



## RDKit in the Modern Biotech

*RDKit User Group Meeting 2018*  
*Ben Tehan & Rob Smith*

© Heptares Therapeutics 2018

*The HEPTARES name, the logo and STAR are trade marks of Heptares Therapeutics Ltd*

*Heptares Therapeutics is a wholly owned subsidiary of Sosei Group Corporation*



# Why RDKit?

- Initially Heptares had a multitude of languages, webserver and toolkits
  - 2009 - 2013
- Languages
  - Fortran, C, C++, Perl, Java, Javascript, Python, Matlab, R, etc...
- Webservers
  - Ruby on Rails, PHP, Javascript, etc...
- Toolkits
  - Indigo, CDK, OEChem, other flavours also available etc...
- What we needed going forward
  - general purpose language which you can find almost anywhere.
  - web applications,
  - desktop apps,
  - network servers,
  - machine learning,
  - easily distributable

# Why RDKit?

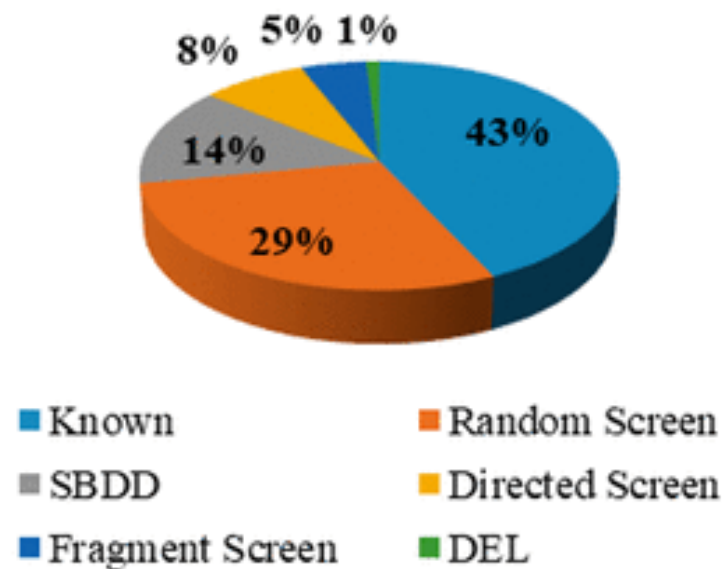
- Heptares Computational Chemistry adopted Python & RDKit as standard
  - 2013 - now
- Languages
  - Python
- Webservers
  - Django
- Toolkits
  - Open Source and active
  - RDKit
- What we needed going forward
  - general purpose language which you can find almost anywhere.
  - web applications,
  - desktop apps,
  - network servers,
  - machine learning,
  - easily distributable



# Drug Discovery Process - lite

- Program initiated (not target selection etc...)
- Cpds identified
  - patent
  - literature
- Leads examined
- Hit expansion
- Hit Analyses

## Lead generation strategies:



# Drug Discovery Process - lite

- Program initiated (not target selection etc...)
- Cpds identified
  - patent
  - literature
- Leads examined
- Hit expansion
- Hit Analyses
- Explore phase
- Patent Analyses
  - Frequency of R-Group analyses
  - FOG analyses
    - Tyrchan et al. JCIM, 2012, 52 (6), pp 1480–1489,
    - DOI: 10.1021/ci3001293
- Similarity
  - Structural (fingerprint)
  - Pharmacophoric
  - Shape-based

# Similarity – Structural Fingerprints

```
# do topological for comparison (path)
mfp = Chem.RDKitFingerprint(mol, maxPath=5, fpSize=1024, nBitsPerHash=2)
PATH5_sim = DataStructs.FingerprintSimilarity(originalfp,mfp)

# do morgan for comparison (circular)
mfp2 = rdMolDescriptors.GetMorganFingerprint(mol,2)
ECFP4_sim=DataStructs.DiceSimilarity(mfp1,mfp2)

# do maccs also (functional group count)
maccsfp2 = MACCSkeys.GenMACCSKeys(mol)
MACCS_sim= DataStructs.FingerprintSimilarity(maccsfp1,maccsfp2)

molHAC = Chem.MolFromSmiles(smi).GetNumAtoms()

# do Fraggles now
res = DataStructs.BulkTverskySimilarity(mfp,queries,0,1,False)

# set some values specified in command line
current_highest_tver=float(tversim)
current_highest_frag=float(fragssim)
fraggles_highest=None

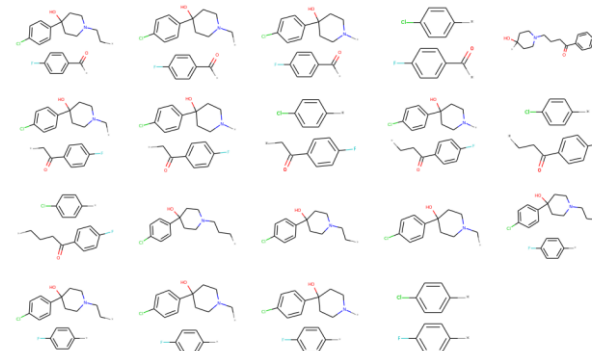
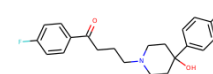
for i in range(0,len(queries)):

    if((float(res[i]) >= float(tversim)) and (float(molHAC) >= float(queryHAC-3)) and (float(molHAC) <= float(queryHAC+4)) ):
        # check greater than tversky similarity cutoff and query matches are of similar size to initial molecule
        # screen passes tversky criteria

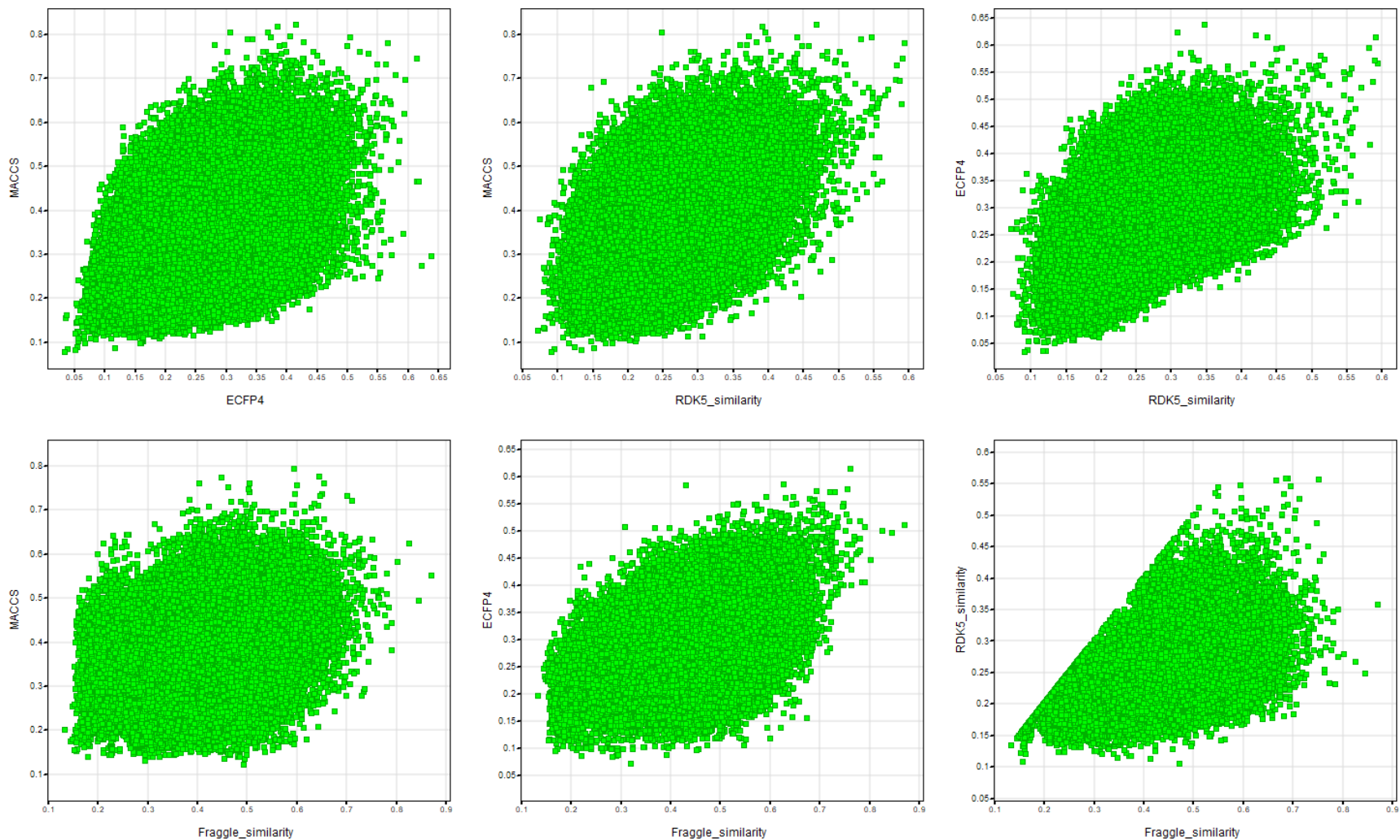
        rdkit_sim,fraggles_sim=FragglesSim.compute_fraggles_similarity_for_subs(mol,original_mol,query_smi,out_fragments[i])

        if(float(fraggles_sim) >= float(current_highest_frag)):
            fraggles_highest=float(fraggles_sim)
            current_highest_frag=fraggles_highest

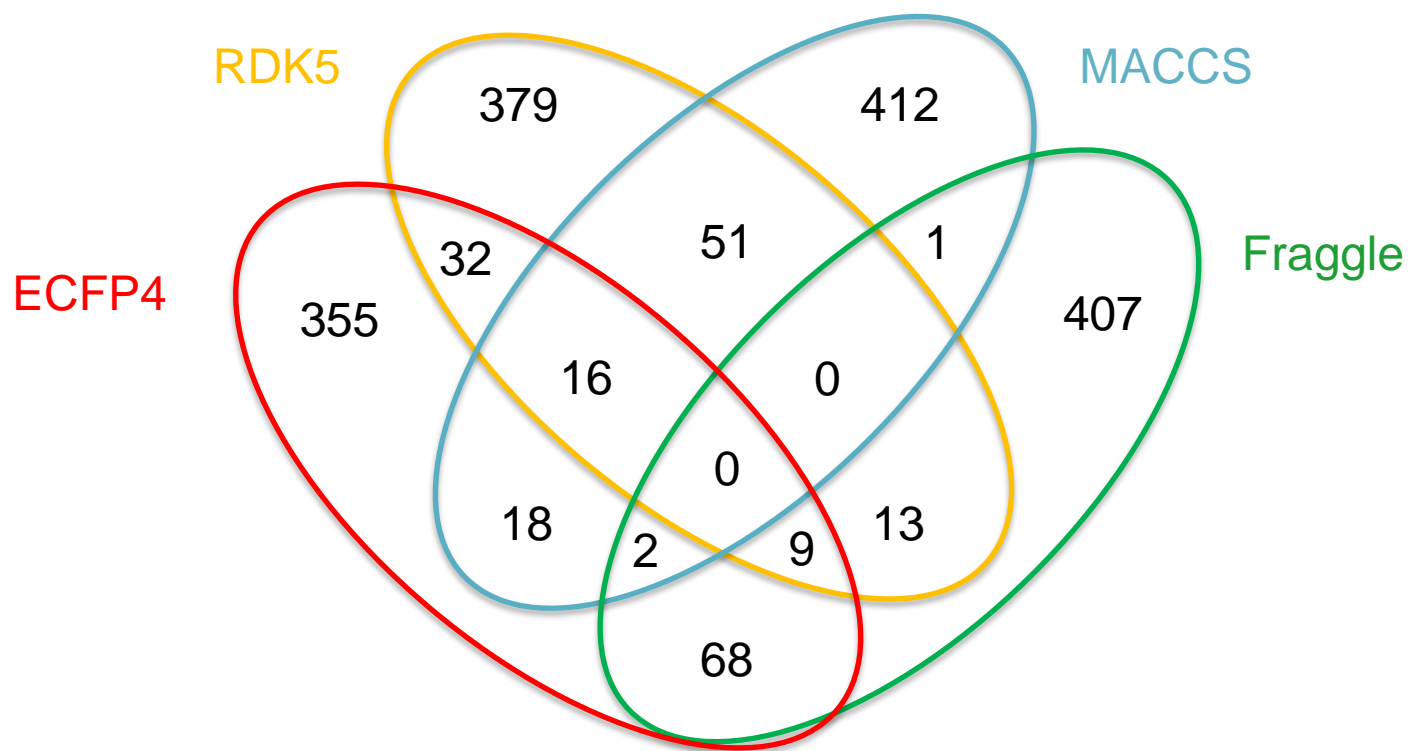
if((fraggles_highest is not None) and (fraggles_highest >= float(fraggles_sim))):
    # got to end of queries and passed tversky criteria
    None
else:
    current_highest_frag='Null'
```



# Similarity – Structural Fingerprints



# Similarity – Structural Fingerprints



Top 500 from each method



# Similarity – All Fingerprints

## RESEARCH ARTICLE

## Open Access



## Comparing structural fingerprints using a literature-based similarity benchmark

Noel M. O'Boyle\* and Roger A. Sayle

### Abstract

**Background:** The concept of molecular similarity is one of the central ideas in cheminformatics, despite the fact that it is ill-defined and rather difficult to assess objectively. Here we propose a practical definition of molecular similarity in the context of drug discovery: molecules A and B are similar if a medicinal chemist would be likely to synthesise and test them around the same time as part of the same medicinal chemistry program. The attraction of such a definition is that it matches one of the key uses of similarity measures in early-stage drug discovery. If we make the assumption that molecules in the same compound activity table in a medicinal chemistry paper were considered similar by the authors of the paper, we can create a dataset of similar molecules from the medicinal chemistry literature. Furthermore, molecules with decreasing levels of similarity to a reference can be found by either ordering molecules in an activity table by their activity, or by considering activity tables in different papers which have at least one molecule in common.

**Results:** Using this procedure with activity data from ChEMBL, we have created two benchmark datasets for structural similarity that can be used to guide the development of improved measures. Compared to similar results from a virtual screen, these benchmarks are an order of magnitude more sensitive to differences between fingerprints both because of their size and because they avoid loss of statistical power due to the use of mean scores or ranks. We measure the performance of 28 different fingerprints on the benchmark sets and compare the results to those from the Riniker and Landrum (J Cheminf 5:26, 2013. doi:10.1186/1758-2946-5-26) ligand-based virtual screening benchmark.

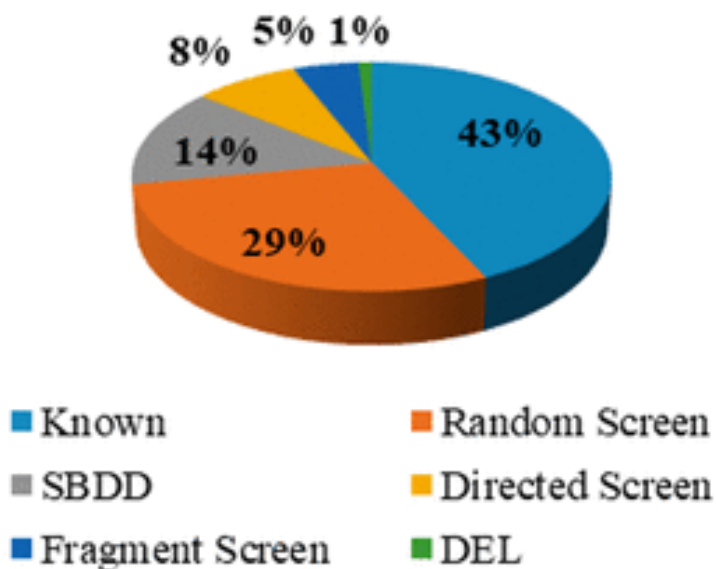
**Conclusions:** Extended-connectivity fingerprints of diameter 4 and 6 are among the best performing fingerprints when ranking diverse structures by similarity, as is the topological torsion fingerprint. However, when ranking very close analogues, the atom pair fingerprint outperforms the others tested. When ranking diverse structures or carrying out a virtual screen, we find that the performance of the ECFP fingerprints significantly improves if the bit-vector length is increased from 1024 to 16,384.

**Keywords:** Similarity searching, Molecular fingerprints, Structural similarity, Similarity benchmark

# Similarity – Shape-based

- Molecular shape based alignment is useful for VS, scaffold hopping, knowledge based ligand identification -> lead generation

## Lead generation strategies:



## Target classes:

- Kinases (30%)
- Other enzymes (23%)
- GPCR's (17%)
- Epigenetic (9%)

## Disease areas:

- Oncology (30%)
- CNS/Pain (18%)
- Infection (14%)
- Metabolic (12%)

Where Do Recent Small Molecule Clinical Development Candidates Come From?  
J. Med. Chem. ASAP, Dean Brown & Jonas Bostrom

# Similarity – Shape-based – Alignment Methods

## ○ AlignMol

- The 3D transformation required to align the specified conformation in the probe molecule to a specified conformation in the reference molecule is computed so that the root mean squared distance between a specified set of atoms is minimized.
- This transform is then applied to the specified conformation in the probe molecule

## ○ GetO3A

- Get an O3A object with atomMap and weights vectors to overlay the probe molecule onto the reference molecule based on **MMFF atom types** and charges

## ○ GetCrippenO3A

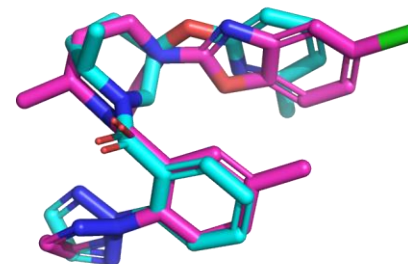
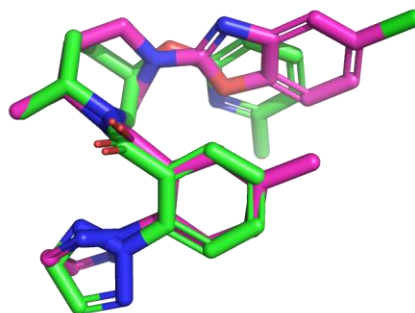
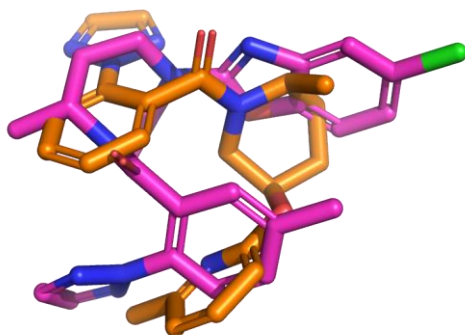
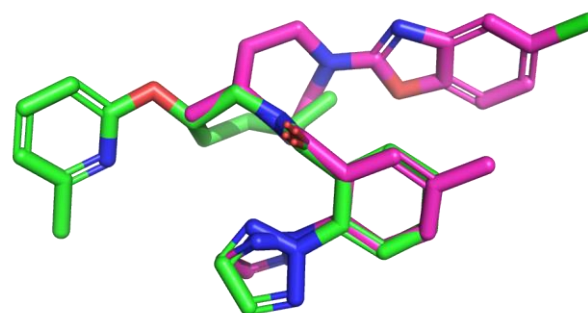
- Get an O3A object with atomMap and weights vectors to overlay the probe molecule onto the reference molecule based on **Crippen logP** atom contributions
- O3A objects then subsequently applied to reference and probe conformer to give O3A rmsd

Greg 2014 – ROCs -> subshape investigation  
<https://iwatobipen.wordpress.com/2014/06/15/try-to-subshape-alignment/>  
It was hard to me to do it using RDKit. Hmm...

# Similarity – Shape-based – Assessing Alignments

- RMSD
  - from alignment methods
- ShapeTanimotoDist
  - Compute the shape tanimoto distance between two molecule based on a predefined alignment
- ShapeProtrudeDist
  - Compute the shape protrude distance between two molecule based on a predefined alignment

minimise rmsd



# Similarity – Shape-based – Assessing Alignments

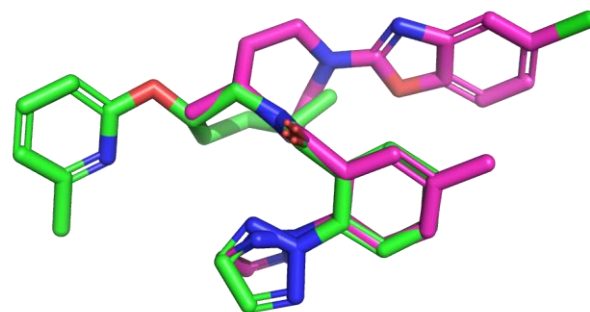
- RMSD

- from alignment methods

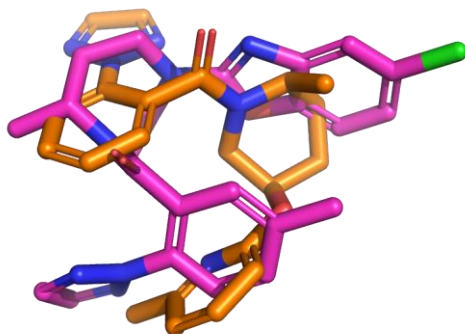
- ShapeTanimotoDist

- Compute the shape tanimoto distance between two molecule based on a predefined alignment

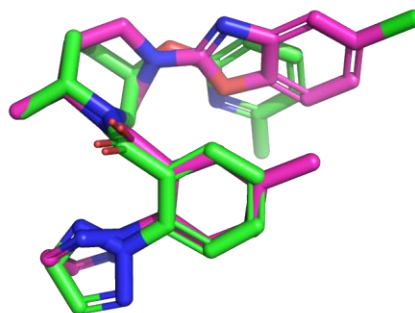
minimise rmsd



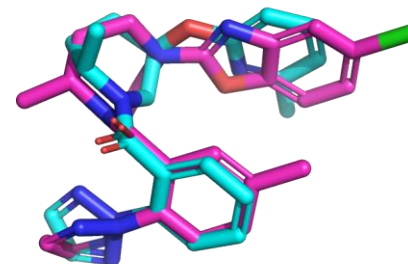
AlignMol



GetO3A



GetCrippenO3A



# Similarity – Shape-based – Python Script

```
for cid in range(0, eachMol.GetNumConformers()):
    try:
        transform_O3A=rdMolAlign.GetO3A(eachMol,eachquery,prbCid=cid,refCid=-1)
        rmsd_O3A=transform_O3A.Align()

        tani_O3A=rdShapeHelpers.ShapeTanimotoDist(eachMol,eachquery,confId1=cid,confId2=-1)

        if tani_O3A < tani_O3A_opt:
            tani_O3A_opt = tani_O3A
            bestconf_tani_O3A = cid

        if rmsd_O3A < rmsd_O3A_opt:
            rmsd_O3A_opt = rmsd_O3A
            rmsd_O3A_opt_tani = tani_O3A
            bestconf_O3A = cid
    except:
        None

    try:
        rmsd_rdkit=rdMolAlign.AlignMol(eachMol,eachquery,prbCid=cid,refCid=-1,atomMap=zip(range(eachMol.GetNumAtoms()),range(eachquery.GetNumAtoms())))
        tani_rdkit=rdShapeHelpers.ShapeTanimotoDist(eachMol,eachquery,confId1=cid,confId2=-1)

        if tani_rdkit < tani_rdkit_opt:
            tani_rdkit_opt = tani_rdkit
            bestconf_tani_rdkit = cid

        if rmsd_rdkit < rmsd_rdkit_opt:
            rmsd_rdkit_opt = rmsd_rdkit
            rmsd_rdkit_opt_tani = tani_rdkit
            bestconf_rdit = cid
    except:
        None
```

– Use futures to distribute jobs over multiple processors



# Matrix comparison – GPCR Bench dataset

ROCs Align-it

AlignMol

GetO3A

GetCrippenO3A

ROCs										Align-it										AlignMol										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETXOS										ETX									
------	--	--	--	--	--	--	--	--	--	----------	--	--	--	--	--	--	--	--	--	----------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-------	--	--	--	--	--	--	--	--	--	-----	--	--	--	--	--	--	--	--	--

## Matrix comparison – GPCR Bench dataset

[illegible]

## Shape tanimoto ranking

- AlignMol
- GetO3A
- GetCrippenO3A

Result highlighted if EF1% > 20%



## Matrix comparison – GPCR Bench dataset

[illegible]

## Shape tanimoto ranking

- AlignMol
- GetO3A
- GetCrippenO3A

RDKit results highlighted if EF1% &gt; 15%

## Matrix comparison – Comparison to ROCs

ROCI: mean												
	ligand	2ds: ROC	0981c: 1	0981a: 2	0981b: 3	0981c: 4	0981a: 5	0981b: 6	0981c: 7	0981a: 8	0981b: 9	0981c: 10
Enzyme	4040-A	2.8 E 8	1.8	1.5	1.5	1.8	1.5	1.8	1.5	1.5	1.5	1.5
ENZYME	4040-B	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-C	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-D	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-E	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-F	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-G	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-H	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-I	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-J	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-K	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-L	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-M	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-N	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-O	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-P	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-Q	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-R	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-S	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-T	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-U	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-V	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-W	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-X	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-Y	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-Z	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AA	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AB	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AC	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AD	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AF	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AG	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AH	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AI	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AJ	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AK	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AL	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AM	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AN	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AO	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AP	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AQ	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AR	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AS	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AT	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AU	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AV	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AW	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AX	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AY	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-AZ	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BA	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BB	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BC	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BD	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BF	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BG	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BH	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BI	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BJ	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BK	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BL	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BM	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BN	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BO	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BP	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BQ	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BR	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BS	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BT	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BU	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BV	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BW	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BX	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BY	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-BZ	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CA	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CB	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CC	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CD	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
ENZYME	4040-CE	1.5 E 8	1.5									

ROCS - omega				O3A tani-shape										Crippen O3A tani-shape									
Receptor	Ligand	EF1%	AUC	rdkit	ETKDG-min	ETKDG	ETKDGv2-min	ETKDGv2	rdkit	ETKDG-min	ETKDG	ETKDGv2-min	ETKDGv2	rdkit	ETKDG-min	ETKDG	ETKDGv2-min	ETKDGv2					
				EF1%	AUC	EF1%	AUC	EF1%	AUC	EF1%	AUC	EF1%	AUC	EF1%	AUC	EF1%	AUC	EF1%	AUC	EF1%	AUC		
average		15.3	0.64	8.8	0.60	9.2	0.61	8.2	0.59	9.3	0.61	8.3	0.59	8.6	0.61	9.0	0.62	8.6	0.60	9.2	0.62	8.6	0.60

CXC04	3000	32.5	8.8	1.4	8.3	23.5	8.3	28.8	4.3	23.8	8.3	23.5	8.3	19.2	8.3	22.2	8.3	42.2	4.3	35.5	8.3	36.4	8.3
CXC04	3101	32.4	8.8	1.4	8.3	23.5	8.3	28.8	4.3	23.8	8.3	23.5	8.3	19.2	8.3	22.2	8.3	42.2	4.3	35.5	8.3	36.4	8.3
CXC04	3101	32.4	8.8	1.4	8.3	23.5	8.3	28.8	4.3	23.8	8.3	23.5	8.3	19.2	8.3	22.2	8.3	42.2	4.3	35.5	8.3	36.4	8.3
EC	EC38A	10.2	8.7	19.2	8.7	47.7	8.7	19.4	8.7	47.7	8.7	19.4	8.7	47.7	8.7	19.4	8.7	7.2	8.7	19.4	8.7	19.4	8.7
EC	39A	3.5	8.5	2.8	5.1	8.4	8.5	2.8	5.1	8.4	8.5	2.8	5.1	8.4	8.5	2.8	5.1	8.4	8.5	2.8	5.1	8.4	8.5
ET8	ET33	1.5	8.8	1.5	8.8	1.5	8.8	1.5	8.8	1.5	8.8	1.5	8.8	1.5	8.8	1.5	8.8	1.5	8.8	1.5	8.8	1.5	8.8
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.4	8.4	8.4	8.4	8.4	3.8	8.7	43.8	8.7	5.3	8.4	8.4	8.4	8.4	8.4
ET8	ET39A	8.5	8.7	8.3	8.7	43.8	8.7	7.4	8.														

[illegible]

- Omega
- RDKit standard
- RDKit ETKDG
- RDKit ETKDGV2

[illegible]

# Drug Discovery Process - lite

- Program initiated (not target selection etc...)
- Cpds identified
  - patent
  - literature
- Leads examined
- Hit expansion
- Hit Analyses
- All phases - continuous
  - PAINS filters
  - Lead-like suitability
    - MPO
    - TEMPO
  - *In-silico* models
    - *Clearance, Solubility, off target panel, etc...*
    - *RF, XGboost, Bayesian etc.*
  - SA scores

# End User Accessibility

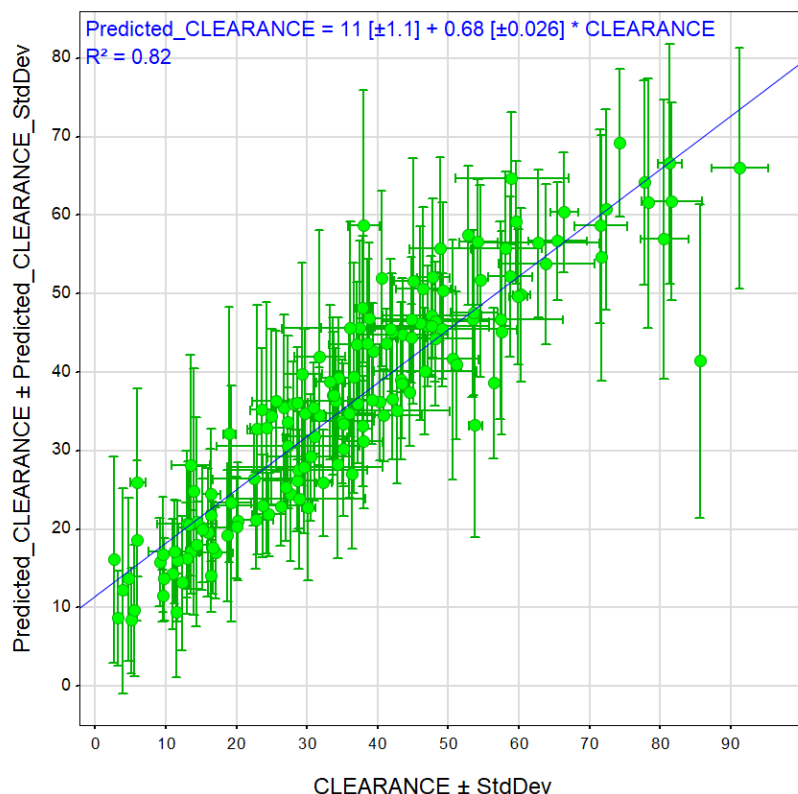
The screenshot displays the Heptares software interface, which is used for analyzing chemical compounds. The main window is titled "Heptares Calculable Properties" and features a menu bar with options like File, Add, Tools, Layout, Select, Scripts, Server, Window, and Help. The interface is divided into several panels:

- Left Panel:** Contains a "Welcome" message and a "haloperidol\_frgs.csv" file. It includes a "Type to filter" search bar, a "Click to edit structure" button, and a "Details Statistics" section showing a table with columns for "Column" and "Value". The table has two rows, with the second row showing "smiles" and "ID" values.
- Table Panel:** A table listing chemical structures (smiles) and their corresponding IDs (frag1, frag9, frag3, etc.). Each row has a checkbox for selection.
- Right Panel:** A list of calculable properties with checkboxes for selection. These include:
  - MW
  - Exact MW
  - cLogP
  - TPSA (NO)
  - Num H Acceptors (HBA)
  - Num H Donors (HBD)
  - CNS MPO (Score Only)
  - CNS TEMPO (Score Only)
  - SS GW Reactive Groups (# Only)
  - SS GW Reactive Groups (# And Component Parts)
  - SS GW Unsuitable Leads (# Only)
  - SS GW Unsuitable Leads (# And Component Parts)
  - Rat Hepatocyte Internal Model
  - Human Hepatocyte Internal Model
  - HTL Lookup
  - CNS MPO (Score Only, No Timeout)
  - CNS MPO (Score And Component Parts)
  - CNS TEMPO (Score And Component Parts)
  - CNS MPO (cachedResults)
  - CNS TEMPO (cachedResults)
  - Compound Lookup
  - ADMET Predictor [NW]
  - ADMET Predictor (All) [NW]
  - SMARTS Queries
- Bottom Panel:** A graph showing a series of green data points plotted on a grid. The x-axis is labeled "CNS MPO (Score Only, No Timeout)" and the y-axis is labeled "CNS MPO (Score And Component Parts)".

The interface also includes a "Calculate" button and a status bar at the bottom showing "rows: 20" and "Used 409.97753 MB".

# Improving Compound Success Through Models

- Rat and Human Clearance Model generation
  - data points from Rat HepClint, and data points for Human HepClint
  - 10 independently generated random forest models implemented via RDKit and SKLearn



```
class randomForestModel():
    def __init__(self, modelToLoad=None, fingerprintType='ECFP6', numberOfModels=10):
        self.__model = []
        if (modelToLoad is not None):
            # Load it in...
            # With the options...
            try:
                with gzip.open(modelToLoad, 'rb') as fileHandle:
                    (self.fingerPrintRoutine, self.fingerPrintDistance, self.__model) = pickle.load(fileHandle)
            except:
                print("Error loading model")
        else:
            self.__model = [RandomForestRegressor() for i in range(0, numberOfModels)]
            if (fingerprintType == 'ECFP6'):
                self.fingerPrintRoutine = generateMorganFingerprintsForMols
                self.fingerPrintDistance = 3
            elif (fingerprintType == 'ECFP4'):
                self.fingerPrintRoutine = generateMorganFingerprintsForMols
                self.fingerPrintDistance = 2
            elif (fingerprintType == 'ECFP2'):
                self.fingerPrintRoutine = generateMorganFingerprintsForMols
                self.fingerPrintDistance = 1
            return None

    def __del__(self):
        return None

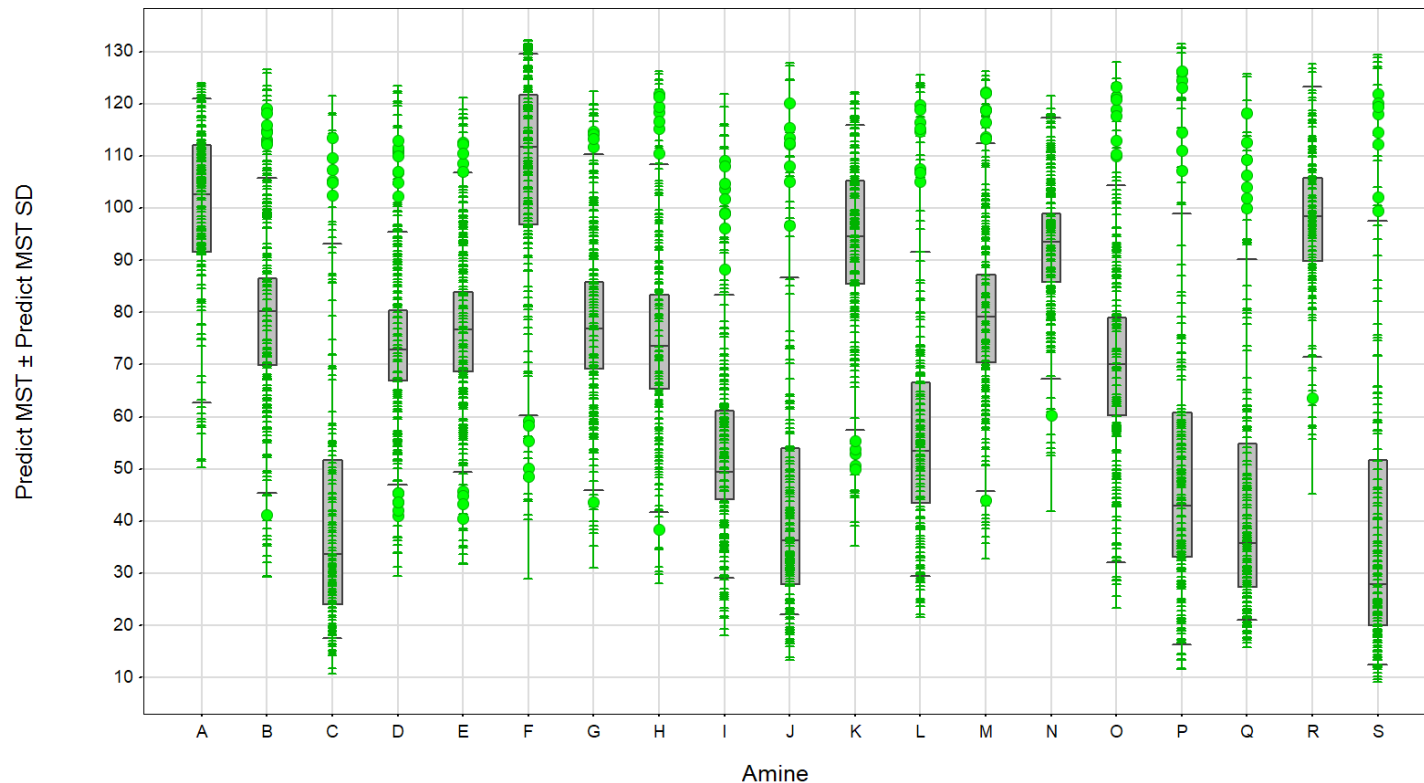
    def saveModel(self, saveFile=None):
        try:
            with gzip.open(saveFile, 'wb') as fileHandle:
                pickle.dump((self.fingerPrintRoutine, self.fingerPrintDistance, self.__model), fileHandle)
        except:
            print("Error saving model")
        return None

    def buildModel(self, propertyValuesToPredict=None, compoundList=None):
        # Build the FPs for training the model
        xFeatures = self.fingerPrintRoutine(compoundList, self.fingerPrintDistance)
        z = [eachModel.fit(xFeatures, propertyValuesToPredict) for eachModel in self.__model]
        return None

    def predictMolecules(self, moleculesToPredict=None):
        import math
        predictions = []
        meanPredictions = []
        sdPredictions = []
```

# Patent Analysis – ‘Fixing’ Sparse Data

- Do we see trends for stability in the amines?



- Amines A, F, K, N, R look interesting

# Accessibility

- All scripts centralised
  - Curated
  - Modified
  - Deposited
- doxygen
  - a tool for writing software reference documentation.
  - Doxygen is free software, released under the terms of the GNU General Public License version 2 (GPLv2).

The screenshot shows a web browser window with the URL `dockingsrv.compchem.heptares.co.uk/pythonModules/html/namespaces.html`. The page title is "Heptares Python Modules 0.1". Below the title is a navigation bar with tabs: "Main Page", "Namespaces" (selected), "Classes", and "Files". There is also a search bar. Below the navigation bar is a sub-navigation bar with "Namespace List" (selected) and "Namespace Members". The main content area is titled "Namespace List" and contains the text "Here is a list of all namespaces with brief descriptions:". Below this is a table with two columns. The first column lists namespaces, and the second column provides brief descriptions. The namespaces listed are: `PythonModules`, `_v0`, `_v0.1`, `chemistry`, `coreGeneration`, `peptides`, `databases`, `SureChEMBL`, `images`, `RDKitImageFromMol`, `molecularDiscovery`, `Moka`, `Waterflap`, `molecularNetworks`, `Corina`, `patents`, `patentAnalytics`, and `schrodinger`. The descriptions are: `.2::chemistry::ChemSketchReader` for `_v0`, and empty for the others.

Namespace	Description
<code>PythonModules</code>	
<code>_v0</code>	<code>.2::chemistry::ChemSketchReader</code>
<code>_v0.1</code>	
<code>chemistry</code>	
<code>coreGeneration</code>	
<code>peptides</code>	
<code>databases</code>	
<code>SureChEMBL</code>	
<code>images</code>	
<code>RDKitImageFromMol</code>	
<code>molecularDiscovery</code>	
<code>Moka</code>	
<code>Waterflap</code>	
<code>molecularNetworks</code>	
<code>Corina</code>	
<code>patents</code>	
<code>patentAnalytics</code>	
<code>schrodinger</code>	



# Drug Discovery Process - lite

- Program initiated (not target selection etc...)
- Cpds identified
  - patent
  - literature
- Leads examined
- Hit expansion
- Hit Analyses
- Exploit phase
  - Matched Pair Analyses
  - Bioisosteric Replacements
    - MMPdb
    - SIB, MedChemica
    - MedChemIdeas
  - Compound Enumeration

# Matched Pair Analyses

- MMP contribution

- Hussain, J., & Rea, C. (2010). "Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets." JCIM, 50(3), 339-348.
- doi:10.1021/ci900450m
- Wagener, M., & Lommerse, J. P. (2006). "The quest for bioisosteric replacements." JCIM, 46(2), 677-685.

- mmpa – github  
forked to

- mmpdb – github

- MedChemIdeas

- MedChemica

- Swiss Institute of Biotechnology
  - Reaxys integration

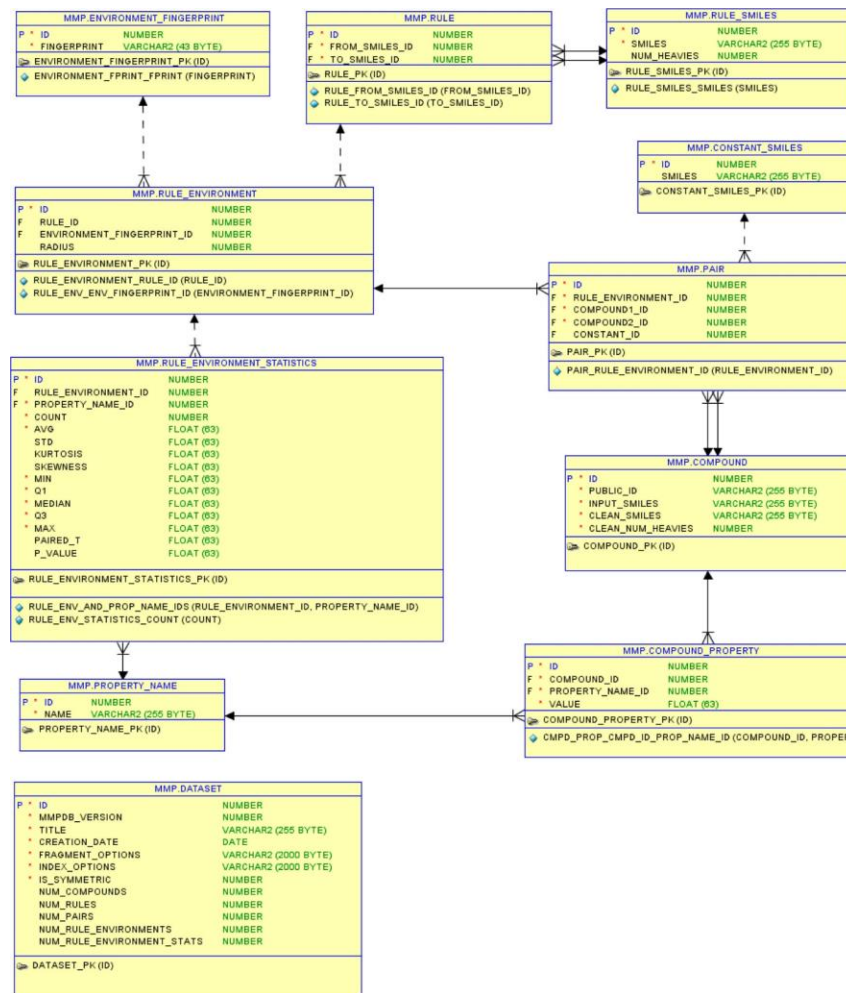
# Matched Pair Analyses

- mmpdb – github contribution

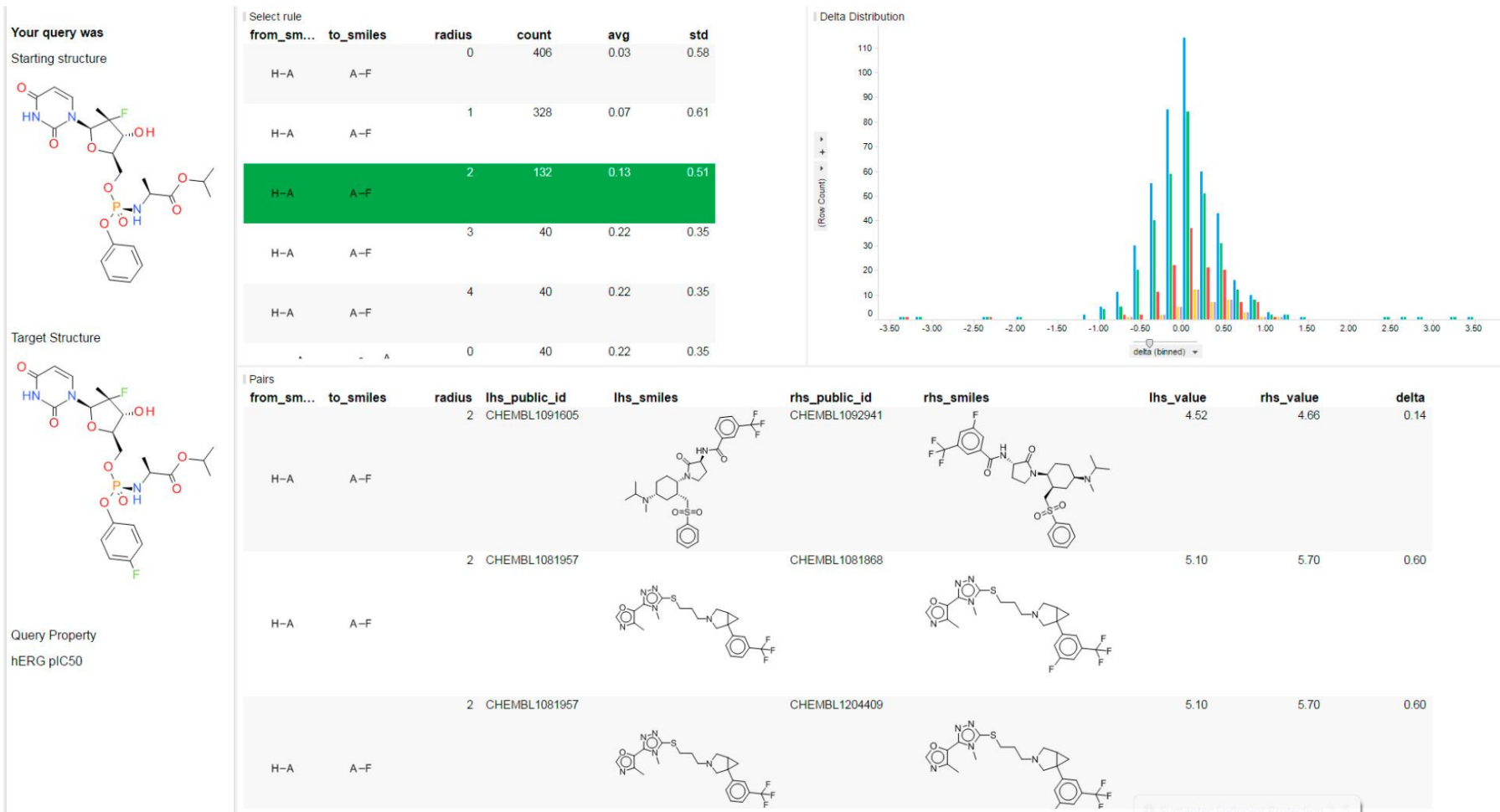
- A. Dalke, J. Hert, C. Kramer.  
mmpdb: An Open-Source Matched  
Molecular Pair Platform for Large  
Multiproperty Data Sets. *JCIM*,  
2018, 58 (5), pp 902–910.,  
doi:10.1021/acs.jcim.8b00173

## ○ Process

- mmpdb fragment
- mmpdb index
- mmpdb transform
- mmpdb predict



# Matched Pair Analyses



## Matched Pair Analyses

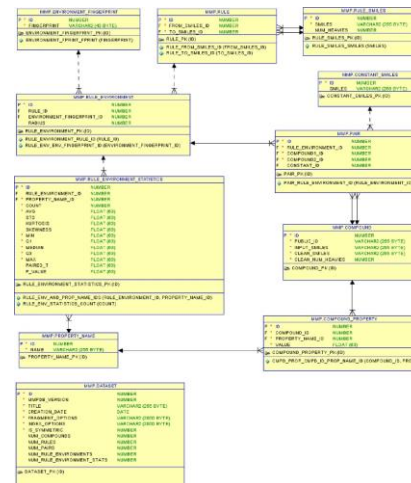
- All transformations all at once
  - with all examples etc...

```
import sqlite3
connection = sqlite3.connect('chembl-herg-var30pct.mmpdb')
cursor = connection.cursor()

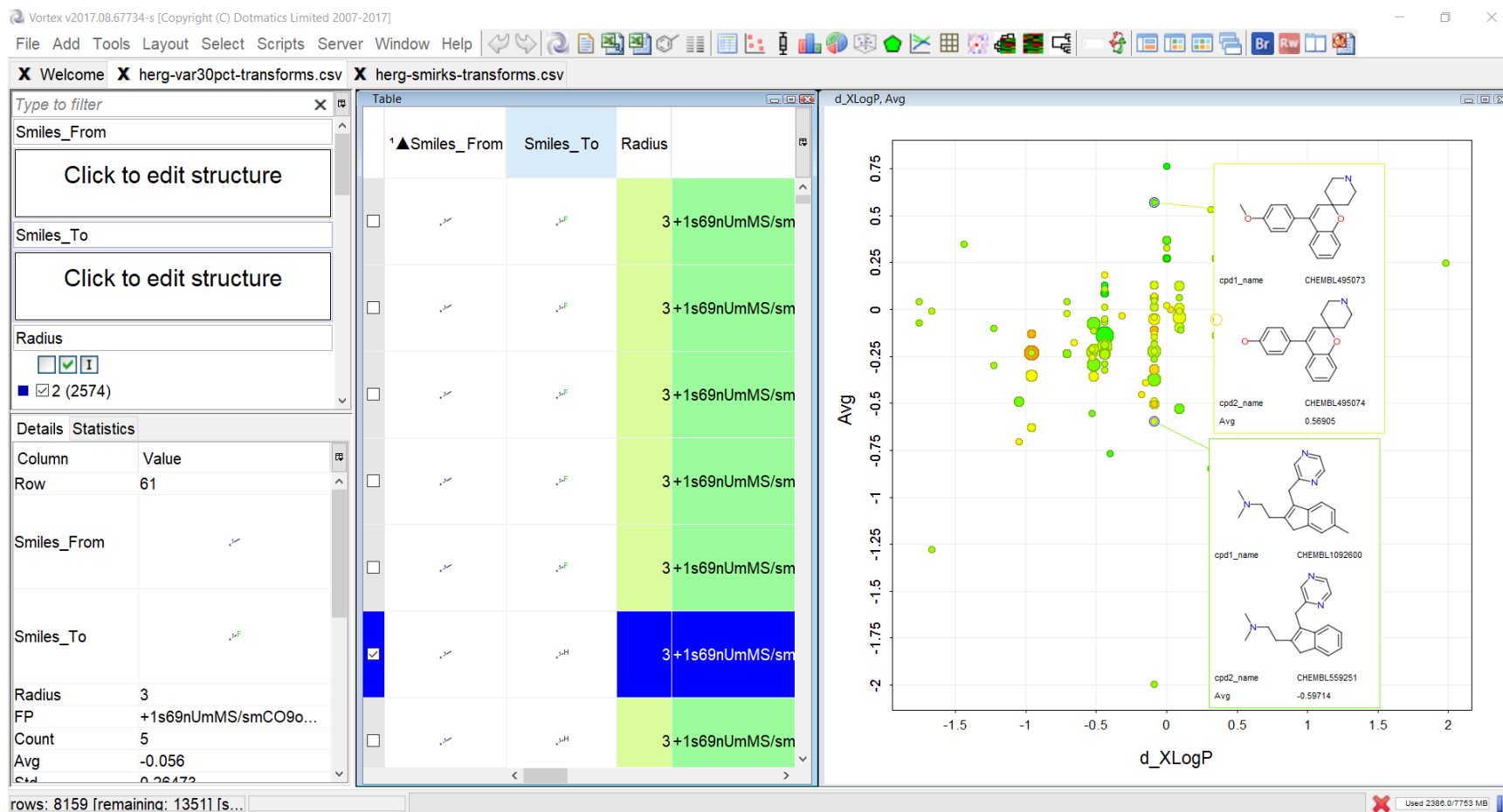
z = cursor.execute("""select
(select t4.smiles from rule_smiles t4 where t4.id = t3.from_smiles_id) as from_smiles,
(select t4.smiles from rule_smiles t4 where t4.id = t3.to_smiles_id) as to_smiles,
t2.radius, t6.fingerprint, t1.count, t1.avg, t1.std, t5.name,
(select t8.input_smiles from compound t8 where t8.id = t7.compound1_id) as cpd1_smiles,
(select t8.public_id from compound t8 where t8.id = t7.compound1_id) as cpd1_name,
(select t8.input_smiles from compound t8 where t8.id = t7.compound2_id) as cpd2_smiles,
(select t8.public_id from compound t8 where t8.id = t7.compound2_id) as cpd2_name
from rule_environment_statistics t1, rule_environment t2, rule t3,
environment_fingerprint t6, property_name t5, pair t7
where t1.count > 1 and t2.radius > 1 and t1.rule_environment_id = t2.id and t2.rule_id = t3.id
and t2.environment_fingerprint_id = t6.id and t5.id = t1.property_name_id and t2.id = t7.rule_environment_id
order by t5.name desc
""")

q=z.fetchall()
```

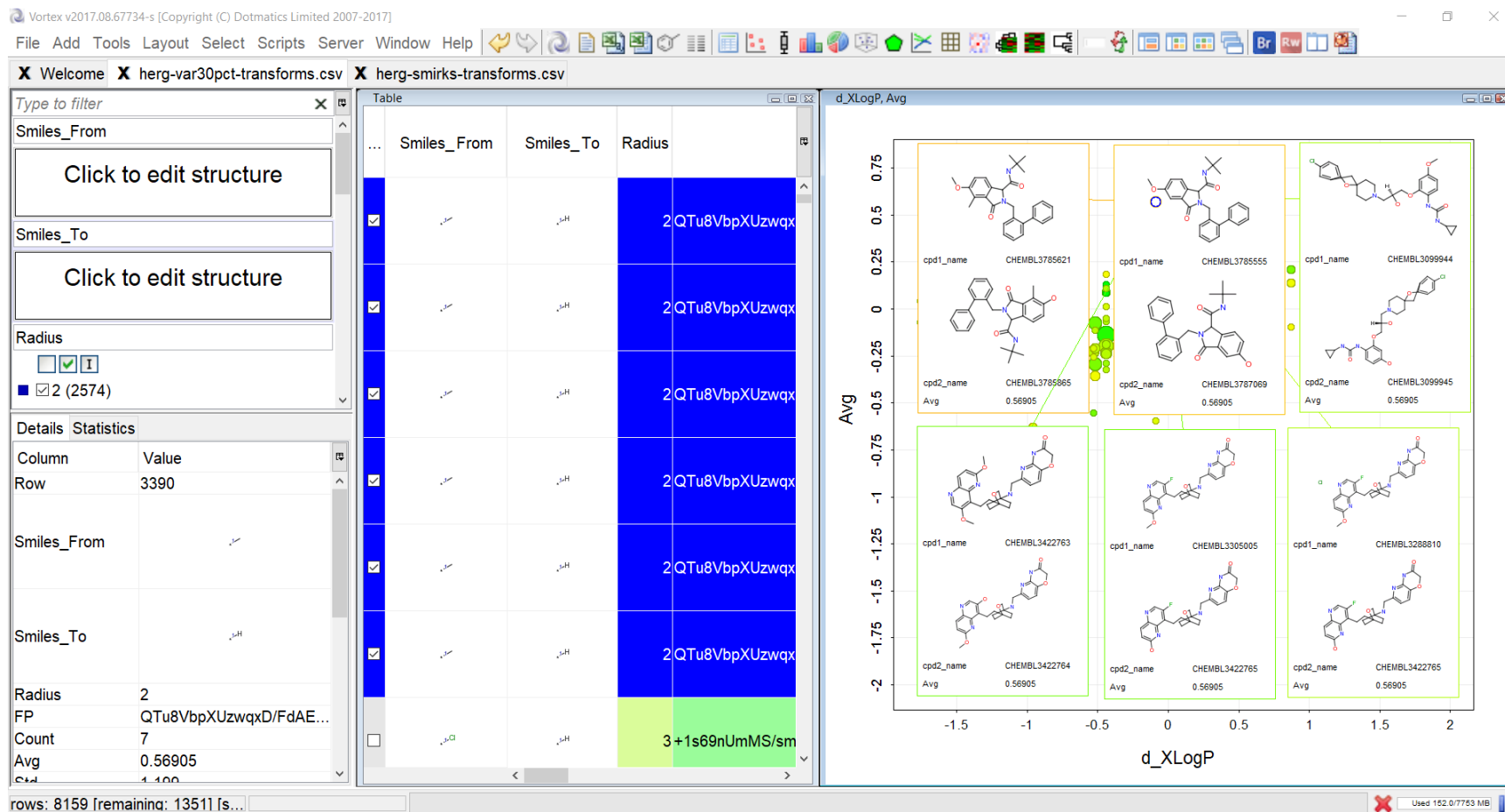
count – number of matched pair occurrences  
radius – fingerprint radius (context)



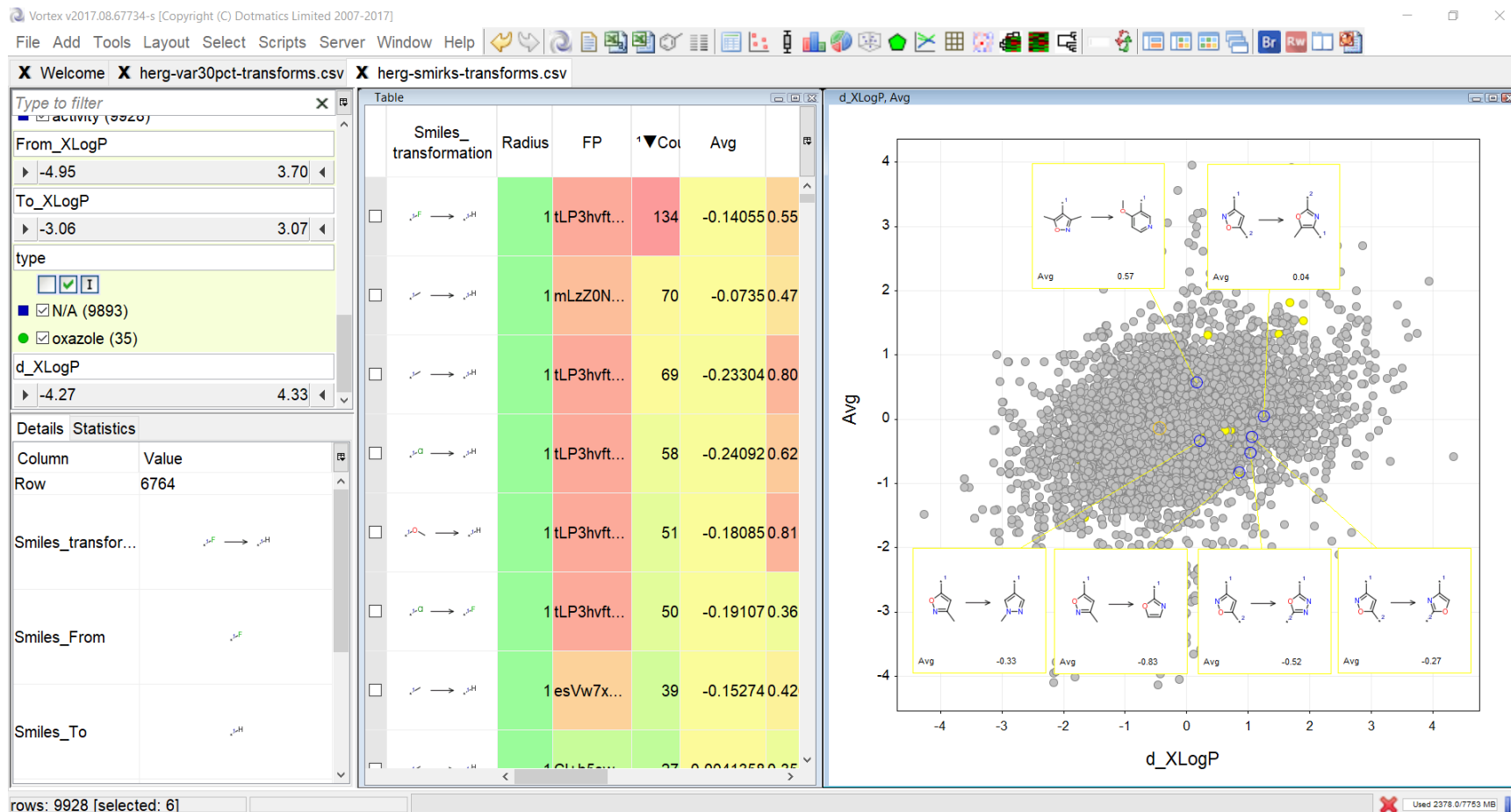
# Matched Pair Analyses



# Matched Pair Analyses



# Matched Pair Analyses



36 oxazole transformations



# Matched Pair Analyses

## ○ Bioisosteres

- see blogspot Feb2018
- encode own transformations

```
'[*:1]C(=O)OH',
'[*:1]ClOC(=O)C=C(OH)O1',
'[*:1]S(=O)(=O)[NH]C',
'[*:1]S(=O)(=O)OH',
'[*:1]C(=O)[NH]C#N',
'[*:1]c1nnn[nH]1',
'[*:1]c1nn[nH]n1',
'[*:1]c1n(C)nc([OH])c1',
'[*:1]c1onc([OH])c1',
'[*:1]c1snc([OH])c1',
'[*:1]C1=CC(=O)[NH]C1(=O)',
'[*:1]C1=NC(=O)[NH]C1(=O)',
'[*:1]ClOC(=O)[NH]C1(=O)',
'[*:1]ClSC(=O)[NH]C1(=O)'
```

- generate via automated means
  - mmpdb, etc...

```
# Isostere replacement
# Greg Landrum Feb 2018 - blogspot
# http://rdkit.blogspot.com/2018/

from rdkit import Chem
from rdkit.Chem import AllChem, Draw

isostere_smis = ('[*:1]S(=O)(=O)NC[*:2]',
                 ' [*:1]C(C(F)(F)F)NC[*:2]',
                 ' [*:1]N1N=NC([*:2])=N1',
                 ' [*:1]Cl=NOC([*:2])=N1')
isosteres = [Chem.MolFromSmiles(x) for x in isostere_smis]

def buildIsostereReaction(start,replacement):
    qps = Chem.AdjustQueryParameters()
    qps.adjustDegree = False
    qps.adjustHeavyDegree = False
    qps.adjustRingCount = False
    qps.aromatizeIfPossible = False
    qps.makeAtomsGeneric = False
    qps.makeBondsGeneric = False
    qps.makeDummiesQueries = True
    start = Chem.AdjustQueryProperties(start,qps)
    replacement = Chem.AdjustQueryProperties(replacement,qps)
    product = AllChem.ChemicalReaction()
    product.AddReactantTemplate(start)
    product.AddProductTemplate(replacement)
    product.Initialize()
    return product

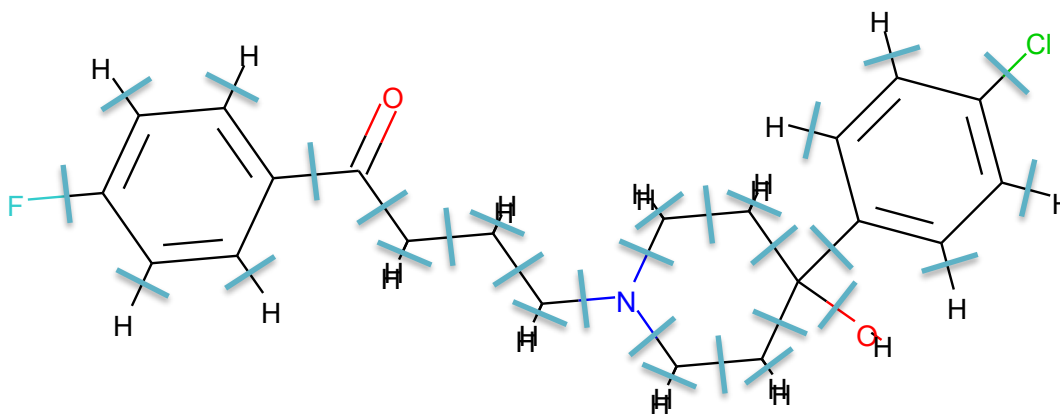
amide = Chem.MolFromSmiles('[*:1]C(=O)NC[*:2]')
isostereReactions =[buildIsostereReaction(amide,x) for x in isosteres]
isostereReactions[2]

def doIsostereReplacement(mol,rxn):
    ps = rxn.RunReactants((mol,))
    res = [x[0] for x in ps]
    return res

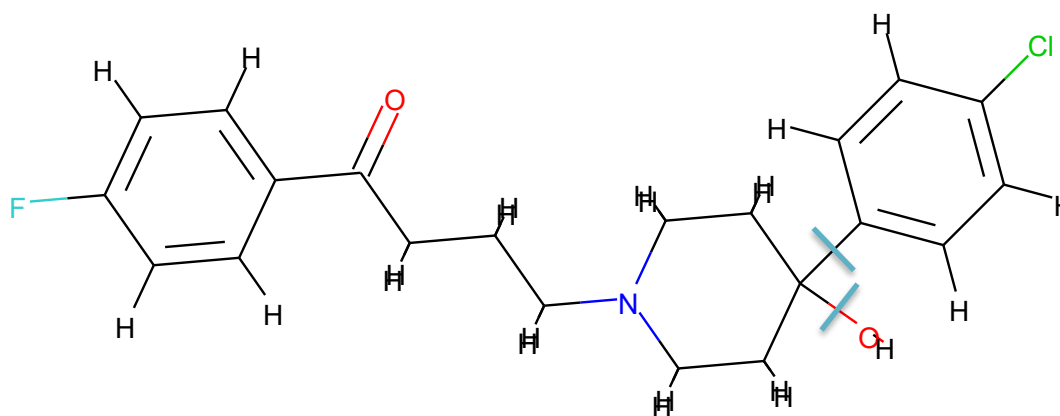
def doIsostereReplacements(mol,rxns):
    res = []
    for i,rxn in enumerate(rxns):
        seenSoFar=set()
        ims = doIsostereReplacement(mol,rxn)
        p0 = rxn.GetProductTemplate(0)
        # save where the match is
        for im in ims:
            smi = Chem.MolToSmiles(im,True)
            if smi not in seenSoFar:
                im.coreMatches = im.GetSubstructMatch(p0)
                res.append(im)
                seenSoFar.add(smi)
    return res

mol =Chem.MolFromSmiles('c1cc(F)ccc1c2(O)CCN(CC2)CCC(=O)Nc3ccccc3')
subs = doIsostereReplacements(mol,isostereReactions)
for sub in subs:
    print(Chem.MolToSmiles(sub,isomericSmiles=True))
```

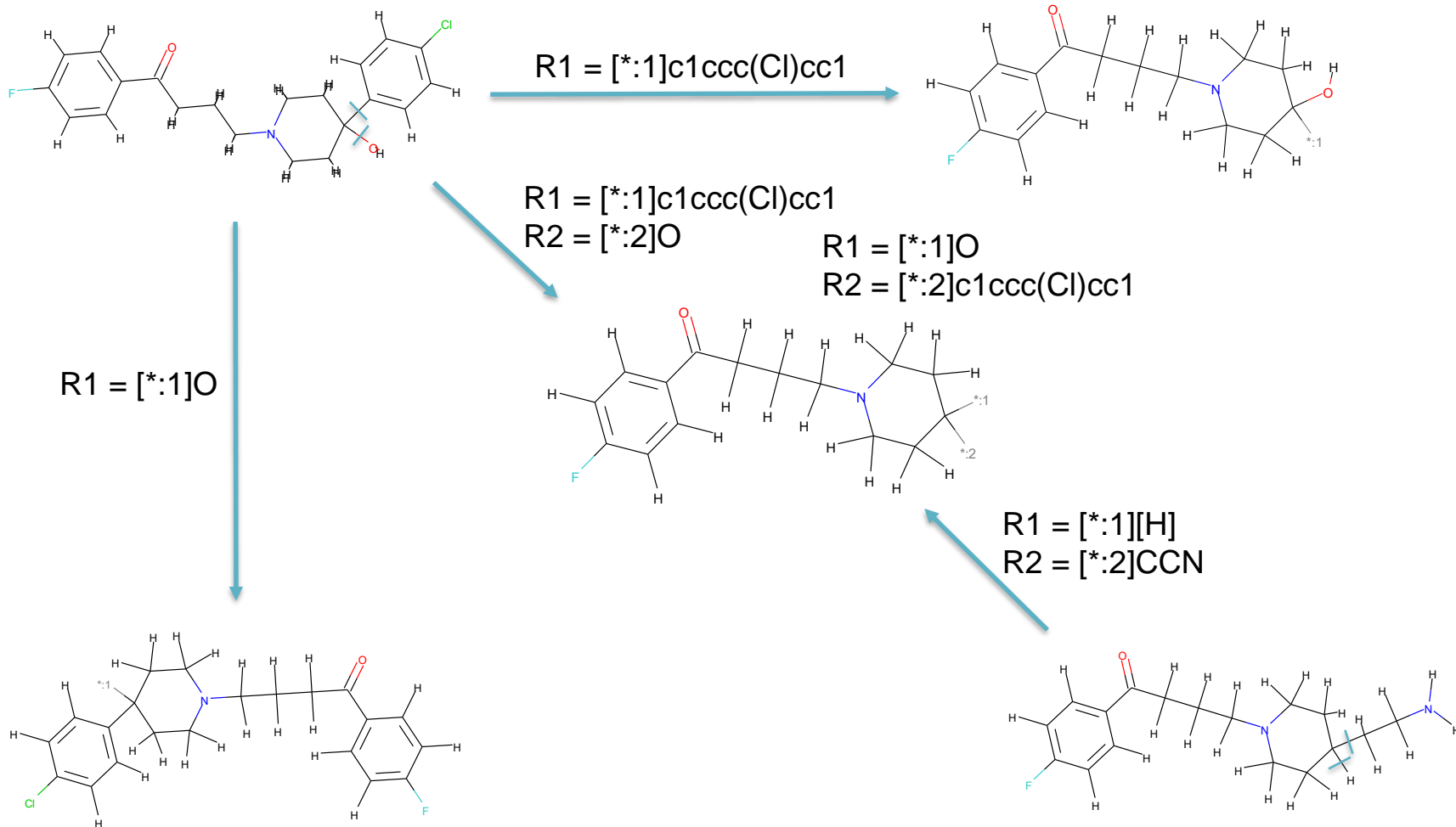
# Med Chem Ideas



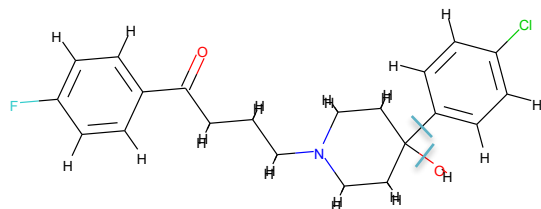
# Med Chem Ideas



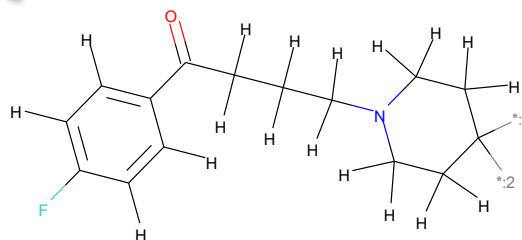
# Med Chem Ideas



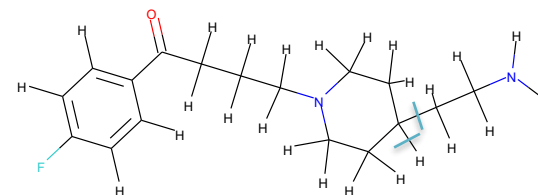
# Med Chem Ideas



R1 = [\*:1]c1ccc(Cl)cc1  
 R2 = [\*:2]O



R1 = [\*:1][H]  
 R2 = [\*:2]CCN











# Drug Discovery Process - lite

- Program initiated (not target selection etc...)
- Cpds identified
  - patent
  - literature
- Leads examined
- Hit expansion
- Hit Analyses
- Exploit phase
  - Matched Series Analyses
    - Free Wilson methods
      - additivity
    - Graph analyses
      - assessing DOA, etc...
  - Compound Docking
    - MCS ligand selection
    - Protein/ligand interactions
  - Preparing collaborative sessions

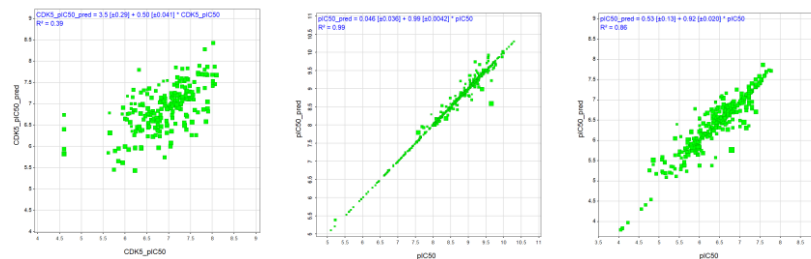
# Matched Series Analysis

- GSK, AZ, Pfizer etc...
- Pat Walters
  - <https://github.com/PatWalters/Free-Wilson>  
(subgraph match)
- Heptares
  - In-house implementation (2013)
  - RDKit, PandasTools, math
  - First principles (ReplaceCore)
  - Rewritten (2015) with numpy

- Error calculation
  - Analysis of each transformation

<sup>1</sup> ▲Rgroup	<sup>2</sup> ▲SMILES_from	<sup>3</sup> ▲SMILES_to	cpd_from	cpd_to	value	ave_value	stdev	<sup>4</sup> ▼occurrences
<input type="checkbox"/> Rgroup1			AZ20637832	AZ20637831	-1.2262	0.5594	0.72641	13
<input type="checkbox"/> Rgroup1			AZ20636864	AZ20636931	-0.58518	0.7082	0.53275	12
<input type="checkbox"/> Rgroup1			AZ20637845	AZ20637843	-0	0.82577	0.37911	11
<input type="checkbox"/> Rgroup1			AZ20637286	AZ20638064	1.4941	0.90835	0.28825	10








- Predictive ability
  - Comparative results to Hongming et al. JCI, 2013, 53, 1324–1336



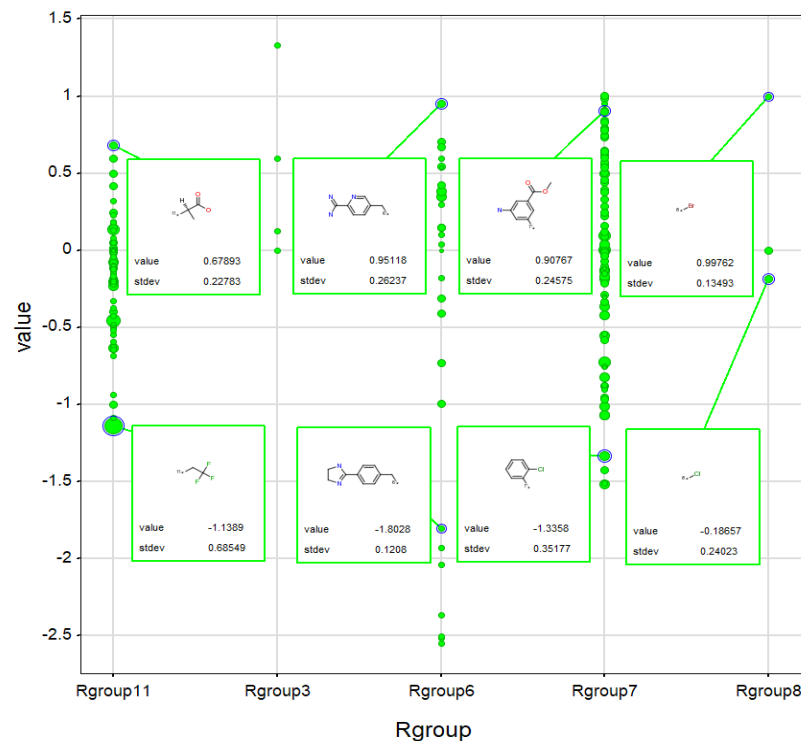
Spencer M. Free and James W. Wilson, A mathematical contribution to structure-activity studies, *J. Med. Chem* 7.4 (1964): 395-399.

# Matched Series Analysis

○ R group decomposition

Table	COMPOUND_NAME	SMILES	pIC50	smiles_Core	smiles_Rgroup3	smiles_Rgroup6	smiles_Rgroup7	smiles_Rgroup8	smiles_Rgroup11
<input type="checkbox"/>	AZ3745867		4.7959						
<input type="checkbox"/>	AZ3745907		6.6021						
<input type="checkbox"/>	AZ3745908		6.6198						
<input type="checkbox"/>	AZ3745909		6.301						
<input type="checkbox"/>	AZ3745910		6.6576						
<input type="checkbox"/>	AZ3745911		6.4318						
<input type="checkbox"/>	AZ3745912		6.5528						
<input type="checkbox"/>	AZ3745913		7.4949						

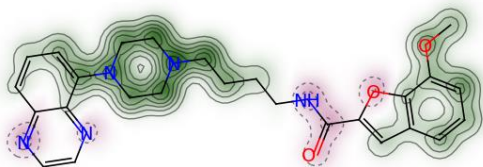
○ R group relative ranking





# Further Processing

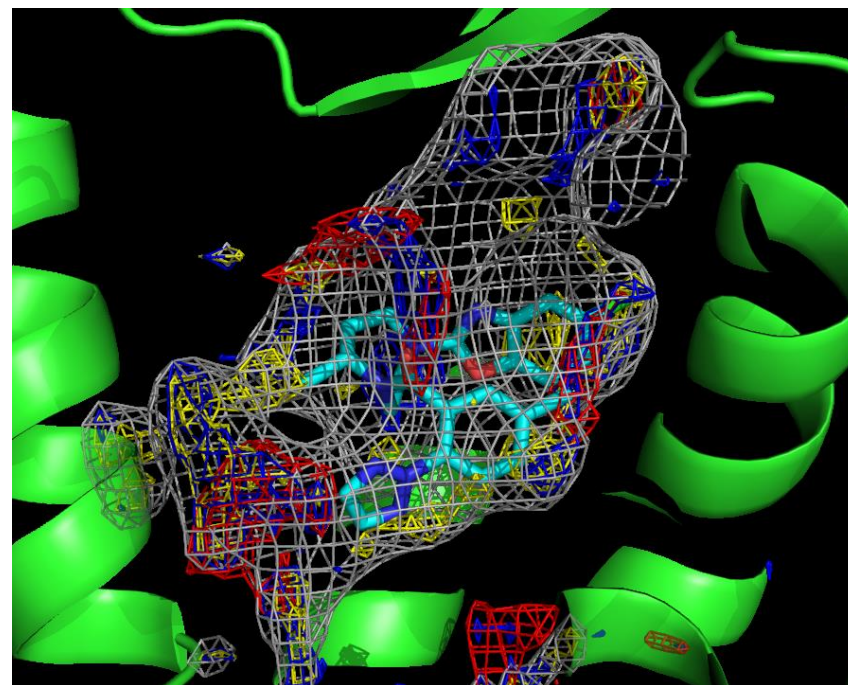
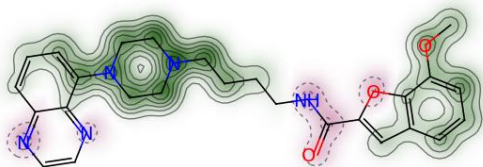
- Compound Docking
  - MCS ligand selection
  - Protein/ligand interactions
- Preparing collaborative sessions
  - Handling molecules
  - Checking duplication
  - Chemdraw -> 3D
- Database collections
  - RDKit cartridge enables substructure & similarity searching of prepared/collated collections



```
def ChemDraw_To_RDKit(fileName='barry-ideas.cdxml'):  
    allRDKitMols = []  
    allTextItems = []  
    if (os.path.splitext(fileName)[1].upper() == '.CDXML'):  
        XMLTree = ElementTree.parse(fileName)  
        XMLRoot = XMLTree.getroot()  
  
        for eachInitialTag in XMLRoot:  
            if (eachInitialTag.tag.upper() == 'PAGE'):  
                # It's a page tag...  
                for eachTag in eachInitialTag:  
                    if (eachTag.tag == 'fragment'):  
                        currentMol = Chem.RWMol()  
                        # Need a conformer to be able to set atom co-ordinates  
                        currentConformer = currentMol.GetConformer(currentMol.AddC  
                        idAtomIdxLookup = {}  
                        idBondIdxLookup = {}  
                        for eachFragmentPortion in eachTag:  
                            if (eachFragmentPortion.tag == 'b'):  
                                # A bond  
                                bondAttributes = eachFragmentPortion.attrib  
                                # Add the bond in  
                                idBondIdxLookup[int(bondAttributes['id'])] = curre  
                                if (bondAttributes.has_key('Order')):  
                                    if (bondAttributes['Order'] == '2'):  
                                        currentMol.GetBondWithIdx(idBondIdxLookup[  
                                    elif (bondAttributes['Order'] == '3'):  
                                        currentMol.GetBondWithIdx(idBondIdxLookup[  
                                else:  
                                    print(bondAttributes)  
                            else:  
                                currentMol.GetBondWithIdx(idBondIdxLookup[int(  
                                if (bondAttributes.has_key('Display')):  
                                    # There's potential wedging on this bond...  
                                    if (bondAttributes['Display'] == 'WedgeHashBe  
                                        # Can we be canny, and just adjust the ato  
                                        currentPos = currentConformer.GetAtomPosit  
                                        currentPos.z = currentPos.z - 0.5  
                                        currentConformer.SetAtomPosition(idAtomIdx  
                                        currentPos = currentConformer.GetAtomPosit  
                                        currentPos.z = currentPos.z + 0.5
```

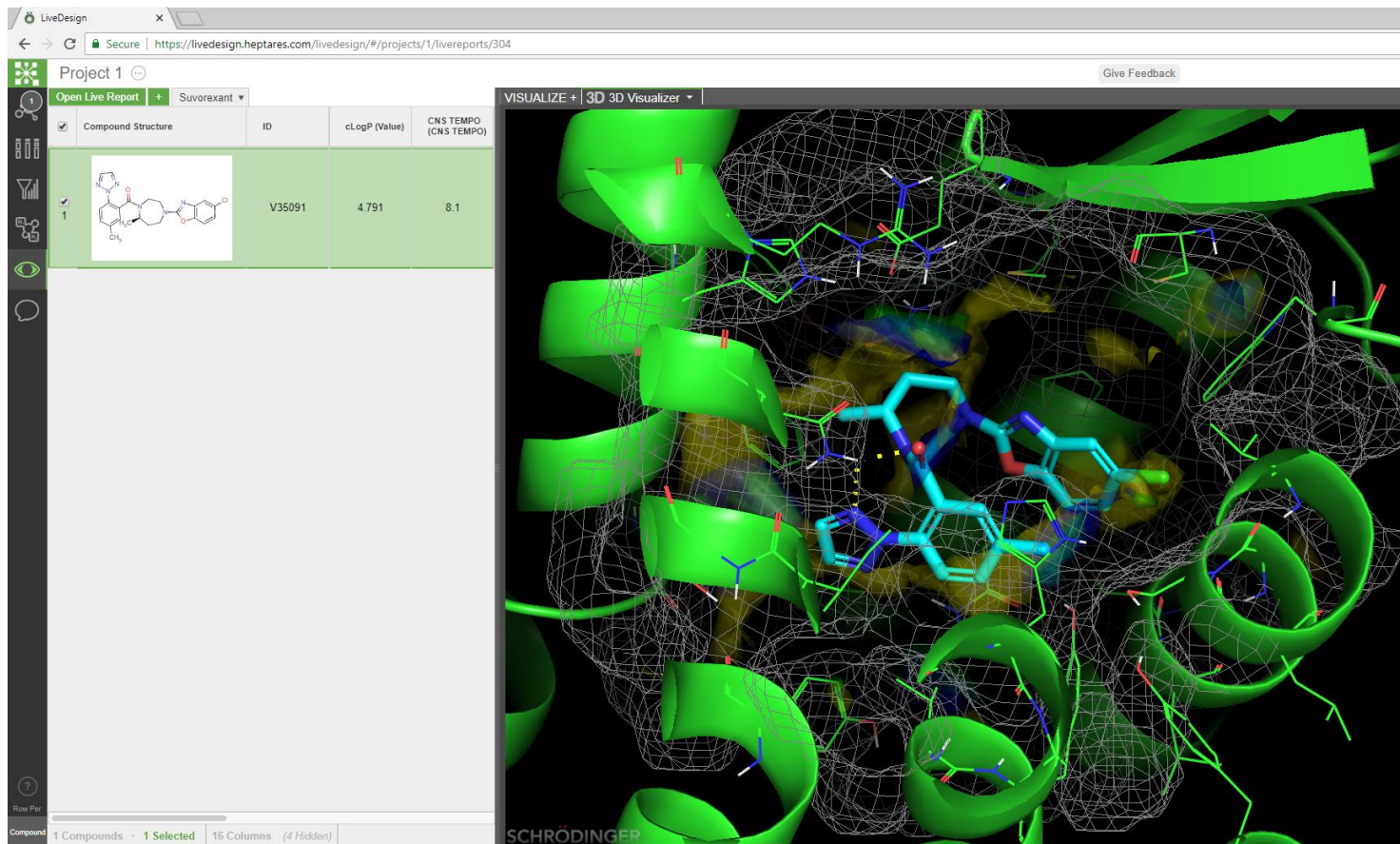
# Further Processing

- Compound Docking
  - MCS ligand selection
  - Protein/ligand interactions
- Preparing collaborative sessions
  - Handling molecules
  - Checking duplication
  - Chemdraw -> 3D
- Database collections
  - RDKit cartridge enables substructure & similarity searching of prepared/collated collections



# Structure Based Drug (Live)Design

- LiveDesign (Schrodinger)
  - RDKit being used extensively



# Acknowledgments

Greg Landrum  
RDKit community

Heptares CompChem (Current)

- Giovanni Bottegoni
- Francesca Deflorian
- Chris de Graaf
- Jon Mason
- Juan-Carlos Mobarec
- Conor Scully
- Rob Smith
- Ben Tehan

Heptares CompChem (Past)

- Andrea Bortolato
- Rob Miller
- Dahlia Weiss

Thank you for listening  
Questions?