# Bioinformatica ed applicazioni di Bioinformatica Strutturale

## Parte 2: Bioinformatica Strutturale

Sotto-disciplina della bioinformatica che comprende:

- ✓rappresentazione,
- ✓deposito,
- ✓ recupero,
- ✓analisi,
- ✓visualizzazione grafica

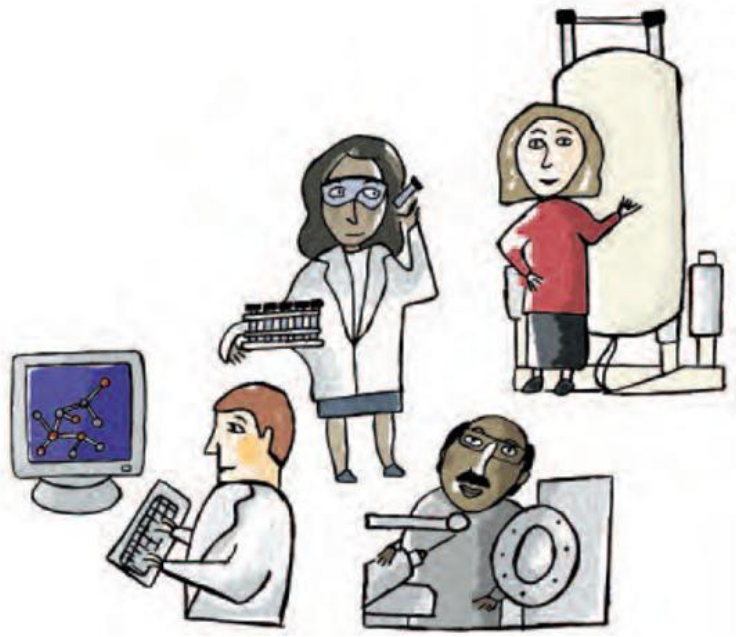
della struttura in scala atomica e spaziale subcellulare di molecole biologiche.

✓ Ha due obiettivi:

1. Fornire metodi generali per manipolare le informazioni sulle macromolecole biologiche;
2. Applicare questi metodi per risolvere problemi in biologia e creare nuove conoscenze

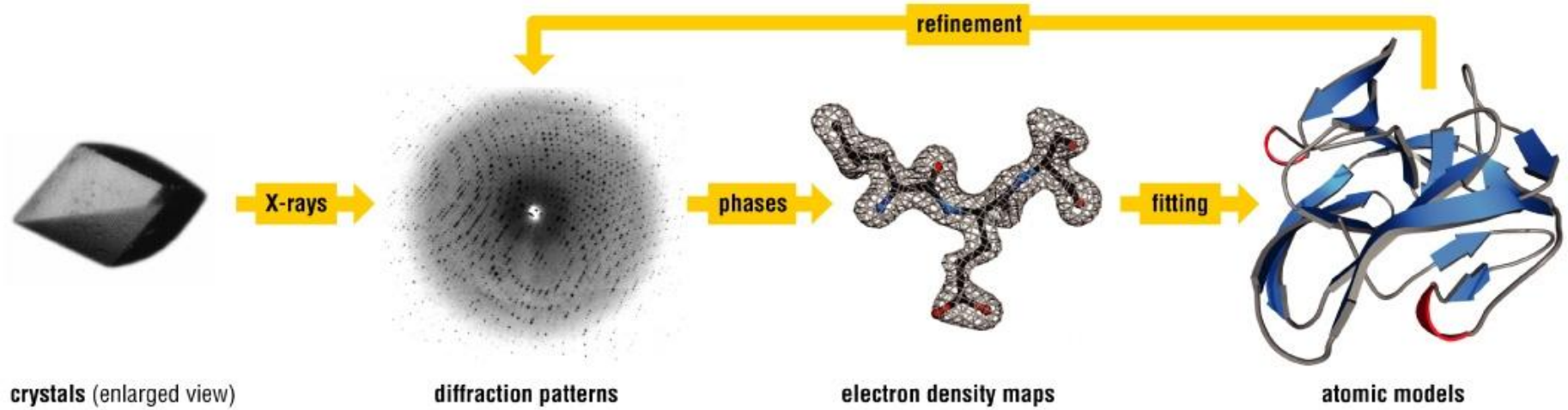
# Origine dati strutturali

- ✓Cristallografia Raggi X
- ✓Spettroscopia NMR
- ✓Criomicroscopia elettronica
- ✓Modelling teorico



# X-Ray

From **Protein Structure and Function** by Gregory A Petsko and Dagmar Ringe

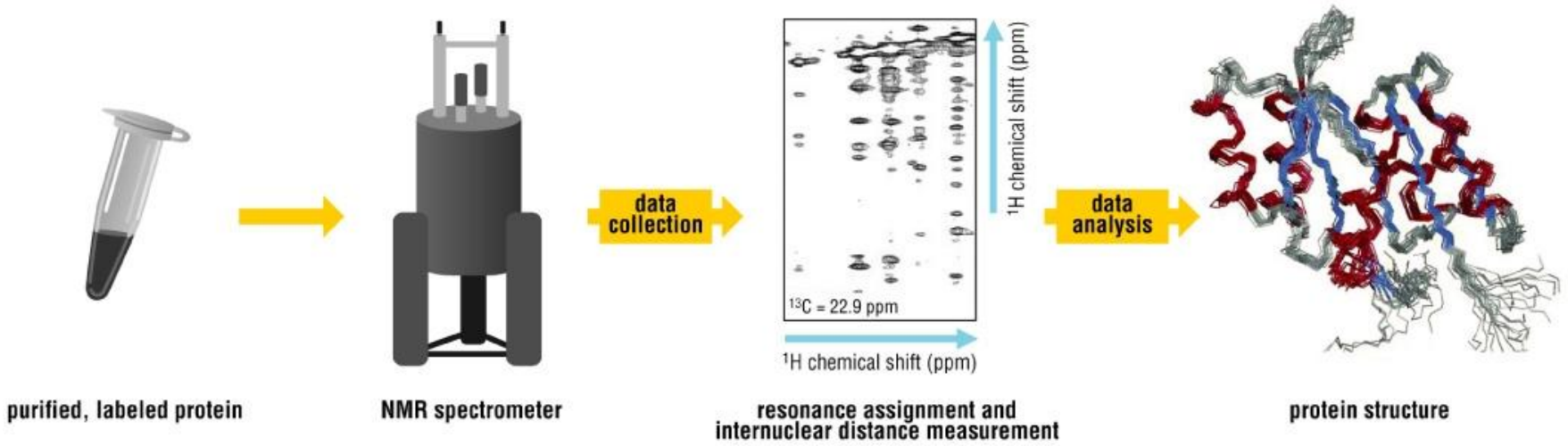





Risoluzione buona ≤ 1.5 Å

# NMR

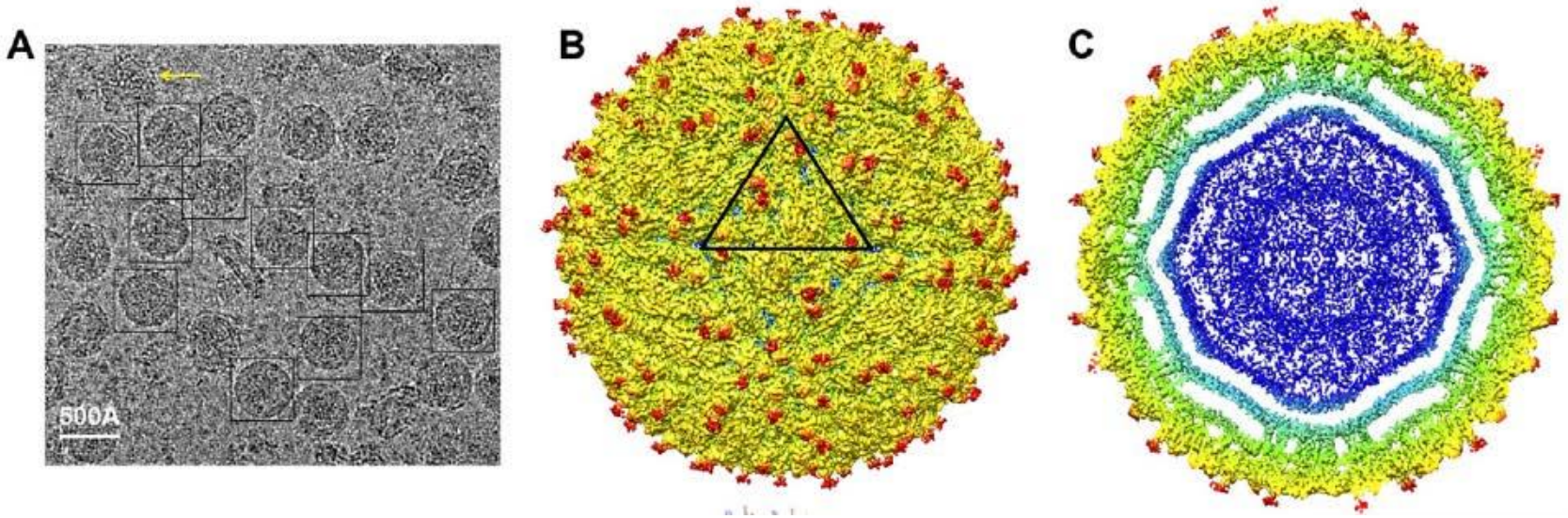
From **Protein Structure and Function** by Gregory A Petsko and Dagmar Ringe






$^{1}H$ NMR, ca 100 aa; $^{15}N$+ $^{13}C$ NMR, ca 350 aa

# Criomicroscopia elettronica

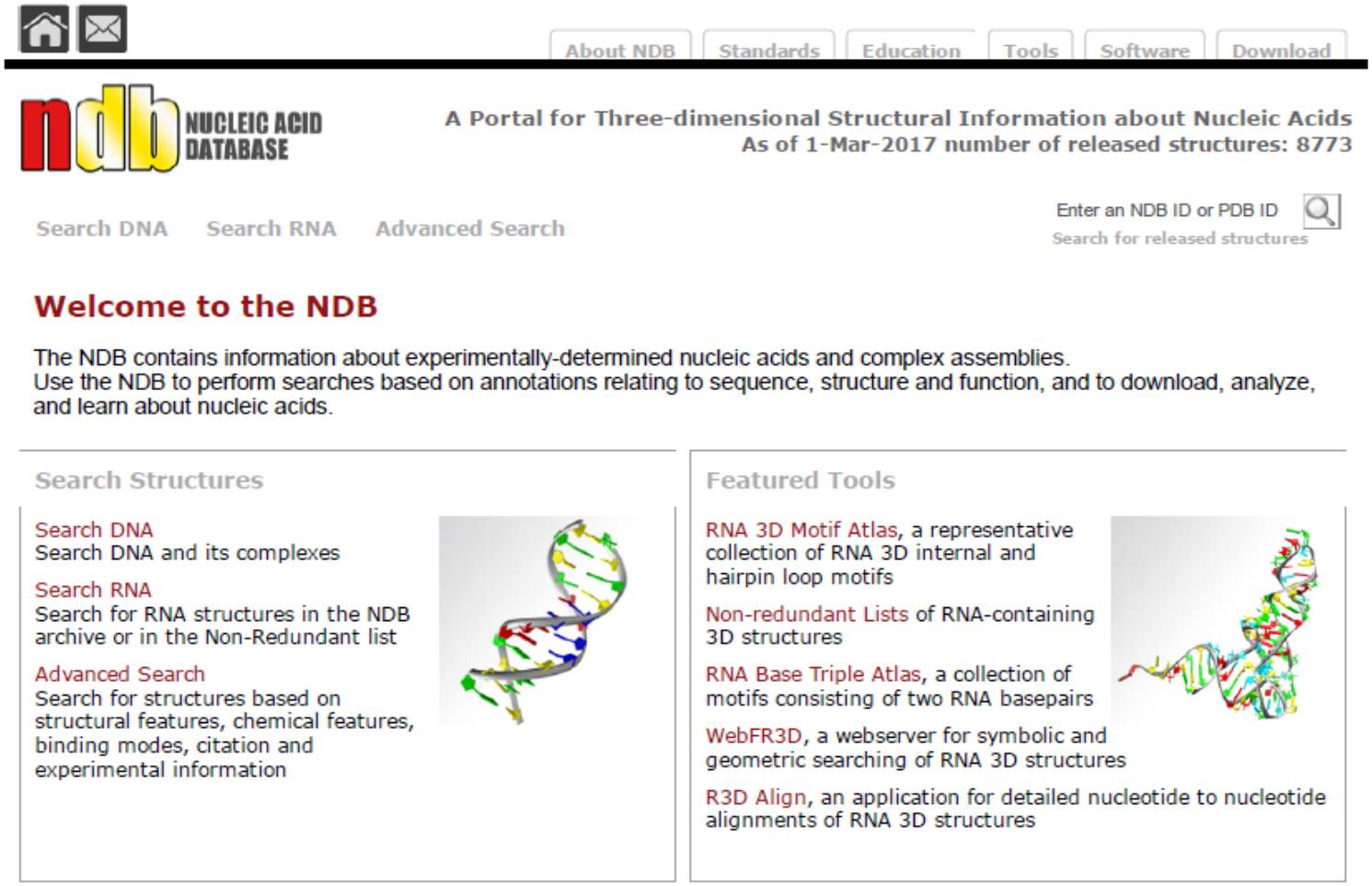



Risoluzione buona 3 ÷ 4 Å

Es: batteriorodopsina, complesso antenna dei cloroplasti, virus; ribosomi (insieme a X-Ray)

# Banca dati Acidi Nucleici

✓1992, specializzata nel campo della struttura di acidi nucleici; contiene strutture di acidi nucleici da soli o complessati con leganti.



# Banche dati per strutture di proteine

- ✓ **RCS PDB** = Research Collaboratory for Structural Bioinformatics Protein Data Bank http://www.rcsb.org/
- ✓ **PDBe** = Protein Data Bank in Europe http://www.ebi.ac.uk/pdbe/
- ✓ **BMRB** = Biological Magnetic Resonace Data Bank http://www.bmrb.wisc.edu/
- ✓ **PDBj** = Protein Data Bank Japan http://pdbj.org/

# Perché è importante conoscere la struttura 3D di una proteina?

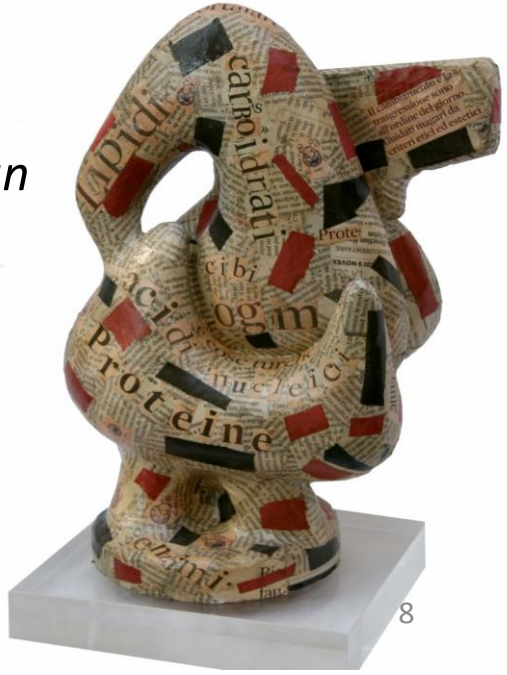
✓Una proteina svolge la sua funzione biologica solo se è ripiegata nella “giusta” struttura

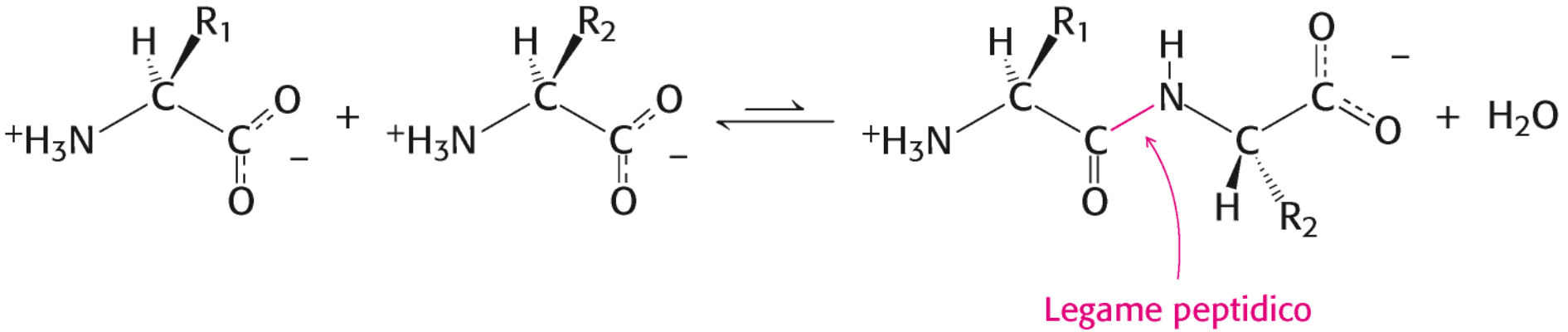
✓Una proteina non ripiegata correttamente o non funziona affatto o funziona male

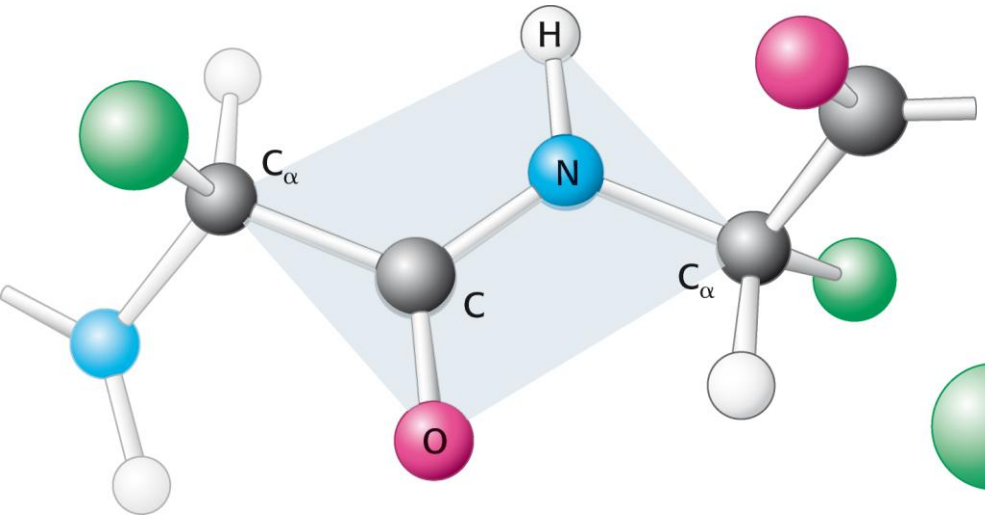
✓Conoscere la struttura è la chiave per avere informazioni sul meccanismo di funzionamento

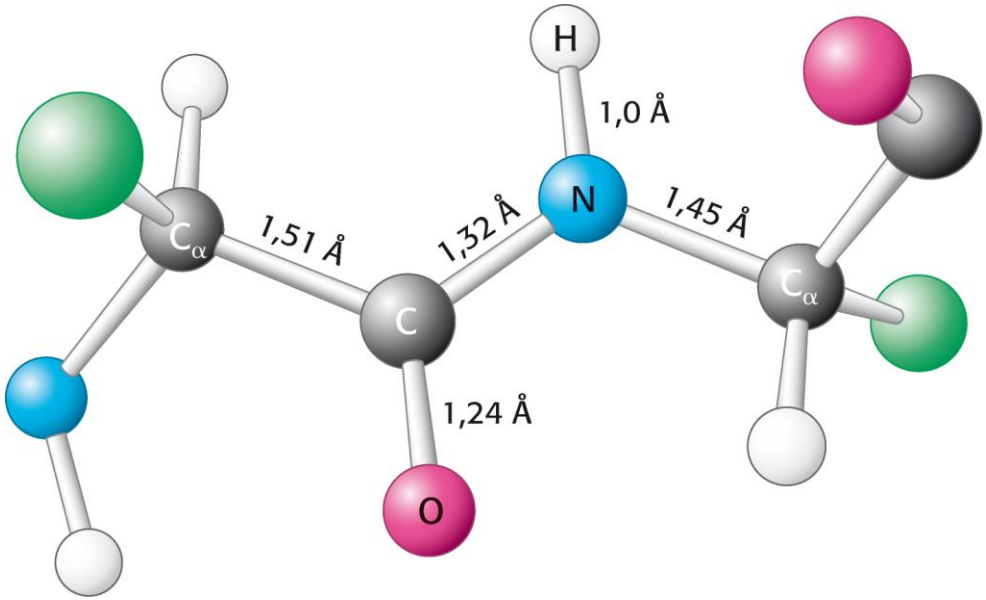
✓L’informazione può essere utilizzata anche per il *drug-design*

✓La struttura non è statica

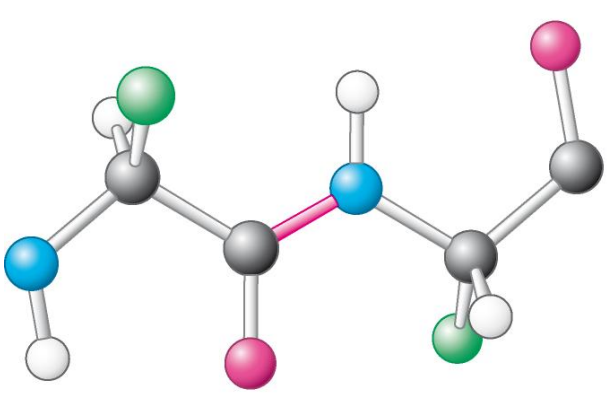



Legame peptidico

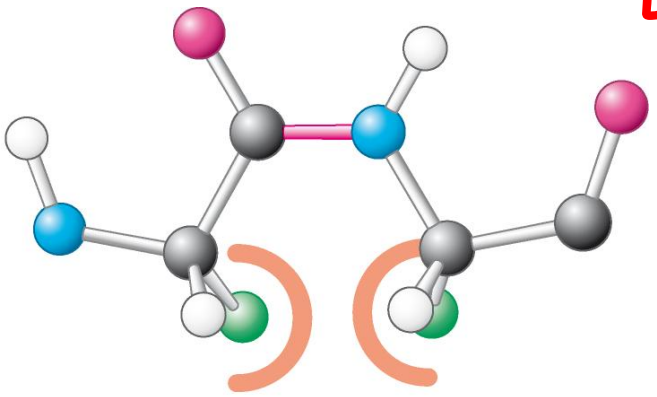

H
Cα
N
C
Cα
O


H
1,0 Å
Cα
1,51 Å
1,32 Å
N
1,45 Å
Cα
C
1,24 Å
O

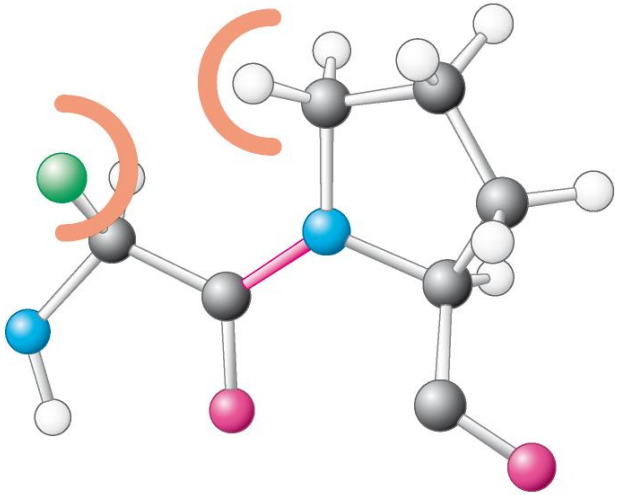
# Legame peptidico



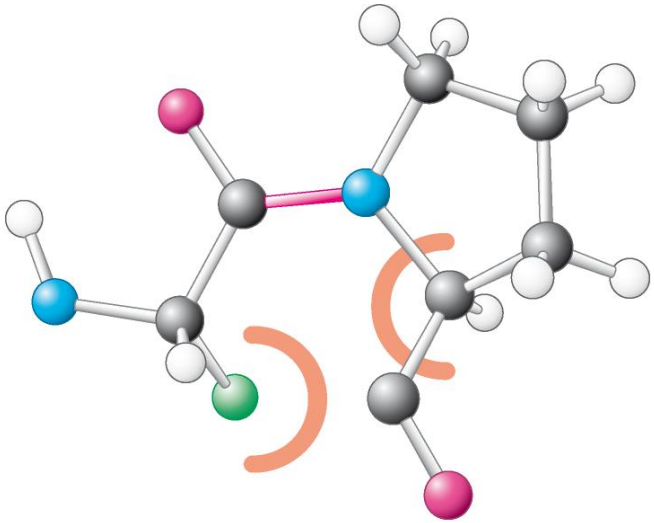
Trans



Cis

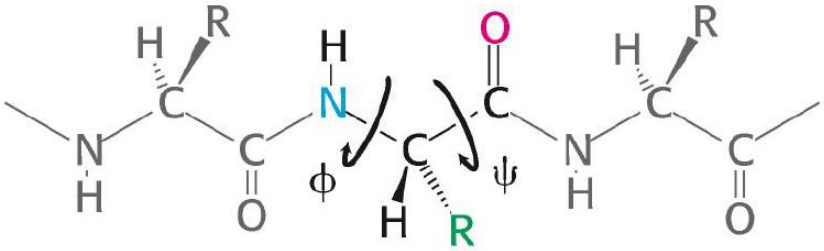


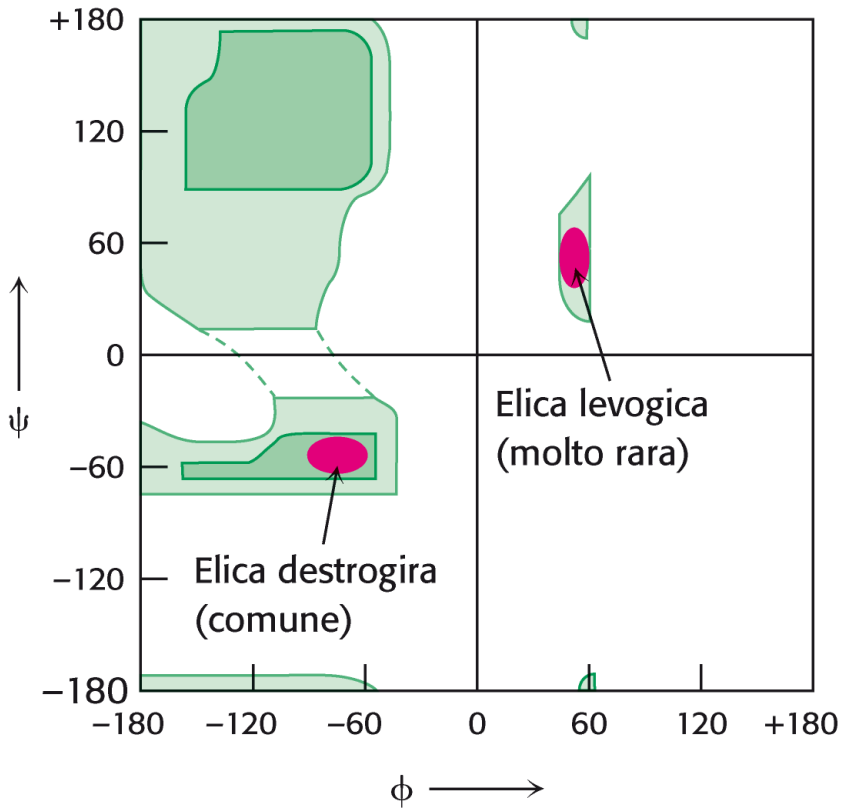
Trans

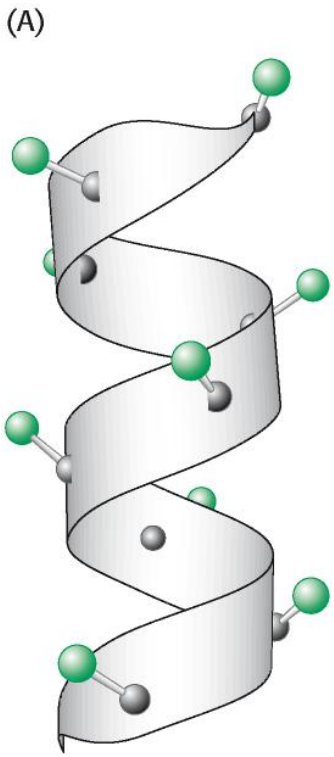


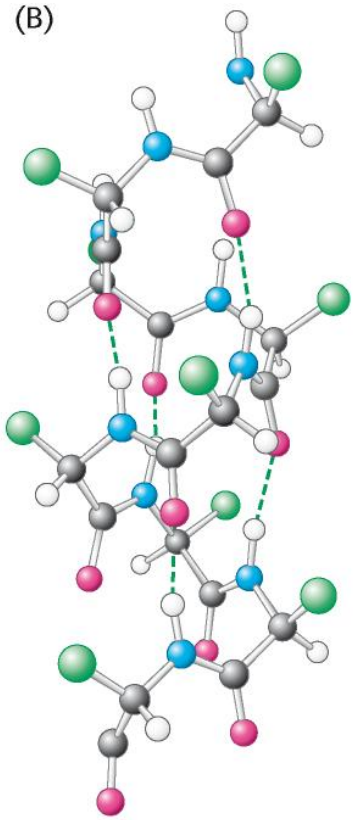
Cis

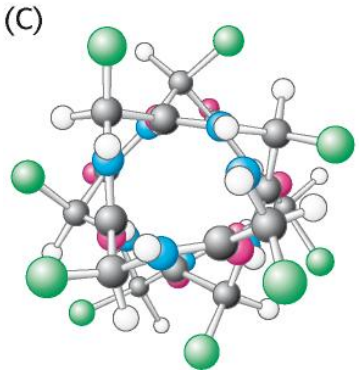
# Alfa elica

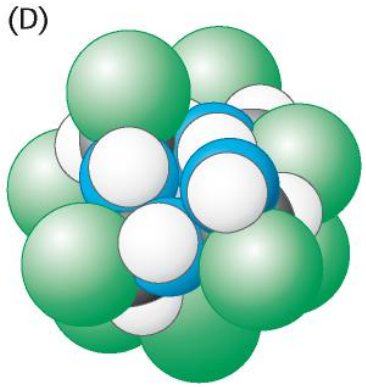

H
R
N
H
C
O
H
N
C
O
C
N
H
C
R
C
O
ϕ
ψ
H
R

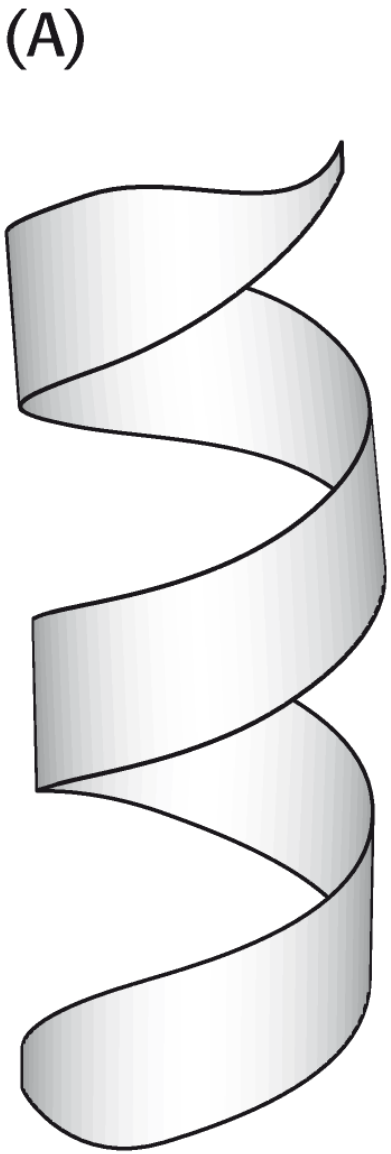

+180
120
60
0
−60
−120
−180
−180
−120
−60
0
60
120
+180
ψ
ϕ
Elica levogica
(molto rara)
Elica destrogira
(comune)

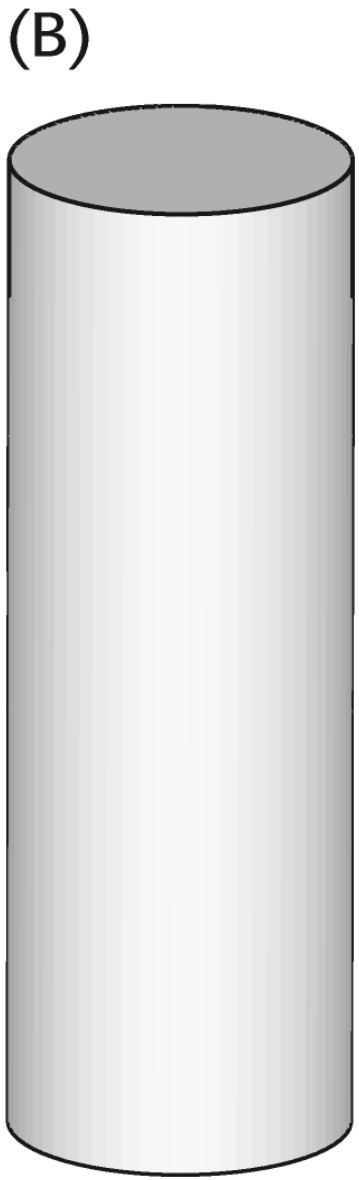

(A)


(B)


(C)

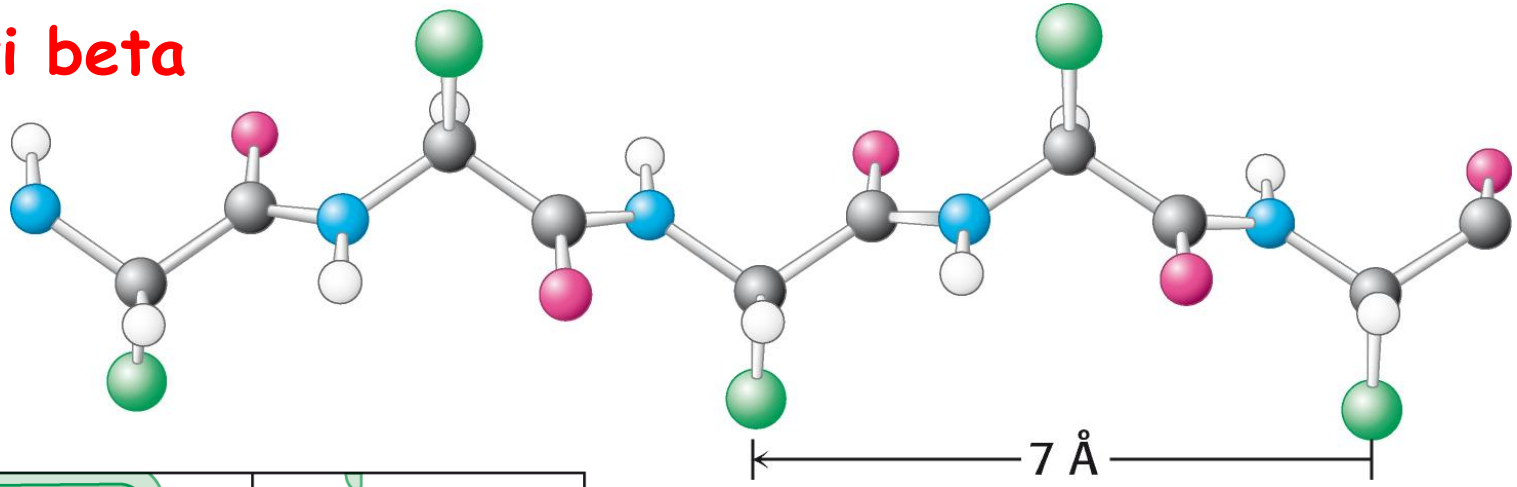

(D)

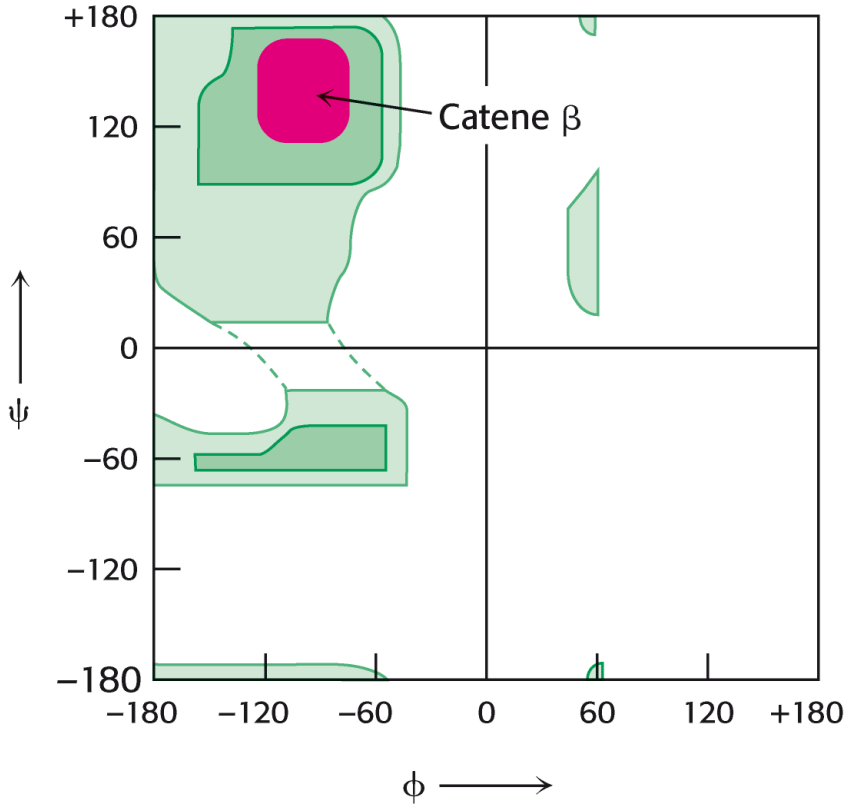


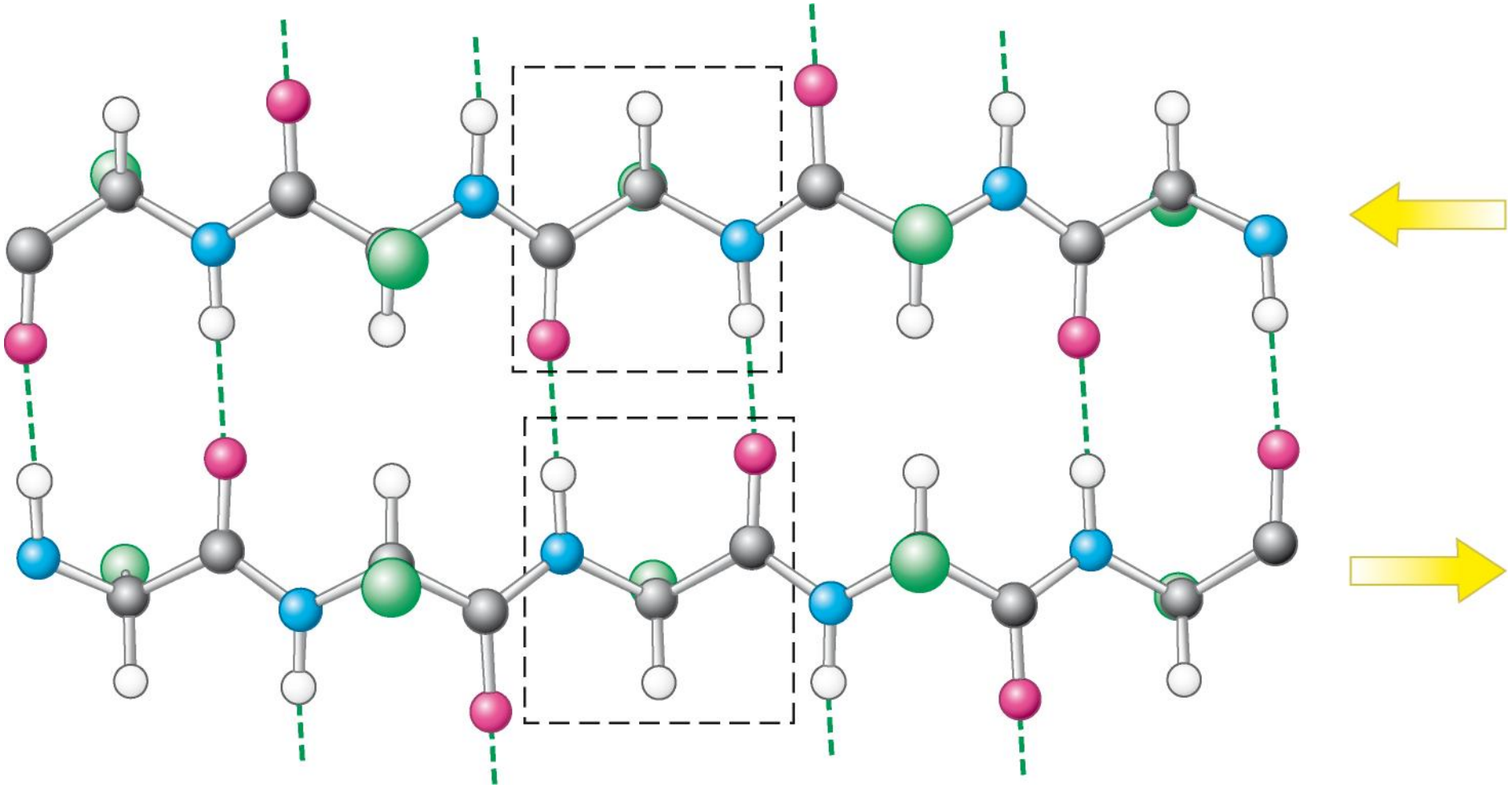

Alfa elica

# Foglietti beta

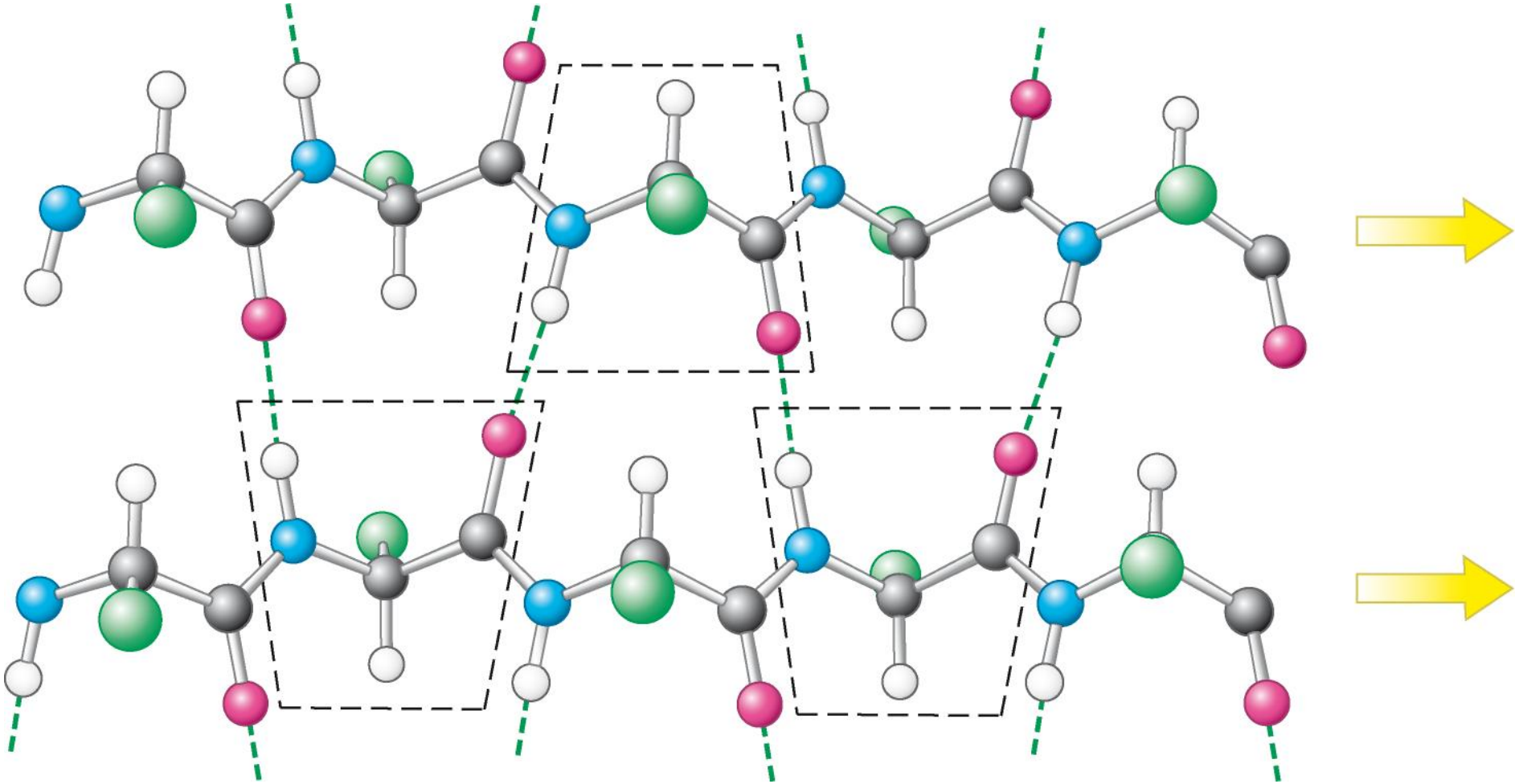


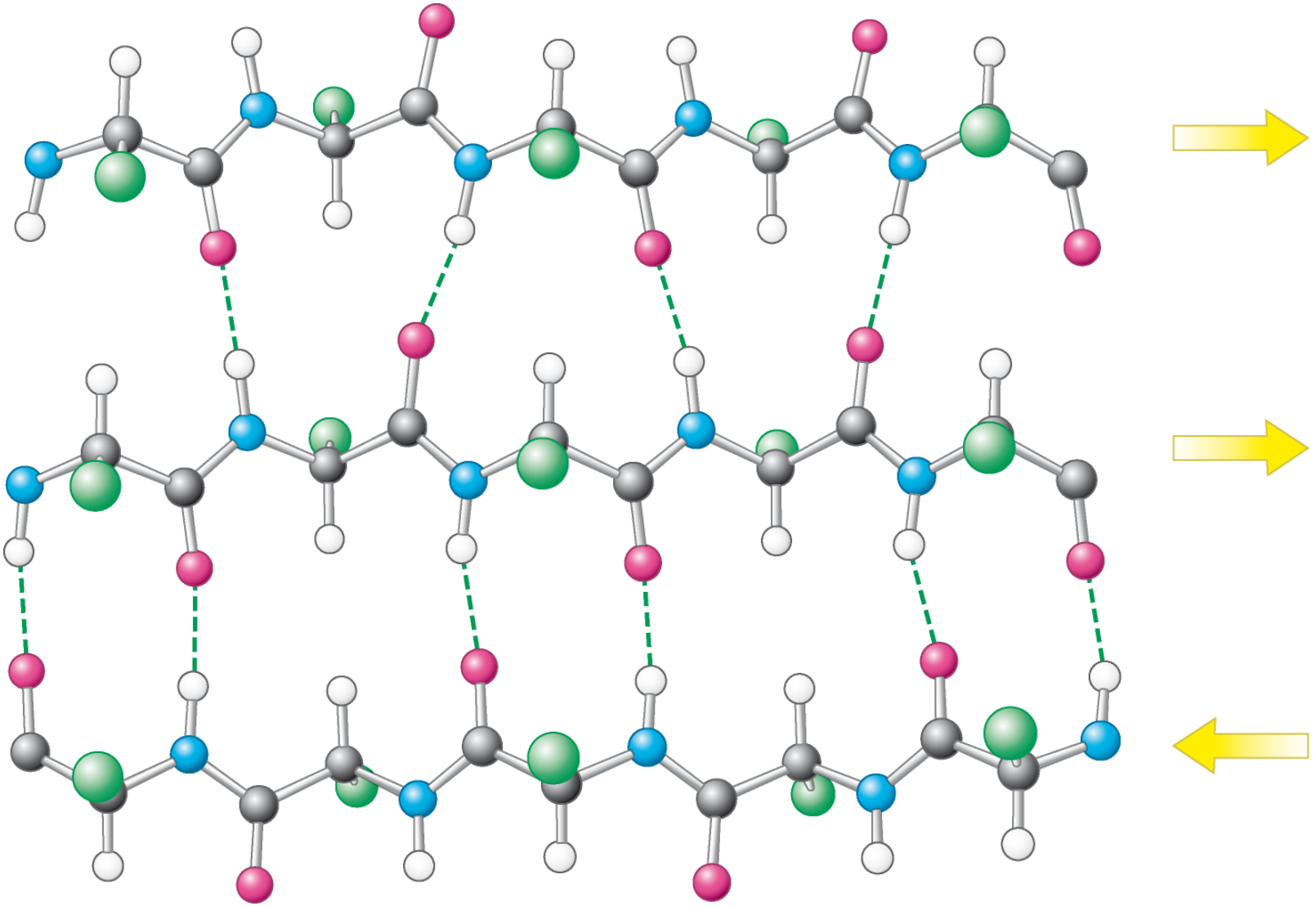
# Foglietti beta antiparalleli



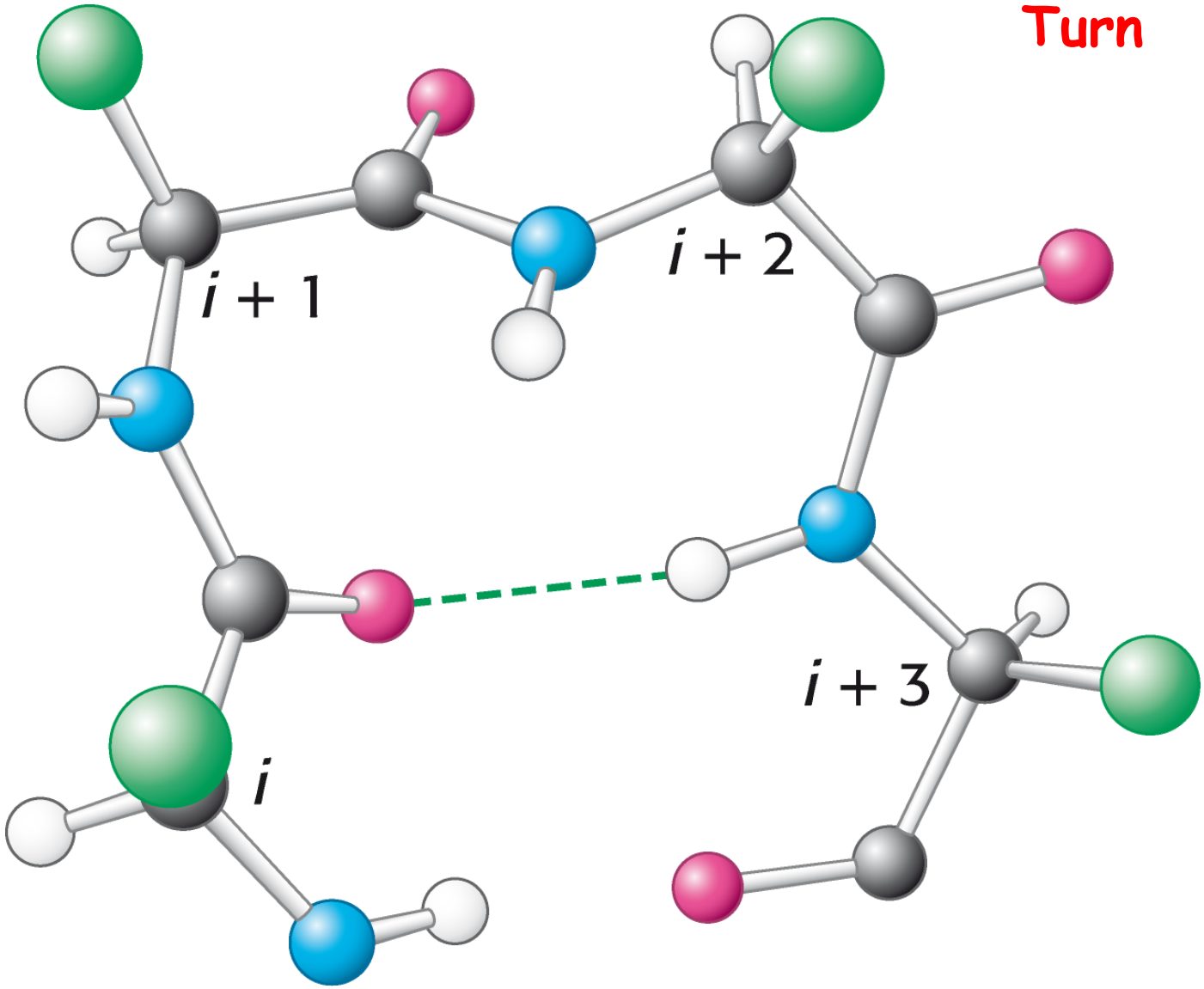
# Foglietti beta paralleli



# Foglietti beta misti



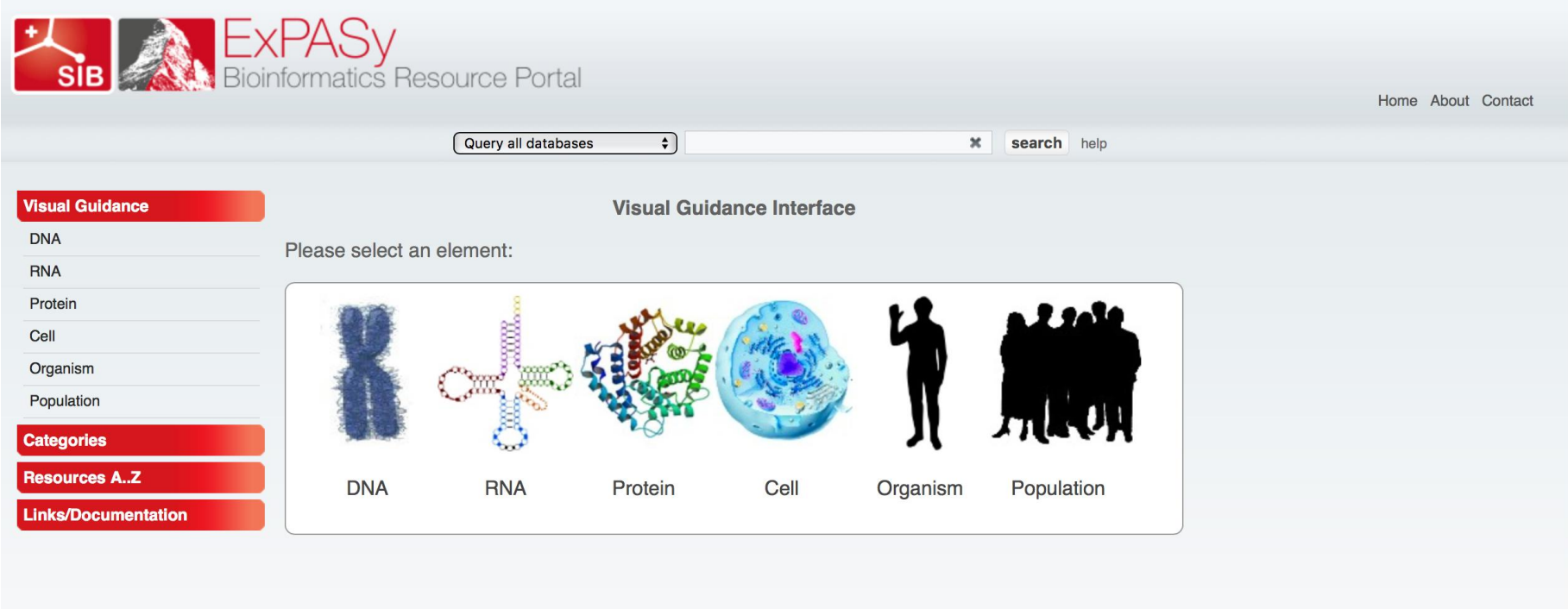
# Turn



**…… e se mancano i dati sperimentali sulla struttura di una proteina?**

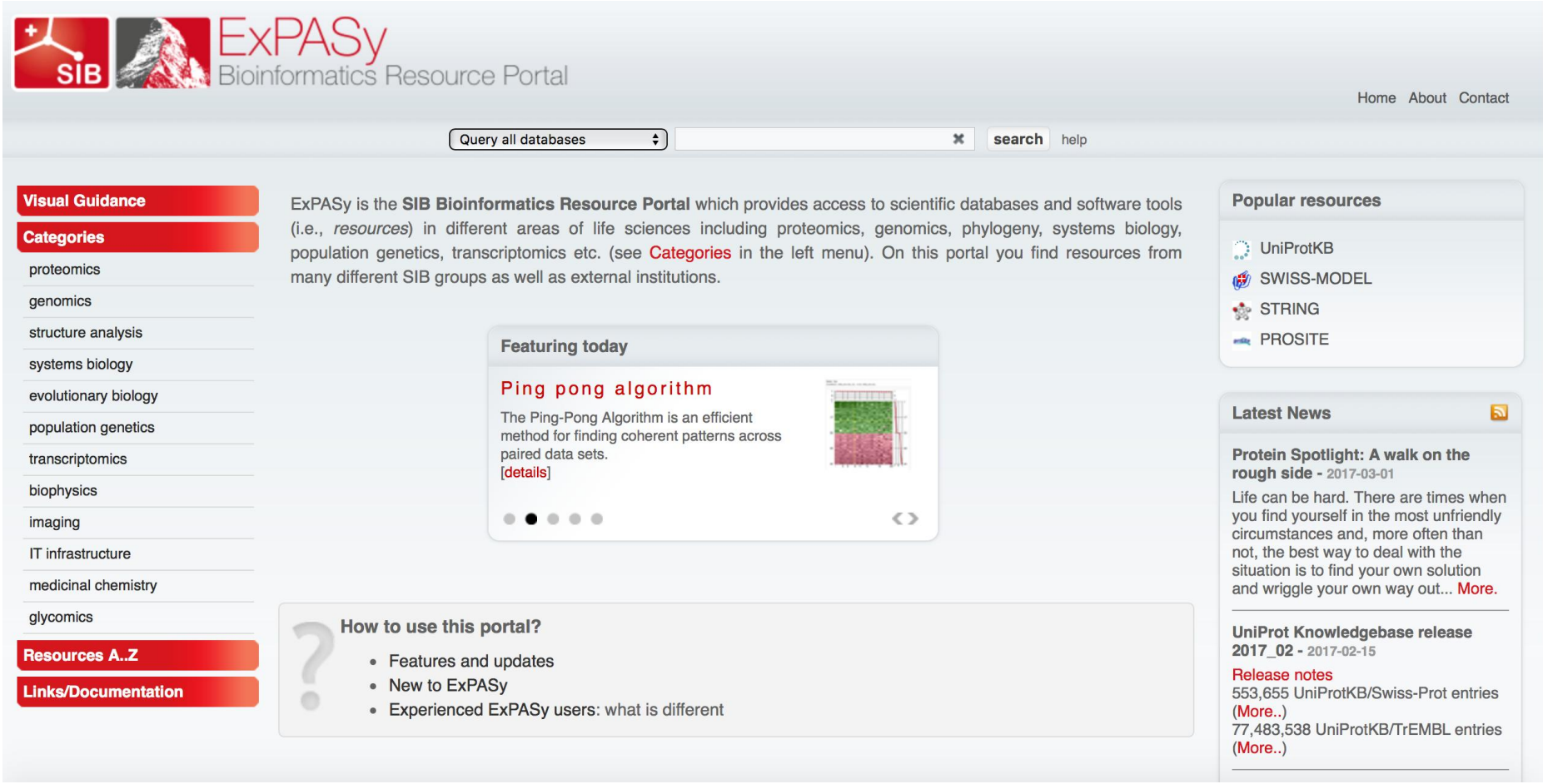
# ExPASy **SIB (Swiss Institute) Bioinformatics Resource Portal**

https://www.expasy.org/



# Predizione della struttura secondaria di una proteina

## ExPASy **SIB (Swiss Institute) Bioinformatics Resource Portal**



ExPASy **SIB (Swiss Institute) Bioinformatics Resource Portal** https://www.expasy.org/

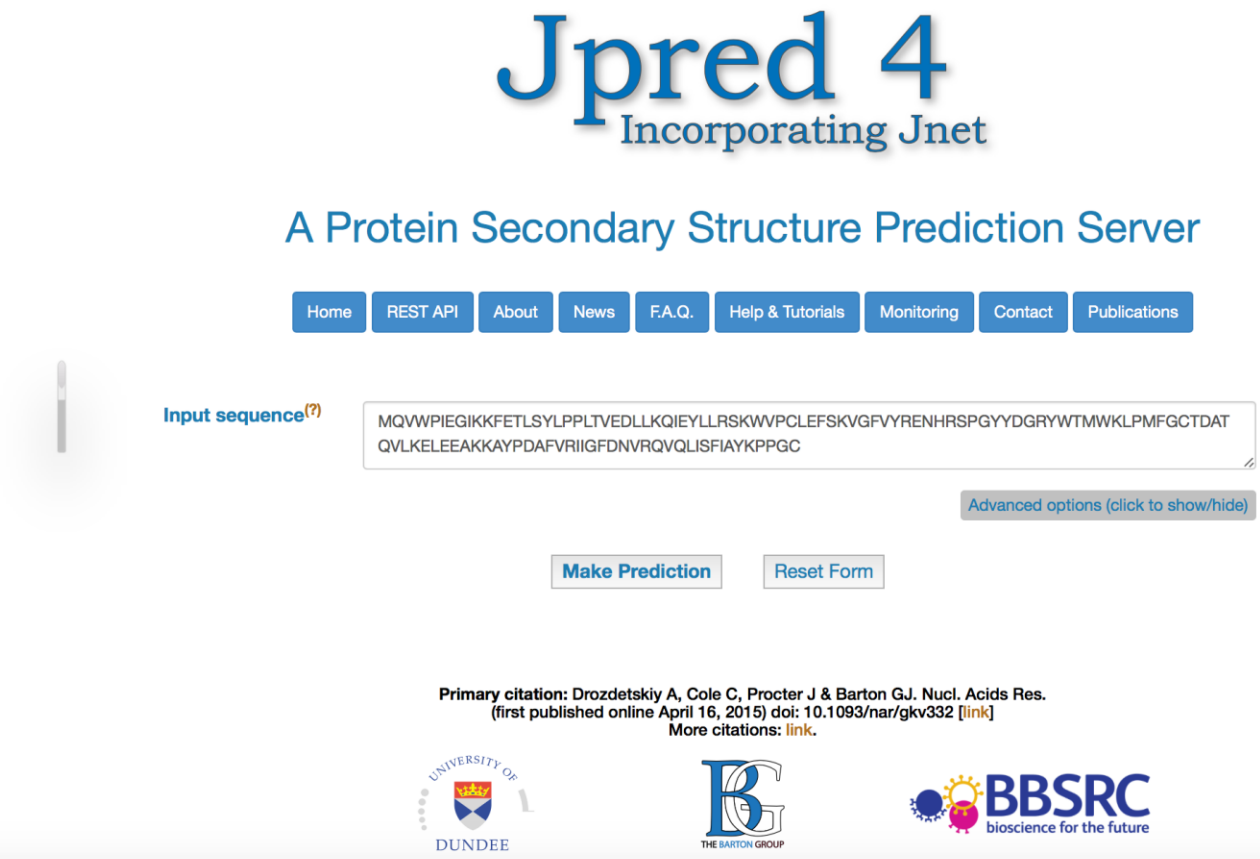
In **Categories**

**proteomics**

**protein stucture**

In **Tools** selezionare un metodo di predizione:

- ✓**GOR**
- ✓**HHPred**
- ✓**Jpred**

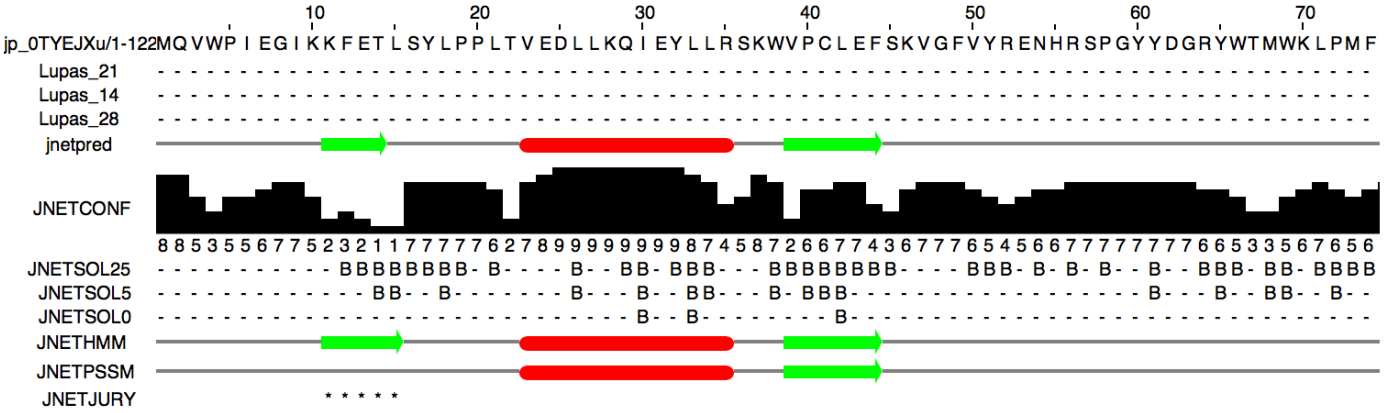


http://www.compbio.dundee.ac.uk/jpred/

# Results

After much trouble and strife, Bob the scheduling penguin has retrieved your results! Rejoice. For your pleasure the following viewing options are available. You may bookmark this page for future reference although data is not kept on the server for more than two days.

- View results summary in SVG - displayed below (details on acronyms used):




- View full results in HTML
- View simple results in HTML
- View results in PDF
- View results in Jalview (Link to a separate page with the Jalview Java Desktop application)
- View everything in a results directory (details on data each file contains are available through README file)
- Get all (but PS) files in TAR.GZ archive
- View results using in-browser Jalview Java applet (light version, limited functionality w.r.t. Jalview Desktop version linked above)

This Jpred prediction was made with following.

# CASP, Critical Assessment of protein Structure Prediction

## Protein Structure Prediction Center
## http://predictioncenter.org/

- ✓ Competizione diffusa in tutto il mondo che ha luogo ogni 2 anni a partire dal 1994 (CASP1 (1994) – CASP12 (2016)
- ✓ Applicazione di metodi predittivi messi alla prova su proteine di struttura nota ma non ancora pubblicata
- ✓ Incentivo per l'avanzamento e lo sviluppo di metodi di identificazione della struttura delle proteine a partire dalla sequenza

# Modelling per omologia

✓E' una tecnica utile quando si vuole predire la struttura di una proteina bersaglio di sequenza nota, correlata ad almeno un'altra proteina di sequenza e struttura nota.

✓Nelle famiglie di proteine, le strutture hanno regioni relativamente costanti ed altre più variabili. Il nucleo della struttura conserva la topologia del folding mentre la periferia può ripiegarsi completamente.

✓Un'unica struttura "genitore" permetterà un modelling ragionevole della regione conservata ma spesso fallirà nel produrre un modello soddisfacente della regione variabile.

✓Il modello è migliore se vi sono più strutture note di proteine correlate da usare come "progenitori."

Alcuni software a disposizione per effettuare il modelling per omologia:

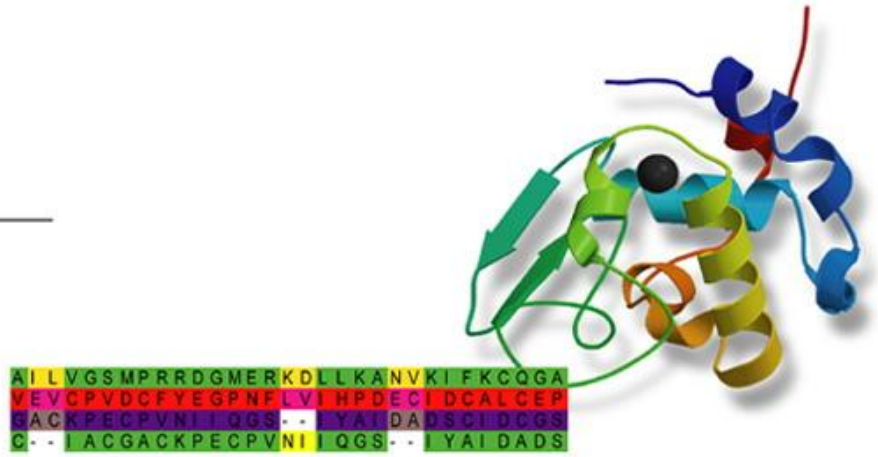
**SWISS-MODEL** (http://swissmodel.expasy.org)

**PHYRE2** (http://www.sbg. bio.ic.ac.uk/phyre2)

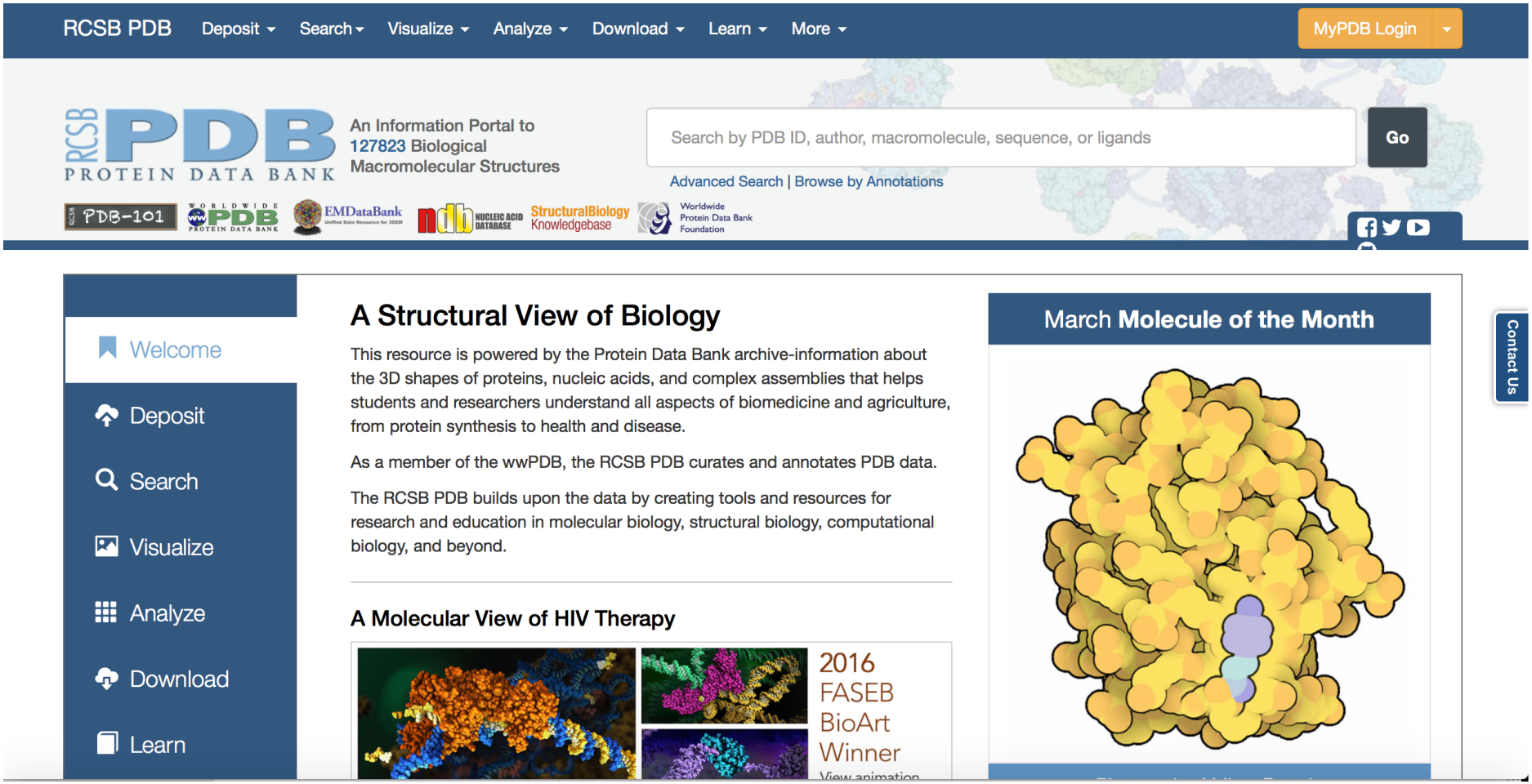
**MODELLER** (https://salilab.org/modeller/)

Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



# Struttura proteine nella Protein Data Bank



# PDB http://www.rcsb.org/
# Processamento dati nella PDB

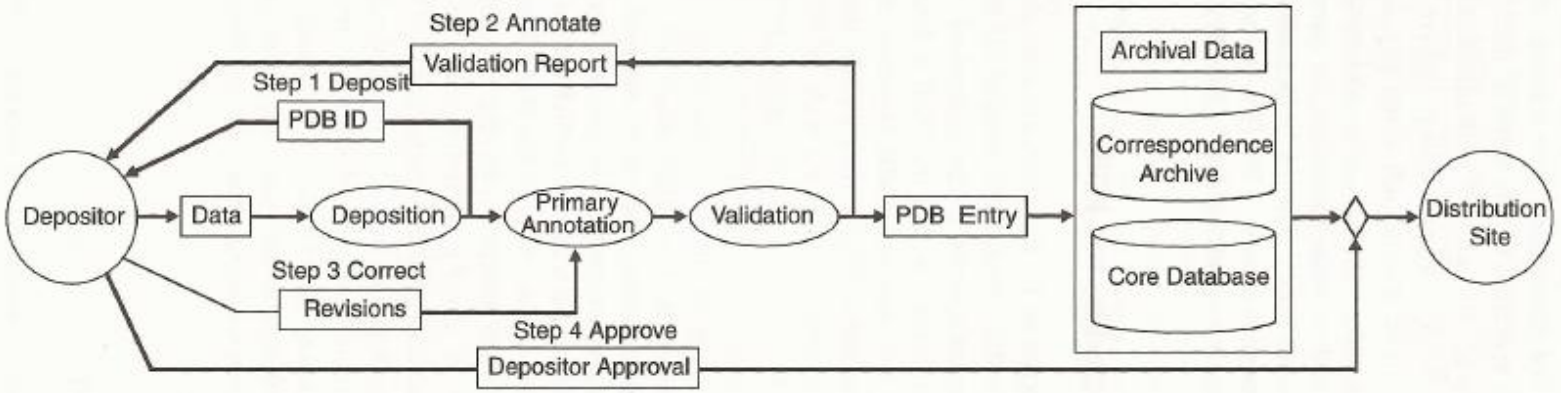


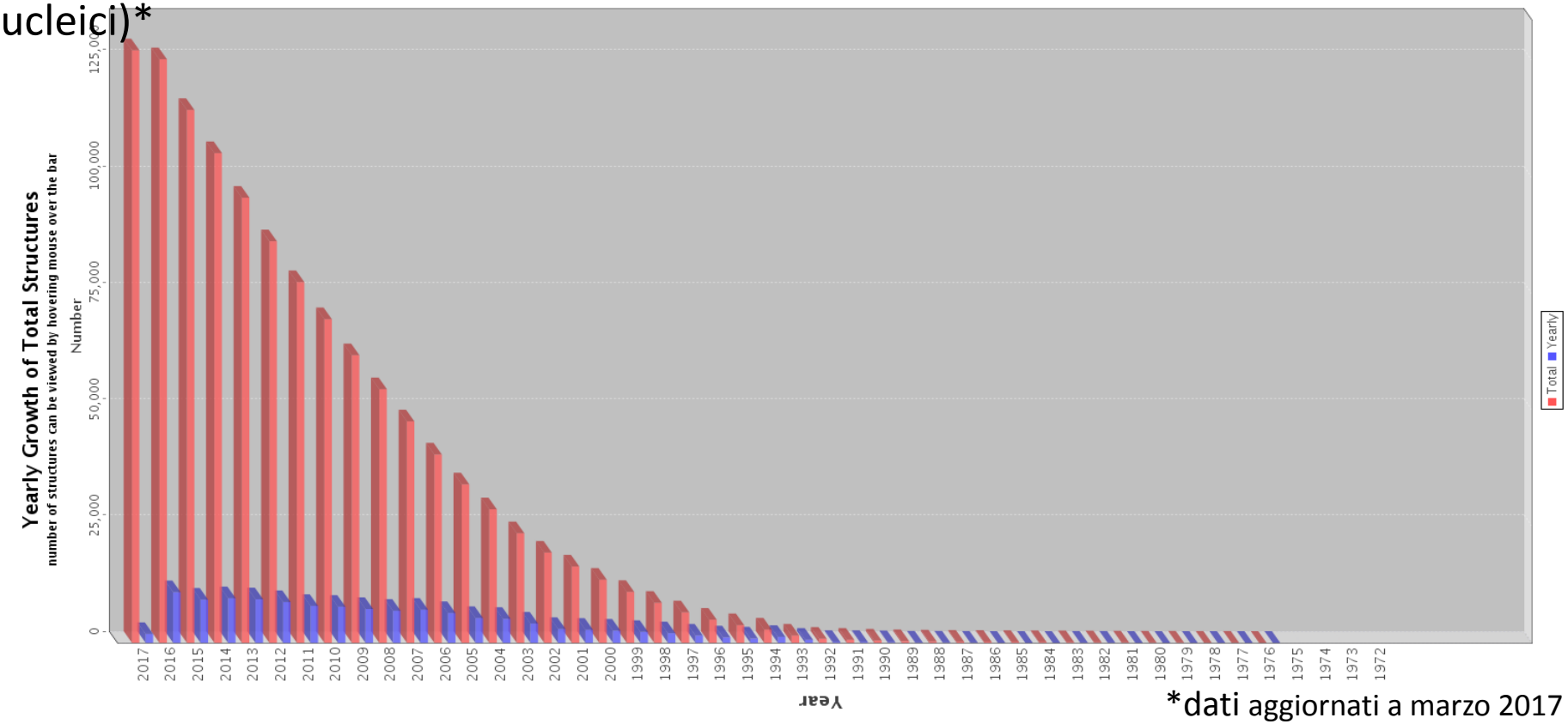

Figure 9.1. The steps involved in PDB data processing. Ellipses represent actions and rectangles define content. Figure reprinted from Berman et al. (2000b) with permission from the International Union of Crystallography.

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne.
(2000) The Protein Data Bank *Nucleic Acids Research*, 28: 235-242.

**1971**, Brookhaven National Laboratory, archivio di strutture cristallografiche di macromolecole biologiche.

**1976**: 7

**2017**: 127823 (**11680** proteine; 1689 DNA; 1252 RNA; 6070 complessi proteine-acidi nucleici)*




*dati aggiornati a marzo 2017

# Contenuto dei dati nella PDB

**Origine:** genere, specie, variante del gene, vettore , ospite, descrizione metodo di sintesi chimica;
**Sequenza**: completa di tutti i componenti macromolecolari
**Struttura chimica** cofattori/gruppi prostetici
**Nomi** di tutti componenti
**Descrizione qualitativa** delle caratteristiche strutturali
**Citazioni bibliografiche** della struttura sottomessa alla PDB
**Coordinate tridimensionali**

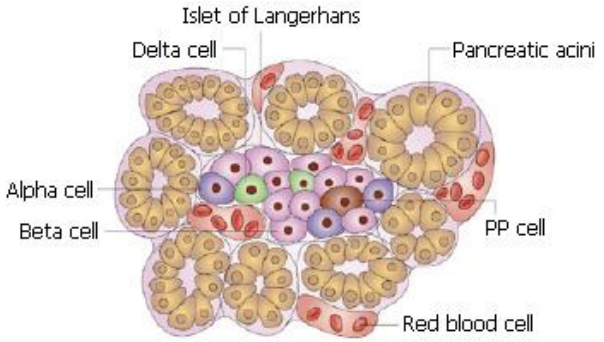
Dati aggiuntivi struttura X-Ray

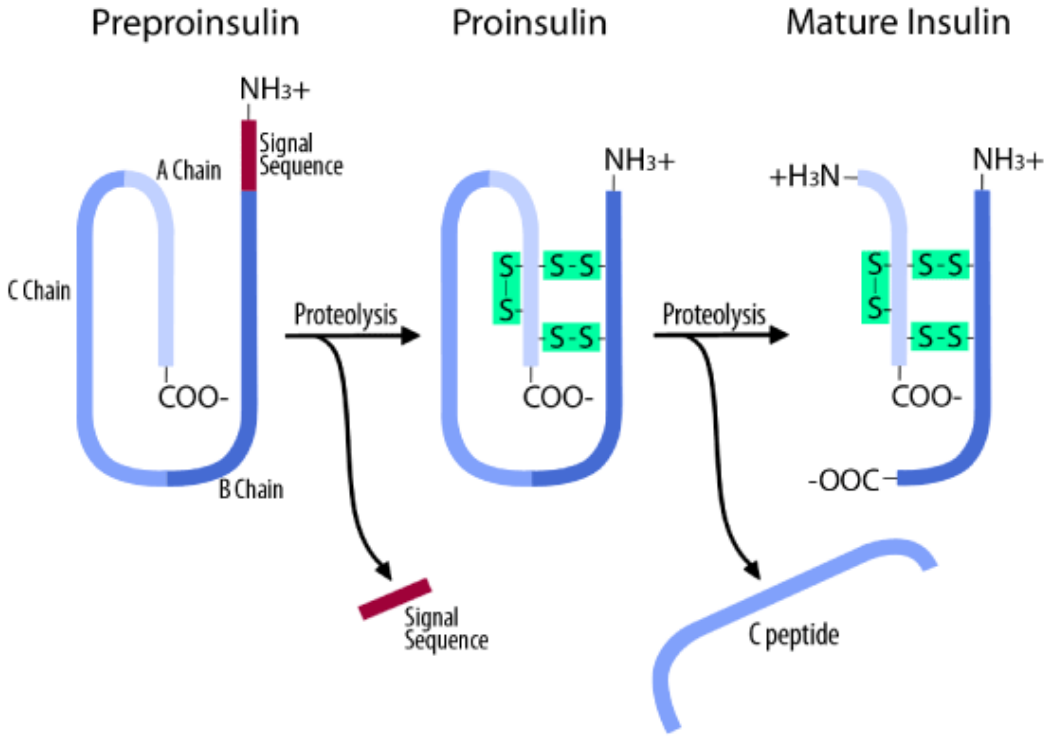
- Fattori di temperatura e posizione di ogni atomo;
- Condizioni di cristallizzazione (pH, solvente, temperatura, metodo)
- Dati del cristallo (dati cella elementare, gruppo spaziale simmetria)
- Informazione sul raffinamento della struttura

Dati aggiuntivi struttura NMR

- N. modello dell'insieme di strutture depositate
- Informazioni sperimentali/strumentali
- Lista esperimenti NMR
- File dei vincoli NMR usati per il calcolo

# Insulina
## (gene *INS*, 110 aa)




Pancreas (isole di *Langerhans*)

**Cellule β: <u>insulina</u>**

Cellule α: glucagone

Cellule δ: somatostatina

Cellule PP: peptide pancreatico PP

**Catena A**

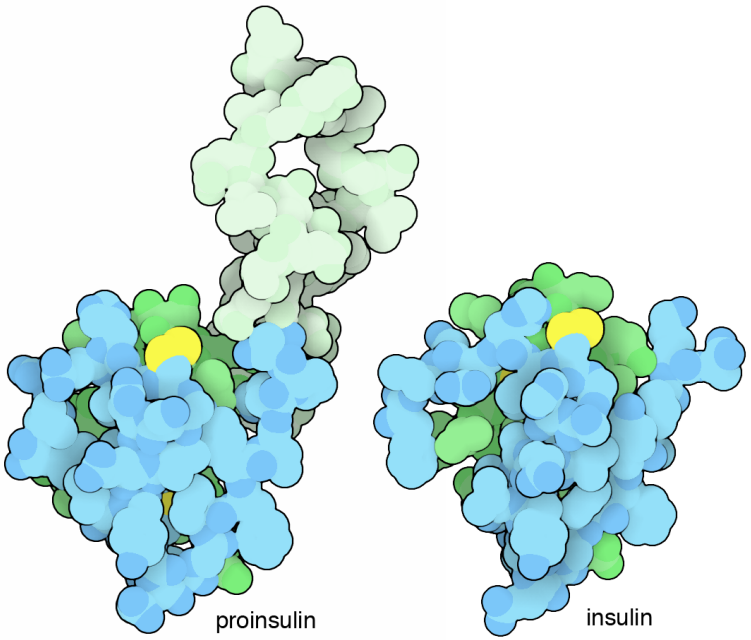
Gly-Ile-Val-Glu-Gln-Cys-Cys-Ala-Ser-Val-Cys-Ser-Leu-Tyr-Gln-Leu-Glu-Asn-Tyr-Cys-Asn

5 10 15 21

**Catena B**

Phe-Val-Asn-Gln-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Ala

5 10 15 20 25 30

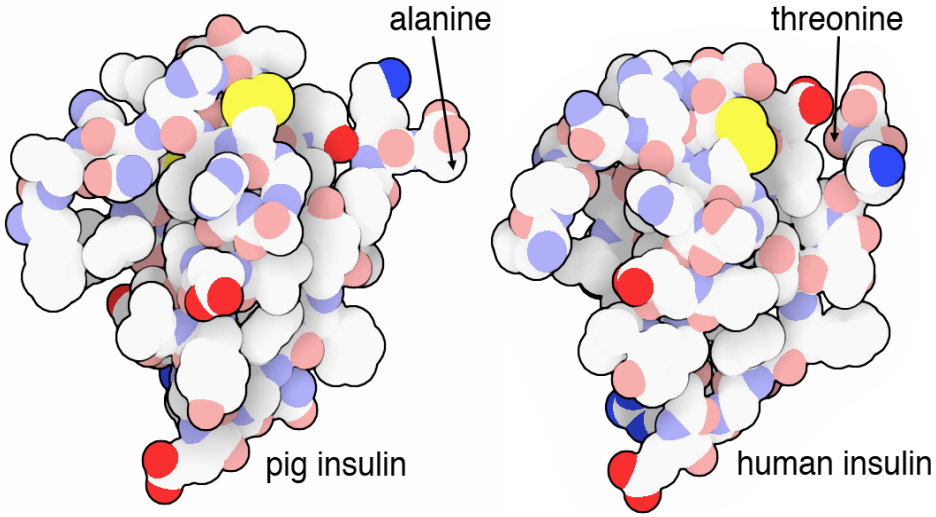



**Catena A**

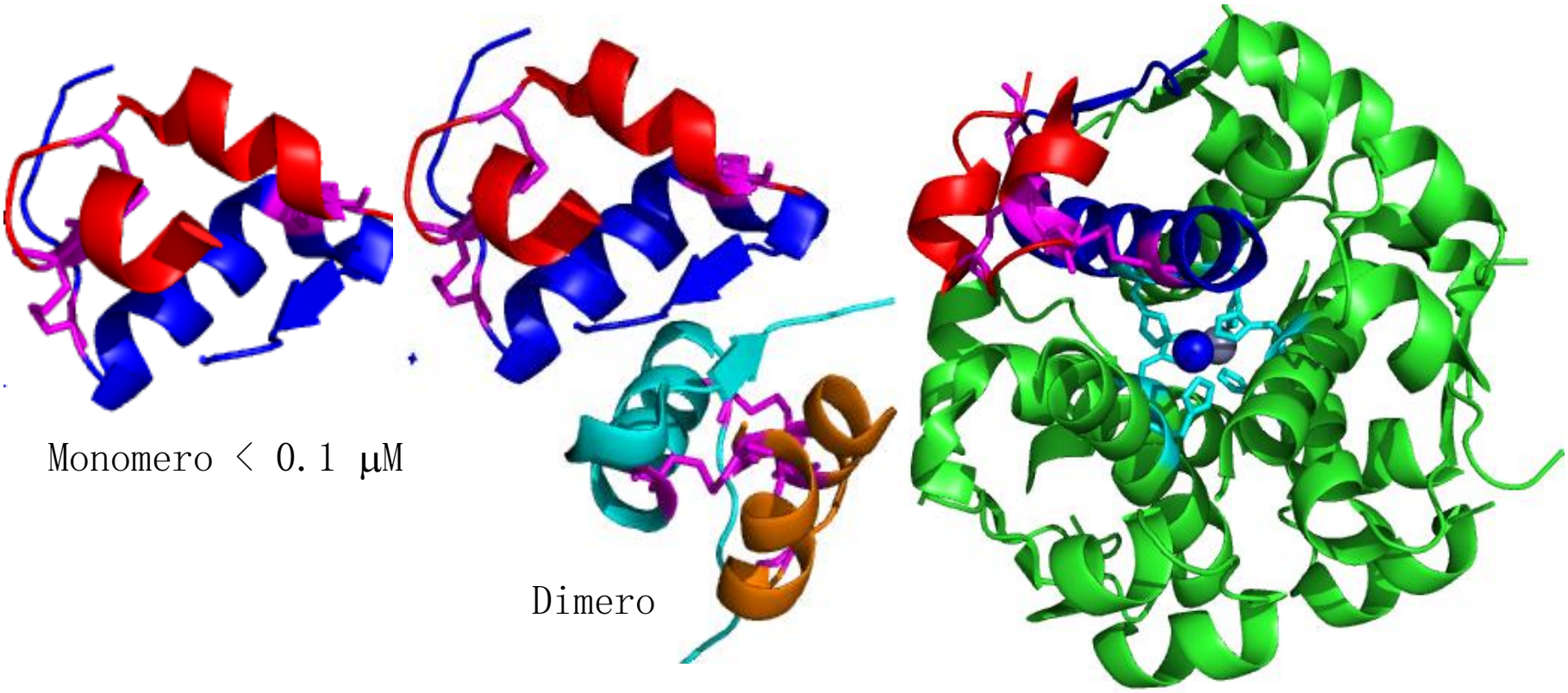
**Catena B**

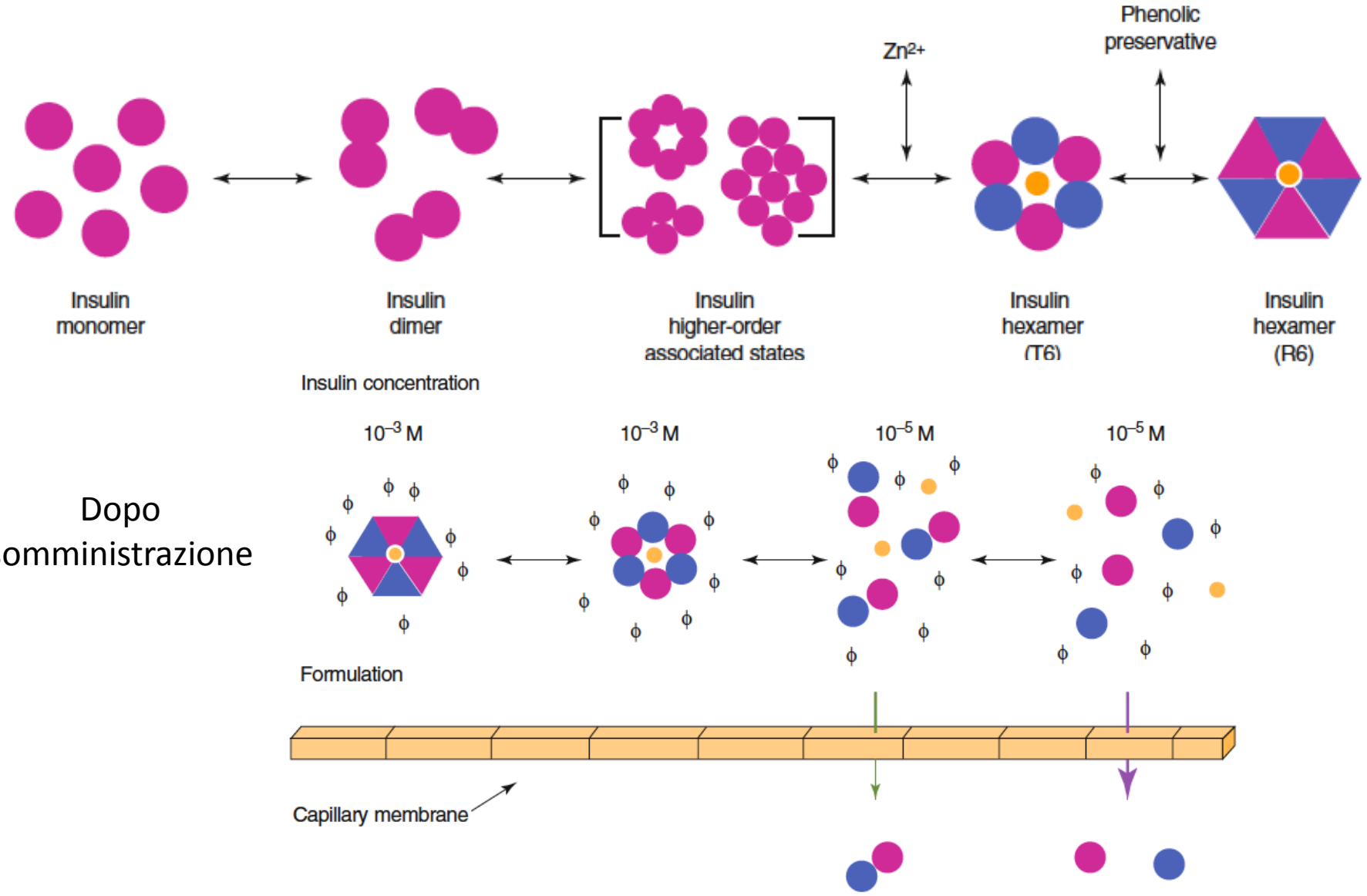
**4INS =** insulina di maiale

**2HIU** = insulina umana



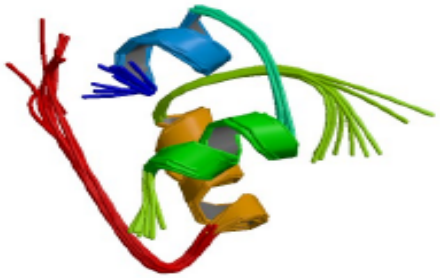

Insulina di maiale (1 aa) e insulina di mucca (3 aa) per la loro somiglianza con quella umana possono essere utilizzate in terapia in quanto riconosciute dalle nostre cellule


Monomero < 0.1 μM
Dimero
Esamero > 0.2 mM o con $Zn^{2+}$




Dopo somministrazione

NMR Ensemble



# 2HIU

NMR STRUCTURE OF HUMAN INSULIN IN 20% ACETIC ACID, ZINC-FREE, 10 STRUCTURES

**DOI:** 10.2210/pdb2hiu/pdb Entry 2HIU **supersedes** 1HIU

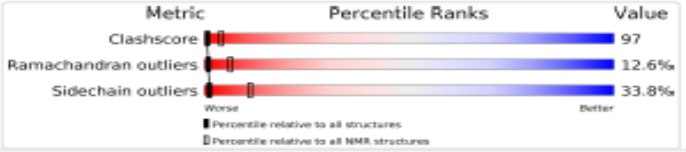
**Classification:** HORMONE
**Deposited:** 1996-10-08 **Released:** 1997-04-01
**Deposition author(s):** Hua, Q.X., Gozani, S.N., Chance, R.E., Hoffmann, J.A., Frank, B.H., Weiss, M.A.
**Organism:** Homo sapiens

**Structural Biology Knowledgebase:** 2HIU (>15 annotations)

**Experimental Data Snapshot**

**Method:** SOLUTION NMR
**Conformers Calculated:** --
**Conformers Submitted:** 10
**Selection Criteria:**

**wwPDB Validation** 3D Report | Full Report




**Macromolecule Content**

- Unique protein chains: 2

Literature

Structure of a protein in a kinetic trap.

Hua, Q.X., Gozani, S.N., Chance, R.E., Hoffmann, J.A., Frank, B.H., Weiss, M.A.

(1995) Nat.Struct.Biol. **2**: 129-138

**PubMed:** 7749917
Primary Citation of Related Structures: 1XGL 2HIU

**PubMed Abstract:**

Macromolecules

**Classification:** HORMONE

**Total Structure Weight:** 5817.68

**Macromolecule Entities**

| Molecule | Chains | Length | Organism | Details |
|---|---|---|---|---|
| INSULIN | A | 21 | Homo sapiens | INS Gene View |
| INSULIN | B | 30 | Homo sapiens | INS Gene View |

RCSB PDB PROTEIN DATA BANK An Information Portal to 127352 Biological Macromolecular Structures

http://www.rcsb.org/pdb/gene/INS

# INS - insulin

## Gene View

The Gene Browser allows to navigate the human genome and investigate the relationship between PDB entries and genes.

Number of PDB entities (unique chains) for this gene: 468

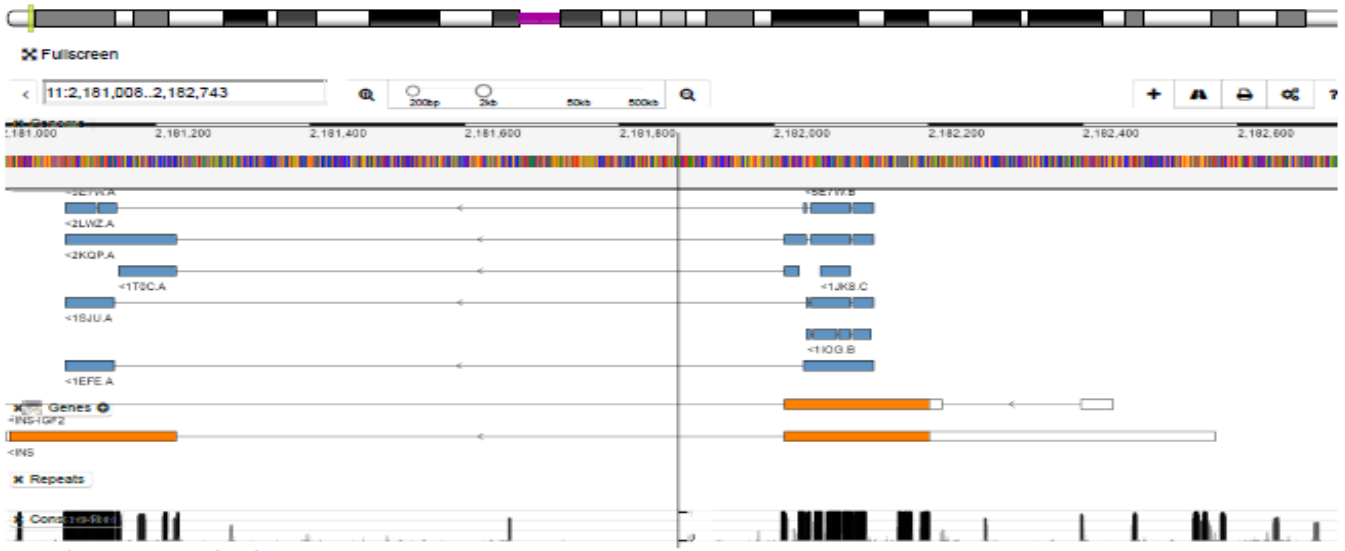
View list of all current human gene IDs

View protein features: Protein Feature View

| Cross References | |
|---|---|
| UniProt: P01308 | HGNC Approved Gene Symbol: INS |
| Previous Symbols: IDDM2, IDDM1 | Ensembl ENSG00000254647 |
| | Previous Names: "insulin-dependent diabetes mellitus 2" |
| Hgncid : HGNC:6081 | Omim: 176730 |
| Refseq: NM_000207 | GenBank: X70508 |

Genomic coordinates: Cytogenetic location: 11p15.5 chr11:2,181,081-2,182,201 Chromosome Location reset view

Genome Assembly: GRCh37




This feature requires an up-to-date web browser.

RCSB PDB PROTEIN DATA BANK An Information Portal to 127823 Biological Macromolecular Structures

# Full protein feature view for P0138

## Insulin - P01308 (INS_HUMAN)

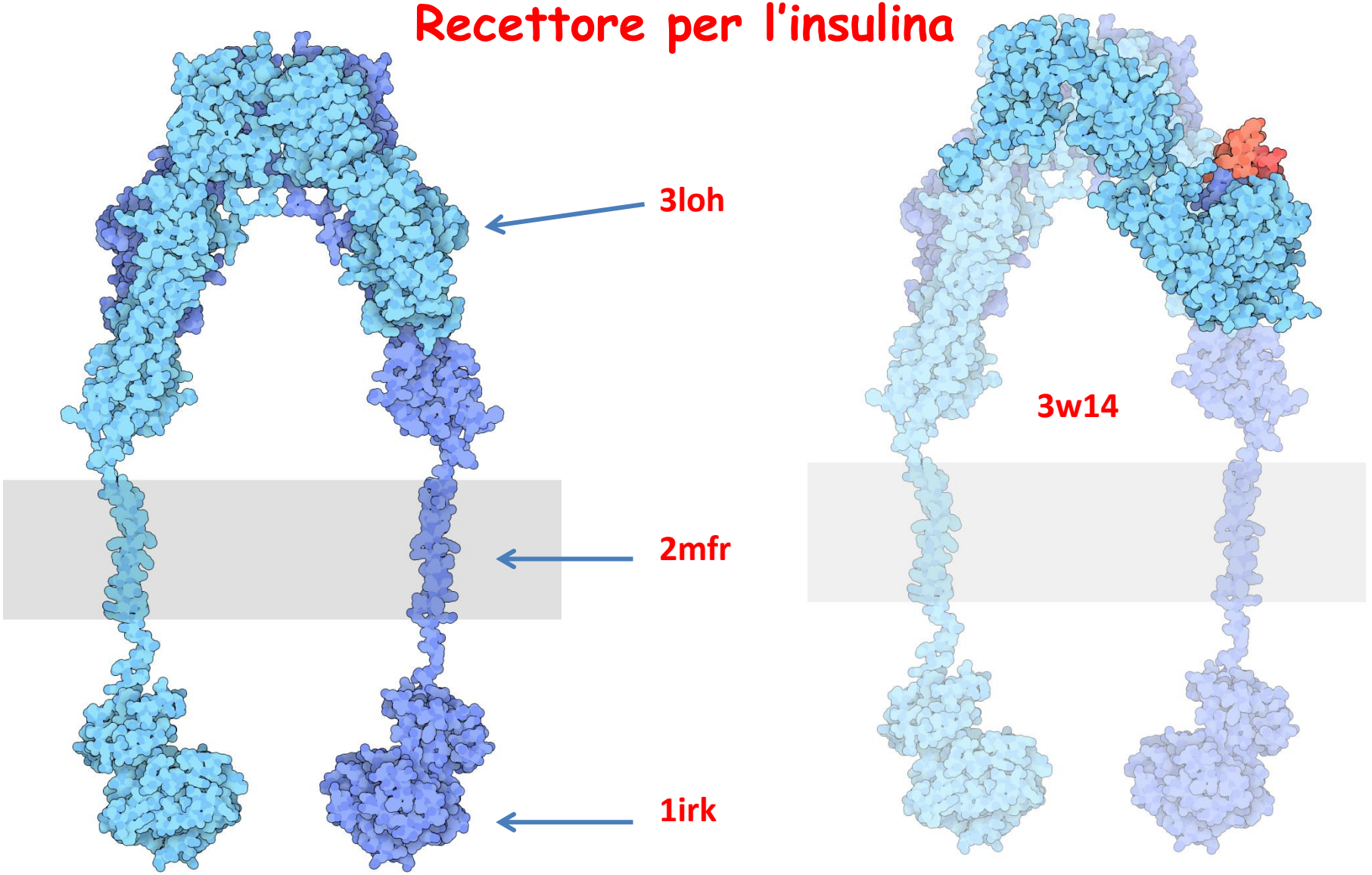
### Protein Feature View of PDB entries mapped to a UniProtKB sequence

Number of PDB entries for P01308: **240**

**Function** Insulin decreases blood glucose concentration. It increases cell permeability to monosaccharides, amino acids and fatty acids. It accelerates glycolysis, the pentose phosphate cycle, and glycogen synthesis in liver. UniProt

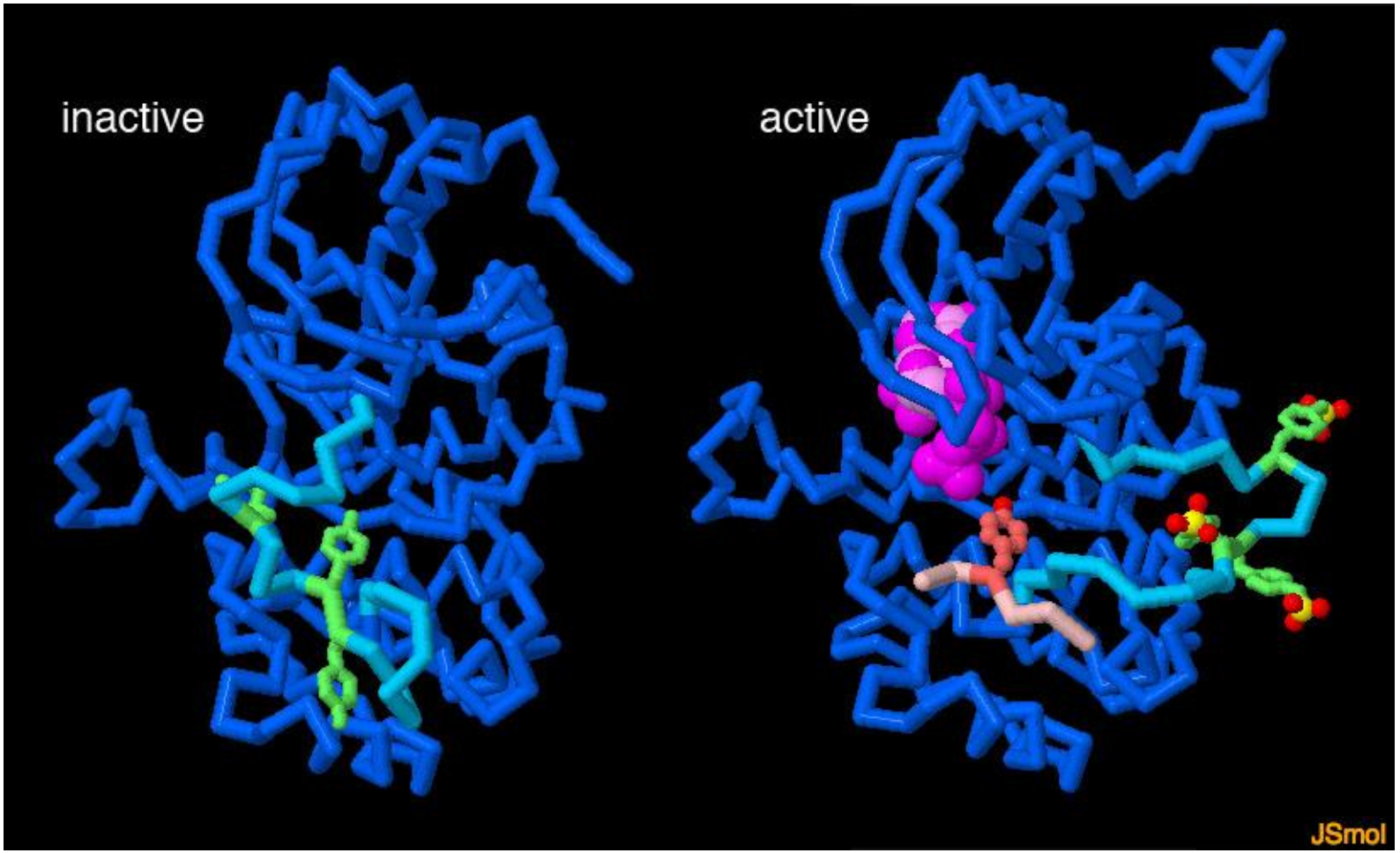
**Subunit Structure** Heterodimer of a B chain and an A chain linked by two disulfide bonds (PubMed:25423173). UniProt

# Recettore per l'insulina



# Dominio tirosin chinasico del recettore nella forma inattiva ed attiva

**1irk** **1ir3**



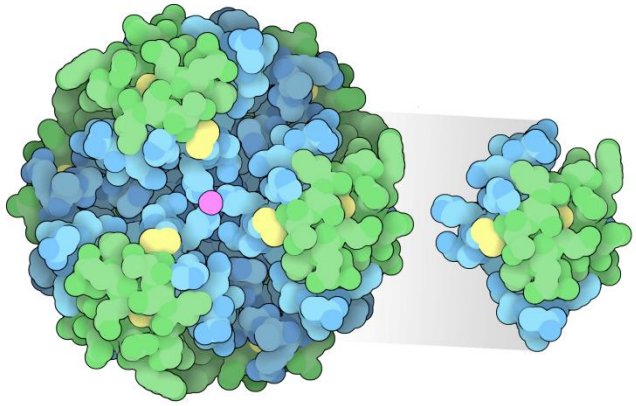
# Insuline "progettate"

- ✓ Prime preparazioni: estrazioni da bestiame (maiali/mucche)
- ✓ Oggi: tecnologia DNA ricombinante da batteri (*E.coli*) che producono insulina umana pura (***Humulin****)
- ✓ L'insulina ricombinante abbassa rapidamente il livello di glucosio del sangue ma i suoi effetti decadono dopo qualche ora -> va bene subito dopo i pasti
- ✓ Progettazione di insuline geneticamente modificate che agiscano in tempi più lunghi e che aiutino a mantenere il controllo basale del glucosio nel sangue

## Insuline di lunga durata

- ✓ L'insulina è immagazzinata nelle cellule β del pancreas come esamero (A6B6) ma si scinde in un monomero attivo quando è trasportata nel sangue e si lega al recettore -> occorre rallentare la formazione del monomero
- ✓ Conservanti (fenolo, cresolo) sono aggiunti ai preparati per stabilizzare l'esamero

Struttura di umana insulina ricombinante (**1trz**)*



# Insuline "progettate"

**Insuline ricombinanti**

| Insuline | 28 | 29 | 30 |
|---|---|---|---|
| WT | Pro | Lys | Thr |
| LISPRO | Lys | Pro | Thr |
| ASPART | Asp | Lys | Thr |

**Azione rapida**

**GLARGINE**

AsnA21Gly + ArgB31ArgB32

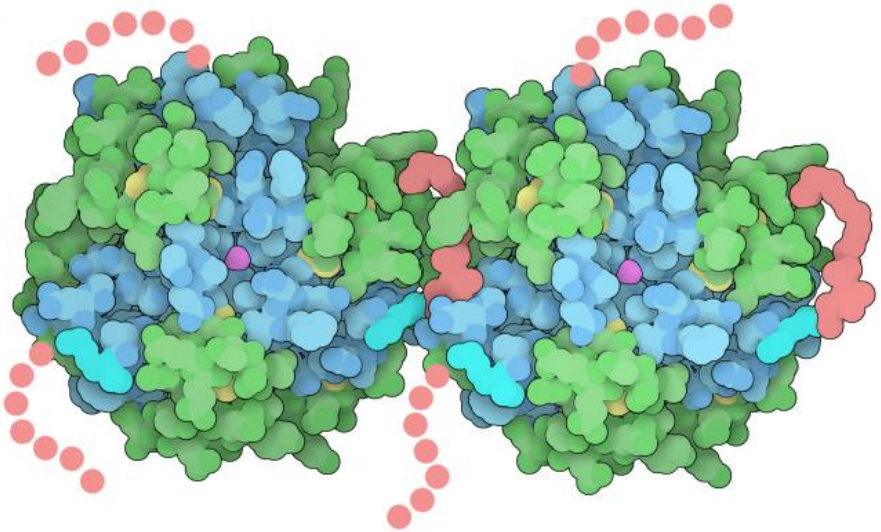
**Azione lenta**

## Insuline di lunga durata

✓Insulina + protamina di pesce = complesso eterogeneo che si scioglie più lentamente

✓**Glargine** (Lantus ®) (PDB entry **4iyf**), (+ 2 Arg e Asn -> Gly) = insulina meno solubile, forma aggregati che rilasciano il monomero più lentamente

✓**Degludec** (Tresiba ®) (PDB entry **4ajx**) = insulina cui sono state attaccate lunghe catene idrocarburiche all'estremità C-ter delle catene B dell'insulina in modo da formare aggregati fra esameri che si sciolgono più lentamente

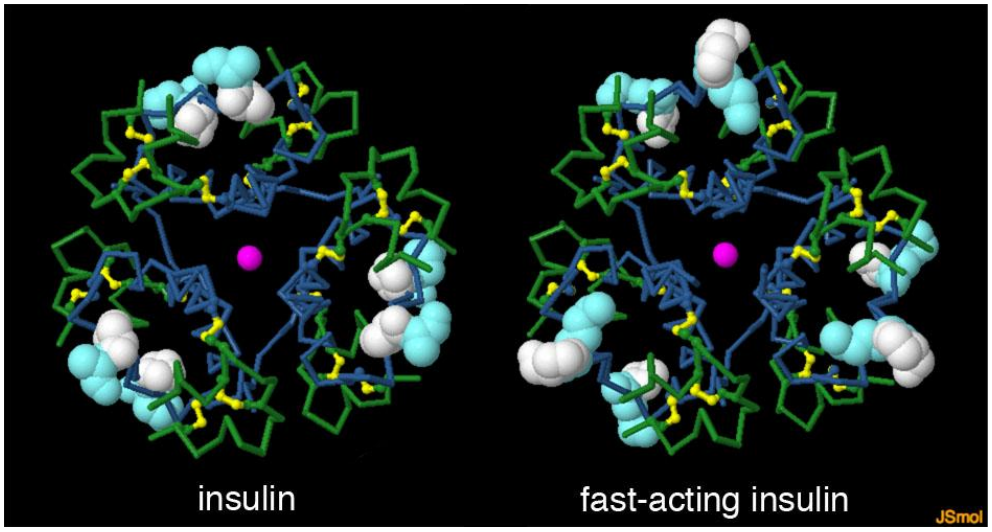


## Insuline di lunga durata + insuline ad azione rapida

✓Combinazione delle 2 insuline per creare un'insulina ad azione rapida che destabilizzi l'esamero così che questo decada rapidamente quando entra nel torrente circolatorio

✓**Lispro** (Humalog ®) (PDB entry **1lph**), coinvolge l'inversione di 2 residui all'estremità C-ter della catena B (Pro-Lys/Lys-Pro) -> si indebolisce l'interazione fra i monomeri, rendendo il mutante 200 volte meno stabile della forma nativa

✓**Aspart** (Novolog ®) (PDB entry **4gbc**). Viene sostituita la Pro in quella stessa posizione con un aa carico (Asp) -> si ha stesso risultato



orphanet http://www.orpha.net/consor4.01/www/cgi-bin/?lng=IT

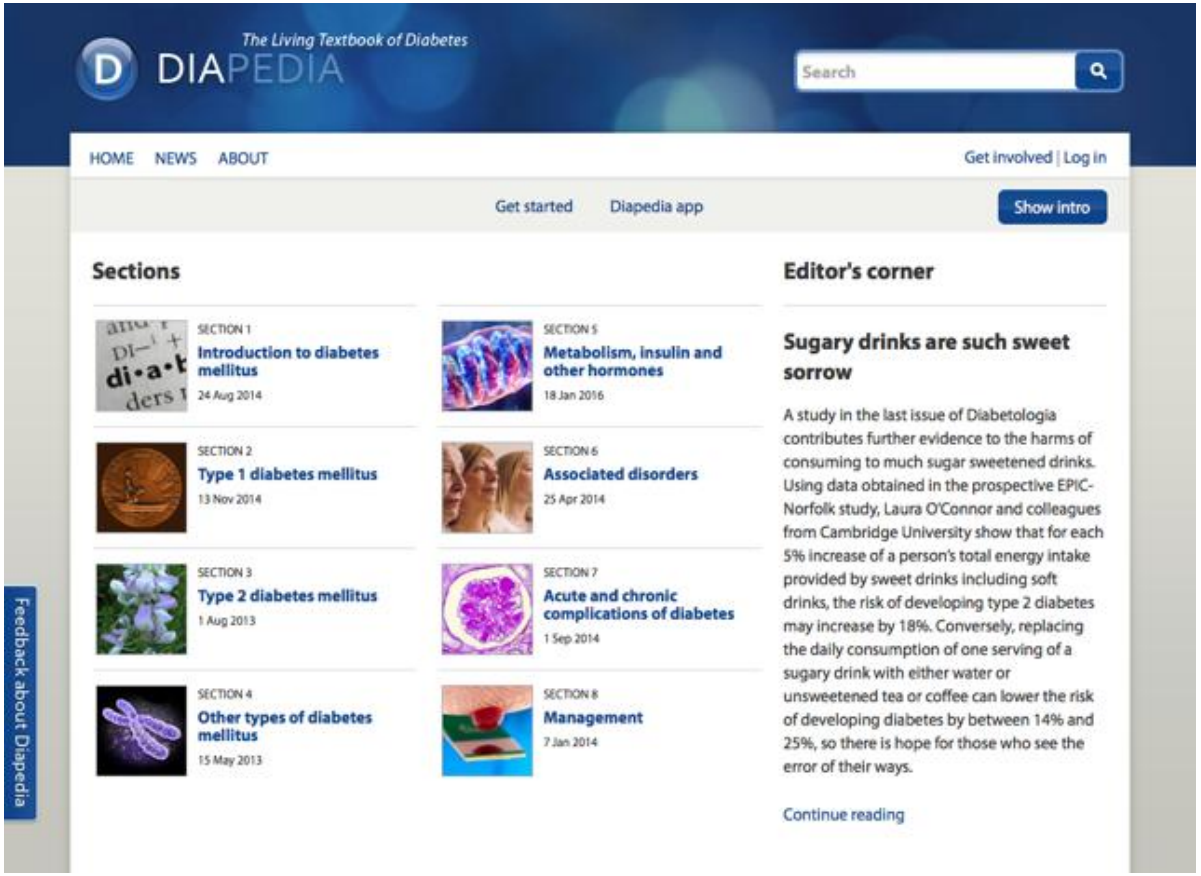
## Diabete mellito

**Tipo I**, giovanile, è insulino-dipendente. Insulina insufficiente o assente (**ORPHA:181371**)

**Tipo II**, adulto, resistenza acquisita all'insulina (**ORPHA:181376**)

World Health Organization http://www.who.int/diabetes/en

https://www.diapedia.org

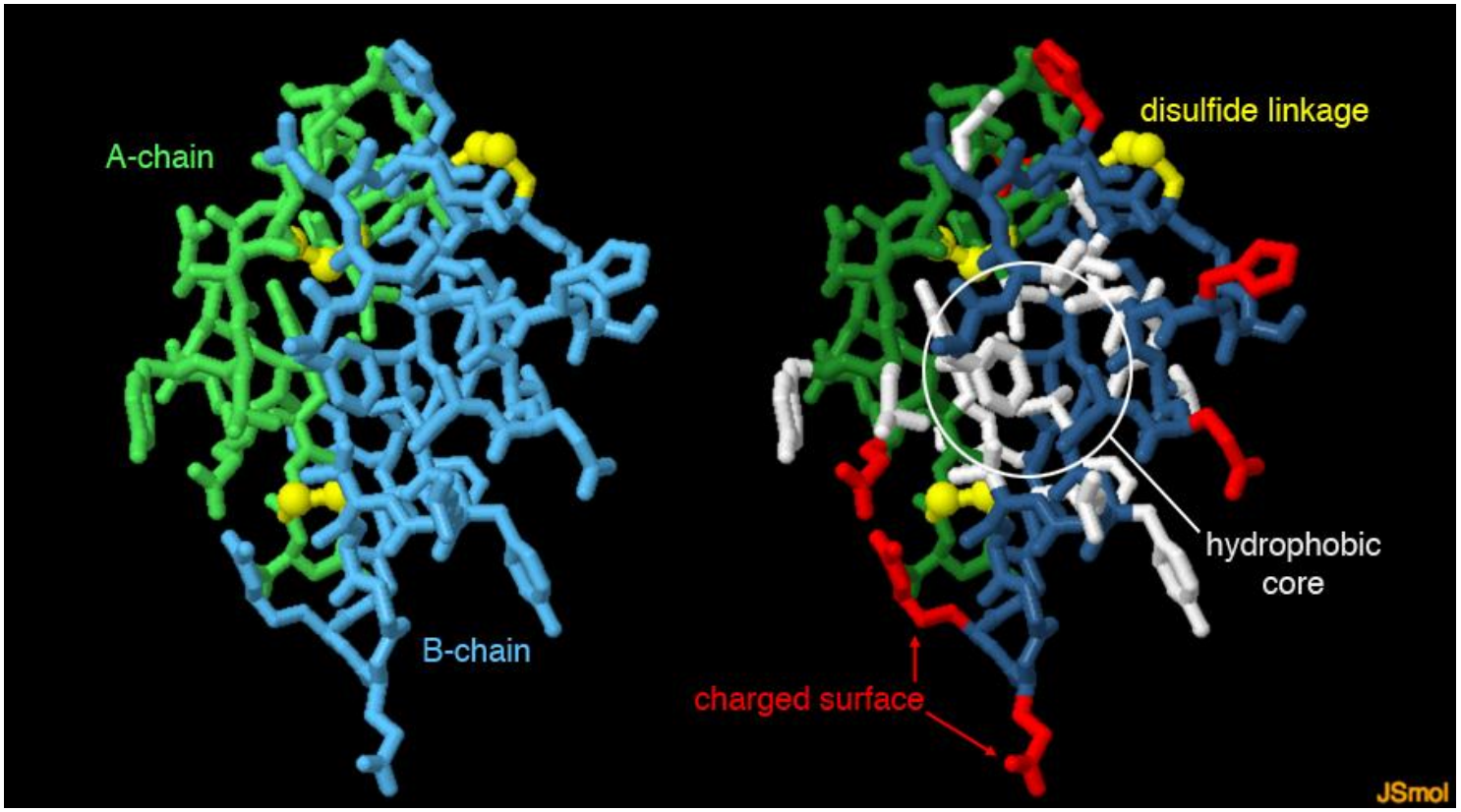


RCSB PDB-101

Nel portale educational della PDB si può trovare materiale didattico utile per esercitazioni sulle varie proteine.

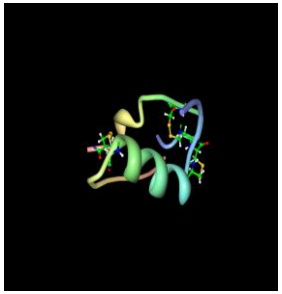
Per esempio all'indirizzo
http://pdb101.rcsb.org/learn/resources/paper_models
Si può scaricare il modello in carta dell'insulina da fare costruire ai ragazzi

# Insulin: 1trz




Insulina = modello perfetto per esplorare la struttura di una proteina
(è semplice da studiare  solo  51 aa,  circa 800 atomi)

# ESERCITAZIONE PRATICA



Andare al sito della PDB

http://www.rcsb.org/pdb/

Dare come parola chiave nella stringa della *search*: **1TRZ** (insulina umana)

✓Vedere se la struttura è stata determinata con cristallografia o NMR

Utilizzare i programmi di grafica in **Standalone Viewers**

**Esplorare il backbone dell'insulina usando il programma *Simple Viewer***

✓Cliccando con il mouse sulla struttura appare il nome ed il numero del residuo toccato con il tipo di atomo selezionato

✓Identificare le estremità N e C terminale della molecola per entrambe le catene A e B

✓Quali sono il primo e l'ultimo amminoacido che appaiono nella struttura?

✓Identificare cosa sono le sfere mostrate

**Visualizzazione della struttura con il programma *Protein Workshop***

✓Vedere i ponti S-S fra quali residui di Cys si formano (selezionare le Cys nelle catene A e B in Visibility sotto Atoms and bonds)

**Guardare la struttura dell'insulina in forma esamerica con *Jsmol* in View in 3D**

✓Selezionare il programma di grafica ***JSmol*** e in *Structure* selezionare ***Biological Assembly 3***

✓Selezionare le diverse modalità di rappresentazione (cartoon/backbone…)/display mode