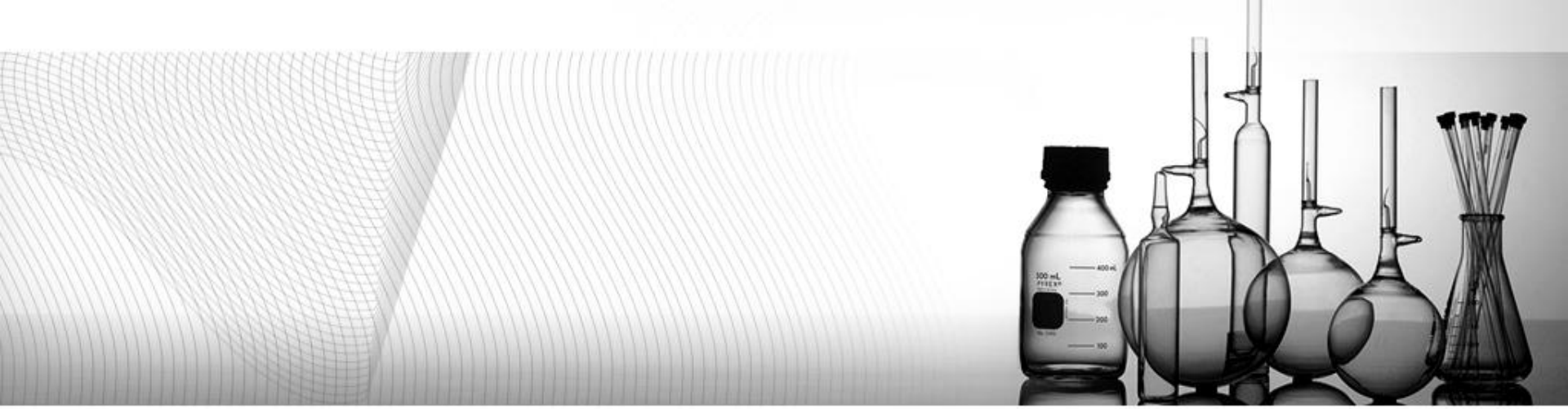


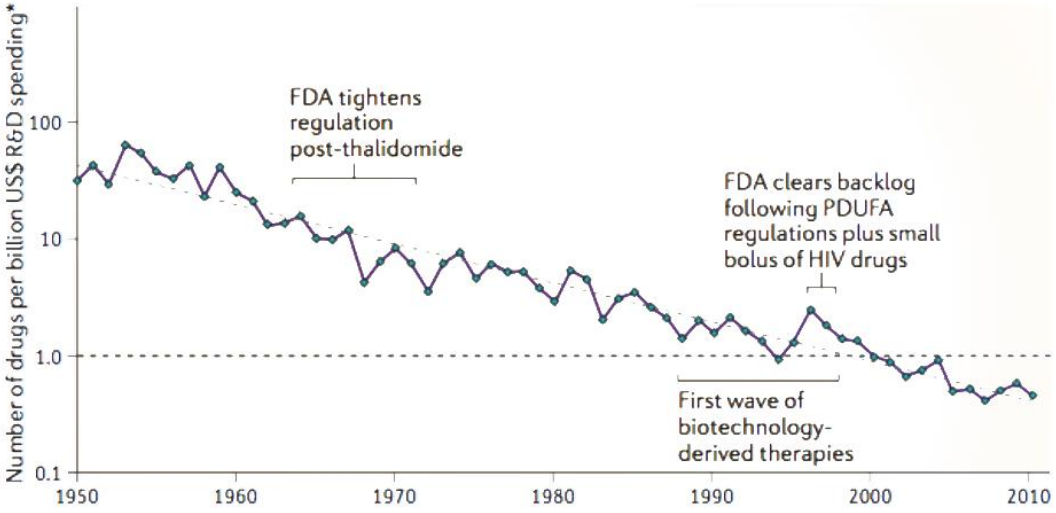
Open PHACTS
Open Pharmacological Space

A novel platform for integrated data-driven drug discovery

Gerhard F. Ecker
Dept Medicinal Chemistry, Univ Vienna
Gerhard.f.ecker@univie.ac.at; www.openphacts.org

Why do we need Open PHACTS?

Pharmaceutical companies currently expend significant effort integrating the vast amount of data publicly available into internal architectures.



Overall trend in R&D efficiency, inflation-adjusted (J. W. Scannel, A. Blanckley, H. Boldon and B. Warrington, *Nat. Rev. Drug Discov.*, 2012, **11**, 191-200, (doi:10.1038/nrd3681))

Currently, pharmaceutical companies assemble their own in-house databases of pharmacological and physicochemical data.

Drug discovery process is hindered by repetition of:

- Data extraction
- Transformation
- Loading stage

The Innovative Medicines Initiative

- EC funded public-private partnership for pharmaceutical research
- Focus on key problems
  - Efficacy, Safety, Education & Training, **Knowledge Management**

The Open PHACTS Project

- Create a *semantic integration hub ("Open Pharmacological Space")…*
- Delivering services to support on-going drug discovery programs in pharma and public domain
- *Not just another project*; Leading academics in semantics, pharmacology and informatics, driven by solid industry business requirements
- 16 academic partners, 8 pharmaceutical companies, 4 biotechs
- Work split into clusters:
  - Tehnical Build (*focus here)*
  - Scientific Drive
  - Community & Sustainability

The Project

Open PHACTS Project Partners

www.openphacts.org

Pfizer Limited – Coordinator
Universität Wien – Managing entity
Technical University of Denmark
University of Hamburg, Center for Bioinformatics
BioSolveIT GmBH
Consorci Mar Parc de Salut de Barcelona
Leiden University Medical Centre
Royal Society of Chemistry
Vrije Universiteit Amsterdam

Spanish National Cancer Research Centre
University of Manchester
Maastricht University
Aqnowledge
University of Santiago de Compostela
Rheinische Friedrich-Wilhelms-Universität Bonn
AstraZeneca
GlaxoSmithKline
Esteve

Novartis
Merck Serono
H. Lundbeck A/S
Eli Lilly
Netherlands Bioinformatics Centre
Swiss Institute of Bioinformatics
ConnectedDiscovery
EMBL-European Bioinformatics Institute
Janssen
OpenLink

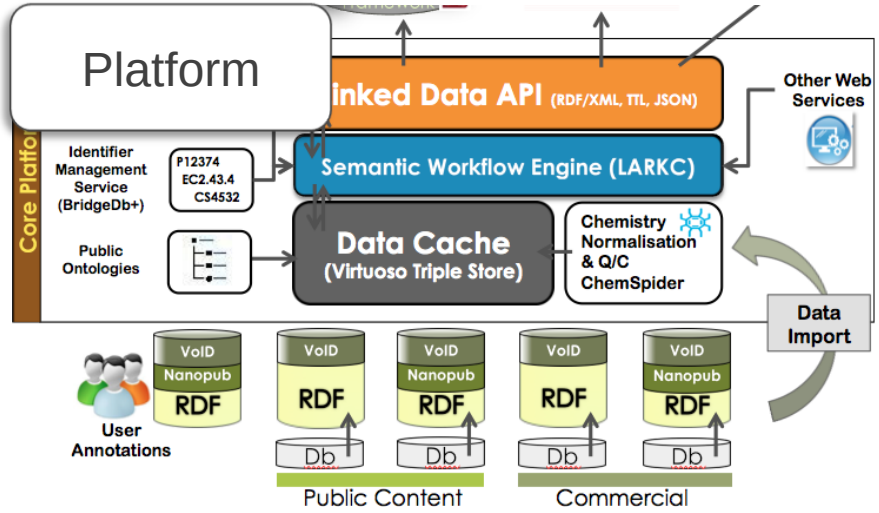
A user-friendly, full featured interface that allows scientists to explore and interrogate integrated biological and chemical data

What will users see?

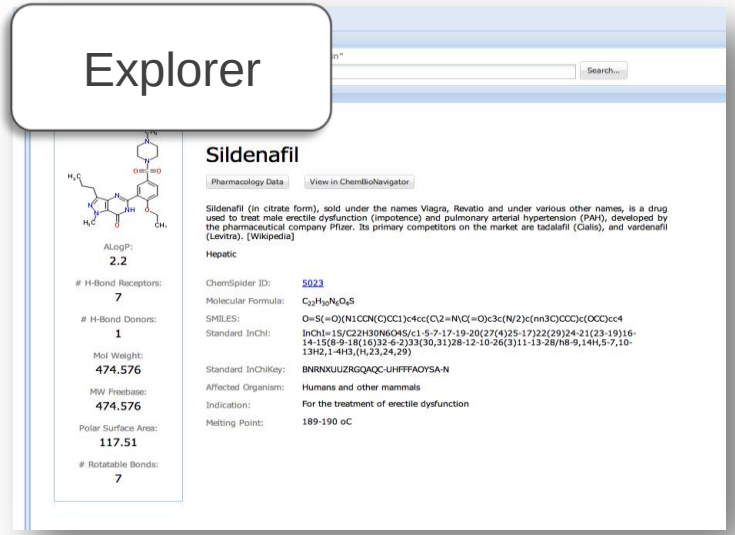
Goals

Open PHACTS
Open Pharmacological Space

Platform



Explorer



Apps

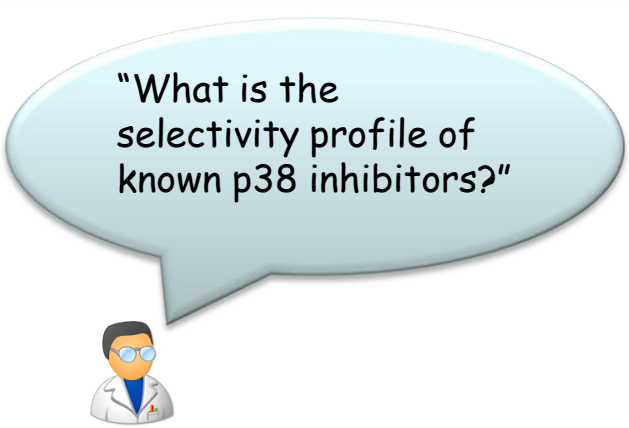
API

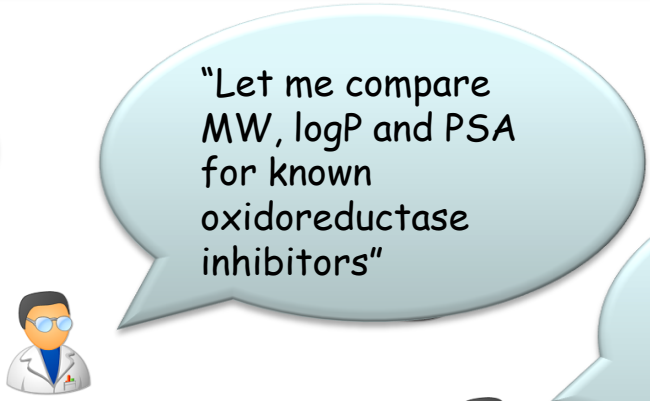
```
?ops_item skos:e
?ops_item skos:e
?cw_uri skos:pre
        void:inD
?equiv_target dc
        ops:targ
        ops:targ
        void:inD
ops:targetOfAssa
?equiv_assay che
        chembl:b
                                    ?std_type ;
```

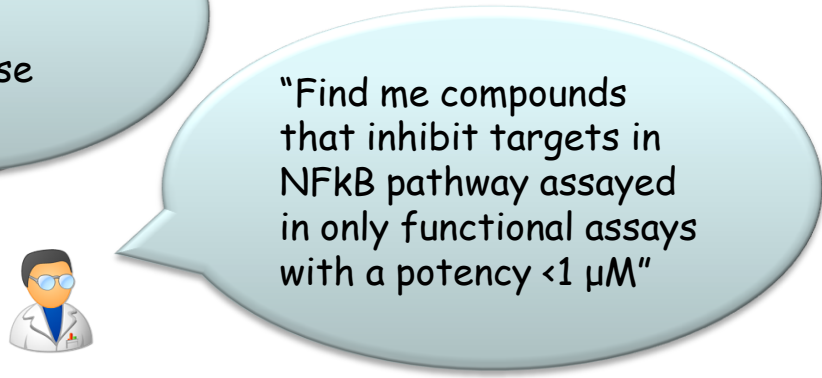
Standards

What do we need?

Open PHACTS
Open Pharmacological Space







ChEMBL | DrugBank | Gene Ontology | Wikipathways | GeneGo

ChEBI | Uniprot | UMLS | GVKBio

ConceptWiki | ChemSpider | TrialTrove | TR Integrity

The Open PHACTS infrastructure can support many different domains & questions

Research Questions

Open PHACTS
Open Pharmacological Space

Number	sum	Nr of 1	Question
15	**12**	**9**	**All oxido,reductase inhibitors active <100nM in both human and mouse**
18	14	8	Given compound X, what is its predicted secondary pharmacology? What are the on and off,target safety concerns for a compound? What is the evidence and how reliable is that evidence (journal impact factor, KOL) for findings associated with a compound?
24	13	8	Given a target find me all actives against that target. Find/predict polypharmacology of actives. Determine ADMET profile of actives.
32	13	8	For a given interaction profile, give me compounds similar to it.
37	13	8	The current Factor Xa lead series is characterised by substructure X. Retrieve all bioactivity data in serine protease assays for molecules that contain substructure X.
38	13	8	Retrieve all experimental and clinical data for a given list of compounds defined by their chemical structure (with options to match stereochemistry or not).
41	13	8	A project is considering Protein Kinase C Alpha (PRKCA) as a target. What are all the compounds known to modulate the target directly? What are the compounds that may modulate the target directly? i.e. return all cmpds active in assays where the resolution is at least at the level of the target family (i.e. PKC) both from structured assay databases and the literature.
44	13	8	Give me all active compounds on a given target with the relevant assay data
46	13	8	Give me the compound(s) which hit most specifically the multiple targets in a given pathway (disease)
59	14	8	Identify all known protein-protein interaction inhibitors

Prioritized Datasets

Open PHACTS
Open Pharmacological Space

ChEMBL

HGNC
HUGO Gene Nomenclature Committee

BETA
WikiPathways
Pathways for the People

DrugBank
Open Data Drug & Drug Target Database

ChemSpider
The free chemical database

ChEBI

UniProt

hmp
human metabolome project

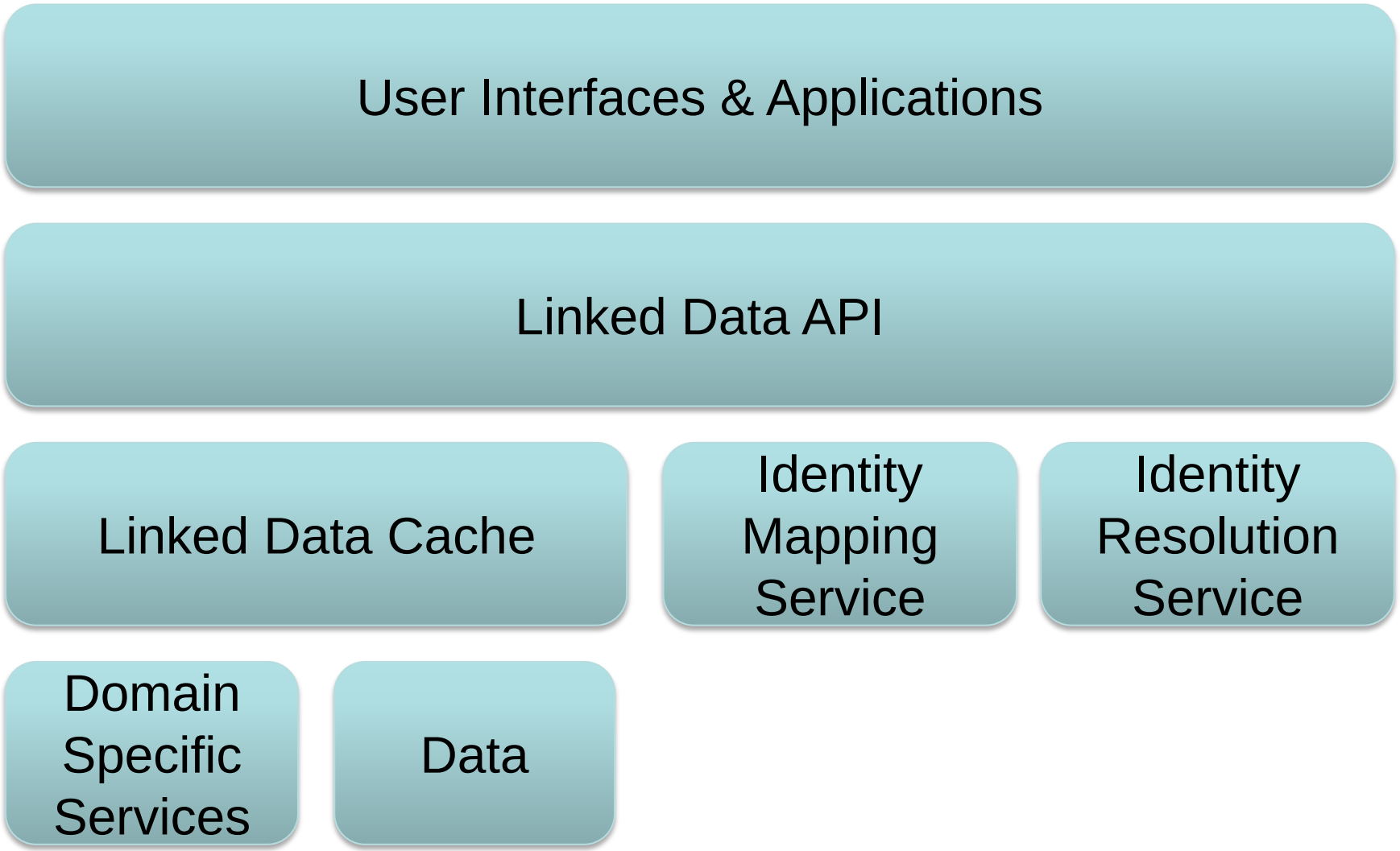
OBI

the Gene Ontology

www.openphacts.org

Architecture

Open PHACTS
Open Pharmacological Space



Quantitative Data Challenges

Open PHACTS
Open Pharmacological Space

STANDARD_TYPE	UNIT_COUNT
AC50	7
Activity	421
EC50	39
IC50	46
ID50	42
Ki	23
Log IC50	4
Log Ki	7
Potency	11
log IC50	0

>5000 types

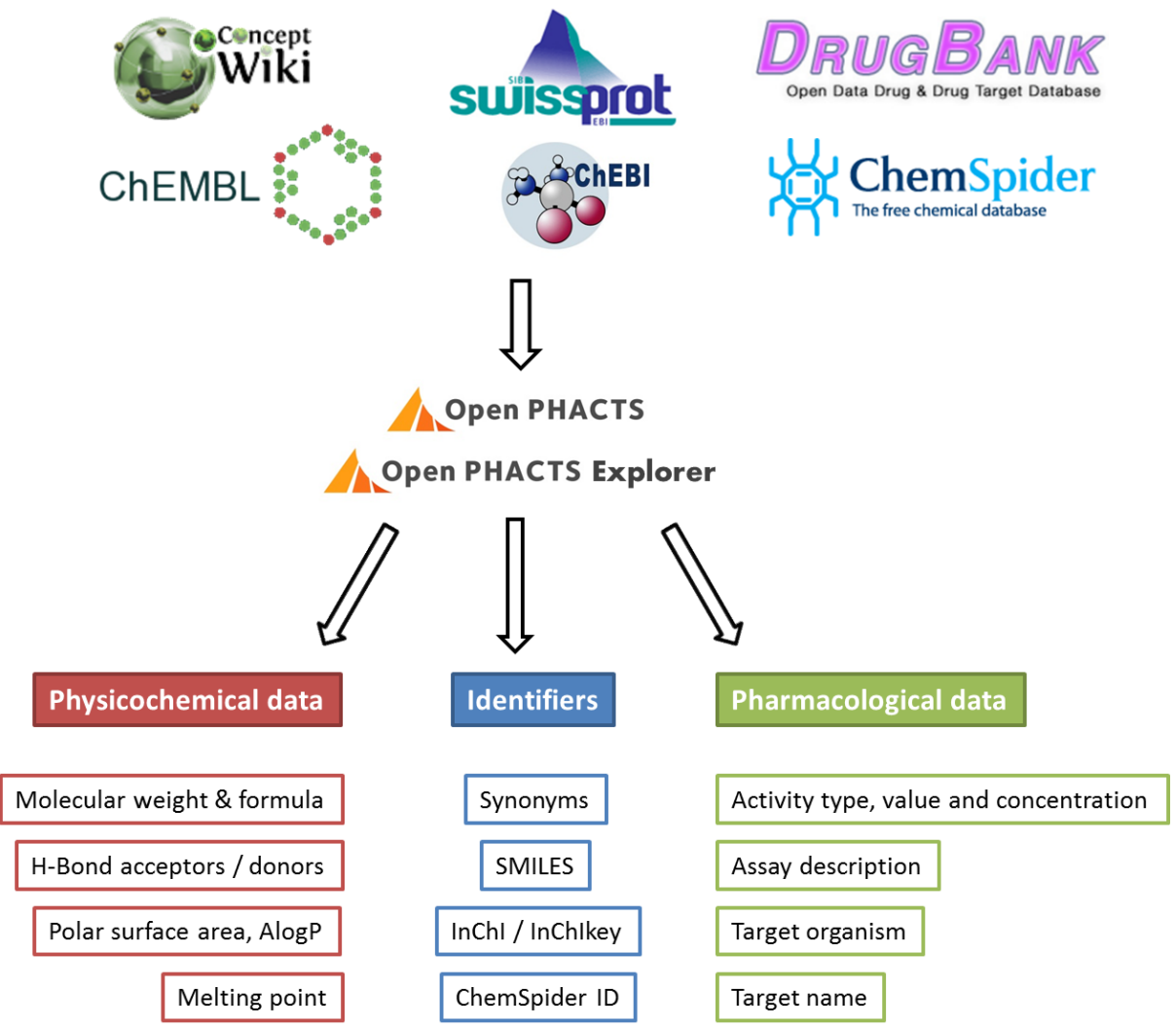
STANDARD_TYPE	STANDARD_UNITS	COUNT(*)
IC50	nM	829448
IC50	ug.mL-1	41000
IC50		38521
IC50	ug/ml	2038
IC50	ug ml-1	509
IC50	mg kg-1	295
IC50	molar ratio	178
IC50	ug	117
IC50	%	113
IC50	uM well-1	52
IC50	p.p.m.	51
IC50	ppm	36
IC50	uM-1	25
IC50	nM kg-1	25
IC50	milliequivalent	22
IC50	kJ m-2	20

Implemented using the Quantities, Dimension, Units, Types Ontology (http://www.qudt.org/)

~ 100 units

Open PHACTS
Open Pharmacological Space

What does Open PHACTS do?



Currently integrated databases

Database	Number of triples (million)
ACD Labs / ChemSpider	161.34
ChEBI	0.91
ChEMBL_v13	146.08
ConceptWiki	3.74
DrugBank	0.52
Enzyme	0.07
Gene Ontology	0.85
SwissProt	156.57
WikiPathways	0.14
TOTAL	**470.21**

Open PHACTS draws together multiple sources of publicly-available pharmacological and chemical data, allowing public access to the information via the Open PHACTS Explorer, an intuitive interface.

Chemistry within Open PHACTS

The challenges associated with handling chemistry data require the support of a publicly accessible platform to **integrate, standardise and host the data**.

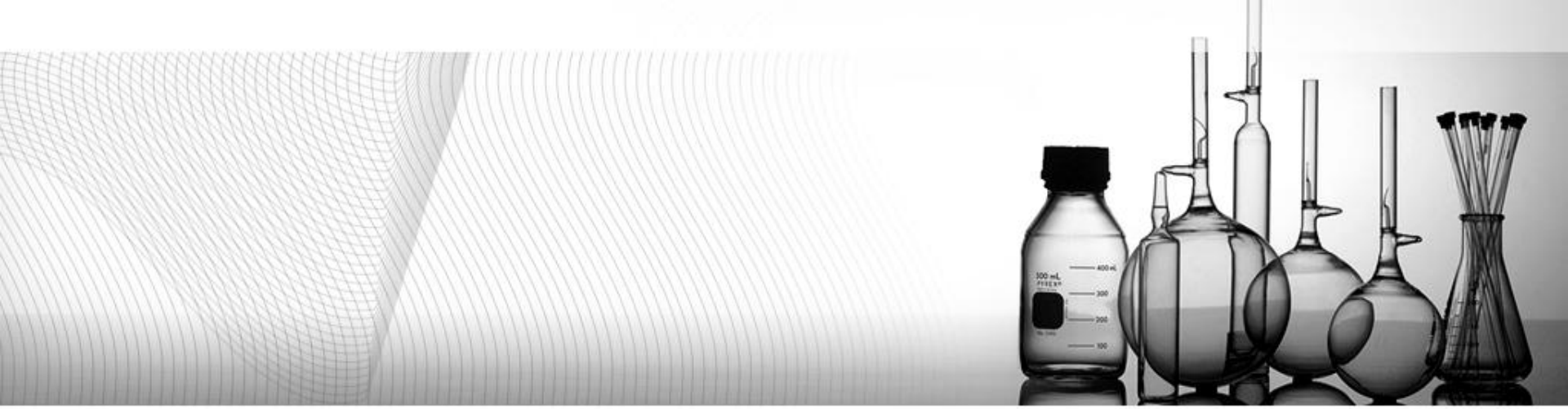
ChemSpider, an online database from the Royal Society of Chemistry hosts the chemical compound collection underpinning Open PHACTS and is responsible for standardising the chemical compounds and providing both regular updates and ongoing data curation.

To serve the Open PHACTS platform, a **structure validation and standardisation platform** (CVSP) has been developed to ensure chemical structures are normalised to rules derived from the FDA structure standardisation guidelines and modified based on input from the EFPIA members.

Example applications

Open PHACTS
Open Pharmacological Space

Advanced analytics	
ChemBioNavigator	Navigating at the interface of chemical and biological data with sorting and plotting options
TargetDossier	Interconnecting Open PHACTS with multiple target centric services. Exploring target similarity using diverse criteria
PharmaTrek	Interactive Polypharmacology space of experimental annotations
UTOPIA	Semantic enrichment of scientific PDFs
Predictions	
GARFIELD	Prediction of target pharmacology based on the Similar Ensemble Approach
eTOX connector	Automatic extraction of data for building predictive toxicology models in eTOX project



Open PHACTS
Open Pharmacological Space

Open PHACTS Explorer

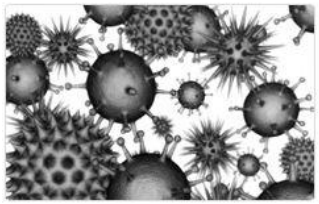
Exploring the Open Pharmacological Space

www.openphacts.org

Open PHACTS
Open Pharmacological Space

HOME | SITEMAP | MEMBER LOGIN

About Open PHACTS | Partners | News & Events | Documents | Open PHACTS Explorer



Menu Explorer

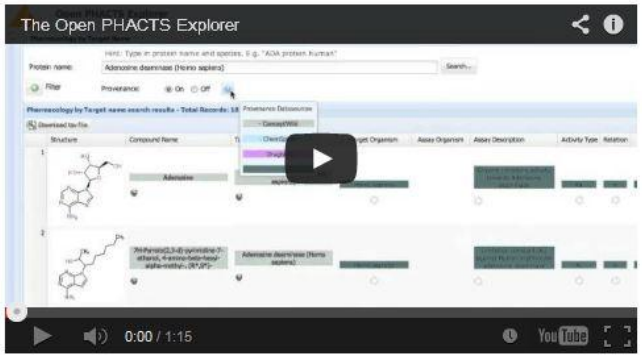
- Welcome
- The Project
- Explorer Release Notes
- Tutorials
- The Data
- Terms and Conditions
  - Terms of Use
  - Privacy Statement
  - Licensing
  - Take-down policy

Welcome to the Open PHACTS Explorer

We are pleased to present the first public release of the beta Open PHACTS Explorer.

The Open PHACTS Explorer allows multiple sources of publicly-available pharmacological and physicochemical data to be intuitively queried, and makes data provenance accessible at every step. The Open PHACTS Explorer was built to answer critical pharmacological questions as defined by academic and pharmaceutical industry scientists.

To get started watch our tutorial introduction, visit the tutorials page or click on the registration link :



This is the first public release of the Open PHACTS Explorer, and we look forward to and value your feedback and comments .

Explorer

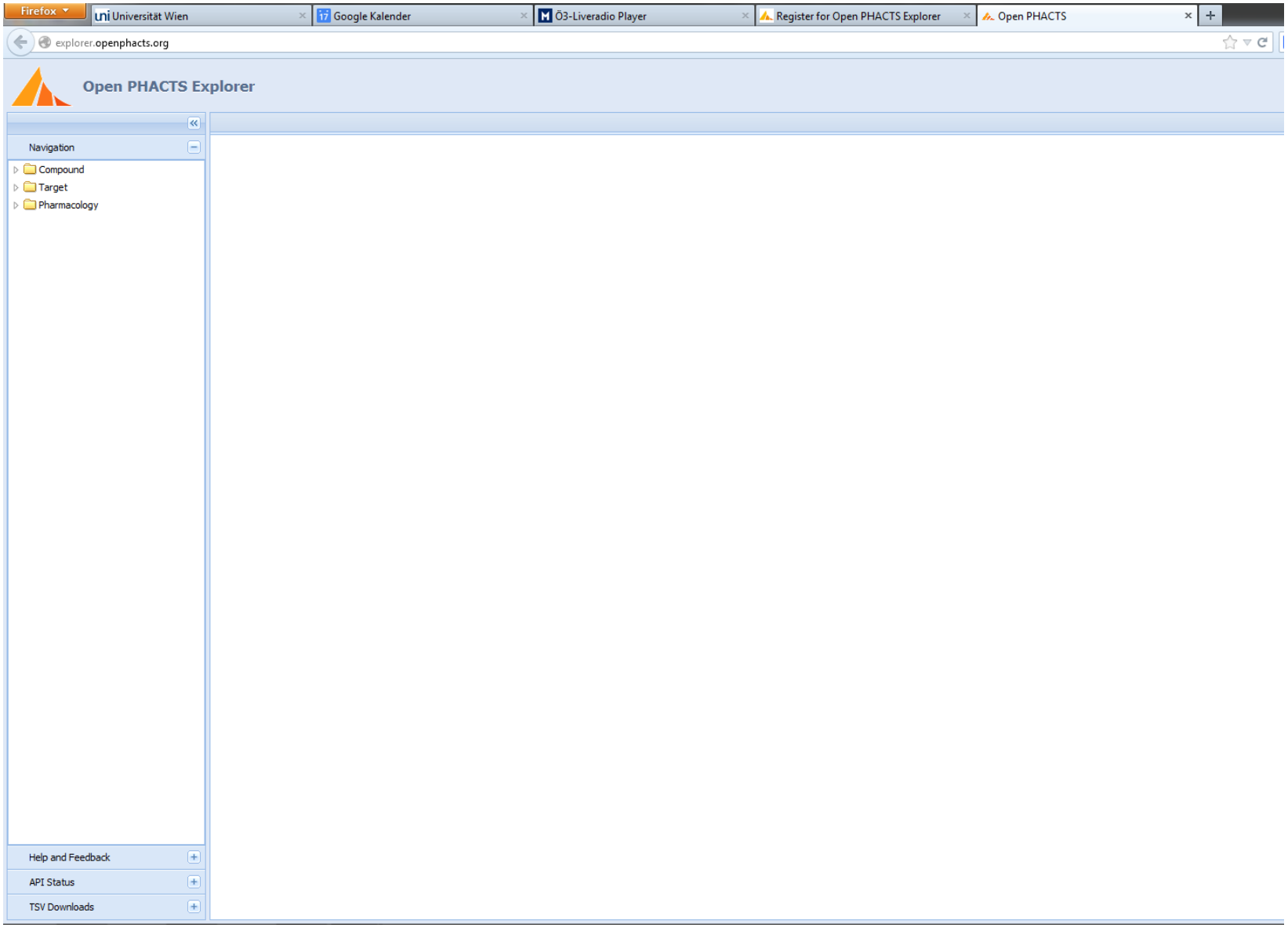
Registration and access to the Open PHACTS Explorer

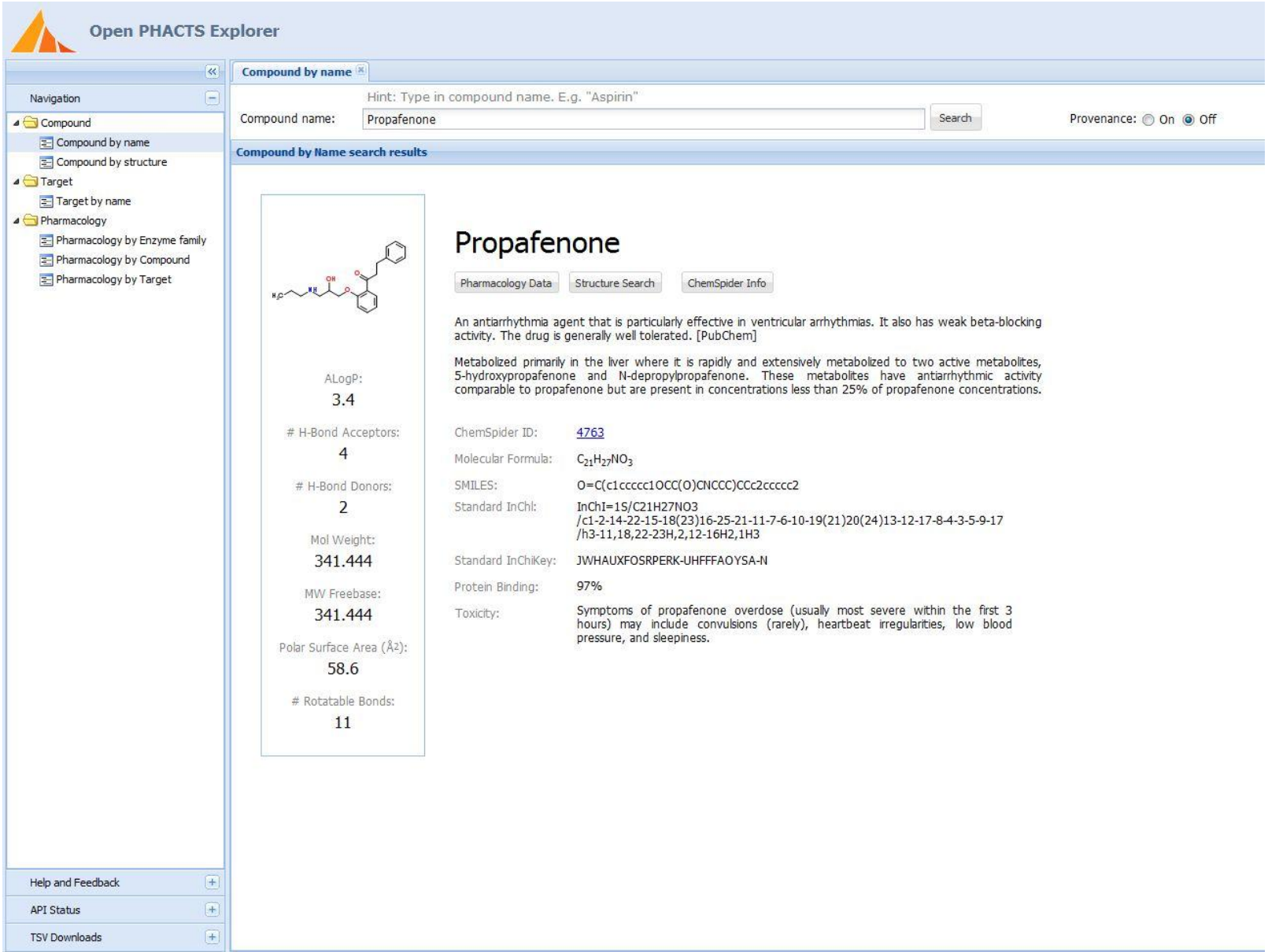
Known issues with beta Open PHACTS Explorer

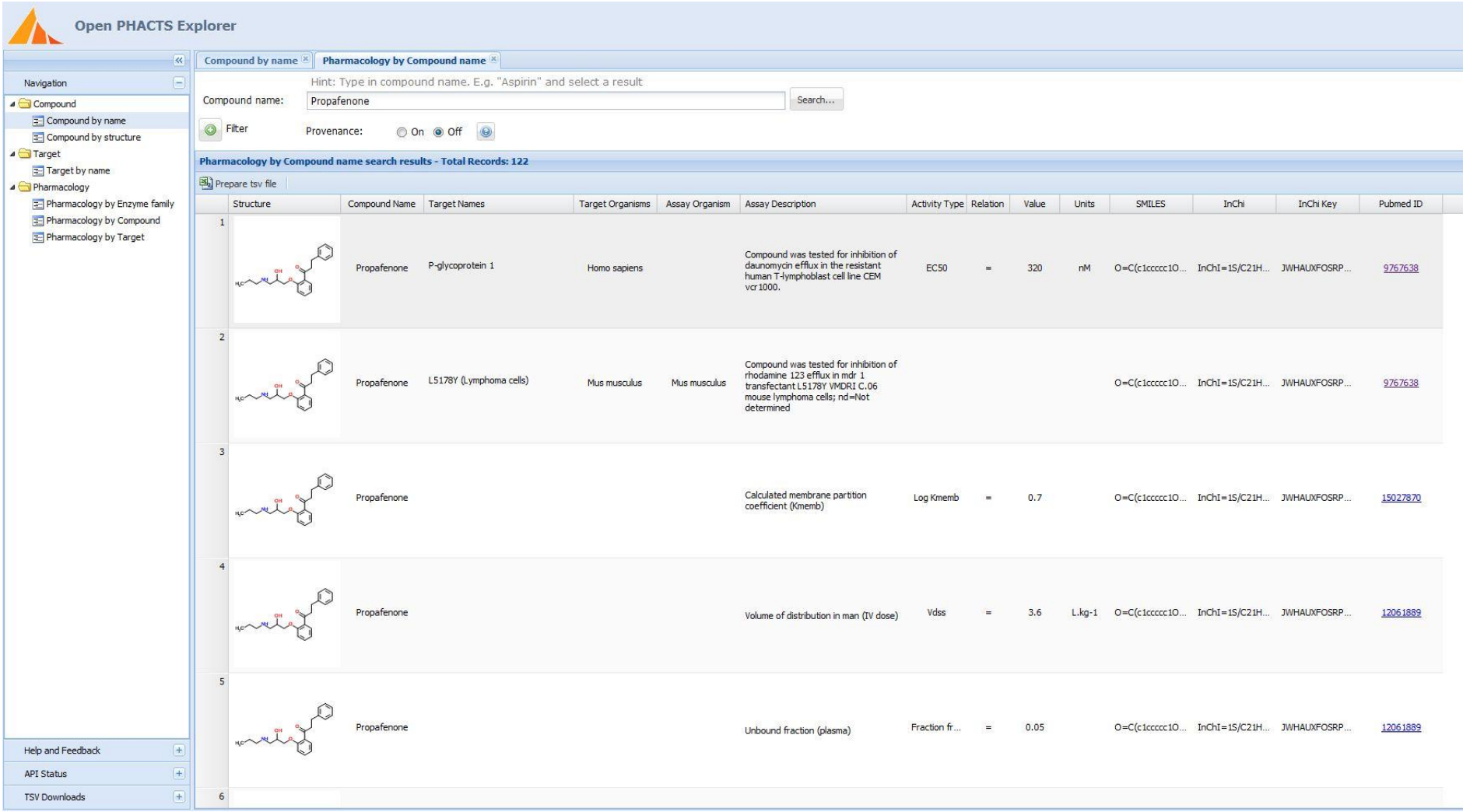
Feedback

efpia | imi

Open PHACTS
Open Pharmacological Space




Open PHACTS Explorer
Navigation
Compound
Compound by name
Compound by structure
Target
Target by name
Pharmacology
Pharmacology by Enzyme family
Pharmacology by Compound
Pharmacology by Target
Help and Feedback
API Status
TSV Downloads
Compound by name
Hint: Type in compound name. E.g. "Aspirin"
Compound name:
Propafenone
Search
Provenance: On Off
Compound by Name search results
ALogP:
3.4
H-Bond Acceptors:
4
H-Bond Donors:
2
Mol Weight:
341.444
MW Freebase:
341.444
Polar Surface Area (Å2):
58.6
Rotatable Bonds:
11
Propafenone
Pharmacology Data
Structure Search
ChemSpider Info
An antiarrhythmia agent that is particularly effective in ventricular arrhythmias. It also has weak beta-blocking activity. The drug is generally well tolerated. [PubChem]
Metabolized primarily in the liver where it is rapidly and extensively metabolized to two active metabolites, 5-hydroxypropafenone and N-depropylpropafenone. These metabolites have antiarrhythmic activity comparable to propafenone but are present in concentrations less than 25% of propafenone concentrations.
ChemSpider ID: 4763
Molecular Formula: C21H27NO3
SMILES: O=C(c1ccccc1OCC(O)CNCCC)CCc2ccccc2
Standard InChI: InChI=1S/C21H27NO3 /c1-2-14-22-15-18(23)16-25-21-11-7-6-10-19(21)20(24)13-12-17-8-4-3-5-9-17 /h3-11,18,22-23H,2,12-16H2,1H3
Standard InChiKey: JWHAUXFOSRPERK-UHFFFAOYSA-N
Protein Binding: 97%
Toxicity: Symptoms of propafenone overdose (usually most severe within the first 3 hours) may include convulsions (rarely), heartbeat irregularities, low blood pressure, and sleepiness.



	Structure	Compound Name	Target Names	Target Organisms	Assay Organism	Assay Description	Activity Type	Relation	Value	Units	SMILES	InChi	InChi Key	Pubmed ID
1		Propafenone	P-glycoprotein 1	Homo sapiens		Compound was tested for inhibition of daunomycin efflux in the resistant human T-lymphoblast cell line CEM vcr1000.	EC50	=	320	nM	O=C(c1ccccc1O...	InChI=1S/C21H...	JWHAUXFOSRP...	9767638
2		Propafenone	L5178Y (Lymphoma cells)	Mus musculus	Mus musculus	Compound was tested for inhibition of rhodamine 123 efflux in mdr 1 transfectant L5178Y VMDRI C.06 mouse lymphoma cells; nd=Not determined					O=C(c1ccccc1O...	InChI=1S/C21H...	JWHAUXFOSRP...	9767638
3		Propafenone				Calculated membrane partition coefficient (Kmemb)	Log Kmemb	=	0.7		O=C(c1ccccc1O...	InChI=1S/C21H...	JWHAUXFOSRP...	15027870
4		Propafenone				Volume of distribution in man (IV dose)	Vdss	=	3.6	L.kg-1	O=C(c1ccccc1O...	InChI=1S/C21H...	JWHAUXFOSRP...	12061889
5		Propafenone				Unbound fraction (plasma)	Fraction fr...	=	0.05		O=C(c1ccccc1O...	InChI=1S/C21H...	JWHAUXFOSRP...	12061889
6														

Open PHACTS Explorer

Compound by name | Pharmacology by Compound name | Compound Structure Search | Target by name

Navigation

- Compound
  - Compound by name
  - Compound by structure
- Target
  - Target by name
- Pharmacology
  - Pharmacology by Enzyme family
  - Pharmacology by Compound
  - Pharmacology by Target

Help and Feedback

API Status

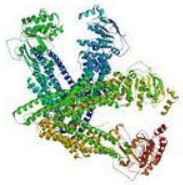
TSV Downloads

Hint: Start typing in protein name and species. E.g. "Adenosine receptor A2a (Homo sapiens)"

Target name: Multidrug resistance protein 1A (Mus musculus) Search

Provenance: On Off

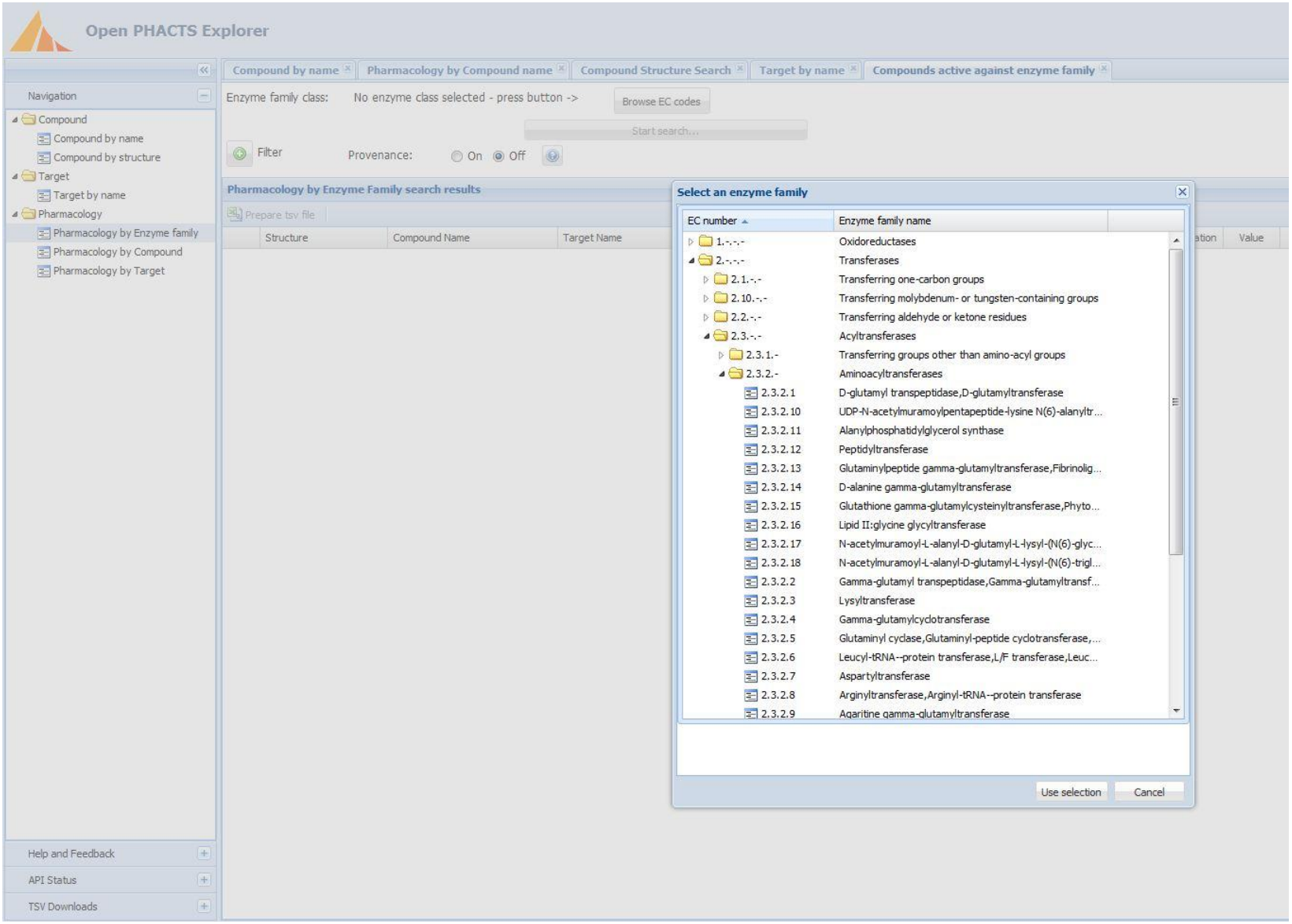
Target Data



Multidrug resistance protein 3 (Mus musculus)

Pharmacology Data

Description:	Multidrug resistance protein 3
Synonyms:	ATP-binding cassette sub-family B member 1A, CHEMBL2573, MDR1A, Multidrug resistance protein 3, P-glycoprotein 3
Specific Function:	Energy-dependent efflux pump responsible for decreased drug accumulation in multidrug-resistant cells.
Keywords:	Hydrolase, Phosphoprotein, 3D-structure, ATP-binding, Cell membrane, Complete proteome, Direct protein sequencing, Glycoprotein, Membrane, Nucleotide-binding, Reference proteome, Repeat, Transmembrane, Transmembrane helix, Transport
PDB Entry:	3G61 3G5U 3G60


Open PHACTS Explorer
Navigation
Compound
Compound by name
Compound by structure
Target
Target by name
Pharmacology
Pharmacology by Enzyme family
Pharmacology by Compound
Pharmacology by Target
Help and Feedback
API Status
TSV Downloads
Compound by name
Pharmacology by Compound name
Compound Structure Search
Target by name
Compounds active against enzyme family
Enzyme family class:
No enzyme class selected - press button ->
Browse EC codes
Start search...
Filter
Provenance:
On
Off
Pharmacology by Enzyme Family search results
Prepare tsv file
Structure
Compound Name
Target Name
Value
Select an enzyme family
EC number
Enzyme family name
1.-.-.- Oxidoreductases
2.-.-.- Transferases
2.1.-.- Transferring one-carbon groups
2.10.-.- Transferring molybdenum- or tungsten-containing groups
2.2.-.- Transferring aldehyde or ketone residues
2.3.-.- Acyltransferases
2.3.1.- Transferring groups other than amino-acyl groups
2.3.2.- Aminoacyltransferases
2.3.2.1 D-glutamyl transpeptidase,D-glutamyltransferase
2.3.2.10 UDP-N-acetylmuramoylpentapeptide-lysine N(6)-alanyltr...
2.3.2.11 Alanylphosphatidylglycerol synthase
2.3.2.12 Peptidyltransferase
2.3.2.13 Glutaminylpeptide gamma-glutamyltransferase,Fibrinolig...
2.3.2.14 D-alanine gamma-glutamyltransferase
2.3.2.15 Glutathione gamma-glutamylcysteinyltransferase,Phyto...
2.3.2.16 Lipid II:glycine glycyltransferase
2.3.2.17 N-acetylmuramoyl-L-alanyl-D-glutamyl-L-lysyl-(N(6)-glyc...
2.3.2.18 N-acetylmuramoyl-L-alanyl-D-glutamyl-L-lysyl-(N(6)-trigl...
2.3.2.2 Gamma-glutamyl transpeptidase,Gamma-glutamyltransf...
2.3.2.3 Lysyltransferase
2.3.2.4 Gamma-glutamylcyclotransferase
2.3.2.5 Glutaminyl cyclase,Glutaminyl-peptide cyclotransferase,...
2.3.2.6 Leucyl-tRNA--protein transferase,L/F transferase,Leuc...
2.3.2.7 Aspartyltransferase
2.3.2.8 Arginyltransferase,Arginyl-tRNA--protein transferase
2.3.2.9 Agaritine gamma-glutamyltransferase
Use selection
Cancel

Open PHACTS
Open Pharmacological Space

Open PHACTS Explorer

Navigation

- Compound
  - Compound by name
  - Compound by structure
- Target
  - Target by name
- Pharmacology
  - Pharmacology by Enzyme family
  - Pharmacology by Compound
  - Pharmacology by Target

Help and Feedback

API Status

TSV Downloads

Compound by name | Pharmacology by Compound name | Compound Structure Search | Target by name | Compounds active against enzyme family | Pharmacology by Target Name

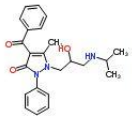
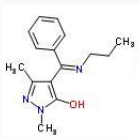
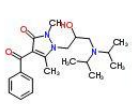
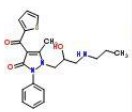
Hint: Type in protein name and species. E.g. "ADA protein human" and select a result

Protein name: Multidrug resistance protein 1 (Homo sapiens) Search...

Filter Provenance: On Off

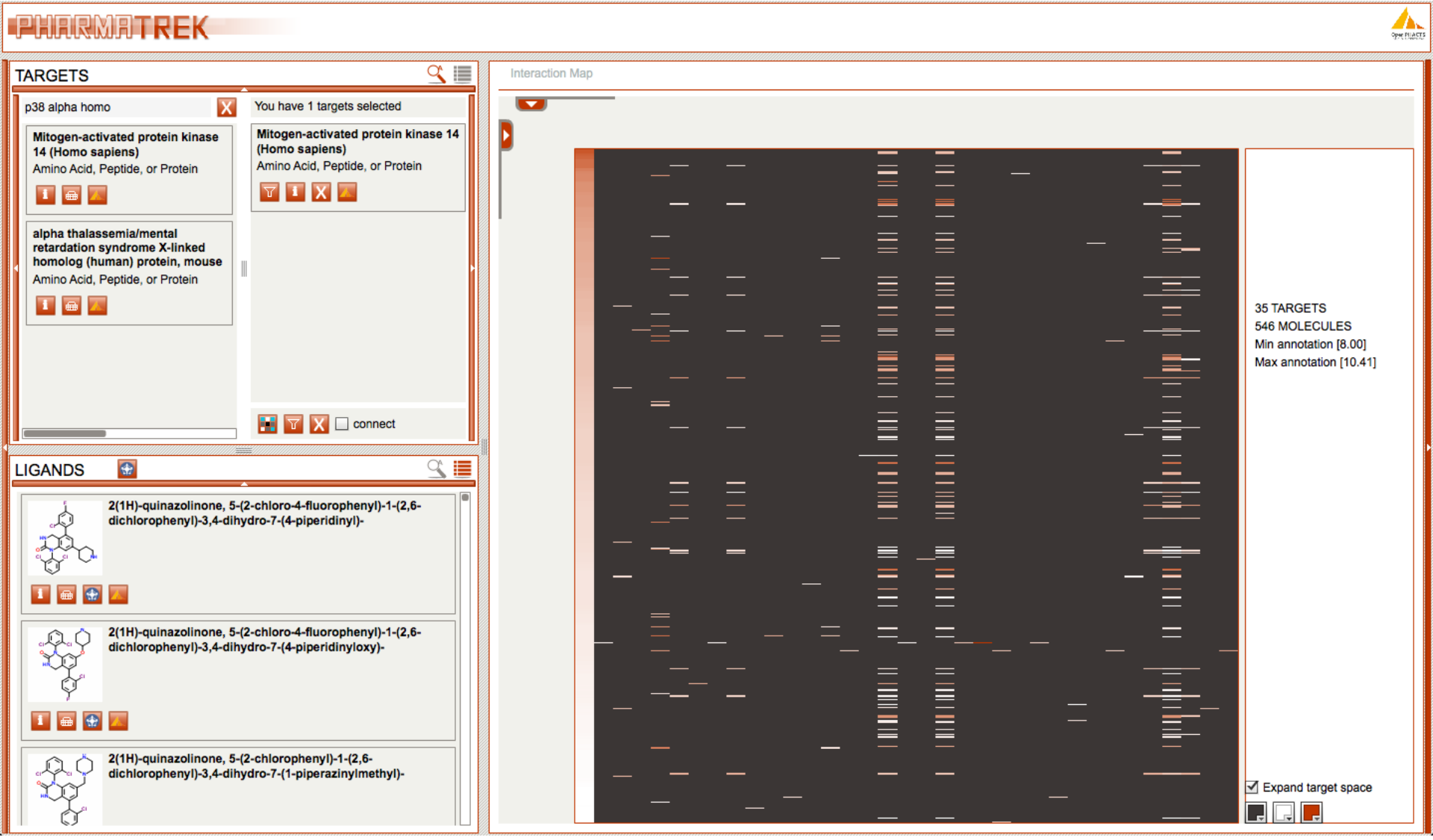
Pharmacology by Target name search results - Total Records: 2563

Prepare tsv file

	Structure	Compound Name	Target Name	Target Organism	Assay Organism	Assay Description	Activity Type	Relation	Value	Units	Mol Weight	SMILES	InChi	InChi Key
1		4-benzoyl-1-[2-hydroxy-3-(propan-2-ylamino)propyl]-5-methyl-2-phenyl-1,2-dihydro-3H-pyrazol-3-one	Multidrug resistance protein 1 (Homo sapiens)	Homo sapiens		Compound was tested for inhibition of daunomycin efflux in the resistant human T-lymphoblast cell line CEM vcr1000.	EC50	=	83180	nM	393.479	O=C2\C(=C(/N(...	InChI=1S/C23H...	VZPAOCHCLZMT...
2		1,3-dimethyl-4-[(E)-phenyl(propylimino)methyl...pyrazol-5-ol	Multidrug resistance protein 1 (Homo sapiens)	Homo sapiens		Compound was tested for inhibition of daunomycin efflux in the resistant human T-lymphoblast cell line CEM vcr1000.	EC50	=	272500	nM	257.331	CCC/N=C(\c1cc...	InChI=1S/C15H...	GJPFQWWTOPA...
3		4-benzoyl-1-[3-(dipropan-2-ylamino)-2-hydroxypropyl]-2,5-dimethyl-1,2-dihydro-3H-pyrazol-3-one	Multidrug resistance protein 1 (Homo sapiens)	Homo sapiens		Compound was tested for inhibition of daunomycin efflux in the resistant human T-lymphoblast cell line CEM vcr1000.	EC50	=	24120	nM	373.489	O=C1\C(=C(/N(...	InChI=1S/C21H...	UYZSJNDVHQW...
4		1-[2-Hydroxy-3-(propylamino)propyl]-5-methyl-2-phenyl-4-(2-thienylcarbonyl)-1,2-dihydro-3H-pyrazol-3-one	Multidrug resistance protein 1 (Homo sapiens)	Homo sapiens		Compound was tested for inhibition of daunomycin efflux in the resistant human T-lymphoblast cell line CEM vcr1000.	EC50	=	72440	nM	399.507	O=C(C=2C(=O)...	InChI=1S/C21H...	JCKJIYACBBEC...
5		1-[2-hydroxy-3-(propan-2-ylamino)propyl]-5-methyl-2-phenyl-4-(3-phenylpropanoyl)-1,2-dihydro-3H-pyrazol-3-one	Multidrug resistance protein 1 (Homo sapiens)	Homo sapiens		Compound was tested for inhibition of daunomycin efflux in the resistant human T-lymphoblast cell line CEM vcr1000.	EC50	=	11750	nM	421.532	O=C2\C(=C(/N(...	InChI=1S/C25H...	DNSZEGKDVYSK...
6														

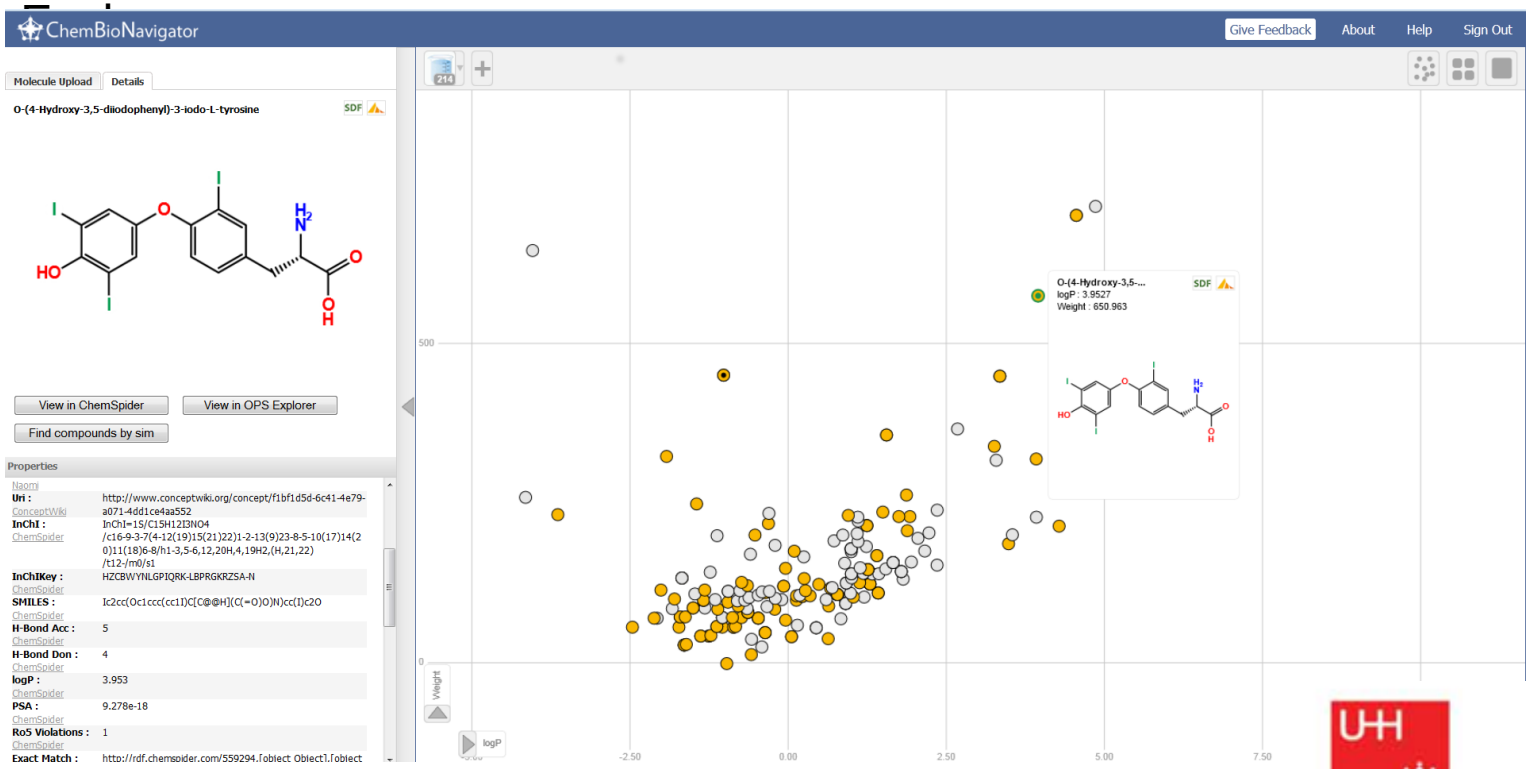
PharmaTrek

Open PHACTS
Open Pharmacological Space



Example Applications: ChemBioNavigator

The ChemBioNavigator allows the user to **visualise the chemical and biological space of a molecule group** in a chemically-aware manner. Individual data points can be investigated further via direct links to ChemSpider and the Open PHACTS

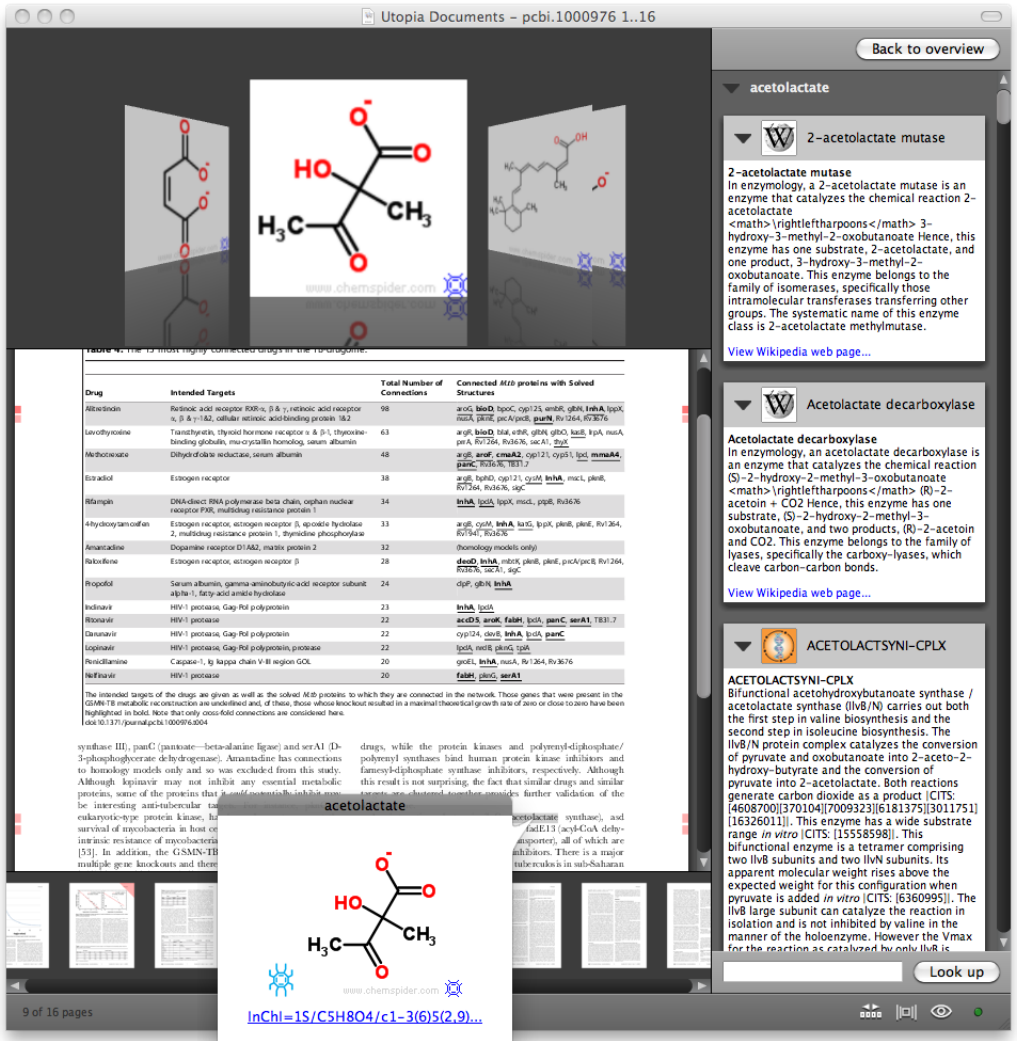


Developed by the University of Hamburg and BioSolveIT

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Utopia Documents

Open PHACTS
Open Pharmacological Space



Outlook

Open PHACTS
Open Pharmacological Space

- Launch of API at Community Workshop, April 22/23, London
- Open PHACTS Training Event, Sep 20, Vienna
- 2nd wave use cases (pathway, disease)
- Sustainability – Open PHACTS Foundation

www.openphacts.org