Chemical Resources

Structure, Bioactivity and Metabolomics

Anne Hersey – ChEMBL Group

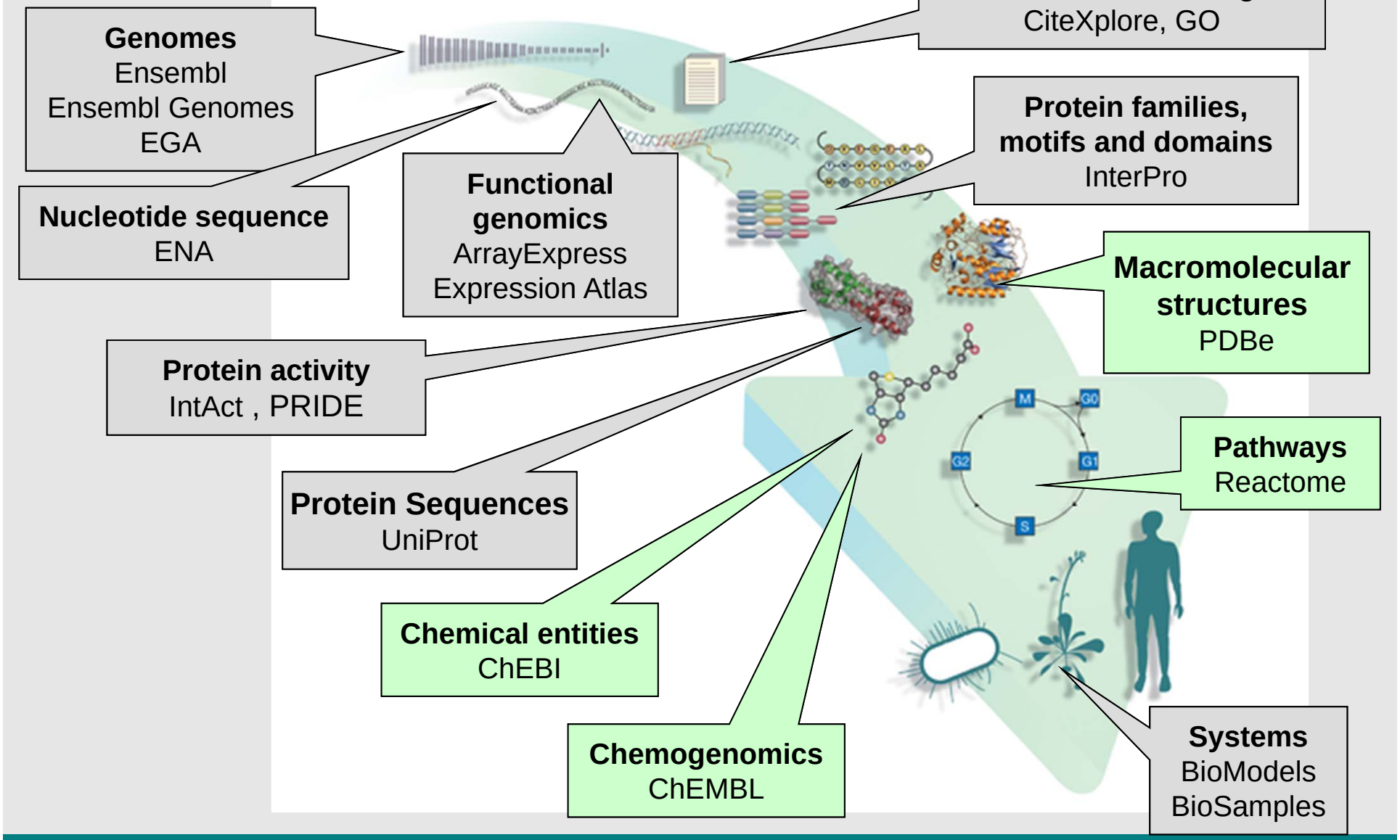
EMBL-EBI

EBI is an Outstation of the European Molecular Biology Laboratory.

Outline

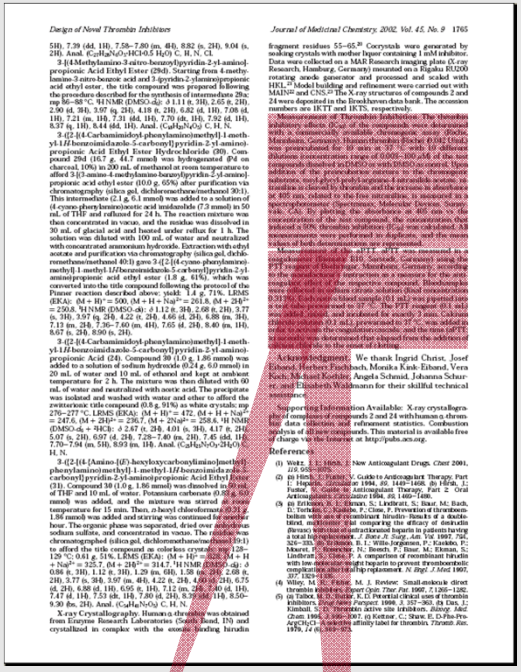
- ChEMBL Background and Content
- Use Cases to show how to find bioactivity data
  - Searching by Target
  - Searching by Structure
  - Finding data on related Targets
  - Data on Disease Processes and Pathways
- Other Resources

EBI Resources

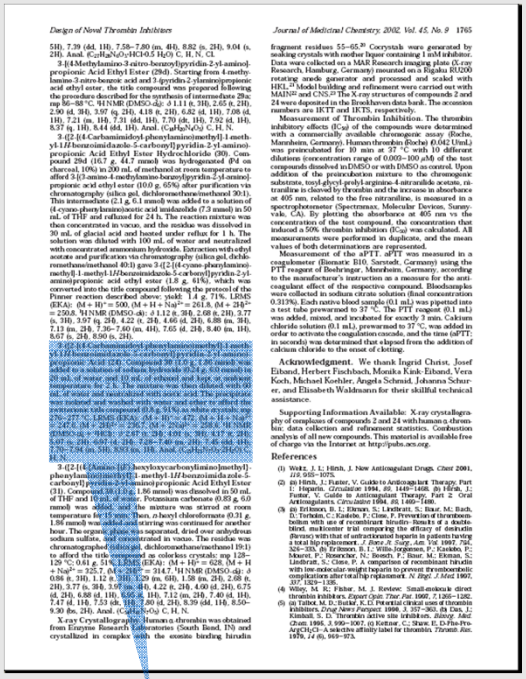


EMBL-EBI

What is the ChEMBL?



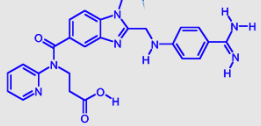
Compound



Target

>Thrombin

MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRRANTFLEEVRKGNLERECVEETCSY
EEAFEALESSTATDVFWAKYTACETARTPRDKLAACLEGNCAEGLGTNYRGHVNITRSGIECQLWRS
RYPHKPEINSTTHPGADLQENFCRNPDSSTTGPWCYTTDPTVRRQECSIPVCGQDQVTVAMTPRSEG
SSVNLSPPLEQCVPDRGQQYQGRLAVTTHGLPCLAWASAQAKALSKHQDFNSAVQLVENFCRNPD
EEGVWCYVAGKPGDFGYCDLNYCEEAVEEETGDGLDEDSDRAIEGRTATSEYQTFFNPRTFGSGE
CGLRPLFEKKSLEDKTERELLESYIDGRIVEGSDAEIGMSPWQVMLFRKSPQELLCGASLISDRW
TAAHCLLYPPWDKNFTENDLLVRIGKHSRTRYERNIEKISMLEKIYIHPRYNWRENLDRDIALMK
KPVAFSDYIHPVCLPDRETAASLLQAGYKGRVTGWGNLKETWTANVGKGQPSVLQVVNLPIVERP
KDSTRIRITDNMFCAGYKPDEGKRGDACEGDSGGPFVMKSPFNNRWYQMGIVSWGEGCDRDGKYG
THVFRLKKWIQKVIDQFGE



Bioactivity

Compound

Assay

SAR Data

Ki=4.5 nM

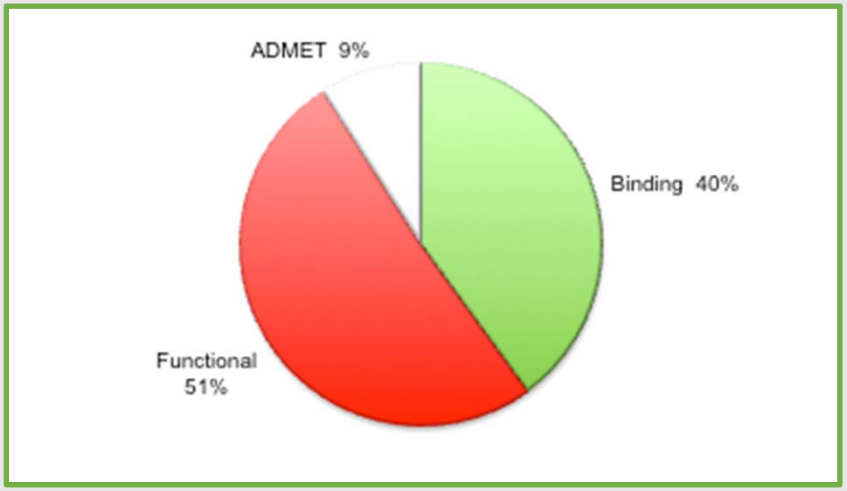
APTT 11 min

EMBL-EBI

ChEMBL Database

- Freely available (searchable and downloadable) resource for drug discovery
- Updated regularly with new data
- Secure searching (*https://www.ebi.ac.uk/chembldb*)

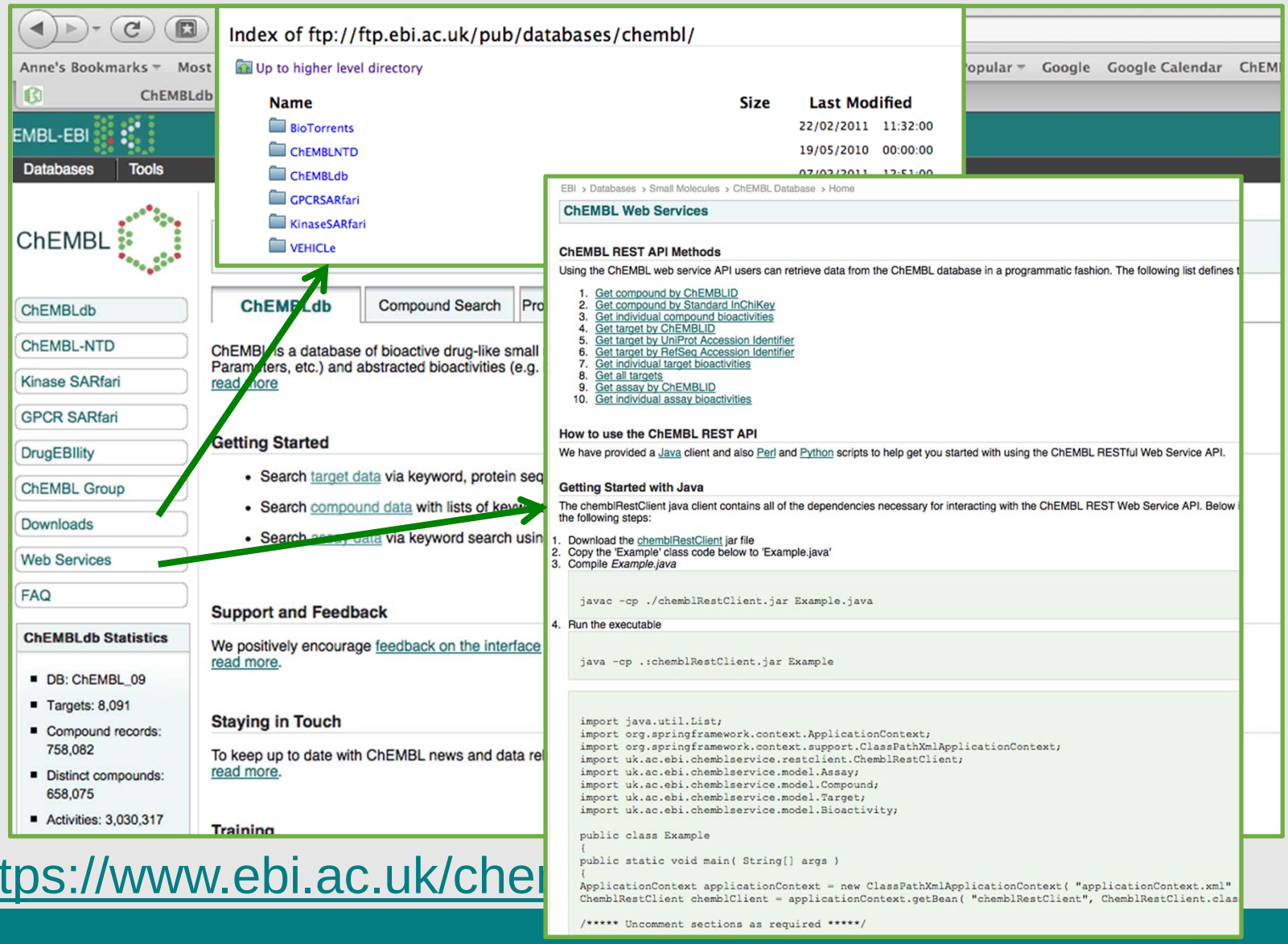
Assays are classified as:
- Binding measurements
- Functional assays
- ADME/toxicity data



ChEMBL11
Targets: 8,603
Compounds: 1,060,258
(~350K from PubChem)
Activities: 5,479,146
Publications: 42,516

5420 proteins
1481 organisms
1431 cell lines

Accessing ChEMBL Data



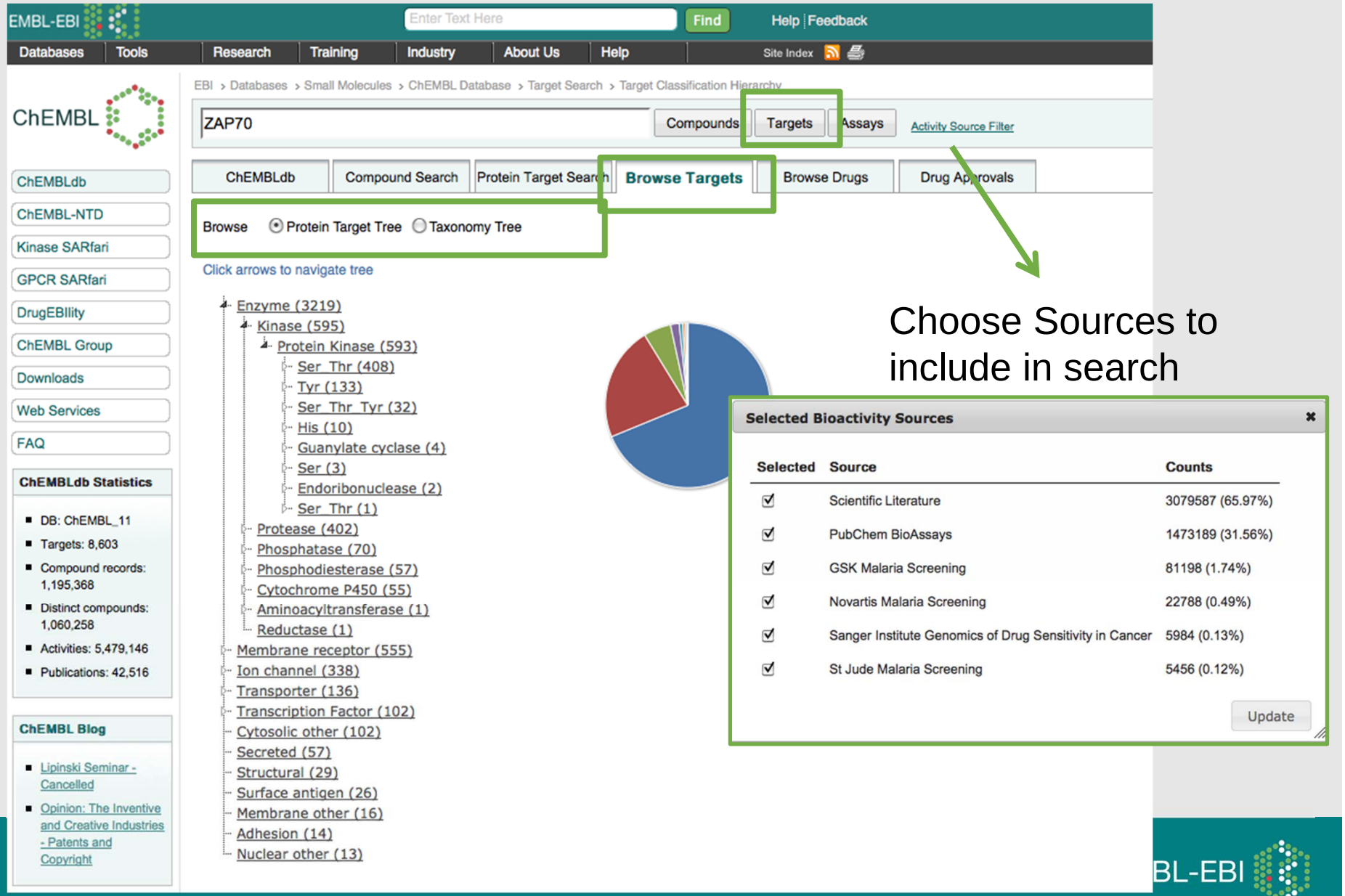
https://www.ebi.ac.uk/chem

EMBL-EBI

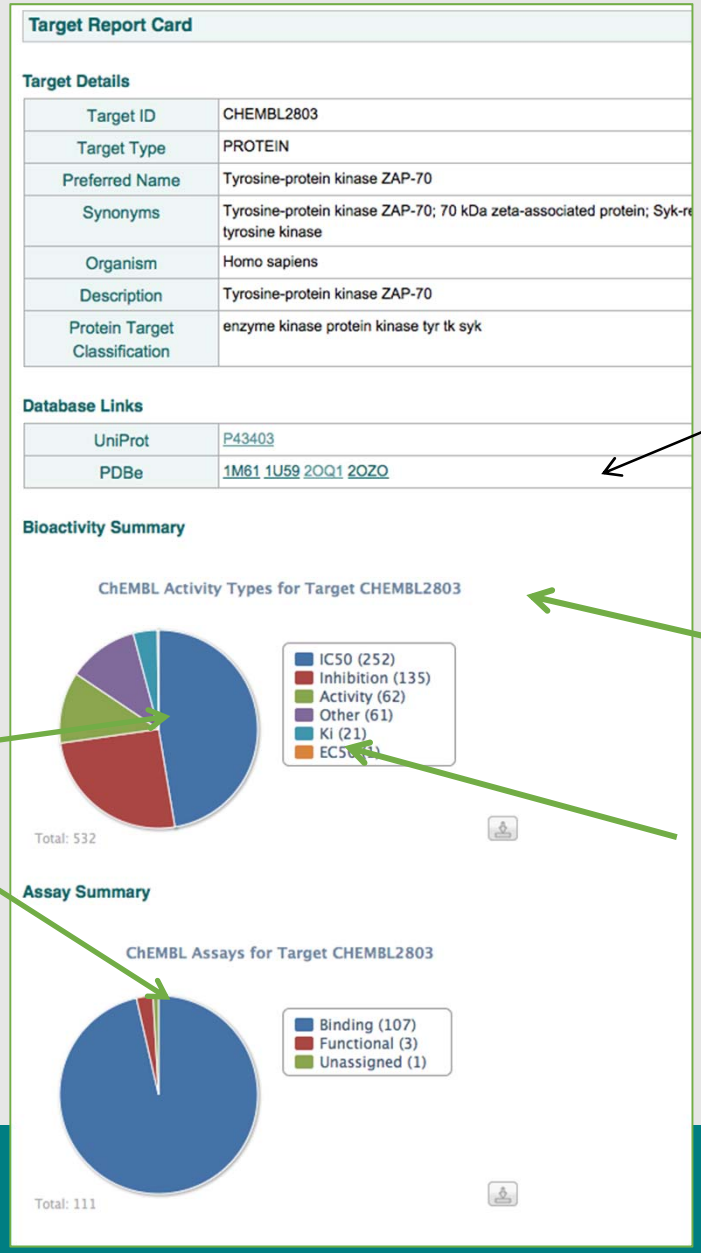
Use Case 1 - Searching by Target

- What is known about chemical structures that bind to a specific protein (ZAP70)?
- What is known about their potency/selectivity/ADMET Properties
- Is there any protein structure data?

Use Case 1 Searching by target in ChEMBL



Retrieving bioactivity details - single target



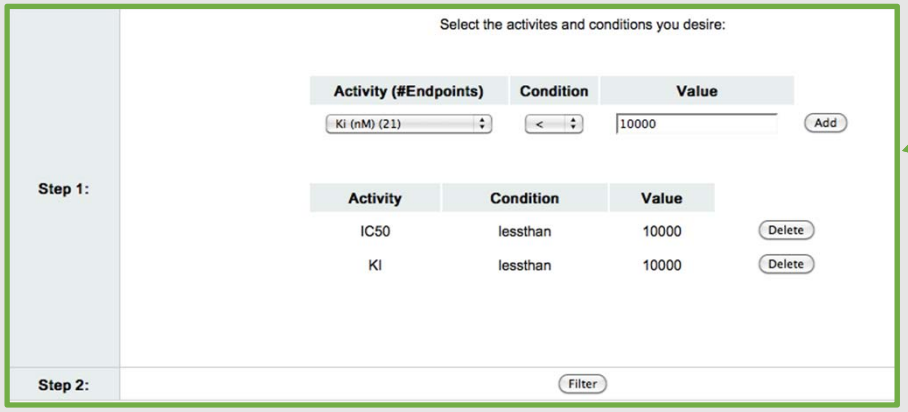
3D Structures

Display all bioactivity data for target

Summary of bioactivity/assay/compound data available for target

Click pie chart to retrieve particular end-points

EMBL-EBI

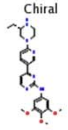
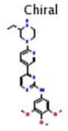
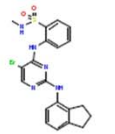
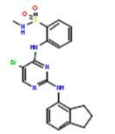
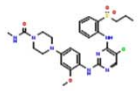
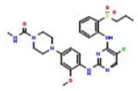


Filter on IC50 <10uM

168 ZAP70 IC50 or Ki values <10uM

ChEMBL Bioactivity Search Results: 168

1 2 3 4 5 6 [Next] [End] Please select....

Parent	Ingredient	Bioactivity	Activity Comment	Operator	Value	Units	Assay ChEMBL ID	Assay Source	Assay Type	Description	ChEMBL Target ID	Target Name	Organism	Target Mapping	Curated By	Reference	Name in Reference
Chiral  CHEMBL112346	Chiral  CHEMBL112346	IC50		=	8	nM	CHEMBL824029	Scientific Literature	B	Inhibition of Zeta-chain (TCR) associated protein kinase 70 kDa phosphorylation of polyGly-Tyr.	CHEMBL2803	Tyrosine-protein kinase ZAP-70	Homo sapiens	Homologous protein	Expert	Bioorg. Med. Chem. Lett., (1999) 9:23:3351	20
 CHEMBL575048	 CHEMBL575048	IC50		=	10	nM	CHEMBL1037893	Scientific Literature	B	Inhibition of human GST-fused ZAP-70 expressed in Sf9 cells	CHEMBL2803	Tyrosine-protein kinase ZAP-70	Homo sapiens	Protein	Intermediate	Eur. J. Med. Chem., (2009) 44:12:4793	26
 CHEMBL573483	 CHEMBL573483	IC50		=	10	nM	CHEMBL1037893	Scientific Literature	B	Inhibition of human GST-fused ZAP-70 expressed in Sf9 cells	CHEMBL2803	Tyrosine-protein kinase ZAP-70	Homo sapiens	Protein	Intermediate	Eur. J. Med. Chem., (2009) 44:12:4793	1

Compound structures

Activity values

Assay details

Target details

References

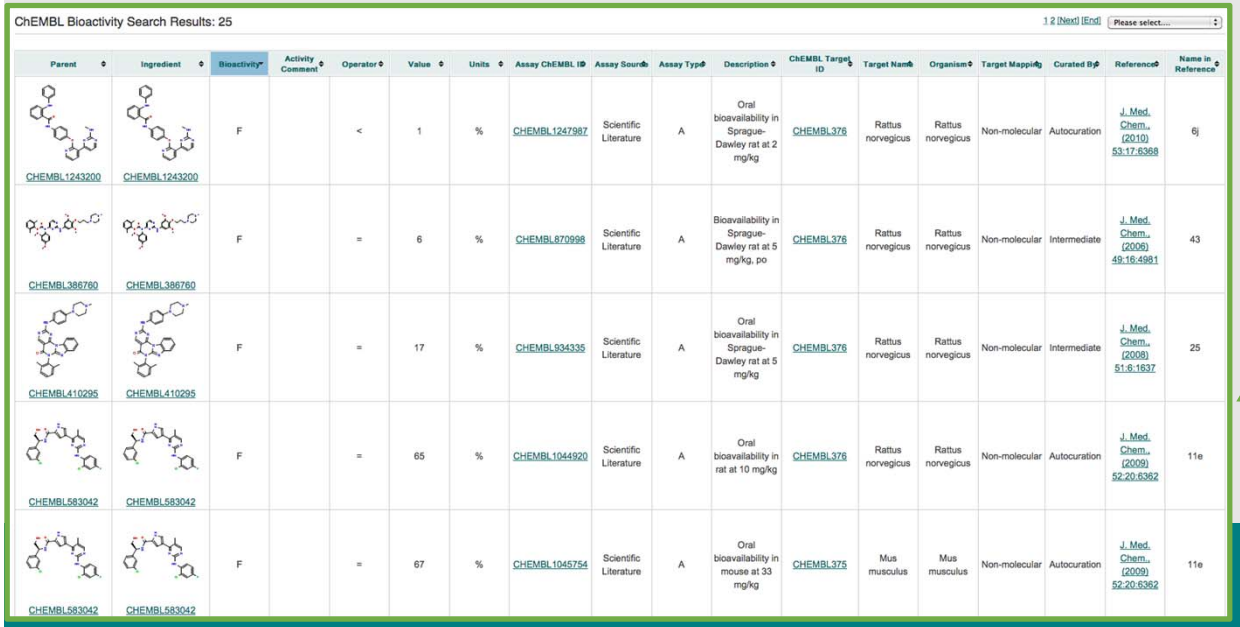
EMBL-EBI

Looking at Selectivity/ADMET Properties

Summary of ChEMBL data for compounds with ZAP70 IC50 <10uM

Protein Target	No of Bioactivity Values
Tyrosine-protein kinase ZAP-70	170
Tyrosine-protein kinase SYK	117
Tyrosine-protein kinase SRC	86
Epidermal growth factor receptor erbB1	54
Protein kinase C beta	52
Tyrosine-protein kinase LCK	52
Protein kinase C alpha	46
Lysyl-tRNA synthetase	43

Summary Example of Protein Target Data


ChEMBL Bioactivity Search Results: 25

1 2 [Next] [End] Please select...

Parent	Ingredient	Bioactivity	Activity Comment	Operator	Value	Units	Assay ChEMBL ID	Assay Source	Assay Type	Description	ChEMBL Target ID	Target Name	Organism	Target Mapping	Curated By	Reference	Name in Reference
CHEMBL1243200	CHEMBL1243200	F		<	1	%	CHEMBL1247987	Scientific Literature	A	Oral bioavailability in Sprague-Dawley rat at 2 mg/kg	CHEMBL376	Rattus norvegicus	Rattus norvegicus	Non-molecular	Autocuration	J. Med. Chem., (2010) 53:17:6368	6j
CHEMBL386760	CHEMBL386760	F		=	6	%	CHEMBL870998	Scientific Literature	A	Bioavailability in Sprague-Dawley rat at 5 mg/kg, po	CHEMBL376	Rattus norvegicus	Rattus norvegicus	Non-molecular	Intermediate	J. Med. Chem., (2006) 49:16:4981	43
CHEMBL410295	CHEMBL410295	F		=	17	%	CHEMBL934335	Scientific Literature	A	Oral bioavailability in Sprague-Dawley rat at 5 mg/kg	CHEMBL376	Rattus norvegicus	Rattus norvegicus	Non-molecular	Intermediate	J. Med. Chem., (2008) 51:6:1637	25
CHEMBL583042	CHEMBL583042	F		=	65	%	CHEMBL1044920	Scientific Literature	A	Oral bioavailability in rat at 10 mg/kg	CHEMBL376	Rattus norvegicus	Rattus norvegicus	Non-molecular	Autocuration	J. Med. Chem., (2009) 52:20:6362	11e
CHEMBL583042	CHEMBL583042	F		=	67	%	CHEMBL1045754	Scientific Literature	A	Oral bioavailability in mouse at 33 mg/kg	CHEMBL375	Mus musculus	Mus musculus	Non-molecular	Autocuration	J. Med. Chem., (2009) 52:20:6362	11e

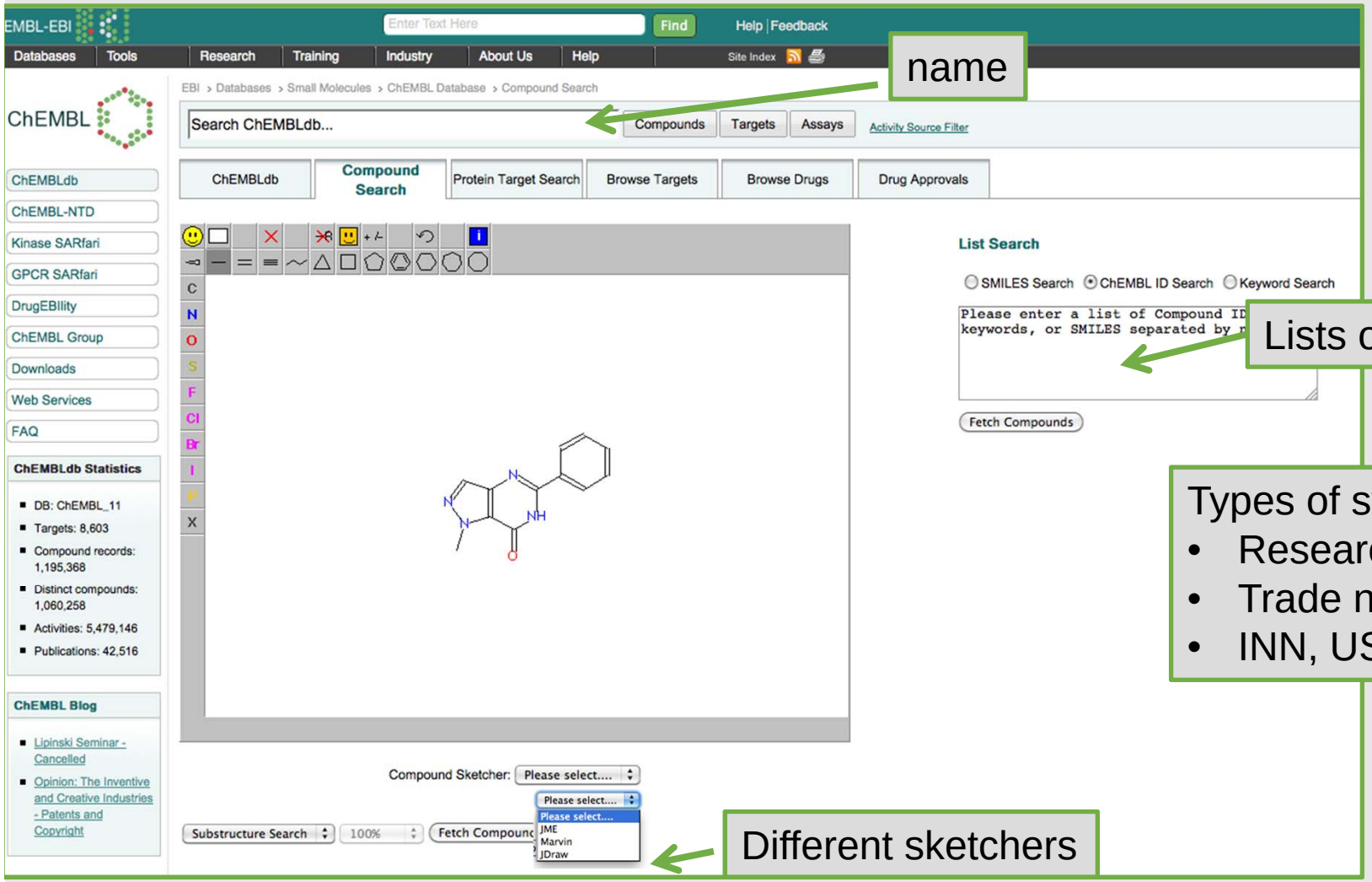
Example of Bioavailability data

EMBL-EBI

Use Case 2 – Searching by Structure

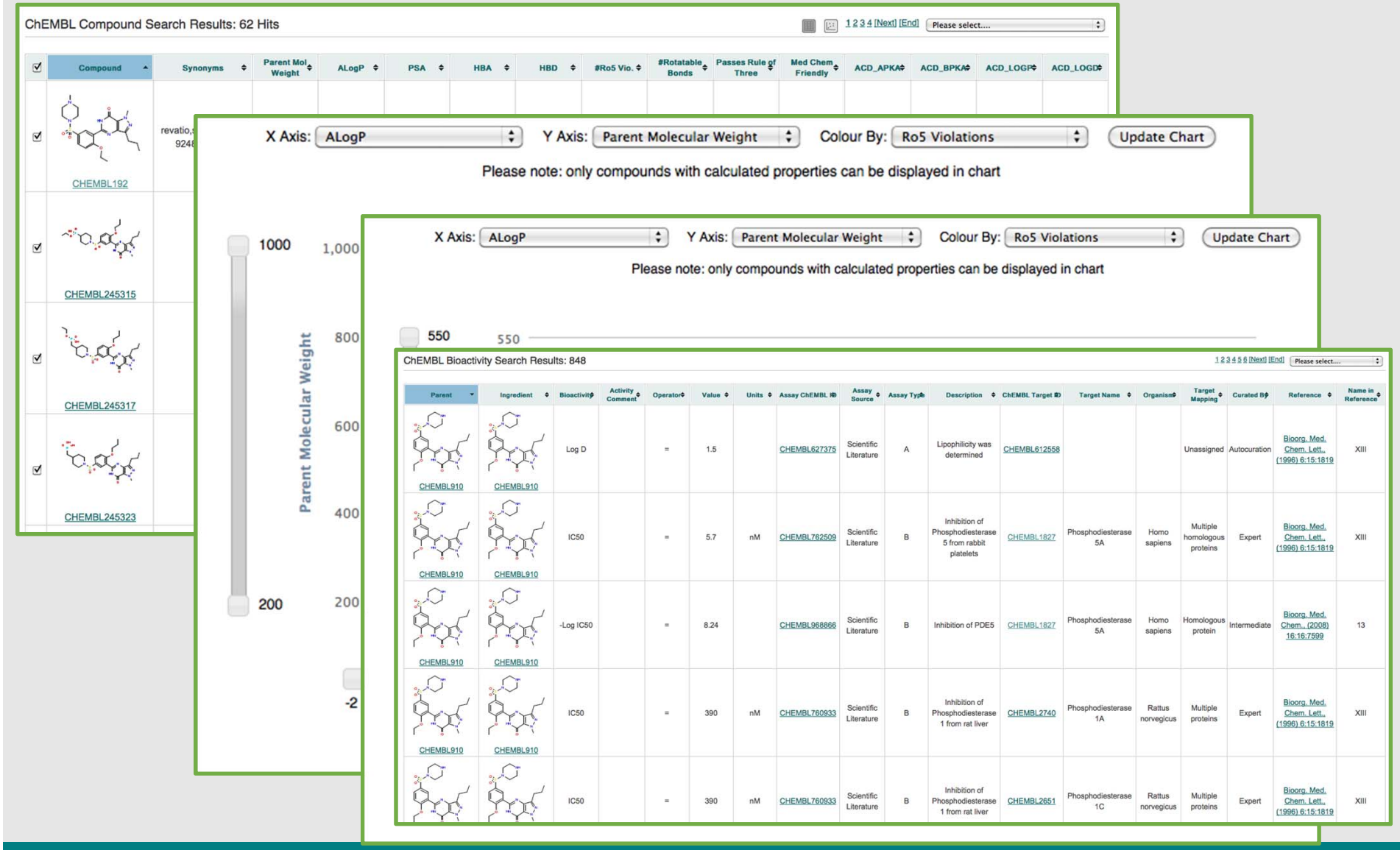
- What compounds contain a particular substructure?
- What is known about their bioactivities?
- Known drugs/clinical Trials

Use Case 2 - Structure Searching



Display/Download Bioactivity Data

Filtering Data on Lipinski Properties etc



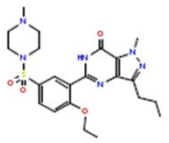
Display Bioactivities of subset

EMBL-EBI

Compound Report Card

Summary

Compound ID	CHEMBL192
Compound Name	sildenafil
Synonyms	UK-92480, Sildenafil, UK-92480-10, Sildenafil Citrate, Sildenafil citrate
Approved Drug	Yes
Trade Names	Viagra, Revatio



CHEMBL192

names

Approved Drug Features

Clinical Trials

Clinical Trials

Number of clinical trials registered at clinicaltrials.gov	324

Parent Properties

Mol. Weight	ALogP	Num Ro5 Violations	Num Rotatable Bonds	Passes Rule-Of-Three	Med Chem Friendly
474.6	2.25	0	7	No	Yes

Parent ACD Properties

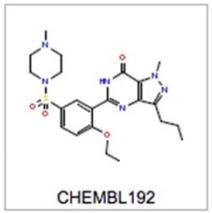
Acidic pKa	Basic pKa	LogP	LogD	Species
10.052	6.027	2.468	2.449	NEUTRAL

Properties

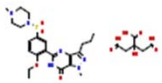
Representations

Molfile	Download MolFile
Molecular Formula	C22H30N6O4S
Canonical SMILES	CCCc1nn(C)c2C(=O)NC(=Nc12)c3cc(ccc3OCC)S(=O)(=O)N4CCN(C)CC4 Download SMILES
Standard InChI	InChI=1S/C22H30N6O4S/c1-5-7-17-19-20(27(4)25-17)22(29)24-21(... Download InChI
Standard InChI Key	BNRNXUUZRGQAQC-UHFFFAOYSA-N Download InChI Key

Molecule Forms



CHEMBL192



CHEMBL1737

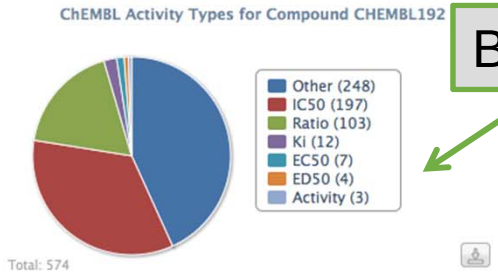
Structure

Database Links

ChEBI	ChEBI:9139
ChemSpider	ChemSpider:BNRNXUUZRGQAQC-UHFFFAOYSA-N
DrugBank	DB00203
PDBe	VIA - VIA (*PDBe Entries*)
PubChem	SID: 26748898 SID: 50085897
Wikipedia	Sildenafil

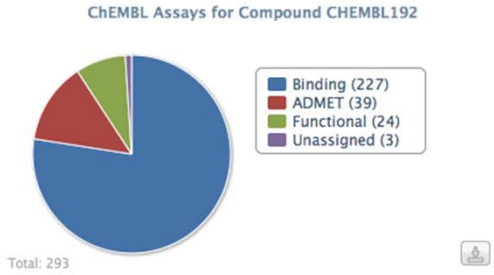
Links

Bioactivity Summary

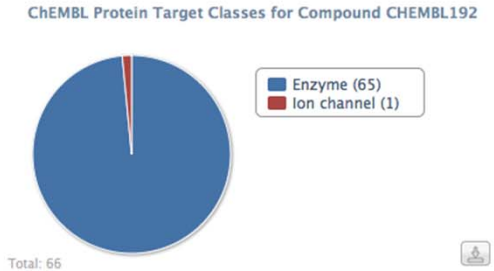


Bioactivities

Assay Summary

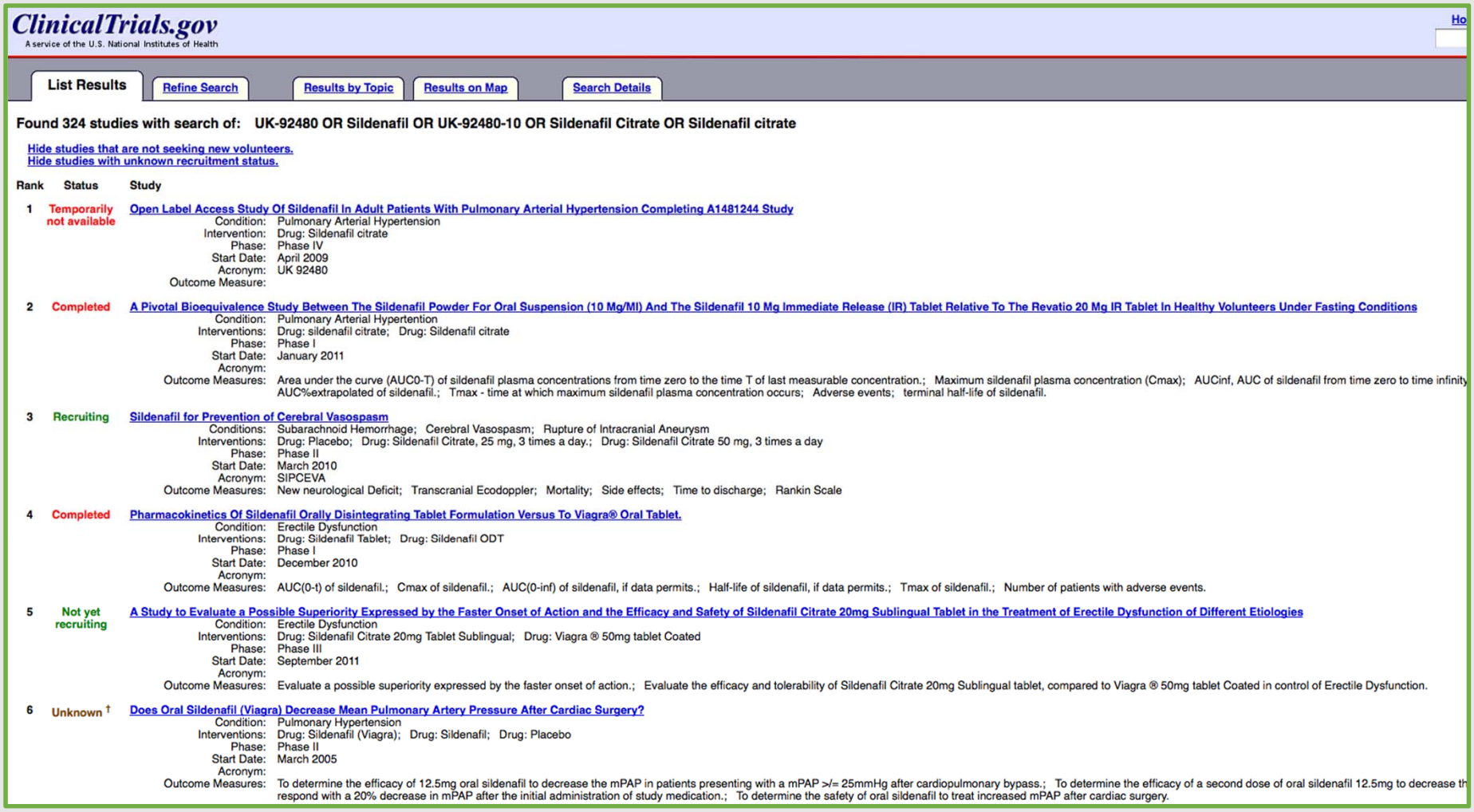


Protein Target Summary



EMBL-EBI

Links to Other Resources



ClinicalTrials.gov
A service of the U.S. National Institutes of Health

List Results | Refine Search | Results by Topic | Results on Map | Search Details

Found 324 studies with search of: UK-92480 OR Sildenafil OR UK-92480-10 OR Sildenafil Citrate OR Sildenafil citrate

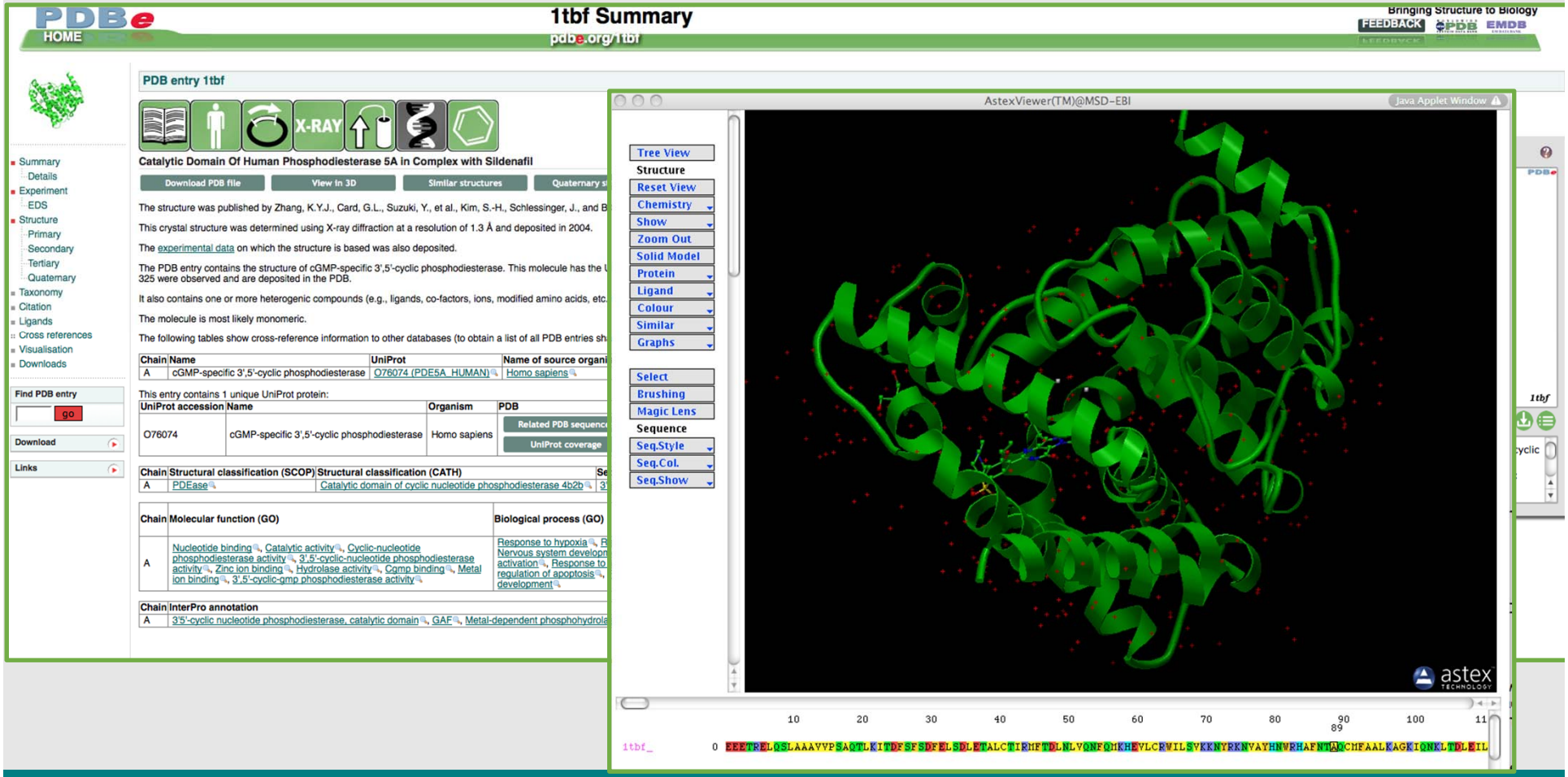
Hide studies that are not seeking new volunteers.
Hide studies with unknown recruitment status.

Rank	Status	Study
1	Temporarily not available	Open Label Access Study Of Sildenafil In Adult Patients With Pulmonary Arterial Hypertension Completing A1481244 Study Condition: Pulmonary Arterial Hypertension Intervention: Drug: Sildenafil citrate Phase: Phase IV Start Date: April 2009 Acronym: UK 92480 Outcome Measure:
2	Completed	A Pivotal Bioequivalence Study Between The Sildenafil Powder For Oral Suspension (10 Mg/Ml) And The Sildenafil 10 Mg Immediate Release (IR) Tablet Relative To The Revatio 20 Mg IR Tablet In Healthy Volunteers Under Fasting Conditions Condition: Pulmonary Arterial Hypertention Interventions: Drug: sildenafil citrate; Drug: Sildenafil citrate Phase: Phase I Start Date: January 2011 Acronym: Outcome Measures: Area under the curve (AUC0-T) of sildenafil plasma concentrations from time zero to the time T of last measurable concentration.; Maximum sildenafil plasma concentration (Cmax); AUCinf, AUC of sildenafil from time zero to time infinity AUC%extrapolated of sildenafil.; Tmax - time at which maximum sildenafil plasma concentration occurs; Adverse events; terminal half-life of sildenafil.
3	Recruiting	Sildenafil for Prevention of Cerebral Vasospasm Conditions: Subarachnoid Hemorrhage; Cerebral Vasospasm; Rupture of Intracranial Aneurysm Interventions: Drug: Placebo; Drug: Sildenafil Citrate, 25 mg, 3 times a day.; Drug: Sildenafil Citrate 50 mg, 3 times a day Phase: Phase II Start Date: March 2010 Acronym: SIPCEVA Outcome Measures: New neurological Deficit; Transcranial Ecodoppler; Mortality; Side effects; Time to discharge; Rankin Scale
4	Completed	Pharmacokinetics Of Sildenafil Orally Disintegrating Tablet Formulation Versus To Viagra® Oral Tablet. Condition: Erectile Dysfunction Interventions: Drug: Sildenafil Tablet; Drug: Sildenafil ODT Phase: Phase I Start Date: December 2010 Acronym: Outcome Measures: AUC(0-t) of sildenafil.; Cmax of sildenafil.; AUC(0-inf) of sildenafil, if data permits.; Half-life of sildenafil, if data permits.; Tmax of sildenafil.; Number of patients with adverse events.
5	Not yet recruiting	A Study to Evaluate a Possible Superiority Expressed by the Faster Onset of Action and the Efficacy and Safety of Sildenafil Citrate 20mg Sublingual Tablet in the Treatment of Erectile Dysfunction of Different Etiologies Condition: Erectile Dysfunction Interventions: Drug: Sildenafil Citrate 20mg Tablet Sublingual; Drug: Viagra ® 50mg tablet Coated Phase: Phase III Start Date: September 2011 Acronym: Outcome Measures: Evaluate a possible superiority expressed by the faster onset of action.; Evaluate the efficacy and tolerability of Sildenafil Citrate 20mg Sublingual tablet, compared to Viagra ® 50mg tablet Coated in control of Erectile Dysfunction.
6	Unknown †	Does Oral Sildenafil (Viagra) Decrease Mean Pulmonary Artery Pressure After Cardiac Surgery? Condition: Pulmonary Hypertension Interventions: Drug: Sildenafil (Viagra); Drug: Sildenafil; Drug: Placebo Phase: Phase II Start Date: March 2005 Acronym: Outcome Measures: To determine the efficacy of 12.5mg oral sildenafil to decrease the mPAP in patients presenting with a mPAP >/= 25mmHg after cardiopulmonary bypass.; To determine the efficacy of a second dose of oral sildenafil 12.5mg to decrease th respond with a 20% decrease in mPAP after the initial administration of study medication.; To determine the safety of oral sildenafil to treat increased mPAP after cardiac surgery.

EMBL-EBI

Links to Other Resources

PDBe - http://www.ebi.ac.uk/pdbe

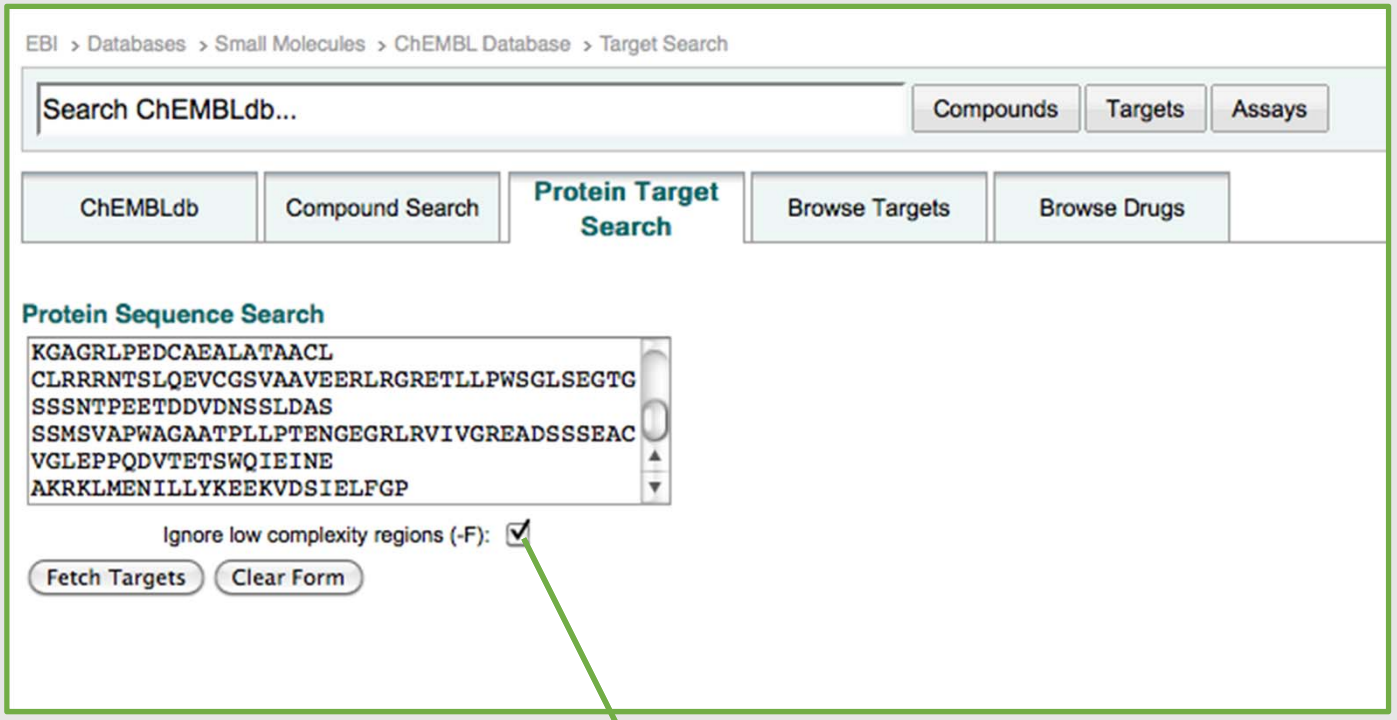


EMBL-EBI

Use Case 3 – Related Targets

- Are there any available data on compounds that bind to proteins related to IRAK2?
- For these compounds what bioactivity data is there on compounds with related sub-structures?
- Is there any crystal structure data on these proteins?

Use Case 3 – Similar Targets



Protein BLAST Results: 250 Hits Parameters: Submitted 1 sequences, searching against 5420 sequences 1 2 3 4 5 6 [Next] [End] Please select....

Query Sequence	ChEMBL ID	Preferred Name	UniProt Accession	Organism	# Compounds	# Endpoints	% Identity	BLAST Score
sp|O43187|IRAK2_HUMAN	CHEMBL3357	Interleukin-1 receptor-associated kinase 1	P51617	Homo sapiens	42	52	27.84	219
sp|O43187|IRAK2_HUMAN	CHEMBL3778	Interleukin-1 receptor-associated kinase 4	Q9NWZ3	Homo sapiens	211	225	27.81	145
sp|O43187|IRAK2_HUMAN	CHEMBL5081	Interleukin-1 receptor-associated kinase 3	Q9Y616	Homo sapiens	68	72	29.34	136
sp|O43187|IRAK2_HUMAN	CHEMBL5604	Dual specificity testis-specific protein kinase 1	Q15569	Homo sapiens	79	84	28.37	73.9

Data on IRAK1,IRAK3 and IRAK4 but not IRAK2

EMBL-EBI

IRAK1, IRAK3 and IRAK4 data

ChEMBL Bioactivity Search Results: 137

Parent	Ingredient	Bioactivity	Activity Comment	Operator	Value	Units	Assay ChEMBL ID	Assay Source	Assay Type	Description	ChEMBL Target ID	T
CHEMBL377383	CHEMBL377383	Ki		>	9900	nM	CHEMBL866115	Scientific Literature	B	Inhibition of Irak4	CHEMBL3778	Ir
CHEMBL1232515	CHEMBL1232515	IC50		=	8	nM	CHEMBL1272500	Scientific Literature	B	Inhibition of IRK4	CHEMBL3778	Ir
CHEMBL1270399	CHEMBL1270399	IC50		=	1800	nM	CHEMBL1272500	Scientific Literature	B	Inhibition of IRK4	CHEMBL3778	Ir
CHEMBL583187	CHEMBL583187	Ki		>	1000	nM	CHEMBL1060312	Scientific Literature	B	Inhibition of IRAK4 by HTRF assay	CHEMBL3778	Ir
CHEMBL583144	CHEMBL583144	Ki		>	1000	nM	CHEMBL1060312	Scientific Literature	B	Inhibition of IRAK4 by HTRF assay	CHEMBL3778	Ir
CHEMBL401633	CHEMBL401633	IC50		=	1	nM	CHEMBL921622	Scientific Literature	B	Inhibition of IRAK4	CHEMBL3778	Interleukin-1 receptor-associated kinase 4

Homo sapiens	Homologous protein	Autocuration	Bioorg. Med. Chem. Lett., (2008) 18:12:3656	rac-45	

EBI › Databases › Small Molecules › ChEMBL Database

Target Report Card

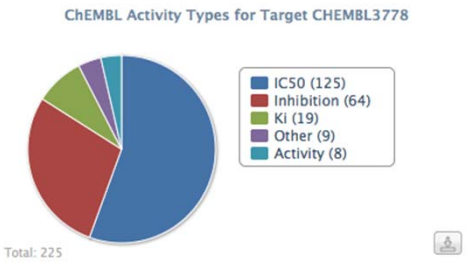
Target Details

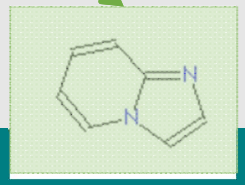
Target ID	CHEMBL3778
Target Type	PROTEIN
Preferred Name	Interleukin-1 receptor-associated kinase 4
Synonyms	Interleukin-1 receptor-associated kinase 4; Renal carcinoma antigen NY-REN-64; IRAK-4
Organism	Homo sapiens
Description	Interleukin-1 receptor-associated kinase 4
Protein Target Classification	enzyme kinase protein kinase ser_thr_tyr tkl irak

Database Links

UniProt	Q9NWZ3
PDBe	2NRU 2NRY 2O8Y 2OIB 2OIC 2OID 3MOP

Bioactivity Summary





Identify sub-structure of interest

What data other available on compounds with this sub-structure?

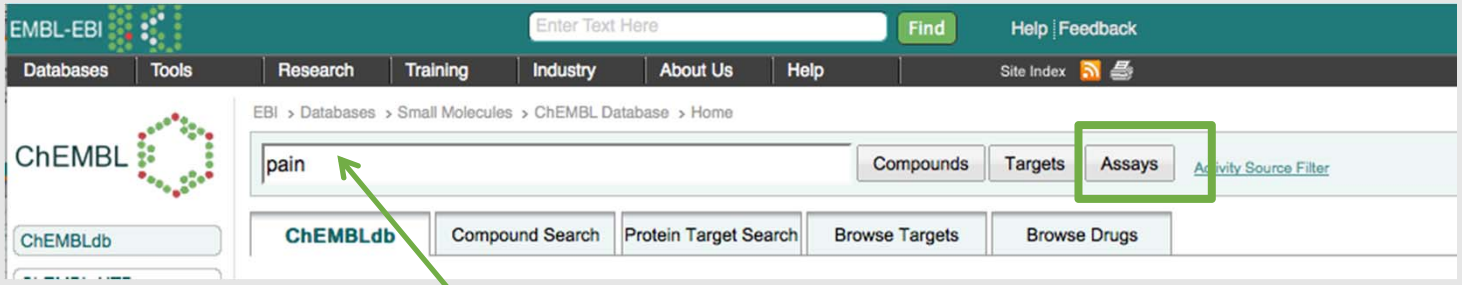
EMBL-EBI

Use Case 4
- Pathways and Disease Processes

- What compounds have been tested in disease models (pain)?
- What compounds interact with a pathway (opioid signalling)?

Use Case 4 – Processes and Pathways

- For performing more general search (e.g., for a disease process, animal model, cell type of interest)



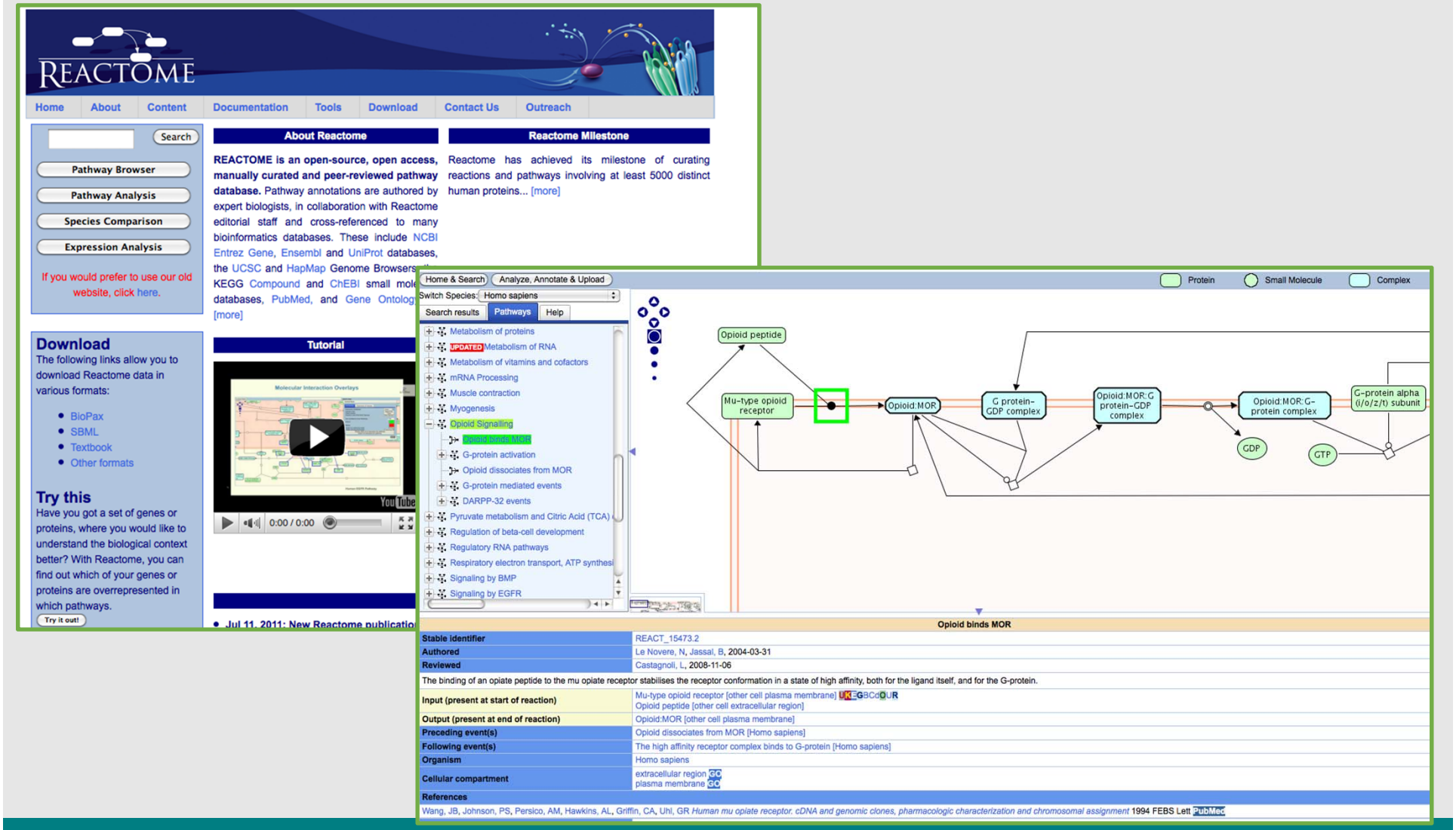
ChEMBL Assay Search Results: 1147 Hits

1 2 3 4 5 6 [Next] [End] Please select....

ChEMBL Assay ID	Assay Source	Assay Type	Assay Organism	Description	Activity Count	Reference
CHEMBL1003518	Scientific Literature	Binding	Carica papaya	Inhibition of papaya papain at 10 ug/ml	2	J. Nat. Prod., (1997) 60:2:184
CHEMBL1003702	Scientific Literature	Functional	Mus musculus	Antinociceptive activity against formalin-induced phase 2 pain in mouse paw assessed as pain behavior at 100 mg/kg, po pretreated 30 mins before fo...	2	J. Med. Chem., (2009) 52:9:3047
CHEMBL1003704	Scientific Literature	Functional	Mus musculus	Antinociceptive activity against acetone-induced chronic constriction injury in mouse paw assessed as duration of pain behavior at 100 mg/kg, po pre...	1	J. Med. Chem., (2009) 52:9:3047
CHEMBL1003952	Scientific Literature	Functional	Mus musculus	Analgesic activity in ICR mouse assessed as pain threshold variation at 0.013 mmol/kg, ip after 30 mins by tail flick test relative to control	3	Eur. J. Med. Chem., (2008) 43:5:1048
CHEMBL1003953	Scientific Literature	Functional	Mus musculus	Analgesic activity in ICR mouse assessed as pain threshold variation at 0.013 mmol/kg, ip after 60 mins by tail flick test relative to control	3	Eur. J. Med. Chem., (2008) 43:5:1048
CHEMBL1003954	Scientific Literature	Functional	Mus musculus	Analgesic activity in ICR mouse assessed as pain threshold variation at 0.013 mmol/kg, ip after 90 mins by tail flick test relative to control	3	Eur. J. Med. Chem., (2008) 43:5:1048
CHEMBL1003955	Scientific Literature	Functional	Mus musculus	Analgesic activity in ICR mouse assessed as pain threshold variation at 0.013 mmol/kg, ip after 120 mins by tail flick test relative	3	Eur. J. Med. Chem., (2008) 43:5:1048

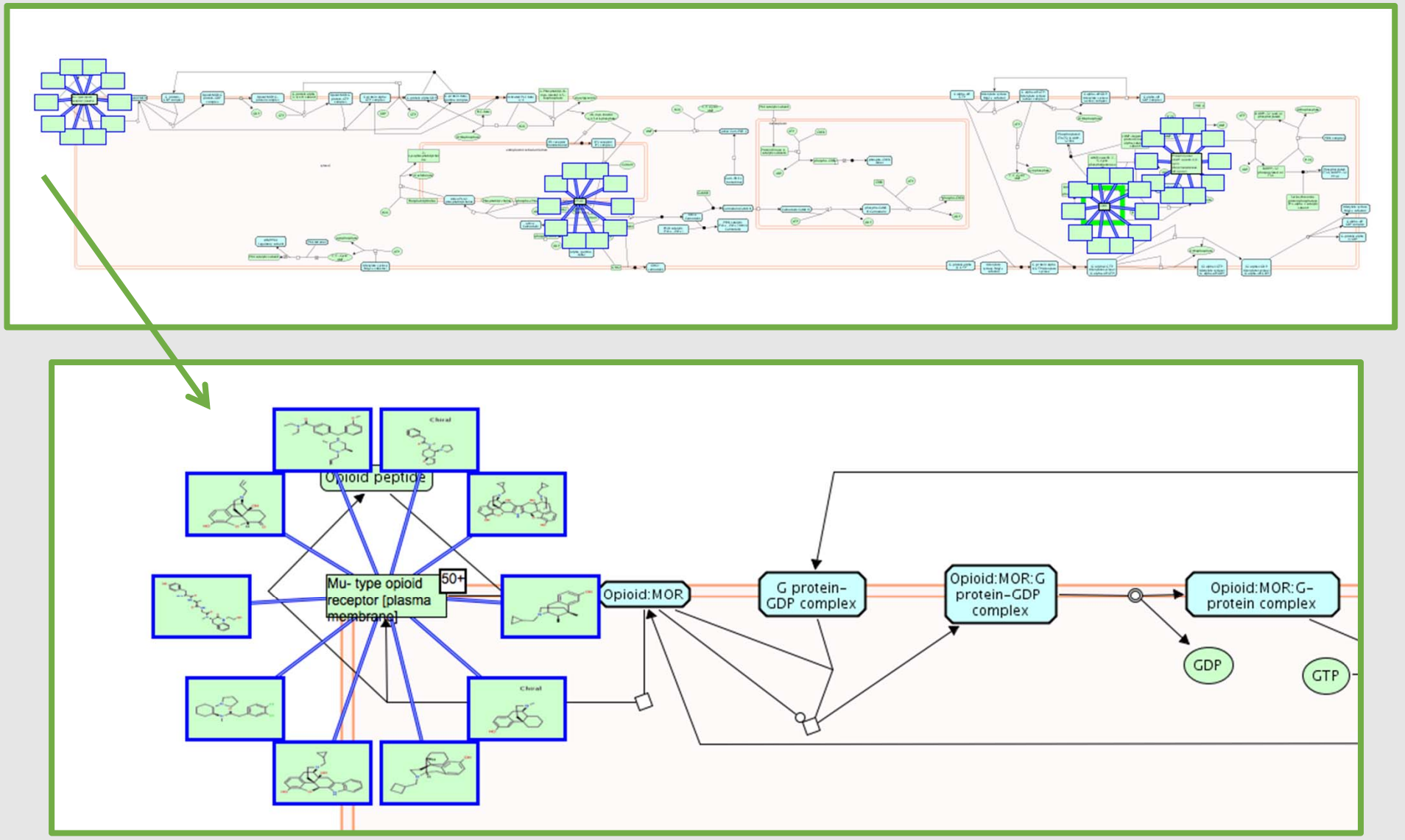
Pathways (Reactome)

http://www.reactome.org/ReactomeGWT/entrypoint.html



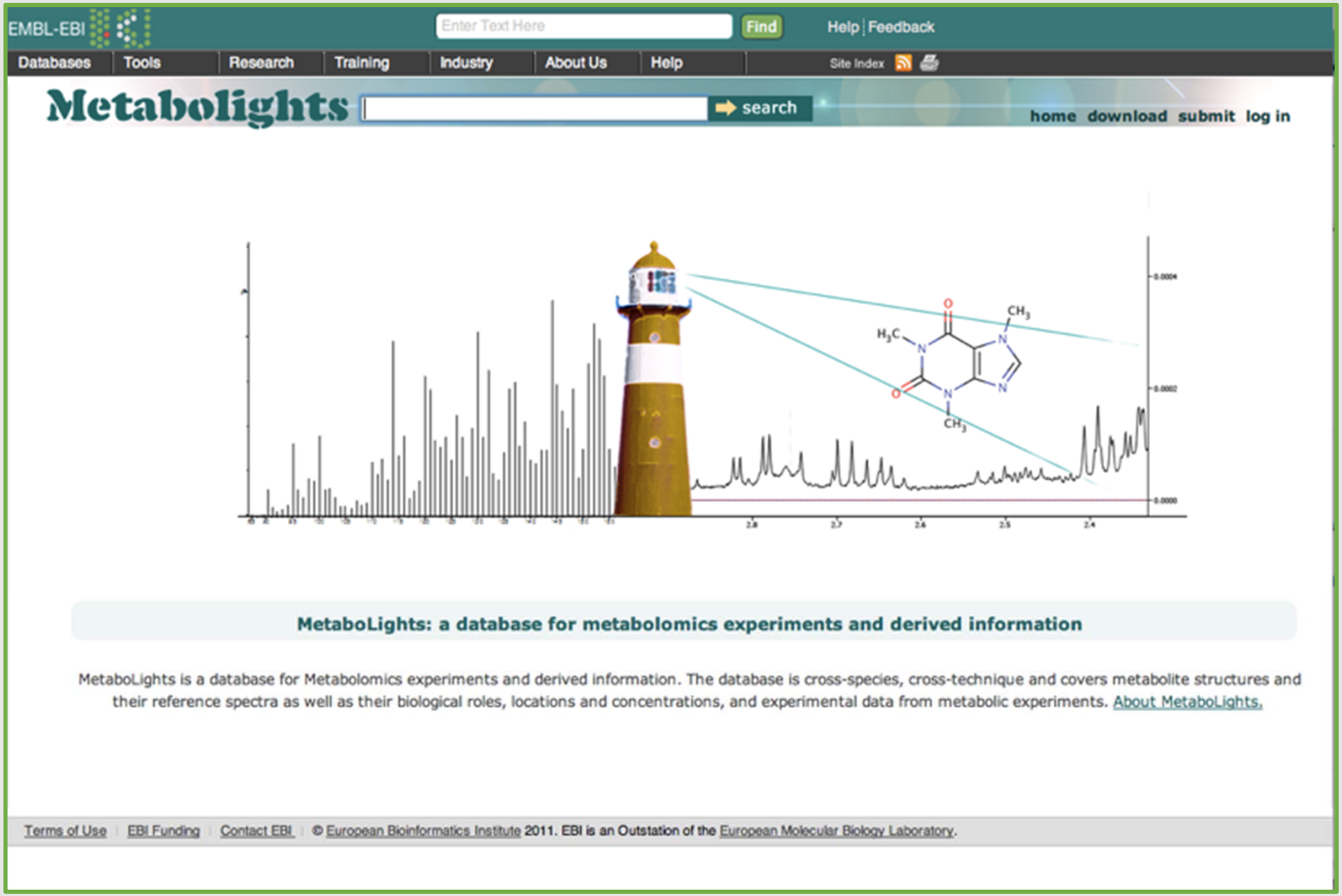
EMBL-EBI

Overlay ChEMBL Compounds that interact with Pathway



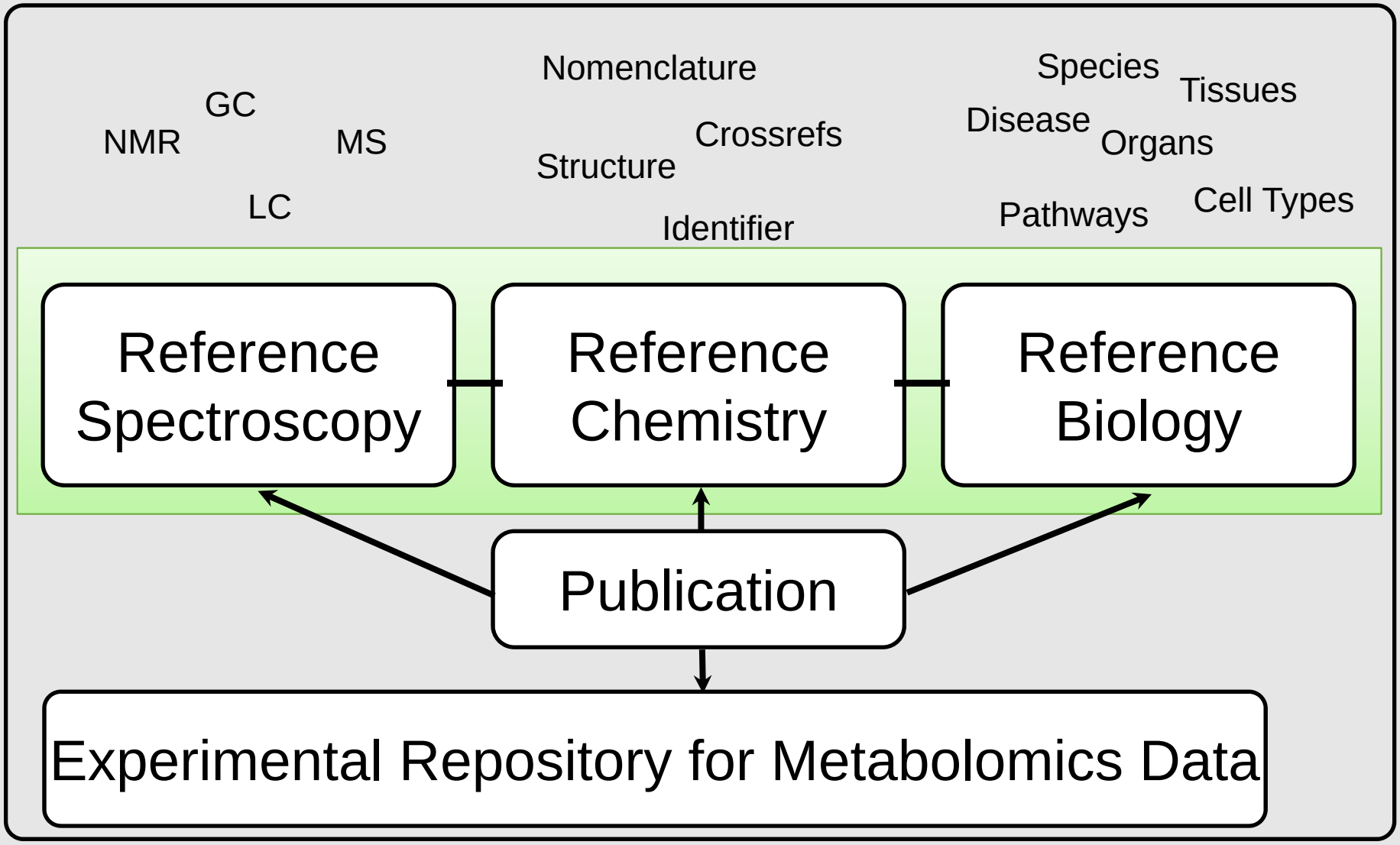
EMBL-EBI

Other Resources

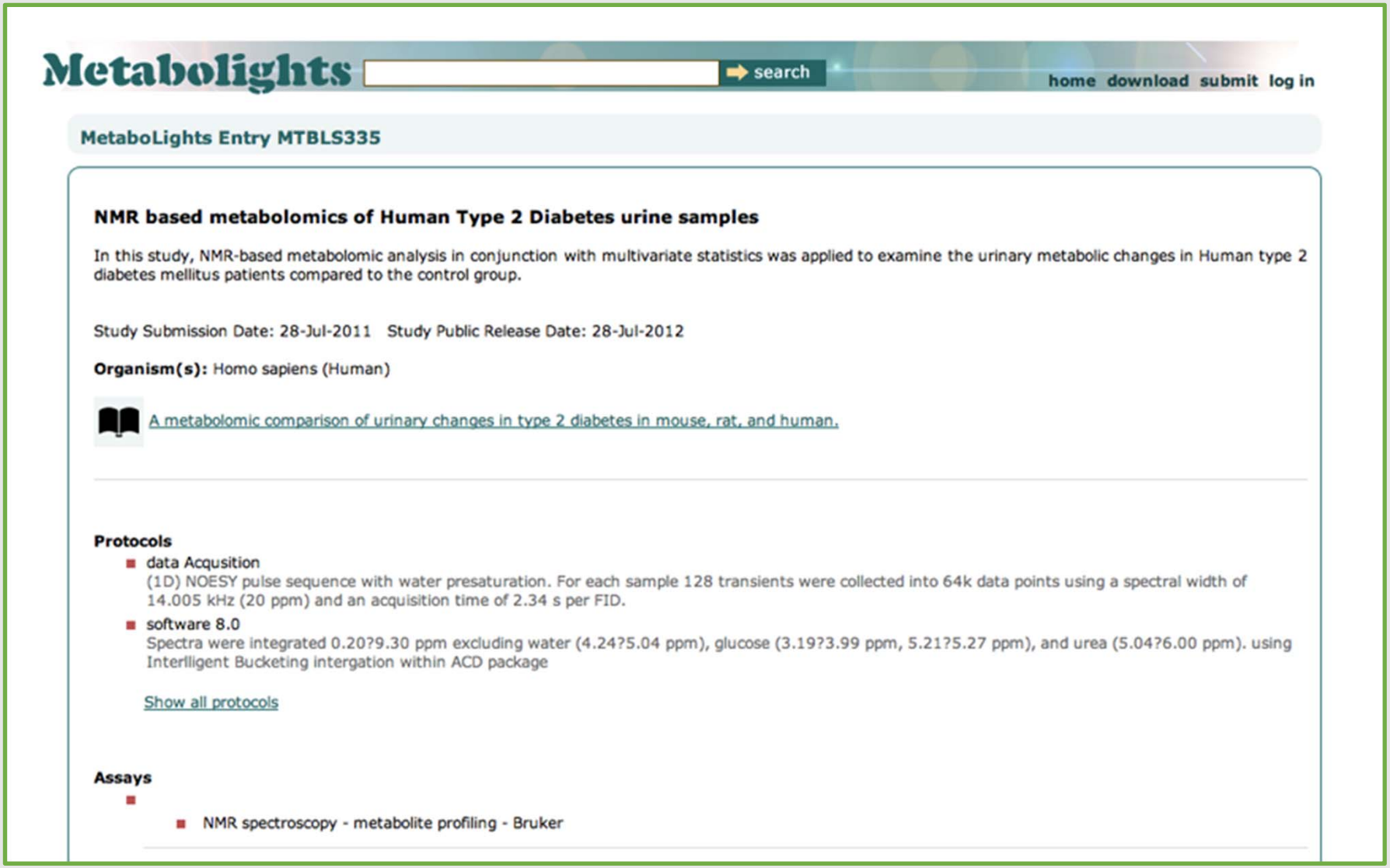


Christoph Steinbeck, Jules Griffin
BBSRC BBR grant BB/I000933/1

EMBL-EBI



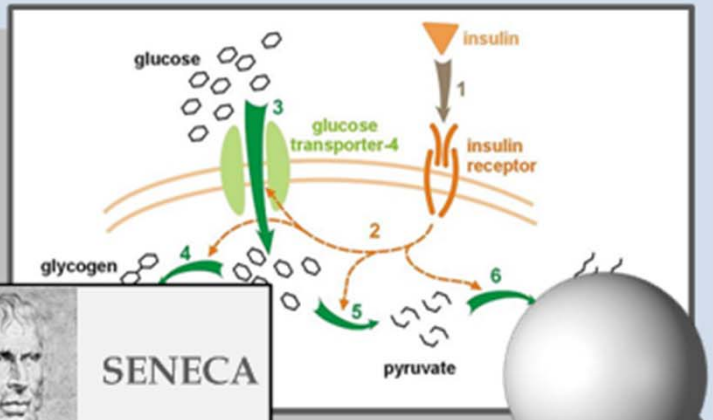
Curated reference information about metabolites, conditions where they are found, concentrations change when conditions change, etc



Will be released to group of testers soon

Steinbeck Group

Metabolism & Natural Products



Chemistry Databases

IntEnz

ChEBI

NMRShiftDB

RHEA

Algorithms & Toolkits

CDK

ORCHEM

Bioclipse

Standards

<CML/>

ChEBI ontology

OBO

Other ChEMBL Resources

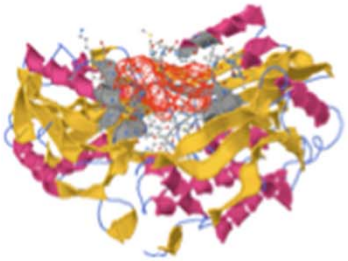
Welcome to Kinase SARfari

DrugEBIlity | Structure Based | Feature Based | Ligand Based

Structure-based DrugEBIlity

This is a structure-based druggablity search engine. Users can survey different types of druggability scores of a given protein structure.

- Druggability: evaluates the suitability of the binding site for small molecules under the Role of Five. [Druggable=1/Undruggable=0]
- Tractability: as above but under more relaxed conditions. [Tractable=1/Intractable=0]
- Ensemble: is the average of druggability score calculated under different models. [ranging from Undruggable:-1.0 to Druggable:+1.0]



Calculation:

- You can calculate structure-based druggability scores by uploading your structure.

Database Seach (Domain-Level):

- If you have a px number (SCOP domain) or a pdb code, you can access its associated data by clicking on the "3D Domain" link.
- If you have an accession, you can access more information by clicking on the "Protein" link.
- If you have a sequence, you may run it against all domain or gene sequences by using the "Blast" link.

This database currently contains:

- 189,702 unique domains
- 20,995 accessions
- 60,244 PDB codes
- 757,964 PDB sites
- 10,222 unique ligand codes

-Updated Aug 2010-

Acknowledgements

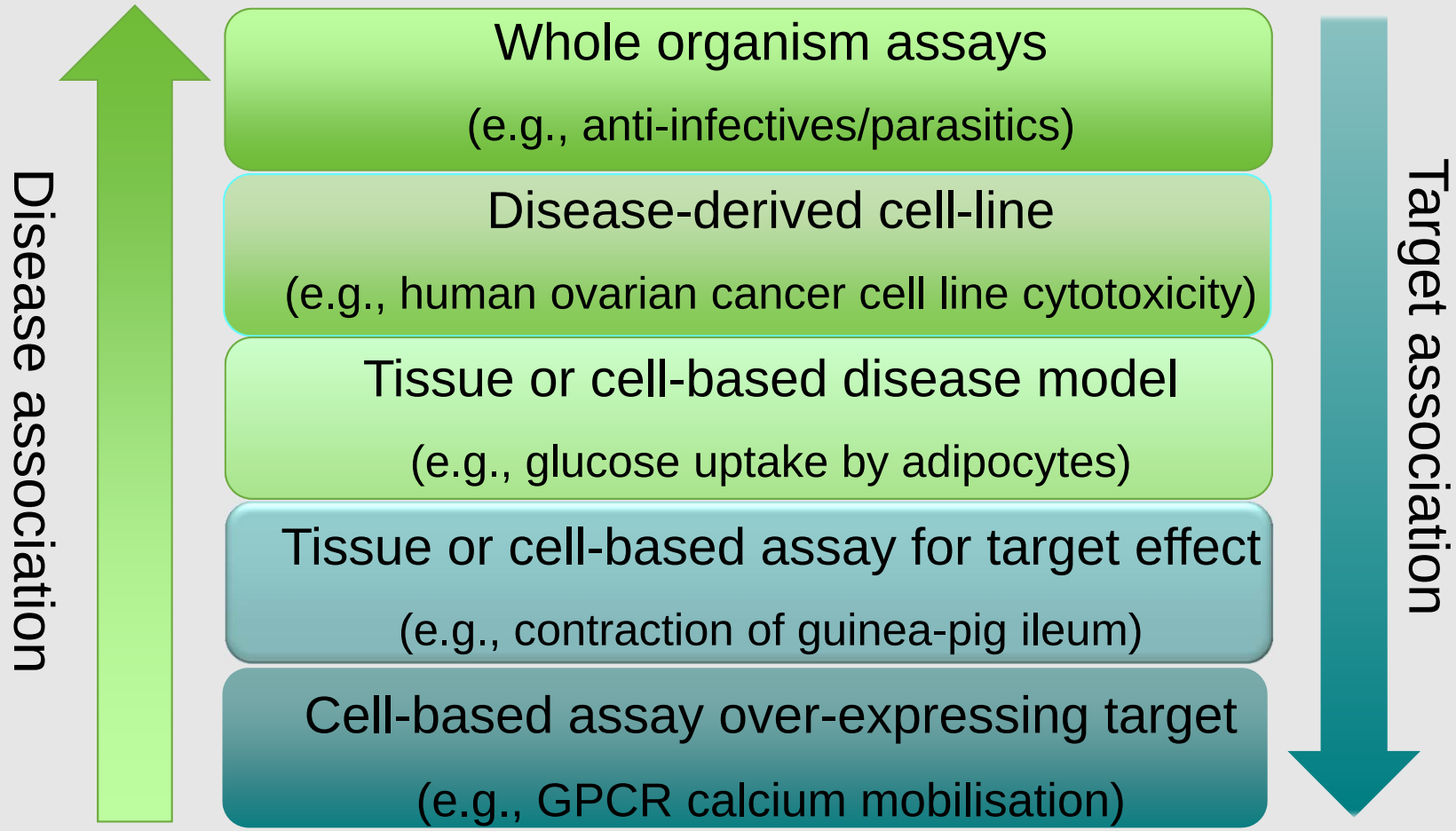
- John Overington and the ChEMBL Group
- Chris Steinbeck and the ChEBI Group
- Numerous Collaborators

Extra Slides

ChEMBL Assays

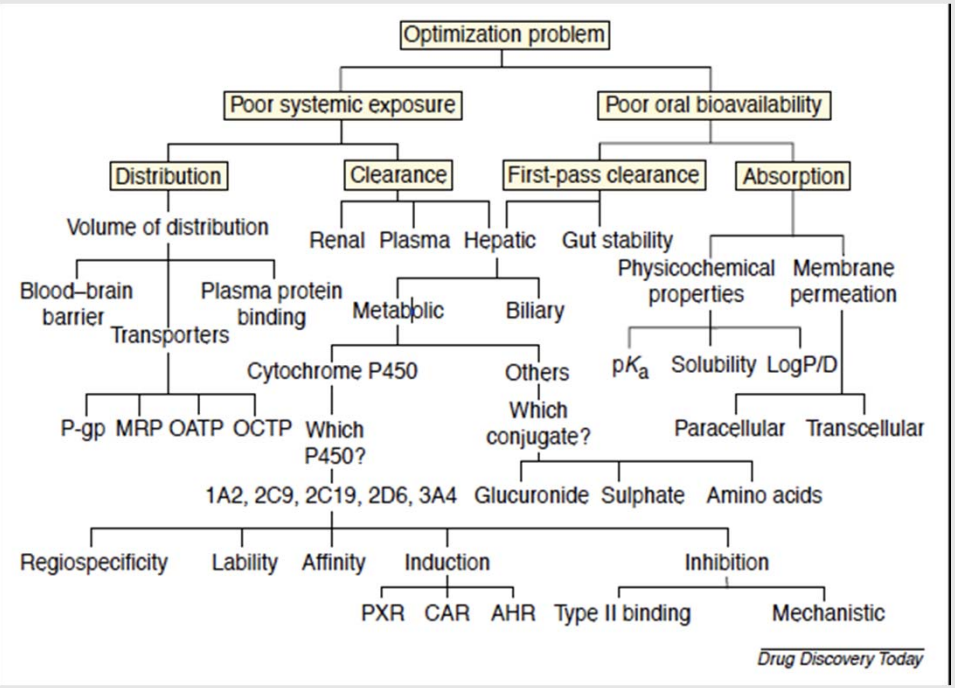
- Binding Assays
- Assays which directly measure the binding of a compound to a particular target
  - E.g., competition binding assays with a radioligand
- Various endpoints measured, but most commonly reported are:
  - IC50 (half maximal inhibitory concentration)
  - Ki (binding affinity)
  - MIC (minimum inhibitory concentration)
  - % Inhibition (of activity)

Functional Assays



ADMET Assays

- Assays measuring:
  Absorption, Distribution, Metabolism, Excretion,Toxicity properties of compounds



- Examples include:
  - Half-life of compound in rats
  - Tissue distribution of compound
  - Levels of metabolites

ChEMBL targets:

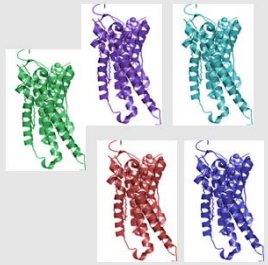
Protein

e.g., PDE5

Protein complex

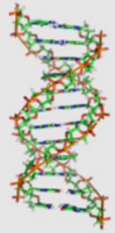
e.g., Nicotinic acetylcholine receptor

Protein family



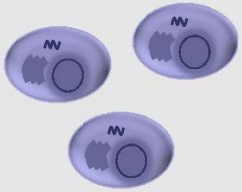
e.g., Muscarinic receptors

Nucleic Acid



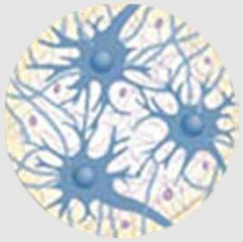
e.g., DNA

Cell Line



e.g., HEK293 cells

Tissue



e.g., Nervous

Sub-cellular Fraction

e.g., Mitochondria

Organism

e.g., Drosophila

And more… (e.g., splice variants, allosteric binding sites, cleaved peptides)

Protein targets

- Each protein target linked to a sequence in UniProt
- Information from UniProt used in ChEMBL to allow searching:
  - Protein name/description
  - Synonyms and gene names
  - Organism (and NCBI Tax ID)
- Proteins in ChEMBL also classified according to family (e.g., Receptor, Kinase, Protease, Transporter etc).
  - Used for searching by target tree

How are targets assigned to assays

- Where the target is a protein complex or protein family, multiple targets are assigned to the assay and one of two flags set:
  - Complex (used if target is protein complex)
  - Multi (used if target is poorly defined)

EXAMPLES

- ‘Compound was evaluated for blocking activity on muscarinic receptors in guinea pig atria’
  - Unclear which muscarinic receptor subtypes relevant (there are 5)
  - Assay mapped to all 5 subtypes (m1, m2, m3, m4 + m5)
  - Multi flag set to 1
- ‘Efficacy against human recombinant GABAA alpha1 in combination with beta3gamma2 expressed in L(tk-) cells by 36Cl ion flux modulation’
  - GABA-A target known to be a protein complex of alpha1/beta3/gamma2 subunits
  - Assay mapped to all three proteins
  - Complex flag set to 1

Confidence scores

- A confidence score is also set to reflect the level of confidence in the target assignment

CONFIDENCE SCORE	DESCRIPTION	TARGET_MAPPING
0	Default value - Target unknown or has yet to be assigned	Unassigned
1	Target assigned is non-molecular	Non-molecular
2	Target assigned is subcellular fraction	Subcellular fraction
3	Target assigned is molecular non-protein target	Molecular (non-protein)
4	Multiple homologous protein targets may be assigned	Multiple homologous proteins
5	Multiple direct protein targets may be assigned	Multiple proteins
6	Homologous protein complex subunits assigned	Homologous protein complex
7	Direct protein complex subunits assigned	Protein complex
8	Homologous single protein target assigned	Homologous protein
9	Direct single protein target assigned	Protein