BIOLOGIA STRUTTURALE PER IL DRUG DESIGN

STRUTTURE E INTERAZIONI TRA BIOMOLECOLE

Andrea Ilari

IBPM CNR

National Research Council of Italy

Programma dettagliato del corso

1. A. Struttura delle proteine.
   B. Relazione struttura-attività nelle proteine alcuni esempi:

- Le cage-like proteins: Ferritine e Dps (DNA binding proteins from starved cells)
- Sorcin (soluble resistance related-calcium binding proteins)
- Proteine di SARS-COV2

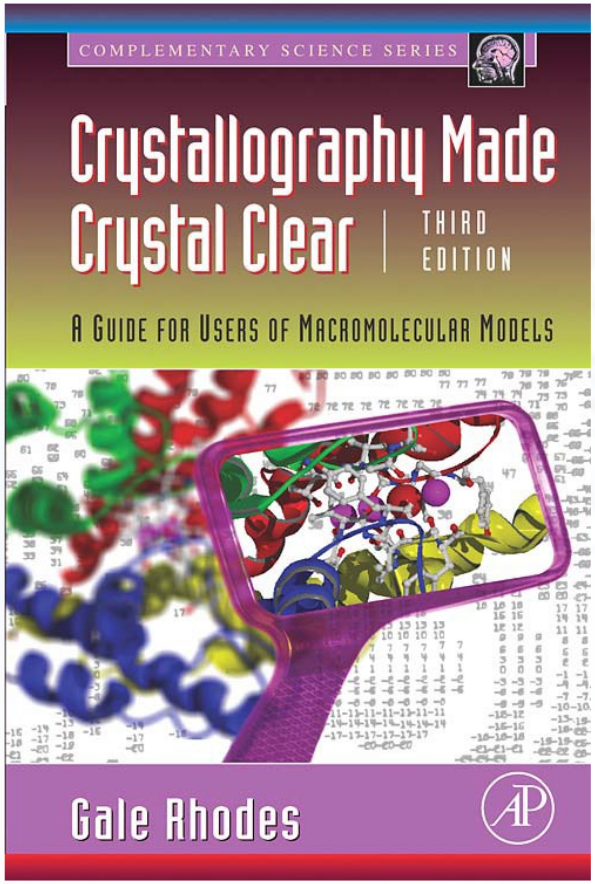
2. Ordine e disordine nelle proteine
3. Metodi per la determinazione delle struttura tridimensionale delle proteine: NMR e CryoEM
4. La cristallografia ai raggi X: la cristallizzazione delle proteine, i cristalli e le loro simmetrie
5. La cristallografia ai raggi X: la diffrazione, la legge di Bragg, lo spazio reciproco, diffrattometri ai raggi X, luce di sincrotrone
6. Prima esercitazione pratica: la cristallizzazione del lisozima.
7. La matematica della cristallografia: la trasformata di Fourier
8. La cristallografia ai raggi X, determinazione della struttura tridimensionale: come si risolve il problema della fase (la soluzione della struttura)
9. Seconda esercitazione pratica: la visualizzazione dei cristalli
10. Analisi strutturale. Come posso usare le informazioni che ottengo dalla struttura macromolecolare.

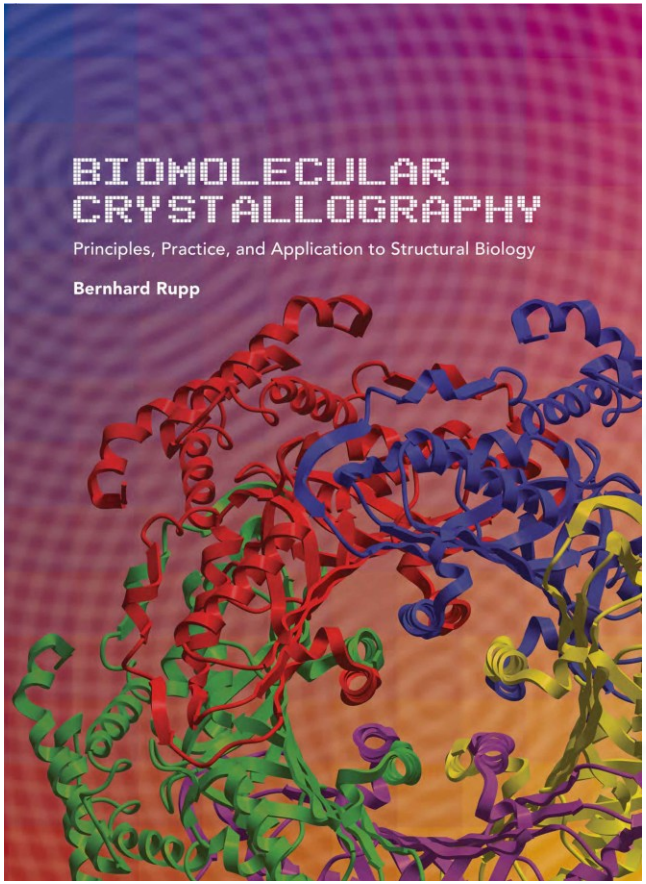
Programma dettagliato del corso

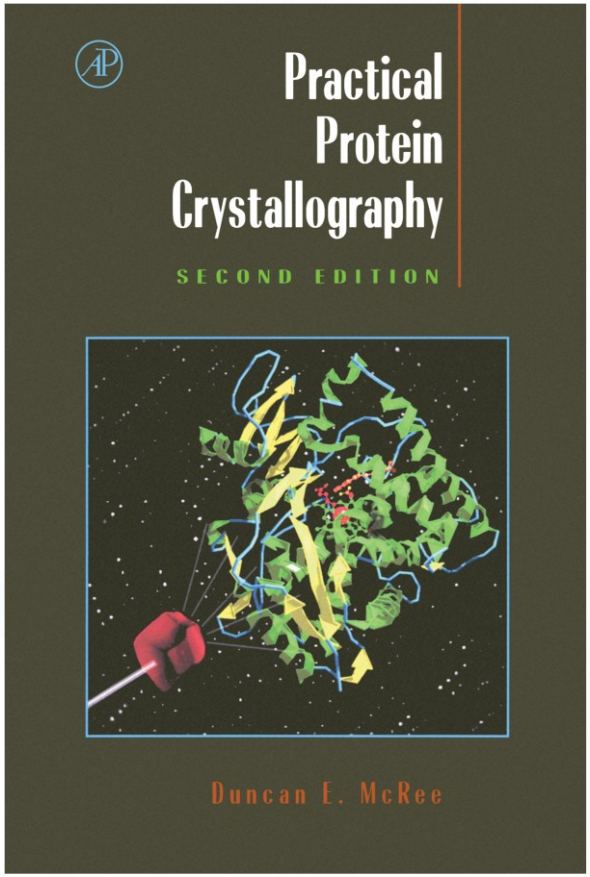
11. Structure-based drug design

12. Come la struttura può servire a progettare nuovi farmaci, un esempio concreto:

La tripanotione reduttasi


COMPLEMENTARY SCIENCE SERIES
Crystallography Made Crystal Clear
THIRD EDITION
A Guide for Users of Macromolecular Models
Gale Rhodes


BIOMOLECULAR CRYSTALLOGRAPHY
Principles, Practice, and Application to Structural Biology
Bernhard Rupp

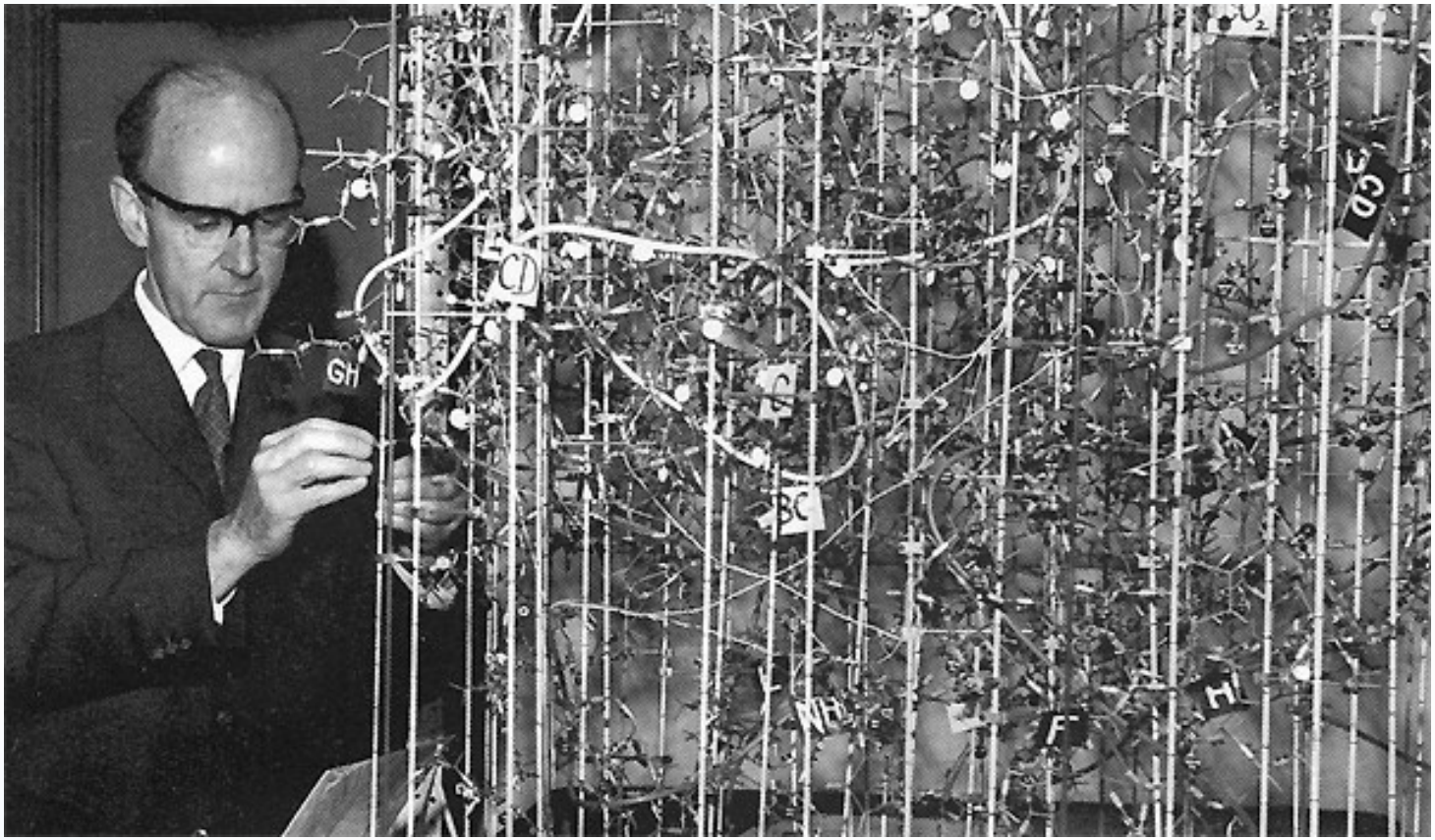

Practical Protein Crystallography
SECOND EDITION
Duncan E. McRee

Lezione 1

Relazione struttura attività nelle proteine

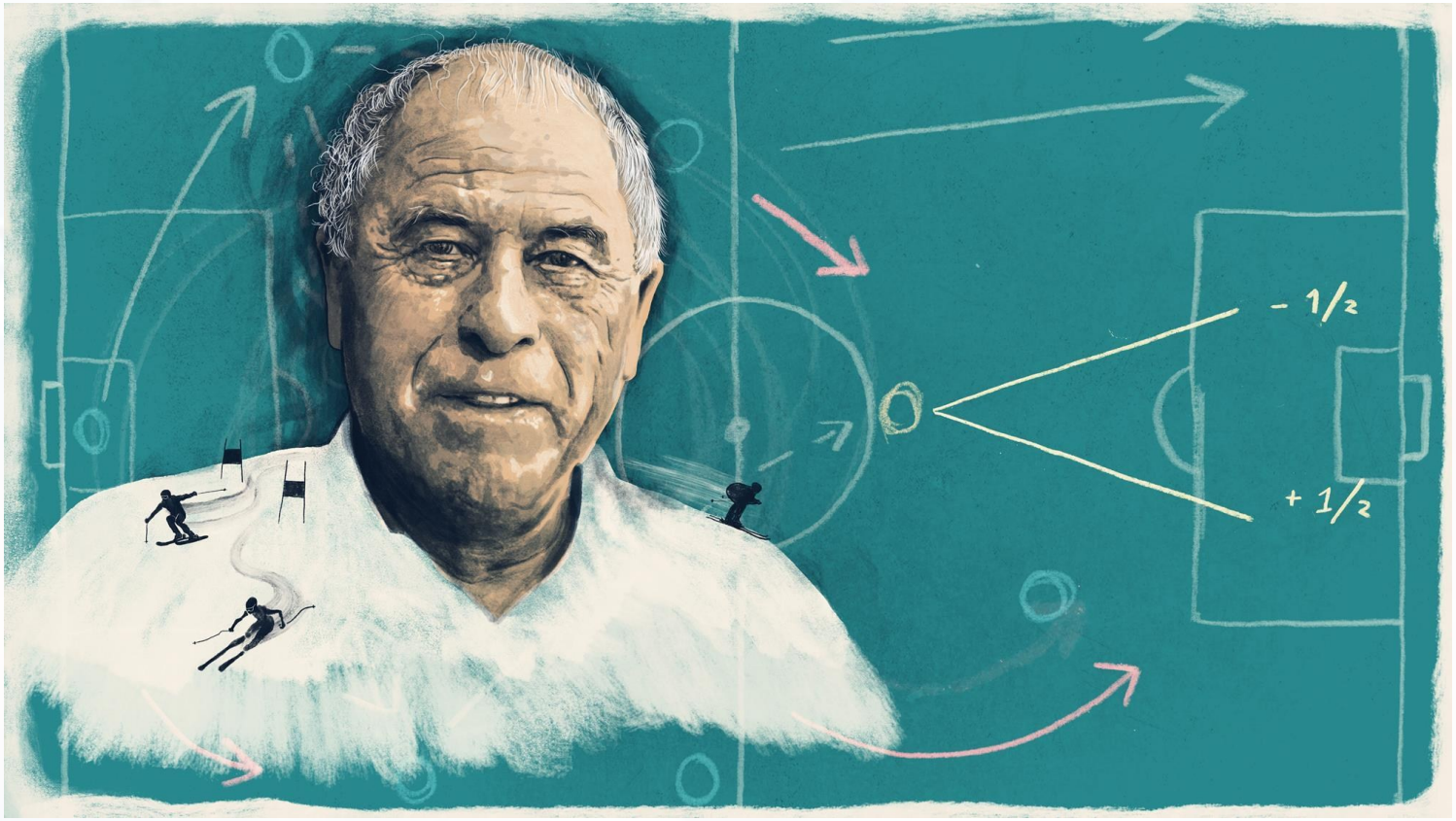
IBPM CNR
National Research Council of Italy

Max Perutz e John Kendrew: a nascita della biologia strutturale e delle moderna Biochimica (premi nobel per la chimica nel 1962)



Struttura dell'Emoglobina

Kurt Wüthrich; NMR to determine the protein structure



for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution.

RISOLUZIONE

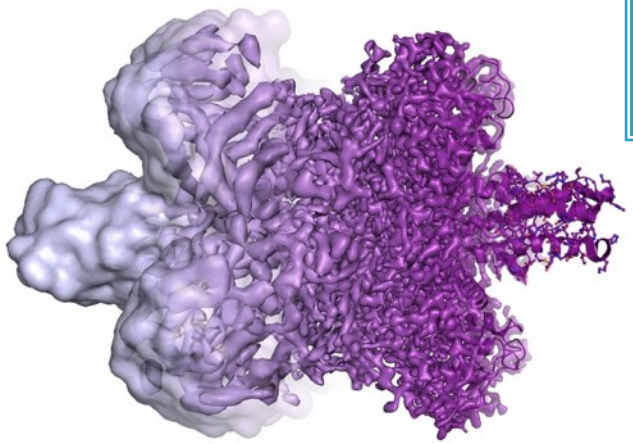
1–2 Å

2–6 Å

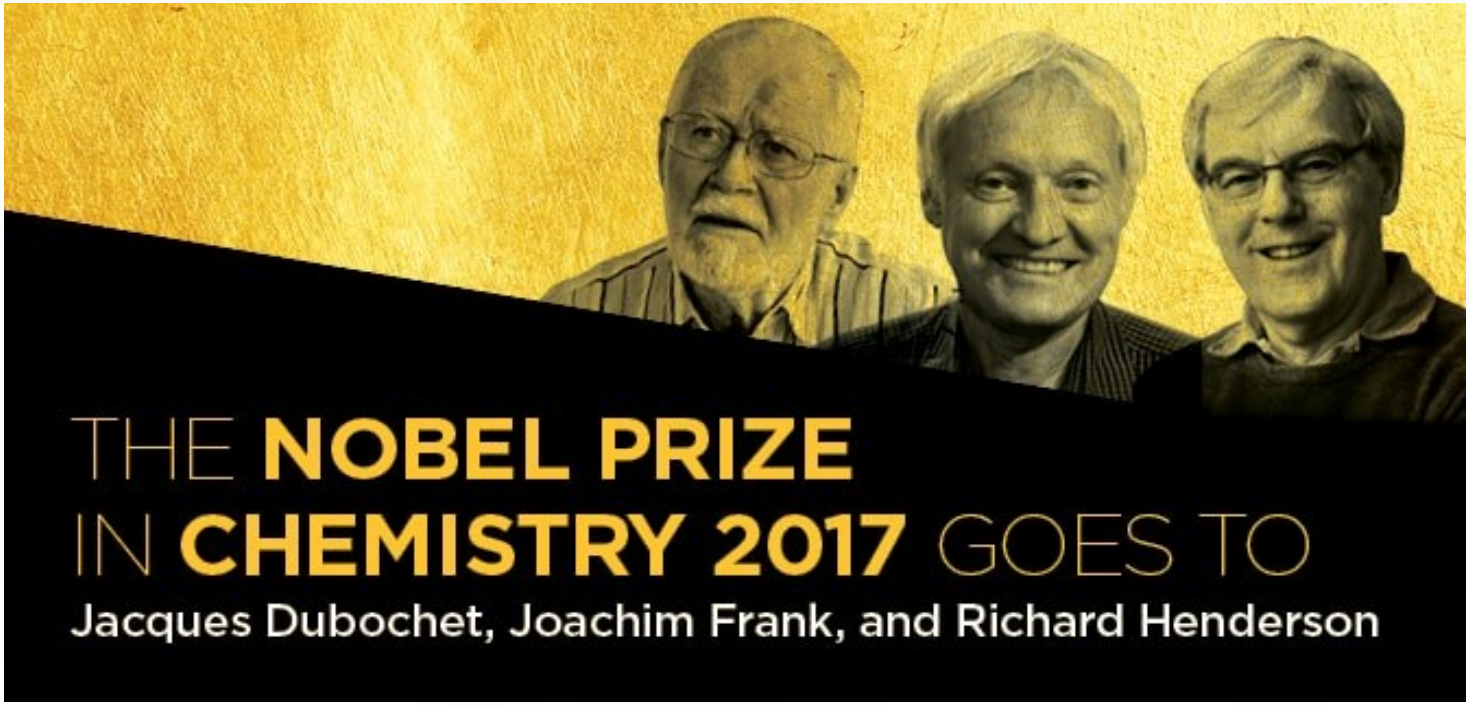
10–20 Å

Congelamento (cryo-EM)

2013

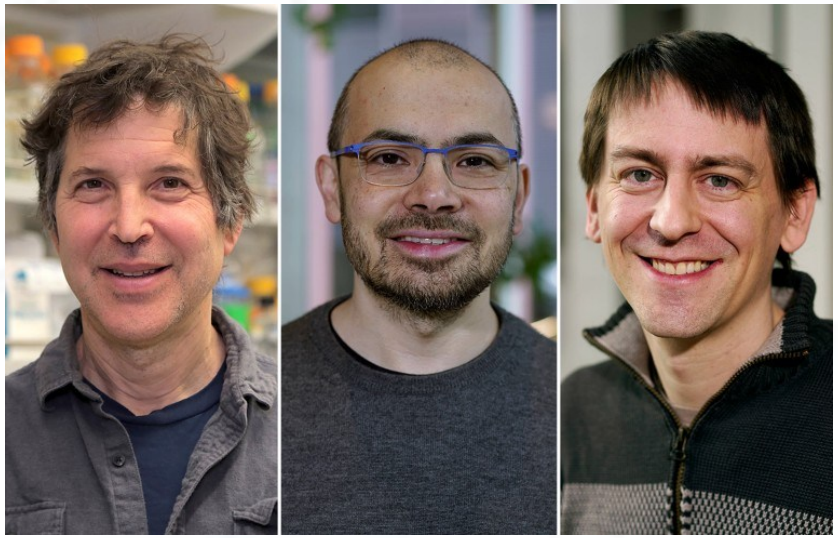


2017



https://www.youtube.com/watch?v=026rzTXb1zw

Chemistry Nobel goes to developers of AlphaFold
AI that predicts protein structures



The 2024 chemistry Nobel was awarded to **John Jumper and Demis Hassabis at Google DeepMind** in London, for developing a game-changing AI tool for predicting protein structures called AlphaFold, and **David Baker, at the University of Washington in Seattle, for his work on computational protein design, which has been bolstered by AI in the past few years.**

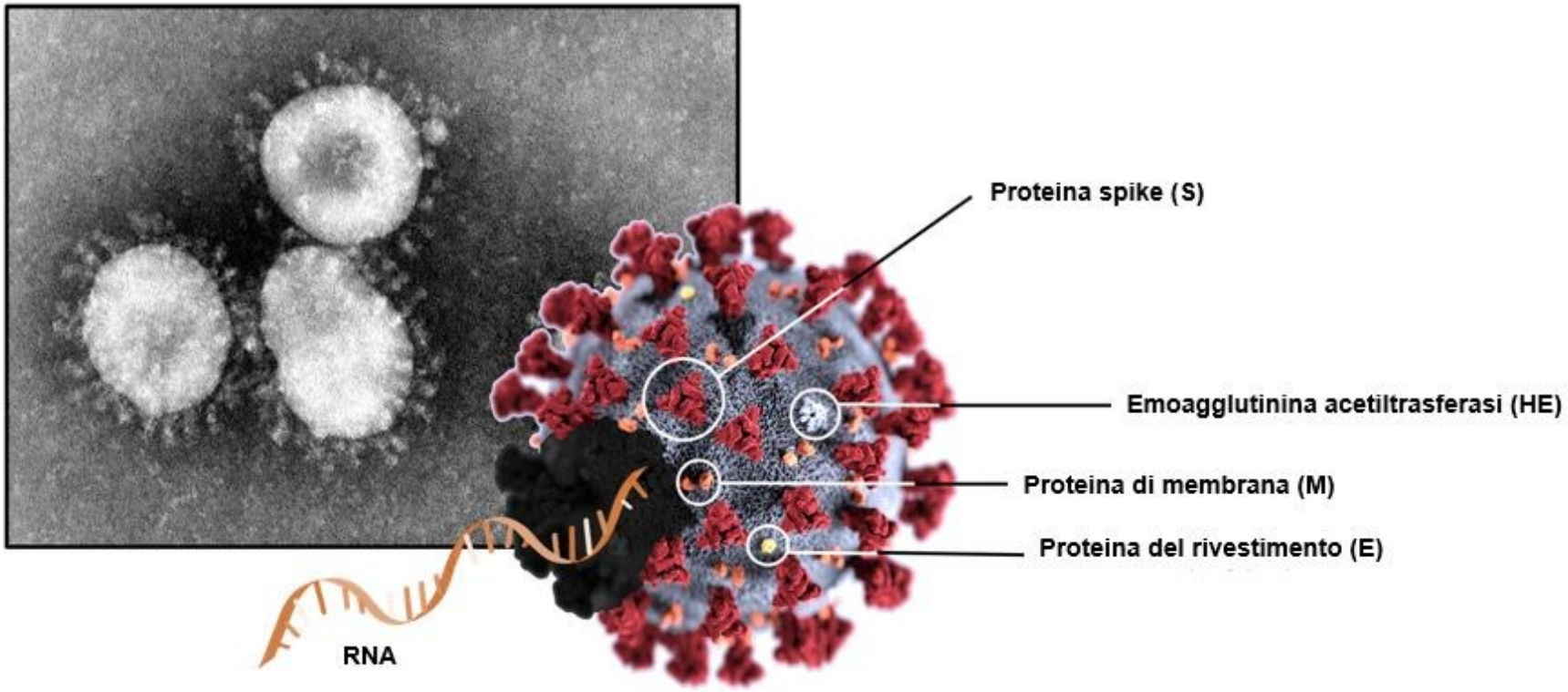


Immagine al microscopio elettronico a trasmissione del coronavirus (fotografia in alto a sinistra). Le particelle virali appaiono bianche circondate da un fondo nero. Sono visibili le proteine Spike che “decorano” la superficie e che conferiscono al virione la caratteristica forma a corona. La ricostruzione grafica al centro dell’immagine mostra l’architettura molecolare del virus SARS-CoV-2. Il capside virale *(in grigio),* che contiene l’RNA a singolo filamento associato alle proteine del nucleocapside, è costituito da un doppio strato lipidico in cui sono inserite le proteine di superficie (S *in rosso*; E *in giallo*; M *in arancio*; HE *in bianco*). [Fotografia in alto a destra: CDC/Dr. Fred Murphy; Sylvia Whitfield; disegno in centro adattato da: CDC/Alissa Eckert, MSMI, Dan Higgins, MAMS]

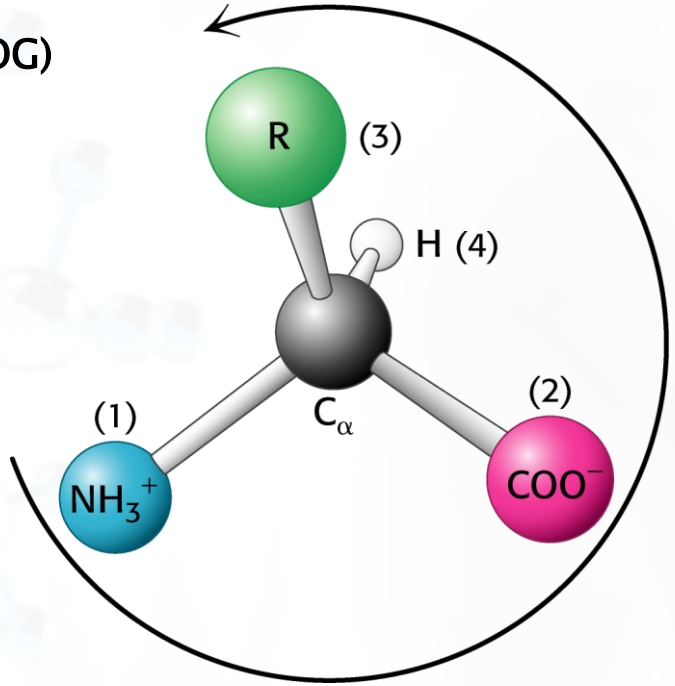
Come rappresentare le molecole chirali:

IL SISTEMA R-S
(CAHN-INGOLD-PRELOG)

Si assegna una priorità ai 3 sostituenti che va col peso atomico:

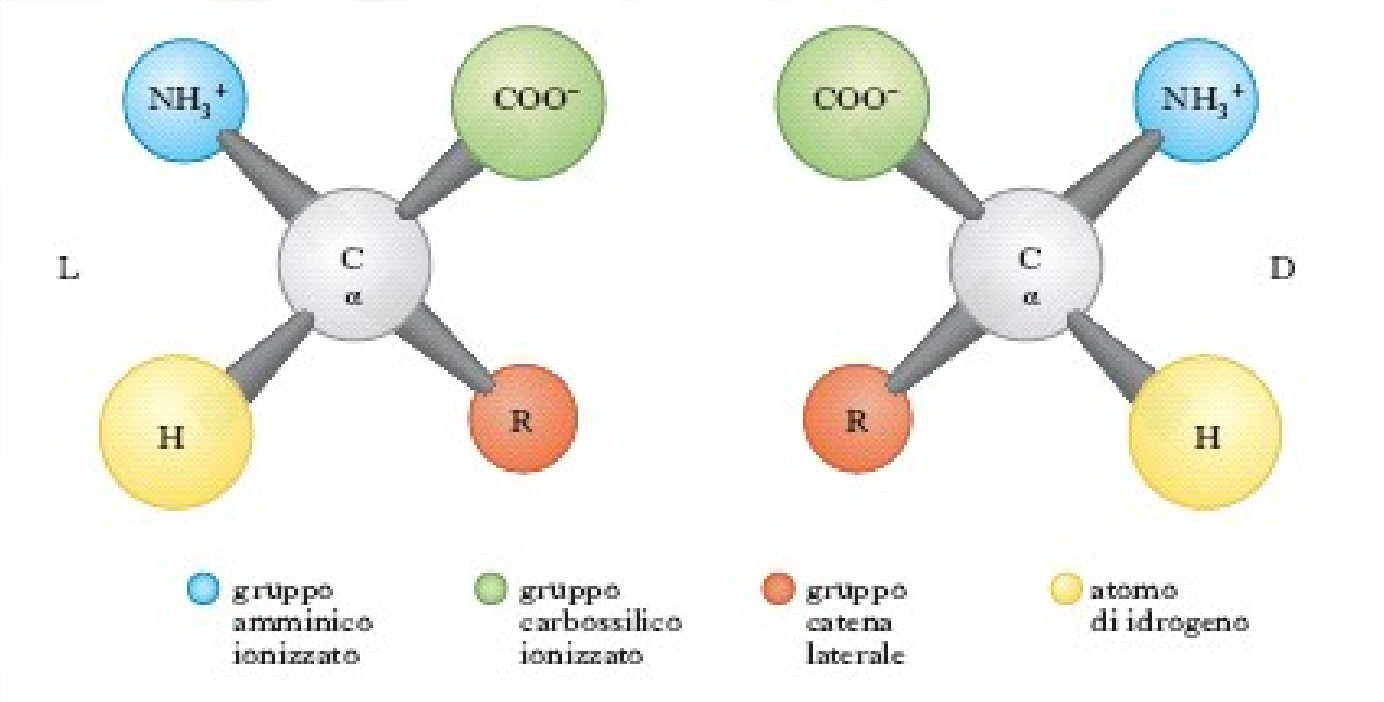
$$SH > OH > NH_2 > COOH > CHO > CH_2OH > C_6H_5 > CH_3 > {}^2H > {}^1H$$

Si orienta la molecola in modo che il sostituente con priorità più bassa (4) sia diretto all'indietro, cioè dietro al piano dell'osservatore (come se andasse "nel foglio").



Osserva l'ordine 1 → 2 → 3Guarda la direzione del percorso dai sostituenti 1 → 2 → 3:Se il movimento è in senso orario, la configurazione è R (rectus).Se il movimento è in senso antiorario, la configurazione è S (sinister).

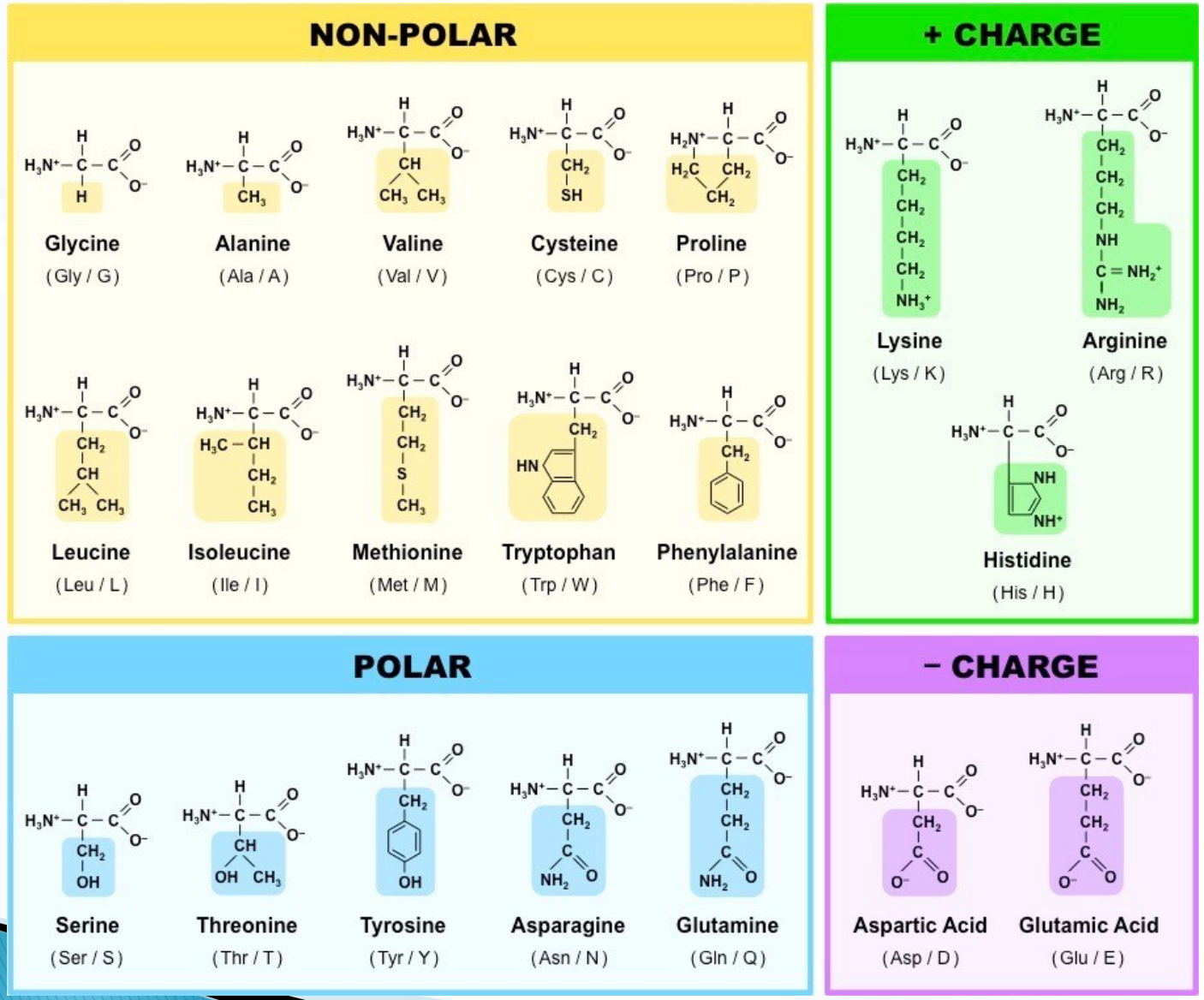
Amminoacidi



D / L → sono convenzioni storiche, nate confrontando la molecola con la gliceraldeide (la molecola di riferimento).

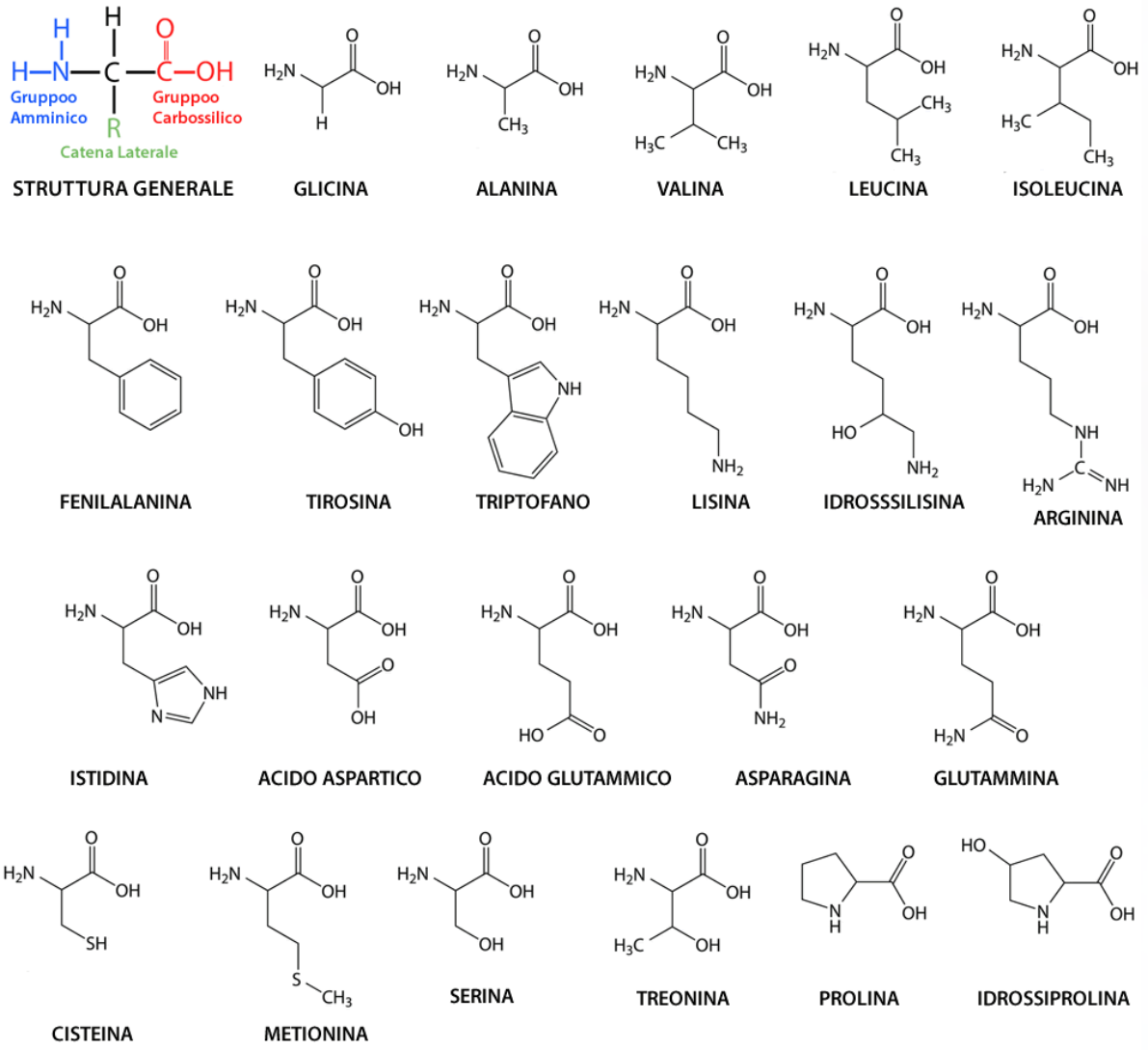
Tutti i 20 L-α-amminoacidi standard hanno carboni α con configurazione *S*, tranne la L-CISTEINA

L e proteine sono formate principalmente da 20 L-amminoacidi



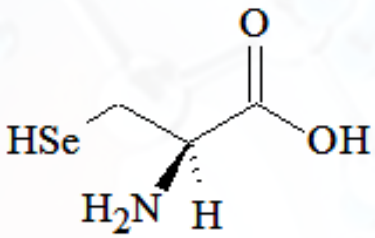
Le strutture degli amminoacidi

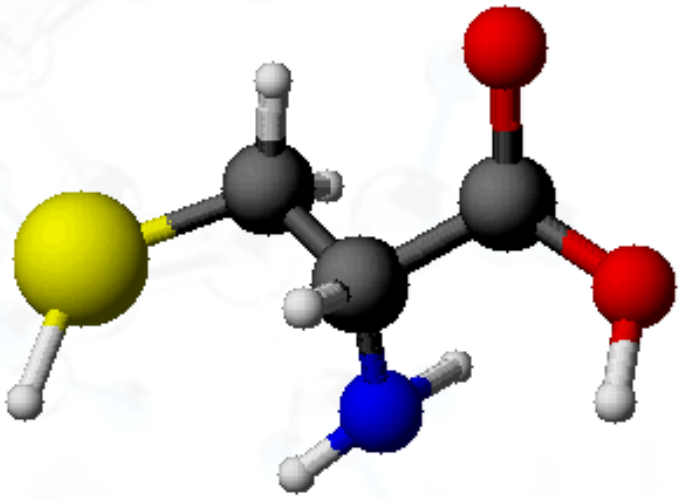
Ai 20 amminoacidi classici
Si aggiungono l'idrossiprolina e l'idrossilisina, ma anche
La selenometionina e la selenocisteina
In cui lo zolfo è sostituido dal selenio



La selenocisteina (il 21° aa)

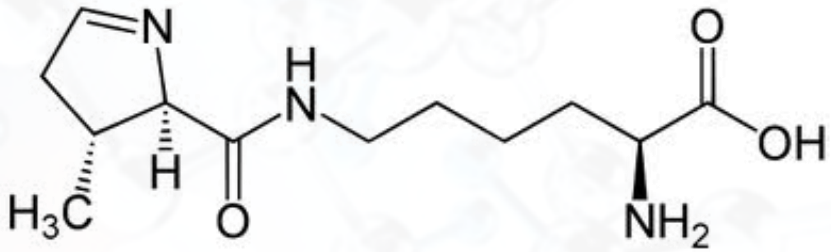
selenoproteine (es. glutatione perossidasi, tioredossina riduttasi)
partecipano a reazioni redox

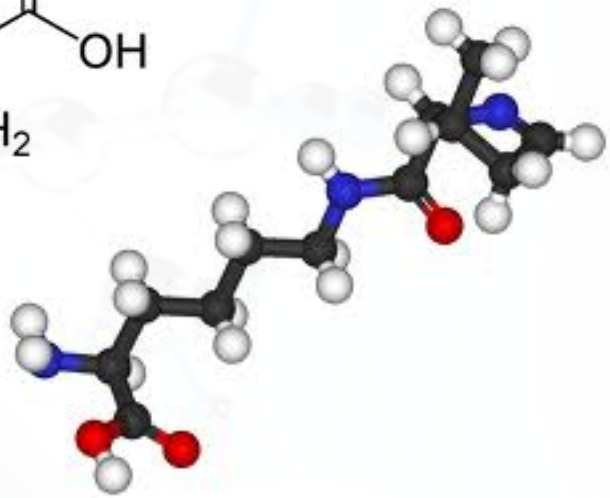




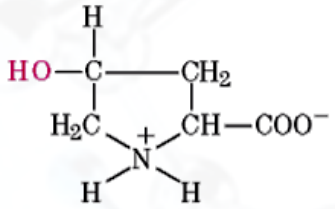
l selenio è più reattivo dello zolfo, quindi la selenocisteina può partecipare più efficacemente a reazioni redox (ossidoriduzioni).

La pirrolisina (proteina degli Archaea, metilammina metiltrasferasi)



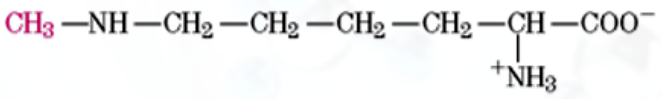


Amminoacidi derivanti da **modificazioni post-traduzionali *irreversibili***



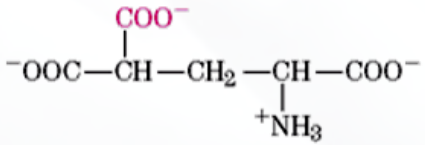
4-Idrossiprolina

Collagene



6-*N*-Metillisina

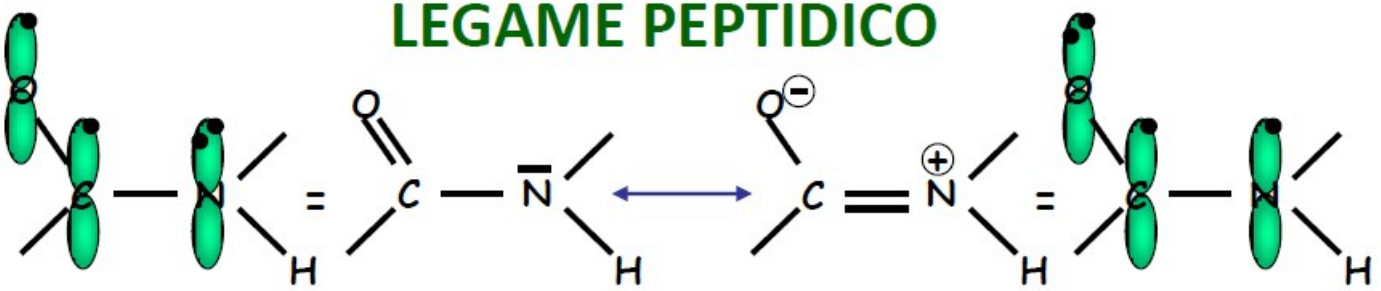
Miosina



γ-Carbossiglutammato

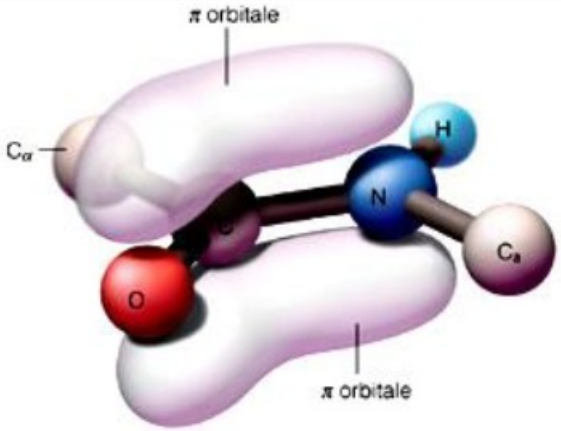
Protrombina

LEGAME PEPTIDICO



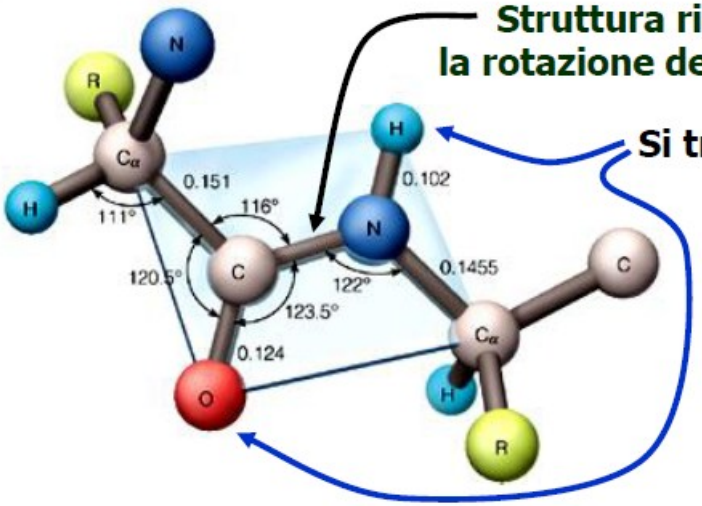
La struttura del legame peptidico è planare in quanto O , C , N e H giacciono sullo stesso piano. Inoltre O , C , N sono ibridizzati sp_2 e quindi ognuno di essi presenta un orbitale p non ibrido. Pertanto il legame peptidico può essere considerato un ibrido di risonanza di due forme.

Il legame peptidico assume per questa ragione un carattere di parziale doppio legame, giustificando la planarità della struttura.

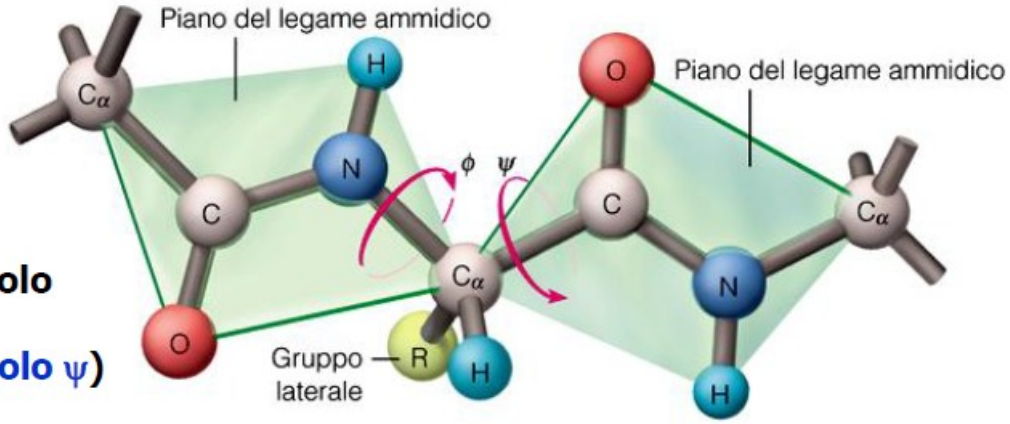


Struttura rigida (non è possibile la rotazione del legame peptidico C-N)

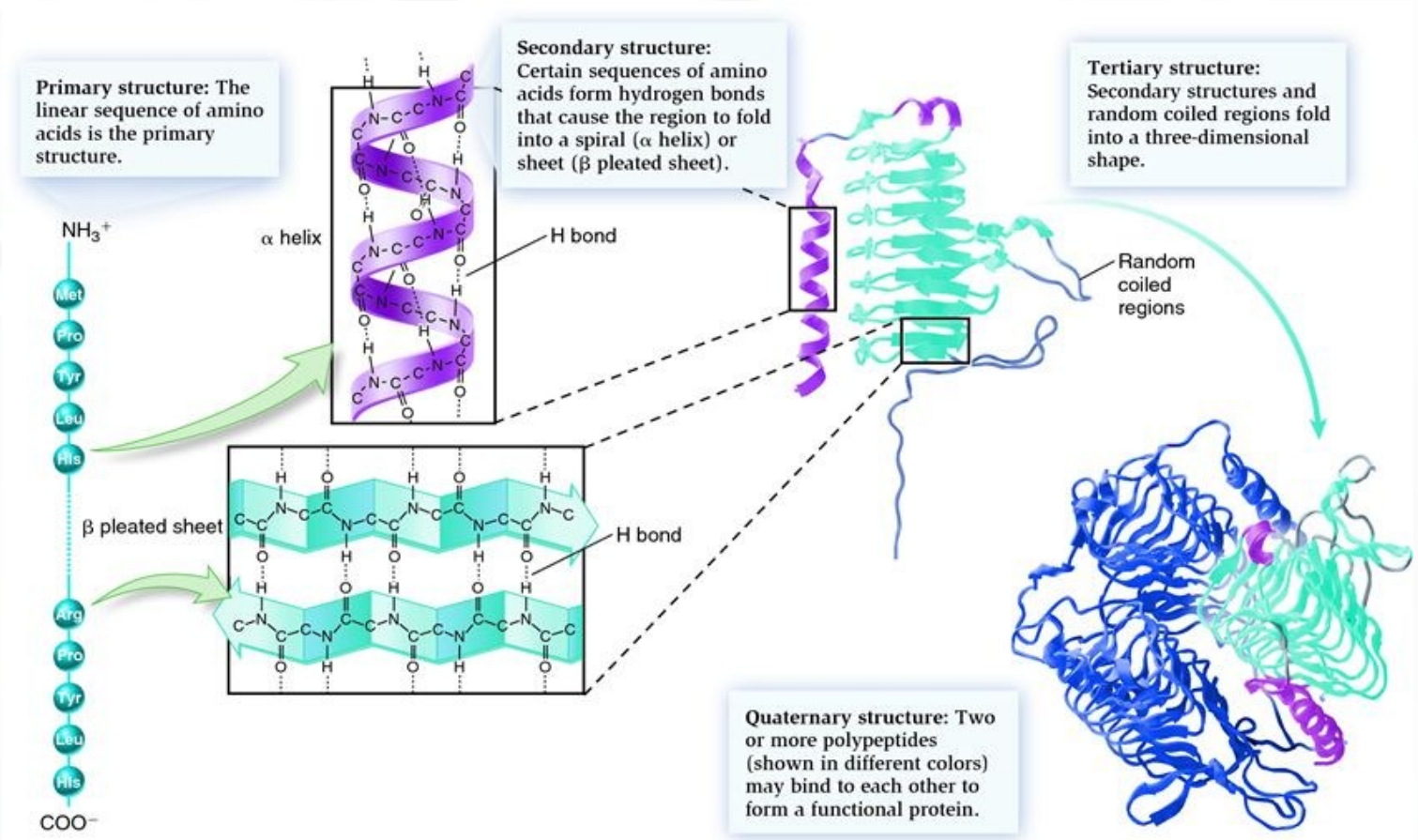
Si trovano in posizione trans



La rotazione è possibile solo intorno ai legami N-C_α (angolo φ) e C_α-C (angolo ψ)



La struttura delle proteine

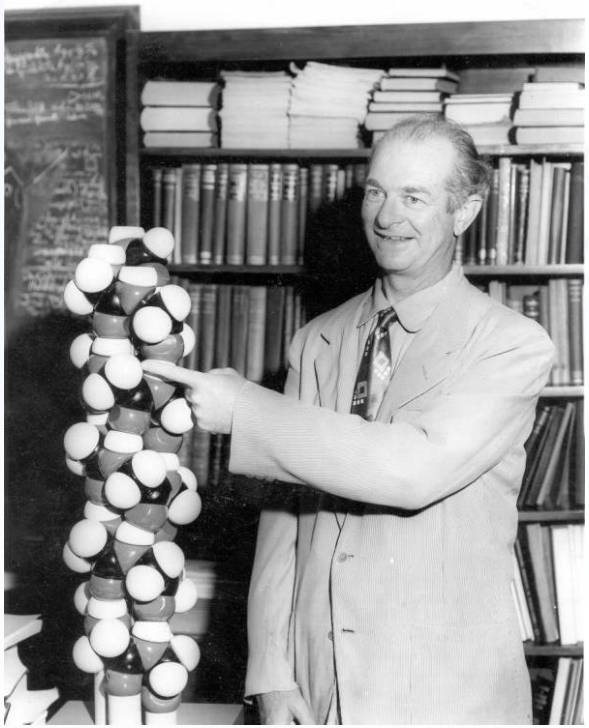


Peptide bond	Average length (Å)	Single bond	Average length (Å)	Hydrogen bond	Average length (Å)
Cα–C	1.525 ± 0.021	C–C	1.540 ± 0.027	O–H - - - O–H	2.8 ± 0.2
C–N	1.329 ± 0.014	C–N	1.489 ± 0.030	N–H - - - O=C	2.9 ± 0.2
N–C	1.458 ± 0.019	C–O	1.420 ± 0.020	O–H - - - O=C	2.8 ± 0.2

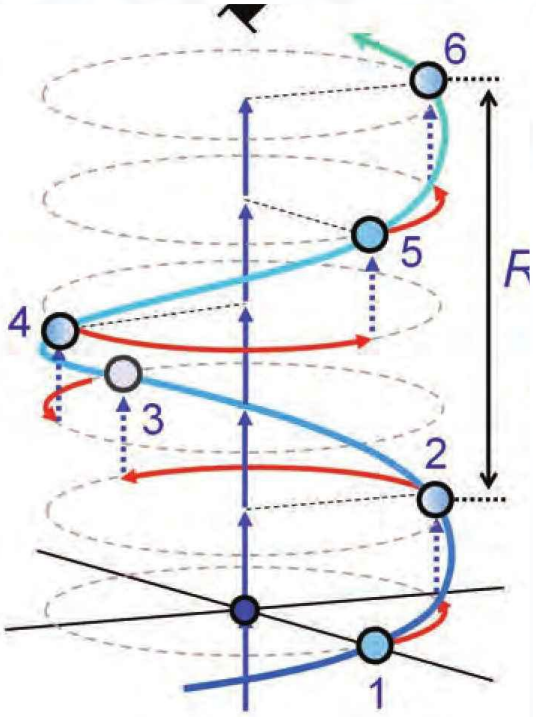
IBPM CNR

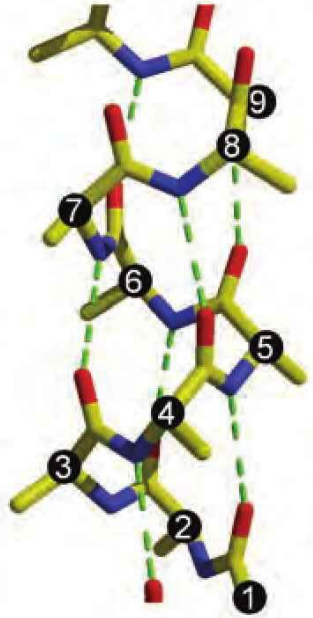
Struttura secondaria: Alfa elica

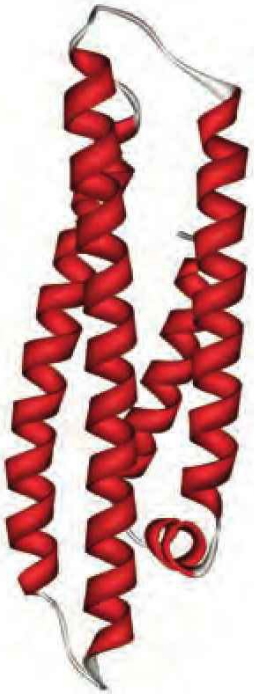
Linus Pauling



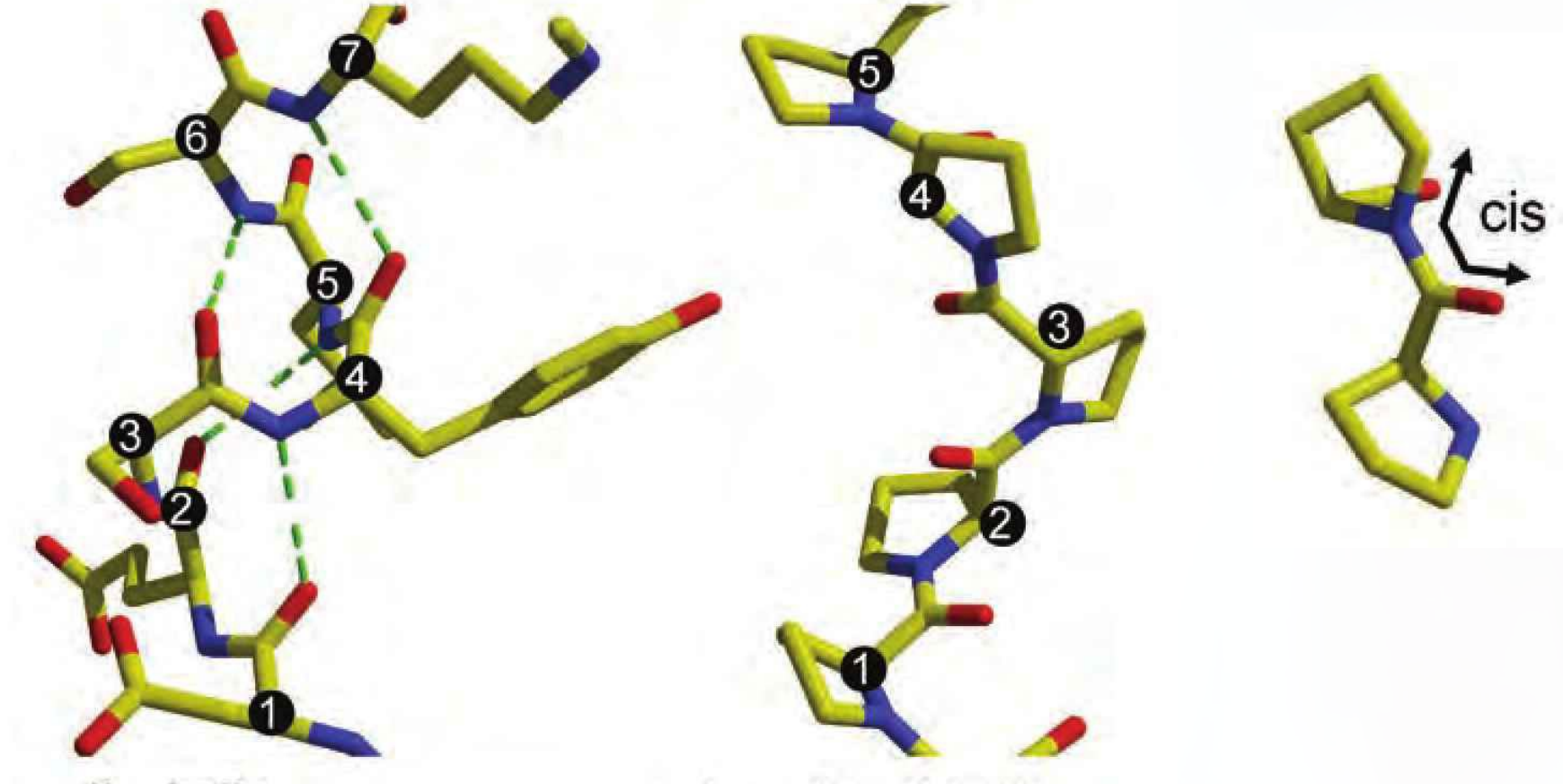
Gli amminoacidi in un'alfa elica sono disposti a formare una struttura ad elica destrorsa in cui ogni residuo Corrisponde a 100° di rotazione e 1.5 Å lungo l'asse dell'elica (il «passo» Dell'alfa elica è 3.6)





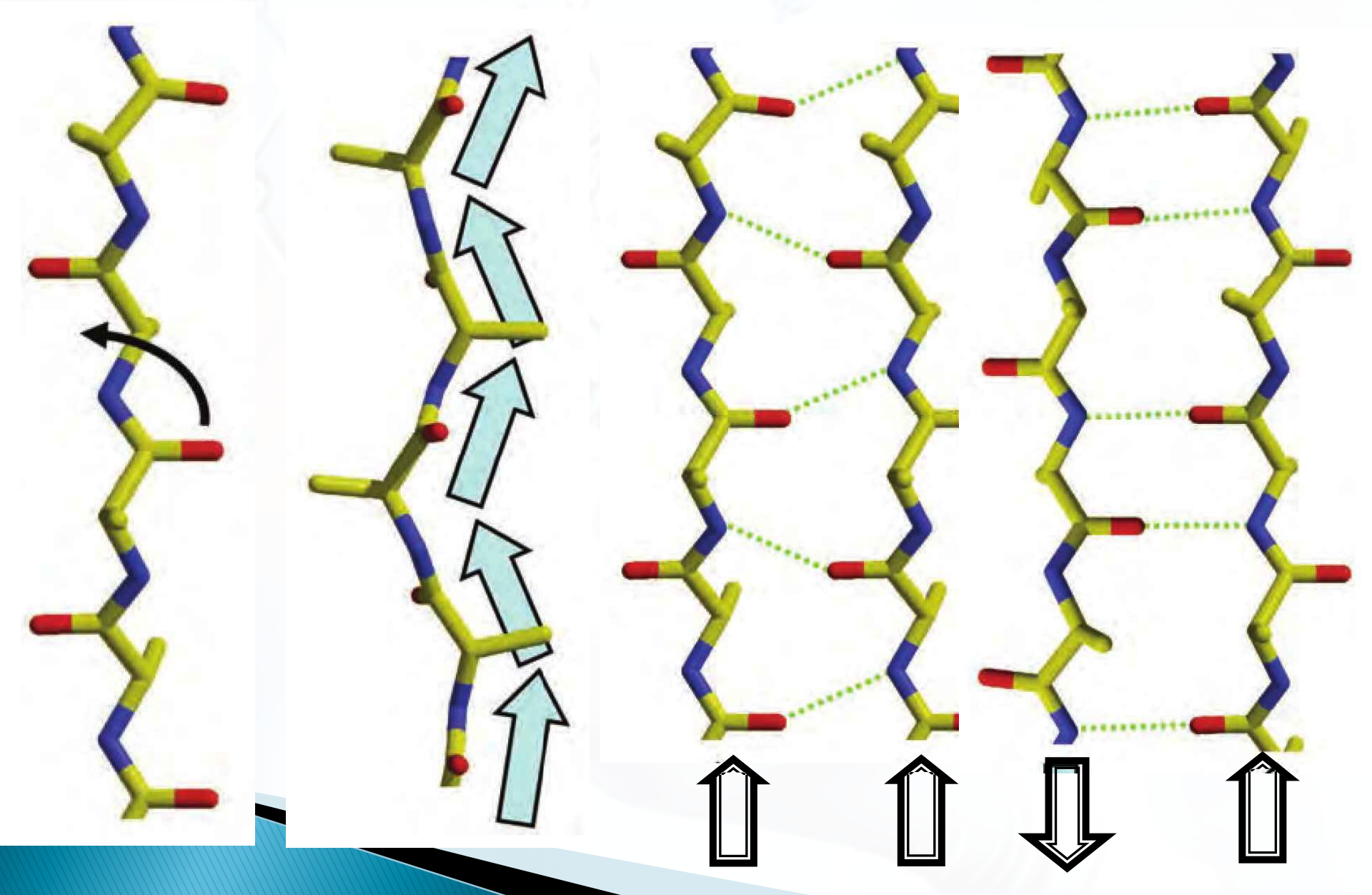


Eliche 3_{10} e eliche di poliprolina di tipo II

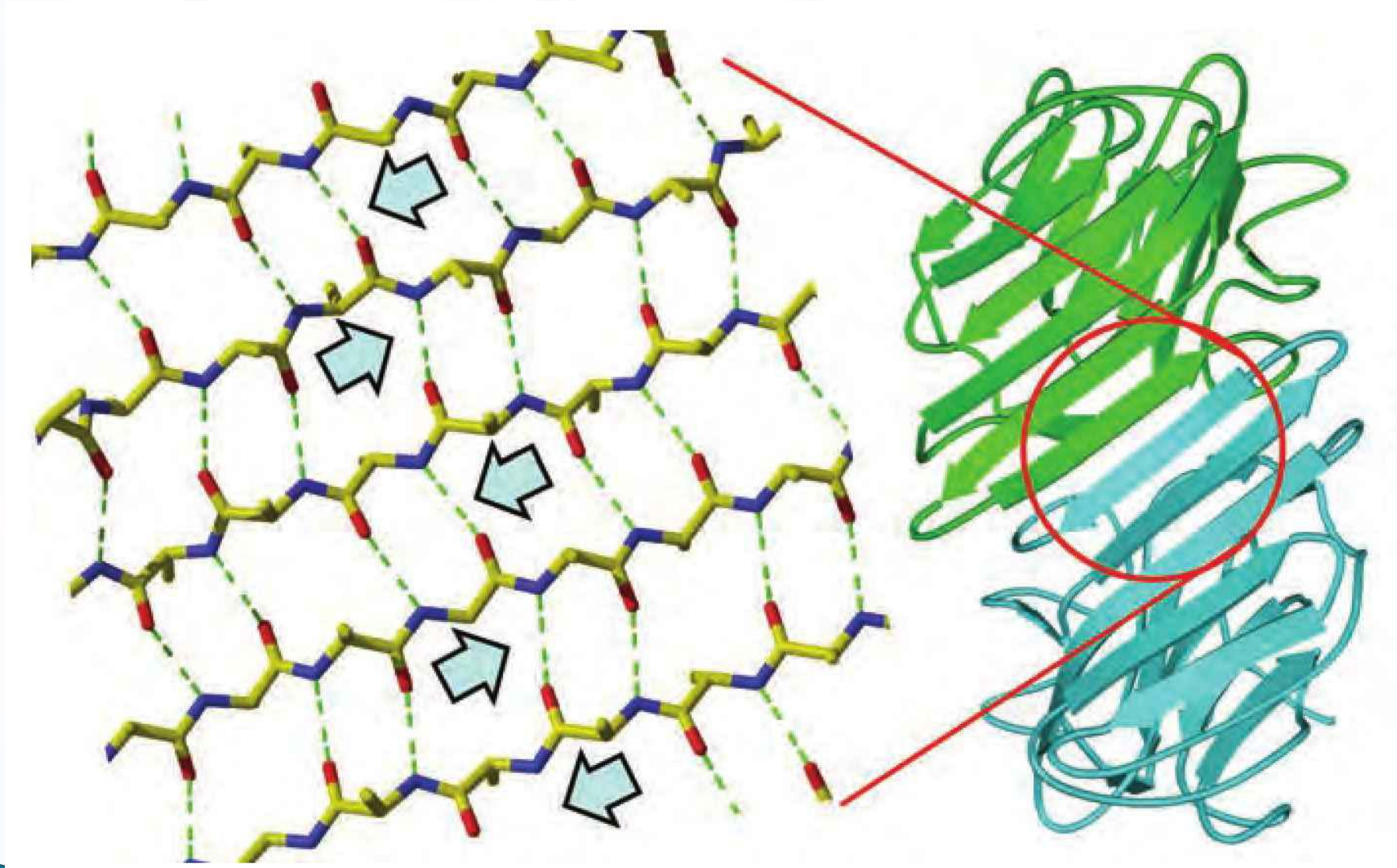


3: ogni giro dell'elica contiene circa 3 residui amminoacidici.
10: il legame a idrogeno si forma tra il gruppo C=O del residuo i e il gruppo N-H del residuo i+3, generando un anello di 10 atomi lungo la catena principale

Foglietti Beta. Formati da filamenti beta disposti parallelamente o antiparallelamente



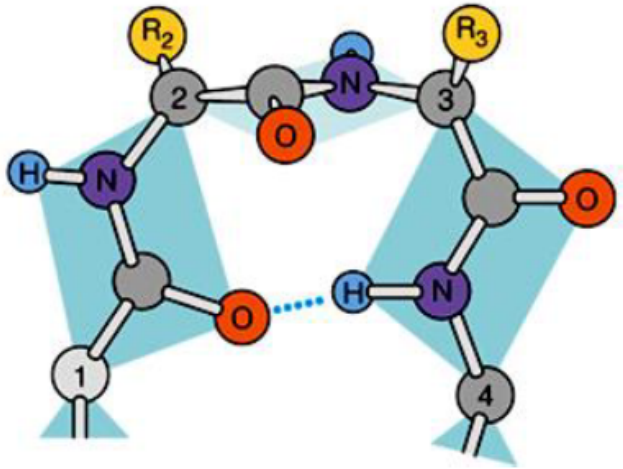
Il «Beta sheet sandwich» nella concavalina A



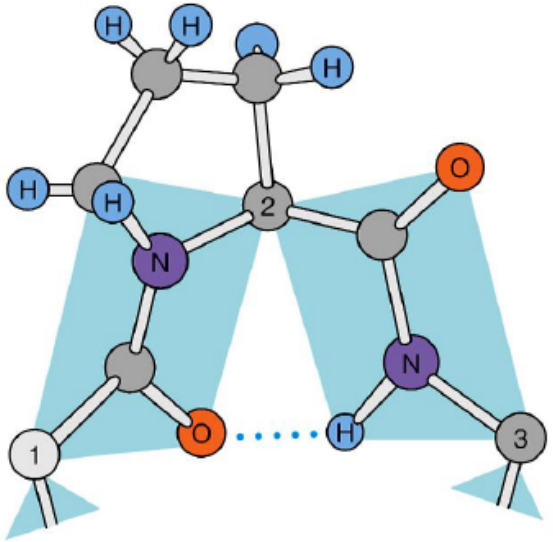
RIPIEGAMENTI NELLA CATENA POLIPEPTIDICA

I punti della catena polipeptidica dove avvengono le ripiegature, sono di solito occupati da alcuni specifici amminoacidi, ad es. glicina, prolina, che ne permettono il ripiegamento grazie alla formazione di legami ad idrogeno in determinate posizioni (ripiegatura β , ripiegatura γ).
La prolina inoltre è in grado di interrompere la struttura ad α-elica.
Curve e ripiegature si presentano di solito alla superficie delle proteine.

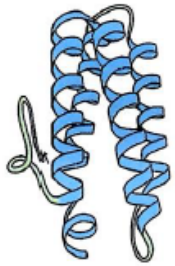
(ripiegatura β)
Punto di inversione determinato dall'amminoacido GLICINA



(ripiegatura γ)
Punto di inversione determinato dall'amminoacido PROLINA

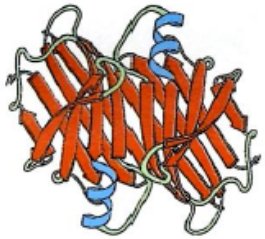


Esempi di proteine aventi percentuali differenti di prevalenza delle due strutture secondarie



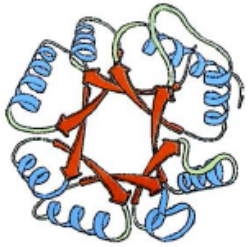
Mioemeritrina

(a) Prevalentemente ad α elica



Prealbumina

(b) Prevalentemente a foglietto β



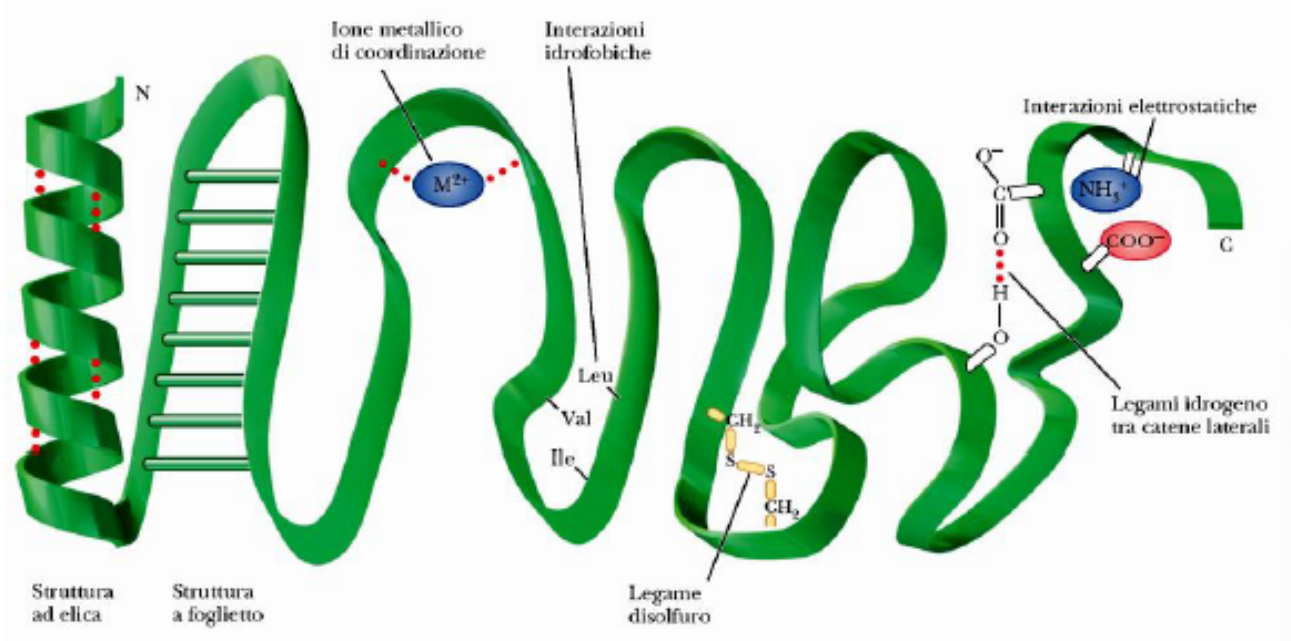
Piruvato chinasi, dominio 1

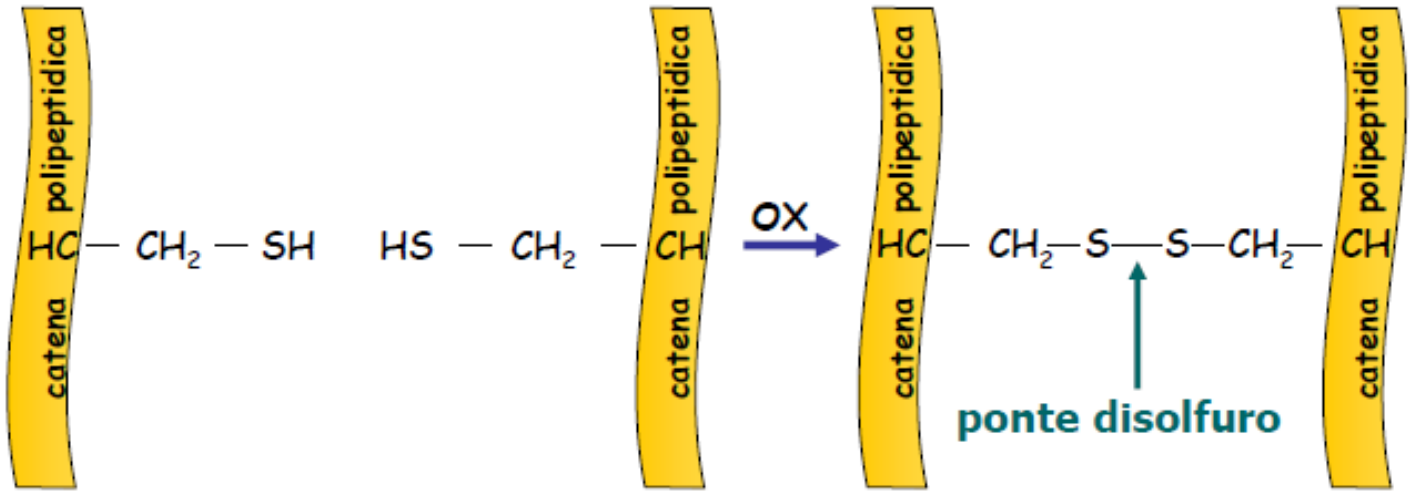
(c) Struttura mista ad α elica e a foglietto β

FORZE CHE STABILIZZANO LA STRUTTURA TERZIARIA DELLE PROTEINE

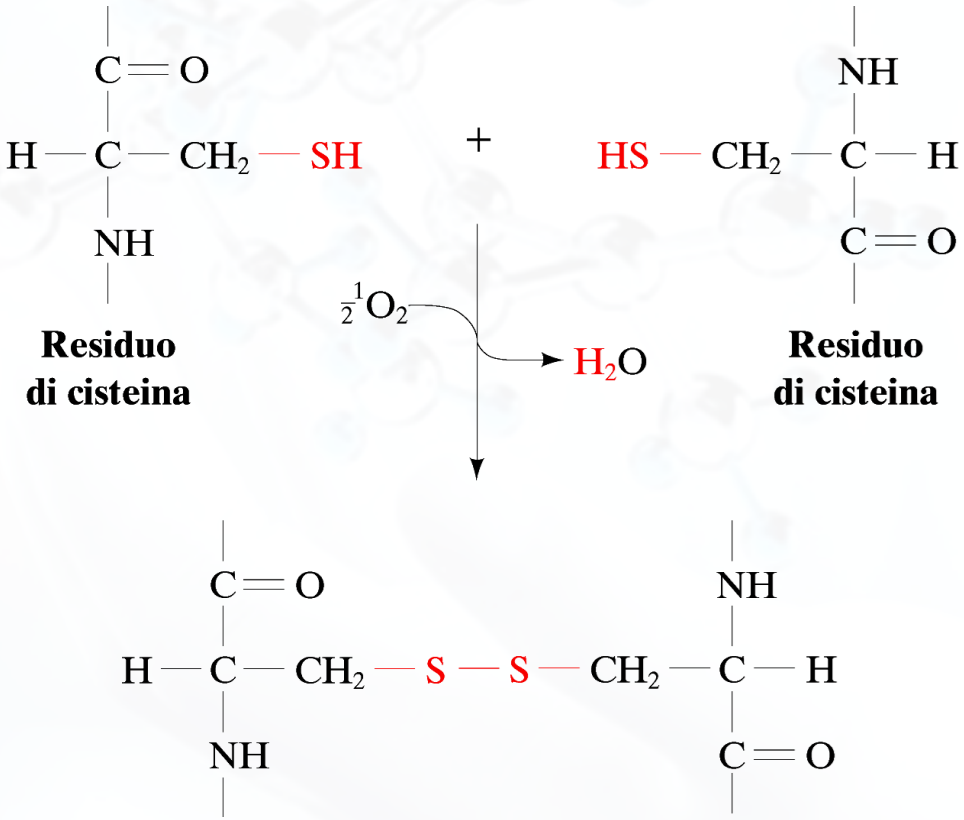
La forma compatta di una proteina globulare (forma nativa) è essenziale per la funzione che tale proteina deve svolgere; pertanto la sua struttura tridimensionale è mantenuta inalterata da tutta una serie di forze che obbligano la catena polipeptidica a conservare quella forma nello spazio.
Tali forze sono:

- **Interazioni di Van der Waals**
- **Ponti salini**
  (attrazioni elettrostatiche tra cariche di segno opposto presenti sulla catena laterale R)
- **Legami ad H intramolecolari**
- **Interazioni idrofobiche**
- **Ponti disolfuro**
  (legami covalenti che si stabiliscono tra due residui di cisteina)

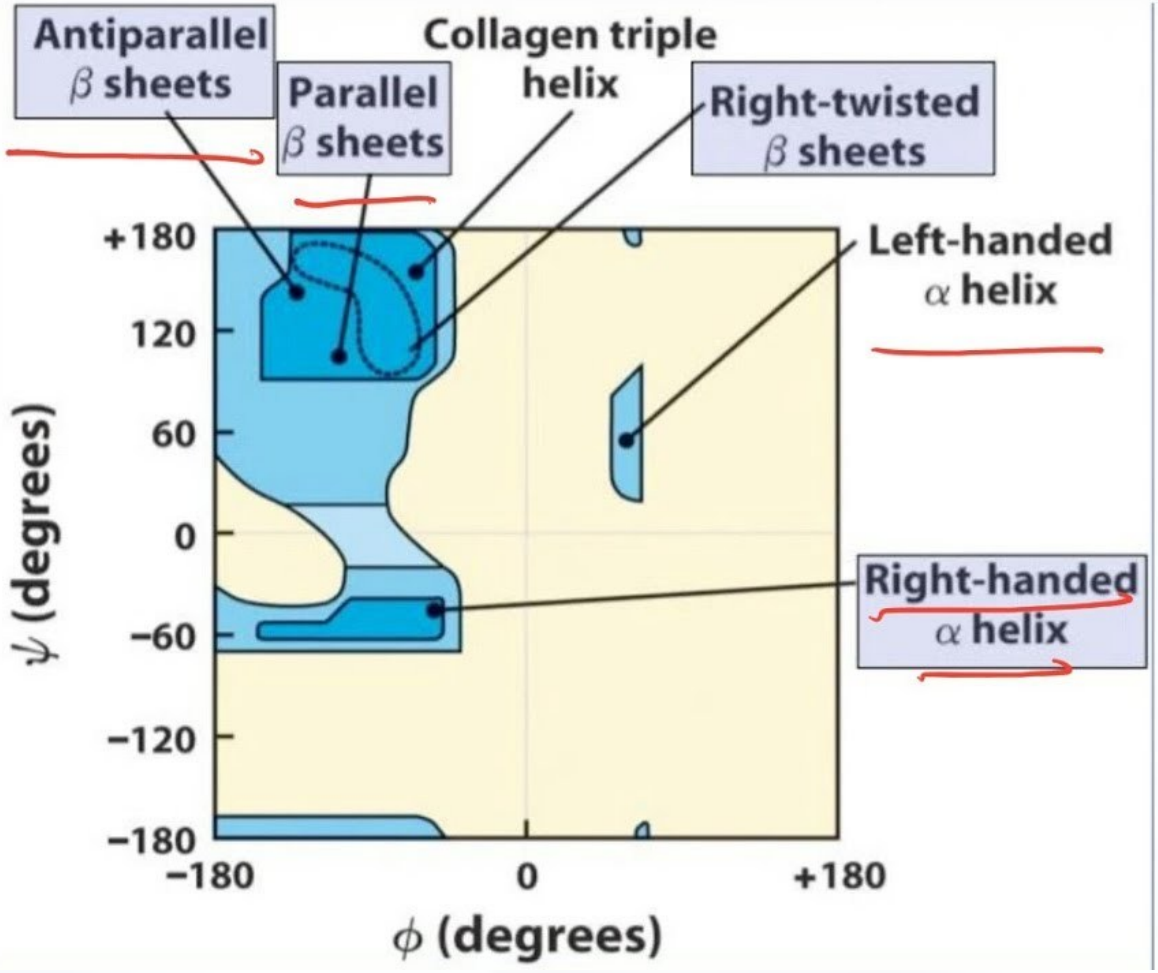


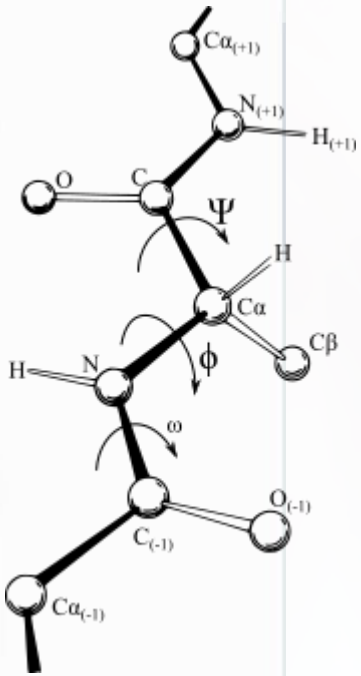


La formazione del ponte disolfuro è una ossidoriduzione:



Ramachandran Plot





Le proteine, in base alla loro forma, possono essere distinte in:

PROTEINE FIBROSE - PROTEINE GLOBULARI

Le proteine fibrose presentano una struttura filamentosa, e rivestono un ruolo strutturale in cellule e tessuti animali. Esse comprendono infatti le più importanti proteine della pelle e del tessuto connettivo e quelle di fibre animali come peli, lana e seta. Tra le proteine fibrose abbiamo le α-cheratine, principali costituenti dei capelli e delle unghia e in buona misura anche della pelle.

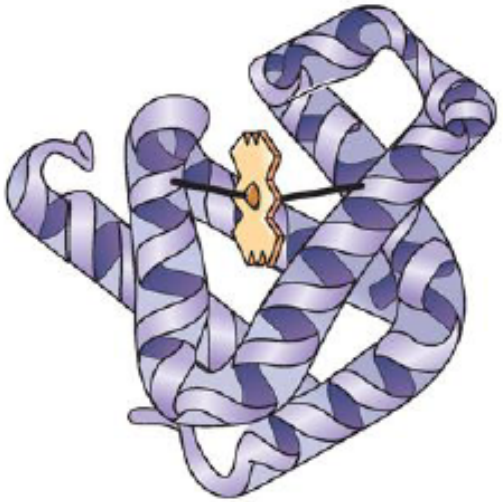
Proteine fibrose:

- insolubili in acqua
- forma presente nel tessuto connettivo
- seta , collagene , α - cheratine



Proteine globulari:

- solubili in acqua
- forma presente nelle proteine cellulari
- struttura tridimensionale (terziaria)

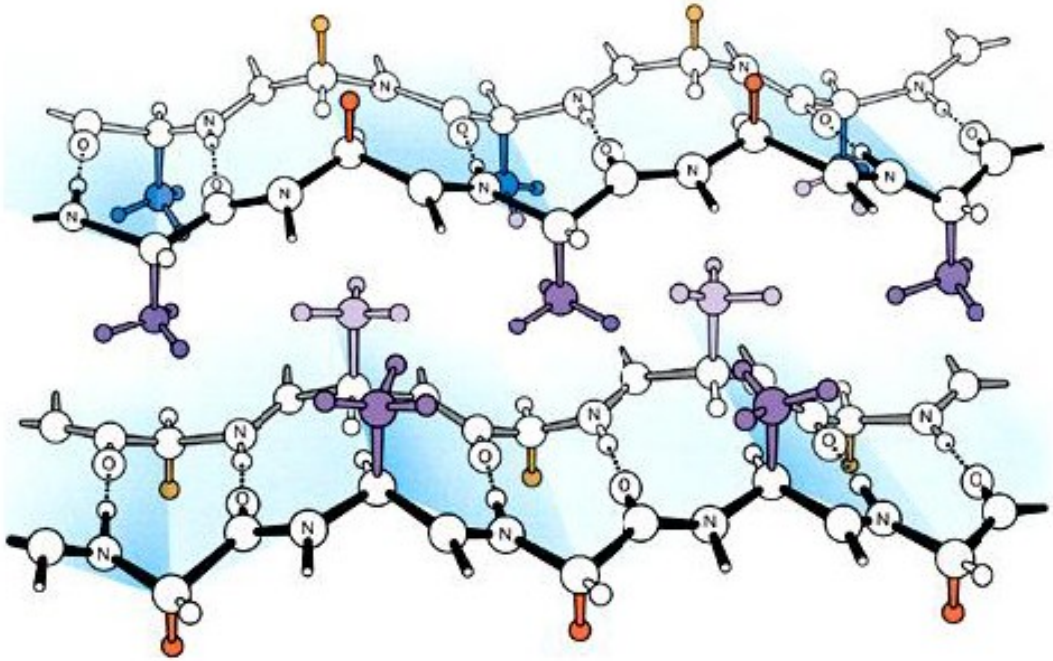


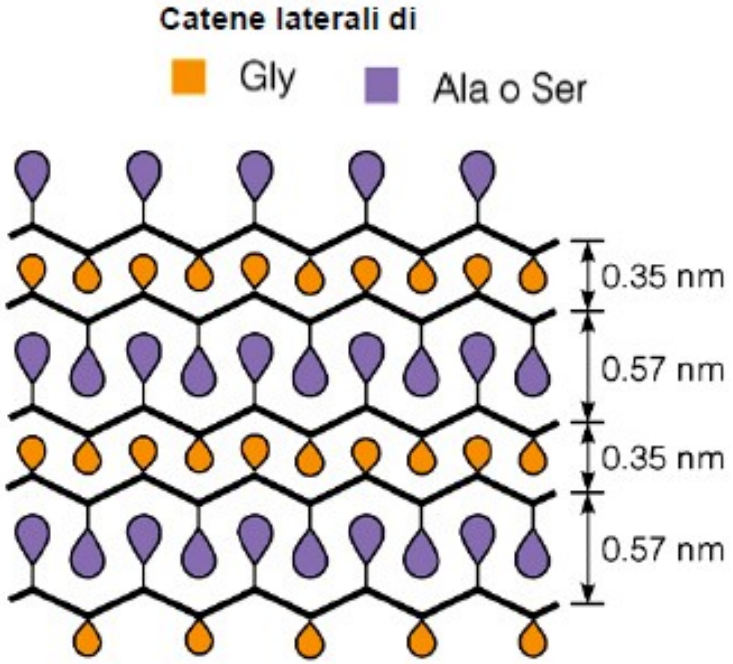
Struttura della FIBROINA

E' un esempio di proteina a foglietto β: costituisce le fibre filate dei bozzoli dei bachi da seta e della tela dei ragni.

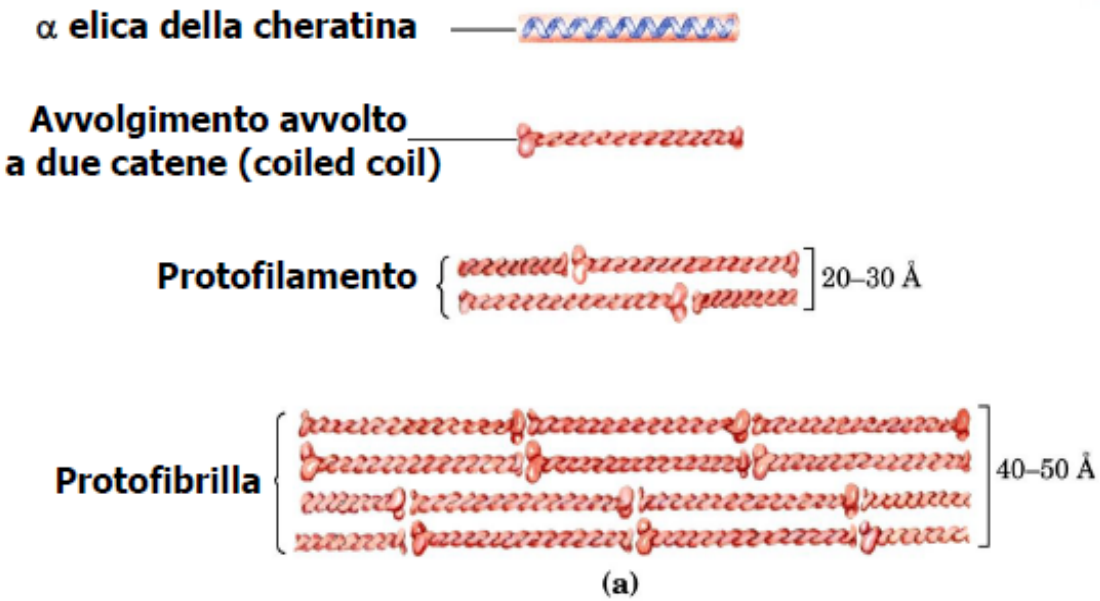


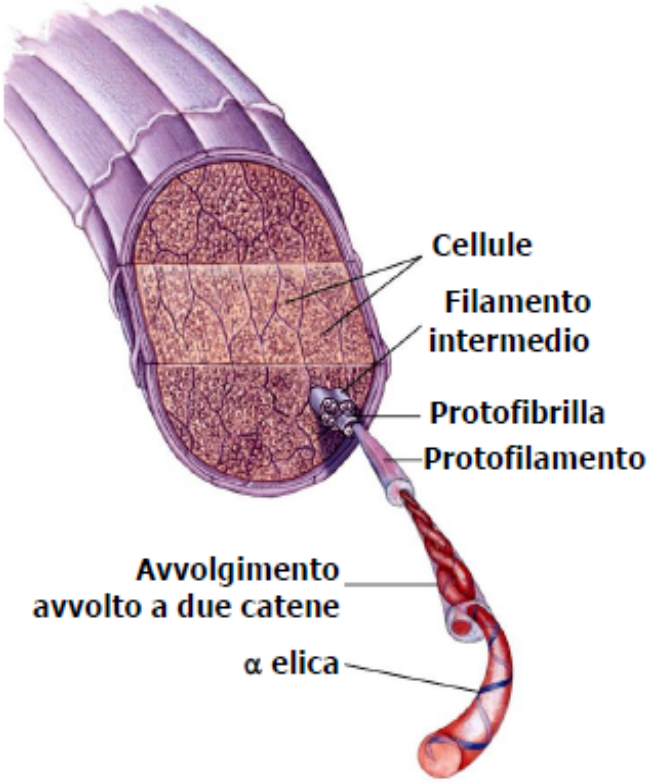
E' formata da lunghe sequenze a foglietto β antiparallelo che decorrono parallele all'asse della fibra. La metà degli amminoacidi è costituita dall'amminoacido glicina che si intercala con la serina o con l'alanina. Questa particolare sequenza rende la fibra compatta e forte, inestensibile e allo stesso tempo flessibile.





STRUTTURA DEL CAPELLO, FORMATO DA UNA TIPICA α-CHERATINA



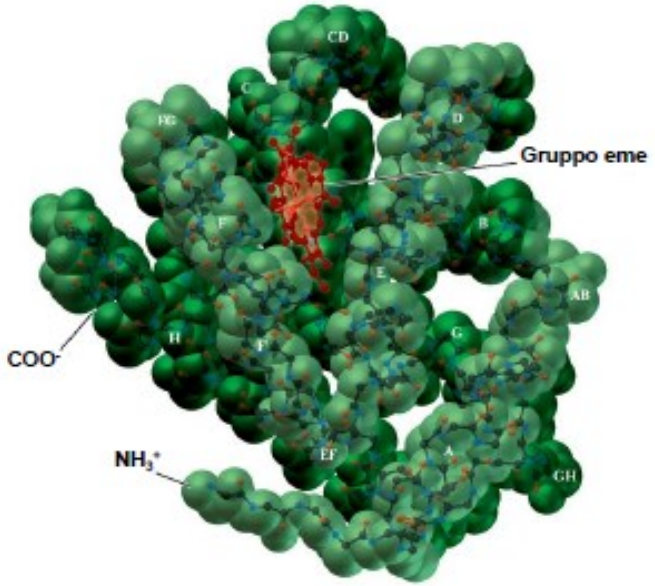


Sezione trasversale di un capello

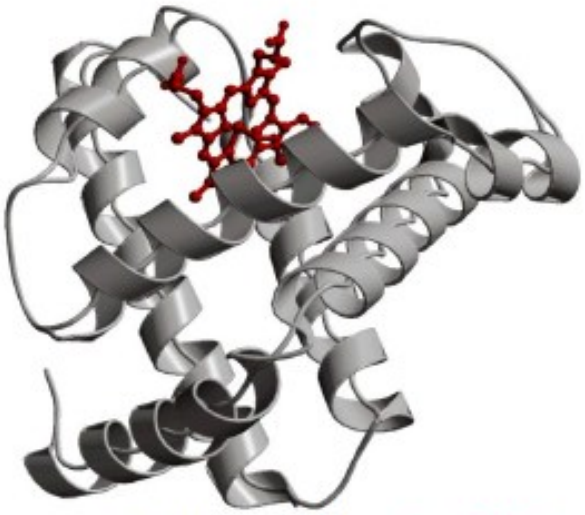
Le strutture filamentose sono formate da più di 300 residui amminoacidici e l'avvolgimento a coppie (coiled coil) è sinistrorso; inoltre in tessuti differenti possono esistere un numero variabile di legami crociati disolfuro.

Un' altra classe di proteine fibrose sono le β-cheratine , che contengono molta più struttura a foglietto β e sono presenti per lo più in uccelli e rettili, in strutture come piume e scaglie.

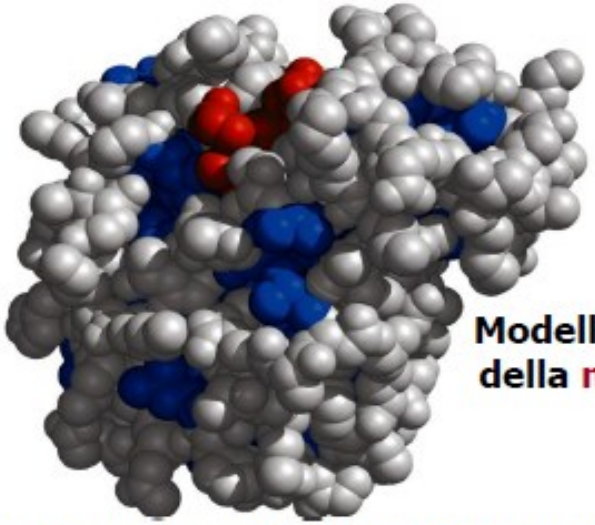
La maggior parte del lavoro chimico di una cellula (sintesi, trasporto, metabolismo), avviene grazie all'azione delle **PROTEINE GLOBULARI**. Le proteine globulari si differenziano da quelle fibrose in quanto presentano un 3° livello di organizzazione strutturale (**struttura terziaria**) che conferisce alla molecola proteica una ben definita e compatta forma tridimensionale.



RIPIEGAMENTO TRIDIMENSIONALE DELLA PROTEINA MIOGLOBINA



Scheletro della mioglobina in forma di nastro, un sistema utile ad evidenziare le regioni con struttura secondaria

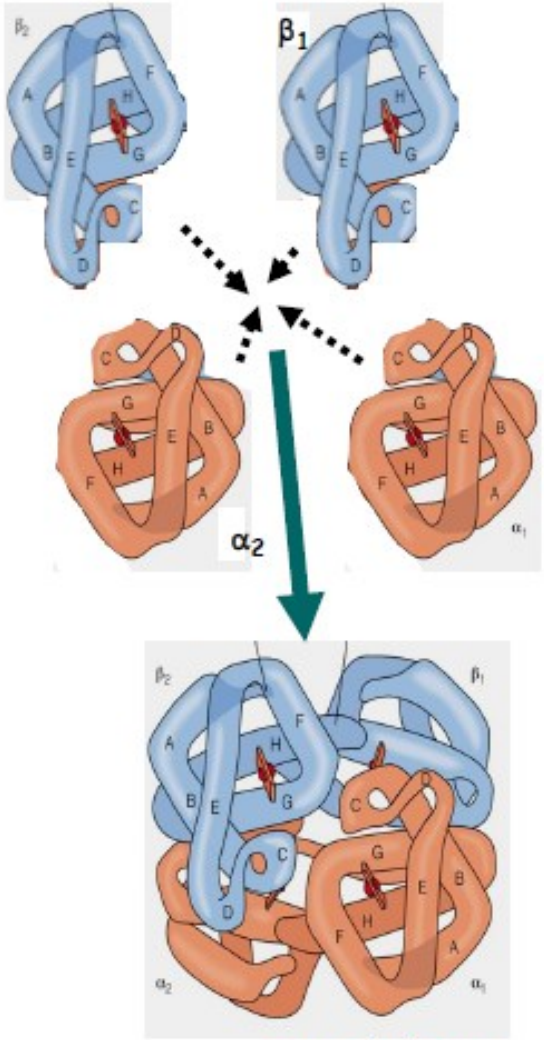


Modello spaziale della mioglobina

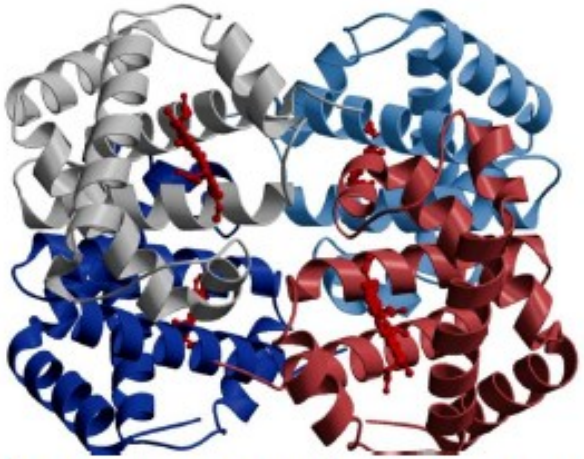
La mioglobina è una proteina che si trova nel tessuto muscolare di molti animali. Essa è utilizzata per immagazzinare O_2; tale funzione è svolta in quanto una parte non proteica della molecola, l'eme, è in grado di legare con alta affinità l'ossigeno e conservarlo fino a quando sarà utilizzato a livello cellulare. La mioglobina è costituita da una sola catena polipeptidica che si ripiega più volte nello spazio e in essa sono presenti varie zone ad α-elica (A - H).

Struttura quaternaria delle proteine

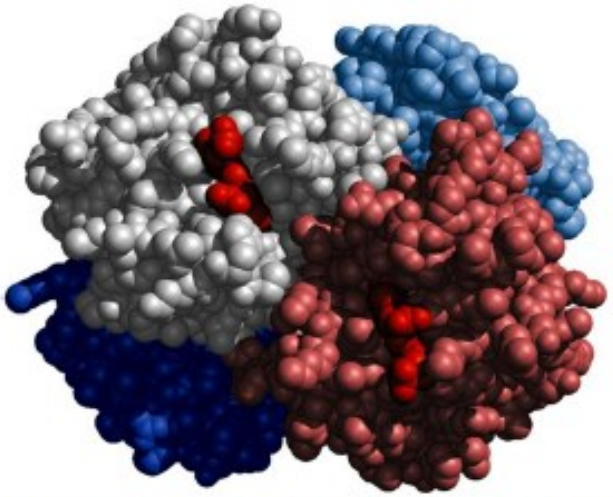
Due o più catene polipeptidiche possono unirsi tra loro formando una proteina multimerica, in questo caso si origina un 4° livello di organizzazione strutturale (struttura quaternaria). Le catene polipeptidiche che formano la proteina sono chiamate subunità e sono tenute insieme dalle stesse forze che stabilizzano la struttura terziaria.



emoglobina



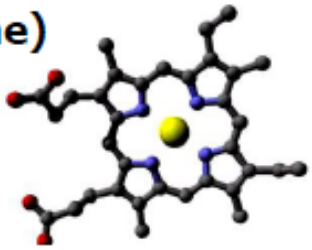
Scheletro a nastro della deossiemoglobina



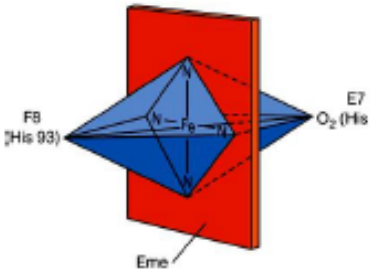
Modello spaziale della deossiemoglobina

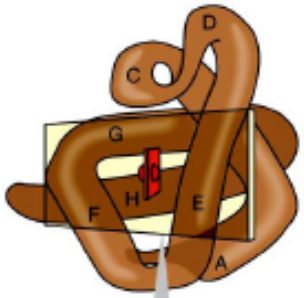
CURVE DI DEOSSIGENAZIONE

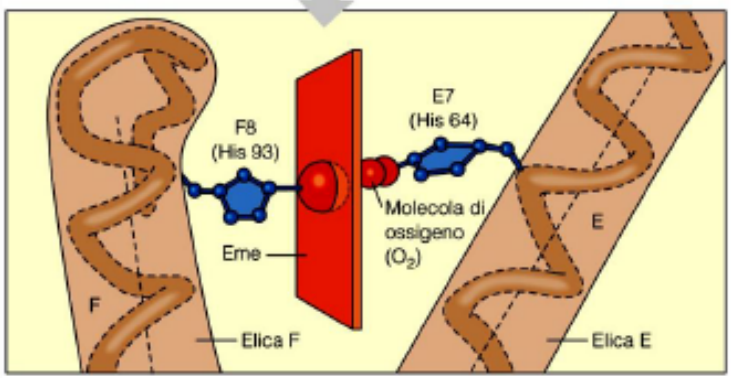
Mioglobina: immagazzina O_2 nelle cellule muscolari (1 gruppo eme)

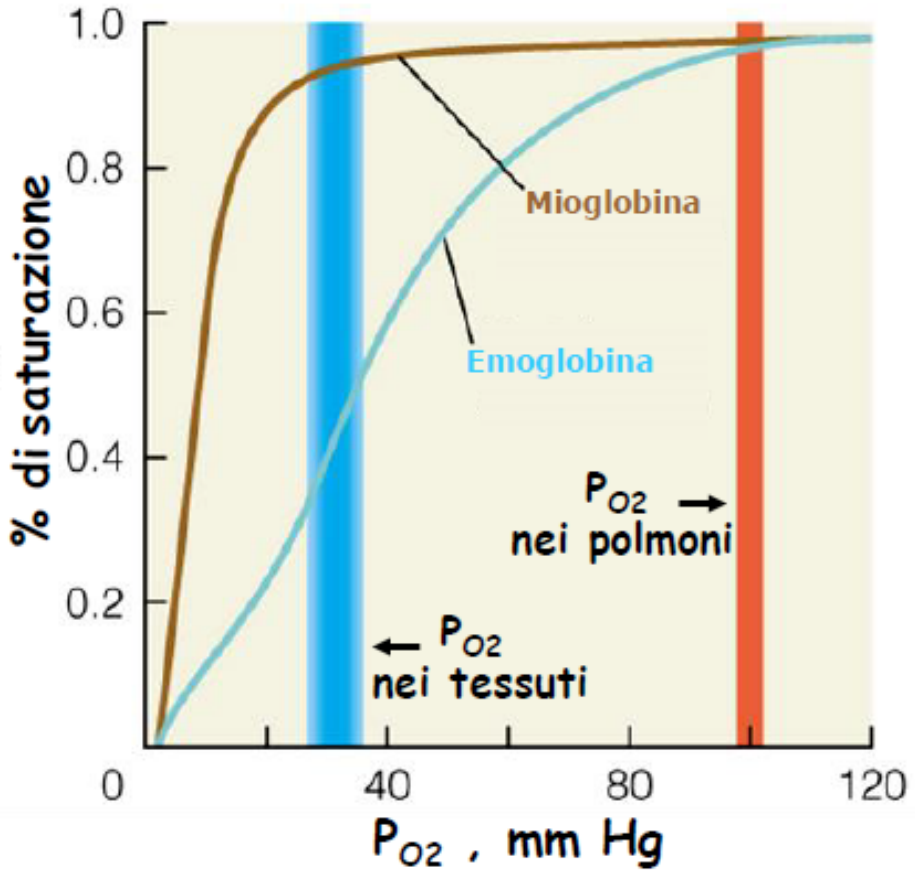


Emoglobina: nelle emazie trasporta O_2 (4 gruppi eme)







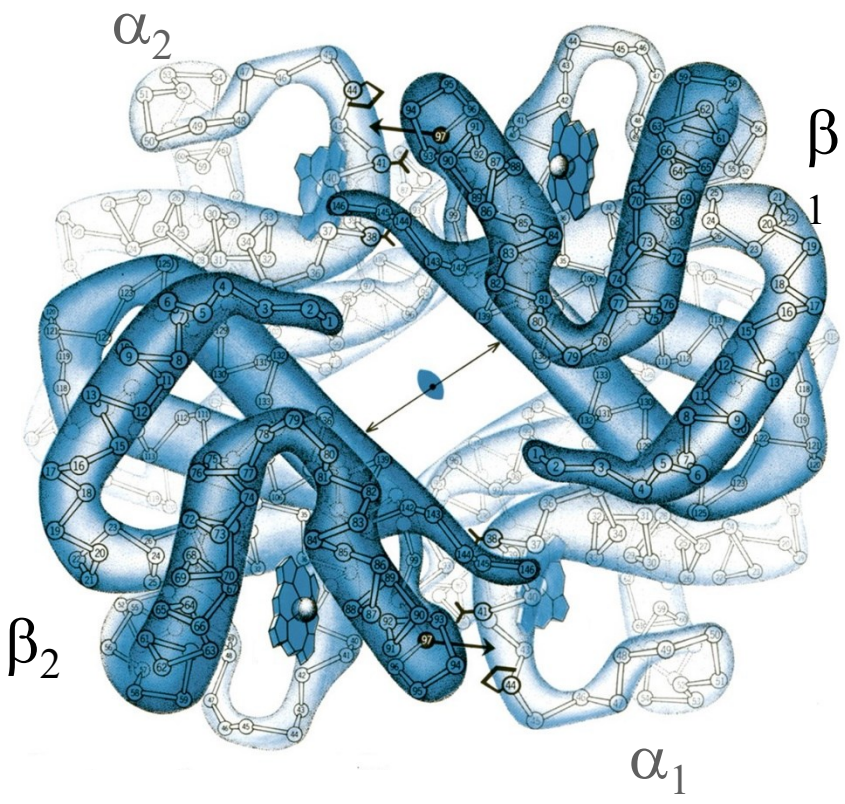


L'emoglobina è un classico esempio di proteina allosterica, intendendo con questo termine un meccanismo di cooperatività tra i quattro gruppi eme presenti in una molecola di emoglobina. La cooperatività consiste nel fatto che quando una molecola di ossigeno si lega ad un gruppo eme di una molecola di emoglobina, gli altri tre gruppi eme presenti nella stessa molecola, acquistano una maggiore affinità per l'ossigeno e quindi possono legare con maggiore facilità altre 3 molecole di ossigeno.

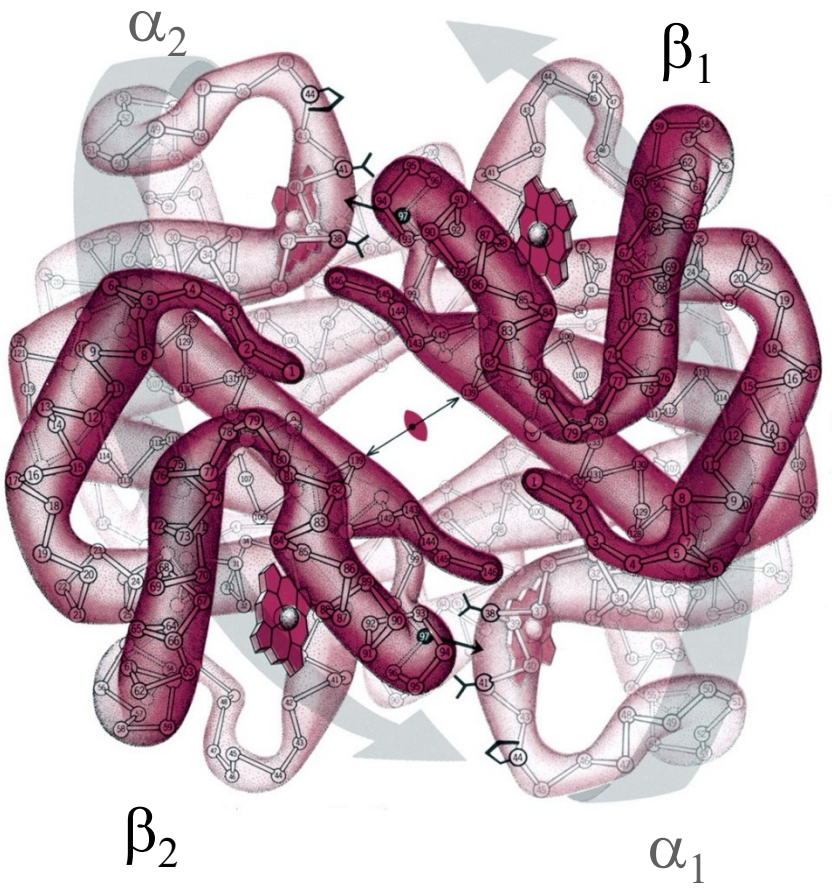
Stati T e R di Hb

Il legame con l'**ossigeno** provoca un **ampio cambiamento conformazionale** dovuto ad un cambiamento delle interazioni dei residui all'interfaccia α_1–β_2 e α_2–β_1

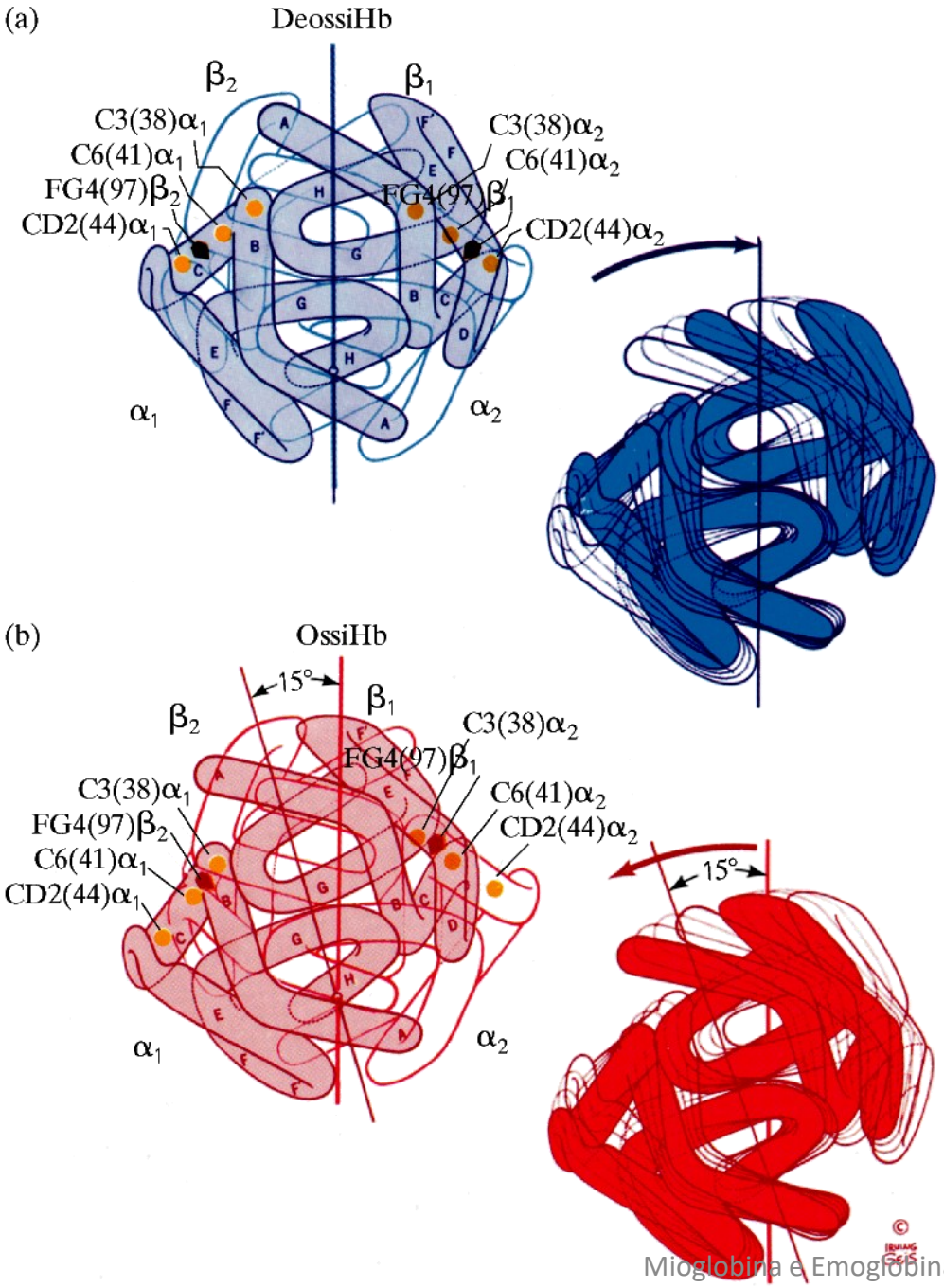
Deossi Hb - **T state**



Ossi Hb - **R state**



Il **canale** si restringe



Il **cambiamento conformazionale** preserva la simmetria e avviene **esclusivamente a livello dell'interfaccia α_1-β_2 (α_2-β_1)** .

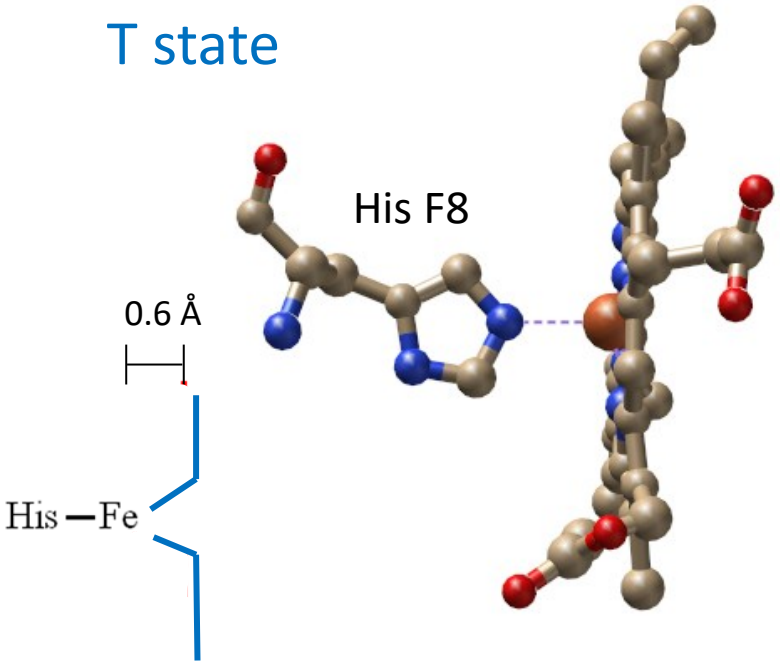
Il protomero α_1–β_1 ruota di 15° rispetto a α_2–β_2 (*rigid body movement*). Alcuni atomi si allontanano fino a 6 Å.

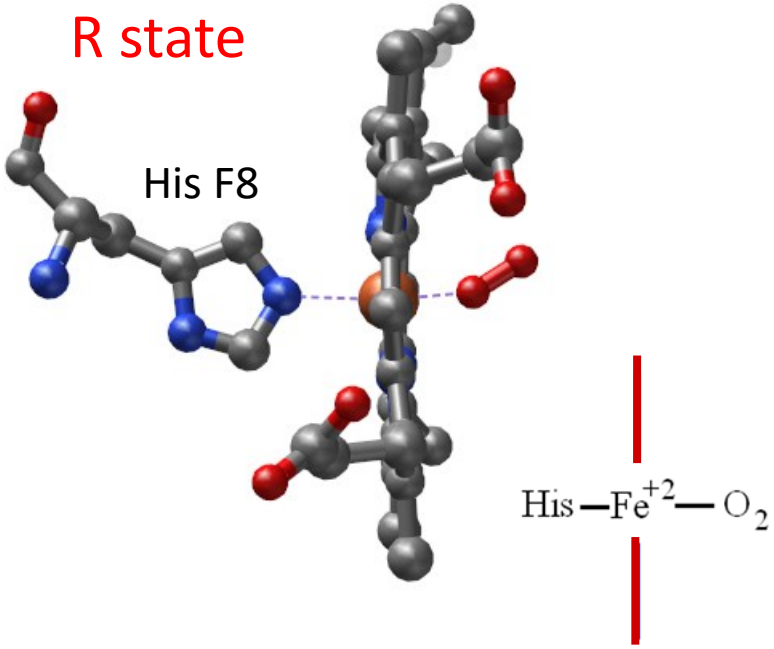
CRACK!

Da cosa è innescato questo riarrangiamento?

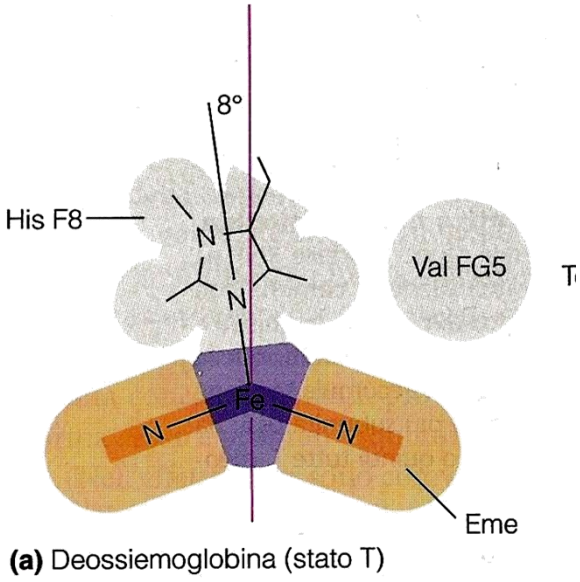
MECCANISMO DI PERUTZ

- Lo stato conformazionale De-ossi è chiamato **T** (*Teso*): è caratterizzato da una bassa affinità per O_2 e il Fe^{+2} si trova spostato 0.6 Å fuori dal piano dell'eme, dalla parte del lato prossimale.
- Lo stato conformazionale Ossi è chiamato **R** (*Rilassato*): è caratterizzato da una alta affinità per O_2 e il Fe^{+2} è complanare con l'eme.

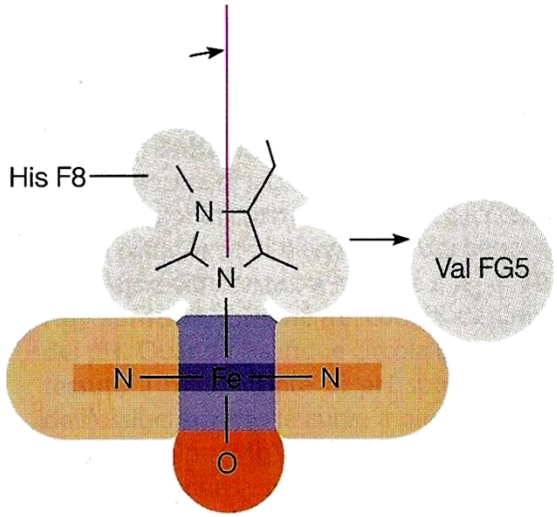




MECCANISMO DI PERUTZ



(a) Deossiemoglobina (stato T)

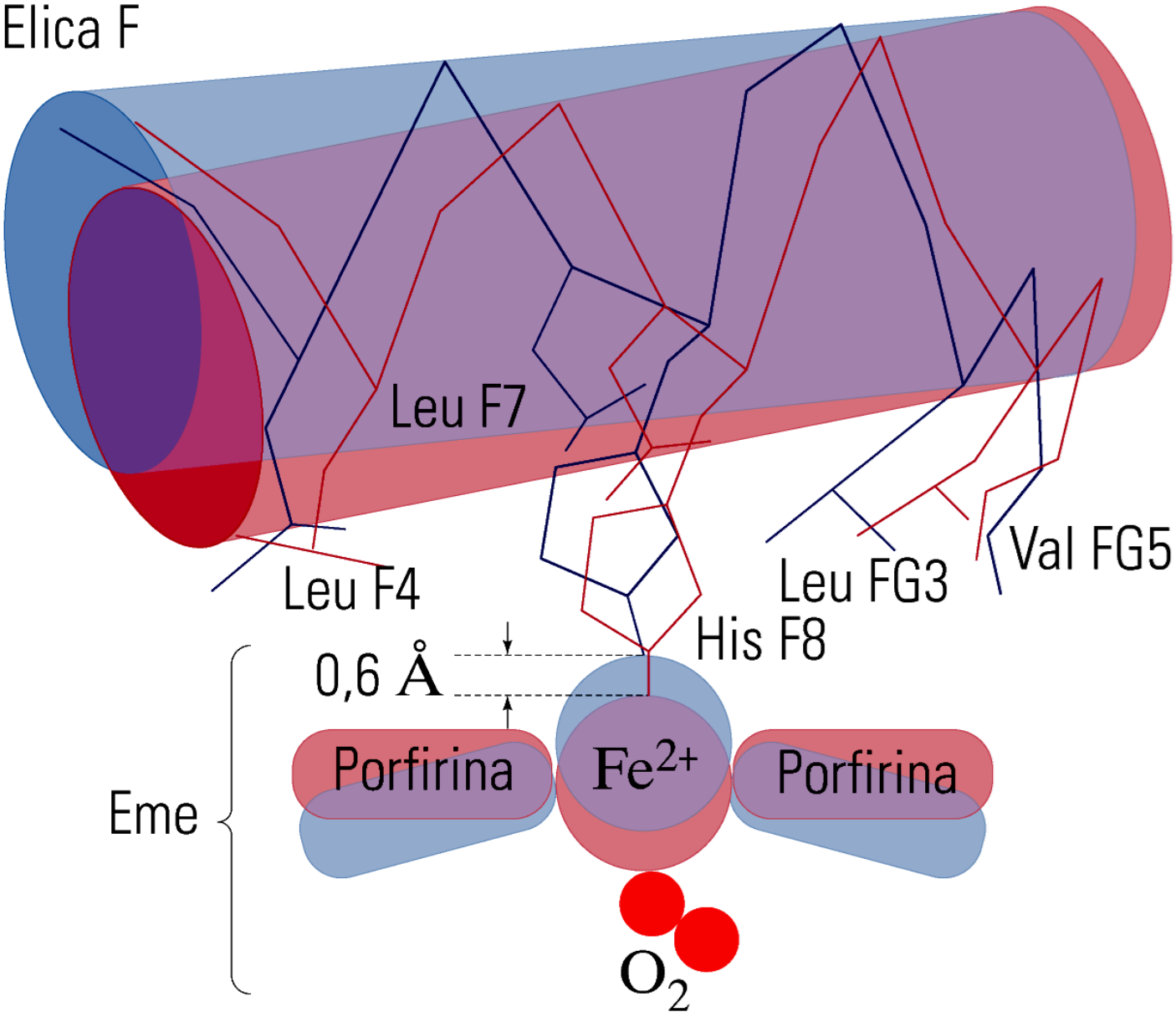


(c) Ossiemoglobina (stato R)

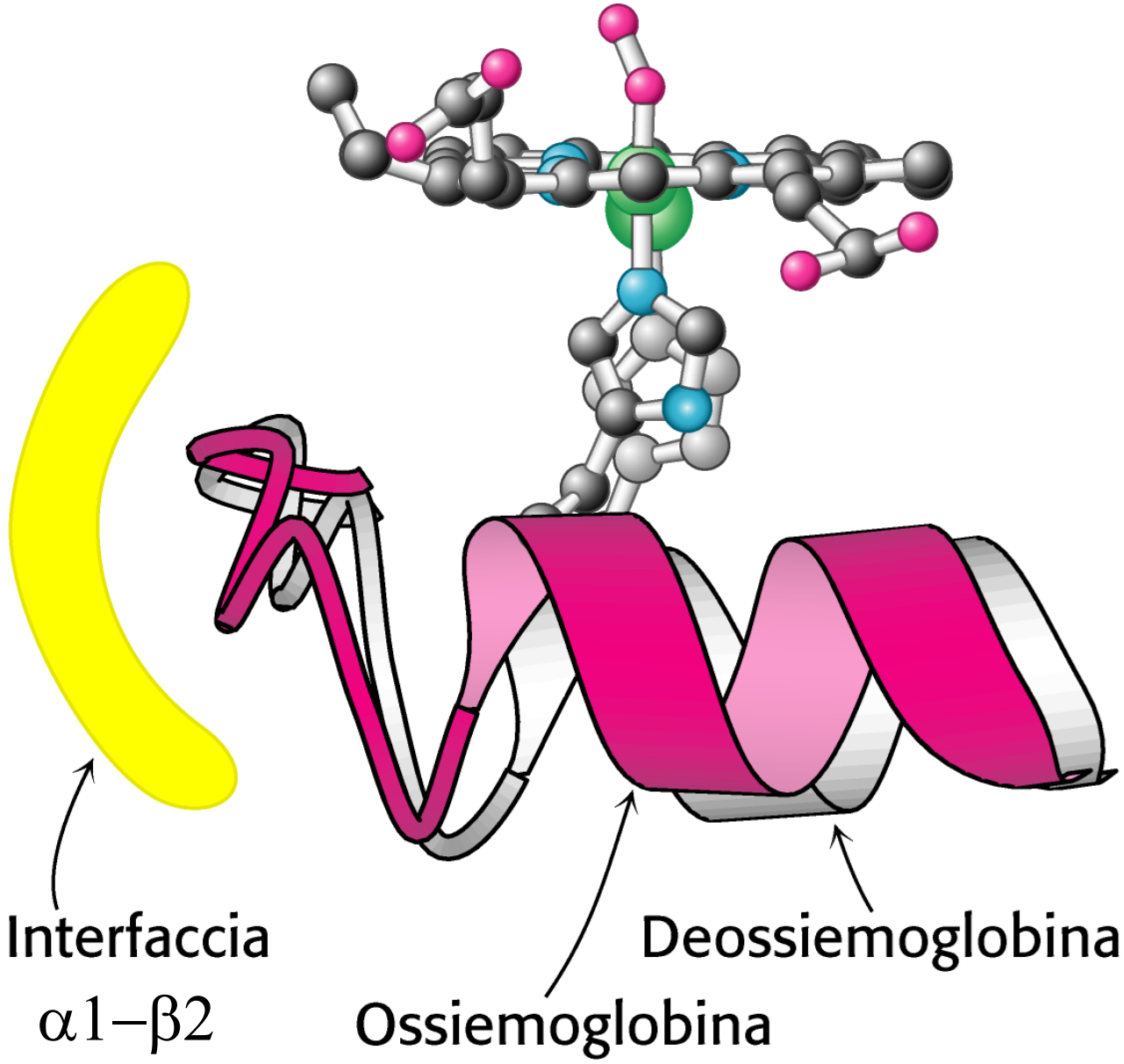
Stato T
- Eme a forma convessa
- Fe 6 fuori dal piano
- His inclinata di 8 °

Stato R
- Eme piatto
- Fe nel piano
- His perpendicolare

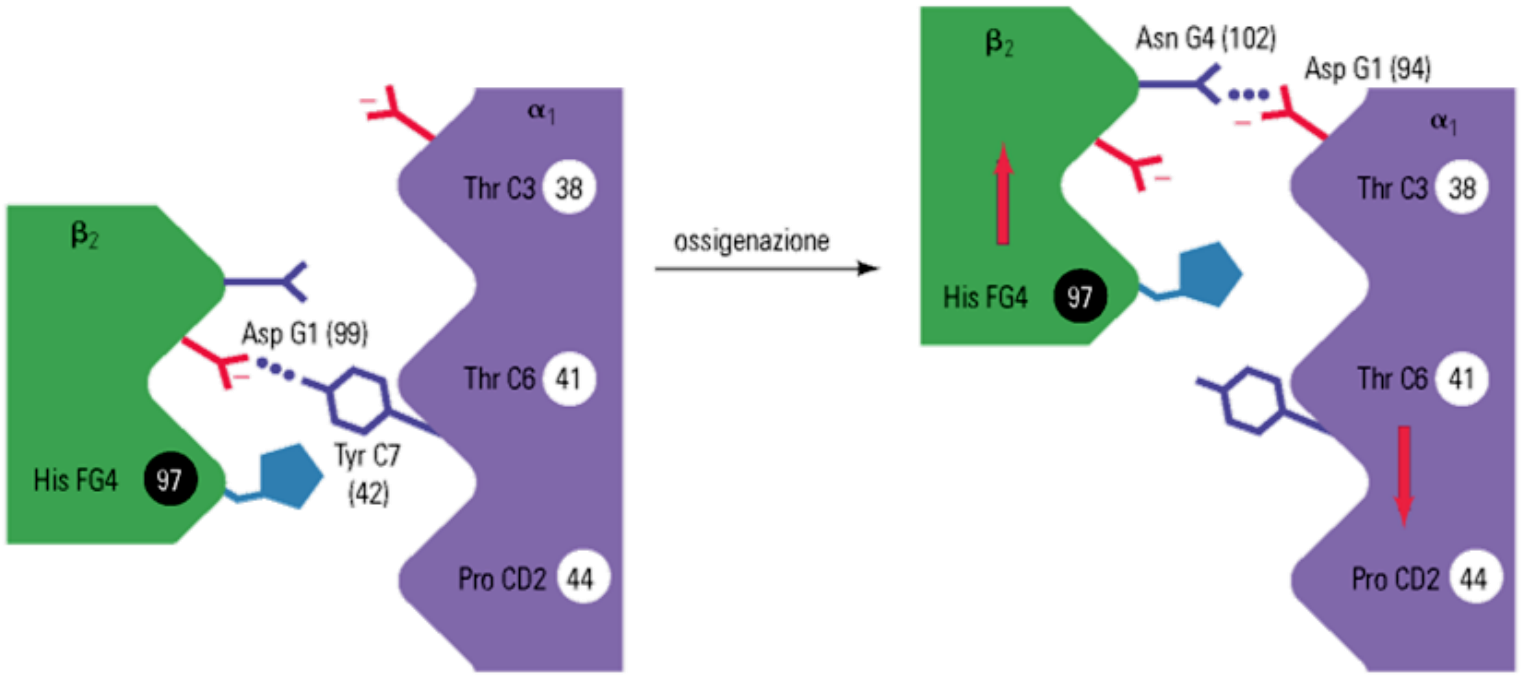
Il legame di O2 ristabilisce la simmetria di coordinazione → 'tira giù' His F8.....



….. His F8 ‘trascina’ l’elica F….



….. l'elica F fa spostare il loop FG che è all'interfaccia tra 2 protomeri αβ



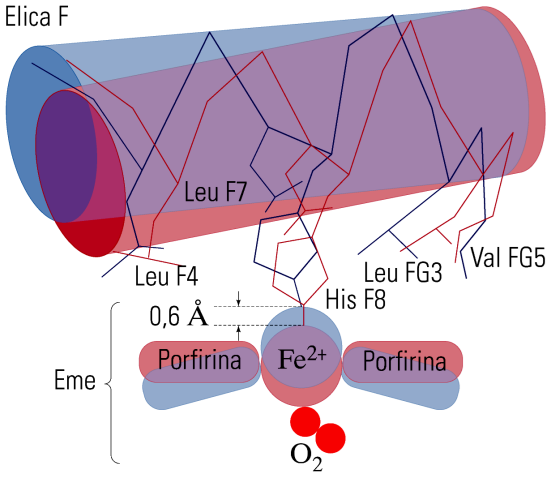
Stato T Click! **Stato R**

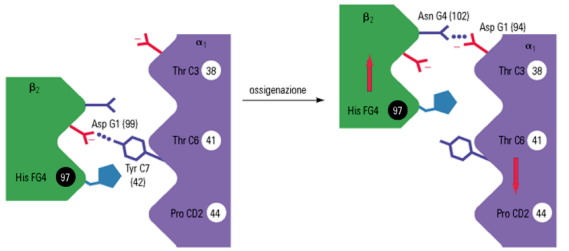
….. Il movimento del loop FG ‘seleziona’ lo stato R

Riassumendo:

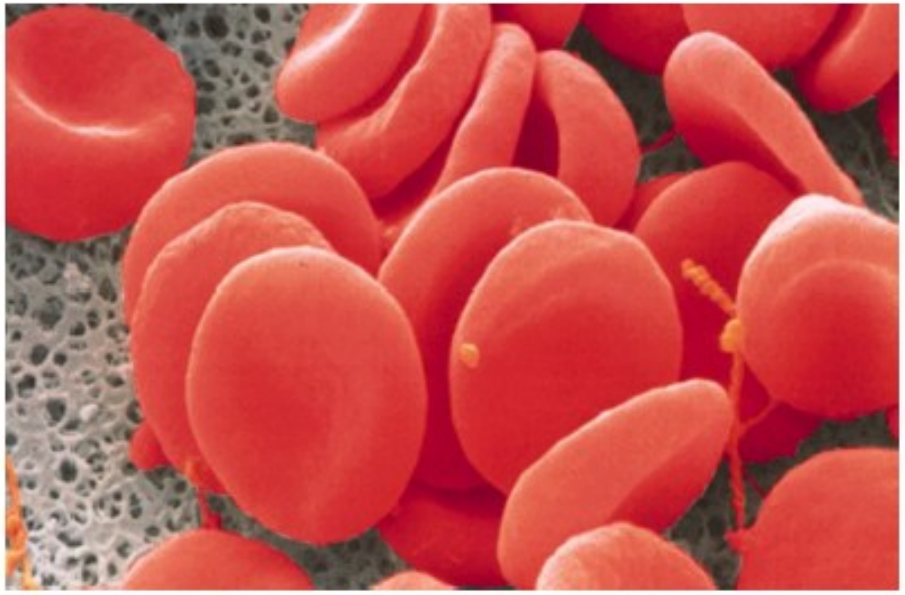
Il legame con O_2 richiede una serie di movimenti altamente coordinati (*causa-effetto*)

- Lo spostamento del Fe di ciascuna subunità causa uno spostamento e **riorientamento dell'His prossimale**, per evitare collisione con l'anello di porfirina e le altre catene laterali
- Il riorientamento dell'His prossimale (His-F8) è accompagnato da una **traslazione dell'elica F verso il piano dell'eme di ~ 1 Å**
- La traslazione dell'elica F è accoppiata ad una variazione quaternaria che **sposta il contatto α_1C–β_2FG** di un giro lungo l'elica α_1C
- la rigidità delle interfacce **α_1–β_1** e **α_2–β_2** fa sì che il cambiamento avvenga contemporaneamente in entrambe le interfacce **α_1–β_2** e **α_2–β_1**

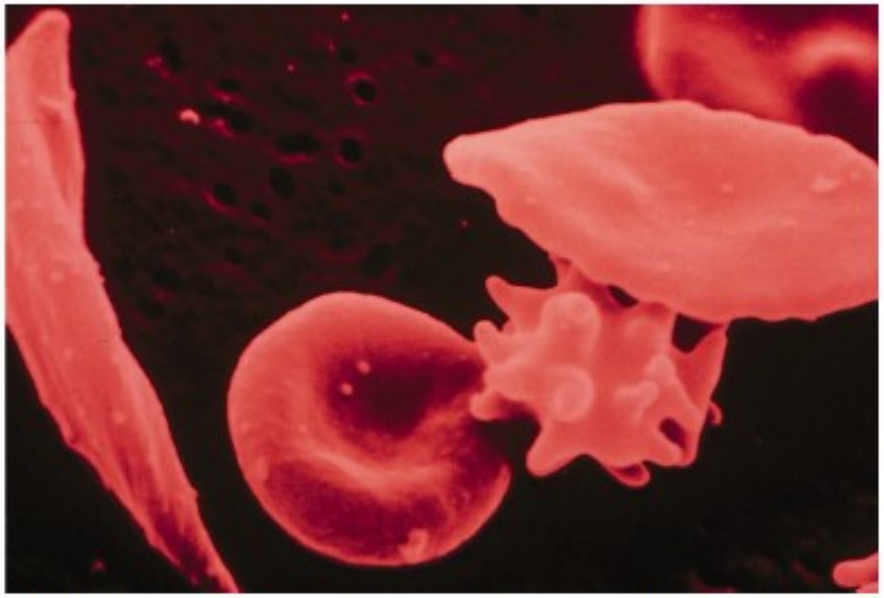




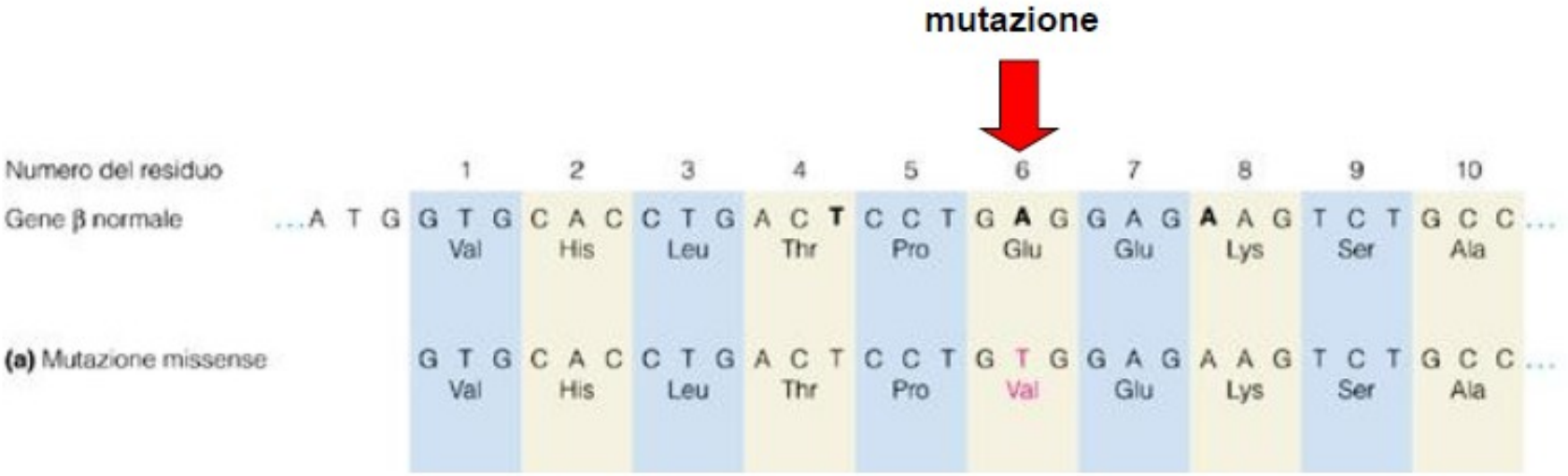
ANEMIA A CELLULE FALCIFORMI



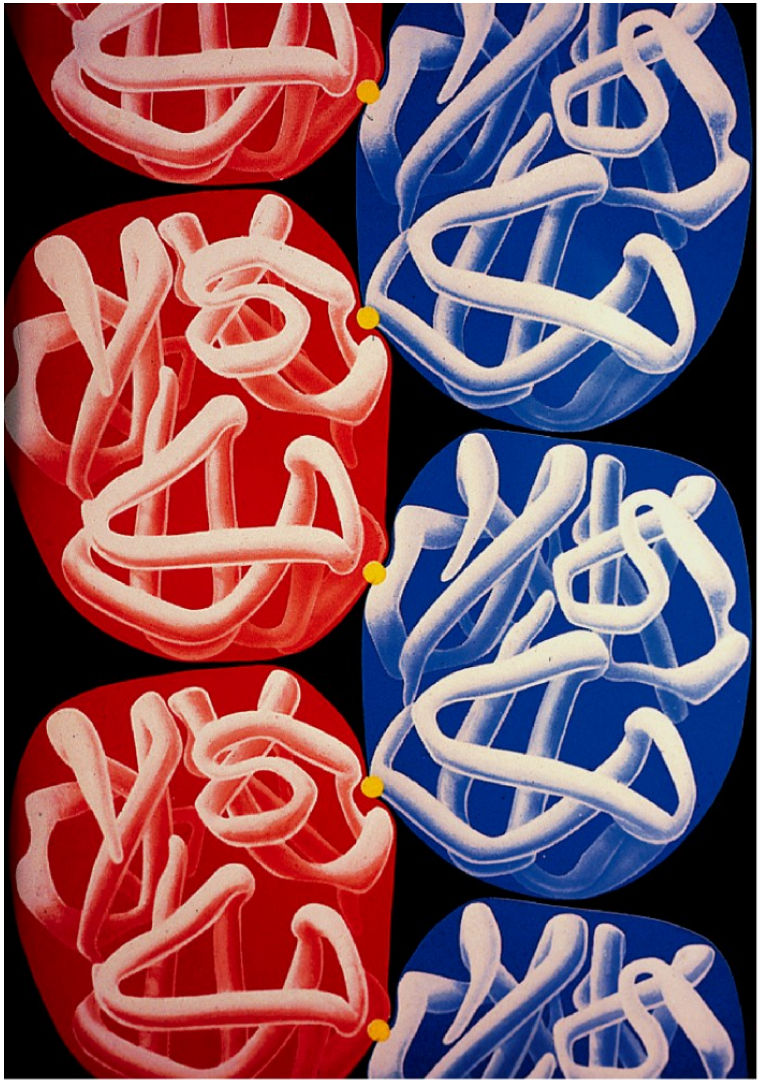
Eritrociti di soggetto normale



Eritrociti di paziente affetto da anemia falciforme

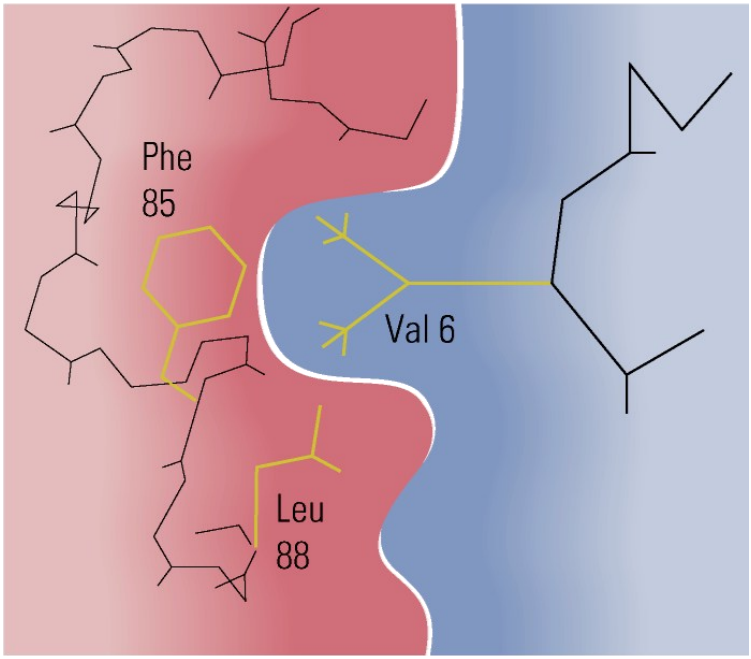


1956 – Vernon Ingram **HbS β6 A3 Glu →Val**



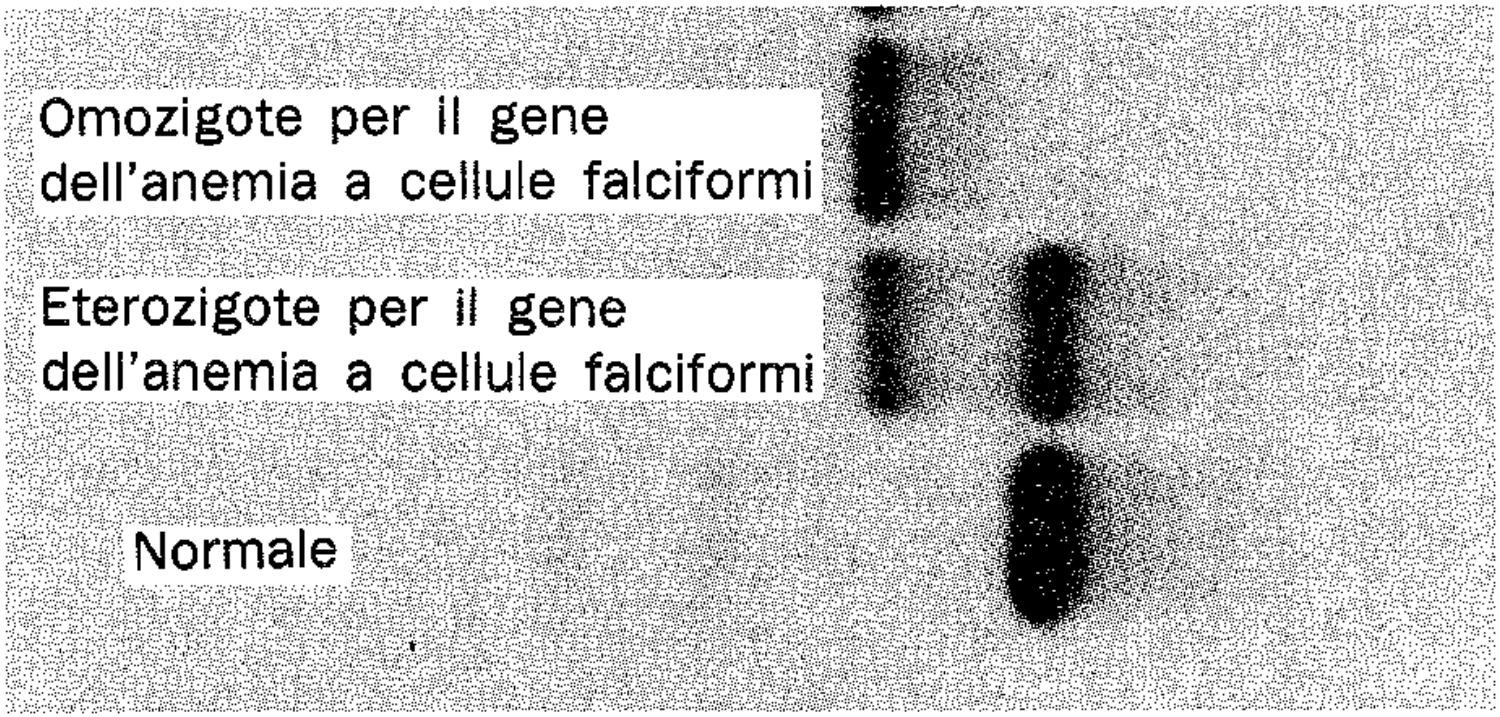
(a)

Interazioni idrofobiche tra catene β
di tetrameri diversi di deossiemoglobina
(forma T)



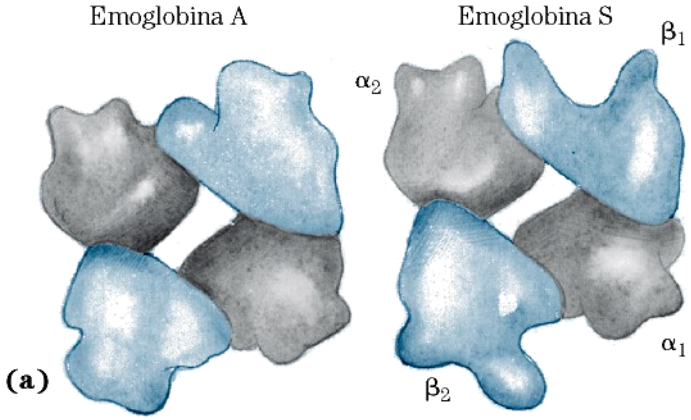
(b)

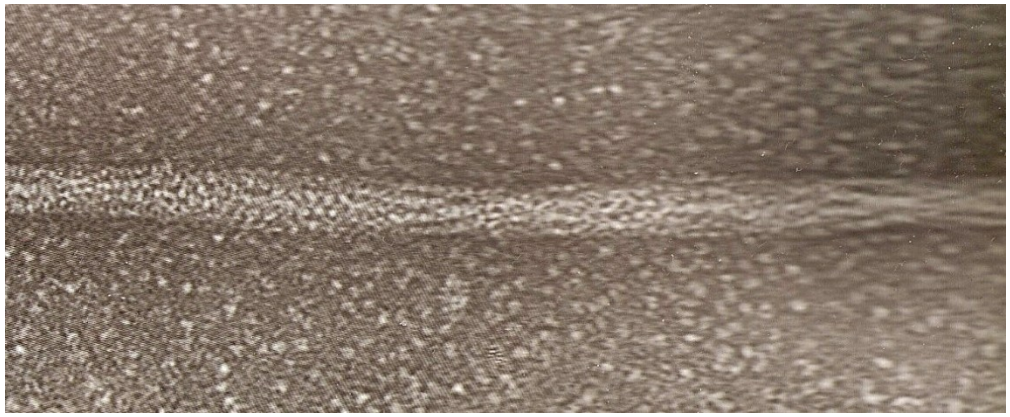
Proprietà elettroforetiche delle emoglobine da individui normali, eterozigoti o omozigoti per il gene dell'anemia falciforme

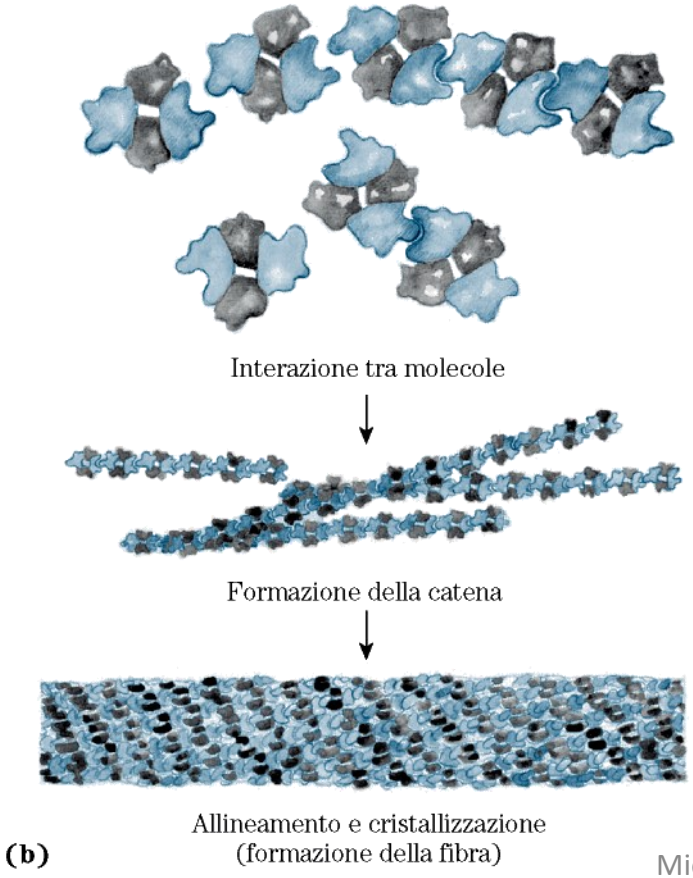


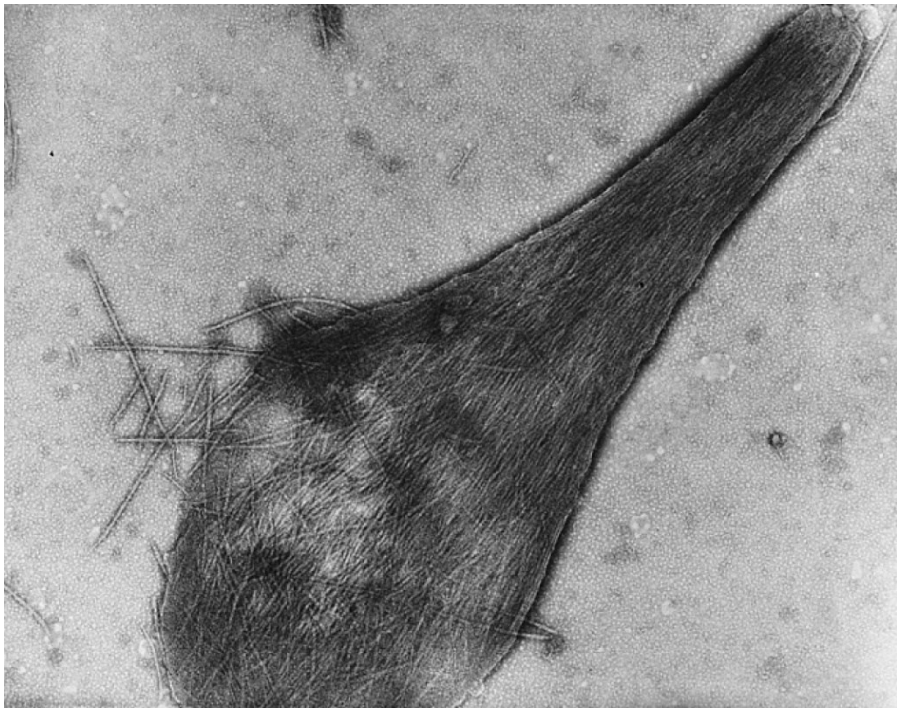
→ Hb normale migra più velocemente verso il polo +

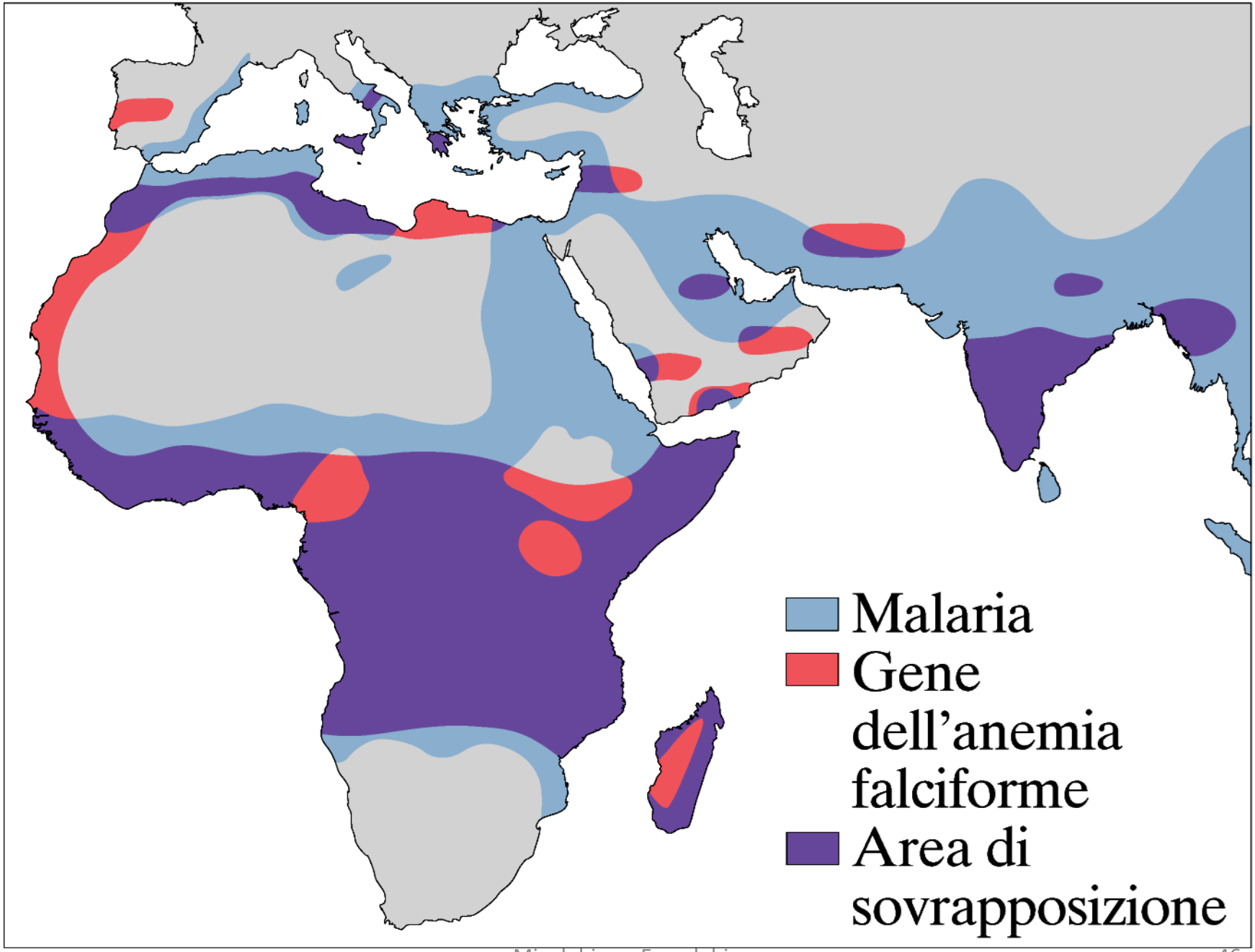
1949 – Linus Pauling HbS = Hb mutante con carica negativa inferiore


Emoglobina A
Emoglobina S
β_1
α_2
α_1
β_2
(a)




Interazione tra molecole
Formazione della catena
Allineamento e cristallizzazione
(formazione della fibra)
(b)




Malaria
Gene dell'anemia falciforme
Area di sovrapposizione

Meccanismo di protezione della HbS rispetto alla malaria

Il parassita della Malaria, il plasmodio, diminuisce di circa 0,4 unità il pH degli eritrociti infettati. La diminuzione di pH promuove la deossigenazione e la polimerizzazione dell'HbS. La distorsione della membrana causa la perdita di K^+. Gli eritrociti danneggiati vanno incontro a lisi e vengono rimossi dalla milza.

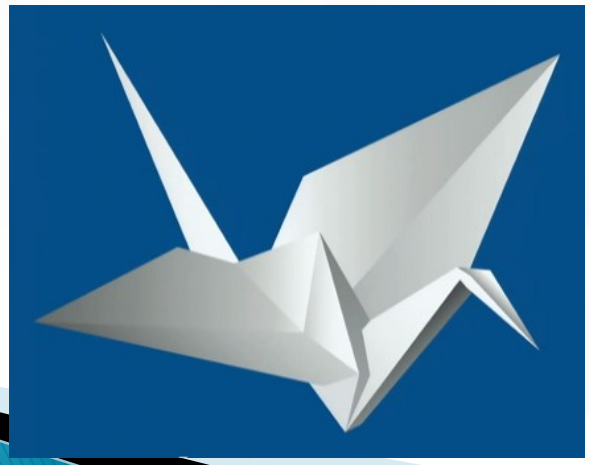
Quando gli eritrociti infettati aderiscono ai capillari (una caratteristica conseguenza dell'infezione, che tende a prevenire l'eliminazione delle cellule infette da parte della milza), la falcizzazione può frammentare per via meccanica il parassita.

Il Plasmodium falciparum sopravvive meno facilmente e si replica di meno negli individui con l'anemia a cellule falciformi

Scopi della biologia strutturale

- Analisi del **dettaglio atomico** delle componenti biologiche di cellule e tessuti → e quindi delle reazioni biologiche
- Comprensione delle relazioni tra **struttura e funzione**
- **Comprensione** dei processi biologici in sede normale, **pre-patologica e patologica**

Solo se so COME FUNZIONA un processo posso CAPIRE cosa è che non lo fa funzionare in condizioni patologiche





PROTEINE

SEQUENZA → STRUTTURA → FUNZIONE

LA **SEQUENZA** DETERMINA LA **STRUTTURA**



LA STRUTTURA DETERMINA LA **FUNZIONE**

Function discovery of cage-like proteins

Oxidative damage

Reactive oxygen species (ROS) are generated as by-products of normal cellular metabolic activities.

All aerobic organisms have evolved a variety of complex defense and repair mechanisms to protect their DNA from oxidative damage because of reactive oxygen species such as $HO\cdot$, O_2 , and H_2O_2.

Pathogenic bacteria have to neutralize H_2O_2 produced by macrophage during infections

Iron storage

Haber-Weiss reaction: $Fe^{2+} + O_2 \rightarrow O_2^{-} + Fe^{3+}$

$$O_2^{-} + e^{-} + 2H^{+} \rightarrow H_2O_2$$

Fenton reaction: $Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^{-} + OH\cdot$

Cage-like protein

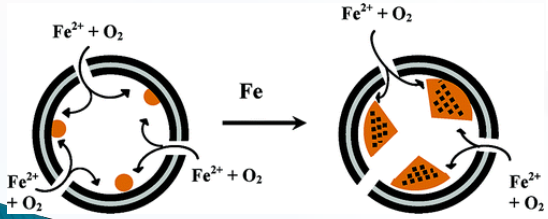
Structural and functional characterization of ferritin and Dps (DNA-binding proteins from starved cells) proteins.

Biological role

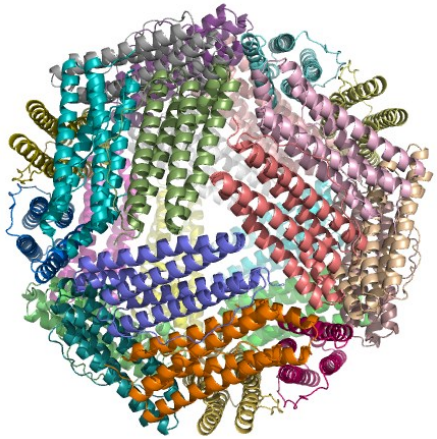
Iron detoxification-sequestration

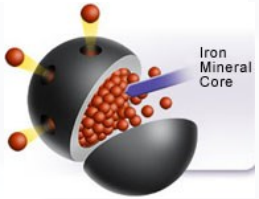
Dps



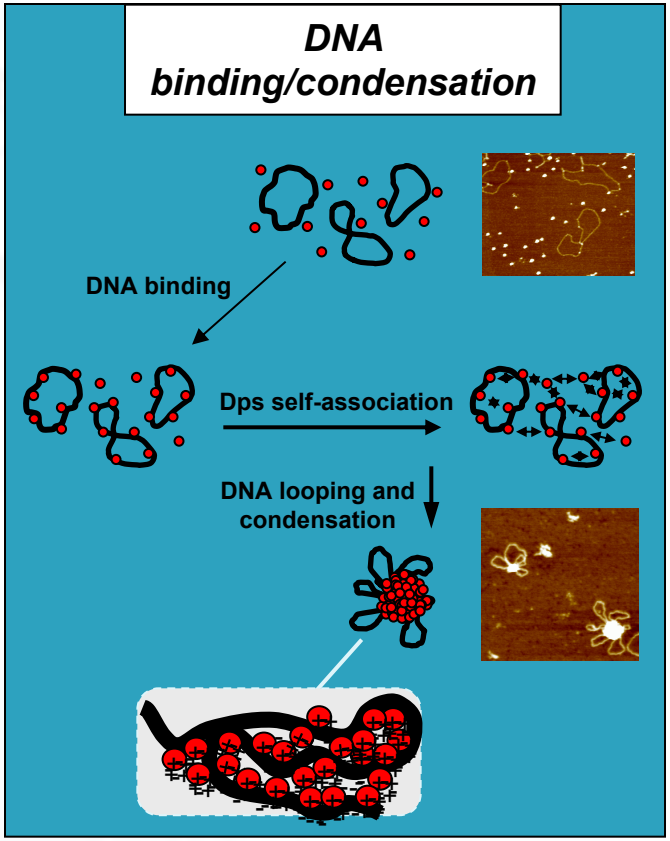


Ferritin





Dps



The dodecameric ferritin from *Listeria innocua* contains a novel intersubunit iron-binding site

Andrea Ilari, Simonetta Stefanini, Emilia Chiancone and Demetrius Tsernoglou

CNR Center of Molecular Biology, Department of Biochemical Sciences "A. Rossi Fanelli", University La Sapienza, Piazzale Aldo Moro 5, 00185 Rome, Italy.

Ferritin is characterized by a highly conserved architecture that comprises 24 subunits assembled into a spherical cage with 432 symmetry. The only known exception is the dodecameric ferritin from *Listeria innocua*. The structure of *Listeria* ferritin has been determined to a resolution of 2.35 Å by molecular replacement, using as a search model the structure of Dps from *Escherichia coli*. The *Listeria* 12-mer is endowed with 23 symmetry and displays the functionally relevant structural features of the ferritin 24-mer, namely the negatively charged channels along the three-fold symmetry axes that serve for iron entry into the cavity and a negatively charged internal cavity for iron deposition. The electron density map shows 12 iron ions on the inner surface of the hollow core, at the interface between monomers related by two-fold axes. Analysis of the nature and stereochemistry of the iron-binding ligands reveals strong similarities with known ferroxidase sites. The *L. innocua* ferritin site, however, is the first described so far that has ligands belonging to two different subunits and is not contained within a four-helix bundle.

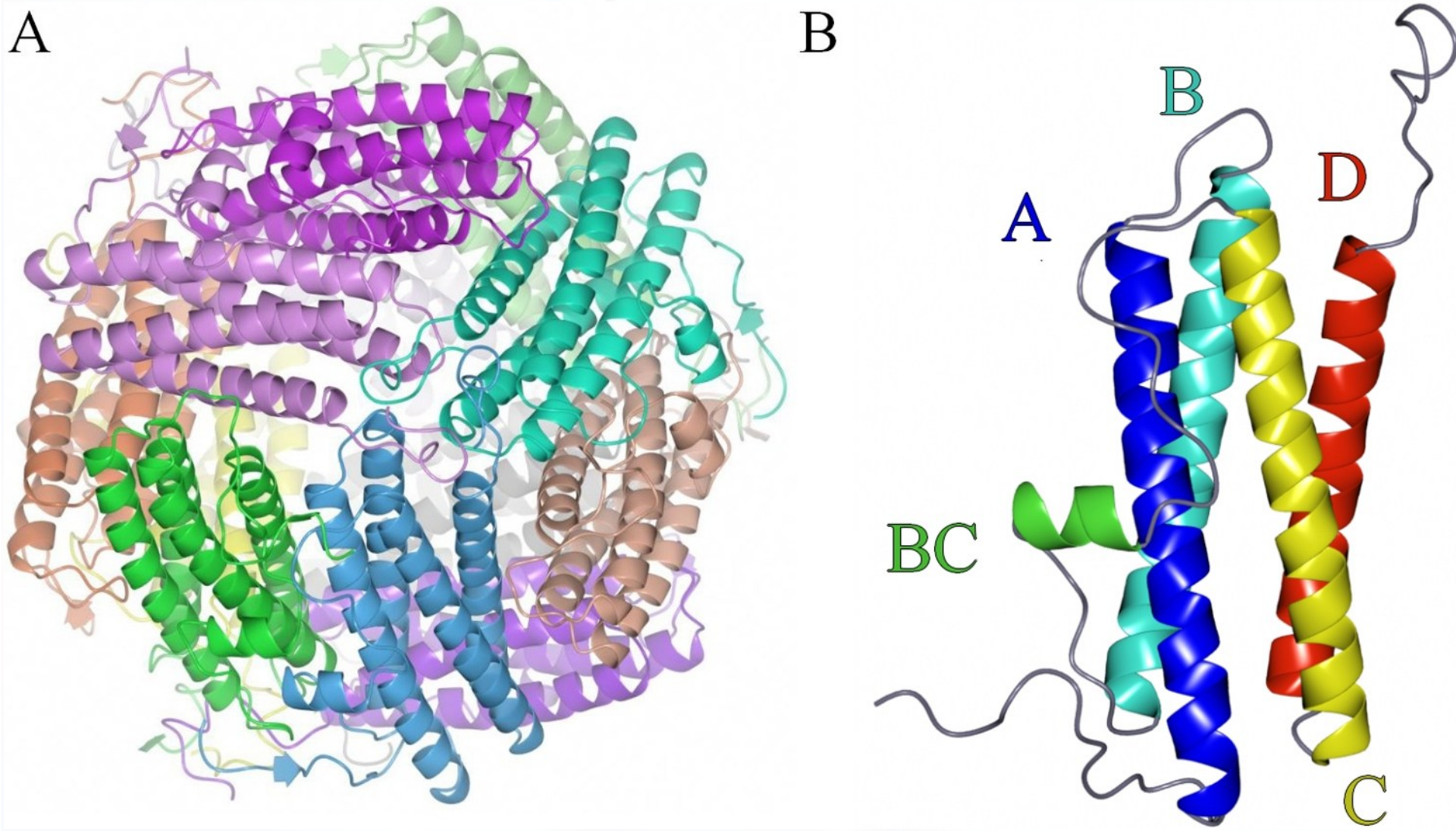
Nearly all forms of life require iron but must counter its unfavorable chemical properties that lead to formation of insoluble ferric-hydroxide polymers and toxic free radical species. Therefore iron is stored in ferritins that sequester the metal in a nontoxic and bioavailable form[1]. All ferritins share the same, highly conserved structure, with the exception of the protein extracted from *Listeria innocua*, the only known ferritin from a Gram-positive bacterium[2,3].

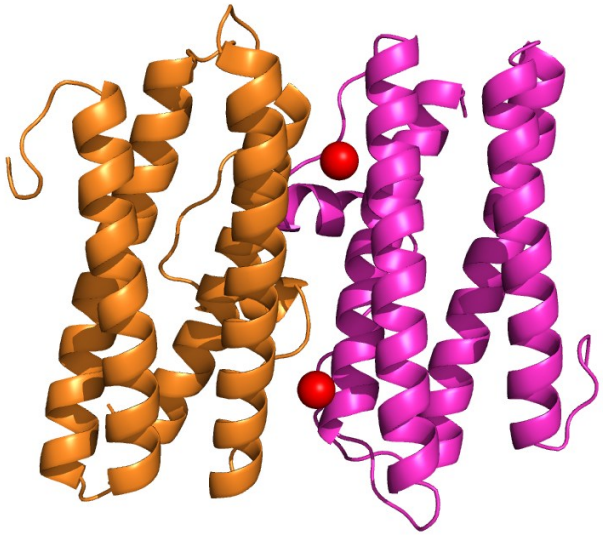
Typically, ferritins are oligomers of 24 identical or similar subunits (M_r 19–21 kDa) that assemble into a spherical shell (M_r 450–500 kDa, external diameter 120 Å) characterized by 432 symmetry[1]. The subunits share the same tertiary fold consisting of a four-helix bundle (helices A–D) with a fifth short helix (E helix) lying at an angle of about 60° to the bundle axis[4–6]. The apoferritin shell can store up to 4,500 iron atoms in the form of ferric hydroxy-phosphate micelles. However, incorporation of iron occurs when the metal is furnished to the protein as Fe^{2+} in the presence of molecular oxygen[1]. All natural ferritins therefore contain H-type subunits that carry within the four-helix bundle a highly conserved ferroxidase center that allows formation of a di-iron species, an intermediate in the iron oxidation and uptake process. Ferritins of higher vertebrates contain an additional type of subunit, called L, that forms hetero-oligomers with the H subunits. L-type chains do not possess a ferroxidase site, but contain a cluster of acidic residues that protrudes from the B helix into the apoferritin cavity and is thought to facilitate nucleation of the iron core[1].

Recently, Bozzi *et al.*[2] isolated from the Gram-positive bacterium *L. innocua* an oligomeric, spherical protein complex containing up to 50–100 iron atoms per oligomer and the functional properties of an authentic ferritin. Thus, at neutral pH values *Listeria* ferritin accelerates Fe^{2+} oxidation about four-fold with respect to auto oxidation[7]. This ferroxidase activity is about one third to one fourth that of mammalian recombinant H-chain ferritin[8,9] and *E. coli* bacterioferritin[10]. As in classical ferritins, the oxidized iron is sequestered inside the protein cav-

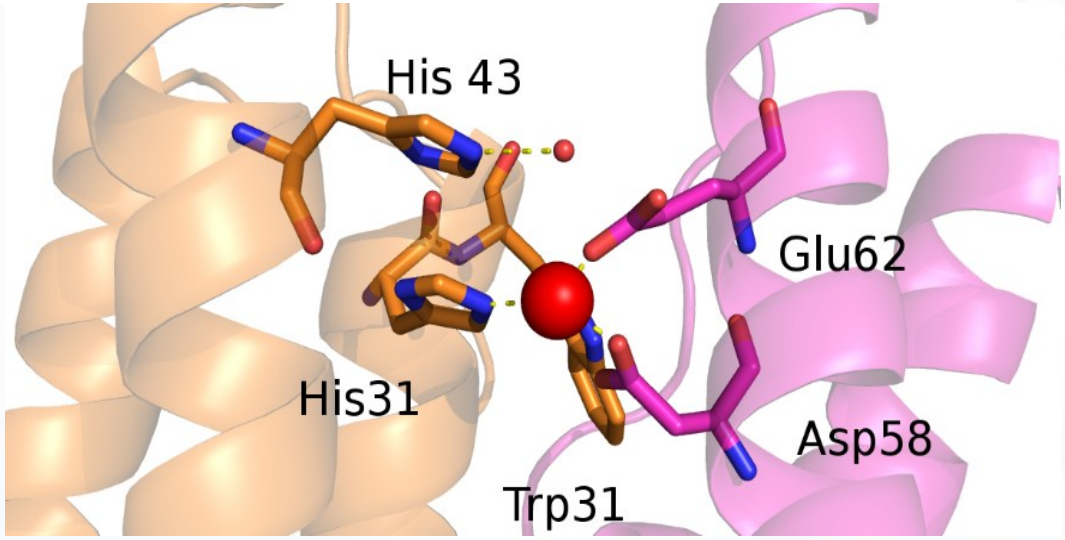
IBPM CNR
National Research Council of Italy

Structure level of DPS



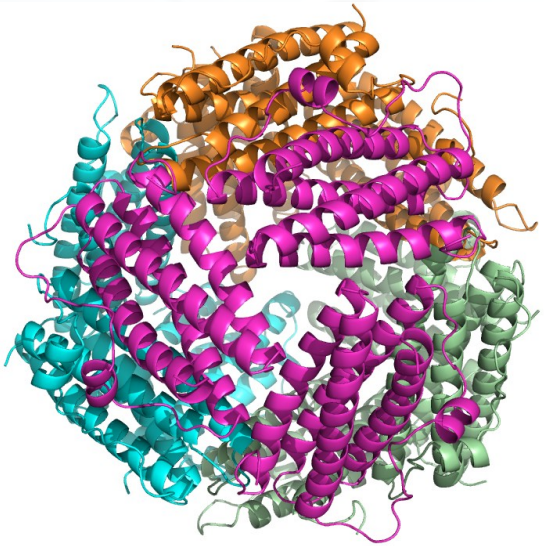


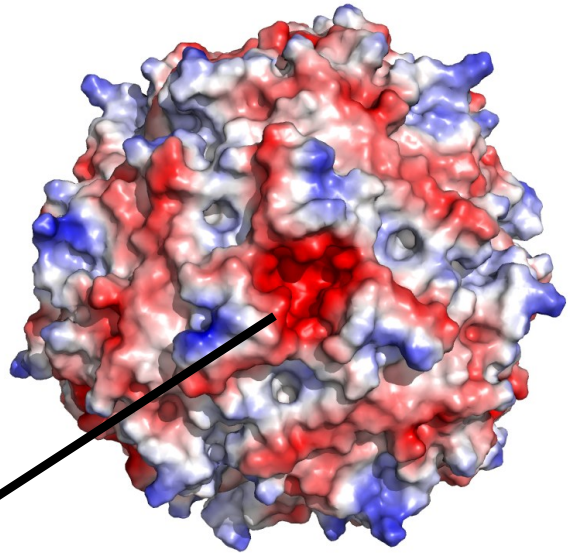
Ferroxidase centre at the interface between the two fold symmetry related subunits

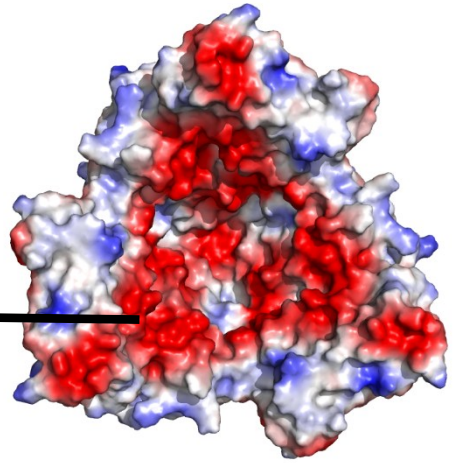


$$2Fe^{2+} + H_2O_2 + 2H_2O \rightarrow 2Fe(III)OOH_{(core)} + 4H^+$$

The L. innocua miniferritin conserves all the structural features necessary for iron entry and storage





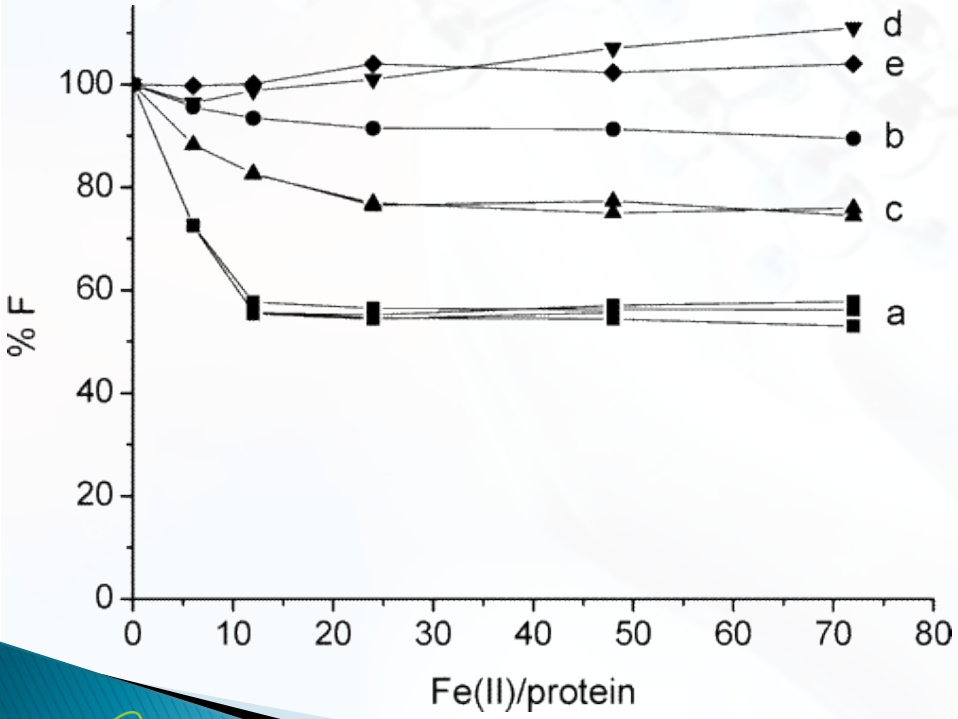


Iron channel at the three-fold symmetry axis

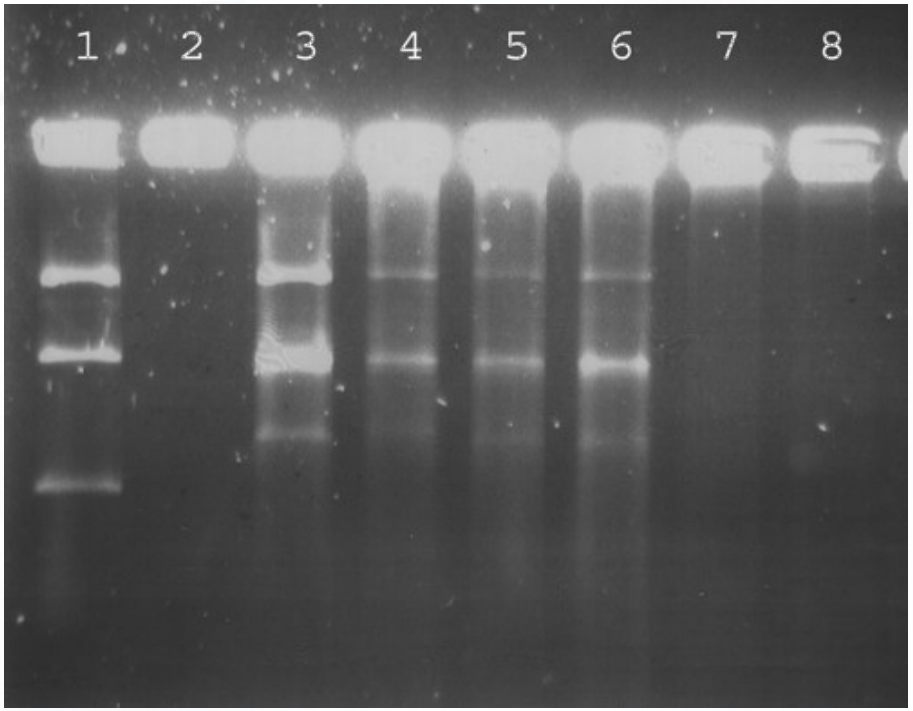
Iron micelle nucleation sites.
Incorporation up to 400 iron ions

Mutant of the Ferroxidase centre do not protect DNA from oxidative damage

Fluorimetric titration in anaerobiosis of L. innocua Dps and its ferroxidase center variants with Fe(II). Curves: (a) wt Dps; (b) H31G; (c) H43G; (d) H31G-H43G; (e) H31G-H43G-D58A. The protein concentration was 1.5 μM in 50 mM Mops-NaOH buffer, pH 7.0. All of the spectra were recorded at 300-400 nm and 25 °C. Excitation was at 290 nm



1. Plasmid
2. Plasmid exposed to FeSO4 and then to H_2O_2;
6. plasmid with Dps exposed to FeSO4 and then to H_2O_2;
7, conditions as in lane 6 but with 3 μM H31G-H43G;
8, conditions as in lanes 6 and 7 but with 3 μM H31G-H43G-D58A.



IBPM CNR

National Research Council of Italy

An improved Bottom-up method



Biomimetic or Bioinspired approach

Application of biological systems operating in nature to the design and synthesis of metal nanoparticles with high functionalities

protein cage architectures

highly symmetrical structures:
spherical shapes enclosing a central hollow space

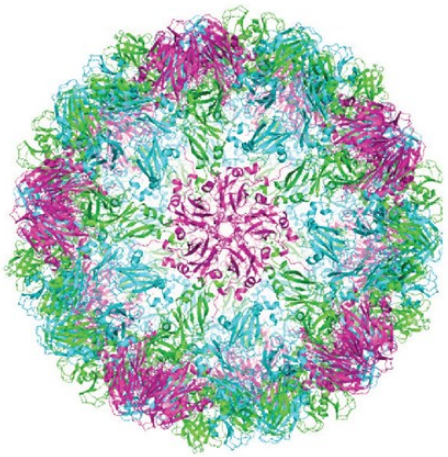
⇒ robust template to restrain particle growth
⇒ coating to prevent NPs aggregation



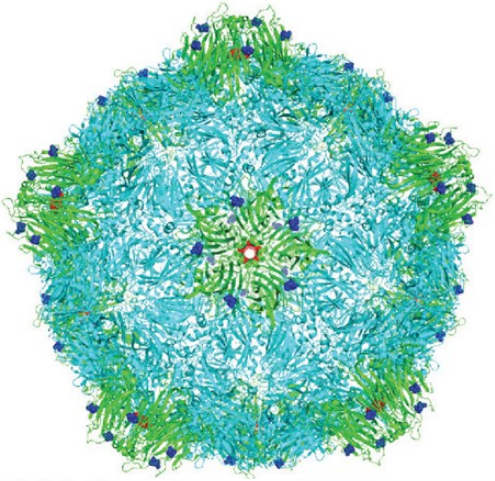
Dps: 4 nm



Ferritin: 8 nm

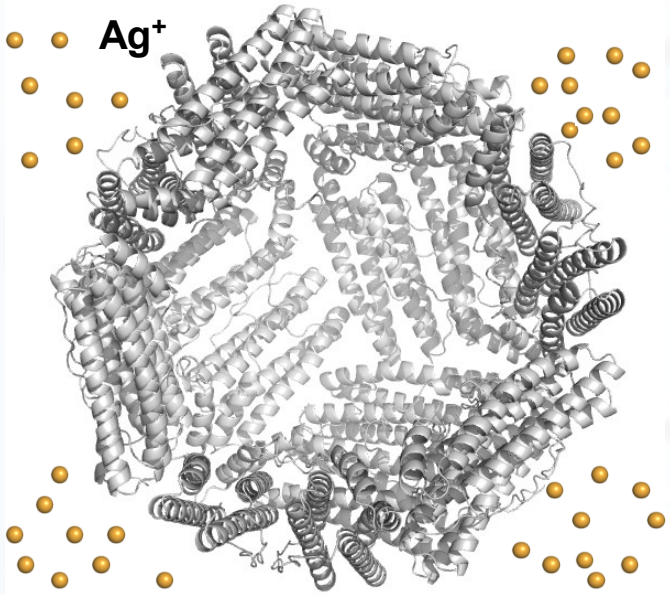


CCMV virus: 22 nm



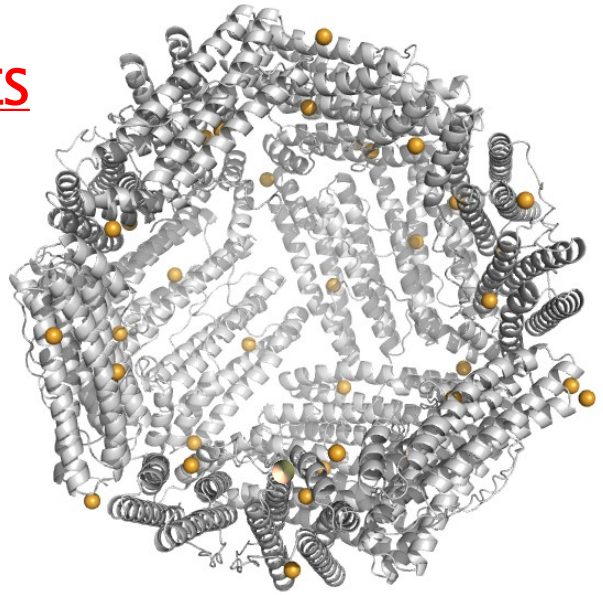
CMV virus: 24 nm

SILVER NANOPARTICLES



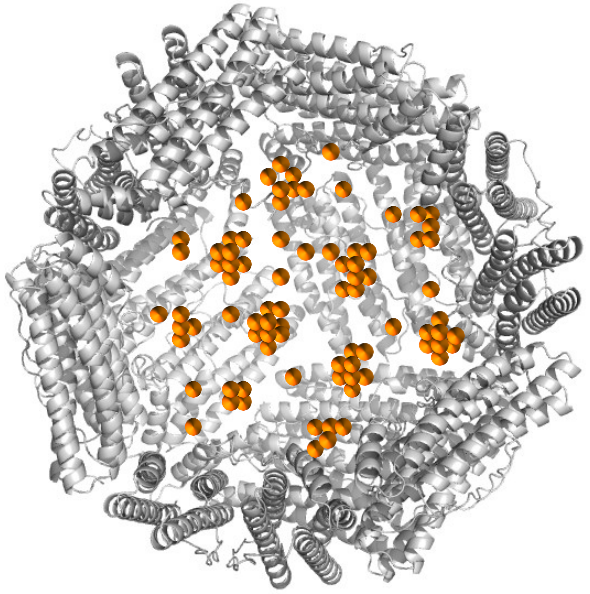
(A)
Metal binding

10 mM Hepes pH 7.8
15 min



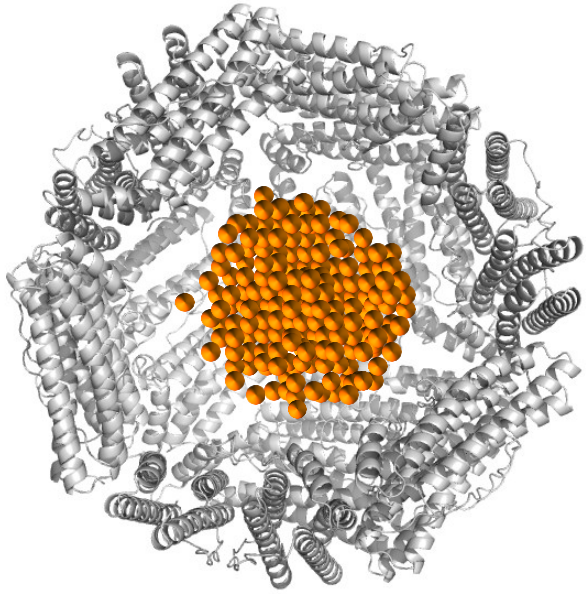
(B)
Redox reaction and nucleation

$NaBH_4$



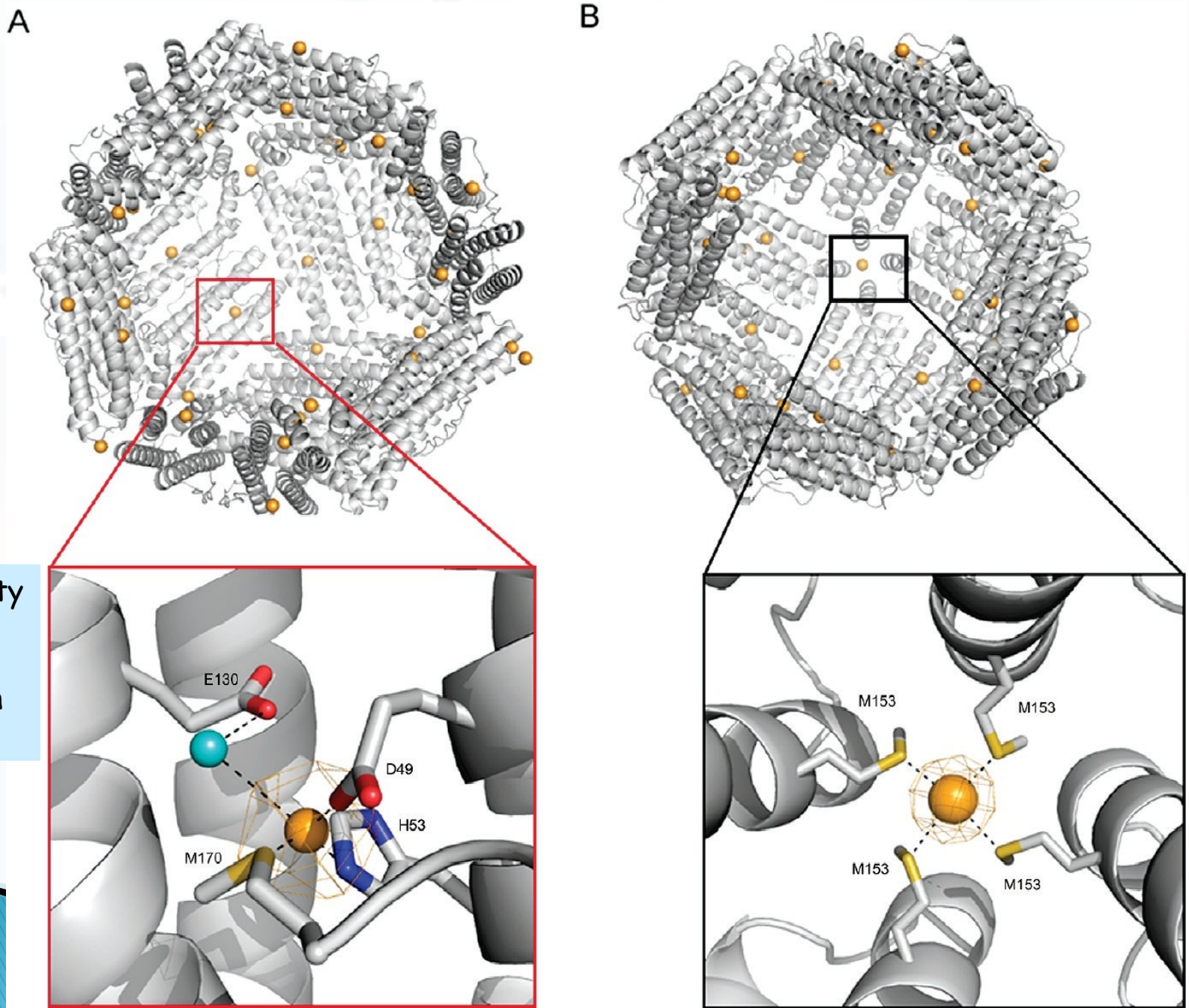
(C)
Ag nanoparticles formation

15 min



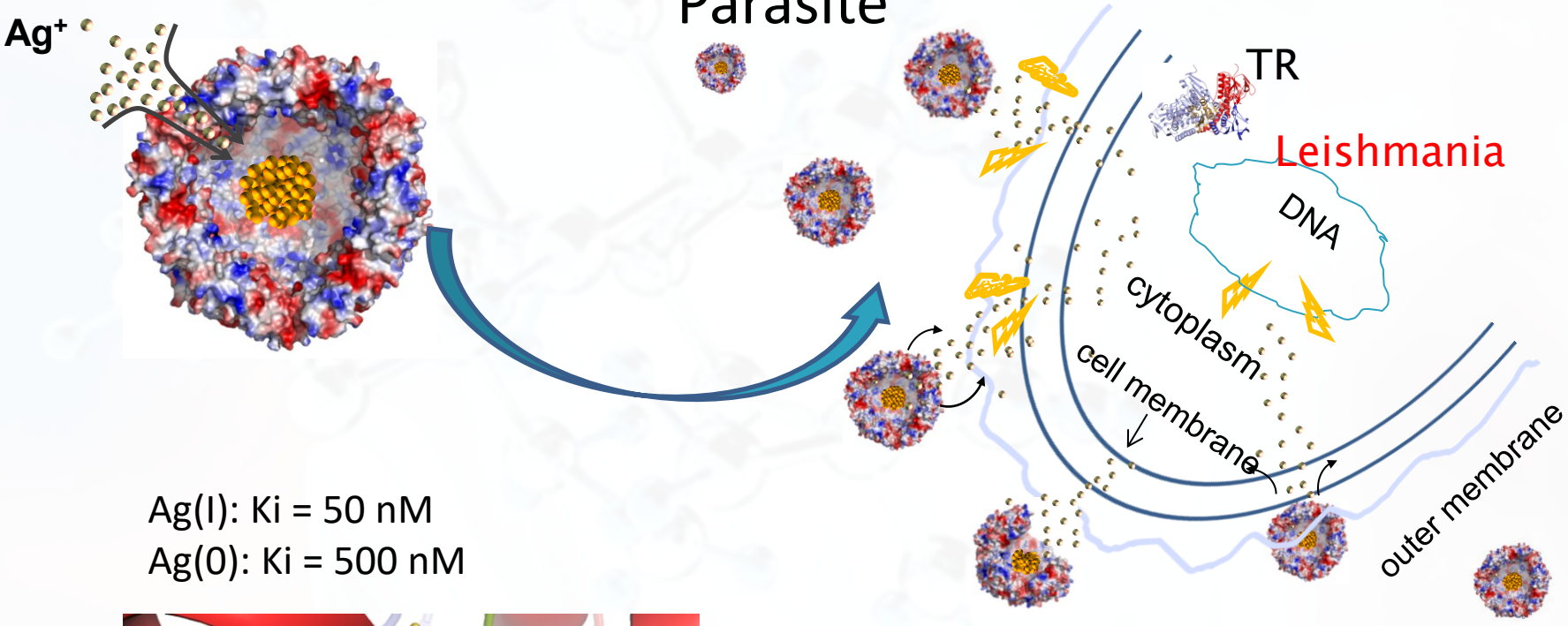


X-ray crystal structure of silver-containing ferritin from *P. furiosus*

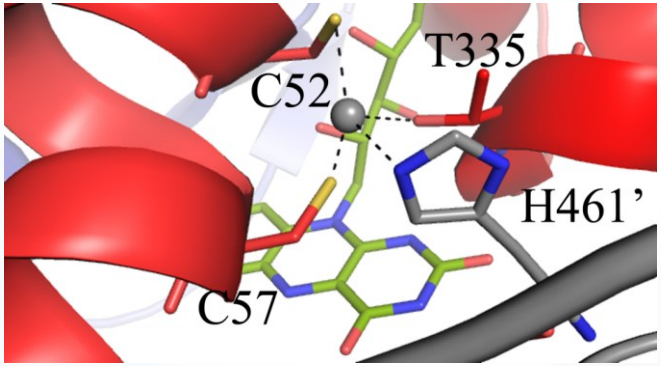


High affinity silver binding/ nucleation sites

Ag inhibits TR with very high efficiency and kill Leishmania Parasite

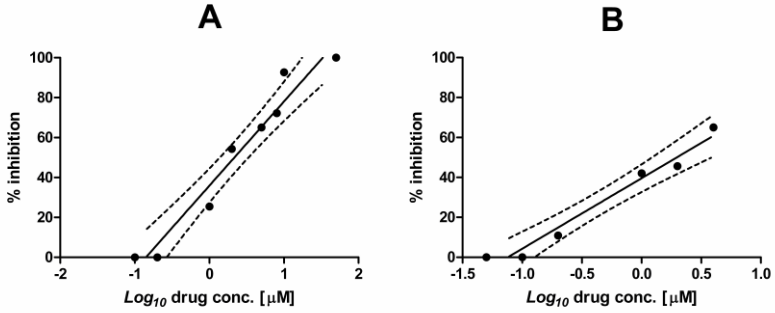


Ag(I): Ki = 50 nM
Ag(0): Ki = 500 nM



Promastigote
Inhibition 2.18 µM

Amastigote
Inhibition 1.76 µM



doi:10.1006/jmbi.2001.5417 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2002) **317**, 447–458

JMB

The Crystal Structure of the Sorcin Calcium Binding Domain Provides a Model of Ca^{2+}-dependent Processes in the Full-length Protein

Andrea Ilari[1], Kenneth A. Johnson[1], Vassilios Nastopoulos[2] Daniela Verzili[1], Carlotta Zamparelli[1], Gianni Colotti[1] Demetrius Tsernoglou[1] and Emilia Chiancone[1]*

[1]*CNR, Centro Studio sulla Biologia Molecolare and Dipartimento di Scienze Biochimiche, Università ''La Sapienza'', P.le Aldo Moro 5 00185, Roma, Italy*

[2]*Department of Chemistry University of Patras, 265 00 Patras, Greece*

**Corresponding author*

Sorcin is a 21.6 kDa calcium binding protein, expressed in a number of mammalian tissues that belongs to the small, recently identified penta-EF-hand (PEF) family. Like all members of this family, sorcin undergoes a Ca^{2+}-dependent translocation from cytosol to membranes where it binds to target proteins. For sorcin, the targets differ in different tissues, indicating that it takes part in a number of Ca^{2+}-regulated processes. The sorcin monomer is organized in two domains like in all PEF proteins: a flexible, hydrophobic, glycine-rich N-terminal region and a calcium binding C-terminal domain. *In vitro*, the PEF proteins are dimeric in their Ca^{2+}-free form, but have a marked tendency to precipitate when bound to calcium. Stabilization of the dimeric structure is achieved by pairing of the uneven EF-hand, EF5. Sorcin can also form tetramers at acid pH.

The sorcin calcium binding domain (SCBD, residues 33-198) expressed in *Escherichia coli* was crystallized in the Ca^{2+}-free form. The structure was solved by molecular replacement and was refined to 2.2 Å with a crystallographic *R*-factor of 22.4 %. Interestingly, the asymmetric unit contains two dimers.

The structure of the SCBD leads to a model that explains the solution properties and describes the Ca^{2+}-induced conformational changes. Phosphorylation studies show that the N-terminal domain hinders phosphorylation of SCBD, i.e. the rate of phosphorylation increased twofold in the absence of the N-terminal region. In addition, previous fluorescence studies indicated that hydrophobic residues are exposed to solvent upon Ca^{2+} binding to full-length sorcin. The model accounts for these data by proposing that Ca^{2+} binding weakens the interactions between the two domains and leads to their reorientation, which exposes hydrophobic regions facilitating the Ca^{2+}-dependent binding to target proteins at or near membranes.

© 2002 Elsevier Science Ltd.

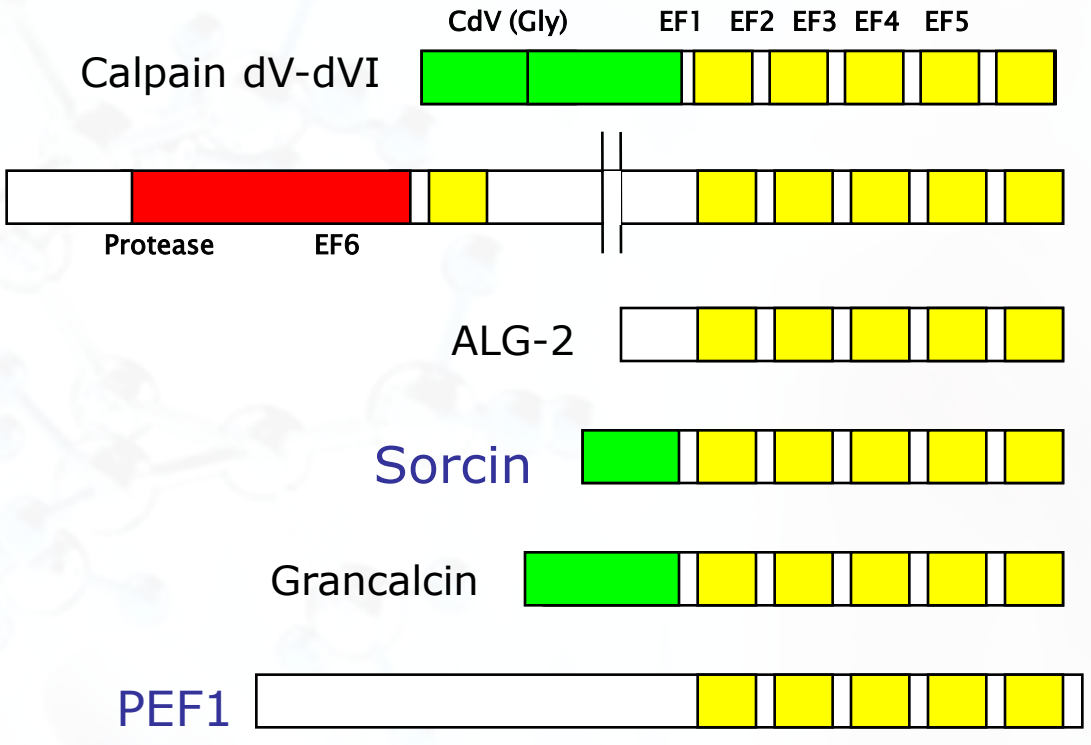
Keywords: Ca^{2+}-binding domain; Ca^{2+}-dependent conformational changes; penta-EF hand family (PEF); phosphorylation site; sorcin

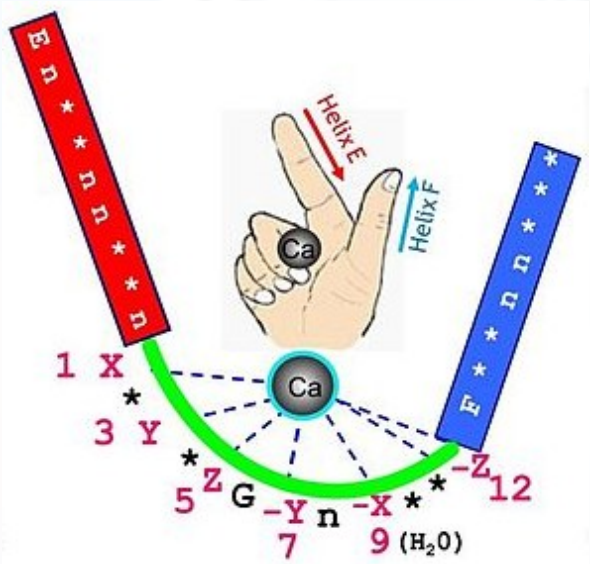
IBPM CNR

National Research Council of Italy

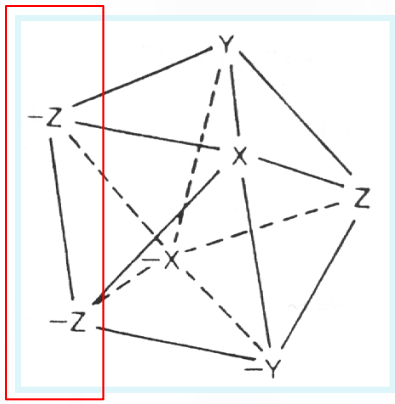
SORCIN (SOluble Resistance-related Calcium binding proteIN) belongs to the penta EF-hand family (PEF)

PEF, formed by two domain:
- **Calcium binding domain (5 EF-hands)**
- **Flexible N-terminal domain**

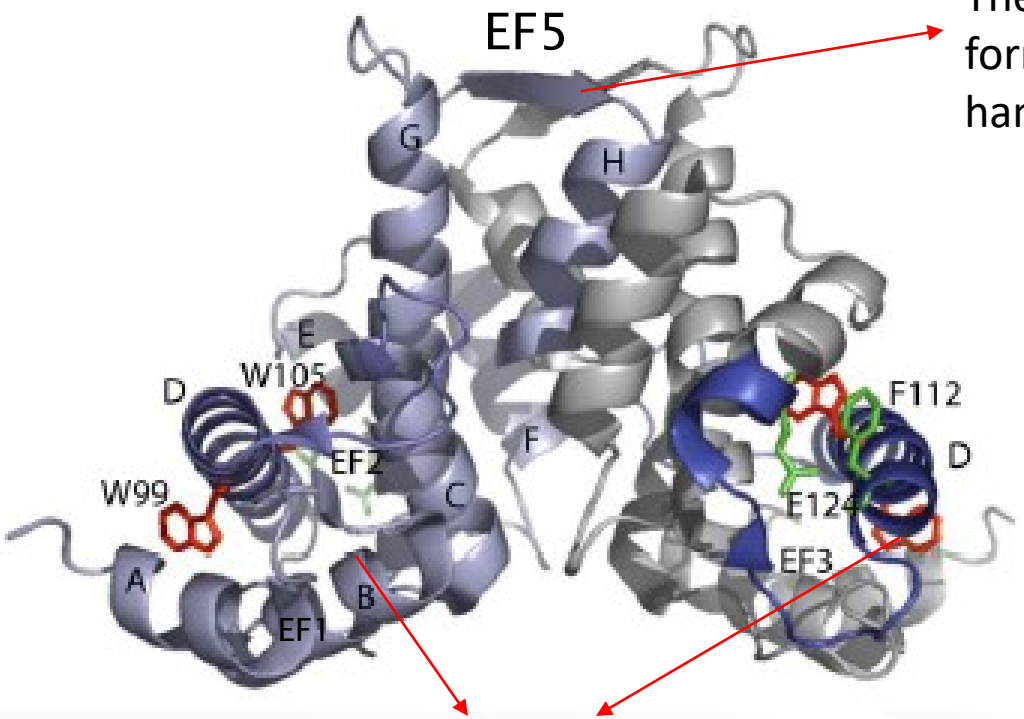








Crystal structure of the hamster SCBD (Sorcin Calcium Binding Domain): sorcin is a homodimer



The monomer-monomer interface is formed by coupling of the two odd EF5 hands of two monomers.

Potentially active EF hand sites:
EF1 and EF2 are structurally coupled;
EF3 is coupled with EF4 (not binding Ca(II))

doi:10.1006/jmbi.2001.5417 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2002) **317**, 447–458

JMB

The Crystal Structure of the Sorcin Calcium Binding Domain Provides a Model of Ca^{2+}-dependent Processes in the Full-length Protein

Andrea Ilari[1], Kenneth A. Johnson[1], Vassilios Nastopoulos[2]
Daniela Verzili[1], Carlotta Zamparelli[1], Gianni Colotti[1]
Demetrius Tsernoglou[1] and Emilia Chiancone[1*]

Sorcin partecipates in several calcium-mediated signal transduction processes

- **Sorcin is an important regulator of cardiac contraction**

- **Sorcin is overexpressed in cancer multidrug resistant cells**
  inhibition of sorcin expression by sorcin-targeting RNA interference led to reversal of drug resistance in different cell lines.

- **Sorcin is one of the most expressed calcium binding proteins in the brain**
  (amygdala, prefrontal cortex, hypothalamus)

Sorcin is an essential protein: Sorcin silencing blocks cell cycle progression in mitosis and induces cell death by triggering apoptosis.

Sorcin interacts with more than 224 proteins as established by "in vitro" Protoarray experiments

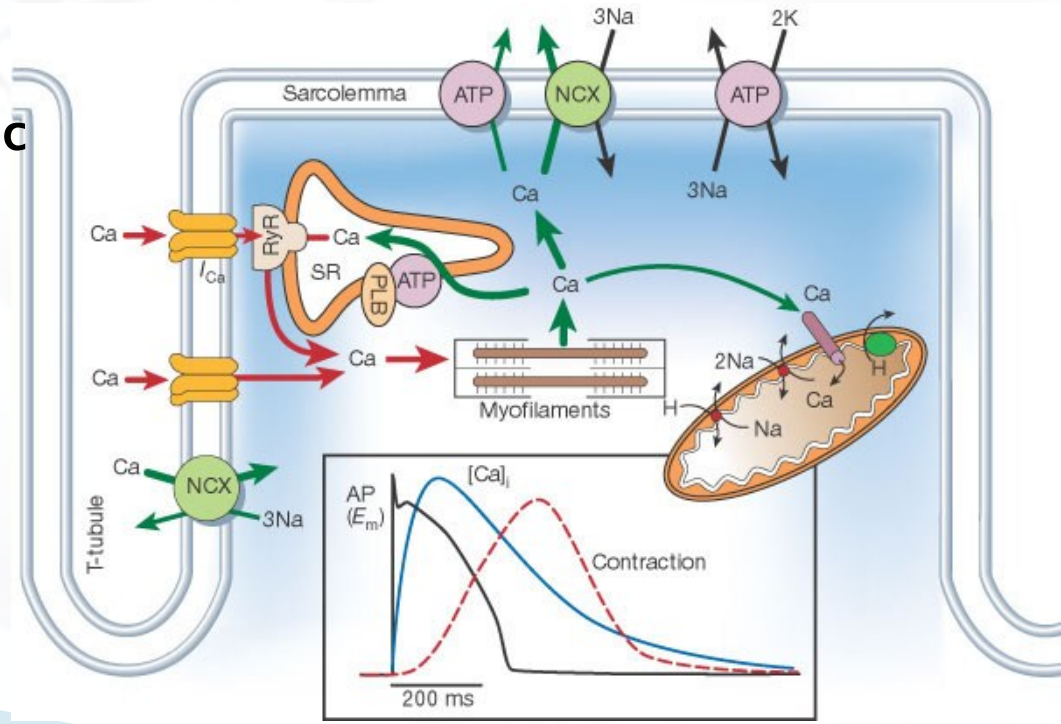
Sorcin network of interactions in the heart has been elucidated

In cardiac and skeletal muscle, sorcin, upon calcium binding, interacts with:

Ryanodine receptor (RyR)	inactivation
Na^+-Ca^{2+} exchanger (NCX)	activation
SERCA (Ca^{2+}-ATPase of the Sarcoplasmic Reticulum)	activation

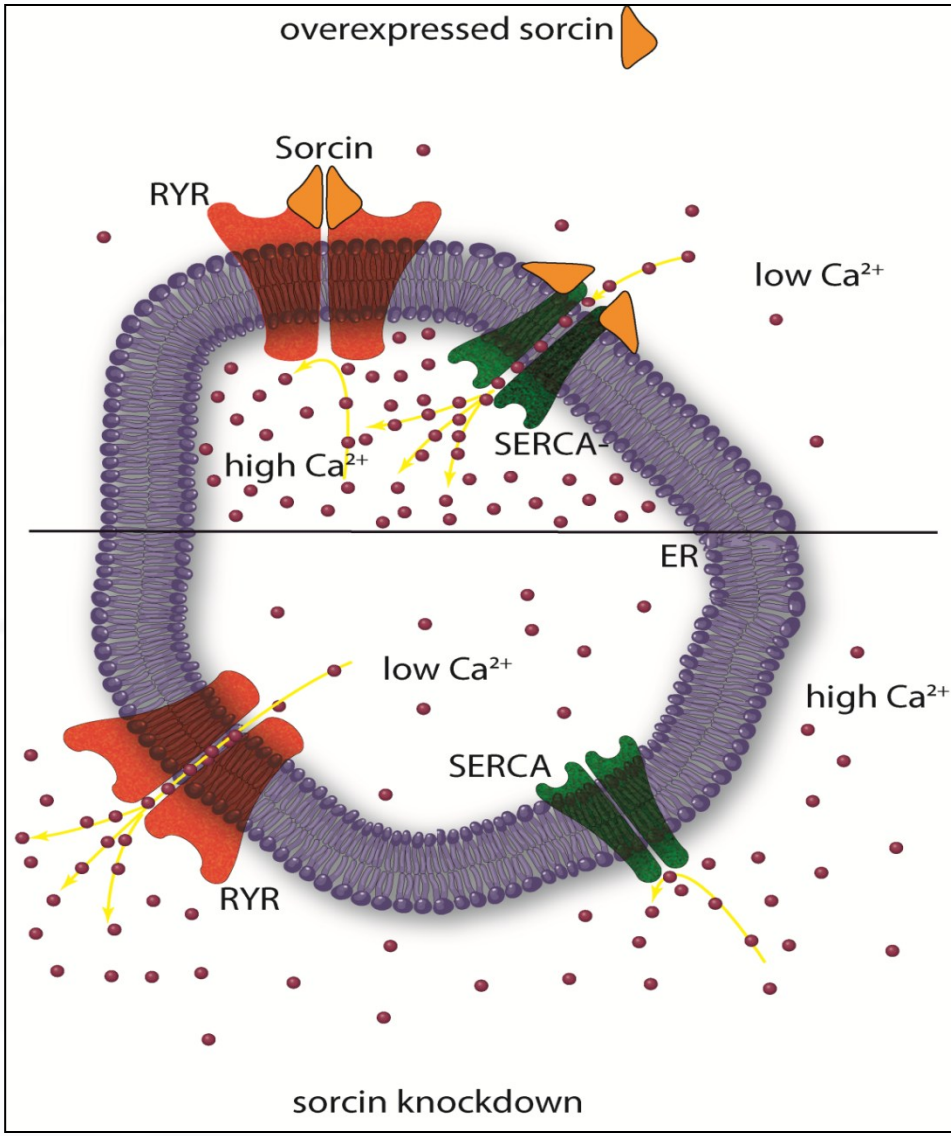
Sorcin activation upon calcium binding diminishes the cytoplasmic concentration of Ca(II)

Heart relaxation



Sorcin expression levels regulate Ca^{2+} concentration and dimensions of ER-derived vesicles.

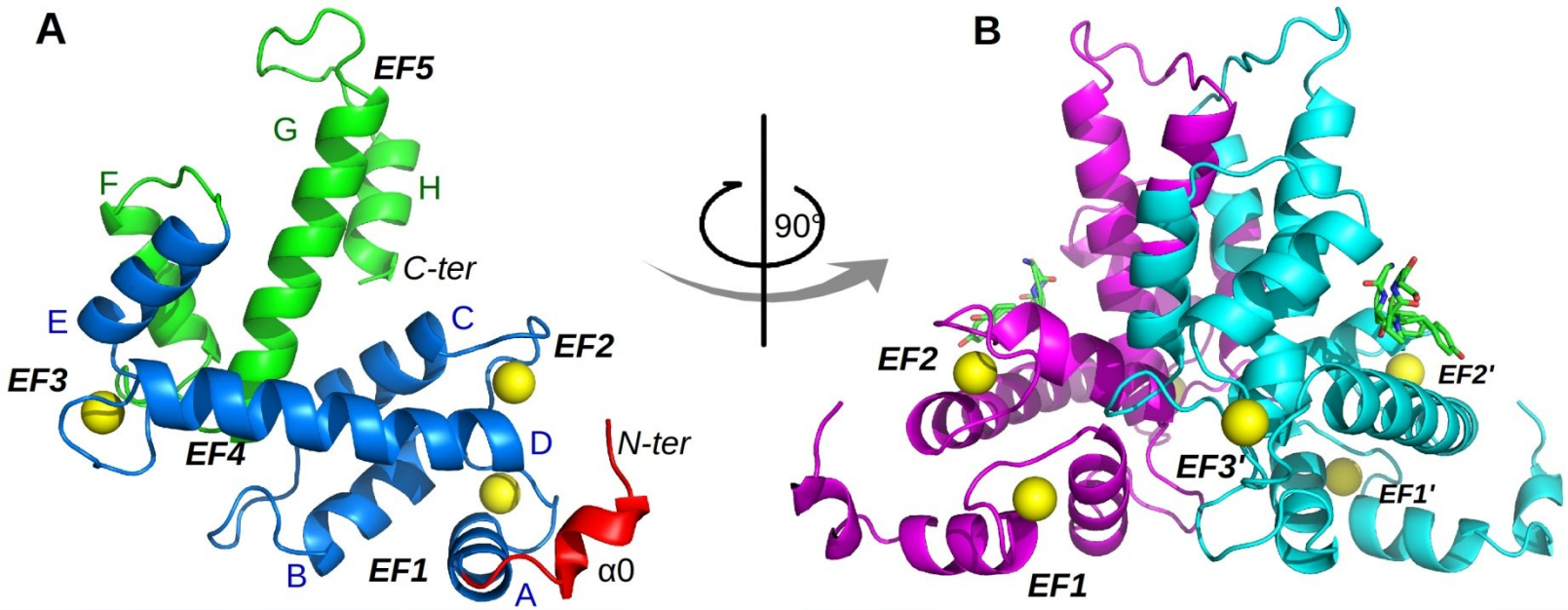
Sorcin increases the concentration Ca (II) in ER by inhibiting RYR and activating SERCA (Sarco-Endoplasmic Reticulum Calcium ATPase).
Sorcin overexpression (top) alters calcium homeostasis
in the cell, increasing the ion content of the ER, and decreasing
cytosolic free Ca (II) concentration.
Sorcin knockdown (bottom) has opposite effects.



Lalioti et al. PLOS ONE 2014, vol. 9, e85438

Sorcin is a homodimer

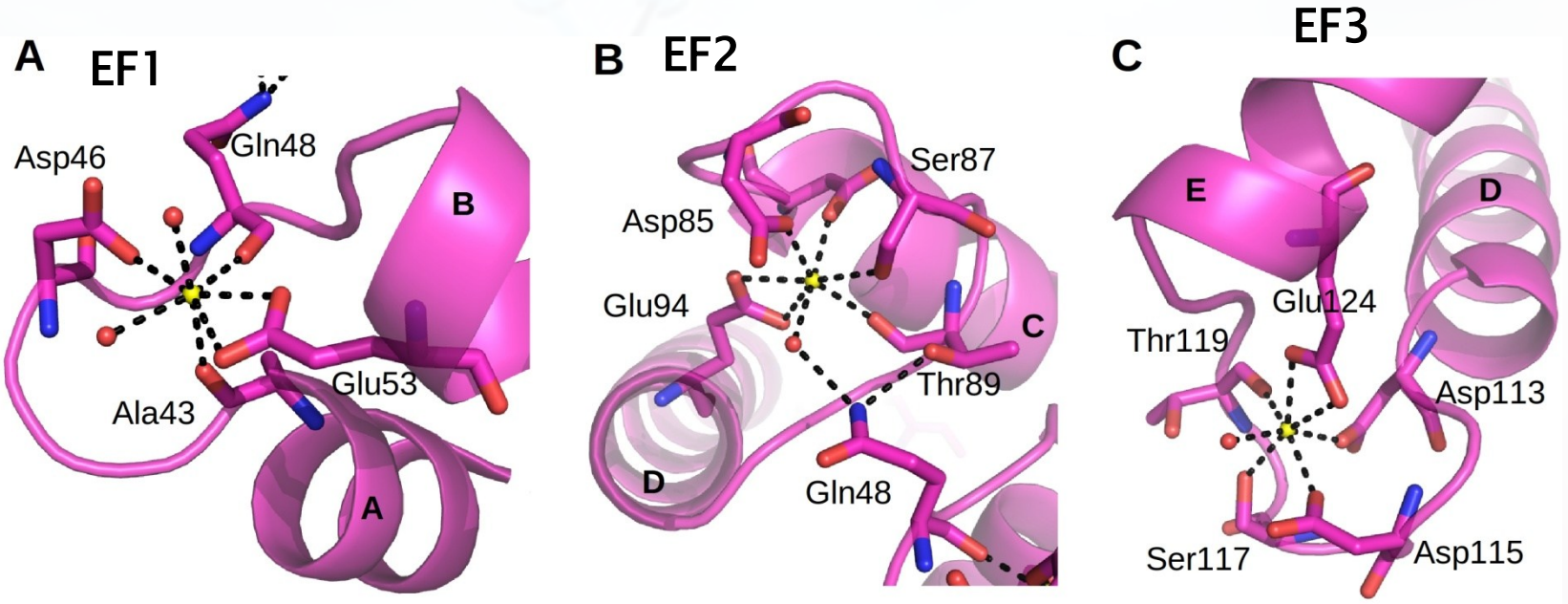
•Sorcin is formed by two domain: SCBD and N-terminal domain (in the structure only few residues are visible (in red))



•SCBD domain is formed by five EF hands and can be divided in two regions EF1-EF3 and EF4-EF5

Ilari et al., Scientific Reports

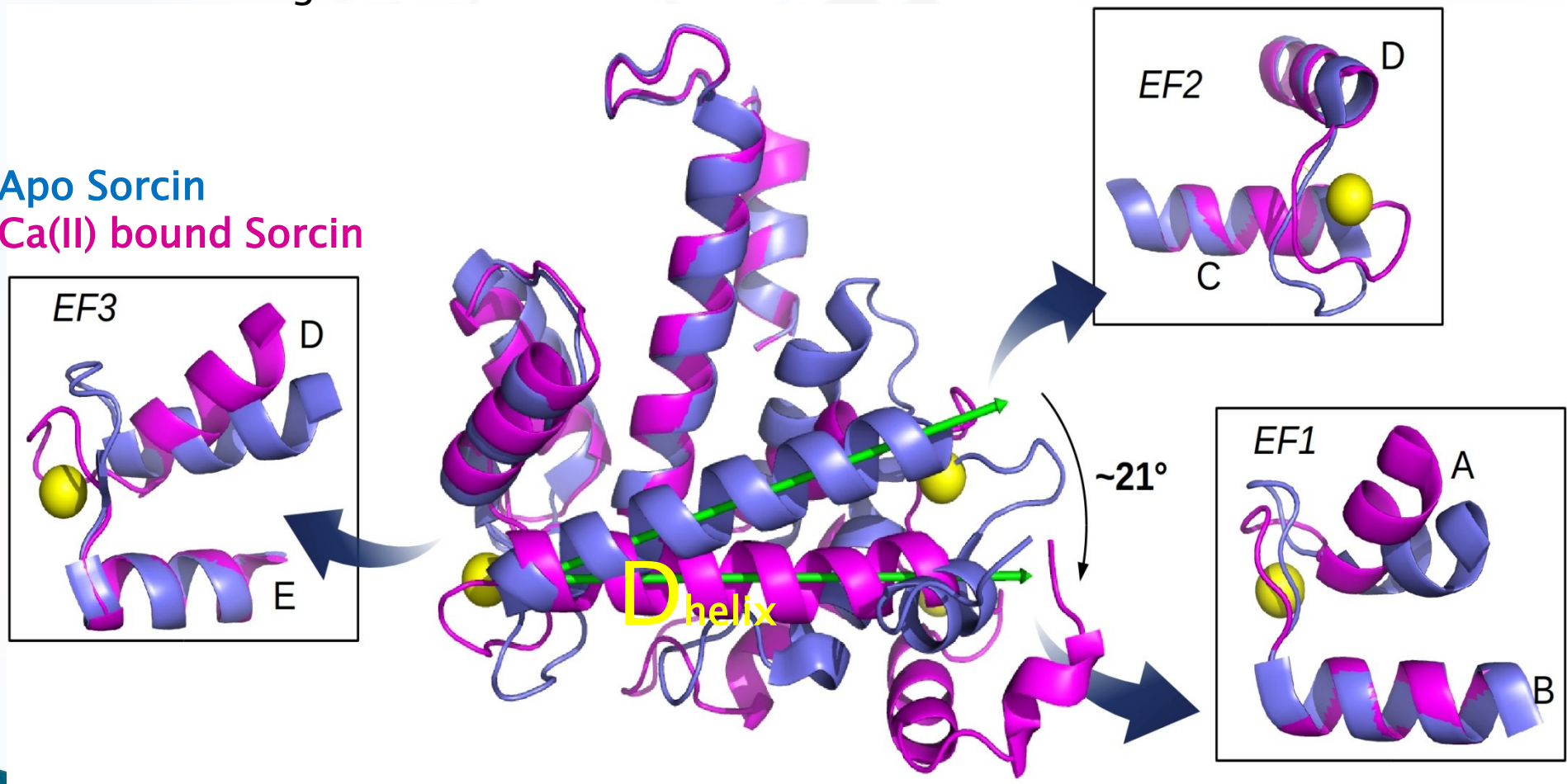
Sorcin can bind only three calcium ions



		x	y	z	-x	-y	-z
CaSor	EF1	Ala43 (O)	H_2O	Asp46 (OD1)	H_2O (Asp50)	Gln48 (O)	Glu53 (OD1-OD2)
	(Å)	2.35	2.45	2.36	2.42	2.30	2.49, 2.59
	EF2	Asp83 (OD1)	Asp85 (OD1)	Ser87 (OG)	H_2O (Gln48)	Thr89 (O)	Glu94 (OD1-OD2)
	(Å)	2.34	2.36	2.57	2.36	2.43	2.51, 2.54
	EF3	Asp113 (OD1)	Asp115 (OD1)	Ser117 (OG)	H2O (Asp121)	Thr119 (O)	Glu124 (OD1-OD2)
	(Å)	2.33	2.39	2.34	2.26	2.28	2.62, 2.53

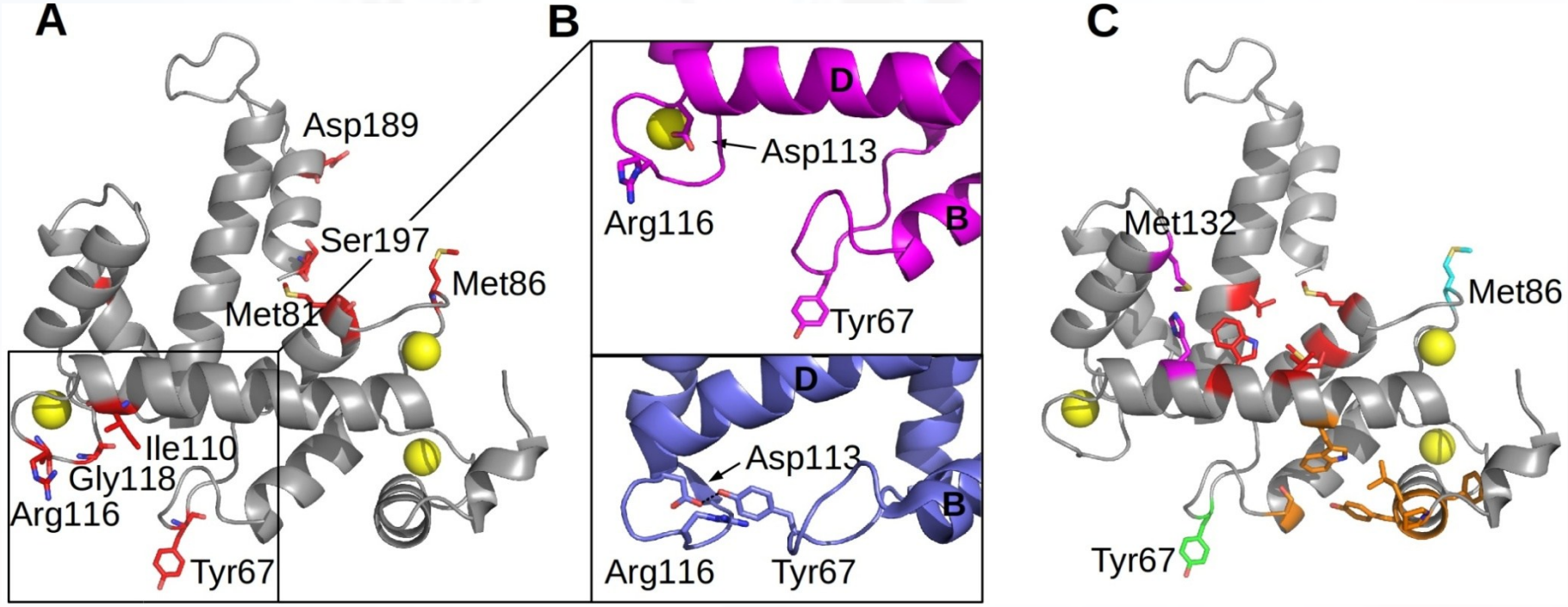
Upon calcium binding sorcin undergoes a big conformational change

The helix D rotate of about 21°
As a consequence the EF1-EF3 region rotate in respect with
The EF4-EF5 region



The binding of Ca(II) to EF3 is responsible for the sorcin big conformational change

In apoSor Tyr67 forms a hydrogen bond with Asp113.
In CaSor the hydrogen bond is broken

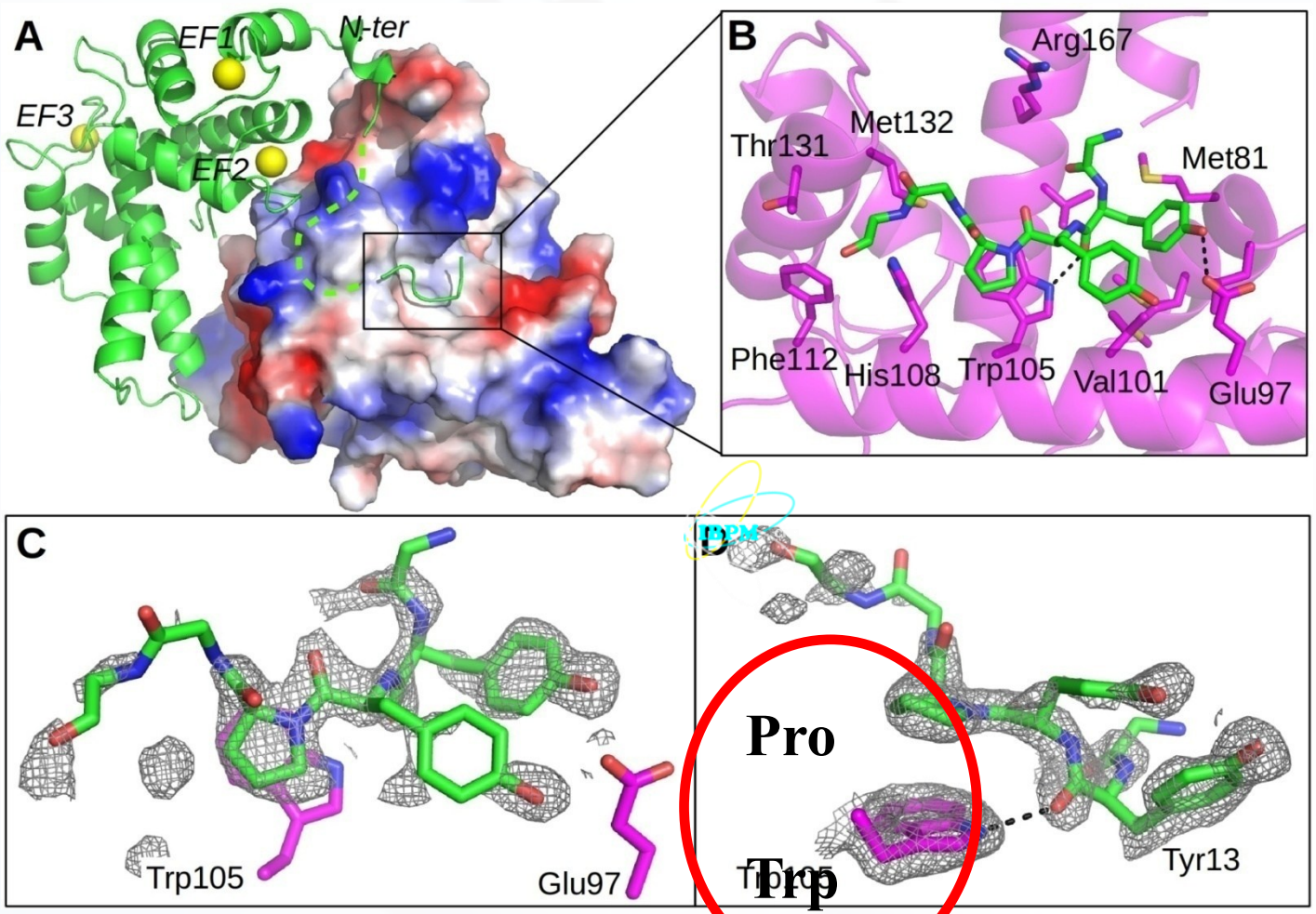


Tyr67 together with Met86 show the strongest variation of Solvent Accessible Surface Area upon Calcium binding

Ca (II) binding determines the opening of three pockets (pocket 1,pocket 2, pocket 3) mediating protein-protein interaction

Ilari et al.Sci Rep. 2015 Nov 18;5:16828.

In CaSor the SCBD domain is bound to a region of the N-terminal domain

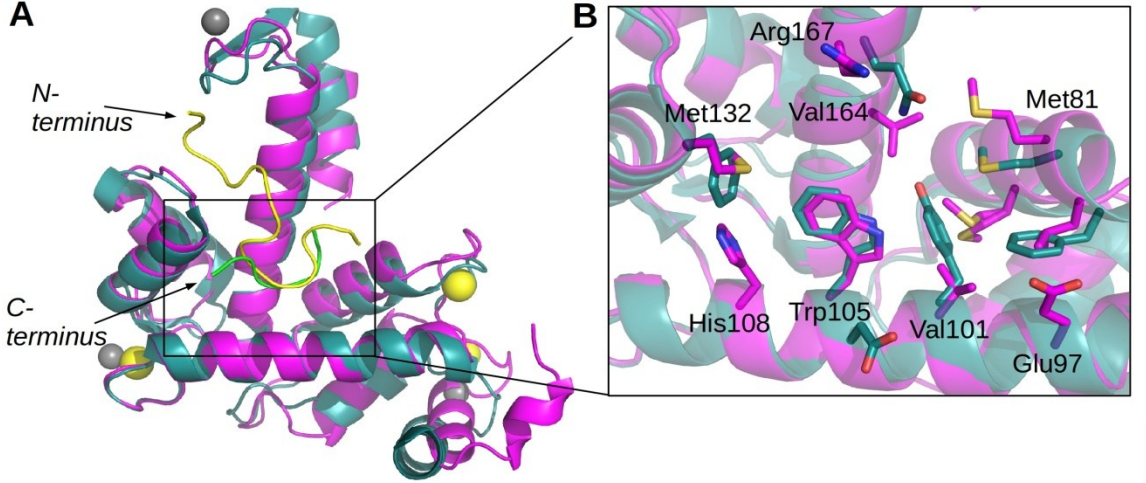


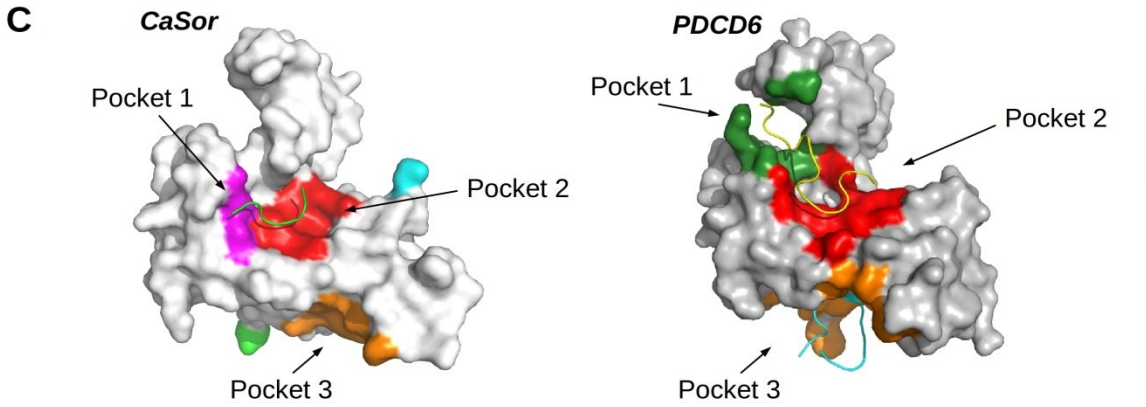
We fitted this electronic density map with the GYYPGG hexapeptide (residues 12-17)

Ilari et al., manuscript in revision Scientific reports

Peptide-binding Pockets 1-3 are present also in PDCD6

The superimposition of CaSor bound to N-terminal peptide and PDCD6 bound to Alix peptide shows that the protein have a similar conformation and that **both peptides bind in pocket 2** (In PDCD6, Val101 substitutes Tyr91)





D

```
                                        A          EF1        B
1    MA-YPGHPGAGGGYYPGGYGGAPG-GPAFPGQTQDPLYGYFAAV-AGQDGQIDADELQRCLTQ   60 Sorcin
1    MAAYSYRP-GPGAGPG------PAAGAALP--DQSFLWNVFQRVDKDRSGVISDTELQQALSN   54 PDCD6

                C          EF2          D          EF3
61   SGIAGGYKPFNLETCRLMVSMLDRDMSGTMGFNEFKELWAVLNGWRQHFISFDTDRSGTV   120 Sorcin
55   ----GTWTPFNPVTVRSIISMFDRENKAGVNFSEFTGVWKYITDWQNVFRTYDRDNSGMI   110 PDCD6

          E          F          EF4          G
121  DPQELQKALTTMGFRLSPQAVNSIAKRYSTNGK--ITFDDYIACCVKLRALTDSFRRRDT   178 Sorcin
111  DKNELKQALSGFGYRLSDQFHDILIRKFDRQGRGQIAFDDFIQGCIVLQRLTDIFRRYDT   170 PDCD6

          EF5        H
179  AQQGVVNFPYDDFIQCVMSV  198 Sorcin
171  DQDGWIQVSYEQYLSMVFSIV 191 PDCD6
```

l'infezione da SARS-CoV-2: il ruolo della biologia strutturale nella conoscenza del virus

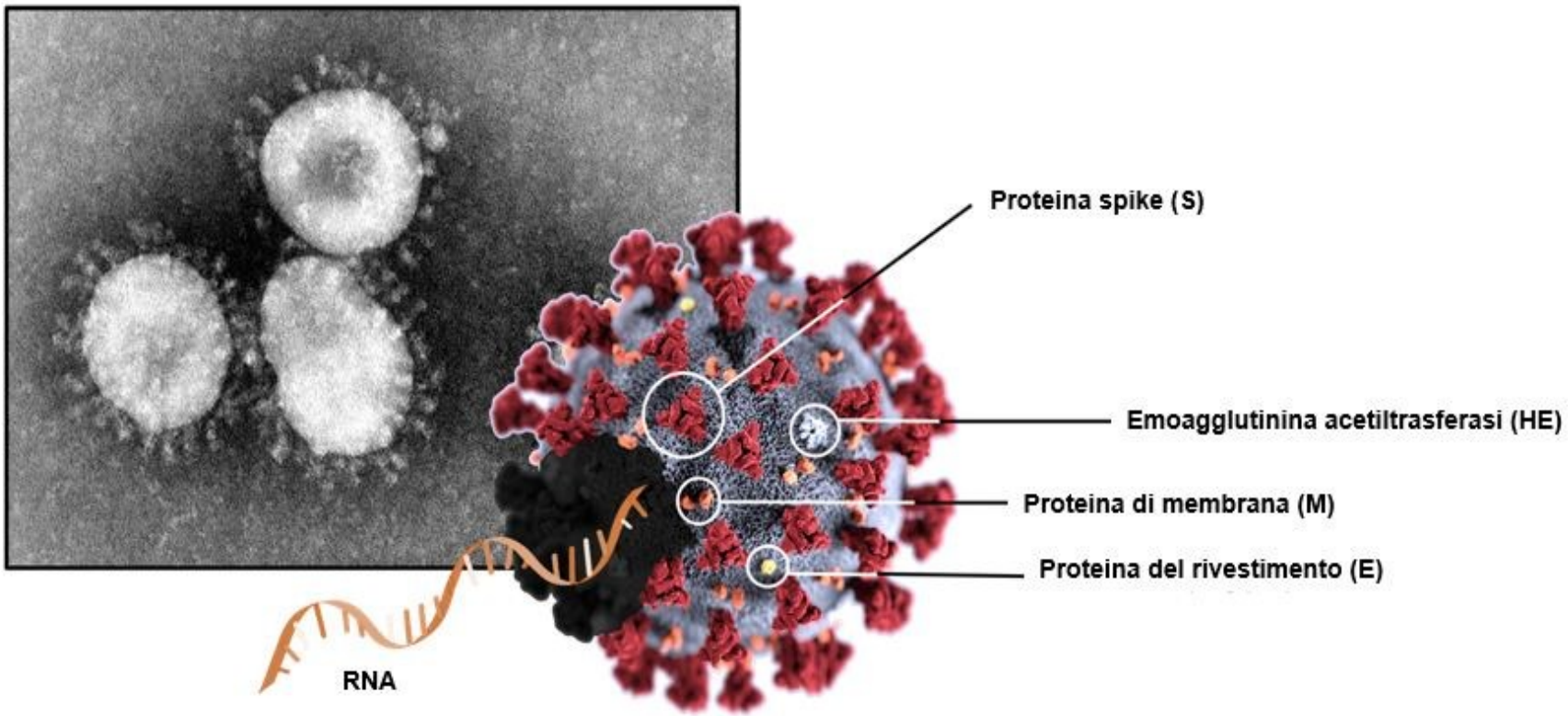


Immagine al microscopio elettronico a trasmissione del coronavirus.Le particelle virali appaiono bianche circondate da un fondo nero. Sono visibili le proteine Spike che "decorano" la superficie e che conferiscono al virione la caratteristica forma a corona Il capside virale (in grigio), che contiene l'RNA a singolo filamento associato alle proteine del nucleocapside, è costituito da un doppio strato lipidico in cui sono inserite le proteine di superficie (S in rosso; E in giallo; M in arancio; HE in bianco).

https://sibbm.zanichelli.it/italiano/2020/08/23/infezione-biologia-strutturale/

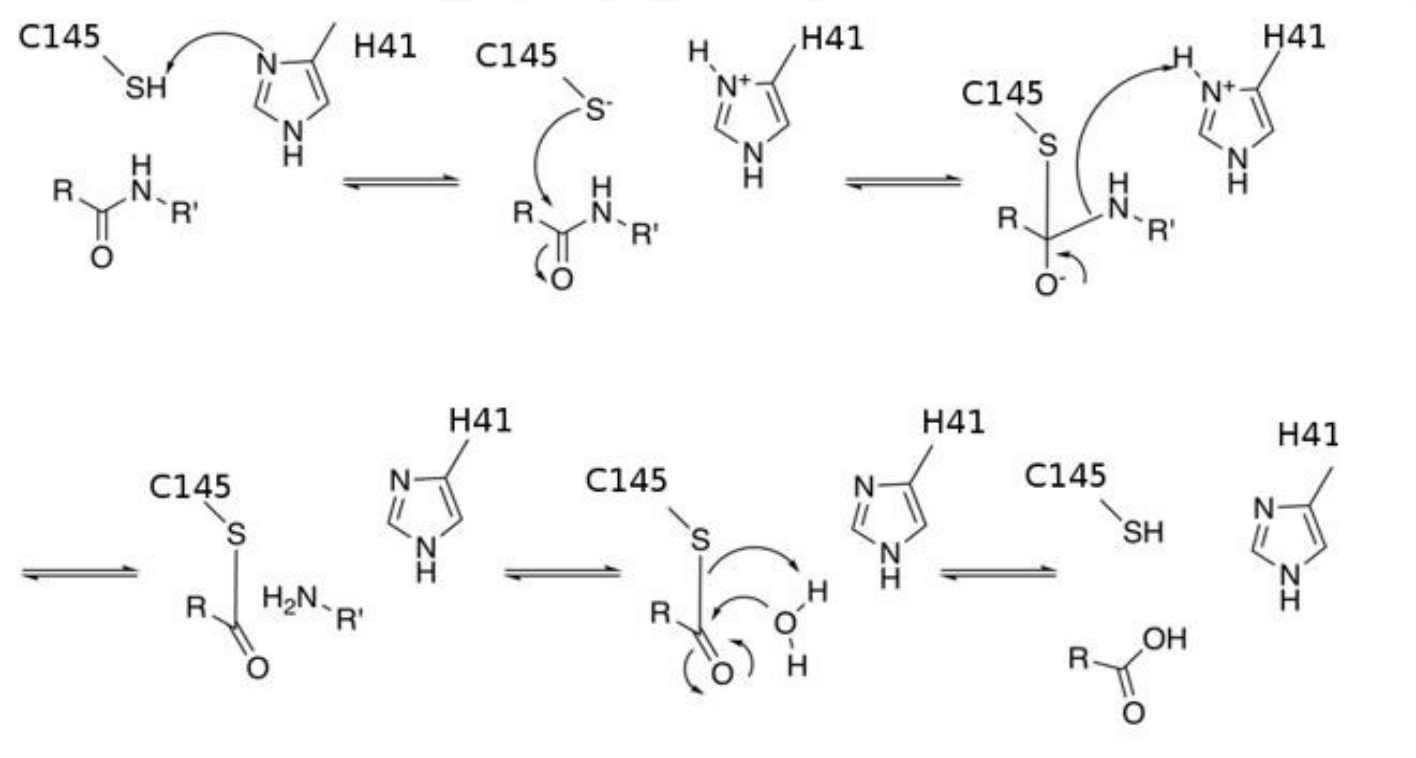
La proteasi principale del CoV-2: l'origine dell'intero arsenale molecolare del virus

Durante l'infezione della cellula ospite, il genoma virale agisce come RNA messaggero e dirige la sintesi di due grandi poliproteine (pp1a e pp1ab) che contengono al loro interno proteine più piccole necessarie alla produzione di nuove particelle virali all'interno delle cellule infettate. Tale insieme di proteine comprende: un complesso di replicazione/trascrizione, diverse proteine strutturali necessarie a costruire virioni e due proteasi . Queste due proteasi giocano un ruolo essenziale poiché tagliano le due grandi poliproteine nelle proteine funzionali più piccole.

La proteasi principale di CoV-2, che effettua il maggior numero di tagli, pesa 33,8 kDa e si chiama M^{pro}, altrimenti conosciuta come proteasi 3C-simile (simile alla chimotripsina). La M^{pro} è fondamentale per la replicazione virale ed è assente nelle cellule umane. Per questa ragione essa rappresenta un buon bersaglio per lo sviluppo di nuovi farmaci antivirali: il blocco delle sue funzioni sarebbe infatti letale per il virus, ma sicuro per gli esseri umani

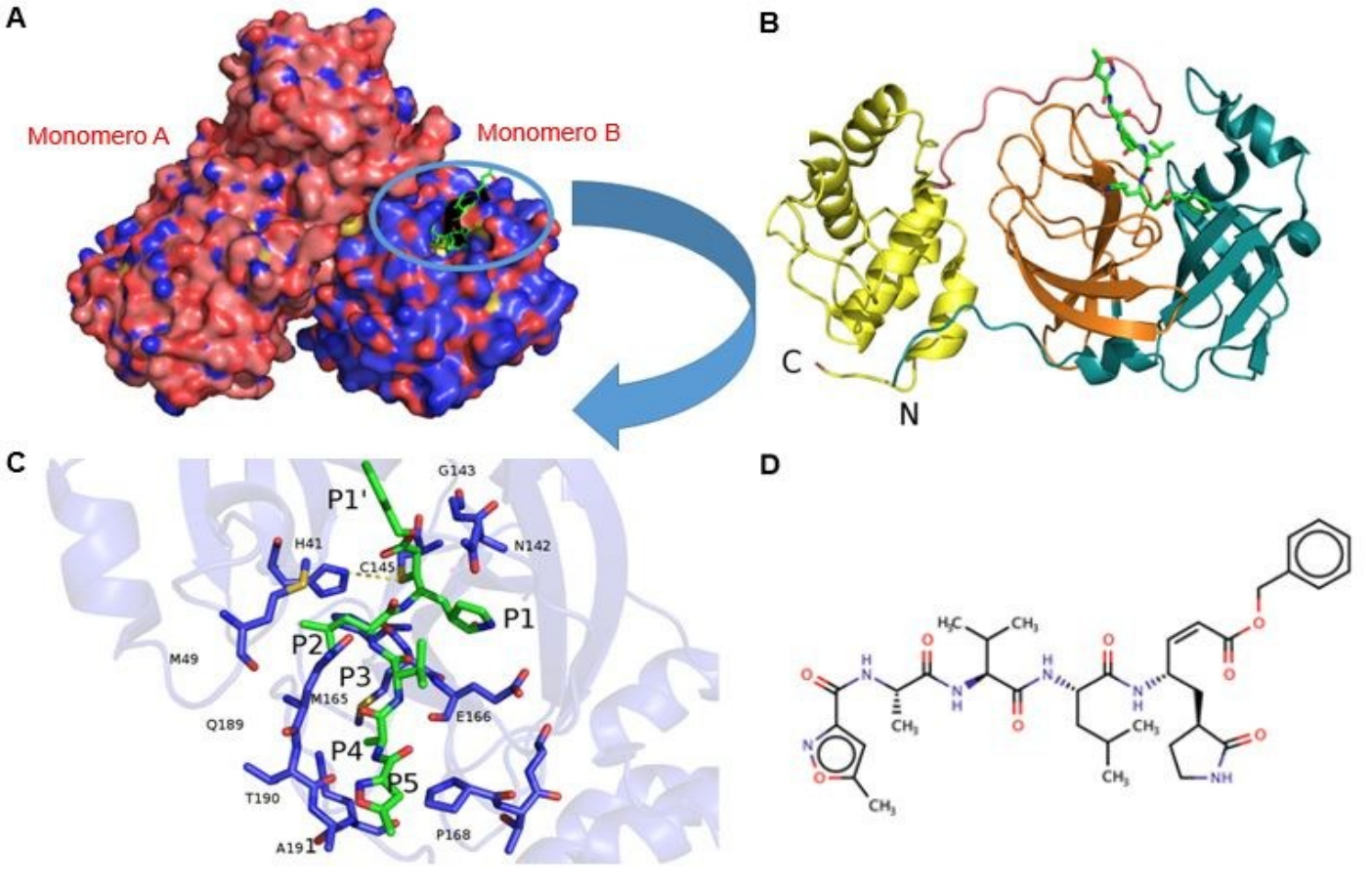
Meccanismo catalitico

La catalisi inizia con la deprotonazione del tiolo della Cys145 da parte dell'His41, che è seguita dall'attacco nucleofilico della cisteina deprotonata al carbonio del legame peptidico della proteina substrato. Viene quindi rilasciata una proteina virale con terminale amminico libero, il residuo di istidina, His41, della proteasi recupera la sua forma deprotonata e si forma un intermedio tioestere che collega il nuovo carbossi-terminale del substrato al tiolo Cys145. Il tioestere è quindi idrolizzato per generare un terminale carbossilico sulla poliproteina rimanente, rigenerando l'enzima libero.



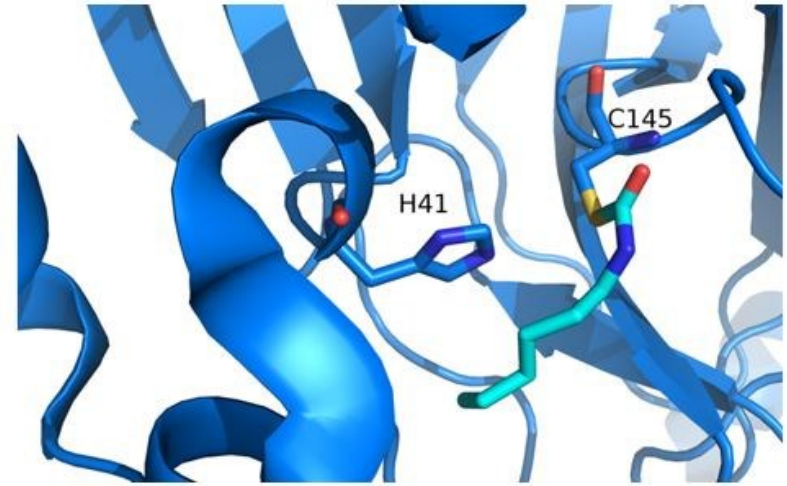
La struttura di M^{pro} in complesso con N3 ha consentito di mappare tutti i siti druggable sulla proteina

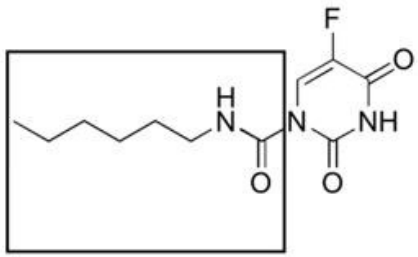
Rappresentazione di superficie dell'omodimero di Mpro. Gli atomi di carbonio dei due monomeri A e B sono colorati rispettivamente in rosa salmone e blu chiaro. Gli atomi di azoto sono colorati in blu, quelli di ossigeno in rosso e quelli di zolfo in giallo. L'inibitore N3 è rappresentato con la struttura a bastoncini ed è colorato in verde. B Rappresentazione "a nastro" del monomero Mpro. I domini I, II e III sono colorati, rispettivamente, in verde chiaro, arancio e giallo. La molecola N3 è colorata in verde. C Ingrandimento del sito di legame per N3. I residui catalici (H41 e C145) e quelli che interagiscono con l'N3 sono rappresentati come bastoncini blu. La posizione P1 corrisponde al residuo immediatamente prima del sito di taglio, seguito da P2, P3, P4, P5, fino alla parte N-terminale del sito di taglio. I siti corrispondenti in Mpro che interagiscono con i residui P1-P5 e P1′ sono denominati S1, S2, S3, S4, S5 and S1′. D Formula strutturale dell'inibitore N3 (N-[(5-Metilisossazol-3-yl)carbonil]alanil-L-valil-N~1~-((1R,2Z)-4-(Benzilossi)-4-osso-1-{[(3R)-2-ossopirrolidin-3-il]metil}But-2-enil)-L-Leucidammide).



Il Virtual Screening su M^{pro} ha portato al'individuazione di un «lead compound» ora in fase clinica

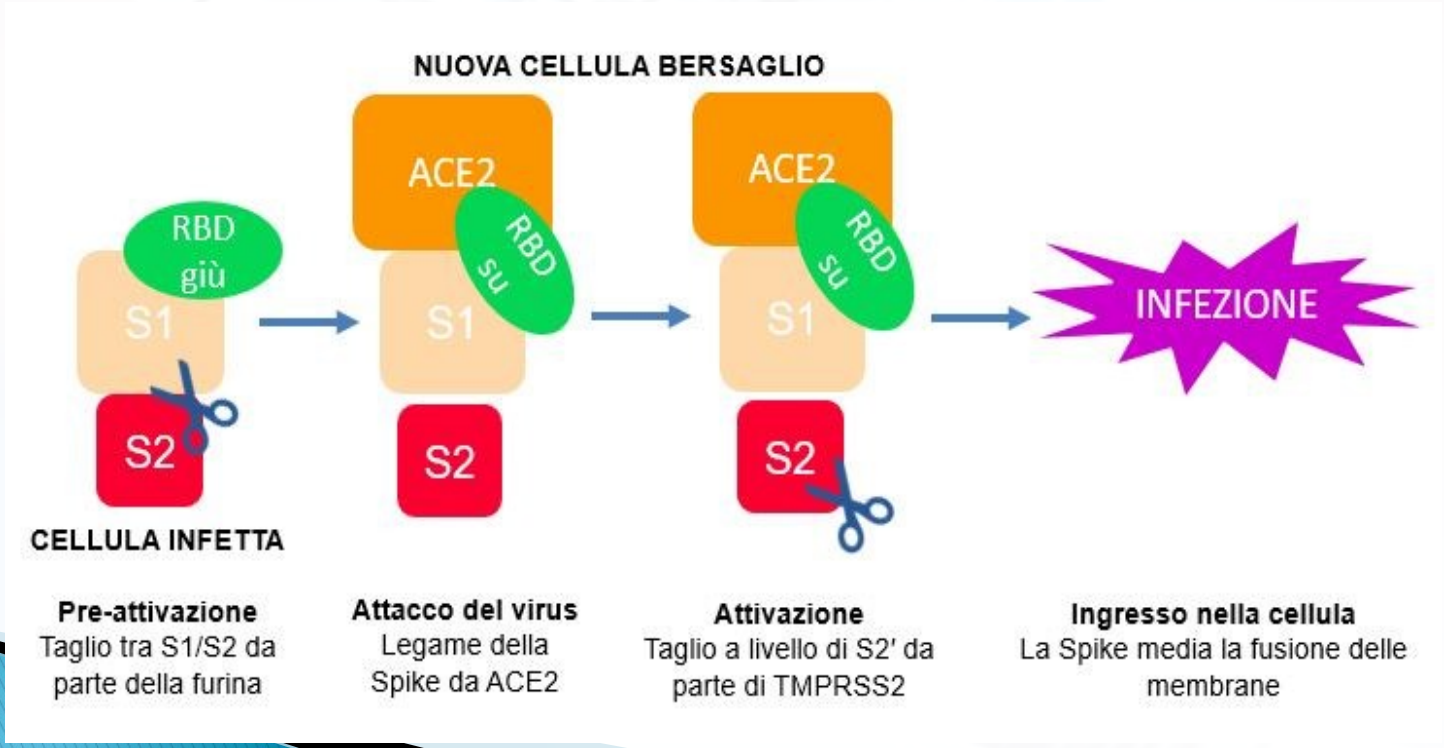
Sono stati analizzati oltre 10000 composti come ligandi possibili della Mpro, inclusi farmaci già approvati, alcuni candidati per le sperimentazioni cliniche e anche dei prodotti naturali. Tra questi, hanno identificato un derivato del 5-fluorouracile, il Carmofur (1-hexylcarbamoyl-5-fluorouracil), capace di legare ed inibire l'M^{pro}). Il Carmofur è stato già approvato come farmaco antitumorale ed è usato per trattare il cancro del colon retto fin dal 1980. Jin e collaboratori hanno risolto la struttura cristallografica dell'Mpro in complesso con il Carmofour, scoprendo le basi molecolari dell'efficacia del composto. In particolare, l'acido grasso di Carmfour lega la Cys145 attraverso un legame covalente, inibendo l'enzima





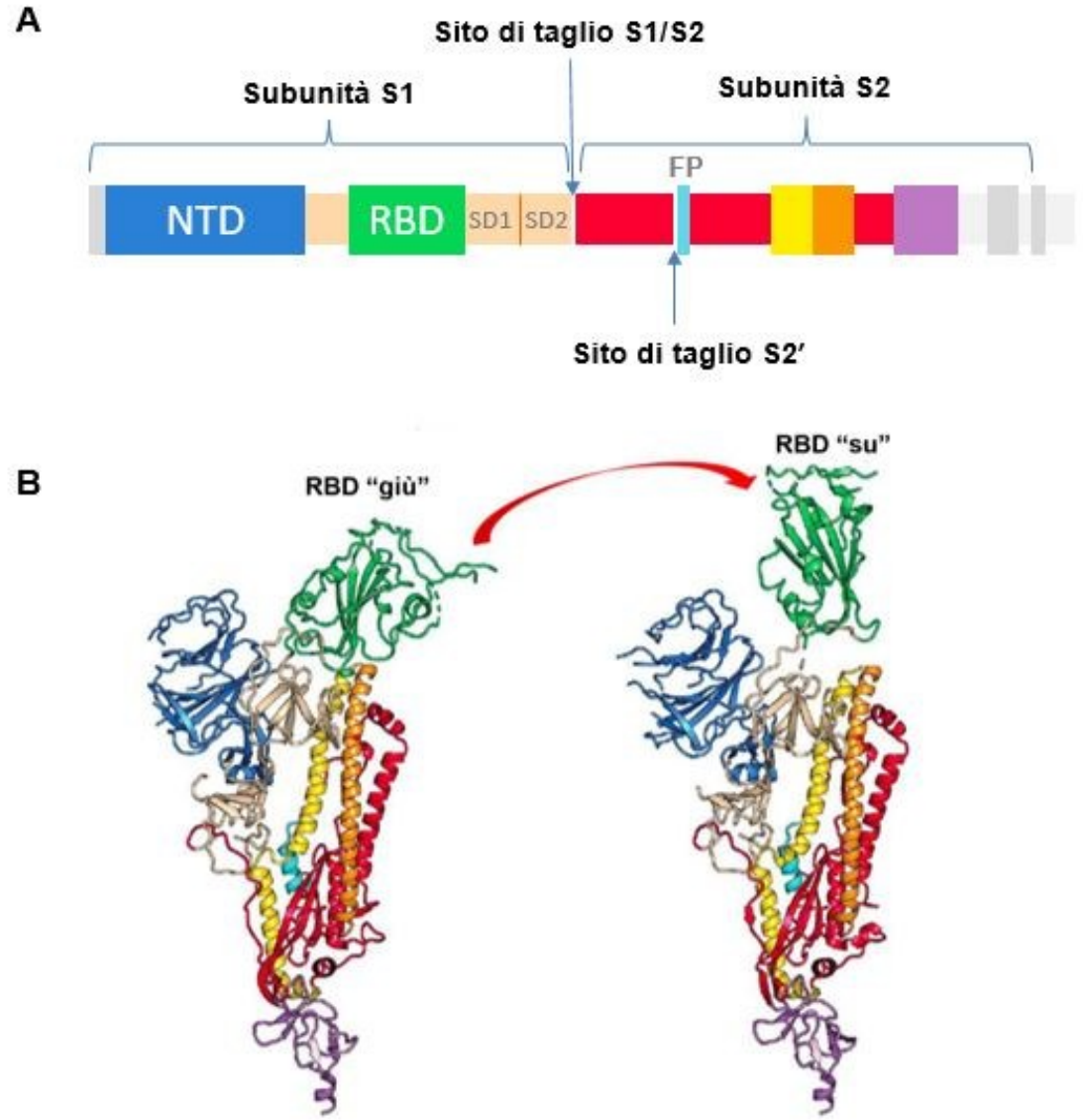
LA PROTEINA SPIKE

I virus evolvono continuamente le proteine della loro superficie per potenziare l'interazione con i recettori sulle cellule ed entrare in esse con maggior efficienza secondo il modello chiave-serratura. Questo è anche il caso della proteina Spike (S) di CoV-2 (la chiave) e del recettore umano Angiotensin Converting Enzyme 2 (hACE2, la serratura).
La proteina S è una delle più interessanti e studiate tra quelle che contribuiscono al legame con il recettore dell'ospite e alla patogenesi virale. La proteina S "decora" la superficie del virus ed è responsabile per l'aspetto a corona della superficie virale, da cui il nome coronavirus. Questa è usata dal virus come una chiave per entrare nelle cellule ospite15. Agisce legando il recettore sulle cellule bersaglio, induce l'endocitosi dei virioni e catalizza la fusione tra le membrane cellulari e virali, assicurando l'ingresso dell'RNA genomico virale nel citoplasma delle cellule

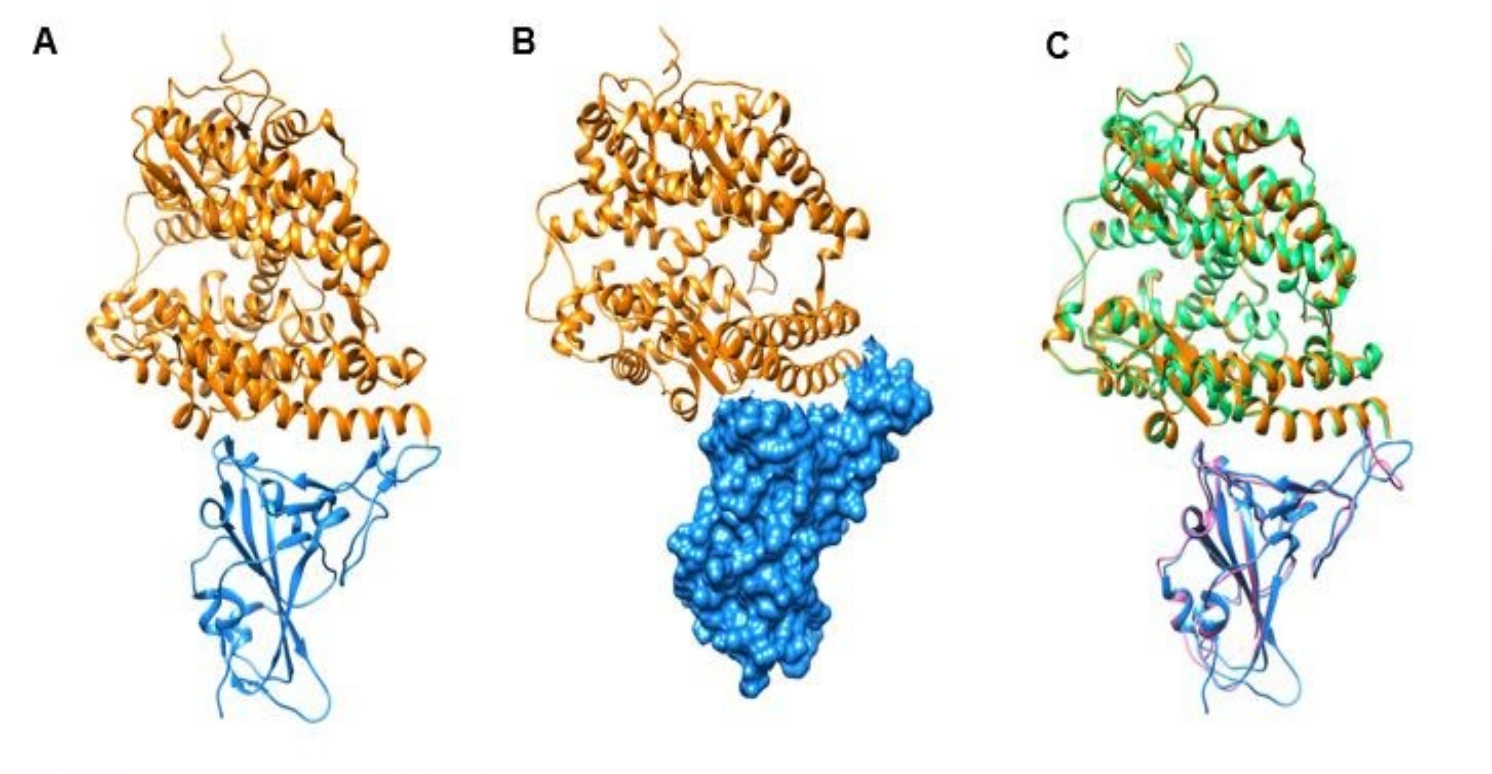


LA proteina Spike ha diversi domini

- un dominio N-terminale (**NTD**, *N-Terminal Domain*);
- due subdomini detti **SD1** e **SD2** (*Subdo main 1* e *2*);
- il dominio di legame al recettore (**RBD**, *Receptor Binding Domain*) responsabile per il legame alla cellula ospite attraverso l'interazione con ACE2 (vedi oltre);



L’RBD si lega all’ACE2 umano



L’RBD è colorato in blu

Tutti i vaccini sono stati sviluppati sulla base delle conoscenze di Biologia Strutturale

Vaccino	Antigene utilizzato	
Pfizer-BioNTech	Spike intera	mRNA
Moderna	Spike intera	mRNA
AstraZeneca	Spike intera	
Johnson & Johnson	Spike intera	
Novavax	Spike intera ricombinante	
Corbevax	RBD ricombinante	
Abdala	RBD ricombinante	
ZF2001 (Zifivax)	RBD dimerico	

Vaccini più economici. Bloccati dall'EMBARGO USA (1 dollaro contro i 30 di MODERNA)

I due vaccini principali sviluppati a Cuba

SOBERANA 02

- RBD ricombinante di SARS-CoV-2.
- RBD coniugato chimicamente alla tossina tetanica (tetanus toxoid, TT).
- Adiuvante: idrossido di alluminio.

SOBERANA Plus

- RBD ricombinante dimerico.
- Dimerizzazione mediante ponte disolfuro intermolecolare.
- Adiuvante: idrossido di alluminio.

- Due dosi di SOBERANA 02: efficacia del 69,7% contro COVID-19 sintomatico.
- Due dosi di SOBERANA 02 più una dose di SOBERANA Plus: efficacia del 92% contro COVID-19 sintomatico.

