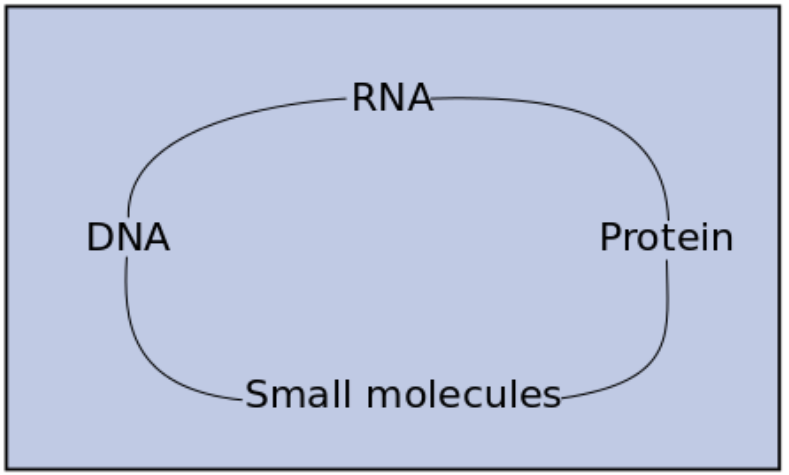
Cheminformatics, Chemical Space and R: Making Sense of FDA and CA Pesticide Reports

Andrew Defries PhD

andrew.defries@gmail.com
https://github.com/andrewdefries

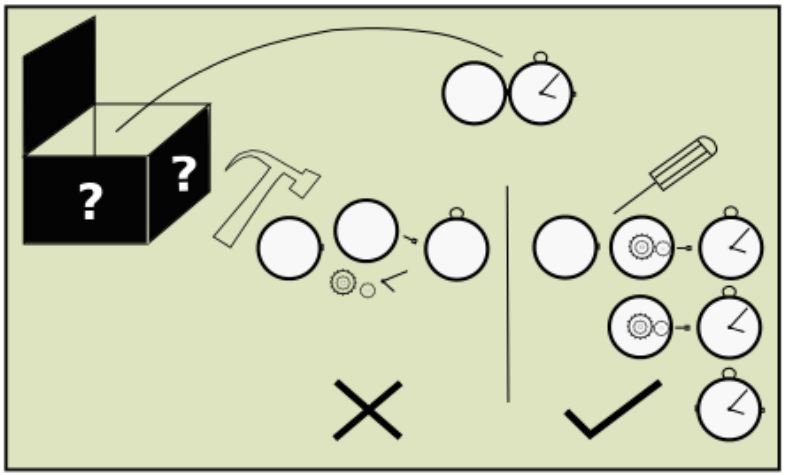
Chemical Space and Bioactivity

The Central Dogma of Biology



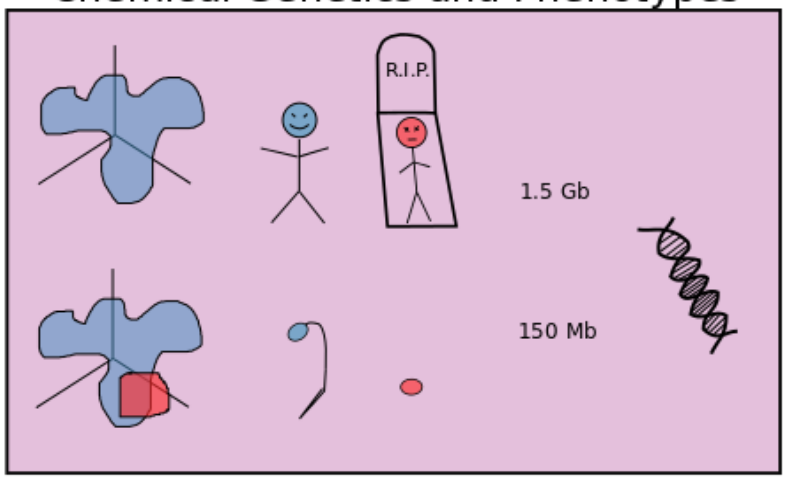
A complex circuit of interactions

The Black Box and the Watch Maker

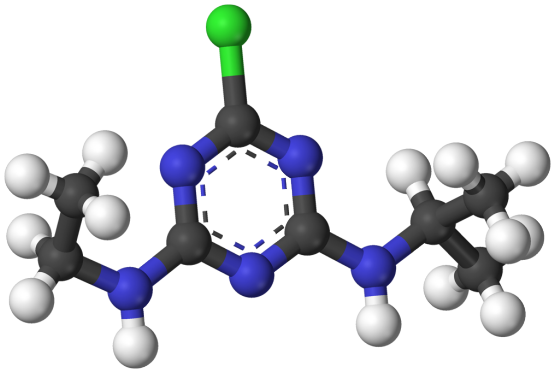


Breaking things and putting them back together

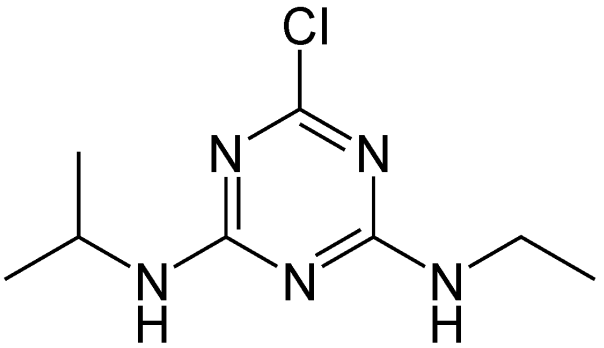
Chemical Genetics and Phenotypes



Cheminformatics



3D-depiction



2D-depiction

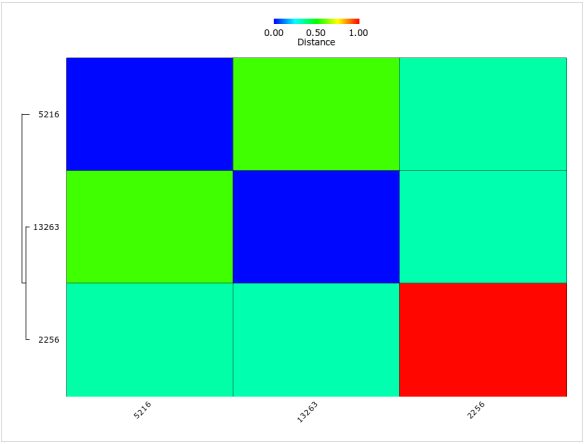
Clc1nc(nc(n1)NC(C)C)NCC

1D-depction

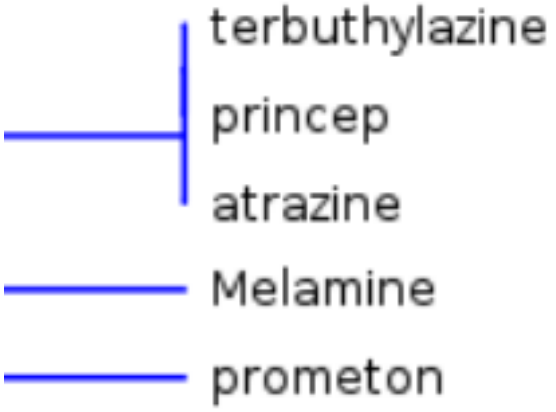
```
2256
  -OEChem-05121409442D

 28 28  0     0  0  0  0  0  0999 V2000
    5.4641    1.8100    0.0000 Cl  0  0  0  0  0  0  0  0  0  0  0  0
    3.7320   -1.1900    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    7.1962   -1.1900    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    5.4641   -1.1900    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    4.5981    0.3100    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    6.3301    0.3100    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    2.8660   -0.6900    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.0000   -1.1900    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.8660    0.3100    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    4.5981   -0.6900    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    8.0622   -0.6900    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    6.3301   -0.6900    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    8.9282   -1.1900    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    5.4641    0.8100    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    3.4030   -0.3800    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    3.7320   -1.8100    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    1.6900   -0.6531    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    1.4631   -1.5000    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    2.3100   -1.7269    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    3.4860    0.3100    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    2.8660    0.9300    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    2.2460    0.3100    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    8.4607   -0.2151    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    7.6636   -0.2151    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    7.1962   -1.8100    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    8.6182   -1.7269    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    9.4651   -1.5000    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
    9.2382   -0.6531    0.0000 H   0  0  0  0  0  0  0  0  0  0  0  0
```

SDF-format

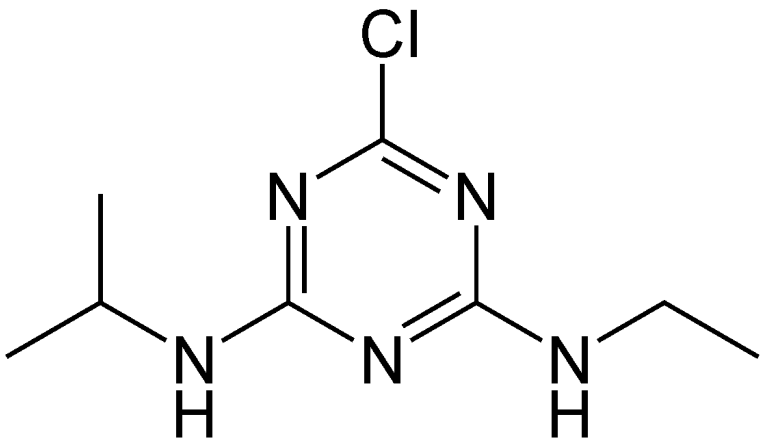


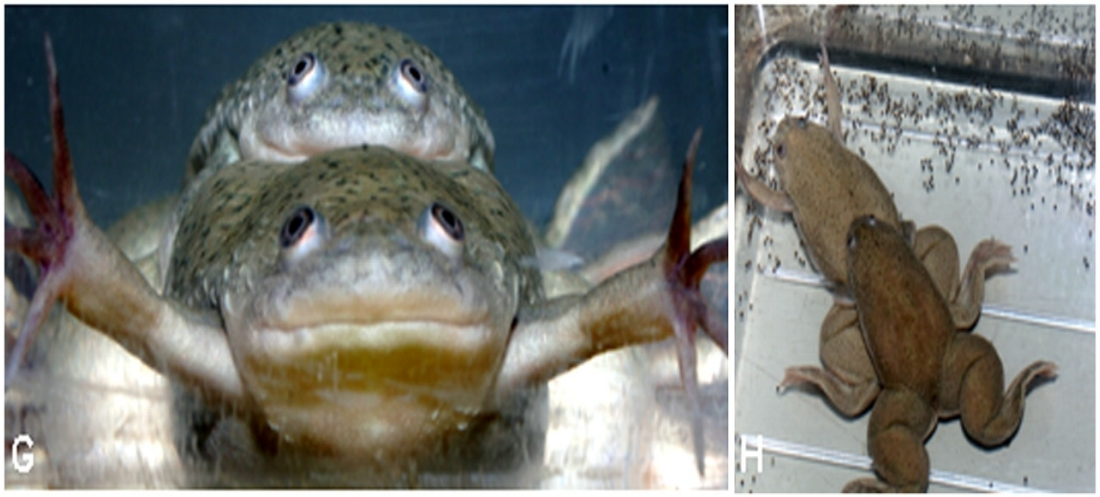
Property based clustering



Similarity based clustering

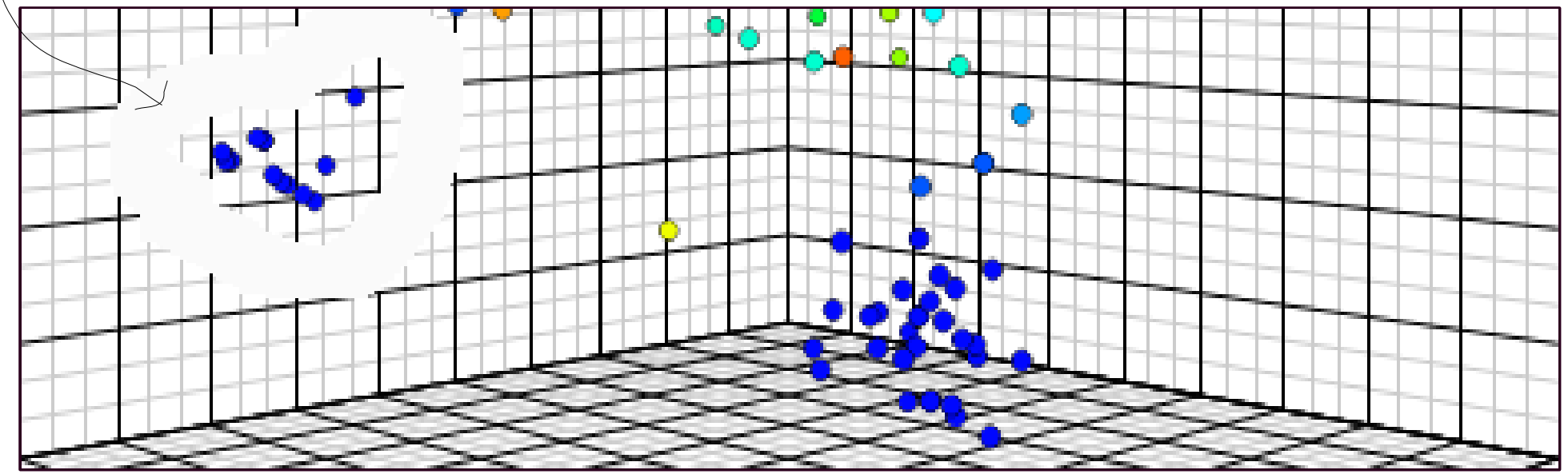
Case Study Atrazine





Atrazine class of herbicides

Hayes et al.,PNAS (2010)



Chemical Space of Global Herbicide Directory

FDA and CA Pesticides

atrazine **1912-24-9**

Use: herbicide
Category: parent
Mol. Form.: C8H14ClN5
LMS Code: 305
Tolerances: 180.220, foreign use
Alternate Names:

9CI	6-chloro-N-ethyl-N'-(1-methylethyl)-1,3,5-triazine-2,4-diamine
ANSI	atrazine
ISO	atrazine
IUPAC	6-chloro-N2-ethyl-N4-isopropyl-1,3,5-triazine-2,4-diamine
Trade	Aatrex
	Atranex
	Atratol
	Gesaprim

```
(2-naphthyloxy)acetic acid              plant growth regulator  Telone
1,3-dichloropropene, (EZ)-              nematocide      Fruitone
1-naphthol              nematocide      Benzac
2,3,5-tri-iodobenzoic acid              nematocide      Weedone
(2,4-dichlorophenoxy)acetic acid                plant growth regulator
4-(2,4-dichlorophenoxy)butyric acid             plant growth regulator
4-pyridinamine          herbicide       Tordon 101
N6-benzyladenine                herbicide       Butyrac
O,S-dimethyl acetylphosphoramidothioate         bactericide     Release
3-dodecyl-1,4-dihydro-1,4-dioxo-2-              herbicide/plant growth
N-((6-chloro-3-pyridyl)methyl)-N'-cyano-N-              insecticide
2-chloro-N-ethoxymethyl-6'-ethylacet-o-         herbicide       Orthene
2-chloro-6-nitro-3-phenoxyaniline               plant growth regulator
cyano-3-phenoxybenzyl (Z)-(1R,3S)-2,2-          plant regulator Acenit
2-chloro-2',6'-diethyl-N-               plant growth regulator  Blazer
ethyl (Z)-N-benzyl-N-[[methyl(1-                plant growth regulator
2-methyl-2-(methylthio)propionaldehyde          safener Rufast
2-mesyl-2-methylpropionaldehyde O-              bird repellent  Lasso
```

STATE OF CALIFORNIA
PESTICIDES SOLD IN CALIFORNIA FOR YEAR: 1991
Combined Disclosed Active Ingredients By Chemical Name

11/19/02

ACTIVE INGREDIENTS	Number of Registrants	Total Pounds
(Z,E)-7,11-HEXADECADIEN-1-YL ACETATE	4	610.02
1,3-DICHLORO-5,5-DIMETHYLHYDANTOIN	16	169,768.76
1,3-DICHLORO-5-ETHYL-5-METHYLHYDANTOIN	16	65,676.96
1,3-DICHLOROPROPENE	5	64,423.97
1-BROMO-3-CHLORO-5,5-DIMETHYL HYDANTOIN	49	1,413,438.55
1-NAPHTHALENEACETAMIDE	9	32.57
2,2-DIBROMO-3-NITRILOPROPIONAMIDE	29	26,663.01
2,4-D	18	103,968.24
2,4-D, ALKANOLAMINE SALTS (ETHANOL AND ISOPROPANOL AMINES)	4	- .41
2,4-D, DIETHANOLAMINE SALT	6	98.01
2,4-D, DIMETHYLAMINE SALT	46	446,277.68
2,4-D, ISOOCTYL ESTER	10	4,824.10
2,4-DP, ISOOCTYL ESTER	4	1,184.10

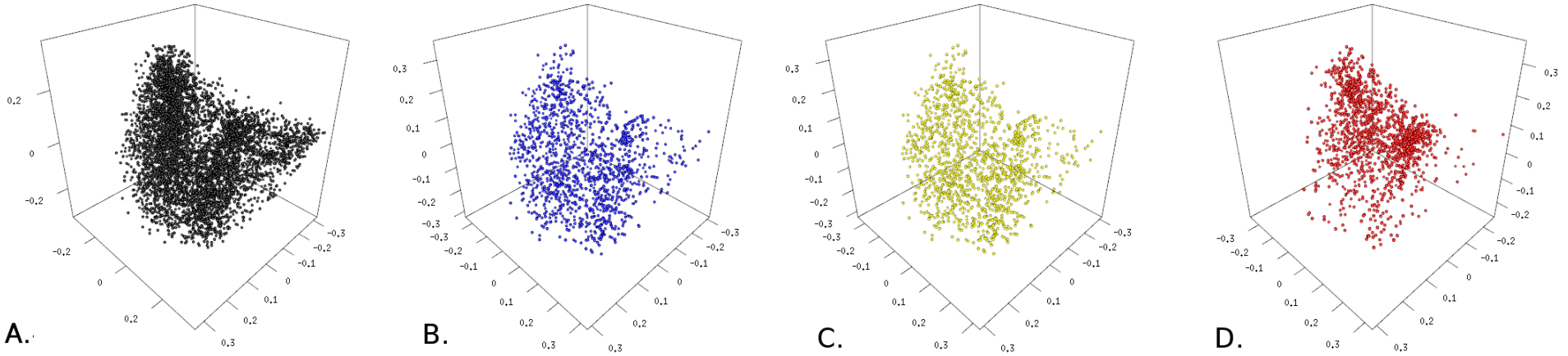
```
"ALUMINUM PHOSPHIDE"                5           136221
"AMITRAZ"                   5           5028 2000
"AMMONIUM SULFATE"                  7           295863
"ARSENIC ACID"              4           621342 2000
"ATRAZINE"                  4           99902 2000
"AVERMECTIN"            6           10901 2000
"AZADIRACHTIN"          8           1861 2000
```

https://github.com/andrewdefries/FDA_Pesticide_Glossary

https://github.com/andrewdefries/CA_Pesticides_1991_2011

sed 's/foo/bar/g'
cut -c x-y

rgl



Merge

FDA
approved drugs

FDA
small molecule drugs

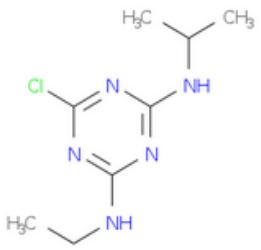
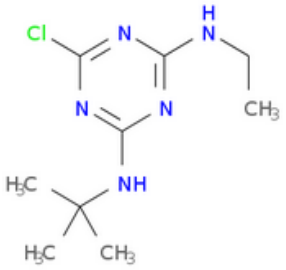
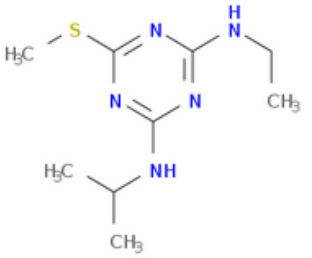
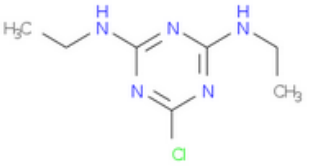
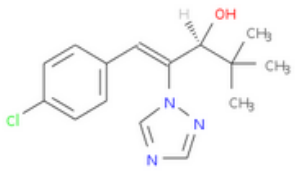
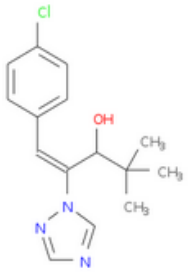
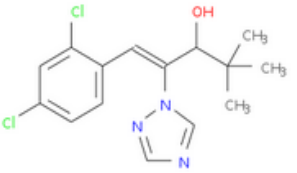
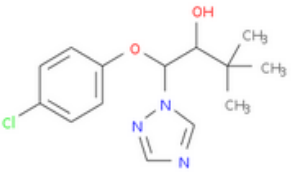
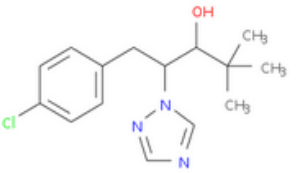
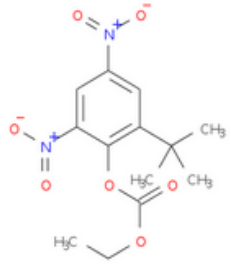
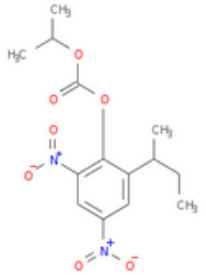
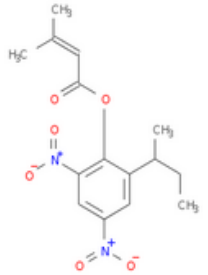
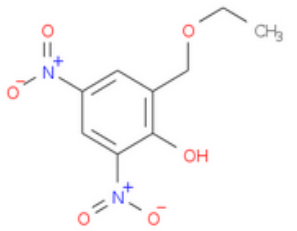
FDA
approved pesticides

```
library(ChemmineR)
sdfset<-read.SDFset()
apset<-sdf2ap(sdfset)
clusters<-cmp.cluster(apset, cutoff=c(0.7))
coord<-cluster.visualize(apset, clusters, size.cutoff=1, dimensions=3, quiet=TRUE)
```

```
library(rgl)
rgl.open() ..
spheres2d(coord$V1, coord$V2, coord$V3, radius=0.005, color="black", alpha=1 ..)
rgl.snapshot
rgl.close()
```

https://github.com/andrewdefries/FDA_Pesticide_Glossary/tree/master/CompareToDrugBank/ChemicalSpaceMontage

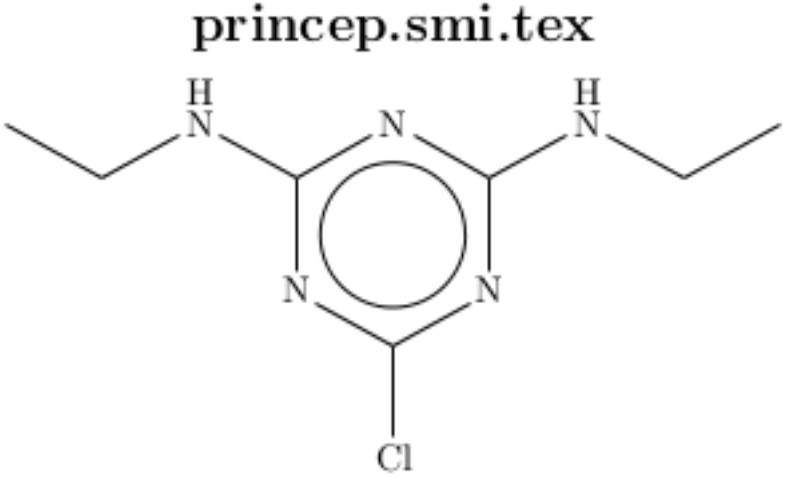
hwriter

1475	atrazine	 atrazine	 terbuthylazine	 iazine-2,4-diamine, N,N'-diethyl-6-(met	 Ametrine	 princep
1476	uniconazole	 henyl)-4,4-dimethyl-2-(1H-1,2,4-triazol	 uniconazole	 diniconazole	 Triadimenol	 Paclobutrazol
1477	2-(1,1-dimethylethyl	 limethylethyl)-4,6-dinitrophenyl ethyl ca	 dinobuton	 Binapacryl	 butralin	 2-(ethoxymethyl)-4,6-dinitrophenol

```
library(hwriter)
nnm<-nearestNeighbors(apset, numNbrs=10)
png<-list.files(pattern="smi.png")
himg=hwriteImage(Draw, table=FALSE)
hwrite(mat, 'out.html', br=TRUE, center=TRUE...)
```

http://andrewdefries.github.io/FDA_Pesticide_Glossary/FDA_Pesticide_Glossary.html

Latex and mol2chemfig



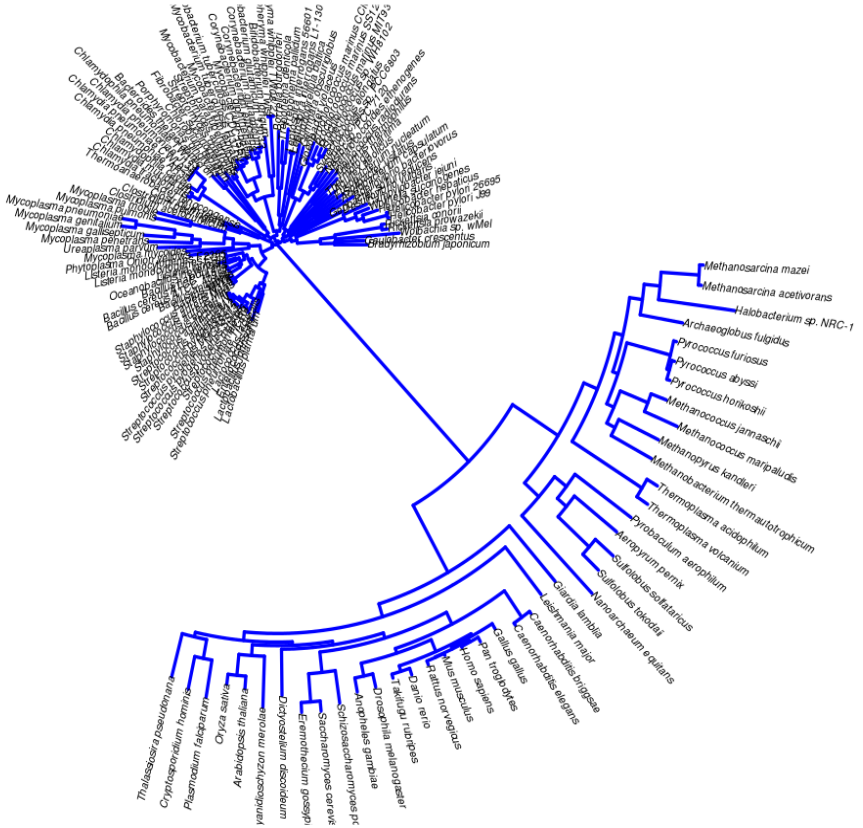
- chemical structure conversion to smi
- smi conversion to tikz
- batch report via R

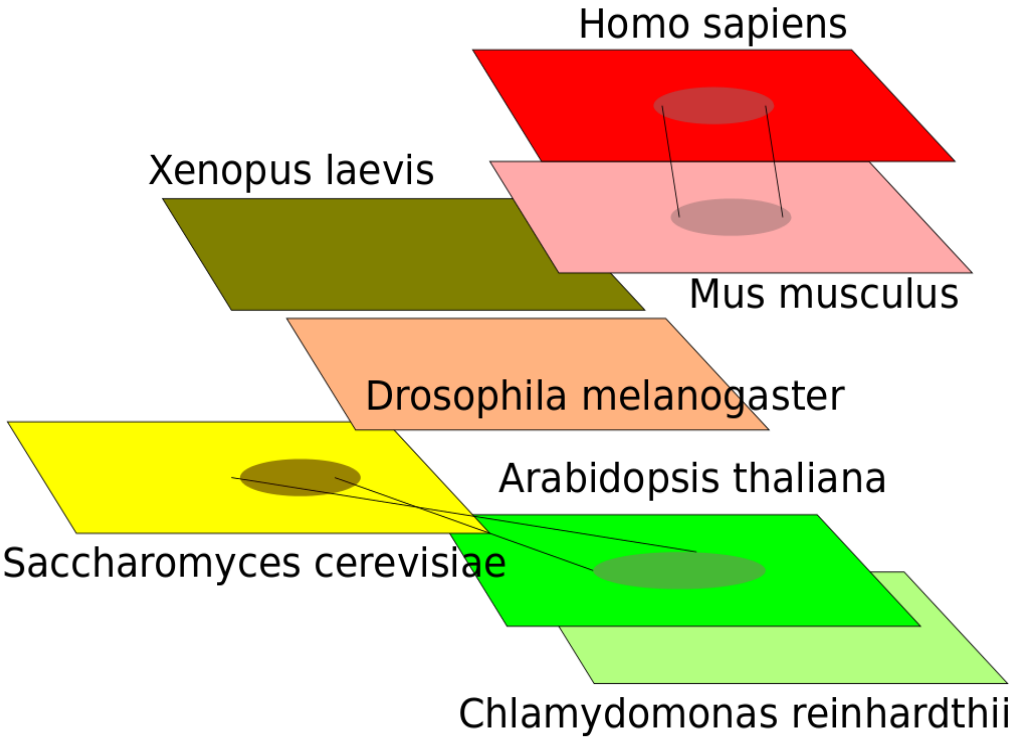
```
library(ChemmineR)
library(ChemmineOB)
write.SMI(smi=smiset, file="out.smi", cid=T)
system("for s in *.smi
do
mol2chemfig -wo $s> $s.tex
done")
```

```
\documentclass{article}
\usepackage{mol2chemfig}
\pagestyle{empty}
\begin{document}
\input{out.smi.tex}
\section{out.smi}
\end{document}
```

https://github.com/andrewdefries/SDF2PDF

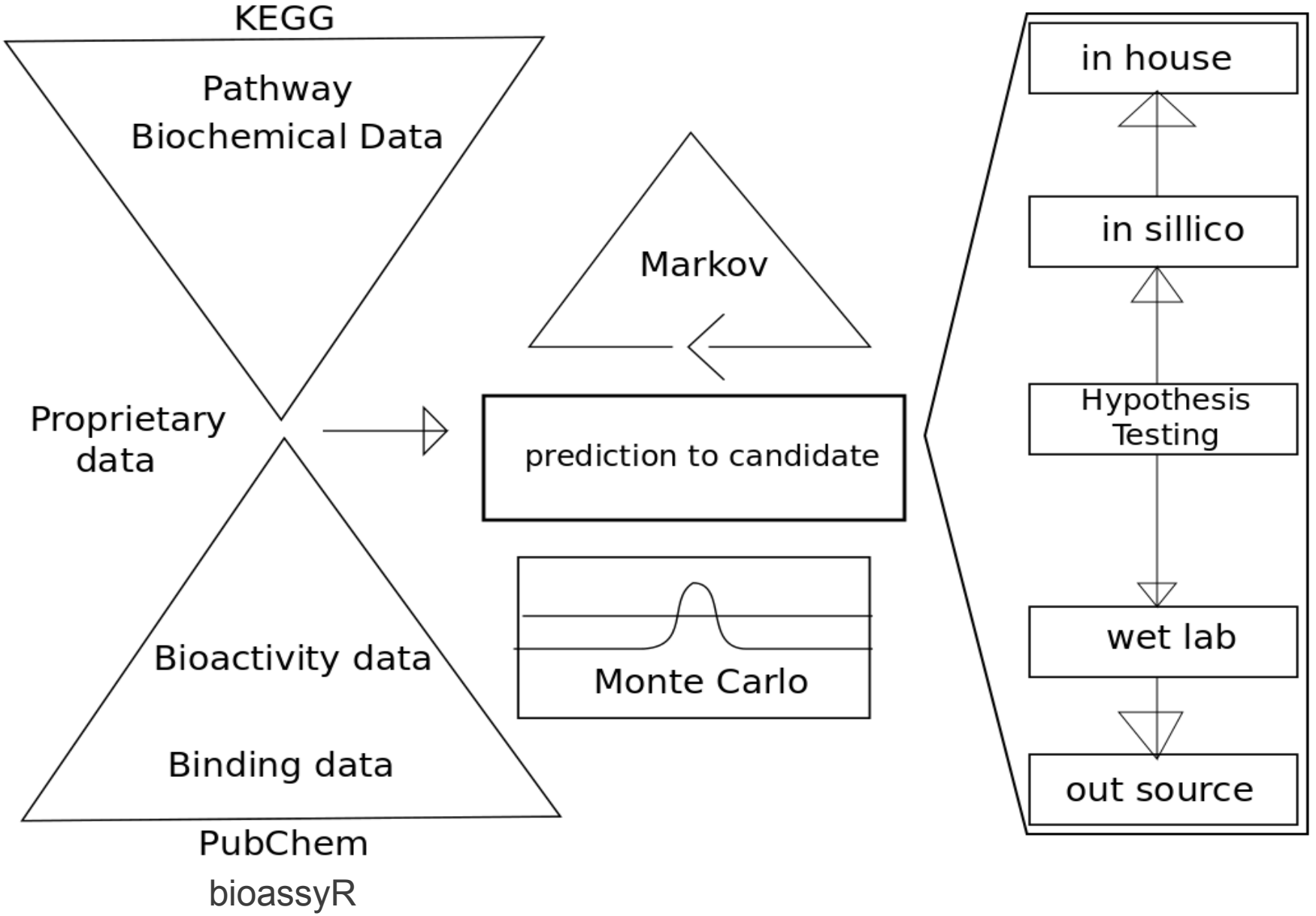
Future Directions





Using existing public data and model organisms
to understand cross-species bioactivity trends

Future Directions



Leveraging big data to make predictions about bioactive compounds

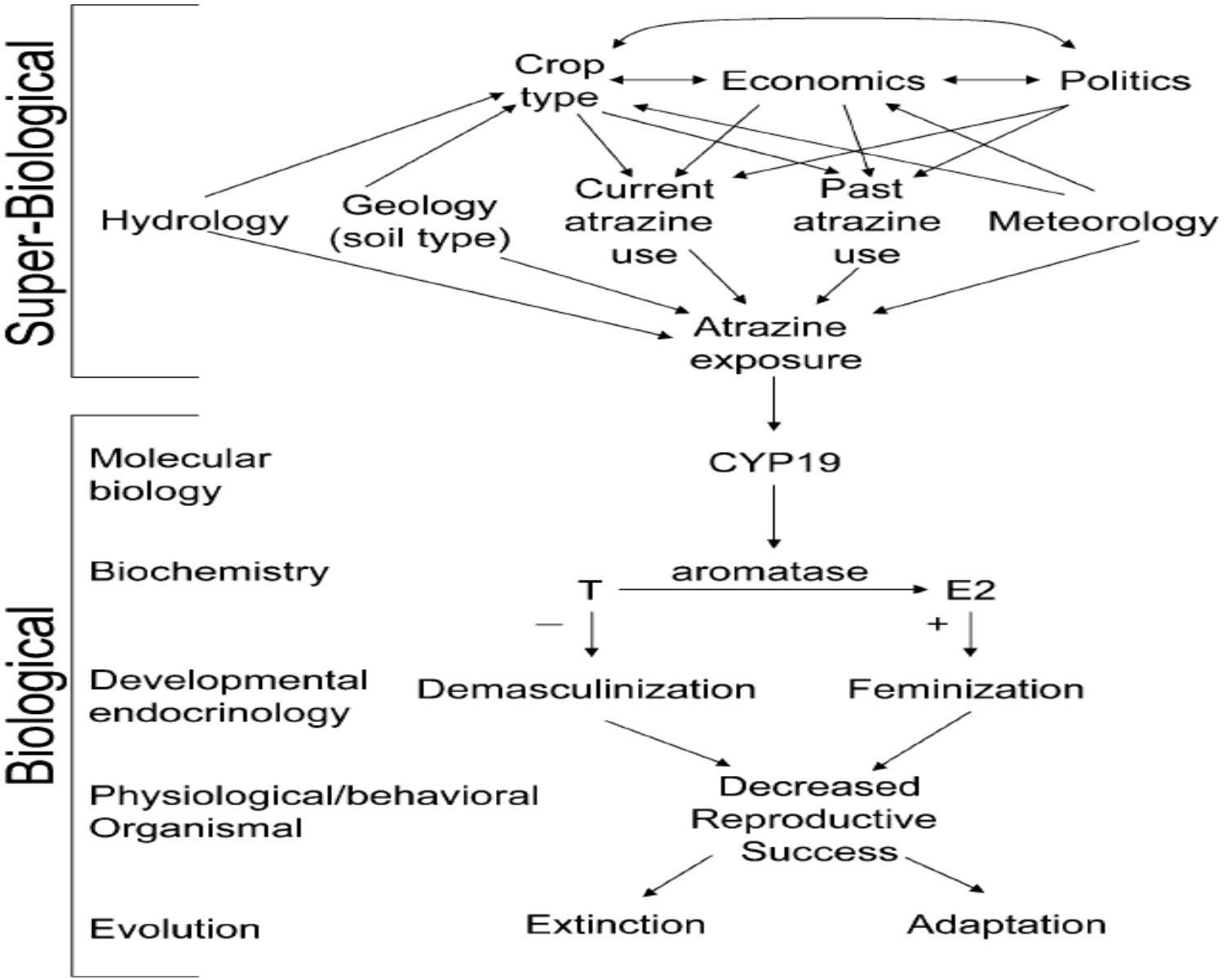
Packages

- ChemmineR
- ChemmineOB
- fmcsR
- hwriter
- rgl
- shiny
- bioassayR
- ctc

- Open Babel
- Latex
- mol2chemfig

- hwriterPlus
- biomaRt
- RCircos
- Inparanoid
- tm

How to “sex” a frog



Hayes et al. Integr. Comp. Biol (2005)