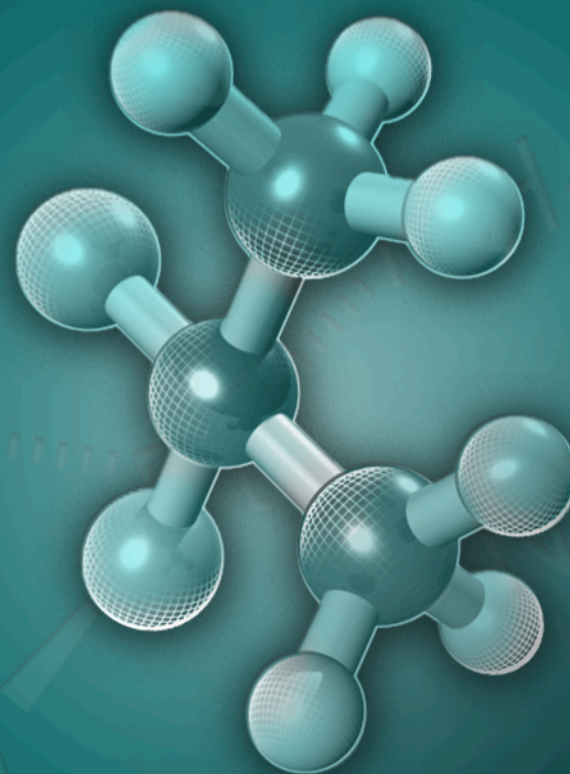


Leveraging Open Chemogenomics Data and Tools with KNIME

George Papadatos

ChEMBL Group

georgep@ebi.ac.uk



What is EMBL-EBI?

- Europe's home for biological **data**, **services**, **research** and **training**
- A trusted data provider for the life sciences
- Part of the European Molecular Biology Laboratory, an intergovernmental research organisation
- International: 570 members of staff from 57 nations

Data resources at EMBL-EBI

Genes, genomes & variation

European Nucleotide Archive
European Variation Archive

Ensembl
Ensembl Genomes

GWAS Catalog
Metagenomics portal

Gene, protein & metabolite expression

RNA Central

ArrayExpress
Expression Atlas

Metabolights
PRIDE

Protein sequences, families & motifs

InterPro

Pfam

UniProt

Molecular structures

Protein Data Bank in Europe
Electron Microscopy Data Bank

Chemical biology

ChEMBL

SureChEMBL

ChEBI

Systems

BioModels

Enzyme Portal

BioSamples

Literature & ontologies

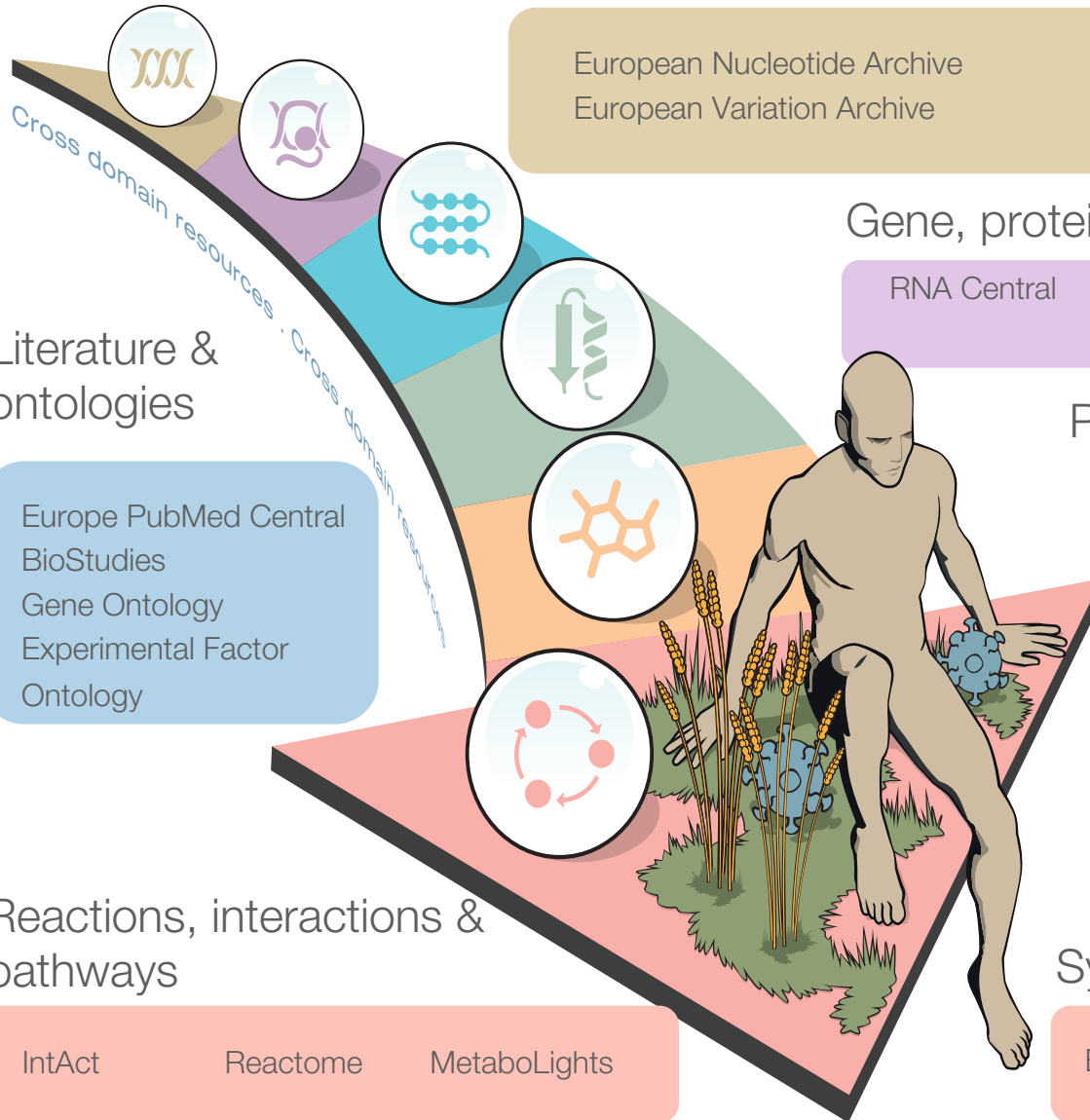
Europe PubMed Central
BioStudies
Gene Ontology
Experimental Factor
Ontology

Reactions, interactions & pathways

IntAct

Reactome

MetaboLights



ChEMBL: Data for drug discovery

1. Scientific facts

3. Insight, tools and resources for translational drug discovery

3822
J. Med. Chem. 2003, 46, 3822-3833

Design of Selective Thrombin Inhibitors Based on the (R)-Phe-Pro-Arg Sequence

John C. Chubbuck,¹ Stuart M. Allen,¹ Alan D. Brown,¹ Paul V. Fish,*¹ Edward Hartsenker,² Stephen J. Edrington,² Keith Turner,² Andrew S. McElroy,¹ John Greenough,¹ Michael P. Pringle,¹ and David J. Kane¹

¹Department of Biomolecular Chemistry, Drug Metabolism, Toxicology, Safety, and Molecular Behaviour, Structure and Design, Pfizer Global Research and Development, Sandwich, Kent CT13 9NJ, United Kingdom

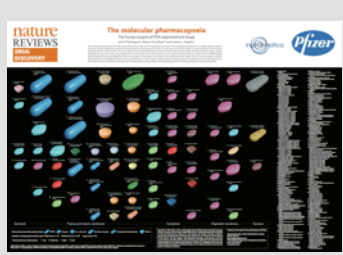
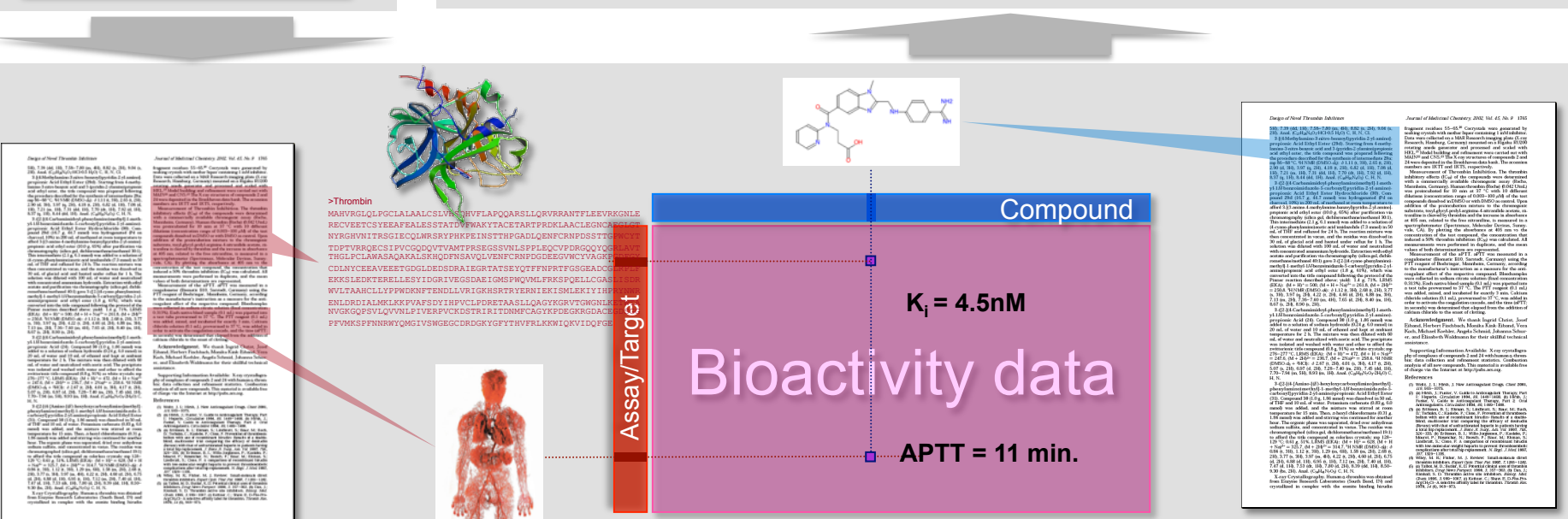
Received December 23, 2002

Rationale and objectives: Thrombin was a central target for the development of novel oral anticoagulant drugs. The objective was to generate a novel class of thrombin inhibitors that were selective for the prothrombinolytic function, as defined by fibrinolytic and heparinase and type III heparinase activity, but that had no activity for the other functions of the enzyme. This was achieved by the design of a novel class of thrombin inhibitors based on the (R)-Phe-Pro-Arg sequence. This sequence is found in the active site of the enzyme, and is thought to be involved in the recognition of the substrate. The design of the inhibitors was based on the (R)-Phe-Pro-Arg sequence, and was achieved by the design of a novel class of thrombin inhibitors based on this sequence.

Methods: The design of the inhibitors was based on the (R)-Phe-Pro-Arg sequence, and was achieved by the design of a novel class of thrombin inhibitors based on this sequence. The inhibitors were synthesized and their activity was determined by a number of assays, including a fibrinolytic assay, a heparinase assay, and a type III heparinase assay. The results of these assays are discussed here.

Results: A series of novel thrombin inhibitors were designed and synthesized. The inhibitors were found to be selective for the prothrombinolytic function, and were found to have no activity for the other functions of the enzyme. The inhibitors were found to be orally active, and were found to be effective in animal models of thrombosis.

Conclusions: The design of the inhibitors was based on the (R)-Phe-Pro-Arg sequence, and was achieved by the design of a novel class of thrombin inhibitors based on this sequence. The inhibitors were found to be selective for the prothrombinolytic function, and were found to have no activity for the other functions of the enzyme. The inhibitors were found to be orally active, and were found to be effective in animal models of thrombosis.

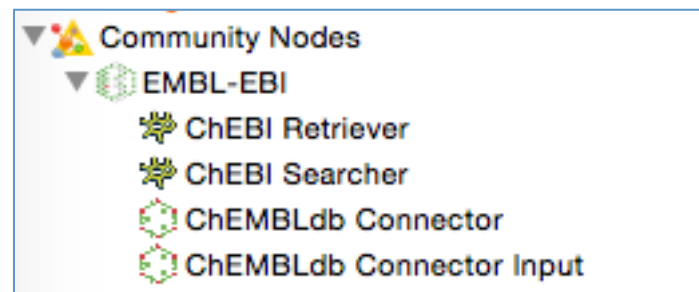



2. Organization, insertion, curation and standardization of pharmacology data

KNIME at the EBI



- KNIME nodes to access ChEBI and ChEMBL databases
 - Trusted community nodes
 - Workflows on Examples server
- Method development and use cases
- Provide KNIME training to scientists and researchers
 - Wellcome Trust drug discovery courses, EMBL courses
- CDK community nodes support



<https://tech.knime.org/book/embl-ebi-nodes-for-knime-trusted-extension>

KNIME and ChEMBL

```

- compounds: {
  chemblid: "CHEMBL941",
  numRo5Violations: 0,
  molecularWeight: 493.60273,
  preferredCompoundName: "IMATINIB",
  alogp: 3.583,
  knownDrug: "Yes",
  medChemFriendly: "Yes",
  rotatableBonds: 7,
  passesRuleOfThree: "No",
  molecularFormula: "C29 H31 N7 O",
  smiles: "CN1CCN(Cc2ccc(cc2)C(=O)Nc3ccc(C)c(Nc4ncnc(n4)c5ccccc53)C4=O)"
}
- <compound>
  <chemblid>CHEMBL941</chemblid>
  <preferredCompoundName>IMATINIB</preferredCompoundName>
  <knownDrug>Yes</knownDrug>
  <medChemFriendly>Yes</medChemFriendly>
  <passesRuleOfThree>No</passesRuleOfThree>
  <molecularFormula>C29 H31 N7 O</molecularFormula>
  <omiles>
    CN1CCN(Cc2ccc(cc2)C(=O)Nc3ccc(C)c(Nc4ncnc(n4)c5ccccc53)C4=O)
  </omiles>
  <species>NEUTRAL</species>
  <numRo5Violations>0</numRo5Violations>
  <rotatableBonds>7</rotatableBonds>
  <molecularWeight>493.60273</molecularWeight>
  <okname>3.583</okname>
  
```

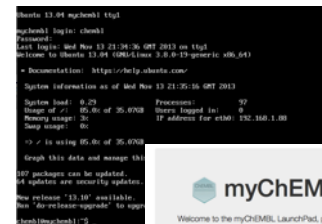
ChEMBL Web Services

14M bioactivities
1.5M structures

ChEMBLdb Connector



Virtual Machine



myChEMBL LaunchPad

Welcome to the myChEMBL LaunchPad, providing access to all resources distributed with the myChEMBL virtual machine.

- ChEMBL Web Services**: Access to a local version of the official ChEMBL Web Services, which connects to the myChEMBL PostgreSQL database (including RDM service cartage labels and tar SQL queries).
- myChEMBL Web Interface**: The myChEMBL Web Interface is a user-friendly application, providing quick access to the myChEMBL data without any prior knowledge of SQL or RDM.
- KNIME Integration**: Learn how to connect the official connectors to myChEMBL, and also how to start processing ChEMBL data with a workflow environment.
- More Information**: For more details on the myChEMBL project, including background, acknowledgments and references.

Local access to ChEMBL data and services



GET Resource



UniChem Web Services
Access ~110M structures from 27 sources

KNIME and ChEMBL

```

- compounds: {
  chemblid: "CHEMBL941",
  numRo5Violations: 0,
  molecularWeight: 493.60273,
  preferredCompoundName: "IMATINIB",
  alogp: 3.583,
  knownDrug: "Yes",
  medChemFriendly: "Yes",
  rotatableBonds: 7,
  passesRuleOfThree: "No",
  molecularFormula: "C29 H31 N7 O",
  smiles: "CN1CCN(Cc2ccc(cc2)C(=O)Nc3ccc(C)c(Nc4ncnc(n4)c5ccccc5)3)3"
}
- <compound>
  <chemblid>CHEMBL941</chemblid>
  <preferredCompoundName>IMATINIB</preferredCompoundName>
  <knownDrug>Yes</knownDrug>
  <medChemFriendly>Yes</medChemFriendly>
  <passesRuleOfThree>No</passesRuleOfThree>
  <molecularFormula>C29 H31 N7 O</molecularFormula>
  <omiles>
    CN1CCN(Cc2ccc(cc2)C(=O)Nc3ccc(C)c(Nc4ncnc(n4)c5ccccc5)3)3
  </omiles>
  <species>NEUTRAL</species>
  <numRo5Violations>0</numRo5Violations>
  <rotatableBonds>7</rotatableBonds>
  <molecularWeight>493.60273</molecularWeight>
  <okaneses>3.583</okaneses>
  
```

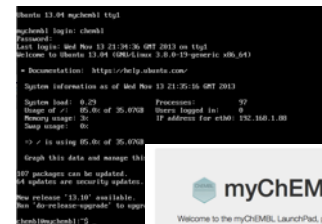
ChEMBL Web Services

14M bioactivities
1.5M structures

ChEMBLdb Connector



Virtual Machine



myChEMBL LaunchPad

Welcome to the myChEMBL LaunchPad, providing access to all resources distributed with the myChEMBL virtual machine.

- ChEMBL Web Services**: Access to a local version of the official ChEMBL Web Services, which connects to the myChEMBL PostgreSQL database (including RDM service cache tables) and our SQL gateway.
- myChEMBL Web Interface**: The myChEMBL Web Interface is a user-friendly application, providing quick access to the myChEMBL data without any prior knowledge of SQL or RDM.
- KNIME Integration**: Learn how to connect KNIME workflows to myChEMBL, and also how to start processing ChEMBL data with a workflow environment.
- Python Notebooks**: A selection of programming tutorials written in Python and processed using interactive Python Notebooks, some examples:
 - myChEMBL introduction
 - Using myChEMBL Web Services
 - Building predictive target models
- More Information**: For more details on the myChEMBL project, including background, acknowledgments and references.

myChEMBL 2014 | ChEMBL Release 18 ChEMBL Release | myChEMBL Support

Local access to ChEMBL data and services

Patent Annotations
4M patent documents
14M structures
260M annotations



OPS_Swagger



GET Resource

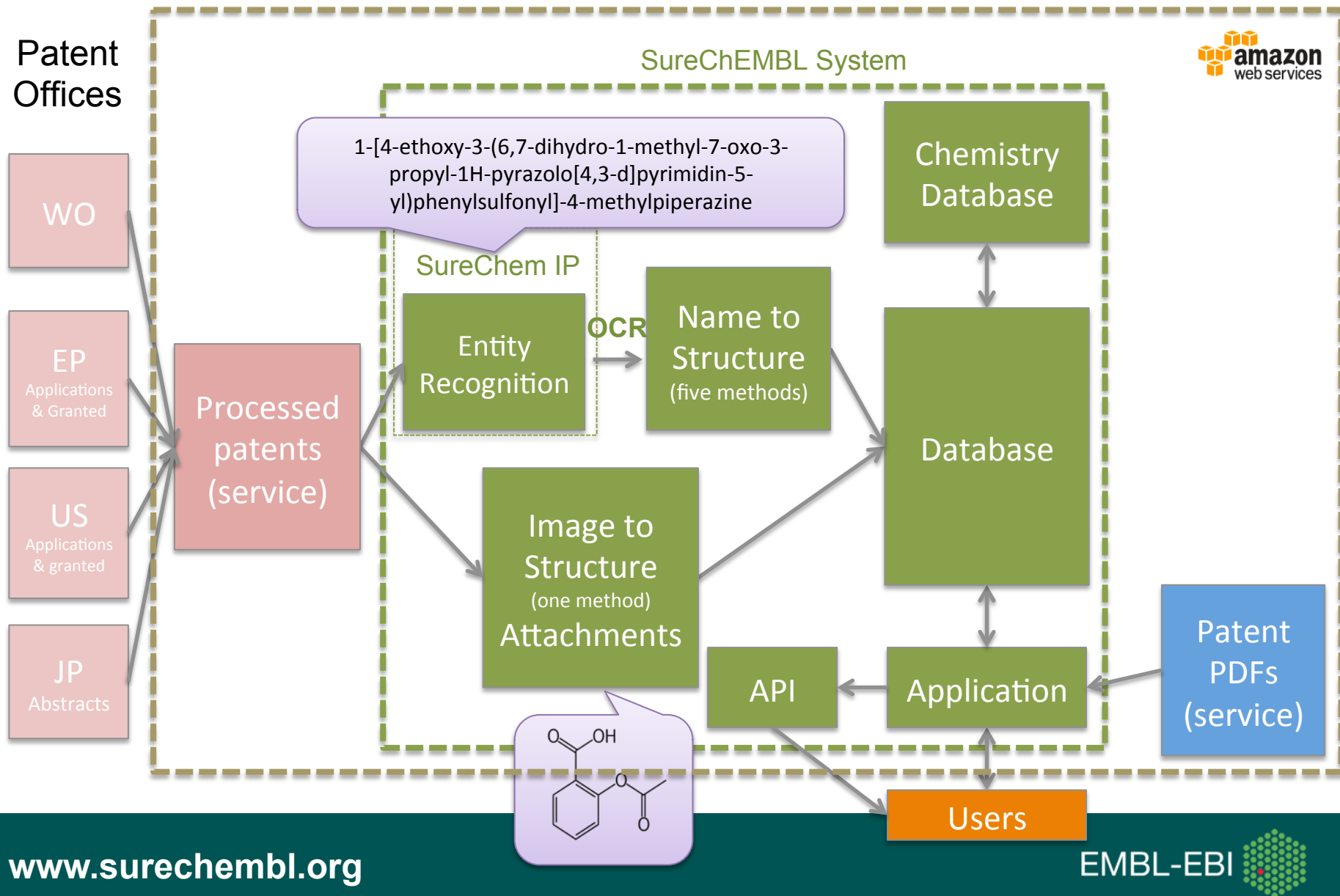


UniChem Web Services
Access ~110M structures from 27 sources

Why looking at patent documents?

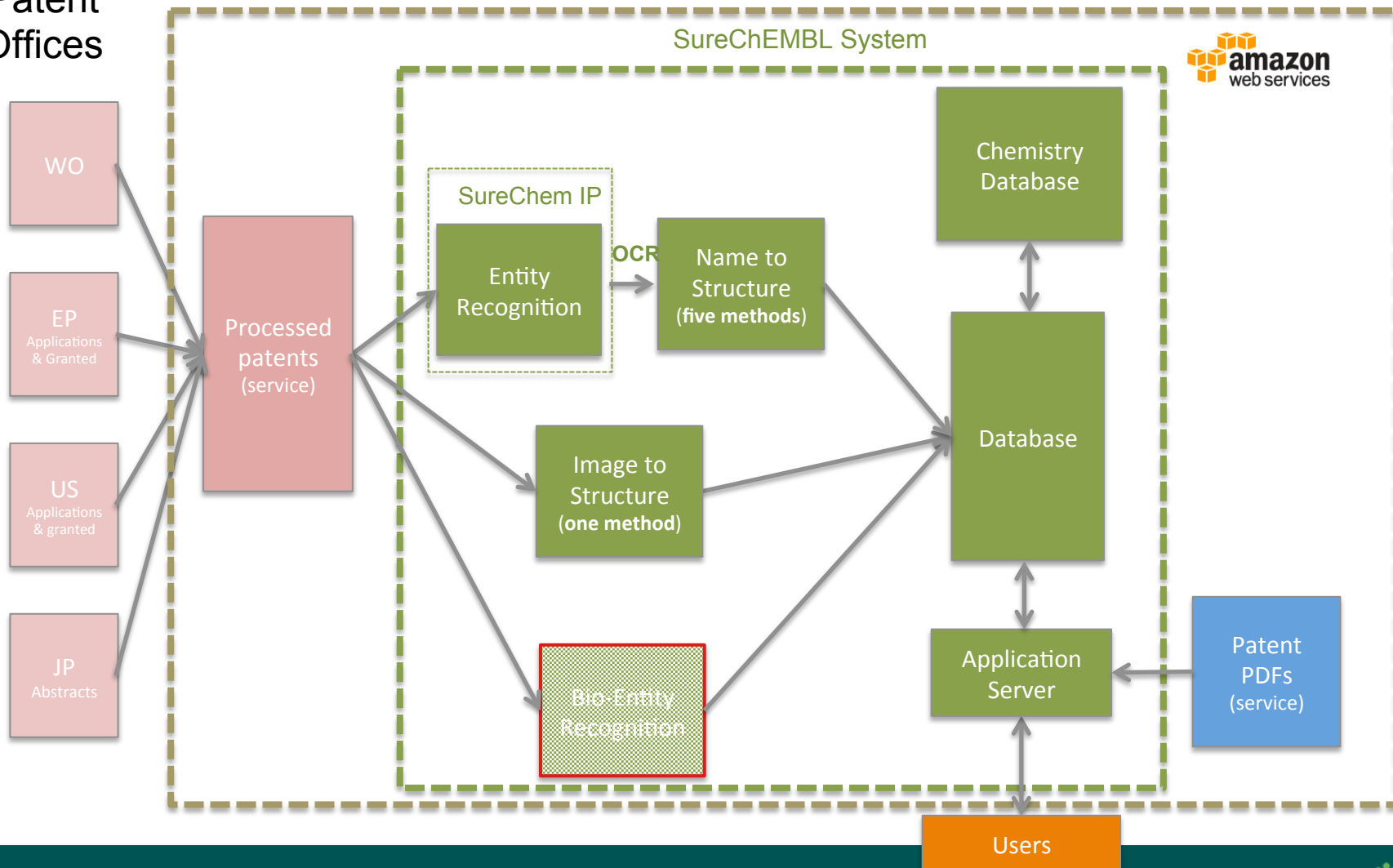
- Patent filing and searching
 - Legal, financial and commercial incentives & interests
 - Prior art, novelty, freedom to operate searches
 - Competitive intelligence
- Unprecedented wealth of knowledge
 - Most of the knowledge will never be disclosed anywhere else
 - Compounds, scaffolds, reactions
 - Biological targets, diseases, indications
 - Average lag of 2-4 years between patent document and journal publication disclosure for chemistry, 4-5 for biological targets

SureChEMBL data processing

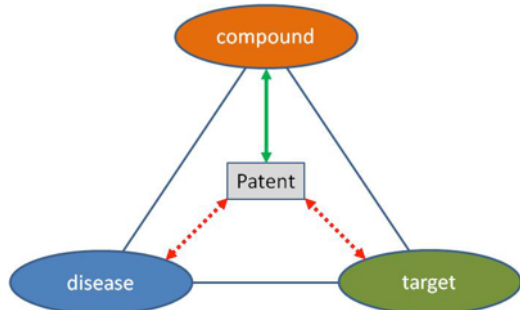


SureChEMBL data processing v2

Patent Offices



SureChEMBL bioannotation



US-9012636-B2

Back to search results

Front-page Claims Description

Your query

Highlight additional recognised chemical terms

Show select Biological annotations Powered by SciBite.com's Termite Engine

- Anatomy**
- Biological process**
- Endogenous cell**
- Gene**
 - HCR1 (19)
 - HCR12 (3)
 - HCR11 (2)
 - CBFA2T3 (1)
 - ALB (1)
- Gene Ontology**
- Indication**
 - Migraine (6)
 - Emesis (4)
 - Hypothalamic Diseases (3)

BACKGROUND OF THE INVENTION

The orexins (hypocretins) comprise two neuropeptides produced in the hypothalamus: the orexin A (OX-A) (a 33 amino acid peptide) and the orexin B (OX-B) (a 28 amino acid peptide) (Sakurai T. et al., Cell, 1998, 92, 573-581). Orexins are found to stimulate food consumption in rats suggesting a physiological role for these peptides as mediators in the central feedback mechanism that regulates feeding behaviour (Sakurai T. et al., Cell, 1998, 92, 573-581). Orexins regulate states of sleep and wakefulness opening potentially novel therapeutic approaches for narcoleptic or insomnia patients (Chenail R. M. et al., Cell, 1999, 98, 437-451). Orexins have also been indicated as playing a role in arousal, reward, learning and memory (Harris, et al., Trends Neurosci., 2006, 29 (10), 571-577). Two orexin receptors have been cloned and characterised in mammals. They belong to the super family of G-protein coupled receptors (Sakurai T. et al., Cell, 1998, 92, 573-581); the orexin-1 receptor (OX1 or OX1R) is selective for OX-A and the orexin-2 receptor (OX2 or OX2R) is capable to bind OX-A as well as OX-B. The physiological actions in which orexins are presumed to participate are thought to be expressed via one or both of OX1 receptor and OX2 receptor as the two subtypes of orexin receptors.

Orexin receptors are found in the mammalian brain and may have numerous implications in pathologies such as depression, anxiety, addiction, obesity, compulsive disorder, affective neurosis, depressive neurosis, anxiety neurosis, dysthymic disorder, bipolar disorder, mood disorder, sexual dysfunction, psychosocial dysfunction; sex disorder; schizophrenia; manic depression; delirium; dementia; severe mental retardation and dyskinesias such as Huntington's disease and Tourette syndrome; eating disorders such as anorexia, bulimia, cachexia, and obesity; addictive feeding behaviours; binge/purge feeding behaviour; cardiovascular diseases; diabetes; appetite/fat disorders; emesis, vomiting, nausea; asthma; cancer; Parkinson's disease; Chabot's syndrome/disease; haemophilic adenoma; prolactinoma; hyperprolactinemia; hyperprolactinoma; tumour/adenoma; hypothalamic disease; inflammatory bowel disease; gastric dyskinesia; gastric ulcers; Froehlich's syndrome; adrenohypophysial disease; hypophysial disease; adrenohypophysial hypofunction; adrenohypophysial hyperfunction; hypothalamic hypogonadism; Kallman's syndrome (anosmia, hypopsmia); functional or psychogenic amenorrhoea; hypopituitarism; hypothalamic hypothyroidism; hypothalamic-pituitary dysfunction; idiopathic hyperprolactinemia; hypothalamic disorders of growth; growth hormone deficiency; dwarfism; gigantism; acromegaly; disturbed biological and circadian rhythms; sleep disturbances associated with narcolepsy such as neurological disorders, neuropathic pain and restless leg syndrome; stroke and brain diseases, acute and congestive heart failure; hypotension; hypertension; urinary retention; osteoporosis; angina pectoris; myocardial infarction; ischemic or haemorrhagic stroke; subarachnoid haemorrhage; ulcers; allergies; benign prostatic hypertrophy; chronic renal failure; renal disease; impaired glucose tolerance; migraine; hyperalgesia; pain; enhanced or exaggerated sensitivity to pain; such as hyperalgesia, neuralgia, and allodynia; acute pain; burn pain; atypical facial pain; neuropathic pain; back pain; complex regional pain syndrome I and II; arthritic pain; sports injury pain; pain related to infection e.g. HIV, post-chemotherapy pain; post-stroke pain; post-operative pain; neuralgia; emesis, nausea, vomiting; conditions associated with visceral pain such as irritable bowel syndrome, and angina; migraine; urinary bladder incontinence e.g. urge incontinence; tolerance to narcotics or withdrawal from narcotics; sleep disorders; sleep apnea; narcolepsy; insomnia; parasomnia; jet lag syndrome; and neurodegenerative disorders including neurological entities such as disinhibition-dementia-parkinsonism-amyotrophy complex; pallido-ponto-sigmal degeneration; epilepsy; seizure disorders and other diseases related to general orexin system dysfunction.

SUMMARY OF THE INVENTION

The present invention is directed to processes for preparing a pyridyl piperidine compound which is an antagonist of orexin receptors and which is useful in the treatment or prevention of neurological and psychiatric disorders and diseases in which orexin receptors are involved.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to a process for preparing a compound of the formula I:

- SciBite's Termite text-mining engine run on 4M life-science patents from SureChEMBL corpus
- Genes (identified by HGNC symbols) and diseases (identified by MeSH IDs) annotated
- Section/frequency information annotated (e.g., in title, abstract, claims, total frequency)
- Relevance score (0-3) to flag important chemical and biological entities and remove noise

Relevance scoring – genes/diseases

- Various features used:
 - Term frequency
 - Position (title, abstract, figure, caption, table)
 - Frequency distribution
- Scores range from 0 – 3
 - 3 – most important entities in the patent
 - 2 – important entities in the patent
 - 1 – mentioned entities in the patent
 - 0 – ambiguous entity/likely annotation error

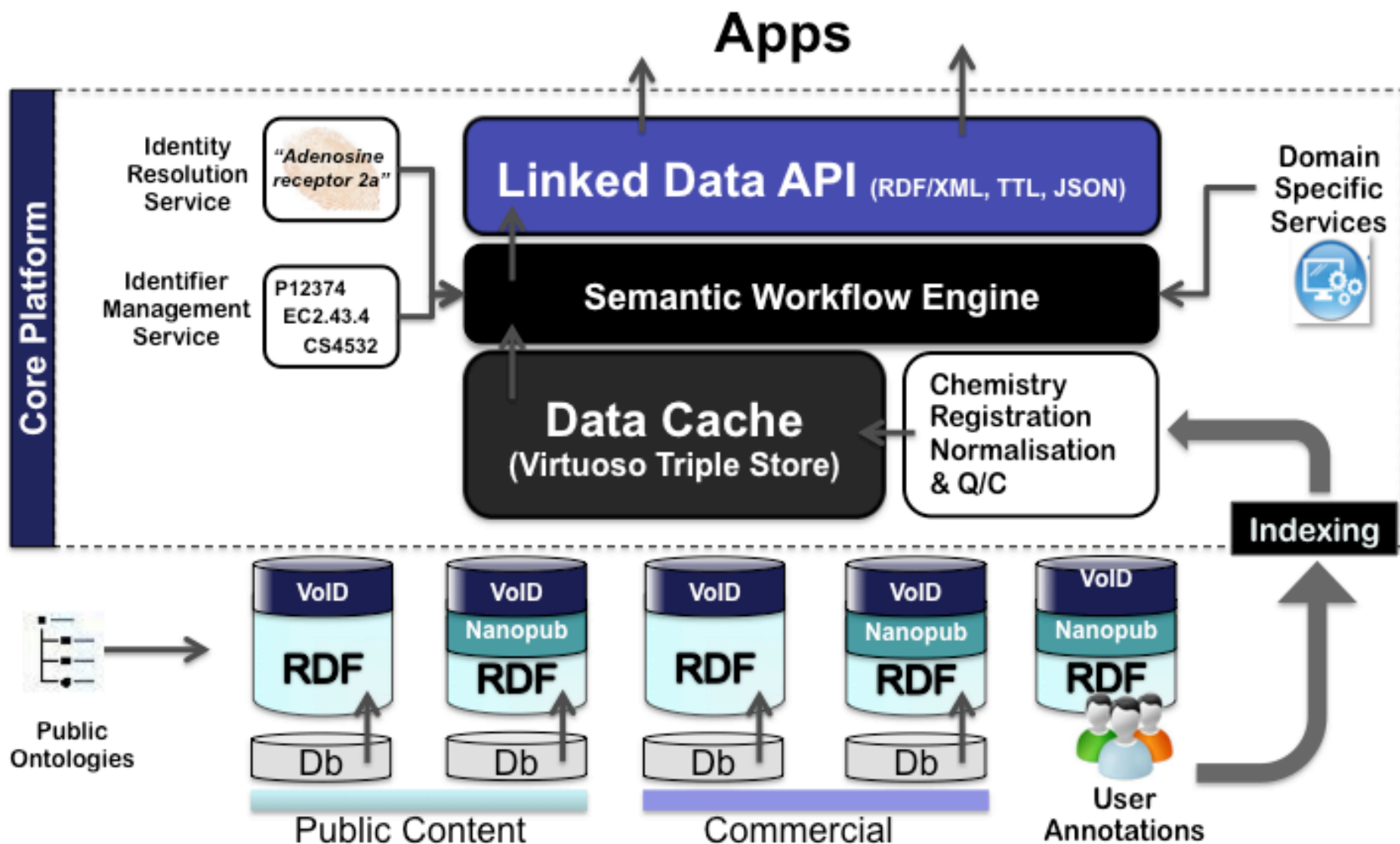
Relevance scoring - compounds

- Main assumptions for relevance:
 1. Very frequent compounds are irrelevant (but if drug-like then that's OK)
 2. Compounds with busy chemical space around them are interesting
 - Use distribution of close analogues (NNs) among compounds found in the **same patent family**
- Scores range from 0 – 3
 - 3 – highest number of NNs: most important entities in the patent
 - 2 – important entities in the patent
 - 1 – few NNs: mentioned entities in the patent
 - 0 – singletons or trivial entities, most likely errors or reagents, solvents, substituents

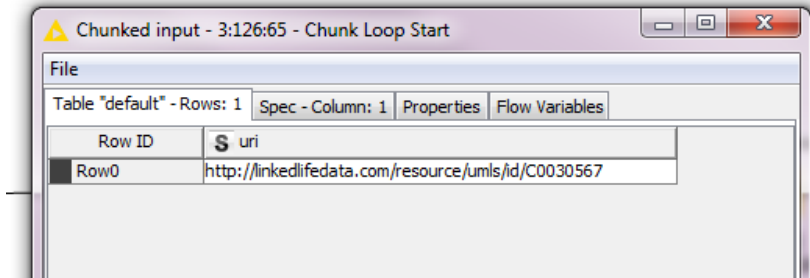
Gotchas & out of scope

- No Markush extraction
- No natural language processing (e.g., ‘compound **x** is an **inhibitor** of target **y**’)
- No extraction of bioactivities
- No chemistry search (yet)
- Patent coverage stops in April 2015
 - Incremental updates TBD
- Patent calls still in dev
 - Old scripts / workflows may break

Open PHACTS Architecture



The Open PHACTS node



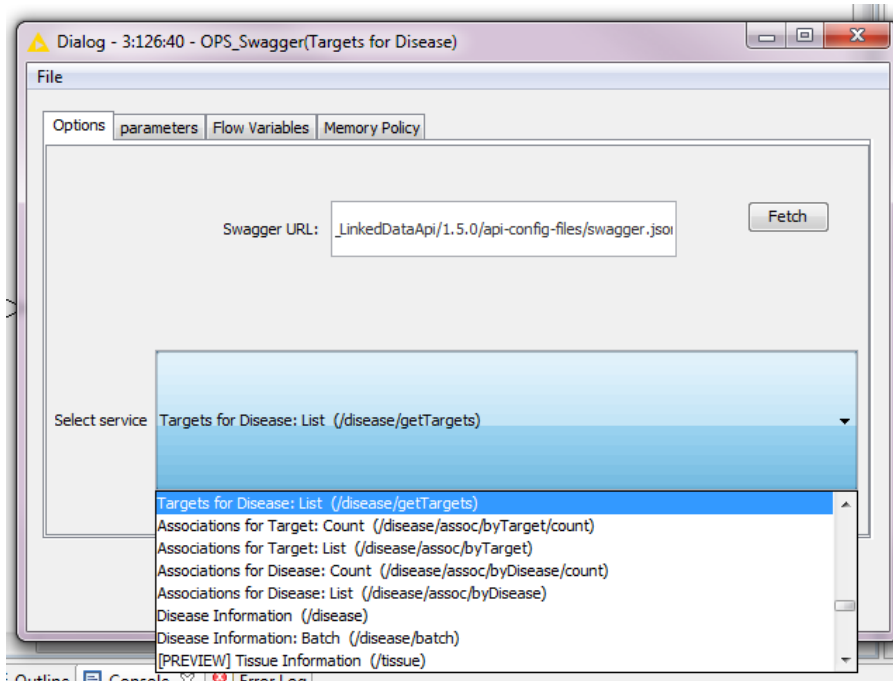
Chunked input - 3:126:65 - Chunk Loop Start

Row ID	uri
Row0	http://linkedlifedata.com/resource/umls/id/C0030567

OPS_Swagger



executable API call to
KREST nodes



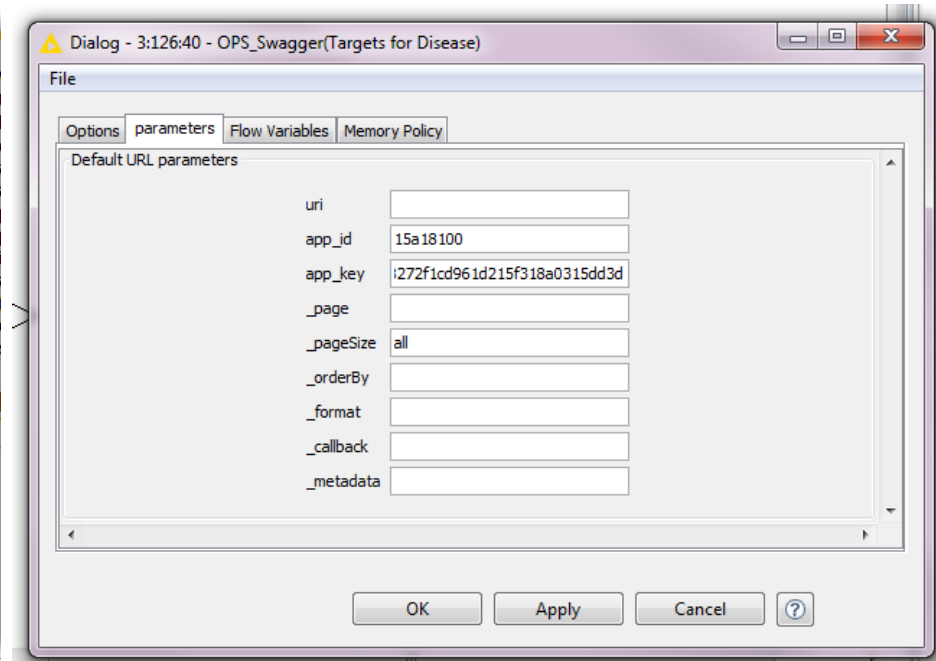
Dialog - 3:126:40 - OPS_Swagger(Targets for Disease)

Options parameters Flow Variables Memory Policy

Swagger URL: Fetch

Select service: **Targets for Disease: List (/disease/getTargets)**

- Targets for Disease: List (/disease/getTargets)
- Associations for Target: Count (/disease/assoc/byTarget/count)
- Associations for Target: List (/disease/assoc/byTarget)
- Associations for Disease: Count (/disease/assoc/byDisease/count)
- Associations for Disease: List (/disease/assoc/byDisease)
- Disease Information (/disease)
- Disease Information: Batch (/disease/batch)
- [PREVIEW] Tissue Information (/tissue)



Dialog - 3:126:40 - OPS_Swagger(Targets for Disease)

Options parameters Flow Variables Memory Policy

Default URL parameters

uri	<input type="text"/>
app_id	15a18100
app_key	i272f1cd961d215f318a0315dd3d
_page	<input type="text"/>
_pageSize	all
_orderBy	<input type="text"/>
_format	<input type="text"/>
_callback	<input type="text"/>
_metadata	<input type="text"/>

OK Apply Cancel ?

Open PHACTS Patent API

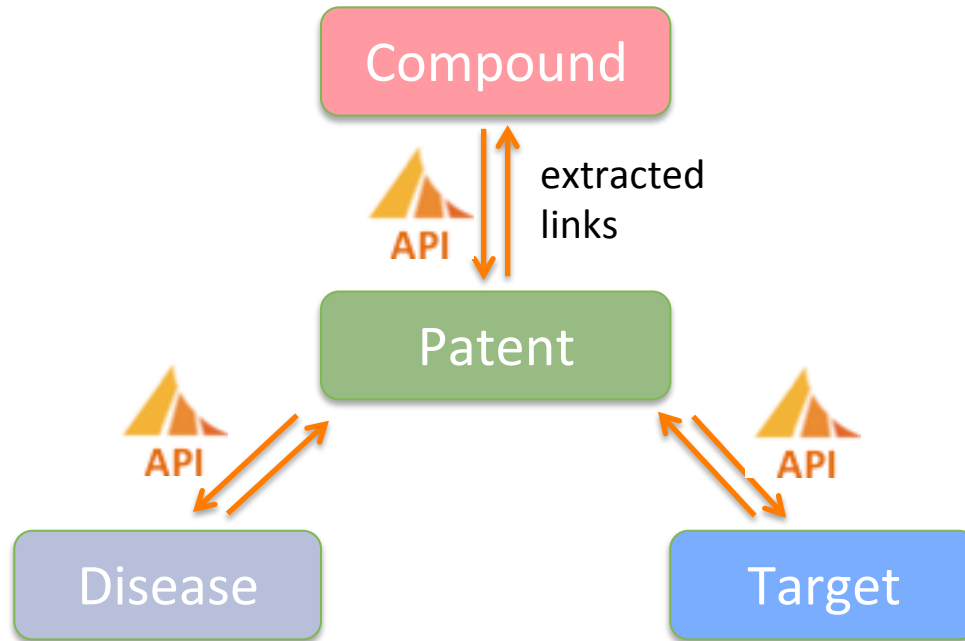
dev.openphacts.org

Patent Show/Hide | List Operations | Expand Operations

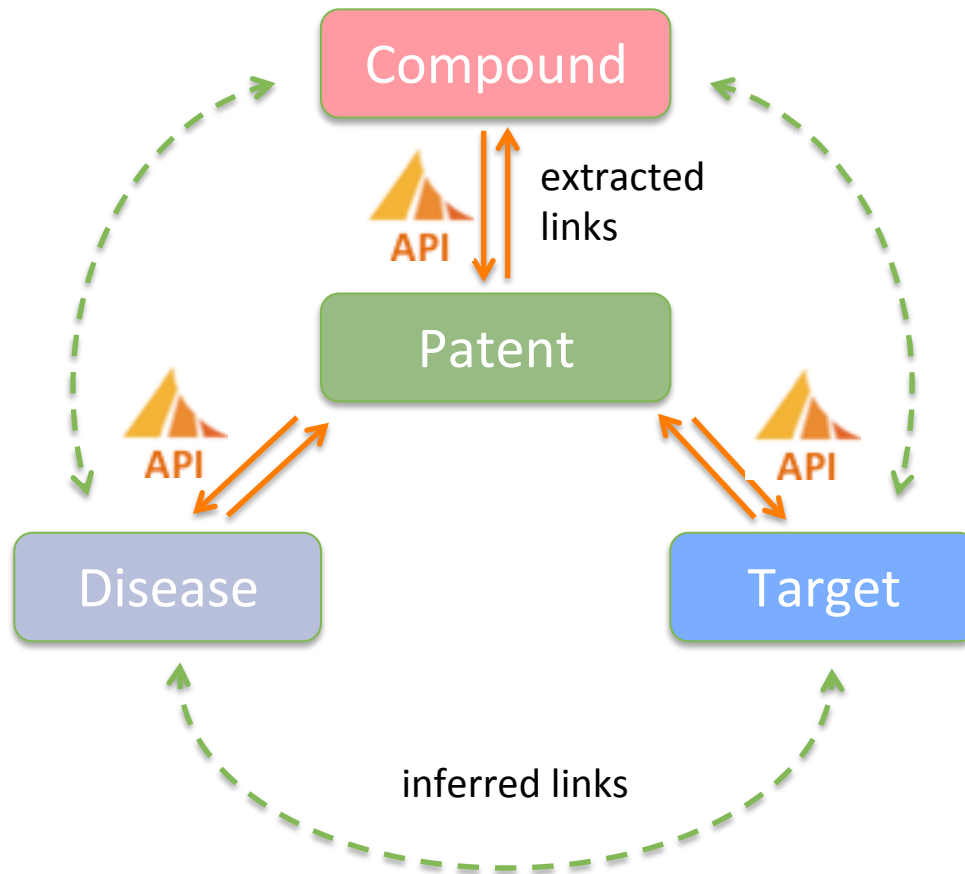
GET	/patent	[PREVIEW] Patent Information
GET	/patent/byCompound	[PREVIEW] Patents for Compound: List
GET	/patent/byCompound/count	[PREVIEW] Patents for Compound: Count
GET	/patent/byDisease	[PREVIEW] Patents for Disease: List
GET	/patent/byDisease/count	[PREVIEW] Patents for Disease: Count
GET	/patent/byTarget	[PREVIEW] Patents for Target: List
GET	/patent/byTarget/count	[PREVIEW] Patents for Target: Count
GET	/patent/entities/count	[PREVIEW] Patent Entities: Count
GET	/patent/entities/pages	[PREVIEW] Patent Entities: List

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

Open PHACTS Patent API

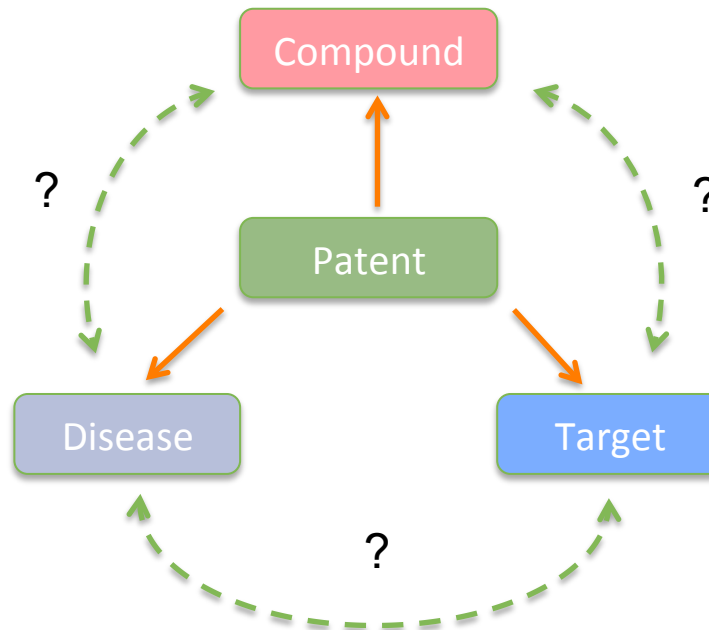


Open PHACTS Patent API



Use case #1: Patent to Entities

1. From a patent get compounds, genes and diseases
 - Frequency and relevance score
2. Filter to remove noise
3. Process and visualise



(12) **United States Patent**
Walter

(10) **Patent No.:** **US 7,718,693 B2**
(45) **Date of Patent:** **May 18, 2010**

(54) **RECEPTOR ANTAGONISTS AND THEIR METHODS OF USE**

WO WO 00/33788 A2 6/2000
WO WO 05/27882 A1 3/2005
WO WO 08/003697 A1 1/2008

(75) Inventor: **Daryl Simon Walter**, Harlow (GB)

(73) Assignee: **Glaxo Group Limited**, Greenford, Middlesex (GB)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **11/772,977**

(22) Filed: **Jul. 3, 2007**

(65) **Prior Publication Data**

US 2008/0009541 A1 Jan. 10, 2008

(30) **Foreign Application Priority Data**

Jul. 6, 2006	(GB)	0613473.8
Nov. 15, 2006	(GB)	0622825.8
Mar. 19, 2007	(GB)	0705263.2
Jun. 13, 2007	(GB)	0711439.0

(51) **Int. Cl.**
A61K 31/4015 (2006.01)
C07D 207/04 (2006.01)

OTHER PUBLICATIONS

Angier et al.; "Some N-Substituted-5-oxo-2-pyrrolidincarboxamides"; *Journal of Organic Chemistry*; 1956; vol. 21, No. 12; pp. 1540-1543.
Zhao et al., *Bioorganic & Medicinal Chemistry*, 7(8), pp. 1647-1654 (1999).
Møss et al., *International Journal of Pharmaceutics*, 52(3), pp. 255-263 (1989).
Tye et al., *Organic & Biomolecular Chemistry*, vol. 2, pp. 813-815 (2004).
Harriman, *Tetrahedron Letters*, 38(32), pp. 5591-5594 (1997).
Bundgaard et al., *Biochemical Society Transactions*, 17(5), pp. 947-949 (1989).
Bundgaard et al., *Journal of Pharmaceutical Sciences*, 78(2), pp. 122-126 (1989).
Møss et al., *Acta Pharmaceutica Nordica*, 4(4), pp. 301-308 (1992).
International Search Report for corresponding application PCT/EP2007/056675 (published as WO 08/003697A1).
Pharmaprojects Review (Jan. 2009) http://www.pharmaprojects.com/therapy_analysis/purin_P2X7_0109.htm.
Chemcats listing of CAS # 1001389-19-0 available from ChemDiv, Inc., San Diego, CA (Jul. 1, 2008).

Primary Examiner—Golam M M Shameem
(74) *Attorney, Agent, or Firm*—Duke M. Fitch; Kathryn L. Sieburth; Lorraine B. Ling

US-7718693-B2

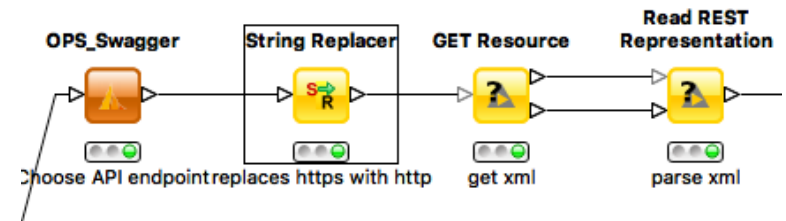
Use case #1: Patent to Entities



- Patent URI:
 - <http://rdf.ebi.ac.uk/resource/surechembl/patent/US-7718693-B2>
- API call:

GET /patent/entities/pages

- 586 entities back





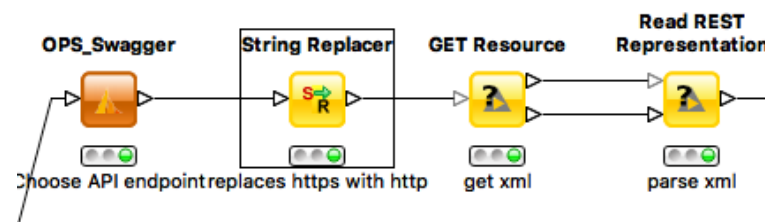
Use case #1: Patent to Entities

- Patent URI:
 - <http://rdf.ebi.ac.uk/resource/surechembl/patent/US-7718693-B2>

- API call:

GET </patent/entities/pages>

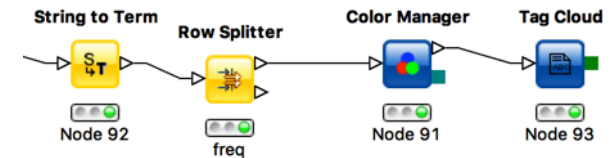
- 586 entities back



S item.href	S type.prefLabel	S smiles	i relevanceScore	i frequency	S occursInPatentTitle	S occursInPatentClaims	S occursInPatentDescription
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL766149	Molecule	<chem>COC(=O)C1CCC(=O)N1C1=CC=CC=C1</chem>	1	4	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/indication/D018771	Disease		2	4	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL288651	Molecule	<chem>OC(=O)[C@@H]1CCC(=O)N1CC1=CC=CC=C1</chem>	1	8	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/indication/D018805	Disease		1	1	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/indication/D018781	Disease		1	1	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL960	Molecule	<chem>C1C(=O)OCC1=CC=CC=C1</chem>	0	1	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL8956	Molecule	<chem>N1N=NC2=C1C=CC=C2</chem>	0	1	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL8249	Molecule	<chem>NC1=CC=C(C=C1)C(O)=O</chem>	0	1	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL2578506	Molecule	<chem>CC(C)N1[C@@H](CCC1=O)C(=O)NCC1=CC=CC=C1</chem>	2	2	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL2586270	Molecule	<chem>CCN1[C@@H](CCC1=O)C(=O)NCC1=CC=CC=C1</chem>	2	1	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL2583583	Molecule	<chem>CC(C)C(OC(=O)CN=C1C[C@@H]2C[C@@H](C2...</chem>	1	4	false	false	true
http://rdf.ebi.ac.uk/resource/surechembl/indication/D017116	Disease		1	1	false	false	true

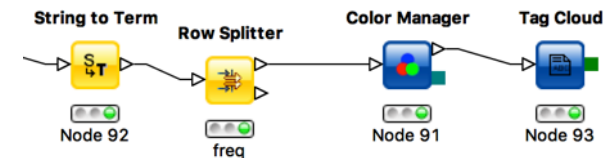
1) Look at target and disease entities

- 178 target and disease entities
- Filter: Relevance score $\geq 2 \rightarrow 23$ remain
- Visualise in tag cloud by frequency



1) Look at target and disease entities

- 178 target and disease entities
- Filter: Relevance score $\geq 2 \rightarrow 23$ remain
- Visualise in tag cloud by frequency



Does it make sense?

Immunologic Deficiency Syndromes
Motor Neuron Disease Bone Diseases
Myocardial Ischemia Visceral Pain
Carcinoma Neuropathic Pain Osteoarthritis
purinergic receptor P2X, ligand-gated ion channel, 7
neurodegenerative disease Pain Inflammation Neuralgia
Nephritis Dementia Alzheimer Disease Hyperalgesia
Paresthesia Rheumatoid Arthritis Headache
Arthralgia
Mild Cognitive Impairment

Does it make sense?

Immunologic Deficiency Syndromes
Motor Neuron Disease Bone Diseases
Myocardial Ischemia Visceral Pain
Carcinoma Neuropathic Pain Osteoarthritis
purinergic receptor P2X, ligand-gated ion channel, 7
neurodegenerative disease Pain Inflammation Neuralgia
Nephritis Dementia Alzheimer Disease Hyperalgesia
Paresthesia Rheumatoid Arthritis Headache
Arthralgia
Mild Cognitive Impairment

(57)

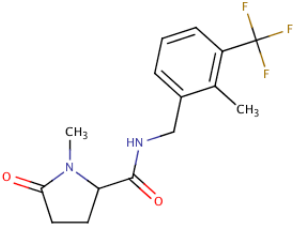
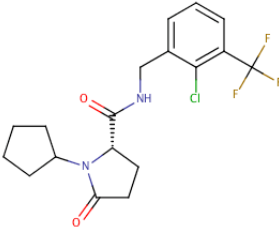
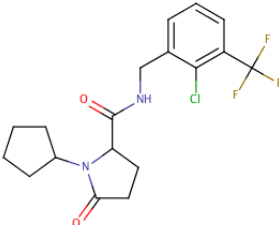
ABSTRACT

The present invention relates to novel oxo-prolinamide derivatives of formula (I) which modulate P2X7 receptor function and are capable of antagonizing the effects of ATP at the P2X7 receptor

and the use of such compounds or pharmaceutical compositions thereof in the treatment of disorders mediated by the P2X7 receptor, for example pain, inflammation and neurodegeneration.

2) Look at compound entities

- 408 compound entities
- Filter: Relevance score $\geq 1 \rightarrow$ 201 remain
- Calculate properties

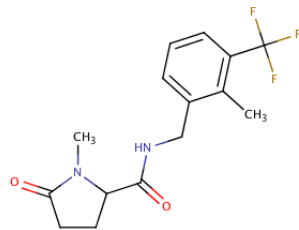
item.href	type.preLabel	relevanceScore	smiles
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL2578453	Molecule	2	
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL2584628	Molecule	2	
http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL2584624	Molecule	2	

2) Look at compound entities

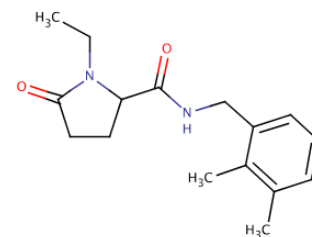
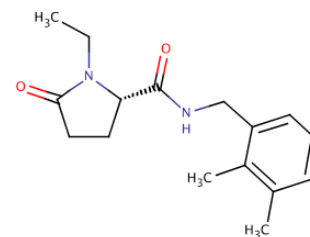
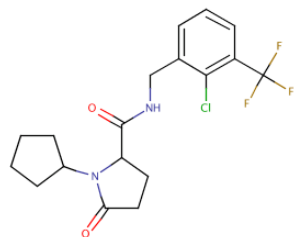
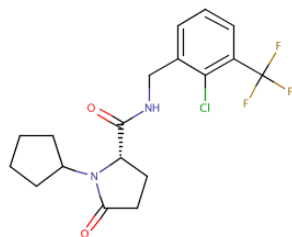
- 408 compound entities
- Filter: Relevance score $\geq 1 \rightarrow$ 201 remain
- Calculate properties

item.href | type.prefLabel | relevanceScore | smiles

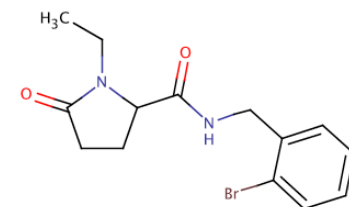
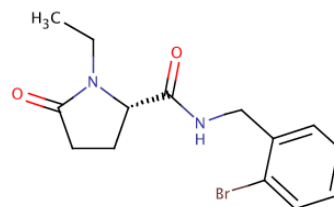
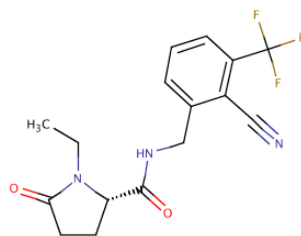
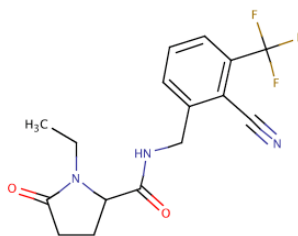
<http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL2578453> Molecule 2



<http://rdf.ebi.ac.uk/resource/surechembl/molecule/5C>



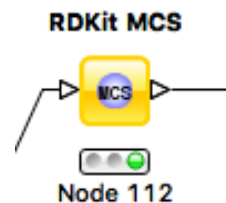
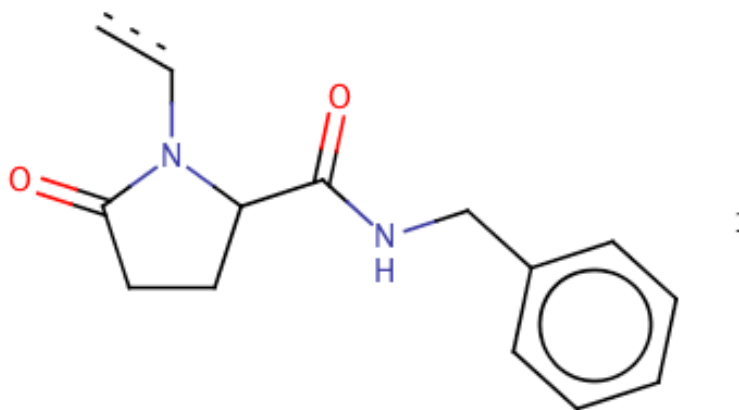
<http://rdf.ebi.ac.uk/resource/surechembl/molecule/5C>



Does it make sense?

- Calculate MCS

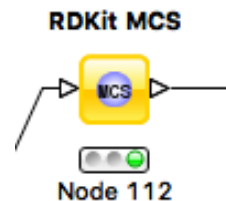
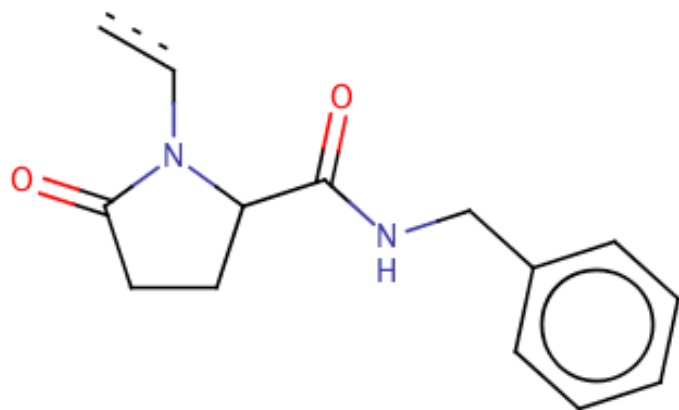
SMA
RTS MCS



Does it make sense?

- Calculate MCS

SMA
RTS MCS

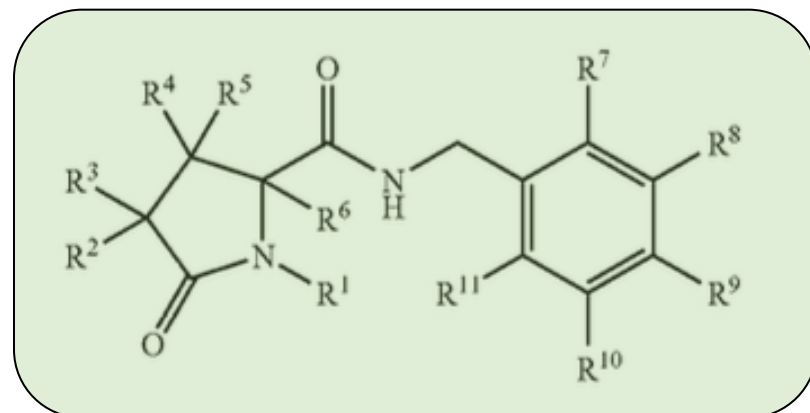


(57)

ABSTRACT

US-7718693-B2

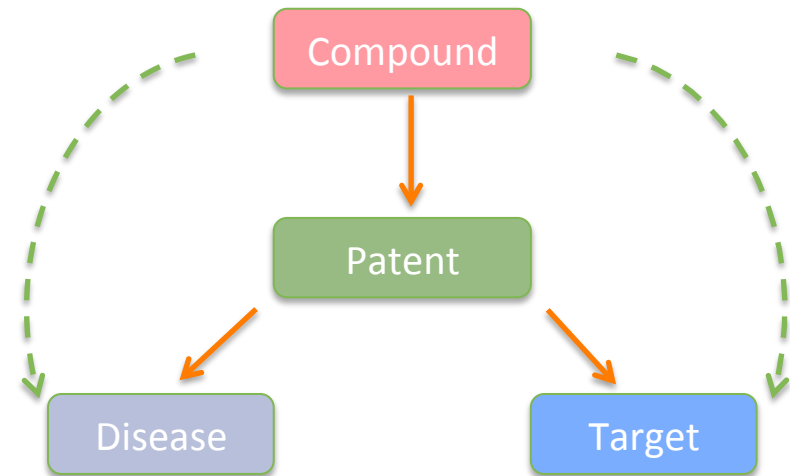
The present invention relates to novel oxo-prolinamide derivatives of **formula (I)** which modulate P2X7 receptor function and are capable of antagonizing the effects of ATP at the P2X7 receptor



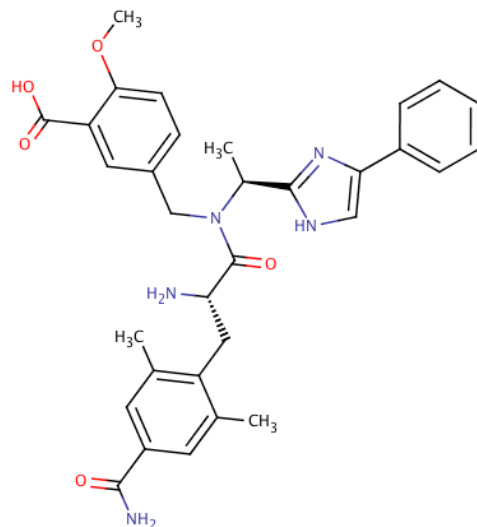
(I)

Use case #2: Drug targets & indications for compound

1. Search patents for a compound (approved drug)
2. Filter to remove noise
 - Frequency, relevance score and classification code
3. For remaining patents, get disease and target entities
4. Filter to remove noise
 - Frequency and relevance score
5. Visualise results



Eluxadoline (JNJ-27018966, VIBERZI)



CHEMBL2159122
FDA Approval: 2015

1) Get patents for Eluxadoline

- UniChem call → SCHEMBL12971682

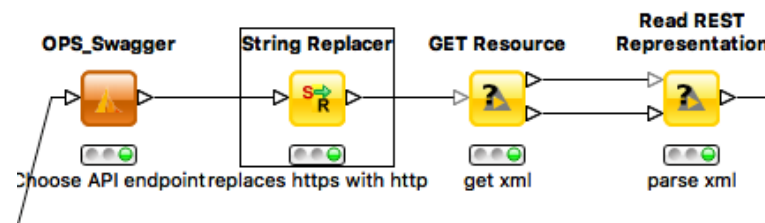
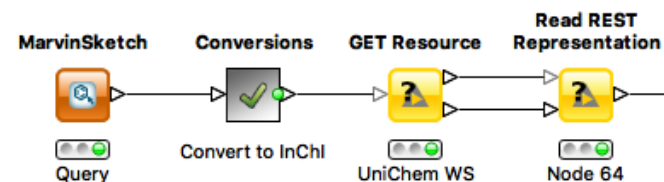
- Compound URI:

- <http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL12971682>

- API call:

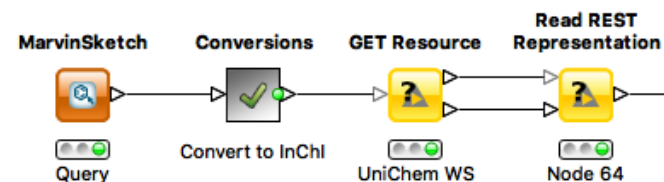
GET /patent/byCompound

- Relevance score ≥ 1 → 17 patents (patentome):



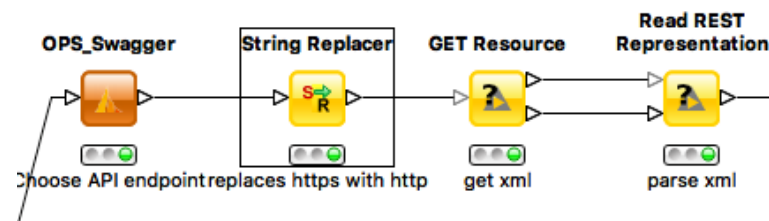
1) Get patents for Eluxadoline

- UniChem call → SCHEMBL12971682
- Compound URI:



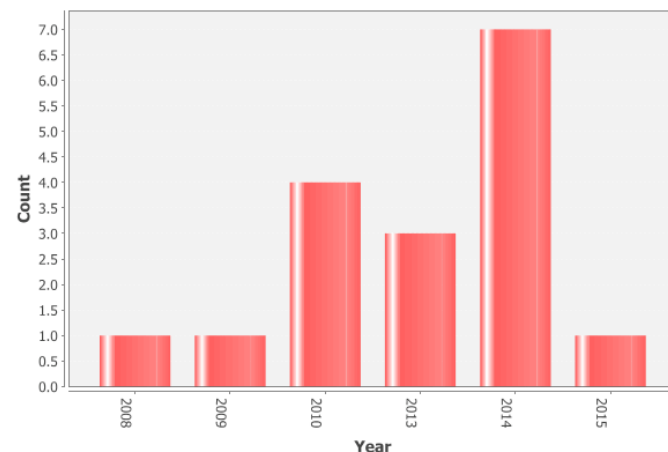
• <http://rdf.ebi.ac.uk/resource/surechembl/molecule/SCHEMBL12971682>

- API call:



- Relevance score ≥ 1 → 17 patents (patentome):

frequency	datePublis...	S title
3	24.Apr.2008	NOVEL COMPOUNDS AS OPIOID RECEPTOR MODULATORS
1	08.Dec.2009	2-tert-Butoxycarbonylamino-3-(4-carbamoyl-2,6-dimethylphe...
1	11.Feb.2010	PROCESS FOR THE PREPARATION OF OPIOID MODULATORS
5	22.Jun.2010	Compounds as opioid receptor modulators
2	31.Aug.2010	Compounds as opioid receptor modulators
3	23.Dec.2010	NOVEL COMPOUNDS AS OPIOID RECEPTOR MODULATORS
6	01.Jan.2013	Compounds as opioid receptor modulators
4	11.Apr.2013	NOVEL COMPOUNDS AS OPIOID RECEPTOR MODULATORS
1	23.Oct.2013	Opioid receptor modulators
1	09.Jan.2014	Novel crystals and process of making 5-(METHYL)-2-METHOX...
8	06.Feb.2014	NOVEL COMPOUNDS AS OPIOID RECEPTOR MODULATORS
1	08.Apr.2014	Crystals and process of making 5-((2-amino-3-(4-carbamoyl-...
8	08.Jul.2014	Compounds as opioid receptor modulators
8	11.Sep.2014	NOVEL COMPOUNDS AS OPIOID RECEPTOR MODULATORS
3	18.Sep.2014	OPIOID RECEPTOR MODULATOR DOSAGE FORMULATIONS
1	02.Oct.2014	OPIOID RECEPTOR MODULATOR DOSAGE FORMULATIONS
1	21.Jan.2015	PROCESS FOR THE PREPARATION OF OPIOID MODULATORS



2) Get target and disease entities for patents

- API call:

```
GET /patent
```

- Classification codes for patents
 - Filter: A61 and C07* → 11 patents remaining

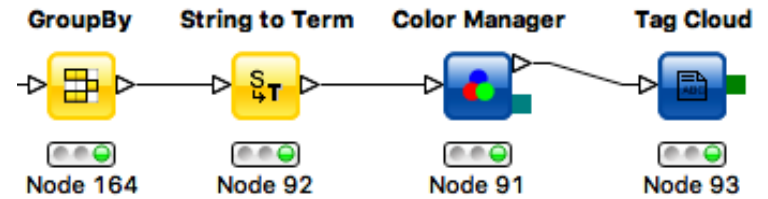
- API call:

```
GET /patent/entities/pages
```

- Filter: Relevance score ≥ 2 , Frequency ≥ 2 → 2 targets, 21 diseases
- Visualise with tag clouds by frequency

* <http://web2.wipo.int/classifications/ipc/ipcpub/>

Results



Relevant targets:

opioid receptor, delta 1
opioid receptor, mu 1

Relevant diseases:

Diabetes Mellitus
Postherpetic Neuralgia
Inflammatory Bowel Diseases
Irritable Bowel Syndrome Inflammation
Gastrointestinal Diseases
Chronic Pain Diarrhea Pain Constipation
Neuralgia Hyperalgesia Crohn Disease
Stroke Neuropathic Pain Ulcerative Colitis
Facial Pain Trigeminal Neuralgia Acute Pain
Peripheral Neuropathies Ileus

Does it make sense?

HIGHLIGHTS OF PRESCRIBING INFORMATION

These highlights do not include all the information needed to use VIBERZI safely and effectively. See full prescribing information for VIBERZI.

VIBERZI (eluxadoline) tablets, for oral use, C-X
Initial U.S. Approval: 2015

INDICATIONS AND USAGE

VIBERZI is a mu-opioid receptor agonist, indicated in adults for the treatment of irritable bowel syndrome with diarrhea (IBS-D). (1)

DOSAGE AND ADMINISTRATION

- The recommended dosage in adults is 100 mg twice daily taken with food. (2)
- The recommended dosage is 75 mg twice daily taken with food in patients who:
 - do not have a gallbladder (2, 5.1)
 - are unable to tolerate the 100 mg dose (2, 6.1)
 - are receiving concomitant OATP1B1 inhibitors (2, 7)
 - have mild or moderate hepatic impairment (2, 8.6)
- Discontinue VIBERZI in patients who develop severe constipation for more than 4 days (2)
- If a dose is missed, take the next dose at the regular time; do not take 2 doses at once (2)

Does it make sense?

HIGHLIGHTS OF PRESCRIBING INFORMATION

These highlights do not include all the information needed to use VIBERZI safely and effectively. See full prescribing information for VIBERZI.

VIBERZI (eluxadoline) tablets, for oral use, C-X
Initial U.S. Approval: 2015

INDICATIONS AND USAGE

VIBERZI is a mu-opioid receptor agonist, indicated in adults for the treatment of irritable bowel syndrome with diarrhea (IBS-D). (1)

DOSAGE AND ADMINISTRATION

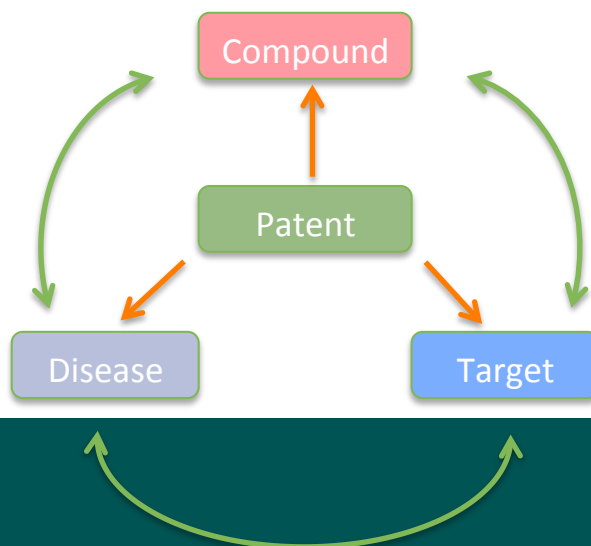
- The recommended dosage in adults is 100 mg twice daily taken with food. (2)
- The recommended dosage is 75 mg twice daily taken with food in patients who:
 - do not have a gallbladder (2, 5.1)
 - are unable to tolerate the 100 mg (2)
 - are receiving concomitant OAT (2)
 - have mild or moderate hepatic impairment (2)
- Discontinue VIBERZI in patients who experience constipation for more than 4 days (2)
- If a dose is missed, take the next dose as directed. Do not take 2 doses at once (2)

opioid receptor, delta 1
opioid receptor, mu 1

Diabetes Mellitus
Postherpetic Neuralgia
Inflammatory Bowel Diseases
Irritable Bowel Syndrome
Inflammation
Gastrointestinal Diseases
Chronic Pain
Diarrhea
Pain
Constipation
Neuralgia
Hyperalgesia
Crohn Disease
Stroke
Neuropathic Pain
Ulcerative Colitis
Facial Pain
Trigeminal Neuralgia
Acute Pain
Peripheral Neuropathies
Ileus

Take home message

- It is now possible to extract and interlink the *key* structures, scaffolds, targets and diseases from med. chem. patent corpus *automatically*
 - By high-throughput text-mining only
 - Thanks to simple heuristics (relevance scores and frequency)
 - Using the Open PHACTS API and **KNIME**
 - For the first time in a free resource in such scale



What next: Other ideas and use cases

- Target validation / Druggability
 - For a target, get me all related and relevant diseases
 - Compare with DisGeNET / CTTV, etc.
 - Any known patented scaffolds for my target?
 - Start a pharmacophore hypothesis for patent busting
- Novelty checking / Due diligence
 - What do we know about this scaffold / compound?
- Add ChEMBL pharmacology, pathway information
- Large-scale data mining
 - Annotated patent chemogenomics space and predictive models in KNIME
 - Anyone?

Availability

- API calls will be released to production soon – available for testing now:
 - <https://dev.openphacts.org/docs/develop>
 - KNIME workflows available on request
- SureChEMBL annotations licensed under CC BY-SA
- SciBite annotations licensed under CC BY-NC-SA
- Check out the Open PHACTS workshop on Friday 9am

ChEMBL web services: streamlining access to drug discovery data and utilities

Mark Davies, Michał Nowotka, George Papadatos, Nathan Dedman, Anna Gaulton, Francis Atkinson, Louisa Bellis and John P. Overington*

Nucleic Acids Research Advance Access published November 7, 2013

Nucleic Acids Research, 2013, 1–8
doi:10.1093/nar/gkt1031

The ChEMBL bioactivity database: an update

A. Patrícia Bento, Anna Gaulton, Anne Hersey, Louisa J. Bellis, Jon Chambers, Mark Davies, Felix A. Krüger, Yvonne Light, Lora Mak, Shaun McGlinchey, Michał Nowotka, George Papadatos, Rita Santos and John P. Overington*

Chambers et al. *Journal of Cheminformatics* 2014, 6:43
<http://www.jcheminf.com/content/6/1/43>



DATABASE

Open Access

UniChem: extension of InChI-based compound mapping to salt, connectivity and stereochemistry layers

Jon Chambers*, Mark Davies, Anna Gaulton, George Papadatos, Anne Hersey and John P. Overington

Challenges 2014, 5, 444–449; doi:10.3390/challe5020444

Communication

ChEMBL Beaker: A Lightweight Web Framework Providing Robust and Extensible Cheminformatics Services

Michał Nowotka, Mark Davies, George Papadatos and John P. Overington *

Bioinformatics Advance Access published November 20, 2013

Database and ontologies

myChEMBL: A virtual machine implementation of open data and cheminformatics tools

Rodrigo Ochoa, Mark Davies, George Papadatos, Francis Atkinson, John P. Overington*
European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK.

Papadatos et al. *Journal of Cheminformatics* 2014, 6:40
<http://www.jcheminf.com/content/6/1/40>



SOFTWARE

Open Access

A document classifier for medicinal chemistry publications trained on the ChEMBL corpus

George Papadatos[†], Gerard JP van Westen^{†*}, Samuel Croset, Rita Santos, Simone Trubian and John P. Overington

OPEN ACCESS

challenges

ISSN 2078-154

www.mdpi.com/journal/challenges

SureChEMBL: a large-scale, chemically annotated patent document database

George Papadatos¹, Mark Davies¹, Nathan Dedman¹, Jon Chambers¹, Anna Gaulton¹, James Siddle², Richard Koks², Sean A. Irvine³, Joe Pettersson⁴, Nicko Goncharoff^{2,*}, Anne Hersey^{1,*} and John P. Overington^{1,*}

¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK, ²Digital Health, 3240, New Zealand and ⁴McKinsey & Company, London

F1000Research

F1000Research 2016, 5(ELIXIR):160 Last updated: 11 FEB 2016



RESEARCH ARTICLE

Patterns of database citation in articles and patents indicate long-term scientific and industry value of biological data resources [version 1; referees: awaiting peer review]

David Bousfield^{1,2}, Johanna McEntyre³, Sameer Velankar³, George Papadatos³, Seung-Hyub Kim³, Florian Graef³, Vid Vartak³,

Senger et al. *J Cheminform* (2015) 7:49
DOI 10.1186/s13321-015-0097-z



RESEARCH ARTICLE

Open Access



Managing expectations: assessment of chemistry databases generated by automated extraction of chemical structures from patents

Stefan Senger^{1*}, Luca Bartek¹, George Papadatos² and Anna Gaulton²

Cheminformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, UK

feature



Drug discovery FAQs: workflows for answering multidomain drug discovery questions

Christine Chichester^{1,5}, christine.chichester@isb-sib.ch, Daniela Digles^{2,5}, Ronald Siebes³, Antonis Loizou³, Paul Groth³ and Lee Harland⁴

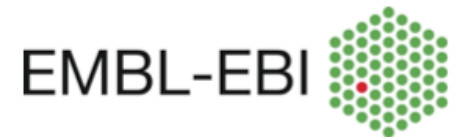
Scientific competency questions as the basis for semantically enriched open pharmacological space development

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 and Gerhard F. Ecker⁴

Acknowledgements

- ChEMBL and SureChEMBL
 - Anna Gaulton
 - Mark Davies
 - Nathan Dedman
 - James Siddle
 - Anne Hersey
- SciBite
 - Lee Harland
- Open PHACTS consortium
 - Nick Lynch
 - Daniela Digles
 - Antonis Loizou

- EBI alumni
 - Edmund Duesbury
 - Stephan Beisken



Technology partners



IFI CLAIMS®
Patent Services

a division of Fairview Research



OPSIN



The Data Are Out There

PubChem

EPA

IUPHAR
International Union of Basic and Clinical Pharmacology

Open PHACTS
Open Pharmacological Space

DisGeNET

Recon X
Reconstruction of the Human Genome

mcule
ONLINE DRUG DISCOVERY PLATFORM

日化辞Web
NikkajiWeb

KEGG
Kyoto Encyclopedia of Genes and Genomes

ZINC

BindingDB

EU-openscreen
Chemical keys for life's locks

ArrayExpress

LINCS

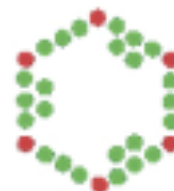
NMRShiftDB

MolPort

SureChEMBL

Concept Wiki

ChEBI ChEMBL



ChemSpider
The free chemical database

swissprot

DRUGBANK
Open Data Drug & Drug Target Database

UniProt

WikiPATHWAYS
Pathways for the People

GENEONTOLOGY
Unifying Biology

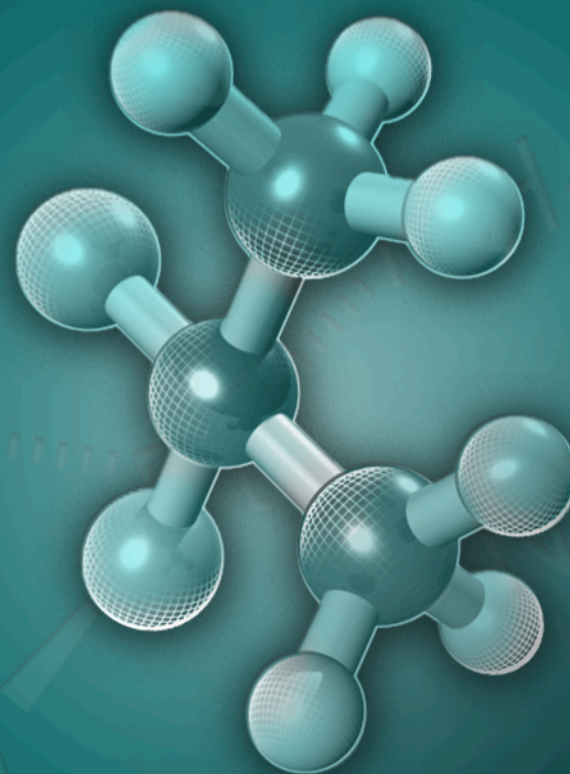
HMDB

Leveraging Open Chemogenomics Data and Tools with KNIME

George Papadatos

ChEMBL Group

georgep@ebi.ac.uk



Back-up slides

Example: All bioactivities for hERG

Dialog - 6:78 - ChEMBLdb Connector

Look-up Compound Search Flow Variables

ChEMBL ID:

Protein Accession:

Search Type:

Bioactivities:

OK Apply Cancel ?



ChEMBLdb Connector



All bioactivities for hERG

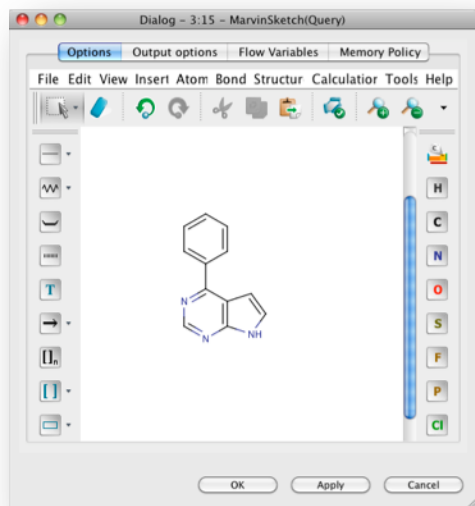
Output Table - 2:57:30 - XPath

Table "default" - Rows: 8869 Spec - Columns: 16 Properties Flow Variables

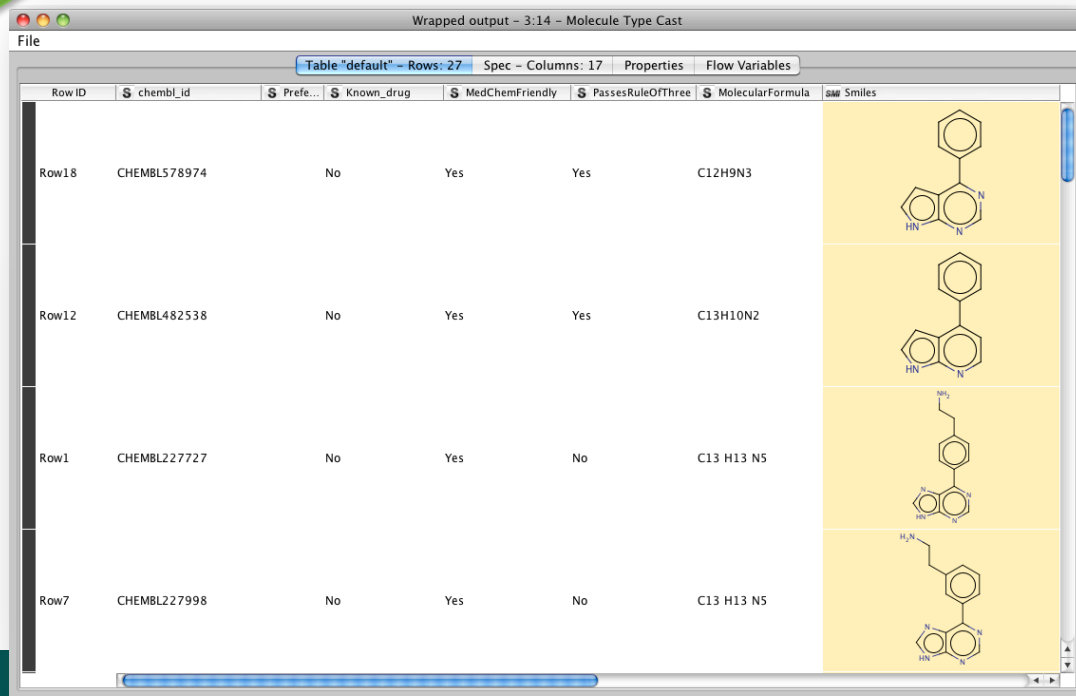
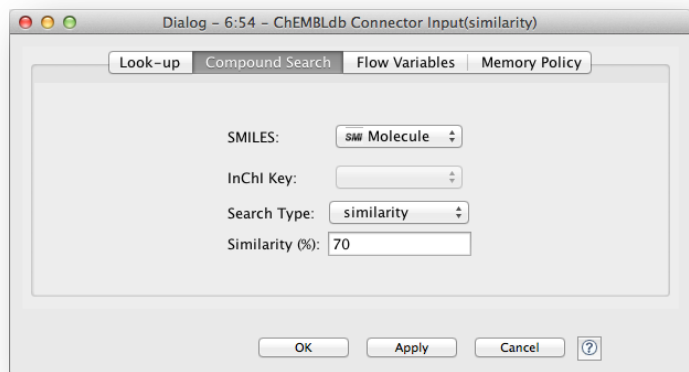
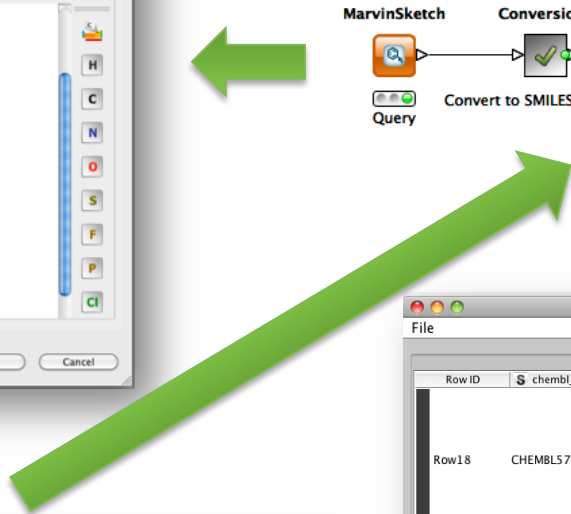
Row ID	\$ parent_cmpd...	\$ ingredient_cm...	\$ target_c...	\$ target_confidence	\$ target_...	\$ reference	\$ name_i...	\$ target_organism	\$ bioacti...	\$ activity...	\$ operator	\$ units	\$	
Row0	CHEMBL384400	CHEMBL384400	CHEMBL240	9	HERG	Biorg. Med. Chem., (2007) 15:5:20...	21e	Homo sapiens	IC50	Unspecified	=	nM	CHE	
Row1	CHEMBL310943	CHEMBL310943	CHEMBL240	7	HERG	J. Med. Chem., (2002) 45:12:2388	26B	Homo sapiens	IP	Unspecified	=	nM	CHE	
Row2	CHEMBL255312	CHEMBL255312	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2008) 18:...	4c cis-en...	Homo sapiens	IC50	Unspecified	=	nM	CHE	
Row3	CHEMBL229171	CHEMBL229171	CHEMBL240	9	HERG	Biorg. Med. Chem., (2007) 15:16:5...	15	Homo sapiens	Inhibition	Unspecified	=	%	CHE	
Row4	CHEMBL229291	CHEMBL229291	CHEMBL240	9	HERG	Biorg. Med. Chem., (2007) 15:16:5...	23d	Homo sapiens	Inhibition	Unspecified	=	%	CHE	
Row5	CHEMBL374865	CHEMBL374865	CHEMBL240	9	HERG	J. Med. Chem., (2007) 50:4:807	20f	Homo sapiens	Activity	Unspecified	>	nM	CHE	
Row6	CHEMBL213914	CHEMBL213914	CHEMBL240	7	HERG	Biorg. Med. Chem. Lett., (2006) 16:...	2	Homo sapiens	Inhibition	Unspecified	=	%	CHE	
Row7	CHEMBL398744	CHEMBL398744	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2006) 16:...	12m	Homo sapiens	Ki	Unspecified	=	nM	CHE	
Row8	CHEMBL299451	CHEMBL299451	CHEMBL240	6	HERG	J. Med. Chem., (1995) 38:11:1877	6f	Homo sapiens	IC50	Unspecified	=	nM	CHE	
Row9	CHEMBL172296	CHEMBL172296	CHEMBL240	7	HERG	Biorg. Med. Chem. Lett., (2004) 14:...	51d	Homo sapiens	Ki	Unspecified	=	nM	CHE	
Row10	CHEMBL232115	CHEMBL232115	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2007) 17:...	8s	Homo sapiens	IC20	Unspecified	=	nM	CHE	
Row11	CHEMBL12713	CHEMBL12713	CHEMBL240	7	HERG	J. Med. Chem., (2002) 45:18:3844	5 (Sertind...	Homo sapiens	IC50	Unspecified	=	nM	CHE	
Row12	CHEMBL43819	CHEMBL43819	CHEMBL240	7	HERG	J. Med. Chem., (2001) 44:10:1603	10	Homo sapiens	Ki	Unspecified	=	nM	CHE	
Row13	CHEMBL378666	CHEMBL378666	CHEMBL240	7	HERG	Biorg. Med. Chem. Lett., (2006) 16:...	7	Homo sapiens	Activity	Unspecified	<=	nM	CHE	
Row14	CHEMBL206209	CHEMBL206209	CHEMBL240	7	HERG	Biorg. Med. Chem. Lett., (2006) 16:...	34	Homo sapiens	Inhibition	Not Deter...	Unspecified	Unspecified	nM	CHE
Row15	CHEMBL12713	CHEMBL12713	CHEMBL240	7	HERG	J. Med. Chem., (2001) 44:11:1627	49 (sertin...	Homo sapiens	IC50	Unspecified	=	nM	CHE	
Row16	CHEMBL179714	CHEMBL179714	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2005) 15:...	10	Homo sapiens	Inhibition	Not Deter...	Unspecified	Unspecified	nM	CHE
Row17	CHEMBL519643	CHEMBL519643	CHEMBL240	7	HERG	Biorg. Med. Chem. Lett., (2005) 15:...	Gransetron	Homo sapiens	Log IC50	Unspecified	=	Unspecified	nM	CHE
Row18	CHEMBL212453	CHEMBL212453	CHEMBL240	7	HERG	Biorg. Med. Chem. Lett., (2006) 16:...	1f	Homo sapiens	Activity	Unspecified	>	nM	CHE	
Row19	CHEMBL367700	CHEMBL367700	CHEMBL240	7	HERG	Biorg. Med. Chem. Lett., (2004) 14:...	28b	Homo sapiens	Ki	Unspecified	=	nM	CHE	
Row20	CHEMBL1423	CHEMBL1423	CHEMBL240	7	HERG	J. Med. Chem., (2003) 46:11:2017	Pimozide	Homo sapiens	IC50	Unspecified	=	nM	CHE	
Row21	CHEMBL424872	CHEMBL424872	CHEMBL240	7	HERG	J. Med. Chem., (2006) 49:15:4455	8c	Homo sapiens	IC50	Unspecified	>	nM	CHE	
Row22	CHEMBL140577	CHEMBL140577	CHEMBL240	7	HERG	Biorg. Med. Chem. Lett., (2004) 14:...	15e	Homo sapiens	Ki	Unspecified	>	nM	CHE	
Row23	CHEMBL229229	CHEMBL229229	CHEMBL240	9	HERG	Biorg. Med. Chem., (2007) 15:16:5...	8l	Homo sapiens	Inhibition	Unspecified	=	%	CHE	
Row24	CHEMBL240864	CHEMBL240864	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2007) 17:...	17	Homo sapiens	Ki	Unspecified	=	nM	CHE	
Row25	CHEMBL245568	CHEMBL245568	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2007) 17:...	1a, SB-64...	Homo sapiens	Log IC50	Unspecified	=	Unspecified	nM	CHE
Row26	CHEMBL240863	CHEMBL240863	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2007) 17:...	16	Homo sapiens	Ki	Unspecified	>	nM	CHE	
Row27	CHEMBL66057	CHEMBL66057	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2006) 16:...	5	Homo sapiens	Activity	Unspecified	=	nM	CHE	
Row28	CHEMBL399352	CHEMBL399352	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2008) 18:...	27	Homo sapiens	Inhibition	Unspecified	=	%	CHE	
Row29	CHEMBL230704	CHEMBL230704	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2007) 17:...	25	Homo sapiens	Ki	Unspecified	=	nM	CHE	
Row30	CHEMBL214377	CHEMBL214377	CHEMBL240	7	HERG	J. Med. Chem., (2006) 49:22:6569	19	Homo sapiens	IC50	Unspecified	=	nM	CHE	
Row31	CHEMBL400917	CHEMBL400917	CHEMBL240	9	HERG	Biorg. Med. Chem. Lett., (2007) 17:...	2	Homo sapiens	IC50	Unspecified	>	nM	CHE	
Row32	CHEMBL353985	CHEMBL353985	CHEMBL240	8	HERG	Biorg. Med. Chem. Lett., (2004) 14:...	KCB-328	Homo sapiens	IC50	Unspecified	=	nM	CHE	

Activity value, assay description, compound, reference

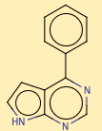
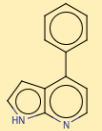
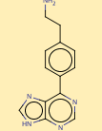
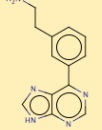
Example: Compound searching in ChEMBL



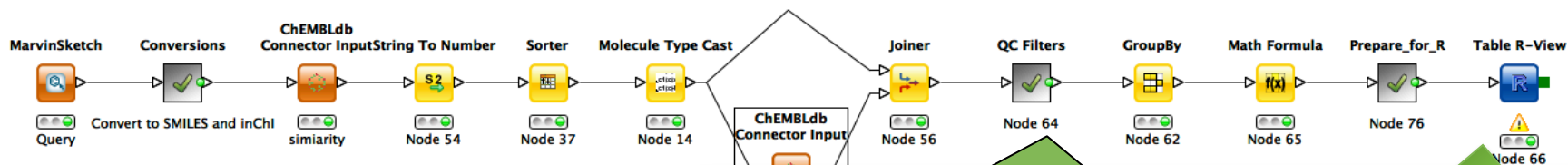
Query



Wrapped output - 3:14 - Molecule Type Cast

Row ID	chembl_id	Prefe...	Known_drug	MedChemFriendly	PassesRuleOfThree	MolecularFormula	Smw Smiles
Row18	CHEMBL578974	No	Yes	Yes	C12H9N3		
Row12	CHEMBL482538	No	Yes	Yes	C13H10N2		
Row1	CHEMBL227727	No	Yes	No	C13 H13 N5		
Row7	CHEMBL227998	No	Yes	No	C13 H13 N5		

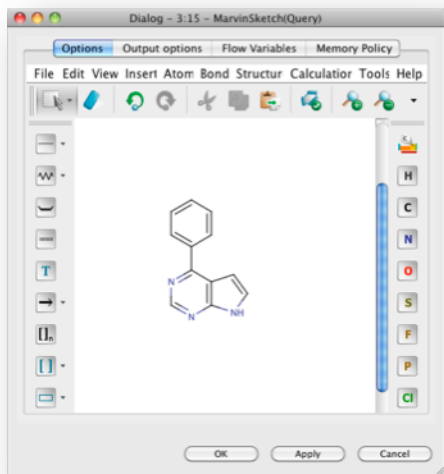
Example: Polypharmacology profile



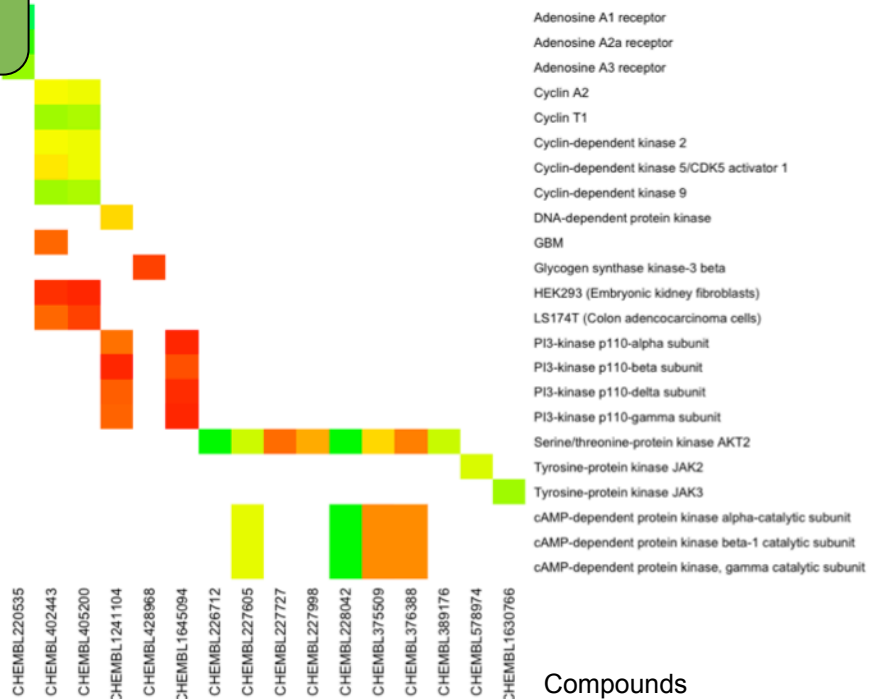
Find NNs

Retrieve bioactivities

Filter, summarise & pivot



Query



Targets

Compounds



myChEMBL LaunchPad

Welcome to the myChEMBL LaunchPad, providing access to all resources distributed with the myChEMBL virtual machine.

Web Interface

This web interface provides quick access to the myChEMBL data without any prior knowledge of SQL or RDKit.

phpPgAdmin Console

Use the console to explore the myChEMBL PostgreSQL database and run SQL queries (**user:** mychembl, **password:** read).

Web Services

Access to a local version of the official ChEMBL Web Services, which connect to the myChEMBL PostgreSQL database.

IPython Notebooks

A selection of programmatic tutorials written in Python and presented using interactive IPython Notebooks.

KNIME Integration

Learn how to connect the KNIME workbench to myChEMBL and also how to start processing ChEMBL data within a workflow environment.

ChEMBL Beaker

Access the functionality of the [RDKit](#) chemical toolkit and the optical structure recognition software [OSRA](#), via a RESTful API.

More Information

For more details on the myChEMBL project, including background, acknowledgements and references.

myChEMBL LaunchPad

192.168.56.101/knime

myChEMBL LaunchPad Home myChEMBL LaunchPad

myChEMBL KNIME Integration Home

🔗 What is KNIME?

KNIME is an open source modular graphical workbench for data processing and analysis. KNIME users can build workflows by executing predefined components, called *nodes*, in a process known as *visual programming*. Among other fields, KNIME has several extensions for life sciences, including bio- and chemoinformatics.

🔗 How do I get KNIME?

Installing KNIME is really easy across all popular platforms (Windows, Mac OS, Linux). Just follow [this](#) link.

🔗 How do I find out more about KNIME?

Here's a useful [page](#). For a basic introduction to KNIME followed by a few chemoinformatics applications, see [this](#) slideshow.

🔗 How to I use KNIME with the myChEMBL VM?

You can access the data in myChEMBL directly from a KNIME desktop running on your host machine. This can be achieved in two ways:

- Direct database connection with the Database Reader node. The node can execute arbitrary SQL SELECT commands and retrieve the results in a KNIME table.
- Calls to the RESTful [web services](#) that are hosted in myChEMBL. This is achieved with the KNIME REST ([KREST](#)) nodes, which are available as community node contribution. The KREST nodes send GET or POST requests to the server and format the response back to a KNIME table.

For an example workflow showcasing the KNIME/myChEMBL connectivity, download [this](#) file to your host machine. Then, open your KNIME desktop on your host machine and click File --> Import KNIME workflow... --> Select archive file --> Browse.. and then navigate to the zip file you've just downloaded and click Finish. If everything worked properly, you will be able to see something like this on your KNIME desktop:

Direct database connection

The input IP of myChEMBL VM is controlled by a workflow variable!
The default is: 192.168.56.101

```

graph LR
    N2[MarvinSketch Node 2] --> N3[TableRow To Variable Node 3]
    N3 --> N72[Java Edit Variable Node 72]
    N72 --> N1[Database Reader Node 1]
    N1 --> N4[Molecule Type Cast Node 4]
    N4 --> N19[2D/3D Scatterplot Node 19]
  
```

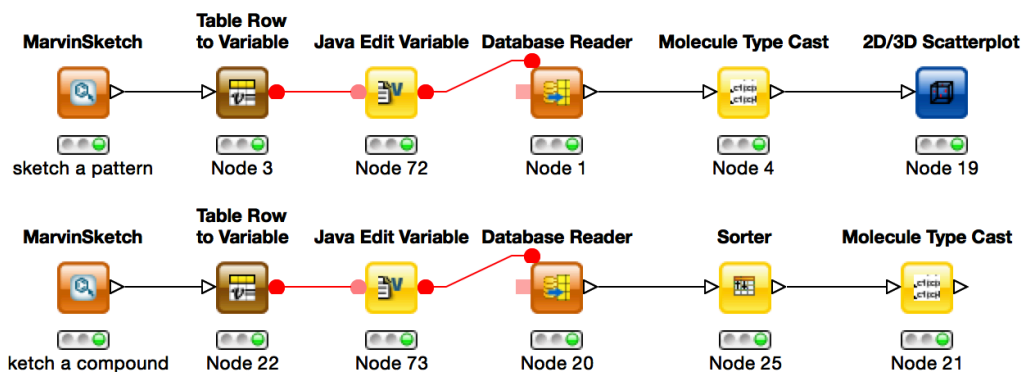
Substructure (SMARTS) Search

Similarity Search

Accessing *local* data and services with myChEMBL

Direct database connection

The input IP of myChEMBI VM is controlled by a workflow variable!
The default is: 192.168.56.101

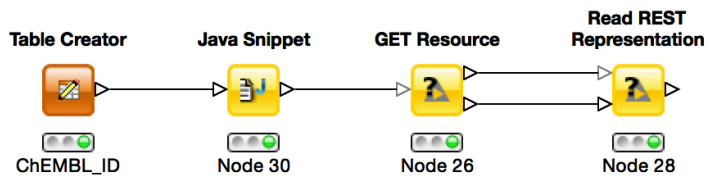


Substructure (SMARTS) Search

Similarity Search

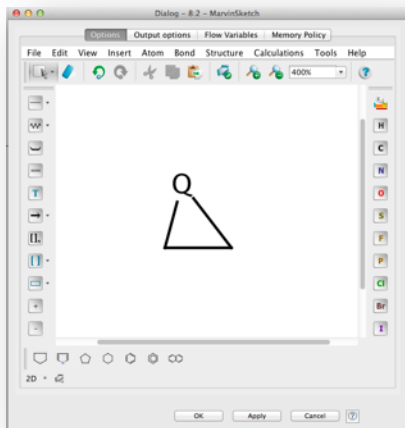
Web service connection

(requires the KREST community nodes
<http://tech.knime.org/book/krest-rest-nodes-for-knime-trusted-extension>)



Get bioactivities for a compound

Using KNIME to connect to myChEMBL



```
SELECT mr.*, md.chembl_id,  
cp.full_mwt, cp.alogp  
from mols_rdkit mr,  
molecule_dictionary md,  
compound_properties cp  
where  
mr.m @> '$${SMolecule}$${}'::qmol  
and  
mr.molregno = md.molregno  
and  
md.molregno = cp.molregno;
```

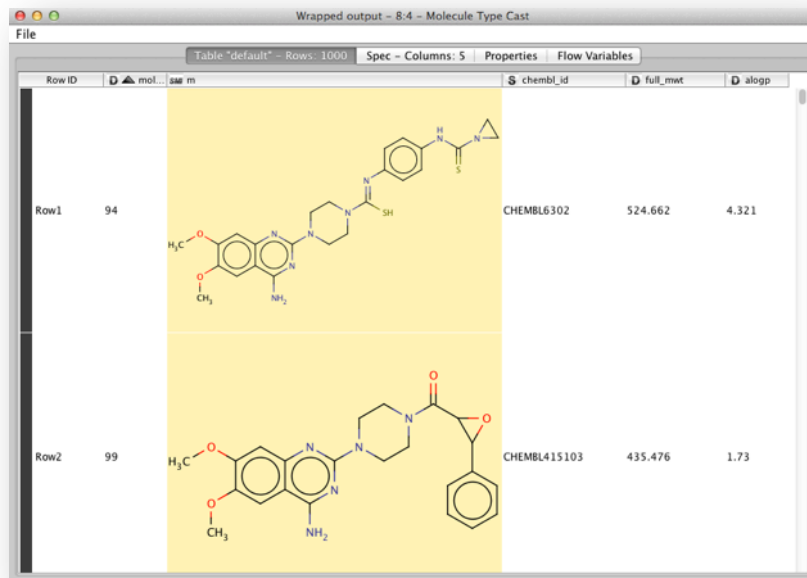
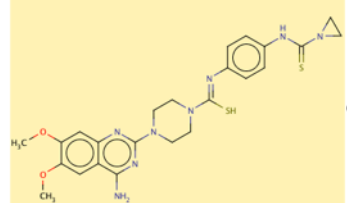
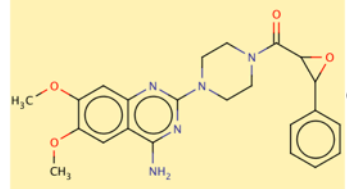
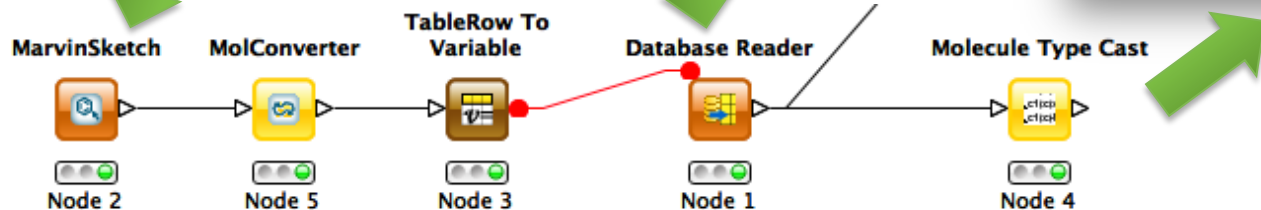


Table "default" - Rows: 1000 Spec - Columns: 5 Properties Flow Variables

Row ID	mol...	mol...	chembl_id	full_mwt	alogp
Row1	94		CHEMBL6302	524.662	4.321
Row2	99		CHEMBL415103	435.476	1.73



Cheminformatics utilities

ChEMBL Beaker API live documentation Explorer

Version: 0.5.24

GET [addHs/:CTAB](#)

GET [atomsInRing/:CTAB/:INDEX/:SIZE](#)

GET [break_bonds/:CTAB](#)

GET [ctab23D/:CTAB](#)

GET [ctab2image/:CTAB](#)

GET [ctab2inchi/:CTAB](#)

GET [ctab2json/:CTAB](#)

GET [ctab2smiles/:CTAB](#)

GET [ctab2svg/:CTAB](#)

GET [descriptors/:CTAB](#)

GET [getNumAtoms/:CTAB](#)

GET [image2ctab/:IMAGE](#)

GET [inchi2ctab/:INCHI](#)

GET [inchi2inchiKey/:INCHI](#)

- Chemical format conversions
- Dynamic image generation
- Image processing (via OSRA)
- Descriptors and property calculations
- Chemical modifications and standardization

Description

Converts SMILES to PNG image. This method accepts `unsafe_base64` encoded string containing single or multiple SMILES optionally containing header line, specific to `*.smi` format. Size is the optional size of image in pixels (default value is 200 px). Legend is optional label in the bottom of image.

Requires

SMILES

Formats

text

Enter a value for SMILES and click GET to test the service!

GET [/smiles2image/?znc5j33CC1](#)

Request URI

`smiles2image/Q04xQ0NOKENjMmNjYyhjYzlpQyg9TyIOYzNjY2MoCyyjKE5jNG5jY2MobjQpYzVjY2NuYzUpYzMpQ0MxIA==`

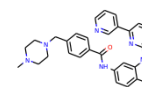
Response Code

200

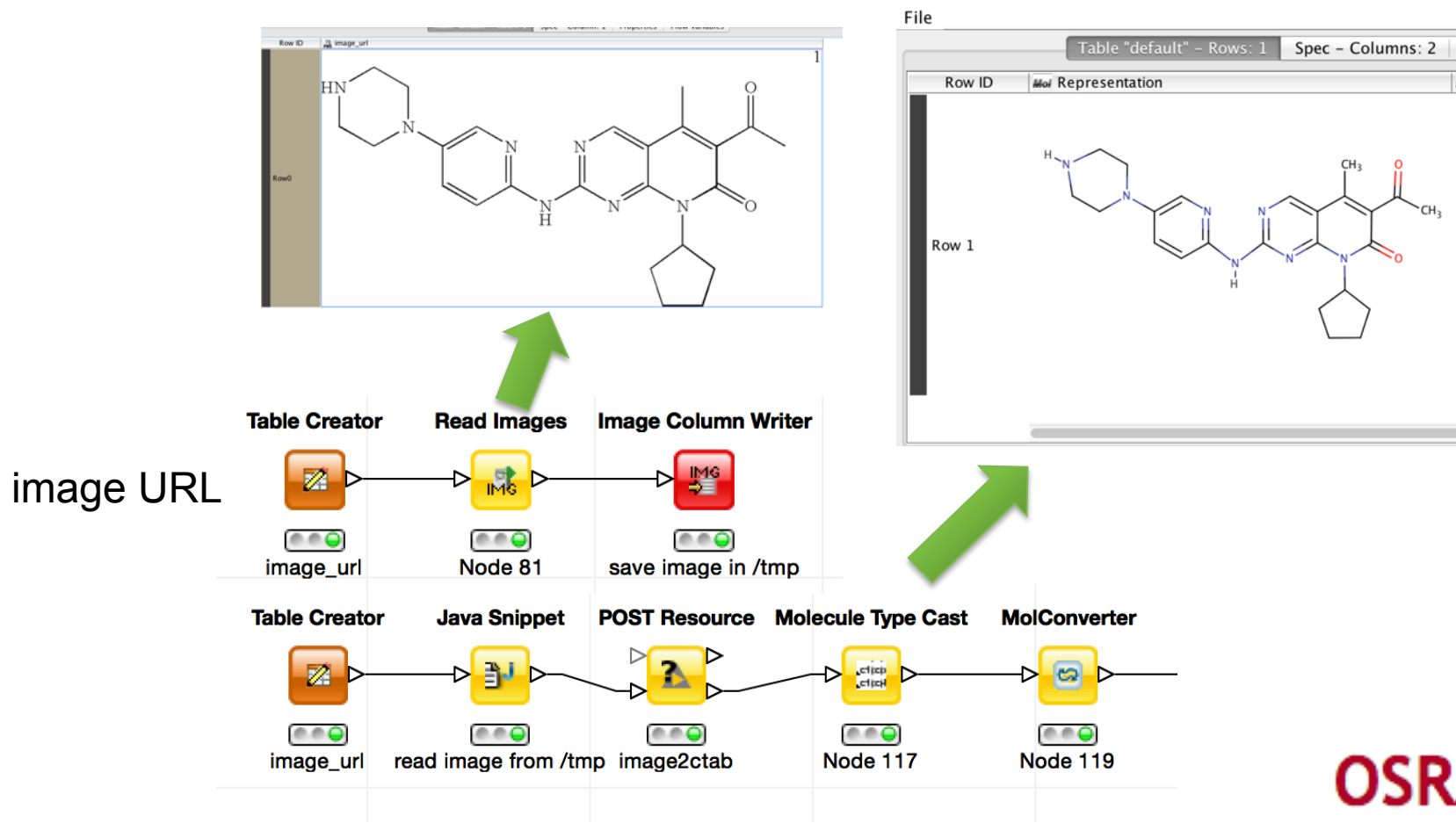
Response

OK

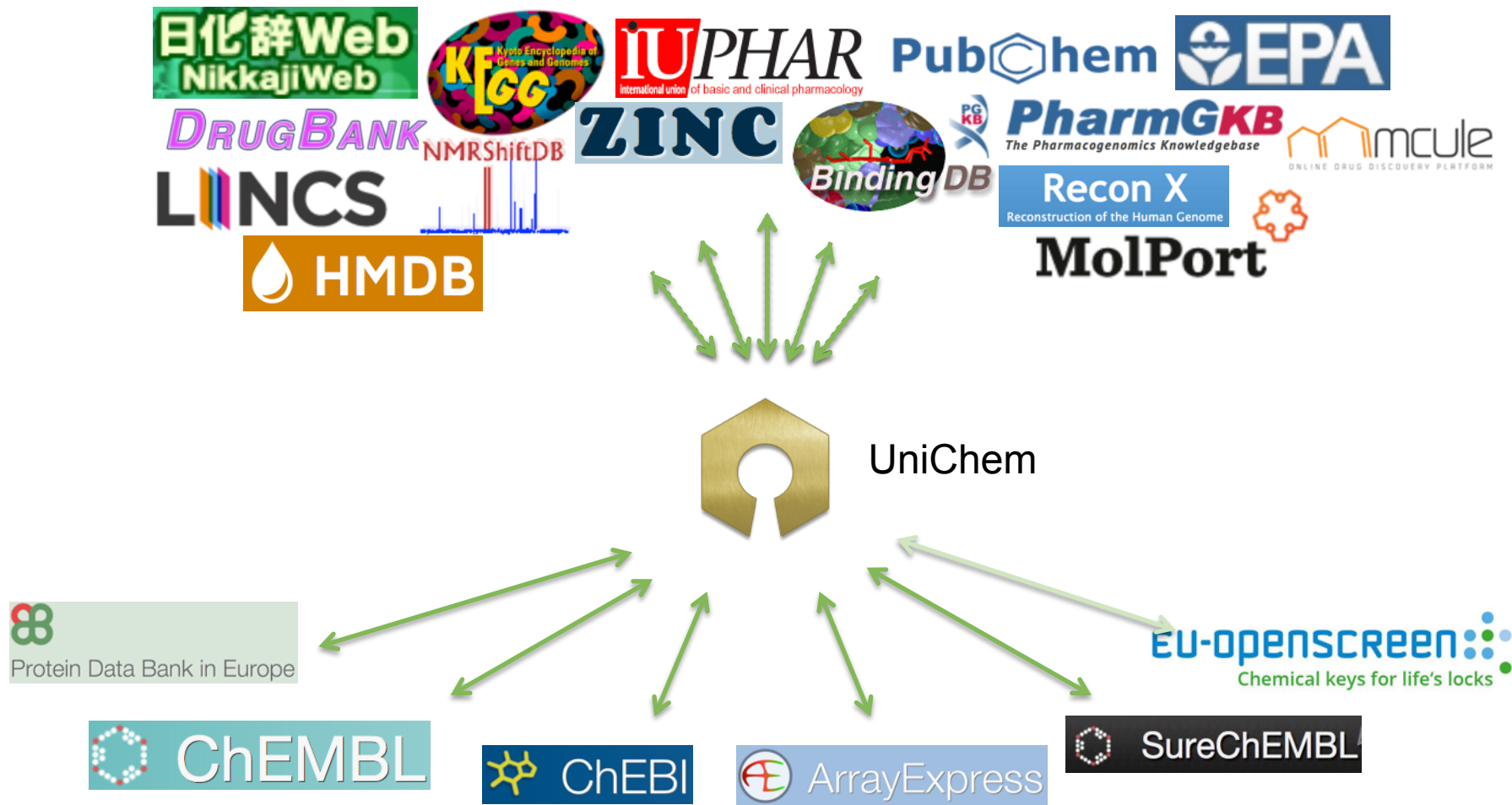
Response Body [Copy](#)



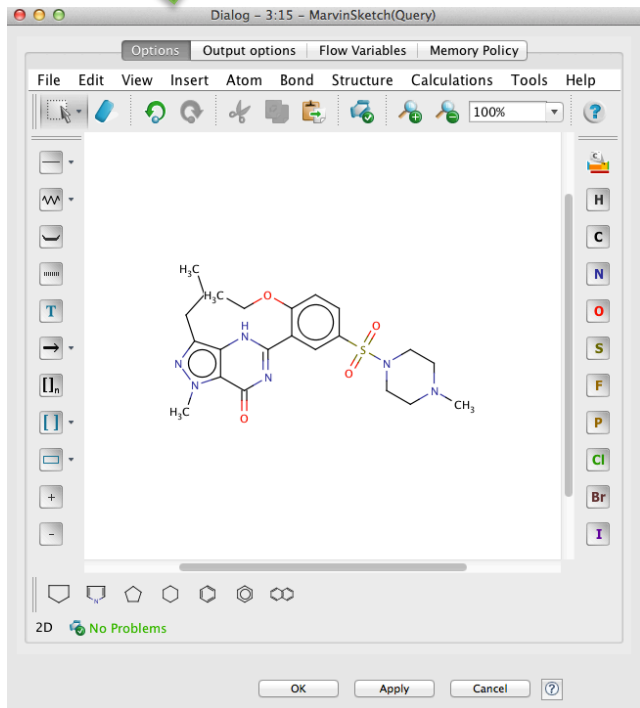
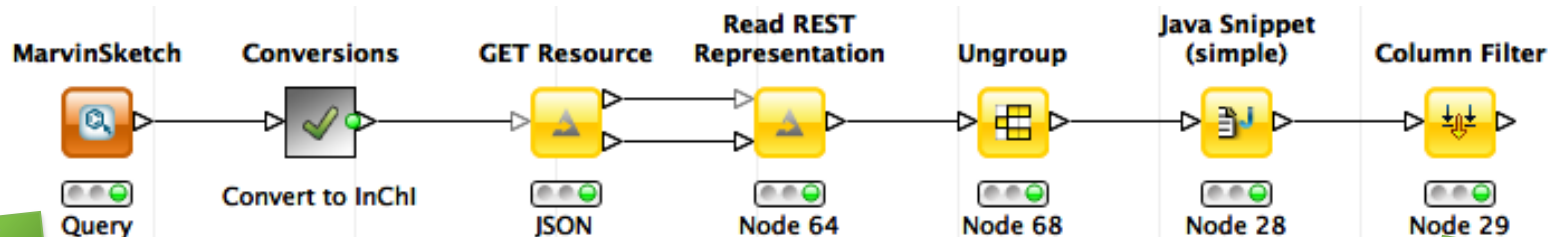
Example: RESTful Image to Structure conversion



UniChem – Compound Mapping across Resources



Novelty checking with UniChem



About Espacenet

Search Results

Refine search → Results

Smart search
Advanced search
Classification search

Quick help

- Can I subscribe to a list of the result list?
- What does the RSS feed with the result list?
- Can I export my results?
- What happens if I click "Download covers"?
- Why is the number of results sometimes only approximated?
- Why is the list limited?

1tb

PDB

1tb

Ligand environment

- ◊ Chain A, residue number 501, type VIA
- ◊ [More information](#) about VIA.
- ◊ [Details](#) of the binding sites of all ligands in this entry.
- ◊ Name: 5-{2-ethoxy-5-[(4-methylpiperazin-1-yl)sulfonyl]phenyl}-1-methyl-3-propyl-1h,6h,7h-

Deutsch Eng

Chan

overs (0)

	IPC:	Publication info:
1/14	A61K31/505	EP0812845 (A1)
5/26	A61K31/519	1997-12-17
7/04	A61K31/635 (+12)	EP0812845 (B1) 1999-07-14