Bioinformática Estrutural

Sequência

Estrutura

Função

Fluxo de informação biológica

Gene ...TTAATAAGT...

⇩ transcrição

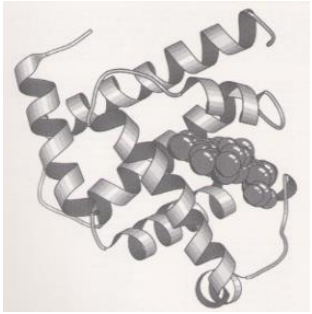
m-RNA ...UUAAUAAGU...

⇩ splicing, tradução

cadeia polipeptídica ...LISVHDN...

⇩ modificações pós-translacionais

proteína



Dogma central da biologia molecular

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

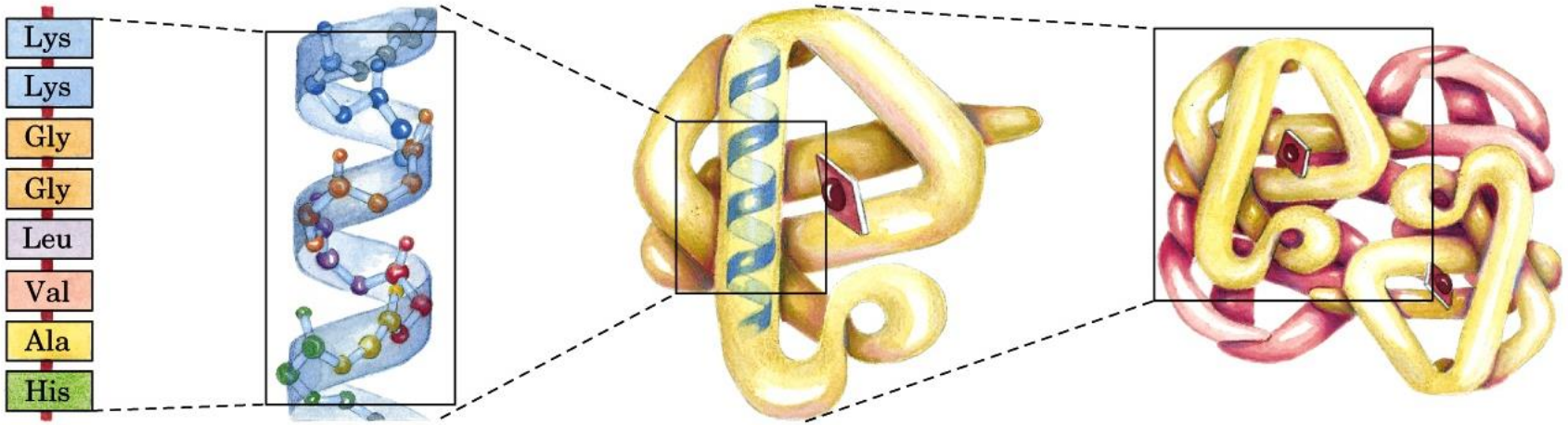
Níveis de organização da estrutura das proteínas

Estrutura primária

Estrutura secundária

Estrutura terciária

Estrutura quaternária



Sequência de aminoácidos

α-hélice

Cadeia polipetídica

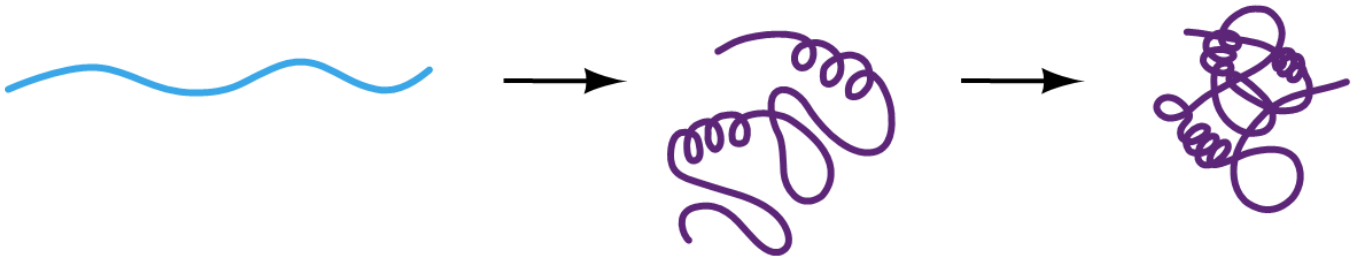
Organização das subunidades

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 polipetídica 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:

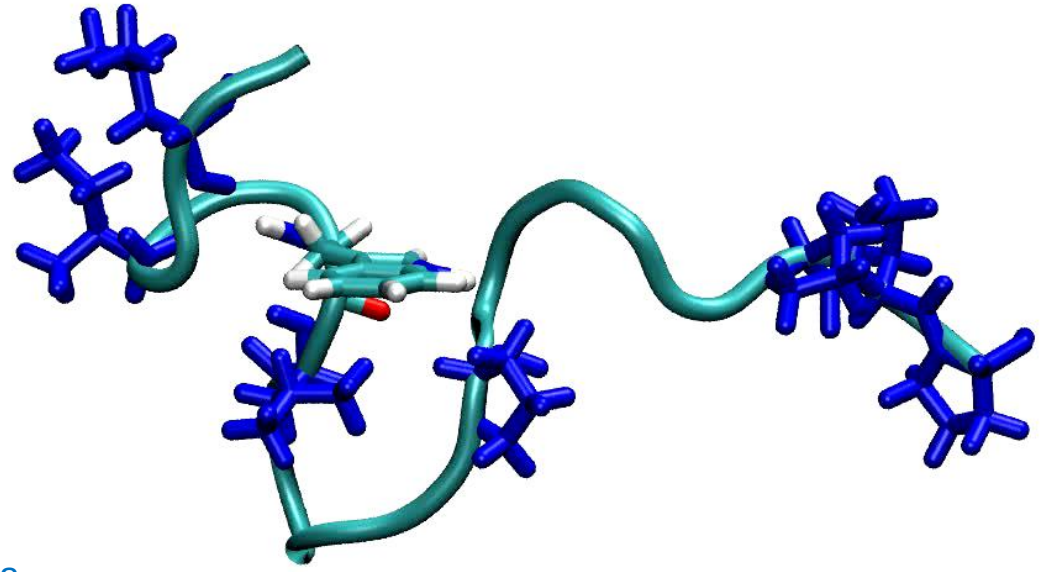


Cadeia polipeptídica linear

Formação de estruturas secundárias: hélices, folhas e loops

Enrolamento da estrutura secundária em três dimensões

Sequência ➞ Estrutura



Simulação do *folding* da mini-protéína **Trp-cage**

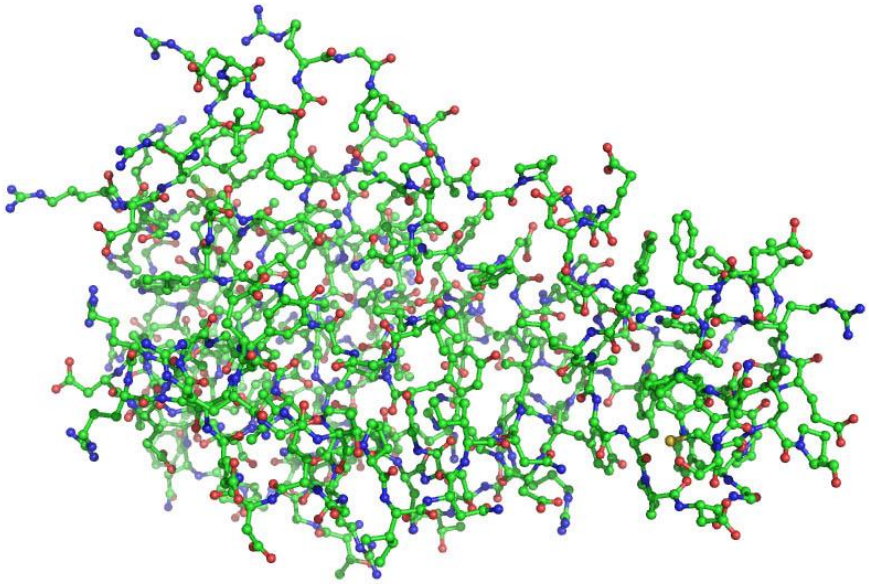
Proteínas mais pequenas ou mais simples frequentemente adquirem a sua estrutura tridimensional *espontaneamente,* sem necessidade de auíxilio de outras proteínas. ,

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 química entre todos os átomos que constituem cada molécula.

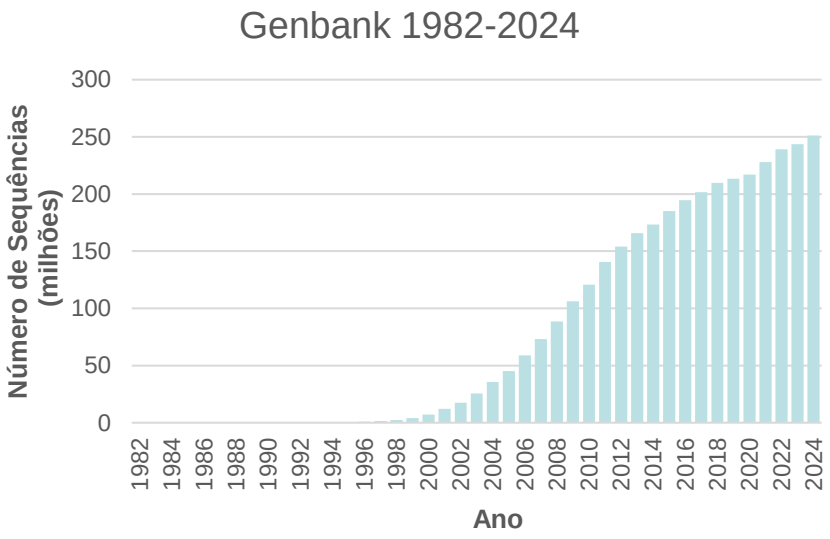
...AVAGGATILVHNQDAGEPAIVLAFG...

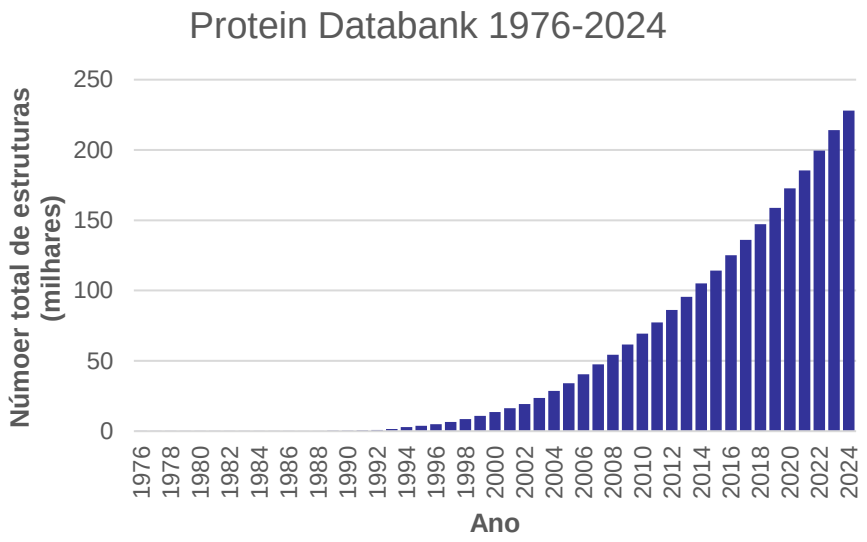
Sequência



Estrutura

Sequência *versus* estrutura



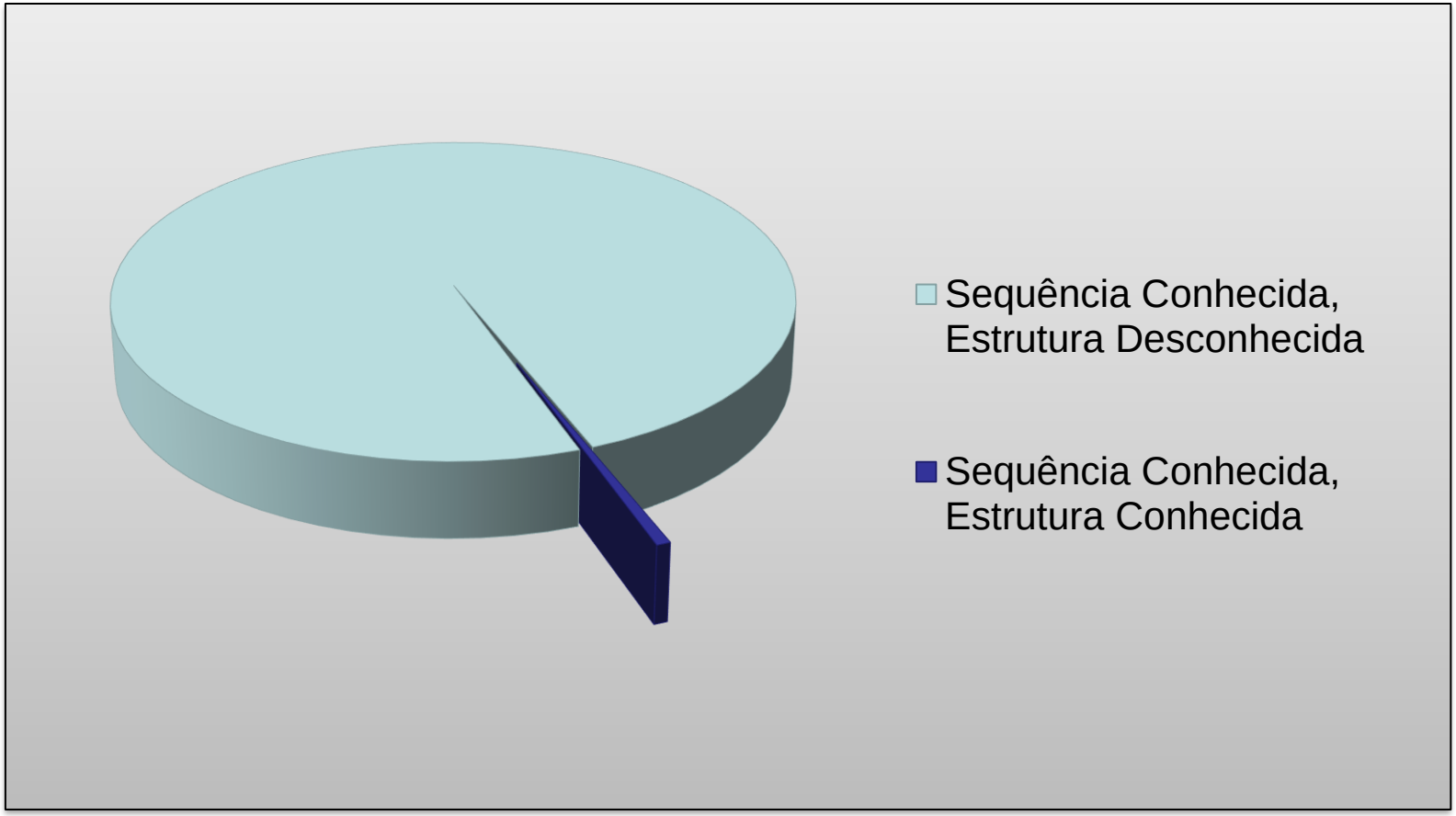


Em 1982: conhecidas 172 estruturas e 606 sequências ...

Hoje (Nov 2024): 227,933 estrutures 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

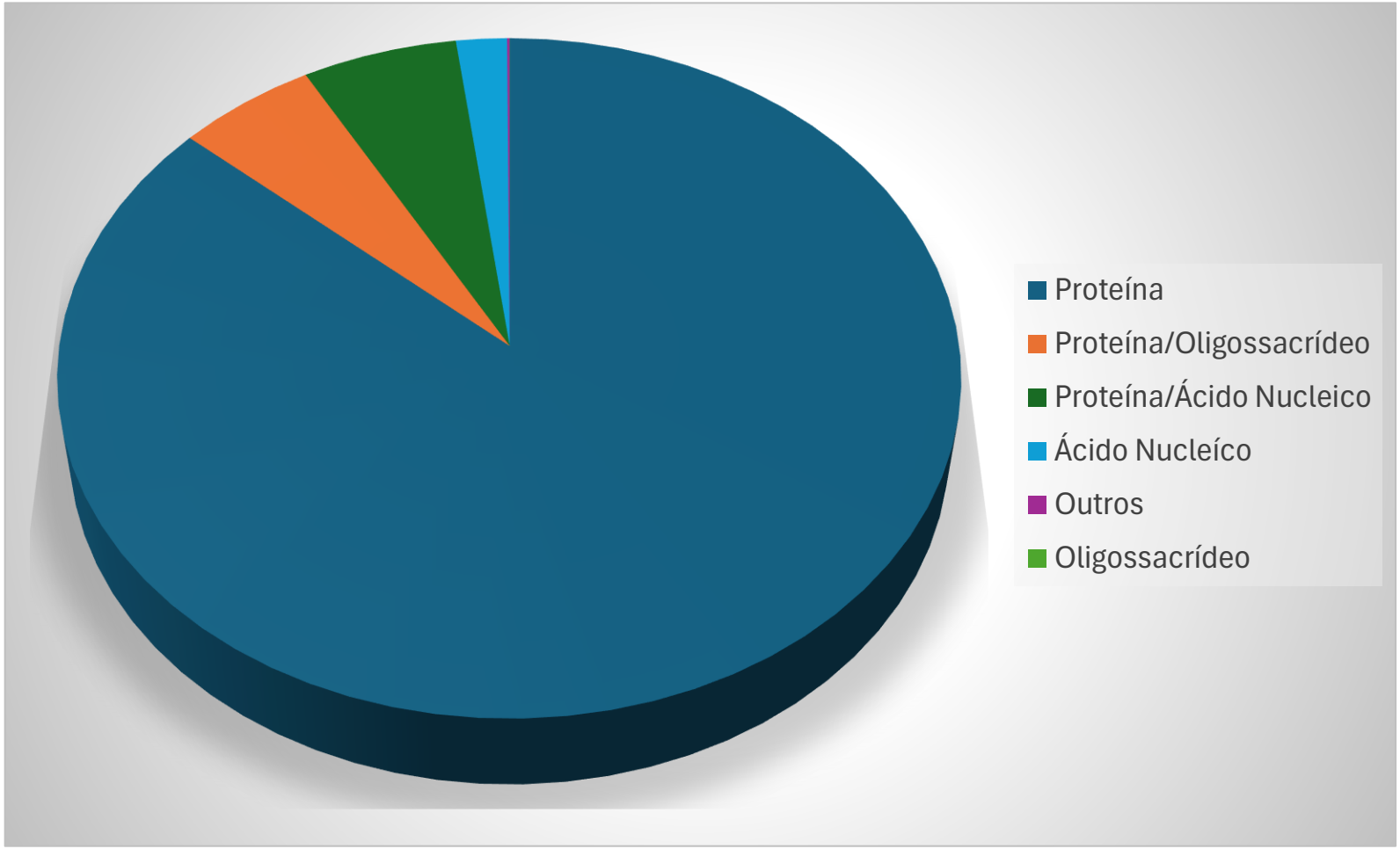
O Protein Data Bank

- 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 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 2/12/2024: **227933**

	Técnica Experimental						
Tipo de Molécula	**X-ray**	**EM**	**NMR**	**Combinação**	**Neutrões**	**Outros**	**Total**
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 Nucleíco	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
Total	189,515	23,645	14,414	238	84	37	**227,933**

Dados de 2/12/2024 em http://www.rcsb.org

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...)

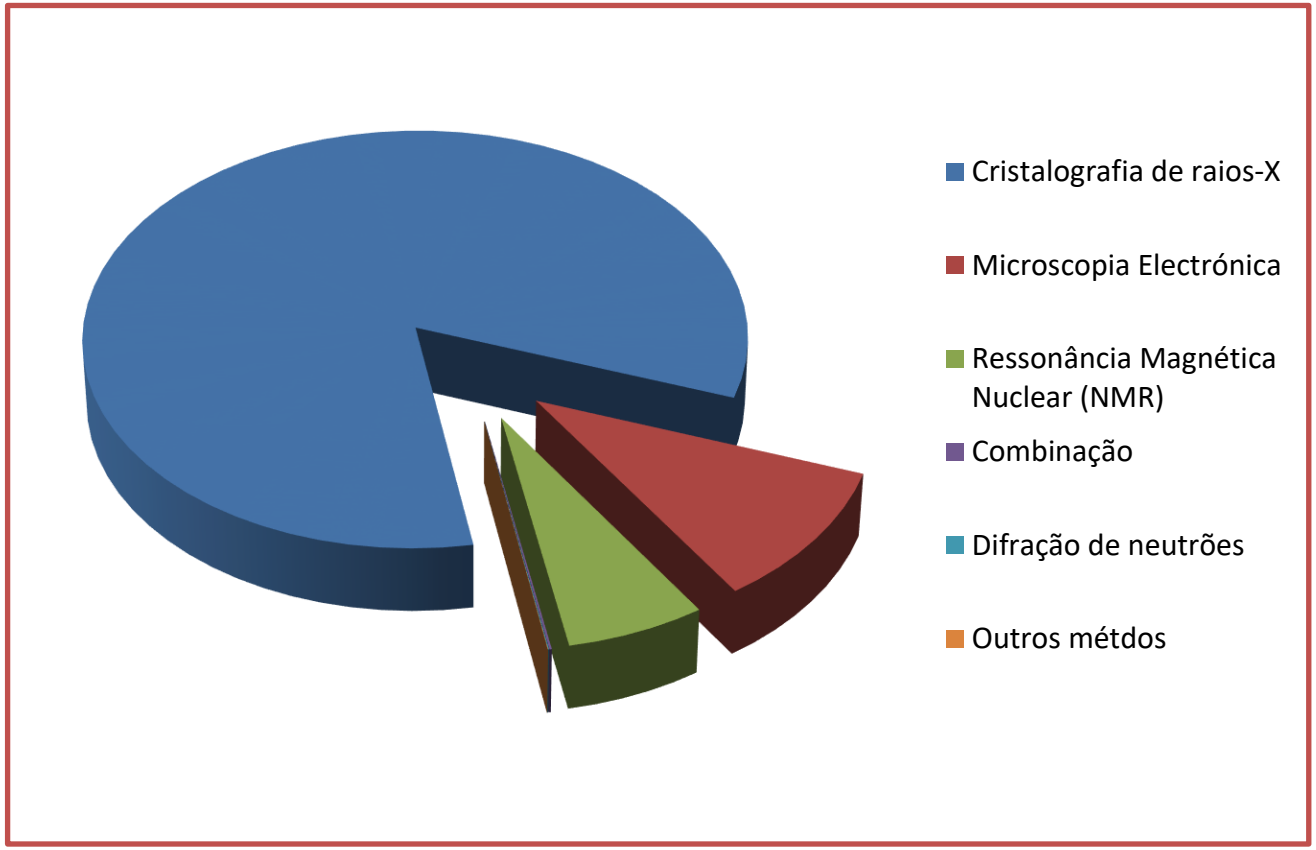
Métodos experimentais

- **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 intense 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.

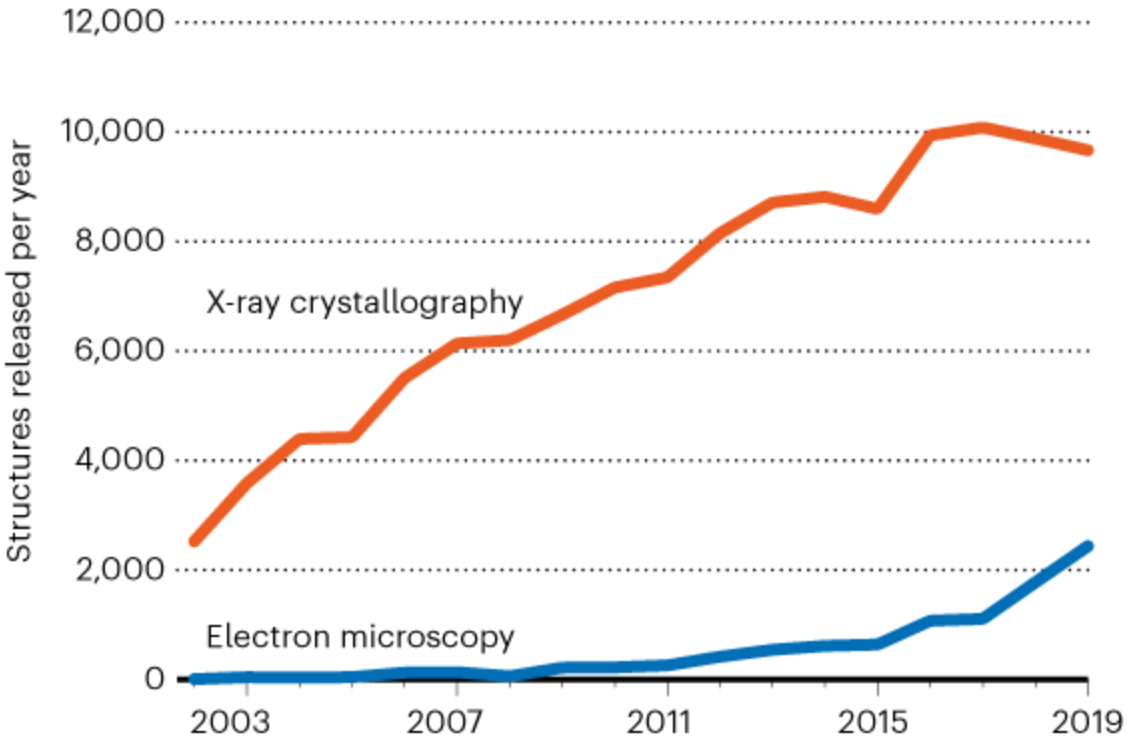
- **Crio-microscopia electrónica:** a amostra da molécula a estudar é congelada rapidamente a cerca de -180 ^{o}C e um feixe de electrões é usado para criar imagens de um enorme número de moleculas da amostra. A análise combinada destas imagens permite resolver a estrutura 3D da molécula.

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



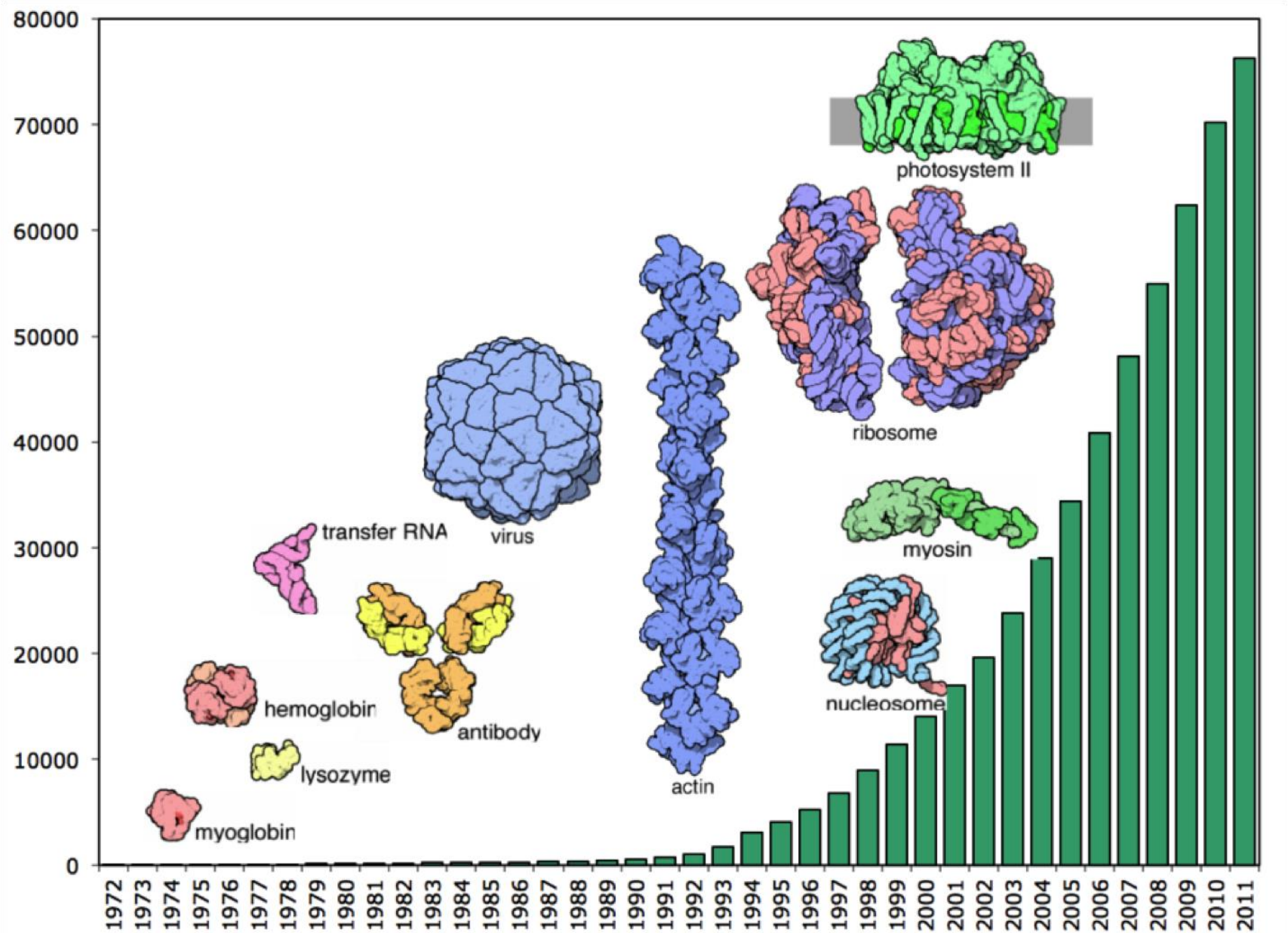
STRUCTURE SLEUTHS

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.

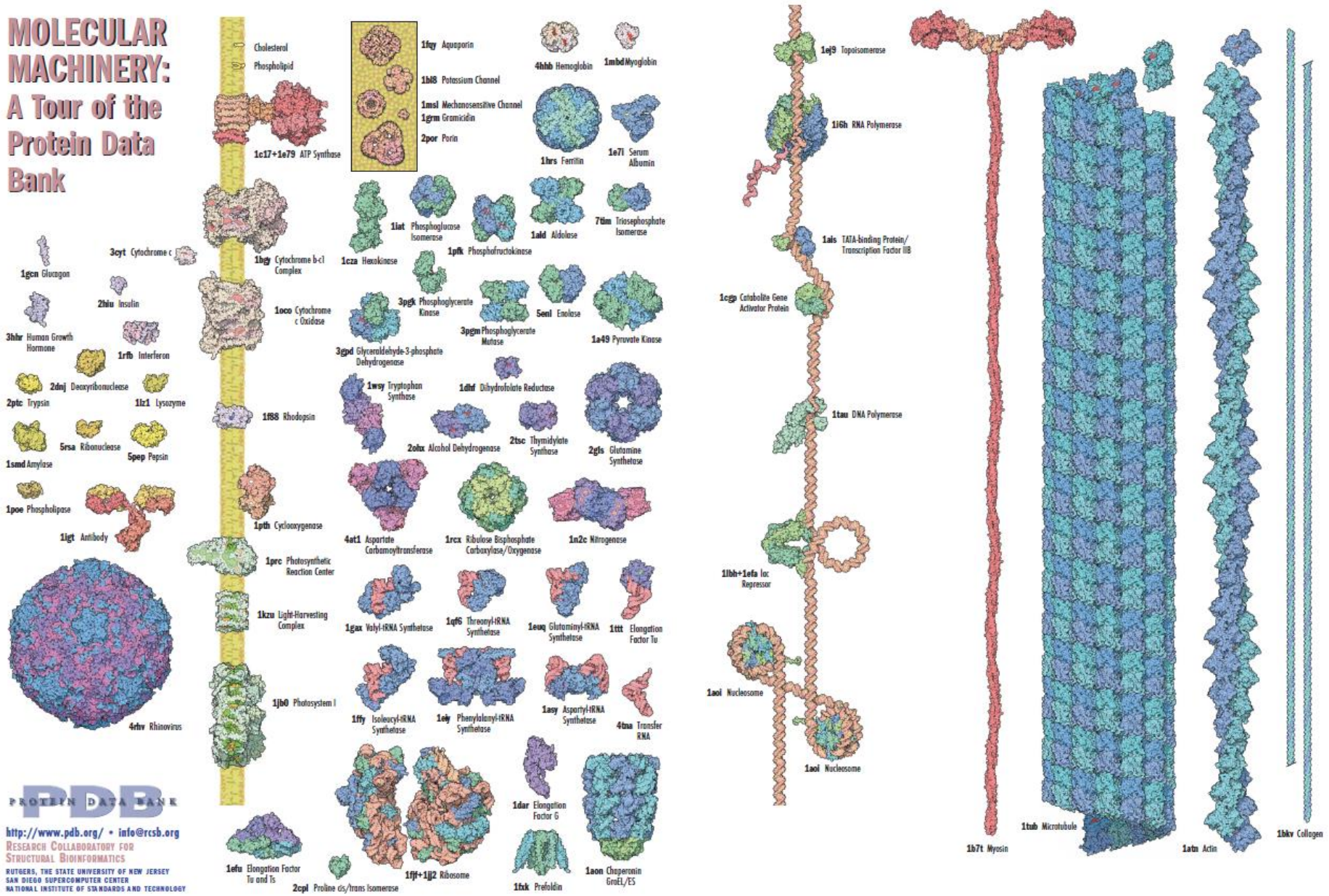


The electron microscopy line shows structures submitted to the Electron Microscopy Data Bank. Nearly all use cryo-EM.

Progresso na determinação das estruturas



O PDB contém uma enorme diversidade estrutural!



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)

lysozyme

Include CSM

Help

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

PDB-101 | wwPDB | EMDataResource | NAKB | wwPDB Foundation

Access Computed St

Welcome
Deposit
Search
Visualize
Analyze
Download
Learn

RCSB Protein Data Bank (RCSB P
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.

of the Month

Flagellar Motor

Latest Entries — As of Tue Nov 26 2024

Features & Highlights

Access SDF/MOL formatted CCD

News | Publications

December 1: World AIDS Day

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

Return Structures grouped by No Grouping Include Computed Structure Models (CSM) Count Clear Search

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

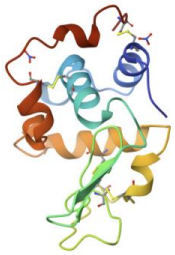
Download File View File

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



Explore in 3D

1JKB

Download File View File

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

1JKC

Download File View File

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

Return Structures grouped by No Grouping Include Computed Structure Models (CSM) Count Clear Search

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)

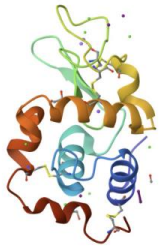
- 1.5 - 2.0 (5)
- 2.0 - 2.5 (2)

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

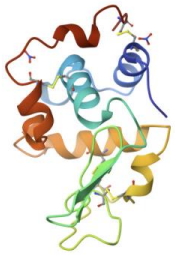
Download File View File

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



Explore in 3D

1JKB

Download File View File

HUMAN LYSOZYME MUTANT WITH GLU 35 REPLACED BY ALA

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

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

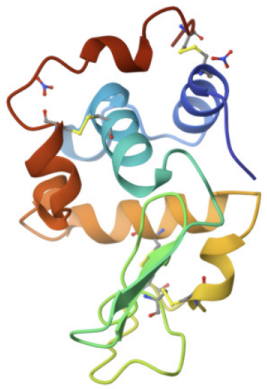
Released	1997-05-15
Method	X-RAY DIFFRACTION 1.66 Å
Organisms	Homo sapiens
Macromolecule	LYSOZYME (protein)
Unique Ligands	NO3

1JKC

Download File View File

https://www.rcsb.org/structure/1JKB

Biological Assembly 1



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

Global Symmetry: Asymmetric - C1

Global Stoichiometry: Monomer - A1

Display Files | Download Files | Data API

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

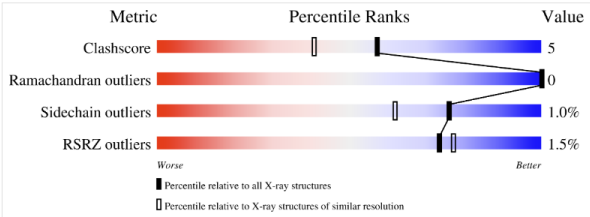
Deposited: 1996-11-13 **Released:** 1997-05-15
Deposition Author(s): Muraki, M., Harata, K., Goda, S., Nagahora, H.

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



- 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, 1JKC, 1JKD

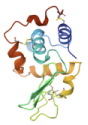
PubMed Abstract:
The importance of van der Waals contact between Glu 35 and Trp 109 to the active-site structure and the catalytic properties of human lysozyme (HL) has been investigated by site-directed mutagenesis. The X-ray analysis of mutant HLs revealed that both the replacement of Glu 35 by Asp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

Macromolecules

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

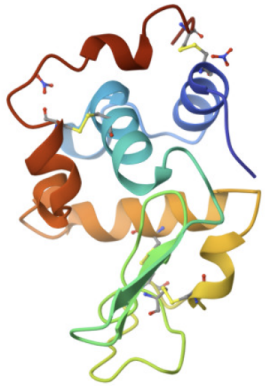
Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details	Image
LYSOZYME	A	130	Homo sapiens	**Mutation(s)**: 1 **Gene Names:** A SYNTHETIC GENE OF HUMAN LYSO **EC:** 3.2.1.17	

UniProt & NIH Common Fund Data Resources

Structure Summary | Structure | Annotations | Experiment | Sequence | Genome | Versions

Biological Assembly 1



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

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

Find Similar Assemblies

Biological assembly 1 assigned by authors.

Display Files | Download Files | Data API

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

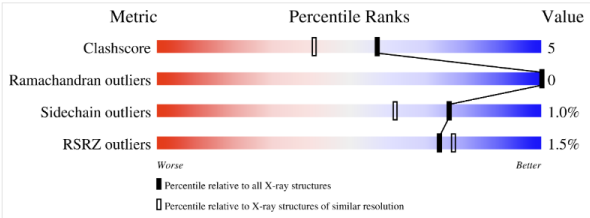
Deposited: 1996-11-13 **Released:** 1997-05-15
Deposition Author(s): Muraki, M., Harata, K., Goda, S., Nagahora, H.

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



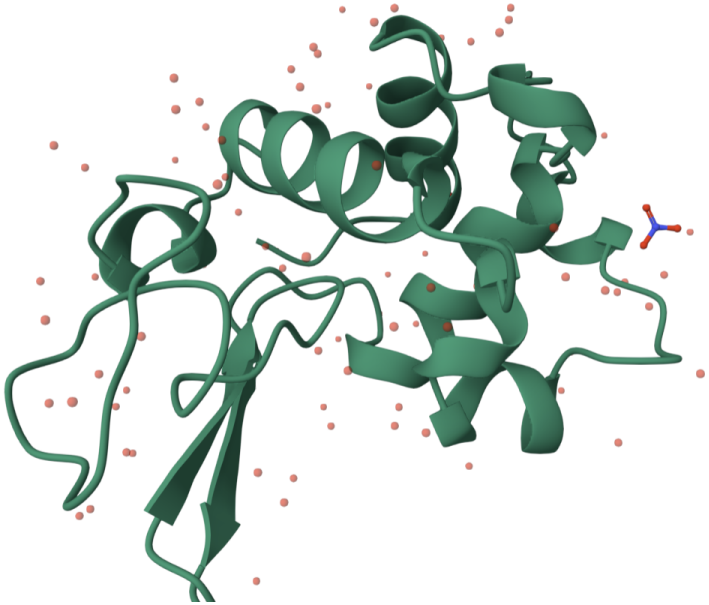
This is version 1.6 of the entry. See complete history.

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
KVFERCELARTLKRLGMDGYRGISLANWMCLAKWASGYNTRATNYNAGDRSTDYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRDPQGIRAWVAWRNRCQN
   121
RDVRQYVQGCGV
```



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	
Water	Ball & Stick	
Ion	Ball & Stick	
Unit Cell P 21 21 21		

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 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

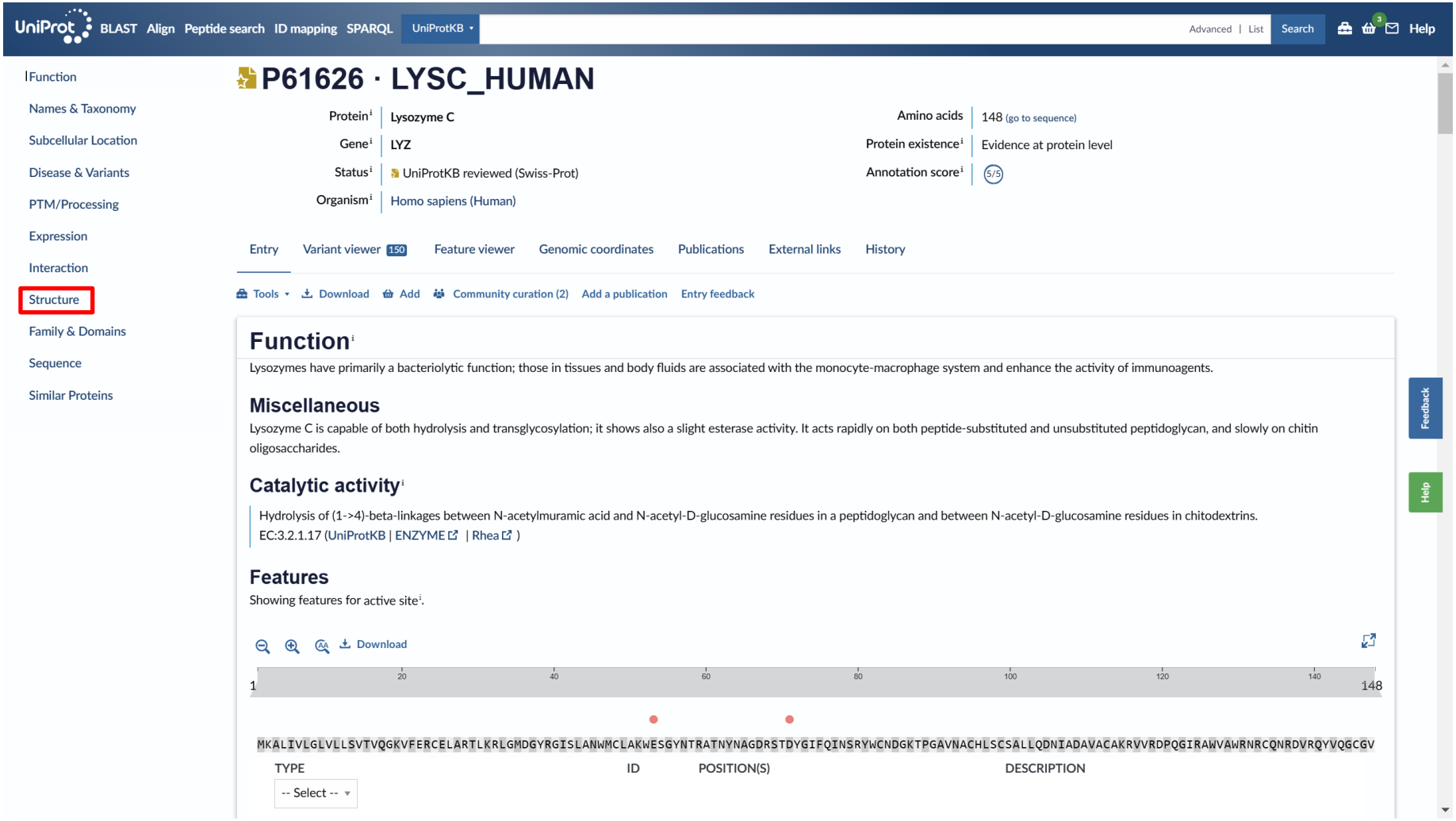
```
HEADER    METAL BINDING PROTEIN                   21-AUG-03   1Q8H
TITLE     CRYSTAL STRUCTURE OF PORCINE OSTEOCALCIN
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: OSTEOCALCIN;
COMPND   3 CHAIN: A
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: SUS SCROFA;
SOURCE   3 ORGANISM_COMMON: PIG
KEYWDS    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
```

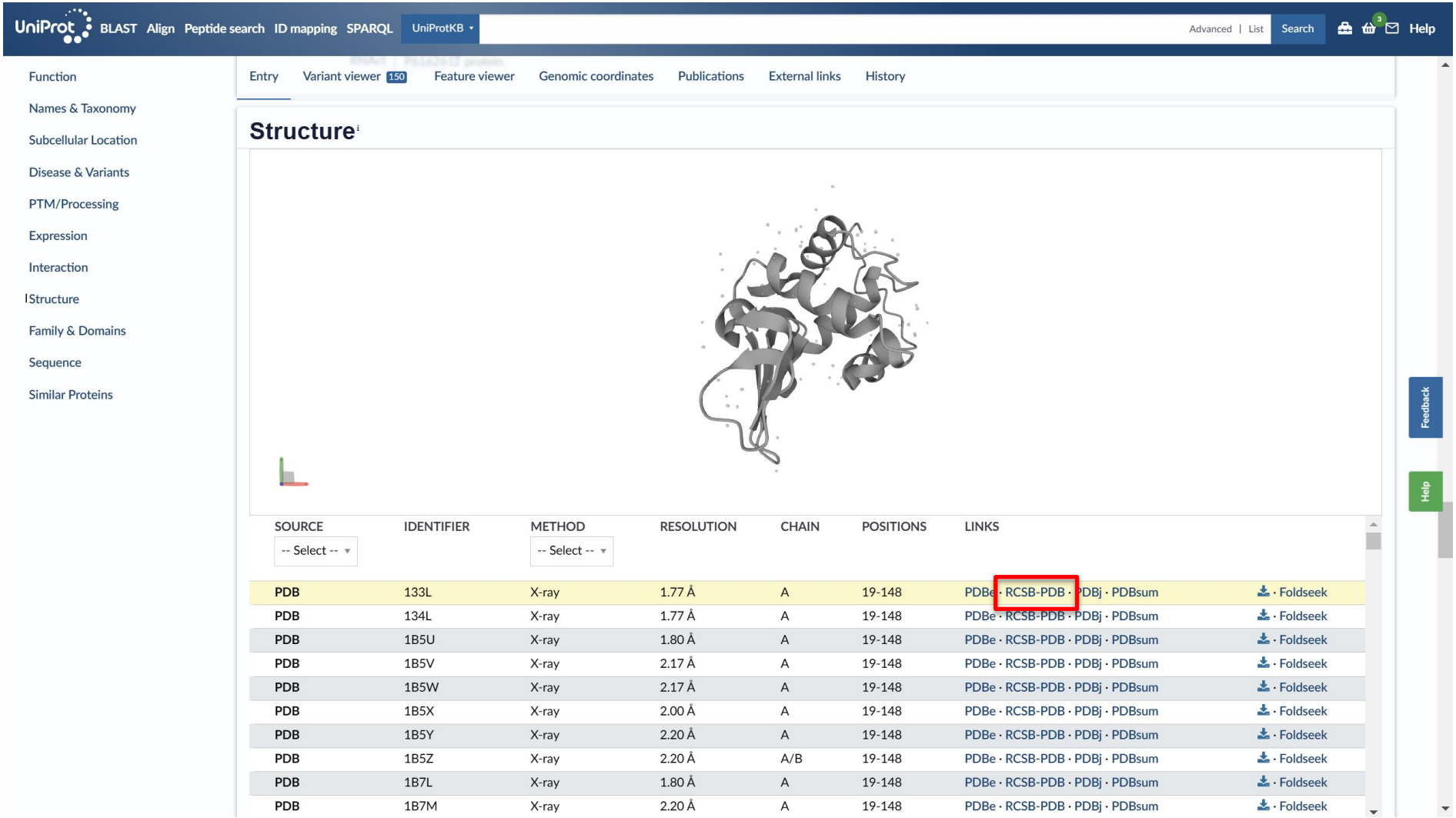
Cabeçalho

Coordenadas

Interligação entre Uniprot e PDB



Interligação entre Uniprot e PDB



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

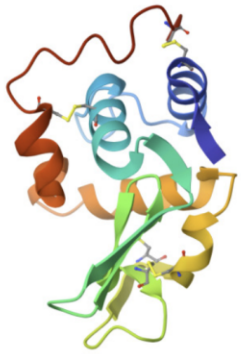
Advanced Search | Browse Annotations

Help

PDB-101 wwPDB EMDataResource NAKB wwPDB Foundation PDB-Dev

Structure Summary | Structure | Annotations | Experiment | Sequence | Genome | Versions

Biological Assembly 1



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

Global Symmetry: Asymmetric - C1

Global Stoichiometry: Monomer - A1

Find Similar Assemblies

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

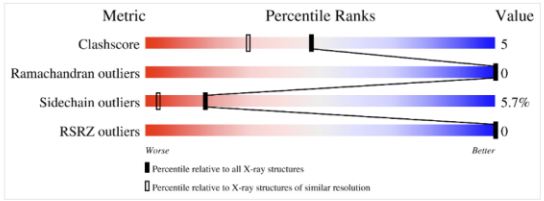
Classification: HYDROLASE(O-GLYCOSYL)
Organism(s): Homo sapiens
Mutation(s): No

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

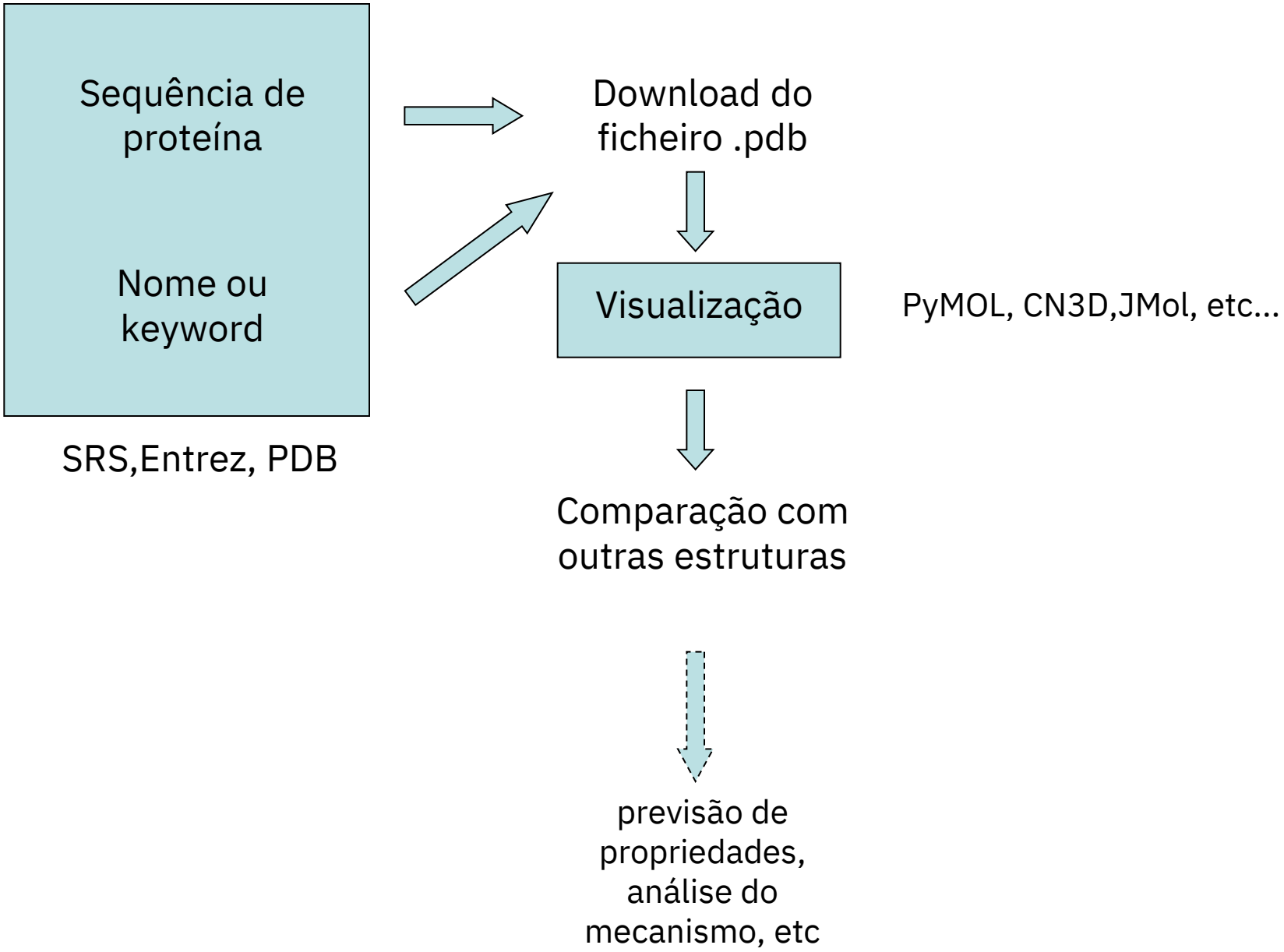
Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.77 Å
R-Value Work: 0.181
R-Value Observed: 0.181

wwPDB Validation 3D Report Full Report



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

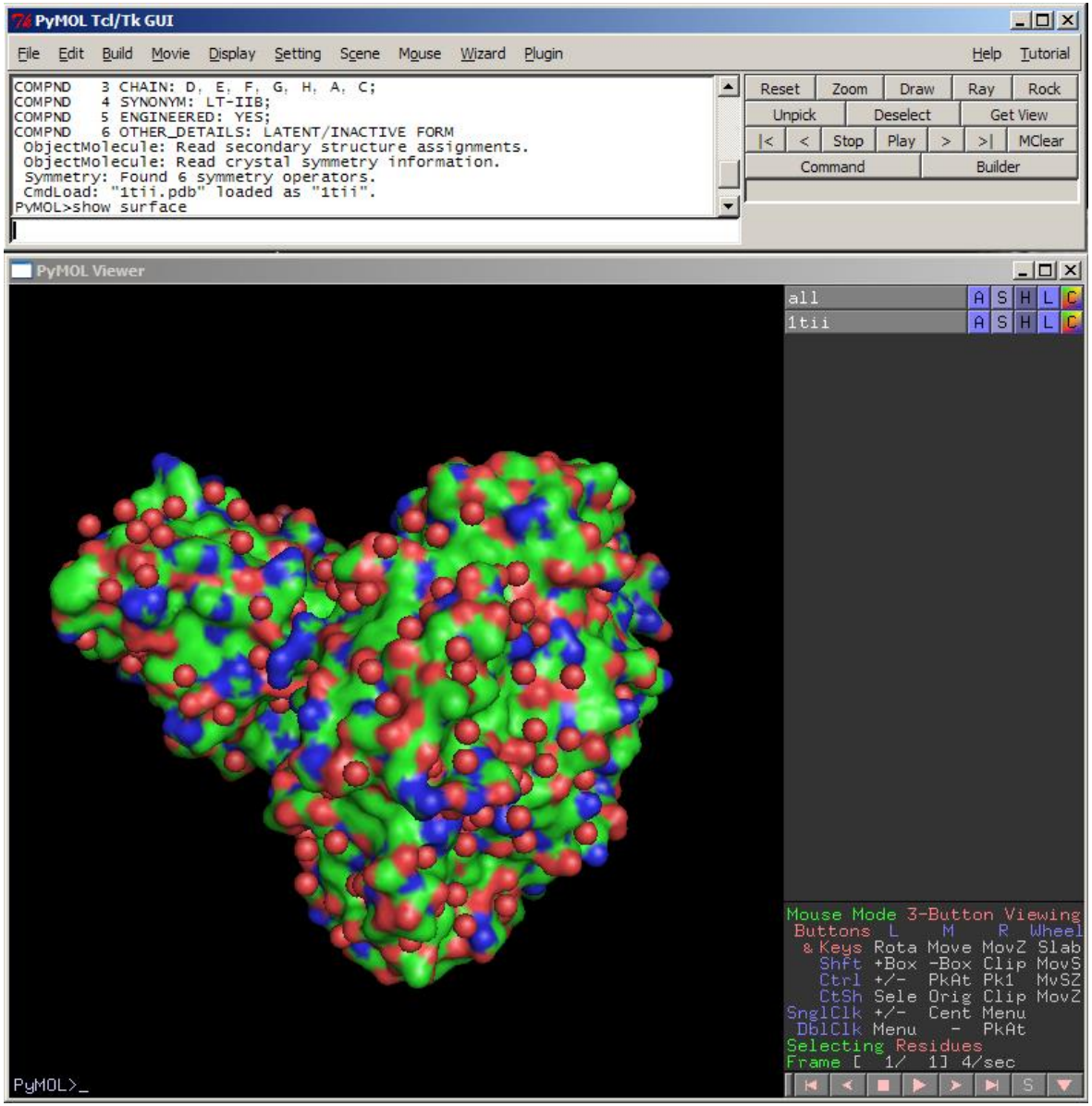
Instaláveis:

- **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/JSMol:** http://jmol.sourceforge.net/

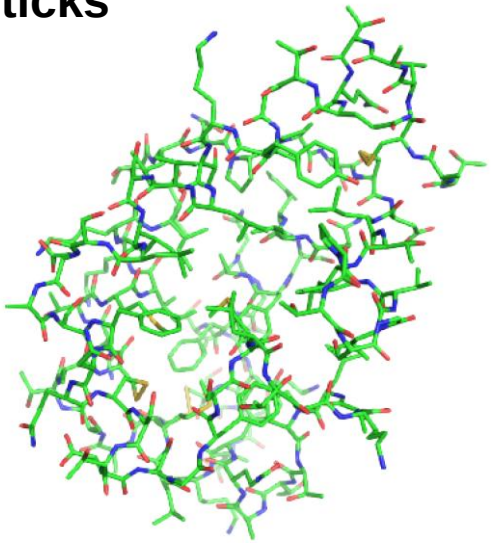
PyMOL



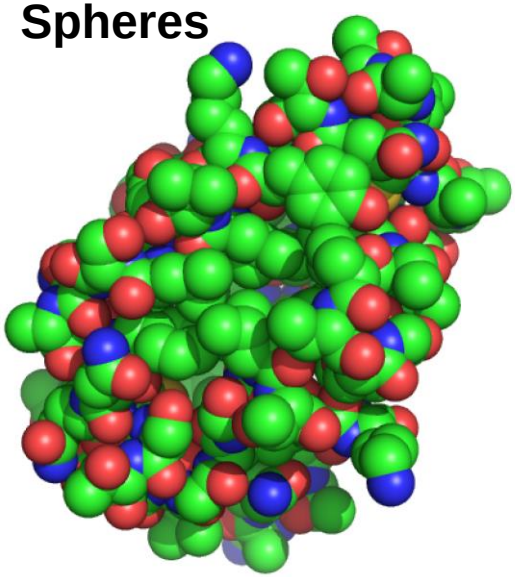
- Open Source
- Acesso livre
- Python / C
- Visualização de macromoléculas
- Animações moleculares
- Comparação de estruturas
- Scripting
- Windows / Linux

http://www.pymol.org

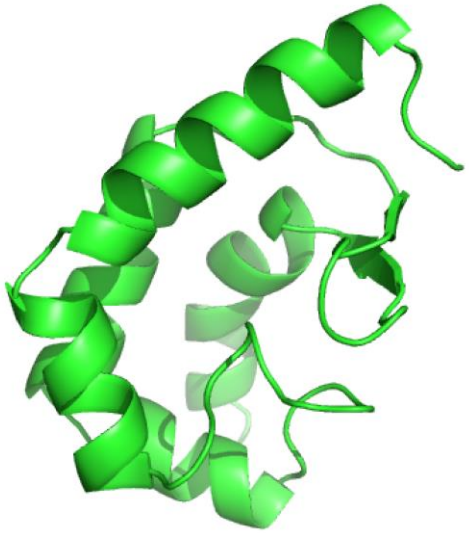
Sticks



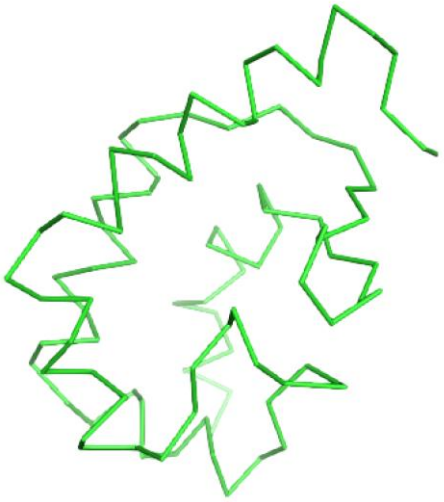
Spheres



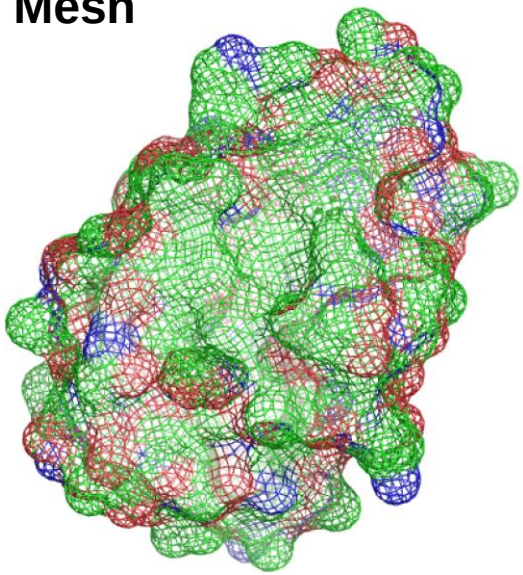
Cartoon



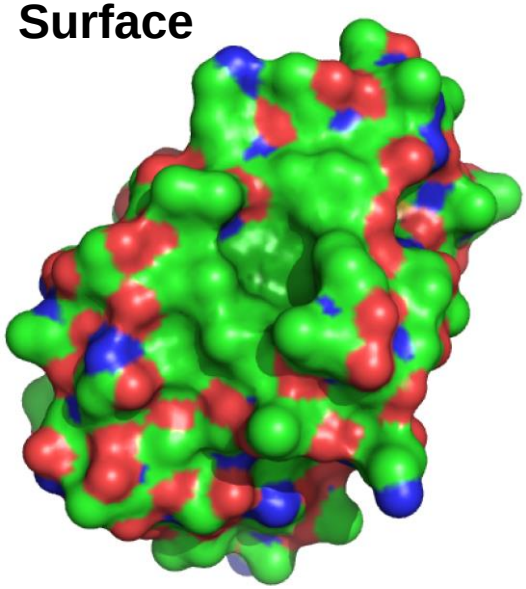
Ribbon



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 → Implica → Similaridade de estrutura

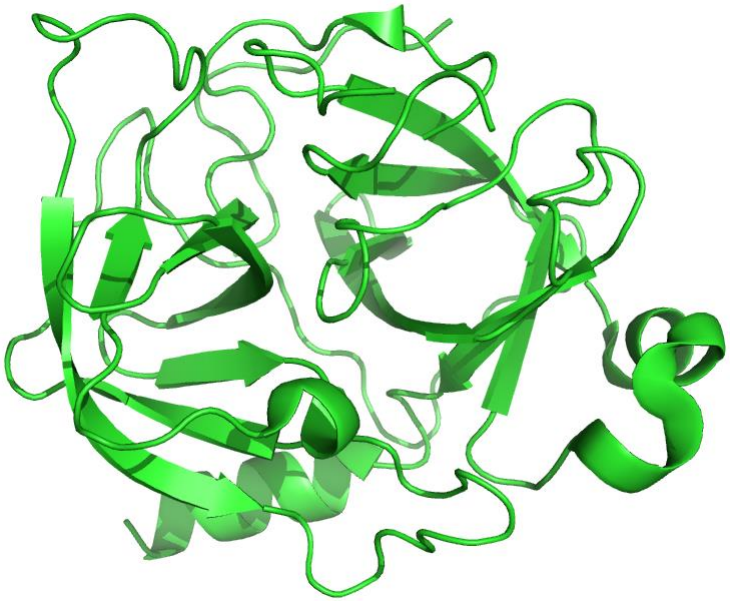
MAS

Similaridade de estrutura → Não implica → Similaridade de sequência

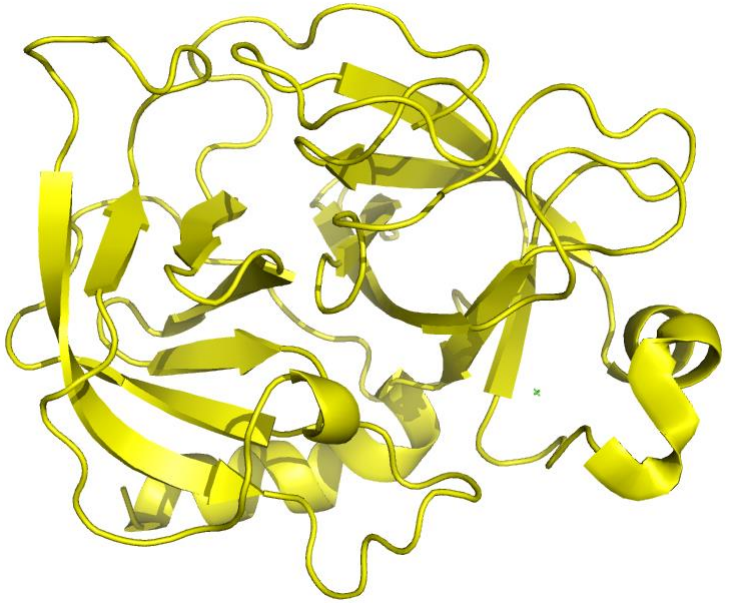
A pressão de selecção evolutiva opera sobre a estrutura (responsável pela função) e não directamente sobre a sequência. Alterações da sequência que conservem a estrutura são geralmente toleradas.

Similaridade estrutural e de sequência

Tripsina bovina



Tripsina *S. griseus*

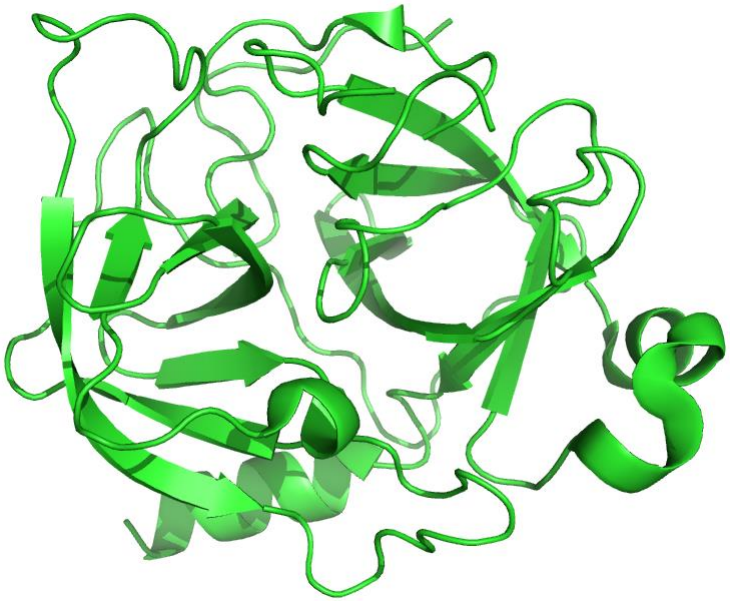


Alinhamento das sequências: **34%** identidade, E(1) **1.4×10^{-17}**

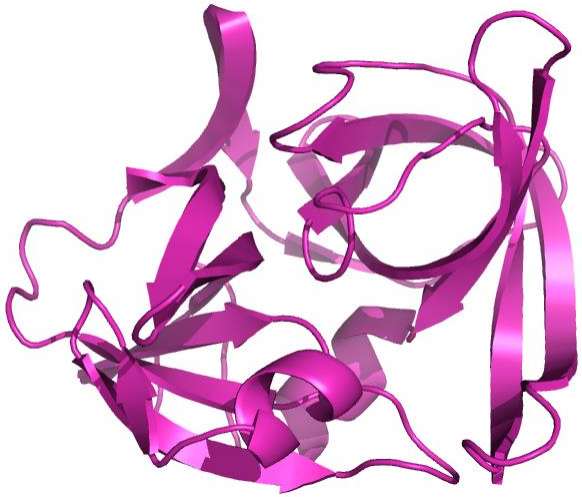
Proteínas homólogas, similaridade de sequências claramente detetável

Similaridade estrutural e de sequência

Tripsina bovina



Protease A *S. griseus*



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

Alinhamento sem significado estatístico

Proteínas homólogas, similaridade de sequências não é detectável

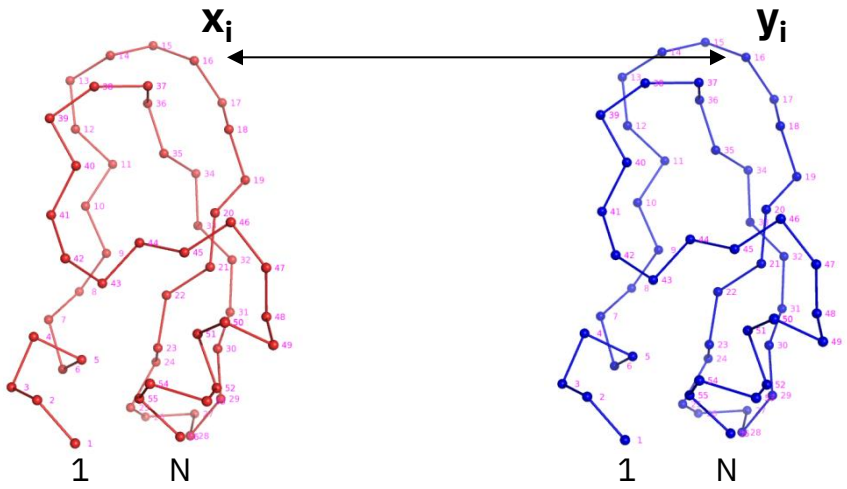
Como quantificar a similaridade estrutural ?

- 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)

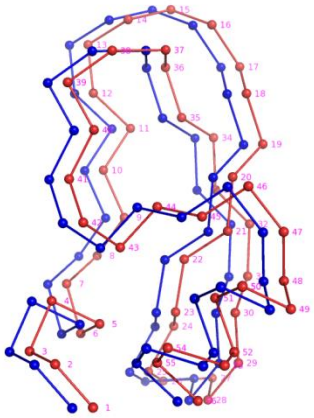
- O valor de RMSD depende da forma como se faz corresponder cada átomo da primeira estrutura a um átomo da segunda. Estabelecer esta correspondência não é um problema trivial, sobretudo para estruturas pouco semelhantes.

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

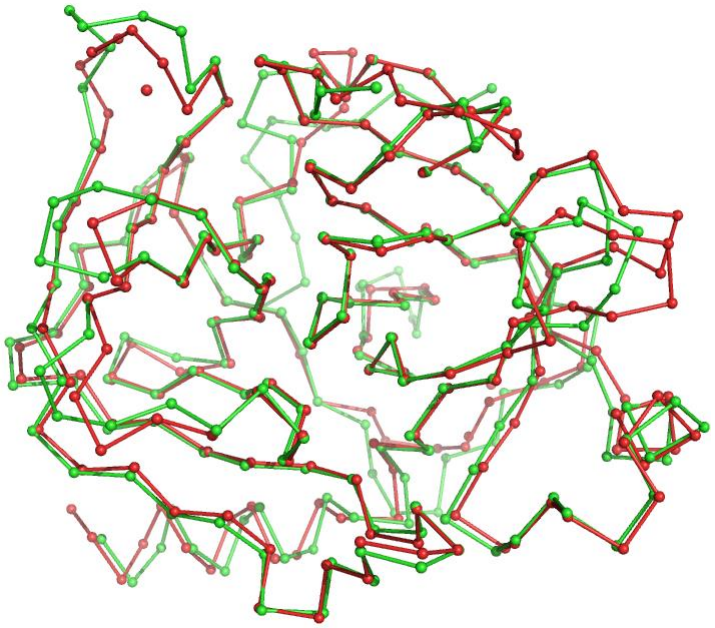
$$RMSD = \sqrt{\frac{\sum_{i}^{N} | \vec{x}_i - \vec{y}_i |^2}{N}}$$

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

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

Relação entre RMSD e identidade de sequência

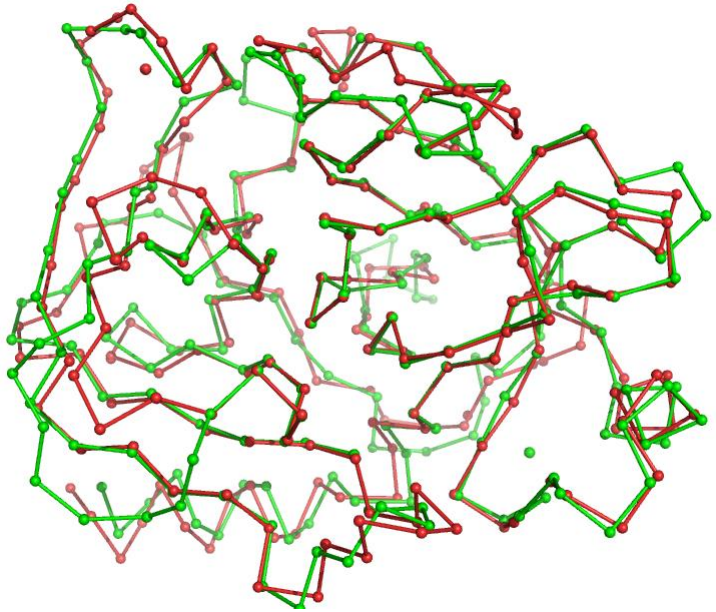
Tripsina humana
versus
Tripsina bovina



RMSD 0.8 Å

40% identidade de sequência

Tripsina humana
versus
Tripsina *S.griseus*

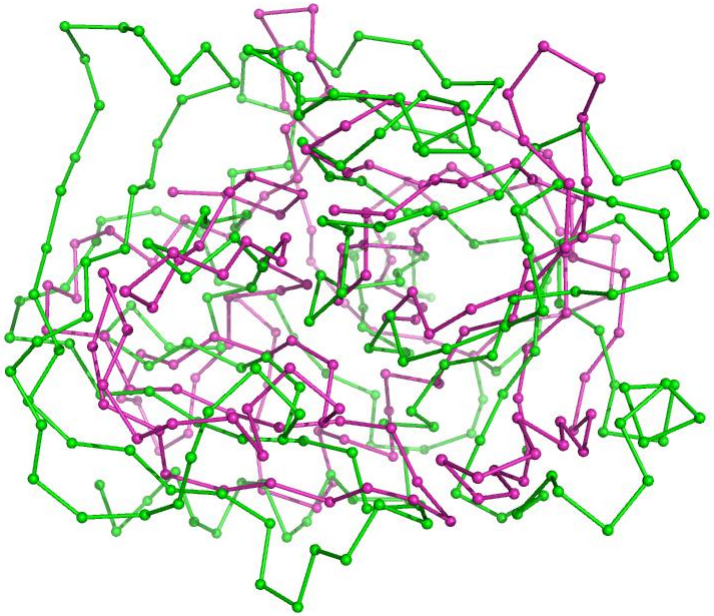


RMSD 1.8 Å

34% identidade de sequência

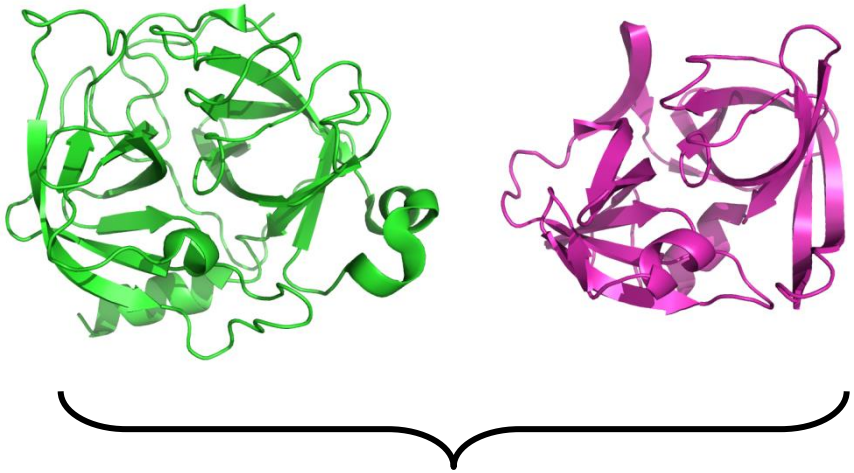
Relação entre RMSD e identidade de sequência

Tripsina humana
versus
Proteinase A *S.griseus*



RMSD 5.7 Å

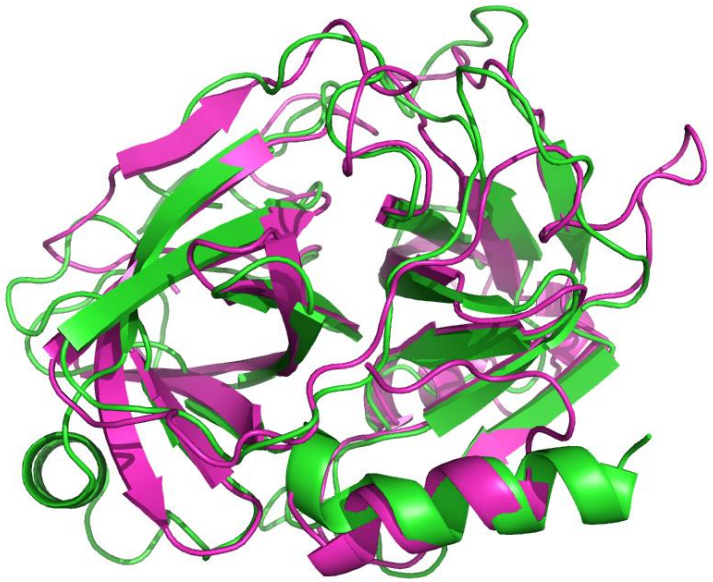
20% identidade de sequência



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

Relação entre RMSD e identidade de sequência

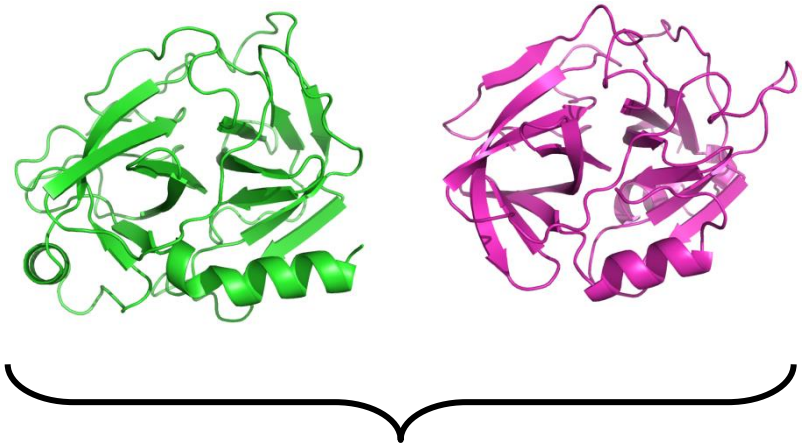
Tripsina humana
versus
Proteinase V8 *S.aureus*



RMSD 2.5 Å

19% identidade de sequência

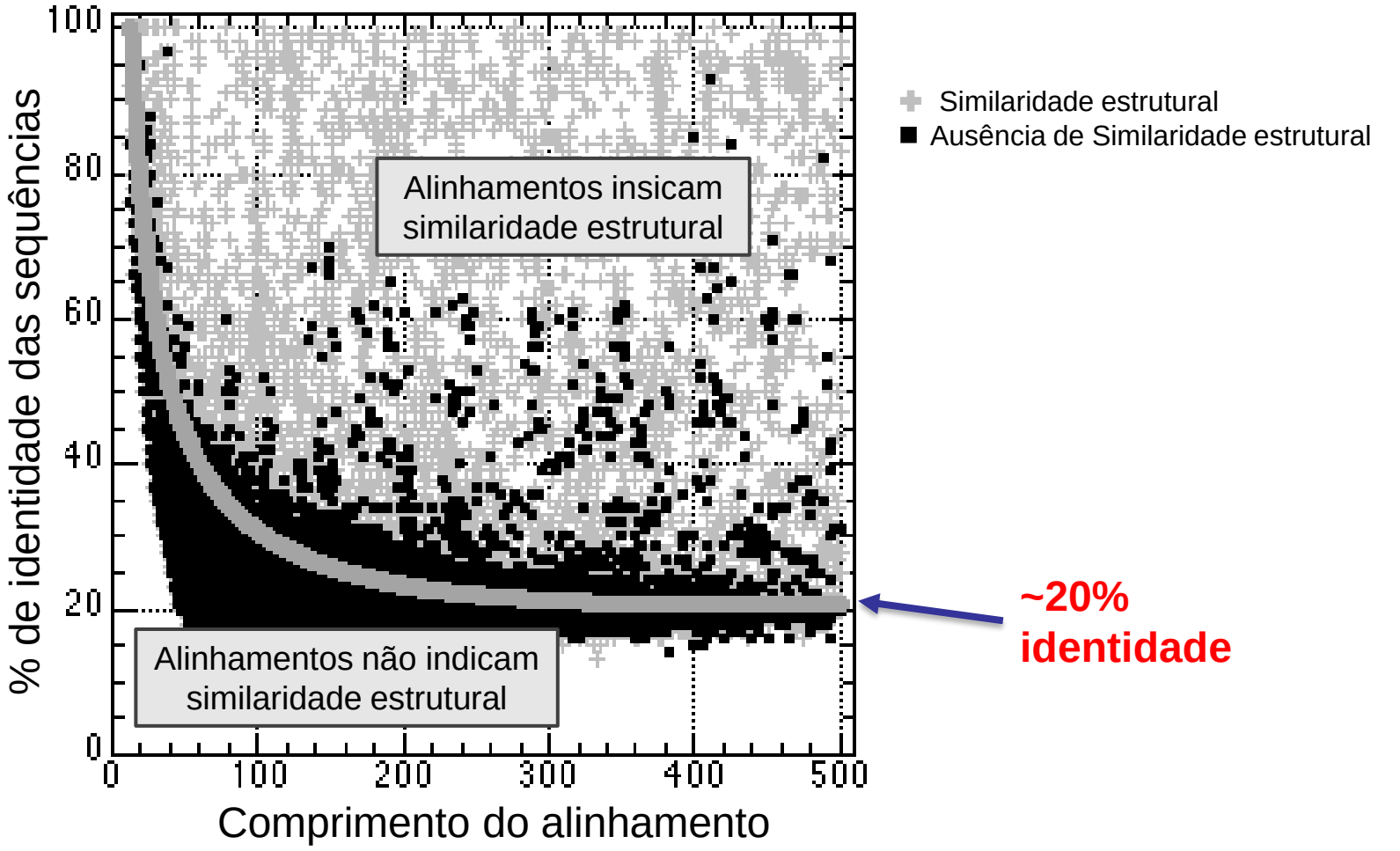
E-value: 8.6×10^2



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

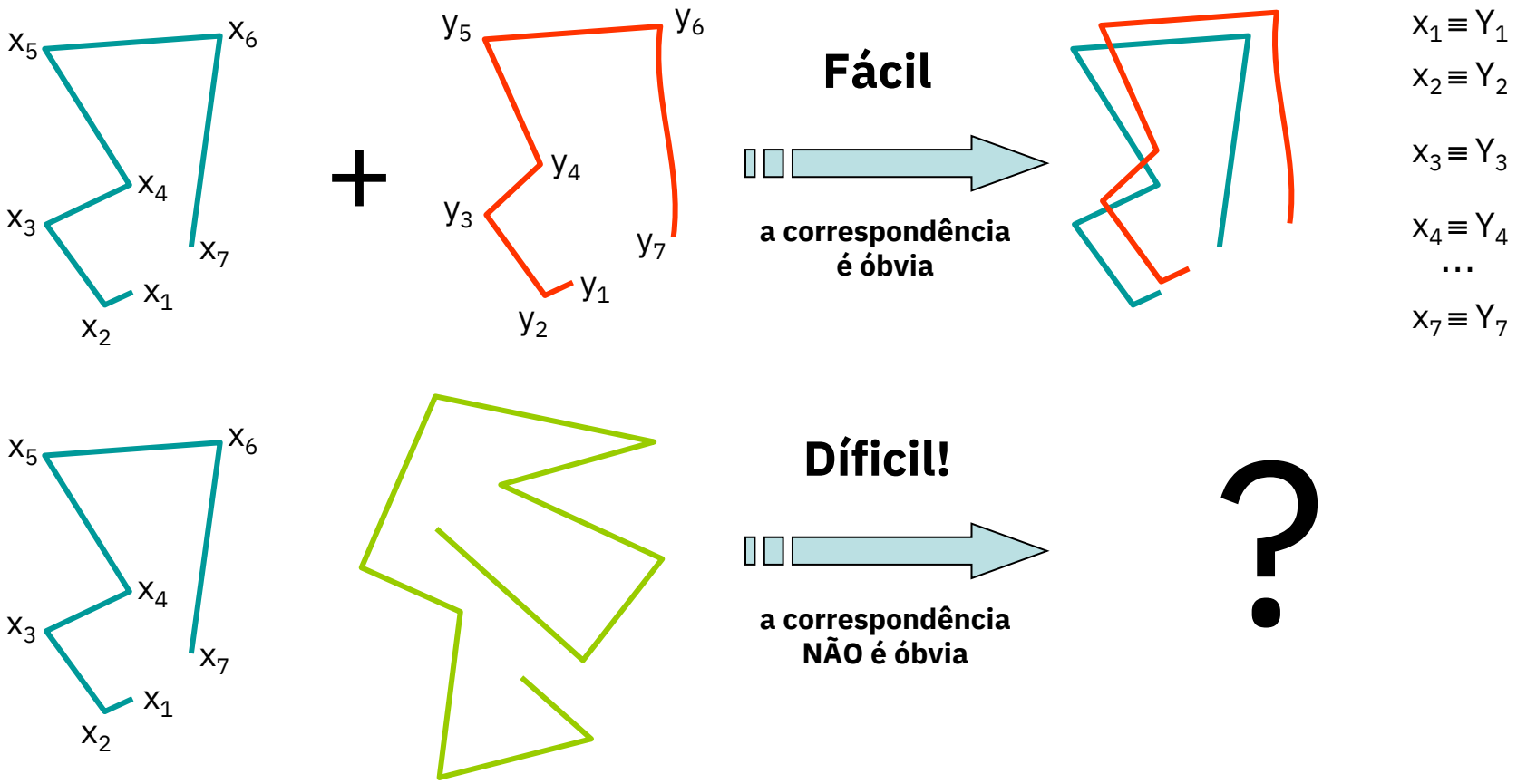


A relação entre a percentagem de identidade e a similaridade estrutural das proteínas depende do comprimento do alinhamento!

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



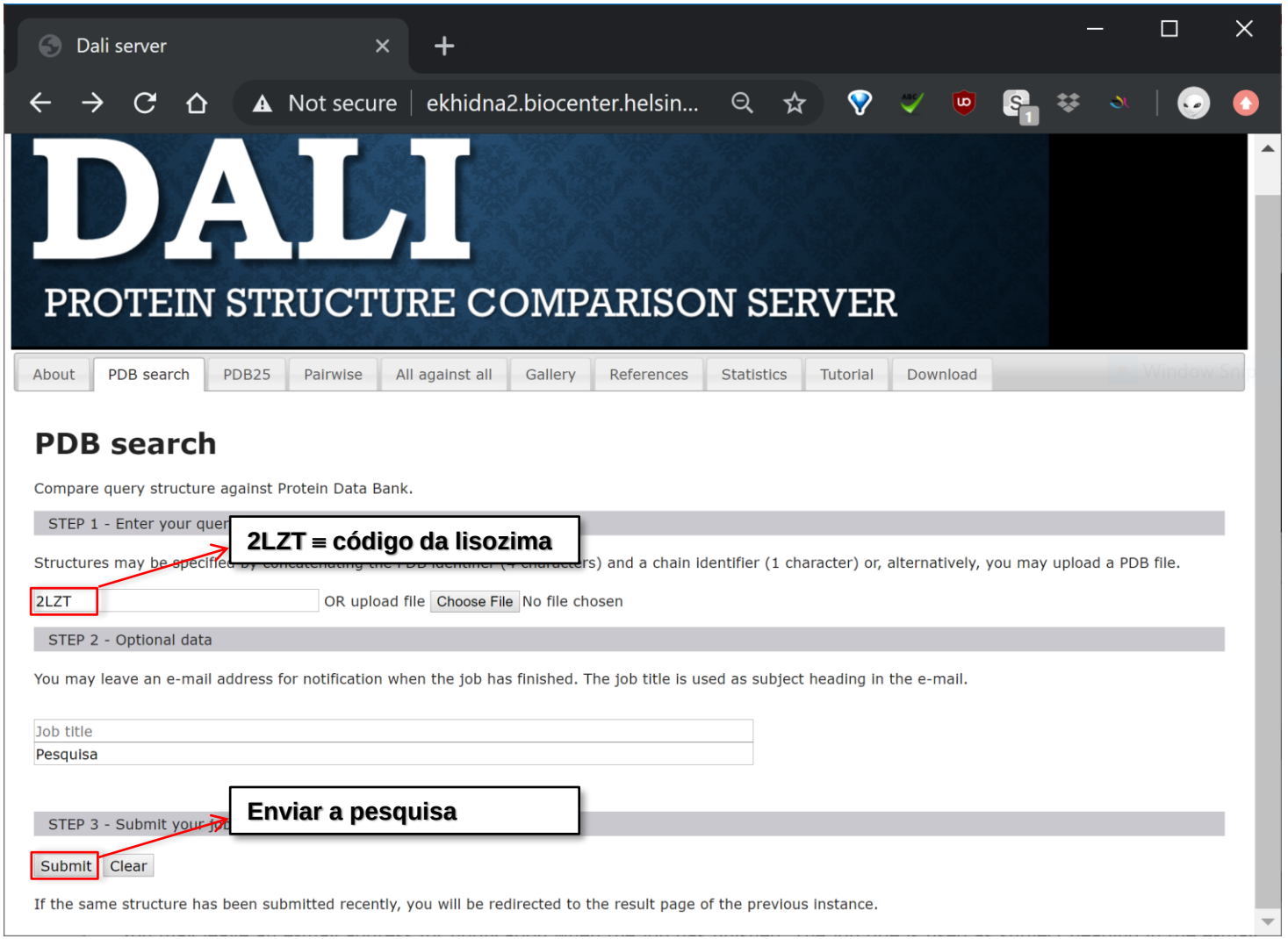
Sites para comparação e pesquisa estrutural

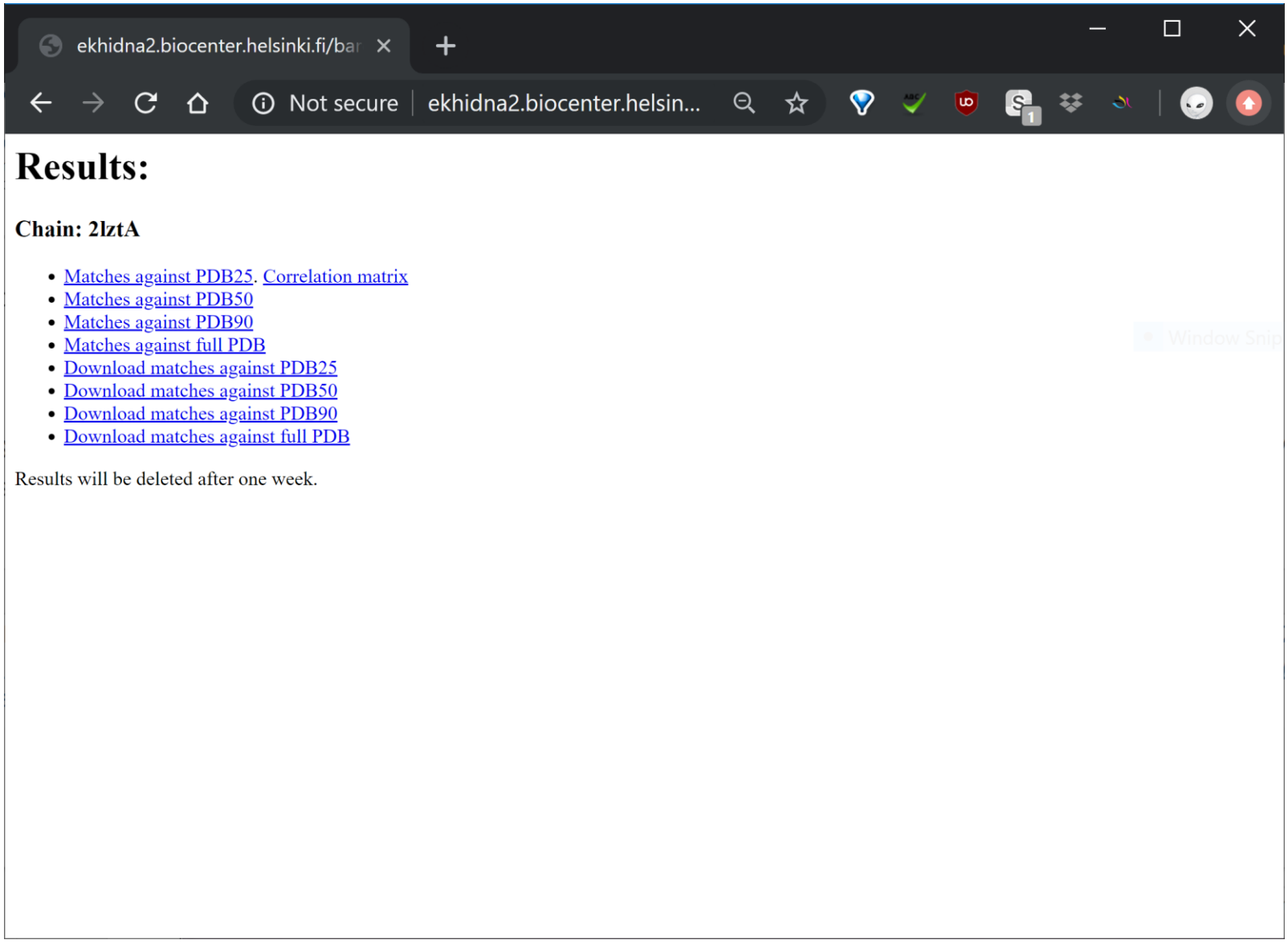
- **PDBeFold @ EBI (P,C, M): https://www.ebi.ac.uk/msd-srv/ssm/**
- **Top Match (C): https://topmatch.services.came.sbg.ac.at/**
- **DALI Server (P,C): http://ekhidna.biocenter.helsinki.fi/dali_server**
- **VAST (P): http://www.ncbi.nlm.nih.gov/Structure/VAST/**
- **VAST+ (P): http://www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi**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



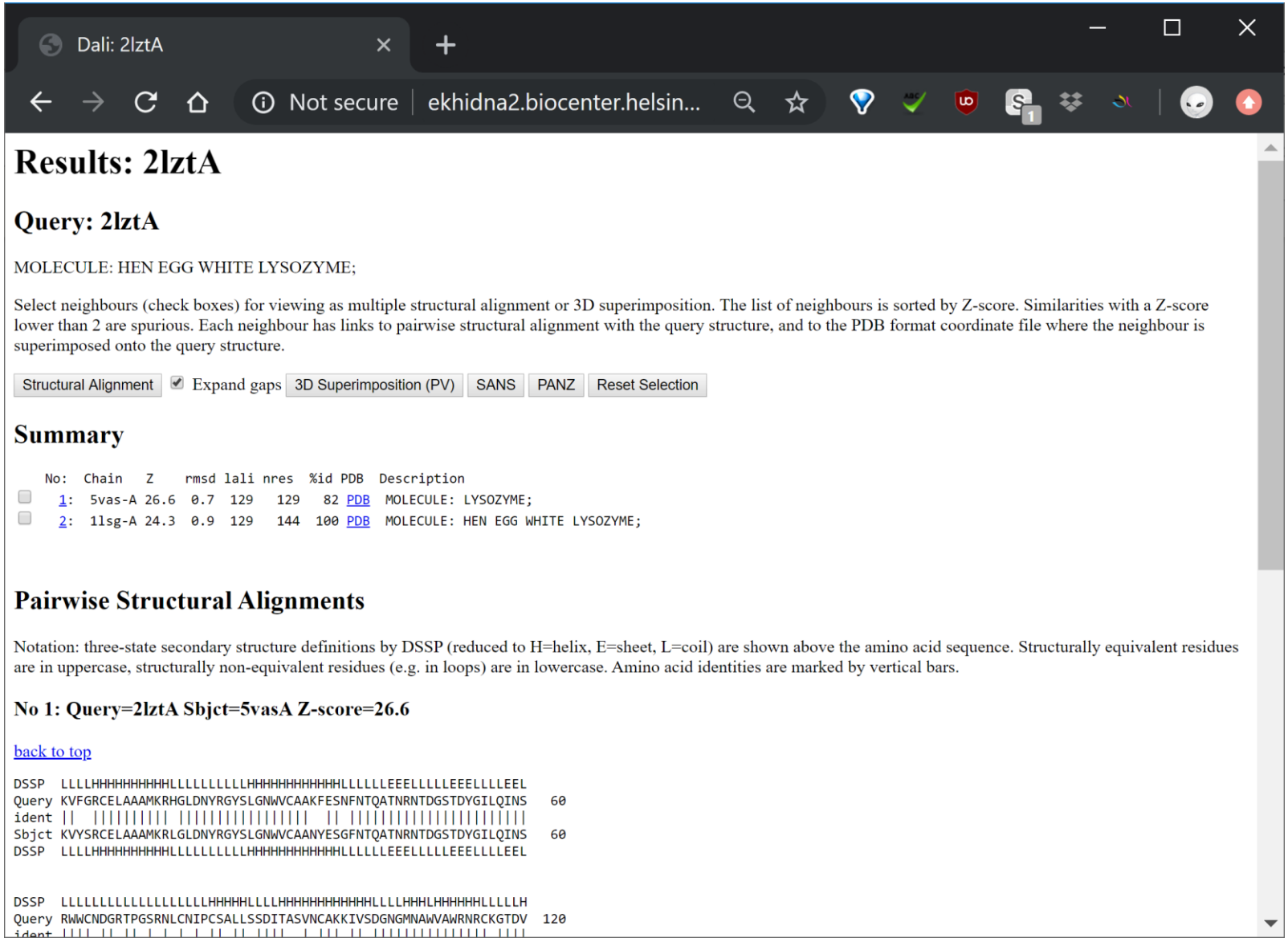


Results:

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.



Results: 2lztA

Query: 2lztA

MOLECULE: HEN EGG WHITE LYSOZYME;

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimposition. The list of neighbours is sorted by Z-score. Similarities with a Z-score lower than 2 are spurious. Each neighbour has links to pairwise structural alignment with the query structure, and to the PDB format coordinate file where the neighbour is superimposed onto the query structure.

Structural Alignment | Expand gaps | 3D Superimposition (PV) | SANS | PANZ | Reset Selection

Summary

	No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
☐	1:	5vas-A	26.6	0.7	129	129	82	PDB	MOLECULE: LYSOZYME;
☐	2:	1lsg-A	24.3	0.9	129	144	100	PDB	MOLECULE: HEN EGG WHITE LYSOZYME;

Pairwise Structural Alignments

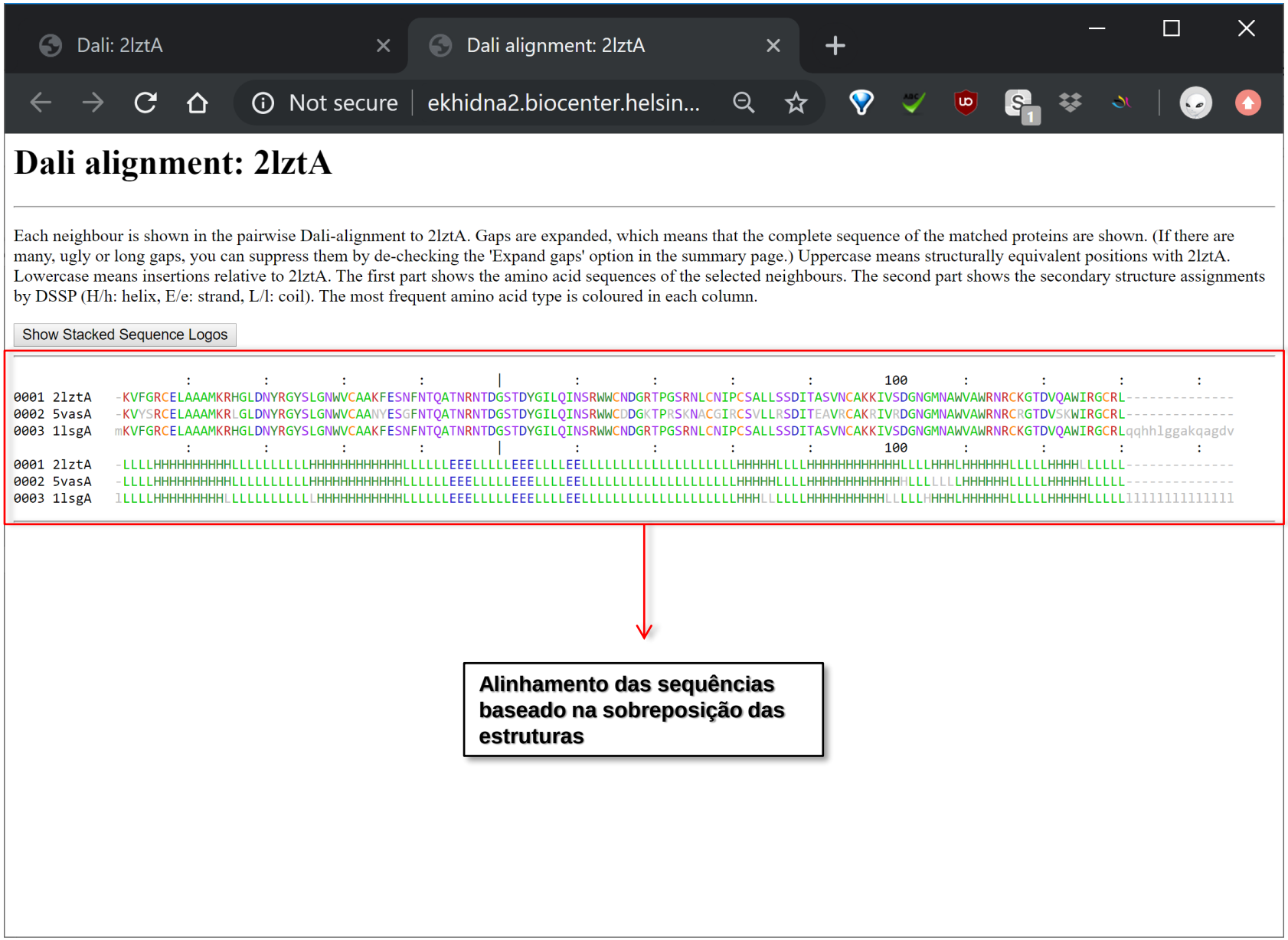
Notation: three-state secondary structure definitions by DSSP (reduced to H=helix, E=sheet, L=coil) are shown above the amino acid sequence. Structurally equivalent residues are in uppercase, structurally non-equivalent residues (e.g. in loops) are in lowercase. Amino acid identities are marked by vertical bars.

No 1: Query=2lztA Sbjct=5vasA Z-score=26.6

back to top

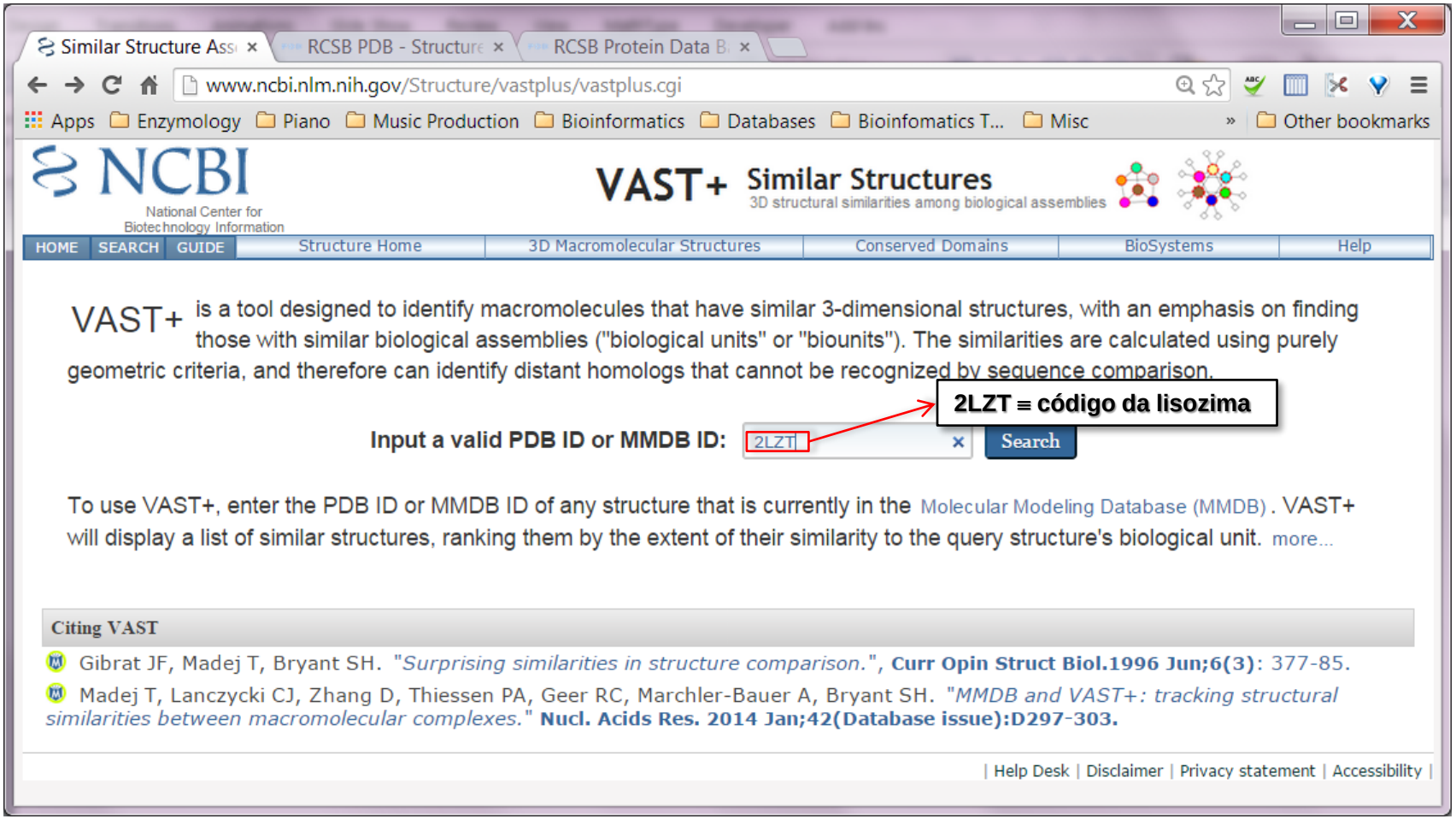
```
DSSP  LLLLHHHHHHHHHHLLLLLLLLLLHHHHHHHHHHHHHLLLLLEEELLLLEEELLLEEL
Query KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINS   60
ident ||  |||||||||| ||||||||||||||||  || ||||||||||||||||||||||||
Sbjct KVYSRCELAAAMKRLGLDNYRGYSLGNWVCAANYESGFNTQATNRNTDGSTDYGILQINS   60
DSSP  LLLLHHHHHHHHHHLLLLLLLLLLHHHHHHHHHHHHHLLLLLEEELLLLEEELLLEEL


DSSP  LLLLLLLLLLLLLLLLLLLHHHHLLLLHHHHHHHHHHHLLLLHHHLHHHHHHLLLLLH
Query RWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTDV  120
ident |||| || || | | | | || || ||||  | ||| || ||||||||||||||| ||||
```


Dali: 2lztA
Dali alignment: 2lztA
Not secure
ekhidna2.biocenter.helsin...
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 -KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTDVQAWIRGCRL----------
0002 5vasA -KVYSRCELAAAMKRLGLDNYRGYSLGNWVCAANYESGFNTQATNRNTDGSTDYGILQINSRWWCDDGKTPRSKNACGIRCSVLLRSDITEAVRCAKRIVRDGNGMNAWVAWRNRCRGTDVSKWIRGCRL----------
0003 1lsgA mKVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTDVQAWIRGCRLqqhhlggakqagdv
0001 2lztA -LLLLHHHHHHHHHHLLLLLLLLLLLHHHHHHHHHHHHLLLLLEEELLLLEEELLLLEELLLLLLLLLLLLLLLLLLLHHHHLLLLHHHHHHHHHHLLLLHHHLHHHHHLLLLLHHHHLLLLL----------
0002 5vasA -LLLLHHHHHHHHHHLLLLLLLLLLLHHHHHHHHHHHHLLLLLEEELLLLLEEELLLLEELLLLLLLLLLLLLLLLLLLHHHHLLLLHHHHHHHHHHHLLLLLLLHHHHHLLLLLHHHHHLLLLL----------
0003 1lsgA lLLLLHHHHHHHHHLLLLLLLLLLLLHHHHHHHHHHHLLLLLEEELLLLLEEELLLLEELLLLLLLLLLLLLLLLLLLHHHLLLLLHHHHHHHHHHLLLLLHHHHHLLLLLlllllllllllll
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**.



VAST+ Similar Struct × | RCSB PDB - Structure × | RCSB Protein Data B × | www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi?uid=2LZT

Apps | Enzymology | Piano | Music Production | Bioinformatics | Databases | Bioinfomatics T... | Misc | Other bookmarks

NCBI National Center for Biotechnology Information

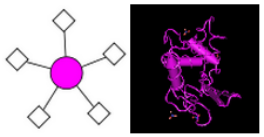
VAST+ Similar Structures

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 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

Original VAST

▼ Display filters

Showing 1 to 10 out of 860 structures | Search within results: PDB ID or MMDB ID [Go]

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

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

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

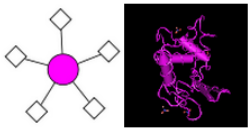
NCBI National Center for Biotechnology Information

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 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

Original VAST

▼ Display filters

Showing 771 to 780 out of 860 structures | **Search within results:** PDB ID or MMDB ID | Go

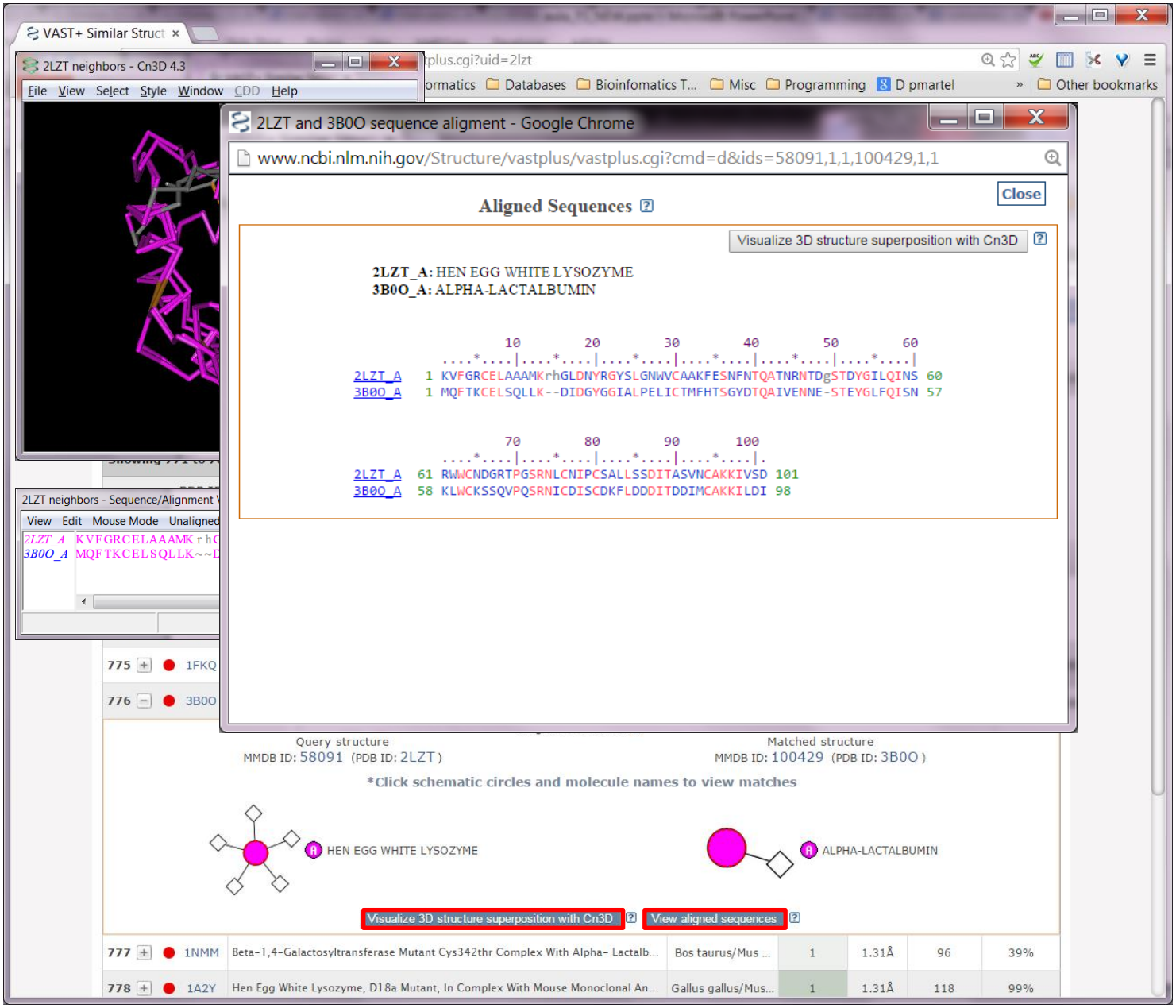
	PDB ID	Description	Taxonomy	Aligned Protein	RMSD	Aligned Residues	Sequence Identity
771	1JA2	Binding Of N-Acetylglucosamine To Chicken Egg Lysozyme: A Powder Diffractio...	Gallus gallus	1	1.29Å	90	100%
772	2FYD	Catalytic Domain Of Bovine Beta 1, 4-Galactosyltransferase In Complex With Alp...	Bos taurus/Mus ...	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	3B0O	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	1NWG	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 | First | Previous | Page 78 of 86 Pages | Next | Last

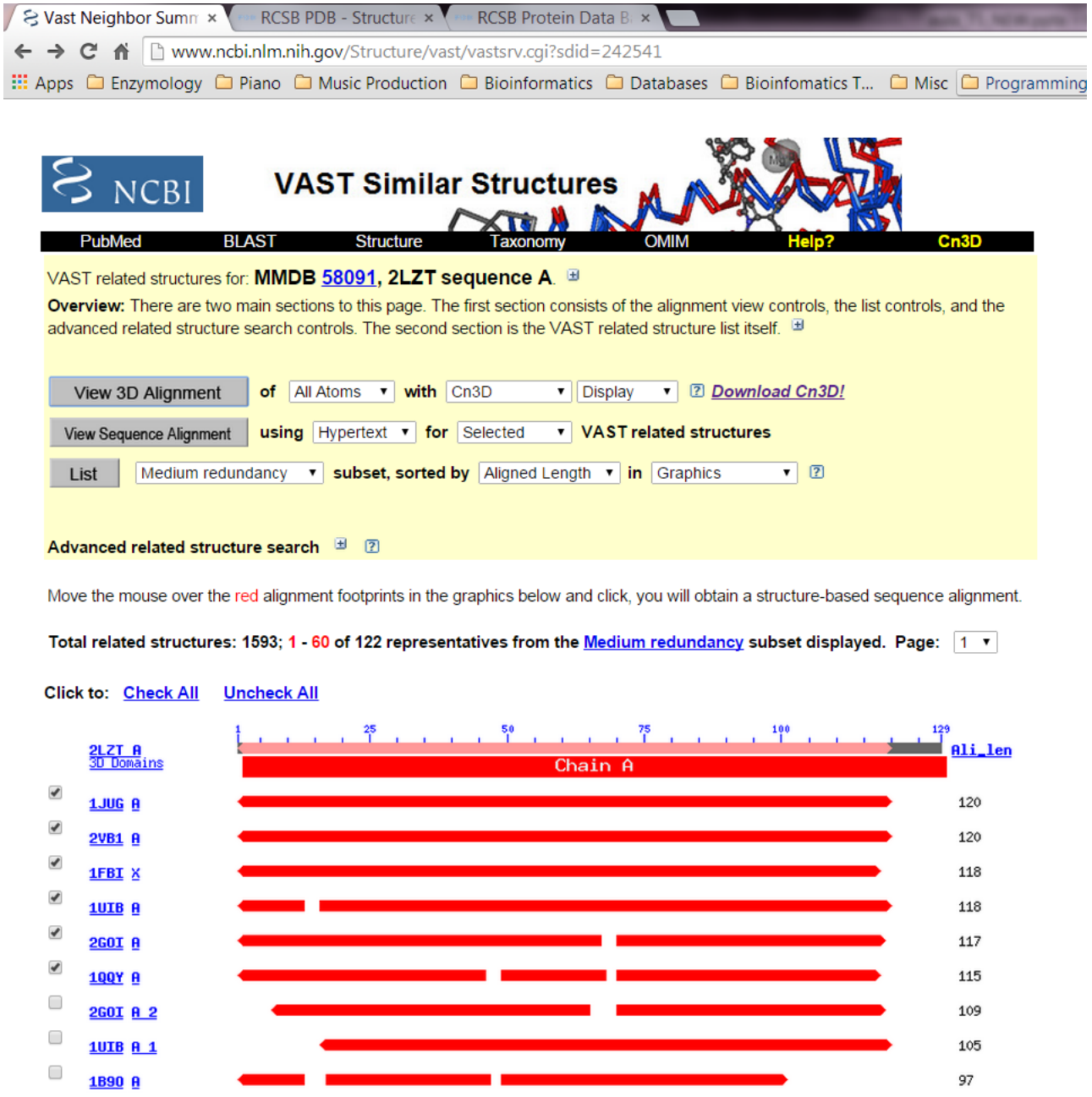
Citing VAST

Gibrat JF, Madej T, Bryant SH. *"Surprising similarities in structure comparison."*, **Curr Opin Struct Biol.1996 Jun;6(3)**: 377-85.

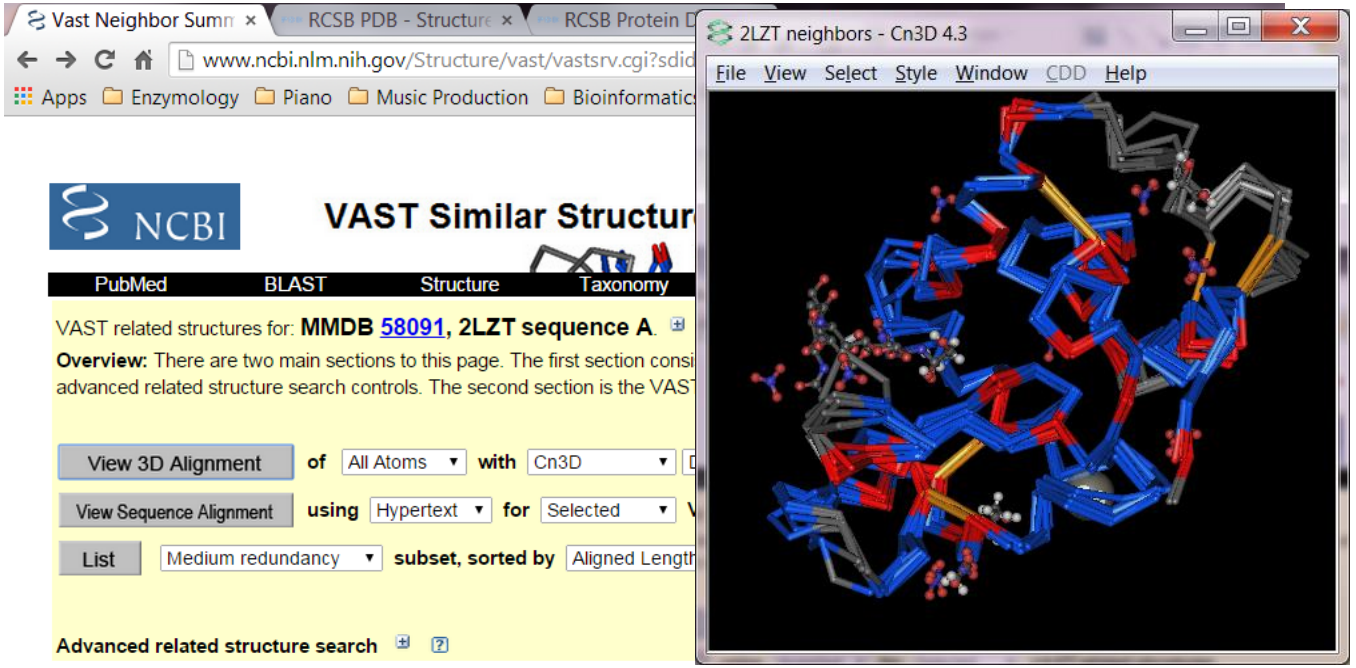
Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, Bryant SH. *"MMDB and VAST+: tracking structural similarities between macromolecular complexes."* **Nucl. Acids Res. 2014 Jan;42(Database issue):D297-303.**


VAST+ Similar Struct
2LZT neighbors - Cn3D 4.3
File View Select Style Window CDD Help
2LZT neighbors - Sequence/Alignment
View Edit Mouse Mode Unaligned
2LZT_A KVFGRCELAAAMK r h G
3B0O_A MQFTKCELSQLLK~~D
2LZT and 3B0O sequence aligment - Google Chrome
www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi?cmd=d&ids=58091,1,1,100429,1,1
Close
Aligned Sequences
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 KVFGRCELAAAMKrhGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDgSTDYGILQINS 60
3B0O_A 1 MQFTKCELSQLLK--DIDGYGGIALPELICTMFHTSGYDTQAIVENNE-STEYGLFQISN 57
70 80 90 100
2LZT_A 61 RWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSD 101
3B0O_A 58 KLWCKSSQVPQSRNICDISCDKFLDDDITDDIMCAKKILDI 98
775 1FKQ
776 3B0O
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
A HEN EGG WHITE LYSOZYME
A ALPHA-LACTALBUMIN
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%

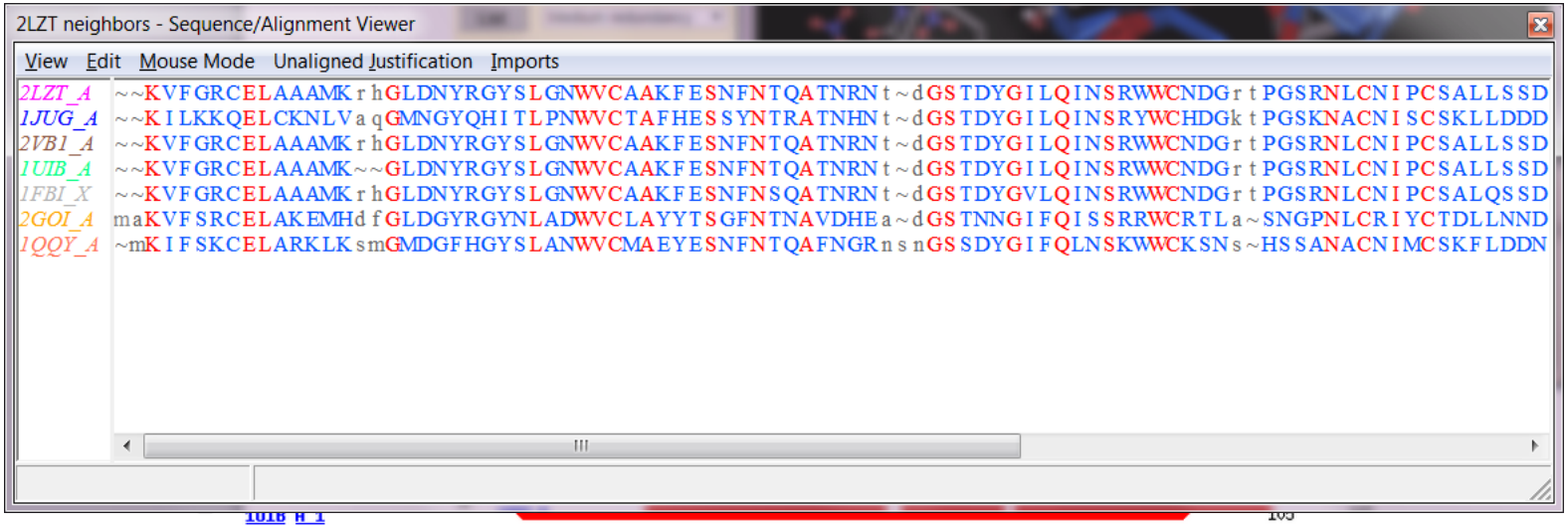
Pesquisa estrutural com (original)VAST



Visualização do alinhamento com o software Cn3D



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



Previsão da estrutura secundária das proteínas

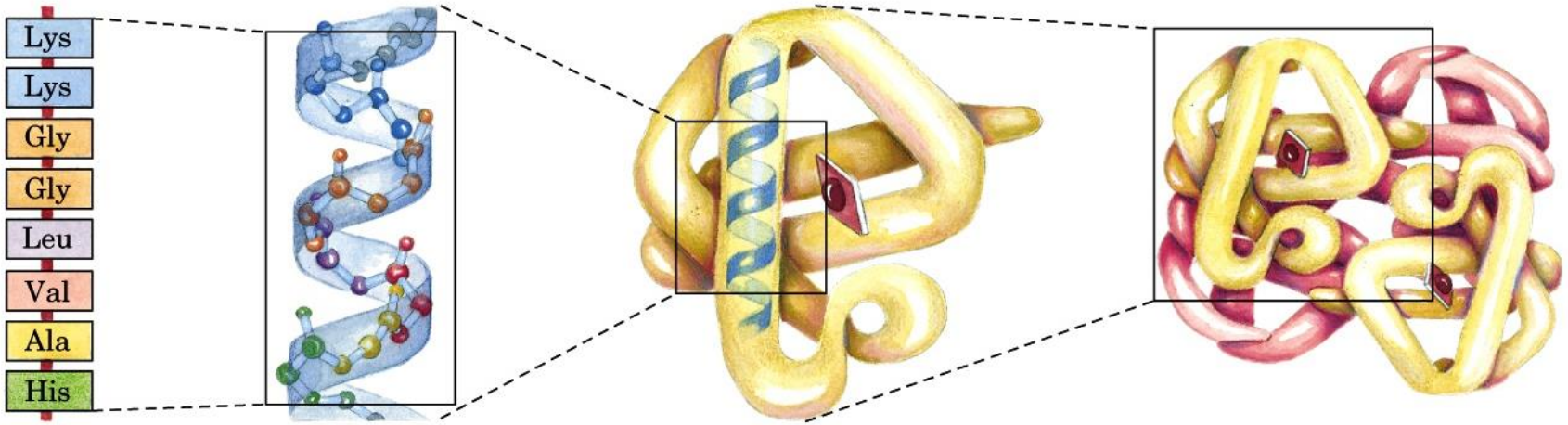
Níveis de organização da estrutura das proteínas

Estrutura primária

Estrutura secundária

Estrutura terciária

Estrutura quaternária



Sequência de aminoácidos

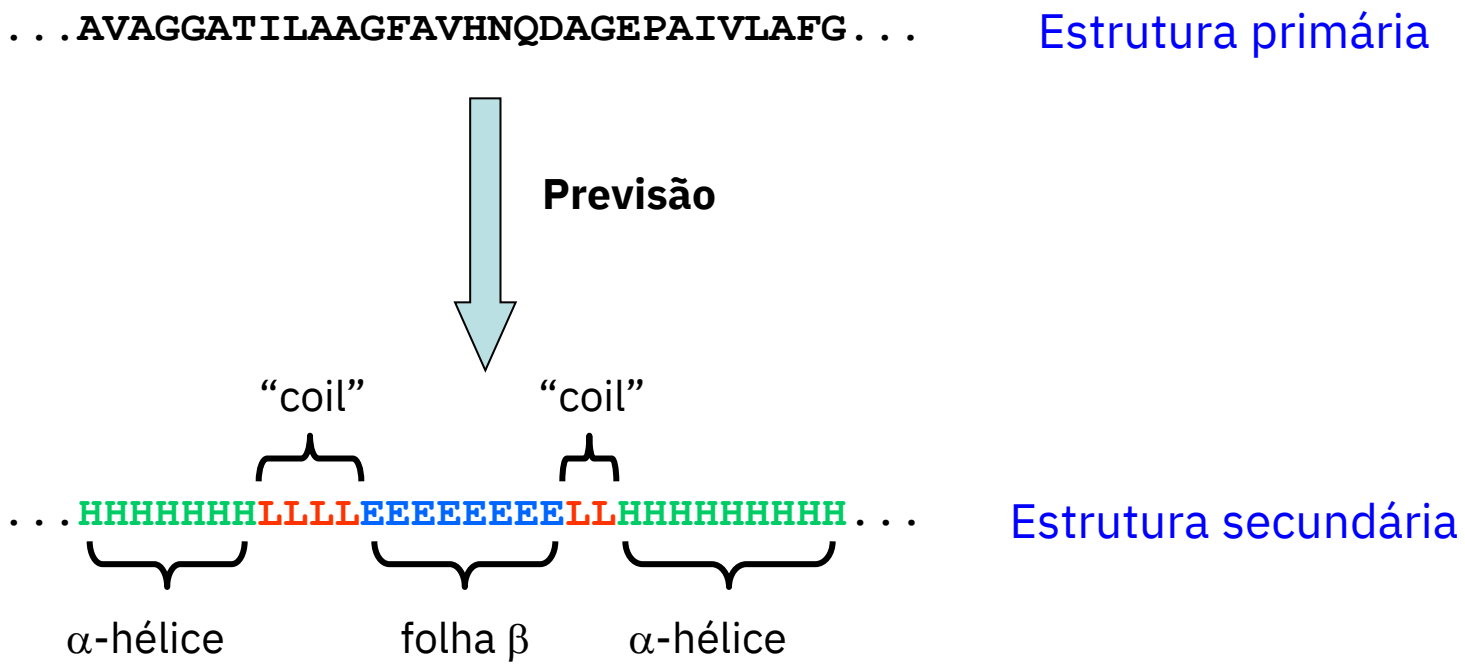
α-hélice

Cadeia polipetídica

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%)
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 abordadagens distintas, tais como neural networks e nearest-neighbor. Produzem os melhores resultados. (Precisão 85-90%)
http://210.44.144.20:82/protein_PSRSM/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_PSRSM/default.aspx

Shuping Zhu and Yihui Liu 2019 *J. Phys.: Conf. Ser.* **1237** 052005

Exemplo de previsão com o programa PHD

Rel: fiabilidade global da previsão (0-9)

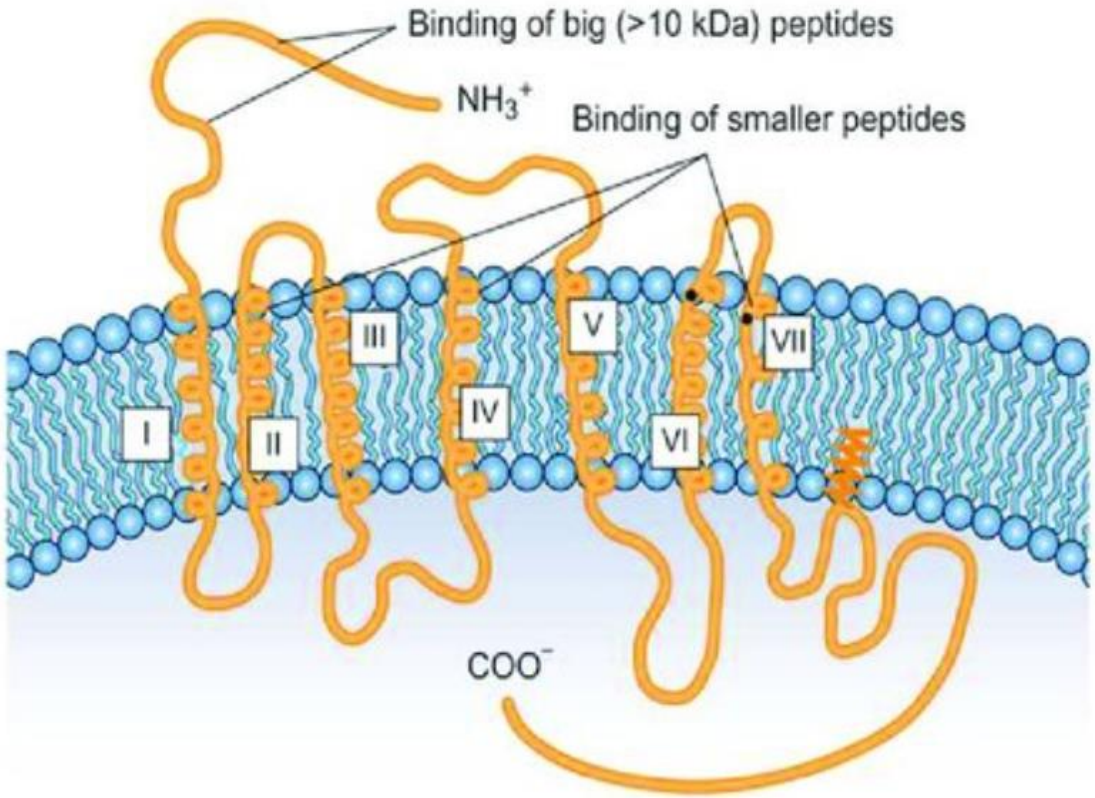
```
             ....,....1....,....2....,....3....,....4....,....5....,....6
        AA   |MERYENLFAQLNDRREGAFVPFVTLGDPGIEQSLKIIDTLIDAGADALELGVPFSDPLAD|
        PHD  |  HHHHHHHHHHH    EEEEEEE     HHHHHHHHHHHHH   EEEE           |
        Rel  |934899999996348872799842489984587999999973996689447677846 89|
detail:
        prH- |036899999987531100000000000167889999999983000000000011132 10
        prE- |000000000000000005798865300000000000000000000178863111000000
        prL- |963100000012368883100123689983211000000001699720036877886789
subset: SUB  |L..HHHHHHHHH..LLL.EEEE...LLLL.HHHHHHHHHHHH.LLLEEE..LLLLL.LLL|


             ....,....7....,....8....,....9....,....10...,....11...,....1
        AA   |GPTIQNANLRAFAAGVTPAQCFEMLALIREKHPTIPIGLLMYANLVFNNGIDAFYARCEQ|
        PHD  | HHHHHHHHHHHH    HHHHHHHHHHHHH      EEEEEE  HHHH  HHHHHHHHHH|
        Rel  |737899999999828962999999999997289993899883422441255999999999|
detail:
        prH- |258899999999841115899999999998410000000002356643269999999 8
        prE- |000000000000000000000000000000000000389988511000000000000000
        prL- |741100000000158873000000000015899861000135542245720000000 0
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)

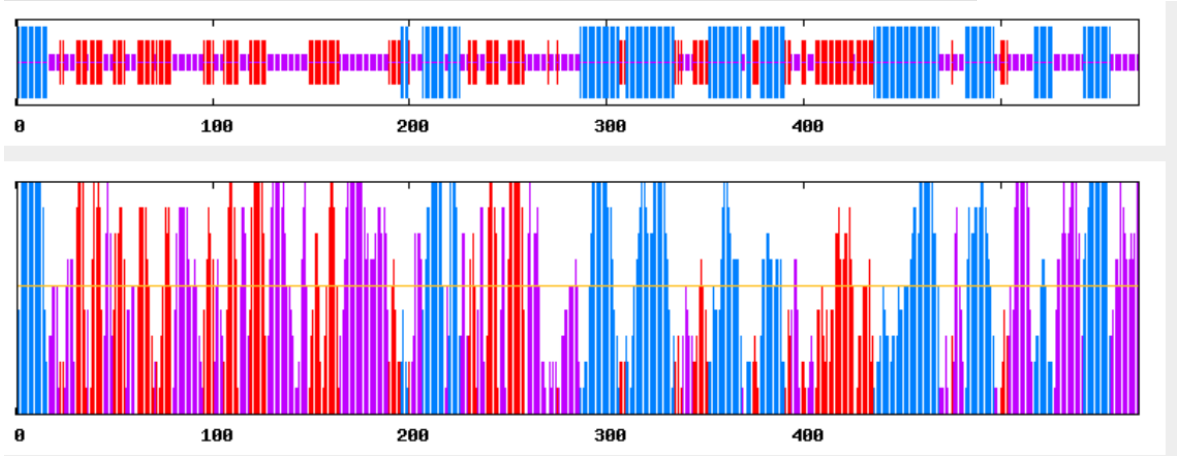
prL: probabilidade do resíduo estar em conformação de “coil” (0-9)


Binding of big (>10 kDa) peptides
NH_3^+
Binding of smaller peptides
I
II
III
IV
V
VI
VII
COO^-

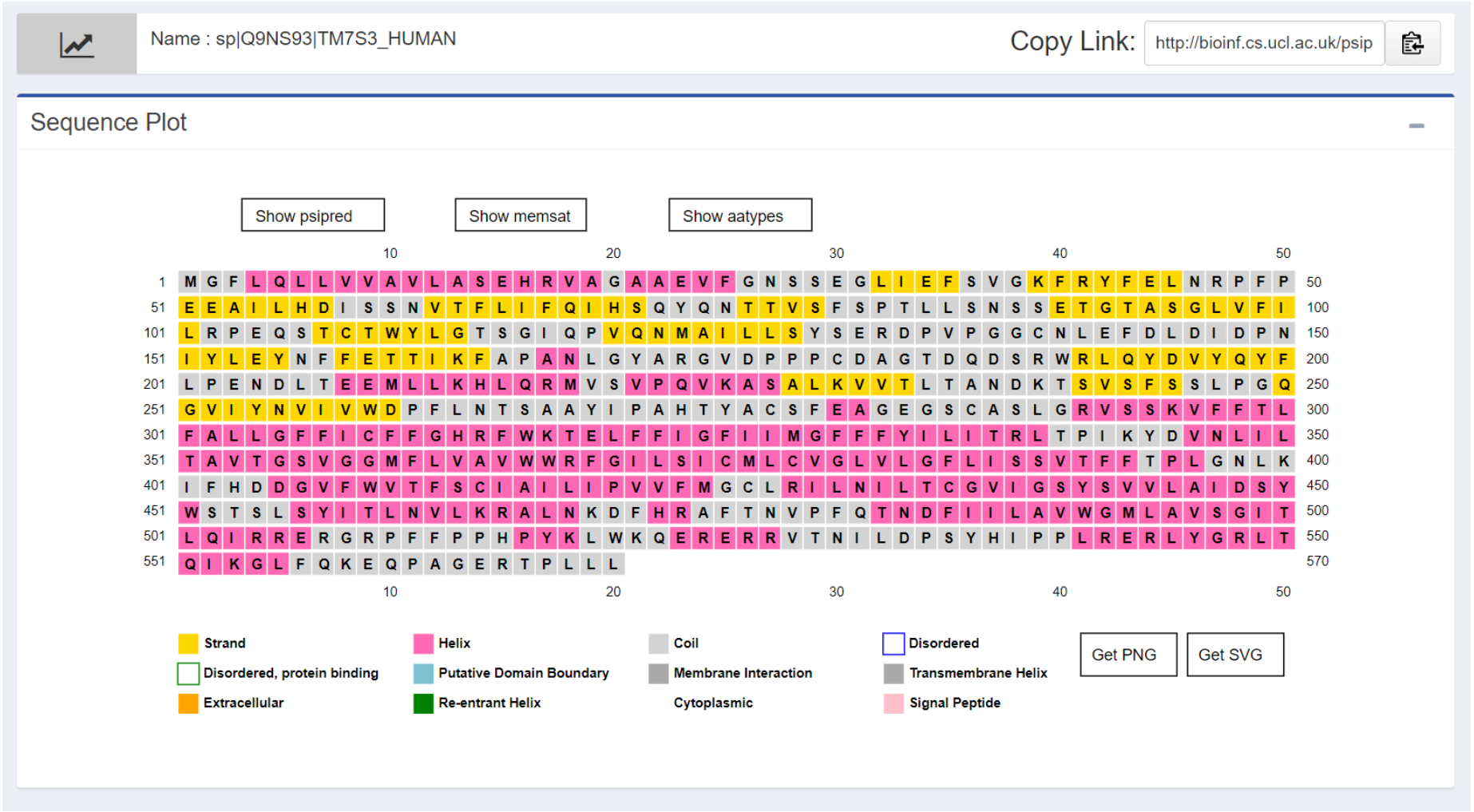
Previsão com o programa PHD

```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MGFLQLLVVAVLASEHRVAGAAEVFGNSSEGLIEFSVGKFRYFELNRPFPEEAILHDISSNVTFLIFQIH
CchHHHHHHHHHHHHHhccCCCCceecCCCCcEEEEeeceEEEEEecCCCceEEEEEcccCCcEEEEEEEe
SQYQNTTVSFSPTLLSNSSETGTASGLVFILRPEQSTCTWYLGTSGIQPVQNMAILLSYSERDPVPGGCN
eceeeEEEEEccCCCCCCCCCCCcccceEEEEcCCcceeEEEEEeCCCCceEEEEEEEecCCCCCCCCCC
LEFDLDIDPNIYLEYNFFETTIKFAPANLGYARGVDPPPCDAGTDQDSRWRLQYDVYQYFLPENDLTEEM
ccccCCCCeeeEEEeeeEEEEEEeccCCCCCCCCCCCCCCCCCCCCCceEEeeehhhheccCCCChHH
LLKHLQRMVSVPQVKASALKVVTLTANDKTSVSFSSLPGQGVIYNVIVWDPFLNTSAAYIPAHTYACSFE
HHHHHHHHcChHHHHhcCCceEEEecCCceEEEEEecCCcEEEEEEEecCCCCCCCcccecccceccc
AGEGSCASLGRVSSKVFFTLFALLGFFICFFGHRFWKTELFFIGFIIMGFFFYILITRLTPIKYDVNLIL
cCCCCCchhhhHHHHHHHHHHHHHHHeeehhhhHHHHHHHHHHHHHHHHHHHHhhheececCccceeeEEE
TAVTGSVGGMFLVAVWWRFGILSICMLCVGLVLGFLISSVTFFTPLGNLKIFHDDGVFWVTFSCIAILIP
eehHHhhhHHHHHHHHHHhcchhheeeehHHHHHHHHHHHhheeecCCcceeecccceeeEeeeeeEEEE
VVFMGCLRILNILTCGVIGSYSVVLAIDSYWSTSLSYITLNVLKRALNKDFHRAFTNVPFQTNDFIILAV
EEEEEEceeeEEEEeehhhhHhhhhhhhHHHHHHHHHHHHHHHHHHHHHcccccceCCCCchHHHHHHH
WGMLAVSGITLQIRRERGRPFFPPHPYKLWKQERERRVTNILDPSYHIPPLRERLYGRLTQIKGLFQKEQ
HHHHHHhccceeeecCcCCCCCCCCCchhhHHHHHhhccCCCCCCCCCCCCcHHHHHHHHHHHHHhccCC
PAGERTPLLL
CCCCCCCCCC
```

```
PHD :
   Alpha helix      (Hh) :   186 is  32.63%
   3₁₀ 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%
```

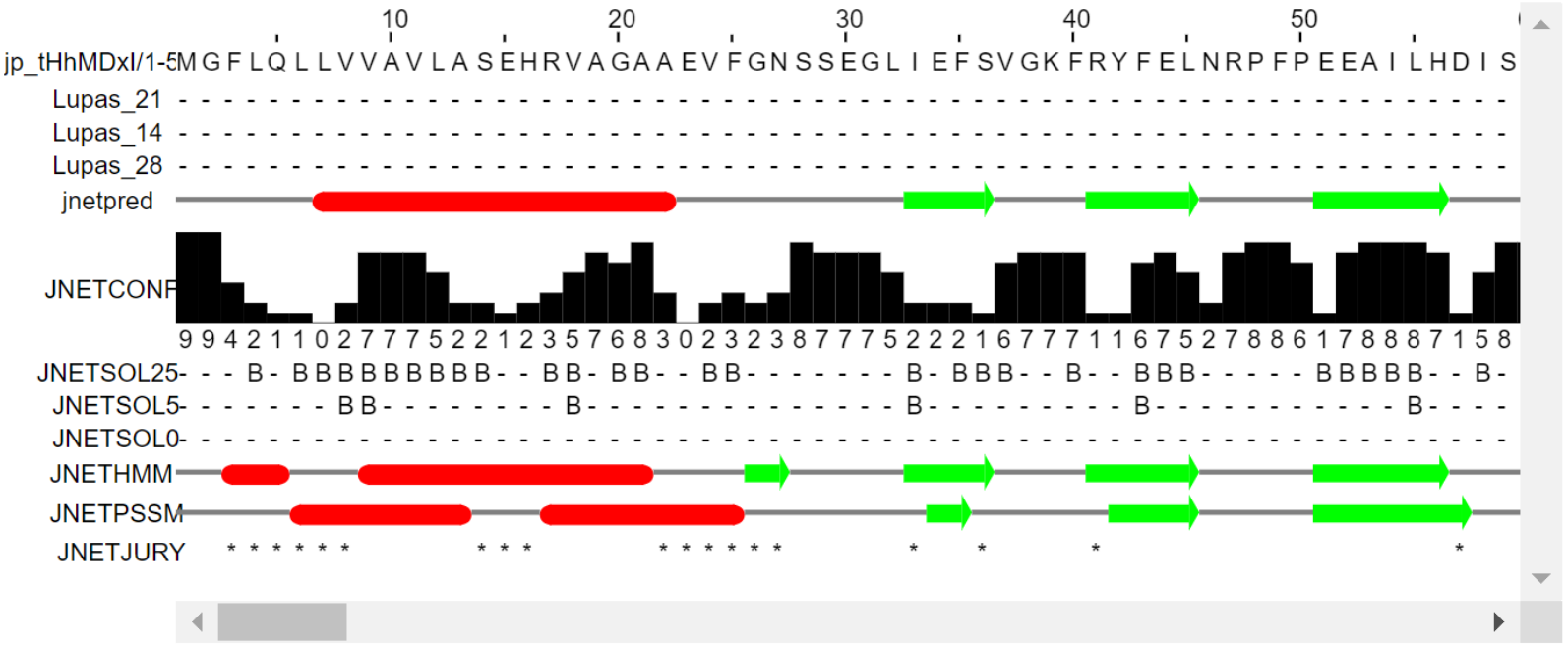


Previsão com o programa PSIPRED



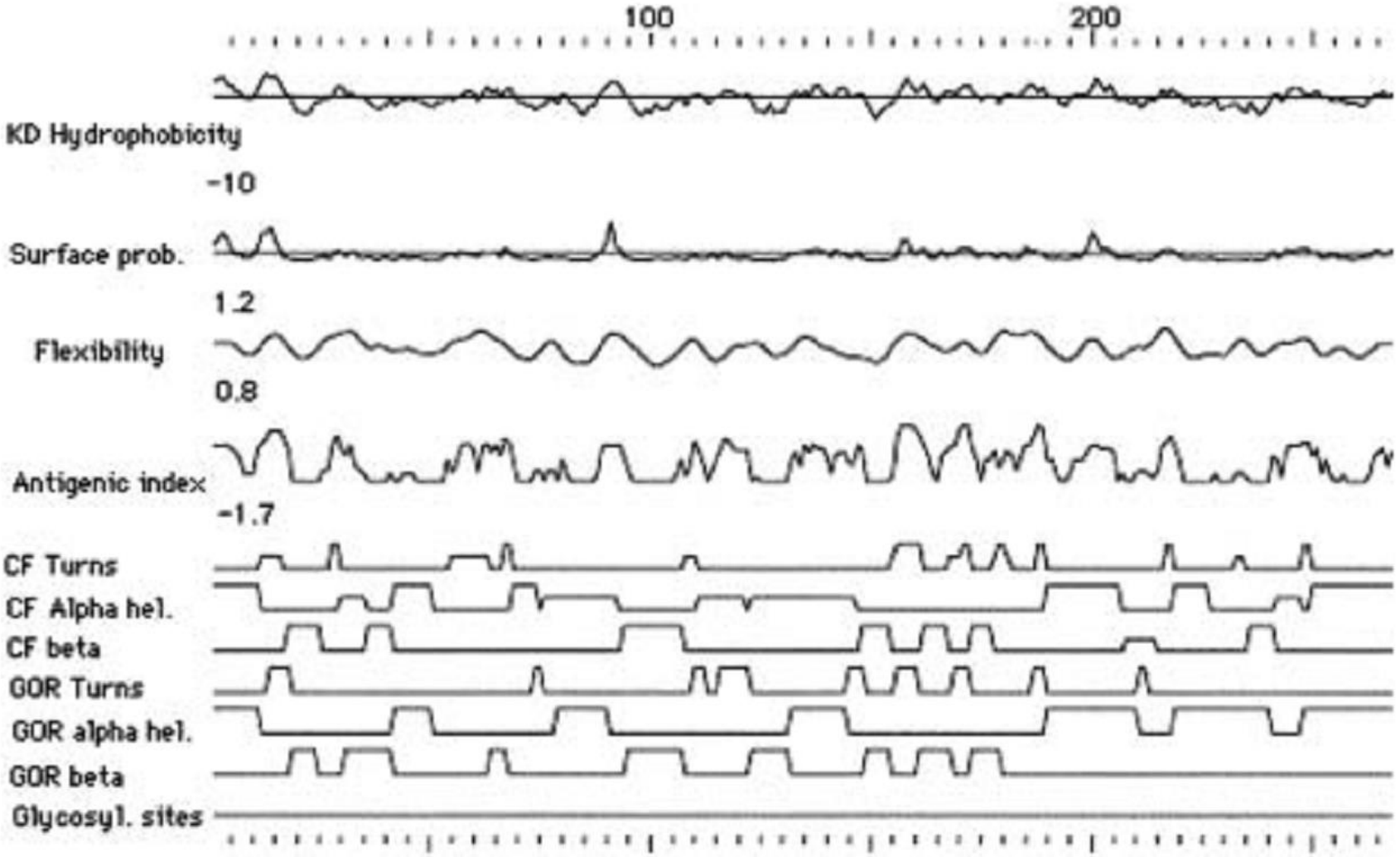
Previsão com o programa Jpred4

- 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).

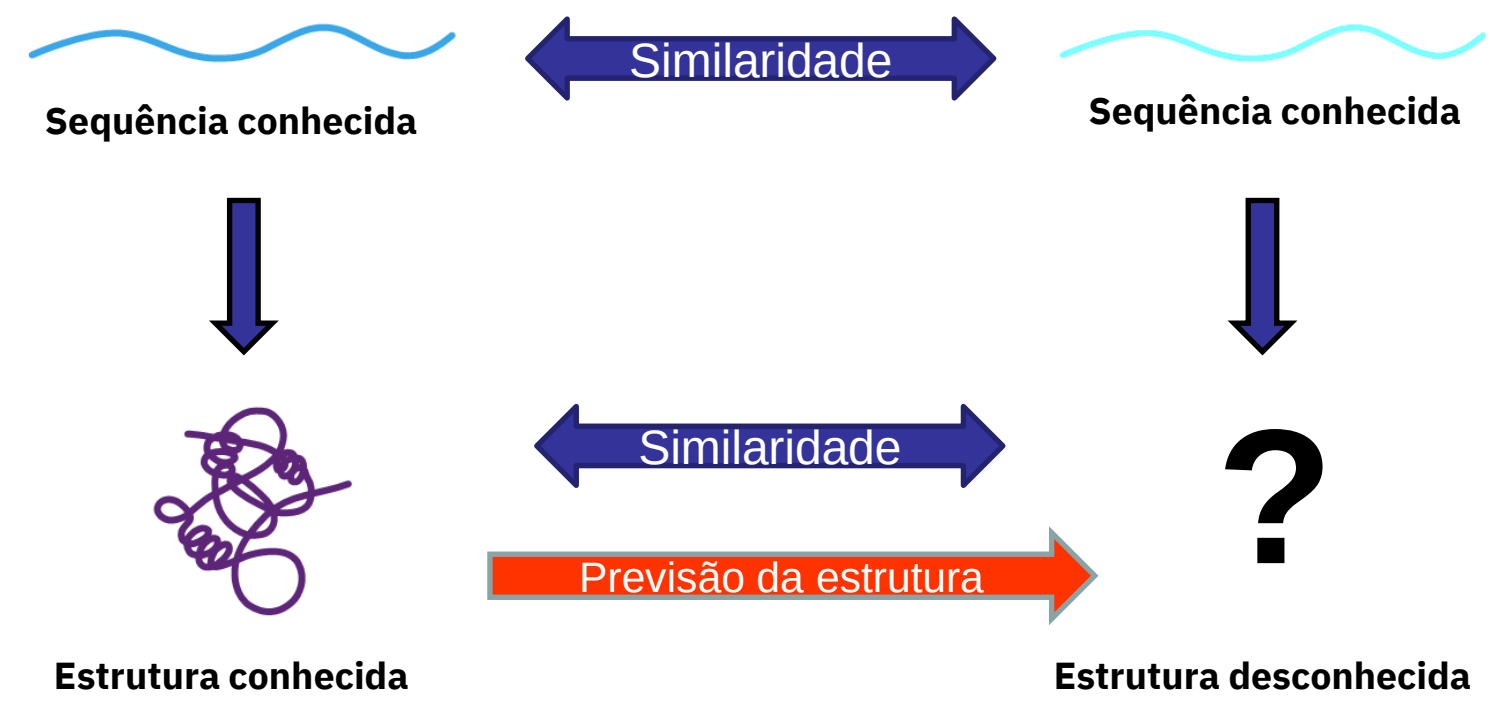
•Dos vários métodos para previsão de estrutura, a modelação por homologia é aquele que dá melhores resultados

•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!

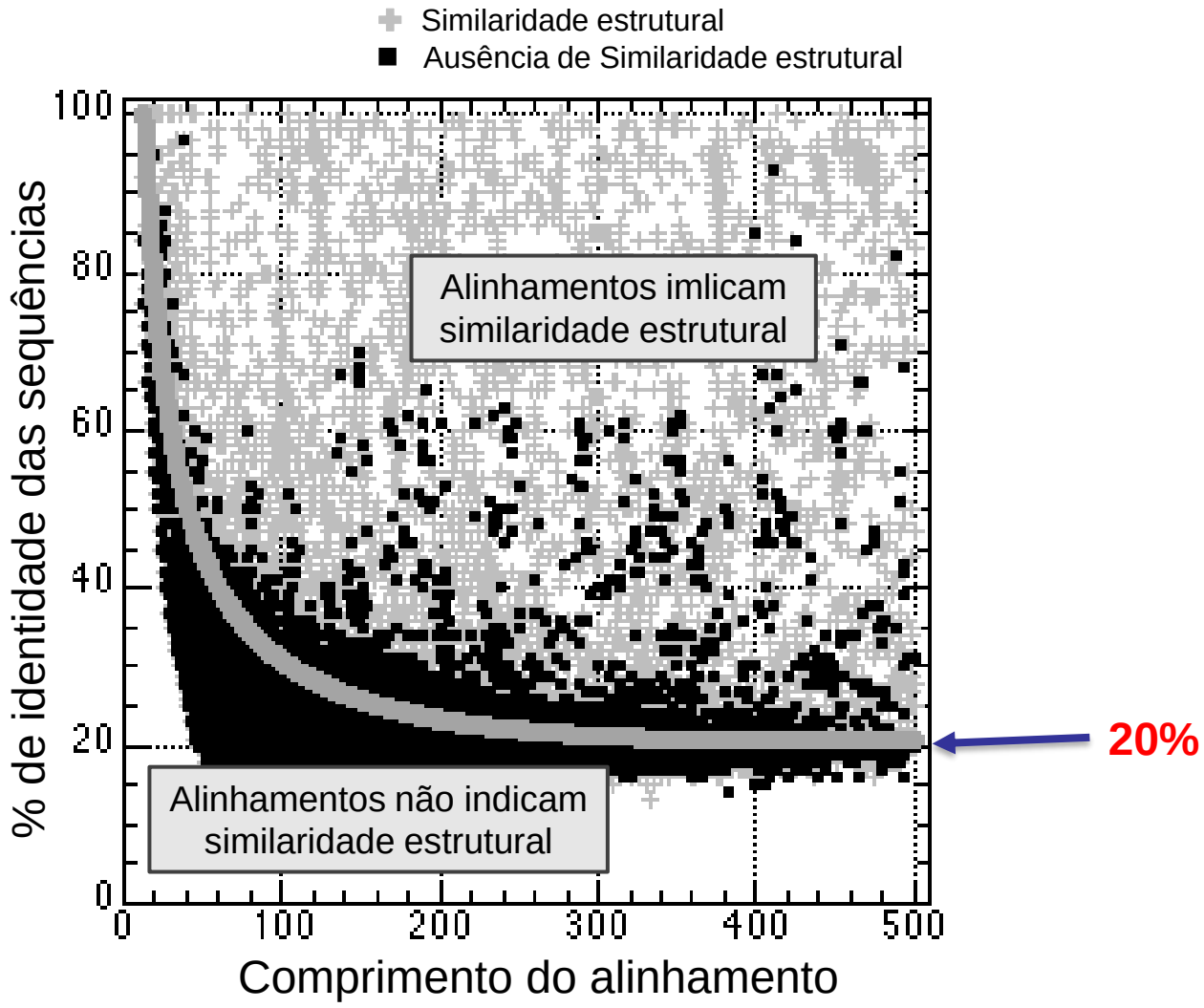
A estrutura das proteínas é determinada pela sua sequência



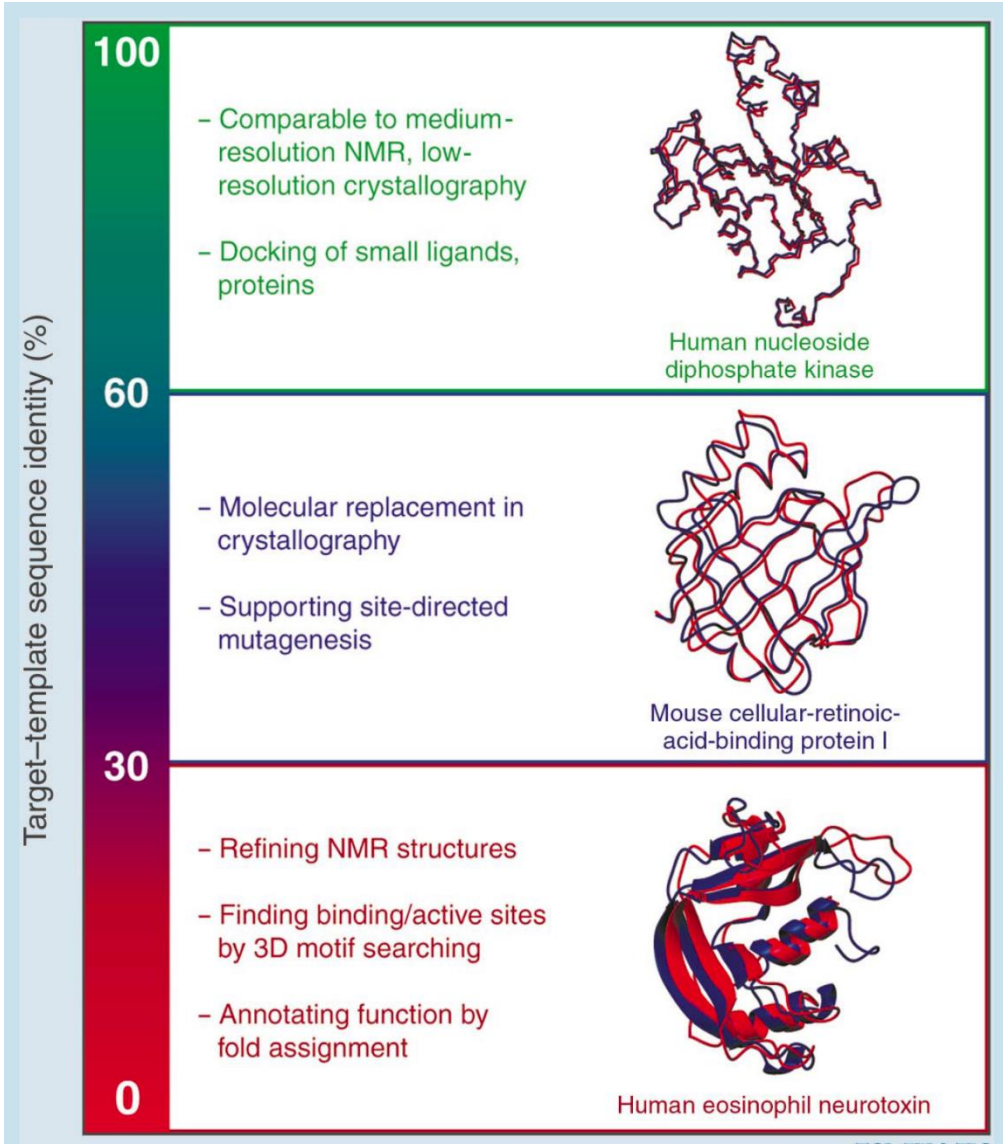
Sequências similares implicam estruturas similares, logo:

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

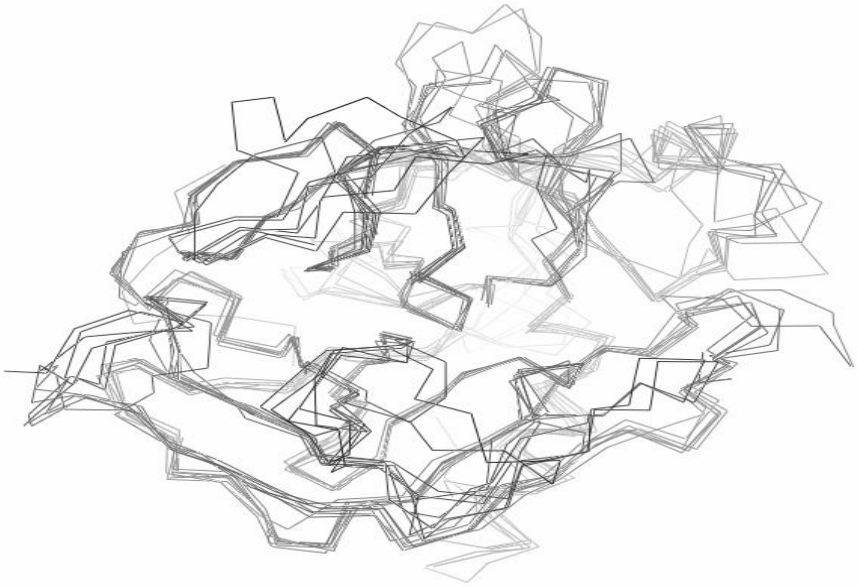
Qual a % de identidade mínima aceitável para existência de similaridade estrutural ?



Impacto da similaridade na qualidade e utilidade do modelo



Alinhamento estrutural das proteases de serina



```
1BOF:A(218:1-218)          IVGGRRARPHAWPFMVSLQL---A-----GGHFCGATLIAPNFVMSAAHCV---------ANVNVR--AVRVVLGAHNLSRREP-TRQV
3BTH:E(223:1-221)          IVGGYTCGANTVPYQVSLNS---G------YHFCGGSLINSQWVVSAAHCY-------------KS--GIQVRLGEDNINVVEG-NEQF
1GMD:A(241:14-239)         IVNGEEAVPGSWPWQVSLQDkt-G------FHFCGGSLINENWVVTAAHCG-------------VTt-SDVVVAGEFDQGSSSE-KIQK
1H8D:H(260:1-248)          IVEGSDAEIGMSPWQVMLFRkspQ------ELLCGASLISDRWVLTAAHCLlyppwdkn----FTenDLLVRIGKHSRTRYERniEKI
1TOC:B(259:1-255)          IVEGQDAEVGLSPWQVMLFRkspQ------ELLCGASLISDRWVLTAAHCLlyppwdkn----FTvdDLLVRIGKHSRTRYERkVEKI
1HYL:A(230:1-227)          IINGYEAYTGLFPYQAGLDI---Tlqdqr-RVWCGGSLIDNKWILTAAHCVhd----------AV--SVVVYLGSAVQYE----GEAV
1TON:_(235:11-232)         ----------SQPWQVAVIN---E--------YLCGGVLIDPSWVITAAHCY-------------SN--NYQVLLGRNNLFKDEP-FAQR
2PKA:B(152:5-149)          ----------------------------------------------------------------------------------------

1BOF:A(218:1-218)          FAVQRIFED-GYD-------------PV-NLLNDIVILQLNGSATINANVQVAQLPA----QGR-----RLGNGVQCLAMGWGL--LGR
3BTH:E(223:1-221)          ISASKSIVHpSYN-------------SN-TLNNDIMLIKLKSAASLNSRVASISLPT------S-----CASAGTQCLISGWGN--TKS
1GMD:A(241:14-239)         LKIAKVFKNsKYN-------------SL-TINNDITLLKLSTAASFSQTVSAVCLPSas----D-----DFAAGTTCVTTGWGL--TRY
1H8D:H(260:1-248)          SMLEKIYIHpRYN-------------WReNLDRDIALMKLKKPVAFSDYIHPVCLPD------RetaasLLQAGYKGRVTGWGN--LKE
1TOC:B(259:1-255)          SMLDKIYIHpRYN-------------WKeNLDRDIALLKLKRPIELSDYIHPVCLPD------KqtaakLLHAGFKGRVTGWGN--RRE
1HYL:A(230:1-227)          VNSERIISHsMFN-------------PD-TYLNDVALIKI-PHVEYTDNIQPIRLPSgeel--N-----NKFENIWATVSGWGQ--SN-
1TON:_(235:11-232)         RLVRQSFRHpDYIplivtrdteqpv-H-DHSNDLMLLHLSEPADITGGVKVIDLPT------K-----EPKVGSTCLASGWGStnPSE
2PKA:B(152:5-149)          ----------------------------DYSHDLMLLRLQSPAKITDAVKVLELPT------Q-----EPELGSTCEASGWGS--IEP

1BOF:A(218:1-218)          ---NR-----GIASVLQELNVTVVT--SLC--------RRSNVCTLVRG---R--QAGVC--FGDSGSPLVC--N------GLIHGIA
3BTH:E(223:1-221)          s--GT-----SYPDVLKCLKAPILSdsSCKsaypgqi-TSNMFCAGYLE---G--GKDSC--QGDSGGPVVC--S------GKLQGIV
1GMD:A(241:14-239)         a--N-------TPDRLQQASLPLLSntNCKkywgtki-KDAMICAGAS-------GVSSC--MGDSGGPLVC--Kknga--WTLVGIV
1H8D:H(260:1-248)          t----------GQPSVLQVVNLPIVErpVCKastriri-TDNMFCAGYKPdegK--RGDAC--EGDSGGPFVM--KspfnnrWYQMGIV
1TOC:B(259:1-255)          t--WTtsVAeVQPSVLQVVNLPLVErpVCKastriri-TDNMFCAGYKPgegK--RGDAC--EGDSGGPFVM--KspynnrWYQMGIV
1HYL:A(230:1-227)          ----------TDTVILQYTYNLVIDndRCAqeyppgiiVESTICGDT-S----D--GKSPC--FGDSGGPFVLsdK------NLLIGVV
1TON:_(235:11-232)         m--VV-----SHDLQ--CVNIHLLSneKCIetykdnv-TDVMLCAGE-----MegGKDTCag--DSGGPLIC--D------GVLQGIT
2PKA:B(152:5-149)          gpdDF-----EFPDEIQCVQLTLLQntFCAdahpdkv-TESMLCAGYLP---G--GKDTC--MGDSGGPLIC--N------GMWQGIT

1BOF:A(218:1-218)          SFVR--GG-CASGLYPDAFAPVAQFVNWIDSIIQ
3BTH:E(223:1-221)          SWGS---G-CAQKNKPGVYTKVCNYVSWIKQTIA
1GMD:A(241:14-239)         SWGS---S-TCSTSTPGVYARVTALVNWVQQTLA
1H8D:H(260:1-248)          SWGE---G-CDRDGKYGFYTHVFRLKKWIQKVID
1TOC:B(259:1-255)          SWGE---G-CDRDGKYGFYTHVFRLKKWIQKVID
1HYL:A(230:1-227)          SFVSga-G-CESG-KPVGFSRVTSYMDWIQQNTG
1TON:_(235:11-232)         SGGA---TpCAKPKTPAIYAKLIKFTSWIKKVMK
2PKA:B(152:5-149)          SWGH---TpCGSANKPSIYTKLIFYLDWIDDTIT
```

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.

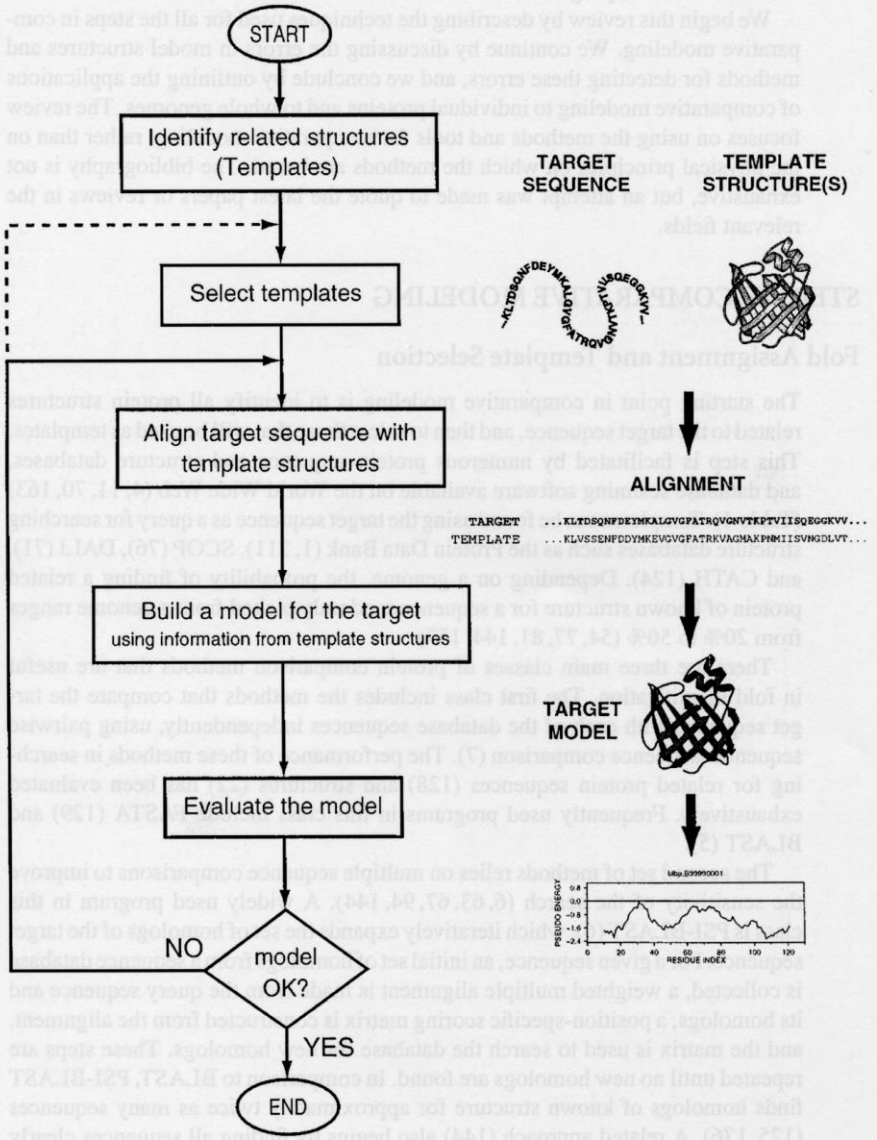
• Optimização das cadeias laterais da proteína por selecção de rotâmeros adequados para cada resíduo e localização.

• Optimização da estrutura dos “loops” existentes no modelo.

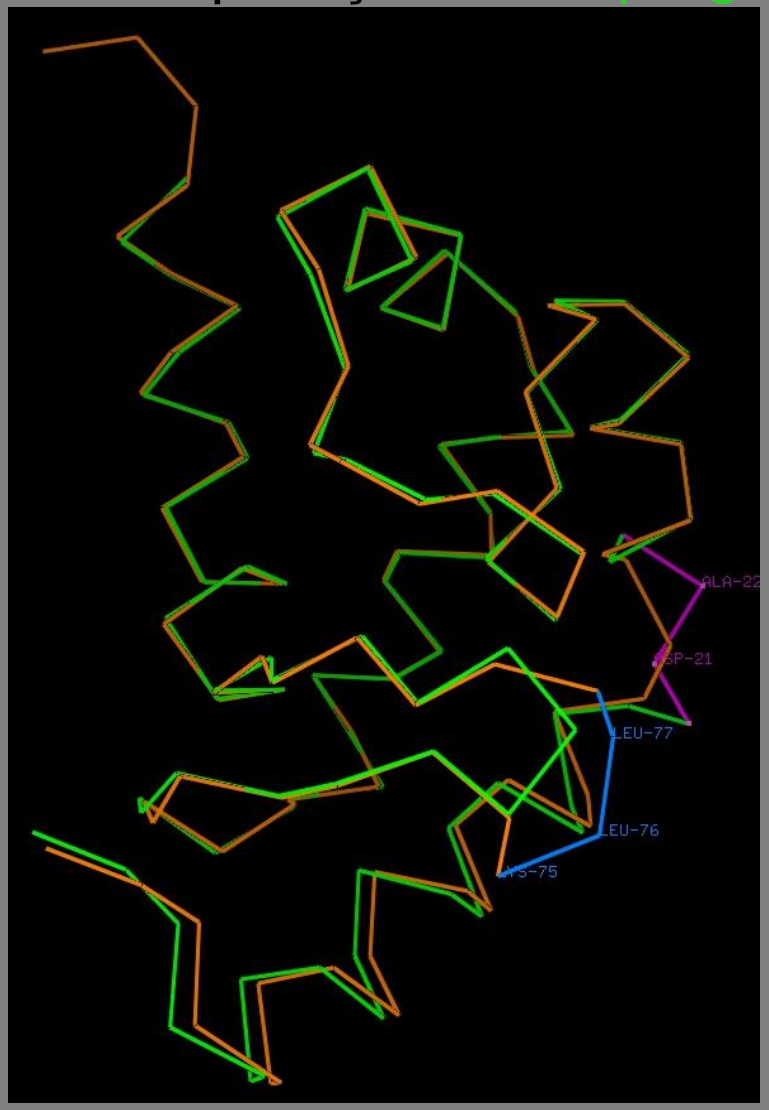
• Optimização global da estrutura por minimização e/ou dinâmica molecular

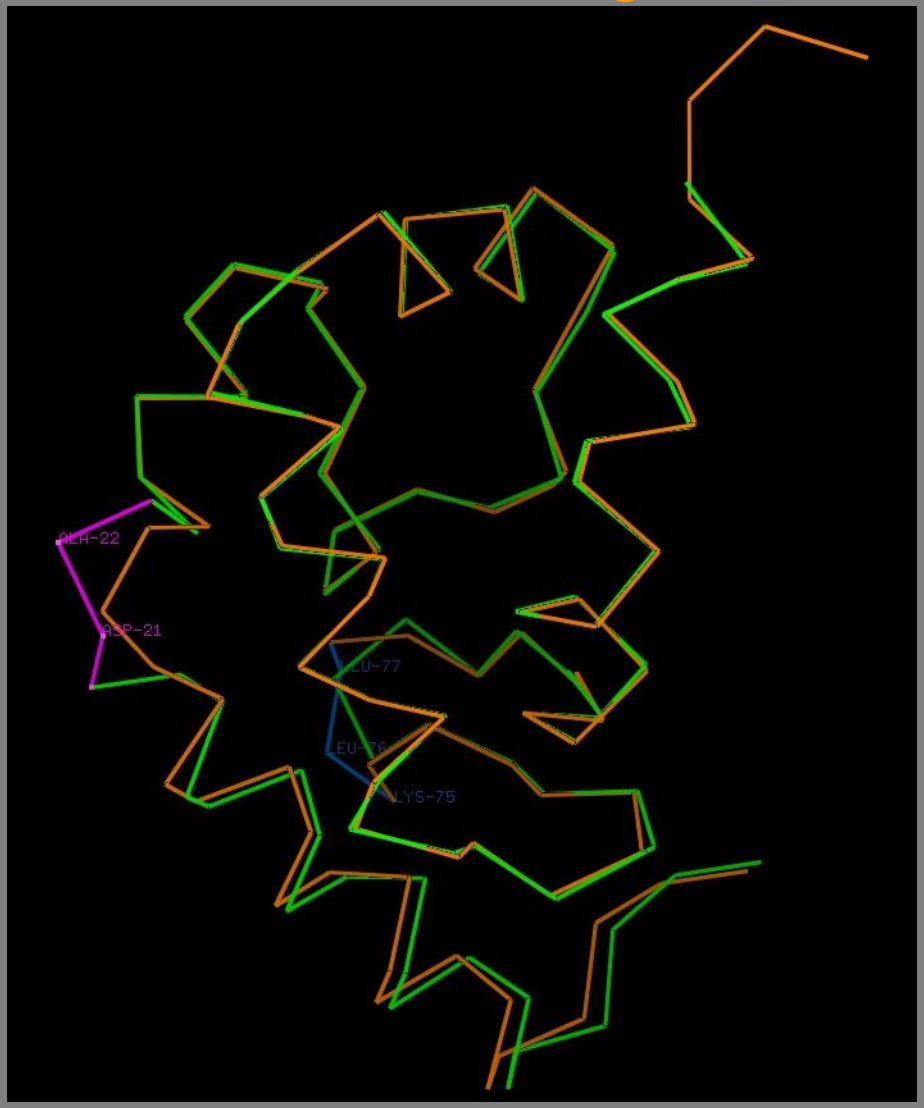
• Validação do modelo por critérios estereoquímicos e fenomenológicos

• Se necessário, corrigir os alinhamentos e voltar a produzir modelos até estes serem correctamente validados


START
Identify related structures (Templates)
Select templates
Align target sequence with template structures
Build a model for the target using information from template structures
Evaluate the model
NO
model OK?
YES
END
TARGET SEQUENCE
TEMPLATE STRUCTURE(S)
ALIGNMENT
TARGET ...KLTDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVV...
TEMPLATE ...KLVSSENFDDYMKEVGVGFATRKVAGMAKPNMIISVNGDLVT...
TARGET MODEL
PSEUDO ENERGY
RESIDUE INDEX

Comparação da criptogeína com o modelo da oligandrina






```
::** .* : :  .*.. :.  *.  *::***:**:....***. *** ** ** *:*:**:*   * .:*:* ...   ..: ****  **.*.  *: :
TACTASQQTAAYKTLVSILSDASFNQCSTDSGYSMLTAKALPTTAQYKLMCASTACNTMIKKIVTLNPPNCDLTVPT---SGLVLNVYSYANGFSNKCSSL---
ATCTDEQFSDSIIKLTPAIG--SVVGCTADSGFSMIPPTGLPTNDQYKKMCKSTNCKTLIKEIKDANVADCELDFSKLLPGSVPLNVYVLANNFDFVCAVVGSA
```

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..)

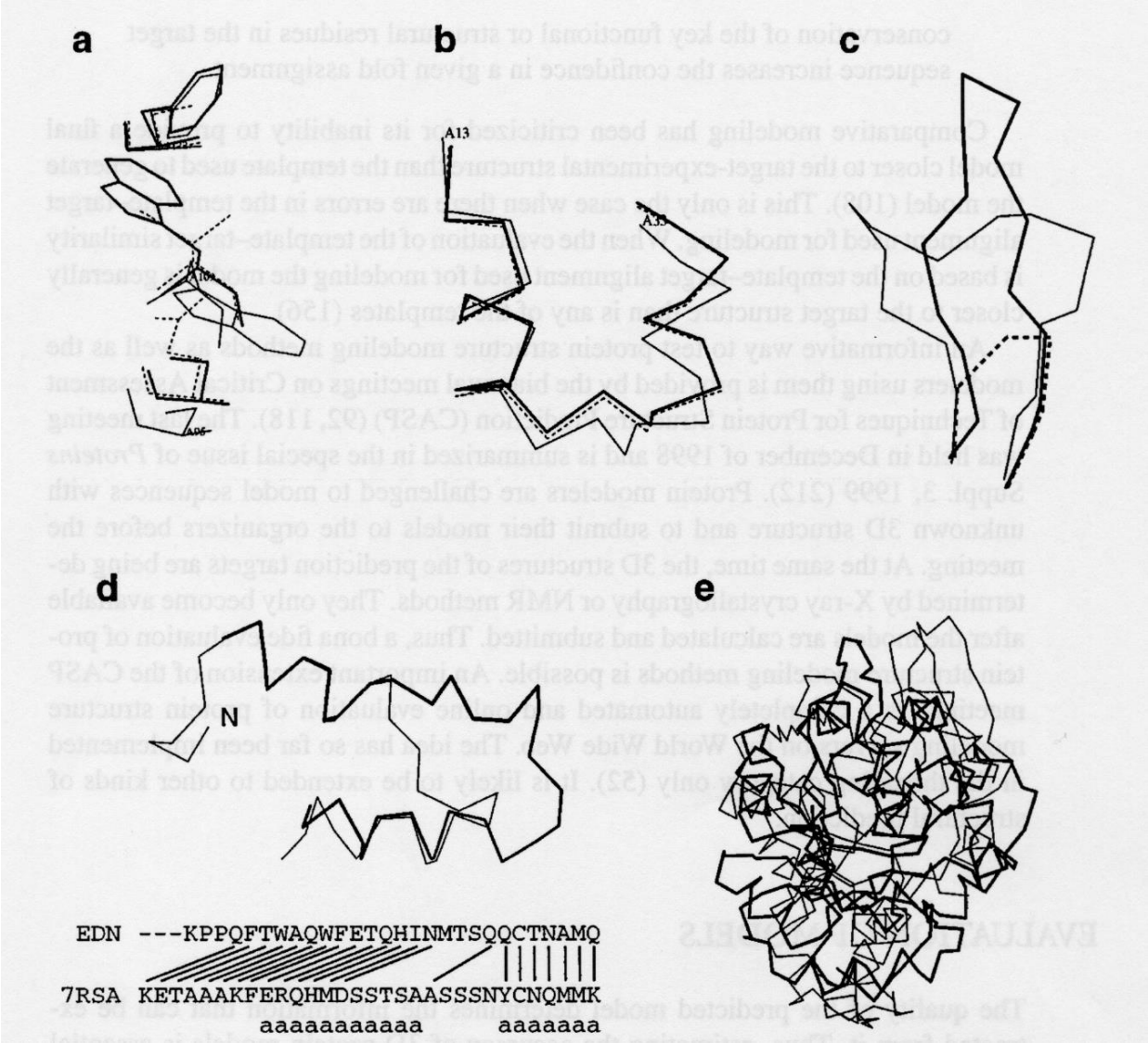
•Distorções e deslocações em zonas correctamente alinhadas. Podem ser devidas à divergência das sequências ou a artefactos na determinação da estrutura, como o empacotamento das moléculas no cristal.

• 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 estutura 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

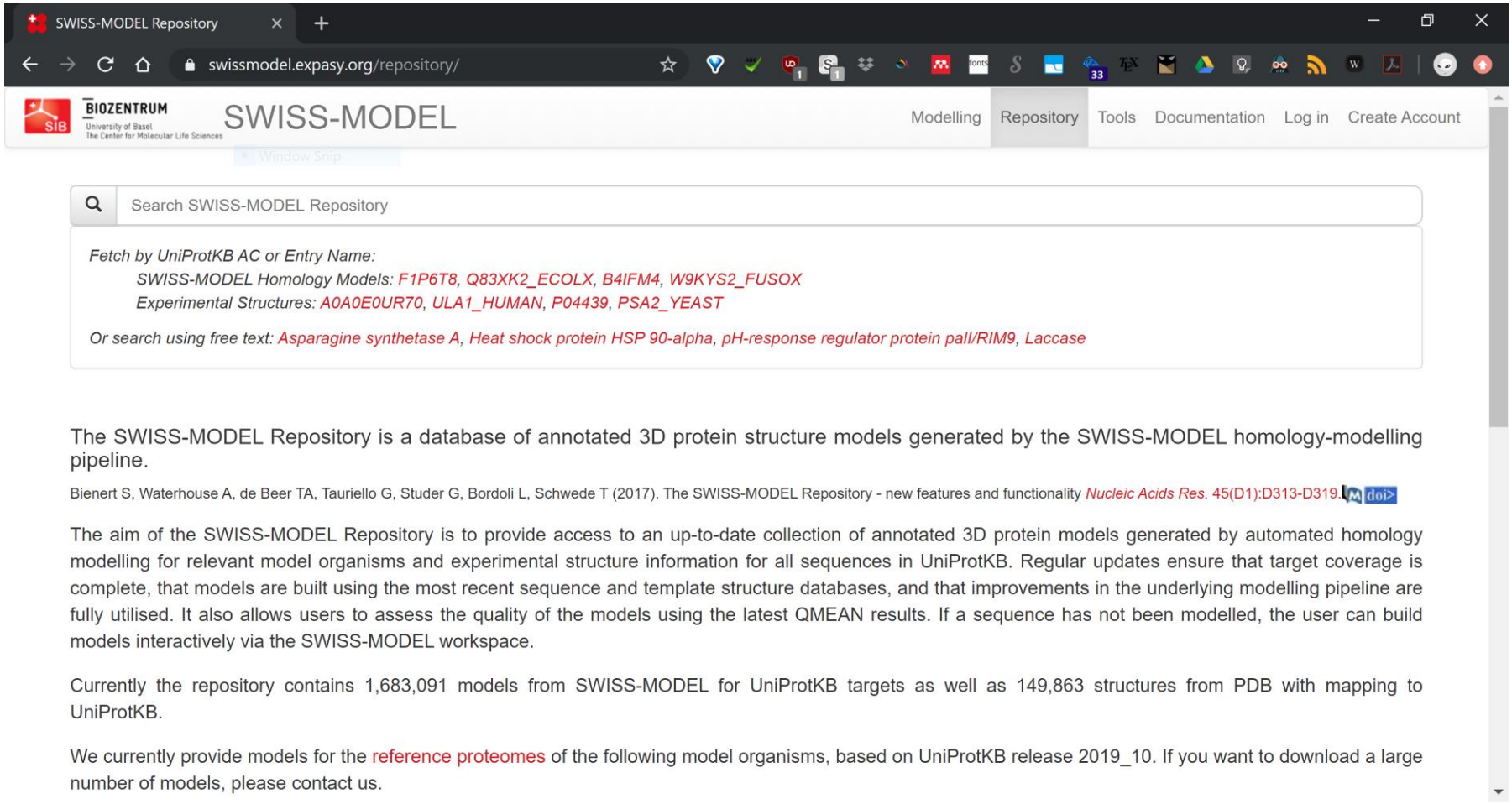
- **SWISS-MODEL -** http://swissmodel.expasy.org
- **Phyre2 -** http://www.sbg.bio.ic.ac.uk/phyre2/
- **I-Tasser -** http://zhanglab.ccmb.med.umich.edu/I-TASSER/
- **Raptor-X -** http://raptorx.uchicago.edu/
- **Hhpred -** http://toolkit.lmb.uni-muenchen.de/hhpred
- **Robetta - https://robetta.bakerlab.org**
- **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 considerer os indicadores de qualidade dos modelos

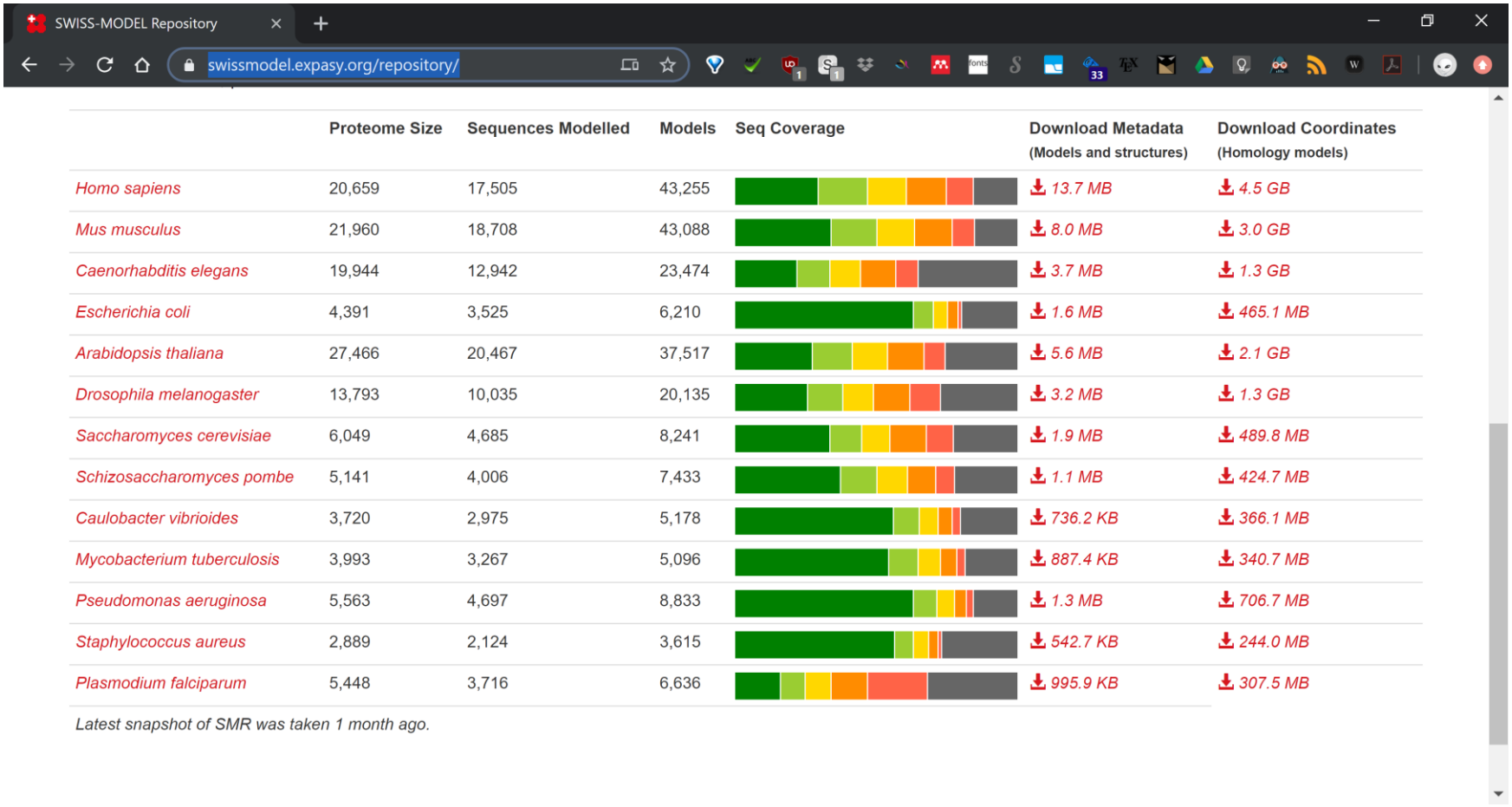
- SWISS Model Repository – https://swissmodel.expasy.org/repository/
- ModBase - https://modbase.compbio.ucsf.edu/i

SWISS MODEL repository



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

SWISS MODEL repository

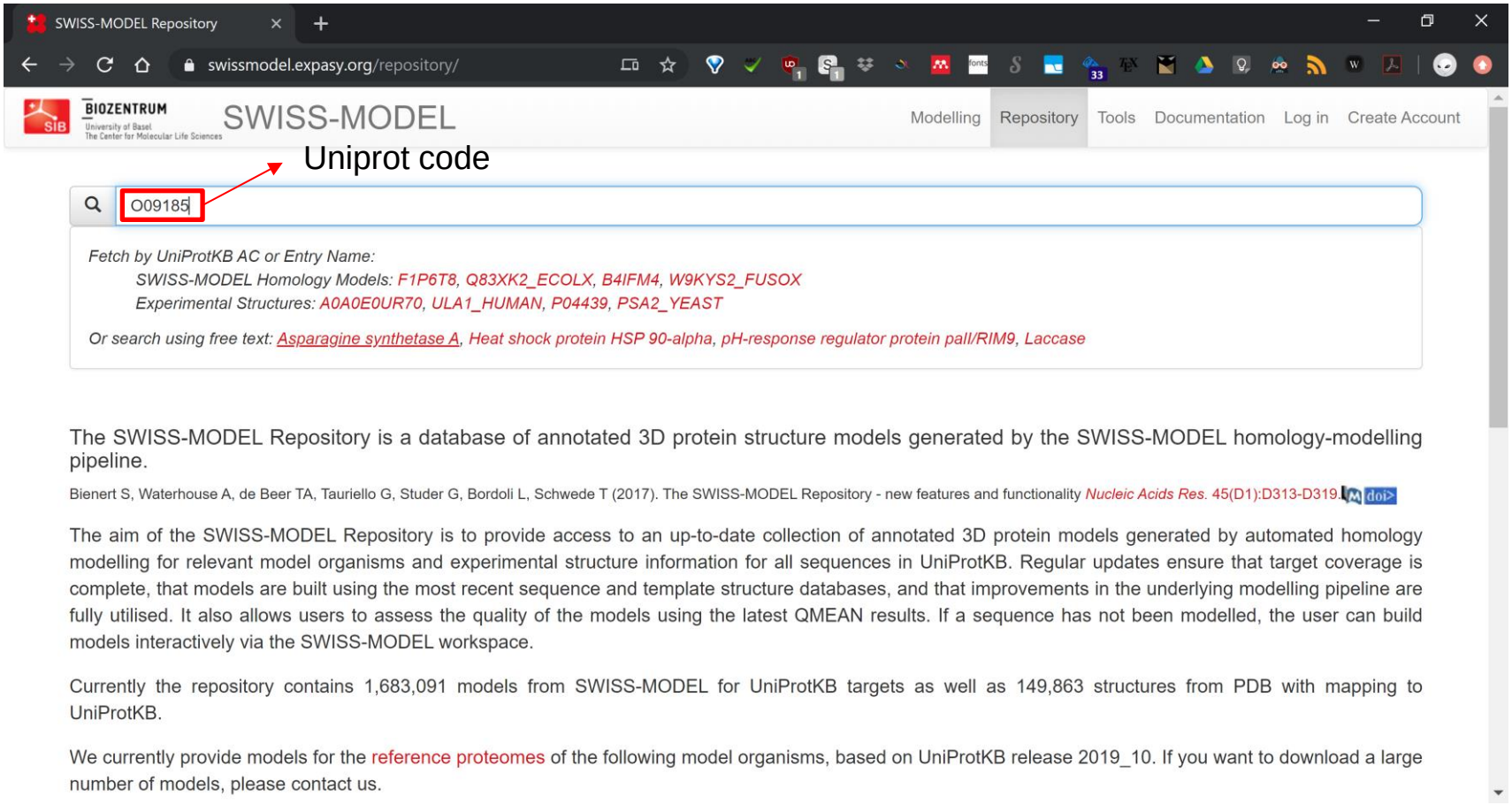


	Proteome Size	Sequences Modelled	Models	Seq Coverage	Download Metadata (Models and structures)	Download Coordinates (Homology models)
Homo sapiens	20,659	17,505	43,255		13.7 MB	4.5 GB
Mus musculus	21,960	18,708	43,088		8.0 MB	3.0 GB
Caenorhabditis elegans	19,944	12,942	23,474		3.7 MB	1.3 GB
Escherichia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

Latest snapshot of SMR was taken 1 month ago.

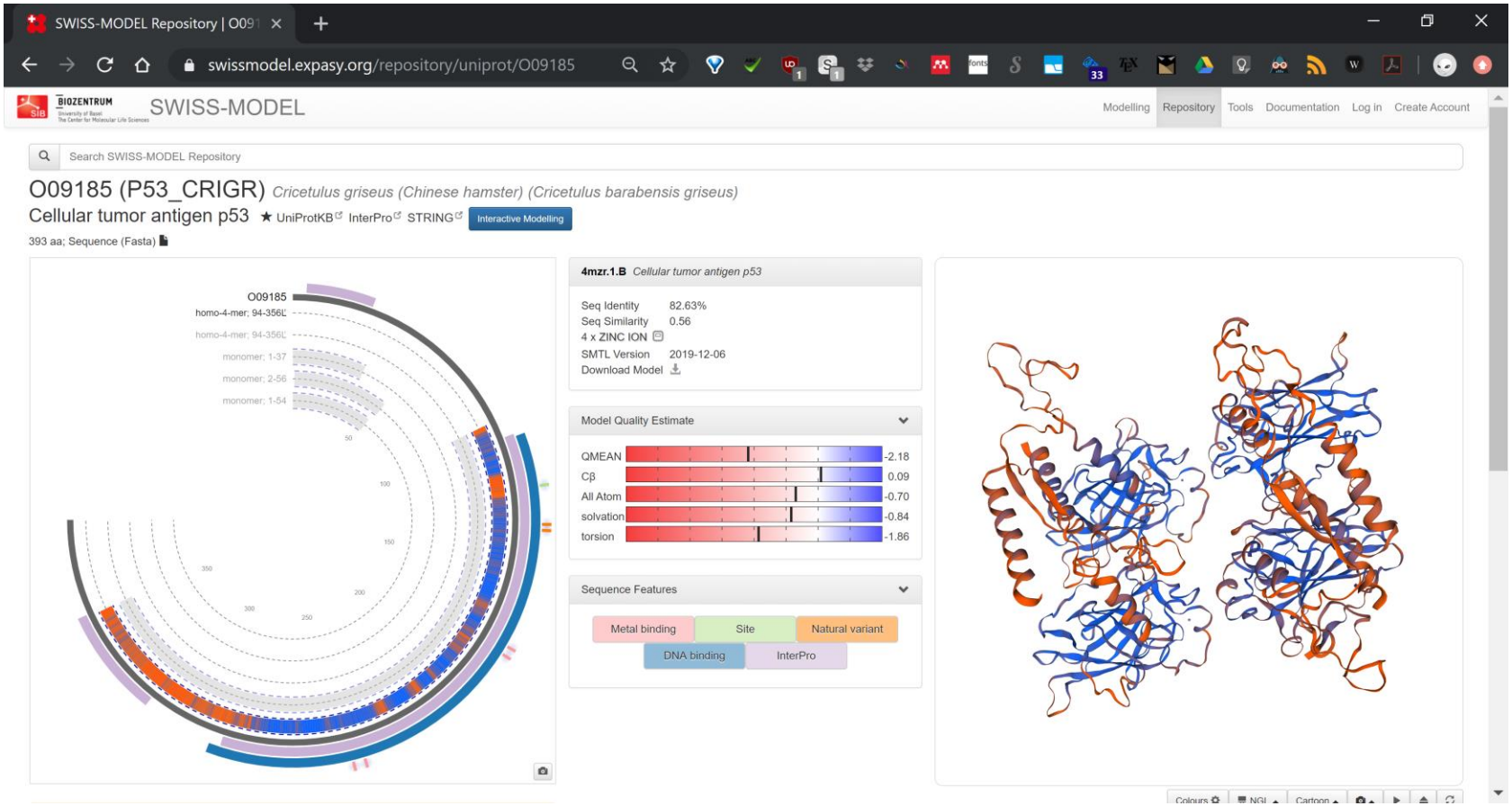
https://swissmodel.expasy.org/repository/

SWISS MODEL repository



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

SWISS MODEL repository

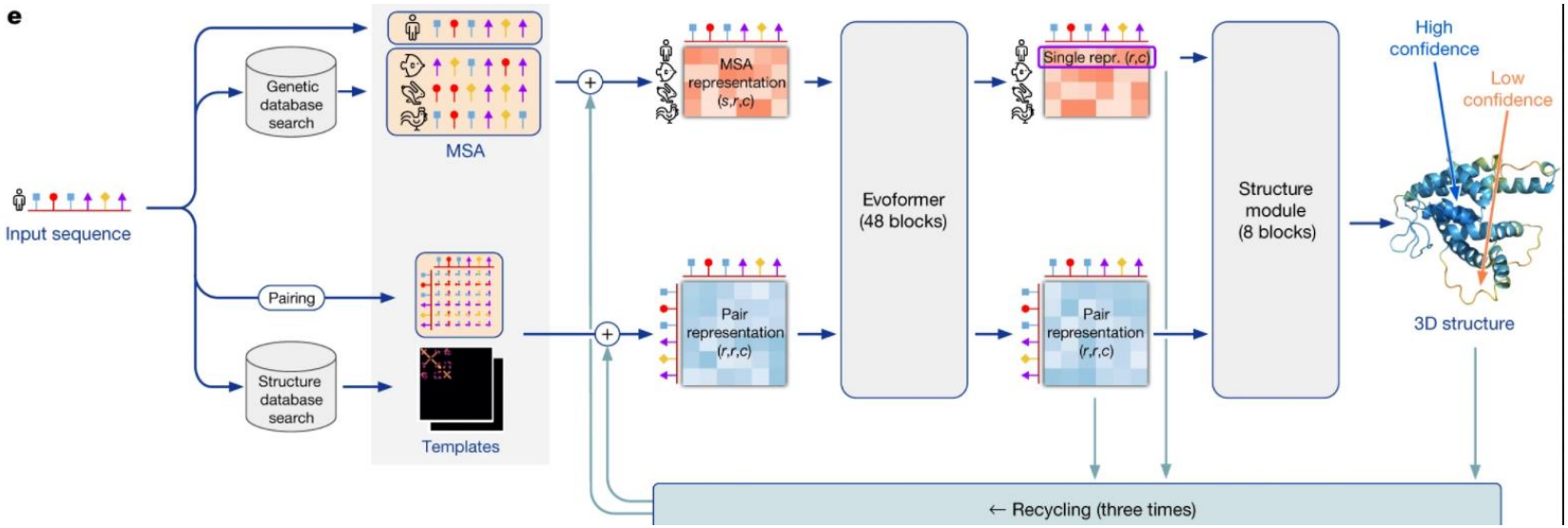


https://swissmodel.expasy.org/repository/

Alpha Fold 2

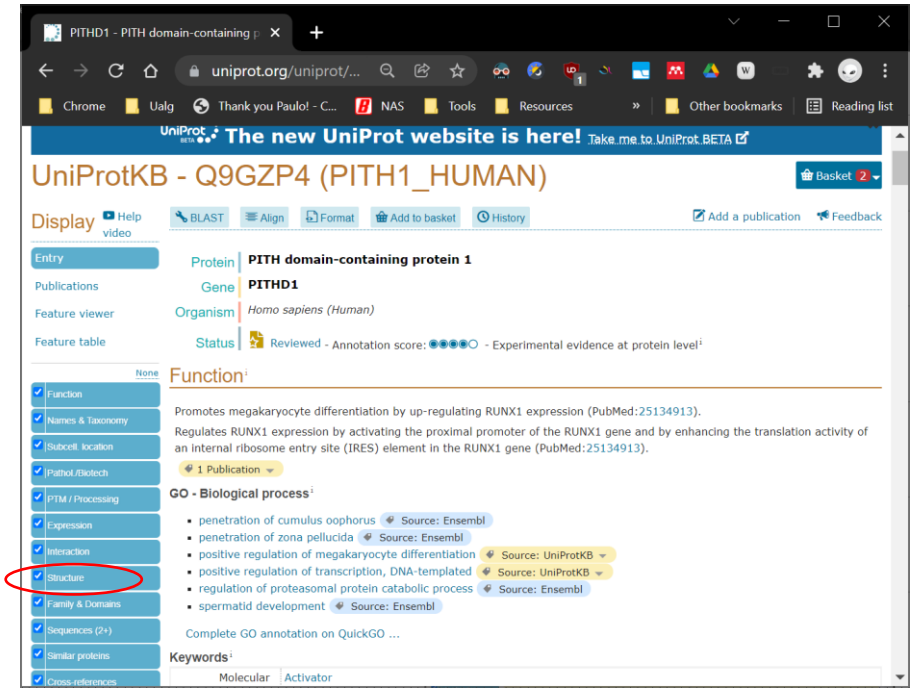
- Em 2020, a Google Deep Mind apresentou o algoritmo Alpha Fold 2 para a previsão da estrutura terciária das proteínas, ultrapassando em larga margem todas as outra ferramentas actualmente disponíveis para este tipo de previsões.

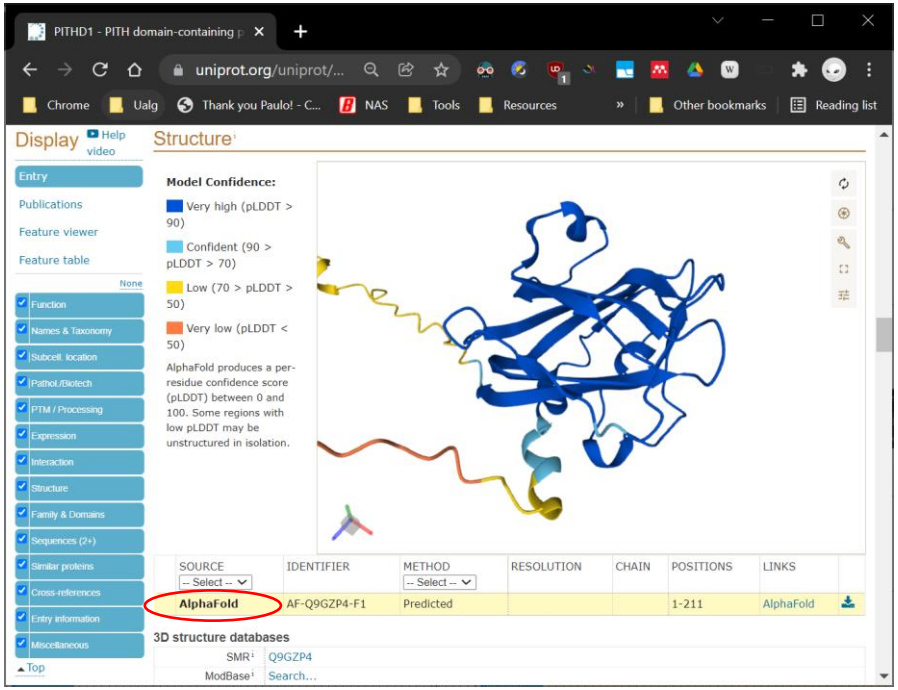
- O Alpha Fold 2 é um *deep neural network,* um algoritmo de aprendizagem máquina que é treinado usando come exemplos as estruturas e sequências de proteínas conhecidas.



Alpha Fold 2 no Uniprot

- A Google Deep Mind, em colaboração com o European Bioinformatics Insitute (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.





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 faramacêutica, utilizadas na descoberta de novos fármacos, c

- Podem conter uma variedade de *descritores moleculares* (estrutura, solubilidade, massa molecular, hidrofobicidade, carga, etc...) e também informação sobre a actividade biológica e até dados de ensaios de actividade

Bases de dados de pequenas moléculas

- PubChem (http://http://pubchem.ncbi.nlm.nih.gov/)
- DrugBank (http://www.drugbank.ca)
- ChEMBL (http://https://www.ebi.ac.uk/chembl/)
- ZINC (http://zinc.docking.org)
- Cambridge Structural Database (http://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**)

  } 9/11/2014
- Permite pesquisa por estrutura, similaridade, etc…

Bases de dados

PubChem

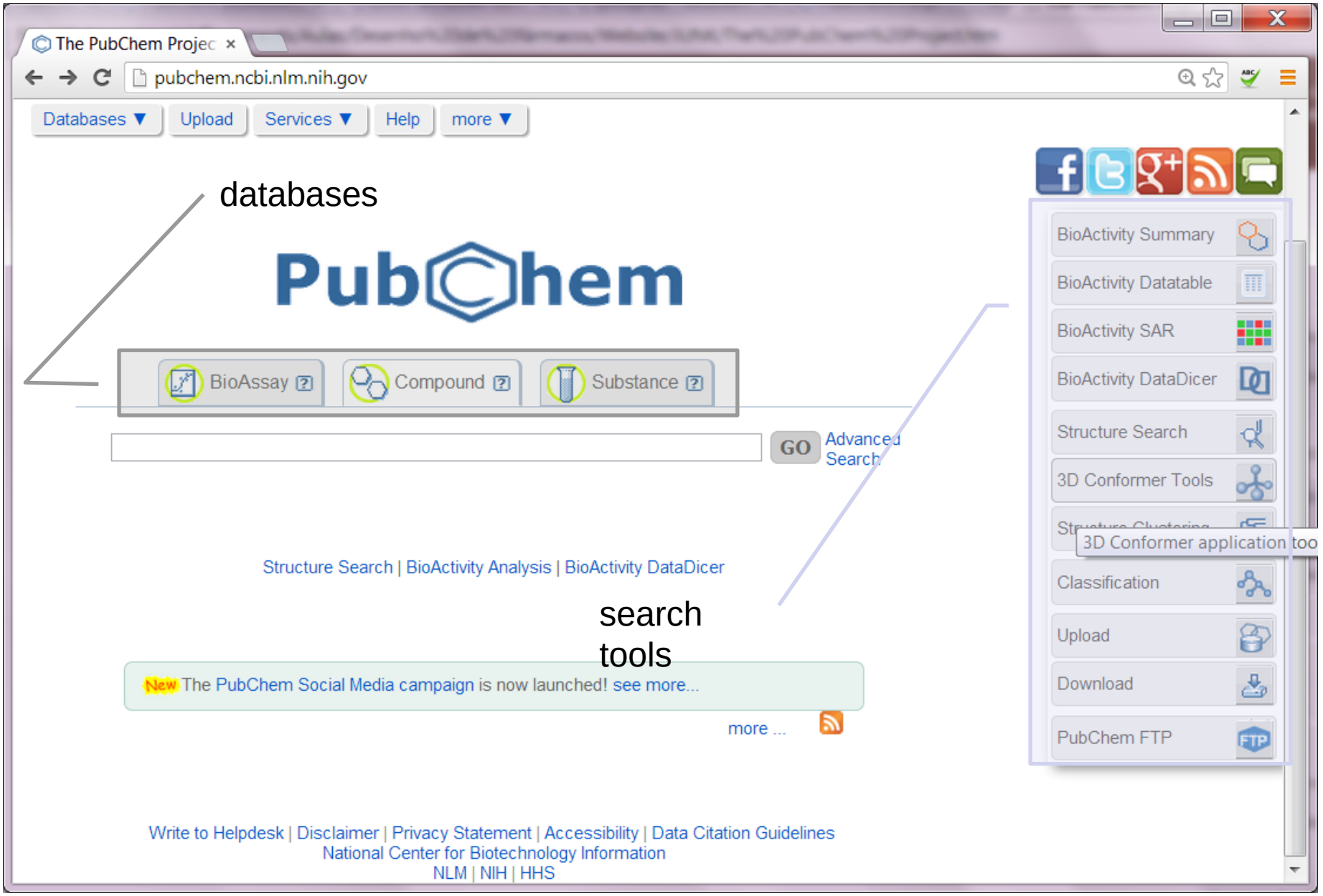
- **PubChem Substance:** cada entrada nesta base de dados contem informação sobre uma *amostra química* de proveniência bem definida, que pode conter ou ou mais compostos. Cada entrada possui referências cruzadas para bibliografia, ensaios biológicos, estruturas de compostos, proteínas, etc...

- **PubChem Compound:** base de estruturas químicas validadas e agrupadas por similaridade. Contem vários descritores e propriedades moleculares pré-calculados (eg: XlogP, MW) que podem ser usados para filtrar as pesquisas. Cada **substância** pode conter um ou mais compostos.

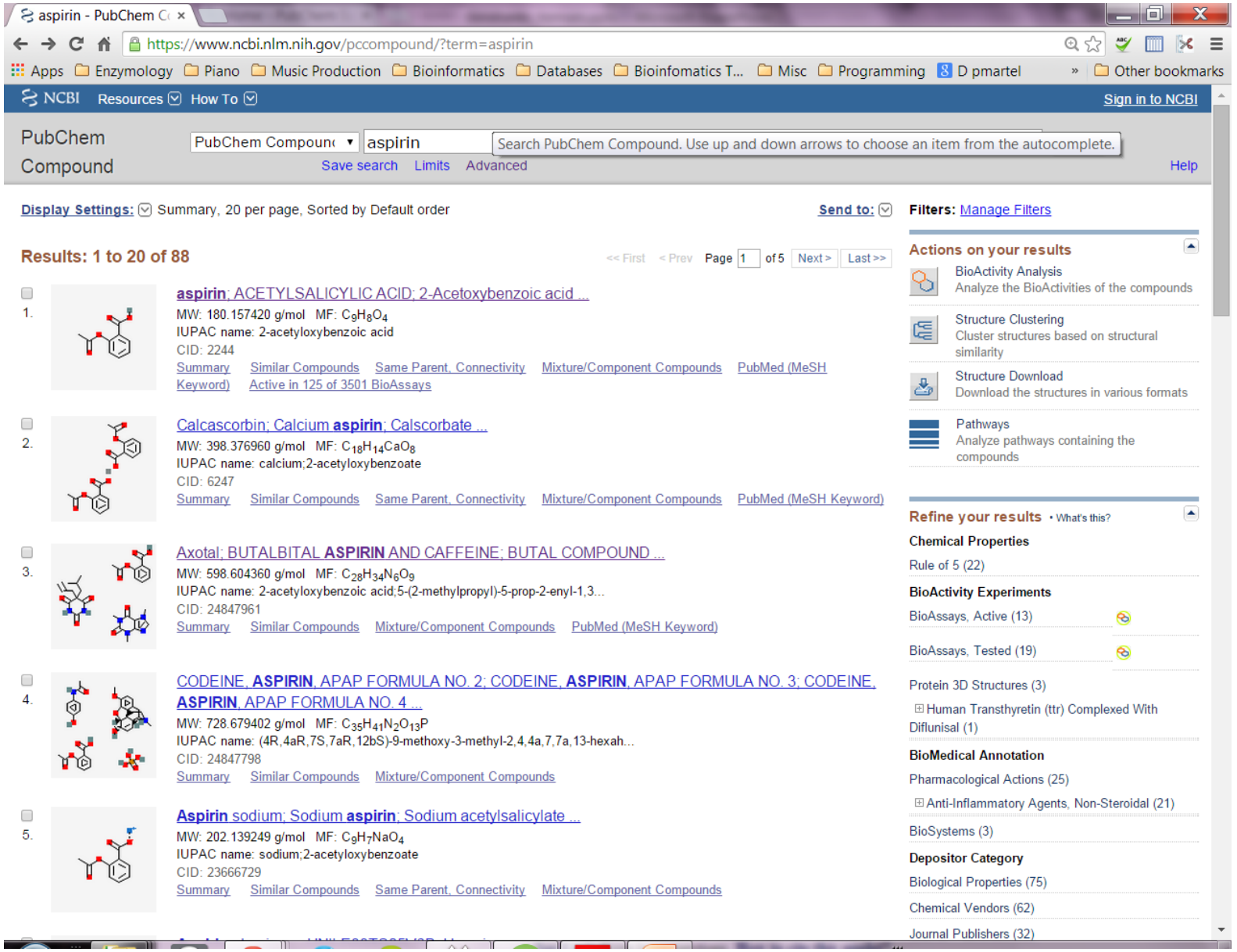
- **PubChem Bioassay:** ensaios de actividade biológicas relativos às entradas de **PubChem Substance**, contendo as descrições e resultados dos ensaios.

Pesquisa no PubChem

- **Compound:** nomes, 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 Projec
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
3D Conformer application too
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



PubChem Compound

Aspirin - PubChem

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

NCBI

PubChem Compound

PubChem Compound | Search

Limits | Advanced search | Help

SHARE

Aspirin - Compound Summary (CID 2244)

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

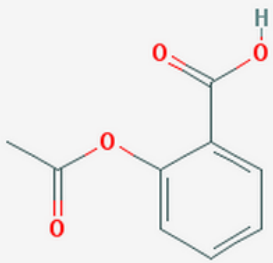
Molecular Formula: $C_9H_8O_4$ **Molecular Weight:** 180.15742 **InChIKey:** BSYNRYMUTXBXSQ-UHFFFAOYSA-N

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

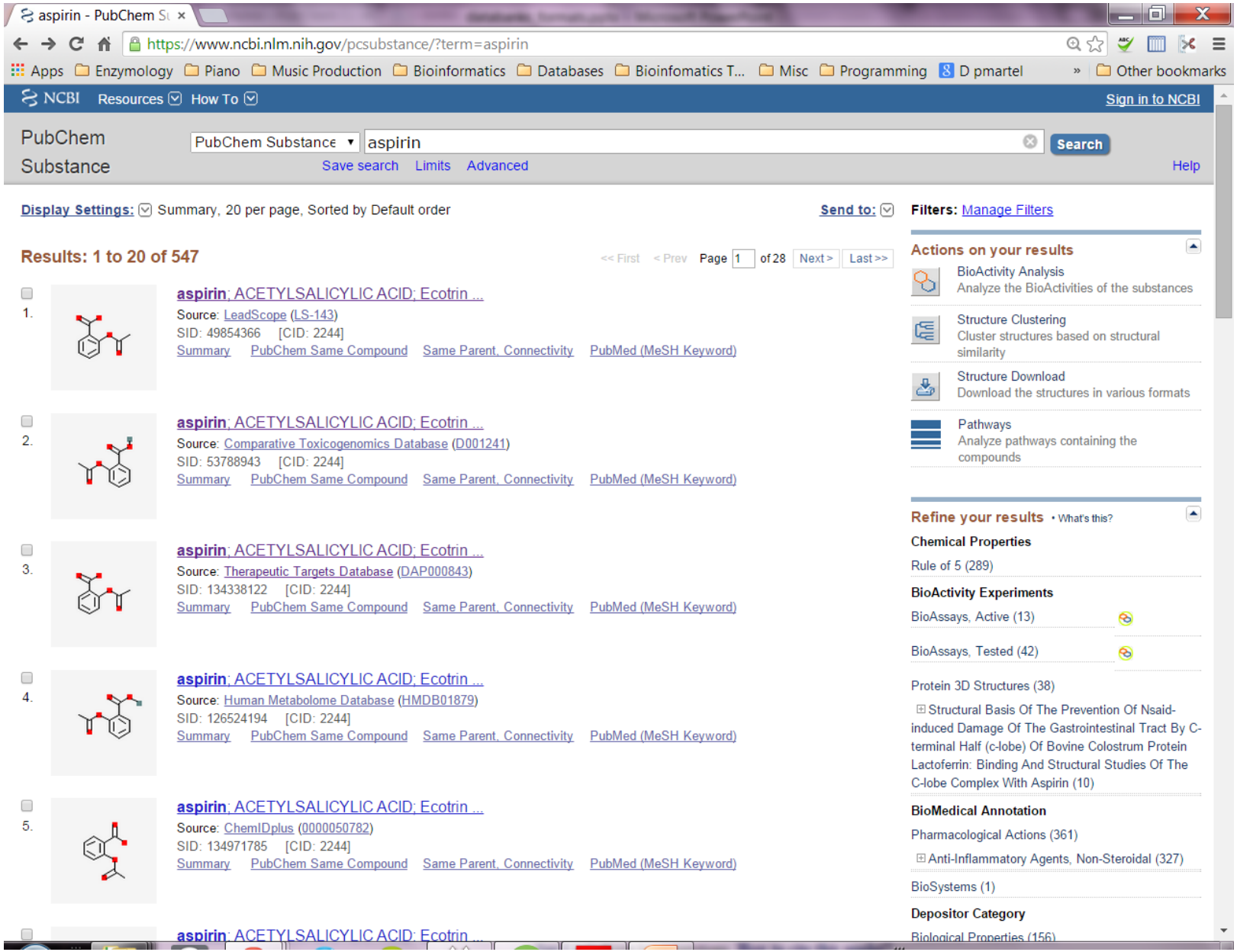
Related Compounds

Same, Connectivity (8)

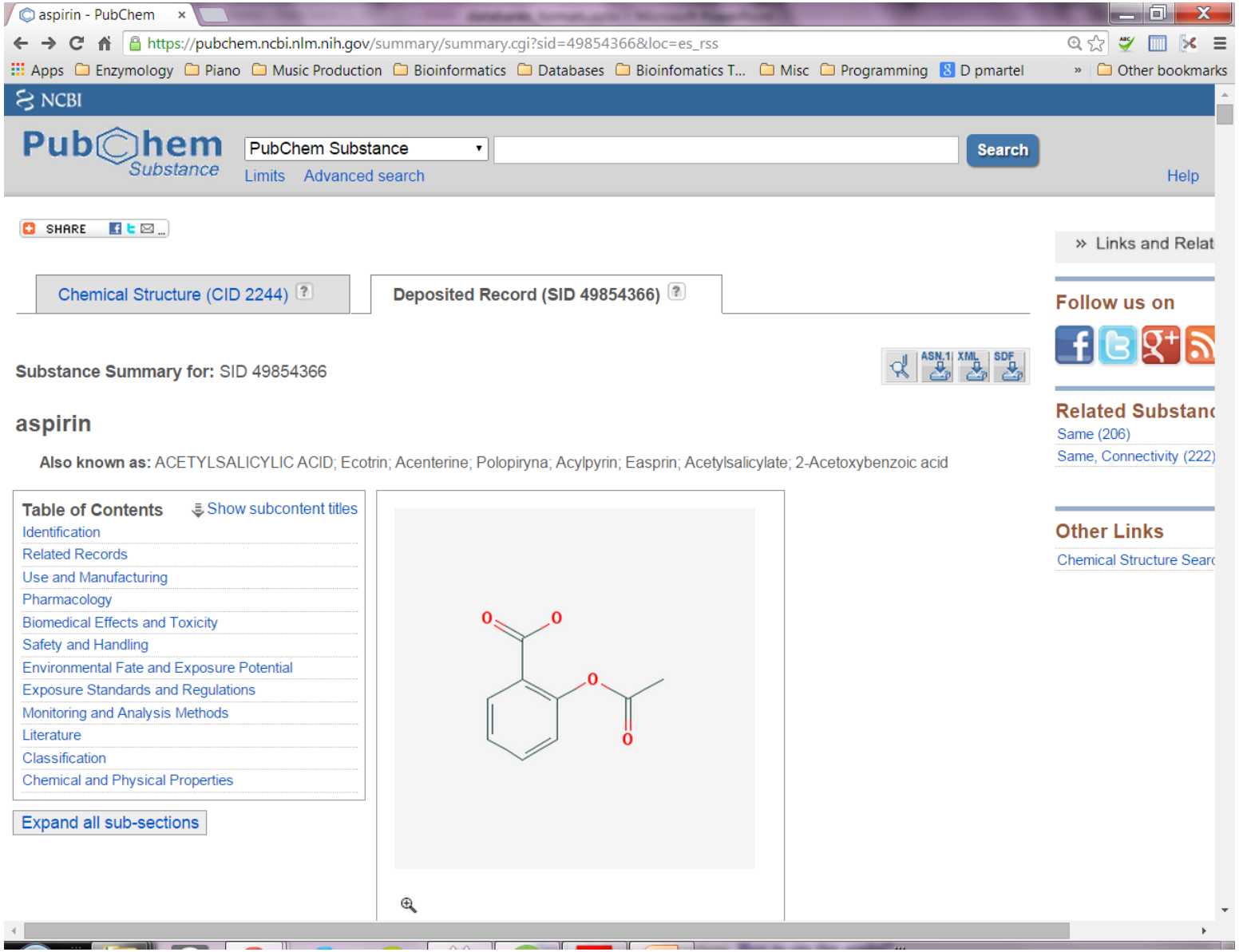
Similar Compounds (3154)

Similar Conformers (8000) View

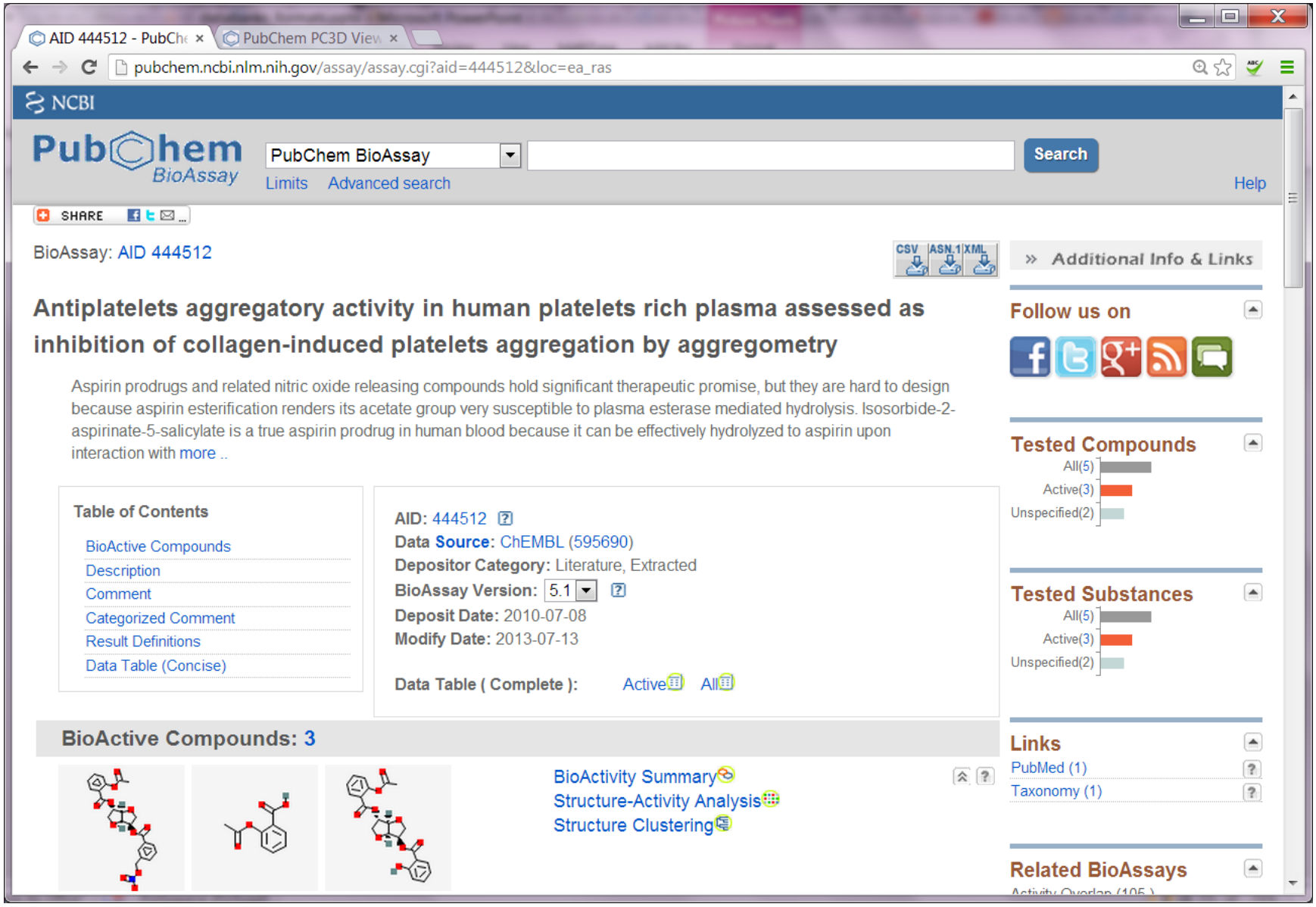
PubChem Substance



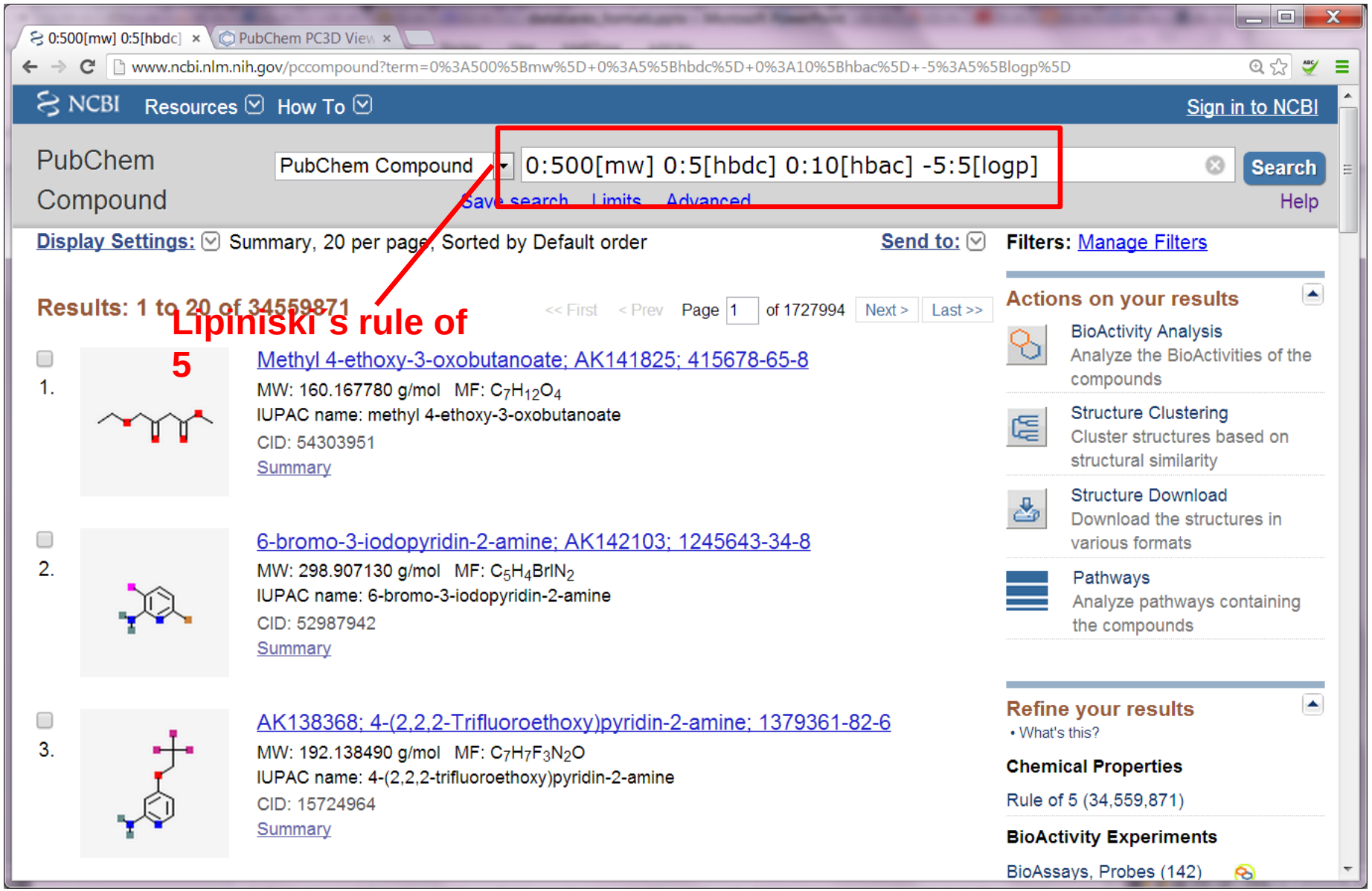
PubChem Substance



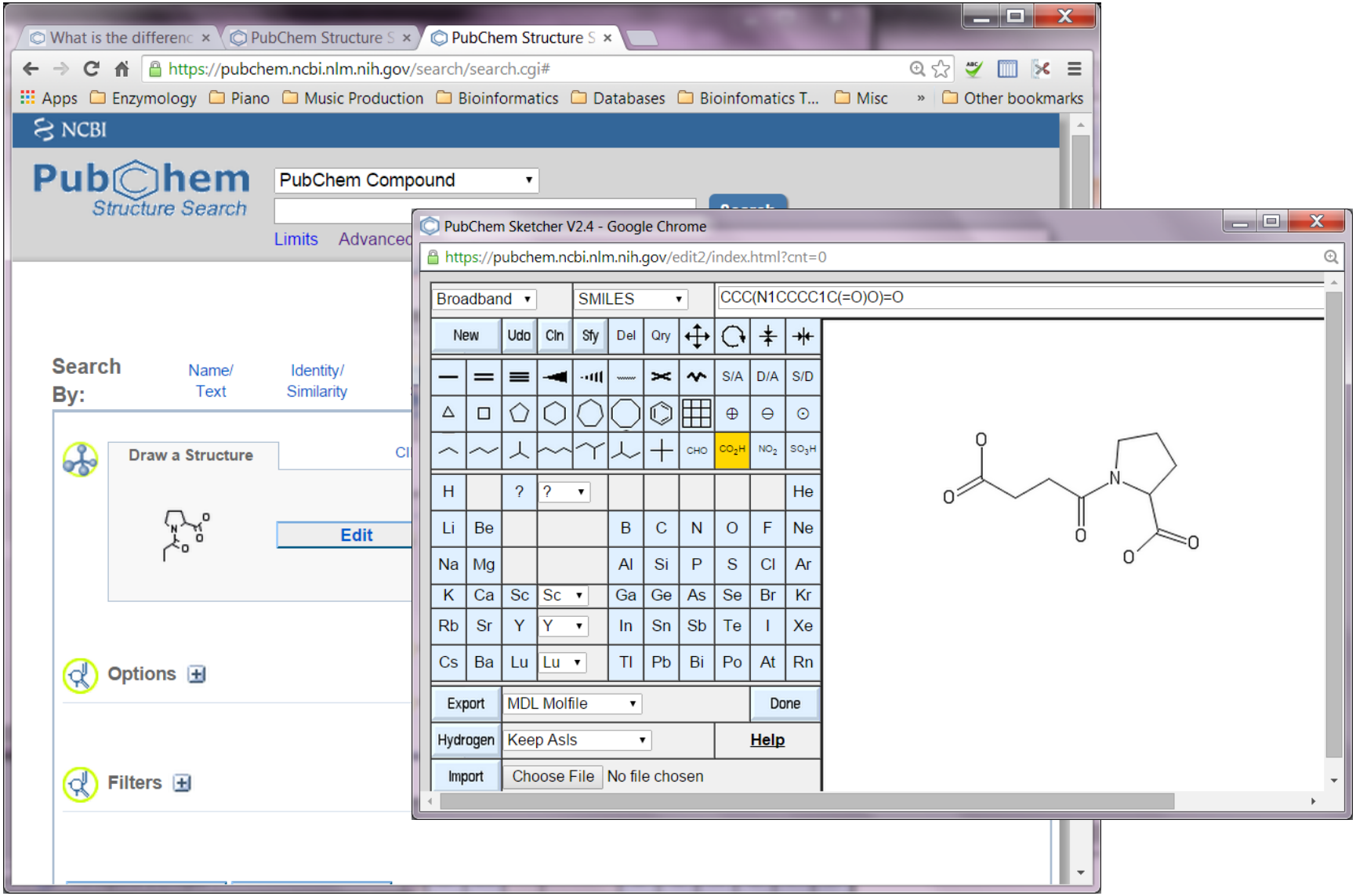
PubChem BioAssay



PubChem – Pesquisa por “Tag”



PubChem – Pesquisa por estrutura

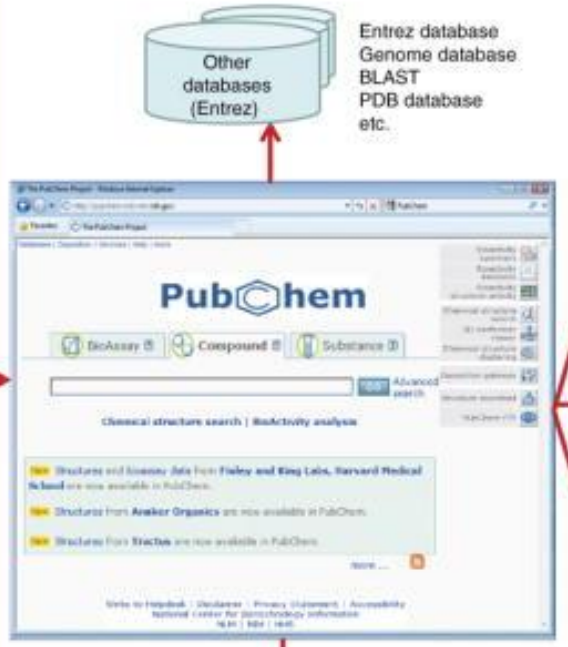


A.

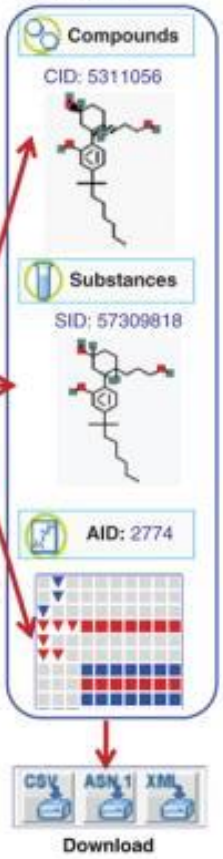
Bioassay Data Source Name	Bioassay count	Substance count
BioAssay Data Deposited by NIH MLPPCN and MLSCN		
NCGC (NIH)	485	398,461
The Scripps Research Institute Molecular Screening Center	483	357,929
Burnham Center for Chemical Genomics	397	400,255
NMMLSC (University of Mexico)	230	348,231
Broad Institute 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
BioAssay Data Deposited by Other Sources		
ChEMBL (European Bioinformatics Institute, EBI)	446,639	551,496
DTP/NCI (NIH)	173	189,809
ChemBank (Broad Institute of Harvard & MIT/Chemical Biology)	106	5,329
SGCOxCompounds (SGC Oxford)	43	319
NINDS Approved Drug Screening Program	34	1,040
BindingDB (CARB)	20	3,285
Diabetic Complications Screening (NIDDK/JDRF)	14	1,040
EPA DSSTox (National Center for Computational Toxicology)	12	4,099
GLIDA, GPCR-Ligand Database	6	19,474
GlaxoSmithKline (GSK)	6	13,533
ProbeDB (NCBI)	5	279
MTDP (CCR, NCI, NIH)	4	99,933
IUPHAR-DB	4	104
Structural Genomics Consortium	2	28
The Genomics Institute of the Novartis Research Foundation (GNF)	1	33,364
Shanghai Institute of Organic Chemistry	1	3,073
Circadian Research, Kay Laboratory (UCSD)	1	1,279
Thermo Scientific Dharmacon RNAi Technologies	1	840
ChemBlock	1	122
CC_PMLSC	1	47
SGCS to Compounds	1	17
Total: 41	449,402	4,985,224

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

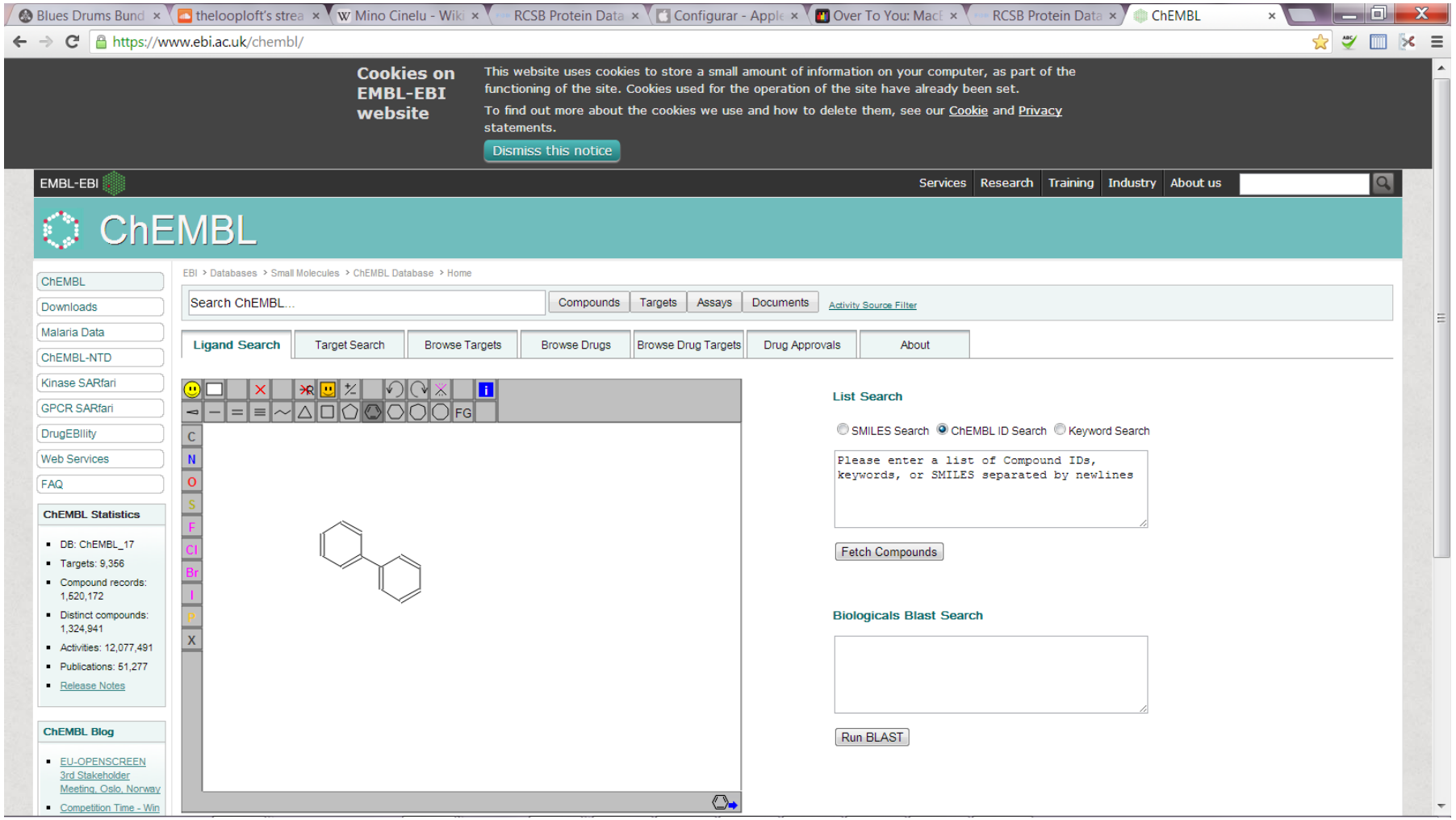
B.



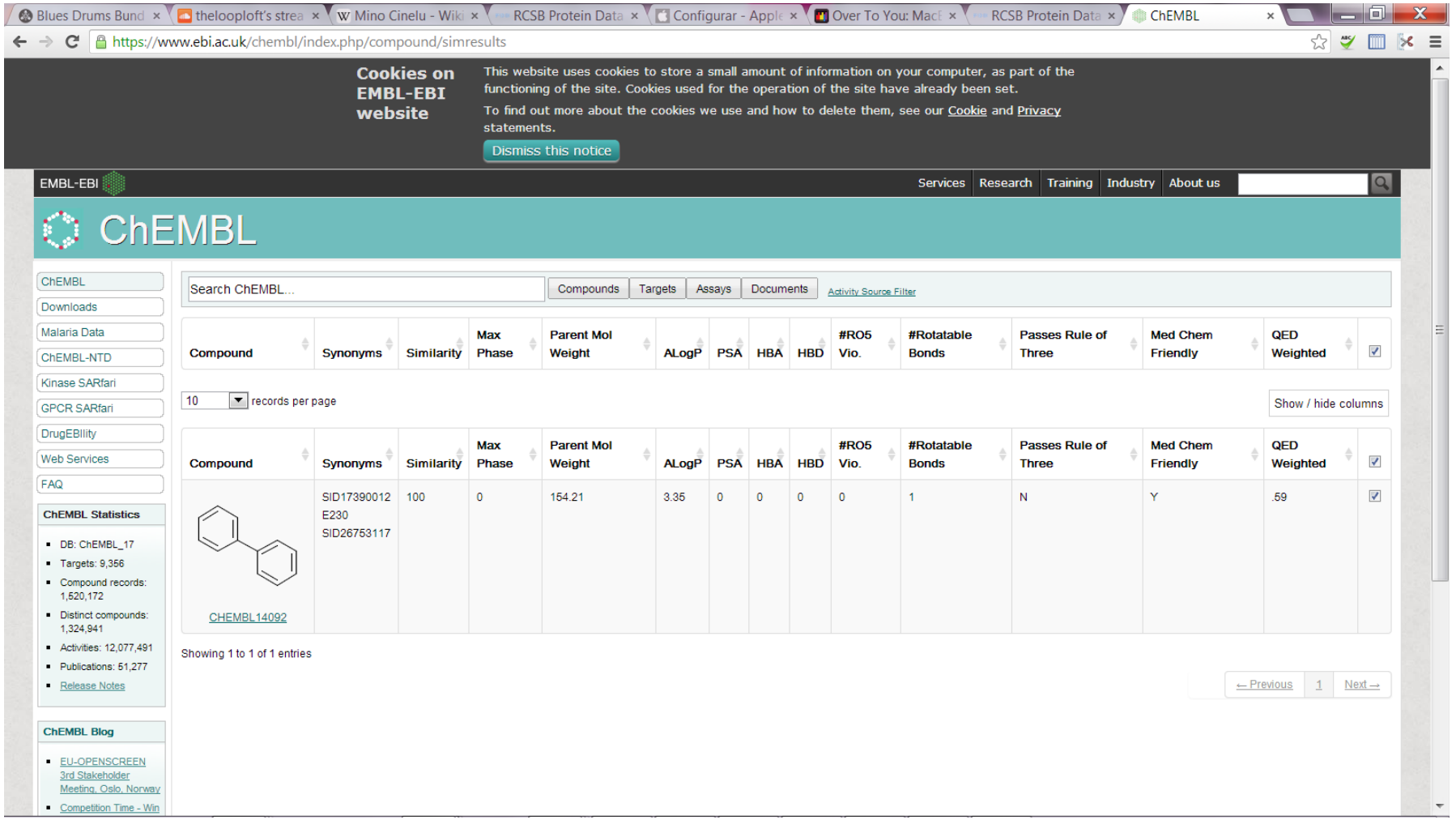
C.



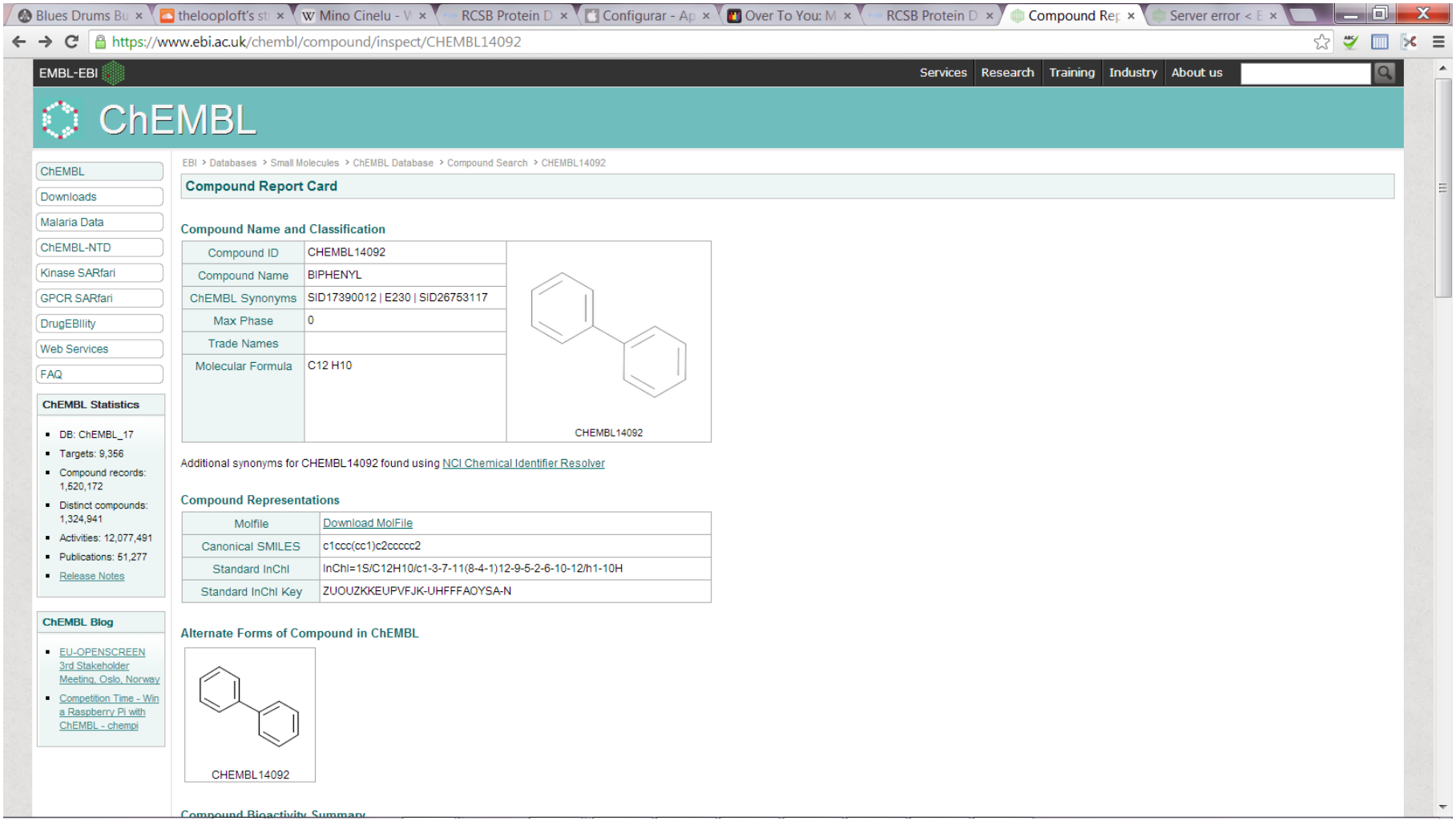
Exemplo de pesquisa estrutural na base ChEMBL



Exemplo de pesquisa estrutural na base ChEMBL



Exemplo de pesquisa estrutural na base ChEMBL



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 DrugBank Search Help / Advanced

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

DrugBank is supported by David Wishart, Departments of Computing Science & Biological Sciences, University of Alberta.

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

More about DrugBank

What's New?

Posted February 17, 2012:

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

0 Comments

News archive

Recent Comments

PEOPLE RECENT POPULAR

DrugBank: Acetylsalic ×

www.drugbank.ca/drugs/DB00945

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

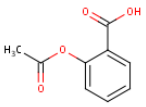
Home | Browse | Search | Downloads | News & Updates | About | Help | Contact Us

Search: Search DrugBank | Search | Help / Advanced

Identification | Taxonomy | Pharmacology | Pharmacoeconomics | Properties | References | Interactions | 0 Comments

targets (3) | enzymes (3) | transporters (3) | carriers (1) | Show Drugs with Similar Structures for All drugs

Identification

Name	**Acetylsalicylic acid**
Accession Number	**DB00945** (APRD00264, EXPT00475)
Type	small molecule
Groups	approved
Description	The prototypical analgesic used in the treatment of mild to moderate pain. It has anti-inflammatory and antipyretic properties and acts as an inhibitor of cyclooxygenase which results in the inhibition of the biosynthesis of prostaglandins. Acetylsalicylic acid also inhibits platelet aggregation and is used in the prevention of arterial and venous thrombosis. (From Martindale, The Extra Pharmacopoeia, 30th ed, p5)
Structure	 Download: MOL | SDF | SMILES | InChI Display: 2D Structure | 3D Structure
Synonyms	• 2-Acetoxybenzenecarboxylic acid • 2-Acetoxybenzoic acid • 2-Carboxyphenyl acetate • A.S.A. • Acetilsalicilico • Acetilum acidulatum • Acetosalic acid • Acetoxybenzoic acid • Acetylsalicylate • Acetylsalicylsaure (GERMAN) • Acetysalicylic acid • Acide acetylsalicylique (FRENCH)