

Struttura di proteine

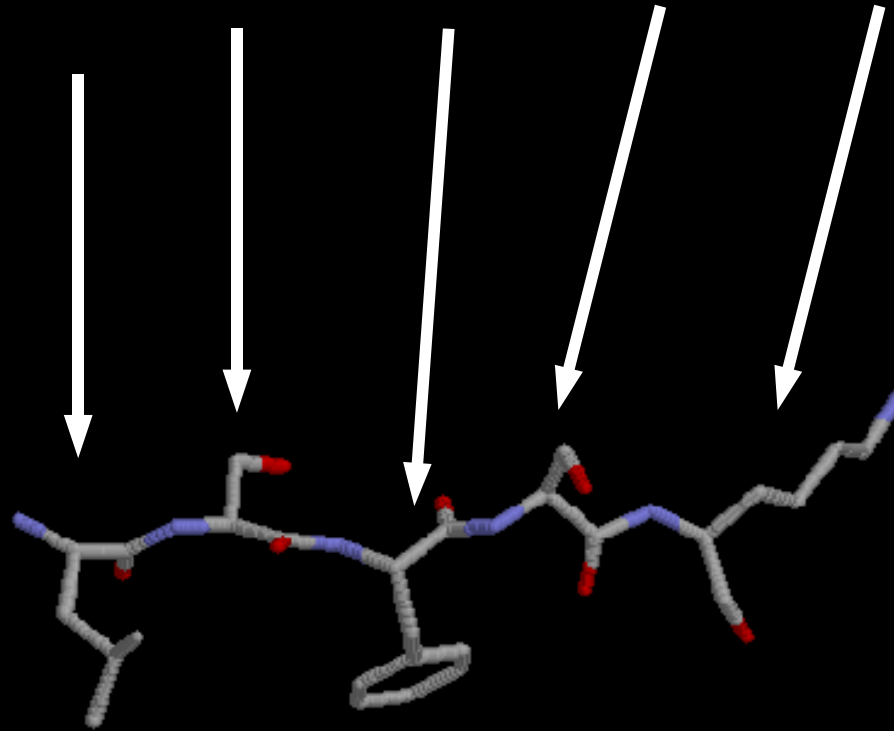
- Introduzione alla struttura delle proteine. Elementi di struttura secondaria, motivi e classificazione
 - Tecniche di determinazione strutturale. Cristallografia a raggi X e Risonanza Magnetica Nucleare
 - Predizione di struttura secondaria. Predizione di struttura terziaria. Modellistica per omologia. Riconoscimento di fold e modellistica ab-initio.
-
-

Proteine

- Le proteine sono molecole, fondamentali per tutte le funzioni cellulari
 - Le proteine sono eteropolimeri lineari costituiti da 20 tipi di ammino acidi. La linearità permette di rappresentare la struttura chimica con una stringa di caratteri (**sequence, primary structure**)
 - Gli ammino acidi sono costituiti da una parte comune (che forma la catena principale delle proteine, **backbone or main chain**) e una parte differente (**side chain** o catena laterale)
-
-

Primary structure, sequence

.... ILE SER PHE SER LYS

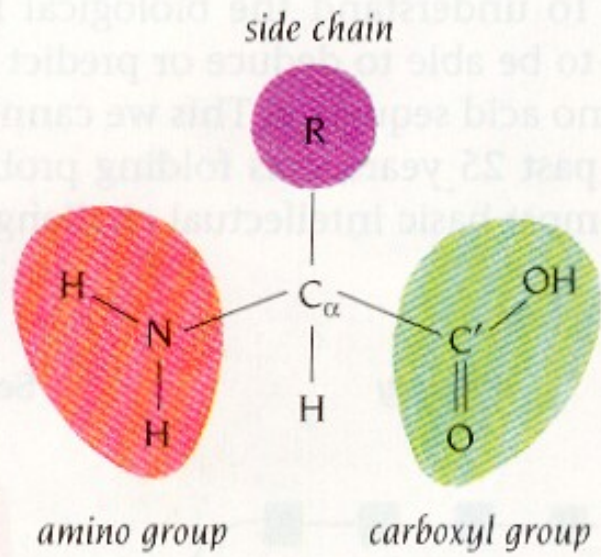


.....ISFSK.....

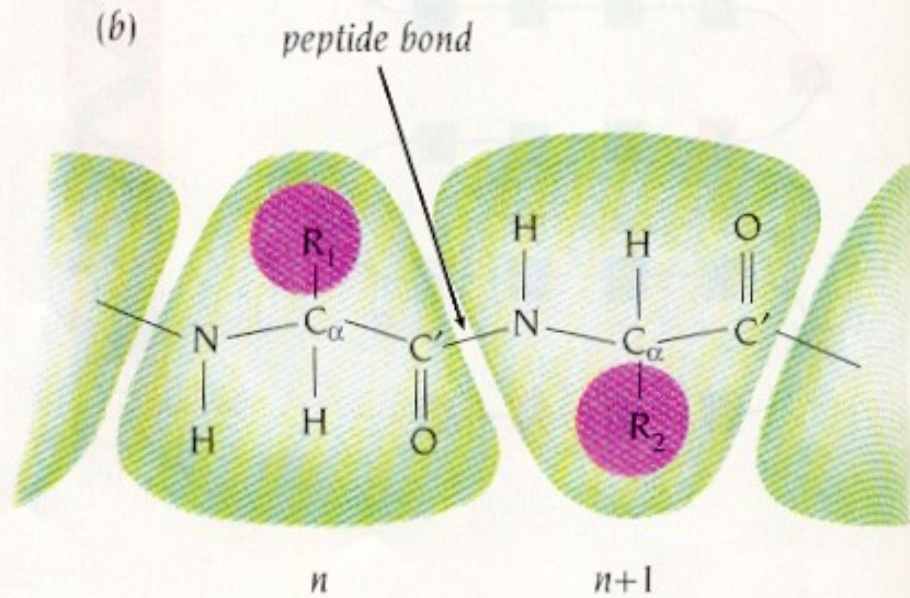
Il legame peptidico

- Due amminoacidi in una proteina sono connessi dal legame peptidico
- I terminali (N- e C-terminale) non sono coinvolti nel legame peptidico e portano i gruppi carichi NH_3^+ e COO^- rispettivamente.

(a)

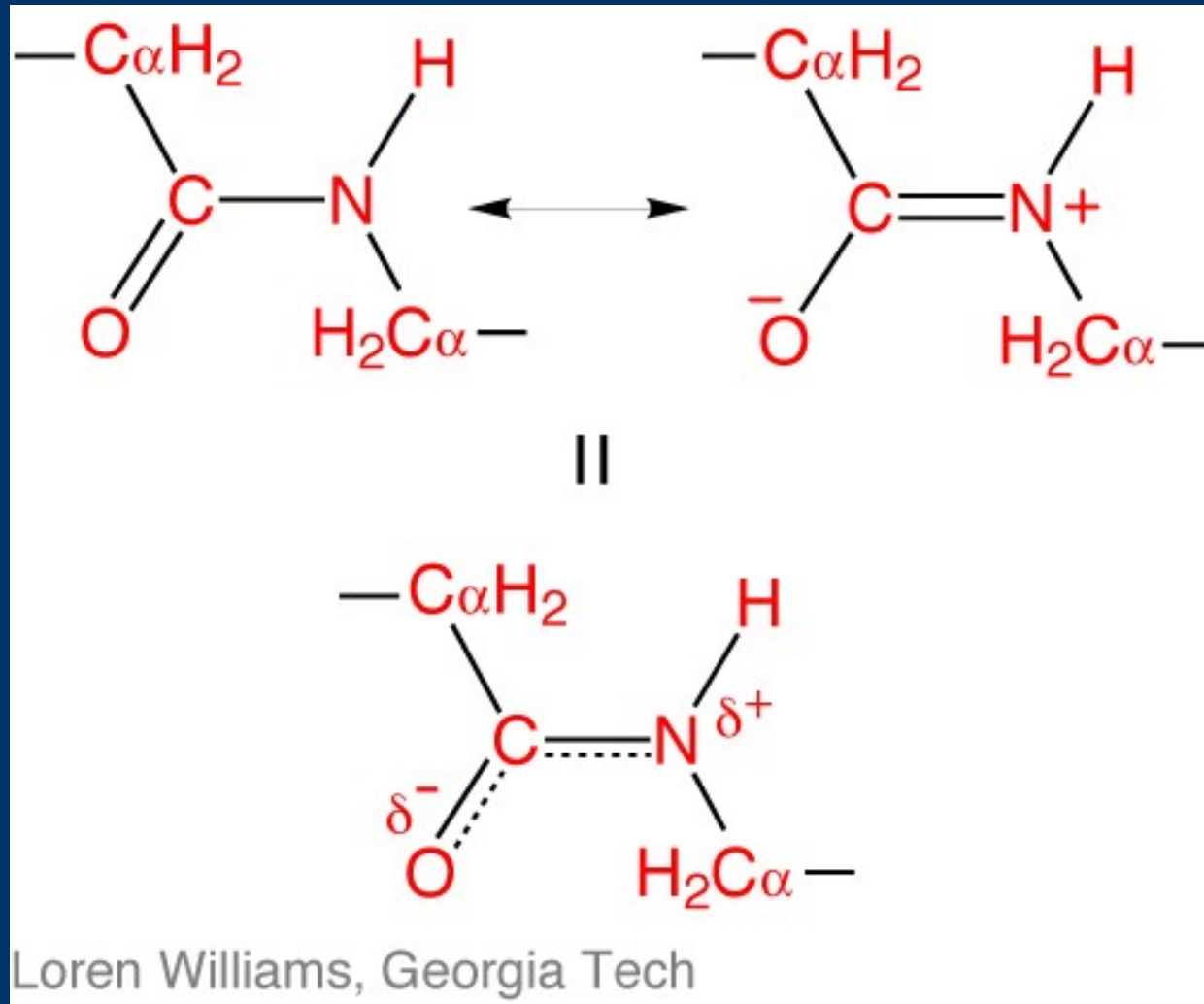


(b)



Il legame peptidico

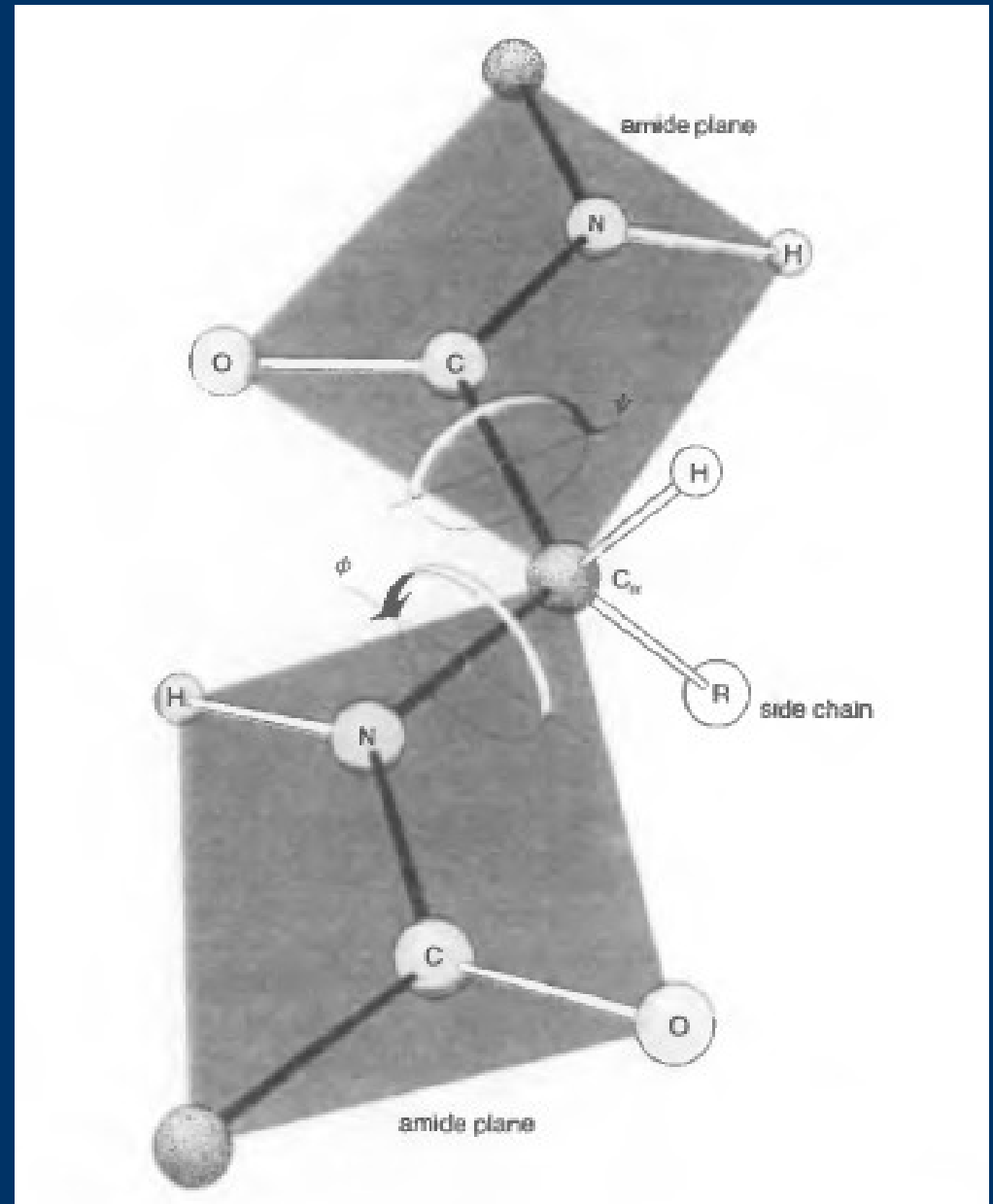
- Gli atomi coinvolti nel legame peptidico e i loro sostituenti formano un piano rigido



Conformazione del backbone

Gli atomi $C\alpha$ -C-N- $C\alpha$ definiscono quindi un piano rigido.

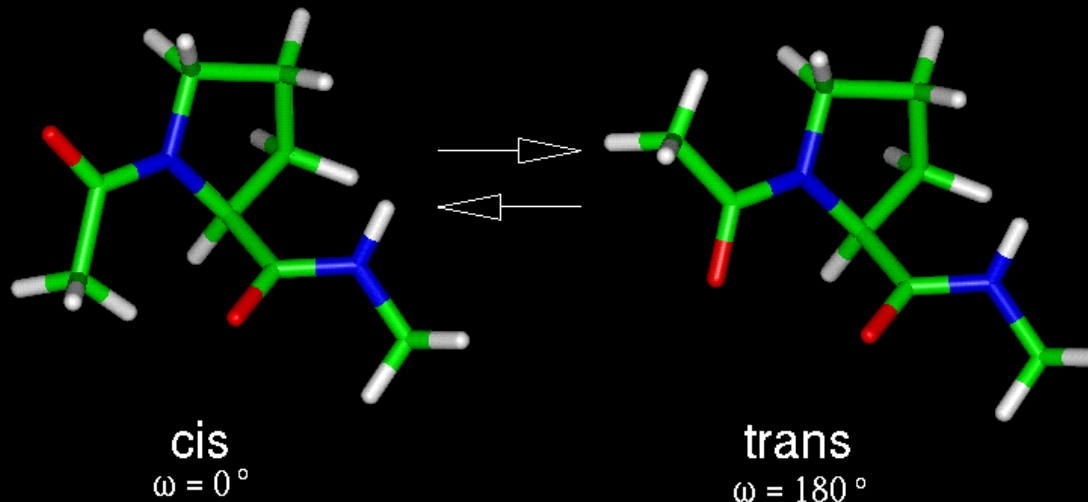
Due configurazioni sono possibili. Quella che si osserva più frequentemente è quella in cui i due $C\alpha$ sono più distanti (trans)



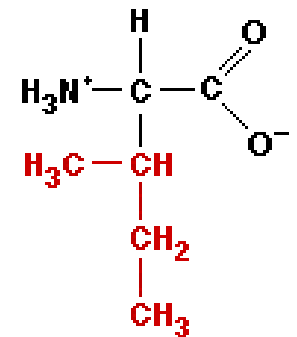
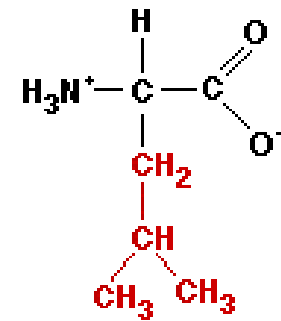
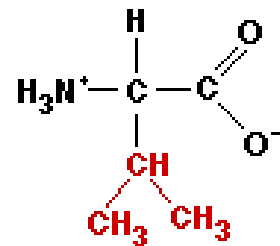
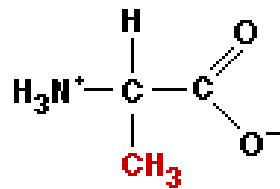
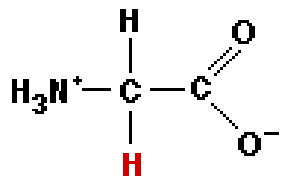
Il legame peptidico

- Per gli ammino acidi diversi dalla Prolina la frequenza relativa della conformazione cis e' ca. dello 0.0004(3)
- Per la prolina questa frequenza e' molto maggiore ca. 0.05(1) perche' il sostituente al C α non e' l'H α ma il C δ .

cis/trans isomerization of a peptidyl–prolyl bond
(Acetyl–prolyl–N–methamid)



Le catene laterali apolari (idrofobiche) si trovano preferenzialmente all'interno delle proteine



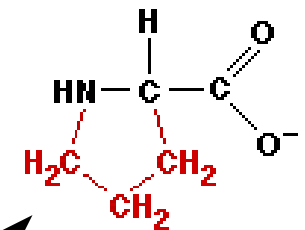
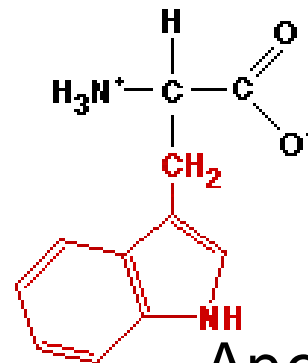
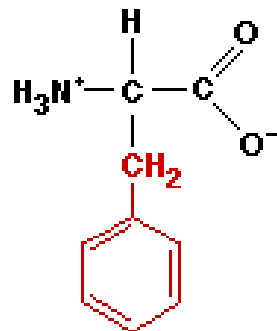
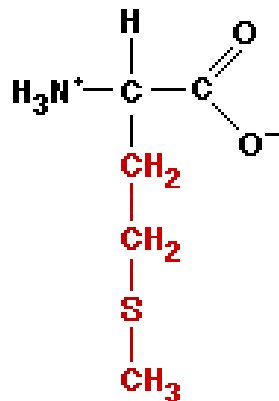
Glycine (Gly)

Alanine (Ala)

Valine (Val)

Leucine (Leu)

Isoleucine (Ile)



Anello chiuso
sul backbone

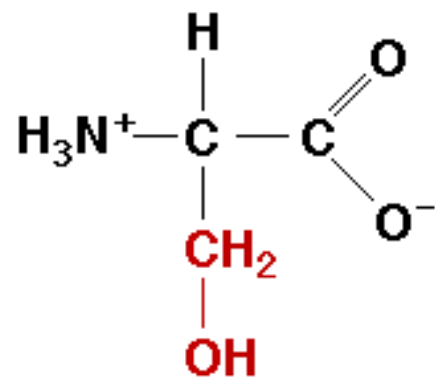
Methionine (Met)

Phenylalanine (Phe)

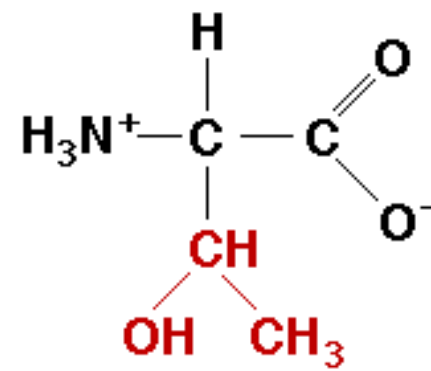
Tryptophan (Trp)

Proline (Pro)

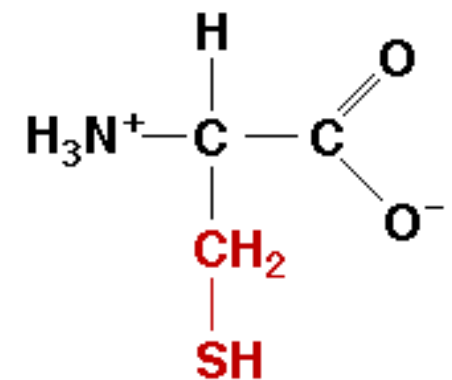
Le catene laterali polari (idrofobiche) possono formare H-bond all'interno e alla superficie di una proteina o essere esposte al solvente



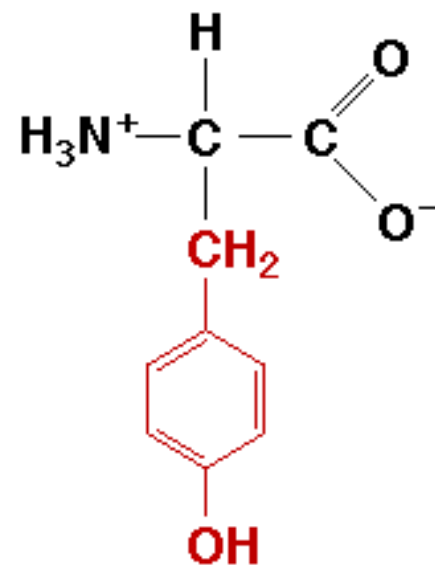
Serine (Ser)



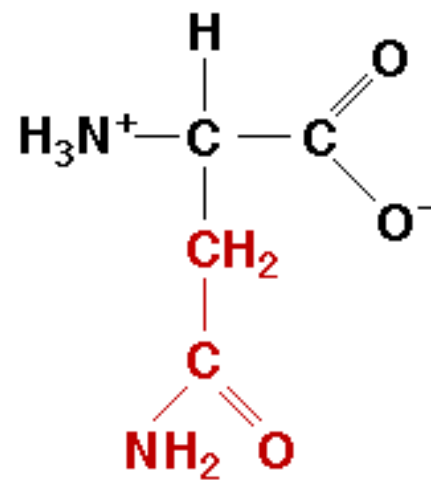
Threonine (Thr)



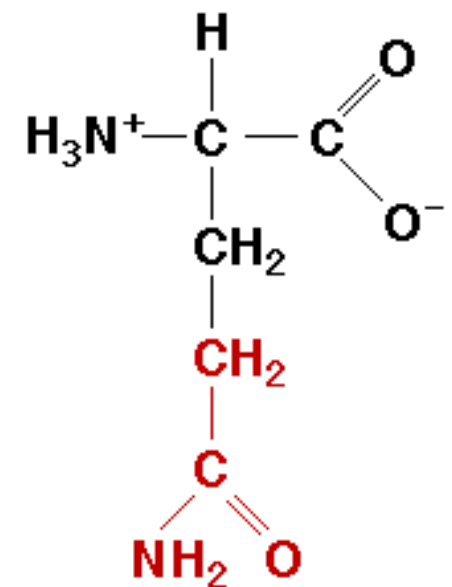
Cysteine (Cys)



Tyrosine (Tyr)

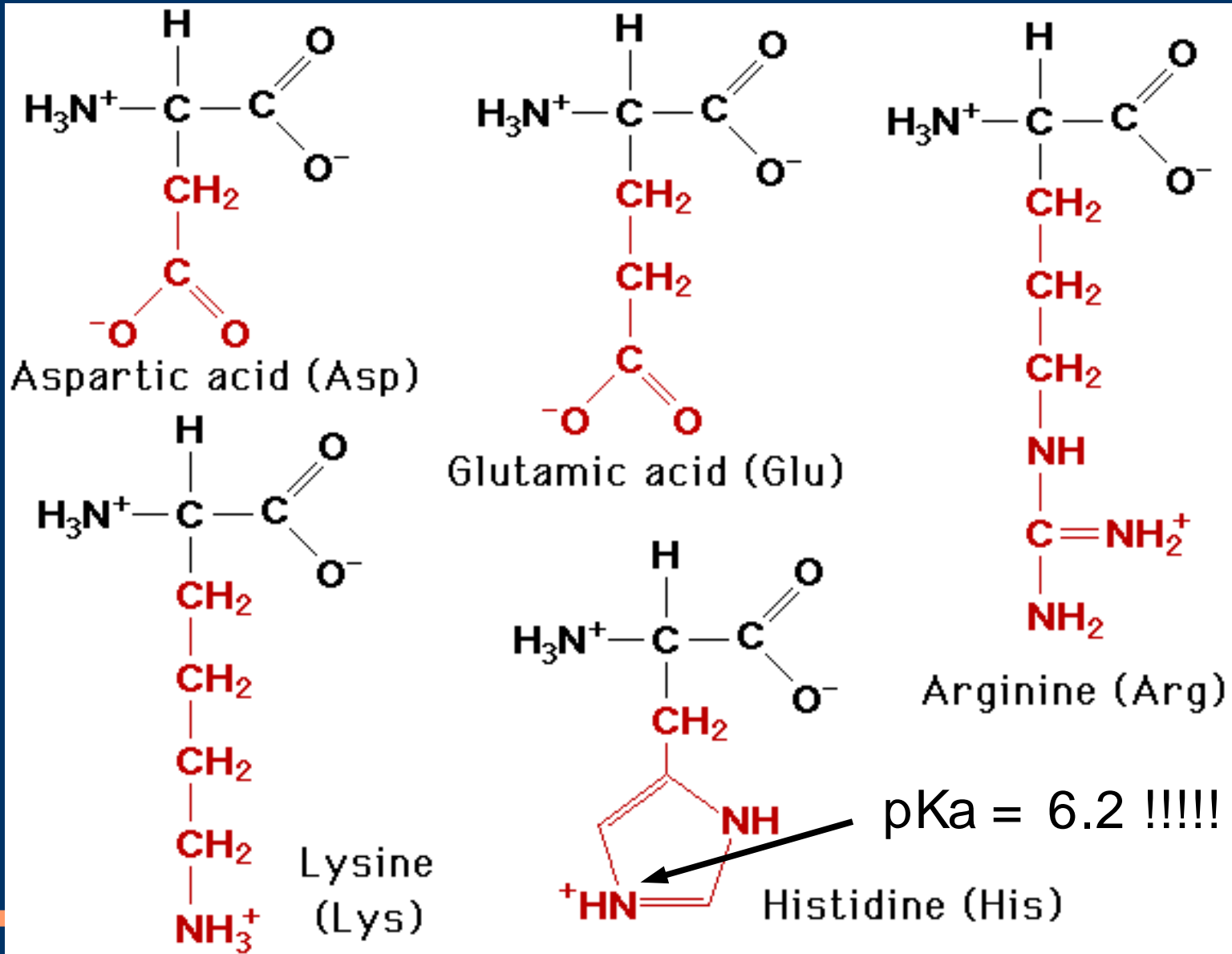


Asparagine (Asn)



Glutamine (Gln)

Le catene laterali cariche si trovano esposte al solvente. Possono formare ponti salini all'interno o alla superficie di una proteina dove in genere formano interazioni a tre residui (ad GLU-ARG-GLU)

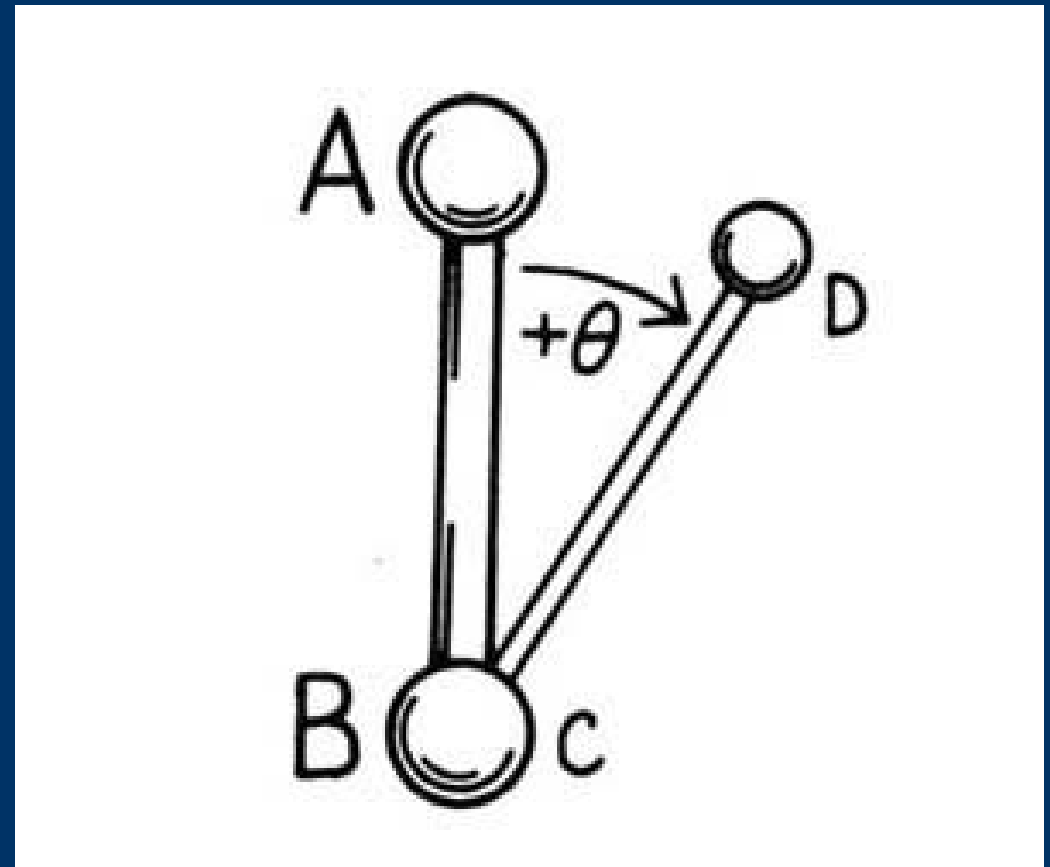


Angoli torsionali

Le lunghezze dei legami sono pressoché costanti, come pure gli angoli di legame ...

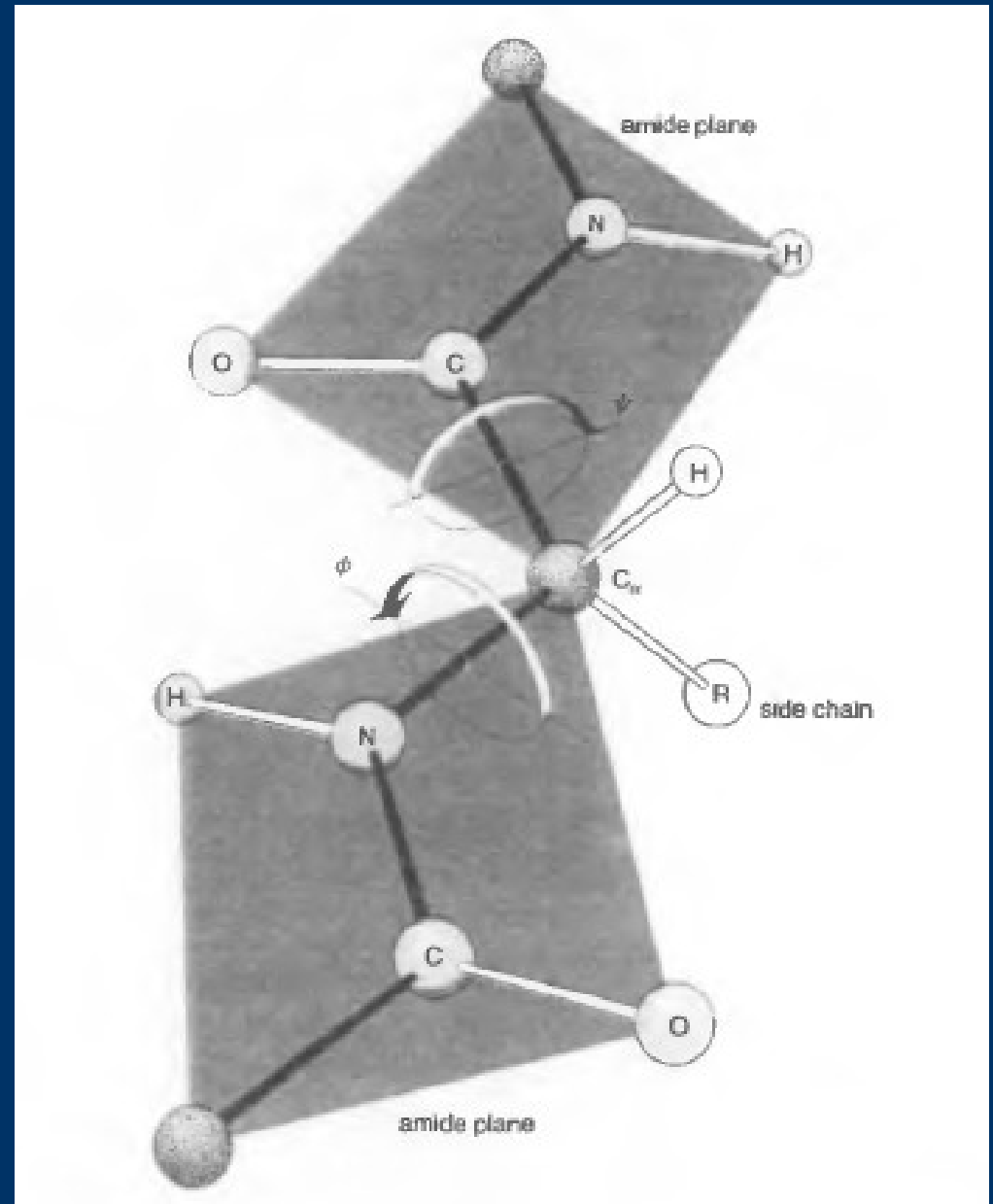
La conformazione del backbone di una proteina è quindi determinata dalle rotazioni attorno ai due legami che coinvolgono l'atomo C α .

Un angolo torsionale è definito come in figura

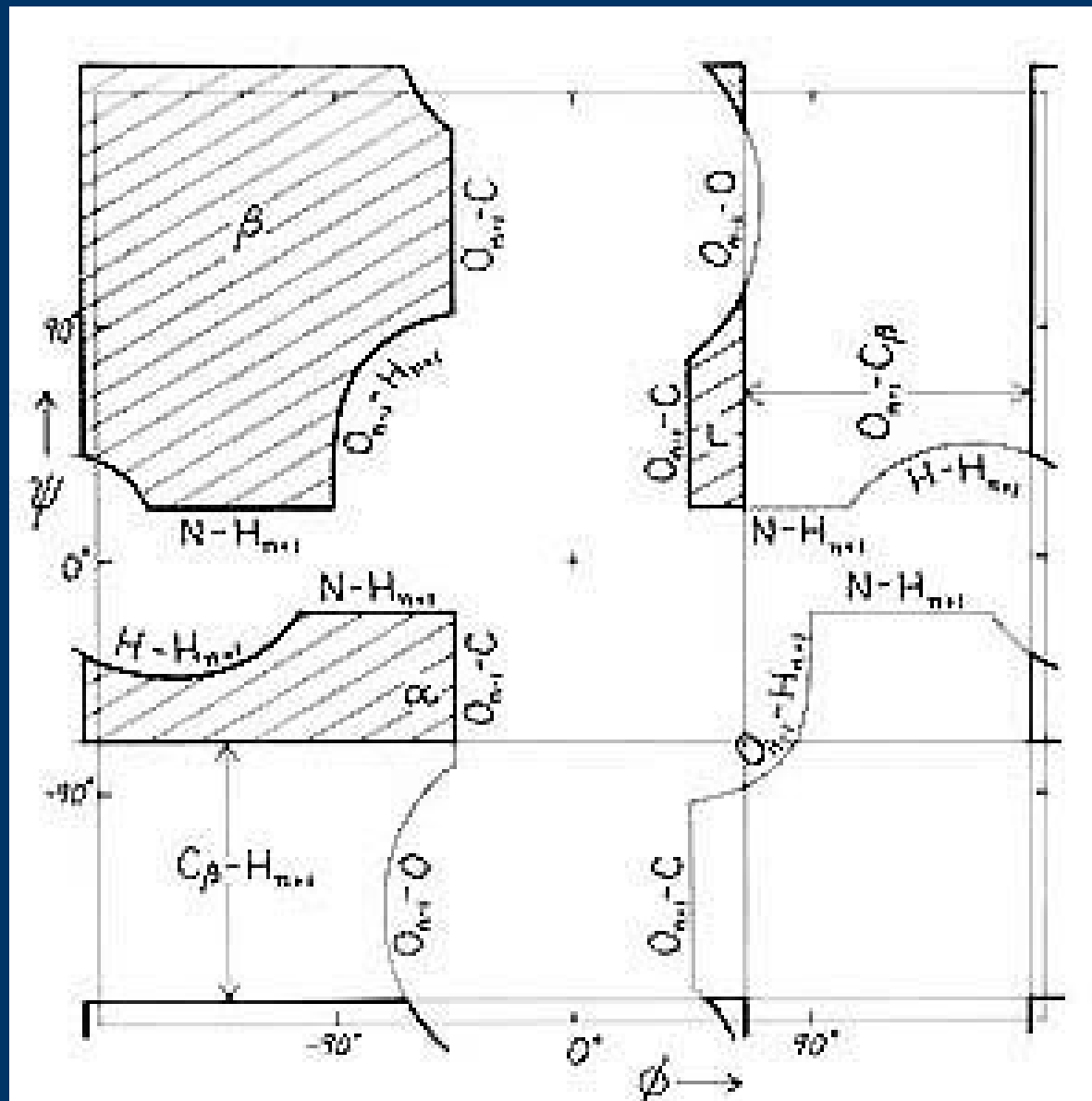


Conformazione del backbone

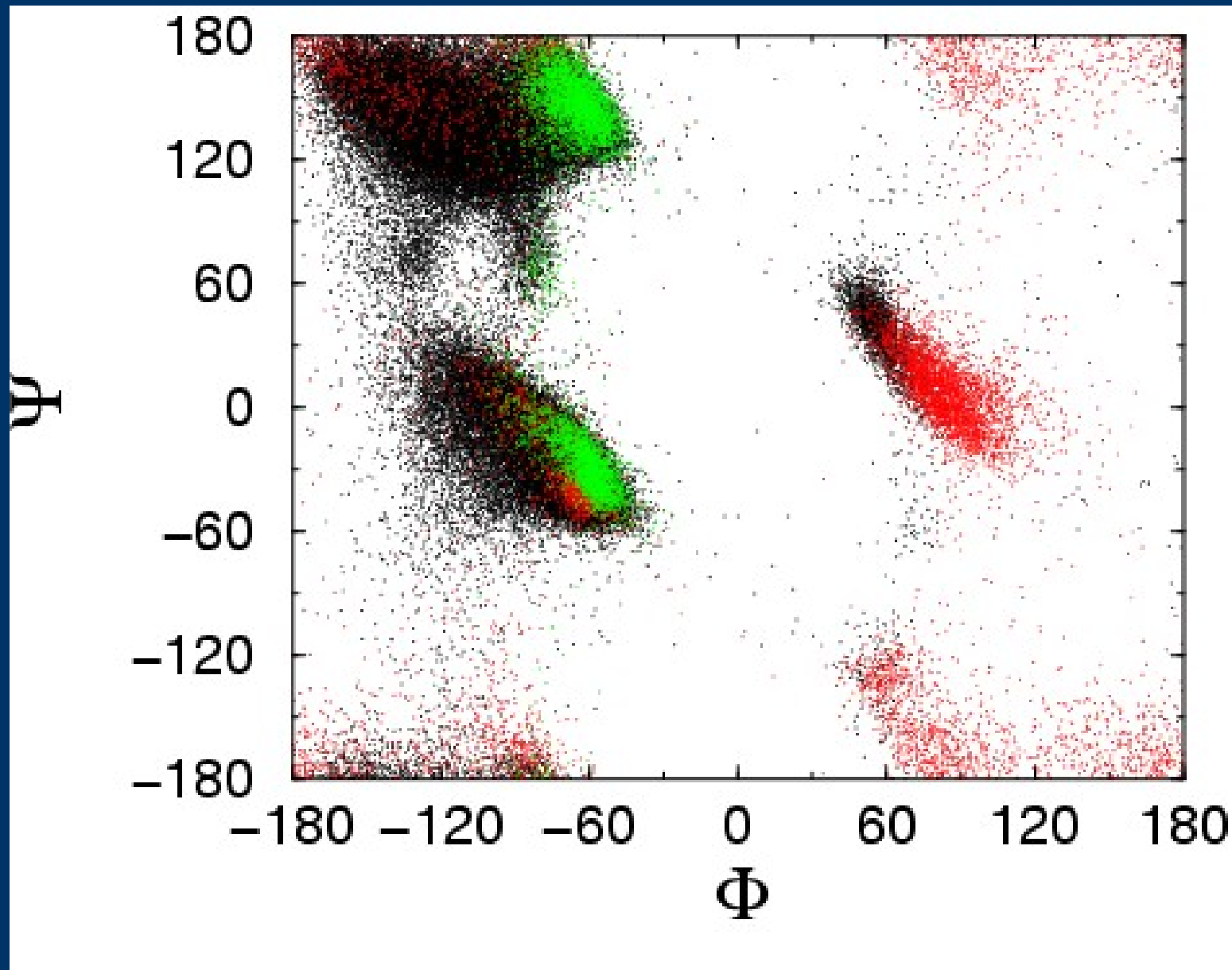
Non tutte le rotazioni sono permesse a causa dell'ingombro sterico del backbone e delle catene laterali.



Derivation diagram



Ramachandran plot (500 strutture)



rosso – glicina
verde – prolina

Struttura di proteine

- Cosa determina la struttura (la conformazione) di una proteina?
 - Anfinsen's hypothesis:
“This hypothesis states that the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, etc.) is the one in which the Gibbs free energy of the *whole system* is lowest”
-
-

Interazioni repulsive

- Interazioni steriche fra atomi non legati
repulsione per distanze inferiori alla somma
dei raggi di van der Waals

H ~ 1.0 Å

C ~ 1.8 Å


N ~ 1.5 Å

O ~ 1.4 Å

P ~ 2.1 Å

S ~ 2.0 Å

Interazioni attrattive

- Interazioni attrattive (dispersive)
 - Le fluttuazioni dinamiche della densità elettronica creano dei dipoli elettrici temporanei che inducono dipoli temporanei nelle molecole vicine. Le interazioni dipolo-dipolo indotto sono debolmente attrattive.
 - $E \sim -A/r^6$
 - La costante A è tipicamente piccola ~ 0.1 kcal/mol
 - Interazioni a corto raggio, molto deboli, ma presenti fra tutti gli atomi
- 

Interazioni elettrostatiche

- Elettronegativita' diversa fra atomi legati
- Gruppi ionizzabili
- Fenomeni di risonanza



- Gli atomi in una molecola possono avere una carica parziale
- Interazione secondo la legge di Coulomb: a lungo raggio e forte

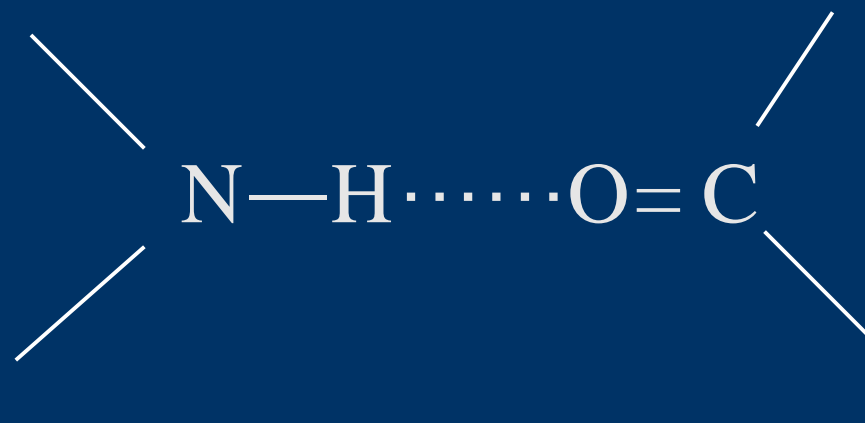
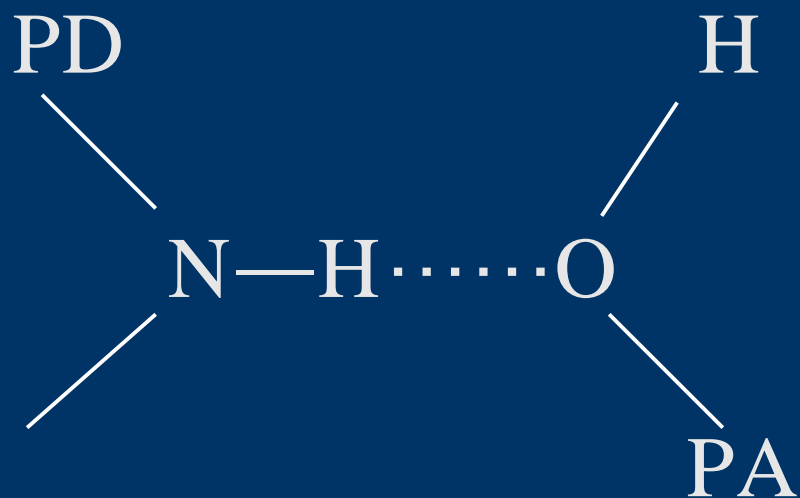
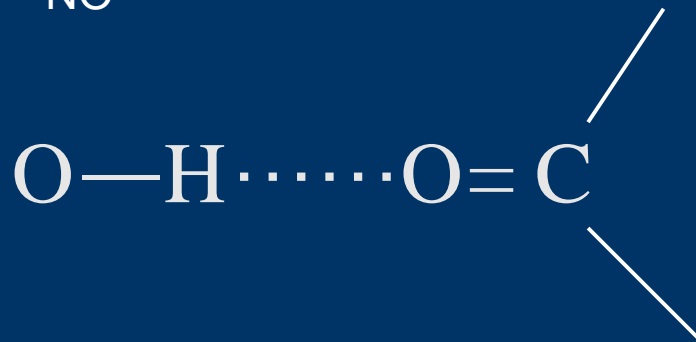
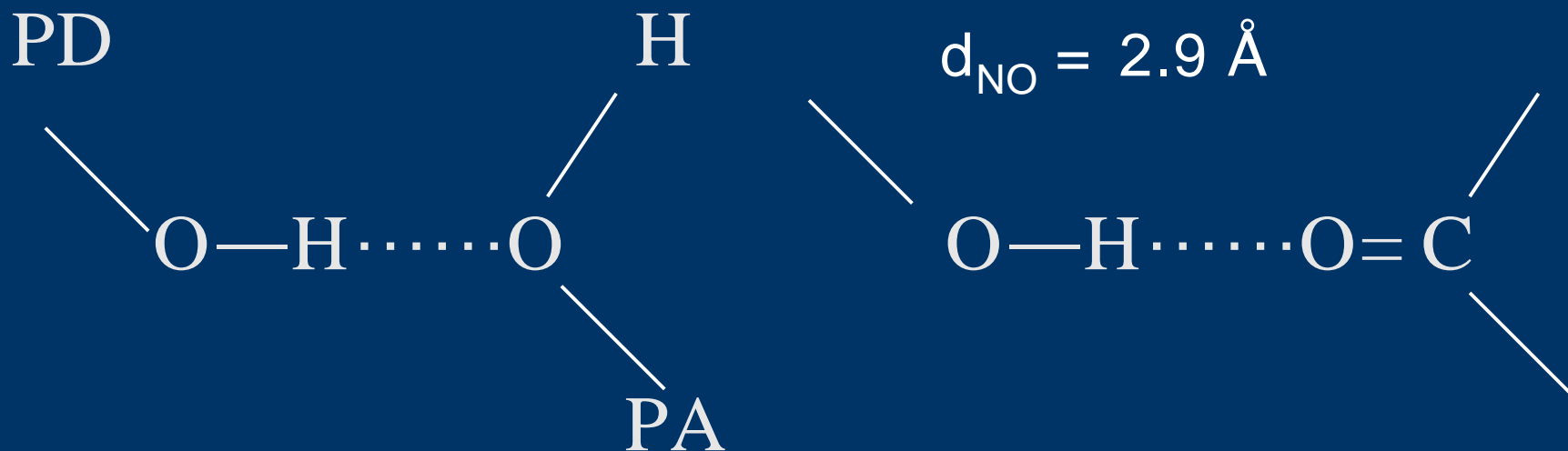
$$E = k \frac{q_1 q_2}{\epsilon r} \quad k = 332 \frac{\text{kcal A}}{e^2 \text{ mol}}$$

ϵ = costante dielettrica
2-4 alcani, molecole organiche
80 acqua

Legami idrogeno

$$d_{OO} = 2.8 \text{ \AA}$$

$$d_{NO} = 2.9 \text{ \AA}$$

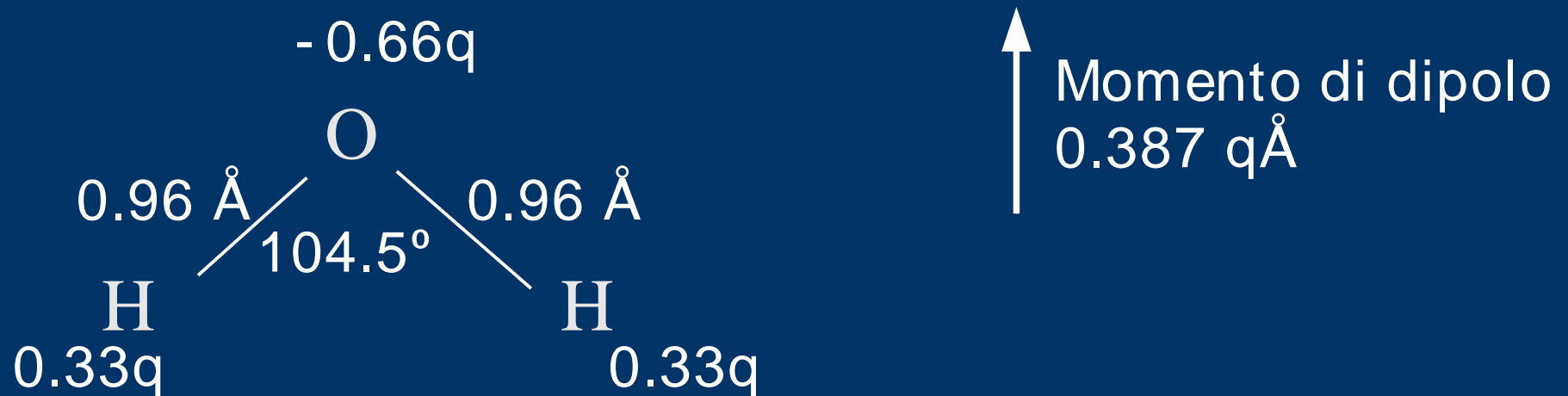


D—H·····A— Donatori e accettori: N, O e S
D H A approx. colineari. Angolo PA-A-H $\sim 115^\circ$

Struttura di proteine

- Cosa determina la struttura (la conformazione) di una proteina?
 - Anfinsen's hypothesis:
“This hypothesis states that the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, etc.) is the one in which the Gibbs free energy of the *whole system* is lowest”
-
-

Acqua



- Può formare due legami idrogeno come accettore e due come donatore
 - Forma strutture temporanee ($\sim 4\text{ps}$) legate con legami idrogeno che coinvolgono più molecole da 4 a 9 (liquido associato)
 - Alta costante dielettrica (~ 80)
-
-

Proprieta' dell'acqua

- Schermo efficiente delle interazioni elettrostatiche
 - Grande energia di solvatazione per gruppi carichi (es. sale)
 - Forma strutture temporanee (~ 4 ps) legate con legami idrogeno che coinvolgono piu' molecole da 4 a 9 (liquido associato).
 - Due importanti conseguenze delle interazioni fra molecole d'acqua:
 - costante dielettrica $\sim 80 \implies$ Grande energia di solvatazione per gruppi carichi (es. sale)
 - Alta tensione superficiale ~ 0.05 kcal/(mol \AA^2) alle interfaccie acqua alcani. “Effetto idrofobico”.
-
-

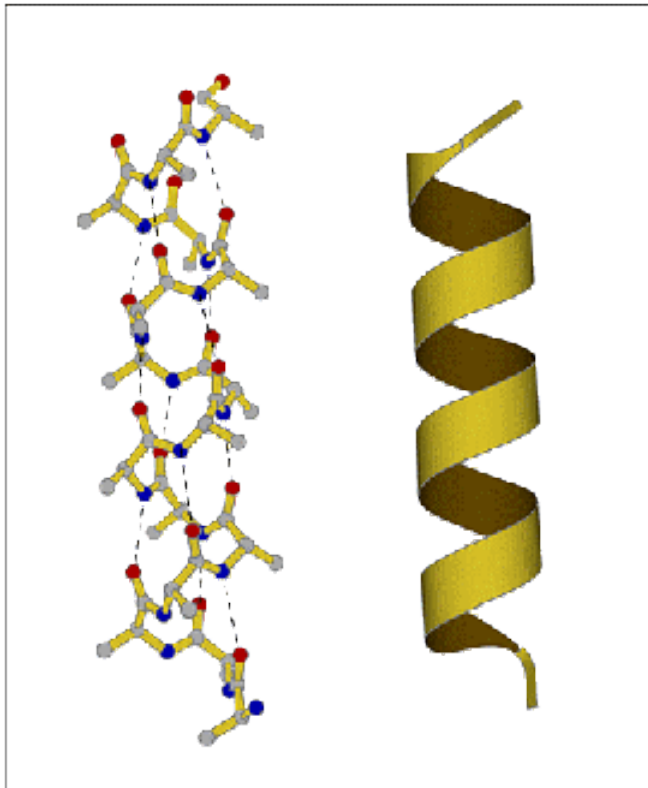
Elementi di struttura secondaria

- Le strutture delle proteine sono estremamente complesse, ma alcuni elementi si trovano frequentemente. Questi elementi rappresentano la cosiddetta struttura secondaria delle proteine
 - Gli elementi principali di struttura secondaria sono le α -eliche e i foglietti β .
 - Oltre a questi si trovano diversi altri elementi ricorrenti spesso stabilizzati da legami idrogeno
-
-

α -eliche



Alpha-helix



- Right-handed helix
- 3.6 residues per turn
- 5.4 Å per turn (“pitch”)
- 1.5 Å per residue (“rise”)
- Hydrogen bond between backbone carbonyl and NH of residue $n+4$
- A macro-dipole with positive N-terminal
- May be amphipatic

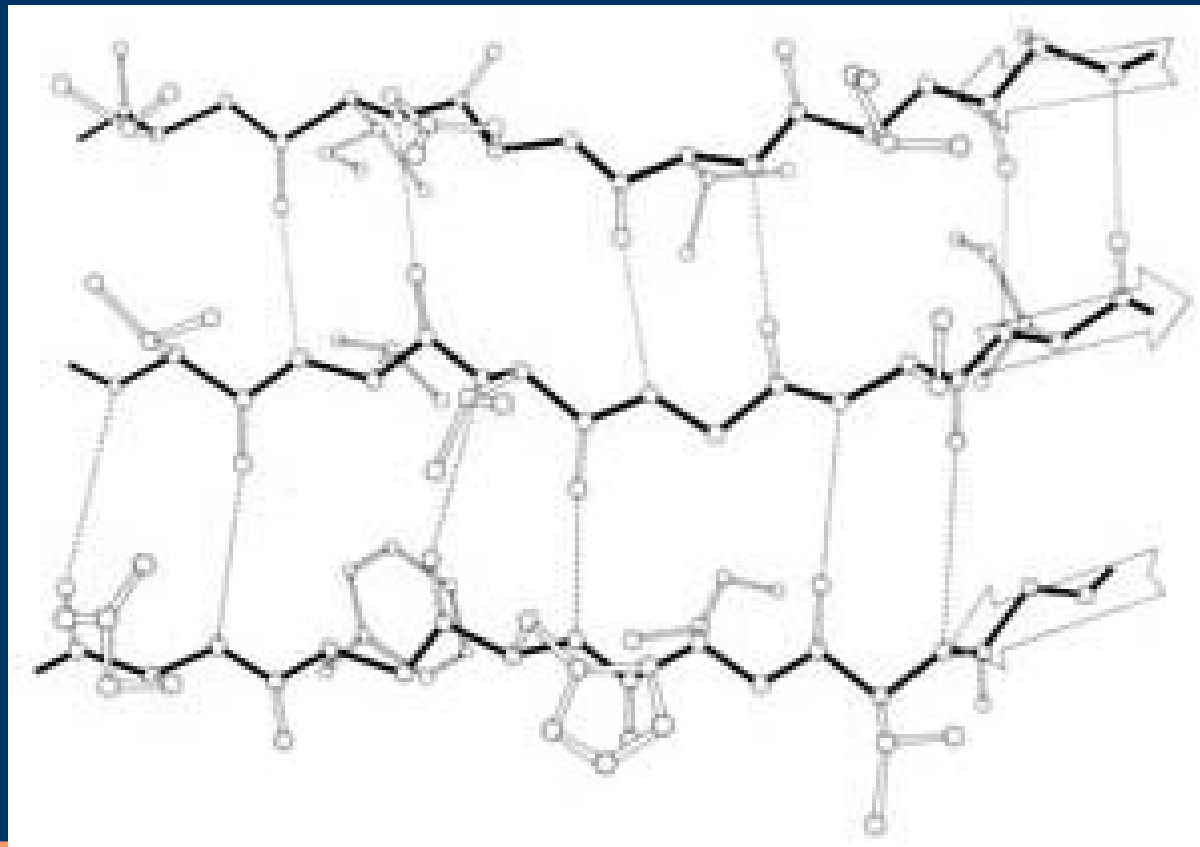
$$\phi = -63, \psi = -43$$

α -eliche

- Ca. il 32% degli amminoacidi si trova in α -elica.
 - La lunghezza media di un'elica e' ca. 12 residui
 - Essendo spesso alla superficie delle proteine hanno spesso una faccia idrofobica ed una idrofilica
 - Gli accettori e i donatori di legami idrogeno non impegnati in legami idrogeno ai due terminali sono spesso coinvolti in motivi detti di "helix capping"
-
-

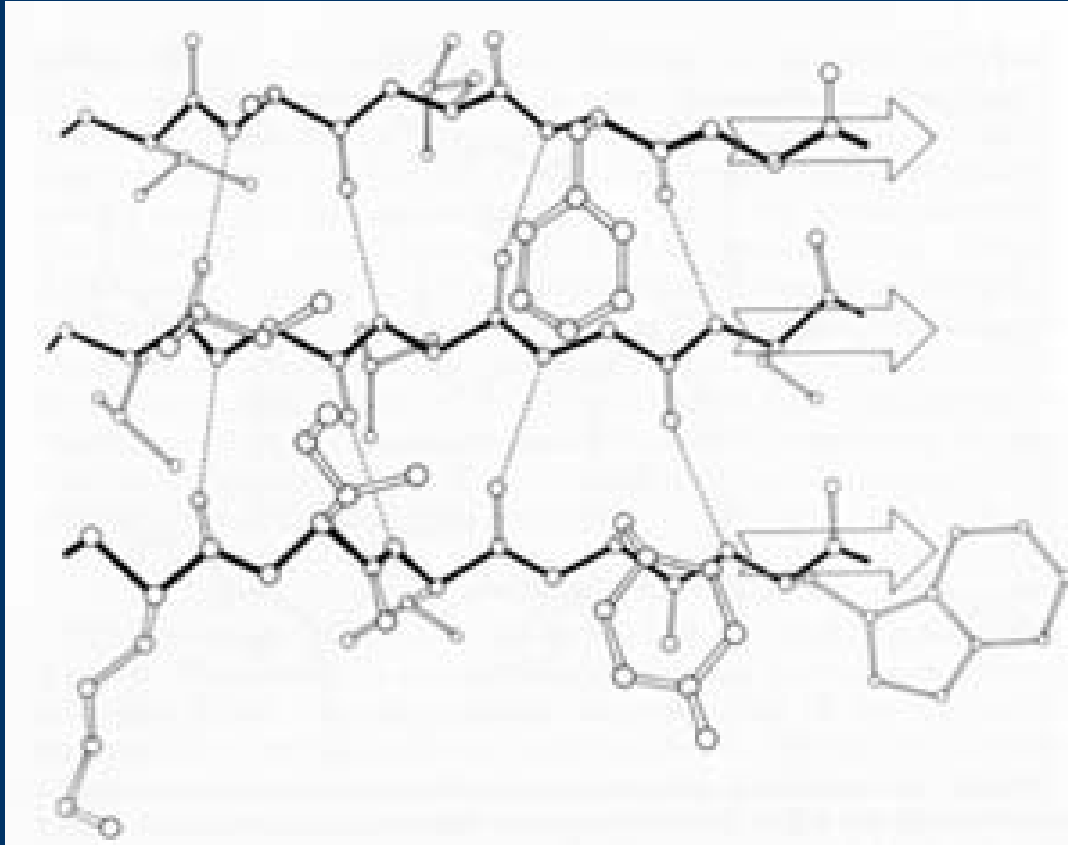
β -sheet

- Un altro tipo di struttura secondaria frequente e' costituita dai beta-sheet (foglietti beta) in cui le catene estese si appaiano con legami idrogeno fra gli atomi del backbone,antiparallelamente:



β -sheet

-o parallelamente:



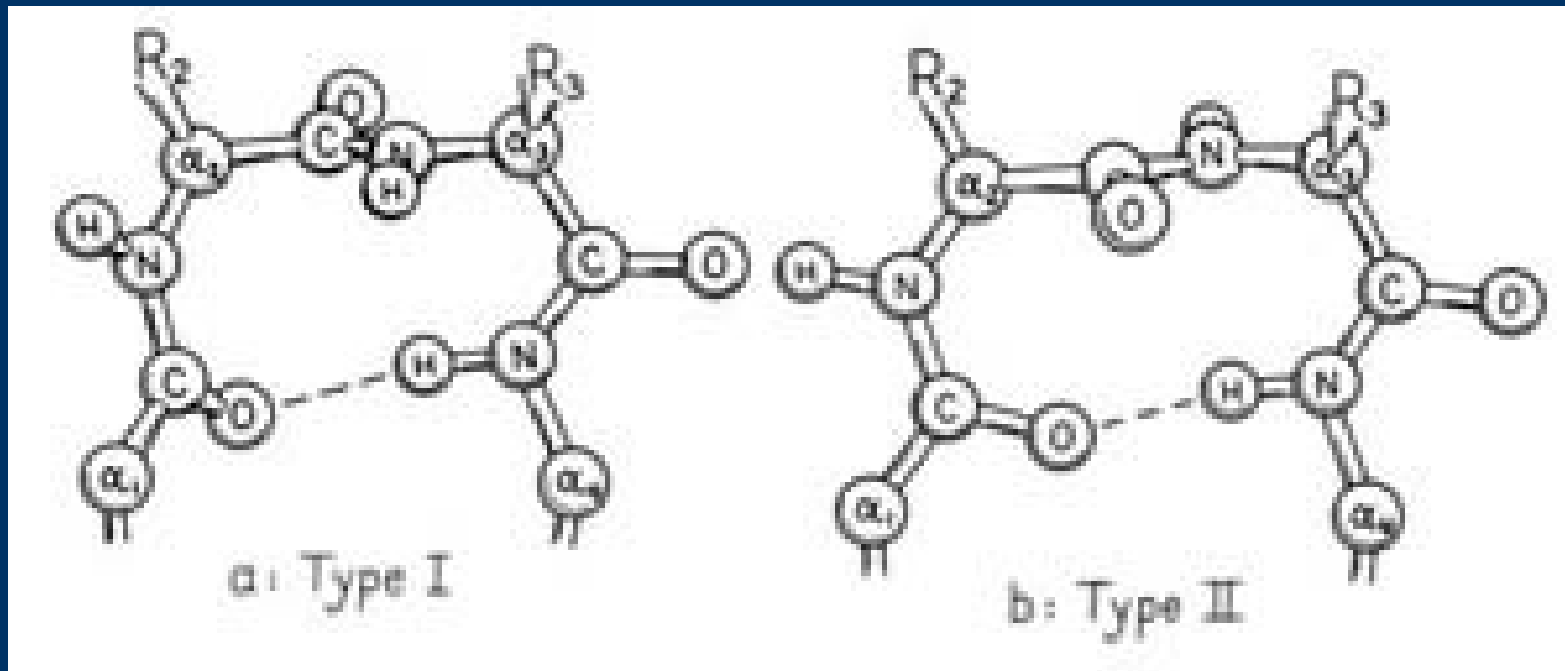
- Gli sheet paralleli sono piu' rigidi e strutturalmente meglio definiti degli sheet antiparalleli.

β -sheet

- Le strutture β sono favorite da residui ramificati al $C\beta$ (Val, Ile, Thr) e da residui ingombranti (Trp)
 - Le catene laterali si alternano da una parte all'altra del foglietto.
 - Se il foglietto e' esposto si puo' avere una alternanza di residui idrofobici e polari
 - Le catene laterali su strands affacciati generalmente sono impaccate strettamente.
 - I foglietti possono essere interamente costituiti da strand paralleli, antiparalleli o misti (ca. il 20%)
-
-

tight turns

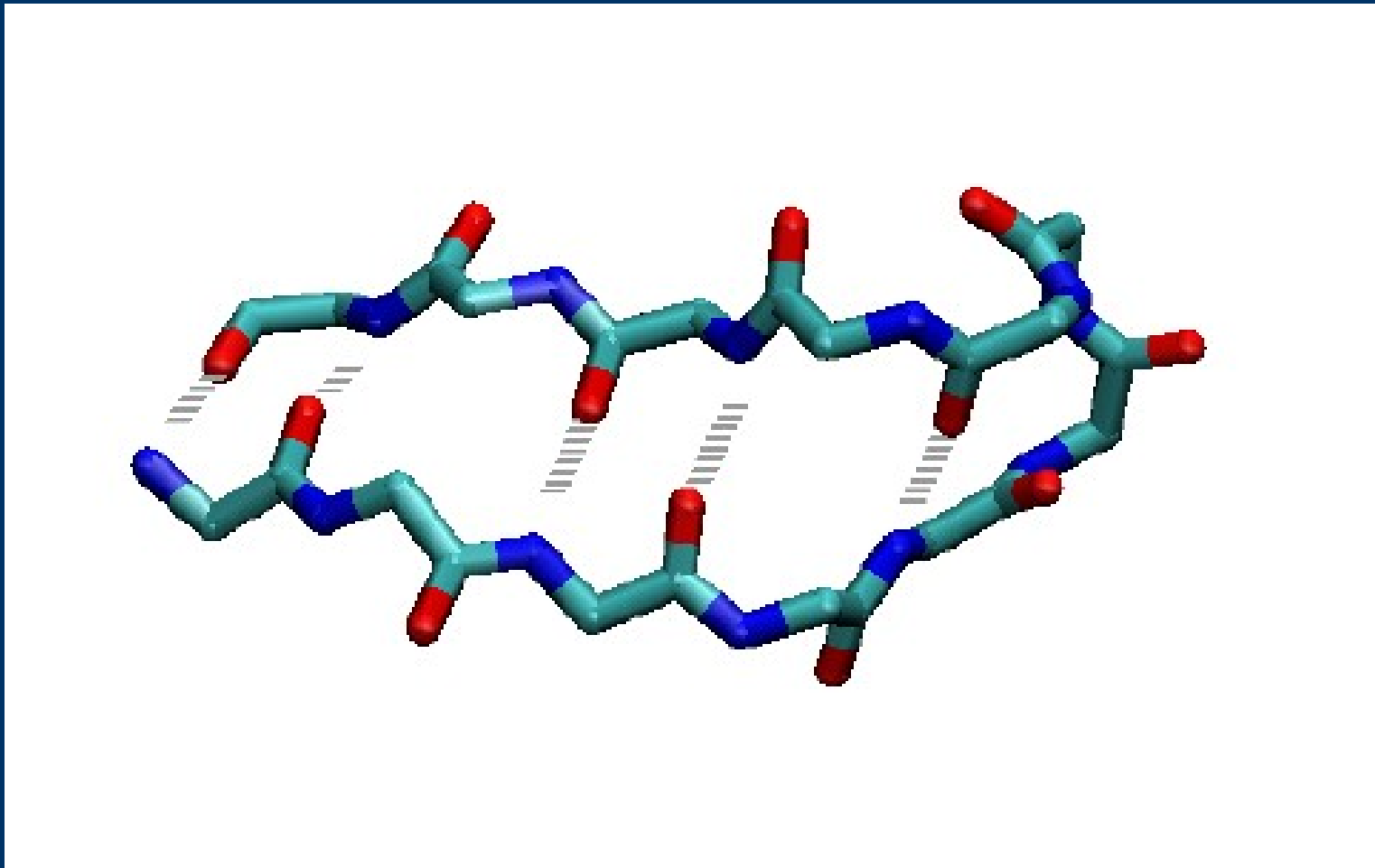
- spesso si trovano cambi di direzione nella catena in corrispondenza di due residui. La conformazione di questi residui prende il nome di turn.



Queste conformazioni spesso richiedono la presenza di glicina o prolina in posizioni chiave.

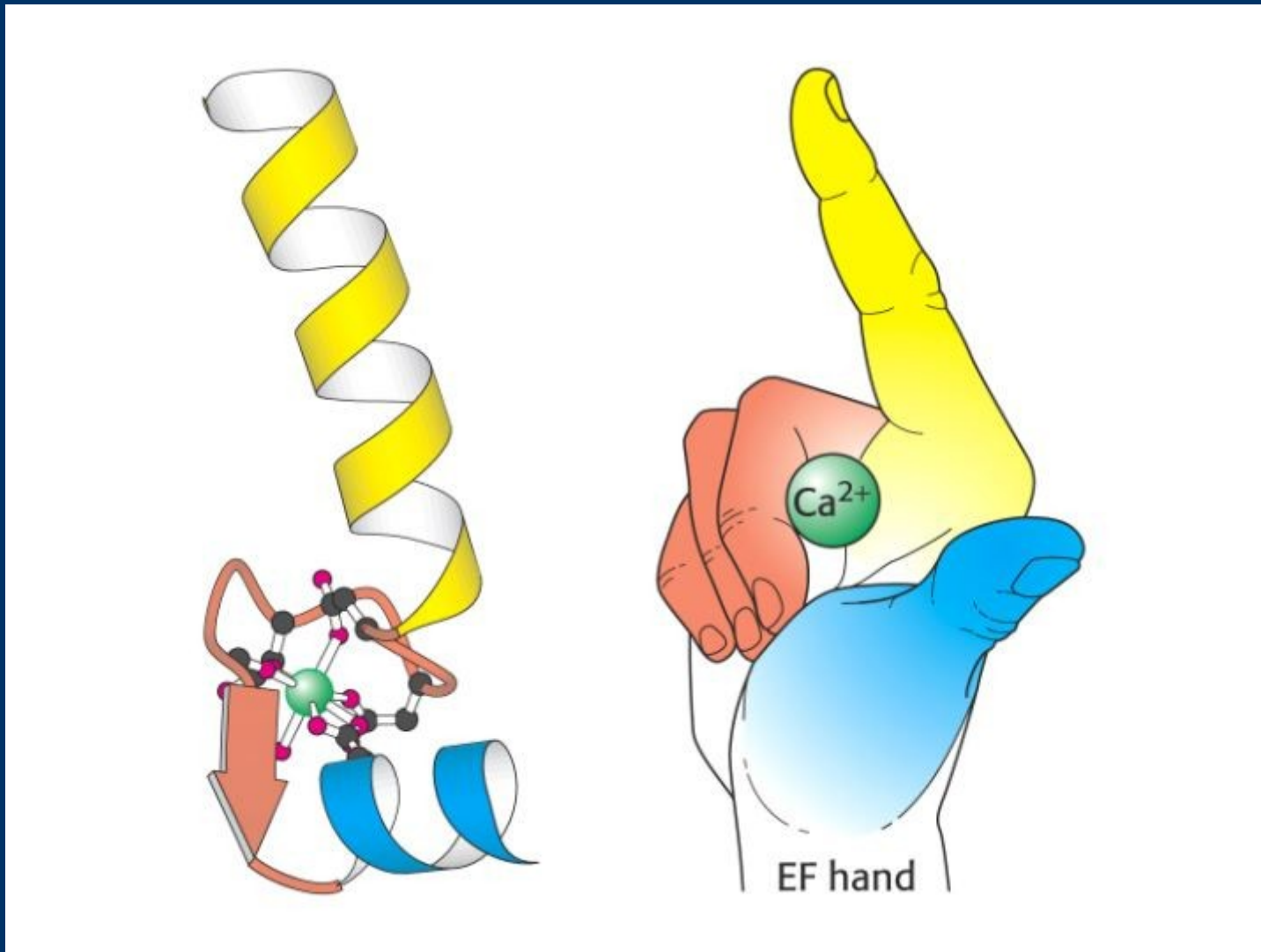
Motivi strutturali

- Elementi di struttura secondaria si possono combinare in semplici motivi. Alcuni esempi
- beta-hairpin:



Motivi

- helix-loop-helix (EF-hand):



Motivi

- Questo motivo come altri sono accompagnati da conservazione in sequenza:

Table 2.2 Amino acid sequences of calcium-binding EF motifs in three different proteins

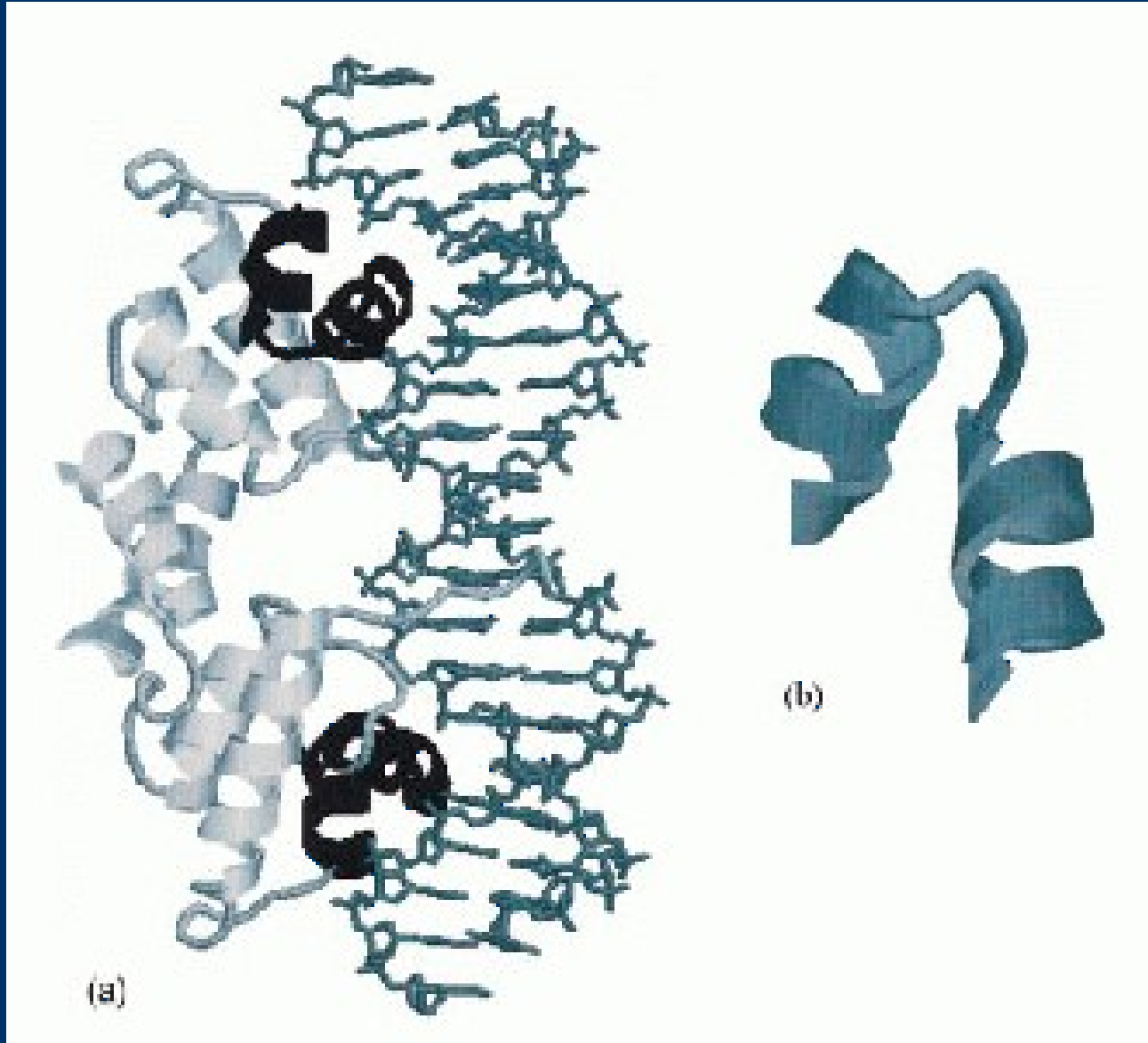
<i>Parvalbumin</i>	V	K	K	A	F	A	I	I	D	Q	D	K	S	G	F	I	E	E	D	E	L	K	L	F	L	Q	N	F
<i>Calmodulin</i>	F	K	E	A	F	S	L	F	D	K	D	G	D	G	T	I	T	T	K	E	L	G	T	V	M	R	S	L
<i>Troponin-C</i>	L	A	D	C	F	R	I	F	D	K	N	A	D	G	F	I	D	I	E	E	L	G	E	I	L	R	A	T



Calcium-binding residues are brown, and residues that form the hydrophobic core of the motif are light green. The helix-loop-helix region shown underneath is colored as in Figure 2.13.

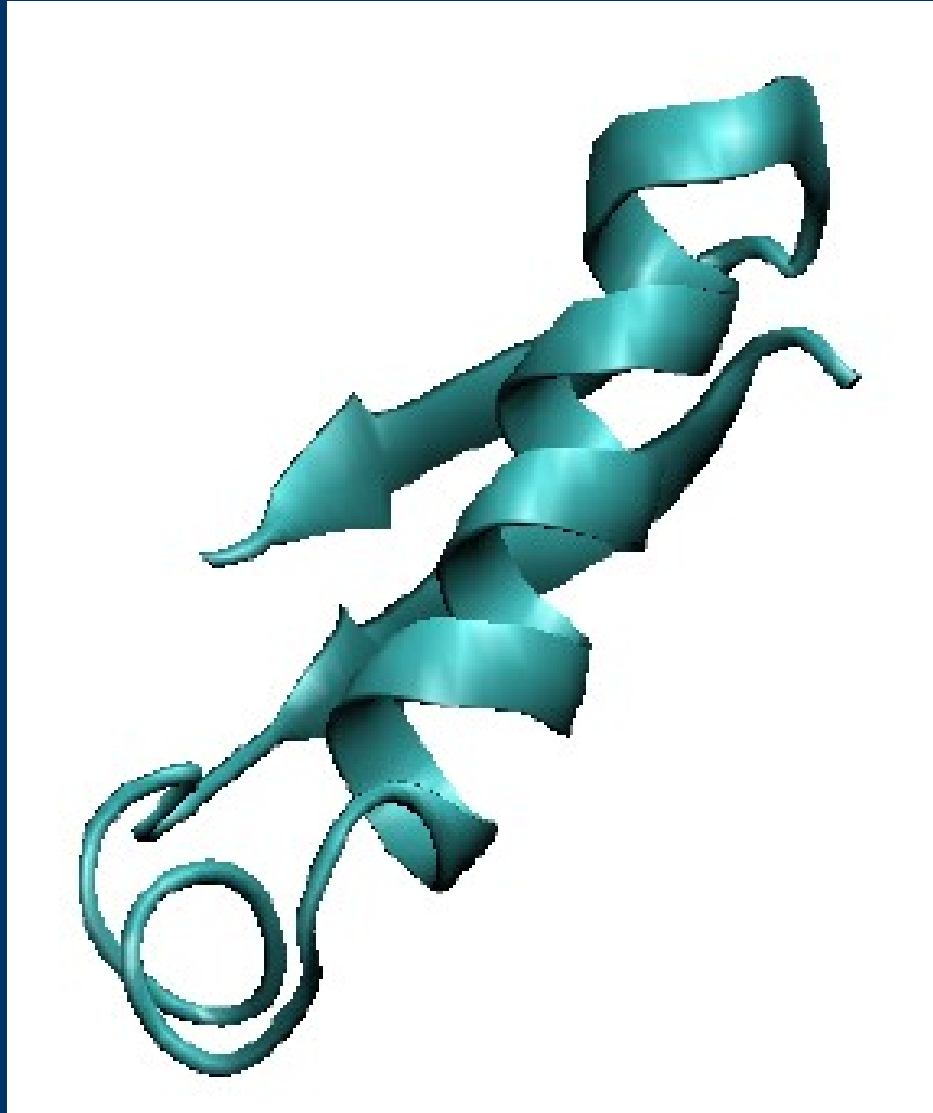
Motivi

- helix-turn-helix, in fattori di trascrizione



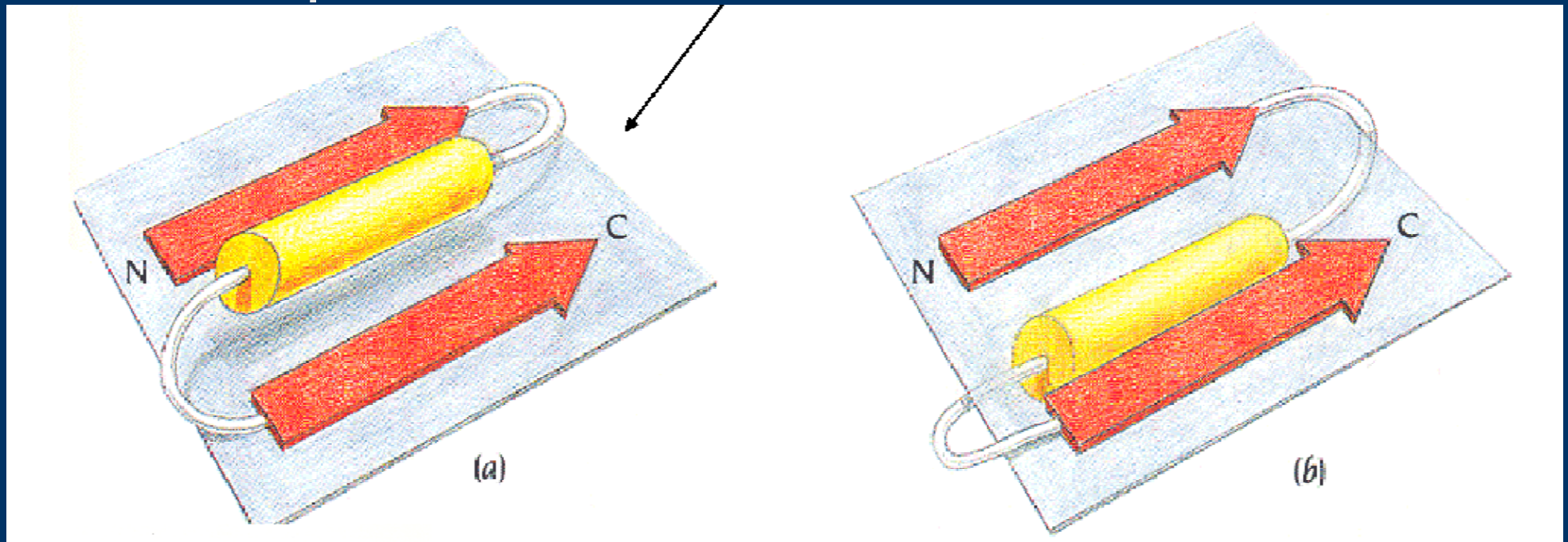
Motivi

- beta-alpha-beta (ad es. nei parallel beta barrel):



Motivi

- Il motivo beta-alpha-beta ha quasi sempre chiralita' destrorsa



Classificazione delle strutture

- Gli elementi di struttura secondaria, i motivi e la loro organizzazione permette di classificare le proteine secondo uno schema gerarchico.
 - Ad esempio nella classificazione di SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/index.html>) i quattro livelli della gerarchia sono:
 - classe (alpha, beta, alpha/beta, alpha+beta)
 - ripiegamento (fold) (ad es. globin)
 - superfamiglia (ad es. alpha-helical ferredoxins)
 - famiglia (ad es. Dihydropyrimidine dehydrogenase, N-terminal domain).
-
-

Classi di strutture

Le strutture di proteine possono essere classificate in un piccolo numero di classi principali

(da SCOP – structural classification of proteins)

Class	f	sfam	fam
All alpha proteins	218	376	608
All beta proteins	144	290	560
Alpha and beta proteins (α/β)	136	222	629
Alpha and beta proteins ($\alpha+\beta$)	279	409	717
Multi-domain proteins	46	46	61
Membrane and cell surf. prot.	47	88	99
Small proteins	75	108	171
Total	945	1539	2845

f = folds

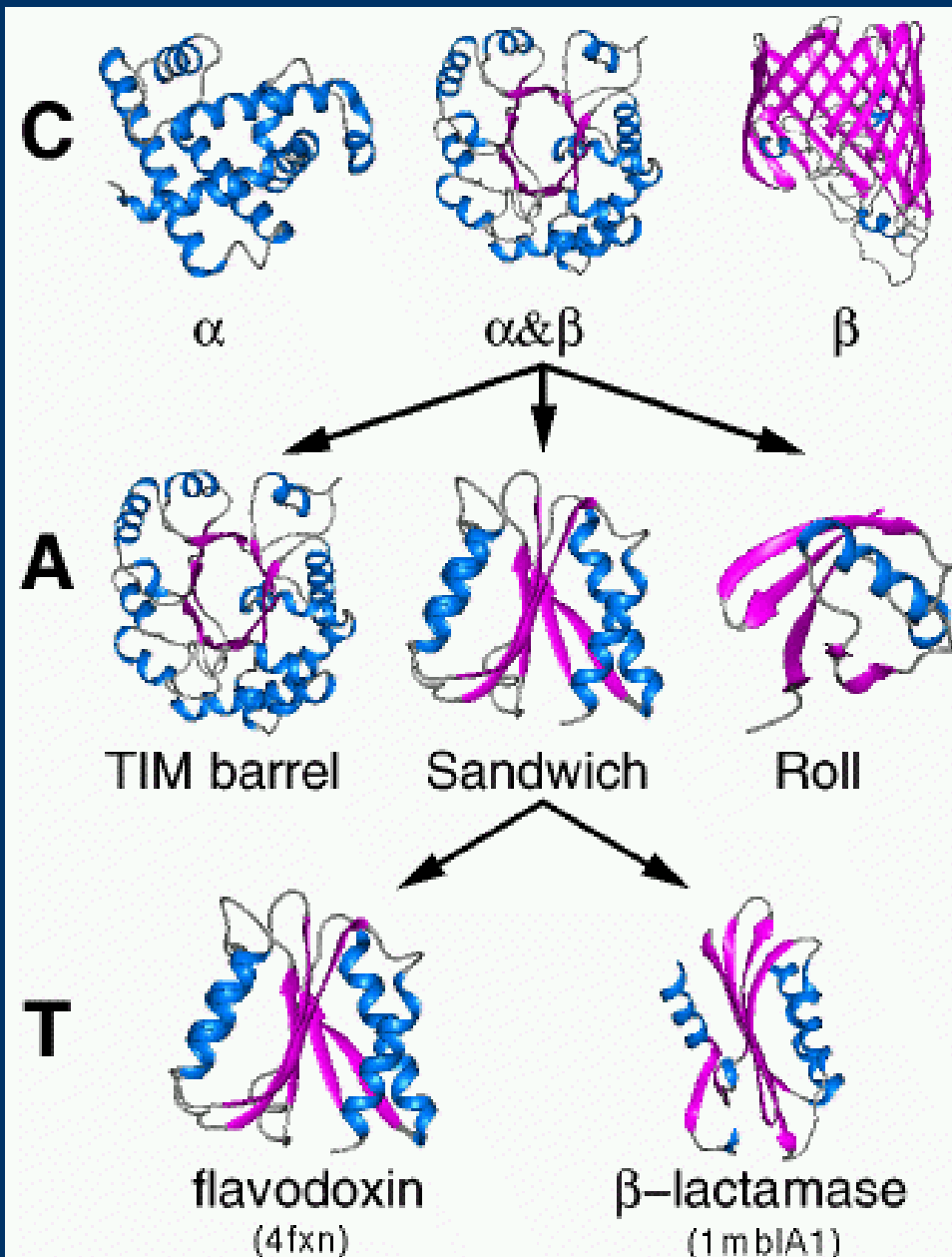
sfam = superfamilies

fam = families

Classificazione delle strutture

- Uno schema simile e' adottato in CATH (<http://cathwww.biochem.ucl.ac.uk/latest/>)
 - Class
 - Architecture
 - Topology
 - Homology
-
-

Classificazione di strutture (CATH)



CLASS

ARCHITECTURE

TOPOLOGY

HOMOLOGY

Struttura e funzione

- Non c'è una corrispondenza univoca fra struttura e funzione, ma in molti casi un ripiegamento è associato ad una funzione specifica.
 - La predizione della struttura può aiutare a ipotizzare la funzione di una proteina
 - La conoscenza della struttura permette di razionalizzare risultati sperimentali o osservazioni circa i mutanti di una proteina
-
-

Struttura di proteine

Le strutture di proteine vengono determinate mediante cristallografia a raggi X o NMR (piccole proteine)

Le strutture possono essere incomplete o definite solo in maniera approssimativa (ad es. solo la traccia dei C_{α})

Le strutture sono depositate come files di testo nella banca dati Protein Data Bank

Il testo che contiene le coordinate atomiche e' visualizzabile con programmi appositi, ad es. SwissPdbViewer o ViewerLite

Vedere le molecole.....

Non possiamo usare normali microscopi per guardare le molecole. La lunghezza d'onda del visibile e' molto maggiore delle distanze interatomiche e non ci sono lenti per la radiazione a bassa lunghezza d'onda

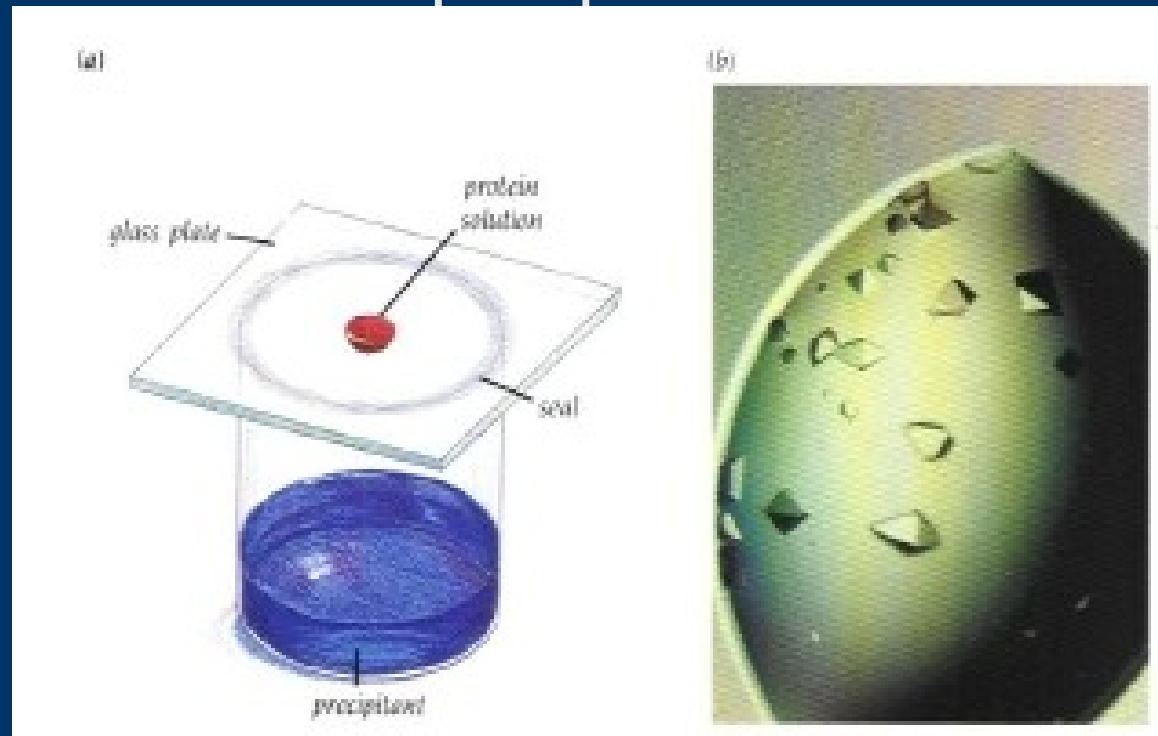
Si usano onde elettromagnetiche con lunghezza d'onda di $\sim 1\text{\AA}$ (raggi X)

La radiazione fa oscillare gli elettroni che a loro volta emettono radiazione

Usando un cristallo il debolissimo segnale di una molecola viene ripetuto in maniera coerente per tutte le molecole del cristallo

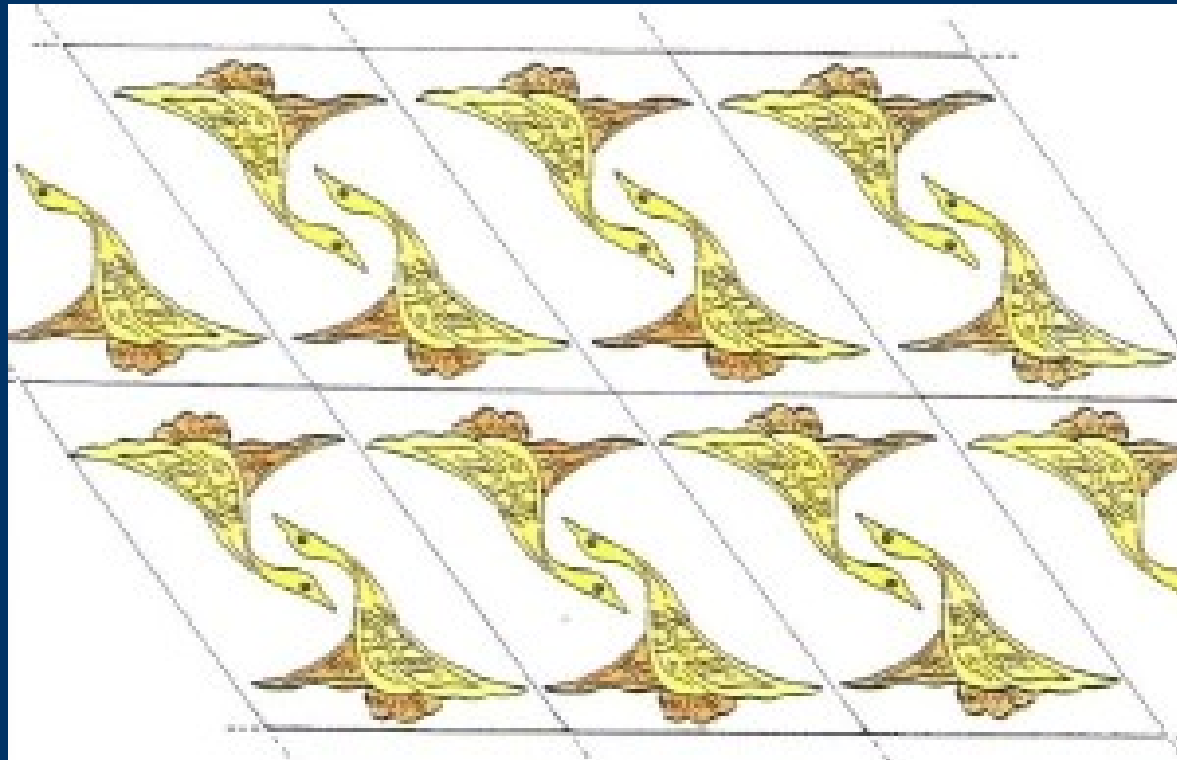
Cristallografia a raggi X

- Il punto di partenza per lo studio strutturale mediante diffrazione di raggi X e' l'ottenimento di un cristallo che contenga le proteine.
- Questo non e' semplice in generale ed estremamente difficile per proteine di membrana.



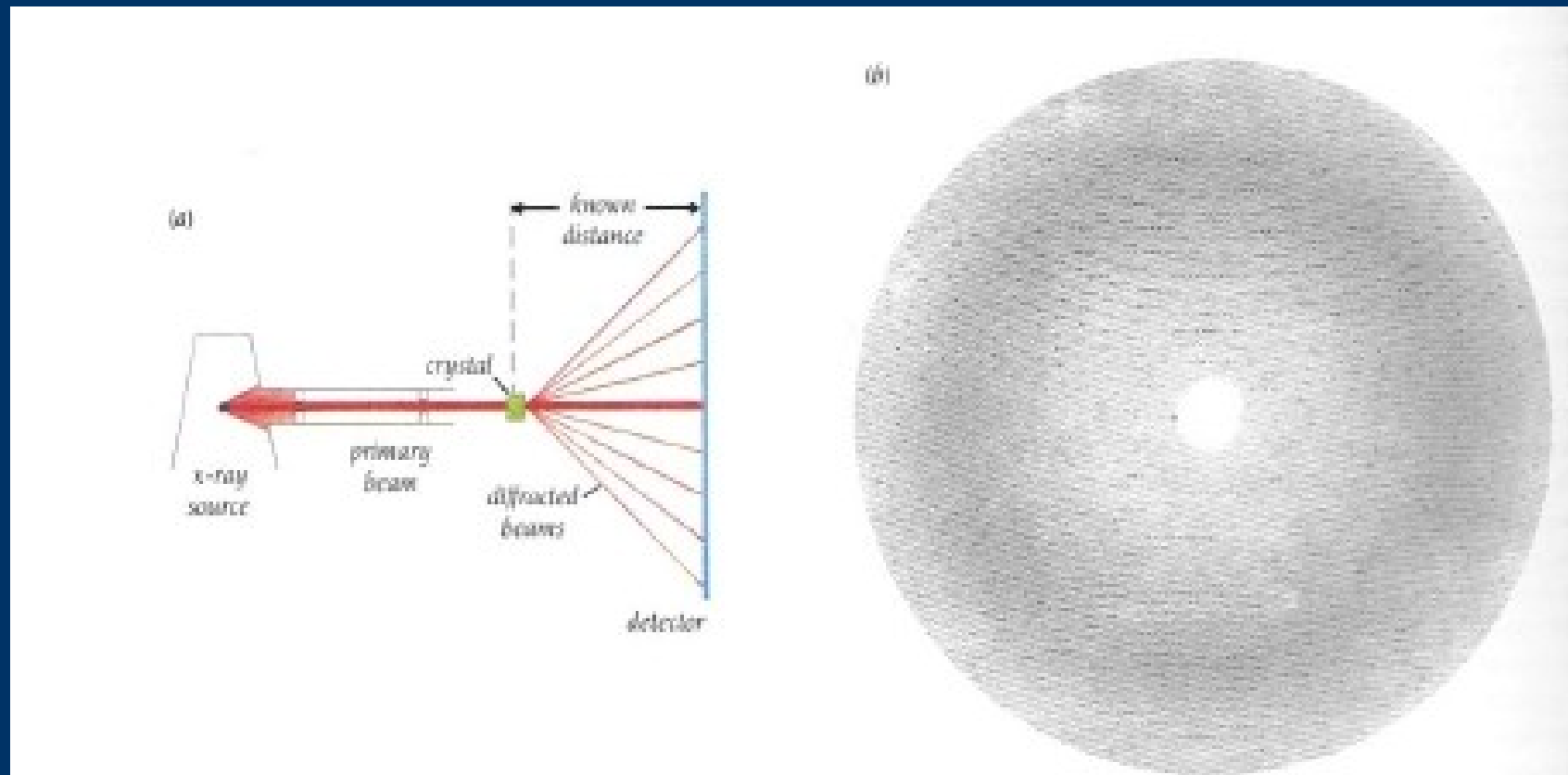
Cristallografia a raggi X

- Un cristallo e' costituito da unita' identiche (unit cell), che contengono una o piu' molecole, che si ripetono nello spazio. Una stessa molecola puo' dare luogo a forme cristallografiche diverse.



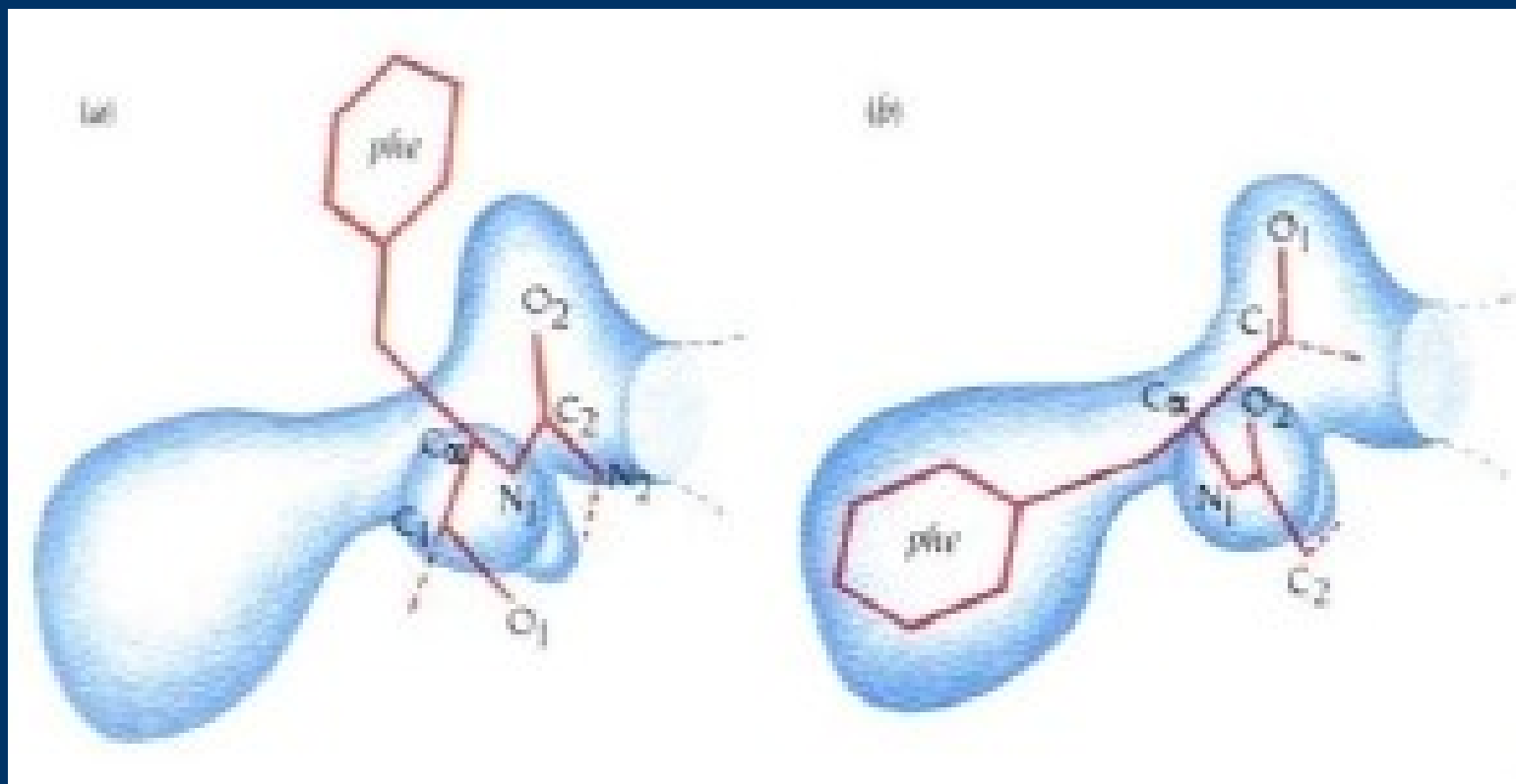
Cristallografia a raggi X

- Sul cristallo della biomolecola viene inviato un fascio di raggi X che viene diffratto. Il campione ruota e il pattern di diffrazione viene registrato.



Cristallografia a raggi X

- Dalla analisi al calcolatore delle macchie di diffrazione viene ricostruita la densità elettronica che riflette la struttura della molecola. Se manca densità elettronica ci possono essere parti mancanti o non accurate



Banche dati di strutture

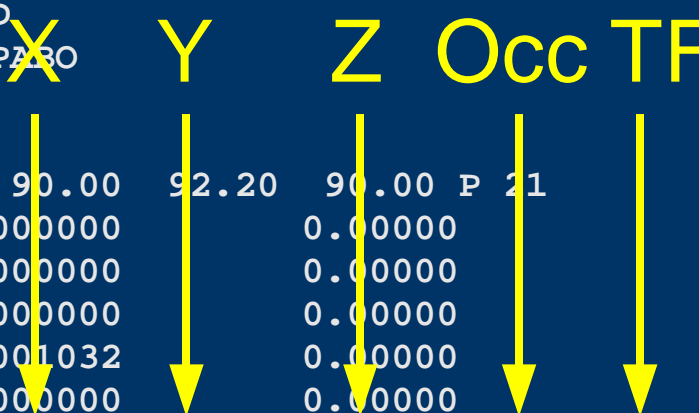
- **PDB** — Protein Data Bank un tempo al Brokhaven National Laboratory ora mantenuta dal Research Collaboratory in Structural Bioinformatics (www.pdb.org)
 - La PDB contiene ca. 50000 strutture (con molta ridondanza) che a loro volta contengono diverse catene di proteine, ma anche acidi nucleici e piccole molecole
 - Le strutture sono ottenute mediante Cristallografia a Raggi X (ca. 42000), NMR (ca. 7000), ma anche in minima parte mediante modellistica.
 - L'informazione e' essenzialmente contenuta nelle coordinate degli atomi della molecola.
-
-

Esempio di file PDB

- 1lmb.pdb

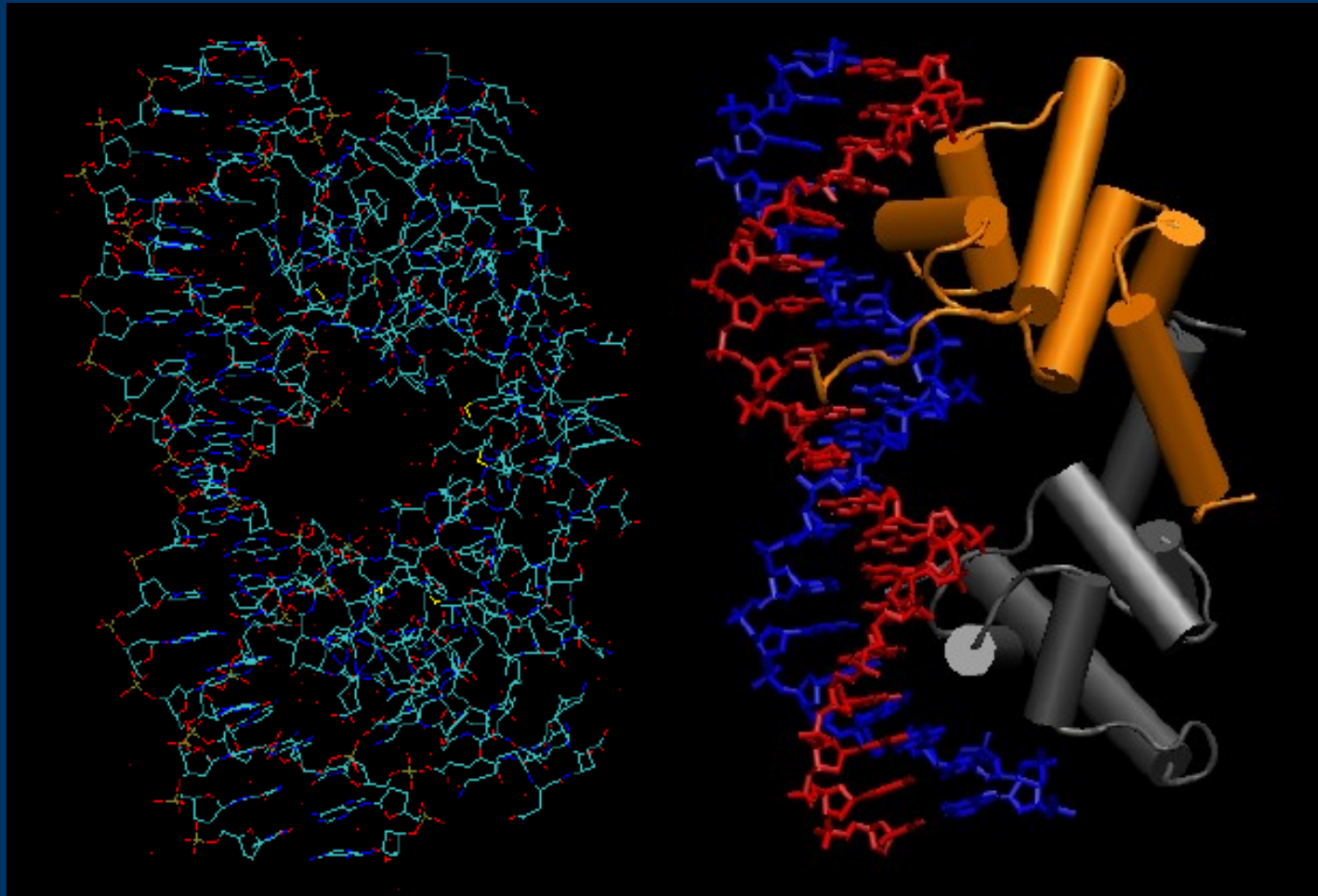
```
HEADER      DNA-BINDING REGULATORY PROTEIN                05-NOV-91   1LMB      1LMB      2
COMPND      LAMBDA REPRESSOR/OPERATOR COMPLEX            1LMB      3
SOURCE      BACTERIOPHAGE (LAMBDA)                      1LMB      4
AUTHOR      L.J.BEAMER,C.O.PABO                          1LMB      5
REVDAT      1   31-OCT-93 1LMB      0                1LMB      6
SPRSDE      15-OCT-93 1LMB      1LRD                1LMB      7
JRNL        AUTH   L.J.BEAMER,C.O.PABO                    1LMB      8
.....
.....
CRYST1      37.220   68.720   57.030   90.00   92.20   90.00 P 21          2  1LMB  110
ORIGX1      1.000000   0.000000   0.000000                0.000000          1LMB  111
ORIGX2      0.000000   1.000000   0.000000                0.000000          1LMB  112
ORIGX3      0.000000   0.000000   1.000000                0.000000          1LMB  113
SCALE1      0.026867   0.000000   0.001032                0.000000          1LMB  114
SCALE2      0.000000   0.014552   0.000000                0.000000          1LMB  115
SCALE3      0.000000   0.000000   0.017548                0.000000          1LMB  116
ATOM        1   O5*    A  1    1        31.960  -0.182   33.011   1.00  53.09          1LMB  117
ATOM        2   C5*    A  1    1        33.188  -0.901   32.936   1.00  47.51          1LMB  118
ATOM        3   C4*    A  1    1        32.885  -2.365   32.858   1.00  47.97          1LMB  119
ATOM        4   O4*    A  1    1        31.900  -2.641   33.907   1.00  42.86          1LMB  120
.....
.....
TER
ATOM        411  O5*    T  2    21       -11.598 -10.089 -18.613   1.00  59.87          1LMB  527
ATOM        412  C5*    T  2    21       -11.458 -10.632 -19.913   1.00  51.63          1LMB  528
ATOM        413  C4*    T  2    21       -10.024 -10.505 -20.364   1.00  47.92          1LMB  529
.....
.....
```

X Y Z Occ TF



Visualizzazione e analisi

- Occorrono programmi specifici per leggere le coordinate e mostrare la struttura (molecular graphics) e analizzare le strutture



Confronto di strutture

Date due strutture per confrontarle devo:

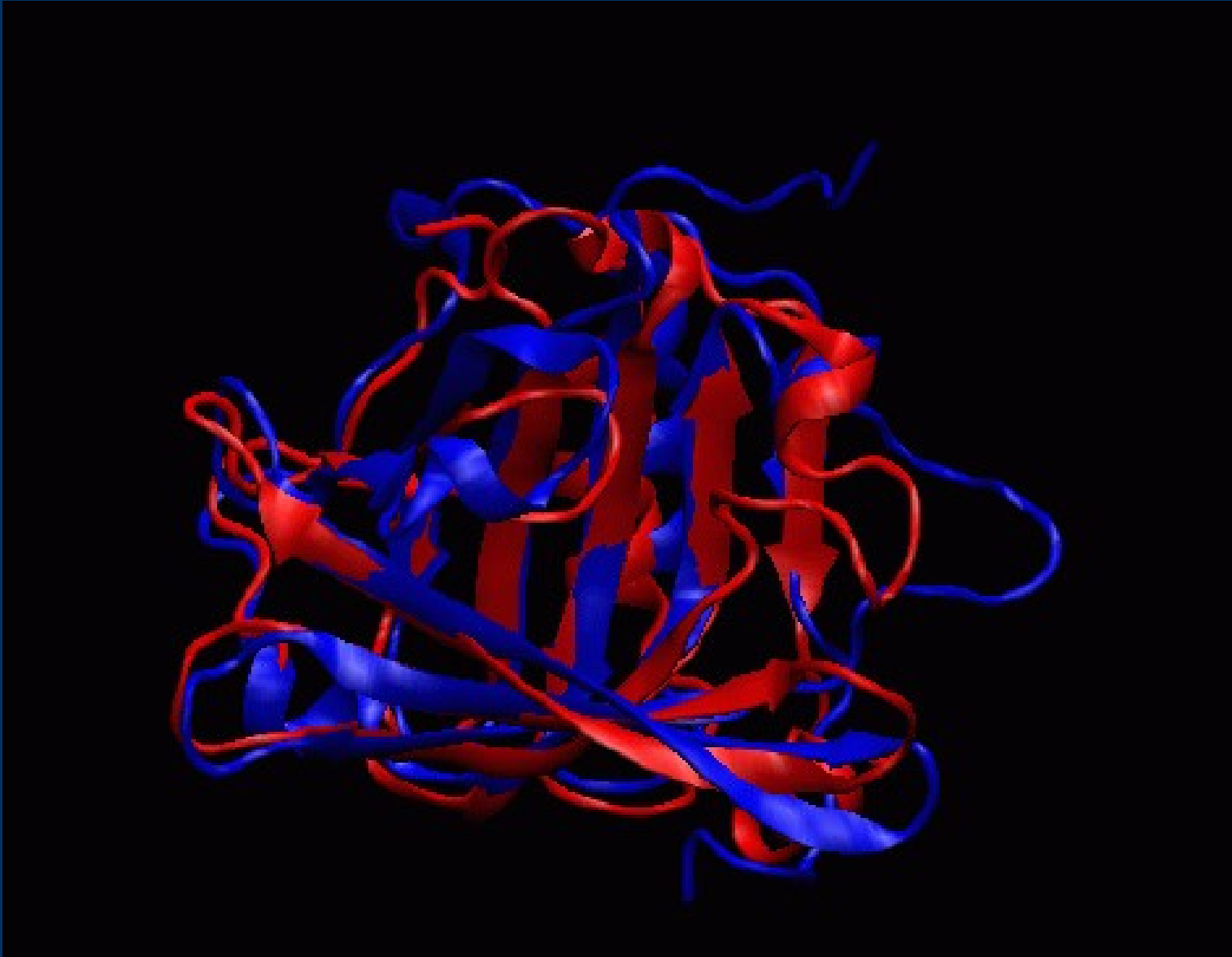
1) stabilire quali parti si corrispondono nelle due strutture (ad es. gli atomi del backbone, oppure gli atomi del backbone delle regioni conservate)

2) trovare la posizione e l'orientazione “migliore” per sovrapporre le due molecole

“migliore” indica quasi sempre tale che la distanza (RMSD) fra gli atomi corrispondenti sia minima

$$RMSD = \sqrt{\frac{\sum_{i=1}^n \|\vec{r}_i^{model} - \vec{r}_i^{native}\|^2}{n}}$$

Confronto di strutture



RMSD =
2.9 Å
146 aa
su 162



Predizione di struttura secondaria

- Accuratezza (a tre stati: α -eliche, β -strand e altro (coil)) di ca. 80%. Accuratezza migliore per le eliche
 - Chou-Fasman: il primo metodo basato sulle propensita' di ogni residuo. Bassa accuratezza (57% a tre stati)
 - GOR (Garnier Osguthorpe Robson) Considera la correlazione fra un residuo e i residui che precedono e seguono (fino a + e -8)
-
-

Predizione di struttura secondaria

Il salto di qualita' nelle predizioni dovuto ad alcuni elementi:

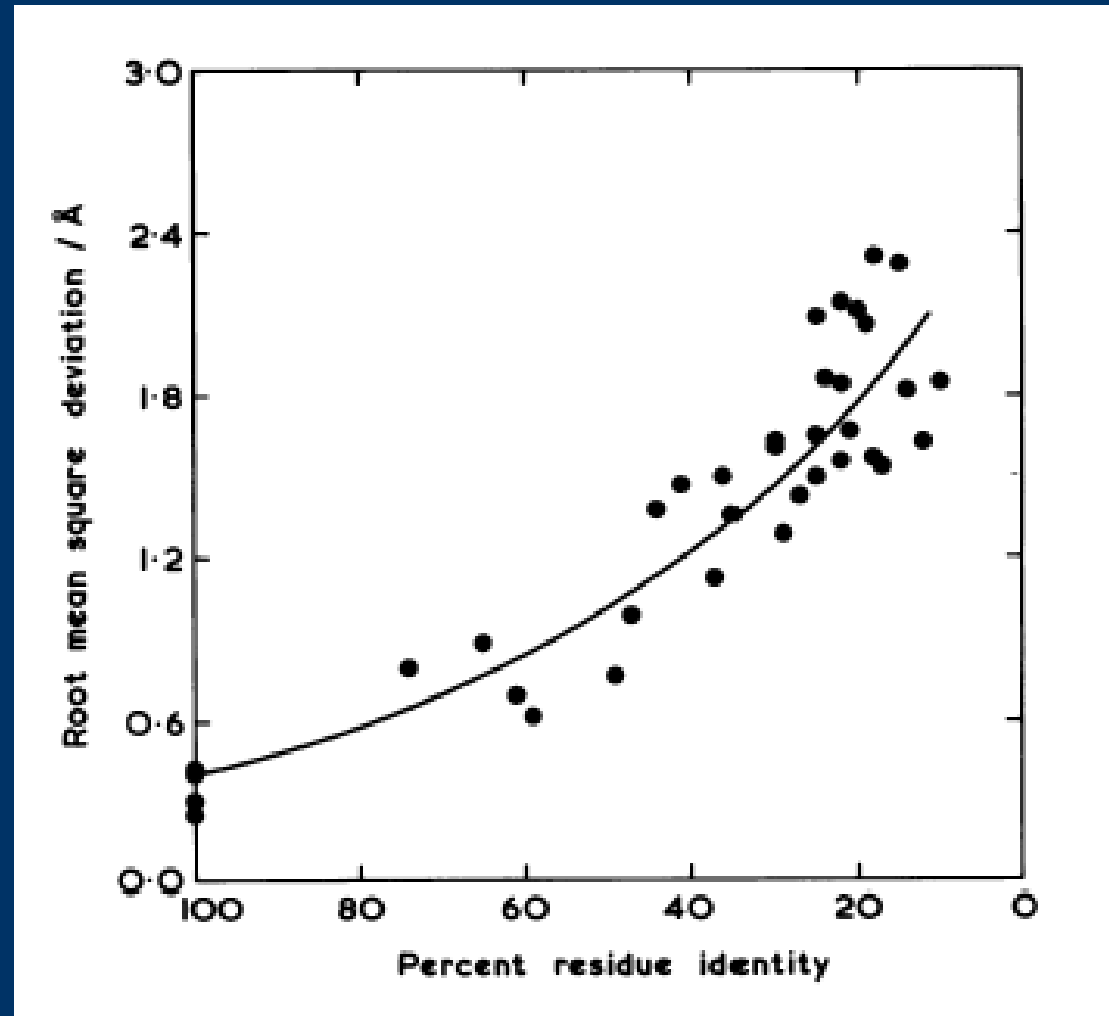
- l'utilizzo di profili al posto delle sequenze singole
- l'utilizzo di reti neurali o metodi statistici sofisticati come HMM.
- l'utilizzo di metodi consensus

Nei test alla cieca recenti la predizione di struttura secondaria raggiunge ca. l'80% di accuratezza a tre stati

Predizione della struttura terziaria

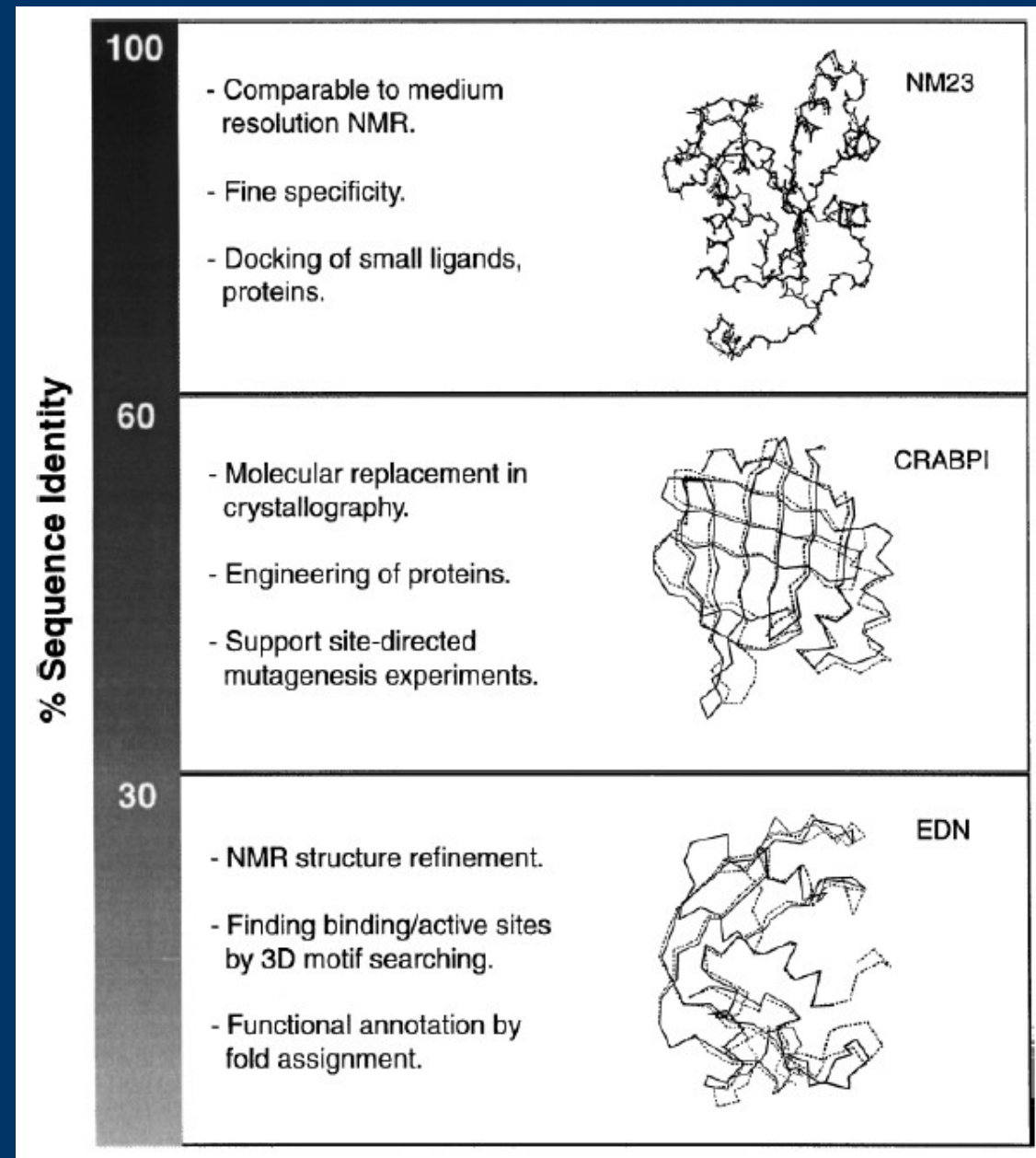
Similarita' di
sequenza vs.
similarita' di
struttura

Chothia e Lesk
(EMBO J. 1986)
hanno osservato
che sequenze con
piu' del 50% di
identita' assumono
la stessa struttura,
mentre sotto il 20%
non si puo' dire
nulla.



Predizione della struttura terziaria

- modellistica per omologia.
Chothia e Lesk (1986): la struttura evolve piu' lentamente della sequenza
- Riconoscimento di fold.
Il numero di folds e' comunque limitato
- Modellistica ab-initio
Le proprieta' strutturali delle proteine sono comprese e simulabili al calcolatore



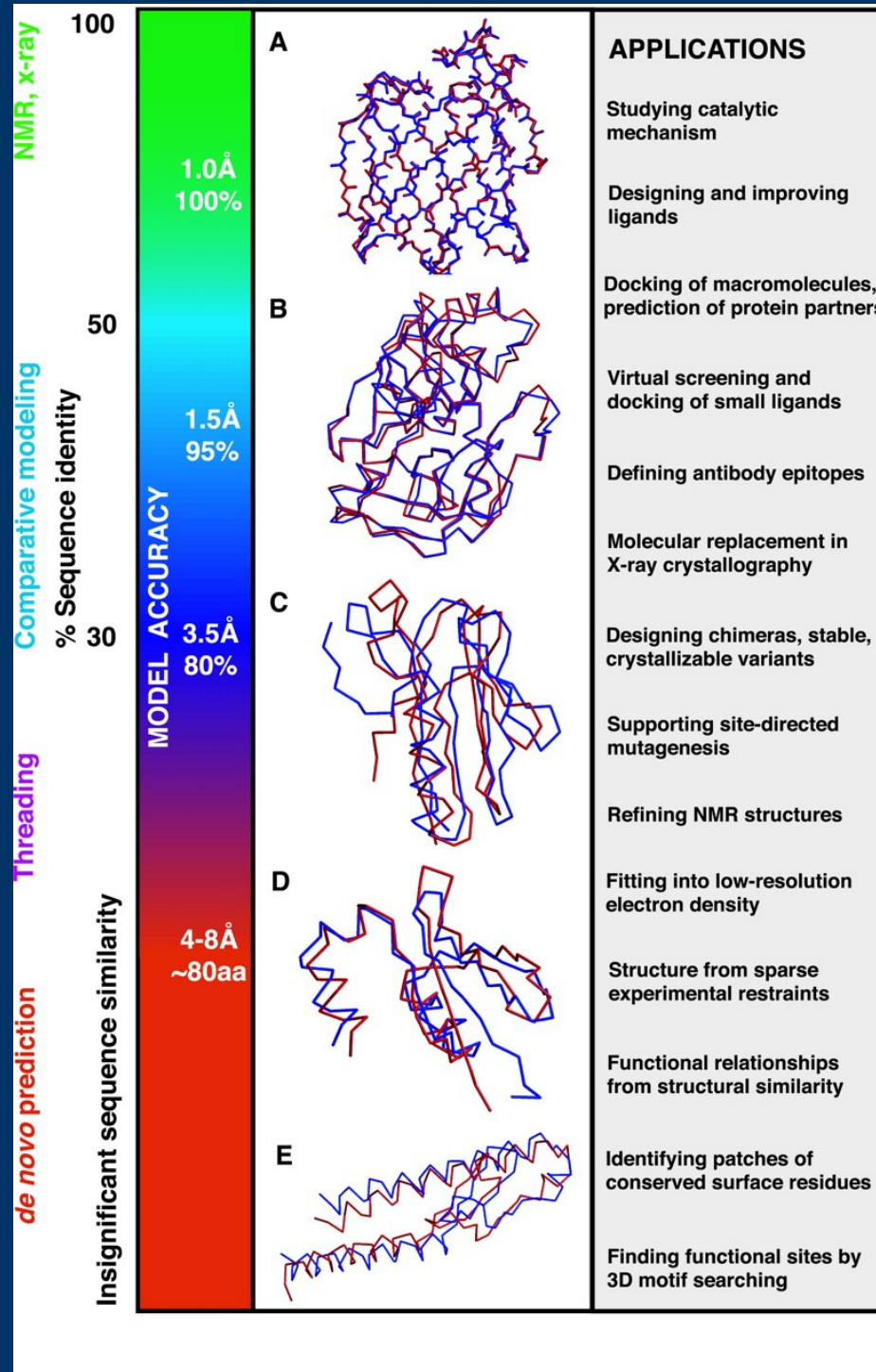
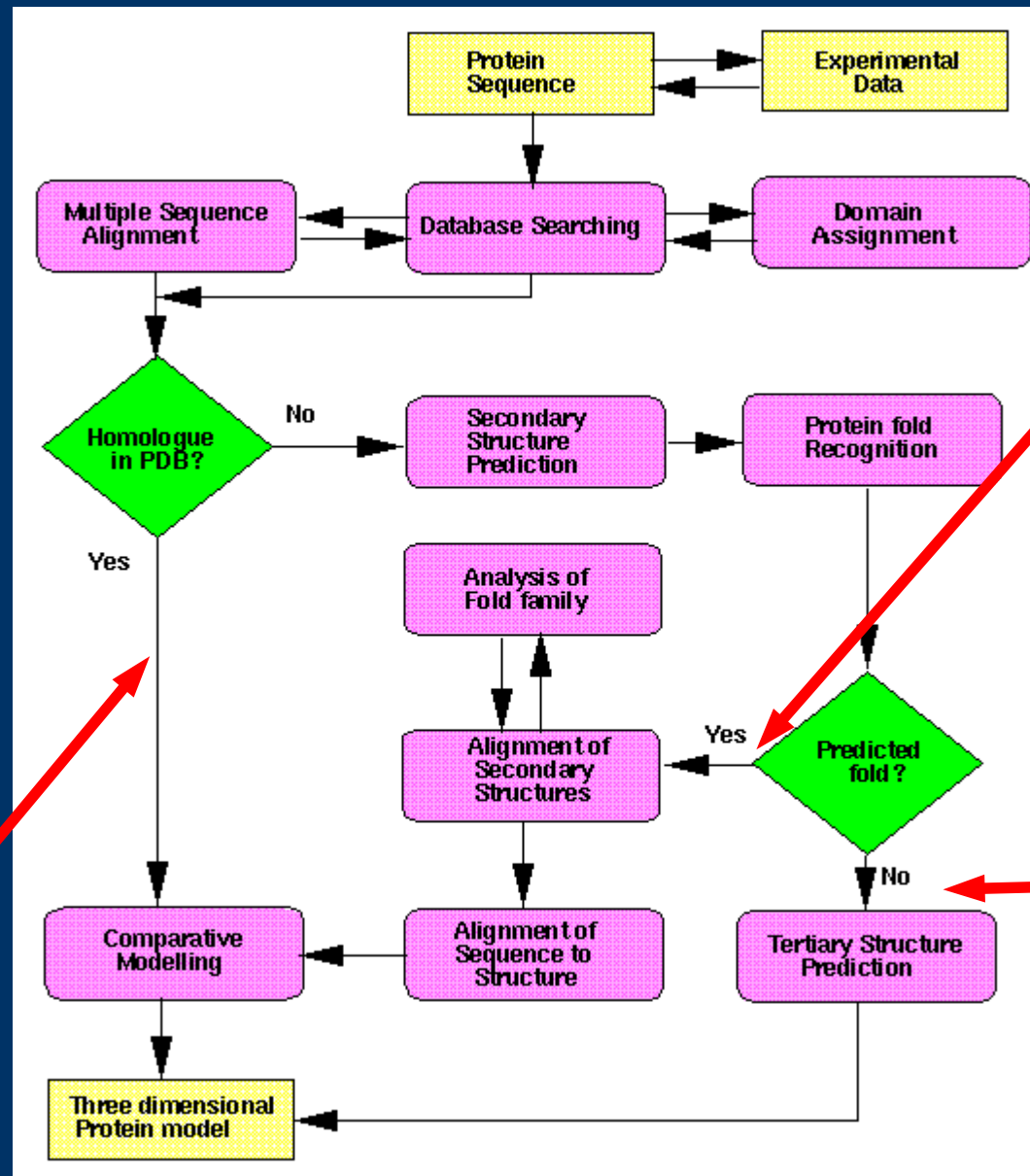


Diagramma di flusso (R.Russell, EMBL)



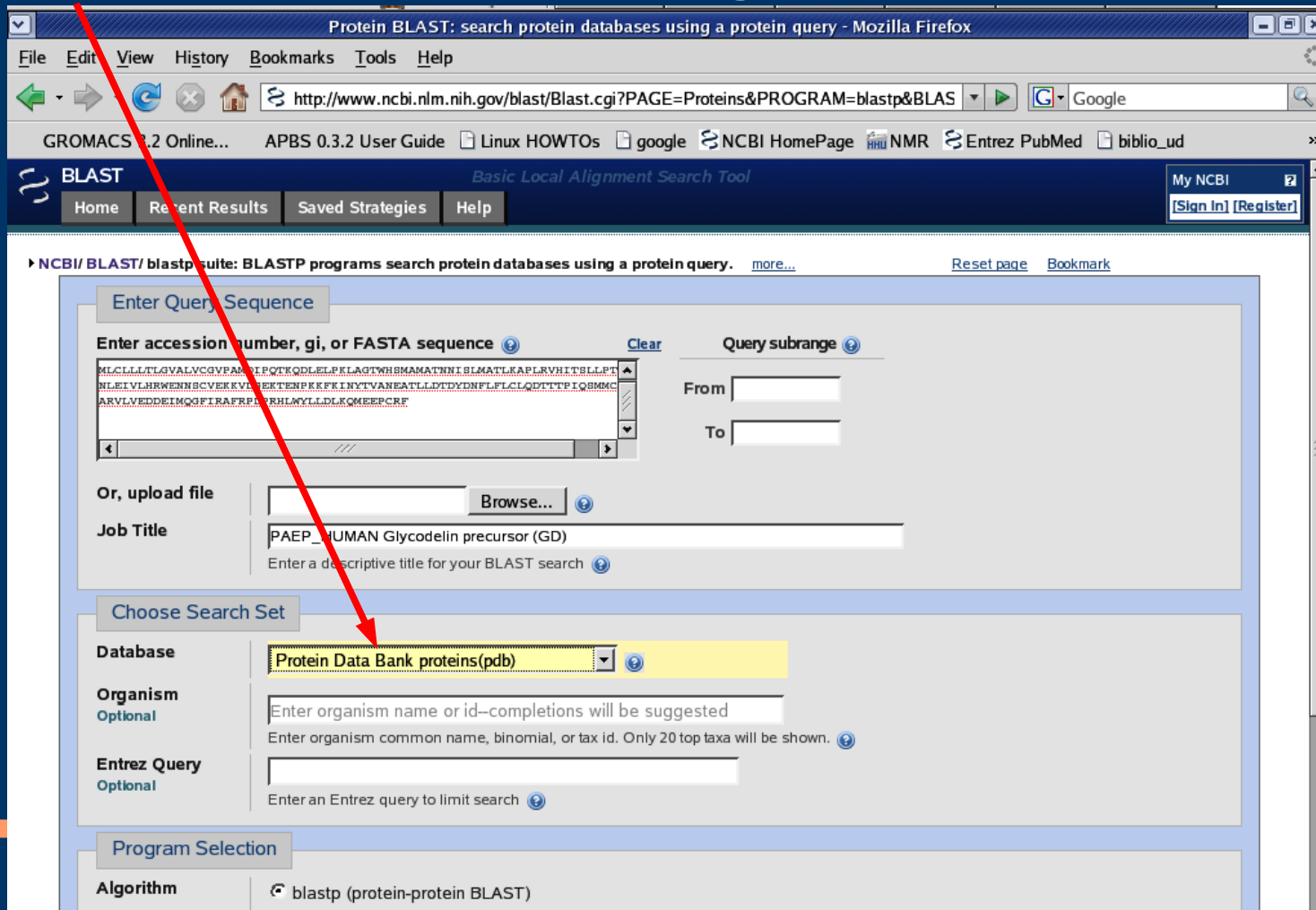
riconoscimento di fold

modellistica per omologia

modellistica ab initio

Comparative (homology) modeling

- Devo trovare con allineamento una struttura stampo (template)
- Eseguo un Blast selezionando la banca dati PDB su ncbi.nih.nlm.gov



The screenshot shows the NCBI BLAST web interface in a Mozilla Firefox browser. The browser's address bar displays the URL: `http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAS`. The page title is "Protein BLAST: search protein databases using a protein query - Mozilla Firefox".

The main content area is titled "NCBI/BLAST/blastp suite: BLASTP programs search protein databases using a protein query." and includes navigation links for "Reset page" and "Bookmark".

The interface is divided into several sections:

- Enter Query Sequence:** A text area containing a protein sequence: `MLCLLLTQVALVCGVPA...PQFKQDLELPKLAGIWHSMAMATNII...LMATLKAPLRVHITSLLEPT...NLEIVLHRWENNBCVEK...EKTENPKKFKIN...YVANEATLLD...YDNFLFLCLQDITTF...IQSMKC...ARVLVEDDEIMQGF...RAFR...RHLWYLLDLKQMEPCRF`. A "Clear" button is next to it. To the right, there are "Query subrange" fields for "From" and "To".
- Or, upload file:** A "Browse..." button for uploading a file.
- Job Title:** A text field containing "PAEP_HUMAN Glycodelin precursor (GD)" and a prompt to "Enter a descriptive title for your BLAST search".
- Choose Search Set:** A section with three main options:
 - Database:** A dropdown menu with "Protein Data Bank proteins(pdb)" selected.
 - Organism (Optional):** A text field with the prompt "Enter organism name or id--completions will be suggested" and a note: "Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown."
 - Entrez Query (Optional):** A text field with the prompt "Enter an Entrez query to limit search".
- Program Selection:** A section with "Algorithm" set to "blastp (protein-protein BLAST)".

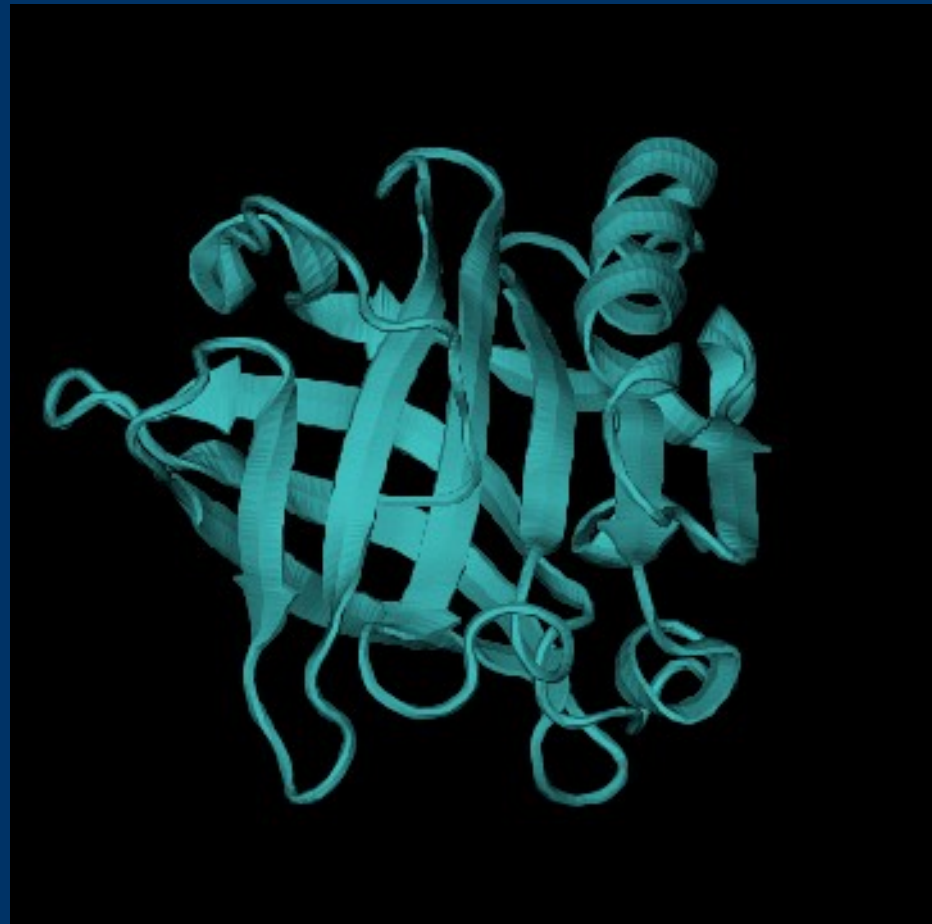
Comparative (homology) modeling

- Allineamento sequenza struttura stampo (template)

PAEP_HUMAN	21	IPQTKQDLELPLKLAGTWHSMAMATNNISLMATLKAPLRVHITSLLPTPED	70
		.: .:.: : : : .:.: :..... :.. . .	
pdb1QG5	3	VTQTMKGLDIQKVAGTWYSLAMAASDISLLDAQSAPLRVYVEELKPTPEG	52
PAEP_HUMAN	71	NLEIVLHRWENNSCVEKKVLGEKTENPKKFKINYTVANEATLLDTDYDNF	120
		: : . : : . : : . : : . :.. :..... :..: .. :	
pdb1QG5	53	DLEILLQKWEDECAQKKIIAEKTKIPAVFKIDALNENKVLVLDTDYKKY	102
PAEP_HUMAN	121	LFLCLQDTHPIQSMCQYLARVLVEDDEIMQGFIRAFRPLPRHLWYLLD	170
		.. : : : : : . . : : : . . : . :. . : :	
pdb1QG5	103	LLFCMENSAEPEQSLVCQCLVRTPEVDDEALEKFDKALKALPMHIRLSFN	152
PAEP_HUMAN	171	LKQMEEPC	178
		.. : .	
pdb1QG5	153	PTQLEEQC	160

Comparative (homology) modeling

- Copia delle coordinate del backbone, generazione delle catene laterali non identiche (il metodo piu' diffuso utilizza una libreria di rotameri)



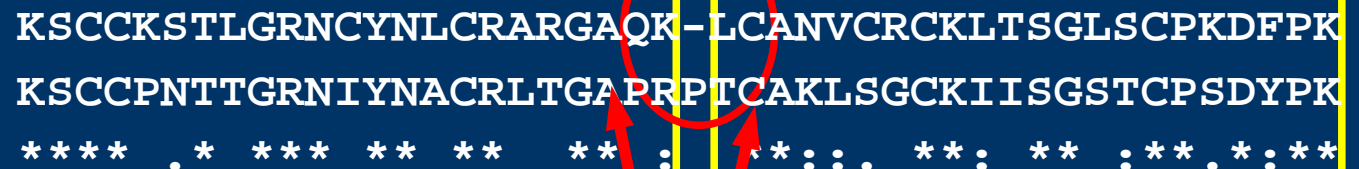
Comparative (homology) modeling

- Un problema e' costituito dai gaps che occorrono principalmente nei loops.

purothionin

viscotoxin

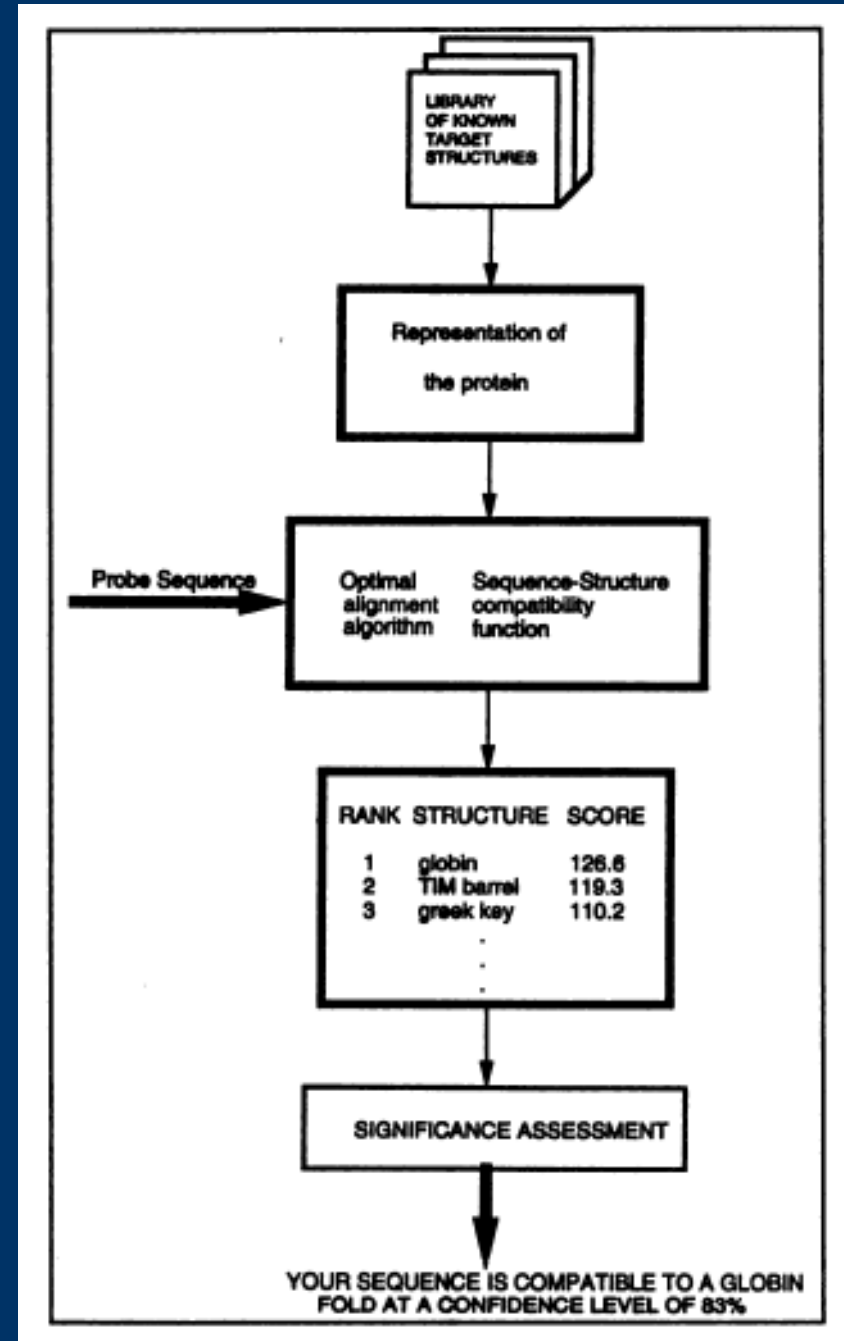
```
KSCCKSTLGRNCYNLCRARGAOK-LCANVCRCKLTSGLSKDFPK  
KSCCPNTTGRNIYNACRLTGAPRPTCAKLSGCKIISGSTCPSDYPK  
**** . * *** ** ** * : * : : . * : * : * : * : *
```



- In questi casi cerco di allineare le sequenze in modo che i gap non occorrono in elementi di struttura secondaria
 - considero le parti attorno al gap piu' simili.
 - Fisso un residuo prima e uno dopo il loop (questi residui sono detti ancore)
 - Costruisco il loop o con algoritmi specifici o copiando le coordinate del backbone da una libreria di frammenti
-
-

Fold recognition

- Se non trovo un omologo in PDB cerco di capire se la sequenza puo' assumere un fold noto
- L'allineamento qui e' sequenza vs. struttura
- Qual e' il fold piu' probabile?



Fold recognition

- Rappresento l'ambiente di ogni amminoacido in una data struttura o in un dato set di strutture allineate con dei parametri: ad es.
 - La struttura secondaria
 - L'accessibilita' al solvente
 - L'idrofobicità degli amminoacidi vicini
-
-

Fold recognition

- Questo mi permette di definire per ogni ammino acido la compatibilità dell'ammino acido con quella posizione nella struttura
 - Questo corrisponde ad un profilo di punteggi che posso usare per un allineamento ad esempio mediante dynamic programming
-
-

Fold recognition 3D-1D profile

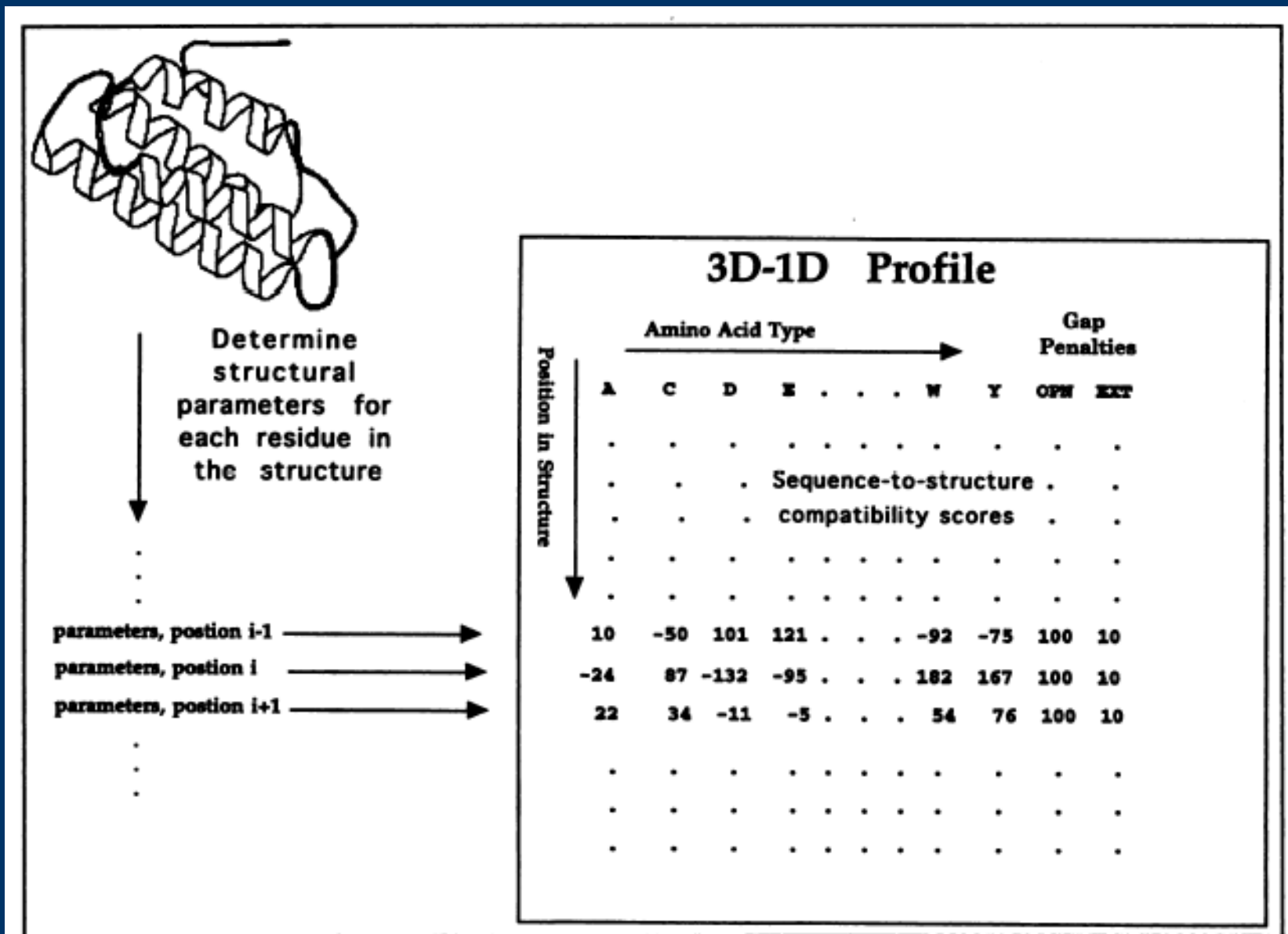


Figure 2. The use of 3D-1D profiles in fold recognition (4). The representation of the protein and a unipositional compatibility function can be combined into a convenient data structure called a *profile*. Using a profile, the application of a dynamic programming algorithm is straightforward.

Ab-initio modeling

E' possibile costruire il fold di una proteina a partire da frammenti in banca dati o da modelli semplificati con tecniche di simulazione di dinamica molecolare o con metodi probabilistici Monte Carlo

Tipicamente si segue un approccio gerarchico:

Modelli semplificati nella fase esplorativa e poi raffinamento con modelli che considerano tutti gli atomi



Risultati recenti ab-initio

ID	L	%a	%b	Round 1	Round 2	Cluster
1b72A	49	69	0	0.8 (0.8)	1.1 (0.9)	1.0 Hox-B1 protein
1shfA	59	5	40	11.1 (9.0)	10.8 (8.5)	10.9 Fyn tyrosine kinase
1tif_	59	22	37	5.3 (2.3)	4.1 (2.8)	3.8 IF3-N
2reb_2	60	61	20	1.2 (0.9)	2.1 (1.6)	1.3 RecA
1r69_	61	63	0	2.1 (2.4)	1.2 (1.5)	1.7 434 repressor
1csp_	67	4	53	5.1 (4.5)	4.7 (4.2)	5.1 Cold-shock protein
1di2A_	69	46	33	2.6 (2.3)	2.6 (2.2)	1.9 RNA binding protein A
1n0uA4	69	43	24	9.9 (8.3)	10.2 (8.1)	2.7 Elongation factor 2
1mla_2	70	34	37	8.4 (7.3)	8.7 (8.1)	7.2 Malonyl-CoA ACP transacylase
1af7__	72	72	0	10.1 (7.9)	10.4 (8.1)	1.7 Cher domain 1
1ogwA_	72	26	33	2.7 (2.3)	1.0 (1.0)	2.6 Ubiquitin
1dcjA_	73	31	27	3.2 (2.2)	2.5 (2.4)	2.0 Yhhp
1dtjA_	74	39	27	1.0 (0.8)	1.2 (0.9)	1.8 KH domain of Nova-2
1o2fB_	77	38	27	10.1 (8.7)	N/A	10.3 Glucose-permease IIBC
1mkyA3	81	32	24	3.2 (3.6)	6.3 (6.1)	3.7 Enga
1tig_	88	35	35	4.1 (4.2)	3.5 (3.4)	2.4 IF3-C

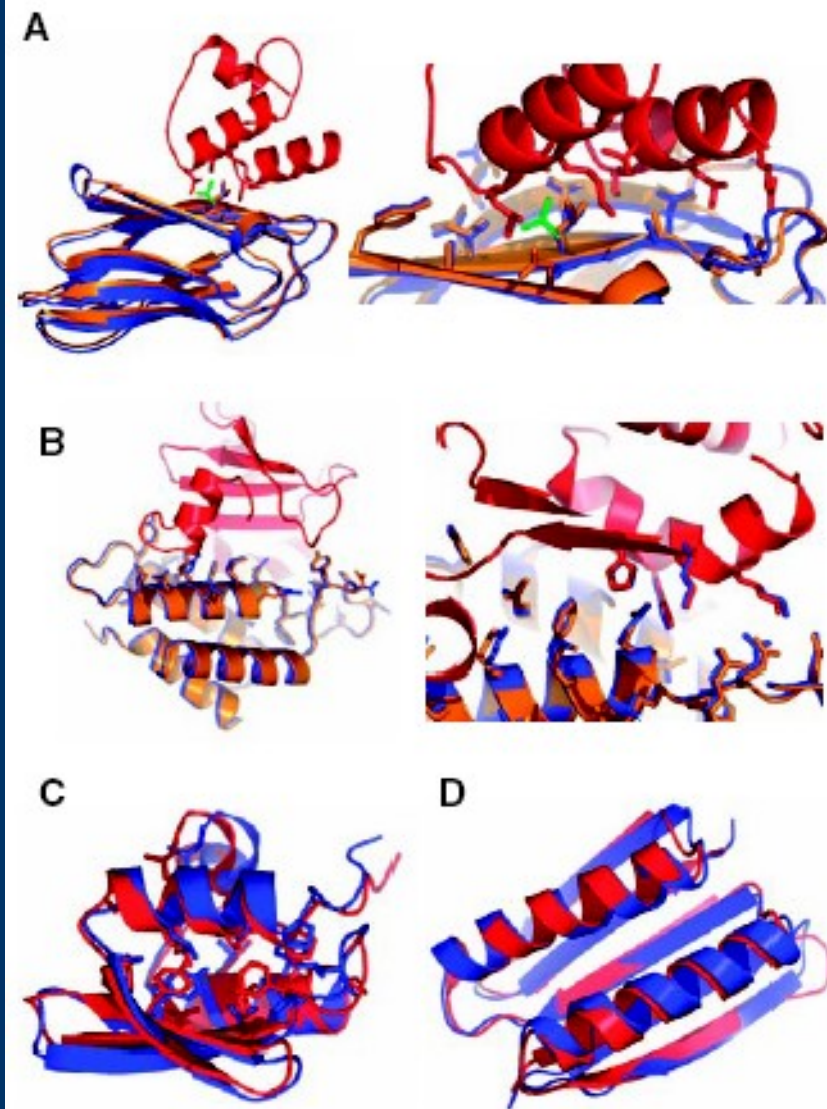


Fig. 3. Examples of high-resolution prediction and design. (A) CAPRI Target 12 [dockerin-cohesin (66); interface residue backbone RMSD = 0.27 Å]. The lowest energy structure in Fig. 2A, main panel, is shown here. The side chain of Leu-83 (green in the free monomer) changes conformation upon binding. Side-chain conformations in red were provided; those in blue were predicted. (B) CAPRI Target 15 [Colicin-Immunity protein D (68); interface residue backbone RMSD = 0.23 Å]. No side-chain information was provided for either partner. (C) CASP6 de novo structure prediction Target 0281 [hypothetical protein from *Thermus thermophilus* Hb8, PDB ID 1whz (69); backbone RMSD = 1.59 Å]. (D) TOP7 (RMSD = 1.2 Å) (42). (A) and (B) are adapted from figure 1 of (70). Blue: models; red and orange: x-ray structures.

Statistical effective energy functions (SEEFs)

- Potenziali derivati dalle distribuzioni osservate nella PDB.

$$W = -kT \log \left(\frac{N_{\text{observed}}}{N_{\text{expected}}} \right)$$

- Spesso usate assieme ad una rappresentazione semplificata (ad es. 2 sfere per ammino acido)
- Valutazione rapida dell'energia, dinamica semplificata

Physical effective energy functions (PEEFs)

- L'energia e' la somma di un termine dovuto al soluto ed un termine dovuto alla solvatazione

$$W = U(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) + W^{solvation}$$

- L'energia di solvatazione tiene conto degli effetti elettrostatici e idrofobici:

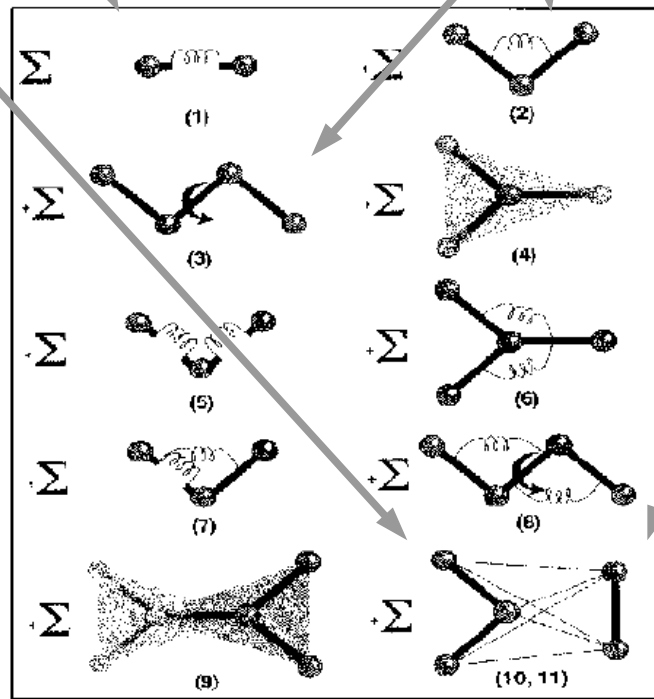
$$W^{solvation} = W^{polar}(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) + \gamma SASA$$

Campi di forza

- Charmm, Gromos, Amber,

$$U = \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{torsions}} k_\phi (1 + \cos(n(\phi - \phi_0)))$$

$$+ \sum_{\text{atom-pairs}} V_{ij} \left(\left(\frac{r_0}{r} \right)^{12} - \left(\frac{r_0}{r} \right)^6 \right) + \sum_{\text{atom-pairs}} \frac{q_i q_j}{\epsilon r} + \text{other terms}$$



Raffinamento (refinement)

- Il raffinamento di un modello si fa usando tecniche di simulazione di dinamica molecolare o di minimizzazione dell'energia per rilassare le zone non modellate bene e cercando la conformazione ad energia piu' bassa
 - Oppure ricostruendo le parti problematiche della struttura
-
-