



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 | 5 |
| 1.2 Struttura delle proteine | 6 |
| 1.2.1 Legami e interazioni molecolari | 7 |
| 1.2.2 Livelli strutturali | 10 |
| 1.2.3 Evoluzione e classificazione | 16 |
| 1.3 Dinamica del ripiegamento | 18 |
| 1.3.1 Geometria ed energetica del ripiegamento | 18 |
| 1.3.2 Ripiegamento assistito | 21 |
| 1.3.3 Misfolding, prioni e malattie | 22 |
| 1.3.4 Controllo qualità e apoptosis | 25 |
| 1.4 Sfide al dogma di Anfinsen: IDP e fold switching | 26 |
| 1.4.1 Considerazioni epistemologiche | 28 |
| 1.5 Il problema del Protein Folding | 29 |
| Bibliografia | 31 |

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.

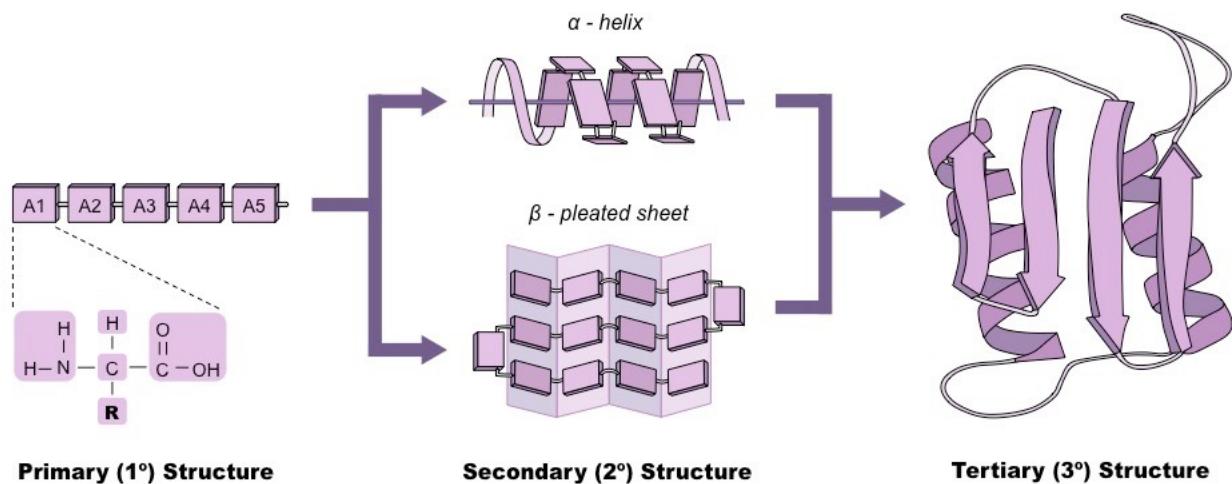


Figura 1.1: Protein folding: dagli amminoacidi alla struttura tridimensionale. Fonte: [2]

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 biologica².

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

²per una trattazione superficiale della mancanza di generalità di questo paradigma si veda la sezione

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 finale³.

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 aminoacidi di cui sono costituite. In altri termini: la struttura nativa, in ambiente fisiologico standard, corrisponde a quella struttura unica, stabile e cinematicamente accessibile avente *minima energia libera*.

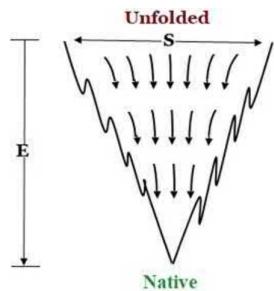


Figura 1.2: Un profilo energetico idealizzato dell'energia libera a forma di imbuto. E=energia, S=entropia.
Fonte: [5]

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
-
- 1.4. Per una scrittura e lettura più agevole della presente tesi si è preferito accettare la visione del paradigma

³ciò non è vero nel 100% dei casi, alcune proteine possono avere più di una conformazione stabile per adempire funzioni diverse (vedi la sezione 1.4) e alcune proteine possono andare incontro a misfolding (vedi la sezione 1.3)

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

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 ($-CH_2 - S-S - CH_2 -$, vedi sez. 1.2.1). 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.

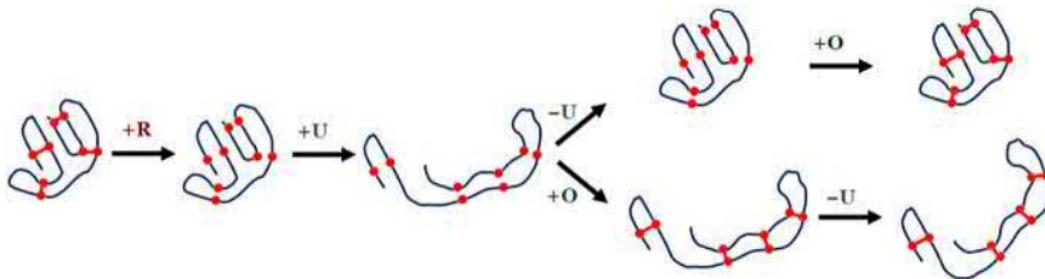


Figura 1.3: Rappresentazione schematica dell'esperimento di Anfinsen. R =reducing agent, U =Urea, O =oxidizing agent, punti rossi=cisteina, linee rosse=ponti disolfuro. Fonte: [5]

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.

Dai lavori di Anfinsen è possibile trarre due ulteriori importanti conclusioni^[9]:

- ha dato via alla grande avventura della ricerca nel campo del protein folding *in vitro*⁴ piuttosto che all'interno di cellule. La struttura nativa non dipendeva quindi dal fatto che la proteina fosse sintetizzata biologicamente con l'aiuto di ribosomi (ed

⁴La locuzione latina *in vivo* significa *nel vivente*. Se il fenomeno biologico viene riprodotto in una provetta si dice *in vitro* mentre se lo si riproduce tramite una simulazione computazionale si dice *in silico*.

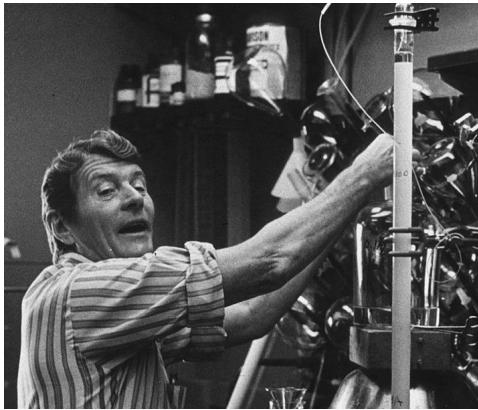


Figura 1.4: C.B. Anfinsen nel suo laboratorio.
Fonte: [7]

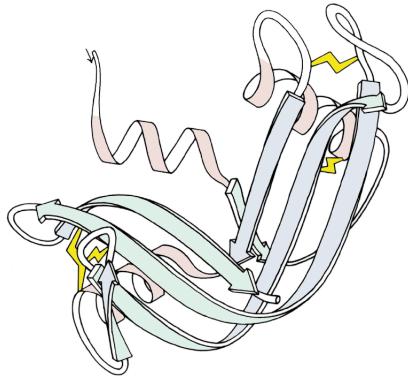


Figura 1.5: Ribonucleasi A, rappresentazione a nastro. In giallo i ponti disolfuro, rosa le α -eliche, verde e azzurro i β -foglietti. Fonte [8]

eventualmente chaperoni molecolari) o che si ripiegasse nuovamente come molecola isolata all'interno di una provetta

- l'evoluzione può agire in modo da cambiare la sequenza amminoacidica ma l'equilibrio del ripiegamento e la cinetica di una data sequenza sono materia della fisica chimica.

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

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).

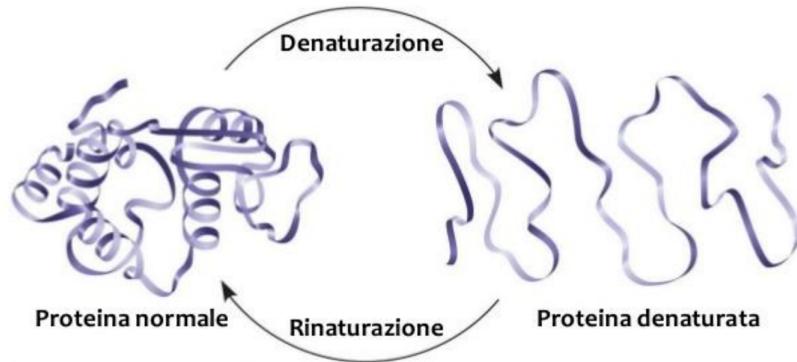


Figura 1.6: Denaturazione e rinaturazione. Fonte: [10]

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:

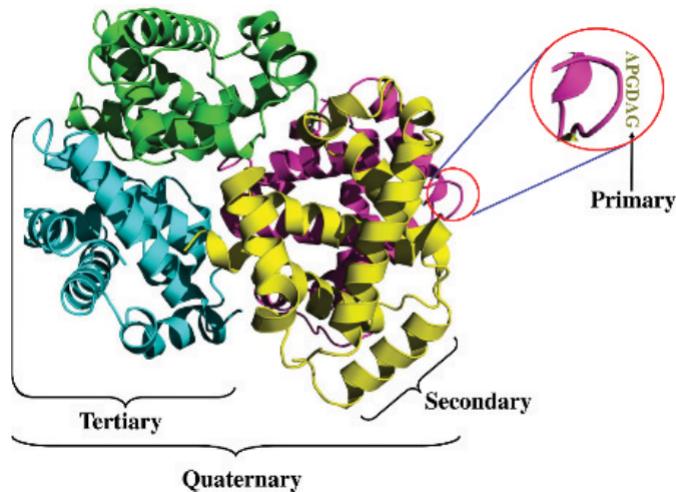


Figura 1.7: Livelli strutturali di una proteina. Fonte: [11]

- *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 ai legami chimici e alle interazioni molecolari.

1.2.1 Legami e interazioni molecolari

La chimica della vita è di un tipo speciale e niente disobeisce alle leggi della chimica e della fisica⁵. È una chimica organica formata da composti carboniosi, in un ambiente acquoso, con temperature "terrestri" e complicata, basata su grandi polimeri. Gli elementi non si possono modificare chimicamente ma gli atomi, particelle più piccole associate ad un elemento, possono risultare incompleti, e grazie a questo formare legami per completarsi. Elementi puri e pienamente completi non trovano spazio nella chimica della vita. Nei viventi solo gli elettroni si spostano⁶ per ricercare stabilità, ovvero per completare il loro guscio orbitale più esterno. Ogni atomo può avere tanti *legami* quanti elettroni gli mancano per completare il suo guscio più esterno. Le *interazioni molecolari* sono forze attrattive o repulsive tra molecole e tra atomi non legati. La *forza di legame* è la misura dell'energia necessaria per romperlo (in kJ/mol o kcal/mol).

- *legame covalente*: prevede la partecipazione di 2 elettroni di valenza fra più atomi ed è il tipo di legame più forte. Due o più atomi tenuti insieme da legami covalenti formano una molecola. C'è una specifica distanza di legame fra i nuclei degli atomi bilanciata tra forze attrattive e repulsive: se sono troppo vicini c'è repulsione mentre se sono troppo lontani non c'è attrazione.
 - *elettronegatività*: spesso gli elettroni in un legame sono condivisi iniquamente. Questo dipende dall'elettronegatività degli atomi, ad esempio l'ossigeno ha elettronegatività 3.4 mentre l'idrogeno 2.1. Quando la differenza di elettronegatività è compresa tra 0.5 e 1.9 la nube elettronica di legame risulta deformata verso l'atomo più elettronegativo, su cui si origina una carica parziale negativa (indicata con δ^-) mentre l'altro atomo acquisisce una carica parziale positiva di uguale valore assoluto. La molecola, divenuta *polare*, si può immaginare ora come un *dipolo* elettrico.
 - *ponti disolfuro*: i legami (o ponti) disolfuro sono legami covalenti tra due atomi di zolfo con energia di legame di 60kcal/mol. Si formano dall'accoppiamento di

⁵Assumendo un paradigma meccanicistico, come accennato all'inizio di questa tesi

⁶Protoni e neutroni si separano solo in condizione estreme: nei reattori nucleari, nel sole, per decadimento radioattivo ...

due gruppi tiolici (-SH). Essendo legami molto forti costituiscono un elemento architettonale fondamentale nella struttura delle proteine. La cisteina presenta un gruppo -SH nella catena laterale e può quindi formare ponti disolfuro.

| Bond Type | Length* (nm) | Strength (kJ/mole) | |
|--|--------------|--------------------|-----------|
| | | In Vacuum | In Water |
| Covalent | 0.10 | 377 [90]** | 377 [90] |
| Noncovalent: ionic bond | 0.25 | 335 [80] | 12.6 [3] |
| Noncovalent: hydrogen bond | 0.17 | 16.7 [4] | 4.2 [1] |
| Noncovalent: van der Waals attraction (per atom) | 0.35 | 0.4 [0.1] | 0.4 [0.1] |

Figura 1.8: Distanza di legame approssimate e forza dei legami chimici. I valori della forza sono riportati in kJ/mol e in [kcal/mol]. Da notare la diminuzione di forza nel legame ionico se in ambiente acquoso.
Fonte: [12]

- *legami non covalenti (interazioni molecolari)*
 - *attrazioni elettrostatiche*: le forze d'attrazione agiscono fra gruppi completamente carichi (legame ionico) e fra i gruppi parzialmente carichi delle molecole polari. Decresce con la distanza. Molto deboli in acqua.
 - * *legame ionico*: l'atomo più elettronegativo strappa completamente un elettrone al suo compagno, si formano due ioni (uno positivo, *catione* e uno negativo *anione*). Si ha quando la differenza di elettronegatività tra i due atomi è maggiore di 1.9.
 - *legame idrogeno*: è una forza dipolo-dipolo che si origina tra molecole contenenti un atomo di idrogeno unito covalentemente (a ossigeno, fluoro o azoto). Un atomo di idrogeno elettropositivo è parzialmente condiviso da due atomi elettronegativi; 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 stabilizzare le molecole. Sono legami deboli singolarmente (1/20 della forza di un legame covalente) ma quando se ne formano simultaneamente molti sono abbastanza forti da fornire un legame stretto (l'acqua bollirebbe a -120°C senza legami idrogeno).
 - *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 di-

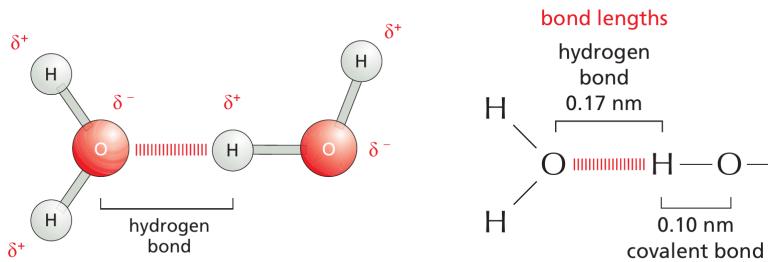


Figura 1.9: Legame idrogeno tra due molecole d'acqua. Fonte: [12]

stanza. Due atomi saranno attratti l'uno dall'altro fino a che la distanza fra i loro nuclei è approssimativamente uguale alla somma dei loro raggi di van der Waals (ad es. per il carbonio il raggio è di 0.2nm)

- *forze idrofobiche*: l'acqua forza insieme i gruppi idrofobici; l'apparente attrazione è in realtà causata da una repulsione dall'acqua, che difende il suo reticolo tenuto insieme da legami idrogeno.

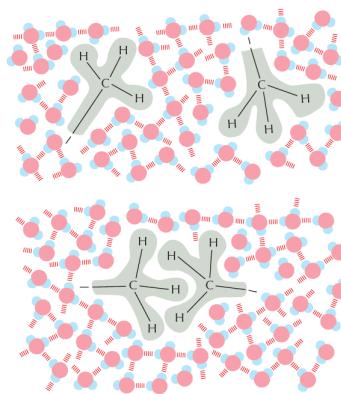


Figura 1.10: Forze idrofobiche. Fonte: [12]

Le sostanze *idrofile* si dissolvono rapidamente nell'acqua poiché le loro molecole formano legami idrogeno con le circostanti molecole d'acqua (nel caso di sostanze polari) o perché queste sono attratte dalle cariche degli ioni (nel caso di sostanze ioniche, es. cloruro di sodio, con ioni Na^+ e Cl^-). Le sostanze *idrofobiche* contengono perlopiù legami non polari e sono solitamente insolubili in acqua. Le molecole d'acqua in questo caso non sono attratte ma possono generarsi forze idrofobiche che raggruppano insieme tali sostanze (come nel nucleo idrofobico delle proteine).

1.2.2 Livelli strutturali

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.

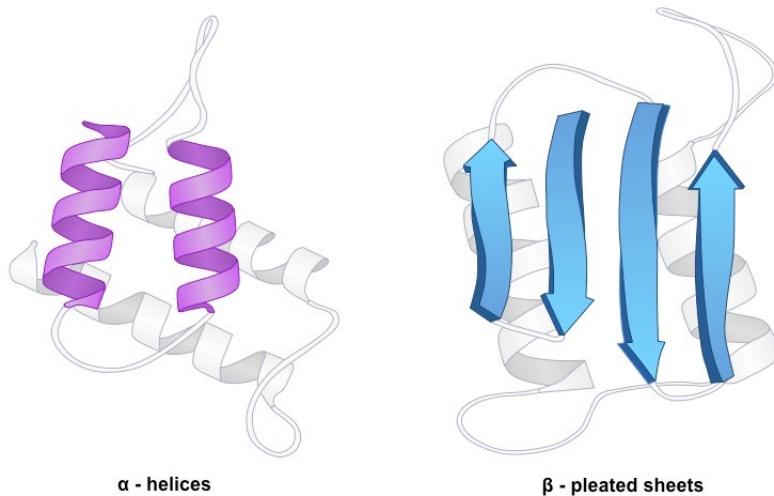


Figura 1.11: Struttura secondaria delle proteine, α -eliche e β -foglietti. Fonte: [2]

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, la struttura secondaria più comune e teorizzata già negli anni '50 da Linus Pauling, 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 a distanza di tre 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 di capelli, lana e unghie hanno formazioni di α -eliche sulla maggior parte della loro lunghezza.

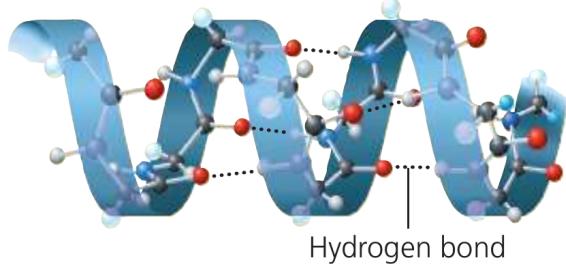


Figura 1.12: Regione di α -elica. Fonte: [10]

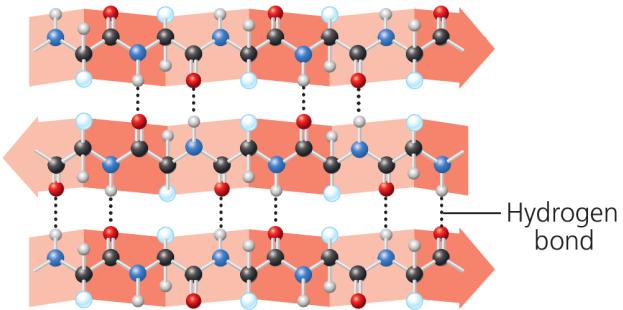


Figura 1.13: Una regione di β -foglietto composto da β -filamenti adiacenti, spesso mostrati come una freccia pieghezzata o piatta puntata in direzione C-terminus. Fonte [10]

Altre proteine fibrose sono invece dominate dai β -foglietti, come le proteine della seta (β -cheratina) e della 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. I β -foglietti tendono a trovarsi all'interno del nucleo della struttura per evitare competizione con le molecole d'acqua per formare legami idrogeno e tendono a favorire residui idrofobici. Si dice che i filamenti sono paralleli quando vanno nella stessa direzione (la freccia che indica la direzione C-terminus è puntata nella stessa direzione).

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

La struttura supersecondaria è riferita alle combinazioni spaziali di strutture secondarie in conformazioni più complesse e alle connessioni che li uniscono. Può essere considerata come esempio di struttura supersecondaria la triplice elica allungata del collagene.

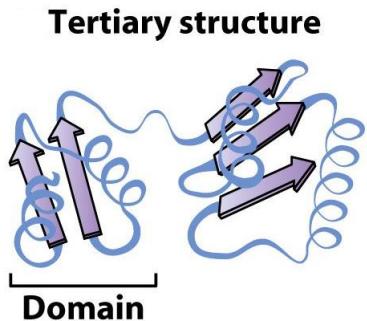


Figura 1.14: Dominio in una proteina. Fonte: [13]

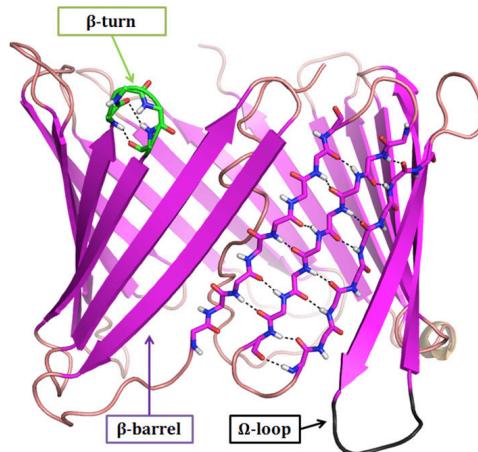


Figura 1.15: Struttura con giri, loop e motivo β -barile. Fonte: [14]

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. Una proteina con vari domini può usare questi per interazioni funzionali con differenti molecole.

Più in generale un *motivo strutturale* è una struttura tridimensionale comune che appare in una varietà di molecole differenti ed evoluzionisticamente scollegate. Nel contesto delle sequenze amminoacidiche si definisce *motivo* un pattern amminoacidico conservato in un gruppo di proteine con attività biochimica simile.

In figura 1.16 sono illustrati alcuni motivi comuni nelle strutture proteiche. Il motivo *elica-loop-elica* ad esempio consiste di due α -eliche collegate da un giro invertito. Un

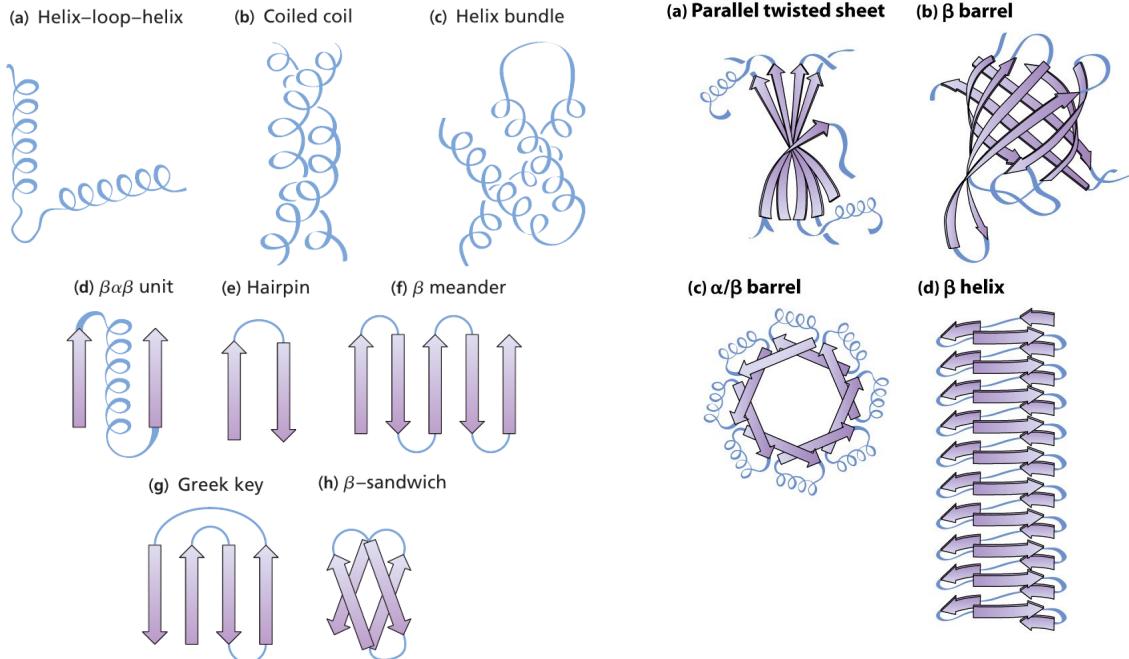


Figura 1.16: Motivi comuni. Fonte [13]

Figura 1.17: Domain folds (ripiegamenti di dominio). Fonte: [13]

motivo simile è l'*elica-giro-elica* dove al posto di un loop si ha un giro che causa un cambio di direzione più netto. Questa particolare conformazione rende questo motivo in grado di legarsi alla scanalatura del DNA e infatti questo motivo si presenta in molte proteine che regolano l'espressione genica.

Il motivo a *simbolo greco* consiste di 4 β -filamenti antiparalleli in un β -foglietto dove l'ordine dei foglietti lungo la catena polipeptidica è 4,1,2,3⁷. In figure 1.15 e 1.17 è illustrato il motivo β -barile composto da β -foglietti ripiegati circolarmente a formare una struttura somigliante ad un barile comune in molte proteine di membrana. I *domain folds*, o ripiegamenti di dominio, sono grandi motivi che costituiscono il nucleo di un dominio.

Giri e *loop* causano cambi di direzione alla backbone della proteina. I loop sono regioni con una struttura tridimensionale fissa ma non regolare. Si trovano generalmente sulla superficie delle proteine. Non sono strutture casuali e non vanno confuse con regioni disordinate o dispiegate. Hanno principalmente lo scopo di connettere strutture secondarie tra loro. È stato ipotizzato che la posizione degli introni nel DNA possa correlare con la locazione dei loop codificati nella proteina^[15].

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

⁷I numeri indicano l'ordine dei filamenti ovvero la loro posizione nel β -foglietto da destra a sinistra

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 possibili solamente fra due cisteine.

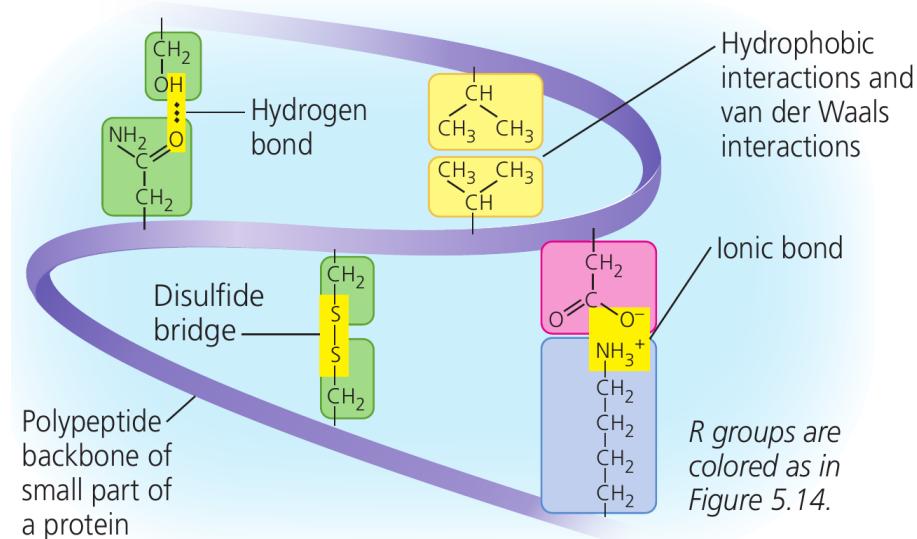


Figura 1.18: I diversi tipi di interazioni che possono contribuire alla struttura terziaria di una proteina.
Fonte: [10]

La glicina assume una speciale posizione tra gli amminoacidi dato che ha il gruppo R più piccolo, un solo atomo di idrogeno (vedi fig. ??): può aumentare la flessibilità locale nella struttura (come infatti accade nel caso dei *giri* sopra accennati).

Prima degli anni '80 il protein folding code (bilancio termodinamico delle forze interatomiche, vedi sez. 1.5) era visto come la somma di molte piccole interazioni (legami idrogeno, interazioni di van der Waals, attrazioni elettrostatiche ...) ma senza nessuna forza dominante^[9]. Negli anni '80, grazie alla modellazione basata sulla meccanica statistica, è emerso un nuovo paradigma: la componente dominante nel folding code sono le forze idrofobiche, il folding code è distribuito sia localmente che non localmente nella sequenza e le strutture secondarie di una proteina sono una conseguenza della struttura terziaria, tanto quanto una causa. Poiché le strutture native sono solamente 5-10kcal/mol

più stabili dei loro stati denaturati è chiaro che nessuna forza intermolecolare può essere ignorata, e per questo la questione su quale forza sia quella dominante non è né semplice né risolta. Tuttavia risulta evidente che solo pochi residui risultano carichi nelle proteine, pertanto le forze elettrostatiche difficilmente possono essere dominanti.

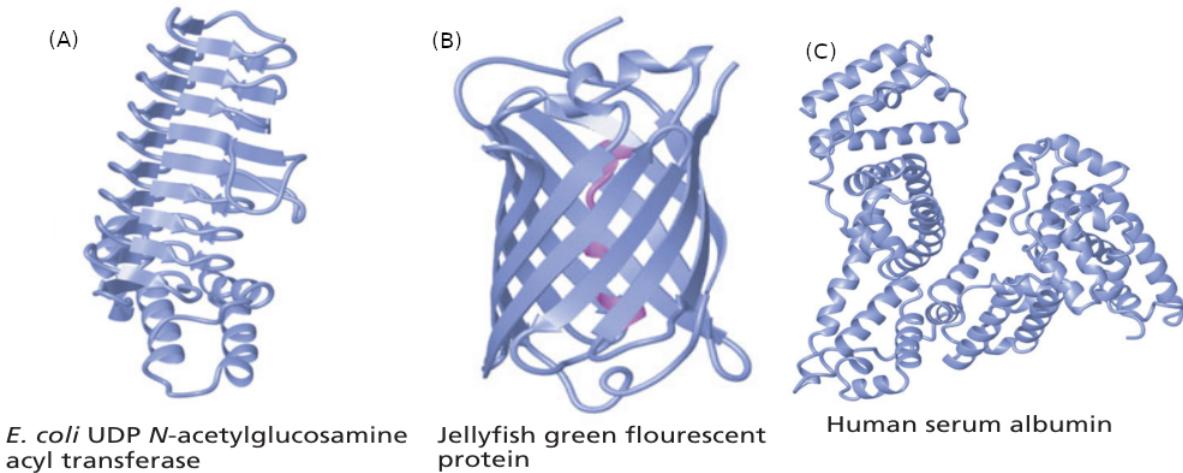


Figura 1.19: Esempi di strutture terziarie in alcune proteine. (A) (classe: all- β) La struttura dell'enzima mostra un classico esempio di β -eliche, struttura abbastanza rara. (B) (classe: all- β) Struttura a β -barile con un' α -elica centrale; i β -filamenti sono anti-parallel. (C) (classe: all- α) Albumina del siero umano. Ha molti domini costituiti da α -eliche a strati e helix bundle. Fonte [13]

Struttura quaternaria

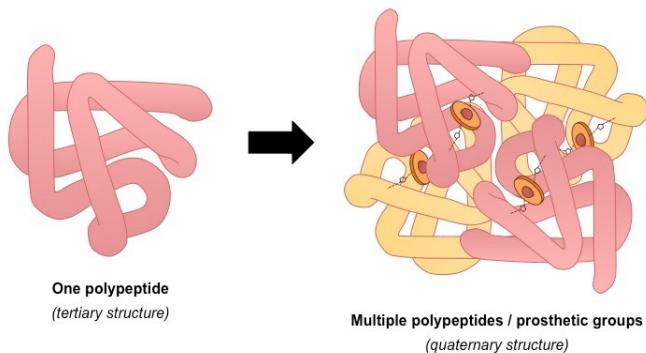


Figura 1.20: Rappresentazione di una struttura quaternaria composta da più polipeptidi e alcuni gruppi prostetici. Fonte [2]

La struttura quaternaria è la forma finale di proteine "assembrate" 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 (vedi anche la cheratina nella

sezione 1.2.3). Un altro esempio è l'emoglobina, proteina globulare formata da 4 subunità polipeptidiche. Le strutture terziarie delle subunità non vengono alterate.

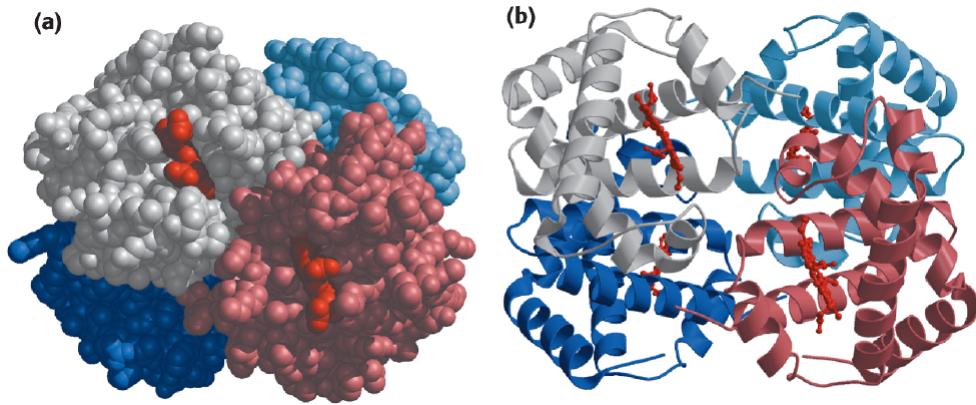


Figura 1.21: Struttura quaternaria della deossiemoglobina, ogni colore rappresenta una diversa subunità. (A) rappresentazione di tipo space-fill. (B) Rappresentazione a nastro. Fonte [13]

La struttura che ne risulta è spesso chiamata *oligomero* e le catene polipeptidiche costituenti sono dette *monomeri*, *protomeri* o *subunità*. Le subunità sono unite mediante legami idrogeno, ionici o forze idrofobiche. I rapporti spaziali tra le subunità sono fissi e la geometria della molecola globale è ben definita. L'unione delle subunità può far emergere proprietà non possedute dai singoli monomeri.

1.2.3 Evoluzione e classificazione

Evoluzione e conservazione

L'evoluzione degli organismi è legata a mutazioni spontanee che avvengono nei loro geni. Le differenze nella struttura primaria sono la “memoria” dei cambiamenti avvenuti a livello genetico nel corso dell’evoluzione. In specie legate da notevole affinità le strutture primarie delle proteine comuni sono simili. Proteine con funzioni analoghe presentano sequenze simili, è molto probabile quindi che queste sequenze si siano evolute a partire da un progenitore comune.

Al contrario di quanto si possa credere la maggior parte dei motivi non ha origini evolutive in comune. Motivi simili sono sorti indipendentemente e semplicemente convergono verso una struttura stabile comune. Il fatto che gli stessi motivi si presentino in centinaia di differenti strutture suggerisce che l'esistenza di un numero limitato di possibili ripiegamenti nell'universo delle strutture proteiche^[16].

Classificazione

La classificazione delle proteine all'interno dei database può essere basata su somiglianze strutturali e/o di sequenza. A livello biologico, in base ai diversi livelli strutturali assunti, le proteine sono classificate in *fibrose* e *globulari*.

Le proteine fibrose sono caratterizzate dalla prevalenza di strutture secondarie rispetto a livelli di organizzazione superiore. Sono costituite da lunghe catene disposte in lunghi fasci o foglietti. La struttura è estremamente ordinata. Svolgono funzioni di protezione e sostegno. Le proteine fibrose costituiscono prevalentemente: pelle, piume, capelli, corna, unghia, squame (con funzione di protezione) e cartilagine, tendini, ossa (con funzione di sostegno). Contengono per la maggior parte residui idrofobici pertanto le proteine fibrose risultano insolubili in acqua. Degli esempi di proteine fibrose sono: cheratina, collagene, elastina, fibroina.

In figura 1.22 è mostrata la struttura di un capello. La proteina dei capelli è l' α -cheratina, una struttura totalmente ad α -eliche. Un paio di queste eliche si attorcigliano per formare una doppia elica. Queste si combinano poi in strutture di ordine superiore chiamate *protofilamenti* e successivamente *protofibrille*. Circa 4 fogli protofibrille ($4 \times 2^3 = 32$ eliche di α -cheratina in tutto) si combinano per formare un filamento intermedio. Un capello è una schiera di filamenti intermedi di α -cheratina.

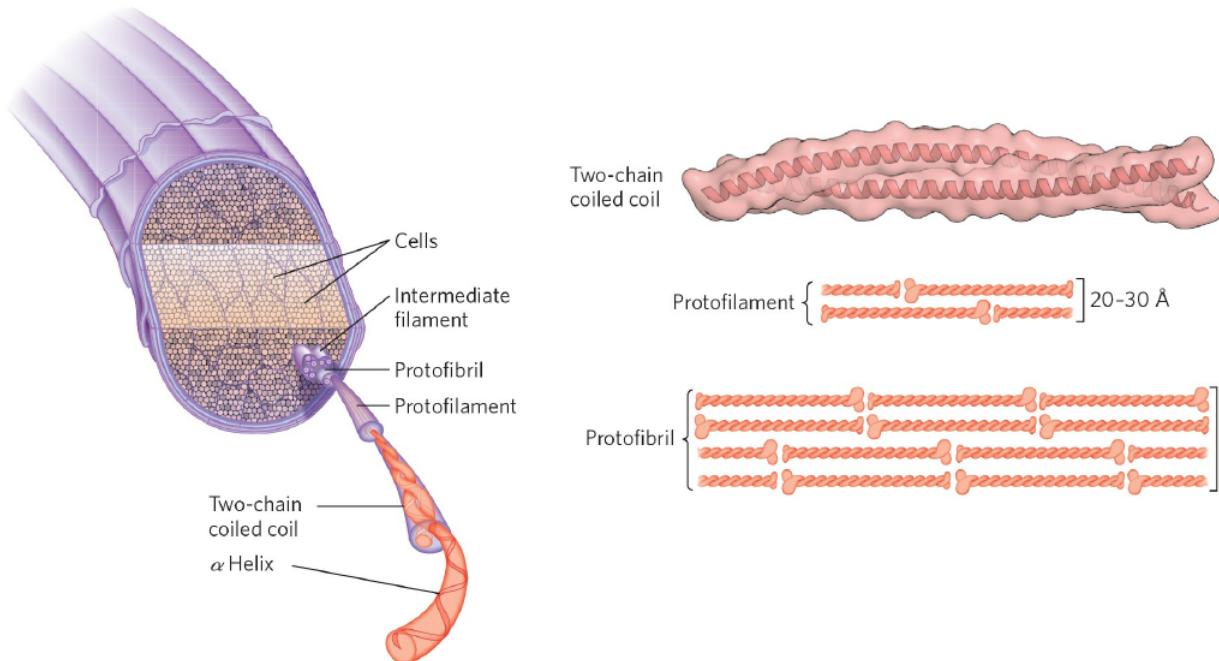


Figura 1.22: Struttura di un capello. Fonte: [17]

Le proteine *globulari* assumono una struttura terziaria e a volte quaternaria. Sono macromolecole compatte di forma all'incirca sferica. Hanno una struttura meno ordinata

rispetto a quelle fibrose. Svolgono funzioni di catalisi, trasporto e regolazione di processi cellulari. Categorie di proteine globulari sono: enzimi, trasportatori di ossigeno e lipidi, alcuni ormoni, recettori di membrana e anticorpi. La struttura è caratterizzata da brevi tratti di α -elica e struttura β , collegate da tratti non organizzati in struttura secondaria. Sono proteine solubili nel citosol.

1.3 Dinamica del ripiegamento

1.3.1 Geometria ed energetica del ripiegamento

Geometria del ripiegamento

Le tante conformazioni possibili della catena polipeptidica sono possibili grazie alla rotazione di essa attorno all'atomo C_α .

Considerando un singolo residuo AA, C funziona da perno unendo due piani ammidici adiacenti. L'angolo di rotazione C -C=O è indicato con Ψ (psi), quello C -NH con Φ (phi). Gli angoli sono calcolati rispetto ad un piano di riferimento che contiene entrambi i legami ammidici (Ψ e $\Phi = 0^\circ$)

conformazioni ideali (in cui non vi sono interazioni tra catene laterali)

-elica: $\Phi = -57^\circ$ $\Psi = -47^\circ$

- parallelo: $\Phi = -119^\circ$ $\Psi = +113^\circ$

-antiparallelo: $\Phi = -139^\circ$ $\Psi = +135^\circ$

Il carattere di doppio legame parziale del legame peptidico impedisce, alle temperature fisiologiche, la rotazione del legame C-N, mentre è possibile la rotazione dei legami C 1-C e C 2-N. Data la NON possibilità di rotazione intorno al legame CO-N si ha possibilità di isomeria cis trans Delle due configurazioni possibili, la trans è quella favorita dal punto di vista energetico (minima repulsione sterica) Oltre il 99% dei legami peptidici delle proteine naturali hanno configurazione trans Fanno eccezione alcuni legami amidici in cui è coinvolto l'N imidico della prolina

In conformazioni diverse, c'è un limite al numero delle combinazioni possibili dei due angoli di rotazione, perché alcune hanno effetti destabilizzanti a causa delle forze di repulsione tra gli O. Vi sono quindi conformazioni più stabili di altre perché favorite da un punto di vista energetico. Conformazioni proibite sono anche quelle in cui si svilupperebbe una forte repulsione tra le catene laterali

- Grafico di Ramachandran

Torsion or dihedral angles (Φ , Ψ) represent the position of the atoms of an amino acid chain where Φ involves the N-C α bond and Ψ involves the C α -C bond. These rotations determine each protein structure. A possible representation of the amino acids of a protein

is $[(\Phi_1, \Psi_1), \dots, (\Phi_n, \Psi_n)]$ where n reflects the residue length. The Ramachandran plot, allows to avoid the possible collisions among atoms.

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

Energetica del ripiegamento

La termodinamica caratterizza gli stati in natura dalla dipendenza da temperatura, pressione, volume e concentrazione chimica.

L'entropia è la quantità di disordine di un sistema, in altri termini è il numero di possibili stati configurazionali del sistema (Ω).

$$S = K_B \ln \Omega$$

Nei sistemi molecolari cambiamenti di entropia (ΔS) rappresentano cambiamenti nella libertà di movimento di atomi appartenenti sia al soluto che al solvente. L'entalpia è l'energia interna al sistema. Cambiamenti positivi di entalpia nelle macromolecole sono associate alla rottura di interazioni non covalenti favorevoli (si assume che i legami covalenti restino invariati). La formazione di legami covalenti diminuisce l'entalpia del sistema e rilascia calore verso l'ambiente. Formando legami si libera energia mentre per romperli si consuma energia.

$$\Delta H \simeq \Delta E$$

$$E = U + K$$

Dove H è l'entalpia, E l'energia interna al sistema, U l'energia potenziale (circa la somma di tutte le interazioni covalenti e non covalenti, ma non quelle non polari) e K l'energia cinetica (associata ai movimenti atomici indotti termicamente).

L'energia libera di Gibbs (G) è l'*energia utile* sotto temperatura e pressione costante. I processi spontanei raggiungono l'equilibrio decrementando la G del sistema a un minimo.

L'energia libera di Gibbs è associata all'entropia e all'entalpia dalla seguente relazione:

$$\Delta G = \Delta H - T\Delta S$$

La seconda legge della termodinamica afferma che in un sistema isolato i processi spontanei raggiungono l'equilibrio incrementando l'entropia del sistema.

Il ripiegamento delle proteine è un processo spontaneo. Esibisce una grande varietà di percorsi, meccanismi e velocità che dipendono da parametri come la composizione della proteina e le condizioni di ripiegamento. A prescindere dall'esatto meccanismo, il processo

di ripiegamento segue sempre la teoria del profilo energetico (*energy landscape theory*), cioè il ripiegamento è sempre accompagnato da una decrescita del numero di conformazioni che possono essere testate dalla proteina, sfuggendo al paradosso di Levinthal.

Una proteina che si ripiega deve procedere da uno stato ad alta energia ed alta entropia a uno stato caratterizzato da bassi valori di energia ed entropia; tale nesso è conosciuto come *imbuto di ripiegamento* (folding funnel, vedi fig. 1.23). Le forze idrofobiche causano la creazione del nucleo idrofobico seppellendo i residui non polari nella struttura nativa, questo vuol dire che l'entropia del solvente acquoso subisce un aumento portando ad una sovraccarenza dell'entropia, ovvero l'entropia del sistema aumenta.

Le proteine non cercano una conformazione fra tutte le possibili conformazioni fino a trovare quella giusta. Piuttosto si ripiegano in una maniera cooperativa in cui ogni step limita ulteriormente le possibilità di ripiegamento degli step seguenti. Il ripiegamento è completato quando la conformazione con più basso livello di energia libera associata alla proteina è trovata. La superficie rugosa del tunnel riflette il fatto che il processo di ripiegamento passa attraverso molti minimi locali separati da barriere da alta-energia.

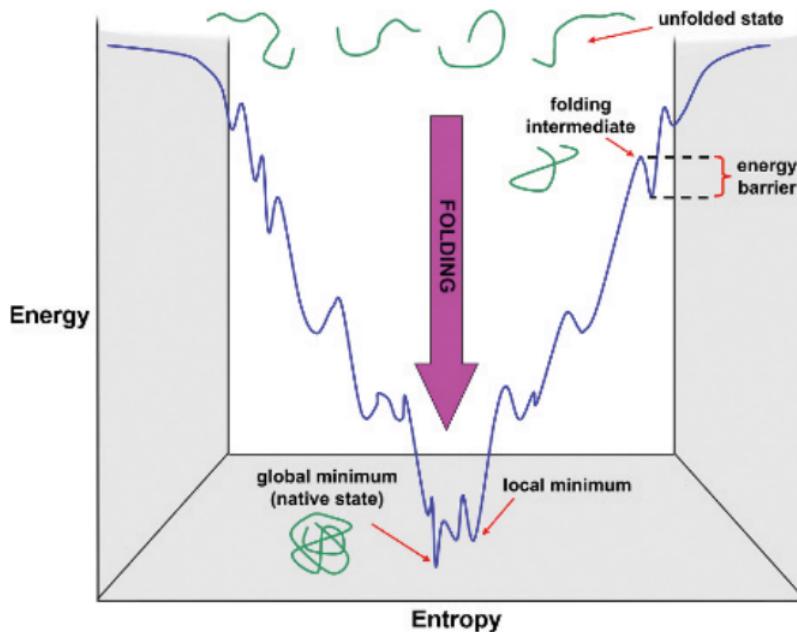


Figura 1.23: Profilo energetico a imbuto del ripiegamento. Fonte [11]

Non c'è un meccanismo di ripiegamento universale, ma una collezione di possibili meccanismi che possono essere usati. La preferenza di una proteina ad usarne uno piuttosto che un altro può dipendere da vari fattori, uno dei quali sembra essere la struttura secondaria. Ad esempio in proteine dominate da α -eliche la struttura è spesso gerarchica.

La velocità di ripiegamento correla non solo con la dimensione della proteina ma anche con la sua topologia nativa. Le proteine *fast-folding*, su scala temporale di nanosecondi,

tendono ad avere grandi proporzioni di elementi secondari locali (α -eliche e giri) laddove quelle *slow-folding* tendono ad avere proporzioni più grandi di elementi globali (β -foglietti).

I fattori termodinamici che danno luogo all'effetto idrofobico sono complessi e non del tutto conosciuti. L'effetto idrofobico è visto come una combinazione dell'effetto di idratazione (effetto entropico) e di interazioni di van der Waals tra molecole di soluto (effetto entalpico).

1.3.2 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.

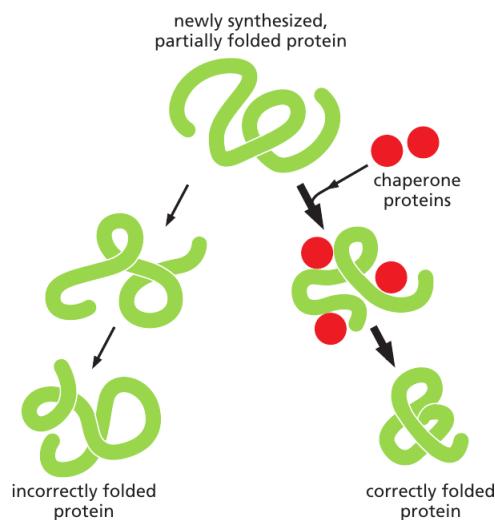


Figura 1.24: Schema della funzione dei chaperoni molecolari. Fonte: [12]

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 corretto ripiegamento delle proteine. I chaperoni molecolari comprendono diverse famiglie di proteine altamente conservative, 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^[18], famosa è la chaperonina procariotica GroEL (vedi fig. 1.25), che può essere assunta come modello di riferimento delle chaperonine.

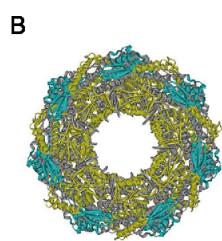
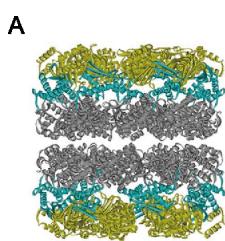


Figura 1.25: Strutture dei complessi GroEL e GroEL-GroES. (B) si può osservare la tipica forma ad anello.

Fonte: [19]

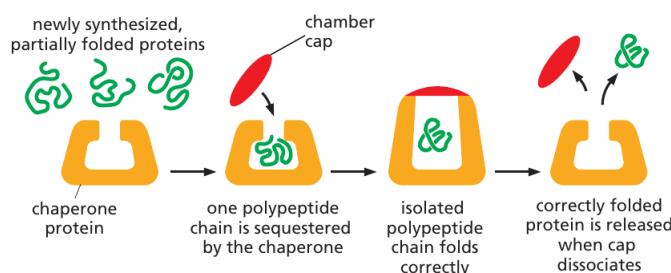


Figura 1.26: Rappresentazione schematica della funzione della camera di isolamento nelle chaperonine. Fonte [12]

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^[12].

1.3.3 Misfolding, prioni e malattie

Misfolding

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.

Malattie

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^[20].

Prioni

I *prioni* (acronimo di "proteinaceous infective **only** particle") sono molecole di natura proteica con la capacità di trasmettere la propria forma mal ripiegata a varianti normali della stessa proteina⁸. 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.

⁸I prioni sono stati studiati e denominati in questo modo dal premio Nobel per la medicina nel 1997 Stanley Prusiner^[21]

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 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 prione 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 **prione**, e P per **proteina**), 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.

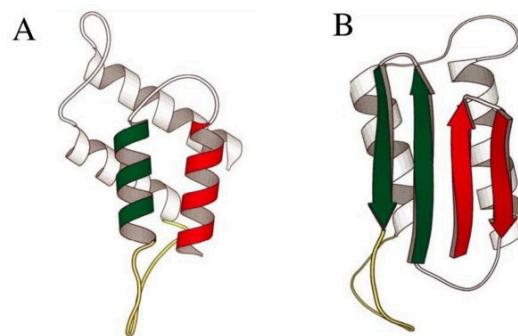


Figura 1.27: (A) Struttura della PrP^c. (B) Struttura della PrP^{sc}. Fonte: [22]

1.3.4 Controllo qualità e apoptosis

Controllo qualità

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'ubiquitina⁹.

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^[12] (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.

Proteasomi

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.

⁹Per "la scoperta della degradazione delle proteine mediata da ubiquitina" è stato assegnato il Premio Nobel per la chimica del 2004 ad Aaron Ciechanover, Avram Hershko ed Irwin Rose.

A causa del ruolo dei proteasomi nella regolazione del ciclo cellulare e dell'apoptosi¹⁰, 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.

Unfolded protein response e Apoptosi

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 sovrapposto. 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'*apoptosis*.

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^[12].

1.4 Sfide al dogma di Anfinsen: IDP e fold switching

Il ripiegamento delle proteine in una cellula è un processo molto complesso che riguarda il trasporto di nuove proteine sintetizzate ad appropriati compartimenti cellulari attraverso targeting, misfolding, stati dispiegati temporanei, modifiche post-traduzione, controllo qualità, aggregazione in complessi, facilitazione dei chaperoni molecolari. Come già spiegato, l'aiuto dei chaperoni molecolari non sfida il dogma di Anfinsen, in quanto non influenza la struttura nativa della proteina.

¹⁰Il termine *apoptosis* 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

La struttura di alcune proteine è difficile da determinare per una semplice ragione: un crescente numero di ricerche biochimiche ha rivelato che un numero significativo di proteine, o regioni di proteine, non hanno una distinta struttura 3D o non ce l'hanno finché non interagiscono con la molecola target oppure cambiano struttura nativa. La loro flessibilità e struttura indefinita è importante per la loro funzione, che potrebbe richiedere legami con differenti target in tempi diversi. Si stima che queste proteine possano contribuire per il 20-30% al proteoma dei mammiferi.

Intrinsically disordered proteins

Una proteina intrinsecamente disordinata (IDP) è una proteina, o una regione di essa, a cui manca una struttura terziaria fissa od ordinata. Le IDP sono comunemente riconosciute come regioni mancanti di densità elettronica in strutture di proteine determinate a cristallografia a raggi X. Molte IDP possono adottare una struttura tridimensionale stabile dopo essersi legate ad altre macromolecole, passando per transizioni disordine-ordine e perciò risultare strutturate per un certo periodo di tempo e non strutturate per altro. Ci sono però anche IDP che svolgono la loro funzione senza assumere mai una forma ordinata attraverso la loro esistenza. Nonostante la loro mancanza di struttura stabile le IDP risultano essere una classe importante e grande di proteine.

Fold switching proteins

Alcune proteine hanno multiple strutture native: possono cambiare la loro forma, rimodelando anche le loro strutture secondarie, in base a fattori esterni. Ad esempio il complesso proteico KaiB cambia ripiegamento durante la giornata agendo da orologio per i cianobatteri. Si possono immaginare le proteine *fold switching* come una sorta di transformer¹¹ dove in un caso la proteina è come un robot che fa una cosa e in un altro caso, in risposta a cambiamenti ambientali, diventa un'automobile e fa qualcos'altro. Il cambio tra strutture alternative è guidato da interazioni della proteina con piccoli ligandi o altre proteine, da modificazioni chimiche (es. fosforilazione) o da cambiamenti nelle condizioni ambientali (temperatura, pH, potenziale di membrana). Ogni struttura alternativa può o corrispondere al minimo globale di energia libera della proteina in certe condizioni o essere cinematicamente intrappolata in un alto minimo locale di energia libera^[25].

Le proteine *fold switching* (FS) differiscono dalle IDP^[24]:

- le FS richiedono che entrambe le loro conformazioni siano determinate, le IDP sono regioni non determinate

¹¹metafora di Lauren Porter^[23], autrice di^[24]

- le IDP sono caratterizzate da sequenze aminoacidiche caratteristiche mentre le FS sono libere da questo vincolo
- le IDP, in isolamento, non si ripiegano cooperativamente mentre le FS si ripiegano, in isolamento, sia cooperativamente che indipendentemente

Per queste ragioni si può affermare che le FS non sono IDP, piuttosto sono un sottinsieme delle proteine globulari le cui strutture stabili cambiano drasticamente in risposta al loro ambiente. I vantaggi di una proteina FS sono legati alla sua bifunzionalità:

- può affrontare velocemente richieste biologiche ovviando al bisogno di risorse cellulari aggiuntive per trascrivere e tradurre un'altra proteina. Un esempio è RfaH: funziona sia da fattore di trascrizione che di traduzione
- regolazione di inattività, può essere bloccata in uno stato di attività o inattività finché non viene innescato uno specifico segnale

È stato stimato che una percentuale tra lo 0.5 e il 4% delle proteine nel PDB cambi ripiegamento^[24].

1.4.1 Considerazioni epistemologiche

La scoperta delle IDP e delle proteine FS ha creato una spaccatura nel dogma della struttura rigida delle proteine, secondo il quale la struttura deve essere fissa al fine di compiere la propria funzione biologica. È interessante ripercorrere velocemente la storia di questo paradigma attraverso un breve excursus storico per poterne mettere in luce le relazioni epistemologiche soggiacenti.

Nel 1894 Fischer ha proposto una metafora di *chiave e serratura* per spiegare come fosse possibile un effetto chimico tra un enzima e un glucoside:

«*Per usare una metafora, vorrei dire che l'enzima e il glucoside devono adattarsi l'uno all'altro come una serratura e una chiave per esercitare un effetto chimico l'uno sull'altro*^{[26][27]}»

Nel 1936 Mirsky e Pauling hanno raccolto una quantità di informazioni sufficiente per concludere:

«*attribuiamo le specifiche proprietà caratteristiche delle proteine native alle loro uniche e definite configurazioni. Consideriamo le proteine denaturate essere caratterizzate dall'assenza di un'unica definita configurazione*^[28]».

Né Mirsky, né Pauling né Hsien Wu (probabilmente il primo a proporre il paradigma struttura unica-funzione) citarono il lavoro di Fischer ma la sua metafora avrebbe supportato pienamente le loro tesi. Perciò ancora prima dell'esperimento di Anfinsen e delle risoluzioni atomiche delle strutture, una specifica e ben ordinata forma tridimensionale

era stata accettata come prerequisito essenziale per la funzione di una proteina.^[27]. Le prime strutture proteiche sono state determinate attraverso la cristallografia negli anni '50, dando ancora più credito all'ipotesi che una struttura fissa fosse necessaria per adempiere la funzione biologica. Queste pubblicazioni hanno contribuito a solidificare il dogma centrale della biologia molecolare. Nonostante questo già nel 1950 si parlava di molte possibilità di strutture per una proteina, in particolare Karush^[30] sull'albumina del siero bovino, inferendo che le interazioni proteina-ligando stabilizzassero il membro più adatto da un insieme di strutture in equilibrio, chiamando questo fenomeno *adattabilità configurazionale*. Successivamente si può ricordare il paradosso di Levinthal negli anni '60, per arrivare negli anni '70 al dogma di Anfinsen, il cui paradigma (sequenza amminoacidica -> struttura tridimensionale -> funzione) è messo in discussione da evidenze di mancanza di generalità. Tuttavia quel sistema di pensiero ha preso piede ed ha pervaso quasi tutti i lavori e teorie successive. Al tempo dell'esperimento di Anfinsen e delle prime risoluzioni atomiche della mioglobina e del lisozima il prerequisito di una forma tridimensionale necessario per la funzione della proteina era già accettato. La successiva valanga di migliaia di strutture determinate ha contribuito ad affossare paradigmi di pensiero alternativi.

La scoperta di segmenti intrinsecamente disordinati è stata compiuta nel 1978, quando ancora erano disponibili le strutture di sole 20 proteine. Alcuni segmenti di proteine non fornivano alcuna densità elettronica distinguibile nonostante fossero essenziali per il funzionamento. Una ragione comune è che gli atomi di quel segmento sono disordinati, ovvero la loro posizione cambia. Con l'avvento del NMR, sempre nel 1978, si è arrivati a identificare intere proteine disordinate e dato che questo metodo è più preciso, la riscoperta di disordine nativo ha avuto un impatto significativo.

Fino al 2000^[29] queste idee non sono apparse nei libri di biochimica, nonostante in 50 anni fossero state pubblicate centinaia di paper sull'importanza della flessibilità e del disordine nelle strutture proteiche.

La parola proteina potrebbe non essere più la parola giusta per identificare il complesso mondo delle macromolecole più importanti della vita^[31].

1.5 Il problema del Protein Folding

Il problema del protein folding è la questione di *come* una sequenza amminoacidica determini la struttura atomica tridimensionale. Il processo del ripiegamento proteico non è così semplice, la maggior parte delle proteine probabilmente passa attraverso stati strutture intermedie sulla loro via per raggiungere la struttura nativa, e il semplice osservare la struttura finale non rivela i passaggi del ripiegamento richiesti per raggiungere quella forma. Il problema del protein folding consiste di 3 puzzle strettamente correlati^[9]:

- *folding code*: la questione termodinamica di quale bilancio delle forze interatomiche determini la struttura della proteina a partire da una data sequenza amminoacidica
- *folding process*: la questione cinetica di quali percorsi alcune proteine usano per ripiegarsi così velocemente
- *protein structure prediction*: si può predire la struttura nativa di una proteina dalla sua sequenza amminoacidica? In altre parole il problema computazionale di come predire la struttura nativa di una proteina dalla sua sequenza amminoacidica

Il problema del protein folding, come si può immaginare, è considerato uno dei problemi più impegnativi degli ultimi 50 anni in biochimica. Un aspetto importante del problema è sottolineato dal paradosso di Levinthal. Nel 1968 Cyrus Levinthal si rese conto che, a causa dell'elevato numero di gradi di libertà di un polipeptide non ripiegato, tale molecola presenterebbe un numero astronomico di possibili conformazioni finali. Se la proteina raggiungesse la sua conformazione finale passando via via attraverso tutte queste configurazioni, sarebbe necessario un tempo ben superiore all'età attualmente stimata dell'universo per raggiungere la configurazione corretta. In realtà, molte piccole proteine si ripiegano spontaneamente in un tempo dell'ordine dei millisecondi o addirittura dei microsecondi. Il tempo di generazione di *E. coli* può essere di circa venti minuti: ciò significa che tutte le proteine essenziali per tale organismo (e presumibilmente di tutti gli altri) possono essere prodotte da zero in un tempo decisamente ristretto, al massimo nell'ordine dei minuti.

Il processo di ripiegamento non è quindi una ricerca all'interno dell'enorme spazio degli stati configurazionali possibili. La differenza enorme che esiste tra il tempo del ripiegamento prevedibile in teoria e quello osservato in realtà è appunto chiamato 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. Per queste ragioni è importante anche rispondere alle altre domande del problema e non fermarsi alla predizione della struttura finale, nonostante questa conoscenza fornisca un grande vantaggio per lo sviluppo di nuovi farmaci e il design di nuove proteine.

Bibliografia

Libri

- [5] S. Pal, *Fundamentals of Molecular Structural Biology*. Academic Press, 2019.
- [10] L. A. Urry, M. L. Cain, S. A. Wasserman, P. V. Minorsky, R. B. Orr e N. A. Campbell, *Campbell Biology*. Pearson, 2021.
- [11] A. Kessel e N. Ben-Tal, *Introduction to proteins: Structure, function and motion*, 2^a ed. Chapman e Hall/CRC, 2018.
- [12] B. Alberts, D. Bray, K. Hopkin et al., *Essential cell biology*, 5^a ed. W. W. Norton e Company, 2019.
- [13] L. A. Moran, H. R. Horton, K. G. Scrimgeour, M. D. Perry e D. Rawn, *Principles of biochemistry*. Pearson London, 2012.
- [17] D. L. Nelson, A. L. Lehninger e M. M. Cox, *Lehninger principles of biochemistry*. Macmillan, 2017.

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.
- [9] K. A. Dill, S. B. Ozkan, M. S. Shell e T. R. Weikl, “The protein folding problem,” *Annu. Rev. Biophys.*, vol. 37, pp. 289–316, 2008.

- [14] J. Murray, N. Laurieri e R. Delgoda, “Chapter 24 - Proteins,” S. Badal e R. Delgoda, cur., pp. 477–494, 2017. DOI: <https://doi.org/10.1016/B978-0-12-802104-0.00024-X>.
- [18] N. A. Ranson, H. E. White e H. R. Saibil, “Chaperonins,” *Biochemical Journal*, vol. 333, n. 2, pp. 233–242, 1998.
- [19] 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.
- [21] 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.
- [22] 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 metalloclothionein 3,” *ADMET and DMPK*, vol. 3, n. 3, pp. 287–295, 2015.
- [24] L. L. Porter e L. L. Looger, “Extant fold-switching proteins are widespread,” *Proceedings of the National Academy of Sciences*, vol. 115, n. 23, pp. 5968–5973, 2018.
- [25] A. E. Varela, K. A. England e S. Cavagnero, “Kinetic trapping in protein folding,” *Protein Engineering, Design and Selection*, vol. 32, n. 2, pp. 103–108, 2019.
- [26] E. Fischer, “Einfluss der Configuration auf die Wirkung der Enzyme,” *Berichte der deutschen chemischen Gesellschaft*, vol. 27, n. 3, pp. 2985–2993, 1894.
- [27] A. K. Dunker, J. D. Lawson, C. J. Brown et al., “Intrinsically disordered protein,” *Journal of molecular graphics and modelling*, vol. 19, n. 1, pp. 26–59, 2001.
- [28] A. E. Mirsky e L. Pauling, “On the structure of native, denatured, and coagulated proteins,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 22, n. 7, p. 439, 1936.
- [29] C. Bracken, M. M. Young e K. Dunker, “Disorder and flexibility in protein structure and function,” pp. 64–66, 2000.
- [30] F. Karush, “Heterogeneity of the binding sites of bovine serum albumin1,” *Journal of the American Chemical Society*, vol. 72, n. 6, pp. 2705–2713, 1950.
- [31] G. Parisi, N. Palopoli, S. C. Tosatto, M. S. Fornasari e P. Tompa, ““Protein” no longer means what it used to,” *Current Research in Structural Biology*, vol. 3, pp. 146–152, 2021.

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).
- [15] “Protein structure prediction - Wikipedia.” (30 dic. 2021), indirizzo: https://en.wikipedia.org/wiki/Protein_structure_prediction (visitato il 27/01/2022).
- [16] L. A. Moran. “Levels of Protein Structure.” (13 mar. 2008), indirizzo: <https://sandwalk.blogspot.com/2008/03/levels-of-protein-structure.html> (visitato il 27/01/2022).
- [20] “Chaperonina - Wikipedia.” (21 ott. 2021), indirizzo: <https://it.wikipedia.org/wiki/Chaperonina> (visitato il 26/01/2022).
- [23] “Lauren Porter and Fold-Switching Proteins.” (14 feb. 2020), indirizzo: <https://www.youtube.com/watch?v=IeX5ebadgiA> (visitato il 28/01/2022).

Altre fonti

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