

PREDIZIONE DELLA STRUTTURA DI PROTEINE

Un obiettivo che si prefiggono i ricercatori nel campo della struttura delle proteine è quello di predire la conformazione tridimensionale partendo dalla sola struttura primaria: da quanto è stato discusso precedentemente è chiaro che il lavoro è enorme. Abbiamo visto che in una catena di soli 100 amminoacidi devono essere variati contemporaneamente 100 ϕ e 100 ψ per trovare la conformazione a più bassa energia in uno spazio a 200 dimensioni, ma questo aspetto è solo una parte del problema. Le considerazioni sulle conformazioni della catena non tengono conto della stabilizzazione da parte del solvente. Inoltre devono essere considerate le possibili conformazioni delle catene laterali ed i loro effetti nella stabilizzazione di una particolare conformazione della catena peptidica. A causa di queste enormi complessità i ricercatori si sono rivolti a procedimenti semiempirici ed inoltre alcuni obiettivi sono stati focalizzati sulla predizione di alcuni elementi fondamentali della struttura secondaria delle proteine come le regioni α -elicoidali o a foglietti β , piuttosto che prendere in considerazione la struttura secondaria e terziaria, contemporaneamente.

Sebbene siano state seguite un certo numero di strategie diverse, tutte sono basate sull'idea che la conformazione di una proteina sia basata principalmente su interazioni locali a corto raggio. Questo significa che un polipeptide corto di sequenza definita avrà una probabilità ragionevole di adottare la stessa conformazione sia che sia isolato sia che faccia parte di una catena proteica più lunga. Alcuni esperimenti confermano questa ipotesi.

Un esempio di predizione della struttura secondaria.

Negli scorsi anni sono stati proposti molti schemi di predizioni di struttura. A titolo di esempio descriveremo lo schema proposto da Chou e Fasman nel 1974.

L'idea di base è quella di studiare un gran numero di strutture cristallografiche di proteine con lo scopo di determinare le preferenze conformazionali dei diversi residui di amminoacidi. Il concetto è che alcuni residui vengono trovati più facilmente di altri in alcune particolari conformazioni: p.es. in α -elica. Infatti, dati su peptidi sintetici indicano che alcuni residui hanno una tendenza più marcata di altri a formare strutture elicoidali. Questa analisi può essere inserita in un contesto più quantitativo calcolando i parametri conformazionali per ciascun residuo. La tabella riporta un esempio di questi calcoli ed è basata sull'analisi della struttura cristallografica di 15 proteine. La seconda colonna riporta la frequenza di trovare un certo residuo nelle strutture α -elicoidali ($f_\alpha = n_\alpha / n$, dove n è il numero totale di residui di un certo tipo, p.es. alanina, trovati nelle 15 proteine ed n_α è il numero di quei residui che si trovano nelle sezioni α -elicoidali. La terza colonna dà il parametro P_α per ciascun residuo e che è definito semplicemente come $P_\alpha = f_\alpha / \langle f_\alpha \rangle$,

dove $\langle f_\alpha \rangle$ è il valor medio di f_α su tutti e 20 i tipi di amminoacidi che possono essere presenti. Si noti che quando $P_\alpha = 1$ quel residuo adotta la conformazione α -elicoidale con la stessa frequenza di un ipotetico residuo medio, mentre se $P_\alpha > 1$ esiste una tendenza maggiore della media ad adottare conformazioni α -elicoidali; il contrario vale per $P_\alpha < 1$. Questo parametro ha qualche somiglianza col parametro s di Zimm e Bragg, relativo alla propagazione di una struttura elicoidale, che discuteremo in seguito. Quando $s = 1$, per un peptide infinitamente lungo, la percentuale di struttura elicoidale in catena è del 50%.

Nella tabella sono riportati anche i parametri f_β , P_β e f_c , P_c che si applicano per l'analisi statistica delle sequenze β e delle sequenze "coil" (non- α , non- β).

E' interessante esaminare i valori di questi parametri nelle regioni elicoidali α e β delle 15 proteine utilizzate per costruire la tabella. $\langle P_\alpha \rangle_{nN}$ definisce il parametro per un residuo nella zona N-terminale dell'elica, ma fuori dell'elica, $\langle P_\alpha \rangle_N$ quello di un residuo nella zona N-terminale dell'elica, ma in all'interno dell'elica, $\langle P_\alpha \rangle_I$ il parametro di un residuo centrale rispetto all'elica, $\langle P_\alpha \rangle_C$ il parametro di un residuo nella zona C-terminale in conformazione elicoidale e $\langle P_\alpha \rangle_{nC}$ quello di un residuo nella zona C-terminale, ma esterno all'elica.. Lo schema mostra che procedendo in entrambe le direzioni dal centro dell'elica verso la periferia c'è una diminuzione del valore di P_α e che proprio alla fine dell'elica il valore di P_α va al di sotto dell'unità . Questo fa nascere l'idea che i terminali delle eliche sono popolati in media da residui che hanno tendenza ad interrompere le eliche e che questi residui agiscono in modo da arrestare un'ulteriore tendenza alla crescita dell'elica. Al contrario le zone centrali delle eliche sono occupati da residui adiacenti che hanno una tendenza superiore alla media a formare un'elica. Questi raggruppamenti potrebbero essere i siti dei nuclei di formazione dell'elica: in seguito l'elica cresce in entrambe le direzioni. Un'analisi simile può essere fatta anche per le strutture β . La tabella riporta la classificazione dei diversi residui di amminoacidi a seconda della loro tendenza a formare o ad interrompere α -eliche e β -eliche.

Queste considerazioni suggeriscono uno schema diretto per la predizione delle strutture secondarie. L'idea è quella di analizzare una sequenza data e di assegnare i parametri conformazionali a ciascun residuo; in seguito si cercano i raggruppamenti di residui che formano eliche fiancheggiate da residui che interrompono le eliche. Facendo uso di questi concetti ci si può costruire un insieme di regole semplici per la predizione delle strutture secondarie a partire dalla sequenza in amminoacidi.