

Mashing Up Drug Discovery

There's an API for that....

Lee Harland July 2013

Open PHACTS <http://openphacts.org>

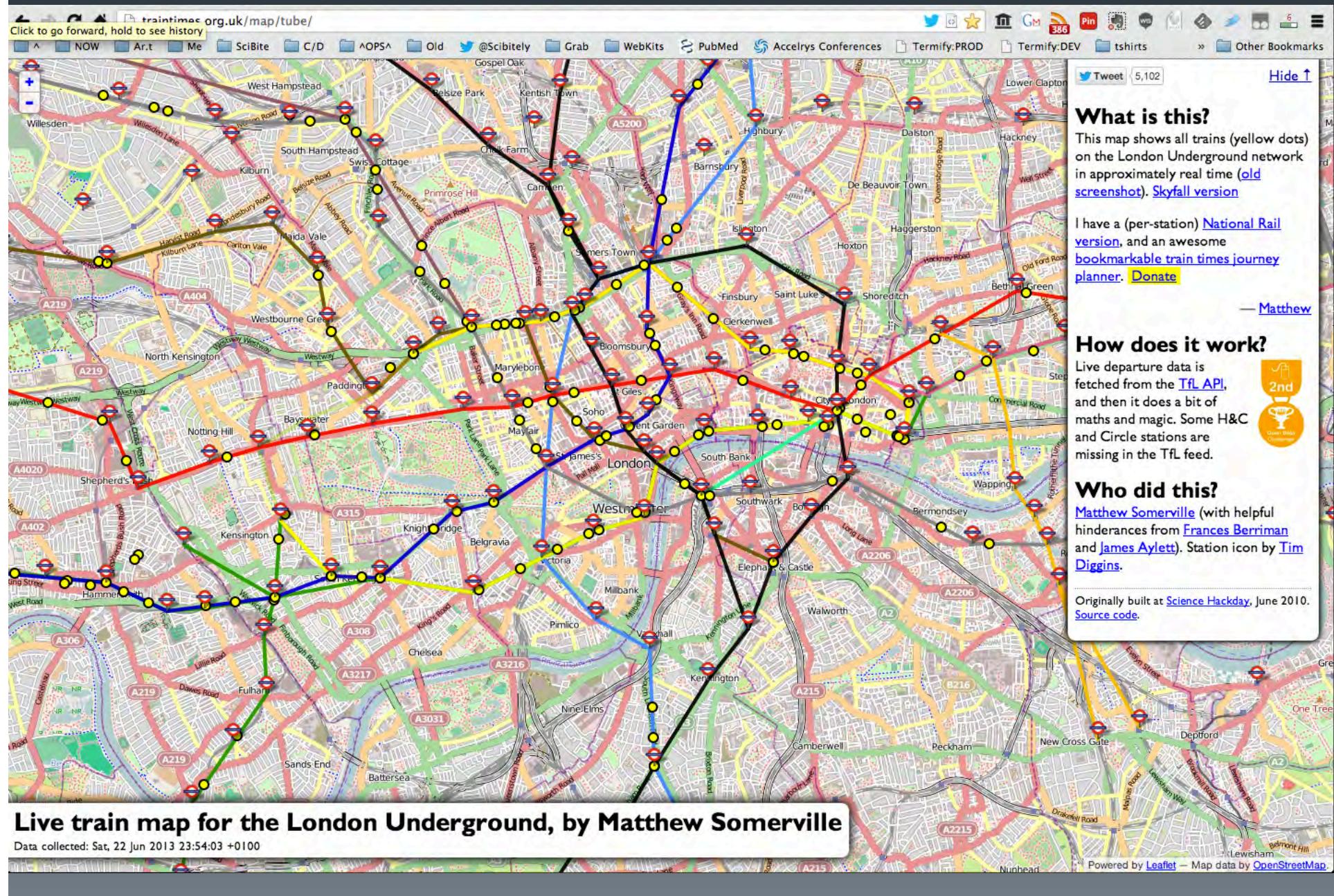
SciBite <http://scibite.com>

ConnectedDiscovery <http://connecteddiscovery.com>

@Scibitely

What does an
Application Program
Interface look like?

```
<application xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://wadl.dev.java.net/2009/02" xmlns:ops="http://ops.epo.org"
  xmlns:xs="http://www.w3.org/2001/XMLSchema" xsi:schemaLocation="http://wadl.dev.java.net/2009/02 ..\schema\wadl.xsd">
  <grammars>
    <include href="..\schema\ops.xsd"/>
  </grammars>
  <resources base="http://ops.epo.org/3.0/rest-services">
    <!-- GET Published data search -->
    <resource path="published-data/search/{publishedConstituents}" id="publishedDataSearch">
      <doc xml:lang="en" title="Published document search"/>
      <param name="publishedConstituents" type="xs:string" required="true" style="template" default="biblio">
        <option value="biblio"/>
        <option value="abstract"/>
        <option value="full-cycle"/>
        <option value="abstract,full-cycle"/>
        <option value="biblio,full-cycle"/>
      </param>
      <method name="GET" href="#wpdSearchGET"/>
      <!-- search query -->
    </resource>
    <!-- GET Published data biblio -->
    <resource path="published-data/{ref-type}/{ref-format}/{number}/{publishedConstituents}" id="publishedDataRetrieval">
      <doc xml:lang="en" title="Published document retrieval GET interface"/>
      <param href="#ref-type"/>
      <param href="#ref-format"/>
      <param name="number" type="xs:string" required="true" style="template" default="EP1000000">
        <doc xml:lang="en">Reference number</doc>
      </param>
      <param name="publishedConstituents" type="xs:string" required="true" style="template" default="biblio">
        <option value="biblio"/>
        <option value="abstract"/>
        <option value="full-cycle"/>
        <option value="abstract,full-cycle"/>
        <option value="biblio,full-cycle"/>
      </param>
      <method name="GET" href="#wpdGETPOST"/>
    </resource>
    <!--Published data fulltext availability inquiry GET -->
    <resource path="published-data/{ref-type}/{ref-format}/{number}/fulltext" id="publishedDataFulltextRequest">
      <doc xml:lang="en" title="Published document fulltext availability inquiry"/>
      <param href="#ref-type"/>
      <param href="#ref-format"/>
      <param name="number" type="xs:string" required="true" style="template" default="EP1000000">
        <doc xml:lang="en">Reference number</doc>
      </param>
      <method name="GET" href="#wpdGETPOST"/>
    </resource>
    <!--Published data claims retrieval GET -->
    <resource path="published-data/{ref-type}/{ref-format}/{number}/claims" id="publishedDataClaimsRequest">
      <doc xml:lang="en" title="Published document claims"/>
      <param href="#ref-type"/>
      <param href="#ref-format"/>
      <param name="number" type="xs:string" required="true" style="template" default="EP1000000">
        <doc xml:lang="en">Reference number</doc>
      </param>
      <method name="GET" href="#wpdGETPOST"/>
    </resource>
    <!--Published data description retrieval GET -->
    <resource path="published-data/{ref-type}/{ref-format}/{number}/description" id="publishedDataDescriptionRequest">
```



[Avian Flu home](#) > [Google Earth](#)**REPORTS CONTENT**[Avian flu home](#)[News](#)[Research](#)[Google earth](#)[Timeline](#)[FAQs](#)**REPORTS LINKS**[Stem cells](#)[Climate change](#)**NPG resources**[Nature](#)[Nature Cell Biology](#)[Nature Medicine](#)news@nature.com[Nature Reviews Drug Discovery](#)[Naturejobs](#)

Track the spread of the H5N1 virus worldwide since 2003 with our award-winning animated Google Earth Mash-Up.

[View the time-enabled maps](#)



Are There Any Bio/Chem/Med?

API Directory – ProgrammableWeb

www.programmableweb.com/apis/directory/1?apicat=Medical

W Ar.t Me SciBite C/D ^OPSA Old @Scibitely Grab WebKits PubMed Accelys Conferences Termify:PROD Termify:DEV

Hot APIs » Twitter YouTube Facebook Google Maps Flickr LinkedIn More » Latest news Today in APIs: Twitter adds Six New Ads API Partners, Microsoft...

Home API News API Directory Mashups Community How-to Contests Subscribe: RSS Email Twitter Facebook

Dashboard Directory Newest Most Popular By Category API Scorecard Add API

Web Services Directory

Hide Filters 

Sort by: Name Date Popularity Category

Keywords Category Company Protocols / Styles

Medical

Date Date Managed By

Filter This List

Viewing 1 to 117 of 117 APIs

| API | Description | Category | Updated |
|---|--|----------|------------|
| Access Plans USA | Health care insurance directory service | Medical | 2013-02-04 |
| Aetna CarePass | Health and drug information portal | Medical | 2012-06-05 |
| AIDSinfo | Government info on HIV/AIDS treatment | Medical | 2011-07-26 |
| Allen Brain Atlas | Neuroscience information resource directory | Medical | 2012-04-05 |
| Assay Depot | Pharmaceutical research service catalog | Medical | 2012-08-17 |
| Beddit | Sleep monitoring service | Medical | 2012-12-19 |
| Bililite | Health data graphing service | Medical | 2013-02-04 |
| BioLabeler | Library of Medicine text extractor and indexing engine | Medical | 2011-12-23 |
| bioNMF | Biomedical data analysis service | Medical | 2012-02-24 |
| Biosemantics ACCCA | Clinical record annotation service | Medical | 2013-02-08 |
| Boliven | Online repository of medical device information | Medical | 2012-01-19 |
| Byomei | Japanese disease reference service | Medical | 2012-10-16 |
| CDC Wonder | CDC health statistics retrieval service | Medical | 2011-11-15 |
| Centers for Disease Control (CDC) Content Syndication | Health information content syndication service | Medical | 2011-09-27 |
| ChemBank | Biomedical database | Medical | 2012-01-17 |
| China Cancer Database | Chinese cancer information database | Medical | 2012-01-31 |

 MASHERY The Premier API Management Solution

 MuleSoft Connect APIs Learn more

Beyond APIs: Mobile-Enabling Developer Ecosystems FORRESTER On-Demand Webinar 

FREE API MANAGEMENT SOLUTION BY 3SCALE SIGNUP NOW! 

Extend your Enterprise with IBM API Management  Download the eKit

  api phany API Management and Delivery Platform data loader.io #1 data loader for Salesforce START AT \$50/MONTH Get Started

Become a ProgrammableWeb Sponsor

Extend your Enterprise with IBM API Management  Download the eKit

<http://programmableweb.com/>

Good Examples

- NLM/NCBI/PubChem - <http://www.ncbi.nlm.nih.gov/>
- EBI (inc ChEMBL, ChEBI etc) – <http://ebi.ac.uk>
- NCI Cactus - <http://cactus.nci.nih.gov/>
- ChemSpider - <http://www.chemspider.com/>
- ChemAxon – e.g. Chemicalize
<http://www.chemicalize.org/>
- OpenTox -
<http://www.opentox.org/dev/apis/api-1.2/structure>
- Altmetric & ImpactStory
- .. Many more web services out there...

Free

VS

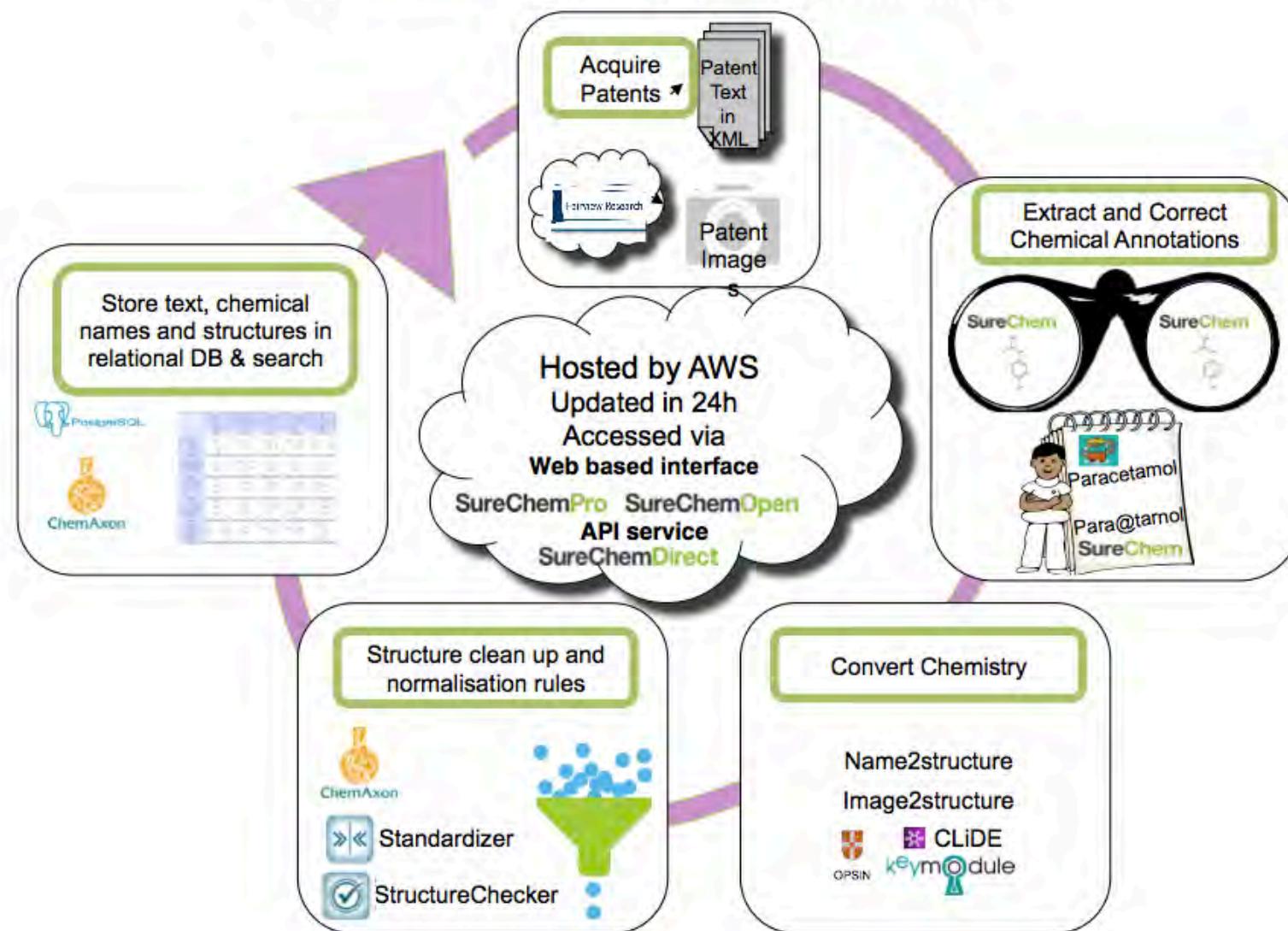
Accessible

Case Studies

1. SureChem

Patent Chemistry

SureChem Chemistry Mining Workflow



SureChemOpen

- ◆ **Free** patent chemistry searching for **all**
- ◆ **Early** access to newly published structures
- ◆ **Easy navigation** through the relevant chemistry in the full text of patent documents
- ◆ Search, View, Print
- ◆ Links to ChemSpider and RSC journals
- ◆ A free viewer for SureChemDirect users

Secured by 128-bit encryption providing same level of security as online banking and other proprietary chemistry searching tools!

SureChemDirect

API

- ◆ Direct access **to chemistry & full text** patent data
- ◆ Data and search hosted on SureChem AWS instance
 - ◆ Dedicated AWS instances available
 - ◆ Secure https line, 128-bit encryption
- ◆ **Next day indexing** of chemistry
- ◆ **Integration** with internal databases and workflow tools (SharePoint, Seurat, SpotFire etc)
- ◆ **Rapid deployment** using **Pipeline Pilot package**
- ◆ Easy manipulation of large volumes of patent chemistry
- ◆ **Full support and documentation**
 - ◆ <http://docs.surechem.com/direct/>

Data Feed

- ◆ Direct access to **chemistry & patent metadata in csv** format
- ◆ Access via secure ftp
- ◆ Batch screening & analysis behind your firewall
- ◆ Daily updates



The image shows a screenshot of the SureChemDirect API documentation. At the top, there are three numbered icons representing different query types: 1. SureChem Query from Text, 2. SureChem Query from Molecule, and 3. SureChem Matrix Query from Text. Below these, the 'SureChem' section contains links for 'View PipelinePilot user guide', 'View SureChemOpen documentation', and 'Submit support ticket'. A 'GitHub' link is also present. The 'Getting Started' section includes an 'Overview' and a 'Normal / JSON response format' section. The 'SureChemDirect API' section has an 'Overview' and a detailed 'Getting Started' section that explains the RESTful web service architecture and supported HTTP verbs.

2. SciBite

Drug Discovery, Right Now.

Therapeutic Intelligence

I care about asthma.

What does the current landscape of (Drugs, Targets, Companies, Trials, Collaborations...) look like?

Where does TI come from?

The image is a collage of screenshots from various websites and platforms, illustrating the sources of data for the TI system. The sources include:

- NCBI: Medical treat ha EP 12621
- EU-CTR: Welcome
- ClinicalTrials.gov: Search for REGISTER
- EudraCT: Examples: Click here
- Nature: A service of CO
- WSJ: Home | News & Co
- EMBL-EBI: The European Bioinformatics Institute, Part of the European Molecular Biology Laboratory
- EMBL-EBI: Explore the EBI: Press releases, Mapping metabolism, Something different
- EMBL-EBI: Press releases
- EMBL-EBI: Mapping metabolism
- EMBL-EBI: Something different
- EMBL-EBI: Data storage in DNA becomes a reality

Twitter-like Newsfeed

SciBite³. Drug Discovery, Now.

Live Pharma/Biotech News. From Molecule To Merger.

Latest News / GlaxoSmithKline Newsletter

Latest articles for **GlaxoSmithKline**, as of Thu Jan 31. Use the toolbox on the right to get updates and filter these articles to find the companies, targets, indications and drugs most interconnected.

Coloured tags show you key topics also mentioned in the article. Click a tag to see its own newsletter, or the '+' link to create a newsletter that specifically mentions the two topics together.

[All!](#) [Popular](#) [News](#) [Literature](#) [Trials](#) [Patents](#) [Grants](#)

Trpv4 Antagonists (Glaxosmithkline Llc)
Patent #EP2549872 (GLAXOSMITHKLINE LLC) ([Today](#))
transient receptor potential c... [GlaxoSmithKline](#)
Patent: [Expand](#) | [View \[This Window\]](#) | [\[New Window\]](#)

Vertex Sales Lag As Hep C Patients Pass On Incivek
FiercePharma ([Today](#))
Hepatitis C Cystic Fibrosis Ribofluranosyl Carboxamide VX-950 PSI 7977 Ivacaftor Gilead Vertex GlaxoSmithKline Abbott Laboratories interferon New Drug
Web/News Item: [Expand](#) | [View \[This Window\]](#) | [\[New Window\]](#)

Boehringer Gets Fda Panel's Thumbs Up On 'modest' New Copd Treatment
FierceBiotech ([1 day ago](#))
Lung Diseases Chronic Obstructive Pulmonary ... Olodaterol Boehringer-Ingelheim GlaxoSmithKline Novartis Theravance glycogen synthase kinase Pharmacovigilance
Web/News Item: [Expand](#) | [View \[This Window\]](#) | [\[New Window\]](#)

A Dose Escalation Study Of Omp-52M51 In Subjects With Solid Tumors
ClinicalTrials.gov Trial Updates Record#NCT01778439 [Sponsor:OncoMed P... ([1 day ago](#))
GlaxoSmithKline OMP-52M51
Clinical Trial: [Expand](#) | [View \[This Window\]](#) | [\[New Window\]](#) | [Read by 3](#) [selleckchem.com/users](#)

The Purpose Of This Study Is To Evaluate The Spirometric Effect (Trough Fev1) Of Umeclidinium/Vilanterol 62.5/25 Mcg Once Daily Compared With Tiotropium 18 Mcg Once Daily

Articles per day [last 90 days]

Analysis Tools

Daily Updates

Everything (except grants) [RSS] [eMail]
 NIH Grants only [RSS]

[Need Help With RSS?]

About This Topic

Quick Help

Geographical View

Connected Topics

Filters & Analysis

Kinase Inhibitors
[www.selleckchem.com](#)
Novel & Potent Kinase Inhibitors Oncology & Cell

Operations

SciBite Main API Methods (v3.1)

EntityBooleanQueryService: Get News For One Or More Entities

/EntityBooleanQueryService [GET](#)

EntityNewsService: Get recent news articles for an input entity

/EntityNewsService [GET](#)

EntitySearchService: Find ID for an entity

/EntitySearchService [GET](#)

EntityWebLinksService: Get HTML Links for interesting sites

/EntityWebLinksService [GET](#)

EntityTextToNewsService: Entity news by text query

/EntityTextToNewsService [GET](#)

DocumentToNewsService: Get news for entities in a document

/DocumentToNewsService [GET](#)

DocumentEntitiesService: Get entities in a document

/DocumentEntitiesService [GET](#)

ListEntitiesService: List all entities

/ListEntitiesService [GET](#)

TermiteService: Run termite NER on a piece of text

/TermiteService [GET](#)

InsightLinksService: Get entities most correlated with an input entity

/InsightLinksService [GET](#)

InsightTopTypesService: Top entities in the news

/InsightTopTypesService [GET](#)

InsightJointInSourceService: Co-occurring entities in a source

/InsightJointInSourceService [GET](#)

InsightLitSpaceService: Co-occurring entities in historical medline abstracts

/InsightLitSpaceService [GET](#)

InsightLitSpacePairService: Co-occurring entity-entity data in historical medline abstracts

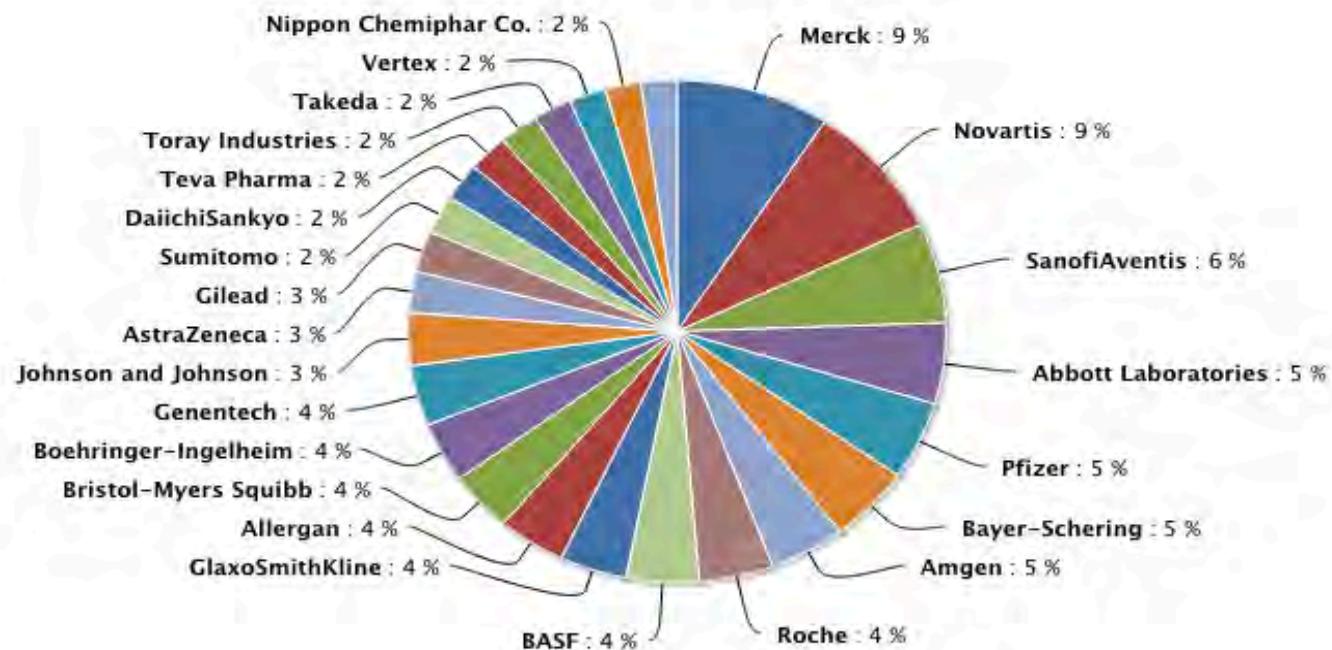
/InsightLitSpacePairService [GET](#)

Slice & Dice With The API

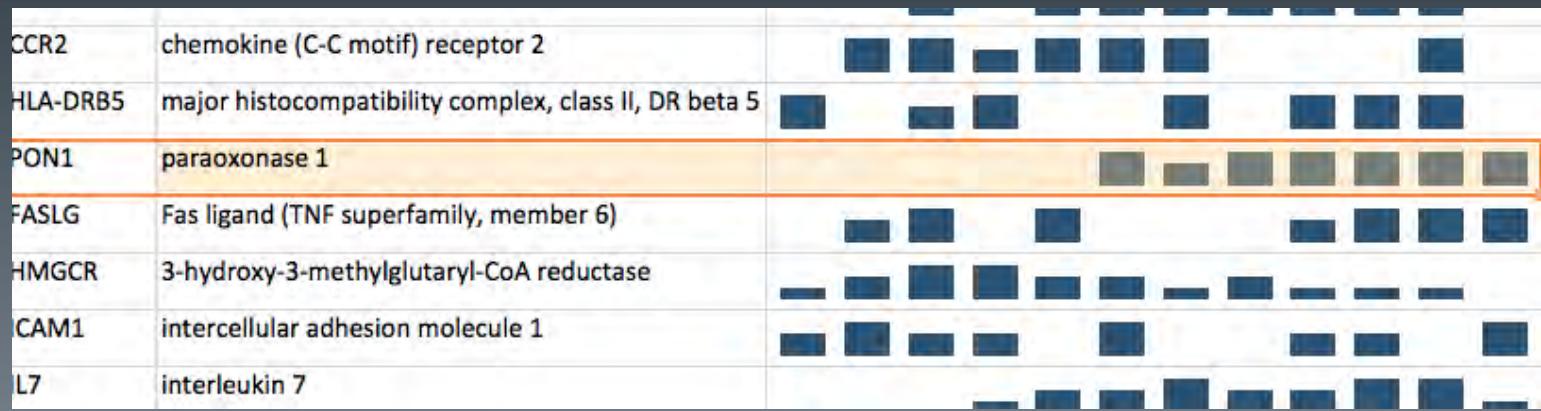
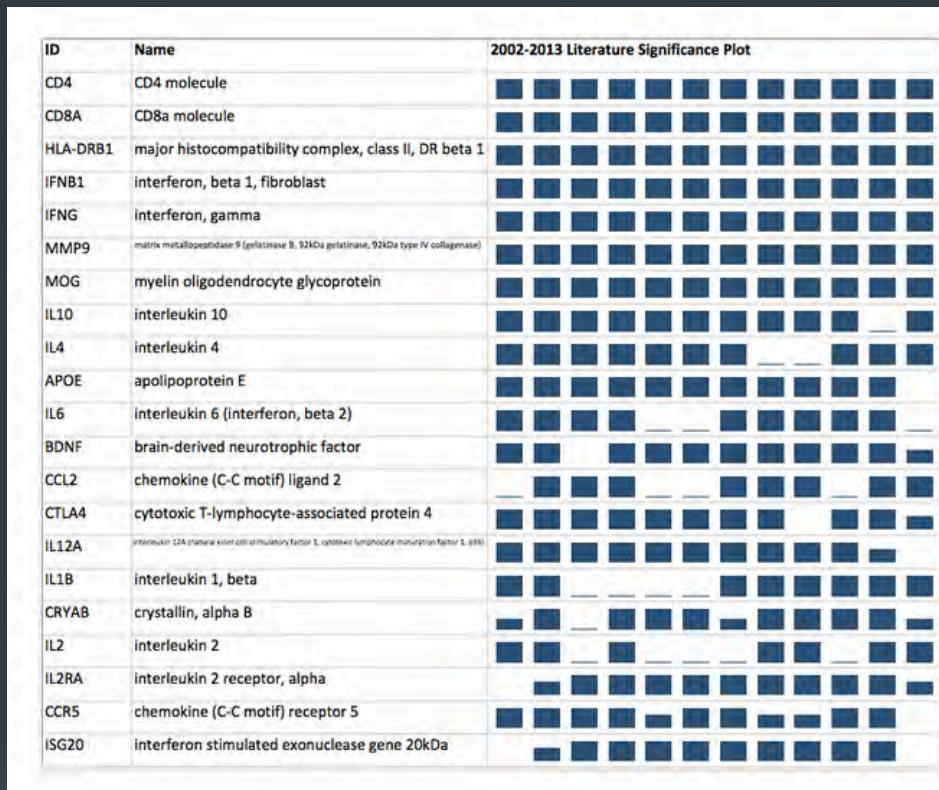


Which companies are filing the most patents?

Patents By Company (Last 6 Months)



Hot Target Barcodes



Indication Opportunity Map

Indications →

Organisations →



3. Open PHACTS

Navigating Pharmacological Space

Open PHACTS Project Partners

Pfizer Limited – Coordinator

Universität Wien – Managing entity

Technical University of Denmark

University of Hamburg, Center for Bioinformatics

BioSolveIT GmbH

Consorti 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

It *Should* be easy to ask...

"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

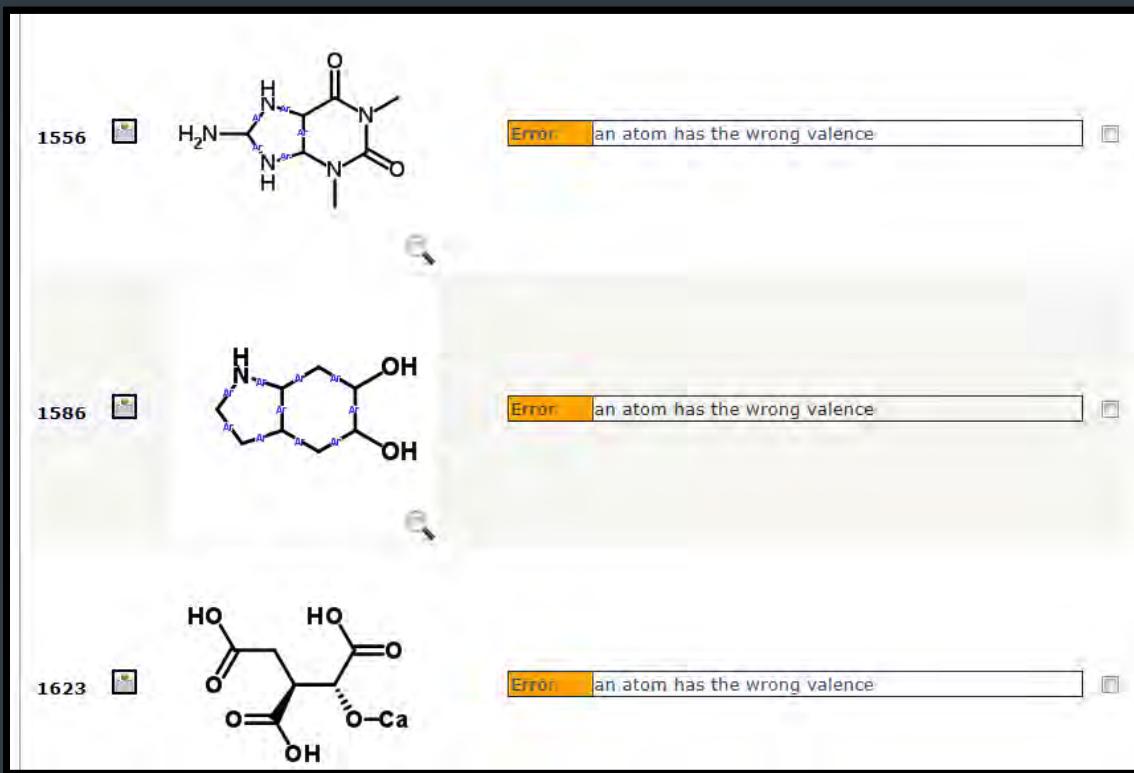
Business Question Driven Approach

| 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 list of compounds defined by their chemical structure (with options to match stereochemistry or not). |
| 41 | 13 | 8 | A project is considering Protein Kinase C Alpha (PRKCA) as a target. What are all the compounds known to modulate the target directly? What are the compounds that may modulate the target directly? i.e. return all cmpds active in assays where the resolution is at least at the level of the target family (i.e. PKC) both from structured assay databases and the literature. |
| 44 | 13 | 8 | Give me all active compounds on a given target with the relevant assay data |
| 46 | 13 | 8 | Give me the compound(s) which hit most specifically the multiple targets in a given pathway (disease) |
| 59 | 14 | 8 | Identify all known protein-protein interaction inhibitors |

... paper coming very soon in DDT



ChemSpider
The free chemical database



- Chemical Registration System
 - Utilizes ChemSpider Validation and Standardization platform including collapsing tautomers
 - Utilizes FDA rule set as basis for standardization (GSK lead)
 - All molecules get a CVSP report – helps us biologists!

dev.openphacts.org

OpenPHACTS API

Chemical Structure Exact Search

InchiKey to URL

Inchi to URL

Chemical Structure Similarity Search

SMILES to URL

Chemical Structure Substructure Search

Get concept description

Map free text to a concept URL based on semantic tag

Map free text to a concept URL

Map URL

Get ChEBI Ontology Class Members

Get ChEBI Ontology Root Classes

Get ChEBI Ontology Class

ChEBI Class Pharmacology Count

/comp

ChEBI Class Pharmacology Complete (DEPRECATED)

ChEBI Ontology Class Pharmacology Paginated

/comp

Compound Information

Compound Pharmacology Count

Compound Pharmacology Complete (DEPRECATED)

Compound Pharmacology Paginated

Data Sources

Get Enzyme Classification Root Classes

Get Enzyme Classification Class

Get Enzyme Classification Class Members

Enzyme Pharmacology Count

Enzyme Pharmacology Complete (DEPRECATED)

Enzyme Pharmacology Paginated

Pathway Information

Activity types

Units for activity type

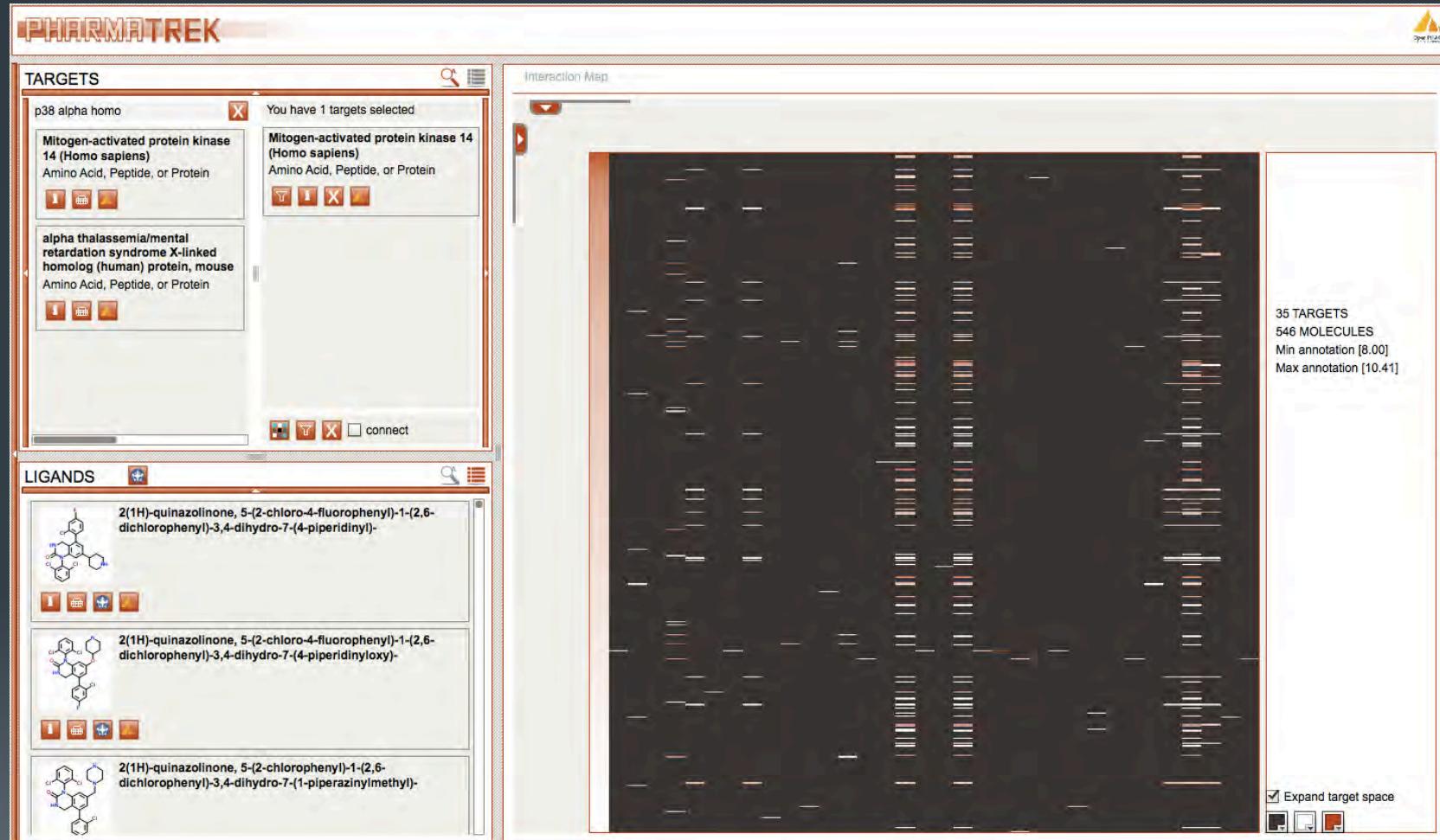
/pha

Target Information

Target Pharmacology Count

Target Pharmacology Complete (DEPRECATED)

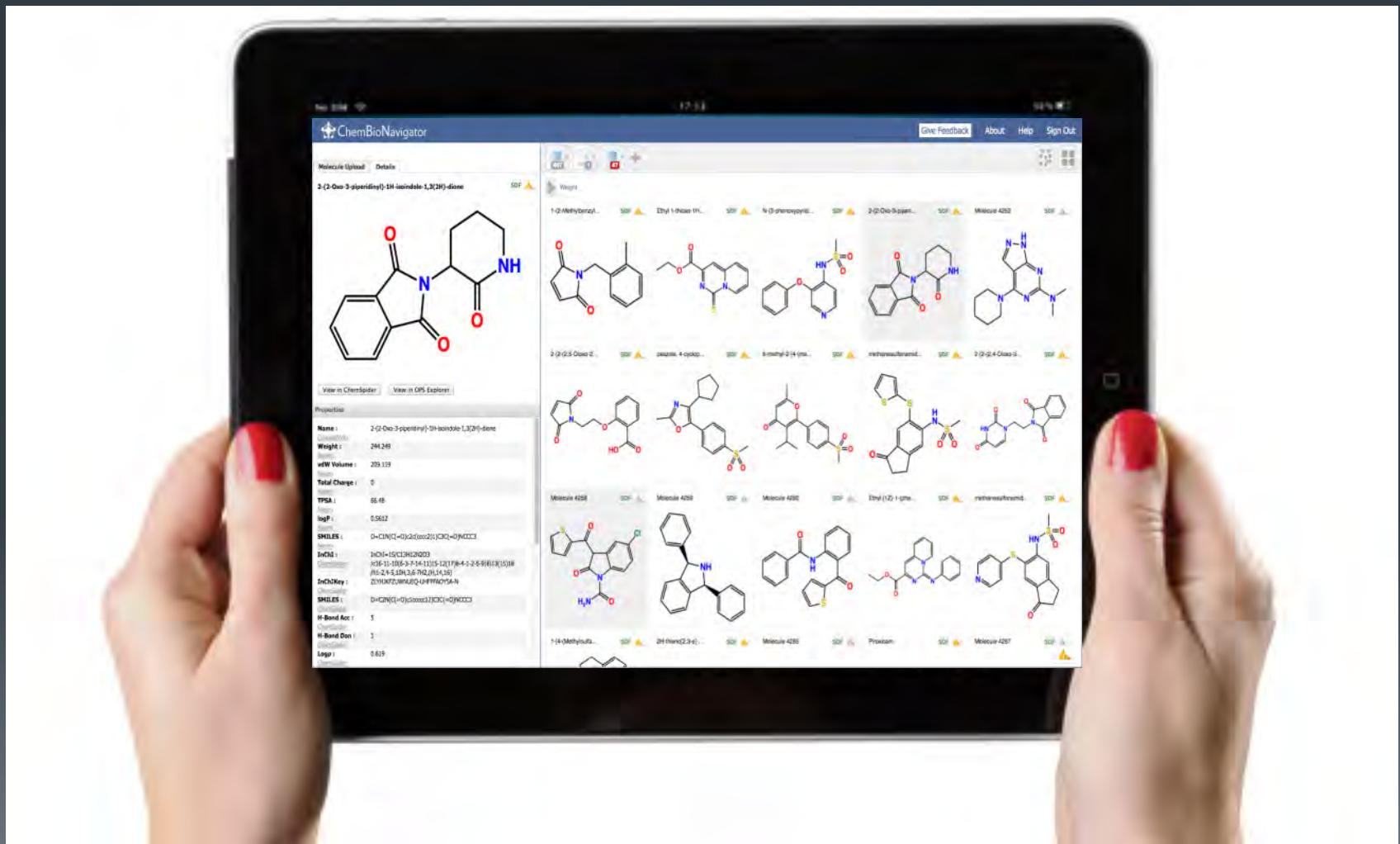
PharmaTrek



<http://pharmatrek.org>

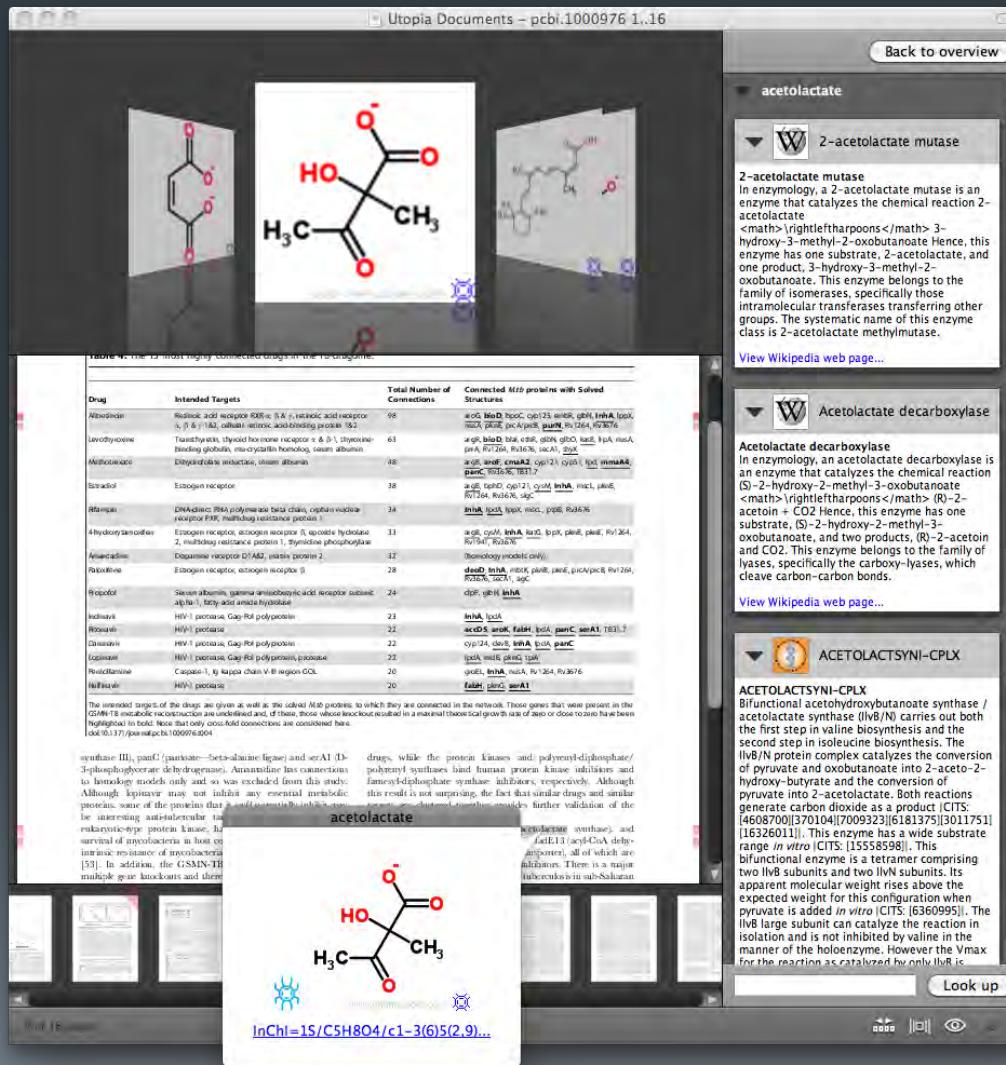
The ChemBio Navigator

<http://chembionavigator.org/>



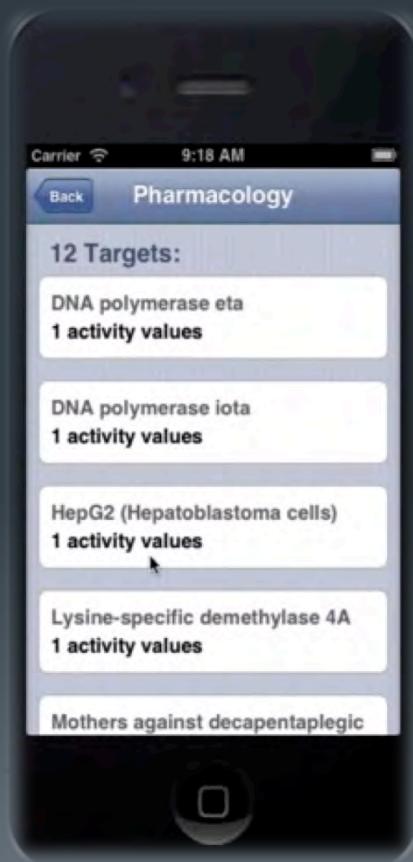
Utopia

<http://getutopia.org>



Hacking The iPhone

<http://www.youtube.com/watch?v=0aGB6YqtuQ0>



Workflow Tools

The image displays two workflow tools: KNIME and Taverna.

KNIME (Top): A screenshot of the KNIME interface. On the left, the project tree shows 'KNIME_project' with various nodes like 'File Reader', 'Java Snippet', and 'Interactive Table'. The main workspace shows a workflow: 'File Reader' → 'Java Snippet' → 'Get Name and Inchi' → 'Get Activity' → 'Activity Parser' → 'Column Filter' → 'Interactive Table'. Below the workflow, descriptive text and icons explain each step. An 'Interactive Table' view is open on the right, showing a table with columns: Name, Inchi, Activity, Units, Relation, and Target. The table contains data for Sorafenib, including its chemical structure and various biological targets.

| 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 |

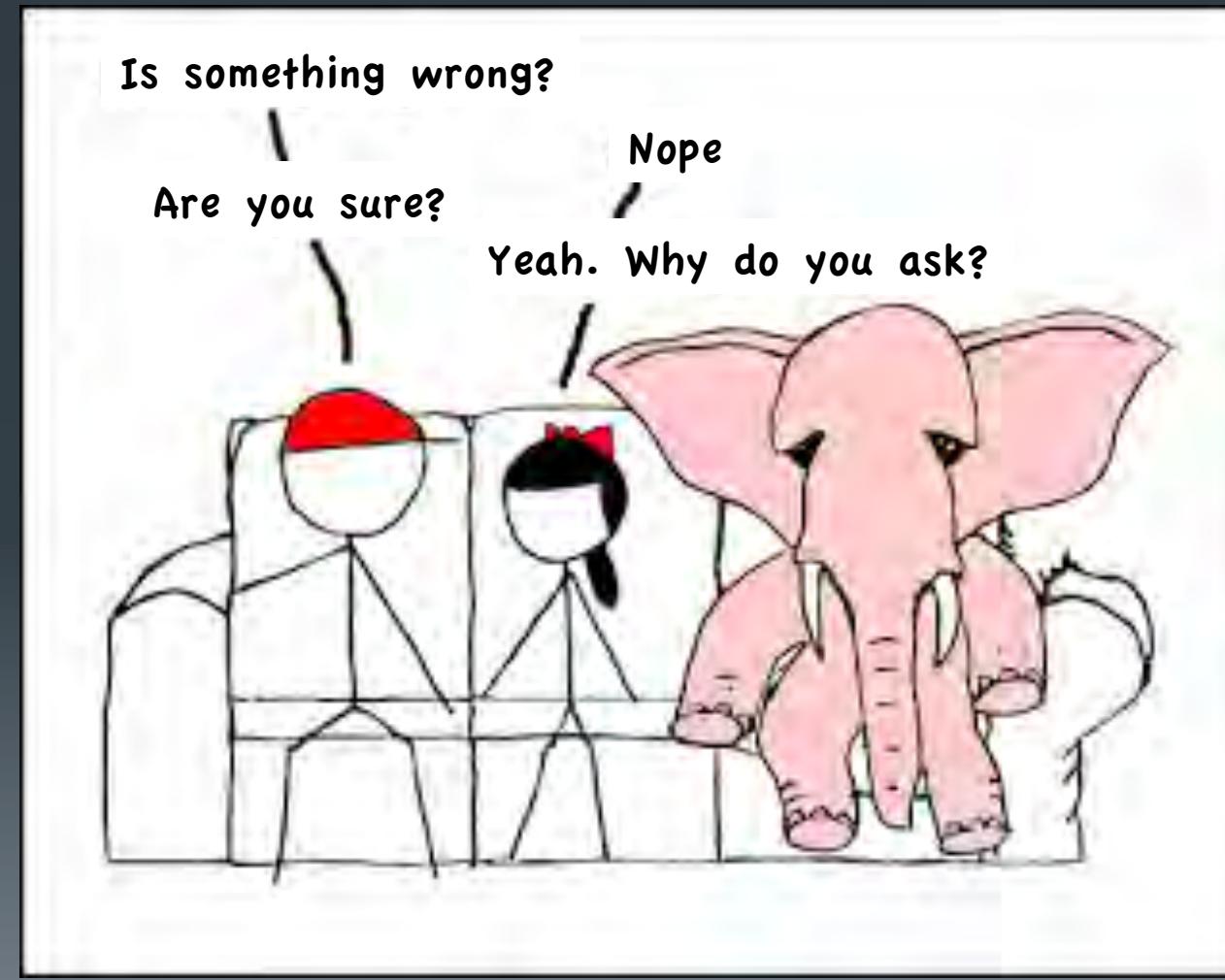
Taverna (Bottom): A screenshot of the Taverna workflow. The top part shows a high-level overview of the workflow steps: 1. Substructure search and retrieve chemspider link, 2. Map compound chemspider link to conceptwiki uri, and 3. Obtain compound pharmacology information. The bottom part shows the detailed workflow nodes: substructure input, Substructure search, map chemspider link to con..., compound pharmacology, Join Data from Cache, Keep Properties, and HTML Table Viewer. The 'map chemspider link to con...' node is highlighted with a yellow box and a descriptive text box: '3. Obtain compound pharmacology information. Join with data(enzyme class) to get bioactivity data for serine protease only'.

+ Google Refine, Helium For Excel, Taverna

Anyone can now ask...

- Who's working on what targets for what indications?
- What chemical matter has been tested against those targets?
- What does that chemical space look like?
- What's been patented in that area?
- What other targets do those compounds hit?
- Who's working on those targets for what indications...?

Security?



Security: Still A Concern For Some

- But more companies moving to cloud externalisation
- Starting to see emergence of security principles (Pistoia, Open PHACTS)
- VM's (local, remote)
- Clever workflows
- Will opportunity overtake fear?

Conclusions

- Accessible API's increasing
- Let people explore, they'll find new use cases
- Can't build a big interface on big data
- Usability, Discovery & Security key challenges

Acknowledgements

- SureChem Limited: <http://surechem.com>
- SciBite Limited: <http://scibite.com>
- Open PHACTS Consortium: <http://openphacts.org> |
<http://dev.openphacts.org> | <http://explorer.openphacts.org>