Scientific Developments in Floes, Toolkits and Applications: Part I

Shyamal Nath

CUP 2023

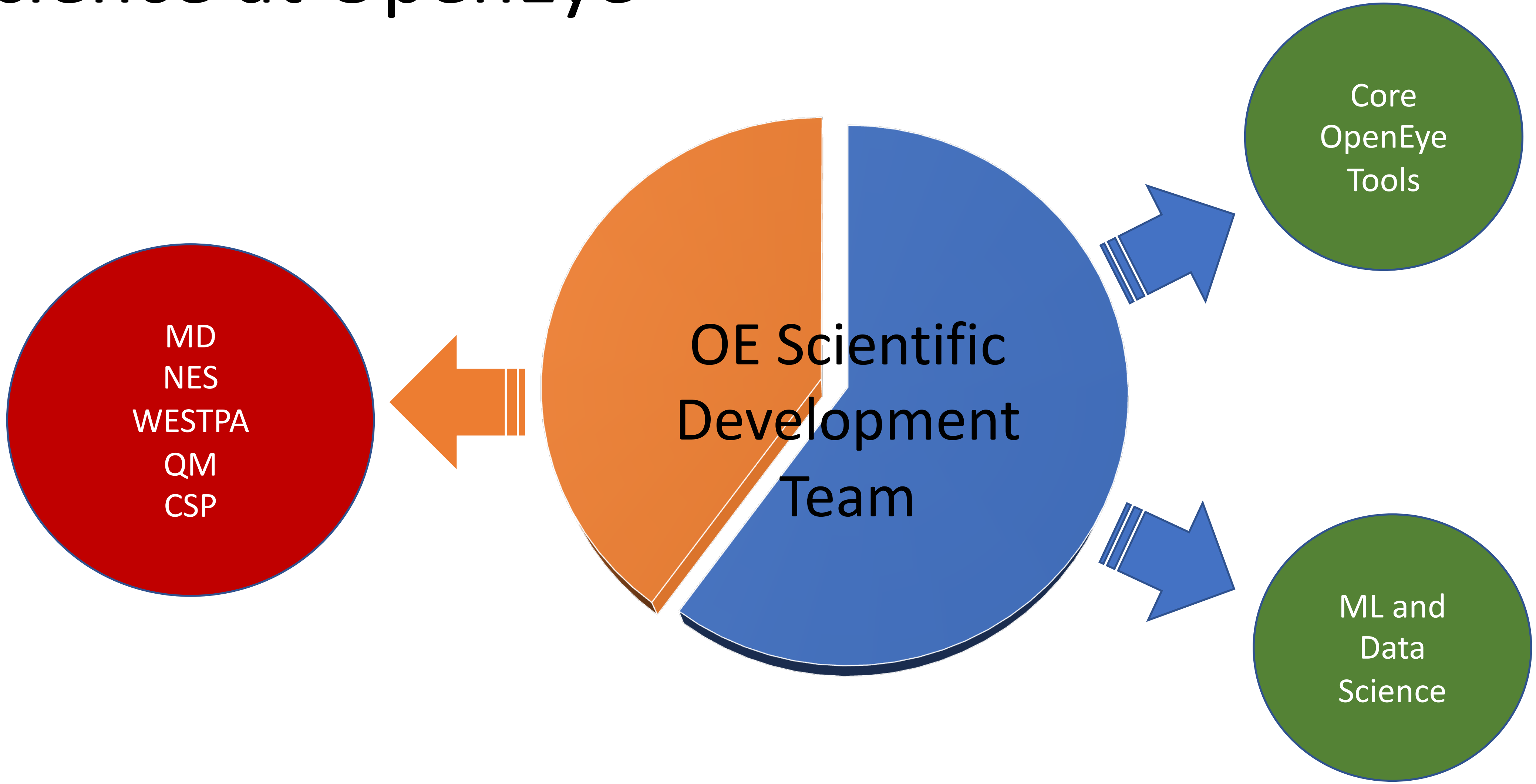
OpenEye SCIENTIFIC

CADENCE MOLECULAR SCIENCES

Introduction

- Six and Half years at OpenEye
- Molecular Modeling, Scientific Development
- Toolkits and Applications Release management
- Scientific Validation

Science at OpenEye



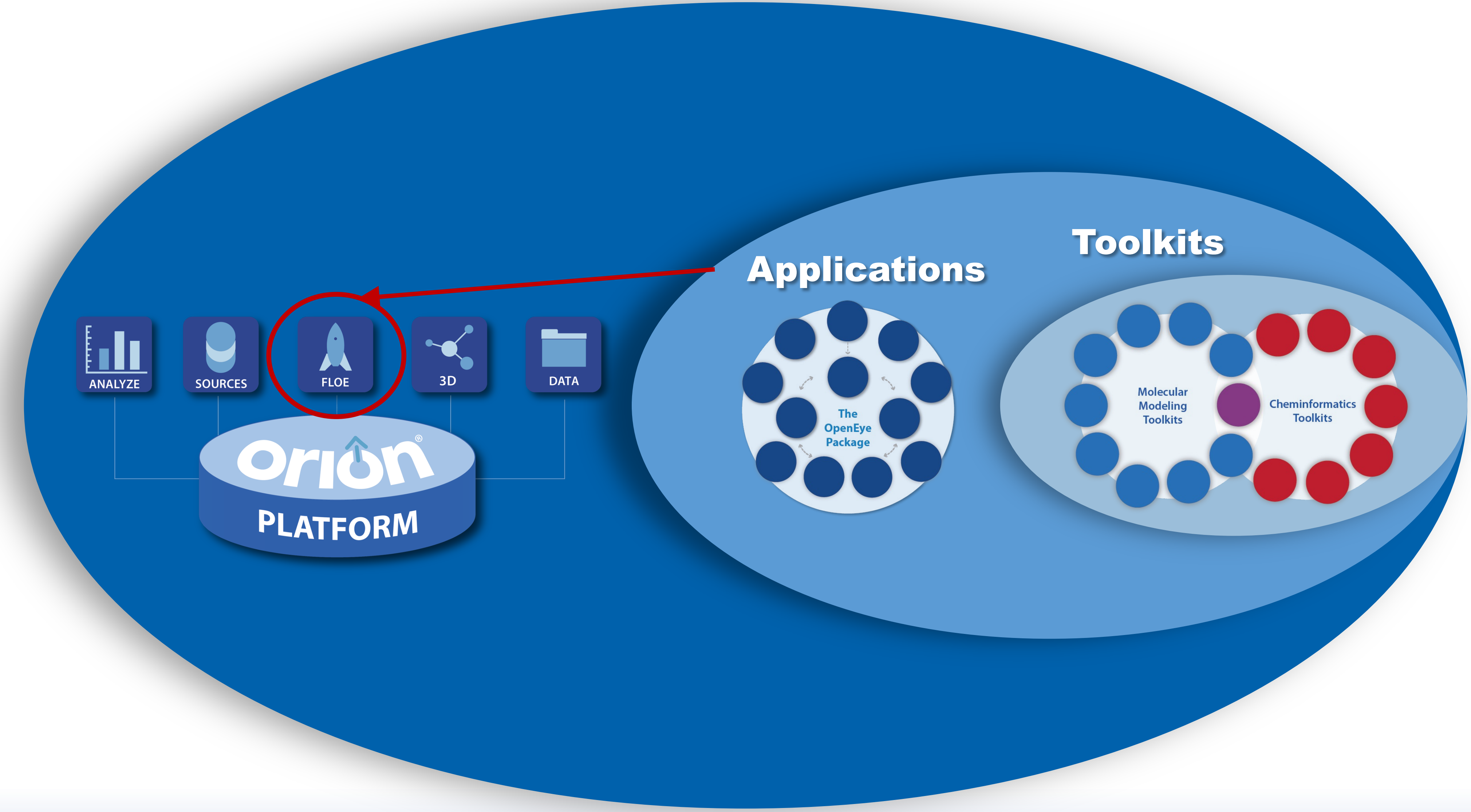
MD
NES
WESTPA
QM
CSP

OE Scientific Development Team

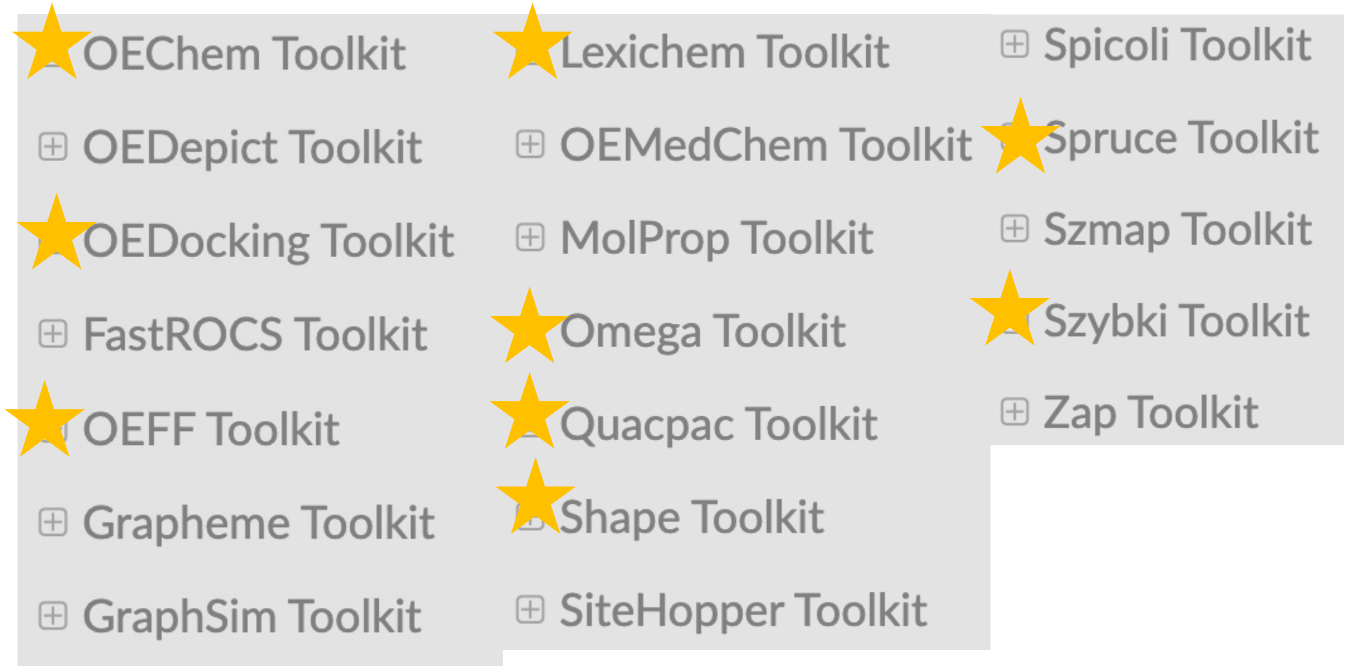
Core OpenEye Tools

ML and Data Science

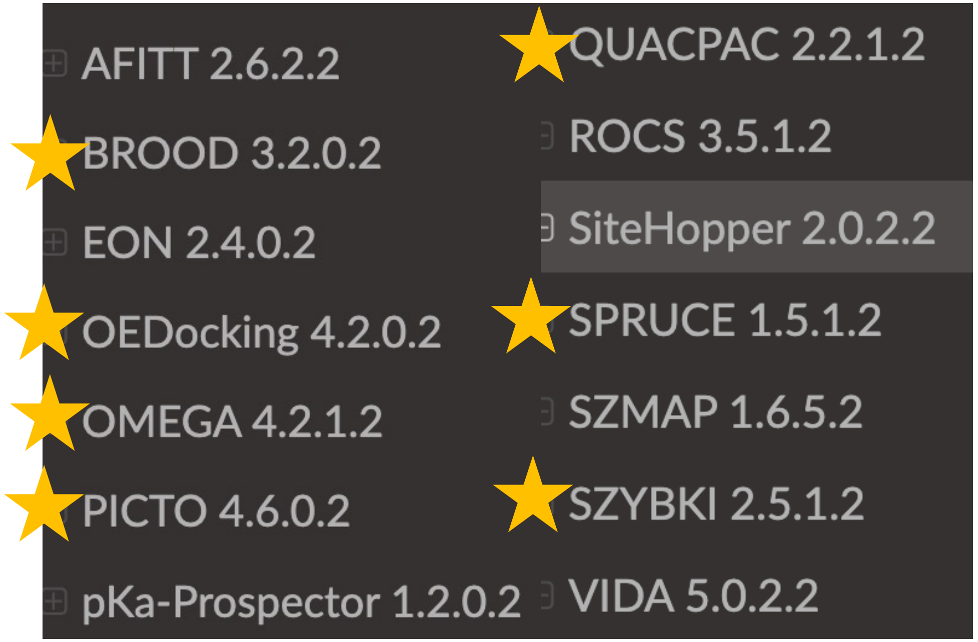
OpenEye Products



Applications and Toolkits



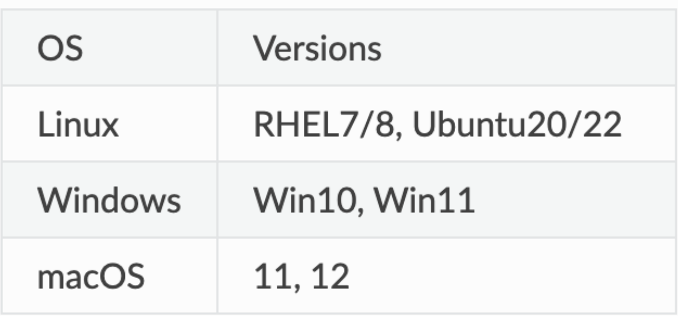
- ★ OEChem Toolkit
- OEDepict Toolkit
- ★ OEDocking Toolkit
- FastROCS Toolkit
- ★ OEFF Toolkit
- Grapheme Toolkit
- GraphSim Toolkit
- ★ Lexichem Toolkit
- OEMedChem Toolkit
- MolProp Toolkit
- ★ Omega Toolkit
- ★ Quacpac Toolkit
- ★ Shape Toolkit
- SiteHopper Toolkit
- Spicoli Toolkit
- ★ Spruce Toolkit
- Szmap Toolkit
- ★ Szybki Toolkit
- Zap Toolkit



- AFITT 2.6.2.2
- ★ BROOD 3.2.0.2
- EON 2.4.0.2
- ★ OEDocking 4.2.0.2
- ★ OMEGA 4.2.1.2
- ★ PICTO 4.6.0.2
- pKa-Prospector 1.2.0.2
- ★ QUACPAC 2.2.1.2
- ROCS 3.5.1.2
- SiteHopper 2.0.2.2
- ★ SPRUCE 1.5.1.2
- SZMAP 1.6.5.2
- ★ SZYBKI 2.5.1.2
- VIDA 5.0.2.2

Applications and Toolkits - Platforms

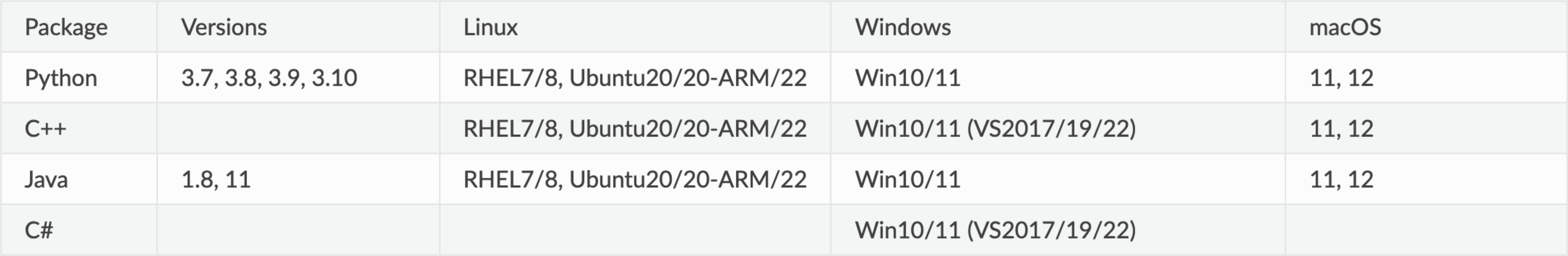
Applications



OS	Versions
Linux	RHEL7/8, Ubuntu20/22
Windows	Win10, Win11
macOS	11, 12

- macOS 12/M1
- AWS Graviton
- Ubuntu 22
- Visual Studios 2022

Toolkits



Package	Versions	Linux	Windows	macOS
Python	3.7, 3.8, 3.9, 3.10	RHEL7/8, Ubuntu20/20-ARM/22	Win10/11	11, 12
C++		RHEL7/8, Ubuntu20/20-ARM/22	Win10/11 (VS2017/19/22)	11, 12
Java	1.8, 11	RHEL7/8, Ubuntu20/20-ARM/22	Win10/11	11, 12
C#			Win10/11 (VS2017/19/22)	

Floe Packages with Core OE Science

- OpenEye Snowball Cubes
- OpenEye Classic Floes (~25 floes)
- OpenEye ChemInfo Floes (~25 floes)
- OpenEye MacroMolecular Data Service (MMDS) Floes (~10 floes)
- OpenEye Biomodeler Floes (~20 floes)
- OpenEye Large Scale Floes (~15 floes)
- OpenEye Generative Design Floes (~15 floes)
- OpenEye Large Scale Reaction Enumeration Floes (~10 floes)

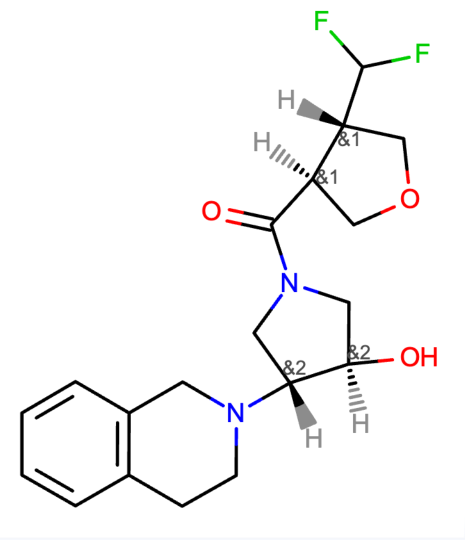
NEW! OpenEye Model Building Floes (~10 floes)

IT'S NOT A MESS, IT'S UNDER CONSTRUCTION OpenEye 3D QSAR Models Floes

OEChem TK: Enhanced Stereo Support

- Added limited support for Enhanced Stereo and CXSMILES in OEChem
  - Used by Enamine, requested by customers
- Support rolled forward into Flipper, OMEGA, Orion

Z5526878619

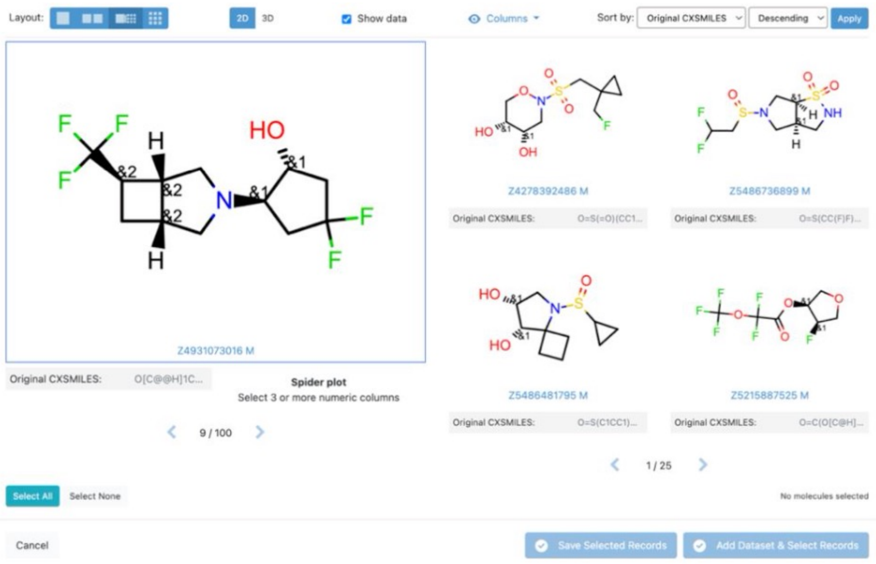


Treated as 4 *independent* stereocenters:

2**4 = **16** stereoisomers

2 independent *groups* of stereocenters:

2**2 = **4** stereoisomers



Original CXSMILES: O[C@@H]1C...

Select 3 or more numeric columns

Original CXSMILES: O=S(=O)(CC1...

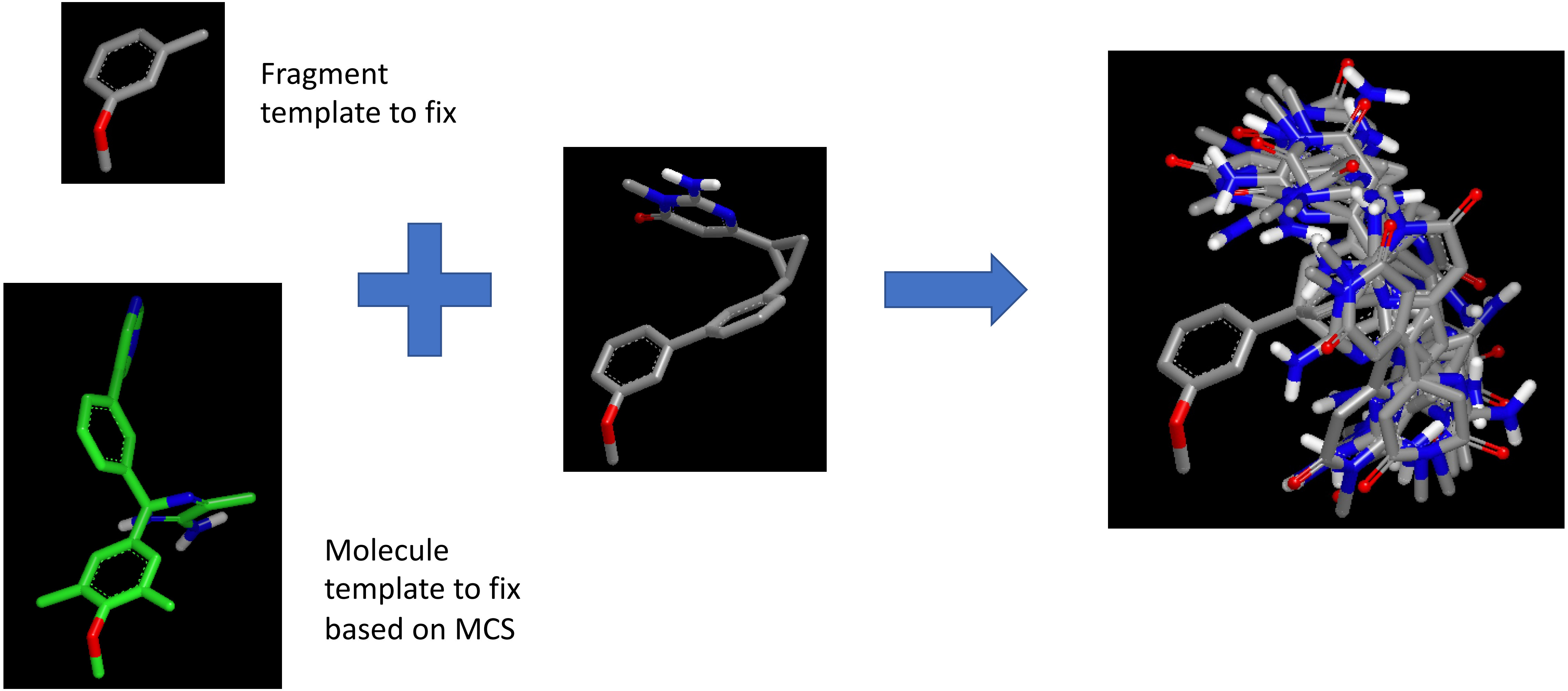
Original CXSMILES: O=S(CC(F)F)...

Original CXSMILES: O=S(C1CC1)...

Original CXSMILES: O=C(O[C@H]...

No molecules selected

OMEGA: MCS based Fix



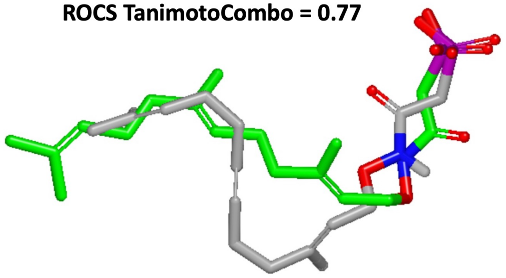
Fragment template to fix

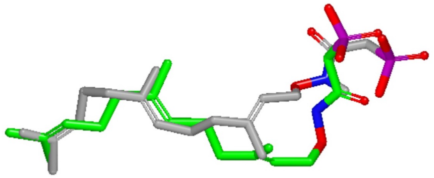
Molecule template to fix based on MCS

Shape: Flexible Overlay

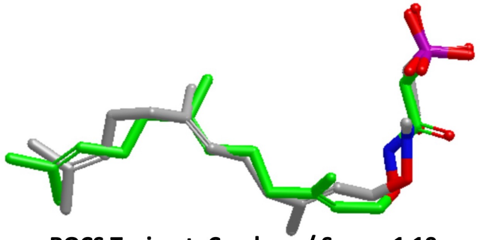
- Flexible Overlay with Shape, color and Forcefield
    - Rigid reference molecule
    - Flexible fit molecule

ROCS TanimotoCombo = 0.77



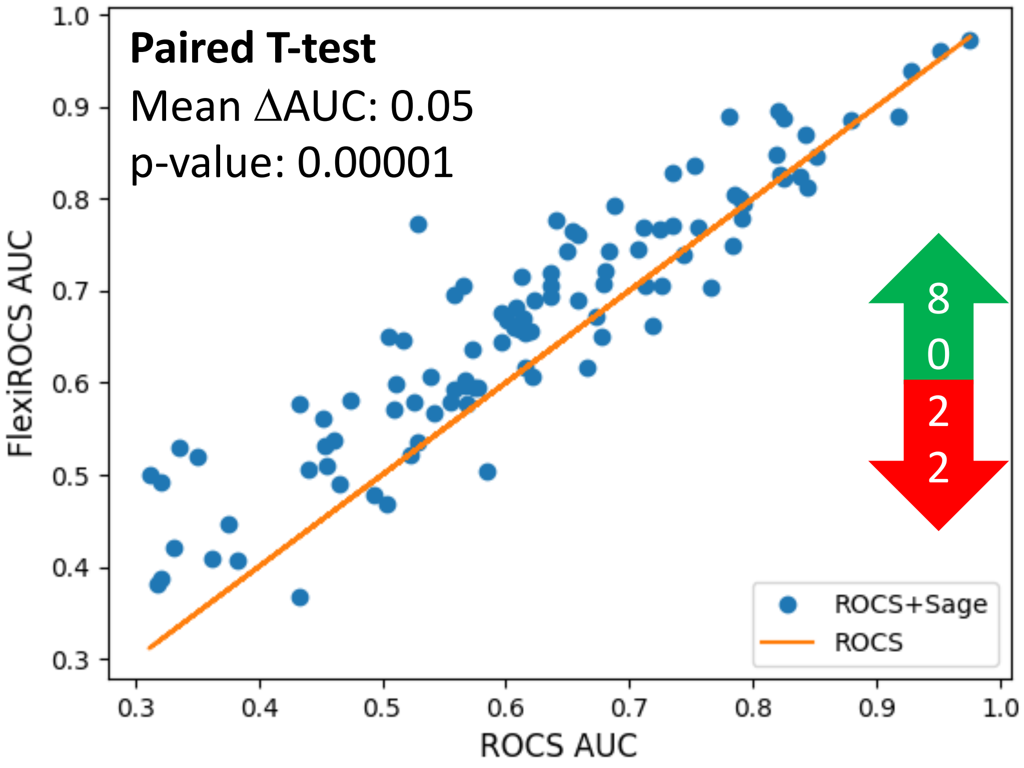


ROCS TanimotoCombo w/ MMFF = 0.91



ROCS TanimotoCombo w/ Sage = 1.19

Virtual Screening with Flexible Overlay (DUDE)



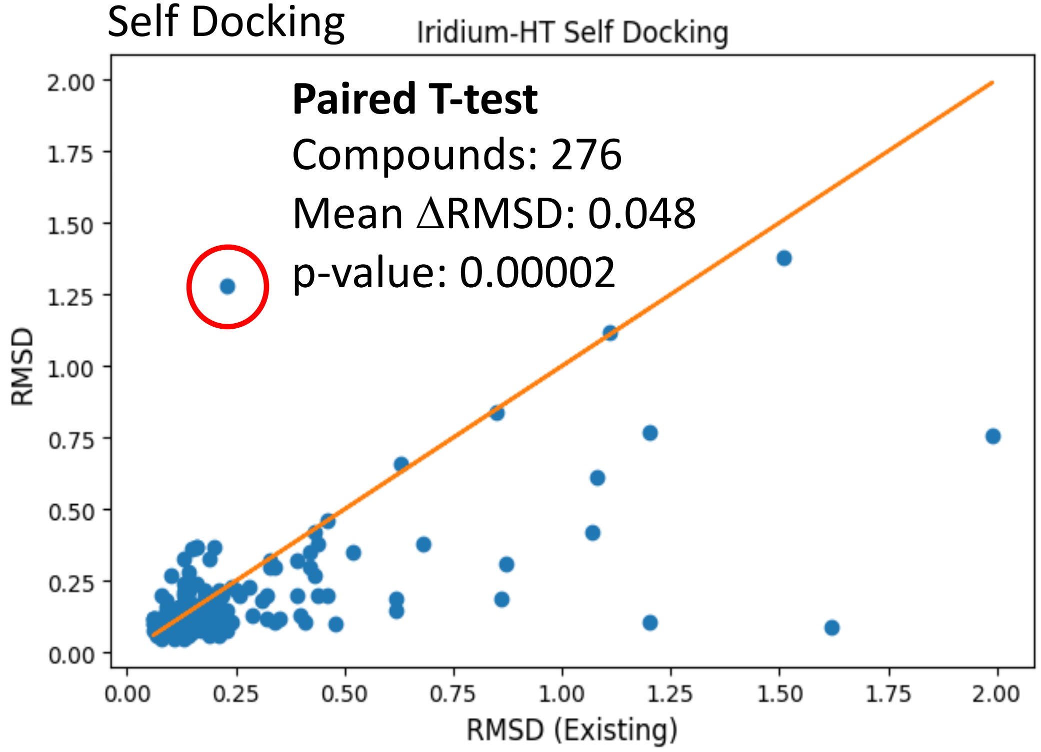
OpenEye SCIENTIFIC
CADENCE MOLECULAR SCIENCES

POSIT: Modified ShapeFit Algorithm

POSTER

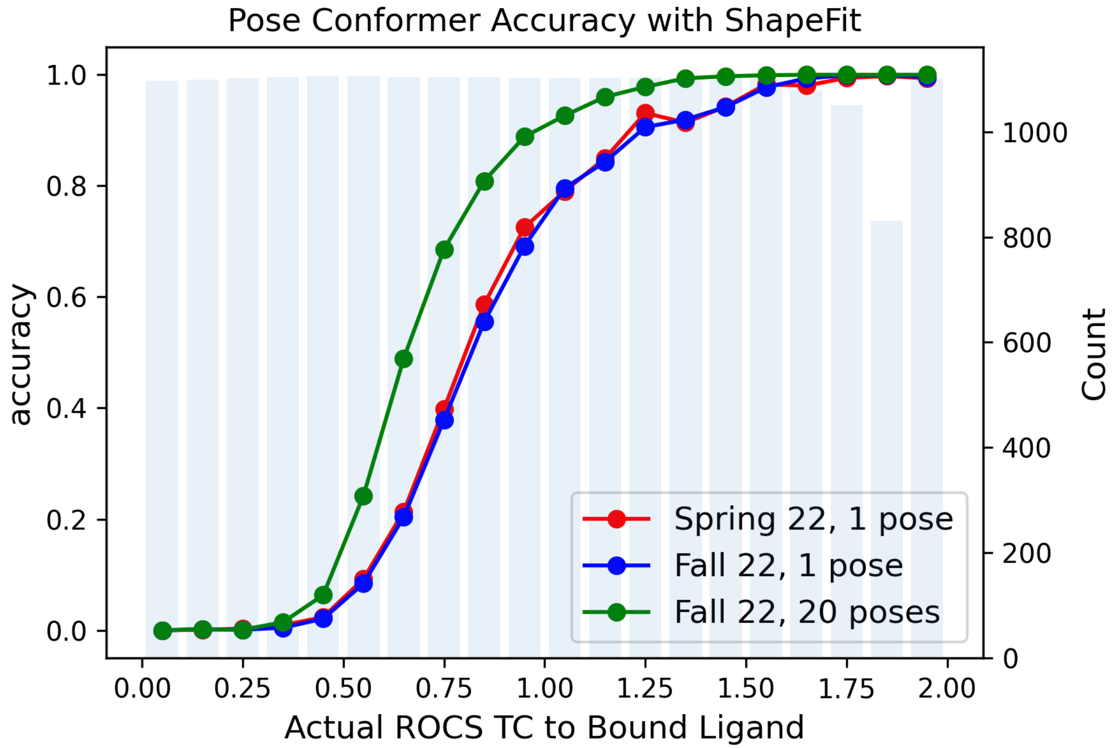
- Simultaneous optimization of force field, shape and color
- Choice of forcefield: MMFF, parsley, Sage (default)
- Multiple Pose generation

Self Docking



Paired T-test
Compounds: 276
Mean ∆RMSD: 0.048
p-value: 0.00002

Cross Docking Experiment



Warren et. al., *Drug Discovery Today* 17, 1281 (2012)

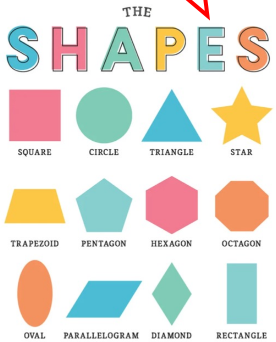
T. Tuccinardi *et al., J. Chem. Inf. Model.*, **50**, 1432 (2010)

Hermite Representation of Shape and ES

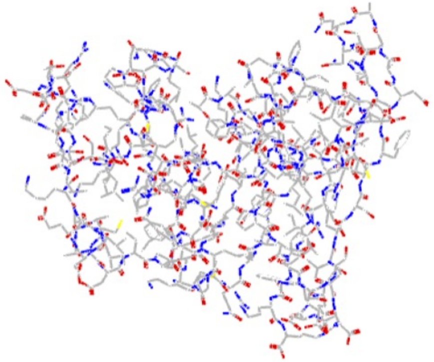
POSTER

Gaussian

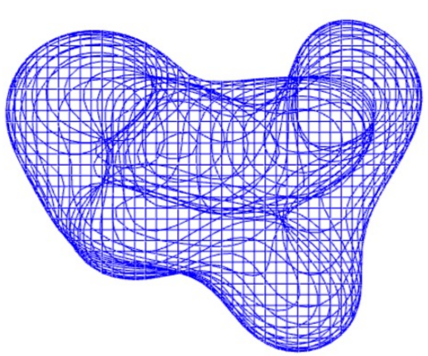
Hermite



$$\text{Shape} = \sum_i C_i \ (\text{Basis Shape})_i$$



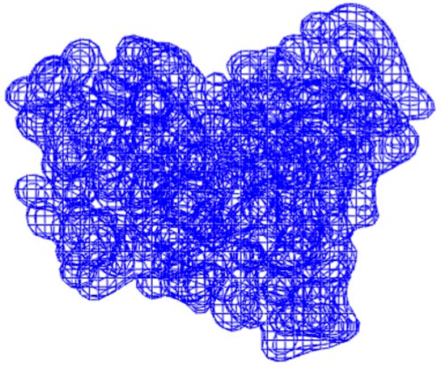
Protein view from VIDA



Low resolution Hermite Prep

NPolyMax = 5

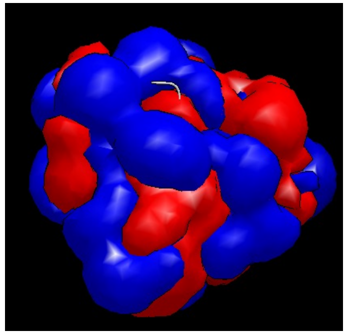
(56 Hermite coefficients kept)



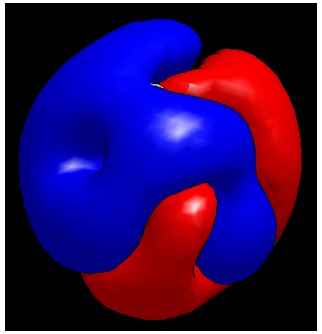
High resolution Hermite Prep

NPolyMax = 30

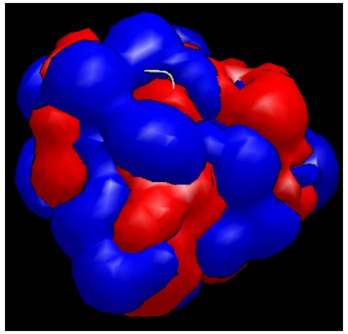
(5456 Hermite coefficients kept)



Gaussian



Hermite 5



Hermite 30

Natoms = ~1450;
sigma = 2.0A

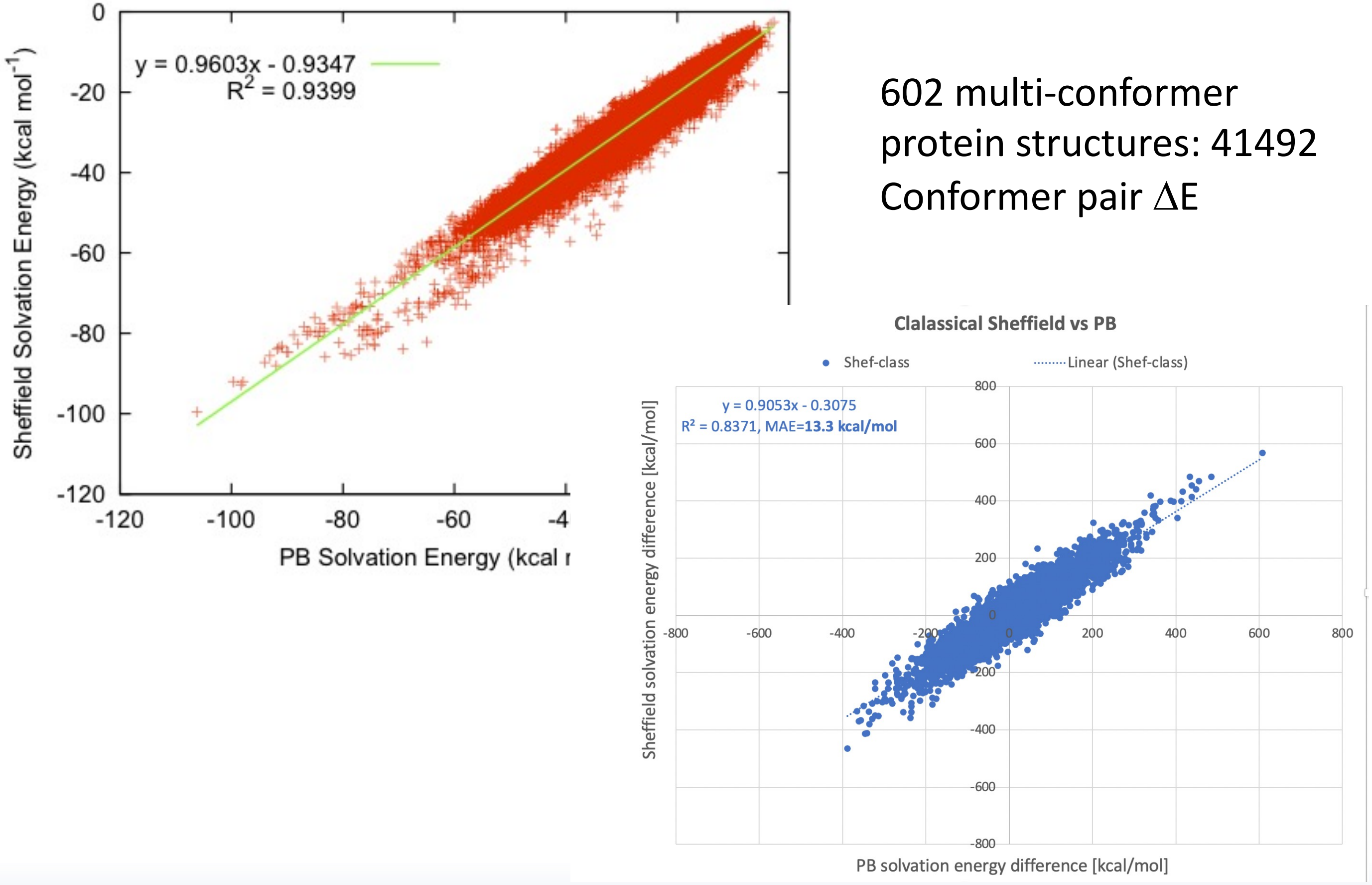
Sheffield Solvation for Proteins

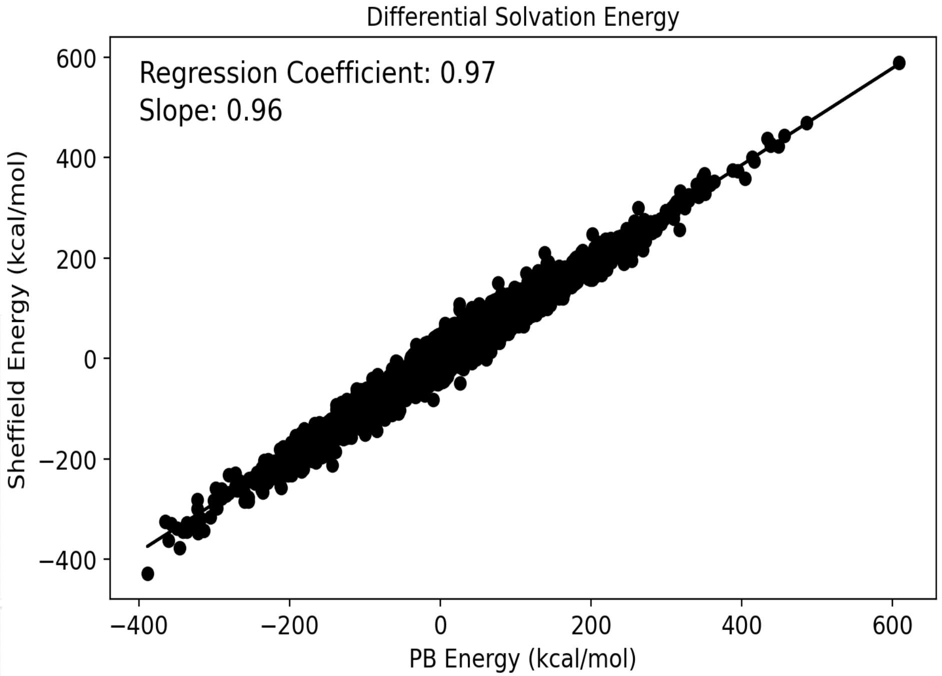
POSTER

$$E_{Shef} = -\frac{f_\epsilon}{4\pi\epsilon_0}\frac{1}{2}\sum_{i,j}\frac{Q_iQ_j}{\sqrt{A\sigma_i\sigma_j + BR_{ij}^2}}$$

$$\sigma_i = \sigma_{0,i} + C\left(1 - \sqrt{\frac{S_i}{S_{0,i}}}\right)$$

602 multi-conformer protein structures: 41492
Conformer pair ΔE





More on Toolkits and Applications

SPRUCE

- BioAssembly factory 5-10x speed-up
- Standardizing inputs before Spruce
- Filtering out Bad output structures

BROOD POSTER

- Two new BROOD Databases based on ChEMBL 31

QUACPAC POSTER

- Multi-state pKa Model

PICTO

- Multiple-molecule handling
- Hydrogen Display
- Atom labeling

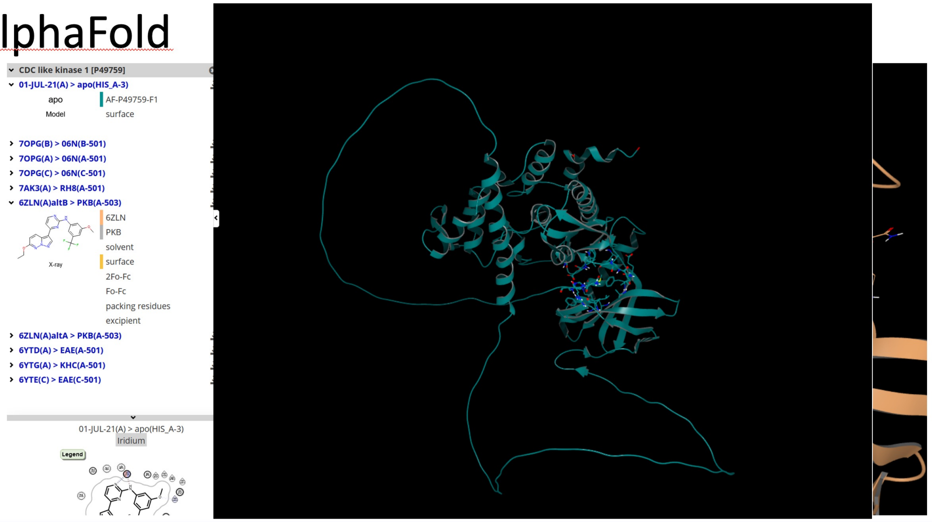
SZYBKI

- Protein-ligand optimization with FF14SB-OpenFF and PB/Sheffield

MacroMolecular Data Service (MMDS)

- 105,779 PDB experiments were prepared -> ~200,000 design units
- 1,325 models from AlphaFold2 were prepared and added
- Organized in MMDS
  - Guide to pharmacology tree
  - Uncategorized tree
  - Currently at 12,728 targets

AlphaFold

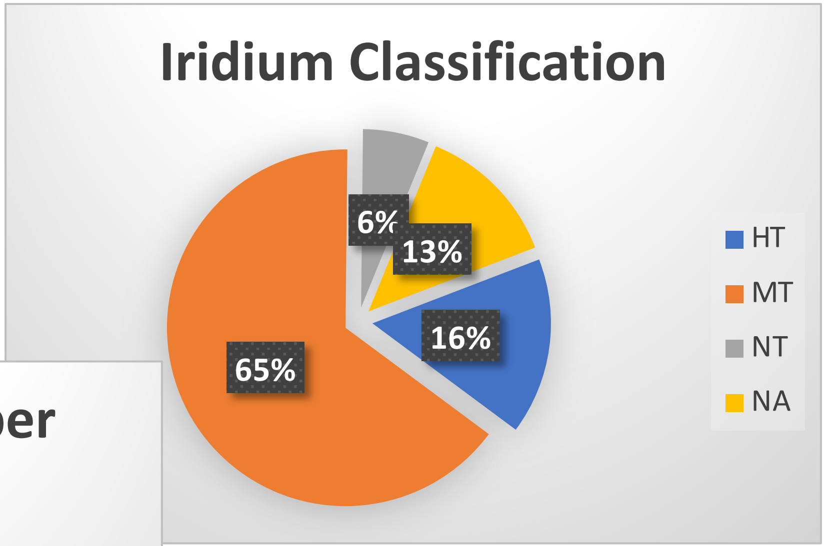


CDC like kinase 1 [P49759]
- 01-JUL-21(A) > apo(HIS_A-3)
  - apo Model: AF-P49759-F1, surface
- 7OPG(B) > 06N(B-501)
- 7OPG(A) > 06N(A-501)
- 7OPG(C) > 06N(C-501)
- 7AK3(A) > RH8(A-501)
- 6ZLN(A)altB > PKB(A-503)
  - X-ray: 6ZLN, PKB, solvent, surface, 2Fo-Fc, Fo-Fc, packing residues, excipient
- 6ZLN(A)altA > PKB(A-503)
- 6YTD(A) > EAE(A-501)
- 6YTG(A) > KHC(A-501)
- 6YTE(C) > EAE(C-501)

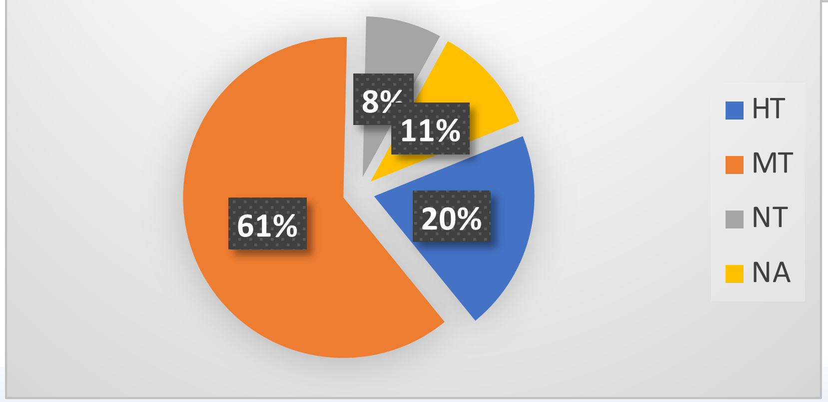
01-JUL-21(A) > apo(HIS_A-3)

Iridium

Iridium Classification



Iridium Classification per PDB code

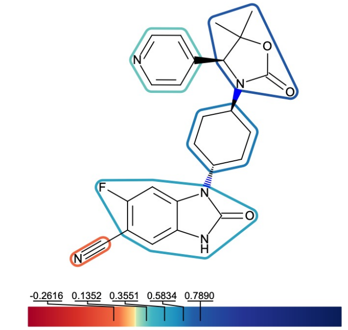


Model Building Floes

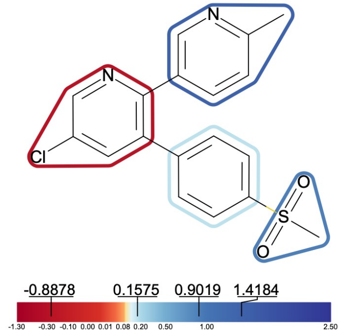
POSTER

- Shipped general Model Building Floes in Orion®
    - General ML Models on 2D Fingerprints
    - Explainable Predictions
    - Confidence and Domain of Application

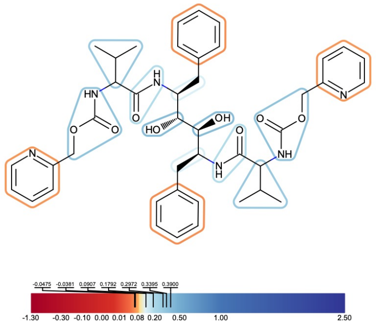
Solubility Pred.: 1.601 log(uM)



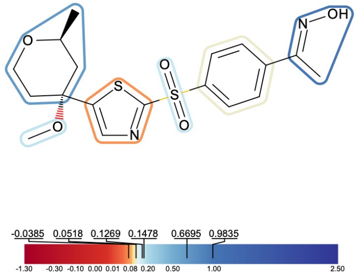
Solubility Pred.: 1.59 log(uM)



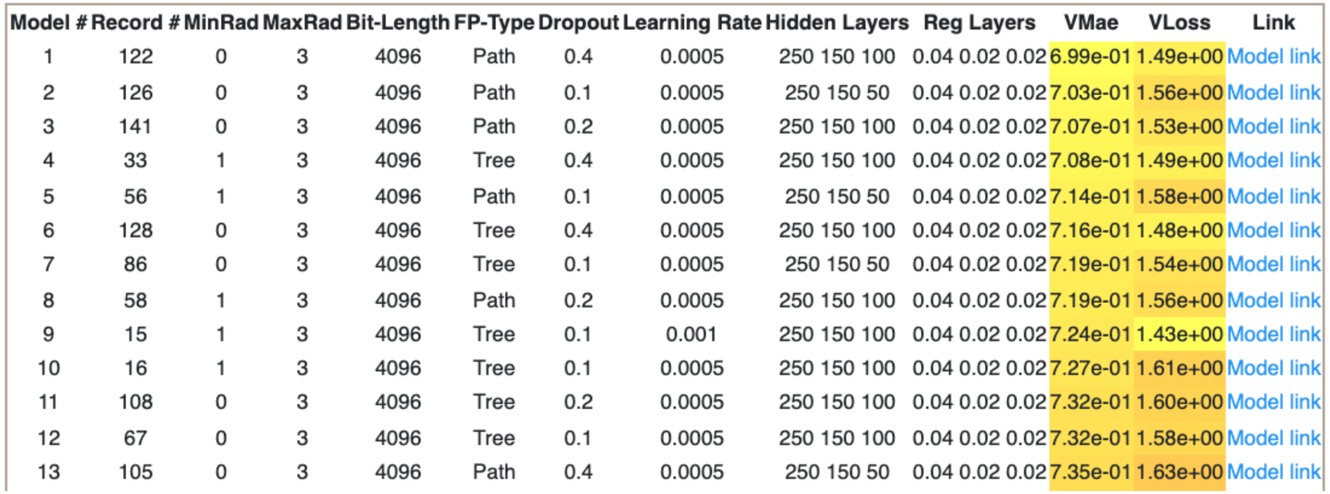
Solubility Pred.: 2.422 log(uM)



Solubility Pred.: 1.941 log(uM)



List of Fully-connected Neural Network Models Generated

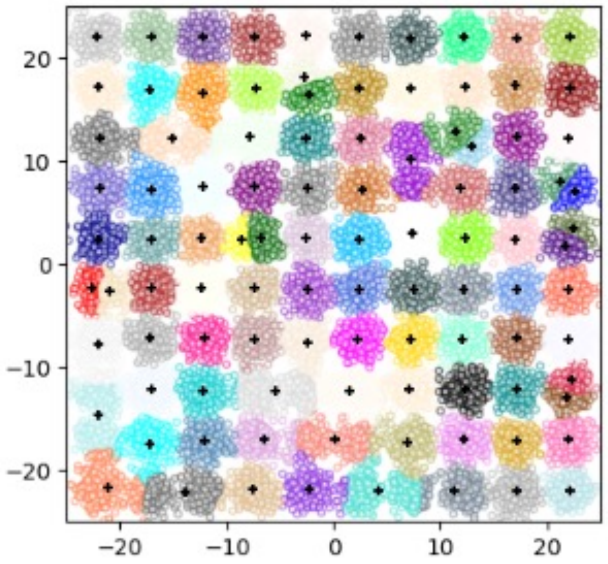


Model #	Record #	MinRad	MaxRad	Bit-Length	FP-Type	Dropout	Learning Rate	Hidden Layers	Reg Layers	VMae	VLoss	Link
1	122	0	3	4096	Path	0.4	0.0005	250 150 100	0.04 0.02 0.02	6.99e-01	1.49e+00	Model link
2	126	0	3	4096	Path	0.1	0.0005	250 150 50	0.04 0.02 0.02	7.03e-01	1.56e+00	Model link
3	141	0	3	4096	Path	0.2	0.0005	250 150 100	0.04 0.02 0.02	7.07e-01	1.53e+00	Model link
4	33	1	3	4096	Tree	0.4	0.0005	250 150 100	0.04 0.02 0.02	7.08e-01	1.49e+00	Model link
5	56	1	3	4096	Path	0.1	0.0005	250 150 50	0.04 0.02 0.02	7.14e-01	1.58e+00	Model link
6	128	0	3	4096	Tree	0.4	0.0005	250 150 100	0.04 0.02 0.02	7.16e-01	1.48e+00	Model link
7	86	0	3	4096	Tree	0.1	0.0005	250 150 50	0.04 0.02 0.02	7.19e-01	1.54e+00	Model link
8	58	1	3	4096	Path	0.2	0.0005	250 150 100	0.04 0.02 0.02	7.19e-01	1.56e+00	Model link
9	15	1	3	4096	Tree	0.1	0.001	250 150 100	0.04 0.02 0.02	7.24e-01	1.43e+00	Model link
10	16	1	3	4096	Tree	0.1	0.0005	250 150 100	0.04 0.02 0.02	7.27e-01	1.61e+00	Model link
11	108	0	3	4096	Tree	0.2	0.0005	250 150 100	0.04 0.02 0.02	7.32e-01	1.60e+00	Model link
12	67	0	3	4096	Tree	0.1	0.0005	250 150 100	0.04 0.02 0.02	7.32e-01	1.58e+00	Model link
13	105	0	3	4096	Path	0.4	0.0005	250 150 50	0.04 0.02 0.02	7.35e-01	1.63e+00	Model link

Clustering Floes

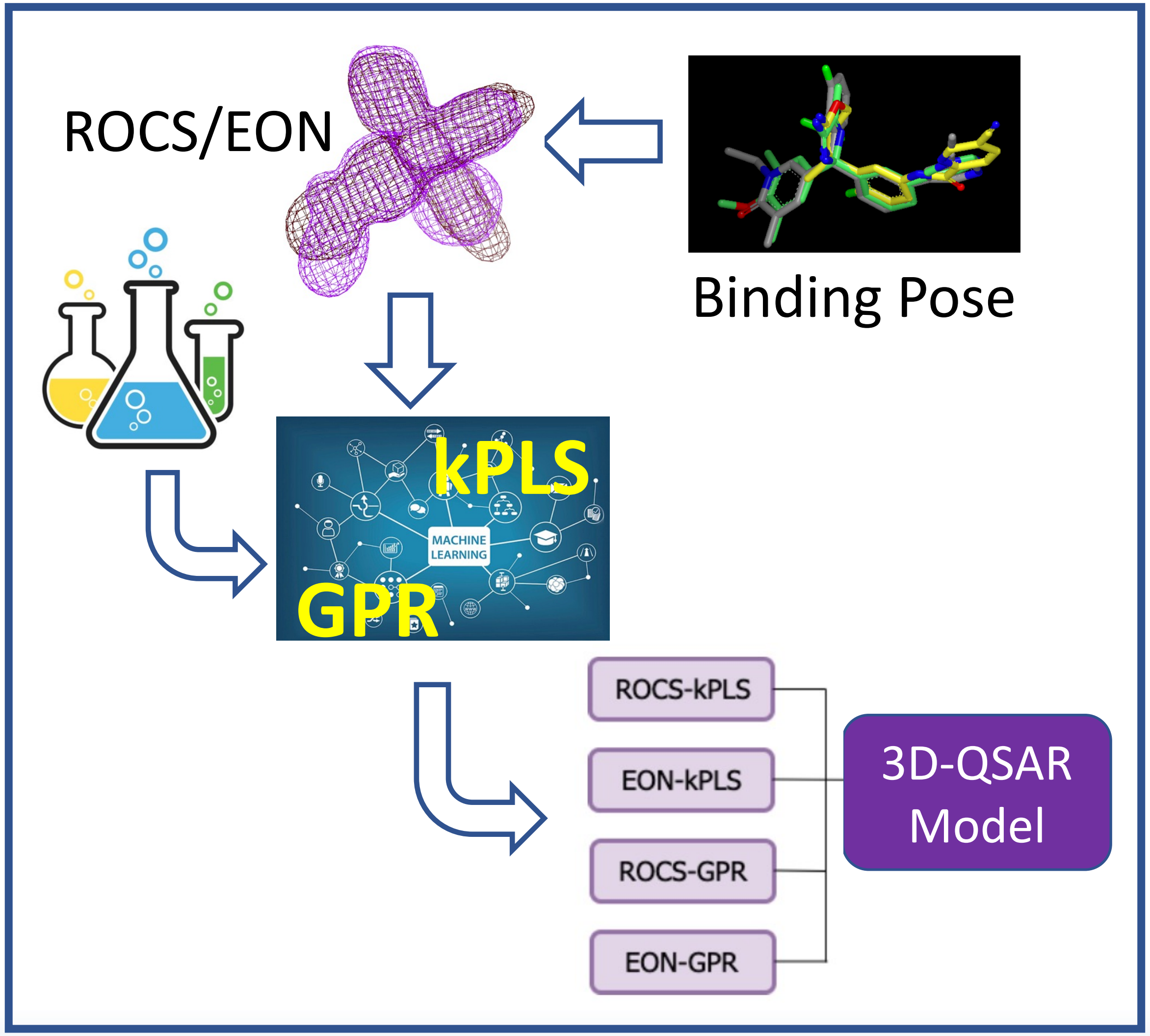
POSTER

- Large overhaul of clustering floes
  - Faster, more methods (K-Medoids)
  - Handle larger datasets
- Problem-focused Floes
  - Diverse Subset
  - Hitlist Clustering
- Large-Scale Clustering
  - Parallelized sphere exclusion
- Utilities
  - Generate similarity matrix
  - Use external similarity matrix



3D QSAR Models for Binding Affinity

POSTER

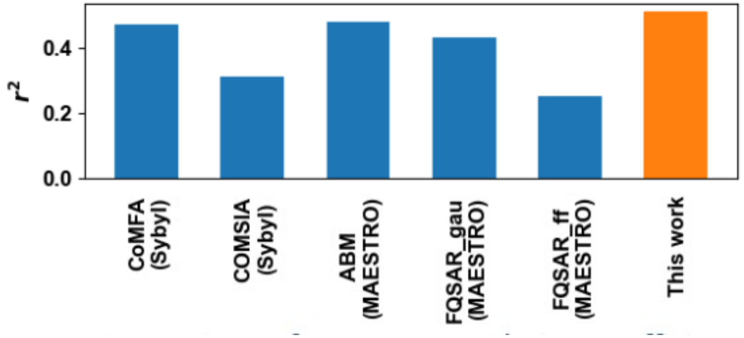


Goals

- Better predictive capability than 2D models
- Molecular design insights
- Results interpretation
- Applicability domain and prediction uncertainty

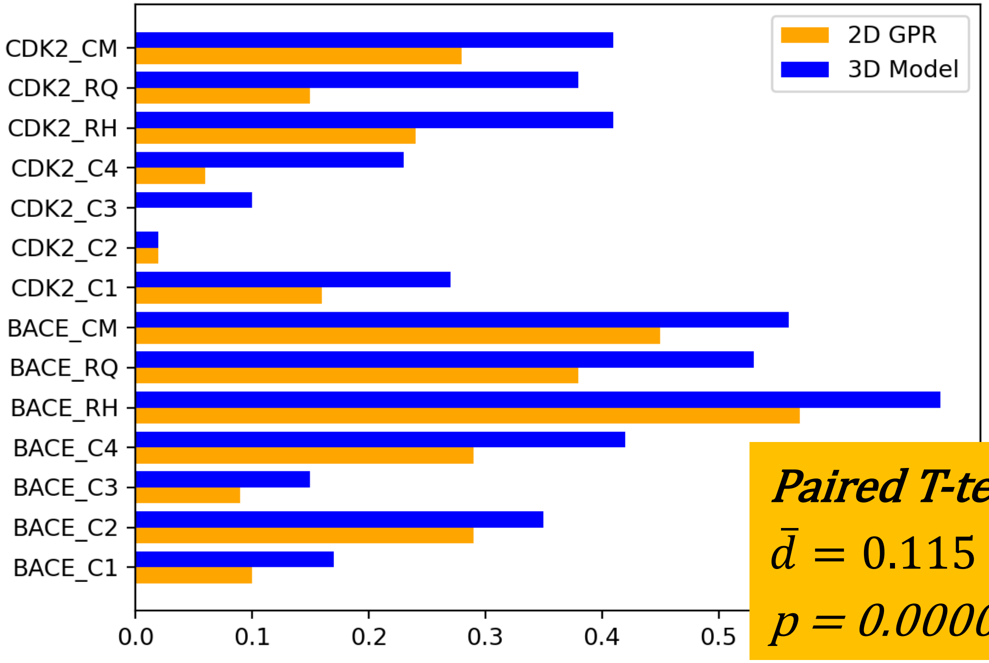
3D QSAR Models for Binding Affinity

External validation/ BACE1 dataset



Subramanian, et al., *J Chem Info Mode*, 2016.

External validation/ prepped dataset

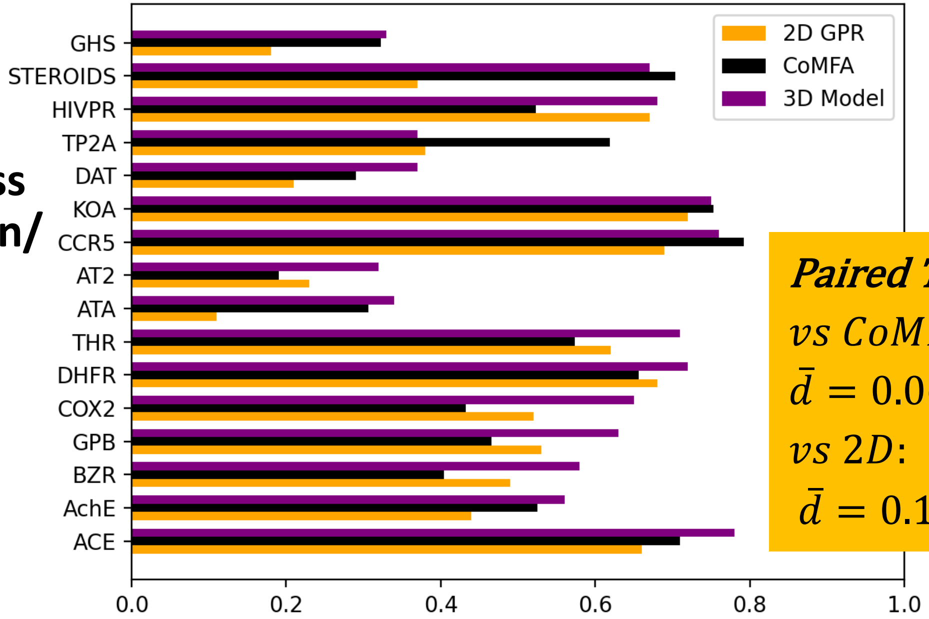


Paired T-test :

$\bar{d} = 0.115$

$p = 0.00001$

LOO Cross validation/ CoMFA dataset



Paired T-tests :

vs CoMFA:

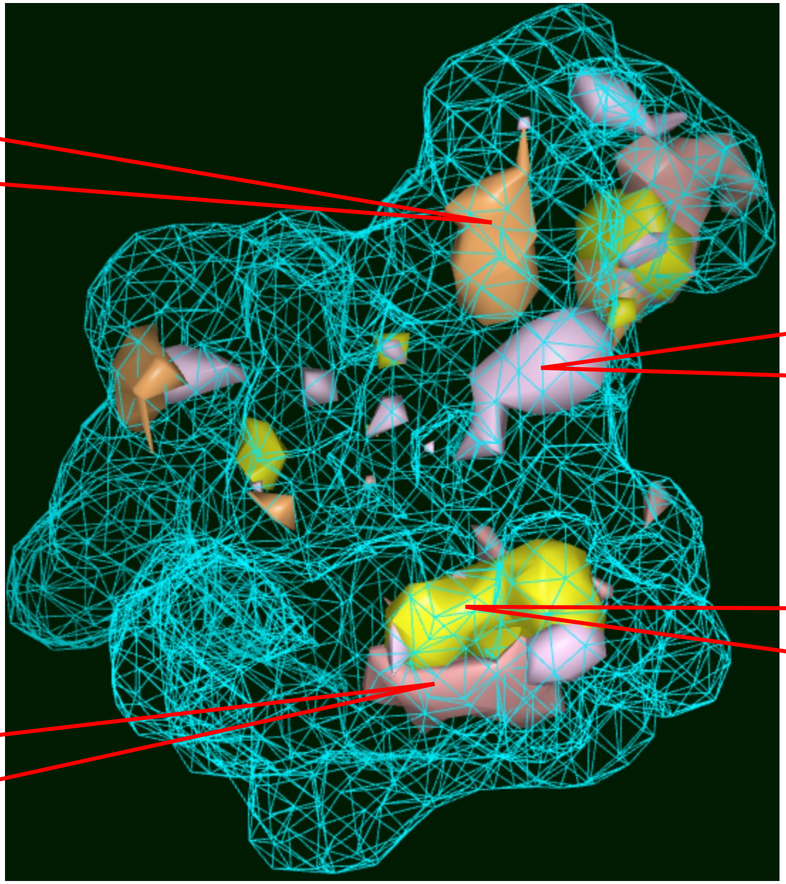
$\bar{d} = 0.06$ ($p = 0.05$)

vs 2D:

$\bar{d} = 0.11$ ($p = 0.0001$)

Journal of Computer-Aided Molecular Design (2019) 33:855–864
https://doi.org/10.1007/s10822-019-00231-x

3D QSAR Models for Binding Affinity



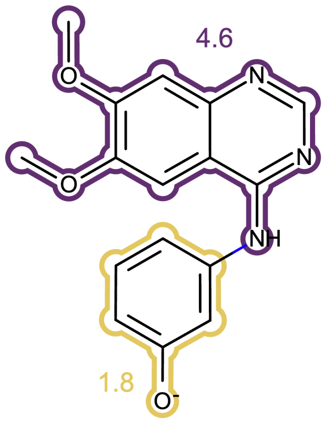
anion preferred region

HB-Acceptor preferred region

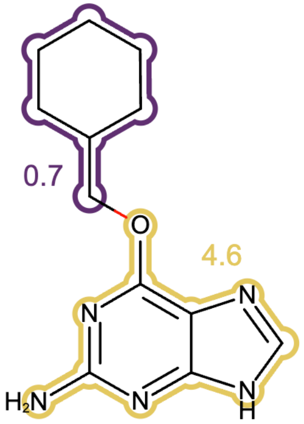
HB-Donor preferred region

cation preferred region

ROCS® Query



Predicted potency: 6.4



Predicted potency: 5.3

Other Floe Packages

Classic-Floes:

- Annotation of interaction hints on docking results dataset for filtering in Orion + interaction fingerprint tanimoto to experimental ligand
- New Single point MMPBSA Floe
- New CHOMP Floe for generating BROOD database
- EON floe extended to accept multiple query molecules

Bio-Modeling Floes:

- New MMDS Data -> New SiteHopper databases to search
- Reduce DU to smallest binding unit (*cost effective for MD)
- Download DU to PDB (outside of ETL floes)

Conclusions

- Significant scientific development our toolkits and products have been performed in 2022 both in advancing our existing products, as well as in some new areas
- With availability of Orion, we are now more posed and committed to advancing both our existing technology and venture into new ones

Thanks!