Bringing Open Source to Drug Discovery

Chris Swain

Cambridge MedChem Consulting

Standing on the shoulders of giants

- There are a huge number of people involved in writing open source software
- It is impossible to acknowledge them all individually
- The slide deck will be available for download and includes 25 slides of details and download links
  - Copy on my website www.cambridgemedchemconsulting.com

Why us Open Source software?

- Allows access to source code
  - You can customise the code to suit your needs
  - If developer ceases trading the code can continue to be developed
  - Outside scrutiny improves stability and security

What Resources are available

- Toolkits
- Databases
- Web Services
- Workflows
- Applications
- Scripts

Toolkits

- OpenBabel
- CDK
- RDkit
- Indigo
- ChemmineR
- Helium
- FROWNS
- Perlmol
- R
- OpenCL
- SciPy
- NumPy
- Pandas
- Flot
- GNU Octave
- OpenMPI

Toolkits

- OpenBabel (htttp://openbabel.org) is a chemical toolbox
  - Ready-to-use programs, and complete programmer's toolkit
  - Read, write and convert over 110 chemical file formats
  - Filter and search molecular files using SMARTS and other methods, KNIME add-on
  - Supports molecular modeling, cheminformatics, bioinformatics
  - Organic chemistry, inorganic chemistry, solid-state materials, nuclear chemistry
  - Written in C++ but accessible from Python, Ruby, Perl, Shell scripts…

Toolkits

- RDKit (http://www.rdkit.org)
  - A collection of cheminformatics and machine-learning software written in C++ and Python.
  - Knime nodes
  - The core algorithms and data structures are written in C ++. Wrappers are provided to use the toolkit from either Python or Java.
  - Additionally, the RDKit distribution includes a PostgreSQL-based cartridge that allows molecules to be stored in relational database and retrieved via substructure and similarity searches.

Toolkits

- The Chemistry Development Kit (CDK. http://sourceforge.net/projects/cdk/)
- A library for bio-, cheminformatics and computational chemistry written in Java
  - Read and write chemical data formats
  - Render chemical structures
  - Algorithms for chemical graph theory
  - QSAR descriptors

Toolkits

- Indigo (http://ggasoftware.com/opensource/indigo)
  - Universal organic chemistry toolkit
  - Cheminformatics algorithms
  - Open-source chemical search engine Bingo is developed on top of the Indigo library
  - Knime nodes
  - Written in C++ but accessible from Python, Ruby, Perl, Shell scripts…

Toolkits

- ChemmineR (http://manuals.bioinformatics.ucr.edu/home/chemminer)
- Helium (http://www.moldb.net/helium.html)
- FROWNS (http://frowns.sourceforge.net)
- Perlmol (http://www.perlmol.org)

Databases

- ChemSpider
- ChEMBL
- SureChemBL
- BindingDB
- Zinc
- PDB
- Crystallography Open Database

Databases

- ChEMBL (https://www.ebi.ac.uk/chembl/)
  - A database of chemical structures and bioactivities
    - 1,566,466 compound records
    - 12,419,715 activities
    - 1,042,374 assays
    - 9,414 targets
    - Subsets for Malaria, Kinase, GPCR, Neglected tropical disease
- SureChemBL(http://www.surechembl.org/)

Databases

- BindingDB (http://www.bindingdb.org)
  - A public, web-accessible database of measured binding affinities. BindingDB contains 1,009,290 binding data, for 6,589 protein targets and 427,325 small molecules
- Zinc (http://zinc.docking.org)
  - A database of commercially-available compounds for virtual screening. ZINC contains over 35 million purchasable compounds in ready-to-dock, 3D formats

Databases

- PDB (http://www.rcsb.org)
  - The Protein Data Bank (PDB) archive is the repository of information about the 3D structures of large biological molecules, including proteins and nucleic acids.
- ChemSpider (http://www.chemspider.com)
  - A free chemical structure database providing fast text and structure search access to over 30 million structures
- Crystallography Open Database (http://www.crystallography.net)

Web Services

- Chemical Identifier Resolver (http://cactus.nci.nih.gov/chemical/structure)
- UniChem (https://www.ebi.ac.uk/unichem/)
- OCHEM (https://ochem.eu)
- Virtual Computational Chemistry Laboratory (http://www.vcclab.org)
- Molinspiration (http://www.molinspiration.com)
- ChemSpider (http://www.chemspider.com/AboutServices.aspx?)
- OpenPHACTS (https://dev.openphacts.org/docs/1.3)
- Beaker (https://pypi.python.org/pypi/chembl_beaker/)

Workflow

- KNIME (http://www.knime.org)
  - Desktop client to design and execute scientific workflows
  - Nodes add extra functinality
- Taverna (http://www.taverna.org.uk)
  - Desktop client to design and execute scientific workflows
  - Integrated with myGRID

Applications within OpenBabel

- obabel - a converter for chemistry and molecular modeling data files
  - 101 file types supported
  - 2D and 3D structure generation
  - Substructure and similarity searching
  - Protonate/deprotonate
  - Generate images
- obminimize - optimize the geometry, minimize the energy for a molecule
- obprop – calculate molecular properties
- obrotate - batch-rotate dihedral angles matching SMARTS
- obrotamer - generate conformer/rotamer coordinates
- obfit - superimpose two molecules based on SMART
- obprobe - create electrostatic probe grid

Programs Using Open Babel

Avogadro - Flexible, cross-platform 3D molecular editor

ASE - Atomic Simulation Environment, a Python framework for building simulation codes, as well as pre- and post-processing tools.

Brabosphere - Brabosphere, a molecular visualization tool for the quantum mechanics package BRABO

CCP1GUI - Free, extensible molecular viewer and editor, including GUI for GAMESS-UK

ChemAzTech - Web/database solution. Manage your chemical database, with product's properties. Draw, edit molecules and interact with common chemical software.

Cheméo - Web search for high quality chemical properties

Chemtool - 2D molecular drawing tool using GTK toolkit

ChemSpotlight - Chemical indexing for Mac OS X

Chemical Structure Project - Open Source Chemical Structure Database

CheS-Mapper - Open Source 3D-viewer for chemical datasets of small molecules

eMolecules - World-wide chemical search

fminer2 - Molecular subgraph mining

GChemPaint - 2D molecular drawing tool for GNOME environment

Ghemical - Molecular editor for GNOME environment

Gnome Chemistry Utils - GTK/Gnome based tools for chemistry software development

Other toolkits have similar lists http://openbabel.org/wiki/Related_Projects

Programs Using Open Babel

hBar Lab: Computer-aided Chemistry On Demand - First web application for performing molecular modeling calculations including quantum chemistry calculations online.

iBabel: GUI for OpenBabel

Kalzium - KDE Chemistry Education project

KMovisto - 3D molecular viewer for Linux

KNIME - Modular data processing (pipeline) environment

lazar - Lazy Structure-Activity Relationships (prediction of toxicity data)

Molekel - Advanced multiplatform 3D molecular viewer

molsketch - 2D molecular drawing tool for Qt/KDE environment (in development)

MyChem - MySQL extension for chemical databases

NanoEngineer-1 - Open source 3D molecular modeler for Windows, Mac and Linux.

NanoHive-1 Nanospace Simulator - Modular simulator for modeling nanometer scale physics and chemistry

OpenMD - Open source molecular dynamics engine

Open3DALIGN - Conformational search and unsupervised molecular alignment

Open3DQSAR - High-throughput computation and chemometric analysis of molecular interaction fields

OSRA - Optical Structure Recognition (graphics into SMILES)

pgchem::tigress - PostgreSQL database engine for molecules

Pharao - Open source pharmacophore generation and alignment

Programs Using Open Babel

Piramid - Open source shape-based alignment using Gaussian volumes

PyRx - Virtual Screening tool that includes Open Babel widget (watch PyRx Screencast - Open Babel to learn more).

SDF2XYZ2SDF - How to exploit TINKER power in cheminformatics projects

Semantic Web Open Babel - Using Open Babel for a chemical Semantic Web

Sieve - Filtering molecule files based on calculated molecular properties, topologies and fragments

SMIREP - SMILES based structure-activity relationship rules

SPORCalc - Substrate Product Occurrence Ratio Calculator - Fingerprint Probabilistic Scoring of Sites of Metabolism

Stripper - Scaffold extraction tool following a number of published scaffold-extraction approaches, including Schuffenhauer, Oprea and Murcko

Toxtree - Estimate toxicity activity using decision trees

V-Sim Visualizes atomic structures such as crystals, grain boundaries, etc.

ViewMol - Molecular viewer and graphical front end for computational chemistry programs

WebBabel - WebBabel is a web application that uses OpenBabel to convert files from one format to another, or browse for files containing molecular structures.

XDrawChem - XWindows/Qt-based 2D molecular drawing tool

YASARA - Yet Another Scientific Artificial Reality Application

Zodiac - Cross-platform suite for Rational drug design

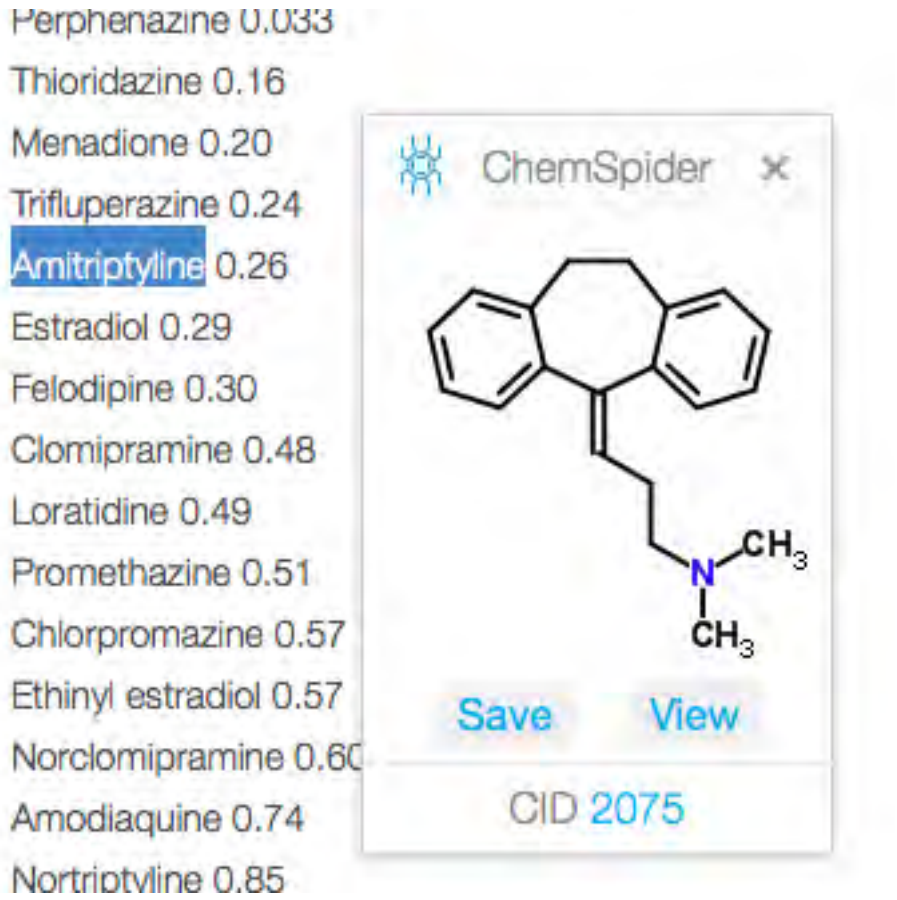
Other Open Source Applications

- GROMACS (http://www.gromacs.org/)
- NWChem (http://www.nwchem-sw.org/index.php)
- PyMOL (http://www.pymol.org/)
- VMD (http://www.ks.uiuc.edu/Research/vmd/)
- OpenMD (http://openmd.org)
- AutoDock Vina (http://vina.scripps.edu)
- Open3DQSAR (http://open3dqsar.sourceforge.net)
- SMARTCyp (http://www.farma.ku.dk/smartcyp/)
- Toxtree (http://sourceforge.net/projects/toxtree/)
- PaDEL (http://padel.nus.edu.sg/software/)

Web Browser Extensions

- Add extra functionality to the web browser
  - Chemspider :- Displays structure of highlighted chemical/drug and links to ChemSpider page.
  - PubChem :- Search PubChem for the highlighted compound
  - eMolecules :- Search eMolecules for the highlighted compound
  - Chemicalize :- Submit the current URL to chemicalize.org
  - DrugBank :- Search DrugBank for the highlighted compound
  - Opsin :- Chemical name to structure and display.
  - ChEMBL :- Displays structure of highlighted chemical/drug and links to ChEMBL page
  - Chemistry Reference Resolver :- This adds a toolbar to your browser as well as adds a right-click menu option for reference resolving

ChemSpider Extension



Scripts

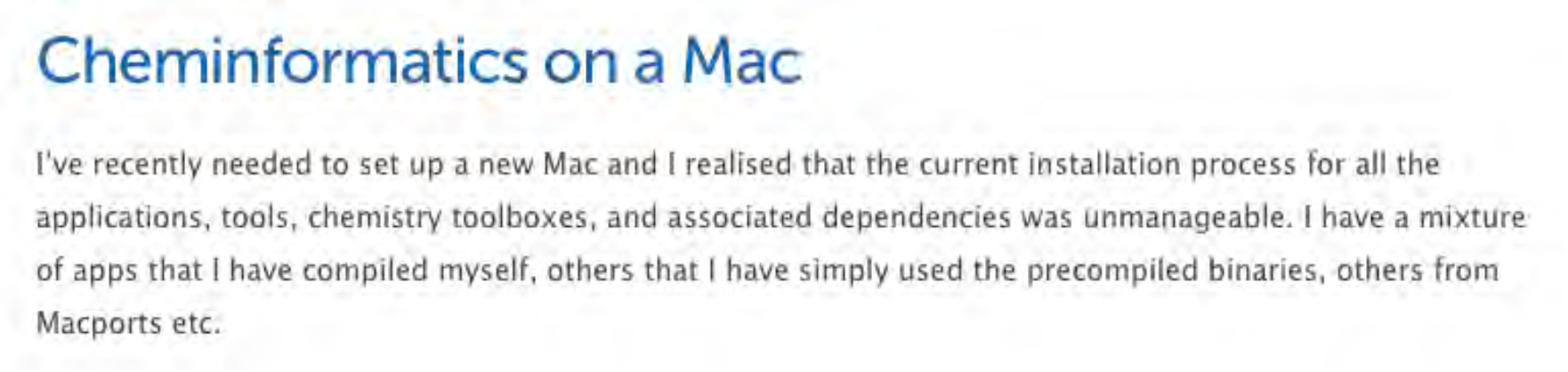
- “Official”
  - SVL Exchange (http://svl.chemcomp.com/index.php),
  - Vortex Script Exchange (https://support.dotmatics.com)
- “Unofficial”
  - Scripts for OpenBabel, Vortex, ChemDraw, RDkit
    - http://www.macinchem.org/reviews/hints_tutorials.php
- Google Apps Scripts for an intuitive interface to organic chemistry Open Notebook
    - http://usefulchem.blogspot.co.uk/2011/06/google-apps-scripts-for-intuitive.html
- Michel Petitjean
  - A variety of cheminformatics scripts
    - http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html
- Jrjohansson/scientific-python-lectures
  - https://github.com/jrjohansson/scientific-python-lectures

More scripts

- Flot (http://www.flotcharts.org) a pure JavaScript plotting library for jQuery
- Pandas (http://pandas.pydata.org) Python Data Analysis Library
- R Scripts (http://cran.r-project.org/doc/contrib/Lemon-kickstart/kr_scrpt.html)

Finding and Installing everything

- Most of the tools are available for all platforms
  - Package managers make installation easier.
  - A step by step guide



Cheminformatics on a Mac

I've recently needed to set up a new Mac and I realised that the current installation process for all the applications, tools, chemistry toolboxes, and associated dependencies was unmanageable. I have a mixture of apps that I have compiled myself, others that I have simply used the precompiled binaries, others from Macports etc.

http://www.macinchem.org/reviews/cheminfo/cheminfoMac.php