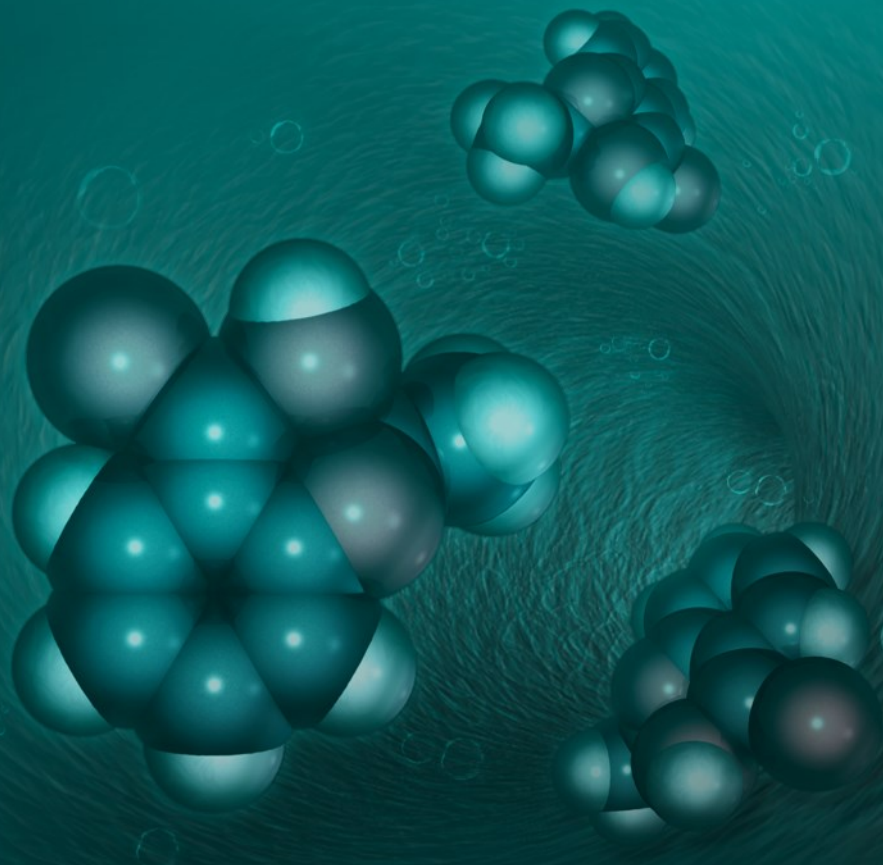


# ChEMBL Database: Open data for use in Toxicity Prediction

Anne Hersey

ChEMBL Group, EMBL-EBI



<https://www.ebi.ac.uk/chembl>



# Exercises

- Target Searching – finding data on a specific protein target
- Compound Searching – finding data on substructural fragments and their associated bioactivity data
- Assay and Document searching – finding data on more complex assays from text searching



# HANDS ON - Target Searching

- Exercise 1
  - You will get to know the ChEMBL Interface and the data returned from target searches
  - You will identify the experimental data available for compounds interacting with specific protein targets



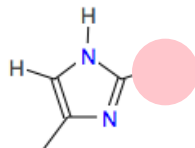
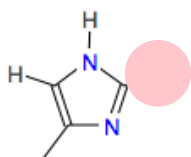
# Exercise 1

- 1) Search for hERG using a keyword search. How many compounds are bioactivity results are available?
- 2) Look at the target report card for hERG.
- 3) Display the Ki bioactivity data
- 4) Download all the Ki data into excel. Look at data (smiles, standard\_values, pChEMBL, properties, cmpds\_chemblid, parent\_cmpd\_chemblid)
- 5) Other ion channels are believed to be associated with Torsade-de-Pointes arrhythmias such as Nav1.5 ((SCN5A, Uniprot\_id: Q14524) and Cav1.2 (CACNA1C, Uniprot\_ID:Q14524). (Cardiovasc. Res., (2011),91,53-61)
- 6) Search for these in different ways in ChEMBL.  
Name searches (different identifiers), Browse targets, Blast Search (hint: you can get the FASTA file from Uniprot <http://www.uniprot.org> )



# HANDS ON - Compound Searching

- Exercise 2
  - You will do some sub-structure searches and compare the bioactivity data for the compounds found in your searches.
  - Specifically we will investigate the difference between Cytochrome P450 inhibition of imidazoles substituted and unsubstituted between the nitrogens



## Exercise 2

1. Draw the unsubstituted imidazole
2. Do a substructure search
3. Select Plot View
4. Filter using property plot for compounds with  $MWT < 500$  and  $ALOGP < 5$   
(Hint: remember to update plot)
5. Select Table View
6. Display Bioactivities
7. Download to excel
8. Use excel filter to select targets (pref\_name column) which are Cytochrome targets
9. Paste cytochrome set to another workbook – label all rows unsubstituted
10. Repeat 1-9 with the substituted imidazole
11. Plot data to see if there are any differences in inhibition (pCHEMBL values are a quick way of looking at this)



# HANDS ON – Assay/Document Searching

## Exercise 3

- You will identify potentially interesting datasets for more complex end-points by searching the assay or document descriptions
- Specifically we will search for:
- Ames test experimental data
- TG-GATES biochemistry data



## Exercise 3 - Ames Search

1. Type Ames into search box and select assays
2. Look at types of data identified (you can filter out any assays that don't look relevant at this point)
3. Select to display bioactivity data
4. Use hide/unhide options to display activity\_comment
5. Download data into excel





## Exercise 3 - Search for TG-GATES data

1. Select just this data source (hint: select activity source filter)
2. To search for all data for this source type \* into search box and select documents
3. Select to view the biochemistry dataset via the document report card
4. Display the bioactivity data
5. Use hide/unhide to view activity comments



# Hands On - Answers



# Exercise 1

Input search term here

Select type of data for search

ChEMBL

hERG

Compounds Targets Assays Documents Activity Source Filter

ChEMBL Target Search Results: 2

10 records per page

Show / hide columns

ChEMBL ID	Preferred Name	UniProt Accession	Target Type	Organism	Compounds	Bioactivities
<a href="#">CHEMBL240</a>	HERG	<a href="#">Q12809</a>	SINGLE PROTEIN	Homo sapiens	10565	13495
<a href="#">CHEMBL2362956</a>	Voltage-gated potassium channel	<a href="#">Q43525</a> , <a href="#">Q9UQ05</a> , <a href="#">Q9ULS6</a> , <a href="#">Q9ULD8</a> , <a href="#">Q9UK17</a> , <a href="#">Q9UJ96</a> , <a href="#">Q9UIX4</a> , <a href="#">Q9NZV8</a> , <a href="#">Q9NSA2</a> , <a href="#">Q9NS40</a> , <a href="#">Q9NR82</a> , <a href="#">Q9H3M0</a> , <a href="#">Q9H252</a> , <a href="#">Q9BQ31</a> , <a href="#">Q96RP8</a> , <a href="#">Q96PR1</a> , <a href="#">Q14721</a> , <a href="#">Q14003</a> , <a href="#">Q12809</a> , <a href="#">Q09470</a> , <a href="#">Q03721</a> , <a href="#">P56696</a> , <a href="#">P51787</a> , <a href="#">P48547</a> , <a href="#">P22460</a> , <a href="#">Q96L42</a> , <a href="#">Q96KK3</a> , <a href="#">Q92953</a> , <a href="#">Q8TDN2</a> , <a href="#">Q8TDN1</a> , <a href="#">Q8TAE7</a> , <a href="#">Q8NCM2</a> , <a href="#">Q6PIU1</a> , <a href="#">Q16322</a> , <a href="#">Q95259</a> , <a href="#">P16389</a> , <a href="#">P17658</a> , <a href="#">P22459</a> , <a href="#">P22001</a> , <a href="#">O43526</a>	PROTEIN FAMILY	Homo sapiens	0	0

Showing 1 to 2 of 2 entries

ChEMBL Statistics

- DB: ChEMBL\_17
- Targets: 9,356
- Compound records: 1,520,172
- Distinct compounds: 1,324,941
- Activities: 12,077,491

Select target to view



# hERG – Target Report Card

## Target Report Card

### Target Name and Classification

Target ID	CHEMBL240
Target Type	SINGLE PROTEIN
Preferred Name	HERG
Synonyms	ERG   ERG-1   ERG1   Eag homolog   Eag-related protein 1   Ether-a-go-go-related gene potassium channel 1   Ether-a-go-go-related protein 1   H-ERG   HERG   KCNH2   Potassium voltage-gated channel subfamily H member 2   Voltage-gated potassium channel subunit Kv11.1   hERG-1   hERG1
Organism	Homo sapiens
Species Group	No
Protein Target Classification	ion channel vgc vgc voltage c kcnh, kv10-12.x (ether-a-go-go)

### Target Components

Component Description	Relationship	Accession
Potassium voltage-gated channel subfamily H member 2	SINGLE PROTEIN	<a href="#">Q12809</a>

### Target Relations

ChEMBL ID	Pref Name	Target Type
<a href="#">CHEMBL2362996</a>	Voltage-gated potassium channel	PROTEIN FAMILY

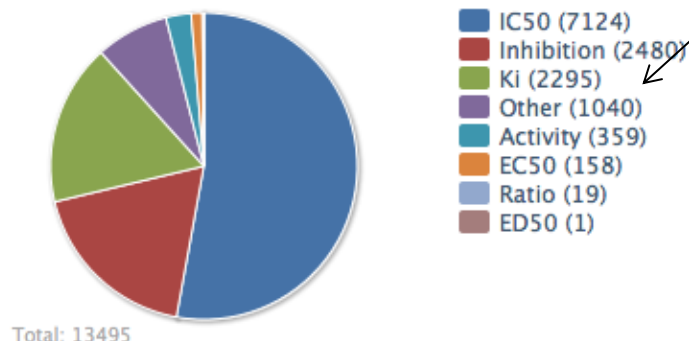
### Approved Drugs

ChEMBL ID	Name	Mechanism of Action	References
<a href="#">CHEMBL1083993</a>	AMIODARONE HYDROCHLORIDE	HERG blocker	<a href="#">DailyMed</a>
<a href="#">CHEMBL473</a>	DOFETILIDE	HERG blocker	<a href="#">DailyMed</a>
<a href="#">CHEMBL1201729</a>	DRONEDARONE HYDROCHLORIDE	HERG blocker	<a href="#">Expert FDA</a>
<a href="#">CHEMBL1200564</a>	IBUTILIDE FUMARATE	HERG blocker	<a href="#">DailyMed</a> <a href="#">PubMed</a>
<a href="#">CHEMBL1700</a>	SOTALOL HYDROCHLORIDE	HERG blocker	<a href="#">DailyMed</a>

## Target Associated Bioactivities

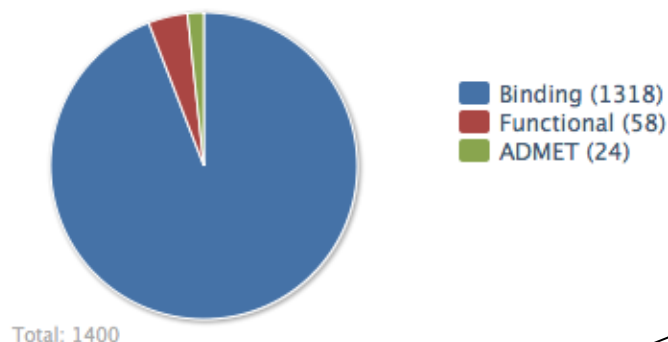
[Click here to see Ki data](#)

### ChEMBL Activity Types for Target CHEMBL240



## Target Associated Assays

### ChEMBL Assays for Target CHEMBL240



## Target Ligand Efficiencies

[Click here to see all data](#)

# Exercise 1- bioactivity data

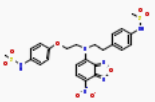
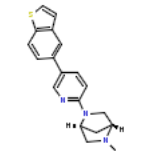
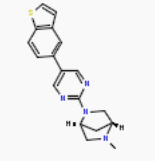
ChEMBL Bioactivity Search Results: 2295

100 records per page

✓ Please select....  
Download All Data (TAB)  
Download All Data (XLS)

Please select....

Show / hide columns

Ingredient	Molweight	Standard Type	Relation	Standard Value	Standard Units	pChEMBL Value	Assay Type	Description	Assay Src Description	Assay Organism	Target Type	Target Name	Target Organism	Reference
 <a href="#">CHEMBL227134</a>	590.63	Ki	=	.39	nM	9.41	B	<a href="#">Displacement of [3H]dofetilide from hERG expressed in HEK293 cells by SPA</a>	Scientific Literature	Homo sapiens	SINGLE PROTEIN	<a href="#">HERG</a>	Homo sapiens	<a href="#">J. Med. Chem., (2007) 50:13:2931</a>
 <a href="#">CHEMBL1085090</a>	321.44	Ki	=	.55	nM	9.26	B	<a href="#">Displacement of [3H]dofetilide from human ERG by whole-cell patch clamp</a>	Scientific Literature	Homo sapiens	SINGLE PROTEIN	<a href="#">HERG</a>	Homo sapiens	<a href="#">Bioorg. Med. Chem. Lett., (2010) 20:12:3636</a>
 <a href="#">CHEMBL1085806</a>	322.43	Ki	=	.56	nM	9.25	B	<a href="#">Displacement of [3H]dofetilide from human ERG by whole-cell patch clamp</a>	Scientific Literature	Homo sapiens	SINGLE PROTEIN	<a href="#">HERG</a>	Homo sapiens	<a href="#">Bioorg. Med. Chem. Lett., (2010) 20:12:3636</a>

Download file contains:

Parent smiles, parent and salt identifiers

Physchem descriptors

Ki values

pChEMBL values -  $-\log_{10}(\text{molar IC}_{50}, \text{XC}_{50}, \text{EC}_{50}, \text{AC}_{50}, \text{Ki}, \text{Kd} \text{ or Potency})$

# Exercise 1 – searching for other ion channels

Name search for target name, ChEMBL\_ID or Uniprot\_ID

The screenshot shows the ChEMBL Targets browser interface. At the top, there is a search bar labeled "Search ChEMBL..." and a navigation menu with tabs for "Compounds", "Targets", "Assays", and "Documents". The "Targets" tab is selected and highlighted with a red box. Below the navigation menu, there are buttons for "Ligand Search", "Target Search", "Browse Targets", "Browse Drugs", "Browse Drug Targets", "Drug Approvals", and "About". The "Browse Targets" button is highlighted with a red box. Below these buttons, there are radio buttons for "Protein Target Tree" (selected) and "Taxonomy Tree". A note says "\* Click arrows or use keyboard arrows (on selected items) to navigate tree". Below this, there are buttons for "Fetch selected targets", "Clear Selections", "Select All", "Collapse All", and "Open All". To the right of these buttons is a "Search Tree" section with a text input field containing "CAC" and a "Search" button, both highlighted with a red box. On the left side, there is a hierarchical tree of target classes. The "Ion channel (487)" category is expanded, showing sub-categories like "LGIC (274)", "VGC (148)", "VGC (125)", "VOLT (119)", "CATIONIC (119)", "K (52)", "CA (38)", "CACN alpha-1, CaVx.x (22)", "CACN alpha-2 delta (15)", "CACNalpha2-delta (2)", "NA (29)", and "c-GMP (1)". The "CACN alpha-1, CaVx.x (22)" category is further expanded, showing sub-categories like "IP3 (14)", "TRP (11)", "RYP (14)", "KIR (8)", "SUR (7)", "K2P (5)", "ASIC (2)", "VIRAL (2)", "Cytosolic other (238)", and "Transporter (163)".

Search or browse for target name in target tree

# Exercise 1 – BLAST search

Find target in Uniprot

<http://www.uniprot.org>

UniProtKB

Search Blast Align Retrieve ID Mapping

Search in: Protein Knowledgebase (UniProtKB) Query: SCN5A Search Advanced Search Clear

1 - 25 of 177 results for SCN5A in UniProtKB sorted by score descending

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway Reduce sequence redundancy to 100%

Results Customize

Show only reviewed (113) (UniProtKB/Swiss-Prot) or unreviewed (84) (UniProtKB/TrEMBL) entries

Restrict term "scn5a" to protein family (3), gene name (63), protein name (3)

Entry	Entry name	Status	Protein names	Gene names	Organism
<input type="checkbox"/> Q14524	SCN5A_HUMAN	★	Sodium channel protein type 5 subunit alpha	SCN5A	Homo sapiens (Human)
<input type="checkbox"/> Q86V90	Q86V90_HUMAN	★	SCN5A protein	SCN5A	Homo sapiens (Human)
<input type="checkbox"/> Q9JYV9	SCN5A_MOUSE	★	Sodium channel protein type 5 subunit alpha	Scn5a	Mus musculus (Mouse)
<input type="checkbox"/> P15389	SCN5A_RAT	★	Sodium channel protein type 5 subunit alpha	Scn5a	Rattus norvegicus (Rat)
<input type="checkbox"/> K3W4N7	K3W4N7_MOUSE	★	Sodium channel protein type 5 subunit alpha	Scn5a	Mus musculus (Mouse)
<input type="checkbox"/> K9J7R3	K9J7R3_XENTR	★	Uncharacterized protein	scn5a	Xenopus tropicalis (Western clawed frog) (Silurana tropicalis)

Select target

Q14524 (SCN5A\_HUMAN) ★ Reviewed, UniProtKB/Swiss-Prot

Last modified September 18, 2013. Version 150. History...

Clusters with 100%, 90%, 50% identity Documents (6) Third-party data

Names Attributes General annotation Ontologies Interactions Ait products Sequence annotation Sequences References Web links Cross-refs Entry Info Documents Customize order

## Names and origin

Protein names

Recommended name:  
Sodium channel protein

```
>sp|Q14524|SCN5A_HUMAN Sodium channel protein type 5 subunit alpha OS=Homo sapiens GN=SCN5A PE=1 SV=2
MANLLPRGTSFRFRFTRESLAAIEKRMAKQAGSTTLQESREGLPEEAPRPOLDLQA
SKKLPDLVONPQELIGEPLEDDPPYSTQKTFIVLMKGTIFRFSATNALVLSFHPFI
SRANKVILVSLFNLGIMCTILTCVMAQADPPWIKVYEVFTATVYFESLVKILAG
PCLIAFTFLRDPNNWLFDSVIMAYTTFVDLGNVSALRTRFRLALKTSIVISGLKTV
GALTQSVKLLADVNLTVFCLSPFALIGLQFLMGNLRHKVRNFTALNTNOSVEADGLV
WESLDLYLSDPENYLLKNGTSDVLLCGNSSDACTCEPGRCLKAGENPDHGYTSFDSFAN
AFLLALFRMTQDCMERLVQQTLSRACKIYMIFFNLVIFLGSPYLVNLIAYVAMAYEEQN
QATTAETTEKEKPFQAMHMLKKEHEALTICVDPYSSASLESMLAPVSHERRKRRK
RMSSGCTSECGDRLPKSDSEDCPRAMNHLSTGRLSRTSMKPRSSSGSIPTPRRDLGSE
ADPADENSTAGESSESHSTLSPVWPLRRTSAQQGSPSPGSAQHALHGKKNSTVDCNGV
VSLLAGOPEATSPGSHLLRPVMLEHPDDTTTSEEPOGPQMLTQAPCVDOGEFPGARQ
RALSAVSVLTSALEESSRHKCPFCNRLAQRILIWECCPLMMSIKQGVKLVMDPFTD
LITWICIVLNTLPMALSHVNMSTSEFEMGLQVCMVFPFPAETFRILALDPYVYFQGG
WIFDSISVILSLMELGSRMSVGLRSPFLRLVFLAKSWPTLWLIKIONSVQALG
NLZVLAIIVFIFAVVGMQLFGKNYSBELRDSGLLPKRWMDPFHAFILIPRLCGEVI
ETMWDCEVSGSGLCLLVLLVHVIGNLVNLIPLALLSSFSADNLTPADEDEMNQLQ
LALARIQRLGRVFKRTTDFCCGLLRORPKPAALAAQOQLPSCIATPYSPPPPTEKVP
RVUNVALGAIPIBINVLLVCLIFLWIFSIHGVNLFACKPGRCINOTRODPLMNTIYVN
KSCQESLNLGELVYKVVNFDVYKAYLALLQVATFGKMDINVAVDSGVTEOPQW
EYNLVNVIYFVPIPIFGSFPFLNLFIQVLIIDNFNQKKLGQDQIFMTEQKKVYNAMK
LGSKKPKPIPRPLNKYQGFIDIVTKAFDVTIMFLICLNMVIMVETDDQSEPEINIL
AKINLLFVAIFPGECIVKLAALRHVYFTNSNIFDFVUVVILSVOTVLSDIQKYFSP
LFPVIRLARIGRLRLRGAQRGIRTLFLPMLMSLPALFNIGLLFLVMPITSIFGMANTA
YVKEAGLDMMFNFQTFANSGCLFQITTSACWDGLLSPINMCPFCBDFPLPNSNGSG
DCGSPAVGILFPTTYIIISFLIVNNYIAIILENFVATEESTPELSEDFDPMVEYIWEK
FDPEATFIEYSVLSDFADALSEPLAIAPNQISINMOLPMVSGDRHICMDILFAFTR
VLGSGEMDALKIQMEKFMANPSKISYEPIT7TLRRKHEEVSAMVIRAFRRHLLQS
LKHSFRLFRQAGSGLSEDAPEERGLIAYVMSNFSRPLGPPSSSISSTSPSYDSV
TRATSDNLQVQGSYSHSEDLADFPSPDRRESIV
```

Select sequence from  
FASTA file

[Compounds](#)[Targets](#)[Assays](#)[Documents](#)[Activity Source Filter](#)[Ligand Search](#)[Target Search](#)[Browse Targets](#)[Browse Drugs](#)[Browse Drug Targets](#)[Drug Approvals](#)[About](#)

## Protein Target Search

☒ Protein Sequence Search ☐ ChEMBL ID Search

TADMTNTAELLEQIPDLGQDVKDPEDCFTEGCVRRPCCAVDTTQAPGKVVWRLRKTCTYH  
IVEHSWFETFIIFMILLSSGALAFEDIYLEERKTIKVLLEYADKMFTYVVFVLEMLLKWVA  
YGFKKYFTNAWCWLDLIVDVSLVSLVANTLGAEMGPIKSLRTLRLRPLRLSRFEGM  
RVVVNALVGAIPSIMNVLLVCLIFWLIFSIMGVNLFAGKFGRCINQTEGDLPLNYTIVNN  
KSCQESLNLTGELYWTKVKVNFNDVVGAGYLALLQVATFKGWMIDIMYAAVDSRGYEEQPQW  
EYNLYMYIYFVIFIGSFFTLNLFIGVIIDNFNQKKKLGQDIFMTEEQKKYYNAMKK  
LGSKKPQKPIRPLNKYQGFIIDVTQAFDVTIMFLICLNMTMMVETDDQSPEKINIL  
AKINLLFVAIFTGECIVKLAALRHYFTNSWNIFDFVVVLSIVGTVLSDIQKYFFSPT  
LFRVIRLARIGRILRLIRGAKGIRTLFALMMSLPALFNIGLLFLVMFIYSIFGMANFA  
YVKWEAGIDDMFNFTFANSMLCLFQITTSAGWDGLLSPILNTGPPYCDPTLPNSNGSRG  
DCGSPAVGILFFTYYIIISFLIVNMYIAILENFSVATEESTEPLSEDDFDMFYEIWEK  
FDPEATQFIEYSVLSDFADALSEPLRIAKPNQISLINMDLPMVSGDRIHCMDILFAFTKR  
VLGESGEMDALKIQMEEKFMAANPSKISYEPITTLRRKHVEVSAMVIQRAFRRLHLLQRS  
LKHASFLFRQQAGSLSEEDAPEREGLIAYVMSENFSRPLGPPSSSISSTSFPPSYDSV  
TRATSDNLQVRGSDYSHSEDLADFPSPDRDRESIV

Blast search results in order of  
similarity to input sequence

Ignore low

[Fetch Targets](#)[Clear Form](#)

ChEMBL BLAST Search Results: 124

[Please select....](#) records per page[Show / hide columns](#)

Query Sequence	ChEMBL ID	Preferred Name	UniProt Accession	Target Type	Organism	# Compounds	# Endpoints	% Identity	BLAST Score	E-Value	
sp Q14524 SCN5A_HUMAN	<a href="#">CHEMBL1980</a>	Sodium channel protein type V alpha subunit	<a href="#">Sodium channel protein type V alpha subunit</a>	SINGLE PROTEIN	Homo sapiens	493	588	100	4177	0e+0	<input checked="" type="checkbox"/>
sp Q14524 SCN5A_HUMAN	<a href="#">CHEMBL2331043</a>	Sodium channel alpha subunit	<a href="#">Sodium channel alpha subunit</a>	PROTEIN FAMILY	Homo sapiens	0	0	100	4177	0e+0	<input checked="" type="checkbox"/>
sp Q14524 SCN5A_HUMAN	<a href="#">CHEMBL1770040</a>	Voltage-gated sodium channel type V alpha polypeptide	<a href="#">Voltage-gated sodium channel type V alpha polypeptide</a>	SINGLE PROTEIN	Oryctolagus cuniculus	1	1	96.19	584	0e+0	<input checked="" type="checkbox"/>
sp Q14524 SCN5A_HUMAN	<a href="#">CHEMBL2362984</a>	Voltage-sensitive sodium channel alpha-subunit	<a href="#">Voltage-sensitive sodium channel alpha-subunit</a>	SINGLE PROTEIN	Pedicular humanus	0	0	55.35	842	0e+0	<input checked="" type="checkbox"/>
sp Q14524 SCN5A_HUMAN	<a href="#">CHEMBL3866</a>	Sodium channel protein type V alpha subunit	<a href="#">Sodium channel protein type V alpha subunit</a>	SINGLE PROTEIN	Rattus norvegicus	69	71	94.01	3792	0e+0	<input checked="" type="checkbox"/>
sp Q14524 SCN5A_HUMAN	<a href="#">CHEMBL2096682</a>	Sodium channel alpha subunits; brain (Types I, II, III)	<a href="#">Sodium channel alpha subunits; brain (Types I, II, III)</a>	PROTEIN FAMILY	Homo sapiens	212	374	62.28	2433	0e+0	<input checked="" type="checkbox"/>
sp Q14524 SCN5A_HUMAN	<a href="#">CHEMBL4187</a>	Sodium channel protein type II alpha subunit	<a href="#">Sodium channel protein type II alpha subunit</a>	SINGLE PROTEIN	Homo sapiens	142	244	62.28	2433	0e+0	<input checked="" type="checkbox"/>
sp Q14524 SCN5A_HUMAN	<a href="#">CHEMBL2095171</a>	Sodium channel alpha subunits; brain (Types I, II, III)	<a href="#">Sodium channel alpha subunits; brain (Types I, II, III)</a>	PROTEIN FAMILY	Rattus norvegicus	204	344	62.07	2427	0e+0	<input checked="" type="checkbox"/>



# Exercise 2 – Substructure search

Ligand Search Target Search Browse Targets Browse Drugs Browse Drug Targets

Draw substructure

JSME Molecular Editor by Peter Ertl and Bruno Bienfait

Substructure Search 100% Fetch Compounds


Search returns 6340 compounds

Select chart view

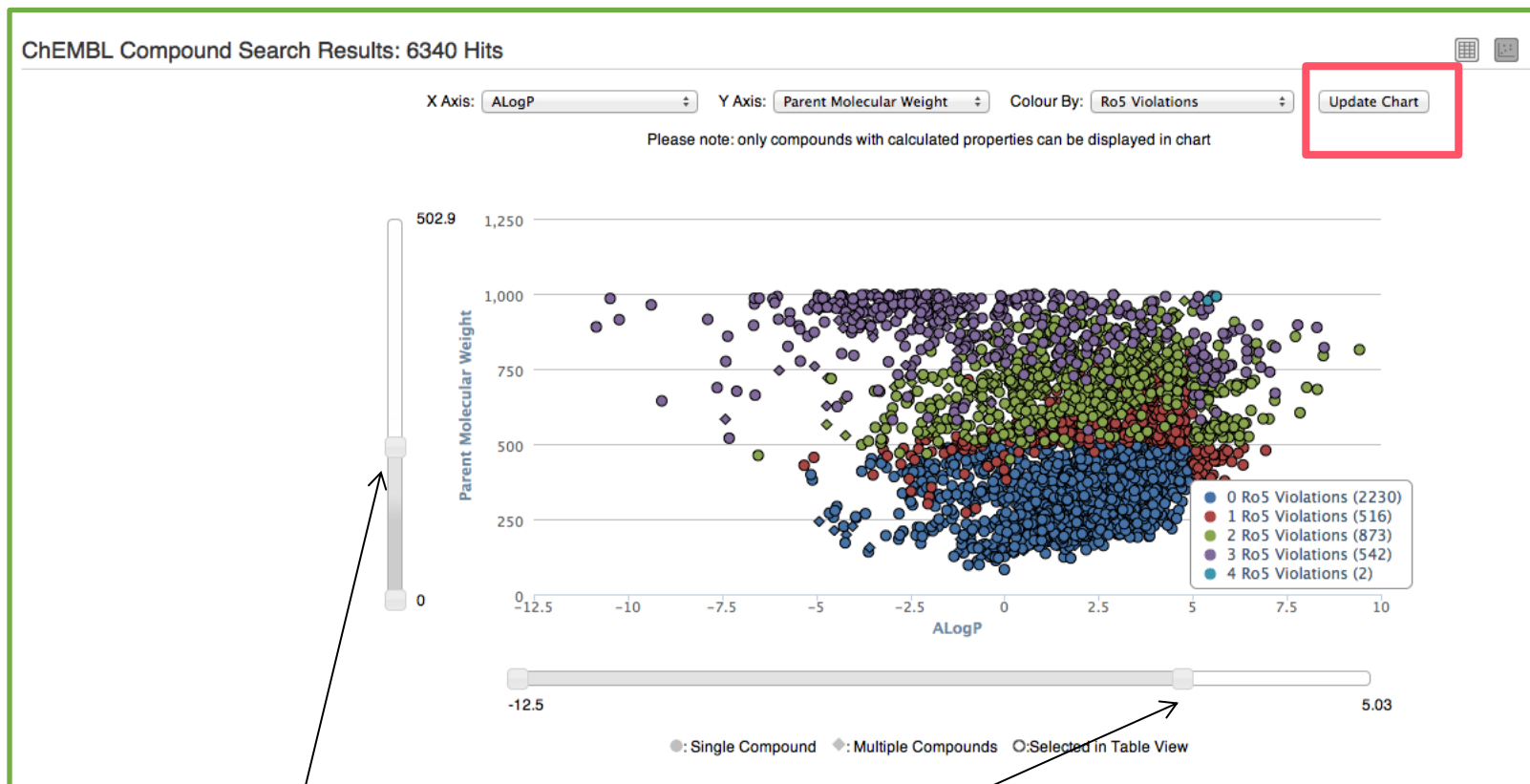
ChEMBL Compound Search Results: 6340 Hits

10 records per page

Show / hide columns

Compound	Synonyms	Max Phase	Parent Mol Weight	ALogP	PSA	HBA	HBD	#RO5 Vio.	#Rotatable Bonds	Passes Rule of Three	Med Chem Friendly	QED Weighted	
 CHEMBL428140		0	2348.73										<input checked="" type="checkbox"/>

# Exercise 2 – Filter on compound properties



Use sliders to filter on MWT and LogP

# Exercise 2 – Filtered dataset

Displaying only compounds with  
MWT<500 and LogP<5

ChEMBL Compound Search Results: 2291 Hits (6340 in Total [clear filter](#))

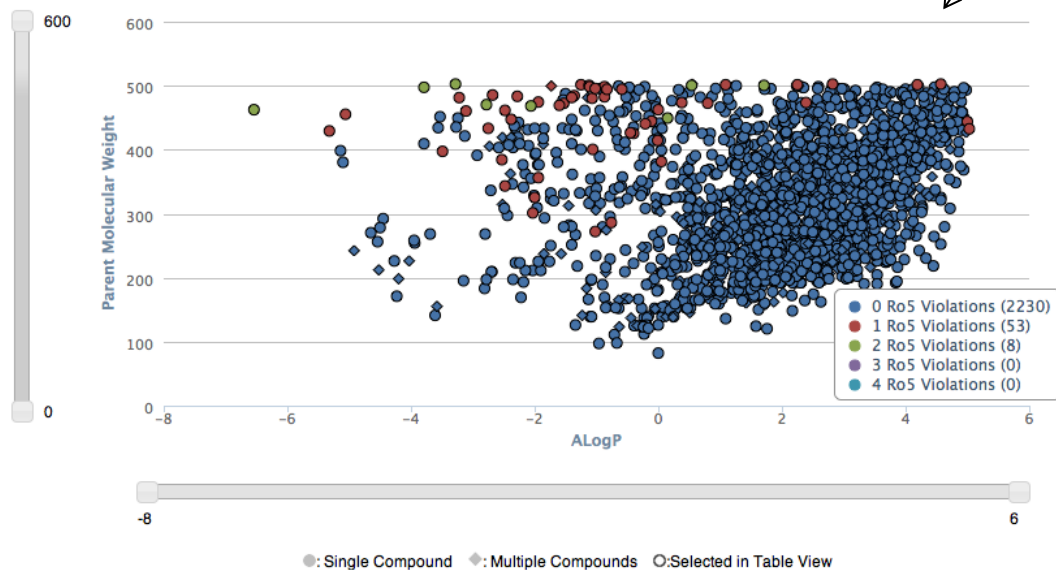
X Axis: ALogP

Y Axis: Parent Molecular Weight

Colour By: Ro5 Violations

Update Chart

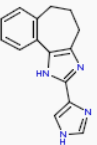
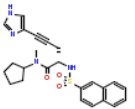
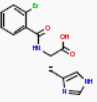
Please note: only compounds with calculated properties can be displayed in chart



# Exercise 2 – Table View of Filtered Compound Set

ChEMBL Compound Search Results: 2291 Hits (6340 in Total [clear filter](#))

10 records per page

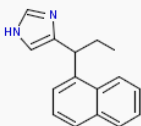
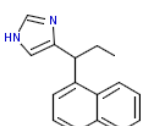
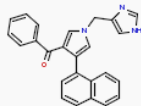
Compound	Synonyms	Max Phase	Parent Mol Weight	ALogP	PSA	HBA	HBD	#RO5 Vio.	#Rotatable Bonds	Passes Rule of Three	Me
 <a href="#">CHEMBL7147</a>		0	250.3	2.83	57.36	2	2	0	1	Y	Y
Chiral  <a href="#">CHEMBL7201</a>		0	450.55	3.26	103.54	4	2	0	8	N	Y
Chiral  <a href="#">CHEMBL7140</a>		0	338.16	1.34	95.08	4	3	0	5	N	Y

# Exercise 2 – Bioactivity data on filtered compounds

Download data into excel

ChEMBL Bioactivity Search Results: 14028

10 records per page

Ingredient	Molweight	Standard Type	Relation	Standard Value	Standard Units	Assay Type	Description	Assay Src Description	Assay Organism	Target Type	Target Name	Target Organism	Reference
 CHEMBL102076	236.31	Ratio	=	3.48		B	<a href="#">Selectivity ratio of Ki of alpha-1 adrenergic receptor to alpha-2 receptor</a>	Scientific Literature		UNCHECKED	<a href="#">Unchecked</a>		<a href="#">J. Med. Chem., (1996) 39:15:3001</a>
 CHEMBL102076	236.31	Ki	=	165	nM	B	<a href="#">Alpha-2 adrenergic receptor binding affinity was tested against membrane preparations of rat brain using 0.2 nM [3H]rauwolscine as radioligand</a>	Scientific Literature		PROTEIN FAMILY	<a href="#">Adrenergic receptor alpha-2</a>	Rattus norvegicus	<a href="#">J. Med. Chem., (1996) 39:15:3001</a>
 CHEMBL102187	377.44	IC50	=	115	nM	F	<a href="#">Inhibition of [3H]FPP incorporation into Ha-ras protein by Farnesyltransferase</a>	Scientific Literature		PROTEIN COMPLEX	<a href="#">Protein farnesyltransferase</a>	Homo sapiens	<a href="#">Bioorg. Med. Chem. Lett., (2001) 11:22:2963</a>

Please select....

✓ Please select....  
Download All Data (TAB)  
Download All Data (XLS)



## Exercise 2 – Select just Cytochrome P450 data in excel

## Excel table of results

Set filter on pref\_name  
Data - filter

Y2453																
	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
1	DESCRIPTION	ASSAY_SF	ASSAY_SF	ASSAY_OI	ASSAY_S1	ASSAY_TA	CURATED	TID	TARGET	TARGET	PROTEIN	PREF_NAME	ORGANISM	CONFIDENCE	TARGET	APD NAME
870	PUBCHEM_E	7	PubChem Bic Homo sapiens			9606	Autocuration	11365	CHEMBL289	SINGLE PRQ P10635	Cytochrome P450 2D6		Homo sapiens	8	Homologous protein	
871	PUBCHEM_E	7	PubChem BioAssays				Autocuration	11365	CHEMBL289	SINGLE PRQ P10635	Cytochrome P450 2D6		Homo sapiens	8	Homologous protein	
872	PUBCHEM_E	7	PubChem BioAssays				Autocuration	11365	CHEMBL289	SINGLE PRQ P10635	Cytochrome P450 2D6		Homo sapiens	8	Homologous protein	
873	PUBCHEM_E	7	PubChem BioAssays				Autocuration	11365	CHEMBL289	SINGLE PRQ P10635	Cytochrome P450 2D6		Homo sapiens	8	Homologous protein	
874	Inhibition of h	1	Scientific Lite  Homo sapiens			9606	Autocuration	11365	CHEMBL289	SINGLE PRQ P10635	Cytochrome P450 2D6		Homo sapiens	9	Protein	
875	PUBCHEM_E	7	PubChem BioAssays				Autocuration	11365	CHEMBL289	SINGLE PRQ P10635	Cytochrome P450 2D6		Homo sapiens	8	Homologous protein	
876	PUBCHEM_E	7	PubChem BioAssays				Autocuration	11365	CHEMBL289	SINGLE PRQ P10635	Cytochrome P450 2D6		Homo sapiens	8	Homologous protein	
877	Inhibition of h	1	Scientific Lite  Homo sapiens			9606	Expert	20056	CHEMBL528	SINGLE PRQ P05181	Cytochrome P450 2E1		Homo sapiens	9	Protein	
878	Inhibitory con	1	Scientific Lite  Homo sapiens			9606	Expert	20056	CHEMBL528	SINGLE PRQ P05181	Cytochrome P450 2E1		Homo sapiens	9	Protein	
879	Inhibitory con	1	Scientific Lite  Homo sapiens			9606	Expert	20056	CHEMBL528	SINGLE PRQ P05181	Cytochrome P450 2E1		Homo sapiens	9	Protein	
880	PUBCHEM_E	7	PubChem BioAssays				Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
881	PUBCHEM_E	7	PubChem BioAssays				Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
882	PUBCHEM_E	7	PubChem BioAssays				Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
883	PUBCHEM_E	7	PubChem BioAssays				Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
884	PUBCHEM_E	7	PubChem BioAssays				Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
885	PUBCHEM_E	7	PubChem BioAssays				Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
886	PUBCHEM_E	7	PubChem Bic Homo sapiens			9606	Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
887	PUBCHEM_E	7	PubChem Bic Homo sapiens			9606	Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
888	PUBCHEM_E	7	PubChem BioAssays				Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
889	PUBCHEM_E	7	PubChem Bic Homo sapiens			9606	Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
890	PUBCHEM_E	7	PubChem Bic Homo sapiens			9606	Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
891	PUBCHEM_E	7	PubChem BioAssays				Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
892	PUBCHEM_E	7	PubChem BioAssays				Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	8	Homologous protein	
893	PUBCHEM_E	7	PubChem Bic Homo sapiens			9606	Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	9	Protein	
894	PUBCHEM_E	7	PubChem Bic Homo sapiens			9606	Autocuration	17045	CHEMBL340	SINGLE PRQ P08684	Cytochrome P450 3A4		Homo sapiens	9	Protein	

After filtering in excel 548 values



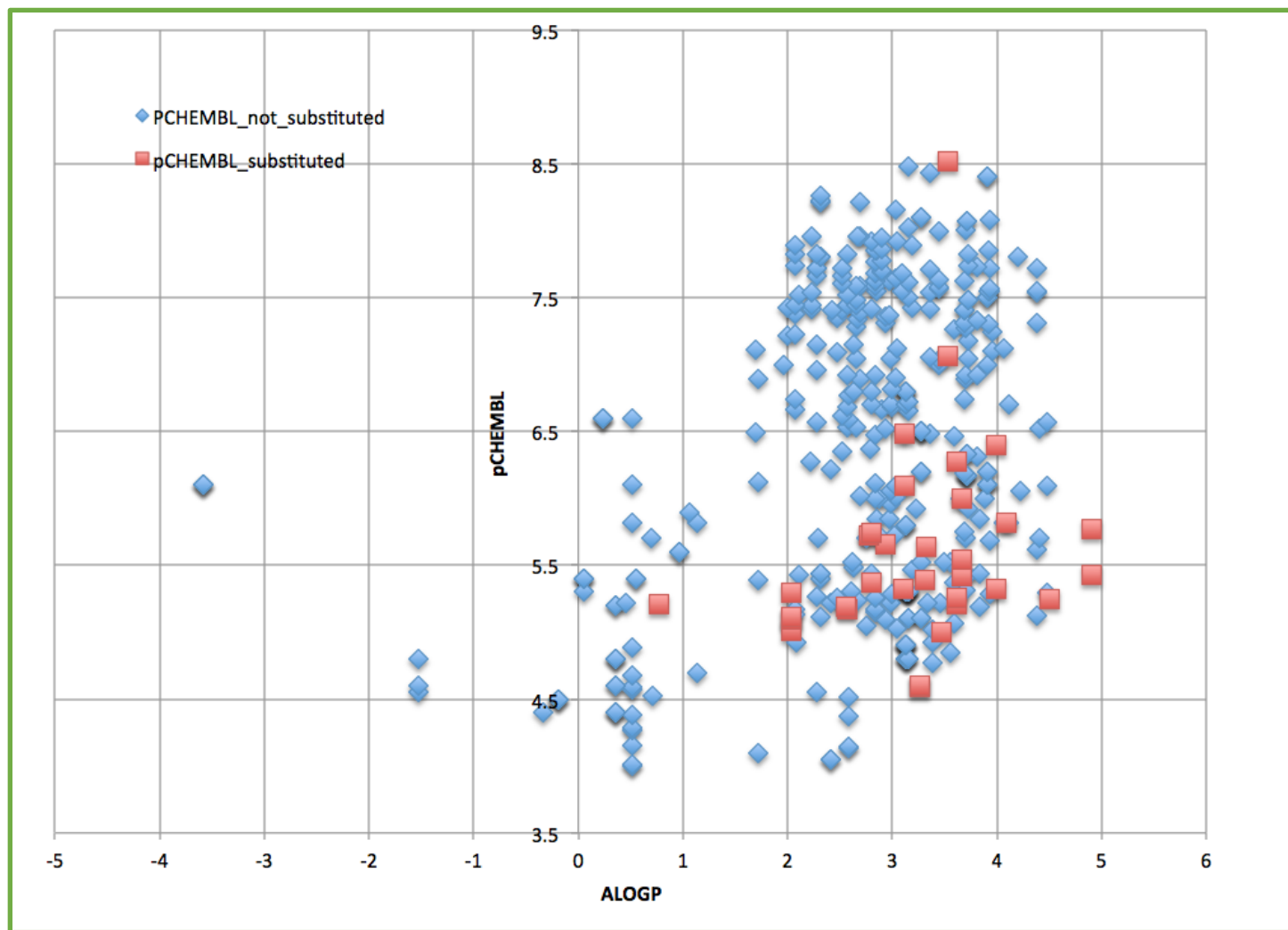
## Exercise 2 - Results from substituted imidazole search

- 1297 compounds
- 1035 compounds after ALOGP <5 and MWT <500 filter
- 8028 bioactivity values
- Filter only on Cytochromes 132 bioactivity values

File: imidazole\_example\_final.xlsx

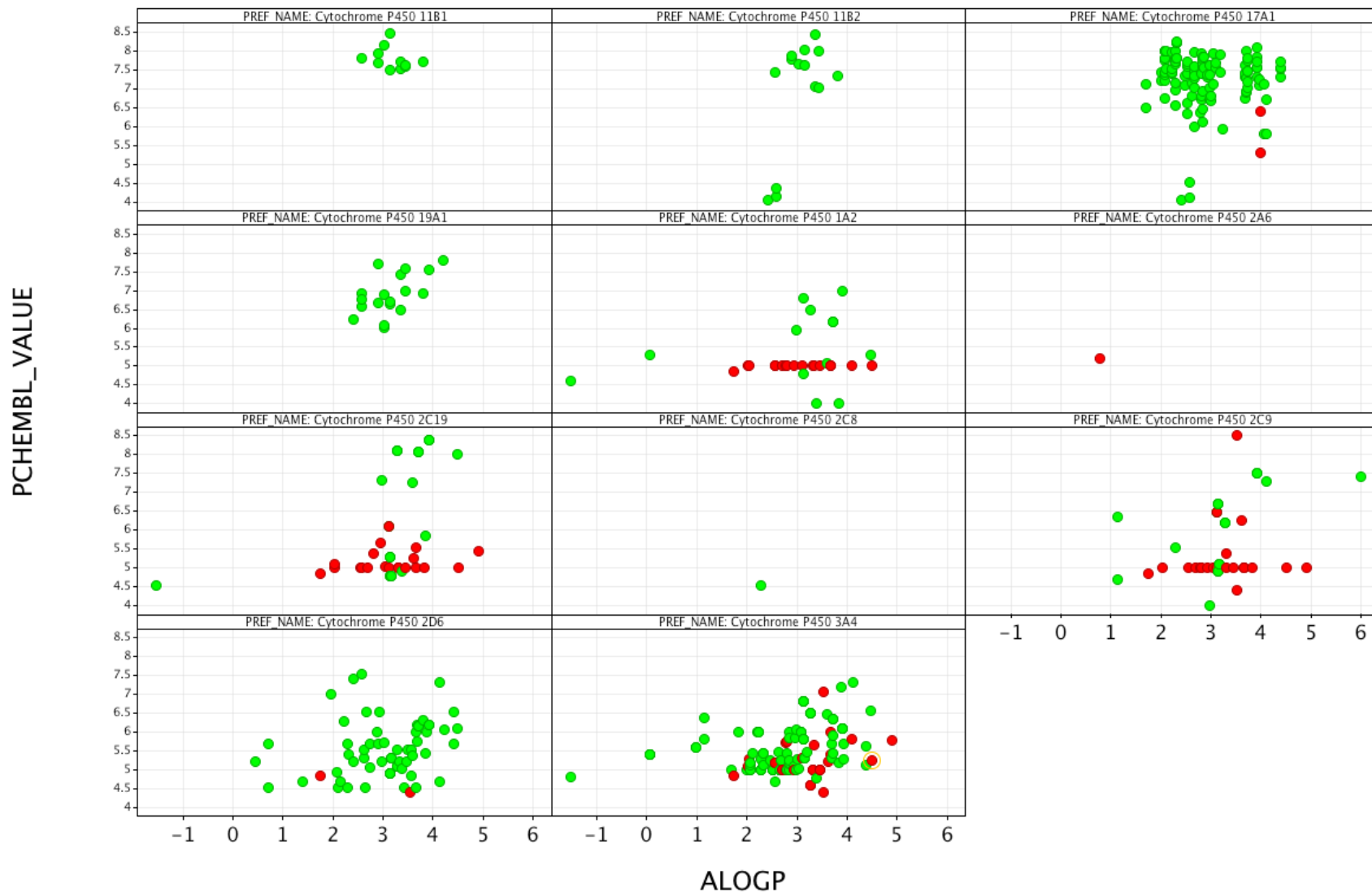


## Exercise 2 – View of Cytochrome P450 data on substituted and unsubstituted Imidazoles





Note also includes assignment of inactive (pCHEMBL=5) for appropriate compounds



# Exercise 3 – Assay search for Ames

Select “display bioactivities” to see data

ChEMBL Assay Search Results: 944

10 records per page

view 100 records per page

Please select....  
Download All (Tab-delimited)  
Download Selected (Tab-delimited)  
Display Bioactivities  
Filter Bioactivities

Please select....

Show / hide columns

ChEMBL Assay ID	Assay Source	Assay Type	Assay Organism	Description	Activity Count	Reference	
<a href="#">CHEMBL799656</a>	Scientific Literature	F	Salmonella enterica subsp. enterica serovar Typhimurium	Mutagenic activity in an Ames test on Salmonella Typhimurium TA98; Activity is log of revertants/nmol	207	<a href="#">J. Med. Chem., (1991) 34:2:786</a>	<input checked="" type="checkbox"/>
<a href="#">CHEMBL2210499</a>	Scientific Literature	A	Salmonella enterica subsp. enterica serovar Typhimurium	Genotoxicity in Salmonella typhimurium TA98 by Ames test in presence of S9 fractions	30	<a href="#">Bioorg. Med. Chem. Lett., (2012) 22:24:7357</a>	<input checked="" type="checkbox"/>
<a href="#">CHEMBL2211370</a>	Scientific Literature	A	Salmonella enterica subsp. enterica serovar Typhimurium	Genotoxicity in Salmonella typhimurium TAMix by Ames test in presence of S9 fractions	30	<a href="#">Bioorg. Med. Chem. Lett., (2012) 22:24:7357</a>	<input checked="" type="checkbox"/>
<a href="#">CHEMBL1262906</a>	Scientific Literature	F	Bacillus anthracis str. Ames	Antimicrobial activity against Bacillus anthracis Ames after 24 hrs	25	<a href="#">Bioorg. Med. Chem., (2010) 18:14:5137</a>	<input checked="" type="checkbox"/>
<a href="#">CHEMBL2210498</a>	Scientific Literature	A	Salmonella enterica subsp. enterica serovar Typhimurium	Genotoxicity in Salmonella typhimurium TA98 by Ames test in absence of S9 fractions	24	<a href="#">Bioorg. Med. Chem. Lett., (2012) 22:24:7357</a>	<input checked="" type="checkbox"/>
<a href="#">CHEMBL2211369</a>	Scientific Literature	A	Salmonella enterica subsp. enterica serovar Typhimurium	Genotoxicity in Salmonella typhimurium TAMix by Ames test in absence of S9 fractions	24	<a href="#">Bioorg. Med. Chem. Lett., (2012) 22:24:7357</a>	<input checked="" type="checkbox"/>

Deselect any data that looks irrelevant

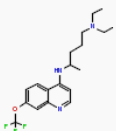
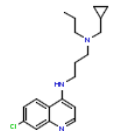
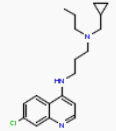
# Exercise 3 – bioactivity results

Select fields to view

ChEMBL Bioactivity Search Results: 2552

Please select....

10 records per page

Ingredient	Molweight	Standard Type	Relation	Standard Value	Standard Units	Assay Type	Description	Assay Src Description	Assay Organism	Target Type	Target Name	Target Organism	Relation
 <a href="#">CHEMBL1087464</a>	369.42	Activity				A	<a href="#">Mutagenic activity in Salmonella Typhimurium TA98 assessed as number of revertants after 48 hrs by Ames test</a>	Scientific Literature	Salmonella enterica subsp. enterica serovar Typhimurium	ADMET	<a href="#">ADMET</a>		
 <a href="#">CHEMBL1087469</a>	331.88	Activity				A	<a href="#">Mutagenic activity in Salmonella Typhimurium TA98 assessed as number of revertants after 48 hrs by Ames test</a>	Scientific Literature	Salmonella enterica subsp. enterica serovar Typhimurium	ADMET	<a href="#">ADMET</a>		
 <a href="#">CHEMBL1087469</a>	331.88	Activity				A	<a href="#">Mutagenic activity in Salmonella Typhimurium TA100 assessed as number of revertants after 48 hrs by Ames test</a>	Scientific Literature	Salmonella enterica subsp. enterica serovar Typhimurium	ADMET	<a href="#">ADMET</a>		

Show / hide columns

☒ Ingredient

☐ Molregno

☐ Parent

☐ Parent Molregno

☐ Molecule Name

☐ Name in Reference

☒ Molweight

☐ AlogP

☐ PSA

☐ #ROS Vio.

☐ Canonical SMILES

☐ Activity Id

☒ Standard Type

☒ Relation

☒ Standard Value

☒ Standard Units

☐ pChEMBL Value

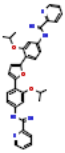
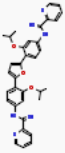
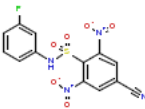
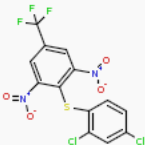
☒ Activity Comment

☐ Data Validity

☐ Potential Duplicate

For complex data **activity comment** often contains qualitative information e.g active/inactive

# Exercise 3 – Bioactivity Results

Ingredient	Molweight	Standard Type	Relation	Standard Value	Standard Units	Activity Comment	Assay Type	Description	Assay Src Description	Assay Organism	Target Type	Target Name	Target Organism	Reference
<a href="#">CHEMBL1275764</a>								<a href="#">metabolic activation system</a>						
 <a href="#">CHEMBL1275764</a>	574.67	Activity				Toxic	A	<a href="#">Induction of mutagenicity in Salmonella serovar Typhimurium TA 100 assessed as number of revertant colonies at 100 ug by Ames assay</a>	Scientific Literature	Salmonella enterica subsp. enterica serovar Typhimurium	ADMET	<a href="#">ADMET</a>		<a href="#">Antimicrob. Agents Chemother., (2010) 54:6:2507</a>
 <a href="#">CHEMBL1275764</a>	574.67	Activity				Toxic	A	<a href="#">Induction of mutagenicity in Salmonella serovar Typhimurium TA 100 assessed as number of revertant colonies at 500 ug by Ames assay</a>	Scientific Literature	Salmonella enterica subsp. enterica serovar Typhimurium	ADMET	<a href="#">ADMET</a>		<a href="#">Antimicrob. Agents Chemother., (2010) 54:6:2507</a>
 <a href="#">CHEMBL1276609</a>	366.28	Activity				Toxic	A	<a href="#">Genotoxicity in Salmonella Typhimurium by Ames test</a>	Scientific Literature	Salmonella enterica subsp. enterica serovar Typhimurium	ADMET	<a href="#">ADMET</a>		<a href="#">Eur. J. Med. Chem., (2010) 45:11:4879</a>
 <a href="#">CHEMBL458102</a>	413.16	Activity				Toxic	A	<a href="#">Genotoxicity in Salmonella Typhimurium by Ames test</a>	Scientific Literature	Salmonella enterica subsp. enterica serovar Typhimurium	ADMET	<a href="#">ADMET</a>		<a href="#">Eur. J. Med. Chem., (2010) 45:11:4879</a>

Results in activity comment field

# Exercise 3 – Activity Source Filter

EBI > Databases > Small Molecules > ChEMBL Database > Home

Search ChEMBL...

Compounds Targets Assays Documents **Activity Source Filter**

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

Select dataset to search

**Selected Bioactivity Sources**

Selected	Source	Assay Counts	Activity Counts
<input type="checkbox"/>	Scientific Literature	724530	4164854 (34.48%)
<input type="checkbox"/>	TP-search Transporter Database	3592	6765 (0.06%)
<input type="checkbox"/>	PubChem BioAssays	2308	7036920 (58.26%)
<input checked="" type="checkbox"/>	Open TG-GATEs	1408	181713 (1.5%)
<input type="checkbox"/>	Sanger Institute Genomics of Drug Sensitivity in Cancer	714	73169 (0.61%)
<input type="checkbox"/>	Millipore Kinase Screening	468	73944 (0.61%)
<input type="checkbox"/>	GSK Published Kinase Inhibitor Set	456	169451 (1.4%)
<input type="checkbox"/>	Guide to Receptors and Channels	344	801 (0.01%)
<input type="checkbox"/>	DrugMatrix in vitro pharmacology assays	132	229944 (1.9%)
<input type="checkbox"/>	Drugs for Neglected Diseases Initiative (DNDi)	62	11554 (0.1%)
<input type="checkbox"/>	MMV Malaria Box	23	5231 (0.04%)
<input type="checkbox"/>	Open Source Malaria Screening	21	226 (0%)
<input type="checkbox"/>	WHO-TDR Malaria Screening	16	5853 (0.05%)
<input type="checkbox"/>	St Jude Malaria Screening	16	5456 (0.05%)
<input type="checkbox"/>	Novartis Malaria Screening	6	27888 (0.23%)
<input type="checkbox"/>	GSK Malaria Screening	6	81198 (0.67%)
<input type="checkbox"/>	GSK Tuberculosis Screening	5	1406 (0.01%)
<input type="checkbox"/>	Harvard Malaria Screening	4	111 (0%)
<input type="checkbox"/>	Deposited Supplementary Bioactivity Data	4	1007 (0.01%)

☐ Check/Uncheck All

Update

# Exercise 3 –TG-GATES Datasets

ChEMBL Document Search Results: 4 Hits

Please select....

10 records per page

Show / hide columns

ChEMBL Doc ID	Journal Title	Doc Source Description	Pubmed ID	DOI	Year	Volume	Article Title	Abstract	
<a href="#">CHEMBL2364496</a>		Open TG-GATES					Open TG-GATES in vivo data: pathology		<input checked="" type="checkbox"/>
<a href="#">CHEMBL2095944</a>		Open TG-GATES					Open TG-GATES in-vivo data: biochemistry	Data generated by the Japanese Toxico Genomics Project. For more information see <a href="http://toxico.nibio.go.jp/open-tggates/search.html">http://toxico.nibio.go.jp/open-tggates/search.html</a> and Uehara T, Ono...	<input checked="" type="checkbox"/>
<a href="#">CHEMBL2095397</a>		Open TG-GATES					Open TG-GATES in-vivo data: hematology	Data generated by the Japanese Toxico Genomics Project. For more information see <a href="http://toxico.nibio.go.jp/open-tggates/search.html">http://toxico.nibio.go.jp/open-tggates/search.html</a> and Uehara T, Ono...	<input checked="" type="checkbox"/>
<a href="#">CHEMBL2095235</a>		Open TG-GATES					Open TG-GATES in-vivo data: organ_weight	Data generated by the Japanese Toxico Genomics Project. For more information see <a href="http://toxico.nibio.go.jp/open-tggates/search.html">http://toxico.nibio.go.jp/open-tggates/search.html</a> and Uehara T, Ono...	<input checked="" type="checkbox"/>

Select to view biochemistry data

# Exercise 3 – TG-GATES Biochemistry Data

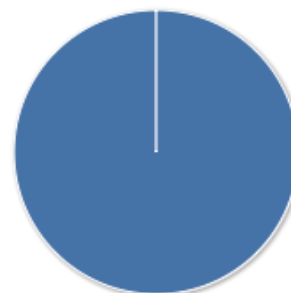
## Document Report Card

Doc ID	CHEMBL2095944
Title	Open TG-GATES in-vivo data: biochemistry
Authors	
Abstract	Data generated by the Japanese Toxicogenomics Project. For more information see <a href="http://toxico.nibio.go.jp/open-tg-gates/search.html">http://toxico.nibio.go.jp/open-tg-gates/search.html</a> and Uehara T, Ono A, Maruyama H, Ohno Y, Urushidani T., The Japanese toxicogenomics project: application to toxicogenomics., Mol Nutr Food Res. 2010 Feb;54(2):218-27

## Protein Target Summary

## Assay Summary

ChEMBL Assays for Doc CHEMBL2095944

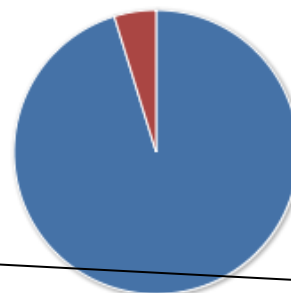


Total: 672



## Bioactivity Summary

ChEMBL Activity Types for Doc CHEMBL2095944



Total: 87675

[Display All Records](#)  
[ChEMBL Widget Details](#)  
[ChEMBL Homepage](#)



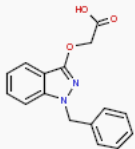
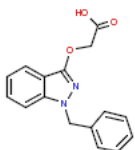
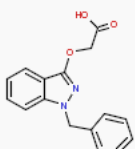
## Compound Summaries

Select to view biochemistry  
bioactivity data from TG-GATES

# Exercise 3 – Bioactivity Results

ChEMBL Bioactivity Search Results: 83500 Please select....

100 records per page Show / hide columns

Ingredient	Molweight	Standard Type	Relation	Standard Value	Standard Units	pChEMBL Value	Activity Comment	Assay Type	Description	Assay Src Description	Assay Organism	Target Type	Target Name	Target Organism	Reference
 <a href="#">CHEMBL1089221</a>	282.29	concentration	=	6419.2	IU/L		N = 5; SD = 4954.63; p-value = 0.07	A	<a href="#">TG_GATES: regimen: single; time: 24 hours; dose: high   dataset: biochemistry; assay: ALP (alkaline phosphatase)</a>	Open TG-GATES	Rattus norvegicus	ORGANISM	<a href="#">Rattus norvegicus</a>	Rattus norvegicus	<a href="#">CHEMBL2095944</a>
 <a href="#">CHEMBL1089221</a>	282.29	concentration	=	5387.2	IU/L		N = 5; SD = 4041.51; p-value = 0.06	A	<a href="#">TG_GATES: regimen: single; time: 9 hours; dose: high   dataset: biochemistry; assay: ALP (alkaline phosphatase)</a>	Open TG-GATES	Rattus norvegicus	ORGANISM	<a href="#">Rattus norvegicus</a>	Rattus norvegicus	<a href="#">CHEMBL2095944</a>
 <a href="#">CHEMBL1089221</a>	282.29	concentration	=	3422.2	IU/L		N = 5; SD = 2691.45; p-value = 0.09	A	<a href="#">TG_GATES: regimen: single; time: 6 hours; dose: middle   dataset: biochemistry; assay: ALP (alkaline phosphatase)</a>	Open TG-GATES	Rattus norvegicus	ORGANISM	<a href="#">Rattus norvegicus</a>	Rattus norvegicus	<a href="#">CHEMBL2095944</a>

Note size of data set: 83500 rows as multiple dose, time points, end points