

Predizione della struttura terziaria

Metodi di predizione

La predizione della struttura tridimensionale è di gran lunga la predizione più complessa che si possa fare su una proteina.

Esistono 3 metodi principali di predizione:

1 - Homology modelling:

se si conoscono proteine simili con struttura nota

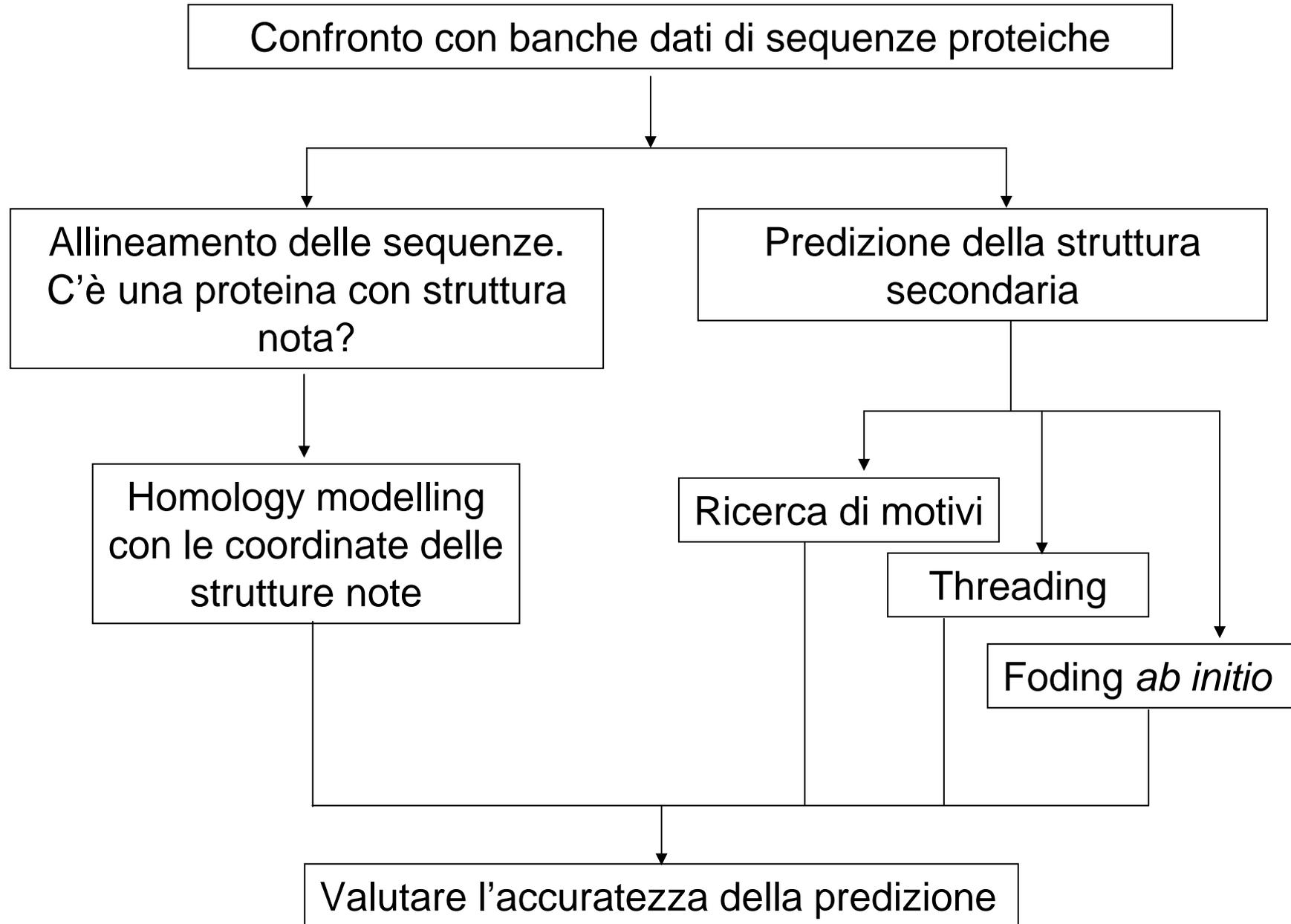
2 - Fold recognition:

se si va alla ricerca di strutture simili, cercando il folding migliore

3 - Ab initio:

simulazioni di ripiegamento in silicio, molto complesse dato che si valutano tutte le possibili interazioni di tutti gli atomi con il solvente

Come approcciarsi al problema



Homology modelling

Assunzione di base: due proteine che presentano una identità maggiore del 30% circa, molto probabilmente derivano da un antenato comune e avranno una struttura simile.

Si va a valutare la

r.m.s.d. (root mean square distance)

cioè la distanza quadratica media tra gli atomi,
generalmente i carboni α ,
ma si può calcolare anche su tutti gli atomi.

Minore è il r.m.s.d. maggiore è la similarità strutturale

=> è stato osservato che proteine che condividono un'identità di sequenza del 50% mantengono circa il 90% dei residui in posizioni conservate

Homology modelling: procedura (1)

L'approccio al modelling per omologia è abbastanza intuitivo, ma è bene seguire delle considerazioni pratiche che possono essere definite

Linee guida per costruire un modello

- 1- Analizzare la struttura secondaria prima di quella tridimensionale, dato che se ci sono regioni fortemente disordinate, è bene escluderle dalla modellazione (sono, per definizione, non modellabili).
- 2 - Identificare in banca dati una proteina a struttura nota che abbia una alta identità (> 50%) con la propria, o identità globale o identità locale.
- 3 - Allineare al meglio le due proteine. Refinire l'allineamento a mano, se necessario. Questa tappa è critica, perché la modellazione verrà fatta sulle proteine allineate.

Homology modelling: procedura (2)

- 4 - Cercare in banca dati il più alto numero di proteine simili tra loro e identificare dopo il multiallineamento le regioni di struttura secondaria conservate (in genere se ci sono dal multiallineamento si vede).
- 5 - Costruire il pre-modello, utilizzando l'allineamento multiplo come guida per la “superimposizione” della struttura nota (stampo)
- 6 - Modellare le regioni altamente variabili (in genere i loop) che connettono regioni a struttura secondaria definita. Esistono database anche dei loop.
- 7 - Aggiustare le catene laterali modellandole su quelle dello stampo. Esistono anche delle librerie di ROTAMERI, cioè isomeri posizionali delle catene laterali con angoli di torsione adeguati alle torsioni del backbone.

Homology modelling: procedura (3)

- 7 – Risolvere, se possibile, i problemi relativi alle collisioni (clashes) dei vari atomi, o manualmente o con programmi che indicano l'energia minima delle strutture.

Alla fine gli atomi di una proteina non dovranno collidere, e la proteina dovrà essere alla minor energia possibile.

Nonostante ci siano delle regole abbastanza precise, l'homology modelling è molto complesso, richiede un grande lavoro, molta esperienza e molto tempo.



E poi serve la verifica sperimentale !!!

Fold recognition (threading)

Il concetto che sta alla base di questa tecnica è:

le proteine presenti in natura hanno un numero finito di strutture possibili, o almeno un numero finito di topologie (ad oggi 1233 secondo la banca dati CATH).

David Heisenberg nel 1991 sviluppò il metodo dei **profili 1D-3D**. Si tratta di catalogare tutti i FOLD possibili in termini di INTORNO per ogni aminoacido di ogni fold.

L'intorno è definito con:

- 1 - Struttura secondaria.
- 2 - Accessibilità al solvente.
- 3 - Tipo di residui circostanti (polari, apolari, idrofobici).

Ogni fold viene descritto come una sequenza (1D) di simboli associati a frequenze di ritrovamento in una data struttura.

Fold recognition (threading)

E' possibile così confrontare una sequenza proteica allineandola con tutti i possibili profili di tutti i fold conosciuti, ricavando un punteggio che valuti il ***best fitting*** della sequenza.

Facendo così si è in grado di identificare strutture di proteine anche molto divergenti tra loro, al punto di non essere riconosciute da nessun programma di allineamento o di similarity search.

Un esempio tipico è l'individuazione della struttura di proteine che hanno la stessa funzione a causa di una **evoluzione convergente**:

originandosi da geni diversi non correlati, la sequenza sarà molto diversa, ma la struttura terziaria, almeno nell'intorno del sito catalitico, deve essere costante per garantire una stessa funzionalità.

Metodo Rosetta

Messo a punto dal gruppo di David Baker nel 1997, non si basa sulle banche dati di fold di riferimento predeterminati, ma segue tre fasi empiriche con alta capacità predittiva:

1 - Divide la sequenza primaria in gruppi (da 3 a 9) residui, ed effettua una ricerca tra le proteine a struttura nota. Si generano così, per ogni frammento, una serie di strutture 3D possibili.

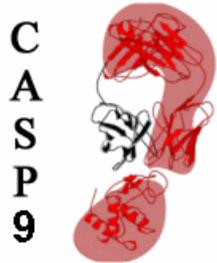
2 - Tutte le possibili combinazioni di strutture 3D locali vengono generate, e considerate inizialmente ugualmente possibili.

3 - Si applicano funzioni di scoring, di minima energia, di comparazione, per assegnare dei punteggi che indicano la qualità di ogni struttura.

Tra i sistemi di fold recognition, è quello che stabilmente ottiene i migliori risultati al CASP.

CASP

Critical Assessment of techniques for protein Structure Prediction
<http://predictioncenter.org/>



E' una "gara" tra i gruppi che sviluppano metodiche di predizione 3D e 2D che si basa sulla predizione delle strutture data una serie di sequenze proteica **e basta**.

Le predizioni migliori vengono valutate sulla base della loro somiglianza con le strutture 3D ottenute sperimentalmente ma tenute segrete fino al giorno delle presentazioni.

Vengono valutati separatamente i tre diversi modi di fare predizioni di struttura.

Il CASP permette di valutare quale è il miglior metodo per generare un modello, tenendo aggiornati tutti sui progressi dei vari gruppi di ricerca e garantendo così agli utenti finali di utilizzare la tecnica predittiva più aggiornata e più affidabile.

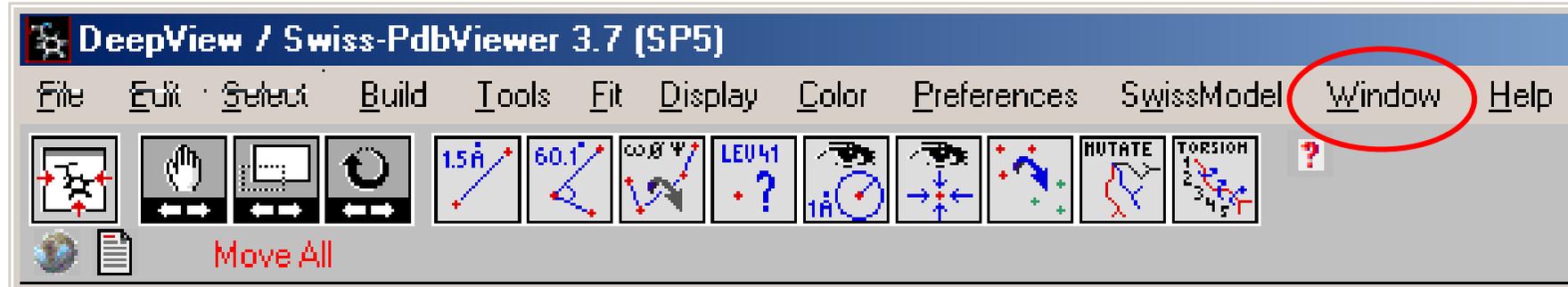
Inoltre, essendo una competizione a premi, stimola la ricerca e la continua innovazione alla ricerca del metodo predittivo migliore.



Menu

- [Home](#)
- [FORCASP Forum](#)
- [PC Login](#)
- [PC Registration](#)
- ▼ [CASP Experiments](#)
 - ▶ [CASP9 \(2010\)](#)
 - [CASP8 \(2008\)](#)
 - [CASP7 \(2006\)](#)
 - [CASP6 \(2004\)](#)
 - [CASP5 \(2002\)](#)
 - [CASP4 \(2000\)](#)
 - [CASP3 \(1998\)](#)
 - [CASP2 \(1996\)](#)
 - [CASP1 \(1994\)](#)
- ▶ [Initiatives](#)
- ▶ [Data Archive](#)
- [Local Services](#)
- [Proceedings](#)
- [Feedback](#)
- [Assessors](#)
- [People](#)
- [Community Resources](#)

Swiss-PDB Viewer



E' molto più complesso e con una interfaccia meno immediata di RasMol, ma ha delle potenzialità enormi da un punto di vista di analisi, modellazione e rendering 3D delle proteine.

Vengono implementati svariati metodi per:

- 1 - Homology modeling
- 2 - Structure alignment
- 3 - Manual refining delle strutture
- 4 - Calcolo delle energie minime
- 5 - Introduzione delle mutazioni strutturali