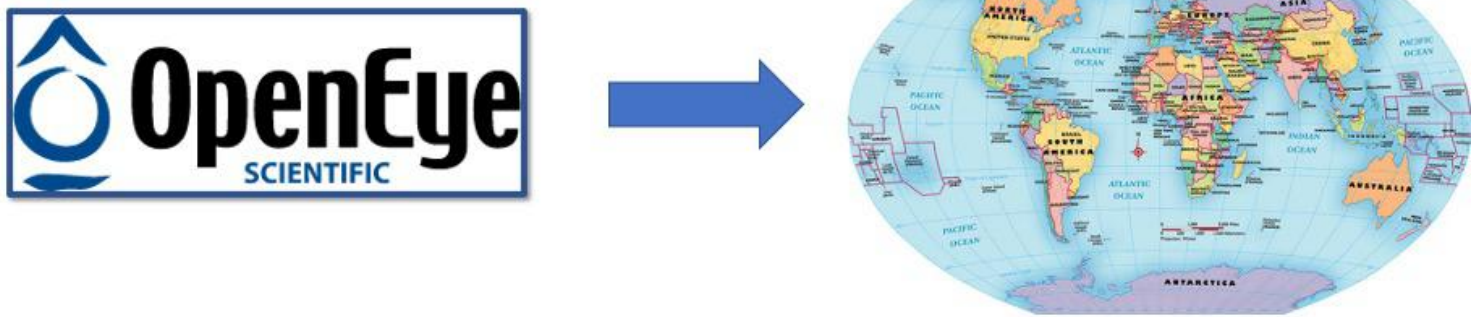
The Why & What of OpenEye

Geoff Skillman, CSO

OpenEye SCIENTIFIC

We’re going to change drug discovery!

Anthony Nicholls, Summer 2000



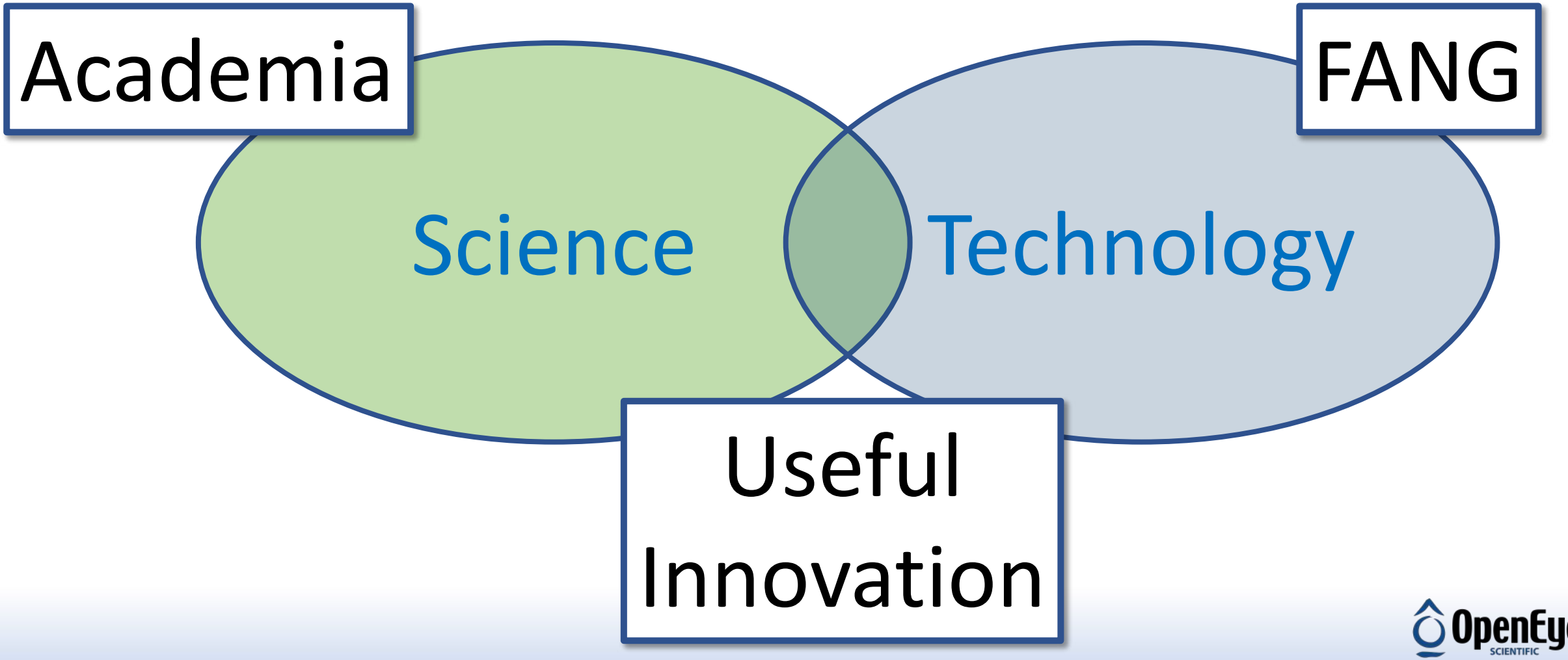
Impact on Drug Discovery Scales

- 7.8 Billion people
  - All of them get seriously ill
- 11% of people have a hospital stay (US)
  - 85% <65 years old
- 48% of people in the US use 1 therapeutic
  - 24% use 3 or more therapeutics
- Discovery Cost >> Manufacturing Cost
- WHO: 336 “essential medicines”

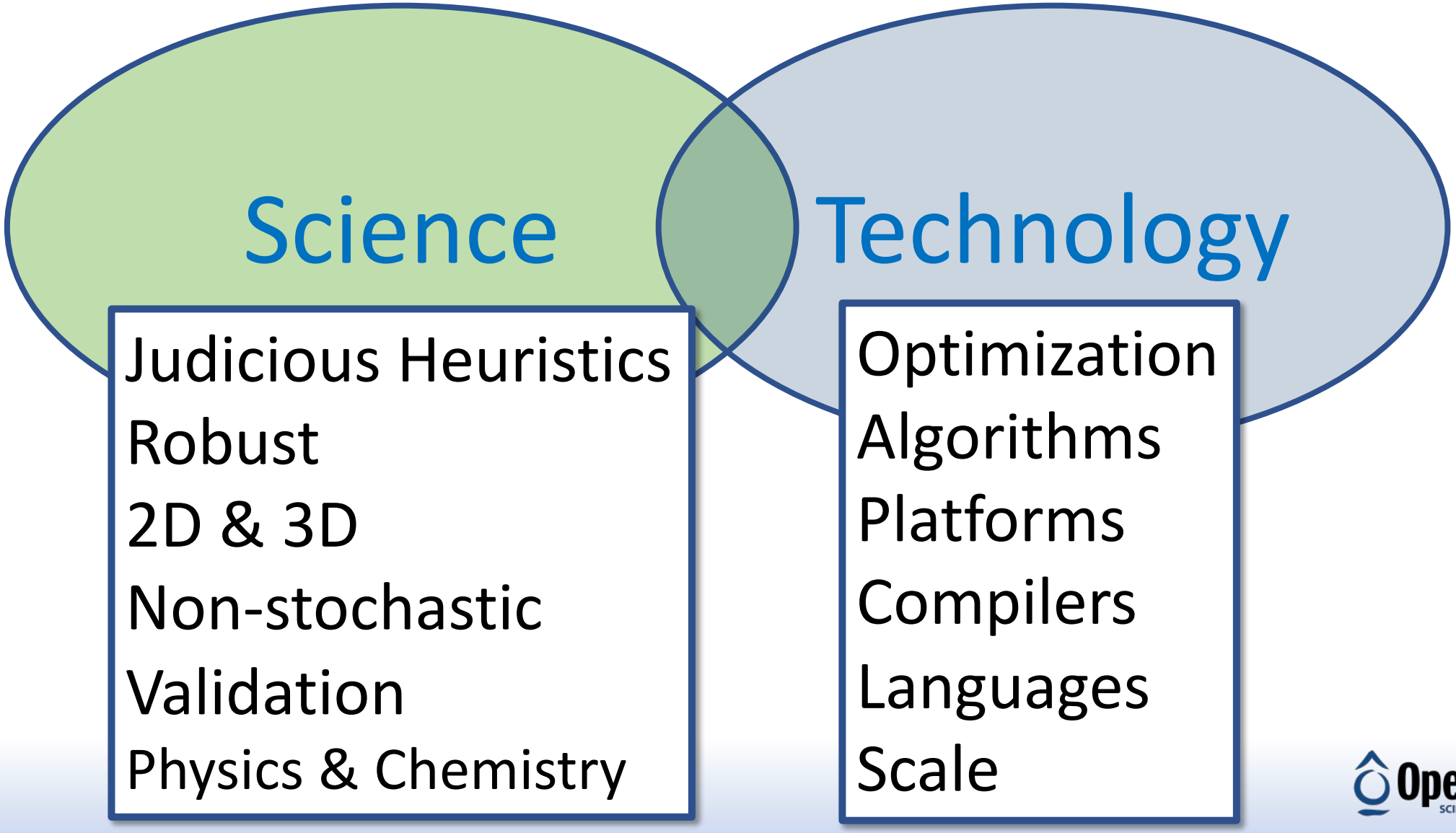
https://www.cdc.gov/nchs/data/hus/2018/038.pdf
https://www.who.int/health-topics/medicines

Why work at a software company?

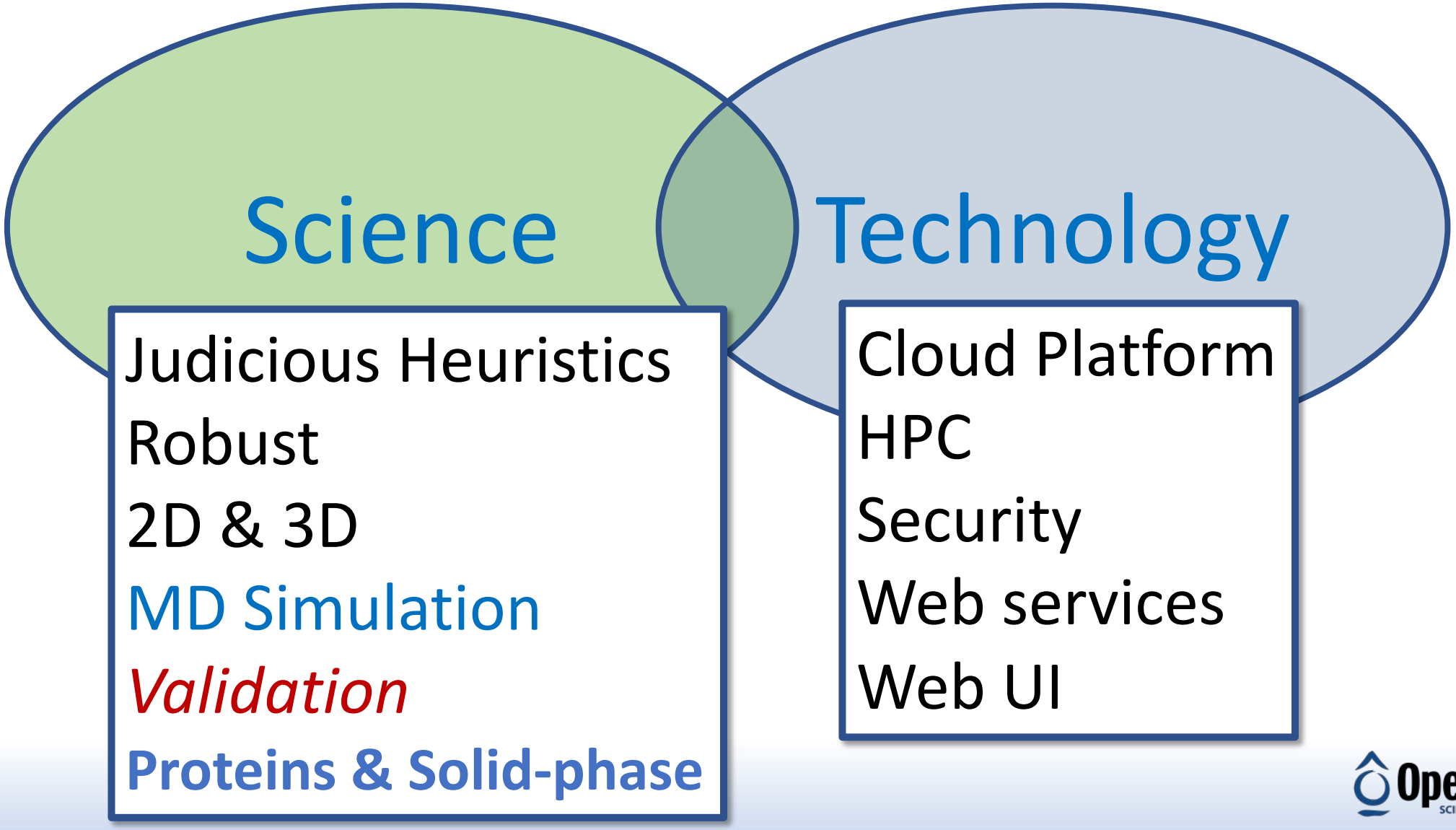
Working at the interface



Old-school problem-solving values

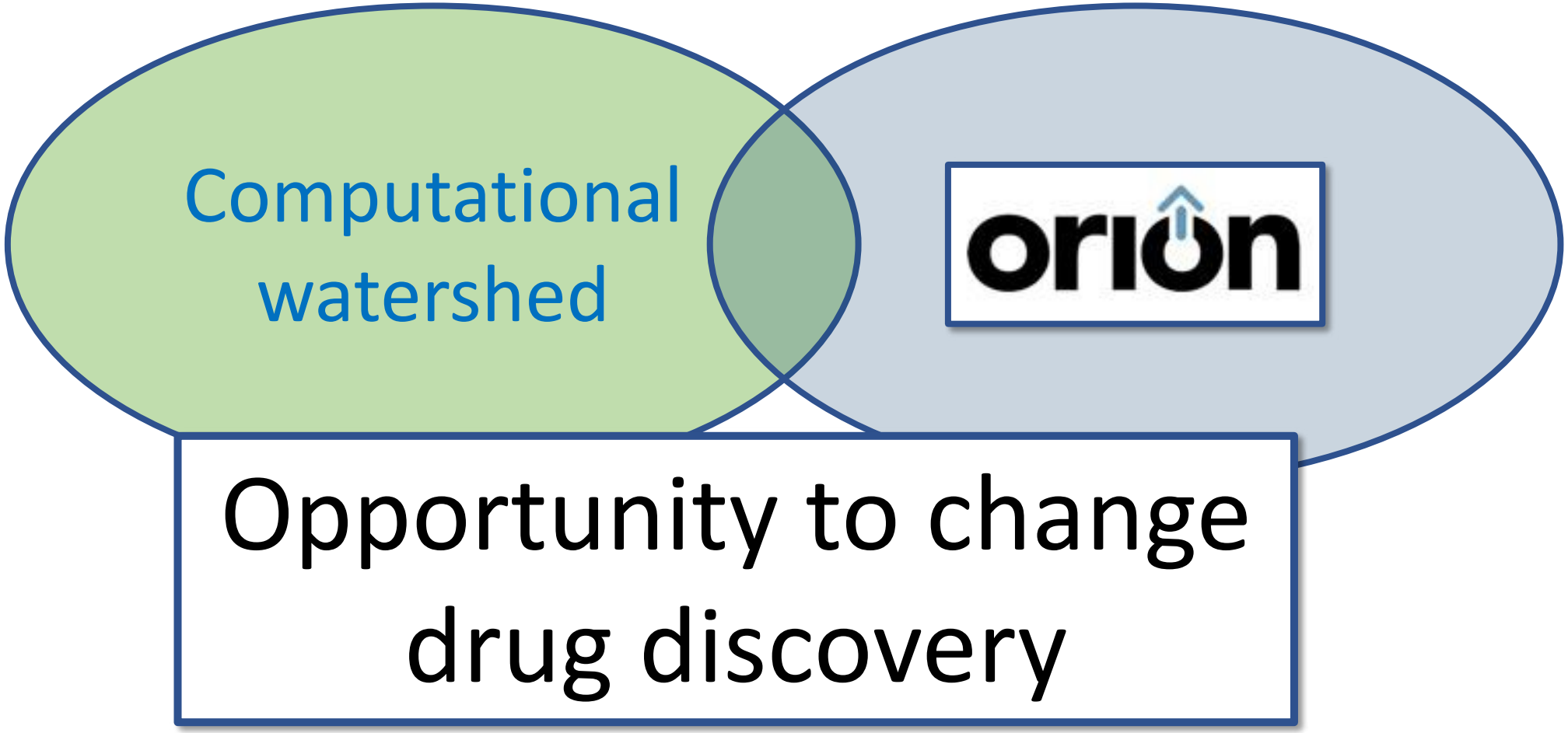


Modern interface values



1. Unprecedented scale
   - Robust, elastic, scalable, affordable
2. Hardware & software
   - Challenge & opportunity
3. Platform
   - Unified technologies
4. 3rd Party tools & opportunities
5. No computation too big

Modern impact



Drug Discovery Realms

- Biology
- Structure
- Complexes
- Mutants
- Selectivity

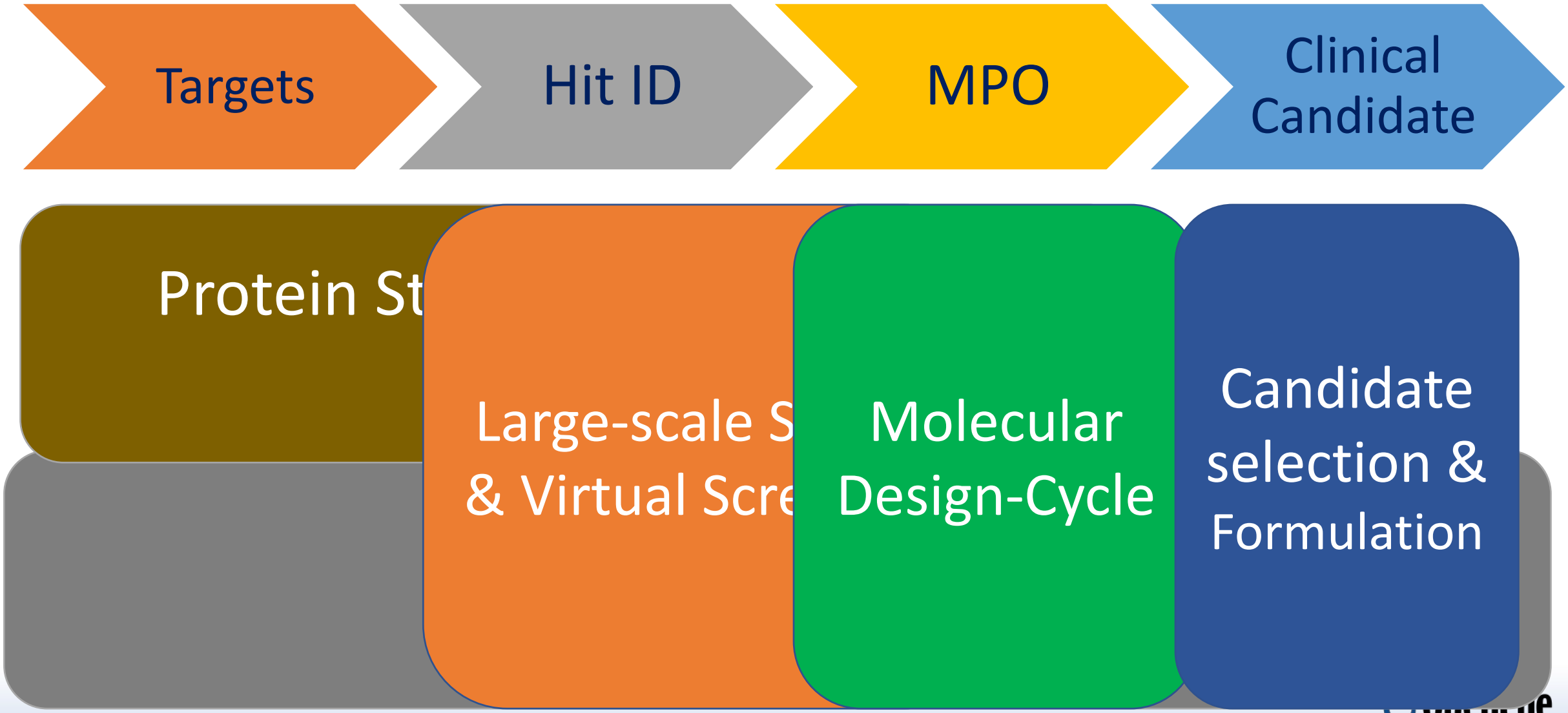
- HTS
- Virtual Screen
- Solubility
- hERG, PGP
- Drugability
- Small-molecule vs
- Antibody vs
- PROTAC vs
- bRo5
- Chemistry

- Potency
- Selectivity
- Bioavailability
- Liver stability
- Clearance
- Permeability
- Metabolites
- Distribution
- Toxicity!

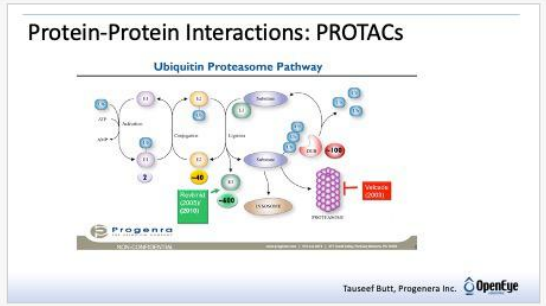
- Candidate selection
- Polymorph risk
- Formulation
- Manufacturing
- Shelf-life
- Hygroscopicity
- Efficacy!

OpenEye SCIENTIFIC

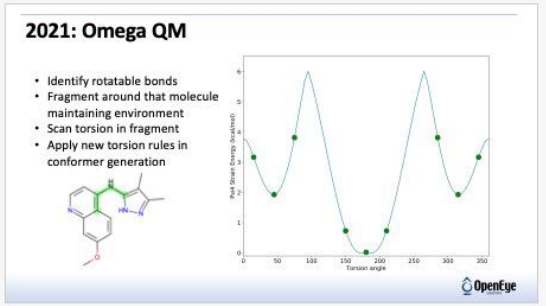
Drug Discovery Realms



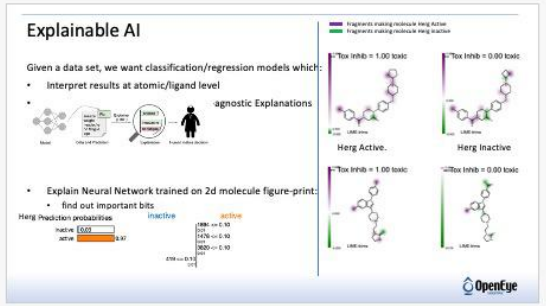
Exciting projects we don’t have time for today



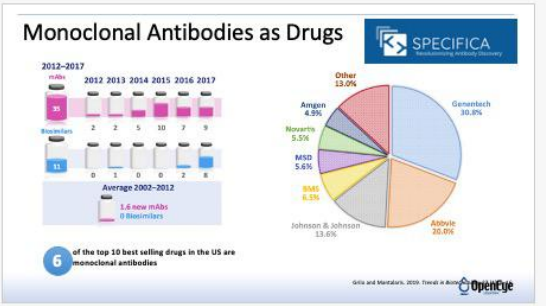
71



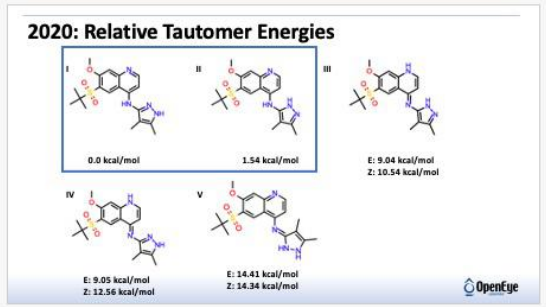
72



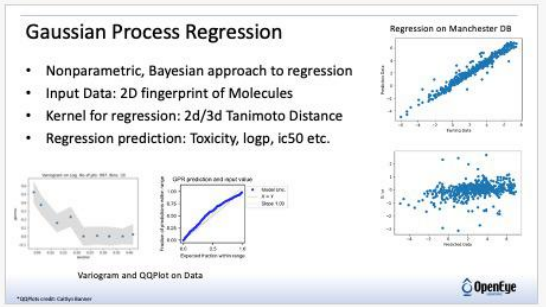
73



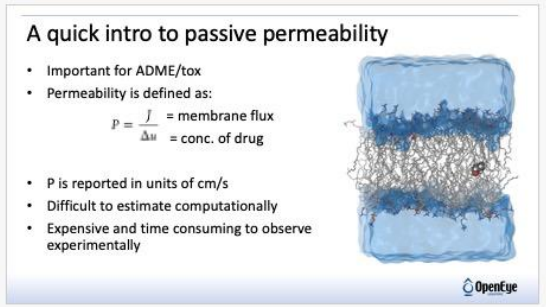
74



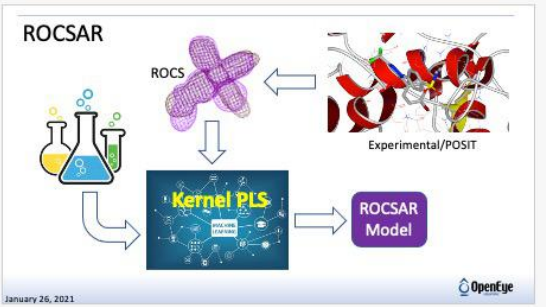
75



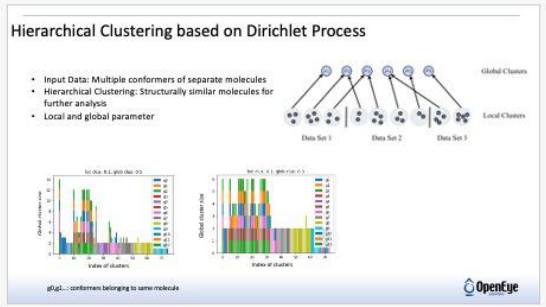
76



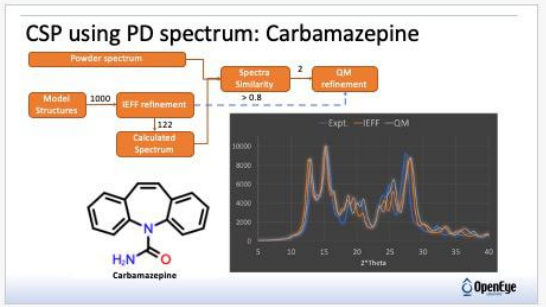
77



78



79



80

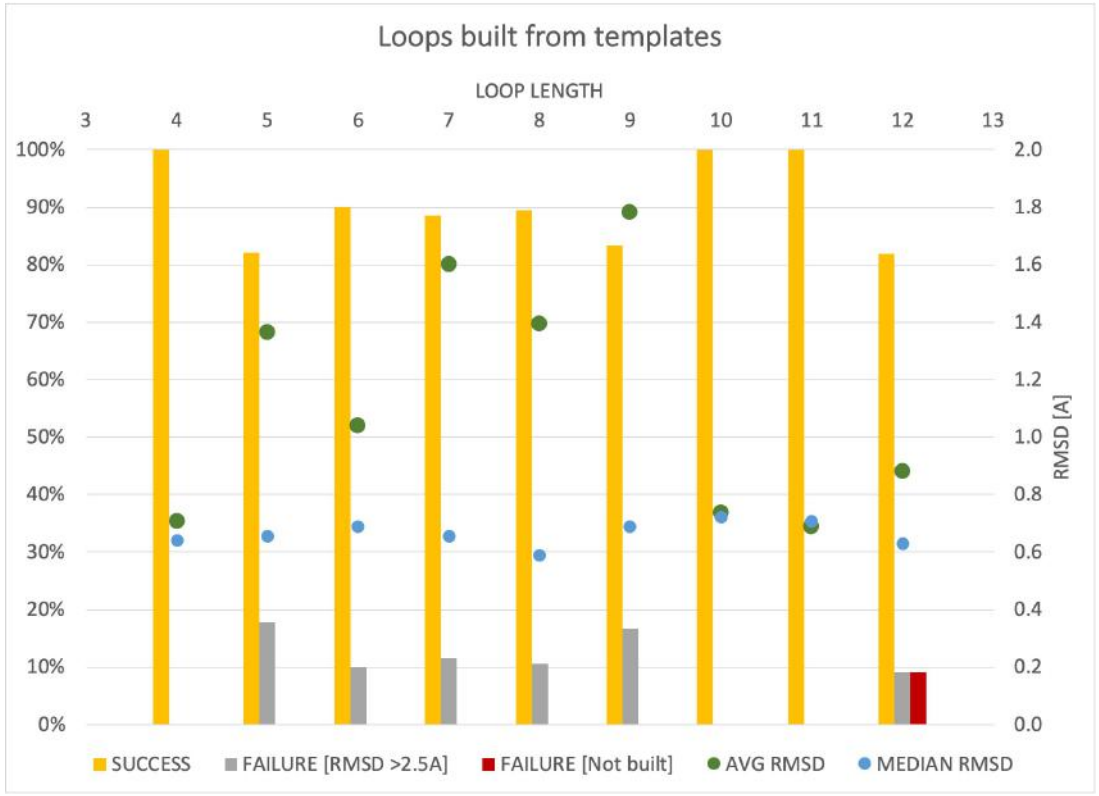
OpenEye SCIENTIFIC

Drug Discovery Realms

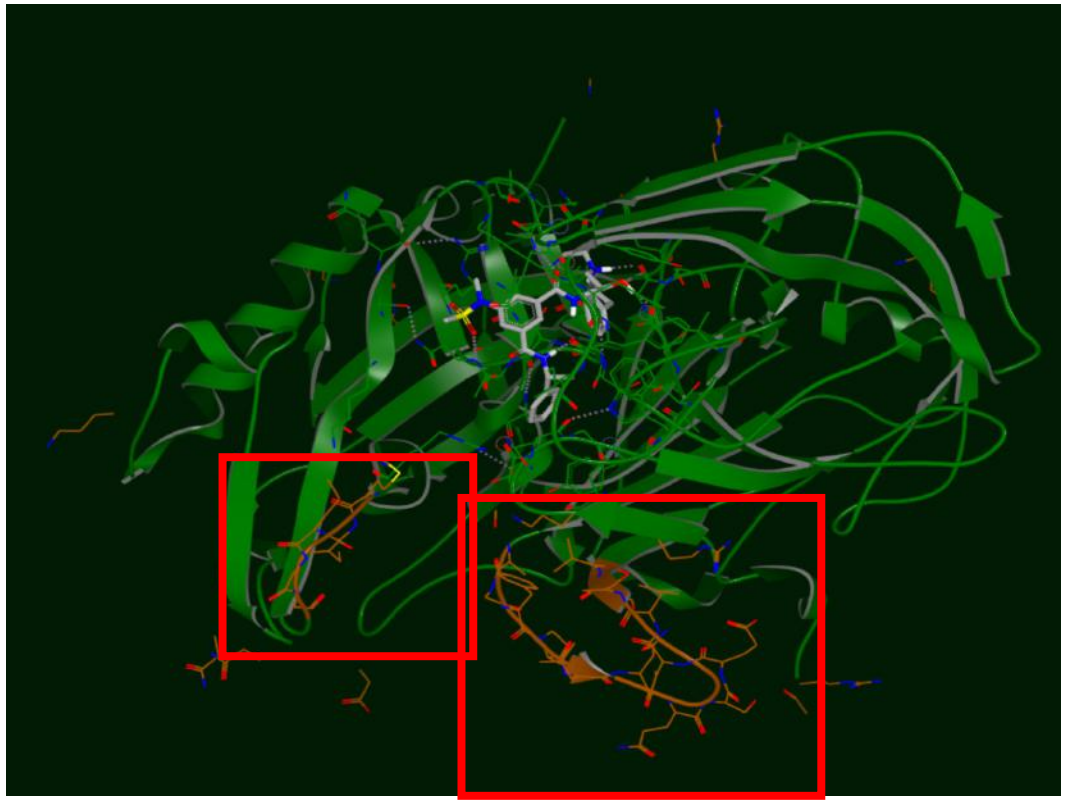
Protein Structure, Modeling & Simulation

Loop modeling

- Enabled by default – expected increased preparation times



Validation set: Rossi et al. (*Prot. Sci.* 2007)



PDB-ID: 3TPP. Modeled pieces in brown

```
DVA between GLU 310 A and THR 314 A
GFPLNQSEVLAS between ALA 157 A and VAL 170 A
```

Guide to Pharmacology

IUPHAR/BPS
Guide to PHARMA

Home | About | Targets | Ligands | Diseases | Resources | Advanced search | Immuno Portal

- GPCRs
- Ion channels
- Nuclear receptors
- Kinases
- Catalytic receptors
- Transporters
- Enzymes
- Other protein targets
- Target search

An expert-driven guid … ets and the substances that act on them.

Quick links

Targets

G protein-couple
Ion channels
Nuclear hormone
Kinases
Catalytic recepto
Transporters
Enzymes
Other protein tar

Ligands

Approved drugs
Synthetic organics
Metabolites
Natural products
Endogenous peptides
Other peptides

… new to Guide to PHARMACOLOGY

…avirus (Covid-19) - view our information page

…tabase version 2020.3 (19 Jun 2020) - full details in our blog post

…st release includes:

…RS-Cov-2 protein family which includes 13 members and is found under …nti-infective targets' in our hierarchy

… emerging oncology target, ACSS2, has been added plus two inhibitor …obe compounds (example 265 and ADG-207).

…dates have also been made to targets across 20+ different protein families

…er 200 ligands are now marked as antibiotic and more than 100 of these have with links to Antibiotic DB

- Improvements have been made to ligand summary pages. Key information, such as synonyms, curator comments, links to activity graph and SMILES and InChI Keys have been prioritised.

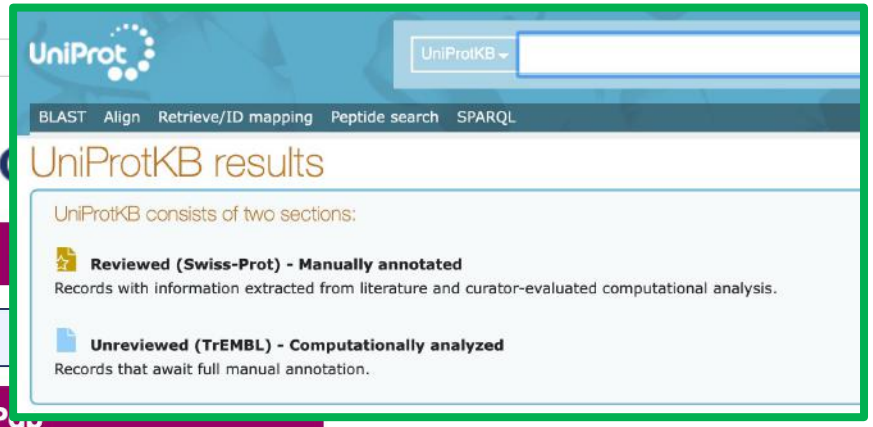
Please read our latest Database Report (April 2020) - download/view PDF

GtoMPdb

IUPHAR/MMV Guide to Malaria Pharmacology

Visit the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY portal
Launched in September 2019 to provide optimised access to GtoPdb data for the malaria research community

The Concise Guide to



Mapping key between GtoP and RCSB is UniprotKB accession #

OpenEye SCIENTIFIC

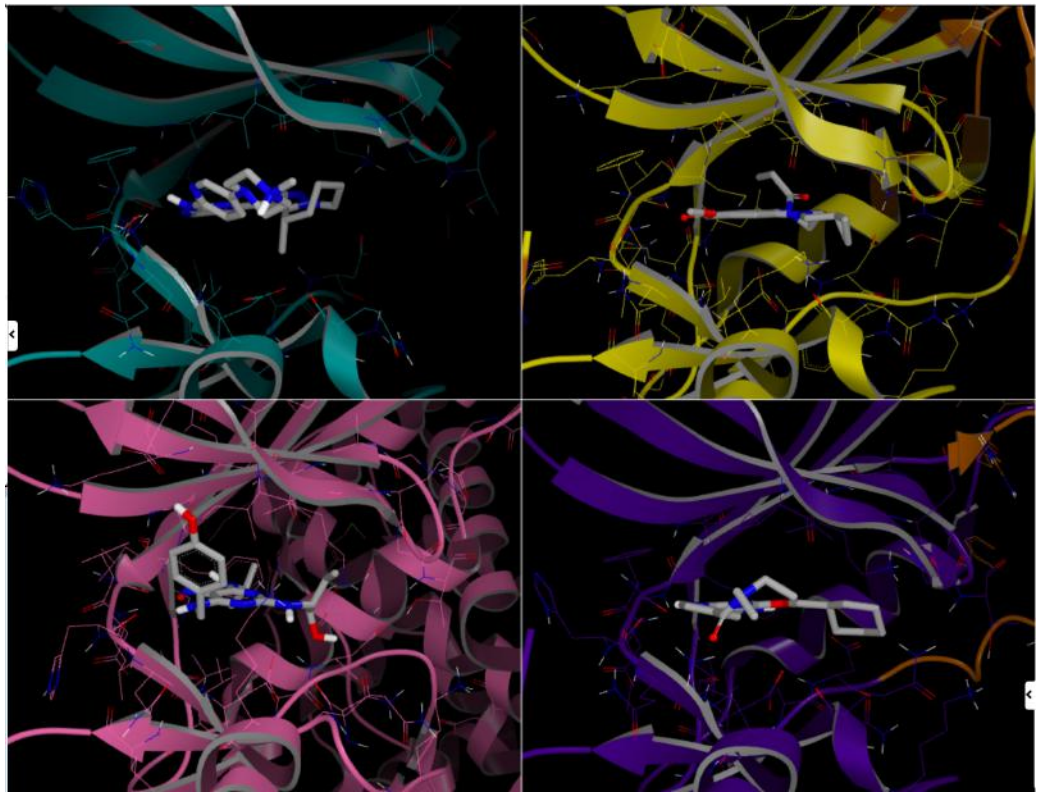
Core RCSB families in Orion

MMDS - Macromolecular Data Service

Sessions | Projects | Experiments

Search by project name... | Expand | Collapse

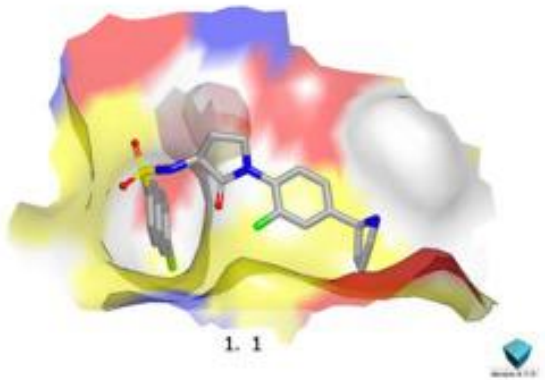
− RCSB	38502 sites
+ Enzymes	[illegible] sites
+ Ion channels	1689 sites
+ Other protein targets	5096 sites
− Receptors	3436 sites
+ Catalytic receptors	430 sites
+ G-protein coupled receptors	612 sites
− Nuclear hormone receptors	2394 sites
+ 0B. DAX-like receptors	13 sites
+ 1A. Thyroid hormone receptors	42 sites
+ 1B. Retinoic acid receptors	50 sites
+ 1C. Peroxisome proliferator-activated receptors	423 sites
+ 1D. Rev-Erb receptors	11 sites
+ 1F. Retinoic acid-related orphans	232 sites
+ 1H. Liver X receptors-like receptors	238 sites
+ 1I. Vitamin D receptors-like receptors	157 sites
+ 2A. Hepatocyte nuclear factor-4 receptors	30 sites
+ 2B. Retinoid X receptors	97 sites
+ 2E. Tailless-like receptors	2 sites
+ 3B. Estrogen-related receptors	86 sites
+ 4A. Nerve growth factor IB-like receptors	45 sites
+ 5A. Fushi tarazu F1-like receptors	44 sites
+ 6A. Germ cell nuclear factor receptors	2 sites
+ Steroid hormone receptors	922 sites
+ Transporters	220 sites



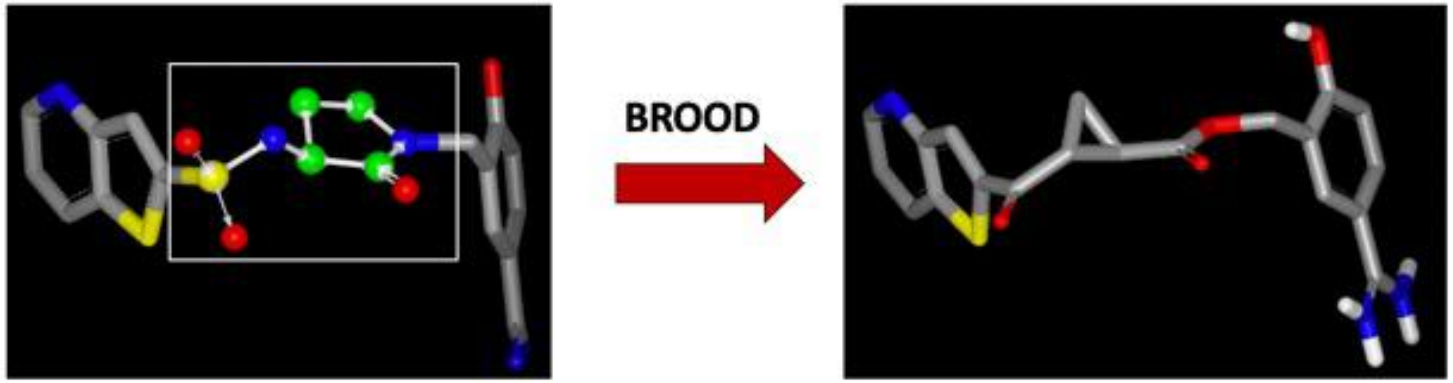
OpenEye SCIENTIFIC

Enabled by GPU Color Optimization

SiteHopper: Shape-based pocket similarity

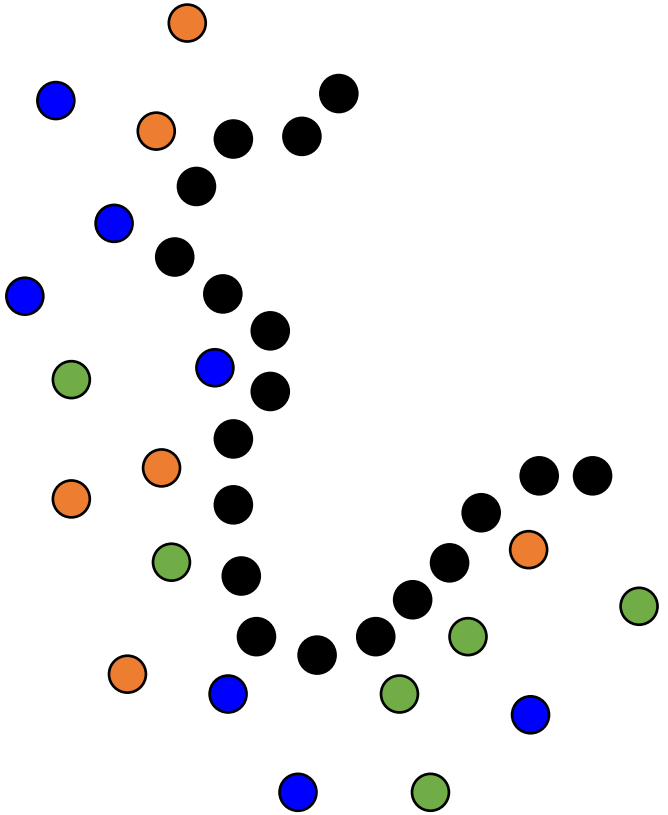


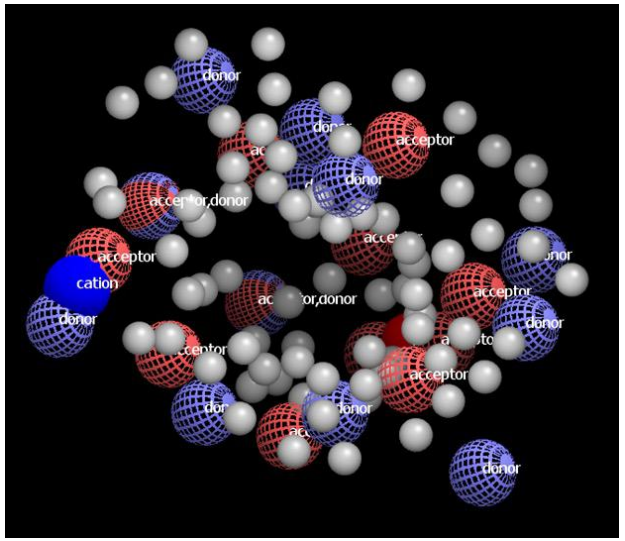
GPU BROOD: Fragment replacement



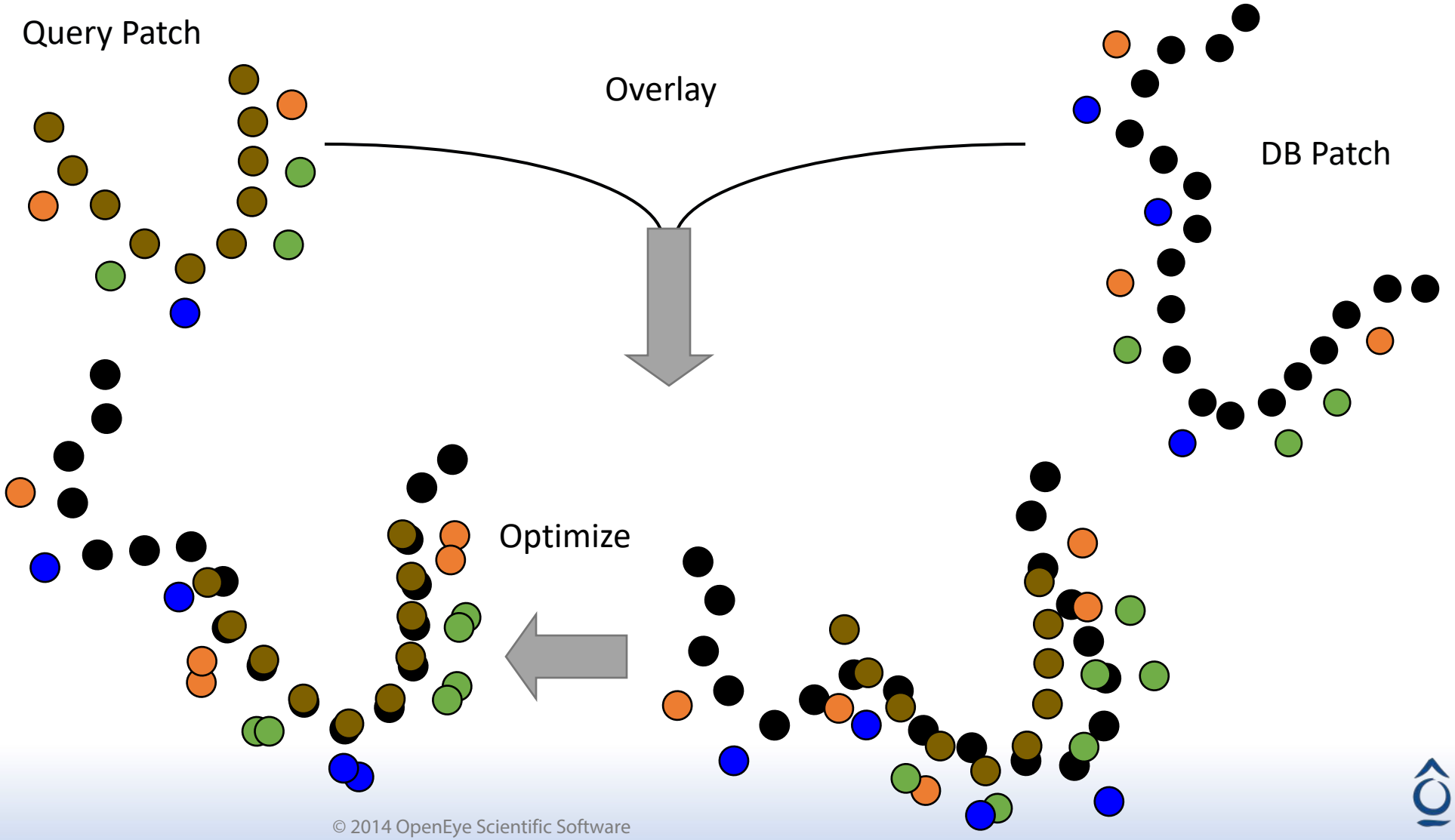
Why compare actives sites

- Organizing target families
- Transferring ligand binding (SAR transfer)
- Analyzing selectivity
- Anticipating off-target effects





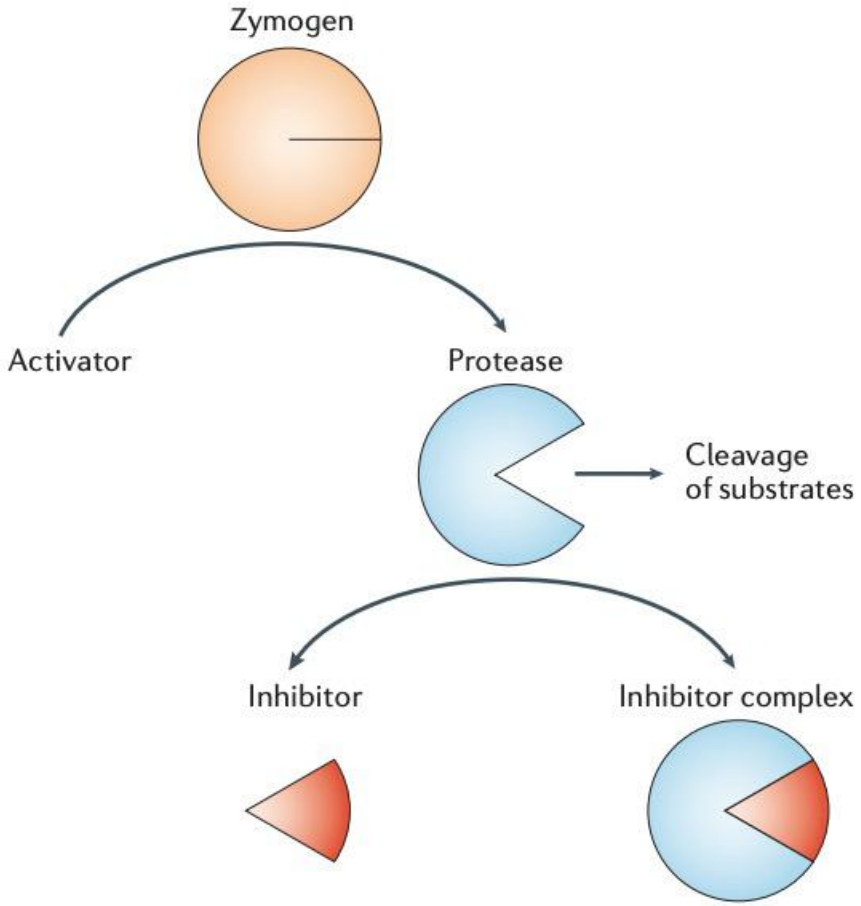
Overlay and alignment: Shape TK



GPU SiteHopper Progress

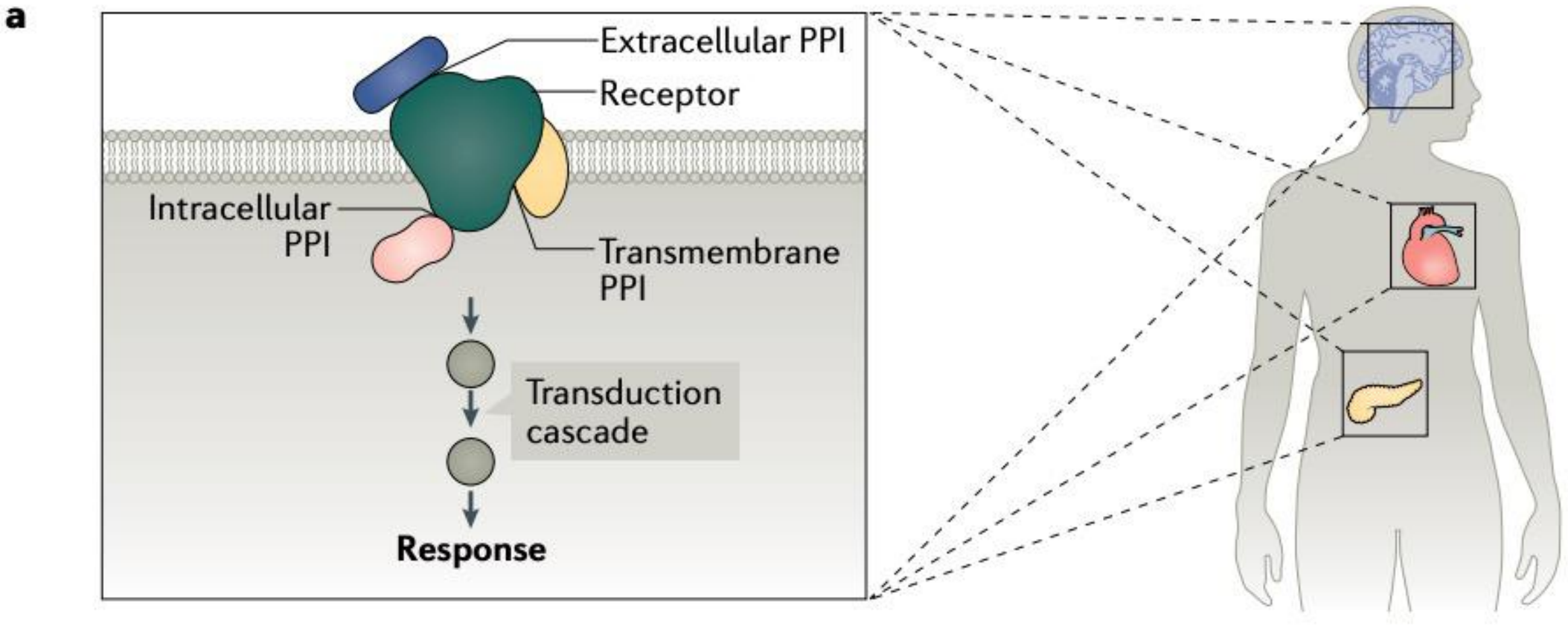
- Orion: Available
  - Floe to create SH Database from GtoP structures
  - Floe to query
- Desktop App: 2021.1 release
- Toolkit: 2021.2 release
- GPU Brood *after*

Enhanced Sampling MD simulations



"Targeting proteases: successes, failures and future prospects, Boris Turk, 5: 785, 2006.

Addressing Biological Complexity

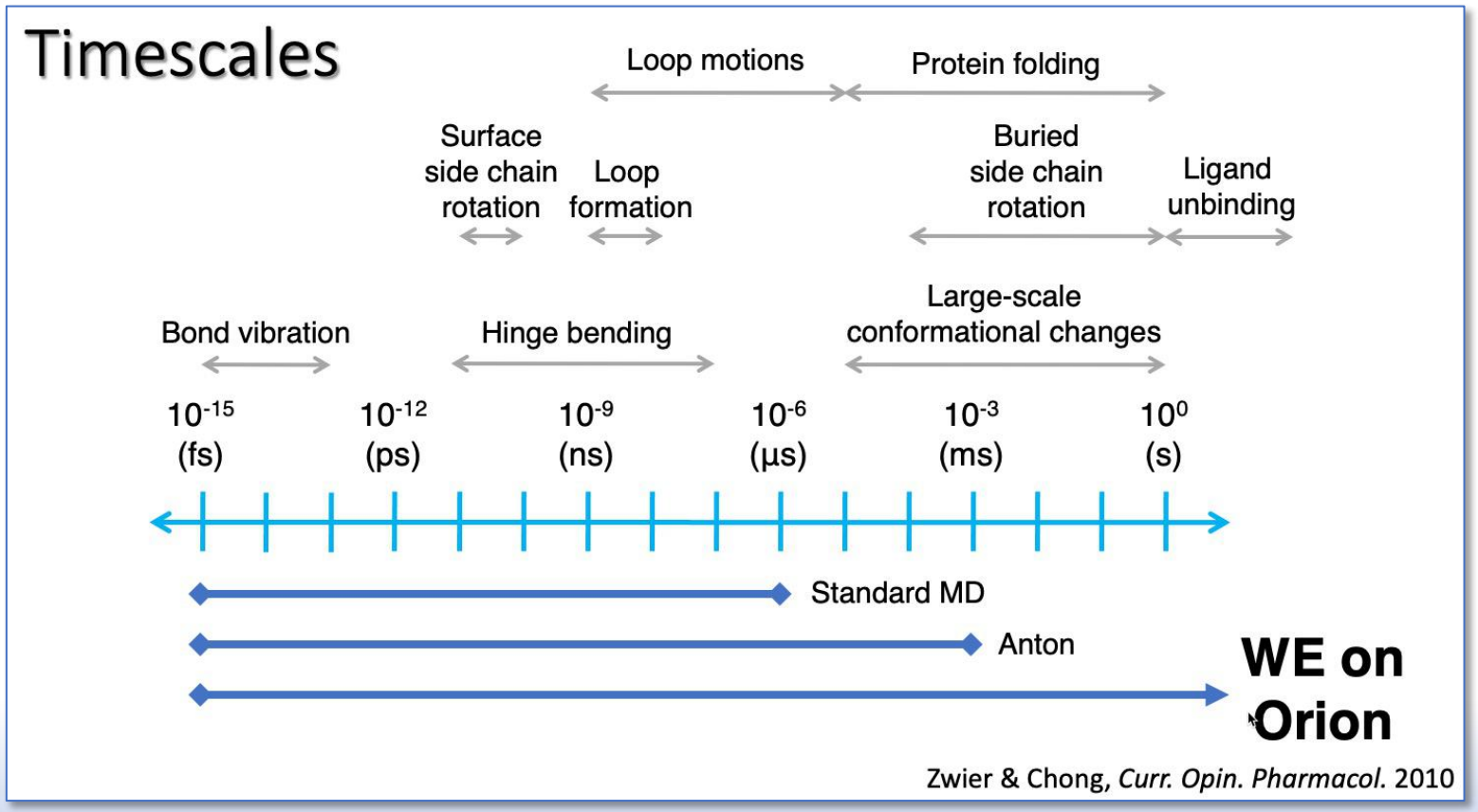


Rosenbaum et al, Nat. Rev. Drug Disc., 19: 884, 2020.

OpenEye SCIENTIFIC

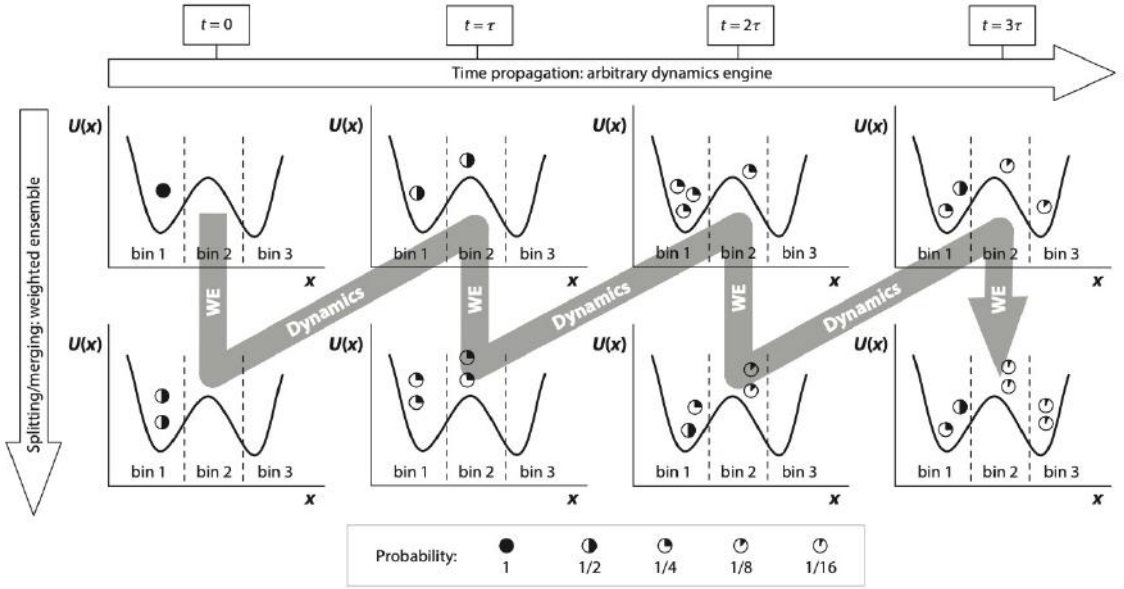
Enhance Sampling MD simulations

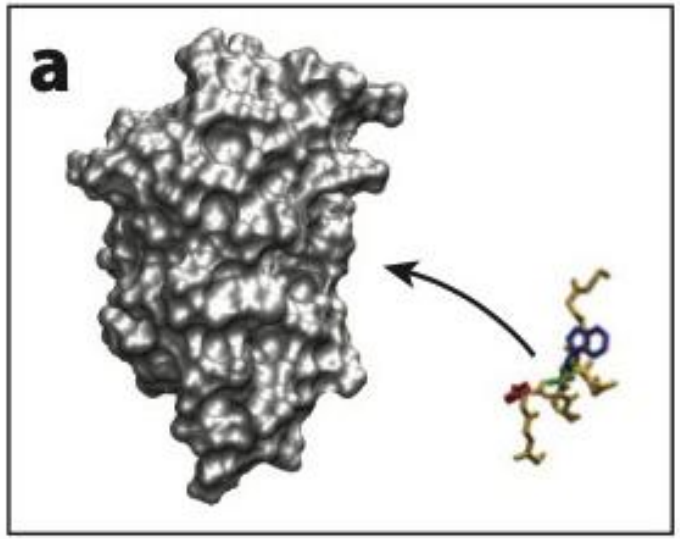
- Tools



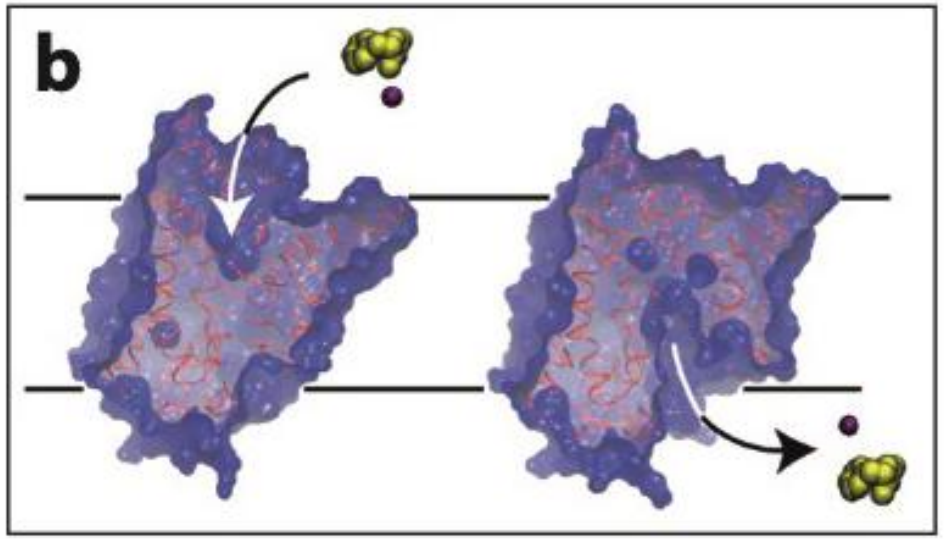
Weighted Ensemble MD

- Unbiased Hamiltonian
- Simulation pathways
- Efficient, selective sampling

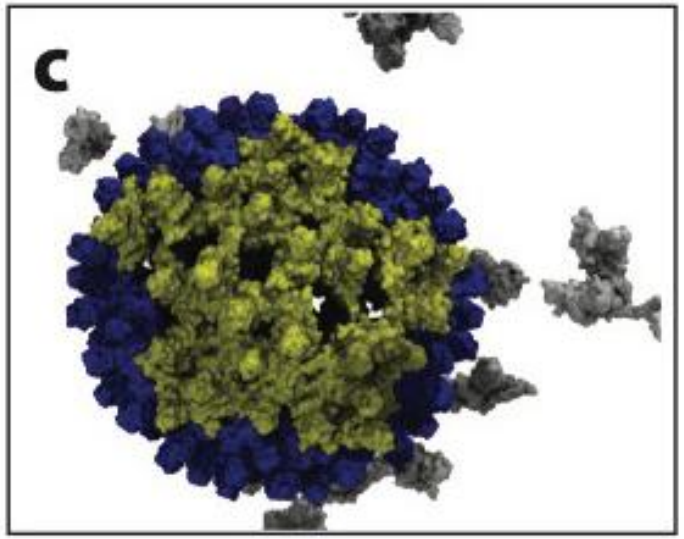




Protein binding



Conformational change



Macromolecular assembly

Zuckerman & Chong, Annu. Rev. Biophys. 46:43-57, 2017

OpenEye SCIENTIFIC

Cloud-based Enhanced-sampling Simulations

- OpenEye goals
  - Tools for protein complexes
  - Enhanced sampling MD for large-scale & rare protein motion
    - WESTPA
    - REMD
  - Understanding conformational states of proteins & complexes
    - Markov-state models relating the states
    - Isoform, mutant, allosteric comparisons
  - Opportunities for selective drug discovery

Drug Discovery Realms

Large-scale Search
& Virtual Screening

Unprecedented Scale

- GigaDocking
  - Docking >3Bn available molecules in 24 hours
  - Multiple successes
  - Service & in Orion

- Robust & Easy
- Databases

Enamine

REAL Database 2.5 Bn
REAL Space 10-20x

mcule
·MORE·THAN·MOLECULES·

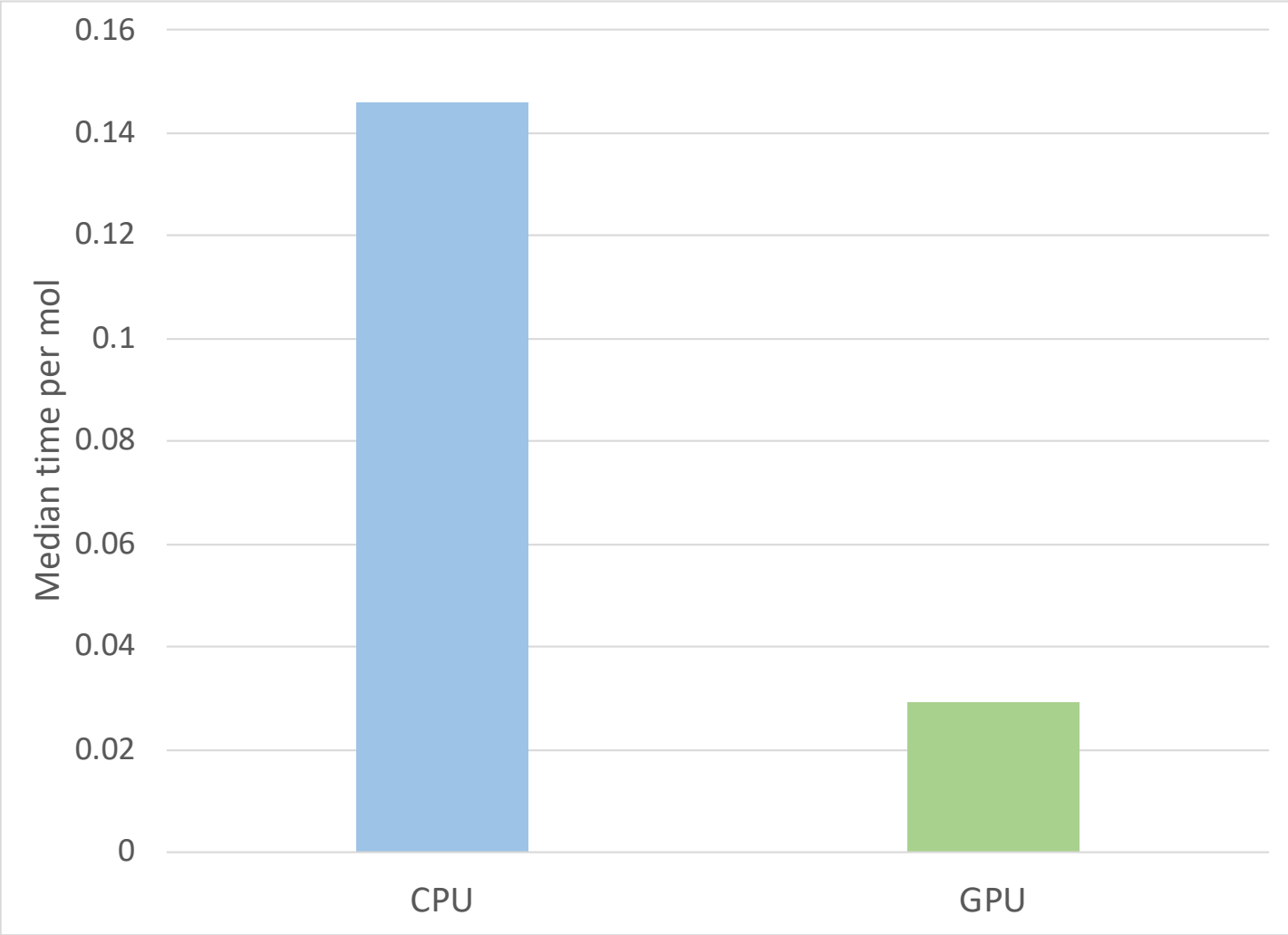
150M – no overlap

WuXi AppTec

1.0Bn – Ro5 filtering

OpenEye SCIENTIFIC

GPU OMEGA improves speed

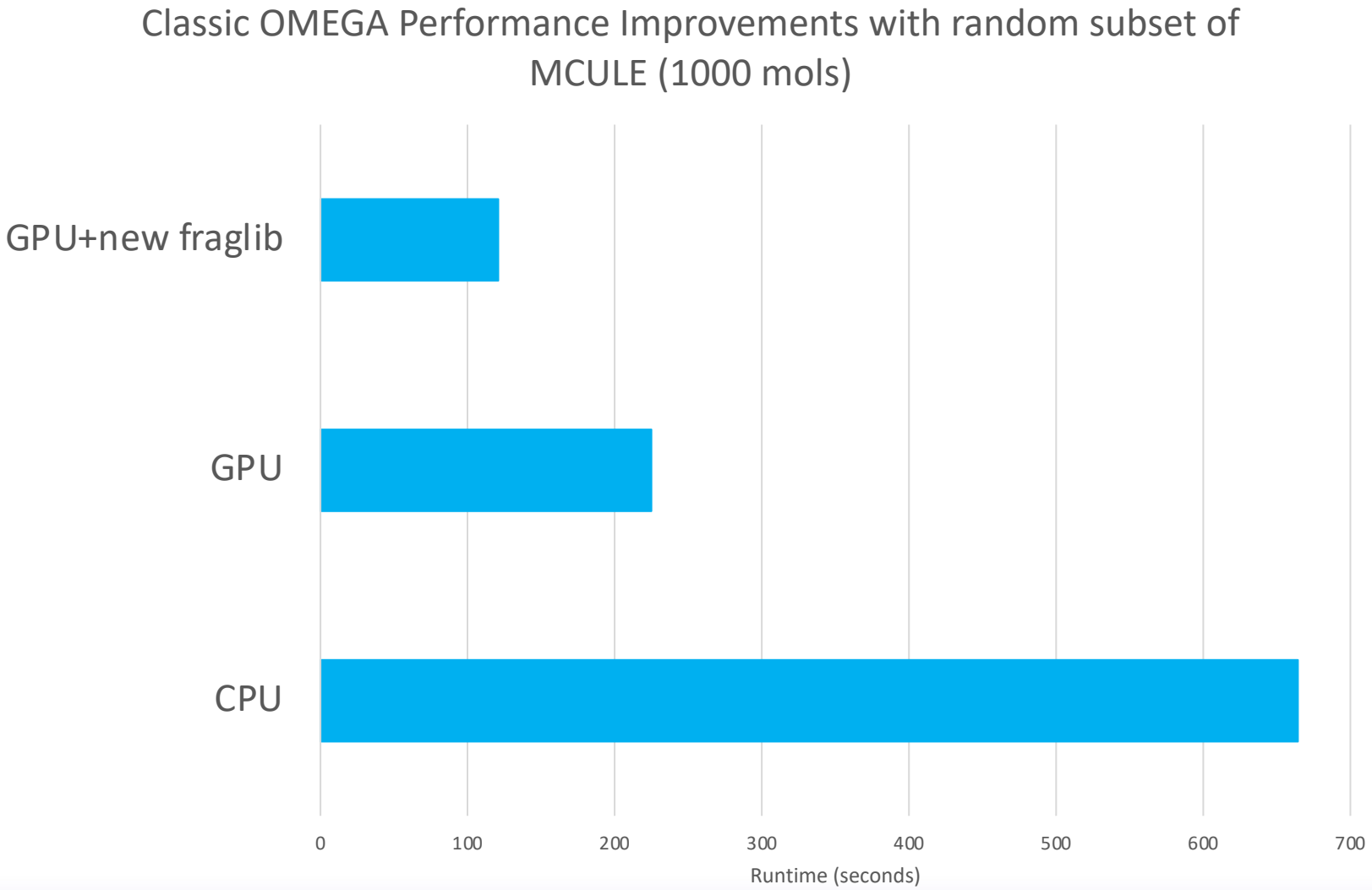


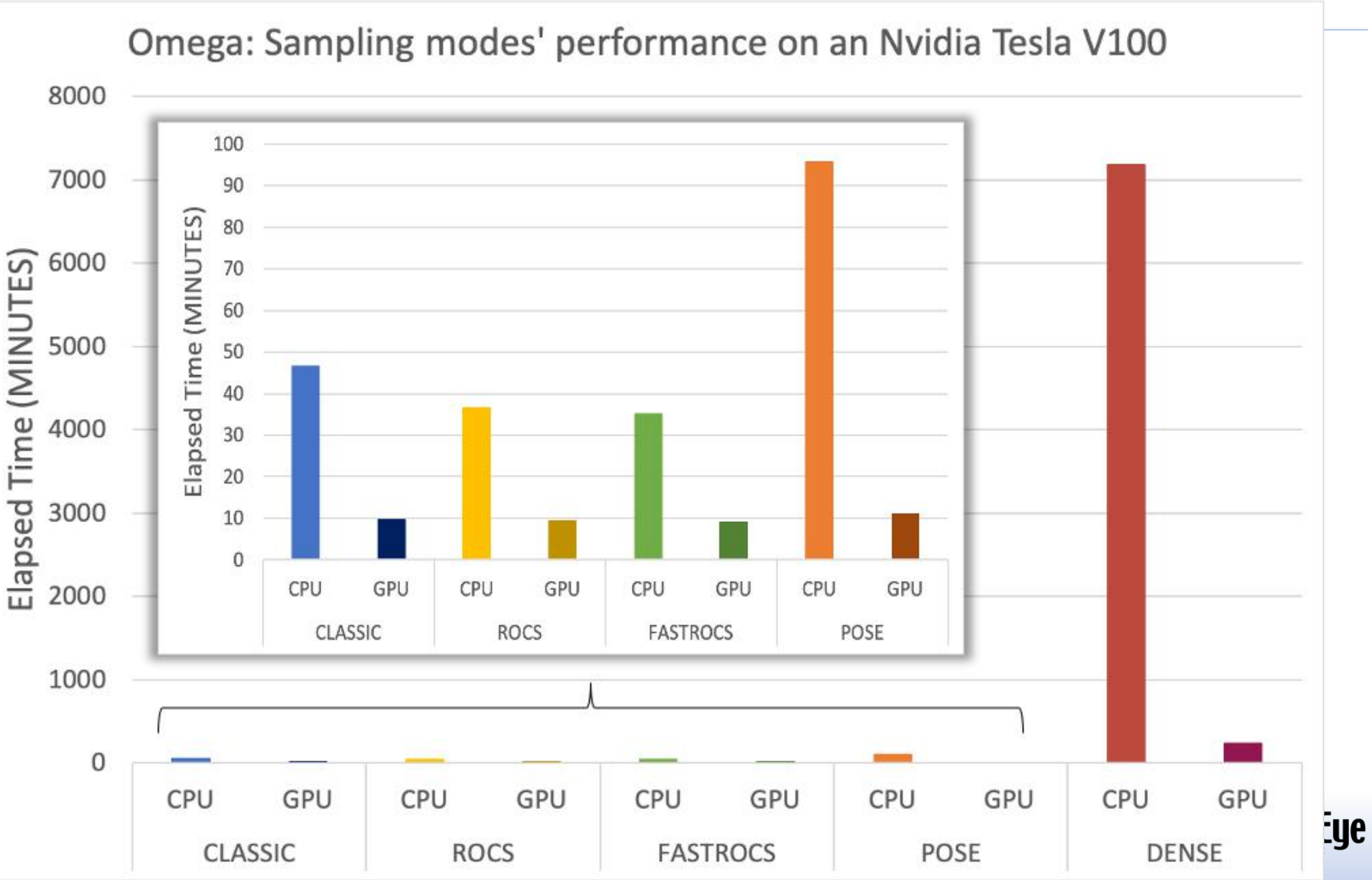
Internal tests:
OMEGA median time = 0.14 secs/molecule
(Ubuntu 16, Intel(R) Xeon(R) Gold 6128 3.40GHz CPU)

FastOMEGA median time = 0.03 secs/molecule
(Ubuntu 16, V100 GPU)

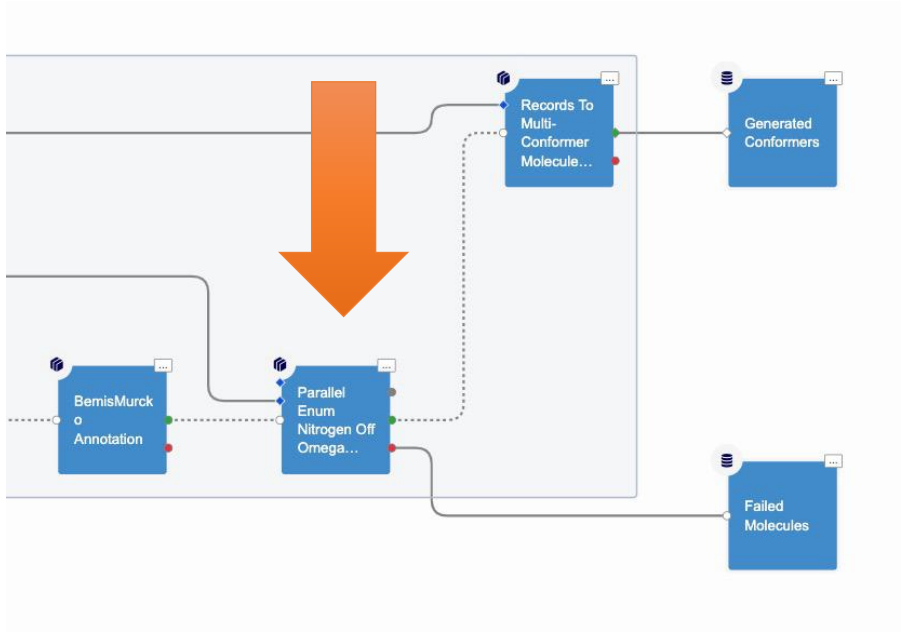
Dataset: Friedrich *et al.*, *J. Chem. Inf. Model.*, **57**, 539 (2017).

Omega new built-in fragment library



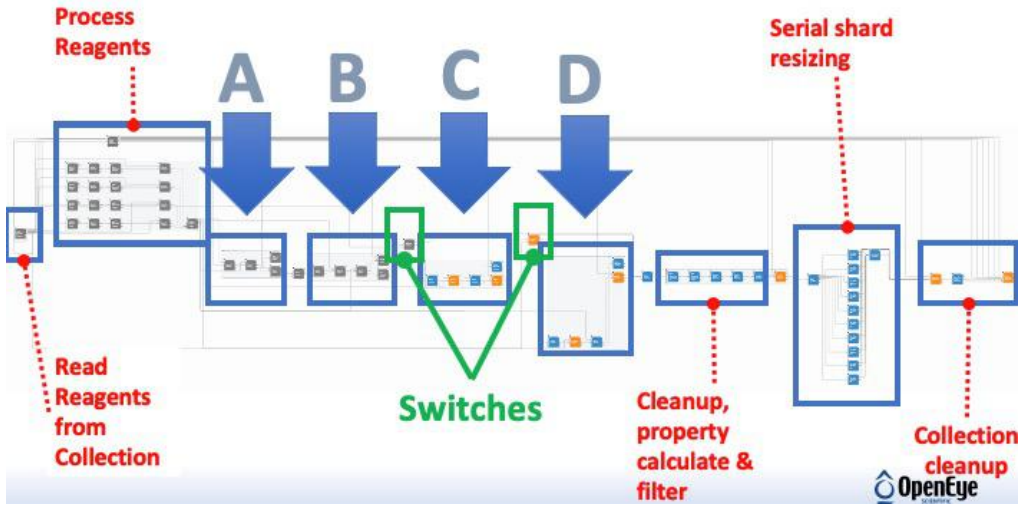

Omega: Sampling modes' performance on an Nvidia Tesla V100
Elapsed Time (MINUTES)
8000
7000
6000
5000
4000
3000
2000
1000
0
CPU
GPU
CLASSIC
ROCS
FASTROCS
POSE
DENSE
Elapsed Time (MINUTES)
100
90
80
70
60
50
40
30
20
10
0
CPU
GPU
CLASSIC
ROCS
FASTROCS
POSE

MT GPU-Omega in Orion – ***~3x less expensive***



Database Prep Floe

2021 FastROCS & Enamine REAL Space 10^10 molecules



OpenEye SCIENTIFIC

Beyond Brut Force

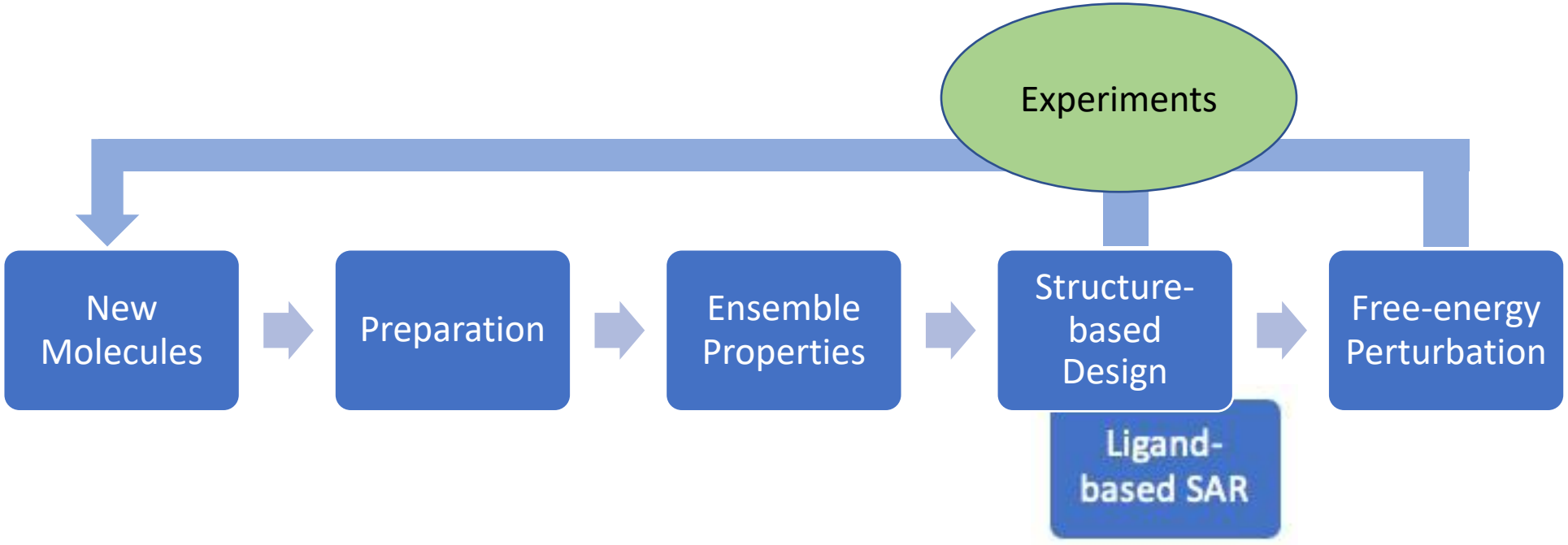
- ROCS & Docking
- Enamine is made from:
  - <150 libraries
  - Each library has 2-4 reagents types A(1..m), B(1..n),[(C(1..p),] D(1..q)]
- Reagent space searching is O(N) rather than O(N^2)
  - memory & search
- Reagent Search
  - Substructure (exact)
  - FP similarity
  - ROCS
  - Docking

OpenEye SCIENTIFIC

Drug Discovery Realms

Molecular Design-Cycle

Simplified Molecular Design

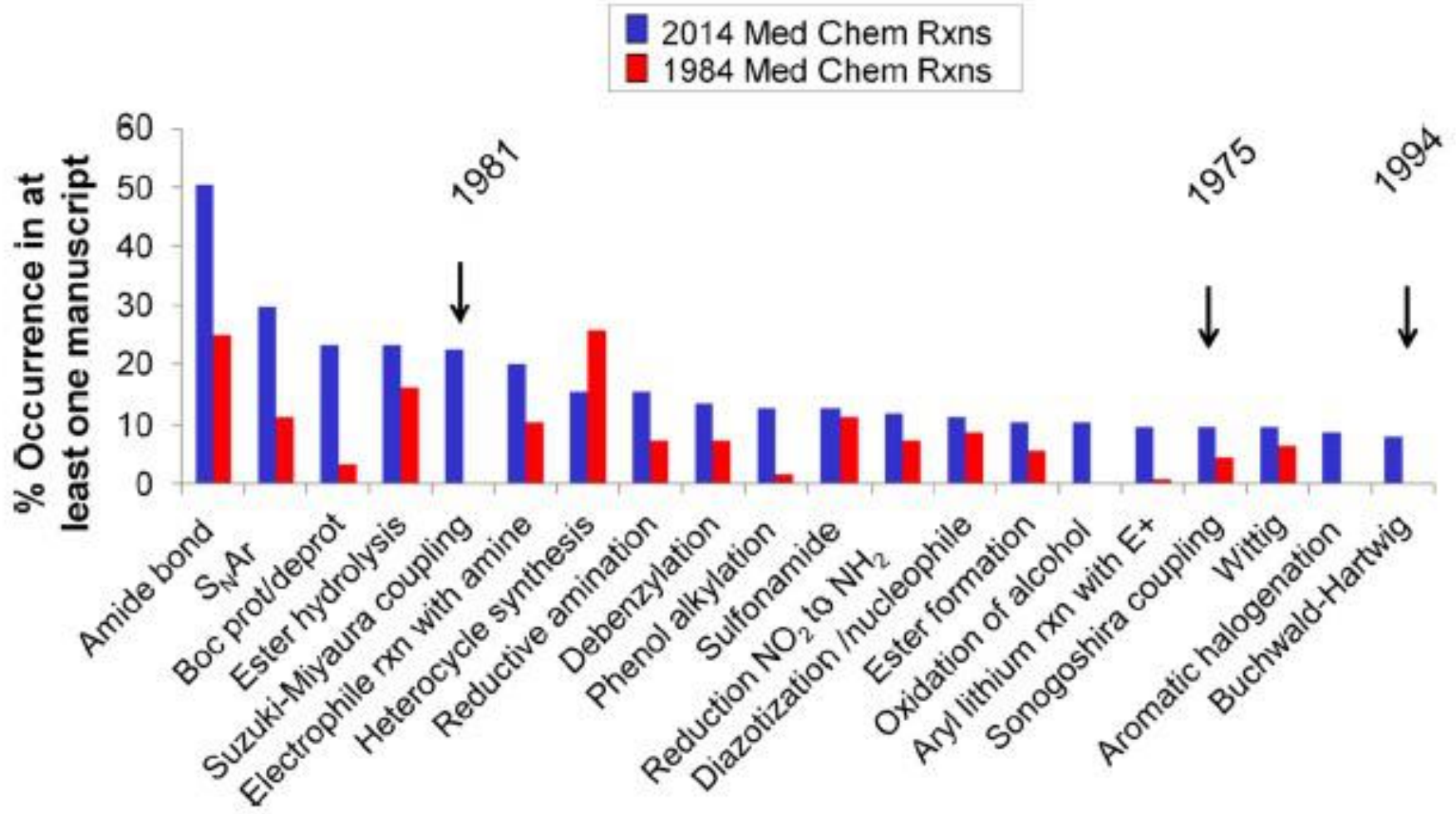


Generative floes

OpenEye SCIENTIFIC

Reactions in the Medicinal chemists' toolbox

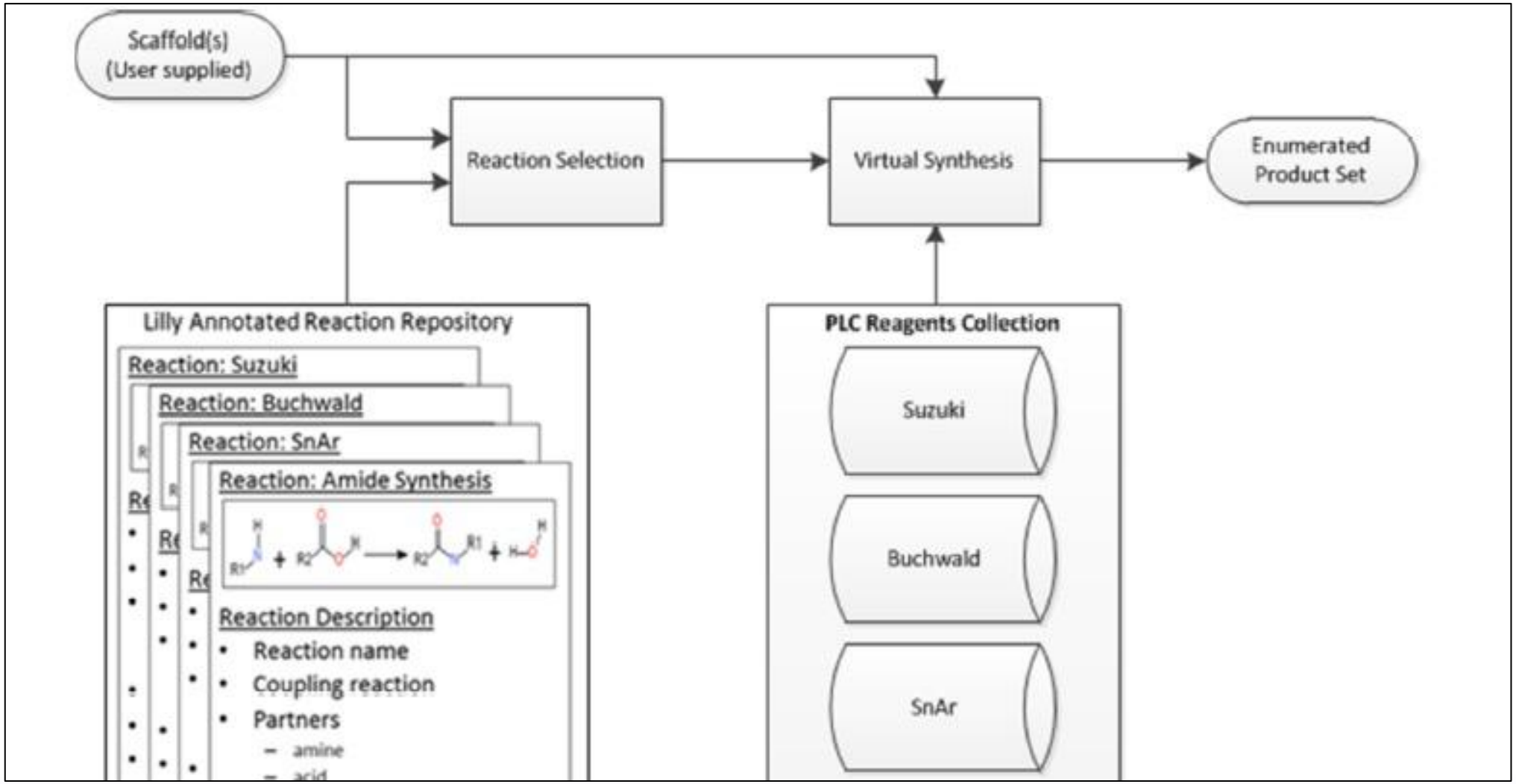
top reactions generally have largely remained the same for the past 3 decades



Brown & Boström, J. Med. Chem., 59(10):4443-4458, 2016.

OpenEye SCIENTIFIC

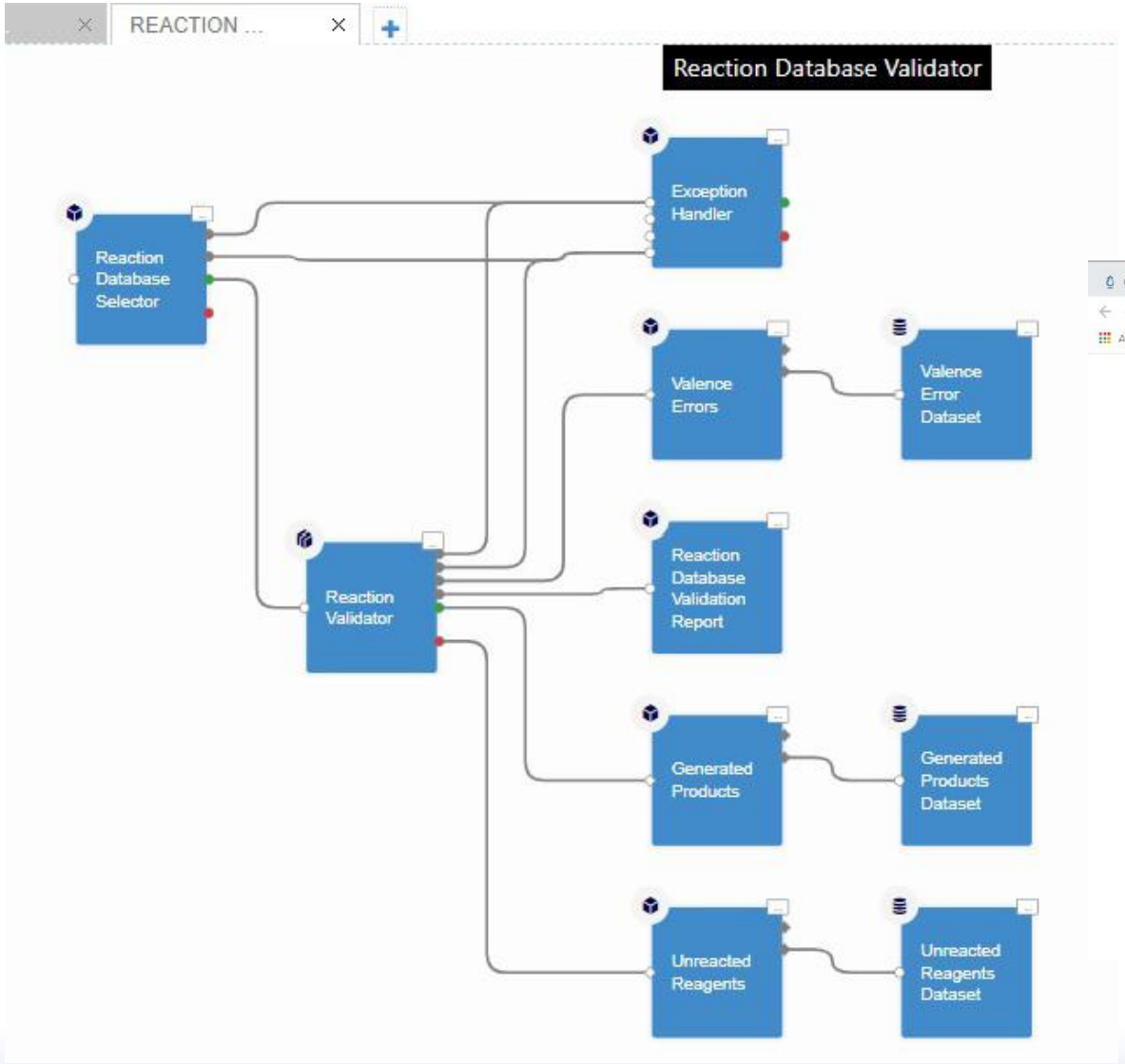
Focused Library Design



Nicolaou, et al., JCIM, 56(7): 1253-1266, 2016

OpenEye SCIENTIFIC

Reaction Database Directory Listing



OpenEye Orion | OpenEye Orion | OpenEye Orion

orion-qa.eyesopen.us

Apps | Staging | QA | Released | Dev | AWS Status | ZIA | ZIASite | Orion - MaaS

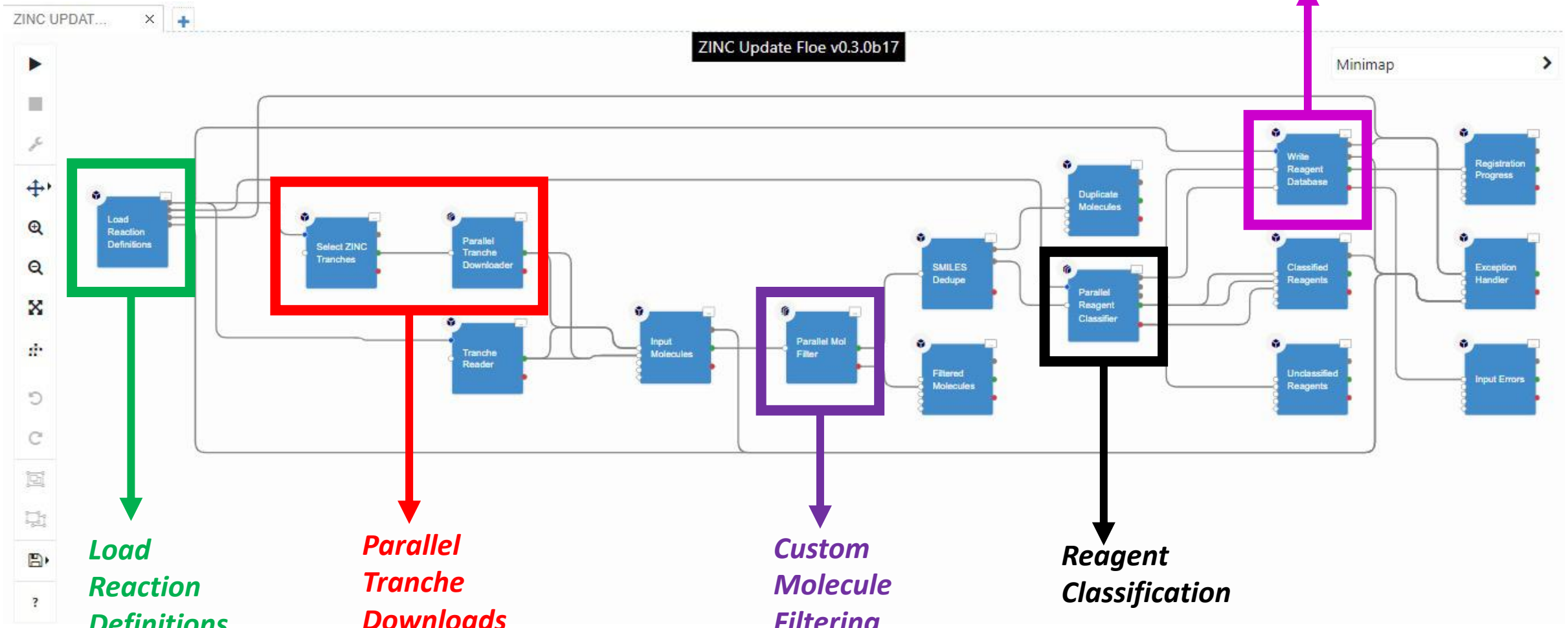
Reaction Database:
phase_III_6M_STATS.db

Reaction Name	Reagent1	Reagent1 Count	Reagent2	Reagent2 Count	Product Count
3-nitrile-pyridine	3-nitrile-pyridine:Diones_2_4	541			541
Buchwald-Hartwig	Buchwald-Hartwig:Halides_aryl	3,337	Buchwald-Hartwig:Amines	182,760	609,870,120
Buchwald_cross_coupling1	Buchwald_cross_coupling1:Amines	257,939	Buchwald_cross_coupling1:Aryl_halides	248,751	64,162,584,189
Buchwald_cross_coupling2	Buchwald_cross_coupling2:Amines	78,758	Buchwald_cross_coupling2:Aryl_halides	248,751	19,591,131,258
Ester_hydrolysis-amide_synthesis1	Ester_hydrolysis-amide_synthesis1:Amines	281,409	Ester_hydrolysis-amide_synthesis1:Esters	63,048	17,742,274,632
Ester_hydrolysis-amide_synthesis2	Ester_hydrolysis-amide_synthesis2:Amines	168,246	Ester_hydrolysis-amide_synthesis2:Esters	63,048	10,607,573,808
Grignard_alcohol	Grignard_alcohol:Ketones_aldehydes	86,355	Grignard_alcohol:Halides_alkyl	611	52,762,905
Grignard_carbonyl	Grignard_carbonyl:Nitriles	38,286	Grignard_carbonyl:Halides_alkyl_aryl	5,072	194,186,592
Heck_non-terminal_vinyl	Heck_non-terminal_vinyl:Non_terminal_vinyls	1,674	Heck_non-terminal_vinyl:Halide_vinyl_aryls	2,384	3,990,816
Heck_terminal_vinyl	Heck_terminal_vinyl:Halide_vinyl_aryls	2	Heck_terminal_vinyl:Terminal_vinyls	2	4
Huisgen_disubst-alkyne	Huisgen_disubst-alkyne:Alkynes_disubstituted	1,926	Huisgen_disubst-alkyne:Alkyl_halides_alcohols	45	86,670
Mitsunobu_imide	Mitsunobu_imide:Alcohols_primary_secondary	104,010	Mitsunobu_imide:Acetylacetamides	2,190	227,781,900
Mitsunobu_phenol	Mitsunobu_phenol:Alcohols_primary_secondary	104,010	Mitsunobu_phenol:Phenols	23,705	2,465,557,050

OpenEye SCIENTIFIC

ZINC Update Floe

sqlite registration (db saved as Orion File resource)



Load Reaction Definitions (from Orion File resource)

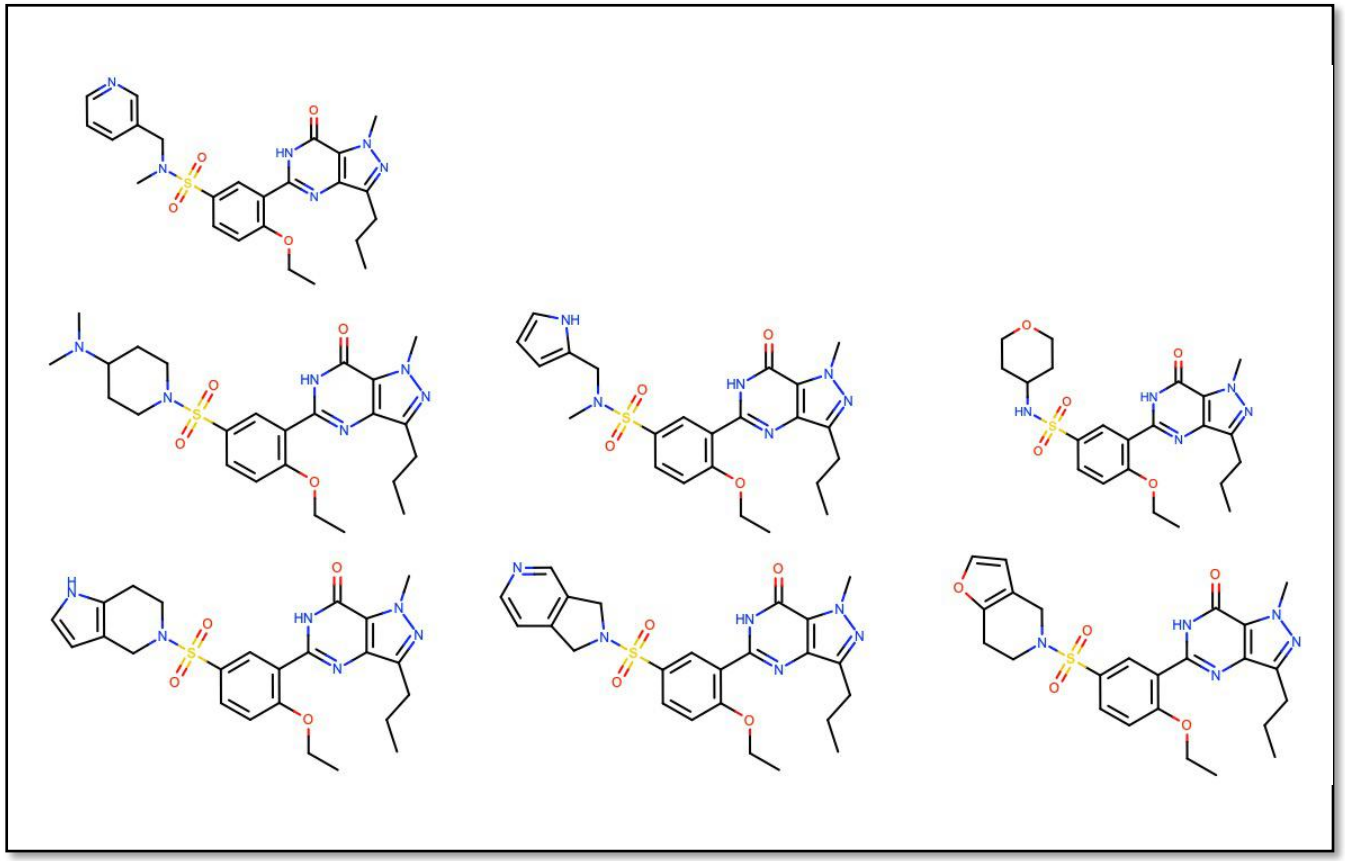
Parallel Tranche Downloads (by id)

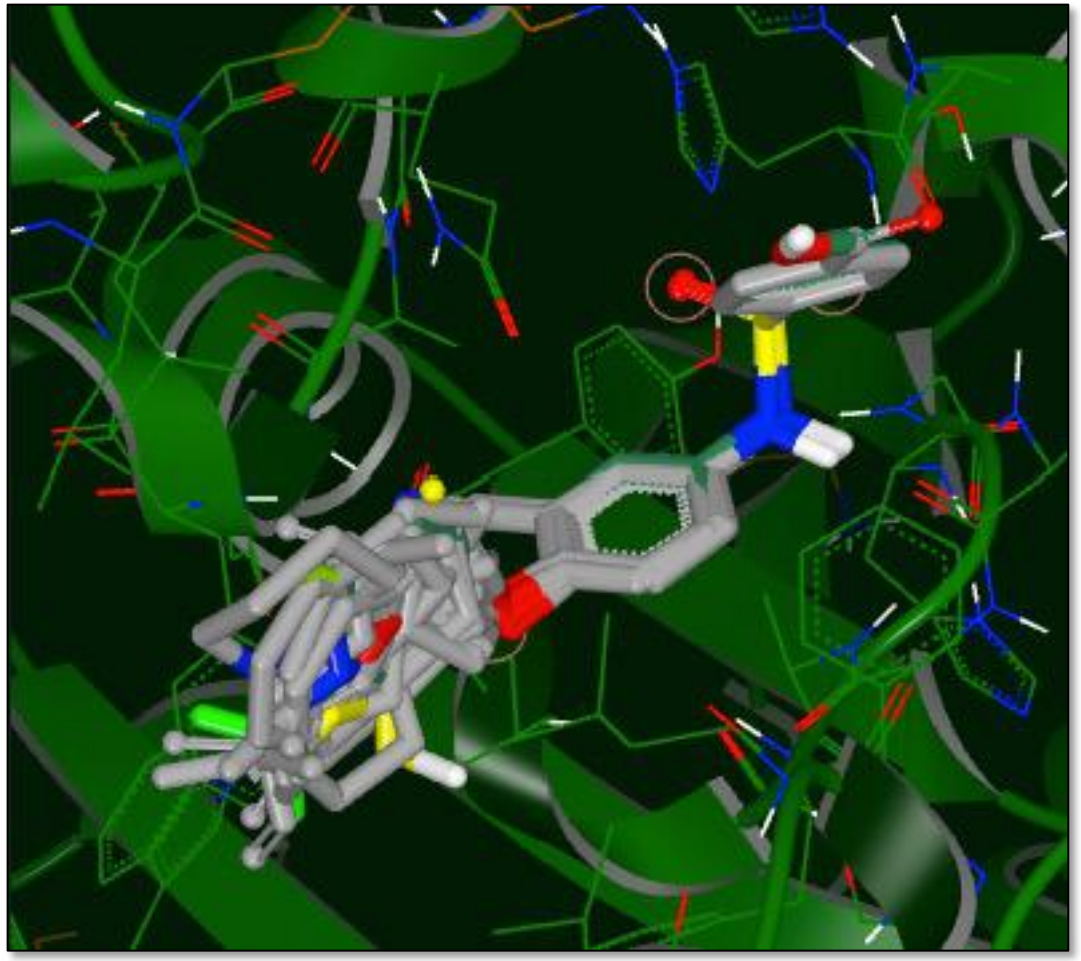
Custom Molecule Filtering (i.e. bad chemistry)

Reagent Classification

OpenEye SCIENTIFIC

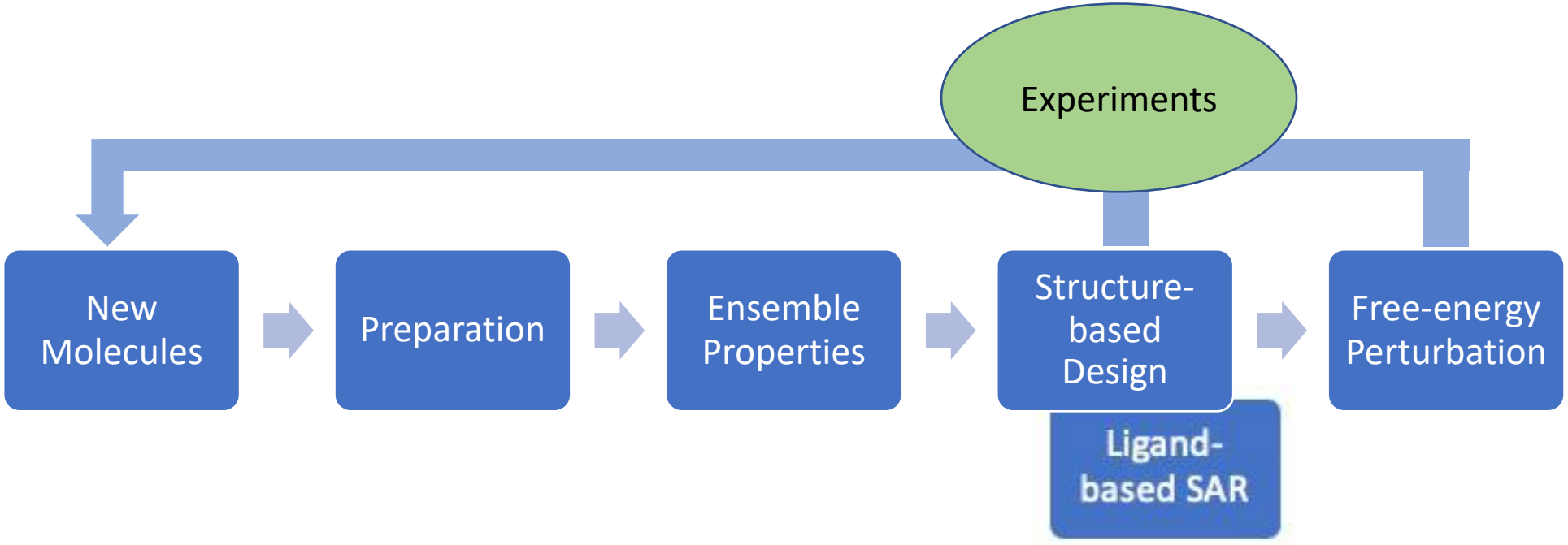
Customer lead → Small library



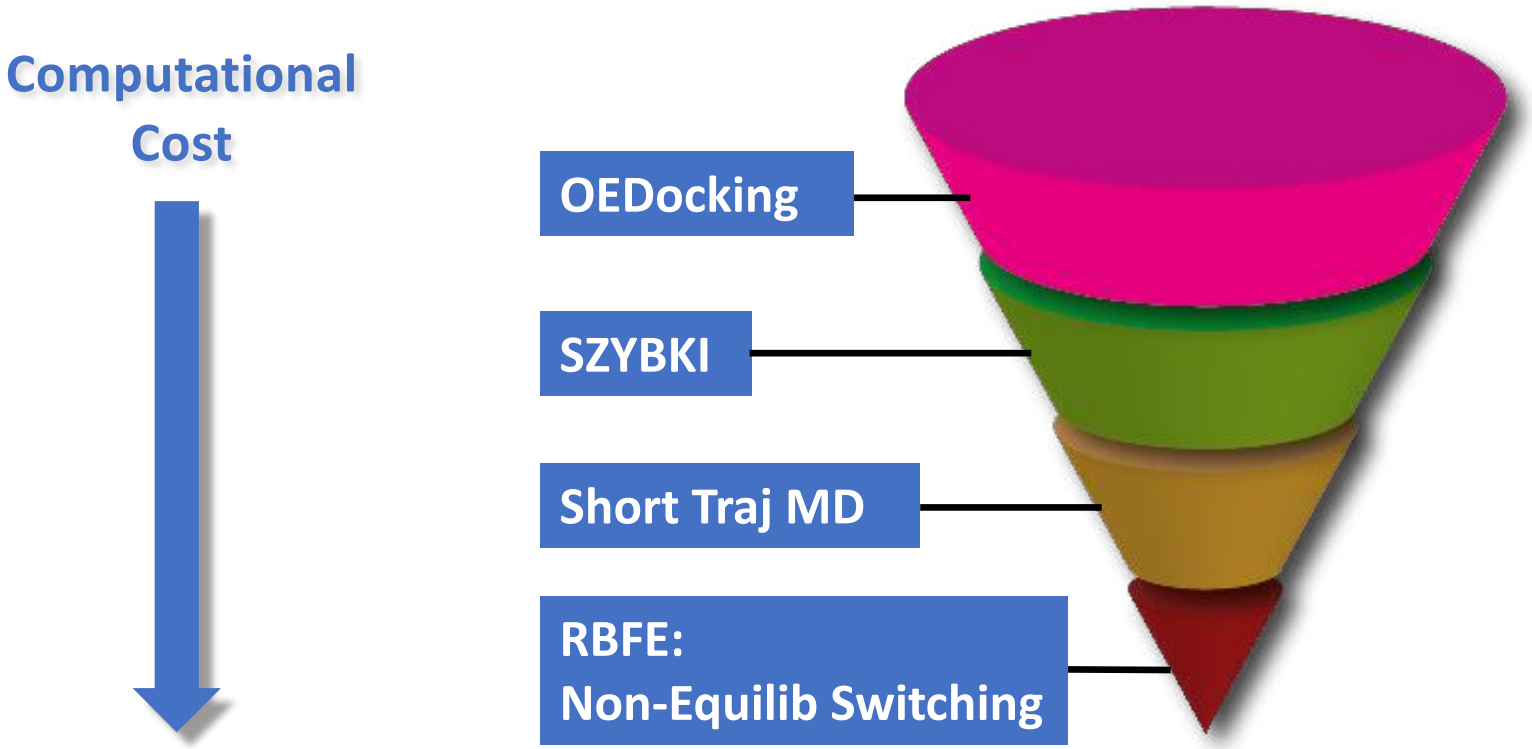


OpenEye SCIENTIFIC

Simplified Molecular Design



MD in Structure-Based Lead Optimization in ORION

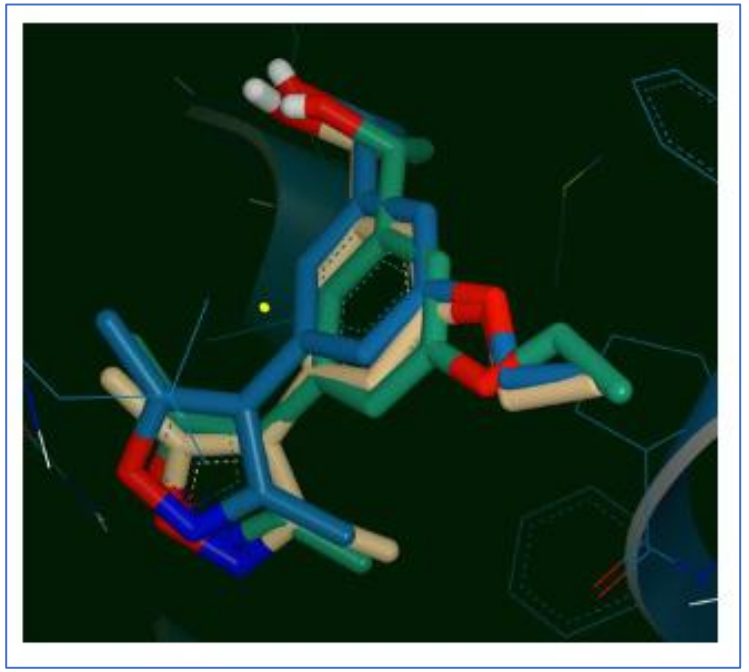


- MD methods can be integrated into traditional workflows in Orion

Amber14sb + Parsley in POSIT

- POSIT
  - Best-in-class Pose-prediction
  - Protein-ligand clashes
- Relaxation
  - Good prospective results
  - Force-field implementation was not robust
- Robust
  - Amber14sb + Parsley
- Large-scale re-validation
  - Spruce cross-docking dataset
  - Tuccinardi method

Why Short Trajectory MD?



- What Christopher always told people:
  - Sample protein, ligand and explicit water
  - 2ns trajectory (short)

  - Examine the ligand trajectory → predict quality
  - Better than docking & cheaper than Free-Energy

- What Ant responded:
  - “Implement that & prove it!”

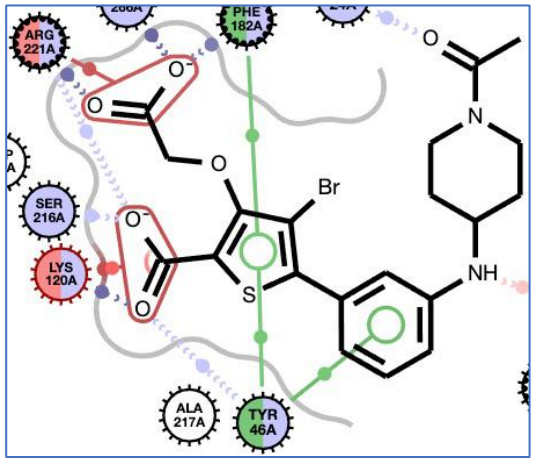
Combining “Design Quality” and “Pose Stability”

- Identify each good OEInteraction, ***i*** in the Initial Pose:

- For each Frame of the trajectory:
  - For each OEInteraction ***i*** from the Initial Pose:
    - If the same OEInteraction is present in that Frame, increment ***i***Count for that OEInteraction

- TrajScore_i = ***i***Count/***n***Frames

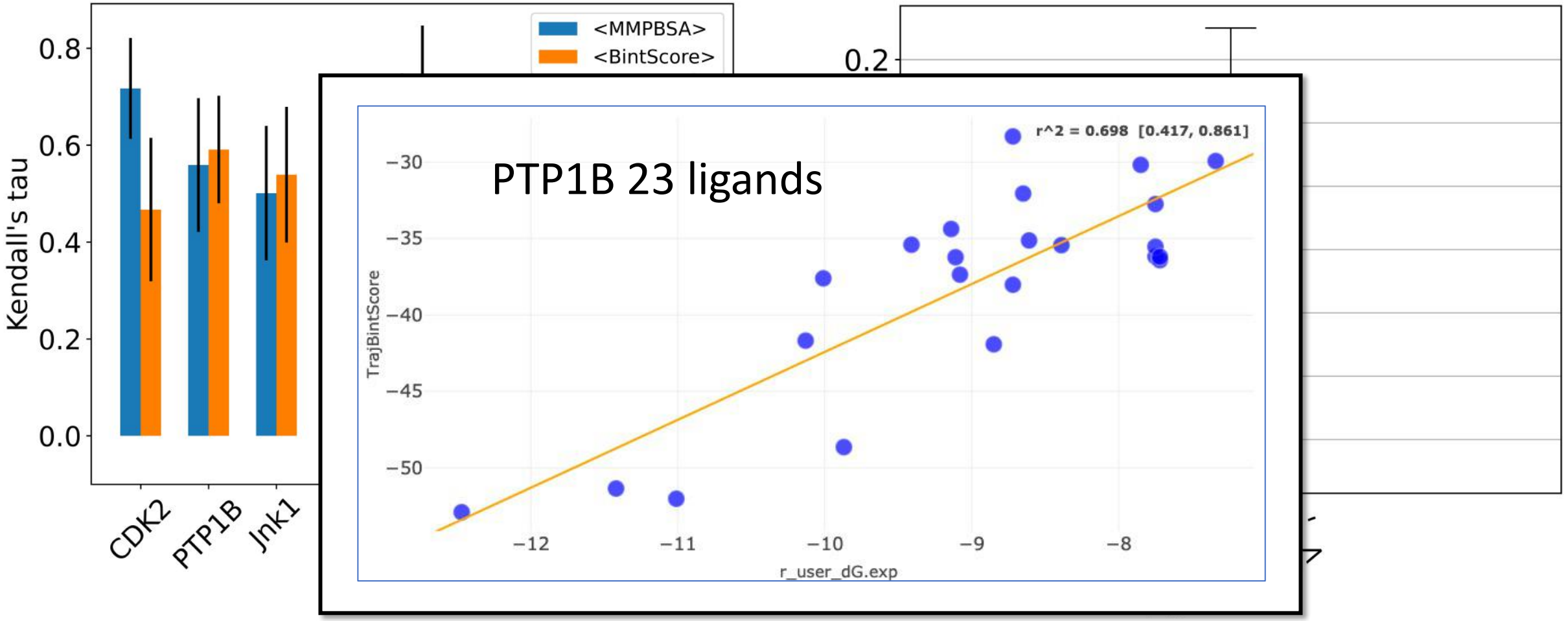
- **BintScore** = $\text{sum}_i(\text{TrajScore}_i)$



- More interactions are better
- More stability is better

OpenEye SCIENTIFIC

MMPBSA vs Bint Score



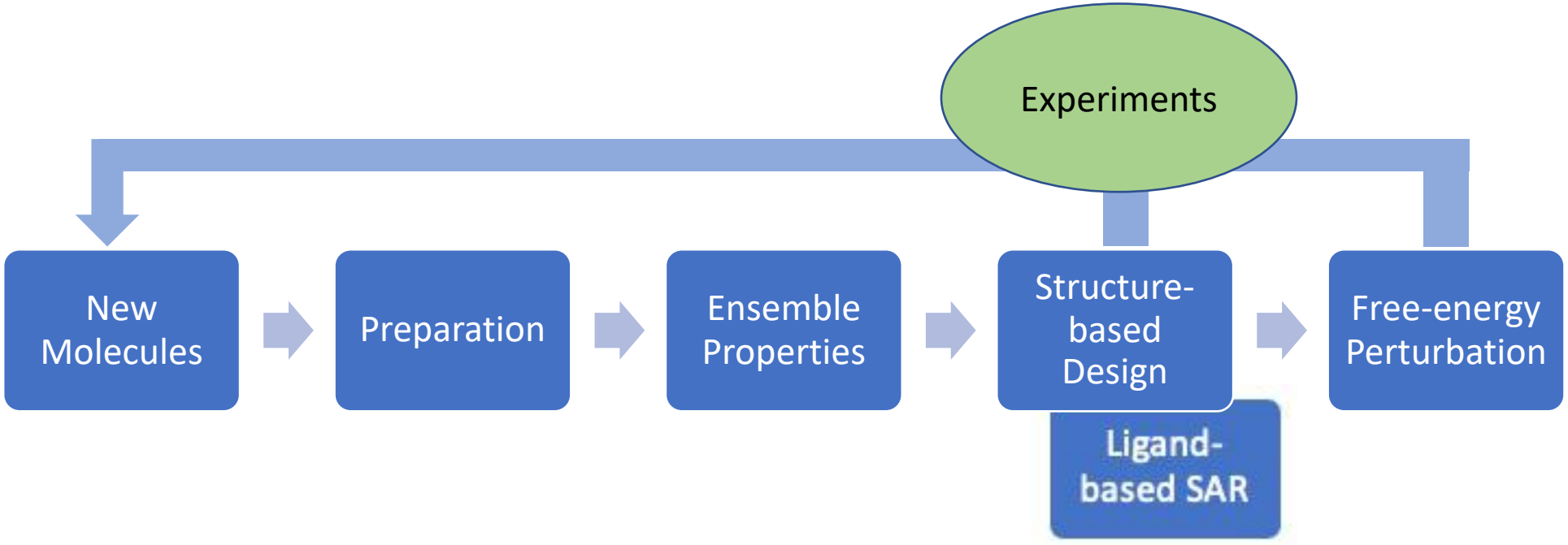
Initial Bint Score Insights

- Bint Score >? MMPBSA

- Simple scoring based on OEInteractions
  - Only **good** OEInteractions
  - Sum initial OEInteractions to estimate affinity
  - Stability of interactions under MD improves the estimate

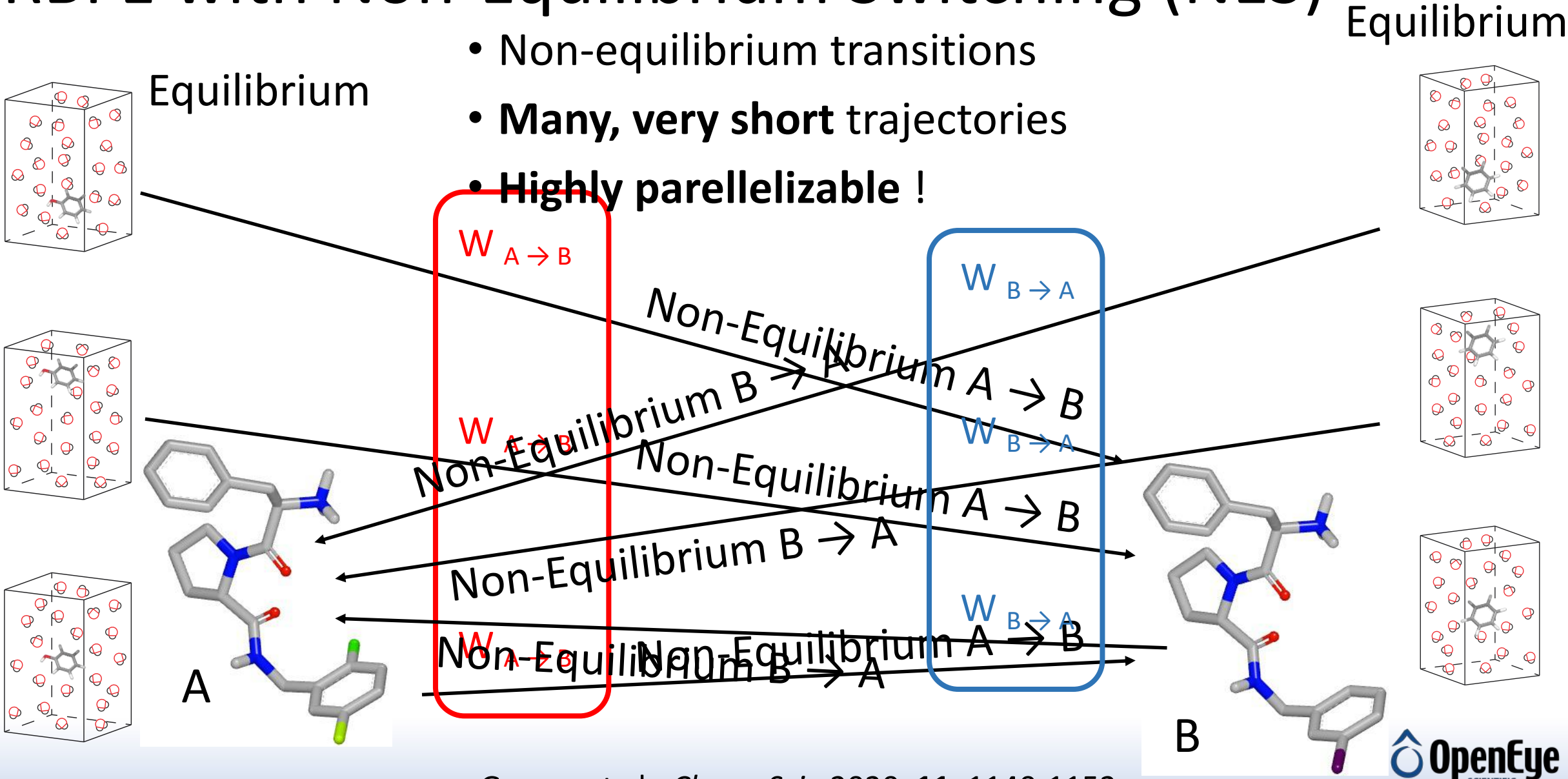
- Learn more from Christopher Bayly's presentation

Simplified Molecular Design



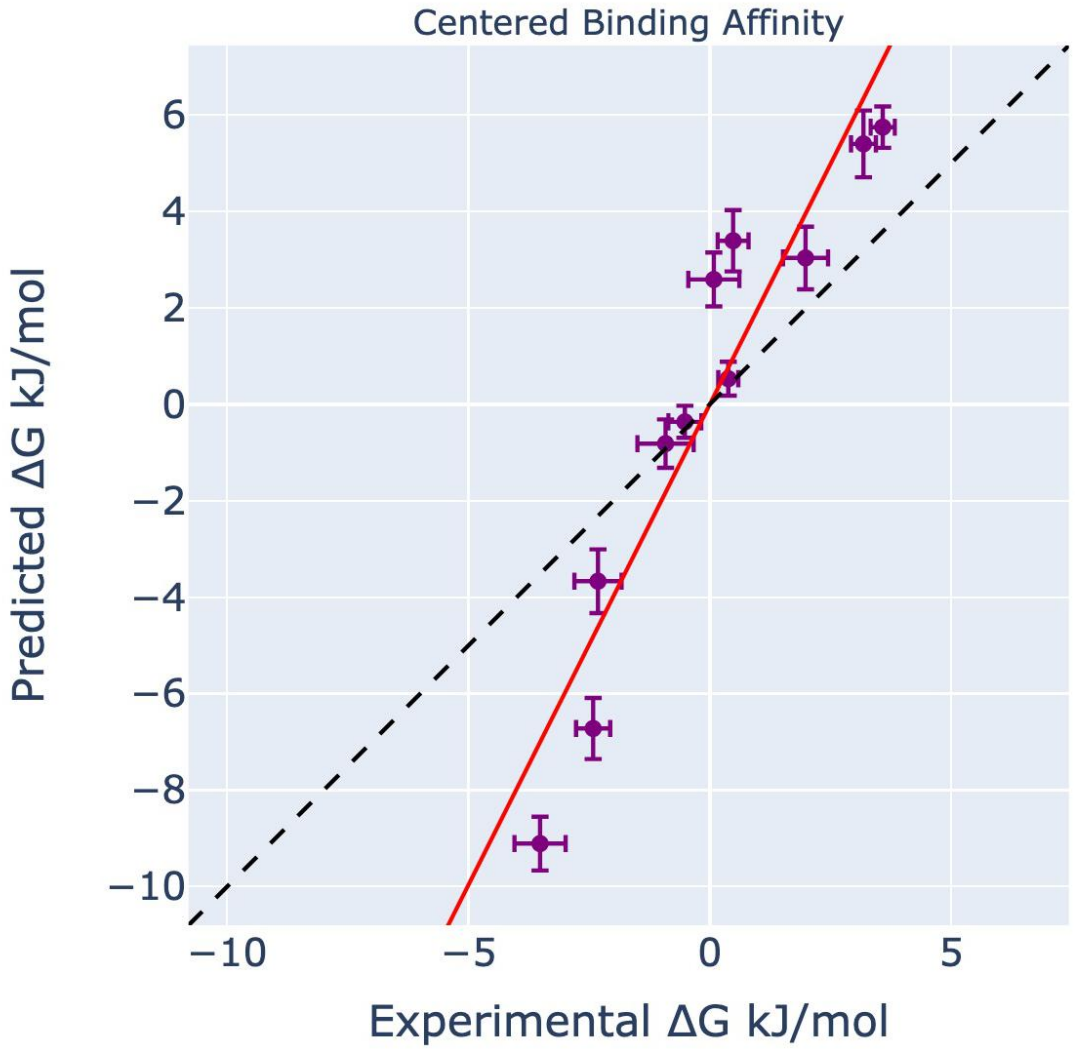
RBFE with Non-Equilibrium Switching (NES)

- Non-equilibrium transitions
- **Many, very short** trajectories
- **Highly parellelizable** !



Gapsys et al., *Chem. Sci.*, 2020, **11**, 1140-1152

RBFE-NES for JACS'15 Thrombin dataset



Method	ΔG kJ/mol
MAE	2.04 [1.08 3.15]
Relative MAE	1.16 [0.72 1.98]
RMSE	2.65 [1.52 3.68]
Relative RMSE	1.22 [0.86 2.03]
Pearson's R^2	0.89 [0.78 0.97]
Spearman's ρ	0.94 [0.89 0.99]
Kendall's τ	0.93 [0.75 1.00]

Hannah Bruce Macdonald, David Hanh https://github.com/choderalab/freeenergyframework

OpenEye SCIENTIFIC

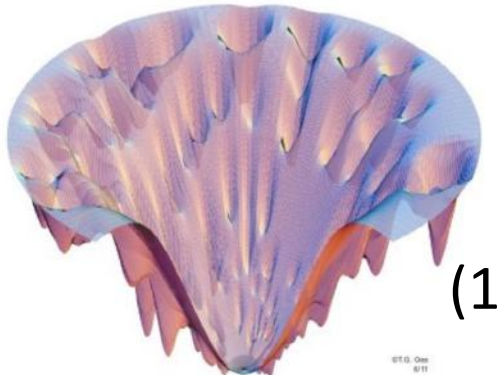
MD Affinity in 2021

- Validation & Testing
- Orion 2021.1
  - Bint floe
  - NES v1.0
    - Ligand-Ligand edges as an input parameter
- Orion 2021.2
  - Lomap for NES
    - Set of ligands as input
  - Preparation exceptions
- Beyond
  - NES science problems (trapped waters & residue conformations)
  - Integrate Mapper choices with Chimeric choices & NES needs

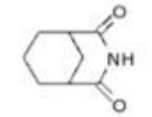
Drug Discovery Realms

Candidate selection & Formulation

Typical physics-based modeling workflow of CSP



(100-1000)



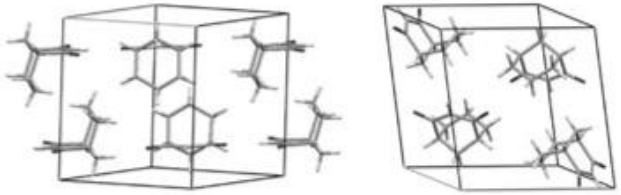
Molecular diagram

Start with 2D structure

Three-dimensional molecular structure

Conformer generation

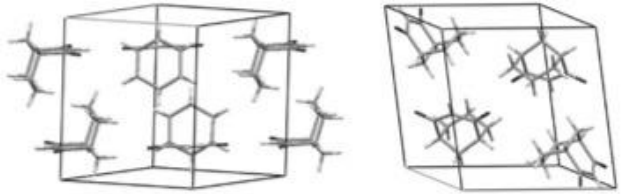
(10^6-10^7)



Many possible close-packed, low-energy crystal structures

Rigid packing -> FF minimization -> FF ranking

(100-1000)

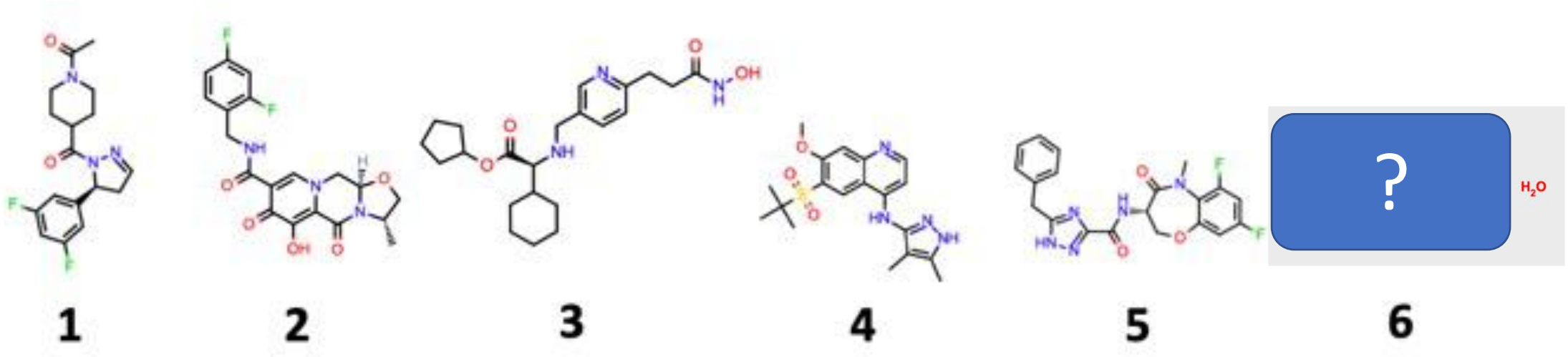


Many possible close-packed, low-energy crystal structures

Re-rank top FF packings using QM

OpenEye SCIENTIFIC

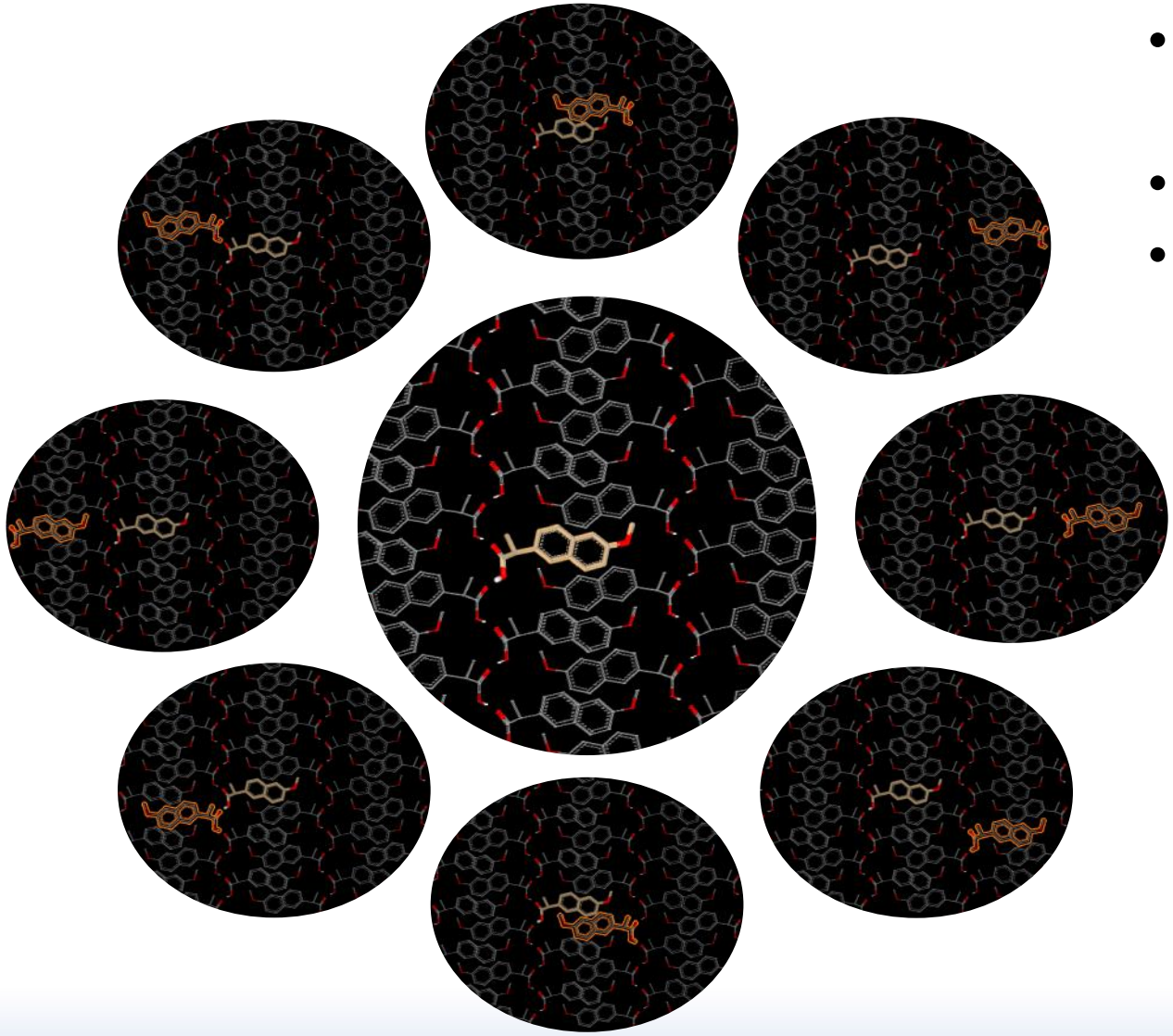
GSK Blind Challenge



Challenge	Dispersion	H-bonds	Flexibility (BSSE)	Space-groups, Tautomers & S	Ring sampling Finite-temp.	Hydrate
Rank	1	1	-	1	1	1
RMS_20	0.18	0.18	No conformer < 1A	0.16	0.23	RMS40=0.47
New	Constrained optimization	Hydroxyl sampling	Iterative conf. packing	Tautomer & torsion sampling	Loose QM opt. & Finite temperature	Water sampling

OpenEye SCIENTIFIC

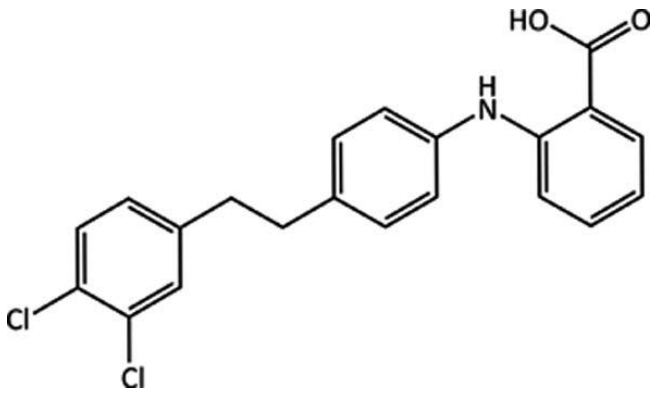
Parallelize Crystal QM

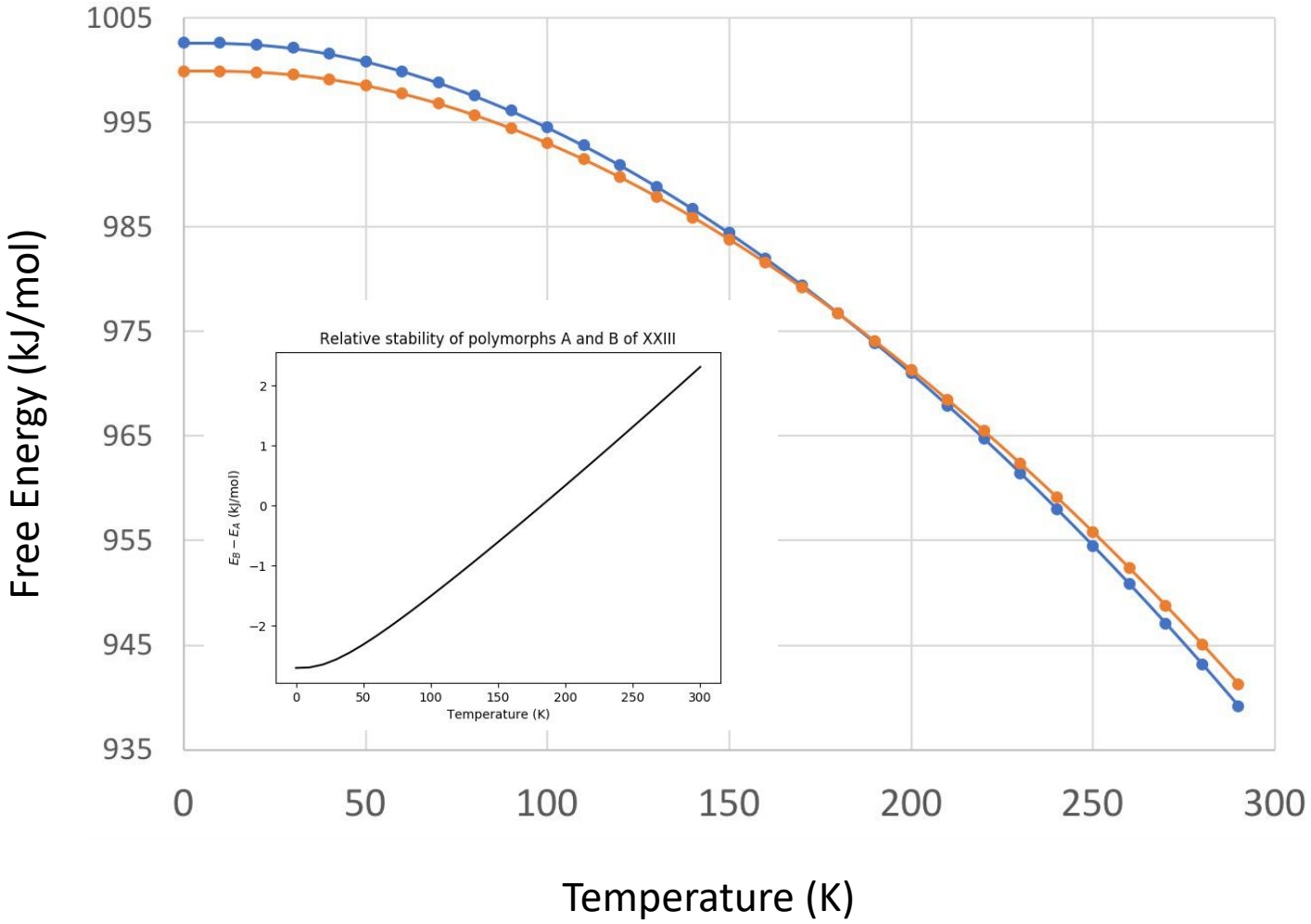


- Crystal expansion as a collection of gas-phase dimers
- Select different levels of all-electron QM
- Treat short/long range differently

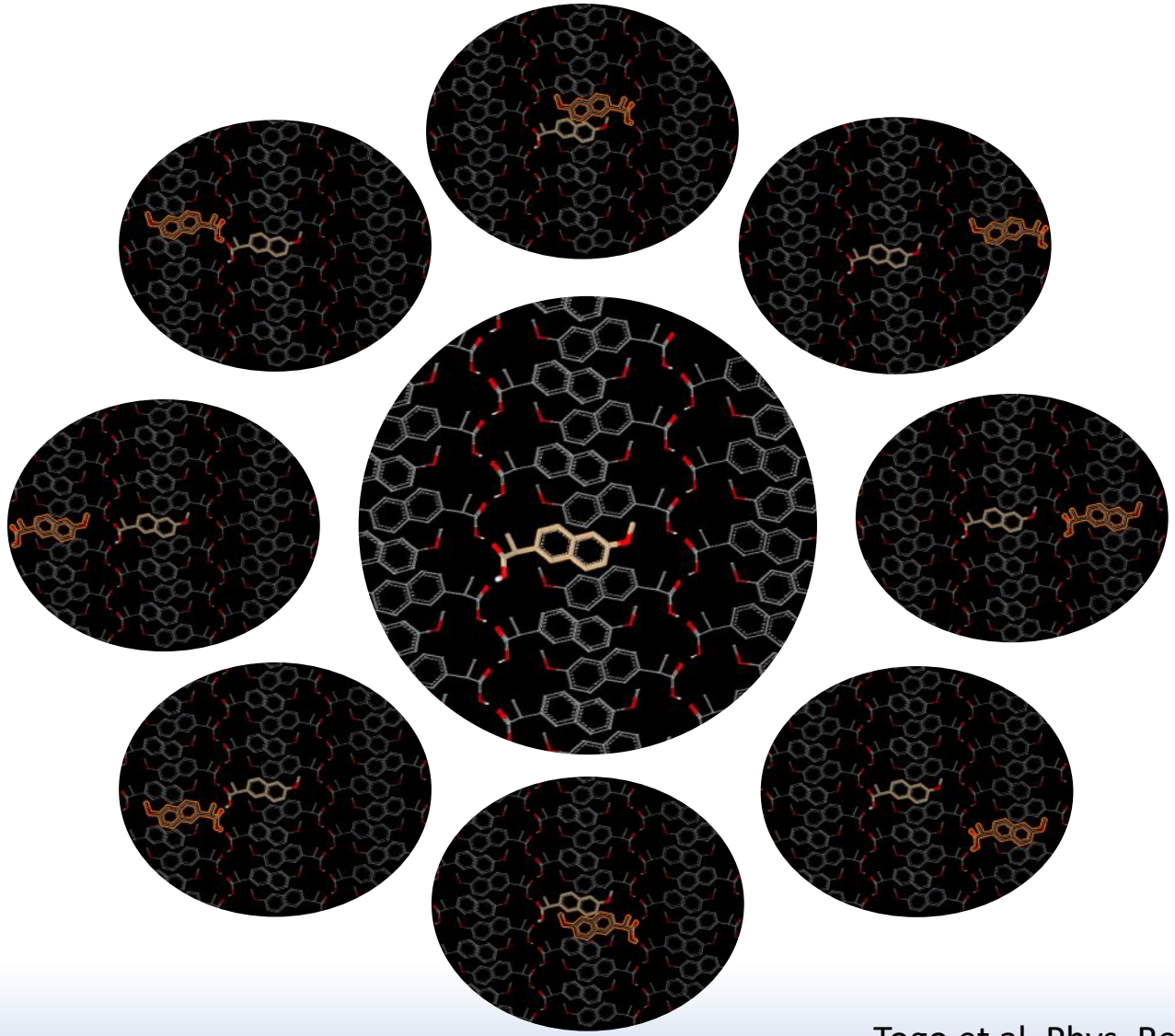
Energy Model	MA%D	MAD (kcal)	ME (kcal)	Rel time (Approx)
TPSS-D3	5.2	0.9	0.2	100,000
PBE-D3	5.75	1.1	0.4	100,000
PBEh-3C	6.3	1.3	0.1	100,000
IEFF+PBE-D3@SR	5.7	1.1	0.5	5,000
DFTB-D3	11.9	2.4	0.1	50
FIT (S. Price)	10.3	2.2	-1.9	1
IEFF (OE)	11.5	2.5	-0.8	1
W99rev (G.Day)	15.7	3.4	-3.3	1

Temperature-dependent relative stability of polymorphs





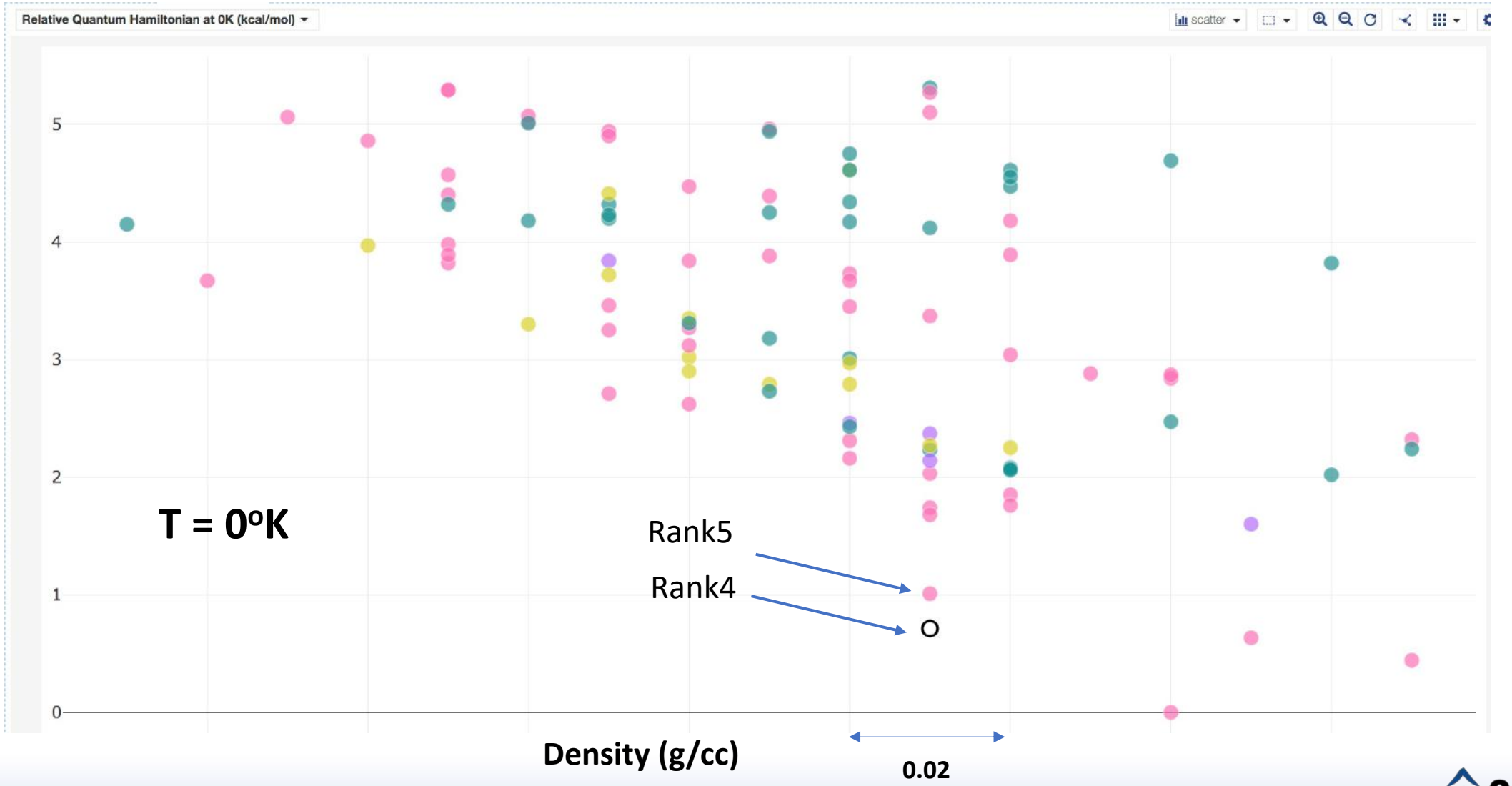
Parallel Hessian and entropy calculation



- Finite-difference Hessian using cluster of dimers calculated in parallel
- Phonon spectrum and entropy calculated using harmonic approximation

Togo et al, Phys. Rev B, 91(9):94306, 2015

OE-GSK 5th Blind Prediction Challenge



OpenEye SCIENTIFIC

From CSP to Formulations

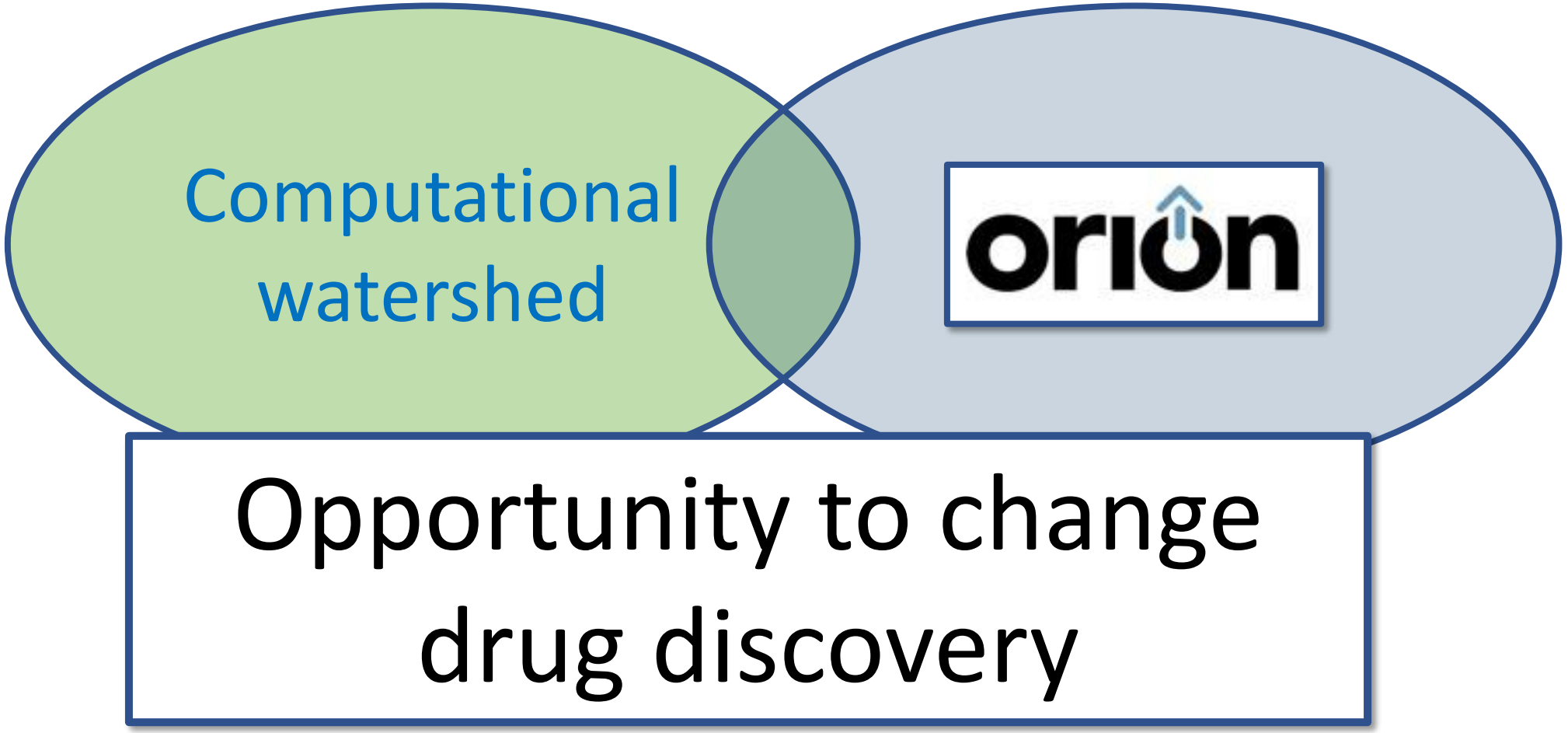
Building Experience

- Varied examples
  - CSP Challenge
  - More Z'=2
    - Salts
    - Hydrates
    - Cocrystals
- Paper
- Collaborative CSP
- CSP as a service

Solutions

- Structure -> polymorph risk
- Salt selection
- Formulation Properties
  - ***Solubility***
  - ***Elasticity, Compressibility***
  - ***Habit (Morphology)***
  - Hygroscopicity
- Solvent selection
- Multiple stoichiometries

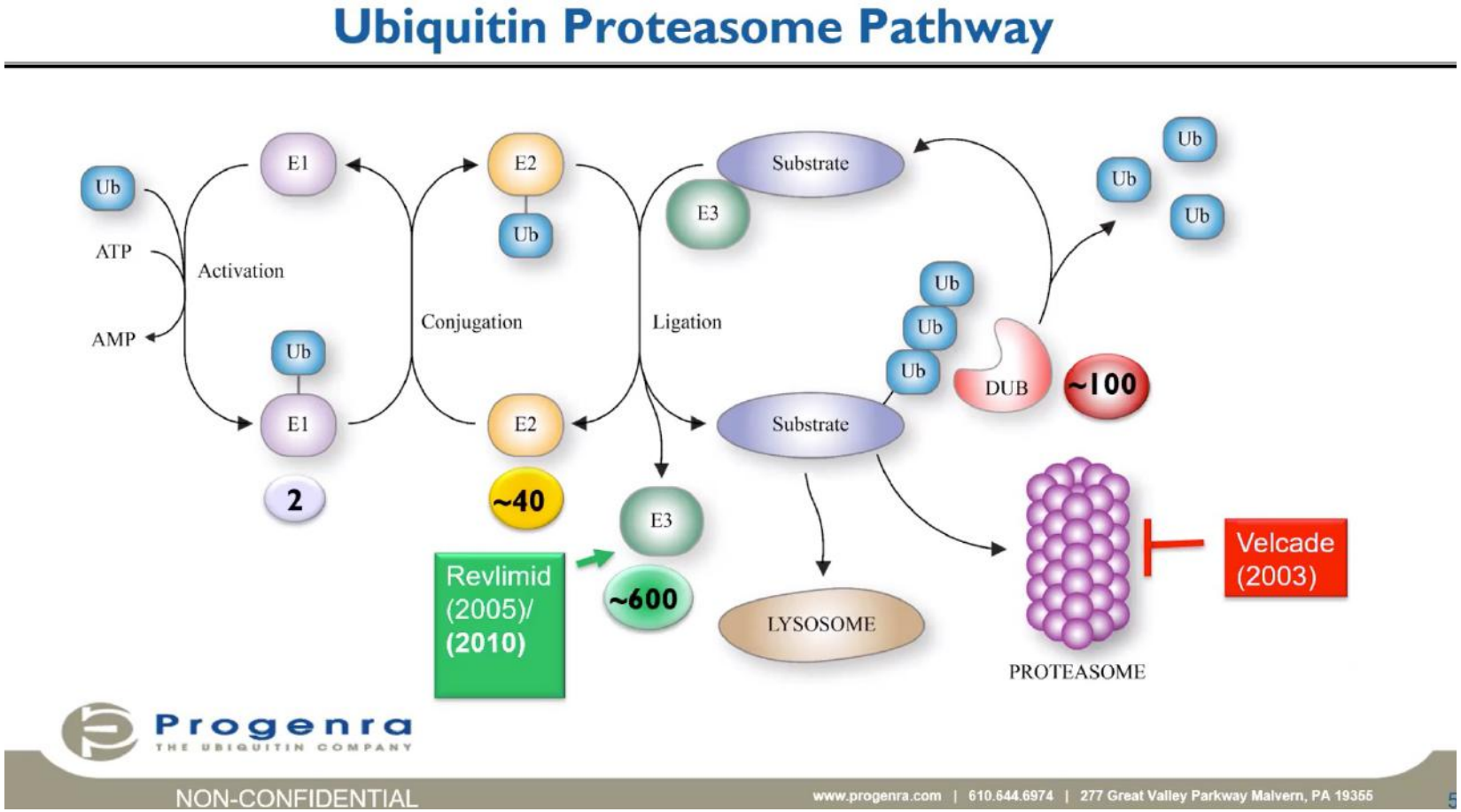
Modern impact



Acknowledgement

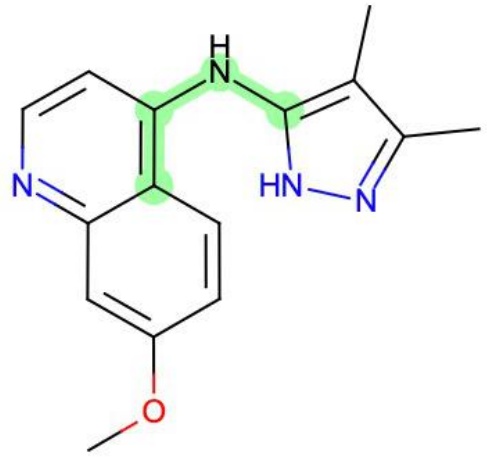
- Burt, Sayan, Nathan
- Perri, Dan
- Shyamal, Kristi, Jun, Stan
- Christopher, Gaetano
- Jesper, David, She, Jeff, Mark
- Hari, Grigory, Caitlin & Tom
- Matt, Phillip, Allison
- Steve, Samaneh, Gunther
- Cloud team
- GSK
  - Colin Edge, Alan Graves, Eric Manas
- NES
  - David Mobley, David Hahn
  - John Chodera
  - Vytautas Gapsys, Bert de Groot
- Enhanced Sampling
  - Lillian Chong

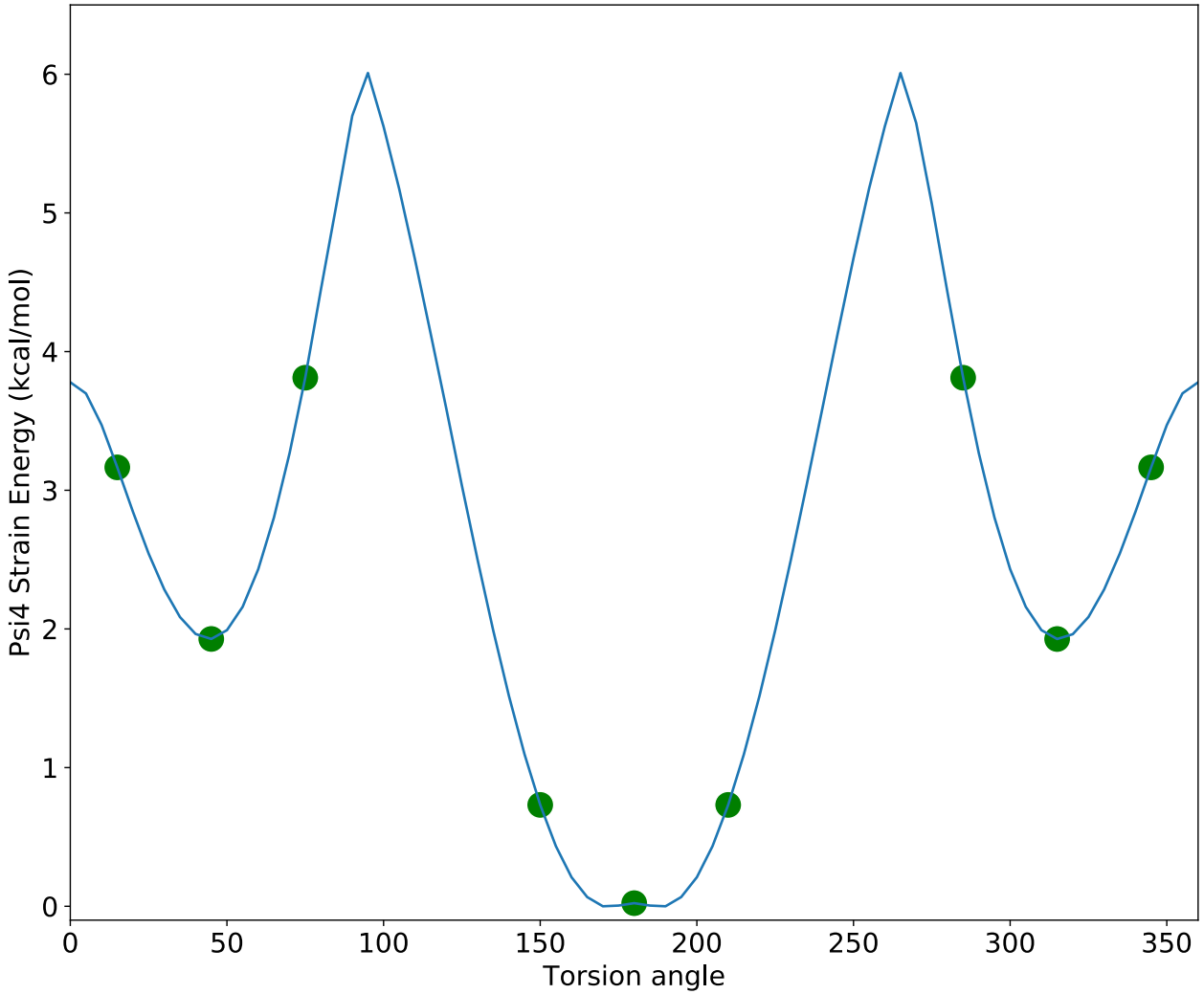
Protein-Protein Interactions: PROTACs



2021: Omega QM

- Identify rotatable bonds
- Fragment around that molecule maintaining environment
- Scan torsion in fragment
- Apply new torsion rules in conformer generation



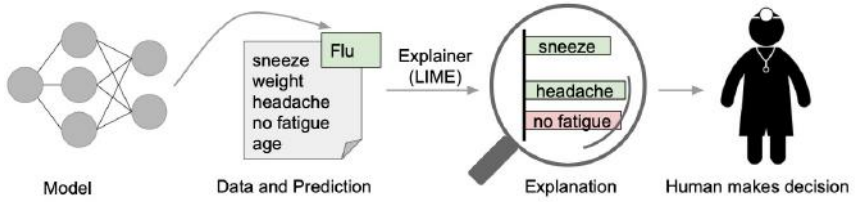


OpenEye SCIENTIFIC

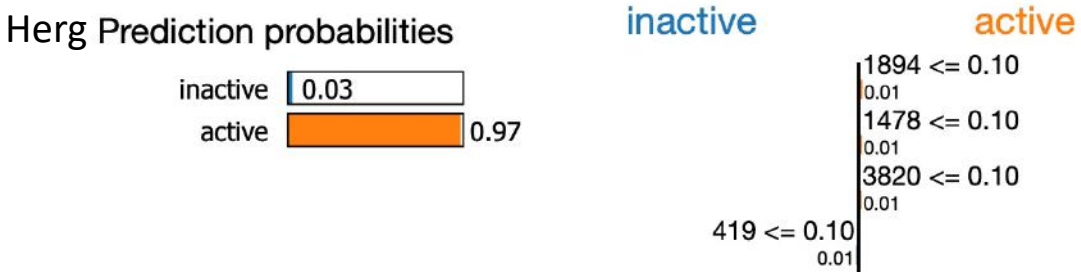
Explainable AI

Given a data set, we want classification/regression models which:

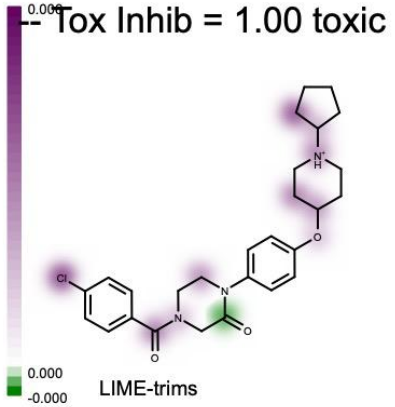
- Interpret results at atomic/ligand level
- -agnostic Explanations



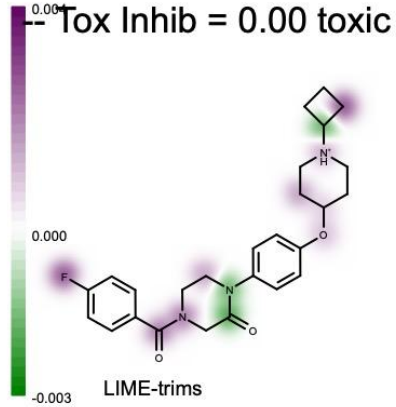
- Explain Neural Network trained on 2d molecule figure-print:
  - find out important bits



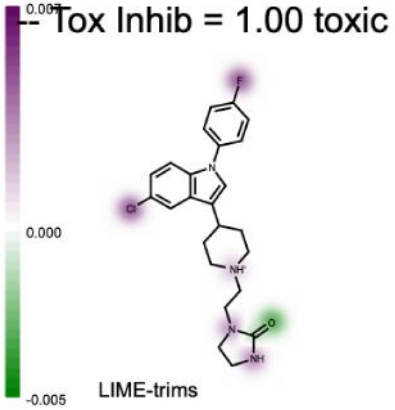
Fragments making molecule Herg Active
Fragments making molecule Herg inactive

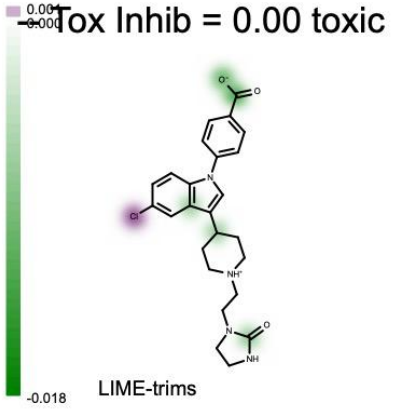


Herg Active.



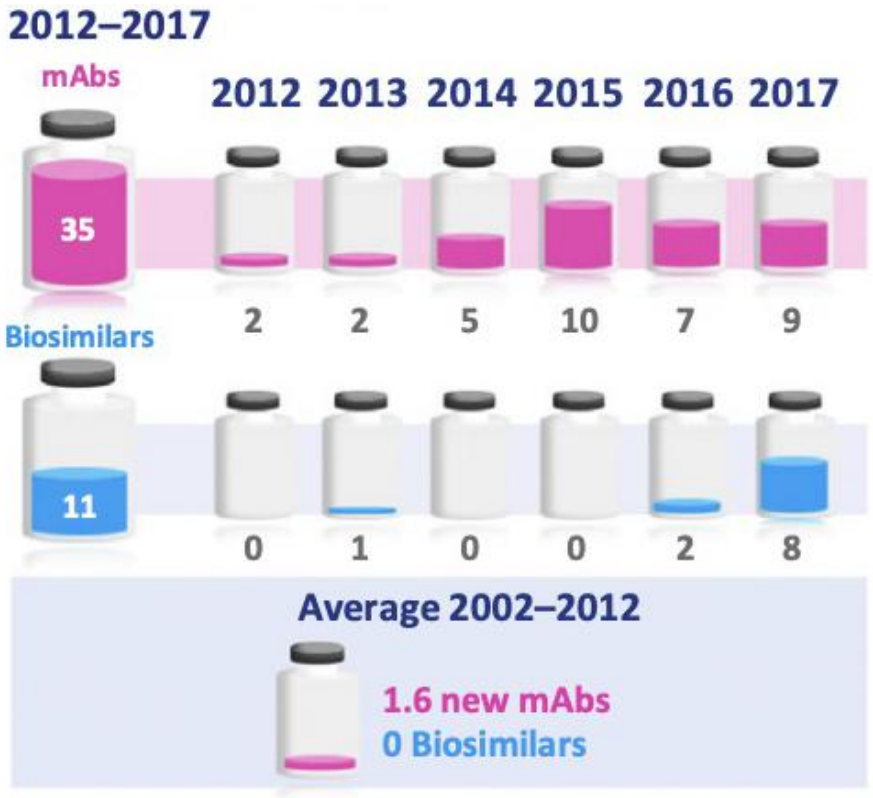
Herg Inactive

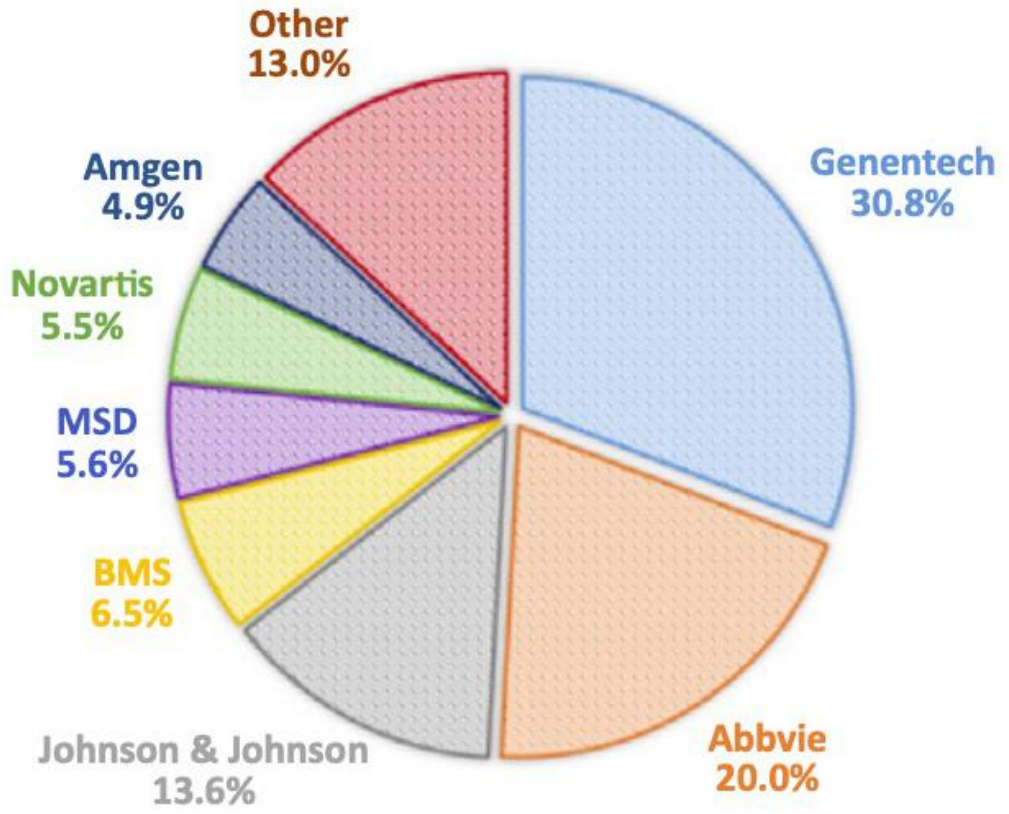




OpenEye SCIENTIFIC

Monoclonal Antibodies as Drugs



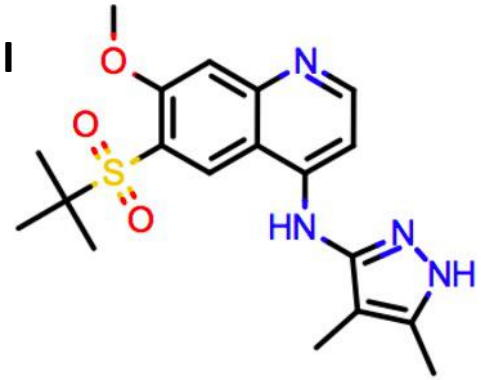


6 **of the top 10 best selling drugs in the US are monoclonal antibodies**

Grilo and Mantalaris. 2019. *Trends in Biotechnology* 37 (1) 9–16

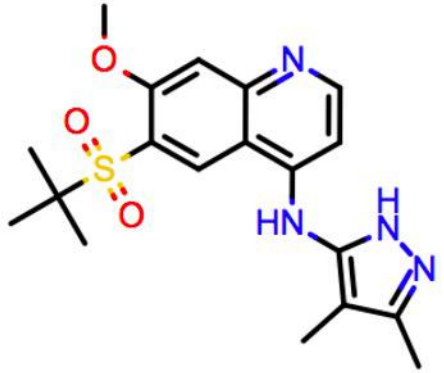
2020: Relative Tautomer Energies

I



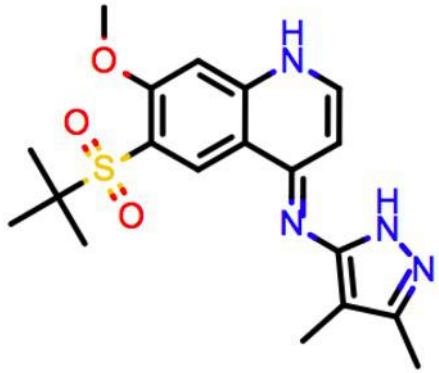
0.0 kcal/mol

II



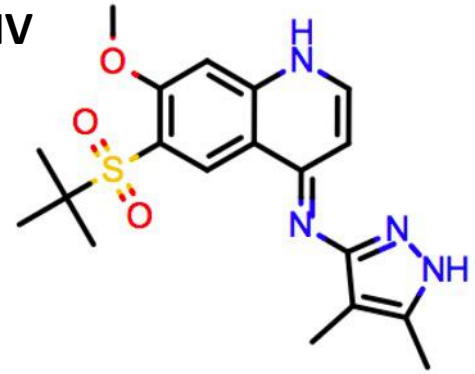
1.54 kcal/mol

III



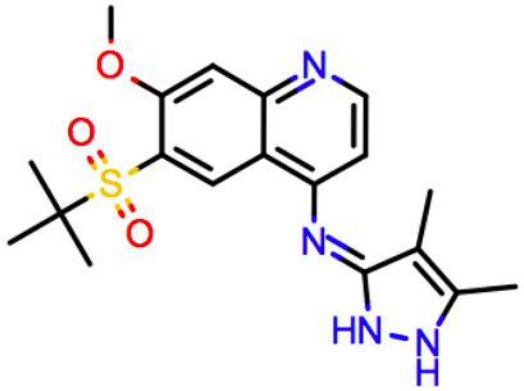
E: 9.04 kcal/mol
Z: 10.54 kcal/mol

IV



E: 9.05 kcal/mol
Z: 12.56 kcal/mol

V



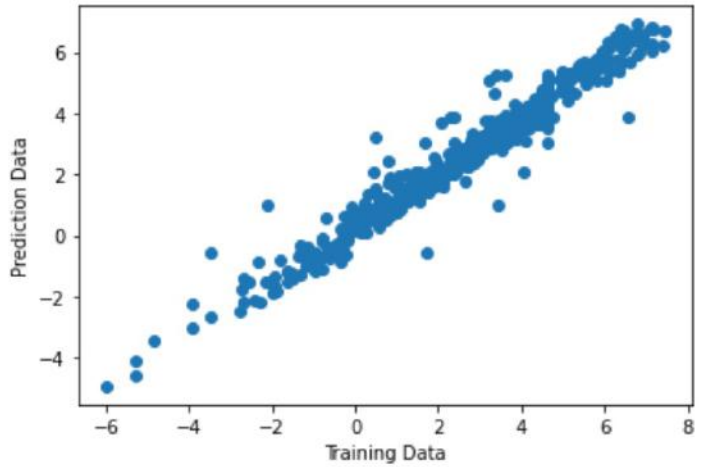
E: 14.41 kcal/mol
Z: 14.34 kcal/mol

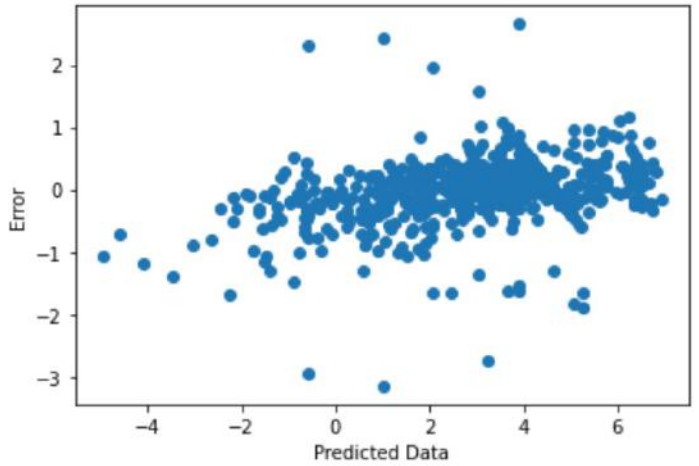
OpenEye SCIENTIFIC

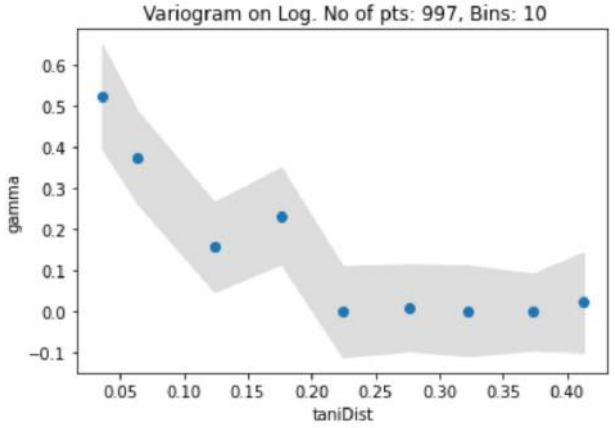
Gaussian Process Regression

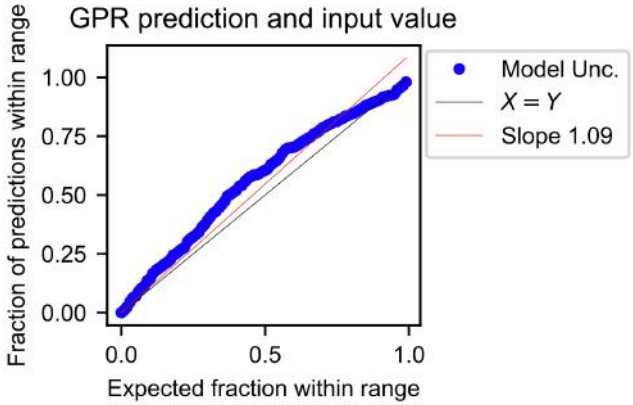
- Nonparametric, Bayesian approach to regression
- Input Data: 2D fingerprint of Molecules
- Kernel for regression: 2d/3d Tanimoto Distance
- Regression prediction: Toxicity, logp, ic50 etc.

Regression on Manchester DB









Variogram and QQPlot on Data

*QQPlots credit: Caitlyn Banner

OpenEye SCIENTIFIC

A quick intro to passive permeability

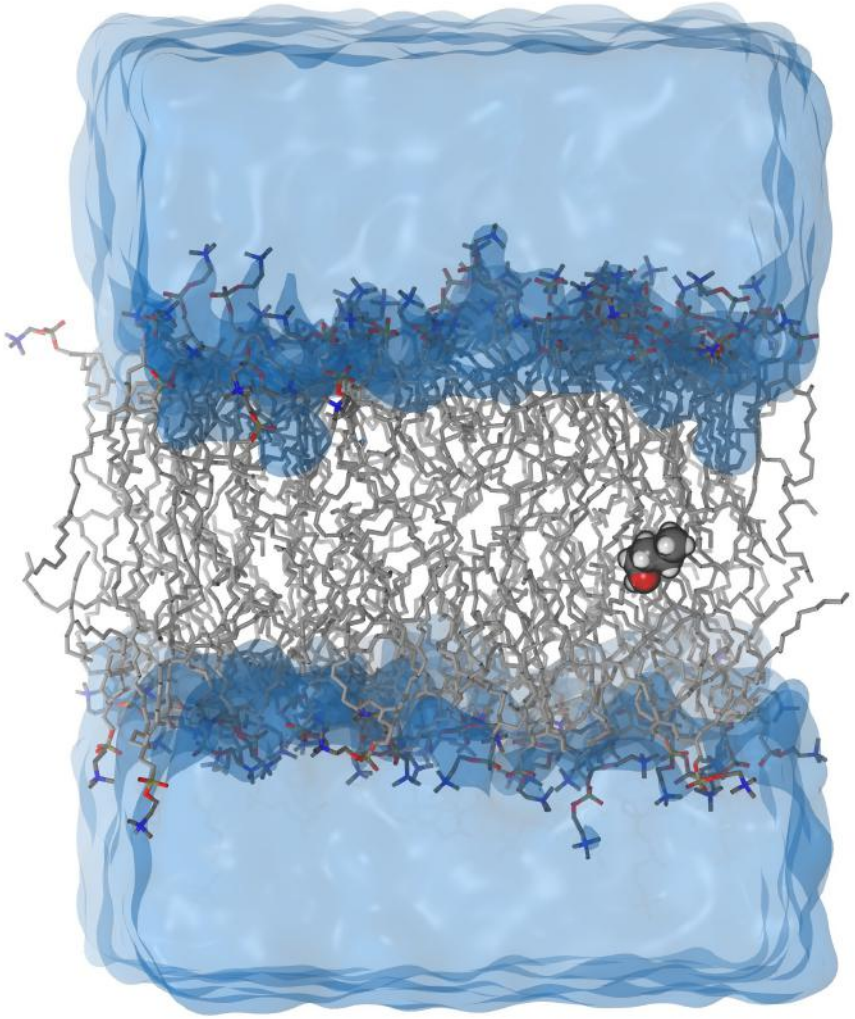
- Important for ADME/tox
- Permeability is defined as:

$$P = \frac{J}{\Delta u}$$

J = membrane flux

Δu = conc. of drug

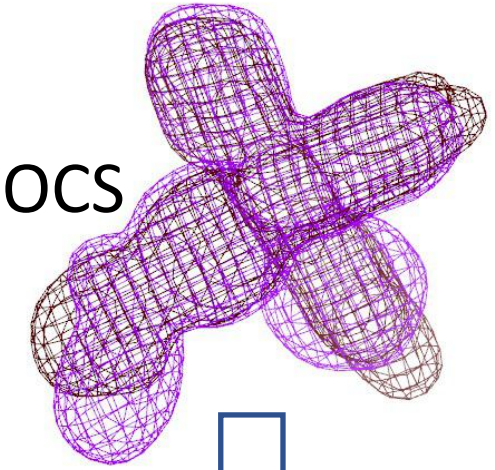
- P is reported in units of cm/s
- Difficult to estimate computationally
- Expensive and time consuming to observe experimentally

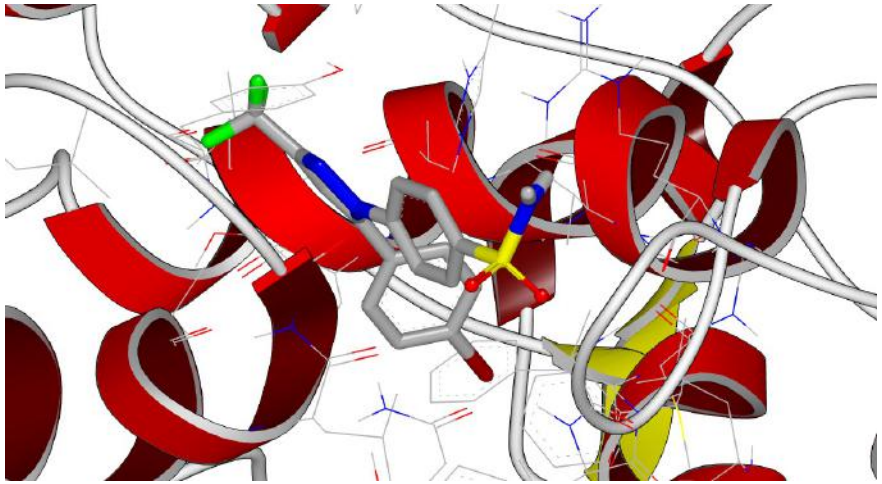


OpenEye SCIENTIFIC

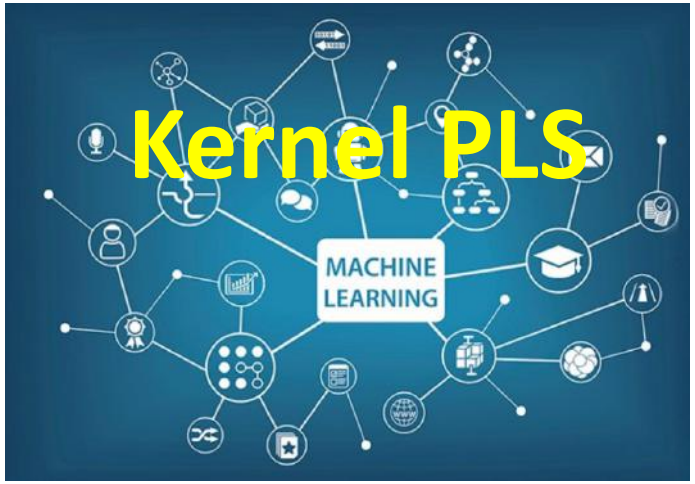
ROCSAR

ROCS





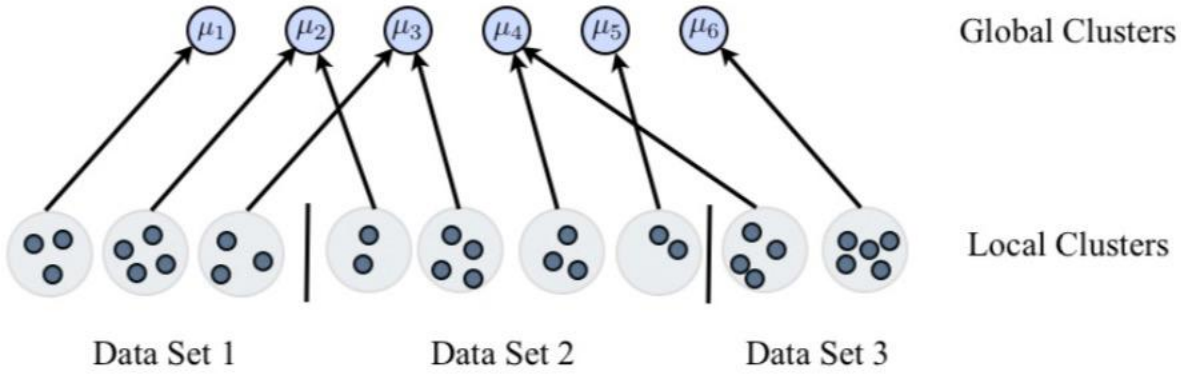
Experimental/POSIT

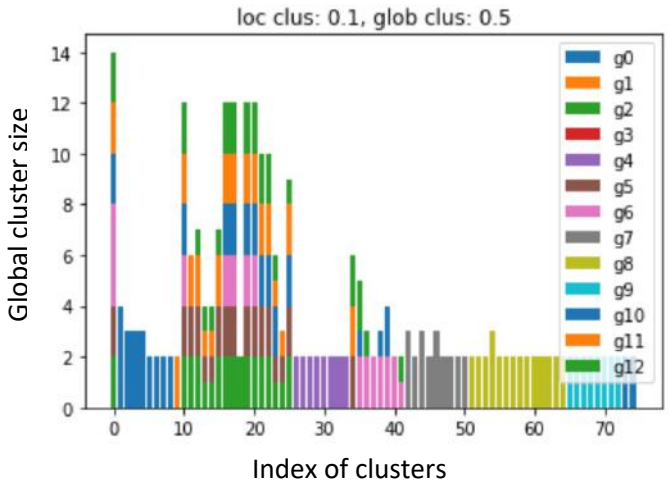


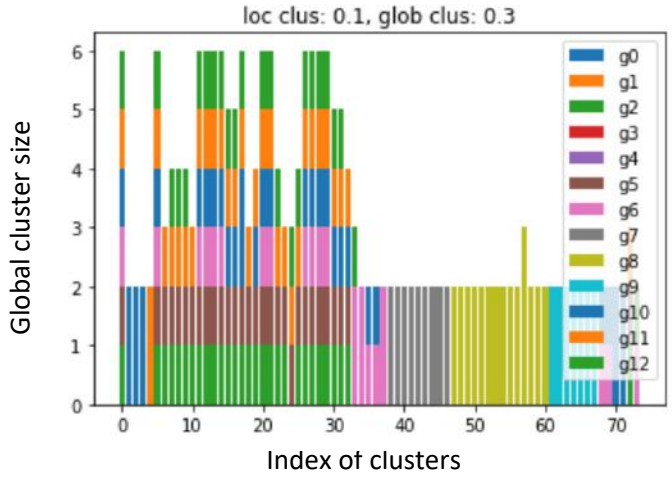
ROCSAR Model

Hierarchical Clustering based on Dirichlet Process

- Input Data: Multiple conformers of separate molecules
- Hierarchical Clustering: Structurally similar molecules for further analysis
- Local and global parameter



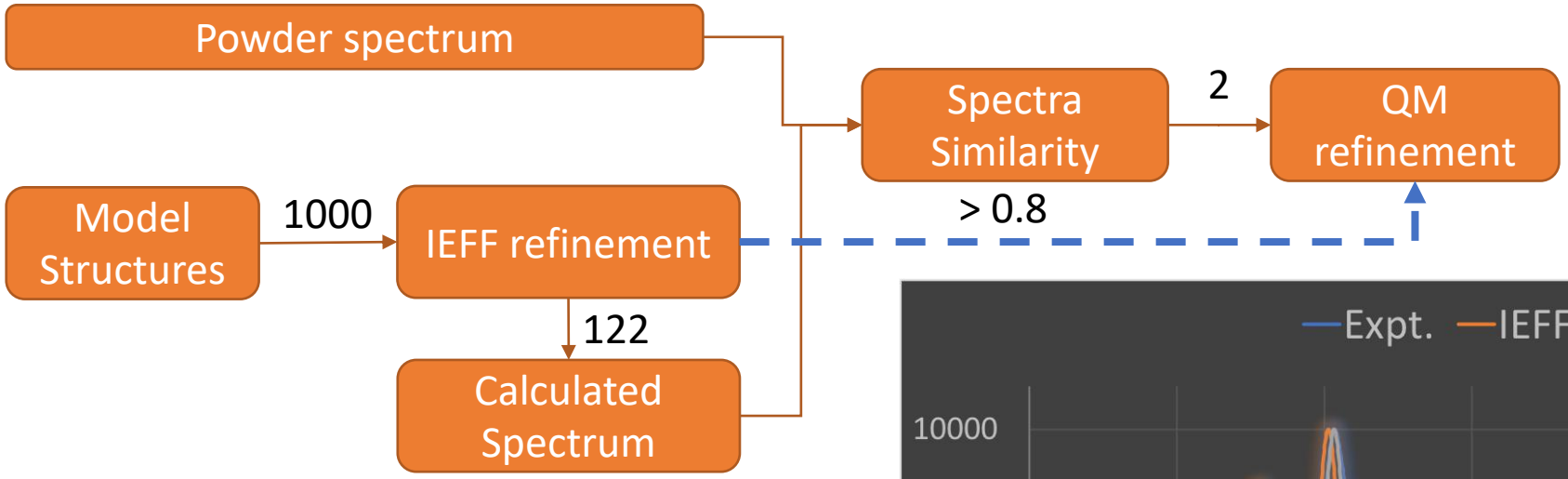


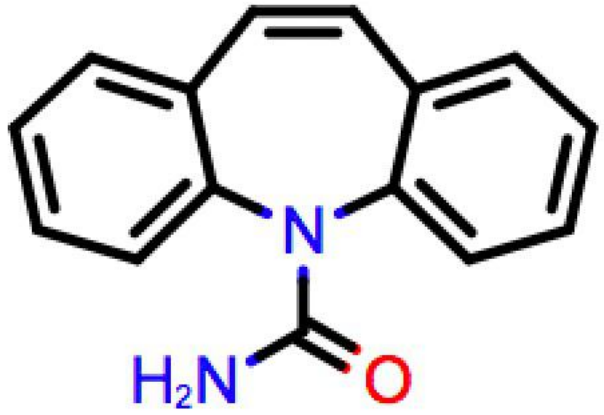


g0,g1…: conformers belonging to same molecule

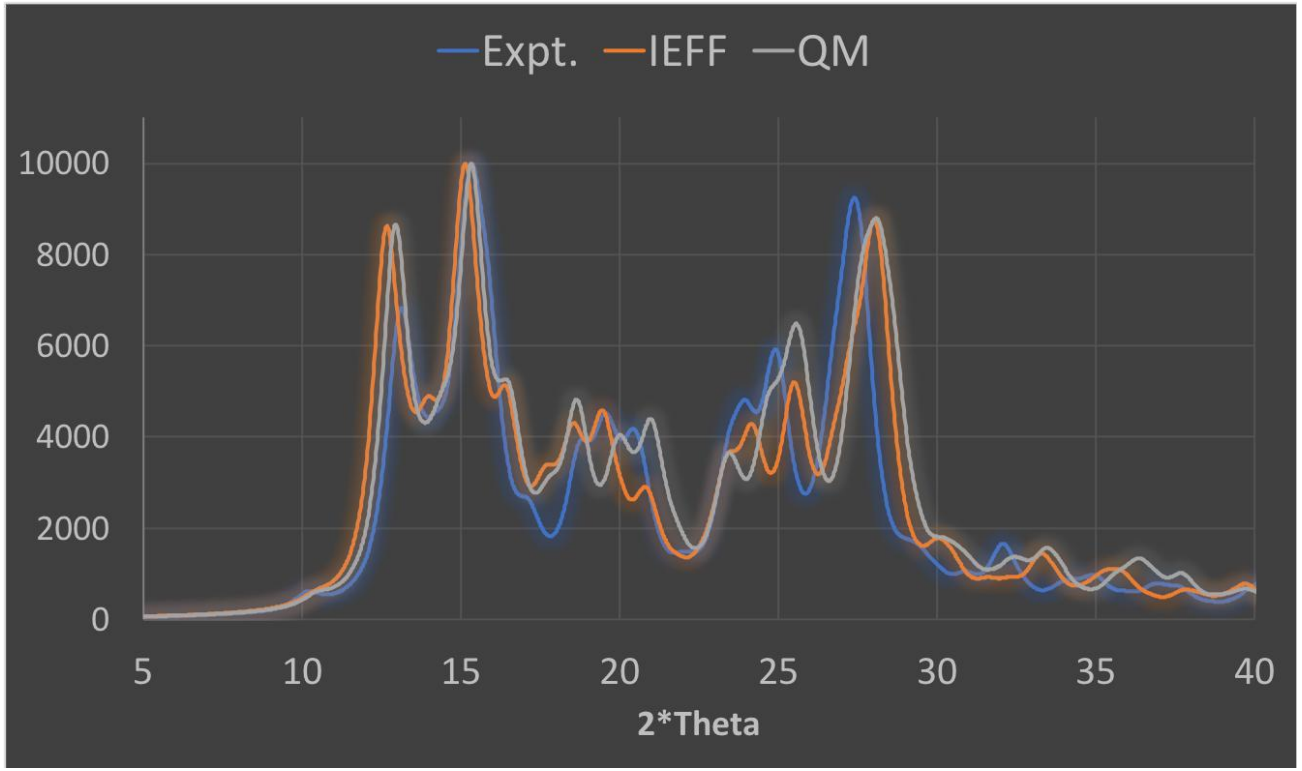
OpenEye SCIENTIFIC

CSP using PD spectrum: Carbamazepine





Carbamazepine



OpenEye SCIENTIFIC