

*Ferramentas e bases de dados  
online para desenho de fármacos*

- Portais: locais de acesso a recurso de vários tipos
- Ferramentas on-line: conversão de formatos, cálculo de propriedades moleculares, visualização, docking,...
- Bases de dados: contêm estruturas moleculares de macromoléculas e moléculas pequenas que podem ser pesquisadas de múltiplas formas
- Podem ser serviços de acesso livre, ou sites comerciais com custos de utilização

# Vantagens da utilização dos serviços online

- Disponíveis em qualquer local
- Custos de manutenção reduzidos
- Custos de licenciamento reduzidos
- Integração de diferentes tipos de software
- Fácil monitorização da utilização
- Computação em *cloud*
- Compatibilidade com múltiplos plataformas informáticas (Win, Mac, Linux, Android, etc)

# O portal Click2Drug

- <http://www.click2drug.org>
- O portal faz parte do Swiss Institute of Bioinformatics
- Contem ~800 links divididos em categorias, incluindo diferentes tipos de software e bases de dados
- Cada link contem uma descrição resumida do serviço oferecido
- Está em permanente actualização

# O portal Click2Drug

SIB Swiss Institute of Bioinformatics

## Click2Drug

Directory Bibliography Encyclopedia Citations Contacts Disclaimer

**Directory of Tools**

- Databases
- Chemical structure rep.
- Molecular modeling
- Homology modeling
- Binding site prediction
- Docking
- Screening
- Target prediction
- Ligand design
- Binding free energy estimation
- QSAR
- ADME Toxicity

**Mobile applications**

**Last additions**

**Tag cloud**

**FAQ**

8+1 / 47 Updated on 7/18/2014. Currently 777 links. Show all links / Hide all links.

Click on the following picture to select tools related to a given activity:

The diagram illustrates the 'In silico drug design pipeline' by Click2Drug. It starts with several input boxes on the left: 'Small molecule databases', '3D structure databases', and 'Molecular fragment databases'. Arrows point from these to 'Prepare/select molecules for screening', which then leads to 'Structure-based (docking)'. This leads to 'Lead design', which has two paths: one to 'Structure-based (design in situ)' and another to 'Structure-based (pharmacophore)'. From 'Lead design', arrows point to 'Virtual High Throughput Screening' (which also receives input from 'Prepare/select molecules for screening') and 'Affinity estimation'. 'Affinity estimation' leads to 'Experimental assay'. 'Experimental assay' also receives input from 'Molecule selection' (which receives input from 'ADME / Toxicity estimation') and 'Lead optimization' (which receives input from 'Ligand-based pharmacophore'). 'Lead optimization' also receives input from 'Structure-based (docking)'. There are dashed arrows from 'Databases to develop ADME / Toxicity models' to 'Lead optimization' and 'Molecule selection'. A red arrow points from 'Experimental assay' back to 'Lead optimization'. On the right, there is a pink oval labeled 'Molecules visualization and handling' with three upward-pointing arrows.

In silico drug design pipeline, by Click2Drug

Show all links Hide all links

# O portal Click2Drug

The screenshot shows a web browser window titled "Directory of in silico" with the URL "www.click2drug.org". The page content is a list of various in silico drug design tools and databases, organized into sections:

- Databases**: ZincDatabase, ChEMBL, Chemspider, Bingo, JChemforExcel, ChemDiff, ProteinDataBank(PDB), BindingMOAD(MotherOfAllDatabase), LigandProtein DataBase(LPDB), TTD, STITCH, SMPDB, ...
- Chemical structure representations**: ChemDraw, MarvinSketch, ACD/ChemSketch, jsMolEditor, Marvinmoleculeeditorandviewer, Ketcher, UCSFChimera, Pymol, OpenStructure, InChi, TriposMol2, PDBformat, OpenBabel, Corina, Indigo, PoseView, DSVisualizer, BINANA, E-Babel, Corinaonline demo, ChemicalIdentifierResolver, ChemMobi, ChemSpotlight, ...
- Molecular Modeling**: CHARMM, GROMACS, Amber, SwissParam, CHARMM-GUI, CHARMMing.org, SwissSideChain, ...
- Homology Modeling**: Modeller, I-TASSER, LOMETS, SWISS-MODEL, SWISS-MODELRepository, Robetta, ...
- Binding site prediction**: MED-SuMo, FINDSITE, fpocket, sc-PDB, CASTp, PocketAnnotatedatabase, 3DLigandSite, metaPocket, PocketAnnotate, ...
- Docking**: Autodock, DOCK, GOLD, SwissDock, DockingServer, 1-ClickDocking, ...
- Screening**: Pharmer, Catalyst, PharmaGist, Blaster, AnchorQuery, istar, ...
- Target prediction**: MolScore-Antivirals, MolScore-Antibiotics, PredictFX, SwissTargetPrediction, SEA, ChemProt, ...
- Ligand design**: GANDI, LUDI, SPROUT, SwissBioisostere, VAMMPIRE, sc-PDB-Frag, e-LEA3D, eDesign, iScreen, ...
- Binding free energy estimation**: ...

# O portal Click2Drug

The screenshot shows a web browser window titled "Directory of in silico" with the URL "www.click2drug.org". The page displays a list of various chemical databases. At the top, there are two buttons: "Show all links" and "Hide all links". Below this, a section titled "Databases" lists several databases: ZincDatabase, ChEMBL, ChemSpider, Bingo, JChemforExcel, ChemDiff, ProteinDataBank(PDB), BindingMOAD(MotherOfAllDatabase), LigandProteinDataBase(LPDB), TTD, STITCH, SMPDB, and others. A section titled "Chemical databases" contains a detailed list of over 20 databases, each with a brief description. The list includes: Zinc Database, ChEMBL, ChemSpider, CoCoCo, DrugBank, PubChem, PubChem Mobile, TCM, Mcule database, WOMBAT, Approved Drugs, ChemSpider Mobile, e-Drug3D, ChemDB/ChemicalSearch, Structural Database (CSD), and SPRESIweb. The bottom of the page shows the footer: "www.click2drug.org/directory\_MolecularModeling.html" and "database of approved drugs and ZINC provided by the CRS4 - Bioinformatics Laboratory, Parco Sardegna Ricerche, Italy."

In silico drug design pipeline, by Click2Drug

Show all links Hide all links

▼ Databases

ZincDatabase, ChEMBL, ChemSpider, Bingo, JChemforExcel, ChemDiff, ProteinDataBank(PDB), BindingMOAD(MotherOfAllDatabase), LigandProteinDataBase(LPDB), TTD, STITCH, SMPDB, ...

Chemical databases

- [Zinc Database](#). Curated collection of commercially available chemical compounds, with 3D coordinates, provided by the Shoichet Laboratory in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF).
- [ChEMBL](#). Curated database of small molecules. Includes interactions and functional effects of small molecules binding to their macromolecular targets, and series of drug discovery databases.
- [ChemSpider](#). Collection of chemical compounds maintained by the Royal Society of Chemistry. Includes the conversion of chemical names to chemical structures, the generation of SMILES and InChI strings as well as the prediction of many physicochemical parameters.
- [CoCoCo](#). Free suite of multiconformational molecular databases for High-Throughput Virtual Screening. It has single and multi conformer databases prepared for HTVS in different formats like Phase, Catalyst, Unity and SDF. Provided by the Department of Pharmaceutical Sciences of the University of Modena and Reggio Emilia.
- [DrugBank](#). Bioinformatics and cheminformatics resource combining detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. Allows searching for similar compounds.
- [PubChem](#). Database of chemical compounds maintained by the National Center for Biotechnology Information (NCBI), along with bioassays results. Allows similar compounds search (2D and 3D).
- [PubChem Mobile](#). Free application to search PubChem databases using chemical names, synonyms, and keywords. For Android.
- [TCM](#). Free small molecular database on traditional Chinese medicine, for virtual screening. It is currently the world's largest TCM database, and contains 170'000 compounds, with 3D mol2 and 2D cdx files, which passed ADMET filters.
- [Mcule database](#). Commercial database of commercially available small molecules. Allows filtering by chemical supplier data (stock availability, price, delivery time, chemical suppliers, catalogs, minimum purity, etc.) and export the whole Mcule database including supplier and procurement related properties. Reduced prices for academic. Provided by Mcule.
- [WOMBAT](#). (World of Molecular Biotactivity). Database of 331,872 entries (268,246 unique SMILES), representing 1,966 unique targets, with biactivity annotations. Compiled by Sunset Molecular Discovery LLC.
- [Approved Drugs](#). The Approved Drugs app contains over a thousand chemical structures and names of small molecule drugs approved by the US Food & Drug Administration (FDA). Structures and names can be browsed in a list, searched by name, filtered by structural features, and ranked by similarity to a user-drawn structure. The detail view allows viewing of a 3D conformation as well as tautomers. Structures can be exported in a variety of ways, e.g. email, twitter, clipboard. For iPad and iPhone. Developed by Molecular Materials Informatics, Inc.
- [ChemSpider Mobile](#). Allows searching the ChemSpider chemical database, provided by the Royal Society of Chemistry. Compounds can be searched by structure or by name, and browsed within the app. Results can be examined by jumping to the web page. Search structures are drawn using the powerful MMDS molecular diagram editor. For iPad. Provided by Molecular Materials Informatics, Inc.
- [e-Drug3D](#). Database mirroring the current content of the U.S. pharmacopoeia of small drugs. Contains 1533 molecular structures with a molecular weight < 2000 (last update: February 2012). Provides SD files (single conformer, tautomers or multiple conformers). Maintained by the Institut de Pharmacologie Moléculaire et Cellulaire, France.
- [ChemDB/ChemicalSearch](#). Find chemicals by various search criteria.
- [Structural Database \(CSD\)](#). Repository for small molecule crystal structures in CIF format. The CSD is compiled and maintained by the Cambridge Crystallographic Data Centre
- [SPRESIweb](#). Integrated database containing over 8.7 million molecules, 4.1 million reactions, 658,000 references and 164,000 patents covering the years 1974 - 2009. Developed by InfoChem.
- [MMSINC](#). Database of non-redundant, annotated and biomedically relevant chemical structures. Includes the analysis of chemical properties, such as ionization and tautomerization processes, and the in silico prediction of 24 important molecular properties in the biochemical profile of each structure. MMSINC supports various types of queries, including substructure queries and the novel 'molecular scissorsin' query. MMSINC is interfaced with other primary data collectors, such as PubChem, Protein Data Bank (PDB), the Food and Drug

www.click2drug.org/directory\_MolecularModeling.html

database of approved drugs and ZINC provided by the CRS4 - Bioinformatics Laboratory, Parco Sardegna Ricerche, Italy.

# Virtual Computational Chemistry Laboratory - VCCLAB

The screenshot shows a web browser window displaying the VCCLAB homepage. The URL in the address bar is <http://www.vcclab.org/lab/>. The page features a header with the text "Virtual Computational Chemistry Laboratory" and a navigation menu with links to Home, About, Partners, Software, Articles, Servers, Download, Web Services, How to cite?, and Contact. A sidebar on the left contains links to Home, About, Partners, Software, Articles, Servers, Download, Web Services, How to cite?, and Contact. The main content area includes a section titled "on-line software" with a list of programs: ALOGPS 2.1\*, ASNN\*, E-BABEL, PNN, PCCLIENT, E-DRAGON, PLS, UFS, and SPC. There is also a note about the accuracy of ALOGPS 2.1\*. At the bottom of the page, there is a copyright notice: "Copyright 2001 -- 2011 http://www.vcclab.org. All rights reserved."

Virtual Computational Chemistry Laboratory

on-line software

- ALOGPS 2.1\* is the most accurate program to predict lipophilicity and aqueous solubility of molecules
- ASNN\* calculates highly predictive non-linear neural network models
- E-BABEL is molecular structure information interchange hub
- PNN produces clearly interpretable analytical non-linear models
- PCCLIENT generates more than 3000 descriptors
- E-DRAGON calculates DRAGON molecular indices
- PLS implements original two-step descriptors selection procedure
- UFS produces a reduced data set that contains no redundancy and a minimal amount of multicollinearity

If you have any questions, problems to run applets, please, contact

PLAY TOP

ON-LINE SOFTWARE

ALOGPS 2.1

ASNN

E-BABEL

PNN

PCCLIENT

E-DRAGON 1.0

PLS

UFS

SPC

Home About Partners Software Articles Servers Download Web Services How to cite? Contact

Copyright 2001 -- 2011 <http://www.vcclab.org>. All rights reserved.

# iDrug: on-line Drug Design Workbench

Screenshot of the iDrug - An Online Interactive Drug Discovery and Design Platform interface.

The interface includes:

- A sidebar navigation tree:
  - Demo (No editable)
  - Pharmacophore
    - Target Navigator (4OH-tamoxifen)
    - Hit Explorer (CDK2: 1AQ1)
    - 18820
    - 18822
  - Similarity
    - HYZ\_2RGP.mol2 (EGFR)
    - 17170
- A central 3D molecular visualization area showing a complex organic molecule.
- A JSmol button below the visualization.
- A feature editor table:

Feature	X	Y	Z	Radius	Enabled
HB DONOR	0	0	0	1	<input type="checkbox"/>
- A dropdown for database selection:
  - Compound Database (MayBridge 60,791)
  - Upload a Database
- Buttons at the bottom: Load, Add Feature, Clear, Save Results.
- A results table on the right side:

Name	Hybrid Score	Feature Score
CHEMBL524457	1.693	0.8
CHEMBL402294	1.719	0.7
CHEMBL525725	1.7	0.7
CHEMBL30432	1.627	0.7
CHEMBL255438	1.557	0.7
CHEMBL498133	1.626	0.7
CHEMBL255866	1.503	0.7
CHEMBL456758	1.527	0.6
CHEMBL67003	1.452	0.6
CHEMBL510845	1.34	0.6
CHEMBL246073	1.51	0.6
CHEMBL257411	1.492	0.6
CHEMBL300000	1.487	0.6

# Virtual Computational Chemistry Laboratory - VCCLAB

The screenshot shows the VCCLAB Servers page at <http://www.vclab.org/servers/>. The page features a header with a molecular model icon, a navigation menu, and a main content area. The main content includes a map of Europe with server locations (Portsmouth, Erlangen, Munich, Kyiv, Milano, Moscow) and a central section titled "VCCLAB Servers" showing a network of clients connected to various servers. A sidebar on the right lists "ON-LINE SOFTWARE" such as ALOGPS 2.1, ASNN, E-BABEL, PNN, PCLIENT, E-DRAGON 1.0, PLS, UFS, and SPC.

**ON-LINE SOFTWARE**

- ALOGPS 2.1
- ASNN
- E-BABEL
- PNN
- PCLIENT
- E-DRAGON 1.0
- PLS
- UFS
- SPC

*Bases de dados e formatos de  
representação de moléculas*

# Bases de dados

- Macromoléculas (Target):
  - Estrutura (Protein Data Bank, PLD, TTD, ModBase)
  - Sequência (Uniprot, Genebank, ...)
- Moléculas pequenas:
  - (PubChem, Drugbank, Cambridge Database, ZINC, ChEMBL, TCM, WOMBAT, ....)

Contém muita informação além da *estrutura/sequência* propriamente dita.

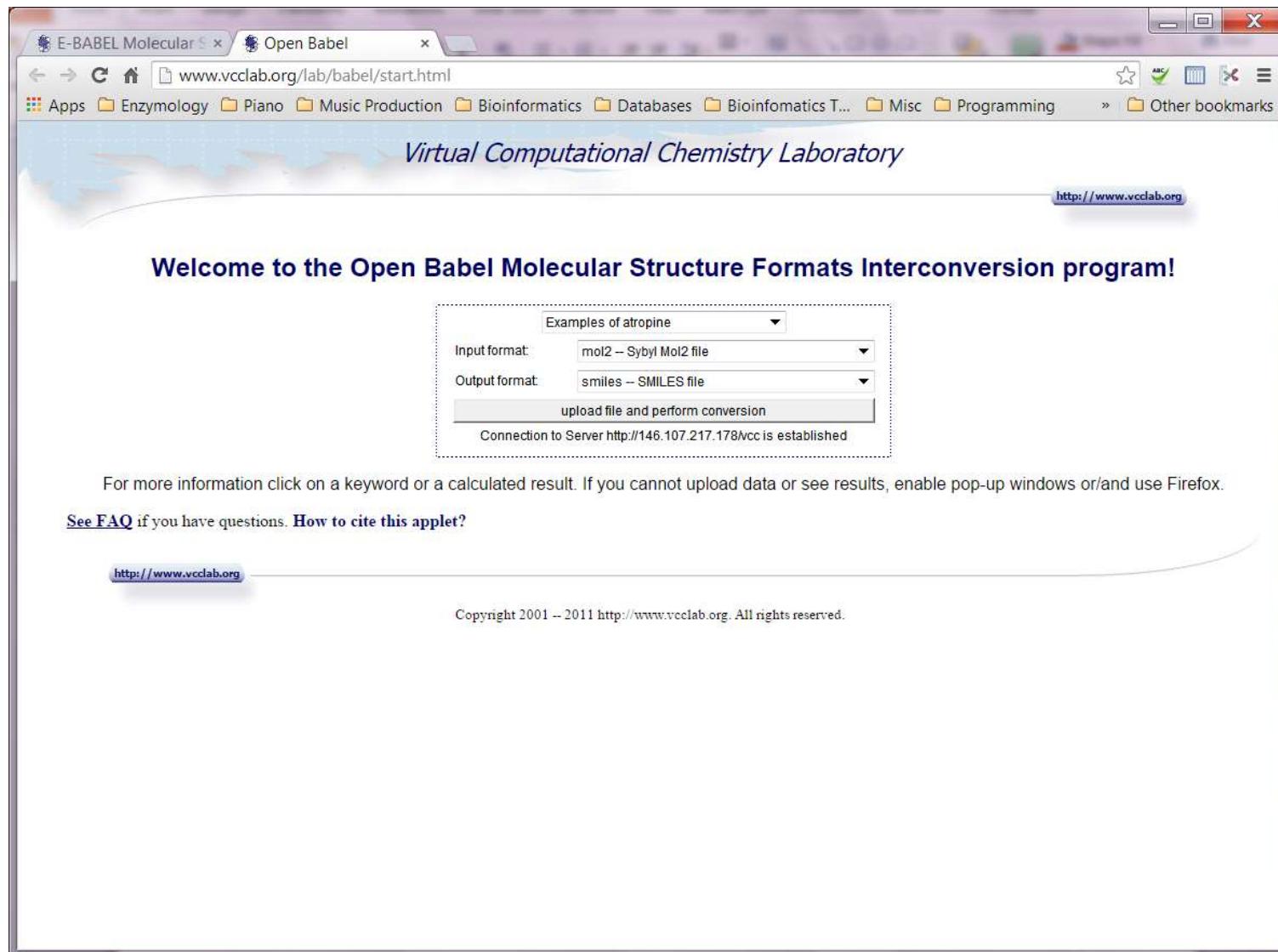
# Formatos de representação

- Estrutura:
  - PDB, MDL, SDF, MOL2, CIF, ASN.1, HIN, Trypos, Sybil, Gaussian, XYZ, CML, XML, SMILES ....
- Sequência:
  - Fasta, SWISSPROT, ASN.1, GCG, GenBank, PIR, Phylip,....

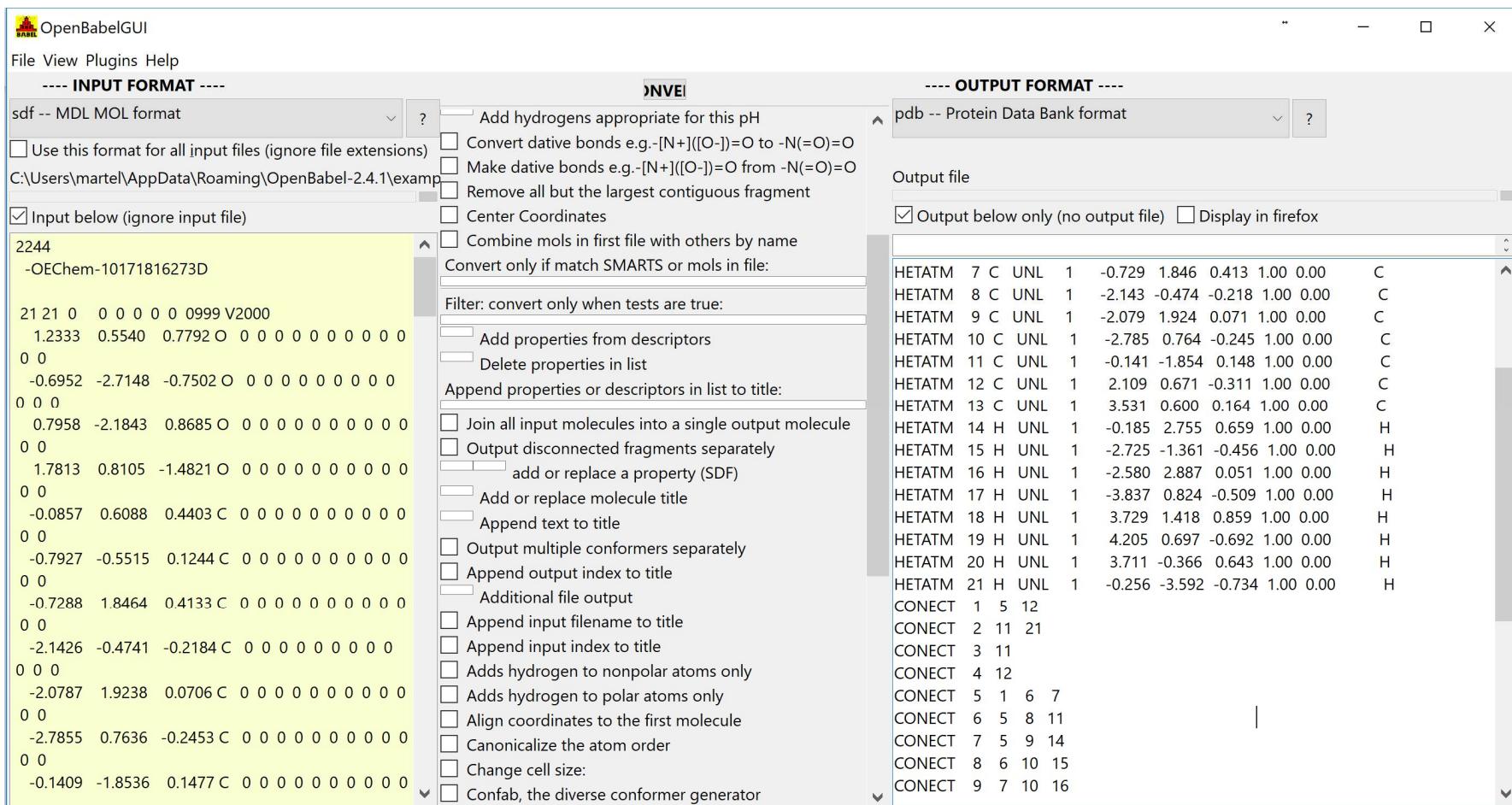
Ferramenta de conversão entre formatos:

**OpenBabel** (<http://openbabel.org>)

# E-Babel: conversão de formatos online



# OpenBabel



# Formato FASTA

- É um formato de representação de sequências biológicas (DNA ou proteína)
- Consiste numa linha de cabeçalho, seguida de linhas contendo a sequência em códigos de 1 letra
- Contem muito pouca informação para além da sequência

# Formato FASTA

>gi|19151|emb|Z14088.1| L.esculentum mRNA for 108 protein

AACAATCATGGCATCTGTGAAGTCGTCGTCGTACATCATCATCATCATTATTCCTTGT  
GTTGTTGATTTGCTTGTGATTGTACTGCAAAGCCAAGTTATCGAGTGTCAACCTAACAGT  
CATGCACCGCGTCACTTACTGGCCTGAACGTCTGCGCCCCATTCCCTGGTCCCAGGCTCACCTAC  
TGCAAGTACGGAGTGTGCAA TGCAGTACAGTCGATTAATCATGACTGTATGTGCAACACT  
ATGCGCATTGCAGCTCAAATTCCAGCTCAG TGCAACCTCCCTCCACTCTCTTGTGCAAAT  
TGAGTTGAGATCAGTGGCCAGCAAGTTACATCTGC TACATGAGCAAATTAAATAATATC  
GTAACAATAAATTAAAGTTGTCTTTTTTTGGTTATGCAAC AGACCAAGGGGTCA  
TGAGAAAAGAGTTGTACTATCATATGATTATCAATAAAAAAAATTATGAG

Cabeçalho

>Q43495|108\_SOLLC Protein 108 precursor - Solanum lycopersicum

MASVKSSSSSSSSSFISLLLILLVIVLQSQVIECQPQQSCTASLTGLNVCAPFLVPGSP  
TASTECCNAVQSINHDCMCNTMRIAAQIPAQCNLPLSCSAN

Sequência

# Formato SWISSPROT

- Representação de sequências de proteína
- Sintaxe complexa com uma variedade de *campos*
- Contem muita informação além da sequência

# Formato SWISSPROT

ID TRY1\_HUMAN      Reviewed;      247 AA.  
AC P07477; A1A509; A6NJ71; B2R5I5; Q5NV57; Q7M4N3; Q7M4N4; Q92955;  
AC Q9HAN4; Q9HAN5; Q9HAN6; Q9HAN7;  
DT 01-APR-1988, integrated into UniProtKB/Swiss-Prot.  
DT 01-APR-1988, sequence version 1.  
DT 18-SEP-2013, entry version 154.  
DE RecName: Full=Trypsin-1;  
DE      EC=3.4.21.4;  
DE AltName: Full=Beta-trypsin;  
DE AltName: Full=Cationic trypsinogen;  
DE AltName: Full=Serine protease 1;  
DE AltName: Full=Trypsin I;  
DE Contains:  
DE      RecName: Full=Alpha-trypsin chain 1;  
DE Contains:  
DE      RecName: Full=Alpha-trypsin chain 2;  
DE Flags: Precursor;  
GN Name=PRSS1; Synonyms=TRP1, TRY1, TRYP1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;  
OC Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RX PubMed=3011602; DOI=10.1016/0378-1119(86)90111-3;  
RA Emi M., Nakamura Y., Ogawa M., Yamamoto T., Nishide T., Mori T.,  
RA Matsubara K.;  
RT "Cloning, characterization and nucleotide sequences of two cDNAs  
RT encoding human pancreatic trypsinogens.";  
RL Gene 41:305-310(1986).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].

(continua)

# Formato SWISSPROT

RX PubMed=8650574; DOI=10.1126/science.272.5269.1755;  
RA Rowen L., Koop B.F., Hood L.;  
RT "The complete 685-kilobase DNA sequence of the human beta T cell  
RT receptor locus.";  
RL Science 272:1755-1762(1996).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE mRNA].  
RC TISSUE=Prostate;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,

• • • •

T STRAND 183 187  
FT STRAND 192 194  
FT STRAND 203 206  
FT STRAND 209 216  
FT STRAND 218 222  
FT STRAND 227 231  
FT HELIX 232 235  
FT HELIX 236 245  
SQ SEQUENCE 247 AA; 26558 MW; DD49A487B8062813 CRC64;  
MNPLLLTFV AAALAAPFDD DDKIVGGYNC EENSPYQVS LNSGYHFCGG  
SLINEQWVVS  
AGHCYKSRIQ VRLGEHNIEV LEGNEQFINA AKIIRHPQYD RKTLNNDIML IKLSSRAVIN  
ARVSTISLPT APPATGTKCL ISGWGNTASS GADYPDELQC LDAPVLSQAK CEASYPGKIT  
SNMFCVGFL E GGKDSCQGDS GGPVVCNGQL QGVVSWGDGC AQKNKPGVYT  
KVYNYVKWI  
NTIAANS  
//

# UniProt, a referência universal para sequências de proteínas

- A fusão das bases de dados PIR, TrEMBL e Swiss-Prot numa única base de dados vem constituir uma referência definitiva para a pesquisa de sequências de proteína.
- Uniprot contem as seguintes subsecções:
  - UniProtKB: contem SwissProt e TrEMBL (translated EMBL)
  - UniParc: contem sequências não-anotadas de várias fontes
  - UniRef: contem sequências agrupadas por similaridade

<http://uniprot.org>



UniProt - Mozilla Firefox

File Edit View History Bookmarks ScrapBook Tools Help del.icio.us

Google Calendar Gmail Wiley InterScience: J... .net Tony Schreiner's We... Prediction of Second... Gmail - Inbox Marés - Portos Princi... >

http://beta.uniprot.org/ uniprot

Downloads Contact Help

UniProt

Search in

Protein Knowledgebase (UniProtKB)

Core Data

Protein Knowledgebase (UniProtKB)  
Sequence Clusters (UniRef)  
Sequence Archive (UniParc)

Supporting Data

Literature citations  
Taxonomy  
Keywords

Information

News  
Documents  
FAQ  
Help

Query

Search Clear Fields »

Search Blast Align Retrieve ID Mapping

the scientific community with a  
ely accessible resource of protein

NEWS

Release 12.6 – Dec 4, 2007  
Complete proteome for  
*Arabidopsis thaliana* in UniProtKB

Statistics for UniProtKB:  
Swiss-Prot · TrEMBL

Forthcoming changes

News archives

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: ★ Swiss-Prot, which is manually annotated and reviewed. ★ TrEMBL, which is automatically annotated and is <b>not</b> reviewed.
UniRef	Sequence clusters, used to speed up similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	<a href="#">Literature citations</a> , <a href="#">taxonomy</a> , <a href="#">keywords</a> and more.

SITE TOUR

The UniProtKB/Swiss-Prot database contains manually curated entries, which are annotated by experts in each field. These entries are the result of a combination of automated analysis and manual curation. The UniProtKB/TrEMBL database contains automatically annotated entries, which are generated from large-scale sequencing projects and other sources. The UniProtKB/UniRef database contains sequence clusters, which are used to speed up similarity searches. The UniProtKB/UniParc database contains a sequence archive, which is used to keep track of sequences and their identifiers. The UniProtKB/UniRef database contains supporting data, such as literature citations, taxonomy, keywords and more.

Learn how to make best use of the tools and data on this site.

PROTEIN SPOTLIGHT

Downloads brochure.pdf MAPMBSaug200... surface.csh T00710308TQ1... pass.pdf Clear

Done FoxyProxy: Ualg

insulin in UniProtKB - Mozilla Firefox

File Edit View History Bookmarks ScrapBook Tools Help deljicio.us

http://beta.uniprot.org/uniprot/?query=insulin&sort=score uniprot

Google Calendar Gmail Wiley InterScience: J... net Tony Schreiner's We... Prediction of Second... Gmail - Inbox Marés - Portos Princi... Index of /cd oranger

UniProtKB Downloads Contact Help

Search in Query

Protein Knowledgebase (UniProtKB) insulin Search Clear Fields »

Search Blast Align Retrieve ID Mapping \*

1 - 25 of 2,876 results for insulin in UniProtKB sorted by score descending

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50% | Customize display Download...

Show only reviewed ★ (UniProtKB/Swiss-Prot) or unreviewed ★ (UniProtKB/TrEMBL) entries

Restrict term "insulin" to protein family, gene name, gene ontology, protein name, strain, taxonomy, tissue, web resource

Page 1 of 116 | Next »

	Accession	Entry Name	Status	Protein Names	Genes	Organism	Length
<input type="checkbox"/>	P06213	INSR_HUMAN	★	<b>Insulin receptor precursor</b> (EC 2.7.10.1) (IR) (CD220 antigen) [Cleaved into: Insulin receptor subunit alpha; Insulin receptor subunit beta]	INSR	Homo sapiens (Human)	1,382
<input type="checkbox"/>	P01308	INS_HUMAN	★	<b>Insulin precursor</b> [Cleaved into: Insulin B chain; Insulin A chain]	INS	Homo sapiens (Human)	110
<input type="checkbox"/>	P35568	IRS1_HUMAN	★	<b>Insulin receptor substrate 1</b> (IRS-1)	IRS1	Homo sapiens (Human)	1,242
<input type="checkbox"/>	P09208	INSR_DROME	★	<b>Insulin-like receptor precursor</b> (EC 2.7.10.1) (DIR) (Dlnr) (dlRH) [Cleaved into: Insulin-like receptor subunit alpha; Insulin-like receptor subunit beta 1; Insulin-like receptor subunit beta 2]	InR (dlnr) (Dir-a) (Inr-a) (CG18402)	Drosophila melanogaster (Fruit fly)	2,144

Downloads brochure.pdf MAPMBSaug200... surface.csh T00710308TQ1... pass.pdf Clear

Done FoxyProxy: Ualg

**Insulin receptor precursor - Homo sapiens (Human) - Mozilla Firefox**

File Edit View History Bookmarks ScrapBook Tools Help del.icio.us

Google Calendar Gmail Wiley InterScience: J... net Tony Schreiner's We... Prediction of Second... Gmail - Inbox Marés - Portos Princi... Index of /cd oranger

http://beta.uniprot.org/uniprot/P06213 uniprot

Contribute Send feedback WikiProteins

Reviewed, UniProtKB/Swiss-Prot P06213 (INSR\_HUMAN)

Last modified November 13, 2007. Version 123. History...

Clusters with 100%, 90%, 50% identity | Documents (7) | Third-party data | Customize display TEXT XML RDF/XML GFF FASTA

Names and origin · General annotation (Comments) · Ontologies · Binary interactions · Alternative products · Sequence annotation (Features) · Sequences · References · Web resources · Cross-references · Entry information · Relevant documents

**Names and origin** Hide | Top

Protein names	<b>Insulin receptor</b> [Precursor] Also known as: <a href="#">EC 2.7.10.1</a> IR CD220 antigen Cleaved into: <a href="#">Insulin receptor subunit alpha</a> <a href="#">Insulin receptor subunit beta</a>
Gene names	Name: <b>INSR</b>
Organism	<b>Homo sapiens (Human)</b>
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo
Protein existence	Evidence at protein level.

**General annotation (Comments)** Hide | Top

Function	This receptor binds insulin and has a tyrosine-protein kinase activity. Isoform <a href="#">Short</a> has a higher affinity for insulin. Mediates the metabolic functions of insulin. Binding to insulin stimulates association of the receptor with downstream mediators including IRS1 and phosphatidylinositol 3'-kinase (PI3K). Can activate PI3K either directly by binding to the p85 regulatory subunit, or indirectly via IRS1.
Catalytic activity	ATP + a [protein]-L-tyrosine = ADP + a [protein]-L-tyrosine phosphate.
Enzyme regulation	Autophosphorylation activates the kinase activity.
Subunit structure	Tetramer of 2 alpha and 2 beta chains linked by disulfide bonds. The alpha chains contribute to the formation of the ligand-binding domain, while the beta chains carry the kinase domain. Interacts with SORBS1 but dissociates from it following insulin stimulation. Binds SH2B2. Interacts with the PTB/PID domains of IRS1

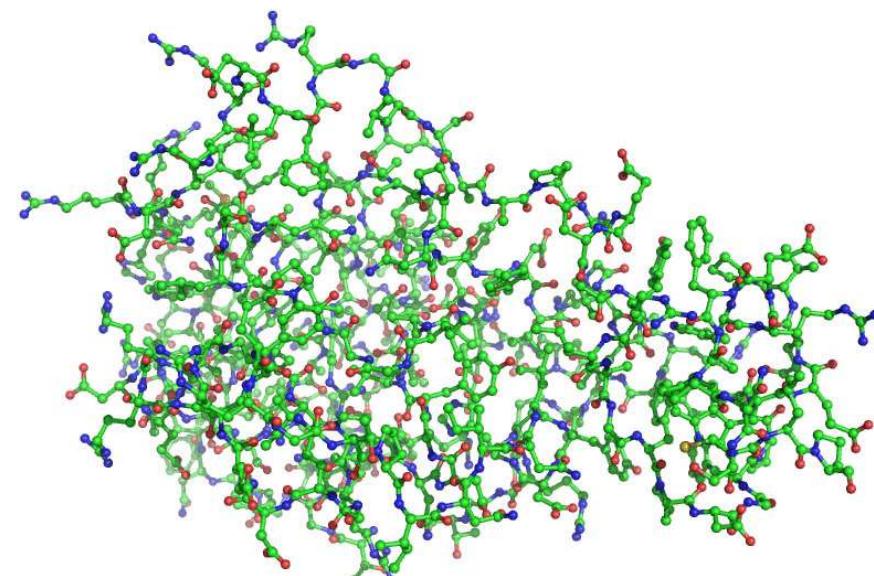
Done FoxyProxy: Ualg

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

Enquanto a sequência de uma proteína ou ácido nucleico é caracterizada simplesmente pela base ou aminoácido que ocorre em cada posição, a descrição duma 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 a molécula

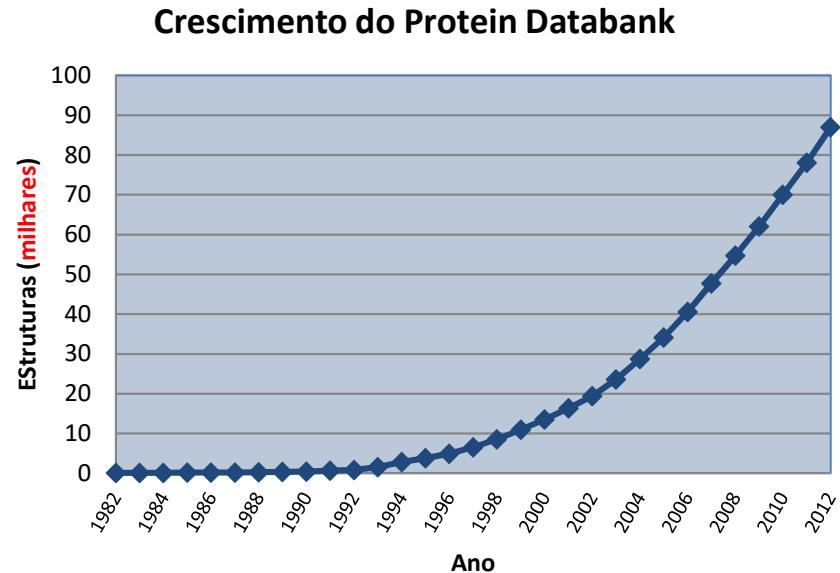
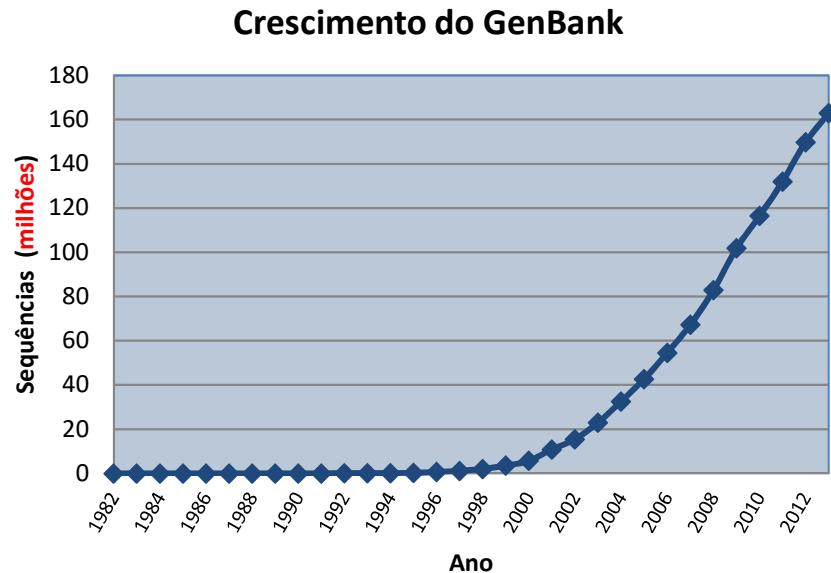
...AVAGGATILVHNQDAGEPAIVLAFG...

Sequência



Estrutura

# Sequência versus estrutura



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

Em 1982: conhecidas 172 estruturas e 315 sequências ...

Hoje (Nov 2014): conhecidas 104,866 estruturas e 274,414,298 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**)!

# 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.
- No caso das proteínas, a topologia de ligação dos 20 aminoácidos standard pode ser assumida *a priori*
- A topologia de outras moléculas, tais como grupos prostéticos, deverá ser especificada
- O formato “tradicional” de representação de estruturas de proteínas é o formato **PDB** (Protein Data Bank file format).
- Para moléculas pequenas usam-se muitos outros formatos, tais como: **cif**, **asn.1**, **mol**, **mdl**, **mol2**, **sdf**, **hin**, ..., ...

# Representação da Aspirina em formato MDL2

```
@<TRIPOS>MOLECULE  
C9H8O4  
21 21 1 0 0  
SMALL  
NO_CHARGES
```

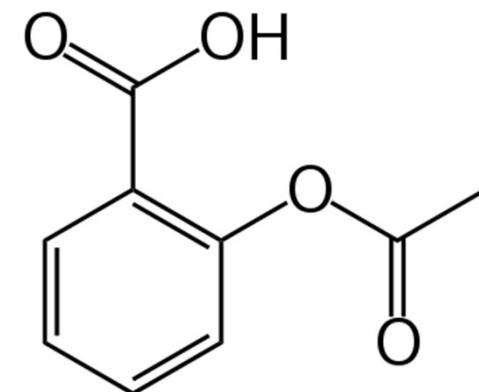
```
@<TRIPOS>ATOM  
1 C1 2.2393 -0.3791 0.2630 C.ar 1 <1> 0.0000  
2 C2 0.8424 1.9231 -0.4249 C.ar 1 <1> 0.0000  
3 C3 2.8709 0.8456 0.2722 C.ar 1 <1> 0.0000  
4 C4 2.1751 1.9935 -0.0703 C.ar 1 <1> 0.0000  
5 C5 -3.4838 0.4953 -0.0896 C.3 1 <1> 0.0000  
6 C6 0.8910 -0.4647 -0.0939 C.ar 1 <1> 0.0000  
7 C7 0.1908 0.6991 -0.4402 C.ar 1 <1> 0.0000  
8 O1 -0.9633 -1.8425 -0.4185 O.2 1 <1> 0.0000  
9 O2 -1.6531 0.8889 1.3406 O.2 1 <1> 0.0000  
10 O3 0.8857 -2.8883 0.2267 O.3 1 <1> 0.0000  
11 C8 0.2090 -1.7720 -0.1069 C.2 1 <1> 0.0000  
12 C9 -2.0185 0.6853 0.2071 C.2 1 <1> 0.0000  
13 O4 -1.1189 0.6285 -0.7886 O.3 1 <1> 0.0000  
14 H1 0.3962 -3.7219 0.2035 H 1 <1> 0.0000  
15 H2 2.7867 -1.2719 0.5268 H 1 <1> 0.0000  
16 H3 0.3069 2.8224 -0.6911 H 1 <1> 0.0000  
17 H4 3.9130 0.9108 0.5482 H 1 <1> 0.0000  
18 H5 2.6781 2.9492 -0.0604 H 1 <1> 0.0000  
19 H6 -3.7360 -0.5623 -0.0120 H 1 <1> 0.0000  
20 H7 -4.0763 1.0637 0.6273 H 1 <1> 0.0000  
21 H8 -3.6988 0.8471 -1.0986 H 1 <1> 0.0000
```

```
@<TRIPOS>BOND
```

1	6	7 ar
2	6	1 ar
3	6	11 1
4	7	2 ar
5	7	13 1
6	1	3 ar
7	11	10 1
8	11	8 2
9	2	4 ar
10	13	12 1
11	12	5 1
12	12	9 2
13	3	4 ar
14	1	15 1
15	2	16 1
16	3	17 1
17	10	14 1
18	4	18 1
19	5	19 1
20	5	20 1
21	5	21 1

Coordenadas

Ligações

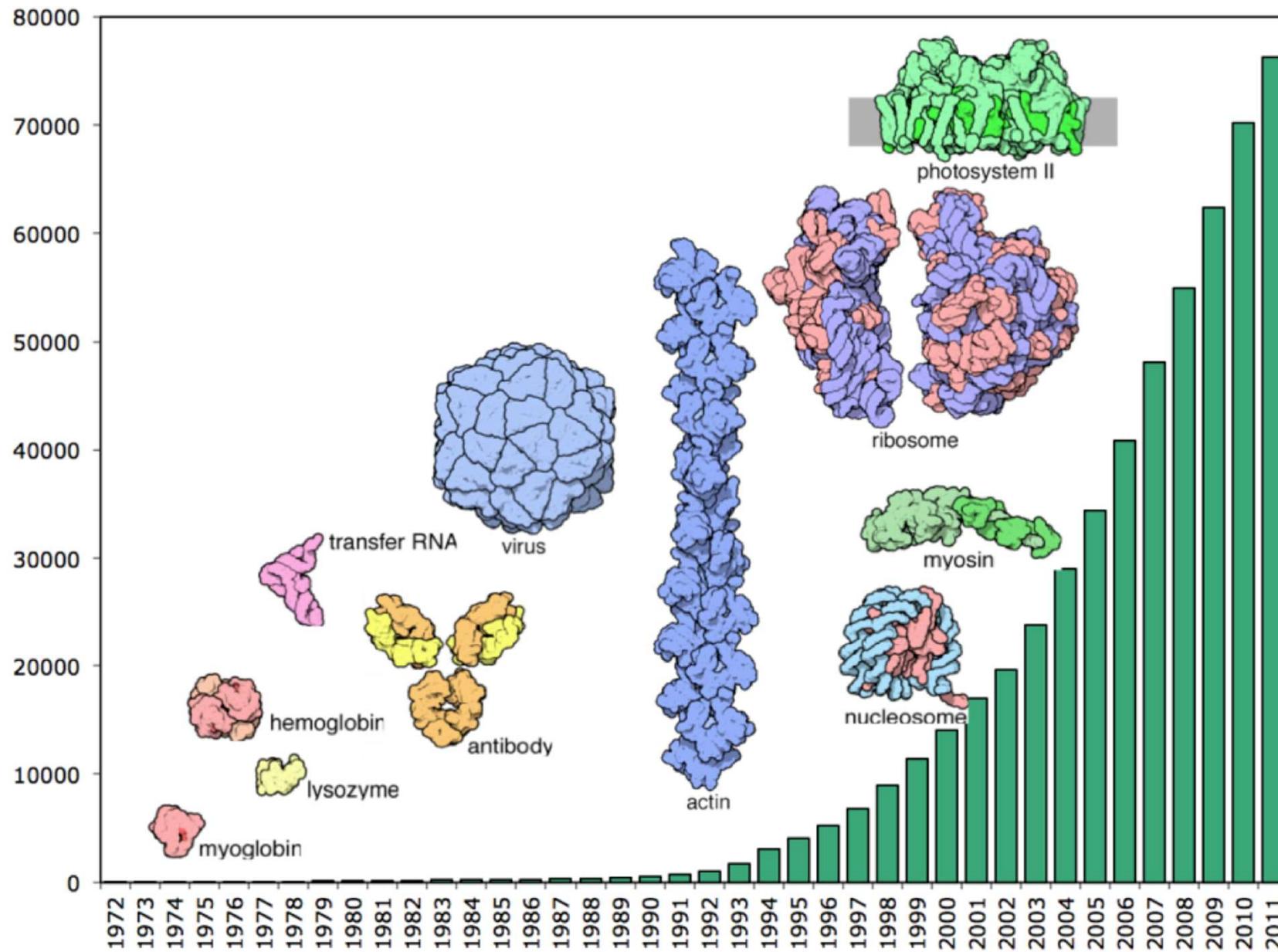


# O Protein Data Bank

- 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 15/02/2018: **137692**

Técnica experimental	Proteínas	Ácidos nucleicos	Complexos Ac.Nuc./Proteína	Outros	Total
Cristalografia de raios X	115449	1905	5889	10	123253
NMR	10642	1234	247	8	12131
Microscopia electrónica	1442	30	498	0	1970
Outras	309	7	8	14	338
Total	127842	3176	6642	32	<b>137692</b>

# Estruturas no Protein Data Bank



# 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, e outros dados associados ao processo de determinação experimental da estruturas
- 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
```

Header

Coordenadas

## Portal de acesso ao PDB

- Acesso ao repositório de estruturas do Protein Databank
- Pesquisa por nomes, sequência, estruturas, ligandos, organismo, método experimental, etc...
- Ferramentas integradas para visualização, comparação de estruturas, análise, etc...

# 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**

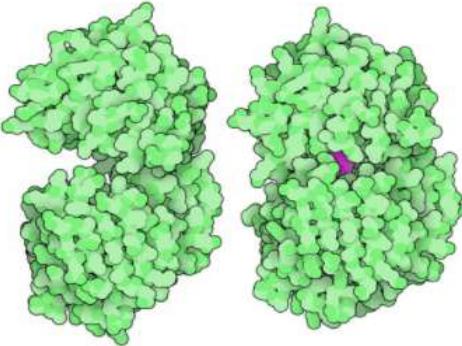
This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

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

termo de pesquisa

The screenshot shows the RCSB Protein Data Bank homepage. At the top, there's a navigation bar with links like 'Home', 'Search', 'Browse', 'About', 'Contact', 'Help', and 'Log In'. Below the navigation is the PDB logo and a link to 'PDB-101'. A banner at the top right says 'An Information Portal to Biological Macromolecular Structures' and 'As of Tuesday Oct 01, 2013 at 5 PM PDT there are 94336 Structures'. The main content area has a search bar with tabs for 'Everything', 'Author', 'Macromolecule', 'Sequence', and 'Ligand'. A red box highlights the 'Everything' tab and the search input field, which contains placeholder text 'e.g., PDB ID, molecule name, author'. To the right of the search bar is a magnifying glass icon. Below the search bar is a 'Search History' link and a 'Previous Results' link. On the left side, there are three sidebar boxes: 'Available on the App Store', 'PDB-101' (with links to 'Structural View of Biology', 'Understanding PDB Data', 'Molecule of the Month', 'Educational Resources', and 'Author Profiles'), and 'MyPDB' (with links to 'Login to your Account', 'Register a New Account', and 'MyPDB Help Page'). On the right side, there's a 'New Features' section with a 'Latest release: September 2013' heading, featuring 'Improved 3D Visualization' (with a 3D molecular model image) and a 'Website Release Archive' dropdown. At the bottom, there's a 'RCSB PDB News' section with links to 'Weekly', 'Quarterly', and 'Yearly' news, and a '2013-10-01 New PDB-101 Animation' link.

# Portal de acesso ao PDB

RCSB PDB - Query Re ×

www.pdb.org/pdb/results/results.do?qr&id=A27431E6&tabtoshow=Current

Search Advanced Browse

Everything Author Macromolecule Sequence Ligand ?

e.g., PDB ID, molecule name, author

Search History (6), Previous Results (102)

**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  
Query Results (102)  
Query History (6)

**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

102 Structure Hits 2 Unreleased Structures 41 Citations 80 Ligand Hits 7 Web Page Hits

**Query Parameters:** Query Details | Save Query to MyPDB

Text Search for: human serum albumin

**Other search suggestions:**

**Query Refinements: Select an item or pie chart ?** Show

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

Enzyme Classification SCOP Classification Protein Symmetry Protein Stoichiometry

**Refine Query with Advanced Search** Show only representatives at Select sequence identity

**1 Related Molecule of the Month articles** Show

Serum Albumin

Showing 1 - 25 of 102 Results Results : 25 Page: 1 of 5

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

**1GNJ** HUMAN SERUM ALBUMIN COMPLEXED WITH CIS-5,8,11,14-EICOSATETRAENOIC ACID (ARACHIDONIC ACID)

Authors: Petitpas, I. Gruene, T. Bhattacharya, A.A. Curry, S.

Release: 2002-01-01 Classification: Plasma Protein

Experiment: X-RAY DIFFRACTION with Residue Count: 585

# Portal de acesso ao PDB

RCSB Protein Data Bank

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

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

A MEMBER OF THE **PDB** | **EMDataBank**  
An Information Portal to Biological Macromolecular Structures  
As of Tuesday Oct 01, 2013 at 5 PM PDT there are 94336 Structures | PDB Statistics | 🔍

Search Advanced Browse

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

Search History (6), Previous Results (102)

**1GNJ** 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  
Query Results (102)  
Query History (6)

**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

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

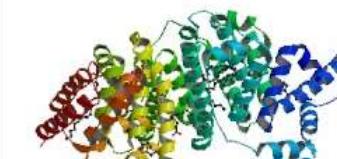
**HUMAN SERUM ALBUMIN COMPLEXED WITH CIS-5,8,11,14-EICOSATETRAENOIC ACID (ARACHIDONIC ACID)**

DOI:10.2210/pdb1gnj/pdb

**Primary Citation**  
Crystal structures of human serum albumin complexed with monounsaturated and polyunsaturated fatty acids.  
Petitpas, I. P., Gruene, T. P., Bhattacharya, A.A. P., Curry, S. P.  
Journal: (2001) J.Mol.Biol. 314: 955  
PubMed: 11743713 🔗  
DOI: 10.1006/jmbi.2000.5208 🔗  
Search Related Articles in PubMed 🔗

**PubMed Abstract:**  
The primary ligands of human serum albumin (HSA), an abundant plasma protein, are non-esterified fatty acids. In vivo, the majority of fatty acids associated with the protein are unsaturated. We present here the first high-resolution crystal structures of HSA complexed with two important unsaturated fatty acids, the monounsaturated oleic acid (C18:1) and the polyunsaturated arachidonic acid (C20:4). Both compounds are observed to occupy the seven binding sites distributed across the protein that are also bound by medium and long-

**Biological Assembly** 🔗  
3D View More Images...  
No symmetry



# Portal de acesso ao PDB

RCSB PDB - Jmol Vie x

www.pdb.org/pdb/explore/jmol.do?structureId=1GNJ&bionumber=1

Browse Search History (6), Previous Results (102)

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 Query Results (102) Query History (6)

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 Drug & Drug Target Mapping

HUMAN SERUM ALBUMIN COMPLEXED WITH CIS-5,8,11,14-EICOSATETRAENOIC ACID (ARACHIDONIC ACID) 1GNJ

Display Files Download Files Share this Page

NOTE: Use your mouse to drag, rotate, and zoom in and out of the structure.



Structure Details

Structure Biological Assembly

Symmetry Type Global Symmetry

Symmetry C1

Stoichiometry A

Select Orientation

Front

Select Display Mode

Secondary Structure

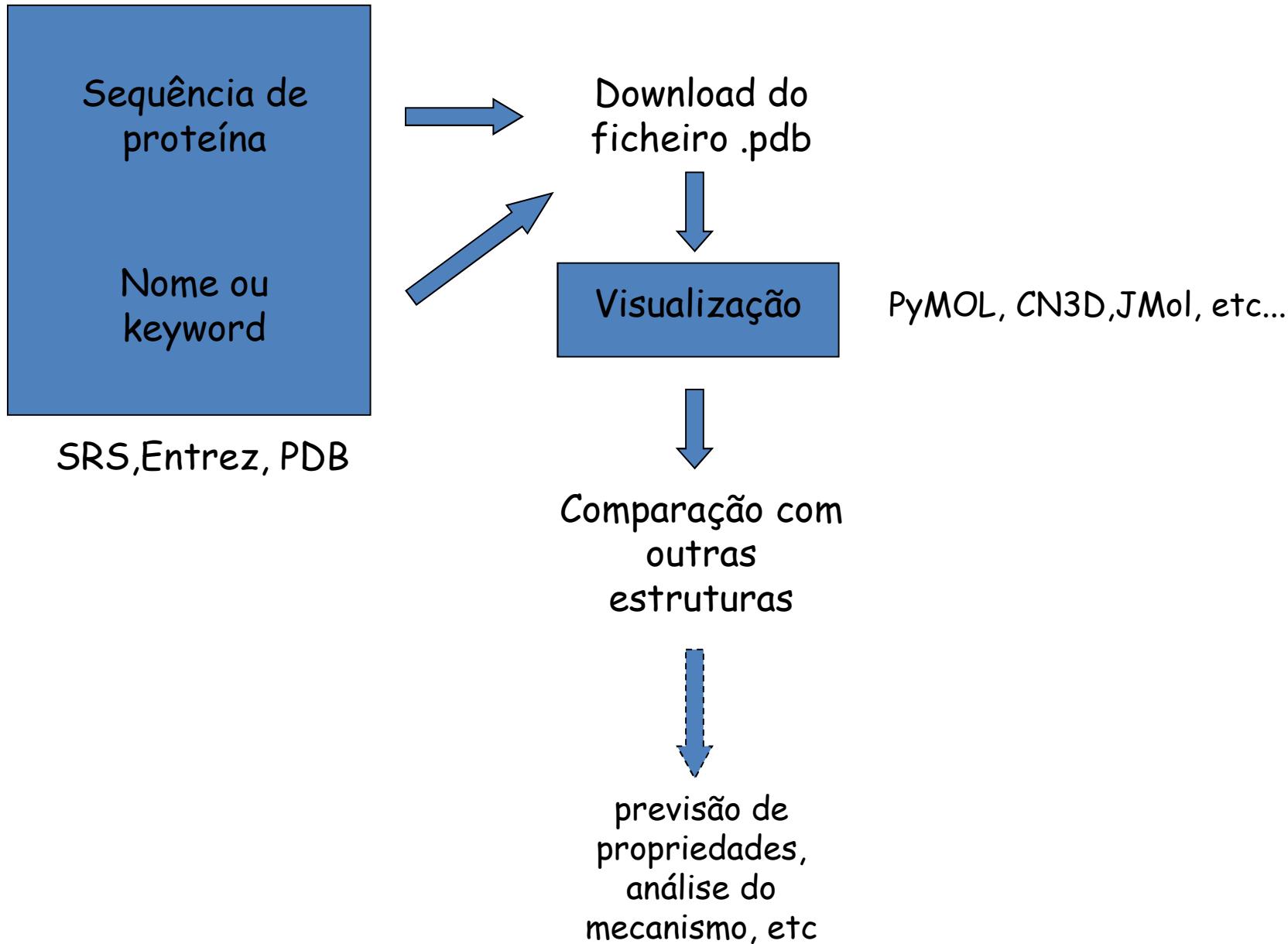
Subunit

Symmetry

Custom View

Export 3D Image

# 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

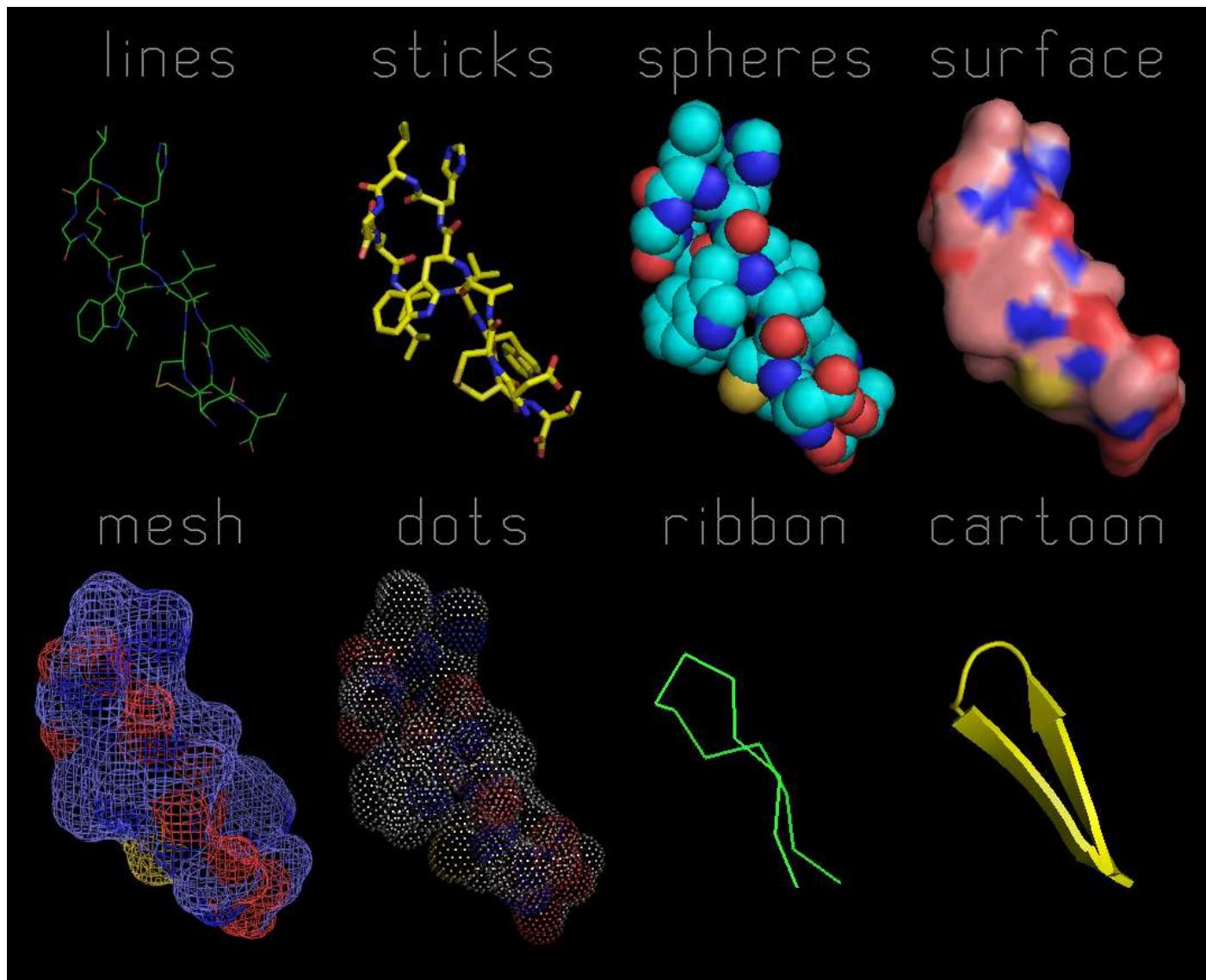
Instaláveis:

- PyMOL: <http://www.pymol.org>
- ICM : <http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>
- QuteMol: <http://qutemol.sourceforge.net/>
- SwissPDB viewer: <http://www.expasy.org/spdbv/>

On-line:

- nglviewr: <http://nglviewer.org/>
- ICMJS: <http://www.molsoft.com>
- Jmol/JSMol: <http://jmol.sourceforge.net/>

# Modos de representação de estruturas





# Bases de dados de pequenas moléculas

- Bases de dados que contêm estruturas de milhares ou milhões de pequenas moléculas , na sua maioria orgânicas
- Ferramenta essencial para o *screening* virtual
- Contêm uma variedade de *descritores* e propriedades das moléculas, umas experimentais, outras calculadas.

# Bases de dados de pequenas moléculas

- PubChem - <http://pubchem.ncbi.nlm.nih.gov>
- DrugBank - <http://www.drugbank.ca>
- ChEMBL - <https://www.ebi.ac.uk/chembl>
- ZINC (purchasable compounds) - <http://zinc.docking.org>
- TCM (traditional chinese medicine) - <http://tcm.cmu.edu.tw>
- CSD (Cambridge Structural Database) -  
<http://webscsd.ccdc.cam.ac.uk>
- ChemDB (database+tools) - <http://www.chemdb.com>
- MOLE DB (molecular descriptors) -  
[http://michem.disat.unimib.it/mole\\_db](http://michem.disat.unimib.it/mole_db)

# PubChem



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

{

9/11/2014

# Bases de dados



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



**PubChem**  
Substance

- Depositor-provided
- Unique Identifier: **SID**



**PubChem**  
Compound

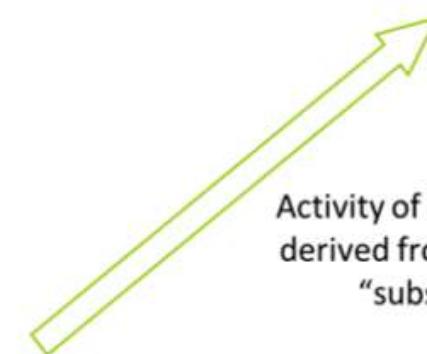
- Unique chemical structures
- Unique Identifier: **CID**

Activity of  
tested  
“substances”



**PubChem**  
BioAssay

- Biological activity test results
- Depositor-provided
- Unique Identifier: **AID**



Activity of “compounds”  
derived from associated  
“substances”



❖ Validate chemical contents

- Atoms defined/real
- Implicit hydrogen
- Functional group
- Atom valence



❖ Normalize representation

- Tautomer invariance
- Aromaticity detection
- Stereochemistry
- Explicit hydrogen



❖ Calculate

- Coordinates
- Properties
- Descriptors



❖ Detect components

- Isolate covalent units
- Neutralize (by  $\pm H^+$  or  $e^-$ )
- Reprocess
- Detect unique components



# Pesquisa



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

The PubChem Project

pubchem.ncbi.nlm.nih.gov

Databases ▾ Upload Services ▾ Help more ▾

**databases**

**PubChem**

BioAssay Compound Substance

GO Advanced Search

Structure Search | BioActivity Analysis | BioActivity DataDicer

New The PubChem Social Media campaign is now launched! see more... more ...

**search tools**

BioActivity Summary  
BioActivity Databable  
BioActivity SAR  
BioActivity DataDicer  
Structure Search  
3D Conformer Tools  
Structure Clustering  
3D Conformer application tool  
Classification  
Upload  
Download  
PubChem FTP

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

# PubChem Compound

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

NCBI Resources How To

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

Save search Limits Advanced Help

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

Results: 1 to 20 of 88

<< First < Prev Page 1 of 5 Next > Last >>

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

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

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

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

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

**Filters:** Manage Filters

**Actions on your results**

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

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

**Chemical Properties**

- Rule of 5 (22)

**BioActivity Experiments**

- BioAssays, Active (13)
- BioAssays, Tested (19)

**Protein 3D Structures (3)**

- Human Transthyretin (ttr) Complexed With Diflunisal (1)

**BioMedical Annotation**

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

**BioSystems (3)**

**Depositor Category**

- Biological Properties (75)
- Chemical Vendors (62)
- Journal Publishers (32)

# PubChem Compound

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

NCBI

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

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

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

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

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

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

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

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

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

**Properties**

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

**BioActivity Data Links**

This Compound  
with Similar Compounds  
with Similar Conformers

**Related Compounds**

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

# PubChem Substance

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

NCBI Resources How To Sign in to NCBI

PubChem Substance aspirin Save search Limits Advanced Search Help

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

Results: 1 to 20 of 547

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

**Actions on your results**

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

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

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

**BioActivity Experiments**  
BioAssays, Active (13)   
BioAssays, Tested (42)   
Protein 3D Structures (38)  
Structural Basis Of The Prevention Of Nsaid-induced Damage Of The Gastrointestinal Tract By C-terminal Half (c-lobe) Of Bovine Colostrum Protein Lactoferrin: Binding And Structural Studies Of The C-lobe Complex With Aspirin (10)

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

**BioSystems** (1)

**Depositor Category**  
Biological Properties (156)

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

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

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

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

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

aspirin: ACETYLSALICYLIC ACID; Ecotrin ...

# PubChem Substance

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

NCBI

PubChem Substance Search Help

SHARE

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

Substance Summary for: SID 49854366

aspirin

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

Table of Contents Show subcontent titles

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

Expand all sub-sections

Chemical Structure:

ASN.1 XML SDF

Follow us on

Related Substances

Same (206)  
Same, Connectivity (222)

Other Links

Chemical Structure Search

# PubChem BioAssay

Screenshot of the PubChem BioAssay interface for BioAssay AID 444512.

The page title is "BioAssay: AID 444512".

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

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

**Table of Contents:**

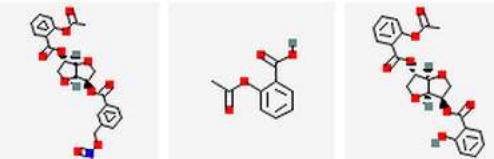
- BioActive Compounds
- Description
- Comment
- Categorized Comment
- Result Definitions
- Data Table (Concise)

**Details:**

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

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

**BioActive Compounds: 3**



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

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

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

**Tested Compounds:**

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

**Tested Substances:**

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

**Links:**

- PubMed (1)
- Taxonomy (1)

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

# PubChem – Pesquisa por “Tag”

Screenshot of the PubChem search interface showing results for compounds matching the "Lipinski's rule of 5" query.

The search bar contains the query: `0:500[mw] 0:5[hbdc] 0:10[hbac] -5:5[logP]`. A red box highlights this search term.

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

**Results: 1 to 20 of 34559871**

**Lipinski's rule of 5** (highlighted in red)

**Actions on your results**

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

**Refine your results**

- What's this?
- Chemical Properties**
- Rule of 5 (34,559,871)
- BioActivity Experiments**
- BioAssays, Probes (142)

**Results List:**

- Methyl 4-ethoxy-3-oxobutanoate; AK141825; 415678-65-8**  
MW: 160.167780 g/mol MF: C<sub>7</sub>H<sub>12</sub>O<sub>4</sub>  
IUPAC name: methyl 4-ethoxy-3-oxobutanoate  
CID: 54303951  
[Summary](#)
- 6-bromo-3-iodopyridin-2-amine; AK142103; 1245643-34-8**  
MW: 298.907130 g/mol MF: C<sub>5</sub>H<sub>4</sub>BrIN<sub>2</sub>  
IUPAC name: 6-bromo-3-iodopyridin-2-amine  
CID: 52987942  
[Summary](#)
- AK138368; 4-(2,2,2-Trifluoroethoxy)pyridin-2-amine; 1379361-82-6**  
MW: 192.138490 g/mol MF: C<sub>7</sub>H<sub>7</sub>F<sub>3</sub>N<sub>2</sub>O  
IUPAC name: 4-(2,2,2-trifluoroethoxy)pyridin-2-amine  
CID: 15724964  
[Summary](#)

# PubChem – Pesquisa por estrutura

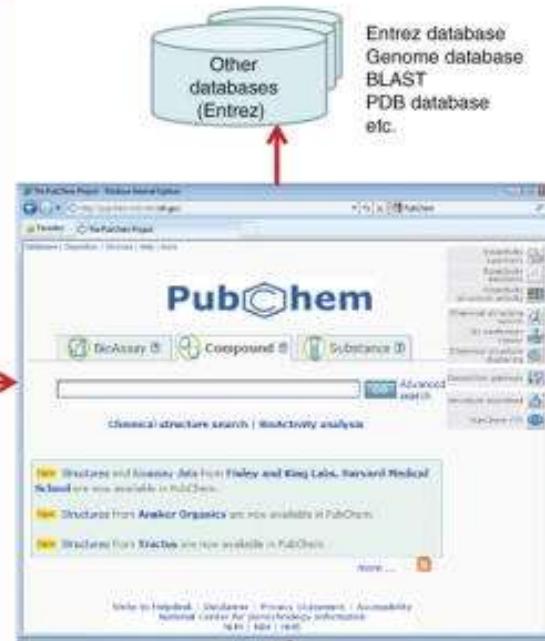
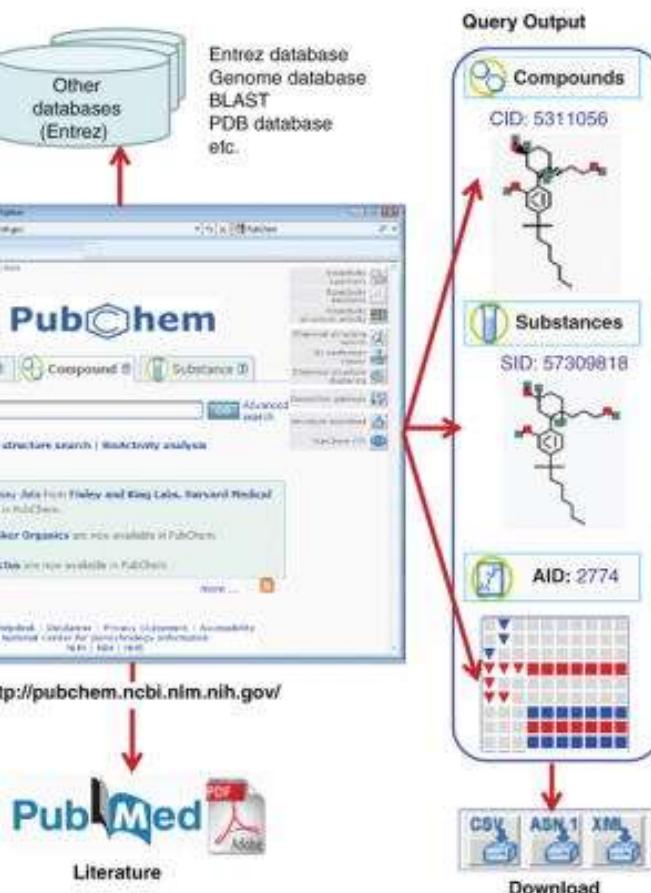
The screenshot shows the PubChem Structure Search interface. On the left, there's a search bar labeled "PubChem Compound" with dropdown menus for "Limits" and "Advanced". Below it is a "Search By:" section with options "Name/Text" and "Identity/Similarity". There are buttons for "Draw a Structure" (with a small chemical sketch icon) and "Edit". Further down are "Options" and "Filters" buttons.

The main area features a "PubChem Sketcher V2.4" window. At the top of this window is a red box highlighting the "SMILES" input field, which contains the string "CCC(N1CCCC1C(=O)O)=O". A red arrow points from the word "SMILES" to this highlighted field. To the right of the input field, the chemical structure of the compound is displayed as a 3D model.

The sketcher interface includes a toolbar with various chemical symbols and a periodic table below it. At the bottom of the sketcher window, there are buttons for "Export" (set to "MDL Molfile"), "Done", "Hydrogen" (set to "Keep AsIs"), and "Import" (with a "Choose File" button).

**A.**

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

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

# ZINC database



- Base de dados de acesso livre
- Contém cerca de 35 milhões de compostos comercialmente disponíveis (purchasable compounds).
- Contém as estruturas tridimensionais dos compostos em formatos de fácil uso para docking e screening virtual
- Possui alguns sub-conjuntos especiais:
  - ZDD – compostos puros aprovados como fármacos pela FDA
  - ZMD – metabolitos primários
  - ZND – derivados de compostos naturais
  - ZBC – compostos biogénicos

.....

....

ZINC01280665 | ZINC Is Not Con X +

Not secure | zinc.docking.org/substance/1280665

Apps AWS LibGen Tut2017 Acad D pmartel Notepad Pasteboard dpaste Trinket Desmos » Other bookmarks

UCSF University of California, San Francisco | About UCSF | Search UCSF | UCSF Medical Center

Shoichet Laboratory docking.org Not Authenticated – sign in

Active cart: Temporary Cart (0 items)

aspirin Go

Synonyms (83) | Vendors (4) | Annotations (20) | Representations (1) | Notes (3) | Targets (0) | Clustered (0) | Reactome (0) | Rings (0) | Analogs (18)

# ZINC01280665

In ZINC since	Heavy atoms	Benign functionality
November 6 <sup>th</sup> , 2005	22	Yes

Popular Name: [Hydrocodone](#)  
Find On: [PubMed](#) – [Wikipedia](#) – [Google](#)  
CAS Numbers: [125-29-1](#), [25968-91-6](#), [34195-34-1](#), [615580-69-3](#)

Other Names:  
[\(-\)-Dihydrocodeinone](#)  
[\(-\)-Dihydrocodeinone; 4,5-alpha-Epoxy-3-methoxy-17-methylmorphinan-6-one](#); [Dihydrocodeinone](#); [Hydrocodon](#); [Hydrocodone](#); [Hydrocone](#); [Hydroconum](#); [Idrocodone](#)  
[\(5alpha\)-17-methyl-3-\(methyloxy\)-4,5-epoxymorphinan-6-one](#)  
[1037-91-8](#)  
[125-29-1](#)

SMILES: C[NH+][C@H]23c4c5ccc(c4O[C@H]2C(=O)CC[C@H]3[C@H]1C5)OC

Download: [MOL2](#) [SDF](#) [SMILES](#) [Flexibase](#)

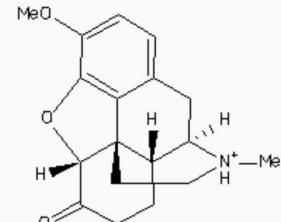
Vendors

<a href="#">American Custom Chemicals Corp.</a>	API0004773, API0002943
<a href="#">IBScreen NP</a>	STOCK1N-54181
<a href="#">Sigma Aldrich (Building Blocks)</a>	H4516 SIGMA
<a href="#">Synthetic Building Blocks</a>	TS-8910, TS69992, TS-70530

Annotations

<a href="#">BindingDB.org</a>	59386689
<a href="#">ChEBI</a>	CHEBI:5779
<a href="#">ChEMBL DrugStore</a>	CHEMBL1200702, CHEMBL1457
<a href="#">ChEMBL12</a>	CHEMBL1200702, CHEMBL1457
<a href="#">ChEMBL10</a>	CHEMBL1457

...docking.org/.../(-)-Dihydrocodeinone; 4,5-alpha-Epoxy-3-methox...



Draw Identity 99% 90% 80% 70%

ZINC01280665 | ZINC Is Not Con X +

Not secure | zinc.docking.org/substance/1280665

Apps AWS LibGen Tut2017 Acad D pmartel Notepad Pasteboard dpaste Trinket Desmos » Other bookmarks

Vendors	Annotations
<a href="#">American Custom Chemicals Corp.</a>	<a href="#">BindingDB.org</a> 59386689 <a href="#">ChEBI</a> CHEBI:5779
<a href="#">IBScreen NP</a>	<a href="#">ChEMBL DrugStore</a> CHEMBL1200702, CHEMBL1457
<a href="#">Sigma Aldrich (Building Blocks)</a>	<a href="#">ChEMBL12</a> CHEMBL1200702, CHEMBL1457
<a href="#">Tetrahedron Building Blocks</a>	<a href="#">ChEMBL19</a> CHEMBL1457, CHEMBL234731, CHEMBL2135756, CHEMBL2062267
	<a href="#">Collaborative Drug Discovery</a> 47623, 326
	<a href="#">DrugBank-approved</a> DB00956
	<a href="#">DrugBank-Street Drugs</a> DB00956
	<a href="#">HMDB Drug</a> HMDB15091
	<a href="#">Human Metabolome Database</a> HMDB15091
	<a href="#">KEGG via PubChem</a> Do8046, Co8024, Do2152, Do3725
	<a href="#">KEGG-C via PubChem</a> Co8024
	<a href="#">KEGG-D via PubChem</a> Do8046, Do2152
	<a href="#">LeadsScope via PubChem</a> LS-92160, LS-54632, LS-187979, LS-174454
	<a href="#">PubChem</a> 44146799, 67043128, 67392166, 11445193, 9830145, 11614091, 11585362, 59253448, 9961898, 5360112, 11247932, 9827514, 5493977, 69808051, 68990195, 11671941, 24847937, 23376922, 49837973, 66813211, 25134193, 49821783, 69725386, 69990013, 46781710, 69267242, 16217589, 6321228, 57512744, 20831824, 66851918, 23616866, 5518315, 67076419, 67505648, 16654983, 23722661,

ZINC01280665 | ZINC Is Not Con X +

Not secure | zinc.docking.org/substance/1280665

Apps AWS LibGen Tut2017 Acad D pmartel Notepad Pasteboard dparse Trinket Desmos Other bookmarks

[SCHEMBL973958](#)

[Sweetlead](#)  
[TTD via PubChem](#)  
[Urine Metabolome](#)

[SWo1815](#)  
[DAPooo253](#)  
[HMDB15091](#)

### Physical Representations

pH range	xlogP	Apolar desolvation (kcal/mol)	Polar desolvation (kcal/mol)	H-bond donors	H-bond acceptors	Net charge	tPSA (Å <sup>2</sup> )	Molecular weight (g/mol)	Rotatable bonds	DL
Reference (pH 7)	1.73	7.78	-61.24	1	4	1	40	300.378	1	↓

### Vendor Notes

Note Type	Comments	Provided By
ALOGPS_SOLUBILITY	7.97e-01 g/l	<a href="#">DrugBank-approved</a>
UniProt Database Links	COR11_PAPSO; COR12_PAPSO; COR13_PAPSO; COR14_PAPSO; COR15_PAPSO	<a href="#">ChEBI</a>
Patent Database Links	EP1518555; EP1557179; EP1584335; EP1604666; EP1604667; EP1623703; EP1639997; EP1685839; EP1700601; EP1741426; EP1776950; EP1782834; EP1797883; EP1813276; EP1815846; EP1820805; EP1829527; EP1829528; EP1849460; EP1897544; EP1897545; EP1938803; EP1961421; EP	<a href="#">ChEBI</a>

### Activity (Go SEA)

#### Analogs ( Draw Identity 99% 90% 80% 70% )

1. [27844230](#)      2. [27844236](#)      3. [27846846](#)      4. [27846853](#)      5. [27846861](#)

6. [27846870](#)      7. [29414775](#)      8. [29463970](#)      9. [40662663](#)      10. [40662666](#)



# Traditional Chinese Medicine (TCM)



- Contem substâncias derivadas de plantas, extractos animais e minerais
- Estruturas tri-dimensionais de compostos presentes nos extractos
- Estruturas disponíveis em formatos 2D e 3D, pré-minimizadas e prontas para usar em docking e screening virtual
- Acesso Livre

搜尋 · 加入會員 · 相關連結 · 聽聽我們

**台灣 中醫藥資料庫**  
Traditional Chinese Medicine Database @ Taiwan

PMF<sub>score</sub> =  $\sum_{i=1}^n \frac{SF_i}{SF_c} \times \sum_{j=1}^m \frac{P_j}{P_{c,j}} \times \frac{SF_i}{SF_c} = a \times pIC_{S0} + b$

PMF<sub>score</sub> =  $\left[ \frac{\sum_{i=1}^n \frac{SF_i}{SF_c} \times \sum_{j=1}^m \frac{P_j}{P_{c,j}} \times \frac{SF_i}{SF_c}}{\sum_{i=1}^n \frac{SF_i}{SF_c}} \right]^{1/(n-1)} \times \left[ \frac{\sum_{i=1}^n \frac{SF_i}{SF_c}}{\sum_{i=1}^n \frac{SF_i}{SF_c}} \right]^{(n-1)/n}$

**化學成份 Chemical Compound**

首頁 > 中醫藥瀏覽 > 化學成份

Chemical Formula: C15H22O5  
 Molecular Weight: 282.332  
 Molecular Volume: 197.91  
 ALogP: 1.998  
 Molecular Polar Surface Area: 53.99  
 Number of Hydrogen Bond Acceptors: 0  
 Number of Hydrogen Bond Donors: 0  
 Number of Rotatable Bonds: 0

Compound: artemisinin

2D結構圖

3D結構圖

( Generate by marvinview )

相關產名: 青蒿 ;

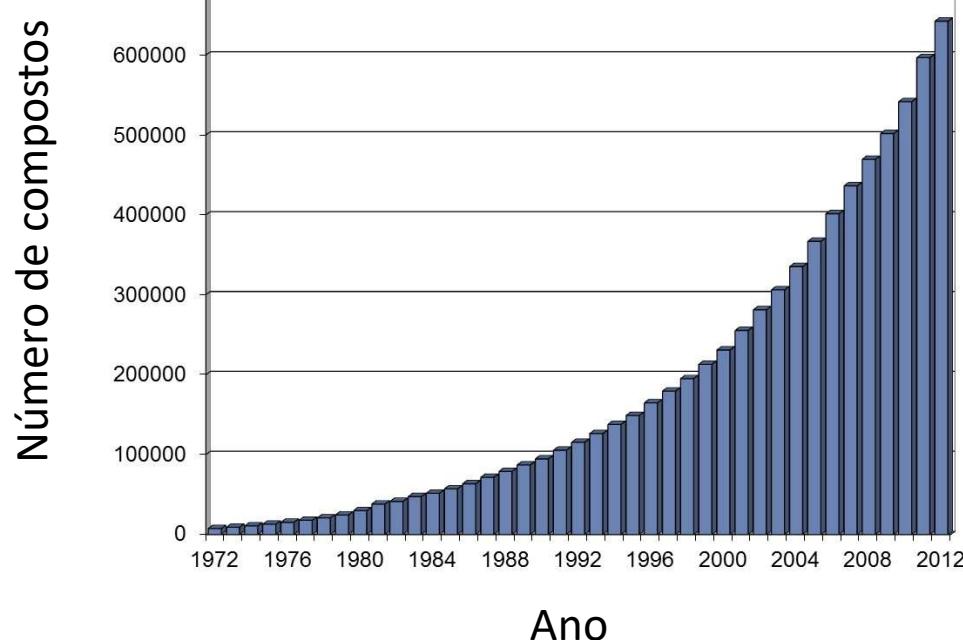
檔案下載: 2D圖 MOL2

# Cambridge Structural Database (CSD)



- Base de dados de estruturas **experimentais** de moléculas pequenas, mantida pelo Cambridge Crystallographic Data Center, UK
- Contem cerca de 700000 compostos **orgânicos** e **organometálicos** determinados por difracção de raios X e de neutrões em cristais individuais e pós.
- É um produto comercial, sem acesso livre. No entanto é possível obter estruturas através de pedidos individuais, desde que para fins não-comerciais.
- É vendida juntamente com o software necessário para a pesquisa, análise e visualização das estruturas (e também o software de docking GOLD)
- Não contém:
  - Polipéptidos e polissacáridos com mais de 24 unidades (ver PDB)
  - Oligonucleótidos
  - Compostos inorgânicos

# CSD - Estatísticas



	Structures	%CSD
Total No. of structures	686 944	100.0
No. of different compounds	628 684	-
No. of literature sources	1 578	-
Organic structures	292 661	42.6
Transition metal present	369 682	53.8
Li – Fr or Be – Ra present	34 433	5.0
Main group metal present	41 711	6.1
3D coordinates present	643 032	93.3
Error-free coordinates	630 329	98.0†
Neutron studies	1 616	0.2
Powder diffraction studies	2 930	0.4
Low/high temp. studies	306 809	44.7
Absolute configuration determined	14 752	2.1
Disorder present in structure	158 127	23.0
Polymorphic structures	20 753	3.0
R-factor < 0.100	645 809	94.0
R-factor < 0.075	585 333	85.2
R-factor < 0.050	378 391	55.1
R-factor < 0.030	78 594	11.4
No. of atoms with 3D coordinates	53 563 990	-

# CSD – Interface WEB

This interactive demo allows you to browse through all 733 entries in the CSD Teaching Database using the standard WebCSD interface. A number of example teaching exercises (including VSEPR and stereochemistry) can be completed using this demo version. A free [hyperlink generator tool](#) is now available.

**ABALEV : (1S,3S)-1,3-bis(4-Bromophenyl)-2-methylpropane-1,3-diol isopropanol solvate**  
V.Gnanadesikan, Y.Horiuchi, T.Ohshima, M.Shibasaki; J.Am.Chem.Soc. (2004), **126**, 7782, doi:[10.1021/ja047906f](https://doi.org/10.1021/ja047906f)

**File Filter Help**

**Find Entry** **ABALEV**

**Entry**

- ABABEL
- ABAFUF
- ABALEV**
- ABCLUA10
- ABEGIY
- ABEKUN
- ABENAX
- ABETOS
- ABIFUM
- ABIKUR
- ABINOS
- ABIZER
- ABUZAY
- ACABAH
- ACABRH02
- ACAJIX
- ACALDA
- ACANIL01
- ACAQUR
- ACARBM01
- ACASED
- ACAZEK
- ACCAAH
- ACCTHP
- ACENYL01
- ACEPOO

< >  
**733 Hits**  
100%  
Stop Search  
Entry loaded

**Hide Viewer**

**Diagram Details Viewer Export Options Help**

**Jmol**

Wireframe ▾ All but C/H ▾  
Hydrogens  Bond types  Disorder   
Packing Options  
None  Unit Cell  3x3x3   
Launch External Viewer

**View Group Symbols Key**

**C<sub>16</sub>H<sub>16</sub>Br<sub>2</sub>O<sub>2</sub>C<sub>3</sub>H<sub>8</sub>O**

**Space Group:** P 2<sub>1</sub>

**a** 10.692(2) **b** 8.858(2) **c** 11.968(2)  
**α** 90 **β** 114.40(1) **γ** 90

**R-Factor:** 3.8%

**Temperature (K):** 200

Added to CSD: 10th November 2004; Last modified: 10th November 2004; Published in WebCSD: 17th September 2009 14:43:54

# CSD – Pedido de estrutura

The screenshot shows a web browser window with the title "Data Request Results" and the URL "www.ccdc.cam.ac.uk/Community/Requestastructure/Pages/DataRequestResponse.aspx". The page displays a search result for query 243822, which returned 1 successful record. The record details a publication in "Journal of Organic Chemistry" (2004), volume 69, issue 4500, doi:10.1021/jo049716t, by Hongbin Li, Hua Yang, J.L. Petersen, Kung K.Wang. Below this, a table summarizes all successful requests, showing one entry for query 243822 with CCDC No. 11.1366(12), Space Group P21/c, and Created On 06/07/2004. A "Download Mode" section offers three options: "Deposited CIF", "Deposited CIF without Structure Factor data", and "Deposited CIF and Structure Factor Files if available". Buttons for "Download Selected", "View in WebCSD", "New Request", and "Email Failed Requests" are also present. A "Conditions of Use of CIFs provided from the CCDC CIF archive" section contains terms of use and a checkbox for agreeing to them. The right sidebar includes links to "Blog", "Awards & Sponsorship", "Deposit a Structure", "Free Services", "Collaborators", "Initiatives", and "Crystal Form Consortium". The top navigation bar includes links for "COMMUNITY", "RESEARCH & CONSULTANCY", "SOLUTIONS", "NEWS & EVENTS", "SUPPORT & RESOURCES", and "THE CCDC".

Website Feedback  >>

COMMUNITY RESEARCH & CONSULTANCY SOLUTIONS NEWS & EVENTS SUPPORT & RESOURCES THE CCDC

Home / Community / Request a Structure / Data Request Results Summary

Your query was: 243822 and returned 1 successful record(s)

Publications

Journal of Organic Chemistry (2004), 69, 4500, doi:10.1021/jo049716t Hongbin Li, Hua Yang, J.L.Petersen, Kung K.Wang

CCDC Structure Summary for All Successful Requests:

Selected	CCDC No	a	b	c	Space Group	Download CIF	View in WebCSD	Created On
<input checked="" type="checkbox"/>	243822	11.1366(12)	6.9872(7)	15.3869(16)	P21/c	<a href="#">Download</a>	ABABEL	06/07/2004

Deposited CIF  
Download Mode Deposited CIF without Structure Factor data  
Deposited CIF and Structure Factor Files if available

You can also download all the selected files at once [Download Selected](#)

[View Selected in WebCSD](#) [View in WebCSD](#) [New Request](#) [Email Failed Requests](#)

Conditions of Use of CIFs provided from the CCDC CIF archive

Individual CIF data sets are provided freely by the CCDC on the understanding that they are used for bona fide research purposes only. They may contain copyright material of the CCDC or of third parties, and may not be copied or further disseminated in any form, whether machine-readable or not, except for the purpose of generating routine backup copies on your local computer system.

I agree to the conditions of use.\*

Your Name \*

Your Email \*

Your Affiliation \*

Blog  
Awards & Sponsorship  
Deposit a Structure  
Free Services  
Collaborators  
Initiatives  
Request a Structure  
Crystal Form Consortium

# Drug Bank

- Base de dados bioinformática e cheminformática
- Contém 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 contém mais de 150 campos

DrugBank

https://www.drugbank.ca

WHAT ARE YOU LOOKING FOR?

Aspirin

Drugs Targets Pathways Indications

The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information.

The latest release of DrugBank (version 5.1.1, released 2018-07-03) contains 11,885 drug entries including 2,528 approved small molecule drugs, 1,184 approved biotech (protein/peptide) drugs, 129 nutraceuticals and over 5,755 experimental drugs. Additionally, 5,132 non-redundant protein (i.e. drug

Acetylsalicylic acid - DrugBank

https://www.drugbank.ca/drugs/DB00945

DRUGBANK

Browse ▾ Search ▾ Downloads About ▾ Help ▾ Blog Contact Us

Drugs

Targets (11) Enzymes (3) Carriers (1) Transporters (3) Biointeractions (16)

# Acetylsalicylic acid

**IDENTIFICATION**

**Name** Acetylsalicylic acid

**Accession Number** DB00945 (APRD00264, EXPT00475)

**Type** Small Molecule

**Groups** Approved, Vet 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**

2-Acetoxybenzenecarboxylic acid

# ChEMBL

- Base de dados mantida e curada manualmente pelo European Bioinformatics Institute (EBI), parte da European Molecular Biology Laboratory (EMBL).
- Contem informação sobre a acção de compostos bioactivos em alvos farmacológicos (drug targets). A informação inclui Ki, Kd, IC50 e EC50.
- Entradas separadas para compostos e alvos.
- A versão mais recente (19) contém 1,411,786 compostos, 10,579 alvos e 12,843,338 ensaios de actividade derivados de 57,156 publicações.
- Contém uma série de ferramentas para análise e filtragem da informação contida na base de dados

# Exemplo de pesquisa estrutural em ChEMBL

The screenshot shows the ChEMBL Database homepage within a web browser. The address bar indicates the URL is https://www.ebi.ac.uk/chembl/b... . The browser interface includes standard controls like back, forward, and search, along with various bookmarked links at the top.

The main header features the ChEMBL logo and navigation links for EMBL-EBI, Services, Research, Training, and About us. A prominent search bar is located on the right side of the header, with the placeholder text "Search in ChEMBL" and a search icon.

Below the header, there's a teal navigation bar with links for UniChem, ChEMBL-NTD, SureChEMBL, Downloads, Web Services, and More. The main content area features a chart titled "Drugs by Usan Year (2015)" showing the number of drugs over time from 1961 to 2016. The chart uses a color scale where darker shades represent higher values (0, 1, 2) and lighter shades represent lower values (3, 4). To the right of the chart, there's a large ChEMBL logo and a descriptive text block about the database.

**Drugs by Usan Year (2015)**

● 0  
● 1  
● 2  
● 3  
● 4

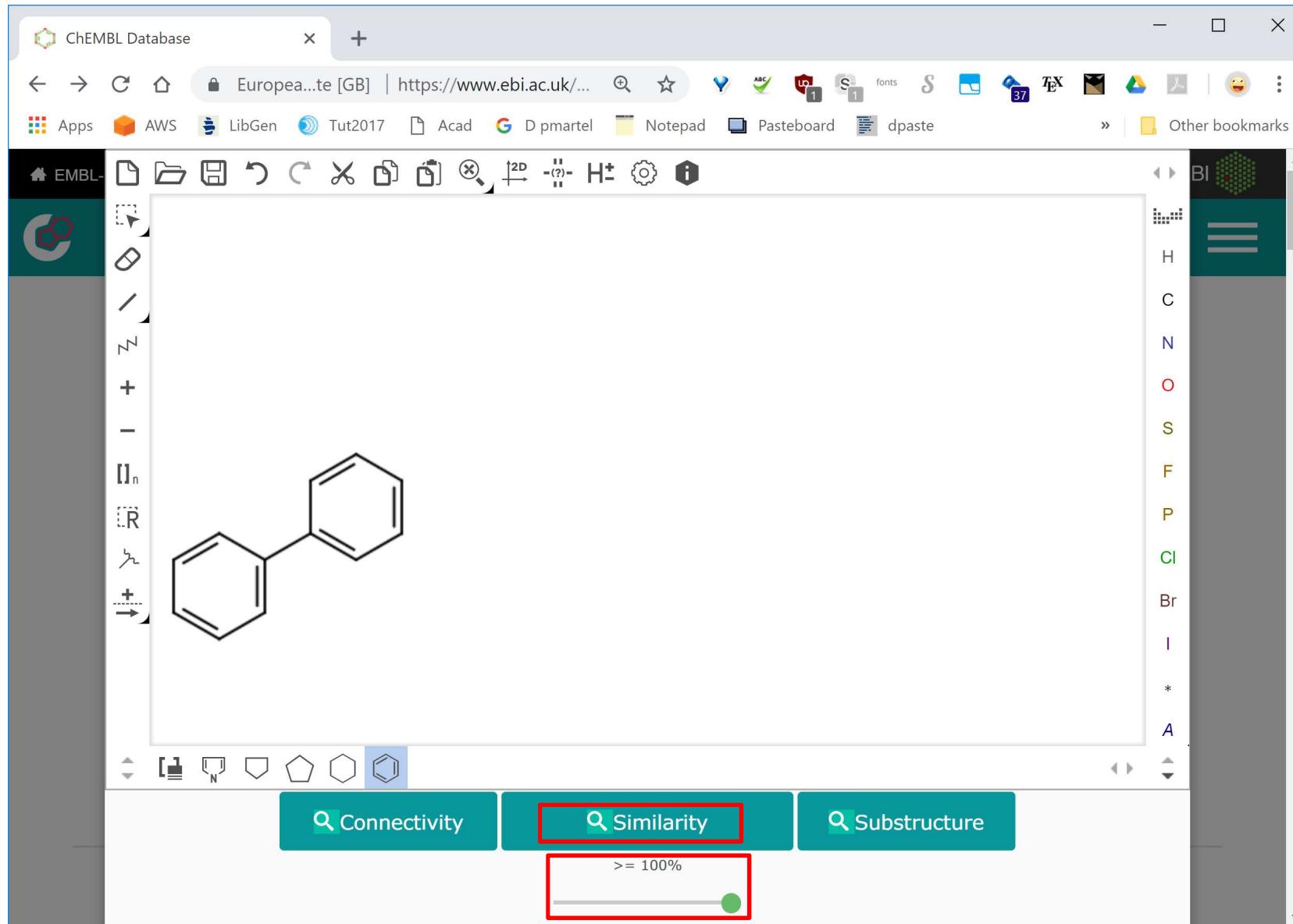
**ChEMBL**

A manually curated database of bioactive molecules with drug-like properties. It brings together chemical, bioactivity and genomic data to aid the translation of genomic information into effective new drugs. See the [interface documentation](#).

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#)

I agree, dismiss this banner

# Exemplo de pesquisa estrutural em ChEMBL



# Exemplo de pesquisa estrutural em ChEMBL

The screenshot shows a web browser window for ChEMBL. The URL is https://www.ebi.ac.uk/chembl/. The search bar contains the query C1=CC=C(C=C1)C1C=CC=CC=1 and a threshold of 100%. The results page displays 1 compound, biphenyl (CHEMBL14092), which is highlighted with a red border. The compound's structure is shown as two fused benzene rings. Below the structure, the identifier CHEMBL14092, name BIPHENYL, and similarity score 100 are listed. The interface includes filters for Type (Small molecule selected), Max Phase (0), #RO5 Violations (0), Molecular Weight ([154.21 to 154.31]), and AlogP.

ChEMBL

Search in ChEMBL

Query: C1=CC=C(C=C1)C1C=CC=CC=1 Threshold: 100% Edit Search

1 Compounds

0 Selected - Select All

Browse Activities

CSV TSV SDF

Filters

Type: Small molecule

Max Phase: 0

#RO5 Violations: 0

Molecular Weight: [154.21 to 154.31]

AlogP

Showing 1-1 out of 1 records

Records per page: 6

Select All

CHEMBL14092

Name: BIPHENYL

Similarity: 100

# Exemplo de pesquisa estrutural em ChEMBL

The screenshot shows a web browser window with the ChEMBL Compound Report Card page. The URL in the address bar is <https://www.ebi.ac.uk/chembl/compound-report-card/CHEMBL14092>. The page title is "Compound Report Card".

The main content area displays the following information for compound CHEMBL14092 (BIPHENYL):

- ID:** CHEMBL14092
- Name:** BIPHENYL
- Max Phase:** 0 Research
- Molecular Formula:** C<sub>12</sub>H<sub>10</sub>
- Molecular Weight:** 154.21
- ChEMBL Synonyms:** E230
- Molecule Type:** Small molecule

On the left side, there is a chemical structure diagram of biphenyl (two benzene rings connected by a single bond). Below the diagram is a toolbar with icons for search, download, and other functions. A red box highlights the minus sign icon in this toolbar.

On the right side, there is a sidebar with a list of links:

- Name And Classification
- Representations
- Sources
- Clinical Data
- Activity Charts
- Literature
- Calculated Properties
- Cross References
- UniChem Cross References
- UniChem Connectivity
- Layer Cross References
- Alternative Forms

# Exemplo de pesquisa estrutural em ChEMBL

The screenshot shows the ChEMBL Compound Report Card interface. A central modal window displays the 3D ball-and-stick model of 4,4'-biphenyl. Below the model, key molecule details are listed:

- Max Phase:** 0 Research
- Molecular Formula:** C<sub>12</sub>H<sub>10</sub>
- Molecular Weight:** 154.21
- ChEMBL Synonyms:** E230
- Molecule Type:** Small molecule

On the left side of the main page, there is a large chemical structure diagram of 4,4'-biphenyl. The right sidebar contains a vertical list of navigation links:

- Name And Classification
- Representations
- Sources
- Clinical Data
- Activity Charts
- Literature
- Calculated Properties
- Cross References
- UniChem Cross References
- UniChem Connectivity
- Layer Cross References
- Alternative Forms

# Exemplo de pesquisa estrutural em ChEMBL

ChEMBL Compound Report Card

Activity Charts

Bioactivity Summary

Assay Summary

Target Summary

Literature

Exclude Alternate Forms Data

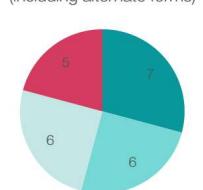
Activity Types for Compound CHEMBL14092 (including alternate forms)



Activity Type	Count
Potency	6
LogP	4
LogP app	2
Papp	2
Ratio	2
logD	1
Binding aff...	1
Concentration	1
IC50	1
Other	5

Exclude Alternate Forms Data

Assays for Compound CHEMBL14092 (including alternate forms)



Assay Type	Count
P - Physicochemical	7
F - Functional	6
B - Binding	6
A - ADME	5

Exclude Alternate Forms Data

Target Classes for Compound CHEMBL14092 (including alternate forms)



Target Class	Count
Enzyme	3
Secreted protein	2
Ion channel	1
Transcript...	1
Unclassifie...	1

# Exemplo de pesquisa estrutural em ChEMBL

The screenshot shows a web browser window with two tabs open. The active tab is titled "Compound Report Card" and displays "UniChem Cross References" for a specific compound. The browser's address bar shows "Europe... [GB] | https://www.ebi.ac....". The bookmarks bar includes links to "ChEMBL", "Europe... [GB]", "LibGen", "Tut2017", "Acad", "Notepad", "Pasteboard", and "dpaste". Below the tabs, there are various icons for file operations like "New", "Open", "Save", "Print", and "Copy/Paste".

Database	Identifier
ACToR	56481-93-7, 92-52-4, 68409-73-4
BindingDB	50168002
Brenda	1885, 107817
ChEBI	17097
ChemicalBook	CB2491271
eMolecules	481835
EPA CompTox Dashboard	DTXSID4020161
FDA SRS	2L9GJK6MGN
Human Metabolome Database	HMDB0034437
IBM Patent System	390A3BB9FB86D9D98D36A1679728E770
KEGG Ligand	C06588
Mcule	MCULE-2274387658
MolPort	MolPort-001-738-537
Nikkaji	J3.929B
NMRShiftDB	10006018
PDBe	BNL

# Exemplo de pesquisa estrutural em ChEMBL

Screenshot of a web browser showing the ChEMBL interface for searching chemical components in the PDB.

The search results for "BIPHENYL" are displayed:

**BNL : Summary**

**Code**: BNL  
**One-letter code**: X  
**Molecule name**: BIPHENYL

**Systematic names**:

Program	Version	Name
ACDLabs	11.02	biphenyl
OpenEye OEToolkits	1.6.1	1,1'-biphenyl

**Formula**: C<sub>12</sub>H<sub>10</sub>  
**Formal charge**: 0  
**Molecular weight**: 154.208 Da

**SMILES**:

Type	Program	Version	Descriptor
SMILES	ACDLabs	11.02	c1cc(ccc1)c2cccc2
SMILES	CACTVS	3.352	c1ccc(cc1)c2cccc2
SMILES	OpenEye OEToolkits	1.7.0	c1ccc(cc1)c2cccc2
Canonical SMILES	CACTVS	3.352	c1ccc(cc1)c2cccc2
Canonical SMILES	OpenEye OEToolkits	1.7.0	c1ccc(cc1)c2cccc2

**IUPAC InChI**: InChI=1S/C12H10/c1-3-7-11(8-4-1)12-9-5-2-6-10-12/h1-10H  
**IUPAC InChI key**: ZUOUZKKEUPVFJK-UHFFFAOYSA-N

**Chemical Components in the PDB**

**wwPDB Information**

Atom count	22 (12 without Hydrogen)
Polymer type	Bound ligand
Type description	NON-POLYMER
Type code	HETAIN
Is modified	No
Standard parent	Not Assigned
Defined at	2003-09-15
Last modified at	2011-06-04
Status	Released
Obsoleted	Not Assigned

**PDBe** is a member of **EMDataBank**

# Exemplo de pesquisa estrutural em ChEMBL

The screenshot shows a web browser window with several tabs open. The active tab is titled "PDBeChem: Ligand Dicti". The main content area is titled "Chemical Components in the PDB" and shows a search result for the molecule "BNL". The search results table has the following columns: Ligand Code, PDB Entry ID, Type, Total, and Distinct. The table lists five entries:

Ligand Code	PDB Entry ID	Type	Total	Distinct
BNL	1ulj	Bound ligand	3	1
BNL	2gbx	Bound ligand	3	1
BNL	2xrz	Bound ligand	12	1
BNL	3gzx	Bound ligand	1	1
BNL	5aew	Bound ligand	9	1

The "1ulj" entry is highlighted with a red box. The browser's address bar shows the URL: "Not secure | www.ebi.ac.uk/pdbe-srv/pdb...". The status bar at the bottom of the browser window indicates "40 fonts" and "40 images".

# Exemplo de pesquisa estrutural em ChEMBL

Screenshot of a web browser showing the Protein Data Bank in Europe (PDB) entry for PDB ID 1ulj. The page displays structural information, experimental validation, and biological context.

**PDBBe > 1ulj**

Biphenyl dioxygenase (BphA1A2) in complex with the substrate

**Source organism:** Rhodococcus jostii RHA1

**Primary publication:**

- Crystal structure of the terminal oxygenase component of biphenyl dioxygenase derived from Rhodococcus sp. strain RHA1.
- Furusawa Y, Nagarajan V, Tanokura M, Masai E, Fukuda M, Senda T
- J. Mol. Biol. 342 1041-52 (2004)
- PMID: 15342255

**X-ray diffraction**  
2.6 Å resolution  
Released: 28 Sep 2004

**Ligands and Environments**

3 bound ligands:

- Fe<sup>+2</sup> (3 x FE2)
- Fe-S (3 x FES)
- Biphenyl (3 x BNL)

No modified residues

**Function and Biology**

**Reaction catalysed:**

$$\text{Biphenyl} + \text{NADH} + \text{O}_2 = (1S,2R)-3\text{-phenylcyclohexa-3,5-diene-1,2-diol} + \text{NAD}^+$$

**Biochemical function:** biphenyl 2,3-dioxygenase activity

**Biological process:** oxidation-reduction process

**Cellular component:** not assigned

**Sequence domains:**

- Aromatic-ring-hydroxylating dioxygenase, alpha subunit
- Ring-hydroxylating dioxygenase beta subunit
- Rieske [2Fe-2S] iron-sulphur domain
- Rieske [2Fe-2S] iron-sulphur domain superfamily
- NTF2-like domain superfamily
- Aromatic-ring-hydroxylating dioxygenase, alpha subunit, C-terminal domain
- Aromatic-ring-hydroxylating dioxygenase, 2Fe-2S-binding site

**Structure domains:**

- Ring hydroxylating alpha subunit ISP domain
- Diahydroxyacetone phosphate binding domain

**Experiments and Validation**

Metric Percentile Ranks Value

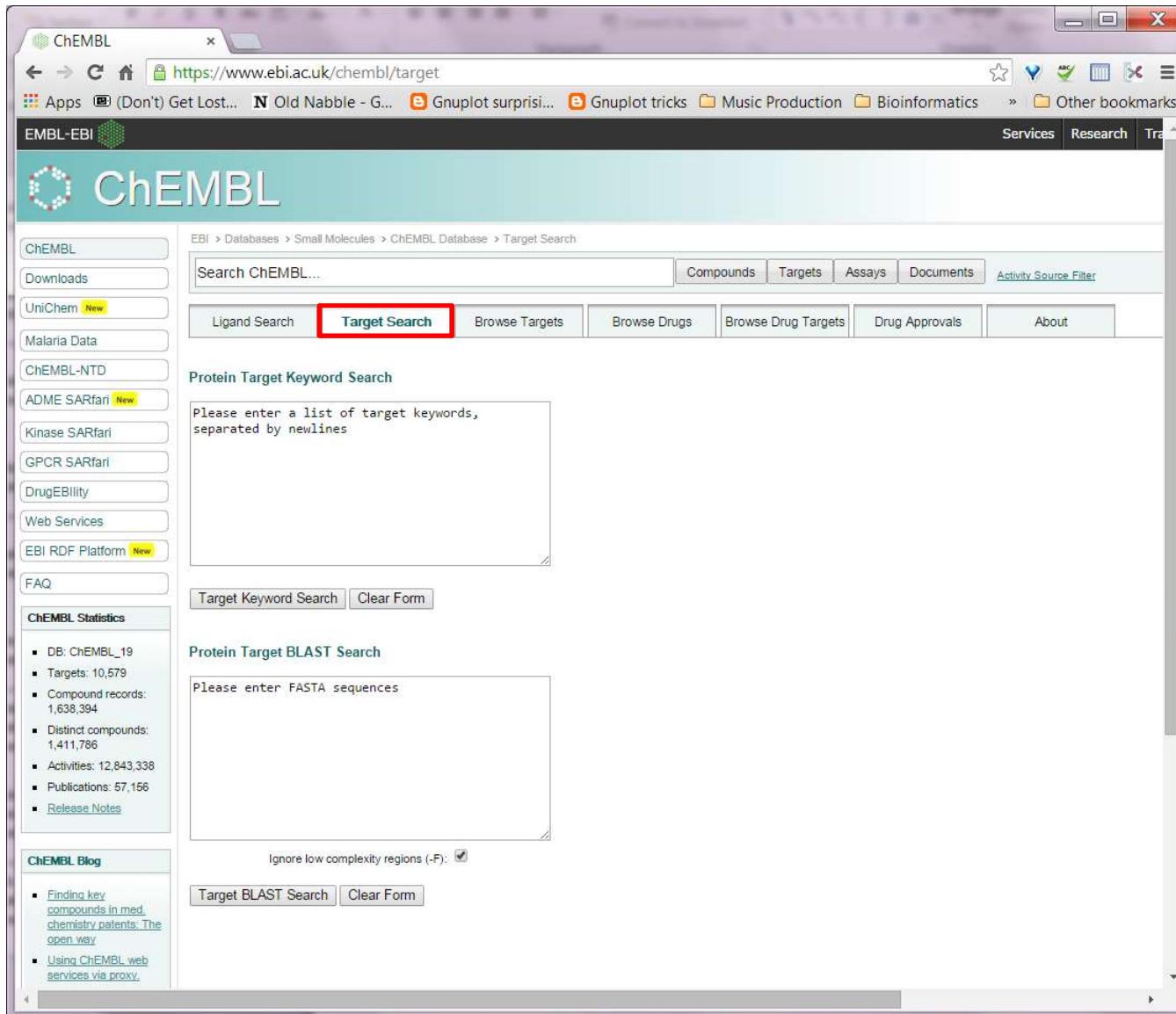
Metric	Percentile Ranks	Value
Rfree	3	0.237
Clashscore	3	0.2%
Ramachandran outliers	5.8%	0.9%
Sidechain outliers	5.8%	0.9%
RSRZ outliers	5.8%	0.9%

**PDB\_REDQ**

The sliders below show the change in model quality between original PDB entry and the PDB\_REDQ entry

Model Geometry Fit model/data

# Pesquisa de targets em ChEMBL



The screenshot shows the ChEMBL Target Search interface. At the top, there is a navigation bar with links to Apps, EBI, and other bookmarked sites. Below the navigation bar, the ChEMBL logo is displayed. The main content area has a header "ChEMBL" and a sub-header "Target Search". A sidebar on the left contains links to various ChEMBL services like ChEMBL, Downloads, UniChem, Malaria Data, ChEMBL-NTD, ADME SARfari, Kinase SARfari, GPCR SARfari, DrugEBility, Web Services, and EBI RDF Platform. Another sidebar on the right lists ChEMBL Statistics: DB: ChEMBL\_19, Targets: 10,579, Compound records: 1,638,394, Distinct compounds: 1,411,786, Activities: 12,843,338, Publications: 57,156, and a link to Release Notes. The central search area features a "Search ChEMBL..." input field and a "Target Search" button, which is highlighted with a red box. Below this are two search sections: "Protein Target Keyword Search" and "Protein Target BLAST Search". Both sections have text input fields for entering target keywords or FASTA sequences. There are also "Target Keyword Search" and "Target BLAST Search" buttons, along with "Clear Form" buttons. A checkbox for "Ignore low complexity regions (-F)" is present in the BLAST search section. The bottom of the page includes a footer with a "ChEMBL Blog" section containing links to posts about key compounds in med. chemistry patents and using ChEMBL web services via proxy.

# Pesquisa de targets em ChEMBL

The screenshot shows the ChEMBL Target Search interface. The URL in the browser is <https://www.ebi.ac.uk/chembl/target>. The search bar contains the keyword "Trypsin". The "Target Search" tab is selected. On the left sidebar, under "ChEMBL Statistics", there is a list of database metrics:

- DB: ChEMBL\_19
- Targets: 10,579
- Compound records: 1,638,394
- Distinct compounds: 1,411,766
- Activities: 12,843,338
- Publications: 57,156
- [Release Notes](#)

Below the statistics, the "ChEMBL Blog" section has two entries:

- Finding key compounds in med. chemistry patents: The open way
- Using ChEMBL web services via proxy.

At the bottom of the search form, there is a checkbox labeled "Ignore low complexity regions (-F):" followed by a checked input field. Below the search form are "Target BLAST Search" and "Clear Form" buttons.

# Pesquisa de targets em ChEMBL

The screenshot shows a web browser window for the ChEMBL website (<https://www.ebi.ac.uk/chembl/index.php/target/results/ids>). The search bar at the top contains the query "Search ChEMBL...". Below the search bar, there are links for "Compounds", "Targets", "Assays", and "Documents", along with an "Activity Source Filter". The main content area displays the "ChEMBL Target Search Results: 22". A dropdown menu "Please select...." is visible. The results table has columns: ChEMBL ID, Preferred Name, UniProt Accession, Target Type, Organism, Compounds, Bioactivities, and a checkbox column. The first row, which is highlighted with a red border, corresponds to the entry: CHEMBL209, Trypsin I, P07477, SINGLE PROTEIN, Homo sapiens, 1673, 1864, and a checked checkbox. The table also includes other entries such as CHEMBL3769 (Trypsin I, P00760), CHEMBL3314 (Alpha-chymotrypsin, P00766), and CHEMBL2095204 (Trypsin, P07477, P35030, P07478). The bottom of the page shows navigation links for "Previous" (1, 2, 3), "Next", and "Showing 1 to 10 of 22 entries".

ChEMBL ID	Preferred Name	UniProt Accession	Target Type	Organism	Compounds	Bioactivities	<input checked="" type="checkbox"/>
CHEMBL209	Trypsin I	P07477	SINGLE PROTEIN	Homo sapiens	1673	1864	<input checked="" type="checkbox"/>
CHEMBL3769	Trypsin I	P00760	SINGLE PROTEIN	Bos taurus	1005	1148	<input checked="" type="checkbox"/>
CHEMBL3314	Alpha-chymotrypsin	P00766	SINGLE PROTEIN	Bos taurus	539	678	<input checked="" type="checkbox"/>
CHEMBL2095204	Trypsin	P07477, P35030, P07478	PROTEIN FAMILY	Homo sapiens	626	657	<input checked="" type="checkbox"/>
CHEMBL2407	Elastase 2A	P08419	SINGLE PROTEIN	Sus scrofa	334	403	<input checked="" type="checkbox"/>
CHEMBL2386	Chymotrypsin C	Q99895	SINGLE PROTEIN	Homo sapiens	323	378	<input checked="" type="checkbox"/>
CHEMBL2366	Trypsin	P00761	SINGLE PROTEIN	Sus scrofa	328	331	<input checked="" type="checkbox"/>
CHEMBL2096988	Thrombin & trypsin	P07478, P00734, P07477, P35030	SELECTIVITY GROUP	Homo sapiens	254	268	<input checked="" type="checkbox"/>
CHEMBL4796	Beta-chymotrypsin	P17538	SINGLE PROTEIN	Homo sapiens	186	261	<input checked="" type="checkbox"/>
CHEMBL2096984	Pancreatic elastase	P00772, P08419	PROTEIN FAMILY	Sus scrofa	184	228	<input checked="" type="checkbox"/>

# Pesquisa de targets em ChEMBL

CHEMBL209 Target x https://www.ebi.ac.uk/chembl/target/inspect/CHEMBL209

Apps (Don't) Get Lost... Old Nabble - G... Gnuplot surprisi... Gnuplot tricks Music Production Bioinformatics Other bookmarks

EMBL-EBI Services Research Training About us

ChEMBL wellcome trust

EBI > Databases > Small Molecules > ChEMBL Database

**Target Report Card**

**Target Name and Classification**

Target ID	CHEMBL209
Target Type	SINGLE PROTEIN
Preferred Name	Trypsin I
Synonyms	Alpha-trypsin chain 1   Alpha-trypsin chain 2   Beta-trypsin   Cationic trypsinogen   PRSS1   Serine protease 1   TRP1   TRY1   TRYPI   Trypsin I   Trypsin-1
Organism	Homo sapiens
Species Group	No
Protein Target Classification	• enzyme > protease > serine protease > serine protease pa clan > serine protease s1a subfamily

**Target Components**

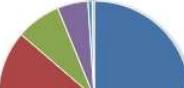
Component Description	Relationship	Accession
Trypsin-1	SINGLE PROTEIN	P07477

**Target Relations**

ChEMBL ID	Pref Name	Target Type
CHEMBL2096988	Thrombin & trypsin	SELECTIVITY GROUP
CHEMBL2095204	Trypsin	PROTEIN FAMILY

**Target Associated Bioactivities**

ChEMBL Activity Types for Target CHEMBL209

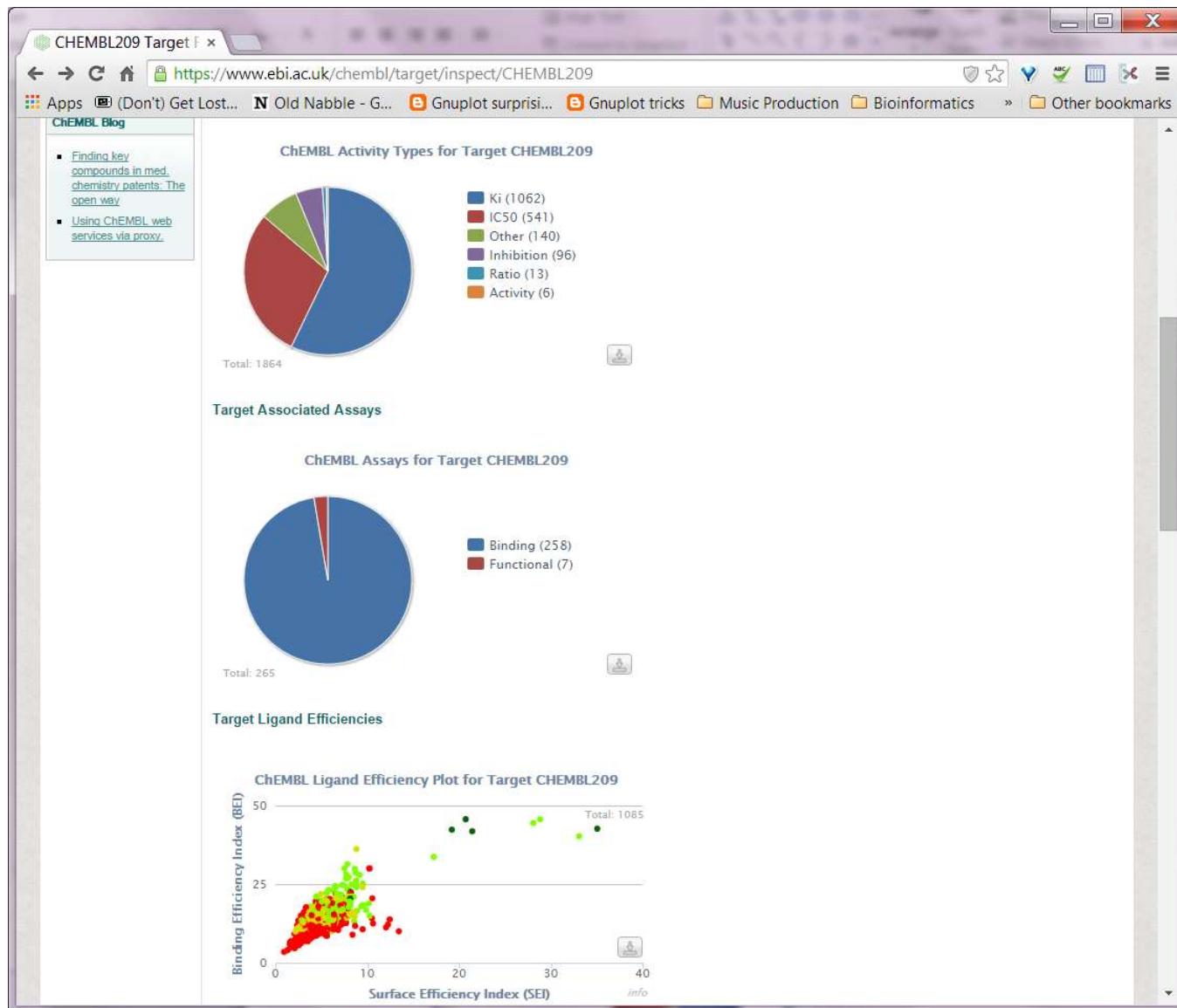


Activity Type	Count
Ki	1062
IC50	541
Other	140
Inhibition	96
Others	11

**ChEMBL Blog**

- Finding key compounds in med. chemistry patents: The open way
- Using ChEMBL web services via proxy.

# Pesquisa de targets em ChEMBL

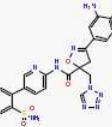


# Pesquisa de targets em ChEMBL

Screenshot of the ChEMBL Bioactivity Search Results page (https://www.ebi.ac.uk/chembl/bioactivity/results/1/cmpd\_chemblid/asc/tabc/display).

The search results show 1068 records per page. The results table includes columns for Ingredient, Molweight, Standard Type, Relation, Standard Value, Standard Units, Assay Type, Description, Assay Src Description, Assay Organism, Target Type, Target Name, Target Organism, and Reference.

Three rows of results are displayed, each showing a compound with a chemical structure (CHEMBL268225), a Ki value of 553.98 nM, and a B assay type. The description for all three rows is "Inhibitory activity against trypsin using human purified enzyme". The assay source is "Scientific Literature" and the target is "SINGLE PROTEIN". The target name is "Trypsin" and the target organism is "Homo sapiens". The reference is "Bioorg. Med. Chem. Lett. (2003) 13:6:1023".

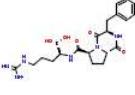
Ingredient	Molweight	Standard Type	Relation	Standard Value	Standard Units	Assay Type	Description	Assay Src Description	Assay Organism	Target Type	Target Name	Target Organism	Reference
	553.98	Ki	>	1600	nM	B	Inhibitory activity against trypsin using human purified enzyme	Scientific Literature		SINGLE PROTEIN	Trypsin	Homo sapiens	<a href="#">Bioorg. Med. Chem. Lett. (2003) 13:6:1023</a>
	553.98	Ki	>	1600	nM	B	Inhibitory activity against trypsin using human purified enzyme	Scientific Literature		SINGLE PROTEIN	Trypsin	Homo sapiens	<a href="#">Bioorg. Med. Chem. Lett. (2003) 13:6:1023</a>
	553.98	Ki	>	1600	nM	B	Inhibitory activity against trypsin using human purified enzyme	Scientific Literature		SINGLE PROTEIN	Trypsin	Homo sapiens	<a href="#">Bioorg. Med. Chem. Lett. (2003) 13:6:1023</a>

# Pesquisa de targets em ChEMBL

Screenshot of the ChEMBL Bioactivity Search Results page (https://www.ebi.ac.uk/chembl/bioactivity/results/1/cmpd\_chemblid/asc/tab/display).

The search results show 1068 entries. The table includes columns for Ingredient, Molweight, Standard Type, Relation, Standard Value, Standard Units, Assay Type, Description, Assay Src Description, Assay Organism, Target Type, Target Name, Target Organism, and Reference.

Key data from the first four rows:

Ingredient	Molweight	Standard Type	Relation	Standard Value	Standard Units	Assay Type	Description	Assay Src Description	Assay Organism	Target Type	Target Name	Target Organism	Reference
		Ki	=	.045	nM	B	<a href="#">Compound was tested in vitro for inhibition of trypsin</a>	Scientific Literature		SINGLE PROTEIN	<a href="#">Trypsin</a>	Homo sapiens	<a href="#">CHEMBL1149934</a>
		Ki	=	.045	nM	B	<a href="#">Inhibition of trypsin</a>	Scientific Literature		SINGLE PROTEIN	<a href="#">Trypsin</a>	Homo sapiens	<a href="#">CHEMBL1151536</a>
		Ki	=	.045	nM	B	<a href="#">Tested for selectivity against trypsin</a>	Scientific Literature		SINGLE PROTEIN	<a href="#">Trypsin</a>	Homo sapiens	<a href="#">CHEMBL1129866</a>
		Ki	=	.045	nM	B	<a href="#">Binding affinity against Trypsin</a>	Scientific Literature		SINGLE PROTEIN	<a href="#">Trypsin</a>	Homo sapiens	<a href="#">J. Med. Chem. (2000) 43:3:305</a>

# Pesquisa de targets em ChEMBL

Compound Report Card <https://www.ebi.ac.uk/chembl/compound/inspect/CHEMBL290376>

EMBL-EBI wellcome trust

ChEMBL

Compound Name and Classification

Compound ID	CHEMBL290376
Compound Name	
ChEMBL Synonyms	DuP-714
Max Phase	0
Trade Names	
Molecular Formula	C <sub>21</sub> H <sub>33</sub> BN <sub>6</sub> O <sub>5</sub>

Chemical structure of DuP-714 (CHEMBL290376):

Additional synonyms for CHEMBL290376 found using NCI Chemical Identifier Resolver

Compound Representations

Molfile	<a href="#">Download MolFile</a>
Canonical SMILES	CC(=O)N[C@H](Cc1ccccc1)C(=O)N2CCC[C@H]2C(=O)N[C@@H](CCNC(=N ...] <a href="#">Download SMILES</a>
Standard InChI	InChI=1S/C21H33BN6O5/c1-14(29)26-16(13-15-7-3-2-4-8-15)20(31 ... <a href="#">Download InChI</a>
Standard InChI Key	FXFYPTZERULUBS-SQNIBIBYSA-N

Alternate Forms of Compound in ChEMBL

Chemical structure of DuP-714 (CHEMBL290376):

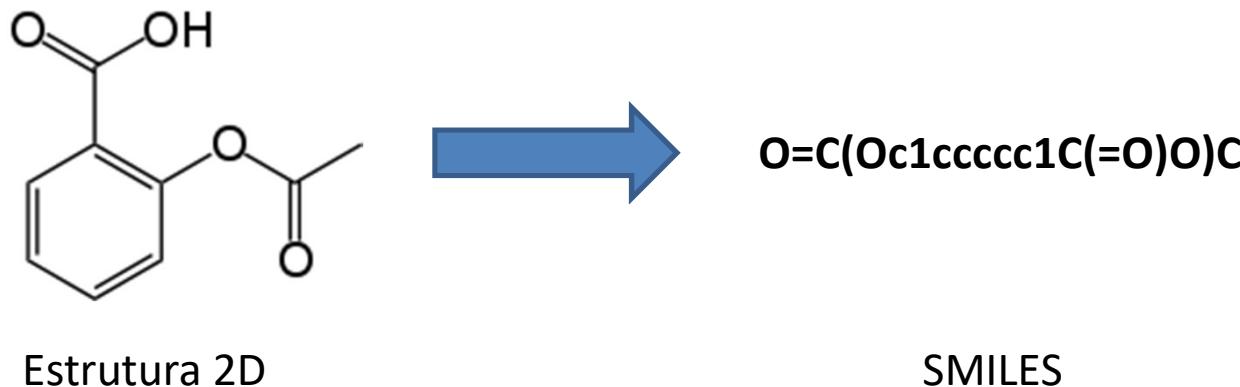
Links:

- Finding key compounds in med. chemistry patents: The open way
- Using ChEMBL web services via proxy

# SMILES

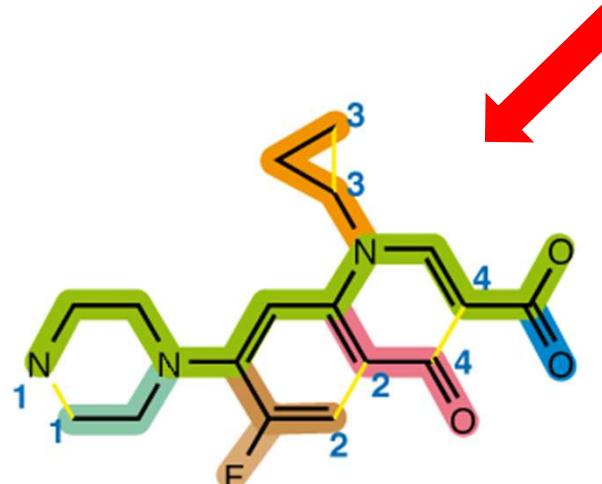
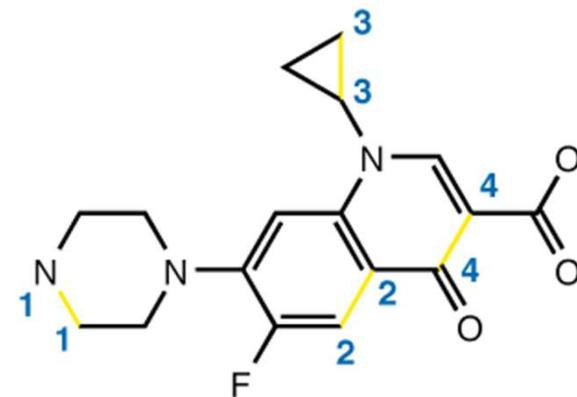
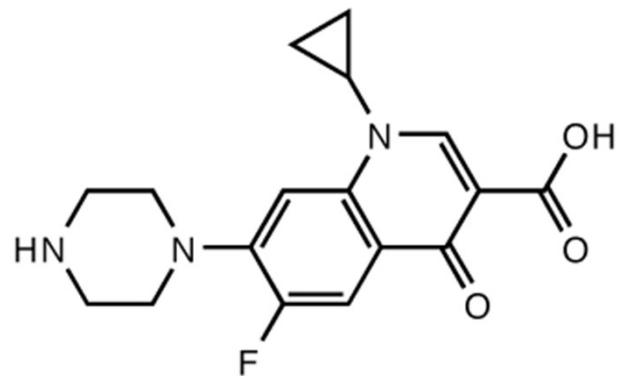
**SMILES** - Simplified Input Molecular Entry Specification

Linguagem que permite a representação de estruturas moleculares 2D na forma de uma sequência (“string”) de caracteres.



Tutorial SMILES: <http://www.daylight.com/>

D. Weininger (1988) *J. Chem. Inf. Comput. Sci.* **28**:31



N1CCN(CC1)C(C(F)=C2)=CC(=C2C4=O)N(C3CC3)C=C4C(=O)O



# SMILES – Regras(1)

Os átomos são representados pelos seus nomes elementais:

C B N O P S Cl Br I H (compostos orgânicos)

- Outros elementos – [Si] [Fe] [Co]
- O hidrogénio é geralmente ignorado: CH<sub>4</sub> → C



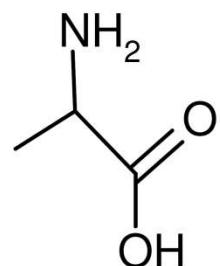
## SMILES – Regras(2)

Átomos e ligações:

- CC      as ligações simples não são representadas
- C=C     ligações duplas
- C#C     ligações triplas
- c:c     ligações entre carbonos aromáticos  
          (geralmente não se representam)
- C@C     qualquer tipo de ligação num anel
- C~C     qualquer tipo de ligação

## SMILES – Regras(3)

As ramificações denotam-se com parêntesis:

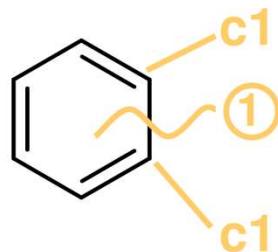


(determinar primeiro a sequência mais longa de ligações)

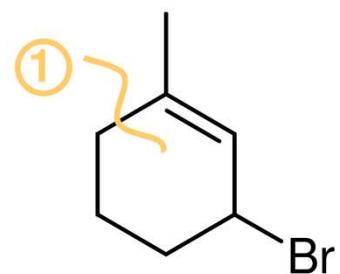
# SMILES – Regras(4)

Compostos cíclicos:

- Encontrar cadeia mais longa
- “abrir” o anel para obter uma cadeia
- numerar carbonos no pontos de abertura



c1ccccc1



CC1=CC(Br)CCC1

# SMILES – Regras(5)

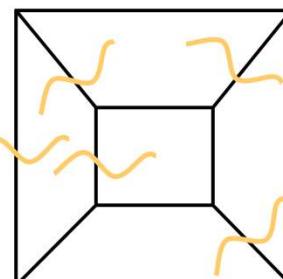
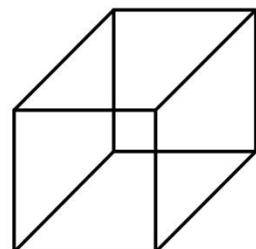
Compostos policíclicos:

- Múltiplos pontos de quebra



c1cc2cccc2cc1

Pode ocorrer fecho de mais do que um anel no mesmo átomo:



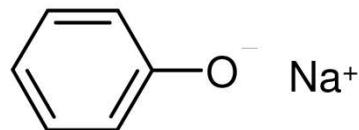
cubano

c12c3c4c1c5c4c3c25

Números maiores que 9 são antecedidos por um ‘%’ : %11

# SMILES – Regras(6)

Compostos ligados não-covalentemente são separados por um “.”



[Na+].[O-]c1ccccc1

Isótopos:

$^{13}\text{C}$  [13C]

$^{13}\text{CH}_4$  [13CH4]

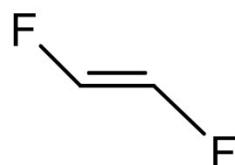
D<sub>2</sub>O [2H]O[2H]

# SMILES – Regras(7)

Configuração em torno de uma ligação dupla:



*cis*



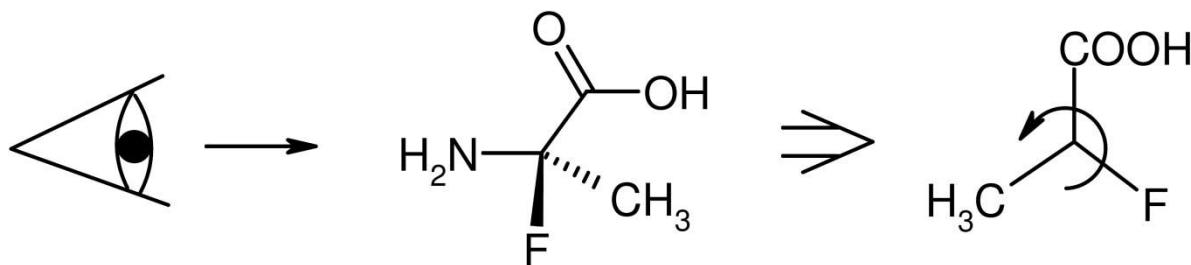
*trans*



Indeterminada

# SMILES – Regras(8)

Quiralidade:



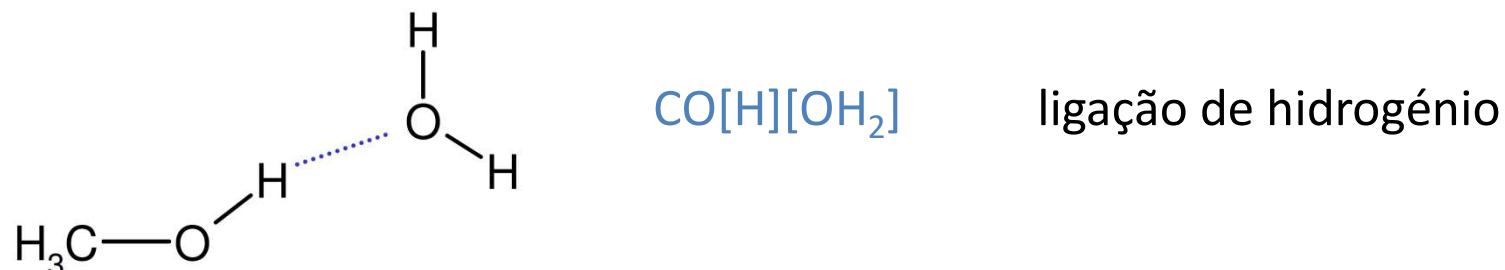
@ - sequência anti-horária de substituintes

@@ - sequência horária de substituintes

N.B. – Ausência de conformidade com o sistema (r,s) de representação absoluta da configuração

# SMILES – Regras(9)

Hidrogénios explícitos:



# SMILES – Regras(9)

As reacções químicas são representadas usando símbolo “>” :

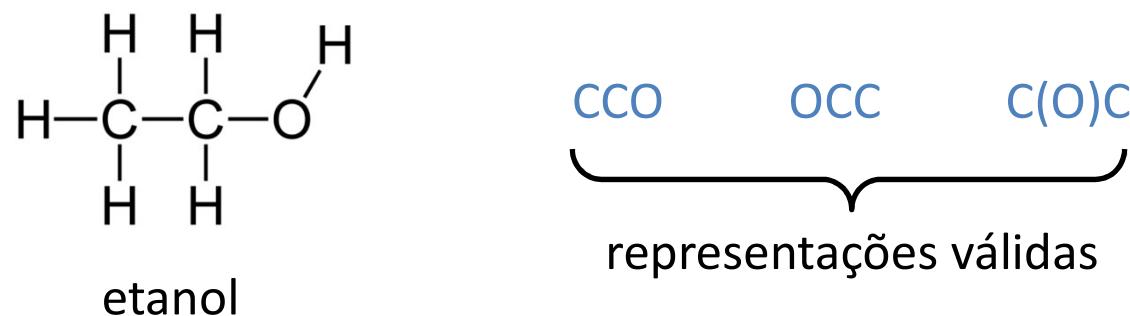
Reagentes > Agentes > Produtos

C=O>O=[O+]-[O-]>O=C=O.O    combustão de metano na presença de ozono

CC([O:1])[OH:2].CC[OH:3]>[H+]>CC([O:1])[O:3]CC.[OH2:2] esterificação ácida do ácido acético e etanol

# SMILES – Software

O problema da geração de SMILES a partir de estruturas não é trivial, pois geralmente existe mais do que uma representação SMILES válida para uma dada estrutura. Exemplo:



Para resolver este problema foram criados algoritmos de *canonização* que permitem gerar um SMILES único para cada molécula – SMILES canónico. Existem diversos packages de software que permitem gerar estes SMILES canónicos:

- Daylight Chemical Information Systems
- OpenEye Scientific Software
- Chemical Computing Group
- Chemistry Development Kit

# SMARTS (1)

SMARTS (SMILES Arbitrary Target Specification):

generalização de SMILES que permite a representação de padrões moleculares. Os padrões são representados dentro de “[]”

Exemplo:

[F,Cl,Br,I]            átomo que pode ser um F, Cl, Br ou I

Átomos:

- c            carbono aromático
- a            átomo aromático (C, N, O, S, ...)
- A            átomo alifático (não-aromático)
- \*            qualquer átomo (ou nenhum)
- [#16]        elemento nº 16 (qualquer tipo de enxofre)
- [rn]          átomo num anel de *n* membros
- [SX2]        enxofre com 2 substituintes    —S—    mas não     $\begin{array}{c} \parallel \\ \text{S} \end{array}$     ou    =S
- [Fe]          átomo de ferro (carga arbitrária)

## SMARTS (2)

Operadores lógicos:

A,B      A ou B

A&B      A e B

A;B      A e B

!A      não A

exemplos:

[F, Cl, Br, I]      F ou Cl ou Br ou I

[!C;R]      átomo aromático e não-alifático num anel

[CH2]      carbono alifático com 2 hidrogénios (metíleno)

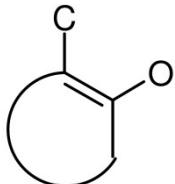
[c,n&H1]      carbono aromático ou NH aromático

[c,n;H1]  
hidrogénio      azoto ou carbono aromático e exactamente um hidrogénio

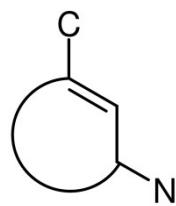
[#7;r5]      qualquer azoto num anel de 5 membros

## SMARTS (3)

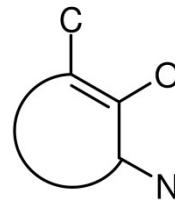
Configuração de substituintes:



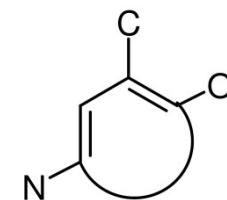
[CaaO]



[CaaaO]



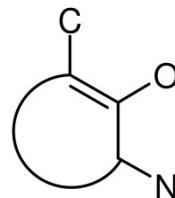
[Caa(O)aN]



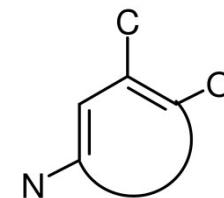
[Ca(aO)aaN]

O ambiente químico de um átomo pode ser especificado da seguinte forma:

C[\$(aaO);\$\$(aaaN)]



ou

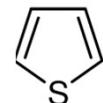


## SMARTS (3)

Configuração de substituintes:

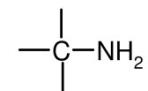
[s,o]1cccc1

tiofenos e furanos



[CX4][NH2]

aminas alifáticas primárias



[C1OC1]

epóxidos



C(=O)[OH,O-,O-.+]

ácido carbónico, carboxilato ou catião

C(=O)[NH1]

ligação peptídica

\*=[OH]

ácidos e enóis

F.F.F.F.F

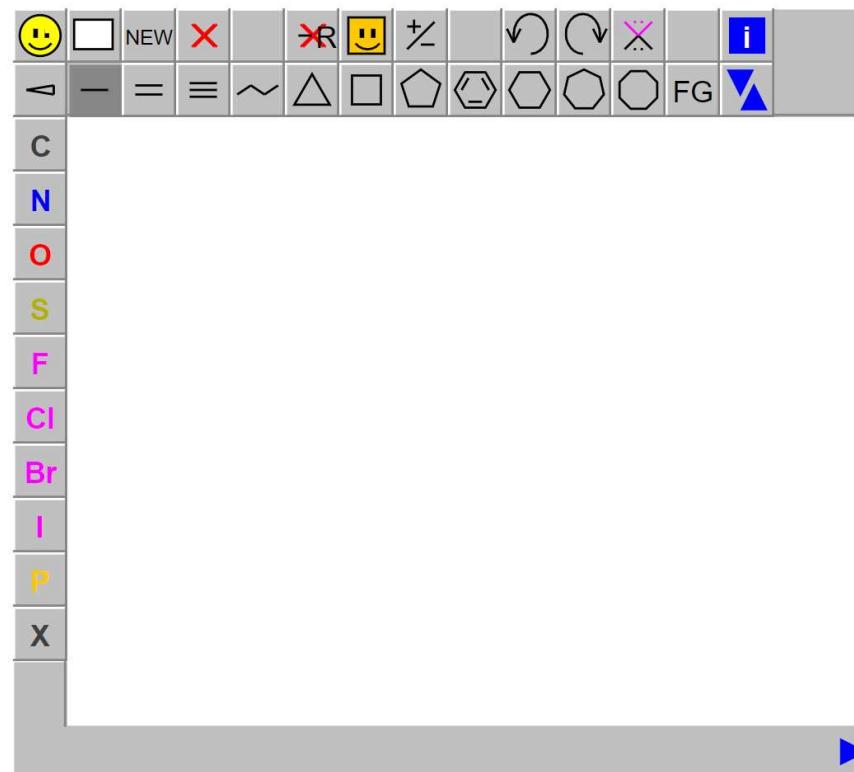
um total de 5 átomos de fluor as

# SMIRKS

- Especificação de reacções
- Superset de SMILES
- Subset de SMARTS
- Possui mecanismos que não existem nas outras duas linguagens

SMIRKS Depiction	Reaction	SMIRKS and Note
	Reacting carbon	<b>[C:1]&gt;&gt;[C:1]</b> Agents aren't allowed in SMIRKS. The format is " reactants >> products ".
	Reacting Carbon (2-Connected)	<b>[C;X2:1]&gt;&gt;[C;X2:1]</b> SMIRKS allows atomic SMARTS expressions. The syntax is: [<SMILES_PART>;<SMARTS_PART>:<MAP>]
	No Reaction	<b>[C;X4H3]-[CH2]C&gt;&gt;[C][CH2]C</b> SMARTS atom specifications may be used for mapped atoms only (i.e. unmapped atoms must be valid SMILES expressions ).
<b>[NO REACTION]</b>	No Reaction	<b>[C;X2:1]~C&gt;&gt;[C;X2:1]=C</b> SMIRKS doesn't allow SMARTS Bond Queries (e.g. ~). Bonds expressions must be valid SMILES.
	Just add water	<b>&gt;&gt;O</b> Upon transformation, all unmapped product-side SMILES get created.

# JSME Molecular Edito & SMILES reader/generator



- Leitura de SMILES, SMARTS, SMIRKS, MOL, SDF
- Geração de SMILES canónicos
- Geração de InCh e InChKey
- Pesquisa de moléculas on-line através da InChKey

[http://peter-ertl.com/jsme/JSME\\_2017-02-26/JSME.html](http://peter-ertl.com/jsme/JSME_2017-02-26/JSME.html)

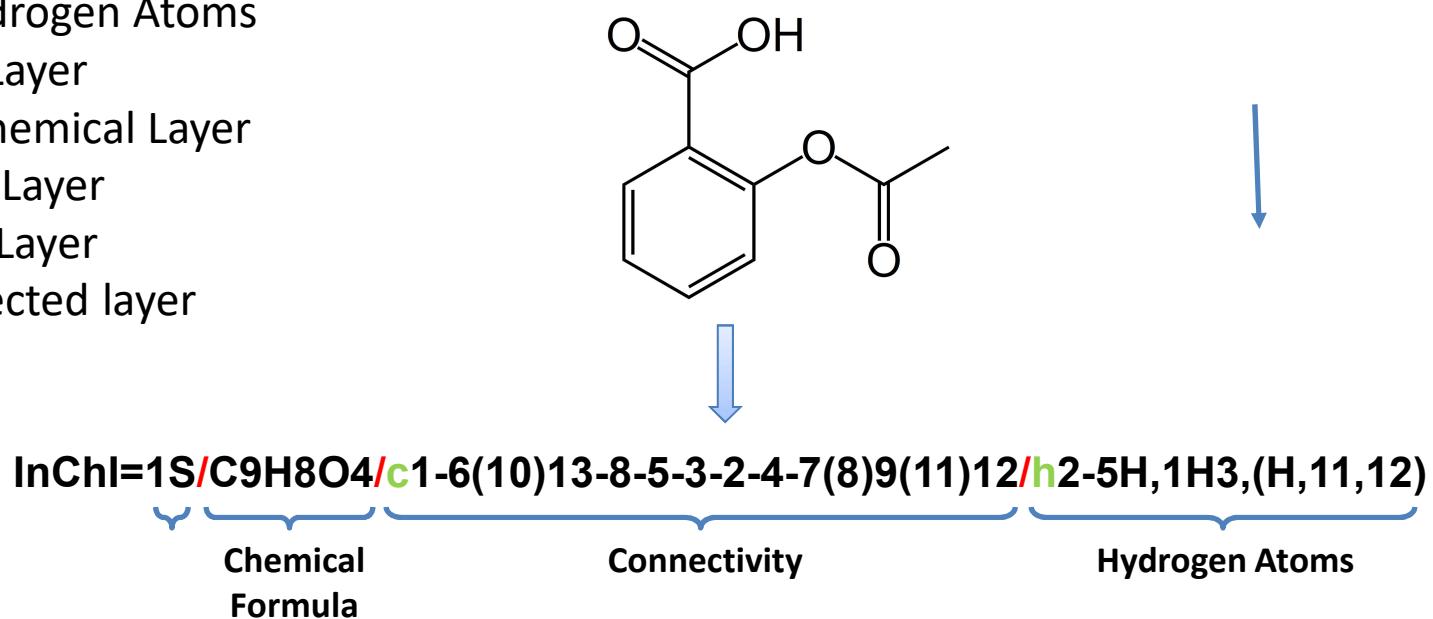
# InChI Representation

## InChI – IUPAC International Chemical Identifier

InChI is a text-based identifier for chemical substances, designed to provide a standard way to provide molecular information

The InChI Identifier describes molecules in terms of different layers of information:

- Main Layer
  - Chemical Formula
  - Atom connections
  - Hydrogen Atoms
- Charge Layer
- Stereochemical Layer
- Isotopic Layer
- Fixed-H Layer
- Reconnected layer

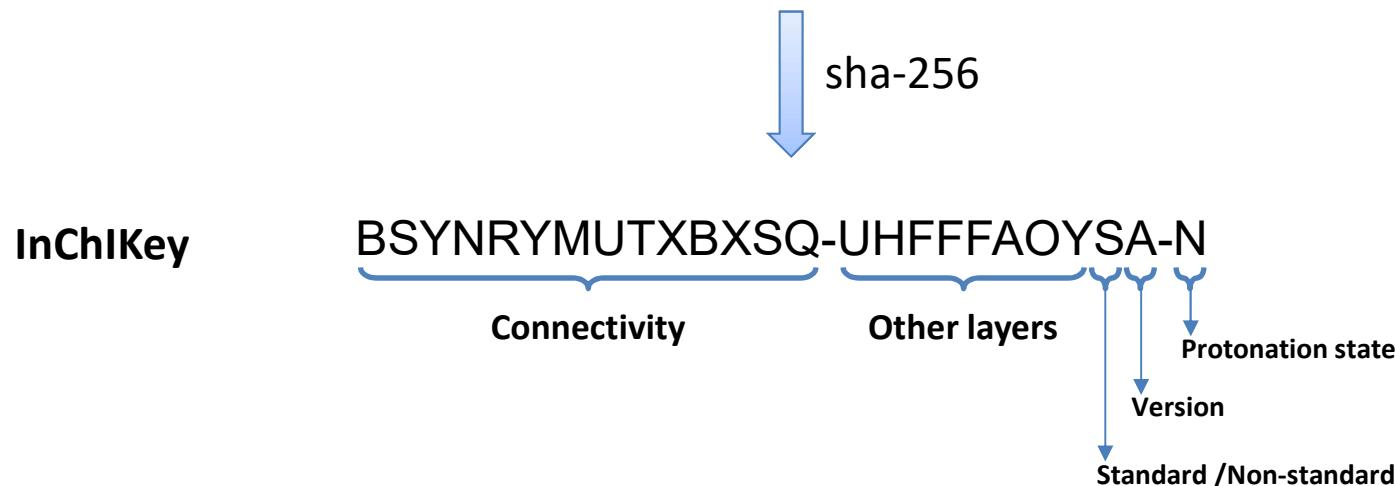


# InChI and InChIKey

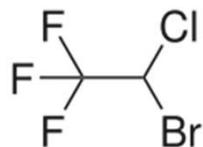
**InChIKey** – Compressed form of the InChI, using a hashing algorithm (sha-256) to produce a quasi-unique string with shorter length.

The InChIKeys are short enough to be used in internet searches (Google, etc)

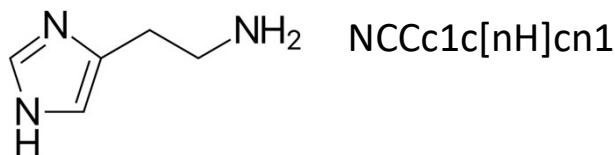
**InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)**



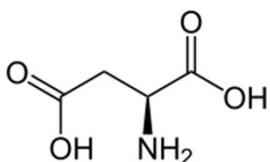
# Exercícios



C(C(F)(F)F)(Cl)Br

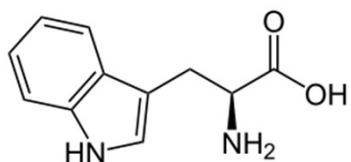


NCCc1c[nH]cn1

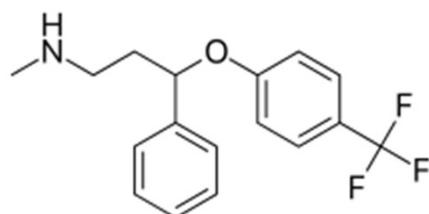


O=C(O)CC(N)C(=O)O

O=C([O-])CC([NH3+])C(=O)[O-]



c1ccc2c(c1)c(c[nH]2)C[C@@H](C(=O)O)N



CNCCC(c1ccccc1)Oc2ccc(cc2)C(F)(F)F