

The Open PHACTS Project: Progress and Future Sustainability

Lee Harland & Bryn Williams-Jones
Open PHACTS / ConnectedDiscovery



Tom Plasterer
AstraZeneca/Open PHACTS Rep





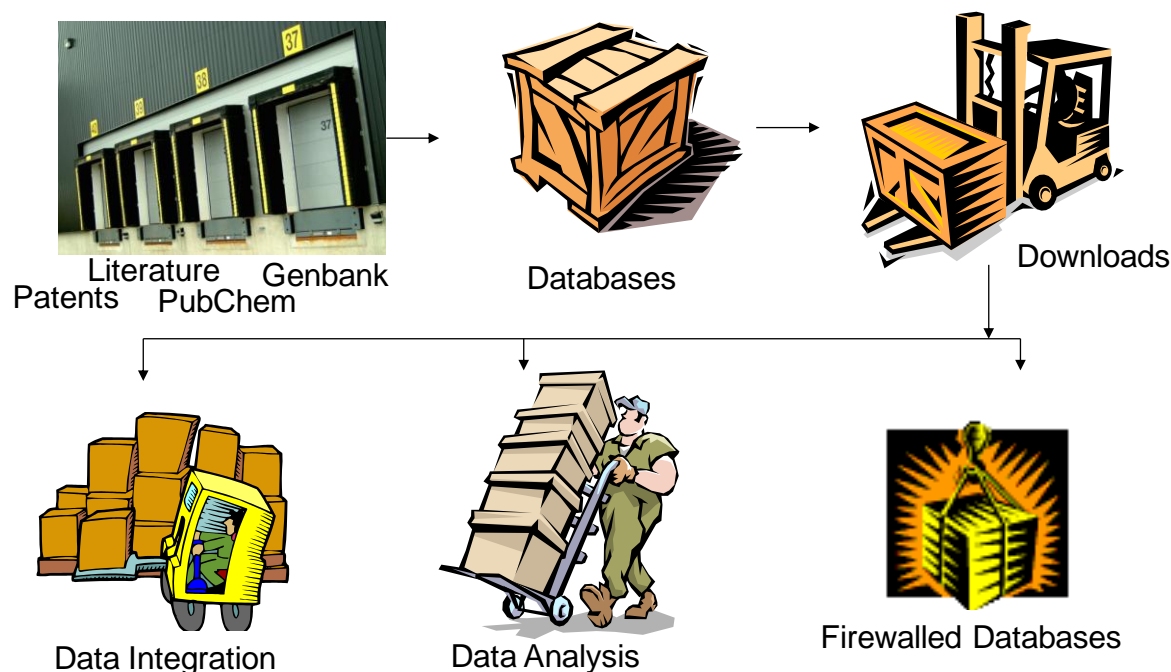
Fundamental issue:

- ✦ There is a *lot* of science outside your walls
- ✦ It's a chaotic space
- ✦ Scientists want to find information quickly and easily
- ✦ Often they just “cant get there” (or don't even know where “there” is)
- ✦ And you have to manage it all (or not)



Pre-competitive Informatics:

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



**X Repeat @
each
company**

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



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
- 23 academic partners, 8 pharmaceutical companies, 3 biotechs
- Work split into clusters:
 - Technical Build
 - Scientific Drive
 - Community & Sustainability

The Project



"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 <1 μ M"



ChEMBL

DrugBank

Gene
Ontology

Wikipathways

GeneGo

ChEBI

UniProt

UMLS

GVKBio

ConceptWiki

ChemSpider

TrialTrove

TR Integrity



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 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



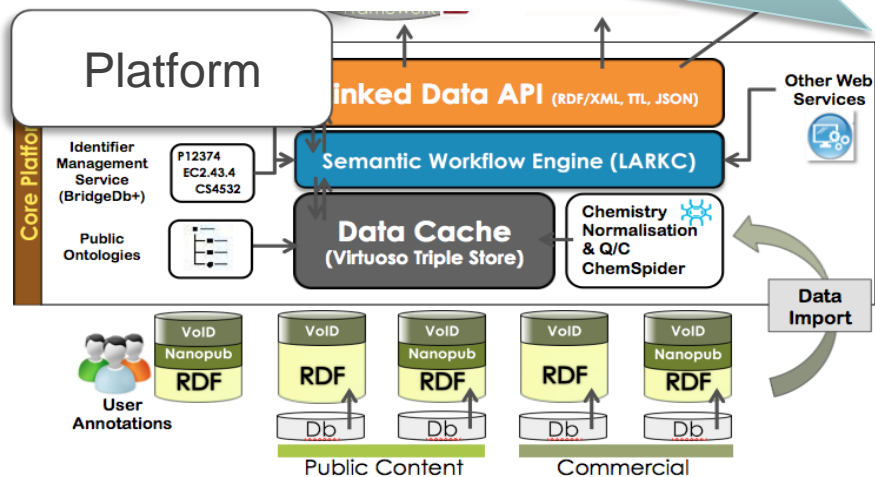
Review

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 Blomberg¹², a, Lee Harland¹³, Bryn Williams-Jones¹³, Gerhard F. Ecker⁴,  



“Provenance Everywhere”



Explorer

Sildenafil

Pharmacology Data | View in ChemBioNavigator

Sildenafil (in citrate form), sold under the names Viagra, Revatio and under various other names, is a drug used to treat male erectile dysfunction (impotence) and pulmonary arterial hypertension (PAH), developed by the pharmaceutical company Pfizer. Its primary competitors on the market are tadalafil (Cialis), and vardenafil (Levitra). [Wikipedia]

Hepatic:

ChemSpider ID: [5023](#)

Molecular Formula: $C_{22}H_{26}N_4O_6S$

SMILES: O=S(=O)(N1CCN(C)CC1)C4=CC(=C)C(=O)C3=CC(=O)C(C)C(C)C4

Standard InChI: O=S(=O)(N1CCN(C)CC1)C4=CC(=C)C(=O)C3=CC(=O)C(C)C(C)C4

Standard InChIKey: BNNRUJZRGQAQ-UHFFFAOYSA-N

Affected Organism: Humans and other mammals

Indication: For the treatment of erectile dysfunction

Melting Point: 189-190 °C

ALogP: 2.2

H-Bond Receptors: 7

H-Bond Donors: 1

Mol Weight: 474.576

MW Freebase: 474.576

Polar Surface Area: 117.51

Rotatable Bonds: 7

Apps

API

?ops_item skos:ec

?ops_item skos:ec

?cw_uri skos:pre

void:ind

?equiv_target do

ops:target

ops:target

void:ind

ops:targetOfAssa

?equiv_assay che

chembl:ch

?std_type ;



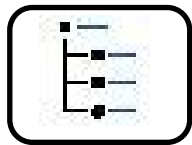


Open PHACTS

Open Pharmacological Space



THE OPEN PHACTS DISCOVERY PLATFORM





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 (OCCRS)	Nov 11, 2013		tbc	
ConceptWiki	version 1.3	NBIC	09 Sept 2013	2,828,966	3,739,884	1



Data Licensing Solution

Chose John Wilbanks as consultant

A framework built around STANDARD well-understood Creative Commons licences – and how they interoperate

Deal with the problems by:



- ✦ Interoperable licences
- ✦ Appropriate terms
- ✦ Declare expectations to users and data publishers
- ✦ One size won't fit all requirements

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							



Its easy to integrate, difficult to integrate well:

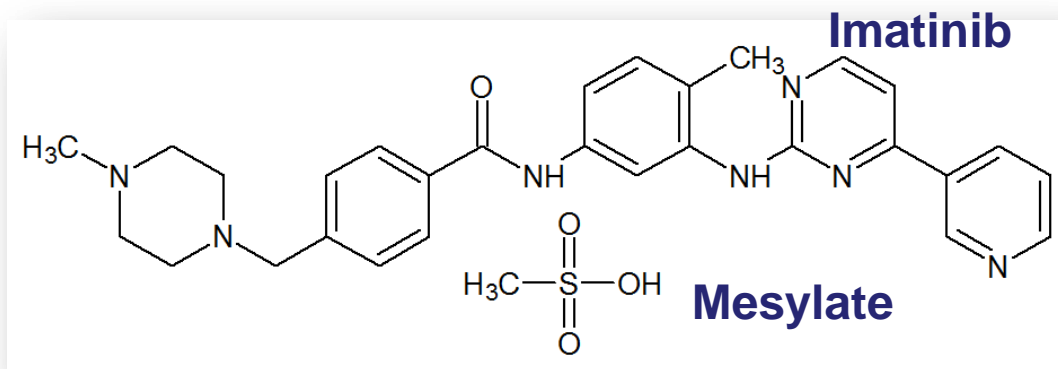
Type a compound name:

glee

- Gleevec
- Gleevec



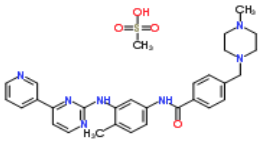
What Is Gleevec?



ChemSpider
The free chemical database

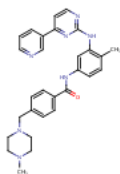
About | More Searches | Web APIs

Gleevec



ChemSpider ID: 123596
Molecular Formula: C₂₉H₃₁N₇O₄S
Average mass: 589.708400
Monoisotopic mass: 589.708400
Systematic name: 4-[[4-methyl-1-piperazinyl]methoxy]phenyl 4-methylpiperazine-1-carboxylate

ChemSpider

Structure	 Download: MOL SDF SMILES InChI Display: 2D Structure 3D Structure
Synonyms	<ul style="list-style-type: none"> Imatinib Mesylate Imatinib Methansulfonate 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: C₂₉H₃₁N₇O₄S
IUPAC name: 4-[[4-methylpiperazin-1-yl]methoxy]phenyl 4-methylpiperazine-1-carboxylate
Active in 205 BioAssays Tested in 1376 BioAssays
CID: 5291
[Similar Compounds](#) [Same Parent, Connectives](#) (MeSH Keyword)

Imatinib mesylate; Gleevec; Glivec ...

MW: 589.708400 g/mol MF: C₃₀H₃₅N₇O₄S
IUPAC name: methanesulfonic acid; 4-[[4-methylpiperazin-1-yl]methoxy]phenyl 4-methylpiperazine-1-carboxylate
Active in 35 BioAssays Tested in 679 BioAssays
CID: 123596
[Similar Compounds](#) [Same Parent, Connectives](#) (MeSH Keyword)

PubChem



Dynamic Equality



chemspider:gleevec

drugbank:gleevec

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




Play!

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

OpenPHACTS API

Chemical Structure Exact Search

/structure/exact [GET](#)

InchiKey to URL

/structure [GET](#)

Inchi to URL

/structure [GET](#)

Chemical Structure Similarity Search

/structure/similarity [GET](#)

SMILES to URL

/structure [GET](#)

Chemical Structure Substructure Search

/structure/substructure [GET](#)

Get concept description

/getConceptDescription [GET](#)

Map free text to a concept URL based on semantic tag

/search/byTag [GET](#)

Map URL

/mapURL [GET](#)

Map free text to a concept URL

/search/freetext [GET](#)

Get ChEBI Ontology Class Members

/compound/chebi/members [GET](#)

Get ChEBI Ontology Root Classes

/compound/chebi/root [GET](#)

Get ChEBI Ontology Class

/compound/chebi/node [GET](#)

ChEBI Class Pharmacology Count


/compound/chebi/pharmacology/count [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. Specefy 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.



APPS




Open PHACTS Explorer

Navigation

- Compound
- Target
- Pharmacology

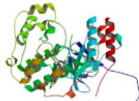
Target by name

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

Target name:

Provenance: ☐ On ☒ Off

Target Data



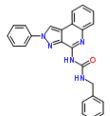
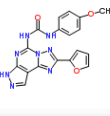
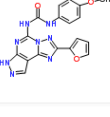
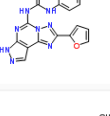

Mitogen-activated protein kinase 14 (Homo sapiens)

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

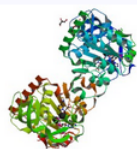
Protein name:

☒ Filter
 Provenance: ☐ On ☒ Off

Pharmacology by Target name search results - Total Records: 7887

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(NCc1cccc...	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...	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 uM	Inhibition	=	98	%	390.356	COc1ccc(cc1)N...	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...	InChI=1S/C
												

<http://explorer.openphacts.org>



Quinone oxidoreductase (Homo sapiens) (4 of 4 results loaded)

[Filter Results](#)

[Create TSV](#)

Compound	Target	Assay	Activity								
Name	Organism	Organism	Description	Type	Relation	Value	Units	Mol Weight	SMILES	InChi	
2,5-bis((6-[ethyl(2-methoxybenzyl)amino]hexyl)amino)cyclohexa-2,5-diene-1,4-dione	<		Binding affinity to NQO1	Km	=	12700	nM	632.876	<chem>CCN(CCCCCCNC1=CC(=O)C(=CC1=O)NC(CCCCCN(CC)CC2=CC=CC=C2OC)CC3=CC=CC=C3OC</chem>	InChi=1S/C38H56N4O4/c1-5-41(29-31-19-11-13-21-37(31)45-3)25-17-9-7-15-23-39-33-27-36(44)34(28-35(33)43)40-24-16-8-10-18-26-42(6-2)30-32-20-12-14-22-38(32)46-4/h11-14,19-22,27-28,39-40H,5-10,15-18,23-26,29-30H2,1-4H3	
2'-Hydroxy-4-bromochalcone	<	Homo sapiens	Induction of quinone reductase activity in human MCF7 cells	Activity				303.151	<chem>C1=CC=C(C(=C1)C(=O)C=C/C2=CC=C(C=C2)Br)O</chem>	InChi=1S/C15H11BrO2/c16-12-8-5-11(6-9-12)7-10-15(18)13-3-1-2-4-14(13)17/h1-10,17H/b10-7+	
2'-hydroxychalcone	<	Homo sapiens	Induction of quinone reductase activity in human MCF7 cells	Activity				224.255	<chem>C1=CC=C(C(=C1)C=C/C(=O)C2=CC=CC=C2O</chem>	InChi=1S/C15H12O2/c16-14-9-5-4-8-13(14)15(17)11-10-12-6-2-1-3-7-12/h1-11,16H/b11-10+	
(2E)-1-(2-hydroxyphenyl)-3-(pyridin-2-yl)prop-2-en-1-one	<	Homo sapiens	Induction of quinone reductase activity in human MCF7 cells	Activity				225.243	<chem>C1=CC=C(C(=C1)C(=O)C=C/C2=CC=CC=N2)O</chem>	InChi=1S/C14H11NO2/c16-13-7-2-1-6-9-8-11-5-3-4-10-15-11/h1-10,12-14,16H/b11-10,13-14	Go to top



Example applications

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



KNIME

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

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

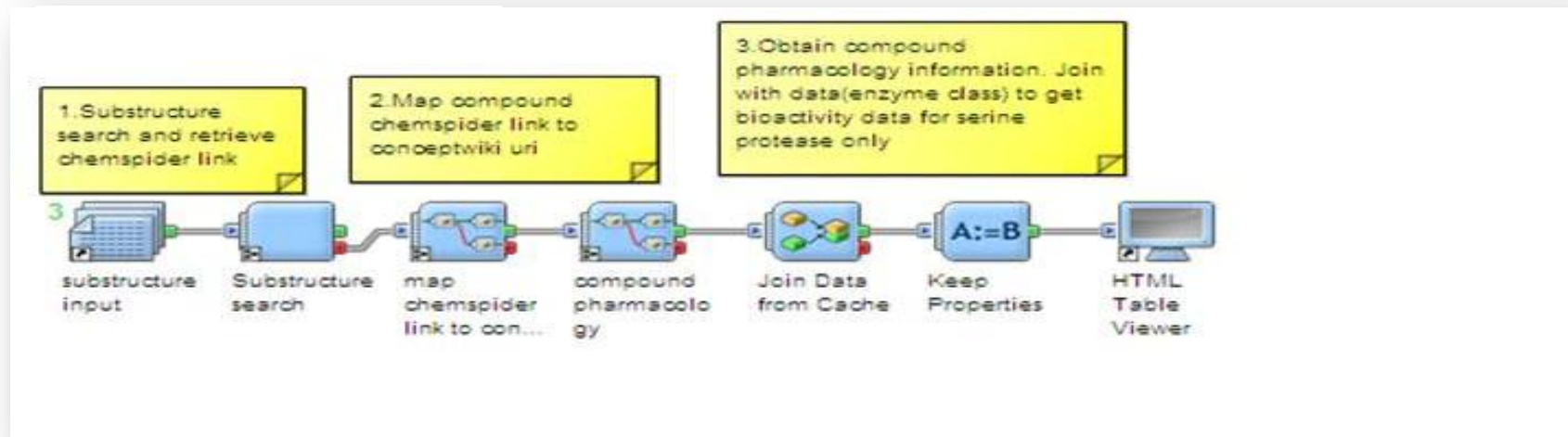
Workflow Diagram:

```

graph LR
    A[File Reader] --> B[Java Snippet]
    B --> C[Get Name and Inchi]
    C --> D[Get Activity]
    D --> E[Activity Parser]
    E --> F[Column Filter]
    F --> G[Interactive Table]
  
```

Node Descriptions:

- 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





Uptake at AstraZeneca: a Use Case

Applying BioAssay Ontology to facilitate HTS analysis

Linda Zander Balderud

Ola Engkvist

Chemistry Innovation Centre, Discovery Sciences

AstraZeneca

Assay Informatics project

Benefits in Adopting BioAssay Ontology (BAO)

- Common language for assay annotation
- Improved project success analyses based on assay technologies
- Better understand the impact of technology artifacts like frequent hitters
- Assay design and screening cascade support during assay development in early projects
- Improved capability to perform combined data mining of internal and public data



FLIPR Tetra High Throughput Cellular Screening System
(from Molecular Devices)



The BioAssay Ontology (BAO)



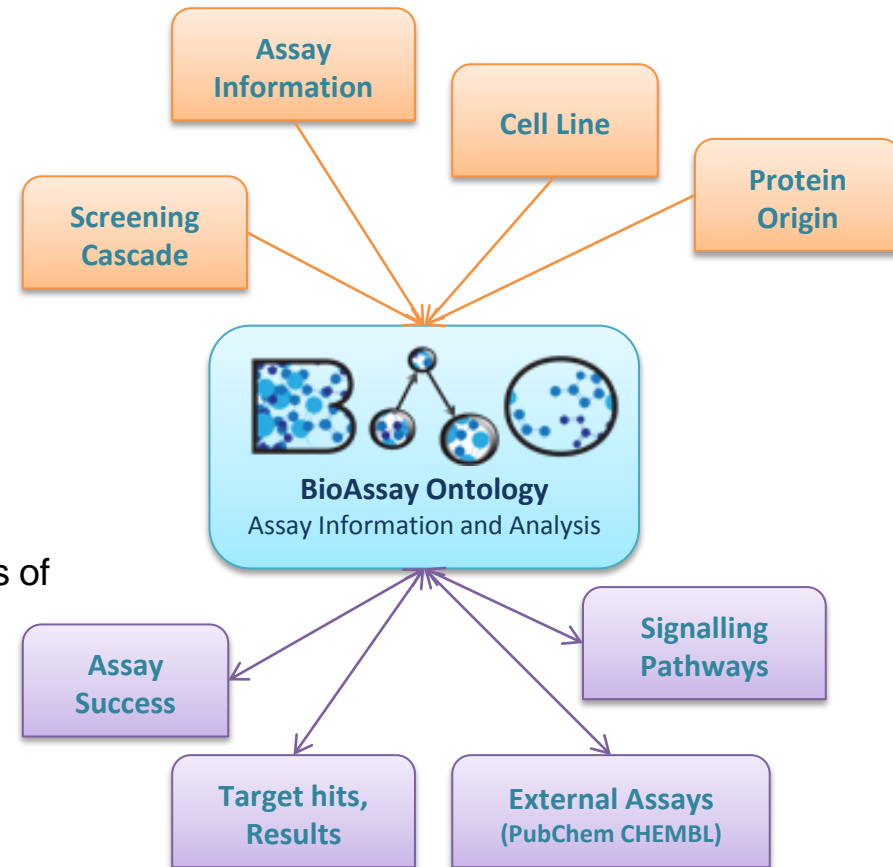
Computational Science, University of Miami, USA

Domain:

- Assay design
- Assay format
- Detection technology
- Meta target
- Endpoint
- Perturbagen

BioAssay Ontology imports:

- **NCBI taxonomy** - organism names and IDs
- **Uniprot** - protein target names and IDs
- **Unit Ontology** - concentration and time unit terms
- **Ontology of Biomedical Investigation** – descriptions of biological assays
- **Gene Ontology** - biological processes
- **Cell Line Ontology** - cell line names
- **CL** – cell types
- **UBERON** – anatomical entities
- **PATO** – cell phenotype
- **SAR connect** – target classifications



Migration to BAO

Annotation of HTS assays

Manual annotation of protocols

HTS assay: reporter gene assay

- Assay method: reporter gene method: beta lactamase induction
- Detection technology: FRET
- Bioassay: beta lactamase assay
- Assay kit: *LiveBLAzer FRET - B/G Loading Kit*
- Wavelength: ex 405 em 460, 535
- Biological process
- Disease

HTS assay: FLIPR

- Assay method: molecular redistribution determination assay
- Detection technology: fluorescence intensity
- Bioassay: calcium redistribution assay
- Assay kit: Fluo-8 No Wash Calcium Assay Kit
- Wavelength: ex 480 em 530
- Biological process
- Disease



Over 900 PubChem assays have been annotated by the BioAssay Ontology team



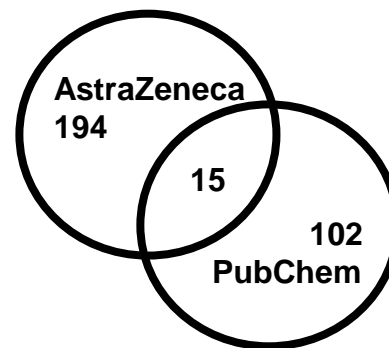
Assay Development Support

Comparison study between AstraZeneca and PubChem HTS assays

412 in-house HTS assays since 2005 have been annotated according to the BioAssay Ontology. The assay design and technology of the annotated assays were analyzed together with 239 primary assays from PubChem. The analyzed PubChem assays are biochemical assays, assays detected by luminescence and/or assays using GPCR targets.

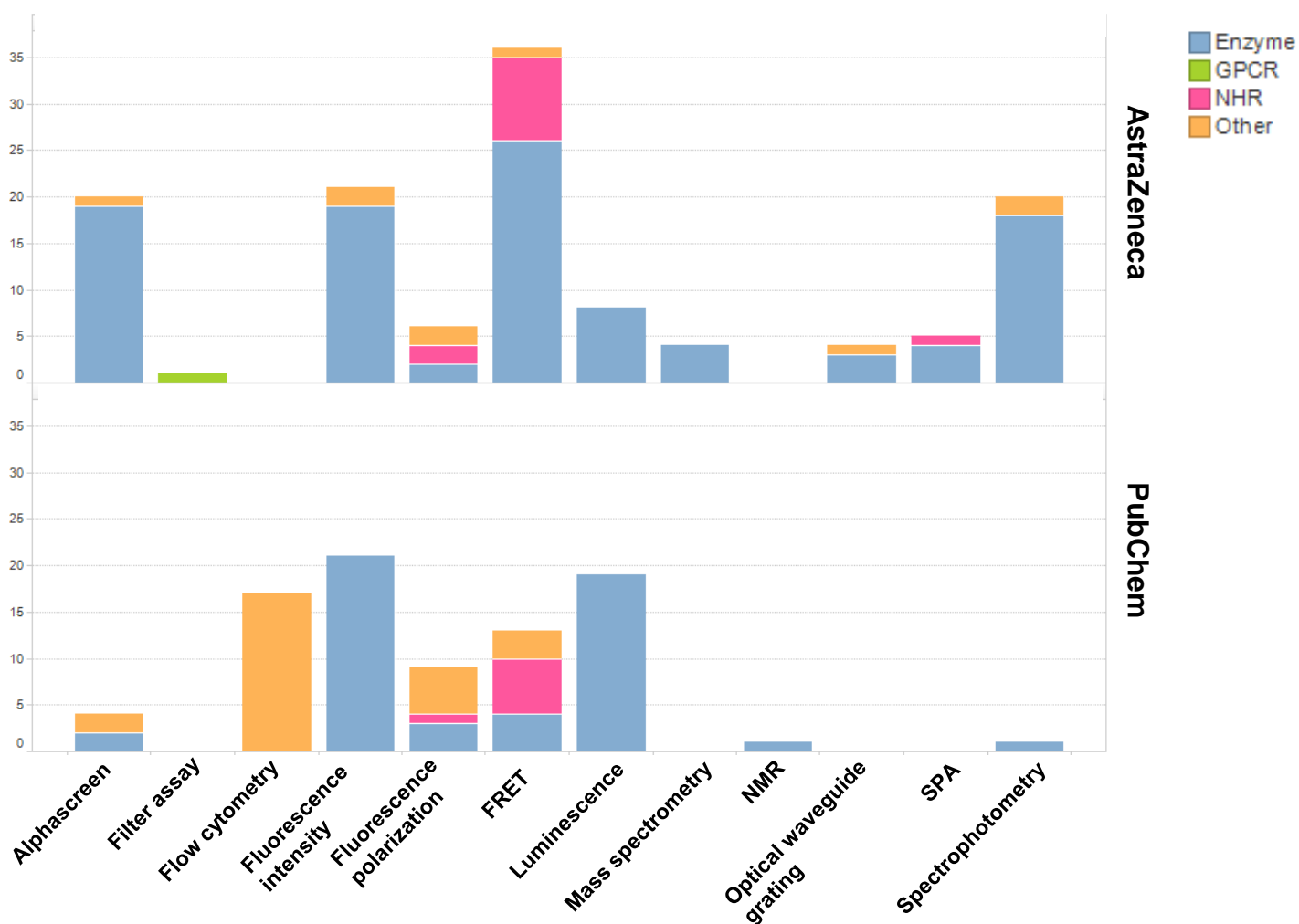
From the annotated assays, 515 assays were using human targets and combined 311 different human targets were represented in the study.

15 of the in-house targets were also screened in at least one PubChem assay. Eight of these were GPCR targets.



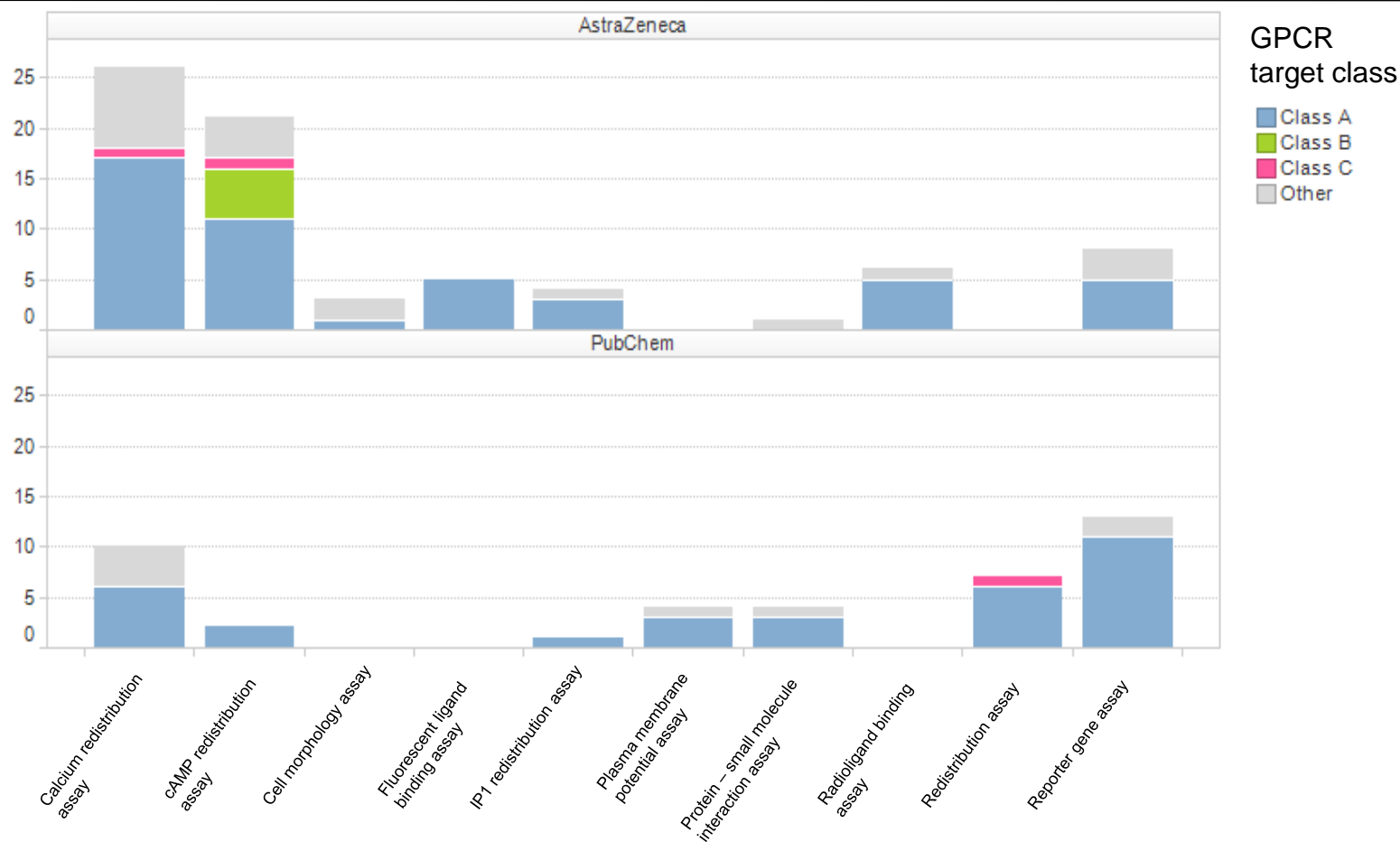
Assay Development Support

Detection Technology of AZ and PubChem Biochemical Assays



Assay Development Support

Assay design of in-house and PubChem GPCR HTS



One explanation for the low usage of cAMP redistribution method among the annotated PubChem assays could be that no class B GPCRs have been screened



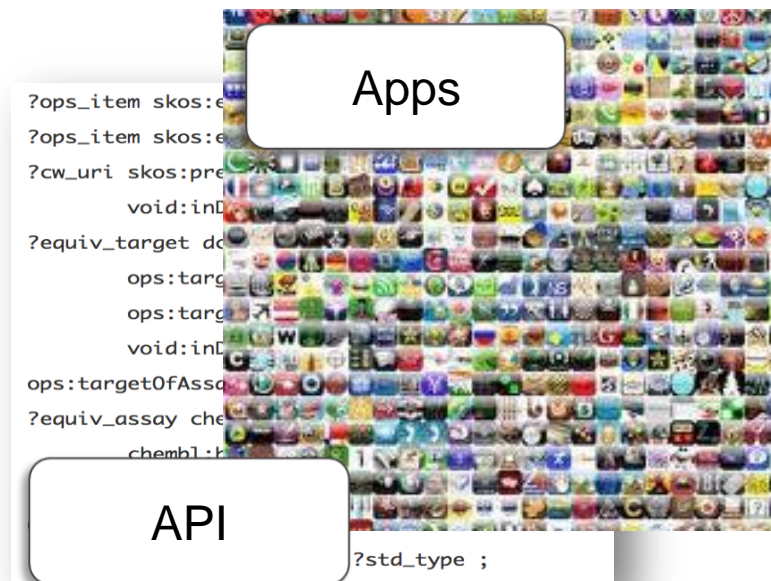
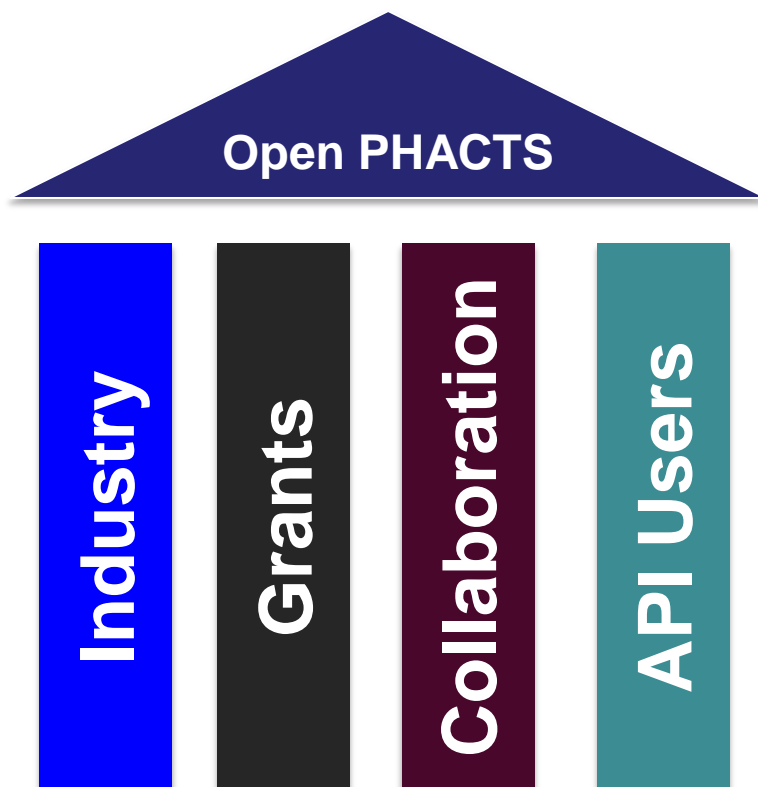


Sustaining The Project

The Open PHACTS Foundation



Kick-Starting Sustainability



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

ChemSpider
The free chemical database**DRUGBANK**
Open Data Drug & Drug Target DatabaseWikiPATHWAYS
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



pmu@openphacts.org



@Open_PHACTS



Open PHACTS