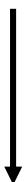


Bioinformática Estrutural

Sequência



Estrutura



Função

Fluxo de informação biológica

Gene ...TTAATAAGT...

↓ transcrição



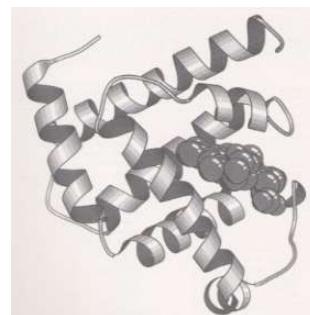
m-RNA ...UUAAUAAGU...

↓ splicing, tradução

cadeia
polipeptídica ...LISVHDN...

↓ modificações pós-translacionais

proteína



Dogma central da
biologia molecular

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

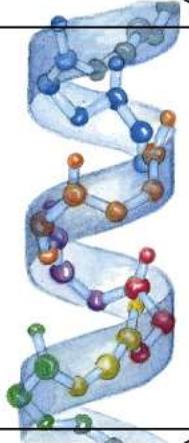
Níveis de organização da estrutura das proteínas

Estrutura primária



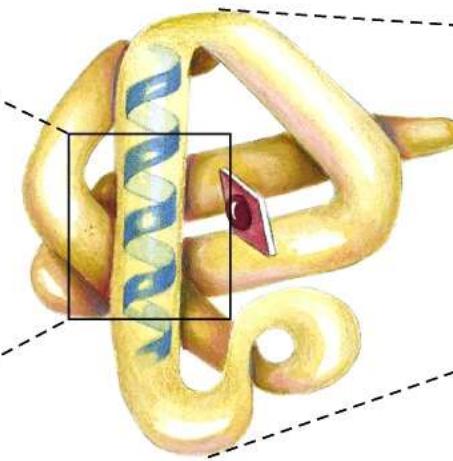
Sequência de aminoácidos

Estrutura secundária



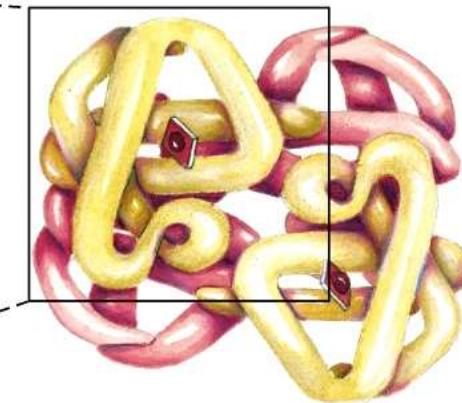
α -hélice

Estrutura terciária



Cadeia polipeptídica

Estrutura quaternária



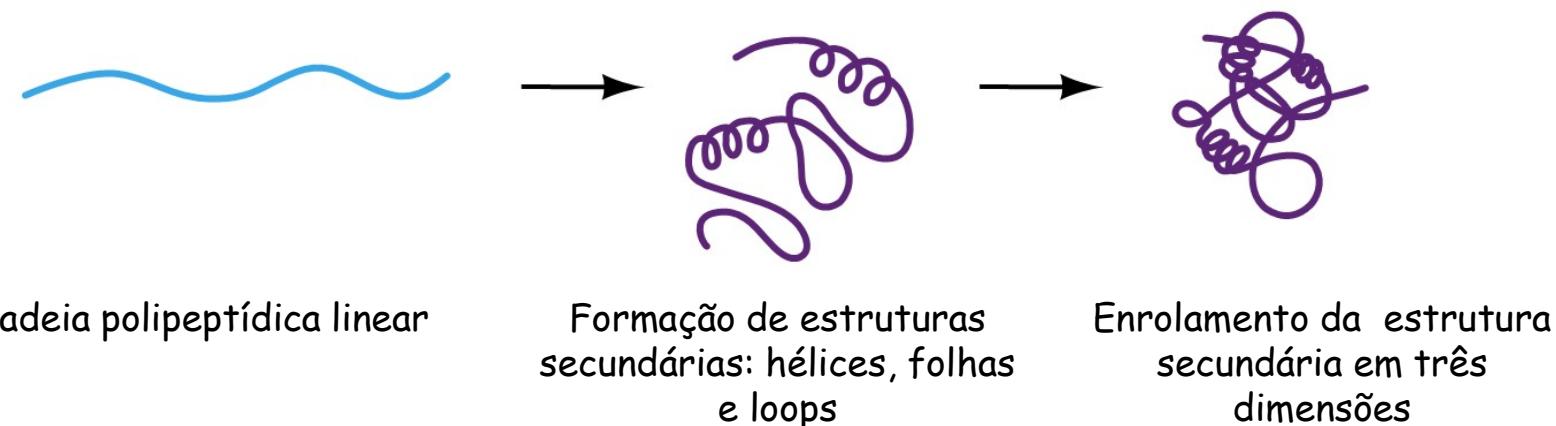
Organização das subunidades

A estrutura das proteínas é determinada pela sua sequência

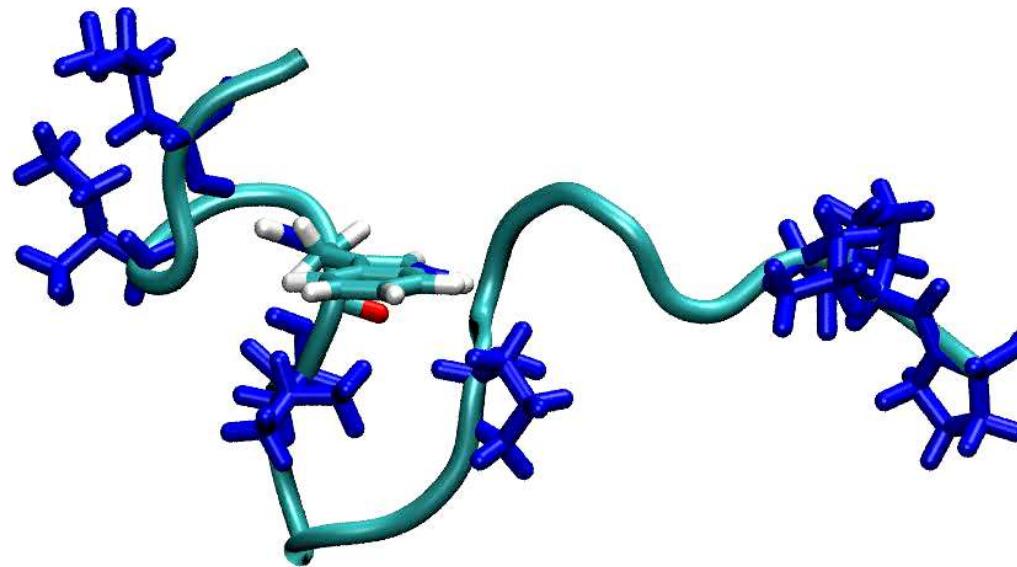
A estrutura tridimensional das proteínas é resultado das interacções entre os átomos que a constituem e o meio aquoso. Em muitos casos a cadeia polipeptídica assume a sua conformação nativa de modo espontâneo, após a síntese ribossomal. Este processo tem o nome de "protein folding".

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

Mecanismo do "folding" das proteínas:



Sequência->Estrutura



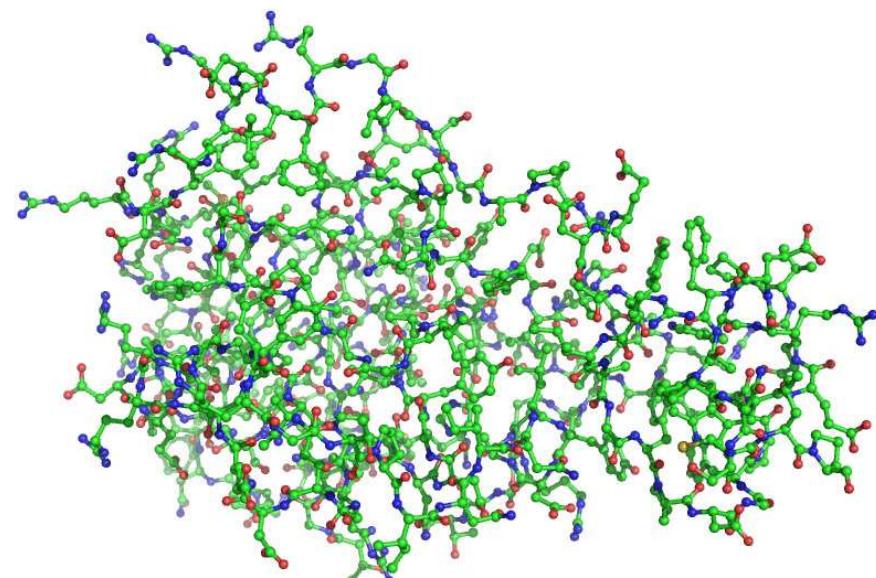
Muitas protéinas adquirem a sua estrutura tridimensional espontâneamente (*folding*)

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

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

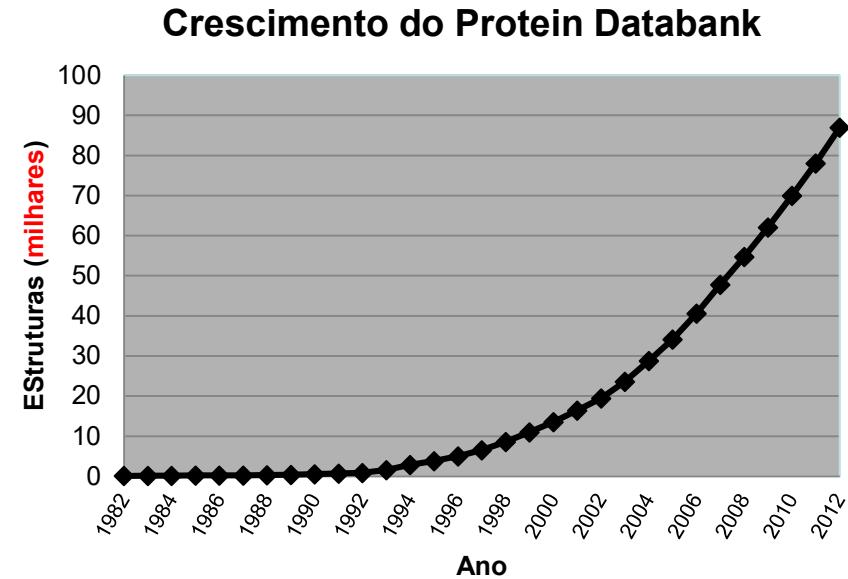
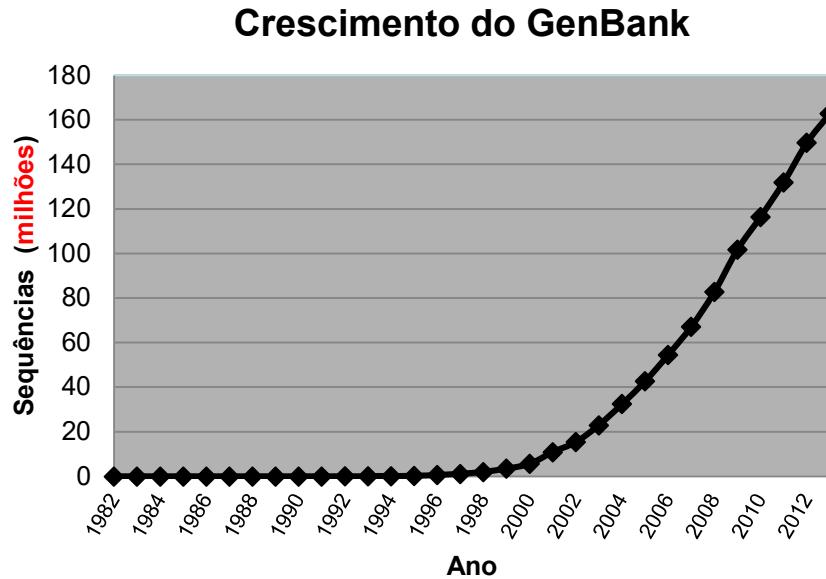
...AVAGGATILVHNQDAGEPAIVLAFG...

Sequência



Estrutura

Sequência versus estrutura



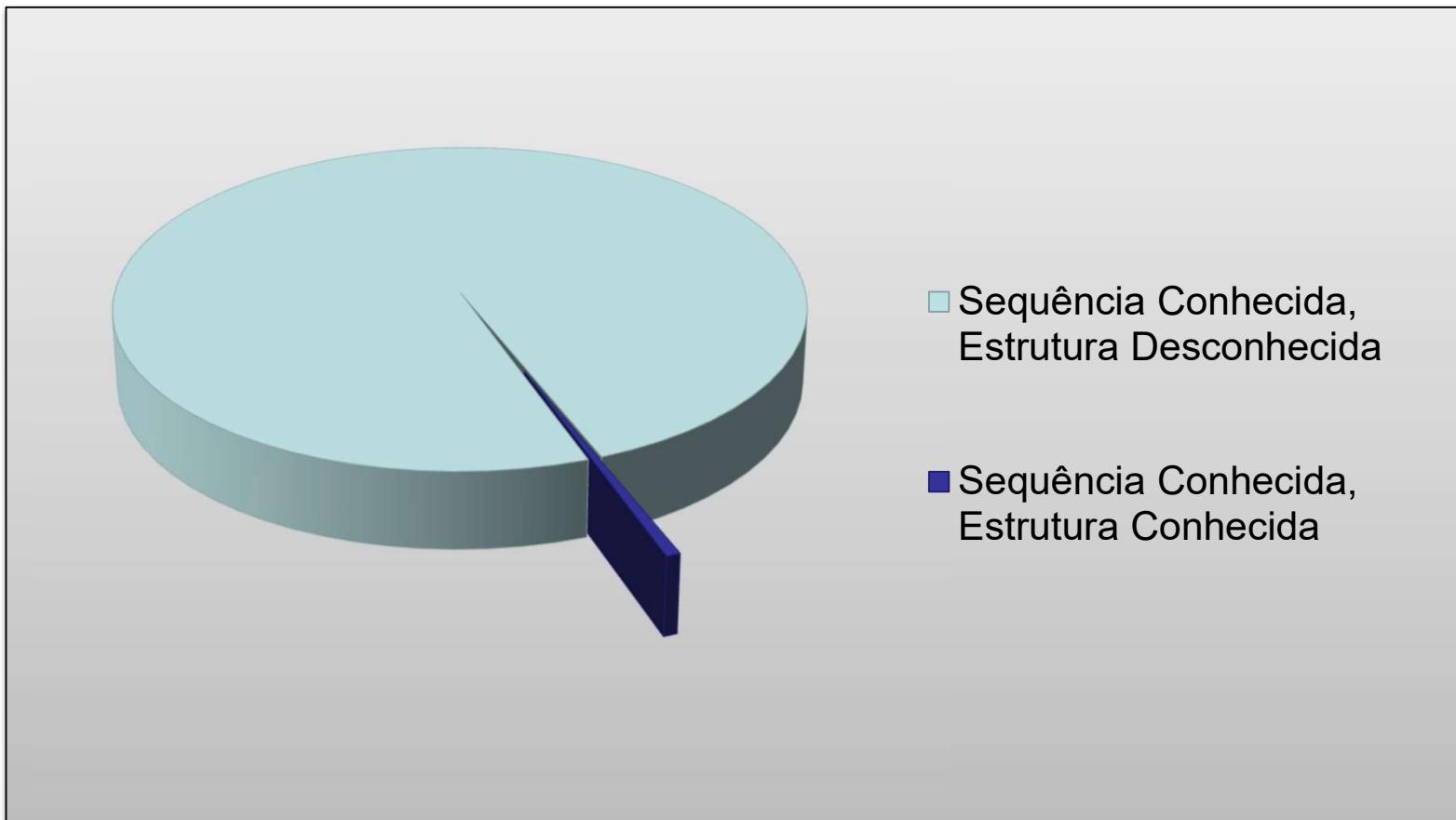
milhões de sequências versus milhares de estruturas!

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

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

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

A maior parte das proteínas conhecidas tem estrutura desconhecida



Importância da previsão estrutural

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

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

Como resolver este problema ?

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

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

I. Bancos de dados de estrutura

Macromoléculas

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

O Protein Data Bank

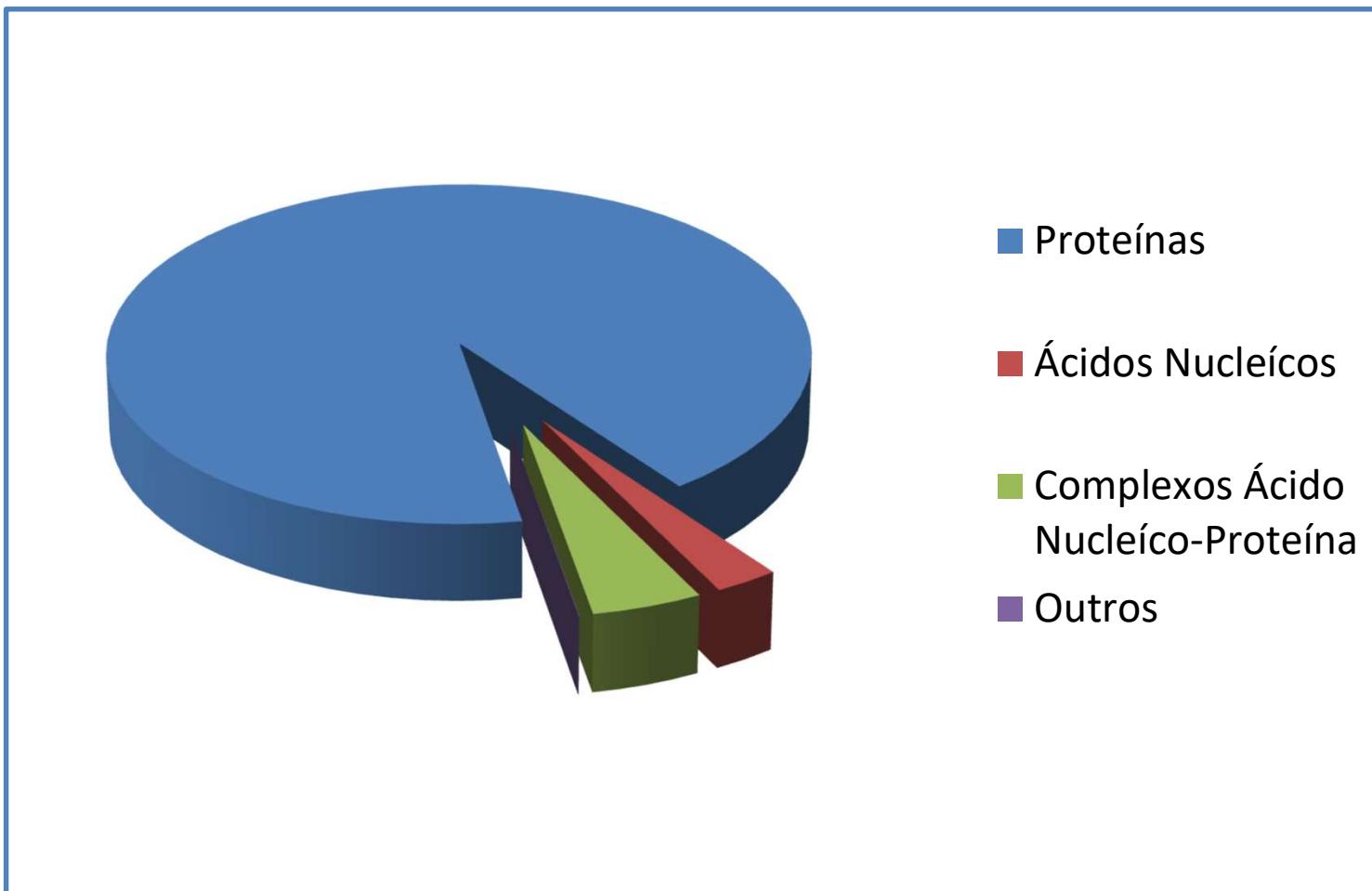
- O Protein Data Bank (PDB) foi criado em 1971 por E.Meyer e W.Hamilton, do Brookhaven National Laboratory (USA), contendo no início 7 estruturas!
- A gestão do PDB foi transferida em 1998 para os membros do RSCB (Research Collaboratory in Structural Bioinformatics) dos quais a Universidade de Rutgers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

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

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

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

O Protein Data Bank contém vários tipos de macromoléculas



De onde provêm a informação estrutural ?

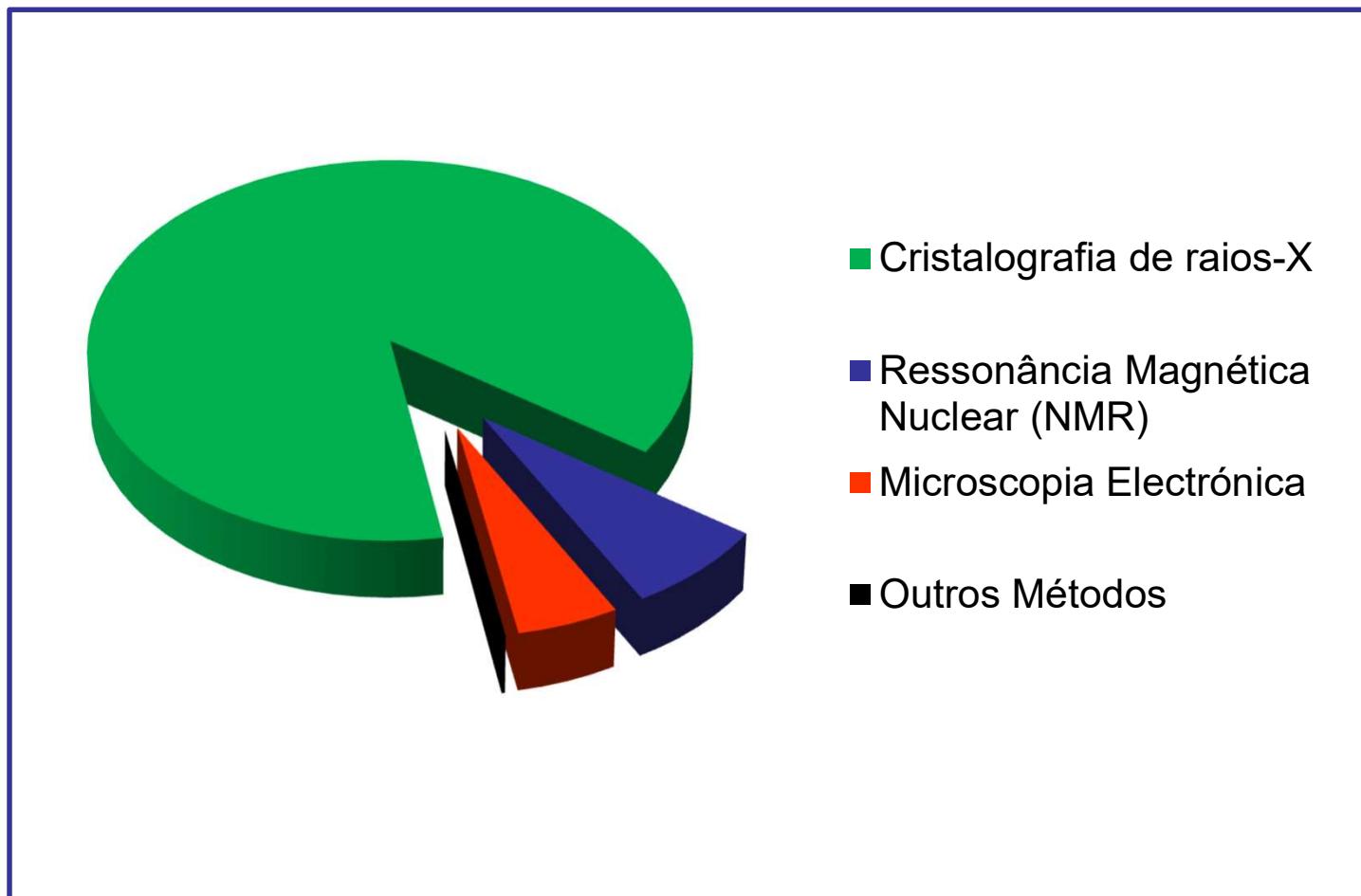
Combinação de vários tipos de conhecimento:

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

Métodos experimentais

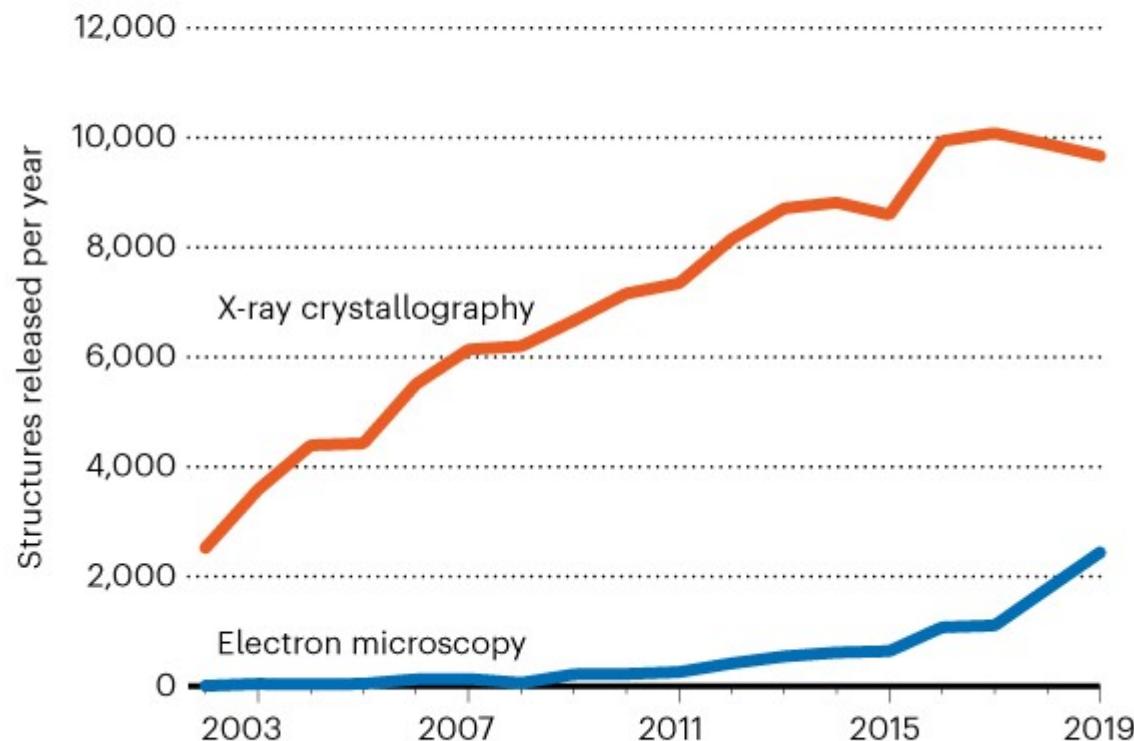
- **Cristalografia de raios X:** a molécula a estudar é purificada e cristalizada a partir de uma solução concentrada. Um feixe de raios X é projectado através do cristal da molécula e o padrão de difracção obtido é usado para resolver a estrutura.
- **Ressonância magnética Nuclear:** a molécula purificada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

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



STRUCTURE SLEUTHS

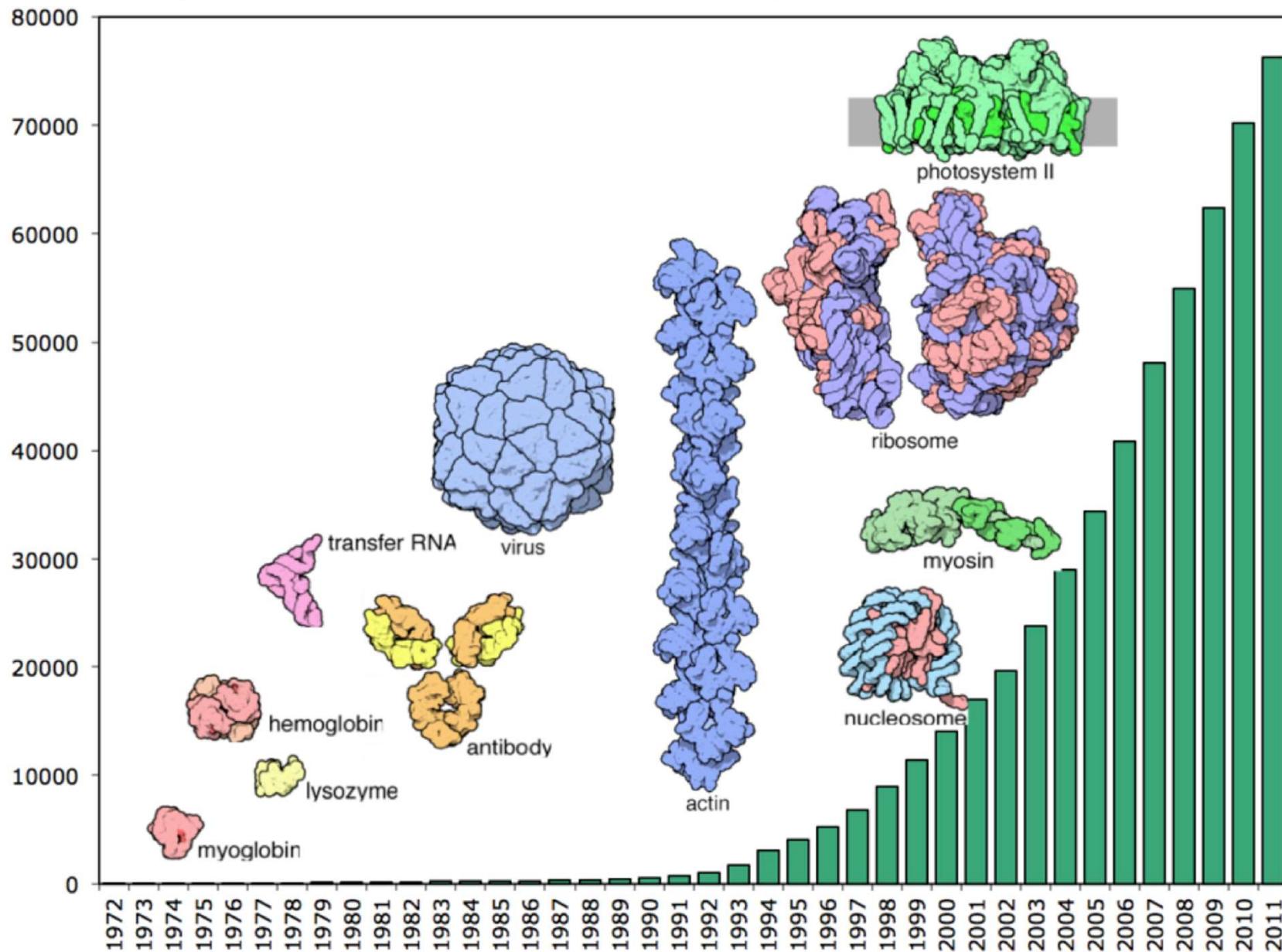
Most structures of proteins and other biological molecules are still solved with X-ray crystallography. But a revolutionary technique called cryo-electron microscopy is catching up, as it becomes more sensitive and widely available.



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

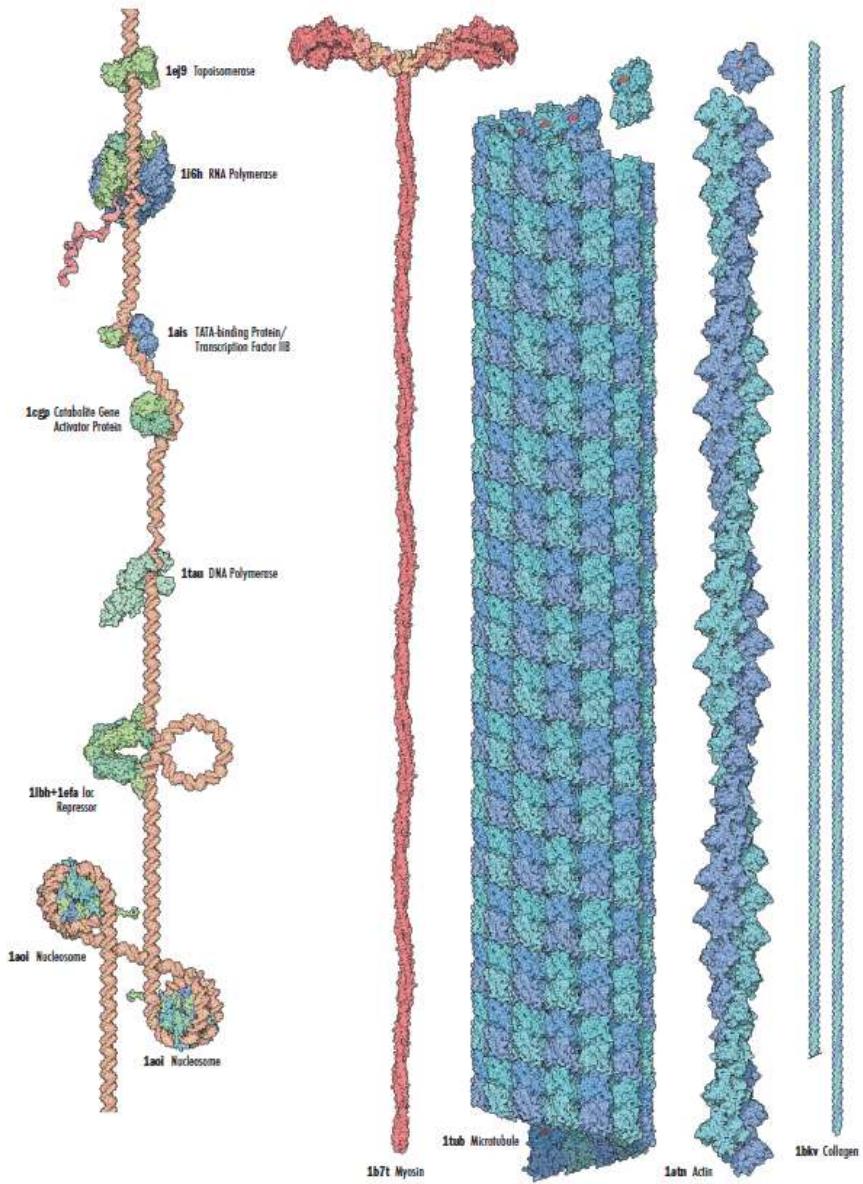
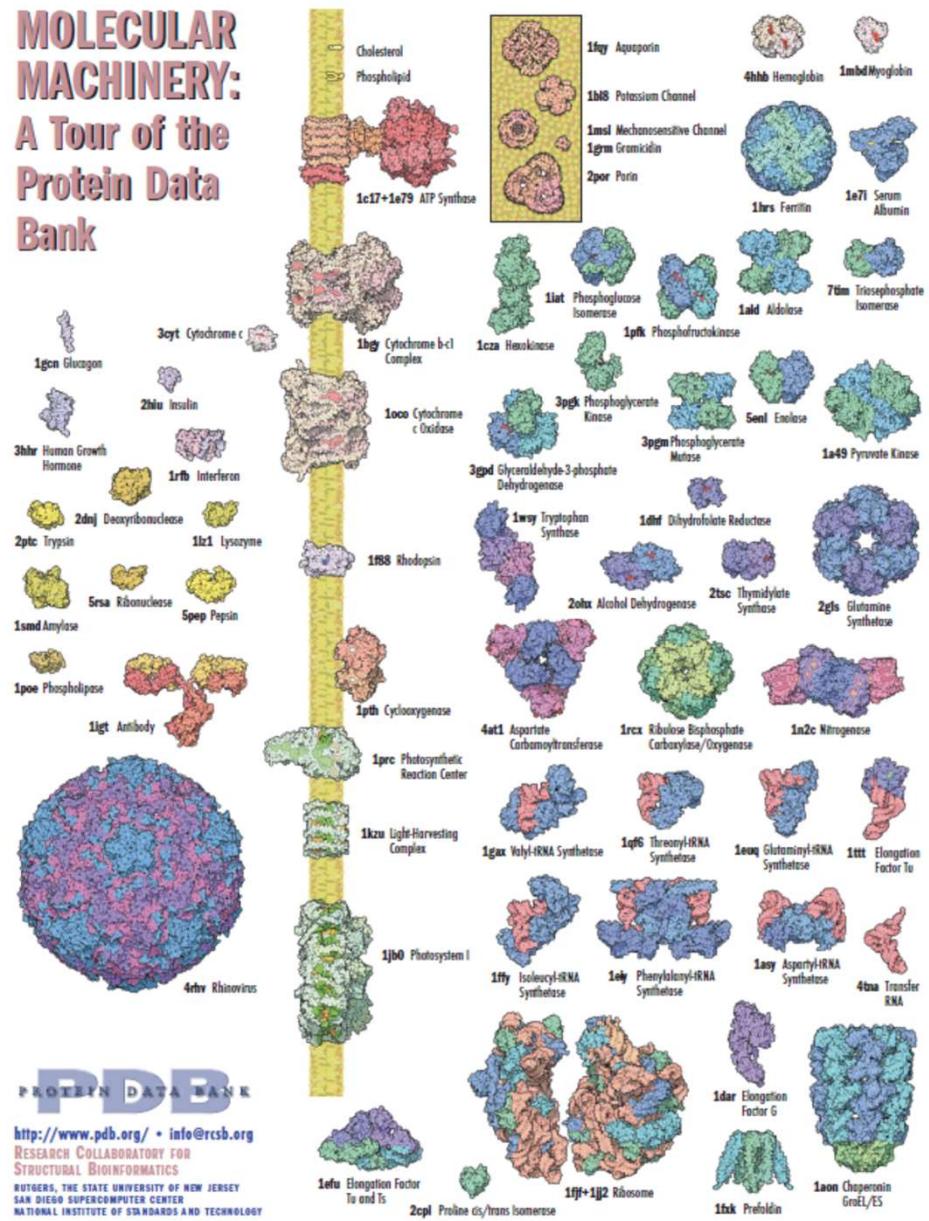
©nature

Progresso na determinação das estruturas



O PDB contém uma enorme diversidade estrutural!

MOLECULAR MACHINERY: A Tour of the Protein Data Bank



Portal de acesso ao PDB

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

158180 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

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

PDB-101 EMDataResource Worldwide Protein Data Bank Foundation

Welcome Deposit Search Visualize Analyze Download Learn

A Structural View of Biology

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

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

Video: How Enzymes Work

VIDEO: HOW ENZYMES WORK

November Molecule of the Month

Phospholipase A2

Latest Entries As of Tuesday Nov 26 2019

6J6B PDB Entry

Borrelia burgdorferi OspA via surface entropy reduction (form2)

Features & Highlights

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

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

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

Join Our Team as a Biocurator

News Publications

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

Introducing Mol* 11/19/2019

New Papers on Molecular Visualization 11/12/2019

New EM map validation in OneDep 11/05/2019

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

PDB Turns 48 10/20/2019

Happy Birthday, Irving Geis 10/18/2019

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

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

<https://www.rcsb.org>

Portal de acesso ao PDB

RCSB PDB: Homepage Paulo

Secure | https://www.rcsb.org

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

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

Janela de pesquisa

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

137692 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

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

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Biology

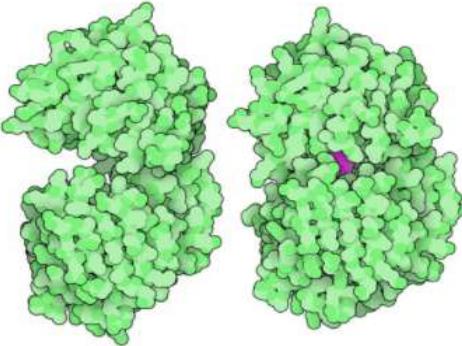
This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

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

New Video: What is a Protein?



February Molecule of the Month



EPSP Synthase and Weedkillers

Contact Us

Latest Entries As of Tuesday Feb 13 2018



Features & Highlights

 New Architecture and Services Enable Faster Access to More Information

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

News Publications

 Meet RCSB PDB at AAAS

Learn how RCSB PDB is Sustaining A Living Digital

Portal de acesso ao PDB

RCSB PDB - Query Re x

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

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

All Categories Author Macromolecule Sequence Ligand ?

Search | All Categories: lysozyme

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

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

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

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

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

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

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

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

Query Details | Save Query to MyPDB

Query Refinements: Select an item or pie chart Show

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

Enzyme Classification SCOP Classification

Refine Query with Advanced Search Remove Similar: Select Percent Similarity

5 Related Molecule of the Month articles Show

PDB Pioneers Nanobodies Adrenergic Receptors Antibodies Lysozyme

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

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

1JKB HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA
Authors: Muraki, M., Harata, K., Goda, S., Nagahora, H.
Release: 1997-05-15 Classification: Lysozyme
Experiment: X-RAY DIFFRACTION with resolution of 1.66 Å Residue Count: 130
Compound: 1 Polymer [Display Full Polymer Details | Display for All Results]
1 Ligand [Display Full Ligand Details | Display for All Results]
Citation: Importance of van der Waals contact between Glu 35 and Trp 109 to the catalytic action of human lysozyme.
(1997) Protein Sci. 6: 473-476 [Display Full Abstract | Display for All Results]
Molecule of the Month:
Search Hit: Classification: LYSOZYME

1JKC HUMAN LYSOZYME MUTANT WITH TRP 109 REPLACED BY PHE
Authors: Muraki, M., Harata, K., Goda, S., Nagahora, H.
Release: 1997-05-15 Classification: Lysozyme

Portal de acesso ao PDB

RCSB Protein Data Bank

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

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

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

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

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

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

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

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

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

HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA

DOI:10.2210/pdb1jkb/pdb

1JKB

Display Files
Download Files
Share this Page

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

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

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

Biological Assembly

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

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

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

Revision History Hide
Mouse over text for details

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

Obter o ficheiro de estrutura em formato PDB

RCSB Protein Data Bank

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

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

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

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

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

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

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

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

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

HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA

DOI:10.2210/pdb1jkb/pdb

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

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

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

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

Source

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

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

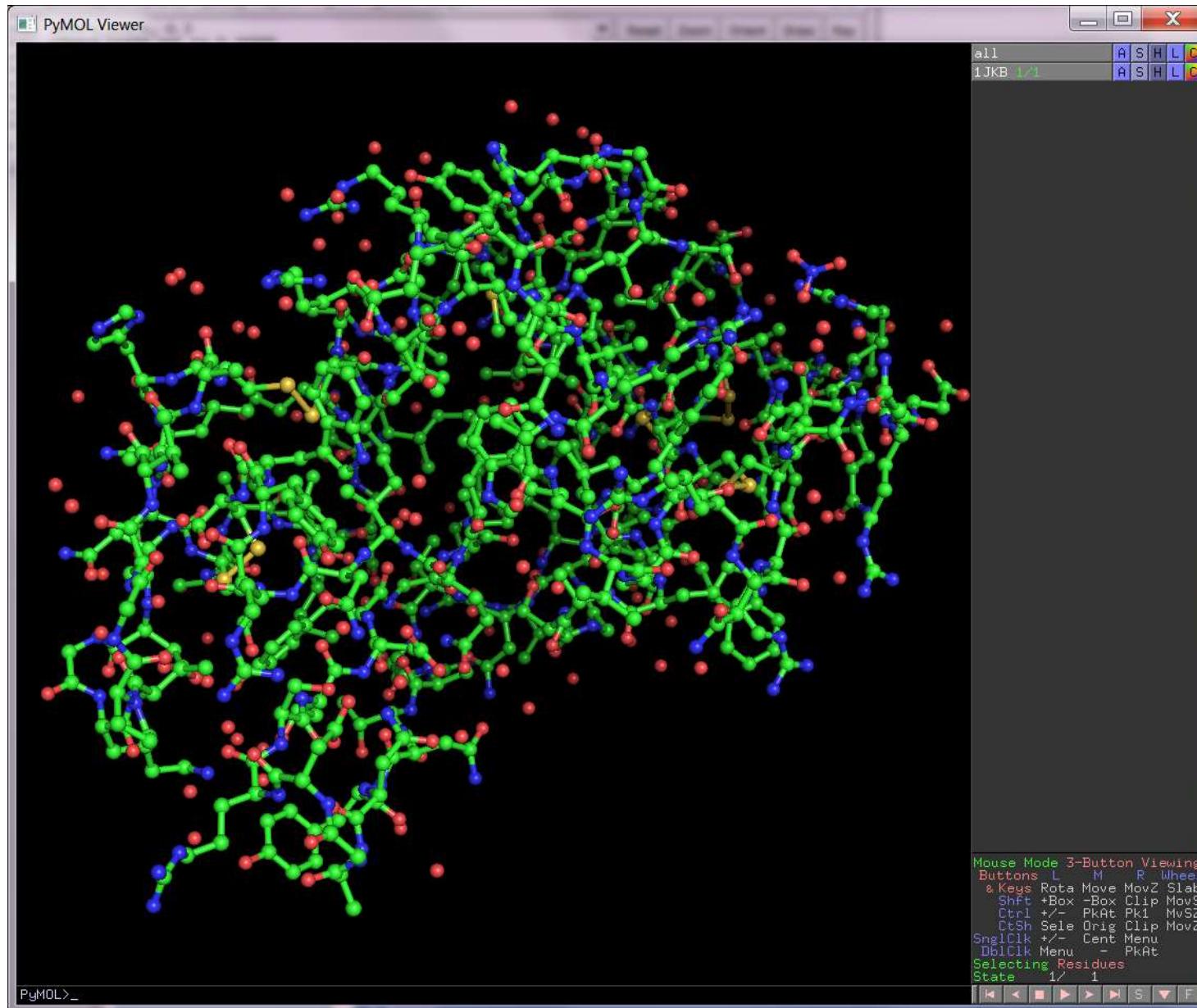
MyPDB Personal Annotations Hide
To save personal annotations, please login to your MyPDB account.

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

Revision History ? Hide
Mouse over text for details

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

Visualizar o ficheiro de estrutura no programa PyMOL



Formatos de representação da estrutura

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

Formato da informação no Protein Data Bank

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

Formato do ficheiro PDB

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

.....

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

.....

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

Cabeçalho

Coordenadas

Interligação entre Uniprot e PDB

The screenshot shows a web browser window displaying the UniProtKB entry for protein P00760 (TRY1_BOVIN). The page has a dark blue header with tabs for 'Search', 'Blast *', 'Align', 'Retrieve', and 'ID Mapping *'. Below the header is a search bar with dropdowns for 'Search in' (set to 'Protein Knowledgebase (UniProtKB)') and 'Query' (empty), along with 'Search', 'Advanced Search», and 'Clear' buttons. The main content area displays the protein's name, P00760 (TRY1_BOVIN), its status as 'Reviewed', and its source, UniProtKB/Swiss-Prot. It also shows the last modification date (October 16, 2013) and version (149). A 'History...' link is available. On the right side, there are links for 'Contribute', 'Send feedback', and 'Read comments (0) or add your own'. Below the main content, there are links for 'Clusters with 100%, 90%, 50% identity', 'Documents (3)', and 'Third-party data'. A row of download links for 'text', 'xml', 'rdf/xml', 'gff', and 'fasta' formats is shown. The navigation menu at the bottom includes 'Names', 'Attributes', 'General annotation', 'Ontologies', 'Interactions', 'Sequence annotation', 'Sequences', 'References', 'Web links', and 'Cross-refs'. The 'Cross-refs' tab is highlighted with a red box. The 'Names and origin' section contains detailed information about the protein's names (Recommended name: Cationic trypsin, EC=3.4.21.4; Alternative name(s): Beta-trypsin), its cleavage into two chains (Alpha-trypsin chain 1 and Alpha-trypsin chain 2), and its organism (Bos taurus (Bovine)). The 'Organism' section also lists the taxonomic identifier (9913 [NCBI]) and the taxonomic lineage (Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Laurasiatheria > Cetartiodactyla > Ruminantia > Pecora > Bovidae > Bovinae > Bos).

Interligação entre Uniprot e PDB

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

Cross-references

Sequence databases

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

3D structure databases

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

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

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

PDB PROTEIN DATA BANK **PDB-101**

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

Search Advanced Browse

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

Search History, Previous Results

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

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

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

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

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

TRYPSIN WITH INHIBITOR AERUGINOSIN 98-B

DOI:10.2210/pdb1aq7/pdb

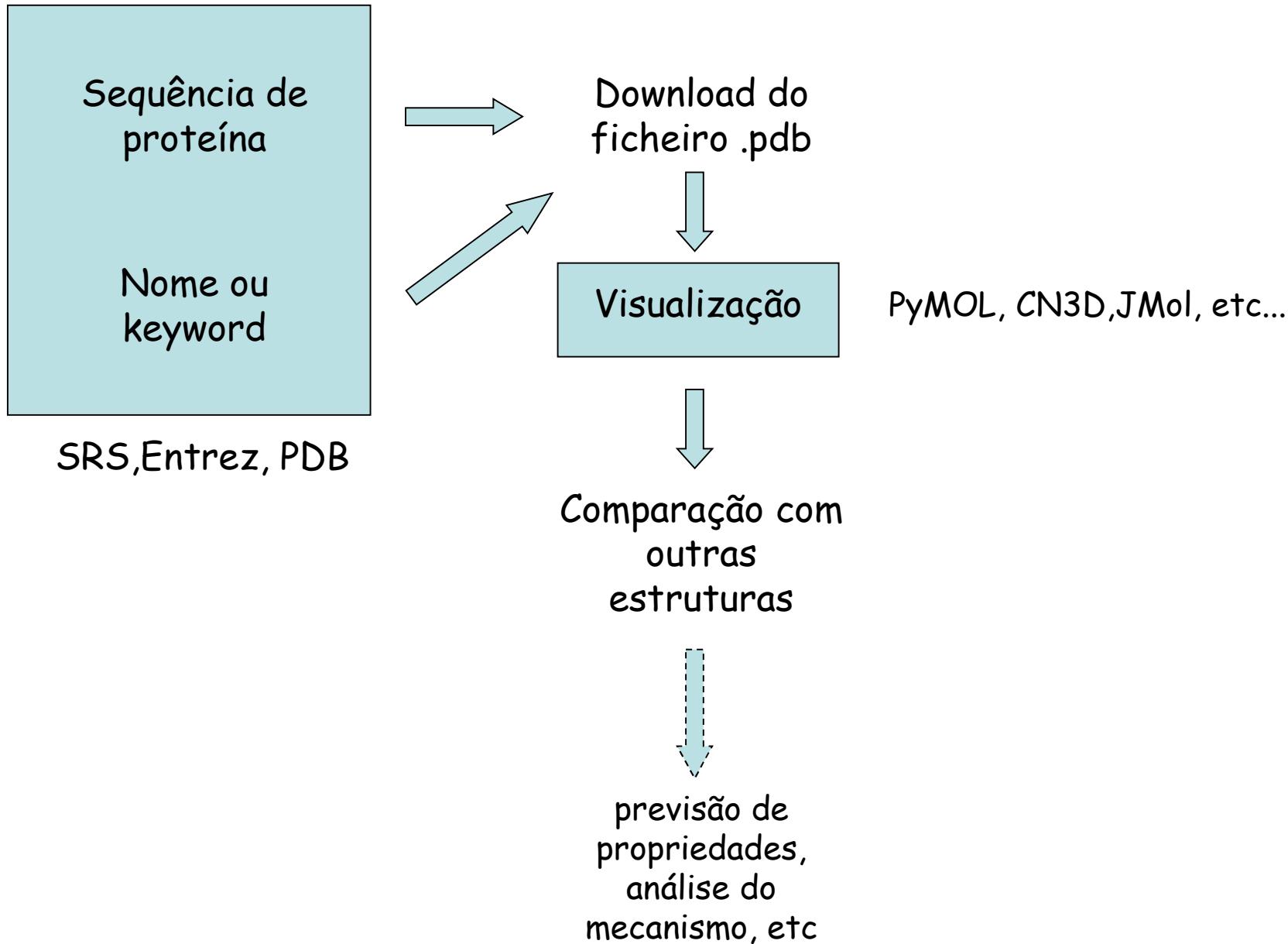
Primary Citation
Atomic Structure of the Trypsin-Aeruginosin 98-B Complex
Sandler, B. , Murakami, M. , Clardy, J.
Journal: (1998) J.Am.Chem.Soc. **120**: 595-596
PubMed ID is not available

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

Molecule:	TRYPSIN
Polymer:	1
Type:	protein
Chains:	A
EC#:	3.4.21.4

Biological Assembly

Visualização de estruturas moleculares

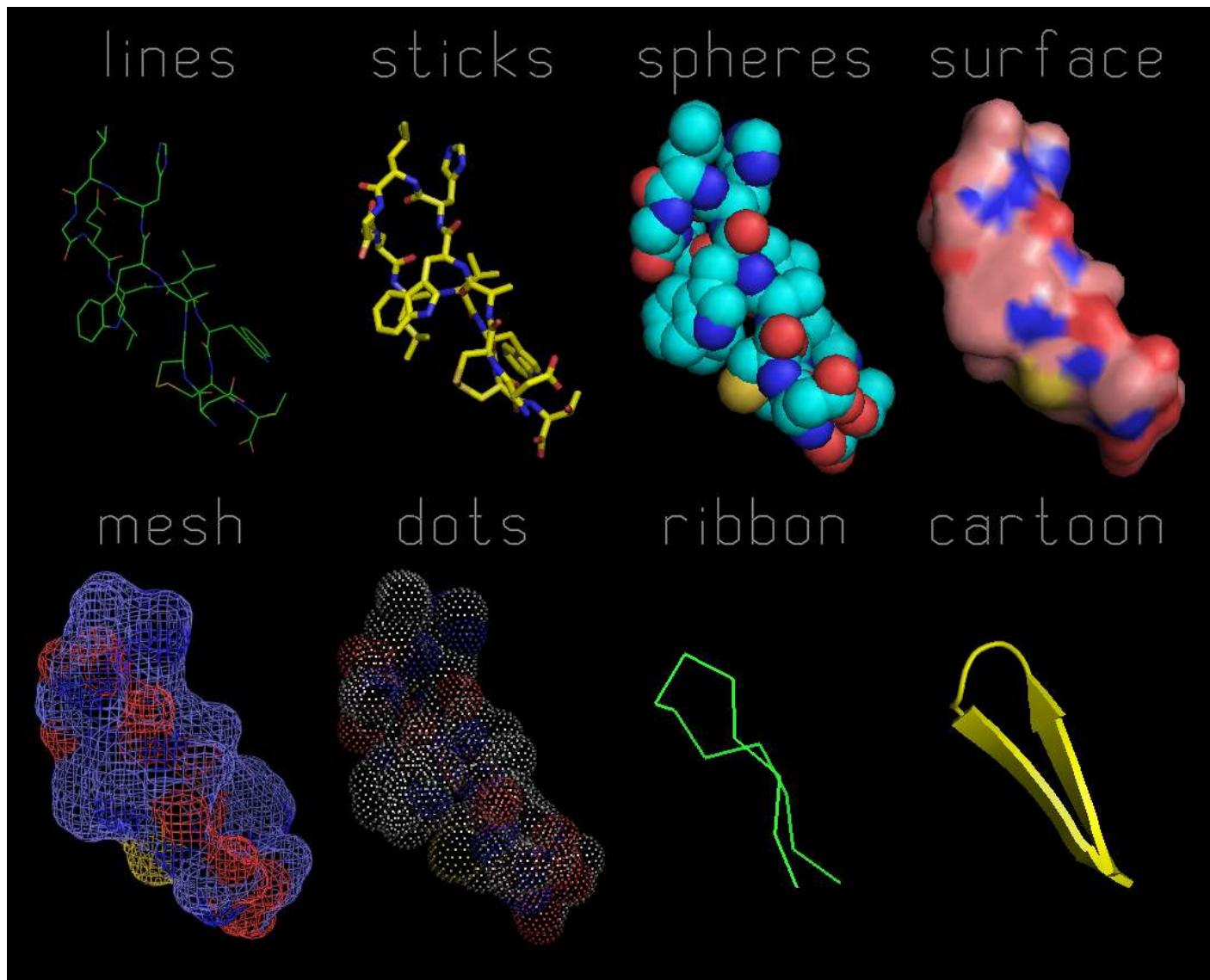


Software para visualização molecular

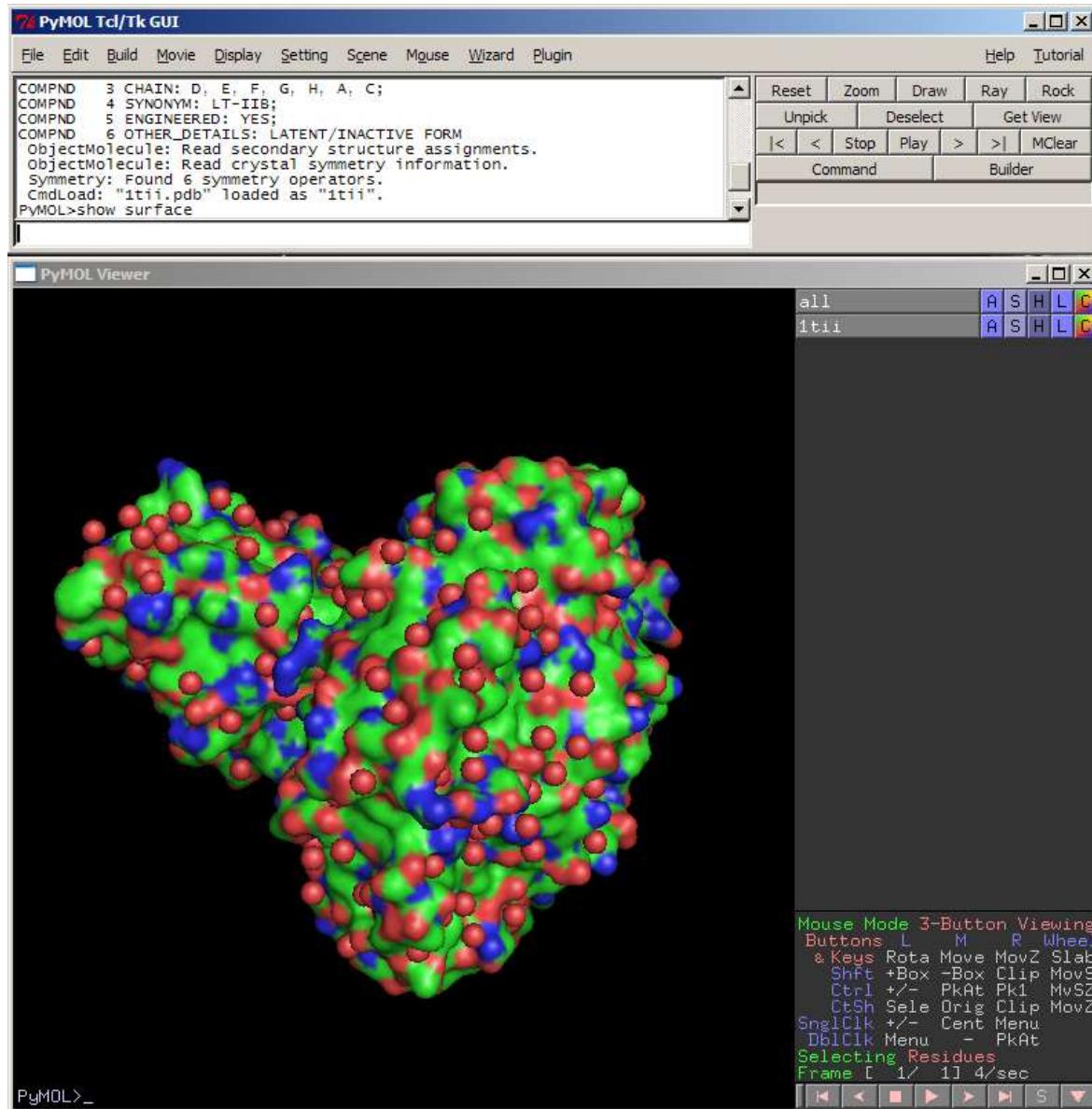
Aplicações de software que permitem a visualização de ficheiros de estrutura molecular (ficheiros PDB e outros formatos), permitindo a análise e cálculo de propriedades moleculares e a comparação de diferentes estruturas

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

Modos de representação de estruturas



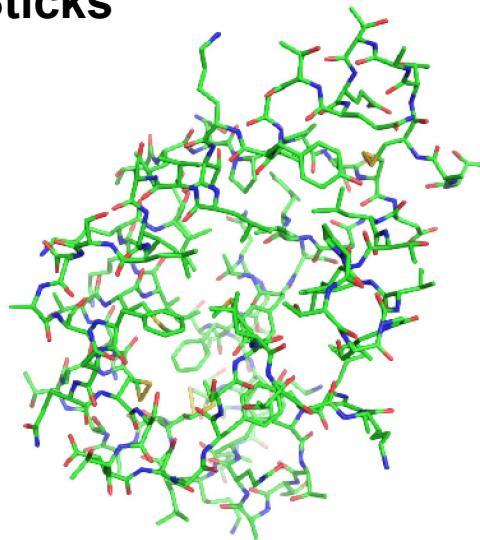
PyMOL



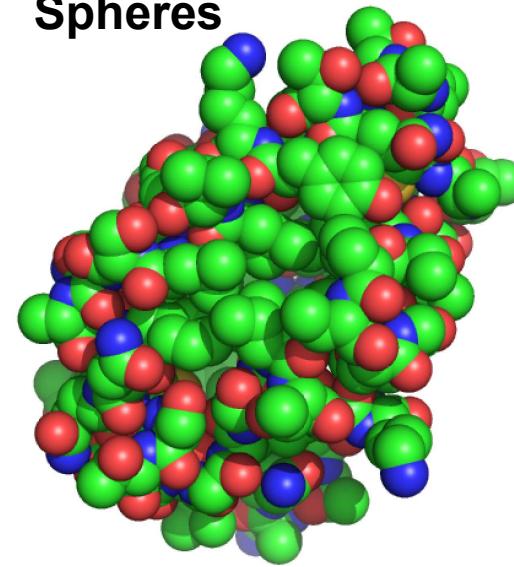
- ❖ Open Source
- ❖ Acesso livre
- ❖ Python / C
- ❖ Visualização de macromoléculas
- ❖ Animações moleculares
- ❖ Comparação de estruturas
- ❖ Scripting
- ❖ Windows / Linux

<http://www.pymol.org>

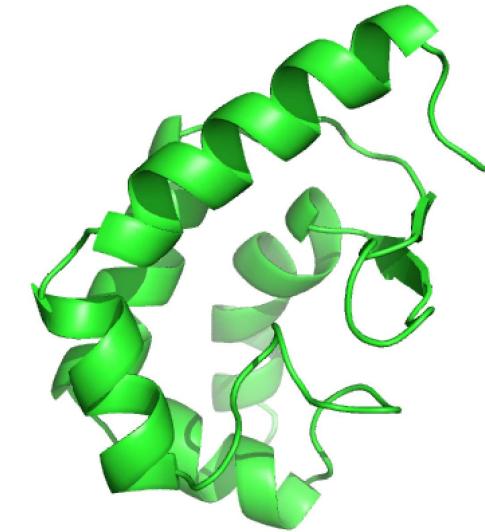
Sticks



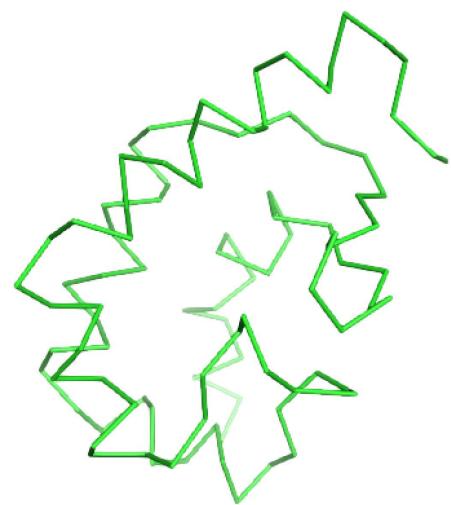
Spheres



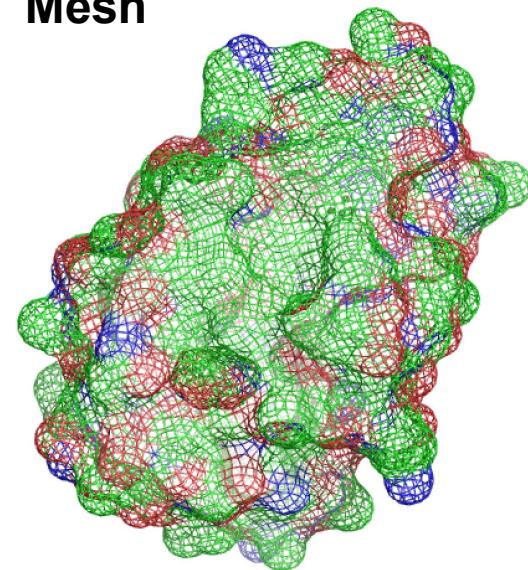
Cartoon



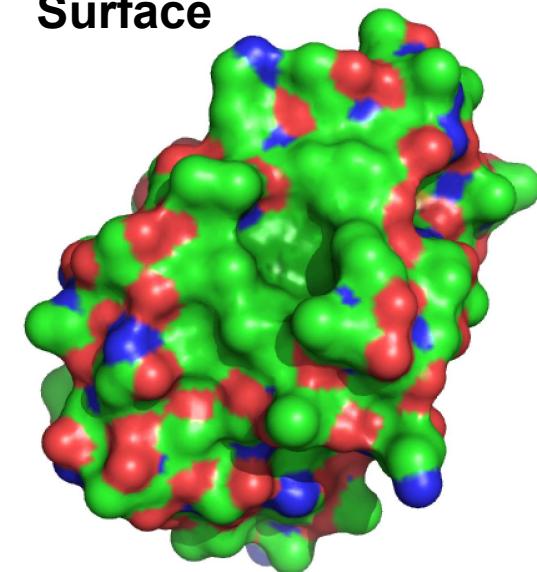
Ribbon



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



MAS



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

Similaridade estrutural e de sequência

Tripsina bovina



Tripsina *S. griseus*



Alinhamento das sequências: 34% identidade, E(1) 1.4×10^{-17}

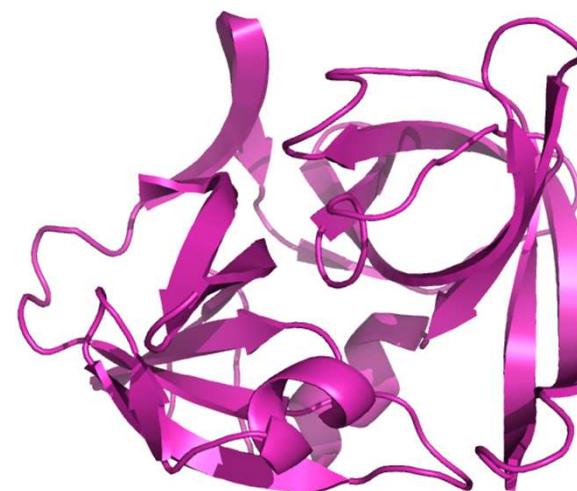
Proteínas homólogas, similaridade de sequências
claramente detectável

Similaridade estrutural e de sequência

Tripsina bovina



Protease A *S. griseus*



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

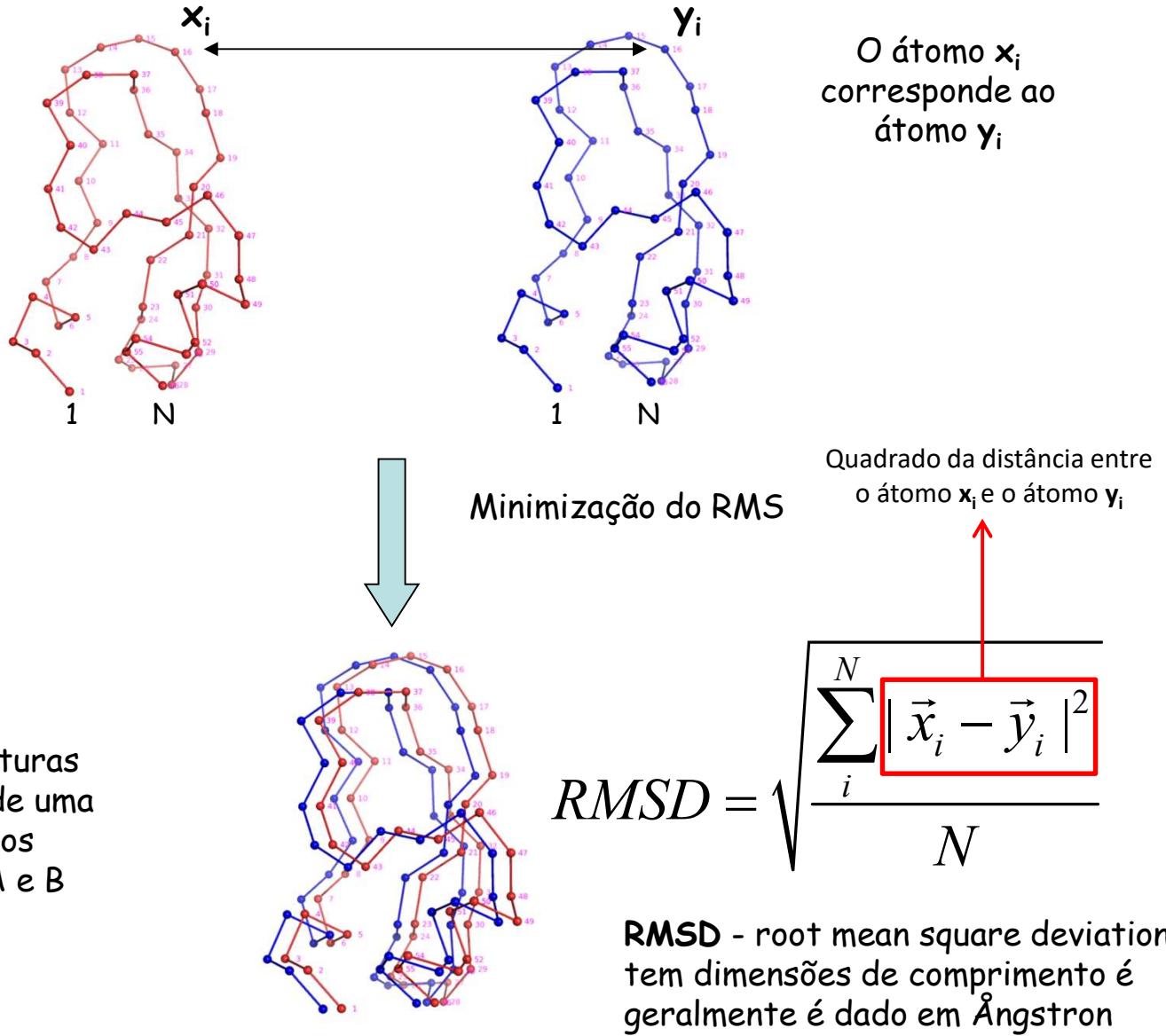
Proteínas homólogas, similaridade de sequências
não é detectável

Alinhamento sem
significado estatístico

Como quantificar a similaridade estrutural ?

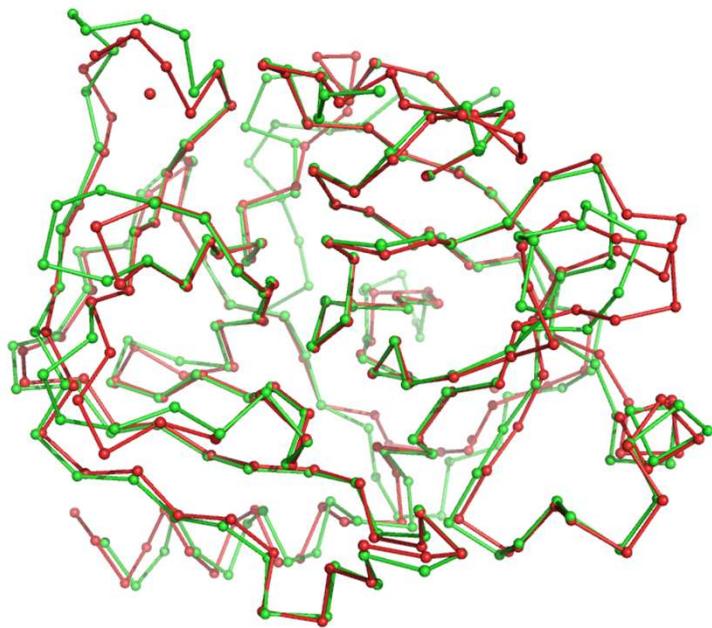
- Tal como a similaridade de sequências, a similaridade de estruturas pode ser quantificada usando diferentes medidas
- O método mais comum consiste em calcular o **desvio quadrático médio (RMSD)** entre pares de átomos das duas estruturas (geralmente expresso em Ångstrons ou nanómetros)
- O valor de 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



Relação entre RMSD e identidade de sequência

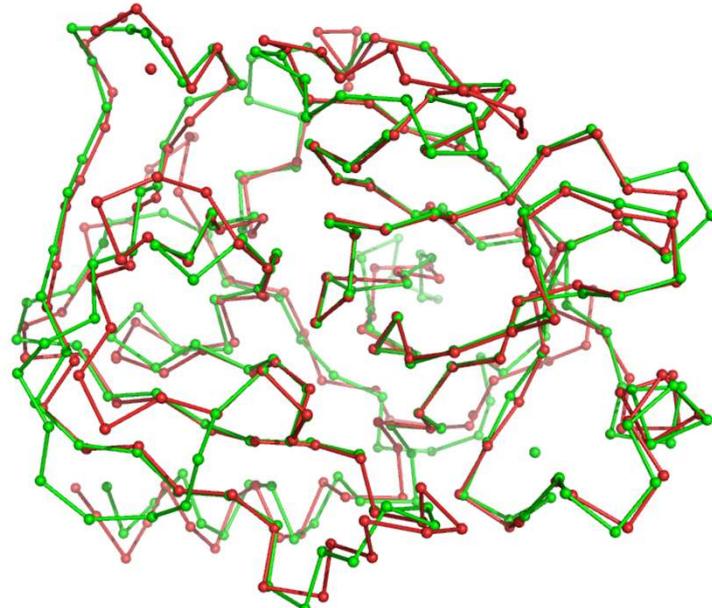
Tripsina humana
versus
Tripsina bovina



RMSD 0.8 Å

40% identidade de sequência

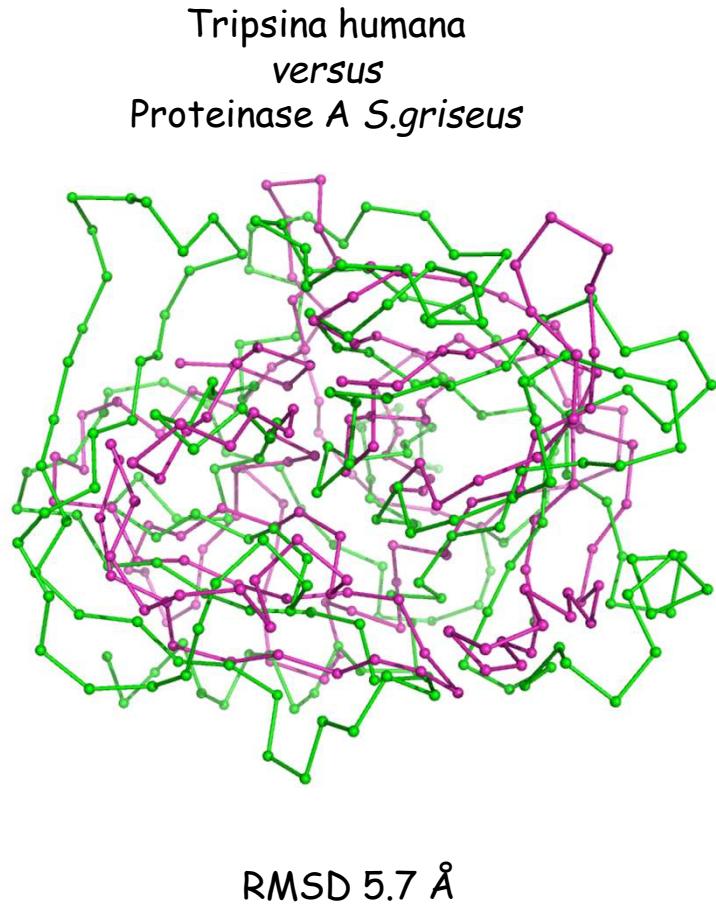
Tripsina humana
versus
Tripsina S.griseus



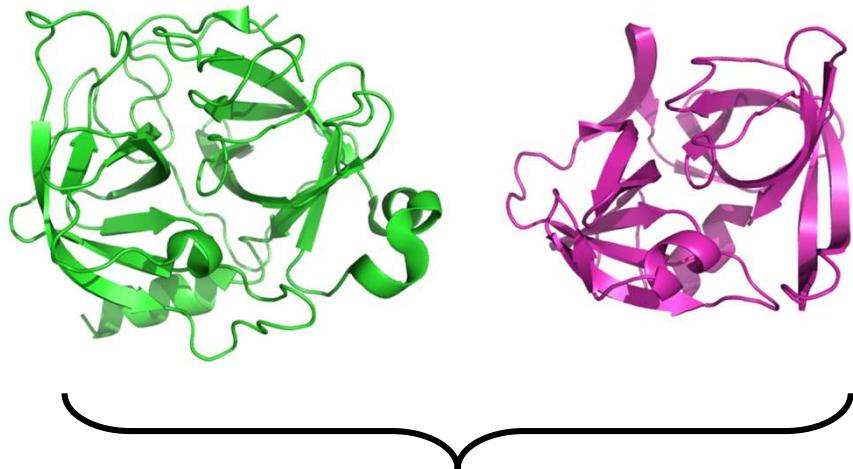
RMSD 1.8 Å

34% identidade de sequência

Relação entre RMSD e identidade de sequência



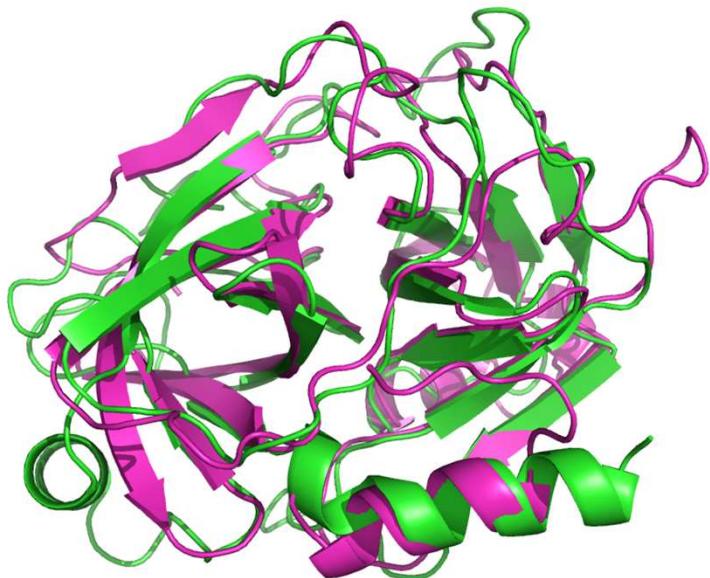
20% identidade de sequência



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

Relação entre RMSD e identidade de sequência

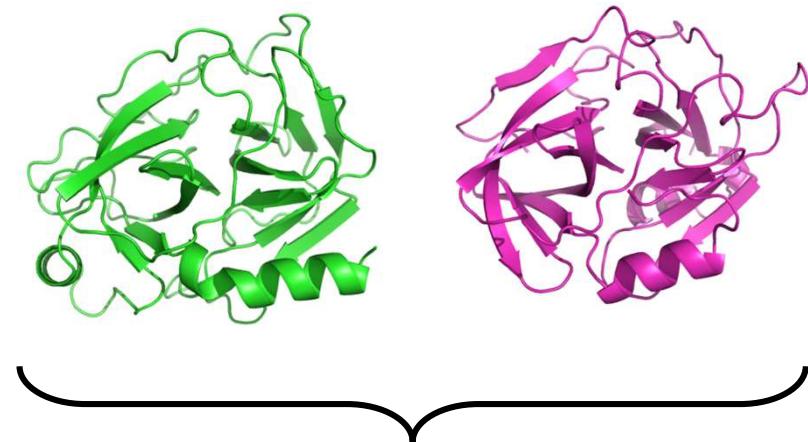
Tripsina humana
versus
Proteinase V8 *S.aureus*



RMSD 2.5 Å

19% identidade de sequência

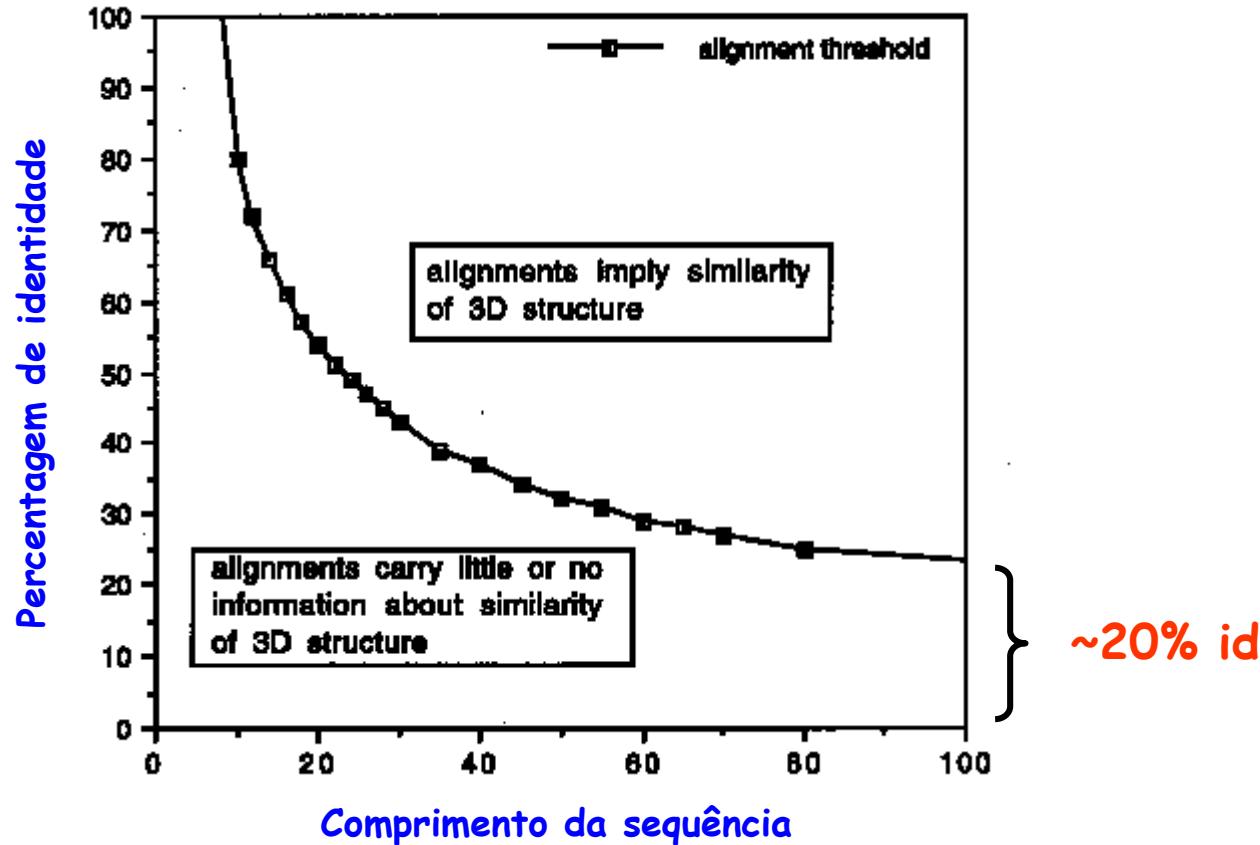
E-value: 8.6×10^2



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

PDB files: 2RA3, 1WCZ

Relação entre RMSD e identidade de sequência

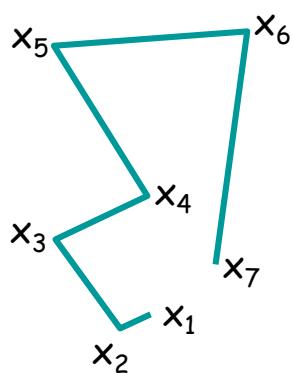


A relação entre a percentagem de identidade e a similaridade estrutural das proteínas depende do comprimento do alinhamento!

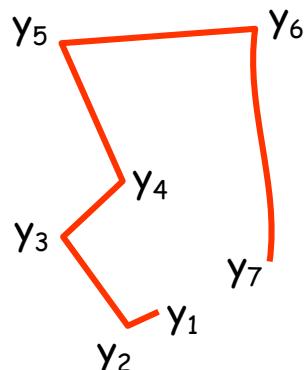
Para identidades inferiores a 20% não é, em geral, possível inferir existência de similaridade estrutural com base no alinhamento das sequências.

Alinhamento estrutural

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



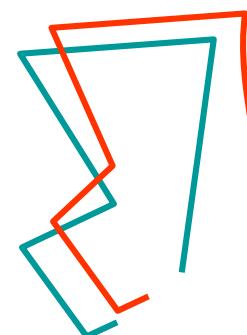
+



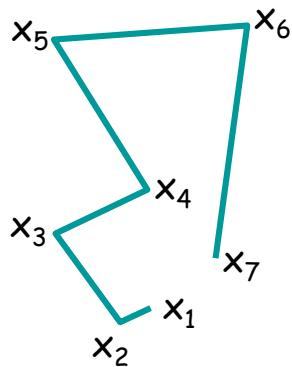
Fácil



a correspondência
é óbvia



$$\begin{aligned}x_1 &\equiv y_1 \\x_2 &\equiv y_2 \\x_3 &\equiv y_3 \\x_4 &\equiv y_4 \dots \\x_7 &\equiv y_7\end{aligned}$$



Difícil!



a correspondência
NÃO é óbvia

?

Sites para comparação e pesquisa estrutural

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

P - pesquisa

C - comparação

M - alinhamento múltiplo

Pesquisa de estruturas: DALI server

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

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

ekhidna2.biocenter.helsinki.fi/bar X +

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

Results:

Chain: 2lztA

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

Results will be deleted after one week.

Dali: 2lztA Dali alignment: 2lztA

Not secure | ekhidna2.biocenter.helsinki.fi

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	-KVFGRCLEAAAMKRHGLDNYRGYSLGNIWVCAAKFESNFNTQATNRNTDGSTDYGIQLQINSRWWCNDGRTPGSRNLNCNIPCSALLSSDITASVNC	AKKIVSDGN	GMNAWAVRNRC	KGTDVQA	WIRGCRL	-
0002 5vasA	-KVYSRCLEAAAMKRLGLDNYRGYSLGNIWVCAANYESGFNTQATNRNTDGSTDYGIQLQINSRWWCDDGKTPRSKNACGIRCSVLLRSDITEAVRC	AKRIVRD	GNGMNAWAVRNRC	RGTDVSKWIRGCRL	-	
0003 1lsgA	mKVFGRCLEAAAMKRHGLDNYRGYSLGNIWVCAAKFESNFNTQATNRNTDGSTDYGIQLQINSRWWCNDGRTPGSRNLNCNIPCSALLSSDITASVNC	AKKIVSDGN	GMNAWAVRNRC	KGTDVQA	WIRGCRL	qgghliggakqagdv
0001 2lztA	-LLLHHHHHHHHHHHHHL	LLLHHHHHHHHHHHHHH	LLLHHHHHHHHHHHHHH	LLLHHHHHHHHHHHHHH	LLLHHHHHHHHHHHHHH	-
0002 5vasA	-LLLHHHHHHHHHHHL	LLLHHHHHHHHHHHHHL	LLLHHHHHHHHHHHHHL	LLLHHHHHHHHHHHHHL	LLLHHHHHHHHHHHHHL	-
0003 1lsgA	1LLLHHHHHHHHHHHL	LLLHHHHHHHHHHHHHL	LLLHHHHHHHHHHHHHL	LLLHHHHHHHHHHHHHL	LLLHHHHHHHHHHHHHL	1111111111111111

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 a web browser window with the URL www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi. The page is titled "VAST+ Similar Structures" and describes the tool for identifying macromolecules with similar 3D structures. A search bar at the top contains the text "2LZT" (highlighted with a red box and arrow). Below the search bar, a paragraph explains how to use VAST+ by entering a PDB ID or MMDB ID. At the bottom, there is a section titled "Citing VAST" with two references. The browser's address bar also shows "RCSB PDB - Structure" and "RCSB Protein Data B".

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

Input a valid PDB ID or MMDB ID:

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

Citing VAST

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

| Help Desk | Disclaimer | Privacy statement | Accessibility |

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

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

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

NCBI
National Center for Biotechnology Information

VAST+ Similar Structures

3D structural similarities among biological assemblies

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

PDB ID or MMDB ID New Search

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

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

Similar Structures Original VAST ▾

▼ Display filters ▾

Showing 1 to 10 out of 860 structures ▾

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

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

Citing VAST

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

VAST+ Similar Struct x

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

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

NCBI National Center for Biotechnology Information

VAST+ Similar Structures 3D structural similarities among biological assemblies

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

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

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

Similar Structures [?] Original VAST [?]

▼ Display filters [?]

Showing 771 to 780 out of 860 structures [?]

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

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

Citing VAST

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

VAST+ Similar Struct x

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

File View Select Style Window CDD Help

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

Aligned Sequences Close

Visualize 3D structure superposition with Cn3D

2LZT_A: HEN EGG WHITE LYSOZYME
3B0O_A: ALPHA-LACTALBUMIN

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

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

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

*Click schematic circles and molecule names to view matches

775 [+] ● 1FKQ
776 [-] ● 3B0O

Visualize 3D structure superposition with Cn3D View aligned sequences

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

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

The screenshot shows the VAST+ software interface. On the left, a window titled '2LZT neighbors - Sequence/Alignment V' displays sequence alignments between 2LZT_A (Hen Egg White Lysozyme) and 3B0O_A (Alpha-Lactalbumin). It includes a sequence viewer with numbered positions 1 through 100, color-coded amino acids, and a 'View' menu. Below the viewer are two entries: 775 (1FKQ) and 776 (3B0O). On the right, a larger window titled '2LZT and 3B0O sequence alignment - Google Chrome' shows the aligned sequences. The top part of the alignment table has columns for positions 10, 20, 30, 40, 50, and 60. The bottom part has columns for 70, 80, 90, and 100. A 'Close' button is at the top right of this window. At the bottom of the alignment table, it says 'Query structure MMDB ID: 58091 (PDB ID: 2LZT)' and 'Matched structure MMDB ID: 100429 (PDB ID: 3B0O)'. Below the alignment table, there's a note: '*Click schematic circles and molecule names to view matches'. Two schematic molecules are shown: one for 'HEN EGG WHITE LYSOZYME' (a central purple circle with four white diamond branches) and one for 'ALPHA-LACTALBUMIN' (a single purple circle with a single white diamond branch). Below these molecules are two buttons: 'Visualize 3D structure superposition with Cn3D' and 'View aligned sequences'. At the very bottom, there are two more entries: 777 (1NMM) and 778 (1A2Y), each with its own set of details.

Pesquisa estrutural com (original)VAST

Vast Neighbor Summary RCSB PDB - Structure RCSB Protein Data Bank

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

Apps Enzymology Piano Music Production Bioinformatics Databases Bioinformatics Tools Misc Programming

NCBI VAST Similar Structures

PubMed BLAST Structure Taxonomy OMIM Help? Cn3D

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

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

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

View Sequence Alignment using Hypertext for Selected VAST related structures

List Medium redundancy subset, sorted by Aligned Length in Graphics

Advanced related structure search

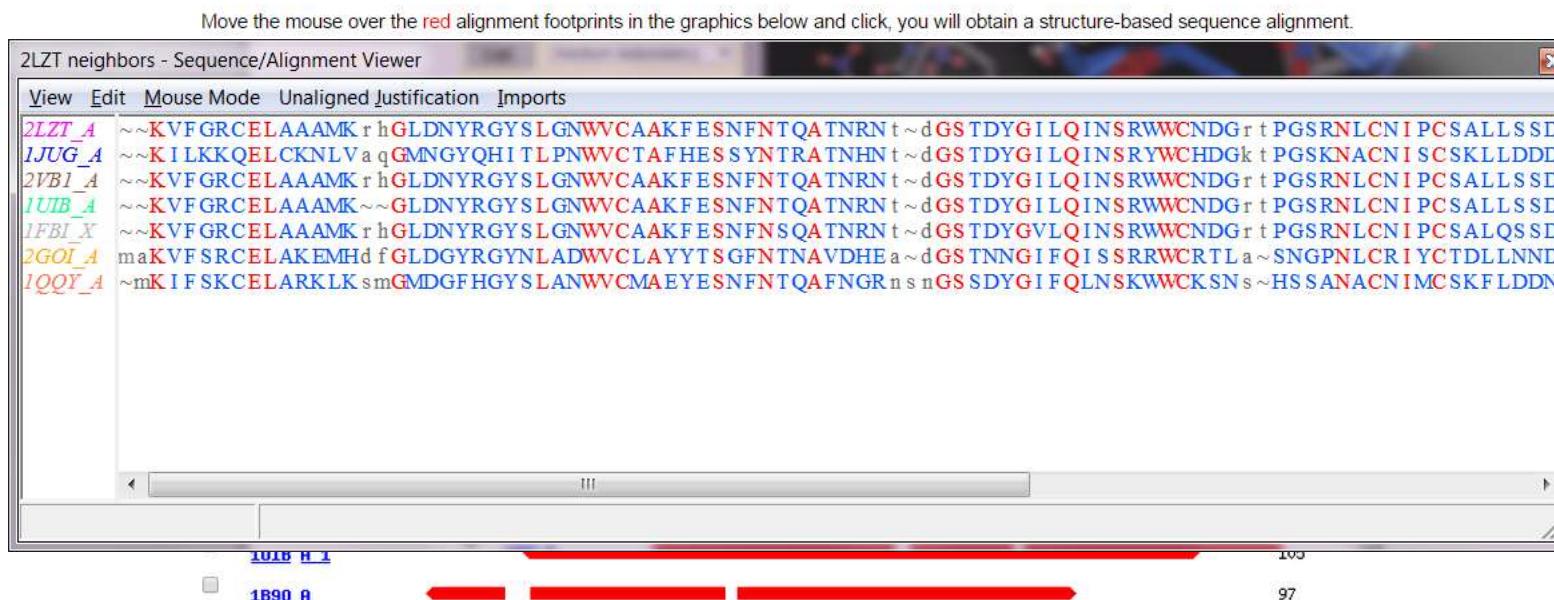
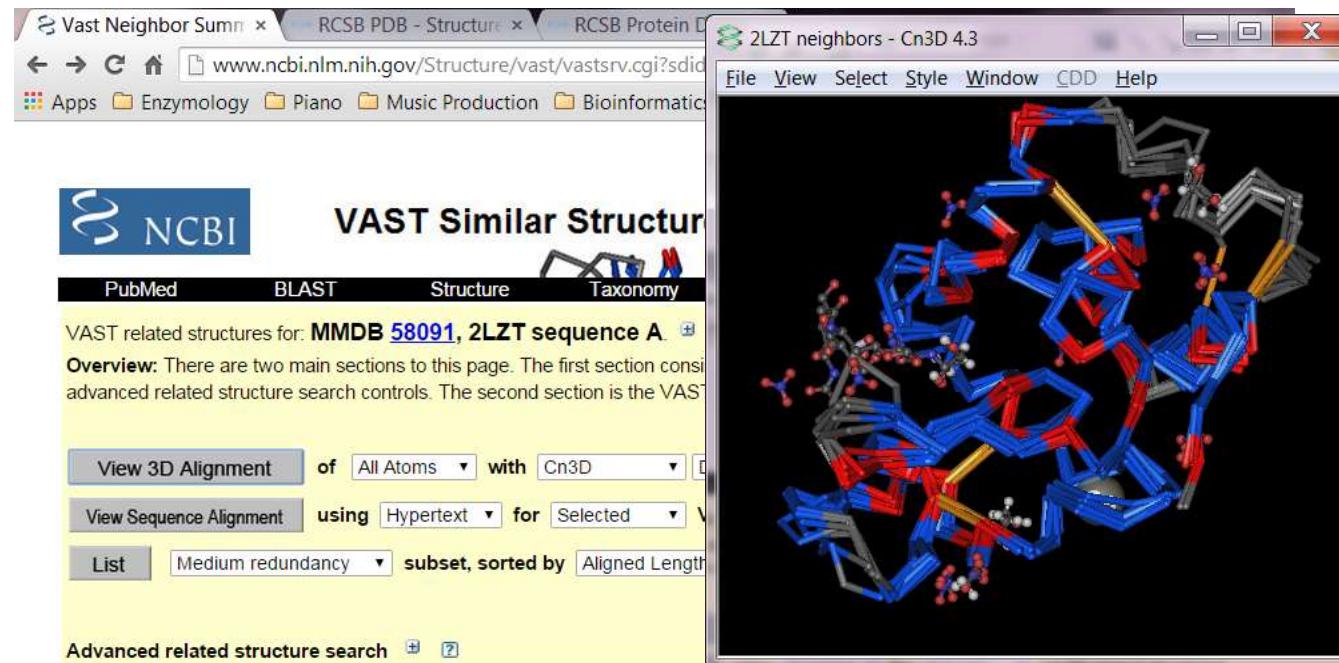
Move the mouse over the red alignment footprints in the graphics below and click, you will obtain a structure-based sequence alignment.

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

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

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

Visualização do alinhamento com o software Cn3D



Previsão da estrutura secundária das proteínas

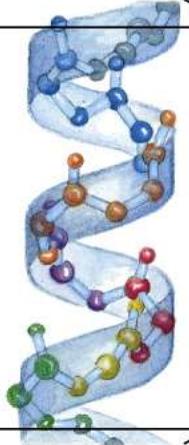
Níveis de organização da estrutura das proteínas

Estrutura primária



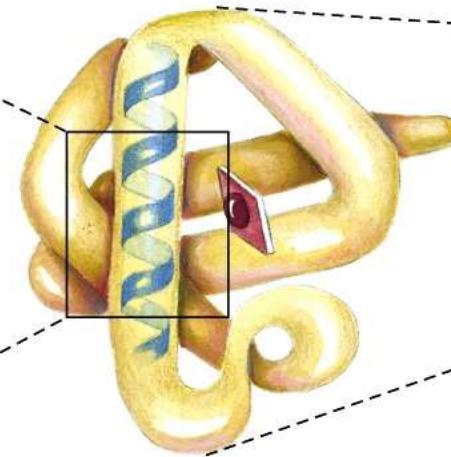
Sequência de aminoácidos

Estrutura secundária



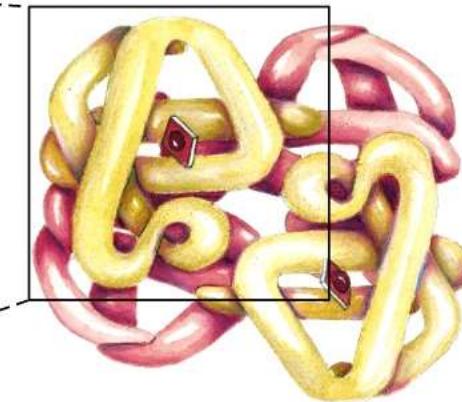
α -hélice

Estrutura terciária



Cadeia polipeptídica

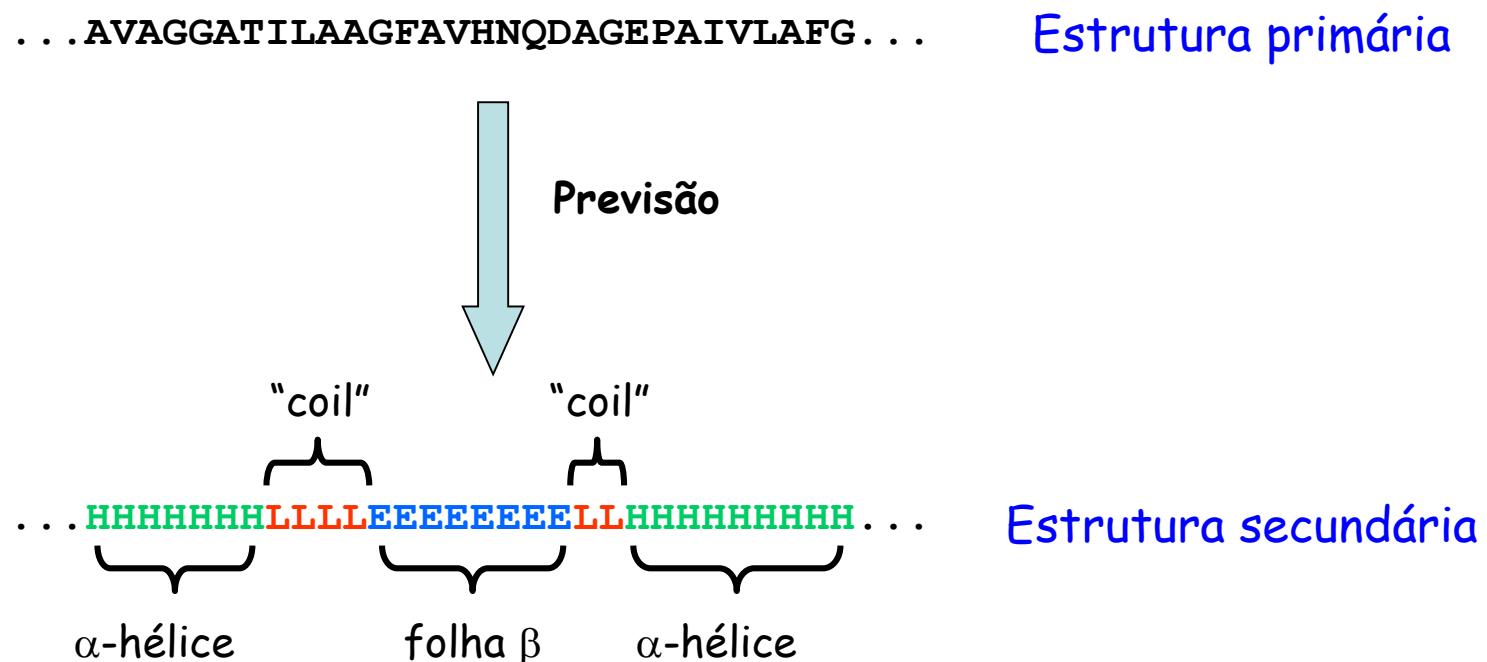
Estrutura quaternária



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.



Métodos de previsão da estrutura secundária

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

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

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

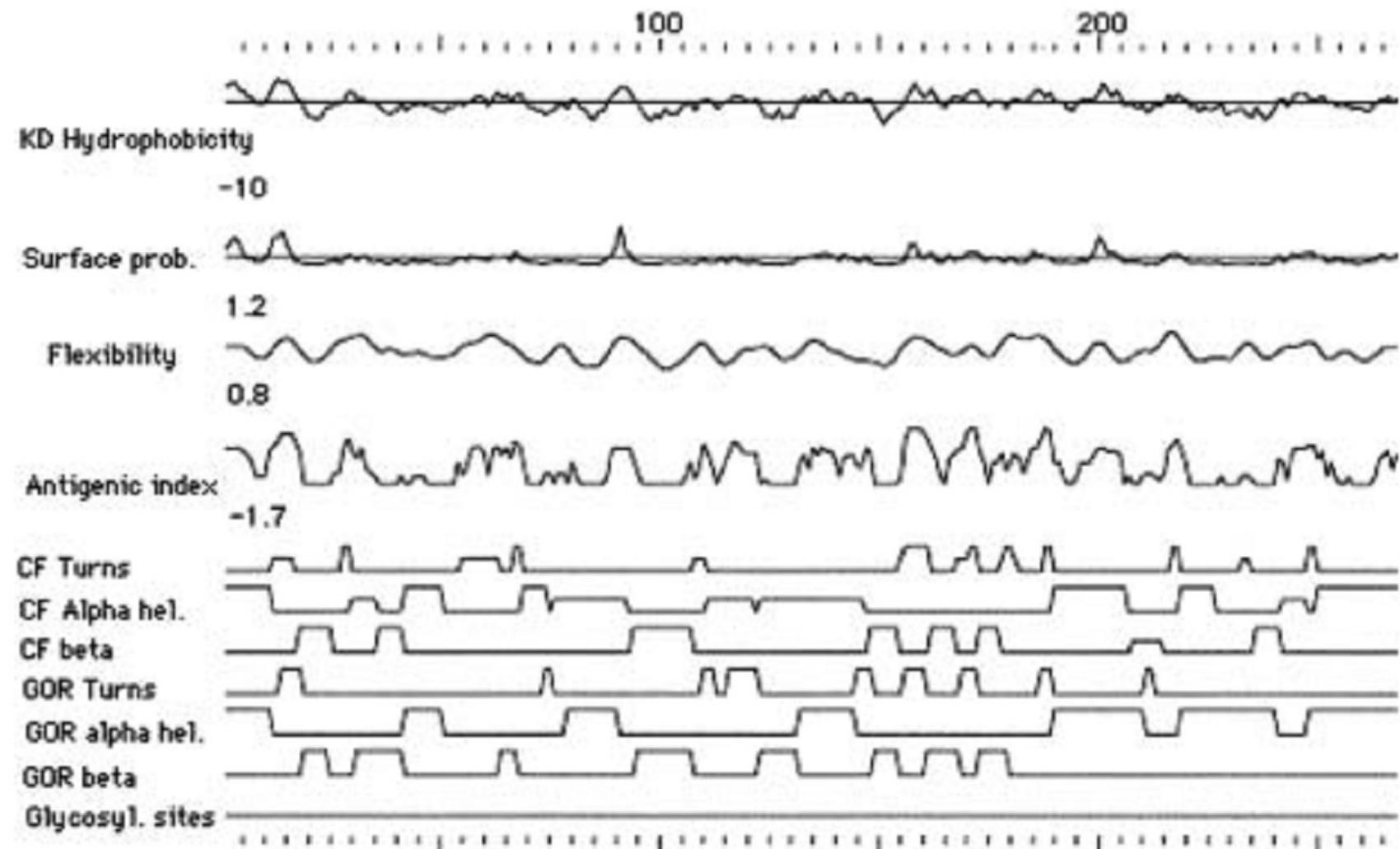
- **Nearest-neighbor** - este método baseia-se na comparação da sequência a prever com sequências de estrutura conhecida. (Precisão: 70-75%)

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

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

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

Previsão GOR e Chou-Fassman com o programa GCG



Exemplo de previsão com o programa PHD

Rel: fiabilidade global da previsão (0-9)

```
.....1.....2.....3.....4.....5.....6
AA |MERYENLFAQLNDRREGAFVPFVTLGDPGIEQSLKIIDTLIDAGADALELGVPFSDPLAD|
PHD | HHHHHHHHHHHH EEEEEEEE HHHHHHHHHHHHHH EEEE |
Rel |93489999999634887279984248998458799999997399668944767784689|
detail:
prH-|03689999998753110000000000001678899999998300000000001113210
prE-|0000000000000000579886530000000000000000000017886311000000
prL-|9631000001236888310012368998321100000001699720036877886789
subset: SUB |L..HHHHHHHHH..LLL.EEEE...LLL.HHHHHHHHHHHH.LLEEE..LLLL.LLL|


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

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

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

prL: probabilidade do resíduo estar em conformação de "coil" (0-9)

Modelação da estrutura terciária por homologia

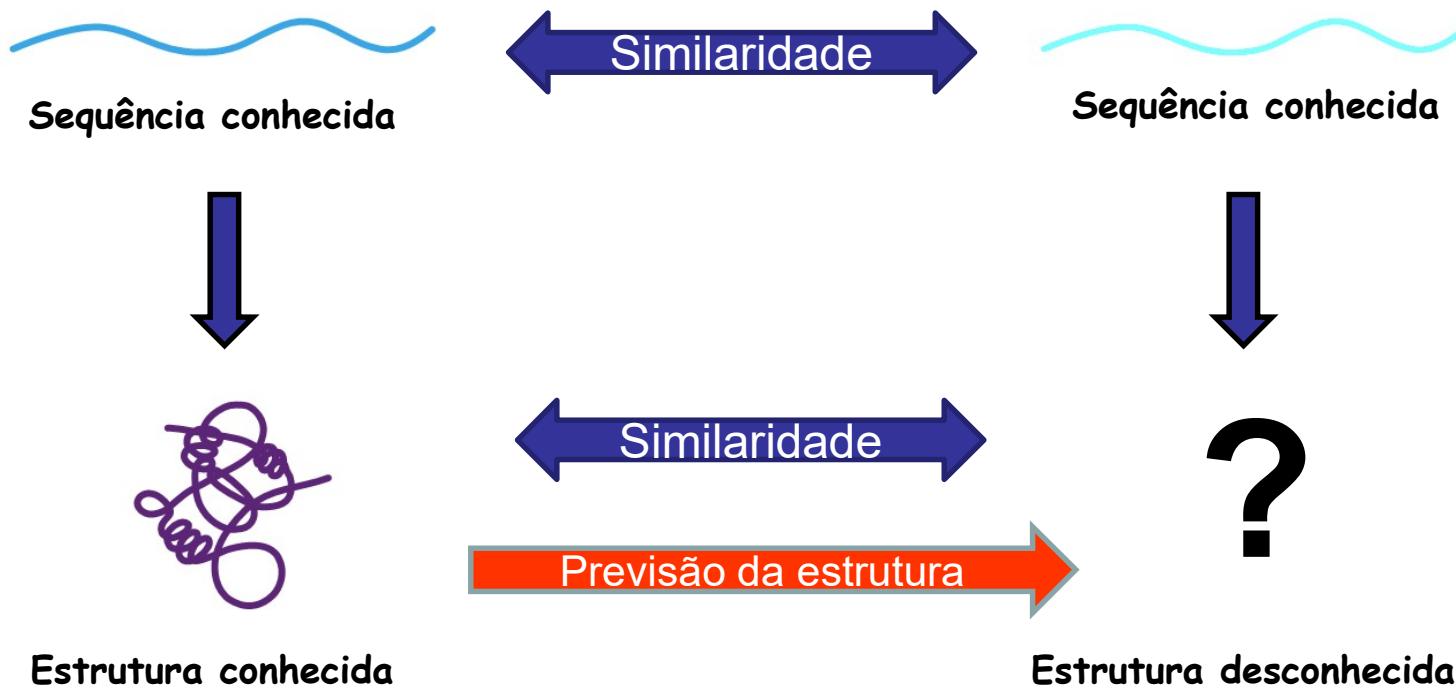
Modelação por homologia

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

Fundamento da Modelação por homologia:

A conservação da sequência está associada à conservação de estrutura!

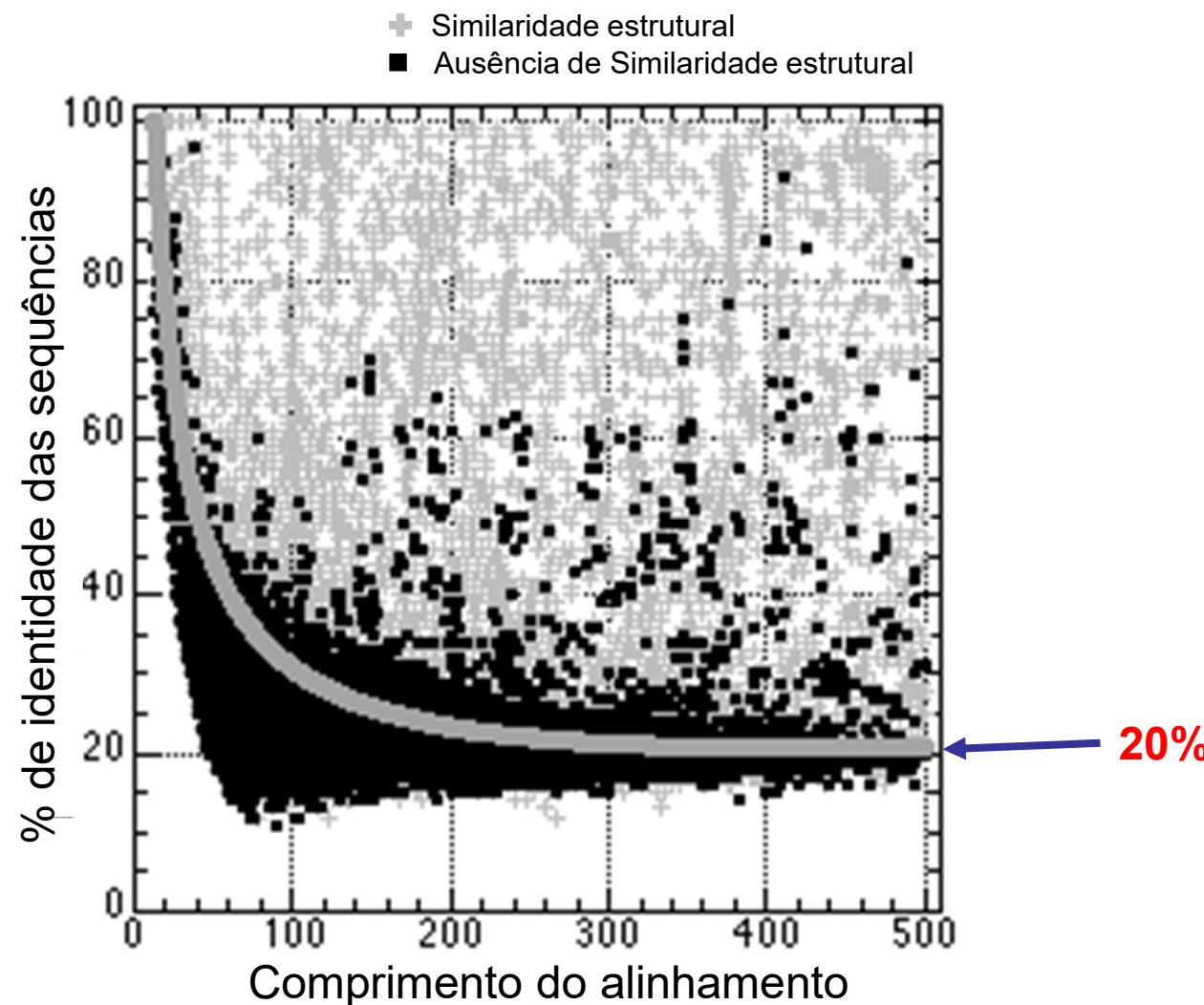
A estrutura das proteínas é determinada pela sua sequência



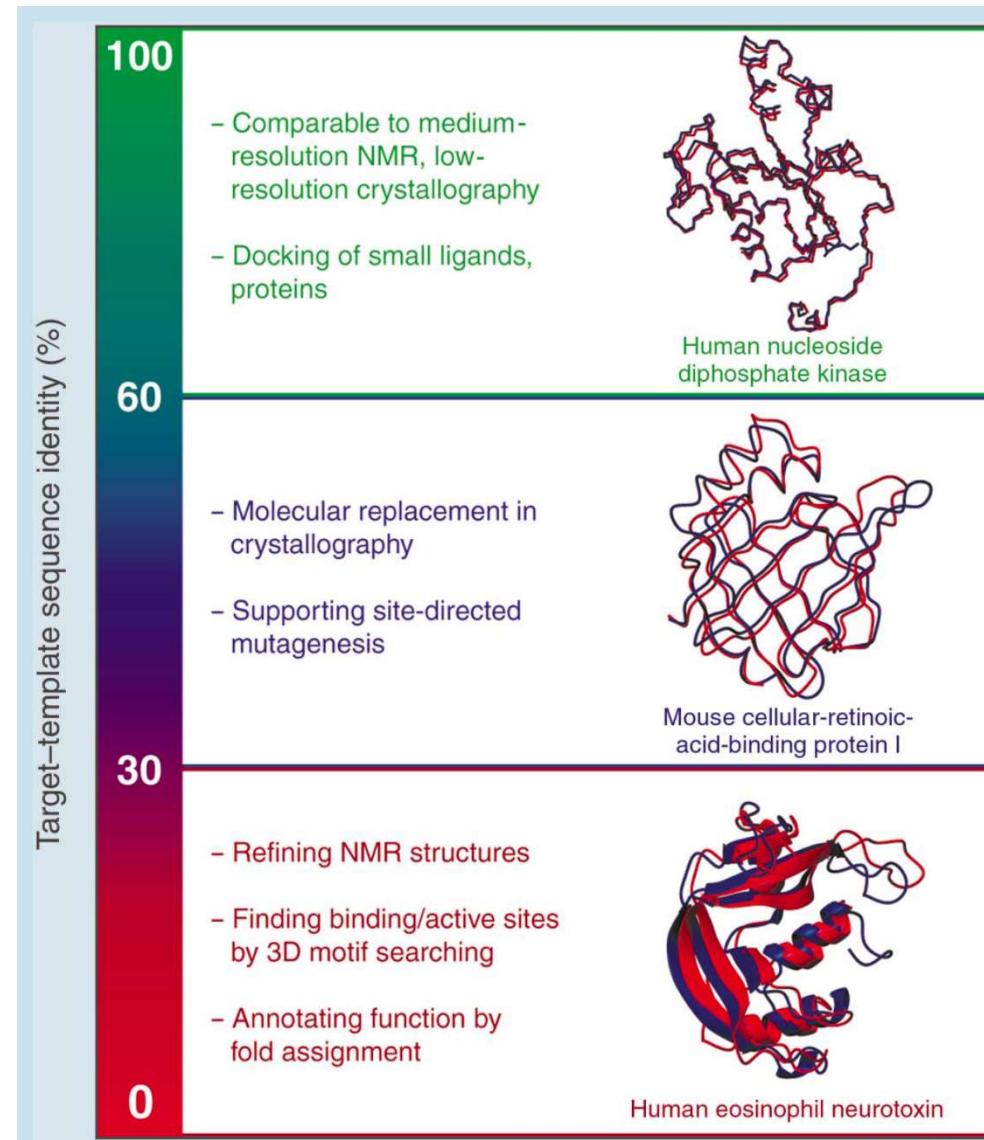
Sequências similares implicam estruturas similares, logo:

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

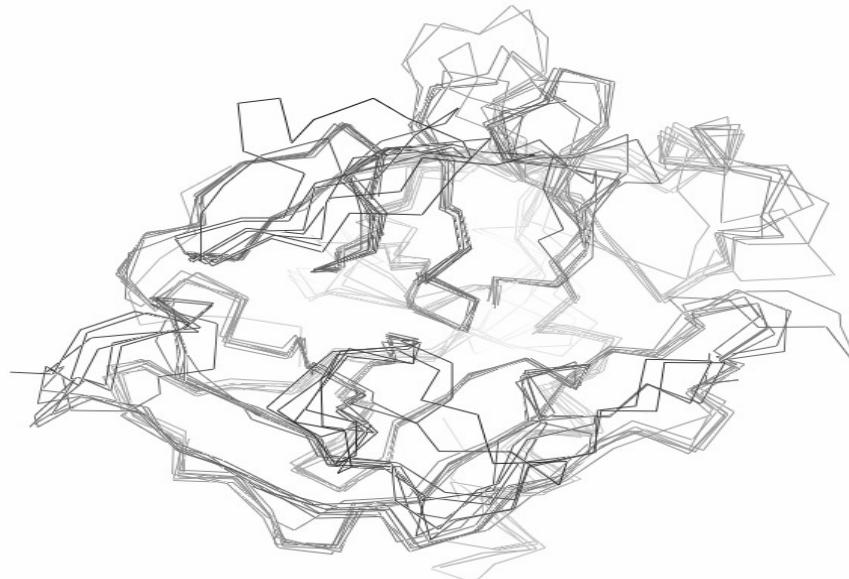
Qual a % de identidade mínima aceitável para existência de similaridade estrutural ?



Impacto da similaridade na qualidade e utilidade do modelo



Alinhamento estrutural das proteases de serina



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

```

IVGGYRRPAPPHAWPFMPLSQLL--A          GGHFGKATLAPTFNVMSSAHCYV-      ANVNYNR - ADRVVLGAGNLNSRPPR-TRD
IVGGYCTGANGVYFQVOVSLSN-   G----- YHHCFGGLSINSQWVSAHCY-      --KS - G IGYFLRGEDDNINVYBEG-N
IVEGGQDSEAVSGLWQVQLSVLL- L-G----- EHGFGGSNSLQVWVSAHCY-      VTL-S DSVVAGFCEDQCSSEK-V
IVEGGQDSEAVSGLWQVQLSVLL- E----- ELLCGAGLSDRWWLTTAAHCY-      FTVdDLVLRIGKHSRTYRVRK-V
IVEGGQDSEAVSGLWQVQLSVLL- kspQ-    ELLCGAGLSDRWWLTTAAHCY-      FTVdDLVLRIGKHSRTYRVRK-V
IINGYBAYTGLFPYQAGLDD-   Tlqddr- RYVWGGSLLIDRWLTTAAHCY-      AV- SSVVYLGSANQYV- -GG
IINGYBAYTGLFPYQAGLDD-   -SQPWQAVAVIN- YLGCGVLLIDPVWTTAAHCY-      SN- NYQVLLGCNLRFEDP-FAC

```

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

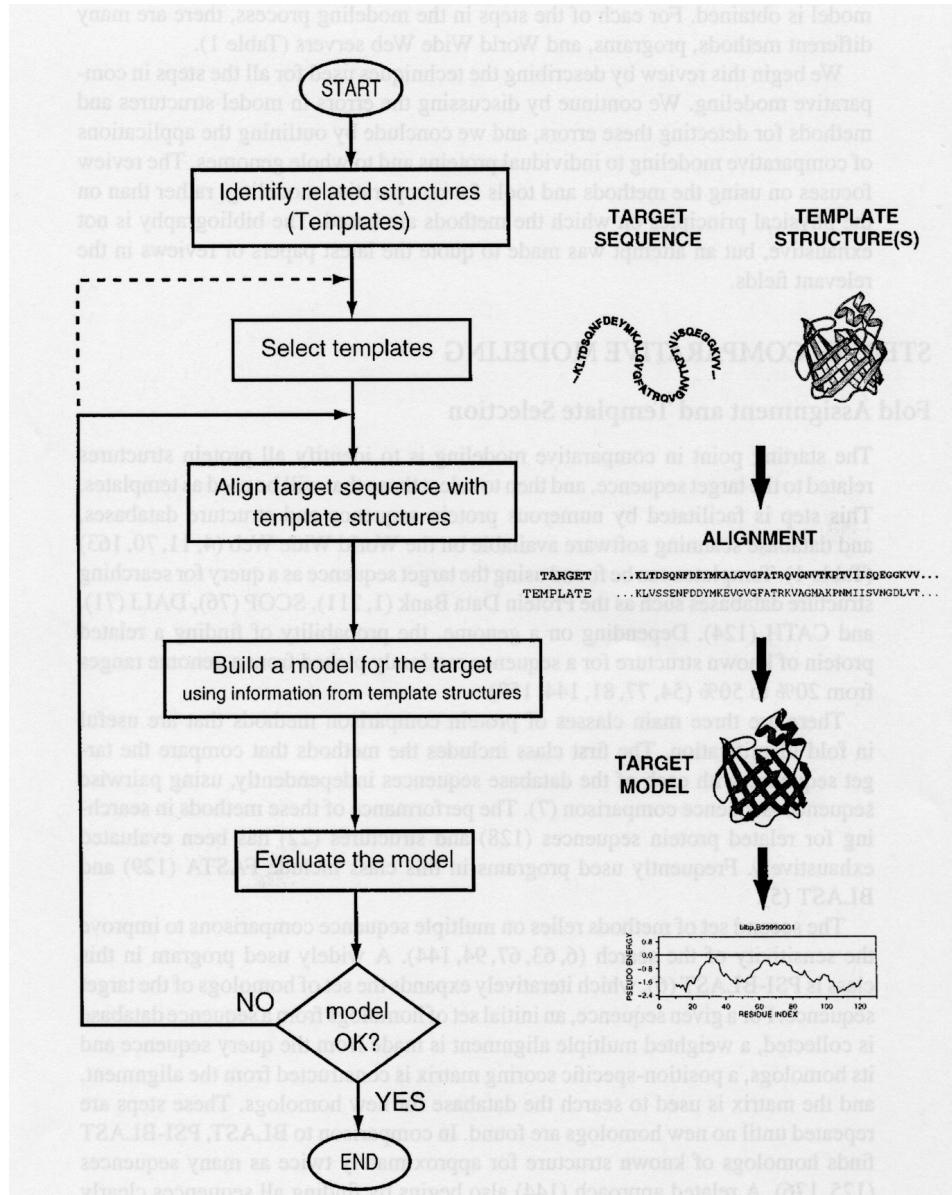
1BOF:**A** (218:1-218)
3ETH:**E** (223:1-221)
1GMD:**A** (241:14-239)
1HSD:**H** (260:1-248)
1TOC:**B** (259:1-255)
1HYL:**A** (230:1-227)
1TON:_ (235:11-232)

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

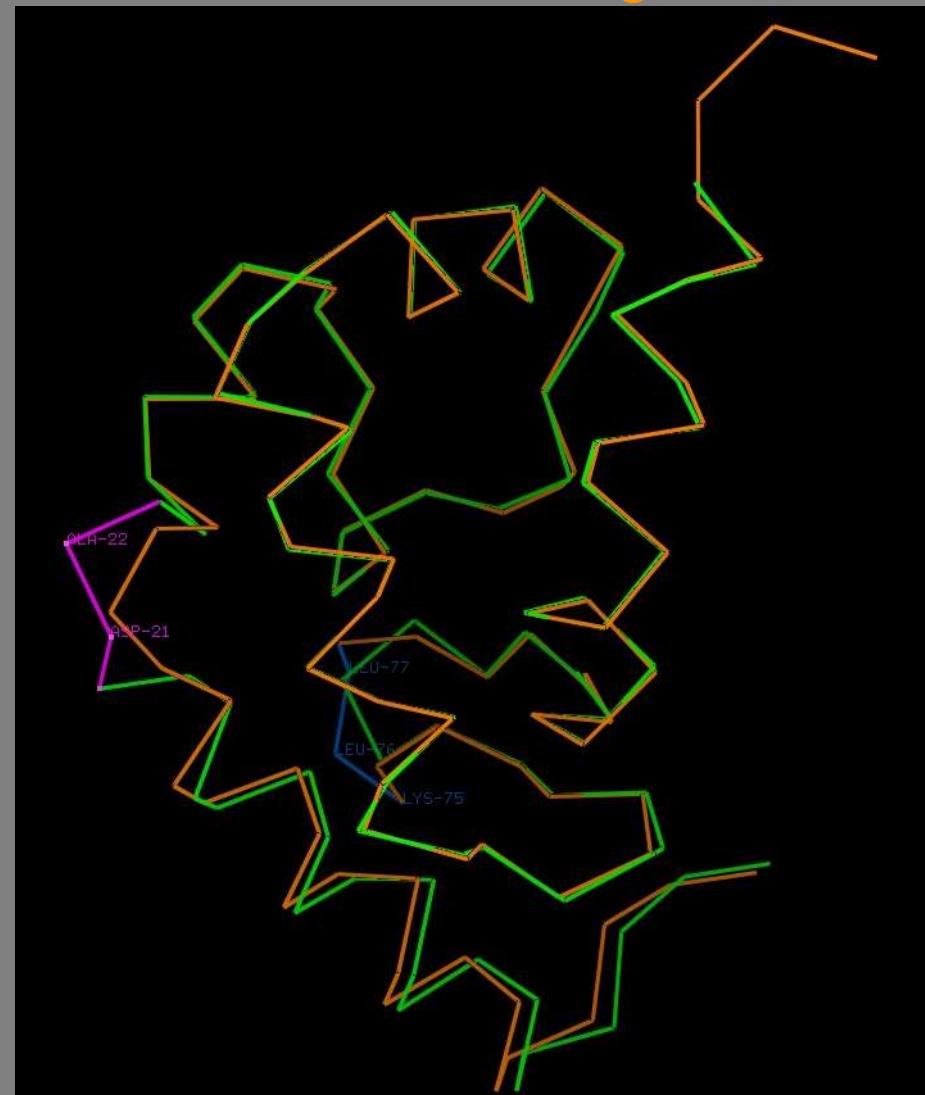
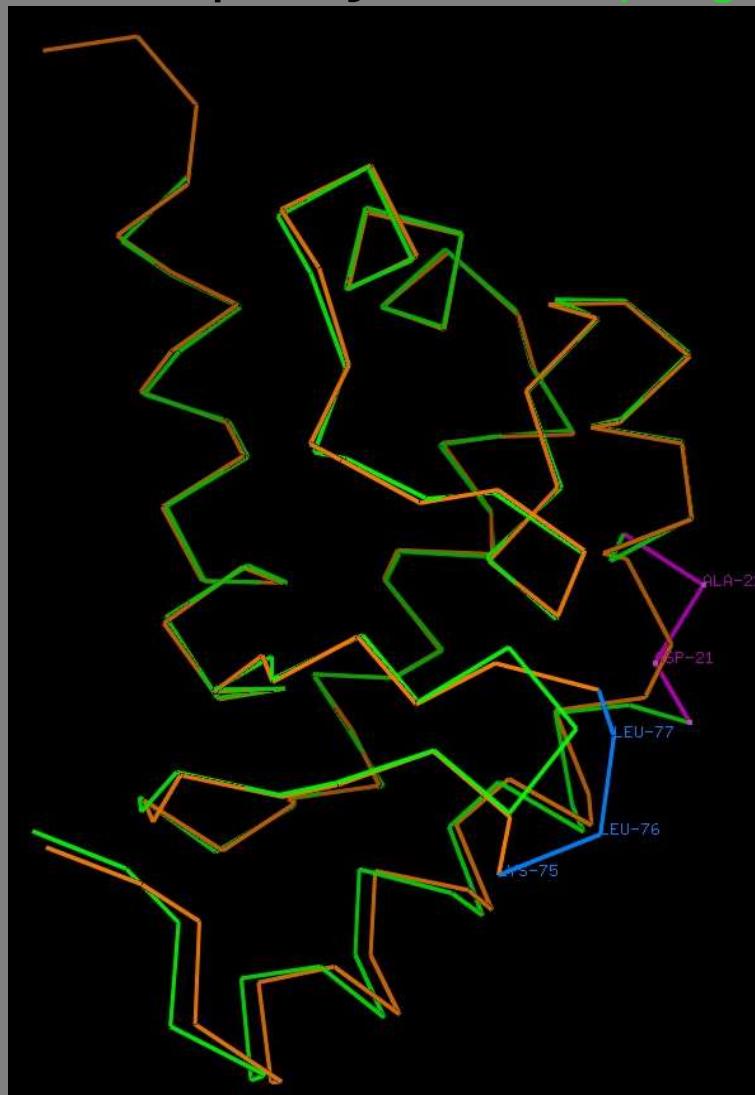
SFVR--GG-CASGLYPDGFAPVVAQEVNWIDSIIIG
SWGS---G-CAQKNNKPGVYTKCVCNYSWIKQIQTIA
SWGS---S-TCTSTSPTGVYVARVTLANVWNSQQTLLA
SWG-----C-CDRDGKYGFTYHVFRLKKWIQKVID
SWG-----G-CDRDGKYGFTYHVFRLKKWIQKVID
SFVSGa---G-CESG-KPVGFSRTVSYMDWIQQNTG
SGGA----TcCAKPKTPAYIYAKLKFRTSWIJKVMV

Passos na modelação por homologia

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



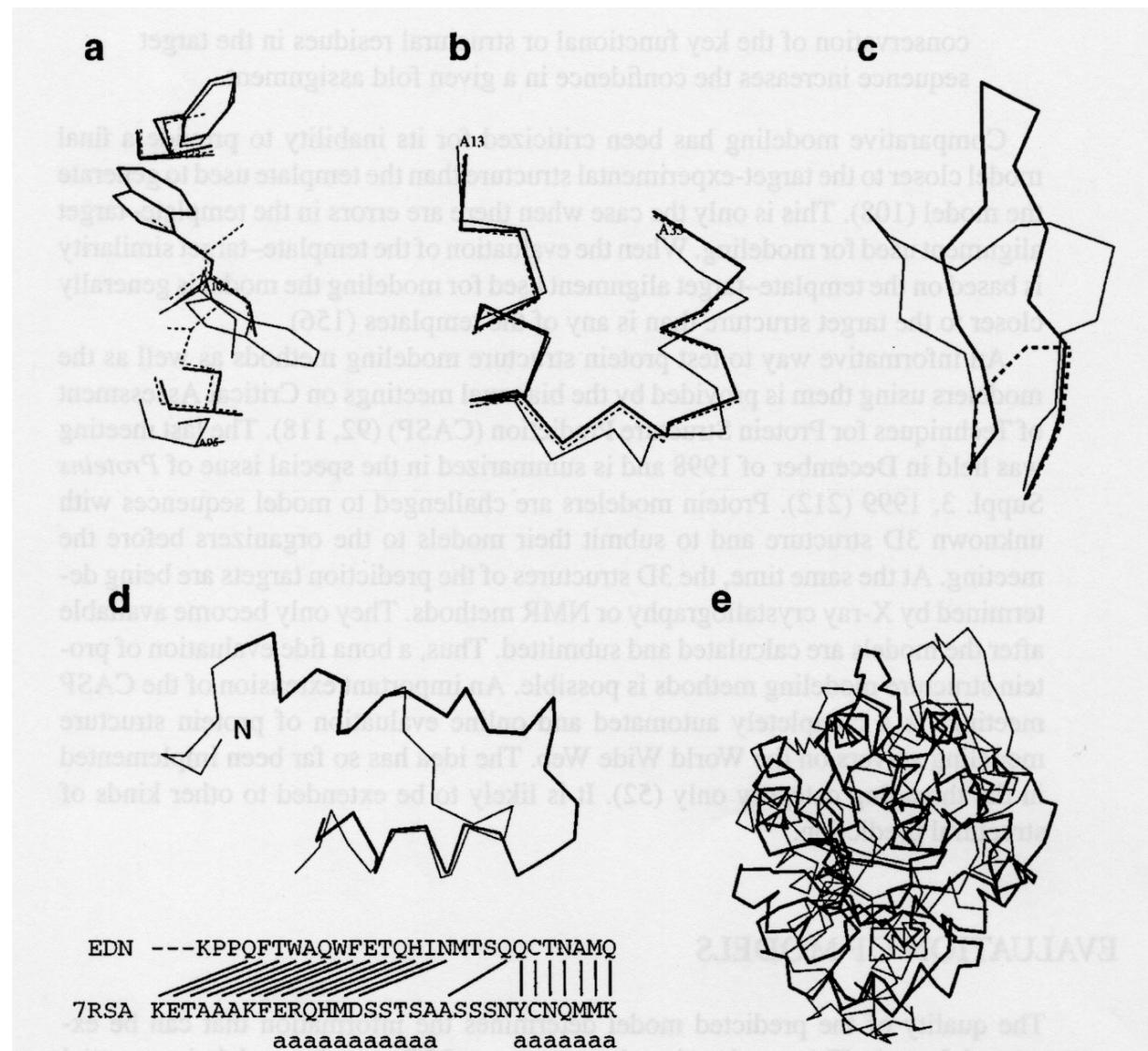
Comparação da criptogeína com o modelo da oligandrina



Erros na modelação por homologia (1)

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

Erros na modelação por homologia (2)



Servidores web para modelação por homologia

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

Bases de modelos pré-calculados

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

SWISS MODEL repository

The screenshot shows a browser window with the title bar "SWISS-MODEL Repository". The address bar contains "swissmodel.expasy.org/repository/". The page itself has a header with the "BIOZENTRUM" logo and "SWISS-MODEL" text. Below the header is a search bar with the placeholder "Search SWISS-MODEL Repository". A section titled "Fetch by UniProtKB AC or Entry Name:" lists "SWISS-MODEL Homology Models: F1P6T8, Q83XK2_ECOLX, B4IFM4, W9KYS2_FUSOX" and "Experimental Structures: AOA0E0UR70, ULA1_HUMAN, P04439, PSA2_YEAST". Another section below says "Or search using free text: Asparagine synthetase A, Heat shock protein HSP 90-alpha, pH-response regulator protein pall/RIM9, Laccase". The top navigation bar includes links for "Modelling", "Repository" (which is highlighted), "Tools", "Documentation", "Log in", and "Create Account". The browser's toolbar is visible at the very top.

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

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

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

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

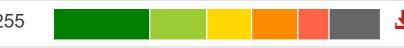
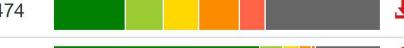
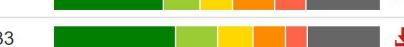
We currently provide models for the **reference proteomes** of the following model organisms, based on UniProtKB release 2019_10. If you want to download a large number of models, please contact us.

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

SWISS MODEL repository

SWISS-MODEL Repository

swissmodel.expasy.org/repository/

	Proteome Size	Sequences Modelled	Models	Seq Coverage	Download Metadata (Models and structures)	Download Coordinates (Homology models)
<i>Homo sapiens</i>	20,659	17,505	43,255		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 a web browser window for the SWISS-MODEL Repository. The address bar contains 'swissmodel.expasy.org/repository/'. The main header includes the BIOZENTRUM logo, the SWISS-MODEL logo, and navigation links for 'Modelling', 'Repository' (which is active), 'Tools', 'Documentation', 'Log in', and 'Create Account'. A red arrow points to the search bar, which contains the UniProt code 'O09185'. Below the search bar, there is a section titled 'Fetch by UniProtKB AC or Entry Name:' followed by a list of homology models and experimental structures. A note at the bottom suggests searching using free text.

UniProt code

O09185

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.

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

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

swissmodel.expasy.org/repository/uniprot/O09185

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

O09185 (P53_CRIGR) *Cricetulus griseus* (Chinese hamster) (*Cricetulus barabensis* griseus)
Cellular tumor antigen p53 ★ UniProtKB[®] InterPro[®] STRING[®]

393 aa; Sequence (Fasta)

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

50 100 150 200 250 300 350

4mzr.1.B Cellular tumor antigen p53

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

Model Quality Estimate

QMEAN	-2.18
C β	0.09
All Atom	-0.70
solvation	-0.84
torsion	-1.86

Sequence Features

Metal binding Site Natural variant
DNA binding InterPro

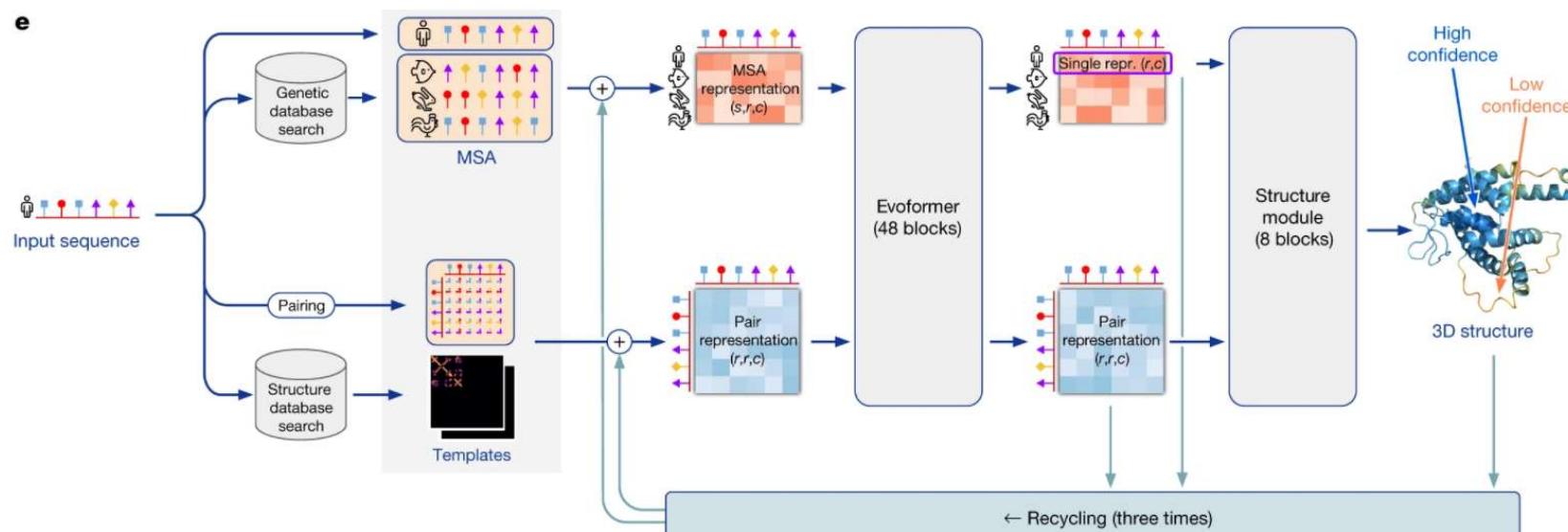
Colours NGI Cartoon

<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.

The image consists of two side-by-side screenshots of the UniProt website. Both screenshots show the entry for PITHD1 (UniProtKB - Q9GZP4 (PITH1_HUMAN)).

Left Screenshot (Entry Page):

- The title is "UniProtKB - Q9GZP4 (PITH1_HUMAN)".
- The protein name is "PITH domain-containing protein 1".
- The gene name is "PITHD1".
- The organism is "Homo sapiens (Human)".
- The status is "Reviewed - Annotation score: 5/5 - Experimental evidence at protein level".
- The "Function" section is expanded, showing annotations for megakaryocyte differentiation, RUNX1 expression, and regulation of transcription.
- The "Structure" checkbox is checked and circled in red.
- The "Keywords" section includes "Molecular Activator".

Right Screenshot (Structure Viewer):

- The title is "PITHD1 - PITH domain-containing protein 1".
- The "Display" tab is selected.
- The "Structure" section is shown, featuring a ribbon model of the protein structure.
- The "Model Confidence" legend indicates confidence levels from "Very high" (blue) to "Very low" (red).
- The "SOURCE" dropdown is set to "AlphaFold".
- The "IDENTIFIER" field shows "AF-Q9GZP4-F1".
- The "METHOD" field shows "Predicted".
- The "RESOLUTION" field shows "1-211".
- The "LINKS" field shows "AlphaFold".
- The "3D structure databases" section lists SMR (Q9GZP4) and ModBase (Search...).