Ferramentas e bases de dados online para desenho de fármacos

- Portais: locais de acesso a recursos 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

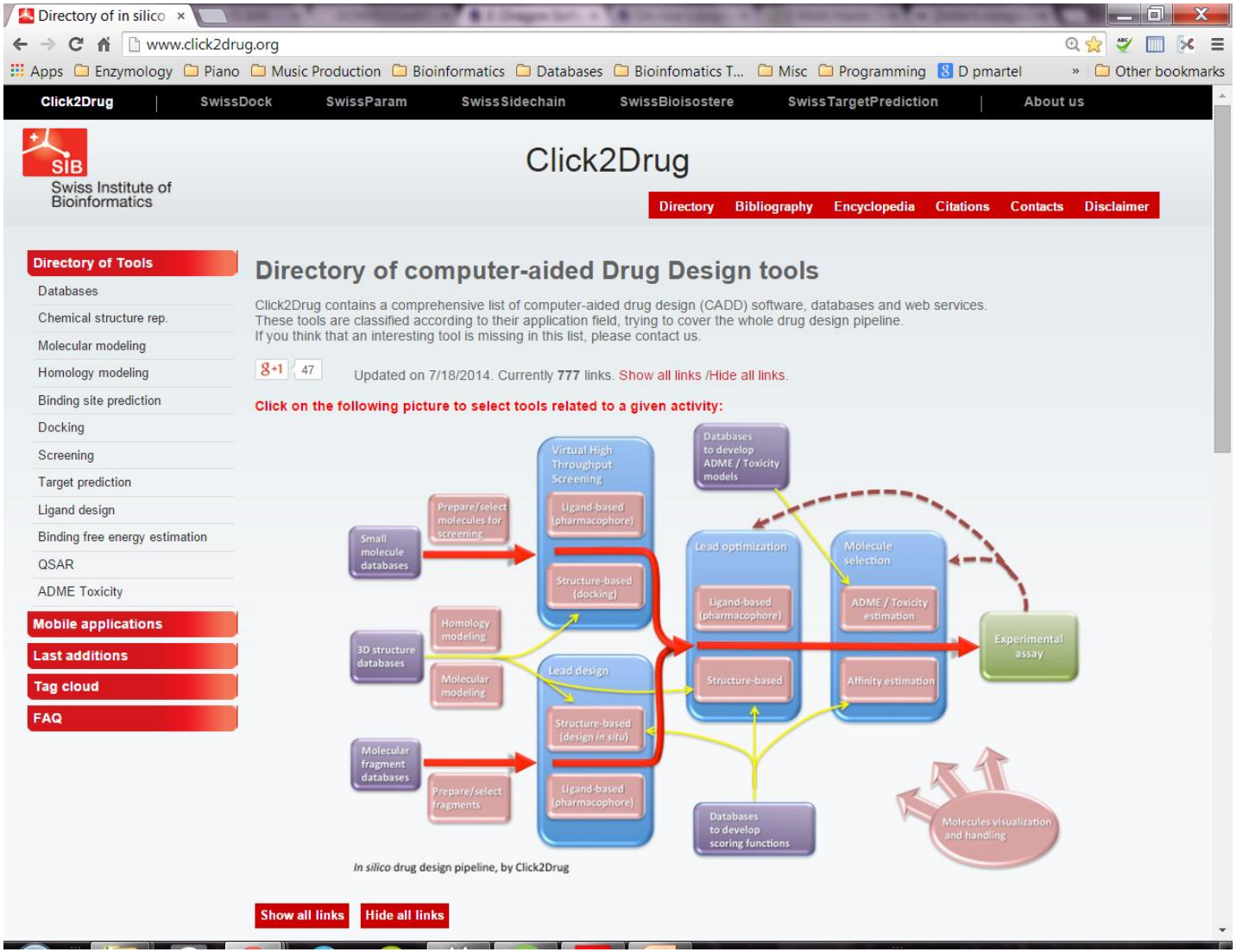
- Disponíveis em qualquer local
- Custos de manutenção reduzidos
- Custos de licenciamento reduzidos
- Integração de diferentes tipos de software
- Fácil monitorização da utilização
- Computação em *cloud*
- Compatibilidade com múltiplos platformas informáticas (Win, Mac, Linux, Android, etc)

Tetko,I.V.(2008) *The Open Applied Informatics J.* **2**:18-21

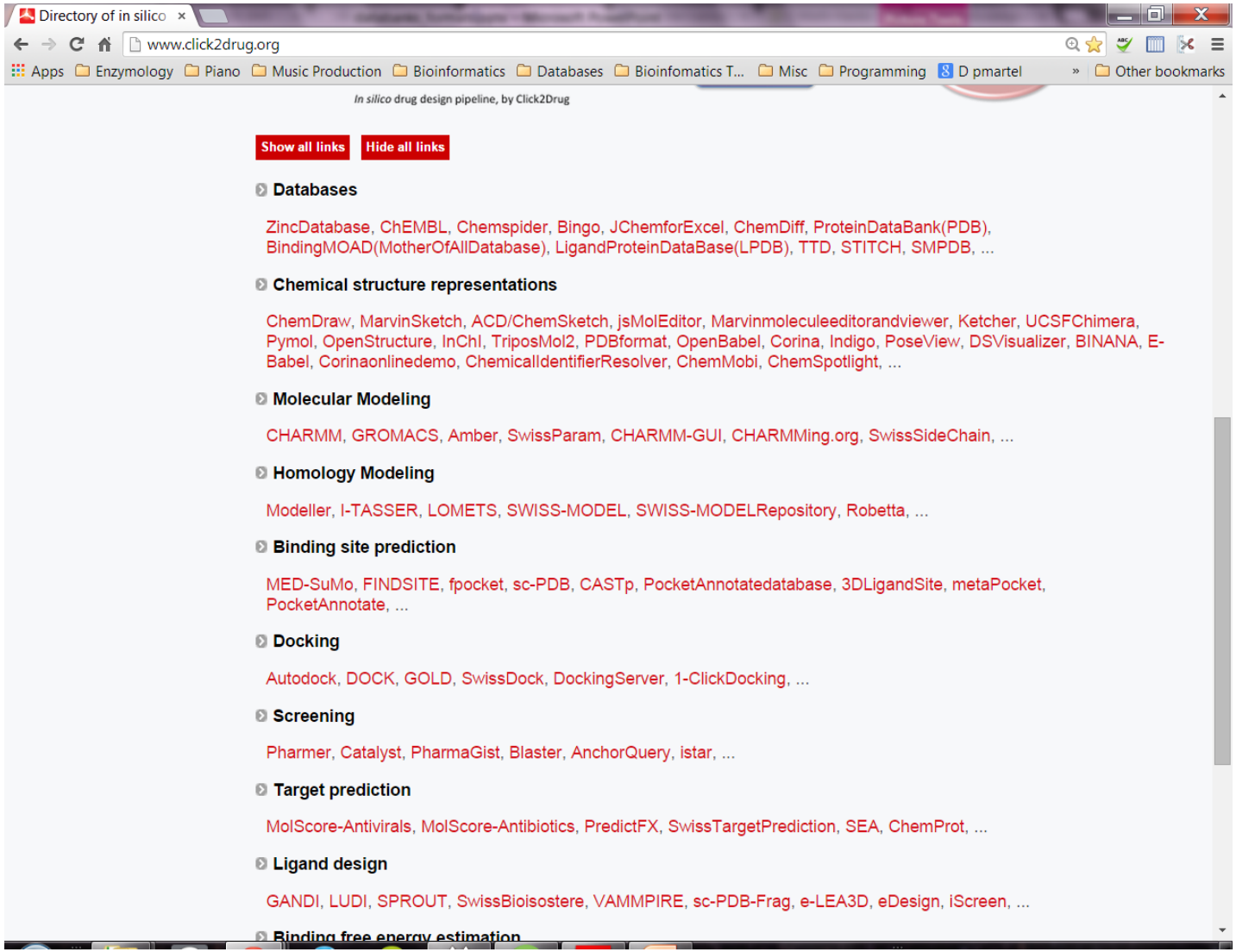
O portal Click2Drug

- 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

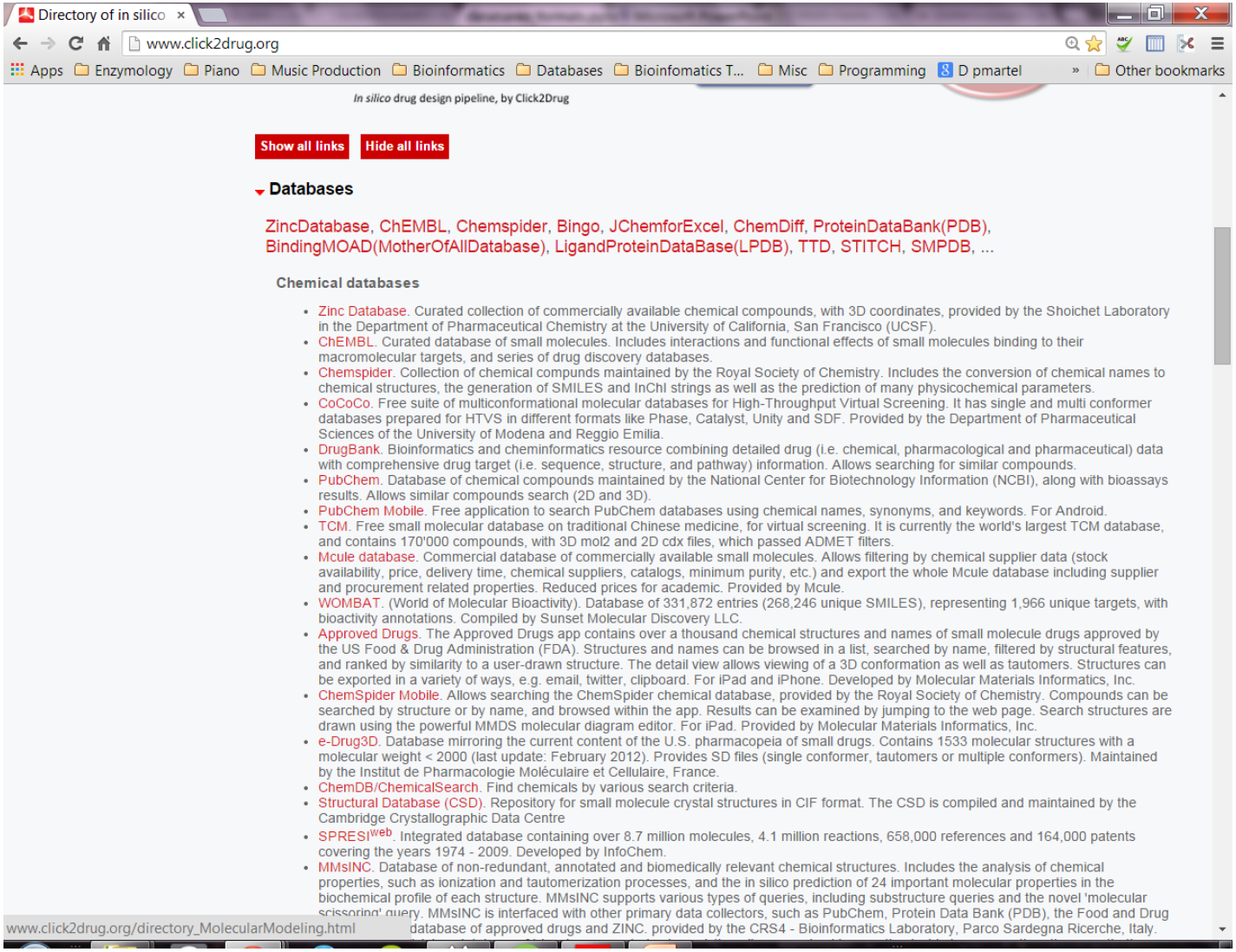
O portal Click2Drug



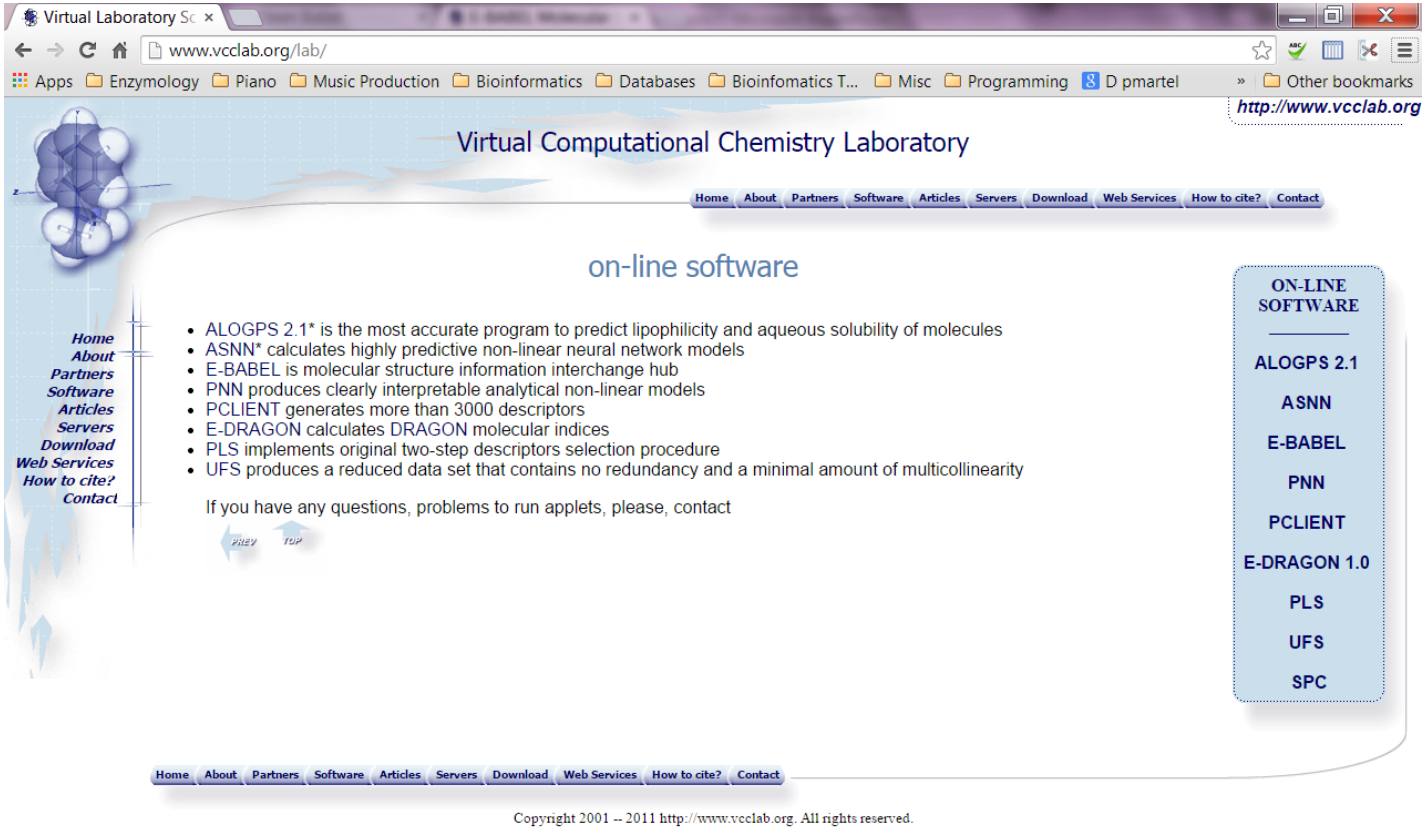
Portal Click2Drug



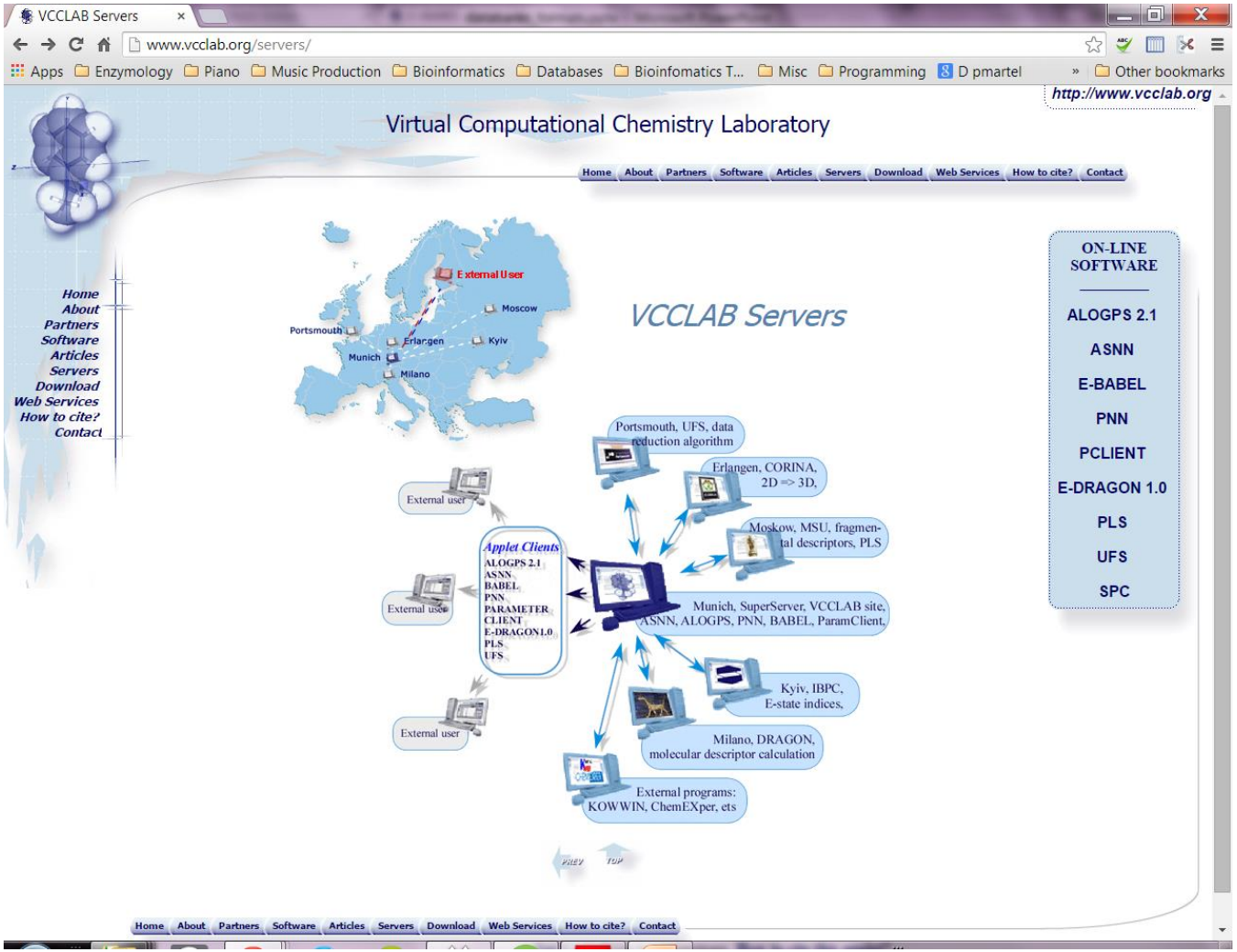
Portal Click2Drug



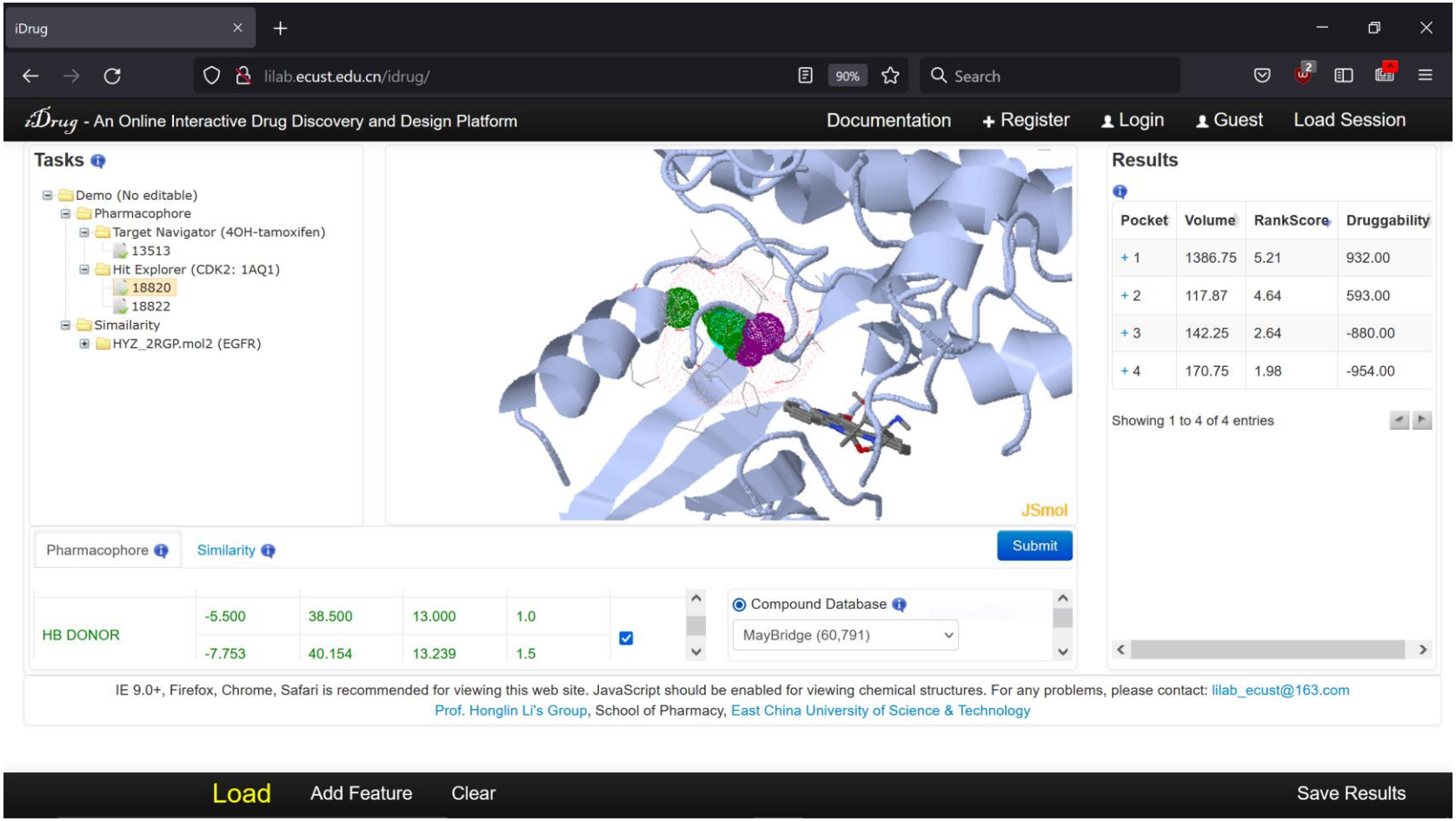
Virtual Computational Chemistry Laboratory - VCCLAB



Virtual Computational Chemistry Laboratory - VCCLAB



iDrug: on-line Drug Design Workbench



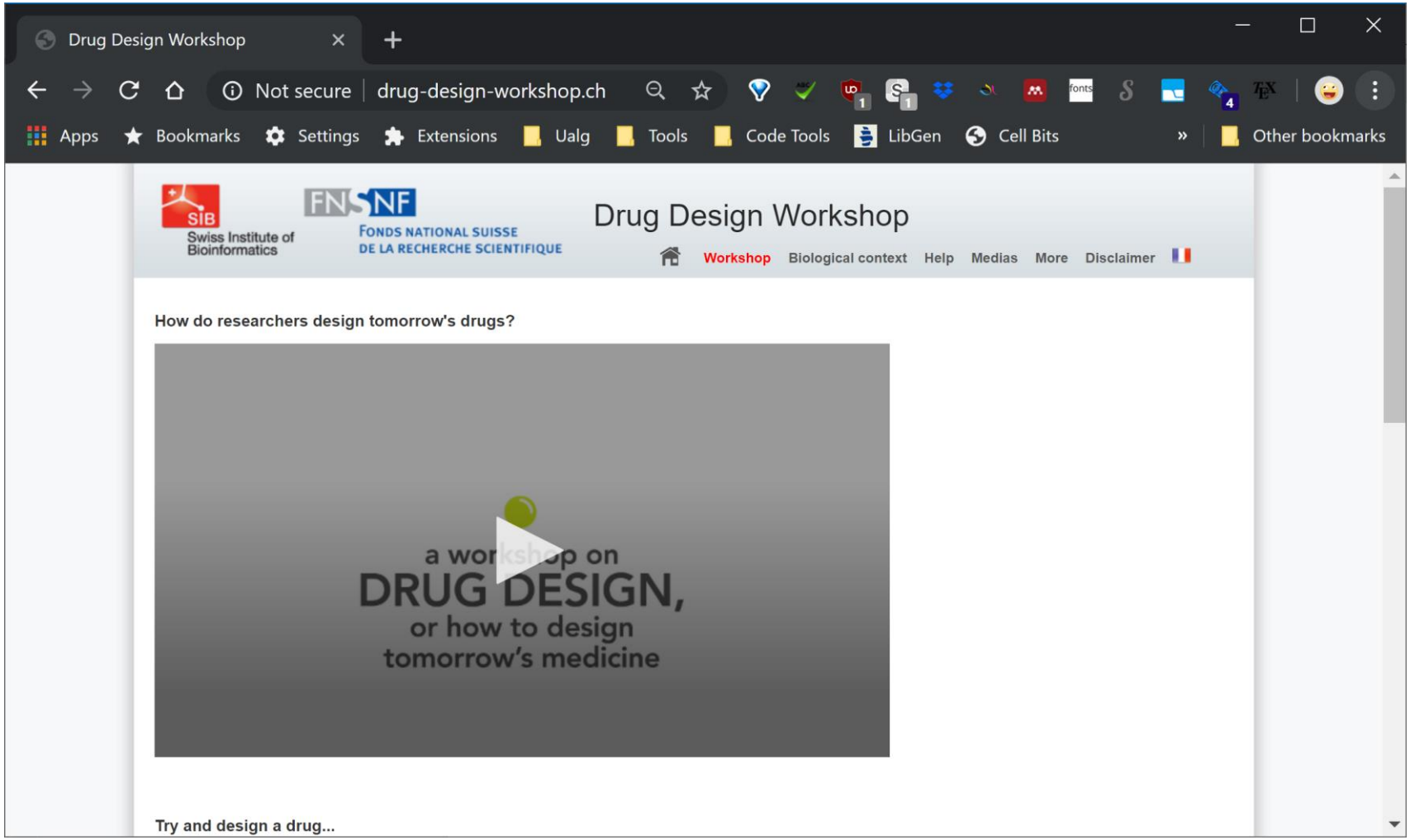
*i*Drug: a web-accessible and interactive drug discovery and design platform

Xia Wang[1], Haipeng Chen[2], Feng Yang[2], Jiayu Gong[2], Shiliang Li[1], Jianfeng Pei[3*], Xiaofeng Liu[1*], Hualiang Jiang[1], Luhua Lai[3] and Honglin Li[1,2*]

Wang *et al. Journal of Cheminformatics* 2014, **6**:28

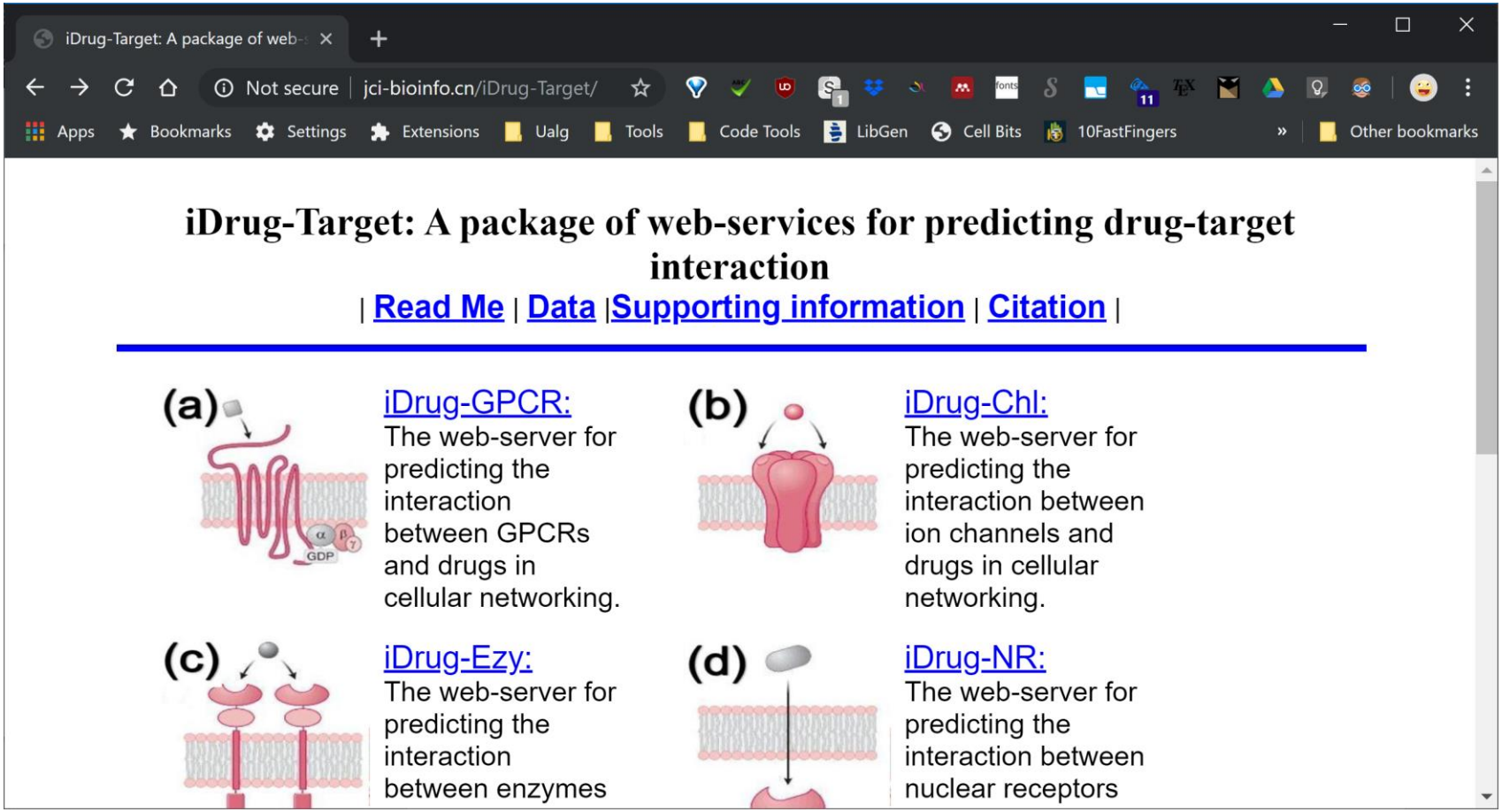
http://lilab.ecust.edu.cn/idrug/

Drug Design Workshop



http://www.drug-design-workshop.ch/

iDrug-Target



http://www.jci-bioinfo.cn/iDrug-Target/

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, ….)

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

Formatos de representação

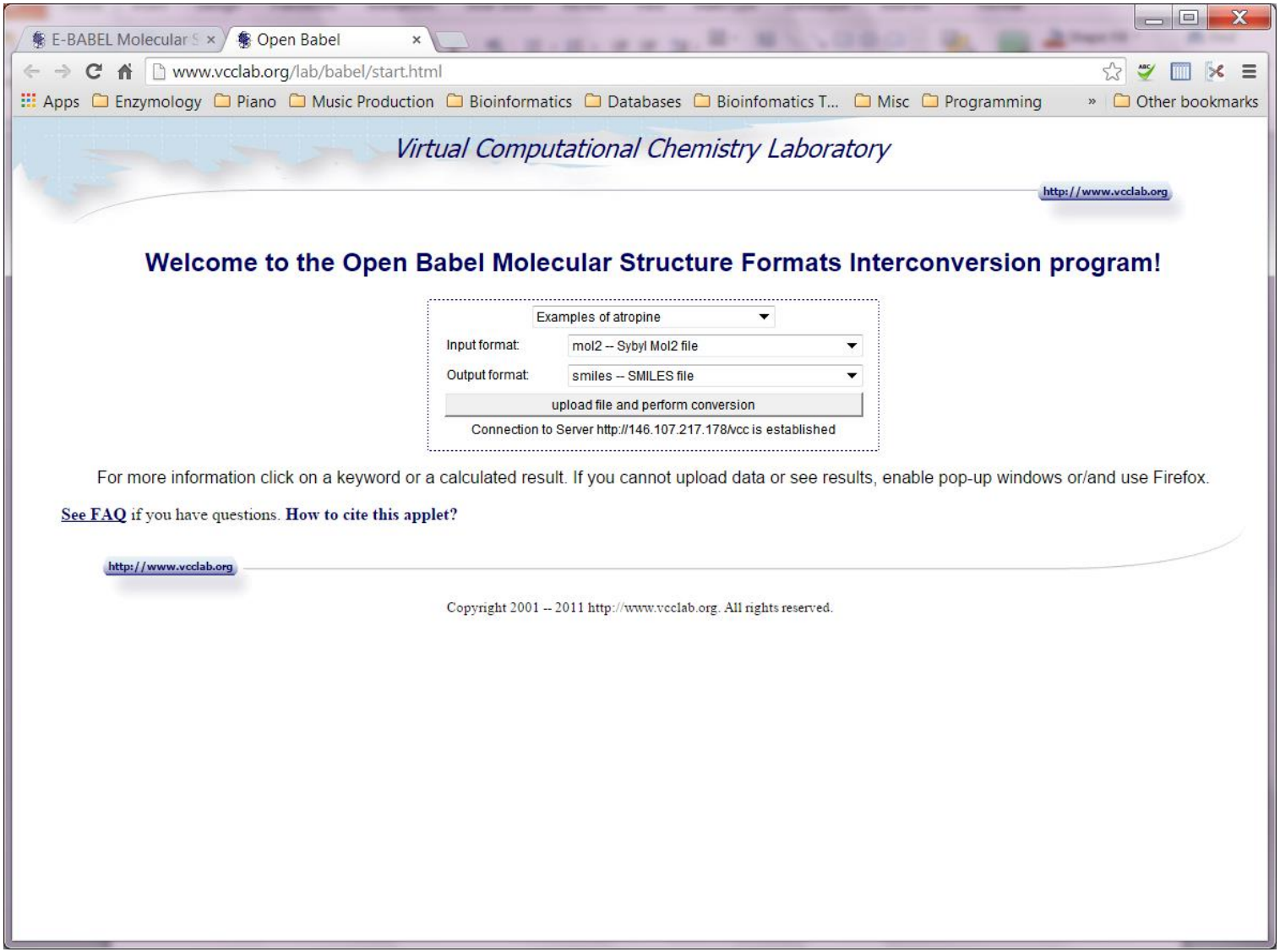
- Estrutura:
  - PDB, MDL, SDF, MOL2, CIF, ASN.1, HIN,Trypos, Sybil, Gaussian, XYZ, CML, XML, SMILES ….

- Sequência:
  - Fasta, SWISSPROT, ASN.1, GCG, GenBank, PIR, Phylip,….

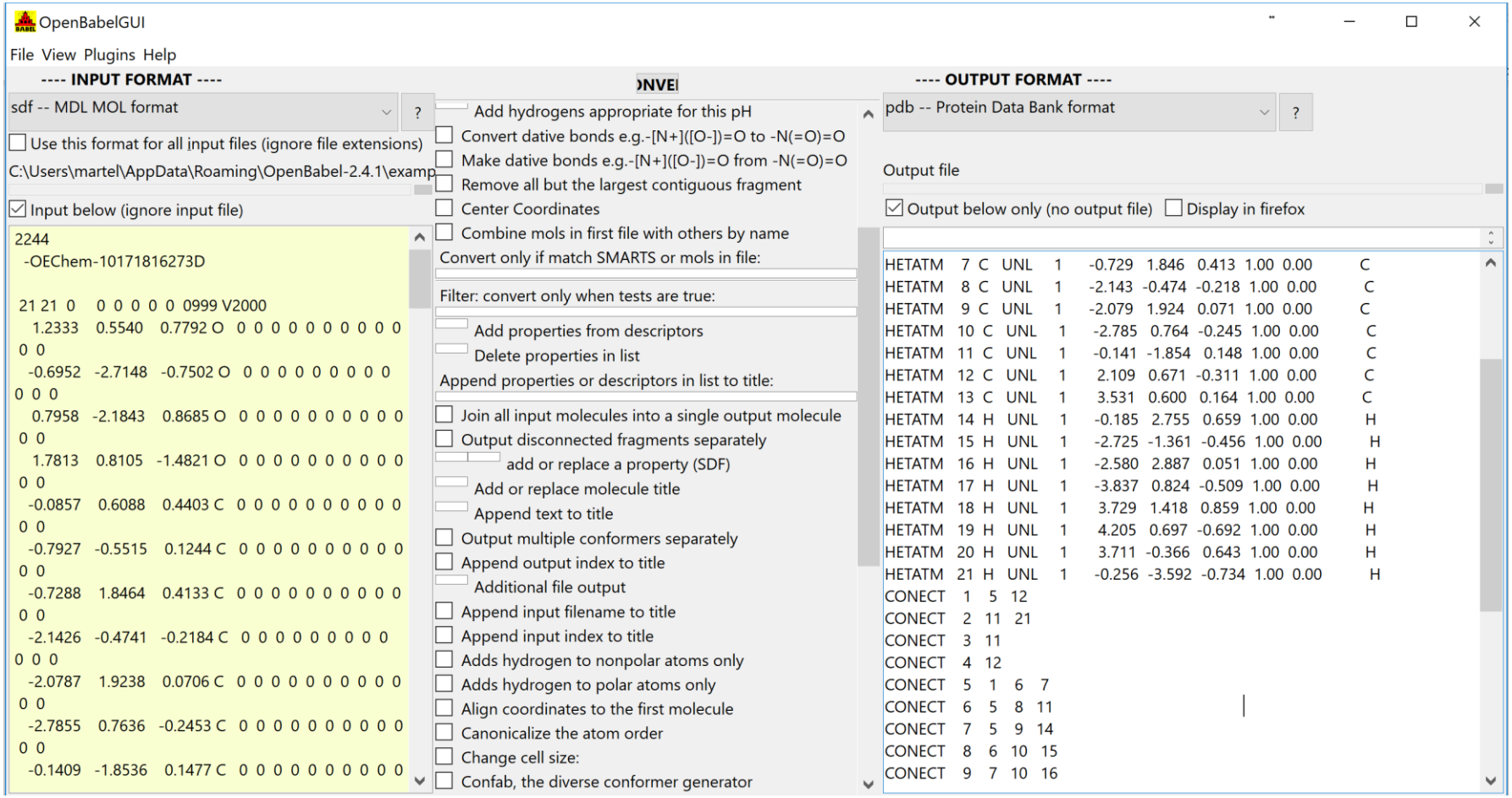
Ferramenta de conversão entre formatos:

OpenBabel (http://openbabel.org)

E-Babel: conversão de formatos online



OpenBabel


OpenBabelGUI
File View Plugins Help
---- INPUT FORMAT ----
sdf -- MDL MOL format
Use this format for all input files (ignore file extensions)
C:\Users\martel\AppData\Roaming\OpenBabel-2.4.1\examp
Input below (ignore input file)
2244
-OEChem-10171816273D
21 21 0 0 0 0 0 0 0999 V2000
1.2333 0.5540 0.7792 O 0 0 0 0 0 0 0 0 0 0 0 0
-0.6952 -2.7148 -0.7502 O 0 0 0 0 0 0 0 0 0 0 0 0
0.7958 -2.1843 0.8685 O 0 0 0 0 0 0 0 0 0 0 0 0
1.7813 0.8105 -1.4821 O 0 0 0 0 0 0 0 0 0 0 0 0
-0.0857 0.6088 0.4403 C 0 0 0 0 0 0 0 0 0 0 0 0
-0.7927 -0.5515 0.1244 C 0 0 0 0 0 0 0 0 0 0 0 0
-0.7288 1.8464 0.4133 C 0 0 0 0 0 0 0 0 0 0 0 0
-2.1426 -0.4741 -0.2184 C 0 0 0 0 0 0 0 0 0 0 0 0
-2.0787 1.9238 0.0706 C 0 0 0 0 0 0 0 0 0 0 0 0
-2.7855 0.7636 -0.2453 C 0 0 0 0 0 0 0 0 0 0 0 0
-0.1409 -1.8536 0.1477 C 0 0 0 0 0 0 0 0 0 0 0 0
ONVE
Add hydrogens appropriate for this pH
Convert dative bonds e.g.-[N+]([O-])=O to -N(=O)=O
Make dative bonds e.g.-[N+]([O-])=O from -N(=O)=O
Remove all but the largest contiguous fragment
Center Coordinates
Combine mols in first file with others by name
Convert only if match SMARTS or mols in file:
Filter: convert only when tests are true:
Add properties from descriptors
Delete properties in list
Append properties or descriptors in list to title:
Join all input molecules into a single output molecule
Output disconnected fragments separately
add or replace a property (SDF)
Add or replace molecule title
Append text to title
Output multiple conformers separately
Append output index to title
Additional file output
Append input filename to title
Append input index to title
Adds hydrogen to nonpolar atoms only
Adds hydrogen to polar atoms only
Align coordinates to the first molecule
Canonicalize the atom order
Change cell size:
Confab, the diverse conformer generator
---- OUTPUT FORMAT ----
pdb -- Protein Data Bank format
Output file
Output below only (no output file)
Display in firefox
HETATM 7 C UNL 1 -0.729 1.846 0.413 1.00 0.00 C
HETATM 8 C UNL 1 -2.143 -0.474 -0.218 1.00 0.00 C
HETATM 9 C UNL 1 -2.079 1.924 0.071 1.00 0.00 C
HETATM 10 C UNL 1 -2.785 0.764 -0.245 1.00 0.00 C
HETATM 11 C UNL 1 -0.141 -1.854 0.148 1.00 0.00 C
HETATM 12 C UNL 1 2.109 0.671 -0.311 1.00 0.00 C
HETATM 13 C UNL 1 3.531 0.600 0.164 1.00 0.00 C
HETATM 14 H UNL 1 -0.185 2.755 0.659 1.00 0.00 H
HETATM 15 H UNL 1 -2.725 -1.361 -0.456 1.00 0.00 H
HETATM 16 H UNL 1 -2.580 2.887 0.051 1.00 0.00 H
HETATM 17 H UNL 1 -3.837 0.824 -0.509 1.00 0.00 H
HETATM 18 H UNL 1 3.729 1.418 0.859 1.00 0.00 H
HETATM 19 H UNL 1 4.205 0.697 -0.692 1.00 0.00 H
HETATM 20 H UNL 1 3.711 -0.366 0.643 1.00 0.00 H
HETATM 21 H UNL 1 -0.256 -3.592 -0.734 1.00 0.00 H
CONECT 1 5 12
CONECT 2 11 21
CONECT 3 11
CONECT 4 12
CONECT 5 1 6 7
CONECT 6 5 8 11
CONECT 7 5 9 14
CONECT 8 6 10 15
CONECT 9 7 10 16

Formato FASTA

- É um formato de representação de sequências biológicas (DNA ou proteína)

- Consiste numa linha de cabeçalho, seguida de linhas contendo a sequência de aminoácidos ou nucleótidos representada em códigos de 1 letra
- Contem muito pouca informação para além da sequência

Formato FASTA

Cabeçalho

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

AACAATCATGGCATCTGTGAAGTCGTCGTCGTCGTCATCATCATCATCATTTATTTCCTTGTT
GTTGTTGATTTTGCTTGTGATTGTACTGCAAAGCCAAGTTATCGAGTGTCAACCTCAACAGT
CATGCACCGCGTCACTTACTGGCCTGAACGTCTGCGCCCCATTCCTGGTCCCAGGCTCACCTAC
TGCAAGTACGGAGTGTTGCAA TGCAGTACAGTCGATTAATCATGACTGTATGTGCAACACT
ATGCGCATTGCAGCTCAAATTCCAGCTCAG TGCAACCTCCCTCCACTCTCTTGTTCTGCAAAT
TGAGTTTGAGATCAGTGGCCAGCAAGTTTACATCTGC TACATGAGCAAATTAAATAATATC
GTAACAATAAATTAAAGTTGTCTTTTTTTTTTTTTTGGTTATGCAAC AGACCAAGGGGGTCA
TGAGAAAAGAGTTTGTACTATCATATGATTATCAATAAAAAAAATTATGAG

>Q43495|108_SOLLC Protein 108 precursor - Solanum lycopersicum
MASVKSSSSSSSSFISLLLLILLVIVLQSQVIECQPQQSCTASLTGLNVCAPFLVPGSP
TASTECCNAVQSINHDCMCNTMRIAAQIPAQCNLPPLSCSAN

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.,
```

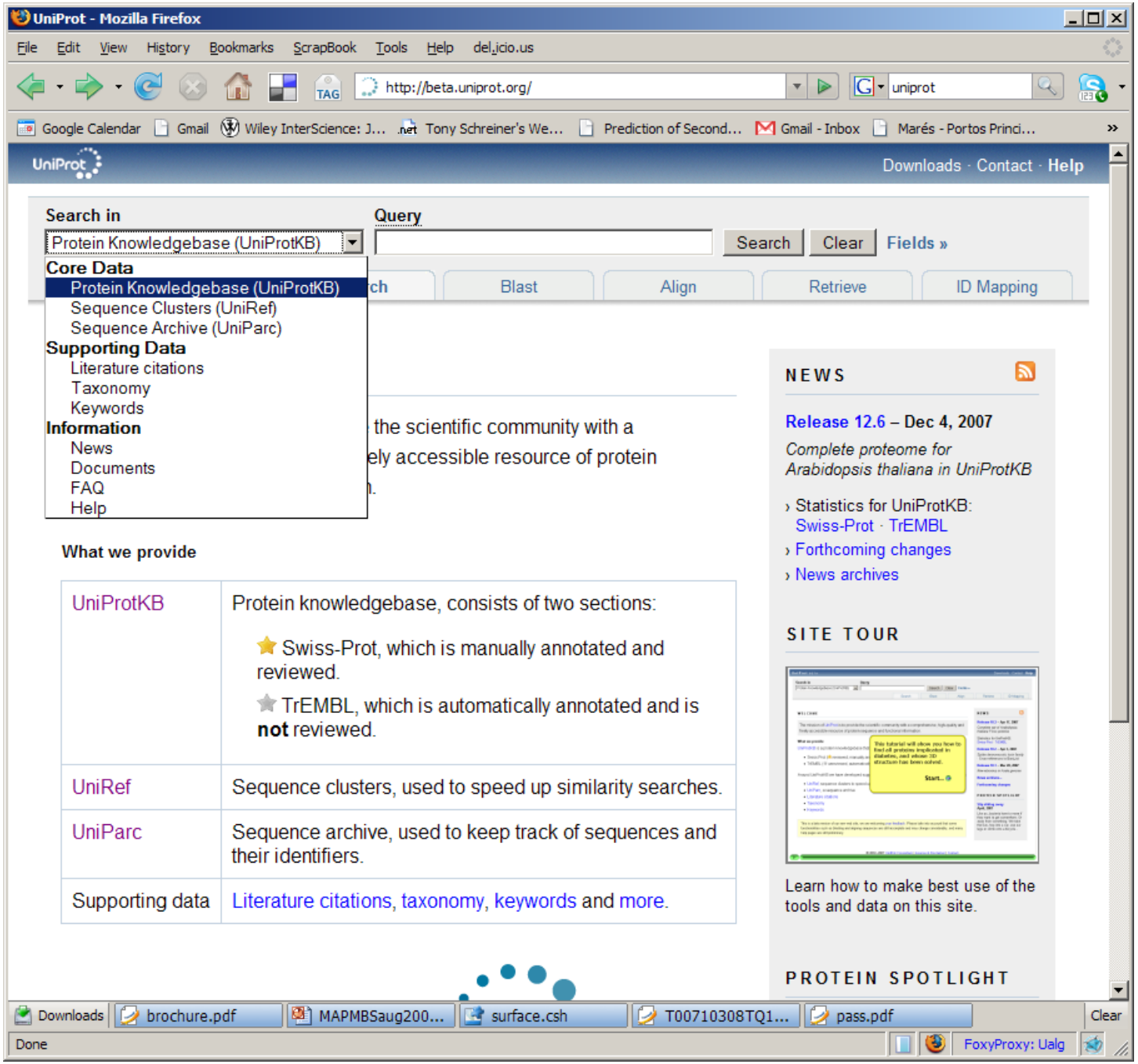
. . . . .

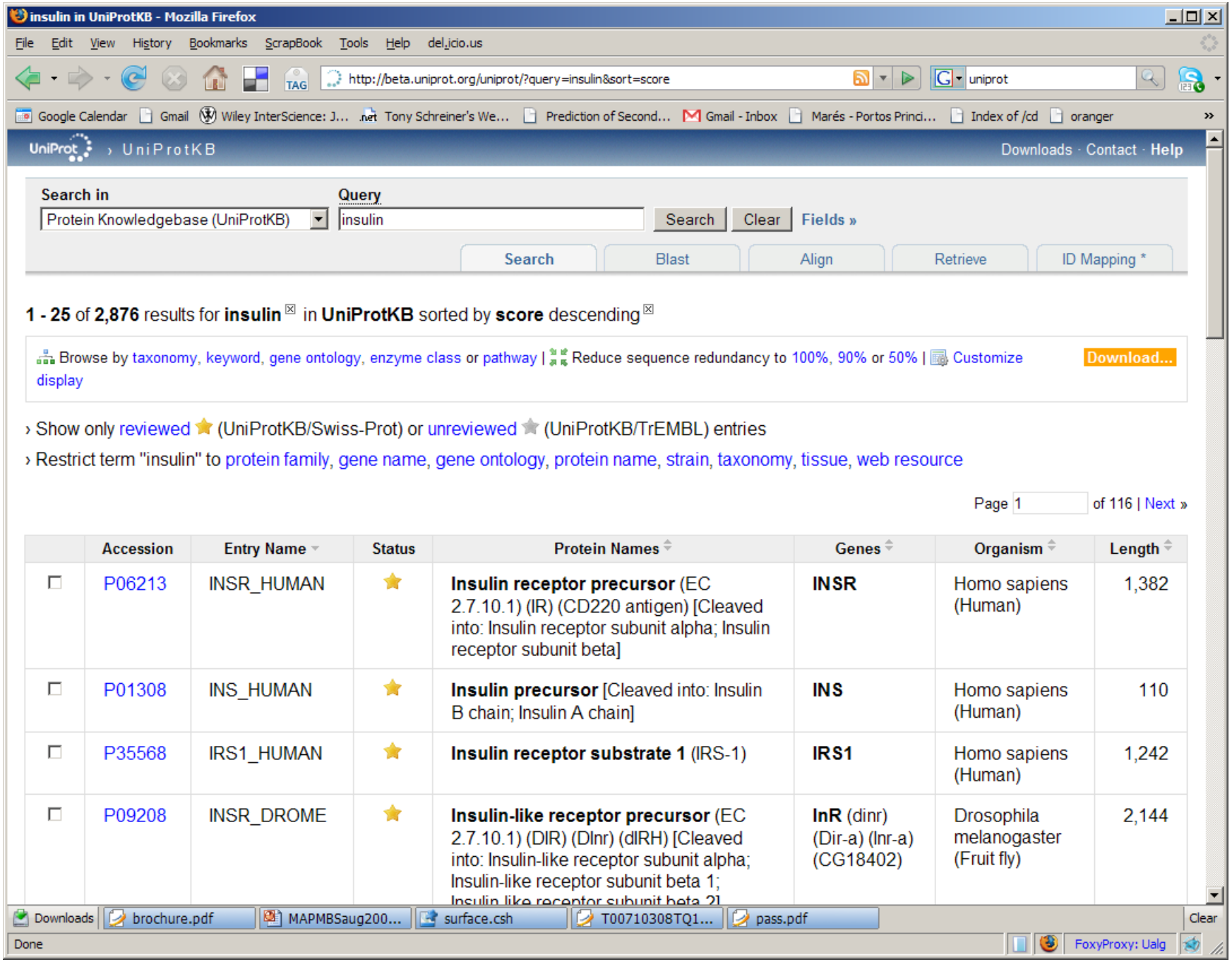
```
T   STRAND     183    187
FT   STRAND     192    194
FT   STRAND     203    206
FT   STRAND     209    216
FT   STRAND     218    222
FT   STRAND     227    231
FT   HELIX      232    235
FT   HELIX      236    245
SQ   SEQUENCE   247 AA;  26558 MW;  DD49A487B8062813 CRC64;
     MNPLLILTFV AAALAAPFDD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

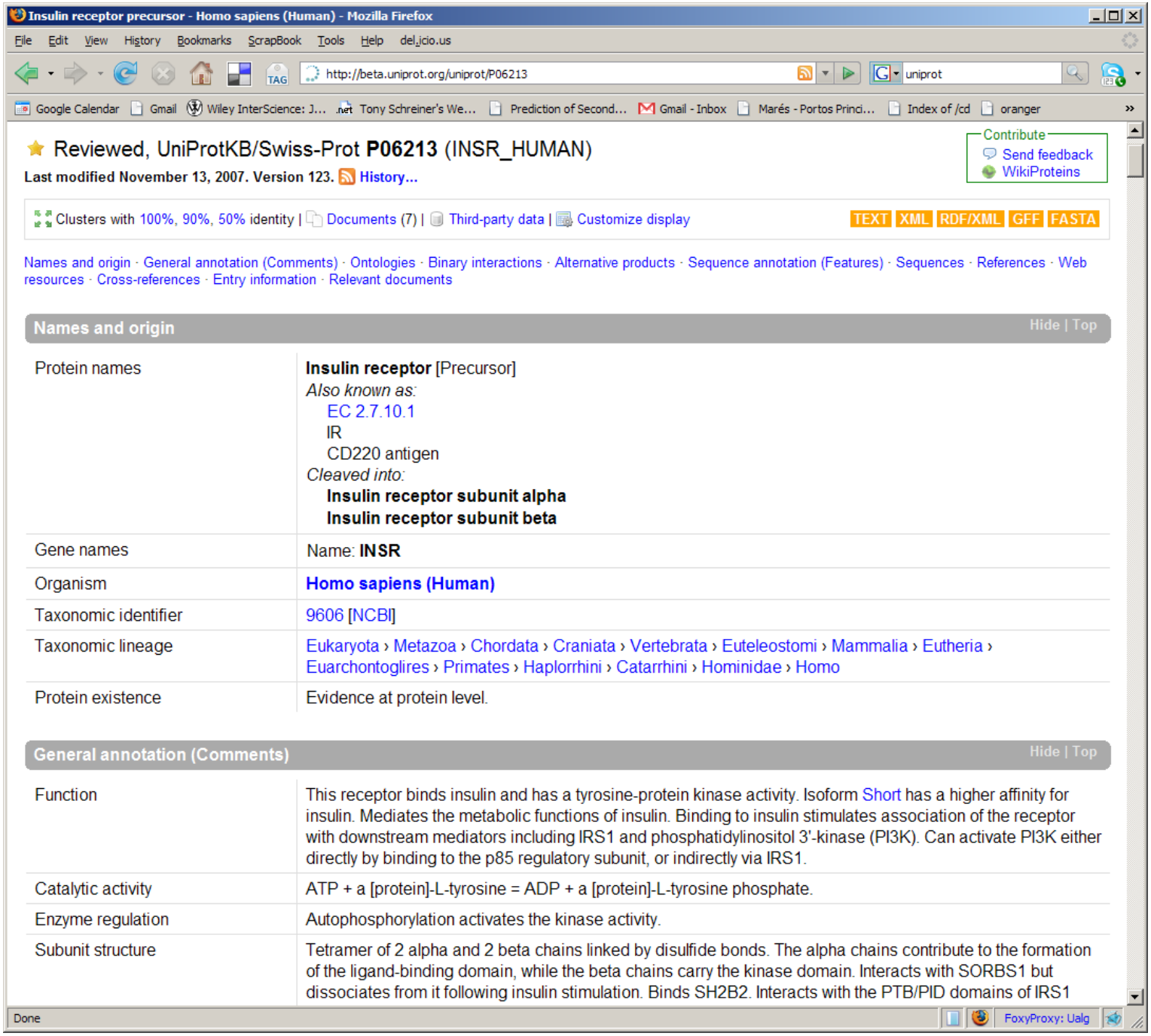
UniProt, a referência universal para sequências de proteínas

- 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
http://beta.uniprot.org/
uniprot
Google Calendar Gmail Wiley InterScience: J... Tony Schreiner's We... Prediction of Second... Gmail - Inbox Marés - Portos Princi...
UniProt
Downloads · Contact · Help
Search in
Protein Knowledgebase (UniProtKB)
Query
Search
Clear
Fields »
Core Data
Protein Knowledgebase (UniProtKB)
Sequence Clusters (UniRef)
Sequence Archive (UniParc)
Supporting Data
Literature citations
Taxonomy
Keywords
Information
News
Documents
FAQ
Help
Blast
Align
Retrieve
ID Mapping
the scientific community with a
ely accessible resource of protein
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 not reviewed.
UniRef
Sequence clusters, used to speed up similarity searches.
UniParc
Sequence archive, used to keep track of sequences and their identifiers.
Supporting data
Literature citations, taxonomy, keywords and more.
NEWS
Release 12.6 – Dec 4, 2007
Complete proteome for Arabidopsis thaliana in UniProtKB
› Statistics for UniProtKB: Swiss-Prot · TrEMBL
› Forthcoming changes
› News archives
SITE TOUR
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

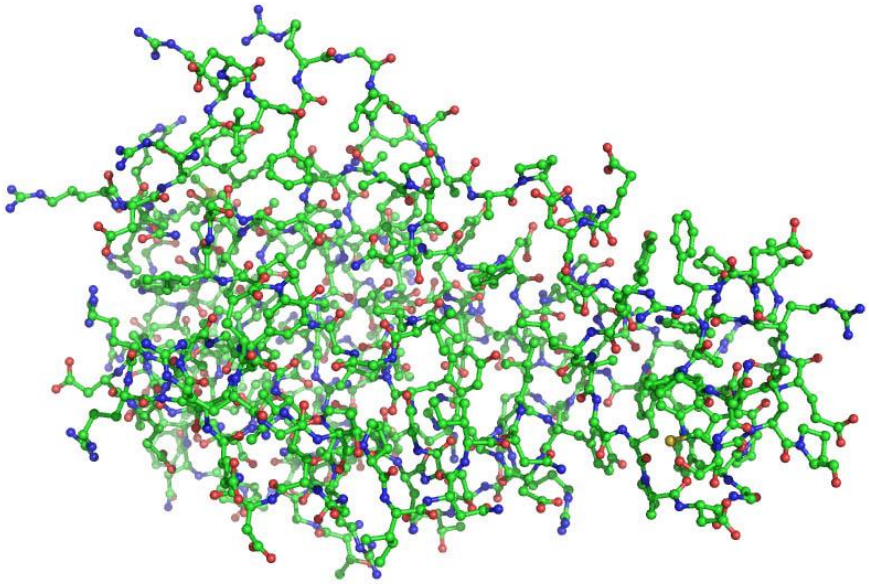

insulin in UniProtKB - Mozilla Firefox
File Edit View History Bookmarks ScrapBook Tools Help del.icio.us
http://beta.uniprot.org/uniprot/?query=insulin&sort=score
uniprot
Google Calendar Gmail Wiley InterScience: J... Tony Schreiner's We... Prediction of Second... Gmail - Inbox Marés - Portos Princi... Index of /cd oranger
UniProt › UniProtKB
Downloads · Contact · Help
Search in
Protein Knowledgebase (UniProtKB)
Query
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
P06213 INSR_HUMAN Insulin receptor precursor (EC 2.7.10.1) (IR) (CD220 antigen) [Cleaved into: Insulin receptor subunit alpha; Insulin receptor subunit beta] INSR Homo sapiens (Human) 1,382
P01308 INS_HUMAN Insulin precursor [Cleaved into: Insulin B chain; Insulin A chain] INS Homo sapiens (Human) 110
P35568 IRS1_HUMAN Insulin receptor substrate 1 (IRS-1) IRS1 Homo sapiens (Human) 1,242
P09208 INSR_DROME Insulin-like receptor precursor (EC 2.7.10.1) (DIR) (Dlnr) (dIRH) [Cleaved into: Insulin-like receptor subunit alpha; Insulin-like receptor subunit beta 1; InR (dinr) (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
http://beta.uniprot.org/uniprot/P06213
uniprot
Google Calendar Gmail Wiley InterScience: J... Tony Schreiner's We... Prediction of Second... Gmail - Inbox Marés - Portos Princi... Index of /cd oranger
Reviewed, UniProtKB/Swiss-Prot P06213 (INSR_HUMAN)
Last modified November 13, 2007. Version 123. History...
Contribute
Send feedback
WikiProteins
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
Insulin receptor [Precursor]
Also known as:
EC 2.7.10.1
IR
CD220 antigen
Cleaved into:
Insulin receptor subunit alpha
Insulin receptor subunit beta
Gene names
Name: INSR
Organism
Homo sapiens (Human)
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 Short 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

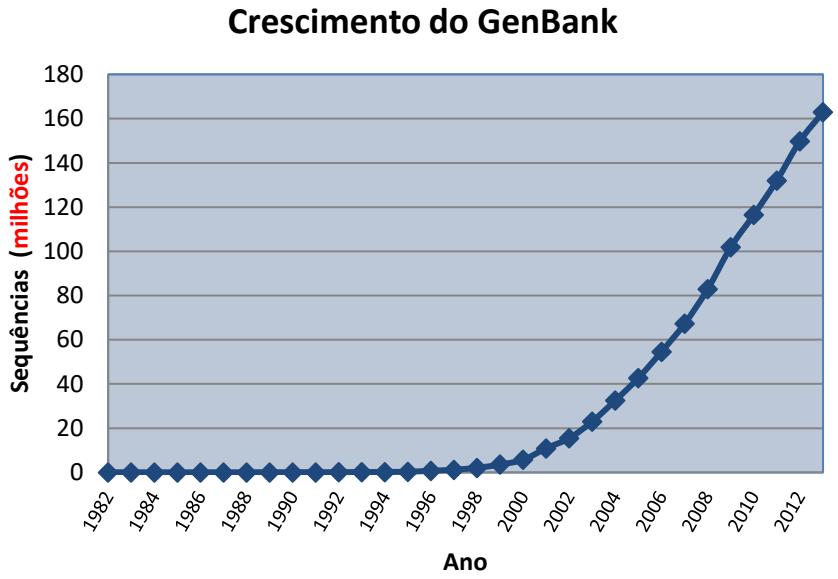
...AVAGGATILVHNQDAGEPAIVLAFG...

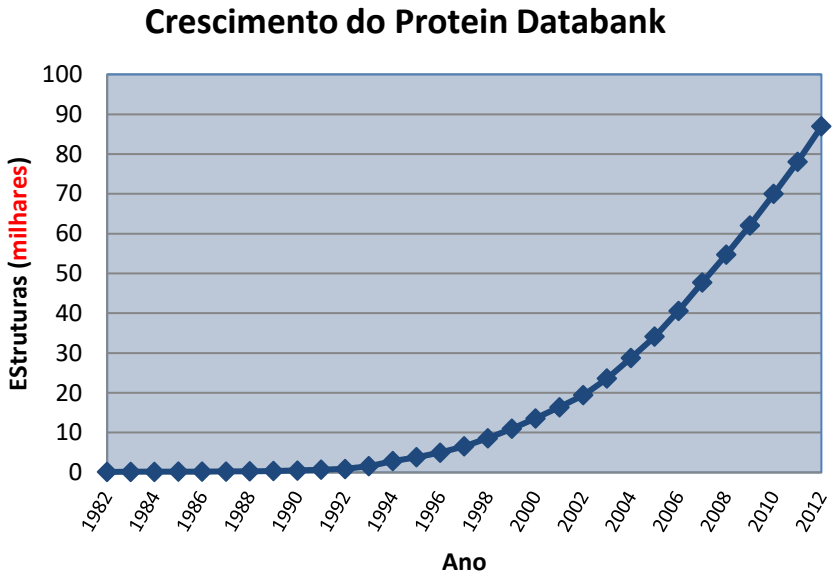


Sequência

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,** …, …

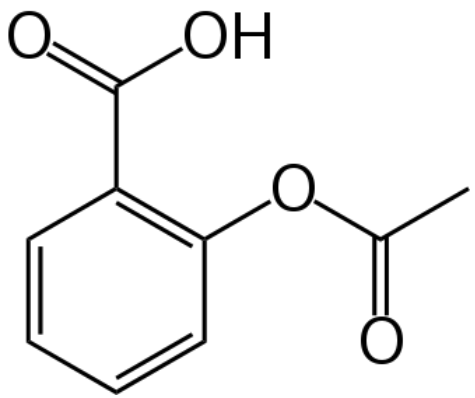
Representação da Aspirina em formato MDL2

```
@<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**.

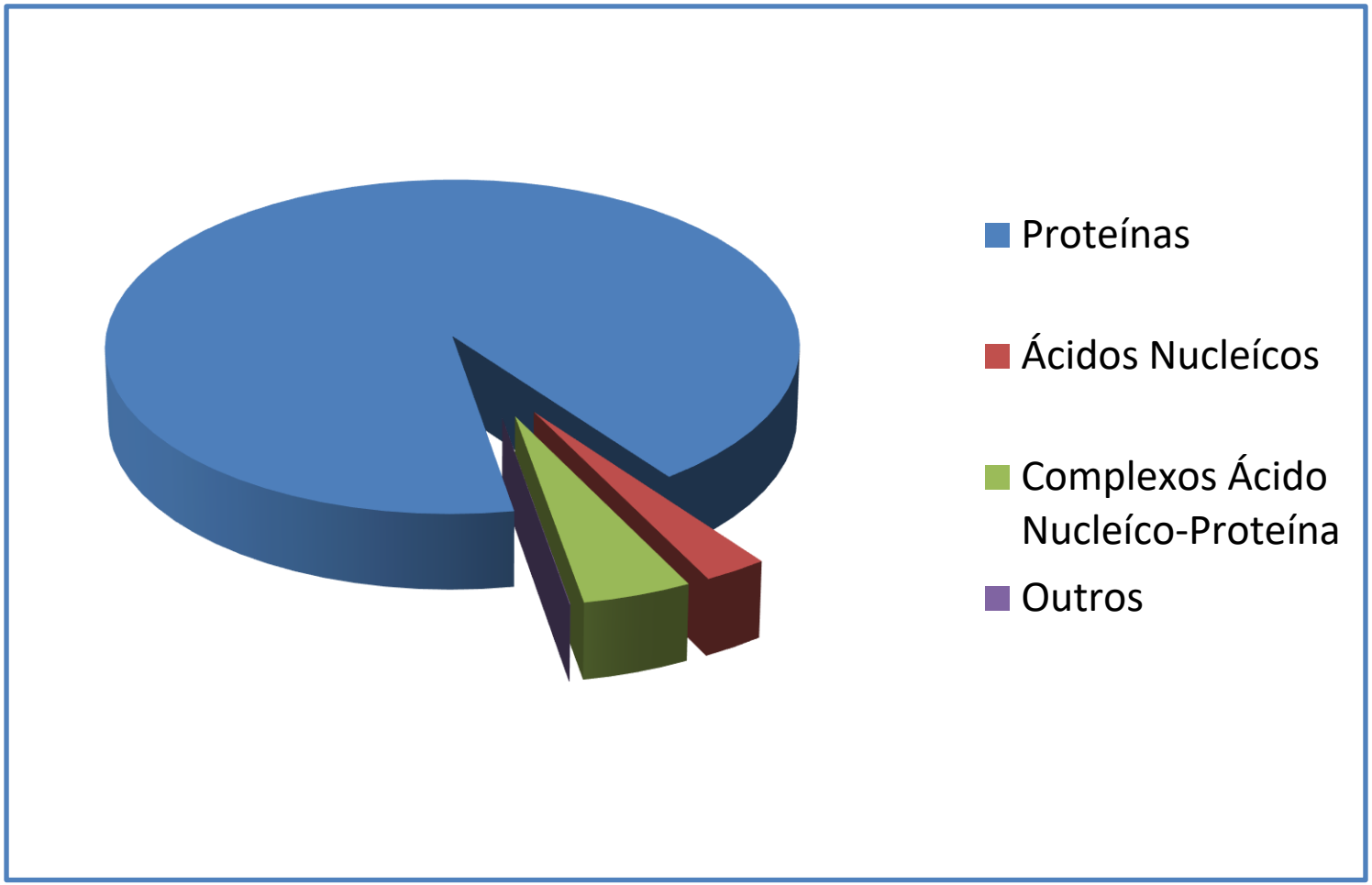
• Contendo inicialmente estruturas de proteínas, o PDB contem hoje em dia outros tipos de moléculas, tais como ácidos nucleicos, lípidos e polissacáridos.

• Número total de estruturas em 9/1/2022: **185610**

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

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

O Protein Data Bank contem vários tipos de macromoléculas



De onde provêm a informação estrutural ?

Combinação de vários tipos de conhecimento:

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

Métodos experimentais

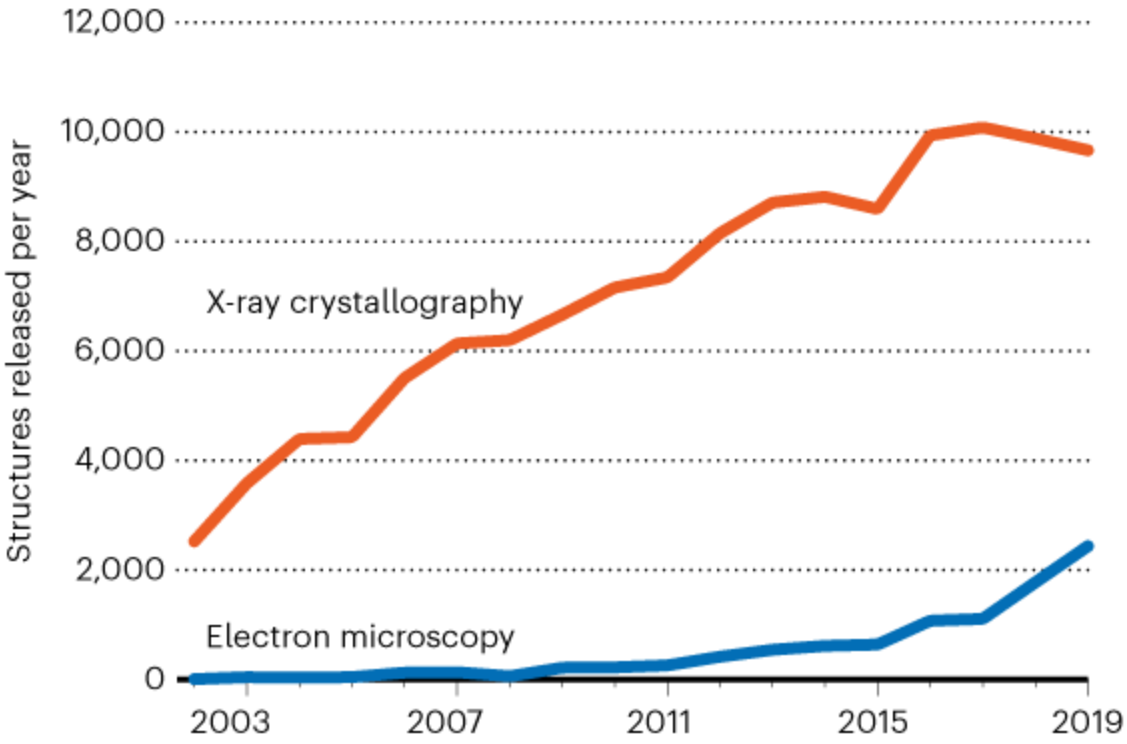
- **Cristalografia de raios X**: a molécula a estudar é purificada e cristalizada a partir de uma solução concentrada. Um feixe de raios X é projectado através do cristal da molécula e o padrão de difracção obtido é usado para resolver a estrutura.

- **Ressonância magnética Nuclear**: a molécula purificada é colocada numa solução aquosa bastante concentrada. A acção de um campo magnético muito intense provoca o desdobramento dos níveis de energia do spin nuclear de alguns elementos (H, ^{13}C, ^{15}N), permitindo o estudo do seu ambiente químico e a determinação da estrutura da macromolécula.

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

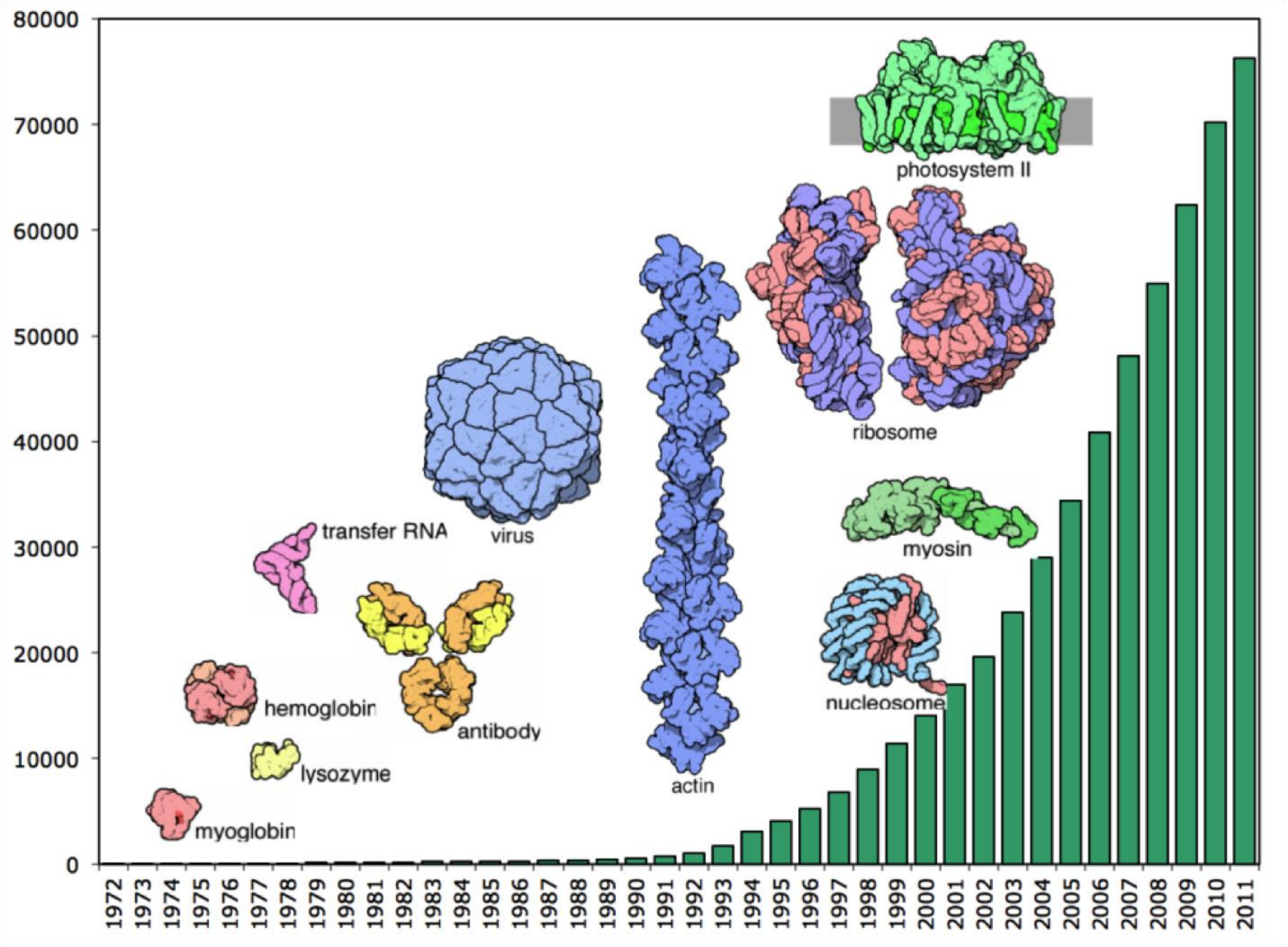
STRUCTURE SLEUTHS

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



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

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 manipulaçãoe 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…

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

RCSB PDB PROTEIN DATA BANK

210,342 Structures from the PDB

1,068,577 Computed Structure Models (CSM)

3D Structures | Enter search term(s), Entry ID(s), or sequence | Include CSM

Advanced Search | Browse Annotations | Help

PDB-101 PDB EMDataResource NAKB NAKB: Nucleic Acid Knowledgebase

New: More Computed Structure Models (CSM) available Learn more

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

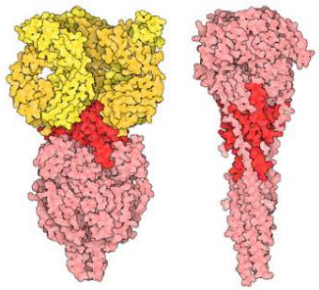
RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

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

Computed Structure Models (CSM) from AlphaFold DB and ModelArchive

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

October Molecule of the Month



RSV Fusion Glycoprotein

Latest Entries

As of Tue Oct 03 2023

8C0V

Structure of the peroxisomal Pex1/Pex6 ATPase complex bound to a substrate in single seam state

Features & Highlights

ASBMB Members: Register Now for Virtual Event

Learn about *Teaching enzymology with the Protein Data Bank: from pandemic to Paxlovid* on October 19.

Register Now for October Virtual Crash Courses on RCSB PDB APIs

Learn about *Leveraging RCSB PDB APIs for Bioinformatics Analyses and Machine Learning*. Part 1 will introduce the Data and Search APIs. Part 2 will offer hands-on exercises.

Coming Soon: PDB Entries with Novel Ligands Distributed Only in PDBx/mmCIF and PDBML File Formats

PDB will run out of 3-character CCD IDs before the end of 2023.

News

Publications

Fall Newsletter Published

In this issue: Fall Training Opportunities; DNS Name Changes; IUCr Report; and more. In the Education Corner, learn about *MedChemBlog: An Innovative Distance Learning Experience for Teaching Medicinal Chemistry*

» 10/08/2023

The Nobel Prize in Physiology or Medicine 2023

Congratulations to Katalin Karikó and Drew Weissman on the award *for their discoveries concerning nucleoside base modifications that enabled the development of effective mRNA vaccines against COVID-19.* PDB-101 offers resources to learn more about

PDB at a Glance 64,698 Structures of Human Sequences 16,701 Nucleic Acid Containing Structures

More Statistics

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

RCSB PDB PROTEIN DATA BANK

210,342 Structures from the PDB

1,068,577 Computed Structure Models (CSM)

3D Structures | Lysozyme | Include CSM | Help

in Structure Keywords
LYSOZYME
Immune system, Lysozyme
in UniProt Molecule Name
Lysozyme
Lysozyme 1
Lysozyme C
Lysozyme C I
Lysozyme C II
Lysozyme C, milk isozyme
Lysozyme C-1
Lysozyme C-2
Lysozyme g
in Additional Structure Keywords
lysozyme inhibitor, g-type lysozyme binding, HYDROLASE

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

New: More Computed S

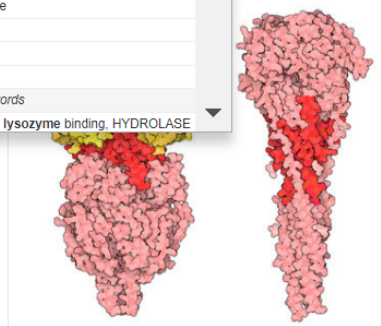
Welcome
Deposit
Search
Visualize
Analyze
Download
Learn

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

- Experimentally-determined 3D structures from the Protein Data Bank (PDB) archive
- Computed Structure Models (CSM) from AlphaFold DB and ModelArchive

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

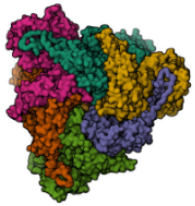
of the Month



RSV Fusion Glycoprotein

Latest Entries

As of Tue Oct 03 2023



8C0V

Structure of the peroxisomal Pex1/Pex6 ATPase complex bound to a substrate in single seam state

Features & Highlights

ASBMB Members: Register Now for Virtual Event

Learn about *Teaching enzymology with the Protein Data Bank: from pandemic to Paxlovid* on October 19.

Register Now for October Virtual Crash Courses on RCSB PDB APIs

Learn about *Leveraging RCSB PDB APIs for Bioinformatics Analyses and Machine Learning*. Part 1 will introduce the Data and Search APIs. Part 2 will offer hands-on exercises.

Coming Soon: PDB Entries with Novel Ligands Distributed Only in PDBx/mmCIF and PDBML File Formats

PDB will run out of 3-character CCD IDs before the end of 2023.

News | Publications

Fall Newsletter Published

In this issue: Fall Training Opportunities; DNS Name Changes; IUCr Report; and more. In the Education Corner, learn about *MedChemBlog: An Innovative Distance Learning Experience for Teaching Medicinal Chemistry*

» 10/08/2023

The Nobel Prize in Physiology or Medicine 2023

Congratulations to Katalin Karikó and Drew Weissman on the award *for their discoveries concerning nucleoside base modifications that enabled the development of effective mRNA vaccines against COVID-19*. PDB-101 offers resources to learn more about

PDB at a Glance | 64,698 Structures of Human Sequences | 16,701 Nucleic Acid Containing Structures | More Statistics

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

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

Search Summary

This query matches 5,868 **Structures**.

Refinements

Structure Determination Methodology

- experimental (5,868)

Scientific Name of Source Organism

- Gallus gallus (1,155)
- Homo sapiens (1,114)
- Tequatrovirus T4 (810)
- Mus musculus (170)
- Escherichia coli (157)
- Escherichia coli K-12 (123)
- synthetic construct (85)
- Bos taurus (83)
- Listeria monocytogenes EGD-e (58)
- Camelus dromedarius (57)

More...

Taxonomy

- Eukaryota (3,192)
- Bacteria (1,723)
- Duplodnaviria (899)
- Archaea (142)
- Riboviria (129)
- other sequences (85)
- Monodnaviria (7)
- Varidnaviria (6)
- unclassified bacterial viruses (3)
- unclassified sequences (2)

More...

Experimental Method

- X-RAY DIFFRACTION (5,721)
- ELECTRON MICROSCOPY (72)
- SOLUTION NMR (35)
- ELECTRON CRYSTALLOGRAPHY (20)
- NEUTRON DIFFRACTION (15)
- POWDER DIFFRACTION (14)
- EPR (4)
- SOLID-STATE NMR (1)

Polymer Entity Type

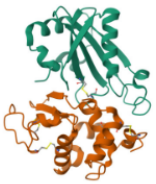
- Protein (5,845)
- DNA (64)
- RNA (33)
- NA-hybrid (4)

Refinement Resolution (Å)

- 0.5 - 1.0 (76)

-- Tabular Repor All Selected

1 to 25 of 5,868 **Structures** Page 1 of 235 2 Sort by ↓ Score



3D View

1UUZ

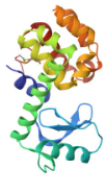
Download File View File

IVY:A NEW FAMILY OF PROTEIN

Abergel, C., Lembo, F., Byrne, D., Maza, C., Claverie, J.M.

(2007) Proc Natl Acad Sci U S A **104**: 6394

Released 2004-01-14
Method X-RAY DIFFRACTION 1.8 Å
Organisms Gallus gallus
Pseudomonas aeruginosa
Macromolecule INHIBITOR OF VERTEBRATE LYSOZYME (protein)
LYSOZYME C (protein)



3D View

1L58

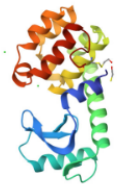
Download File View File

ANALYSIS OF THE INTERACTION BETWEEN CHARGED SIDE CHAINS AND THE ALPHA-HELIX DIPOLE USING DESIGNED THERMOSTABLE MUTANTS OF PHAGE T4 LYSOZYME

Nicholson, H., Matthews, B.W.

To be published

Released 1991-10-15
Method X-RAY DIFFRACTION 1.65 Å
Organisms Tequatrovirus T4
Macromolecule T4 LYSOZYME (protein)
Unique Ligands BME



3D View

1L36

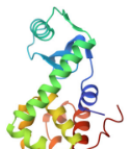
Download File View File

TOWARD A SIMPLIFICATION OF THE PROTEIN FOLDING PROBLEM: A STABILIZING POLYALANINE ALPHA-HELIX ENGINEERED IN T4 LYSOZYME

Zhang, X.-J., Baase, W.A., Matthews, B.W.

(1991) Biochemistry **30**: 2012-2017

Released 1991-10-15
Method X-RAY DIFFRACTION 1.7 Å
Organisms Tequatrovirus T4
Macromolecule LYSOZYME (protein)
Unique Ligands BME, CL



1L54

Download File View File

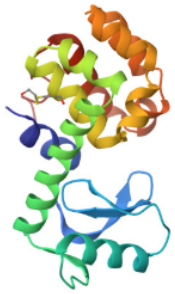
THE STRUCTURAL AND THERMODYNAMIC CONSEQUENCES OF BURYING A CHARGED RESIDUE WITHIN THE HYDROPHOBIC CORE OF T4 LYSOZYME

Daopin, S., Matthews, B.W.

(1991) Biochemistry **30**: 11521-11529

Released 1991-10-15

Biological Assembly 1



3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

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

Find Similar Assemblies

Biological assembly 1 assigned by authors.

Macromolecule Content

- Total Structure Weight: 18.71 kDa
- Atom Count: 1,462
- Modelled Residue Count: 164
- Deposited Residue Count: 164
- Unique protein chains: 1

Display Files | Download Files | Data API

1L58

ANALYSIS OF THE INTERACTION BETWEEN CHARGED SIDE CHAINS AND THE ALPHA-HELIX DIPOLE USING DESIGNED THERMOSTABLE MUTANTS OF PHAGE T4 LYSOZYME

PDB DOI: https://doi.org/10.2210/pdb1L58/pdb

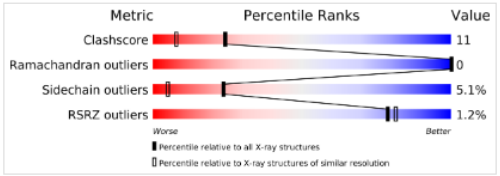
Classification: HYDROLASE (O-GLYCOSYL)
Organism(s): Tequatrovirus T4
Mutation(s): No

Deposited: 1991-05-06 **Released:** 1991-10-15
Deposition Author(s): Nicholson, H., Matthews, B.W.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.65 Å
R-Value Observed: 0.167

wwPDB Validation 3D Report | Full Report



This is version 1.3 of the entry. See complete history.

Literature

Download Primary Citation

Nicholson, H., Becktel, W., Matthews, B.W.

To be published.

Macromolecules

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

Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details	Image

Macromolecules

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

Entity ID: 1

Molecule	Chains	Sequence Length	Organism	Details	Image
T4 LYSOZYME	A	164	Tequatrovirus T4	Mutation(s): 0 Gene Names: E EC: 3.2.1.17	

UniProt

Find proteins for P00720 *(Enterobacteria phage T4)* Explore P00720 Go to UniProtKB: P00720

Entity Groups

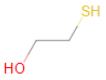
Sequence Clusters	30% Identity 50% Identity 70% Identity 90% Identity 95% Identity 100% Identity
UniProt Group	P00720

Protein Feature View Expand

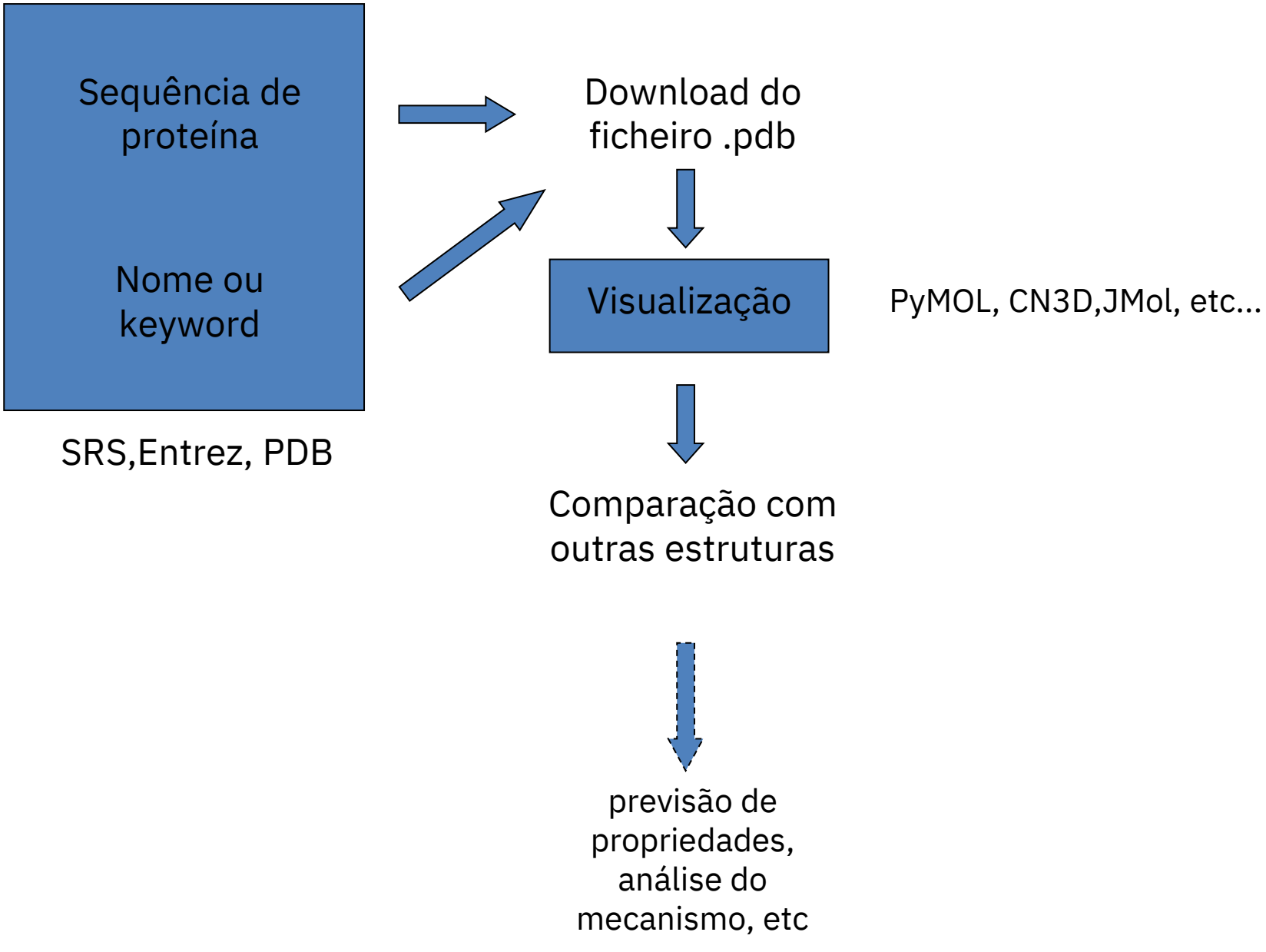
Reference Sequence 1L58_1

Small Molecules

Ligands 1 Unique

ID	Chains	Name / Formula / InChI Key	2D Diagram	3D Interactions
BME Query on BME Download Ideal Coordinates CCD File Download Instance Coordinates	B [auth A]	**BETA-MERCAPTOETHANOL** $C_2 H_6 O S$ DGVVWUTYPXICAM-UHFFFAOYSA-N		Ligand Interaction

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

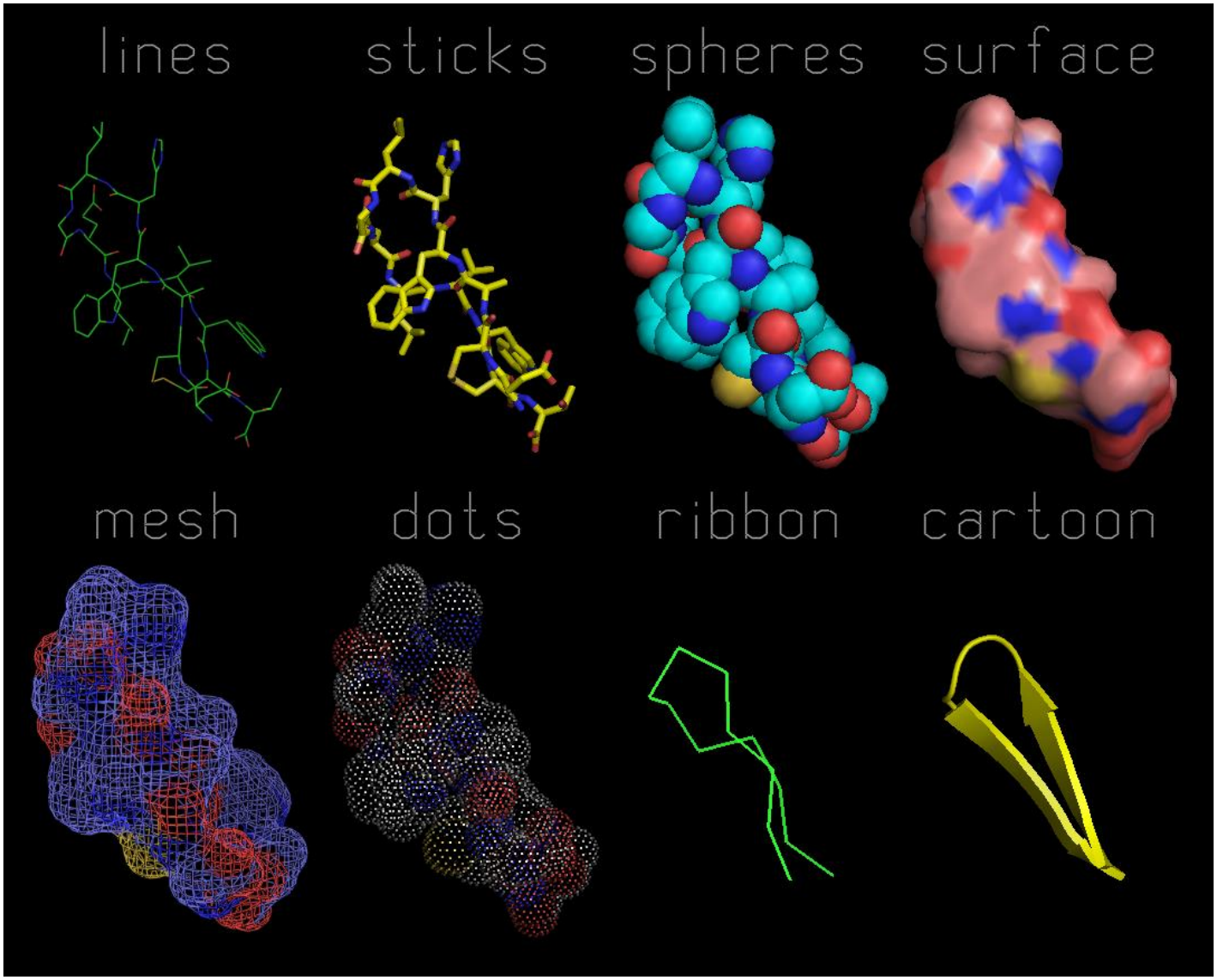
Instaláveis:

- **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://webcsd.ccdc.cam.ac.uk
- ChemDB (database+tools) - http://www.chemdb.com
- MOLE DB (molecular descriptors) - 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 interligadas:
  - Compound (**111,050,847**)
  - Substance (**277,194,318**)
  - Bioassay (**1,391,562**)

  } **21/11/2021**
- Permite pesquisa por estrutura, similaridade, etc…

PubChem Data (21/11/2021)

Data Collection	Live Count	Description
Compounds	111,050,847	Unique chemical structures extracted from contributed PubChem Substance records
Substances	277,194,318	Information about chemical entities provided by PubChem contributors
BioAssays	1,391,562	Biological experiments provided by PubChem contributors
Bioactivities	292,633,795	Biological activity data points reported in PubChem BioAssays
Genes	103,715	Gene targets tested in PubChem BioAssays and those involved in PubChem Pathways
Proteins	96,561	Protein targets tested in PubChem BioAssays and those involved in PubChem Pathways
Taxonomy	531,241	Organisms of targets tested in PubChem BioAssays and those involved in PubChem Pathways
Pathways	238,597	Interactions between chemicals, genes, and proteins
Literature	33,307,005	Scientific publications with links in PubChem
Patents	28,543,965	Patents with links in PubChem
Data Sources	824	Organizations contributing data to PubChem

21/11/2021

https://pubchemdocs.ncbi.nlm.nih.gov/statistics

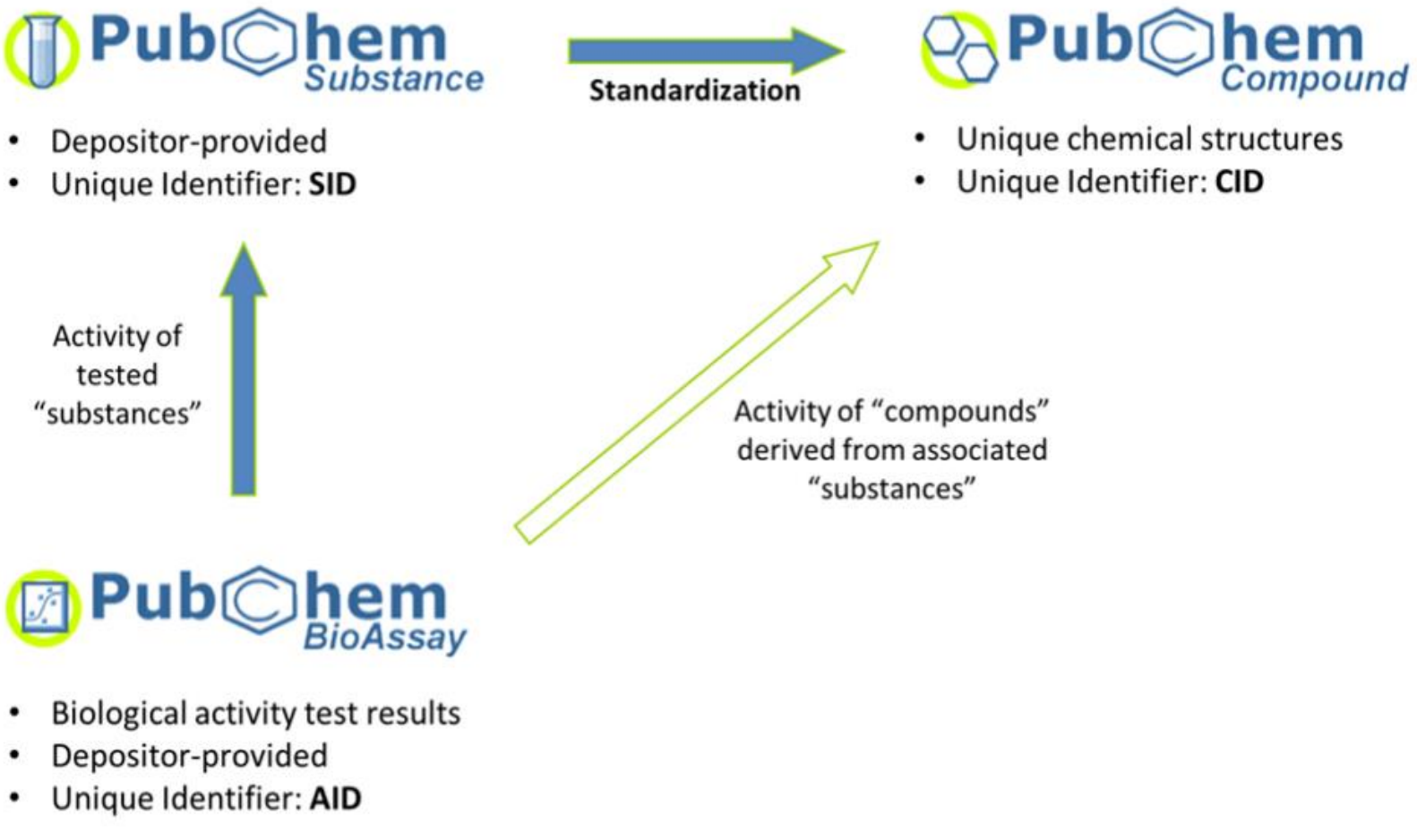
Bases de dados

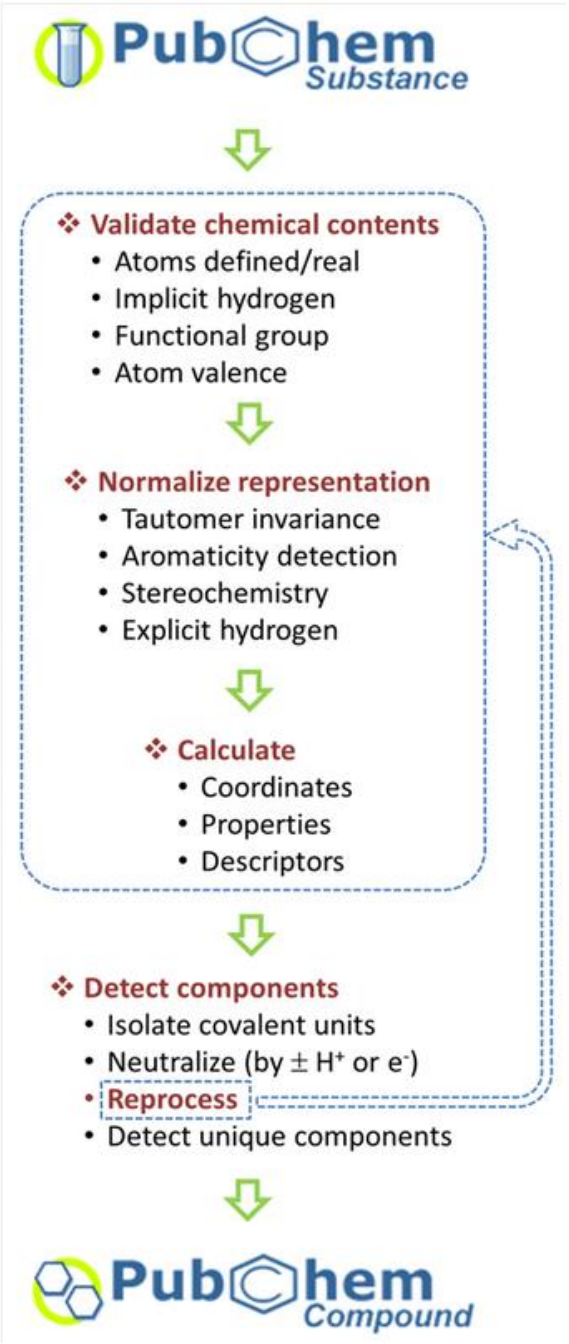
PubChem

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

- **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.

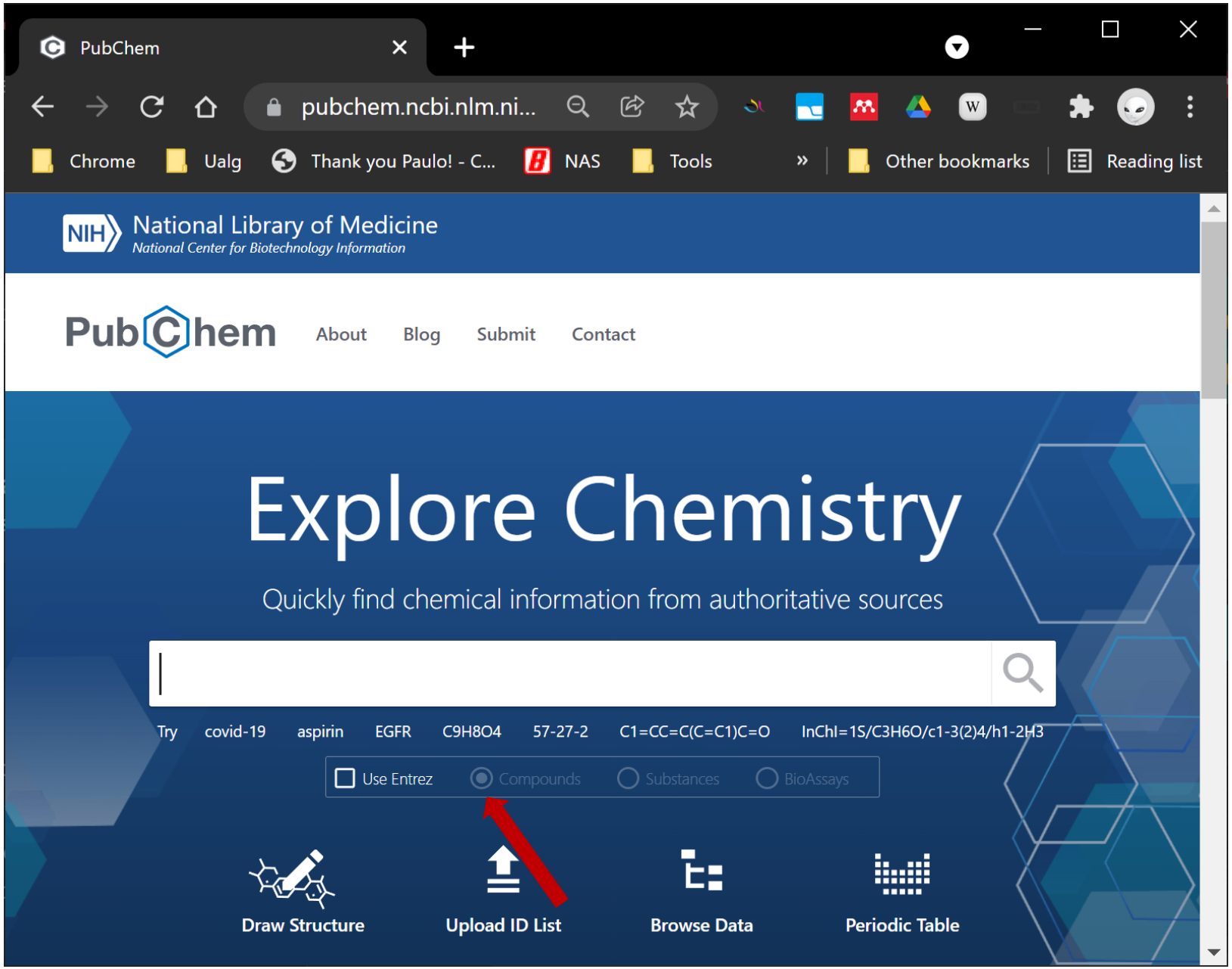

PubChem Substance
Standardization
PubChem Compound
• Depositor-provided
• Unique Identifier: SID
• Unique chemical structures
• Unique Identifier: CID
Activity of tested "substances"
Activity of "compounds" derived from associated "substances"
PubChem BioAssay
• Biological activity test results
• Depositor-provided
• Unique Identifier: AID

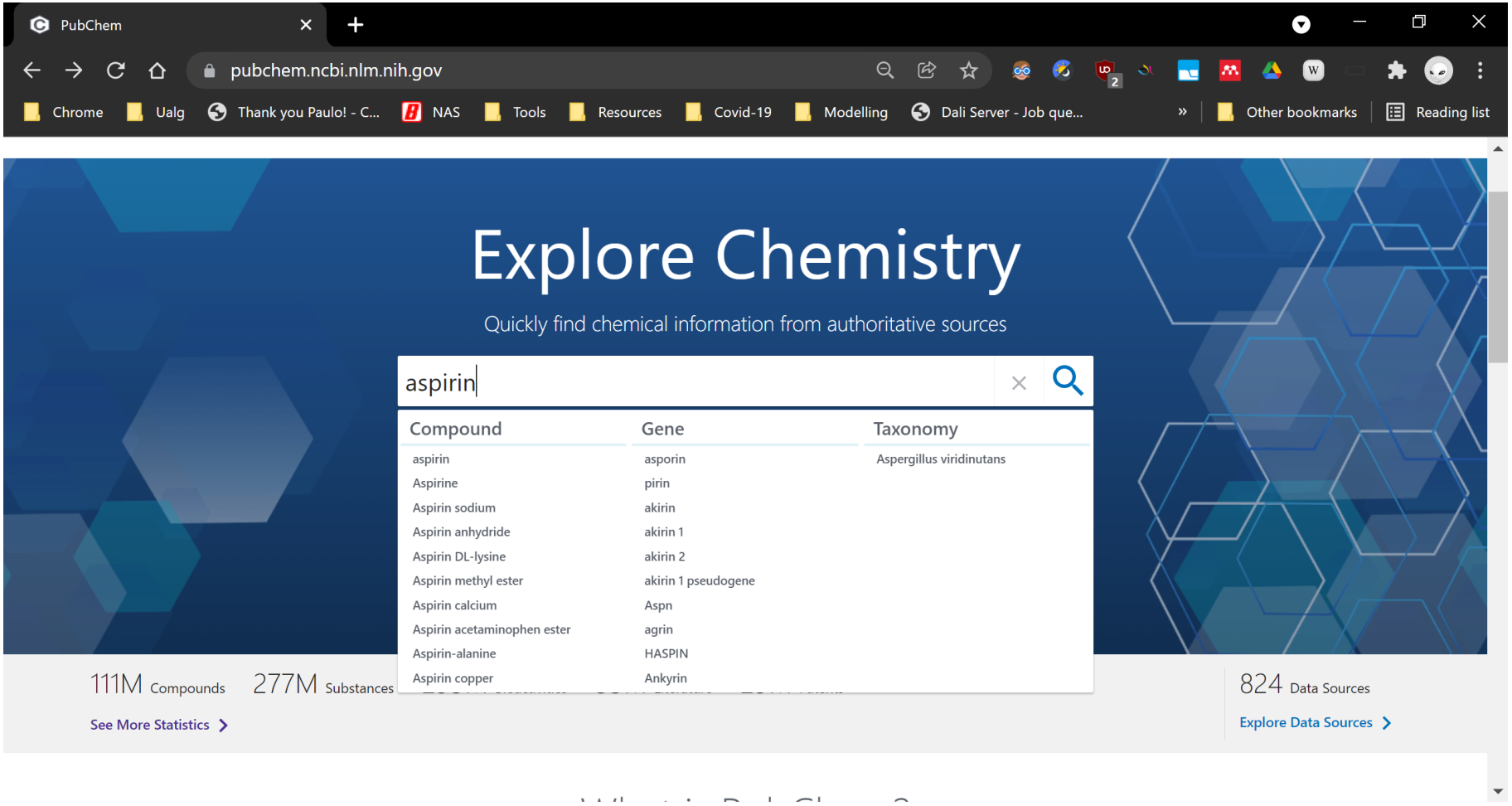

PubChem Substance
❖ Validate chemical contents
• Atoms defined/real
• Implicit hydrogen
• Functional group
• Atom valence
❖ Normalize representation
• Tautomer invariance
• Aromaticity detection
• Stereochemistry
• Explicit hydrogen
❖ Calculate
• Coordinates
• Properties
• Descriptors
❖ Detect components
• Isolate covalent units
• Neutralize (by ± H+ or e-)
• Reprocess
• Detect unique components
PubChem Compound

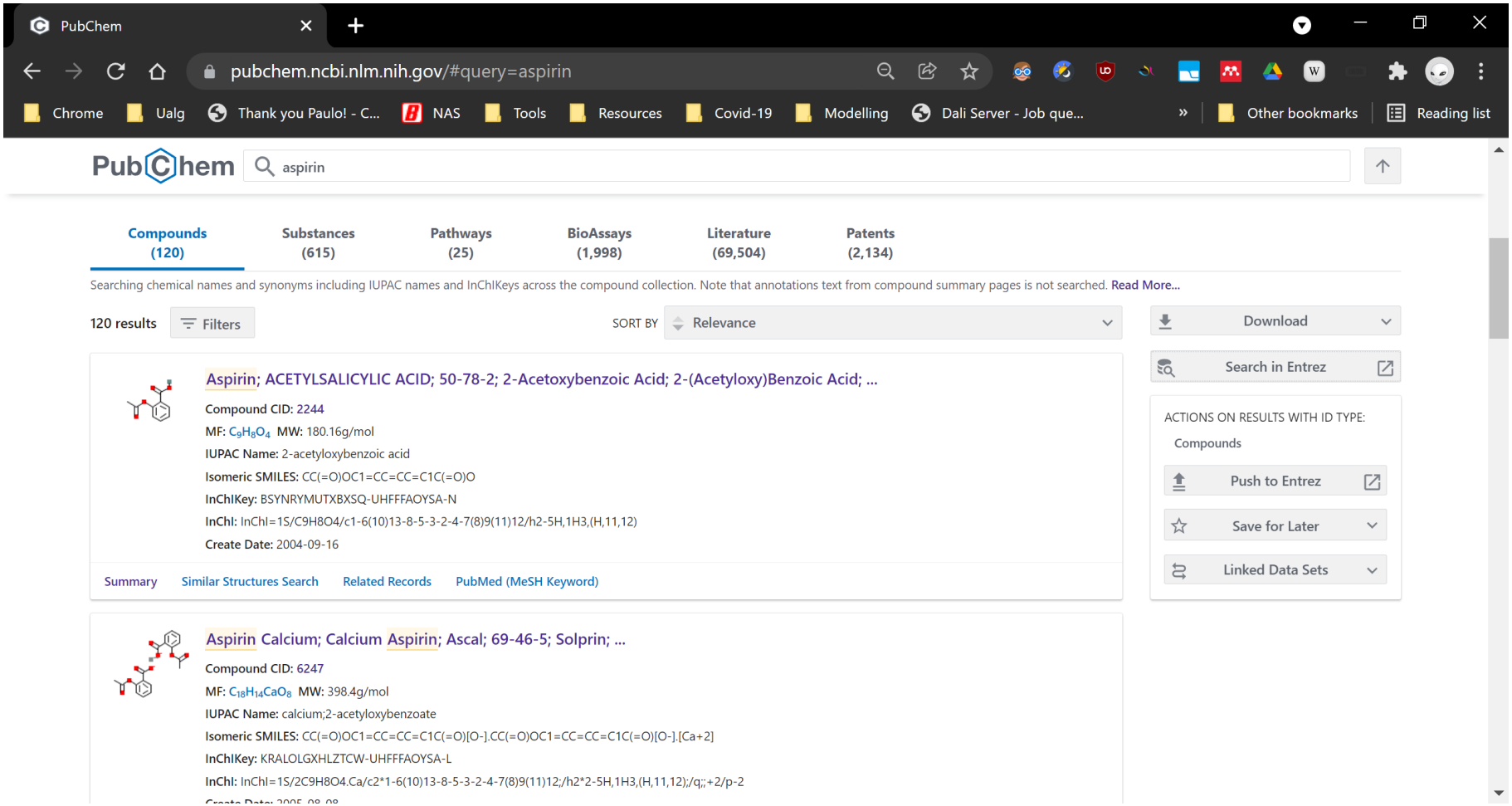
Pesquisa

PubChem

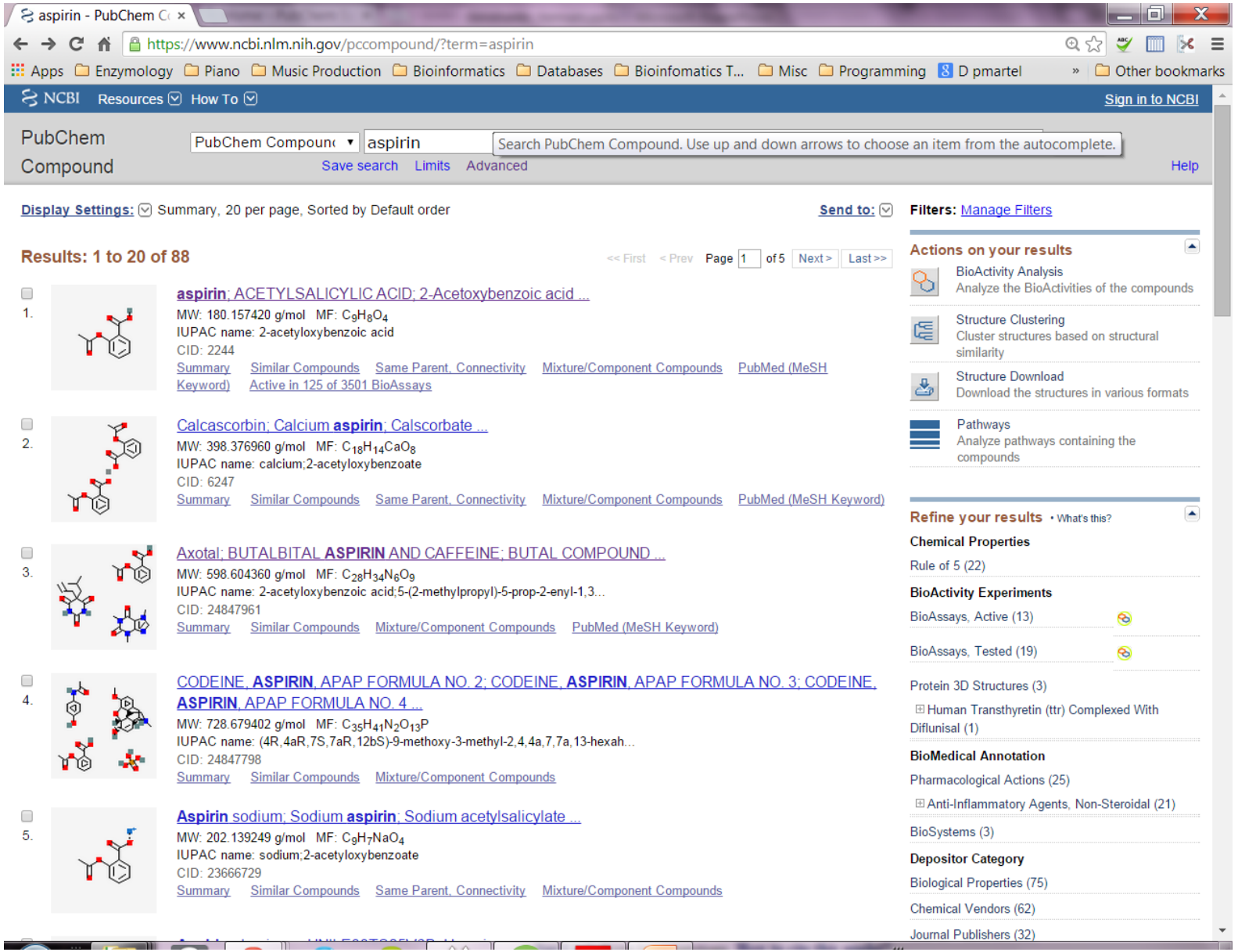
- **Compound:** nomes, sinónimos ou keywords.
- **Substance:** nomes, sinónimos, keywords
- **Bioassay:** pesquisa de termos nas descrição do ensaio
- **Entrez:** pesquisar usando as ferramentas do NCBI
- **Estrutura:** pesquisar por similaridade de estrutura
- **Ferramentas de análise:** SAR maps, tabelas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


PubChem
pubchem.ncbi.nlm.ni...
Chrome
Ualg
Thank you Paulo! - C...
NAS
Tools
Other bookmarks
Reading list
National Library of Medicine
National Center for Biotechnology Information
PubChem
About
Blog
Submit
Contact
Explore Chemistry
Quickly find chemical information from authoritative sources
Try
covid-19
aspirin
EGFR
C9H8O4
57-27-2
C1=CC=C(C=C1)C=O
InChI=1S/C3H6O/c1-3(2)4/h1-2H3
Use Entrez
Compounds
Substances
BioAssays
Draw Structure
Upload ID List
Browse Data
Periodic Table


PubChem
pubchem.ncbi.nlm.nih.gov
Chrome
Ualg
Thank you Paulo! - C...
NAS
Tools
Resources
Covid-19
Modelling
Dali Server - Job que...
Other bookmarks
Reading list
Explore Chemistry
Quickly find chemical information from authoritative sources
aspirin
Compound
aspirin
Aspirine
Aspirin sodium
Aspirin anhydride
Aspirin DL-lysine
Aspirin methyl ester
Aspirin calcium
Aspirin acetaminophen ester
Aspirin-alanine
Aspirin copper
Gene
asporin
pirin
akirin
akirin 1
akirin 2
akirin 1 pseudogene
Aspn
agrin
HASPIN
Ankyrin
Taxonomy
Aspergillus viridinutans
111M Compounds
277M Substances
See More Statistics
824 Data Sources
Explore Data Sources


PubChem
pubchem.ncbi.nlm.nih.gov/#query=aspirin
Chrome
Ualg
Thank you Paulo! - C...
NAS
Tools
Resources
Covid-19
Modelling
Dali Server - Job que...
Other bookmarks
Reading list
PubChem
aspirin
Compounds (120)
Substances (615)
Pathways (25)
BioAssays (1,998)
Literature (69,504)
Patents (2,134)
Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that annotations text from compound summary pages is not searched. Read More...
120 results
Filters
SORT BY
Relevance
Download
Search in Entrez
ACTIONS ON RESULTS WITH ID TYPE:
Compounds
Push to Entrez
Save for Later
Linked Data Sets
Aspirin; ACETYLSALICYLIC ACID; 50-78-2; 2-Acetoxybenzoic Acid; 2-(Acetyloxy)Benzoic Acid; ...
Compound CID: 2244
MF: C9H8O4 MW: 180.16g/mol
IUPAC Name: 2-acetyloxybenzoic acid
Isomeric SMILES: CC(=O)OC1=CC=CC=C1C(=O)O
InChIKey: BSYNRYMUTXBXSQ-UHFFFAOYSA-N
InChI: InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)
Create Date: 2004-09-16
Summary
Similar Structures Search
Related Records
PubMed (MeSH Keyword)
Aspirin Calcium; Calcium Aspirin; Ascal; 69-46-5; Solprin; ...
Compound CID: 6247
MF: C18H14CaO8 MW: 398.4g/mol
IUPAC Name: calcium;2-acetyloxybenzoate
Isomeric SMILES: CC(=O)OC1=CC=CC=C1C(=O)[O-].CC(=O)OC1=CC=CC=C1C(=O)[O-].[Ca+2]
InChIKey: KRALOLGXHLZTCW-UHFFFAOYSA-L
InChI: InChI=1S/2C9H8O4.Ca/c2*1-6(10)13-8-5-3-2-4-7(8)9(11)12;/h2*2-5H,1H3,(H,11,12);/q;;+2/p-2

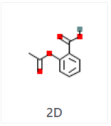
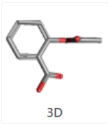
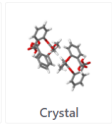
PubChem Compound



Aspirin | HC9H7O4 - PubChem

pubchem.ncbi.nlm.nih.gov/compound/2244

Chrome | Ualg | Thank you Paulo! - C... | NAS | Tools | Resources | Covid-19 | Modelling | Dali Server - Job que... | Other bookmarks | Reading list

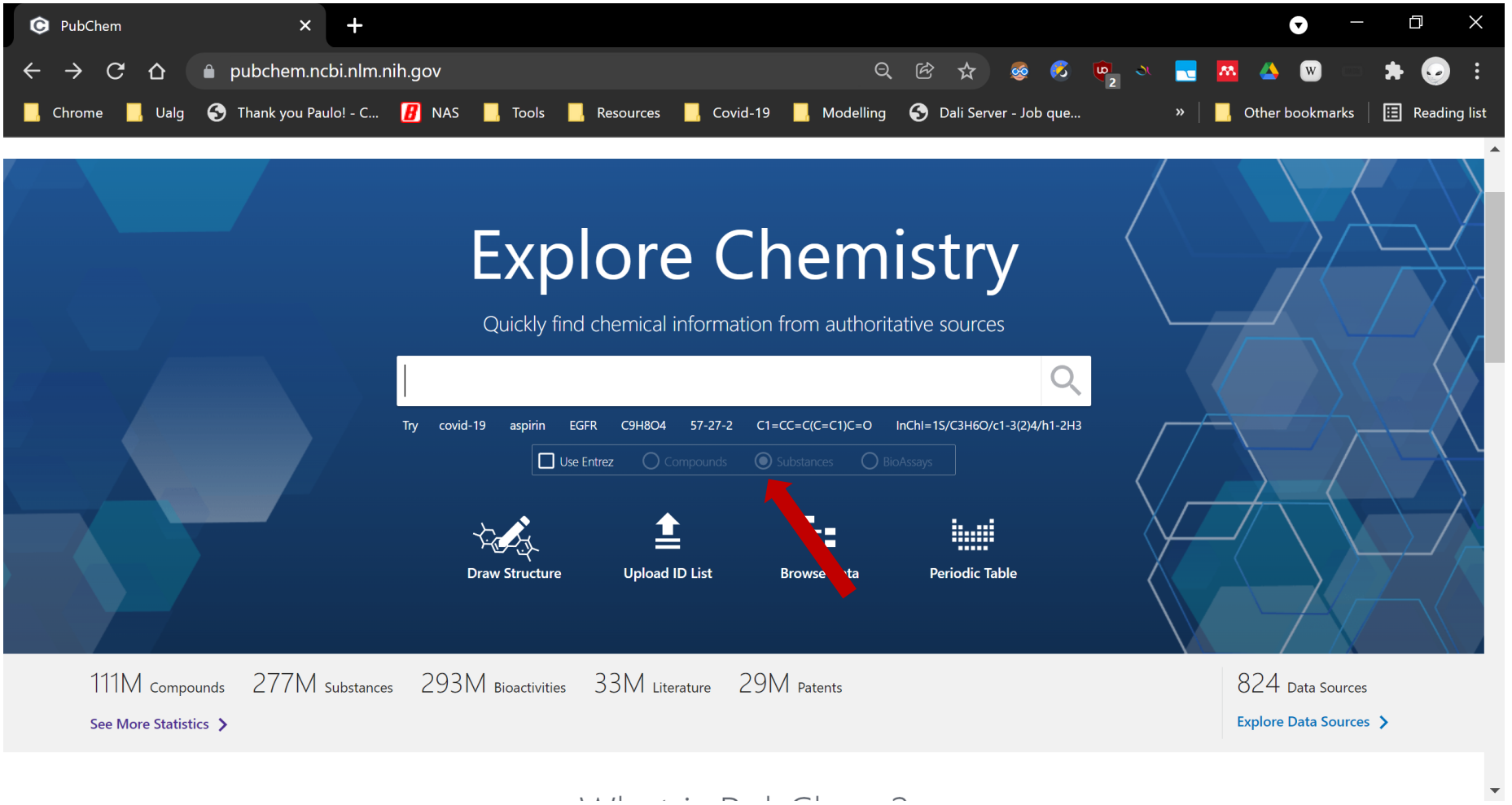
PubChem CID	2244
Structure	 2D  3D  Crystal Find Similar Structures
Chemical Safety	 Irritant Laboratory Chemical Safety Summary (LCSS) Datasheet
Molecular Formula	$C_9H_8O_4$ or $CH_3COOC_6H_4COOH$ or $HC_9H_7O_4$
Synonyms	aspirin ACETYLSALICYLIC ACID 50-78-2 2-Acetoxybenzoic acid 2-(Acetyloxy)benzoic acid More...
Molecular Weight	180.16
Dates	Modify 2021-11-20 Create 2004-09-16

Aspirin or acetylsalicylic acid is perhaps the most commonly used analgesic and antipyretic medication worldwide, having been in clinical use for over 100

Cite | Download

CONTENTS

- Title and Summary
- 1 Structures
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Spectral Information
- 5 Related Records
- 6 Chemical Vendors
- 7 Drug and Medication Information
- 8 Pharmacology and Biochemistry
- 9 Use and Manufacturing
- 10 Identification
- 11 Safety and Hazards
- 12 Toxicity
- 13 Associated Disorders and Diseases
- 14 Literature
- 15 Patents


PubChem
pubchem.ncbi.nlm.nih.gov
Chrome
Ualg
Thank you Paulo! - C...
NAS
Tools
Resources
Covid-19
Modelling
Dali Server - Job que...
Other bookmarks
Reading list
Explore Chemistry
Quickly find chemical information from authoritative sources
Try
covid-19
aspirin
EGFR
C9H8O4
57-27-2
C1=CC=C(C=C1)C=O
InChI=1S/C3H6O/c1-3(2)4/h1-2H3
Use Entrez
Compounds
Substances
BioAssays
Draw Structure
Upload ID List
Periodic Table
111M Compounds
277M Substances
293M Bioactivities
33M Literature
29M Patents
See More Statistics
824 Data Sources
Explore Data Sources

PubChem

pubchem.ncbi.nlm.nih.gov/#query=aspirin&tab=substance

Chrome | Ualg | Thank you Paulo! - C... | NAS | Tools | Resources | Covid-19 | Modelling | Dali Server - Job que... | Other bookmarks | Reading list

PubChem aspirin

Isomeric SMILES: CC(=O)OC1=CC=CC=C1C(=O)O
InChIKey: BSYNRYMUTXBXSQ-UHFFFAOYSA-N
InChI: InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)
Create Date: 2004-09-16

Summary | Similar Structures Search | Related Records | PubMed (MeSH Keyword)

Compounds (120) | **Substances (615)** | Pathways (25) | BioAssays (1,998) | Literature (69,504) | Patents (2,134)

Searching chemical names and synonyms in the substance records submitted by PubChem's contributors. Read More...

615 results | Filters | SORT BY Relevance

Download | Search in Entrez

ACTIONS ON RESULTS WITH ID TYPE:
(•) Substances
( ) Compounds

Push to Entrez | Save for Later | Linked Data Sets

Aspirin; ACETYLSALICYLIC ACID; 2-Acetoxybenzoic Acid; Acetylsalicylate; Acylpyrin; ...
Substance SID: 49854366 **Compound CID:** 2244
Data Source: LeadScope **External ID:** LS-143
Data Source Category: Legacy Depositors; Subscription Services
Deposit Date: 2008-07-09 **Last Modified Date:** 2011-04-18

Aspirin; ACETYLSALICYLIC ACID; 2-Acetoxybenzoic Acid; Acetylsalicylate; Acylpyrin; ...
Substance SID: 319061566 **Compound CID:** 2244
Data Source: ToxPlanet **External ID:** ToxPlanet-NTAtNzgtMjIyNDQ=
Data Source Category: Subscription Services
Deposit Date: 2016-11-25 **Last Modified Date:** 2019-02-10

SID 319061566 - PubChem

pubchem.ncbi.nlm.nih.gov/substance/319061566

Chrome | Ualg | Thank you Paulo! - C... | NAS | Tools | Resources | Covid-19 | Modelling | Dali Server - Job que... | Other bookmarks | Reading list

SUBSTANCE RECORD

2-(Acetyloxy)benzoic acid

Cite | Download

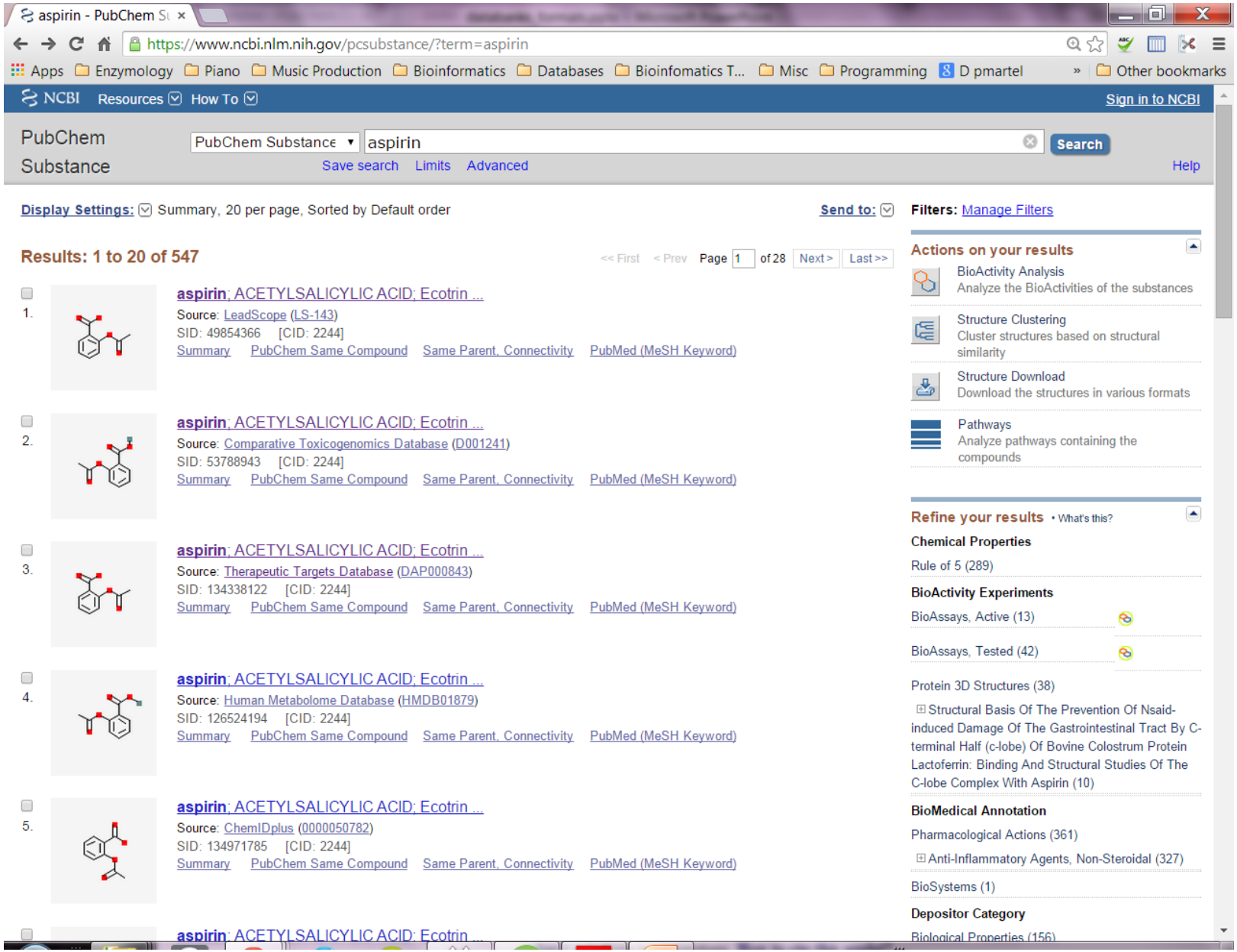
CONTENTS

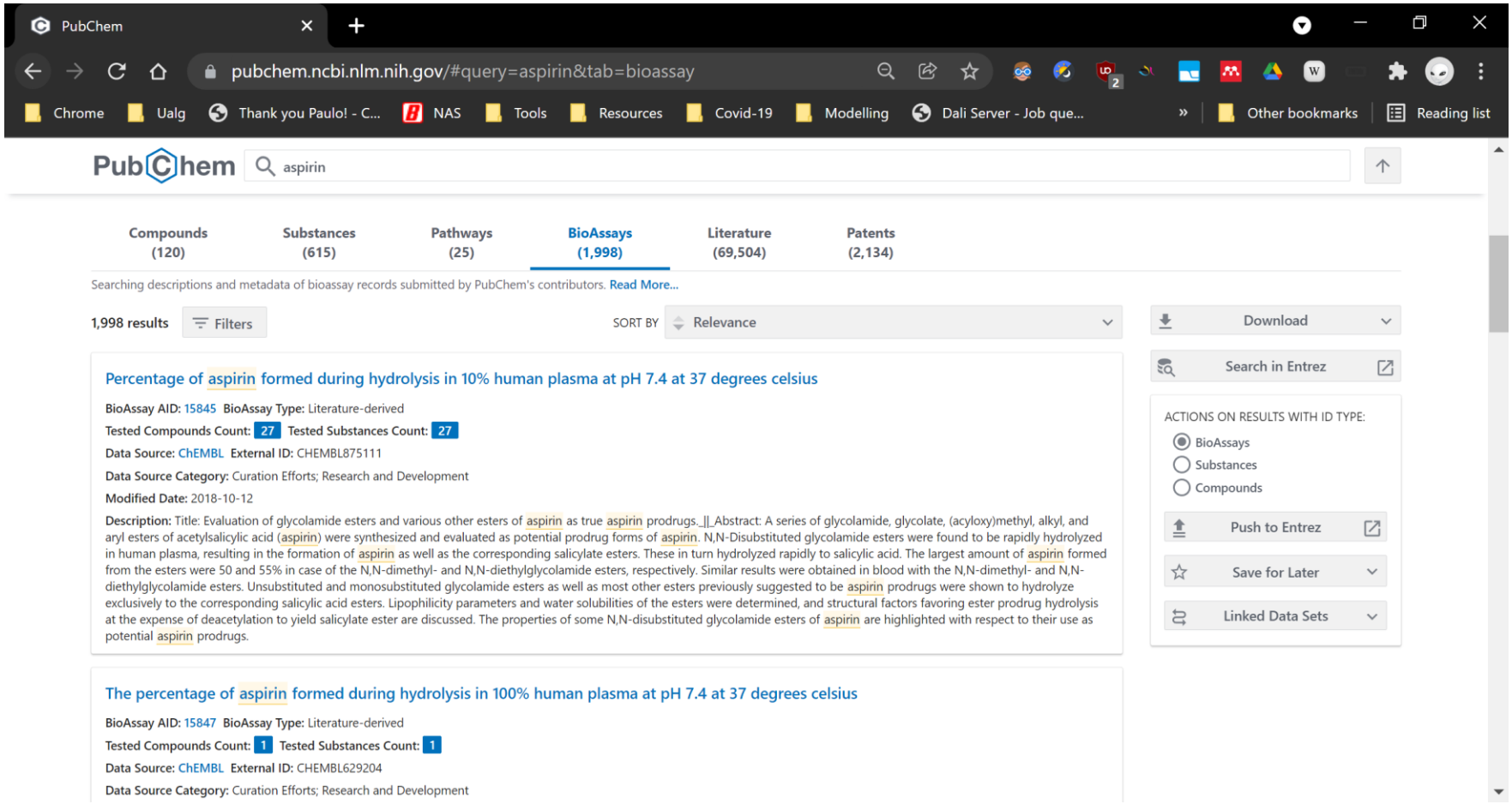
Title and Summary
1 2D Structure
2 Identity
3 Related Records
4 Information Sources

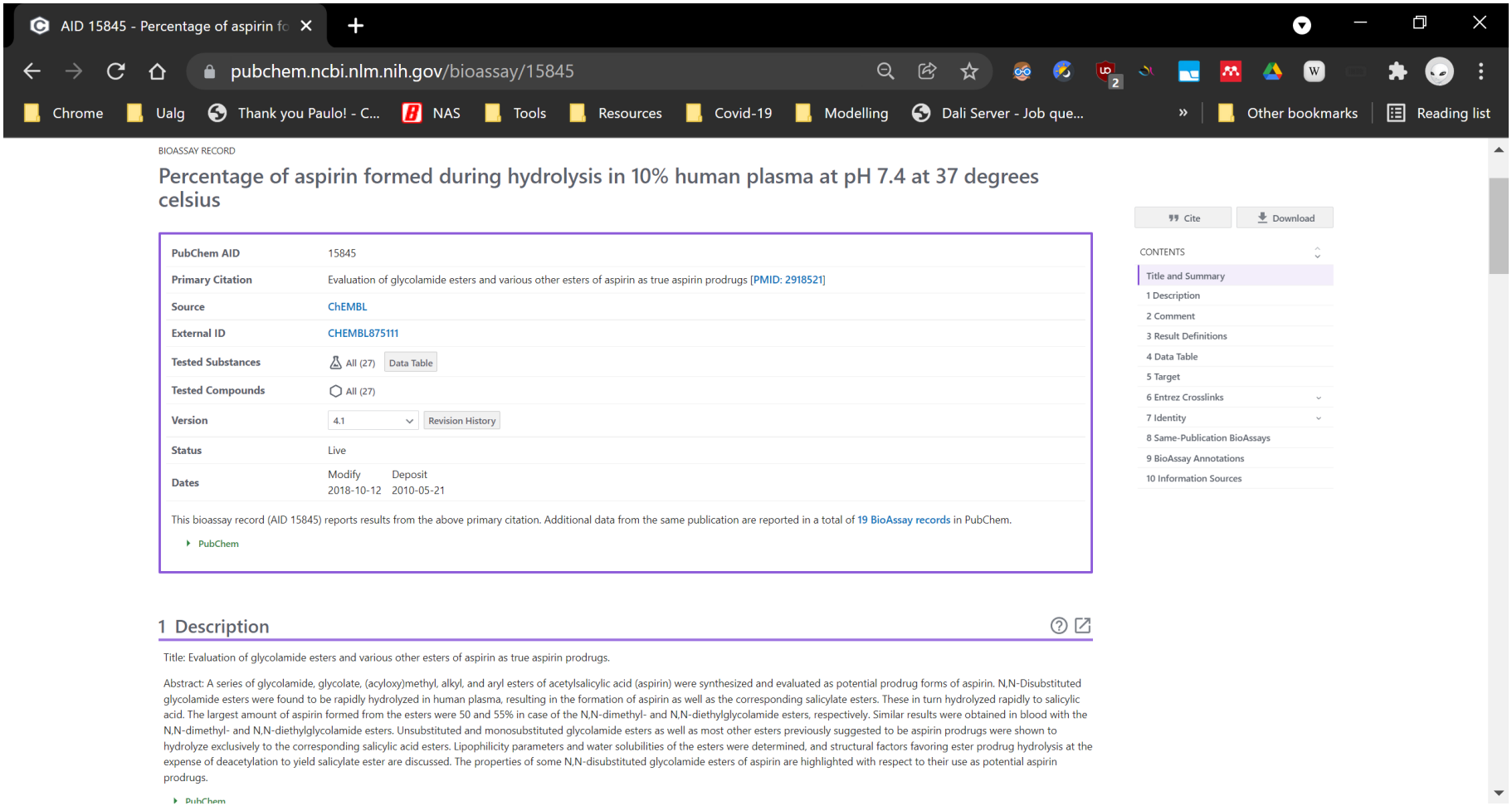
PubChem SID	319061566
Structure	 2D
Source	ToxPlanet
External ID	ToxPlanet-NTAtNzgtMjIyNDQ=
Source Category	Subscription Services
Version	1 Revision History
Status	Live
Related Compounds	PubChem CID CID 2244 (Aspirin)
Dates	Available 2016-11-25 Deposit 2016-11-25

Please note that the substance record is presented as provided to PubChem by the source (depositor). For standardized chemical structure and/or annotation information, please visit the summary page for CID 2244.

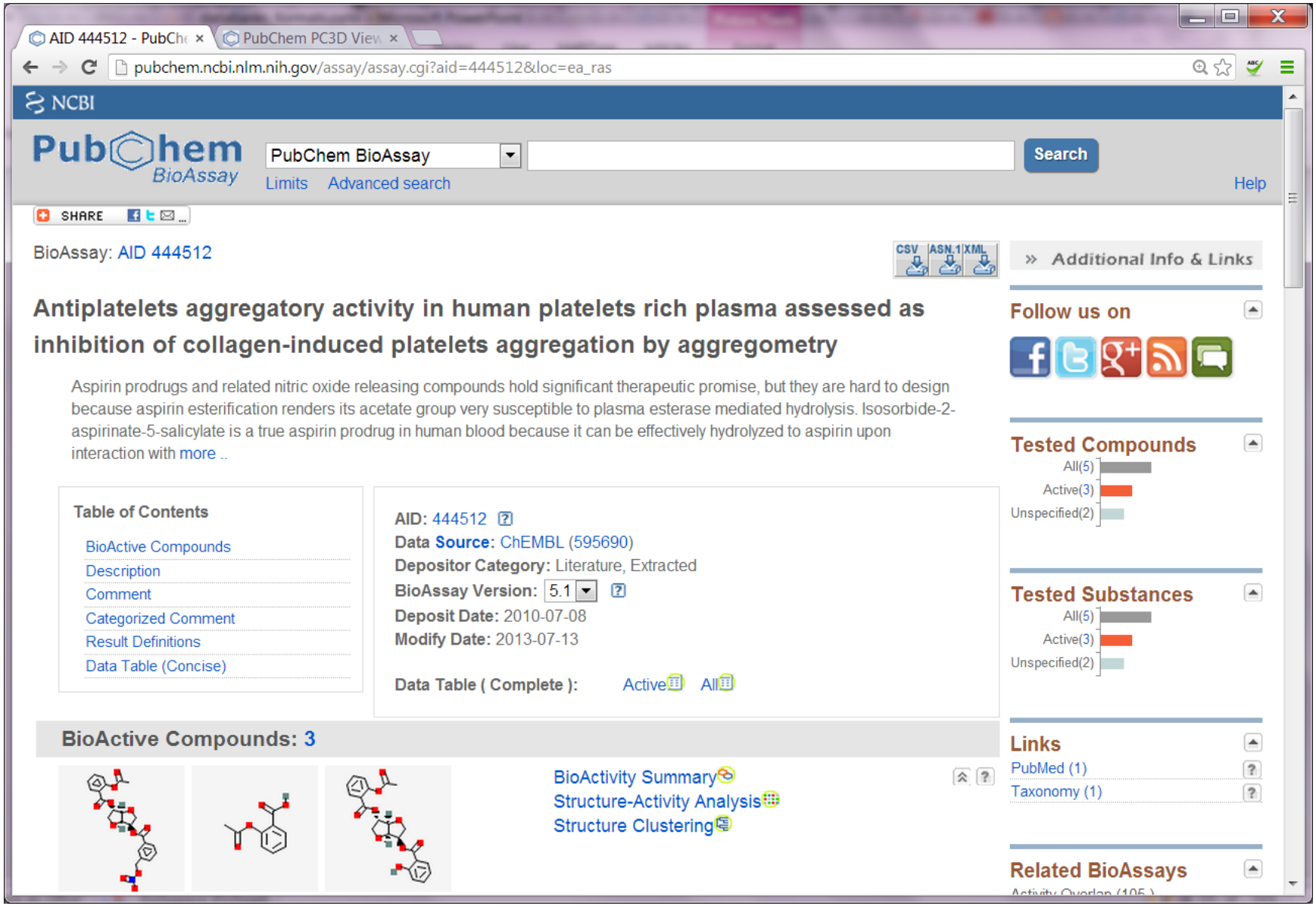
PubChem Substance



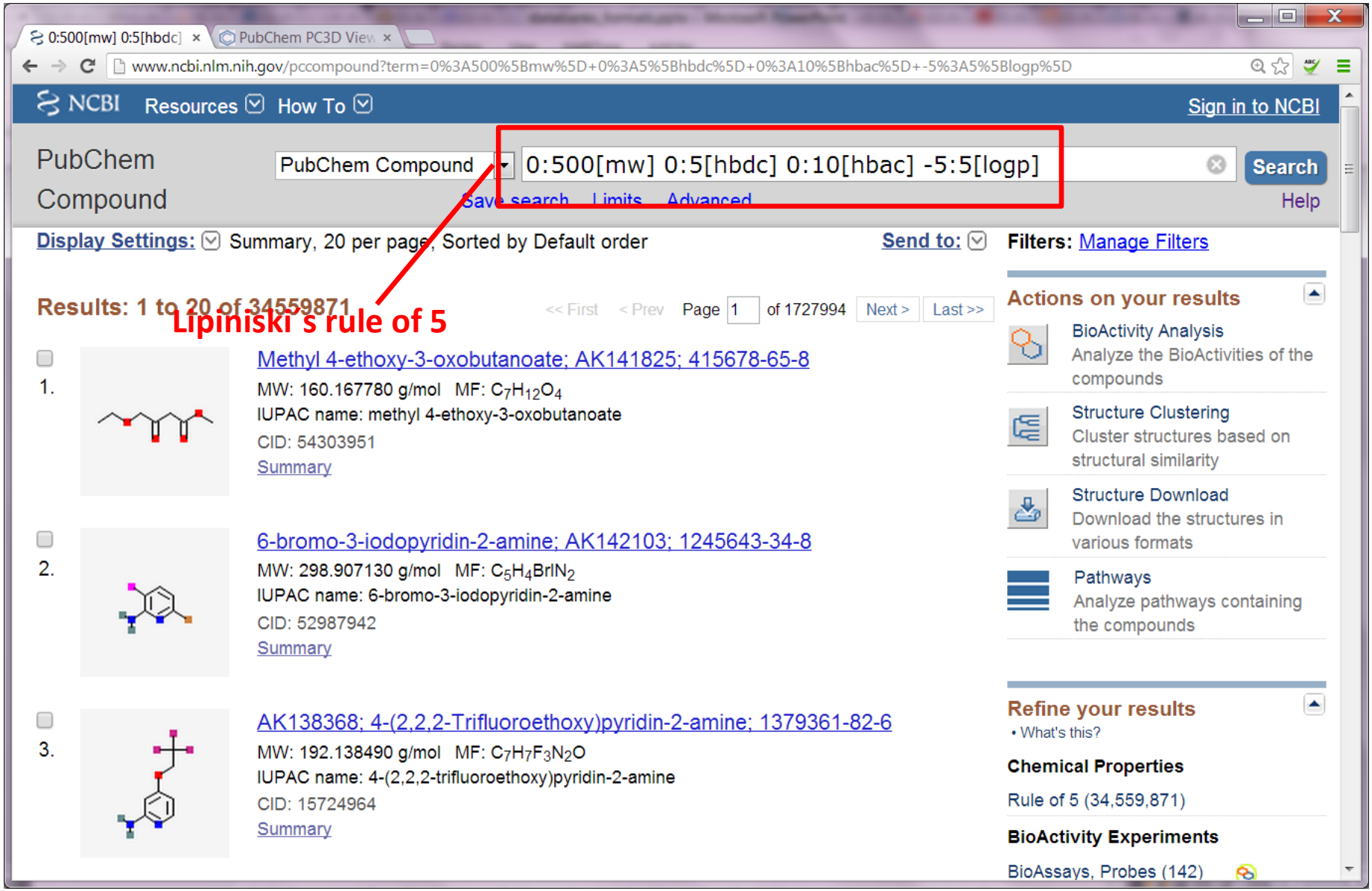

PubChem
pubchem.ncbi.nlm.nih.gov/#query=aspirin&tab=bioassay
Chrome
Ualg
Thank you Paulo! - C...
NAS
Tools
Resources
Covid-19
Modelling
Dali Server - Job que...
Other bookmarks
Reading list
PubChem
aspirin
Compounds (120)
Substances (615)
Pathways (25)
BioAssays (1,998)
Literature (69,504)
Patents (2,134)
Searching descriptions and metadata of bioassay records submitted by PubChem's contributors. Read More...
1,998 results
Filters
SORT BY
Relevance
Download
Search in Entrez
ACTIONS ON RESULTS WITH ID TYPE:
BioAssays
Substances
Compounds
Push to Entrez
Save for Later
Linked Data Sets
Percentage of aspirin formed during hydrolysis in 10% human plasma at pH 7.4 at 37 degrees celsius
BioAssay AID: 15845 BioAssay Type: Literature-derived
Tested Compounds Count: 27 Tested Substances Count: 27
Data Source: ChEMBL External ID: CHEMBL875111
Data Source Category: Curation Efforts; Research and Development
Modified Date: 2018-10-12
Description: Title: Evaluation of glycolamide esters and various other esters of aspirin as true aspirin prodrugs.||Abstract: A series of glycolamide, glycolate, (acyloxy)methyl, alkyl, and aryl esters of acetylsalicylic acid (aspirin) were synthesized and evaluated as potential prodrug forms of aspirin. N,N-Disubstituted glycolamide esters were found to be rapidly hydrolyzed in human plasma, resulting in the formation of aspirin as well as the corresponding salicylate esters. These in turn hydrolyzed rapidly to salicylic acid. The largest amount of aspirin formed from the esters were 50 and 55% in case of the N,N-dimethyl- and N,N-diethylglycolamide esters, respectively. Similar results were obtained in blood with the N,N-dimethyl- and N,N-diethylglycolamide esters. Unsubstituted and monosubstituted glycolamide esters as well as most other esters previously suggested to be aspirin prodrugs were shown to hydrolyze exclusively to the corresponding salicylic acid esters. Lipophilicity parameters and water solubilities of the esters were determined, and structural factors favoring ester prodrug hydrolysis at the expense of deacetylation to yield salicylate ester are discussed. The properties of some N,N-disubstituted glycolamide esters of aspirin are highlighted with respect to their use as potential aspirin prodrugs.
The percentage of aspirin formed during hydrolysis in 100% human plasma at pH 7.4 at 37 degrees celsius
BioAssay AID: 15847 BioAssay Type: Literature-derived
Tested Compounds Count: 1 Tested Substances Count: 1
Data Source: ChEMBL External ID: CHEMBL629204
Data Source Category: Curation Efforts; Research and Development

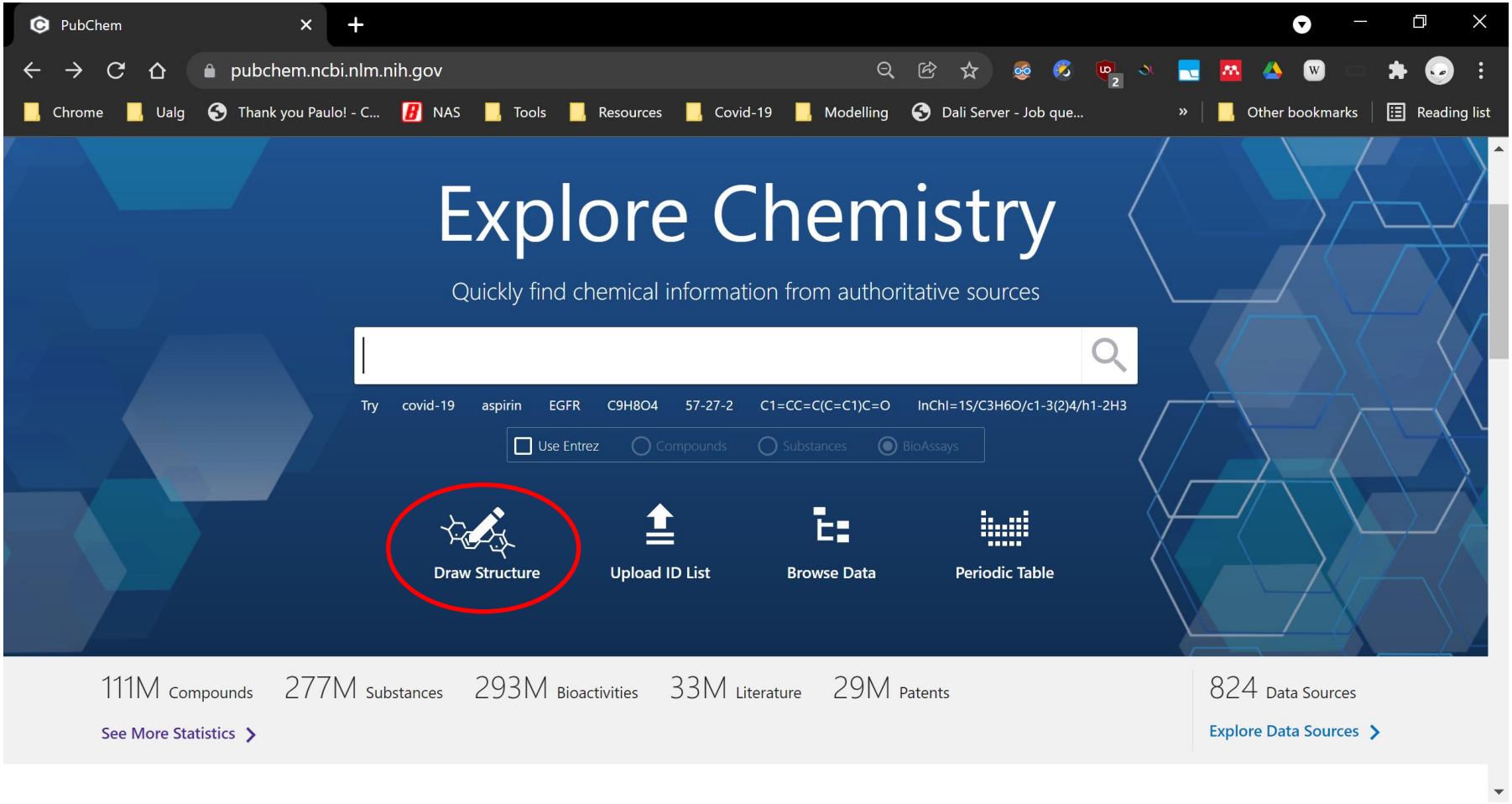

AID 15845 - Percentage of aspirin fo
pubchem.ncbi.nlm.nih.gov/bioassay/15845
Chrome
Ualg
Thank you Paulo! - C...
NAS
Tools
Resources
Covid-19
Modelling
Dali Server - Job que...
Other bookmarks
Reading list
BIOASSAY RECORD
Percentage of aspirin formed during hydrolysis in 10% human plasma at pH 7.4 at 37 degrees celsius
PubChem AID 15845
Primary Citation Evaluation of glycolamide esters and various other esters of aspirin as true aspirin prodrugs [PMID: 2918521]
Source ChEMBL
External ID CHEMBL875111
Tested Substances All (27) Data Table
Tested Compounds All (27)
Version 4.1 Revision History
Status Live
Dates Modify 2018-10-12 Deposit 2010-05-21
This bioassay record (AID 15845) reports results from the above primary citation. Additional data from the same publication are reported in a total of 19 BioAssay records in PubChem.
PubChem
Cite
Download
CONTENTS
Title and Summary
1 Description
2 Comment
3 Result Definitions
4 Data Table
5 Target
6 Entrez Crosslinks
7 Identity
8 Same-Publication BioAssays
9 BioAssay Annotations
10 Information Sources
1 Description
Title: Evaluation of glycolamide esters and various other esters of aspirin as true aspirin prodrugs.
Abstract: A series of glycolamide, glycolate, (acyloxy)methyl, alkyl, and aryl esters of acetylsalicylic acid (aspirin) were synthesized and evaluated as potential prodrug forms of aspirin. N,N-Disubstituted glycolamide esters were found to be rapidly hydrolyzed in human plasma, resulting in the formation of aspirin as well as the corresponding salicylate esters. These in turn hydrolyzed rapidly to salicylic acid. The largest amount of aspirin formed from the esters were 50 and 55% in case of the N,N-dimethyl- and N,N-diethylglycolamide esters, respectively. Similar results were obtained in blood with the N,N-dimethyl- and N,N-diethylglycolamide esters. Unsubstituted and monosubstituted glycolamide esters as well as most other esters previously suggested to be aspirin prodrugs were shown to hydrolyze exclusively to the corresponding salicylic acid esters. Lipophilicity parameters and water solubilities of the esters were determined, and structural factors favoring ester prodrug hydrolysis at the expense of deacetylation to yield salicylate ester are discussed. The properties of some N,N-disubstituted glycolamide esters of aspirin are highlighted with respect to their use as potential aspirin prodrugs.
PubChem

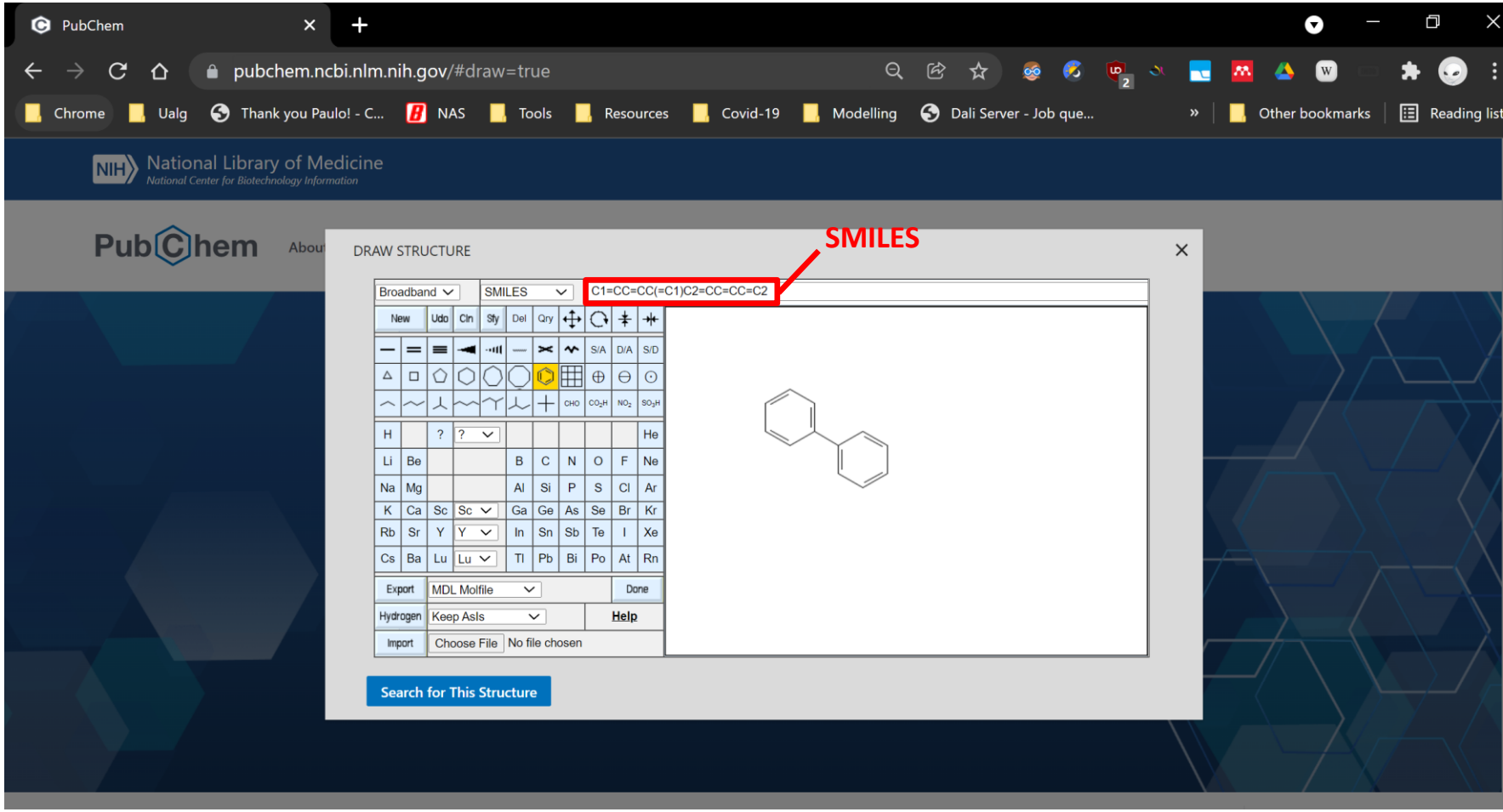
PubChem BioAssay



PubChem – Pesquisa por “Tag”




PubChem
pubchem.ncbi.nlm.nih.gov
Chrome
Ualg
Thank you Paulo! - C...
NAS
Tools
Resources
Covid-19
Modelling
Dali Server - Job que...
Other bookmarks
Reading list
Explore Chemistry
Quickly find chemical information from authoritative sources
Try covid-19 aspirin EGFR C9H8O4 57-27-2 C1=CC=C(C=C1)C=O InChI=1S/C3H6O/c1-3(2)4/h1-2H3
Use Entrez
Compounds
Substances
BioAssays
Draw Structure
Upload ID List
Browse Data
Periodic Table
111M Compounds
277M Substances
293M Bioactivities
33M Literature
29M Patents
See More Statistics
824 Data Sources
Explore Data Sources

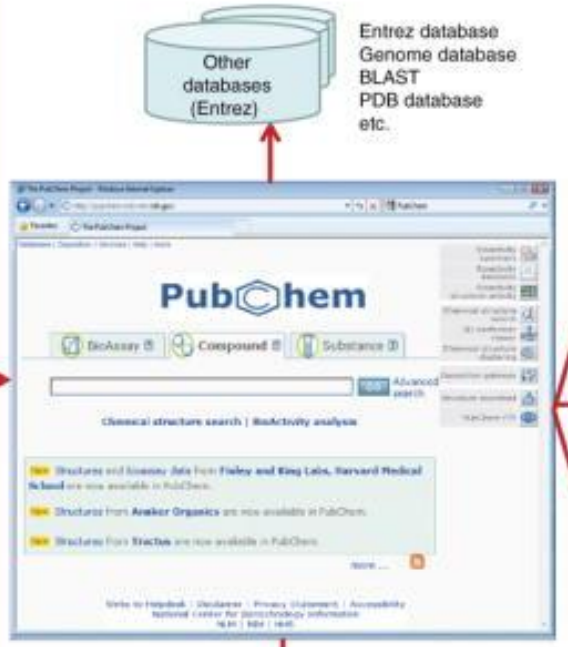

PubChem
pubchem.ncbi.nlm.nih.gov/#draw=true
Chrome
Ualg
Thank you Paulo! - C...
NAS
Tools
Resources
Covid-19
Modelling
Dali Server - Job que...
Other bookmarks
Reading list
National Library of Medicine
National Center for Biotechnology Information
PubChem
DRAW STRUCTURE
SMILES
C1=CC=CC(=C1)C2=CC=CC=C2
Broadband
SMILES
New
Udo
Cln
Sty
Del
Qry
Export
MDL Molfile
Done
Hydrogen
Keep Asls
Help
Import
Choose File
No file chosen
Search for This Structure

A.

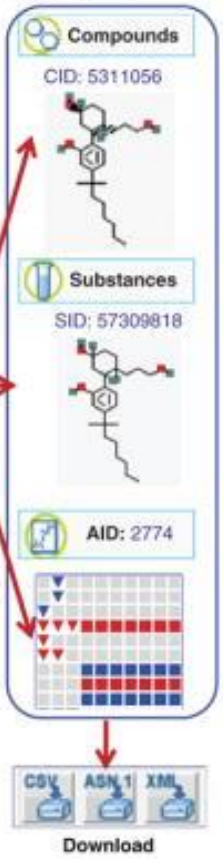
Bioassay Data Source Name	Bioassay count	Substance count
BioAssay Data Deposited by NIH MLPPCN and MLSCN		
NCGC (NIH)	485	398,461
The Scripps Research Institute Molecular Screening Center	483	357,929
Burnham Center for Chemical Genomics	397	400,255
NMMLSC (University of Mexico)	230	348,231
Broad Institute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

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

B.



C.



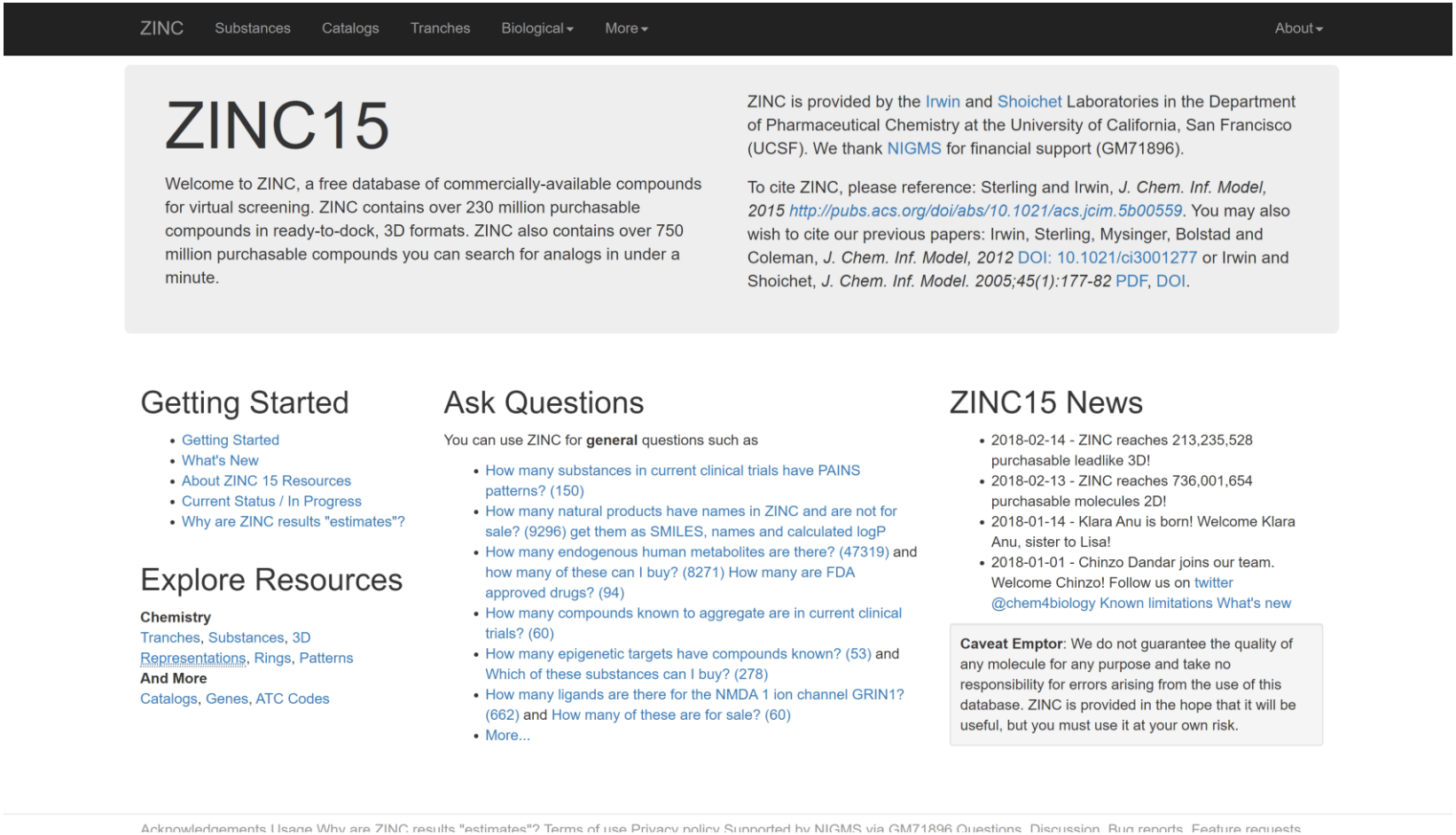
ZINC database

- Base de dados de acesso livre
- Contem cerca de 230 milhões de compostos comercialmente disponíveis (purchasable compounds) com as respectivas estruturas 3D em formatos de fácil uso para docking e screening virtual
- Contem cerca de 740 milhões de compostos comercialmente disponíveis com estruturas 2D, que podem ser usados para pesquisar análogos.
- Possui alguns sub-conjuntos especiais:
    - ZDD – compostos puros aprovados como fármacos pela FDA
    - ZMD – metabolitos primários
    - ZND – derivados de compostos naturais
    - ZBC – compostos biogénicos

    …..

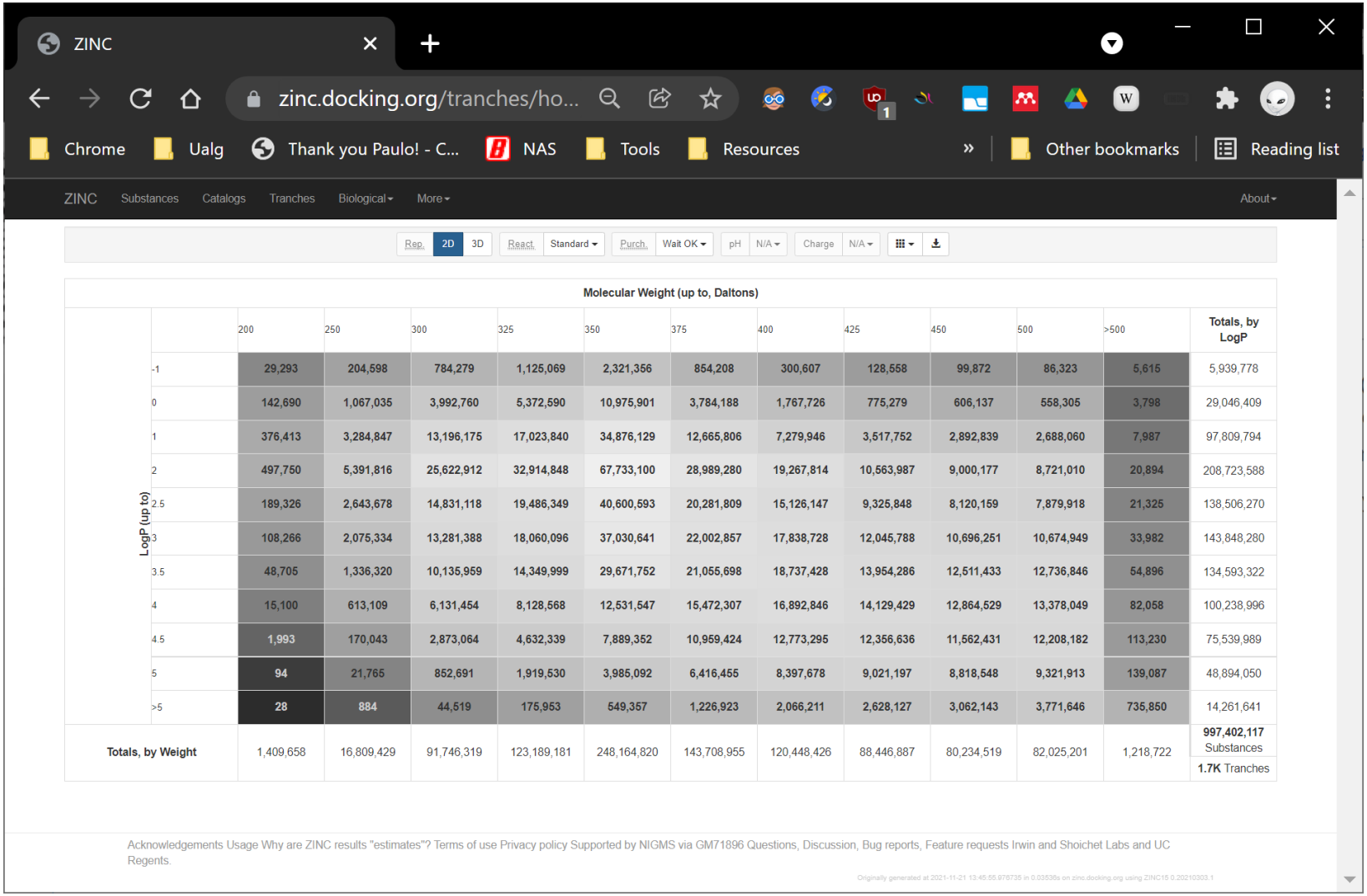
    ….

ZINC database



http://zinc.docking.org/

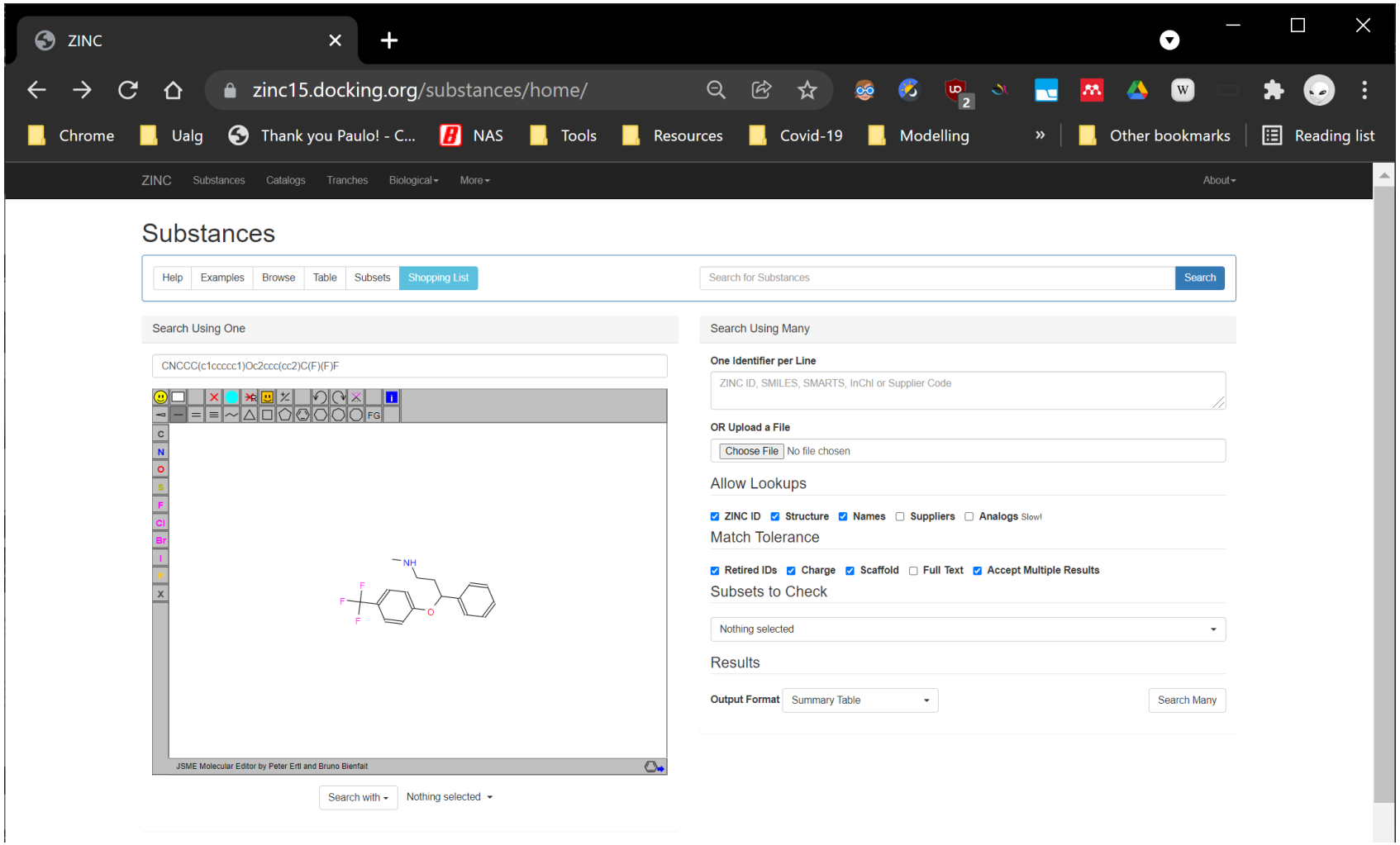
ZINC tranches



Molecular Weight (up to, Daltons)

LogP (up to)	200	250	300	325	350	375	400	425	450	500	>500	Totals, by LogP
-1	29,293	204,598	784,279	1,125,069	2,321,356	854,208	300,607	128,558	99,872	86,323	5,615	5,939,778
0	142,690	1,067,035	3,992,760	5,372,590	10,975,901	3,784,188	1,767,726	775,279	606,137	558,305	3,798	29,046,409
1	376,413	3,284,847	13,196,175	17,023,840	34,876,129	12,665,806	7,279,946	3,517,752	2,892,839	2,688,060	7,987	97,809,794
2	497,750	5,391,816	25,622,912	32,914,848	67,733,100	28,989,280	19,267,814	10,563,987	9,000,177	8,721,010	20,894	208,723,588
2.5	189,326	2,643,678	14,831,118	19,486,349	40,600,593	20,281,809	15,126,147	9,325,848	8,120,159	7,879,918	21,325	138,506,270
3	108,266	2,075,334	13,281,388	18,060,096	37,030,641	22,002,857	17,838,728	12,045,788	10,696,251	10,674,949	33,982	143,848,280
3.5	48,705	1,336,320	10,135,959	14,349,999	29,671,752	21,055,698	18,737,428	13,954,286	12,511,433	12,736,846	54,896	134,593,322
4	15,100	613,109	6,131,454	8,128,568	12,531,547	15,472,307	16,892,846	14,129,429	12,864,529	13,378,049	82,058	100,238,996
4.5	1,993	170,043	2,873,064	4,632,339	7,889,352	10,959,424	12,773,295	12,356,636	11,562,431	12,208,182	113,230	75,539,989
5	94	21,765	852,691	1,919,530	3,985,092	6,416,455	8,397,678	9,021,197	8,818,548	9,321,913	139,087	48,894,050
>5	28	884	44,519	175,953	549,357	1,226,923	2,066,211	2,628,127	3,062,143	3,771,646	735,850	14,261,641
Totals, by Weight	1,409,658	16,809,429	91,746,319	123,189,181	248,164,820	143,708,955	120,448,426	88,446,887	80,234,519	82,025,201	1,218,722	**997,402,117** Substances **1.7K** Tranches

ZINC substance search



ZINC

zinc15.docking.org/substances/resolved/

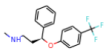
Chrome | Ualg | Thank you Paulo! - C... | NAS | Tools | Resources | Covid-19 | Modelling | Other bookmarks | Reading list

ZINC Substances Catalogs Tranches Biological More About

« 1 » 2 / substances Filters Lookup

ZINC	Summary

ZINC1530637



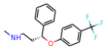
1
CNCCC(c1ccccc1)O...
InChIKey:Same-Sca...

fluoxetine

Tranche DIAA
Subsets anodyne bb fda for-sale in-stock natural-products

Purchasability In-Stock (145 vendors, 82 annotated)
Activity 18 activities from 156 observation(s)
in 5 class(es) with 17 gene(s)
Studies 74 citations
186 clinical trial(s)

ZINC1530638



1
CNCCC(c1ccccc1)O...
InChIKey:Same-Sca...

fluoxetine

Tranche DIAA
Subsets anodyne bb fda for-sale in-stock

Purchasability In-Stock (146 vendors, 80 annotated)
Activity 18 activities from 157 observation(s)
in 5 class(es) with 17 gene(s)
Studies 75 citations
186 clinical trial(s)

« 1 » 2 / substances Filters Lookup

Acknowledgements Usage Why are ZINC results "estimates"? Terms of use Privacy policy Supported by NIGMS via GM71896 Questions, Discussion, Bug reports, Feature requests Irwin and Shoichet Labs and UC Regents.

ZINC1530637 (Fluoxetine)

zinc15.docking.org/substances/ZINC000001530...

Chrome | Ualg | Thank you Paulo! - C... | NAS | Tools | Resources | Covid-19 | Modelling | Other bookmarks | Reading list

ZINC Substances Catalogs Tranches Biological More About

/ substances / ZINC000001530637

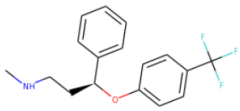
ZINC1530637 (Fluoxetine)

In: anodyne bb fda for-sale in-stock natural-products

Google Wikipedia PubMed

Added	Availability	Since	Mwt	logP	Download
2004-10-06	In-Stock	2015-08-07	309.331	4.435	

Mol Formula	Rings	Heavy Atoms	Hetero Atoms	Fraction sp^3	Tranche
C17H18F3NO	2	22	5	0.29	DIAA



Draw

SMILES	CNCC[C@H](Oc1ccc(C(F)(F)F)cc1)c1ccccc1
InChI	InChI=1S/C17H18F3NO/c1-21-12-11-16(13-5-3-2-4-6-13)22-15-9-7-14(8-10-15)17(18,19)20/h2-10,16,21H,11-12H2,1H3/t16-/m0/s1
InChI Key	RTHCYVBBDHJXIQ-INIZCTEOSA-N

Available 3D Representations

pH range	Net charge	H-bond donors	H-bond acceptors	tPSA	Rotatable bonds	Apolar desolvation	Polar desolvation	Download
Reference	1	1	1	25	6	9.47	-44.22	

Vendors (79 Total) — 145 Items Total

AK Scientific Economical	J10277
Chemodex Economical	F0141
Chem Scene Economical	CS-1838
KeyOrganics Bioactives	KS-1061
MedChem Express Economical	HY-B0102A
Molport SC Economical	MolPort-001-683-482, MolPort-003-666-535, MolPort-009-194-198
Oakwood Economical	375072

Annotated Catalogs (44 Total) — 82 Items Total

NIH Clinical Collection	MLS002589965
Prestwick Chemical	Prestw-511
SMDC Iconix	131498
SMDC Pharmakon	131498
Tocriscreen	0927
Ambinter	Amb17614490, Amb2608851, Amb534674, Amb6297426
BindingDB.org	30130, 50331514, 81875
ChEBI	CHEBI:86990, CHEBI:86992, CHEBI:86995, CHEBI:86997
ChEMBL20	CHEMBL1169388, CHEMBL1201082, CHEMBL1257031, CHEMBL41

ZINC53 (Aspirin)

zinc15.docking.org/substanc...

Chrome | Ualg | Thank you Paulo! - C... | NAS | Tools | Resources | Other bookmarks | Reading list

ZINC | Substances | Catalogs | Tranches | Biological | More | About

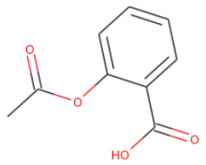
/ substances / ZINC000000000053

ZINC53 (Aspirin)

In: anodyne bb fda for-sale in-stock natural-products

Google Wikipedia PubMed

Added	Availability	Since	Mwt	logP	Download
2005-09-27	In-Stock	2015-08-07	180.159	1.31	



Draw

Mol Formula	Rings	Heavy Atoms	Hetero Atoms	Fraction sp^3	Tranche
C9H8O4	1	13	4	0.11	ADAA

SMILES	CC(=O)Oc1ccccc1C(=O)O
InChI	InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)
InChI Key	BSYNRYMUTXBXSQ-UHFFFAOYSA-N

Available 3D Representations

pH range	Net charge	H-bond donors	H-bond acceptors	tPSA	Rotatable bonds	Apolar desolvation	Polar desolvation	Download
Reference	-1	0	4	66	2	6.58	-56.82	

Traditional Chinese Medicine (TCM)

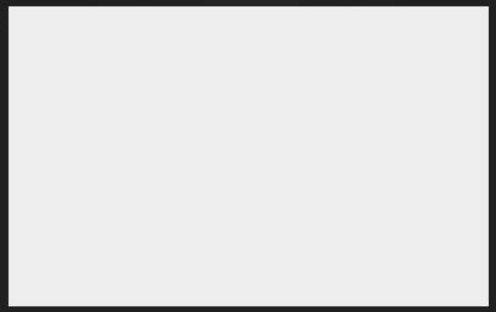
- Contem substâncias derivadas de plantas, extractos animais e minerais
- Estruturas tri-dimensionais de compostos presentes nos extractos
- Estruturas disponíveis em formatos 2D e 3D, pré-minimizadas e prontas para usar em docking e screening virtual
- Acesso Livre

化學成份 Chemical Compoound

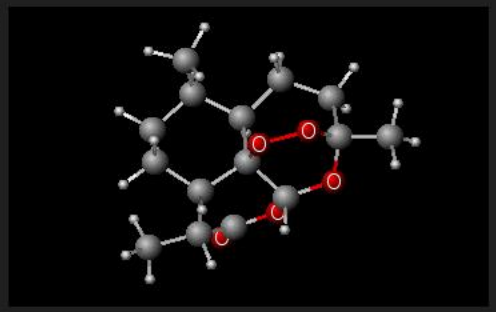
檔案下載 2D圖 MOL2

Compound artemisinin

2D結構圖



3D結構圖

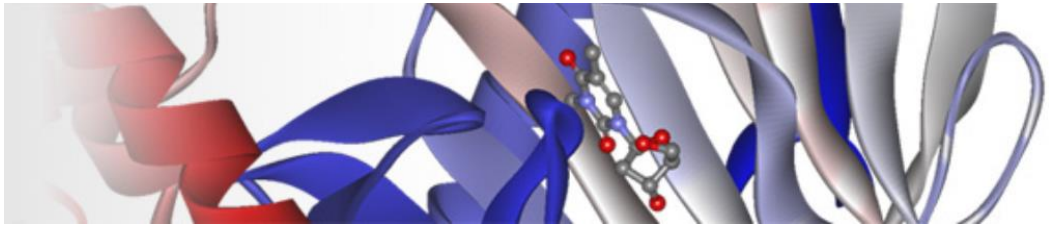


(Generate by marvinview)

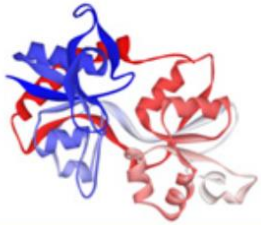
Chemical Formula	C15H22O5
Molecular Weight	282.332
Molecular Volume	197.91
ALogP	1.998
Molecular Polar Surface Area	53.99
Number of Hydrogen Bond Acceptors	0
Number of Hydrogen Bond Donors	0
Number of Rotatable Bonds	0

相關種名 青蒿 ;

檔案下載 2D圖 MOL2



iScreen



Upload User's Protein

TCM DataBase
(>20000 Compounds)

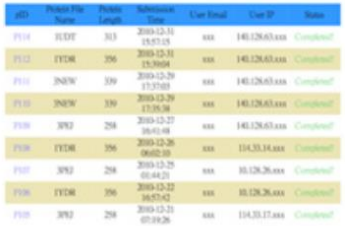
virtual screening

PLANTS Docking Module

De Novo

E-LEA3D Module

Web result page



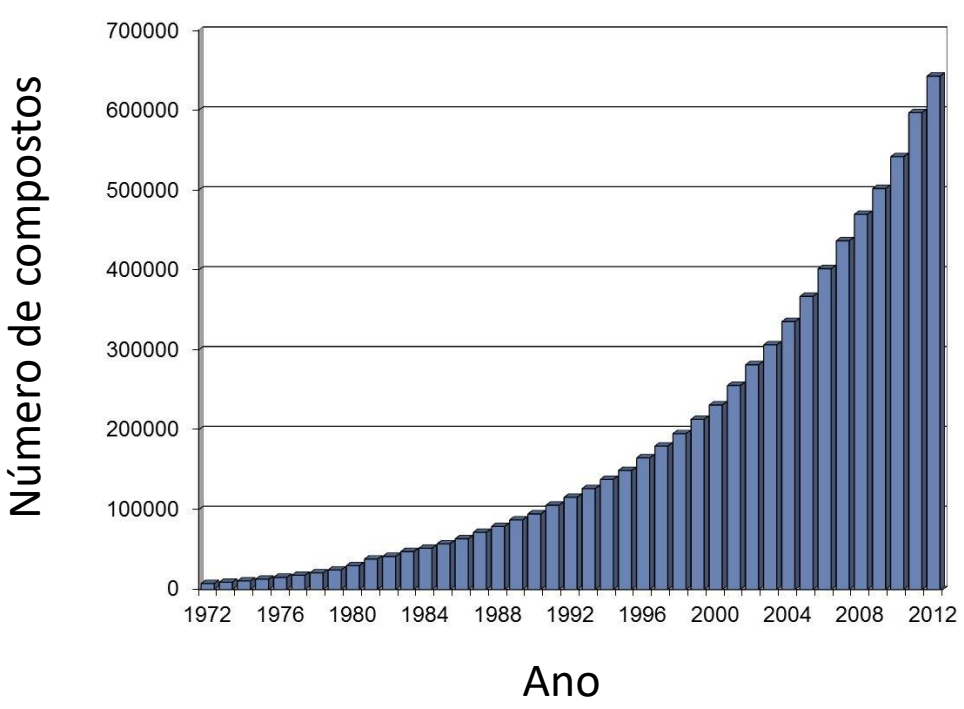
http://iscreen.cmu.edu.tw/intro.php

Cambridge Structural Database (CSD)

www.ccdc.cam.ac.uk

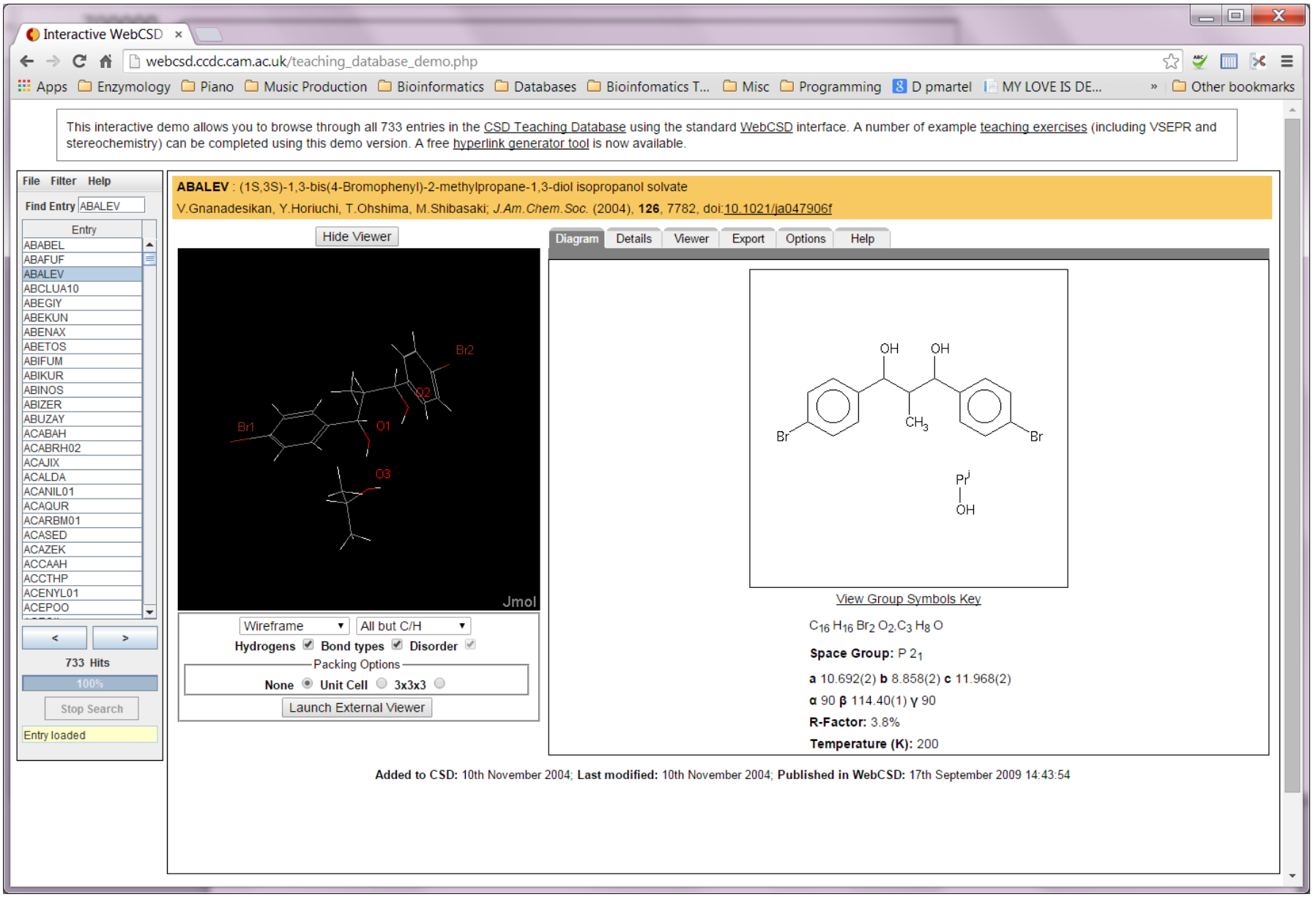
- Base de dados de estruturas **experimentais** de moléculas pequenas, mantida pelo Cambridge Crystallographic Data Center, UK
- Contem cerca de 700000 compostos **orgânicos** e **organometálicos** determinados por difracção de raios X e de neutrões em cristais individuais e pós.
- É um produto comercial, 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

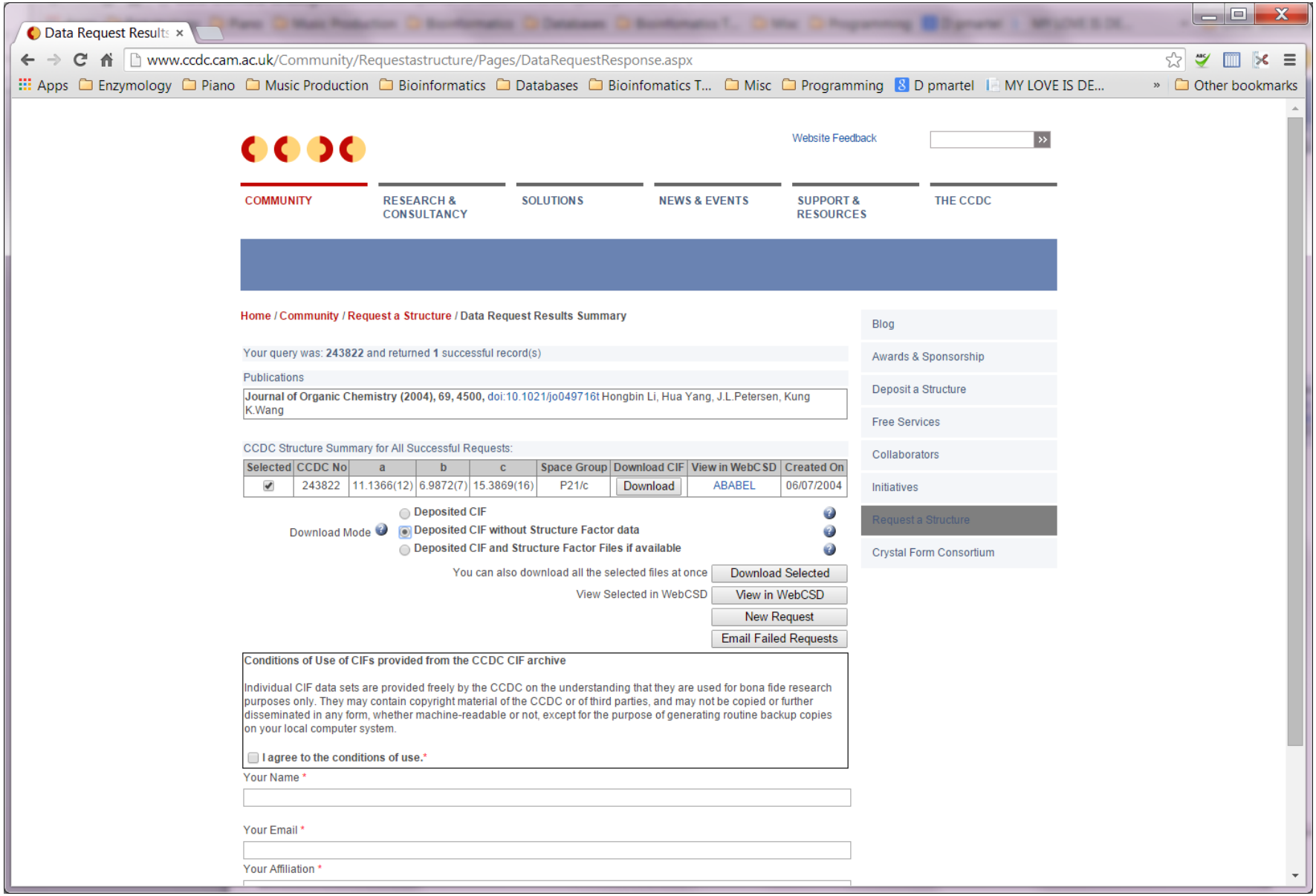


	Structures	%CSD
Total No. of structures	686 944	100.0
No. of different compounds	628 684	-
No. of literature sources	1 578	-
Organic structures	292 661	42.6
Transition metal present	369 682	53.8
Li – Fr or Be – Ra present	34 433	5.0
Main group metal present	41 711	6.1
3D coordinates present	643 032	93.3
Error-free coordinates	630 329	98.0†
Neutron studies	1 616	0.2
Powder diffraction studies	2 930	0.4
Low/high temp. studies	306809	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



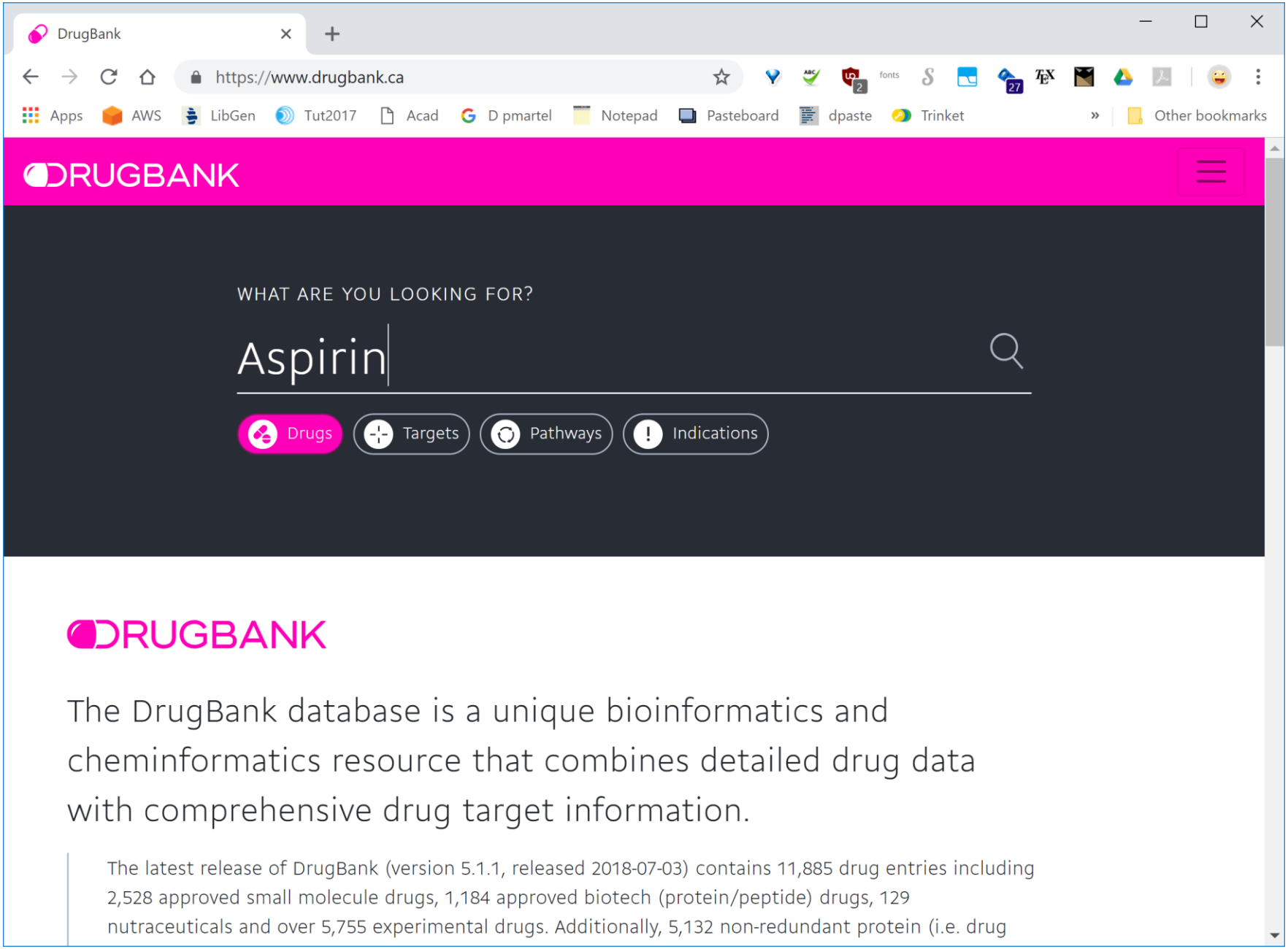
CSD – Pedido de estrutura



Drug Bank

- Base de dados bioinformática e cheminformática
- Version 5.1.4 (2019-07-02):
    - 13441 compostos
    - 2618 fármacos aprovados pela FDA
    - 1340 fármacos biológicas (proteínas/péptidos)
    - 130 nutraceutials
    - 6335 fármacos em fase experimental
    - 5157 proteínas (alvos/enzimas/transporters/carriers)

- Cada entrada (DrugCard) contém mais de 200 campos
- As entradas combinam informação sobre o fármaco (química, farmacológica e farmacêutica) com informação sobre o alvo (sequência, estrutura e via metabólica)


DrugBank
https://www.drugbank.ca
Apps
AWS
LibGen
Tut2017
Acad
D pmartel
Notepad
Pasteboard
dpaste
Trinket
Other bookmarks
DRUGBANK
WHAT ARE YOU LOOKING FOR?
Aspirin
Drugs
Targets
Pathways
Indications
DRUGBANK
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

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

DRUGBANK

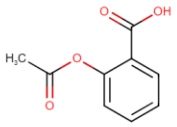
Browse Search Downloads About Help Blog Contact Us

Drugs

Acetylsalicylic acid

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

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	

3D Download Similar Structures

Synonyms 2-Acetoxybenzenecarboxylic acid

ChEMBL

• Base de dados mantida e curada manualmente pelo European Bioinformatics Institute (EBI), parte do European Molecular Biology Laboratory (EMBL).

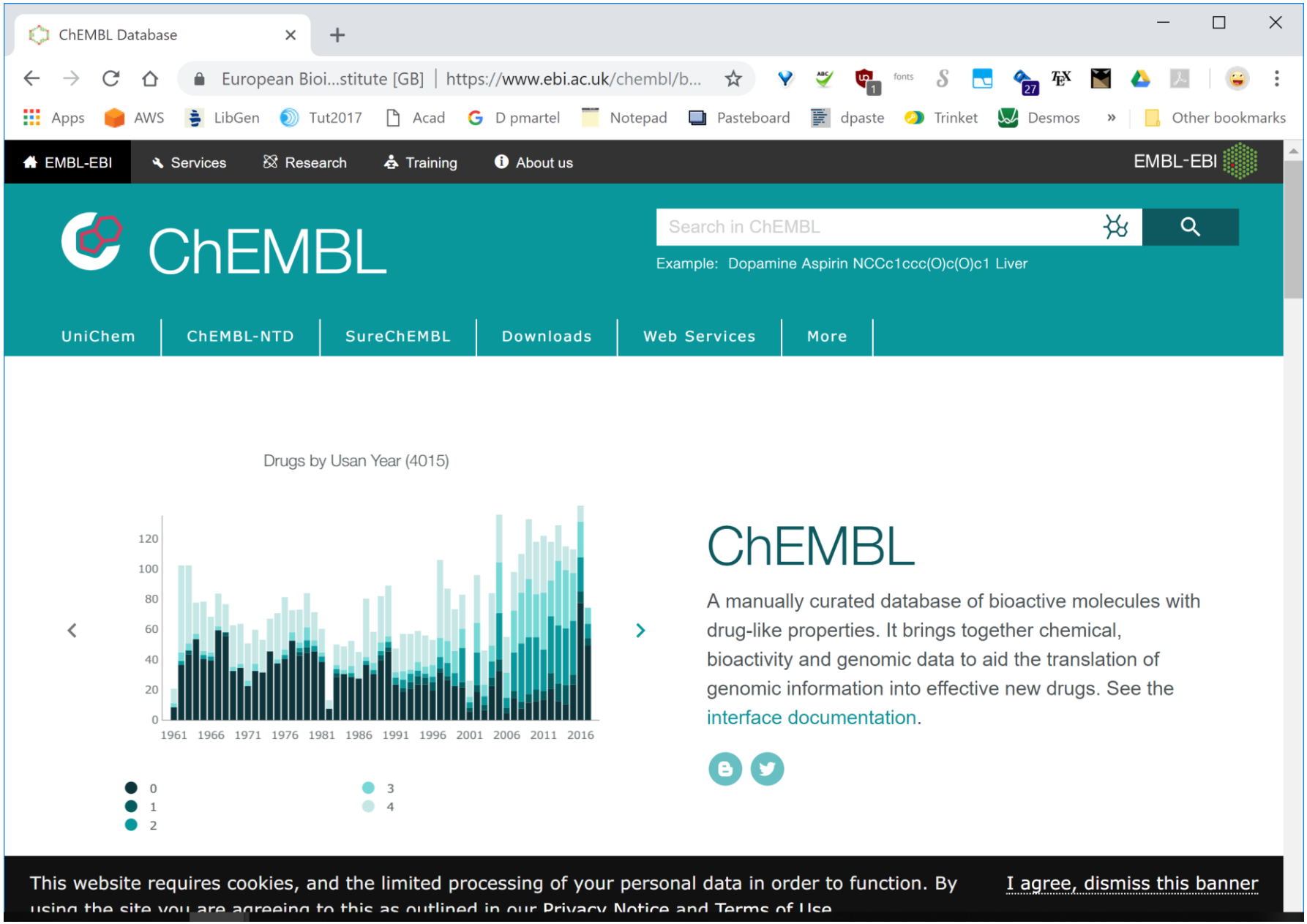
• Contem informação sobre a acção de compostos bioactivos em alvos farmacológicos (drug targets). A informação inclui Ki, Kd, IC50 e EC50.

• Entradas separadas para compostos e alvos.

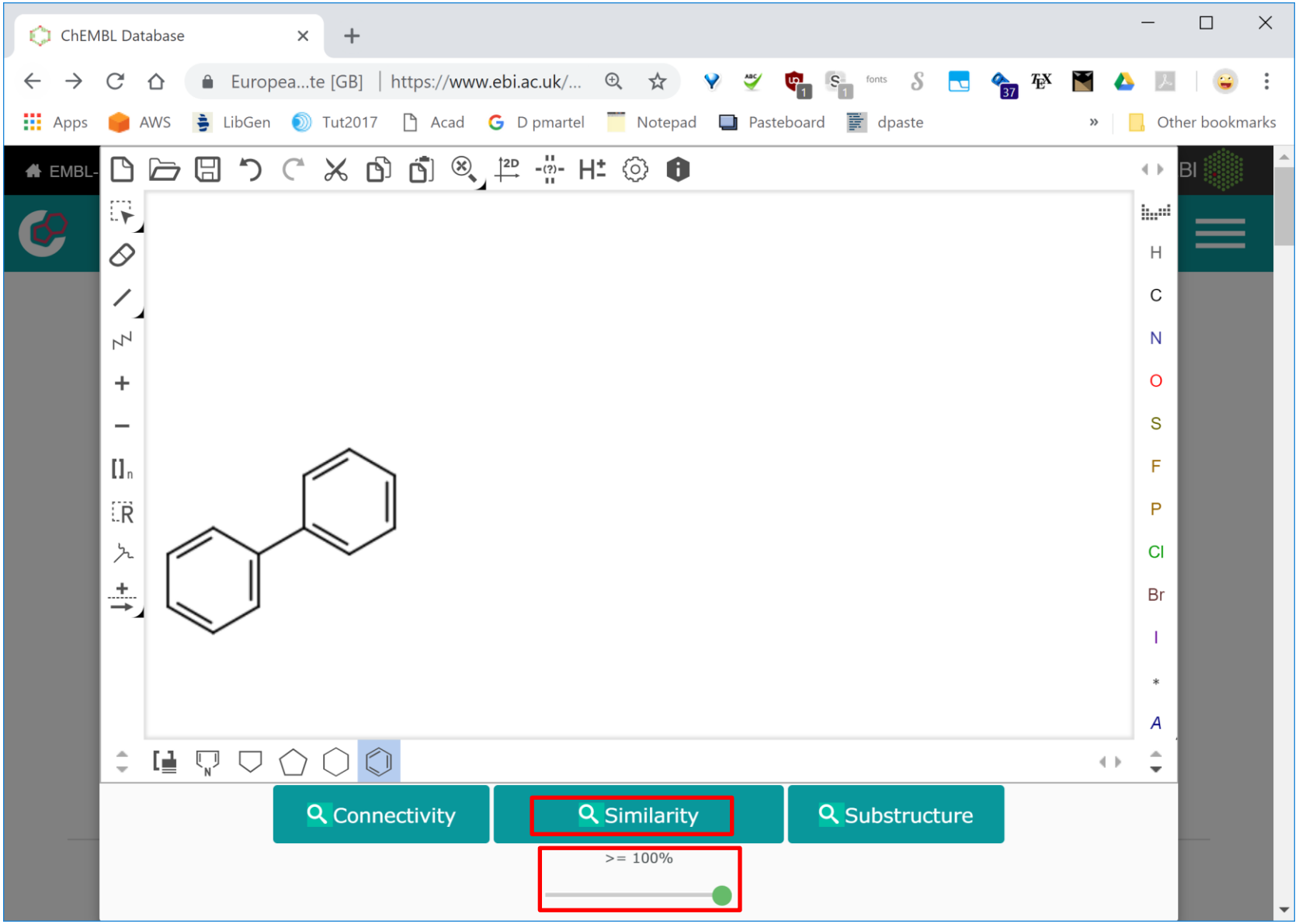
• A versão mais recente (v. 29, 1/07/2021) contém 2,105,464 compostos, 14,454 alvos e 18,635,916 ensaios de actividade derivados de 81,544 publicações.

• Contém uma série de ferramentas para análise e filtragem da informação contida na base de dados

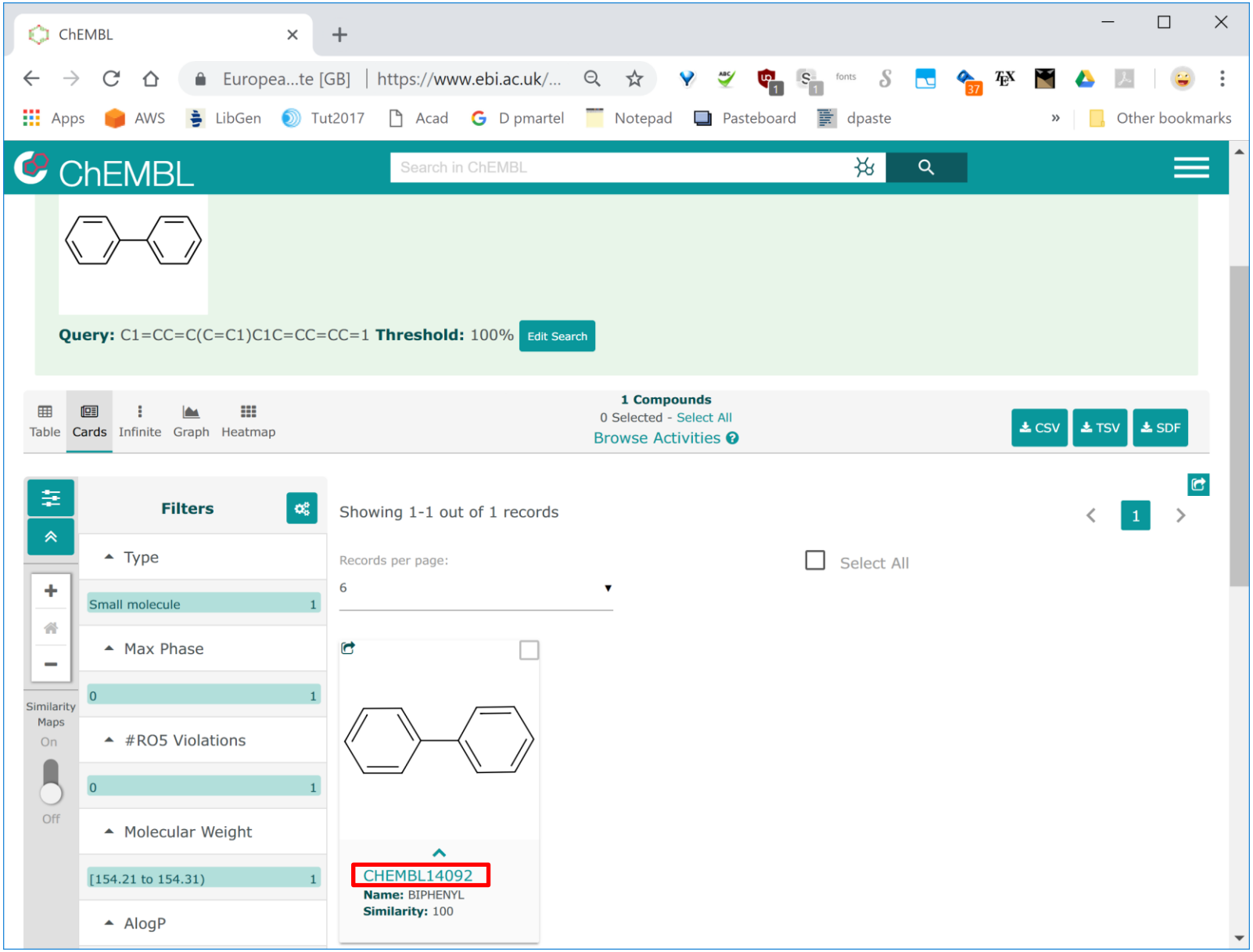
Exemplo de pesquisa estrutural em ChEMBL



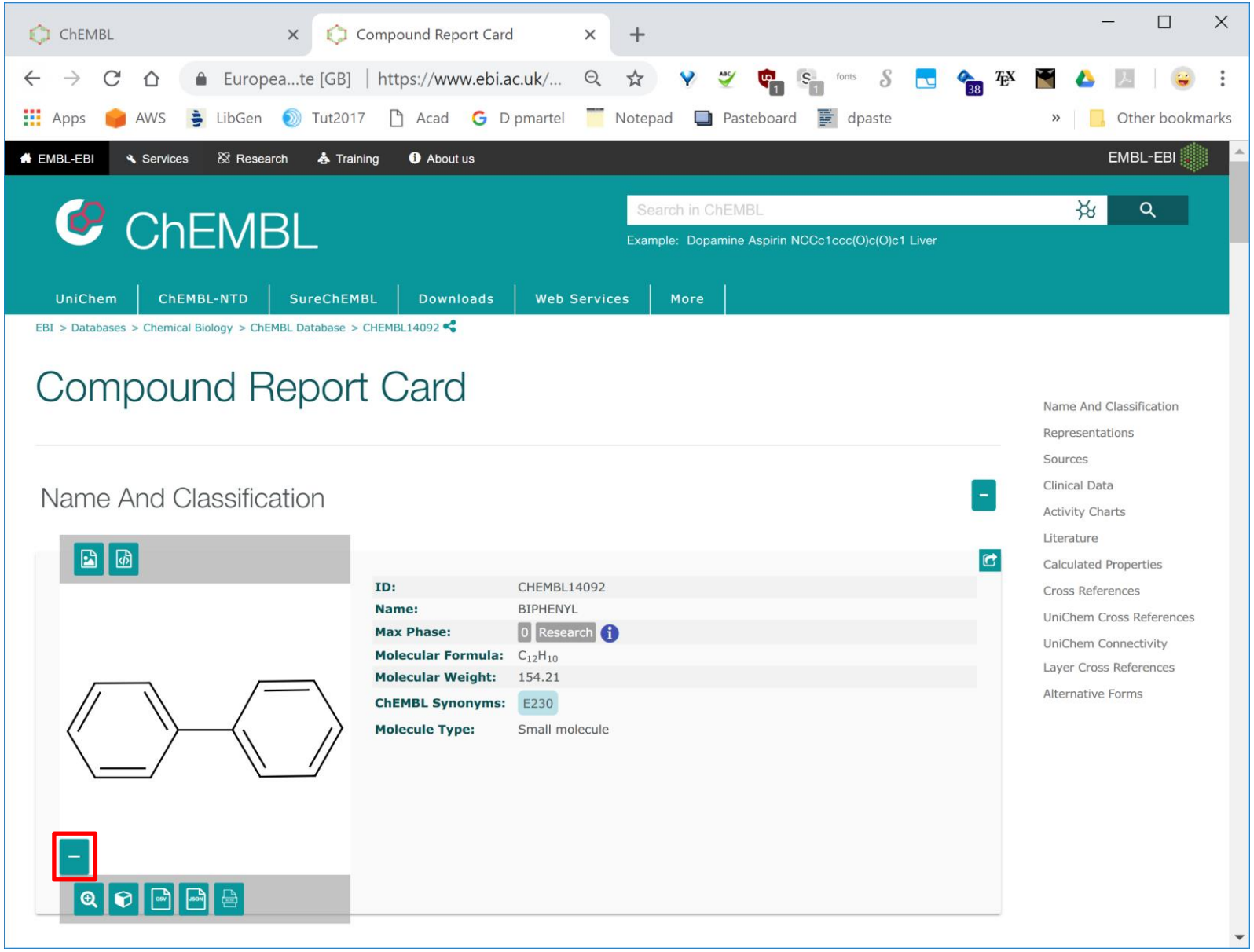
Exemplo de pesquisa estrutural em ChEMBL



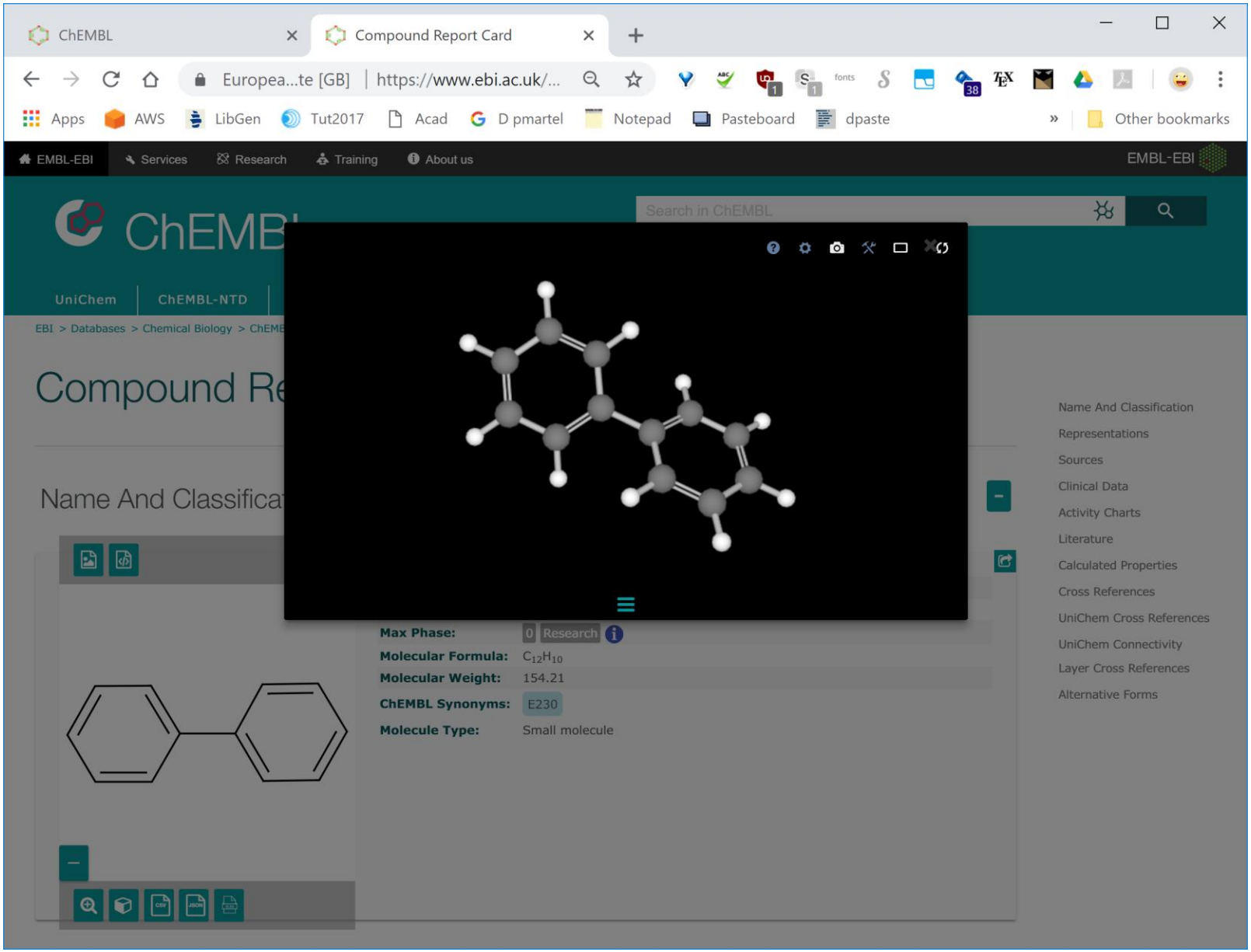
Exemplo de pesquisa estrutural em ChEMBL



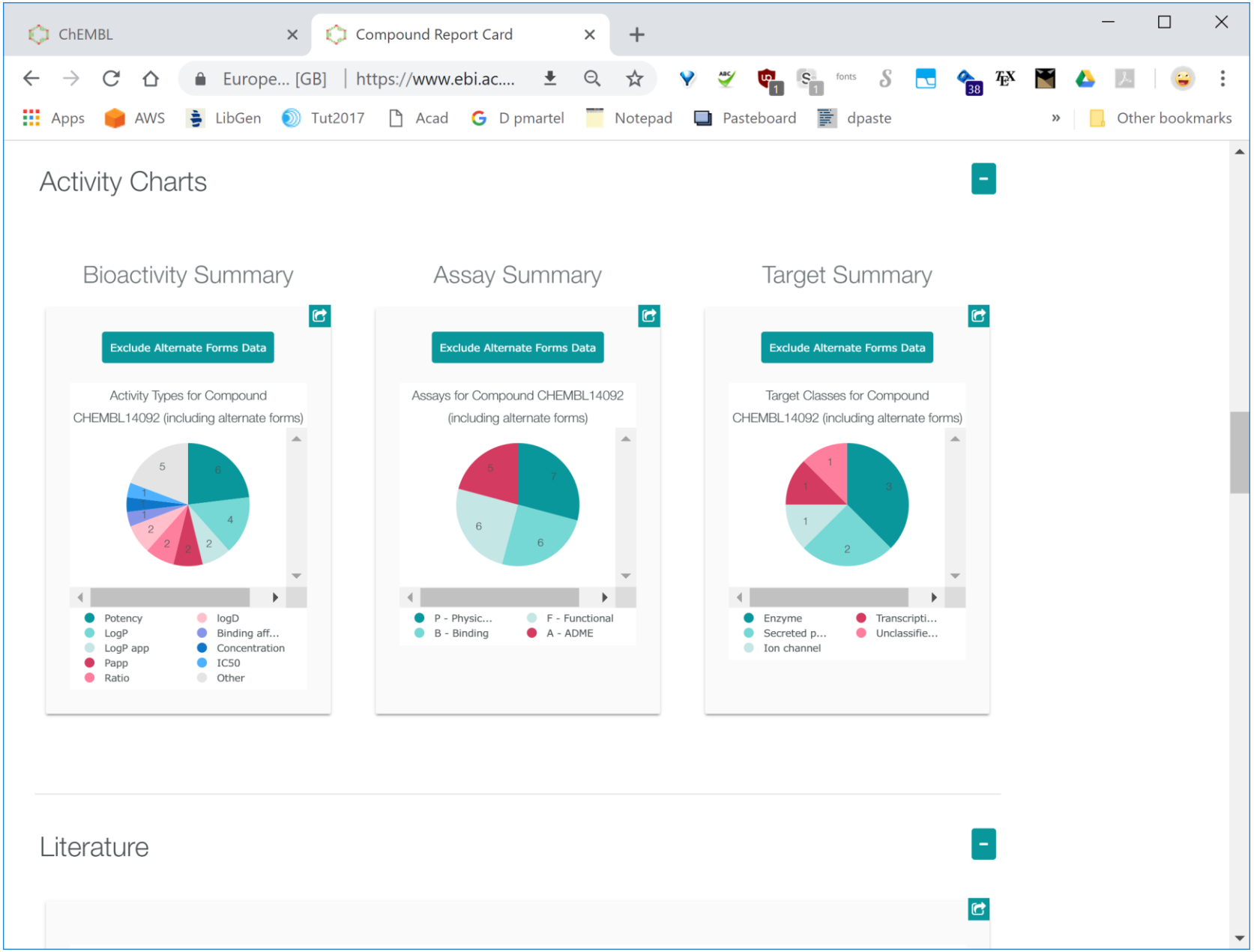
Exemplo de pesquisa estrutural em ChEMBL



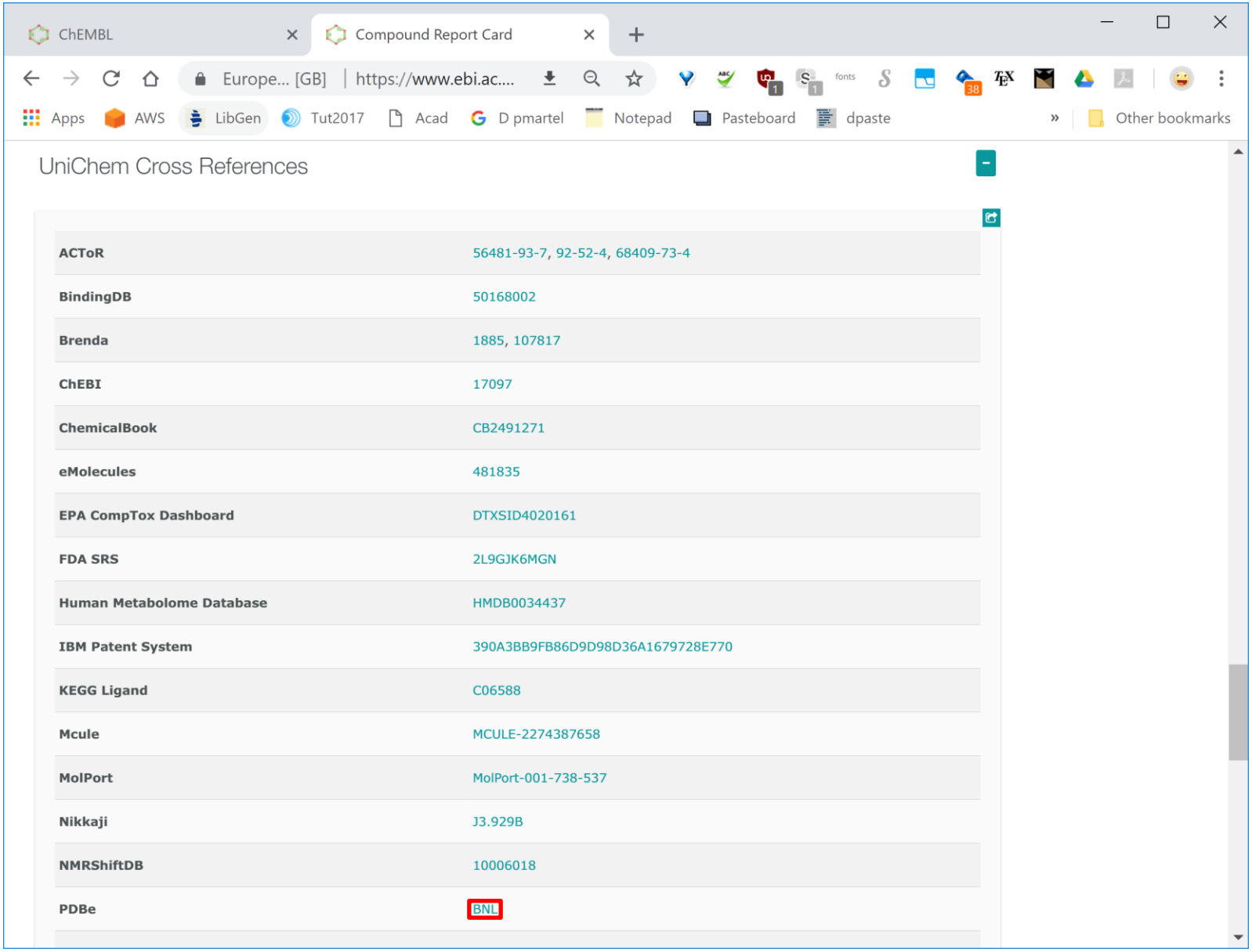
Exemplo de pesquisa estrutural em ChEMBL



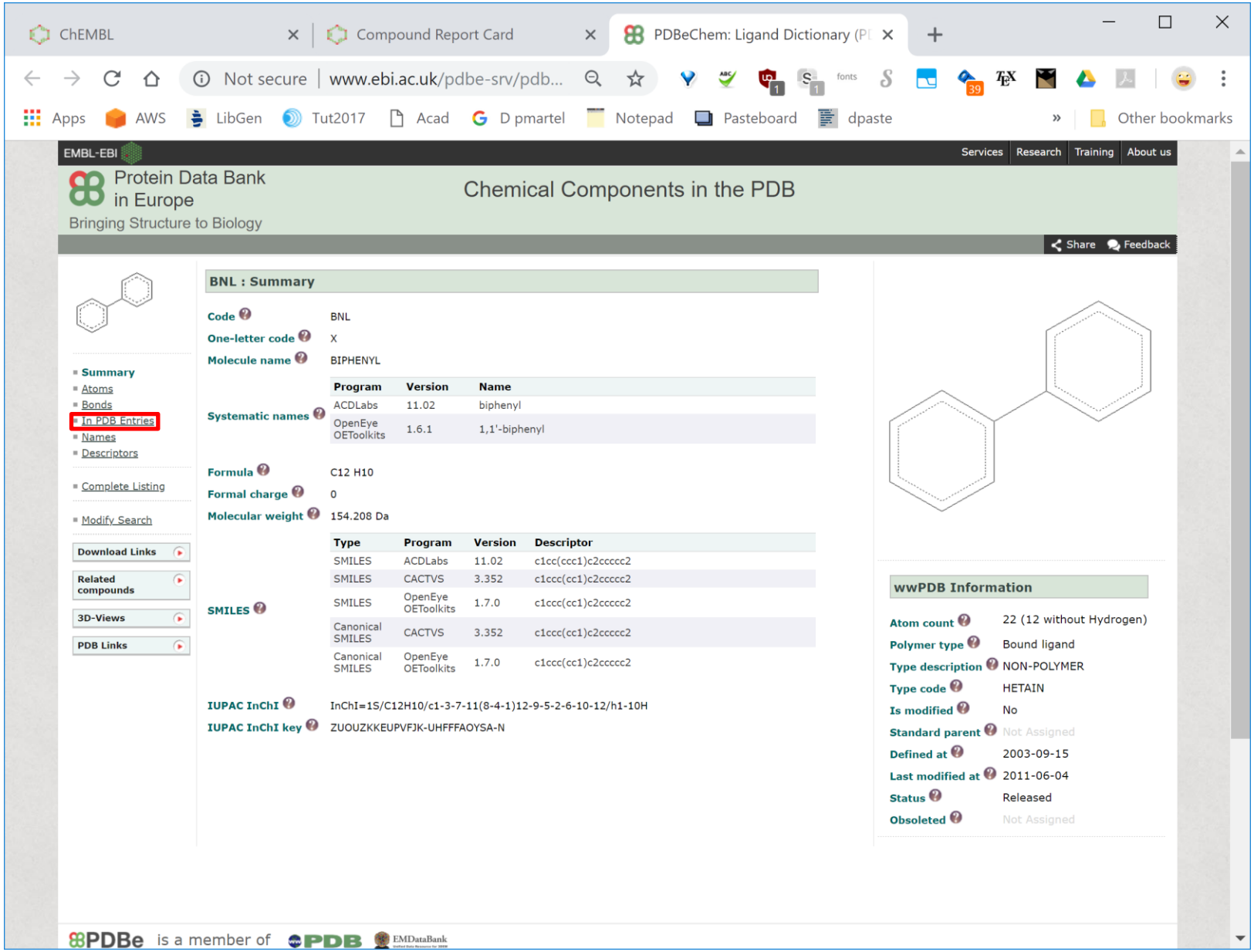
Exemplo de pesquisa estrutural em ChEMBL



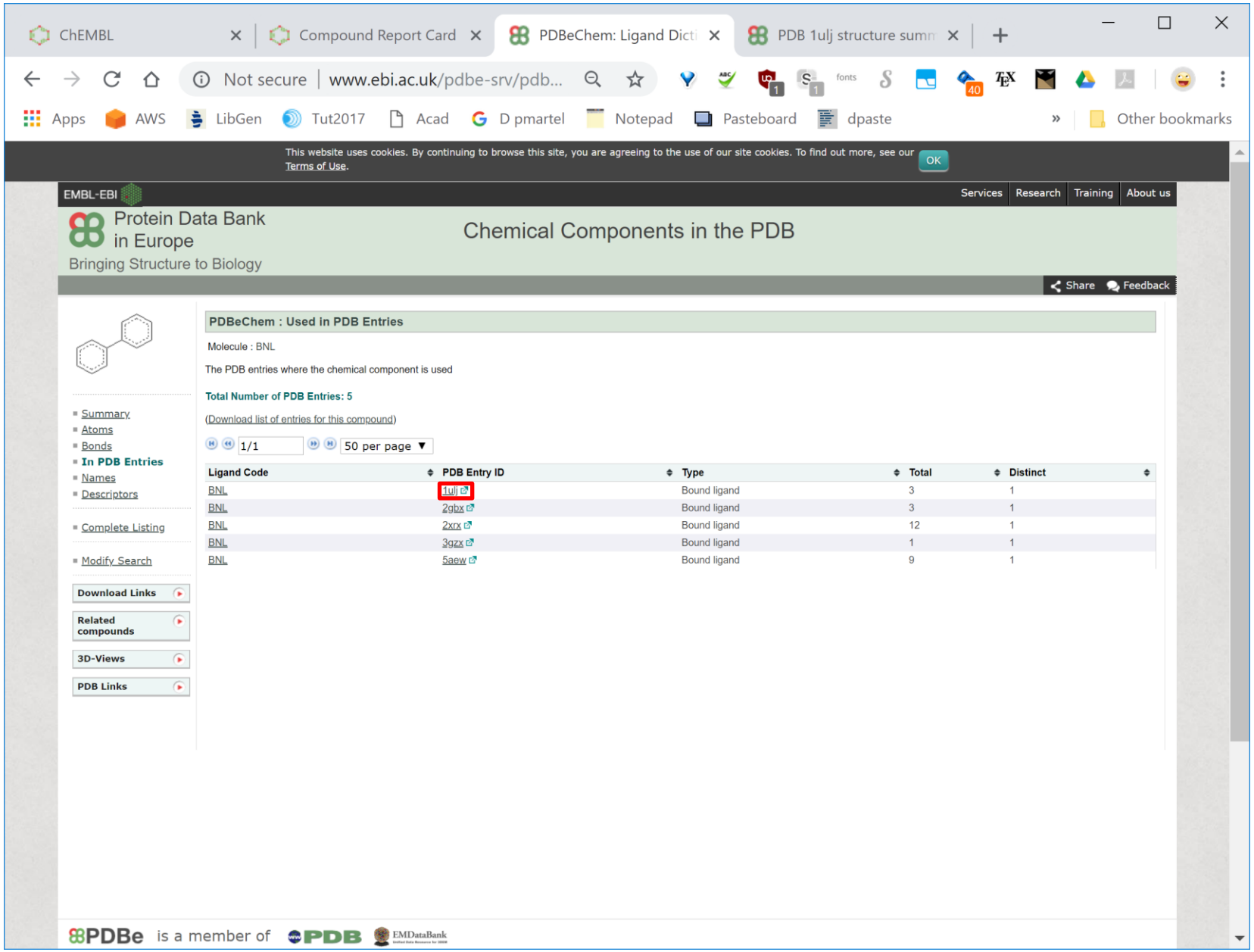
Exemplo de pesquisa estrutural em ChEMBL



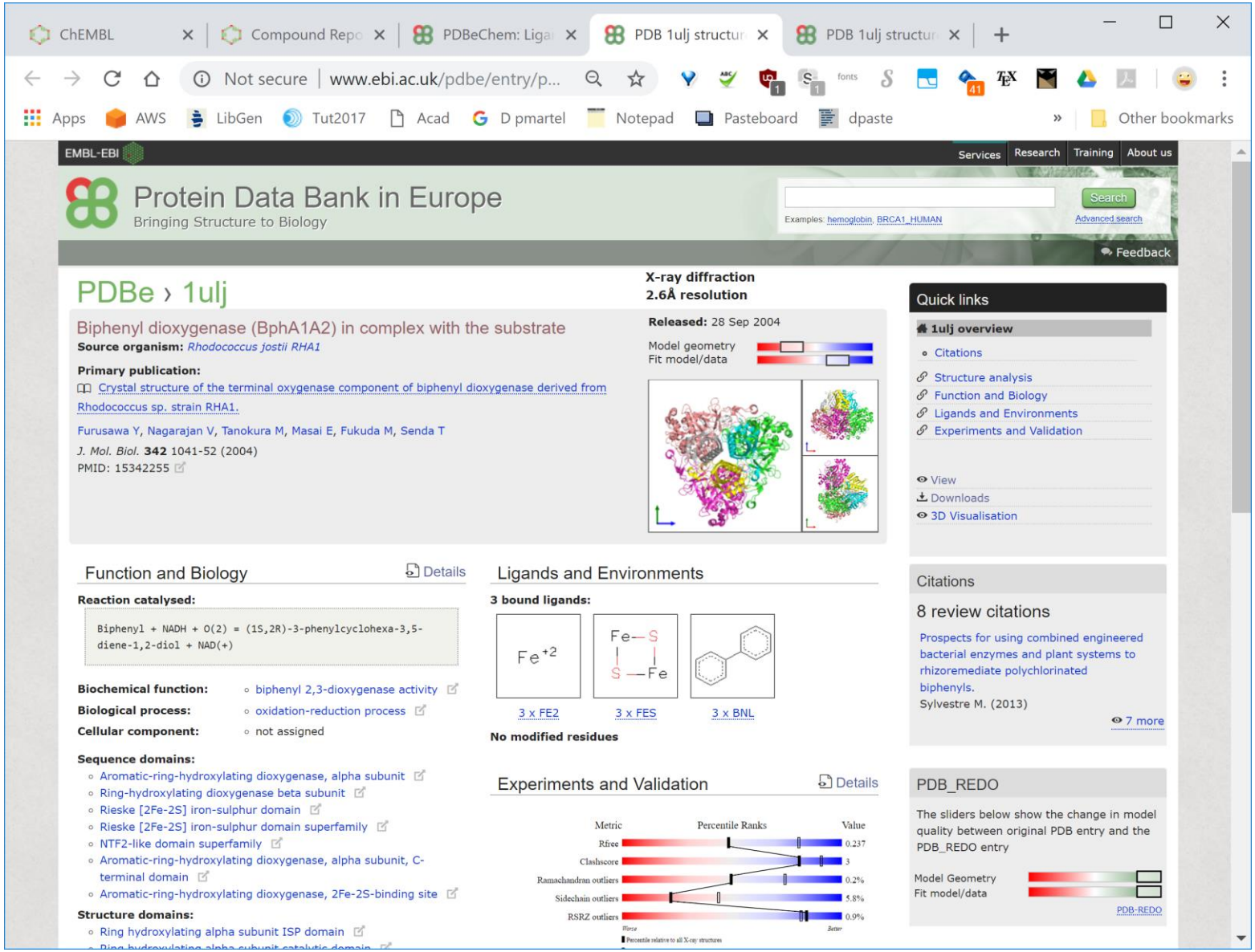
Exemplo de pesquisa estrutural em ChEMBL



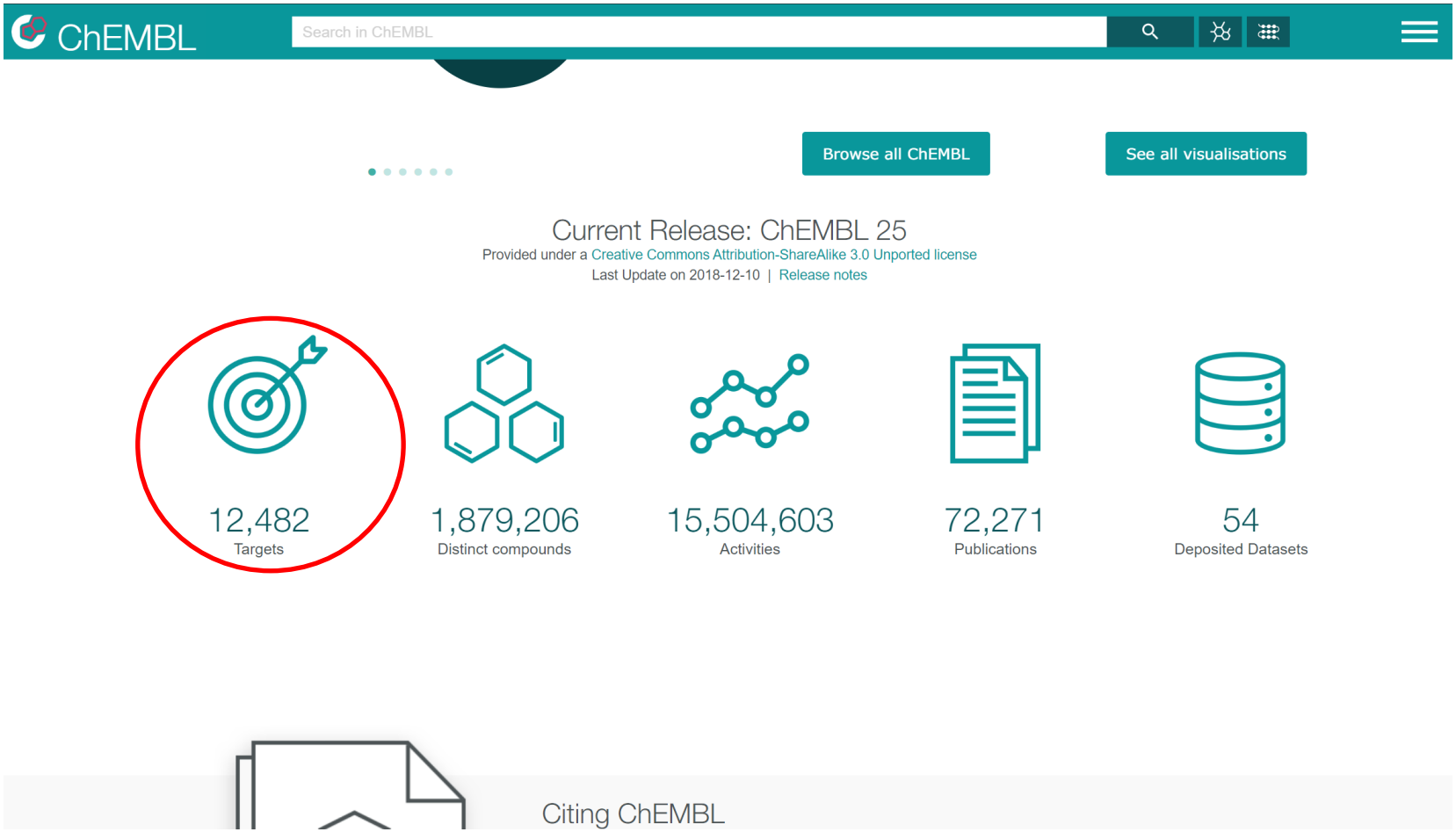
Exemplo de pesquisa estrutural em ChEMBL



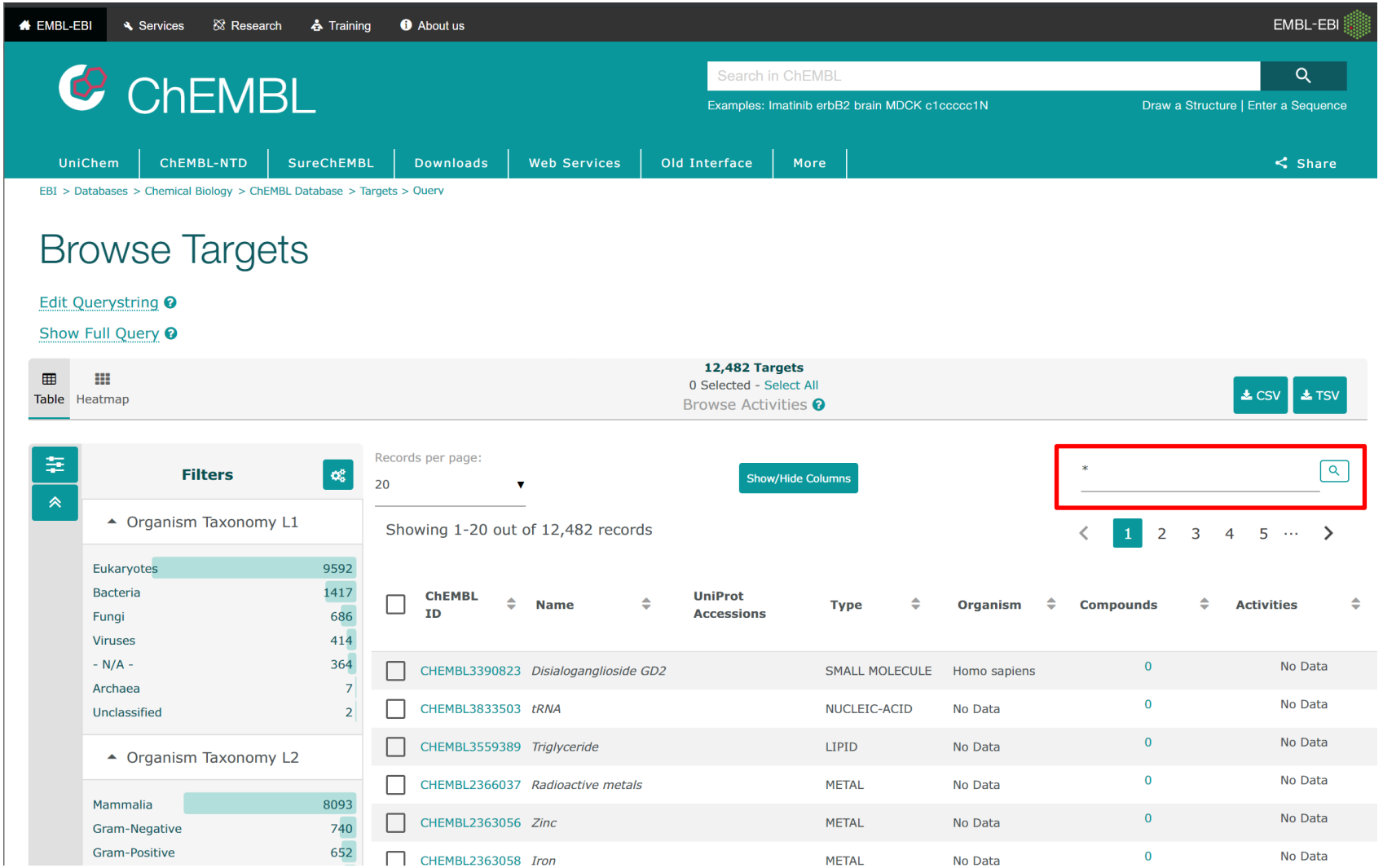
Exemplo de pesquisa estrutural em ChEMBL



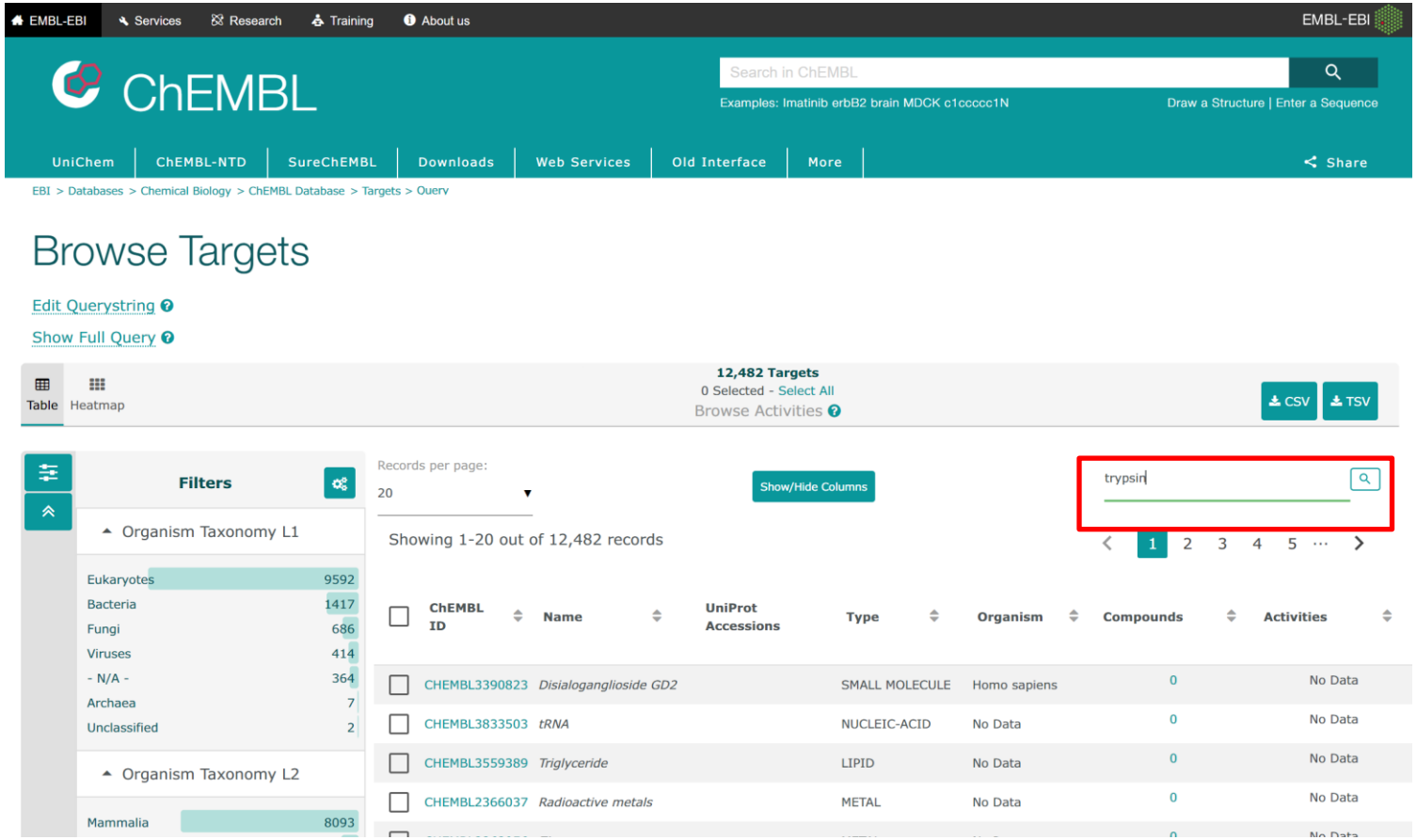
Pesquisa de targets em ChEMBL



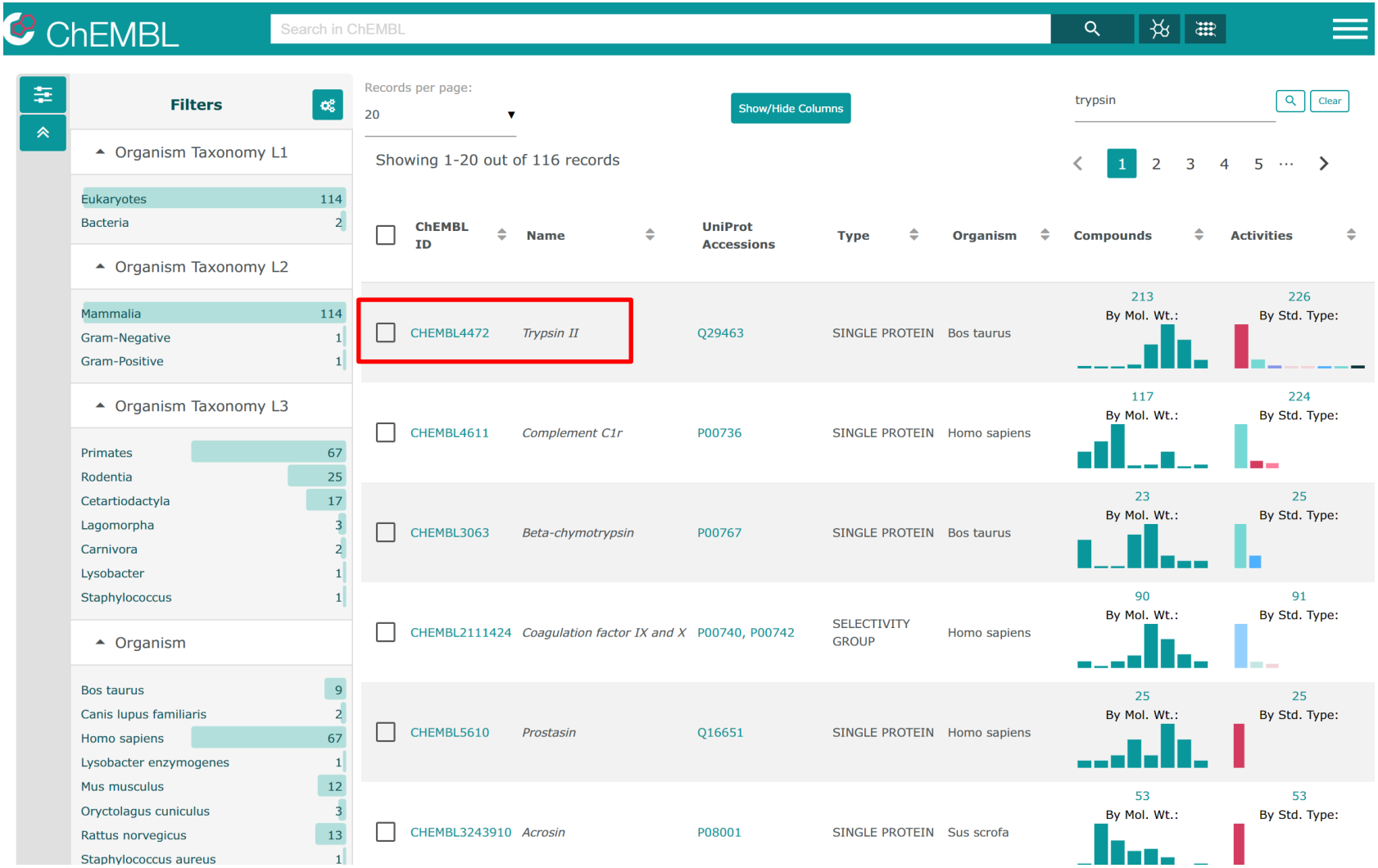
Pesquisa de targets em ChEMBL



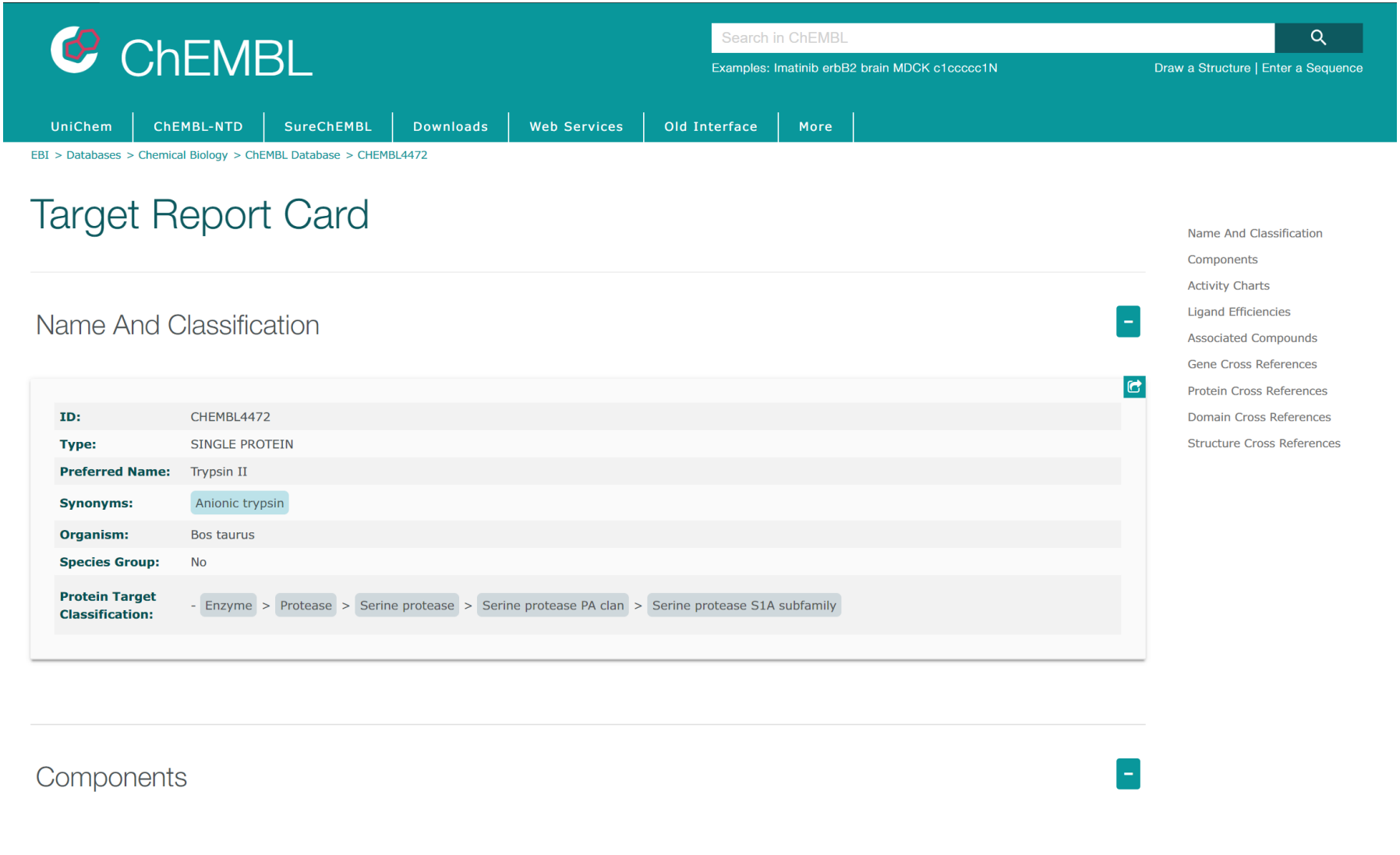
Pesquisa de targets em ChemBL



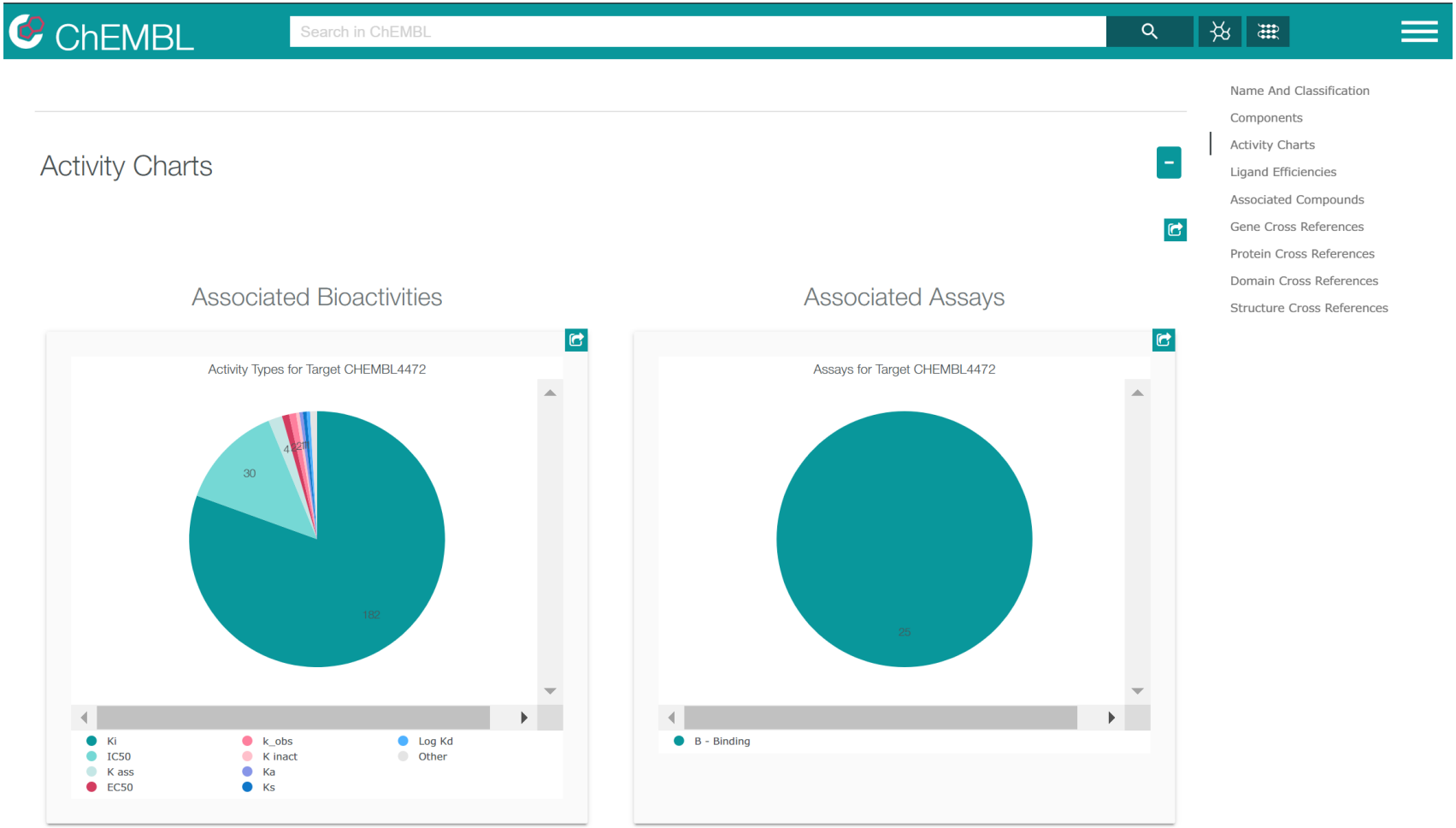
Pesquisa de targets em ChEMBL



Pesquisa de targets em ChEMBL

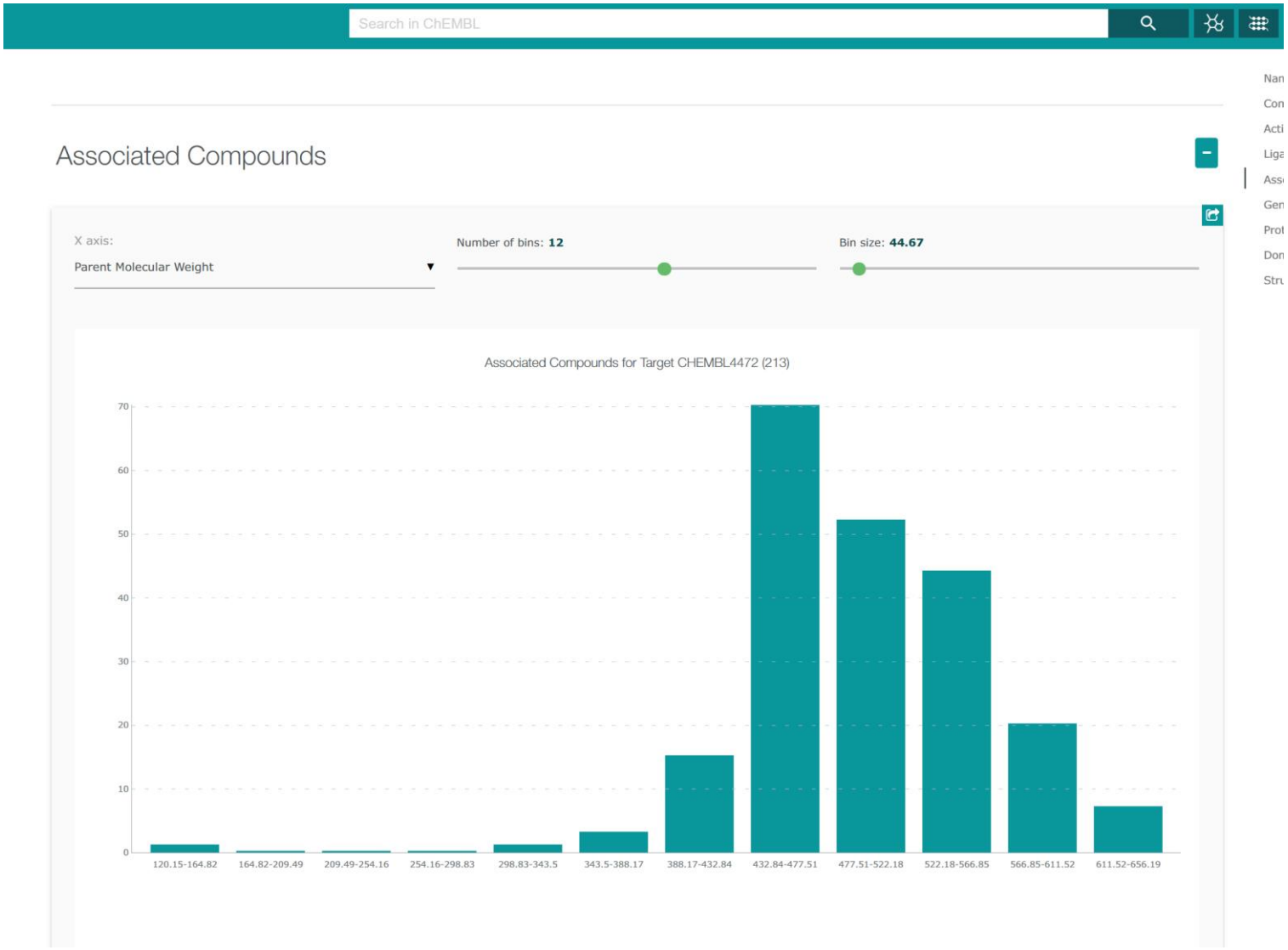


Pesquisa de targets em ChEMBL



Pesquisa de targets em ChEMBL

Pesquisa de targets em ChEMBL



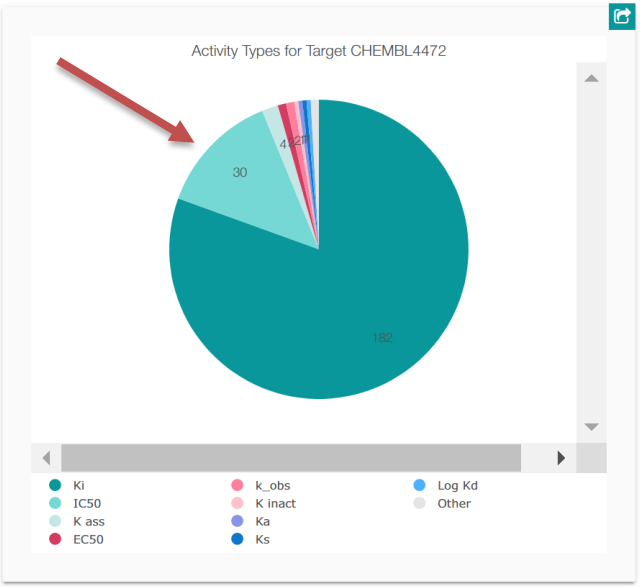
ChEMBL

Search in ChEMBL

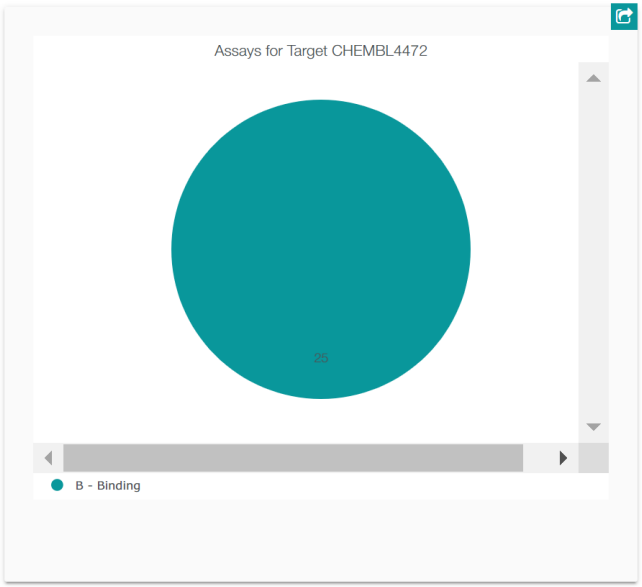
Name And Classification
Components
Activity Charts
Ligand Efficiencies
Associated Compounds
Gene Cross References
Protein Cross References
Domain Cross References
Structure Cross References

Activity Charts

Associated Bioactivities



Associated Assays



Browse Activities

Edit Querystring

Show Full Query

Table

30 Activities

0 Selected - Select All

Browse Compounds

CSV | TSV

Filters

Standard Type

IC50 30

Target Type

SINGLE PROTEIN 30

Organism Taxonomy L1

Eukaryotes 30

Organism Taxonomy L2

Mammalia 30

Organism Taxonomy L3

Cetartiodactyla 30

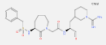
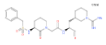
Target Organism

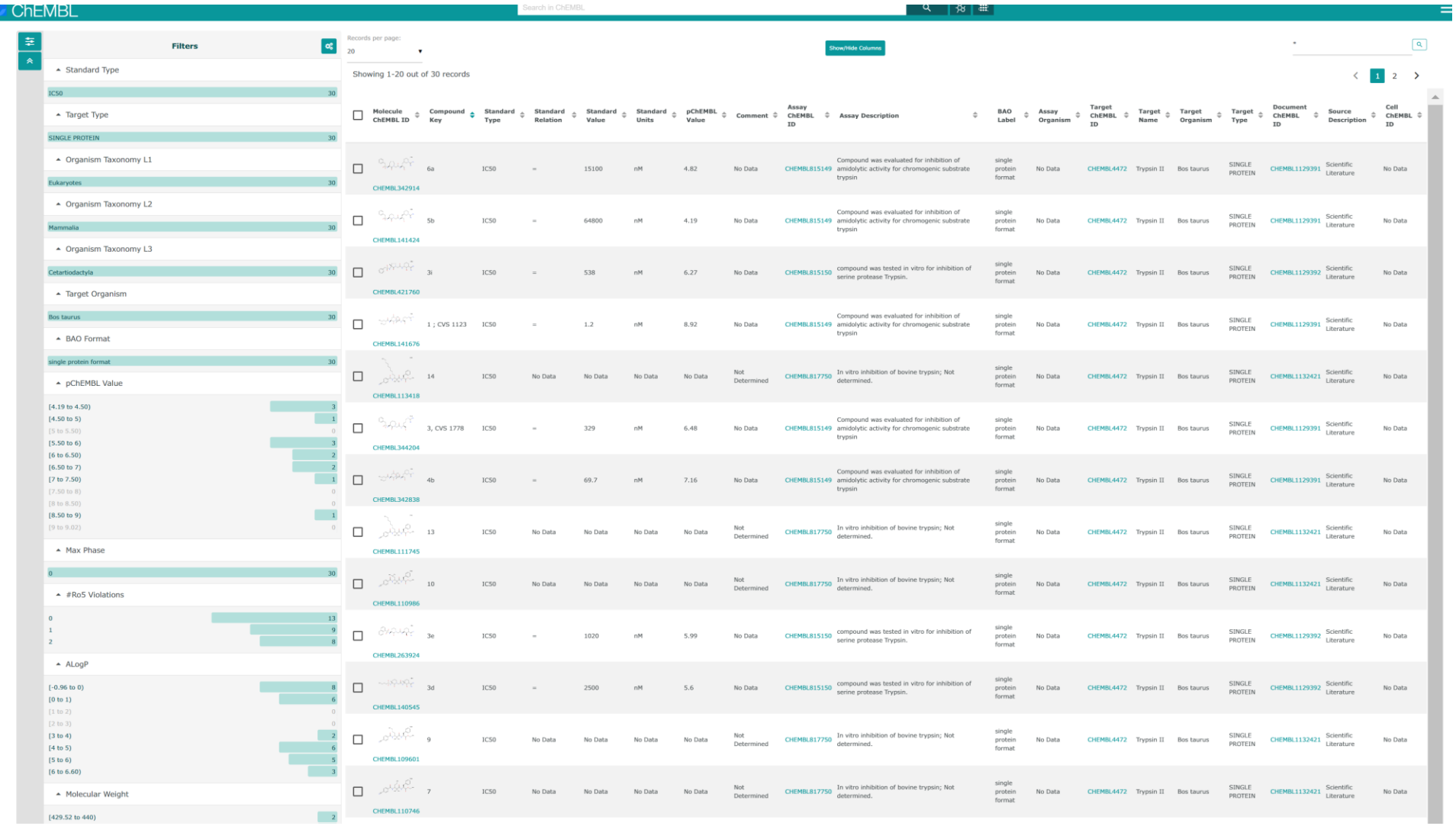
Records per page: 20

Show/Hide Columns

Showing 1-20 out of 30 records

1 2

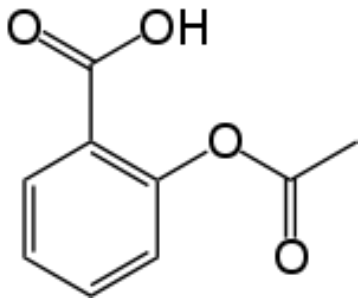
Molecule ChEMBL ID	Compound Key	Standard Type	Standard Relation	Standard Value	Standard Units	pChEMBL Value	Comment	Assay ChEMBL ID	Assay Descri
 CHEMBL342914	6a	IC50	=	15100	nM	4.82	No Data	CHEMBL815149	Compou evaluate inhibitio amidoly activity chromo substrat
 CHEMBL141424	5b	IC50	=	64800	nM	4.19	No Data	CHEMBL815149	Compou evaluate inhibitio amidoly activity chromo substrat


ChEMBL
Search in ChEMBL
Filters
Standard Type
IC50 30
Target Type
SINGLE PROTEIN 30
Organism Taxonomy L1
Eukaryotes 30
Organism Taxonomy L2
Mammalia 30
Organism Taxonomy L3
Cetartiodactyla 30
Target Organism
Bos taurus 30
BAO Format
single protein format 30
pChEMBL Value
[4.19 to 4.50) 3
[4.50 to 5) 1
[5 to 5.50) 0
[5.50 to 6) 3
[6 to 6.50) 2
[6.50 to 7) 2
[7 to 7.50) 1
[7.50 to 8) 0
[8 to 8.50) 0
[8.50 to 9) 1
[9 to 9.02) 0
Max Phase
0 30
#Ro5 Violations
0 13
1 9
2 8
ALogP
[-0.96 to 0) 8
[0 to 1) 6
[1 to 2) 0
[2 to 3) 0
[3 to 4) 2
[4 to 5) 6
[5 to 6) 5
[6 to 6.60) 3
Molecular Weight
[429.52 to 440) 2
Records per page:
20
Show/Hide Columns
Showing 1-20 out of 30 records
1 2
Molecule ChEMBL ID
Compound Key
Standard Type
Standard Relation
Standard Value
Standard Units
pChEMBL Value
Comment
Assay ChEMBL ID
Assay Description
BAO Label
Assay Organism
Target ChEMBL ID
Target Name
Target Organism
Target Type
Document ChEMBL ID
Source Description
Cell ChEMBL ID
CHEMBL342914 6a IC50 = 15100 nM 4.82 No Data CHEMBL815149 Compound was evaluated for inhibition of amidolytic activity for chromogenic substrate trypsin single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1129391 Scientific Literature No Data
CHEMBL141424 5b IC50 = 64800 nM 4.19 No Data CHEMBL815149 Compound was evaluated for inhibition of amidolytic activity for chromogenic substrate trypsin single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1129391 Scientific Literature No Data
CHEMBL421760 3i IC50 = 538 nM 6.27 No Data CHEMBL815150 compound was tested in vitro for inhibition of serine protease Trypsin. single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1129392 Scientific Literature No Data
CHEMBL141676 1 ; CVS 1123 IC50 = 1.2 nM 8.92 No Data CHEMBL815149 Compound was evaluated for inhibition of amidolytic activity for chromogenic substrate trypsin single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1129391 Scientific Literature No Data
CHEMBL113418 14 IC50 No Data No Data No Data No Data Not Determined CHEMBL817750 In vitro inhibition of bovine trypsin; Not determined. single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1132421 Scientific Literature No Data
CHEMBL344204 3, CVS 1778 IC50 = 329 nM 6.48 No Data CHEMBL815149 Compound was evaluated for inhibition of amidolytic activity for chromogenic substrate trypsin single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1129391 Scientific Literature No Data
CHEMBL342838 4b IC50 = 69.7 nM 7.16 No Data CHEMBL815149 Compound was evaluated for inhibition of amidolytic activity for chromogenic substrate trypsin single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1129391 Scientific Literature No Data
CHEMBL111745 13 IC50 No Data No Data No Data No Data Not Determined CHEMBL817750 In vitro inhibition of bovine trypsin; Not determined. single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1132421 Scientific Literature No Data
CHEMBL110986 10 IC50 No Data No Data No Data No Data Not Determined CHEMBL817750 In vitro inhibition of bovine trypsin; Not determined. single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1132421 Scientific Literature No Data
CHEMBL263924 3e IC50 = 1020 nM 5.99 No Data CHEMBL815150 compound was tested in vitro for inhibition of serine protease Trypsin. single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1129392 Scientific Literature No Data
CHEMBL140545 3d IC50 = 2500 nM 5.6 No Data CHEMBL815150 compound was tested in vitro for inhibition of serine protease Trypsin. single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1129392 Scientific Literature No Data
CHEMBL109601 9 IC50 No Data No Data No Data No Data Not Determined CHEMBL817750 In vitro inhibition of bovine trypsin; Not determined. single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1132421 Scientific Literature No Data
CHEMBL110746 7 IC50 No Data No Data No Data No Data Not Determined CHEMBL817750 In vitro inhibition of bovine trypsin; Not determined. single protein format No Data CHEMBL4472 Trypsin II Bos taurus SINGLE PROTEIN CHEMBL1132421 Scientific Literature No Data

SMILES

SMILES - Simplified Input Molecular Entry Specification

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



Estrutura 2D

O=C(Oc1ccccc1C(=O)O)C

SMILES

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

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

Fluoxetine

From Wikipedia, the free encyclopedia
(Redirected from Prozac)

"Prozac" redirects here. For other uses, see Prozac (disambiguation).

Fluoxetine, sold under the brand names **Prozac** and **Sarafem** among others, is an antidepressant of the selective serotonin reuptake inhibitor (SSRI) class.[2] It is used for the treatment of major depressive disorder, obsessive–compulsive disorder (OCD), bulimia nervosa, panic disorder, and premenstrual dysphoric disorder.[2] It is also approved for treatment of major depressive disorder in adolescents and children 8 years of age and over.[6] It has also been used to treat premature ejaculation.[2] Fluoxetine is taken by mouth.[2]

Common side effects include indigestion, trouble sleeping, sexual dysfunction, loss of appetite, dry mouth, and rash. Serious side effects include serotonin syndrome, mania, seizures, an increased risk of suicidal behavior in people under 25 years old, and an increased risk of bleeding.[2] Discontinuation syndrome is less likely to occur with fluoxetine than with other antidepressants, but it still happens in many cases. Fluoxetine taken during pregnancy is associated with significant increase in congenital heart defects in the newborns.[7][8] It has been suggested that fluoxetine therapy may be continued during breastfeeding if it was used during pregnancy or if other antidepressants were ineffective.[9]

Fluoxetine was discovered by Eli Lilly and Company in 1972, and entered medical use in 1986.[10] It is on the World Health Organization's List of Essential Medicines.[11] It is available as a generic medication.[2] In 2019, it was the 20th most commonly prescribed medication in the United States, with more than 27 million prescriptions.[12][13] Lilly also markets fluoxetine in a fixed-dose combination with olanzapine as olanzapine/fluoxetine (Symbyax).[14]

Contents [hide]

1 Medical uses
- 1.1 Depression
- 1.2 Obsessive–compulsive disorder
- 1.3 Panic disorder
- 1.4 Bulimia nervosa
- 1.5 Premenstrual dysphoric disorder
- 1.6 Impulsive aggression
- 1.7 Special populations

2 Adverse effects
- 2.1 Sexual dysfunction
- 2.2 Discontinuation syndrome
- 2.3 Pregnancy
- 2.4 Suicide
- 2.5 QT prolongation

3 Overdose
4 Interactions

The efficacy of fluoxetine in the treatment of obsessive–compulsive disorder (OCD) was demonstrated in two randomized multicenter phase III clinical trials. The pooled results of these trials demonstrated that 47% of completers treated with the highest dose were "much improved" or "very much improved" after 13 weeks of treatment, compared to 11% in the placebo arm of the trial.[3] The American Academy of Child and Adolescent Psychiatry state that SSRIs, including fluoxetine, should be used as first-line therapy in children, along with cognitive behavioral therapy (CBT), for the treatment of moderate to severe OCD.[29]

Panic disorder [edit]

The efficacy of fluoxetine in the treatment of panic disorder was demonstrated in two 12-week randomized multicenter phase III clinical trials that enrolled patients diagnosed with panic disorder, with or without agoraphobia. In the first trial, 42% of subjects in the fluoxetine-treated arm were free of panic attacks at the end of the study, vs. 28% in the placebo arm. In the second trial, 62% of fluoxetine treated patients were free of panic attacks at the end of the study, vs. 44% in the placebo arm.[3]

Bulimia nervosa [edit]

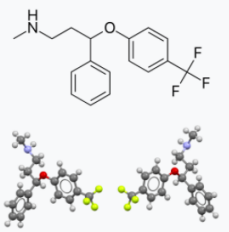
A 2011 systematic review discussed seven trials which compared fluoxetine to a placebo in the treatment of bulimia nervosa, six of which found a statistically significant reduction in symptoms such as vomiting and binge eating.[30] However, no difference was observed between treatment arms when fluoxetine and psychotherapy were compared to psychotherapy alone.

Premenstrual dysphoric disorder [edit]

Fluoxetine is used to treat premenstrual dysphoric disorder, a condition where individuals have affective and somatic symptoms monthly during the luteal phase of menstruation.[31][32] Taking fluoxetine 20 mg/d can be effective in treating PMDD,[33][34] though doses of 10 mg/d have also been prescribed effectively.[35][36]

Impulsive aggression [edit]

Fluoxetine

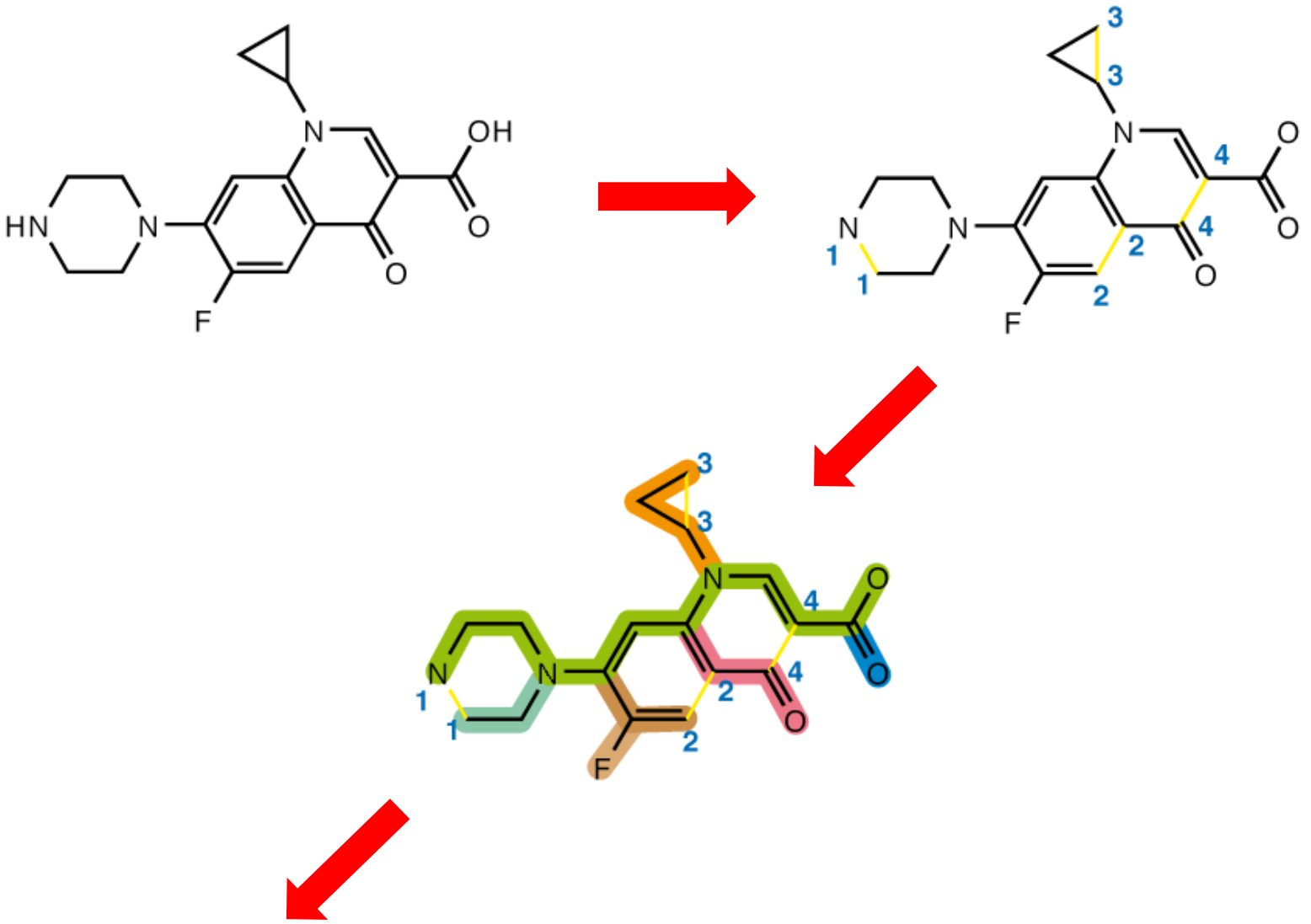


Fluoxetine (top),
(*R*)-fluoxetine (left), (*S*)-fluoxetine (right)

Clinical data	
Pronunciation	/flu ˈɒksəti:n/
Trade names	Prozac, Sarafem, Adofen, others
AHFS/Drugs.com	Monograph
MedlinePlus	a689006
License data	EU EMA: by INN US DailyMed: Fluoxetine US FDA: Fluoxetine
Pregnancy category	AU: C
Addiction liability	None[1]
Routes of administration	By mouth
Drug class	Selective serotonin reuptake inhibitor (SSRI)[2]
ATC code	N06AB03 (WHO) QN06AB03 (WHO) as HCl: CHEBI:5119 ✓
ChEMBL	ChEMBL41 ✓ as HCl: ChEMBL1201082 ✓
CompTox Dashboard (EPA)	DTXSID7023067
ECHA InfoCard	100.125.370
Chemical and physical data	
Formula	$C_{17}H_{18}F_3NO$
Molar mass	309.332 g·mol^{-1}
3D model (JSmol)	Interactive image
Chirality	Racemic mixture
Melting point	179 to 182 °C (354 to 360 °F)
Boiling point	395 °C (743 °F)
Solubility in water	14

SMILES [hide]
CNCCC(c1ccccc1)Oc2ccc(cc2)C(F)(F)F

InChI [hide]
InChI=1S/C17H18F3NO/c1-21-12-11-16(13-5-3-2-4-6-13)22-15-9-7-14(8-10-15)17(18,19)20/h2-10,16,21H,11-12H2,1H3 ✓
Key:RTHCYVBBDHJXIQ-UHFFFAOYSA-N ✓
(verify)



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: CH4 $\rightarrow$ C

CC	$H_3C\text{-}CH_3$
C=C	$H_2C{=}CH_2$
C#C	$HC{\equiv}CH$
CCO	$H_3C\text{-}CH_2OH$

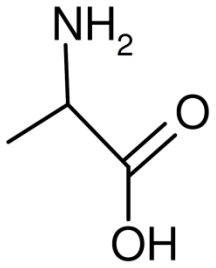
SMILES – Regras(2)

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:



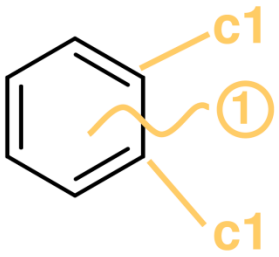
CC(N)C(=O)O

(determinar primeiro a sequência mais longa de ligações)

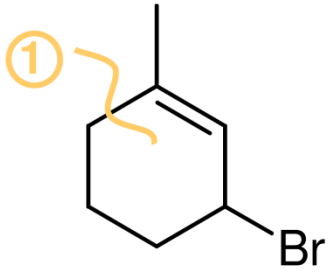
SMILES – Regras(4)

Compostos cíclicos:

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



c1ccccc1

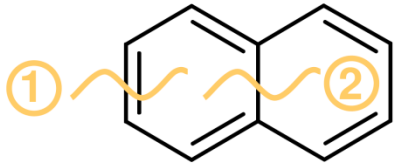


CC1=CC(Br)CCC1

SMILES – Regras(5)

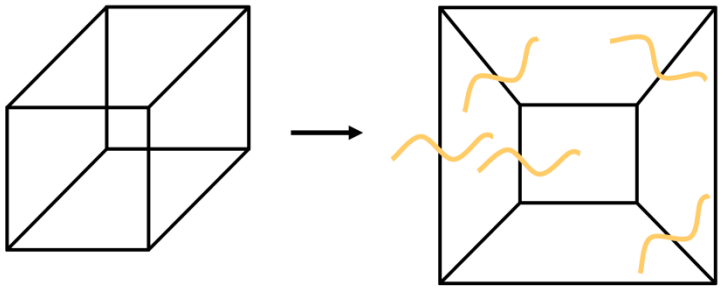
Compostos policíclicos:

- Múltiplos pontos de quebra



c1cc2ccccc2cc1

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



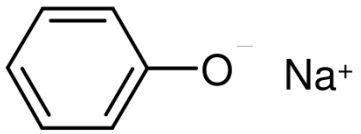
cubano

c12c3c4c1c5c4c3c25

Números maiores que 9 são antecedidos por um '%' : %11

SMILES – Regras(6)

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



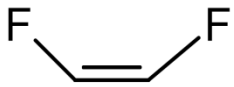
[Na+].[O-]c1ccccc1

Isótopos:

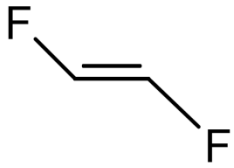
^{13}C	[13C]
$^{13}CH_4$	[13CH4]
D_2O	[2H]O[2H]

SMILES – Regras(7)

Configuração em torno de uma ligação dupla:



F/C=C\F *cis*

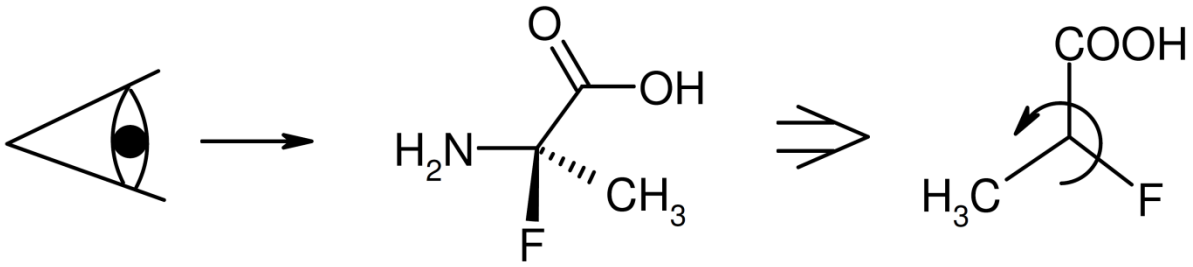


F/C=C/F *trans*

FC=CF Indeterminada

SMILES – Regras(8)

Quiralidade:



N[C@](C)(F)C(=O)O

@ - sequência anti-horária de substituintes
@@ - sequência horária de substituintes

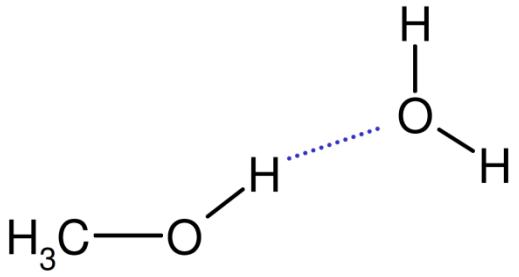
N.B. – Ausência de conformidade com o sistema (**r**,**s**) de representação absoluta da configuração

SMILES – Regras(9)

Hidrogénios explícitos:

H+ [H+] protão

H_2 [H][H]



CO[H][OH_2] ligação de hidrogénio

SMILES – Regras(9)

As reacções químicas são representadas usando símbolo “>” :

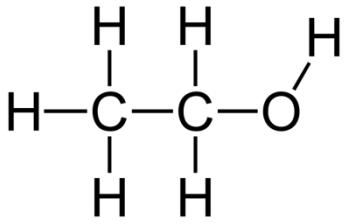
Reagentes > Agentes > Produtos

C.O=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] esterificaçãoácidda do ácida 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:



etanol

CCO OCC C(O)C

representações válidas

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)
[r*n*]	átomo num anel de *n* membros
[SX2]	enxofre com 2 substituintes mas não  ou
[Fe]	átomo de ferro (carga arbitrária)

SMARTS (2)

Operadores lógicos:

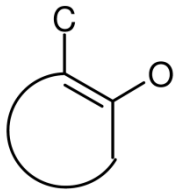
A,B	A ou B
A&B	A e B
A;B	A e B
!A	não A

exemplos:

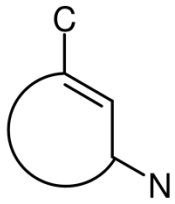
[F, Cl, Br, I]	F ou Cl ou Br ou I
[!C;R]	átomo aromático e não-alifático num anel
[CH2]	carbono alifático com 2 hidrogénios (metileno)
[c,n&H1]	carbono aromático ou NH aromático
[c,n;H1]	azoto ou carbono aromático e exactamente um hidrogénio
[#7;r5]	qualquer azoto num anel de 5 membros

SMARTS (3)

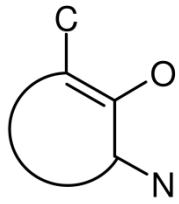
Configuração de substituintes:



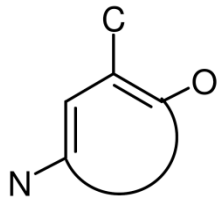
[CaaO]



[CaaaO]



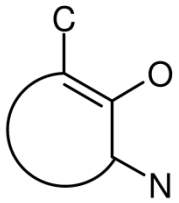
[Caa(O)aN]



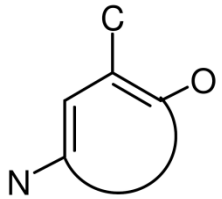
[Ca(aO)aaN]

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

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

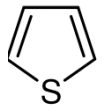
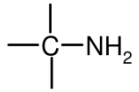


ou



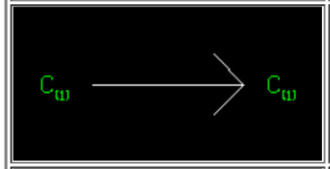
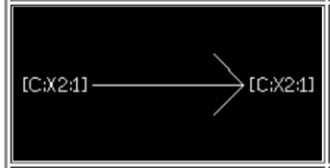
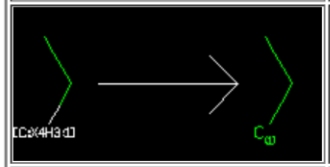
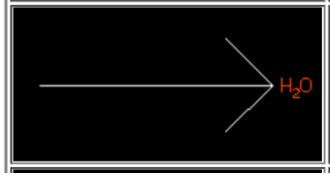
SMARTS (3)

Configuração de substituintes:

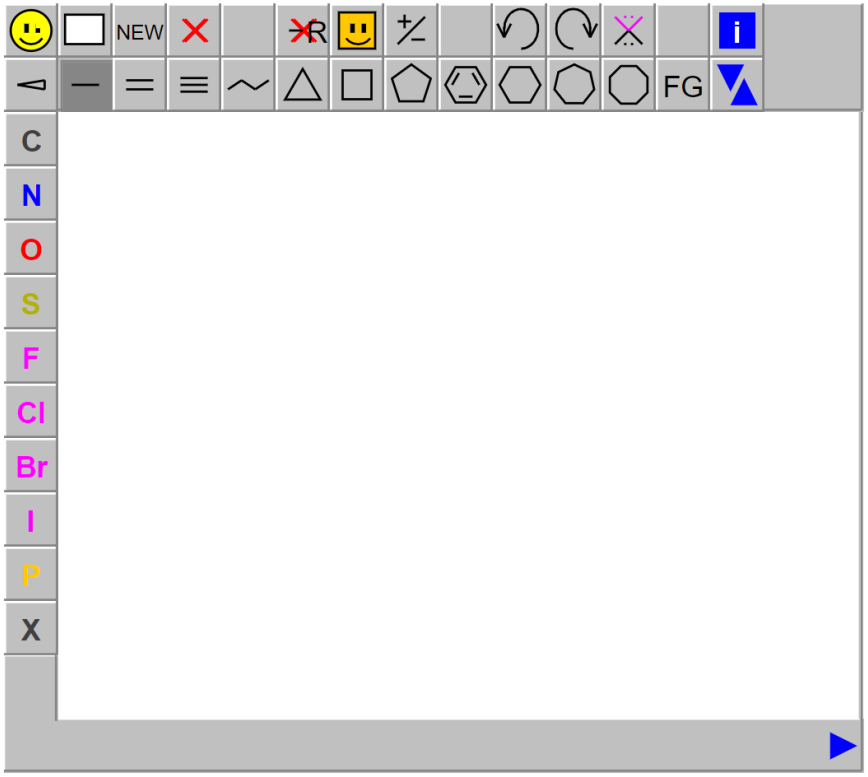
[s,o]1cccc1	tiofenos e furanos 
[CX4][NH2]	aminas alifáticas primárias 
[C1OC1]	epóxidos 
C(=O)[OH,O-,O-.+]	ácido carbónico, carboxilato ou catião
C(=O)[NH1]	ligação peptídica
=[OH]	ácidos e enóis
F.F.F.F.F	um total de 5 átomos de fluor as

SMIRKS

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

SMIRKS Depiction	Reaction	SMIRKS and Note
	Reacting carbon	**[C:1]>>[C:1]** Agents aren't allowed in SMIRKS. The format is " reactants >> products ".
	Reacting Carbon (2-Connected)	**[C;X2:1]>>[C;X2:1]** SMIRKS allows atomic SMARTS expressions. The syntax is: [<SMILES_PART>;<SMARTS_PART>:<MAP>]
	No Reaction	**[C;X4H3]-[CH2]C>>[C][CH2]C** SMARTS atom specifications may be used for mapped atoms only (i.e. unmapped atoms must be valid SMILES expressions).
[NO REACTION]	No Reaction	**[C;X2:1]~C>>[C;X2:1]=C** SMIRKS doesn't allow SMARTS Bond Queries (e.g. ~). Bonds expressions must be valid SMILES.
	Just add water	**>>O** 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 InChI e InChKey
- Pesquisa de moléculas on-line através da InChKey

https://jsme-editor.github.io/dist/JSME_test.html

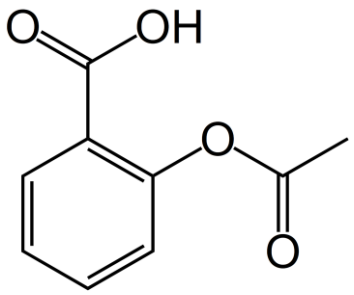
InChI Representation

InChI – IUPAC International Chemical Identifier
Developed by IUPAC and NIST 2000-2005 (pronounced “In Key”)

InChI is a text-based identifier for chemical substances, designed to offer 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=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)

1-version number
S-standardized InChI

Chemical Formula

Connectivity

Hydrogen Atoms

InChI and InChIKey

InChI's are too long and complex to reliably work as search keywords in database/internet searaches.

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

Different InChI's can produce the same InChiKey, but that's an extremely rare event.

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

sha-256

InChIKey BSYNRYMUTXBXSQ-UHFFFAOYSA-N

Connectivity **Other layers** **Protonation state** **Version** **Standard /Non-standard**

InChI and InChIKey

InChI's are too
keywords in da

InChIKey – Com
algorithm (sha-
with shorter le

Different InChI'
extremely rare

InChI=1S/C9H8

string

an

H,11,12)

CHAR	PROTONS	CHAR	PROTONS
N	0		
M	-1	O	+1
L	-2	P	+2
K	-3	Q	+3
J	-4	R	+4
I	-5	S	+5
H	-6	T	+6
G	-7	U	+7
F	-8	V	+8

InChIKey

Connectivity

Other layers

Protonation state

Version

Standard /Non-standard

Exercícios