

Open PHACTS

Open Pharmacological Space

A Data Platform for Drug Discovery

Paul Groth (@pgroth)

<http://www.few.vu.nl/~pgroth>



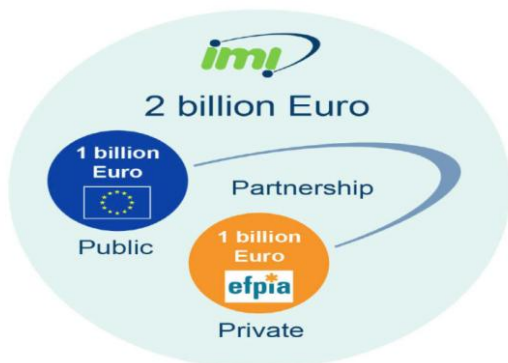


- 1. WHY**
- 2. THE PLATFORM**
- 3. APPS**
- 4. THE FUTURE**



Open PHACTS

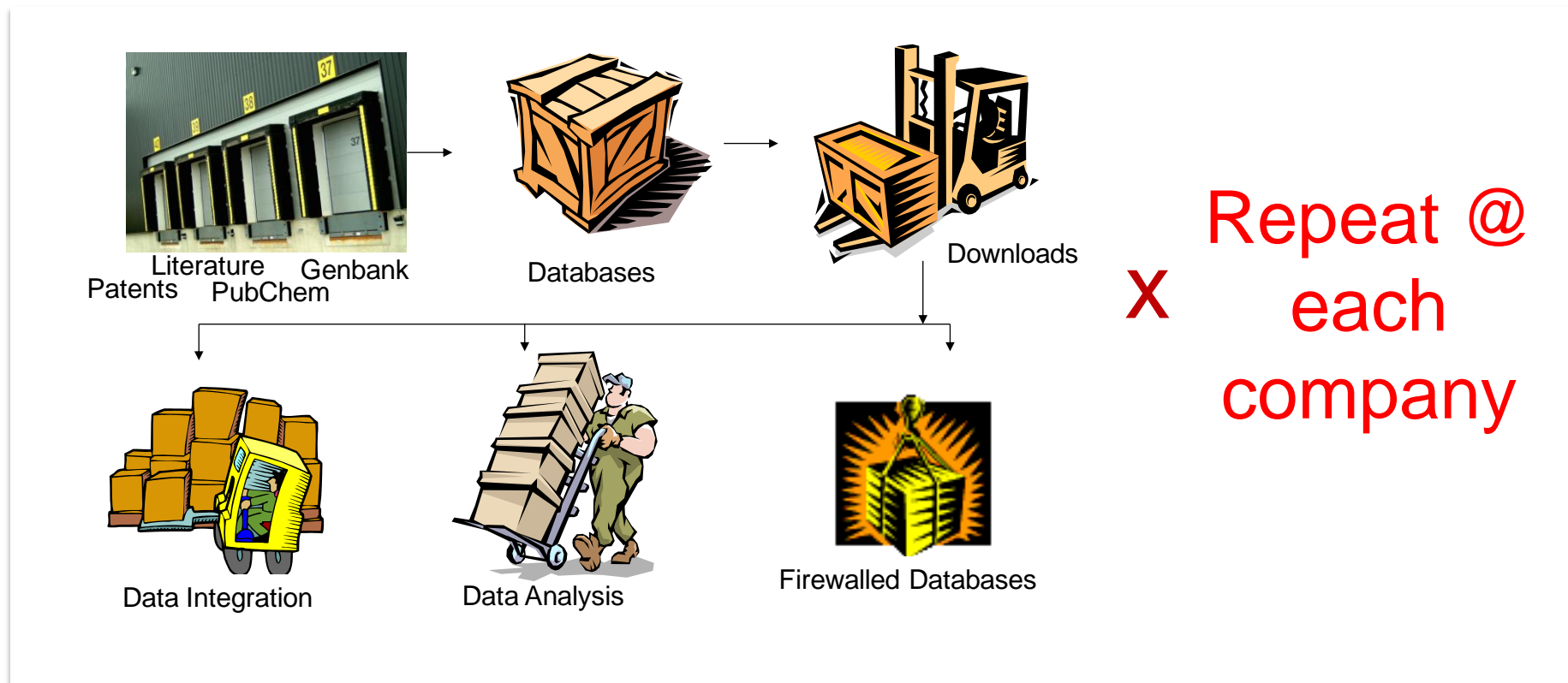
Open Pharmacological Space





Pre-competitive Informatics:

Pharma are all accessing, processing, storing & re-processing external research data



Lowering industry firewalls: pre-competitive informatics in drug discovery
Nature Reviews Drug Discovery (2009) 8, 701-708 doi:10.1038/nrd2944



Business Question Driven Approach

Number	sum	Nr of 1	Question
15	12	9	All oxidoreductase 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 substructure (with options to match)
41	13	8	A project is considering Protein X. Retrieve all active compounds known to modulate the target directly? i.e. return all compounds that bind to the target at a certain level of the target family (i.e. PK)
44	13	8	Give me all active compounds of a given target
46	13	8	Give me the compound(s) which are most similar to a given compound (disease)
59	14	8	Identify all known protein-protein interactions



Drug Discovery Today

Volume 18, Issues 17–18, September 2013, Pages 843–852



Review

Scientific competency questions as the basis for semantically enriched open pharmacological space development

Kamal Azzaoui¹, Edgar Jacoby¹⁴, Stefan Senger², Emiliano Cuadrado Rodríguez³, Mabel Loza³, Barbara Zdrzil⁴, Marta Pinto⁴, Antony J. Williams⁵, Victor de la Torre⁶, Jordi Mestres⁷, Manuel Pastor⁷, Olivier Taboureau⁸, Matthias Rarey⁹, Christine Chichester¹⁰, Steve Pettifer¹¹, Niklas Blomberg^{12, a}, Lee Harland¹³, Bryn Williams-Jones¹³, Gerhard F. Ecker⁴.  



"What is the selectivity profile of known p38 inhibitors?"



"Let me compare MW, logP and PSA for known oxidoreductase inhibitors"



"Find me compounds that inhibit targets in NFkB pathway assayed in only functional assays with a potency <math>< 1 \mu\text{M}</math>"



ChEMBL

DrugBank

Gene
Ontology

Wikipathways

GeneGo

ChEBI

UniProt

UMLS

GVKBio

ConceptWiki

ChemSpider

TrialTrove

TR Integrity



▶ 1st Hackathon



▶ Hackathon 2

▶ Focused User Feedback



▶ Hackathon 3

▶ Alpha for public release

▶ Platform On Hosting Provider

▶ Revised Platform Implementation



▶ Prototype

▶ Hosting Selected



▶ Hosting Partner On-board

▶ Linked Data API

▶ Usathon

▶ Public Release

▶ Drive Team In-Place

▶ Project Start

▶ Tech Team Kickoff

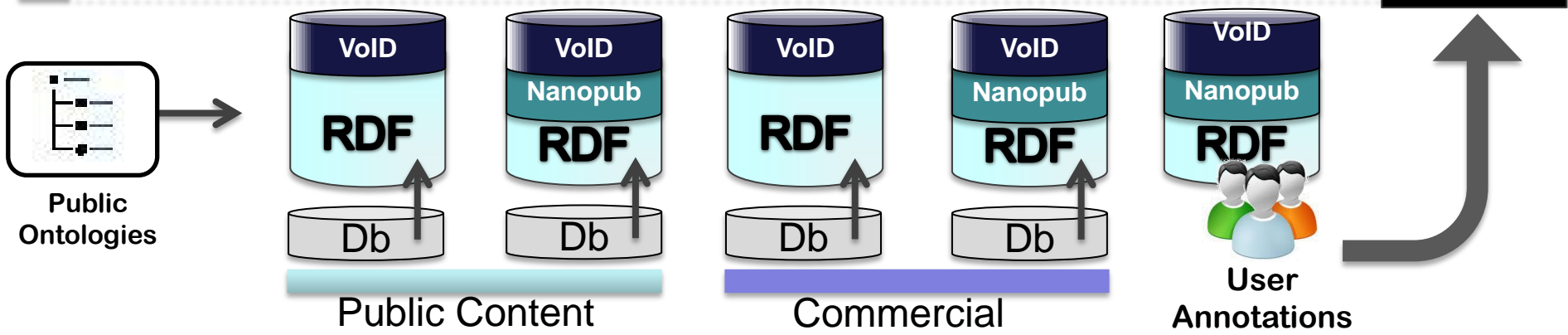
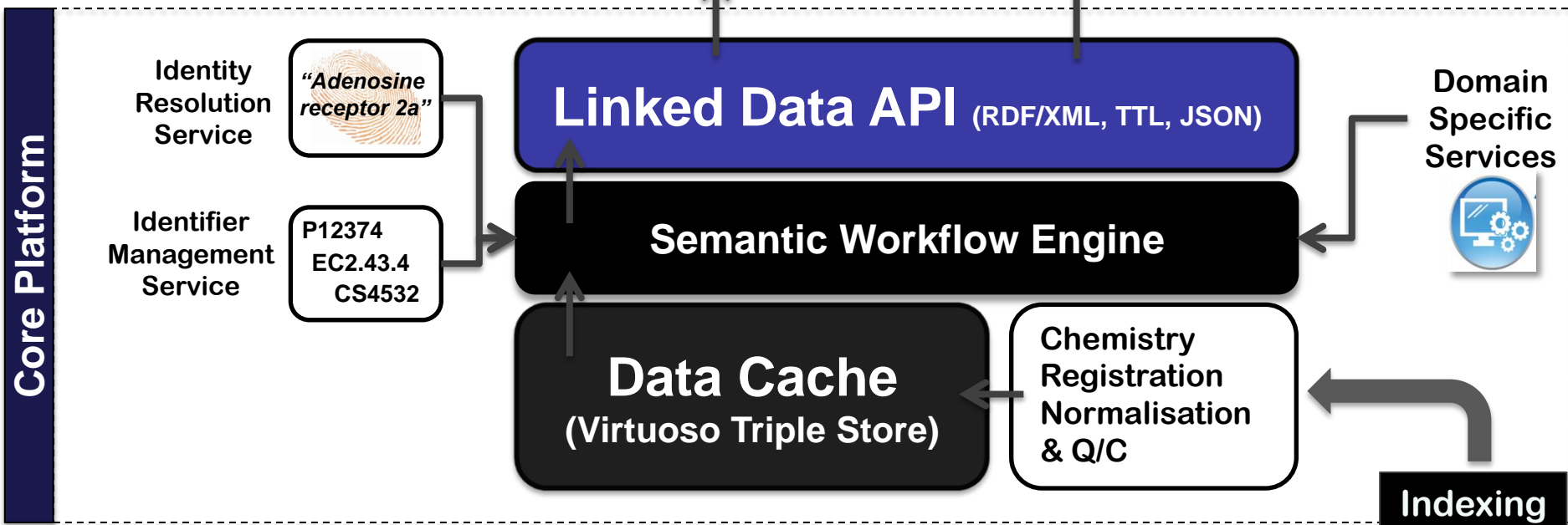
Open Phacts



THE OPEN PHACTS DISCOVERY PLATFORM

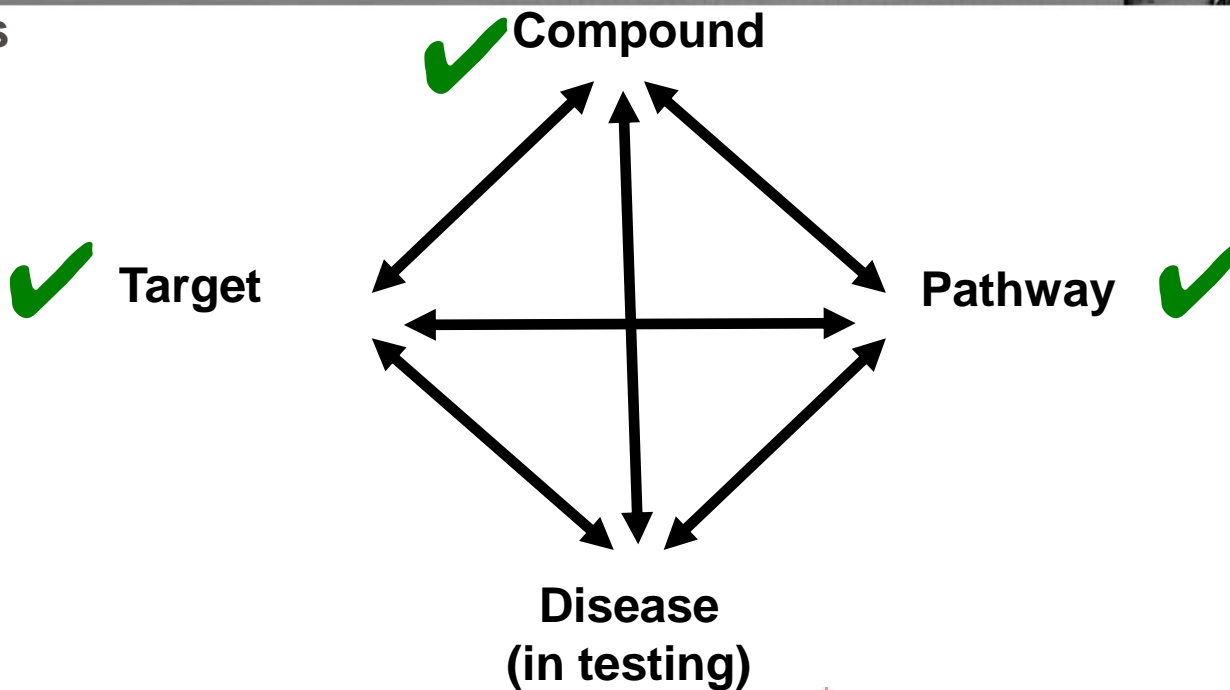


Apps

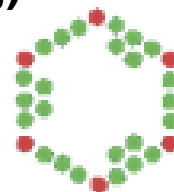




Data Sources



ChEMBL



WikiPATHWAYS
Pathways for the People





Play!

<https://dev.openphacts.org/>

OpenPHACTS API

Chemical Structure Exact Search	<code>/structure/exact</code> GET
InchiKey to URL	<code>/structure</code> GET
Inchi to URL	<code>/structure</code> GET
Chemical Structure Similarity Search	<code>/structure/similarity</code> GET
SMILES to URL	<code>/structure</code> GET
Chemical Structure Substructure Search	<code>/structure/substructure</code> GET
Get concept description	<code>/getConceptDescription</code> GET
Map free text to a concept URL based on semantic tag	<code>/search/byTag</code> GET
Map URL	<code>/mapURL</code> GET
Map free text to a concept URL	<code>/search/freetext</code> GET
Get ChEBI Ontology Class Members	<code>/compound/chebi/members</code> GET
Get ChEBI Ontology Root Classes	<code>/compound/chebi/root</code> GET
Get ChEBI Ontology Class	<code>/compound/chebi/node</code> GET
ChEBI Class Pharmacology Count	<code>/compound/chebi/pharmacology/count</code> GET

PARAMETER	VALUE	DESCRIPTION
app_id	<input type="text"/>	Your access application id
app_key	<input type="text"/>	Your access application key
searchOptions.Molecule	<input type="text" value="(required)"/>	A SMILES string. E.g. <chem>CC(=O)Oc1ccccc1C(=O)O</chem>
searchOptions.SimilarityType	<input type="text"/>	0: Tanimoto ; 1: Tversky ; 2: Euclidian
searchOptions.Threshold	<input type="text"/>	Double <= 1.0
commonOptions.Complexity	<input type="text"/>	(Not supported at the moment) 0: Any ; 1: Single ; 2: Multi
commonOptions.Isotopic	<input type="text"/>	(Not supported at the moment) 0: Any ; 1: Labeled ; 2: NotLabeled
commonOptions.HasSpectra	<input type="text"/>	(Not supported at the moment) Boolean
commonOptions.HasPatents	<input type="text"/>	(Not supported at the moment) Boolean
resultOptions.Limit	<input type="text"/>	Integer. Search limit. Specify how many results return back during the search. Default value: -1 .
resultOptions.Start	<input type="text"/>	Integer. Return results starting the index. Default value: 0
resultOptions.Length	<input type="text"/>	Integer. How many results should be returned starting from Start index. Default value: -1.



Secure Cloud Hosted + Virtualized

Triple Store

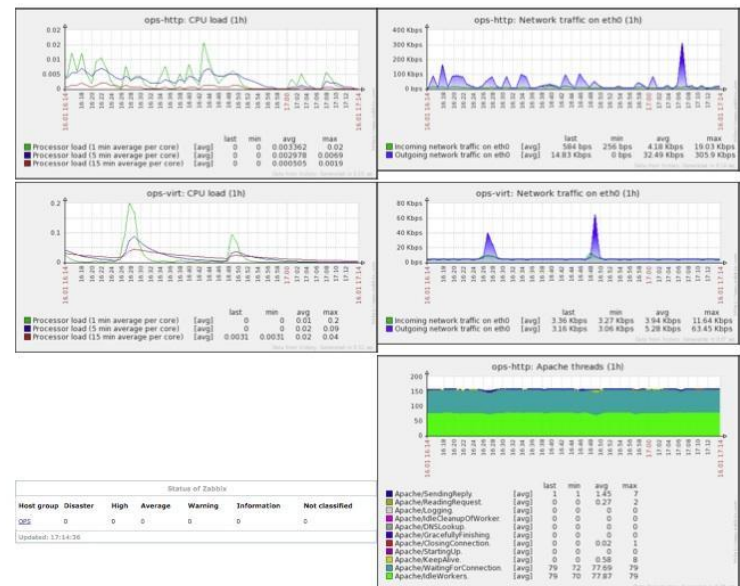
- Virtuoso 7 column store
- Scale to > 100 billion triples

Network

- AMX-IS
- Extensive memcache
- Monitored

Hardware (development)

- 2 x Intel Xeon E5-2640 - 384 GB
- DDR3 1333MHz RAM - 1.5 TB
- SSD - 3TB 7200rpm





Dealing With The *Really* Tough Parts

Data Licensing

John
Wilbanks
<http://del-fi.org/>



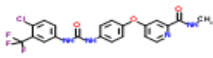
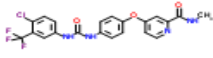
Compatibility chart		Terms that may be used for a derivative work or adaptation						
		BY	BY-NC	BY-NC-ND	BY-NC-SA	BY-ND	BY-SA	PD
Status of original work	PD	■	■	■	■	■	■	■
	BY	■	■	■	■	■	■	
	BY-NC		■	■	■			
	BY-NC-ND							
	BY-NC-SA				■			
	BY-ND							
	BY-SA						■	



Filter Provenance: On Off

Pharmacology by Compound name search results - Total Records

Prepare full result set download Download CSV-file Download SD-file

Structure	Compound name	Target Organism	Target Name
45 	Sorafenib	Homo sapiens	Ephrin type-A receptor 7
46 	Sorafenib	Homo sapiens	MLDQJTXFUGDVEO Stem cell growth factor receptor

Provenance Datasources

- ConceptWiki
- ChemSpider
- Drugbank
- ChEMBL

The Open PHACTS VoID Editor

Input: Dataset characteristics

VoID Metadata

VoID Title

Example VoID Document

VoID Description

Example description of the VoID document

VoID Number

Provenance

Please select the origin of the data from the following options and then complete the resulting form.

- Using Original Data Location
- Downloaded a copy of the original data
- Data format conversion, e.g. RDBMS -> RDF
- Data derived from another source

Downloaded from

Provide a URI for the location where the data was downloaded from.

Dataset Version

If appropriate, provide the version number for the original dataset, e.g. for ChEMBL this could be '13'.

Date downloaded

Enter the date when the data was downloaded.

Downloaded by

Enter an identifier for the person or organisation who downloaded the data.

Provenance everywhere



Its easy to integrate, difficult to integrate well:

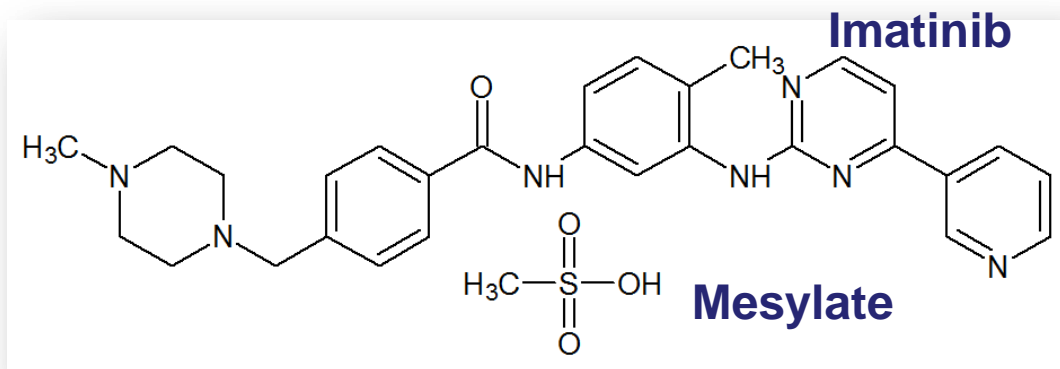

Type a compound name:

glee

- Gleevec
- Gleevec



What Is Gleevec?

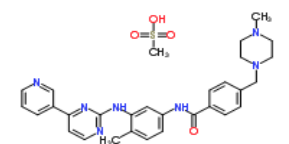



ChemSpider
The free chemical database

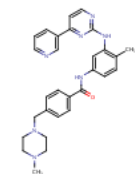
About | More Searches | Web APIs

Gleevec

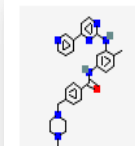
ChemSpider ID: 123596
Molecular Formula: C₂₉H₃₁N₇O₄S
Average mass: 589.708400
Monoisotopic mass: 589.708400
Systematic name: 4-[[4-(4-methyl-1-piperazinyl)methyl]phenyl]pyrimidin-2-yl]pyridin-3-ylmethanesulfonamide



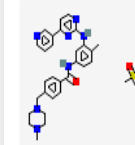
ChemSpider

Structure	 Download: MOL SDF SMILES InChI Display: 2D Structure 3D Structure
Synonyms	<ul style="list-style-type: none"> Imatinib Mesylate Imatinib Methanesulfonate STI-571
Brand names	<ul style="list-style-type: none"> Gleevec Glivec

Drugbank



[Imatinib; 152459-95-5; sti-571 ...](#)
 MW: 493.602740 g/mol MF: C₂₉H₃₁N₇O
 IUPAC name: 4-[[4-(4-methylpiperazin-1-yl)methyl]phenyl]pyrimidin-2-yl]pyridin-3-ylmethanesulfonamide
 Active in 205 BioAssays | Tested in 1376 BioAssays
 CID: 5291
[Similar Compounds](#) | [Same Parent, Connectivity](#) | [\(MeSH Keyword\)](#)

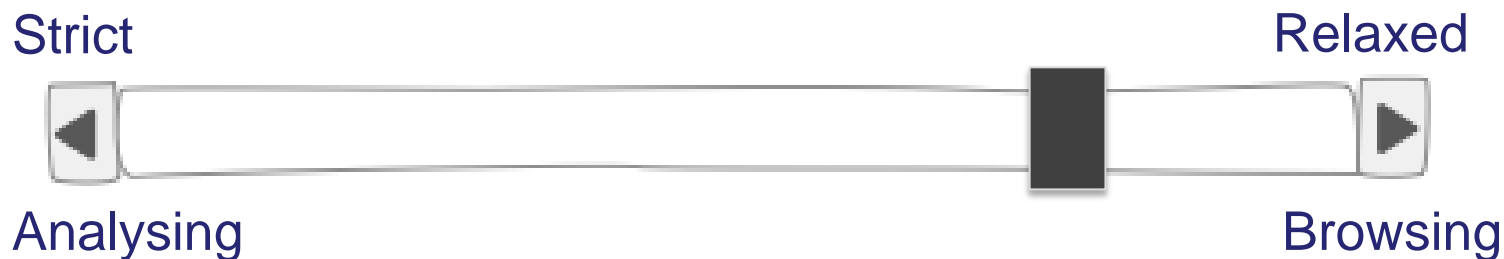


[Imatinib mesylate; Gleevec; Glivec ...](#)
 MW: 589.708400 g/mol MF: C₃₀H₃₅N₇O₄S
 IUPAC name: methanesulfonic acid; 4-[[4-(4-methylpiperazin-1-yl)methyl]phenyl]pyrimidin-2-yl]pyridin-3-ylmethanesulfonamide
 Active in 35 BioAssays | Tested in 679 BioAssays
 CID: 123596
[Similar Compounds](#) | [Same Parent, Connectivity](#) | [\(MeSH Keyword\)](#)

PubChem



Dynamic Equality



chemspider:gleevec

drugbank:gleevec

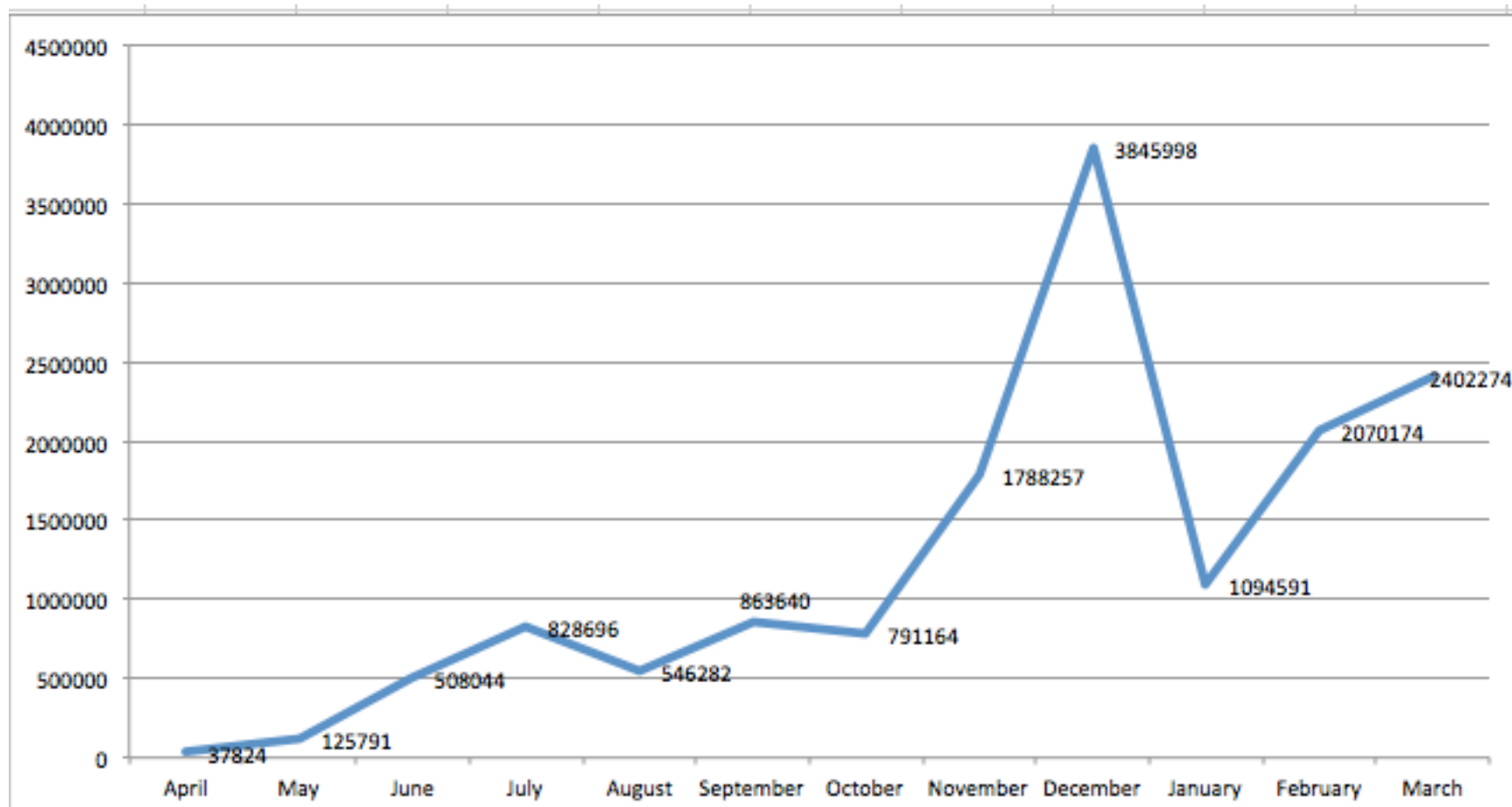
```
LinkSet#1 {  
  chemspider:gleevec hasParent imatinib ...  
  drugbank:gleevec exactMatch imatinib ...  
}
```



APPS



API Hits (April 2013 – March 2014)





Open PHACTS

Browse and search the data within the Open PHACTS Discovery Platform.

✦ Developed by the **University of Manchester** and **University of Vienna**



ChemBioNavigator

Visualise the chemical and biological space of a molecule group in a chemically-aware manner.

✦ Developed by the **University of Hamburg** and **BioSolveIT GmbH**

PHARMATREK

Navigate pharmacological space in a flexible and interactive way.

✦ Developed by the **Consorti Mar Parc de Salut de Barcelona (PSMAR)**

SciBite



Connects the latest news and events in Pharma and Biotech directly to pharmacology data within the Open PHACTS platform.

✦ Developed by **SciBite Limited**



utopia

Allows the semantic enrichment of scientific articles in PDF format.

✦ Developed by the **University of Manchester**



GARfield

Intuitive predicts target pharmacology based on the Similar Ensemble Approach.

✦ Developed by the **Technical University of Denmark**



collector

Extracts data to build QSAR predictive models with data from the eTOX project.

✦ Developed by **PSMAR** as part of the **eTOX project**



accelrys[®]

Pipeline Pilot

A repository of useful Pipeline Pilot components and workflows has been developed.

🏠 **Open PHACTS - Pipeline Pilot Community**



KNIME

A KNIME repository of components and workflows has been developed.

🏠 **Open PHACTS - KNIME Community**



Excel

Queries the Open PHACTS API from Microsoft's Excel spreadsheet software.

✦ Developed by the **University of Vienna**



AQknowledge[™]

Semantics for Science

Identifies significant entities in scientific text, and provides links to Open PHACTS Explorer.

✦ Developed by **AQknowledge**



he

Helium for Excel Community Edition contains three functions that use the Open PHACTS API.

✦ Developed by **Ceiba Solutions**



Open PHACTS Explorer

Navigation: Compound, Target, Pharmacology

Target by name: Mitogen-activated protein kinase 14 (Homo sapiens) Search Provenance: On Off

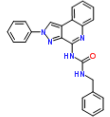
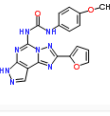
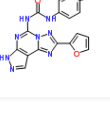
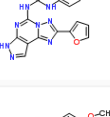
Hint: Start typing in protein name and species. E.g. "Adenosine receptor A2a (Homo sapiens)"

Mitogen-activated protein kinase 14 (Homo sapiens)

Protein name: Adenosine receptor A3 (Homo sapiens) Search...
 Filter Provenance: On Off

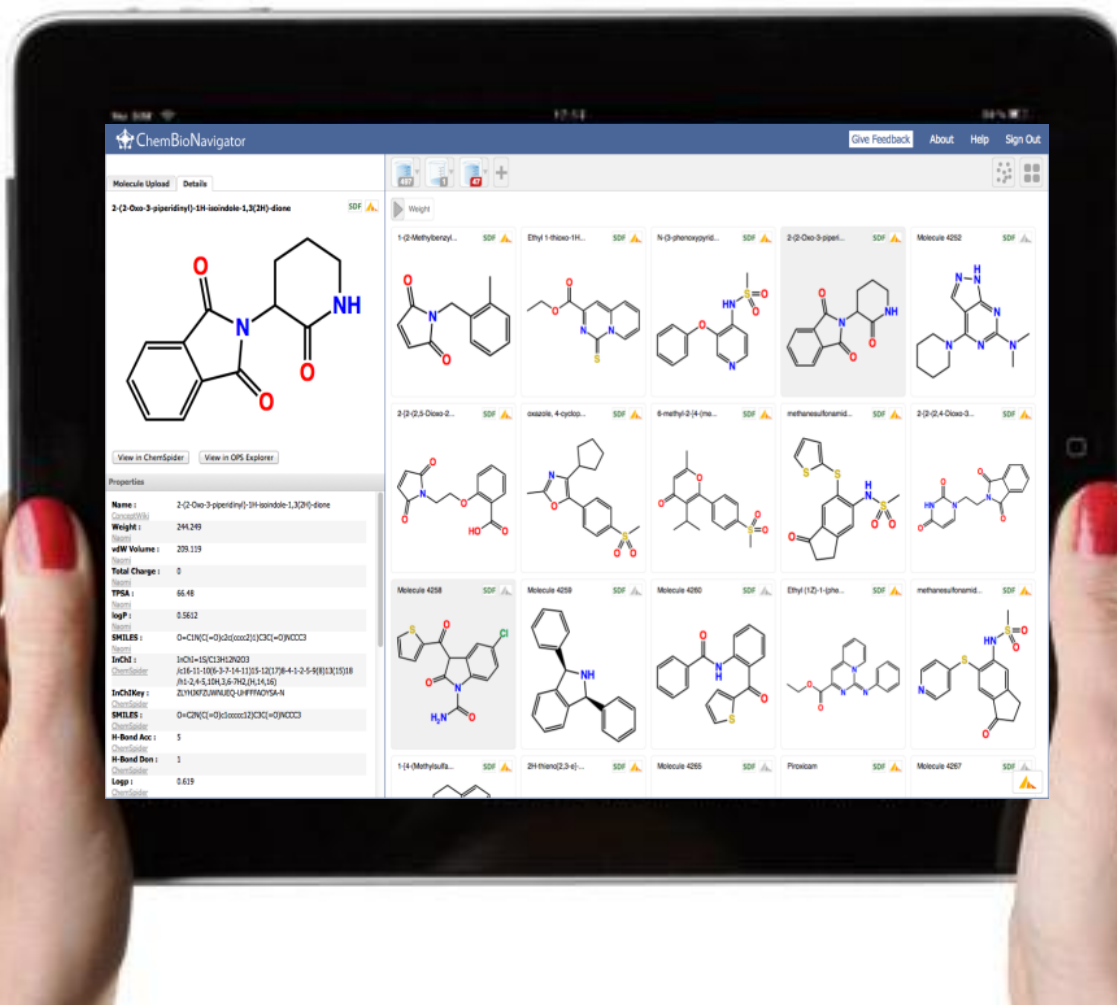
Pharmacology by Target name search results - Total Records: 7887

[Prepare tsv file](#)

Structure	Compound Name	Target Name	Target Organism	Assay Organism	Assay Description	Activity Type	Relation	Value	Units	Mol Weight	SMILES	InChi
	urea, N-(phenylmethyl)-N'-(2-phenyl-2H-pyrazolo[3,4-c]quinolin-4-yl)-	Adenosine receptor A3 (Homo sapiens)	Homo sapiens		Displacement of specific [¹²⁵ I]AB-MECA binding at human adenosine A3 receptor expressed in CHO cells	Ki	=	8.3	nM	393.441	O=C(NC(c1ccccc1)NC2=CN3C=CC=CN3C2=CC=C4C=CC=CC45)C6=CC=CC=C6	InChi=1S/C
	1-[2-(furan-2-yl)-7H-pyrazolo[4,3-e][1,2,4]triazolo[1,5-c]pyrimidin-5-yl]-3-(4-methoxyphenyl)urea	Adenosine receptor A3 (Homo sapiens)	Homo sapiens	Homo sapiens	Displacement of [³ H]MRE3008-F20 from human adenosine A3 receptor expressed in CHO cells; range 0.08-0.27	Ki	=	0.14	nM	390.356	COc1ccc(cc1)N[C@@H]2C=NC3=C(N2)C=CC=C3C4=CC=CC=C45	InChi=1S/C
	1-[2-(furan-2-yl)-7H-pyrazolo[4,3-e][1,2,4]triazolo[1,5-c]pyrimidin-5-yl]-3-(4-methoxyphenyl)urea	Adenosine receptor A3 (Homo sapiens)	Homo sapiens	Homo sapiens	Percent reversal of 100 nM IB-MECA-inhibited cAMP accumulation in CHO cells expressing human A3 adenosine receptor at 1 μM	Inhibition	=	98	%	390.356	COc1ccc(cc1)N[C@@H]2C=NC3=C(N2)C=CC=C3C4=CC=CC=C45	InChi=1S/C
	1-[2-(furan-2-yl)-7H-pyrazolo[4,3-e][1,2,4]triazolo[1,5-c]pyrimidin-5-yl]-3-(4-methoxyphenyl)urea	Adenosine receptor A3 (Homo sapiens)	Homo sapiens	Homo sapiens	Inhibition of cAMP accumulation in CHO cells expressing human A3 adenosine receptor	IC50	=	1.8	nM	390.356	COc1ccc(cc1)N[C@@H]2C=NC3=C(N2)C=CC=C3C4=CC=CC=C45	InChi=1S/C

Feedback

ChemBioNavigator



ChemBioNavigator | Give Feedback | About | Help | Sign Out

Molecule Upload | Details

2-(2-Oxo-3-piperidinyl)-3H-isindole-1,3(2H)-dione SDF

View in ChemSpider | View in OPS Explorer

Properties

Name: 2-(2-Oxo-3-piperidinyl)-3H-isindole-1,3(2H)-dione
 Weight: 244.249
 MW Volume: 209.119
 Total Charge: 0
 TPSA: 66.48
 logP: 0.5632
 SMILES: O=C1N(C(=O)c2ccccc2)C(=O)N1CCCC

Grid of molecules:

- 1-(2-Methylbenzyl)-... SDF
- Ethyl 1-thio-1H... SDF
- N-(3-phenoxypyrid... SDF
- 2-(2-Oxo-3-piper... SDF
- Molecule 4252 SDF
- 2-(2-(2,5-Dioxo-2... SDF
- osazolin, 4-cydo... SDF
- 6-methyl-2(4-phen... SDF
- methanesulfonamid... SDF
- 2-(2-(2,4-Dioxo-3... SDF
- Molecule 4258 SDF
- Molecule 4259 SDF
- Molecule 4260 SDF
- Ethyl (1Z)-1-phen... SDF
- methanesulfonamid... SDF
- 1-(4-(Methylsulfa... SDF
- 2H-thiazolo[2,2-e]... SDF
- Molecule 4265 SDF
- Proxizan SDF
- Molecule 4267 SDF

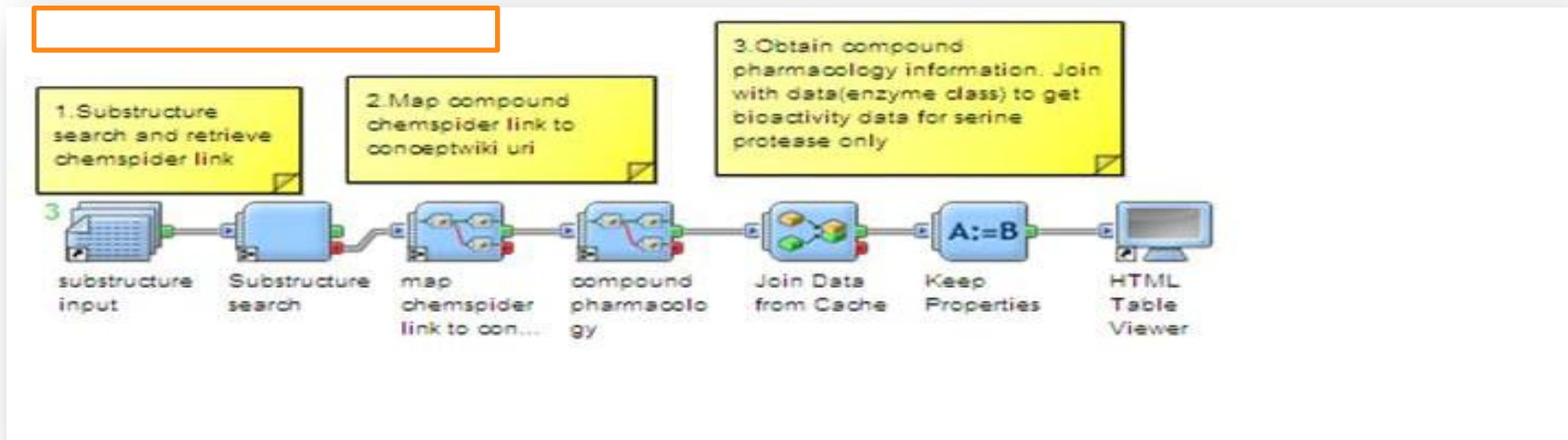
KNIME

Table View - 0:31 - Interactive Table (7 x 6)

Name	Inchi	Activity	Units	Relation	Target
.. Sorafenib	MLDQJTXFUGDVEO-UHFFFAOYSA...	3400	nM	=	Serine/threonine-protein kinase PLK4
.. Sorafenib	MLDQJTXFUGDVEO-UHFFFAOYSA...	250	nM	=	MAP kinase signal-integrating kinase 2
.. Sorafenib	MLDQJTXFUGDVEO-UHFFFAOYSA...	5.4	uM	=	HCT-116 (Colon carcinoma cells)
.. Sorafenib	MLDQJTXFUGDVEO-UHFFFAOYSA...	1700	nM	=	Ephrin type-B receptor 1
.. Sorafenib	MLDQJTXFUGDVEO-UHFFFAOYSA...	3300	nM	=	Dual specificity mitogen-activated protein kinase kin.
.. Sorafenib	MLDQJTXFUGDVEO-UHFFFAOYSA...	6200	nM	=	Cyclin-dependent kinase 5

Workflow Nodes:

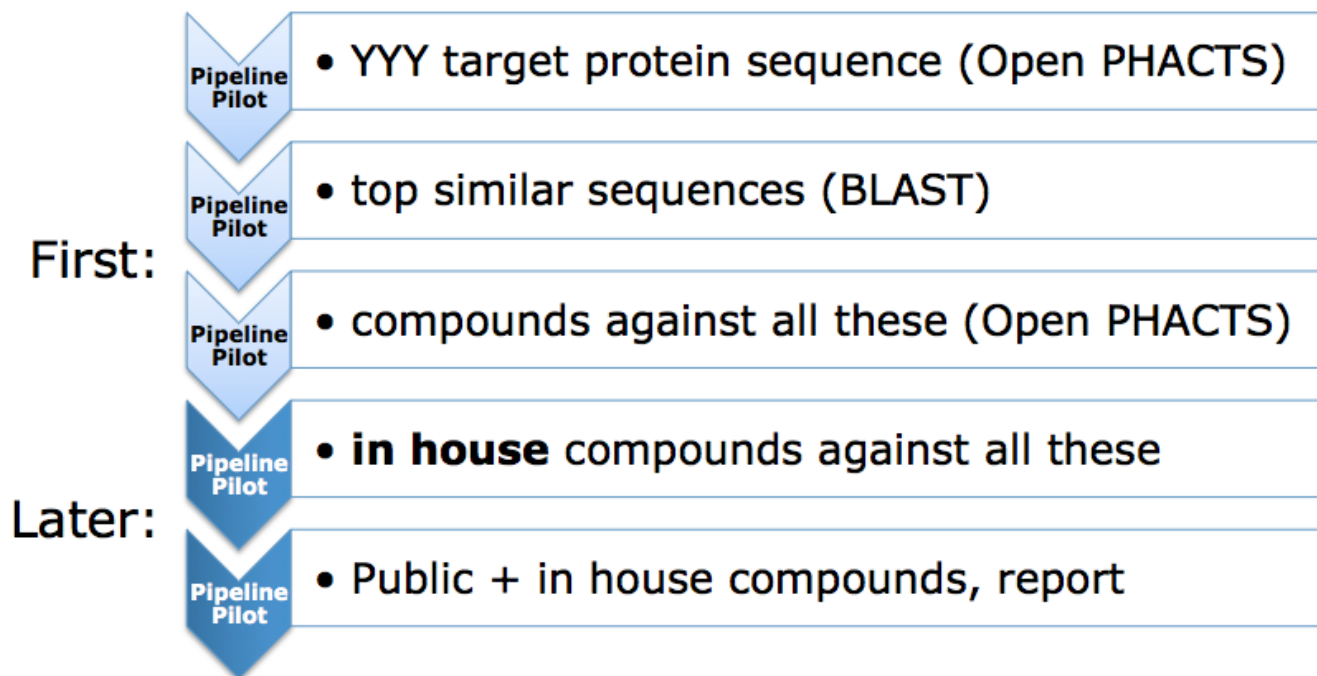
- File Reader**: Simply gets the URL [I dont know how to get it to start otherwise!]
- Java Snippet**: Fetch JSON from web
- Get Name and Inchi**: Name & Inchi Grabber
- Get Activity**: Now turn the activity JSON into rows
- Activity Parser**: For each activity row, extract the columns we want
- Column Filter**: Tidy Up: Remove Processing Columns Now
- Interactive Table**: Node 31





Open PHACTS Use Case: Neuroscience / Oncology

- Which compounds are associated with YYY and related targets to design a focused set?





THE FUTURE



Open PHACTS

Open Pharmacological Space



Pharma
Companies

Life
Science
Companies

Data
Providers

Connecting
Communities

App
Developers

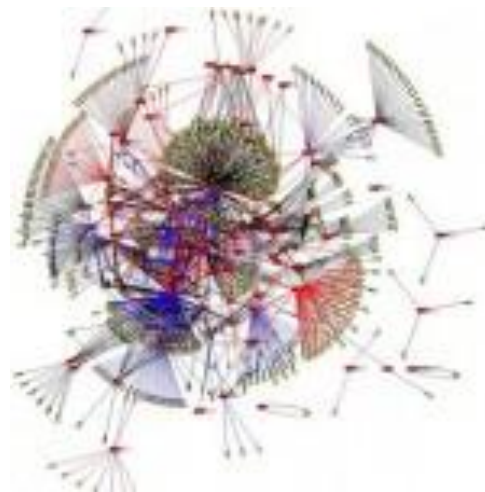
Academic
Research

Next Gen
IT

✦ “Software is free like puppies are free - they both need money for maintenance”



✦ ...and more resource for future development



The Open PHACTS Foundation

OPF is a not-for-profit membership organisation, supporting the Open PHACTS Discovery Platform:

A sustainable, open, vibrant and interoperable information infrastructure for applied life science research and development.

To reduce the barriers to drug discovery in industry, academia and for small businesses, the **Open PHACTS Discovery Platform** provides tools and services to interact with multiple integrated and publicly available data sources. To integrate this data, extensive cross-referencing of scientific concepts is needed across all databases.

The Open PHACTS Foundation ensures the sustainability of the **Open PHACTS Discovery Platform** infrastructure and acts as a hub for relevant scientific research and development.



ChEMBL



DRUGBANK
Open Data Drug & Drug Target Database



WIKIPATHWAYS
Pathways for the People



Key Resources


 [Open PHACTS API](#)

 [Open PHACTS Repository](#)

Subscribe to the Foundation Newsletter

Subscribe

Contact us

 Email:
info@openphactsfoundation.org

 Twitter: [@Open PHACTS](#)

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
 The Open PHACTS Foundation



pmu@openphacts.org



@Open_PHACTS



Open PHACTS

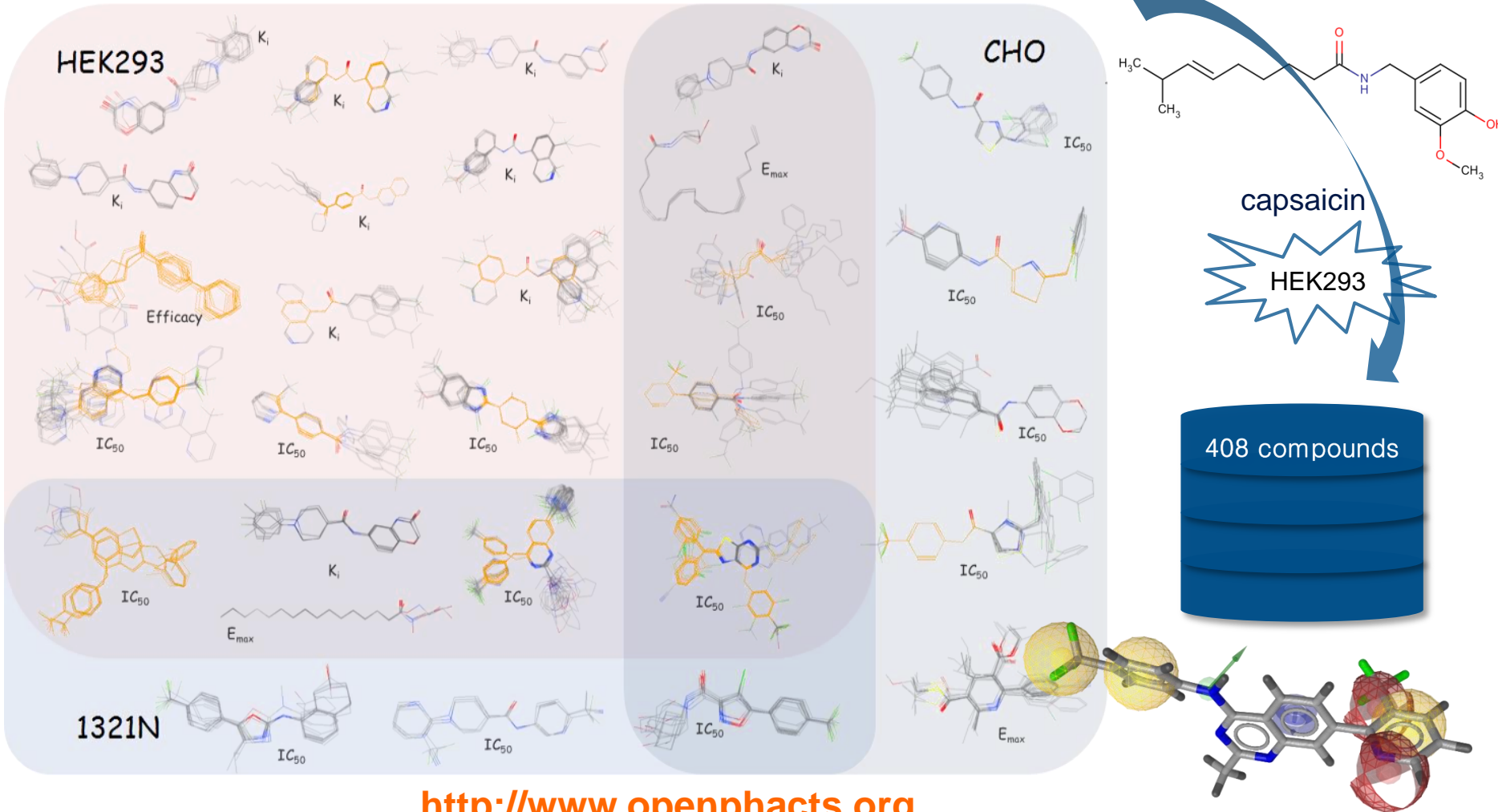
Backup

Present Content

Statistics of Datasets Loaded into Open PHACTS Version 1.3

Source	Version	Supplier	Downloaded	Initial Records	Triples	Properties
ChEMBL	ChEMBL 16 RDF	EBI	25 June 2013	1,247,403 (~1,236,686 compounds, 9844 targets, 6243 target components, 873 protein classes)	304,420,681	77
DrugBank	Aug 2008	Bio2Rdf (www4.wiwiiss.fu-berlin.de)	08 Aug 2012	19,628 (~14,000 drugs, 5000 targets)	517,584	74
SwissProt, UniParc, UniRef	2013_06	SIB	2013_06		533,394,147	82
ENZYME	2013_07	SIB	2013_07	6,187	47,661	2
ChEBI	Release 104	EBI	19 June 2013	40,575	40,575	2
GeneOntology	Jan 21, 2013	GO	21 Jan 2013	38,137	1,265,273	26
GOA	2013	GO	09 Sept 2013	various species	23,489,501	15
WikiPathways	v0. ? 1_20130710	Maastricht	10 July 2013	946	1,449,981	34
ChemSpider		Open PHACTS Chemistry Registry (OCSR)	Nov 11, 2013		tbc	
ConceptWiki	version 1.3	NBIC	09 Sept 2013	2,828,966	3,739,884	1

hTRPV1 → 2328 ligands from Open PHACTS



moe 2012.10

File Edit Select Render Protein Compute Window Help

SVL DBV SEQ Cancel System

Open
LigX
Constrain
Close
Center
SiteView
Hydrogens
Hide
Show
Ligand
Surface
Measure
Builder
Sketch
Minimize
Select
Extend
Delete

Database Viewer: d:/1304 - april 2013/example2.mdb

File Edit Display Compute OpenPHACTS Window Help

SVL DBV MOE Cancel

mol

- Pharmacology By Target
- Pharmacology By Compound
- Pharmacology By Enzyme Family
- Pharmacology By ChEBI Class

Input CHEMBL-target ID

Please input the CHEMBL ID of your target








OK Cancel

0 entries, 1 field, 0 selected, all visible.

Database Viewer: d:/1304 - april 2013/example.mdb

File Edit Display Compute OpenPHACTS Window Help

SVL DBV MOE Cancel

	smiles	http://data.kasabi.com/databe...	pmid	full_mwt	http://www.conceptwiki.org/	prefLabel	prefLabel
1		http://data.kasabi.com/da	9767638	393.4790	http://www.conceptwiki.	Multidrug resistance protein 1 (Homo sapie	Multidrug
2		http://data.kasabi.com/da	9767638	257.3310	http://www.conceptwiki.	Multidrug resistance protein 1 (Homo sapie	Multidrug
3		http://data.kasabi.com/da	9767638	373.4890	http://www.conceptwiki.	Multidrug resistance protein 1 (Homo sapie	Multidrug
4		http://data.kasabi.com/da	9767638	399.5070	http://www.conceptwiki.	Multidrug resistance protein 1 (Homo sapie	Multidrug
5		http://data.kasabi.com/da	9767638	421.5320	http://www.conceptwiki.	Multidrug resistance protein 1 (Homo sapie	Multidrug
6		http://data.kasabi.com/da	9767638	393.4790	http://www.conceptwiki.	Multidrug resistance protein 1 (Homo sapie	Multidrug
7		http://data.kasabi.com/da	9767638	421.5320	http://www.conceptwiki.	Multidrug resistance protein 1 (Homo sapie	Multidrug

2563 entries, 0 selected, all visible. 19 fields, 0 selected, all visible.

MMFF94x PBC

Atoms Ribbon Contacts Fog ZClip