

Modellamento di proteine per omologia: teoria ed esempi

Anna Marabotti



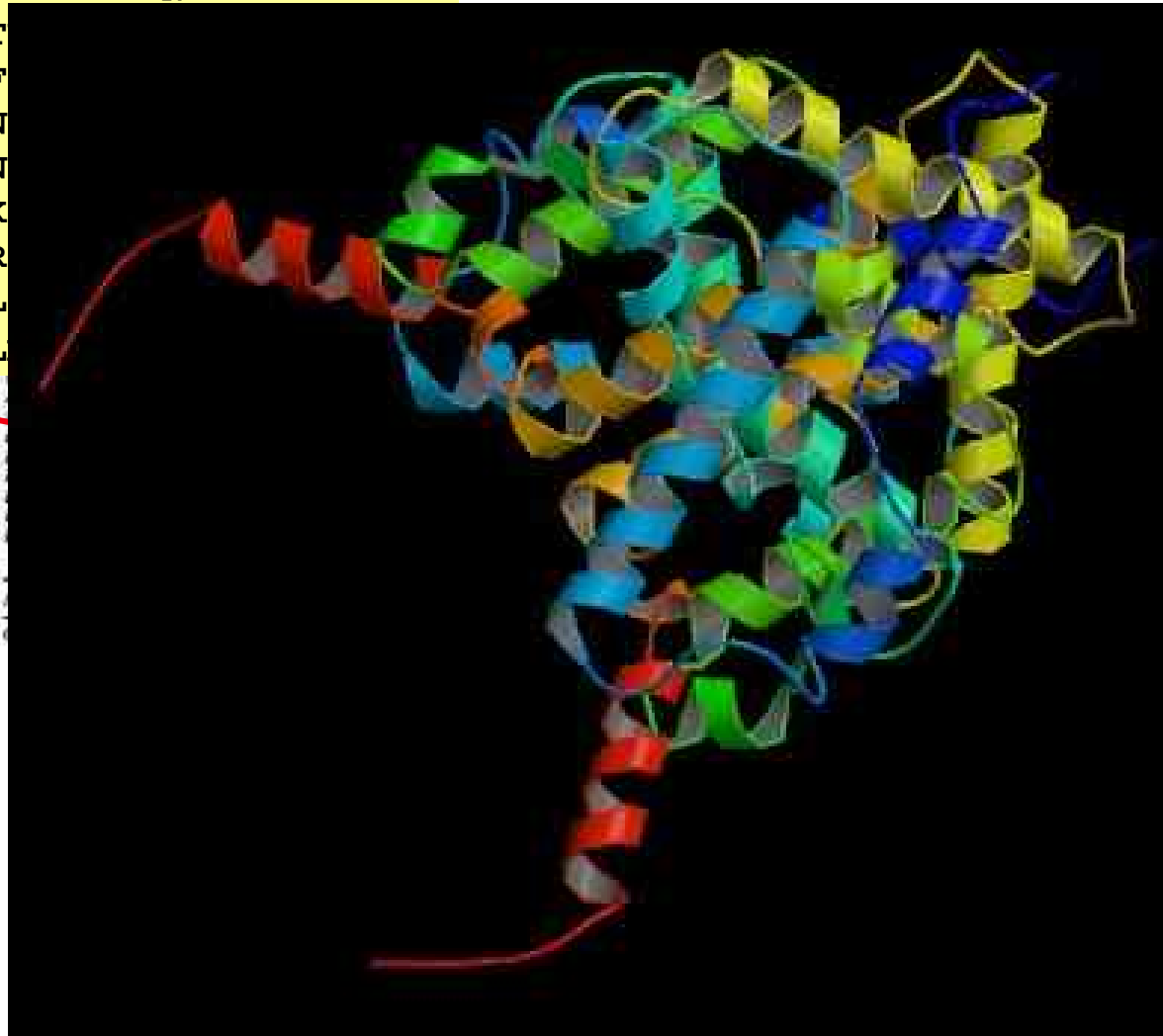
CRISCEB - Seconda Università di Napoli



Istituto di Scienze dell'Alimentazione - CNR – Avellino

Dal gene alla funzione...

MGSSHHHHHSSGLVPRGSHKREAVQEER
QRGKDRNENEVESTSSANEDMPVERILEA
ELAVEPKTET
ICQAADKQLF
QVILLRAGWN
LATGLHVHRN
SKMRDMQMDK
SNPAEVEALR
GRFAKLLLR
IGDTPIDTFL



Journal of Molecular Crystalline and Liquid Crystals, Vol. 710, 775

5'-DNA template-3'

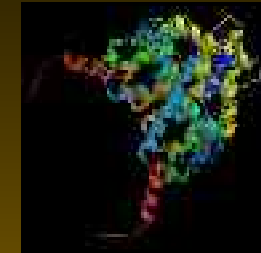
ggccgc ggccggcat
ttctc caccagggtg
cccctc gctgcaccgc
catcag caccctgagc
cccctc gggcccccac
cagccc ccagctcagc
cctggg cctcaatggc
caccaa gcacatctgc
cagctg cgaggggtgc
ctgccc cgacaacaag
ccgcta ccagaagtgc
gcgtgg caaggaccgg
ggccgt ggagaggatc
ggaggc aaacatgggg
agcagc cgacaacaag
agagct gcccttggac
cgccctc cttctccac
gcacgt ccaccggaac
gacgga gcttgtgtcc
ggcgc catcgtcctc
ggcgct gagggagaag
gcagcc gggaaaggtc
caaatg cctggaacat
ccttat ggagatgctg
tgccca ccggttctgg
tccctg cccttctctg
ctaaga gatgtgttgt
tggctt tctgagcag
cccacc gggctctcag
cggccc cagccctgga
tcgatg ctggttttca

gaattc

Importanza della struttura delle proteine

Dalla struttura si può ipotizzare la funzione

```
MGSSHHHHHSSGLVPRGSHKREAVQEERQRGKDRNENEVES  
TSSANEDMPVERILEAELAVEPKTETETYVEANMGLNPSSPNDP  
VTNICQAADKQLFTLVEWAKRIPHFSSELPLDDQVILLRAGWN  
ELLIASFSHRSIAVKDGIILLATGLHVHRNSAHSAGVGAIFDR  
VLTELVSKMRDMQMDKTELGCLRAIVLFPDPSKGLSNPAEVE  
ALREKVYASLEAYCKHKYPEQPGRFAKLLLRPALRSIGLKC  
LEHLFFFKLIQDTPIDTFLMEMLEAPHQMT
```

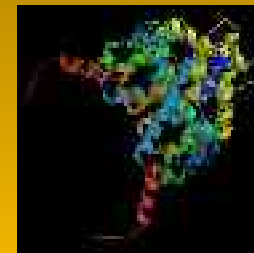
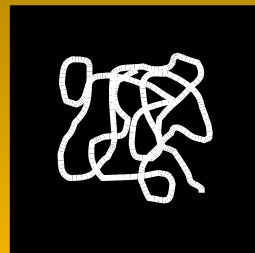


Progettazione di modifiche, studio di effetti su struttura e funzione

```
MGSSHHHHHSSGLVPRGSHKREAVQEERQRGKDRNENEVES  
TSSANEDMPVERILEAELAVEPKTETETYVEANMGLNPSSPNDP  
VTNICQAADKQLFTLVEWAKRIPHFSEWPLDDQVILLRAGWN  
ELLIASFSHRSIAVKDGIILLATGLHVHRNSAHSAGVGAIFDR  
VLTELVSKMRDMQMDKTELGCLRAIVLFPDPSKGLSNPAEVE  
ALREKVYASLEAYCKHKYPEQPGRFAKLLLRPALRSIGLKC  
LEHLFFFKLIQDTPIDTFLMEMLEAPHQMT
```



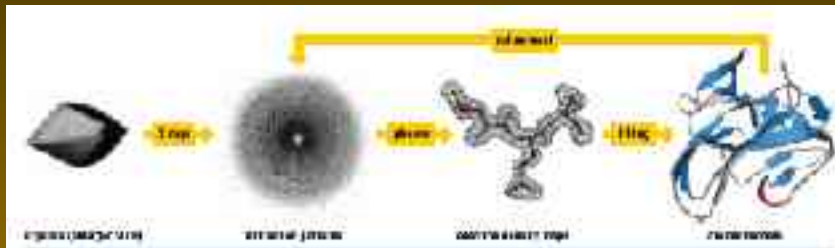
Comprensione dei meccanismi di "folding"



Perché predire la struttura delle proteine?

Limiti dei metodi sperimentali:

Raggi X:



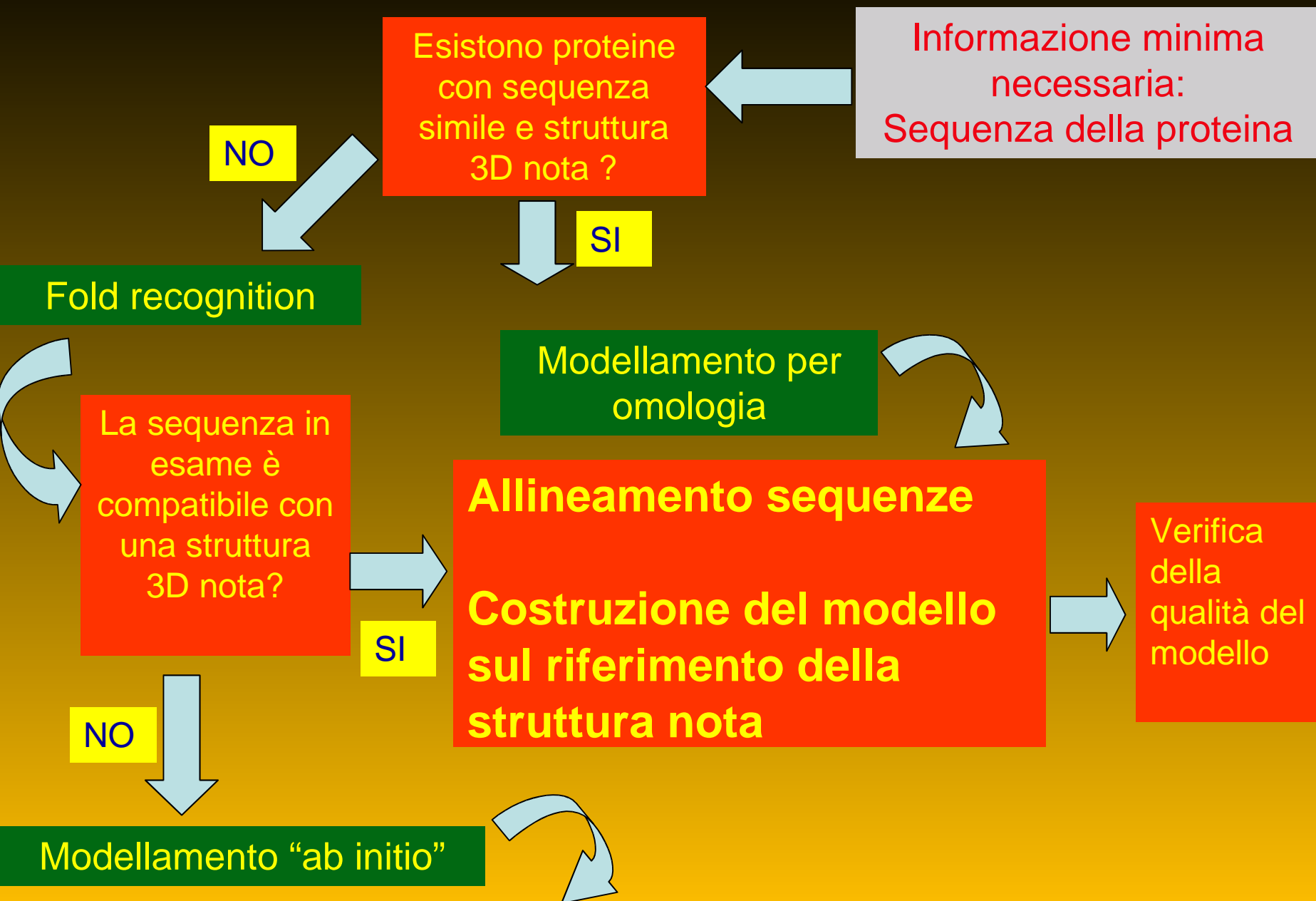
NMR:



- Difficile ottenere cristalli idonei
- Non adatti a proteine di membrana (non cristallizzano bene)
- Necessità di attrezzature particolari (sincrotrone)
- Lunghi tempi (anni)

- Necessità di soluzioni molto concentrate (rischio di aggregazione e precipitazione)
- Necessità di attrezzature particolari (risonanza magnetica)
- Non adatti a proteine molto grandi (limite massimo: ca 20-30 kDa)

Come predire la struttura ?



Modellamento per omologia

Quali le condizioni ideali?

Template	LSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASEDL LS+GEWQ VL+VW KVEAD+AGHGQ++LIRLF HPETLEKFD+FKHLKTEAEMKASEDL
Model	LSDGEWQQVLNVWVGKVEADIAGHGQEV LIRLFTGHPETLEKFDKFKHLKTEAEMKASEDL

Template	KKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPG KKHG VLTALG ILLKKKGHHEAELKPLAQSHATKHKIPIKYLEFIS+AIIHVLHS+HPG
Model	KKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPG

Template	AFVPTPAERDDLASIMKLLTKYDNLFETSFPYSMGWGAPTGSEAGANWNHWQLHAHY T A+R DLA +KKL ++YDNLF+ SFPYSMGWGAP E HWQLHAH+Y
Model	ITDLTDAQRSDLALAKKLTSTRYDNLFQCSFPYSMGWGAPFNGEEN - - - QHWQLHAHFY

Template	PPLLRSATVRKFMVGYEMLAQAQRDLTPEQAAERLRALPEVHY PPLLRSATVRKFMVGYEMLA+ QRDLT EQAAERLRA+ ++H+
Model	PPLLRSATVRKFMVGYEMLAETQRDLTAEQAAERLRAVSDIHF

Template	NVIWYNKQIFKKYGIDPSSIKTIDDLFAVAEKLKENGVTFFALGDRNKWPATQVFE - VML N +W N + KK G+ S T+D+ F A+K+K G P A G + W VFE V L
Model	NWLWANPDVLKKGSGV - - SLPTTLDEFDAADKIKAAGFIPLAHGGQ - PWQDATVFEAVAL

Template	VAVGGVDYIEKFINGEISANDPTLKKVLEYFVKYV - - - EYSNPDHAAKTWDEACALVYKG +G DY + F+ ++ N + K++E F K+ +Y + + + W+ A ++V G
Model	DVLGSEDYNAKAFV - - DLDMNVLSGDKMVEVFTKFKKMRDYIDSNAQGRDWNVATSMVING

Template	LSSGEDKWLRLRKAQEVLLNSPLQEEHN - - FPPDHY - - - - - GLPAVCDLFA + E K + E L N P HN P+ Y G P V D FA
Model	VQEAIEKPFV - - - - - ESLSNCPKHGLHNPLKAPERYLMYGDISAKAANMLGQPVVSDFFA

Template	CMQLKVDKEKW + +V + +
Model	KLIQRVSPKSY

>75%

Modelling backbone, catene laterali, nessun loop da modellare

40-75%

Modelling backbone, inserzione loop

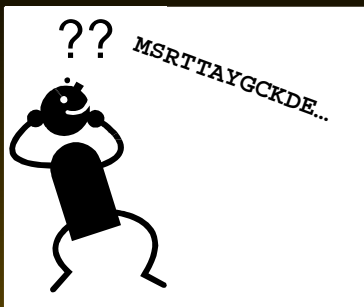
20-40%

Difficile modelling del backbone, larghe porzioni strutturali senza riferimento.

<20%

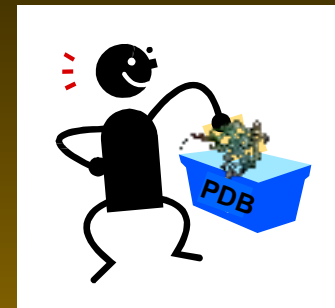
Modelling per omologia non applicabile

Modellamento per omologia



Ricerca in banche dati strutturali di proteine simili alla nostra e utili come template.

Analisi della sequenza



Allineamento di sequenza tra proteina da modellare e proteine somiglianti (compresi i template strutturali)

Modellamento della struttura della proteina



Analisi della sequenza di amminoacidi

```
MGSSHHHHHHS SGLVPRGSHKREAVQEERQRGKDRNENEVESTSSANEDMPVERILEAELAVEPKTETY  
VEANMGLNPSSPNDPVTNICQAADKQLFTLVEWAKRIPHFSELPLDDQVILLRAGWNELLIASFHSRSI  
AVKDGILLATGLHVHRNSAHSAGVGAI FDRVLTEL VSKMRDMQMDKTELGCLRAIVLFNPDSKGLSNPA  
EVEALREKVYASLEAYCKHKYPEQPGRFAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLE  
APHQMT
```

Analisi della struttura primaria e confronto con le sequenze presenti nelle banche dati

BANCHE DATI:

- Sequenze di nucleotidi (GenBank, EMBL)
- Sequenze di amminoacidi (UniProt, PIR, GenPept)
- Strutture terziarie e quaternarie di proteine (PDB)

-

BLAST (Basic Local Alignment Search Tool)

Blastp: Paragona una sequenza amminoacidica a un database di sequenze di proteine

Adatto sia per la ricerca per similarità di sequenze amminoacidiche nelle banche dati, sia per allineare due sequenze simili tra loro.

BLAST OUTPUT			Score	E
Sequences producing significant alignments:			(bits)	Value
sp P28814 BARW_HORVU	BARWIN		265	3e-71
sp P43082 HEVL_ARATH	HEVEIN-LIKE PROTEIN PRECURSOR		196	2e-50
sp P29062 PR4A_TOBAC	PATHOGENESIS-RELATED PROTEIN PR-4A PRECURSOR		194	8e-50
sp P02877 HEVE_HEVBR	HEVEIN PRECURSOR (MAJOR HEVEIN)		187	1e-47
sp P09761 WIN1_SOLTU	WOUND-INDUCED PROTEIN WIN1 PRECURSOR		183	8e-47
sp Q02243 WIN_SOYBN	WOUND-INDUCED PROTEIN		137	1e-32
sp Q07447 RECA_PSEPU	RECA PROTEIN		29	3.9
sp P13944 CA1C_CHICK	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR (FIB...		28	8.8

Bit Score ed E-Value sono parametri usati per la valutazione della significatività dell'allineamento:

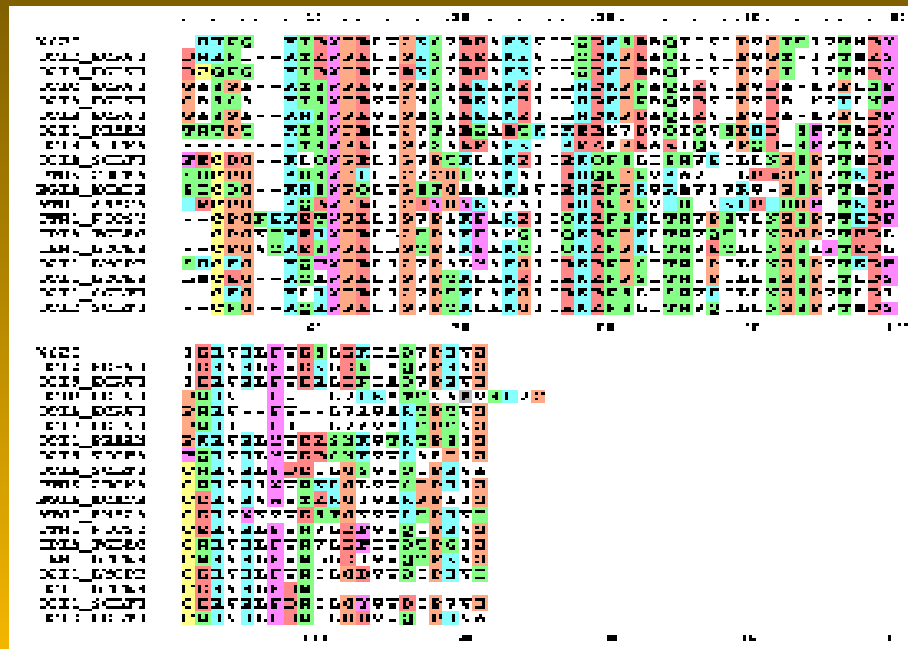
- lo Score indica quanta somiglianza c'è tra i due segmenti di sequenza allineate;
- l'E-value, in termini statistici, ci dice quanti altri segmenti della banca dati ci si aspetta che abbiano quello stesso Score.

Allineamento di sequenze

E' IL PASSAGGIO CRUCIALE E IL PIU' DELICATO NELLA COSTRUZIONE DEL MODELLO PER OMOLOGIA

L'allineamento multiplo consente di avere informazioni sulla conservazione e variabilità di ogni posizione all'interno di una famiglia proteica e quindi aiuta a definire meglio la posizione di inserzioni e delezioni.

Programmi usati: ClustalW, T-Coffee, MUSCLE, PROBCONS etc.



Allineare le sequenze non basta...

IL MIGLIORE ALLINEAMENTO PER IL MODELLAMENTO PER OMOLOGIA NON È QUELLO CHE ALLINEA SOLO GLI AMMINOACIDI SIMILI TRA LORO!!

Occorre tenere conto di fattori strutturali, informazioni funzionali e in generale di tutte le informazioni disponibili raccolte sulla proteina.

Alcune informazioni utili da considerare sono:

- Pattern di sequenza: porzione di sequenza in cui sono presenti determinati amminoacidi posti a determinate distanze tra loro. Può caratterizzare una particolare funzione della proteina.

...W-I/L-R-[Hydro]- [Polar]-G-[Any]...

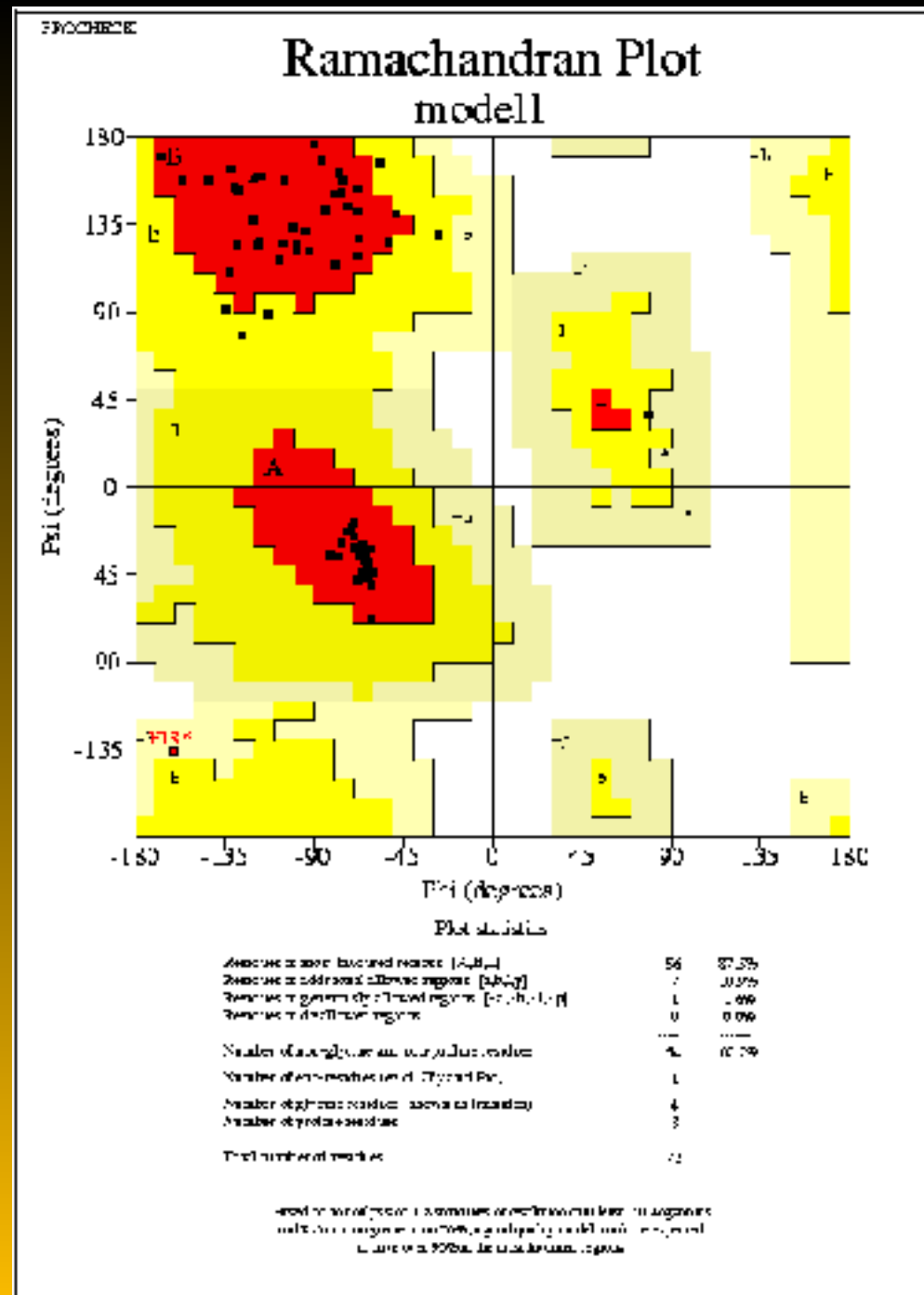
- Amminoacidi coinvolti nell'interazione con ligandi o all'interno di siti catalitici

Es. triade catalitica Ser-proteasi (Ser, Asp, His)

- Informazioni sulle strutture secondarie delle proteine.

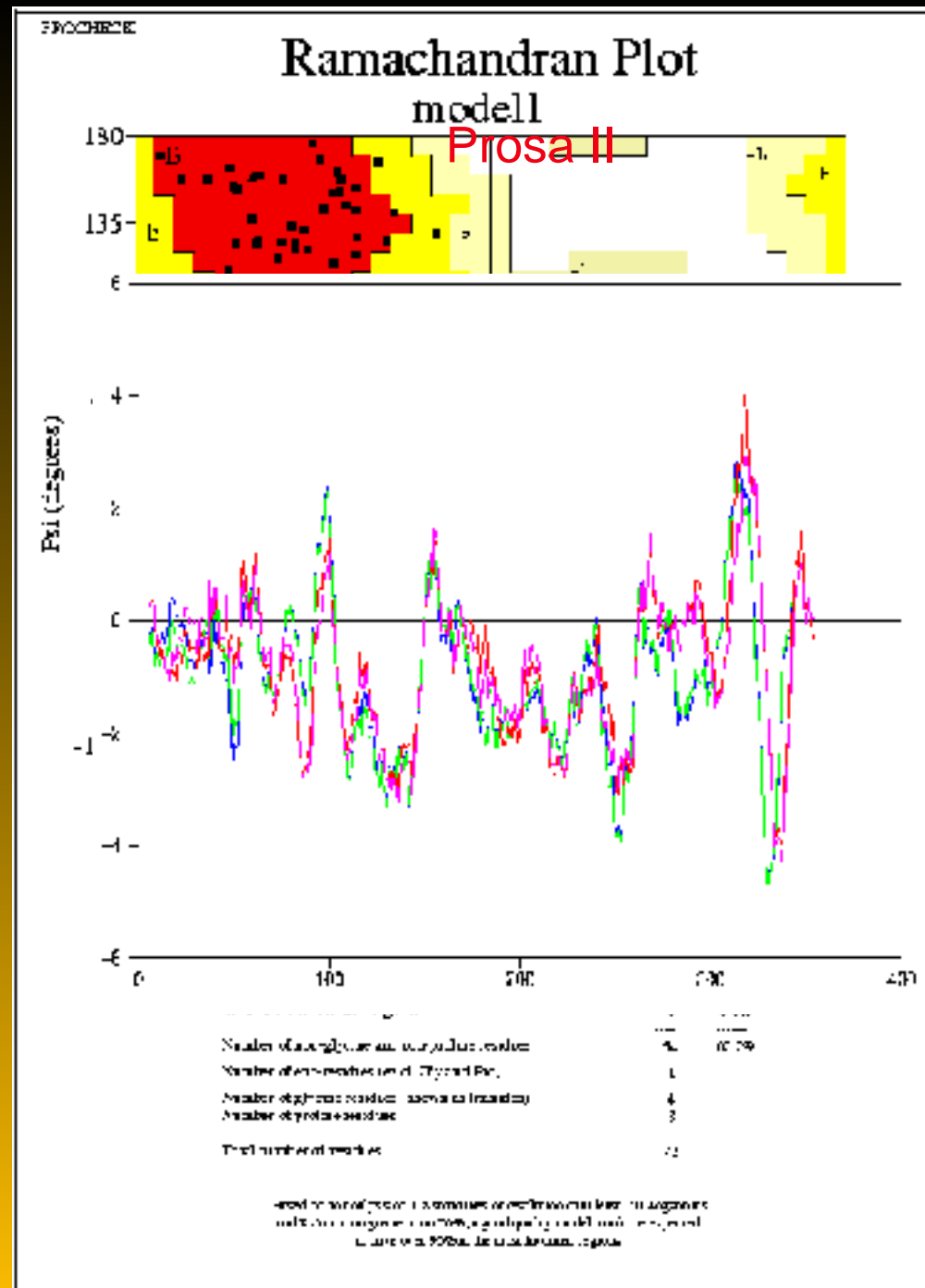
Valutazione del modello

- Caratteristiche stereo-chimiche del backbone



Valutazione del modello

- Caratteristiche stereo-chimiche del backbone
- Compatibilità con proprietà strutturali es. del template



Valutazione del modello

- Caratteristiche stereo-chimiche del backbone
- Compatibilità con proprietà strutturali es. del templato
- Altri confronti con proprietà determinate sperimentalmente (es. CD, funzione etc)



•Selezione del miglior modello

