2.6: Chemical Resolvers, Molecular Editors and Visualization

Introduction

Chemical resolvers can take one form of molecular representation and convert it to another. That is, they can resolve what the compound is from its represenation. They can be web based services or software applications. A molecular editor is in essence a type of resolver, that has a graphical editor interface where human's can draw molecules. But under the hood, it is using cheminformatics representations like connection tables. Database services like PubChem and ChemSpider also have integrated editors and resolvers and so the distinction across these is a bit fuzzy.

Chemical Resolvers

We will define chemical resolvers as programs that can resolve a chemical structure from a representation, and then use that to transform it to another representation or provide information on the chemical.

Web-Based Resolvers

These are services that typically offer both a GUI (for humans) and an API (for machines). This list is not comprehensive, and will grow as time allows.

CIR

The **Chemical Identifier Resolver** (CIR) is a service of the Computer-Aided Drug Design (CADD) group of the Chemical Biology Laboratory (CBL) of the National Cancer Institute (NCI) in Maryland US. The direct link to CIR is here:

https://cactus.nci.nih.gov/chemical/structure

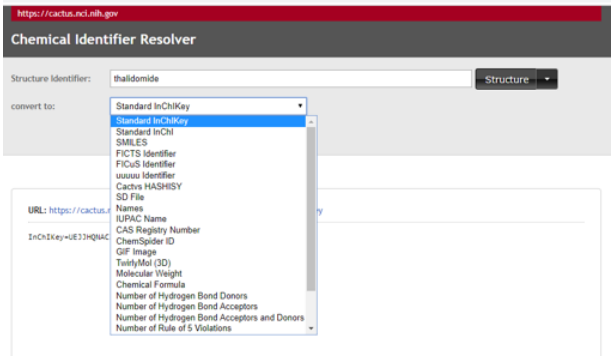

Figure 2.6.1: The GUI interface for the NCI/CADD CIR (https://cactus.nci.nih.gov/chemical/structure).

The CIR service also offers a variety of API interfaces and we already explored one of them with the InChILayersExplorer of activity 2.4.1 (section 2.4.3.3.1), where an Excel spreadsheet used the CIR service to convert a name to an InChI.

There are also a variety of other resources available through the **CA**DD Group **C**hemiformatics **T**ools and **U**ser **S**ervices (CACTUS) that students are encouraged to explore.

https://cactus.nci.nih.gov/

OPSIN

The **Open Parser for Systematic IUPAC Nomenclature** (OPSIN) is run by the Centre for Molecular Informatics and the University of Cambridge in England, the URL is:

https://opsin.ch.cam.ac.uk/

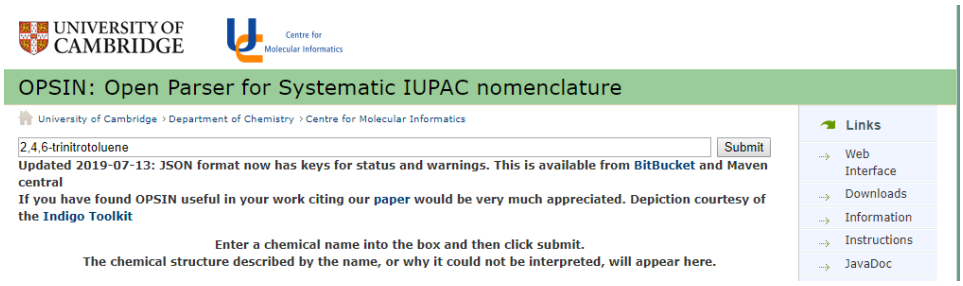

Figure 2.6.2: OPSIN resolver, (https://opsin.ch.cam.ac.uk/)

One of the nice things of the OPSIN resolver is that if you have an incorrect IUPAC name, it stops where it can't parse the name and tell's you where the problem is. For example, patents will often use IUPAC names and believe it or not, they are often misspelled and wrong! So lets look at this name here,

1-[4-(2-methoxyethy)phenoxy]-3-(propan-2-amino)propan-2-ol.

Do you see what is wrong with it? Well pasting it into OPSIN gives you an error and an idea where to look!

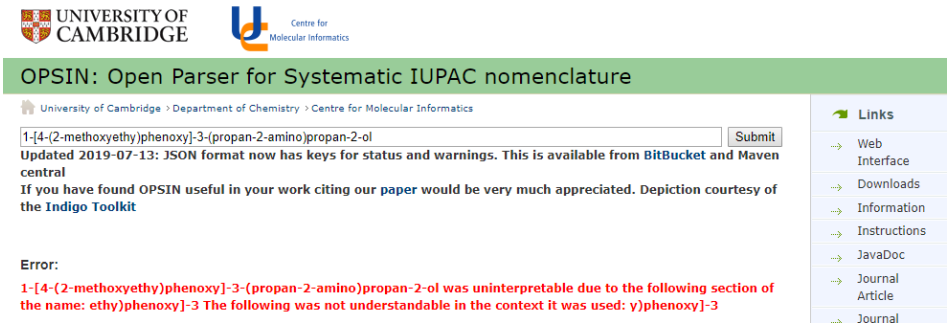

Figure 2.6.3: OPSIN resolver showing an error in resolving the IUPAC name.

The beauty is the error message occurs when the resolver could no longer parse the word and it got stumped at the y), which should have been yl), as in 1-[4-(2-methoxyethyl)phenoxy]-3-(propan-2-ylamino)propan-2-ol.

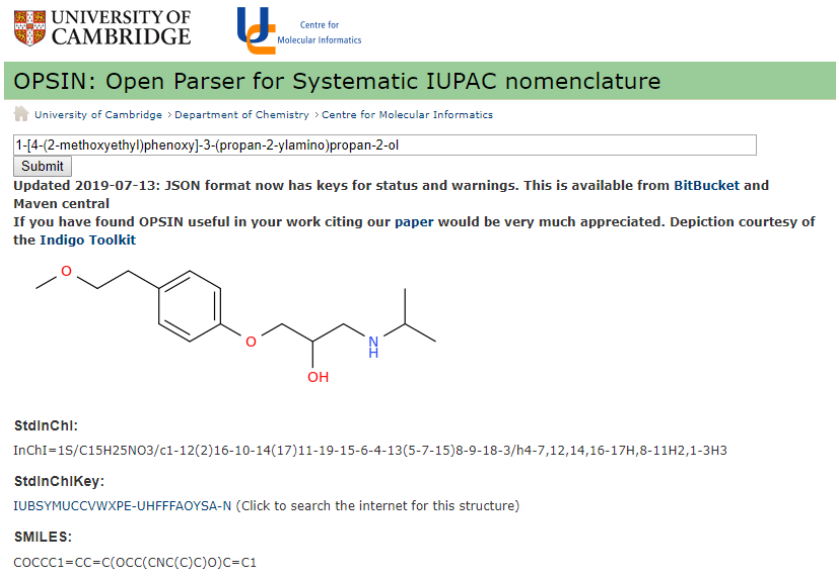

Figure 2.6.4: Discovering the type in figure 3 allowed us to correct the IUPAC name and we now have identifiers for this compound.

The above figures show how the OPSIN GUI can be used. In activity X we will look at a Google Spreadsheet that uses the OPSIN resolver to convert a list of IUPAC names to InChIKeys and then use those keys to link directly to PubChem on those chemicals,

and thus instantly get access to information on a list of compounds. Lets manually do that now. That is, we append the InChI key to following URL stem

https://pubchem.ncbi.nlm.nih.gov/compound/INCHIKEY

and so the following link gets us to information on this compound

https://pubchem.ncbi.nlm.nih.gov/compound/IUBSYMUCCVWXPE-UHFFFAOYSA-N

bringing us to the following information on PubChem

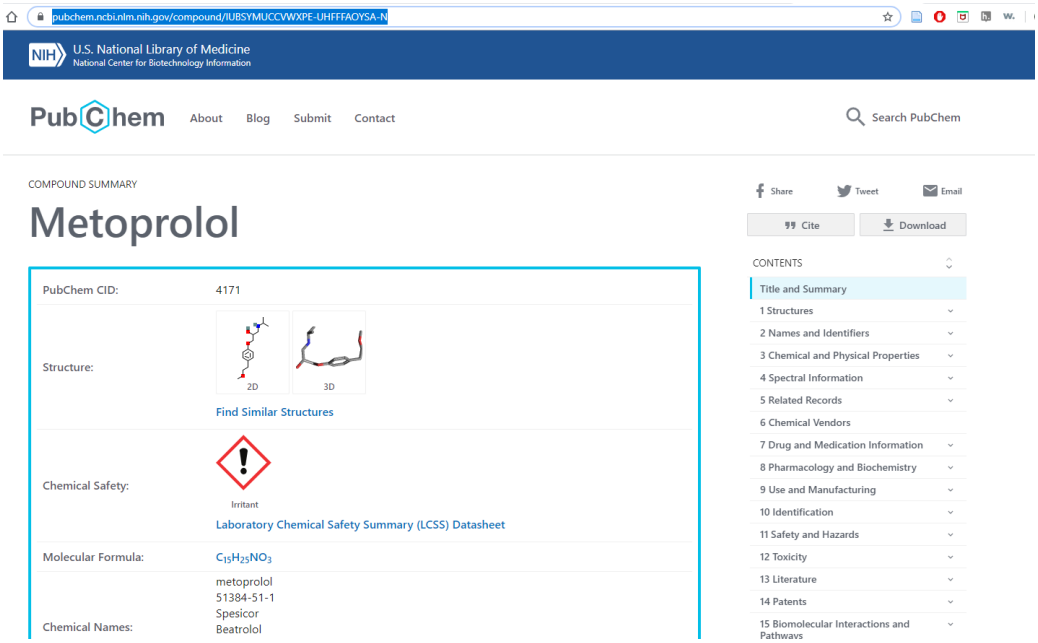

Figure 2.6.5: After using OPSIN to correct the incorrect IUPAC name we can deduce the compound is Metoprolol and obtain information on it.

IUPAC Name2PubChem

Go to the InChI OER (https://www.inchi-trust.org/oer/) and filter Content Type/Spreadsheet with File Type/Google Sheet, and click on the hit for the "IUPAC Name2PubChem". From here you will find a Google sheet that automates the above process for a list of chemicals. Video 2.6.1 shows how you can go to the OPSIN chemical resolver and figure out how to create a Google sheet that performs these functions.

Video 2.6.1: 6:47 min YouTube video describing how to connect an column of IUPAC names to information in PubChem (https://youtu.be/oDxMUJ0dNWw)

If you download the module from the InChI OER you will see the code described in the video which you can cut and paste into a spreadsheet. This module uses two webservices functions, the IMPORTDATA function and the HYPERLINK function, along with the concatenate function for a string of text, &. It is very important that you develop the skill of looking at code you have never seen and then finding out what it does through web searches.

Resolver Programs

OK, Resolver Programs may not be the best title for this section. These are software packages that convert molecular file formats, and are often called tool kits. These have many functionalities that can be used offline, in contrast to web services, which need internet access.

Open Babel

Open Babel is an open source Chemistry Toolbox and does much more than just convert structural formats and can be downloaded from the following URL.

http://openbabel.org

In figure 2.6.6 we see using Open Babel to convert a SMILES string to a mol file with 3D coordinates. The Open Babel Documentation can give you a feel for some of the things you can do with Open Babel, http://openbabel.org/docs/current/OpenBabel.pdf .

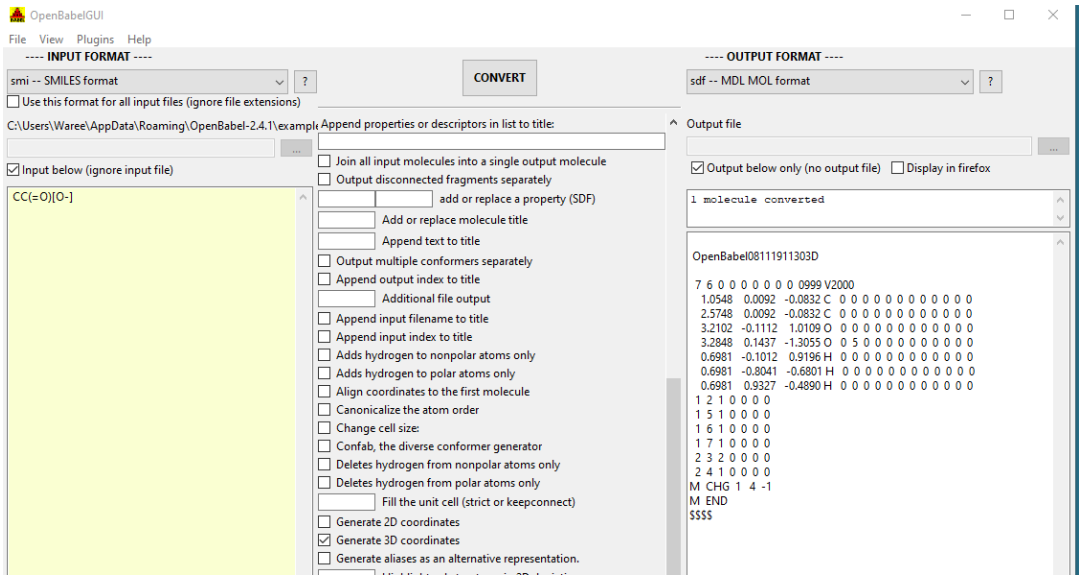

Figure 2.6.6: Using Open Babel to convert between file types, here converting a SMILES string to a 3D mol file with coordinates.

Molecular Editors

Web Services

1. MolView - site maintained by Herman Bergwerf
   - http://molview.org/
2. PubChem Sketcher
   - https://pubchem.ncbi.nlm.nih.gov/#draw=true
3. ChemSpider Structure Search
   - https://www.chemspider.com/StructureSearch.aspx
4. Chemagic
   - https://chemagic.org/molecules/amini.html
5. Wikipedia Chemical Structure Explorer
   - http://www.cheminfo.org/Wikipedia/
   - see JChem Inf. article (https://jcheminf.biomedcentral.com/articles/10.1186/s13321-015-0061-y)

Software Packages

Dr. Tamas Gunda has posted a decent resource: Chemical Drawing Programs - The Comparison of Accelrys (Biovia) Draw, ChemBioDraw (ChemDraw), DrawIt, ChemDoodle and Chemistry 4-D Draw (http://www.gunda.hu/dprogs/index.html)

Open Source

1. JChemPaint - open source (LGLP license):
   - https://jchempaint.github.io/
2. BIOVIA Draw – free for academic use
   - http://accelrys.com/products/collaborative-science/biovia-draw/draw-no-fee.php
3. ChemSketch Freeware - free for academic and personal use, requires registration and download:
   - http://www.acdlabs.com/resources/freeware/chemsketch/
4. BIO-RAD - Chemical Structure drawing, spectral analysis & more
   - http://www.knowitall.com/academic/
5. JSME - open source (BSD license):
   - http://peter-ertl.com/jsme/

Fee-Based

1. ChemDraw – requires subscription and download
   - Basic drawing package: http://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemDraw/ChemDrawPrime/Default.aspx
2. ChemDoodle – requires purchase and download
   - https://www.chemdoodle.com/
   - Free Trial: https://www.chemdoodle.com/free-trial/
3. ChemSketch – requires purchase and download
   - http://www.acdlabs.com/products/draw_nom/draw/chemsketch/
   - Free Trial at above link
4. Chemistry 4-D Draw
   - http://www.cheminnovation.com/products/chem4d.asp

Molecular Visualization

1. Jmol and Jsmol
   - http://jmol.sourceforge.net/
2. Avogadro
   - https://avogadro.cc/
