

# Bioinformática Estrutural

Sequência



Estrutura



Função

# Fluxo de informação biológica

Gene      ...TTAATAAGT...

↓ transcrição



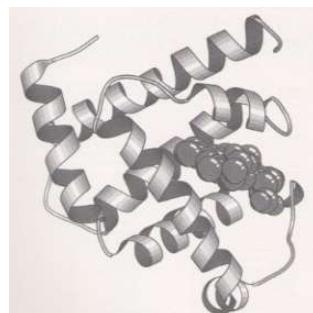
m-RNA    ...UUAAUAAGU...

↓ splicing, tradução

cadeia  
polipeptídica    ...LISVHDN...

↓ modificações pós-translacionais

proteína



Dogma central da  
biologia molecular

Excepções: vírus de RNA,  
priões, ribozimas (?)

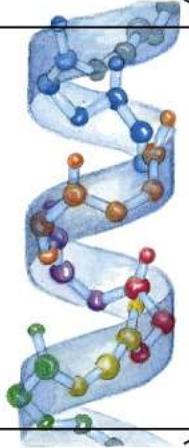
# Níveis de organização da estrutura das proteínas

Estrutura primária



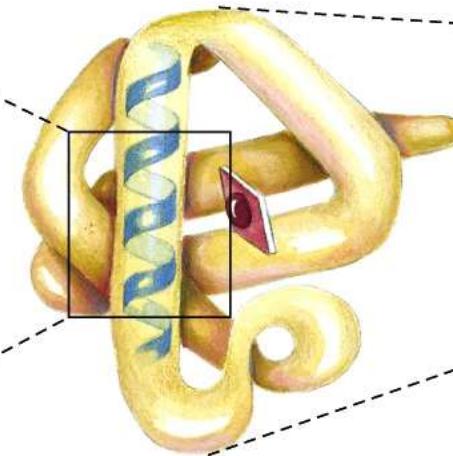
Sequência de aminoácidos

Estrutura secundária



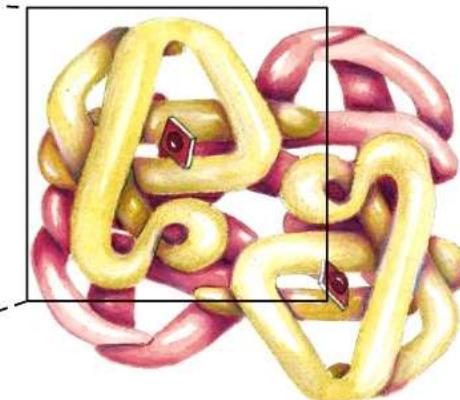
$\alpha$ -hélice

Estrutura terciária



Cadeia polipeptídica

Estrutura quaternária



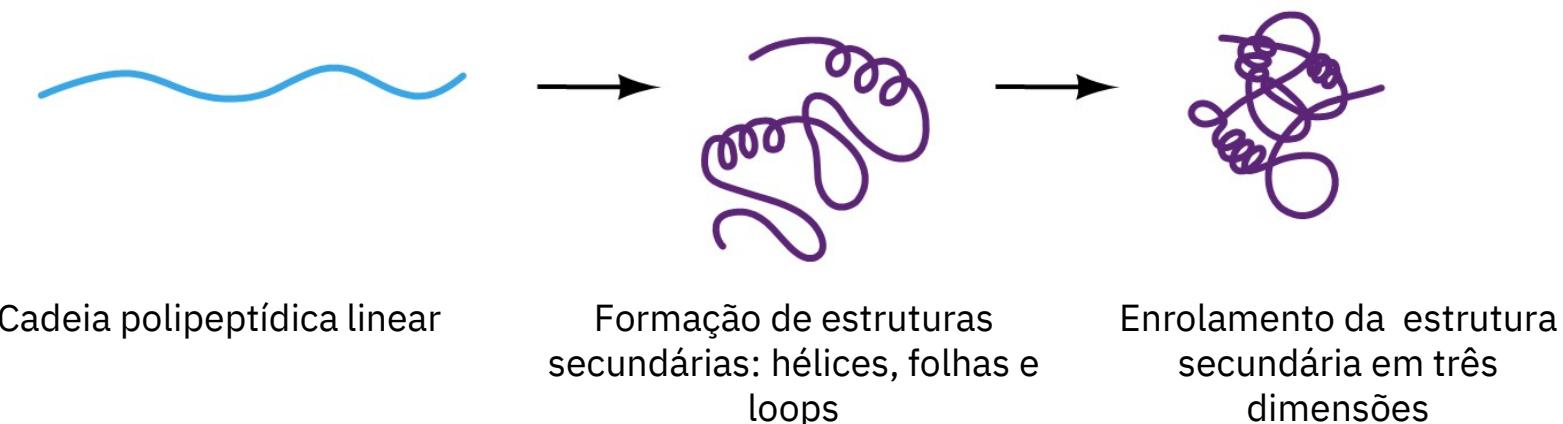
Organização das subunidades

# A estrutura das proteínas é determinada pela sua sequência

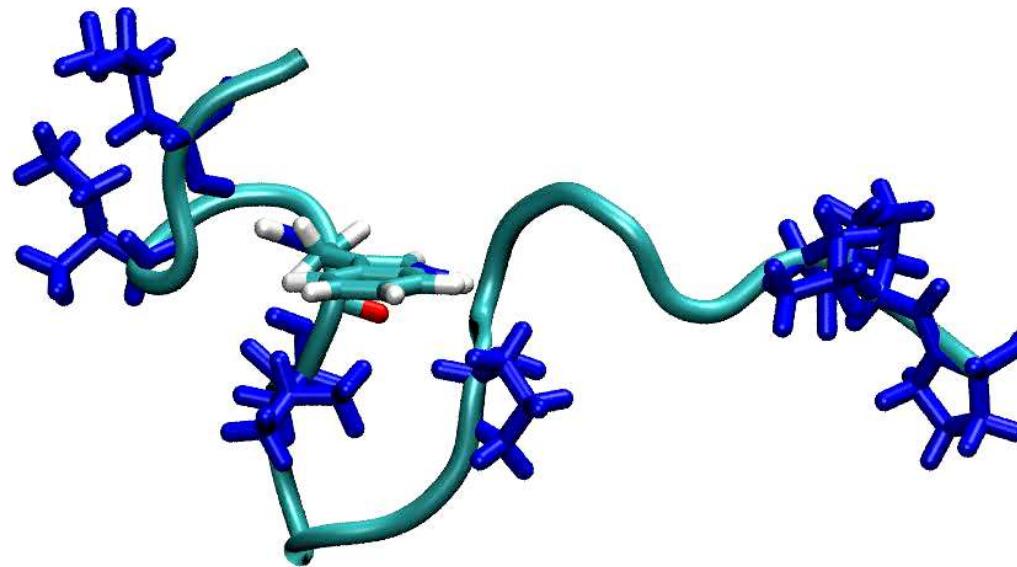
A estrutura tridimensional das proteínas é resultado das interacções entre os átomos que a constituem e o meio aquoso. Em muitos casos a cadeia polipeptídica assume a sua conformação *nativa* de modo espontâneo, após a síntese ribossomal. Este processo tem o nome de “protein folding”.

*A previsão da estrutura tridimensional das proteínas a partir da sua sequência é um dos problemas fundamentais da biologia molecular!*  
*(Folding problem)*

## Mecanismo do “folding” das proteínas:



# Sequência->Estrutura



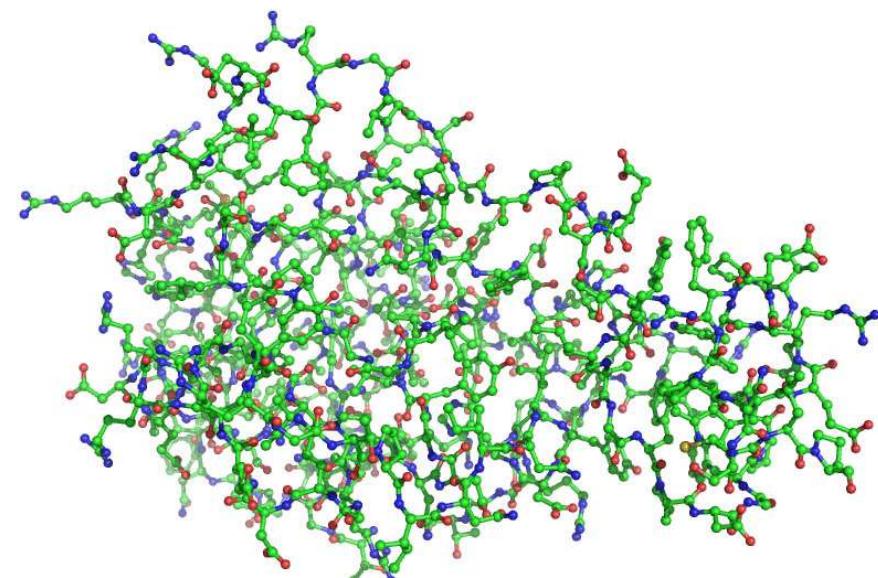
Muitas protéinas adquirem a sua estrutura tridimensional *espontâneamente (folding)*

# A determinação da estrutura é muito mais complexa que a determinação da sequência

Enquanto a sequência de uma proteína ou ácido nucleico é caracterizada simplesmente pela base ou aminoácido que ocorre em cada posição, a descrição da estrutura molecular implica a indicação da posição de cada átomo no espaço tridimensional, bem como a especificação das ligações química entre todos os átomos que constituem cada molécula.

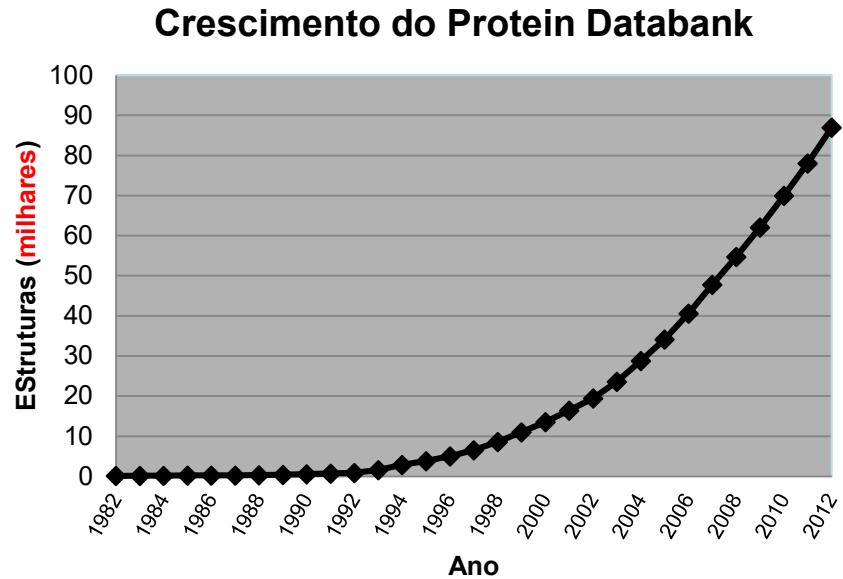
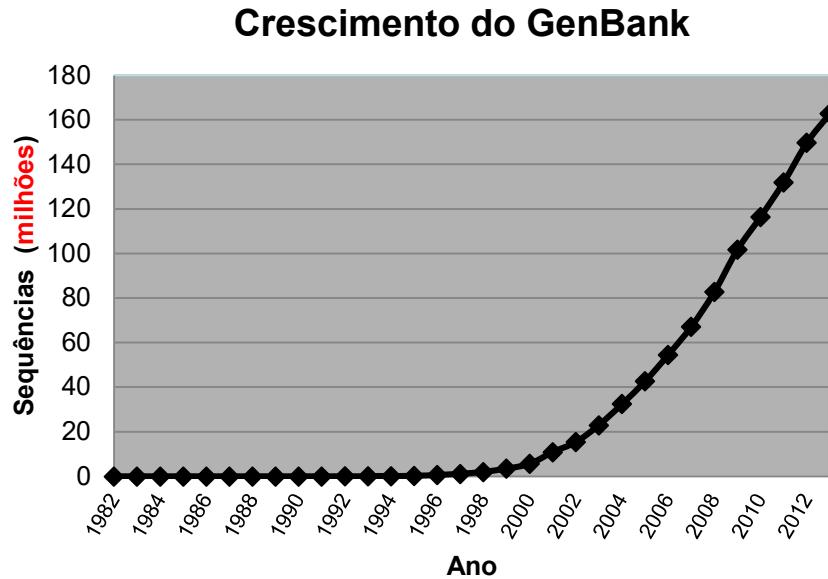
...AVAGGATILVHNQDAGEPAIVLAFG...

Sequência



Estrutura

# Sequência versus estrutura



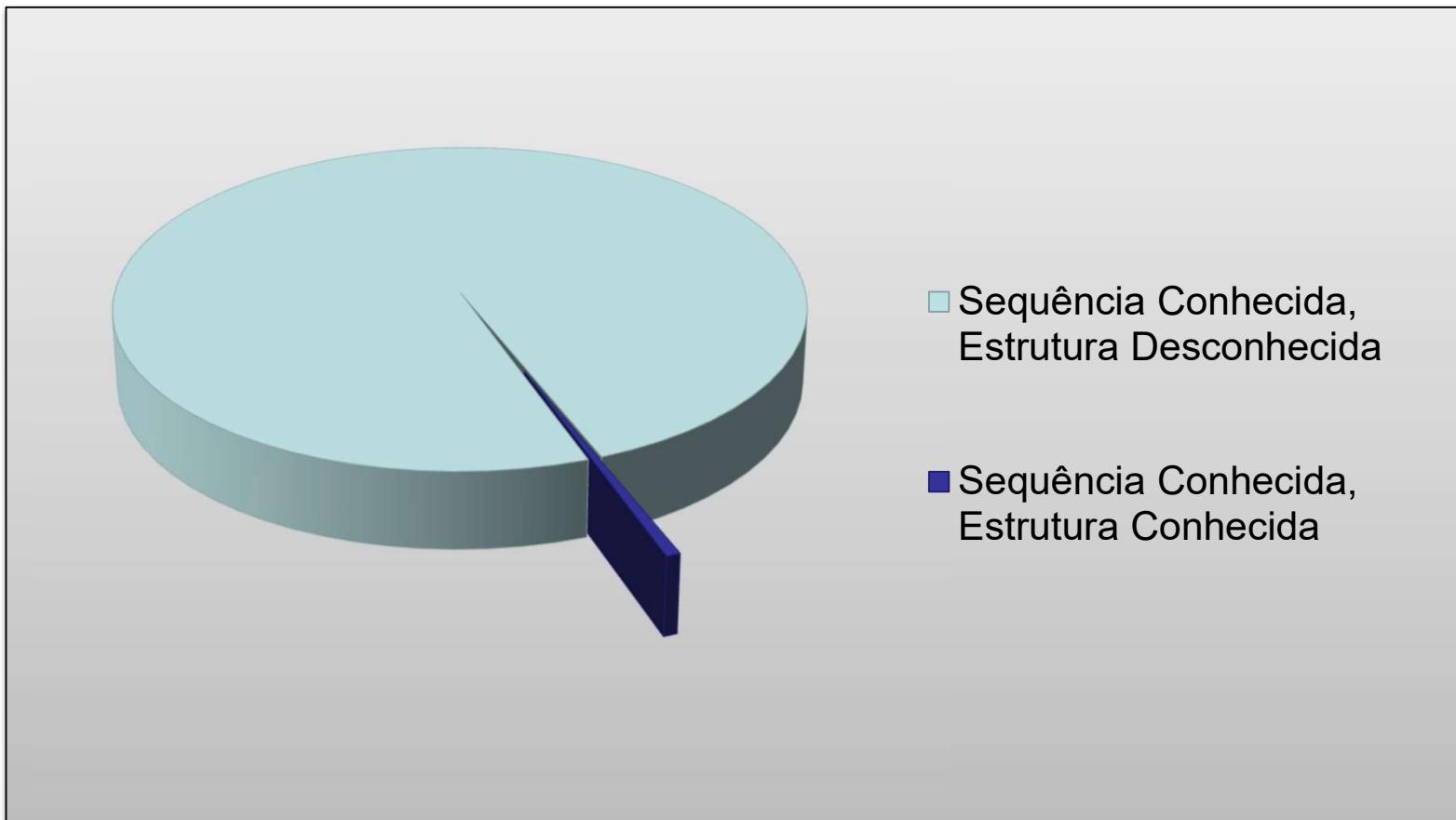
**milhões** de sequências versus **milhares** de estruturas!

**Em 1982:** conhecidas 172 estruturas e 606 sequências ...

**Hoje (Oct 2019):** conhecidas **158,180** structures e **216,763,706** sequências!!

**Conclusão:** A determinação das sequências faz-se a um ritmo muito superior ao das estruturas (cada vez temos mais proteínas de **sequência conhecida** e **estrutura desconhecida**)!

A maior parte das proteínas conhecidas tem estrutura desconhecida



# Importância da previsão estrutural

O elevado e sempre crescente número de sequências de proteínas sem estrutura conhecida torna necessário arranjar métodos mais rápidos de determinação da estrutura tridimensional das proteínas...

Os métodos de determinação da estrutura não têm capacidade de acompanhar o ritmo da determinação das sequências, e provavelmente nunca terão!

## **Como resolver este problema ?**

A estrutura tridimensional das proteínas tem que ser *prevista* a partir da sua sequência. No caso geral este é um problema de difícil solução, mas existem muitas situações em que pode ser resolvido com grande precisão.

A previsão da estrutura tridimensional das proteínas é, portanto, um dos *problemas fundamentais da bioinformática*.

# I. Bancos de dados de estrutura

# Macromoléculas

- O desenvolvimento das técnicas de determinação da estrutura molecular levou à acumulação de um número considerável de estruturas de proteínas (~100000)
- A maior parte das estruturas foram determinadas pelos métodos de difracção (cristalografia) de raios X ou então por ressonância magnética nuclear (RMN)
- A principal base de dados de estruturas de proteínas é o Protein Databank (PDB) <http://www.rcsb.org>

# O Protein Data Bank

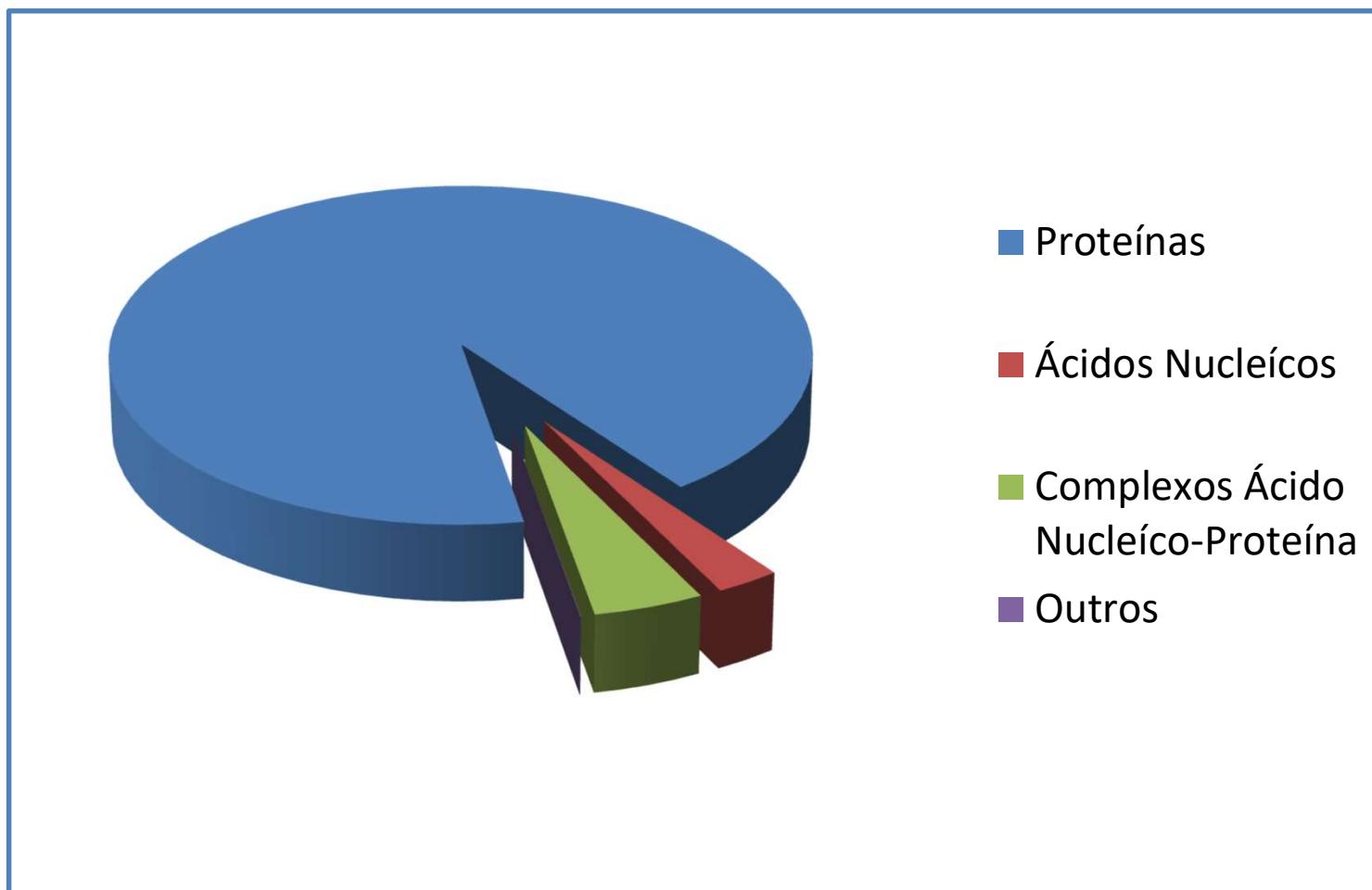
- O Protein Data Bank (PDB) foi criado em 1971 por E.Meyer e W.Hamilton, do Brookhaven National Laboratory (USA), contendo no início 7 estruturas!
- A gestão do PDB foi transferida em 1998 para os membros do RSCB (Research Collaboratory in Structural Bioinformatics) dos quais a Universidade de Rutgers é o site principal. O PDB (<http://www.rcsb.org>) é um banco de dados de acesso **livre**.
- Contendo inicialmente estruturas de proteínas, o PDB contem hoje em dia outros tipos de moléculas, tais como ácidos nucleicos, lípidos e polissacáridos.
- Número total de estruturas em 9/1/2022: **185610**

Técnica experimental	Proteínas	Ácidos nucleicos	Complexos Ac.Nuc./Proteína	Outros	Total
Cristalografia de raios X	151958	2387	7575	161	162081
NMR	11881	1391	274	37	13583
Microscopia electrónica	7477	61	2101	3	9642
Outras	102	3	3	4	109
Combinação	183	8	8	1	195
Total	171601	3850	9953	206	<b>185610</b>

Dados de 9/1/2022 em <http://www.rcsb.org>

Molecular Type	X-ray	NMR	EM	Multiple n	Other M.	Total
<b>Protein</b>	151958	11881	7477	183	102	171601
Protein/NA	7575	274	2101	3	0	9953
Nucleic acid (only)	2387	1391	61	8	3	3850
Others	161	37	3	1	4	206
<b>Total</b>	<b>162242</b>	<b>13620</b>	<b>9645</b>	<b>196</b>	<b>113</b>	<b>185816</b>

O Protein Data Bank contém vários tipos de macromoléculas



# De onde provêm a informação estrutural ?

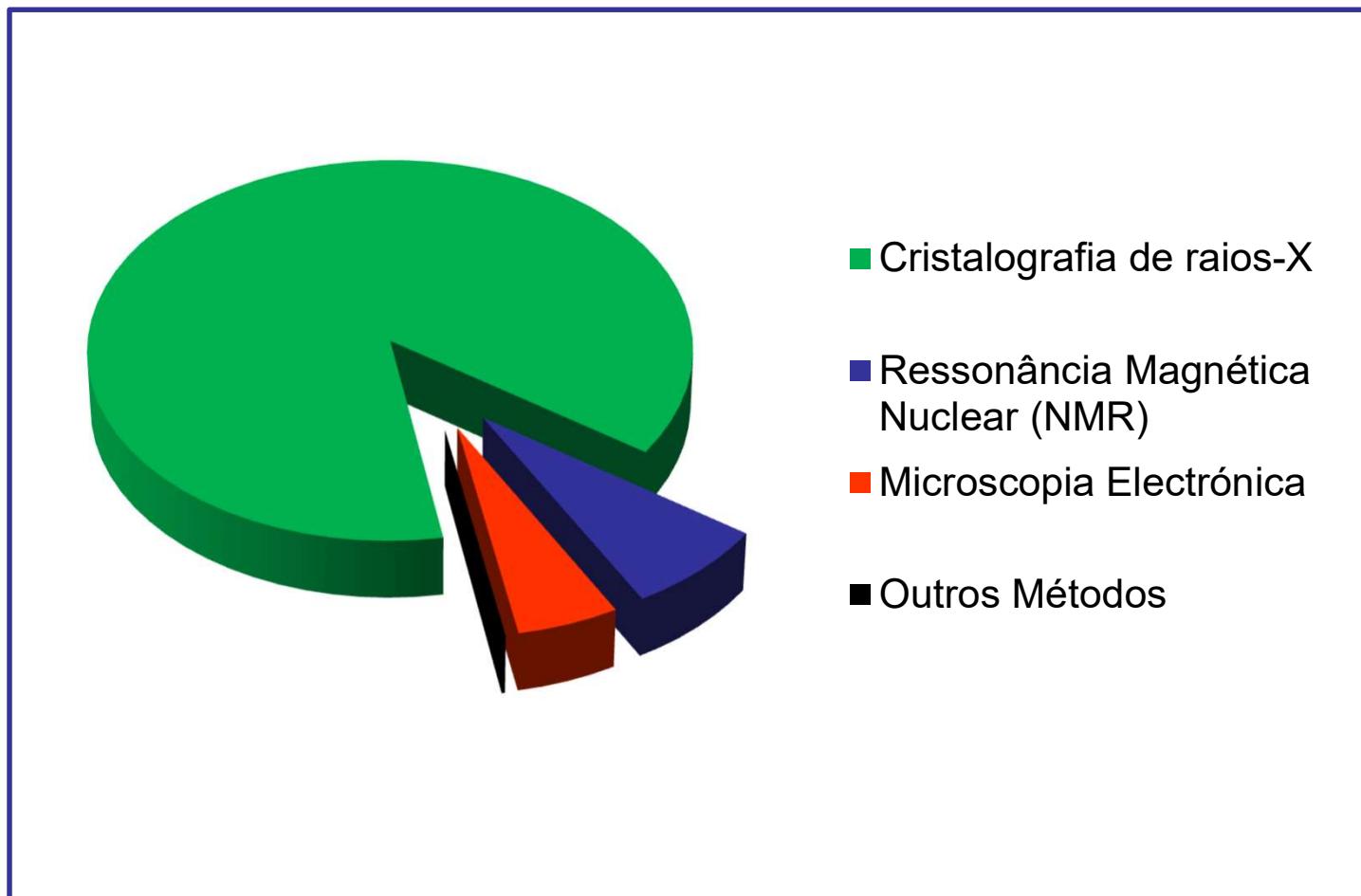
## **Combinação de vários tipos de conhecimento:**

- Teoria da ligação química
- Geometria de moléculas pequenas
- Métodos experimentais para a determinação da estrutura:
  - ❖ Cristalografia de raios X
  - ❖ Ressonância Magnética Nuclear (NMR)
  - ❖ Outros métodos (microscopia, difracção de neutrões, etc...)

# Métodos experimentais

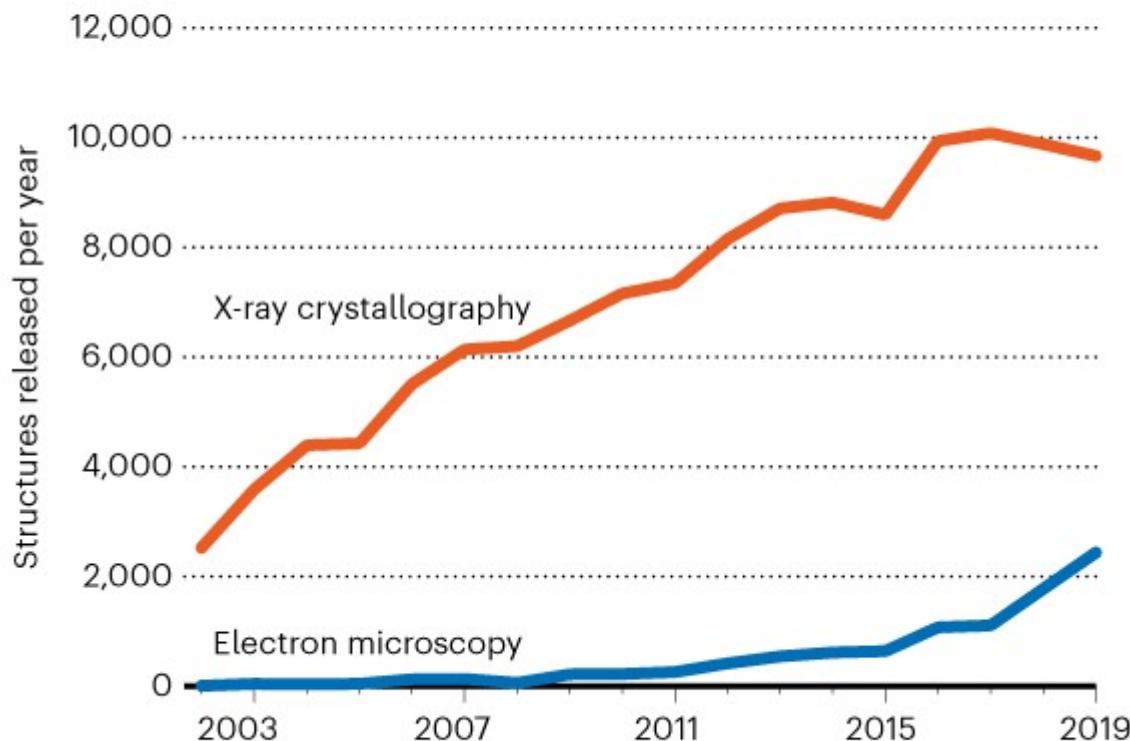
- **Cristalografia de raios X:** a molécula a estudar é purificada e cristalizada a partir de uma solução concentrada. Um feixe de raios X é projectado através do cristal da molécula e o padrão de difracção obtido é usado para resolver a estrutura.
- **Ressonância magnética Nuclear:** a molécula purificada é colocada numa solução aquosa bastante concentrada. A acção de um campo magnético muito intenso provoca o desdobramento dos níveis de energia do spin nuclear de alguns elementos ( $H$ ,  $^{13}C$ ,  $^{15}N$ ), permitindo o estudo do seu ambiente químico e a determinação da estrutura da macromolécula.
- **Crio-microscopia electrónica:** a amostra da molécula a estudar é congelada rapidamente a cerca de  $-180\text{ }^{\circ}\text{C}$  e um feixe de electrões é usado para criar imagens de um enorme número de moléculas da amostra. A análise combinada destas imagens permite resolver a estrutura 3D da molécula.

A maioria da estruturas do PDB são obtidas por cristalografia de raios X



## STRUCTURE SLEUTHS

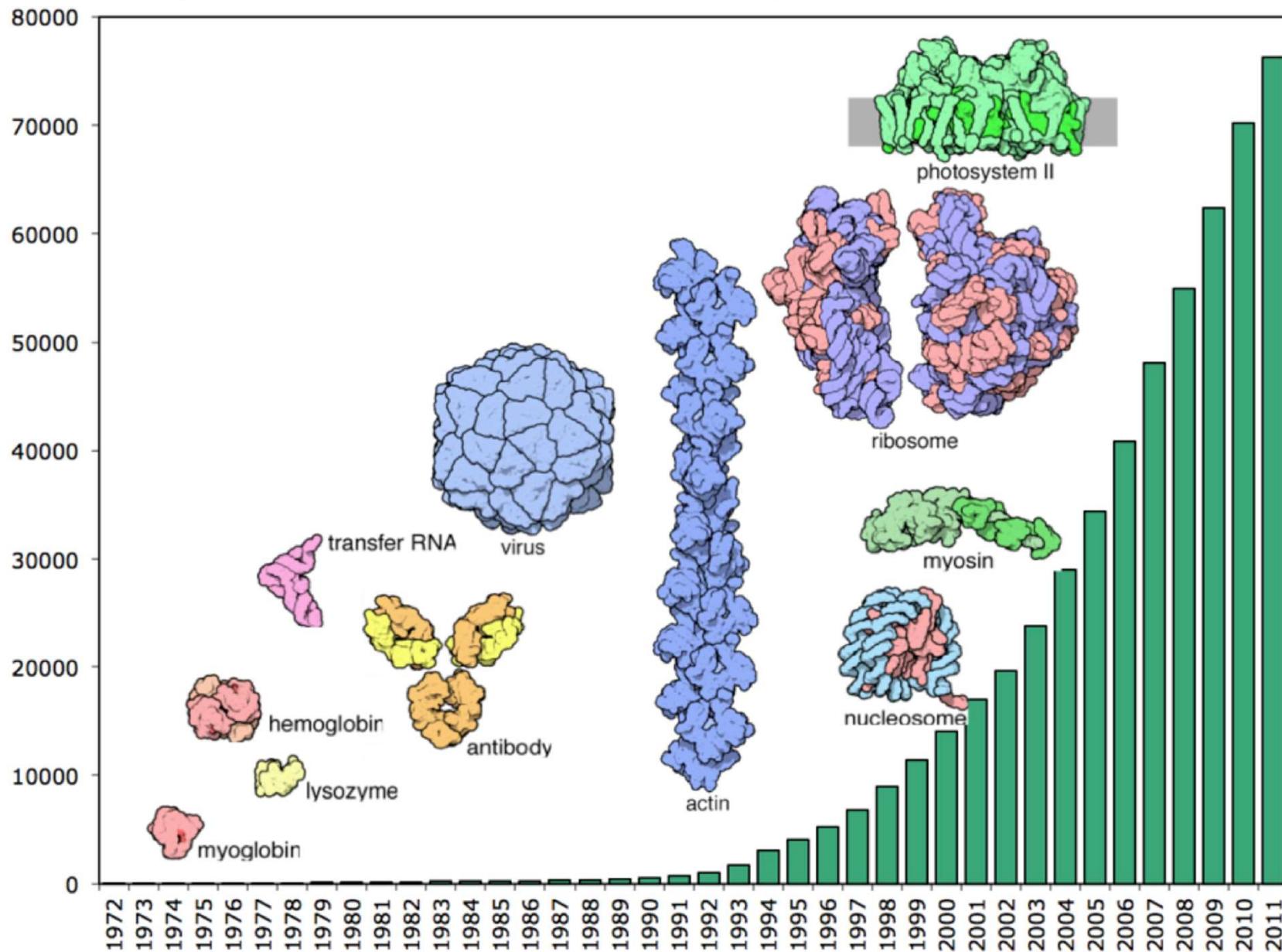
Most structures of proteins and other biological molecules are still solved with X-ray crystallography. But a revolutionary technique called cryo-electron microscopy is catching up, as it becomes more sensitive and widely available.



The electron microscopy line shows structures submitted to the Electron Microscopy Data Bank. Nearly all use cryo-EM.

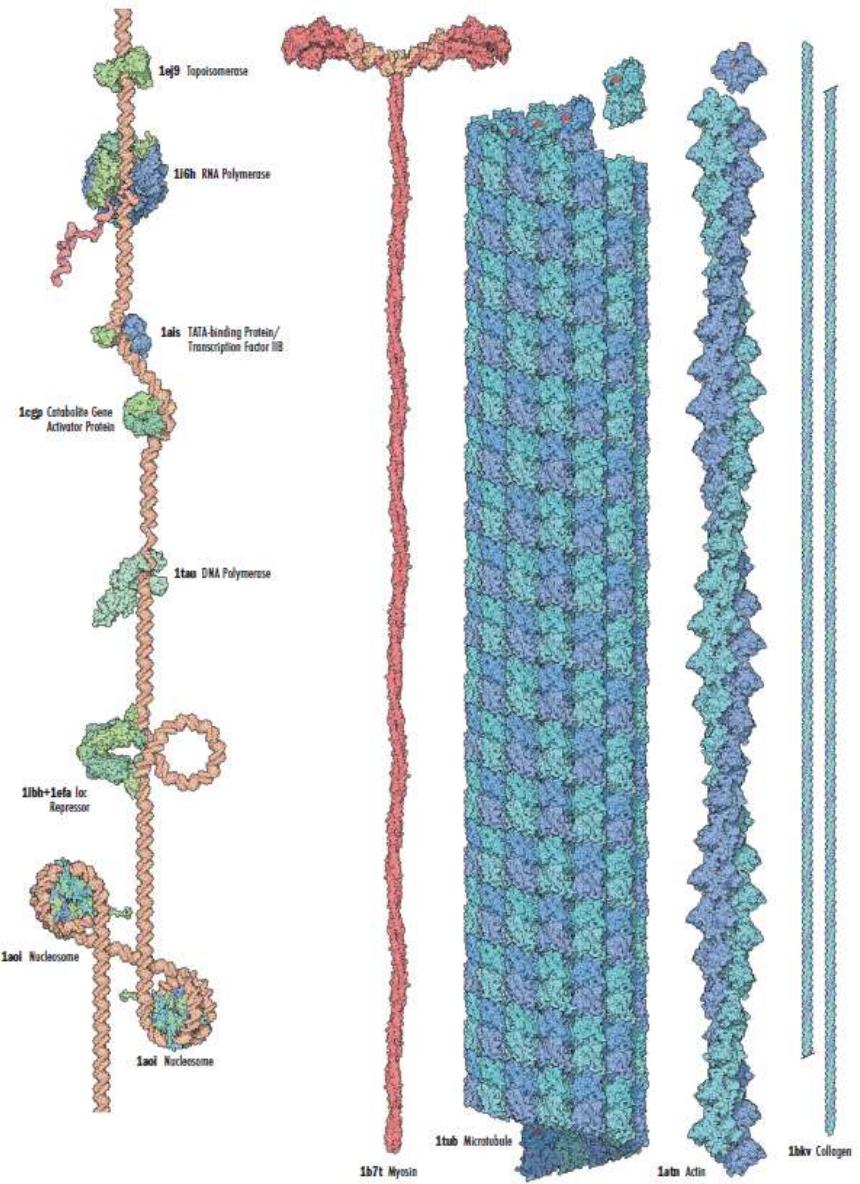
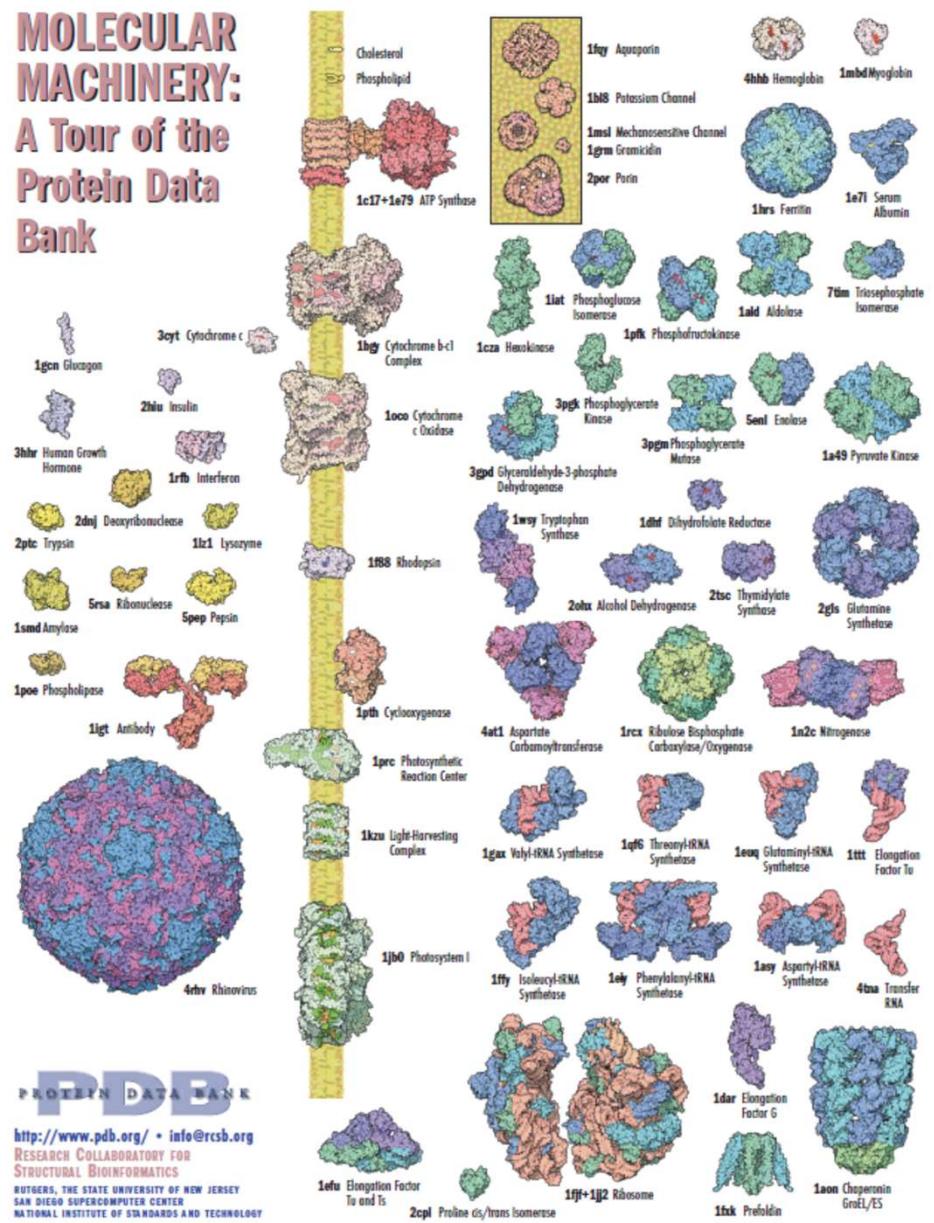
©nature

# Progresso na determinação das estruturas



# O PDB contém uma enorme diversidade estrutural!

## MOLECULAR MACHINERY: A Tour of the Protein Data Bank



# Portal de acesso ao PDB

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

158180 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or ligands Go Advanced Search | Browse by Annotations

PDB-101 EMDataResource Worldwide Protein Data Bank Foundation

Welcome Deposit Search Visualize Analyze Download Learn

A Structural View of Biology

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

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data. The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Video: How Enzymes Work

VIDEO: HOW ENZYMES WORK

November Molecule of the Month

Phospholipase A2

Latest Entries As of Tuesday Nov 26 2019

6J6B PDB Entry

Borrelia burgdorferi OspA via surface entropy reduction (form2)

Features & Highlights

New EM map validation in OneDep Additional validation for electron microscopy maps helps users identify potential discrepancies.

Improved resolution of DOIs for PDB entries Access new wwPDB summary pages for released PDB entries with PDB DOIs

Mandatory PDBx/mmCIF format files submission for MX depositions Submission of PDBx/mmCIF format files for crystallographic depositions to the PDB will be mandatory from July 1<sup>st</sup> 2019 onward. PDB format files will no longer be accepted for deposition of structures solved by MX techniques.

Join Our Team as a Biocurator

News Publications

Education Corner: Gaming Structural Biology for General Audiences (Part 2) Learn about Deep Learning, Citizen Science & Puppies from Diamond Light Source's Michele Darrow 11/26/2019

Introducing Mol\* 11/19/2019

New Papers on Molecular Visualization 11/12/2019

New EM map validation in OneDep 11/05/2019

Education Corner: Gaming Structural Biology for General Audiences (Part 1) 10/29/2019

PDB Turns 48 10/20/2019

Happy Birthday, Irving Geis 10/18/2019

From the Bench to Molecule of the Month 10/15/2019

PDB at a Glance 48974 Distinct Protein Sequences | 44467 Structures of Human Sequences | 11504 Nucleic Acid Containing Structures | More Statistics

<https://www.rcsb.org>

# Portal de acesso ao PDB

RCSB PDB: Homepage Paulo

Secure | https://www.rcsb.org

Apps Acad AWS Free Cell Bits PDF Calendario 2017-18 VTC NiceHash LuckPool Suprnova WhatToMine D pmartel Other bookmarks

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

**Janela de pesquisa**

Search by PDB ID, author, macromolecule, sequence, or ligand   Go Advanced Search | Browse by Annotations

137692 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

PDB-101 Worldwide Protein Data Bank EMDDataBank Nucleic Acid Database Worldwide Protein Data Bank Foundation

**Welcome**

**Deposit**

**Search**

**Visualize**

**Analyze**

**Download**

**Learn**

**A Structural View of Biology**

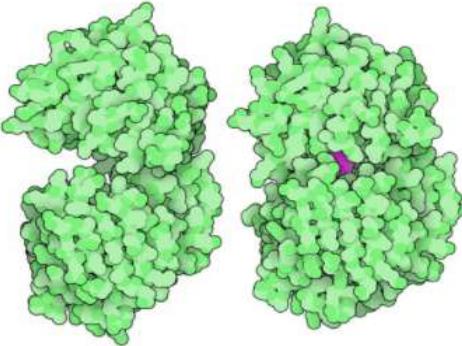
This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

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

New Video: What is a Protein?



**February Molecule of the Month**



EPSP Synthase and Weedkillers

Contact Us

**Latest Entries** As of Tuesday Feb 13 2018



**Features & Highlights**

 New Architecture and Services Enable Faster Access to More Information

Explore the improved display of PDB Statistics, structure funding information, and 3D views of

**News** Publications

 Meet RCSB PDB at AAAS

Learn how RCSB PDB is Sustaining A Living Digital

# Portal de acesso ao PDB

RCSB PDB - Query Re x

www.pdb.org/pdb/results/results.do?tabtoshow=Current&qrId=B13E001B

Gnuplot surprisi... Gnuplot tricks Old Nabble - G... Music Production Bioinformatics 10.10.20.1/form... + Pocket Other bookmarks

All Categories Author Macromolecule Sequence Ligand ?

Search | All Categories: lysozyme

PDB-101 Hide Structural View of Biology Understanding PDB Data Molecule of the Month Educational Resources Author Profiles

MyPDB Hide Login to your Account Register a New Account Query Results (1346) Query History (1)

Home Hide News & Publications Usage/Reference Policies Deposition Policies Website FAQ Deposition FAQ Contact Us About Us Careers External Links Sitemap New Website Features

Deposition Hide All Deposit Services Electron Microscopy X-ray | NMR Validation Server BioSync Beamlines/Facilities Related Tools

Tools Hide Download Files Compare Structures File Formats RESTful Web Services Widgets

Help Hide Launch Help System Display Settings Video Tutorials Glossary of Terms RCSB PDB Mobile

1346 Structure Hits 72 Unreleased Structures 408 Citations 278 Ligand Hits 51 Web Page Hits

Query Parameters: Text Search for: lysozyme Other search suggestions: show

Query Details | Save Query to MyPDB

Query Refinements: Select an item or pie chart Show

Organism Taxonomy Exp. Method X-ray Resolution Release Date Polymer Type

Enzyme Classification SCOP Classification

Refine Query with Advanced Search Remove Similar: Select Percent Similarity

5 Related Molecule of the Month articles Show

PDB Pioneers Nanobodies Adrenergic Receptors Antibodies Lysozyme

Showing 1 - 25 of 1346 Results Results : 25 Page: 1 of 54

Filter: Check All View: Detailed Reports: Select one... Sort: Relevance

**1JKB** HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA

Authors: Muraki, M., Harata, K., Goda, S., Nagahora, H.

Release: 1997-05-15 Classification: Lysozyme

Experiment: X-RAY DIFFRACTION with resolution of 1.66 Å Residue Count: 130

Compound: 1 Polymer [ Display Full Polymer Details | Display for All Results ]

Citation: Importance of van der Waals contact between Glu 35 and Trp 109 to the catalytic action of human lysozyme. (1997) Protein Sci. 6: 473-476 [ Display Full Abstract | Display for All Results ]

Molecule of the Month:

Search Hit: Classification: LYSOZYME

**1JKC** HUMAN LYSOZYME MUTANT WITH TRP 109 REPLACED BY PHE

Authors: Muraki, M., Harata, K., Goda, S., Nagahora, H.

Release: 1997-05-15 Classification: Lysozyme

# Portal de acesso ao PDB

RCSB Protein Data Bank

www.pdb.org/pdb/explore/explore.do?structureId=1JKB

Gnuplot surprises... Gnuplot tricks Old Nabble - G... Music Production Bioinformatics 10.10.20.1/form... + Pocket Other bookmarks

**PDB-101** Hide  
Structural View of Biology  
Understanding PDB Data  
Molecule of the Month  
Educational Resources  
Author Profiles

**MyPDB** Hide  
Login to your Account  
Register a New Account  
Query Results (1346)  
Query History (1)

**Home** Hide  
News & Publications  
Usage/Reference Policies  
Deposition Policies  
Website FAQ  
Deposition FAQ  
Contact Us  
About Us  
Careers  
External Links  
Sitemap  
New Website Features

**Deposition** Hide  
All Deposit Services  
Electron Microscopy  
X-ray | NMR  
Validation Server  
BioSync Beamlines/Facilities  
Related Tools

**Tools** Hide  
Download Files  
Compare Structures  
File Formats  
RESTful Web Services  
Widgets

**Help** Hide  
Launch Help System  
Display Settings  
Video Tutorials  
Glossary of Terms  
RCSB PDB Mobile

**Summary** Sequence Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Geometry Links

**HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA**

DOI:10.2210/pdb1jkb/pdb

**1JKB**

**Display Files**  
**Download Files**  
**Share this Page**

**Primary Citation**  
Importance of van der Waals contact between Glu 35 and Trp 109 to the catalytic action of human lysozyme.  
Muraki, M., Goda, S., Nagahora, H., Harata, K.  
Journal: (1997) Protein Sci. 6: 473-476  
PubMed: 9041653  
PubMedCentral: PMC2143631  
DOI: 10.1002/pro.5560060227  
Search Related Articles in PubMed

**PubMed Abstract:**  
The importance of van der Waals contact between Glu 35 and Trp 109 to the active-site structure and the catalytic properties of human lysozyme (HL) has been investigated by site-directed mutagenesis. The X-ray analysis of mutant HLs revealed that both...  
[ Read More & Search PubMed Abstracts ]

**Molecular Description**  
Classification: LYSOZYME  
Structure Weight: 14786.71  
Molecule: LYSOZYME  
Polymer: 1 Type: protein Length: 130  
Chains: A  
EC#: 3.2.1.17  
Mutation: E35A  
Organism: Homo sapiens  
UniProtKB: Protein Feature View | Search PDB | P61626

**Biological Assembly**

**View in 3D** More Images...  
Biological assembly 1 assigned by authors  
Downloadable viewers:  
Simple Viewer Protein Workshop  
Kiosk Viewer

**MyPDB Personal Annotations** Hide  
To save personal annotations, please login to your MyPDB account.

**Deposition Summary**  
Authors: Muraki, M., Harata, K., Goda, S., Nagahora, H.  
Deposition: 1996-11-13  
Release: 1997-05-15  
Last Modified (REVDAT): 2009-02-24

**Revision History** Hide  
Mouse over text for details

P61626  
Mol. Processing signal peptide Lysozyme C  
E.C. 3.2.1.17:Lysozyme  
UP Sites  
PDB Sites  
Sestruc  
1JKB.A  
+

Source

# Obter o ficheiro de estrutura em formato PDB

RCSB Protein Data Bank

www.pdb.org/pdb/explore/explore.do?structureId=1JKB

Gnuplot surprises... Gnuplot tricks Old Nabble - G... Music Production Bioinformatics 10.10.20.1/form... + Pocket Other bookmarks

PDB-101 Hide  
Structural View of Biology  
Understanding PDB Data  
Molecule of the Month  
Educational Resources  
Author Profiles

MyPDB Hide  
Login to your Account  
Register a New Account  
Query Results (1346)  
Query History (1)

Home Hide  
News & Publications  
Usage/Reference Policies  
Deposition Policies  
Website FAQ  
Deposition FAQ  
Contact Us  
About Us  
Careers  
External Links  
Sitemap  
New Website Features

Deposition Hide  
All Deposit Services  
Electron Microscopy  
X-ray | NMR  
Validation Server  
BioSync Beamlines/Facilities  
Related Tools

Tools Hide  
Download Files  
Compare Structures  
File Formats  
RESTful Web Services  
Widgets

Help Hide  
Launch Help System  
Display Settings  
Video Tutorials  
Glossary of Terms  
RCSB PDB Mobile

HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA

DOI:10.2210/pdb1jkb/pdb

Primary Citation  
Importance of van der Waals contact between Glu 35 and Trp 109 to the catalytic action of human lysozyme.  
Muraki, M., Goda, S., Nagahora, H., Harata, K.  
Journal: (1997) Protein Sci. 6: 473-476  
PubMed: 9041653  
PubMedCentral: PMC2143631  
DOI: 10.1002/pro.5560060227  
Search Related Articles in PubMed

PubMed Abstract:  
The importance of van der Waals contact between Glu 35 and Trp 109 to the active-site structure and the catalytic properties of human lysozyme (HL) has been investigated by site-directed mutagenesis. The X-ray analysis of mutant HLs revealed that both...  
[ Read More & Search PubMed Abstracts ]

Molecular Description  
Classification: Lysozyme  
Structure Weight: 14786.71  
Molecule: LYSOZYME  
Polymer: 1 Type: protein Length: 130  
Chains: A  
EC#: 3.2.1.17  
Mutation: E35A  
Organism: Homo sapiens  
UniProtKB: Protein Feature View | Search PDB | P61626

Uniprot PDB  
P61626  
Molec. Processing  
E.C.  
UP Sites  
PDB Sites  
Sectruct  
1JKB.A  
Signal peptide  
Lysozyme C  
3.2.1.17: Lysozyme

Source

FASTA Sequence  
PDB File (Text) (highlighted)  
PDB File (gz)  
mmcIF File  
mmcIF File (gz)  
PDBML/XML File  
PDBML/XML File (gz)  
Structure Factor (Text)  
Structure Factor (gz)  
Biological Assembly (gz) (A)

Biological assembly 1 assigned by authors  
Downloadable viewers:  
Simple Viewer Protein Workshop  
Kiosk Viewer

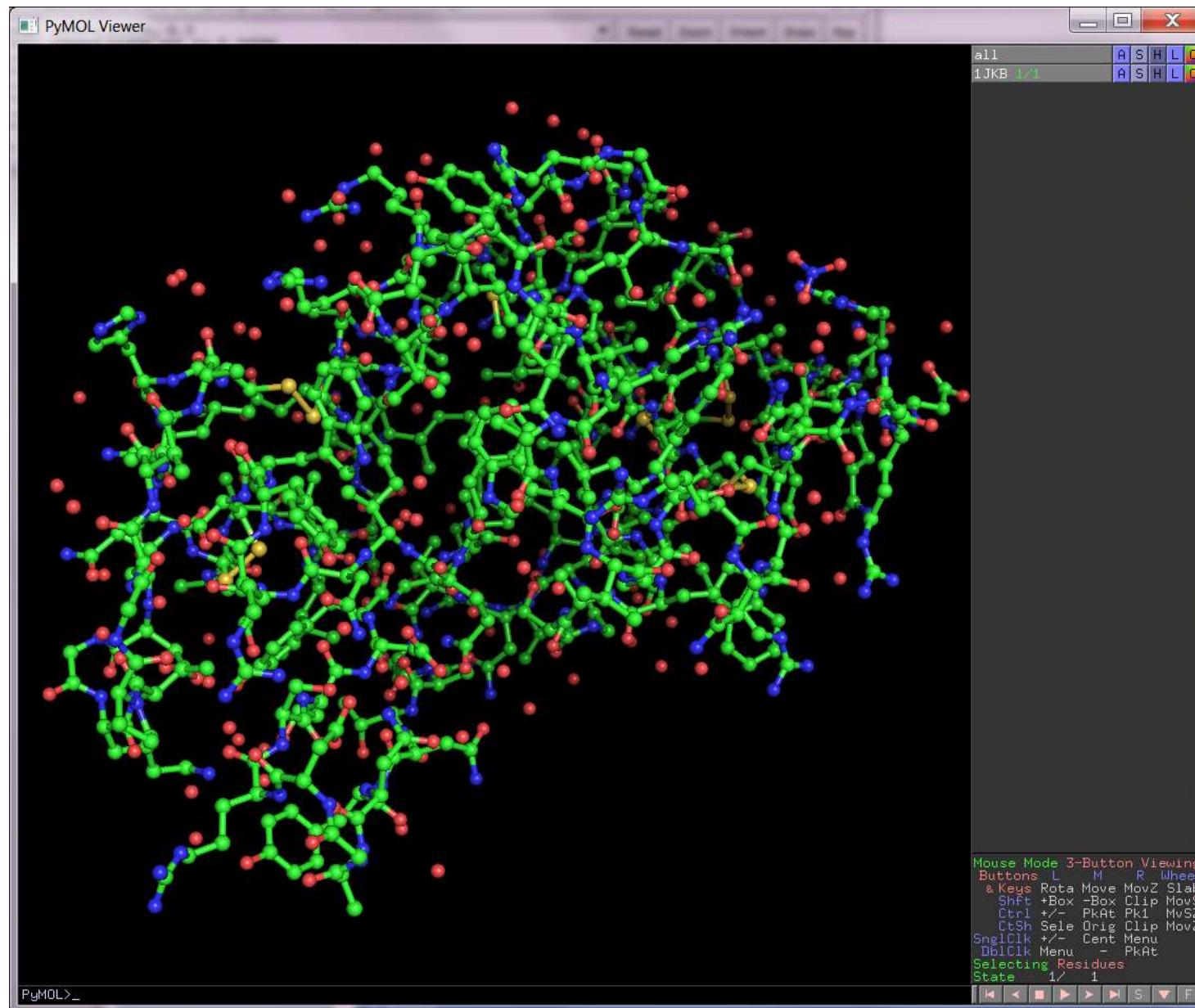
MyPDB Personal Annotations Hide  
To save personal annotations, please login to your MyPDB account.

Deposition Summary Hide  
Authors: Muraki, M., Harata, K., Goda, S., Nagahora, H.  
Deposition: 1996-11-13  
Release: 1997-05-15  
Last Modified (REVDAT): 2009-02-24

Revision History ? Hide  
Mouse over text for details

www.pdb.org/pdb/download/downloadFile.do?fileFormat=pdb&compression=NO&structureId=1JKB

# Visualizar o ficheiro de estrutura no programa PyMOL



# Formatos de representação da estrutura

- A representação da estrutura molecular em bancos de dados passa pela descrição das **coordenadas atómicas**, do **tipo de átomo**, e das **ligações químicas** presentes.
- A descrição do tipo de átomos e ligações que os unem designa-se como **topologia** da molécula.
- No caso das proteínas, a topologia dos 20 aminoácidos standard pode ser assumida *a priori*, pois a estrutura dos aminoácidos é conhecida
- A topologia de outras moléculas, tais como grupos prostéticos , deverá ser especificada
- O formato “tradicional” de representação de estrutura no Protein Data Bank é o formato **PDB**.

# Formato da informação no Protein Data Bank

- A informação contida no Protein Databank inclui coordenadas atómicas, topologias de ligação (descrição das ligações químicas), nomes dos átomos e grupos químicos, dados associados ao processo de determinação experimental da estruturas e outras informações sobre a função, ligandos, propriedades, etc...
- Presentemente a informação no PDB está disponível nos seguintes formatos:
  - **pdb file:** O formato “flat file”, um tipo de ficheiro chamado “ficheiro PDB”. Estes ficheiros são os mais utilizados pelos softwares de manipulação e visualização de estruturas e têm geralmente a extensão “.pdb”
  - **mmCIF:** - um formato mais poderoso e estruturado que o ficheiro PDB, ainda não tendo sido largamente adoptado
  - **XML:** - extended mark-up language, um formato muito geral de representação de informação, compatível com um vasto número de aplicações de software.

# Formato do ficheiro PDB

```
HEADER      METAL BINDING PROTEIN          21-AUG-03   1Q8H
TITLE       CRYSTAL STRUCTURE OF PORCINE OSTEOCALCIN
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: OSTEOCALCIN;
COMPND     3 CHAIN: A
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: SUS SCROFA;
SOURCE      3 ORGANISM_COMMON: PIG
KEYWDS     HELIX-TURN-HELIX-TURN-HELIX, PAPER-CLIP, HYDROXYAPATITE
KEYWDS     2 CRYSTAL SURFACE BINDING PROTEIN, CALCIUM BINDING PROTEIN,
KEYWDS     3 BONE GLA PROTEIN
EXPDTA    X-RAY DIFFRACTION
AUTHOR     Q.Q.HOANG, F.SICHERI, A.J.HOWARD, D.S.YANG
REVDAT    1 11-NOV-03 1Q8H 0
JRNL       AUTH Q.Q.HOANG, F.SICHERI, A.J.HOWARD, D.S.YANG
JRNL       TITL BONE RECOGNITION MECHANISM OF PORCINE OSTEOCALCIN
JRNL       TITL 2 FROM CRYSTAL STRUCTURE.
JRNL       REF  NATURE V. 425 977 2003
JRNL       REFN ASTM NATUAS UK ISSN 0028-0836
REMARK    1
REMARK    2
REMARK    2 RESOLUTION. 2.00 ANGSTROMS.
REMARK    3
REMARK    3 REFINEMENT.
REMARK    3 PROGRAM : CNS 1.1
REMARK    3 AUTHORS : BRUNGER, ADAMS, CLORE, DELANO, GROS, GROSSE-
```

.....

ATOM	1	N	PRO	A	13	10.210	29.966	44.935	1.00	38.06	N
ATOM	2	CA	PRO	A	13	9.718	29.013	43.919	1.00	37.33	C
ATOM	3	C	PRO	A	13	9.566	29.662	42.541	1.00	37.52	C
ATOM	4	O	PRO	A	13	9.275	30.855	42.444	1.00	38.00	O
ATOM	5	CB	PRO	A	13	8.383	28.488	44.434	1.00	37.68	C
ATOM	6	CG	PRO	A	13	7.919	29.624	45.336	1.00	36.60	C
ATOM	7	CD	PRO	A	13	9.196	30.126	45.995	1.00	36.47	C
ATOM	8	N	ASP	A	14	9.777	28.879	41.483	1.00	36.83	N
ATOM	9	CA	ASP	A	14	9.671	29.384	40.116	1.00	36.13	C

.....

```
MASTER      299      0      6      3      0      0      0      6      378      1      38      4
END
```

Cabeçalho

Coordenadas

# Interligação entre Uniprot e PDB

The screenshot shows a web browser window displaying the UniProtKB entry for protein P00760 (TRY1\_BOVIN). The page has a dark blue header with tabs for 'Search', 'Blast \*', 'Align', 'Retrieve', and 'ID Mapping \*'. Below the header is a search bar with dropdowns for 'Search in' (set to 'Protein Knowledgebase (UniProtKB)') and 'Query' (empty), along with 'Search', 'Advanced Search», and 'Clear' buttons. The main content area displays the protein's name, P00760 (TRY1\_BOVIN), its status as 'Reviewed', and its source, UniProtKB/Swiss-Prot. It also shows the last modification date (October 16, 2013) and version (149). A 'History...' link is available. On the right side, there are links for 'Contribute', 'Send feedback', and 'Read comments (0) or add your own'. Below the main content, there are links for 'Clusters with 100%, 90%, 50% identity', 'Documents (3)', and 'Third-party data'. A row of download links for 'text', 'xml', 'rdf/xml', 'gff', and 'fasta' formats is also present. The navigation menu at the bottom includes 'Names', 'Attributes', 'General annotation', 'Ontologies', 'Interactions', 'Sequence annotation', 'Sequences', 'References', 'Web links', and 'Cross-refs'. The 'Cross-refs' tab is highlighted with a red box. The 'Names and origin' section contains detailed information about the protein's names (Recommended name: Cationic trypsin, EC=3.4.21.4; Alternative name(s): Beta-trypsin), its cleavage into two chains (Alpha-trypsin chain 1 and Alpha-trypsin chain 2), and its organism (Bos taurus (Bovine)). The 'Organism' section also lists the taxonomic identifier (9913 [NCBI]) and the taxonomic lineage (Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Laurasiatheria > Cetartiodactyla > Ruminantia > Pecora > Bovidae > Bovinae > Bos).

# Interligação entre Uniprot e PDB

Screenshot of a web browser showing the Uniprot entry for P00760. The page displays cross-references and 3D structure databases.

**Cross-references**

**Sequence databases**

<input checked="" type="radio"/> EMBL	BC134797 mRNA. Translation: AAI34798.1.
<input type="radio"/> GenBank	BC146041 mRNA. Translation: AAI46042.1.
<input type="radio"/> DDBJ	D38507 mRNA. Translation: BAA07516.1.
IPI	IPI00706427.
PIR	TRBOTR. A90164.
RefSeq	NP_001107199.1. NM_001113727.1.
UniGene	Bt.91423.

**3D structure databases**

Entry	Method	Resolution (Å)	Chain	Positions	PDBsum
1AQ7	X-ray	2.20	A	24-246	[»]
1AUJ	X-ray	2.10	A	24-246	[»]
1AZ8	X-ray	1.80	A	24-246	[»]
1BJU	X-ray	1.80	A	24-246	[»]
1BJV	X-ray	1.80	A	24-246	[»]
1BTP	X-ray	2.20	A	18-246	[»]
1BTW	X-ray	1.70	A	18-246	[»]
1BTX	X-ray	1.70	A	18-246	[»]
1BTY	X-ray	1.50	A	18-246	[»]
1BTZ	X-ray	2.00	A	18-246	[»]

Blues Drums Bundle   theloophloft's stream   Mino Cinelu - Wikipedia   RCSB Protein Data Bank   Configurar - Apple S   Over To You: MacBo   RCSB Protein Data Bank

www.pdb.org/pdb/explore/explore.do?structureId=1AQ7

**PDB** PROTEIN DATA BANK   **PDB-101**

A MEMBER OF THE **PDB** | EMDDataBank  
An Information Portal to Biological Macromolecular Structures  
As of Tuesday Oct 29, 2013 at 5 PM PDT there are 95113 Structures | PDB Statistics |

Search Advanced Browse

Everything Author Macromolecule Sequence Ligand ?  
e.g., PDB ID, molecule name, author

Search History, Previous Results

**1AQ7** Display Files ▾ Download Files ▾ Share this Page ▾

**PDB-101** Hide  
Structural View of Biology  
Understanding PDB Data  
Molecule of the Month  
Educational Resources  
Author Profiles

**MyPDB** Hide  
Login to your Account  
Register a New Account  
MyPDB Help Page

**Home** Hide  
News & Publications  
Usage/Reference Policies  
Deposition Policies  
Website FAQ  
Deposition FAQ  
Contact Us  
About Us  
Careers  
External Links  
Sitemap  
New Website Features

**Summary** 3D View Sequence Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Geometry Links

**TRYPSIN WITH INHIBITOR AERUGINOSIN 98-B**

DOI:10.2210/pdb1aq7/pdb

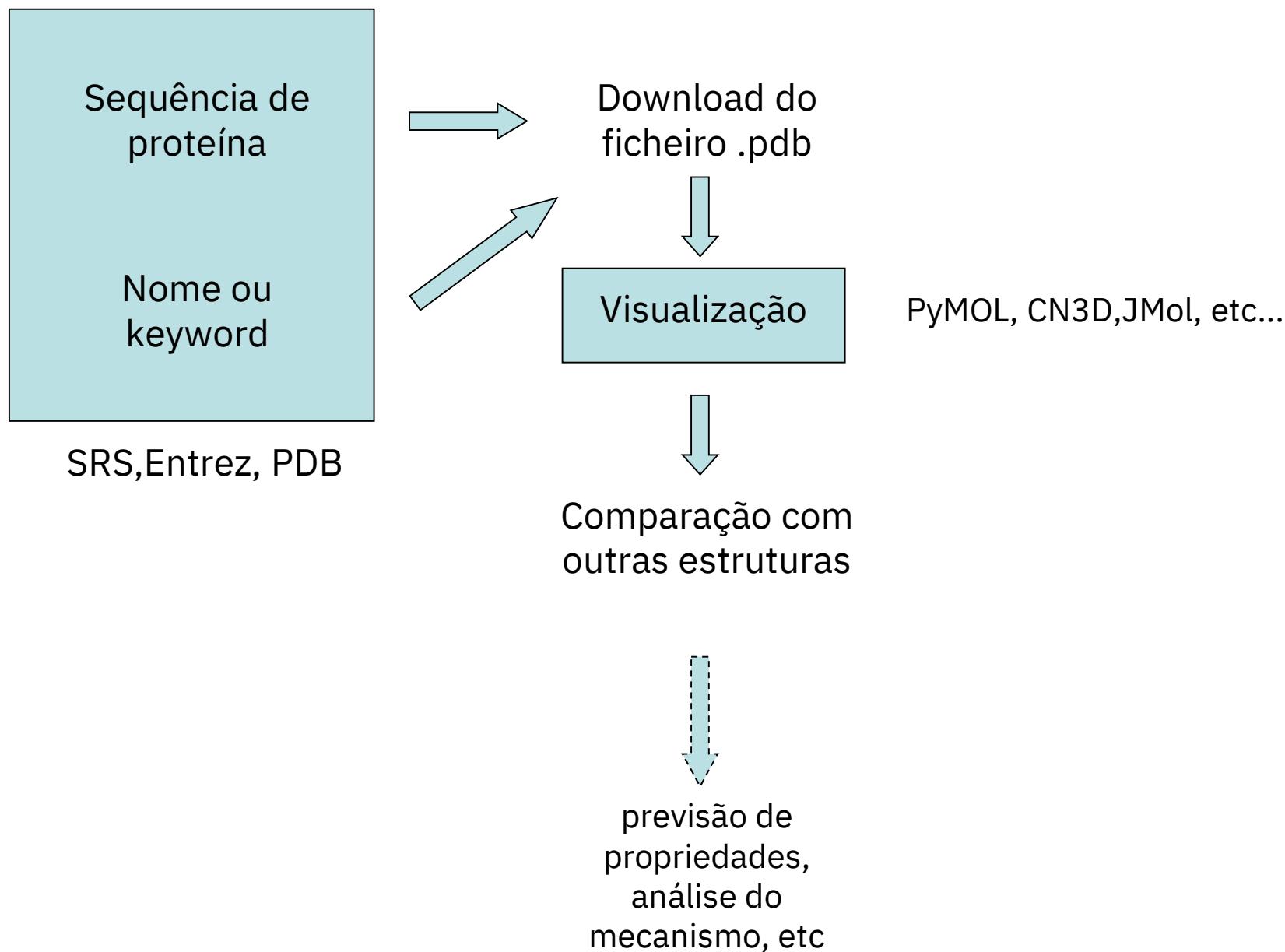
**Primary Citation**  
**Atomic Structure of the Trypsin-Aeruginosin 98-B Complex**  
Sandler, B. , Murakami, M. , Clardy, J.   
Journal: (1998) J.Am.Chem.Soc. **120**: 595-596  
PubMed ID is not available

**Molecular Description** Hide  
**Classification:** Hydrolase/hydrolase Inhibitor   
**Structure Weight:** 23979.18

<b>Molecule:</b>	TRYPSIN
<b>Polymer:</b>	1
<b>Type:</b>	protein
<b>Chains:</b>	A
<b>EC#:</b>	3.4.21.4

**Biological Assembly**

# Visualização de estruturas moleculares

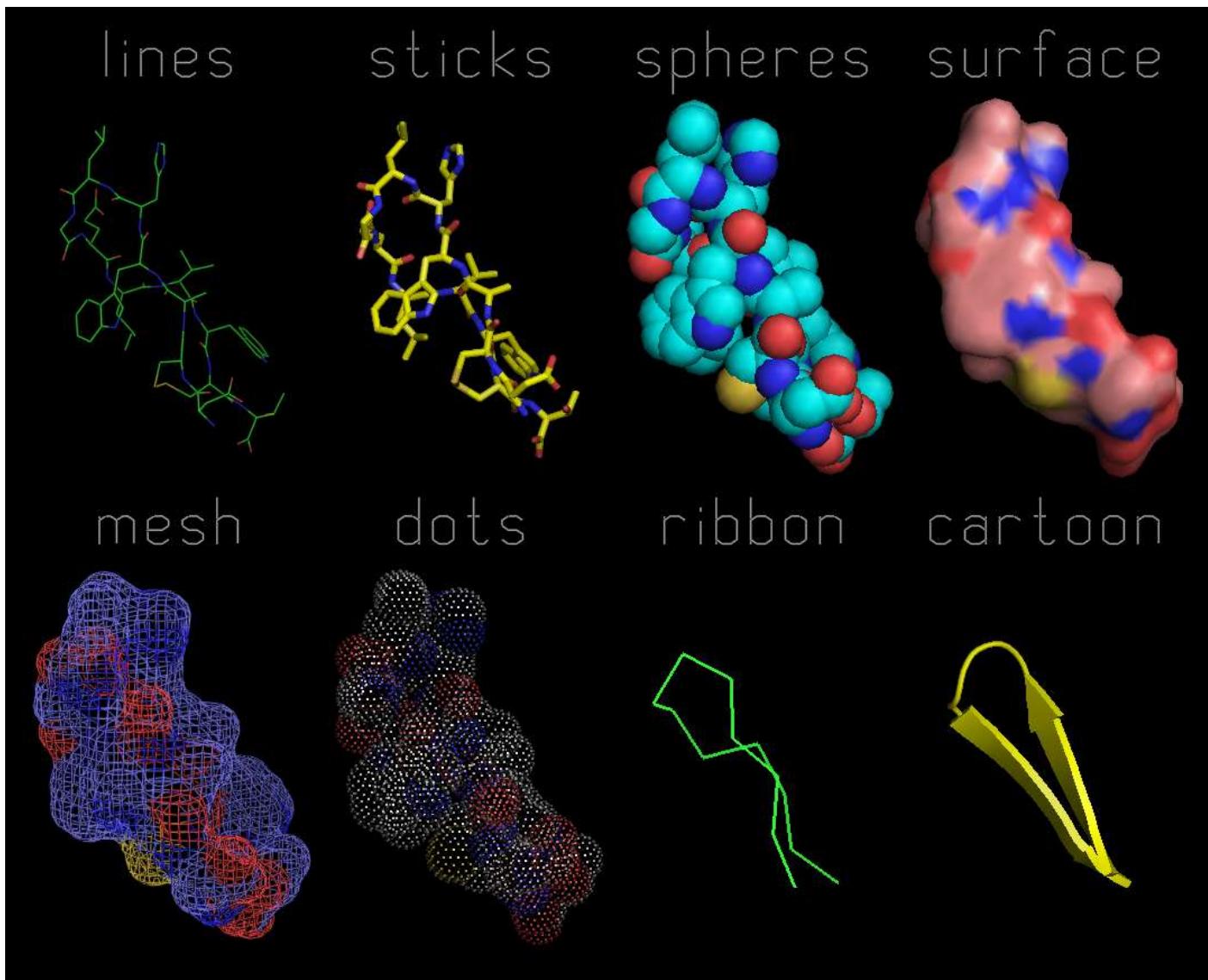


# Software para visualização molecular

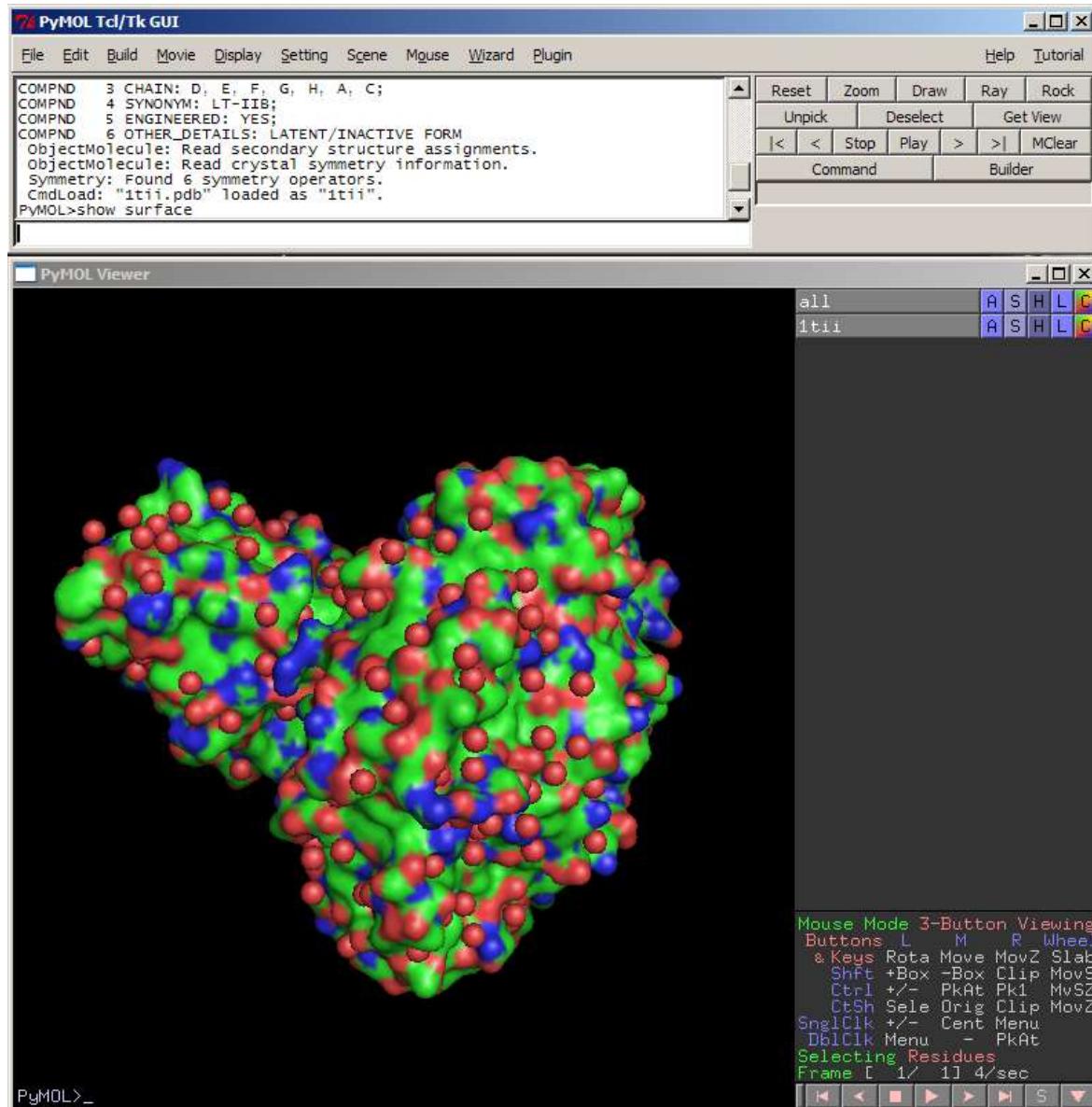
Aplicações de software que permitem a visualização de ficheiros de estrutura molecular (ficheiros PDB e outros formatos), permitindo a análise e cálculo de propriedades moleculares e a comparação de diferentes estruturas

- **PyMOL:** <http://www.pymol.org>
- **Cn3D:** <http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>
- **JMol:** <http://jmol.sourceforge.net/>
- **SwissPDB viewer:** <http://www.expasy.org/spdbv/>
- **Rasmol:** <http://www.rasmol.org/>

# Modos de representação de estruturas



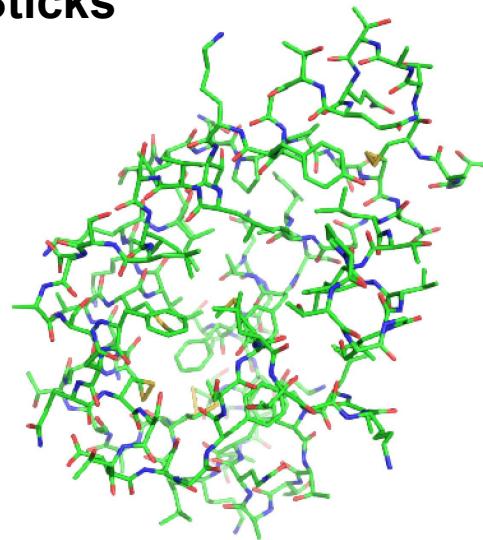
# PyMOL



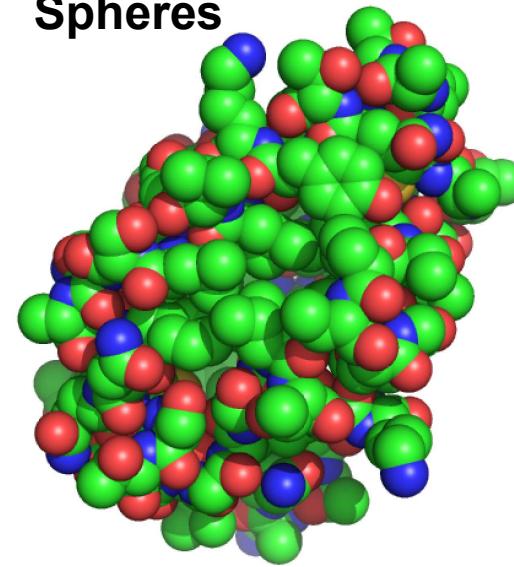
- ❖ Open Source
- ❖ Acesso livre
- ❖ Python / C
- ❖ Visualização de macromoléculas
- ❖ Animações moleculares
- ❖ Comparaçao de estruturas
- ❖ Scripting
- ❖ Windows / Linux

<http://www.pymol.org>

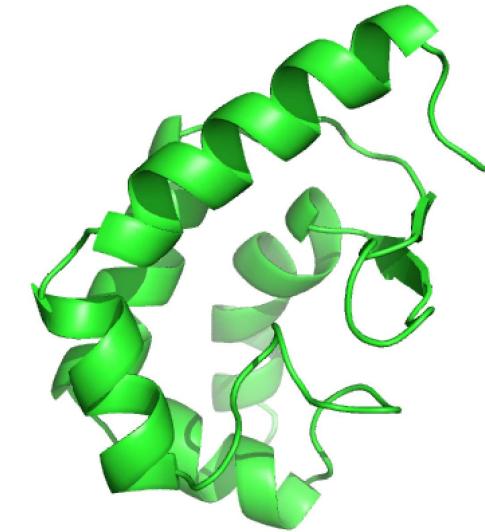
**Sticks**



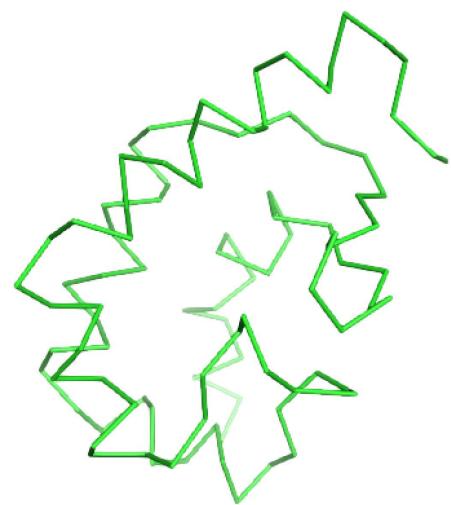
**Spheres**



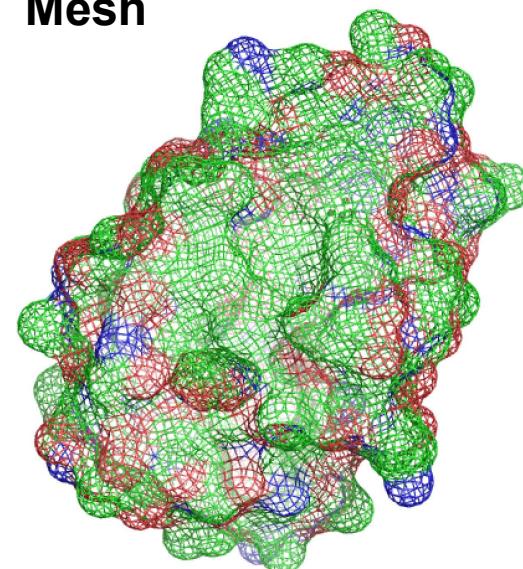
**Cartoon**



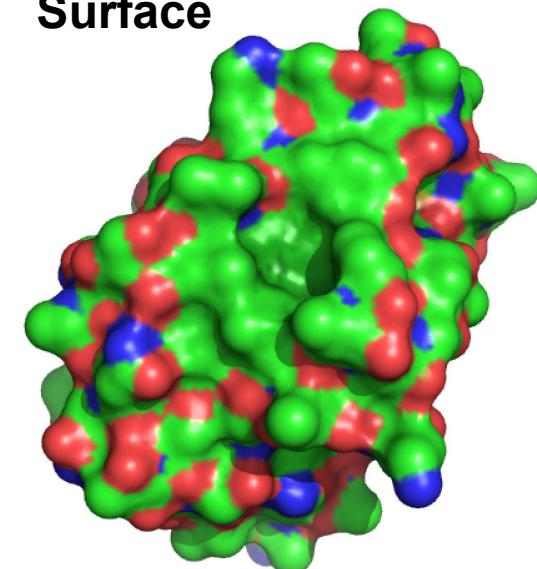
**Ribbon**



**Mesh**



**Surface**



## II. Alinhamento e pesquisa estrutural de proteínas

# Comparação de estruturas

- A estrutura tridimensional das proteínas pode ser comparada e o seu grau de **similaridade estrutural** avaliado (tal como comparamos as sequências).
- Existe uma relação clara entre **similaridade de estrutura** e **similaridade de sequência**: proteínas de sequência similar têm quase sempre estruturas similares.
- **A estrutura é mais conservada que a sequência**: proteínas de estrutura similar podem não ter sequências similares.

# A estrutura das proteínas é mais conservada que a sua sequência



**MAS**



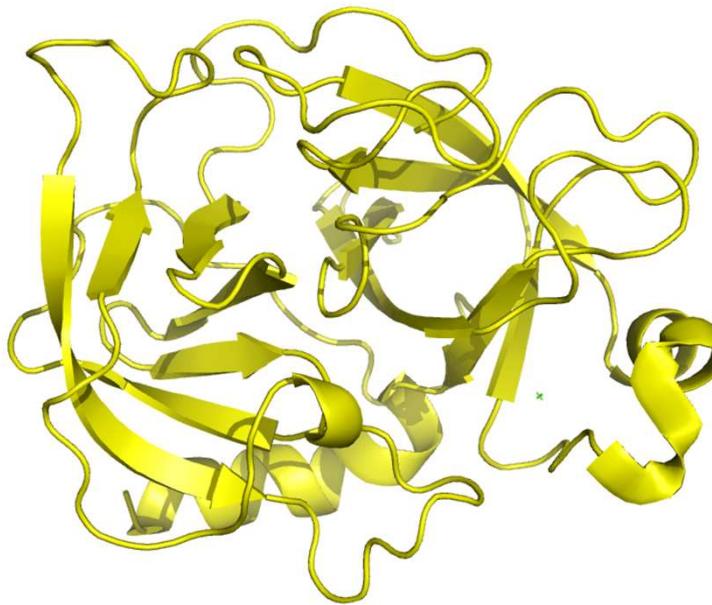
A pressão de selecção evolutiva opera sobre a estrutura (responsável pela função) e não directamente sobre a sequência. Alterações da sequência que conservem a estrutura são geralmente toleradas.

# Similaridade estrutural e de sequência

Tripsina bovina



Tripsina *S. griseus*

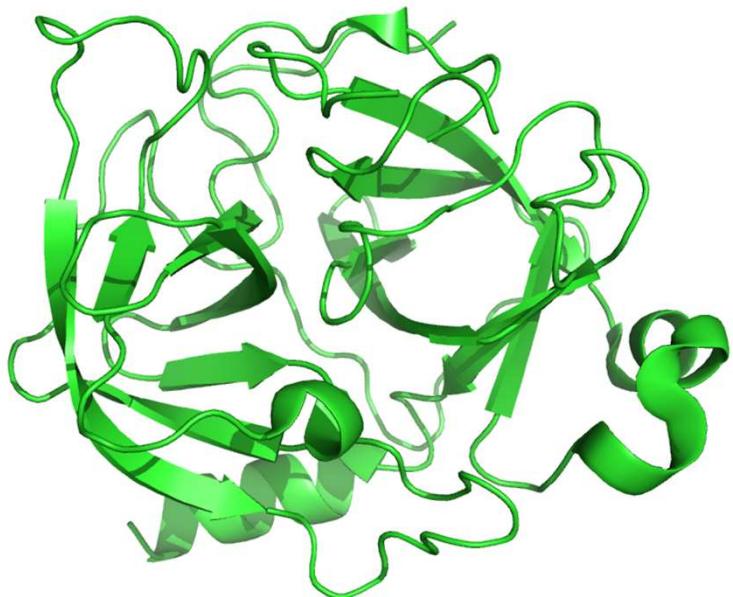


Alinhamento das sequências: **34%** identidade, E(1)  **$1.4 \times 10^{-17}$**

**Proteínas homólogas, similaridade de sequências  
claramente detectável**

# Similaridade estrutural e de sequênci

Tripsina bovina



Protease A *S. griseus*



Alinhamento das sequências: 20% identidade, E(1) **0.28**

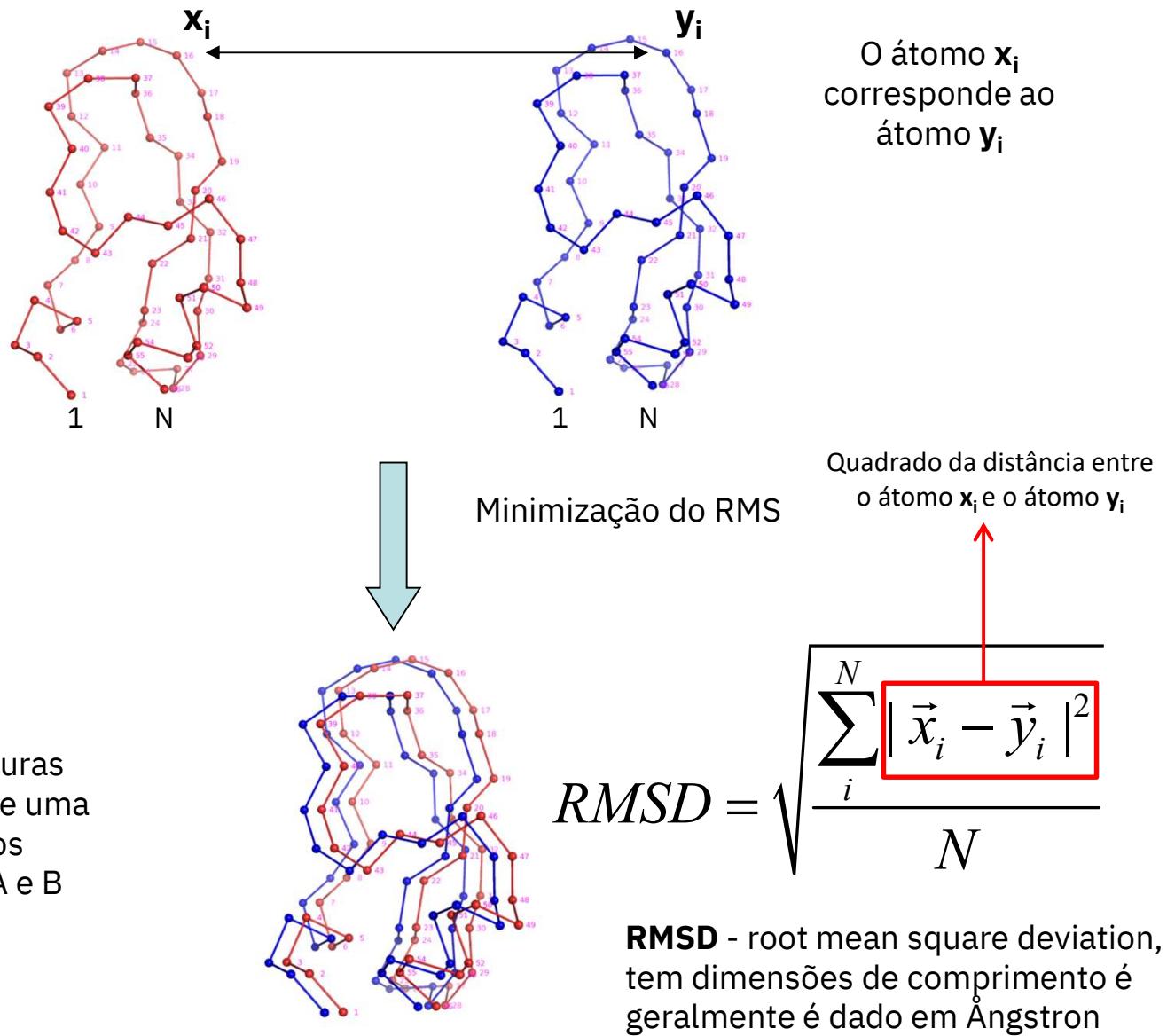
**Proteínas homólogas, similaridade de sequências  
não é detectável**

Alinhamento sem  
significado estatístico

# Como quantificar a similaridade estrutural ?

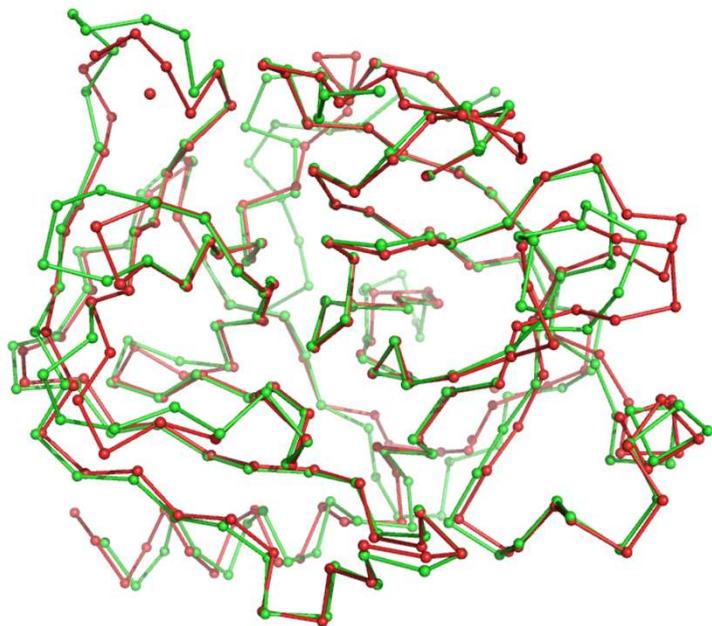
- Tal como a similaridade de sequências, a similaridade de estruturas pode ser quantificada usando diferentes medidas
- O método mais comum consiste em calcular o **desvio quadrático médio (RMSD)** entre pares de átomos das duas estruturas (geralmente expresso em Ångstrons ou nanómetros)
- O valor de RMSD depende da forma como se faz corresponder cada átomo da primeira estrutura a um átomo da segunda. Estabelecer esta correspondência não é um problema trivial, sobretudo para estruturas pouco semelhantes.

# Comparação de estruturas



# Relação entre RMSD e identidade de sequência

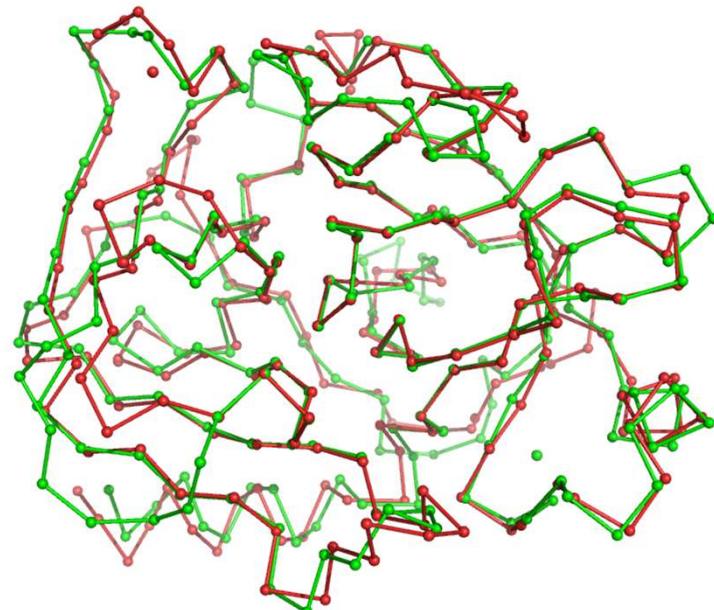
Tripsina humana  
*versus*  
Tripsina bovina



RMSD 0.8 Å

40% identidade de sequência

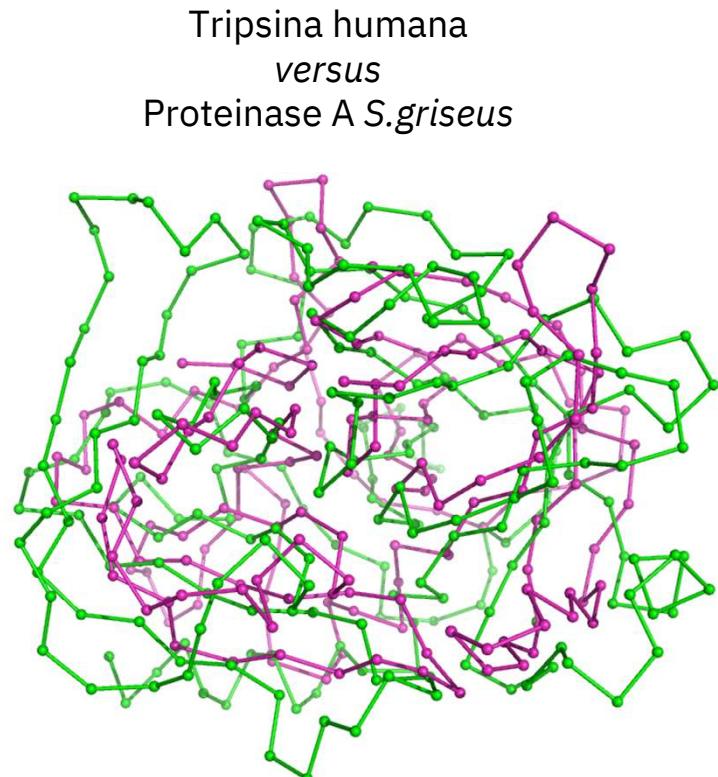
Tripsina humana  
*versus*  
Tripsina S.griseus



RMSD 1.8 Å

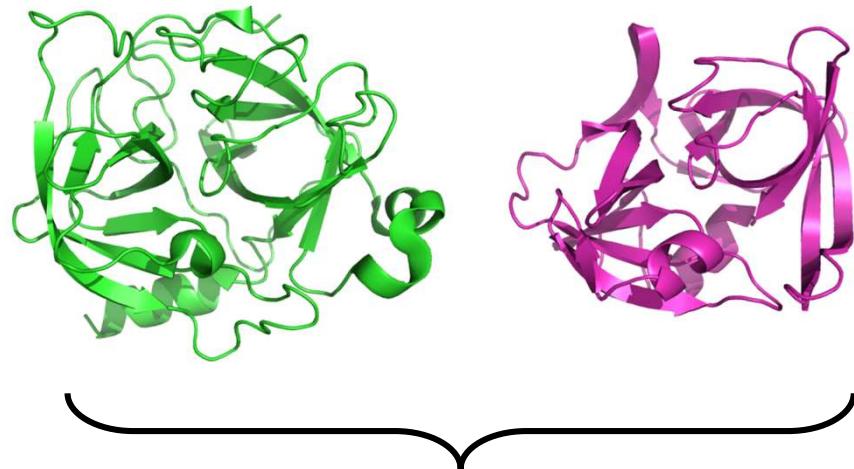
34% identidade de sequência

# Relação entre RMSD e identidade de sequência



RMSD 5.7 Å

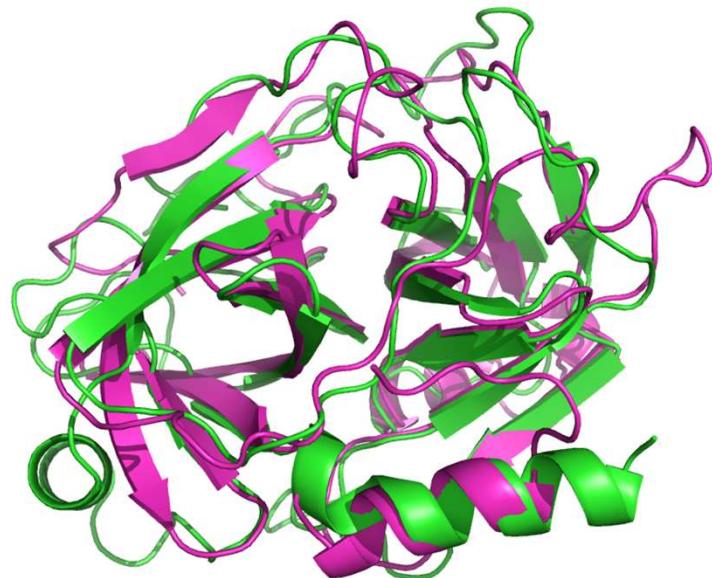
20% identidade de sequência



As duas proteínas têm clara  
semelhança estrutural, mas não é  
detectável por comparação de  
sequências

# Relação entre RMSD e identidade de sequência

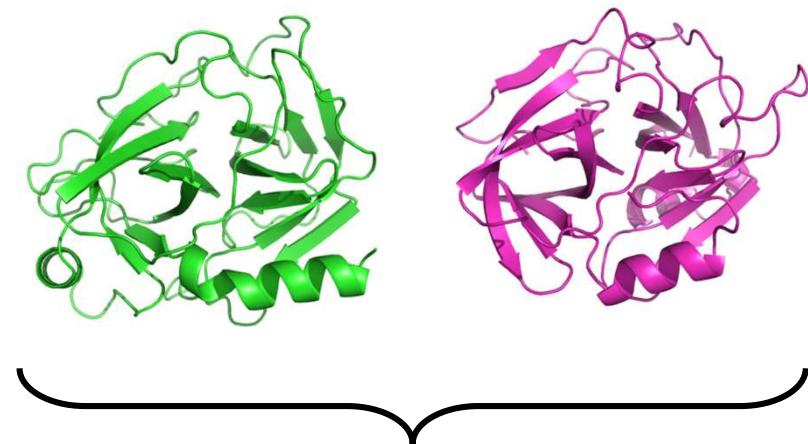
Tripsina humana  
*versus*  
Proteinase V8 *S.aureus*



RMSD 2.5 Å

19% identidade de sequência

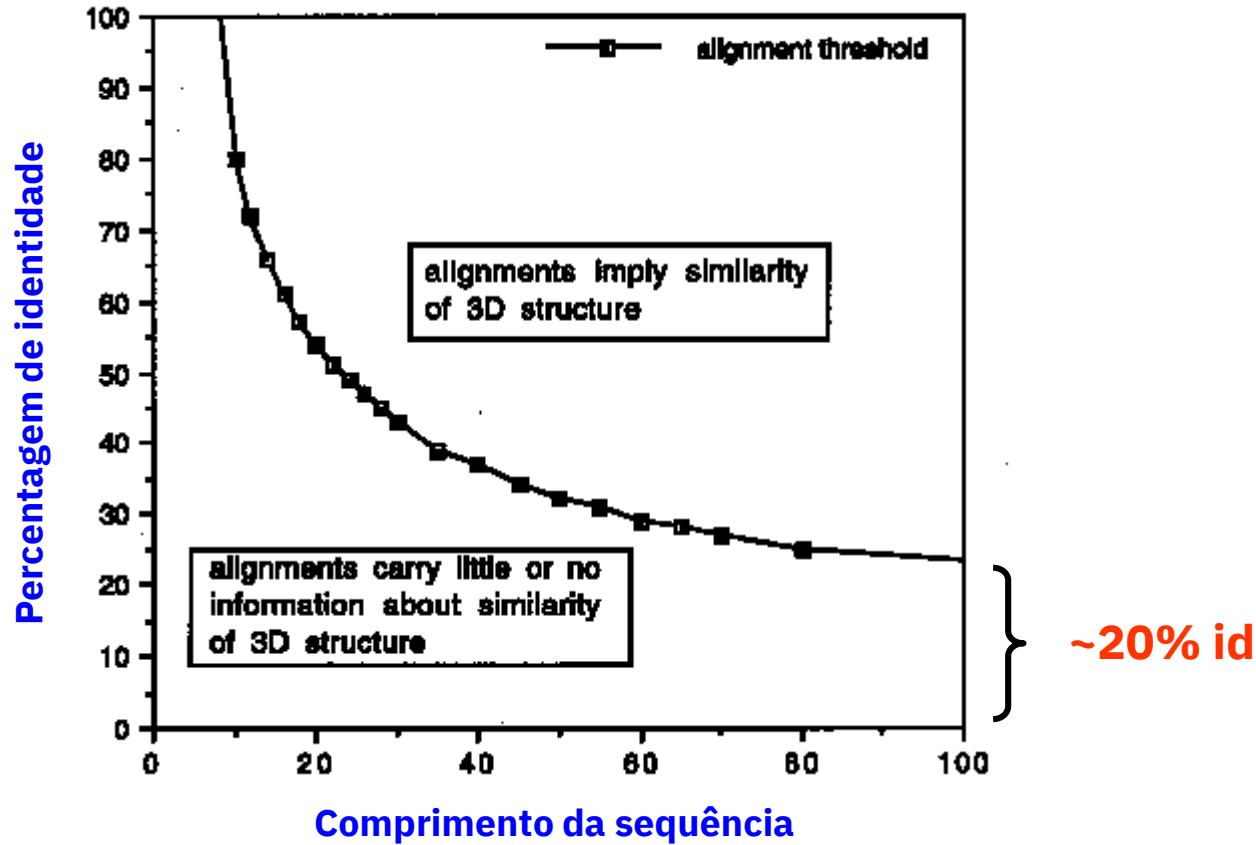
E-value:  $8.6 \times 10^2$



As duas proteínas têm clara  
semelhança estrutural, mas esta  
não é detectável por comparação  
das duas sequências

PDB files: 2RA3, 1WCZ

# Relação entre RMSD e identidade de sequência

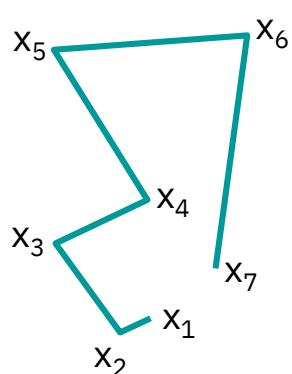


A relação entre a percentagem de identidade e a similaridade estrutural das proteínas depende do comprimento do alinhamento!

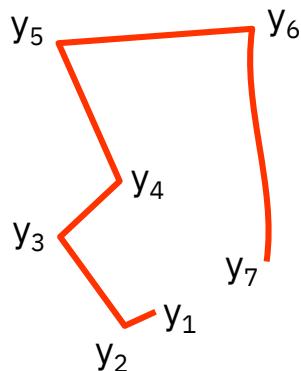
Para identidades inferiores a 20% não é, em geral, possível inferir existência de similaridade estrutural com base no alinhamento das sequências.

# Alinhamento estrutural

O alinhamento estrutural é em geral muito mais difícil que o alinhamento de sequências, pois é necessário estabelecer a correspondência entre os átomos que minimiza o RMS

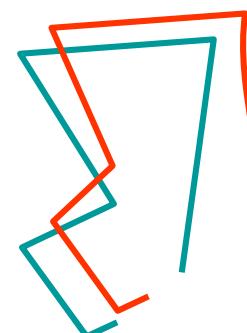


+

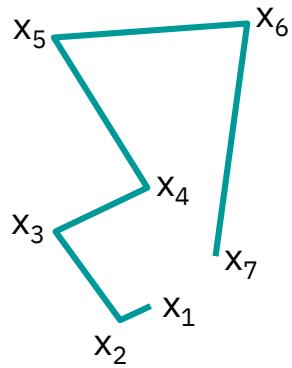


**Fácil**

a correspondência  
é óbvia



$x_1 \equiv Y_1$   
 $x_2 \equiv Y_2$   
 $x_3 \equiv Y_3$   
 $x_4 \equiv Y_4$   
...  
 $x_7 \equiv Y_7$



**Difícil!**

a correspondência  
NÃO é óbvia

?

## Sites para comparação e pesquisa estrutural

- PDBeFold @ EBI (P,C, M): <https://www.ebi.ac.uk/msd-srv/ssm/>
- Top Match (C): <https://topmatch.services.came.sbg.ac.at/>
- DALI Server (P,C): [http://ekhidna.biocenter.helsinki.fi/dali\\_server](http://ekhidna.biocenter.helsinki.fi/dali_server)
- VAST (P): <http://www.ncbi.nlm.nih.gov/Structure/VAST/>
- VAST+ (P): <http://www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi>
- Deep Align (M) - <http://raptorx.uchicago.edu/DeepAlign/submit/>

P – pesquisa

C – comparação    M – alinhamento múltiplo

# Pesquisa de estruturas: DALI server

Pretendemos encontrar estruturas semelhantes a uma determinada estrutura, neste caso a uma estrutura do PDB (do enzima lisozima) cujo código é **2LZT**.

The screenshot shows the DALI server interface. At the top, the title "DALI PROTEIN STRUCTURE COMPARISON SERVER" is displayed. Below it, a navigation bar includes links for About, PDB search, PDB25, Pairwise, All against all, Gallery, References, Statistics, Tutorial, and Download. The main section is titled "PDB search" and contains instructions: "Compare query structure against Protein Data Bank." A text input field labeled "STEP 1 - Enter your query" contains the text "2LZT ≡ código da lisozima". A red arrow points to this input field. Below it, another input field contains "2LZT" and has a red border, with the text "OR upload file Choose File No file chosen" next to it. A "STEP 2 - Optional data" section follows, with a "Job title" input field containing "Pesquisa". A red arrow points to the "Enviar a pesquisa" button at the bottom of this section. The final "STEP 3 - Submit your job" section contains "Submit" and "Clear" buttons, with a red border around the "Submit" button. A note at the bottom states: "If the same structure has been submitted recently, you will be redirected to the result page of the previous instance."

ekhidna2.biocenter.helsinki.fi/bar X +

Not secure | ekhidna2.biocenter.helsinki... Q ☆

Results:

Chain: 2lztA

- [Matches against PDB25. Correlation matrix](#)
- [Matches against PDB50](#)
- [Matches against PDB90](#)
- [Matches against full PDB](#)
- [Download matches against PDB25](#)
- [Download matches against PDB50](#)
- [Download matches against PDB90](#)
- [Download matches against full PDB](#)

Results will be deleted after one week.





# Pesquisa de estruturas similares no VAST+

Pretendemos encontrar estruturas semelhantes a uma determinada estrutura, neste caso a uma estrutura do PDB (do enzima lisozima) cujo código é **2LZT**.

The screenshot shows a web browser window with three tabs open: "Similar Structure Assi", "RCSB PDB - Structure", and "RCSB Protein Data B". The main content area is titled "VAST+ Similar Structures" and describes the tool as identifying macromolecules with similar 3D structures, focusing on biological assemblies. A search bar at the top contains the text "2LZT" with a red arrow pointing to it from the left. Below the search bar, a button labeled "Search" is visible. The page also includes sections for "Citing VAST" with two references and links for "Help Desk", "Disclaimer", "Privacy statement", and "Accessibility".

VAST+ is a tool designed to identify macromolecules that have similar 3-dimensional structures, with an emphasis on finding those with similar biological assemblies ("biological units" or "biounits"). The similarities are calculated using purely geometric criteria, and therefore can identify distant homologs that cannot be recognized by sequence comparison.

Input a valid PDB ID or MMDB ID:  2LZT = código da lisozima

To use VAST+, enter the PDB ID or MMDB ID of any structure that is currently in the Molecular Modeling Database (MMDB). VAST+ will display a list of similar structures, ranking them by the extent of their similarity to the query structure's biological unit. [more...](#)

Citing VAST

Gibrat JF, Madej T, Bryant SH. "Surprising similarities in structure comparison.", *Curr Opin Struct Biol.* 1996 Jun;6(3): 377-85.  
Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, Bryant SH. "MMDB and VAST+: tracking structural similarities between macromolecular complexes." *Nucl. Acids Res.* 2014 Jan;42(Database issue):D297-303.

| Help Desk | Disclaimer | Privacy statement | Accessibility |

VAST+ Similar Struct × RCSB PDB - Structure × RCSB Protein Data Bank ×

www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi?uid=2LZT

Apps Enzymology Piano Music Production Bioinformatics Databases Bioinformatics Tools Misc Other bookmarks

**NCBI**  
National Center for Biotechnology Information

## VAST+ Similar Structures

3D structural similarities among biological assemblies

HOME SEARCH GUIDE Structure Home 3D Macromolecular Structures Conserved Domains BioSystems Help

PDB ID or MMDB ID New Search

**Refinement Of Triclinic Lysozyme. II. The Method Of Stereochemically Restricted Least-Squares**

MMDB ID: 58091 (PDB ID: 2LZT)  
 Biological unit 1: monomeric  
 Source organism: Gallus gallus  
 Number of proteins: 1 (HEN EGG WHITE LYSOZYME)  
 Number of chemicals: 5 (Nitrate Ion (5) ▾)

Similar Structures Original VAST ▾

▼ Display filters ▾

Showing 1 to 10 out of 860 structures ▾

PDB ID		Description	Taxonomy	Aligned Protein	RMSD	Aligned Residues	Sequence Identity
1	1LZN	Neutron Structure Of Hen Egg-White Lysozyme	Gallus gallus	1	0.10 Å	120	100%
2	4LZT	Atomic Resolution Refinement Of Triclinic Hen Lysozyme At 295k	Gallus gallus	1	0.10 Å	120	100%
3	1V7S	Triclinic Hen Lysozyme Crystallized At 313k From A D2O Solution	Gallus gallus	1	0.11 Å	120	100%
4	1LKS	Hen Egg White Lysozyme Nitrate	Gallus gallus	1	0.12 Å	120	100%
5	2F2N	Triclinic Hen Egg Lysozyme Cross-linked By Glutaraldehyde	Gallus gallus	1	0.12 Å	120	100%
6	2F30	Triclinic Cross-linked Lysozyme Soaked With 4.5m Urea	Gallus gallus	1	0.17 Å	120	100%
7	4MWK	Triclinic Hewl Co-crystallised With Cisplatin, Studied At A Data Collection Temp...	Gallus gallus	1	0.22 Å	120	100%
8	2F4G	Triclinic Cross-linked Lysozyme Soaked In Bromoethanol 1m	Gallus gallus	1	0.24 Å	120	100%
9	4MWM	Triclinic Hewl Co-crystallised With Cisplatin, Studied At A Data Collection Temp...	Gallus gallus	1	0.25 Å	120	100%
10	2VB1	Hewl At 0.65 Angstrom Resolution	Gallus gallus	1	0.31 Å	120	100%

Show 10 structures First Previous Page 1 of 86 Pages Next Last

Citing VAST

Gibrat JF, Madej T, Bryant SH. "Surprising similarities in structure comparison.", *Curr Opin Struct Biol.* 1996 Jun;6(3): 377-85.  
 Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, Bryant SH. "MMDB and VAST+: tracking structural similarities between macromolecular complexes." *Nucl. Acids Res.* 2014 Jan;42(Database issue):D297-303.

VAST+ Similar Struct x

www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi?uid=2lzt

Apps Enzymology Piano Music Production Bioinformatics Databases Bioinformatics T... Misc Programming D pmartel Other bookmarks

**NCBI** National Center for Biotechnology Information

**VAST+ Similar Structures** 3D structural similarities among biological assemblies

HOME SEARCH GUIDE Structure Home 3D Macromolecular Structures Conserved Domains BioSystems Help PDB ID or MMDB ID New Search

Refinement Of Triclinic Lysozyme. II. The Method Of Stereochemically Restricted Least-Squares

MMDB ID: 58091 (PDB ID: 2LZT)  
 Biological unit 1: monomeric  
 Source organism: Gallus gallus  
 Number of proteins: 1 (HEN EGG WHITE LYSOZYME)  
 Number of chemicals: 5 (Nitrate Ion (5) ▾)

Similar Structures [?] Original VAST [?]

▼ Display filters [?]

Showing 771 to 780 out of 860 structures [?]

PDB ID	Description	Taxonomy	Aligned Protein	RMSD	Aligned Residues	Sequence Identity
771 [+] 1JA2	Binding Of N-Acetylglucosamine To Chicken Egg Lysozyme: A Powder Diffractio...	Gallus gallus	1	1.29 Å	90	100%
772 [+] 2FYD	Catalytic Domain Of Bovine Beta 1, 4-Galactosyltransferase In Complex With Alp...	Bos taurus/Mus ...	1	1.30 Å	96	39%
773 [+] 1NQI	Crystal Structure Of Lactose Synthase, A 1:1 Complex Between Beta1,4- Galacto...	Bos taurus/Mus ...	1	1.30 Å	97	38%
774 [+] 1AM7	Lysozyme From Bacteriophage Lambda	Enterobacteria p...	1	1.30 Å	37	22%
775 [+] 1FKQ	Recombinant Goat Alpha-Lactalbumin T29v	Capra hircus	1	1.30 Å	96	45%
776 [+] 3B00	Crystal Structure Of Alpha-lactalbumin	Homo sapiens	1	1.30 Å	98	40% <span style="border: 2px solid red; padding: 2px;">(highlighted)</span>
777 [+] 1NMM	Beta-1,4-Galactosyltransferase Mutant Cys342thr Complex With Alpha- Lactalb...	Bos taurus/Mus ...	1	1.31 Å	96	39%
778 [+] 1A2Y	Hen Egg White Lysozyme, D18a Mutant, In Complex With Mouse Monoclonal An...	Gallus gallus/Mus...	1	1.31 Å	118	99%
779 [+] 1NWG	Beta-1,4-Galactosyltransferase Complex With Alpha- Lactalbumin And N-Butan...	Bos taurus/Mus ...	1	1.31 Å	97	38%
780 [+] 1PZY	W314a-Beta1,4-Galactosyltransferase-I Complexed With Alpha-Lactalbumin In ...	Bos taurus/Mus ...	1	1.31 Å	96	39%

Show 10 structures First Previous Page 78 of 86 Pages Next Last

Citing VAST

Gibrat JF, Madej T, Bryant SH. "Surprising similarities in structure comparison.", *Curr Opin Struct Biol.* 1996 Jun;6(3): 377-85.  
 Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, Bryant SH. "MMDB and VAST+: tracking structural similarities between macromolecular complexes." *Nucl. Acids Res.* 2014 Jan;42(Database issue):D297-303.

VAST+ Similar Struct x

2LZT neighbors - Cn3D 4.3 plus.cgi?uid=2lzt

File View Select Style Window CDD Help

2LZT and 3B0O sequence alignment - Google Chrome www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi?cmd=d&ids=58091,1,1,100429,1,1

Aligned Sequences Close

Visualize 3D structure superposition with Cn3D

2LZT\_A: HEN EGG WHITE LYSOZYME  
3B0O\_A: ALPHA-LACTALBUMIN

	10	20	30	40	50	60
2LZT_A	1 KVFGRCLEAAAMKrhGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDgSTDYGILQINS	60				
3B0O_A	1 MQFTKCELSQLLK--DIDGYGGIALPELICTMFHHTSGYDTQAIVENNE-STEYGLFQISN	57				
	70	80	90	100		
2LZT_A	61 RWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSD	101				
3B0O_A	58 KLWCKSSQVPQSRNICDISCDKFLDDDITDDIMCAKKILDI	98				

Query structure  
MMDB ID: 58091 (PDB ID: 2LZT)

Matched structure  
MMDB ID: 100429 (PDB ID: 3B0O)

\*Click schematic circles and molecule names to view matches

775 [+] ● 1FKQ  
776 [-] ● 3B0O

Visualize 3D structure superposition with Cn3D View aligned sequences

777 [+] ● 1NMM Beta-1,4-Galactosyltransferase Mutant Cys342thr Complex With Alpha-Lactalb... Bos taurus/Mus ... 1 1.31Å 96 39%

778 [+] ● 1A2Y Hen Egg White Lysozyme, D18a Mutant, In Complex With Mouse Monoclonal An... Gallus gallus/Mus... 1 1.31Å 118 99%

The screenshot shows the VAST+ software interface. On the left, a window titled '2LZT neighbors - Sequence/Alignment V' displays sequence alignments between 2LZT\_A (Hen Egg White Lysozyme) and 3B0O\_A (Alpha-Lactalbumin). It includes a sequence viewer with numbered positions 1 through 100, color-coded amino acids, and a 'View' menu. Below the viewer are two entries: 775 (1FKQ) and 776 (3B0O). On the right, a larger window titled '2LZT and 3B0O sequence alignment - Google Chrome' shows the aligned sequences. The top part of the alignment table has columns for positions 10, 20, 30, 40, 50, and 60. The bottom part has columns for 70, 80, 90, and 100. A 'Close' button is at the top right of this window. At the bottom of the alignment table, it says 'Query structure MMDB ID: 58091 (PDB ID: 2LZT)' and 'Matched structure MMDB ID: 100429 (PDB ID: 3B0O)'. Below the alignment table, there's a note: '\*Click schematic circles and molecule names to view matches'. Two schematic molecules are shown: one for 'HEN EGG WHITE LYSOZYME' (a central purple circle with four white diamond branches) and one for 'ALPHA-LACTALBUMIN' (a single purple circle with a single white diamond branch). Below these molecules are two buttons: 'Visualize 3D structure superposition with Cn3D' and 'View aligned sequences'. At the very bottom, there are two more entries: 777 (1NMM) and 778 (1A2Y), each with its own set of details.

# Pesquisa estrutural com (original)VAST

Vast Neighbor Summary x RCSB PDB - Structure x RCSB Protein Data Bank x

www.ncbi.nlm.nih.gov/Structure/vast/vastsrv.cgi?sdid=242541

Apps Enzymology Piano Music Production Bioinformatics Databases Bioinformatics Tools Misc Programming

**NCBI VAST Similar Structures**

PubMed BLAST Structure Taxonomy OMIM Help? Cn3D

VAST related structures for: MMDB 58091, 2LZT sequence A.

**Overview:** There are two main sections to this page. The first section consists of the alignment view controls, the list controls, and the advanced related structure search controls. The second section is the VAST related structure list itself.

View 3D Alignment of All Atoms with Cn3D Display Download Cn3D!

View Sequence Alignment using Hypertext for Selected VAST related structures

List Medium redundancy subset, sorted by Aligned Length in Graphics

Advanced related structure search

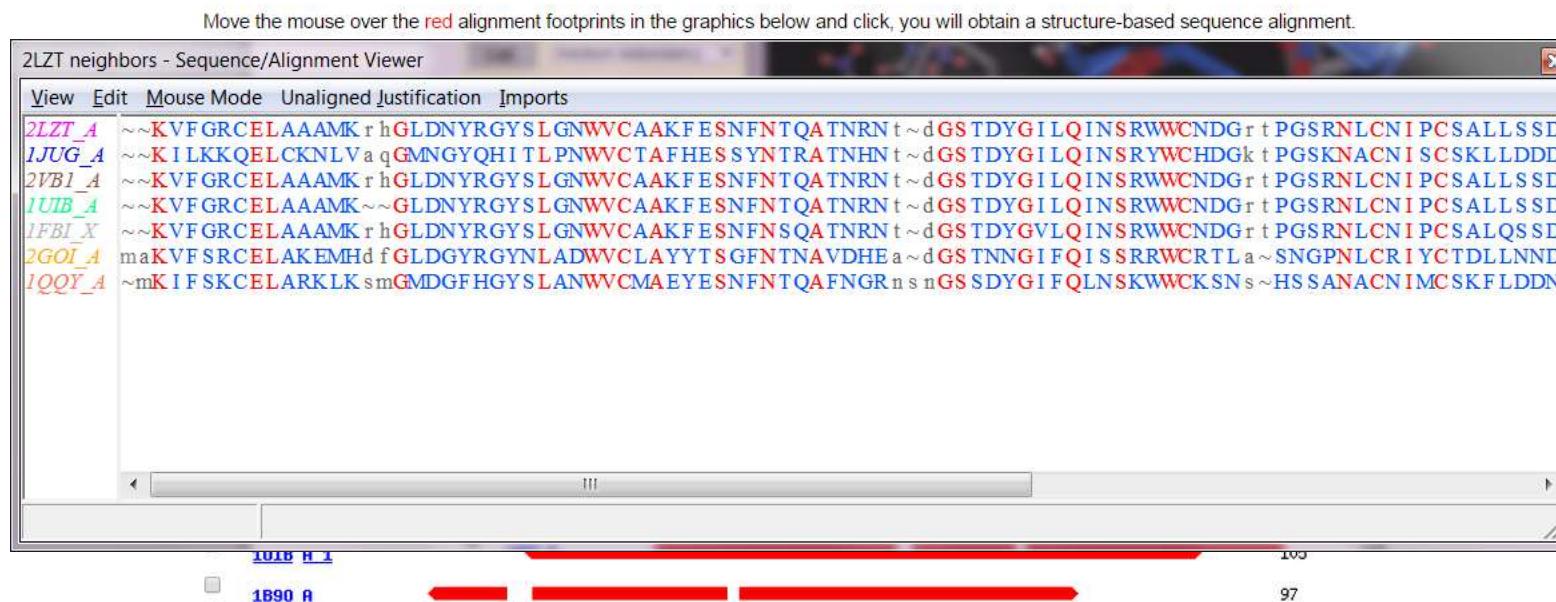
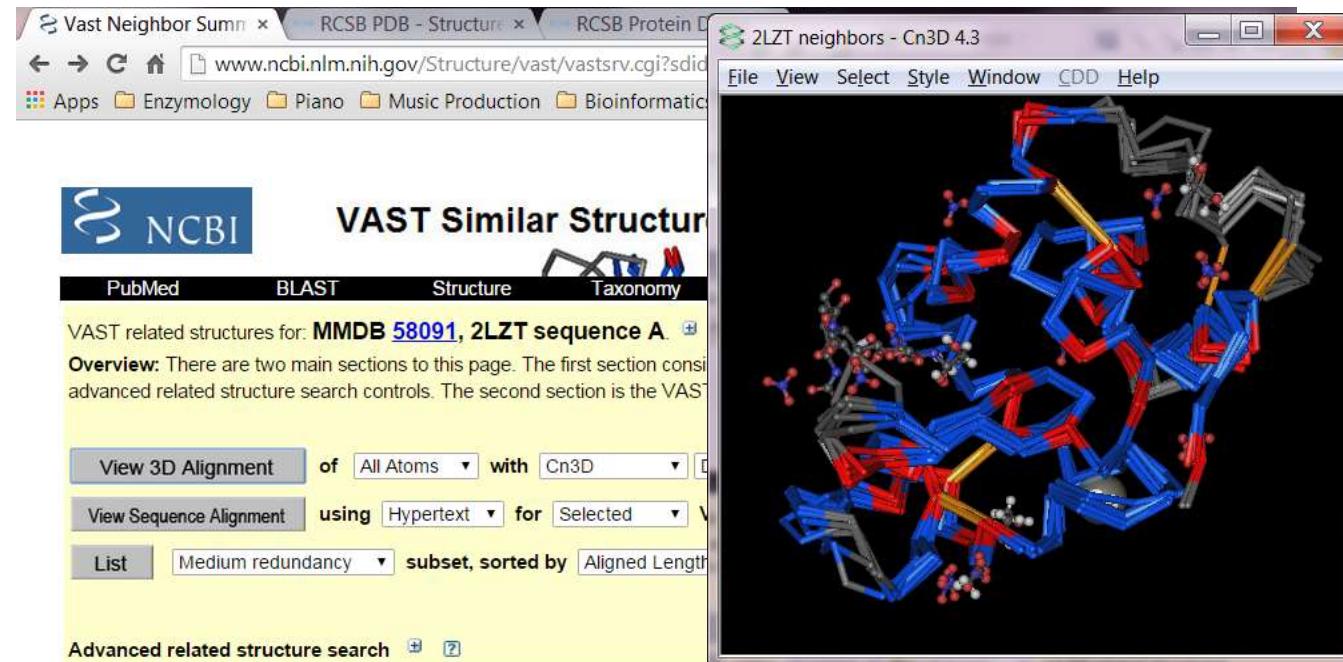
Move the mouse over the red alignment footprints in the graphics below and click, you will obtain a structure-based sequence alignment.

Total related structures: 1593; 1 - 60 of 122 representatives from the Medium redundancy subset displayed. Page: 1

Click to: [Check All](#) [Uncheck All](#)

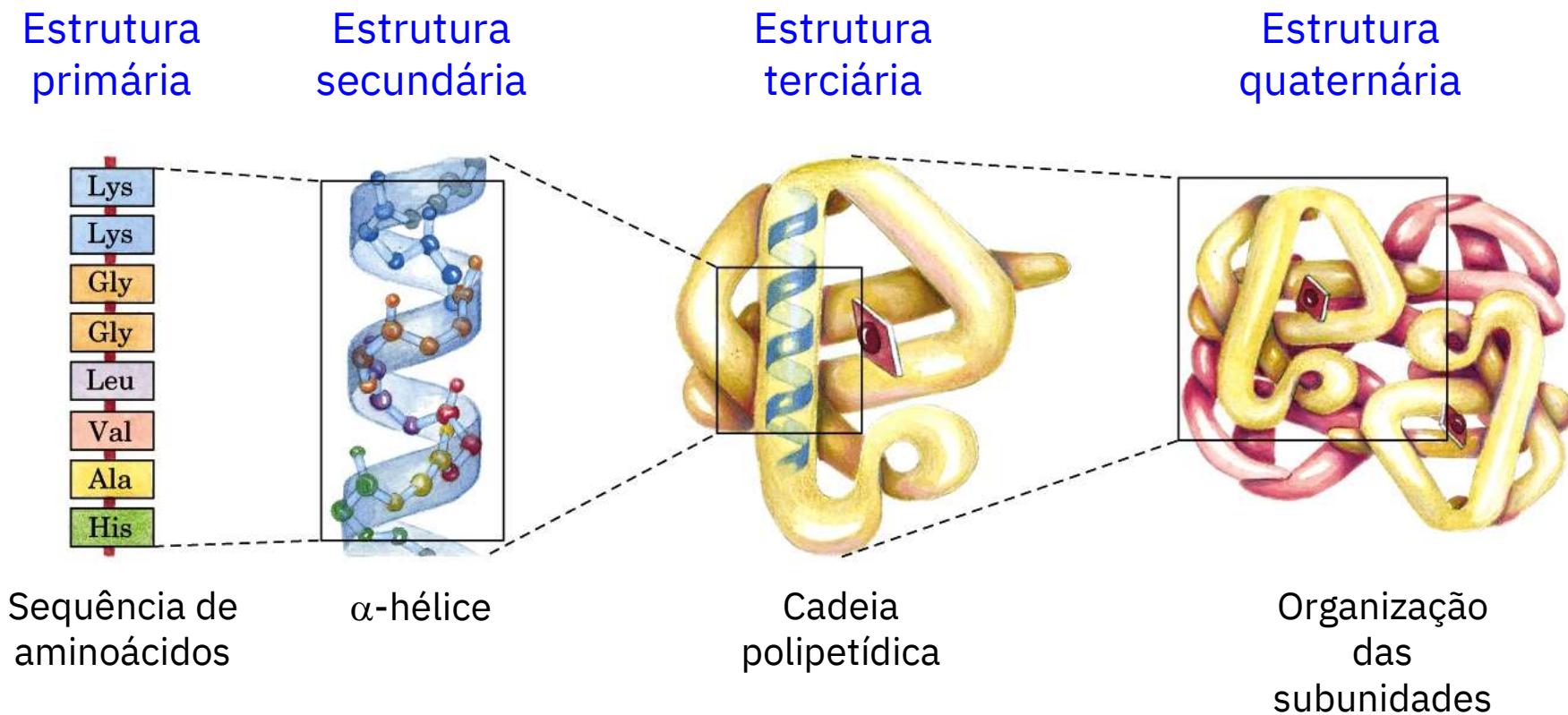
Structure ID	Aligned Length
2LZT_A	129
3D Domains	Ali_len
<input checked="" type="checkbox"/> 1JUG_A	120
<input checked="" type="checkbox"/> 2VB1_A	120
<input checked="" type="checkbox"/> 1FB1_X	118
<input checked="" type="checkbox"/> 1U1B_A	118
<input checked="" type="checkbox"/> 2GOT_A	117
<input checked="" type="checkbox"/> 1QQY_A	115
<input type="checkbox"/> 2GOT_A_2	109
<input type="checkbox"/> 1U1B_A_1	105
<input type="checkbox"/> 1B90_A	97

# Visualização do alinhamento com o software Cn3D



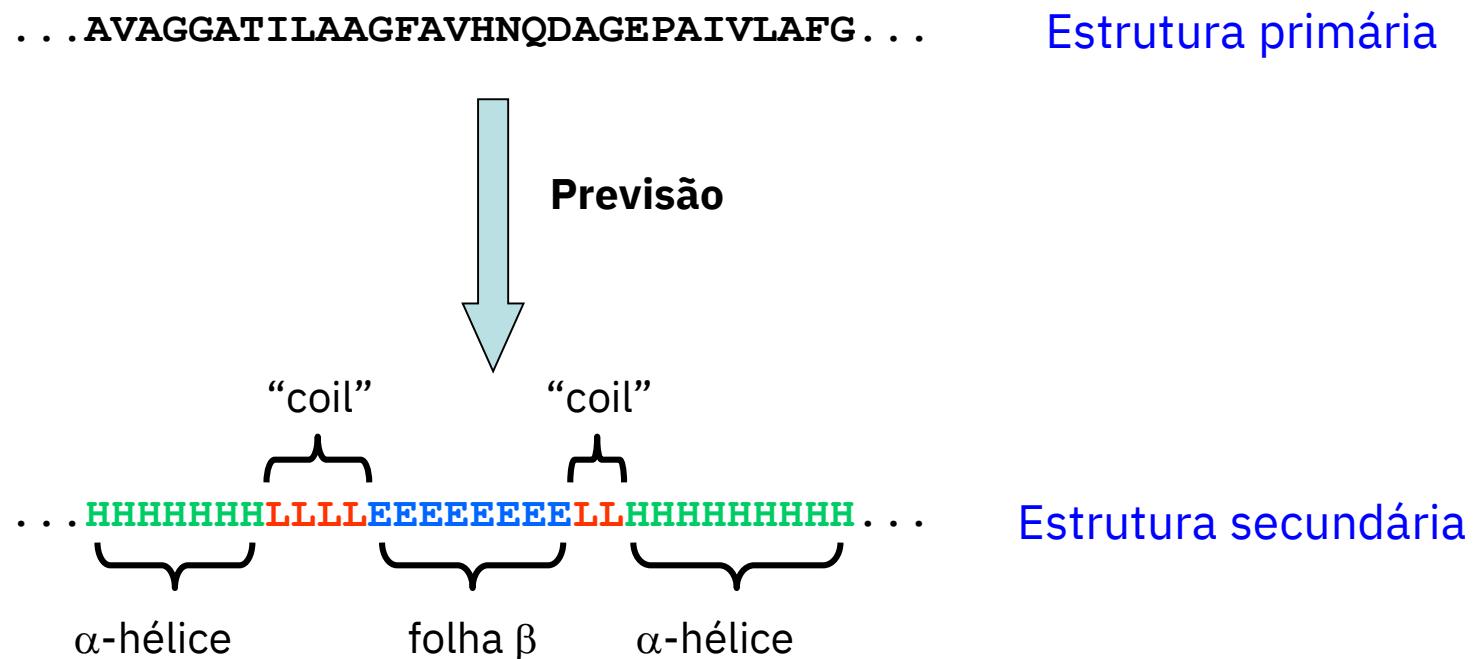
# Previsão da estrutura secundária das proteínas

# Níveis de organização da estrutura das proteínas



# O problema da previsão da estrutura secundária

Dada a **sequência** de uma proteína, pretende-se identificar as regiões dessa proteína que adotam diferentes tipos de **estrutura secundária**. Este problema é consideravelmente mais simples que deduzir a estrutura tridimensional completa da proteína (previsão da estrutura terciária). Atualmente conseguem-se precisões na ordem dos 75%-85%, dependendo do tipo de proteínas em análise.



# Métodos de previsão da estrutura secundária

- **Chou-Fasman & GOR** - baseiam-se na análise das frequências de cada um dos 20 aminoácidos nos vários tipos de estrutura secundária. (Precisão: 50-60%)
- **NN (Neural network)** - Usam um modelo de **rede neural** que é treinada para aprender a reconhecer a estrutura secundária a partir da sequência de aminoácidos. A rede neural é primeiramente “ensinada” com um conjunto de sequências e respectivas estruturas secundárias (training set), passando depois a ser capaz de prever a estrutura para sequências que não fazem parte do training set. (Precisão: ~70-85%)

<http://www.predictprotein.org/> (O programa PHD faz parte das ferramentas neste site)

<http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html> (**NNPredict**)

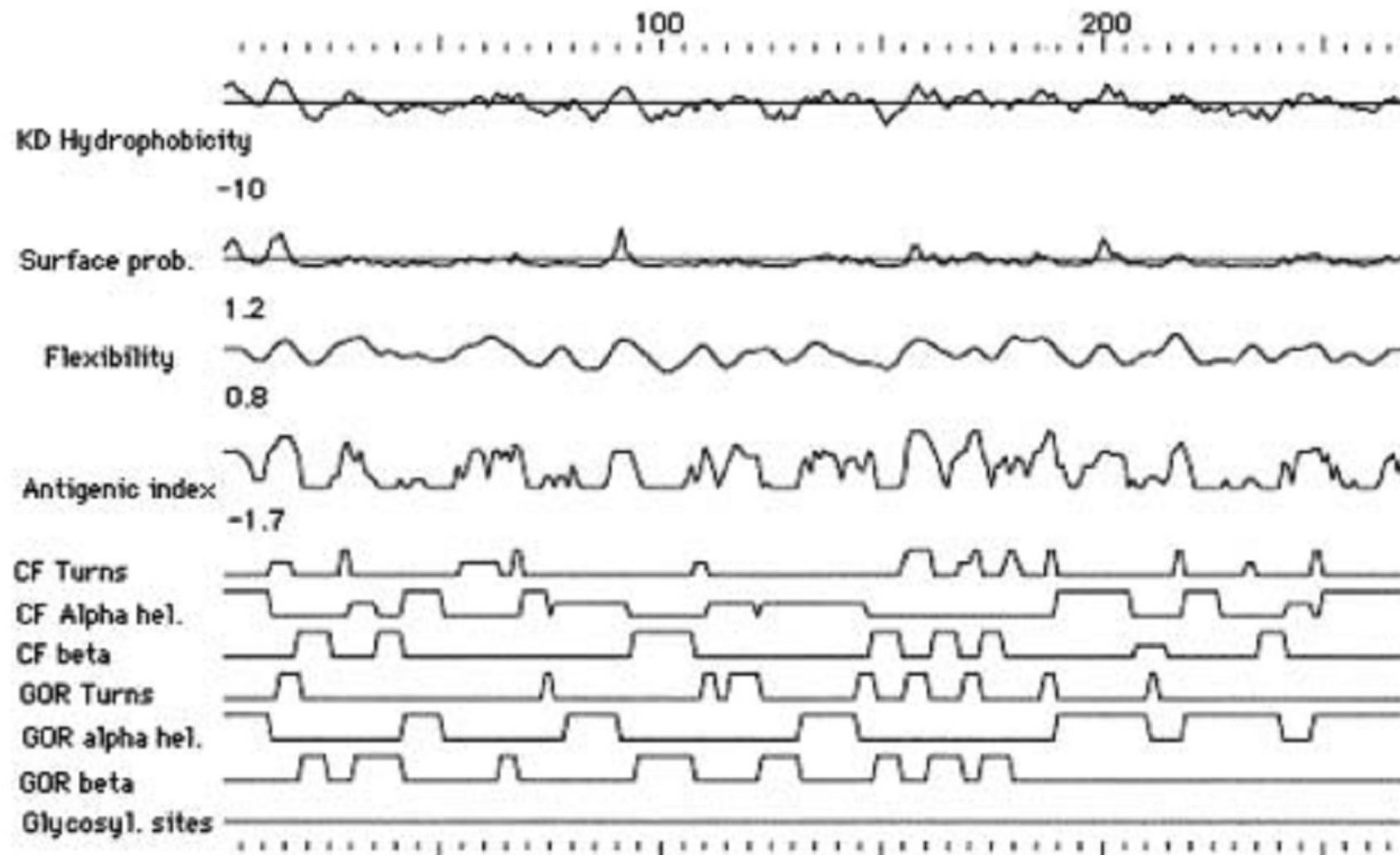
- **Nearest-neighbor** - este método baseia-se na comparação da sequência a prever com sequências de estrutura conhecida. (Precisão: 70-75%)

<http://bioweb.pasteur.fr/seqanal/interfaces/predator.html> (**PREDATOR**)

<http://bioweb.pasteur.fr/seqanal/interfaces/nssp.html> (**NSSP**)

Program	Web address	Method	Reference
Baylor College of Medicine (BCM) DSC J-Pred structure prediction server	<a href="http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html">http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html</a> <a href="http://www.bmm.icnet.uk/dsc/">http://www.bmm.icnet.uk/dsc/</a> <a href="http://jura.ebi.ac.uk:8888/">http://jura.ebi.ac.uk:8888/</a>	collection of methods and linked to other servers linear discrimination NNSSP, DSC, Predator, Mulpred, <sup>b</sup> Zpred, <sup>c</sup> Jnet, <sup>e</sup> and PHD	see Web site and text King et al. (1997) Cuff et al. (1998); and see text
NNPRED	<a href="http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html">http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html</a>	neural networks enhanced to detect sequence periodicity	Kneller et al. (1990)
NPS@ server, MLR combination for secondary structure prediction <sup>a</sup>	<a href="http://pbil.ibcp.fr/NPSA/">http://pbil.ibcp.fr/NPSA/</a>	combination of prediction methods using multivariate linear regression to optimize the predictions	Guermeur et al. (1999)
Protein Sequence Analysis (PSA) System <sup>d</sup>	<a href="http://bmrc-www.bu.edu/psa/index.html">http://bmrc-www.bu.edu/psa/index.html</a>	discrete space models (hidden Markov models) for patterns of $\alpha$ helices, $\beta$ strands, tight turns, and loops in specific structural classes	Stultz et al. (1993, 1997); White et al. (1994)
PREDATOR	<a href="http://www.embl-heidelberg.de/argos/predator/predator_info.html">http://www.embl-heidelberg.de/argos/predator/predator_info.html</a>	based on analysis of long- and short-range amino acid interactions and alignments of sequence pairs	Frishman and Argos (1995, 1996, 1997)
Predict Protein server	<a href="http://www.embl-heidelberg.de/predictprotein/predictprotein.html">http://www.embl-heidelberg.de/predictprotein/predictprotein.html</a> ; see also mirror sites	neural networks of multiple sequence alignment	Rost and Sander (1994); Rost (1996)
PSSP	<a href="http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html">http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html</a>	nearest neighbor enhanced by non-intersecting local and multiple sequence alignments	Salamov and Solovyev (1995, 1997)
Simpa96 SOPM, SOPMA	<a href="http://pbil.ibcp.fr/NPSA/">http://pbil.ibcp.fr/NPSA/</a> <a href="http://pbil.ibcp.fr/NPSA/">http://pbil.ibcp.fr/NPSA/</a>	nearest-neighbor method nearest-neighbor method based on sequence alignments	Levin (1997) Geourjon and Deleage (1994, 1995)
SSP	<a href="http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html">http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html</a>	linear discriminant analysis based on amino acid composition of local and adjacent regions	see H option for this program on Web page
UCLA-DOE structure prediction server	<a href="http://www.doe-mbi.ucla.edu/people/frsver/frsver.html">http://www.doe-mbi.ucla.edu/people/frsver/frsver.html</a>	collection of methods and linked to other servers	Fischer and Eisenberg (1996)

## Previsão GOR e Chou-Fassman com o programa GCG



# Exemplo de previsão com o programa PHD

**Rel:** fiabilidade global da previsão (0-9)

```
.....1.....2.....3.....4.....5.....6
AA |MERYENLFAQLNDRREGAFVPFVTLGDPGIEQSLKIIDTLIDAGADALELGVPFSDPLAD|
PHD | HHHHHHHHHHHH EEEEEEEE HHHHHHHHHHHHHH EEEE |
Rel |93489999999634887279984248998458799999997399668944767784689|
detail:
prH-|03689999998753110000000000001678899999998300000000001113210
prE-|0000000000000000579886530000000000000000000017886311000000
prL-|963100000123688831001236899832110000000169972003687786789
subset: SUB |L..HHHHHHHHH..LLL.EEEE...LLL.HHHHHHHHHHHH.LLEEE..LLLL.LLL|


.....7.....8.....9.....10.....11.....1
AA |GPTIQNALRAFAAGVTPAQCFCMLALIREKHPTIPIGLMYANLVFNNGIDAFYARCEQ|
PHD | HHHHHHHHHHHH HHHHHHHHHHHH EEEEEEE HHHH HHHHHHHHHH |
Rel |73789999999982896299999999972899938998834224412559999999999|
detail:
prH-|25889999999984111589999999998410000000000235664326999999998
prE-|000000000000000000000000000000000000000000000000000000000000
prL-|741100000000158873000000000015899861000135542245720000000000
subset: SUB |L.HHHHHHHHHHHH.LLL.HHHHHHHHHHHH.LLLL.EEEE.....LHHHHHHHHHHH|
```

**prH:** probabilidade do resíduo estar em conformação de hélice (0-9)

**prE:** probabilidade do resíduo estar em conformação de folha beta (0-9)

**prL:** probabilidade do resíduo estar em conformação de “coil” (0-9)

# Modelação da estrutura terciária por homologia

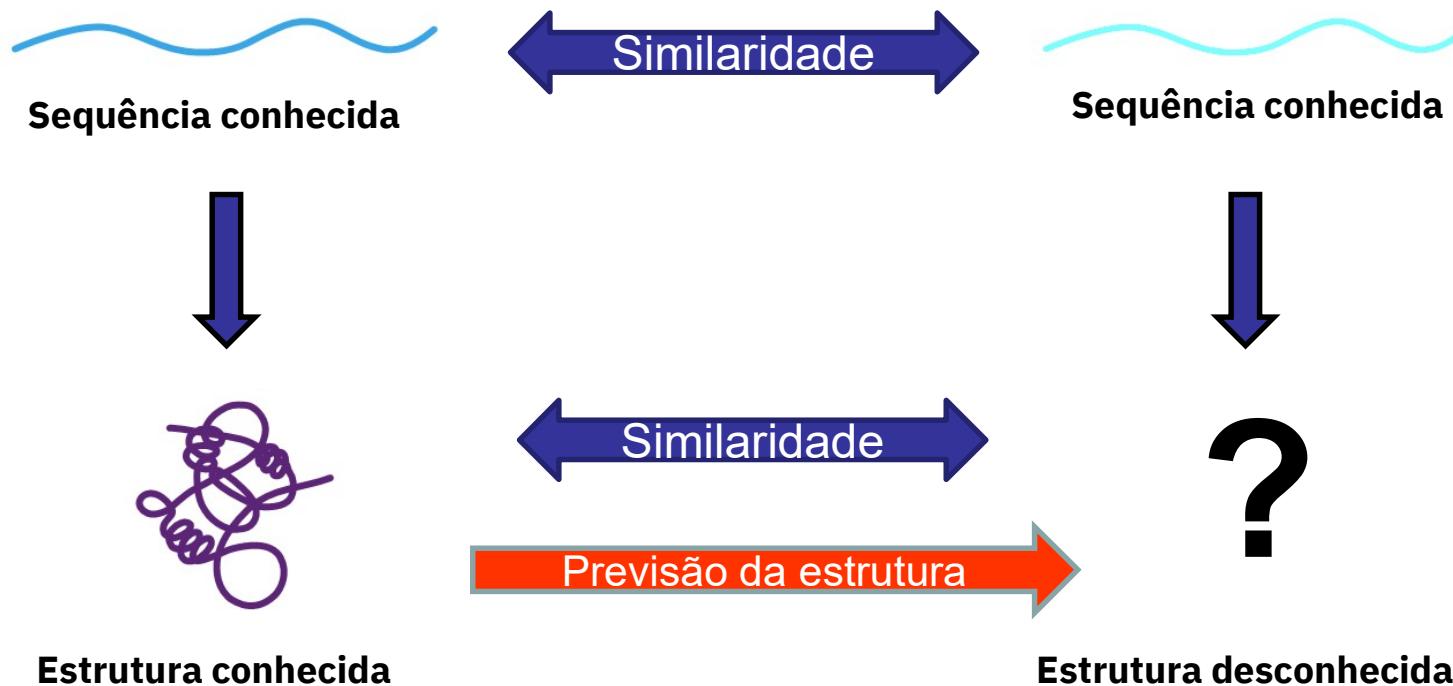
## Modelação por homologia

- A previsão da estrutura tridimensional de uma proteína a partir da sua sequência é extremamente importante, já que o número de sequências conhecidas (~1000000) excede largamente o de estruturas (~20000).
- Dos vários métodos para previsão de estrutura, a modelação por homologia é aquele que dá melhores resultados
- Para se poder construir um modelo por homologia fiável é necessário que a sequência a modelar apresente uma percentage de identidade com uma proteína de estrutura conhecida de pelo menos 30-40% !

Fundamento da Modelação por homologia:

**A conservação da sequência está associada à conservação de estrutura!**

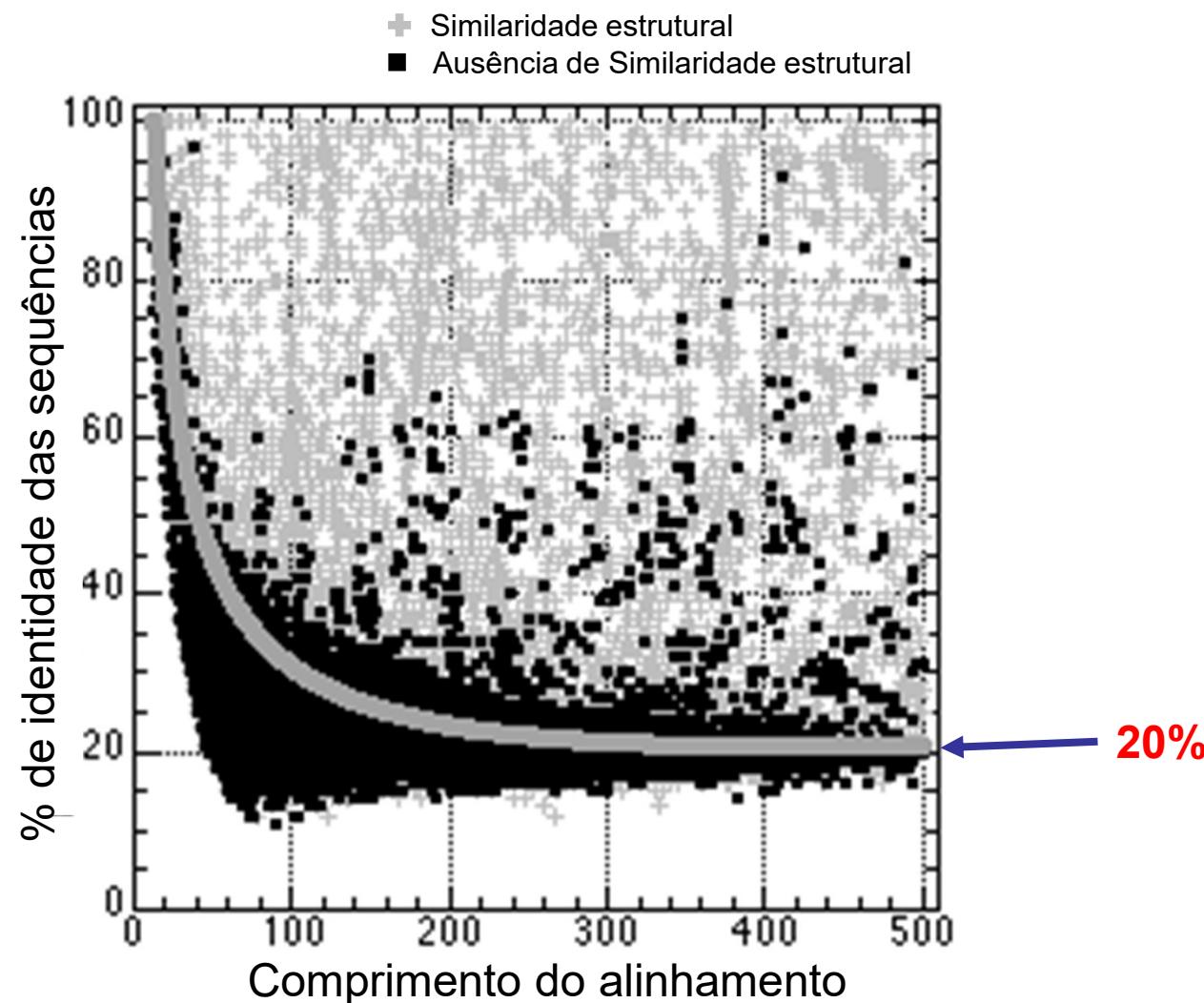
# A estrutura das proteínas é determinada pela sua sequência



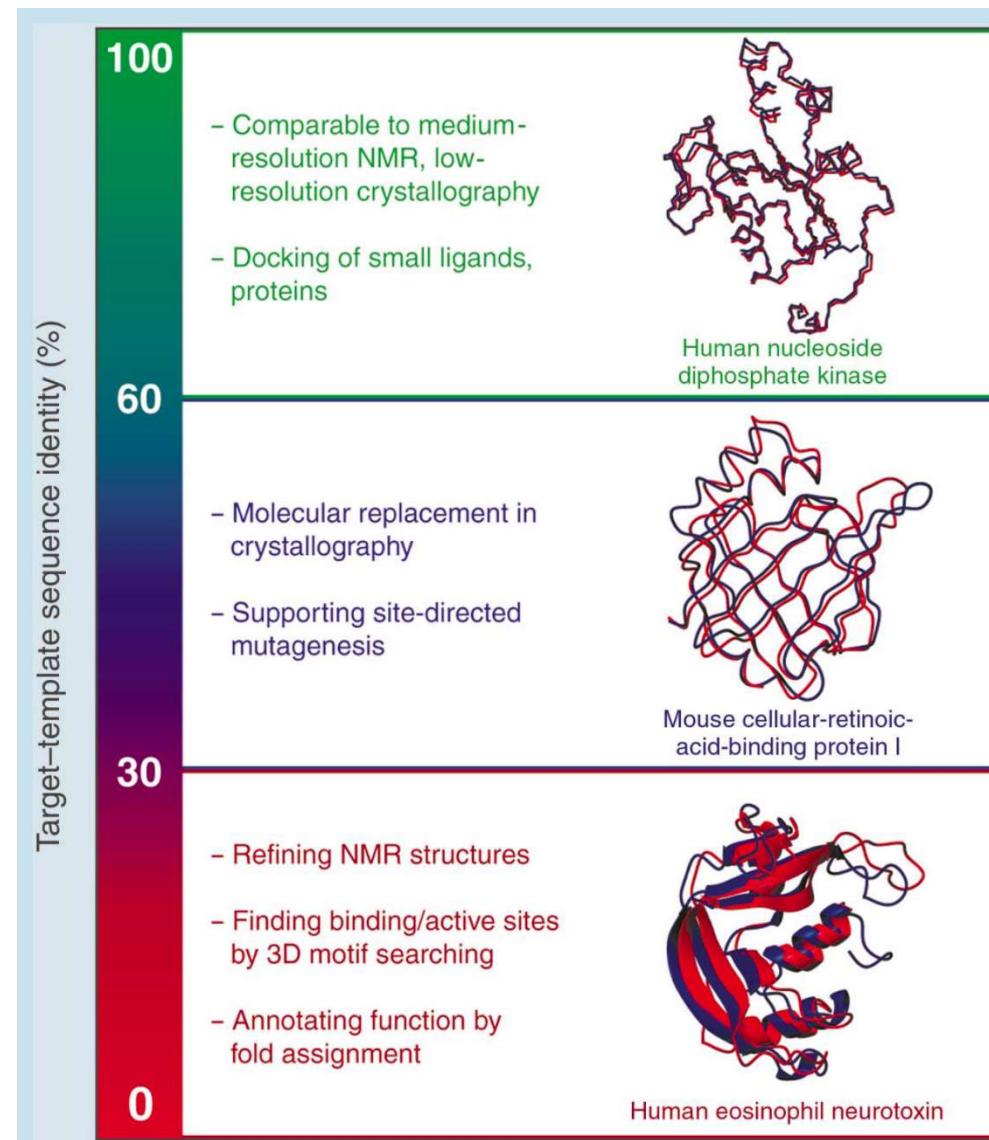
**Sequências similares implicam estruturas similares, logo:**

*A estrutura desconhecida de uma proteína pode ser prevista (construída), a partir da estrutura tridimensional de uma proteína de sequência suficientemente semelhante.*

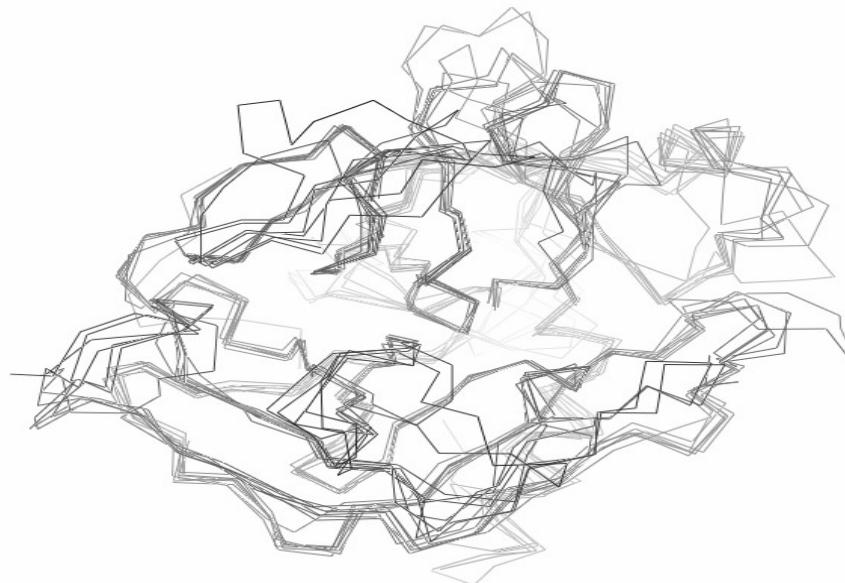
Qual a % de identidade mínima aceitável para existência de similaridade estrutural ?



# Impacto da similaridade na qualidade e utilidade do modelo



# Alinhamento estrutural das proteases de serina



1EOF:A {218:1-218)  
3ETH:E {223:1-221)  
1GMD:A {241:14-239)  
1HSD:H {260:1-248)  
1TOC:B {259:1-255)  
1HYL:A {230:1-227)  
1TON:\_ {235:11-232)  
2PKA:B {152:5-149)

```

IV GGG RAR PHA WP M PMS LQL - A - GNGHFGATL APN VPM SVA MSAHCY - ANVNYR - AVR VVL GANNLRS PRRP - TRL
IV CGG TCG ANGT PVY CVS LNS - D - YH HCCGGL LNS QVWV SVA MSAHCY - - K - G LVQ FBL EDN INVV BG - NED
IV EGV EAVG SVA WQV CQV LNS - G - EY HCCGGL LNS QVWV SVA MSAHCY - VTL - SDU VVAG EFC QSS SVA MSAHCY
IV EGGDQ EAVG SVA WQV CQV LNS - E - ELLGCGL LSD IRW LTAAHC RLjyppm dkmn - FTv dD LVL RIG KHS RTR YVR EK V
IV EGGDQ EAVG SVA WQV CQV LNS - P - ELLGCGL LSD IRW LTAAHC RLjyppm dkmn - FTv dD LVL RIG KHS RTR YVR EK V
I IGGY EAY TGL FPV QAG LDI - T - Tl qdq - RVWGGSL IRND KWI LTAAHC Vnd - AV - SVV YLG SAVQYB - - G
I IGGY EAY TGL FPV QAG LDI - P - YLGGCVL IDN SWV ITAAHC Y - SN - NYQ VLL RGGN LRFEDP - F
I IGGY EAY TGL FPV QAG LDI - S - SQPQWQVAVIN - - -

```

1EOF:A(218:1-218)  
3ETH:E(223:1-221)  
1GMD:A(241:14-239)  
1HSD:H(260:1-248)  
1TOC:B(259:1-255)  
1HYL:A(230:1-227)  
1TON:\_(235:11-232)  
2PKA:B(152:5-149)

```

FAVOP I FED -G-YD          -PV -NLNNL D V ILQ LNLQSAT IANVQVQVLAPE -QEP -GLP
ISASKS IVHps YN          -SN -TLNND M AASAS LNS RVASIS LPT -S -CAGTQCCLISWGWL -TIN
LRIKARW FNSYKX YN         -ST -TNDNQ QFNSYKX PFSQTV CVP LPSA -D -DAFGAT TIVTGWL -GTL
SLRDLR D LLLK LKRP EYLSD YIHPVCP LD -R -SET ASAD TIVTGWL -GTL
SMLDK I YI HpB YN         -WKR NELDRD L ALLKL KLRP FIELSD YIHPVCP LD -KqT -AKL LHAGP KFR VTVGWN -RN
VN SRS I IS HMFN P         -PD -TYLNOVLA ITI -PHV YTDNQI RI LPSS geel -N -NKF ENI WYVQW GQ -RN
RLVRQSVR FHDp YI P1v1tndtqP -D -D XSHMOLM LRLPSQ PAR ITAVK YLE LET -Q -BPELGS TCASGSWGS -II

```

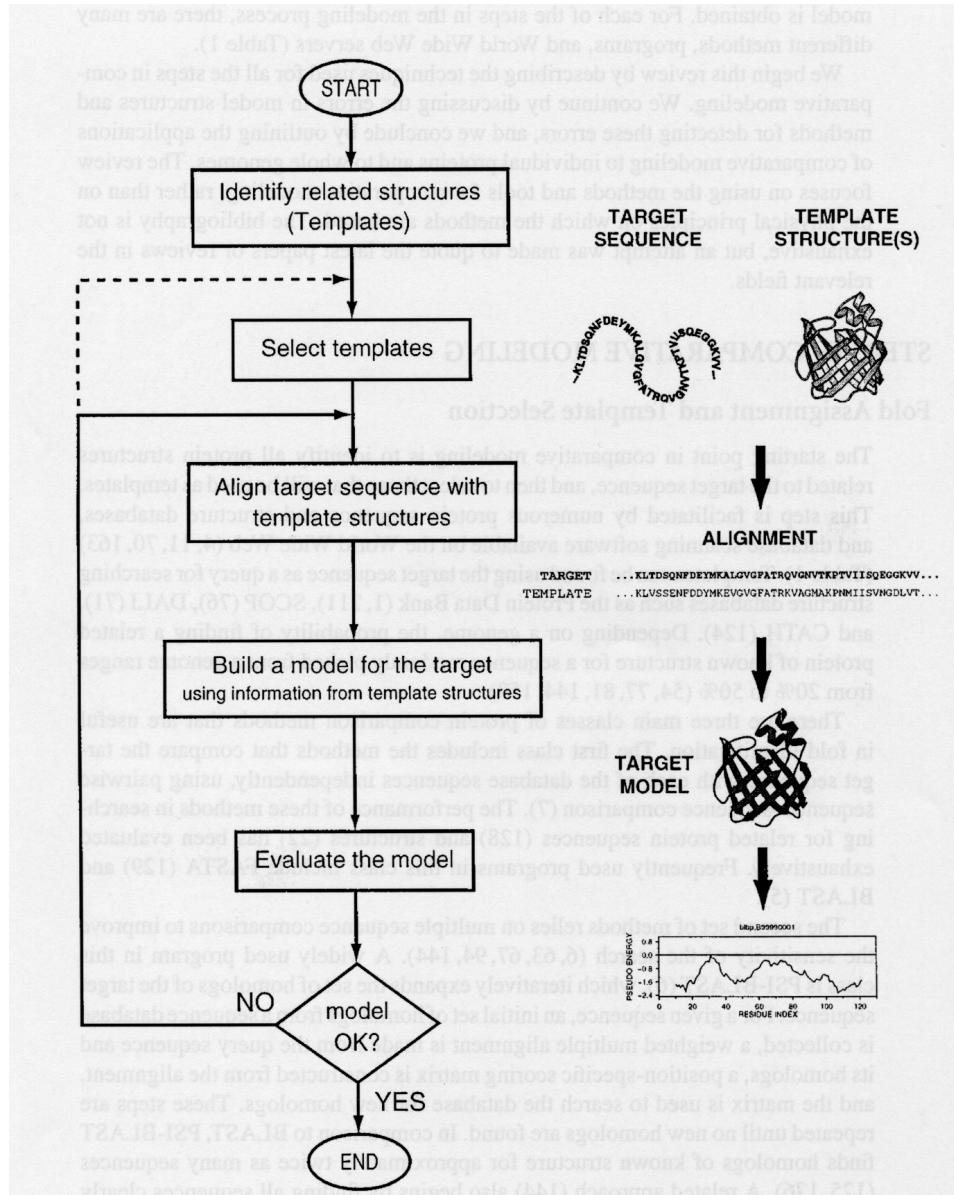
1EOF:A {218:1-218)  
3ETH:E {223:1-221)  
1GMD:A {241:14-239)  
1HSD:H {260:1-248)  
1TOC:B {259:1-255)  
1HYL:A {230:1-227)  
1TON:\_ {235:11-232)

1EOF:A {218:1-218}  
3BTM:E {223:1-221}  
1GMD:A {241:14-239}  
1HSD:H {260:1-248}  
1TOC:B {259:1-255}  
1HYL:A {230:1-227}  
1TON:\_ {235:11-232}

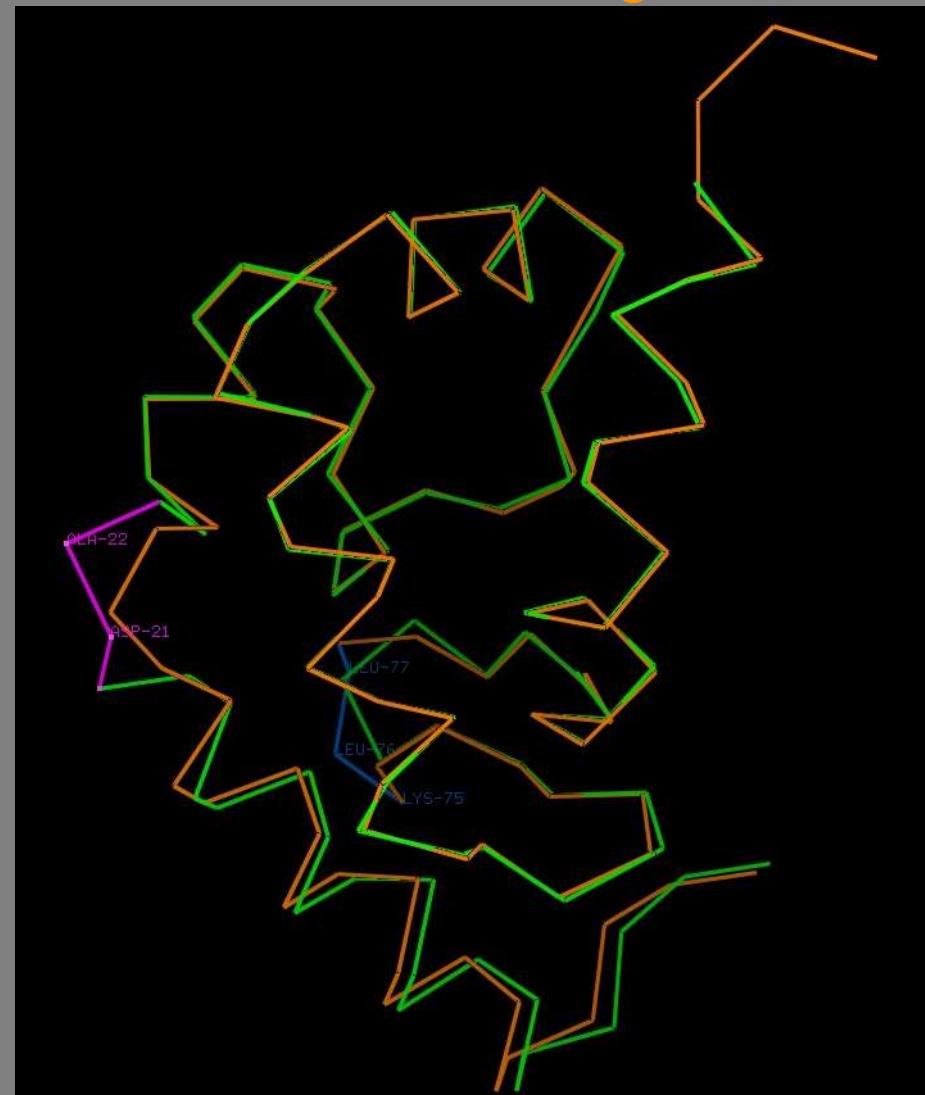
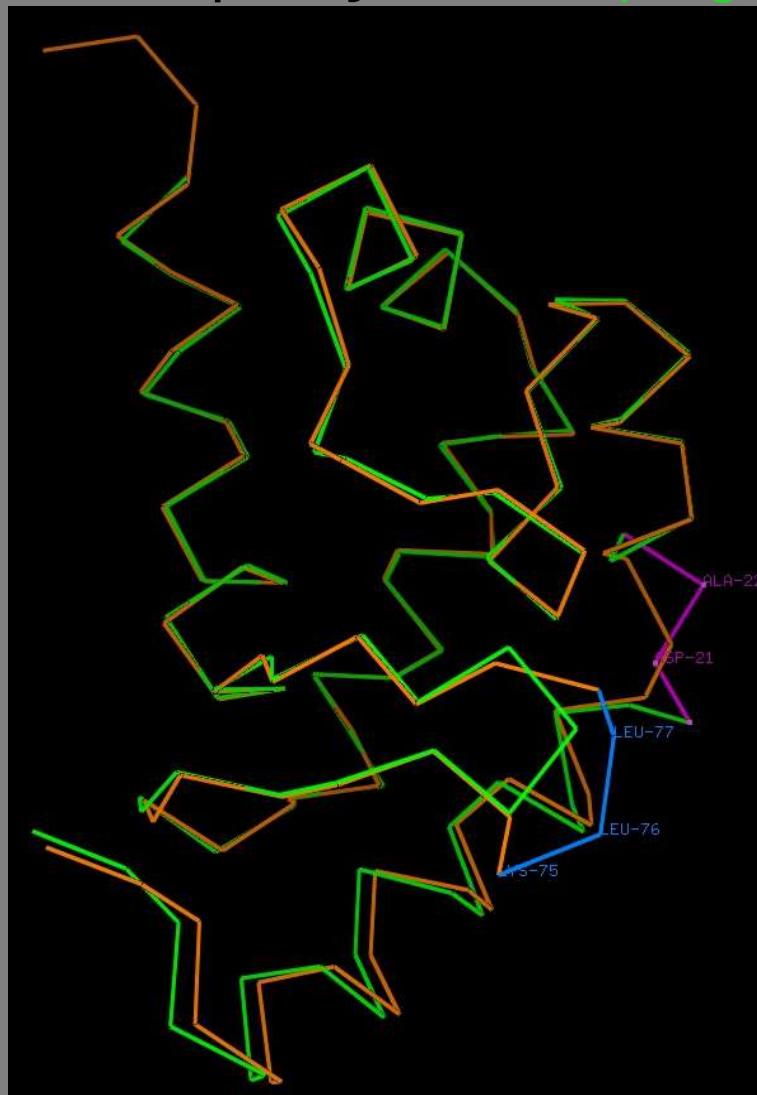
SFVR--GG-CASGLYPDGFAPVVAQEVNWIDSIIIG  
SWGS---G-CAQKNNKPGVYTKCVCNVSWSIKQTIA  
SWGS---S-TCTSTSPTGVYVARVTLANWVNSQQTLA  
SWGE---G-CDRDGKYGFTYHVFRLKWKIQVKIVD  
SWGE---G-CDRDGKYGFTYHVFRLKWKIQVKIVD  
SFVSGa-G-CESG-KPVGFSRTVSYMDWIQQNTG  
SGGA---TpCAKPKTPAYIARLKFITSWIKVVM

## Passos na modelação por homologia

- Alinhamento estrutural das proteínas de estrutura conhecida homólogas da proteína que se pretende modelar. I inspecção visual do alinhamento e eventuais correções.
- Alinhamento da sequência da proteína a modelar contra o *profile*, ou conjunto, das sequências alinhadas no passo anterior
- Construção do modelo tridimensional da proteína através das restrições impostas pela correspondência entre os resíduos alinhados com o conjunto das estruturas.
- Optimização das cadeias laterais da proteína por selecção de rotâmeros adequados para cada resíduo e localização.
- Optimização da estrutura dos “loops” existentes no modelo.
- Optimização global da estrutura por minimização e/ou dinâmica molecular
- Validação do modelo por critérios estereoquímicos e fenomenológicos
- Se necessário, corrigir os alinhamentos e voltar a produzir modelos até estes serem correctamente validados



# Comparação da criptogeína com o modelo da oligandrina

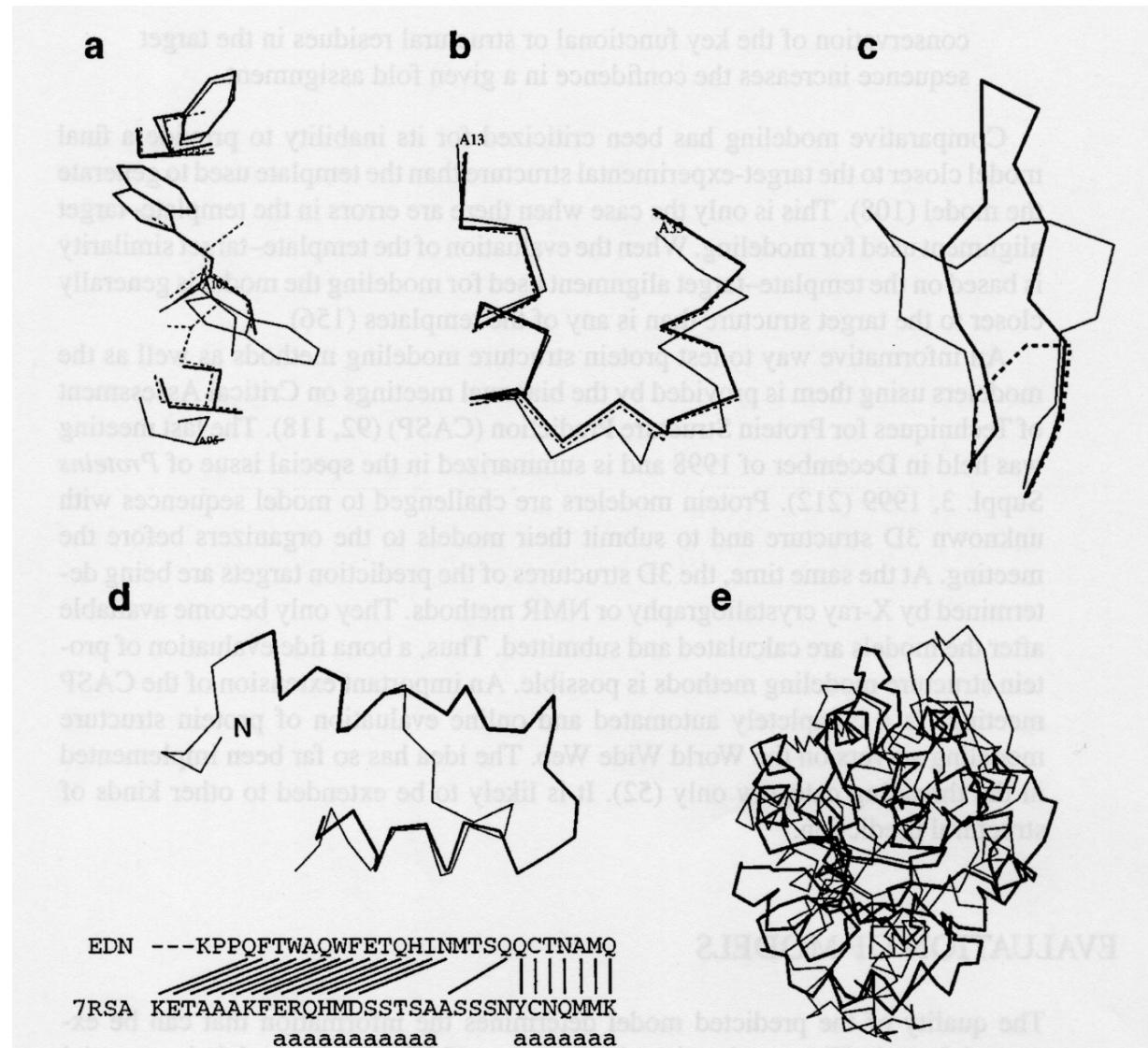


TA  
CTA  
S  
Q  
T  
A  
A  
Y  
K  
T  
L  
V  
S  
I  
L  
S  
**D**  
A  
S  
F  
N  
Q  
C  
S  
T  
D  
S  
G  
Y  
S  
M  
L  
T  
A  
K  
A  
L  
P  
T  
A  
Q  
Y  
K  
L  
M  
C  
A  
S  
T  
A  
C  
N  
T  
M  
I  
K  
K  
I  
V  
T  
L  
N  
P  
P  
N  
C  
D  
L  
T  
V  
P  
T  
—  
—  
S  
G  
L  
V  
L  
N  
V  
Y  
S  
Y  
A  
N  
G  
F  
S  
N  
K  
C  
S  
S  
L  
—  
—  
A  
T  
C  
T  
D  
E  
Q  
F  
S  
D  
S  
I  
I  
K  
L  
T  
P  
A  
I  
G  
—  
—  
S  
V  
G  
C  
T  
A  
D  
S  
G  
F  
S  
M  
I  
P  
P  
T  
G  
L  
P  
T  
N  
D  
Q  
Y  
K  
K  
M  
C  
K  
S  
T  
N  
C  
K  
T  
L  
I  
K  
E  
I  
K  
D  
A  
N  
V  
A  
D  
C  
E  
L  
F  
S  
**K**  
**L**  
**P**  
**G**  
**S**  
**V**  
**P**  
**L**  
**N**  
**V**  
**V**  
**L**  
**A**  
**N**  
**F**  
**D**  
**F**  
**V**  
**C**  
**A**  
**V**  
**G**  
**S**

## Erros na modelação por homologia (1)

- Empacotamento das cadeias laterais incorrecto. Quando a divergência de sequências se torna elevada verificam-se diferenças no empacotamento do “core” da proteína. Erros graves se ocorrerem em zonas ligadas à função (centros activos, etc..)
- Distorções e deslocações em zonas correctamente alinhadas. Podem ser devidas à divergência das sequências ou a artefactos na determinação da estrutura, como o empacotamento das moléculas no cristal.
- Erros em regiões para as quais não há correspondência nas moléculas de estrutura conhecida - “loops”. São as regiões mais difíceis de modelar. Para sequências pequenas (<9 aa.), certos métodos podem determinar correctamente a conformação do “backbone” da proteína.
- Erros devidos a um alinhamento incorrecto das sequências. São a principal fonte de erros na modelação por homologia, quando a percentagem de identidade é < 30 %. Usar um número grande de sequências para melhorar o alinhamento.
- Escolha incorrecta da estrutura ou estruturas a usar como base para a construção do modelo. Este problema ocorre para identidades muito baixas, < 25%

## Erros na modelação por homologia (2)



# Servidores web para modelação por homologia

- **SWISS-MODEL** - <http://swissmodel.expasy.org>
- **Phyre2** - <http://www.sbg.bio.ic.ac.uk/phyre2/>
- **I-Tasser** - <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>
- **Raptor-X** - <http://raptorgx.uchicago.edu/>
- **Hhpred** - <http://toolkit.lmb.uni-muenchen.de/hhpred>
- **Robetta** - <https://robbetta.bakerlab.org>
- **ModWeb** - <https://modbase.compbio.ucsf.edu/modweb/>

# Bases de modelos pré-calculados

- Repositórios que contêm modelos calculados de forma sistemática para uma larga fracção das sequências conhecidas
  - Forma simples e rápida de obter um modelo para uma proteína de estrutura desconhecida
  - Geralmente “seguros” para similaridades de sequência > 70-75%
  - Podem ser refinados ou gerados para diferentes “templates”
  - Importante considerer os indicadores de qualidade dos modelos
- 
- SWISS Model Repository – <https://swissmodel.expasy.org/repository/>
  - ModBase - <https://modbase.compbio.ucsf.edu/i>

# SWISS MODEL repository

The screenshot shows a web browser window for the SWISS-MODEL Repository at [swissmodel.expasy.org/repository/](https://swissmodel.expasy.org/repository/). The page features a header with the BIOZENTRUM logo, UniProtKB AC or Entry Name search fields, and links for Modelling, Repository (selected), Tools, Documentation, Log in, and Create Account. Below the header, it displays statistics: 1,683,091 models from SWISS-MODEL for UniProtKB targets and 149,863 structures from PDB with mapping to UniProtKB. It also lists specific UniProtKB entries: F1P6T8, Q83XK2\_ECOLX, B4IFM4, W9KYS2\_FUSOX; AOA0E0UR70, ULA1\_HUMAN, P04439, PSA2\_YEAST. A note indicates that the repository provides models for reference proteomes of various organisms.

The SWISS-MODEL Repository is a database of annotated 3D protein structure models generated by the SWISS-MODEL homology-modelling pipeline.

Bienert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L, Schwede T (2017). The SWISS-MODEL Repository - new features and functionality *Nucleic Acids Res.* 45(D1):D313-D319. [\[doi\]](#)

The aim of the SWISS-MODEL Repository is to provide access to an up-to-date collection of annotated 3D protein models generated by automated homology modelling for relevant model organisms and experimental structure information for all sequences in UniProtKB. Regular updates ensure that target coverage is complete, that models are built using the most recent sequence and template structure databases, and that improvements in the underlying modelling pipeline are fully utilised. It also allows users to assess the quality of the models using the latest QMEAN results. If a sequence has not been modelled, the user can build models interactively via the SWISS-MODEL workspace.

Currently the repository contains 1,683,091 models from SWISS-MODEL for UniProtKB targets as well as 149,863 structures from PDB with mapping to UniProtKB.

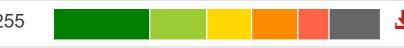
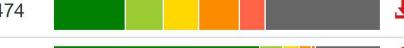
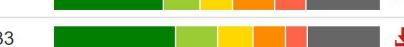
We currently provide models for the **reference proteomes** of the following model organisms, based on UniProtKB release 2019\_10. If you want to download a large number of models, please contact us.

<https://swissmodel.expasy.org/repository/>

# SWISS MODEL repository

SWISS-MODEL Repository

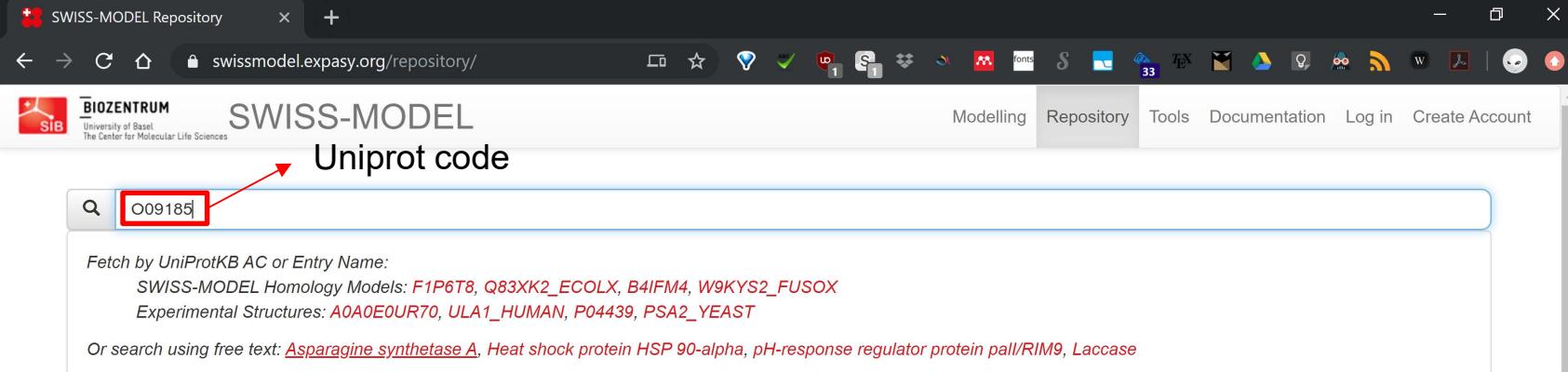
swissmodel.expasy.org/repository/

	Proteome Size	Sequences Modelled	Models	Seq Coverage	Download Metadata (Models and structures)	Download Coordinates (Homology models)
<i>Homo sapiens</i>	20,659	17,505	43,255		<a href="#">13.7 MB</a>	<a href="#">4.5 GB</a>
<i>Mus musculus</i>	21,960	18,708	43,088		<a href="#">8.0 MB</a>	<a href="#">3.0 GB</a>
<i>Caenorhabditis elegans</i>	19,944	12,942	23,474		<a href="#">3.7 MB</a>	<a href="#">1.3 GB</a>
<i>Escherichia coli</i>	4,391	3,525	6,210		<a href="#">1.6 MB</a>	<a href="#">465.1 MB</a>
<i>Arabidopsis thaliana</i>	27,466	20,467	37,517		<a href="#">5.6 MB</a>	<a href="#">2.1 GB</a>
<i>Drosophila melanogaster</i>	13,793	10,035	20,135		<a href="#">3.2 MB</a>	<a href="#">1.3 GB</a>
<i>Saccharomyces cerevisiae</i>	6,049	4,685	8,241		<a href="#">1.9 MB</a>	<a href="#">489.8 MB</a>
<i>Schizosaccharomyces pombe</i>	5,141	4,006	7,433		<a href="#">1.1 MB</a>	<a href="#">424.7 MB</a>
<i>Caulobacter vibrioides</i>	3,720	2,975	5,178		<a href="#">736.2 KB</a>	<a href="#">366.1 MB</a>
<i>Mycobacterium tuberculosis</i>	3,993	3,267	5,096		<a href="#">887.4 KB</a>	<a href="#">340.7 MB</a>
<i>Pseudomonas aeruginosa</i>	5,563	4,697	8,833		<a href="#">1.3 MB</a>	<a href="#">706.7 MB</a>
<i>Staphylococcus aureus</i>	2,889	2,124	3,615		<a href="#">542.7 KB</a>	<a href="#">244.0 MB</a>
<i>Plasmodium falciparum</i>	5,448	3,716	6,636		<a href="#">995.9 KB</a>	<a href="#">307.5 MB</a>

Latest snapshot of SMR was taken 1 month ago.

<https://swissmodel.expasy.org/repository/>

# SWISS MODEL repository



The screenshot shows a web browser window for the SWISS-MODEL Repository. The URL in the address bar is [swissmodel.expasy.org/repository/](https://swissmodel.expasy.org/repository/). The page header includes the BIOZENTRUM logo, the SWISS-MODEL logo, and navigation links for Modelling, Repository, Tools, Documentation, Log in, and Create Account. Below the header is a search bar with a magnifying glass icon. The search term 'O09185' is entered into the search bar and highlighted with a red box. An arrow points from the text 'Uniprot code' to this red box. Below the search bar, there is a section titled 'Fetch by UniProtKB AC or Entry Name:' which lists homology models and experimental structures. It also provides a free text search option.

The SWISS-MODEL Repository is a database of annotated 3D protein structure models generated by the SWISS-MODEL homology-modelling pipeline.

Biernert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L, Schwede T (2017). The SWISS-MODEL Repository - new features and functionality *Nucleic Acids Res.* 45(D1):D313-D319. [\[M\]](#) [\[do>\]](#)

The aim of the SWISS-MODEL Repository is to provide access to an up-to-date collection of annotated 3D protein models generated by automated homology modelling for relevant model organisms and experimental structure information for all sequences in UniProtKB. Regular updates ensure that target coverage is complete, that models are built using the most recent sequence and template structure databases, and that improvements in the underlying modelling pipeline are fully utilised. It also allows users to assess the quality of the models using the latest QMEAN results. If a sequence has not been modelled, the user can build models interactively via the SWISS-MODEL workspace.

Currently the repository contains 1,683,091 models from SWISS-MODEL for UniProtKB targets as well as 149,863 structures from PDB with mapping to UniProtKB.

We currently provide models for the [reference proteomes](#) of the following model organisms, based on UniProtKB release 2019\_10. If you want to download a large number of models, please contact us.

<https://swissmodel.expasy.org/repository/>

# SWISS MODEL repository

swissmodel.expasy.org/repository/uniprot/O09185

BIOZENTRUM University of Basel The Center for Molecular Life Sciences SWISS-MODEL

O09185 (P53\_CRIGR) *Cricetulus griseus* (Chinese hamster) (*Cricetulus barabensis* griseus)  
Cellular tumor antigen p53 ★ UniProtKB<sup>®</sup> InterPro<sup>®</sup> STRING<sup>®</sup>

393 aa; Sequence (Fasta)

009185  
homo-4-mer; 94-356L  
homo-4-mer; 94-356L  
monomer; 1-37  
monomer; 2-56  
monomer; 1-54

50 100 150 200 250 300 350

4mzr.1.B Cellular tumor antigen p53

Seq Identity 82.63%  
Seq Similarity 0.56  
4 x ZINC ION  
SMTL Version 2019-12-06  
Download Model

Model Quality Estimate

QMEAN	-2.18
C $\beta$	0.09
All Atom	-0.70
solvation	-0.84
torsion	-1.86

Sequence Features

Metal binding Site Natural variant  
DNA binding InterPro

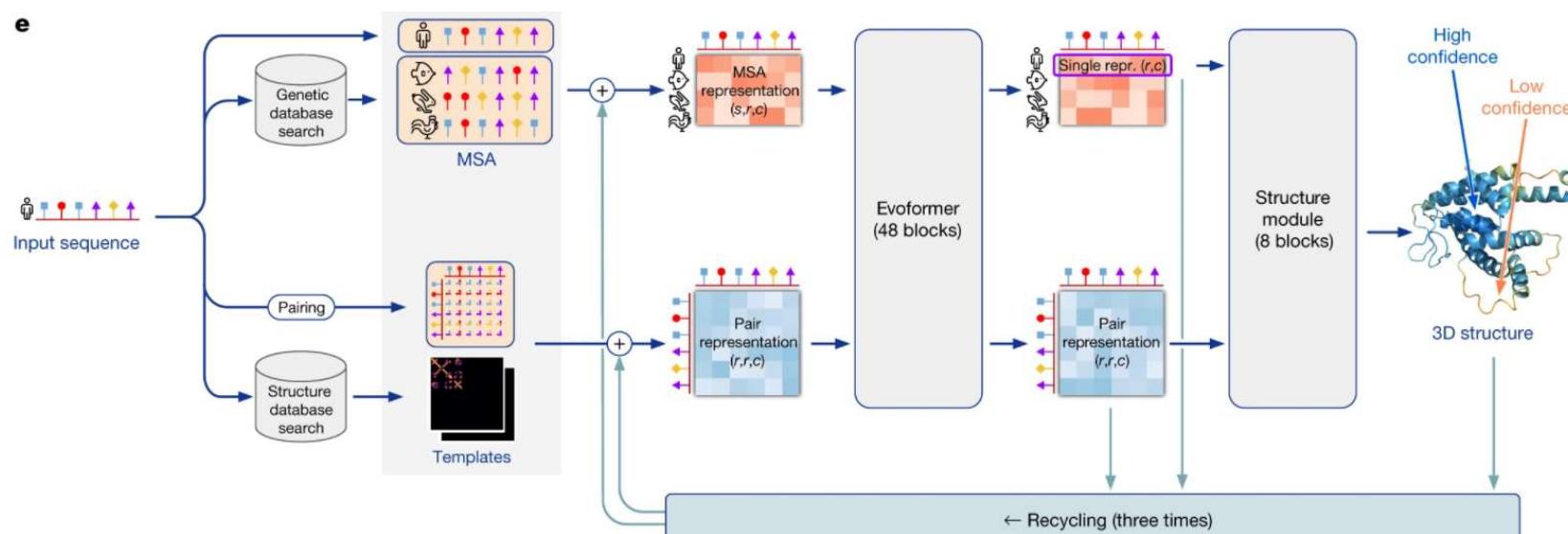
Colours NGL Cartoon

<https://swissmodel.expasy.org/repository/>

# Alpha Fold 2



- Em 2020, a Google Deep Mind apresentou o algoritmo Alpha Fold 2 para a previsão da estrutura terciária das proteínas, ultrapassando em larga margem todas as outras ferramentas actualmente disponíveis para este tipo de previsões.
- O Alpha Fold 2 é um *deep neural network*, um algoritmo de aprendizagem máquina que é treinado usando como exemplos as estruturas e sequências de proteínas conhecidas.



# Alpha Fold 2 no Uniprot

- A Google Deep Mind, em colaboração com o European Bioinformatics Institute (EBI), produziu previsões para a estrutura terciária de todas as proteínas do proteoma humano. Estas previsões encontram-se disponíveis nas respectivas entradas do banco de dados Uniprot.

The image consists of two side-by-side screenshots of the UniProt website. Both screenshots show the entry for PITHD1 (Q9GZP4) from the UniProtKB database.

**Left Screenshot (Display View):**

- Entry:** PITH domain-containing protein 1
- Gene:** PITHD1
- Organism:** Homo sapiens (Human)
- Status:** Reviewed - Annotation score: 5/5 - Experimental evidence at protein level!
- Function:** Promotes megakaryocyte differentiation by up-regulating RUNX1 expression (PubMed:25134913). Regulates RUNX1 expression by activating the proximal promoter of the RUNX1 gene and by enhancing the translation activity of an internal ribosome entry site (IRES) element in the RUNX1 gene (PubMed:25134913).
- GO - Biological process:**
  - penetration of cumulus oophorus (Source: Ensembl)
  - penetration of zona pellucida (Source: Ensembl)
  - positive regulation of megakaryocyte differentiation (Source: UniProtKB)
  - positive regulation of transcription, DNA-templated (Source: UniProtKB)
  - regulation of proteasomal protein catabolic process (Source: Ensembl)
  - spermatid development (Source: Ensembl)
- Structure:** (This section is circled in red.)
- Keywords:** Complete GO annotation on QuickGO ...
- Molecular Activator:**

**Right Screenshot (Structure View):**

- Model Confidence:** A legend indicates confidence levels: Very high (pLDDT > 90), Confident (90 > pLDDT > 70), Low (70 > pLDDT > 50), and Very low (pLDDT < 50). It also notes that AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.
- Structure:** A 3D ribbon model of the protein structure, colored by confidence.
- Table:** A table showing the 3D structure databases.

SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
AlphaFold	AF-Q9GZP4-F1	Predicted			1-211	AlphaFold
- 3D structure databases:** SMR: Q9GZP4, ModBase: Search...

# Moléculas pequeñas

# Bases de dados de pequenas moléculas

- Bases de dados que contêm estruturas de milhares ou milhões de pequenas moléculas, na sua maioria compostos orgânicos, sintéticos ou de origem natural
- Ferramentas essenciais para indústria farmacêutica, utilizadas na descoberta de novos fármacos, c
- Podem conter uma variedade de *descritores moleculares* (estrutura, solubilidade, massa molecular, hidrofobicidade, carga, etc...) e também informação sobre a actividade biológica e até dados de ensaios de actividade

# Bases de dados de pequenas moléculas

- PubChem (<http://pubchem.ncbi.nlm.nih.gov/>)
- DrugBank (<http://www.drugbank.ca>)
- ChEMBL ([http://https://www.ebi.ac.uk/chembl/](https://www.ebi.ac.uk/chembl/))
- ZINC (<http://zinc.docking.org>)
- Cambridge Structural Database ([http://http://www.ccdc.cam.ac.uk](http://www.ccdc.cam.ac.uk))
- Traditional Chinese Medicine (<http://tcm.cmu.edu.tw>)

# PubChem



- Conjunto de bases de dados mantido pelo National Institute for Biotechnology Information (NCBI), parte da rede dos National Institutes of Health (NIH), nos EUA.
- Três bases de dados centrais contendo substâncias, compostos químicos e ensaios de actividade para diferentes sistemas biológicos
- Contem moléculas com menos de 1000 átomos e menos de 1000 ligações químicas
- 3 bases de dados
  - Compound (**62,041,347**)
  - Substance (**178431037**)
  - Bioassay (**1112105**)
- Permite pesquisa por estrutura, similaridade, etc...

9/11/2014

# Bases de dados



- **PubChem Substance:** cada entrada nesta base de dados contem informação sobre uma *amostra química* de proveniência bem definida, que pode conter ou ou mais compostos. Cada entrada possui referências cruzadas para bibliografia, ensaios biológicos, estruturas de compostos, proteínas, etc...
- **PubChem Compound:** base de estruturas químicas validadas e agrupadas por similaridade. Contem vários descritores e propriedades moleculares pré-calculados (eg: XlogP, MW) que podem ser usados para filtrar as pesquisas. Cada **substância** pode conter um ou mais compostos.
- **PubChem Bioassay:** ensaios de actividade biológicas relativos às entradas de **PubChem Substance**, contendo as descrições e resultados dos ensaios.

# Pesquisa no PubChem



- **Compound:** nomes, sinônimos ou keywords.
- **Substance:** nomes, sinônimos, keywords
- **Bioassay:** pesquisa de termos nas descrição do ensaio
- **Entrez:** pesquisar usando as ferramentas do NCBI
- **Estrutura:** pesquisar por similaridade de estrutura
- **Ferramentas de análise:** SAR maps, tabelas customizáveis, etc...

The PubChem Project

pubchem.ncbi.nlm.nih.gov

Databases ▾ Upload Services ▾ Help more ▾

databases

# PubChem

BioAssay Compound Substance

GO Advanced Search

Structure Search | BioActivity Analysis | BioActivity DataDicer

search tools

New The PubChem Social Media campaign is now launched! [see more...](#)

more ...

Facebook Twitter Google+ RSS Chat

BioActivity Summary

BioActivity Databatable

BioActivity SAR

BioActivity DataDicer

Structure Search

3D Conformer Tools

Structure Clustering

3D Conformer application tool

Classification

Upload

Download

PubChem FTP

Write to Helpdesk | Disclaimer | Privacy Statement | Accessibility | Data Citation Guidelines  
National Center for Biotechnology Information  
NLM | NIH | HHS

This screenshot shows the main interface of the PubChem website. At the top, there's a navigation bar with links for 'Databases', 'Upload', 'Services', 'Help', and 'more'. Below the header is the 'PubChem' logo. Underneath the logo is a search bar with three buttons: 'BioAssay', 'Compound', and 'Substance'. To the right of the search bar are two buttons: 'GO' and 'Advanced Search'. Below the search area, there are links for 'Structure Search', 'BioActivity Analysis', and 'BioActivity DataDicer'. A callout bubble at the bottom left says 'New The PubChem Social Media campaign is now launched!' with a link to 'see more...'. On the right side, there's a sidebar with social media icons (Facebook, Twitter, Google+, RSS, Chat) and a list of tools: BioActivity Summary, BioActivity Databatable, BioActivity SAR, BioActivity DataDicer, Structure Search, 3D Conformer Tools, Structure Clustering, Classification, Upload, Download, and PubChem FTP. At the bottom of the page, there are links for writing to the helpdesk, disclaimer, privacy statement, accessibility, data citation guidelines, and information about the National Center for Biotechnology Information, NLM, NIH, and HHS.

# PubChem Compound

aspirin - PubChem C x https://www.ncbi.nlm.nih.gov/pccompound/?term=aspirin

NCBI Resources How To

PubChem Compound aspirin Search PubChem Compound. Use up and down arrows to choose an item from the autocomplete.

Save search Limits Advanced Help

Display Settings: Summary, 20 per page. Sorted by Default order

Results: 1 to 20 of 88

Page 1 of 5

Send to: Filters: Manage Filters

Actions on your results

- BioActivity Analysis Analyze the BioActivities of the compounds
- Structure Clustering Cluster structures based on structural similarity
- Structure Download Download the structures in various formats
- Pathways Analyze pathways containing the compounds

Refine your results • What's this?

Chemical Properties Rule of 5 (22)

BioActivity Experiments BioAssays, Active (13) BioAssays, Tested (19)

Protein 3D Structures (3) Human Transthyretin (ttr) Complexed With Diflunisal (1)

Biomedical Annotation Pharmacological Actions (25) Anti-Inflammatory Agents, Non-Steroidal (21)

BioSystems (3)

Depositor Category Biological Properties (75) Chemical Vendors (62) Journal Publishers (32)

1. aspirin: ACETYLSALICYLIC ACID; 2-Acetoxybenzoic acid ...  
MW: 180.157420 g/mol MF: C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>  
IUPAC name: 2-acetoxybenzoic acid  
CID: 2244  
[Summary](#) [Similar Compounds](#) [Same Parent Connectivity](#) [Mixture/Component Compounds](#) [PubMed \(MeSH Keyword\)](#) Active in 125 of 3501 BioAssays

2. Calascorbin: Calcium aspirin; Calascorbate ...  
MW: 398.376960 g/mol MF: C<sub>18</sub>H<sub>14</sub>CaO<sub>8</sub>  
IUPAC name: calcium;2-acetoxybenzoate  
CID: 6247  
[Summary](#) [Similar Compounds](#) [Same Parent Connectivity](#) [Mixture/Component Compounds](#) [PubMed \(MeSH Keyword\)](#)

3. Axotal: BUTALBITAL ASPIRIN AND CAFFEINE; BUTAL COMPOUND ...  
MW: 598.604360 g/mol MF: C<sub>28</sub>H<sub>34</sub>N<sub>6</sub>O<sub>9</sub>  
IUPAC name: 2-acetoxybenzoic acid;5-(2-methylpropyl)-5-prop-2-enyl-1,3...  
CID: 24847961  
[Summary](#) [Similar Compounds](#) [Mixture/Component Compounds](#) [PubMed \(MeSH Keyword\)](#)

4. CODEINE, ASPIRIN, APAP FORMULA NO. 2; CODEINE, ASPIRIN, APAP FORMULA NO. 3; CODEINE, ASPIRIN, APAP FORMULA NO. 4 ...  
MW: 728.679402 g/mol MF: C<sub>35</sub>H<sub>41</sub>N<sub>2</sub>O<sub>13</sub>P  
IUPAC name: (4R,4aR,7S,7aR,12bS)-9-methoxy-3-methyl-2,4,4a,7,7a,13-hexah...  
CID: 24847798  
[Summary](#) [Similar Compounds](#) [Mixture/Component Compounds](#)

5. Aspirin sodium; Sodium aspirin; Sodium acetylsalicylate ...  
MW: 202.139249 g/mol MF: C<sub>9</sub>H<sub>7</sub>NaO<sub>4</sub>  
IUPAC name: sodium;2-acetoxybenzoate  
CID: 23666729  
[Summary](#) [Similar Compounds](#) [Same Parent Connectivity](#) [Mixture/Component Compounds](#)

# PubChem Compound

Aspirin - PubChem    pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=2244#x27

NCBI

**PubChem Compound**  **Search** **Help**

**SHARE** [f](#) [t](#) [e](#) ...

**Aspirin - Compound Summary (CID 2244)**

Also known as: ACETYLSALICYLIC ACID, 2-Acetoxybenzoic acid, Acylpyrin, Ecotrin, Acenterine, Polopiryna, Acetosal, Colfarit, Enterosarein

Molecular Formula: C<sub>9</sub>H<sub>8</sub>O<sub>4</sub> Molecular Weight: 180.15742 InChIKey: BSYNRYMUTXBXSQ-UHFFFAOYSA-N

The prototypical analgesic used in the treatment of mild to moderate pain. It has anti-inflammatory and antipyretic properties and acts as an inhibitor of cyclooxygenase which results in the inhibition of the biosynthesis of prostaglandins. Aspirin also inhibits platelet aggregation and is used in the prevention of arterial and venous thrombosis. (From Martindale, The Extra Pharmacopoeia, 30th ed, p5) From: MeSH

**Table of Contents** [Show subcontent titles](#)

- Identification
- Related Records
- Use and Manufacturing
- Pharmacology
- Biomedical Effects and Toxicity
- Safety and Handling
- Environmental Fate and Exposure Potential
- Exposure Standards and Regulations
- Monitoring and Analysis Methods
- Literature
- Patents
- Biomolecular Interactions and Pathways
- Biological Test Results
- Classification
- Chemical and Physical Properties

**2D Structure** **3D Conformer**

**Properties**

Compound ID: 2244  
Molecular Weight: 180.15742 [g/mol]  
Molecular Formula: C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>  
XLogP3: 1.2  
H-Bond Donor: 1  
H-Bond Acceptor: 4

**BioActivity Data Links**

This Compound  
with Similar Compounds  
with Similar Conformers

**Related Compounds**

Same, Connectivity (8)  
Similar Compounds (3154)  
Similar Conformers (8000) [View](#)

# PubChem Substance

aspirin - PubChem S x https://www.ncbi.nlm.nih.gov/pcsubstance/?term=aspirin

NCBI Resources How To

PubChem Substance aspirin Save search Limits Advanced Search Help

Display Settings: Summary, 20 per page, Sorted by Default order

Results: 1 to 20 of 547

<< First < Prev Page 1 of 28 Next > Last >>

**Send to:** Filters: Manage Filters

**Actions on your results**

- BioActivity Analysis Analyze the BioActivities of the substances
- Structure Clustering Cluster structures based on structural similarity
- Structure Download Download the structures in various formats
- Pathways Analyze pathways containing the compounds

**Refine your results** • What's this?

**Chemical Properties**  
Rule of 5 (289)

**BioActivity Experiments**  
BioAssays, Active (13)   
BioAssays, Tested (42)

**Protein 3D Structures** (38)  
 Structural Basis Of The Prevention Of Nsaid-induced Damage Of The Gastrointestinal Tract By C-terminal Half (c-lobe) Of Bovine Colostrum Protein Lactoferrin: Binding And Structural Studies Of The C-lobe Complex With Aspirin (10)

**BioMedical Annotation**  
Pharmacological Actions (361)  
 Anti-Inflammatory Agents, Non-Steroidal (327)

**BioSystems** (1)

**Depositor Category**  
Biological Properties (156)

1. aspirin: ACETYLSALICYLIC ACID; Ecotrin ...  
Source: LeadScope (LS-143)  
SID: 49854366 [CID: 2244]  
[Summary](#) [PubChem Same Compound](#) [Same Parent, Connectivity](#) [PubMed \(MeSH Keyword\)](#)

2. aspirin: ACETYLSALICYLIC ACID; Ecotrin ...  
Source: Comparative Toxicogenomics Database (D001241)  
SID: 53788943 [CID: 2244]  
[Summary](#) [PubChem Same Compound](#) [Same Parent, Connectivity](#) [PubMed \(MeSH Keyword\)](#)

3. aspirin: ACETYLSALICYLIC ACID; Ecotrin ...  
Source: Therapeutic Targets Database (DAP000843)  
SID: 134338122 [CID: 2244]  
[Summary](#) [PubChem Same Compound](#) [Same Parent, Connectivity](#) [PubMed \(MeSH Keyword\)](#)

4. aspirin: ACETYLSALICYLIC ACID; Ecotrin ...  
Source: Human Metabolome Database (HMDB01879)  
SID: 126524194 [CID: 2244]  
[Summary](#) [PubChem Same Compound](#) [Same Parent, Connectivity](#) [PubMed \(MeSH Keyword\)](#)

5. aspirin: ACETYLSALICYLIC ACID; Ecotrin ...  
Source: ChemIDplus (0000050782)  
SID: 134971785 [CID: 2244]  
[Summary](#) [PubChem Same Compound](#) [Same Parent, Connectivity](#) [PubMed \(MeSH Keyword\)](#)

aspirin: ACETYLSALICYLIC ACID; Ecotrin ...

# PubChem Substance

aspirin - PubChem https://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?sid=49854366&loc=es\_rss

NCBI

PubChem Substance Search Help

SHARE

Chemical Structure (CID 2244) Deposited Record (SID 49854366)

Substance Summary for: SID 49854366

aspirin

Also known as: ACETYLSALICYLIC ACID; Ecotrin; Acenterine; Polopiryna; Acylpyrin; Easprin; Acetylsalicylate; 2-Acetoxybenzoic acid

Table of Contents Show subcontent titles

- Identification
- Related Records
- Use and Manufacturing
- Pharmacology
- Biomedical Effects and Toxicity
- Safety and Handling
- Environmental Fate and Exposure Potential
- Exposure Standards and Regulations
- Monitoring and Analysis Methods
- Literature
- Classification
- Chemical and Physical Properties

Expand all sub-sections

Chemical Structure:

ASN.1 XML SDF

Follow us on

Related Substances

Same (206)  
Same, Connectivity (222)

Other Links

Chemical Structure Search

# PubChem BioAssay

Screenshot of the PubChem BioAssay interface for BioAssay AID 444512.

The page title is "BioAssay: AID 444512".

The main title of the assay is "Antiplatelets aggregatory activity in human platelets rich plasma assessed as inhibition of collagen-induced platelets aggregation by aggregometry".

A brief description states: "Aspirin prodrugs and related nitric oxide releasing compounds hold significant therapeutic promise, but they are hard to design because aspirin esterification renders its acetate group very susceptible to plasma esterase mediated hydrolysis. Isosorbide-2-aspirinate-5-salicylate is a true aspirin prodrug in human blood because it can be effectively hydrolyzed to aspirin upon interaction with more ...".

**Table of Contents:**

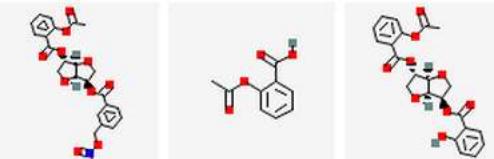
- BioActive Compounds
- Description
- Comment
- Categorized Comment
- Result Definitions
- Data Table (Concise)

**Details:**

- AID: 444512
- Data Source: ChEMBL (595690)
- Depositor Category: Literature, Extracted
- BioAssay Version: 5.1
- Deposit Date: 2010-07-08
- Modify Date: 2013-07-13

**Data Table (Complete):** Active (3) All (5)

**BioActive Compounds: 3**



**BioActivity Summary:** [View](#) [?](#)

**Structure-Activity Analysis:** [View](#) [?](#)

**Structure Clustering:** [View](#) [?](#)

**Links:**

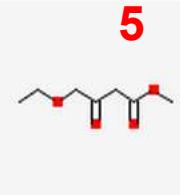
- PubMed (1)
- Taxonomy (1)

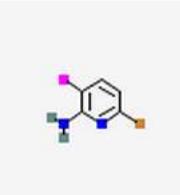
**Related BioAssays:** Activity Overlap (105)

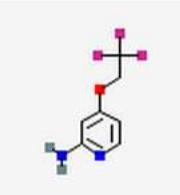
# PubChem – Pesquisa por “Tag”

Lipinski's rule of

0:500[mw] 0:5[hbdc] 0:10[hbac] -5:5[logP]

1.  Methyl 4-ethoxy-3-oxobutanoate; AK141825; 415678-65-8  
MW: 160.167780 g/mol MF: C<sub>7</sub>H<sub>12</sub>O<sub>4</sub>  
IUPAC name: methyl 4-ethoxy-3-oxobutanoate  
CID: 54303951  
[Summary](#)

2.  6-bromo-3-iodopyridin-2-amine; AK142103; 1245643-34-8  
MW: 298.907130 g/mol MF: C<sub>5</sub>H<sub>4</sub>BrIN<sub>2</sub>  
IUPAC name: 6-bromo-3-iodopyridin-2-amine  
CID: 52987942  
[Summary](#)

3.  AK138368; 4-(2,2,2-Trifluoroethoxy)pyridin-2-amine; 1379361-82-6  
MW: 192.138490 g/mol MF: C<sub>7</sub>H<sub>7</sub>F<sub>3</sub>N<sub>2</sub>O  
IUPAC name: 4-(2,2,2-trifluoroethoxy)pyridin-2-amine  
CID: 15724964  
[Summary](#)

**Actions on your results**

-  BioActivity Analysis  
Analyze the BioActivities of the compounds
-  Structure Clustering  
Cluster structures based on structural similarity
-  Structure Download  
Download the structures in various formats
-  Pathways  
Analyze pathways containing the compounds

**Refine your results**

- What's this?

**Chemical Properties**

Rule of 5 (34,559,871)

**BioActivity Experiments**

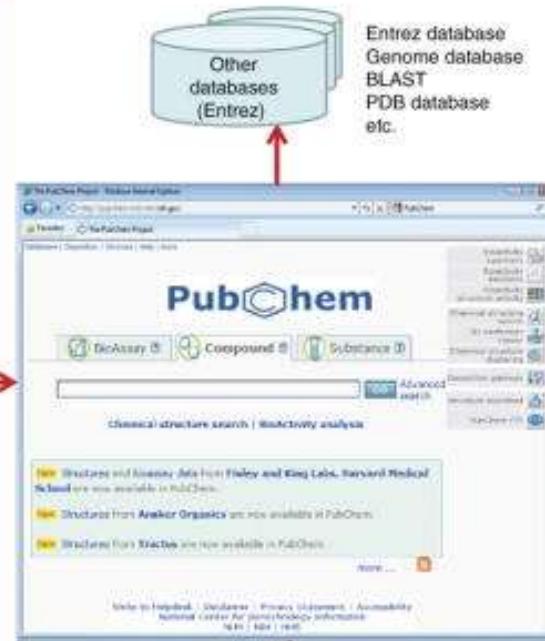
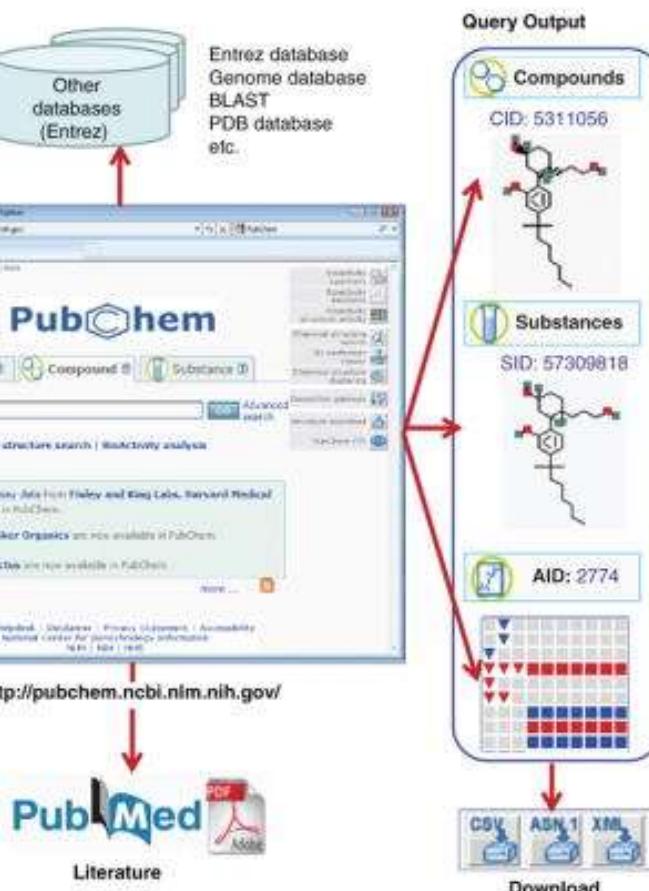
BioAssays, Probes (142)

# PubChem – Pesquisa por estrutura

The screenshot shows the PubChem Structure Search interface. On the left, there's a search bar labeled "PubChem Compound" with dropdown menus for "Limits" and "Advanced". Below it is a "Search By:" section with "Name/Text" and "Identity/Similarity" options. A "Draw a Structure" button is present, with a small chemical structure sketch shown below it. To the right of the search bar is a "PubChem Sketcher V2.4" window. The sketcher has a toolbar at the top with icons for New, Undo, Redo, Cut, Copy, Paste, Delete, Query, and various selection tools. Below the toolbar is a grid of chemical symbols and structures. The SMILES input field in the sketcher contains the string CCC(N1CCCC1C(=O)O)=O. To the right of the sketcher, the resulting chemical structure is displayed as a 2D diagram: CCC(N1CCCC1C(=O)O)=O.

**A.**

Bioassay Data Source Name	Bioassay count	Substance count
<b>BioAssay Data Deposited by NIH MLPPCN and MLSCN</b>		
NCGC (NIH)	485	398,461
The Scripps Research Institute Molecular Screening Center	483	357,929
Burnham Center for Chemical Genomics	397	400,255
NMMLSC (University of Mexico)	230	348,231
Broad Institute of MIT and Harvard	179	334,761
Vanderbilt Screening Center for GPCRs, Ion Channels & Transporters	101	223,904
SRMLSC (Southern Research Institute)	89	226,666
Johns Hopkins Ion Channel Center	74	305,806
University of Pittsburgh Molecular Library Screening Center	70	222,637
Southern Research Specialized Biocontainment Screening Center	63	339,742
PCMD (Penn Center for Molecular Discovery)	57	226,345
Emory University Molecular Libraries Screening Center	54	370,189
Columbia University Molecular Screening Center	33	197,177
<b>BioAssay Data Deposited by Other Sources</b>		
ChEMBL (European Bioinformatics Institute, EBI)	446,639	551,496
DTP/NCI (NIH)	173	189,809
ChemBank (Broad Institute of Harvard & MIT/Chemical Biology)	106	5,329
SGCOxCompounds (SGC Oxford)	43	319
NINDS Approved Drug Screening Program	34	1,040
BindingDB (CARB)	20	3,285
Diabetic Complications Screening (NIDDK/JDRF)	14	1,040
EPA DSSTox (National Center for Computational Toxicology)	12	4,099
GLIDA, GPCR-Ligand Database	6	19,474
GlaxoSmithKline (GSK)	6	13,533
ProbeDB (NCBI)	5	279
MTDP (CCR, NCI, NIH)	4	99,933
IUPHAR-DB	4	104
Structural Genomics Consortium	2	28
The Genomics Institute of the Novartis Research Foundation (GNF)	1	33,364
Shanghai Institute of Organic Chemistry	1	3,073
Circadian Research, Kay Laboratory (UCSD)	1	1,279
Thermo Scientific Dharmacon RNAi Technologies	1	840
ChemBlock	1	122
CC_PMLSC	1	47
SGCS to Compounds	1	17
Total: 41	449,402	4,985,224

<http://pubchem.ncbi.nlm.nih.gov/sources/>**B.****C.**

# Exemplo de pesquisa estrutural na base ChEMBL

Screenshot of the ChEMBL website (<https://www.ebi.ac.uk/chembl/>) showing a search interface for small molecules.

The page includes a "Cookies on EMBL-EBI website" notice at the top.

Navigation bar: Services, Research, Training, Industry, About us.

Search bar: Search ChEMBL... (Compounds, Targets, Assays, Documents, Activity Source Filter).

Search tools: Ligand Search, Target Search, Browse Targets, Browse Drugs, Browse Drug Targets, Drug Approvals, About.

Chemical drawing tools: A toolbar with various chemical drawing and selection icons.

ChEMBL Statistics (left sidebar):

- DB: CHEMBL\_17
- Targets: 9,366
- Compound records: 1,520,172
- Distinct compounds: 1,324,941
- Activities: 12,077,491
- Publications: 51,277
- [Release Notes](#)

ChEMBL Blog (left sidebar):

- [EU-OPENSCREEN 3rd Stakeholder Meeting, Oslo, Norway](#)
- [Competition Time - Win](#)

Central content area:

- List Search:** Radio buttons for SMILES Search, CHEMBL ID Search, Keyword Search. Input field: Please enter a list of Compound IDs, keywords, or SMILES separated by newlines. Button: Fetch Compounds.
- Biologicals Blast Search:** Input field. Button: Run BLAST.

Chemical structure displayed in the center: c1ccc(cc1)-c2ccc(cc2) (2-methylbenzene).

# Exemplo de pesquisa estrutural na base ChEMBL

Blues Drums Bund × thelooploft's strea × W Mino Cinelu - Wiki × RCSB Protein Data × Configurar - Apple × Over To You: Mac × RCSB Protein Data × ChEMBL ×

https://www.ebi.ac.uk/chembl/index.php/compound/simresults

**Cookies on EMBL-EBI website**

This website uses cookies to store a small amount of information on your computer, as part of the functioning of the site. Cookies used for the operation of the site have already been set.

To find out more about the cookies we use and how to delete them, see our [Cookie](#) and [Privacy](#) statements.

[Dismiss this notice](#)

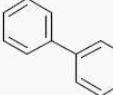
EMBL-EBI

**ChEMBL**

Search ChEMBL... Compounds Targets Assays Documents Activity Source Filter

Compound Synonyms Similarity Max Phase Parent Mol Weight ALogP PSA HBA HBD #RO5 Vio. #Rotatable Bonds Passes Rule of Three Med Chem Friendly QED Weighted

10 records per page Show / hide columns

Compound	Synonyms	Similarity	Max Phase	Parent Mol Weight	ALogP	PSA	HBA	HBD	#RO5 Vio.	#Rotatable Bonds	Passes Rule of Three	Med Chem Friendly	QED Weighted
 CHEMBL14092	SID17390012 E230 SID26753117	100	0	154.21	3.35	0	0	0	0	1	N	Y	.59

Showing 1 to 1 of 1 entries

← Previous 1 Next →

# Exemplo de pesquisa estrutural na base ChEMBL

Screenshot of the ChEMBL Compound Report Card for compound CHEMBL14092.

The page shows the following details:

**Compound Name and Classification**

Compound ID	CHEMBL14092
Compound Name	BiPHENYL
ChEMBL Synonyms	SID17390012   E230   SID26753117
Max Phase	0
Trade Names	
Molecular Formula	C <sub>12</sub> H <sub>10</sub>

**Chemical Structure:** BiPHENYL (Biphenyl)

**Additional synonyms for CHEMBL14092 found using NCI Chemical Identifier Resolver**

**Compound Representations**

Molfile	<a href="#">Download MolFile</a>
Canonical SMILES	c1ccc(cc1)c2ccccc2
Standard InChI	InChI=1S/C12H10/c1-3-7-11(8-4-1)12-9-5-2-6-10-12/h1-10H
Standard InChI Key	ZUOUZKKEUPVFJK-UHFFFAOYSA-N

**Alternate Forms of Compound in ChEMBL**

**Chemical Structure:** BiPHENYL (Biphenyl)

**Compound Bioactivity Summary**

# Drug Bank

- Base de dados bioinformática e cheminformática
- Contem actualmente informação sobre 6711 compostos
- Contém 1447 fármacos aprovados pela FDA
- Combina informação sobre o fármaco (química, farmacológica e farmacêutica) com informação sobre o alvo (sequência, estrutura e via metabólica)
- Cada entrada contem mais de 150 campos

DrugBank

www.drugbank.ca

Gnuplot surprise... Gnuplot tricks Old Nabble - G... Music Production Bioinformatics 10.10.20.1/form... + Pocket Other bookmarks

# DRUGBANK

Open Data Drug & Drug Target Database

Home Browse Search Downloads News & Updates About Help Contact Us

Search: Search DrugBank Search Help / Advanced

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains 6711 drug entries including 1447 FDA-approved small molecule drugs, 131 FDA-approved biotech (protein/peptide) drugs, 85 nutraceuticals and 5080 experimental drugs. Additionally, 4227 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each DrugCard entry contains more than 150 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

DrugBank is supported by [David Wishart](#), Departments of [Computing Science](#) & [Biological Sciences](#), [University of Alberta](#).

DrugBank is also supported by [The Metabolomics Innovation Centre](#), a Genome Canada-funded core facility serving the scientific community and industry with world-class expertise and cutting-edge technologies in metabolomics.

[More about DrugBank](#)

What's New?

Posted February 17, 2012:

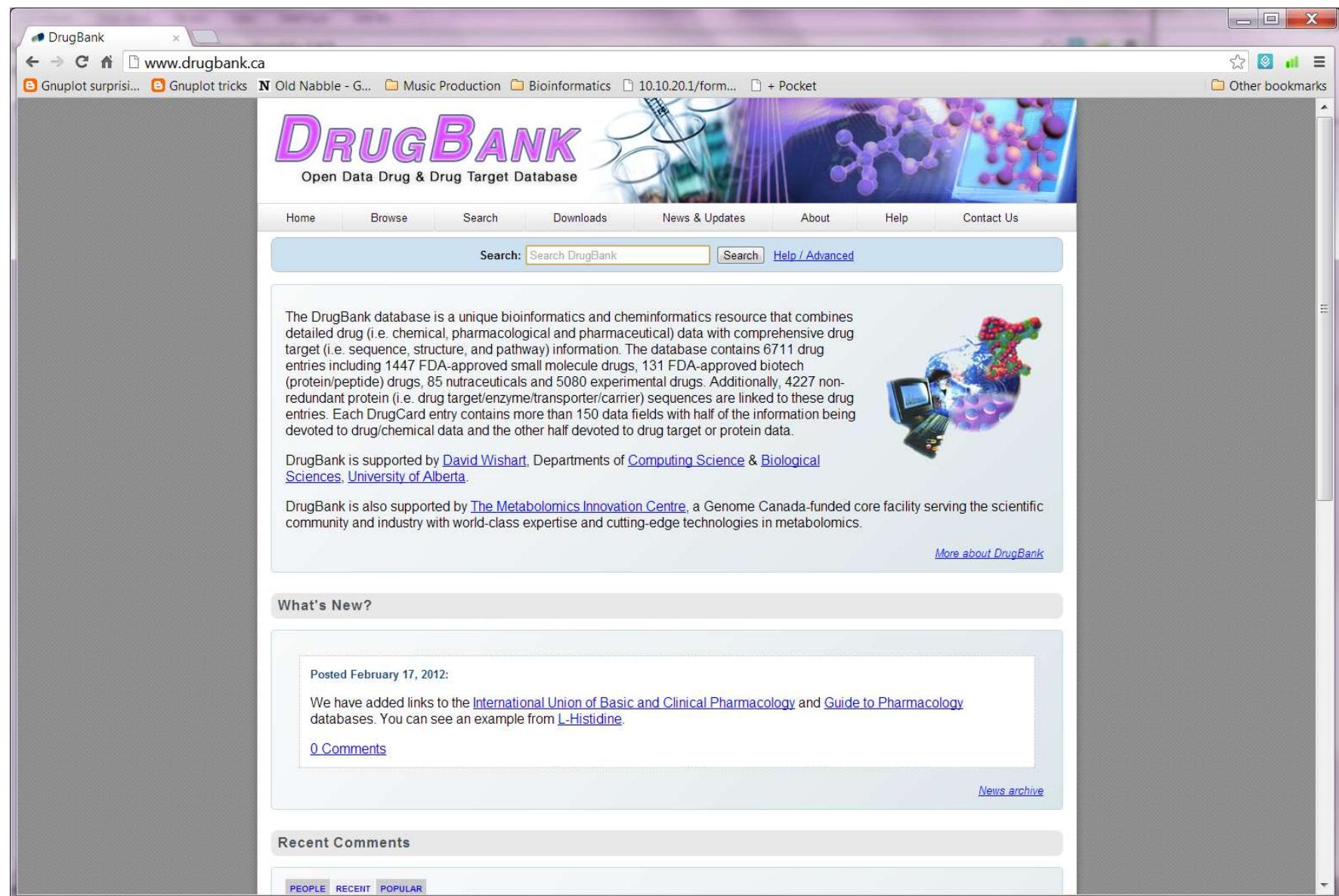
We have added links to the [International Union of Basic and Clinical Pharmacology](#) and [Guide to Pharmacology](#) databases. You can see an example from [L-Histidine](#).

[0 Comments](#)

[News archive](#)

Recent Comments

PEOPLE RECENT POPULAR



DrugBank: Acetylsalicylic acid

www.drugbank.ca/drugs/DB00945

Gnuplot surprise... Gnuplot tricks Old Nabble - G... Music Production Bioinformatics 10.10.20.1/form... + Pocket Other bookmarks

# DRUGBANK

Open Data Drug & Drug Target Database

Home Browse Search Downloads News & Updates About Help Contact Us

Search: Search DrugBank Search Help / Advanced

Identification Taxonomy Pharmacology Pharmacoeconomics Properties References Interactions 0 Comments

targets (3) enzymes (3) transporters (3) carriers (1)

Show Drugs with Similar Structures for All drugs

**Identification**

Name	Acetylsalicylic acid
Accession Number	DB00945 (APRD00264, EXPT00475)
Type	small molecule
Groups	approved
Description	The prototypical analgesic used in the treatment of mild to moderate pain. It has anti-inflammatory and antipyretic properties and acts as an inhibitor of cyclooxygenase which results in the inhibition of the biosynthesis of prostaglandins. Acetylsalicylic acid also inhibits platelet aggregation and is used in the prevention of arterial and venous thrombosis. (From Martindale, The Extra Pharmacopoeia, 30th ed, p5)

**Structure**

Download: MOL | SDF | SMILES | InChI  
Display: 2D Structure | 3D Structure

**Synonyms**

- 2-Acetoxybenzenecarboxylic acid
- 2-Acetoxybenzoic acid
- 2-Carboxyphenyl acetate
- A.S.A.
- Acetilsalicilico
- Acetilum acidulatum
- Acetosalic acid
- Acetoxybenzoic acid
- Acetylsalicylate
- Acetylsalicylaure (GERMAN)
- Acetylsalicylic acid
- Acide acetylsalicylique (FRENCH)