

# Rappresentazione e utilizzo di modelli cristallografici

Laurea Magistrale in Biotecnologie Mediche  
Curriculum Nanobiotecnologie

A.A. 2021-22

---

# Uso delle coordinate in contesti differenti

I modelli molecolari ottenuti per via sperimentale sono utilizzati essenzialmente per stabilire una **relazione tra la funzione della macromolecola e la sua struttura.**

Il termine funzione è quanto mai ampio, assume significati anche molto diversi (attività catalitica, funzioni recettoriali, interazione con piccole molecole...)

Inoltre, le coordinate ottenute sperimentalmente possono trovare molti usi:

- Per il modelling di proteine omologhe
- Per studi di natura teorica a partire da un modello sperimentale (Dinamica Molecolare)
- Per studi di Docking
- Per studi di Virtual Screening (Structure Based)

In tutti questi casi, la disponibilità di coordinate sperimentali è indispensabile!

# Accesso ai dati strutturali di macromolecole

Il risultato del processo di determinazione strutturale per via cristallografica, è costituito dall'insieme di coordinate e fattori termici di tutti gli atomi contenuti nell'unità asimmetrica.

Alle coordinate degli atomi è associato l'insieme dei fattori di struttura da cui è stato possibile costruire il modello molecolare.

A partire dalle coordinate degli atomi e dai fattori di struttura è possibile ricostruire non solo il modello molecolare, ma anche verificare quanto il modello molecolare costruito è coerente con i dati di diffrazione sperimentalmente determinato.

Con il progredire e il consolidarsi della cristallografia di macromolecole come metodo di validità generale per la determinazione strutturale, il numero di strutture proteiche disponibili aumentò sensibilmente.

# Protein Data Bank

# Protein Data Bank (PDB)

Fu presto evidente che era necessario un archivio delle strutture proteiche determinate, per questo motivo nel 1971 nacque il **Protein (Structure) Data Bank** noto come **PDB** originariamente ospitato presso i laboratori di Brookhaven.



## Celebrating 50 Years of the Protein Data Bank Archive

In 1971, the structural biology community established the single worldwide archive for macromolecular structure data—the Protein Data Bank (PDB). From its inception, the PDB has embraced a culture of open access, leading to its widespread use by the research community. PDB data are used by hundreds of data resources and millions of users exploring fundamental biology, energy, and biomedicine.

In 2021, RCSB PDB and the wwPDB celebrates the 50th Anniversary of the PDB with symposia, materials, and more.

# Cosa è il Protein Data Bank

Con gli anni, l'aumentare del numero delle strutture molecolari determinate e con la maggiore disponibilità di mezzi informatici, il PDB è diventato qualcosa di più di un semplice archivio.

Al giorno d'oggi il PDB assolve diversi compiti:

- **Archivio delle strutture di macromolecole biologiche** determinate sperimentalmente (coordinate, dati sperimentali)
- **Organismo di validazione delle strutture** determinate sperimentalmente
- **Strumento di analisi delle strutture** macromolecolari

Visto l'elevato numero di strutture macromolecolari depositate, **il PDB è dotato di un potente motore di ricerca.**



*<http://www.rcsb.org/>*

# PDB – Deposizione e Validazione

# PDB – Deposizione dei dati



## wwPDB OneDep System

Existing deposition


Deposition ID

Password

[Log in](#)

[Forgot Password](#)

**PROTEIN DATA BANK**



PDB50: Celebrating 50 Years of the PDB archive  
Visit [wwpdb.org/pdb50](http://wwpdb.org/pdb50) for related symposia and resources

Al termine del processo di ottimizzazione della struttura, a seguito della sua validazione il modello generato viene depositato presso il **P**rotein structure **D**ata **B**ank (PDB), insieme ai fattori di struttura (dati sperimentali) che sono alla base della sua genesi.

**Affinché una struttura macromolecolare (biologica) sia pubblicata e discussa in una rivista scientifica, la struttura deve essere stata depositata presso il PDB.**



# PDB – Quali dati vengono depositati

A seconda del metodo sperimentale utilizzato (raggi-X, NMR, cryo-EM) vengono depositati dati diversi.

Per una struttura cristallografica (elenco non esaustivo):

- **Coordinate e fattori termici di TUTTI gli atomi presenti nel modello cristallografico ottimizzato**
- **Fattori di struttura UTILIZZATI per la costruzione e ottimizzazione del modello**
- Condizioni di Cristallizzazione
- Dati Cristallografici (Gruppo spaziale, Cella Unitaria)
- Condizioni Sperimentali dell'esperimento di Diffrazione (lunghezza d'onda, beamline)
- Risultati e indicatori di qualità della riduzione dei dati (Risoluzione, R-merge, completezza...)
- Metodo usato per la risoluzione della struttura (MIR, MAD, MR...)
- Risultati della procedura di refinement (R-work, R-free...)
- Informazioni biochimiche sulla macromolecola (funzione della macromolecola)
- Organismo di origine
- Sistema di Espressione
- Sequenza della proteina (se proteina)

*Esiste un periodo di 'quarantena' di 1 anno dal momento della deposizione, dopodiché le coordinate, i fattori di struttura e i dati depositati sono accessibili a tutti.*

# PDB – Deposizione e Validazione

Nel corso della procedura di Deposizione dei dati, **il PDB esegue automaticamente una ‘validazione’ indipendente del modello e dei fattori di struttura** che si intende depositare.

The screenshot shows the wwPDB validation system interface. On the left, there is a login section with a 'Validation ID' input field, a 'Password' input field, a 'Log in' button, and a 'Forgot Password' button. Below these is a 'Deposition server' button. On the right, there is a welcome message: 'Welcome to the wwPDB validation system!'. Below the message, there are instructions: 'This server runs the performs the same validation as you would observe during the deposition process. This service is designed to help you check your model and experimental files prior to start of deposition. To continue with an existing validation session, please login on the left. To start a new validation, please complete the form below. Upon completion, you will be emailed login information specific to your new validation.' Below the instructions, there are two input fields: 'Your e-mail address' and 'Password (optional, or we will provide one)'. The password field has a note: 'This is a shared "group password" (6 to 16 alphanumeric characters)'. There are information icons (i) next to the 'Validation ID', 'Password', 'Your e-mail address', and 'Password' fields.

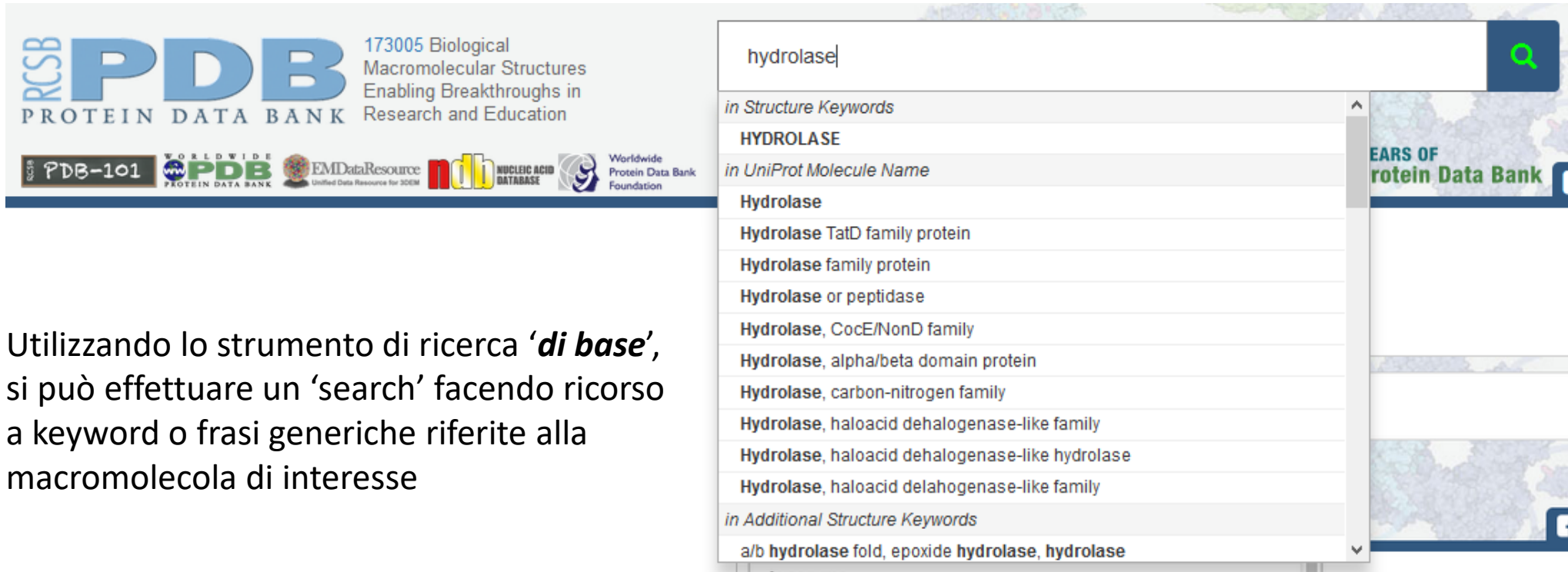
I **‘ligandi’** presenti nella struttura, ovvero tutte quelle componenti chimiche di natura **NON-macromolecolare** (ioni e molecole inorganiche e organiche eccetto l’acqua), sono confrontate con un database dei ligandi già depositati.

Se si tratta di un nuovo ligando vengono richiesti dati chimici e strutturali e il nuovo ligando è aggiunto al database con un suo codice identificativo di 3 caratteri alfanumerici.

# PDB – Ricerca di una struttura

# PDB – Search: Basic Search

Il PDB offre la possibilità di cercare nel database delle strutture (contenente ad oggi circa 170000 strutture depositate).



The screenshot displays the PDB search interface. At the top left, the PDB logo is accompanied by the text "173005 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education". Below the logo are several partner logos including PDB-101, Worldwide PDB, EMDatabase, ndb, and Worldwide Protein Data Bank Foundation. The search bar contains the text "hydrolase" and a green search icon. A dropdown menu is open, showing search results categorized into three sections: "in Structure Keywords" with the result "HYDROLASE"; "in UniProt Molecule Name" with results including "Hydrolase", "Hydrolase TatD family protein", "Hydrolase family protein", "Hydrolase or peptidase", "Hydrolase, CocE/NonD family", "Hydrolase, alpha/beta domain protein", "Hydrolase, carbon-nitrogen family", "Hydrolase, haloacid dehalogenase-like family", "Hydrolase, haloacid dehalogenase-like hydrolase", and "Hydrolase, haloacid delahogenase-like family"; and "in Additional Structure Keywords" with the result "a/b hydrolase fold, epoxide hydrolase, hydrolase".

Utilizzando lo strumento di ricerca **'di base'**, si può effettuare un 'search' facendo ricorso a keyword o frasi generiche riferite alla macromolecola di interesse

# PDB – Search: Accesso per argomenti

Una possibilità alternativa per cercare la/le struttura/e di interesse consiste nell'accedere al database sulla base di **keywords** (argomenti)

## Browse by Annotation

PDB entries have been annotated by various ontologies and hierarchical classification schemes.

- [Anatomical Therapeutic Chemical Classification System](#)
- [Biological Process](#)
- [CATH](#)
- [Cellular Component](#)
- [Enzyme Classification](#)
- [Genome Location](#)
- [Membrane Protein](#)
- [MeSH \(Medical Subject Headings\)](#)
- [Molecular Function](#)
- [SCOP](#)
- [Protein Symmetry](#)
- [Source Organism](#)

Start Browsing

# PDB – Search: Advanced Search

La ricerca può essere anche di una certa complessità.

**Il PDB è dotato di un motore di ricerca piuttosto sofisticato.**

Si può eseguire una ricerca sofisticata sulla base di diversi argomenti.

Advanced Search Query Builder

Attribute ?

-- Enter and/or select a field name --

- Full Text
- ID(s) and Keywords
- Deposition
- Chemical Components
- Deposited Entry Features
- Polymer Molecular Features
- Polymer Instance (Chain) Features
- Nonpolymer Molecular Features
- Oligosaccharide/Branched Molecular Features
- Assembly Features
- Biologically Interesting Molecules (BIRD)
- Methods
- X-ray Method Details
- X-ray Data Collection Details
- Cell Dimensions and Space Group
- Crystal Properties
- EM Method Details
- NMR Method Details
- Linked External Resources
- Publications Primary
- Publications All

Sequence ?

Sequence Motif ?

Structure Similarity ?

Chemical ?

Add Group

About About Us Help

wwPDB

RCSB PDB is funded by the National Science Foundation and the National Institutes of Health

# PDB – Search: Advanced Search

Advanced Search Query Builder

Attribute

Drug Target Name x contains phrase Add + NOT Count ✕

Add Field Add Subgroup Remove Group

Add Group

Sequence

AND MTTQAPTFTQPLQSVVLEG. Enter a sequence containing a minimum of 20 residues, OR enter a PDB ID in the text box below and select from the sequence list. The second option is useful for finding sequences that are similar to a sequence from a given structure and chain. ⋮

PDB ID 1MBN Target Protein E-Value Cutoff 0.1 Identity Cutoff 0 % (Integer only) Count Clear

Sequence Motif

AND MQTIF

Target Protein Mode Simple Count Clear

Structure Similarity

AND PDB ID 1MBN Count Clear

Chemical

AND C12 H28 N4 O. Note that a Chemical Formula Search is case-sensitive. For example: "NIC4" (Nitrogen, Iodine, Carbon) and "NiC4" (Nickel, Carbon), will yield different results. ⋮

Query Type Formula Match Subset Count Clear

Display Results as Structures Count Clear 🔍

# PDB – Risultati di una ricerca

Il risultato di una ricerca è costituito da una lista di strutture, più o meno attinenti agli argomenti dei campi usati per la ricerca.

Il risultato della ricerca può essere ulteriormente ridotto, sulla base di ulteriori limitazioni (es: organismo che produce la proteina).

The screenshot shows the PDB search results for the query "hydroxysteroid dehydrogenase". The page includes the PDB logo, search bar, and navigation options. A red box highlights the "Refinements" section, which allows filtering results by organism. The "Refinements" section is currently empty, and a red arrow points to it from the text on the left. The main results area shows a list of structures, with the first entry highlighted: "3GMD: Structure-based design of 7-Azaindole-pyrrolidines as inhibitors of 11beta-Hydroxysteroid-Dehydrogenase type I". The entry includes a 3D ribbon diagram of the protein structure, a "3D View" button, and a table of metadata including release date, method, organism, macromolecule, and unique ligands.

**Refinements** Clear All

SCIENTIFIC NAME OF SOURCE ORGANISM Clear

- Homo sapiens (50623)
- Mus musculus (7214)
- Escherichia coli (6191)
- synthetic construct (4733)
- Escherichia coli K-12 (3615)
- Rattus norvegicus (3249)
- Bos taurus (3085)
- Saccharomyces cerevisiae (2822)
- Saccharomyces cerevisiae S288C (1963)
- Gallus gallus (1889)

[More...](#)

TAXONOMY Clear

**3GMD** Download File View File

Structure-based design of 7-Azaindole-pyrrolidines as inhibitors of 11beta-Hydroxysteroid-Dehydrogenase type I

Valeur, E., Lepifre, F., Roche, D., Christmann-Franck, S., Hillertz, P., Musil, D.

(2012) Bioorg Med Chem Lett **22**: 5909-5914

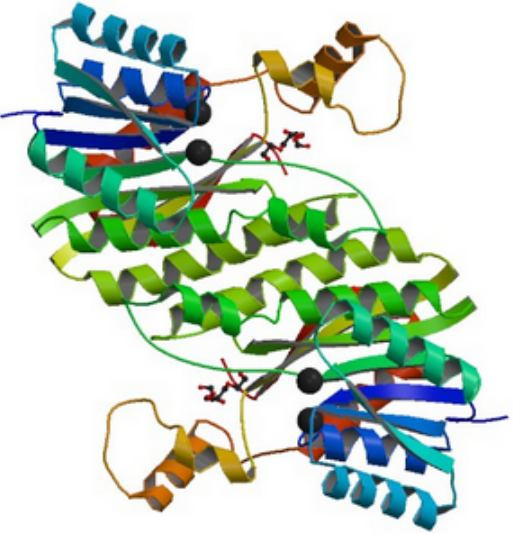
<b>Released</b>	2010-11-17
<b>Method</b>	X-RAY DIFFRACTION 2.28 Å
<b>Organisms</b>	Mus musculus
<b>Macromolecule</b>	Corticosteroid 11-beta-dehydrogenase isozyme 1 (protein)
<b>Unique Ligands</b>	2M3, NDP, SO4



# PDB – Coordinate e fattori di struttura

# PDB – Accesso ai dati

Biological Assembly 1 ?



3D View: Structure | Electron Density | Ligand Interaction

Standalone Viewers  
Protein Workshop | Ligand Explorer

## 3IS3

Crystal structure of 17beta-Hydroxysteroid dehydrogenase (Cochliobolus lunatus)

DOI: [10.2210/pdb3IS3/pdb](https://doi.org/10.2210/pdb3IS3/pdb)

Classification: [OXIDOREDUCTASE](#)

Organism(s): [Cochliobolus lunatus](#)

Expression System: [Escherichia coli](#)

Deposited: 2009-08-25 Released: 2010-08-18

Deposition Author(s): [Cassetta, A.](#), [Lamba, D.](#), [Krastanova, I.](#)

Experimental Data Snapshot

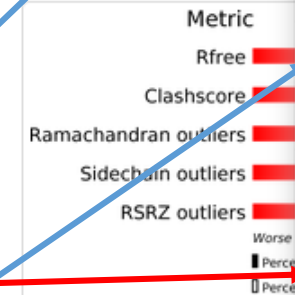
Method: X-RAY DIFFRACTION

Resolution: 1.48 Å

R-Value Free: 0.188

R-Value Work: 0.175

wwPDB Validation



Coordinate e fattori di struttura

Display Files ▾ Download Files ▾

- FASTA Sequence
- PDB Format
- PDB Format (gz)
- PDBx/mmCIF Format
- PDBx/mmCIF Format (gz)
- PDBML/XML Format (gz)
- Biological Assembly 1
- Biological Assembly 2
- Structure Factors (CIF)
- Structure Factors (CIF - gz)
- 2fo-fc Map (DSN6)
- fo-fc Map (DSN6)
- Map Coefficients (MTZ format)

# Il file PDB

Il file PDB (come comunemente chiamato) è un file di testo che **contiene coordinate, fattori termici e alcune informazioni di carattere generale sulla natura della struttura cristallografica a cui si riferisce**.

Alcune informazioni sono comunque reperibili in forma più semplice e immediata direttamente dal sito web del PDB.

Contiene inoltre dati specifici su:

- Acquisizione e qualità dei dati (**R-merge**)
- Procedura di affinamento
- Qualità dell'affinamento (fattori **R-work e R-free**)
- Qualità del modello (**outliers** da validazione)

```
REMARK 2
REMARK 2 RESOLUTION.      1.48 ANGSTROMS.
REMARK 3
REMARK 3 REFINEMENT.
REMARK 3   PROGRAM       : REFMAC 5.5.0094
REMARK 3   AUTHORS      : MURSHUDOV, VAGIN, DODSON
REMARK 3
REMARK 3   REFINEMENT TARGET : MAXIMUM LIKELIHOOD
REMARK 3
REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.48
REMARK 3 RESOLUTION RANGE LOW  (ANGSTROMS) : 23.86
REMARK 3 DATA CUTOFF              (SIGMA(F)) : 0.000
REMARK 3 COMPLETENESS FOR RANGE          (%) : 98.4
REMARK 3 NUMBER OF REFLECTIONS          : 51846
REMARK 3
REMARK 3 FIT TO DATA USED IN REFINEMENT.
REMARK 3 CROSS-VALIDATION METHOD           : THROUGHOUT
REMARK 3 FREE R VALUE TEST SET SELECTION : RANDOM
REMARK 3 R VALUE          (WORKING + TEST SET) : 0.176
REMARK 3 R VALUE          (WORKING SET)       : 0.175
REMARK 3 FREE R VALUE              : 0.188
REMARK 3 FREE R VALUE TEST SET SIZE  (%) : 10.100
REMARK 3 FREE R VALUE TEST SET COUNT   : 5215
REMARK 3
```

# Il file PDB

Il file PDB contiene le informazioni cristallografiche:

- Cella unitaria
- Gruppo Spaziale
- Origine

Inoltre contiene le coordinate dei diversi atomi (atomo e residuo)

E il relativo **fattore termico**

**Isotropo** [sfera]

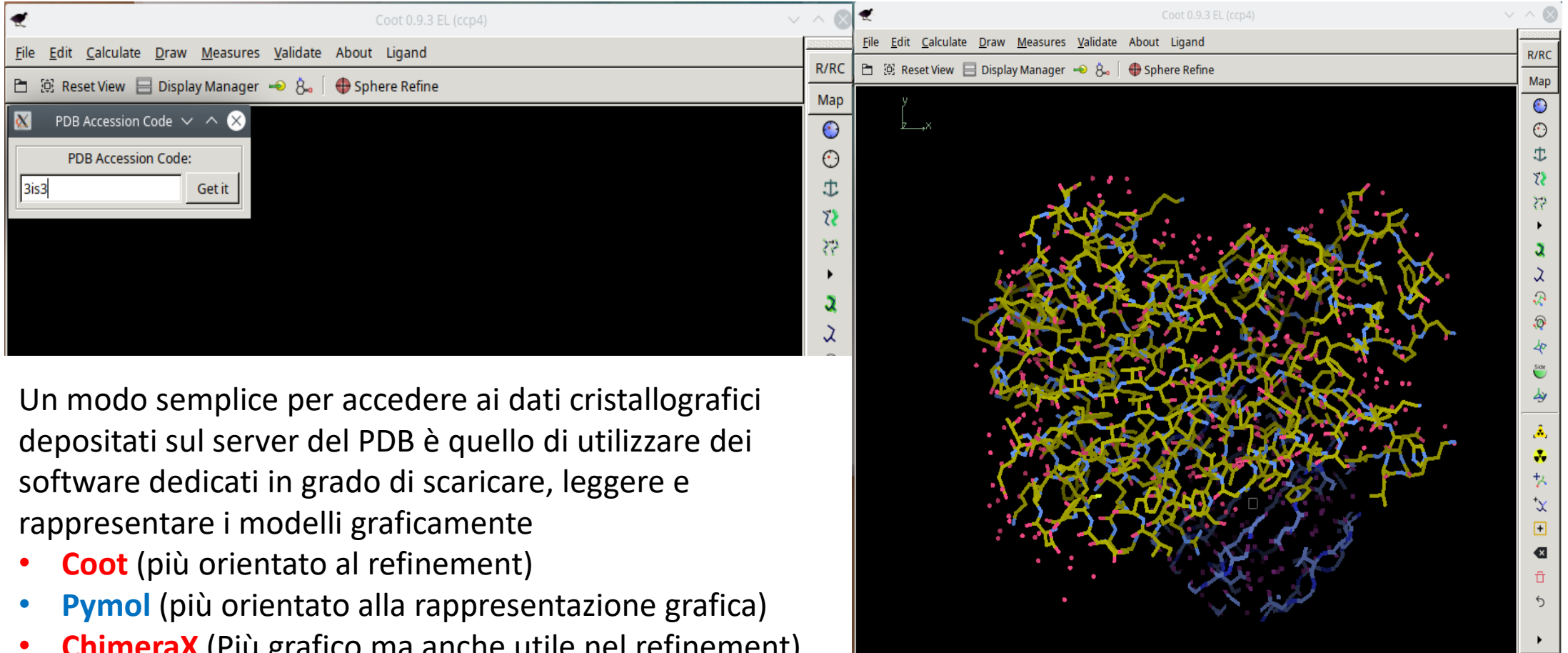
**Anisotropo** [ellissoide]

	Cella unitaria						Gruppo spaziale					
CRYST1	67.914	67.914	268.347	90.00	90.00	90.00	I	41	2	2	16	
ORIGX1	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000						
ORIGX2	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000						
ORIGX3	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000						
SCALE1	0.014725	0.000000	0.000000	0.000000	0.000000	0.000000						
SCALE2	0.000000	0.014725	0.000000	0.000000	0.000000	0.000000						
SCALE3	0.000000	0.000000	0.003727	0.000000	0.000000	0.000000						
ATOM	1	N	ATYR	A	11	-33.371	19.479	2.962	0.50	34.97		N
ANISOU	1	N	ATYR	A	11	3446	5787	4055	2107	-766	-466	N
ATOM	2	N	BTYR	A	11	-28.685	23.311	-4.764	0.50	40.27		N
ANISOU	2	N	BTYR	A	11	4968	6177	4156	2620	-869	213	N
ATOM	3	CA	ATYR	A	11	-31.953	19.022	3.088	0.50	33.26		C
ANISOU	3	CA	ATYR	A	11	3343	5436	3858	1960	-719	-430	C
ATOM	4	CA	BTYR	A	11	-27.982	22.196	-4.052	0.50	38.00		C
ANISOU	4	CA	BTYR	A	11	4639	5854	3947	2411	-828	136	C
ATOM	5	C	ATYR	A	11	-31.012	19.779	2.148	0.50	33.38		C
ANISOU	5	C	ATYR	A	11	3518	5357	3808	2022	-730	-334	C

I fattori termici sono raramente utilizzati, anche se di estrema importanza per comprendere la mobilità dei vari aminoacidi.

**Le coordinate sono la parte più usata del file PDB.**

# Accedere ai file nel PDB



Un modo semplice per accedere ai dati cristallografici depositati sul server del PDB è quello di utilizzare dei software dedicati in grado di scaricare, leggere e rappresentare i modelli graficamente

- **Coot** (più orientato al refinement)
- **Pymol** (più orientato alla rappresentazione grafica)
- **ChimeraX** (Più grafico ma anche utile nel refinement)
- **Rasmol** (molto semplice ma funziona sempre!)

# Accedere alle mappe di densità

The image displays two windows from the Coot 0.9.3 EL (ccp4) software. The left window, titled "PDB Validation Report for 3IS3", provides a detailed analysis of the structure's quality. It includes a table of metrics with percentile ranks and values, and a list of outliers with their respective scores.

Metric	Percentile Ranks	Value
Rfree		0.21
Clashscore		2.24
Ramachandran Outliers		0.39
Sidechain Outliers		0.00
RSRZ Outliers		5.00

Legend:  
■ Percentile relative to all x-ray structures  
□ Percentile relative to x-ray structures of similar resolution

EDS Resolution: 1.48

- Residue A 11 TYR:  
Bad RSRZ 5.582
- Residue A 11 TYR:  
Clash atom O score: 0.53  
Clash atom O score: 0.42  
Clash atom C score: 1.04  
Clash atom C score: 0.53  
Bad RSRZ 5.582
- Residue A 12 ILE:  
Bad RSRZ 4.070

The right window shows a 3D density map of the protein structure, with atoms represented as sticks and density as a blue mesh. The interface includes a menu bar (File, Edit, Calculate, Draw, Measures, Validate, About, Ligand) and a toolbar with various tools like "Reset View", "Display Manager", and "Sphere Refine".

Tramite gli stessi software è anche possibile accedere alle mappe di densità e ai dati di Validazione

# PDB – Informazioni Presenti

# La singola struttura nel PDB



The screenshot displays the PDB entry page for 3IS3. On the left, a 3D ribbon diagram of the protein structure is shown. The right side contains the following information:

- 3IS3**
- Crystal structure of 17beta-Hydroxysteroid dehydrogenase (Apo form) from fungus *Cochliobolus lunatus*
- DOI: [10.2210/pdb3IS3/pdb](https://doi.org/10.2210/pdb3IS3/pdb)
- Classification: [OXIDOREDUCTASE](#)
- Organism(s): [Cochliobolus lunatus](#)
- Expression System: [Escherichia coli](#)
- Deposited: 2009-08-25 Released: 2010-08-18
- Deposition Author(s): [Cassetta, A.](#), [Lamba, D.](#), [Krastanova, I.](#)

**Experimental Data Snapshot**

- Method: X-RAY DIFFRACTION
- Resolution: 1.48 Å
- R-Value Free: 0.188
- R-Value Work: 0.175

**wwPDB Validation**

Metric	Percentile Ranks	Value
Rfree		0.211
Clashscore		2
Ramachandran outliers		0.4%
Sidechain outliers		0
RSRZ outliers		5.0%

Legend:   
■ Percentile relative to all X-ray structures   
□ Percentile relative to X-ray structures of similar resolution

Nel PDB, ogni singola struttura è **identificata da un codice alfanumerico (A-Z, 0-9) a 4 cifre**. La singola struttura, in questo caso con codice 3IS3, si presenta come un sommario del contenuto completo della 'scheda', per la struttura in esame (Nome e funzione della proteina, organismo di appartenenza, proprietà cristallografiche, qualità della struttura).



# PDB – Informazioni: Riferimenti Bibliografici

Tra le informazioni accessibili nella scheda di una struttura presente nel PDB, vi sono anche i riferimenti bibliografici dello studio in cui la struttura (in questo caso 3QWF) è stata presentata e discussa.

Literature Download Primary Citation ▾

Structural Studies on a Fungal 17Beta-Hydroxysteroid Dehydrogenase  
[Cassetta, A.](#), [Krastanova, I.](#), [Kristan, K.](#), [Brunskole Svegelj, M.](#), [Lamba, D.](#), [Lanisnik Rizner, T.](#), [Stojan, J.](#)  
(2012) *Biochem J* **441**: 151-160  
PubMed: [21929506](#) Search on PubMed  
DOI: [10.1042/BJ20110567](#)  
Primary Citation of Related Structures:  
[3QWF](#), [3QWI](#)

**PubMed Abstract:**  
The 17β-HSD (17β-hydroxysteroid dehydrogenase) from the filamentous fungus *Cochliobolus lunatus* (17β-HSDcl) is a NADP(H)-dependent enzyme that preferentially catalyses the interconversion of inactive 17-oxosteroids and their active 17β-hydroxy counterparts. 17β-HSDcl belongs to the SDR (short-chain dehydrogenase/reductase) superfamily. It is currently the only fungal 17β-HSD member that has been described and represents one of the model enzymes of the cP1 classical subfamily of NADPH-dependent SDR enzymes. A thorough crystallographic analysis has been performed to better understand the structural aspects of this subfamily and provide insights into the evolution of the HSD enzymes. The crystal structures of the 17β-HSDcl apo, holo and coumestrol-inhibited ternary complex, and the active-site Y167F mutant reveal subtle conformational differences in the substrate-binding loop that probably modulate the catalytic activity of 17β-HSDcl. Coumestrol, a plant-derived non-steroidal compound with oestrogenic activity, inhibits 17β-HSDcl [IC50 2.8 μM; at 100 μM substrate (4-oestrene-3,17-dione)] by occupying the putative steroid-binding site. In addition to an extensive hydrogen-bonding network, coumestrol binding is stabilized further by π-π stacking interactions with Tyr212. A stopped-flow kinetic experiment clearly showed the coenzyme dissociation as the slowest step of the reaction and, in addition to the low steroid solubility, it prevents the accumulation of enzyme-coenzyme-steroid ternary complexes.

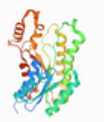
**Organizational Affiliation:**  
Istituto di Cristallografia - UOS Trieste, Consiglio Nazionale delle Ricerche, Area Science Park-Basovizza, S.S. 14 km 163.5, I-34149 Trieste, Italy. [alberto.cassetta@ic.cnr.it](mailto:alberto.cassetta@ic.cnr.it)

# PDB - informazioni: Sommario

Macromolecules

Find similar proteins by: [Sequence](#) (by identity cutoff) | [Structure](#)

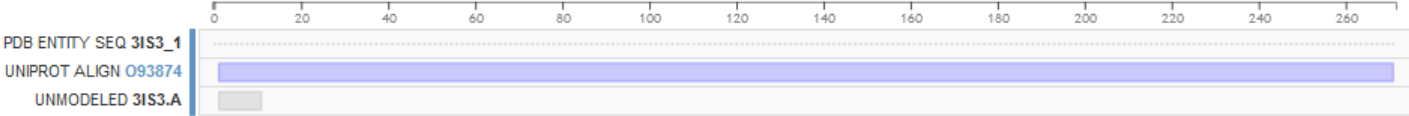
**Entity ID: 1**

Molecule	Chains	Sequence Length	Organism	Details	Image
17beta-hydroxysteroid dehydrogenase	A	270	<a href="#">Curvularia lunata</a>	Mutation(s): 0 Gene Names: <a href="#">17HSDcl</a> EC: <a href="#">1.1.1.62</a>	

Find proteins for [O93874](#) (*Cochliobolus lunatus*)      Explore [O93874](#)      Go to UniProtKB: [O93874](#)

**Protein Feature View** [Expand](#)

Reference Sequence



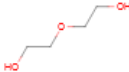
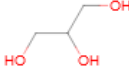
PDB ENTITY SEQ 3IS3\_1  
UNIPROT ALIGN O93874  
UNMODELED 3IS3.A

Sono Presenti informazioni generali sulla proteina: N° di aminoacidi, organismo di provenienza, proprietà biochimiche, link ad ulteriori database (UniProtKB, ...)

# PDB – Informazioni: Ligandi

Sono presenti informazioni sulla tipologia dei ligandi eventualmente presenti nella struttura.

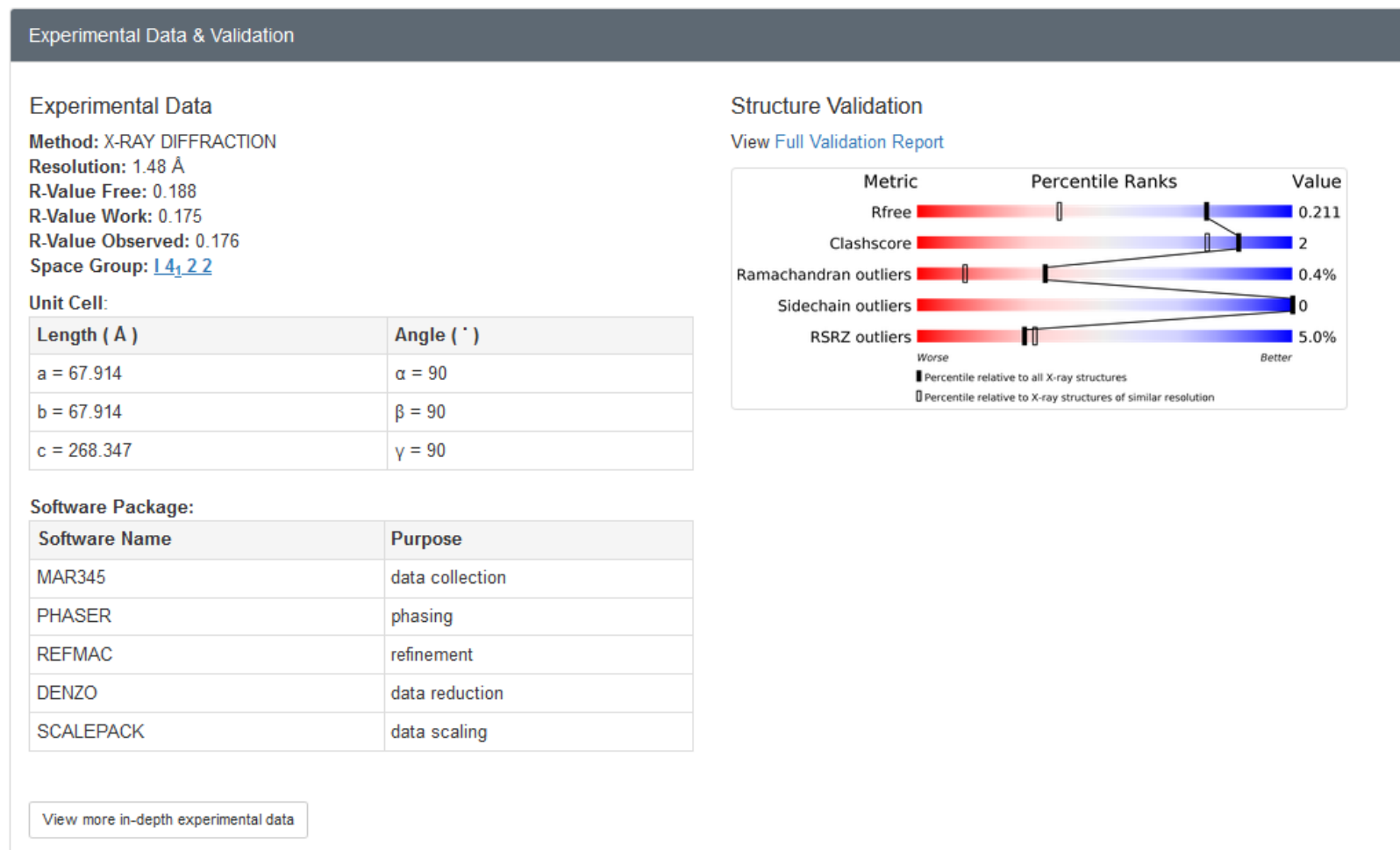
[E' considerato ligando qualsiasi molecola organica o inorganica che non sia parte della proteina, dell'oligonucleotide o che non sia acqua]

Small Molecules				
Ligands 3 Unique				
ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
<a href="#">PEG</a> <a href="#">Query on PEG</a>  <a href="#">Download Ideal Coordinates CCD File</a> <a href="#">Download Instance Coordinates</a>	A	DI(HYDROXYETHYL)ETHER C <sub>4</sub> H <sub>10</sub> O <sub>3</sub> MTHSVFCYNBDYFN-UHFFFAOYSA-N		<a href="#">Ligand Interaction</a>
<a href="#">GOL</a> <a href="#">Query on GOL</a>  <a href="#">Download Ideal Coordinates CCD File</a> <a href="#">Download Instance Coordinates</a>	A	GLYCEROL C <sub>3</sub> H <sub>8</sub> O <sub>3</sub> PEDCQBHIVMGVHV-UHFFFAOYSA-N		<a href="#">Ligand Interaction</a>
<a href="#">CL</a> <a href="#">Query on CL</a>  <a href="#">Download Ideal Coordinates CCD File</a> <a href="#">Download Instance Coordinates</a>	A	CHLORIDE ION Cl VEXZGXHMUGYJMC-UHFFFAOYSA-M	Cl <sup>-</sup>	<a href="#">Ligand Interaction</a>

# PDB - informazioni: Sommario dei dati cristallografici

E' presente un sommario dei:

- Dati Cristallografici
- Indicatori di qualità (R/Rfree)
- Esito della Validazione
- Software utilizzato



# PDB - Informazioni: Condizioni sperimentali

3IS3

Display Files - Download Files -

Crystal structure of 17beta-Hydroxysteroid dehydrogenase (Apo form) from fungus Cochliobolus lunatus

## X-RAY DIFFRACTION

### Crystallization

Crystallization Experiments				
ID	Method	pH	Temperature	Details
1	VAPOR DIFFUSION, HANGING DROP	8	277	20% (W/V) PEG 6000, 20% (V/V) Glycerol, 0.1M Tris, pH 8.0, VAPOR DIFFUSION, HANGING DROP, temperature 277K

Crystal Properties	
Matthews coefficient	Solvent content
2.67	53.99

### Diffraction

Diffraction Experiment									
ID #	Crystal ID	Scattering Type	Data Collection Temperature	Detector	Detector Type	Details	Collection Date	Monochromator	Protocol
1	1	x-ray	100	IMAGE PLATE	MAR scanner 345 mm plate	Platinum Coated three segment toroidal mirror	2005-07-20	M	SINGLE WAVELENGTH

Radiation Source					
ID #	Source	Type	Wavelength List	Synchrotron Site	Beamline
1	SYNCHROTRON	ELETTRA BEAMLINE 5.2R	1.2	ELETTRA	5.2R

Per ogni singola struttura sono riportate:

- Condizioni di cristallizzazione
- Strumentazione utilizzata (**beamline**) per l'acquisizione dei dati
- Temperatura
- Lunghezza d'onda
- Detector

# PDB – Informazioni: Acquisizione dei dati

## Data Collection

Overall												
ID #	Resolution (High)	Resolution (Low)	Percent Possible (Observed)	R Merge I (Observed)	R Sym I (Observed)	Net I Over Average Sigma (I)	Redundancy	Number Reflections (All)	Number Reflections (Observed)	Observed Criterion Sigma (F)	Observed Criterion Sigma (I)	B (Isotropic) From Wilson Plot
1	1.48	28.8	99.6	0.068	0.068	19.2	13.5	52140	52140		-3	20.8

Highest Resolution Shell									
ID #	Resolution (High)	Resolution (Low)	Percent Possible (All)	Percent Possible (Observed)	R Merge I (Observed)	R-Sym I (Observed)	Mean I Over Sigma (Observed)	Redundancy	Number Unique Reflections (All)
	1.48	1.51	98.4		0.494	0.494	4.1	6.4	1894

Sono riportate le informazioni sull'acquisizione dei dati di diffrazione: Risoluzione, Numero di Fattori di struttura, Ridondanza, Completezza,  $\langle I/\sigma(I) \rangle$ , R-merge

# PDB – informazioni: Risoluzione e ottimizzazione

## Refinement

Statistics															
Diffraction ID	Structure Solution Method	Cross Validation method	Starting model	Resolution (High)	Resolution (Low)	Number Reflections (All)	Number Reflections (Observed)	Number Reflections (R-Free)	Percent Reflections (Observed)	R-Factor (All)	R-Factor (Observed)	R-Work	R-Free	R-Free Selection Details	Mean Isotropic B
X-RAY DIFFRACTION	MOLECULAR REPLACEMENT	THROUGHOUT	1G00	1.48	23.86	51846	51846	5215	98.36	0.17633	0.17633	0.17495	0.18828	RANDOM	24.853

Anisotropic B[3][3]
0.01

Sono indicate una serie di informazioni sulla risoluzione del problema della fase (***Structure Solution***) e ottimizzazione (***Refinement***):

- Metodo usato per la risoluzione strutturale (e dettagli)
- Risoluzione e numero di fattori di struttura utilizzati nel corso del refinement
- **R-work**
- **R-free**
- B (isotropo) medio

# PDB – informazioni: qualità del modello

Sono presenti informazioni anche sulla qualità del modello, essenzialmente come deviazioni quadratiche medie (rmsd) di parametri geometrici comunemente utilizzati come restraints nel corso dell'ottimizzazione del modello.

Una descrizione più estesa dell'intero processo di validazione è disponibile nel **'Validation Report'**

RMS Deviations	
Key	Refinement Restraint Deviation
r_dihedral_angle_2_deg	36.394
r_dihedral_angle_3_deg	11.113
r_dihedral_angle_4_deg	9.426
r_dihedral_angle_1_deg	6.041
r_scangle_it	3.529
r_scbond_it	2.227
r_angle_refined_deg	1.422
r_mcangle_it	1.304
r_mcbond_it	0.773
r_chiral_restr	0.104

Show All Keys

Non-Hydrogen Atoms Used in Refinement	
Non-Hydrogen Atoms	Number
Protein Atoms	1964
Nucleic Acid Atoms	
Solvent Atoms	296
Heterogen Atoms	15



# PDB – Report di Validazione

# PDB Validazione

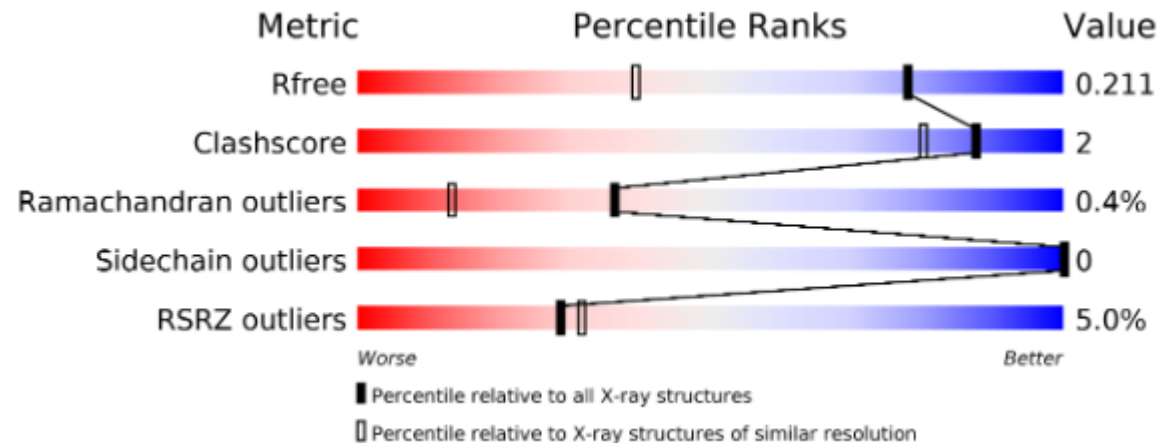


Full wwPDB X-ray Structure Validation Report [i](#)

**Per ogni struttura proteica depositata presso il PDB esiste un report di validazione che ne illustra eventuali criticità.**

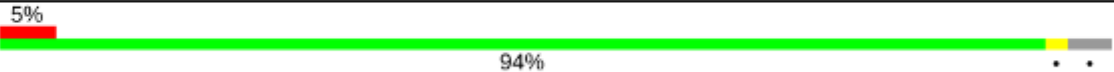
Prima di essere accettata una struttura deve rispondere a dei parametri minimi di qualità.

Per ogni struttura è immediatamente visualizzato un indicatore compatto di qualità:

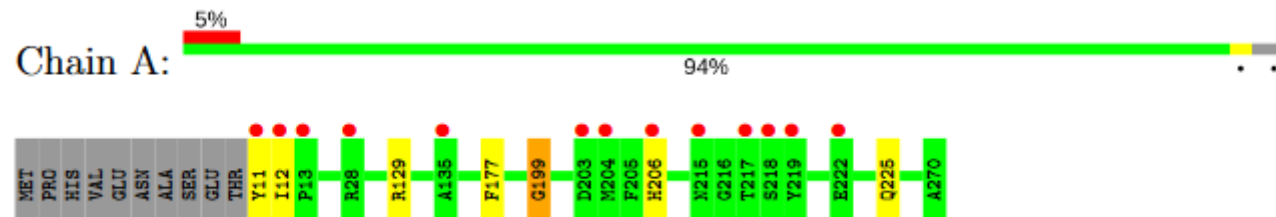


# PDB Validazione

Nel report è anche indicata la qualità della struttura, residuo per residuo

Mol	Chain	Length	Quality of chain
1	A	270	

Sono indicati quali sono i residui 'problematici' indicati con un codice a colori di problematicità crescente (verde=OK poi giallo, arancio, rosso). Un 'pallino' rosso indica una densità elettronica non ottimale.



# PDB Validazione

Nel report sono chiaramente indicati quali sono i parametri geometrici meritevoli di attenzione speciale (**outliers**), in quanto potenzialmente problematici dal punto di vista geometrico (**Z > 5**).

Inoltre sono individuati i contatti corti (collisioni) presenti nel modello.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	177	PHE	CB-CG-CD2	-5.54	116.92	120.80
1	A	129	ARG	NE-CZ-NH2	-5.42	117.59	120.30

All (9) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:11[B]:TYR:C	1:A:12[B]:ILE:HD13	1.77	1.04
1:A:12[B]:ILE:HD13	1:A:12[B]:ILE:N	1.85	0.90
1:A:12[B]:ILE:N	1:A:12[B]:ILE:CD1	2.57	0.66
1:A:11[B]:TYR:C	1:A:12[B]:ILE:CD1	2.65	0.53
1:A:11[B]:TYR:O	1:A:12[B]:ILE:HD13	2.09	0.53
1:A:199:GLY:HA2	3:A:273:PEG:H11	1.91	0.51
3:A:273:PEG:H42	5:A:539:HOH:O	2.18	0.43
1:A:11[B]:TYR:O	1:A:12[B]:ILE:CD1	2.67	0.42
1:A:206:HIS:HE1	1:A:225:GLN:OE1	2.02	0.41

# PDB Validazione

All (13) RSRZ outliers are listed below:

Sono infine riportati quali **residui hanno una bassa correlazione con la densità elettronica osservata.**

Una bassa correlazione tra modello e densità osservata, è in genere è indice di densità elettronica 'debole' ma può essere anche un comportamento 'dinamico' dei residui indicati (flessibilità) piuttosto che ad un vero errore nel modello.

Sono inoltre riportati quei **residui con valori dei fattori termici eccessivamente grandi.**

Anche in questo si può avere un errore nel modello o un comportamento 'dinamico' degli atomi coinvolti.

Mol	Chain	Res	Type	RSRZ
1	A	11[A]	TYR	5.6
1	A	12[A]	ILE	4.1
1	A	203	ASP	3.6
1	A	13[A]	PRO	3.5
1	A	219	TYR	3.4
1	A	218	SER	3.0
1	A	204	MET	3.0
1	A	217	THR	2.8
1	A	206	HIS	2.3
1	A	28	ARG	2.3
1	A	215	ASN	2.3
1	A	222	GLU	2.2
1	A	135	ALA	2.0

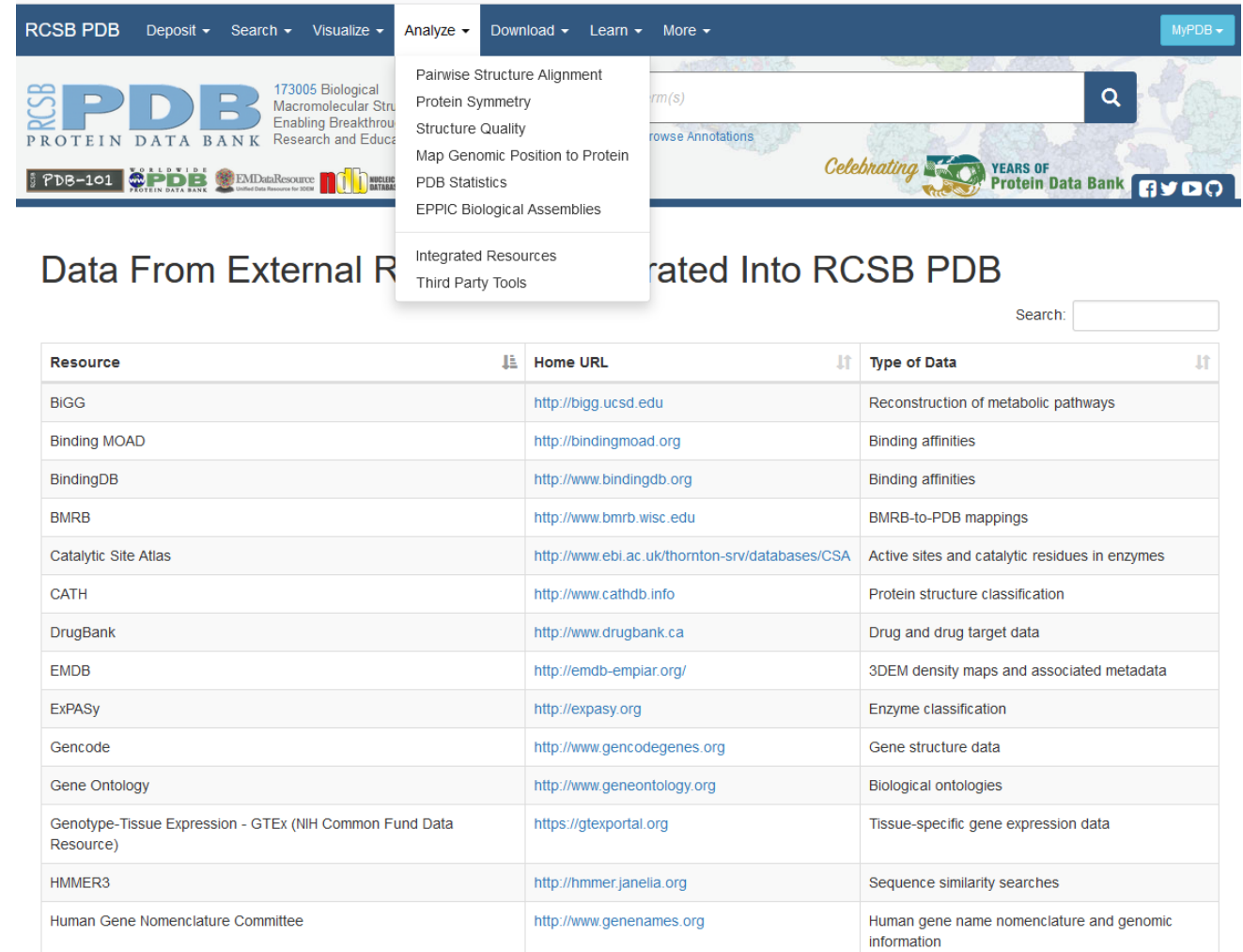
# Rappresentazioni grafiche di una macromolecola biologica

# PDB - Analisi

Il PDB integra alcuni strumenti di analisi della struttura:

- Geometria (distanze, angoli)
- Struttura quaternaria (se presente)
- Allineamento con altre strutture

In genere per l'analisi della struttura si ricorre ad altri software capaci di analisi più complesse (es: **Pymol**, **ChimeraX**, server specifici)



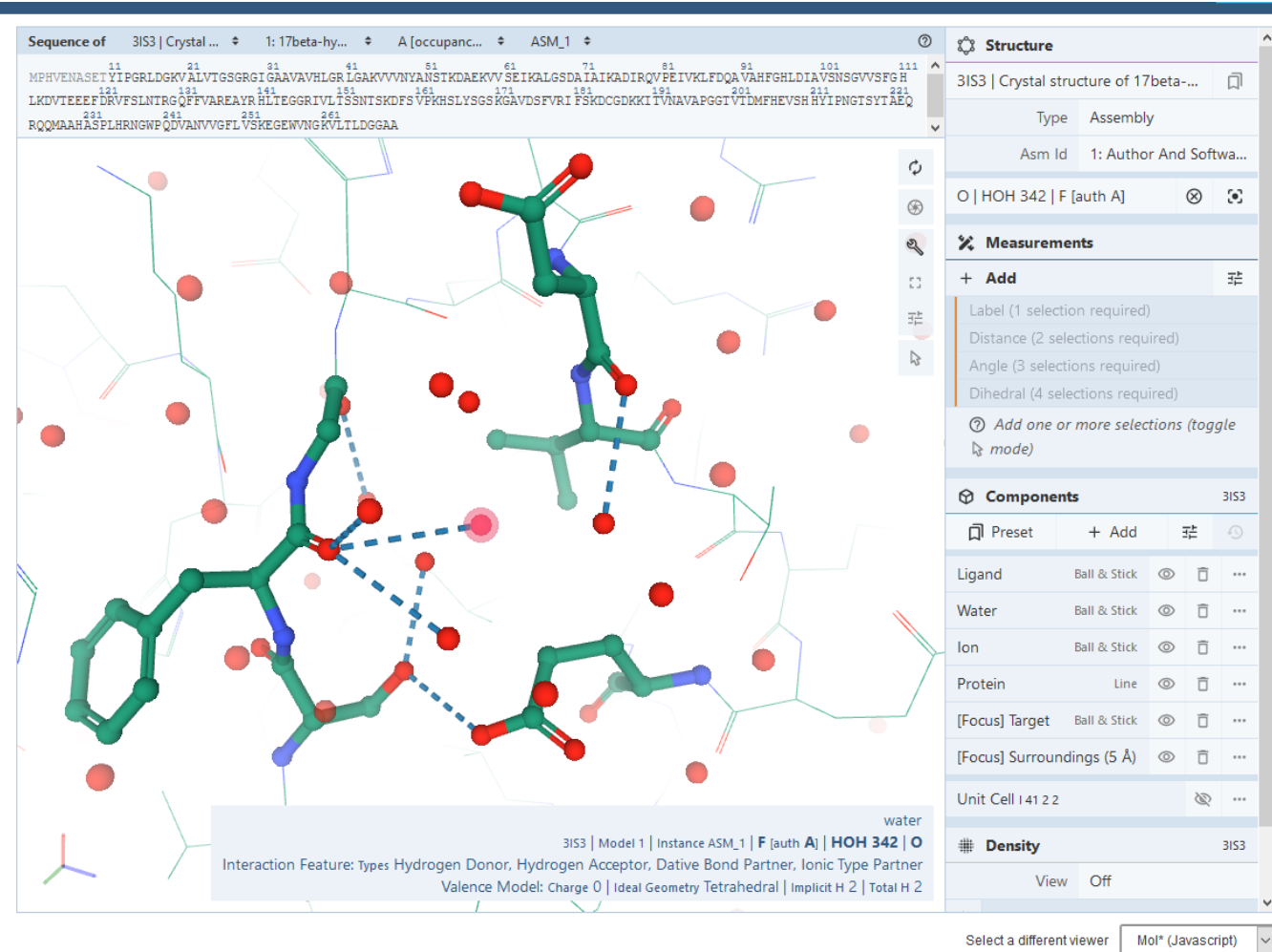
The screenshot shows the RCSB PDB website interface. The 'Analyze' dropdown menu is open, listing various analysis tools: Pairwise Structure Alignment, Protein Symmetry, Structure Quality, Map Genomic Position to Protein, PDB Statistics, EPPIC Biological Assemblies, Integrated Resources, and Third Party Tools. Below the menu, a table titled 'Data From External Resources Integrated Into RCSB PDB' is displayed. The table has columns for Resource, Home URL, and Type of Data. A search bar is located to the right of the table title.

Resource	Home URL	Type of Data
BIGG	<a href="http://bigg.ucsd.edu">http://bigg.ucsd.edu</a>	Reconstruction of metabolic pathways
Binding MOAD	<a href="http://bindingmoad.org">http://bindingmoad.org</a>	Binding affinities
BindingDB	<a href="http://www.bindingdb.org">http://www.bindingdb.org</a>	Binding affinities
BMRB	<a href="http://www.bmrwisc.edu">http://www.bmrwisc.edu</a>	BMRB-to-PDB mappings
Catalytic Site Atlas	<a href="http://www.ebi.ac.uk/thornton-srv/databases/CSA">http://www.ebi.ac.uk/thornton-srv/databases/CSA</a>	Active sites and catalytic residues in enzymes
CATH	<a href="http://www.cathdb.info">http://www.cathdb.info</a>	Protein structure classification
DrugBank	<a href="http://www.drugbank.ca">http://www.drugbank.ca</a>	Drug and drug target data
EMDB	<a href="http://emdb-empiar.org/">http://emdb-empiar.org/</a>	3DEM density maps and associated metadata
ExPASy	<a href="http://expasy.org">http://expasy.org</a>	Enzyme classification
Gencode	<a href="http://www.gencodegenes.org">http://www.gencodegenes.org</a>	Gene structure data
Gene Ontology	<a href="http://www.geneontology.org">http://www.geneontology.org</a>	Biological ontologies
Genotype-Tissue Expression - GTEx (NIH Common Fund Data Resource)	<a href="https://gtexportal.org">https://gtexportal.org</a>	Tissue-specific gene expression data
HMMER3	<a href="http://hmm.janelia.org">http://hmm.janelia.org</a>	Sequence similarity searches
Human Gene Nomenclature Committee	<a href="http://www.genenames.org">http://www.genenames.org</a>	Human gene name nomenclature and genomic information

# PDB – Visualizzazione

Il sito web del PDB integra un utile strumento di visualizzazione della struttura.

Il software di visualizzazione integrato possiede una certa flessibilità anche se sono preferiti software esterni (Pymol, ChimeraX)

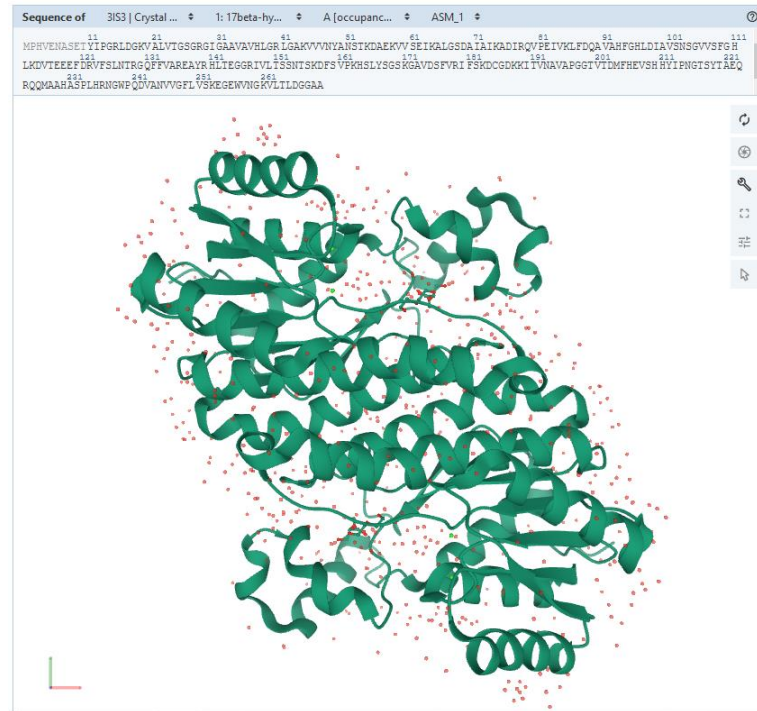




# PDB - Visualizzazione

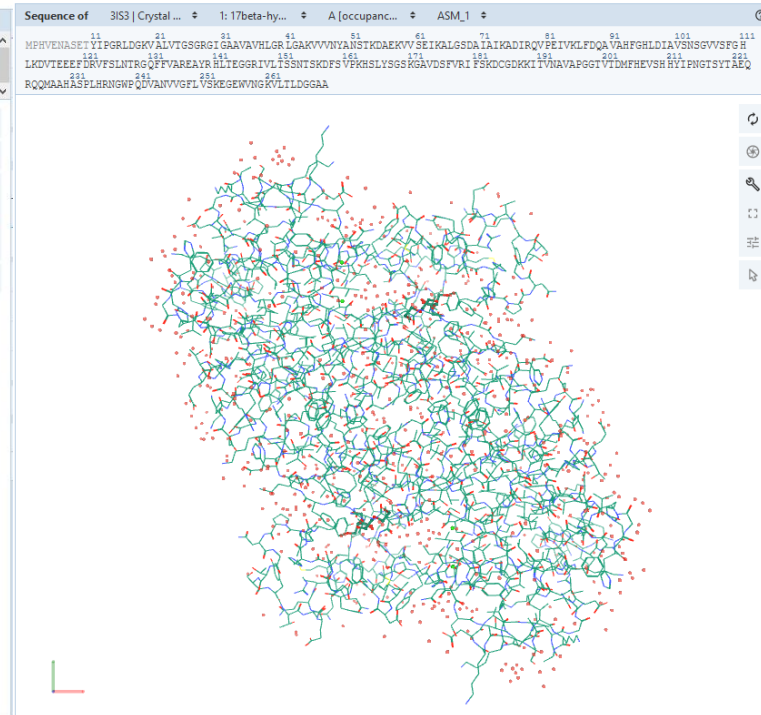
3IS3

Crystal structure of 17beta-Hydroxysteroid dehydrogenase (Apo form) from fungus Cochliobolus lunatus



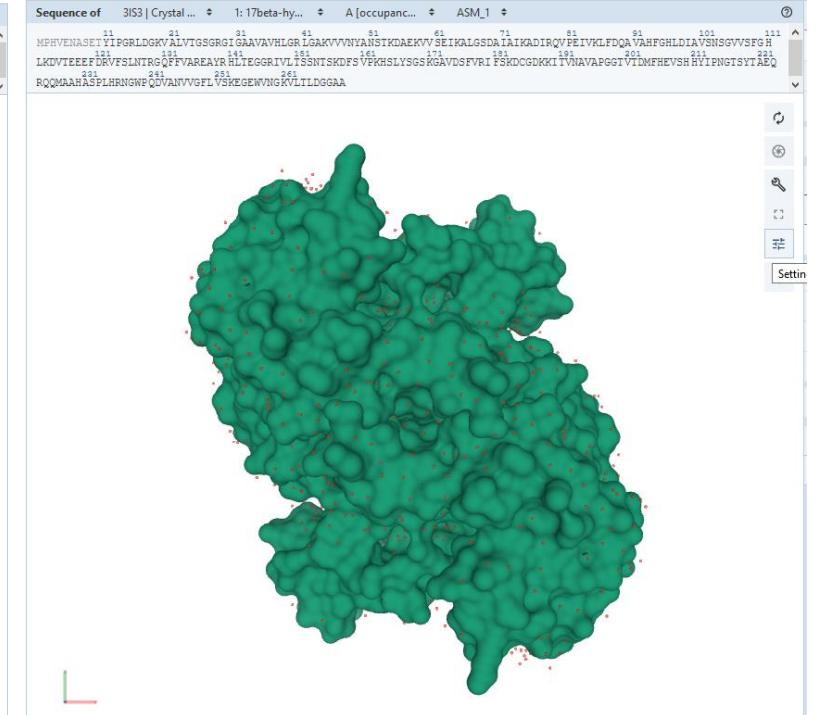
3IS3

Crystal structure of 17beta-Hydroxysteroid dehydrogenase (Apo form) from fungus Cochliobolus lunatus



3IS3

Crystal structure of 17beta-Hydroxysteroid dehydrogenase (Apo form) from fungus Cochliobolus lunatus



Le strutture biologiche macromolecolari possono essere rappresentate in vari modi capaci di comunicare **informazione di natura diversa**.

# Wireframe

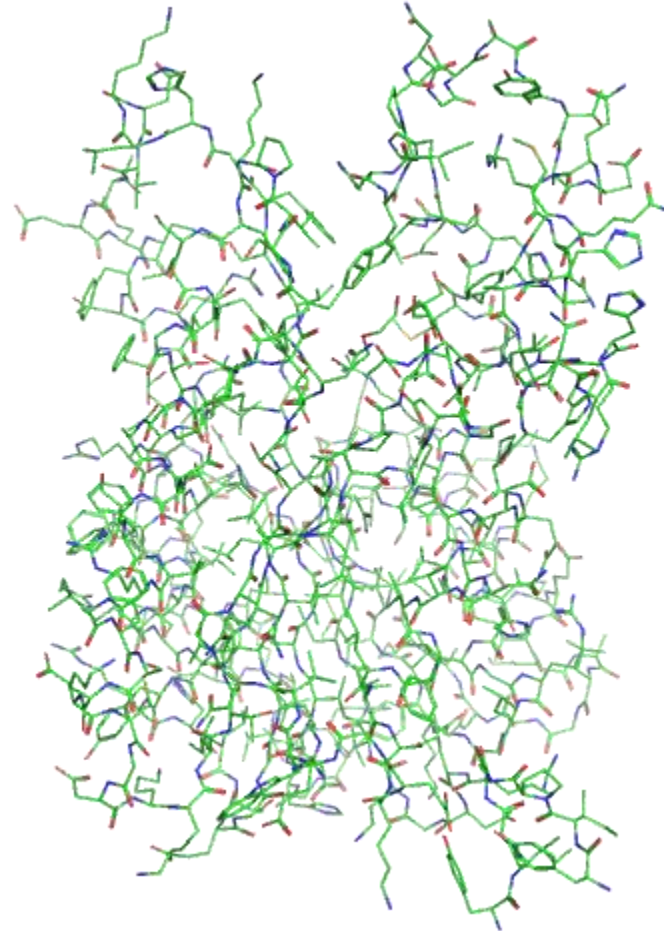
E' una rappresentazione molto diffusa.

Dal punto di vista computazionale è poco intensiva e quindi molto semplice, anche se nel complesso un po' confusa.

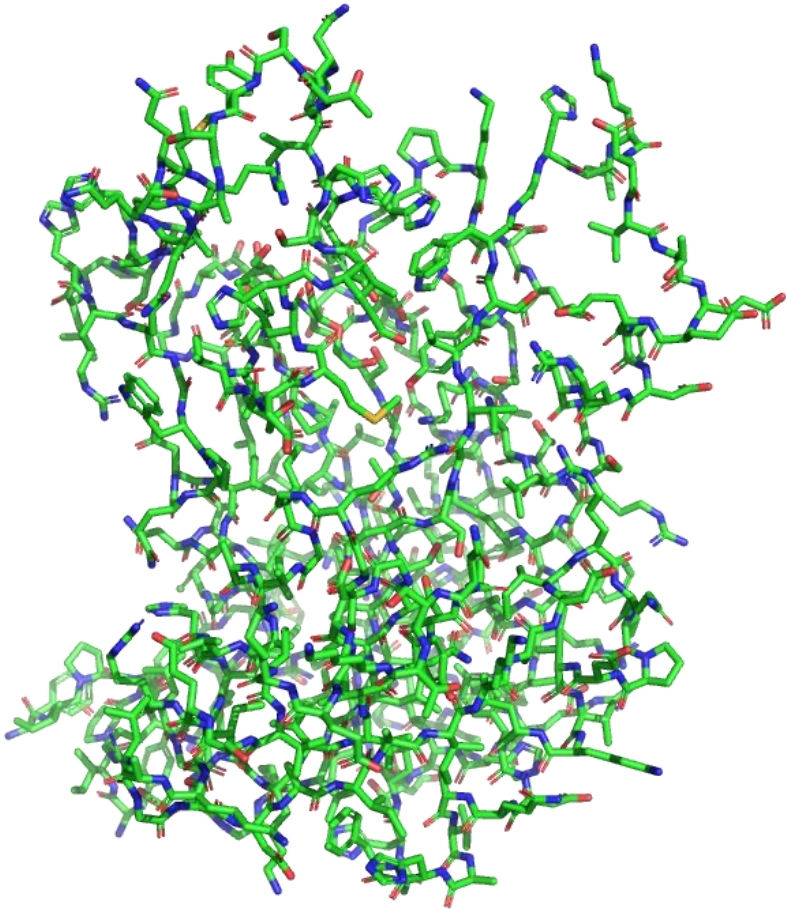
La colorazione dipende dal tipo di proprietà che si intende rappresentare:

- Specie atomica
- Fattore termico (mobilità)
- Accessibilità al solvente
- Conservazione tra strutture omologhe
- ...

**L' uso del colore è un buon modo per associare ad un residuo (o atomo) una proprietà specifica.**



# Rappresentazioni 'Baton' e 'Ball and Stick'



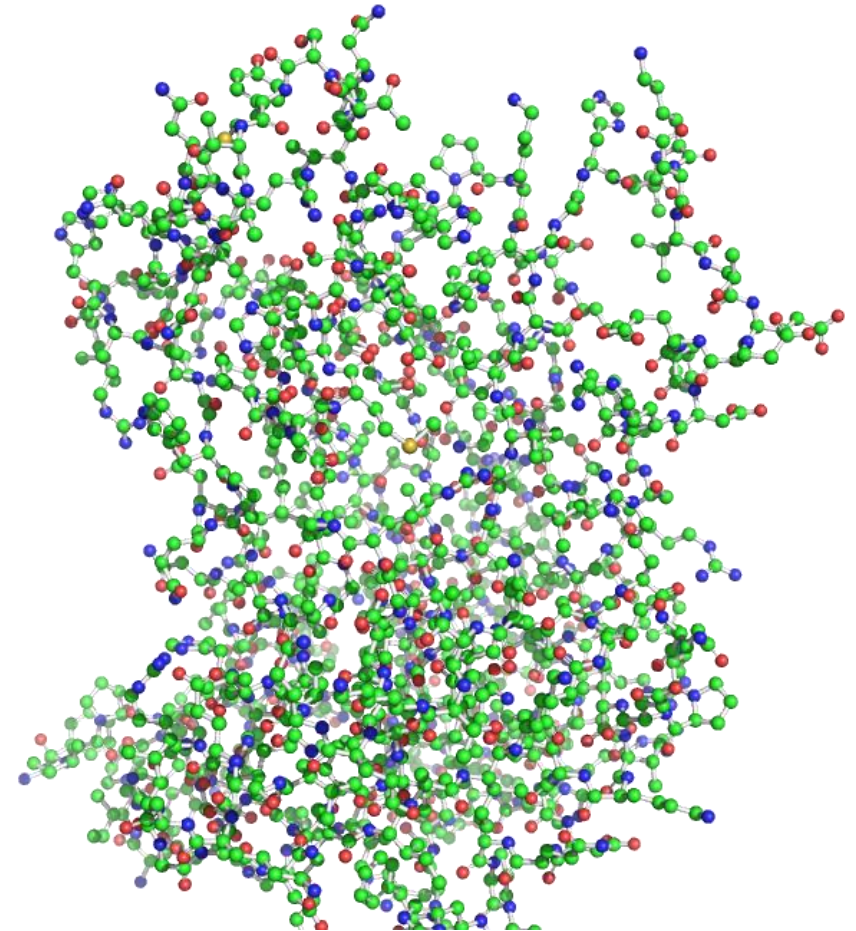
**BATON**

Sono rappresentazioni eccessivamente complesse per l'intera molecola.

Funzionano bene quando si vuole rappresentare parti limitate della struttura.

I colori sono tipicamente associati alla specie atomica.

**Sono spesso usati per rappresentare ligandi.**



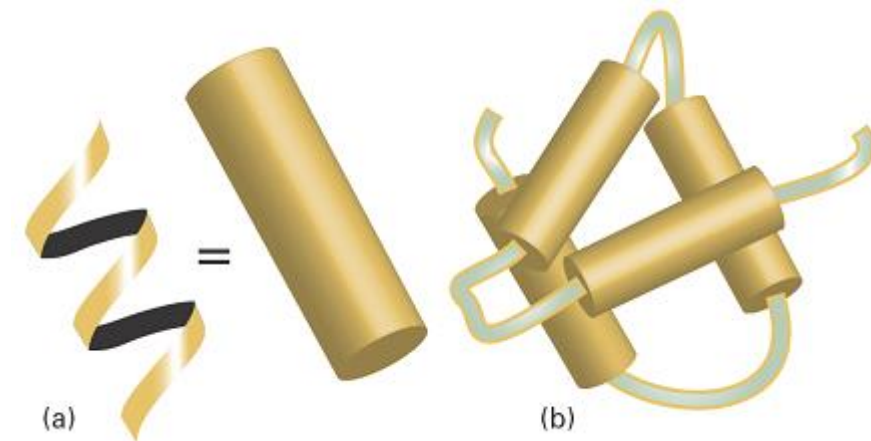
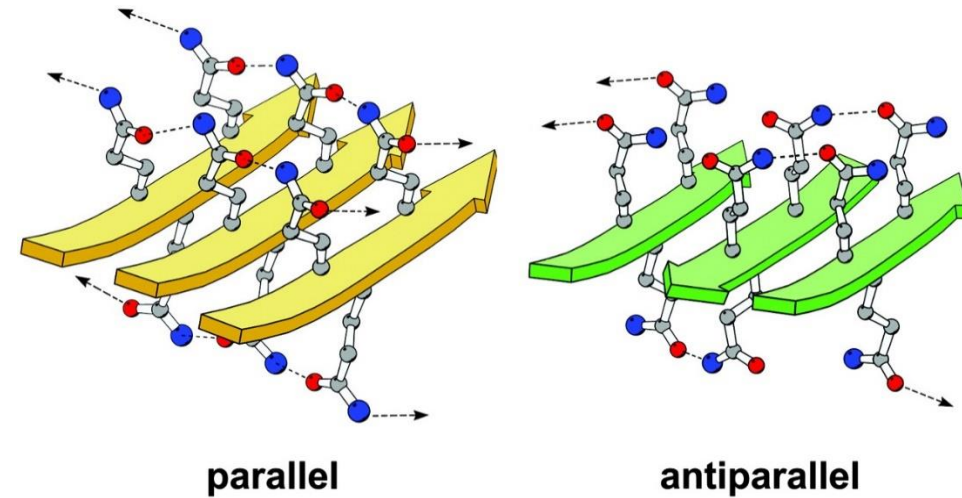
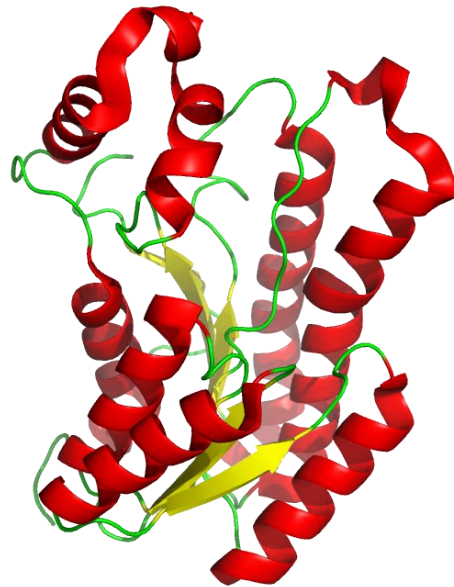
**BALL AND STICK**

# Rappresentazione della struttura – ‘Cartoon’

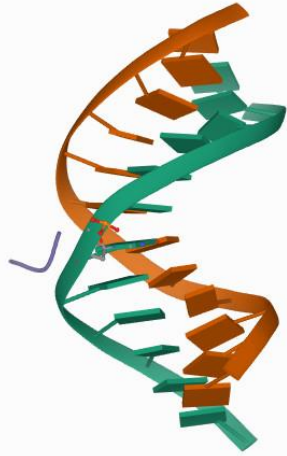
E' la classica rappresentazione del **folding** della macromolecola.

mette in evidenza gli elementi di struttura secondaria:

- Eliche (cilindri o spirale)
- Foglietti- $\beta$  (paralleli o anti-paralleli)
- loops



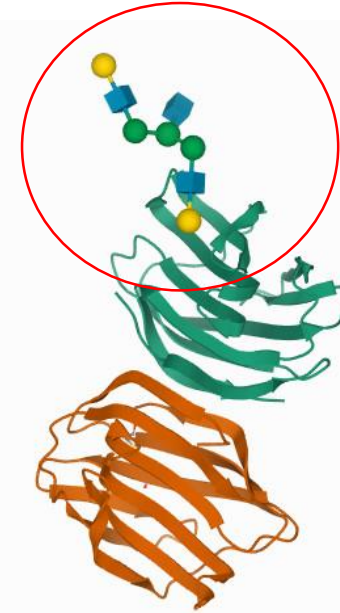
# Rappresentazione della struttura – altro



RNA/DNA



Proteina/DNA

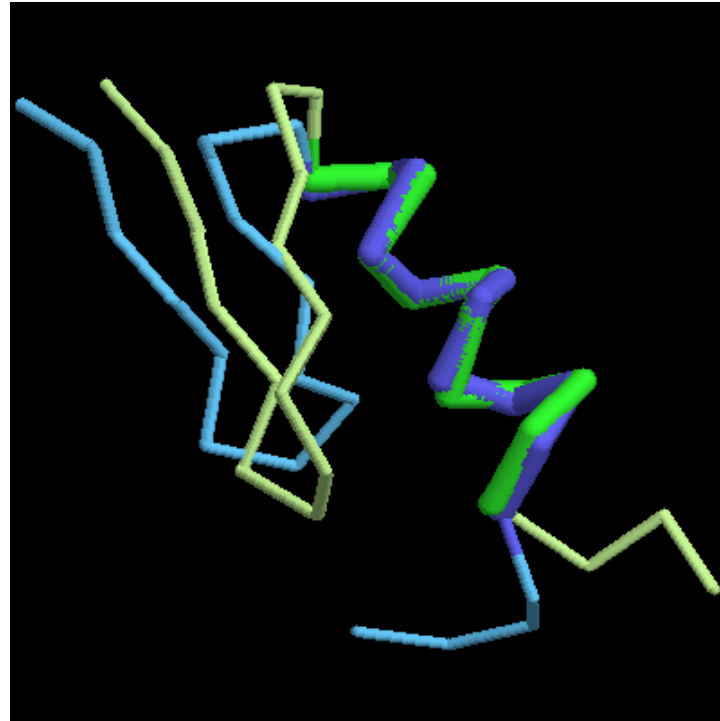


Oligosaccaridi

# Rappresentazione della struttura – traccia $C\alpha$ (Backbone)

Una forma più semplice ma meno informativa è quella di rappresentare solamente i  $C\alpha$  dei vari residui connessi tra loro da un segmento.

**E' un tipo di rappresentazione utile quando si sovrappongono modelli diversi per compararne le differenze.**



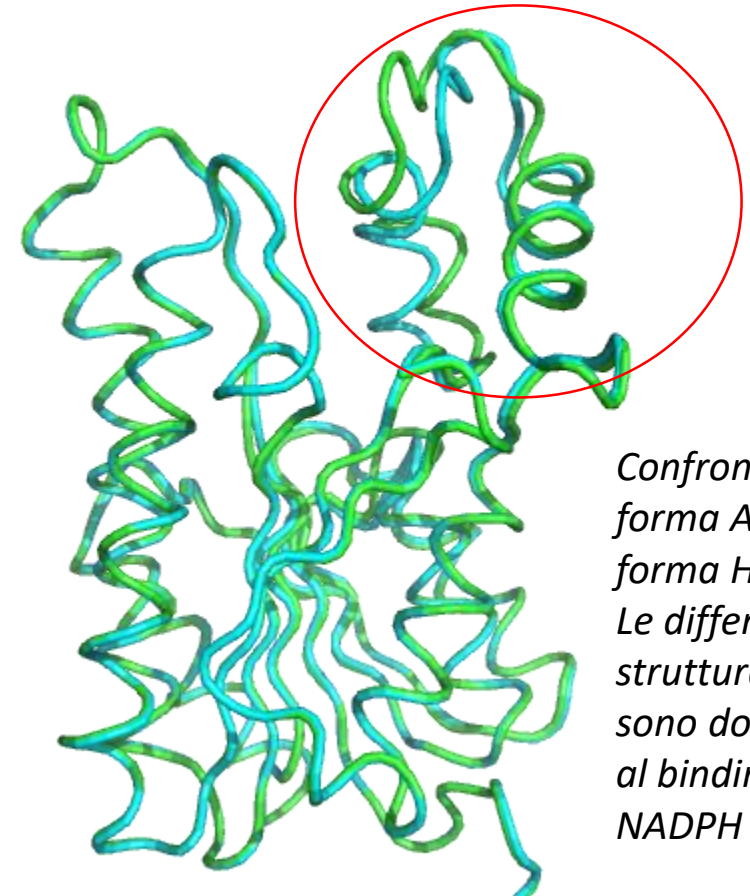
# Sovrapposizione di Strutture

La sovrapposizione tra due strutture (simili), viene effettuata cercando di minimizzare la distanza tra C $\alpha$  equivalenti.

Un protocollo (automatico) per l'allineamento:

1. I C $\alpha$  equivalenti sono stabiliti a seguito di un allineamento di sequenza.
2. Le due strutture sono quindi riportate ad un'origine comune e allineate in modo simile.
3. Viene quindi minimizzata la quantità  $\sum d^2$  dove  $d$  è la distanza tra C $\alpha$  equivalenti (la sommatoria è estesa a tutti i C $\alpha$  equivalenti)
4. Il risultato dell'allineamento è espresso graficamente e numericamente dalla deviazione quadratica media (r. m. s. d. espressa in Å) delle posizioni dei C $\alpha$  equivalenti e sovrapposti.

E' estremamente utile per **mostrare la similitudine strutturale globale** di due macromolecole, o per **evidenziare differenze locali**.

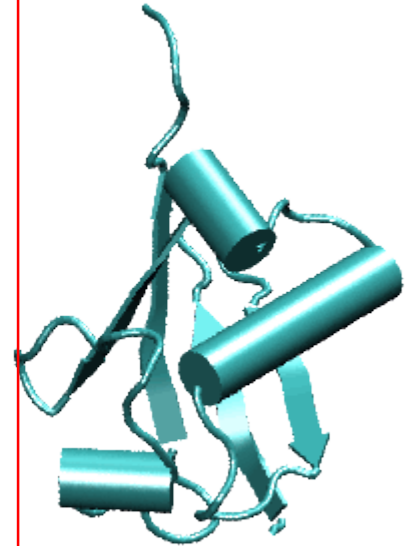
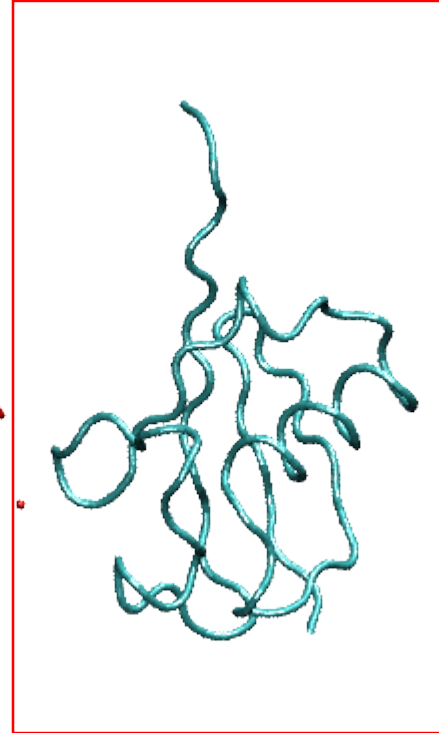
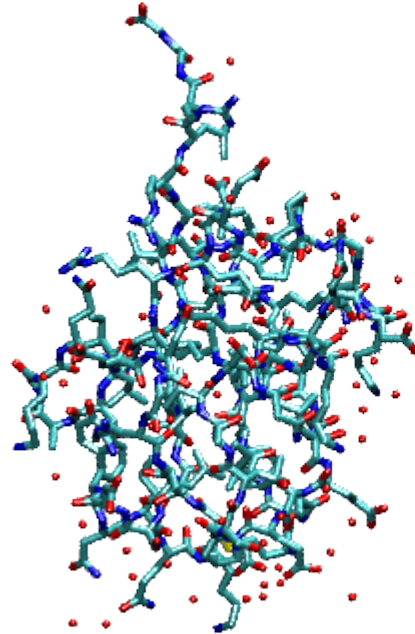


*Confronto tra  
forma Apo e  
forma Holo.  
Le differenze  
strutturali  
sono dovute  
al binding di  
NADPH*

# Rappresentazione – Worm-Like (o liquorice)

La rappresentazione worm-like non introduce nessuna informazione ulteriore rispetto a quella dei  $C\alpha$ , su cui è basata.

E' un po' più aggraziata...





# Rappresentazione della struttura – ‘putty’

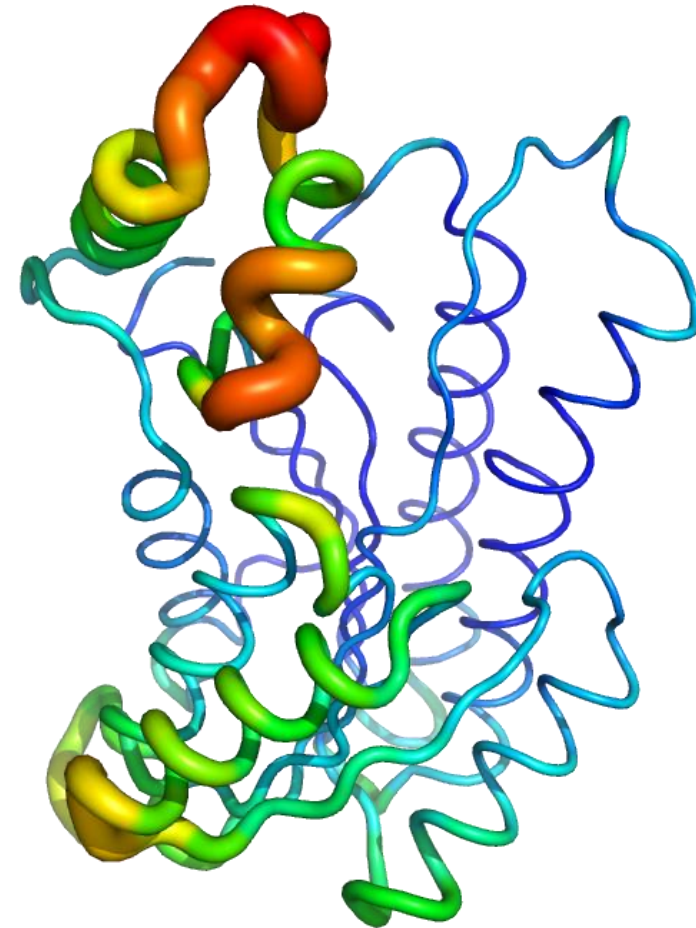
Questa rappresentazione, nota come *putty*, è nata da uno specifico software (*pymol*); oltre a rappresentare la molecola come worm-like, associava lo spessore e il colore del ‘tubo’ al valore del fattore termico B.

Più è grande B, più lo spessore è grande e il colore tende al rosso.

B piccoli sono associati a piccoli spessori e colore blu.

E’ una rappresentazione utile per mostrare la ***flessibilità***, associata al valore di B, delle varie parti della molecola.

**In generale si possono associare colore e spessore a proprietà diverse dal valore di B, facendone uno strumento versatile per ‘mappare’ delle proprietà sulla struttura** (es: conservazione del residuo, esposizione al solvente...)



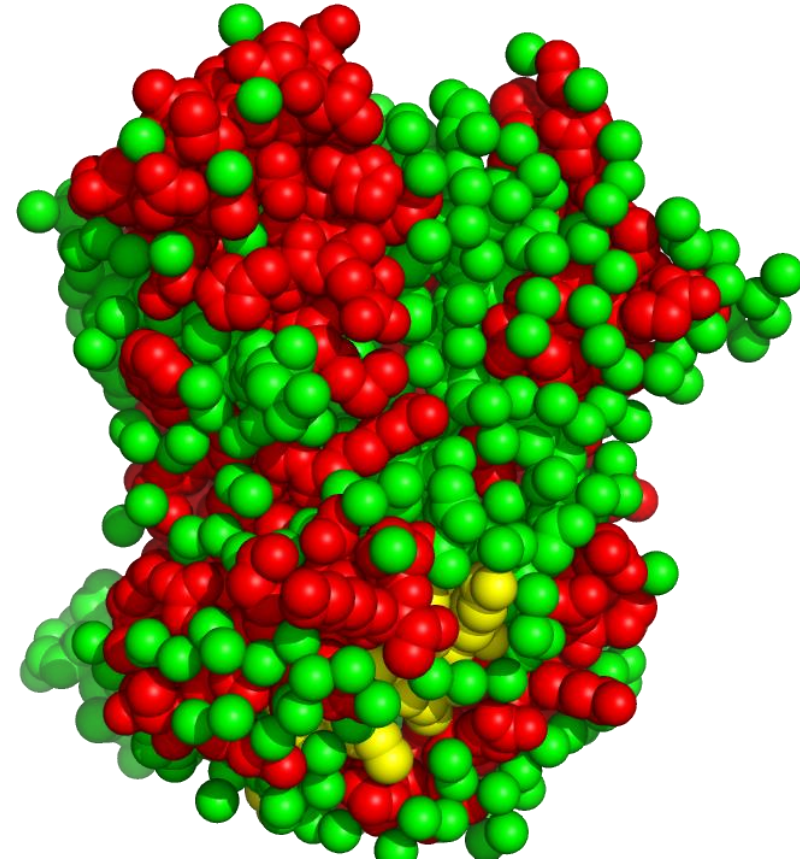
# Rappresentazioni Volumetriche

# Il modello a sfere (space-filling)

Il modello a sfere disegna una sfera (raggio di van der Waals) in corrispondenza di ogni atomo.

Rende l'idea degli **aspetti volumetrici** della proteina, ma non è molto usata per la molecola intera (è preferita la *superficie accessibile al solvente*).

Può essere interessante per un ligando nel sito di binding (fornisce un'idea dello spazio occupato).



# Rappresentazione delle strutture: superfici

Un modo molto efficace di **'mappare'** le proprietà atomiche della proteina ed ottenere così una sorta di cartografia della sua superficie, è quella di costruire una superficie della molecola e riportare sulla superficie la proprietà di interesse.

Distinguiamo tra:

- ***Superficie di van der Waals***
- ***Superficie Accessibile al Solvente***

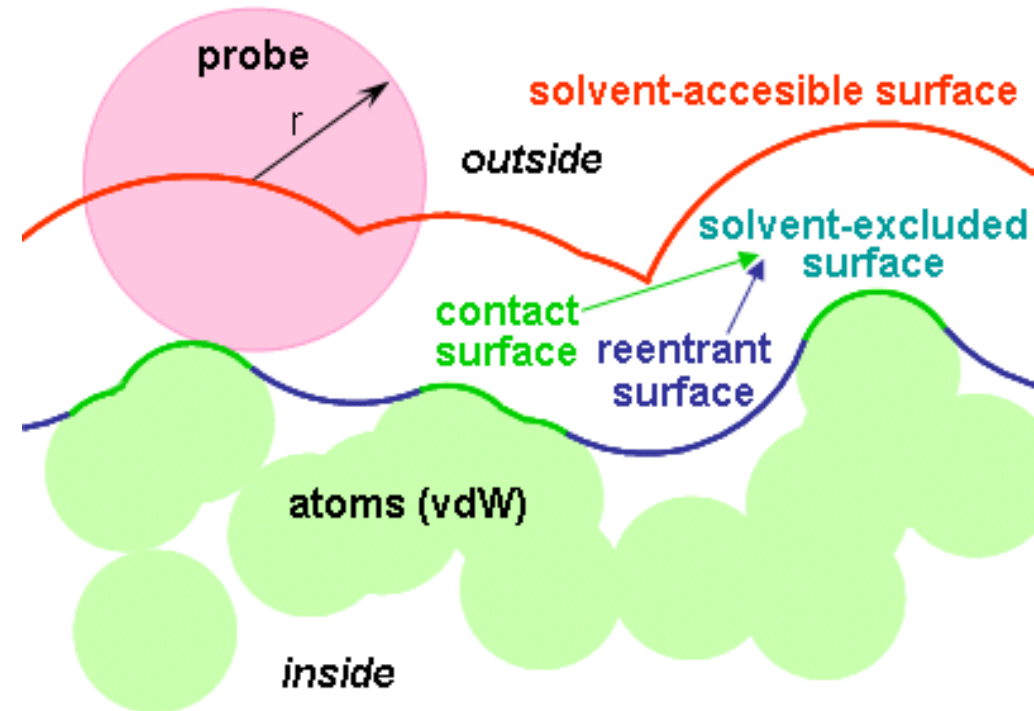
# Superficie di van der Waals

La superficie di van der Waals è determinata dal raggio di van der Waals dei singoli atomi.

Il raggio di van der Waals determina il raggio di un atomo concepito come sfera rigida e 'impenetrabile'.

La convoluzione delle superfici di van der Waals di tutti gli atomi, la **superficie di van der Waals**, definisce un volume che può essere considerato come il volume della molecola.

La superficie di van der Waals non è particolarmente utile, né troppo utilizzata in biologia strutturale.



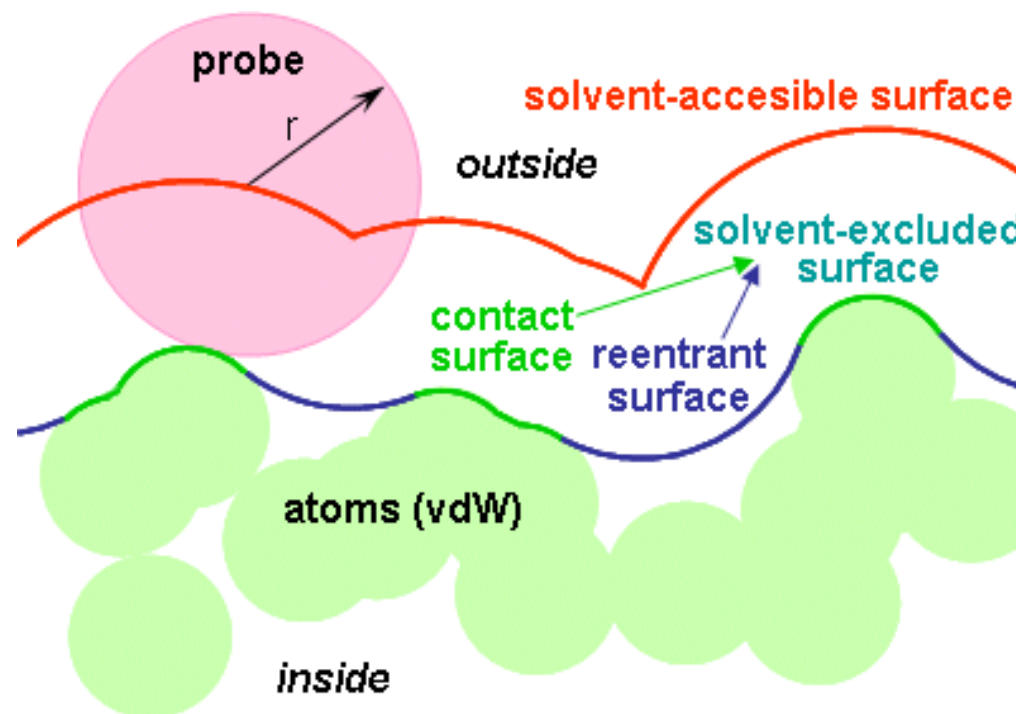
# Superficie accessibile ed esclusa al solvente

La superficie accessibile al solvente (SASA) si ottiene disegnando una superficie ottenuta a partire da una sfera (probe) di raggio  $r$ , rappresentativa di una molecola di acqua, che viene fatta rotolare sugli atomi accessibili alla sfera stessa.

Nel rotolare la sfera definisce due superfici:

La **superficie esclusa al solvente**, definita dalla convoluzione dei punti generati della sfera (rotolante sulla macromolecola) più vicini alla molecola di cui si vuole descrivere la superficie. Questi punti possono toccare gli atomi (se questi sono esposti) o meno (se lo spazio è troppo 'ristretto'; punti 'rientranti').

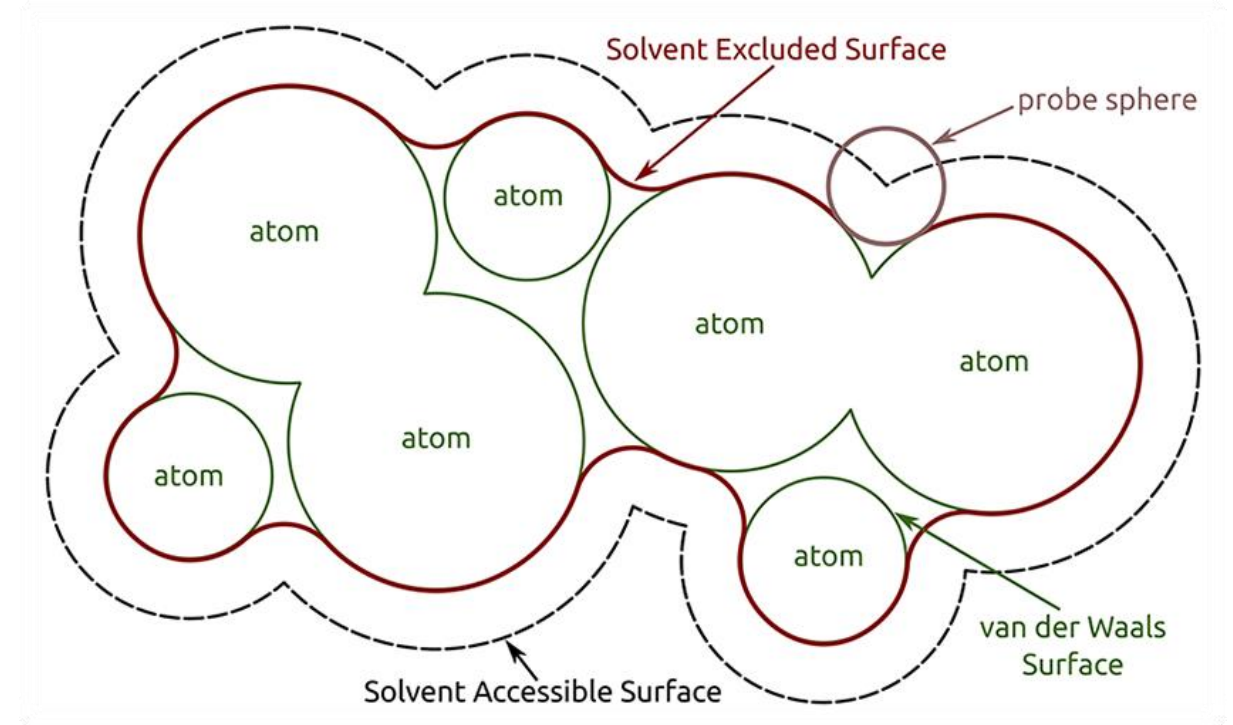
La **superficie accessibile al solvente** è definita dalla convoluzione dei punti generati dal centro della sfera, man mano che rotola sulla superficie macromolecolare.



# Superficie accessibile al solvente

La superficie accessibile al solvente è in generale più interessante della superficie di van der Waals, in quanto ***rappresenta la superficie della proteina che può effettivamente venire in contatto con altre molecole*** (almeno di solvente).

La superficie accessibile al solvente è più grande della superficie di van der Waals

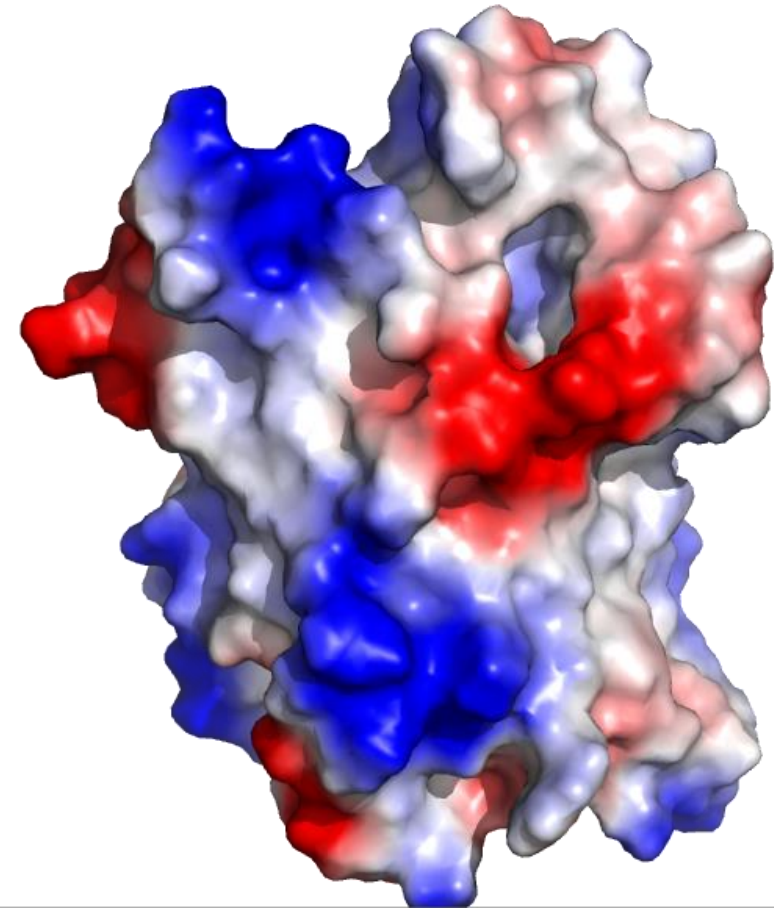


# Superficie – potenziale elettrostatico

Una proprietà molto interessante che viene spesso mappata sulla superficie è il **potenziale elettrostatico**, che può essere calcolato a partire dalle coordinate e dal tipo di atomi.

I diversi tipi atomici avranno cariche elettriche differenti, nel complesso una proteina può essere vista come una distribuzione di cariche elettriche associate quindi ad un potenziale elettrostatico.

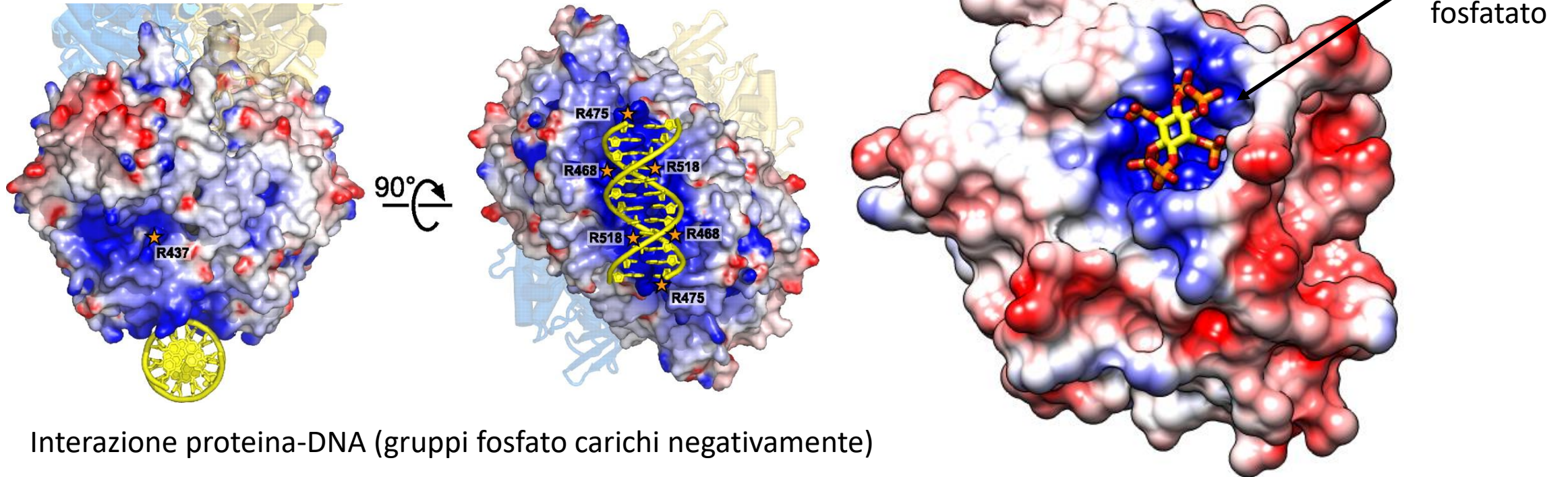
***Sulla superficie è riportato Il valore del potenziale elettrostatico corrispondente*** (sotto forma scala di colore)





# Importanza del potenziale elettrostatico

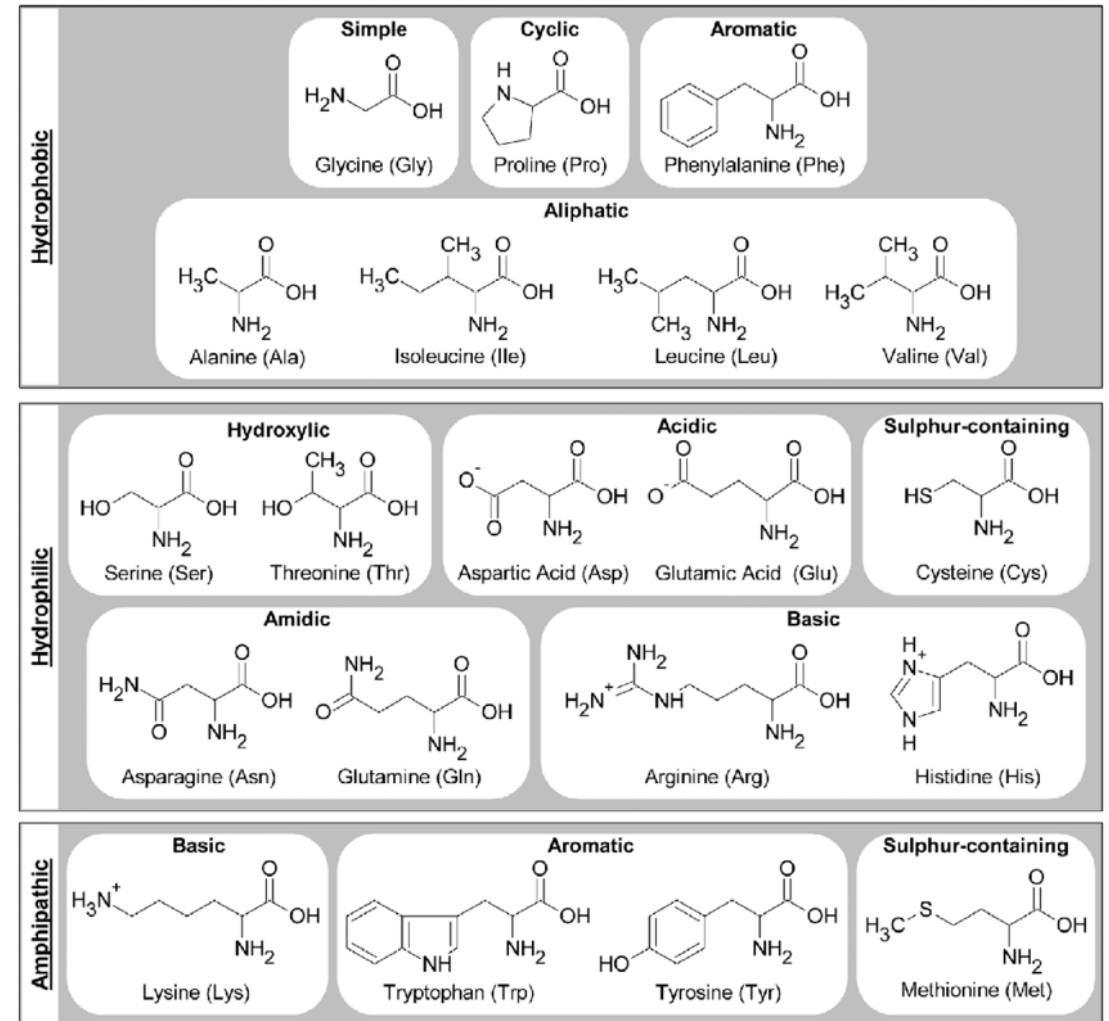
Questa proprietà rende conto della capacità della superficie di interagire con molecole dotate di carica elettrica discreta o semplicemente polari.



# Idrofobicità

L'idrofobicità rende conto della tendenza di una molecola ad interagire favorevolmente con l'acqua (intesa come solvente) o meno.

Amino acid	Index value	Amino acid	Index value
A	0.62	M	0.64
C	0.29	N	-0.78
D	-0.9	P	0.12
E	-0.74	Q	-0.85
F	1.19	R	-2.53
G	0.48	S	-0.18
H	-0.4	T	-0.05
I	1.38	V	1.08
K	-1.5	W	0.81
L	1.06	Y	0.26

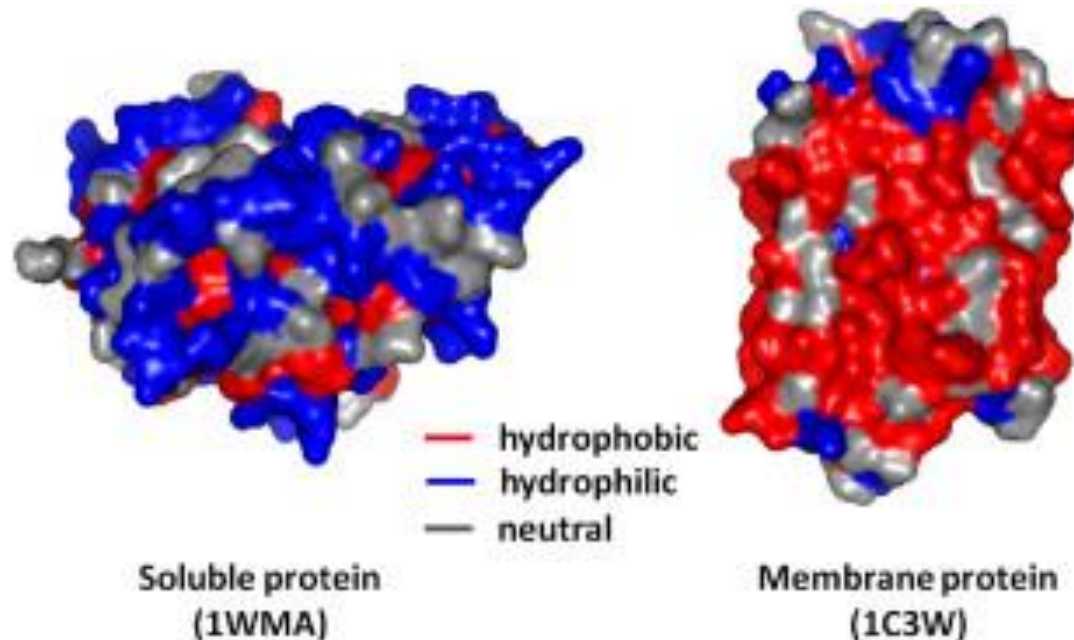


# Superficie - Idrofobicità

L'*idrofbicità/idrofilicità* è una proprietà importante e viene 'mappata' sulla SASA.

L'idrofobicità di parti diversi della macromolecola rendono conto di eventuali interazioni con altre proteine, ligandi o superfici (lipofiliche o meno).

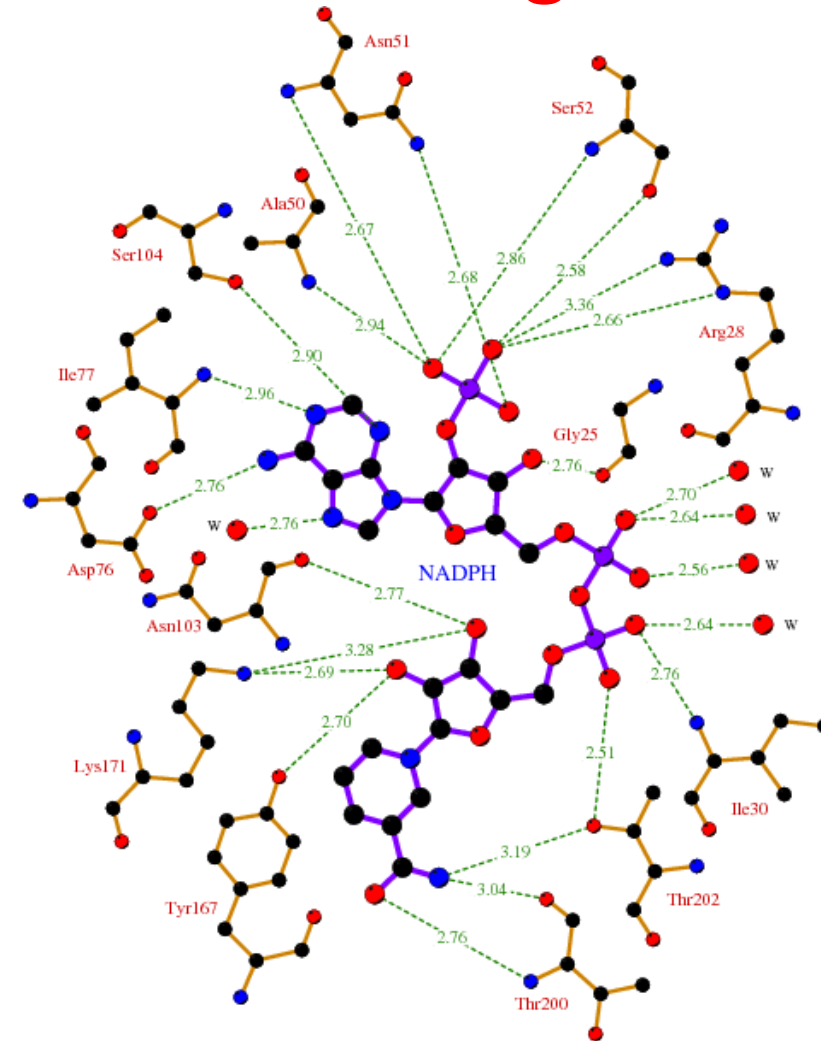
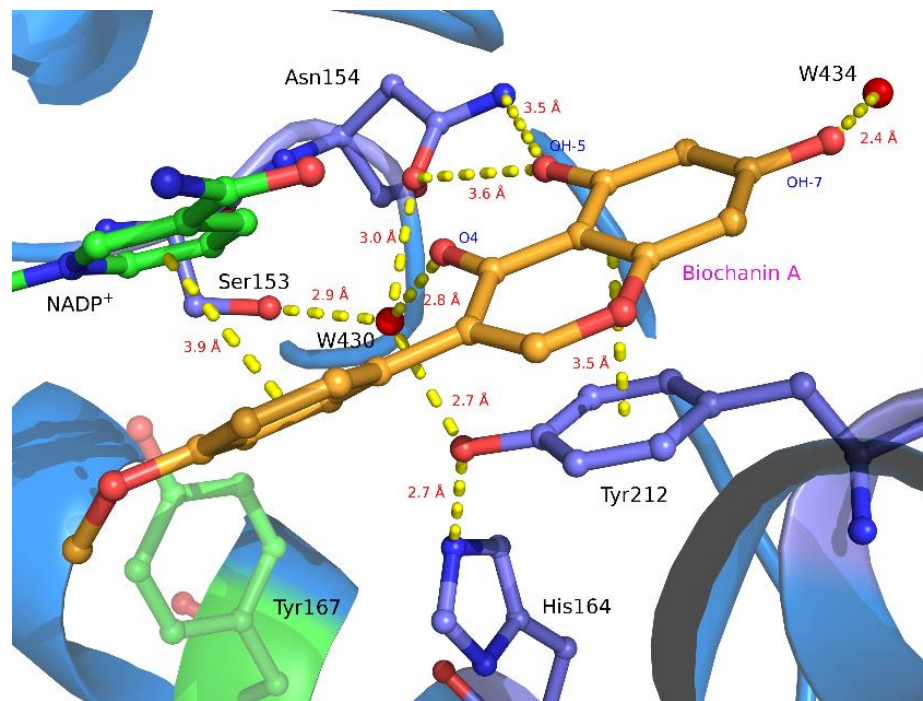
Una proteina solubile interagisce favorevolmente con l'acqua. Ci saranno residui idrofilici in superficie



La parte idrofobica interagisce con il doppio strato lipidico della membrana cellulare (apolare)

# Rappresentazione - Interazione con ligandi

Nel caso di formazione di complessi tra la proteina e un ligando, la determinazione strutturale del complesso in genere fornisce utilissime informazioni sul meccanismo di interazione.



# Topologia delle strutture proteiche

# Classificazione Strutturale delle proteine - dominio

Esistono diverse definizioni di cosa è un dominio nel contesto di una struttura:

- 1) La minima parte di un gene che ha una specifica proprietà funzionale
- 2) Regioni che mostrano similitudine nella sequenza
- 3) Una regione proteica dotata di una specifica proprietà funzionale
- 4) Regione di una proteina che ricorre in altre proteine non correlate**
- 5) Una regione distinta e compatta di una proteina
- 6) ...

Esistono di conseguenza diversi modi di classificare strutturalmente le proteine, gli approcci più usati sono:

- **CATH**
- **SCOP**

# CATH



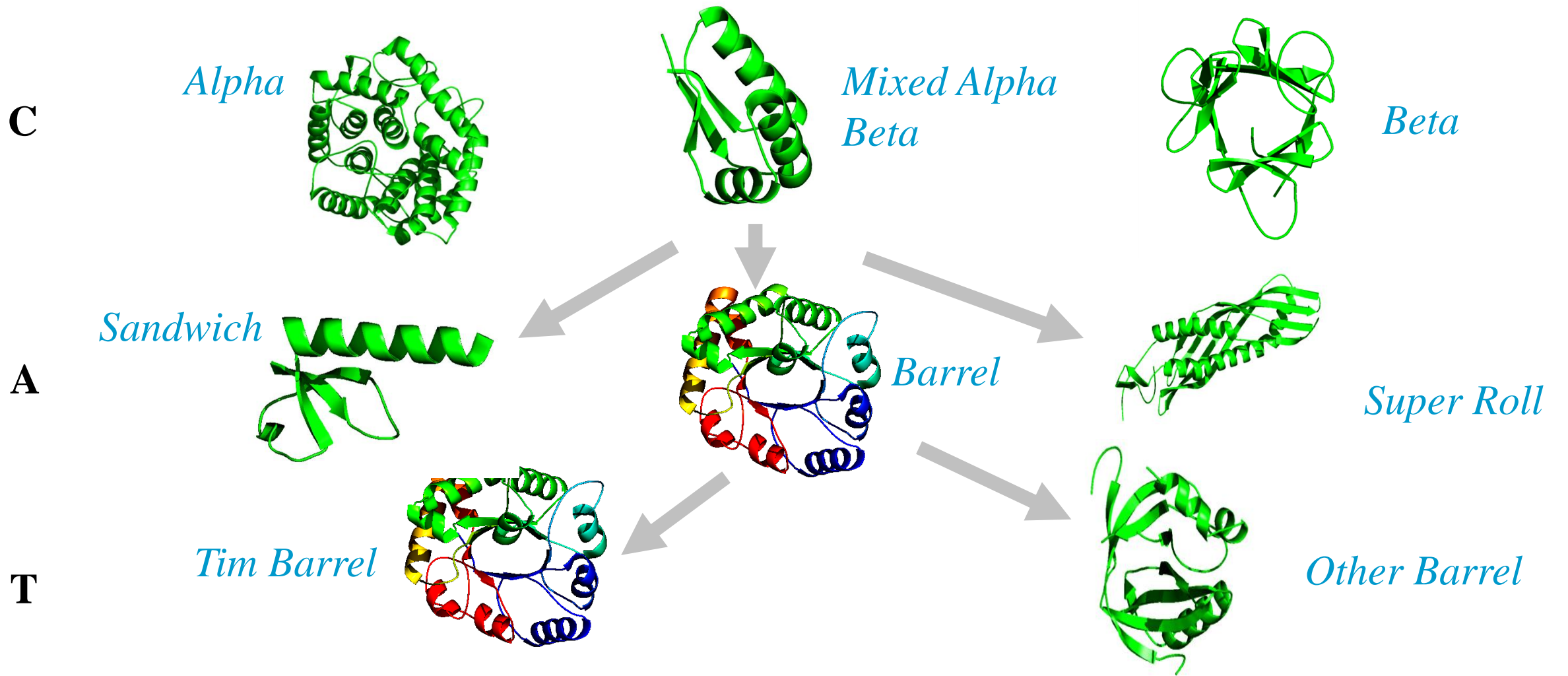
CATH organizza e classifica le strutture presenti nel Protein Data Bank

## What is CATH?

The CATH database is a hierarchical domain classification of protein structures in the Protein Data Bank. Protein structures are classified using a combination of automated and manual procedures. There are four major levels in this hierarchy:

- Class - structures are classified according to their secondary structure composition (mostly alpha, mostly beta, mixed alpha/beta or few secondary structures).
- Architecture - structures are classified according to their overall shape as determined by the orientations of the secondary structures in 3D space but ignores the connectivity between them.
- Topology (fold family) - structures are grouped into fold groups at this level depending on both the overall shape and connectivity of the secondary structures.
- Homologous superfamily - this level groups together protein domains which are thought to share a common ancestor and can therefore be described as homologous.

# CATH





# SCOP

## SCOP 2

SCOP: Structural Classification of Proteins

Anche SCOP provvede ad una classificazione generale ed ampia delle proteine (class, fold, domain...), non limitandosi alle sole strutture del PDB (come le proteine fibrose, proteine non strutturate)

### Browse by structural class

- All alpha proteins
- All beta proteins
- Alpha and beta proteins(a/b)
- Alpha and beta proteins(a+b)
- Small proteins

### Browse by protein type

- Globular proteins
- Membrane proteins
- Fibrous proteins
- Non-globular/Intrinsically unstructured proteins

# Topologia della struttura e PDB

Nel PDB, per ogni struttura sono identificati i domini presenti

Per ogni dominio è descritto la topologia generale della struttura

## 3IS3

Display Files Download Files

Crystal structure of 17beta-Hydroxysteroid dehydrogenase (Apo form) from fungus Cochliobolus lunatus

Domain Annotation: SCOP Classification

[SCOP Database \(version: 2.0\) Homepage](#)

Chains	Domain Info	Class	Fold	Superfamily	Family	Domain	Species
A	d3is3a_	<a href="#">Alpha and beta proteins (a/b)</a>	<a href="#">NAD(P)-binding Rossmann-fold domains</a>	<a href="#">NAD(P)-binding Rossmann-fold domains</a>	<a href="#">automated matches</a>	<a href="#">automated matches</a>	<a href="#">(Curvularia lunata)</a> [TaxId: 5503]

Domain Annotation: CATH

[CATH Database \(version 4.2.0\) Homepage](#)

Chain	Domain	Class	Architecture	Topology	Homology
A	3.40.50.720	<a href="#">Alpha Beta</a>	<a href="#">3-Layer(aba) Sandwich</a>	<a href="#">Rossmann fold</a>	<a href="#">NAD(P)-binding Rossmann-like Domain</a>

Gene Product Annotation

[Gene Product Annotation](#)

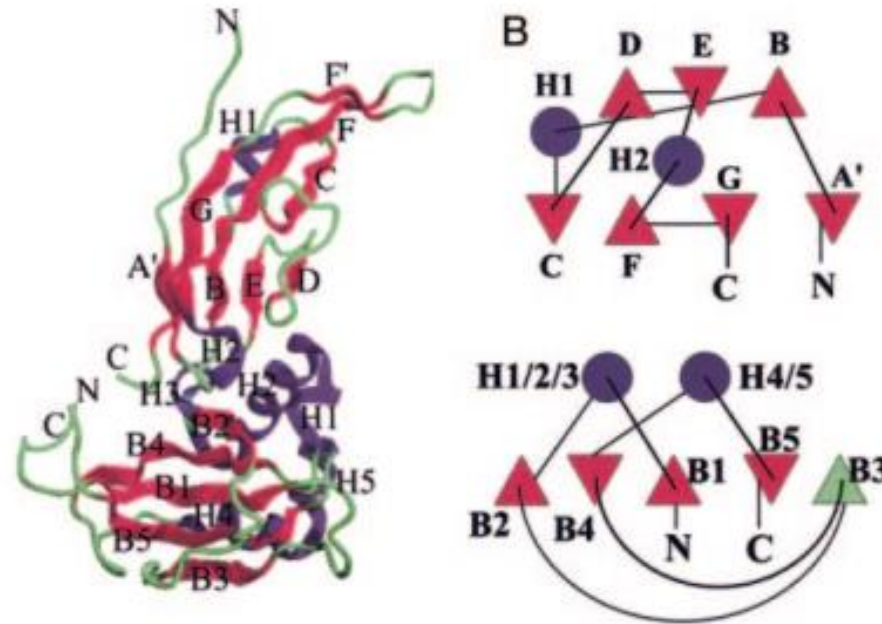
Chains	Polymer	Molecular Function	Biological Process	Cellular Component
A	17beta-hydroxysteroid dehydrogenase	<ul style="list-style-type: none"><li><a href="#">catalytic activity</a></li><li><a href="#">oxidoreductase activity</a></li></ul>		

# La struttura cristallografica in un articolo scientifico

# Contesto del lavoro cristallografico (1)

## The Structure of the N-terminal Region of Murine Skeletal Muscle $\alpha$ -Dystroglycan Discloses a Modular Architecture\*

Dystroglycan (DG) is a cell surface receptor consisting of two subunits:  $\alpha$ -dystroglycan, extracellular and highly glycosylated, and  $\beta$ -dystroglycan, spanning the cell membrane. It is a pivotal member of the dystrophin-glycoprotein complex and is involved in a wide variety of important cellular processes such as the stabilization of the muscle fiber sarcolemma or the clustering of acetylcholine receptors. We report the 2.3-Å resolution crystal structure of the murine skeletal muscle N-terminal  $\alpha$ -DG region, which confirms the presence of two autonomous domains; the first finally identified as an Ig-like and the second resembling ribosomal RNA-binding proteins. Solid-phase laminin binding assays show the occurrence of protein-protein type of interactions involving the Ig-like domain of  $\alpha$ -DG.



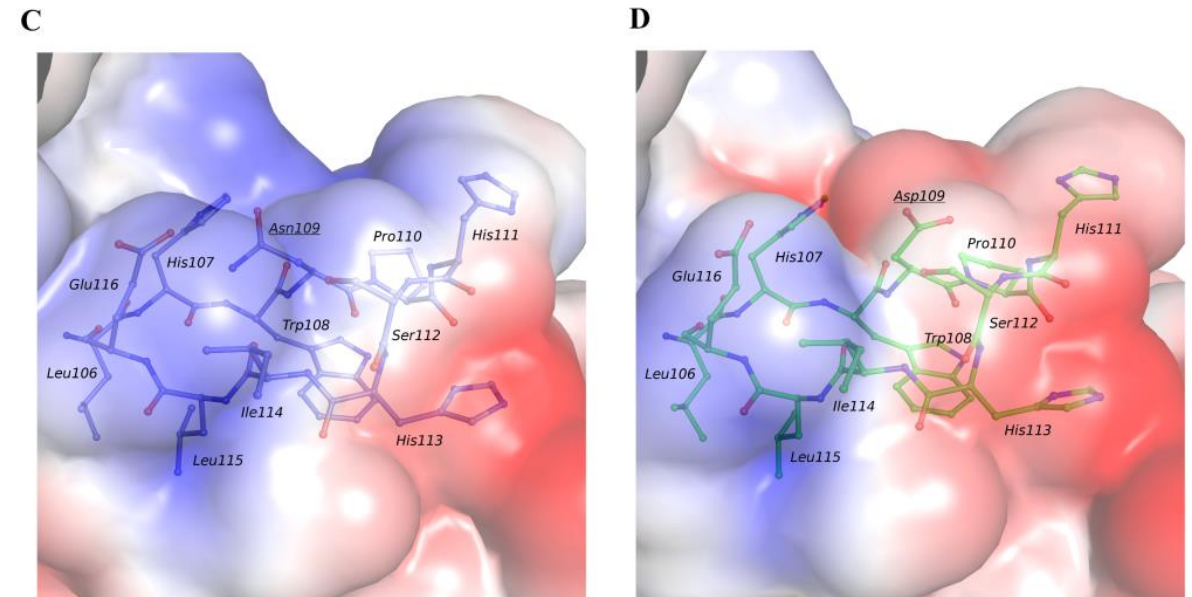
La proteina è completamente nuova. L'architettura molecolare (FOLD) è la parte più rilevante

# Contesto del lavoro cristallografico (2)

## The effect of the pathological V72I, D109N and T190M missense mutations on the molecular structure of $\alpha$ -dystroglycan

### Abstract

Dystroglycan (DG) is a highly glycosylated protein complex that links the cytoskeleton with the extracellular matrix, mediating fundamental physiological functions such as mechanical stability of tissues, matrix organization and cell polarity. A crucial role in the glycosylation of the DG  $\alpha$  subunit is played by its own N-terminal region that is required by the glycosyltransferase LARGE. Alteration in this O-glycosylation deeply impairs the high affinity binding to other extracellular matrix proteins such as laminins. Recently, three missense mutations in the gene encoding DG, mapped in the  $\alpha$ -DG N-terminal region, were found to be responsible for hypoglycosylated states, causing congenital diseases of different severity referred as primary dystroglycanopathies. To gain insight on the molecular basis of these disorders, we investigated the crystallographic and solution structures of these pathological point mutants, namely V72I, D109N and T190M. Small Angle X-ray Scattering analysis reveals that these mutations affect the structures in solution, altering the distribution between compact and more elongated conformations. These results, supported by biochemical and biophysical assays, point to an altered structural flexibility of the mutant  $\alpha$ -DG N-terminal region that may have repercussions on its interaction with LARGE and/or other DG-modifying enzymes, eventually reducing their catalytic efficiency.

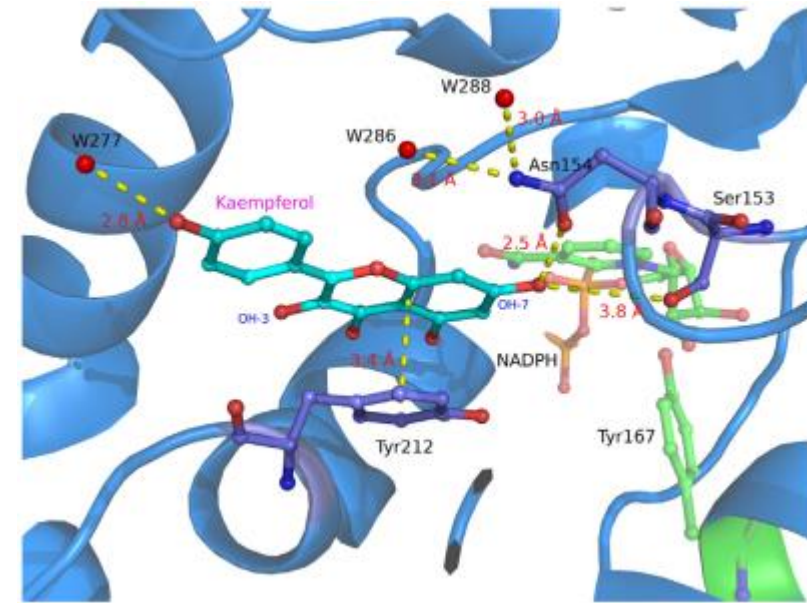


Studio su mutanti: gli effetti strutturali globali e soprattutto locali, dovuti alla mutazione, sono la parte rilevante

# Contesto del lavoro cristallografico (3)

## Structural basis for inhibition of 17 $\beta$ -hydroxysteroid dehydrogenases by phytoestrogens: The case of fungal 17 $\beta$ -HSDcl

Phytoestrogens are plant-derived compounds that functionally and structurally mimic mammalian estrogens. Phytoestrogens have broad inhibitory activities toward several steroidogenic enzymes, such as the 17 $\beta$ -hydroxysteroid dehydrogenases (17 $\beta$ -HSDs), which modulate the biological potency of androgens and estrogens in mammals. However, to date, no crystallographic data are available to explain phytoestrogens binding to mammalian 17 $\beta$ -HSDs. NADP(H)-dependent 17 $\beta$ -HSD from the filamentous fungus *Cochliobolus lunatus* (17 $\beta$ -HSDcl) has been the subject of extensive biochemical, kinetic and quantitative structure–activity relationship studies that have shown that the flavonols are the most potent inhibitors. In the present study, we investigated the structure–activity relationships of the ternary complexes between the holo form of 17 $\beta$ -HSDcl and the flavonols kaempferol and 3,7-dihydroxyflavone, in comparison with the isoflavones genistein and biochanin A. Crystallographic data are accompanied by kinetic analysis of the inhibition mechanisms for six flavonols (3-hydroxyflavone, 3,7-dihydroxyflavone, kaempferol, quercetin, fisetin, myricetin), one flavanone (naringenin), one flavone (luteolin), and two isoflavones (genistein, biochanin A). The kinetics analysis shows that the degree of hydroxylation of ring B significantly influences the overall inhibitory efficacy of the flavonols. A distinct binding mode defines the interactions between 17 $\beta$ -HSDcl and the flavones and isoflavones. Moreover, the complex with biochanin A reveals an unusual binding mode that appears to account for its greater inhibition of 17 $\beta$ -HSDcl with respect to genistein. Overall, these data provide a blueprint for identification of the distinct molecular determinants that underpin 17 $\beta$ -HSD inhibition by phytoestrogens.



Complesso proteina-ligando: l'interazione tra molecole è la parte rilevante

---

# Espressione e purificazione

## Mutagenesis and primers

Traditionally, we have worked on murine dystroglycan since it displays a very high degree of identity (93%) with human  $\alpha$ -DG. We introduced an additional mutation, R166H, within the N-terminal domain of  $\alpha$ -DG in order to make it more resistant to proteolysis [13]; both the wild-type and the mutated variants herein analyzed harbor this mutation. The murine  $\alpha$ -DG (50–313)R166H (hereinafter WT) DNA was cloned into a bacterial vector, pHis-Trx, for the expression of the protein as thioredoxin fusion product, also containing an N-terminal 6xHis tag and a thrombin cleavage site, as previously described [13]. The point mutation T190M was introduced into the WT DNA construct using the QuikChange site-directed mutagenesis kit (Stratagene) and appropriate primers. Briefly, the mutation was inserted within the coding region of the WT N-terminal domain cloned in the vector pHis-Trx [13] for prokaryotic expression aimed at crystallographic analysis, and within the entire cDNA sequence of the murine DG cloned in the pEGFP-N1 plasmid for expression in eukaryotic cells [19], respectively, using the following primers:

Forward: 5' -CCAGTGACTGTCCTTATGGTGATTCTGGATGCT-3'

Reverse: 5' -AGCATCCAGAATCACCATAAGGACAGTCACTGG-3'

The construct DG<sup>T190M</sup>-pEGFP-N1 allows to express DG with a Green Fluorescent Protein (GFP) fused at the C-terminus of  $\beta$ -DG. Moreover, a Myc tag is present, inserted after Lys498, within the C-terminal domain of  $\alpha$ -DG [19]. All constructs were verified by automated sequencing.

## Fusion protein expression and purification

The recombinant  $\alpha$ -DG(50–313)R166H T190M (hereinafter T190M) fusion protein was expressed in *Escherichia coli* BL21(DE3) Codon Plus RIL strain and purified using nickel affinity chromatography. The fragment of interest was obtained upon thrombin cleavage. Further purification steps were carried out using anion exchange and gel filtration chromatography. Namely, after thrombin cleavage, the flow through of a HiTrap Chelating column (GE Healthcare) was applied onto a Hi-Trap Q HP column (GE Healthcare) pre-equilibrated with buffer A (25 mM Tris-HCl pH 7.5). T190M was eluted with a linear gradient of 0–0.5 M NaCl in buffer A. The fractions containing T190M were pooled, concentrated with Amicon Ultra 15 (Millipore) and loaded on Superdex 200 10/300 GL (GE Healthcare) pre-equilibrated with 25 mM Tris-HCl pH 7.5, 0.15M NaCl, at a flow rate of 0.4 mL/min: the core fractions of the peak were supplemented with 2.5% Glycerol and concentrated by Microcon GM10 (Millipore). The purity of the protein was confirmed by Tricine/SDS-PAGE [20].

The possibility that the methyl sulfide (thioether) of Met 190 would be oxidized in T190M, or in general that the recombinant protein would harbor some additional undesired modifications, has been ruled out via mass spectrometry analysis exploiting an Orbitrap instrument (Thermo Scientific). The corresponding mass value obtained also confirmed the presence of a disulfide bridge between Cys180 and Cys262.

Per quanto non direttamente legate all'analisi strutturale, questa parte del lavoro può fornire utili indicazioni su alcuni aspetti della struttura:

- Presenza o meno di un tag di purificazione
- Qualità della purificazione
- Giustificare la presenza nella struttura di ioni/ligandi eventualmente già presenti nei buffer utilizzati
- Escludere o meno che talune trasformazioni (bio)chimiche siano avvenute
- ...

# Cristallizzazione e acquisizione dati

La conoscenza delle condizioni di cristallizzazione può spiegare la presenza di determinate specie chimiche nel modello finale. Lo stesso può accadere per il crioprotettore usato per il congelamento dei cristalli. La conoscenza del pH può avere una certa rilevanza sul senso della struttura. Se riportato, il contenuto di acqua (Coefficiente di Matthews) può essere utile per comprendere alcune proprietà della struttura (fattori termici elevati, per esempio)

Crystals of T190M were grown by using the vapor diffusion hanging drop method, following the protocol used for the crystallization of WT [13]. Drops were prepared by mixing 1  $\mu\text{L}$  of the protein solution (5.25 mg/mL in 25 mM Tris, 150 mM NaCl and 2.5% glycerol; pH 7.5) with 1  $\mu\text{L}$  of the precipitant solution (0.8 M citrate buffer; pH 7.0) and equilibrated against the reservoir (1 mL) at 4°C.

Data collection was carried out at the European Synchrotron Radiation Facility Grenoble (France), ID23-1 beamline, using a detector Pilatus-6M (Dectris) and the wavelength of 1.00 Å. Data collection was carried out at -173°C. Before being exposed to the X-ray beam, crystals were quickly dipped in cryoprotectant solution (25% ethylene glycol, 1.0 M citrate buffer, pH 7.0) and flash frozen in liquid nitrogen.

Indexing, integration and data reduction of the diffraction data were carried out by using a combination of XDS [21] and CCP4 [22] programs. Data reduction statistics are reported in Table 1.

## Data collection

For cryoprotection, crystals were soaked for a few seconds in a reservoir solution containing 20% (wt/vol) glycerol. The crystals were mounted in a cryoloop and subsequently flash-frozen in liquid nitrogen. X-ray data were collected at 100 K on beamline ID29 at the European Synchrotron Radiation Facility (Grenoble, France). A native data set was collected to 2.6 Å resolution. The crystal belongs to space group  $H32$ , with unit cell parameters  $a=130.2$  Å,  $b=130.2$  Å, and  $c=306.2$  Å. There are two molecules in the asymmetric unit that have a  $V_M$  of  $2.9 \text{ \AA}^3 \text{ Da}^{-1}$  and a solvent content of 58%.<sup>27</sup> Crystals of SeMet-EstA showed a well-defined Se K absorption edge by fluorescence scanning. A single SeMet-EstA crystal was used for MAD data collection at the peak (0.9791 Å), inflection (0.9793 Å), and remote (0.9557 Å) wavelengths up to 2.6 Å resolution. Data were indexed and integrated with MOSFLM<sup>28</sup> and scaled using SCALA.<sup>29</sup>

La conoscenza della lunghezza e delle condizioni di acquisizione possono risultare utili per spiegare alcune caratteristiche delle mappe di densità elettronica (qualità generale, danno da radiazione)



# Determinazione e raffinamento della Struttura

Il dettaglio sul metodo di risoluzione strutturale può essere importante.

Ad esempio a questo livello viene riportato **quale e quanta parte della proteina studiata è stata inclusa nel modello finale**. Sono in genere descritte le parti del modello finale non identificate in virtù della loro flessibilità e conseguente assenza di mappa 2Fo-Fc .

La descrizione della modalità di raffinamento strutturale è molto importante per capire:

- ***Quale modello di coordinate è stato utilizzato***
- ***Quale modello di fattori termici è stato utilizzato***
- Le modalità di raffinamento in relazione alla risoluzione

Sono anche illustrati dettagli sulla **qualità** e la **validazione** del modello.

## Structure determination and refinement

The EstA structure was solved by MAD phasing with MAD data from the SeMet-EstA crystal and the native data set, using HKL2MAP.<sup>30</sup> Eight selenium sites in the asymmetric unit of the crystal were found and used to calculate phases to 2.6 Å resolution (Table 2). However, initial density maps were of generally poor quality and not suitable for tracing the structure. Phases were improved using RESOLVE,<sup>31,32</sup> allowing the identification of NCS. NCS averaging and solvent flattening were performed using the program DM<sup>33</sup> of the CCP4 suite, giving an electron density map of better quality. Autobuilt models from RESOLVE and ARP/wARP<sup>34</sup> were combined to give a starting model comprising 170 residues from a total of 380, with only 20 residues assigned into sequence properly. The structure was then manually rebuilt in Coot<sup>35</sup> and refined using CNS<sup>36</sup> and REFMAC.<sup>37</sup> Strict NCS restraints were applied during the earlier stages of the refinement and released at later stages. In the final stages of refinement, solvent molecules were added using ARP-wARP<sup>34</sup> and manually inspected in Coot.<sup>35</sup> A sulfate ion from the reservoir solution was clearly visible at high contour level in the omit map. The final refinement  $R_{\text{work}}$  is 19.7%, and  $R_{\text{free}}$  is 26.7%.

The structure of the EstA-paraoxon complex was determined by molecular replacement, which was performed with MOLREP using the native structure as model. The complex structure was rebuilt in Coot<sup>35</sup> and refined using REFMAC.<sup>37</sup> The  $F_o - F_c$  electron density map and omit density map displayed clear density for paraoxon and were used to assign the head of the paraoxon molecule. No water molecules were picked for this low-resolution data set. The model with paraoxon was refined to a final  $R_{\text{work}}$  of 22.0% and an  $R_{\text{free}}$  of 26.2%.

In the final models (native and paraoxon-complexed crystal structures), residues 16–20 of both molecules and residues 249–250 of chain B were missing because of poor density. The geometry of both models was monitored using PROCHECK,<sup>38</sup> with the native and paraoxon complex models having 85.2% and 82.2% of their residues in the most favored regions of the Ramachandran plot, respectively. Cartoon representations were generated using PyMOL.<sup>8</sup>

# La Tabella 1

La cosiddetta ‘**Table 1**’ contiene una descrizione sintetica dei risultati della determinazione strutturale (cristallografica)

Non ha un formato fisso e può contenere dati in quantità e forma diversa

**Table 2.** Data collection, phasing, and refinement statistics

	Native	SeMet peak	SeMet inflection	SeMet remote	EstA paraoxon complex
<i>Data collection</i>					
Wavelength (Å)	1.0000	0.9791	0.9793	0.9757	1.0000
Resolution range (Å)	50–2.6 (2.74–2.6) <sup>a</sup>	50–2.6 (2.74–2.6)	50–2.6 (2.74–2.6)	50–2.6 (2.74–2.6)	50–3.0 (3.16–3.0)
Space group	H32	H32	H32	H32	H32
Unit cell parameters (Å)					
<i>a</i>	130.2	131.0	131.0	131.0	130.5
<i>b</i>	130.2	131.0	131.0	131.0	130.5
<i>c</i>	306.2	306.8	306.8	306.8	304.5
Observed reflections	173,066	353,388	358,408	358,161	160,768
Unique reflections	31,079	31,295	31,587	31,457	20,405
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
$R_{\text{merge}}^b$	0.087 (0.430)	0.086 (0.498)	0.082 (0.412)	0.075 (0.437)	0.152 (0.643)
$\langle I/\sigma(I) \rangle$	14.2 (3.6)	22.3 (4.6)	22.1 (4.4)	24.8 (5.0)	13.4 (3.0)
Redundancy	5.6 (5.7)	11.4 (11.7)	11.3 (11.4)	11.4 (11.3)	7.9 (8.1)
<i>MAD phasing</i>					
Number of Se sites		8			
Figure of merit		0.5			
<i>Refinement</i>					
Resolution range (Å)	20–2.6 (2.66–2.6)				35–3.0 (3.08–3.0)
Number of reflections	28,373				19,311
$R_{\text{work}}$ (%)	19.7 (31.2)				22.0 (30.7)
$R_{\text{free}}$ (%)	26.7 (35.5)				26.2 (35.6)
Average <i>B</i> -factors					
Protein	36.8				45.8
Water	33.2				—
Ligand	49.5				39.6
RMSD					
Bond lengths (Å)	0.018				0.016
Bond angles (°)	1.88				1.77

<sup>a</sup> Values in parentheses correspond to the highest-resolution shell.

<sup>b</sup>  $R_{\text{merge}} = \sum_h \sum_l |I_{hl} - \langle I_h \rangle| / \sum_h \sum_l I_{hl}$ , where  $I_l$  is the  $l$ th observation of reflection  $h$ , and  $\langle I_h \rangle$  is the weighted average intensity for all observations  $l$  of reflection  $h$ .

# La Tabella 1 – Dati Cristallografici

Sono sempre riportati

- Gruppo spaziale
- Cella unitaria
- Risoluzione (d)
  
- Inoltre sono riportati dati importanti per valutare la qualità della struttura
- Completezza
- Ridondanza
- $I/\sigma(I)$

Di particolare importanza per valutare la qualità dei dati è l'**R-merge**, che indica l'accordo tra fattori di struttura uguali per simmetria (equivalenti). In genere ha un valore inferiore al 0.2 (20%)

	Native
<i>Data collection</i>	
Wavelength (Å)	1.0000
Resolution range (Å)	50–2.6 (2.74–2.6) <sup>a</sup>
Space group	H32
Unit cell parameters (Å)	
<i>a</i>	130.2
<i>b</i>	130.2
<i>c</i>	306.2
Observed reflections	173,066
Unique reflections	31,079
Completeness (%)	100.0 (100.0)
$R_{\text{merge}}^b$	0.087 (0.430)
$\langle I/\sigma(I) \rangle$	14.2 (3.6)
Redundancy	5.6 (5.7)

In parentesi ci sono i valori calcolati per un gruppo di riflessi selezionati in un intervallo di risoluzione più alta.

# La Tabella 1 – Fasi sperimentali

Se le fasi sono ottenute per via sperimentale, sono riportati ***i dati sperimentali relativi alle ulteriori acquisizioni*** (in questo caso MAD).

Inoltre sono riportati informazioni (numero di derivati o diffusori anomali) e indicatori di qualità sul procedimento di ‘fasatura’ (in questo caso la figura di merito)

	SeMet peak	SeMet inflection	SeMet remote
<i>Data collection</i>			
Wavelength (Å)	0.9791	0.9793	0.9757
Resolution range (Å)	50–2.6 (2.74–2.6)	50–2.6 (2.74–2.6)	50–2.6 (2.74–2.6)
Space group	H32	H32	H32
Unit cell parameters (Å)			
<i>a</i>	131.0	131.0	131.0
<i>b</i>	131.0	131.0	131.0
<i>c</i>	306.8	306.8	306.8
Observed reflections	353,388	358,408	358,161
Unique reflections	31,295	31,587	31,457
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
$R_{\text{merge}}^b$	0.086 (0.498)	0.082 (0.412)	0.075 (0.437)
$\langle I/\sigma(I) \rangle$	22.3 (4.6)	22.1 (4.4)	24.8 (5.0)
Redundancy	11.4 (11.7)	11.3 (11.4)	11.4 (11.3)
<i>MAD phasing</i>			
Number of Se sites	8		
Figure of merit	0.5		

Questa struttura è stata determinata con MAD eseguito su una proteina ‘sostituita’ con Seleno-Metionina

# La Tabella 1 - Raffinamento

Sono riportati in modo sintetico alcuni dati importanti per l'ottimizzazione come la risoluzione e il numero di fattori di struttura utilizzati.

Di grande importanza sono i valori di **R-work** e **R-free**, indicatori di qualità dell'intero procedimento di raffinamento strutturale. In genere l'R-work è minore del 30%, l'R-free del 35%.

Anche in questo caso sono calcolati i valori per un gruppo di fattori di struttura selezionati alla risoluzione più elevata (indicati in parentesi).

<i>Refinement</i>	
Resolution range (Å)	20–2.6 (2.66–2.6)
Number of reflections	28,373
$R_{\text{work}}$ (%)	19.7 (31.2)
$R_{\text{free}}$ (%)	26.7 (35.5)
Average <i>B</i> -factors	
Protein	36.8
Water	33.2
Ligand	49.5
RMSD	
Bond lengths (Å)	0.018
Bond angles (°)	1.88

---

# La Tabella 1 - Validazione

- In Tabella 1 sono in genere riportati anche i valori numerici di alcuni indicatori di qualità:
- RMSD di distanze e angoli
- % di residui favoriti nel Ramachandran plot
- Clashcore (secondo MolProbity) [raro]
- Valori dei fattori termici (non sono veri indicatori di qualità però ci dicono quanto è flessibile la proteina)

RMS (bonds) (Å)	0.004
RMS (angles) (°)	0.72
Ramachandran favored (%)	96.48
Ramachandran allowed (%)	3.52
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	11.15
Average B-factor	31.60
Macromolecules	31.86
Solvent	24.03

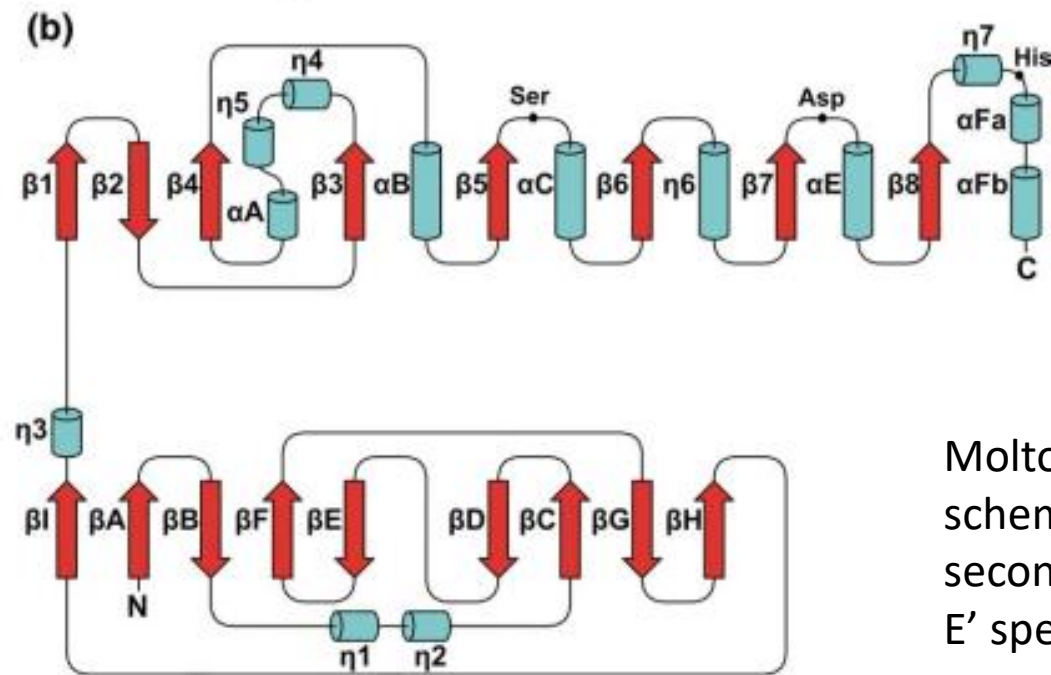
I dati riportati per la frazione di fattori di struttura a risoluzione più alta (es: R-merge, R-free...), rendono conto della validità della scelta (soggettiva) effettuata sul limite massimo di risoluzione.

In un certo senso costituiscono una validazione della risoluzione utilizzata.

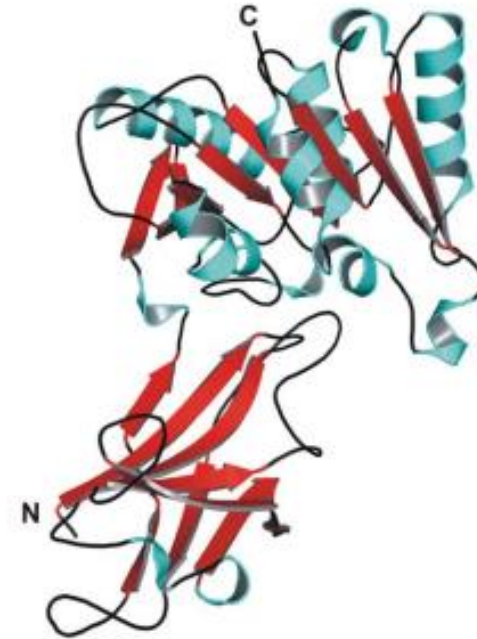
---

# La topologia di una nuova struttura

Se il lavoro cristallografico è rivolto alla determinazione di una nuova struttura, la sua struttura generale viene riportata graficamente



(a)

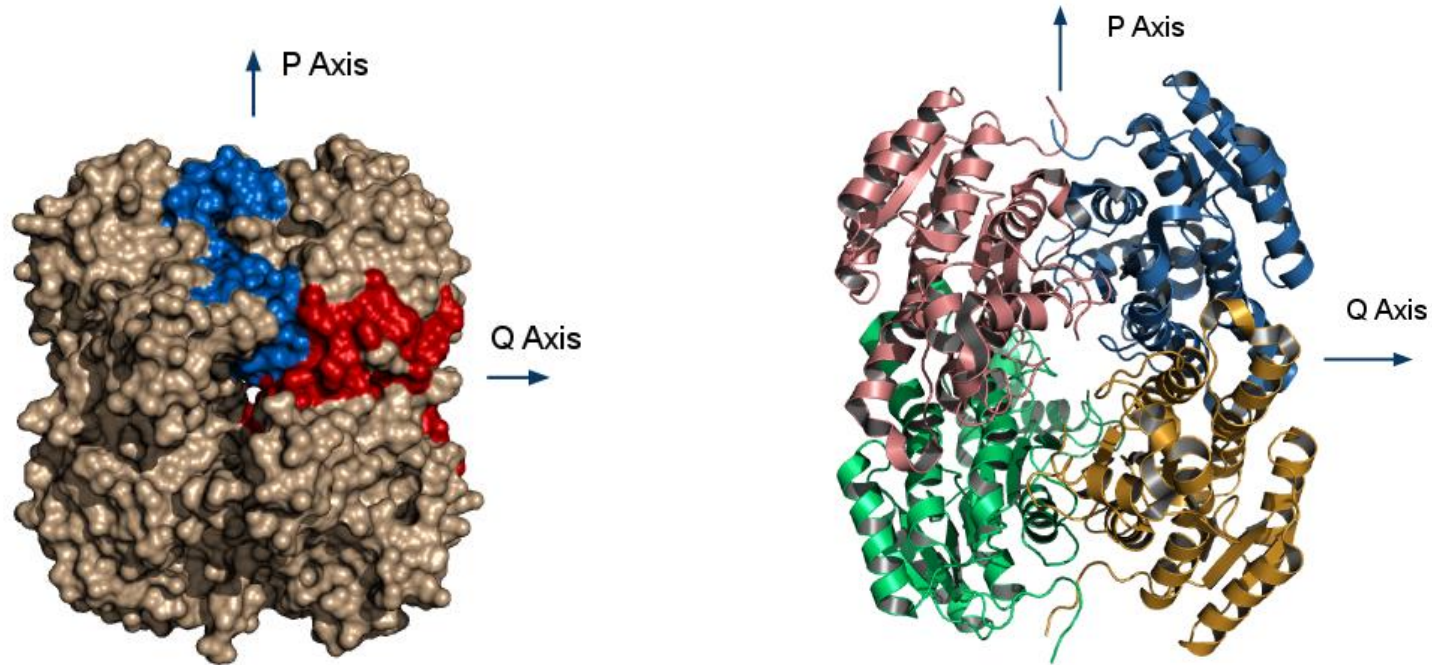


Molto spesso è anche riportata una descrizione grafica schematica, dell'organizzazione degli elementi di struttura secondaria.

E' spesso riportato il fold di riferimento, utilizzando strumenti come **DALI** (<http://ekhidna2.biocenter.helsinki.fi/dali/>)

# Struttura Quaternaria

Se la proteina è dotata di una struttura quaternaria, ovvero è costituita da un insieme di oligomeri, questa organizzazione è descritta nel lavoro e riportata graficamente.

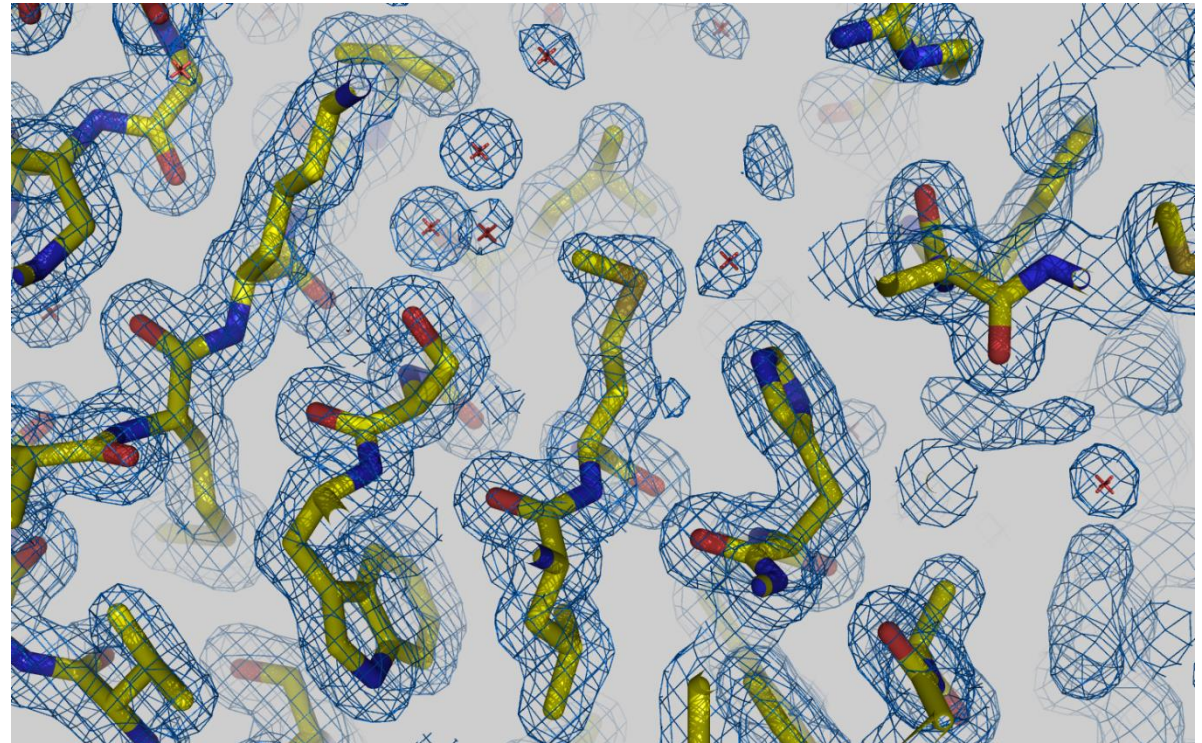




# La qualità della mappa di densità elettronica

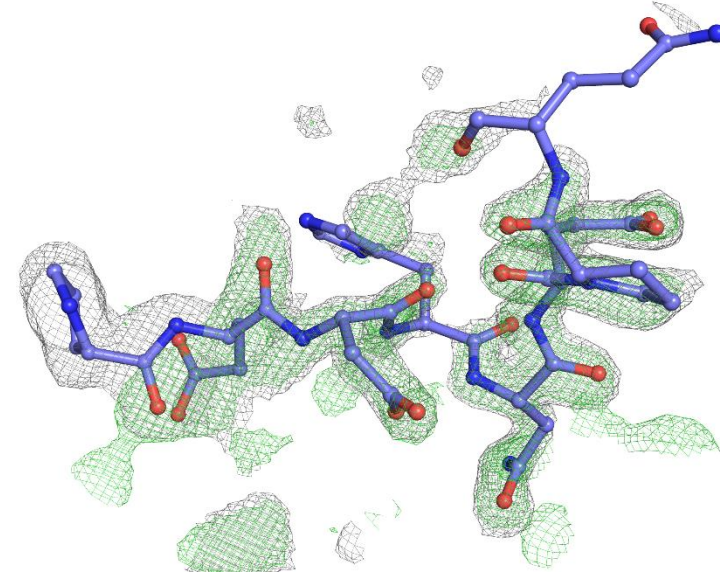
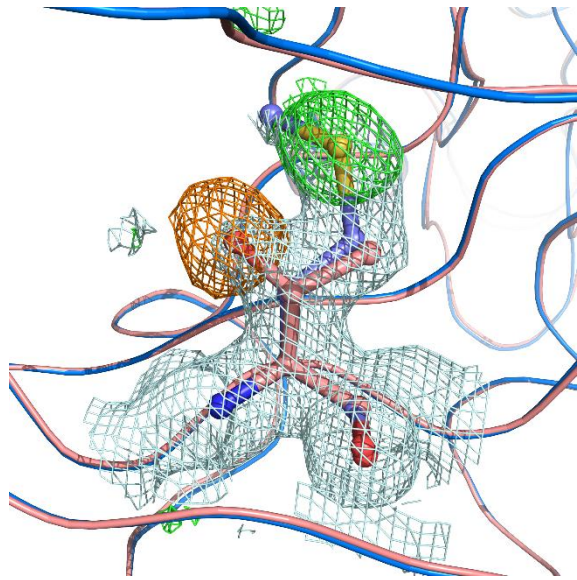
In assenza di accesso diretto alle coordinate e ai fattori di struttura (PDB), la visualizzazione dell'accordo tra modello e mappa  $2F_o-F_c$ , mappa rappresentativa della  $\rho(r)$  (densità elettronica), fornisce l'unica indicazione sull'accordo tra modello ottimizzato e dati sperimentali.

La densità viene spesso riportata per le zone di maggiore interesse biologico (sito attivo, siti di binding...)



# La qualità della mappa di densità elettronica

La mappa 2Fo-Fc (e anche la Fo-Fc) sovrapposta al modello è anche utilizzata per mettere in evidenza criticità nel modello, come ad esempio la flessibilità di alcune parti della proteina (mappa 2Fo-Fc debole o assente)



L'utilizzo delle mappe Fo-Fc può invece essere utilizzato per mostrare differenze tra i modelli.

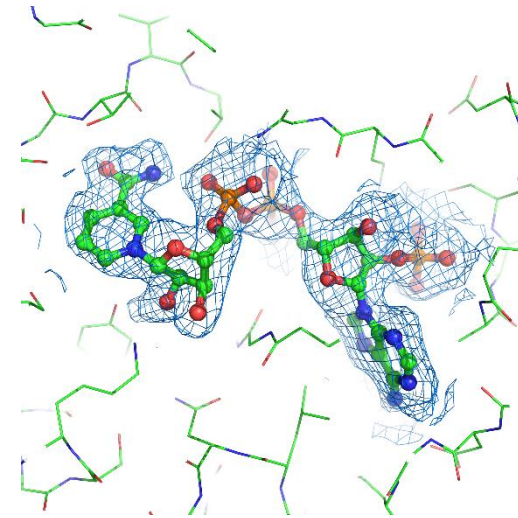
Nel caso indicato abbiamo una mutazione Thr→Met, utilizzando il modello WT (con Thr) la mappa Fo-Fc dimostra che il gruppo OH della Thr (picco rosso, negativo) non è presente, mentre è invece presente il gruppo S-CH<sub>3</sub> (picco verde, positivo).

# La qualità del ligando (omit map)

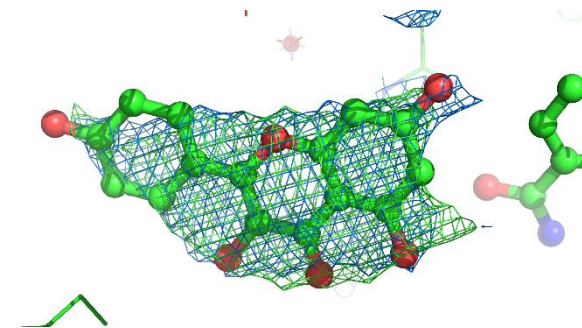
Se nel lavoro è discussa l'interazione con un ligando, viene spesso riportata in modalità grafica il modello finale del ligando stesso sovrapposto alla sua mappa  $2F_o - F_c$  calcolata escludendo il contributo degli atomi del ligando medesimo (**omit-map**).

In sostanza nel **calcolo della omit-map**, nel calcolare gli  $F_c$  non includo gli atomi del ligando. Se il ligando è davvero presente, la mappa che definisce il ligando deve essere di qualità accettabile e il modello finale del ligando e la mappa  $2F_o - F_c$  così calcolata devono essere coerenti tra loro.

In genere ligandi fortemente interagenti con la proteina ( $K_d < \mu\text{M}$ ) e in complessi ottenuti per co-cristallizzazione, mostrano omit-map di qualità migliore



$K_d < \text{nM}$



$K_d \sim \mu\text{M}$

# Analisi critica dei risultati

- Se possibile **valutare sempre il modello in relazione ai dati sperimentali** (modello + fattori di struttura/mappe di densità elettronica).
- Valutare bene i dati riportati in Tabella 1 (**qualità generale della struttura**)
- **Analizzare criticamente i Materiali e Metodi** in relazione a quanto dichiarato nell'analisi strutturale
- Valutare con attenzione le figure, anche in relazione all'informazione che vogliono trasmettere
- **Valutare criticamente il rapporto tra qualità della struttura e informazione trasmessa**: la risoluzione della struttura è coerente con il risultato dichiarato?
- Se il lavoro riguarda un complesso proteina-ligando, **valutare se il ligando è effettivamente presente e se la sua posizione è ragionevole** (esistono strumenti di validazione disponibili in rete: <http://www.ruppweb.org/twilight/default.htm>)
- **Non dimenticare mai, nel caso della diffrazione di raggi-X, che mi trovo in un cristallo!** La conformazione della macromolecola potrebbe risultare alterata.