



SAPIENZA  
UNIVERSITÀ DI ROMA

Facoltà di Scienze Matematiche, Fisiche e Naturali  
Corso di Laurea Magistrale in Fisica

ID number

Advisor

Co-Advisor

Academic Year

---

Sapienza University of Rome

© . All rights reserved

This thesis has been typeset by  $\text{\LaTeX}$  and the Sapthesis class.

Author's email:





# Contents



# Introduction

Le proteine rappresentano una classe di biomolecole essenziali alla vita, costituendo la base funzionale e strutturale di tutti gli organismi viventi. Composte da catene di amminoacidi connesse tramite legami peptidici, esse si distinguono per la loro capacità di assumere una straordinaria varietà di conformazioni tridimensionali. Questa caratteristica conferisce alle proteine una flessibilità funzionale senza precedenti, permettendo loro di partecipare a una vasta gamma di processi biologici.

Tra le principali funzioni delle proteine si annoverano:

- **Catalisi enzimatica:** Le proteine enzimatiche accelerano le reazioni chimiche, riducendo l'energia di attivazione e regolando i processi metabolici cellulari.
- **Ruoli strutturali:** Alcune proteine, come il collagene o la cheratina, forniscono supporto meccanico e integrità strutturale ai tessuti.
- **Trasporto:** Molecole come l'emoglobina facilitano il trasporto di ossigeno, mentre altre proteine trasportano nutrienti e ioni attraverso le membrane cellulari.
- **Regolazione e segnalazione:** Le proteine svolgono ruoli chiave nella trasduzione dei segnali cellulari e nella regolazione dell'espressione genica.

Le proteine, dunque, sono sia i costruttori sia i regolatori della vita, operando con precisione nei vari contesti fisiologici.

## 0.1 Sintesi Proteica

La sintesi proteica è il processo attraverso il quale le cellule costruiscono proteine a partire dall'informazione genetica contenuta nel DNA. Questo meccanismo fondamentale è essenziale per la vita, in quanto le proteine svolgono una varietà di funzioni chiave all'interno degli organismi viventi, tra cui ruoli strutturali, catalitici, di segnalazione e trasporto. La sintesi proteica si suddivide in due fasi principali: *trascrizione* e *traduzione*, ognuna delle quali rappresenta un passo critico nel trasferimento dell'informazione genetica dal DNA alla proteina funzionale.

### 0.1.1 Fase 1: La Trascrizione

La trascrizione è il primo passo del processo, durante il quale l'informazione genetica contenuta nel DNA viene copiata in una molecola di RNA messaggero (*mRNA*).

Questo processo avviene nel nucleo delle cellule eucariotiche (o nel citoplasma delle cellule procariotiche) e può essere suddiviso nelle seguenti fasi:

1. **Inizio:** Un complesso enzimatico, chiamato RNA polimerasi, si lega a una regione specifica del DNA, detta *promotore*, situata all'inizio del gene da trascrivere.
2. **Allungamento:** L'RNA polimerasi "legge" la sequenza di nucleotidi del DNA stampo e sintetizza una molecola di mRNA complementare, utilizzando basi azotate specifiche: Adenina (A), Uracile (U, al posto della Timina presente nel DNA), Guanina (G) e Citosina (C).
3. **Terminazione:** Quando l'RNA polimerasi raggiunge una sequenza del DNA chiamata *sequenza di terminazione*, il processo si arresta e l'mRNA è rilasciato.

L'mRNA sintetizzato rappresenta una copia temporanea dell'informazione genetica, pronta per essere utilizzata durante la fase successiva.

### 0.1.2 Fase 2: La Traduzione

La traduzione è il processo attraverso il quale l'mRNA viene decodificato per costruire una proteina. Questo avviene nei ribosomi, complessi molecolari presenti nel citoplasma o associati al reticolo endoplasmatico rugoso. Le principali fasi della traduzione sono:

1. **Inizio:** La traduzione inizia quando il ribosoma si lega a una regione specifica dell'mRNA, individuando il primo codone di inizio (AUG), che codifica per l'amminoacido metionina.
2. **Allungamento:** I *tRNA* (RNA di trasporto) trasportano specifici amminoacidi al ribosoma. Ogni tRNA ha un anticodone complementare al codone dell'mRNA, garantendo che l'amminoacido corretto venga aggiunto alla catena polipeptidica in costruzione. Gli amminoacidi vengono uniti tramite legami peptidici.
3. **Terminazione:** Quando il ribosoma raggiunge uno dei codoni di stop (UAA, UAG, UGA), la sintesi si arresta, e la proteina completa viene rilasciata.

### 0.1.3 Il Codice Genetico

Il codice genetico è il sistema di corrispondenza tra le triplette di basi nucleotidiche (codoni) dell'mRNA e gli amminoacidi. Esso è:

- **Universale:** Il codice genetico è quasi identico in tutti gli organismi viventi.
- **Ridondante:** Molti amminoacidi sono codificati da più di un codone.
- **Non ambiguo:** Ogni codone codifica per un unico amminoacido o per un segnale di stop.



### 0.1.4 Importanza della Sintesi Proteica

La sintesi proteica è cruciale per la vita. Attraverso questo processo, le cellule producono enzimi, ormoni, proteine strutturali e altre molecole che regolano o facilitano tutte le funzioni vitali. L'efficienza e la precisione della sintesi proteica sono garantite da un sistema di controllo altamente evoluto, che include la correzione degli errori durante la trascrizione e la traduzione, nonché l'azione di molecole accessorie come i chaperoni, che aiutano le proteine a ripiegarsi nella loro struttura tridimensionale funzionale.

### 0.1.5 Conclusione

La sintesi proteica è uno dei processi biologici più complessi e affascinanti, che dimostra la straordinaria capacità delle cellule di convertire l'informazione genetica in molecole funzionali. Questo meccanismo non solo garantisce il mantenimento della vita, ma rappresenta anche un punto centrale per molte applicazioni in biotecnologia e medicina, dalla produzione di farmaci alla terapia genica.

## Conformazioni proteiche e modi funzionali

La struttura tridimensionale di una proteina è determinata dalla sequenza di amminoacidi (struttura primaria) e dalle interazioni chimiche che stabilizzano la struttura secondaria, terziaria e, in molti casi, quaternaria. Queste interazioni portano alla formazione di strutture funzionali in grado di riconoscere substrati specifici o interagire con altre molecole.

Un concetto fondamentale nello studio delle proteine è quello dei *modi funzionali*, ovvero le configurazioni dinamiche che una proteina può assumere per adattarsi ai cambiamenti ambientali o alle interazioni con partner molecolari. Tali cambiamenti strutturali sono spesso descritti come un insieme di microstati energetici, ciascuno con una probabilità associata.

## Distribuzioni di probabilità nelle proteine

Le fluttuazioni strutturali delle proteine possono essere studiate in termini di distribuzioni di probabilità (Probability Density Functions, PDF). Una PDF descrive la probabilità di trovare una proteina in un dato stato conformazionale o con un certo valore di energia. Le proteine non sono statiche, ma vivono in un continuo dinamico tra diversi stati, ciascuno caratterizzato da una differente energia libera.

L'analisi delle PDF trova applicazione in numerosi contesti, tra cui:

- Studio della stabilità proteica e delle transizioni conformazionali.
- Descrizione dell'affinità di legame tra proteina e ligando.
- Modellizzazione dei processi di ripiegamento proteico (*protein folding*).

## Allostericità: una proprietà emergente delle proteine

Un fenomeno straordinario legato alla dinamicità proteica è l'*allostericità*, ovvero la capacità di una proteina di trasmettere un segnale o un cambiamento conformazionale da un sito all'altro. Questo meccanismo fondamentale consente a molte proteine di svolgere le loro funzioni biologiche in modo altamente regolato, rappresentando una chiave di volta per la comprensione della biochimica cellulare e per lo sviluppo di nuovi farmaci. L'allostericità gioca un ruolo cruciale nella regolazione dell'attività enzimatica, nella trasduzione del segnale e nella comunicazione intracellulare, fungendo da "interruttore molecolare" che risponde a stimoli ambientali e intracellulari.

### Definizione e principi fondamentali dell'allostericità

L'allostericità, dal greco *allos* (altro) e *stereos* (struttura), si riferisce a un fenomeno in cui l'interazione di una molecola (effettore) con un sito specifico di una proteina, noto come *sito allosterico*, provoca una modifica conformazionale che influenza l'attività funzionale di un altro sito, solitamente il *sito attivo*. Questo processo non avviene attraverso interazioni dirette tra i due siti, ma mediante cambiamenti nella rete di interazioni intramolecolari che regolano la struttura e la dinamica della proteina.

Dal punto di vista termodinamico, l'allostericità può essere descritta come una modulazione della distribuzione dei microstati energetici di una proteina. In altre parole, l'interazione con un effettore allosterico altera l'insieme degli stati conformazionali disponibili, facilitando o inibendo l'accesso a configurazioni funzionali.

Matematicamente, il fenomeno può essere rappresentato come una variazione nella popolazione dei microstati  $\{s_i\}$  descritta da una distribuzione di Boltzmann ponderata:

$$P(s_i) = \frac{e^{-\Delta G(s_i)/k_B T}}{\sum_j e^{-\Delta G(s_j)/k_B T}}$$

dove  $\Delta G(s_i)$  rappresenta l'energia libera associata al microstato  $s_i$ ,  $k_B$  è la costante di Boltzmann e  $T$  è la temperatura assoluta.

L'effetto allosterico può essere quantificato anche considerando il cambiamento dell'energia libera associata all'interazione con l'effettore. Se si considera una proteina con due stati principali,  $R$  (rilassato) e  $T$  (teso), l'equilibrio allosterico può essere espresso in termini di costante di equilibrio:

$$L = \frac{[T]}{[R]}$$

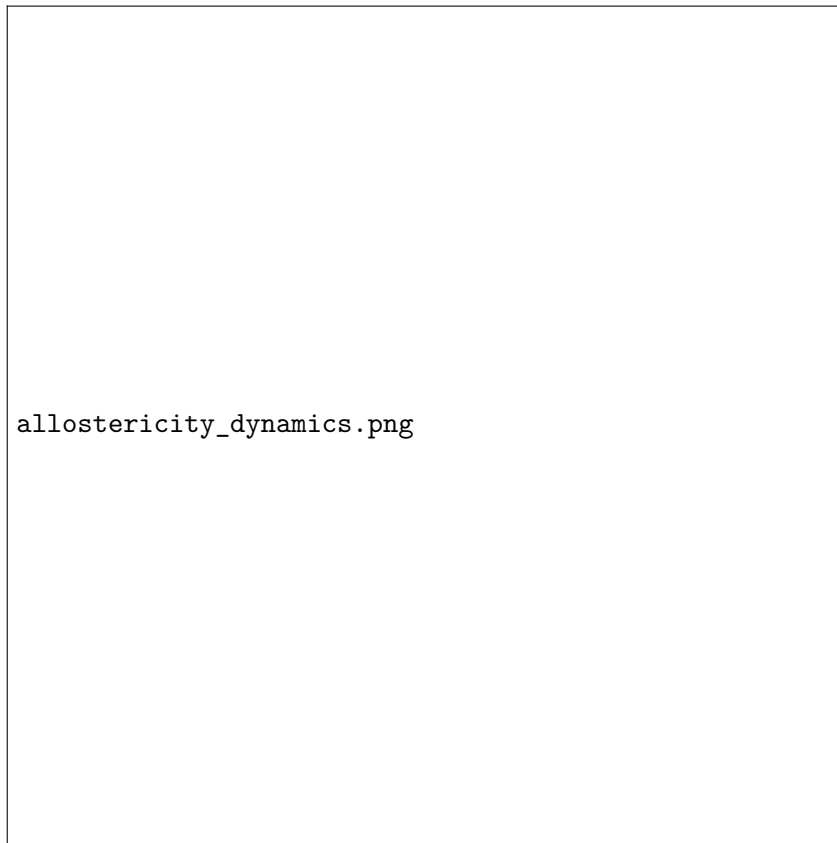
dove  $[T]$  e  $[R]$  rappresentano le concentrazioni relative dei due stati conformazionali. La presenza di un effettore modifica  $L$ , stabilizzando uno stato rispetto all'altro.

Questi modelli forniscono una base quantitativa per comprendere come le interazioni allosteriche modulino le funzioni biologiche delle proteine, come nel caso dell'emoglobina, che sfrutta l'allostericità per modulare il trasporto dell'ossigeno.

## Allostericità e dinamica proteica

La moderna comprensione dell'allostericità si basa sempre più sull'idea che il fenomeno sia intrinsecamente legato alla dinamicità proteica. Le proteine non sono strutture rigide, ma ensemble di stati conformazionali che fluttuano su scale temporali e spaziali. Gli effettori allosterici alterano questa rete dinamica, rimodellando i percorsi energetici che collegano i vari stati funzionali. Le proteine possono essere rappresentate come un insieme di stati conformazionali  $\{C_i\}$ , dove  $i$  indica un particolare stato conformazionale. La distribuzione delle popolazioni degli stati è descritta da una funzione di partizione  $Z = \sum_i e^{-\beta E_i}$ , dove  $E_i$  è l'energia associata allo stato  $C_i$  e  $\beta = \frac{1}{k_B T}$  è il fattore termico.

Un effettore allosterico può essere rappresentato come una perturbazione che modifica le energie degli stati conformazionali  $\{E_i\} \rightarrow \{E'_i\}$ , rimodellando i percorsi energetici e alterando la probabilità  $P_i$  di ogni stato, dove  $P_i = \frac{e^{-\beta E'_i}}{Z'}$ , con  $Z' = \sum_i e^{-\beta E'_i}$ .



**Figure 0.1.** Rappresentazione schematica dell'allostericità: l'interazione con un effettore allosterico (sito A) induce un cambiamento conformazionale che influenza un sito distante (sito B).

Questa visione dinamica si integra con approcci computazionali, come la dinamica molecolare e le reti di interazioni residue, che permettono di identificare percorsi specifici attraverso i quali l'informazione allosterica si propaga all'interno della

proteina.

### Esempi biologici di allostericità

L'emoglobina, una proteina presente nei globuli rossi, è un classico esempio di allostericità.

Essa trasporta ossigeno dai polmoni ai tessuti e il suo funzionamento dipende da cambiamenti conformazionali che influenzano la capacità di legare l'ossigeno.

L'emoglobina può assumere due stati principali:

- **\*\*Stato T (teso)\*\***: con bassa affinità per l'ossigeno.
- **\*\*Stato R (rilassato)\*\***: con alta affinità per l'ossigeno.

Quando una molecola di ossigeno si lega, l'intera struttura della proteina cambia, passando gradualmente dallo stato T allo stato R.

Questo fenomeno, chiamato **\*\*cooperatività\*\***, rende più facile il legame delle successive molecole di ossigeno.

Nei polmoni, dove l'ossigeno è abbondante, l'emoglobina assume lo stato R per caricarsi.

Nei tessuti, invece, fattori come l'anidride carbonica e i protoni stabilizzano lo stato T, favorendo il rilascio dell'ossigeno.

L'allostericità consente all'emoglobina di adattarsi dinamicamente alle necessità dell'organismo, catturando e rilasciando ossigeno in modo efficiente.

- **Emoglobina:** Uno degli esempi più noti, l'allostericità nell'emoglobina consente il rilascio cooperativo dell'ossigeno ai tessuti. Gli ioni  $H^+$  e il 2,3-bisfosfoglicerato agiscono come effettori negativi, stabilizzando lo stato T e favorendo il rilascio dell'ossigeno.
- **Recettori accoppiati a proteine G (GPCR):** Questi recettori trasducono segnali extracellulari attraverso meccanismi allosterici che coinvolgono cambiamenti conformazionali tra il dominio extracellulare e quello intracellulare.
- **Enzimi regolatori:** Molti enzimi, come l'aspartato transcarbamilasi, sono regolati allostericamente per coordinare le vie metaboliche in risposta ai bisogni cellulari.

### Implicazioni dei rapporti di causalità dell'allostericità nella cura del corpo umano

L'allostericità è un principio fondamentale che regola il comportamento delle proteine e degli enzimi nel corpo umano, permettendo una risposta dinamica ai cambiamenti ambientali e fisiologici. Comprendere i rapporti di causalità nei meccanismi allosterici offre nuove prospettive per migliorare la diagnosi, la prevenzione e il trattamento di molte malattie. Tra le principali implicazioni si annoverano:

- **Sviluppo di terapie allosteriche mirate:** Identificando i circuiti di regolazione allosterica alterati in specifiche malattie, è possibile progettare farmaci che modulano selettivamente i percorsi disfunzionali. Per esempio, modulatori allosterici dell'emoglobina possono migliorare il trasporto di ossigeno nei pazienti affetti da anemia falciforme.

- **Prevenzione delle malattie neurodegenerative:** Nelle patologie come Alzheimer e Parkinson, disfunzioni nei meccanismi allosterici possono portare all'aggregazione patologica di proteine. Interventi farmacologici che stabilizzano le forme native delle proteine attraverso la modulazione allosterica rappresentano una promettente strategia terapeutica.
- **Regolazione del metabolismo:** Molti enzimi metabolici sono regolati allostericamente per mantenere l'equilibrio energetico cellulare. La correzione di disfunzioni allosteriche in enzimi chiave, come la glucokinasi nel diabete, può migliorare significativamente la gestione della glicemia.
- **Riduzione degli effetti collaterali:** I farmaci allosterici, agendo su siti regolatori meno conservati, possono ridurre il rischio di effetti collaterali rispetto ai farmaci che colpiscono i siti attivi. Questo approccio è particolarmente utile in oncologia, dove una maggiore selettività è cruciale.
- **Personalizzazione della medicina:** La comprensione dei rapporti causali tra l'allostericità e la funzione proteica può guidare la medicina di precisione, permettendo di modulare specifici percorsi regolatori in base alle esigenze molecolari di ciascun paziente.

L'analisi dei rapporti di causalità nell'allostericità non solo approfondisce la comprensione della biologia molecolare, ma rappresenta un ponte verso terapie innovative, migliorando la qualità della vita e aprendo la strada a soluzioni terapeutiche più efficaci e sicure.

## Conclusioni

L'allostericità è una proprietà emergente delle proteine che illustra la complessità e l'eleganza della regolazione molecolare nei sistemi biologici. Grazie alla combinazione di approcci sperimentali e computazionali, stiamo iniziando a comprendere meglio i meccanismi alla base di questo fenomeno, con promettenti applicazioni in campo medico e biotecnologico. Continuare a esplorare l'allostericità non solo ci aiuterà a rispondere a domande fondamentali sulla biologia molecolare, ma aprirà anche nuove strade per l'innovazione terapeutica.

## La proteina PDZ: una panoramica

La proteina PDZ rappresenta una delle classi più studiate di domini proteici modulari, essenziali per il mantenimento dell'organizzazione spaziale e funzionale nelle cellule. Il nome "PDZ" deriva dai primi tre membri della famiglia in cui è stato identificato: *Post-synaptic density protein 95 (PSD-95)*, *Discs large tumor suppressor (Dlg1)* e *Zona occludens-1 (ZO-1)*. I domini PDZ sono fondamentali per la formazione di complessi multiproteici e per l'ancoraggio di proteine bersaglio a regioni specifiche della membrana cellulare.

## Struttura e caratteristiche principali del dominio PDZ

Il dominio PDZ è composto da circa 80-90 residui amminoacidici, che si organizzano in una struttura a barile  $\beta$  ( *$\beta$ -sandwich*) formata da sei filamenti  $\beta$  alternati e due  $\alpha$ -eliche. Questa conformazione crea una tasca idrofobica specifica per il legame con i ligandi, tipicamente localizzati nelle regioni C-terminali delle proteine target.

Le caratteristiche principali del dominio PDZ includono:

- **Specificità di legame:** Il dominio PDZ riconosce sequenze specifiche di 4-5 amminoacidi, note come *motivi PDZ-binding*, situate alla fine delle proteine bersaglio.
- **Modularità:** I domini PDZ possono essere combinati in proteine più grandi, consentendo una vasta gamma di interazioni con altre proteine.
- **Versatilità funzionale:** Grazie alla loro struttura conservata e al riconoscimento sequenza-specifico, i domini PDZ partecipano a processi chiave come la segnalazione cellulare, l'organizzazione sinaptica e la polarità cellulare.

## Funzioni biologiche del dominio PDZ

I domini PDZ giocano un ruolo centrale in diversi processi biologici. Tra le loro funzioni principali si annoverano:

### 1. Organizzazione della segnalazione cellulare

Il dominio PDZ funge da piattaforma di ancoraggio per i complessi proteici coinvolti nella trasduzione del segnale. Ad esempio, proteine PDZ come PSD-95 si associano ai recettori del glutammato nei neuroni, contribuendo alla formazione di sinapsi funzionali e alla modulazione della plasticità sinaptica.

### 2. Mantenimento della polarità cellulare

Le proteine contenenti domini PDZ, come ZO-1 e Dlg1, sono fondamentali per mantenere la polarità apico-basolaterale nelle cellule epiteliali. Attraverso l'interazione con le giunzioni strette, queste proteine assicurano una corretta compartimentalizzazione delle membrane cellulari.

### 3. Traffico di proteine e membrana

I domini PDZ regolano il trasporto intracellulare di proteine bersaglio verso compartimenti specifici. Un esempio è rappresentato dalla proteina NHERF1, che interagisce con il recettore della paratiroide per regolare il riassorbimento del calcio nei reni.

### 4. Sinapsi e plasticità neuronale

Nei neuroni, i domini PDZ sono coinvolti nell'organizzazione delle proteine postsinaptiche. La formazione di complessi proteici attraverso interazioni PDZ-ligando è essenziale per la stabilizzazione dei recettori di membrana e per la trasmissione sinaptica efficace.

## Il dominio PDZ e la dinamica allosterica

Un aspetto emergente nello studio delle proteine PDZ è la loro capacità di esibire dinamiche allosteriche. Sebbene il dominio PDZ sia relativamente piccolo, presenta una notevole plasticità conformazionale che gli consente di rispondere a stimoli molecolari attraverso cambiamenti strutturali.

### Meccanismo di allostericità

L'allostericità nel dominio PDZ si manifesta quando il legame di un ligando nella tasca canonica induce modifiche conformazionali che influenzano la capacità di legare altri partner molecolari. Questo fenomeno è particolarmente evidente nei domini PDZ delle proteine multi-PDZ, dove l'interazione con un ligando può modulare la funzione di altri domini.

### Implicazioni biologiche

L'allostericità nei domini PDZ è cruciale per processi come:

- *Coordinazione sinaptica*: Le modifiche conformazionali nei domini PDZ di PSD-95 regolano l'associazione con recettori postsinaptici.
- *Risposta a stimoli esterni*: Il dominio PDZ può agire come sensore molecolare, adattando la sua funzione a variazioni ambientali o fisiologiche.

## 0.2 Introduzione

La proteina **PTP1E** (*Protein Tyrosine Phosphatase 1E*), nota anche come *PTPL1*, è una fosfatasi tirosinica umana che svolge un ruolo fondamentale nella regolazione della segnalazione cellulare. Essa contiene diversi domini PDZ, moduli proteici responsabili delle interazioni con altre proteine, essenziali per il mantenimento dell'organizzazione funzionale e strutturale della cellula.

Le strutture **3LNX** e **3LNY**, depositate nel *Protein Data Bank* (PDB), rappresentano il *secondo dominio PDZ* di questa proteina, studiato in due contesti diversi: 3LNX si riferisce al dominio isolato, mentre 3LNY lo descrive in complesso con un peptide derivato da RA-GEF2 (*Rap Guanine Nucleotide Exchange Factor 2*).

## 0.3 Domini PDZ

I domini PDZ sono regioni modulari che riconoscono sequenze specifiche di altre proteine, spesso alla loro estremità carbossilica. Questi domini:

- Facilitano la formazione di complessi multiproteici.
- Ancorano proteine bersaglio a regioni specifiche della cellula.
- Regolano processi critici come la segnalazione cellulare e la polarità.

## 0.4 Struttura 3LNX: Dominio PDZ2 isolato

La struttura **3LNX** rappresenta il dominio PDZ2 della proteina PTP1E studiato in forma isolata. Determinata tramite **diffrazione a raggi X** con una risoluzione di **1,64 Å**, questa struttura è stata analizzata per comprendere le sue proprietà intrinseche, tra cui:

- La conformazione del dominio in assenza di ligandi.
- Le dinamiche strutturali che lo rendono adatto a legarsi a specifici ligandi.

### 0.4.1 Dettagli tecnici

- **Organismo:** *Homo sapiens*.
- **Metodo:** Diffrazione a raggi X.
- **Risoluzione:** 1,64 Å.
- **Data di deposito:** 3 febbraio 2010.
- **Autori:** Zhang, J., Chang, A., Ke, H., Phillips Jr., G.N., Lee, A.L. (Center for Eukaryotic Structural Genomics, CESSG).

## 0.5 Struttura 3LNY: Dominio PDZ2 in complesso con un peptide

La struttura **3LNY** rappresenta il dominio PDZ2 di PTP1E in complesso con un peptide derivato dalla proteina RA-GEF2. Determinata tramite **diffrazione a raggi X** con una risoluzione di **1,30 Å**, questa struttura ha permesso di studiare le interazioni molecolari tra il dominio PDZ e il suo ligando.

### 0.5.1 Analisi della struttura 3LNY

L'analisi della struttura ha rivelato che:

- Il dominio PDZ2 mantiene una conformazione simile a quella osservata nella struttura isolata (3LNX).
- Cambiamenti significativi avvengono nella *dinamica delle catene laterali* degli amminoacidi, suggerendo un meccanismo allosterico basato sulla flessibilità.
- La specificità del legame è mediata da interazioni chiave tra il dominio PDZ2 e il peptide di RA-GEF2.

### 0.5.2 Dettagli tecnici

- **Organismo:** *Homo sapiens*.
- **Metodo:** Diffrazione a raggi X.



- **Risoluzione:** 1,30 Å.
- **Data di deposito:** 3 febbraio 2010.
- **Data di rilascio:** 23 marzo 2010.
- **Autori:** Zhang, J., Chang, A., Ke, H., Phillips Jr., G.N., Lee, A.L. (Center for Eukaryotic Structural Genomics, CESC).

## 0.6 Significato biologico

Le strutture 3LNX e 3LNY forniscono informazioni cruciali sui meccanismi di funzionamento del dominio PDZ2:

- **3LNX** ha permesso di comprendere la conformazione intrinseca del dominio PDZ2.
- **3LNY** ha evidenziato le dinamiche molecolari e le interazioni specifiche con un peptide ligando.

Questi studi offrono una base per comprendere meglio il ruolo del dominio PDZ2 nella segnalazione cellulare e nella formazione di complessi proteici.

## 0.7 Conclusioni

Lo studio delle strutture 3LNX e 3LNY ha migliorato la nostra comprensione delle proprietà strutturali e funzionali del dominio PDZ2 nella proteina PTP1E. Tali conoscenze sono fondamentali per esplorare il ruolo di questa proteina nella regolazione cellulare e nella sua potenziale implicazione in malattie.

Per ulteriori dettagli, consulta il *Protein Data Bank* ai seguenti link:

- 3LNX: <https://www.rcsb.org/structure/3LNX>.
- 3LNY: <https://www.rcsb.org/structure/3LNY>.

## Implicazioni biomediche delle proteine PDZ

Le proteine contenenti domini PDZ sono coinvolte in numerose patologie, tra cui:

- **Tumori:** Alterazioni nelle interazioni PDZ-ligando possono contribuire alla perdita di polarità cellulare e alla progressione tumorale.
- **Malattie neurodegenerative:** Le disfunzioni nei complessi PDZ sono associate a patologie come la malattia di Alzheimer, dove la segnalazione sinaptica è compromessa.
- **Malattie cardiovascolari:** Domini PDZ regolano i canali ionici e i recettori coinvolti nel controllo della pressione sanguigna e nella contrattilità cardiaca.

## Conclusioni

Il dominio PDZ rappresenta un modello ideale per studiare le interazioni proteina-proteina e i meccanismi allosterici alla base delle funzioni cellulari. La sua modularità, specificità di legame e versatilità funzionale lo rendono un bersaglio fondamentale per lo sviluppo di terapie innovative volte a modulare l'attività proteica in contesti patologici.

## 0.8 Introduzione

I domini **PDZ**, come quelli descritti dalle strutture **3LNX** e **3LNY**, sono moduli proteici che giocano un ruolo cruciale nella regolazione delle interazioni proteina-proteina e nella trasmissione dei segnali intracellulari. Comprendere il loro **meccanismo allosterico** e la loro **localizzazione nei tessuti umani** fornisce informazioni fondamentali per la biologia cellulare e per lo sviluppo di terapie mirate.

## 0.9 Meccanismo Allosterico dei Domini PDZ

Il meccanismo **allosterico** permette ai domini PDZ di trasmettere segnali a lunga distanza attraverso cambiamenti dinamici o strutturali indotti dal legame con un ligando. Questo fenomeno è fondamentale per la loro funzione:

### 0.9.1 1. Ruolo nella segnalazione cellulare

- I domini PDZ riconoscono e si legano a specifici ligandi proteici, regolando la formazione di complessi multiproteici essenziali per la trasduzione del segnale.
- Il meccanismo allosterico consente ai PDZ di adattarsi a diversi segnali chimici o fisici, influenzando la funzionalità cellulare.

### 0.9.2 2. Applicazioni terapeutiche

#### Disturbi neurologici:

- Le funzioni sinaptiche mediate dai domini PDZ di proteine come PTP1E sono cruciali per la plasticità e la comunicazione neuronale.
- Comprendere il meccanismo allosterico offre opportunità per sviluppare terapie per malattie neurodegenerative (es. Alzheimer) o disturbi neuropsichiatrici.

#### Cancro:

- La proteina PTP1E agisce come regolatore negativo della crescita cellulare. Alterazioni nei domini PDZ possono favorire la proliferazione tumorale.
- Modulatore allosterici potrebbero rappresentare una strategia innovativa per il trattamento del cancro.

## 0.10 Localizzazione dei Domini PDZ nel Corpo Umano

I domini PDZ, inclusi quelli della proteina **PTP1E**, sono espressi in diversi tessuti e sistemi del corpo umano. La loro localizzazione è strettamente legata alla funzione biologica della proteina.

### 0.10.1 1. Sistema Nervoso Centrale

- La proteina PTP1E, contenente il dominio PDZ2, è abbondantemente espressa nei **neuroni**.
- È coinvolta nella formazione e regolazione delle sinapsi, contribuendo alla plasticità sinaptica e alla trasmissione dei segnali tra i neuroni.
- Le sue interazioni mediante PDZ sono essenziali per l'organizzazione spaziale dei recettori sinaptici e delle molecole di segnalazione.

### 0.10.2 2. Tessuti Epiteliali

- Nei tessuti epiteliali (es. pelle, intestino, fegato), i domini PDZ regolano la **polarità cellulare** e l'adesione tra le cellule.
- Contribuiscono al mantenimento della barriera epiteliale e alla comunicazione intercellulare, processi cruciali per l'integrità tissutale.

### 0.10.3 3. Sistema Immunitario

- La proteina PTP1E è espressa anche in alcune cellule immunitarie, dove i suoi domini PDZ modulano i segnali di attivazione cellulare.
- Queste funzioni sono importanti per la regolazione delle risposte immunitarie.

### 0.10.4 4. Altri tessuti

- I domini PDZ sono presenti in cellule di tessuti muscolari e connettivi, dove contribuiscono a mantenere la stabilità strutturale e a regolare la trasduzione del segnale meccanico.

## 0.11 Significato Biologico e Applicazioni

Lo studio delle strutture 3LNX e 3LNY ha fornito importanti informazioni sul funzionamento dei domini PDZ e sulle loro interazioni con ligandi. Questo ha implicazioni fondamentali in:

- **Biologia cellulare:** Comprendere il ruolo dei domini PDZ nella regolazione della crescita, adesione e segnalazione cellulare.
- **Medicina:** Sviluppare farmaci mirati per modulare l'attività di proteine contenenti PDZ, con applicazioni per il trattamento di malattie neurologiche, tumorali e autoimmuni.

## 0.12 Conclusioni

I domini PDZ, come quelli descritti da 3LNX e 3LNY, sono fondamentali per l'organizzazione e la funzione cellulare in diversi tessuti del corpo umano. Il loro meccanismo allosterico e la loro capacità di mediare interazioni proteina-proteina li rendono bersagli promettenti per lo sviluppo di terapie innovative.

Per ulteriori dettagli sulle strutture, consulta il *Protein Data Bank* ai seguenti link:

- 3LNX: <https://www.rcsb.org/structure/3LNX>.
- 3LNY: <https://www.rcsb.org/structure/3LNY>.

## 0.13 Fondamenti dei Modelli GNM

I *Gaussian Network Models* (GNM) rappresentano un approccio fondamentale per lo studio delle fluttuazioni isotropiche delle proteine. Originariamente proposti come un'estensione dei modelli di reti elastiche (*Elastic Network Models*, ENM), i GNM sfruttano la semplicità matematica delle matrici di adiacenza per descrivere le dinamiche globali delle proteine con un costo computazionale ridotto.

Questa sezione si focalizza sull'analisi matematica e sull'applicazione dei GNM, evidenziando i miglioramenti introdotti rispetto ai modelli ENM tradizionali e discutendo il loro impatto nello studio delle proprietà dinamiche proteiche come i fattori  $B$  cristallografici e le transizioni conformazionali.

I GNM rappresentano una proteina come una rete elastica di nodi (i residui proteici), dove ogni nodo è connesso ai suoi vicini entro una certa distanza di cutoff. L'energia potenziale totale della rete è data da:

$$U = \frac{\gamma}{2} \sum_{i,j} \Gamma_{ij} \|\mathbf{r}_i - \mathbf{r}_j\|^2, \quad (0.1)$$

dove:

- $\gamma$  è la costante di elasticità (uniforme per tutte le connessioni),
- $\Gamma_{ij}$  è l'elemento della matrice di Kirchhoff, che rappresenta la connettività della rete,
- $\mathbf{r}_i$  è la posizione del nodo  $i$ .

La matrice di Kirchhoff è definita come:

$$\Gamma_{ij} = \begin{cases} -1 & \text{se } i \neq j \text{ e i nodi sono connessi,} \\ 0 & \text{se } i \neq j \text{ e i nodi non sono connessi,} \\ \sum_{k \neq i} |\Gamma_{ik}| & \text{se } i = j. \end{cases} \quad (0.2)$$

### 0.13.1 Fluttuazioni Isotropiche e Fattori B

Le fluttuazioni isotropiche medie di ciascun residuo  $i$  sono proporzionali alla diagonale della pseudo-inversa della matrice di Kirchhoff:

$$\langle \Delta \mathbf{r}_i^2 \rangle = k_B T [\Gamma^+]_{ii}, \quad (0.3)$$

dove:

- $k_B$  è la costante di Boltzmann,
- $T$  è la temperatura,
- $\Gamma^+$  è la pseudo-inversa di  $\Gamma$ .

I fattori  $B$  cristallografici, che rappresentano le fluttuazioni termiche osservate sperimentalmente, sono correlati a  $\langle \Delta \mathbf{r}_i^2 \rangle$  tramite:

$$B_i = \frac{8\pi^2}{3} \langle \Delta \mathbf{r}_i^2 \rangle. \quad (0.4)$$

### 0.13.2 Dipendenze dalla Distanza

Un miglioramento significativo rispetto ai modelli GNM tradizionali è l'introduzione di interazioni dipendenti dalla distanza:

$$k(r) \propto \frac{1}{r^\alpha}, \quad (0.5)$$

dove  $\alpha$  è un parametro che può variare a seconda delle condizioni del sistema:

- $\alpha = 2$  per piccole deformazioni,
- $\alpha = 6$  o  $7$  per deformazioni maggiori.

Questo approccio rimuove la necessità di un cutoff esplicito, aumentando la predittività dei modelli e migliorando la corrispondenza con i dati sperimentali.

## 0.14 Analisi delle Transizioni Conformazionali

I modelli GNM non solo predicono le fluttuazioni locali, ma catturano anche i modi collettivi di movimento che spesso corrispondono a transizioni conformazionali funzionali. Questi modi normali sono ottenuti diagonalizzando la matrice di Kirchhoff:

$$\Gamma = Q \Lambda Q^T, \quad (0.6)$$

dove:

- $Q$  è la matrice dei vettori propri,
- $\Lambda$  è la matrice diagonale dei valori propri.

I modi con valori propri piccoli rappresentano movimenti collettivi su larga scala, essenziali per comprendere i cambiamenti conformazionali.

## 0.15 Validazioni Sperimentali

Il confronto tra i fattori  $B$  calcolati ( $B_{\text{calc}}$ ) e quelli sperimentali ( $B_{\text{exp}}$ ) è una misura diretta dell'accuratezza del modello. La correlazione è espressa come:

$$R = \frac{\sum_i (B_{\text{exp},i} - \bar{B}_{\text{exp}})(B_{\text{calc},i} - \bar{B}_{\text{calc}})}{\sqrt{\sum_i (B_{\text{exp},i} - \bar{B}_{\text{exp}})^2 \sum_i (B_{\text{calc},i} - \bar{B}_{\text{calc}})^2}}. \quad (0.7)$$

Gli studi hanno dimostrato che modelli senza cutoff producono valori  $R$  più elevati rispetto ai modelli tradizionali.

## 0.16 Applicazioni Biologiche

I GNM sono utilizzati per:

- Predire le regioni più mobili nelle proteine.
- Identificare i modi di movimento funzionali correlati a transizioni conformazionali.
- Interpretare gli effetti di mutazioni sulle dinamiche proteiche.
- Analizzare i cambiamenti indotti da ligandi, utilizzando un'estensione dei GNM che include regolatori allosterici.

## 0.17 Conclusioni

I *Gaussian Network Models* forniscono una base matematica semplice ed efficace per studiare le dinamiche delle proteine. L'introduzione di interazioni dipendenti dalla distanza rappresenta un significativo passo avanti, eliminando parametri arbitrari e migliorando la predizione di fenomeni biologici complessi.

## 0.18 La Teoria dei Network e la Matrice di Kirchhoff nei Gaussian Network Models

I *Gaussian Network Models* (GNM) si basano sulla rappresentazione delle proteine come *grafi non orientati*. In questa formulazione, i residui proteici (o altre unità strutturali) sono rappresentati come nodi del grafo, mentre le interazioni tra di essi, definite entro una soglia di distanza  $r_c$ , sono rappresentate come archi. La teoria dei grafi fornisce un linguaggio matematico potente per modellare queste relazioni, dove la connettività della rete è catturata dalla *matrice di adiacenza*  $A$  e dalla *matrice di Kirchhoff*  $\Gamma$ , che gioca un ruolo centrale nei GNM.

### 0.18.1 Matrice di Adiacenza e Matrice di Kirchhoff

La *matrice di adiacenza*  $A$  è una matrice simmetrica  $N \times N$ , dove  $N$  è il numero di nodi (residui). Gli elementi di  $A$  sono definiti come:

$$A_{ij} = \begin{cases} 1 & \text{se } \|\mathbf{r}_i^0 - \mathbf{r}_j^0\| \leq r_c \text{ e } i \neq j, \\ 0 & \text{altrimenti.} \end{cases} \quad (0.8)$$

Qui,  $r_c$  rappresenta la distanza di cutoff entro la quale due nodi sono considerati connessi.

La *matrice di Kirchhoff*  $\Gamma$ , detta anche matrice Laplaciana del grafo, è definita come:

$$\Gamma_{ij} = \begin{cases} -A_{ij} & \text{se } i \neq j, \\ \sum_k A_{ik} & \text{se } i = j. \end{cases} \quad (0.9)$$

### 0.18.2 Proprietà della Matrice di Kirchhoff

La matrice di Kirchhoff ha le seguenti proprietà fondamentali:

- **Simmetria:**  $\Gamma$  è simmetrica poiché  $A$  è simmetrica.
- **Valore proprio nullo:** Il più piccolo valore proprio di  $\Gamma$  è zero, corrispondente al vettore proprio costante  $\mathbf{1}$ , riflettendo il fatto che il grafo è connesso.
- **Pseudo-inversa:** La pseudo-inversa  $\Gamma^+$  è utilizzata per calcolare le fluttuazioni isotropiche:

$$\langle \Delta \mathbf{r}_i^2 \rangle = k_B T [\Gamma^+]_{ii}. \quad (0.10)$$

### 0.18.3 Connessione ai Modelli GNM

La matrice di Kirchhoff è essenziale per il calcolo dei modi normali e delle fluttuazioni isotropiche nei GNM. Diagonalizzando  $\Gamma$ , otteniamo:

$$\Gamma = Q \Lambda Q^T, \quad (0.11)$$

dove  $Q$  è la matrice dei vettori propri e  $\Lambda$  è la matrice diagonale dei valori propri. I valori propri piccoli rappresentano modi di movimento collettivi, mentre i valori propri grandi rappresentano fluttuazioni locali.

I *Gaussian Network Models* sfruttano questa struttura per predire le proprietà dinamiche delle proteine, come le fluttuazioni atomiche ( $\langle \Delta \mathbf{r}_i^2 \rangle$ ) e i fattori  $B$ :

$$B_i = \frac{8\pi^2}{3} \langle \Delta \mathbf{r}_i^2 \rangle. \quad (0.12)$$

## Chapter 1

# Modi normali nelle proteine

### Introduzione ai modi normali

La teoria dei modi normali (Normal Mode Analysis, NMA) è una tecnica fondamentale per analizzare le fluttuazioni dinamiche delle proteine a livello atomico. Questa metodologia consente di studiare le oscillazioni collettive delle strutture tridimensionali delle proteine attorno a uno stato di equilibrio, modellando il sistema come una rete di oscillatori armonici. I modi normali rappresentano i movimenti collettivi indipendenti che compongono la dinamica della proteina.

Matematicamente, la teoria dei modi normali si basa sull'approssimazione di un minimo locale della superficie di energia potenziale. In questa sezione, verrà introdotto il formalismo matematico alla base dell'NMA, con esempi applicati alla dinamica delle proteine.

### 1.1 Il contributo unico della meccanica statistica nello studio delle proteine

#### 1.1.1 Prospettive complementari: fisica e biologia

Le proteine sono macchine molecolari fondamentali per la vita, responsabili di una vasta gamma di funzioni biologiche. Lo studio delle proteine è tradizionalmente dominato dalla biologia, che offre una comprensione dettagliata dei meccanismi biochimici e strutturali. Tuttavia, l'approccio della fisica, in particolare attraverso la meccanica statistica, fornisce strumenti e prospettive che sono complementari e spesso innovative rispetto a quelli biologici.

#### Approccio della biologia

La biologia si concentra principalmente su:

- La struttura atomistica delle proteine, studiata attraverso tecniche sperimentali come la cristallografia a raggi X, la spettroscopia NMR e la microscopia crioelettronica.
- I meccanismi biochimici delle interazioni proteina-ligando.



- La regolazione genetica e cellulare dell'espressione proteica.

Questi approcci forniscono informazioni essenziali ma spesso statiche, limitandosi a snapshot di una proteina in uno stato particolare.

### **Contributo della fisica**

I fisici, al contrario, applicano i principi della meccanica statistica e della teoria dei sistemi complessi per esplorare:

- Le dinamiche proteiche su scale temporali e spaziali estese.
- Le proprietà emergenti delle reti di interazioni residue.
- La natura probabilistica dei moti molecolari, andando oltre le descrizioni deterministiche.

Questo approccio permette di comprendere non solo "cosa" fa una proteina, ma anche "come" e "perché" compie determinate funzioni.

#### **1.1.2 Meccanica statistica: un linguaggio universale per sistemi complessi**

La meccanica statistica fornisce un linguaggio universale per descrivere sistemi con molte parti interagenti, come le proteine. Le principali forze dell'approccio della fisica sono:

#### **Riduzione della complessità**

Le proteine sono sistemi altamente complessi, con migliaia di atomi che interagiscono attraverso forze chimiche, elettriche e meccaniche. I fisici utilizzano modelli semplificati, come il modello di rete gaussiana (GNM) o i modelli di campo medio, per ridurre la complessità senza perdere i fenomeni essenziali. Ad esempio:

$$V_{GNM} = \frac{g}{2} \sum_{i,j} K_{ij} \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j$$

rappresenta un modello semplice ma potente per descrivere le fluttuazioni residue.

#### **Approccio probabilistico**

Mentre i biologi spesso si concentrano su stati stabili (ad esempio, conformazioni native), i fisici esplorano l'intero spazio delle configurazioni possibili, utilizzando distribuzioni di probabilità per descrivere i moti:

$$P(\mathbf{x}) \propto e^{-\beta V(\mathbf{x})}$$

Questo approccio permette di studiare transizioni conformazionali rare ma biologicamente rilevanti.

### Teoremi di fluttuazione e risposta

La fisica introduce teoremi generali, come il teorema fluttuazione-dissipazione, che collega le fluttuazioni spontanee alle risposte del sistema:

$$R_{ij}(t) = C_{ij}(t)C_{ij}^{-1}(0)$$

Questi strumenti consentono di analizzare come una proteina risponde a perturbazioni esterne, con implicazioni dirette per l'allosteria e la funzione.

#### 1.1.3 Un esempio concreto: l'allosteria

L'allosteria è un fenomeno chiave per comprendere la regolazione proteica. I biologi la descrivono spesso in termini di "interruttori" conformazionali, mentre i fisici la trattano come un problema di trasporto di informazione attraverso una rete. Questo approccio consente di:

- Identificare percorsi ottimali di comunicazione tra il sito regolatore e il sito attivo.
- Quantificare l'efficienza del trasferimento di informazione usando strumenti come il trasferimento di entropia:

$$TE_{j \rightarrow i}(t) = H[x_i(t)|x_i(0)] - H[x_i(t)|x_i(0), x_j(0)]$$

- Esplorare fenomeni fuori equilibrio, che spesso giocano un ruolo cruciale nelle dinamiche allosteriche.

#### 1.1.4 Vantaggi dell'approccio fisico

Il contributo dei fisici non si limita alla modellizzazione teorica, ma comprende anche:

- La capacità di progettare esperimenti computazionali per testare ipotesi difficili da esplorare sperimentalmente.
- L'uso di simulazioni multiscala per collegare dinamiche atomistiche con fenomeni macroscopici.
- La formulazione di principi generali che possono essere applicati a proteine molto diverse tra loro.

#### 1.1.5 Collaborazione interdisciplinare

Nonostante le differenze metodologiche, fisici e biologi condividono l'obiettivo di comprendere il funzionamento delle proteine. La combinazione di tecniche sperimentali avanzate, come la microscopia crioelettronica, con modelli teorici basati sulla fisica rappresenta una delle frontiere più promettenti nella biologia molecolare.

## 1.2 Conclusioni

Il contributo della fisica allo studio delle proteine va ben oltre la semplice modellizzazione. La capacità di ridurre la complessità, formulare teoremi generali e applicare strumenti matematici avanzati offre nuove prospettive sui fenomeni biologici. In particolare, la meccanica statistica permette di esplorare dinamiche fuori equilibrio, interazioni residue e proprietà emergenti che sono difficili da affrontare con i soli strumenti della biologia tradizionale. Questo approccio interdisciplinare non solo arricchisce la nostra comprensione teorica, ma apre anche nuove possibilità per applicazioni pratiche, come la progettazione di farmaci e la bioingegneria.

## 1.3 La meccanica statistica del non equilibrio: una finestra sulla realtà biologica

### 1.3.1 Differenze tra equilibrio e non equilibrio

La meccanica statistica tradizionale si basa sull'assunto che i sistemi raggiungano uno stato di equilibrio termodinamico, caratterizzato dalla distribuzione di Boltzmann:

$$P(\mathbf{x}) \propto e^{-\beta V(\mathbf{x})}$$

dove  $\beta = (k_B T)^{-1}$  è il fattore termico e  $V(\mathbf{x})$  è il potenziale del sistema. In questa condizione, le proprietà macroscopiche possono essere dedotte dalle distribuzioni stazionarie e le fluttuazioni sono bilanciate da processi dissipativi.

Tuttavia, la maggior parte dei sistemi biologici, incluse le proteine, opera lontano dall'equilibrio. Questi sistemi sono continuamente alimentati da energia chimica (ad esempio, ATP o gradienti di concentrazione), generando comportamenti dinamici che non possono essere descritti adeguatamente dalla fisica dell'equilibrio.

### Caratteristiche dei sistemi fuori equilibrio

I sistemi fuori equilibrio presentano le seguenti caratteristiche distintive:

- **\*\*Rottura del bilancio dettagliato\*\***: Le transizioni tra stati non rispettano più la simmetria  $P(i \rightarrow j) = P(j \rightarrow i)$ , generando correnti probabilistiche.
- **\*\*Produzione di entropia\*\***: I sistemi fuori equilibrio producono entropia a un tasso non nullo, riflettendo il consumo di energia.
- **\*\*Dinamiche dipendenti dal tempo\*\***: Le distribuzioni di probabilità possono variare nel tempo, portando a stati stazionari non canonici o oscillazioni periodiche.
- **\*\*Cicli stocastici\*\***: In molti sistemi biologici, le transizioni tra stati seguono traiettorie cicliche, come nel caso delle proteine motorie.

### 1.3.2 Teoria del non equilibrio e sistemi biologici

La meccanica statistica del non equilibrio estende i concetti dell'equilibrio per includere la dinamica, la causalità e il consumo di energia. Questo approccio è particolarmente utile nello studio delle proteine, dove le interazioni non lineari e il rumore termico giocano un ruolo cruciale.

#### Distribuzioni fuori equilibrio

A differenza dell'equilibrio, dove la distribuzione di probabilità è data dalla funzione di Boltzmann, i sistemi fuori equilibrio sono caratterizzati da distribuzioni stazionarie generalizzate:

$$P_{\text{st}}(\mathbf{x}) \propto e^{-\phi(\mathbf{x})}$$

dove  $\phi(\mathbf{x})$  include contributi energetici, dissipativi e di flusso.

#### Flussi stazionari e bilancio dell'entropia

Un elemento distintivo del non equilibrio è la presenza di flussi stazionari  $\mathbf{J}(\mathbf{x})$ , definiti come:

$$\mathbf{J}(\mathbf{x}) = P_{\text{st}}(\mathbf{x})\mathbf{v}(\mathbf{x})$$

dove  $\mathbf{v}(\mathbf{x})$  è il campo di velocità probabilistica. Questi flussi indicano che il sistema è mantenuto lontano dall'equilibrio da fonti esterne di energia.

La produzione di entropia è data da:

$$\dot{S}_{\text{tot}} = \int \mathbf{J}(\mathbf{x}) \cdot \nabla \phi(\mathbf{x}) d\mathbf{x}$$

e rappresenta una misura quantitativa del grado di non equilibrio.

### 1.3.3 Applicazioni alla dinamica delle proteine

Le proteine, come l'ubiquitina, operano spesso in condizioni lontane dall'equilibrio. Questo è evidente nei seguenti aspetti:

- **Transizioni conformazionali**: Le proteine attraversano stati di alta energia per svolgere le loro funzioni, guidate da legami chimici o interazioni con ligandi.
- **Dinamiche allosteriche**: La trasmissione di segnali tra siti distanti nella proteina richiede un flusso continuo di energia e informazione.
- **Fluttuazioni amplificate**: Le dinamiche residue sono influenzate da rumore termico che agisce in modo asimmetrico a causa della struttura non equilibrata.

#### Esempio: modello di rete gaussiana fuori equilibrio

Nel modello di rete gaussiana (GNM) fuori equilibrio, l'energia include termini aggiuntivi per il flusso dissipativo:

$$V_{\text{neq}}(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{K} \mathbf{x} + \mathbf{f}_{\text{ext}} \cdot \mathbf{x}$$

dove  $\mathbf{f}_{\text{ext}}$  rappresenta forze esterne che mantengono il sistema lontano dall'equilibrio.

### 1.3.4 Non equilibrio e causalità

Uno dei contributi più importanti della meccanica statistica del non equilibrio è la possibilità di studiare la causalità in sistemi dinamici. Indicatori come il trasferimento di entropia sono particolarmente sensibili alle condizioni fuori equilibrio:

$$TE_{j \rightarrow i}(t) = H[x_i(t)|x_i(0)] - H[x_i(t)|x_i(0), x_j(0)]$$

Fuori dall'equilibrio,  $TE_{j \rightarrow i}$  incorpora informazioni aggiuntive sui flussi stazionari e sulle interazioni energetiche non conservative.

### 1.3.5 Vantaggi dell'approccio del non equilibrio

L'approccio fuori equilibrio offre numerosi vantaggi rispetto ai modelli tradizionali:

- **Realismo biologico**: Molti processi cellulari non possono essere descritti adeguatamente in equilibrio.
- **Comprensione dei meccanismi**: Le proprietà emergenti, come l'allosteria e la cooperatività, dipendono dal non equilibrio.
- **Previsioni quantitative**: Modelli non equilibrati possono predire fenomeni dinamici come cicli molecolari e fluttuazioni amplificate.

### 1.3.6 Criticità e sfide

Nonostante il suo potenziale, l'applicazione della meccanica statistica del non equilibrio presenta alcune sfide:

- **Modellizzazione complessa**: Richiede modelli più dettagliati e computazionalmente intensivi.
- **Dati sperimentali**: È spesso difficile ottenere dati quantitativi per validare i modelli.
- **Teoria generale**: A differenza dell'equilibrio, non esiste una teoria unificata per il non equilibrio.

## 1.4 Conclusioni

La meccanica statistica del non equilibrio rappresenta una frontiera fondamentale nello studio delle proteine e di altri sistemi biologici complessi. Il suo approccio dinamico e flessibile permette di superare i limiti dei modelli tradizionali, offrendo una visione più realistica della realtà biologica. In particolare, il non equilibrio fornisce strumenti essenziali per studiare la causalità, i flussi di informazione e i meccanismi emergenti in sistemi lontani dall'equilibrio termodinamico.

## 1.5 Modello matematico dei modi normali

### 1. Energia potenziale della proteina

L'energia potenziale di una proteina può essere descritta come una funzione  $V(\mathbf{x})$ , dove  $\mathbf{x} = (x_1, x_2, \dots, x_{3N})$  rappresenta le coordinate dei  $N$  atomi nella struttura tridimensionale. In prossimità di un minimo locale, l'energia può essere approssimata tramite uno sviluppo di Taylor di secondo ordine:

$$V(\mathbf{x}) \approx V(\mathbf{x}_0) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 V}{\partial x_i \partial x_j} \Delta x_i \Delta x_j,$$

dove  $\mathbf{x}_0$  è la configurazione di equilibrio e  $\Delta \mathbf{x} = \mathbf{x} - \mathbf{x}_0$  rappresenta la deviazione rispetto all'equilibrio. Il termine  $\frac{\partial^2 V}{\partial x_i \partial x_j}$  definisce la matrice hessiana  $\mathbf{H}$  dell'energia potenziale:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 V}{\partial x_1^2} & \frac{\partial^2 V}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 V}{\partial x_1 \partial x_{3N}} \\ \frac{\partial^2 V}{\partial x_2 \partial x_1} & \frac{\partial^2 V}{\partial x_2^2} & \cdots & \frac{\partial^2 V}{\partial x_2 \partial x_{3N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 V}{\partial x_{3N} \partial x_1} & \frac{\partial^2 V}{\partial x_{3N} \partial x_2} & \cdots & \frac{\partial^2 V}{\partial x_{3N}^2} \end{bmatrix}.$$

### 2. Risoluzione del problema agli autovalori

Il movimento della proteina può essere descritto tramite la seconda legge di Newton:

$$M \frac{d^2 \mathbf{x}}{dt^2} = -\mathbf{H} \Delta \mathbf{x},$$

dove  $M$  è una matrice diagonale che contiene le masse atomiche. Dividendo per  $M$  e trasformando il sistema in coordinate ridotte, si ottiene l'equazione del moto:

$$\frac{d^2 \mathbf{q}}{dt^2} = -\mathbf{H}' \mathbf{q},$$

dove  $\mathbf{H}' = M^{-1/2} \mathbf{H} M^{-1/2}$  è la matrice hessiana ridotta. La soluzione di questo sistema si ottiene diagonalizzando  $\mathbf{H}'$ :

$$\mathbf{H}' \mathbf{v}_i = \lambda_i \mathbf{v}_i,$$

dove  $\lambda_i$  sono gli autovalori (frequenze quadratiche dei modi normali) e  $\mathbf{v}_i$  sono gli autovettori (i modi normali).

## 1.6 Interpretazione fisica dei modi normali

Gli autovalori  $\lambda_i$  rappresentano le frequenze quadratiche associate ai modi normali, mentre gli autovettori  $\mathbf{v}_i$  descrivono la direzione del movimento per ciascun modo. I modi normali con autovalori piccoli (basse frequenze) corrispondono a movimenti collettivi lenti e di grande ampiezza, spesso associati a processi funzionali delle proteine, come l'apertura e chiusura di tasche enzimatiche o il movimento di domini.

## 1. Modi rigidi e vibrazionali

I primi sei modi normali, corrispondenti agli spostamenti traslazionali e rotazionali rigidi della proteina, hanno autovalori nulli o prossimi allo zero. Gli altri modi descrivono oscillazioni vibrazionali interne della proteina.

## 2. Fluttuazioni termiche

Le fluttuazioni termiche di una proteina possono essere descritte utilizzando i modi normali. La deviazione media quadratica di un atomo  $i$  in equilibrio è proporzionale all'inverso degli autovalori:

$$\langle (\Delta x_i)^2 \rangle \propto \sum_k \frac{(v_k)_i^2}{\lambda_k},$$

dove  $(v_k)_i$  è la componente del  $k$ -esimo autovettore associata all'atomo  $i$ .

## 1.7 Applicazioni della teoria dei modi normali alle proteine

### 1. Analisi della dinamica funzionale

La teoria dei modi normali è stata utilizzata per comprendere i movimenti collettivi delle proteine, come l'apertura e la chiusura delle tasche enzimatiche. Ad esempio, per l'enzima adenilato ciclasi, i modi normali a basse frequenze mostrano oscillazioni che facilitano l'accesso del substrato al sito attivo.

### 2. Calcolo dei fattori B

I fattori B, osservati nella cristallografia a raggi X, rappresentano la flessibilità atomica in una struttura proteica. Tramite NMA, è possibile prevedere i fattori B teorici:

$$B_i = \frac{8\pi^2}{3} \langle (\Delta x_i)^2 \rangle.$$

### 3. Disegno di farmaci

La dinamica delle proteine, studiata tramite NMA, è essenziale per il design di farmaci. I movimenti lenti e collettivi identificati tramite modi normali possono rivelare siti allosterici potenziali per l'intervento farmacologico.

### 4. Stabilità e ripiegamento proteico

I modi normali possono essere utilizzati per studiare il ripiegamento proteico e le transizioni conformazionali. Ad esempio, analizzando le frequenze dei modi, è possibile identificare i movimenti correlati alla destabilizzazione di strutture secondarie.

## 1.8 Modelli semplificati: elastic network model (ENM)

Per ridurre la complessità computazionale, sono stati sviluppati modelli semplificati come l'Elastic Network Model (ENM). In questi modelli, gli atomi della proteina sono rappresentati come nodi connessi da molle, e l'energia potenziale è descritta come:

$$V = \frac{1}{2} \sum_{i < j} k_{ij} (d_{ij} - d_{ij}^0)^2,$$

dove  $d_{ij}$  è la distanza tra gli atomi  $i$  e  $j$ ,  $d_{ij}^0$  è la distanza di equilibrio, e  $k_{ij}$  è la costante elastica.

L'ENM semplifica il calcolo della matrice hessiana, rendendo l'analisi dei modi normali computazionalmente più efficiente.

## 1.9 Conclusioni

La teoria dei modi normali offre una descrizione dettagliata delle dinamiche proteiche, fornendo informazioni preziose sui movimenti collettivi e sulle fluttuazioni termiche. Applicata a contesti biologici, come il design di farmaci e lo studio della funzione enzimatica, l'NMA rappresenta uno strumento potente per comprendere il comportamento dinamico delle proteine.



## Chapter 2

# Gaussian Network Models (GNM) nelle proteine

### Introduzione ai modelli di rete gaussiani

Il Gaussian Network Model (GNM) è un approccio semplificato ma efficace per analizzare le dinamiche collettive delle proteine basandosi sulla loro struttura tridimensionale. Questo modello si colloca nella categoria degli *elastic network models* (ENM) e rappresenta una potente astrazione per descrivere le fluttuazioni attorno alla configurazione nativa di una proteina. Per costruire una descrizione rigorosa, è essenziale passare da una rappresentazione generale della dinamica molecolare, basata sull'Hamiltoniana, a un sistema equivalente che utilizza molle armoniche per modellare le interazioni tra i residui.

### 2.1 Dall'Hamiltoniana alla descrizione elastica

La dinamica completa di una proteina può essere rappresentata da un'Hamiltoniana generica, che descrive l'energia totale del sistema come somma dell'energia cinetica e dell'energia potenziale:

$$H = T + V, \quad (2.1)$$

dove  $T$  rappresenta l'energia cinetica e  $V$  l'energia potenziale. Nello specifico, l'energia cinetica è espressa come:

$$T = \frac{1}{2} \sum_{i=1}^N m_i \|\dot{\mathbf{r}}_i\|^2, \quad (2.2)$$

con  $m_i$  massa del residuo  $i$ ,  $\dot{\mathbf{r}}_i$  velocità del residuo  $i$ , e  $N$  il numero totale di residui. L'energia potenziale  $V$  descrive le interazioni tra i residui e, in linea di principio, include termini complessi che considerano legami covalenti, interazioni non covalenti e forze di esclusione sterica.

### Semplificazione tramite approccio armonico

Per studiare le fluttuazioni attorno alla configurazione nativa, si assume che le deviazioni dalle posizioni di equilibrio siano piccole. In questo regime, l'energia potenziale può essere approssimata tramite un'espansione di Taylor al secondo ordine attorno alla configurazione nativa  $\{\mathbf{r}_i^0\}$ :

$$V(\{\mathbf{r}_i\}) \approx V_0 + \frac{1}{2} \sum_{i,j=1}^N \frac{\partial^2 V}{\partial \mathbf{r}_i \partial \mathbf{r}_j} \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j, \quad (2.3)$$

dove  $V_0$  è il valore dell'energia potenziale nella configurazione nativa e  $\Delta \mathbf{r}_i = \mathbf{r}_i - \mathbf{r}_i^0$  rappresenta la deviazione della posizione del residuo  $i$  rispetto all'equilibrio. Il termine di ordine zero può essere ignorato poiché non contribuisce alla dinamica, e il termine di ordine uno si annulla poiché la forza è nulla all'equilibrio. Pertanto, l'energia potenziale è dominata dal termine quadratico.

La matrice delle seconde derivate, detta matrice Hessiana  $\mathbf{H}$ , cattura la curvatura locale dell'energia potenziale:

$$H_{ij} = \frac{\partial^2 V}{\partial \mathbf{r}_i \partial \mathbf{r}_j}. \quad (2.4)$$

### Modello a molle armoniche

In un GNM, l'Hamiltoniana viene ulteriormente semplificata assumendo che i residui siano collegati da molle armoniche, con una costante elastica uniforme  $\gamma$ . In questo modello, l'energia potenziale è scritta come:

$$V(\{\Delta \mathbf{r}_i\}) = \frac{\gamma}{2} \sum_{i=1}^N \sum_{j=1}^N \Gamma_{ij} (\Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j), \quad (2.5)$$

dove  $\mathbf{\Gamma}$  è la matrice di Kirchoff, definita come:

$$\Gamma_{ij} = \begin{cases} -1 & \text{se } r_{ij} \leq R_c \text{ e } i \neq j, \\ \sum_{k \neq i} A_{ik} & \text{se } i = j, \end{cases} \quad (2.6)$$

e  $A_{ij}$  è la matrice di adiacenza della rete, che rappresenta le connessioni elastiche tra i residui basate su un cutoff di distanza  $R_c$ :

$$A_{ij} = \begin{cases} 1 & \text{se } r_{ij} \leq R_c \text{ e } i \neq j, \\ 0 & \text{altrimenti.} \end{cases} \quad (2.7)$$

In questo contesto,  $\gamma$  rappresenta una costante elastica uniforme che caratterizza la rigidità delle molle.

### Energia totale del sistema

L'Hamiltoniana finale, che combina l'energia cinetica e quella potenziale armonica, è quindi data da:

$$H = \frac{1}{2} \sum_{i=1}^N m_i \|\dot{\mathbf{r}}_i\|^2 + \frac{\gamma}{2} \sum_{i=1}^N \sum_{j=1}^N \Gamma_{ij} (\Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j). \quad (2.8)$$

Questa forma descrive completamente il sistema in termini di una rete elastica di nodi (i residui) connessi da molle (le interazioni elastiche). Le equazioni del moto derivate da questa Hamiltoniana permettono di calcolare le fluttuazioni dinamiche e le correlazioni tra i residui.

## 2.2 Fluttuazioni e correlazioni dinamiche

Le fluttuazioni medie quadratiche  $\langle (\Delta R_i)^2 \rangle$  di ciascun residuo e le correlazioni  $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle$  tra coppie di residui sono direttamente ottenibili risolvendo le equazioni di equilibrio. La matrice inversa di  $\mathbf{\Gamma}$ , detta *matrice delle correlazioni inverse*, è fondamentale per queste analisi:

$$\langle \Delta \mathbf{R} \Delta \mathbf{R}^T \rangle \propto \mathbf{\Gamma}^{-1}. \quad (2.9)$$

Le fluttuazioni medie quadratiche per un residuo  $i$  sono date da:

$$\langle (\Delta R_i)^2 \rangle = \left[ \mathbf{\Gamma}^{-1} \right]_{ii}, \quad (2.10)$$

mentre le correlazioni tra i residui  $i$  e  $j$  sono:

$$\langle \Delta R_i \cdot \Delta R_j \rangle = - \left[ \mathbf{\Gamma}^{-1} \right]_{ij}. \quad (2.11)$$

## 2.3 Conclusioni

Gli indicatori dinamici, come la correlazione, la risposta lineare e l'entropia di produzione, forniscono strumenti matematici potenti per analizzare i meccanismi allosterici delle proteine. Combinando questi approcci, è possibile ottenere una comprensione dettagliata delle reti dinamiche che governano la funzione proteica, aprendo nuove prospettive per la progettazione di farmaci e per l'ingegneria delle proteine.

## Chapter 3

# Indicatori per l'analisi delle dinamiche proteiche e allosteriche

### 3.0.1 Il concetto di indicatori dinamici

Gli indicatori dinamici rappresentano strumenti matematici e fisici fondamentali per analizzare e caratterizzare le proprietà dinamiche di sistemi complessi, come le proteine, le reti molecolari o altri sistemi biologici. Essi si basano su una descrizione approfondita delle fluttuazioni, delle interazioni e dei processi stocastici che governano il comportamento collettivo dei componenti del sistema. L'importanza degli indicatori dinamici risiede nella loro capacità di estrarre informazioni quantitative dalle dinamiche sottostanti, fornendo un quadro dettagliato delle relazioni tra struttura e funzione.

**Definizione generale e motivazioni** Gli indicatori dinamici sono definiti come quantità matematiche o fisiche che catturano aspetti specifici delle dinamiche di un sistema. A differenza delle analisi statiche, che considerano una singola configurazione della proteina o una struttura cristallografica, questi strumenti forniscono informazioni sulle fluttuazioni temporali, sui moti collettivi e sulle interazioni spaziali tra diverse parti del sistema.

In particolare, essi sono utili per:

- **Quantificare le fluttuazioni locali e globali:** Le proteine non sono strutture rigide, ma oscillano costantemente attorno alla loro configurazione nativa a causa delle fluttuazioni termiche. Gli indicatori dinamici permettono di misurare e caratterizzare l'ampiezza di queste fluttuazioni, evidenziando regioni flessibili o rigidamente vincolate.
- **Esplorare le correlazioni dinamiche:** I residui all'interno di una proteina possono muoversi in modo sincronizzato o indipendente. Gli indicatori dinamici aiutano a determinare il grado di cooperazione tra questi movimenti, identificando moti collettivi che possono essere cruciali per la funzione biologica.

- **Valutare la trasmissione di informazioni:** Le proteine spesso trasmettono segnali conformazionali da una regione all'altra, un fenomeno noto come allostericità. Gli indicatori dinamici forniscono strumenti per quantificare il trasferimento di informazioni dinamiche attraverso la struttura proteica.

**Fluttuazioni locali e globali: una panoramica** Le fluttuazioni dinamiche di una proteina possono essere analizzate a diversi livelli di dettaglio:

- **Fluttuazioni locali:** Queste descrivono i moti di piccoli gruppi di atomi o residui. Indicatori come le fluttuazioni medie quadratiche ( $\langle(\Delta R_i)^2\rangle$ ) forniscono una misura della mobilità di ciascun residuo, mettendo in evidenza le regioni flessibili come loop o terminali liberi.
- **Fluttuazioni globali:** Questi moti coinvolgono interi domini o subunità della proteina. Gli indicatori dinamici, come i modi normali a basse frequenze, permettono di identificare movimenti collettivi che supportano funzioni biologiche chiave, come l'apertura e chiusura di tasche enzimatiche.

**Correlazioni dinamiche e interazioni tra residui** Le correlazioni dinamiche rappresentano un altro aspetto critico delle dinamiche proteiche. Queste misurano il grado di sincronia nei movimenti di due residui o regioni distinte. Indicatori come la matrice delle correlazioni dinamiche  $C_{ij}(t)$  sono fondamentali per:

- Identificare reti di residui cooperativi, che spesso formano il cuore dei meccanismi allosterici.
- Distinguere tra movimenti sincroni (correlazioni positive) e movimenti opposti (anticorrelazioni negative).

Questi concetti sono particolarmente rilevanti per analizzare proteine multidominio, dove il coordinamento tra diversi segmenti è essenziale per la funzione.

**Trasferimento di informazione e comunicazione dinamica** Un ulteriore elemento chiave degli indicatori dinamici è la loro capacità di quantificare il trasferimento di informazioni tra diverse parti della proteina. Questo è cruciale per comprendere i meccanismi allosterici, in cui eventi locali, come il legame di un ligando, causano cambiamenti conformazionali che si propagano attraverso la struttura. Indicatori come la *transfer entropy* ( $TE$ ) permettono di:

- Quantificare il flusso di informazione da un residuo a un altro.
- Identificare i "nodi" critici della rete dinamica che mediano la comunicazione tra regioni distanti.

**Connessione con i processi stocastici** Il comportamento dinamico di una proteina è governato da processi stocastici, che combinano forze deterministiche, come quelle elastiche, e fluttuazioni casuali termiche. Gli indicatori dinamici sono strettamente legati a questa descrizione probabilistica, poiché derivano da equazioni del moto che incorporano:

- **Fluttuazioni termiche:** Modellate come rumore bianco gaussiano.
- **Reti elastiche:** Che descrivono le interazioni tra i residui vicini.

Questo quadro matematico consente di collegare gli indicatori dinamici a grandezze fisiche misurabili, come i fattori di Debye-Waller (*B-factors*) ottenuti dalla cristallografia a raggi X.

**Relazione con le proprietà spettrali della matrice di Kirchhoff** Un aspetto cruciale degli indicatori dinamici è la loro relazione con le proprietà spettrali della matrice di Kirchhoff, che descrive la connettività elastica del sistema. Gli autovalori  $\lambda_k$  e gli autovettori  $\mathbf{u}_k$  della matrice di Kirchhoff permettono di decomporre i movimenti dinamici in *modi normali*, ciascuno caratterizzato da una frequenza e un'ampiezza specifica. Gli indicatori dinamici possono quindi essere interpretati come combinazioni ponderate di questi modi normali, fornendo una comprensione più profonda delle dinamiche collettive.

**Applicazioni pratiche degli indicatori dinamici** Gli indicatori dinamici sono stati applicati con successo in numerosi studi per:

- Identificare regioni flessibili o rigide in enzimi, che sono spesso correlate a siti attivi o di legame.
- Analizzare i meccanismi di apertura e chiusura di canali ionici.
- Progettare proteine ingegnerizzate con proprietà dinamiche desiderate.
- Sviluppare farmaci mirati contro regioni funzionali identificate dinamicamente.

**Limiti e sfide** Nonostante il loro successo, gli indicatori dinamici presentano alcune limitazioni:

- La loro interpretazione dipende fortemente dal modello utilizzato per descrivere la dinamica proteica (es. rete elastica vs dinamica molecolare).
- Alcuni indicatori richiedono una grande quantità di dati o calcoli intensivi, specialmente in sistemi complessi o di grandi dimensioni.
- La scelta dei parametri (es. cutoff di distanza, costanti elastiche) può influenzare significativamente i risultati.

**Conclusioni** Gli indicatori dinamici offrono un potente set di strumenti per analizzare le dinamiche proteiche, fornendo informazioni critiche sulle interazioni e sui meccanismi funzionali. La loro combinazione con modelli computazionali avanzati e dati sperimentali rappresenta una direzione promettente per future ricerche in biologia strutturale e biofisica.

### 3.0.2 Tipologie di indicatori dinamici

Nel contesto delle dinamiche proteiche, gli indicatori dinamici rappresentano strumenti essenziali per quantificare e comprendere i moti collettivi e locali di una proteina, analizzando non solo le interazioni statiche ma anche quelle temporali e causali. Questi indicatori consentono di esplorare le relazioni tra struttura, dinamica e funzione, fornendo informazioni fondamentali su come le proteine rispondono a stimoli interni o esterni e su come comunicano all'interno della loro rete molecolare.

**Il ruolo degli indicatori dinamici nelle proteine** Gli indicatori dinamici sono progettati per catturare diversi aspetti delle fluttuazioni e delle interazioni proteiche:

- **Aspetto locale:** caratterizzano il comportamento di singoli residui o piccoli gruppi di residui.
- **Aspetto collettivo:** descrivono i moti che coinvolgono regioni più ampie della proteina o addirittura l'intera struttura.
- **Causalità:** determinano in che misura il movimento di una parte della proteina influenzi direttamente un'altra.
- **Allostericità:** identificano come le modifiche locali (ad esempio, il legame con un ligando) si traducano in cambiamenti funzionali globali.

La classificazione degli indicatori dinamici è utile per separare le informazioni relative ai diversi livelli di organizzazione della dinamica proteica. Nel contesto delle reti elastiche e delle descrizioni stocastiche, possiamo identificare tre categorie principali di indicatori dinamici: **correlazioni dinamiche**, **risposta lineare** e **trasferimento di informazione**. Ognuno di essi ha un ruolo specifico nella comprensione della dinamica proteica.

#### 1. Correlazioni dinamiche

Le correlazioni dinamiche rappresentano la sincronia nei movimenti di coppie di residui e sono descritte matematicamente dalla matrice delle correlazioni temporali  $C_{ij}(t)$ :

$$C_{ij}(t) = \langle x_i(0)x_j(t) \rangle, \quad (3.1)$$

dove  $x_i(t)$  e  $x_j(t)$  rappresentano i movimenti dei residui  $i$  e  $j$  nel tempo, e  $\langle \cdot \rangle$  denota la media sull'ensemble. Le correlazioni dinamiche hanno diverse interpretazioni:

- **Correlazioni positive:** indicano che i residui si muovono in modo cooperativo nella stessa direzione.
- **Correlazioni negative:** rappresentano movimenti opposti, spesso associati a cambiamenti conformazionali significativi.
- **Correlazioni nulle:** suggeriscono che i movimenti di due residui sono indipendenti.

Le correlazioni dinamiche sono particolarmente utili per identificare *reti di residui cooperativi*, che sono spesso coinvolte in meccanismi funzionali chiave come l'allostericità. Ad esempio, le correlazioni elevate tra un sito attivo e un sito distante possono indicare la presenza di un canale di comunicazione dinamico che trasmette segnali conformazionali attraverso la proteina.

**Esempi applicativi delle correlazioni dinamiche** Le correlazioni dinamiche sono state utilizzate per analizzare enzimi come l'HIV-1 proteasi, dove le correlazioni tra i flap e il core enzimatico hanno rivelato meccanismi di apertura e chiusura essenziali per la catalisi. Analogamente, nei canali ionici, le correlazioni dinamiche hanno permesso di identificare residui critici per la regolazione del flusso ionico.

## 2. Risposta lineare

La risposta lineare misura come un residuo risponde a perturbazioni esterne applicate ad altri residui. Formalmente, la risposta dinamica è definita come:

$$R_{ij}(t) = \frac{\delta \langle x_i(t) \rangle}{\delta F_j(0)}, \quad (3.2)$$

dove  $F_j(0)$  rappresenta una forza esterna applicata al residuo  $j$  al tempo iniziale.

La risposta lineare è un indicatore di *sensibilità dinamica*: descrive la misura in cui un residuo è influenzato dai cambiamenti in altre parti della proteina. In particolare, permette di identificare residui che agiscono come nodi critici nella rete di comunicazione interna della proteina.

**Interpretazione fisica della risposta lineare** La risposta  $R_{ij}(t)$  è strettamente legata alla matrice di Kirchhoff  $\mathbf{K}$  e ai suoi autovalori/autovettori. I residui che mostrano una risposta elevata a perturbazioni esterne sono spesso coinvolti in meccanismi di trasmissione del segnale. Inoltre:

- Una risposta alta suggerisce una connessione dinamica forte tra due residui.
- Una risposta bassa implica una connessione debole o un'assenza di comunicazione diretta.

**Utilizzo della risposta lineare nell'allostericità** Nei sistemi allosterici, la risposta lineare può essere utilizzata per prevedere quali residui trasmettono segnali conformazionali dal sito di legame all'intera proteina. Ad esempio, in una proteina come l'emoglobina, la risposta lineare permette di mappare il percorso dinamico attraverso cui il legame dell'ossigeno in un sito attivo induce cambiamenti strutturali cooperativi.

## 3. Trasferimento di informazione

Il trasferimento di informazione è un indicatore fondamentale per quantificare la causalità nelle dinamiche proteiche. Formalmente, si basa sulla misura di *transfer entropy* (TE), definita come:

$$TE_{j \rightarrow i}(t) = \langle \log ($$



$\overline{)P[x_i(t)|x_i(0), x_j(0)]P[x_i(t)|x_i(0)], (3.3)}$  dove  $P[x_i(t)|x_i(0), x_j(0)]$  è la probabilità condizionata del movimento del residuo  $i$  dato lo stato iniziale di  $i$  e  $j$ , e  $P[x_i(t)|x_i(0)]$  esclude l'influenza di  $j$ .

**Causalità e comunicazione dinamica** A differenza delle correlazioni, che descrivono solo sincronia, il trasferimento di informazione identifica relazioni causali tra residui. Un valore elevato di  $TE_{j \rightarrow i}(t)$  implica che i movimenti del residuo  $j$  influenzano direttamente i movimenti del residuo  $i$ . Questo è cruciale per comprendere come le proteine trasmettono segnali conformazionali attraverso reti di interazioni dinamiche.

**Importanza nell'allostericità** Nel contesto dell'allostericità, il trasferimento di informazione permette di tracciare i percorsi dinamici che collegano un evento locale (come il legame di un ligando) a cambiamenti globali nella struttura proteica. Ad esempio:

- Nelle chinasi, il trasferimento di informazione può rivelare come il legame di ATP regoli l'attività catalitica attraverso moti cooperativi.
- Nei recettori accoppiati a proteine G (GPCR), l'analisi del trasferimento di informazione evidenzia i residui chiave coinvolti nella trasduzione del segnale attraverso la membrana.

**Calcolo e applicazione pratica** Il calcolo del trasferimento di informazione si basa sull'analisi di traiettorie stocastiche e richiede una buona conoscenza delle proprietà dinamiche del sistema. È stato applicato con successo per identificare regioni allosteriche in numerosi sistemi proteici, aiutando a progettare farmaci che mirano a residui specifici per modulare la funzione proteica.

**Conclusioni** Le tre tipologie principali di indicatori dinamici – correlazioni dinamiche, risposta lineare e trasferimento di informazione – forniscono una gamma completa di strumenti per analizzare le dinamiche proteiche. Ciascuno di essi contribuisce a un aspetto specifico della comprensione delle relazioni dinamiche e causali nelle proteine, con applicazioni dirette nella biologia strutturale, nell'ingegneria proteica e nel design di farmaci. La loro combinazione offre una visione dettagliata e multilivello delle interazioni che governano i processi biologici.

[a4paper,12pt]article [utf8]inputenc amsmath amssymb graphicx geometry hyperref a4paper, margin=1in



SAPIENZA  
UNIVERSITÀ DI ROMA

# Causalità nei Sistemi Biologici: Analisi di Correlazione, Risposta ed Entropia nella Proteina Ubiquitina

Facoltà di Scienze Matematiche, Fisiche e Naturali  
Corso di Laurea Magistrale in Fisica

ID number

Advisor

Co-Advisor

Academic Year

---

**Causalità nei Sistemi Biologici: Analisi di Correlazione, Risposta ed Entropia  
nella Proteina Ubiquitina**

Sapienza University of Rome

© . All rights reserved

This thesis has been typeset by  $\text{\LaTeX}$  and the Sapthesis class.

Author's email:

# Contents

## 3.1 Introduzione

### 3.1.1 Causalità nei sistemi biologici

La causalità è un concetto fondamentale per comprendere i meccanismi che regolano i sistemi biologici. Contrariamente alla correlazione, che descrive un'associazione statistica tra due variabili, la causalità implica un legame diretto di causa-effetto. In una proteina, questo significa che un cambiamento in una parte della struttura può direttamente influenzare un'altra regione.

La comprensione della causalità è essenziale per:

- Identificare le relazioni funzionali tra i residui.
- Determinare i percorsi allosterici.
- Progettare interventi mirati, ad esempio nella progettazione di farmaci.

### 3.1.2 Differenza tra correlazione e causalità

Mentre la correlazione indica semplicemente che due variabili si muovono insieme, la causalità implica che una variabile è responsabile del cambiamento nell'altra. Questo è particolarmente importante nei sistemi biologici, dove interazioni complesse possono generare correlazioni spurie.

## 3.2 Indicatori per l'analisi della causalità

### 3.2.1 Correlazione

La correlazione è definita come la misura dell'associazione lineare tra due variabili. Per una proteina, è calcolata utilizzando la matrice di covarianza:

$$C_{ij} = \langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j \rangle$$

dove  $\Delta \mathbf{r}_i$  rappresenta lo spostamento del residuo  $i$  dalla sua posizione di equilibrio.

### 3.2.2 Funzione di risposta

La funzione di risposta misura la reazione di una variabile a una perturbazione:

$$R_{ij}(t) = \frac{\delta \langle x_i(t + \tau) \rangle}{\delta x_j(\tau)}$$

Questo è particolarmente utile per determinare relazioni causali dirette tra residui.

### 3.2.3 Trasferimento di entropia

Il trasferimento di entropia è un indicatore asimmetrico che misura la causalità:

$$TE_{j \rightarrow i}(t) = H[x_i(t)|x_i(0)] - H[x_i(t)|x_i(0), x_j(0)]$$

dove  $H$  è l'entropia di Shannon. Questa quantità permette di distinguere i "donatori" e i "ricevitori" di informazione.

### 3.3 Processi stocastici nella dinamica proteica

#### 3.3.1 Modello di rete gaussiana (GNM)

Il modello di rete gaussiana rappresenta la dinamica delle proteine come un sistema armonico di nodi e connessioni. L'equazione del moto per il sistema è:

$$\gamma \dot{x}_i = -g \sum_j K_{ij} x_j + \sqrt{2\gamma k_B T} \xi_i(t)$$

dove:

- $\gamma$  è il coefficiente di attrito.
- $g$  è il parametro di rigidità.
- $\xi_i(t)$  è un processo stocastico gaussiano con media zero.

#### Fluttuazione-dissipazione

Utilizzando il teorema di fluttuazione-dissipazione, possiamo scrivere la funzione di risposta in termini di correlazioni:

$$R_{ij}(t) = C_{ij}(t) C_{ij}^{-1}(0)$$

#### 3.3.2 Processo di Ornstein-Uhlenbeck

Il modello GNM è un caso particolare del processo di Ornstein-Uhlenbeck, descritto dall'equazione:

$$dx_t = -\theta x_t dt + \sigma dW_t$$

dove:

- $\theta$  rappresenta la velocità di ritorno verso l'equilibrio.
- $\sigma$  controlla l'ampiezza delle fluttuazioni.
- $W_t$  è un moto browniano standard.

#### Soluzione analitica

La soluzione del processo di Ornstein-Uhlenbeck è:

$$x_t = x_0 e^{-\theta t} + \int_0^t e^{-\theta(t-s)} \sigma dW_s$$

### 3.4 Dimostrazioni matematiche

#### 3.4.1 Derivazione della funzione di risposta

Partendo dall'equazione del moto:

$$\gamma \dot{x}_i = -g \sum_j K_{ij} x_j$$

e applicando la trasformata di Laplace, otteniamo:

$$R_{ij}(s) = \frac{\delta_{ij}}{\gamma s + gK_{ij}}$$

Invertendo la trasformata, si ottiene la funzione di risposta nel dominio del tempo.

### 3.4.2 Trasferimento di entropia

Considerando un processo gaussiano, il trasferimento di entropia può essere espresso come:

$$TE_{j \rightarrow i} = \frac{1}{2} \ln \left( \frac{\det C_{ii}}{\det(C_{ii} - C_{ij}C_{jj}^{-1}C_{ji})} \right)$$

## 3.5 Risultati e applicazioni

### 3.5.1 Caso dell'ubiquitina

L'analisi delle dinamiche dell'ubiquitina mostra che:

- Le correlazioni identificano regioni con moti coordinati.
- La funzione di risposta rivela relazioni causali dirette.
- Il trasferimento di entropia evidenzia percorsi di comunicazione asimmetrici.

## 3.6 Conclusioni

L'integrazione di più metodologie consente di comprendere meglio la causalità nei sistemi proteici. La combinazione di correlazione, funzione di risposta e trasferimento di entropia rappresenta l'approccio più robusto.

## .1 Appendice: Derivazioni e dettagli tecnici

### .1.1 Teorema di fluttuazione-dissipazione

Dimostrazione dettagliata del legame tra correlazione e risposta.

## Riferimenti