

# Bioinformática Estrutural

Sequência



Estrutura



Função

# Fluxo de informação biológica

Gene ...TTAATAAGT...

↓ transcrição

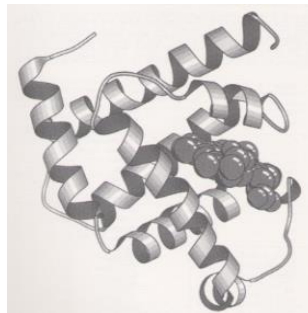
m-RNA ...UUAUAAGU...

↓ splicing, tradução

cadeia polipeptídica ...LISVHDN...

↓ modificações pós-translacionais

proteína



Dogma central da  
biologia molecular

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

# Níveis de organização da estrutura das proteínas

Estrutura primária

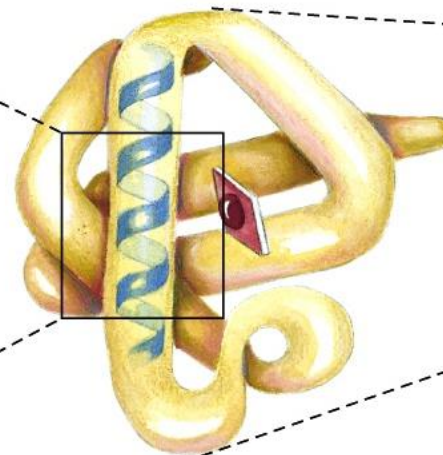
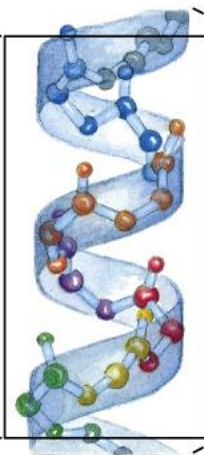
Estrutura secundária

Estrutura terciária

Estrutura quaternária



Sequência de aminoácidos

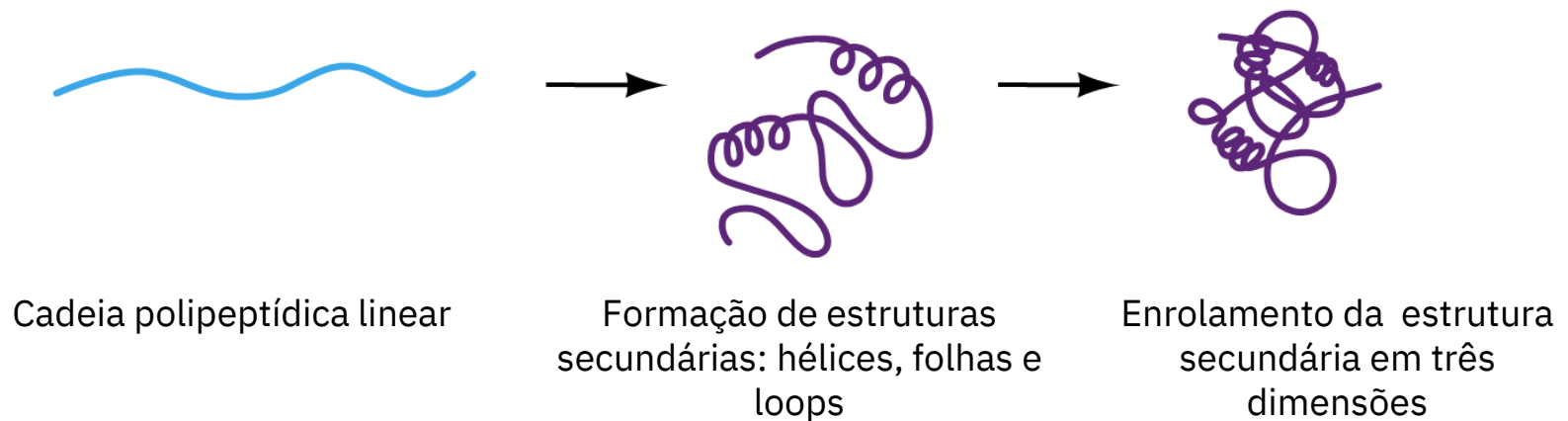


# A estrutura das proteínas é determinada pela sua sequência

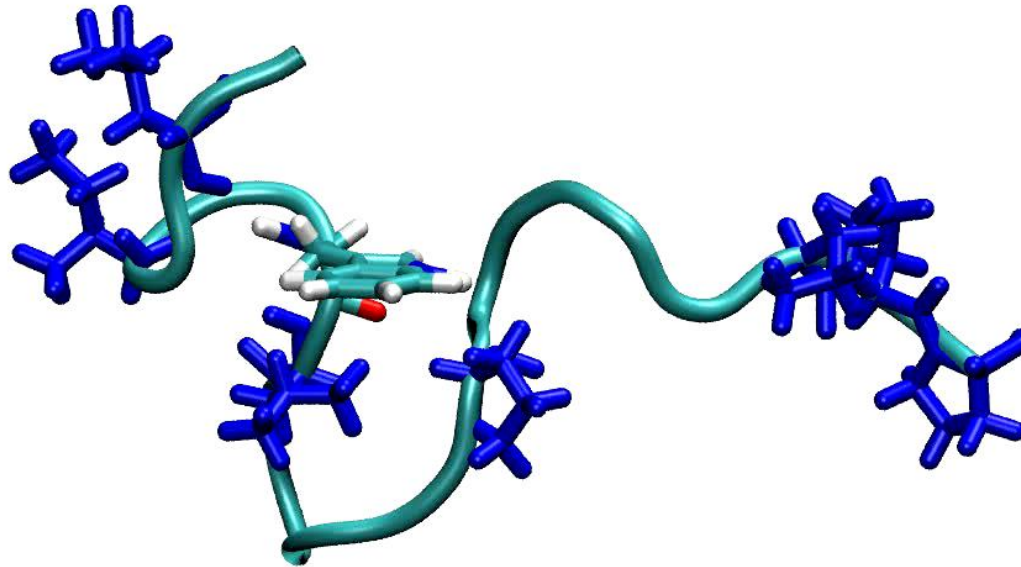
A estrutura tridimensional das proteínas é resultado das interações entre os átomos que a constituem e o meio aquoso. Em muitos casos a cadeia polipeptídica assume a sua conformação *nativa* de modo espontâneo, após a síntese ribossomal. Este processo tem o nome de “protein folding”.

*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



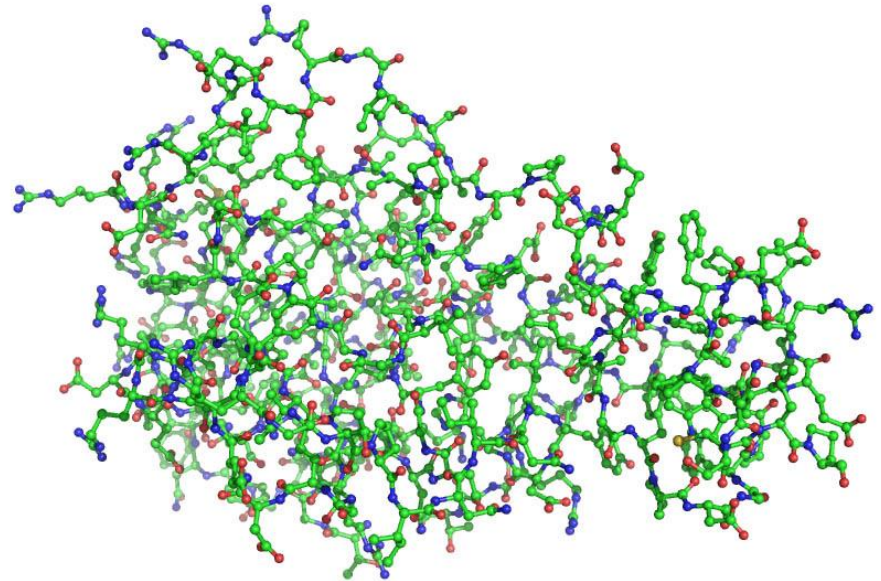
Simulação do *fold*ing da mini-proteína **Trp-cage**

Proteínas mais pequenas ou mais simples frequentemente adquirem a sua estrutura tridimensional *espontaneamente*, sem necessidade de auxílio de outras proteínas. ,

# 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 químicas entre todos os átomos que constituem cada molécula.

...AVAGGATILVHNQDAGEPAIVLAFG...

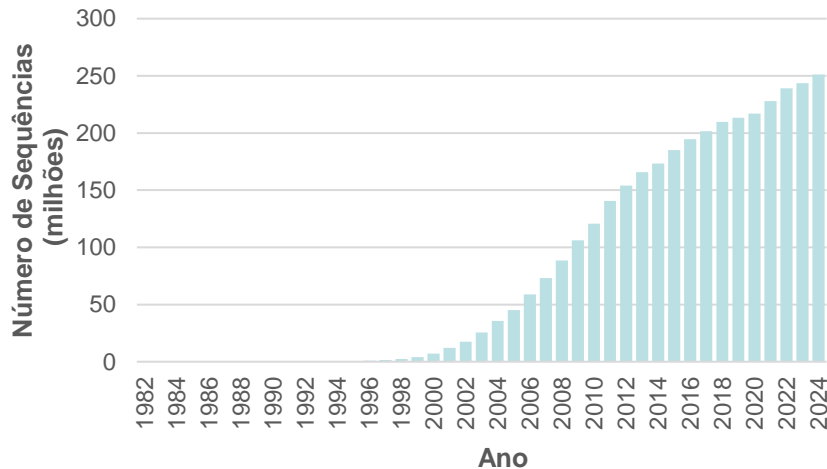


Sequência

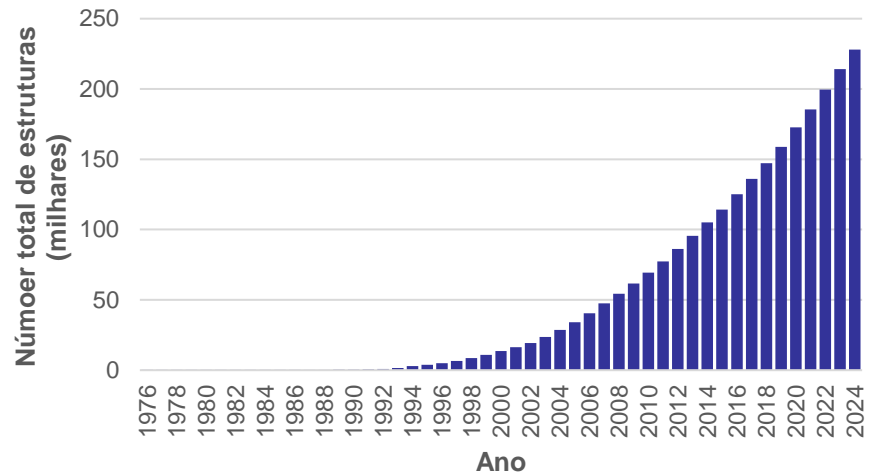
Estrutura

# Sequência *versus* estrutura

Genbank 1982-2024



Protein Databank 1976-2024



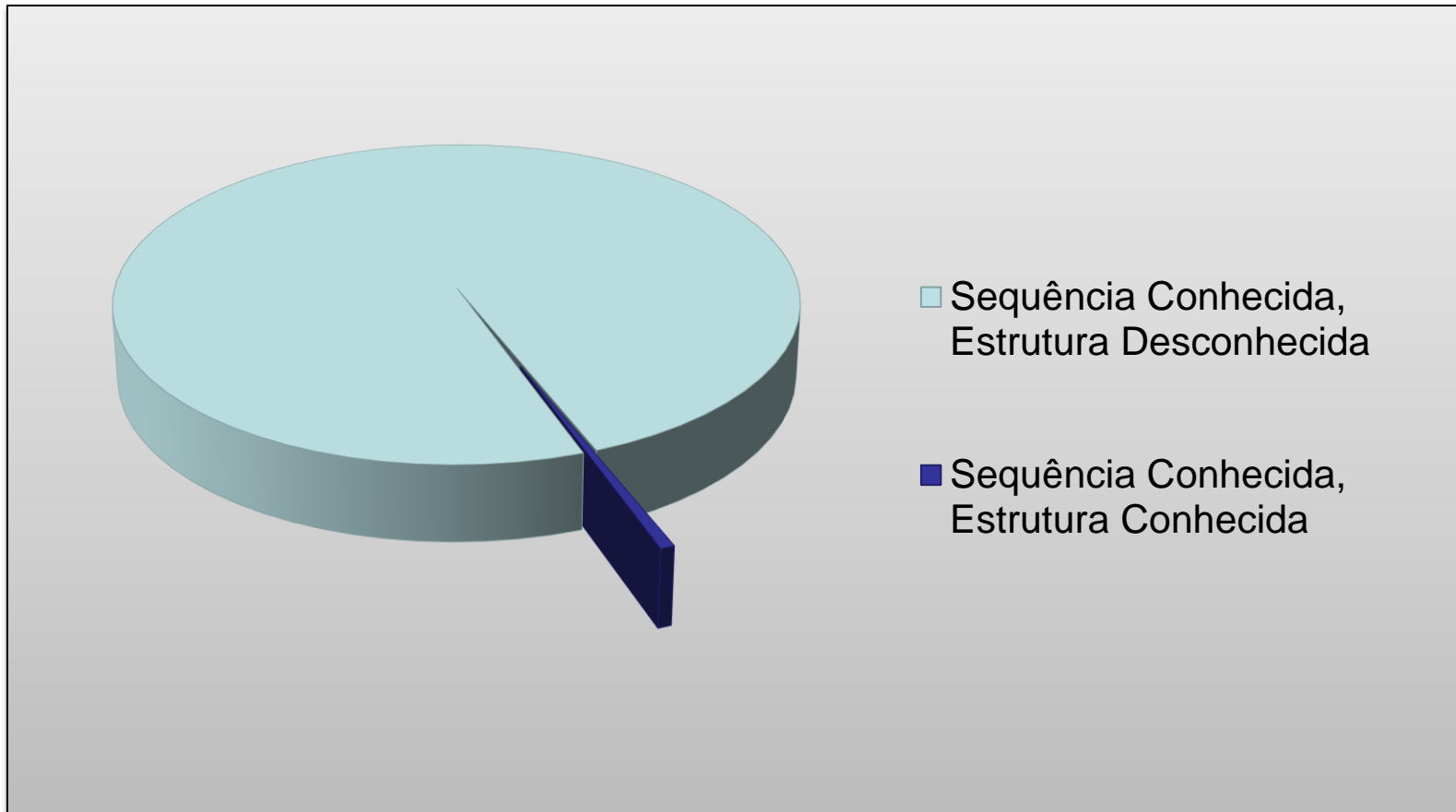
**Em 1982:** conhecidas 172 estruturas e 606 sequências ...

**Hoje (Nov 2024): 227,933** estruturas no Protein Data Bank e **252,347,664** de sequências no GeneBank (mas são conhecidas mais de 2 **biliões**)!!

**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

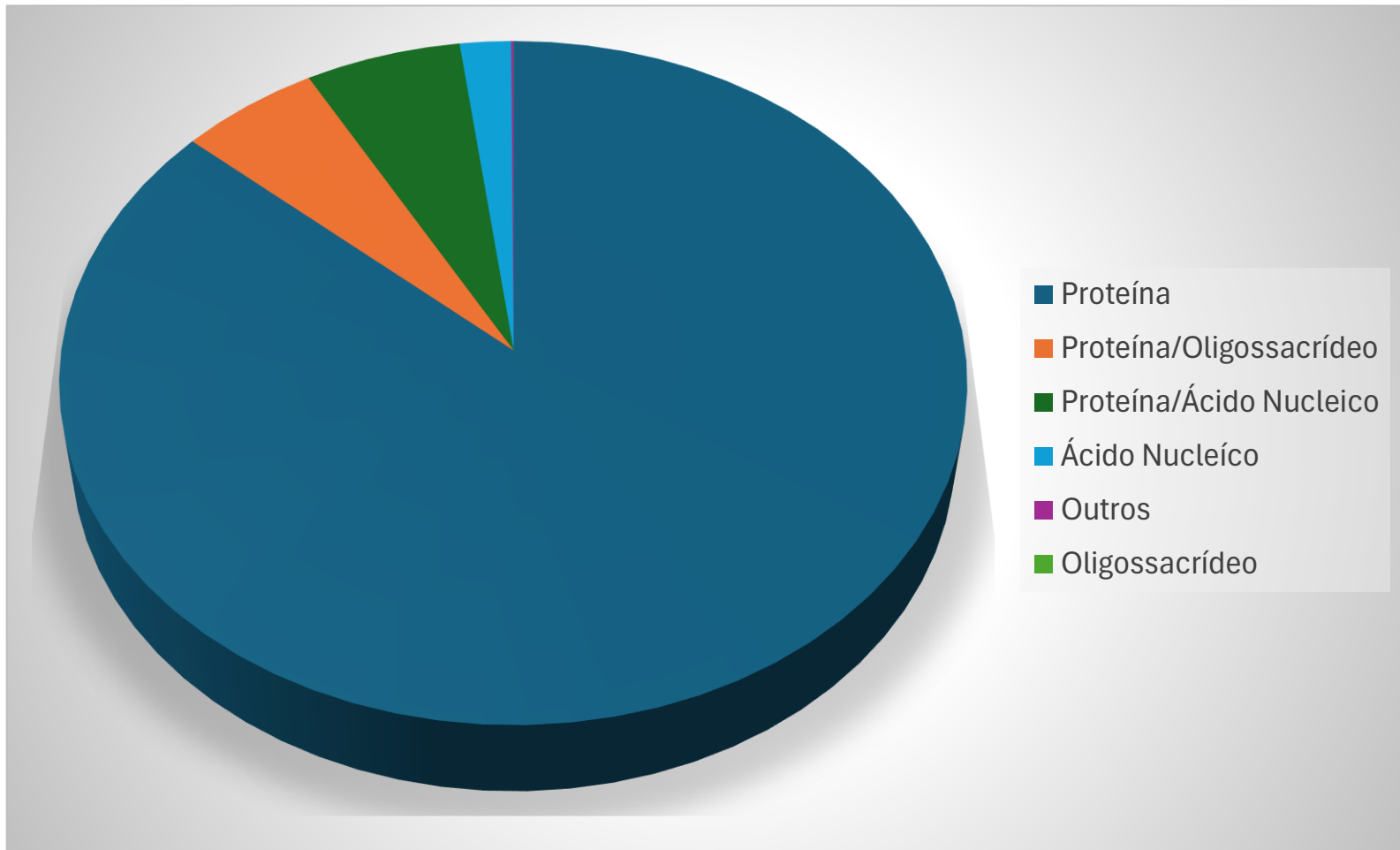
- O desenvolvimento das técnicas de determinação da estrutura molecular levou à acumulação de um número considerável de estruturas de proteínas (~220000)
- Estas estruturas foram determinadas por meio de várias técnicas experimentais, sendo as mais importantes:
  - Difração (cristalografia) de raios X
  - Ressonância magnética nuclear (RMN)
  - Crio-microscopia eletrónica (Cryo-EM)
- A principal base de dados de estruturas de proteínas é o Protein Databank (PDB), que pode ser acedido no portal <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 RCSB (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 contém hoje em dia outros tipos de moléculas, tais como ácidos nucleicos, lípidos e polissacáridos.
- Número total de estruturas em 2/12/2024: **227933**

Tipo de Molécula	Técnica Experimental						Total
	X-ray	EM	NMR	Combinação	Neutrões	Outros	
Proteína	168,016	16,009	12,543	208	79	32	196,887
Proteína/Oligossacrídeo	9,682	2,699	34	8	2	0	12,425
Proteína/Ácido Nucleico	8,764	4,785	286	7	0	0	13,842
Ácido Nucleico	2,872	142	1,512	14	3	1	4,544
Outros	170	10	33	0	0	0	213
Oligossacrídeo	11	0	6	1	0	4	22
<b>Total</b>	<b>189,515</b>	<b>23,645</b>	<b>14,414</b>	<b>238</b>	<b>84</b>	<b>37</b>	<b>227,933</b>

# O Protein Data Bank contem vários tipos de macromoléculas



# De onde provêm a informação estrutural ?

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

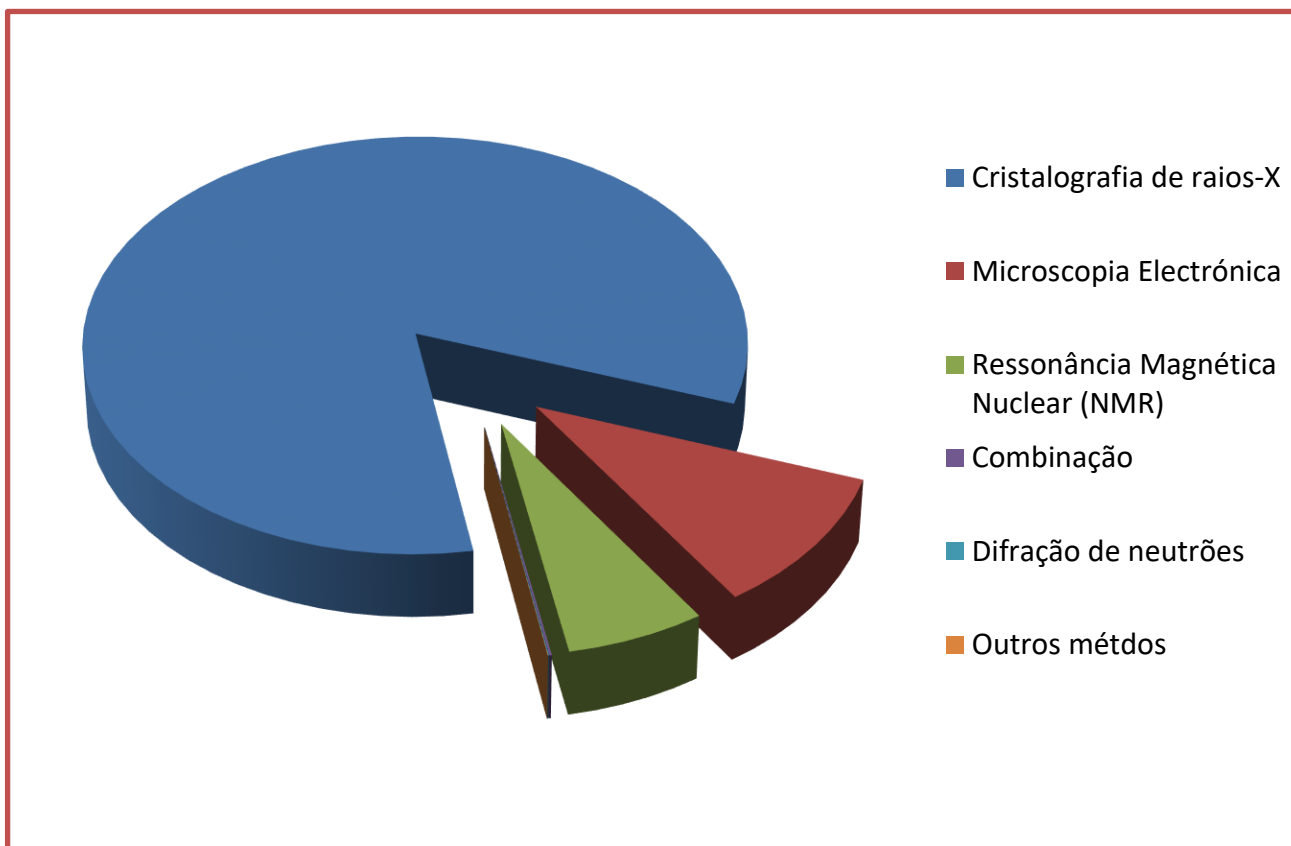
- Teoria da ligação química
- Geometria de moléculas pequenas
- Métodos experimentais para a determinação da estrutura:
  - ❖ Cristalografia de raios X
  - ❖ Crio-microscopia electrónica (Cryo-EM)
  - ❖ Ressonância Magnética Nuclear (NMR)
  - ❖ Outros métodos (difração de neutrões, SAXS, etc...)



# Métodos experimentais

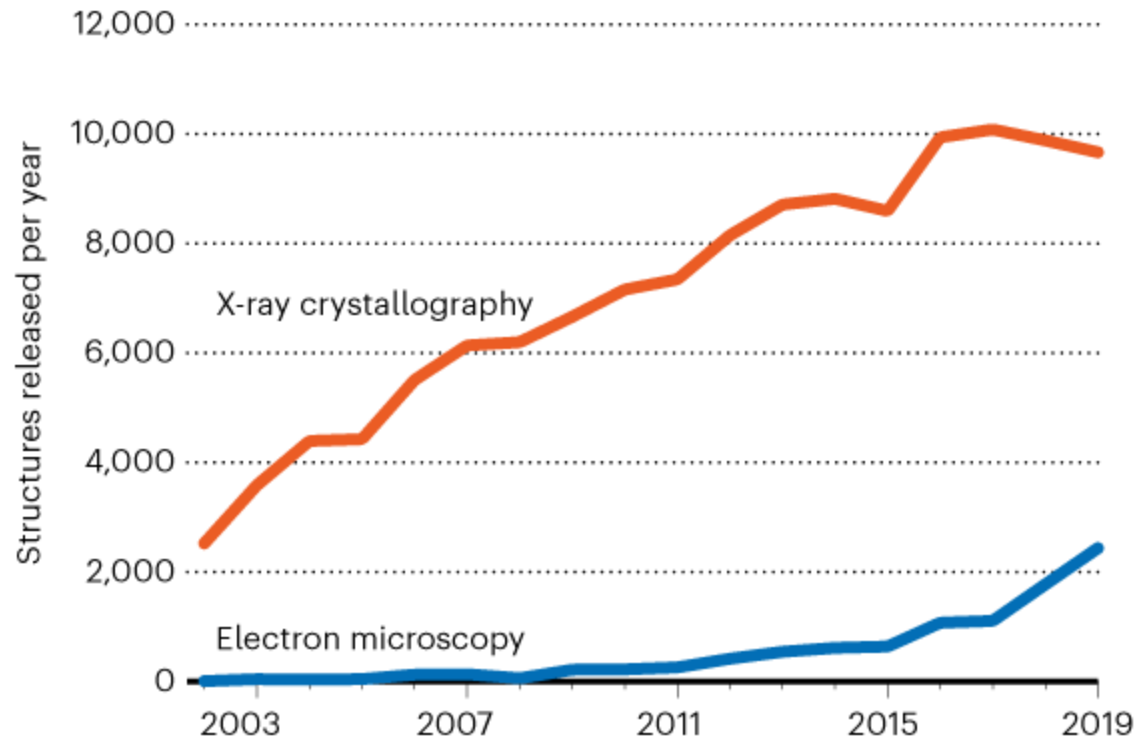
- **Cristalografia de raios X:** a molécula a estudar é purificada e cristalizada a partir de uma solução concentrada. Um feixe de raios X é projectado através do cristal da molécula e o padrão de difracção obtido é usado para resolver a estrutura.
- **Ressonância magnética Nuclear:** a molécula purificada 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,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ), 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 das estruturas do PDB são obtidas por cristalografia de raios X



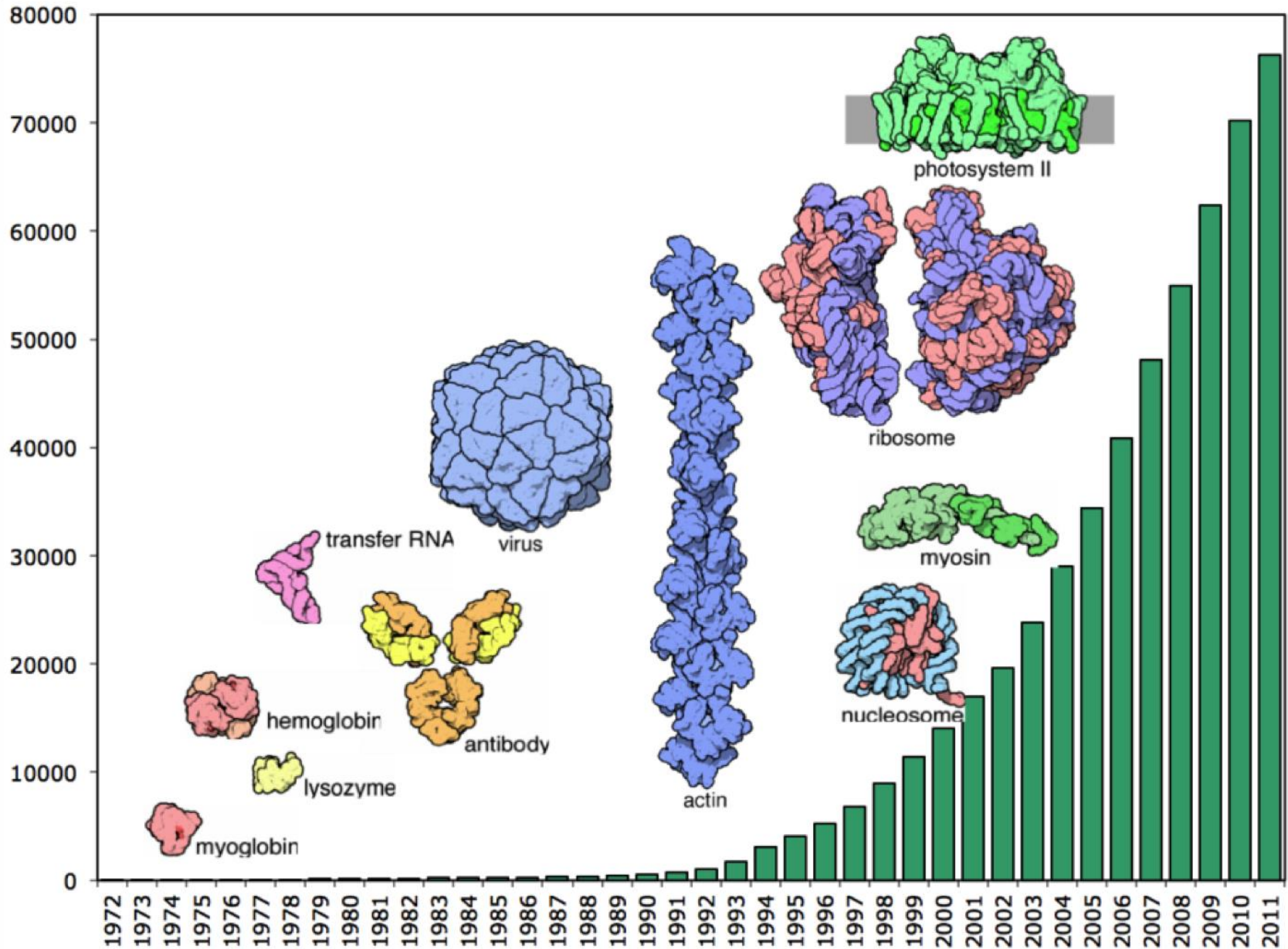
## STRUCTURE SLEUTHS

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.



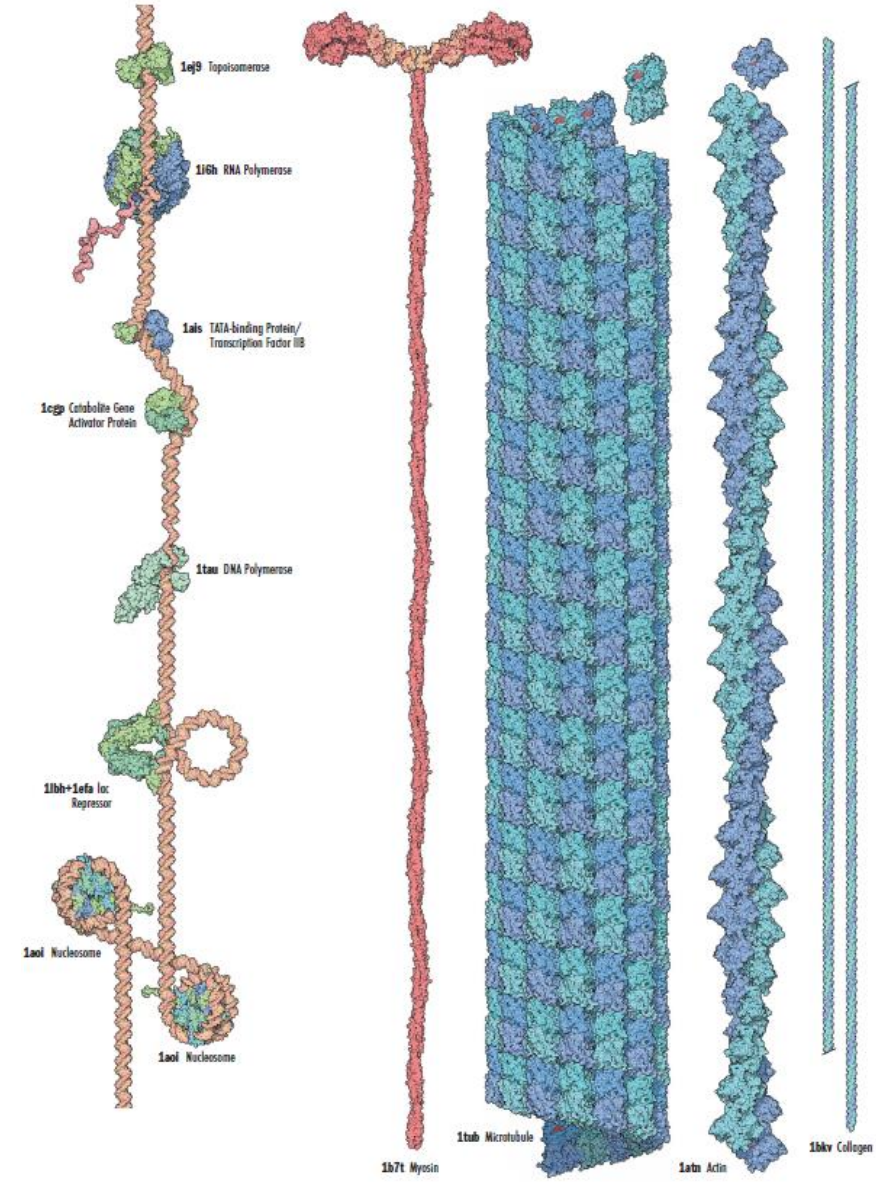
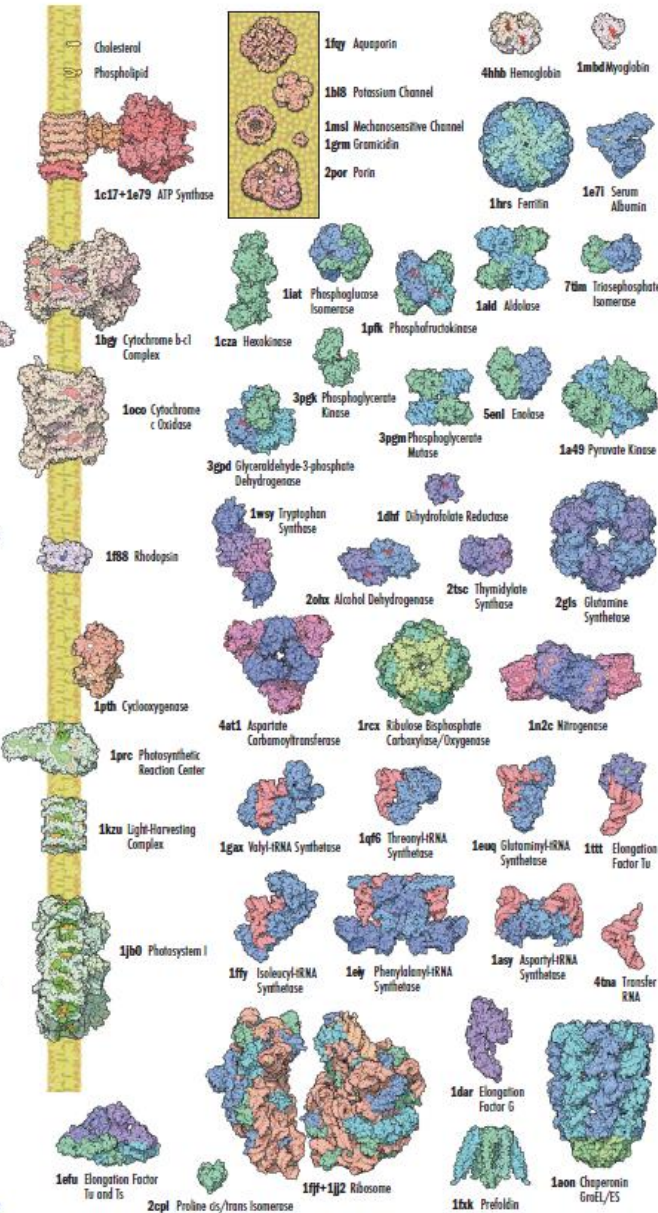
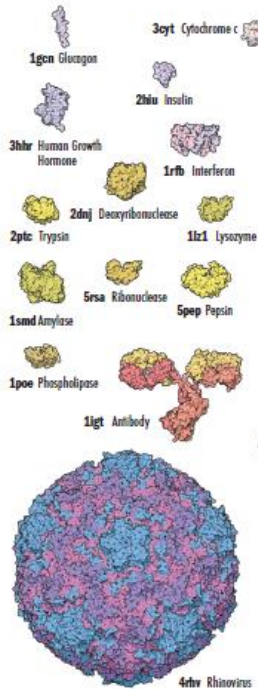
The electron microscopy line shows structures submitted to the Electron Microscopy Data Bank. Nearly all use cryo-EM.

# 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 Protein Data Bank

<https://www.rcsb.org>

RCSB PDB Deposit Search Visualize Analyze Download Learn About Careers COVID-19

Help Contact us MyPDB

RCSB PDB  
PROTEIN DATA BANK

227,933 Structures from the PDB

1,068,577 Computed Structure Models (CSM)

Enter search term(s), Entry ID(s), or sequence

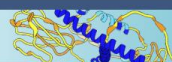
Include CSM



Advanced Search | Browse Annotations

Help

PDB-101 PDB EMDataResource NAKB wwPDB Foundation PDB-Dev



Access Computed Structure Models (CSMs) of available model organisms [Learn more](#)

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

Experimentally-determined 3D structures from the **Protein Data Bank (PDB)** archive

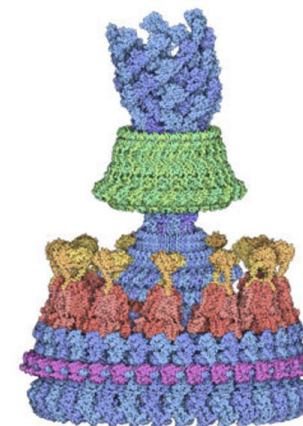
**Computed Structure Models (CSM)** from AlphaFold DB and ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

Explore NEW Features

PDB-101 Training Resources

## December Molecule of the Month



Flagellar Motor

Latest Entries

As of Tue Nov 26 2024

Features & Highlights

News

Publications

Access SDF/MOL formatted CCD

December 1: World AIDS Day

lysozyme Include CSM ?

- in Structure Keywords
  - LYSOZYME
  - Immune system, Lysozyme
- in UniProt Molecule Name
  - Lysozyme
  - Lysozyme 1
  - Lysozyme C
  - Lysozyme C I
  - Lysozyme C II
  - Lysozyme C, milk isozyme
  - Lysozyme C-1
  - Lysozyme C-2
  - Lysozyme M1
  - Lysozyme RrrD
- in Additional Structure Keywords

- Welcome
- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

RCSB Protein Data Bank (RCSB PDB) provides science and education by providing visualization, and analysis of:

- Experimentally-determined 3D Bank (PDB) archive
- Computed Structure Models ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

Explore NEW Features



PDB-101 Training Resources

of the Month



Flagellar Motor

Latest Entries As of Tue Nov 26 2024

Features & Highlights

News Publications

- Access SDF/MOL formatted CCD
- December 1: World AIDS Day

Search Summary This query matches 7 Structures.

Refinements

Structure Determination Methodology

experimental (7)

Scientific Name of Source Organism

- Homo sapiens (4)
- Gallus gallus (2)
- Camelus dromedarius (1)
- Tachygllossus aculeatus (1)

Taxonomy

Eukaryota (7)

Experimental Method

X-RAY DIFFRACTION (7)

Polymer Entity Type

Protein (7)

Refinement Resolution (Å)

- 1.5 - 2.0 (5)
- 2.0 - 2.5 (2)

Release Date

- 1995 - 1999 (5)
- 2000 - 2004 (2)

-- Tabular Report -- All Selected

1 to 7 of 7 Structures Page 1 of 1 25 Sort by Score



Explore in 3D

1GWD

Tri-iodide derivative of hen egg-white lysozyme

Evans, G., Bricogne, G.

(2002) Acta Crystallogr D Biol Crystallogr 58: 976

Released 2002-06-06  
 Method X-RAY DIFFRACTION 1.77 Å  
 Organisms Gallus gallus  
 Macromolecule LYSOZYME C (protein)  
 Unique Ligands CL, CMO, EDO, IOD, NA

Download File View File



Explore in 3D

1JKB

HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA

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

(1997) Protein Sci 6: 473-476

Released 1997-05-15  
 Method X-RAY DIFFRACTION 1.66 Å  
 Organisms Homo sapiens  
 Macromolecule LYSOZYME (protein)  
 Unique Ligands NO3

Download File View File

1JKC

Download File View File



Search Summary This query matches 7 Structures.

Refinements

Structure Determination Methodology

experimental (7)

Scientific Name of Source Organism

- Homo sapiens (4)
- Gallus gallus (2)
- Camelus dromedarius (1)
- Tachyglossus aculeatus (1)

Taxonomy

Eukaryota (7)

Experimental Method

X-RAY DIFFRACTION (7)

Polymer Entity Type

Protein (7)

Refinement Resolution (Å)

- 1.5 - 2.0 (5)
- 2.0 - 2.5 (2)

Release Date

- 1995 - 1999 (5)
- 2000 - 2004 (2)

-- Tabular Report -- All Selected

1 to 7 of 7 Structures Page 1 of 1 25 Sort by Score



Explore in 3D

1GWD

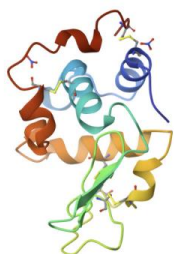
Tri-iodide derivative of hen egg-white lysozyme

Evans, G., Bricogne, G.

(2002) Acta Crystallogr D Biol Crystallogr 58: 976

Released 2002-06-06  
 Method X-RAY DIFFRACTION 1.77 Å  
 Organisms Gallus gallus  
 Macromolecule LYSOZYME C (protein)  
 Unique Ligands CL, CMO, EDO, IOD, NA

Download File View File



Explore in 3D

1JKB

HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA

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

(1997) Protein Sci 6: 473-476

Released 1997-05-15  
 Method X-RAY DIFFRACTION 1.66 Å  
 Organisms Homo sapiens  
 Macromolecule LYSOZYME (protein)  
 Unique Ligands NO3

Download File View File

1JKC

Download File View File

Biological Assembly 1 ?



# 1JKB

HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA

PDB DOI: <https://doi.org/10.2210/pdb1JKB/pdb>

**Classification:** **LYSOZYME**  
**Organism(s):** Homo sapiens  
**Expression System:** Saccharomyces cerevisiae  
**Mutation(s):** Yes ⓘ

**Deposited:** 1996-11-13 **Released:** 1997-05-15  
**Deposition Author(s):** Muraki, M., Harata, K., Goda, S., Nagahora, H.

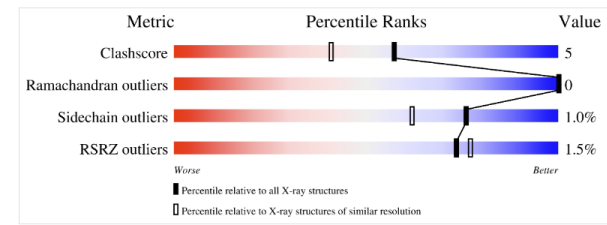
### Experimental Data Snapshot

**Method:** X-RAY DIFFRACTION  
**Resolution:** 1.66 Å  
**R-Value Free:** 0.202  
**R-Value Work:** 0.178  
**R-Value Observed:** 0.178

**Starting Model:** experimental  
[View more details](#)

### wwPDB Validation ⓘ

3D Report Full Report



**Explore in 3D:** [Structure](#) | [Sequence Annotations](#) | [Electron Density](#) | [Validation Report](#) | [Ligand Interaction \(NO3\)](#)

**Global Symmetry:** Asymmetric - C1 ⓘ  
**Global Stoichiometry:** Monomer - A1 ⓘ

- Total Structure Weight: 14.79 kDa ⓘ
- Atom Count: 1,122 ⓘ
- Modelled Residue Count: 130 ⓘ
- Deposited Residue Count: 130 ⓘ
- Unique protein chains: 1

### 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.](#)

(1997) Protein Sci **6**: 473-476

PubMed: [9041653](#) [Search on PubMed](#) [Search on PubMed Central](#)

DOI: <https://doi.org/10.1002/pro.5560060227>

Primary Citation of Related Structures:

[1JKA](#), [1JKB](#), [1JJC](#), [1JKD](#)

#### 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 the replacement of Glu 35 by Asp or Ala, and the replacement of Trp 109 by Phe or Ala result...

[View More](#)


#### Organizational Affiliation:

Biomolecules Department, National Institute of Bioscience and Human-Technology, Ibaraki, Japan. [muraki@nibh.go.jp](mailto:muraki@nibh.go.jp)

## Macromolecules

Find similar proteins by: [Sequence](#) (by identity cutoff) | [3D Structure](#)

### Entity ID: 1


Molecule	Chains ⓘ	Sequence Length	Organism	Details	Image
LYSOZYME	A	130	<a href="#">Homo sapiens</a>	Mutation(s): 1 ⓘ Gene Names: <a href="#">A SYNTHETI</a> <a href="#">C GENE OF HUMAN LYSO</a> EC: <a href="#">3.2.1.17</a>	

### UniProt & NIH Common Fund Data Resources

Structure Summary Structure Annotations Experiment Sequence Genome Versions

Display Files Download Files Data API

Biological Assembly 1 ?



Explore in 3D: [Structure](#) | [Sequence Annotations](#) | [Electron Density](#) | [Validation Report](#) | [Ligand Interaction \(NO3\)](#)

**Global Symmetry:** Asymmetric - C1 **i**  
**Global Stoichiometry:** Monomer - A1 **i**

[Find Similar Assemblies](#)

Biological assembly 1 assigned by authors.

# 1JKB

HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA

**PDB DOI:** <https://doi.org/10.2210/pdb1JKB/pdb>

**Classification:** **LYSOZYME**  
**Organism(s):** *Homo sapiens*  
**Expression System:** *Saccharomyces cerevisiae*  
**Mutation(s):** Yes **i**

**Deposited:** 1996-11-13 **Released:** 1997-05-15  
**Deposition Author(s):** Muraki, M., Harata, K., Goda, S., Nagahora, H.

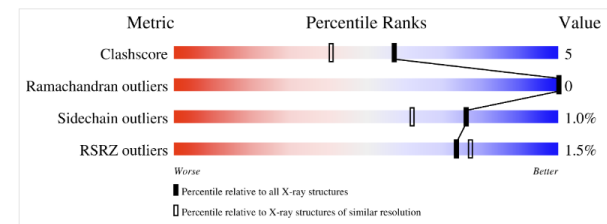
### Experimental Data Snapshot

**Method:** X-RAY DIFFRACTION  
**Resolution:** 1.66 Å  
**R-Value Free:** 0.202  
**R-Value Work:** 0.178  
**R-Value Observed:** 0.178

**Starting Model:** experimental  
[View more details](#)

### wwPDB Validation **i**

3D Report Full Report



This is version 1.6 of the entry. See complete [history](#).

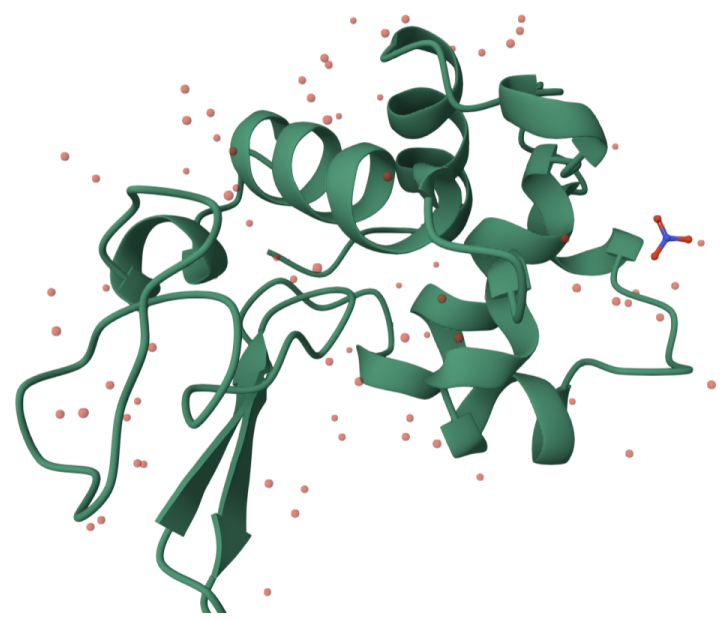
# HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA

Help

Sequence of 1JKB | HUMA... Chain 1: LYSOZYME A

```

1      11      21      31      41      51      61      71      81      91     101     111
KVFERCELARTLKRLGMDGYRGISLANWMCLAKWASGYNTRATNYNAGDRSTDYGI FQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVDRDPQGI RAWVAWRNRCQN
121
RDVQRQYVQGGCV
    
```



**Structure**

1JKB | HUMAN LYSOZYME MUTANT ...

Type Assembly

Asm Id 1: Author Defined Asse...

Dynamic Bonds Off

Nothing Focused

**Measurements**

**Structure Motif Search**

**Components** 1JKB

Preset	+ Add		
Polymer	Cartoon	<input type="checkbox"/>	...
Water	Ball & Stick	<input type="checkbox"/>	...
Ion	Ball & Stick	<input type="checkbox"/>	...
Unit Cell P 21 21 21		<input type="checkbox"/>	...

**Density**

**Quality Assessment**

**Assembly Symmetry**

**Export Models**

**Export Animation**

**Export Geometry**

# Formatos de representação da estrutura

- A representação da estrutura molecular em bancos de dados passa pela descrição das **coordenadas atômicas**, 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

- Function
- Names & Taxonomy
- Subcellular Location
- Disease & Variants
- PTM/Processing
- Expression
- Interaction
- Structure**
- Family & Domains
- Sequence
- Similar Proteins

## P61626 · LYSC\_HUMAN

Protein<sup>i</sup> | Lysozyme C Amino acids | 148 [\(go to sequence\)](#)  
Gene<sup>i</sup> | LYZ Protein existence<sup>i</sup> | Evidence at protein level  
Status<sup>i</sup> |  UniProtKB reviewed (Swiss-Prot) Annotation score<sup>i</sup> |  5/5  
Organism<sup>i</sup> | Homo sapiens (Human)

Entry Variant viewer **150** Feature viewer Genomic coordinates Publications External links History

 Tools

### Function<sup>i</sup>

Lysozymes have primarily a bacteriolytic function; those in tissues and body fluids are associated with the monocyte-macrophage system and enhance the activity of immunogens.

### Miscellaneous

Lysozyme C is capable of both hydrolysis and transglycosylation; it shows also a slight esterase activity. It acts rapidly on both peptide-substituted and unsubstituted peptidoglycan, and slowly on chitin oligosaccharides.

### Catalytic activity<sup>i</sup>

Hydrolysis of (1->4)-beta-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in a peptidoglycan and between N-acetyl-D-glucosamine residues in chitodextrins.  
EC:3.2.1.17 (UniProtKB | ENZYME [E](#) | Rhea [E](#) )

### Features




Showing features for active site<sup>i</sup>.



MKALIVLGLVLLSVTVQGKVFERCELARTLKRLLGMDGYRGISLANWMLAKWESGYNTRATNYNAGDRSTDYGFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVDRDPQGIRAWVAWRNRCQNRDVRQYVQCGV


TYPE	ID	POSITION(S)	DESCRIPTION
-- Select --			











# Interligação entre Uniprot e PDB



UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB  Advanced | List Search    Help

Function Variant viewer **150** Feature viewer Genomic coordinates Publications External links History

## Structure<sup>i</sup>



SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
<input type="text" value="-- Select --"/>		<input type="text" value="-- Select --"/>				
PDB	133L	X-ray	1.77 Å	A	19-148	PDBe · <b>RCSB-PDB</b> · PDBj · PDBsum  · Foldseek
PDB	134L	X-ray	1.77 Å	A	19-148	PDBe · RCSB-PDB · PDBj · PDBsum  · Foldseek
PDB	1B5U	X-ray	1.80 Å	A	19-148	PDBe · RCSB-PDB · PDBj · PDBsum  · Foldseek
PDB	1B5V	X-ray	2.17 Å	A	19-148	PDBe · RCSB-PDB · PDBj · PDBsum  · Foldseek
PDB	1B5W	X-ray	2.17 Å	A	19-148	PDBe · RCSB-PDB · PDBj · PDBsum  · Foldseek
PDB	1B5X	X-ray	2.00 Å	A	19-148	PDBe · RCSB-PDB · PDBj · PDBsum  · Foldseek
PDB	1B5Y	X-ray	2.20 Å	A	19-148	PDBe · RCSB-PDB · PDBj · PDBsum  · Foldseek
PDB	1B5Z	X-ray	2.20 Å	A/B	19-148	PDBe · RCSB-PDB · PDBj · PDBsum  · Foldseek
PDB	1B7L	X-ray	1.80 Å	A	19-148	PDBe · RCSB-PDB · PDBj · PDBsum  · Foldseek
PDB	1B7M	X-ray	2.20 Å	A	19-148	PDBe · RCSB-PDB · PDBj · PDBsum  · Foldseek

 Feedback  Help

# Interligação entre Uniprot e PDB

RCSB PDB Deposit Search Visualize Analyze Download Learn About Careers COVID-19 Help Contact us MyPDB

RCSB PDB PROTEIN DATA BANK 227,933 Structures from the PDB 1,068,577 Computed Structure Models (CSM) Enter search term(s), Entry ID(s), or sequence Include CSM Help

Advanced Search | Browse Annotations

PDB-101 PDB EMDataResource NAKB wwPDB Foundation PDB-Dev

Structure Summary Structure Annotations Experiment Sequence Genome Versions

Biological Assembly 1 ? Display Files Download Files Data API



**133L**  
ROLE OF ARG 115 IN THE CATALYTIC ACTION OF HUMAN LYSOZYME. X-RAY STRUCTURE OF HIS 115 AND GLU 115 MUTANTS

PDB DOI: <https://doi.org/10.2210/pdb133L/pdb>

Classification: **HYDROLASE(O-GLYCOSYL)**  
Organism(s): *Homo sapiens*  
Mutation(s): No

Deposited: 1993-06-01 Released: 1993-10-31  
Deposition Author(s): Harata, K., Muraki, M., Jigami, Y.

**Experimental Data Snapshot**  
Method: X-RAY DIFFRACTION  
Resolution: 1.77 Å  
R-Value Work: 0.181  
R-Value Observed: 0.181

**wwPDB Validation** 3D Report Full Report

Metric	Percentile Ranks	Value
Clashscore		5
Ramachandran outliers		0
Sidechain outliers		5.7%
RSRZ outliers		0

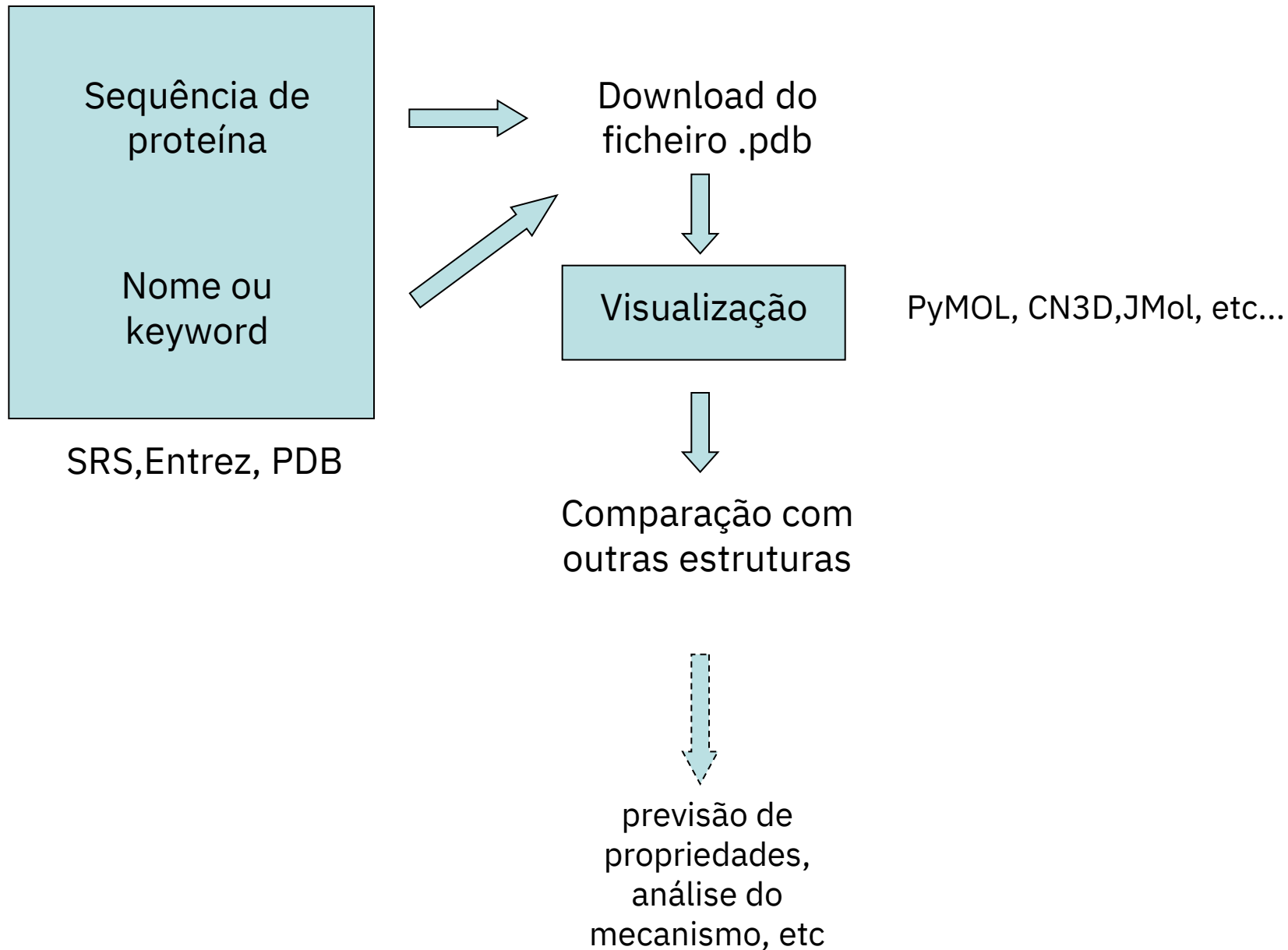
Worse Better  
■ Percentile relative to all X-ray structures  
■ Percentile relative to X-ray structures of similar resolution

Explore in 3D: Structure | Sequence Annotations | Electron Density | Validation Report

Global Symmetry: Asymmetric - C1  
Global Stoichiometry: Monomer - A1

Find Similar Assemblies

# 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

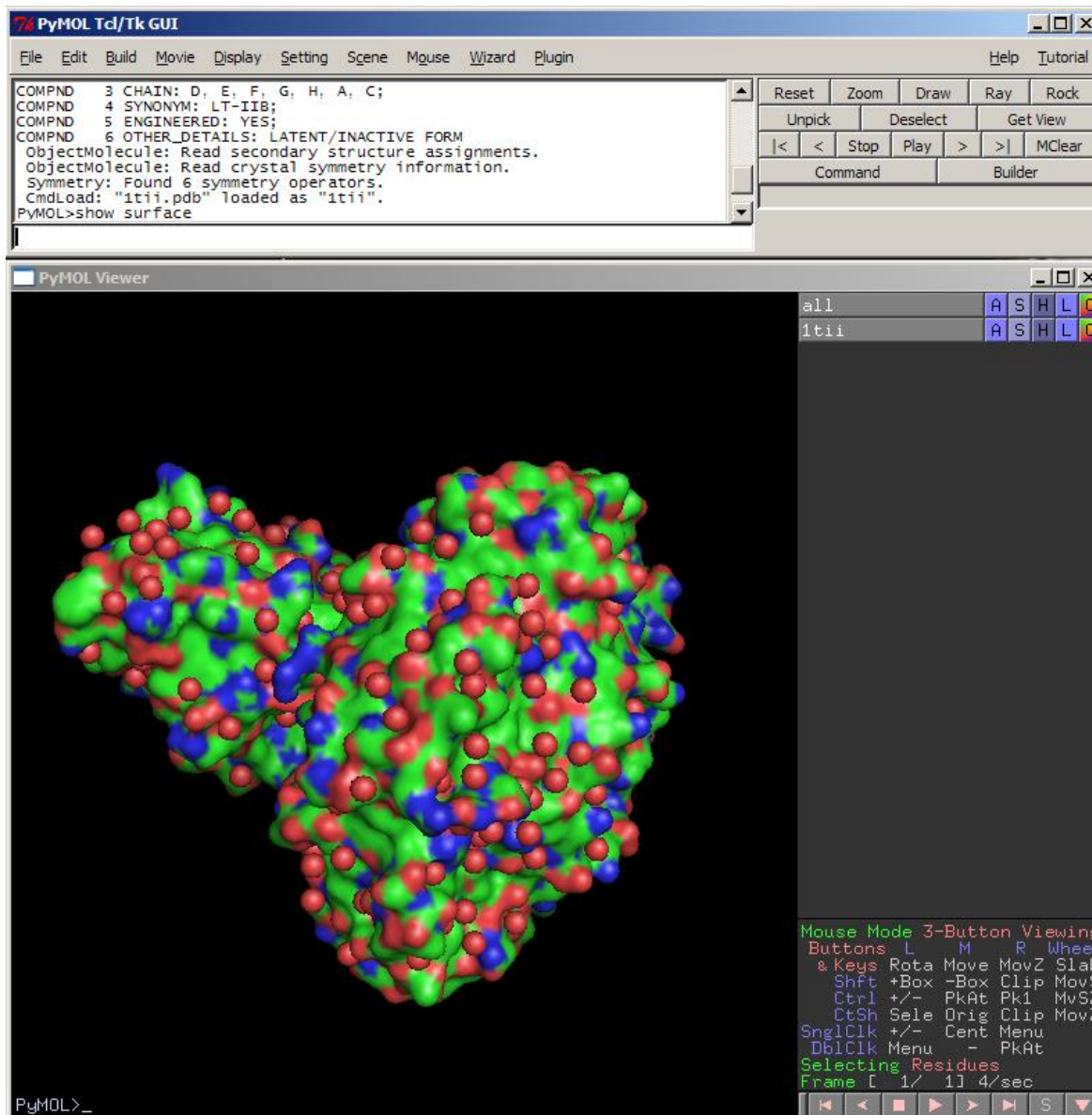
## Instaláveis:

- PyMOL: <http://www.pymol.org>
- VMD : <http://www.ks.uiuc.edu/Research/vmd/>
- Chimera/ChimeraX: <https://www.cgl.ucsf.edu/chimera/>
- SwissPDB viewer: <http://www.expasy.org/spdbv/>

## On-line (web apps):

- Mol\* : <https://molstar.org/viewer/>
- nglviewr: <http://nglviewer.org/>
- ICMJS: <http://www.molsoft.com>
- Jmol/JS Mol: <http://jmol.sourceforge.net/>

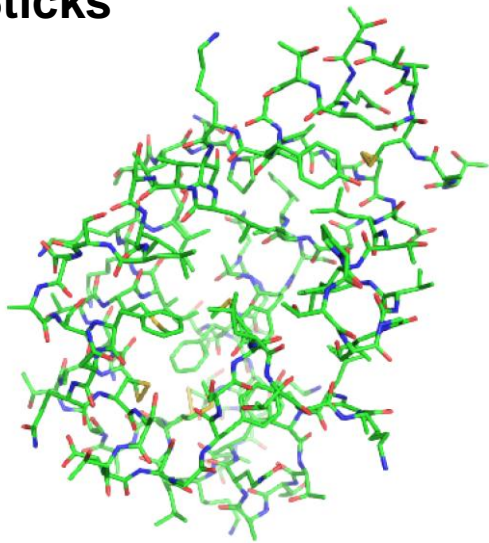
# PyMOL



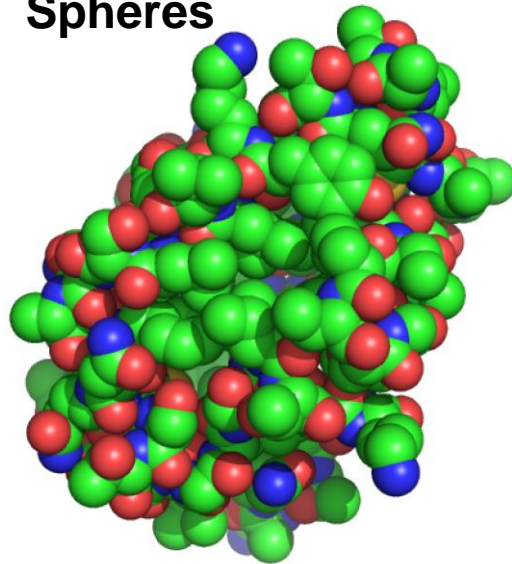
- ❖ Open Source
- ❖ Acesso livre
- ❖ Python / C
- ❖ Visualização de macromoléculas
- ❖ Animações moleculares
- ❖ Comparação de estruturas
- ❖ Scripting
- ❖ Windows / Linux

<http://www.pymol.org>

**Sticks**



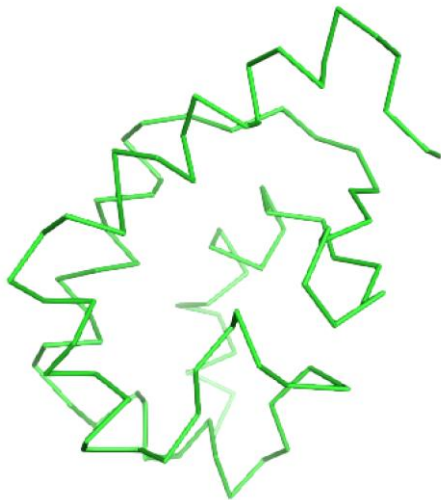
**Spheres**



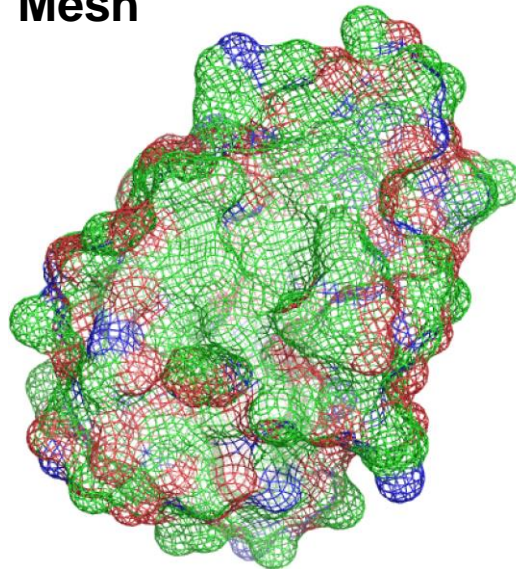
**Cartoon**



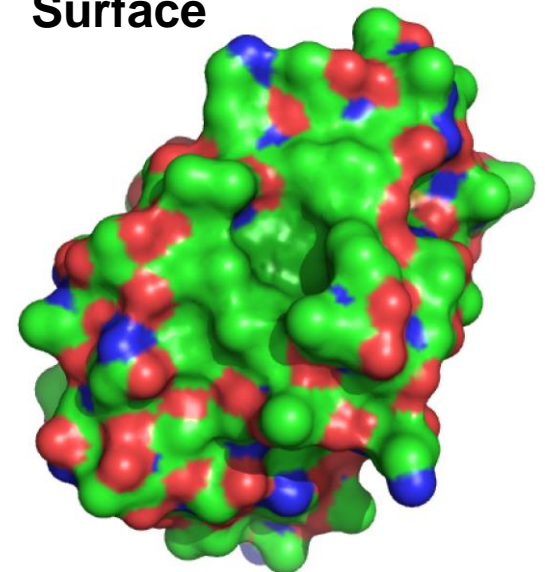
**Ribbon**



**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

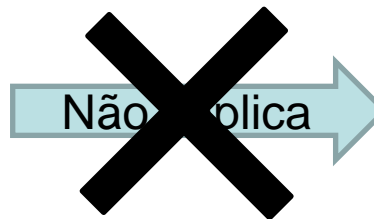
Similaridade de  
sequência



Similaridade de  
estrutura

**MAS**

Similaridade de  
estrutura



Similaridade de  
sequência

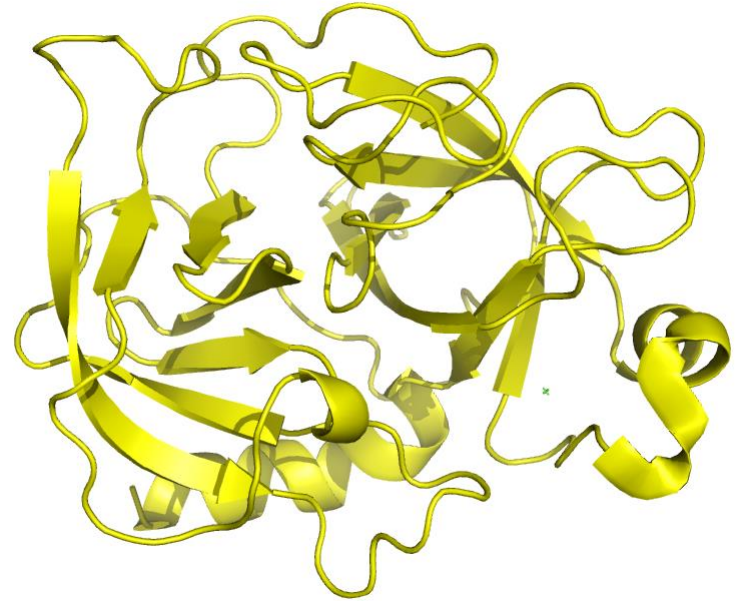
A pressão de selecção evolutiva opera sobre a estrutura (responsável pela função) e não directamente sobre a sequência. Alterações da sequência que conservem a estrutura são geralmente toleradas.

# Similaridade estrutural e de sequência

Tripsina bovina



Tripsina *S. griseus*

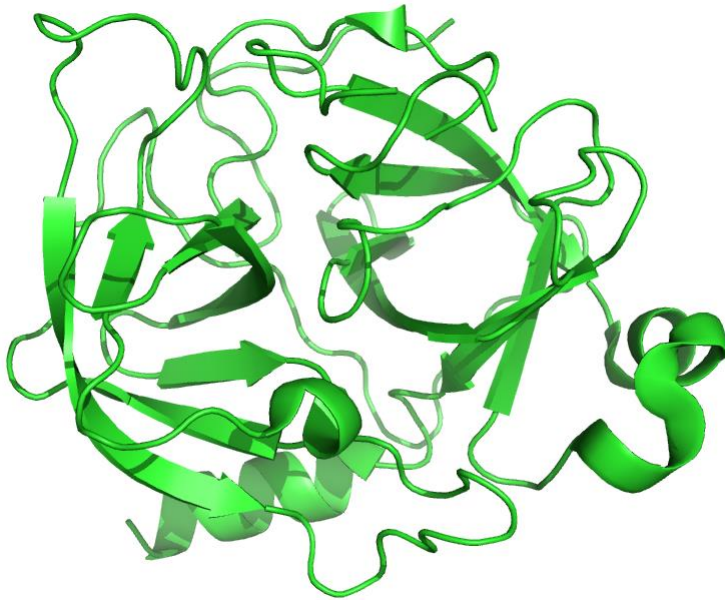


Alinhamento das sequências: **34%** identidade,  $E(1) \mathbf{1.4 \times 10^{-17}}$

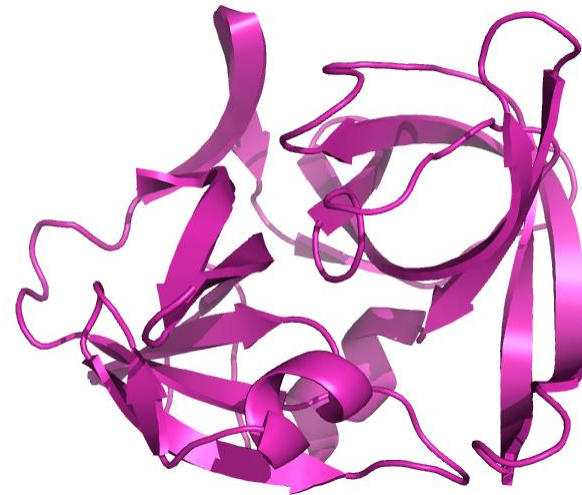
**Proteínas homólogas, similaridade de sequências  
claramente detetável**

# Similaridade estrutural e de sequência

Tripsina bovina



Protease A *S. griseus*



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

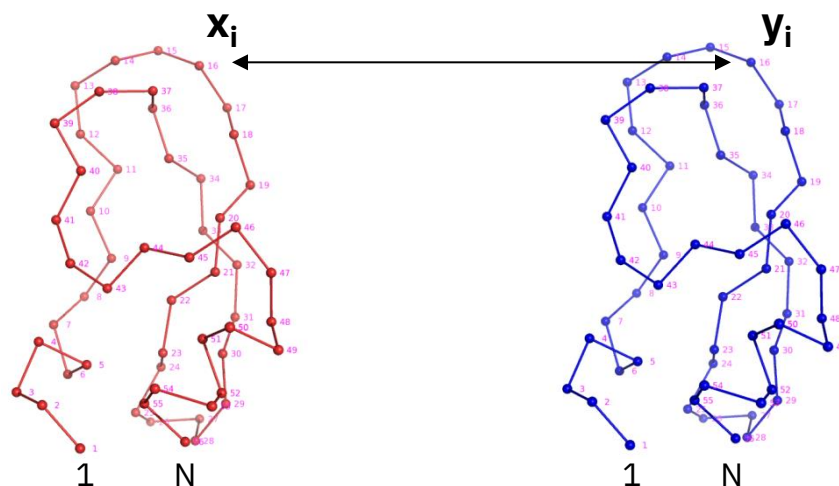
**Proteínas homólogas, similaridade de sequências  
não é detectável**

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

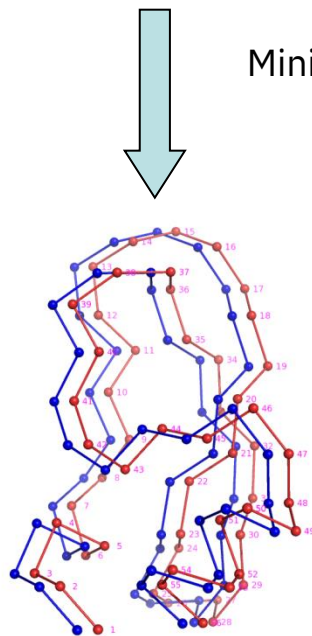


O átomo  $x_i$   
corresponde ao  
átomo  $y_i$

Minimização do RMS

Quadrado da distância entre  
o átomo  $x_i$  e o átomo  $y_i$

A comparação de estruturas  
pressupõe a definição de uma  
correspondência entre os  
átomos das moléculas A e B

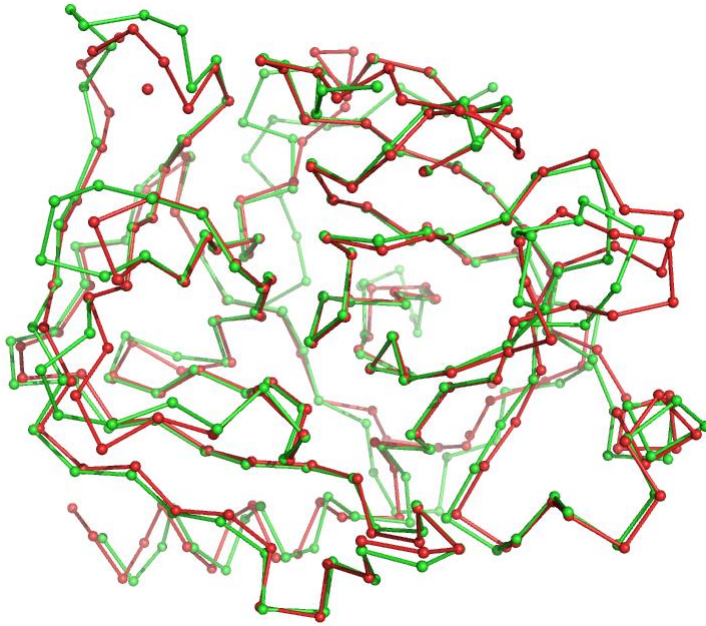


$$RMSD = \sqrt{\frac{\sum_i^N |\vec{x}_i - \vec{y}_i|^2}{N}}$$

**RMSD** - root mean square deviation,  
tem dimensões de comprimento é  
geralmente é dado em Ångstron

# Relação entre RMSD e identidade de sequência

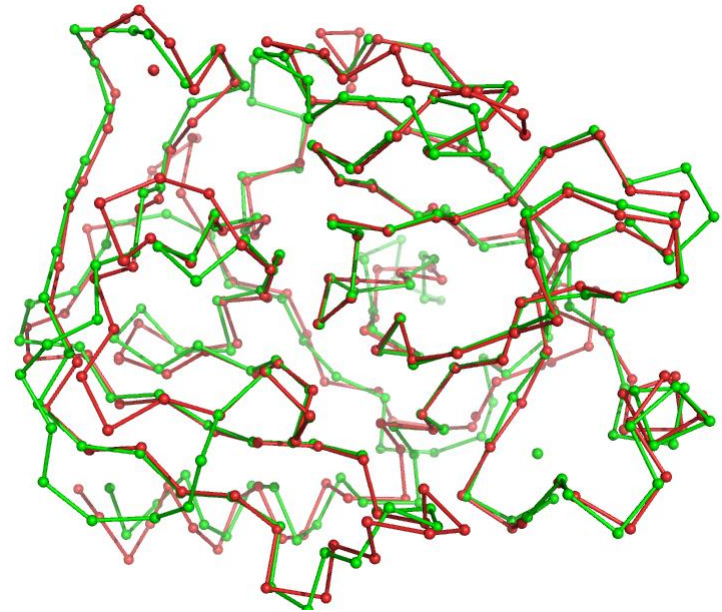
Tripsina humana  
*versus*  
Tripsina bovina



RMSD 0.8 Å

40% identidade de sequência

Tripsina humana  
*versus*  
Tripsina *S.griseus*

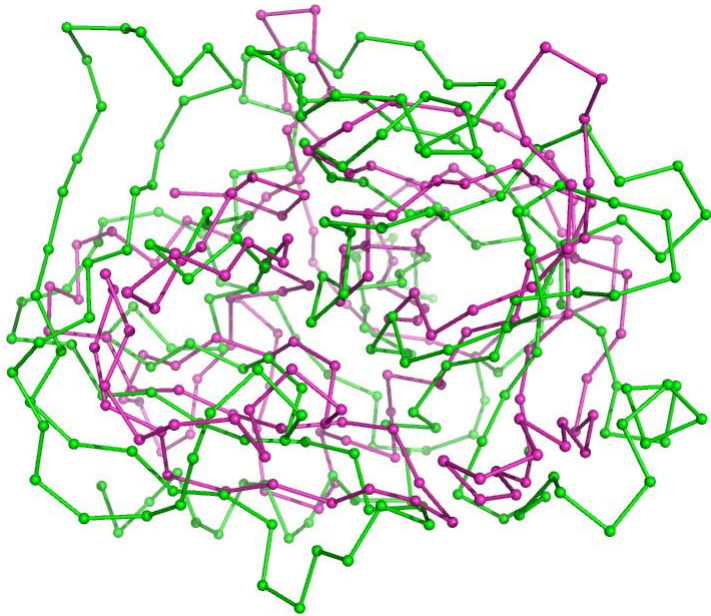


RMSD 1.8 Å

34% identidade de sequência

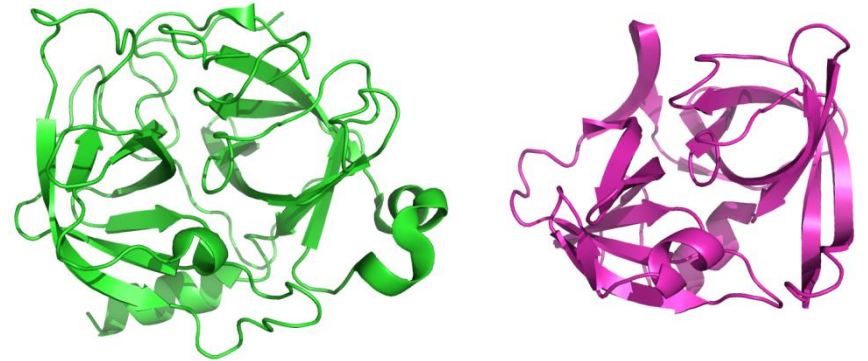
# Relação entre RMSD e identidade de sequência

Tripsina humana  
*versus*  
Proteinase A *S.griseus*



RMSD 5.7 Å

20% identidade de sequência

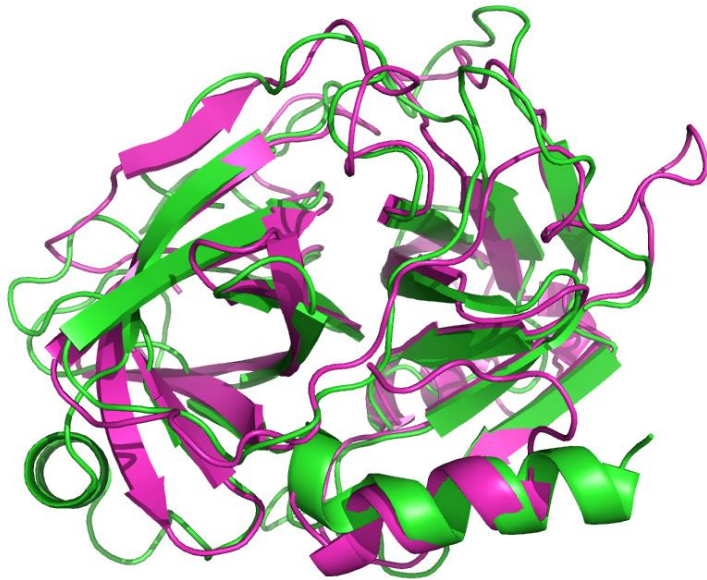


As duas proteínas têm clara  
semelhança estrutural, mas não é  
detectável por comparação de  
sequências



# Relação entre RMSD e identidade de sequência

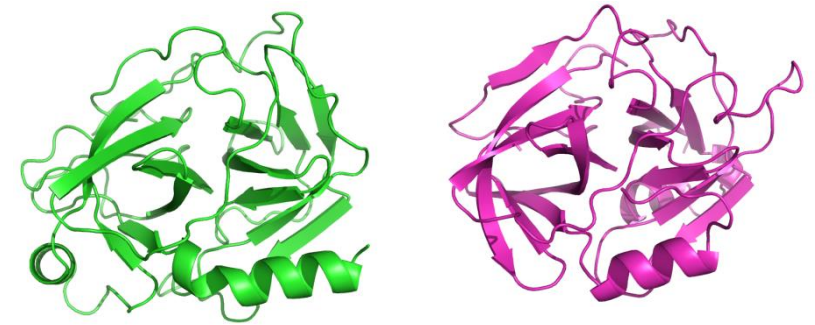
Tripsina humana  
*versus*  
Proteinase V8 *S.aureus*



RMSD 2.5 Å

19% identidade de sequência

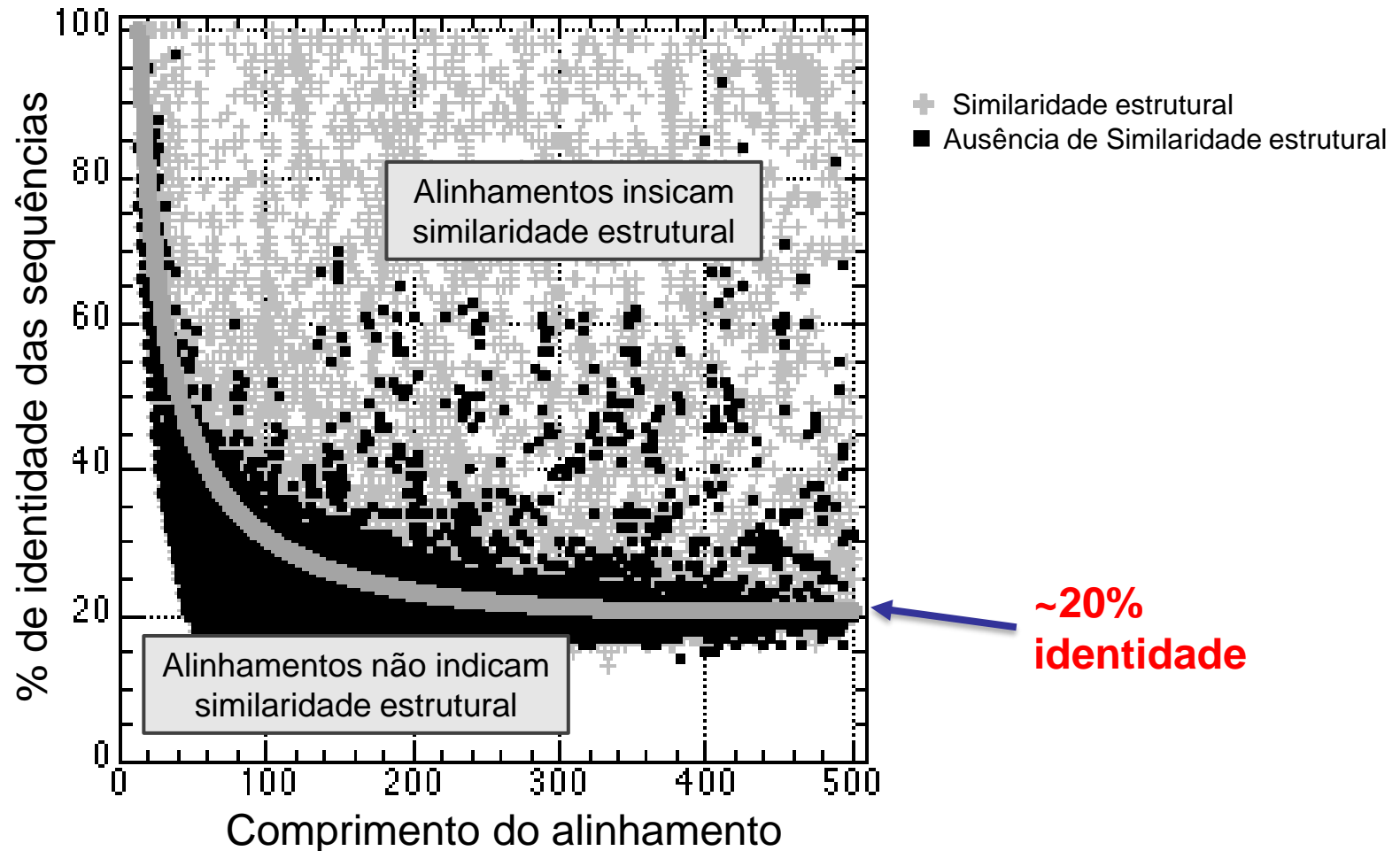
E-value:  $8.6 \times 10^2$



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 seqüência

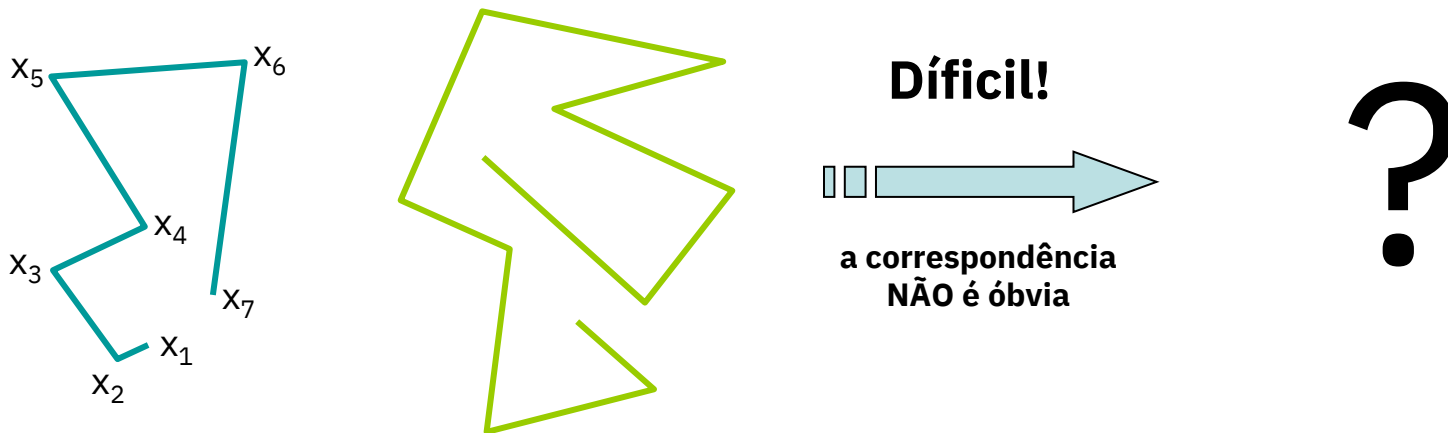
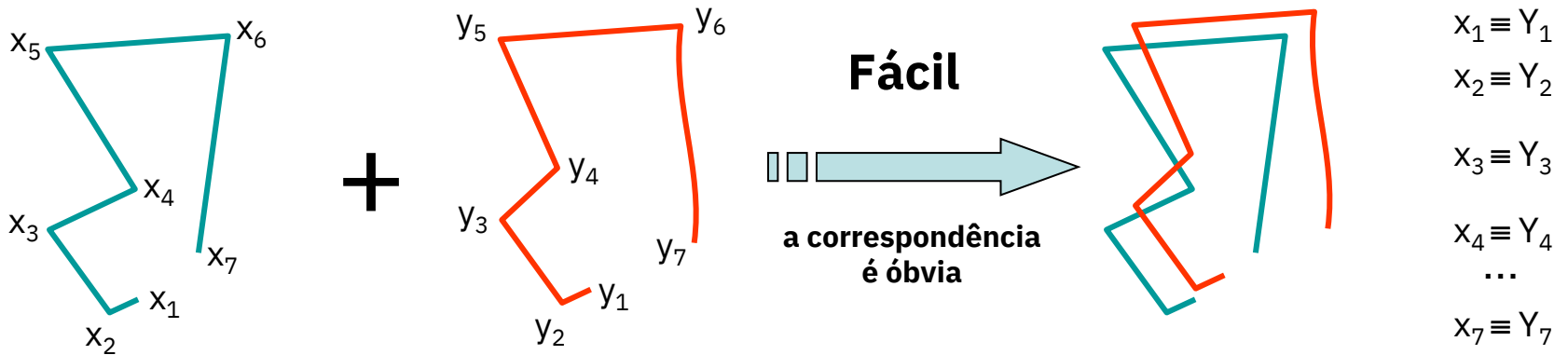


A relação entre a percentagem de identidade e a similaridade estrutural das proteínas depende do comprimento do alinhamento!

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

# 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



# 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

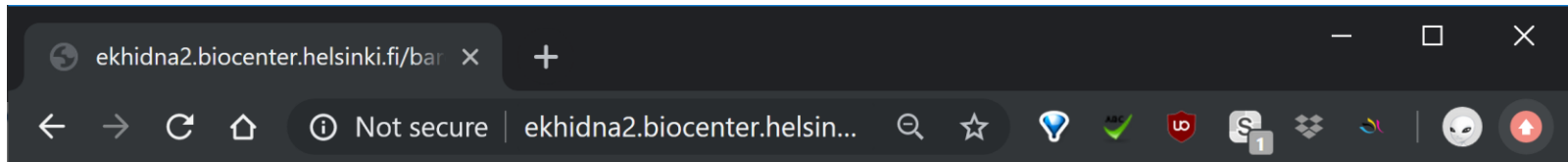
Pretendemos encontrar estruturas semelhantes a uma determinada estrutura, neste caso a uma estrutura do PDB (do enzima lisozima) cujo código é **2LZT**.

The screenshot shows the DALI server web interface in a browser window. The browser's address bar shows the URL `ekhidna2.biocenter.helsinki.fi`. The page title is "DALI PROTEIN STRUCTURE COMPARISON SERVER". The navigation menu includes "About", "PDB search", "PDB25", "Pairwise", "All against all", "Gallery", "References", "Statistics", "Tutorial", and "Download".

The "PDB search" section is active. It contains the following elements:

- STEP 1 - Enter your query:** A text input field containing "2LZT". A red box highlights the input, and a callout box points to it with the text "2LZT ≡ código da lisozima".
- Instructions:** "Structures may be specified by concatenating the PDB identifier (4 characters) and a chain identifier (1 character) or, alternatively, you may upload a PDB file."
- Upload options:** "OR upload file" with a "Choose File" button and "No file chosen" text.
- STEP 2 - Optional data:** A section for providing an email address and job title. The "Job title" field contains "Pesquisa".
- STEP 3 - Submit your job:** A "Submit" button (highlighted with a red box) and a "Clear" button. A callout box points to the "Submit" button with the text "Enviar a pesquisa".

At the bottom, a note states: "If the same structure has been submitted recently, you will be redirected to the result page of the previous instance."



## 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.

Window Snip



## Dali alignment: 2lztA

Each neighbour is shown in the pairwise Dali-alignment to 2lztA. Gaps are expanded, which means that the complete sequence of the matched proteins are shown. (If there are many, ugly or long gaps, you can suppress them by de-checking the 'Expand gaps' option in the summary page.) Uppercase means structurally equivalent positions with 2lztA. Lowercase means insertions relative to 2lztA. The first part shows the amino acid sequences of the selected neighbours. The second part shows the secondary structure assignments by DSSP (H/h: helix, E/e: strand, L/l: coil). The most frequent amino acid type is coloured in each column.

Show Stacked Sequence Logos

```
0001 2lztA  -KVFGRCLEAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCACKKIVSDGNGMNAWAWNRCKGTDVQAWIRGCRLE-----
0002 5vasA  -KVYSRCELAAMKRIGLDNYRGYSLGNWVCAANYESGFNTQATNRNTDGSTDYGILQINSRWWCDGKTPRSKNACGIRCSVLLRSDITEAVRCAKRIVRDGNGMNAWAWNRCKGTDVSKWIRGCRLE-----
0003 1lsgA  mKVFGRCLEAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCACKKIVSDGNGMNAWAWNRCKGTDVQAWIRGCRLEqqhhlggakqagdv
0001 2lztA  -LLLLHHHHHHHHHHLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL
0002 5vasA  -LLLLHHHHHHHHHHLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL
0003 1lsgA  lLLLLHHHHHHHHHHLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL
```

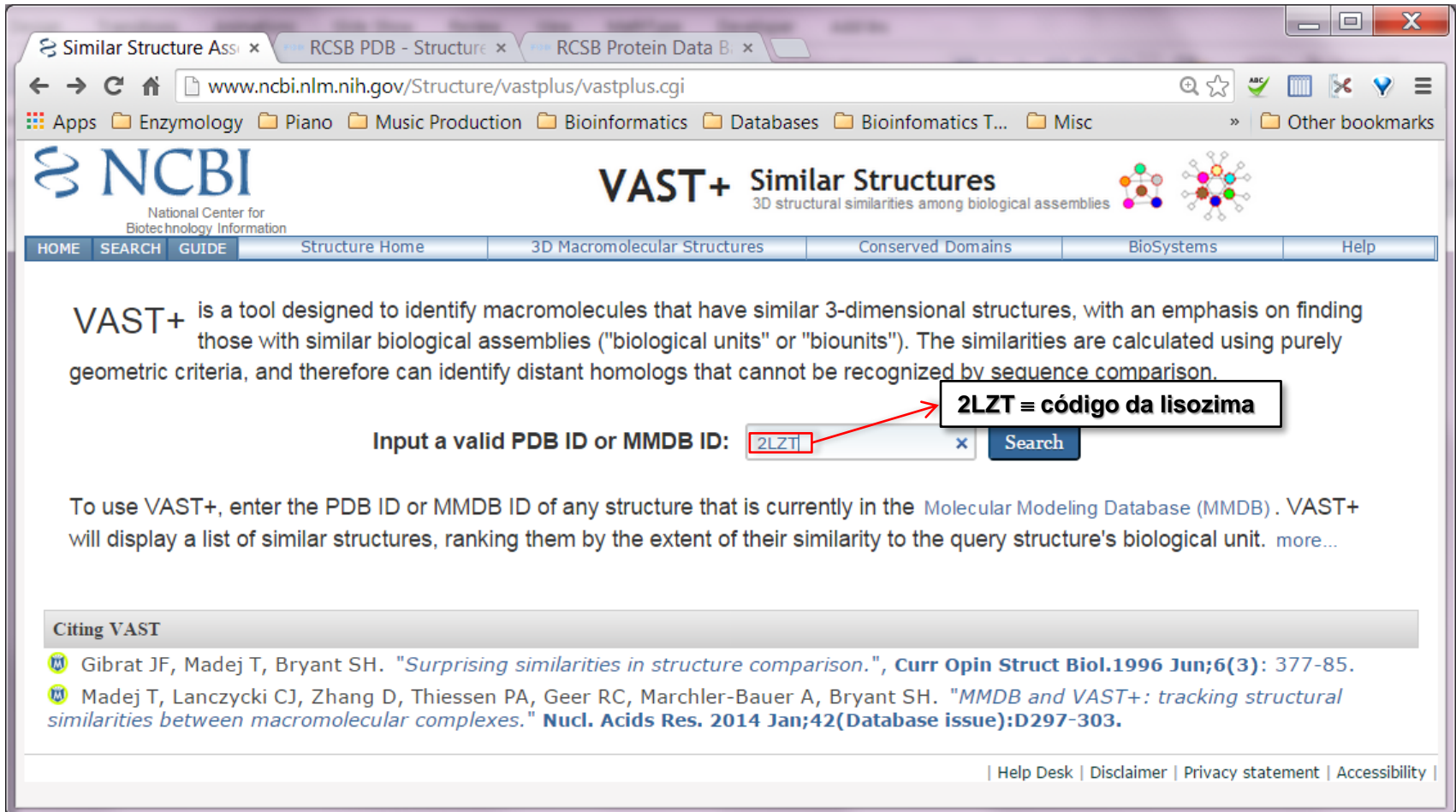


**Alinhamento das sequências baseado na sobreposição das estruturas**



# 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 the VAST+ web interface. The browser address bar displays [www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi](http://www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi). The page header includes the NCBI logo and the title "VAST+ Similar Structures" with the subtitle "3D structural similarities among biological assemblies". A navigation menu contains links for HOME, SEARCH, GUIDE, Structure Home, 3D Macromolecular Structures, Conserved Domains, BioSystems, and Help.

The main content area features a descriptive paragraph: "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."

Below the text is a search input field labeled "Input a valid PDB ID or MMDB ID:" containing the text "2LZT". A red arrow points from this input field to a callout box containing the text "2LZT = código da lisozima". A "Search" button is located to the right of the input field.

At the bottom of the page, there is a "Citing VAST" section with two references:

- 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.

The footer contains links for Help Desk, Disclaimer, Privacy statement, and Accessibility.

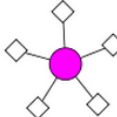
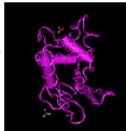
VAST+ Similar Structures

3D structural similarities among biological assemblies

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

PDB ID or MMDB ID

### Refinement Of Triclinic Lysozyme. Ii. The Method Of Stereochemically Restrained 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

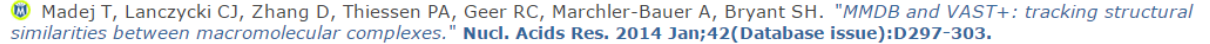
Showing 1 to 10 out of 860 structures

PDB ID	Description	Taxonomy	Aligned Protein	RMSD	Aligned Residues	Sequence Identity
1 <input type="button" value="+"/> ● 1LZN	Neutron Structure Of Hen Egg-White Lysozyme	Gallus gallus	1	0.10Å	120	100%
2 <input type="button" value="+"/> ● 4LZT	Atomic Resolution Refinement Of Triclinic Hew Lysozyme At 295k	Gallus gallus	1	0.10Å	120	100%
3 <input type="button" value="+"/> ● 1V7S	Triclinic Hen Lysozyme Crystallized At 313k From A D2o Solution	Gallus gallus	1	0.11Å	120	100%
4 <input type="button" value="+"/> ● 1LKS	Hen Egg White Lysozyme Nitrate	Gallus gallus	1	0.12Å	120	100%
5 <input type="button" value="+"/> ● 2F2N	Triclinic Hen Egg Lysozyme Cross-linked By Glutaraldehyde	Gallus gallus	1	0.12Å	120	100%
6 <input type="button" value="+"/> ● 2F30	Triclinic Cross-Linked Lysozyme Soaked With 4.5m Urea	Gallus gallus	1	0.17Å	120	100%
7 <input type="button" value="+"/> ● 4MWK	Triclinic Hewl Co-crystallised With Cisplatin, Studied At A Data Collection Temp...	Gallus gallus	1	0.22Å	120	100%
8 <input type="button" value="+"/> ● 2F4G	Triclinic Cross-Linked Lysozyme Soaked In Bromoethanol 1m	Gallus gallus	1	0.24Å	120	100%
9 <input type="button" value="+"/> ● 4MWM	Triclinic Hewl Co-crystallised With Cisplatin, Studied At A Data Collection Temp...	Gallus gallus	1	0.25Å	120	100%
10 <input type="button" value="+"/> ● 2VB1	Hewl At 0.65 Angstrom Resolution	Gallus gallus	1	0.31Å	120	100%

Show 10 structures   Page 1 of 86 Pages

#### Citing VAST





VAST+ Similar Struct x

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

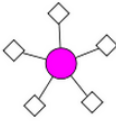
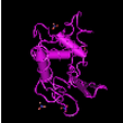
NCBI National Center for Biotechnology Information

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

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

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



Similar Structures

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 Diffraction...	Gallus gallus	1	1.29Å	90	100%
772	2FYD Catalytic Domain Of Bovine Beta 1, 4-Galactosyltransferase In Complex With Alp...	Bos taurus/Mus ...	1	1.30Å	96	39%
773	1NQI Crystal Structure Of Lactose Synthase, A 1:1 Complex Between Beta1,4- Galacto...	Bos taurus/Mus ...	1	1.30Å	97	38%
774	1AM7 Lysozyme From Bacteriophage Lambda	Enterobacteria p...	1	1.30Å	37	22%
775	1FKQ Recombinant Goat Alpha-Lactalbumin T29v	Capra hircus	1	1.30Å	96	45%
776	3B00 Crystal Structure Of Alpha-lactalbumin	Homo sapiens	1	1.30Å	98	40%
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%
779	1NWX Beta-1,4-Galactosyltransferase Complex With Alpha- Lactalbumin And N-Butan...	Bos taurus/Mus ...	1	1.31Å	97	38%
780	1PZY W314a-Beta1,4-Galactosyltransferase-I Complexed With Alpha-Lactalbumin In ...	Bos taurus/Mus ...	1	1.31Å	96	39%

Show 10 structures   Page 78 of 86 Pages

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 3B00 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  
3B00\_A: ALPHA-LACTALBUMIN

```

      10      20      30      40      50      60
2LZT_A  1  KVFGRCELAAAMKrhGLDNYRGYSLGNMCAAKFESNFNTQATNRNTDgSTDYGILQINS 60
3B00_A  1  MQFTKCELSQLLK--DIDGVGGIALPELICTFHTSGYDTQAIVENNE-STEYGLFQISN 57

```

```

      70      80      90      100
2LZT_A  61  RWMcNDGRTPGSRNLCNIPCSALLSSDITASVNCACKIVSD 101
3B00_A  58  KLWCKSSQVPSRNICDISCDKFLDDITDDIMCAKKILDI 98

```

2LZT neighbors - Sequence/Alignment View Edit Mouse Mode Unaligned

2LZT\_A KVFGRCELAAAMKrhG...  
3B00\_A MQFTKCELSQLLK~~~

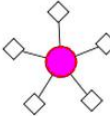
775 [+](#) ● 1FKQ

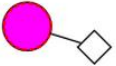
776 [-](#) ● 3B00

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

Matched structure  
MMDB ID: 100429 (PDB ID: 3B00)

\*Click schematic circles and molecule names to view matches

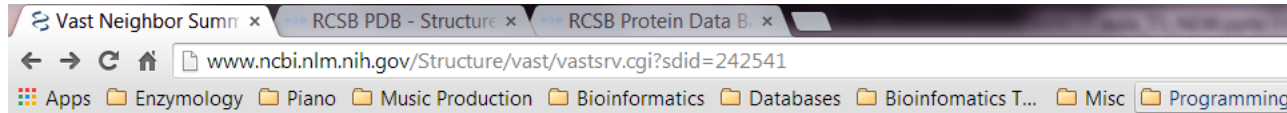
 HEN EGG WHITE LYSOZYME

 ALPHA-LACTALBUMIN

[Visualize 3D structure superposition with Cn3D](#) [View aligned sequences](#)

777 <a href="#">+</a> ● 1NMM	Beta-1,4-Galactosyltransferase Mutant Cys342thr Complex With Alpha-Lactalb...	Bos taurus/Mus ...	1	1.31Å	96	39%
778 <a href="#">+</a> ● 1A2Y	Hen Egg White Lysozyme, D18a Mutant, In Complex With Mouse Monoclonal An...	Gallus gallus/Mus...	1	1.31Å	118	99%

# Pesquisa estrutural com (original)VAST



## 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.

of  with   [Download Cn3D!](#)

using  for  VAST related structures

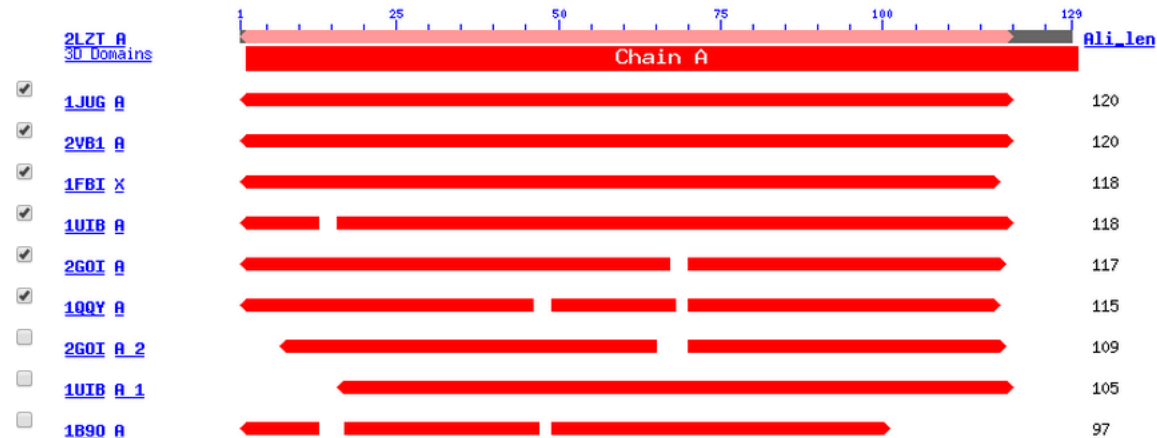
subset, sorted by  in

[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:

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



# Visualização do alinhamento com o software Cn3D

The image shows two overlapping windows. The background window is the NCBI VAST Similar Structures page for MMDDB 58091, 2LZT sequence A. It features navigation tabs for PubMed, BLAST, Structure, and Taxonomy. The 'Structure' tab is active, showing options to 'View 3D Alignment' (set to 'All Atoms' and 'Cn3D') and 'View Sequence Alignment' (set to 'Hypertext' and 'Selected'). Below these are controls for 'List' (Medium redundancy) and 'subset, sorted by' (Aligned Length). An 'Advanced related structure search' link is also present.

The foreground window is titled '2LZT neighbors - Cn3D 4.3'. It displays a 3D molecular model where two protein structures are overlaid. One structure is shown in blue and red, while the other is in grey. The alignment is visualized with red and blue sticks connecting corresponding atoms between the two structures.

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

The '2LZT neighbors - Sequence/Alignment Viewer' window displays a sequence alignment between the 2LZT sequence and several other protein sequences. The alignment is color-coded to show conserved regions. The sequences are:

- 2LZT\_A: ~KVFGRCELAAAMK r hGLDNYRGS L GNWVCAAKFESNFNTQATNRN t ~dGS TDYG I LQ INSRWWCNDG r t PGSRNLCNI PCSALLSSD
- IJUG\_A: ~K I LKKQELCKNLV a qGMNGYQHI TLPNWVCTAFHES SYNTRATNHN t ~dGS TDYG I LQ INSRWCHDGk t PGSKNACNI SC SKLLDDD
- 2VBI\_A: ~KVFGRCELAAAMK r hGLDNYRGS L GNWVCAAKFESNFNTQATNRN t ~dGS TDYG I LQ INSRWWCNDG r t PGSRNLCNI PCSALLSSD
- IUIB\_A: ~KVFGRCELAAAMK ~GLDNYRGS L GNWVCAAKFESNFNTQATNRN t ~dGS TDYG I LQ INSRWWCNDG r t PGSRNLCNI PCSALLSSD
- IFBI\_X: ~KVFGRCELAAAMK r hGLDNYRGS L GNWVCAAKFESNFNSQA TNRN t ~dGS TDYGVLQ INSRWWCNDG r t PGSRNLCNI PCSALQSSD
- 2GOI\_A: maKVF SRCELAKEMhd fGLDGYRGNLADWVCLAYYTSGFNTNAVDHE a ~dGS TNNG I FQ I S SRRWCRTL a ~SNGPNLCRI YCTDLLNND
- IQQY\_A: ~mK I F SKCELARKLK smGMDGFHGYS LANWVCAEYESNFNTQAFNGR ns nGS SDYG I FQLN SKWCK SN s ~HS SANACNI IMC SKFLDDD

At the bottom of the window, there are two red horizontal bars representing alignment footprints. The first bar is labeled 'IUIB H 1' and the second bar is labeled '1890 H'. A scroll bar is visible at the bottom of the alignment viewer.

IUIB H 1

1890 H

# Previsão da estrutura secundária das proteínas

# Níveis de organização da estrutura das proteínas

Estrutura primária

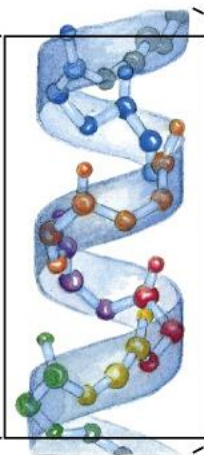
Estrutura secundária

Estrutura terciária

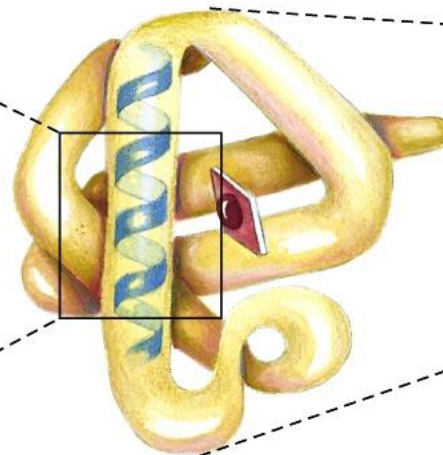
Estrutura quaternária



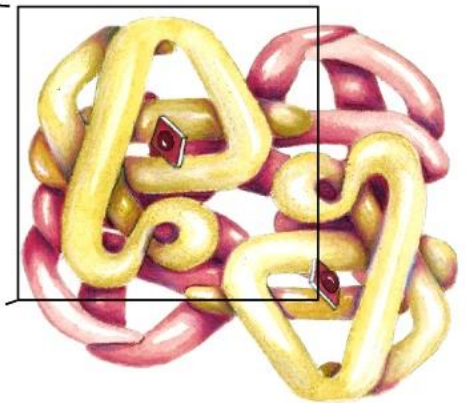
Sequência de aminoácidos



$\alpha$ -hélice



Cadeia polipeptídica



Organização das subunidades

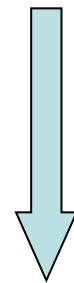


# O problema da previsão da estrutura secundária

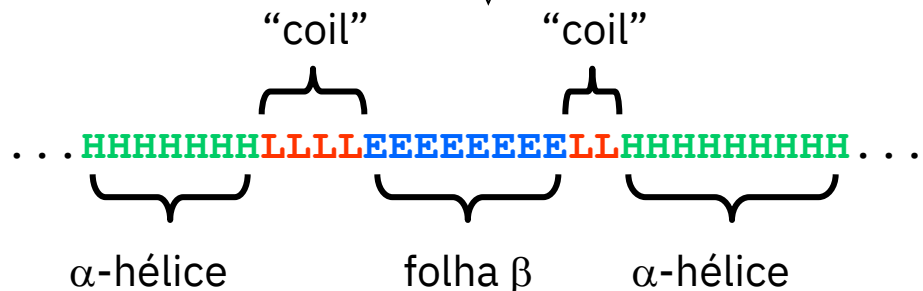
Dada a **sequência** de uma proteína, pretende-se identificar as regiões dessa proteína que adotam diferentes tipos de **estrutura secundária**. Este problema é consideravelmente mais simples que deduzir a estrutura tridimensional completa da proteína (previsão da estrutura terciária). Atualmente conseguem-se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

...AVAGGATILAAGFAVHNQDAGEPAIVLAFG...

Estrutura primária



Previsão



Estrutura secundária

# 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%)

[https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_phd.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_phd.html)  
(**PHD**)

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

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

- **Métodos híbridos:** combinam abordagens distintas, tais como neural networks e nearest-neighbor. Produzem os melhores resultados. (Precisão 85-90%)

[http://210.44.144.20:82/protein\\_PSRS/default.aspx](http://210.44.144.20:82/protein_PSRS/default.aspx) (**PSRSM**)

# Ferramentas on-line para a previsão da estrutura secundária de proteínas e péptidos

- **Jpred4:** <https://www.compbio.dundee.ac.uk/jpred4/>
- **RaptorX:** <http://raptorx.uchicago.edu/StructurePrediction/predict/>
- **Psipred:** <http://bioinf.cs.ucl.ac.uk/psipred/>
- **Hhpred:** <https://toolkit.tuebingen.mpg.de/tools/hhpred>
- **Spider3:** <https://sparks-lab.org/server/spider3/>
- **PRSSM:** [http://210.44.144.20:82/protein\\_PSRSMS/default.aspx](http://210.44.144.20:82/protein_PSRSMS/default.aspx)

# Exemplo de previsão com o programa PHD

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

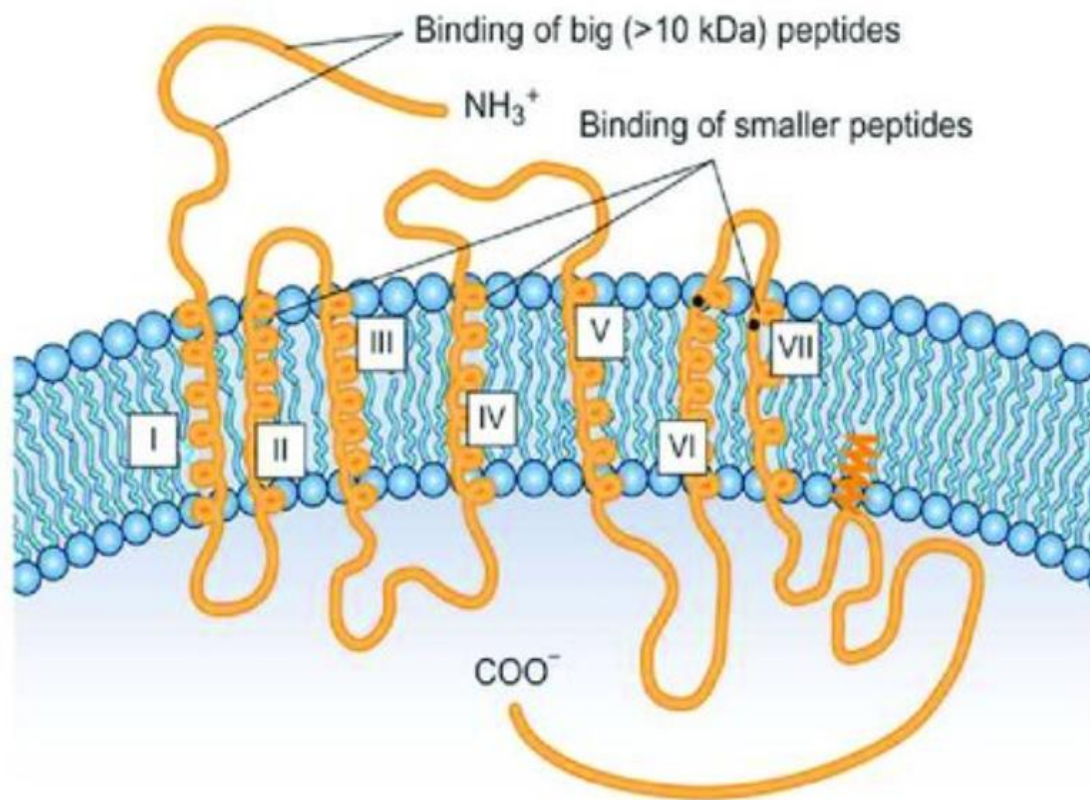
```
      . . . . . 1 . . . . . 2 . . . . . 3 . . . . . 4 . . . . . 5 . . . . . 6
AA   | MERYENLFAQLNDRREGAFVVPFVTLGDPGIEQSLKI IDTLIDAGADALELGVPFSDPLAD |
PHD  |  HHHHHHHHHHHH      EEEEEEE      HHHHHHHHHHHHHH      EEEE      |
Rel  | 934899999996348872799842489984587999999997399668944767784689 |
detail:
prH- | 03689999998753110000000000001678899999998300000000001113210
prE- | 000000000000000000579886530000000000000000000000000178863111000000
prL- | 963100000012368883100123689983211000000001699720036877886789
subset: SUB | L..HHHHHHHHH..LLL.EEEE...LLL.HHHHHHHHHHHH.LLLEE..LLLLL.LLL |

      . . . . . 7 . . . . . 8 . . . . . 9 . . . . . 10 . . . . . 11 . . . . . 1
AA   | GPTIQNANLRAFAAGVTPAQCFEMLALIREKHPTIPIGLLMYANLVFNNGIDAFYARCEQ |
PHD  |  HHHHHHHHHHHH      HHHHHHHHHHHHHH      EEEEE HHHH HHHHHHHHHHH |
Rel  | 7378999999998289629999999999728999389988342244125599999999 |
detail:
prH- | 2588999999984111589999999998410000000000235664326999999998
prE- | 00000000000000000000000000000000000000000000000000038998851100000000000000
prL- | 7411000000001588730000000000001589986100013554224572000000000
subset: SUB | L.HHHHHHHHHHH.LLL.HHHHHHHHHHHH.LLLL.EEEEE.....LHHHHHHHHHH |
```

**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)



# Previsão com o programa PHD

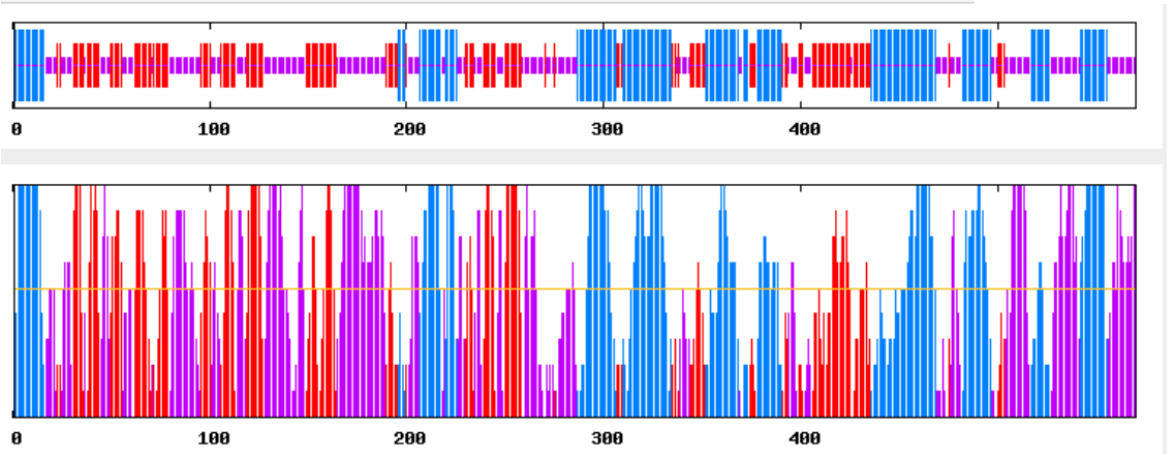
```

      10      20      30      40      50      60      70
      |      |      |      |      |      |      |
MGFLQLLVAVLASEHRVAGAAEVFGNSSEGLIEFVSGKFRYFELNRPFPPEAILHDISSNVTFLIFQIH
CcHHHHHHHHHHHHhhccCCCceecCCCCcEEEEececEEEEeCCcCeEEEEcccCcEEEEEEe
SQYQNTTVSFSPDLLSNSSETGTASGLVILRPEQSTCTWYLGTSIQPVQNMAILLSYSERDPVPGGCN
ecceeeEEEEcCCcCCCCCCCCccccEEEEcCCcceeEEEEeCCcCeEEEEEEecCCCCCCCCC
LEFDLDIDPNIYLEYNFFETTIFKAPANLGYARGVDPCCDAGTDQDSRWRLQYDVYQYFLPENDLTEEM
ccccCCCCeEEeEEeEEEEEecCCCCCCCCCCCCCCCCCCCCcEeEehhhhecCCCCHH
LLKHLQRMVSVQVKASALKVVTLTANDKTSVFSSSLPGQGVIIYNVIVWDPFLNTSAAYIPAHTYACSF
HHHHHHHhCcHHHHhCCcEEeCCcEEEEEcCCcEEEEEEEcCCCCCccccccccccc
AGEGSCASLGRVSSKVFFTLFFALLGFFICFFGHRFWKTELEFFIGFIIMGFFFYILITRLTPIKYDVLIL
cCCCCcchhhHHHHHHHHHHHHhhheeehhhhHHHHHHHHHHHHHHHHhhheecCccceeeEEE
TAVTGSVGMFLVAVWWRFGILSICMLCVGLVGLFISSVTFFTPLGNLKFHDDGVFWVTFSCIAIILIP
eehhHhhHHHHHHHHHHhchhhheeeHHHHHHHHHHhhheecCCcceeccccceeeEeeeeEEEE
VVMGCLRILNILTCLGIVIGSYSVLAIDSYWSTLSYITLNLKRALNKDFHRAFTNVPFQTNDFIILAV
EEEEEEcceeEEEEehhhhhHhhhhhhhhHhhHHHHHHHHHHHHHHHHcccccecCCcchHHHHHH
WGLAVSGITLQIRRRERGRPFPPHPYKWKQERERRVTNILDPSYHIPPLRERLYGRLTQIKGLFQKEQ
HHHHHHhcccccecCCCCCCCCcchhhHHHHhhccCCCCCCCCCcHHHHHHHHHHHHhccCC
PAGERTPLLL
CCCCCCCCC

```

PHD :

Alpha helix	(Hh) :	186 is	32.63%
3 <sub>10</sub> helix	(Gg) :	0 is	0.00%
Pi helix	(Ii) :	0 is	0.00%
Beta bridge	(Bb) :	0 is	0.00%
Extended strand	(Ee) :	164 is	28.77%
Beta turn	(Tt) :	0 is	0.00%
Bend region	(Ss) :	0 is	0.00%
Random coil	(Cc) :	220 is	38.60%
Ambiguous states (?)	:	0 is	0.00%
Other states	:	0 is	0.00%



# Previsão com o programa PSIPRED



Name : sp|Q9NS93|TM7S3\_HUMAN

Copy Link: <http://bioinf.cs.ucl.ac.uk/psip>

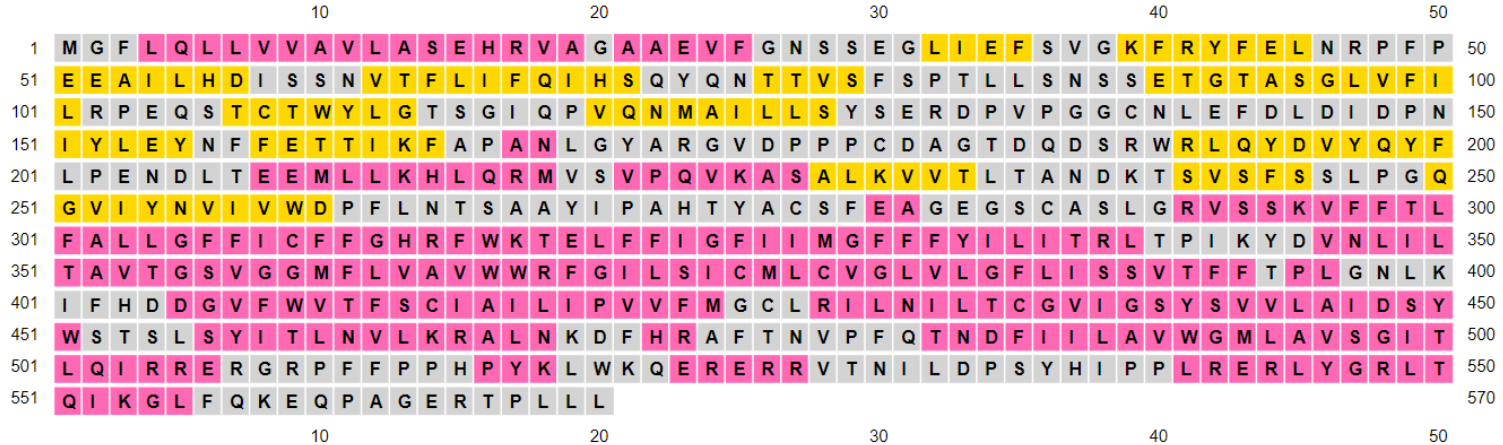


## Sequence Plot

Show psipred

Show memsat

Show aatypes



Strand

Helix

Coil

Disordered

Disordered, protein binding

Putative Domain Boundary

Membrane Interaction

Transmembrane Helix

Extracellular

Re-entrant Helix

Cytoplasmic

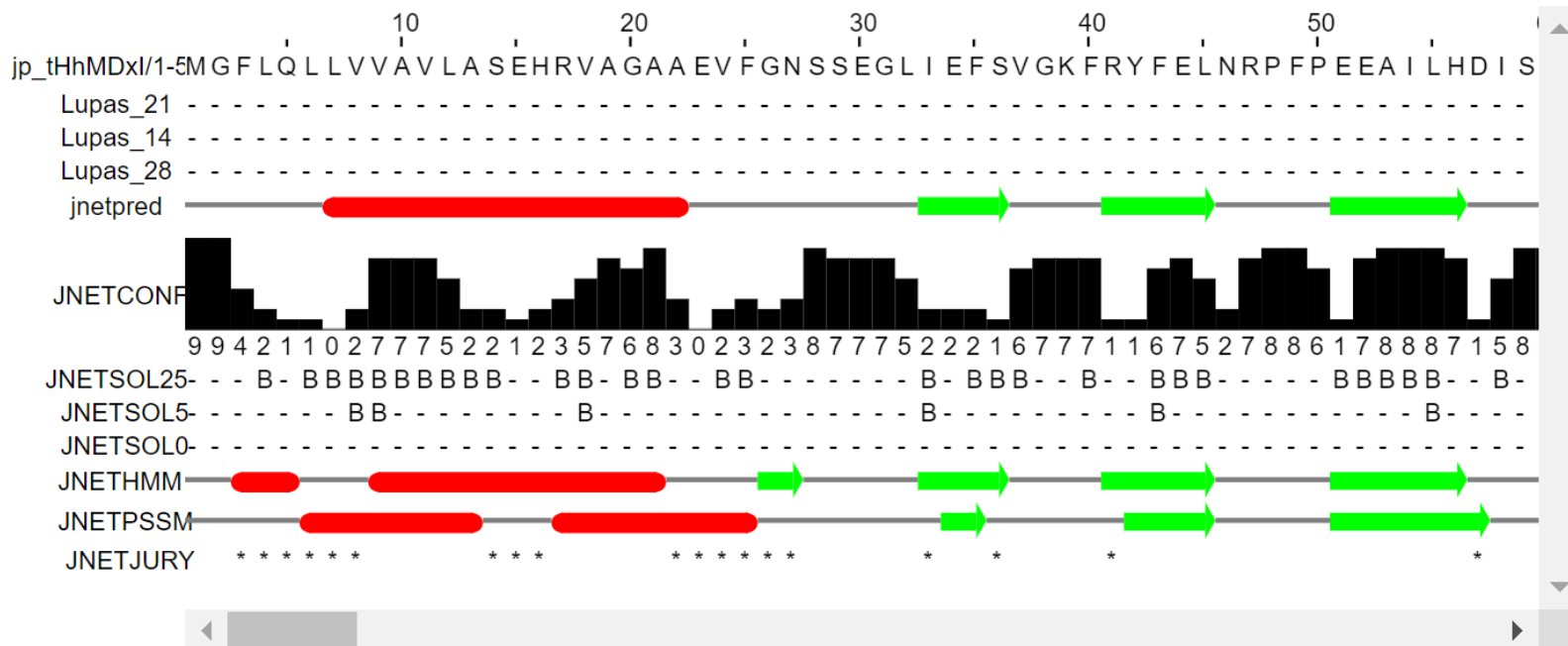
Signal Peptide

Get PNG

Get SVG

# Previsão com o programa Jpred4

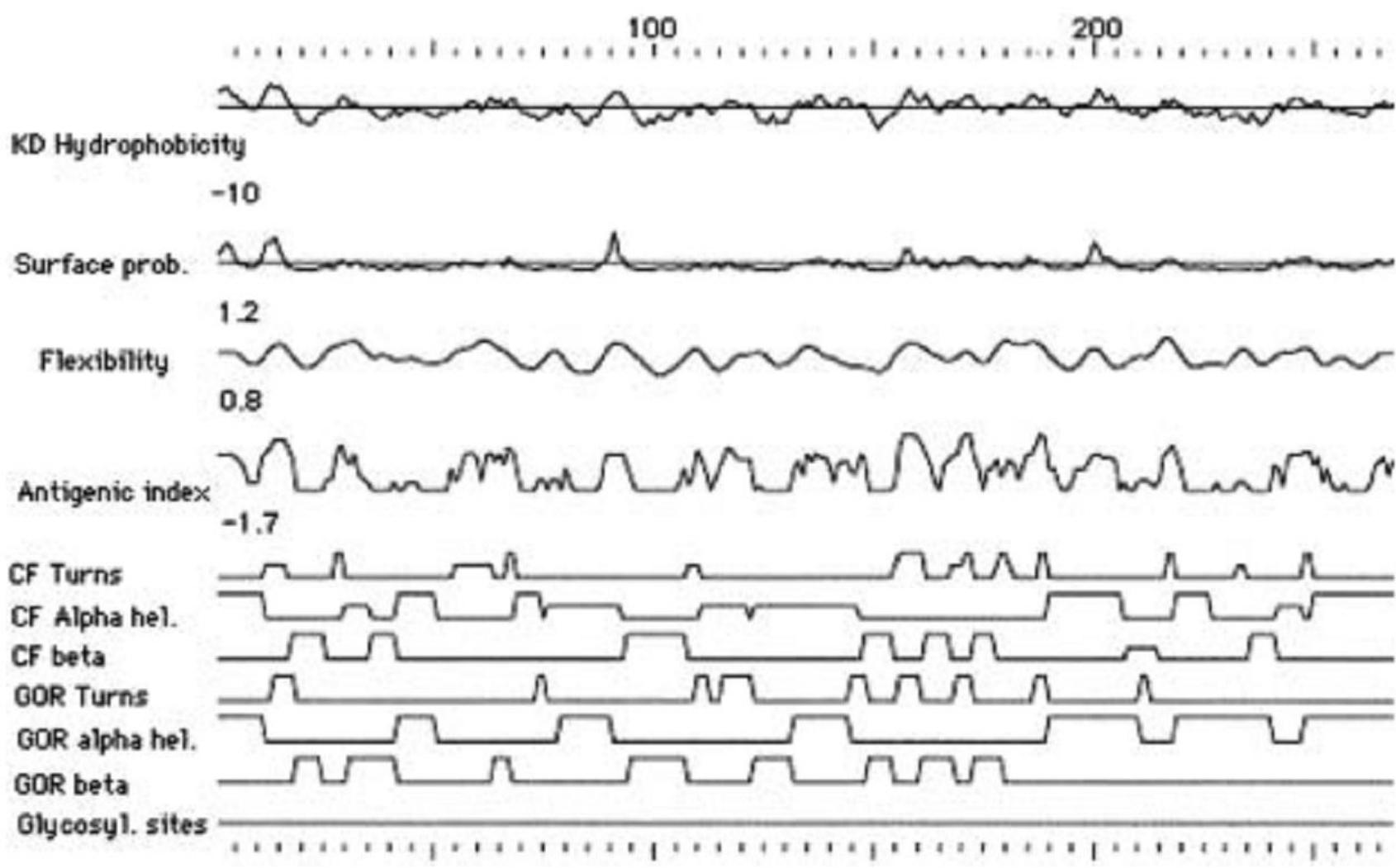
- View results summary in SVG - displayed below (details on acronyms used):



- View full results in HTML



# Previsão GOR e Chou-Fassman com o programa GCG



Modelação da estrutura  
terciária por homologia

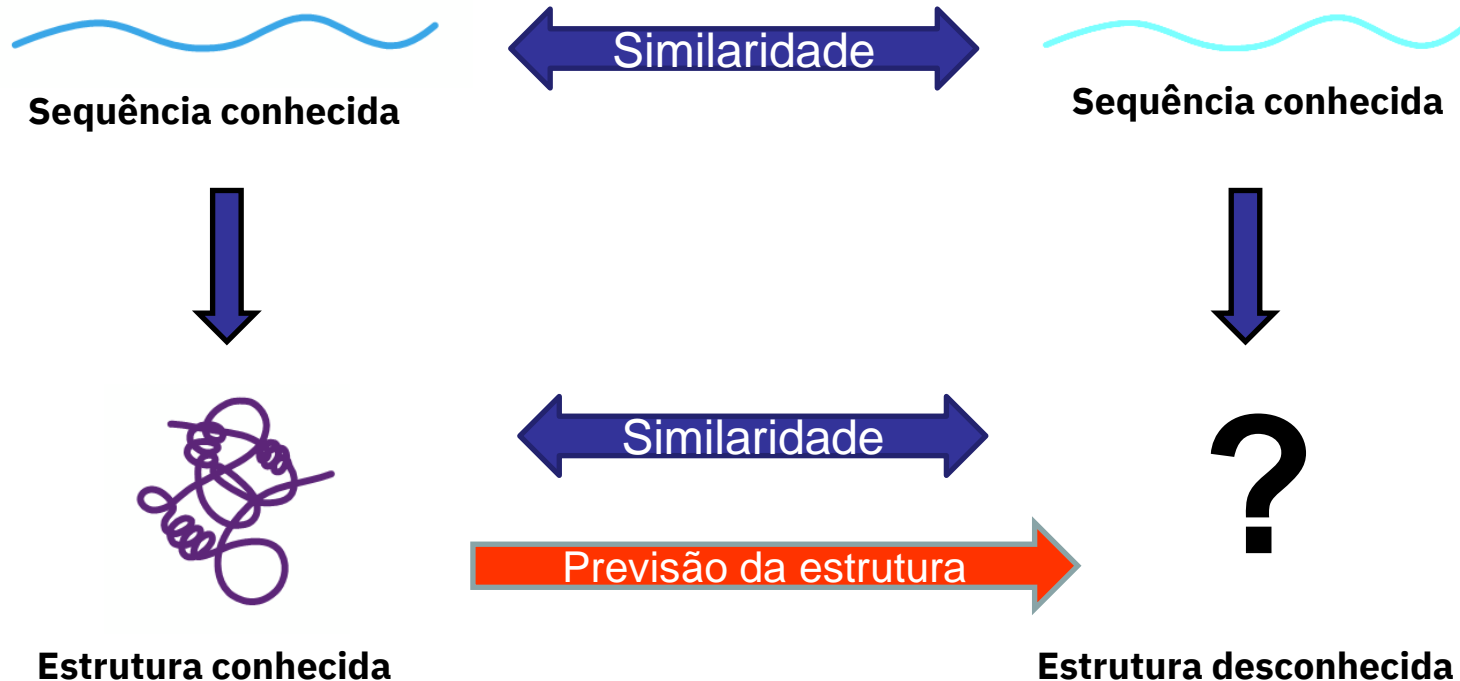
# Modelação por homologia

- A previsão da estrutura tridimensional de uma proteína a partir da sua sequência é extremamente importante, já que o número de sequências conhecidas (~200000000) excede largamente o de estruturas (~200000).
- 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 fiavel 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!**

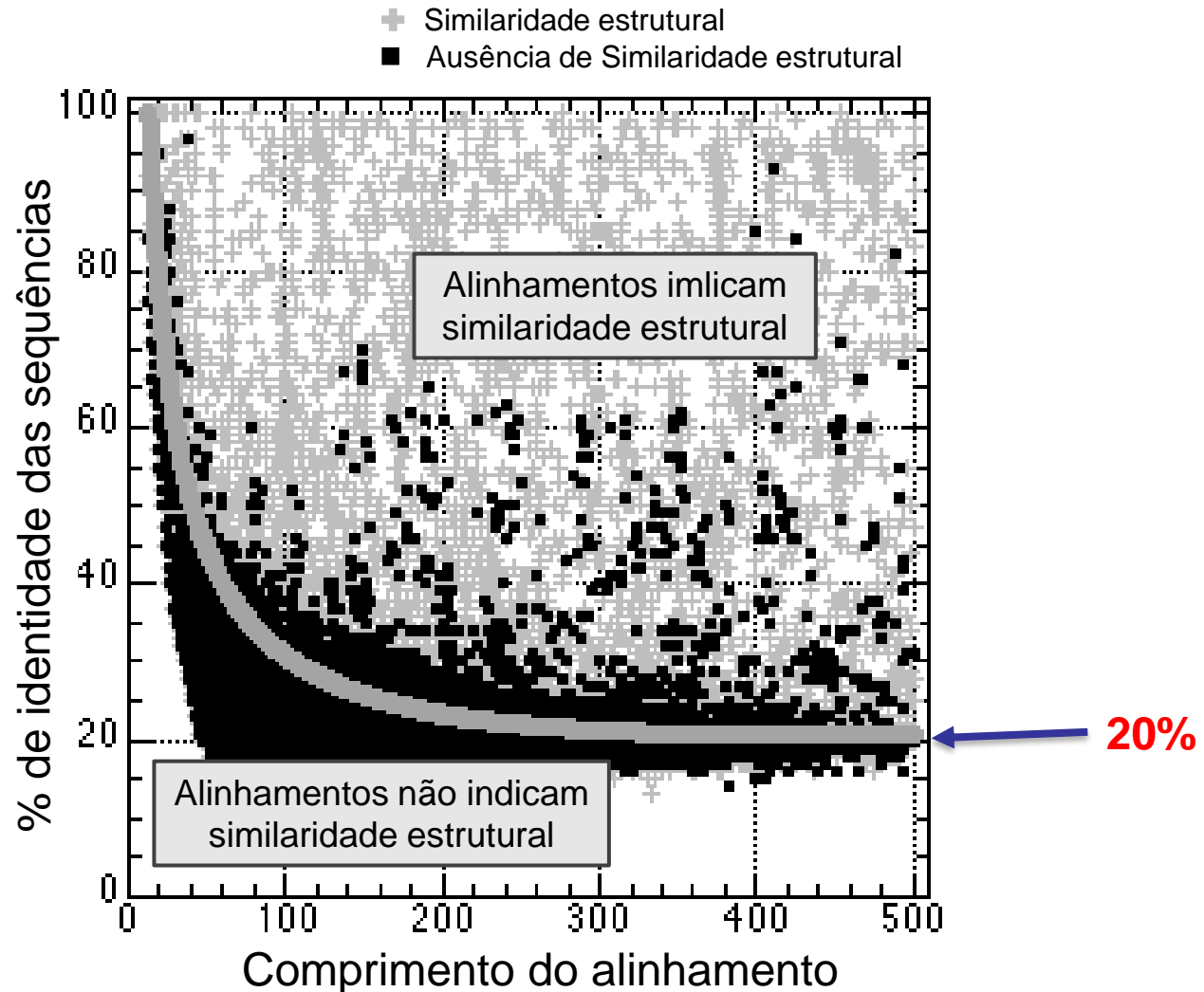
# A estrutura das proteínas é determinada pela sua sequência



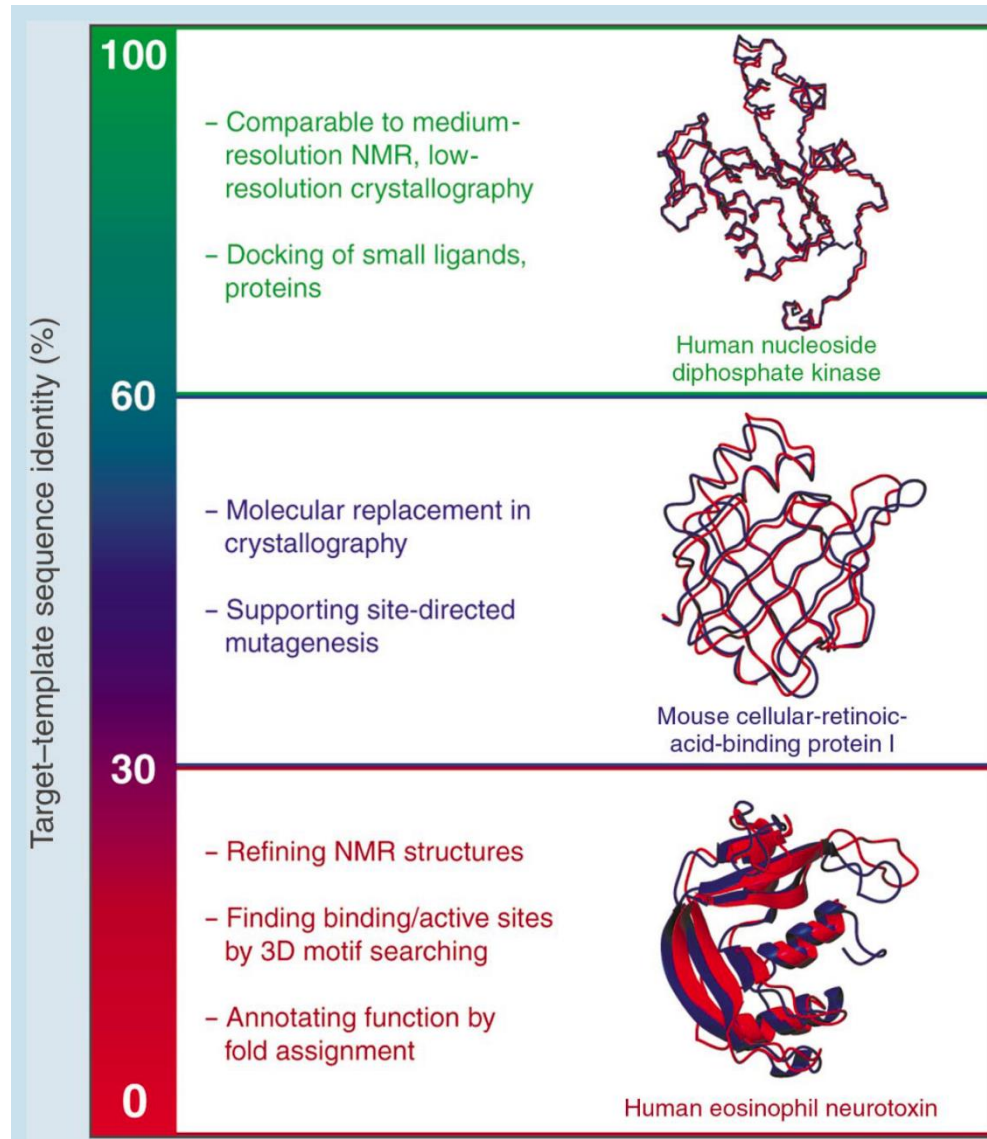
**Sequências similares implicam estruturas similares, logo:**

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

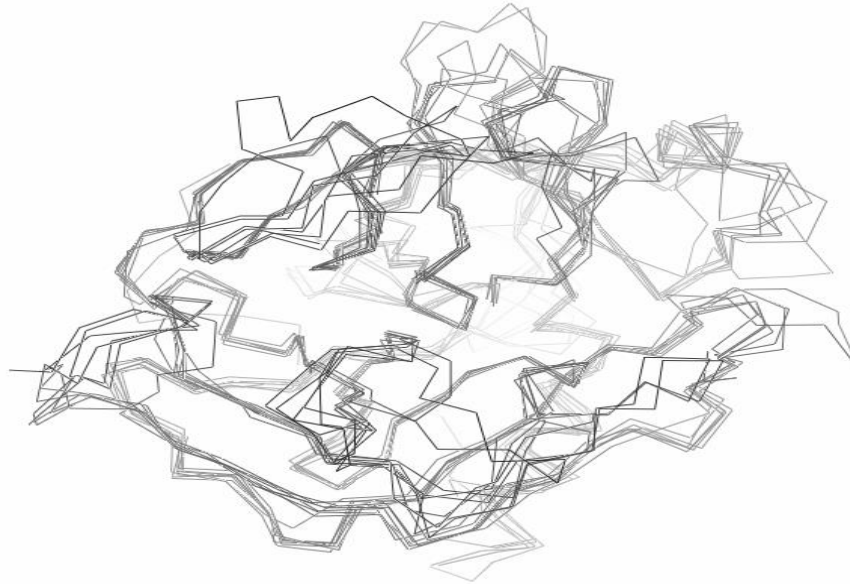
# Qual a % de identidade mínima aceitável para existência de similaridade estrutural ?



# Impacto da similaridade na qualidade e utilidade do modelo



# Alinhamento estrutural das proteases de serina



1BOF:A (218:1-218)  
 3BTH:E (223:1-221)  
 1GMD:A (241:14-239)  
 1H8D:H (260:1-248)  
 1T0C:B (259:1-255)  
 1HYL:A (230:1-227)  
 1TON:\_ (235:11-232)  
 2PKA:B (152:5-149)

IVGGRRAPHAWPFMVS LQL---A-----GGHFCGATLIAPNFVMSAAHCV-----ANVNVR--AVRVVLGAHNLSRREP-TRQV  
 IVGGYTCGANTVRYQVSLNS---G-----YHFCGGSLINSQWVSAAHCY-----KS--G IQVRLGEDN INVVEG-NEQF  
 IVNGEAVPGSWEWQVSLQQL-G-----FIFCGGSLINERWVITAACG-----VTL-SVVVAGEPDQSSSE-KIQK  
 IVRGSDAEIGMSPWQVMLFRkspQ-----ELLCGASLISDRWVLTAAHCLlyppw dkn-----FTe nD LLV RIGKHS RTR YER IERI  
 IVRGSDAEIGMSPWQVMLFRkspQ-----ELLCGASLISDRWVLTAAHCLlyppw dkn-----FTv dD LLV RIGKHS RTR YER IERI  
 IINGYEA Y TGLFFYQAGLDT---Tlqqq---RWVCGGSLIDRWVLTAAHCLlyppw dkn-----AV--SVVVVAGEPDQSSSE-KIQK  
 -----SQPWQAVIN---E-----YLCGGVLDPSWVITAACY-----SN--NYQVLLGRNNLEKDEP-FAQR

1BOF:A (218:1-218)  
 3BTH:E (223:1-221)  
 1GMD:A (241:14-239)  
 1H8D:H (260:1-248)  
 1T0C:B (259:1-255)  
 1HYL:A (230:1-227)  
 1TON:\_ (235:11-232)  
 2PKA:B (152:5-149)

FAVQRIFED-GYD-----PV-NLLNDIVILQLNGSATINAVQVQALPA---QGR-----RLGNVQCLAMGWGL--LGR  
 ISASKSIVHPSYN-----N-TLNNDIMLTKLNSAASLNSRVAASISLPT---S---CASAGTQCLISGWN---TKS  
 LKIAKVFKNsKYN-----SL-TLNNDIPLLLKLSAASFSQTVSAVCLPSas---D-----DFAAGTTCVTTGWGL--TRY  
 SMLKRYIYHPRYN-----WRENLDRIALMLRKKPVAFSDYIH PVC LPO-----Ret aas LLQAGY KGRV TGWGN--LKE  
 SMLKRYIYHPRYN-----WRENLDRIALMLRKKPVAFSDYIH PVC LPO-----Kqt aak LLHAGY KSRV TGWGN--RRE  
 VNSERIIISHMFN-----PD-TYLNDAVALKLI-PHVEYTDNIQPIRLPSgeel---N---NKFENIWA TVSGWGQ--SN-  
 RLVRQSPFhp dYIpllvtn dteqpv-H-DHSNDMLLHLEPADITGGVKVIDLPT-----K-----EPKVGSTCLASGWGS t nPSE  
 -----DYSHDLMLLR LQS PAR ITD AVK VLE LPT-----Q-----EPKVGSTCLASGWGS t nPSE

1BOF:A (218:1-218)  
 3BTH:E (223:1-221)  
 1GMD:A (241:14-239)  
 1H8D:H (260:1-248)  
 1T0C:B (259:1-255)  
 1HYL:A (230:1-227)  
 1TON:\_ (235:11-232)  
 2PKA:B (152:5-149)

---NR-----GIASVLQELNVTVVIT--SLC-----RHSNVC TLVRG---R--QAGVC--FGDSGGPLVC--N-----GLIHG IA  
 s--GT-----SYPDVLRKGLKAPILSdsSCSsyp p q l--TSNMFCAGYLE---G--GKDC--QDSGGPFVVC--S-----GKLQAG IV  
 a--N-----TPDRLQASLPLLSa tNCNKv q t k l--KDMICAGAS-----GVSSC--MGDSGGPLVC--Kkn ga--WTLV G IV  
 t-----GQPSVLQVNLPIV Rr p VC k d s t r i r i--TDNMFCAGY K P d e g K--RGDAC--EGDSGGPFVVM--Ksp f n n i W Y Q M G IV  
 ---WTLv a a e V Q P S V L Q V V N L P L V R r p V C k a s t r i r i--TDNMFCAGY K P d e g K--RGDAC--EGDSGGPFVVM--Ksp y n n i W Y Q M G IV  
 ---G-----DTVILQYTWLVYIDa r C k a g y p p q v s t i c g d t s--D--GKDC--FGDSGGPFVVC--N-----NLIIGVY  
 m--VV-----SHDLQ--CVN IHLLSa r C t e t y k d n--TDVMLCAGE---MeggKDT Cag--DSGGPLIC--D-----GVLQGI IT  
 q p d R F-----EFPDIEQCVQLTL L Q n F C A d a h p k v--TESMLCAGYLP---G--GRDTC--MGDSGGPLIC--N-----GMWQGIT

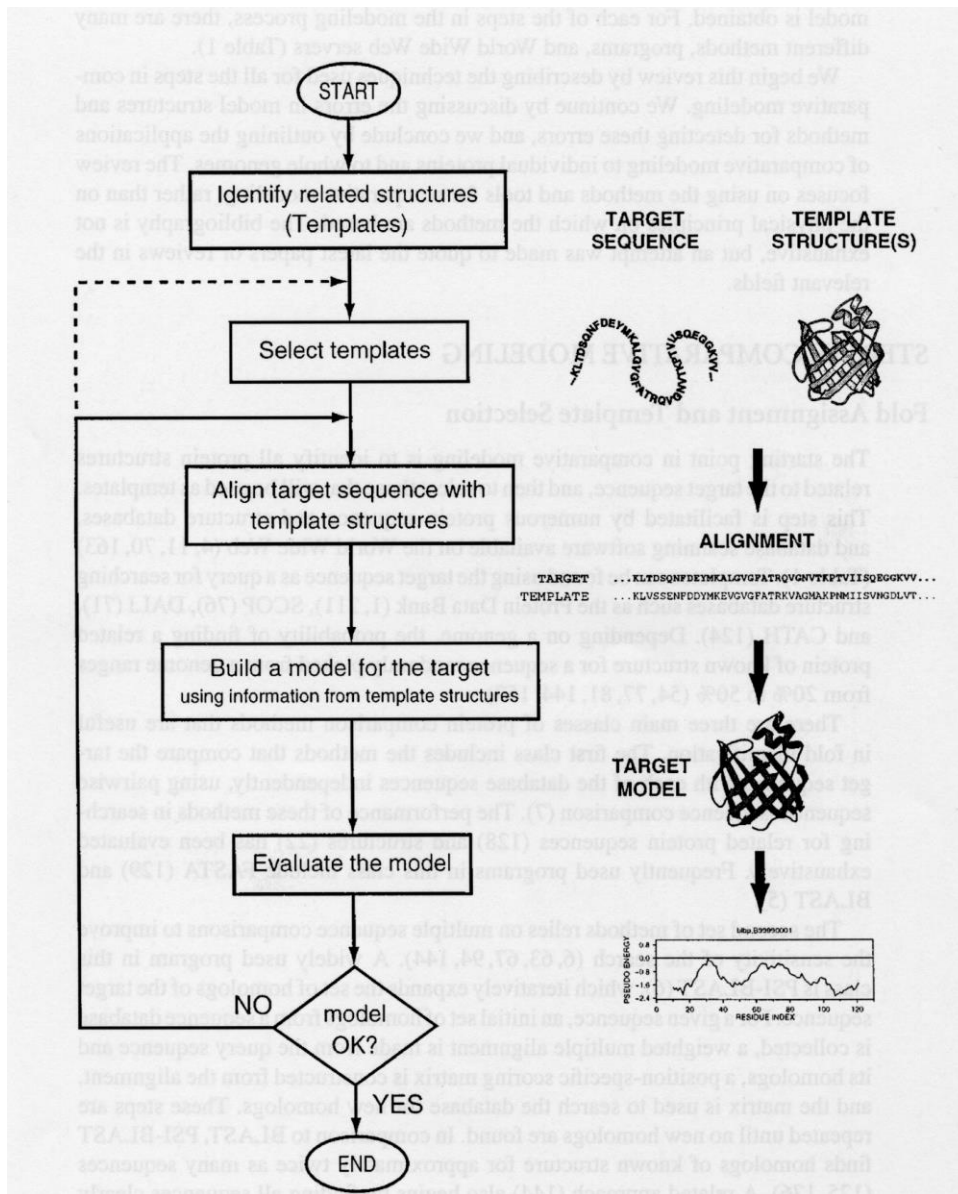
1BOF:A (218:1-218)  
 3BTH:E (223:1-221)  
 1GMD:A (241:14-239)  
 1H8D:H (260:1-248)  
 1T0C:B (259:1-255)  
 1HYL:A (230:1-227)  
 1TON:\_ (235:11-232)  
 2PKA:B (152:5-149)

SFVR--GG-CASGLYDAPFAVAQFVNNIDSI IQ  
 SWGS---G-CAQKNKPGVYTKVCNVYSWI RQT IA  
 SWGS---S-TCSTSTPGVYARVTLVNVNQQTLA  
 SWGE---G-CDRDGRYGFYTHVFLRKRKI QNVI D  
 SWGE---G-CDRDGRYGFYTHVFLRKRKI QNVI D  
 SFVSGa-G-CESG-KPVGFSRVTSYMDVIQQNIG  
 SGGa---TpCARPKTPA IYAKLIFTSWI KVMYK  
 SWGH---TpCGSANKPS IYKLIFFYLWDIDDTT

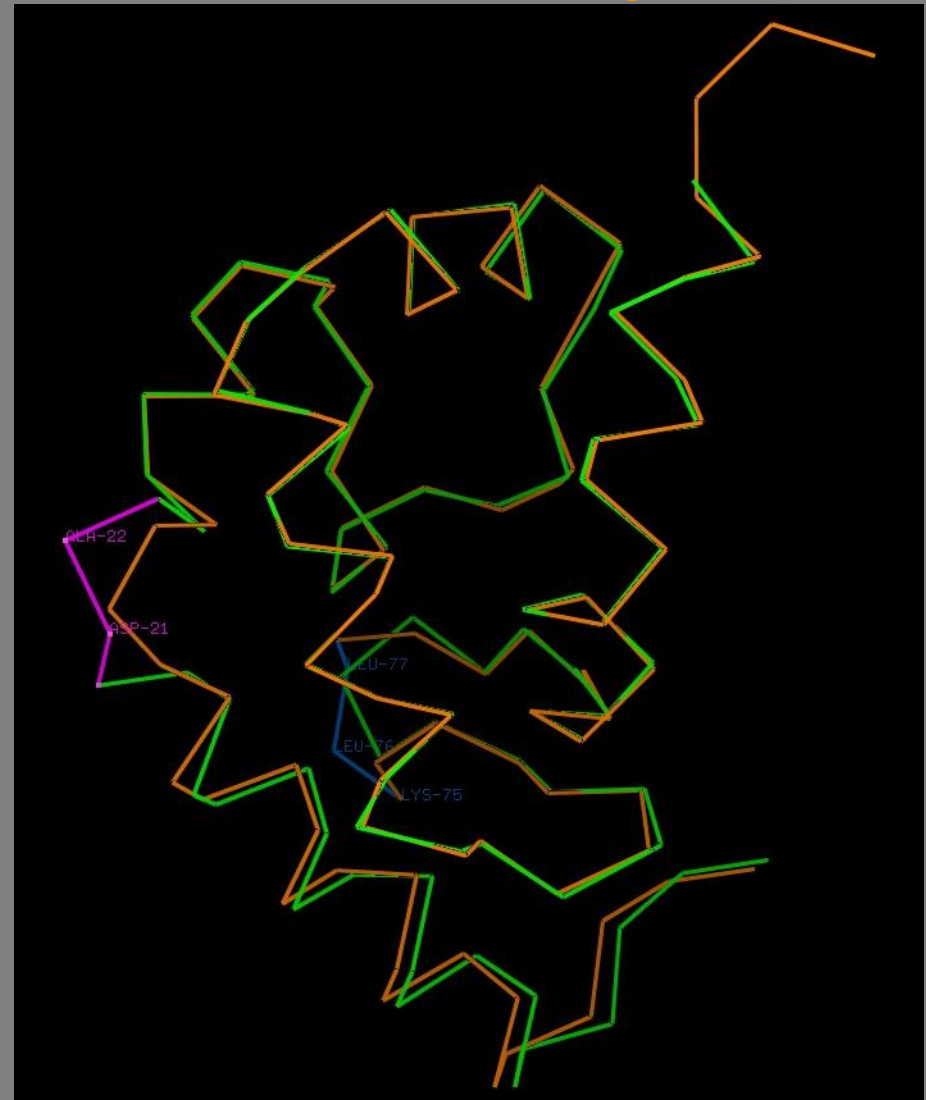
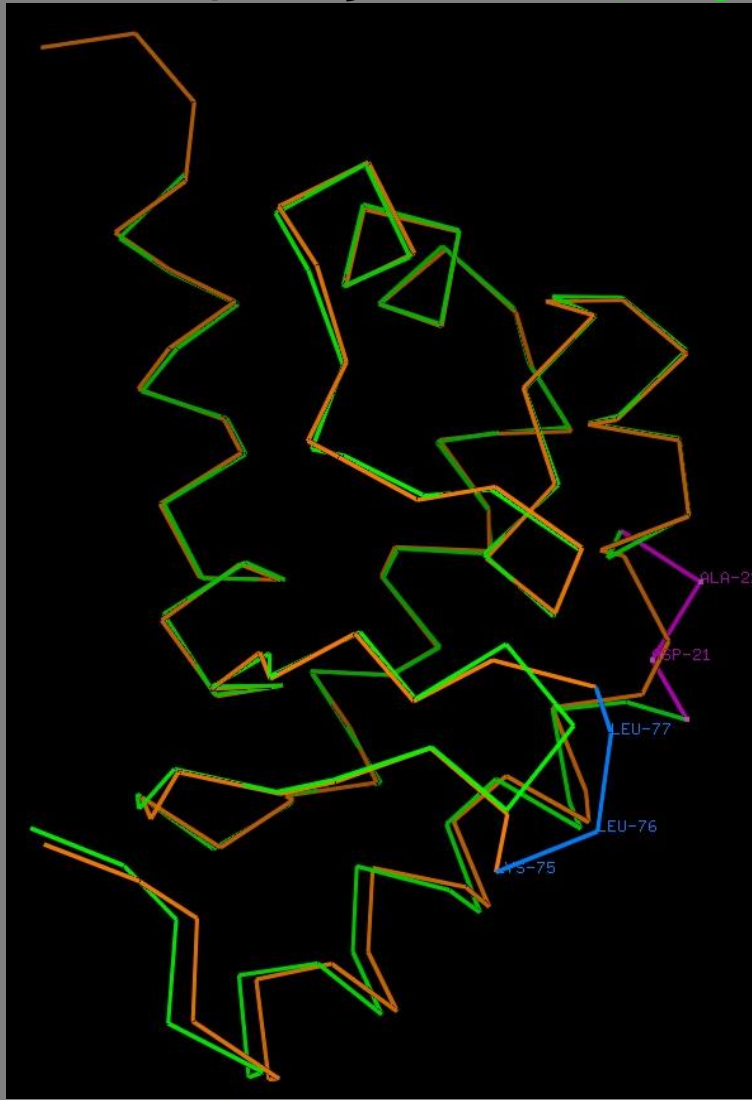
# Passos na modelação por homologia

- Alinhamento estrutural das proteínas de estrutura conhecida homólogas da proteína que se pretende modelar. Inspeção 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**



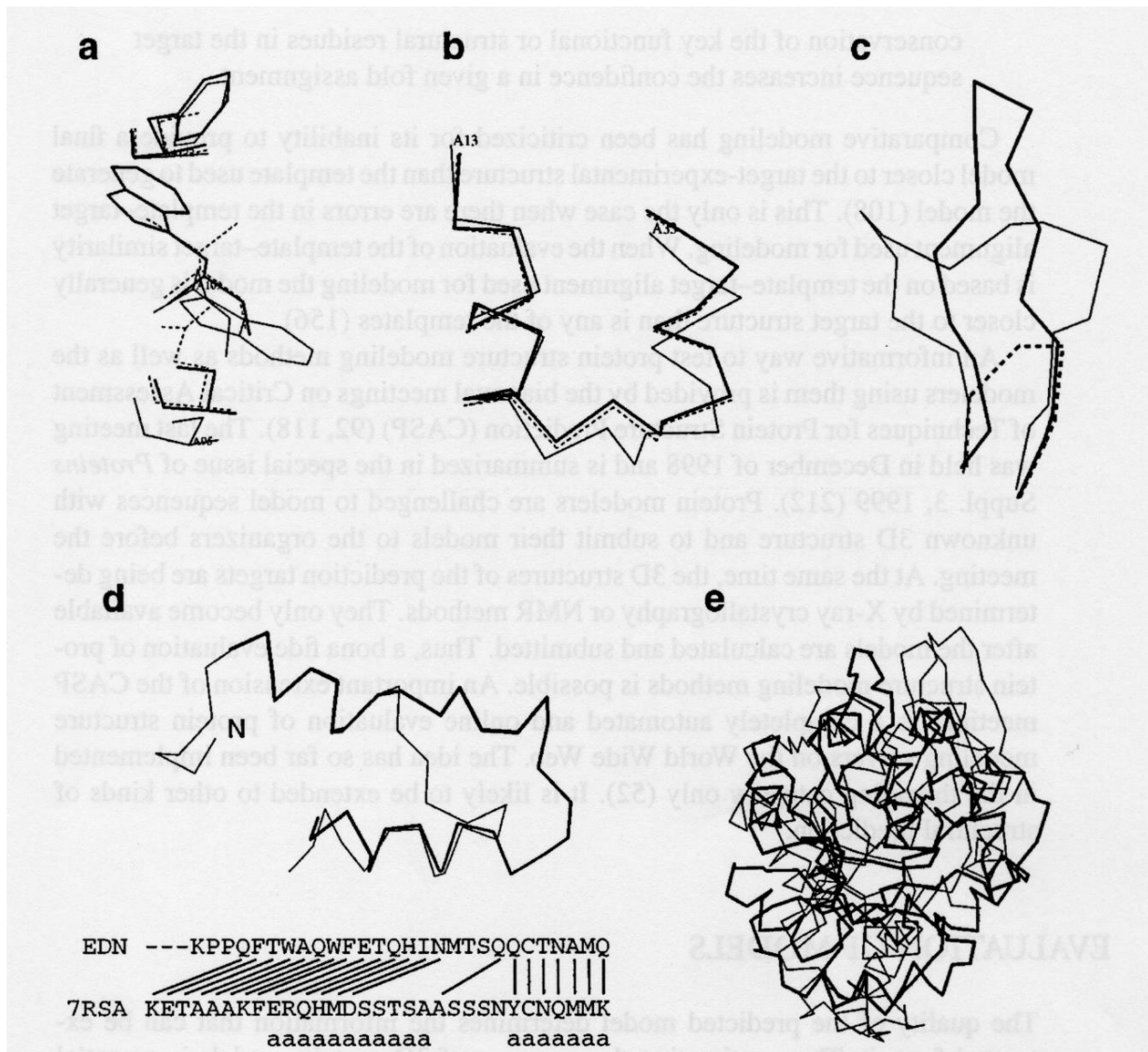
```

::** . * : : . * . . : . * . * : : ** : ** : . . . . . ** . ** * * * * * : : ** : * * . . . . . : ** * * * * . * : :
TACTASQQTAAYKTLVLSILSDA SFNQCSTDSGYSMLTAKALPTTAQYKLMCASTACNTMIKKIVTLNPPNCDLTVPT---SGLVLNVYSYANGFSNKCSSL---
ATCTDEQFSDSIKLTPAIG--SVVGCTADSGFSMIPPTGLPTNDQYKKMCKSTNCKTLIKEIKDANVADCELDKSKLLP GSVPLNVYVLANNFDFVCAVVGSA
    
```

# 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)



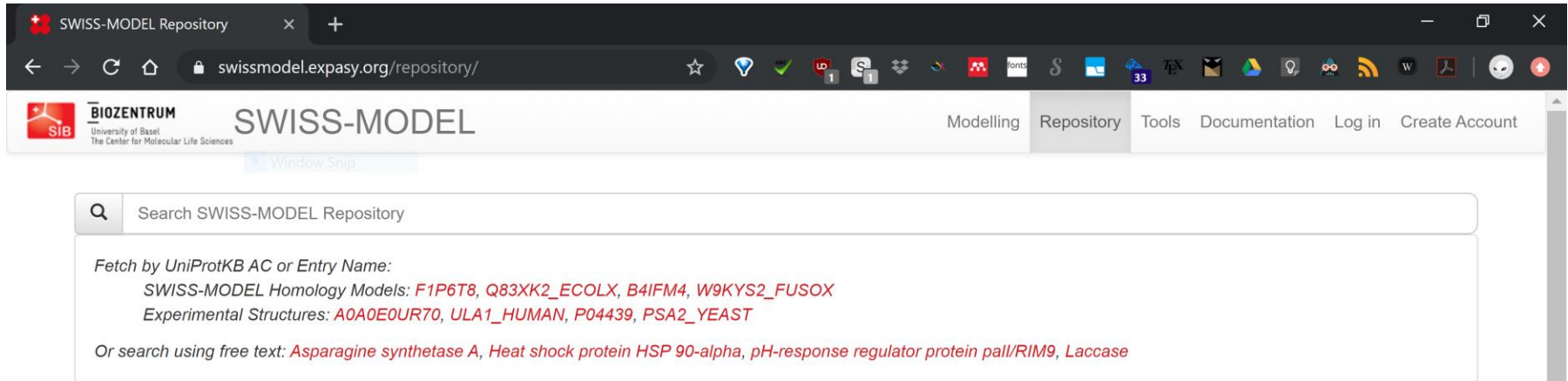
# Servidores web para modelação por homologia

- **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://raptorx.uchicago.edu/>
- **Hhpred** - <http://toolkit.lmb.uni-muenchen.de/hhpred>
- **Robetta** - <https://robetta.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 considerar 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 the SWISS-MODEL Repository website. The browser address bar displays [swissmodel.expasy.org/repository/](https://swissmodel.expasy.org/repository/). The page header includes the BIOCENTRUM logo (University of Basel, The Center for Molecular Life Sciences) and the SWISS-MODEL logo. Navigation links for Modelling, Repository, Tools, Documentation, Log in, and Create Account are visible. A search bar is present with the text "Search SWISS-MODEL Repository". Below the search bar, there are suggestions for fetching data by UniProtKB AC or Entry Name, listing Homology Models (F1P6T8, Q83XK2\_ECOLX, B4IFM4, W9KYS2\_FUSOX) and Experimental Structures (AOA0E0UR70, ULA1\_HUMAN, P04439, PSA2\_YEAST). A note suggests searching using free text: Asparagine synthetase A, Heat shock protein HSP 90-alpha, pH-response regulator protein pall/RIM9, Laccase.

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](https://doi.org/10.1093/nar/nkx104)

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



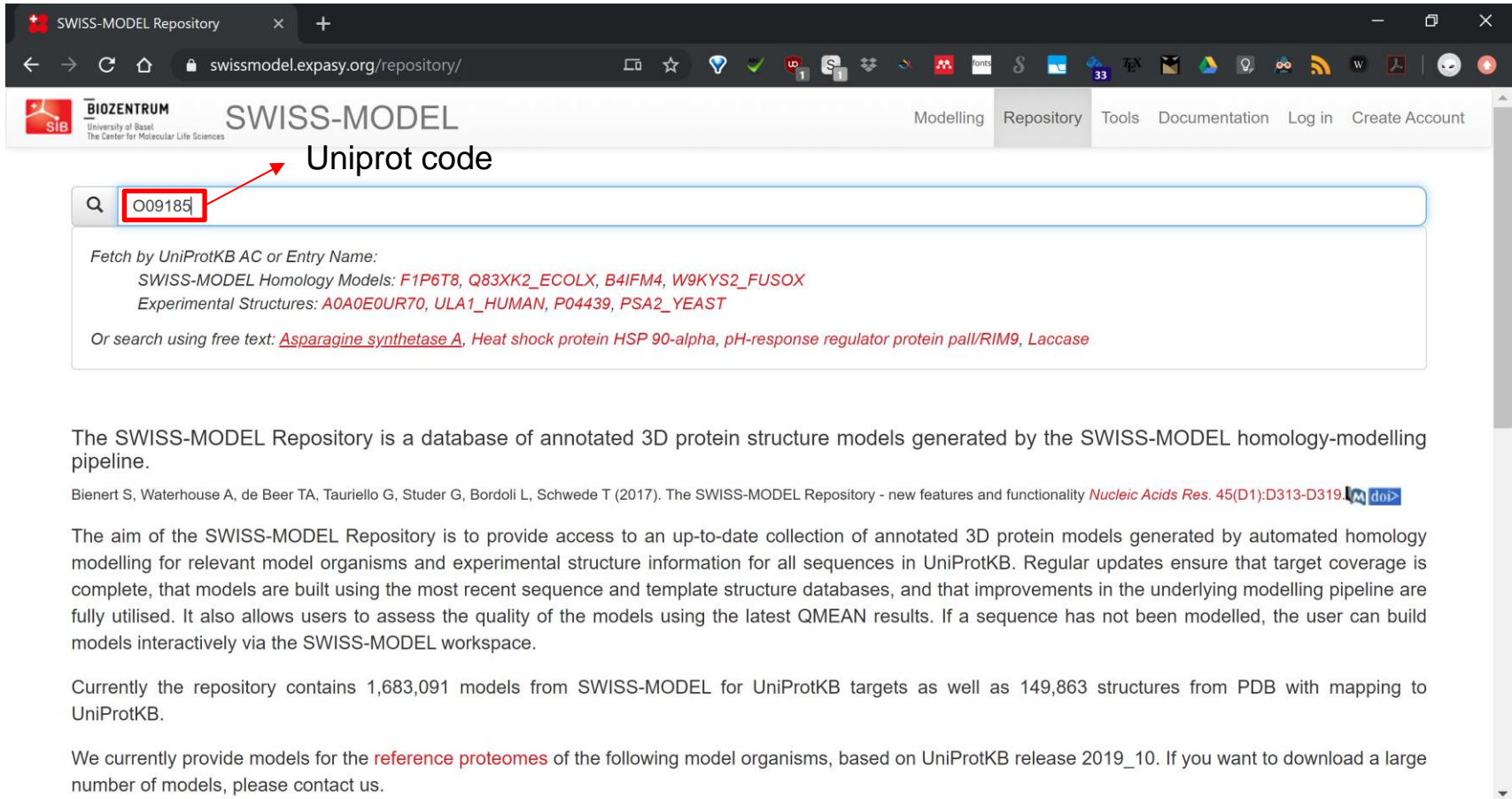
	Proteome Size	Sequences Modelled	Models	Seq Coverage	Download Metadata (Models and structures)	Download Coordinates (Homology models)
<i>Homo sapiens</i>	20,659	17,505	43,255		↓ 13.7 MB	↓ 4.5 GB
<i>Mus musculus</i>	21,960	18,708	43,088		↓ 8.0 MB	↓ 3.0 GB
<i>Caenorhabditis elegans</i>	19,944	12,942	23,474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

Latest snapshot of SMR was taken 1 month ago.

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



# SWISS MODEL repository



The screenshot shows the SWISS-MODEL Repository website. The browser address bar displays [swissmodel.expasy.org/repository/](https://swissmodel.expasy.org/repository/). The website header includes the logo for BIOZENTRUM SIB (University of Basel, The Center for Molecular Life Sciences) and the text "SWISS-MODEL". Navigation links for "Modelling", "Repository", "Tools", "Documentation", "Log in", and "Create Account" are visible. A search bar contains the UniProt code "O09185", which is highlighted with a red box and an arrow pointing to the text "Uniprot code" above it. Below the search bar, the results are displayed:

Fetch by UniProtKB AC or Entry Name:  
SWISS-MODEL Homology Models: [F1P6T8](#), [Q83XK2\\_ECOLX](#), [B4IFM4](#), [W9KYS2\\_FUSOX](#)  
Experimental Structures: [A0A0E0UR70](#), [ULA1\\_HUMAN](#), [P04439](#), [PSA2\\_YEAST](#)

Or search using free text: [Asparagine synthetase A](#), [Heat shock protein HSP 90-alpha](#), [pH-response regulator protein pall/RIM9](#), [Laccase](#)

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 | O09185

swissmodel.expasy.org/repository/uniprot/O09185

BIOZENTRUM University of Basel The Center for Molecular Life Sciences

SWISS-MODEL

Modelling Repository Tools Documentation Log in Create Account

Search SWISS-MODEL Repository

O09185 (P53\_CRIGR) *Cricetulus griseus* (Chinese hamster) (*Cricetulus barabensis griseus*)  
Cellular tumor antigen p53 ★ UniProtKB InterPro STRING Interactive Modelling

393 aa; Sequence (Fasta)

**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
CB	0.09
All Atom	-0.70
solvation	-0.84
torsion	-1.86

**Sequence Features**

- Metal binding
- Site
- Natural variant
- DNA binding
- InterPro

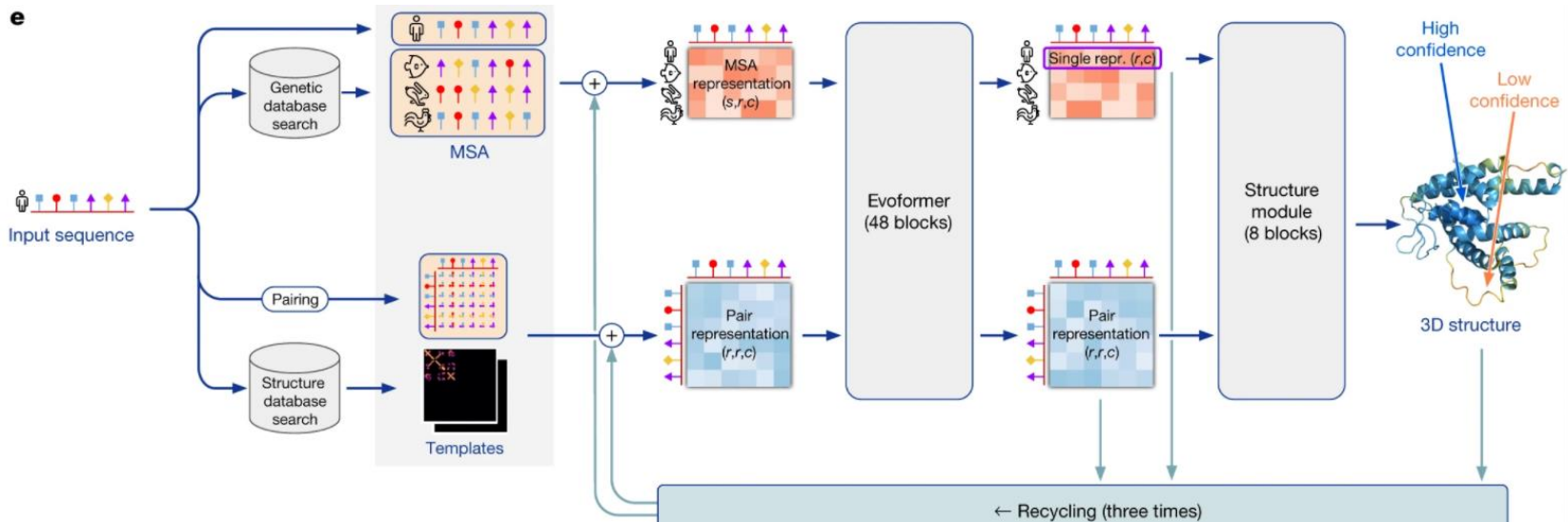
Colours NGI Cartoon

<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.

UniProtKB - Q9GZP4 (PITH1\_HUMAN)

Protein: PITH domain-containing protein 1  
Gene: PITHD1  
Organism: Homo sapiens (Human)  
Status: Reviewed - Annotation score: ●●●●○ - Experimental evidence at protein level<sup>1</sup>

**Function<sup>1</sup>**

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<sup>1</sup>**

- 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

Keywords: Molecular, Activator

Structure

**Model Confidence:**

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.

SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
AlphaFold	AF-Q9GZP4-F1	Predicted			1-211	AlphaFold

3D structure databases

SMR<sup>2</sup>: Q9GZP4  
ModBase<sup>1</sup>: Search...

Moléculas pequenas

# 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 datos de pequeñas moléculas

- PubChem (<http://pubchem.ncbi.nlm.nih.gov/>)
- DrugBank (<http://www.drugbank.ca>)
- ChEMBL (<https://www.ebi.ac.uk/chembl/>)
- ZINC (<http://zinc.docking.org>)
- Cambridge Structural Database (<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 contém informação sobre uma *amostra química* de proveniência bem definida, que pode conter um 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. Contém 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, sinónimos ou keywords.
- **Substance:** nomes, sinónimos, 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 x  
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 ...

BioActivity Summary  
BioActivity Datatable  
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

# PubChem Compound

aspirin - PubChem C... x

https://www.ncbi.nlm.nih.gov/pccompound/?term=aspirin

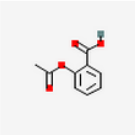
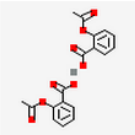
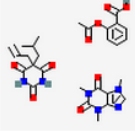
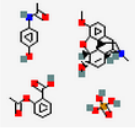
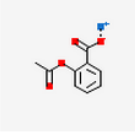
Apps Enzymology Piano Music Production Bioinformatics Databases Bioinformatics T... Misc Programming D pmarcel » Other bookmarks

NCBI Resources How To Sign in to NCBI





PubChem Compound 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 Send to: Filters: Manage Filters

Results: 1 to 20 of 88 << First < Prev Page 1 of 5 Next > Last >>

-  [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](#)
-  [Calcascorbin; Calcium aspirin; Calcascorbate ...](#)  
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 \(MeSH Keyword\)](#)
-  [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 \(MeSH Keyword\)](#)
-  [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](#)
-  [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](#)



**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)

# PubChem Compound

Aspirin - PubChem

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

NCBI

PubChem Compound

Search

Limits Advanced search Help

SHARE

Aspirin - Compound Summary (CID 2244)

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

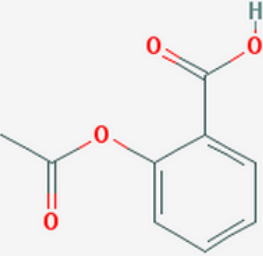
Molecular Formula:  $C_9H_8O_4$  Molecular Weight: 180.15742 InChIKey: BSYNRYMUTXBXSQ-UHFFFAOYSA-N

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



Links and Related Information

Follow us on

Properties

Compound ID: 2244

Molecular Weight: 180.15742 [g/mol]

Molecular Formula:  $C_9H_8O_4$

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 Substance

https://www.ncbi.nlm.nih.gov/pcsubstance/?term=aspirin

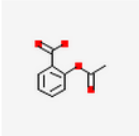
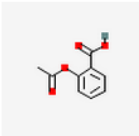
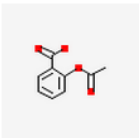
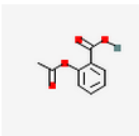
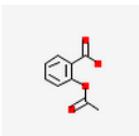
NCBI Resources How To Sign in to NCBI

PubChem Substance  Search





Save search Limits Advanced Help

Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: Manage Filters

Results: 1 to 20 of 547 << First < Prev Page 1 of 28 Next > Last >>

-  **aspirin: ACETYLSALICYLIC ACID; Ecotrin ...**  
Source: [LeadScope \(LS-143\)](#)  
SID: 49854366 [CID: 2244]  
[Summary](#) [PubChem Same Compound](#) [Same Parent, Connectivity](#) [PubMed \(MeSH Keyword\)](#)
-  **aspirin: ACETYLSALICYLIC ACID; Ecotrin ...**  
Source: [Comparative Toxicogenomics Database \(D001241\)](#)  
SID: 53788943 [CID: 2244]  
[Summary](#) [PubChem Same Compound](#) [Same Parent, Connectivity](#) [PubMed \(MeSH Keyword\)](#)
-  **aspirin: ACETYLSALICYLIC ACID; Ecotrin ...**  
Source: [Therapeutic Targets Database \(DAP000843\)](#)  
SID: 134338122 [CID: 2244]  
[Summary](#) [PubChem Same Compound](#) [Same Parent, Connectivity](#) [PubMed \(MeSH Keyword\)](#)
-  **aspirin: ACETYLSALICYLIC ACID; Ecotrin ...**  
Source: [Human Metabolome Database \(HMDB01879\)](#)  
SID: 126524194 [CID: 2244]  
[Summary](#) [PubChem Same Compound](#) [Same Parent, Connectivity](#) [PubMed \(MeSH Keyword\)](#)
-  **aspirin: ACETYLSALICYLIC ACID; Ecotrin ...**  
Source: [ChemIDplus \(0000050782\)](#)  
SID: 134971785 [CID: 2244]  
[Summary](#) [PubChem Same Compound](#) [Same Parent, Connectivity](#) [PubMed \(MeSH Keyword\)](#)

**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)

# PubChem Substance

aspirin - PubChem x

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

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

NCBI

PubChem Substance

PubChem Substance Limits Advanced search 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

ASN.1 XML SDF

Follow us on

Related Substance

- Same (206)
- Same, Connectivity (222)

Other Links

Chemical Structure Search

# PubChem BioAssay

AID 444512 - PubChem x PubChem PC3D View x

pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=444512&loc=ea\_ras

NCBI

PubChem BioAssay

Search

Limits Advanced search Help

SHARE

BioAssay: [AID 444512](#)

CSV ASN.1 XML

>> Additional Info & Links

## Antiplatelets aggregatory activity in human platelets rich plasma assessed as inhibition of collagen-induced platelets aggregation by aggregometry

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 \(Concise\)](#)

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](#) [All](#)

### BioActive Compounds: 3

BioActivity Summary  
Structure-Activity Analysis  
Structure Clustering

### Tested Compounds

All(5)  
Active(3)  
Unspecified(2)

### Tested Substances

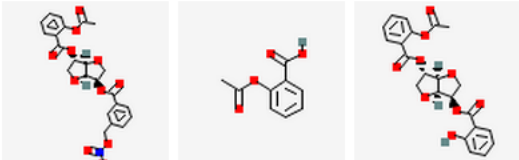
All(5)  
Active(3)  
Unspecified(2)

### Links

[PubMed \(1\)](#)  
[Taxonomy \(1\)](#)

### Related BioAssays

[Activity Overlap \(105\)](#)





# PubChem – Pesquisa por “Tag”

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

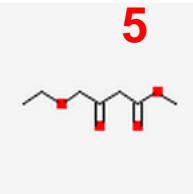
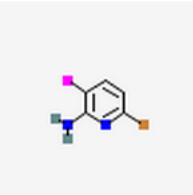
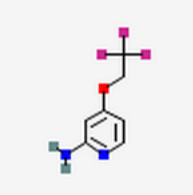
NCBI Resources How To Sign in to NCBI

PubChem Compound PubChem Compound 0:500[mw] 0:5[hbdc] 0:10[hbac] -5:5[logp] Search Help





Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: Manage Filters

Results: 1 to 20 of 34559871 << First < Prev Page 1 of 1727994 Next > Last >>

**Lipinski's rule of 5**

-  **5** [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](#)
-  [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](#)
-  [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 image shows a screenshot of a web browser displaying the PubChem website and a window for PubChem Sketcher V2.4. The browser window shows the PubChem Structure Search page with the URL <https://pubchem.ncbi.nlm.nih.gov/search/search.cgi#>. The PubChem Sketcher V2.4 window is open over the search page, showing the SMILES string CCC(N1CCCC1C(=O)O)=O and a chemical structure diagram of the corresponding molecule, which is a cyclic amide derivative.

The PubChem Sketcher V2.4 interface includes a toolbar with various drawing tools (New, Undo, Clone, Style, Delete, Query, etc.), a periodic table, and a drawing area. The drawing area shows the chemical structure of the molecule defined by the SMILES string. The periodic table is partially visible, showing elements from H to Rn.

The PubChem website interface shows the search options (Name/Text, Identity/Similarity) and the "Draw a Structure" button. The chemical structure shown in the "Draw a Structure" section is a cyclic amide derivative, which matches the structure shown in the PubChem Sketcher V2.4 window.

**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
<b>Total: 41</b>	<b>449,402</b>	<b>4,985,224</b>

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

**B.**

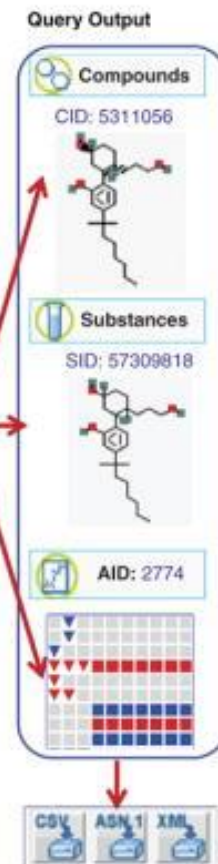


<http://pubchem.ncbi.nlm.nih.gov/>



Literature

**C.**



Download

# Exemplo de pesquisa estrutural na base ChEMBL

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

← → ↻ <https://www.ebi.ac.uk/chembl/> ☆ 🌱 📄 🗄️ ☰

**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](#)

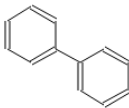
EMBL-EBI Services Research Training Industry About us

## ChEMBL

EBI > Databases > Small Molecules > ChEMBL Database > Home

Search ChEMBL... Compounds Targets Assays Documents [Activity Source Filter](#)

**Ligand Search** Target Search Browse Targets Browse Drugs Browse Drug Targets Drug Approvals About



**List Search**

SMILES Search  ChEMBL ID Search  Keyword Search

Please enter a list of Compound IDs, keywords, or SMILES separated by newlines

**Biologicals Blast Search**

**ChEMBL Statistics**

- DB: ChEMBL\_17
- Targets: 9,356
- Compound records: 1,620,172
- Distinct compounds: 1,324,941
- Activities: 12,077,491
- Publications: 51,277
- [Release Notes](#)

**ChEMBL Blog**

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

# Exemplo de pesquisa estrutural na base ChEMBL

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

← → ↻ <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 Services Research Training Industry About us

## ChEMBL

ChEMBL Downloads Malaria Data ChEMBL-NTD Kinase SARfari GPCR SARfari DrugEBlity Web Services FAQ

ChEMBL Statistics

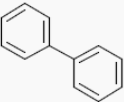
- DB: ChEMBL\_17
- Targets: 9,356
- Compound records: 1,520,172
- Distinct compounds: 1,324,841
- Activities: 12,077,491
- Publications: 51,277
- [Release Notes](#)

ChEMBL Blog

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

Search ChEMBL... Compounds Targets Assays Documents [Activity Source Filter](#)

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	<input checked="" type="checkbox"/>
 <a href="#">CHEMBL14092</a>	SID17390012 E230 SID26753117	100	0	154.21	3.35	0	0	0	0	1	N	Y	.59	<input checked="" type="checkbox"/>

Showing 1 to 1 of 1 entries

[← Previous](#) 1 [Next →](#)

# Exemplo de pesquisa estrutural na base ChEMBL

Blues Drums Bu x thelooploft's st x Mino Cinelu - V x RCSB Protein D x Configurar - Ap x Over To You: M x RCSB Protein D x Compound Rep x Server error < E x

← → ↻ <https://www.ebi.ac.uk/chembl/compound/inspect/CHEMBL14092> ☆ 🌱 📄 🔍 ☰

EMBL-EBI **ChEMBL** Services Research Training Industry About us

ChEMBL Downloads Malaria Data ChEMBL-NTD Kinase SARfari GPCR SARfari DrugEBility Web Services FAQ

**ChEMBL Statistics**

- DB: ChEMBL\_17
- Targets: 9,366
- Compound records: 1,520,172
- Distinct compounds: 1,324,941
- Activities: 12,077,491
- Publications: 61,277
- [Release Notes](#)

**ChEMBL Blog**

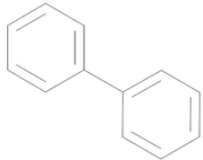
- [EU-OPENSOURCE 3rd Stakeholder Meeting, Oslo, Norway](#)
- [Competition Time - Win a Raspberry Pi with ChEMBL - chempi](#)

EBI > Databases > Small Molecules > ChEMBL Database > Compound Search > CHEMBL14092

### Compound Report Card

**Compound Name and Classification**

Compound ID	CHEMBL14092
Compound Name	BIPHENYL
ChEMBL Synonyms	SID17390012   E230   SID26753117
Max Phase	0
Trade Names	
Molecular Formula	C12 H10



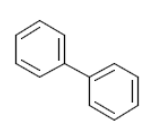
CHEMBL14092

Additional synonyms for CHEMBL14092 found using [NCI Chemical Identifier Resolver](#)

**Compound Representations**

Molfile	<a href="#">Download MolFile</a>
Canonical SMILES	<chem>c1ccc(cc1)c2ccccc2</chem>
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**



CHEMBL14092

**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 surprisi... 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 [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: Acetylsalic x  
 www.drugbank.ca/drugs/DB00945  
 Gnuplot surprisi... Gnuplot tricks N Old Nabble - G... Music Production Bioinformatics 10.10.20.1/form... + Pocket Other bookmarks

# DRUGBANK

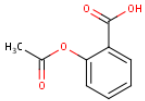
Open Data Drug & Drug Target Database

Home Browse Search Downloads News & Updates About Help Contact Us

Search:  Search [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	<b>Acetylsalicylic acid</b>
Accession Number	<b>DB00945</b> (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	 <p>Download: <a href="#">MOL</a>   <a href="#">SDF</a>   <a href="#">SMILES</a>   <a href="#">InChI</a>            Display: <a href="#">2D Structure</a>   <a href="#">3D Structure</a></p>
Synonyms	<ul style="list-style-type: none"> <li>• 2-Acetoxybenzenecarboxylic acid</li> <li>• 2-Acetoxybenzoic acid</li> <li>• 2-Carboxyphenyl acetate</li> <li>• A.S.A.</li> <li>• Acetilsalicilico</li> <li>• Acetilum acidulatum</li> <li>• Acetosalic acid</li> <li>• Acetoxybenzoic acid</li> <li>• Acetylsalicylate</li> <li>• Acetylsalicylsäure (GERMAN)</li> <li>• Acetylsalicylic acid</li> <li>• Acide acetylsalicylique (FRENCH)</li> </ul>