Scientific Developments in Floes, Toolkits, & Applications
Part 2

Jesper Sørensen

CUP 2023

OpenEye SCIENTIFIC
CADENCE MOLECULAR SCIENCES

Floe Packages with OE Science

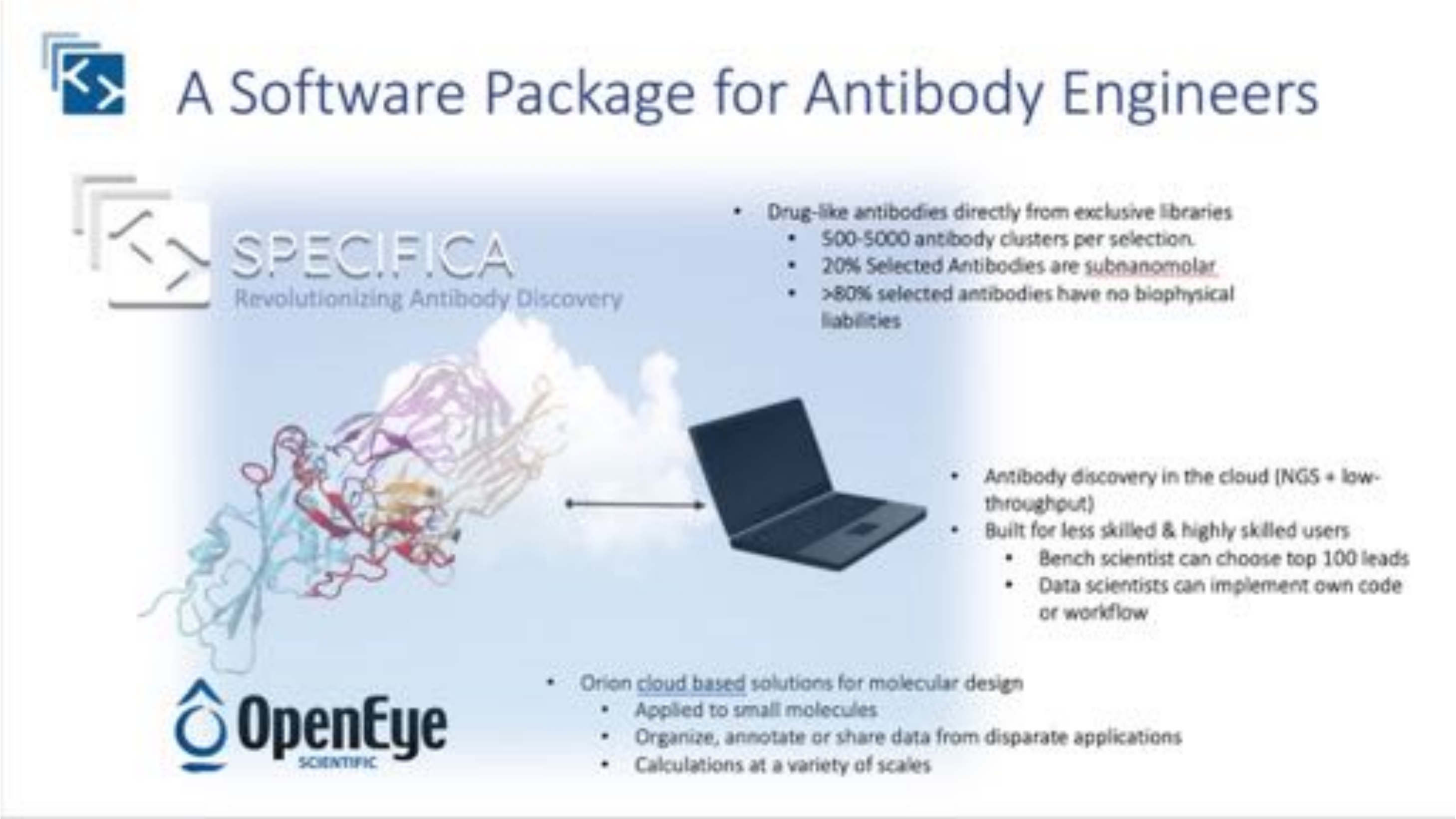
- OpenEye Snowball Cubes
- OpenEye Classic Floes
- OpenEye ChemInfo Floes
- OpenEye MMDS Floes
- OpenEye Biomodeler Floes
- **OpenEye Large Scale Floes**
- **OpenEye Generative Design Floes**
- **OpenEye Large Scale Rxn Enum Floes**
- **OpenEye Model Building Floes**
- OpenEye 3D-QSAR Model Floes

- OpenEye QM Psi4 Floes
- **OpenEye MD Affinity Floes**
- **OpenEye Permeability Floes**

Additional Suites & Modules

- OpenEye / Specifica AbXtract Module
- OpenEye QM Gaussian Floes
- **OpenEye Crystal Math Floes**

Antibody Discovery Suite



A Software Package for Antibody Engineers

SPECIFICA
Revolutionizing Antibody Discovery

- Drug-like antibodies directly from exclusive libraries
  - 500-5000 antibody clusters per selection.
  - 20% Selected Antibodies are subnanomolar.
  - >80% selected antibodies have no biophysical liabilities

- Antibody discovery in the cloud (NGS + low-throughput)
- Built for less skilled & highly skilled users
  - Bench scientist can choose top 100 leads
  - Data scientists can implement own code or workflow

OpenEye SCIENTIFIC

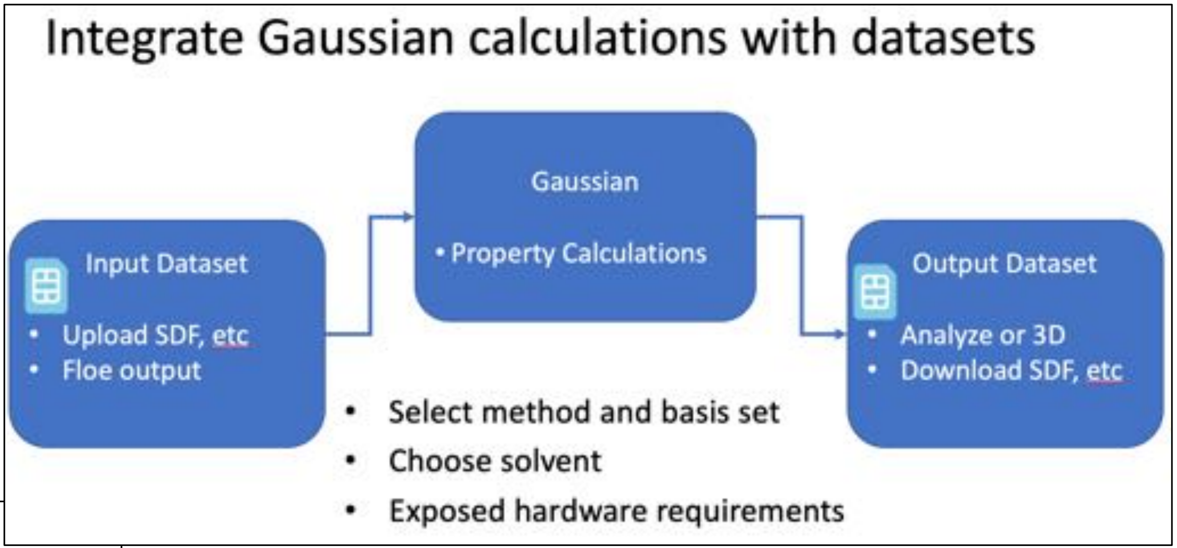
- Orion cloud based solutions for molecular design
  - Applied to small molecules
  - Organize, annotate or share data from disparate applications
  - Calculations at a variety of scales

DATA COMMONS

Gaussian Module

More methods:
- Semi-empirical
- Dispersion corrected

More solvent support:
- Water, DMSO, and octanol solvents
- IEFPCM, CPCM, and SMD models



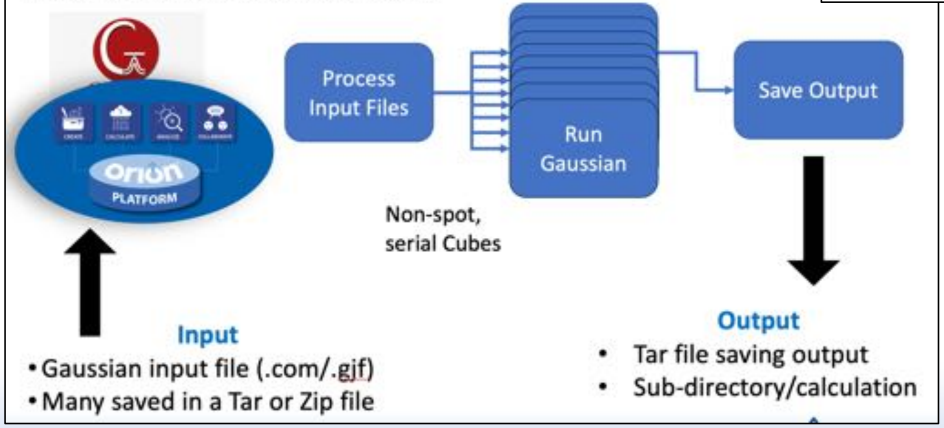
Integrate Gaussian calculations with datasets

Input Dataset
- Upload SDF, etc
- Floe output

Gaussian
- Property Calculations

Output Dataset
- Analyze or 3D
- Download SDF, etc

- Select method and basis set
- Choose solvent
- Exposed hardware requirements



Run Gaussian on Orion

Process Input Files

Run Gaussian

Save Output

Non-spot, serial Cubes

Input
- Gaussian input file (.com/.gjf)
- Many saved in a Tar or Zip file

Output
- Tar file saving output
- Sub-directory/calculation

Floes
- Single Point Energy
- Geometry Optimization
- Torsion Scanning
- Conformer Ensemble Analysis

Strengthening our Core

Standings

2022 SUMMER COED 7v7

Strengthening our Core

But really…

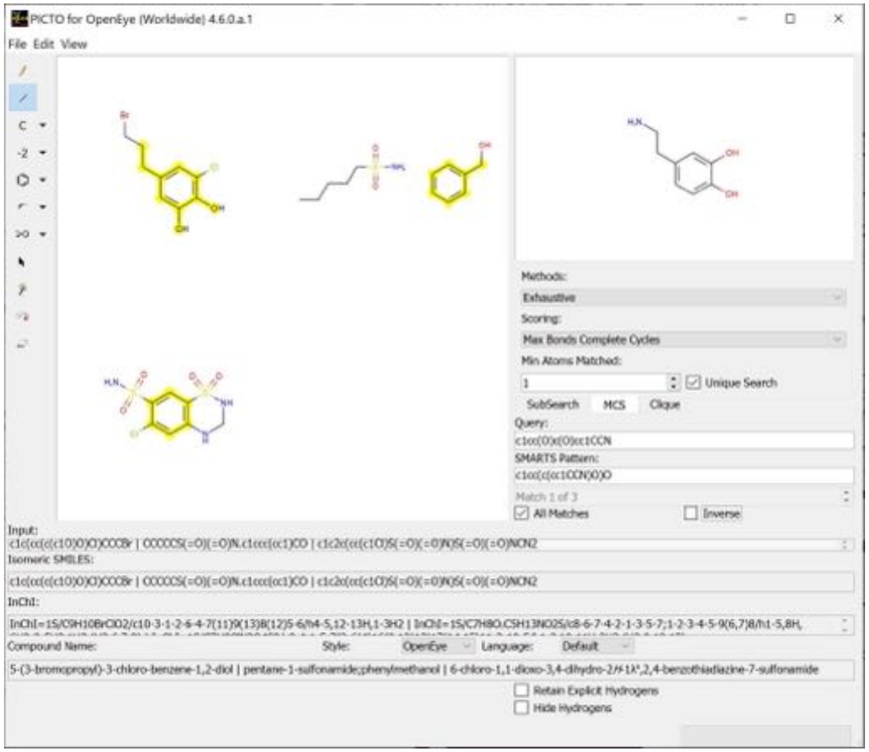
- Physics
- Shape & Electrostatics
- Statistics
- Cheminformatics
- Molecular modelling
- Biomodelling
- Simulation

- We could go faster
  - and make it cheaper

- We can do more (sampling) with Orion
  - Reliable and robust compute
  - Sometimes more sampling is necessary

PICTO

- MCS and Clique Search (in addition to SubSearch)
- Show all-matches / inverted matches
- Visual feedback for invalid SMARTs
- Retain explicit hydrogens
- Switching aromatic-model
- InChi Inputs and Outputs
- Export SMILES



Methods: Exhaustive

Scoring: Max Bonds Complete Cycles

Min Atoms Matched: 1 ☑ Unique Search

Query: c1cc(O)c(O)cc1CCN

SMARTS Pattern: c1cc(c(cc1CCN)O)O

Match 1 of 3

☑ All Matches ☐ Inverse

Input: c1c(cc(c(c1O)O)Cl)CCCBr | CCCCCS(=O)(=O)N.c1ccc(cc1)CO | c1c2c(cc(c1Cl)S(=O)(=O)N)S(=O)(=O)NCN2

Isomeric SMILES: c1c(cc(c(c1O)O)Cl)CCCBr | CCCCCS(=O)(=O)N.c1ccc(cc1)CO | c1c2c(cc(c1Cl)S(=O)(=O)N)S(=O)(=O)NCN2

InChI: InChI=1S/C9H10BrClO2/c10-3-1-2-6-4-7(11)9(13)8(12)5-6/h4-5,12-13H,1-3H2 | InChI=1S/C7H8O.C5H13NO2S/c8-6-7-4-2-1-3-5-7;1-2-3-4-5-9(6,7)8/h1-5,8H,

Compound Name: Style: OpenEye Language: Default

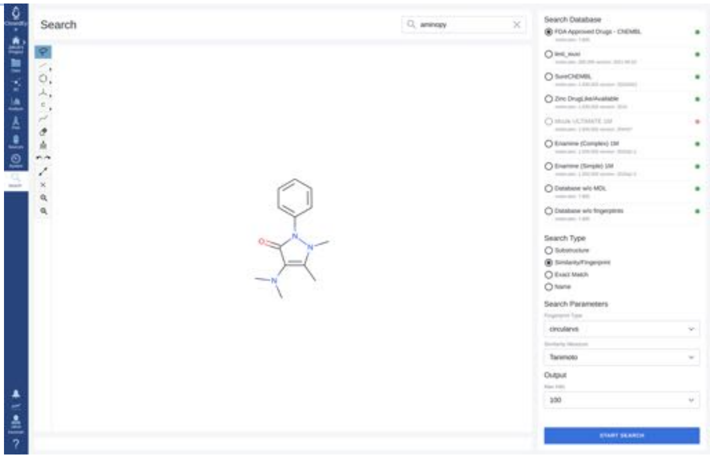
5-(3-bromopropyl)-3-chloro-benzene-1,2-diol | pentane-1-sulfonamide;phenylmethanol | 6-chloro-1,1-dioxo-3,4-dihydro-2H-1λ⁶,2,4-benzothiadiazine-7-sulfonamide

☐ Retain Explicit Hydrogens

☐ Hide Hydrogens

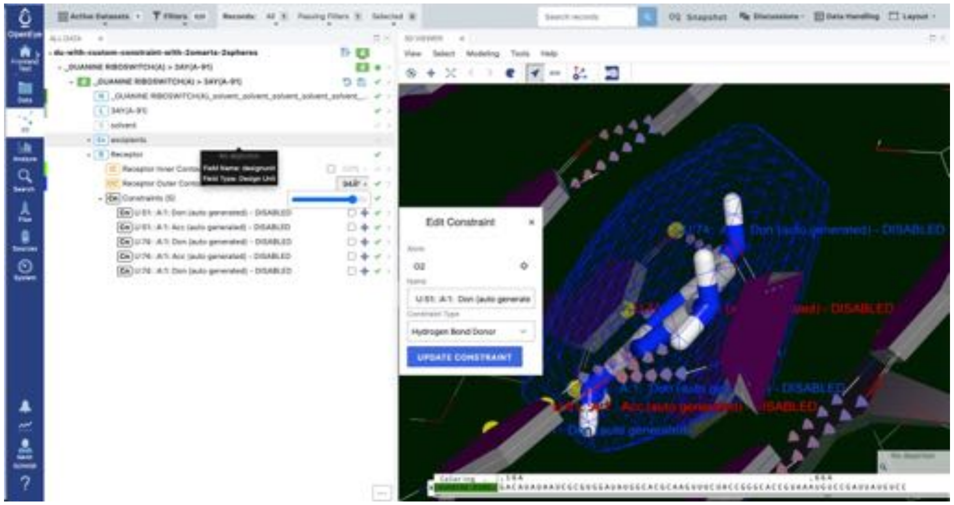
Improved UI in Orion

- MaaS Searches
    - Use new tightly-integrated MolSearch backend/UI in Orion



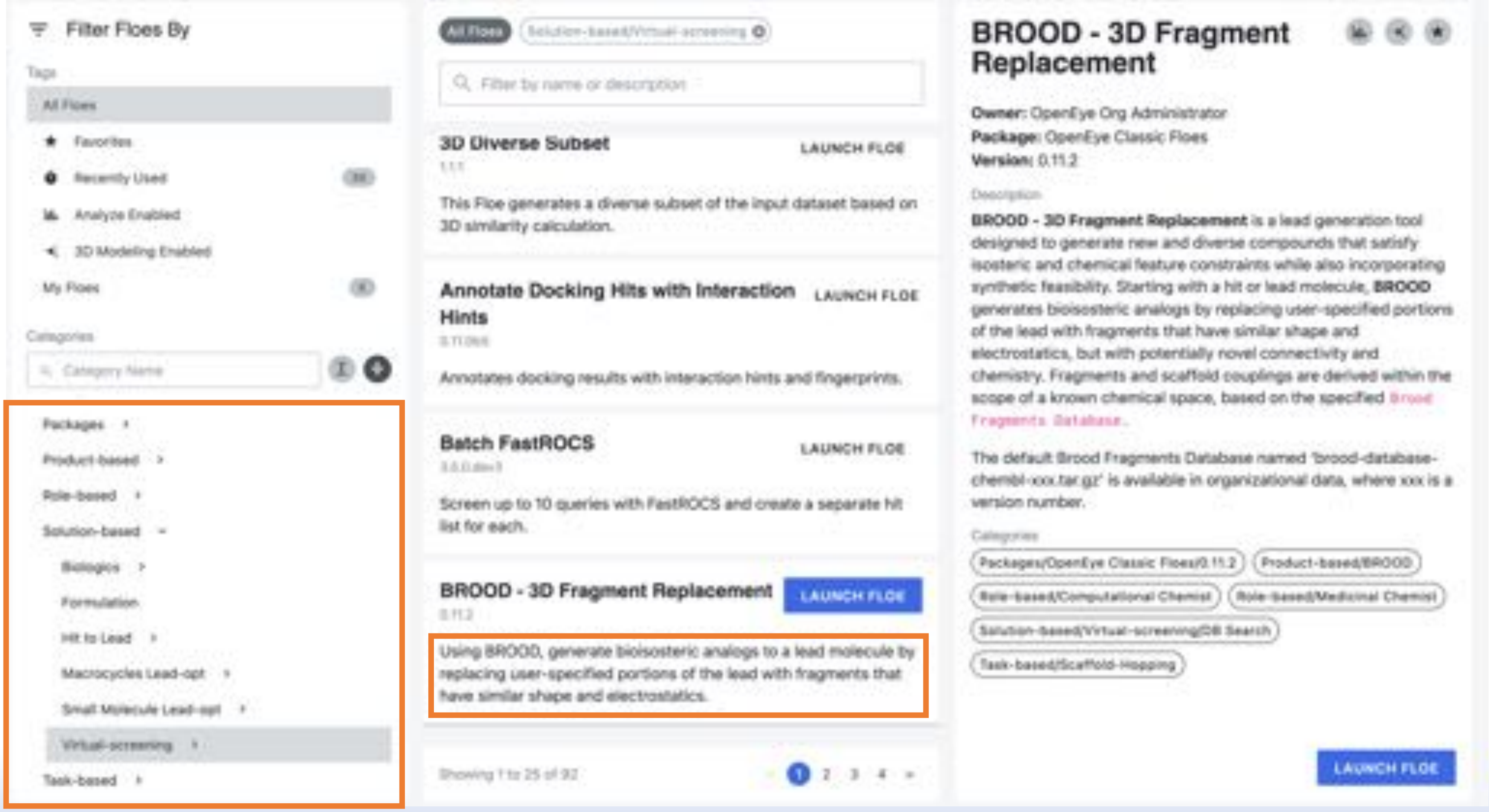
- MakeReceptor GUIs on Orion

- Transition some desktop GUIs to Orion



🤜🤛 lots of work by and collaboration with Cloud teams

Floe Categories



3D Diverse Subset
1.1.1

This Floe generates a diverse subset of the input dataset based on 3D similarity calculation.

Annotate Docking Hits with Interaction Hints
0.11.0b6

Annotates docking results with interaction hints and fingerprints.

Batch FastROCS
3.5.0.dev1

Screen up to 10 queries with FastROCS and create a separate hit list for each.

BROOD - 3D Fragment Replacement
0.11.2

Using BROOD, generate bioisosteric analogs to a lead molecule by replacing user-specified portions of the lead with fragments that have similar shape and electrostatics.

BROOD - 3D Fragment Replacement

Owner: OpenEye Org Administrator
Package: OpenEye Classic Floes
Version: 0.11.2

Description

BROOD - 3D Fragment Replacement is a lead generation tool designed to generate new and diverse compounds that satisfy isosteric and chemical feature constraints while also incorporating synthetic feasibility. Starting with a hit or lead molecule, **BROOD** generates bioisosteric analogs by replacing user-specified portions of the lead with fragments that have similar shape and electrostatics, but with potentially novel connectivity and chemistry. Fragments and scaffold couplings are derived within the scope of a known chemical space, based on the specified `Brood Fragments Database`.

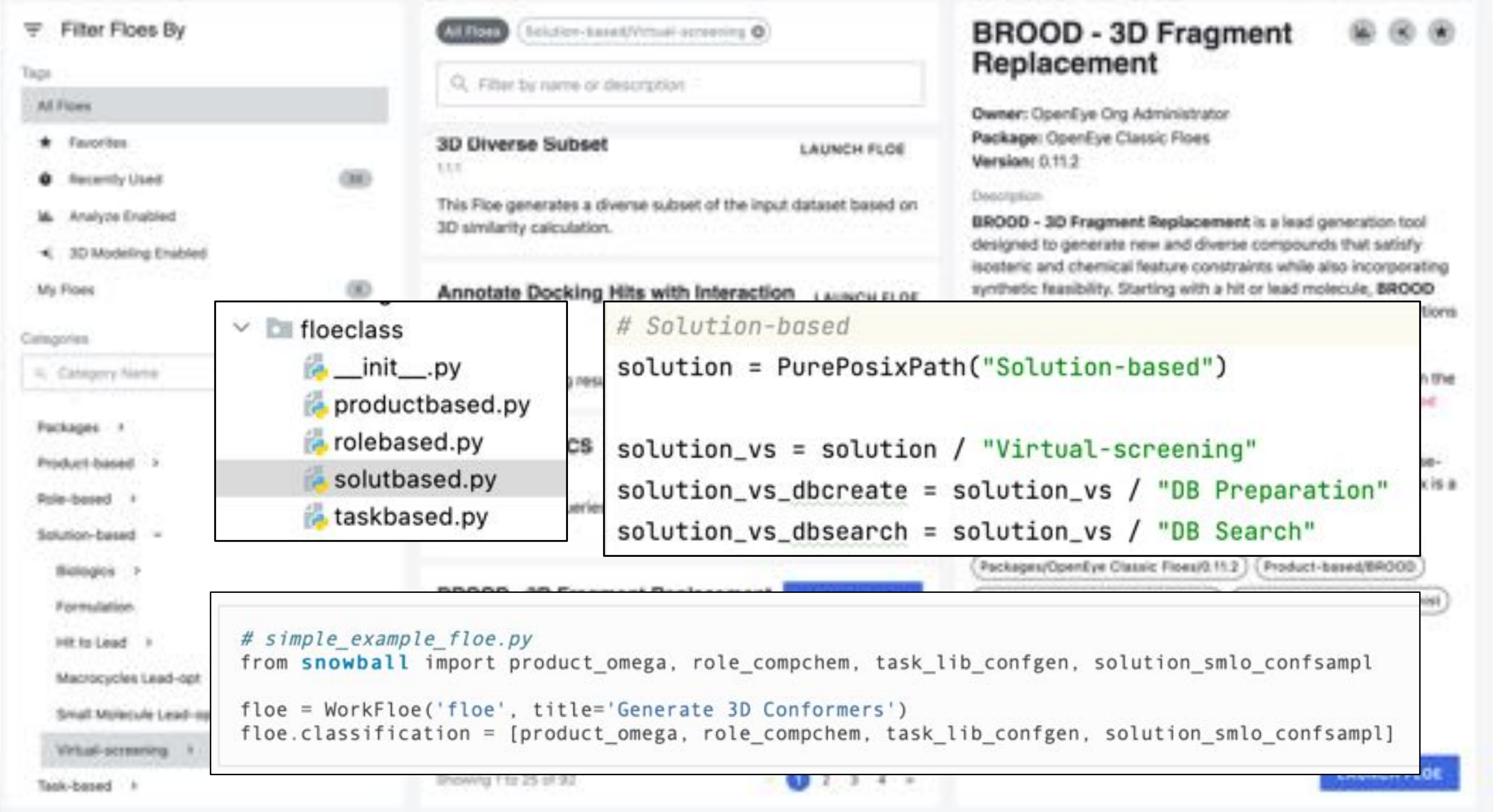
The default Brood Fragments Database named 'brood-database-chembl-xxx.tar.gz' is available in organizational data, where xxx is a version number.

Categories

- Packages/OpenEye Classic Floes/0.11.2
- Product-based/BROOD
- Role-based/Computational Chemist
- Role-based/Medicinal Chemist
- Solution-based/Virtual-screening/DB Search
- Task-based/Scaffold-Hopping

LAUNCH FLOE

Floe Categories – for programmers



3D Diverse Subset

This Floe generates a diverse subset of the input dataset based on 3D similarity calculation.

Annotate Docking Hits with Interaction

BROOD - 3D Fragment Replacement

Owner: OpenEye Org Administrator
Package: OpenEye Classic Floes
Version: 0.11.2

Description

BROOD - 3D Fragment Replacement is a lead generation tool designed to generate new and diverse compounds that satisfy isosteric and chemical feature constraints while also incorporating synthetic feasibility. Starting with a hit or lead molecule, **BROOD**

```
floeclass
    __init__.py
    productbased.py
    rolebased.py
    solutbased.py
    taskbased.py
```

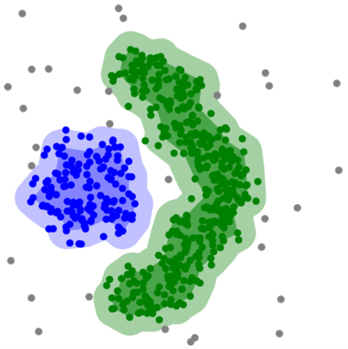
```
# Solution-based
solution = PurePosixPath("Solution-based")

solution_vs = solution / "Virtual-screening"
solution_vs_dbcreate = solution_vs / "DB Preparation"
solution_vs_dbsearch = solution_vs / "DB Search"
```

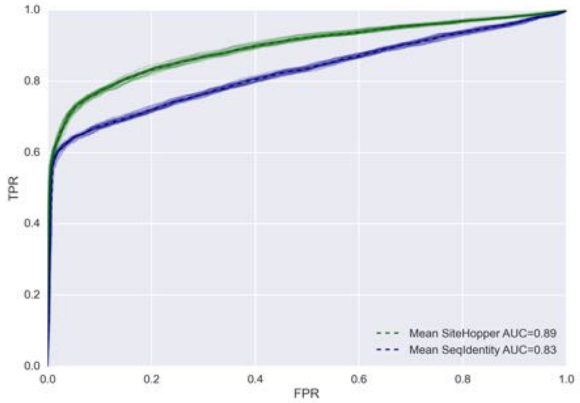
```
# simple_example_floe.py
from snowball import product_omega, role_compchem, task_lib_confgen, solution_smlo_confsampl

floe = WorkFloe('floe', title='Generate 3D Conformers')
floe.classification = [product_omega, role_compchem, task_lib_confgen, solution_smlo_confsampl]
```

OpenEye SCIENTIFIC
CADENCE MOLECULAR SCIENCES

Data Science Cube Package

- Dependencies
  - Scikit-learn
  - Numpy
  - SciPy
  - Pandas
  - Matplotlib
- Contents
  - Clustering
  - Test/Train Splits
  - Statistics/Metrics

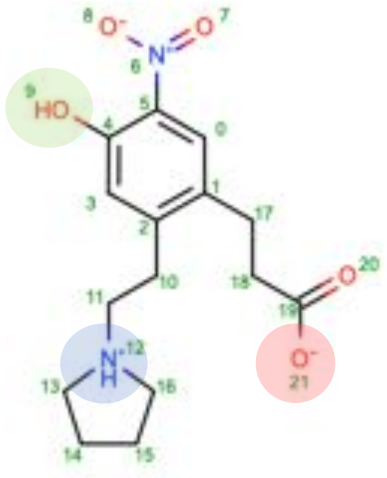


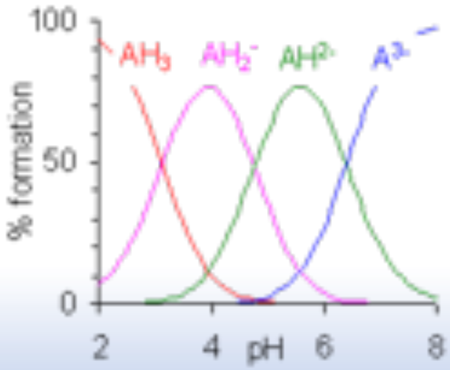


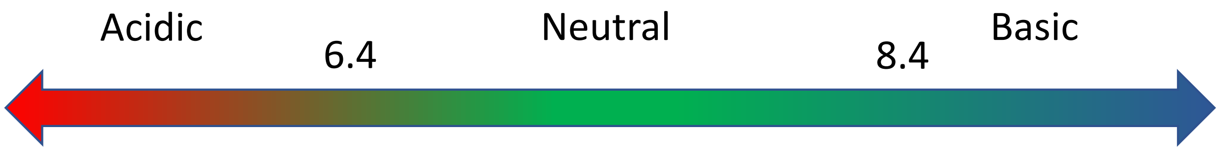
Predicting the pKa of ionizable groups

POSTER

- predict pKA of ionizable groups using Gaussian Process Regression
  - Use OE strengths in Physics, 3D to build features
- Handle molecules with one or more ionizable groups
- Use micro-pKas to predict macro-pKa (experiment)







Acidic

6.4

Neutral

8.4

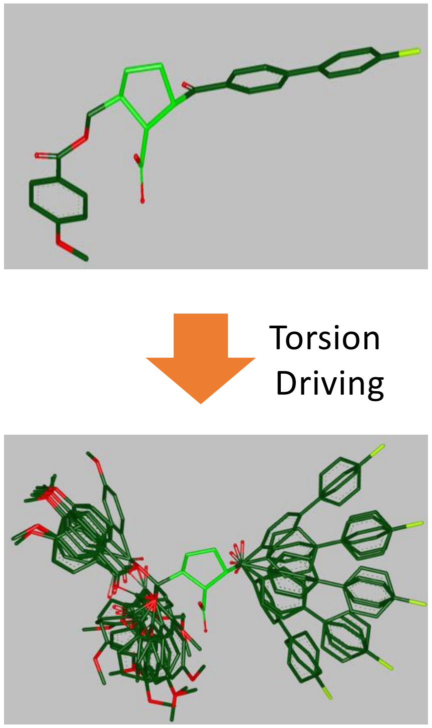
Basic

OpenEye SCIENTIFIC

CADENCE MOLECULAR SCIENCES

New OMEGA Torsion Library

POSTER



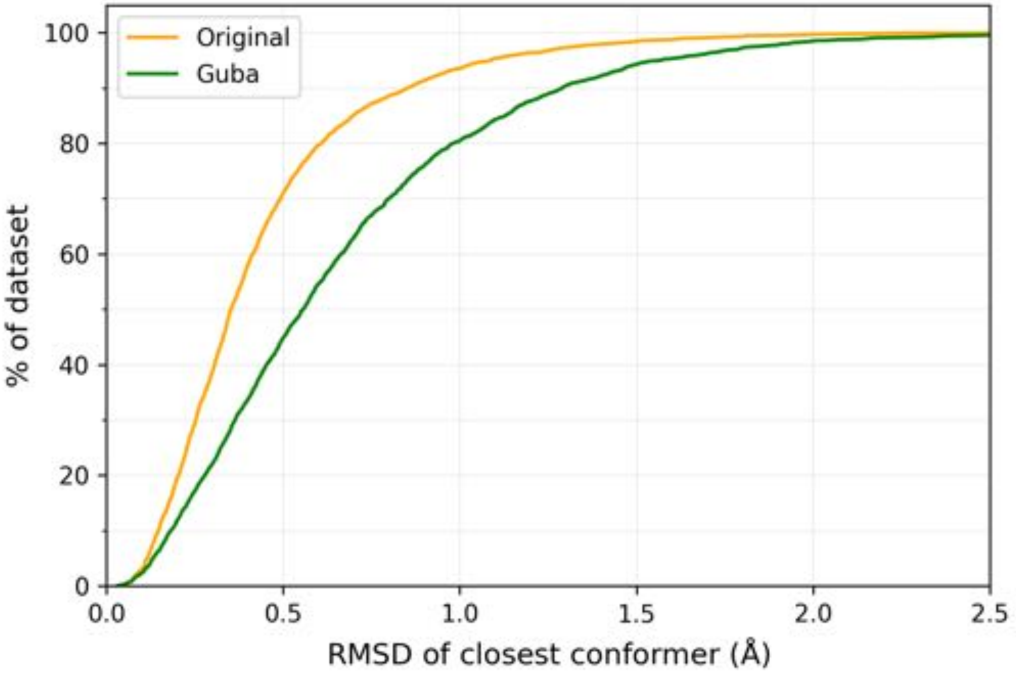
Torsion Driving

- **Torsion Library ->** Rules for torsion driving
- **New Torsion Library**
  - Torsion scanning using QM
  - More sets of specific rules

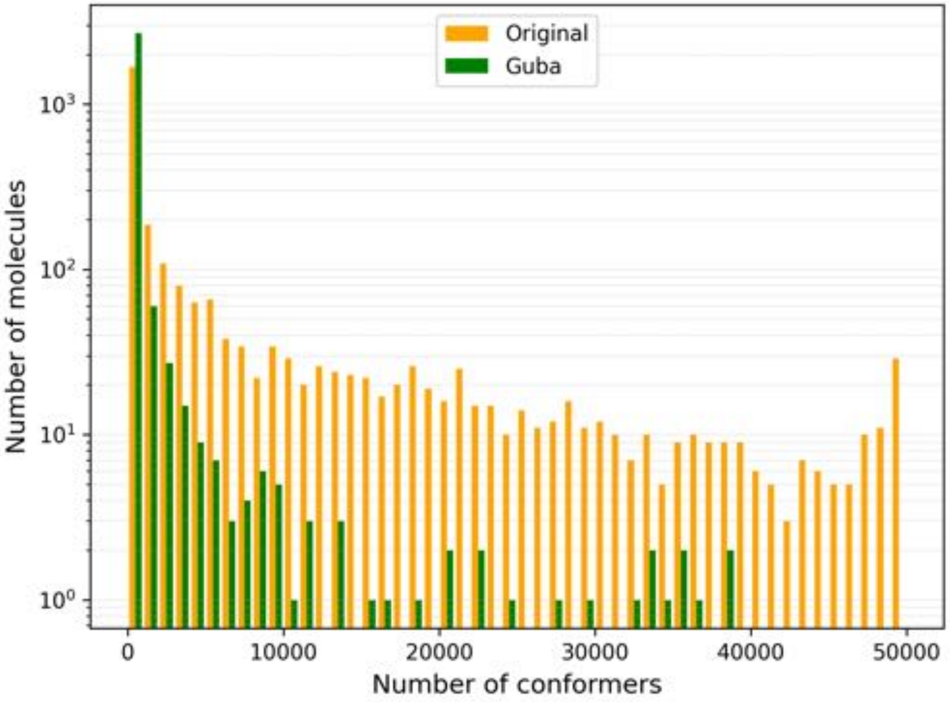
Comparing Torsion Libraries

Benchmarking with Platinum Diverse Dataset

POSTER



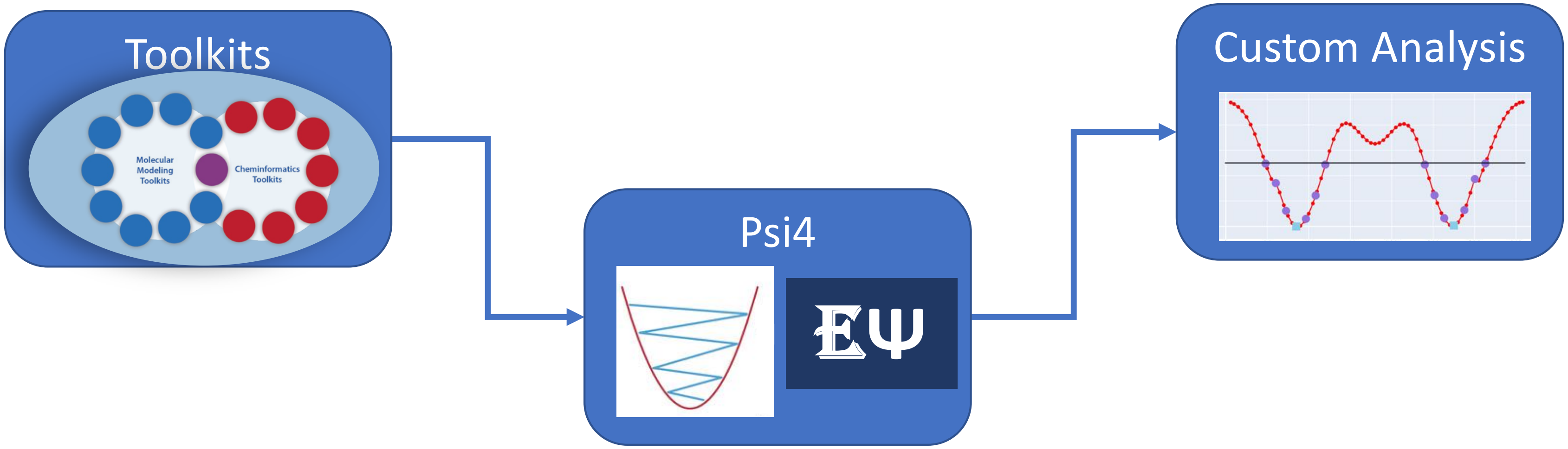
Original library can give a lower RMSD than Guba for a larger percent of the dataset



Original library generates a significantly larger ensemble for the molecules compared to Guba

OpenEye SCIENTIFIC
CADENCE MOLECULAR SCIENCES

Floes integrate toolkits, QM, and custom analysis



Toolkits

Psi4

Custom Analysis

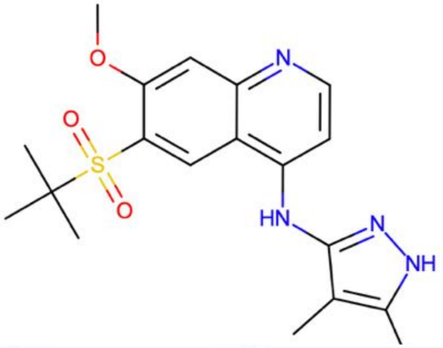
Torsion Scanning

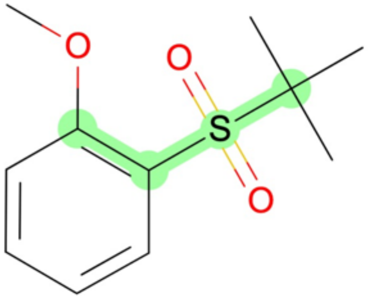
- Select one single bond
- Scan all rotatable bonds
- Generate custom Omega rules

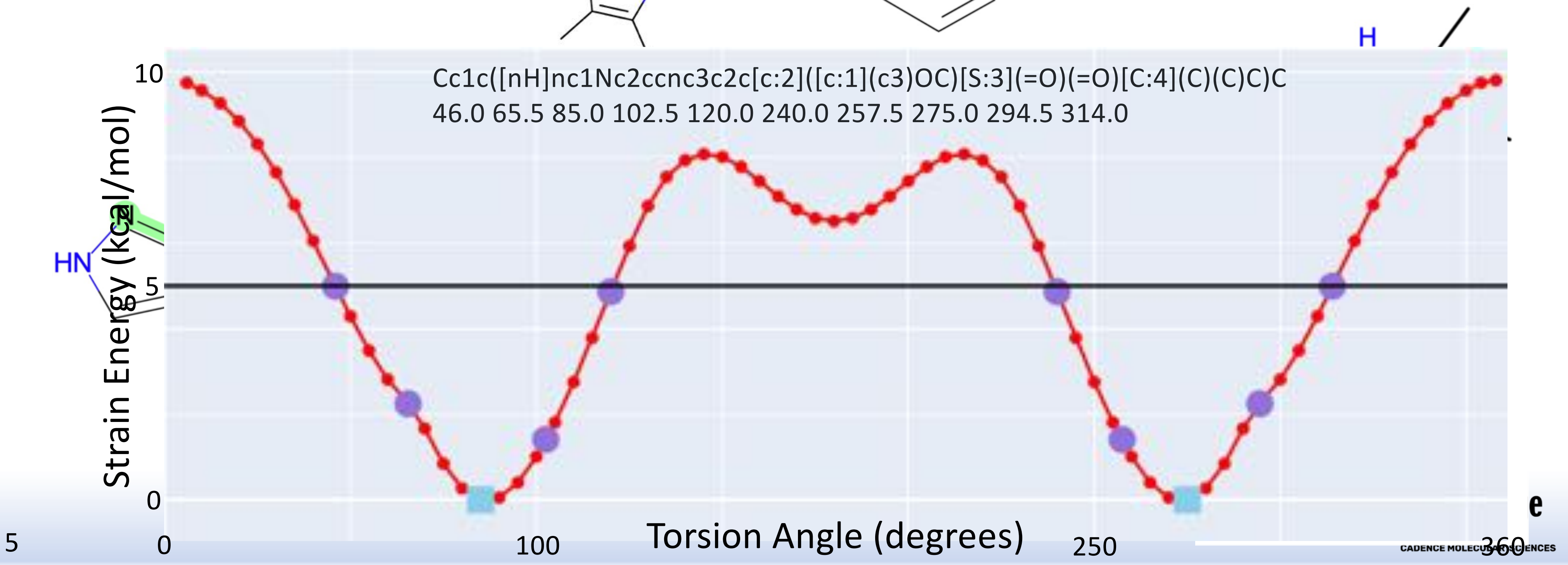
Conformer Analysis

- Diverse ensemble
- Local minima search
- Compare tautomers

Fragmented torsion scans for custom rules



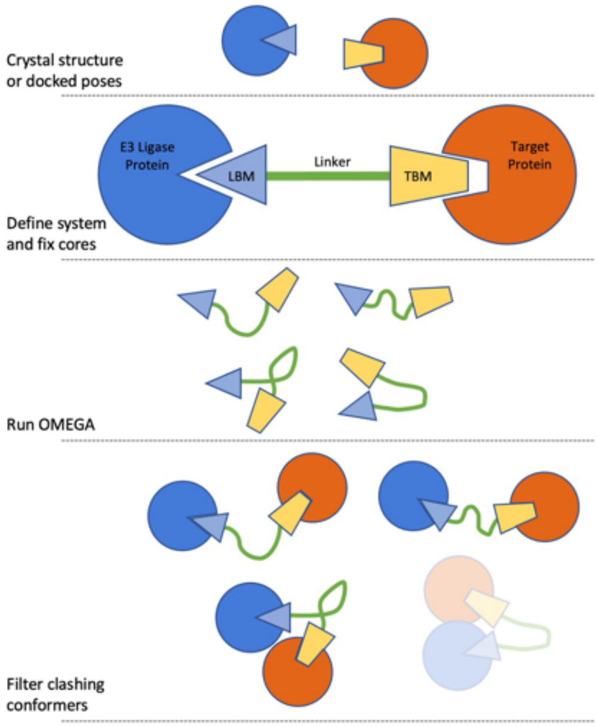


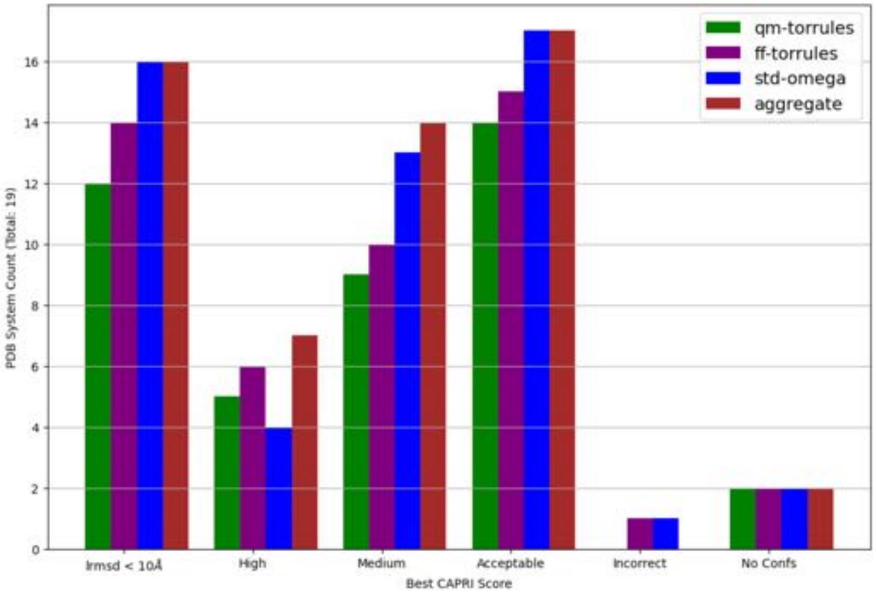

Cc1c([nH]nc1Nc2ccnc3c2c[c:2]([c:1](c3)OC)[S:3](=O)(=O)[C:4](C)(C)C

46.0 65.5 85.0 102.5 120.0 240.0 257.5 275.0 294.5 314.0

PROTAC conformer sampling

POSTER

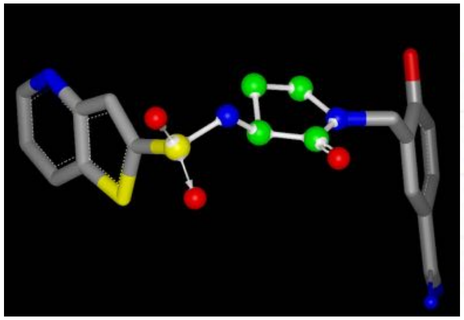


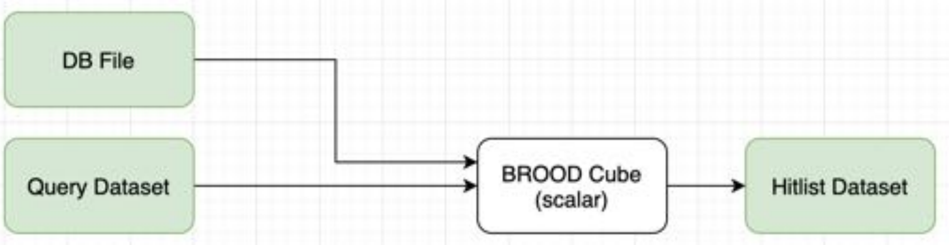


Generating new and modifying torsion rules to focus sampling on linker

OpenEye
SCIENTIFIC
CADENCE MOLECULAR SCIENCES

Orion-Native BROOD Floe



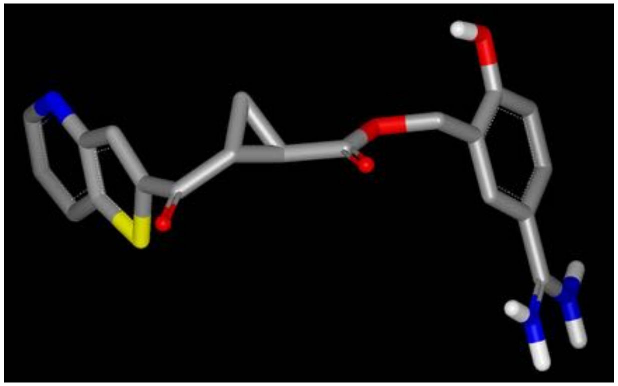


DB File

Query Dataset

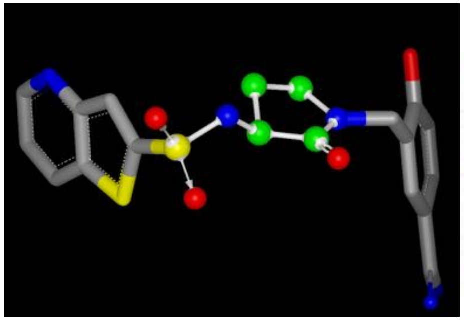
BROOD Cube (scalar)

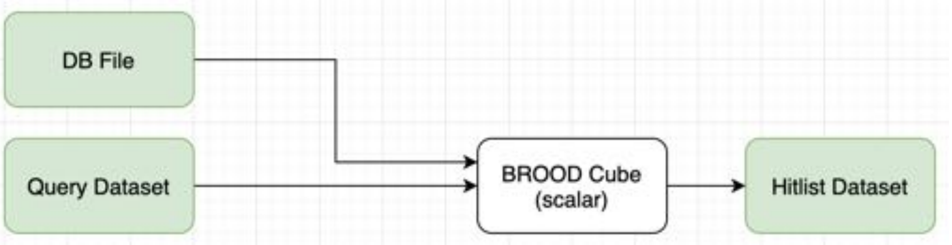
Hitlist Dataset



Existing

Orion-Native BROOD Floe



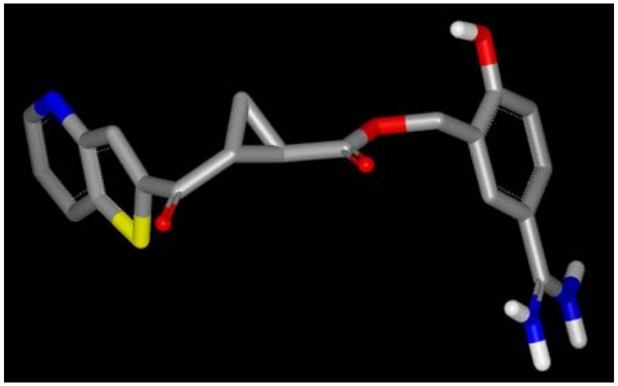


DB File

Query Dataset

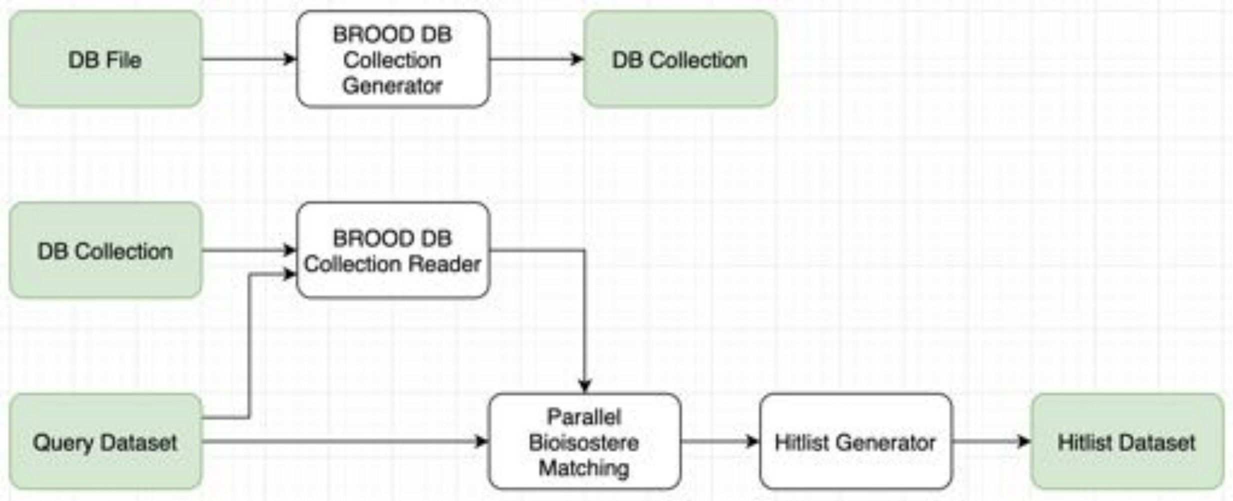
BROOD Cube (scalar)

Hitlist Dataset



Existing

Proposed



DB File

BROOD DB Collection Generator

DB Collection

DB Collection

BROOD DB Collection Reader

Query Dataset

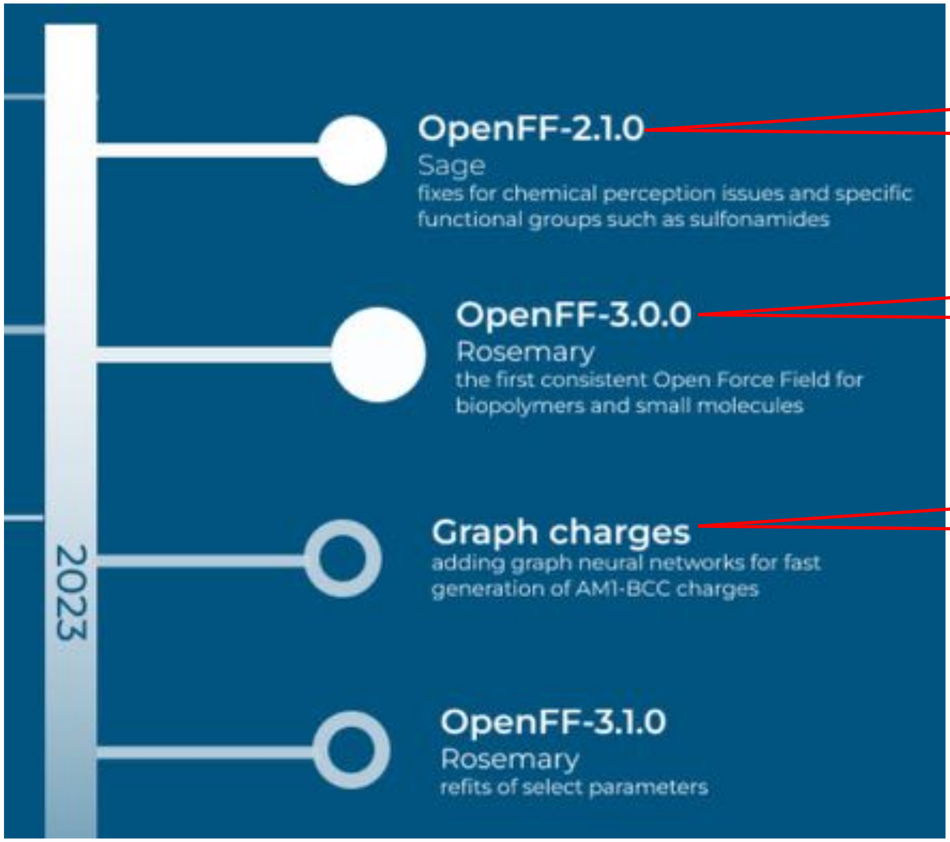
Parallel Bioisostere Matching

Hitlist Generator

Hitlist Dataset

POSTER

OpenFF Forcefield support



OpenFF-2.1.0
Sage
fixes for chemical perception issues and specific functional groups such as sulfonamides

OpenFF-3.0.0
Rosemary
the first consistent Open Force Field for biopolymers and small molecules

Graph charges
adding graph neural networks for fast generation of AM1-BCC charges

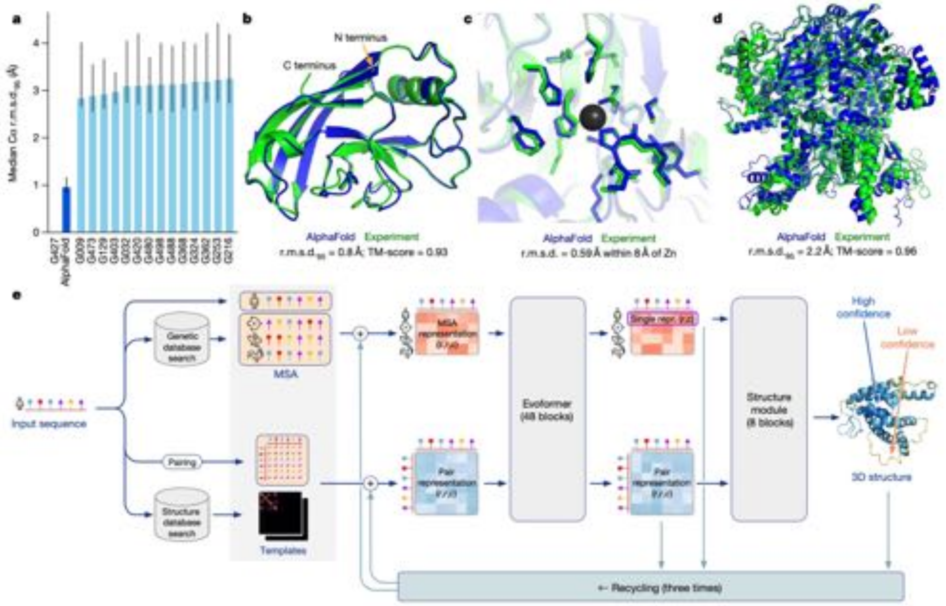
OpenFF-3.1.0
Rosemary
refits of select parameters

Improving Existing FF: specific parameters

New FF: Additional functional form, Protein parameters

Requires third party ML code integration to toolkit

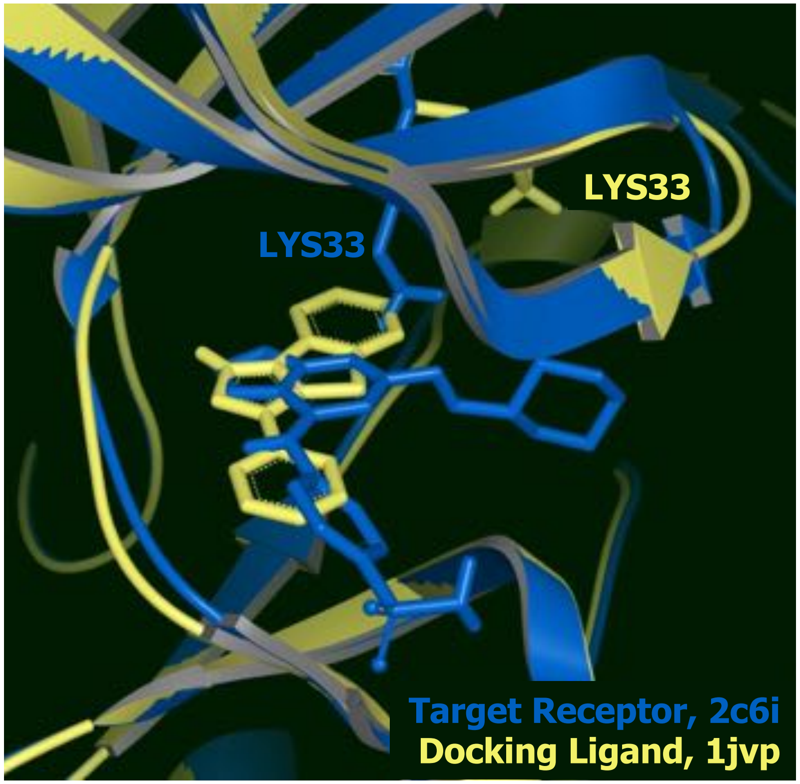
Floes for AI Structure Prediction



- Use pre-trained models (AlphaFold, OpenFold, ESMFold, …)
- Training for-purpose models
- Allows users to generate structure for supplied protein sequence

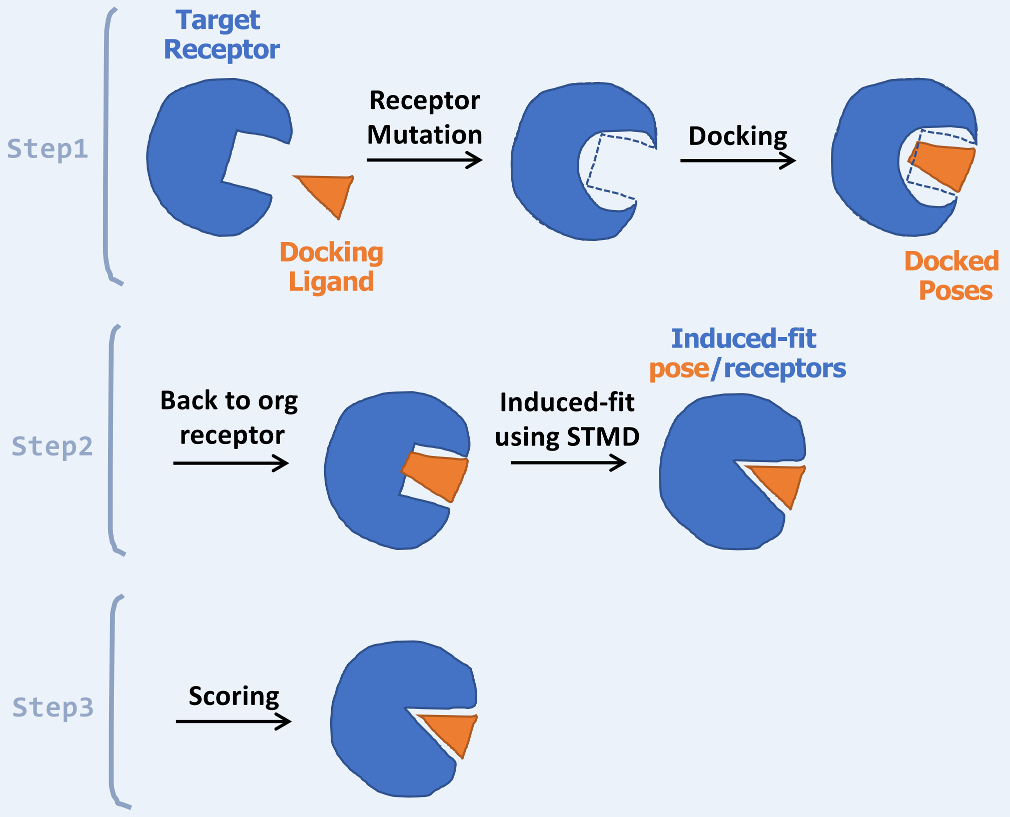
Jumper et al., Nature, 596, 583-589 (2001)

Induced-Fit Posing (IFP) Floe



- Due to the differences in the shapes of the ligands and their ways of binding, simple docking of the 1jvp (yellow) ligand to the 2c6i (blue) receptor without (1) receptor mutation and (2) induced fit using MD can't place the ligand in the right pose.

IFP : 3-step protocol for the correct prediction of the binding pose of a ligand which requires receptor conformation change



Step1

Target Receptor

Docking Ligand

Receptor Mutation →

Docking →

Docked Poses

Step2

Back to org receptor →

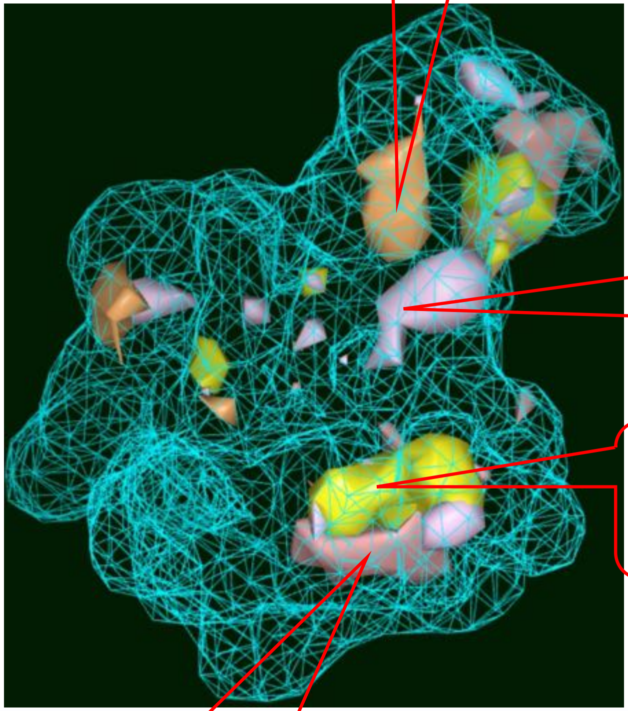
Induced-fit using STMD →

Induced-fit pose/receptors

Step3

Scoring →

Exploring new ways to build queries



anion preferred region

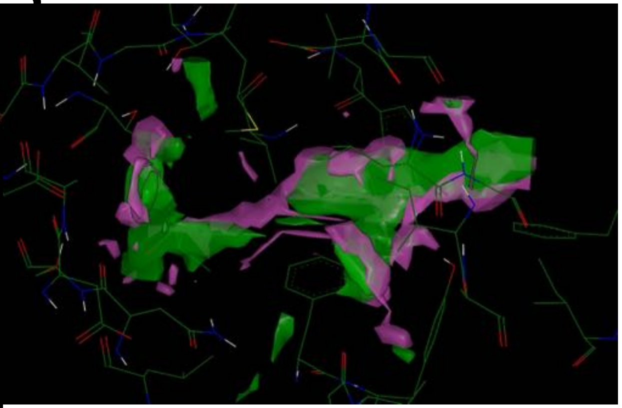
HB-Acceptor preferred region

HB-Donor preferred region

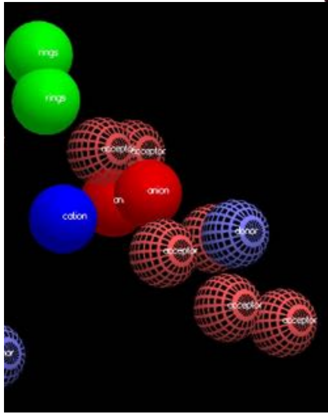
cation preferred region

from 3D-QSAR Model

from Binding Pockets

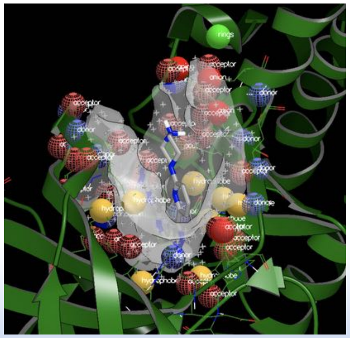


Shape from SZMAP grids

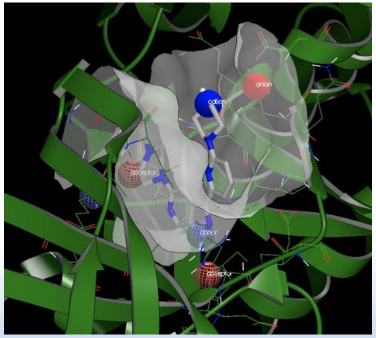


Color from interactions of known/posed ligands

SiteHopper Patch



Experimental Patch

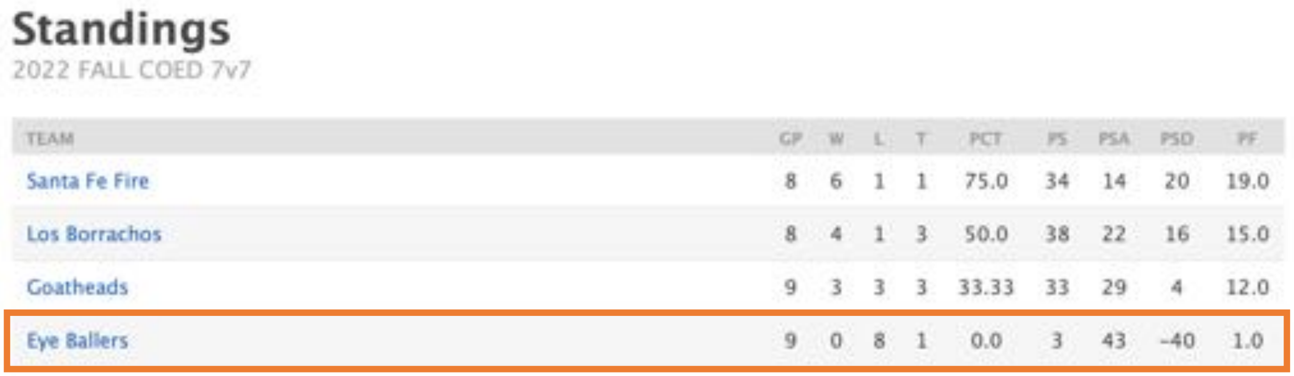


OpenEye SCIENTIFIC

CADENCE MOLECULAR SCIENCES

Status

- Even when you focus, some things are hard



Standings

2022 FALL COED 7v7

TEAM	GP	W	L	T	PCT	PS	PSA	PSD	PF
Santa Fe Fire	8	6	1	1	75.0	34	14	20	19.0
Los Borrachos	8	4	1	3	50.0	38	22	16	15.0
Goatheads	9	3	3	3	33.33	33	29	4	12.0
Eye Ballers	9	0	8	1	0.0	3	43	-40	1.0

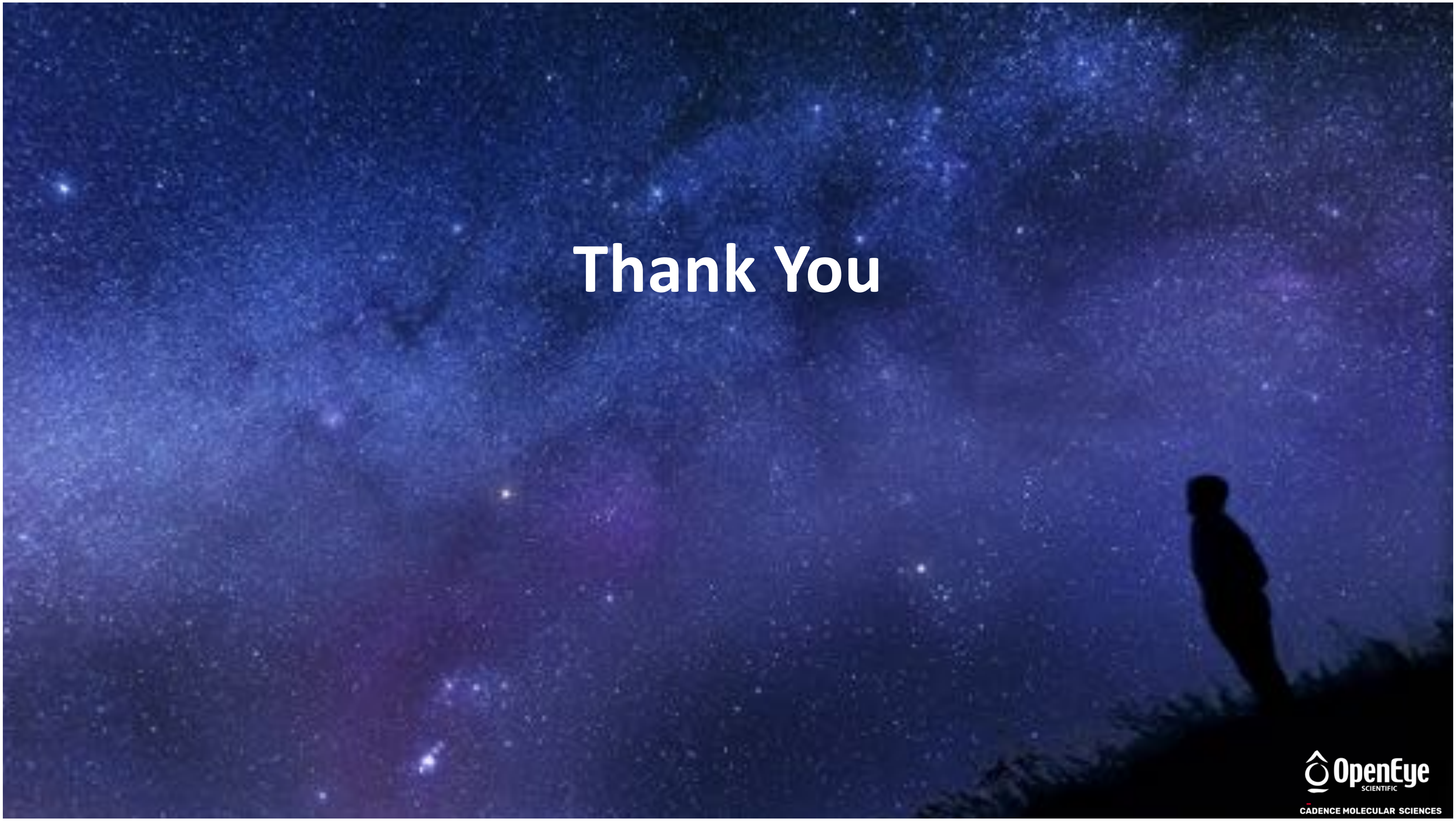
Conclusions

- 2022 moved core OE forward on many different fronts
  - TKs, Apps, Floes

- We are continuing to roll out updates in 2023
  - Improve core tech
  - Improve user interface and experience
  - Give users new, faster, and better solutions

- Exciting new things on horizon to roll out of partnerships to all

Work of many people across several groups, with collaboration between science and cloud groups, enabled by all of OpenEye



Thank You

OpenEye SCIENTIFIC

CADENCE MOLECULAR SCIENCES