



UNIVERSITÀ DI PISA

Corso di Laurea Triennale in Informatica (L-31)

TESI DI LAUREA

**Protein Folding: dai metodi classici per la
predizione della struttura di proteine alla
rivoluzione di AlphaFold**

Relatore

Prof. Paolo Milazzo

Correlatore

Prof. Mario Pirchio

Candidato

Ludovico Venturi

ANNO ACCADEMICO 2020/2021

Indice

1	Protein Folding	2
1.1	Postulato di Anfinsen	3
1.1.1	Esperimento di Anfinsen	4
1.1.2	Denaturazione	4
1.2	Struttura delle proteine	6
1.2.1	Interazioni molecolari	6
1.2.2	Geometria dei legami?	11
1.3	Ripiegamento assistito	11
1.3.1	Misfolding e malattie	13
1.3.2	Controllo qualità e proteasomi	14
1.3.3	Unfolded protein response	16
1.4	Eccezioni al postulato di Anfinsen	16
1.4.1	Intrinsically disordered proteins	16
1.4.2	Fold switching proteins	16
1.4.3	box: Filosofia della scienza	17
1.5	Il problema del Protein Folding	17
1.5.1	Limiti al ripiegamento	17
1.5.2	Energetica del ripiegamento	17
	Bibliografia	19

Capitolo 1

Protein Folding

«la forma è l'immagine plastica della funzione»¹

La correlazione tra forma e funzione si rivela fondamentale nel caso delle proteine. Un canale ionico neuronale permette il passaggio di ioni grazie alla sua forma a canale; una ferritina cattura e immagazzina gli ioni ferro grazie alla sua forma a sfera cava.

Il ripiegamento delle proteine (*protein folding*) è il processo di ripiegamento molecolare attraverso il quale a partire dalla sequenza lineare amminoacidica le proteine ottengono la loro struttura tridimensionale, chiamata forma *nativa*, che permette loro di svolgere la relativa funzione biologica.

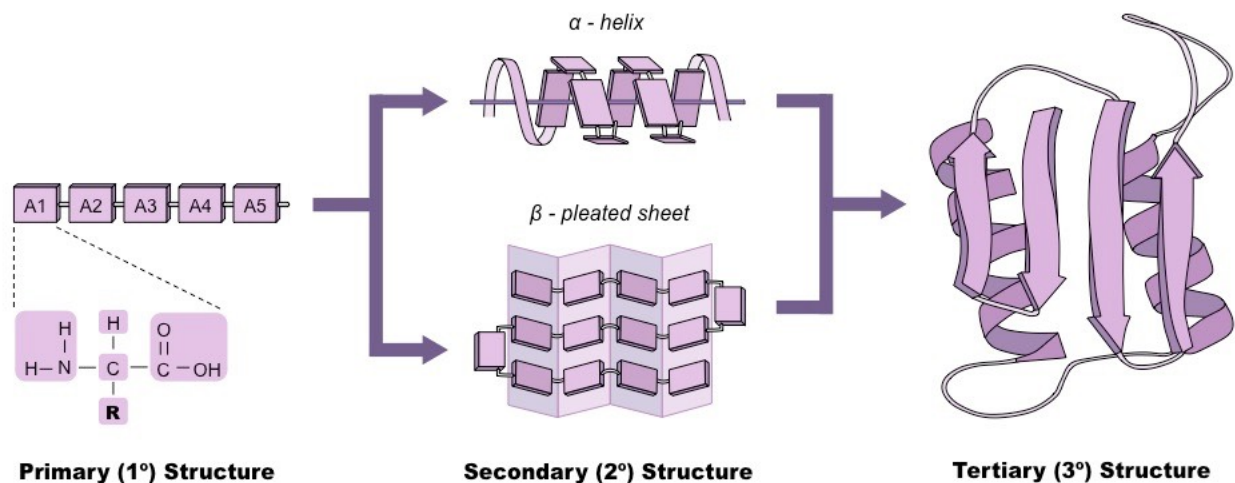


Figura 1.1: Protein folding: dagli amminoacidi alla struttura tridimensionale. Fonte: [2]

¹A. Ruffini, *Fisiogenia, la biodinamica dello sviluppo ed i fondamentali problemi morfologici dell'embriologia generale*, 1925

Il ripiegamento nella forma tridimensionale avviene spontaneamente sia durante la sintesi proteica nei ribosomi sia al termine di questa. Una specifica proteina si ripiegherà nello stesso modo e avrà la stessa struttura finale².

La prima teoria del ripiegamento proteico è stata proposta negli anni venti del 20° secolo da Hsien Wu^[3], in relazione al processo di denaturazione (vedi sezione 1.1.2). È però Anfinsen, premio Nobel per la chimica, negli anni '60 a compiere un fondamentale passo nella comprensione del processo del ripiegamento proteico^[4].

1.1 Postulato di Anfinsen

Il postulato di Anfinsen (conosciuto anche come *dogma* o *ipotesi termodinamica* di Anfinsen) afferma che la struttura nativa delle proteine (almeno quelle globulari) è determinata solamente dalla sequenza di amminoacidi di cui sono costituite. In altri termini: la struttura nativa, in ambiente fisiologico standard, corrisponde a quella struttura unica, stabile e cineticamente accessibile avente *minima energia libera*.

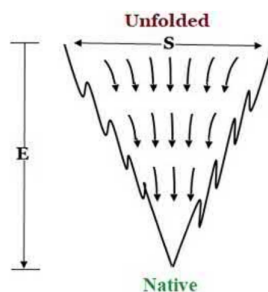


Figura 1.2: Un "panorama" idealizzato dell'energia libera a forma di imbuto. E =energia, S =entropia. Fonte: [5]

Vi sono quindi 3 condizioni:

1. *unicità*, la sequenza non deve possedere altre configurazioni dotate di energia libera comparabile
2. *stabilità*, piccoli cambiamenti nell'ambiente circostante non possono produrre cambiamenti nella configurazione a energia minima. Ciò può essere descritto come una superficie parabolica di energia libera con lo stato nativo corrispondente al punto di minimo (visivamente simile ad un imbuto, vedi fig.1.2); la superficie di energia libera nelle vicinanze dello stato nativo deve essere abbastanza ripida ed elevata

²ciò non è vero nel 100% dei casi, alcune proteine possono avere più di una conformazione stabile per adempiere funzioni diverse (vedi la sezione 1.4.2) e alcune proteine possono andare incontro a misfolding (vedi la sezione 1.3)

3. *accessibilità cinetica*, il percorso nella superficie di energia libera dallo stato *unfolded* a *folded* deve essere ragionevolmente piano

1.1.1 Esperimento di Anfinsen

L'esperimento, compiuto nel 1957^[6], consisteva nella denaturazione e rinaturazione della ribonucleasi A, dimostrando che il secondo processo era possibile senza agenti ausiliari. L'enzima in questione è formato da 124 amminoacidi, tra cui 8 cisteine che formano 4 ponti disolfuro ($-CH_2 - S-S - CH_2-$). È stato usato un agente riducente per scindere questi ponti e l'urea per denaturare la proteina: questa non mostrava più alcuna attività enzimatica. A questo punto se l'urea era rimossa prima, seguita dall'aggiunta di un agente ossidante per consentire ai ponti disolfuro di riformarsi, la ribonucleasi A riacquistava spontaneamente la sua struttura terziaria e il prodotto ottenuto risultava praticamente indistinguibile dalla proteina nativa di partenza, riottenendo piena attività biologica.

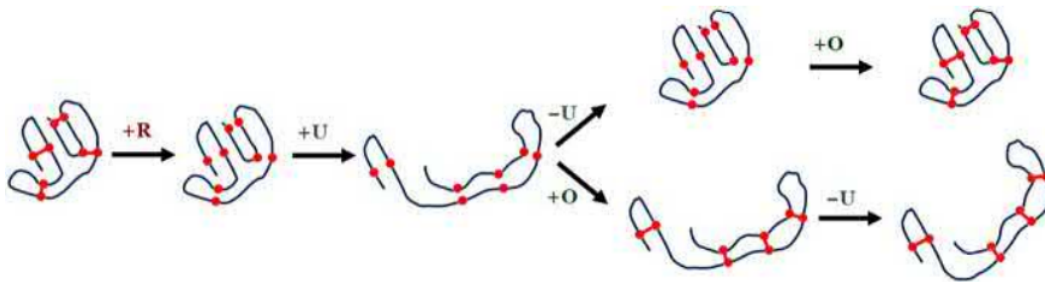


Figura 1.3: Rappresentazione schematica dell'esperimento di Anfinsen. R=reducing agent, U=Urea, O=oxidizing agent, punti rossi=cisteina, linee rosse=ponti disolfuro. Fonte: [5]

I ponti disolfuro si riformano nella stessa posizione della proteina nativa nonostante ci siano 105 modi possibili per ricombinarli. Se invece veniva prima aggiunto l'agente ossidante e poi tolta l'urea il prodotto ottenuto era un miscuglio di molte delle possibili 105 configurazioni, raggiungendo solamente l'1% dell'attività enzimatica.

1.1.2 Denaturazione

La denaturazione delle proteine è il fenomeno relativo all'alterazione della struttura nativa dovuto a variazioni di temperatura, pH o contatto con determinate sostanze chimiche. La denaturazione è un processo che porta alla perdita di ordine e quindi ad un aumento di entropia. La struttura primaria rimane invariata, data la stabilità dei legami peptidici. A causa della denaturazione le proteine perdono la loro funzione biologica e possono esporre e rendere reattivi alcuni gruppi funzionali che possono causare l'aggregazione di più proteine.

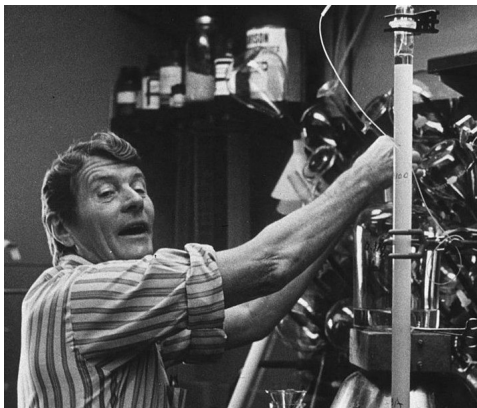


Figura 1.4: C.B. Anfinsen nel suo laboratorio.
Fonte: [7]

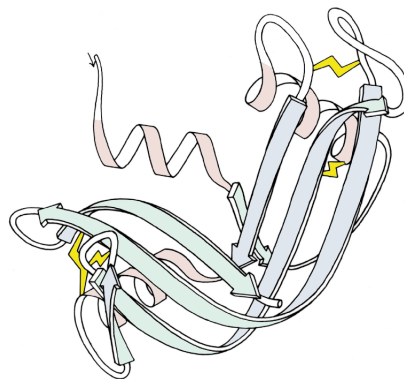


Figura 1.5: Ribonucleasi A, rappresentazione a nastro. In giallo i ponti disolfuro, rosa le α -eliche, verde e azzurro i β -foglietti. Fonte [8]

Può avvenire che una volta rimosso l'agente denaturante la proteina ritorni allo stato di partenza (*rinaturazione*) ma spesso il processo è irreversibile.

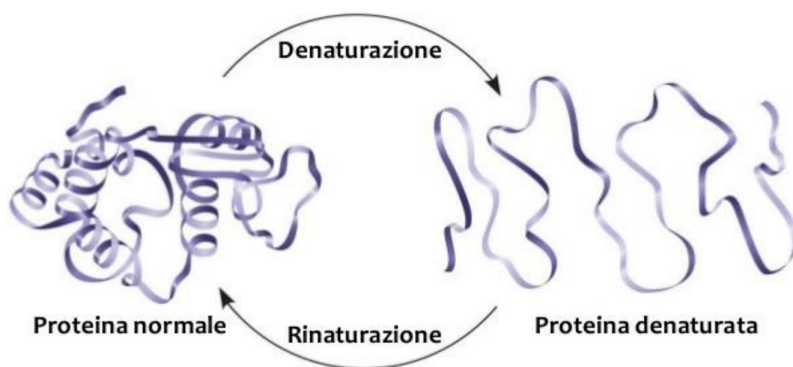


Figura 1.6: Denaturazione e rinaturazione. Fonte: [9]

La proprietà di certe sostanze chimiche (es. urea) di denaturare una molecola proteica si deve alla loro capacità di legare transientemente, attraverso legami deboli, come ad esempio legami idrogeno, i residui amminoacidici costituenti la proteina. Questi legami vengono termodinamicamente preferiti a quelli intramolecolari o intermolecolari con l'acqua. Ciò comporta l'impossibilità per la proteina di mantenere la propria struttura tridimensionale e quindi questa si denatura.

Applicazioni nella vita quotidiana di questo fenomeno sono la cottura dei cibi (basti pensare all'albumina nell'uovo) e la permanente ai capelli (denaturazione dell' α -cheratina, rompendo e riformando ponti disolfuro).

1.2 Struttura delle proteine

Da un punto di vista chimico le proteine sono di gran lunga, tra quelle conosciute, le molecole strutturalmente più complesse e sofisticate funzionalmente. È possibile studiare la loro struttura individuando successivi livelli di organizzazione:

- *struttura primaria*: la sequenza ordinata degli amminoacidi
- *struttura secondaria*: regioni ripetitive locali stabilizzate da legami idrogeno tra atomi della backbone (α -eliche e β -foglietti)
- *struttura supersecondaria*: combinazione di strutture secondarie e connessioni (motivi, domini, loop, giri ...)
- *struttura terziaria*: forma tridimensionale di una singola catena polipeptidica, risultante dalle interazioni dei residui
- *struttura quaternaria*: forma finale di proteine "assemblate" da 2 o più catene polipeptidiche già ripiegate

Prima di passare ad analizzare ogni livello della struttura delle proteine è utile un veloce sguardo alle interazioni molecolari.

1.2.1 Interazioni molecolari

- *legame covalente*: prevede la compartecipazione di 2 elettroni di valenza fra più atomi ed è il tipo di legame più forte (kcal/mol). Due o più atomi tenuti insieme da legami covalenti formano una molecola
 - *elettronegatività*:
- *legami non covalenti*
 - *legame ionico*: l'atomo più elettronegativo strappa completamente un elettrone al suo compagno, si formano due ioni (uno positivo, *catione* e uno negativo, *anione*)
 - *legame idrogeno*: un atomo di idrogeno si trova fra due atomi elettronegativi vicini; ad es. nell'acqua gli atomi di idrogeno (parzialmente positivi) si trovano fra due atomi di ossigeno (parzialmente negativi). L'idrogeno, legato a uno dei due atomi di ossigeno, permette all'altro di avvicinarsi e di stabilizzarsi
 - *interazioni di van der Waals*: nelle molecole apolari gli elettroni si possono accumulare in modo asimmetrico, formando regioni momentaneamente polari che permettono così una temporanea stabilizzazione fra molecole a breve distanza
- *effetto idrofobico*:
-

- *ponte disolfuro*, lo zolfo di una cisteina si lega allo zolfo della seconda cisteina.
- una forza dominante o tante piccole forze?
- Cambio di paradigma da metà anni '80

Struttura primaria

La struttura primaria delle proteine è la sequenza ordinata degli amminoacidi che ne determina la conformazione nativa. La posizione nella sequenza di specifici amminoacidi è un fattore fondamentale per la determinazione di quali porzioni della proteina andranno a legarsi formando globalmente la struttura finale. La nota importante, basata sul dogma di Anfinsen, è che la sequenza amminoacidica di ogni proteina contiene l'informazione che specifica sia la struttura nativa che la via per raggiungere quello stato. Questo comunque non vuol dire che strutture simili si ripieghino in modo simile.

Struttura secondaria

La struttura secondaria riguarda le regioni ripetitive locali stabilizzate da legami idrogeno tra atomi della backbone: (α -eliche e β -foglietti).

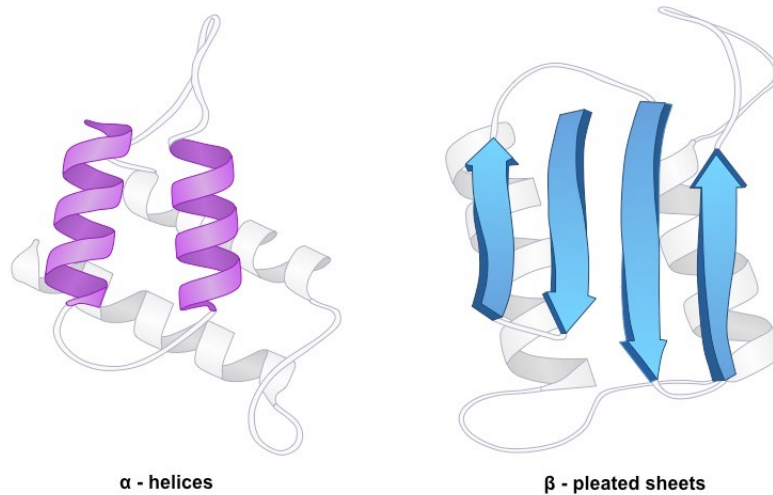


Figura 1.7: Struttura secondaria delle proteine, α -eliche e β -foglietti. Fonte: [2]

Questo livello di organizzazione è una conseguenza dei legami a idrogeno tra gli amminoacidi appartenenti a una stessa catena, o tra gli amminoacidi di catene diverse. All'interno della backbone del polipeptide gli atomi di ossigeno hanno una parziale carica negativa e gli atomi di idrogeno attaccati all'azoto hanno una parziale carica positiva perciò possono formarsi legami idrogeno fra questi atomi. Individualmente sarebbero deboli

legami ma poiché sono ripetuti molte volte su di una regione relativamente lunga di una catena polipeptidica possono fare da supporto per una particolare conformazione.

Nella struttura ad α -elica gli amminoacidi sono avvolti in una spirale tenuta insieme da legami idrogeno ogni 4 amminoacidi. Tra l'atomo di idrogeno legato all'azoto di ogni legame peptidico e l'ossigeno del gruppo carbossilico del legame peptidico sovrastante (che si trova appunto a distanza di quattro amminoacidi lungo la catena) si instaura un legame a idrogeno. Tuttavia se gli amminoacidi che si succedono lungo un tratto di catena proteica hanno gruppi R voluminosi, come avviene nella prolina, o gruppi R dotati della stessa carica elettrica, come avviene negli amminoacidi lisina e arginina, l' α -elica non può formarsi, a causa delle forze di repulsione che si generano tra i residui. Alcune proteine fibrose, come l' α -cheratina, la proteina strutturale dei capelli, la lana e le unghie hanno formazioni di α -eliche sulla maggior parte della loro lunghezza.

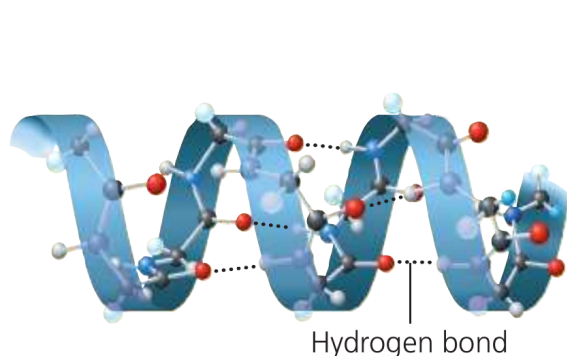


Figura 1.8: Regione di α -elica. Fonte: [9]

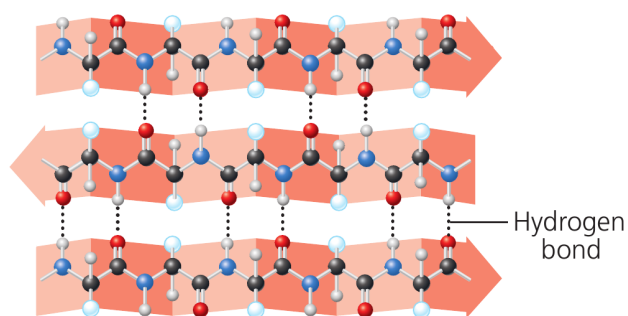


Figura 1.9: Una regione di β -foglietto composto da β -filamenti adiacenti, spesso mostrati come una freccia pieghettata o piatta puntata in direzione C-terminus. Fonte [9]

Altre proteine fibrose sono invece dominate dai β -foglietti, come le proteine della seta (β -cheratina) e la tela prodotta dai ragni. In queste conformazioni due o più segmenti della catena polipeptidica giacenti lato su lato (chiamati β -filamenti) sono connessi da tre o più legami idrogeno. Si definisce β -filamento una sequenza peptidica di amminoacidi (tipicamente 5-10) che si dispone linearmente ed è in grado di formare legami idrogeno. Ciascuna delle catene è totalmente estesa e presenta una conformazione a zig-zag, dovuta alla geometria dei legami attorno a ciascun atomo di carbonio e di azoto nella catena. In questo caso, i legami a idrogeno si formano tra gli amminoacidi di due catene adiacenti. I gruppi amminici di uno scheletro peptidico formano legame con quelli carbossilici del filamento opposto. In ogni singolo filamento i residui si dispongono perpendicolarmente al piano del foglietto, puntando alternativamente verso l'alto e verso il basso.

Nella vita quotidiana, se tiriamo per i due estremi una fibra di lana questa si allunga: si stanno rompendo i legami idrogeno e le eliche si allontanano sempre di più, ma lasciando la presa i legami idrogeno si riformano e le eliche ricompaiono nella struttura. Se invece

tiriamo la seta si può osservare che non è elastica: i foglietti di cui è composta la sua struttura non sono smantellabili senza rompere anche i legami covalenti della backbone.

Struttura supersecondaria: Motivi, Domini, Loop, Giri ...

La struttura supersecondaria è riferita alle combinazioni spaziali di strutture secondarie in conformazioni più complesse e alle connessioni che li uniscono.

I *motivi* (motifs) e *domini* (domains) sono regioni tridimensionali della catena polipeptidica formate da differenti strutture secondarie adibite a svolgere una determinata funzione per la proteina di cui fanno parte. Tuttavia sono differenti in quanto i motivi non mantengono la loro forma se separati dalla proteina laddove i domini la mantengono. Questo perché i motivi e il resto della proteina sono più vicini e si vengono così a formare legami idrogeno che permettono ai motivi di mantenere la struttura. I domini sono sì legati alla backbone della proteina ma non abbastanza vicini alla restante parte della formazione proteica da stabilire legami, pertanto se vengono separati non perdono la loro struttura e possono mantenere la loro funzione.

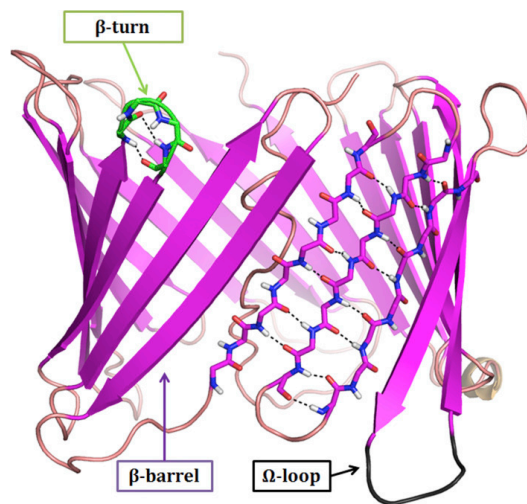


Figura 1.10: Giri e Loop. Fonte: [10]

I looppppppdasdas Protein loops are patternless regions which connect two regular secondary structures. They are generally located on the protein's surface in solvent exposed areas and often play important roles, such as interacting with other biological objects. Despite the lack of patterns, loops are not completely random structures

Nelle strutture secondarie e terziarie si trovano spesso bruschi cambiamenti di direzione nella struttura: i *giri* (turns). Queste nette svolte sono possibili grazie agli amminoacidi prolina e glicina. Il gruppo R della prolina si ripiega verso il gruppo amminico, distorcendo la catena naturalmente. Si forma però uno stretto spazio a causa del giro: l'amminoacido

con gruppo R meno voluminoso è ovviamente la glicina ed è per questo che si trovano insieme nei giri.

Struttura terziaria

La struttura terziaria è la struttura tridimensionale globale risultante dalle interazioni tra i residui successivamente alle conformazioni locali della struttura secondaria ed è quindi la descrizione del risultato del processo di ripiegamento proteico. Un tipo di interazione importante è quella idrofobica che induce i residui non polari (e quindi idrofobici) a raggrupparsi al centro della catena polipeptidica, formando un nucleo idrofobico. La forma della proteina può venire rinforzata dai ponti disolfuro, legami covalenti fra due cisteine.

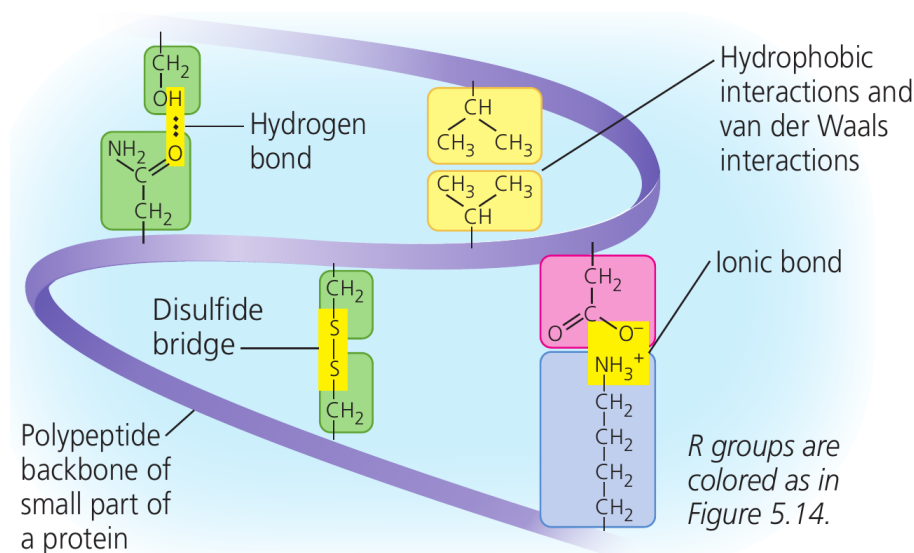


Figura 1.11: I diversi tipi di interazioni che possono contribuire alla struttura terziaria di una proteina.
Fonte: [9]

Struttura quaternaria

La struttura quaternaria è la forma finale di proteine "assemblate" da 2 o più catene polipeptidiche già ripiegate. Il collagene ne è un esempio poiché è formata da 3 polipeptidi quasi interamente a spirale che si attorcigliano l'uno sull'altro formando un'elica tripla ancora più larga, dando alle lunghe fibre una grande forza. Un altro esempio è l'emoglobina, proteina globulare formata da 4 subunità polipeptidiche. Le strutture terziarie delle subunità non vengono alterate.

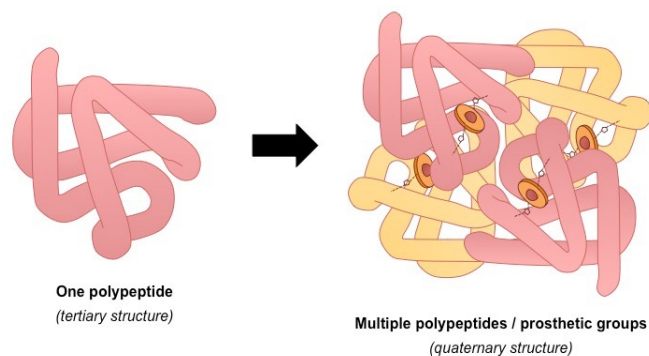


Figura 1.12: Rappresentazione di una struttura quaternaria composta da più polipeptidi e alcuni gruppi prostetica. Fonte [2]

1.2.2 Geometria dei legami?

- Domini, Residui, Motivi, Giri, Loops, Turns

1.3 Ripiegamento assistito

All'interno delle cellule le proteine più piccole si ripiegano indipendentemente, mentre proteine più grandi sono assistite principalmente da complessi chiamati *chaperoni molecolari*. È importante notare che l'assistenza è cinetica in natura: non aggiunge nuove informazioni necessarie alla proteina per ripiegarsi, pertanto il dogma di Anfinsen non viene contraddetto. Ciò che fanno questi complessi è creare un ambiente nel quale le proteine possano ripiegarsi senza "distrazioni" dovute a interazioni con altre entità (ad esempio evitando l'aggregazione con altre proteine) e senza rimanere bloccate in conformazioni intermedie durante il loro percorso di ripiegamento. In poche parole sono misure di protezione della cellula.

Più in dettaglio i chaperoni molecolari svolgono le seguenti funzioni:

1. assistono il corretto ripiegamento delle catene polipeptidiche (lunghe) appena sintetizzate
2. dirigono l'assemblaggio di complessi multienzimatici
3. donano una "seconda chance" a proteine danneggiate favorendone la rinaturazione
4. partecipano nella parziale denaturazione durante il trasporto di proteine attraverso membrane di mitocondri o cloroplasti

Tutti i compartimenti cellulari delle cellule eucariotiche (nucleo, citosol, reticolo endoplasmatico, mitocondri e cloroplasti) hanno il proprio set di chaperoni che assicura un

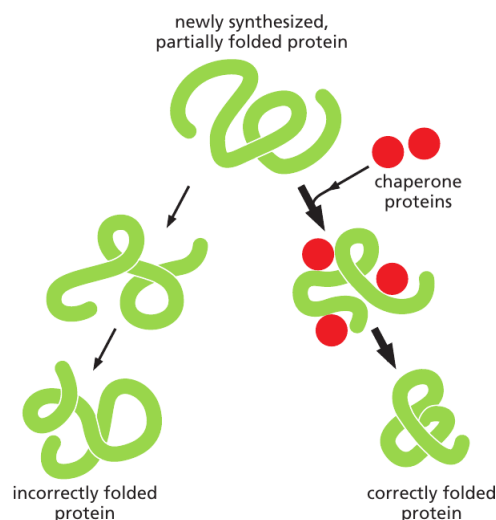


Figura 1.13: Schema della funzione dei chaperoni molecolari. Fonte: [11]

corretto ripiegamento delle proteine. I chaperoni molecolari comprendono diverse famiglie di proteine altamente conservate, tra cui le Hsp (Heat shock protein), proteine espresse in grande quantità sotto condizioni di alto stress, per contrastarne l'effetto denaturante. Queste ultime sono state classificate in base al loro peso molecolare, ad es. Hsp60 dove "60" indica 60kDa. Le Hsp60 vengono chiamate anche *chaperonine* e sono una famiglia di chaperoni molecolari a doppio anello che agiscono da "camera di isolamento" per il ripiegamento di altre proteine^[12], famosa è la chaperonina procariotica GroEL (vedi fig.1.14), che può essere assunta come modello di riferimento delle chaperonine.

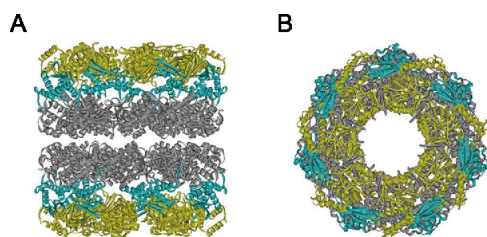


Figura 1.14: Strutture dei complessi GroEL e GroEL-GroES. (B) si può osservare la tipica forma ad anello. Fonte: [13]

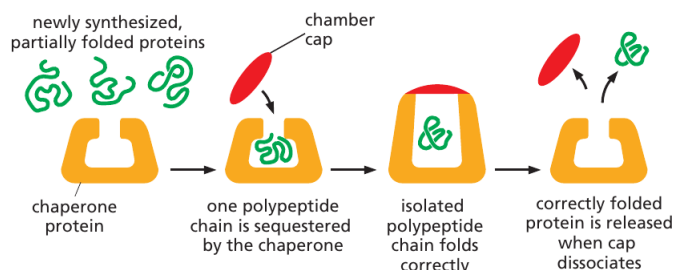


Figura 1.15: Rappresentazione schematica della funzione della camera di isolamento nelle chaperonine. Fonte [11]

Sebbene i mitocondri (e i cloroplasti) abbiano il loro genoma e creino le loro proteine, la maggior parte delle proteine che questi organelli usano sono codificate dai geni nel nucleo e importati dal citosol. Ogni proteina viene quindi parzialmente denaturata per effettuare il trasporto. I chaperoni molecolari all'interno di questi organelli aiutano a tirare le proteine attraverso le due membrane e a ripiegarle una volta all'interno^[11].

1.3.1 Misfolding e malattie

Il *misfolding* è il fenomeno dell'errato ripiegamento di una proteina, ovvero quando una proteina non può raggiungere il suo stato nativo. Ciò può accadere per mutazioni alla sua sequenza amminoacidica, anche per un solo amminoacido differente (come nel caso dell'anemia falciforme) o per fattori esterni. Le proteine mal ripiegate tipicamente contengono β -foglietti organizzati in una struttura denominata cross- β , disposizione molto stabile e insolubile, generalmente resistente alla proteolisi. Il mal ripiegamento di alcune proteine può innescare ulteriori mal ripiegamenti e la conseguente accumulazione di proteine mal ripiegate in aggregati (od oligomeri) che possono guadagnare tossicità attraverso le interazioni intermolecolari. L'incremento dei livelli di proteine aggregate può portare alla formazione di *amiloidi*, strutture fibrillari formate da deposizioni di materiale proteico insolubile. L'errato ripiegamento delle proteine è alla base quindi di molte patologie umane, definite malattie da misfolding, categorizzabili in due gruppi:

- malattia causata dalla perdita o degradazione della proteina o dall'errato trasporto intracellulare
- malattie causate dall'accumulo, intra od extra-cellulare, di proteine aggregate (ad esempio le malattie da prione)

Molti tipi di tumore diventano chemio-resistenti perché iper-esprimono alcune Hsp, come la Hsp70 e la Hsp90. Le Hsp sono presenti anche in quantità elevatissime nel cervello dei pazienti con malattia di Alzheimer e morbo di Parkinson. Tuttavia si crede che la loro aumentata espressione non sia lesiva di per sé ma rappresenti piuttosto una risposta difensiva agli elevati livelli di stress che caratterizzano queste patologie. Ci sono molti morbi associati a mutazioni nei geni codificanti i chaperoni. Alterazioni genetiche delle chaperonine possono portare a patologie umane che in genere colpiscono molti organi ed apparati contemporaneamente^[14].

I *prioni* (acronimo di "**p**roteinaceous **i**nfective **o**nly particle") sono molecole di natura proteica con la capacità di trasmettere la propria forma mal ripiegata a varianti normali della stessa proteina.³ Il ruolo ipotizzato di una proteina come agente infettivo è in contrasto con tutti gli altri agenti infettivi conosciuti, come i viroidi, virus, batteri, funghi, parassiti: tutti contengono acidi nucleici (DNA, RNA o entrambi) mentre le proteine sono composte di soli amminoacidi.

I prioni formano amiloidi che si accumulano nei tessuti e sono associati a danni di questi e alla morte cellulare. I prioni sono attualmente considerati i più probabili agenti delle

³I prioni sono stati studiati e denominati in questo modo dal premio Nobel per la medicina nel 1997 Stanley Prusiner^[15]

encefalopatie spongiformi trasmissibili (TSE) dei mammiferi. Nel *morbo della mucca pazza* (encefalopatia spongiforme bovina), malattia neurologica degenerativa e irreversibile, vi è il ruolo di un prione a causare mal ripiegamenti di alcune proteine native causando la formazione di strutture amiloidi fatali (al microscopio le dense placche fibrose appaiono come buchi, da qui il caratteristico aspetto "a spugna"). Tutte le malattie da prioni sono attualmente inguaribili e letali, con un periodo di incubazione che dura generalmente vari anni.

Gli aggregati di prioni sono stabili e questa stabilità strutturale consente loro di essere immuni alla maggior parte dei trattamenti conosciuti. L'organismo infettato non ha modo di degradarli: a differenza di virus e batteri i prioni rimangono intatti anche in presenza di trattamenti come sterilizzazione, forti dosi di radiazioni ionizzanti, uso di formaldeide, varechina, acqua bollente e a differenza delle altre proteine sono resistenti alla maggior parte delle proteasi.

La proteina di cui sono fatti i prioni, *PrP* (protease-resistant-protein, Pr per **pr**ione, e P per **pro**teina), si trova in tutto il corpo, anche negli individui sani, ed è altamente conservata nei mammiferi. Tuttavia, la *PrP* trovata nel materiale infettante ha una struttura diversa. Nell'uomo la *PrP^c* (cellulare, forma normale) è codificata da un solo gene, *PRNP*. La *PrP^{sc}* (scrapie, forma patologica) differisce dalla proteina naturale *PrP^c* per la conformazione tridimensionale: la *PrP^c* ha una struttura più aperta contenente 3 segmenti ad α -eliche e pochi β -foglietti; la *PrP^{sc}* invece ha una struttura più compatta e stabile e presenta un aumento di β -foglietti.

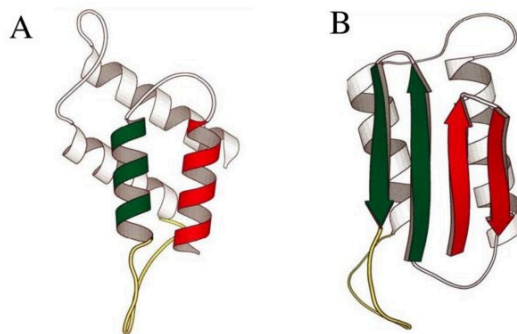


Figura 1.16: (A) Struttura della *PrP^c*. (B) Struttura della *PrP^{sc}*. Fonte: [16]

1.3.2 Controllo qualità e proteasomi

L'uscita delle proteine dal reticolo endoplasmatico (RE) è controllata per assicurare la qualità delle proteine. Sebbene alcune proteine siano appositamente create e destinate a funzionare nel RE la maggior parte delle proteine che entrano nel RE sono destinate ad altri luoghi. Queste vengono impacchettate nelle vescicole di trasporto e gemmano per

fondersi con l'apparato del Golgi. Ma l'uscita dal RE è altamente selettiva: le proteine che falliscono a ripiegarsi nella forma nativa e quelle che non si assemblano correttamente sono attivamente conservate nel RE attraverso i legami con i chaperoni molecolari che risiedono lì. Questi trattengono le proteine nel RE finché non si verifica il corretto ripiegamento o assemblaggio. Se questo non si verifica o fallisce ancora le proteine sono esportate nel citosol dove sono degradate da un *proteasoma*. Le proteine da degradare sono contraddistinte dal loro legame con l'ubiquitina⁴.

Ad esempio molecole di anticorpi sono composte da 4 catene polipeptidiche che si assemblano in completi anticorpi nel RE. Gli anticorpi parzialmente assemblati sono conservati nel RE finché tutte e 4 le catene non sono pronte. Le molecole di anticorpi che falliscono ad assemblarsi vengono degradate.

Nonostante l'indubbia utilità di questo meccanismo di controllo a volte questo può rivelarsi dannoso per l'organismo. Ad esempio la mutazione predominante che causa la *fibrosi cistica*, comune malattia genetica che comporta seri danni polmonari, produce una proteina di trasporto della membrana plasmatica leggermente mal ripiegata. Tuttavia questa potrebbe funzionare normalmente se raggiungesse la membrana plasmatica ma viene bloccata nel RE e successivamente degradata^[11] (per usare una metafora si può immaginare la situazione di un condannato alla pena di morte innocente). Le conseguenze sono terribili. La nota da sottolineare è che in questa malattia la mutazione non causa un'inattivazione di una proteina importante ma la proteina attiva è scartata dalle cellule prima che questa possa avere l'opportunità di funzionare.

I proteasomi sono complessi di *proteasi* (enzima in grado di catalizzare la rottura del legame peptidico delle proteine) che degradano le proteine mal ripiegate attraverso reazioni di *proteolisi*. Sono presenti nelle cellule di tutti gli eucarioti e procarioti. La struttura e la funzione di questi complessi è altamente conservata.

A causa del ruolo dei proteasomi nella regolazione del ciclo cellulare e dell'apoptosi⁵, sono oggi un bersaglio rilevante nelle terapie antitumorali. Farmaci inibitori nella terapia antiretrovirale interferiscono con il ciclo replicativo del virus HIV proprio bloccando l'attività dell'enzima della proteasi.

⁴Per "la scoperta della degradazione delle proteine mediata da ubiquitina" è stato assegnato il Premio Nobel per la chimica del 2004 ad Aaron Ciechanover, Avram Herskho ed Irwin Rose.

⁵Il termine *apoptosi* indica una forma di morte cellulare programmata (un'auto-distruzione). Al contrario della necrosi, che è una forma di morte cellulare risultante da un acuto stress o trauma cellulare, l'apoptosi è portata avanti in modo ordinato e regolato, richiede consumo di energia (ATP) e generalmente porta a un vantaggio durante il ciclo vitale dell'organismo (è infatti chiamata da alcuni morte altruista o morte pulita). Durante il suo sviluppo, ad esempio, l'embrione umano presenta gli abbozzi di mani e piedi "palmati": affinché le dita si differenzino, è necessario che le cellule che costituiscono le membrane interdigitali muoiano

1.3.3 Unfolded protein response

La dimensione del RE è controllato dalla "richiesta" per il ripiegamento delle proteine. Il meccanismo di controllo nel RE, eseguito dai chaperoni molecolari, può essere sopraffatto. Quando succede le proteine mal ripiegate si accumulano nel RE. Se l'accumulo è abbastanza grande, questo innesca un complesso programma chiamato *unfolded protein response* (UPR). Questo programma incita la cellula a produrre più RE, inclusi più chaperoni molecolari, e altre proteine riguardanti il controllo qualità. L'UPR permette alla cellula di regolare la dimensione del RE per gestire propriamente il volume delle proteine in entrata. In alcuni casi tuttavia anche un RE espanso non riesce a gestire la richiesta e l'UPR indirizza la cellula verso l'*apoptosi*.

Una situazione del genere può avvenire negli adulti in cui insorge il diabete. I tessuti diventano gradualmente resistenti all'effetto dell'insulina. Per compensare questa resistenza le cellule che secernono insulina nel pancreas ne producono ancora di più. Si arriva infine alla situazione in cui il loro RE arriva ad una capacità massima e viene innescato l'UPR e di conseguenza la morte cellulare. Col tempo sempre più cellule secernenti insulina sono eliminate e la richiesta per quelle sopravvissute aumenta rendendole sempre più vulnerabili a questo meccanismo, esacerbando ulteriormente la malattia^[11].

1.4 Eccezioni al postulato di Anfinsen

– IDP [soft computing articolo]— The structure of some proteins is difficult to determine for a simple reason: A growing body of biochemical research has revealed that a significant number of proteins, or regions of proteins, do not have a distinct 3-D structure until they interact with a target protein or other molecule. Their flexibility and indefinite structure are important for their function, which may require binding with different targets at different times. These proteins, which may account for 20–30 mammalian proteins, are called intrinsically disordered proteins and are the focus of current research.

1.4.1 Intrinsically disordered proteins

1.4.2 Fold switching proteins

Some proteins have multiple native structures, and change their fold based on some external factors. For example, the KaiB protein complex switches fold throughout the day, acting as a clock for cyanobacteria. It has been estimated that around 0.5–4% of PDB proteins switch folds.[7] The switching between alternative structures is driven by interactions of the protein with small ligands or other proteins, by chemical modifications (such

as phosphorylation) or by changed environmental conditions, such as temperature, pH or membrane potential. Each alternative structure may either correspond to the global minimum of free energy of the protein at the given conditions or be kinetically trapped in a higher local minimum of free energy.[8]

— porter, youtube — “I study proteins and proteins have been thought to have one structure that has one function or fold. I’m studying this group of proteins called fold switching proteins. So, they can actually change their structures and their functions in response to changes in the cell. So, you can kind of imagine fold switching proteins are like a transformer where in one case the protein is like a robot that does one thing and then in another case, in response to changes in our bodies, it becomes a car and can do something else. And an advantage to this is it can respond really quickly to changes in our bodies

1.4.3 box: Filosofia della scienza

1.5 Il problema del Protein Folding

- cos’è il problema del protein folding: non solo la struttura finale
 - 3 sottoproblemi: folding code, protein structure prediction e folding process
 - le domande di principio del protein folding
- paradosso di Levinthal

Come accennato nella sezione 1.3, la malattia di Alzheimer, la fibrosi cistica e altre malattie neurodegenerative sono associate al mal ripiegamento delle proteine. La conoscenza dei fattori di mal ripiegamento e la comprensione del processo di ripiegamento proteico potrebbero aiutare nello sviluppo di cure per queste malattie. La conoscenza della struttura delle proteine fornisce un grande vantaggio per lo sviluppo di nuovi farmaci e il design di nuove proteine.

1.5.1 Limiti al ripiegamento

- Limiti al ripiegamento: angoli di torsione e piano di Ramachandran

1.5.2 Energetica del ripiegamento

- processo spontaneo: energia di Gibbs, entalpia, entropia

— Biochemists now know the amino acid sequence for about 160 million proteins, with about 4.5–5 million added each month, and the three-dimensional shape for about 40,000. Researchers have tried to correlate the primary structure of many proteins with their three-dimensional structure to discover the rules of protein folding. Unfortunately, however, the protein-folding process is not that simple. Most proteins probably go through several intermediate structures on their way to a stable shape, and looking at the mature structure does not reveal the stages of folding required to achieve that form

– strumenti [soft computing articolo]— Even when scientists have a correctly folded protein in hand, determining its exact three-dimensional structure is not simple, for a single protein has thousands of atoms. The method most commonly used to determine the 3-D structure of a protein is X-ray crystallography

Bibliografia

Libri

- [5] S. Pal, *Fundamentals of Molecular Structural Biology*. Academic Press, 2019.
- [9] L. A. Urry, M. L. Cain, S. A. Wasserman, P. V. Minorsky, R. B. Orr e N. A. Campbell, *Campbell Biology*. Pearson, 2021.
- [11] B. Alberts, D. Bray, K. Hopkin et al., *Essential cell biology*, 5^a ed. W. W. Norton e Company, 2019.

Articoli

- [3] H. Wu e E. Yang, “Studies on denaturation of proteins. XL Effect of hydrogen ion concentration on rate of denaturation of egg albumin by urea. A theory of denaturation,” *Chin J Physiol*, vol. 5, pp. 301–344, 1931.
- [4] C. B. Anfinsen, “The formation and stabilization of protein structure.,” *Biochemical Journal*, vol. 128, n. 4, p. 737, 1972.
- [6] C. B. Anfinsen, E. Haber, M. Sela e F. White Jr, “The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 47, n. 9, p. 1309, 1961.
- [12] N. A. Ranson, H. E. White e H. R. Saibil, “Chaperonins,” *Biochemical Journal*, vol. 333, n. 2, pp. 233–242, 1998.
- [13] R. Iizuka e T. Funatsu, “Chaperonin GroEL uses asymmetric and symmetric reaction cycles in response to the concentration of non-native substrate proteins,” *Biophysics and Physicobiology*, vol. 13, pp. 63–69, 2016.
- [15] S. B. Prusiner, M. R. Scott, S. J. DeArmond e F. E. Cohen, “Prion protein biology,” *cell*, vol. 93, n. 3, pp. 337–348, 1998.

- [16] B. Ruttkay-Nedecky, E. Sedlackova, D. Chudobova et al., “Prion protein and its interactions with metal ions (Cu^{2+} , Zn^{2+} , and Cd^{2+}) and metallothionein 3,” *ADMET and DMPK*, vol. 3, n. 3, pp. 287–295, 2015.

Risorse Online

- [2] “Protein Structure .BioNinja.” (15 apr. 2021), indirizzo: <https://ib.bioninja.com.au/standard-level/topic-2-molecular-biology/24-proteins/protein-structure.html> (visitato il 27/01/2022).
- [7] “Christian B. Anfinsen - Digital Collections - National Library of Medicine.” (1 gen. 2022), indirizzo: <http://resource.nlm.nih.gov/101408166> (visitato il 25/01/2022).
- [8] “File:RibonucleaseA SS paleRib.png - Wikimedia Commons.” (16 mar. 2012), indirizzo: https://commons.wikimedia.org/wiki/File:RibonucleaseA_SS_paleRib.png (visitato il 25/01/2022).
- [14] “Chaperonina - Wikipedia.” (21 ott. 2021), indirizzo: <https://it.wikipedia.org/wiki/Chaperonina> (visitato il 26/01/2022).

Altre fonti

- [1] A. Ruffini, *Fisiogenia, la biodinamica dello sviluppo ed i fondamentali problemi morfologici dell'embriologia generale*, 1925.
- [10] J. Murray, N. Laurieri e R. Delgoda, “Chapter 24 - Proteins,” in *Pharmacognosy*, S. Badal e R. Delgoda, cur., Boston: Academic Press, 2017, pp. 477–494, ISBN: 978-0-12-802104-0. DOI: <https://doi.org/10.1016/B978-0-12-802104-0.00024-X>. indirizzo: <https://www.sciencedirect.com/science/article/pii/B978012802104000024X>.