ChemSpider

Building a Structure Centric Community for Chemists

Antony Williams

CINF: Chemometrics: From Teaching to Research; April 5th 2008, ACS New Orleans
In terms of informatics, what do students need to learn/understand?

- The value of managing data and information
- How to capture, manage and share their own data/information
  - Text, charts, structures, data, “models”
- How to “search for data/information” in all its flavors
  - Local, organizational and worldwide searching
  - For chemistry specifically – “chemical structures”
  - Sorting the “wheat from the chaff” and validating online information
  - And, and, and…
How might a chemist use the web?

- Searching for “information” about a compound
  - Articles
  - Patents
  - Safety/tox data
  - Analytical data
  - and...

- Text-based searching is limited in scope – too many identifiers, too broad

- Chemists ask for a “Structure Searchable” web as a path to information...
Free access website for chemists to research structure based information
- Structure/substructure searches
- Text-based searches
- Prediction of properties
- Web service-based integration

Platform for deposition, curation, integration of data
- Structures, analytical data, annotations, links to resources
- Annotation and curation of data in real-time

A platform to assist discovery?
Example Search 1

- Is there any information about "Quesnoin"?

OR...

- Type in the name (and there may be many) or other identifier
- Paste the InChI String, InChIKey or SMILES
- Draw the structure
### Example Search 1

**INHERENT PROPERTIES, IDENTIFIERS AND REFERENCES**

- **ChemSpider ID:** 21105581
- **Empirical Formula:** C_{20}H_{30}O_{4}
- **Molecular Weight:** 334.4498
- **Nominal Mass:** 334 Da
- **Average Mass:** 334.4498 Da
- **Monoisotopic Mass:** 334.214409 Da

**Systematic Name (OpenEye):**

- SMILES:
  ```
  OC(=O)[C@@]5(C)CCC[O@@]1(C)[C@@H]5CC[O@@]24C[C@@H]3C[C@@O][O](CC[C@@H]12)[C@@H]4O3
  ```
- **InChI:** InChI=1/C20H30O4/c1-17-6-3-7-18(2,16(21)22)13(17)4-8-19-10-12-11-20(23,15(19)24-12)9-5-14(17)19/h12-15,23H,3-11H2,1-2H3,(H,21,22)/t12-,13-,14-,15+,17+,18-,19+,20-/m0/s1
- **InChIKey:** KVZXTIZQSVUTI-8FJL7GHCBZ

**DATA SOURCE(S)**

<table>
<thead>
<tr>
<th>Data Source</th>
<th>External ID(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antony_Williams</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**NAMES AND SYNONYMS**

Legend: Validated by Experts, **Validated by Users**, Non-Validated, Removed by Users, Redirected by Users, Redirect Approved by Experts

- **Quesnoin**
  - (1R,4S,5S,9S,10S,13S,15S,17R)-13-Hydroxy-5,9-dimethyl-16-oxapentacyclo[13.2.1.0^3,10^7,0^4,9,0^5,13,17^7]octadecane-5-carboxylic acid
  - 4H-2,3b-methanophenanthro[1,2-b]furan-6-carboxylic acid, tetradecahydro-11a-hydroxy-6,9a-dimethyl-, (2S,3aR,3bR,5aS,6S,9aS,9bS,11aS)-
### Example Search 1

#### Predicted Properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogP</td>
<td>ACD/LogP: 3.03</td>
</tr>
<tr>
<td>ACD/LogD (pH 5.5)</td>
<td>2.1</td>
</tr>
<tr>
<td>ACD/BCF (pH 5.5)</td>
<td></td>
</tr>
<tr>
<td>ACD/KOC (pH 5.5)</td>
<td></td>
</tr>
<tr>
<td>#H bond acceptors</td>
<td>4</td>
</tr>
<tr>
<td>#Ffreely Rotating Bonds</td>
<td>2</td>
</tr>
<tr>
<td>Index of Refraction</td>
<td>1.587</td>
</tr>
<tr>
<td>Molar Volume</td>
<td>265.4 cm³</td>
</tr>
<tr>
<td>Surface Tension</td>
<td>54.1 dyne/cm</td>
</tr>
<tr>
<td>Flash Point</td>
<td>175.1 Celsius</td>
</tr>
<tr>
<td>Boiling Point</td>
<td>497.4 Celsius at 760 mmHg</td>
</tr>
<tr>
<td>#H bond donors</td>
<td></td>
</tr>
<tr>
<td># of Rule of 5 Violations</td>
<td>0</td>
</tr>
<tr>
<td>ACD/LogD (pH 7.4)</td>
<td>0.3</td>
</tr>
<tr>
<td>ACD/BCF (pH 7.4)</td>
<td></td>
</tr>
<tr>
<td>ACD/KOC (pH 7.4)</td>
<td></td>
</tr>
<tr>
<td>Polar Surface Area</td>
<td>66.76 Å²</td>
</tr>
<tr>
<td>Molar Refractivity</td>
<td>89.31 cm³</td>
</tr>
<tr>
<td>Polarizability</td>
<td>35.4 $10^{-24}$cm³</td>
</tr>
<tr>
<td>Density</td>
<td>1.25 g/cm³</td>
</tr>
<tr>
<td>Enthalpy of Vaporization</td>
<td>88.14 kJ/mol</td>
</tr>
<tr>
<td>Vapour Pressure</td>
<td>5.62E-12 mmHg at 25 Celsius</td>
</tr>
</tbody>
</table>

#### Supplemental Information

**Description**

Quesnoin, a novel unique pure organic compound, was isolated from amber discovered in the Oise River area of the Paris basin (France) and dated at 55 million years old. 1H and 13C NMR indicated an unknown diterpene skeleton, quesnane. The absolute configurations of the eight chiral centers of quesnoin were determined to be 4S, 5S, 8R, 9S, 10S, 13S, 14R, and 16S. The work indicated that the climate of the Paris basin might have been tropical in the early Eocene period, 55 million years ago.

**Links & References**


[DOI: 10.1021/jo701544k]
Example Search 2

- What compounds have a mass of 300+/\pm 0.001?

- or search a combination of intrinsic/predicted properties
### Example Search 2

132 hits found in 2.34 seconds
Nominal Mass >= 299 AND Nominal Mass <= 301 AND Monoisotopic Mass >= 299.999 and Monoisotopic Mass <= 300.001 AND SingleComponent AND NonIsotopic

<table>
<thead>
<tr>
<th>ID</th>
<th>Structure</th>
<th>Empirical Formula</th>
<th>Molecular Weight</th>
<th>Monoisotopic Mass, Da</th>
<th>LogP</th>
<th>ACD/LogD (pH 5.5)</th>
<th>ACD/LogD (pH 7.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>207470</td>
<td><img src="image1" alt="Structure" /></td>
<td>C₁₂H₁₃BrO₄</td>
<td>301.1332</td>
<td>299.999714</td>
<td></td>
<td>ACD/LogP: 3.65</td>
<td>XLogP: 2.50</td>
</tr>
<tr>
<td>159022</td>
<td><img src="image2" alt="Structure" /></td>
<td>C₁₂H₁₃BrO₄</td>
<td>301.1332</td>
<td>299.999714</td>
<td></td>
<td>ACD/LogP: 4.10</td>
<td>XLogP: 3.50</td>
</tr>
<tr>
<td>159020</td>
<td><img src="image3" alt="Structure" /></td>
<td>C₁₂H₁₃BrO₄</td>
<td>301.1332</td>
<td>299.999714</td>
<td></td>
<td>ACD/LogP: 2.96</td>
<td>XLogP: 3.50</td>
</tr>
</tbody>
</table>
Building a Structure Centric Community for Chemists

Complex Search

- **Search by Structure**
  - Exact
  - Substructure

- **Search by Identifier**

- **Search by Elements**

- **Search by Properties**

- **Search by Calculated Properties**

- **Search by Data Source, Data Source Type or Focused Library**

- **Search by LASSO Similarity**

- **Options**
  - Single/Multi-component
    - Search Any
    - Search Single-Component Structures Only
    - Search Multi-Component Structures Only
  - Isotopically Labeled
    - Search Any
    - Search Isotopically Labeled Structures Only
    - Disregard Isotopically Labeled Structures

- **Additional Filters**
  - Filter only those having spectra associated
  - Filter only those having patents link

- **Search Hits Limit**
  - 1000
<table>
<thead>
<tr>
<th>Page 1</th>
<th>Page 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Don J Durzan</td>
<td>Q. Goo, J. Golik</td>
</tr>
<tr>
<td>... - S12 Meeting Abstract Nitric oxide, cell death and increased [Taxol] recovery Durzan Don J 1 <a href="mailto:djdurzan@ucdavis.edu">djdurzan@ucdavis.edu</a> 1Department of Plant Sciences ...</td>
<td>... (UUC) 2'-Carbamate Taxol Acta Crystallographica Section CCrystal Structure Communications010-2701 organic compounds Volume 51 Part 2 Pages 295-298 February 1995 2 ...</td>
</tr>
<tr>
<td>Hemant K Parekh, Mahesh Adikari, Bharath Vennapusa</td>
<td>F. Gueritte-Voegelin, D. Guenard, L. Monguillot, P. Potier, J. Guilhem, M. Cesario, C. Pascard</td>
</tr>
<tr>
<td>... partitioning of cili with the cellular microtubules: a possible mechanism of development of Taxol resistance in human ovarian carcinoma cells Parekh Hemant K 1 hemant.parekh@ ...</td>
<td>... (UUC) Structure of a synthetic Taxol precursor: N-tert-butoxycarbonyl-10-deacetyl-N-debenzoylTaxol Acta Crystallographica Section CCrystal Structure Communications010 ...</td>
</tr>
<tr>
<td>Search Term: taxol</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td></td>
</tr>
</tbody>
</table>

NCBI Entrez [Search]

1650 hits found in 30.14 seconds

"paclitaxel"[MeSH Terms] OR taxol[Title] OR taxol[Abstract]

1 2 3 4 5 6 7 8 9 10...


Shannon KB, Carman JC, Moree CB, Tamburier JS, Salmon ED. **Taxol-stabilized Microtubules Can Position the Cytokinetin Furrow in Mammalian Cells.** Mol Biol Cell. 2005 Sep; 16; 4423-4435.
ChemSpider Data

- The database contains close to 20 million compounds obtained from almost 100 data sources. 11 new data sets are waiting to be deposited – 1.2 million compounds
  - Chemical vendors
  - Publishers
  - Commercial Database Vendors
  - US and international patents
  - Structure aggregators
  - Scraped from websites
  - Deposited by users
The Quality of Data Online...

- Aggregating and connecting to data online opens up quality issues
- Structure-identifier associations are “dirty”
- Manual curation of small databases is enough work – what about millions of structures?
- Structures are far from perfect. What is a “correct structure”?
  - Full stereochemistry?
  - Historical timeline of structure?
  - Who is the authority?
Quality is a Major Issue - Search Butanol

Depositor-Supplied Synonyms: (Total: 107)

n-butanol
1-butanol
Butyl alcohol
n-butyl alcohol
butanol
1-hydroxybutane
Methylolpropane
Propylcarbinol
Propylmethanol
Hernostyp

n-Butanol butanolen
Tetrabutoxytitanium
Tyzor BP
Butanol [French]
Butyl orthotitanate
Tin tetrabutanolate
Tyzor TBT
BuOH
Tetrabutoxyzirconium
Butanolen [Dutch]
Tetrabutyl zirconate

Titanium tetrabutoxide
Titanium tetrabutylate
Butyl titanate (IV)
Zirconium tetrabutoxide
Butyl zirconate (IV)
Titanium, tetrabutoxy-
Tetrabutyl orthotitanate
Butyl alcohol (natural)
FEMA Number 2178
Normal primary butyl alcohol
Zirconic acid butyl ester
TETRABUTYL TITANATE
Manually curating data is a challenge for millions of structures

Regular expression checking can clean up a lot

- Search for Chloride and check molecular formula for Cl
- Check InChI for stereochemistry and remove identifiers

Crowdsourcing means providing a platform to curate, annotate and tag data in real time

Provide levels of curator administration to prevent vandalism
Curators - An Active Community

- Daily crowdsourced curation underway
  - 40 curation emails per day
  - 100 identifiers per day removed, approved or added

Building a Structure Centric Community for Chemists
Multi-level Curation and Approval

Building a Structure Centric Community for Chemists
Anyone can “Post Comments” associated with a structure.

To curate or add more data to a structure we require login to track.
But, when registered and logged in...

- Ability to curate and add to the database
  - Add structures
  - “Clean” structures
  - Add data (spectra, CIFs, images)
  - Add links to other pages (URLs)
  - Add publication details
Adding to the Database - Structure

New Record

Convert from String (Name, SMILES, InChI, etc)


Load from File (MOL or SDF)

Description & Tags (Hide Details)
Description:

Tags:

Identifiers (Hide Details)

Links (Hide Details)
[Add Citation] [Add URL(s)] [Add From PubMed] [Add DOI]
Adding New Text Data

Add Publication

- Title of Paper:
- Main Author(s):
- Journal:
- DOI:
- PubMed ID:
- URL(s):
- Free text/description:

Not all fields below are obligatory but we prefer as much information as possible. Please enter at least the Title of the Paper with the DOI or PMID.

Add Identifier

Please enter identifier(s) one per line

Add URL

Enter one or more link URLs, one link per line.

Building a Structure Centric Community for Chemists
Supplemental Information

Description
Cubane (C8H8) is a synthetic hydrocarbon molecule that consists of eight carbon atoms arranged at the corners of a cube, with one hydrogen atom attached to each carbon molecule. It is one of the Platonic hydrocarbons. Cubane is a solid crystalline substance. The cubane molecule was first synthesized in 1964 by Dr. Philip Eaton, a professor of chemistry at the University of Chicago. Before its synthesis, researchers believed that cubic carbon-based molecules could only exist in theory. It was believed that cubane would be impossible to synthesize because the unusually sharp 90-degree bonding angle of the carbon atoms would be too highly strained and hence unstable. Surprisingly, once formed, cubane is actually quite kinetically stable due to a lack of readily available decomposition paths.

Tags
cubane
prismane

Links & References
Wikipedia Page
Cubane chemistry at Imperial College London
Computational Chemistry Wiki
It’s a Lot of Work to Deposit…

- Scraping and indexing is underway…with permission
- We average about 500 journal articles per week at present
- WITH STRUCTURES
- We are indexing HighWire now – “HighWire-hosted publishers have collectively made 1,873,044 articles free.”
- Then we will connect structures to indexed chemical names for a structure searchable literature database
Can ChemSpider Enable Discovery?

- Yes, chemists can search by text, structure, substructure or properties to find associated information – a one stop portal and...

<table>
<thead>
<tr>
<th>ID</th>
<th>Structure</th>
<th>Empirical Formula</th>
<th>Molecular Weight</th>
<th>Monoisotopic Mass, Da</th>
<th>LogP</th>
<th>ACD/LogD (pH 5.5)</th>
<th>ACD/LogD (pH 7.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>467489</td>
<td><img src="image" alt="Structure" /></td>
<td>$\text{C}<em>{13}\text{H}</em>{13}\text{F}\text{N}<em>{4}\text{O}</em>{2}$</td>
<td>252.2449</td>
<td>252.102254</td>
<td>1.38</td>
<td>-1.02</td>
<td>1.11</td>
</tr>
<tr>
<td>467721</td>
<td><img src="image" alt="Structure" /></td>
<td>$\text{C}<em>{16}\text{H}</em>{16}\text{O}_{3}$</td>
<td>184.2322</td>
<td>184.109944</td>
<td>ACD/LogP: 1.29 XLogP: 1.50</td>
<td>0.1</td>
<td>-1.69</td>
</tr>
</tbody>
</table>

Building a Structure Centric Community for Chemists
ChemSpider – Research in Progress

- Supporting Open Notebook Science as a repository
- For the purpose of online virtual screening
- Applying descriptors of various types to filter a database of 20 million compounds

In progress:
- Utilizing SimBioSys’ LASSO Descriptor
- Collaboration based on ECCR’s ChemModLab (see next talk)
LASSO uses 23 kinds of Interactive Surface Point Descriptors and
- is conformation independent
- screens at 1 million structures/min
- is proven to enrich screened databases
- provides scaffold hopping

- Hbond Donors (5 kinds)
- Acceptors (5 kinds)
- Ambivalent H donor/acceptor
- Aromatic Pi-stacking (5 kinds)
- Hydrophobic (3 kinds)
- Metal ions
- Misc (Sulfur, Halogens)

http://dx.doi.org/10.1007/s10822-007-9164-5
LASSO Descriptors on ChemSpider

- 40 target receptors chosen
  - From the Database of Useful Decoys dataset
  - [http://dud.docking.org/](http://dud.docking.org/)
  - Brian Shoichet, UCSF

- Wide range of receptor classes
- Each target family had 10s-100s of known actives
- Actives used as query files for LASSO
- LASSO similarity descriptors generated across all 40 targets and deposited on ChemSpider
LASSO Descriptors on ChemSpider

14 Million molecules have descriptors (so far)

<table>
<thead>
<tr>
<th>Category</th>
<th>Target</th>
<th>PDB Code</th>
<th>LASSO Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other Enzymes</td>
<td>PARP, poly(ADP-ribose) polymerase</td>
<td>1efy</td>
<td>0.74</td>
</tr>
<tr>
<td>Kinases</td>
<td>PDGFRb, platelet derived growth factor receptor kinase</td>
<td>N/A</td>
<td>0.70</td>
</tr>
<tr>
<td>Other Enzymes</td>
<td>COX-2, cyclooxygenase-2</td>
<td>1cx2</td>
<td>0.21</td>
</tr>
<tr>
<td>Kinases</td>
<td>EGFr, epidermal growth factor receptor</td>
<td>1m17</td>
<td>0.09</td>
</tr>
<tr>
<td>Nuclear Hormone Receptors</td>
<td>PPARg, peroxisome proliferator activated receptor</td>
<td>1fm9</td>
<td>0.08</td>
</tr>
<tr>
<td>Other Enzymes</td>
<td>HIVRT, HIV reverse transcriptase</td>
<td>1rt1</td>
<td>0.02</td>
</tr>
<tr>
<td>Nuclear Hormone Receptors</td>
<td>RXRa, retinoic X receptor R</td>
<td>1mvc</td>
<td>0.02</td>
</tr>
<tr>
<td>Kinases</td>
<td>P38 MAP, P38 mitogen activated protein</td>
<td>1kv2</td>
<td>0.02</td>
</tr>
<tr>
<td>Kinases</td>
<td>VEGFr2, vascular endothelial growth factor receptor</td>
<td>1vr2</td>
<td>0.02</td>
</tr>
<tr>
<td>Serine Proteases</td>
<td>FXa, factor Xa</td>
<td>1fr0</td>
<td>0.02</td>
</tr>
<tr>
<td>Metalloenzymes</td>
<td>PDE5, phosphodiesterase 5</td>
<td>1xp0</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Building a Structure Centric Community for Chemists
Choose relationships of target receptors based on LASSO descriptors for searching

For example, ask the question “What are the top 1000 molecules with similar LASSO descriptors to the actives for the Estrogen Receptor”
Building a Structure Centric Community for Chemists

<table>
<thead>
<tr>
<th>Nuclear Hormone Receptors</th>
<th>Kinases</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR, androgen receptor</td>
<td>CDK2, cyclindependent kinase 2</td>
</tr>
<tr>
<td>ER, estrogen receptor; agonist</td>
<td>EGFr, epidermal growth factor receptor</td>
</tr>
<tr>
<td>ER, estrogen receptor; antagonist</td>
<td>FGFr1, fibroblast growth factor receptor kinase</td>
</tr>
<tr>
<td>GR, glucocorticoid receptor</td>
<td>HSP90, human heat shock protein 90</td>
</tr>
<tr>
<td>MR, mineralocorticoid receptor</td>
<td>P38 MAP, P38 mitogen activated protein</td>
</tr>
<tr>
<td>PPARg, peroxisome proliferator activated receptor</td>
<td>PDGFr, platelet derived growth factor receptor kinase</td>
</tr>
<tr>
<td>PR, progesterone receptor</td>
<td>SRC, tyrosine kinase SRC</td>
</tr>
<tr>
<td>RXRa, retinoic X receptor R</td>
<td>TK, thymidine kinase</td>
</tr>
<tr>
<td></td>
<td>VEGFr2, vascular endothelial growth factor receptor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Serine Proteases</th>
<th>Metalloenzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>FXa, factor Xa</td>
<td>ACE, angiotensin-converting enzyme</td>
</tr>
<tr>
<td>Thrombin</td>
<td>ADA, adenosine deaminase</td>
</tr>
<tr>
<td>Trypsin</td>
<td>COMT, catechol O-methyltransferase</td>
</tr>
<tr>
<td></td>
<td>PDE5, phosphodiesterase 5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Folate Enzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DHFR, dihydrofolate reductase</td>
</tr>
<tr>
<td>SART</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Enzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AChE, acetylcholinesterase</td>
</tr>
</tbody>
</table>
Example for ER enrichment

- Review the KIERBL Dataset from EPA’s DSSTox
- Ran LASSO on this dataset to extract descriptors
- Aggregated the ER LASSO values for the ChemSpider structure collection (14 million subset)
- With 14 million “decoys” where are the known actives relative to matching LASSO descriptors? This gives an “enrichment number”
60% of the actives were recovered in the top 1% of the database.

“Environmental binders” are weak binders

The top ranked compounds may well be active ER binders

Likely candidates for experimental investigation

Building a Structure Centric Community for Chemists
Screening for weak ER binders
ER Screening Study Highlight

- ER-agonist set recovers high % actives
  - Exceptional results: virtual screening is looking for ligands that bind 1-3 orders of magnitude better than the natural ligand (NOT 3 to 5 orders worse!!!!)
  - Extremely sensitive with clearly positive enrichment
  - Using the set found in Chemspider could enrich a dataset for subsequent structure-based and/or in vitro screening!
Data are 1 Week Old

Work to be Done Yet To Validate Further

- Validation of LASSO over 14 million structures – “non-druglike” molecules – should we curate out?
- Run LASSO descriptors on remaining members of database
- Use PhysChem filters at time of Searching (already pre-calculated and in properties)
- Validate on real examples from drug discovery
The ChemSpider database is a rich source of structural diversity especially “Commercial Chemistry”

- With data source specific searching commercial ligands can be used for virtual screening
ChemSpider’s Virtual Screening Set to ChemModLab

- 2M commercial compounds from ChemSpider
- State of the art statistical modeling
- Identification of a set of potentially active molecules
Where to from here? **Short term**

- Integrated text and structure/substructure searching of the Open Access literature and Pubmed is in development (close)
- Batch-deposition system for large structure files (close)
- Web-based scraping of structure-based information
- Enhanced web services layer to integrate searches
- Open Notebook Science support
- Deposit latest SureChem Patent Database
- Reaction handling and deposition
Where to from here? **Mid-term**

- **Spidering for Chemistry** – extract data from articles, webpages and data sources AND stay within copyright
- **WiChempedia project** – wiki-layers on top of ChemSpider, alongside Wikipedia curation project
- **Deeper integration** to text-based searching and conversion of chemical names to structures for online structure searching:
  - Improved integration with NCBI Entrez system
  - Deliver “dedicated websites” for specific publishers
Where to from here? **Mid-Term**

- An extensible datamodel “on the fly” allows us to easily expand to integrate abstract data to structures.
- We are a connection network NOT just a database of information.
- Enhanced searching – ALL information in our data model will be searchable. An expanding list of data entities is growing.
- Social networking/Web 2.0. We will soon release a forum for discussions and collaboration around structures and reactions.
Seeking…

- Individuals/groups interested in making their students theses, reports, publications structure/substructure searchable online
- Individuals/groups interested in building a free access online repository of chemical reactions
- Individuals/groups interested in making their analytical data available online for the masses
- Individuals with structure-based algorithms to connect
- Individuals who would like to support the ChemSpider curation process
Long-term?

- Ask me at ACS Philly..
Acknowledgments

- **LASSO**: Rocky Goldsmith and Aniko Simon (SimBioSys’ LASSO can be seen at Booth 432)
- **NISS**: Stan Young, Jackie Hughes-Oliver & ECCR Team
- **The ChemSpider Advisory Group** (almost 30 people)
- **Commercial Software Providers** (SureChem, OpenEye, ChemAxon, SimBioSys, ACD/Labs, Microsoft)
- **Open Source community** (Jmol, JSpecView)
- **ChemSpider Depositors, Curators and Users**
- **And…**
ChemSpider is self-funded

Sponsorship contributions made by Waters, Thermo, Indigo Molecular Models, SimBioSys, SureChem and Otava Chemicals. Thanks!
Happy Birthday ChemSpider

- “1 year old today” – Released Sunday Spring ACS 2007
- Now averages 4000 unique visitors per day
- >3000 web service calls per day – increasing as more vendors integrate
- Moving quickly from text to structure searching

Building a Structure Centric Community for Chemists
Out of 1 Year Beta Release Today
New Website Release (today/tonight...)
Contact

antony.williams@chemspider.com