

Bioinformatica ed applicazioni di Bioinformatica Strutturale

Parte 2: Bioinformatica Strutturale

Sotto-disciplina della bioinformatica che comprende:

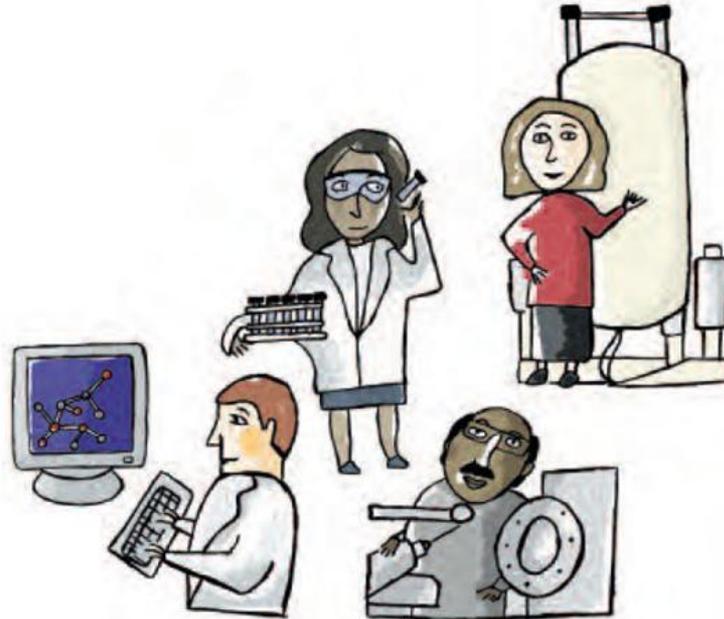
- ✓ rappresentazione,
- ✓ deposito,
- ✓ recupero,
- ✓ analisi,
- ✓ visualizzazione grafica

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

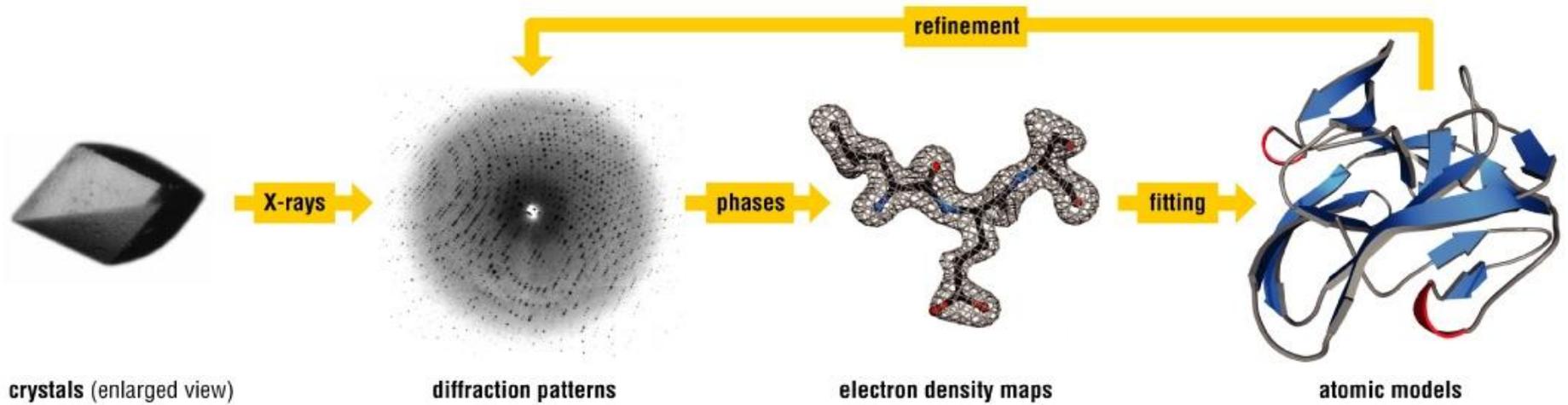
Origine dati strutturali

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



X-Ray

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

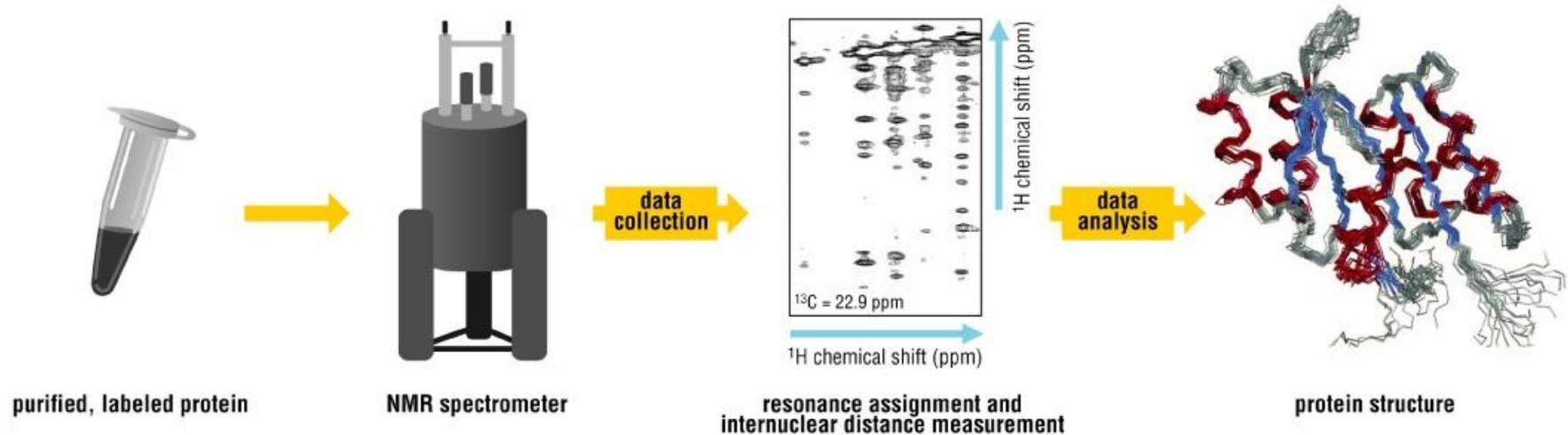


© 1999–2004 New Science Press

Risoluzione buona $\leq 1.5 \text{ \AA}$

NMR

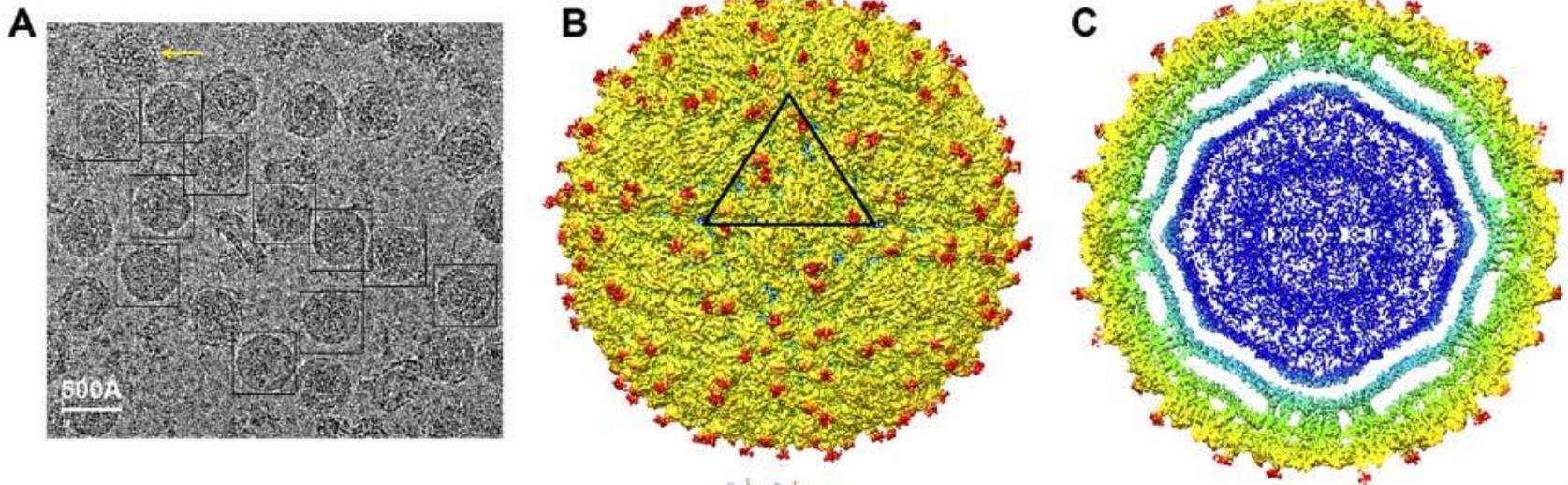
From **Protein Structure and Function** by Gregory A Petsko and Dagmar Ringe



© 1999–2004 New Science Press

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

Criomicroscopia elettronica



Risoluzione buona $3 \div 4 \text{ \AA}$

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.

[About NDB](#)[Standards](#)[Education](#)[Tools](#)[Software](#)[Download](#)

A Portal for Three-dimensional Structural Information about Nucleic Acids
As of 1-Mar-2017 number of released structures: 8773

[Search DNA](#) [Search RNA](#) [Advanced Search](#)

Enter an NDB ID or PDB ID 
Search for released structures

Welcome to the NDB

The NDB contains information about experimentally-determined nucleic acids and complex assemblies. Use the NDB to perform searches based on annotations relating to sequence, structure and function, and to download, analyze, and learn about nucleic acids.

Search Structures

[Search DNA](#)
Search DNA and its complexes

[Search RNA](#)
Search for RNA structures in the NDB archive or in the Non-Redundant list

[Advanced Search](#)
Search for structures based on structural features, chemical features, binding modes, citation and experimental information



Featured Tools

[RNA 3D Motif Atlas](#), a representative collection of RNA 3D internal and hairpin loop motifs

[Non-redundant Lists](#) of RNA-containing 3D structures

[RNA Base Triple Atlas](#), a collection of motifs consisting of two RNA basepairs

[WebFR3D](#), a webserver for symbolic and geometric searching of RNA 3D structures

[R3D Align](#), an application for detailed nucleotide to nucleotide alignments of RNA 3D structures



Helen M. Berman, Rutgers University, Wilma K. Olson, Rutgers University, David Beveridge, Wesleyan University

ndbadmin@ndbserver.rutgers.edu

©1995-2015 The Nucleic Acid Database Project | Rutgers, The State University of New Jersey

Banche dati per strutture di proteine



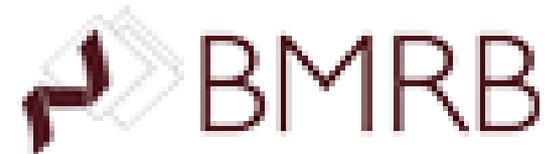
- ✓ **RCS PDB** = Research Collaboratory for Structural Bioinformatics Protein Data Bank <http://www.rcsb.org/>



- ✓ **PDBe** = Protein Data Bank in Europe <http://www.ebi.ac.uk/pdbe/>



- ✓ **BMRB** = Biological Magnetic Resonance Data Bank <http://www.bmrb.wisc.edu/>

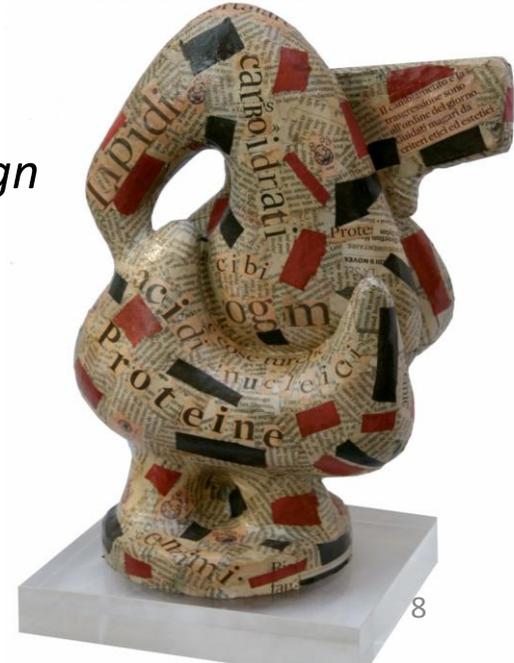


- ✓ **PDBj** = Protein Data Bank Japan <http://pdbj.org/>

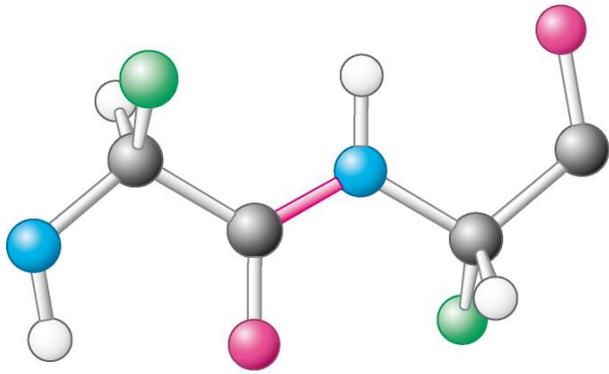


Perché è importante conoscere la struttura 3D di una proteina?

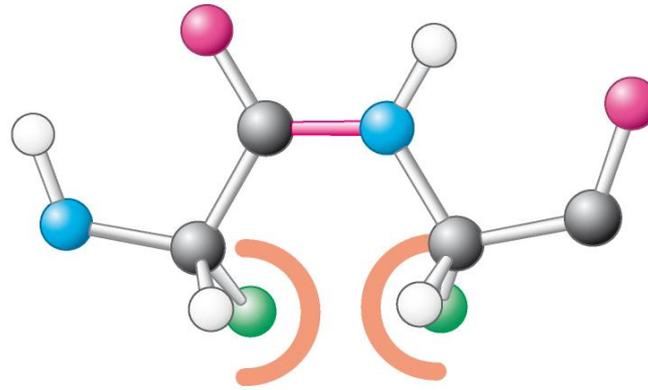
- ✓ Una proteina svolge la sua funzione biologica solo se è ripiegata nella “giusta” struttura
- ✓ Una proteina non ripiegata correttamente o non funziona affatto o funziona male
- ✓ Conoscere la struttura è la chiave per avere informazioni sul meccanismo di funzionamento
- ✓ L’informazione può essere utilizzata anche per il *drug-design*
- ✓ La struttura non è statica



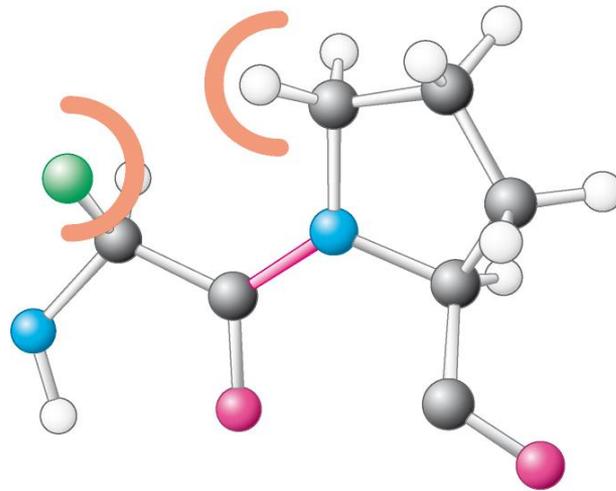
Legame peptidico



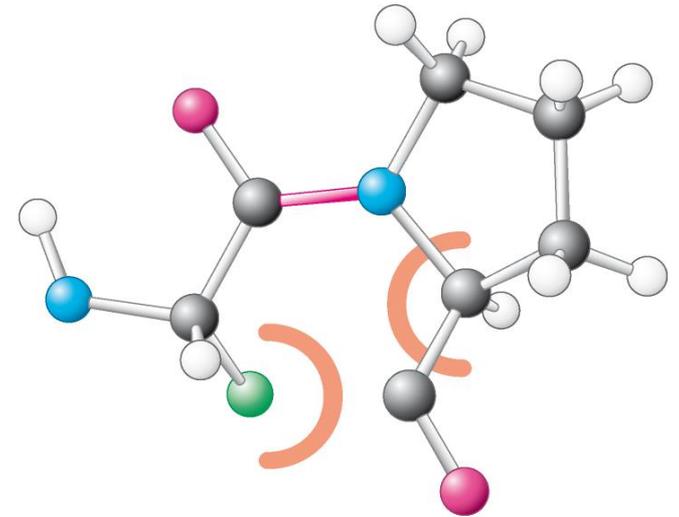
Trans



Cis

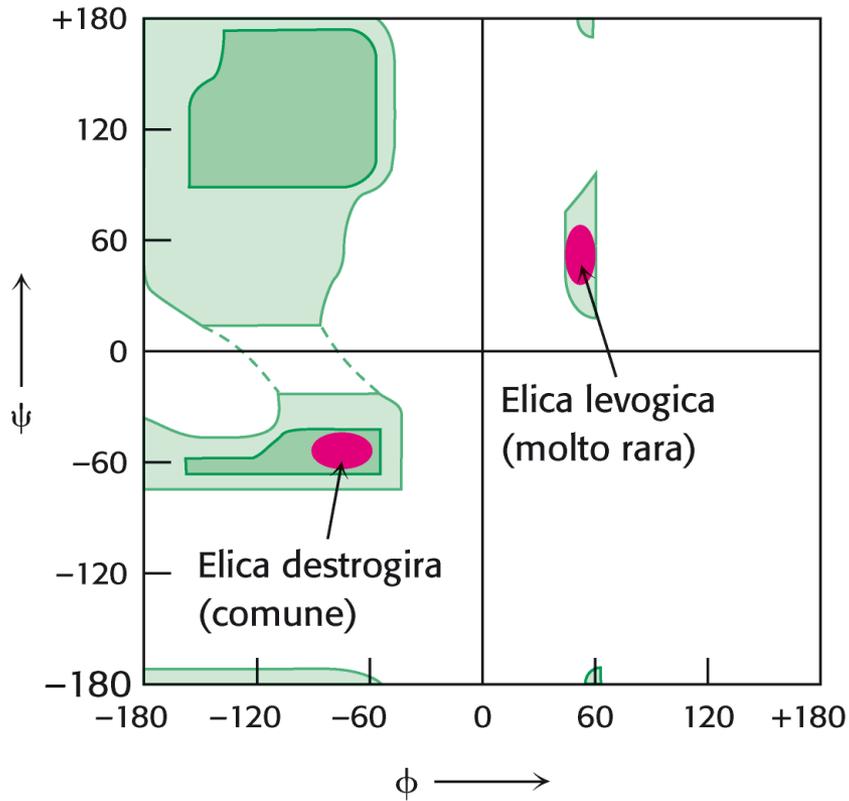
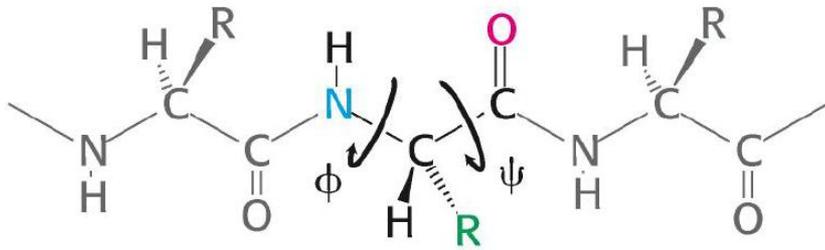


Trans

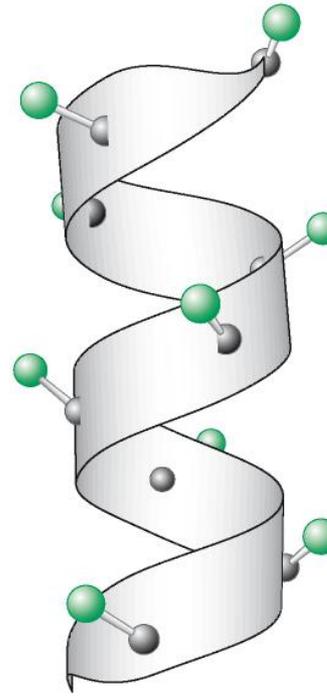


Cis

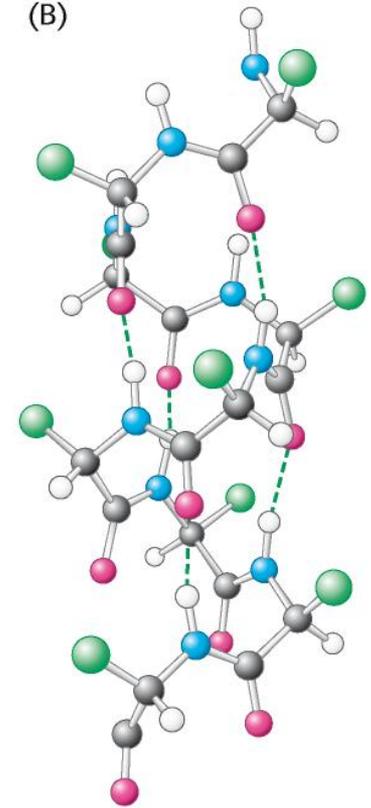
Alfa elica



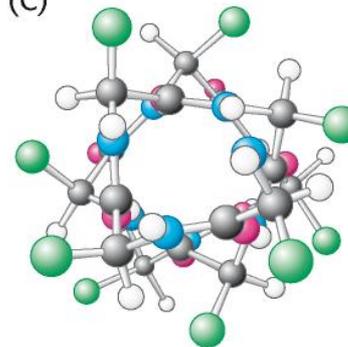
(A)



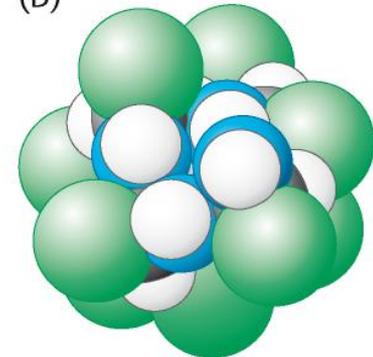
(B)



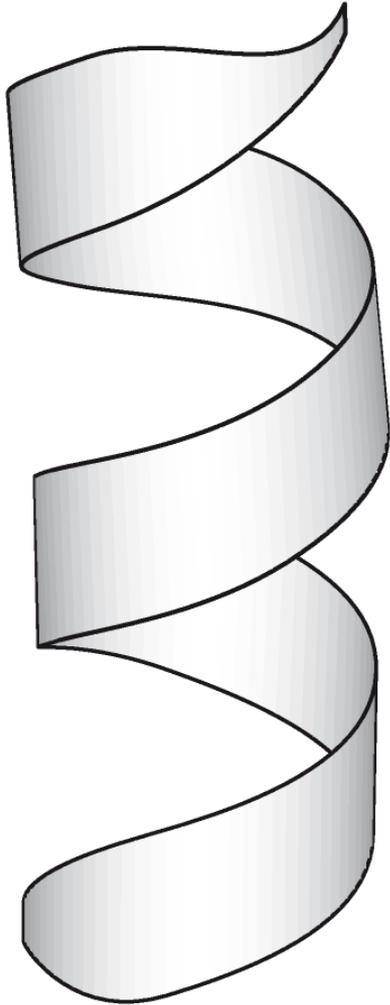
(C)



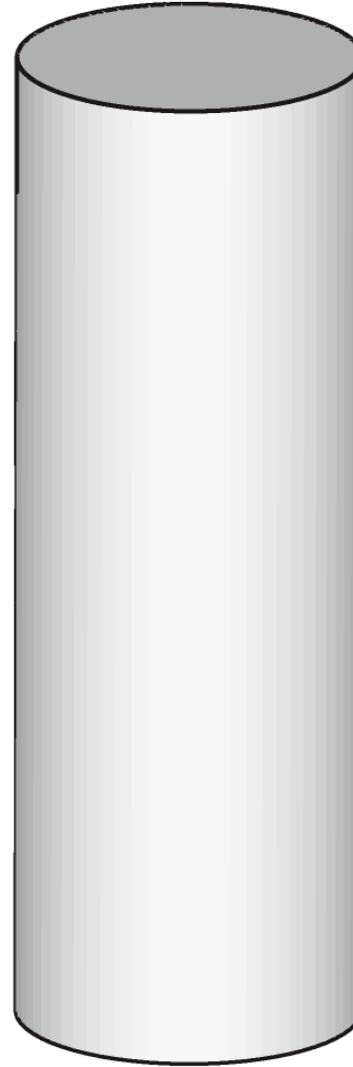
(D)



(A)

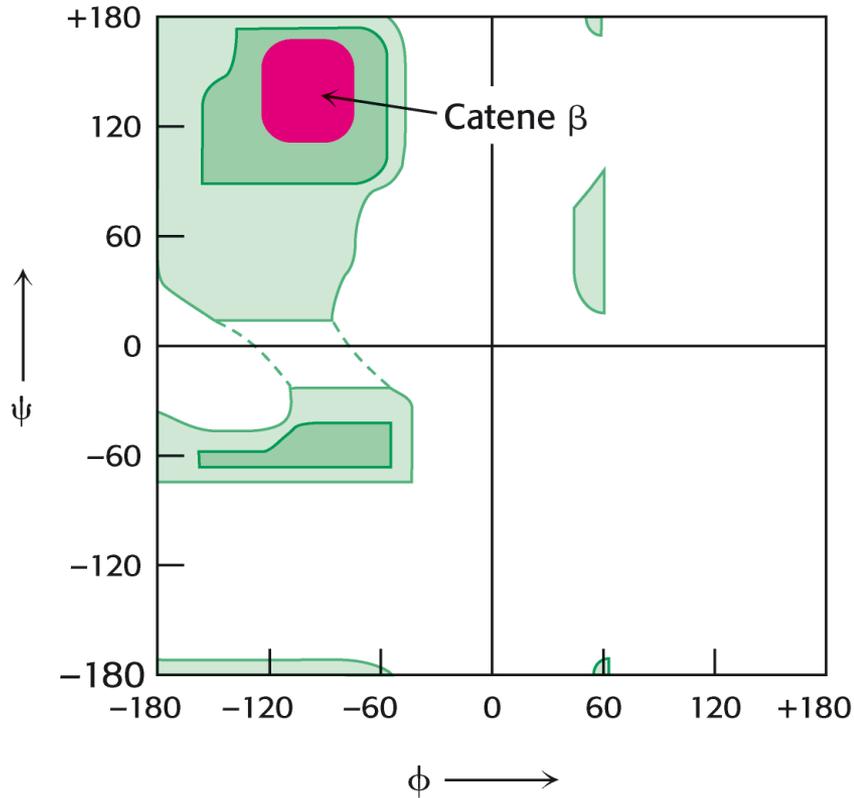
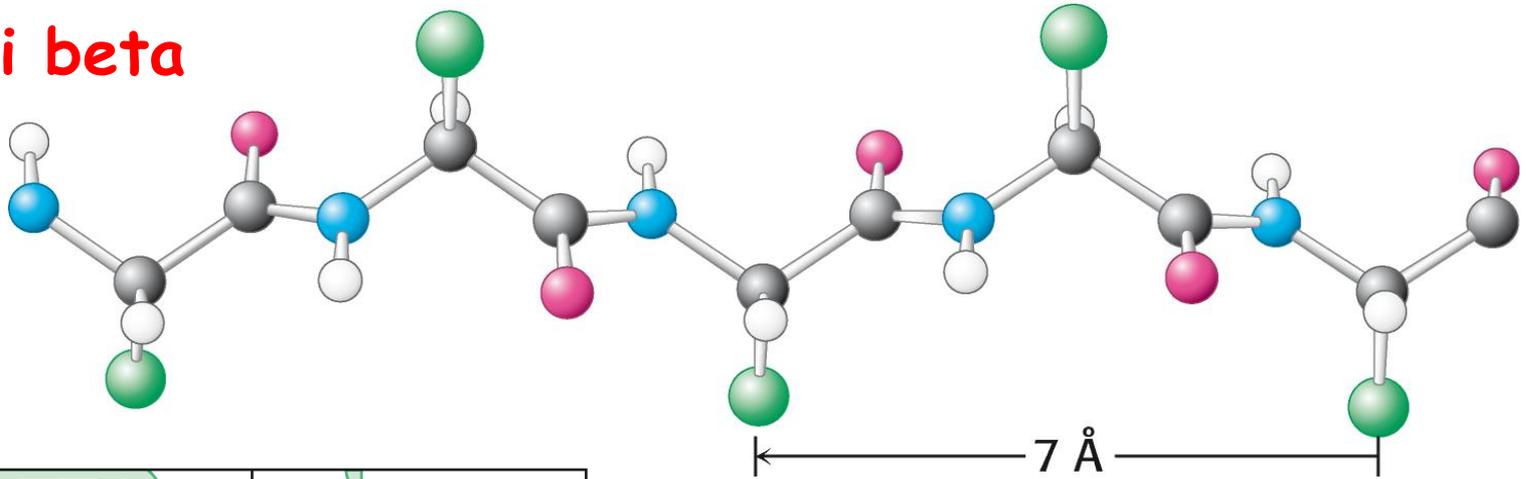


(B)

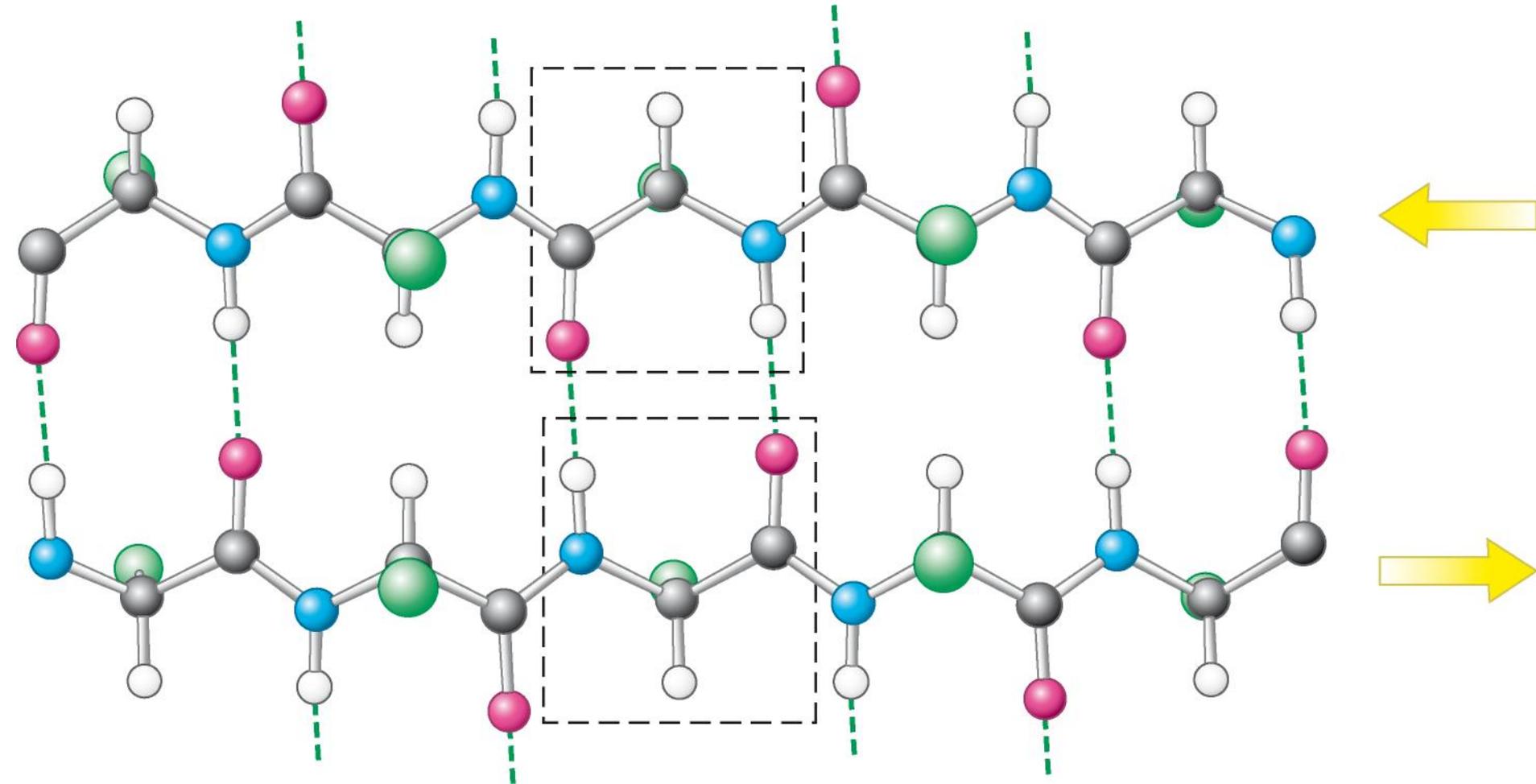


Alfa elica

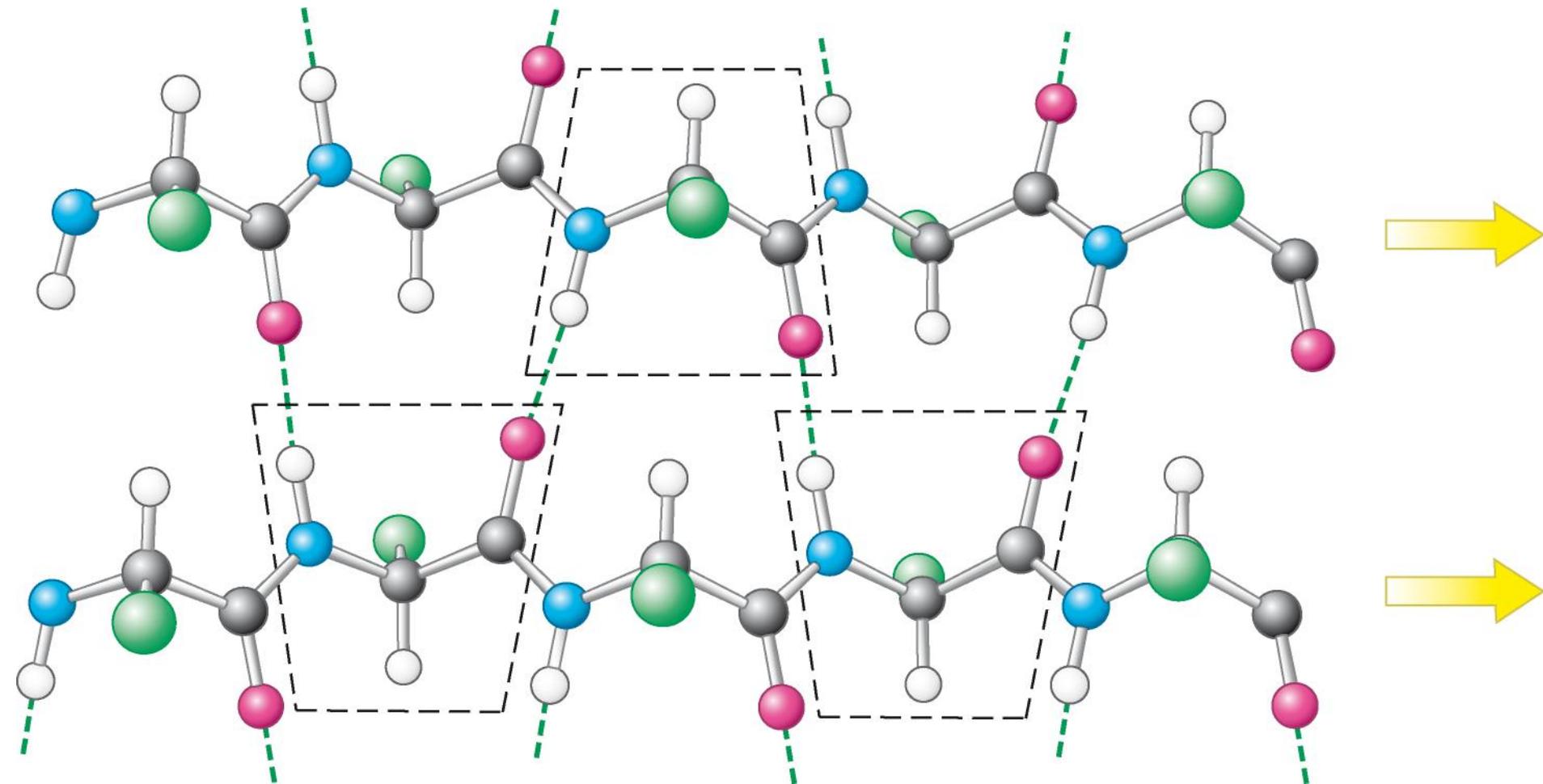
Foglietti beta



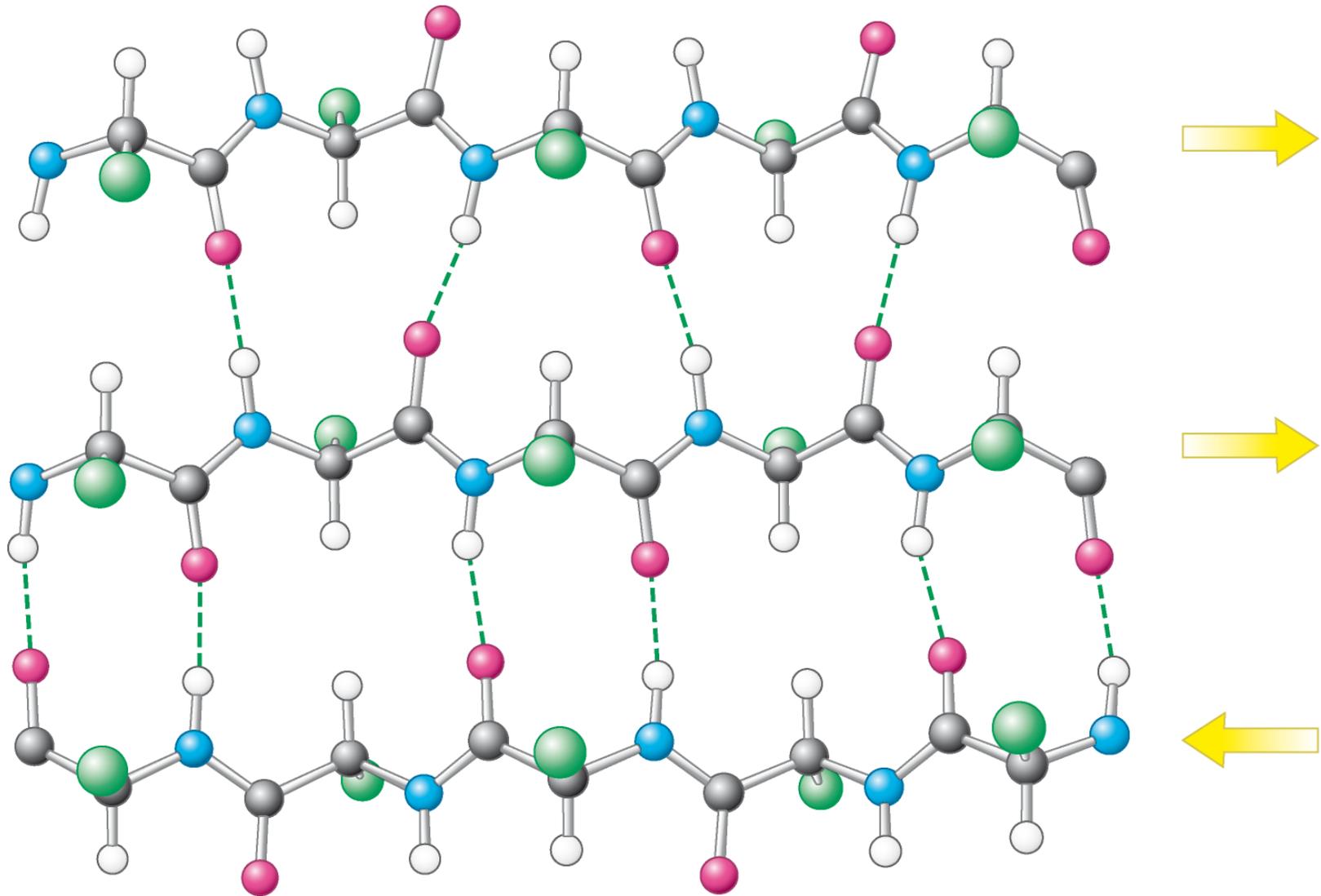
Foglietti beta antiparalleli

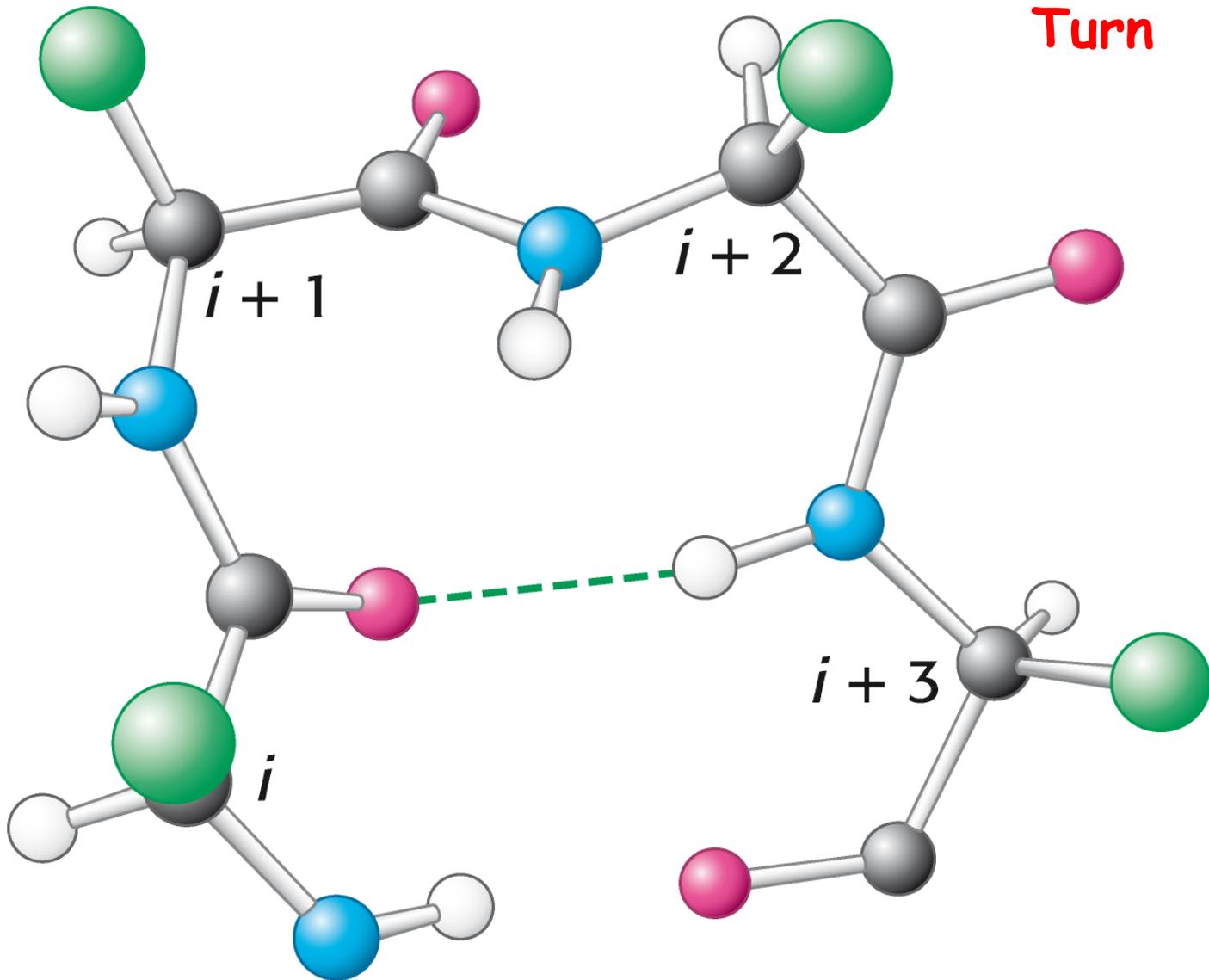


Foglietti beta paralleli



Foglietti beta misti





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



ExPASy SIB (Swiss Institute) Bioinformatics Resource Portal

<https://www.expasy.org/>

Visual Guidance

DNA
RNA
Protein
Cell
Organism
Population

Categories
Resources A..Z
Links/Documentation

Visual Guidance Interface

Please select an element:

DNA RNA Protein Cell Organism Population

Predizione della struttura secondaria di una proteina

ExPASy SIB (Swiss Institute) Bioinformatics Resource Portal

The screenshot shows the ExPASy SIB Bioinformatics Resource Portal homepage. At the top left, there are logos for SIB and ExPASy. A search bar is located at the top center with a dropdown menu set to "Query all databases". On the left side, there is a vertical navigation menu with categories like "Visual Guidance", "Categories", "Resources A..Z", and "Links/Documentation". The main content area features a "Featuring today" section with a slide about the "Ping pong algorithm". On the right side, there are sections for "Popular resources" (UniProtKB, SWISS-MODEL, STRING, PROSITE) and "Latest News" (Protein Spotlight: A walk on the rough side, UniProt Knowledgebase release 2017_02).

Visual Guidance

Categories

- proteomics
- genomics
- structure analysis
- systems biology
- evolutionary biology
- population genetics
- transcriptomics
- biophysics
- imaging
- IT infrastructure
- medicinal chemistry
- glycomics

Resources A..Z

Links/Documentation

ExPASy is the **SIB Bioinformatics Resource Portal** which provides access to scientific databases and software tools (i.e., *resources*) in different areas of life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. (see **Categories** in the left menu). On this portal you find resources from many different SIB groups as well as external institutions.

Featuring today

Ping pong algorithm

The Ping-Pong Algorithm is an efficient method for finding coherent patterns across paired data sets. [\[details\]](#)

Popular resources

- UniProtKB
- SWISS-MODEL
- STRING
- PROSITE

Latest News

Protein Spotlight: A walk on the rough side - 2017-03-01

Life can be hard. There are times when you find yourself in the most unfriendly circumstances and, more often than not, the best way to deal with the situation is to find your own solution and wriggle your own way out... [More.](#)

UniProt Knowledgebase release 2017_02 - 2017-02-15

Release notes
553,655 UniProtKB/Swiss-Prot entries ([More..](#))
77,483,538 UniProtKB/TrEMBL entries ([More..](#))

In Categories

proteomics

protein structure

In Tools selezionare un metodo di predizione:

✓ GOR

✓ HHPred

✓ Jpred

Jpred 4

Incorporating Jnet

A Protein Secondary Structure Prediction Server

[Home](#) [REST API](#) [About](#) [News](#) [F.A.Q.](#) [Help & Tutorials](#) [Monitoring](#) [Contact](#) [Publications](#)

Input sequence^(?)

```
MQVWPIEGIKKFETLSYLPPLTVEDLLKQIEYLLRSKWVPCLEFSKVGfVYRENHRSPGYDGRYWTMWKLPfMGCTDAT  
QVLKELEEAKKAYPDAFVRIIGFDNVRQVQLISFIAYKPPGC
```

[Advanced options \(click to show/hide\)](#)

[Make Prediction](#)

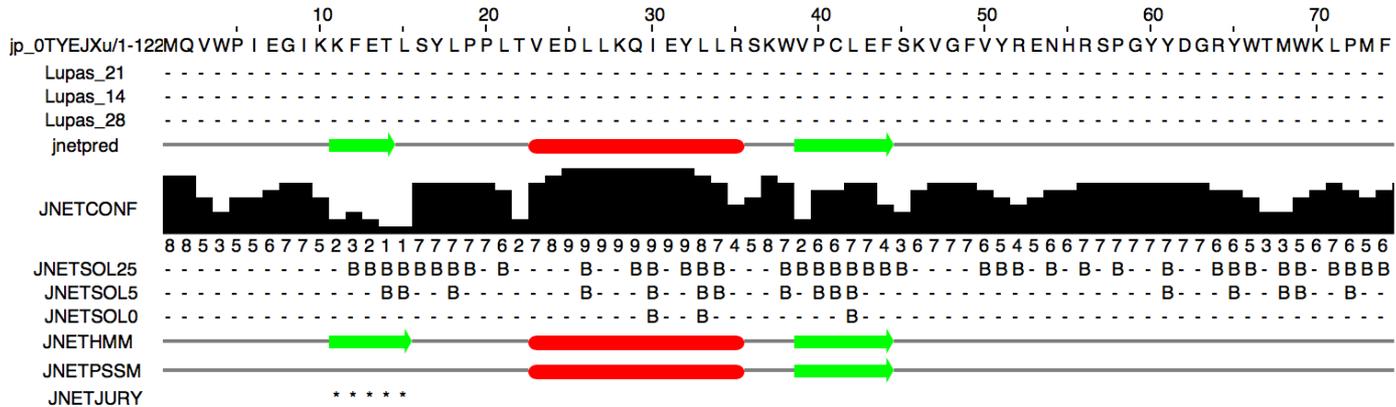
[Reset Form](#)

Primary citation: Drozdetskiy A, Cole C, Procter J & Barton GJ. Nucl. Acids Res.
(first published online April 16, 2015) doi: 10.1093/nar/gkv332 [\[link\]](#)
More citations: [link](#).

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



```
AI L V G S M P R R D G M E R K D L L K A N V K I F K C Q G A  
V E V C P V D C F Y E G P N F L V I H P D E C I D C A L C E P  
G A C K P E C P V N I I Q G S - - Y A I D A D S C I D C G S  
C - - I A C G A C K P E C P V N I I Q G S - - I Y A I D A D S
```

Struttura proteine nella Protein Data Bank

RCSB PDB Deposit Search Visualize Analyze Download Learn More

MyPDB Login

RCSB PDB PROTEIN DATA BANK

An Information Portal to
127823 Biological
Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands

Go

Advanced Search | Browse by Annotations

RCSB PDB-101

WORLDWIDE PDB PROTEIN DATA BANK

EMDataBank

NUCLEIC ACID DATABASE

StructuralBiology Knowledgebase

Worldwide Protein Data Bank Foundation



Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

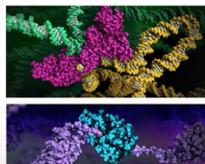
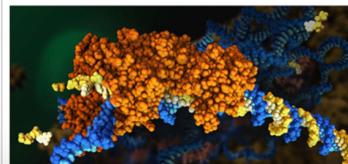
A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

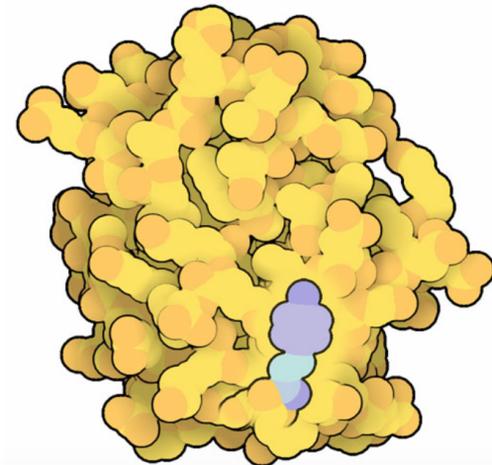
A Molecular View of HIV Therapy



2016
FASEB
BioArt
Winner

View animation

March Molecule of the Month



Contact Us

PDB <http://www.rcsb.org/> Processamento dati nella PDB

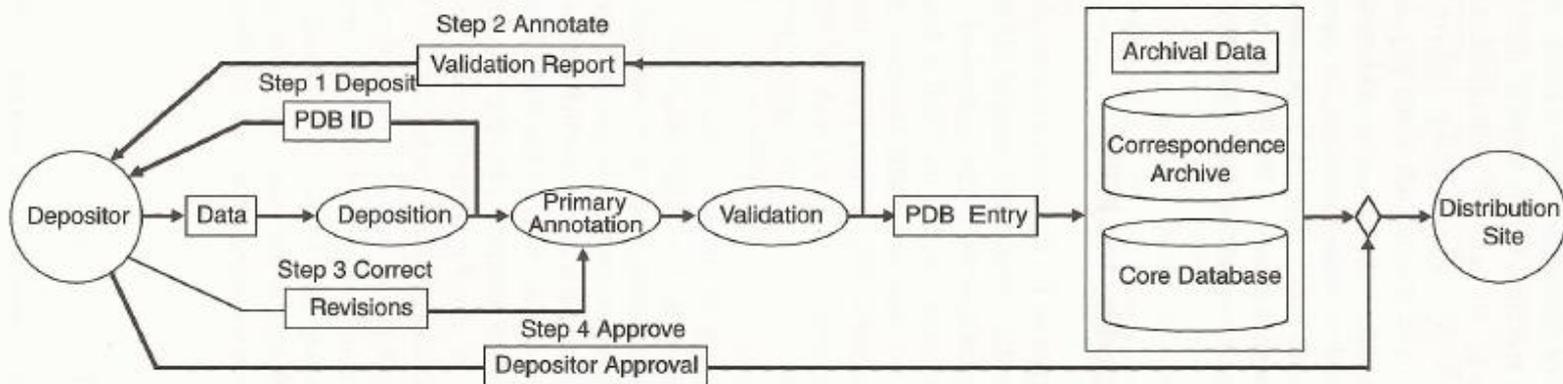
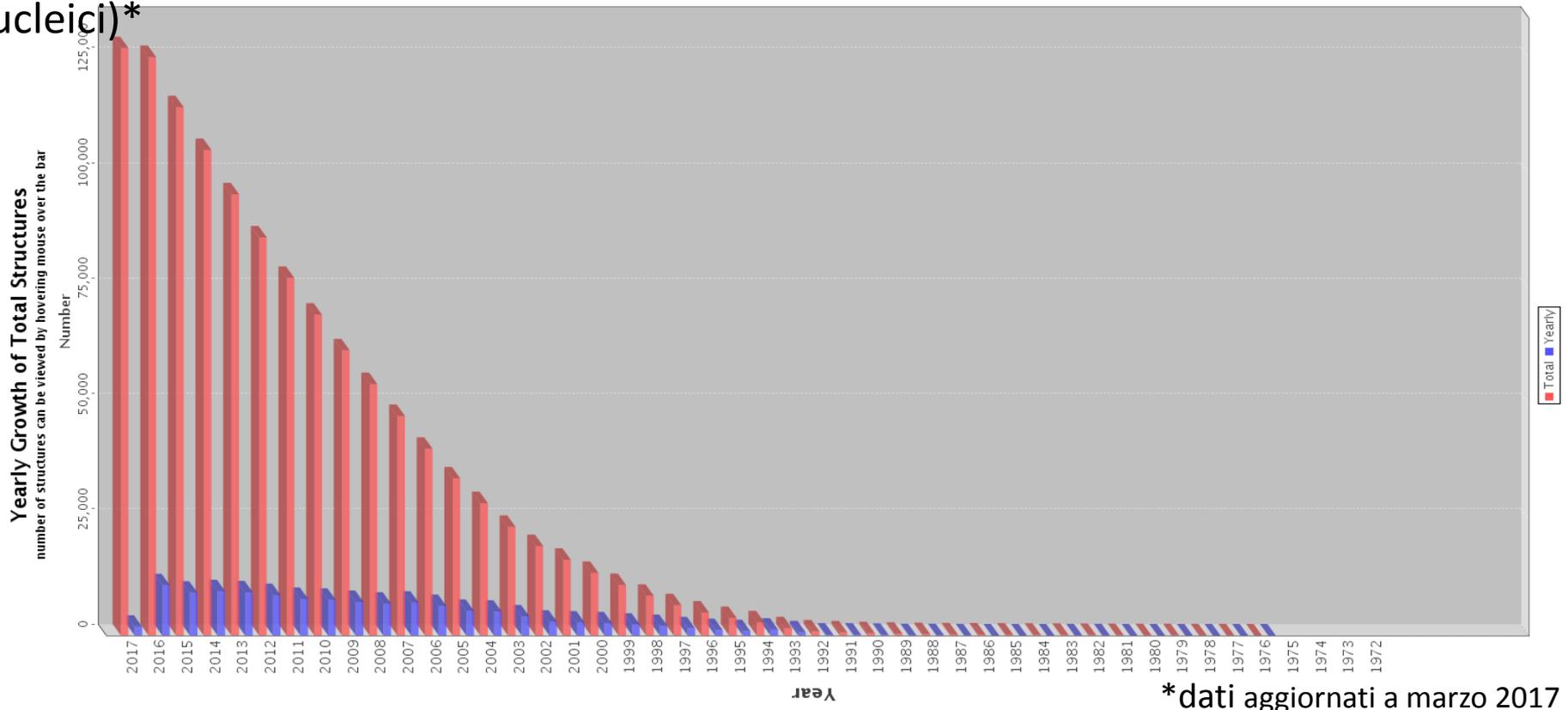


Figure 9.1. The steps involved in PDB data processing. Ellipses represent actions and rectangles define content. Figure reprinted from Berman et al. (2000b) with permission from the International Union of Crystallography.

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)*



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

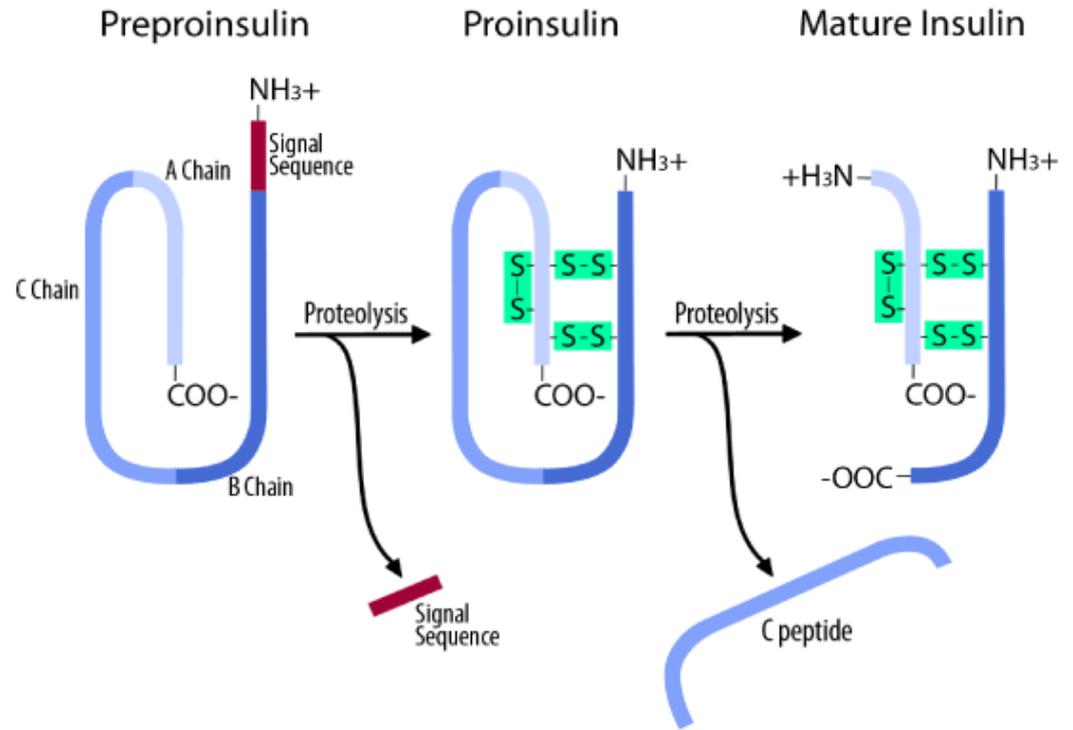
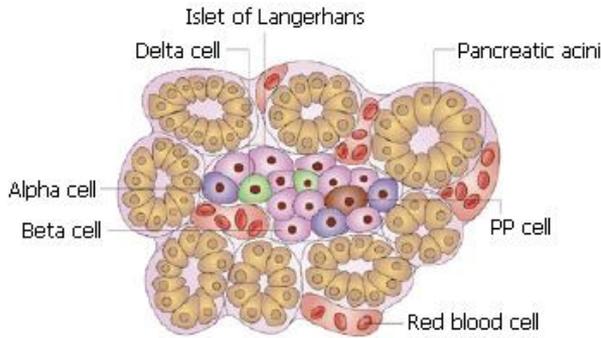
Dati aggiuntivi struttura X-Ray

- 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 : insulina

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

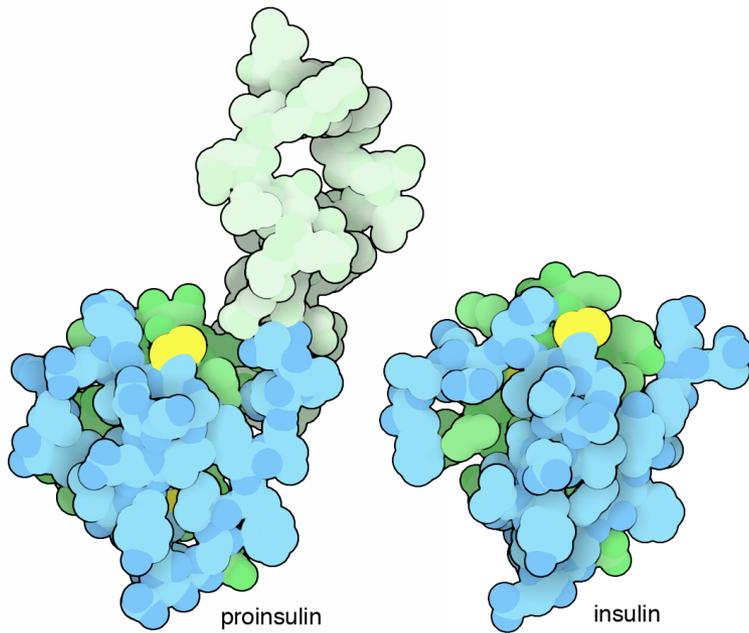
Catena B

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

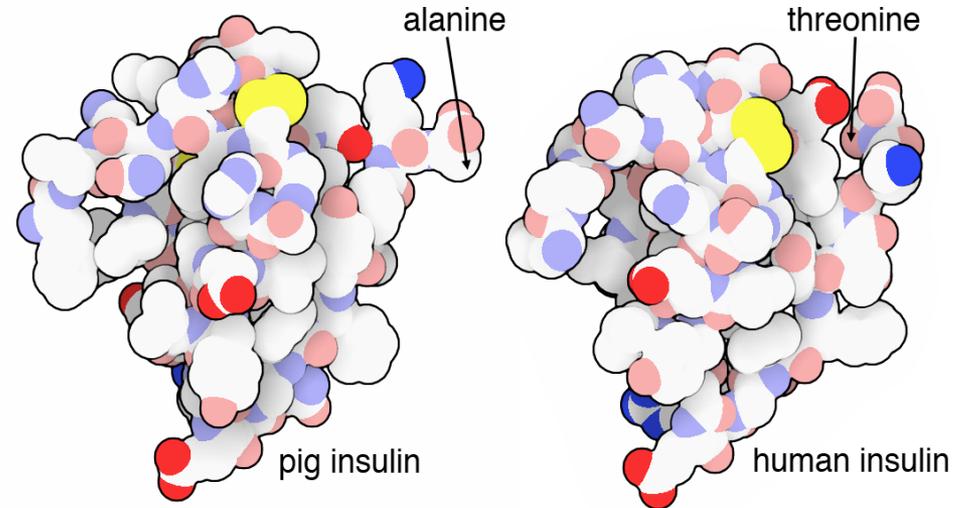
Catena A

Catena B

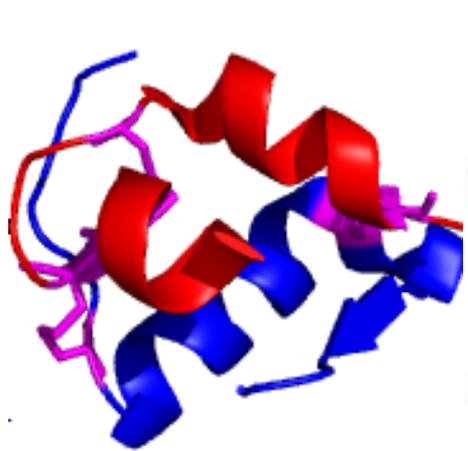


4INS = insulina di maiale

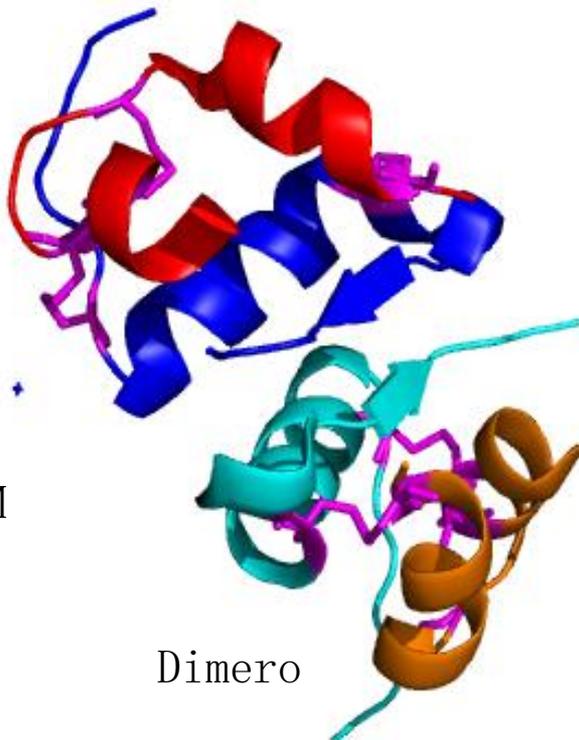
2HIU = insulina umana



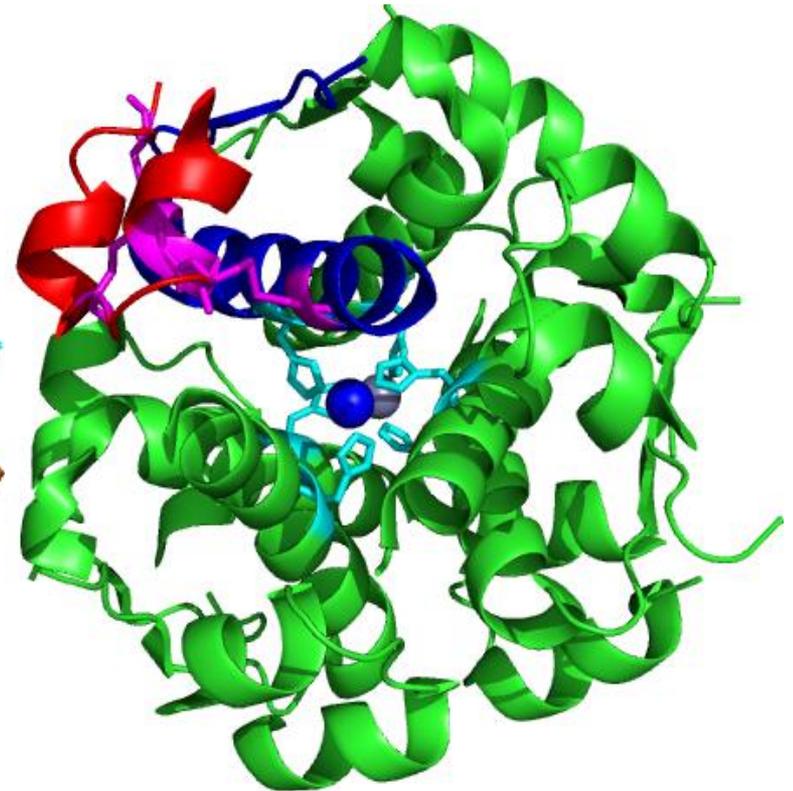
Insulina di maiale (1 aa) e insulina di mucca (3 aa) per la loro somiglianza con quella umana possono essere utilizzate in terapia in quanto riconosciute dalle nostre cellule



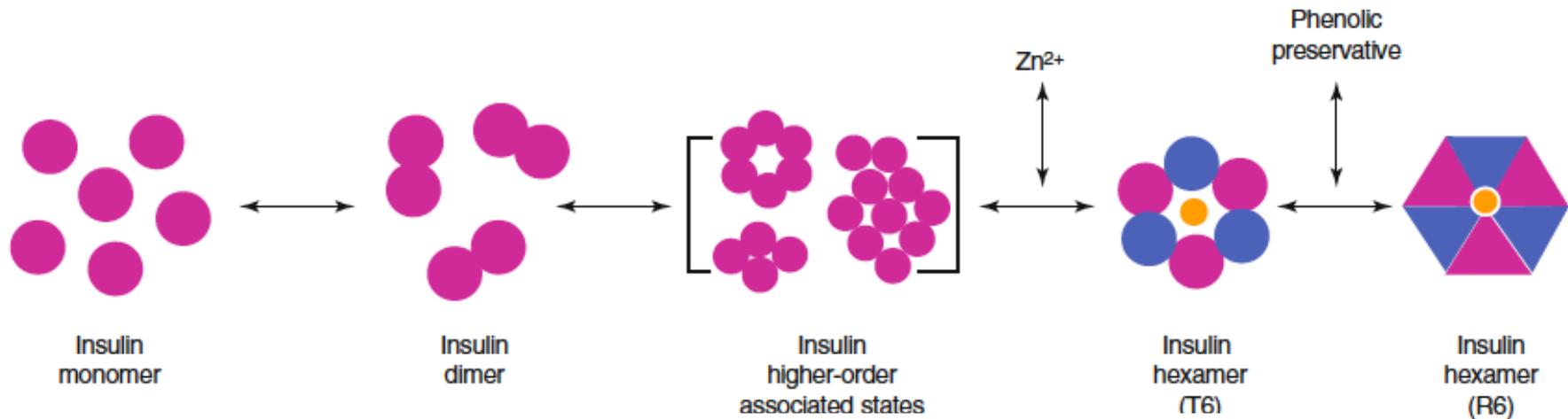
Monomero $< 0.1 \mu\text{M}$



Dimero



Esamero $> 0.2 \text{ mM}$ o con Zn^{2+}



Insulin concentration

10⁻³ M

10⁻³ M

10⁻⁵ M

10⁻⁵ M

Dopo somministrazione

Formulation



Capillary membrane



2HIU

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

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

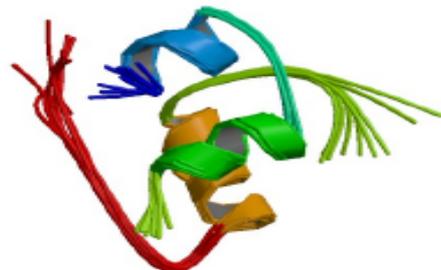
Classification: **HORMONE**

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)



Macromolecule Content

- Unique protein chains: 2

Experimental Data Snapshot

Method: SOLUTION NMR

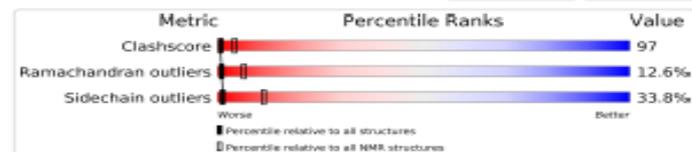
Conformers Calculated: --

Conformers Submitted: 10

Selection Criteria:

wwPDB Validation

[3D Report](#) [Full Report](#)



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

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
Hgncid : HGNC:6081	Previous Names: "Insulin-dependent diabetes mellitus 2"
Refseq: NM_000207	Omlm: 176730
	GenBank: X70508

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

Genome Assembly [GRCh37](#)



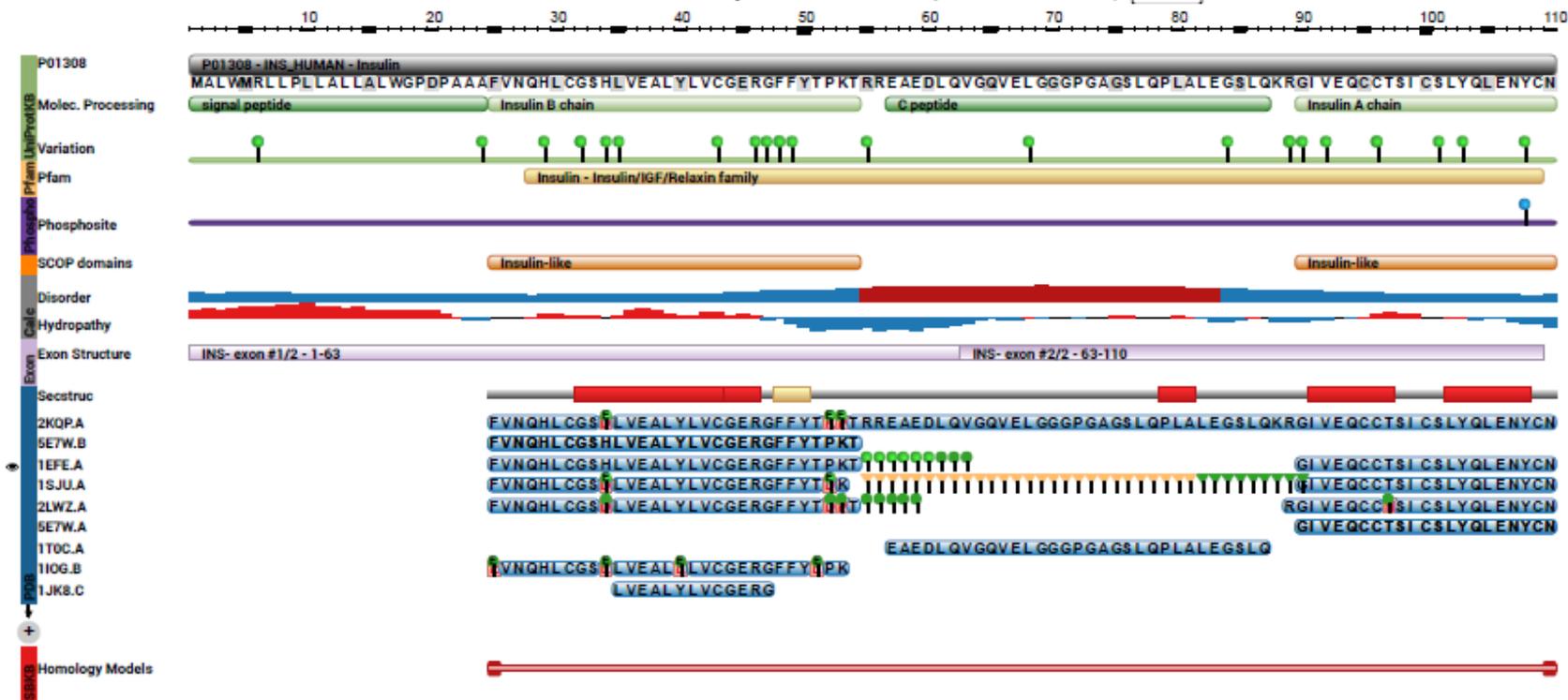
Insulin - P01308 (INS_HUMAN)

Protein Feature View of PDB entries mapped to a UniProtKB sequence

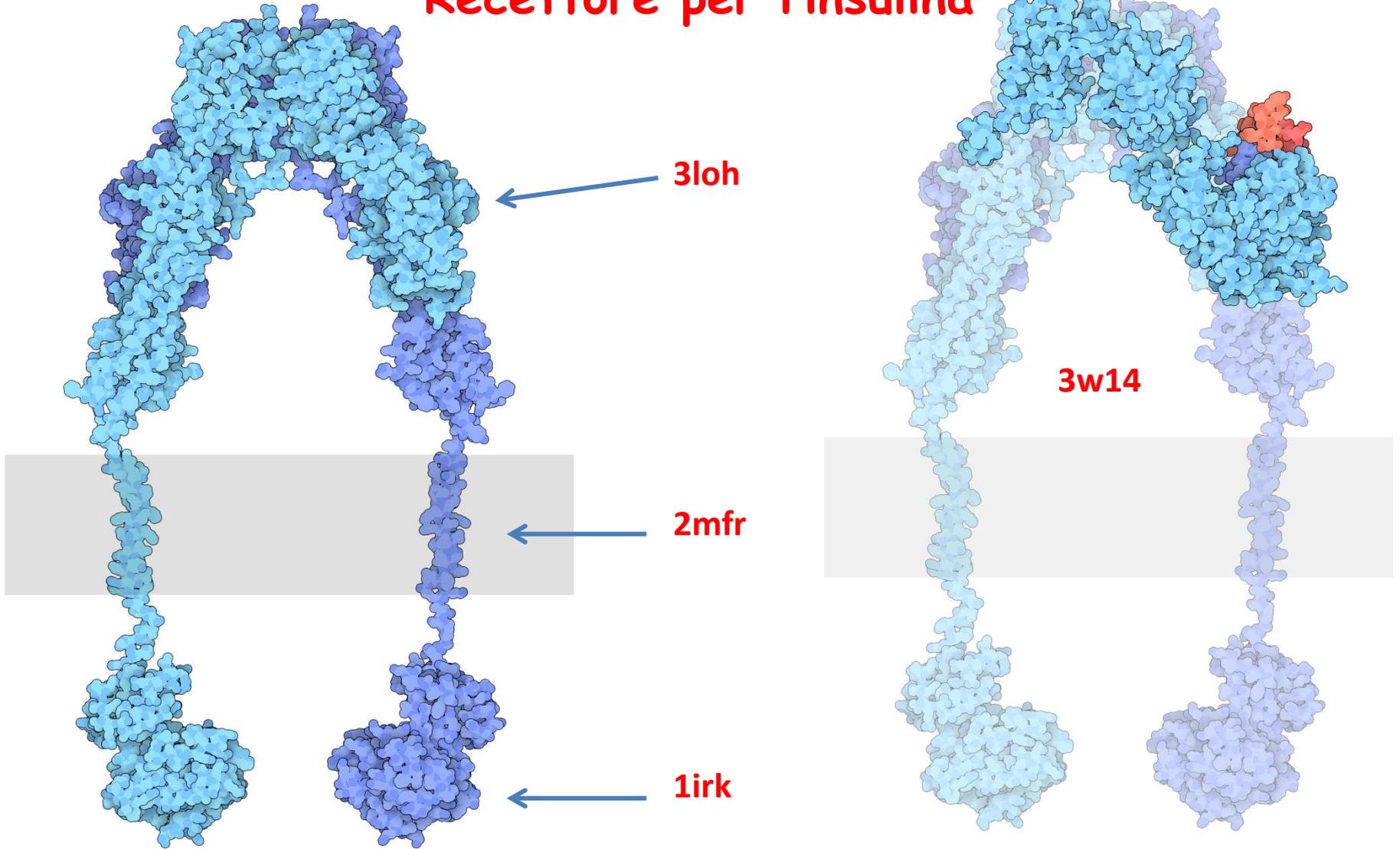
Number of PDB entries for P01308: 240

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](#)



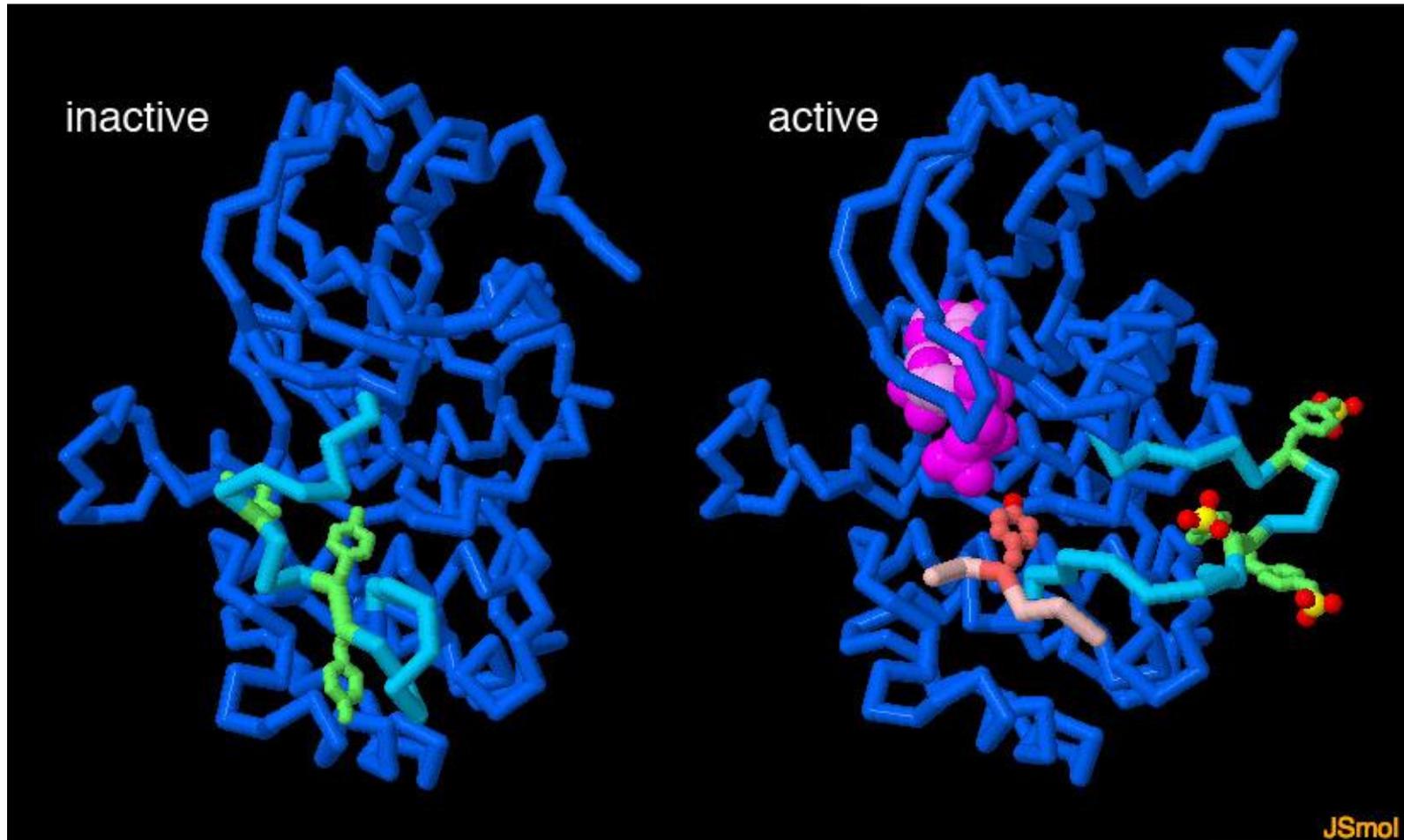
Recettore per l'insulina



Dominio tirosin chinasi del recettore nella forma inattiva ed attiva

1irk

1ir3



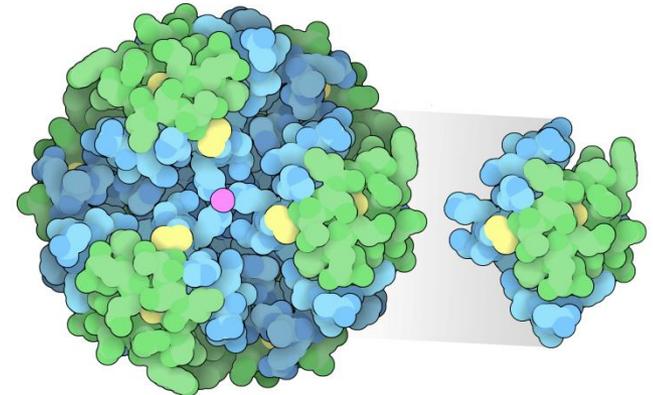
Insuline "progettate"

- ✓ 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

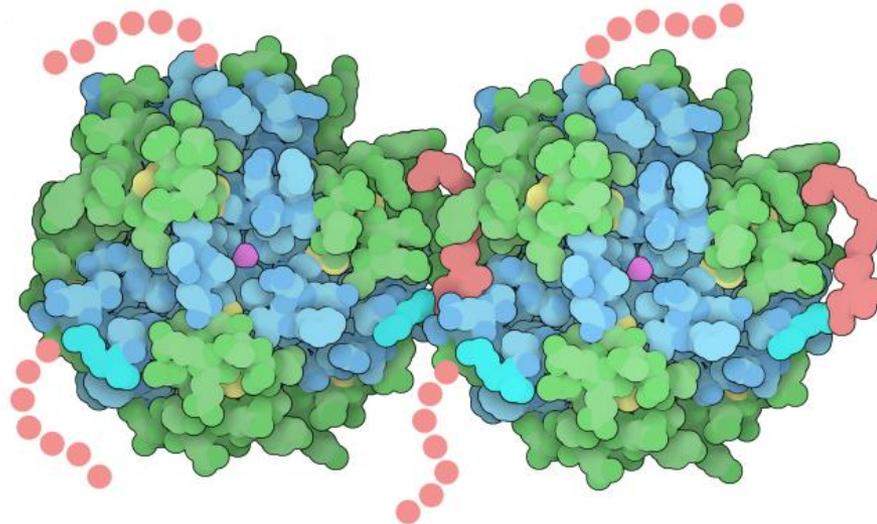
GLARGINE

AsnA21Gly + ArgB31ArgB32

Azione lenta

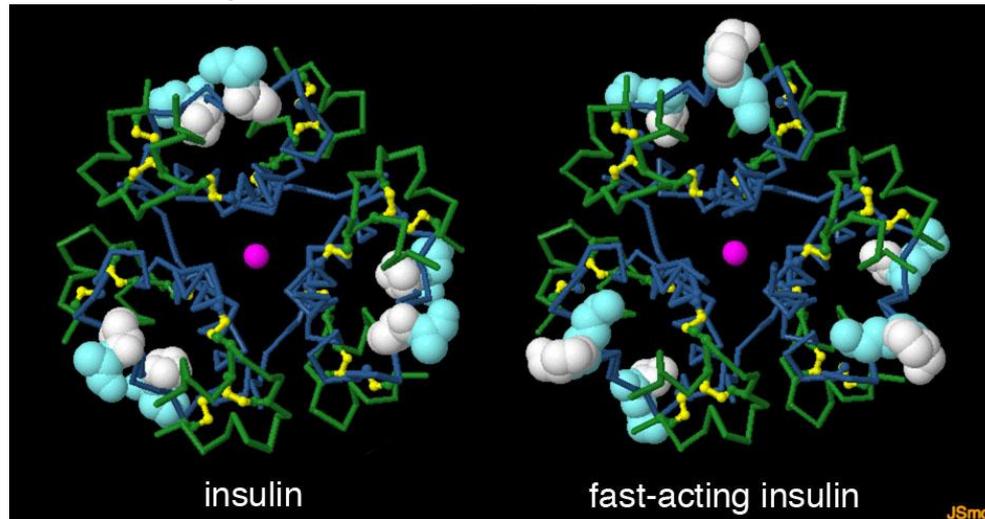
Insuline di lunga durata

- ✓ 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



Diabete mellito

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

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



<http://www.who.int/diabetes/en>

<https://www.diapedia.org>

A screenshot of the Diapedia website, titled "The Living Textbook of Diabetes". The page features a dark blue header with the Diapedia logo and a search bar. Below the header, there are navigation links for "HOME", "NEWS", and "ABOUT", along with "Get started" and "Diapedia app" buttons. The main content area is divided into two columns: "Sections" and "Editor's corner". The "Sections" column lists eight sections with corresponding images and dates: "Introduction to diabetes mellitus" (24 Aug 2014), "Type 1 diabetes mellitus" (13 Nov 2014), "Type 2 diabetes mellitus" (1 Aug 2013), "Other types of diabetes mellitus" (15 May 2013), "Metabolism, insulin and other hormones" (18 Jan 2016), "Associated disorders" (25 Apr 2014), "Acute and chronic complications of diabetes" (1 Sep 2014), and "Management" (7 Jan 2014). The "Editor's corner" section features an article titled "Sugary drinks are such sweet sorrow" with a sub-headline "A study in the last issue of Diabetologia contributes further evidence to the harms of consuming too much sugar sweetened drinks." The article text discusses the EPIC-Norfolk study and the impact of sugar intake on diabetes risk. A "Continue reading" link is provided at the bottom of the article. A vertical "Feedback about Diapedia" button is located on the left side of the page.



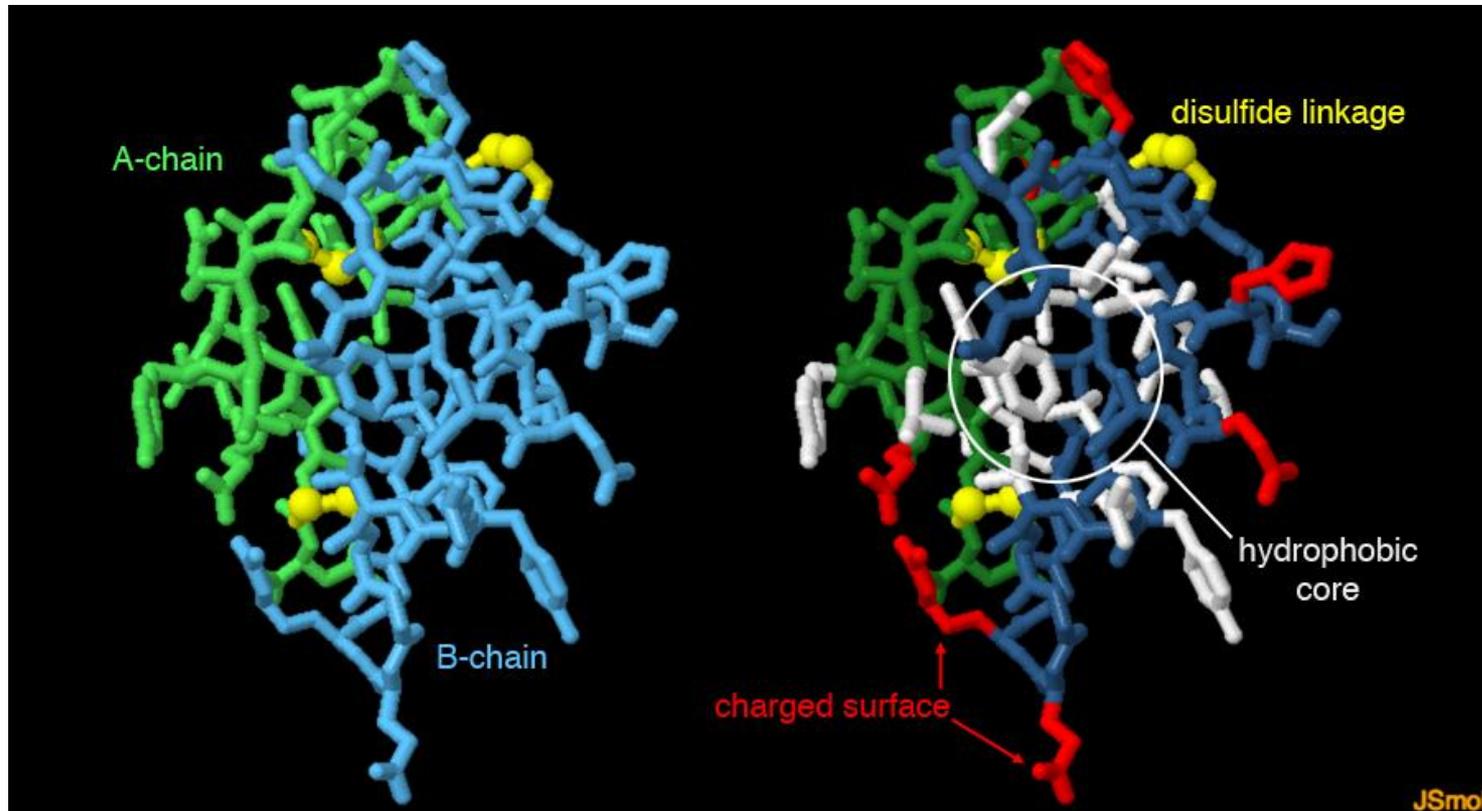
Nel portale educational della PDB si può trovare materiale didattico utile per esercitazioni sulle varie proteine.

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)

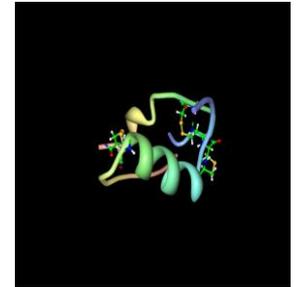
ESERCITAZIONE PRATICA

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