovare in entrambi i tipi di proteine.

In questo elaborato ci concentreremo su di un particolare tipo di proteina intrinsecamente disordinata chiamata *Ddx4<sup>N1</sup>*. Questa, come vedremo, in ambiente acquoso è in grado di effettuare una separazione di fase liquido-liquido. Rimandiamo però lo studio dettagliato di questi meccanismi al terzo capitolo, poiché solo allora avremo acquisito gli strumenti necessari alla comprensione.

## Capitolo 2

# Teoria di Flory-Huggins

La teoria di Flory-Huggins, si trova documentata in [7], è un modello matematico atto a spiegare la termodinamica delle soluzioni di polimeri.

E' una teoria di campo medio che bene spiega la separazione di fase per proteine idrofobiche in cui le interazioni tra i monomeri sono prevalentemente di contatto e a corto raggio d'azione. Infatti sebbene il modello riesca a computare sensatamente il contributo energetico dovuto all'entropia di mescolamento e il contributo energetico di contatto, non tratta delle interazioni elettrostatiche che invece sono presenti nelle IDPs. Quindi verrà esposta la teoria classica di Flory-Huggins e in seguito verranno attuate alcune modifiche al fine di adattarla allo studio della proteina in esame.

### 2.1 Entropia di mescolamento

Abbiamo visto nella sezione (1.2) come funziona a grandi linee la separazione di fase e quali sono i parametri da considerare al fine di comprendere questo meccanismo. Vedremo ora nel dettaglio come andare a computare l'entropia di mescolamento. Richiamiamo la situazione precedentemente illustrata. Consideriamo in via del tutto generale il mescolamento di due specie A e B come mostrato in figura e assumiamo per il momento che le due sostanze mescolate formino un liquido omogeneo di fase unica.

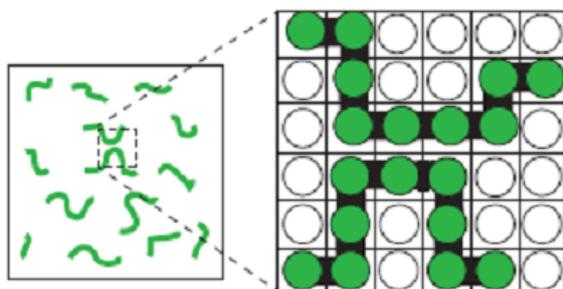


Figura 3. Mescolamento omogeneo di due sostanze. Si forma una soluzione avente una fase unica.

Per liquido omogeneo si intende appunto un miscuglio uniforme in cui tutte le componenti sono perfettamente mischiate su scala molecolare. Assumiamo anche che il volume non cambi durante il mescolamento delle due specie nella nostra soluzione.

Quindi il volume totale sarà  $V_A + V_B$ . Le frazioni di volume sono già state definite nelle equazioni (1.5) e (1.6).

Queste mostrano il mix di due tipi di molecole di eguale volume molecolare, tuttavia un simile

mescolamento è possibile anche se una o entrambe le specie sono polimeri.

Per il caso generale andiamo quindi a definire con  $v_0$  il volume unitario, cioè il volume del sito reticolare.

Di conseguenza una molecola di specie A avrà un volume

$$v_A = N_A v_0 \quad (2.1)$$

e una molecola di specie B

$$v_B = N_B v_0 \quad (2.2)$$

Laddove  $N_A$  e  $N_B$  rappresentano il numero di siti reticolari occupati dalle rispettive molecole. In base al numero di  $N_i$ , possiamo distinguere tre situazioni differenti:

- **soluzioni regolari.** Sono miscugli di specie aventi piccoli volumi molecolari,  $N_A = N_B = 1$
- **soluzioni polimeriche.** Sono miscugli di macromolecole identiche,  $N_A = N \gg 1$  e  $N_B = 1$
- **miscele polimeriche.** Sono miscugli di macromolecole di differenti specie chimiche,  $N_A \gg 1$  e  $N_B \gg 1$ .

Il sistema di volume  $V_A + V_B$  occupa quindi

$$M = \frac{V_A + V_B}{v_0} \quad (2.3)$$

siti reticolari, mentre tutte le molecole di specie A ne occupano

$$\frac{V_A}{v_0} = M \phi_A \quad (2.4)$$

L'entropia  $S$  è determinata dal prodotto della costante di Boltzmann  $k_B$  con il logaritmo naturale del numero di modi  $\Omega$  di disporre le molecole nel sito reticolare:

$$S = k_B \ln \Omega \quad (2.5)$$

Il numero di stati traslazionali di una singola molecola non è altro che il numero di posizioni indipendenti che una molecola può assumere nel reticolo ed è quindi uguale al numero di siti reticolari. Dunque in un miscuglio omogeneo di A e B, ogni molecola ha

$$\Omega_{AB} = M \quad (2.6)$$

dove  $M$  è dato dall'equazione 2.3.

Invece il numero di stati  $\Omega_A$  di ogni molecola di specie A è uguale al numero di siti reticolari occupati da questa, ovvero

$$\Omega_A = M \phi_A \quad (2.7)$$

Mettendo insieme le due equazioni sopra scritte otteniamo la variazione di entropia nel mescolamento per una singola molecola di specie A:

$$\Delta S_A = k_B \ln \Omega_{AB} - k_B \ln \Omega_A = k_B \ln \left( \frac{\Omega_{AB}}{\Omega_A} \right) = -k_B \ln \phi_A \quad (2.8)$$

E' da notare che la variazione di entropia sopra scritta è sempre positiva, questo perché la frazione di volume è sempre minore dell'unità. Ovviamente è possibile ottenere una relazione analoga per una molecola di specie B. Per calcolare l'entropia di mescolamento totale dobbiamo sommare il contributo derivante da ogni molecola. Perciò avremo:

$$\Delta S_{mix} = n_A \Delta S_A + n_B \Delta S_B = -k_B (n_A \ln \phi_A + n_B \ln \phi_B) \quad (2.9)$$

Con  $n_A = \frac{M\phi_A}{N_A}$  e  $n_B = \frac{M\phi_B}{N_B}$  numero di molecole di ogni specie.

Scriviamo infine un'espressione per l'entropia di mescolamento per sito reticolare:

$$\Delta S'_{mix} = \frac{\Delta S_{mix}}{M} = -k_B \left( \frac{\phi_A}{N_A} \ln \phi_A + \frac{\phi_B}{N_B} \ln \phi_B \right) \quad (2.10)$$

Ovviamente l'espressione cambierà se avremo a che fare con soluzioni regolari piuttosto che con soluzioni o miscele di polimeri.

## 2.2 Energia interna di mescolamento

In questo paragrafo vedremo come computare l'energia interna di mescolamento in modo da poter calcolare l'energia libera di Helmholtz e di conseguenza costruire i diagrammi di fase.

Le interazioni tra due specie possono essere attrattive o repulsive. La teoria di Flory-Huggins scrive l'energia di mixing in termini di tre interazioni di accoppiamento ( $u_{AA}, u_{AB}, u_{BB}$ ) tra siti reticolari adiacenti occupati dalle due specie.

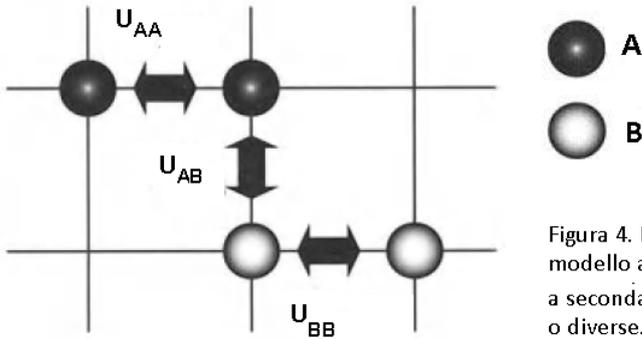


Figura 4. Forze di contatto tra due specie A e B nel modello a reticolo. Vengono definite tre forze  $U_{AA}, U_{BB}, U_{AB}$  a seconda che l'interazione avvenga tra specie identiche o diverse.

Definiamo ora l'energia media  $U_i$  di accoppiamento, per una specie di tipo i, come la media pesata delle energie di interazione con i monomeri adiacenti. La probabilità che uno dei primi vicini

sia un monomero di una data specie è assunta essere la frazione di volume corrispondente alla specie in questione, cioè  $\phi_A$  o  $\phi_B$  a seconda dei casi. Quindi le frazioni di volume rappresentano i pesi che compaiono nelle seguenti equazioni, relative ad A e B rispettivamente:

$$U_A = u_{AA}\phi_A + u_{AB}\phi_B \quad (2.11)$$

$$U_B = u_{BB}\phi_B + u_{AB}\phi_A \quad (2.12)$$

Ogni sito reticolare ha  $z$  primi vicini perciò l'interazione media di un monomero di tipo A con tutti i suoi  $z$  primi vicini risulta  $zU_A$ .

L'energia media per monomero sarà invece la metà di questo valore,  $\frac{zU_A}{2}$ , per il fatto che ogni interazione viene contata due volte (una volta per il monomero in questione e una per il suo vicino). L'energia corrispondente per sito reticolare occupato da una specie B avrà analoga forma. Dato che il numero di siti occupati da una specie A è  $M\phi_A$  e quello di una specie B è  $M\phi_B$  sommando tutti i contributi di tutte le interazioni otteniamo l'energia totale della nostra soluzione:

$$U = \frac{zM}{2}[U_A\phi_A + U_B\phi_B] \quad (2.13)$$

Poiché tra le due frazioni di volume vale la relazione  $\phi = \phi_A = 1 - \phi_B$ , combinando le espressioni 2.11-2.13 possiamo scrivere:

$$U = \frac{zM}{2}[u_{AA}\phi^2 + 2u_{AB}\phi(1 - \phi) + u_{BB}(1 - \phi)^2] \quad (2.14)$$

L'energia di interazione reticolare per una specie A prima del mescolamento è  $zu_{AA}$  poiché in questa situazione ogni monomero di specie A è circondato soltanto da monomeri della stesso tipo. Quindi l'energia totale di una specie A prima del mescolamento sarà:

$$\frac{zM}{2}u_{AA}\phi \quad (2.15)$$

mentre per B:

$$\frac{zM}{2}u_{BB}(1 - \phi) \quad (2.16)$$

Quindi l'energia totale di entrambe le specie prima del mescolamento sarà la somma delle due:

$$U_0 = \frac{zM}{2}[u_{AA}\phi + u_{BB}(1 - \phi)] \quad (2.17)$$

Arriviamo così all'espressione che ci interessa, ovvero alla differenza di energia tra lo stato mescolato e quello di partenza:

$$\Delta U_{mix} = U - U_0 = \frac{zM}{2}\phi(1 - \phi)(2u_{AB} - u_{AA} - u_{BB}) \quad (2.18)$$

che può essere riscritta in termini di differenza di energia per sito reticolare:

$$\Delta U'_{mix} = \frac{U - U_0}{M} = \frac{z}{2}\phi(1 - \phi)(2u_{AB} - u_{AA} - u_{BB}) \quad (2.19)$$

### 2.3 Energia libera e parametro di interazione

Il parametro di interazione di Flory  $\chi$  è una grandezza adimensionale definita per caratterizzare la differenza delle energie di interazione nella soluzione mescolata. Ha la seguente espressione:

$$\chi = \frac{z}{2} \frac{(2u_{AB} - u_{AA} - u_{BB})}{k_B T} \quad (2.20)$$

Questo ci permette di riscrivere l'equazione 2.19 in forma più compatta:

$$\Delta U'_{mix} = \chi \phi(1 - \phi) k_B T \quad (2.21)$$

Equazione che unita con la 2.10 ci consente di calcolare l'energia libera di Helmholtz del mescolamento per sito reticolare:

$$\Delta F'_{mix} = \Delta U'_{mix} - T \Delta S'_{mix} = k_B T \left[ \frac{\phi}{N_A} \ln \phi + \frac{1 - \phi}{N_B} \ln(1 - \phi) + \chi \phi(1 - \phi) \right] \quad (2.22)$$

I primi due termini dell'equazione sono contributi di origine entropica e agiscono sempre favorendo il mescolamento. L'ultimo termine invece ha origine energetica e può essere positivo (si oppone al mescolamento), nullo (nel caso di miscugli ideali), o negativo (favorisce il mescolamento) a seconda del segno di  $\chi$ . Si intuisce quindi l'importanza che riveste la conoscenza del parametro di interazione. Questo valore viene determinato per via sperimentale e scritto in una espressione dipendente dalla temperatura:

$$\chi(T) \cong A + \frac{B}{T} \quad (2.23)$$

Il termine A viene spesso definito come termine entropico, mentre il secondo, che esprime la dipendenza di  $\chi$  dalla temperatura, rappresenta il termine entalpico. Dal segno di A e B dipenderà il segno di  $\chi$ . Un valore di A positivo favorirà il mescolamento delle due sostanze opponendosi quindi alla separazione di fase. Un valore positivo di B agirà invece favorendo la separazione e inibendo il mescolamento.

### 2.4 Determinazione dei diagrammi di fase

Avendo ora a disposizione l'espressione per l'energia libera di mix possiamo passare alla determinazione dei diagrammi di fase.

Abbiamo visto nella sezione 1.2 che la stabilità di una soluzione in cui sono presenti due specie chimiche differenti dipende dalla curvatura della funzione  $\Delta F'_{mix}$  (equazione 1.7, 1.8). Andiamo quindi a calcolare la derivata seconda dell'energia rispetto alla frazione di volume:

$$\frac{\partial^2 \Delta F'_{mix}}{\partial \phi^2} = \frac{\partial^2 \Delta U'_{mix}}{\partial \phi^2} - T \frac{\partial^2 \Delta S'_{mix}}{\partial \phi^2} = k_B T \left[ \frac{1}{N_A \phi} + \frac{1}{N_B (1 - \phi)} \right] - 2\chi k_B T \quad (2.24)$$

Da questa si nota come, a temperature finite,  $\Delta F'_{mix}$  sia convessa agli estremi della composizione per il fatto che in queste due situazioni ( $\phi \rightarrow 0$  o  $\phi \rightarrow 1$ ) la derivata seconda del contributo entropico diverge.

Quindi agli estremi vi è una zona di stabilità.

Man mano che la temperatura viene abbassata il termine entropico diminuisce permettendo all'energia repulsiva di contribuire.

Nel caso in cui  $\chi > 0$  (il secondo termine dell'equazione 2.24 è quindi negativo) al di sotto di una temperatura critica  $T_C$  vi è un range di composizione nel quale sussistono fasi separate aventi una energia libera minore dello stato omogeneo. Chiaramente vi saranno più stati separati aventi energia minore dello stato omogeneo di mix, ma lo stato di equilibrio sarà definito dalla minore energia libera. Al fine di minimizzare la propria energia il sistema sceglie le due composizioni che hanno una linea retta comune con la pendenza più bassa possibile, che altro non è che la tangente comune di  $\Delta F'_{mix}$ .

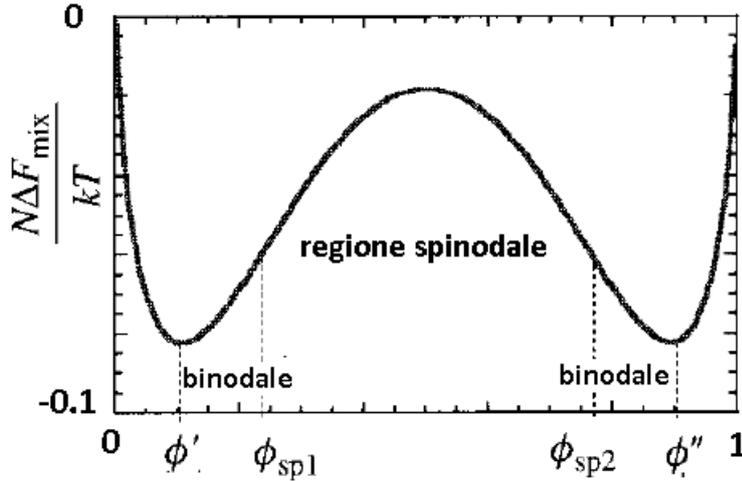


Figura 5. Dipendenza dell'energia libera dalla composizione per un miscuglio polimerico simmetrico ( $N_A = N_B = N$ ). E' possibile riconoscere la regione spinodale in cui la composizione è instabile e la regione binodale in cui è metastabile. Figura presa da [7]

Facendo riferimento alla figura, si vede che per tutte le composizioni che si trovano nella zona di mescolamento (miscibility gap), cioè tra  $\phi'$  e  $\phi''$ , il sistema riesce a minimizzare la propria energia separandosi in due fasi, che sono appunto  $\phi'$  e  $\phi''$ .

Dentro la miscibility gap ci sono regioni instabili e metastabili riconoscibili dal cambio di curvatura della funzione energia libera. Nello specifico tra i due punti di flesso la derivata seconda dell'energia è negativa e quindi lo stato mescolato è instabile. Questo fa sì che anche le più piccole fluttuazioni del sistema abbassino il valore dell'energia libera conducendo alla separazione di fase. Questa zona viene chiamata **regione spinodale**.

Viene invece definita **regione binodale**, la zona metastabile che si trova tra i punti di flesso e le due composizioni di equilibrio. La derivata seconda è ivi positiva e pertanto il mix è localmente stabile; infatti questa configurazione è stabile per piccole fluttuazioni nella composizione. Pertanto sono richieste larghe oscillazioni per raggiungere l'equilibrio termodinamico.

Andiamo ora a considerare la dipendenza dell'energia libera di mix  $\Delta F'_{mix}$  dalla temperatura. Così facendo possiamo costruire un diagramma di fase in modo da riassumere il comportamento della fase  $\phi$ , mostrando le regioni di stabilità, instabilità e metastabilità. I confini di fase sono determinati dalla tangente comune dell'energia libera nelle due composizioni che corrispondono alle fasi di equilibrio:

$$\left(\frac{\partial \Delta F'_{mix}}{\partial \phi}\right)_{\phi=\phi'} = \left(\frac{\partial \Delta F'_{mix}}{\partial \phi}\right)_{\phi=\phi''} \quad (2.25)$$

La derivata dell'energia libera ha espressione:

$$\frac{\partial \Delta F'_{mix}}{\partial \phi} = k_B T \left( \frac{\ln \phi}{N_A} + \frac{1}{N_A} - \frac{\ln(1-\phi)}{N_B} - \frac{1}{N_B} + \chi(1-2\phi) \right) \quad (2.26)$$

Per il caso in cui  $N_A = N_B = N$  la tangente comune è orizzontale quindi:

$$\left( \frac{\partial \Delta F'_{mix}}{\partial \phi} \right)_{\phi=\phi'} = \left( \frac{\partial \Delta F'_{mix}}{\partial \phi} \right)_{\phi=\phi''} = \frac{\partial \Delta F'_{mix}}{\partial \phi} = k_B T \left( \frac{\ln \phi}{N} - \frac{\ln(1-\phi)}{N} + \chi(1-2\phi) \right) = 0 \quad (2.27)$$

Ora invertendo la relazione precedente determiniamo un'equazione per il parametro di interazione  $\chi_b$  che fa riferimento alla regione binodale:

$$\chi_b = \frac{\ln[\phi/(1-\phi)]}{(2\phi-1)N} \quad (2.28)$$

e usando la dipendenza sperimentale del parametro di interazione dalla temperatura (Eq. 2.23) ricaviamo una relazione per la temperatura in funzione della composizione:

$$T_b = \frac{B}{\ln(\phi/(1-\phi))/(2\phi-1)N - A} \quad (2.29)$$

Abbiamo detto che la regione spinodale è la regione che si trova tra i due punti di flesso. Quindi per determinare una relazione analoga a quella determinata per la regione binodale bisogna determinare i punti di flesso ponendo la derivata seconda dell'energia libera uguale a zero:

$$\frac{\partial^2 \Delta F'_{mix}}{\partial \phi^2} = k_B T \left( \frac{1}{N_A \phi} + \frac{1}{N_B(1-\phi)} - 2\chi \right) = 0 \quad (2.30)$$

Da questa ricaviamo una relazione per parametro di interazione nella regione spinodale:

$$\chi_s = \frac{1}{2} \left( \frac{1}{N_A \phi} + \frac{1}{N_B(1-\phi)} \right) \quad (2.31)$$

E analogamente a prima un'espressione per la temperatura in funzione della composizione:

$$T_s = \frac{B}{\frac{1}{2} \left( \frac{1}{N_A \phi} + \frac{1}{N_B(1-\phi)} \right) - A} \quad (2.32)$$

Queste espressioni ci permettono di costruire così i seguenti diagrammi di fase (con  $N_A = N_B$  in 2.31, 2.32) :

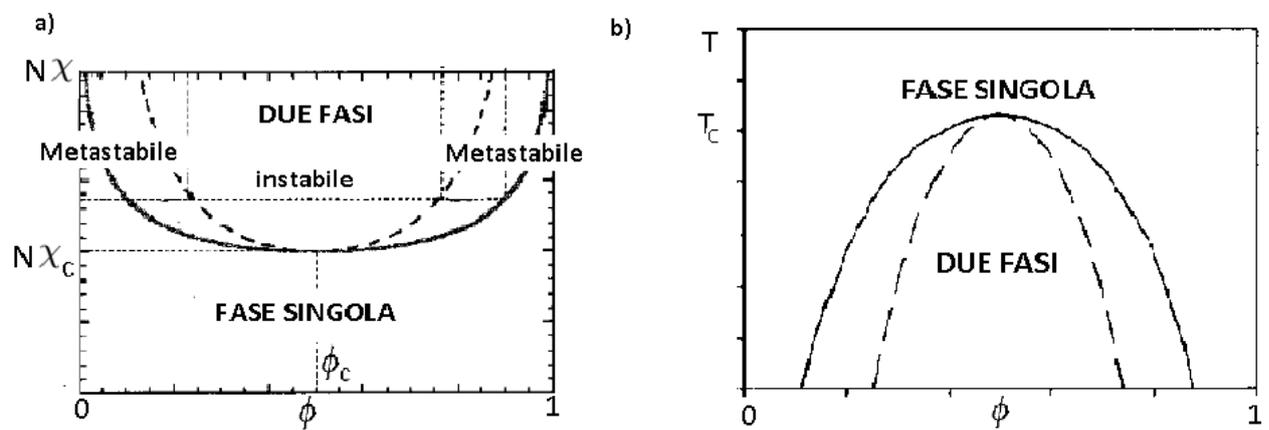


Figura 6. Diagrammi di fase per un miscuglio simmetrico ( $N_A = N_B$ ). In a) viene rappresentato il parametro di interazione in funzione della concentrazione. Si riconoscono le zone instabili e metastabili. In b) troviamo la dipendenza della temperatura dalla concentrazione. Vediamo che al di sotto di una temperatura critica  $T_c$  vi è la coesistenza di due fasi separate. Figura presa da [7]

## Capitolo 3

# RPA (Random Phase Approximation)

Nel primo paragrafo di questo capitolo verrà effettuata una modifica alla teoria di FH mentre nel secondo andremo a vedere, a grandi linee, la teoria della RPA che ci permetterà di andare a calcolare l'interazione elettrostatica presente tra i residui amminoacidici carichi. Vedremo infatti, nel dettaglio nel capitolo 4, come la separazione di fase di una proteina dipende fortemente dalla sua sequenza di amminoacidi. Riuscire quindi a computare correttamente le interazioni presenti ci permette di comprenderne il comportamento.

### 3.1 Entropia di configurazione

Seguendo i ragionamenti e la teoria ben esposta in [5] consideriamo una soluzione acquosa in cui sono presenti dei polianfoliti (cioè polimeri aventi una carica) e vediamo come calcolare l'entropia di configurazione di un siffatto sistema. Nella soluzione vi saranno quindi, oltre ai residui amminoacidici della proteina, controioni e sale. Ogni catena polianfolitica è composta da  $N$  residui amminoacidici aventi carica (in unità di  $e$ )  $\sigma_i = (\sigma_1, \sigma_2, \dots, \sigma_N)$  e quindi aventi valori  $(+1, -1, 0)$ . La maggior parte del calcolo dell'entropia di questo sistema è basata sul modello reticolare di FH, in cui il volume totale è partizionato in siti aventi volume  $a^3$ , dell'ordine di una molecola del solvente.

Perciò come visto nella sezione 2.1 il numero totale di siti reticolari sarà  $M = V/a^3$ .

Allora l'energia libera totale  $F$  per sito reticolare in unità di  $k_B T$  è data da:

$$f \equiv \frac{F a^3}{V k_B T} = -s + f_{int}, \quad (3.1)$$

dove  $-s$  è l'entropia che deriva dalla trattazione di FH e  $f_{int}$  è l'energia interna dovuta alle interazioni elettrostatiche che calcoleremo utilizzando appunto la RPA. Le quantità che compaiono nell'espressione 3.1 sono quantità adimensionali. Il termine  $f_{int}$  viene spesso caratterizzato come il contributo "entalpico" del sistema, anche se in realtà dovrebbe contenere anche termini entropici come ad esempio le interazioni idrofobiche che dipendono dalla temperatura. Vogliamo ora generalizzare il termine entropico già visto nella sezione 2.1 (Eq. 2.5) per un insieme di  $m$  differenti tipi di polimeri che indichiamo con  $i = 1, 2, \dots, m$  ognuno con  $N_i$  monomeri di dimensione  $r_i a^3$ , dove le molecole del solvente, controioni e sale sono polimeri con  $N_i = 1$ .

In questo sistema il numero totale di siti reticolari è dato da :

$$M = \frac{V}{a^3} = \sum_{k=1}^m r_k n_k N_k, \quad (3.2)$$

dove  $n_i$  è il numero di polimeri di tipo  $i$  nella soluzione. Definiamo ora la concentrazione per un monomero di tipo  $i$  in unità di  $a^3$  :

$$\phi_i = \rho_i a^3 = \frac{n_i N_i}{M} \quad (3.3)$$

dove  $\rho_i$  rappresenta  $\rho_m, \rho_s$  e  $\rho_c$ , cioè le concentrazioni rispettivamente dei monomeri, sale e controioni. Nello specifico  $\rho_c = \rho_m |\sum_i \sigma_i|/N$  perché il numero di controioni è uguale alla carica netta dei polianfoliti.

Riprendendo la trattazione di FH separiamo il numero di stati  $\Omega$  in un contributo che discende dalla libertà traslazione del primo monomero di ogni polimero ( $\Omega_{CM}$ ) e in un altro che tiene conto della libertà conformazionale di ogni polimero con una fissata posizione del primo monomero  $\Omega_{conf}$  così che  $\Omega = \Omega_{CM} \Omega_{conf}$ .

Da ora in poi considereremo  $r_i = 1$  per ogni  $i$ . Il termine di traslazione, che corrisponde al numero di modi di arrangiare  $\sum_i n_i$  monomeri quando ognuno di essi è il primo monomero di un polimero, risulta essere:

$$\Omega_{CM} = \frac{M!}{(M - \sum_{i=1}^m n_i)! \prod_{i=1}^m n_i!}, \quad (3.4)$$

Per quanto riguarda invece il termine conformazionale, in un reticolo in cui ogni sito ha  $z$  primi vicini, si scrive la formula:

$$\Omega_{conf} = (z-1)^{N_1} \prod_{l=0}^{N_2-1} \left(1 - \frac{\sum_{i=1}^m n_i + l}{M}\right) = \frac{(z-1)^{N_1}}{M^{N_2}} (M - \sum_{i=1}^m n_i)!, \quad (3.5)$$

dove  $N_1 = N_2 = \sum_i n_i (N_i - 1)$  è il numero totale di siti (monomeri) che possono essere riempiti dopo la collocazione del primo monomero.

Quindi ora avendo i due contributi possiamo scrivere una espressione per  $\Omega$ , espressione in cui va tenuto conto di una possibile violazione del volume escluso; dopo alcune semplificazioni si arriva alla formula di interesse:

$$\Omega = \Omega_{CM} \Omega_{conf} = \frac{M!}{\prod_{i=1}^m n_i!} \frac{(z-1)^{\sum_{i=1}^m n_i (N_i - 1)}}{M^{\sum_{i=1}^m n_i (N_i - 1)}} \quad (3.6)$$

Quindi possiamo scrivere l'espressione dell'entropia di configurazione per sito reticolare:

$$-s = -\frac{1}{M} \ln \Omega = \sum_{i=1}^m \frac{n_i}{M} \ln n_i + \quad (\text{termini lineari in } n_i) = \sum_{i=1}^m \frac{\phi_i}{N_i} \ln \phi_i + \quad (\text{termini lineari in } \phi_i) \quad (3.7)$$

E' possibile dimostrare, ma noi non lo faremo, come i termini lineari in  $\phi_i$  non influiscano sulla separazione di fase. Scriviamo quindi un'espressione finale per l'entropia di mescolamento:

$$-s = \frac{\phi_m}{N} \ln \phi_m + \phi_c \ln \phi_c + 2\phi_s \ln \phi_s + \phi_w \ln \phi_w \quad (3.8)$$

dove vale la relazione (ponendo  $r_i = 1$ )  $\phi_m = 1 - \phi_c - 2\phi_s - \phi_w$

### 3.2 Energia elettrostatica

Abbiamo visto nella teoria di FH che l'energia interna che compare nel calcolo dell'energia libera di Helmholtz era rappresentata dalle energie di contatto presenti tra gli elementi della catena. Nell'equazione 3.1 invece  $f_{int}$  è costituita dalle interazioni elettrostatiche che avvengono tra i polimeri carichi all'interno della nostra soluzione. Si suppone quindi che le interazioni a lungo range siano molto più forti di quelle di contatto. Per risolvere questo problema, è possibile utilizzare la teoria RPA, per i dettagli si vedano [4] e [5], che fornisce per la forza elettrostatica la seguente espressione:

$$f_{el} = \frac{1}{2} \int_0^\infty \frac{d^3(ka)^3}{(2\pi)^3} \{ \ln[\det(1 + \hat{G}_k \hat{U}_k)] - Tr(\hat{\rho} \hat{U}_k) \}, \quad (3.9)$$

dove  $\hat{G}_k$  rappresenta una matrice di correlazione di densità in assenza di interazioni tra i monomeri (quindi tiene conto solo dei vincoli geometrici come la connettività della catena),  $\hat{U}_k$  è la matrice che computa le interazioni elettrostatiche tra le cariche elettriche e  $\hat{\rho}$  è la matrice delle densità delle specie presenti nella soluzione; sono tutte matrici  $(N+2) \times (N+2)$ . In generale, in un sistema di polianfoliti e monomeri ionici, la matrice  $\hat{G}_k$  racchiude al suo interno quattro categorie di correlazione: monomero-monomero (MM), ione-ione (II) e monomero-ione (MI) con le rispettive matrici  $\hat{G}_{MM}$ ,  $\hat{G}_{II}$ ,  $\hat{G}_{MI}$ . Quindi

$$\hat{G}_k = \begin{pmatrix} \hat{G}_{MM}(k) & \hat{G}_{MI}(k) \\ \hat{G}_{IM}(k) & \hat{G}_{II}(k) \end{pmatrix}, \quad (3.10)$$

tuttavia nel nostro caso trattiamo tutte le molecole presenti come catene gaussiane in cui le correlazioni esistono solo tra monomeri dello stesso polimero tenuti assieme dalla connettività della catena. Di conseguenza tutti gli elementi delle matrici  $\hat{G}_{IM}(k)$  e  $\hat{G}_{MI}(k)$  sono zero. In altre parole in questo modello non ci sono correlazioni tra monomeri e ioni.

La matrice  $\hat{G}_{II}(k)$  cioè la matrice di correlazione ione-ione è una matrice  $2 \times 2$  diagonale:

$$\hat{G}_{II} = \hat{\rho}_I = \begin{pmatrix} \rho_+ & 0 \\ 0 & \rho_- \end{pmatrix}, \quad (3.11)$$

dove  $\rho_+$  e  $\rho_-$  sono le densità degli ioni positivi e negativi.

Invece la matrice di correlazione monomero-monomero viene calcolata come prodotto del fattore di struttura di una catena Gaussiana per la densità dei polimeri:

$$\hat{G}_{MM}(k) = \frac{\rho_m}{N} \hat{G}_M(k). \quad (3.12)$$

Gli elementi di matrice di  $\hat{G}_M(k)$  sono dati da :

$$\hat{G}_M(k)_{ij} = e^{-(kb)^2 |i-j|/6}, \quad (3.13)$$

laddove  $b$  rappresenta la lunghezza dei legami polimerici delle catene gaussiane. Quindi la matrice  $\hat{G}_M(k)_{ij}$  tiene conto delle correlazioni tra l' $i$ -esimo e il  $j$ -esimo monomero.

Mettendo insieme le due relazioni precedenti otteniamo:

$$\hat{G}_k = \begin{pmatrix} (\rho_m/N) \hat{G}_M(k) & 0 \\ 0 & \hat{\rho}_I \end{pmatrix}. \quad (3.14)$$

La matrice  $\hat{U}_k$  dell'equazione 3.9 ha invece espressione:

$$\hat{U}_k = \frac{4\pi l_B}{k^2[1 + (kb)^2]} |q\rangle \langle q| = \lambda(k) |q\rangle \langle q|, \quad (3.15)$$

dove  $l_B = e^2/(4\pi\varepsilon_0\varepsilon_r k_B T)$  è la lunghezza di Bjerrum, che rappresenta la distanza alla quale l'interazione elettrostatica tra due cariche è comparabile in magnitudo all'energia termica  $k_B T$ ,  $\varepsilon_0$  e  $\varepsilon_r$  sono rispettivamente la permittività del vuoto e quella relativa,  $\langle q|$  è il vettore colonna per le cariche dei monomeri e degli ioni monovalenti con componenti  $|q_i\rangle = \sigma_i$  per  $1 \leq i \leq N$ ,  $q_{N+1} = 1$  e  $q_{N+2} = -1$

Il determinante dell'equazione (3.9) può essere semplificato, utilizzando l'identità di Sylvester, come:

$$\det(1 + \hat{G}_k \hat{U}_k) = 1 + \lambda(k) \langle q| \hat{G}_k |q\rangle = 1 + \lambda(k) \left( 2\rho_s + \rho_c + \frac{\rho_m}{N} \langle \sigma| \hat{G}_M(k) | \sigma \rangle \right) \quad (3.16)$$

e nel limite per  $k \rightarrow \infty$

$$\ln[\det(1 + \hat{G}_k \hat{U}_k)] \rightarrow \lambda(k) \langle q| \hat{G}_k |q\rangle. \quad (3.17)$$

Ora per comodità di calcolo si definisce una temperatura ridotta  $T^*$

$$T^* \equiv b/I_B = 4\pi\varepsilon_0\varepsilon_r k_B T b / e^2 \quad (3.18)$$

e l'integrazione di  $f_{el}$  viene riscritta in termini del numero d'onda ridotto  $k' = kb$ . Utilizzando inoltre le densità riscalate  $\phi_m = \rho_m a^3$ ,  $\phi_c = \rho_c a^3$  e  $\phi_s = \rho_s a^3$ , possiamo scrivere un'espressione per  $f_{el}$  in termini di variabili adimensionali,

$$f_{el} = \int_0^\infty \frac{dk' k'^2}{4\pi^2} \left\{ \frac{1}{\eta} \ln[1 + \eta G(k')] - G(k') \right\}, \quad (3.19)$$

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

$$G(k') = \frac{4\pi}{k'^2(1 + k'^2)T^*} (2\phi_s + \phi_c + \frac{\phi_m}{N} \langle \sigma| \hat{G}_M(k) | \sigma \rangle). \quad (3.20)$$

## Capitolo 4

# Risultati per la proteina *Ddx4*

In questo capitolo applichiamo la teoria della RPA alla proteina *Ddx4*. Come accennato nel paragrafo (1.3) questa rientra nella categoria delle proteine intrinsecamente disordinate, proteine nelle quali vi è una maggior presenza di elementi polari e carichi. L'obiettivo sarà quello di mettere in luce come il comportamento di questa dipenda dalla disposizione dei residui carichi all'interno della sequenza amminoacidica.

### 4.1 *Ddx4*<sup>N1</sup> e *Ddx4*<sub>CS</sub>

Lo studio dei meccanismi che regolano il comportamento della proteina *Ddx4* è di notevole importanza poiché si è recentemente osservato che questa è in grado di effettuare una separazione di fase liquido-liquido in cui si ottengono due fasi solubili in ambiente acquoso aventi due differenti concentrazioni. Durante questo processo si formano delle gocce viscosi essenziali per la vita cellulare poiché costituenti di organelli privi di membrana che svolgono rilevanti funzioni biologiche. Per evidenziare quanto sopra affermato, cioè che la separazione di fase dipenda fortemente dalla collocazione degli amminoacidi carichi, studiamo, come fatto in [3] e [4], la proteina originale *Ddx4*<sup>N1</sup> e la sua proteina mutante *Ddx4*<sub>CS</sub>, cioè la proteina che si ottiene scambiando all'interno della sequenza la posizione di alcuni dei residui carichi. In figura mostriamo i primi 236-residui amminoacidici delle due sequenze:

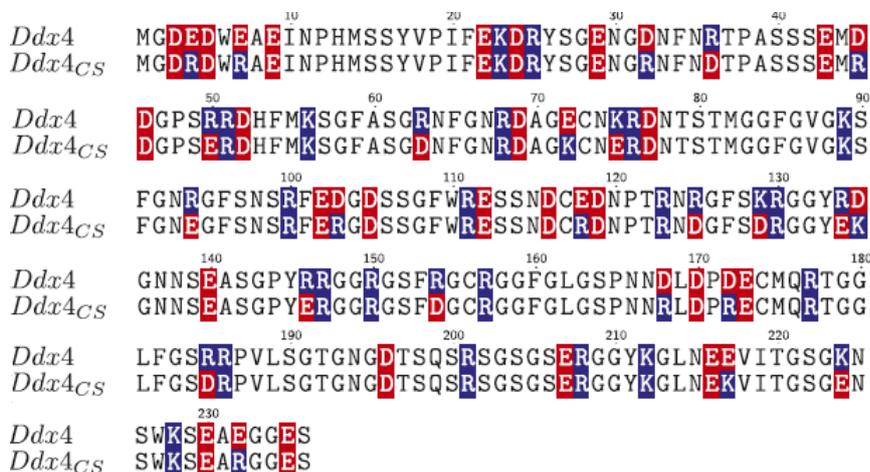


Figura 7. I primi 236 residui amminoacidici della proteina *Ddx4* e della sua mutante *Ddx4*<sub>CS</sub>, ottenuta scambiando la posizione di alcuni residui carichi tra loro. Figura presa da [1]

L'energia libera della proteina *Ddx4*, che calcoliamo mettendo insieme l'entropia di configurazione (Eq. 3.8) e l'energia elettrostatica (Eq. 3.19), è funzione dei polianfoliti e della densità di sale,

$$f = f(\phi_m, \phi_s). \quad (4.1)$$

Consideriamo inizialmente una situazione semplice in cui la concentrazione di sali è nulla ( $\phi_s = 0$ ). Inoltre assumiamo che la lunghezza di legame  $b$  equivalga alla lunghezza del solvente  $a$ , per cui nell'equazione 3.19  $\eta = 1$ . Quindi nella nostra soluzione saranno presenti i monomeri  $\phi_m$  della proteina e i contro-ioni  $\phi_c = |\sum_i \sigma_i| \phi_m / N$ .

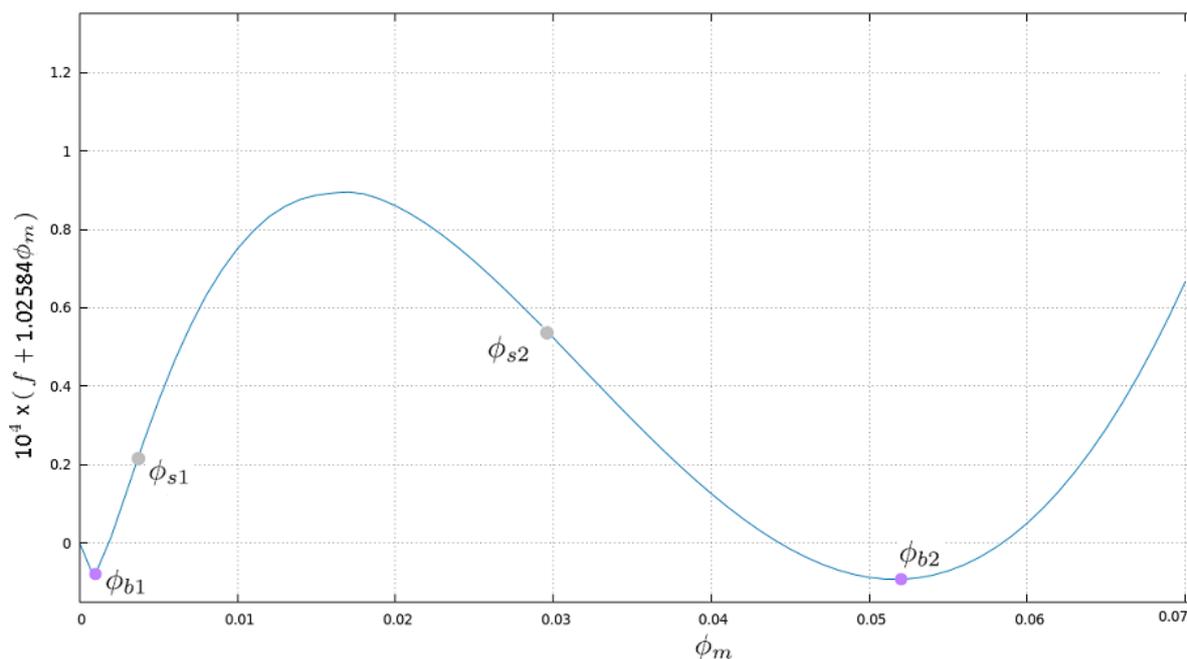


Figura 8. Energia libera in funzione della concentrazione a  $T^*=0.286$  per la proteina Ddx4

Come si può vedere in figura, per una temperatura fissata  $T^*$  la funzione energia libera mostra due minimi locali, uno a  $\phi_m \approx 0.006$  e l'altro a  $\phi_m \approx 0.052$ .

I due punti  $\phi_m = \phi_{s1}, \phi_{s2}$ , soddisfano la condizione

$$f'' = \frac{d^2 f}{d\phi_m^2} = 0. \quad (4.2)$$

Tra questi due punti, che definiscono la regione spinodale, l'energia libera è concava e una qualsiasi concentrazione  $\phi_m$  dello stato mescolato non è stabile perché l'energia libera di questo è maggiore di quella di un generico stato separato.

Per determinare un diagramma di fase in cui si mettono in relazione temperature e concentrazione dovremo risolvere il sistema di equazioni 4.3:

$$\begin{cases} f'(\phi_m^\alpha) = f'(\phi_m^\beta) \\ f(\phi_m^\alpha) - \phi_m^\alpha f'(\phi_m^\alpha) = f(\phi_m^\beta) - \phi_m^\beta f'(\phi_m^\beta) \end{cases} \quad (4.3)$$

Per ogni temperatura sarà possibile trovare due fasi  $\phi_m^\alpha$  e  $\phi_m^\beta$  che soddisfano il sistema e che

rappresentano le due concentrazioni che coesistono all'equilibrio. Possiamo quindi costruire un grafico in cui mostriamo la regione binodale per la proteina  $Ddx4^{N1}$  e per la  $Ddx4_{CS}$ .

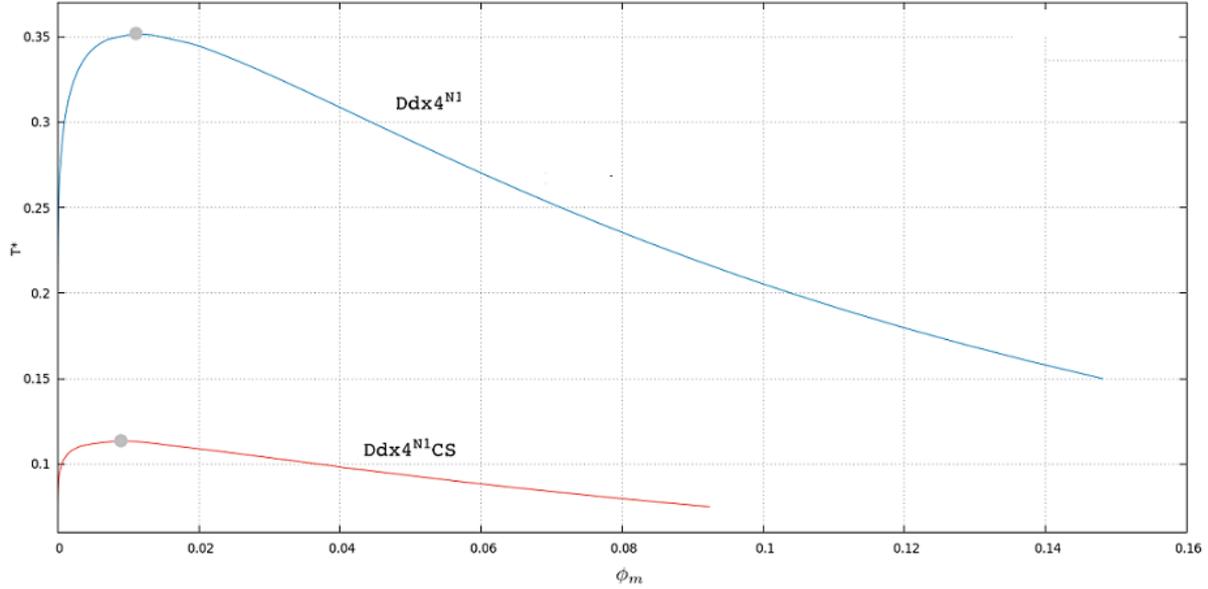


Figura 9. Separazione di fase per le proteine Ddx4 e Ddx4 Cs.

Come si può facilmente notare la temperatura critica  $T_{cr}^*$  e la concentrazione critica  $\phi_{cr}$  sono molto diverse per le due sequenze, evidente segno che il comportamento dipende fortemente dalla posizione dei residui carichi e non dalla carica netta della proteina (che è appunto la stessa nei due casi). Nello specifico la  $T_{cr}^*$  della proteina mutata risulta essere all'incirca 1/3 di quella originale, e allo stesso tempo anche la  $\phi_{cr}$  risulta essere minore nella sequenza mutata.

Sperimentalmente, a temperatura ambiente, si osserva infatti che la proteina originale effettua una separazione di fase mentre la mutante no. Nel caso della proteina originale la coesistenza di due fasi separate si osserva in un evidente cambio di torbidità della soluzione man mano che viene abbassata la temperatura, fino alla formazione di gocce con una concentrazione di  $Ddx4$  relativamente alta.

## 4.2 $Ddx4$ in presenza di NaCl

Andiamo ora a vedere il comportamento della nostra proteina in una soluzione acquosa in presenza di sale. Limitiamo le considerazioni a concentrazioni di sale in cui  $\phi_s = 0.0018, 0.0027, 0.0036,$  e  $0.0054$  che corrispondono rispettivamente a  $[NaCl]=100, 150, 200,$  e  $300$  mM. La relazione che lega la frazione di volume dei residui amminoacidici alla concentrazione della proteina espressa in unità molarli è data da  $\phi_m = 236 [Ddx4]/(55.5M)$  poiché  $[H_2O]=55.5M$ . Focalizziamo la nostra attenzione nel range di concentrazione  $[Ddx4]=0-400 \mu M$ . Utilizzando le equazioni (3.8) e (3.19) per il calcolo dell'energia interna e risolvendo poi il sistema di equazioni (4.3) abbiamo costruito i diagrammi di fase per la proteina in presenza di sale.

Come si può facilmente notare la presenza di Nacl all'interno della soluzione influenza la separazione di fase; maggiore è la quantità minore sarà la temperatura critica al di sotto della quale

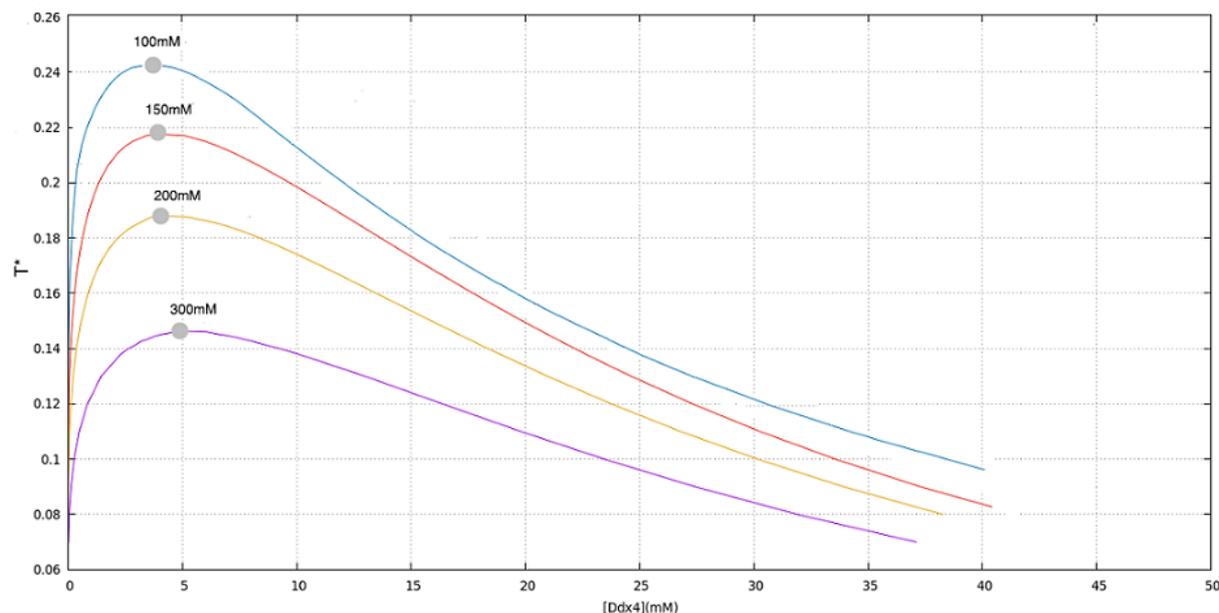


Figura 10. Diagrammi di fase per la proteina Ddx4 a diverse concentrazioni di NaCl (100,150,200 e 300 mM)

avviene la separazione in due fasi distinte.

Diversamente dal grafico in cui si erano confrontate le due sequenze per la *Ddx4* qui abbiamo posto sull'asse delle ascisse la concentrazione espressa in mM.

Si presenta ora un problema, non banale, nella scelta del valore da attribuire alla permittività relativa  $\varepsilon_r$ . Infatti per poter passare da temperature adimensionali a temperature in Kelvin necessitiamo di questo valore da inserire nell'equazione (3.18).

La nostra soluzione è costituita da proteine immerse in soluzione acquosa.

La permittività dell'acqua è  $\varepsilon_w \approx 80$  mentre quella della proteina è  $\varepsilon_p \approx 2 - 4$  pertanto abbiamo a che fare con un mezzo in cui la permittività può variare tra questi due valori.

Tuttavia se si sceglie un qualsiasi valore costante di  $\varepsilon_r$  (compreso tra 4 e 80) si ricavano temperature troppo basse, addirittura negative (in gradi Celsius), segno che il modello così applicato, sebbene riesca a costruire diagrammi di fase sensati da un punto di vista strutturale, pone la proteina in situazioni fisiche non accettabili.

### 4.3 Permittività dipendente dalla concentrazione

Un primo tentativo di risolvere la questione consiste nel pensare che la permittività relativa della nostra soluzione dipenda dalla concentrazione  $\phi_m$  di monomeri all'interno della soluzione e che quindi non sia un valore costante.

Un modo per stimare l'effettiva  $\varepsilon_r$  è quello di considerare due piani carichi di larghezza infinita con in mezzo proteine e solvente. Per una soluzione in cui  $\phi_m$  è la frazione di volume dei monomeri e  $1 - \phi_m$  è quella dell'acqua il campo elettrico  $E$  tra i due piani è dato dalla legge di Gauss:

$$\varepsilon_0 E = \phi_m \frac{\sigma}{\varepsilon_p} + (1 - \phi_m) \frac{\sigma}{\varepsilon_w} = \frac{\sigma}{\varepsilon_{eff}}, \quad (4.3)$$

pertanto la permittività relativa efficace sarà:

$$\varepsilon_{eff}(\phi_m) = \frac{\varepsilon_p \varepsilon_w}{\phi_m \varepsilon_w + (1 - \phi_m) \varepsilon_p}, \quad (4.4)$$

il cui andamento è mostrato in figura. L'andamento è sensato e molto simile a quello previsto dalla teoria di Clausius-Mossotti [5].

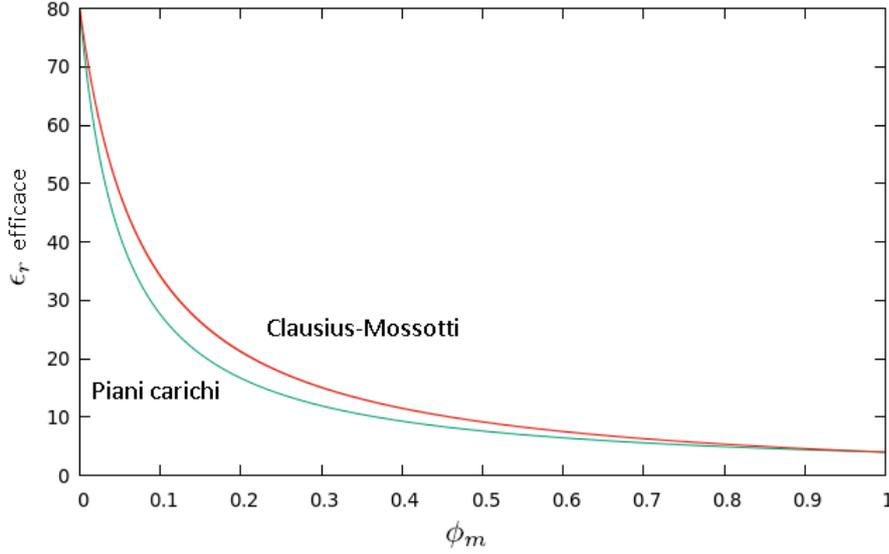


Figura 11. Permettività relativa efficace in funzione della concentrazione. In rosso la curva ottenuta con il modello di Clausius-Mossotti mentre in azzurro quella ottenuta con il modello a piani carichi.

Andiamo ora ad inserire quanto ottenuto nella teoria RPA modificando l'espressione di  $I_B$  e di conseguenza ridefinendo una temperatura ridotta  $T_0^*$  in termini della sola permittività del vuoto  $\varepsilon_0$ :

$$T_0^* = 4\pi\varepsilon_0 k_B T b / e^2 \quad (4.5)$$

Inoltre dobbiamo modificare l'espressione (3.19) della forza elettrostatica. Infatti poiché ora  $G(k)$  non è più lineare in  $\phi_m$  non possiamo più usarlo allo scopo di eliminare la divergenza "ultravioletta" perché altrimenti verrebbero alterati i comportamenti di fase.

Di conseguenza, per il caso  $\varepsilon_r(\phi_m)$  l'espressione modificata risulta:

$$f_{el} = \int_0^\infty \frac{dk' k'^2}{4\pi^2} \left\{ \frac{1}{\eta} \ln[1 + \eta G_1(k')] - G_2(k') \right\}, \quad (4.6)$$

dove

$$G_1(k') = \frac{4\pi}{k'^2(1 + k'^2)T_0^* \varepsilon_r(\phi_m)} \left( 2\phi_s + \phi_c + \frac{\phi_m}{N} \langle \sigma | \hat{G}_M(k') | \sigma \rangle \right) \quad (4.7)$$

$$G_2(k') = \frac{4\pi}{k'^2(1 + k'^2)T_0^* \varepsilon_r(\phi_m)} \left( 2\phi_s + \phi_c + \frac{\phi_m}{N} \sum_i |\sigma_i| \right). \quad (4.8)$$

Utilizzando queste equazioni per il calcolo dell'energia interna e risolvendo il sistema di equazioni 4.3 per i confini di fase, presentiamo i risultati che si ottengono per una soluzione in assenza di sale. In particolare in figura viene mostrato un confronto tra le curve che si ottengono utilizzando rispettivamente un  $\varepsilon_r$  costante e uno dipendente dalla concentrazione.

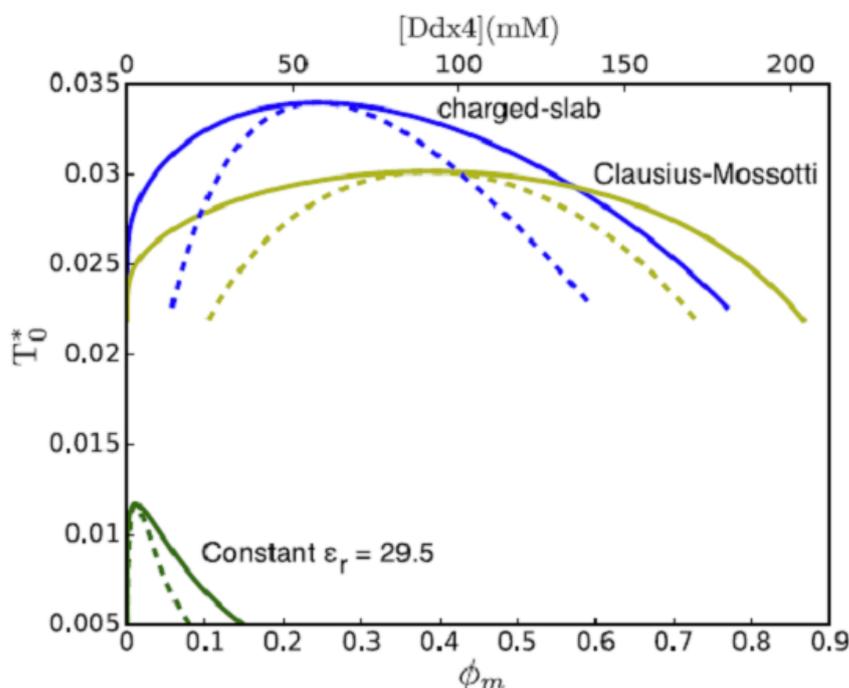


Figura 12. Separazione di fase (in assenza di sale) con  $\varepsilon_r=29.5$  e dipendente dalla concentrazione. Figura presa da [4]

Come si può vedere le proprietà di fase vengono notevolmente modificate. In particolare si osserva un significativo shift sia sull'asse relativo alle concentrazioni che su quello delle temperature. Inoltre vengono notevolmente ampliati i confini di fase.

Tuttavia, sebbene la dipendenza della permittività relativa dalla concentrazione sia un'ipotesi sensata, andando ad invertire la relazione (4.5) troviamo ancora una volta temperature fisicamente non plausibili; questa volta si ottengono temperature non negative, bensì troppo elevate, dell'ordine dei migliaia di Kelvin.

#### 4.4 RPA + Flory-Huggins

La teoria RPA ci ha permesso di razionalizzare il differente comportamento della proteina  $Ddx4^{N1}$  e di quella mutante, differenze che non vengono spiegate dalla teoria di Flory-Huggins. Tuttavia abbiamo notato come la teoria RPA vada incontro ad un problema nel passaggio da

temperature dimensionali a temperature in Kelvin. Abbiamo anche visto che utilizzando una  $\varepsilon_{eff}(\phi_m)$  dipendente dalla concentrazione il problema di fondo non viene risolto. Una possibile soluzione è quella di utilizzare la teoria RPA, per tener conto delle interazioni elettrostatiche, unita a quella di Flory-Huggins per computare le forze di contatto all'interno della catena. Pertanto all'energia elettrostatica  $f_{el}$  aggiungiamo il contributo (già definito in (2.21), ma che riportiamo per completezza) dato da  $f_{FH} = \chi\phi_m(1 - \phi_m) = \left(\frac{B}{T^*} + A\right)\phi_m(1 - \phi_m)$ , con  $B=0.15$  e  $A=-0.3$  (valori ottenuti tramite fit con i dati sperimentali disponibili come fatto in [5]).

Otteniamo così:

$$f_{int} = f_{el} + f_{FH}, \quad (4.9)$$

Quindi risolvendo il sistema (4.3) possiamo costruire i diagrammi di fase della proteina *Ddx4* in presenza di sale.

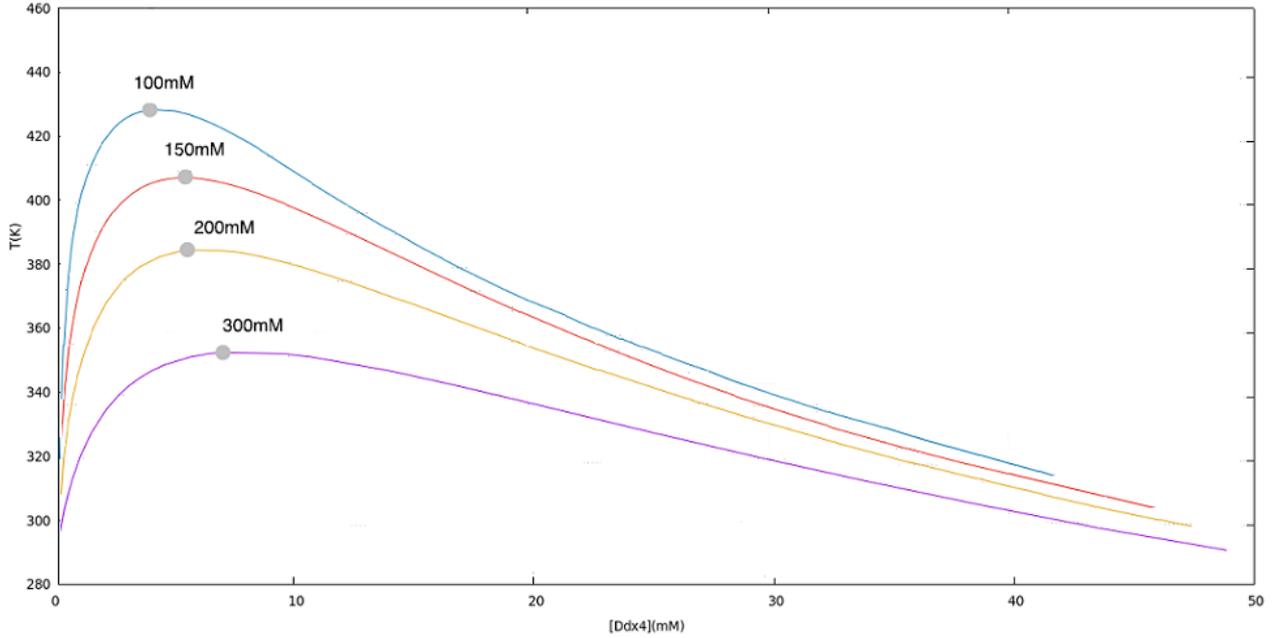


Figura 13. Diagrammi di fase per la proteina *Ddx4* a diverse concentrazioni di sale. Grafici ottenuti utilizzando RPA + FH

Come si può notare con questo ampliamento della teoria si ottengono valori di temperatura sensati. In questo caso per la conversione in Kelvin si utilizza  $\varepsilon_r = 29.5$ , valore preso da [5] e derivante da un fit con i dati sperimentali disponibili per la proteina.

## Capitolo 5

# Confronto con i dati sperimentali e conclusioni

Nel paragrafo precedente abbiamo visto che un modo per poter risolvere il problema del passaggio da temperature adimensionali a temperature reali è quello di utilizzare la teoria RPA assieme alla teoria di Flory-Huggins. In figura mostriamo un confronto tra le curve teoriche ottenute con questo modello e i dati sperimentali per la proteina *Ddx4* (estrapolati da [6]).

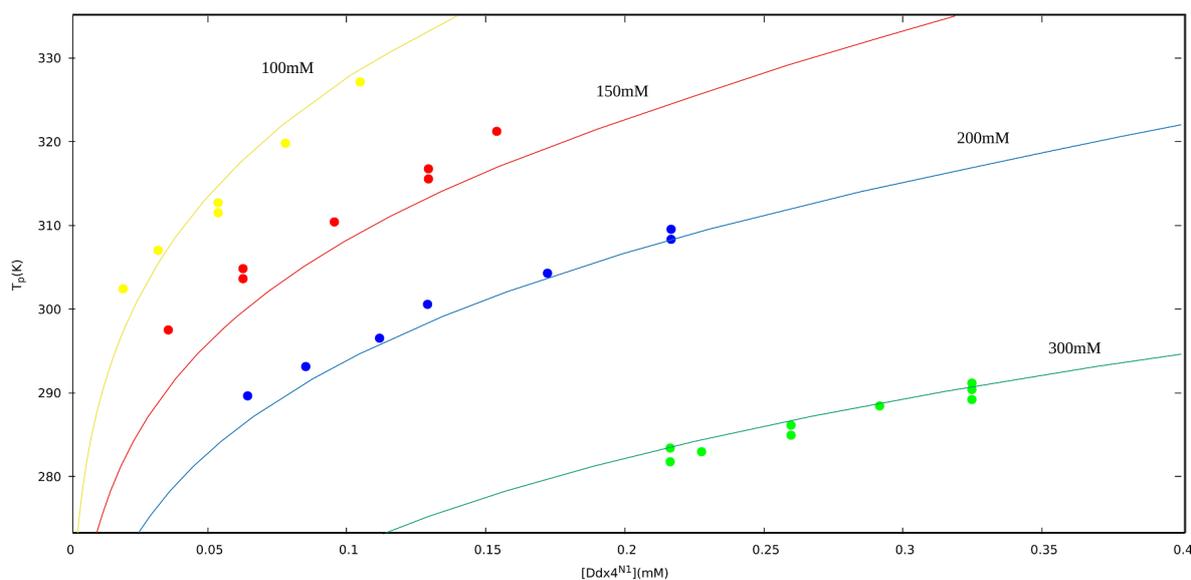


Figura 14. Confronto tra dati sperimentali e curve teoriche ottenute con RPA + FH.

Vediamo che le curve teoriche seguono piuttosto bene l'andamento dei dati sperimentali, e riproducono in maniera ottimale il grafico a pag.188 di [5], evidente segno che il modello in questione dia una rappresentazione realistica delle interazioni presenti all'interno della proteina. Tuttavia bisogna dire che nella conversione per le temperature viene utilizzato un valore della permittività relativa  $\epsilon_r$  che è il risultato di un fit con i dati sperimentali.

Infatti, sebbene questo sia un valore sensato per una bassa concentrazione di IDPs, un modello più completo consisterebbe nel riuscire a computarlo in maniera autonoma, senza cioè far utilizzo dei dati sperimentali al fine di ottenere un risultato valido. E' anche vero però che, come abbiamo visto, inserire nella teoria un valore di permittività dipendente dalla concentrazione di

monomeri non risolve il problema della conversione delle temperature. Quindi, sebbene questa ipotesi di dipendenza sia molto sensata e meriterebbe di essere approfondita sembra non essere compatibile con la teoria RPA qui esposta.

Dunque per concludere mettiamo in evidenza i risultati ottenuti per la proteina *Ddx4*. Abbiamo visto che per proteine intrinsecamente disordinate, in cui vi è una forte presenza di amminoacidi carichi, è necessario andare oltre la teoria classica di Flory-Huggins per il calcolo dell'energia interna poiché questa non tratta delle interazioni elettrostatiche che hanno luogo tra i residui amminoacidici.

La teoria RPA è in grado di computare con successo queste interazioni e di mettere in luce come la collocazione dei residui carichi all'interno della sequenza influenzi e modifichi il comportamento di fase della proteina. Tuttavia, nonostante la RPA da sola sia in grado di evidenziare questa forte dipendenza e quella dovuta alla quantità di sale presente in soluzione, i risultati compatibili con una situazione fisicamente accettabile risultano essere quelli che si ottengono andando ad unire questo modello con la teoria di Flory-Huggins che considera le interazioni interne di contatto tra gli elementi della catena proteica. Questo sembra essere un buon modo di procedere poiché si è in grado di ottenere curve teoriche che sono in buon accordo con i dati sperimentali.

# Bibliografia

- [1] Jacob P Brady, Patrick J Farber, Ashok Sekhar, Yi-Hsuan Lin, Rui Huang, Alaji Bah, Timothy J Nott, Hue Sun Chan, Andrew J Baldwin, Julie D Forman-Kay, et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences*, 114(39):E8194–E8203, 2017.
- [2] Alexei V Finkelstein and Oleg Ptitsyn. *Protein physics: a course of lectures*. Elsevier, 2016.
- [3] Johnny Habchi, Peter Tompa, Sonia Longhi, and Vladimir N Uversky. Introducing protein intrinsic disorder. *Chemical reviews*, 114(13):6561–6588, 2014.
- [4] Yi-Hsuan Lin, Julie D Forman-Kay, and Hue Sun Chan. Sequence-specific polyampholyte phase separation in membraneless organelles. *Physical review letters*, 117(17):178101, 2016.
- [5] Yi-Hsuan Lin, Jianhui Song, Julie D Forman-Kay, and Hue Sun Chan. Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. *Journal of Molecular Liquids*, 228:176–193, 2017.
- [6] Timothy J Nott, Evangelia Petsalaki, Patrick Farber, Dylan Jarvis, Eden Fussner, Anne Plochowitz, Timothy D Craggs, David P Bazett-Jones, Tony Pawson, Julie D Forman-Kay, et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell*, 57(5):936–947, 2015.
- [7] M Rubenstein and RH Colby. *Polymer physics*: Oxford university press, 2003.