

# Building a Chemical Database

In this tutorial we will build a database of COX-2 Inhibitors, perform chemical searching and clustering.

**Background** COX-2 selective inhibitor is a form of non-steroidal anti-inflammatory drug (NSAID) that directly targets COX-2, an enzyme responsible for inflammation and pain. Targeting selectivity for COX-2 reduces the risk of peptic ulceration, and is the main feature of celecoxib, rofecoxib and other members of this drug class.

**Controversy** After several COX-2 inhibiting drugs were approved for marketing, data from clinical trials revealed that COX-2 inhibitors caused a significant increase in heart attacks and strokes, with some drugs in the class having worse risks than others. Rofecoxib (commonly known as Vioxx) was taken off the market in 2004 because of these concerns and celecoxib (Celebrex) and traditional NSAIDS received boxed warnings on their labels.

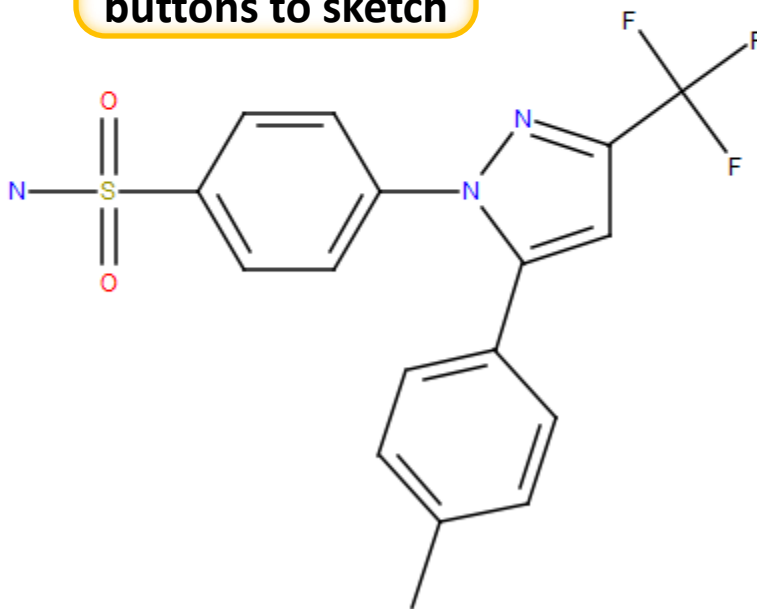


# Sketch the COX-2 inhibitor Celebrex

1. Click on the Molecular Editor button

2. Use the bond and atom buttons to sketch

3. Monitor Properties



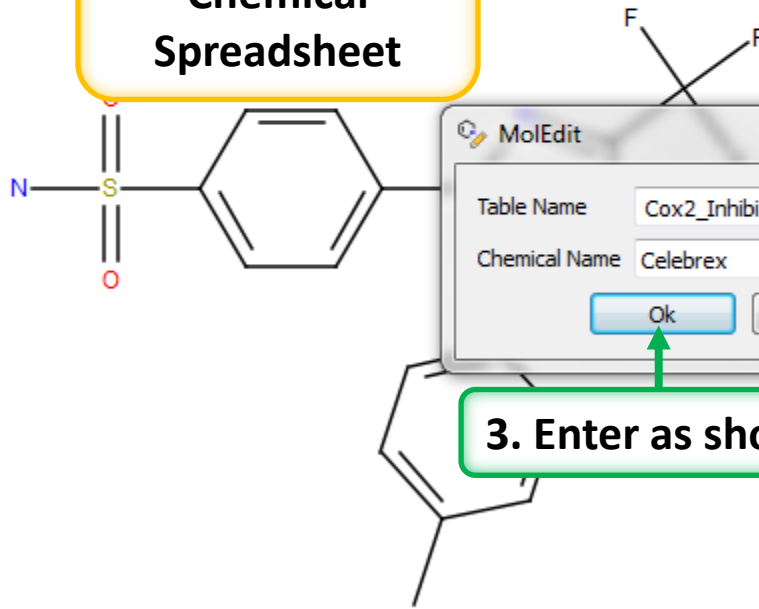
Name	Value
1 Formula	C17H14F3N3
2 Smiles	O=S(=O)(C(C=
3 IUPAC	4-(5-p-tolyl-3-(
4 InChI	InChI=1S/C17
5 InChIKey	RZEKVGVFHLEC
6 MolWeight	381.0759
7 HBA	4
8 HBD	2
9 RotB	4
10 DrugLikeness	-1.03104
11 MolArea	354.755
12 MoldHf	-146.26
13 MolLogP	3.96221
14 MolLogS	-6.15579
15 MolPSA	63.6646
16 Volume	313.115
17 Bad Groups	
18 Groups	Phenyl Pyrazole Halo Sulfonamide

Add the name of the drug and export to chemical spreadsheet.

1. Add Name "Celebrex"

2. Export to Chemical Spreadsheet

3. Enter as shown



MolEdit

Table Name: Cox2\_Inhibitors

Chemical Name: Celebrex

Ok Cancel

Name	Value
1 Formula	C17H14F3N3
2 Smiles	O=S(=O)(C(C=
3 IUPAC	4-(5-p-tolyl-3-(
4 InChI	InChI=1S/C17
5 InChIKey	RZEKVGVHFLEC
MolWeight	381.0759
HBA	4
HBD	2
RotB	4
DrugLikeness	-1.03104
MolArea	354.755
MoldHf	-146.26
13 MolLogP	3.96221
14 MolLogS	-6.15579
15 MolPSA	63.6646
16 Volume	313.115
17 Bad Groups	
18 Groups	Phenyl Pyrazole Halo Sulfonamide

Observe new chemical spreadsheet.

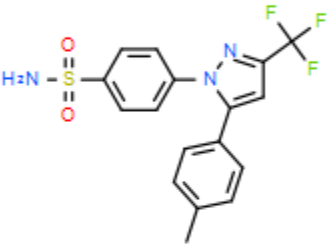
Cox2\_Inhibitors Molsoft icm 3.8-1 [NewProject \*] (1 table)

File Edit View Bioinfo Tools Homology Chemistry Docking MolMechanics Win

search ligedit

PDB Search 1xbb in All Fields Append To Result pdbReadNmrMode all occupancyDisplay none

Workspace... Cox2\_Inhibitors

mol	Molecule Name
 1	Celebrex

no selection

tables

Cox2\_I

All

GO

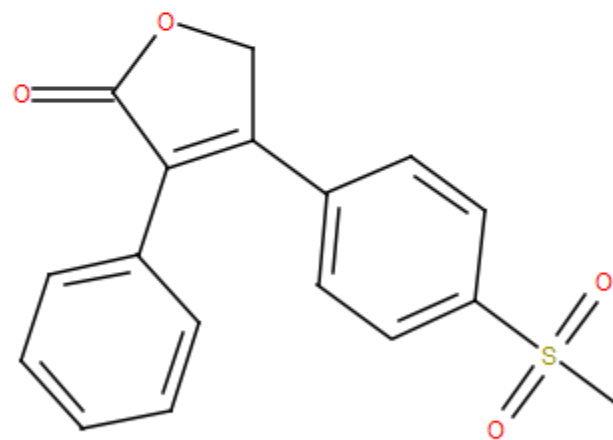
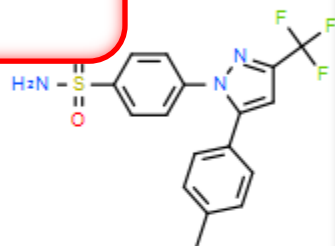
Chemical structure of Celebrex (2-(4-(4-chlorophenyl)-5-(trifluoromethyl)-1H-imidazol-2-yl)benzenesulfonamide) is shown in the spreadsheet cell.

Add a new row to the chemical spreadsheet and add Vioxx.

1. Right click  
"Insert New  
Row"

2. Double click and sketh  
Vioxx shown (right)

3. Close editor and the  
chemical will be added to  
the spreadsheet



Now you should have two chemicals in your spreadsheet.

New Molsoft icm 3.8-1 [NewProject \*] (1 table)

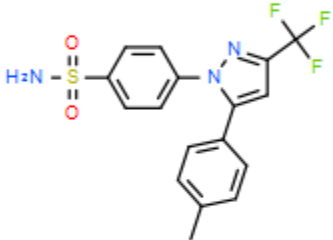
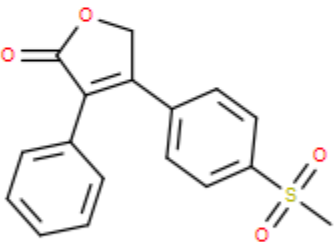
File Edit View Bioinfo Tools Homology Chemistry Docking MolMechanics Win

search ligedit

PDB Search 1xbb All Fields Append To Result pdbReadNmrMode all occupancyDisplay none

Workspace... no selection tables Cox2\_II

Cox2\_Inhibitors

mol	Molecule Name
	Celebrex
	Vioxx

All

Chemical structures shown: Celebrex (1) and Vioxx (2). The Vioxx entry is highlighted with a black border.

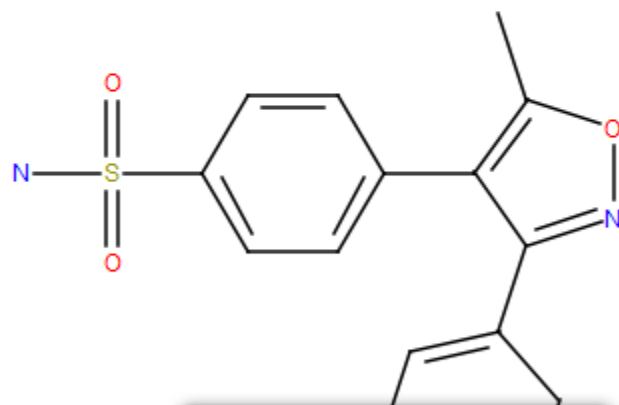
Add the COX-2 inhibitor Bextra to the spreadsheet using SMILES string.

1. Click on the Ligand Editor button

2. Edit/Add Smiles

3. Paste SMILES string

4. Export chemical to spreadsheet



Molecule Editor

Enter SMILES string (e.g. 'C(=O)O')

CN1C=CC2=C(C1)C(=O)N2C3=CC=CC=C3S(=O)(=O)N

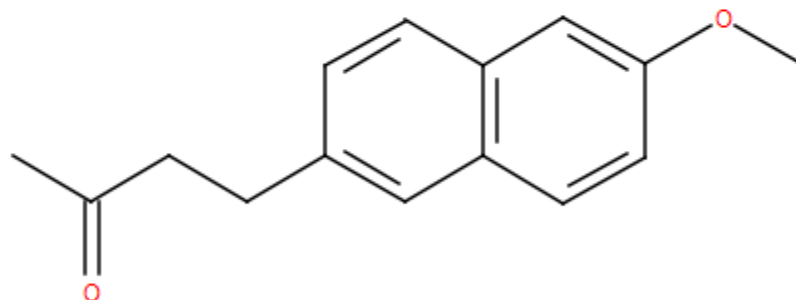
OK Cancel

	Value
1	Formula C <sub>16</sub> H <sub>14</sub> N <sub>2</sub> O <sub>3</sub>
2	Smiles O=S(=O)(N)C(c1ccc(cc1)C2=CN(C)C=C2)C3=CC=CC=C3
3	IUPAC 4-(5-methyl-3-pyrrolo[2,1-b]isoxazol-2-yl)benzenesulfonamide
4	InChI InChI=1S/C16H14N2O3
5	InChIKey LNPDTQAFDNK
6	MolWeight 314.0725
7	HBA 5
8	HBD 2
9	RotB 3
10	DrugLikeness 0.0330088
11	MolArea 316.735
12	MoldHf -14.1406
13	MolLogP 3.06622
14	MolLogS -4.98608
15	MolPSA 72.5302
16	Volume 273.547
17	Bad Groups
18	Groups Phenyl Isoxazole Sulfonamide

Use the chemical dictionary to read in the drug Relafen (Nabumetone).

1. Click and start typing "Nabumetone"

2. Export to Chemical Spreadsheet



		C15 H16 O2	
2	Smiles	O(C(=CC(=C(C	
3	IUPAC	4-(8-methoxy-t	
4	InChI	InChI=1S/C15H	
5	InChIKey	BLXXJMDCKKHM	
6	MolWeight	228.1150	
7	HBA	2	
8	HBD	0	
9	RotB	4	
10	DrugLikeness	-0.746927	
11	MolArea	276.562	
12	MoldHf	-51.083	
13	MolLogP	3.82632	
14	MolLogS	-4.57031	
15	MolPSA	20.8457	
16	Volume	240.325	
17	Bad Groups		
		Ketone (Carbor	
		Ether	
18	Groups		



Extract a COX-2 inhibitor from a PDB file.

1. Search "4COX"

2. Right click on "indomethacin" and choose the option "Extract Ligand"

3. Enter as shown

The screenshot displays the MolScribe software interface. The top menu bar includes File, Edit, View, Bioinfo, Tools, Homology, Chemistry, Docking, MolMechanics, and Win. The toolbar contains various icons for file operations, search, and visualization. The main workspace shows a 3D molecular model of a protein-ligand complex. The left sidebar, titled 'Workspace Panel', lists objects and tables. The 'objects' list includes '4cox' and several sub-objects (a, b, c, d, ahem, aimn, bhem). The 'aimn' object, representing indomethacin, is highlighted. A context menu is open over 'aimn', with 'Extract Ligand' selected. A dialog box titled 'Extract ligand into chemical table' is open, showing options for 'as 2D drawing' (selected) and 'keep 3D coordinates', a checked 'append To Existing Table' option, and a dropdown menu set to 'Cox2\_Inhibitors'. The 'Ok' button is highlighted.

no selection

objects (1 item)

- 4cox [1\*] XR; 2.9Å
  - a 552 A 3 gl 13 sites PGH2\_M
  - b 552 A 3 gl 9 sites PGH2\_MOU
  - c 552 A 3 gl 8 sites PGH2\_MOU
  - d 552 A 3 gl 8 sites PGH2\_MOU
  - ahem H protoporphyrin ix conta
  - aimn H indomethacin
  - bhem H protoporphyrin ix conta

tables (1 item)

- Cox2\_Inhibitors 4 rows 2 cols 0 header

Extract ligand into chemical table

as 2D drawing  keep 3D coordinates

append To Existing Table Cox2\_Inhibitors

Ok Cancel

1 non-ICM Obj

Insert additional chemical properties to the spreadsheet.

1. Chemistry/Calculate Properties

4cox Molsoft icm 3.8-1 [NewProject \*] (1 object 1 table)

File Edit View Bioinfo Tools Homology Chemistry Docking MolMechanics Win

display light labels

Workspace Panel

no selection

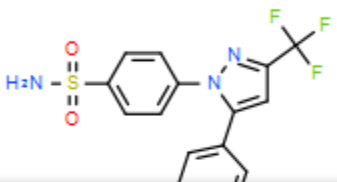
objects (1 item)

- 4cox [1\*] XR; 2.9A
- a 552 A 3 gl 13 sites F
- b 552 A 3 gl 9 sites PGH:
- c 552 A 3 gl 8 sites PGH:

mol

Molecule Name

Celebrex



1

Calculate Chemical Properties

Function

Function Multiple Functions

Arguments

Main argument: mol

New column location

Insert  after  before  in-place

at end

column mol

New column name

name

Action

Function	Name	Category	Description
<input checked="" type="checkbox"/> MolWeight(mol)	molWeight	Chemical	Molecular w...
<input type="checkbox"/> MolFormula(mol)	molFormula	Chemical	Chemical fo...
<input type="checkbox"/> IupacName(mol)	iupacName	Chemical	IUPAC nom...
<input checked="" type="checkbox"/> MolLogP(mol)	molLogP	Chemical	Octanol wat...
<input checked="" type="checkbox"/> MolLogS(mol)	molLogS	Chemical	Water solub...
<input type="checkbox"/> MolPSA(mol)	molPSA	Chemical	Polar surfac...
<input type="checkbox"/> MolVolume(mol)	molVolume	Chemical	Molecular v...
<input type="checkbox"/> MoldHf(mol)	moldHf	Chemical	Heats of for...
<input type="checkbox"/> DrugLikeness(mol)	drugLikeness	Chemical	Empirical dr...
<input type="checkbox"/> Smiles(mol,mode='unique')	smiles	Chemical	SMILES/SM...
<input type="checkbox"/> BadGroups(mol)	badGroups	Chemical	Unwanted o...
<input type="checkbox"/> Nof_Atoms(mol,atom='*')	nof_Atoms	Chemical	Number of ...
<input type="checkbox"/> Nof_Molecules(mol)	nof_Molecules	Chemical	Number of i...
<input type="checkbox"/> Nof_Fragments(mol,fragment='*')	nof_Fragments	Chemical	Number of f...
<input type="checkbox"/> Nof_Chirals(mol,type='any')	nof_Chirals	Chemical	Number of c...
<input type="checkbox"/> Nof_Rings(mol)	nof_Rings	Chemical	Number of r...
<input type="checkbox"/> Max_Ring_Size(mol)	max_Ring_Size	Chemical	Largest ind...

Add To List

OK Cancel

Search ChEMBL for COX2 inhibitors.



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

**ChEMBL Statistics**

- DB: ChEMBL\_18
- Targets: 9,414
- Compound records: 1,566,998
- Distinct compounds: 1,359,508
- Activities: 12,419,715

EBI > Databases > Small Molecules > ChEMBL Database > Home

cox2 Compounds Targets Assays Documents Activity Sources

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

1. Search by Target "cox2"

List Search

SMILES Search ChEMBL ID Search Keyword Search

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

Fetch Compounds

Biologicals Blast Search

[Empty search box]

Open up ChEMBL record ID 230.

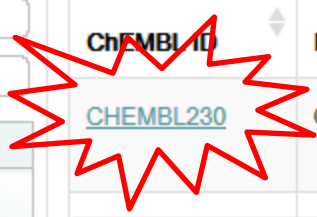
- ChEMBL
- Downloads
- Malaria Data
- ChEMBL-NTD
- Kinase SARfari
- GPCR SARfari
- DrugEBility
- Web Services
- FAQ

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

ChEMBL Target Search Results: 11 Please select...

10 records per page Show / hide

ChEMBL ID	Preferred Name	UniProt Accession	Target Type	Organism	Compounds	Bioactivit
<a href="#">CHEMBL230</a>	Cyclooxygenase-2	<a href="#">P35354</a>	SINGLE PROTEIN	Homo sapiens	5282	8829
ChEMBL ID	Preferred Name	UniProt Accession	Target Type	Organism	Compounds	Bioactivit
<a href="#">CHEMBL4102</a>	Cyclooxygenase-2	<a href="#">P79208</a>	SINGLE PROTEIN	Ovis aries	878	1267
<a href="#">CHEMBL2094253</a>	Cyclooxygenase	<a href="#">P23219</a>	PROTEIN	Homo sapiens	853	1146



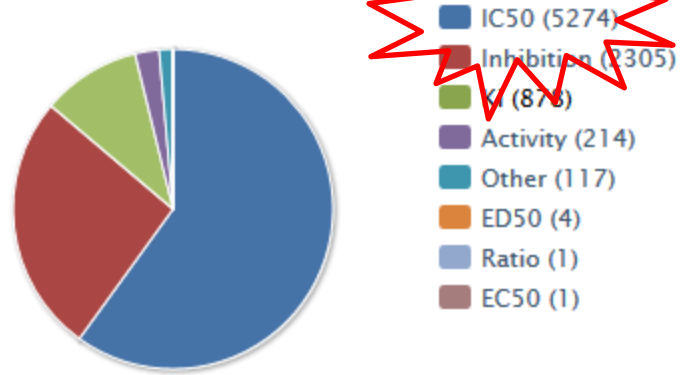
**ChEMBL Statistics**

- DB: ChEMBL\_18
- Targets: 9,414
- Compound records: 1,566,998
- Distinct compounds: 1,359,508
- Activities: 12,419,715
- Publications: 53,298

<a href="#">CHEMBL622</a>	ETODOLAC	Cyclooxygenase-2 inhibitor	<a href="#">PubMed</a>
<a href="#">CHEMBL599</a>	MELOXICAM	Cyclooxygenase-2 inhibitor	<a href="#">PubMed</a>
<a href="#">CHEMBL1070</a>	NABUMETONE	Cyclooxygenase-2 inhibitor	<a href="#">DailyMed</a>
<a href="#">CHEMBL122</a>	ROFECOXIB	Cyclooxygenase-2 inhibitor	<a href="#">FDA</a>
<a href="#">CHEMBL865</a>	VALDECOXIB	Cyclooxygenase-2 inhibitor	<a href="#">DailyMed</a>

Target Associated Bioactivities

ChEMBL Activity Types for Target CHEMBL230



Total: 8829

Target Associated Assays

ChEMBL Assays for Target CHEMBL230



Compounds

Targets

Assays

Documents

[Activity Source Filter](#)

s: 878

Please select...

Please select...

[Download All Bioactivity Data \(Tab-delimited\)](#)[Download All Bioactivity Data \(XLS\)](#)[Show / hide columns](#)

Relation	Standard Value	Standard Units	Assay Type	Description	Assay Src Description	Assay Organism	Target Type	Target Name	Target Organism	Reference
			B	<a href="#">DRUGMATRIX: Cyclooxygenase COX-2 enzyme inhibition (substrate: Arachidonic acid)</a>	DrugMatrix in vitro pharmacology assays		SINGLE PROTEIN	<a href="#">Cyclooxygenase-2</a>	Homo sapiens	<a href="#">CHEMBL1909046</a>
			B	<a href="#">DRUGMATRIX:</a>	DrugMatrix in		SINGLE	<a href="#">Cyclooxygenase-2</a>	Homo	<a href="#">CHEMBL1909046</a>

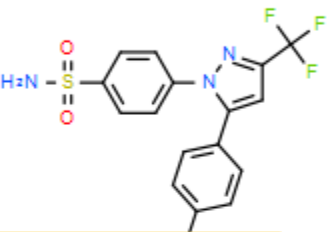
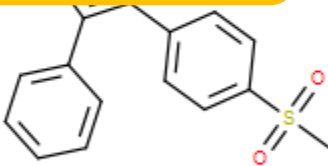
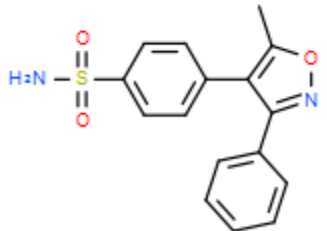
Download the data in XLS format and then open in Excel and save in comma separated value form (.csv)

Read in the .csv file into ICM (File/Open)

1. File/Open

2. Check this option

3. Click OK

mol	Molecule Name	drugLikeness	molWeight	molLogP	molLogS	nof HBA	nof HBD
	Celebrex	-1.031	381.1	3.962	-6.156	6	2
	Bextra	0.0	228.1	3.826	-4.57	3	0
	Nabumetone	-0.7469	228.1	3.826	-4.57	3	0

CSV reading options

top two lines

```
CMPD_CHEMBLID,MOLREGNO,PARENT_CMPD_CHEMBLID,PARENT_MOLREGN..  
CHEMBL24441,32459,CHEMBL24441,32459,BETAHISTINE,BETAHISTIN..
```

Read first row as field names

Ignore text in lines after hash

Treat empty fields in num columns as ND

Skip first lines

Ok Cancel

Convert the SMILES string into 2D sketch.

1. Chemistry/Convert/ Smiles to 2D

Smiles to 2D

Table: chembl\_cox2 Smiles Column: CANONICAL\_SMILES

Keep Original Column

Ok Cancel Help

2. Enter as shown

	.LOGP	PSA	NUM RO5 VIOLATIONS	CANONICAL SMILES	ACTIVITY ID	STANDARD TYPE	REL
1	.68	24.92	0	CNCCc1ccccc1	7613656	Ki	
2	.63	64.79	0	CCCOc1ccc(cc1N)C(=O)OCCN(CC)CC	7732956	Ki	
3	.9	138.53	0	Cc1cnc(cn1)C(=O)NCCc2ccc(cc2)S(=O)	7668268	Ki	
4	.98	202.26	1	CC(C)[C@H](NC(=O)N(C)Cc1csc(n1)C(C)C)C(=O)N[C@H]	7809126	Ki	
5	.4	78.49	0	CN1C(=O)CN=C(c2ccccc2)c3cc(ccc13)[N+](=O)[O-]	7722813	Ki	
6	.99	62.66	0	CN(C1CCCC1)C(=O)CCCCO2ccc3nc(O)ccc3c2	7688611	Ki	
7							
8	.78	40.46	0				
9	.9	36.86	0				
10	.75	91.76	1				
11	.1	123.79	1				
12	.94	56.73	0				
13							
14	.61	37.29	0	CC(C)Cc1ccc(cc1)C(C)C(=O)O	7662384	Ki	
15	.21	64.09	0	CCN1C=C(C(=O)O)C(=O)c2ccccc2	7727701	Ki	
16	.88	29.54	0	COc1ccccc1C(=O)N2CCC(Cc3ccccc3)C2	7636680	Ki	
17	.75	63.32	0	NC[C@@H]1CC[C@H](CC1)C(=O)O	7785816	Ki	
18	.5	26.02	0	NC12CC3CC(CC(C3)C1)C2	7625398	Ki	
19	.41	72.88	0	CC1CN(CCN1)c2cc3N(C=C(C(=O)O)C(=O)c3c(C)c2F)C4CC	7676976	Ki	
20	.54	68.36	0	C[C@@H](CO)NC(=O)[C@H]1CN(C)C(=O)N1	7667088	Ki	
21	.31	63.4	0	NC(=O)CN1CCCC1=O	7713348	Ki	
22				CN[C@H](CC(C)C)C(=O)N[C@H]1[C@H]CNC1=O	7787337	Ki	
23	.13	65.84	0	O=C1N=CN=C2NNC=C12	7623701	Ki	
24	.22	72.83	0	CC[C@H](C)C(=O)O[C@H]1C[C@@H](C)C=C2C=C[C@H]C2	7753250	Ki	
25	.2	118.21	1	CN1C[C@@H](C)C(=O)N1	7698463	Ki	



# Chemical Searching

Let us compare our small spreadsheet of COX-2 drugs with the larger set of inhibitors from ChEMBL.

1. Chemistry/Compare two sets

Compare Two Sets

First Table: Cox2\_Inhibitors      Second Table: chembl\_cox2

exact     similarity

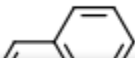

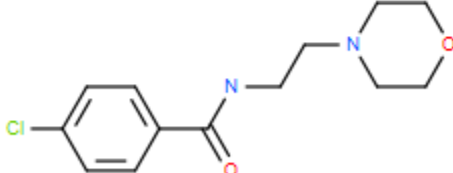

Distance: 0.5

Select Matching Records     Show Only Different Records

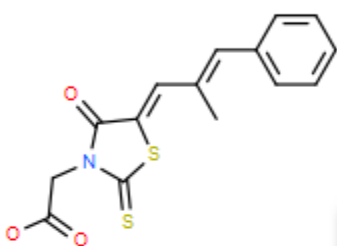
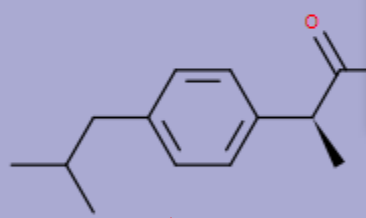
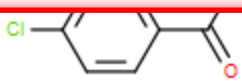
Hint  
Double-click on the tab of the second table to see them side by side

Ok    Cancel    Help

2. Enter as shown

CMPD CHEMBLID	mol	MOLREGNO	PARENT CMPD CHEMBLID	PARENT MOLREGNO	MOL PREI
CHEMBL56337		88882	CHEMBL56337	88882	EPALREST
CHEMBL175				317909	DEXIBUPF
CHEMBL86304		139099	CHEMBL86304	139099	MOCLOBE
CHEMBL1076014				611501	TOLEPERY

Let us make a new spreadsheet from the comparison hits. Selected compounds will be highlighted in blue.

CMPD CHEMBLID	mol	MOLREGNO	PARENT CMPD CHEMBLID	PARENT MOLREGNO	MOL PREI
CHEMBL56337		88882	CHEMBL56337	88882	EPALREST
CHEMBL175					09 DEXIBUPF
CHEMBL86304		139099	CHEMBL86304	139099	MOCLOBE

1. Right click and choose Copy rows to ICM Table

Table

New table name:

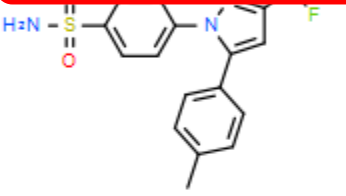
similar\_0.5

OK Cancel

2. Enter a new table name for the similar compounds

Save the spreadsheets in SDF format and delete Cox2\_Inhibitors and similar\_0\_5

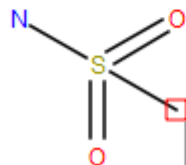
1. Right click "Save As" .sdf.

mol	Molecule Name	drugLikeness	molWeight	molLogP	molLogS	nof HBA	nof HBD
1		-1.031	381.1	3.962	-6.156	6	2
2	Vioxx	-0.7455	314.1	2.205	-4.957	7	0
3	Bextra	0.03301	314.1	3.066	-4.986	7	2
	Nabumetone	-0.7469	228.1	3.826	-4.57	3	0

1. Click here to  
open Chemical  
Search window

CMPD ID	MOLREGNO	PARENT CMPD CHEMBLID	PARENT MOLREGNO	MOL PREI
CHEMBL24441	32459	CHEMBL24441	32459	BETAHIST
CHEMBL1196	209007	CHEMBL1196	209007	PROPARA
CHEMBL1073	139286	CHEMBL1073	139286	GLIPIZIDE

We will start by sketching a sulfonamide group and setting ring membership as R1 and an Attachment Point on the Carbon atom.



1. Right click on atom

2. Select Attachment point

3. Select ring membership R1

- Hydrogens
- Ring membership
  - Off
  - As Drawn R
  - R0
  - R1
  - R2
  - R3
- Ring size
- Charge
- Isotope
- Hybridization
- Aromaticity
- Connectivity

- Attachment point
- Edit Atom Label
- Delete

Table File Molcart

chembl\_cox2

Query Options

Search type Substructure

Max distance 0.4

Number of matches any

Match stereo  Ignore salt

Keep ring membership

100000

result  Overwrite

Highlight match  Rotate by match

Display as grid  Store query

Append

Hide after Search Search

Your query should look like this.

1. Our query is the sketch

2. Table to be searched

3. Select substructure search

4. Results will return to a new table called "result"

5. Click Search

The screenshot shows the ICM Chemical Search interface. At the top, the title bar reads "ICM Chemical Search: [new file \*] Molcart not connected". The main window is divided into several sections:

- Query Input:** A dropdown menu is set to "<draw>".
- Data Source:** A dropdown menu is set to "chembl\_cox2".
- Search Type:** A dropdown menu is set to "Substructure".
- Parameters:** A similarity threshold of "0.4" and "any" matches are selected. There are checkboxes for "Match stereo", "Ignore salt", "Selected only", and "Keep ring membership". The "Maximum # of hits" is set to "100000".
- Results:** Radio buttons for "Count hits only", "Hide unmatched", "Select in source", and "Save results to:". The "Save results to:" option is selected, and a dropdown menu is set to "result". There are checkboxes for "Overwrite", "Highlight match", "Display as grid", "Rotate by match", and "Store query". An "Append" checkbox is at the bottom.
- Search Button:** A "Search" button is located at the bottom right.

In the center, a chemical structure of a sulfonamide group is shown with the SMILES string [C\*]R(=O)S(=O)(=O)N. The interface also includes a sidebar with a list of elements (C, N, O, F, P, S, Cl, Br, I, B) and a table at the bottom with columns for "FieldName" and "Relation".

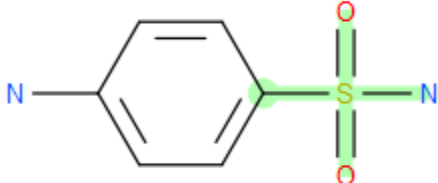
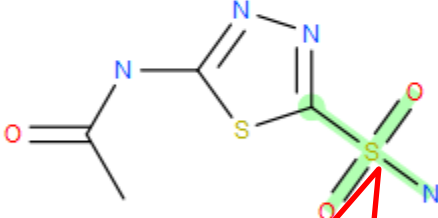
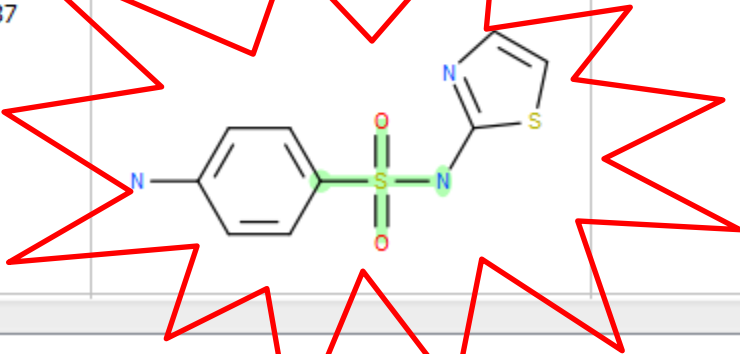
Results are shown with the query highlighted in green. We are finding chemicals with sulfonamide linked between two substituents e.g Phenyl-Sulfonamide-Thiazole shown - let us filter these out (next step).

cheminformatics\_tutorial Molsoft icm 3.8-1 [H:\icmd\man\cheminformatics\_tutorial

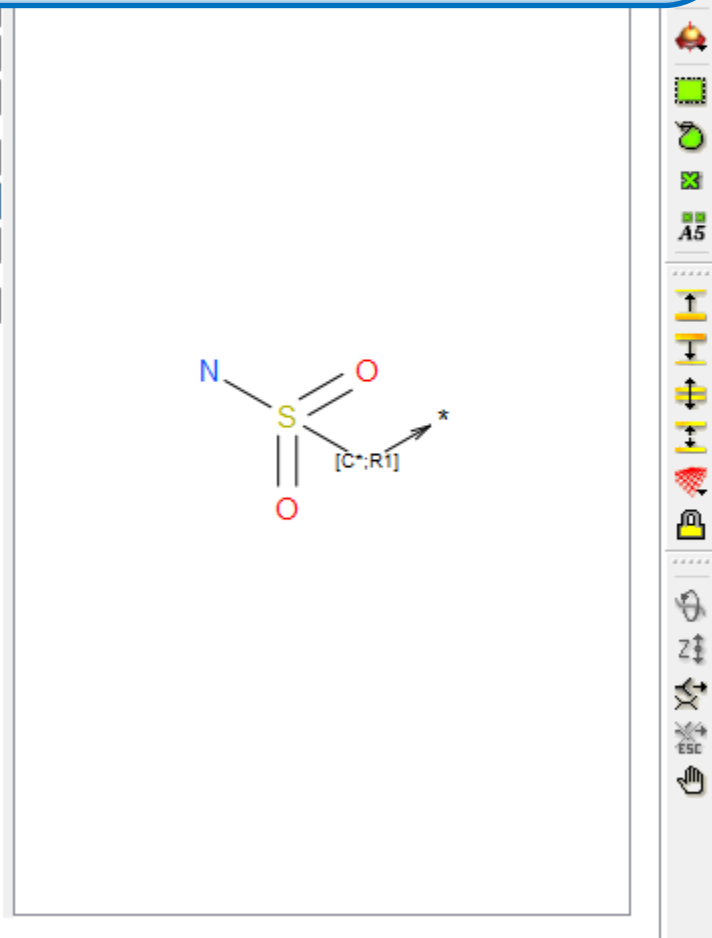
File Edit View Bioinfo Tools Homology Chemistry Docking MolMed

display light labels meshes search ligedit

chembl\_cox2 result

CMPD	CHEMBLID	mol	MOLREGNO
1	CHEMBL21		
2	CHEMBL20		1125 CHEI
3	CHEMBL437		2095 CHEI

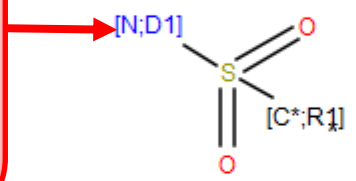
1 non-ICM Obj





By changing the connectivity of the Nitrogen we can find only terminal sulfonamide groups similar to the drug Celebrex.

1. Right click on the Nitrogen atom and choose Connectivity "As Drawn"



Data Source

Table File Molcart

chembl\_cox2

Query Options

Search type Substructure

Max distance 0.4

Number of matches any

Match stereo  Ignore salt

Selected only  Keep ring membership

Maximum # of hits 100000

Results

Count hits only  Select in source

Hide unmatched  Save results to:

Table File Molcart

result  Overwrite

Highlight match  Rotate by match

Display as grid  Store query

Append

Hide after Search

Search

FieldName	Relation	Value
-----------	----------	-------

Exclude fragment list:

Text search

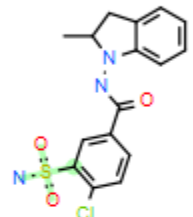
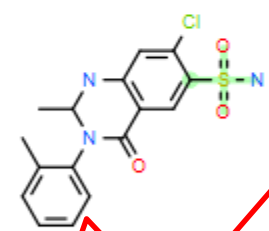
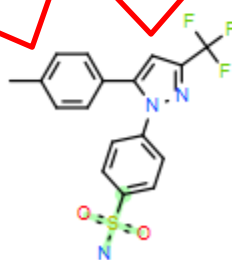
By changing the connectivity of the Nitrogen we can find only terminal sulfonamide groups similar to the drug Celebrex.

cheminformatics\_tutorial Molsoft icm 3.8-1 [H:\icmd\man\cheminformatics\_tutorial

File Edit View Bioinfo Tools Homology Chemistry Docking MolMed

display light labels meshes search ligedit

chembl\_cox2 result

CMPD	CHEMBLID	mol	MOLREGNO	PAR
10	CHEMBL406		racemic	619 CHEI
11	CHEMBL878		racemic	73548 CHEI
12	CHEMBL118			18694 CHEI

Header Filter Tree Scaffold

[N:D1] S O O [C\*:R1]

1 non-ICM Obj

Now let us expand the search a bit by including sulfonyl groups as well as sulfonamides – this will identify Vioxx like molecules as well.

1. Right click on the Nitrogen atom and choose Element/Custom and add "N,C"

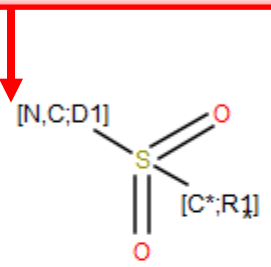


Table File Molcart

chembl\_cox2

Query Options

Search type Substructure

Max distance 0.4

Number of matches any

Match stereo  Ignore salt

Selected only  Keep ring membership

Maximum # of hits 100000

Results

Count hits only  Select in source

Hide unmatched  Save results to:

Table File Molcart

result  Overwrite

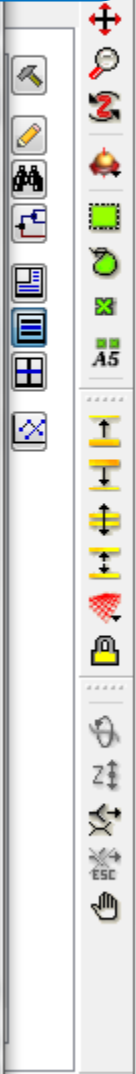
Highlight match  Rotate by match

Display as grid  Store query

Append

Hide after Search

Search



FieldName	Relation	Value
-----------	----------	-------

[Exclude fragment list:](#)

Text search



display light labels meshes search ligedit



chembl\_cox2 result

CMPD	CHEMBLID	mol	MOLREGNO	PAR
7				
8	CHEMBL122		21697	CHEI
9	CHEMBL1055		67010	CHEI
		racemic	134333	CHEI

Header Filter Scaffold

[N:C;D1]  
[C\*:R1]



# Chemical Clustering

**Cluster the ChEMBL\_Cox2 table by chemical substructure.**

**Clustering Parameters**

Table: chembl\_cox2

Name:

Take Labels from Column: MOL\_PREF\_NAME

Keep the existing trees

Add column with cluster numbers, name: d

Build for the selected 72 rows

Descriptors:

the "mol" column and all numerical columns     the mol column only (2D pharmacophore)

the selected columns only     use existing distance matrix:

the "mol" column only (linear fingerprints)

Method:

Tree    Memory required: 1 MB

Parameters:

Linkage type: UPGMA

Keep distance matrix

Compressed hierarchical clustering

OK    Cancel    Help

**1. Click here to cluster**

**2. Choose options as shown**

CMPD	CHEMBLID	mol
1	CHEMBL24441	<chem>N</chem>
2	CHEMBL1196	<chem>CCCCO</chem>
3	CHEMBL1073	<chem>C1=NC=NC=C1O</chem>

MOLREGNO	MOL PRE
32459	BETAHIST
209007	PROPARA
139286	GLIPIZIDE

The cluster is colored by branch and is fully interactive. Selections can be made in the tree or table. The cluster distance can be changed by dragging the bar as shown.

The screenshot shows the Molsoft icm 3.8-1 interface. The main window displays a table of compounds and a dendrogram. The table has columns for 'CMPD CHEMBLID', 'mol', 'MOLREGNO', and 'PA'. The dendrogram on the right shows a hierarchical clustering of compounds, with branches colored in yellow, cyan, and pink. A red box with the text '1. Drag to change distance' and an arrow points to a vertical bar in the dendrogram, indicating that the cluster distance can be adjusted by dragging this bar.

CMPD CHEMBLID	mol	MOLREGNO	PA
CHEMBL24441	<chem>CNCCc1cccnc1</chem>	32459	CHI
CHEMBL1201353	chiral <chem>CN(C)C[C@H](c1ccc(Cl)cc1)c2cccnc2</chem>	675304	CHI
CHEMBL3	chiral <chem>CN1CCCC1[C@@H]2C=CN=CC=C2</chem>	115	CHI

1. Drag to change distance

BETAHISTINE  
DEXCHLORPH..  
NICOTINE  
TACRINE  
CICLOPIROX  
DEFERIPRONE  
L-MIMOSINE  
PRALIDOXIM..  
CETYLPIRID..  
ISONIAZID  
IPRONIAZID  
NIACIN  
NIACINAMIDE  
METYRAPONE  
RISEDRONIC..  
IDOQUINOL  
OXYQUINOLINE  
CARBAMAZEP..  
PARAQUAT D..  
MILRINONE  
INAMRINONE  
PROPARACAINE  
PROCAINE  
TETRACAINE  
BUTAMBEN  
BENZOCAINE

1 non-ICM Obj

Use the right click selection tools to find clusters containing known COX-2 drugs.

cheminformatics\_tutorial Molsoft icm 3.8-1 [H:\icmd\man\cheminformatics\_tutorial



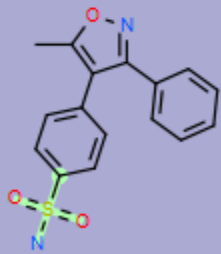
File Edit View Bioinfo Tools Homology Chemistry Docking MolMed

display light labels meshes search ligedit

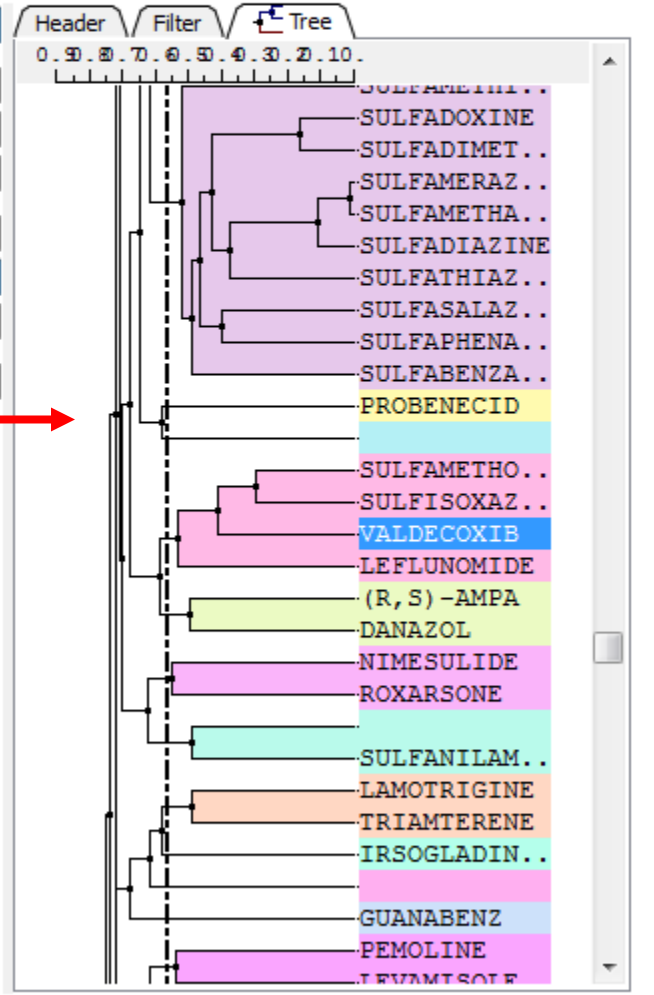
LH

RES ATOM VAR SITE

chembl\_cox2

	CMPD CHEMBLID	mol	MOLREGNO	PA
573	CHEMBL443			2247 CHI
574	CHEMBL453			
575	CHEMBL865			67010 CHI

1. Right click here for more clustering options



60

