

Rappresentazione e utilizzo di modelli cristallografici

Laurea Magistrale in Biotecnologie Mediche
Curriculum Nanobiotecnologie

A.A. 2022-23

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
50 Years

PDB50: Celebrating 50 Years of the PDB archive
Visit www.pdb.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' field, a 'Password' field, a 'Log in' button, and a 'Forgot Password' button. Below these is a 'Deposition server' button. On the right, a welcome message states: 'Welcome to the wwPDB validation system! 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 message 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)'. Information icons (i) are present next to the Validation ID, Password, and e-mail address fields.

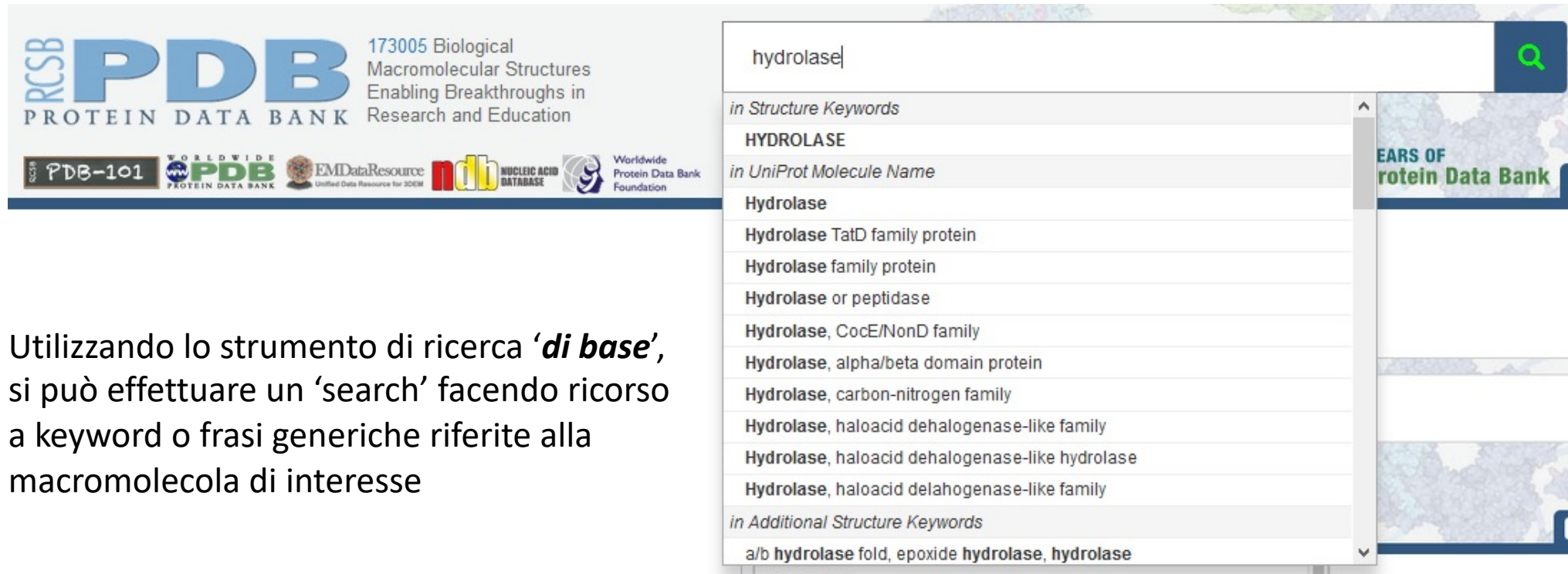
I **‘ligandi’** presenti nella struttura, ovvero tutte quelle componenti chimiche di natura **NON-macromolecolare** (ioni e molecole inorganiche e organiche eccetto l’acqua), sono confrontate con un database dei ligandi già depositati.

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 Basic Search interface. On the left, the PDB logo is accompanied by the text '173005 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education'. Below this, several partner logos are visible: RCSB PDB-101, Worldwide PDB, EMD Data Resource, NDB, and the Worldwide Protein Data Bank Foundation. The main search area features a text input field containing the word 'hydrolase' and a green magnifying glass icon. A dropdown menu is open below the search bar, listing suggestions categorized by search type: 'in Structure Keywords' (HYDROLASE), 'in UniProt Molecule Name' (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, Hydrolase, haloacid dehalogenase-like family), and 'in Additional Structure Keywords' (a/b hydrolase fold, epoxide hydrolase, hydrolase). The background of the interface shows a faint protein structure visualization.

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.

The screenshot displays the 'Advanced Search Query Builder' interface. On the left, there are expandable sections for 'Attribute', 'Sequence', 'Sequence Motif', 'Structure Similarity', and 'Chemical'. The 'Attribute' section is currently expanded, showing a list of searchable fields. A search input field at the top of this list contains the placeholder text '-- Enter and/or select a field name --'. Below the input field, a scrollable list of attributes is visible, including: 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, and Publications All. At the bottom of the interface, there is a footer with navigation links for 'About', 'About Us', 'Help', and 'wwPDB', along with a statement: '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. ...

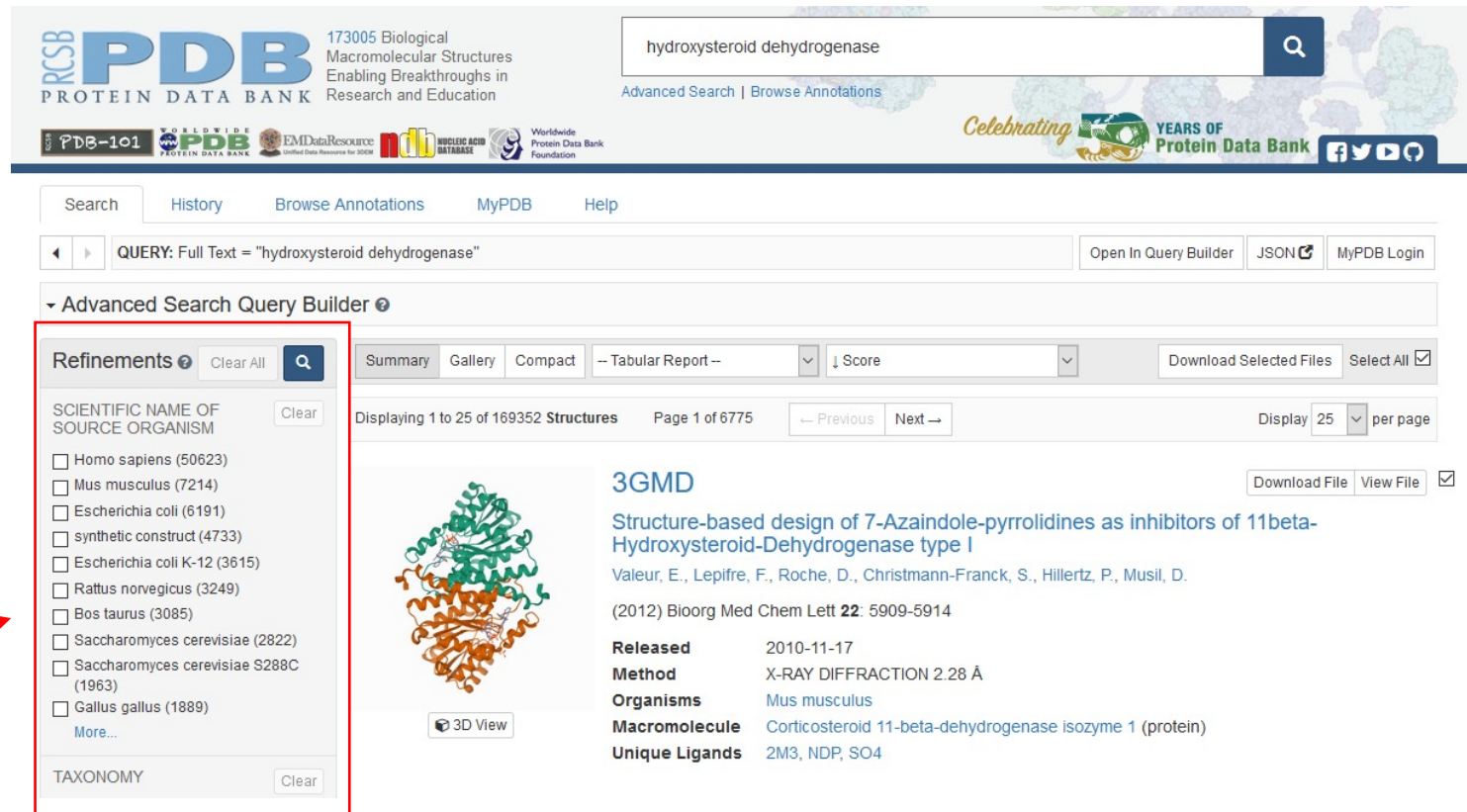
Query Type ? Formula ? ☐ Match Subset ? Count Clear

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).

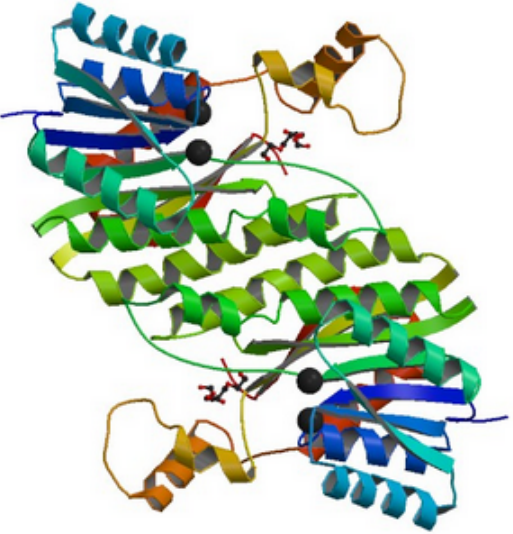


The screenshot displays the PDB search results for the query "hydroxysteroid dehydrogenase". The page includes a search bar at the top with the query text and a search button. Below the search bar, there are navigation links: Search, History, Browse Annotations, MyPDB, and Help. The search results are displayed in a table format, showing the first 25 of 169352 structures. The table includes columns for the structure ID, name, and release date. A sidebar on the left, titled "Refinements", allows users to filter results by various criteria, including "SCIENTIFIC NAME OF SOURCE ORGANISM". This sidebar is highlighted with a red box, and a red arrow points from the text "ulteriori limitazioni" to it. The sidebar lists several organisms with checkboxes: 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), and Gallus gallus (1889). There is also a "More..." link and a "TAXONOMY" section. The main content area shows a 3D model of the protein structure and a summary of the search results, including the number of structures found and the page number.

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

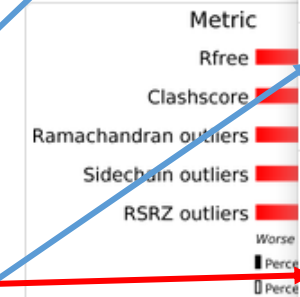
Method: X-RAY DIFFRACTION

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]

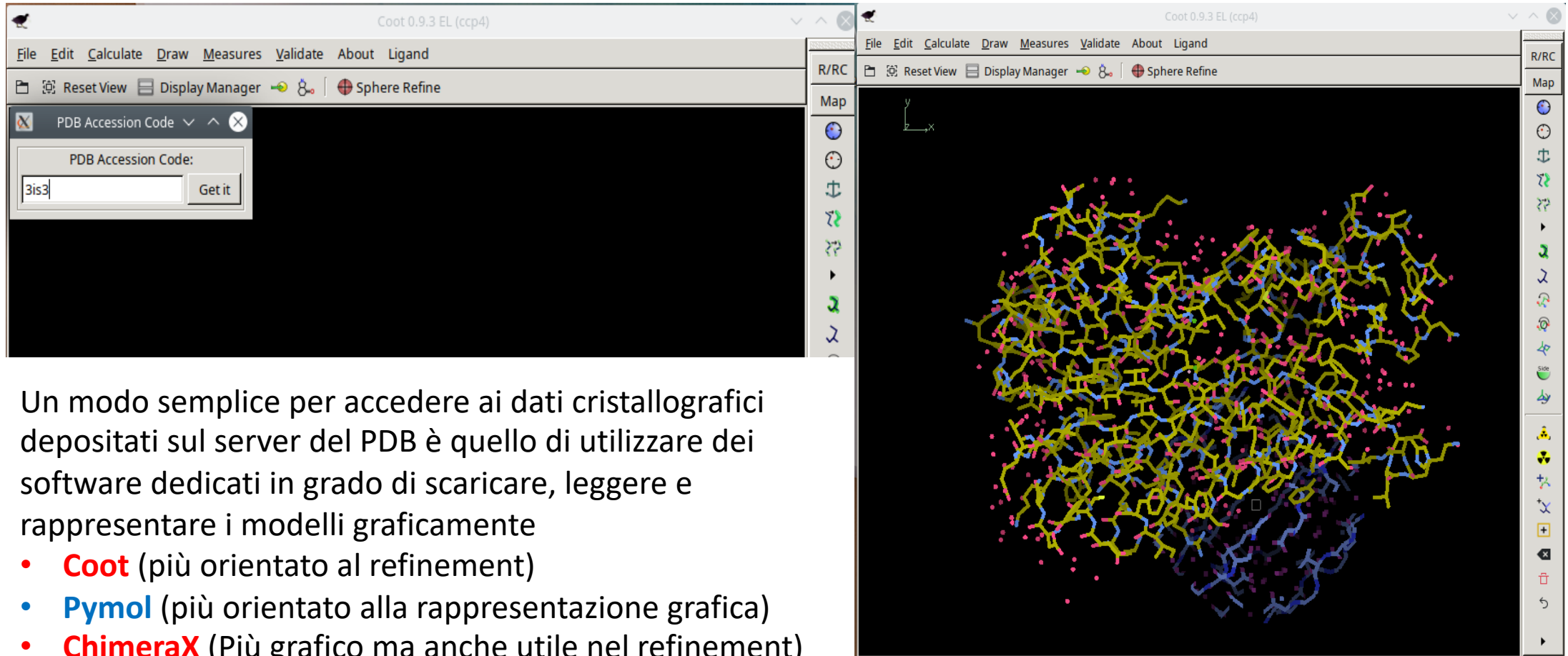
Anisotropo [ellissoide]

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 screenshot displays the Coot 0.9.3 EL (ccp4) interface. On the left, a 'PDB Validation Report for 3IS3' window is open, showing a table of metrics and their percentile ranks. The metrics include Rfree (0.21), Clashscore (2.24), Ramachandran Outliers (0.39), Sidechain Outliers (0.00), and RSRZ Outliers (5.00). Below the table, specific validation details for Residue A 11 TYR and Residue A 12 ILE are shown. On the right, the main Coot window displays a 3D density map of the protein structure, with atoms represented as spheres and bonds as sticks. The map is colored by density, with blue representing the highest density and red representing the lowest. The Coot window also shows a menu bar with options like File, Edit, Calculate, Draw, Measures, Validate, About, and Ligand, and a toolbar with various tools for map manipulation.

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

Worse Better

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

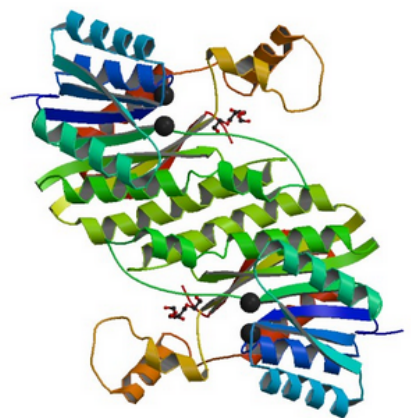
Close

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

Biological Assembly 1 ?



3D View: Structure | Electron Density | Ligand Interaction

Standalone Viewers
Protein Workshop | Ligand Explorer

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

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

Worse  Better

■ 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 17 β -HSD (17 β -hydroxysteroid dehydrogenase) from the filamentous fungus *Cochliobolus lunatus* (17 β -HSDcl) is a NADP(H)-dependent enzyme that preferentially catalyses the interconversion of inactive 17-oxo-steroids and their active 17 β -hydroxy counterparts. 17 β -HSDcl belongs to the SDR (short-chain dehydrogenase/reductase) superfamily. It is currently the only fungal 17 β -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 17 β -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 17 β -HSDcl. Coumestrol, a plant-derived non-steroidal compound with oestrogenic activity, inhibits 17 β -HSDcl [IC₅₀ 2.8 μ M; at 100 μ M 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

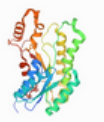
Hide Full Abstract

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	Curvularia lunata	Mutation(s): 0 ⓘ Gene Names: 17HSDcI EC: 1.1.1.62	

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

Protein Feature View [Expand](#)

Reference Sequence: [3IS3_1](#) ▼

0 20 40 60 80 100 120 140 160 180 200 220 240 260

PDB ENTITY SEQ [3IS3_1](#)

UNIPROT ALIGN [O93874](#)

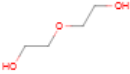
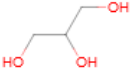
UNMODELED [3IS3.A](#)

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.

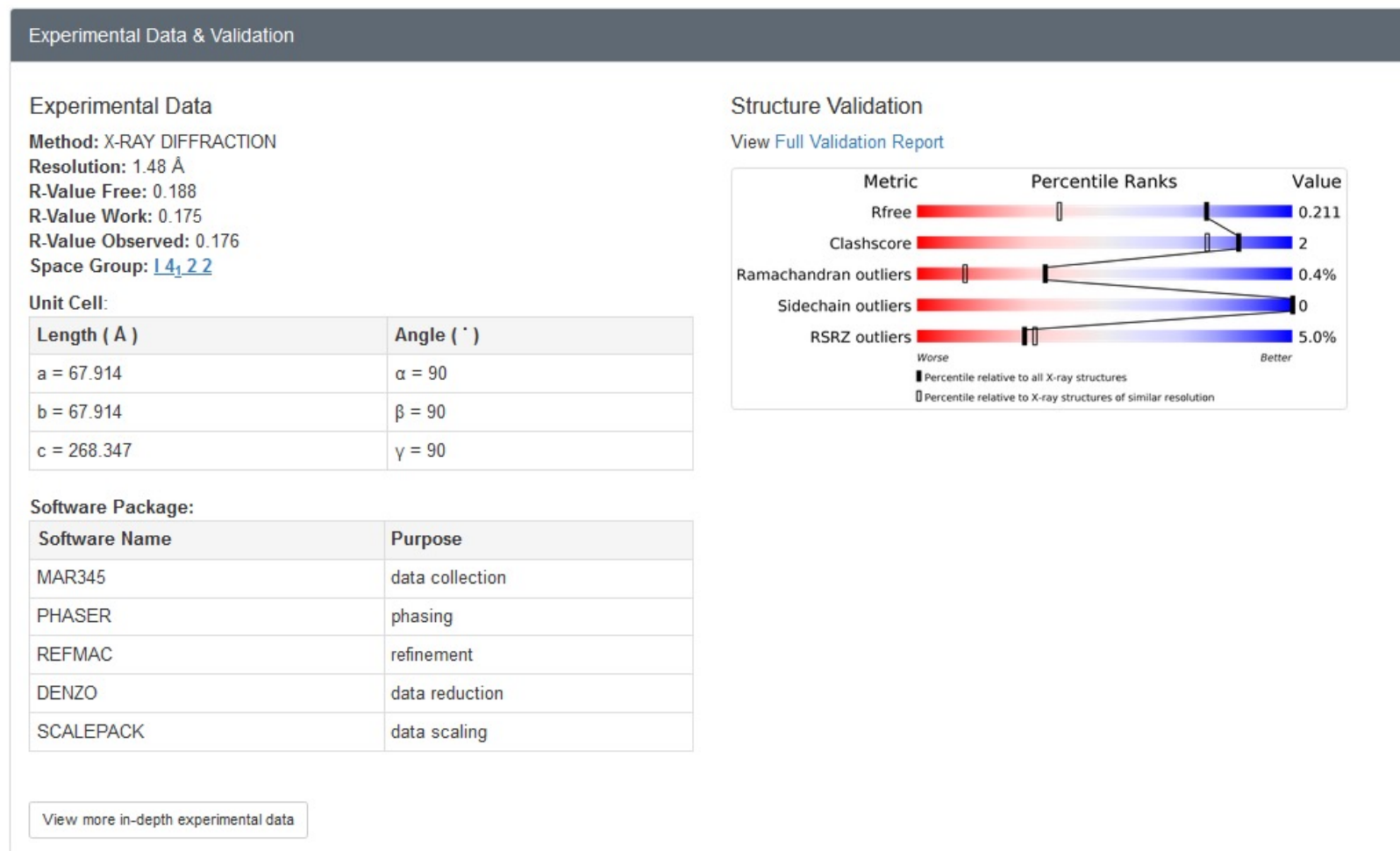
[E' considerato ligando qualsiasi molecola organica o inorganica che non sia parte della proteina, dell'oligonucleotide o che non sia acqua]

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

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	1G0O	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

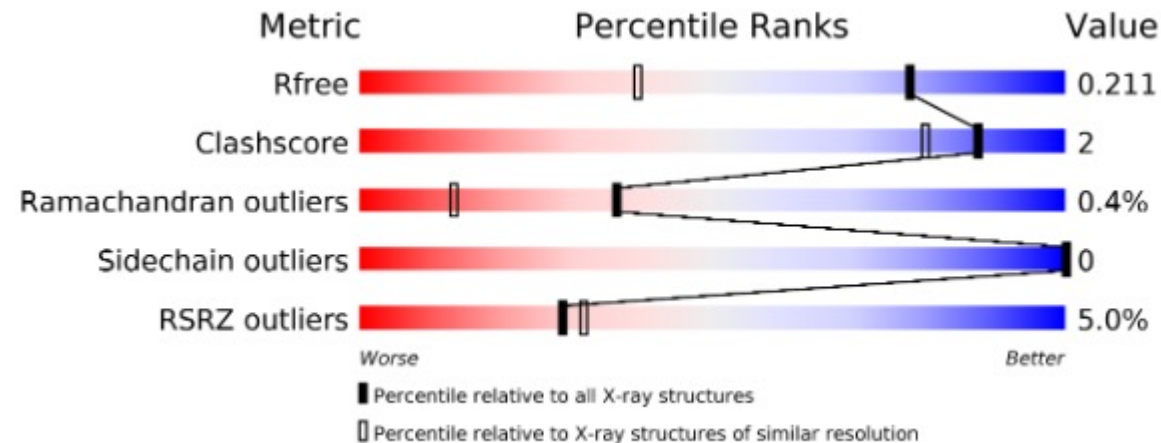


Full wwPDB X-ray Structure Validation Report [i](#)

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

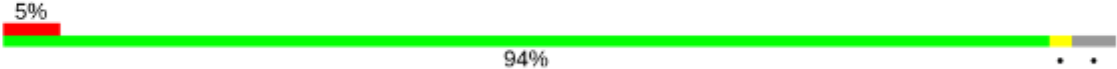
Prima di essere accettata una struttura deve rispondere a dei parametri minimi di qualità.

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

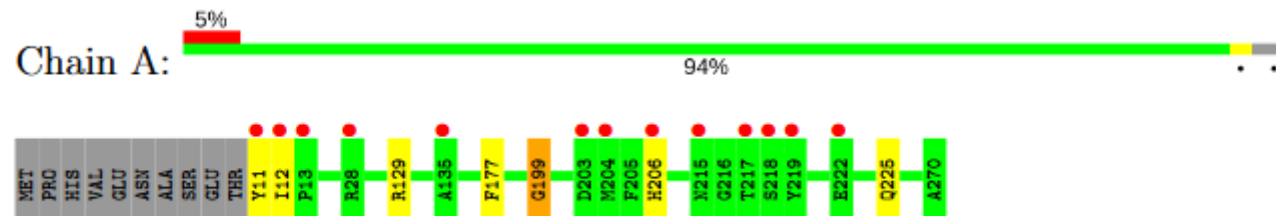


PDB Validazione

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($^{\circ}$)	Ideal($^{\circ}$)
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

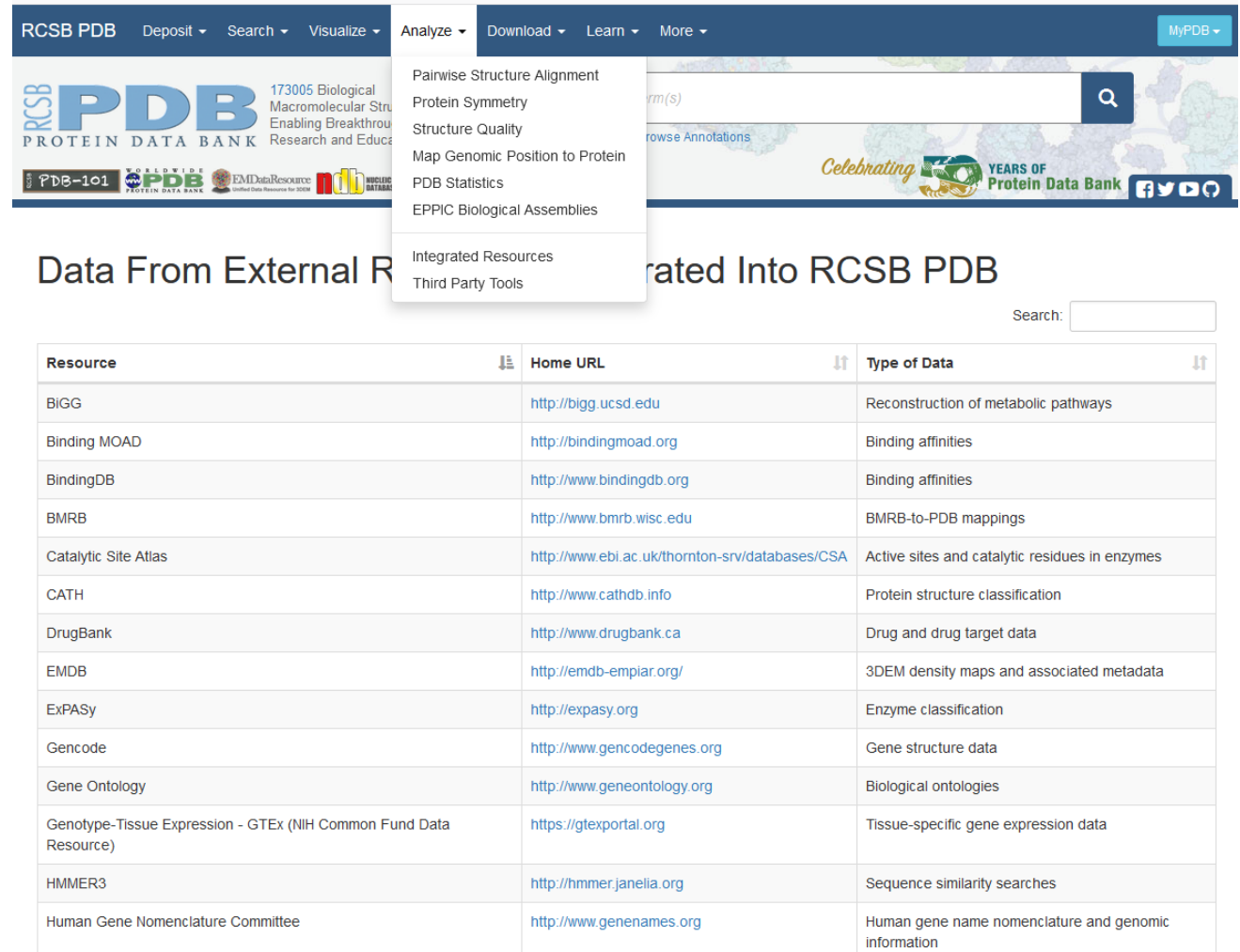
Rappresentazioni grafiche di una macromolecola biologica

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)



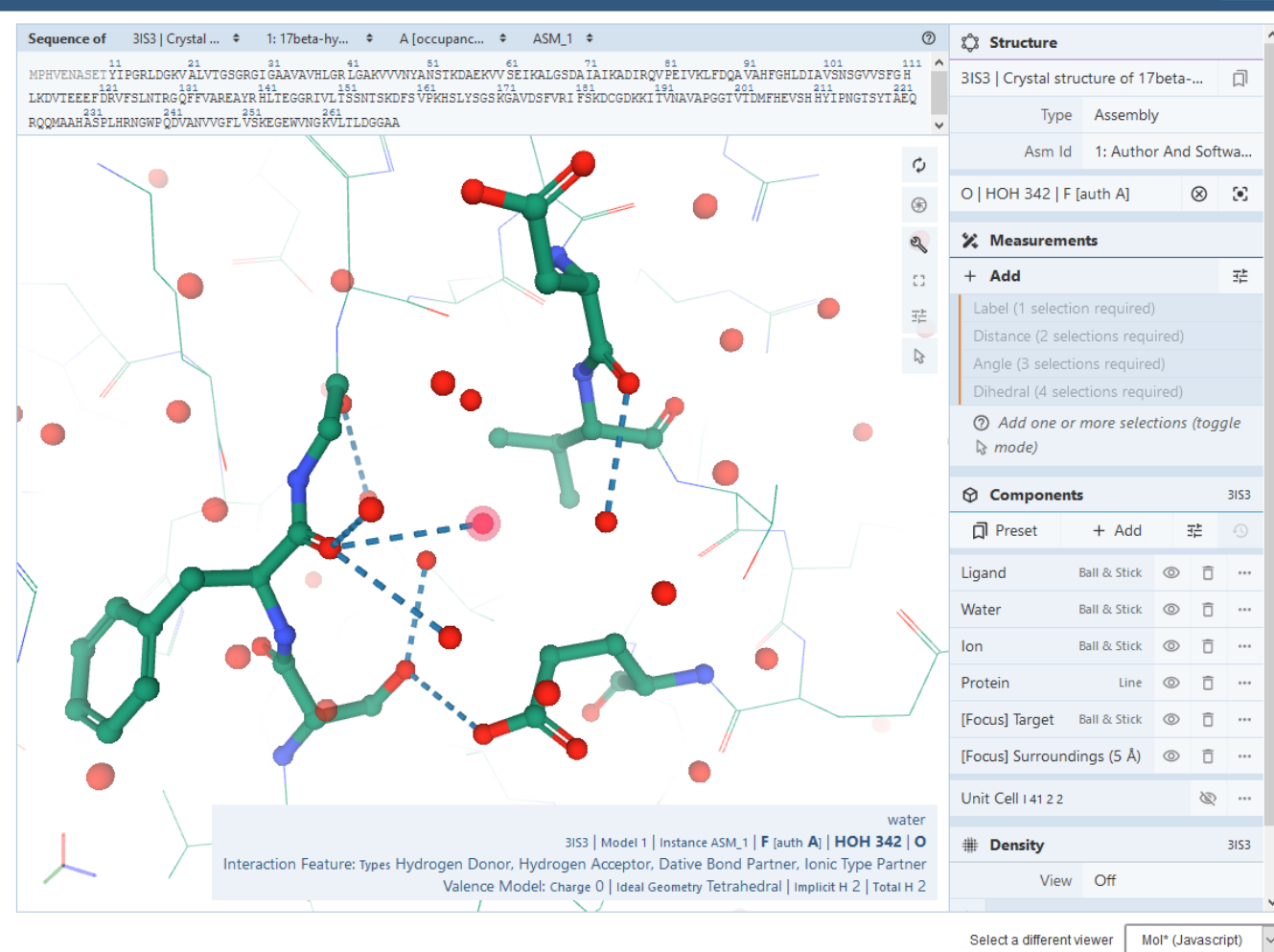
The screenshot shows the RCSB PDB website interface. The 'Analyze' dropdown menu is open, displaying various analysis tools. Below the menu, a table titled 'Data From External Resources Integrated Into RCSB PDB' lists various external resources and their associated data types.

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

PDB – Visualizzazione

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

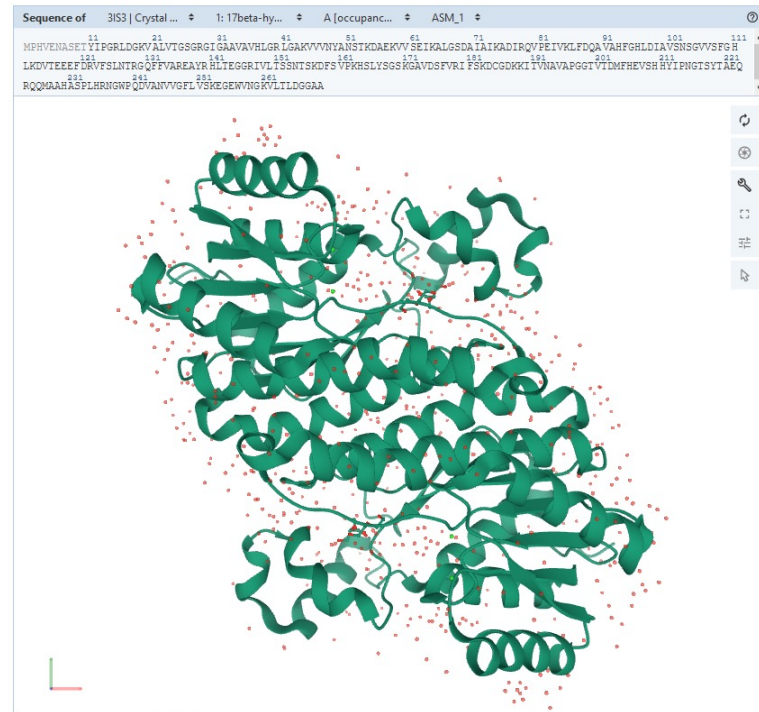
Il software di visualizzazione integrato possiede una certa flessibilità anche se sono preferiti software esterni (Pymol, ChimeraX)



PDB - Visualizzazione

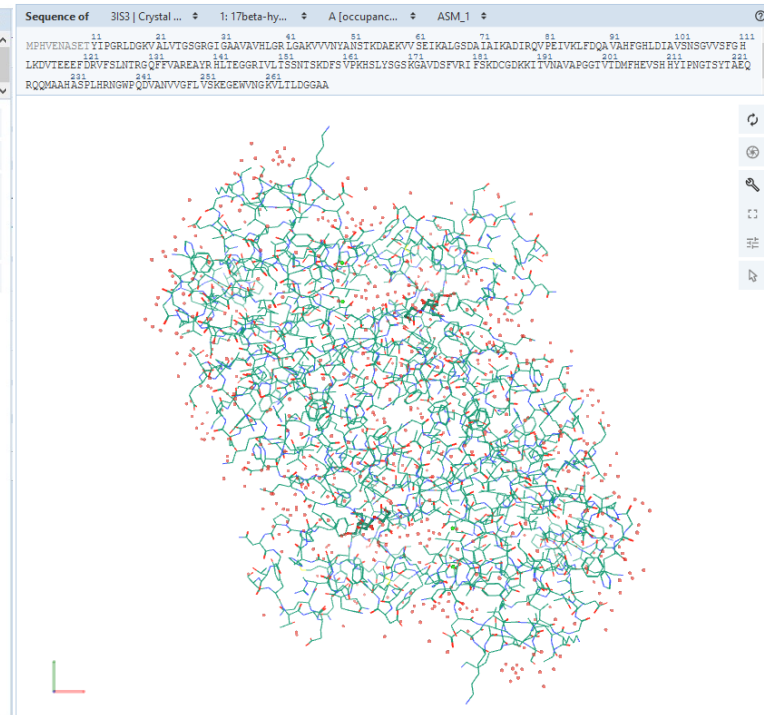
3IS3

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



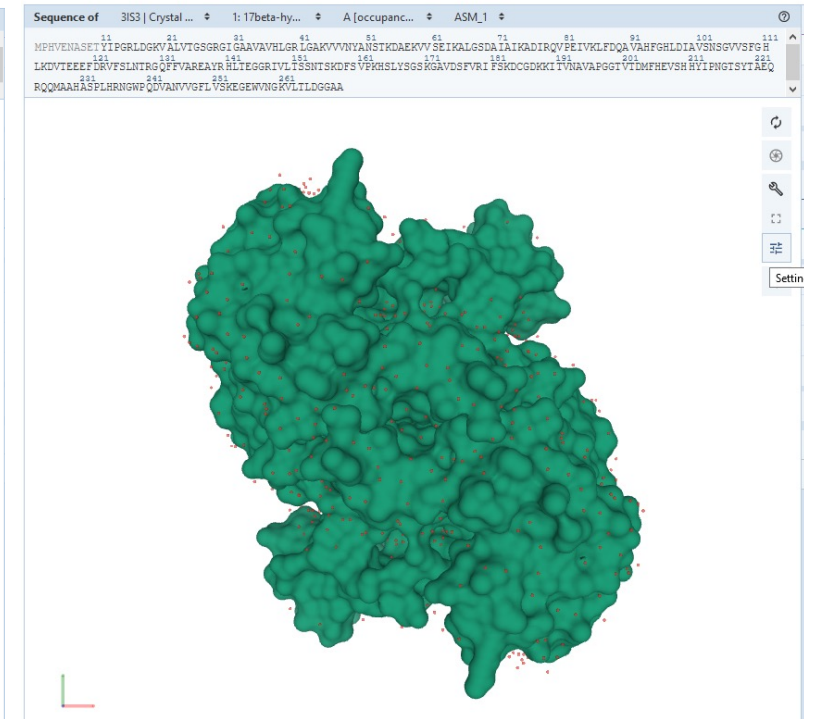
3IS3

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 posso essere rappresentate in vari modi capaci di comunicare **informazione di natura diversa**.

Wireframe

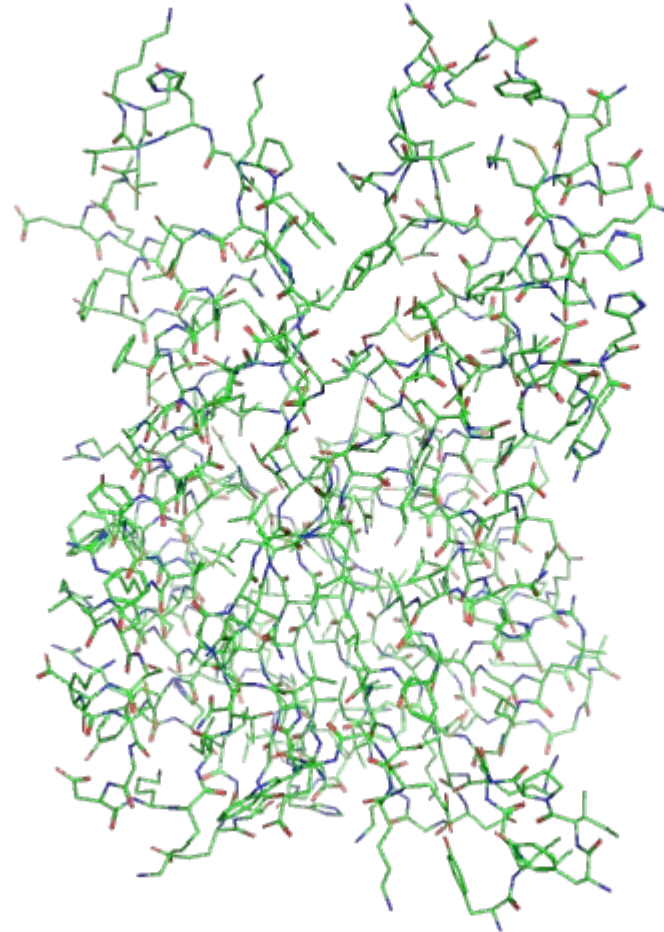
E' una rappresentazione molto diffusa.

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

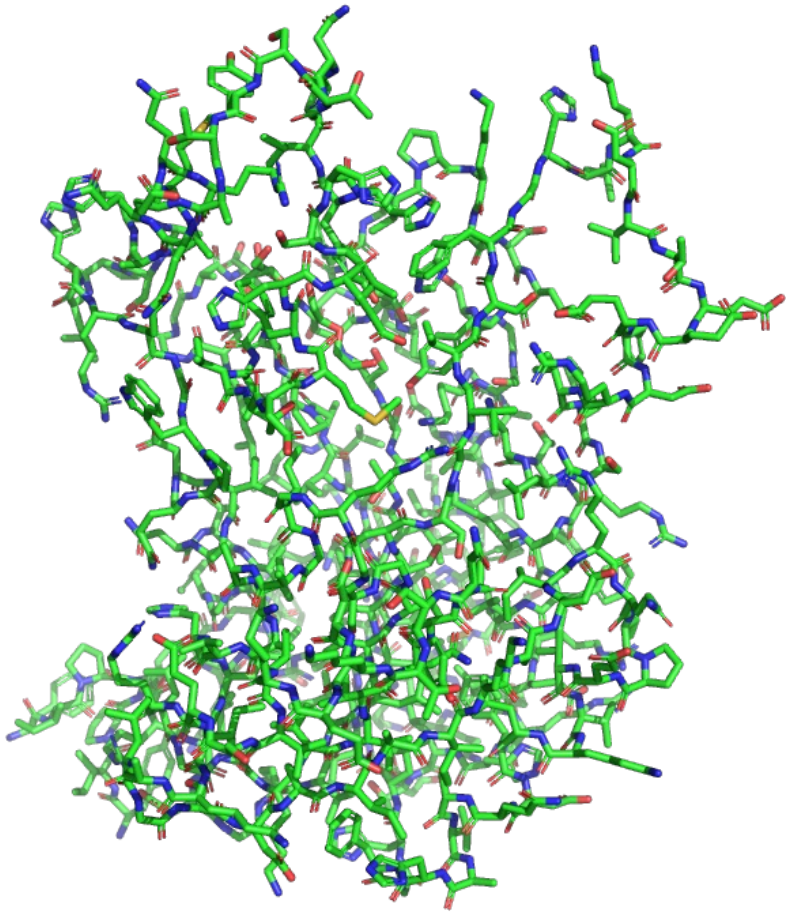
La colorazione dipende dal tipo di proprietà che si intende rappresentare:

- 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'



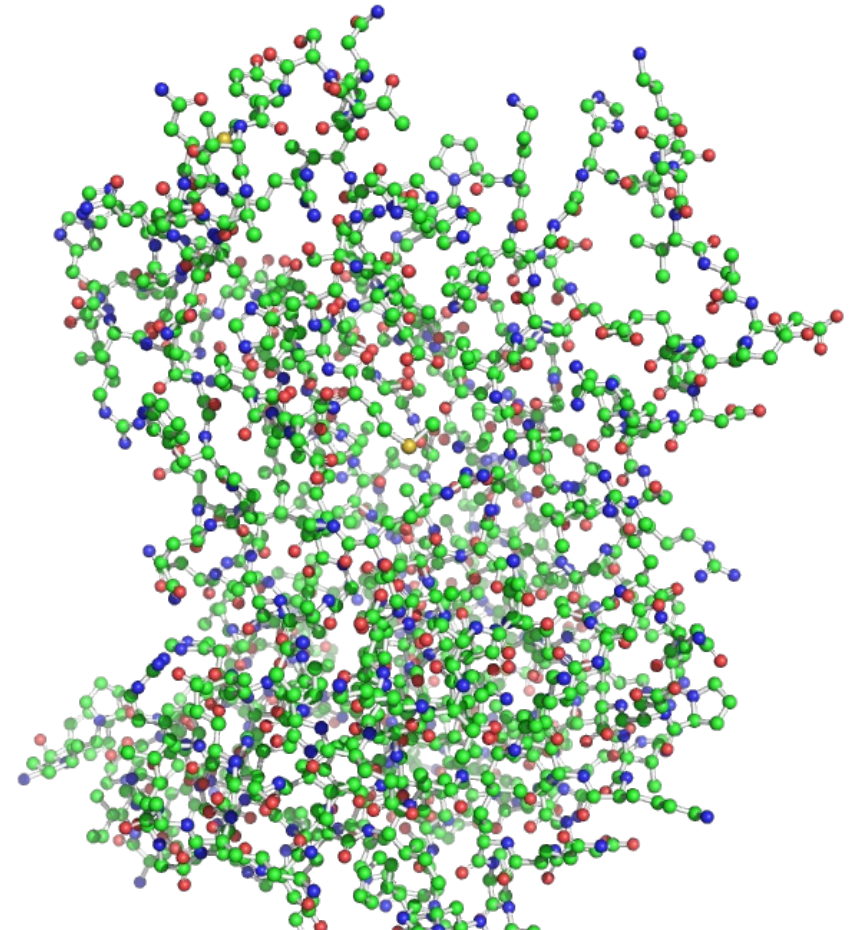
BATON

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.



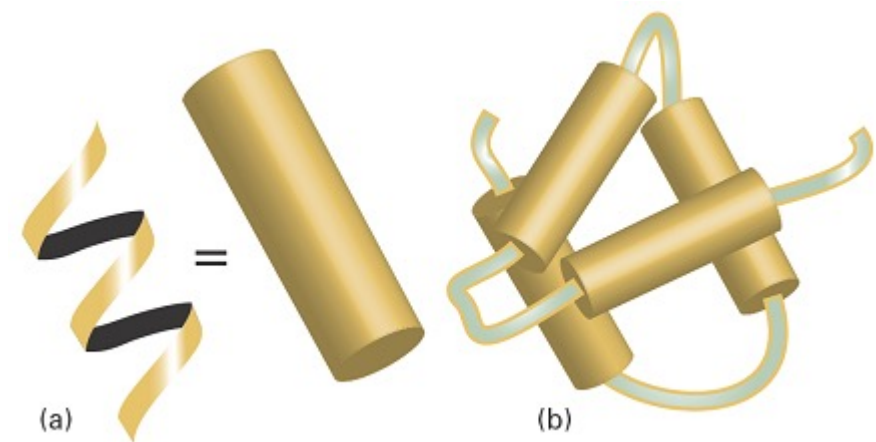
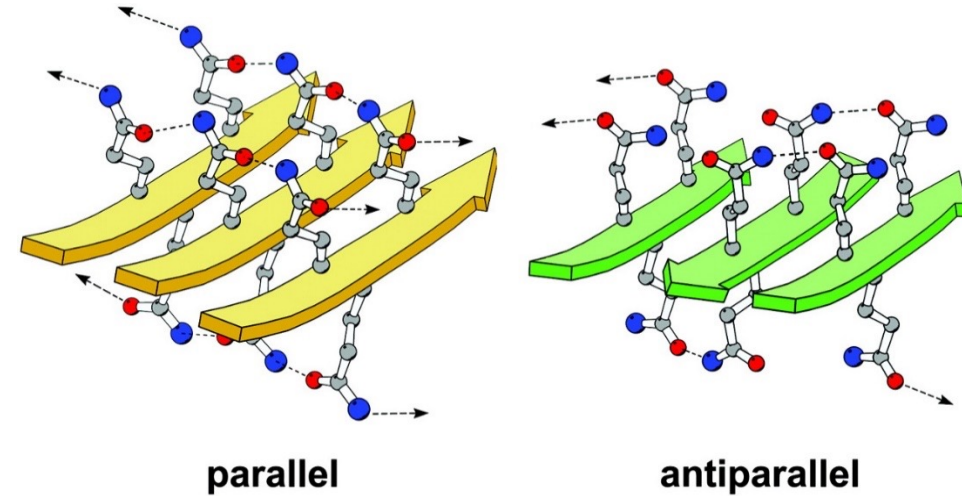
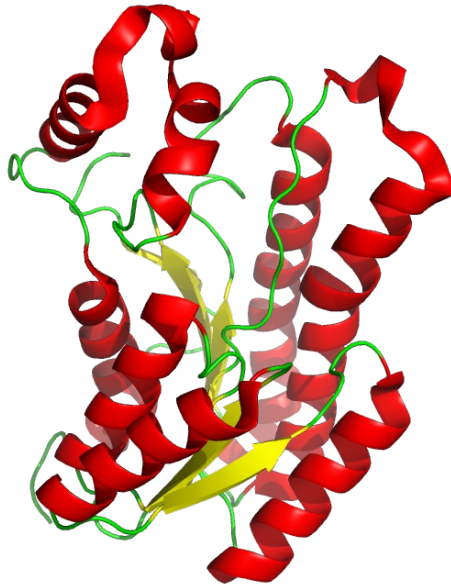
BALL AND STICK

Rappresentazione della struttura – ‘Cartoon’

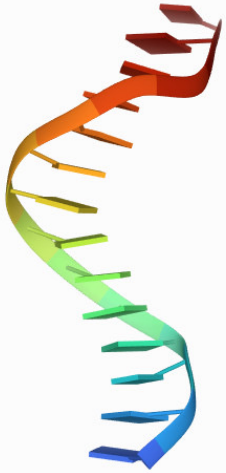
E' la classica rappresentazione del **folding** della macromolecola.

mette in evidenza gli elementi di struttura secondaria:

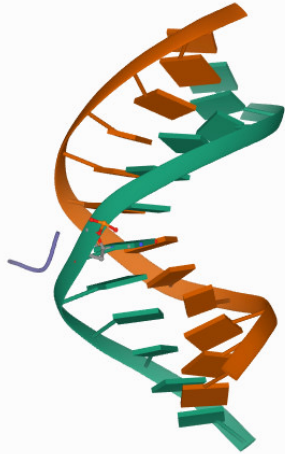
- Eliche (cilindri o spirale)
- Foglietti- β (paralleli o anti-paralleli)
- loops



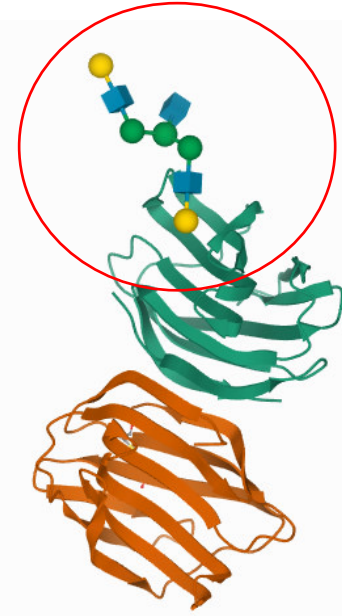
Rappresentazione della struttura – altro



RNA/DNA



Proteina/DNA

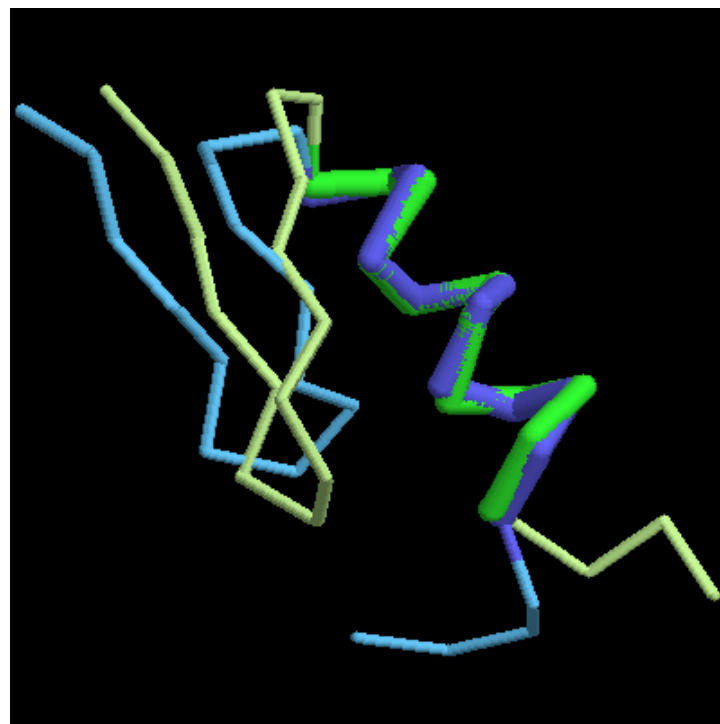


Oligosaccaridi

Rappresentazione della struttura – traccia C α (Backbone)

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

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



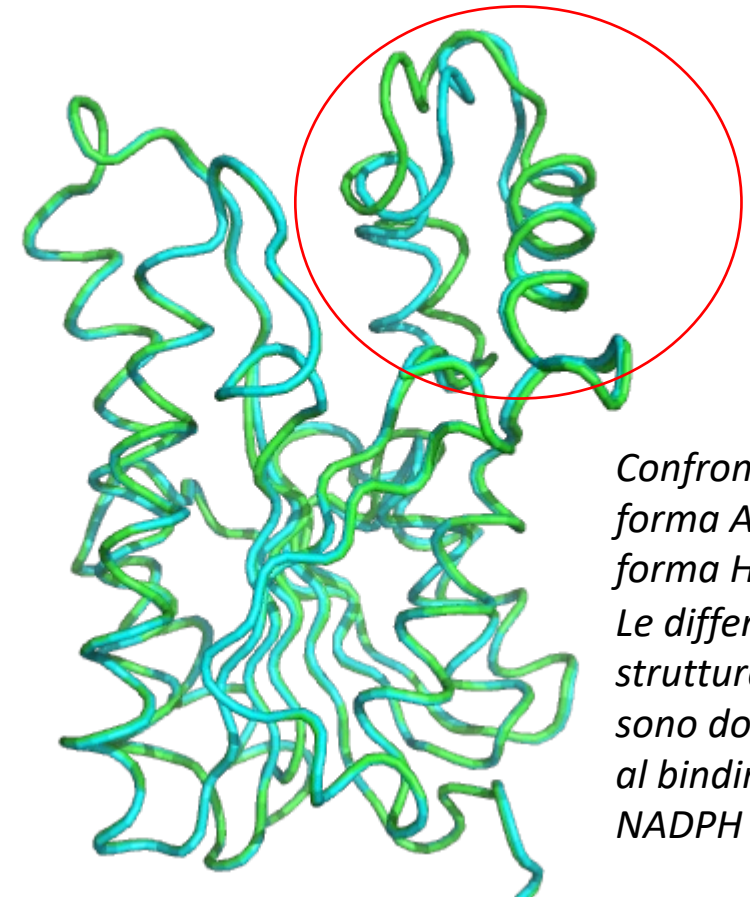
Sovrapposizione di Strutture

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

Un protocollo (automatico) per l'allineamento:

1. I C α 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 α equivalenti (la sommatoria è estesa a tutti i C α 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 α equivalenti e sovrapposti.

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

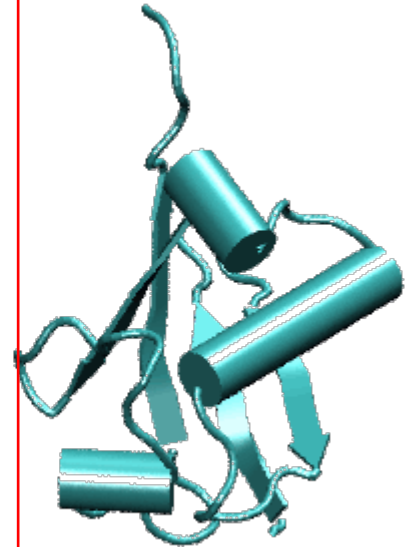
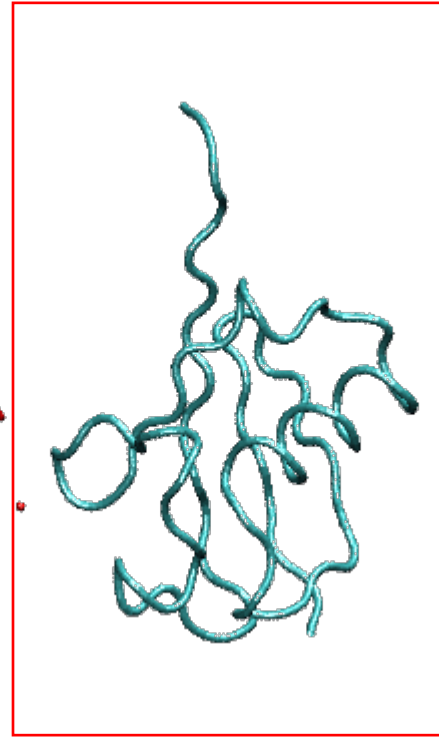
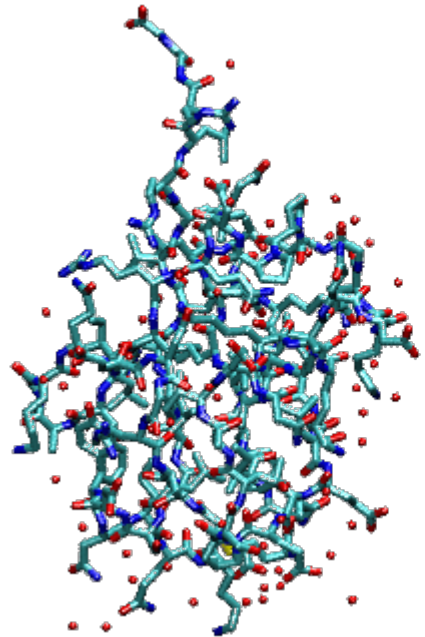


*Confronto tra
forma Apo e
forma Holo.
Le differenze
strutturali
sono dovute
al binding di
NADPH*

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...

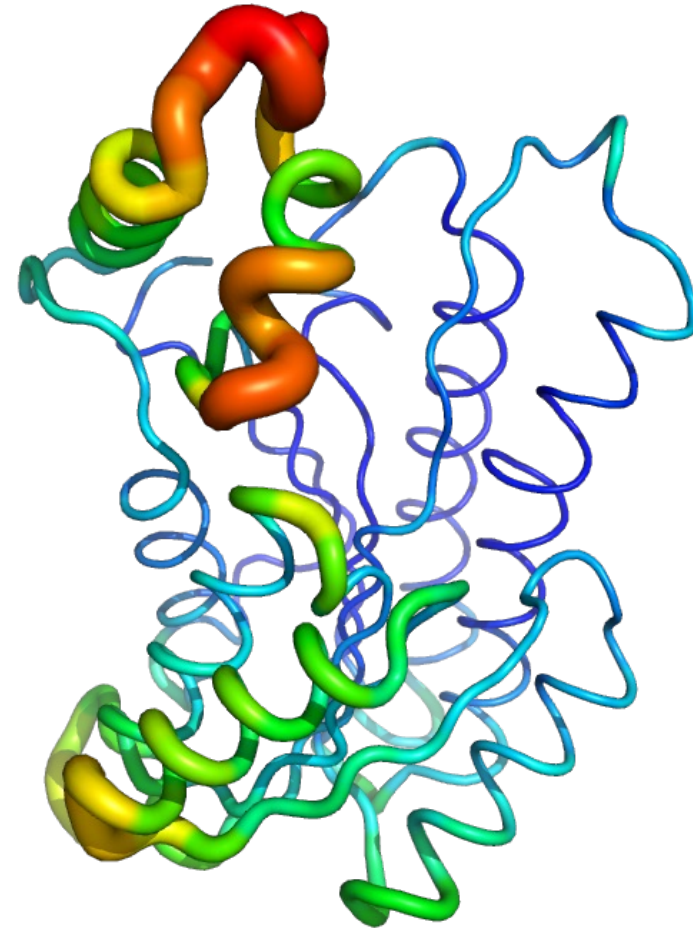


Rappresentazione della struttura – ‘putty’

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...)



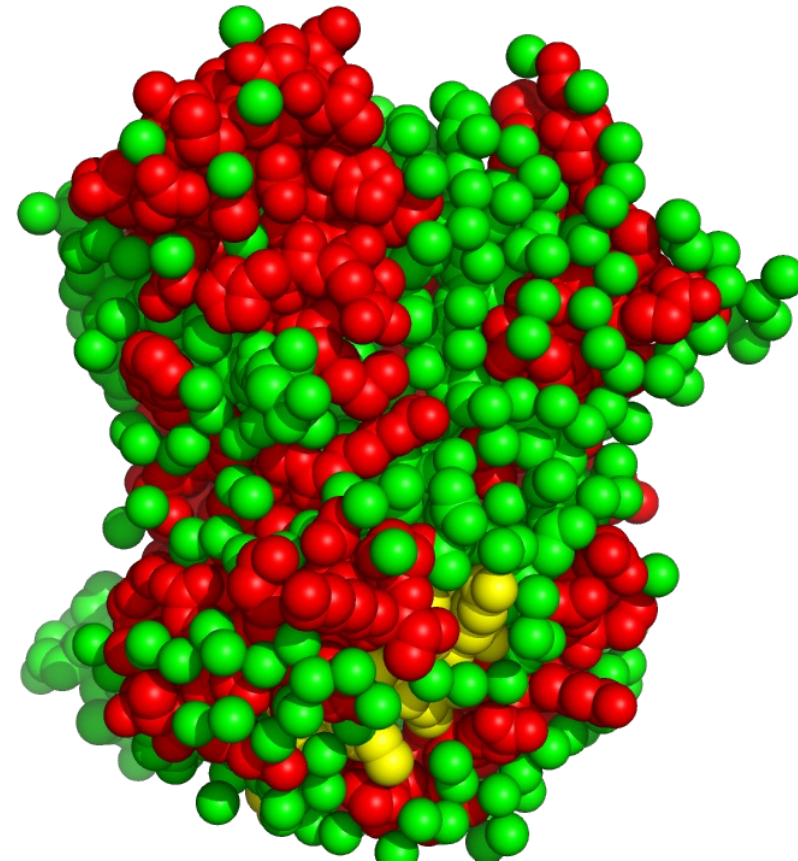
Rappresentazioni Volumetriche

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***

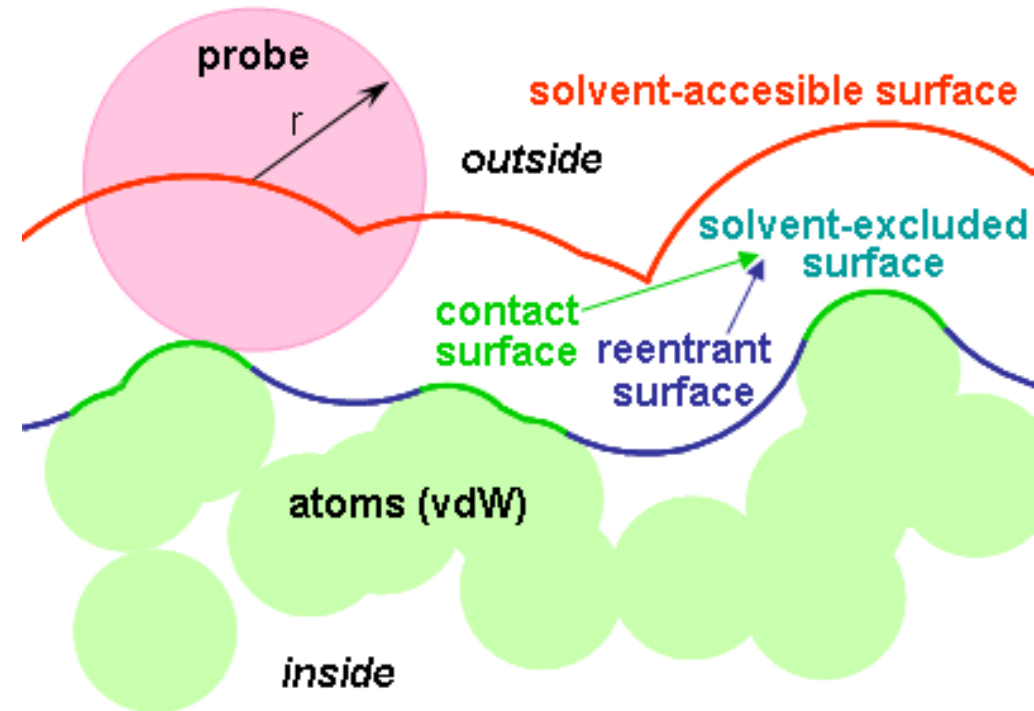
Superficie di van der Waals

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.



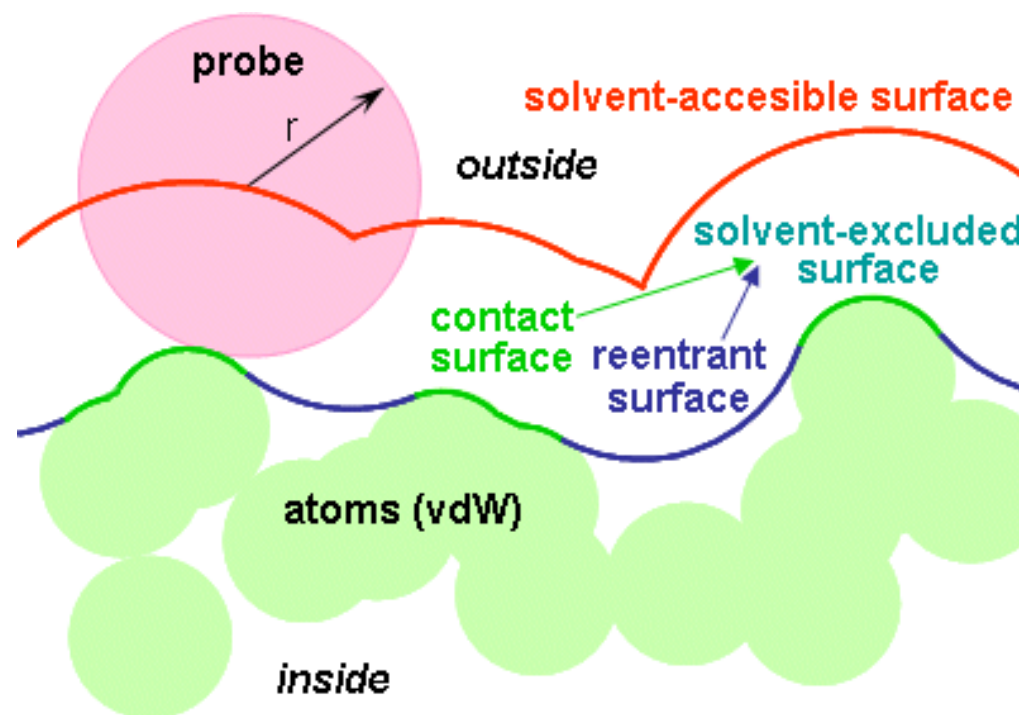
Superficie accessibile ed esclusa al solvente

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 dalla 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').

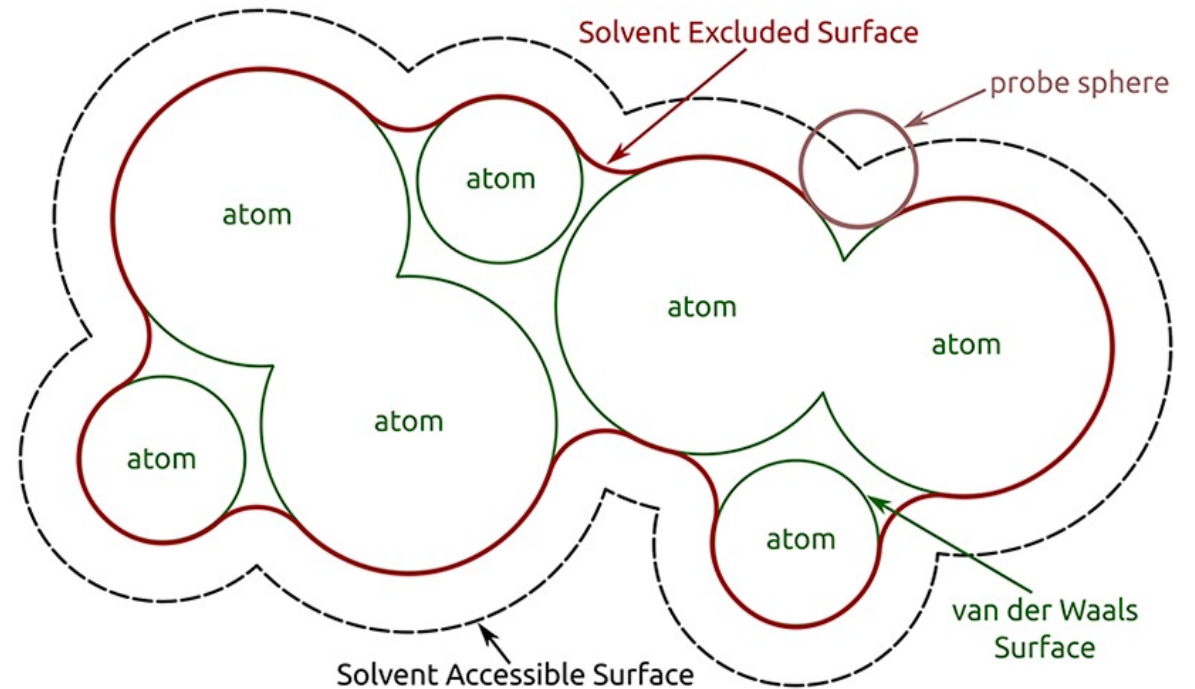
La **superficie accessibile al solvente** è definita dalla convoluzione dei punti generati dal centro della sfera, man mano che rotola sulla superficie macromolecolare.



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

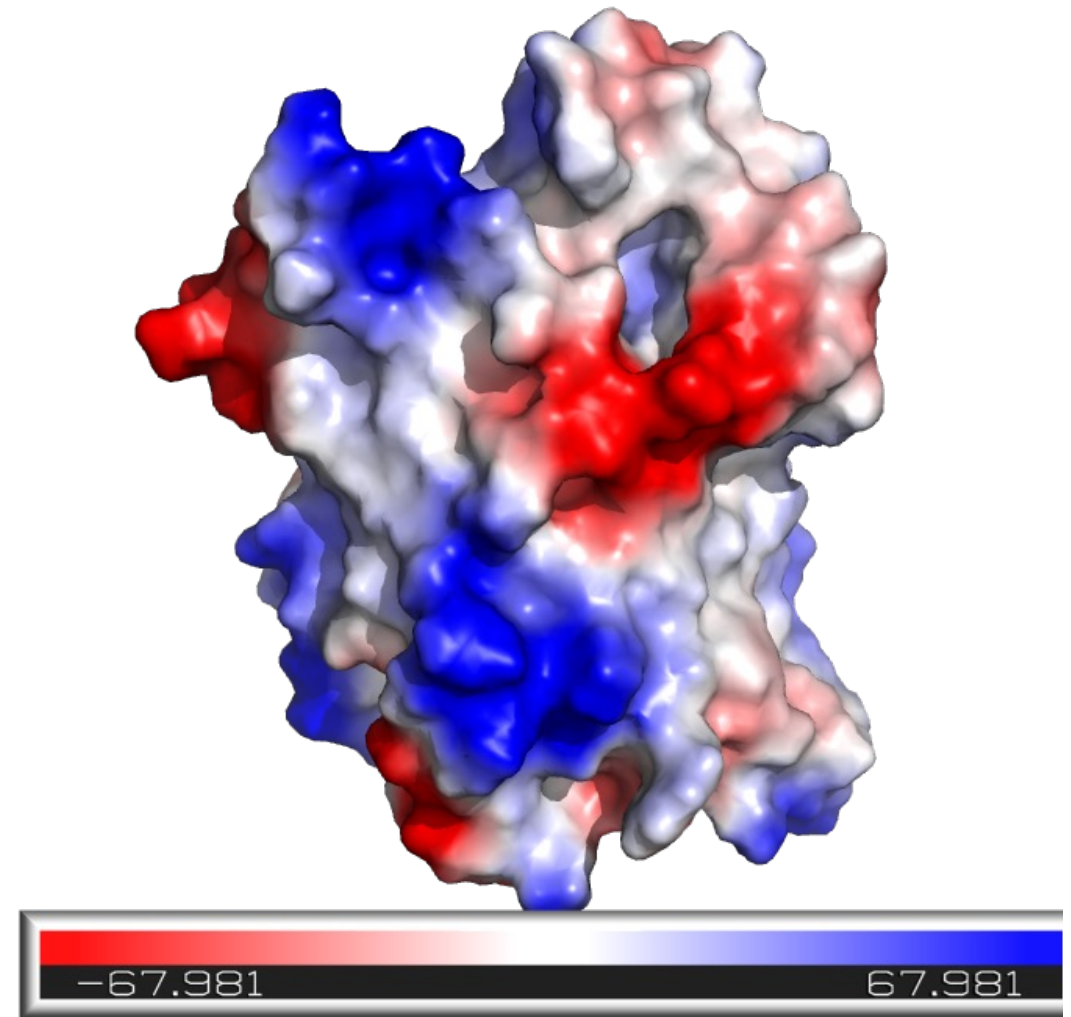


Superficie – potenziale elettrostatico

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.

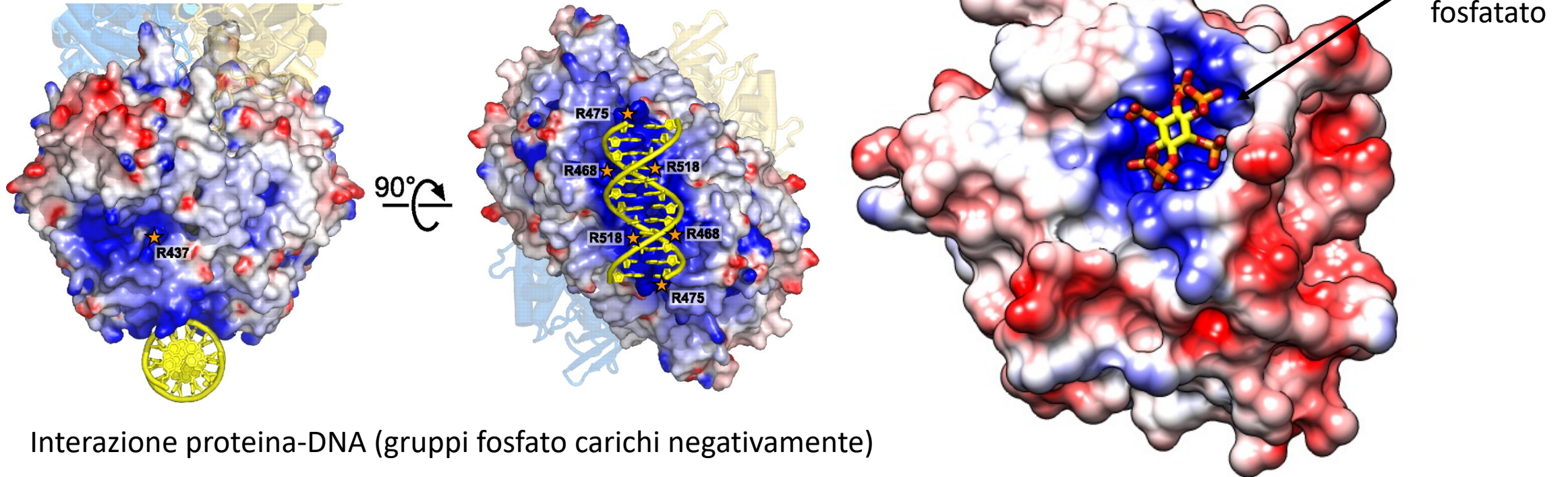
I diversi tipi atomici avranno cariche elettriche differenti, nel complesso una proteina può essere vista come una distribuzione di cariche elettriche associate quindi ad un potenziale elettrostatico.

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.

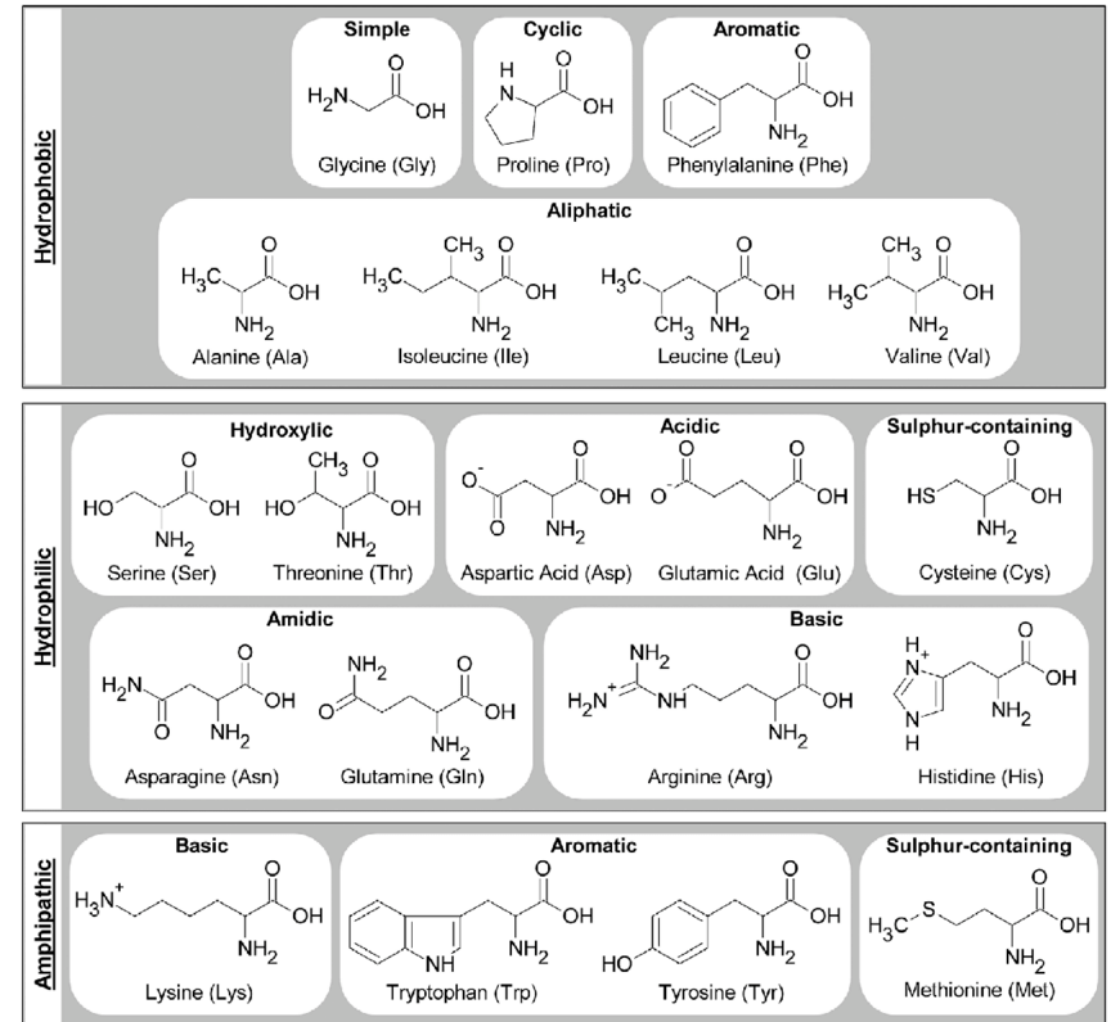


Interazione proteina-DNA (gruppi fosfato carichi negativamente)

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

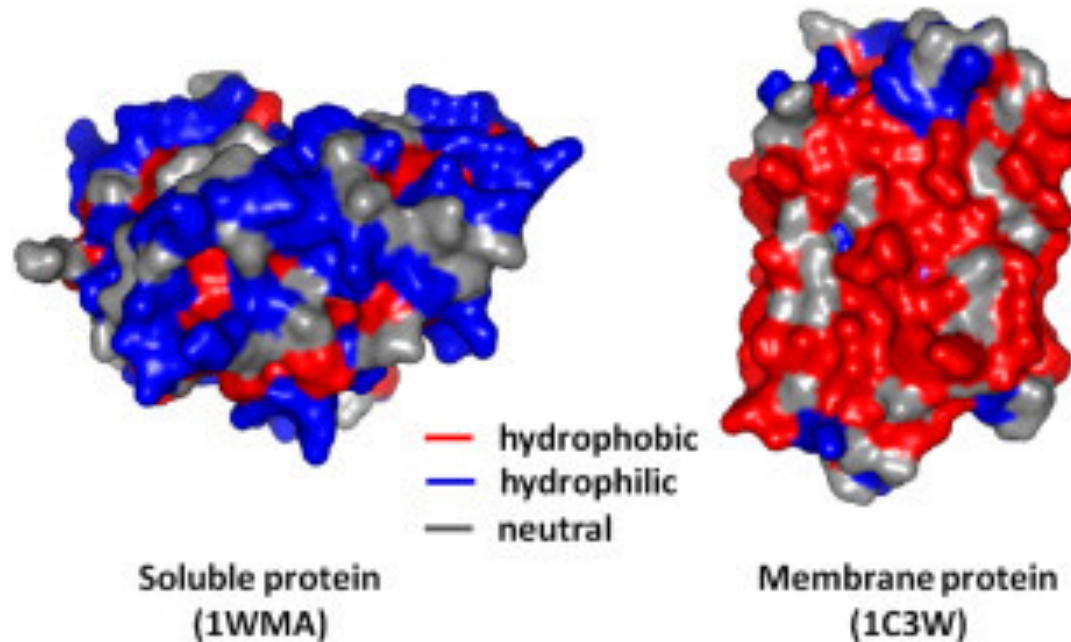


Superficie - Idrofobicità

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.

