PubChem:
An Information Resource
Linking Chemistry and Biology

Evan E. Bolton, Ph.D.

ACS National Meeting
San Francisco, CA
September 10, 2006
The Molecular Libraries Roadmap: An Integrated Initiative

- Instrumentation
- Assay Development
- Chemical Diversity
- Predictive ADMET

Technology Development

Screening

Molecular Libraries Screening Centers Network (MLSCN)

Informatics

Cheminformatics Research Centers

Compound Repository (MLSMR)

PubChem
PubChem Goals ...

... Archive molecular structure and bioassay data from the Molecular Libraries Screening Center Network

... Provide search, retrieval and data analysis tools to optimize utility of these results
Further optimize research utility by including other public sources of chemical structure and bioactivity information and by integration with other NIH Biomedical information resources whenever possible.
PubChem Goals ...

... An on-line resource providing comprehensive information on the biological activities of small molecules

... Accessible to molecular biologists

... Useful to computational / medicinal chemists and other specialists
PubChem Approach ...

... “GenBank model”

... direct depositions by investigators

... highly automated (low database cost)

... 25 year precedents in biology

... less precedent in chemistry
PubChem Contents …

... Contributed substance records

... with chemical structure

... chemical names and comments

... links to contributor web sites

... contributed links to other NCBI biomedical databases
PubChem Contents ...

... Contributed bioassay records

... with assay description / protocol

... links to tested substances

... summary and detailed test results

... links to contributor web sites and other NCBI databases
PubChem Contents ...

... 279 Bioassays Contributed

... 691,387 Bioassay Test Results

... 12,793,312 Substances Contributed

... 7,995,947 Unique Compound Structures

... 54 Depositing Organizations
Growth in PubChem Substances

Date

Count

Compound

Substance

May-05, Jun-05, Jul-05, Aug-05, Sep-05, Oct-05, Nov-05, Dec-05, Jan-06, Feb-06, Mar-06, Apr-06, May-06, Jun-06, Jul-06
Growth in PubChem Users per Day
PubChem Retrieval System …

… Optimize “discoverability” for molecular biologists by integrating PubChem into NCBI’s Entrez / PubMed Search Engine

… Chemical structure search

… Bioassay result search

… Exploratory structure-activity tools
NCBI’s Entrez Search Engine ...
Entrez Links and Neighbors ...

1,500,000 users ...
50,000,000 hits ...
... per day
Search for “Kaempferol” ...
Compounds in PubChem ...
Compound in PubChem ...

CID 576 — PubChem Compound Summary — Microsoft Internet Explorer

Compound Summary:

CID: 576
Substances: 14 Links
All: 14 Links
Same: 12 Links
Mixture: 2 Links

BioActivity: 7 Links
PubMed: 15 Links
Protein Structure: 1 Link
NLM Toxicology: Link
Similar Compounds: 73 Links
Structure Search
High-throughput assays for promiscuous inhibitors

Nature Chemical Biology, 1, 146-148 (2005)
doi: 10.1038/nchembio718

High-throughput assays for promiscuous inhibitors

Brian Y Feng1,2, Anang Shelat1,2, Thompson N Doman2, R Kip Guy1 and Brian K Shoichet1

High-throughput screening (HTS) searches large libraries of chemical compounds for those that can modulate the activity
Link to Protein 3D Structures ...

Crystal Structure Of Quercetin 2,3-Dioxygenase Anaerobically Complexed With The Substrate Kaempferol [mmdbId:21069]
Link to Protein 3D Structure ...
Similar Compounds in PubChem ...

1. CID: 5280863
   - kaempferol, Kaempferol ...
   - IUPAC: 3,5,7-trihydroxy-2-(4-hydroxyphenyl)chromen-4-one
   - MW: 286.236 | MF: C15H10O6

2. CID: 5280343
   - quercetin, Quercetine ...
   - IUPAC: 2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-chromen-4-one
   - MW: 302.236 | MF: C15H10O7

3. CID: 5281672
   - myricetin, Camabiscetin ...
   - IUPAC: 3,5,7-trihydroxy-2-(3,4,5-trihydroxyphenyl)chromen-4-one
   - MW: 318.235 | MF: C15H10O8
A Similar Compound ...
Link to Protein 3D Structures ...
Compare Protein / Ligand Complexes ...
Link to Contributor Site ...
Link to Another Structure ...
Molecular components in the MMDB structure are listed below. The icons indicate macromolecular chains, 3D domains, protein classifications and ligands. Please hold the mouse over each icon for more information on the component.
View Protein Family Alignment ...
Links from Compounds to PubMed ...

Medical Subject Annotations: (Total: 2) Display: Next 1 | All

Quercetin
A flavonol widely distributed in plants. It is an antioxidant, like many other phenolic heterocyclic compounds. Glycosylated forms include RUTIN and quercetin.

PubMed via MeSH Choose by Subheadings:
- administration and dosage
- analysis
- blood
- diagnostic use
- isolation and purification
- pharmacology
- secretion
- toxicity
- adverse effects
- antagonists and inhibitors
- chemical synthesis
- genetics
- metabolism
- physiology
- standards
- urine

Depositor-Supplied Synonyms: (Total: 74)
Links from Compounds to PubMed ...
Link to BioAssays where “Active” ...

<table>
<thead>
<tr>
<th>No.</th>
<th>AID</th>
<th>Assay Description</th>
<th>Source</th>
<th>Total substances tested:</th>
<th>Active:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>396</td>
<td>Compound Screen Assay, Human YWHAB</td>
<td>SGCOxCompounds</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>395</td>
<td>Compound Screen Assay, Human STK16</td>
<td>SGCOxCompounds</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>393</td>
<td>Compound Screen Assay, Human PIM1</td>
<td>SGCOxCompounds</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>4</td>
<td>386</td>
<td>Compound Screen Assay, Human DIRAS</td>
<td>SGCOxCompounds</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>383</td>
<td>Compound Screen Assay, Human CLK3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A BioAssay where “Active” ...
A BioAssay where “Active” ...

This is a cell-free, enzymatic assay for inhibition of HIV-2 ribonuclease H (RNase H) activity. This assay has typically been run in dose-response format as a secondary assay to RNase H.

The substrate is an 18 nucleotide RNA/DNA duplex labeled with 6-FAM (fluorescein) on the 3’ end of the RNA and DABCYL on the 5’ end of the DNA. As a duplex, the fluorescein fluorescence is quenched by the close proximity of DABCYL. After HIV-2 RNase H is added to the reaction the enzyme cleaves the RNA four nucleotides from the 3’ end of the RNA. This RNA fragment dissociates into solution and the fluorescence of the label is no longer quenched. The reaction is terminated by the addition of EDTA which removes the required Mg²⁺ from solution. Fluorescence is an indication of enzymatic activity and reduced
A BioAssay where "Active" ...
Entrez Links and Neighbors ...

1,500,000 users ... 50,000,000 hits ... ... per day
Optimize “discoverability” for molecular biologists by integrating PubChem into NCBI’s Entrez / PubMed Search Engine

Chemical structure search

Bioassay result search

Exploratory structure-activity tools
Structure Search Tool ...
With Sketcher and Search Options ...
Search Results Pass to Entrez ...
BioAssay Search Tool ...
Select Experimental Results ...

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Value Count</th>
<th>Min::Max</th>
<th>Lower Query</th>
<th>Upper Query</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Submitter</td>
<td>48125</td>
<td></td>
<td></td>
<td></td>
<td>Select one</td>
</tr>
<tr>
<td>2</td>
<td>Submission Date</td>
<td>48125</td>
<td></td>
<td></td>
<td></td>
<td>Select one</td>
</tr>
<tr>
<td>3</td>
<td>Activity Direction</td>
<td>48125</td>
<td></td>
<td></td>
<td></td>
<td>Select one</td>
</tr>
<tr>
<td>4</td>
<td>Activity Qualifier</td>
<td>48125</td>
<td></td>
<td></td>
<td></td>
<td>Select one</td>
</tr>
<tr>
<td>5</td>
<td>Qualified AC50</td>
<td>48125</td>
<td>6.06e-08 :: 0.00769</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Log of AC50</td>
<td>48125</td>
<td>-7.22 :: -2.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Hill Coefficient</td>
<td>6656</td>
<td>-2.97 :: 32.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Curve R2</td>
<td>6656</td>
<td>-0.66 :: 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Data Type</td>
<td>48125</td>
<td></td>
<td></td>
<td></td>
<td>Select one</td>
</tr>
<tr>
<td>10</td>
<td>Compound Type</td>
<td>48125</td>
<td></td>
<td></td>
<td></td>
<td>Select one</td>
</tr>
</tbody>
</table>
For Selected PubChem Compounds ...

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>3.077μM</td>
<td>48125 -100 :: 79.3</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Activity at 6.881μM</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Activity at 15.386μM</td>
<td>48125 -103.7 :: 102.4</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Activity at 0.034mM</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Activity at 0.077mM</td>
<td>48125 -106.5 :: 327.9</td>
<td></td>
</tr>
</tbody>
</table>

Additional Filters:
- Result Filter:
- Substance Filter
## PubChem BioAssay Results ...

<table>
<thead>
<tr>
<th>#</th>
<th>Structure</th>
<th>SID</th>
<th>CID</th>
<th>Score</th>
<th>Outcome</th>
<th>Log of AC50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Structure 1" /></td>
<td>4252010</td>
<td>2232224</td>
<td>52</td>
<td>Active</td>
<td>-5.21</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Structure 2" /></td>
<td>4256225</td>
<td>837064</td>
<td>48</td>
<td>Active</td>
<td>-4.81</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="Structure 3" /></td>
<td>4252054</td>
<td>2231736</td>
<td>21</td>
<td>Inactive</td>
<td>-2.11</td>
</tr>
<tr>
<td>4</td>
<td><img src="image4.png" alt="Structure 4" /></td>
<td>4257232</td>
<td>876802</td>
<td>21</td>
<td>Inactive</td>
<td>-2.11</td>
</tr>
</tbody>
</table>

Selected BioAssay Result Count: 5 out of 48125  Page: 1 of 1
Quercetin Structure-Activity
BioAssay Summary Tool ...
BioAssay Summary Tool ...

<table>
<thead>
<tr>
<th>AID</th>
<th>Active</th>
<th>Inactive</th>
<th>Total Tested</th>
<th>Data</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>382</td>
<td>11</td>
<td></td>
<td>11</td>
<td>Data</td>
<td>Compound Screen Assay, Human CLK1</td>
</tr>
<tr>
<td>395</td>
<td>8</td>
<td></td>
<td>8</td>
<td>Data</td>
<td>Compound Screen Assay, Human STK16</td>
</tr>
<tr>
<td>354</td>
<td>7</td>
<td>6</td>
<td>13</td>
<td>Data</td>
<td>NCTR Estrogen Receptor Binding Database (NCTRER)</td>
</tr>
<tr>
<td>115</td>
<td>5</td>
<td>43</td>
<td>49</td>
<td>Data</td>
<td>NCI human tumor cell line growth inhibition assay. Data for the SR Leukemia cell line</td>
</tr>
<tr>
<td>383</td>
<td>4</td>
<td></td>
<td>4</td>
<td>Data</td>
<td>Compound Screen Assay, Human CLK3</td>
</tr>
<tr>
<td>367</td>
<td>4</td>
<td></td>
<td>4</td>
<td>Data</td>
<td>HIV-2 RNase H Inhibition</td>
</tr>
<tr>
<td>179</td>
<td>4</td>
<td>53</td>
<td>57</td>
<td>Data</td>
<td>NCI AIDS Antiviral Assay</td>
</tr>
<tr>
<td>125</td>
<td>4</td>
<td>53</td>
<td>57</td>
<td>Data</td>
<td>NCI human tumor cell line growth inhibition assay. Data for the HL-60(TB) Leukemia cell line</td>
</tr>
<tr>
<td>121</td>
<td>4</td>
<td>52</td>
<td>56</td>
<td>Data</td>
<td>NCI human tumor cell line growth inhibition assay. Data for the K-562 Leukemia cell line</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ACE Inhibitors Structure-Activity ...
Structure Clustering Tool ...

66 compounds were collapsed into 23 rows with a Tanimoto score of 0.9.

Display the tree with structures
Members of One Structure Cluster ...
With Confirmed ACE Inhibitor ...
MKP-1 Inhibitors Structure-Activity ...

BioAssay Summary

AID: 442
Name: In vitro MKP-1 Phosphatase Dose Response Confirmation and Secondary Selectivity/Specificity Assay
Data Source: University of Pittsburgh Molecular Library Screening Center (MH-76391 MLSCN MKP-1 Dose Response Assay)

Test Results: Show Select Plot

Links: 
Compounds: All: 6 Active: 6
Substances: All: 6 Active: 6
Actives for MKP-1 BioAssay ...

CID: 5389004
MLS000071484, SMR000038306
MW: 322.404 | MF: C19H22N4O

CID: 5280880
prostaglandin A2, Medullin ...
IUPAC: (Z)-7-[(1R,2S)-2-[(E,3S)-3-hydroxyoct-1-enyl]-5-oxo-1-cyclopent-3-enyl]hept-5-enoic acid
MW: 334.450 | MF: C20H30O4

CID: 2094474
ZINC02632041, MLS000058168 ...
IUPAC: [2-[4-amino-1-methyl-3-(2-methylpropyl)-2,6-dioxopyrimidine-5-yl]-2-oxo-ethyl] 2-(2-furyl)quinoline-4-carboxylate
MW: 476.481 | MF: C25H24N4O6
Structurally Diverse ...
Active in other BioAssays ...

![Image of a table showing various BioAssay details]

- Total BioAssays: 34
- Total Pages: 2

<table>
<thead>
<tr>
<th>AID</th>
<th>Active</th>
<th>Inactive</th>
<th>Total Tested</th>
<th>Data</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>443</td>
<td>6</td>
<td></td>
<td>6</td>
<td>Data</td>
<td>Cdc25B Catalytic Domain Protein Tyrosine Phosphatase Dose Response Confirmation and Secondary Selectivity/Specificity Assay</td>
</tr>
<tr>
<td>442</td>
<td>6</td>
<td></td>
<td>6</td>
<td>Data</td>
<td>In vitro MKP-1 Phosphatase Dose Response Confirmation and Secondary Selectivity/Specificity Assay</td>
</tr>
<tr>
<td>374</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>Data</td>
<td>In vitro Primary HTS Assay for MKP-1</td>
</tr>
<tr>
<td>431</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>Data</td>
<td>Fluorescent HTS Cytotoxicity/Cell viability assay (HPDE-C7K cells)</td>
</tr>
<tr>
<td>430</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>Data</td>
<td>Fluorescent HTS Cytotoxicity/Cell viability assay (HPDE-C7 cells)</td>
</tr>
<tr>
<td>425</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>Data</td>
<td>MKP-3 in vitro HTS assay</td>
</tr>
<tr>
<td>429</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>Data</td>
<td>HTS for Tumor Hsp90 Inhibitors</td>
</tr>
<tr>
<td>428</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>Data</td>
<td>Measurement of GPCR-mediated thallium flux through GIRK channels</td>
</tr>
<tr>
<td>412</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>Data</td>
<td>Discovery of Novel Allosteric Agonists of the M4</td>
</tr>
</tbody>
</table>
Expand to Similar Structures ...
Structure Clustering Tool...
One Structure Cluster ...
Mostly Active in Related BioAssays ...
Though not the Expected Mechanism ...

The oxidative mechanism of action of ortho-quinone inhibitors of protein-tyrosine phosphatase alpha is mediated by hydrogen peroxide.


Drug Discovery, SUGEN Inc., 230 East Grand Avenue, South San Francisco, CA 94080, USA. michael.bova@elan.com

Here, we report the identification and characterization of five ortho-quinone inhibitors of PTPalpha. We observed that the potency of these inhibitors is strongly affected by the presence of hydrogen peroxide in the reaction mixture.
Other Analysis Tools Coming ...