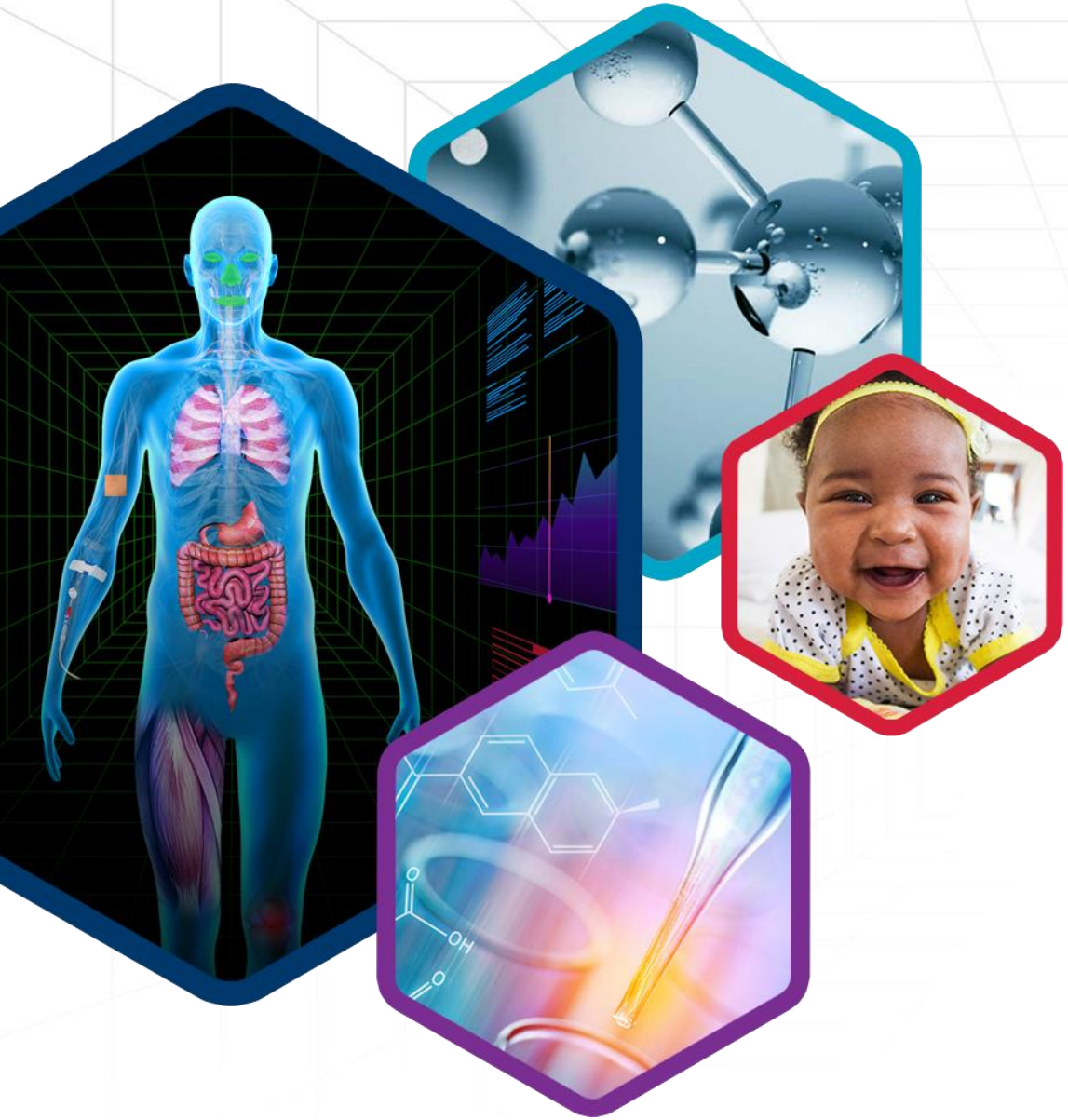


SimulationsPlus

Cognigen | DILIsym Services | Lixoft

AI-driven Lead Discovery

Michael S. Lawless, David Miller, and Marvin Waldman

Spring ACS 2022

San Diego CA

AI-driven Lead Discovery

- Computer based *de novo* drug design background
- The AIDD Module in ADMET Predictor®
- Using chemically intelligent SMIRKS transformations to generate molecules
- Synthetic feasibility assessment
- Incorporating pharmacokinetic and ADMET liability predictions
- Penalizing out of scope predictions
- Using capping values when objective function value is “good enough”
- Using external programs to compute objective functions
- Selecting Pareto optimal compounds
- AIDD module demonstration

Computer Aided *De Novo* Drug Design

Structure Based Design (SBD)
- Ludi
- MCSS/Hook
- Sprout

"It ain't just activity anymore"
- Lipinski Rule of 5
- ADMET
- Drug design in multi-objective

Multi-objective LBD & SBD
- Typically, multi-objective combined into a single function
- EA-Inventor
- Muse
- In-house pharma programs

- Pareto-based optimization
- Deep learning generative algorithms

Generating and Selecting/Scoring Virtual Molecules

Generating virtual molecules

- Elementary “Transforms”
  - Modify molecules with small changes or fragment additions/deletions
    - Change/Add single atom or bond
    - Add common fragments
      - Carboxylic acid, sulfonamide, phenyl, etc.
    - Synthetic Feasibility?
      - Filter molecules based on structural alerts, e.g. hemiacetal, peroxide, etc.
  - Combinatorial chemistry libraries
    - Known synthetic reactions and building blocks
    - Chemical Diversity/Novelty?
    - SMIRKS-based
  - Deep Learning
    - Generative algorithms based on SMILES or graph representations

Molecule selecting/scoring approaches

- Multi-objective Criteria
  - Weighted Sum or “Combining” Function
  - Pareto optimal
  - Criteria
    - Activity/Docking scoring
    - ADMET liabilities
    - Similarity to known active/lead
    - Synthetic accessibility
    - Drug-likeness
    - Chemical filters

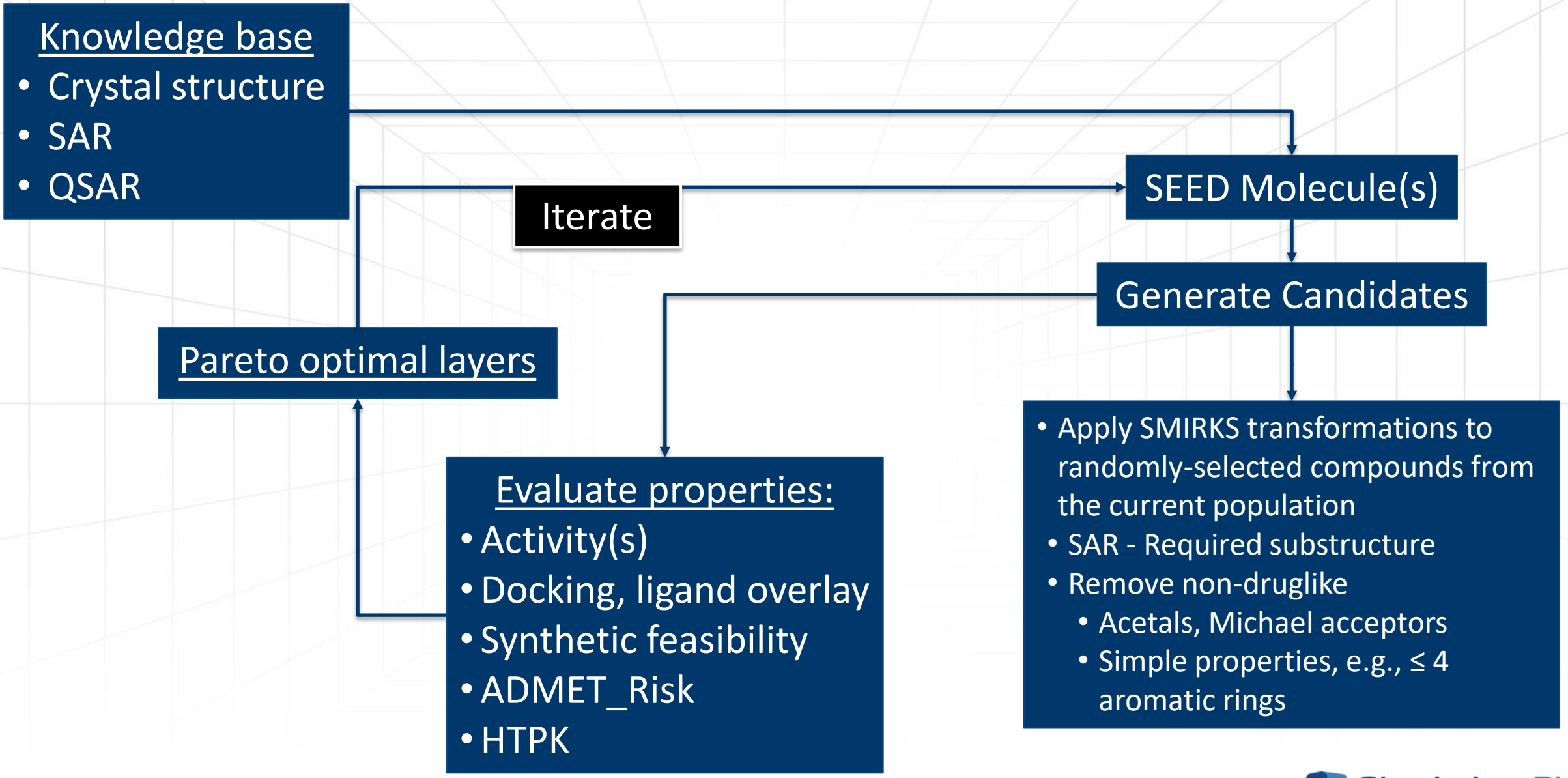
AIDD Module in ADMET Predictor®

Goal – design compounds that have:

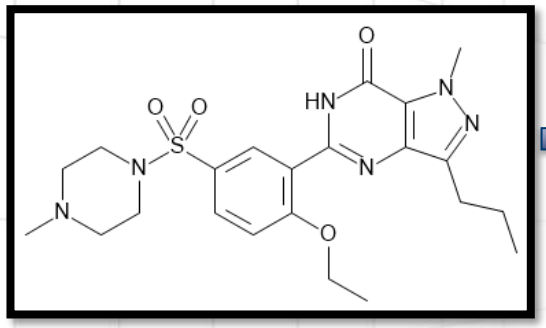
- High potency and selectivity at the primary target
- High synthetic feasibility
- Good ADMET and pharmacokinetic (PK) properties

Can be used for hit finding, lead optimization, and scaffold hopping e.g., patent busting

AIDD Workflow



Generate Molecules Using Transform Rules

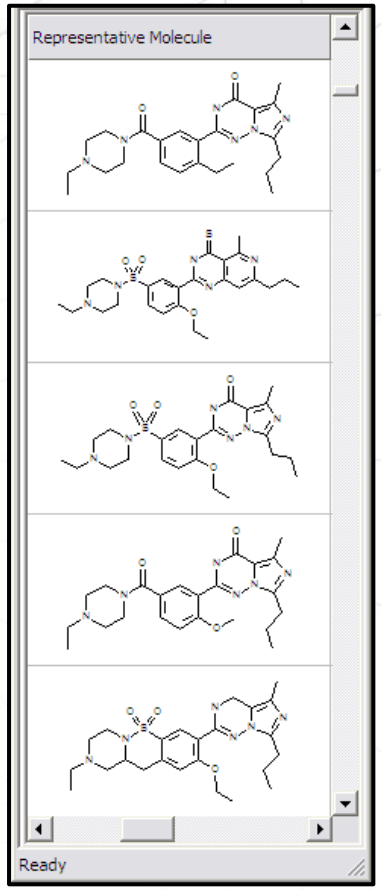


SMIRKS Transforms

- Bioisosteric replacements
- Reactions from literature or in-house expertise
- Chemically-intelligent "mutations"

Users can modify the default rules

Generated analogs

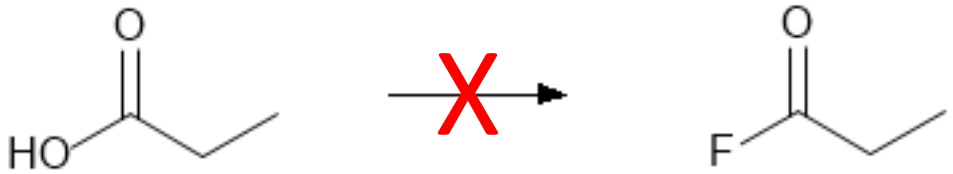


Stewart K.D. et al. *Bioorg. Med. Chem.* **2006**, 14, 7011-7022.

SimulationsPlus
Cognigen | DILIsym Services | Lixoft

Chemically Intelligent Transforms

- Example: Convert Non-fluorine to fluorine
  - Simple version: [!#9:1;D1_S]>>[#9:1]
  - Problem (Need to avoid)

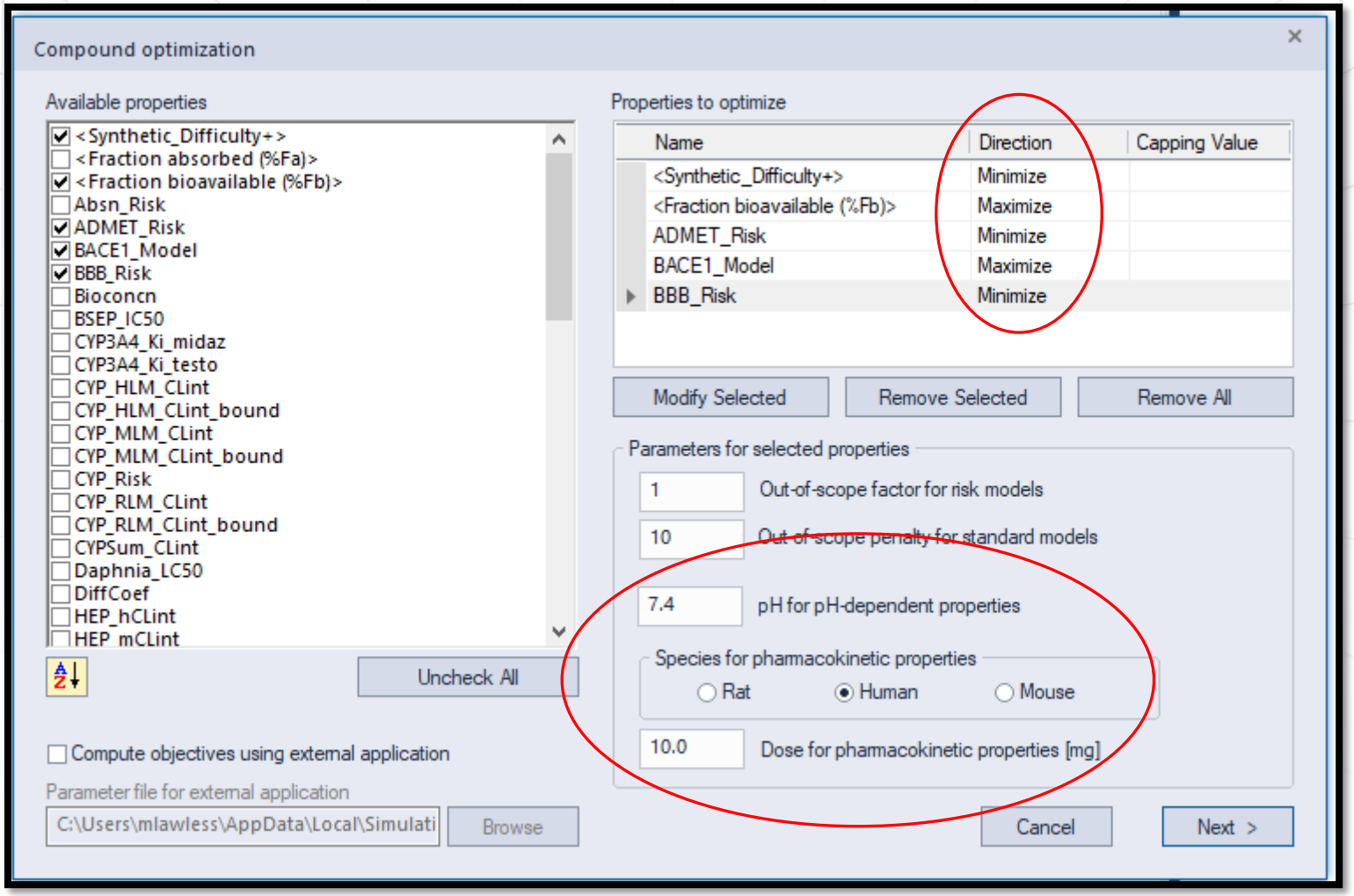


  - Improved SMIRKS: [!#9;D1_S$(*~[#6])!$(*C=[O,N,S]):1]>>[#9:1]
  - Currently ~150 transforms

SimulationsPlus
Cognigen | DILIsym Services | Lixoft

Properties

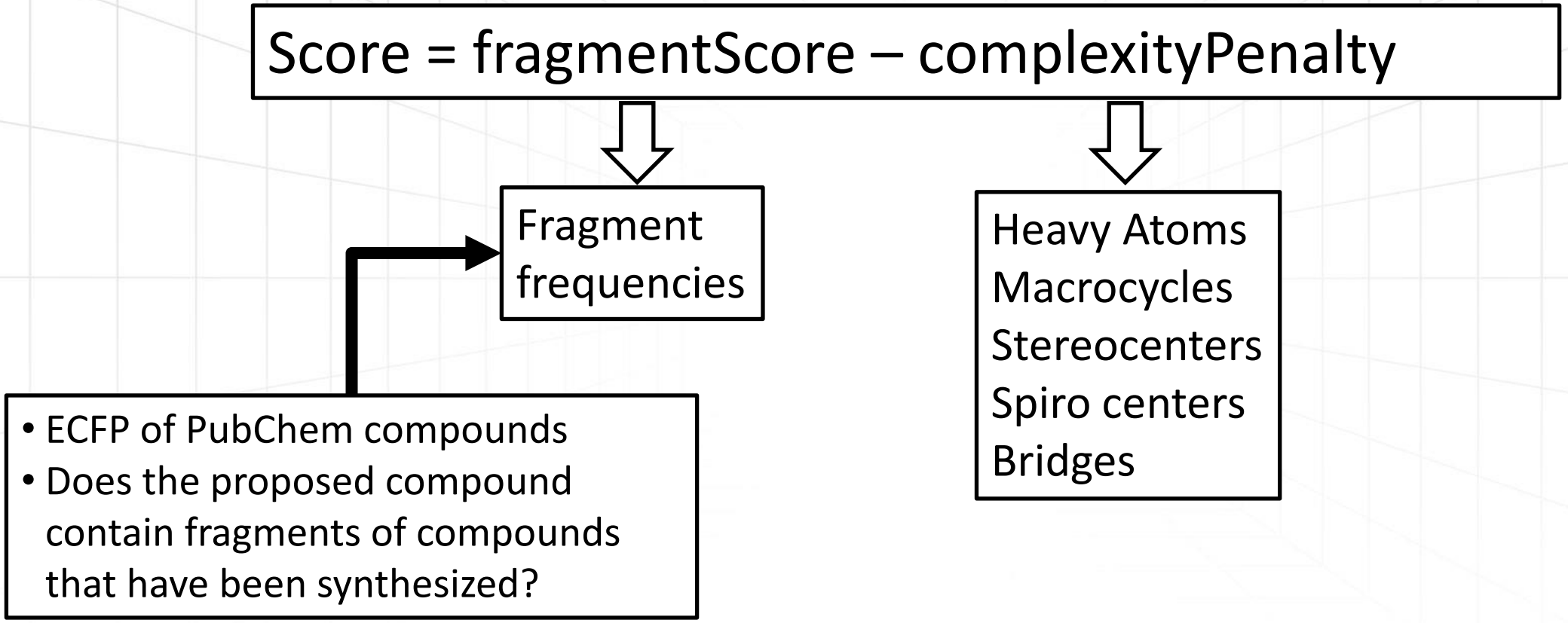
~50 Built-in models
%Fa, %Fb
Synthetic Difficulty+
User Models



SimulationsPlus
Cognigen | DILIsym Services | Lixoft

Synthetic Feasibility Assessment

Based on method from Ertl and Shuffenhauer, *J. Cheminfo,* **2009**, 1**,** 8.

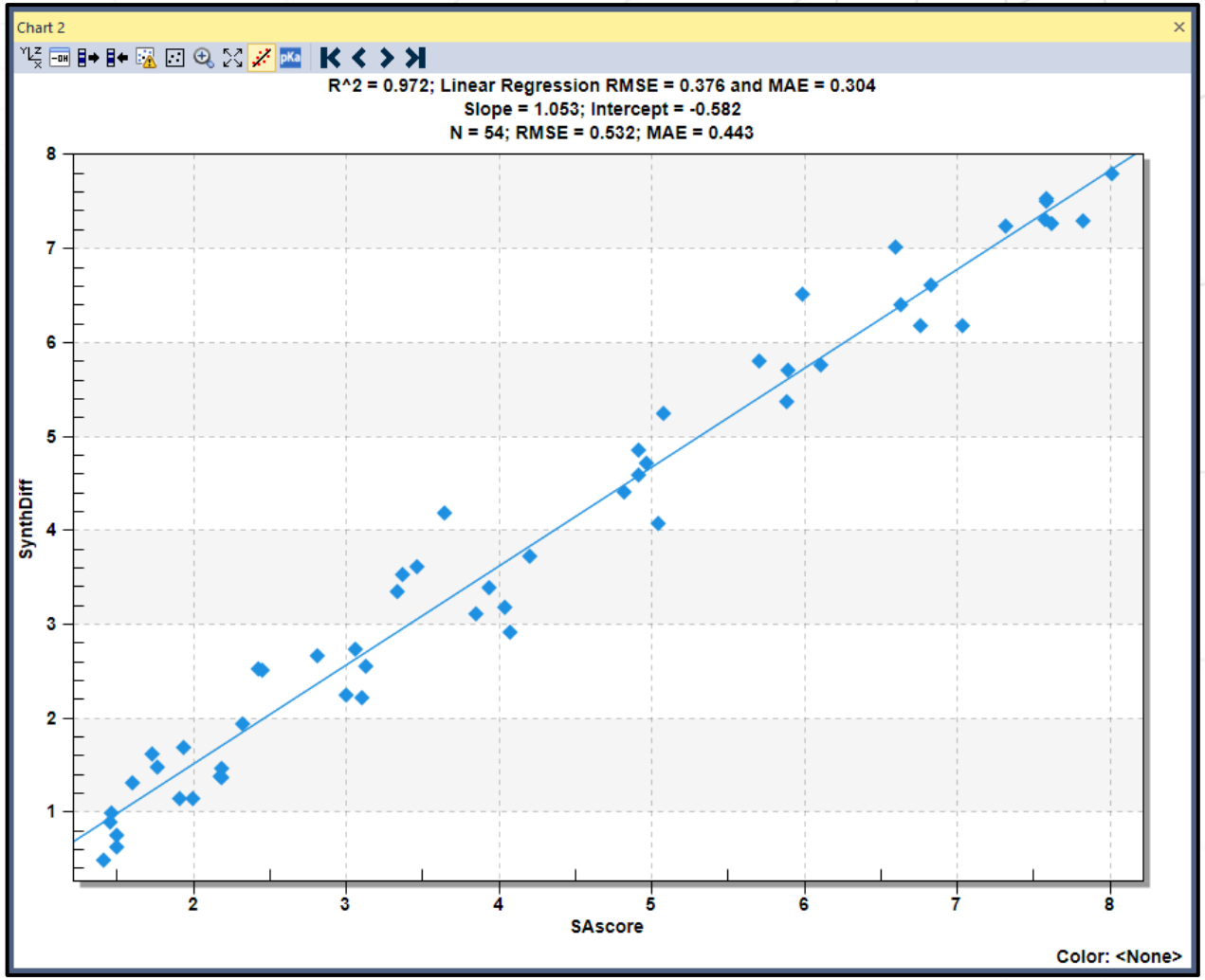


SimulationsPlus
Cognigen | DILIsym Services | Lixoft

Synthetic Accessibility/Difficulty

	SA Ertl[1]	SynthDiff[2]
Training	~1 million	~47 million
Outer Layer	Any	aromatic vs. aliphatic
Complexity	Same	Same
Range	1-10	0-10

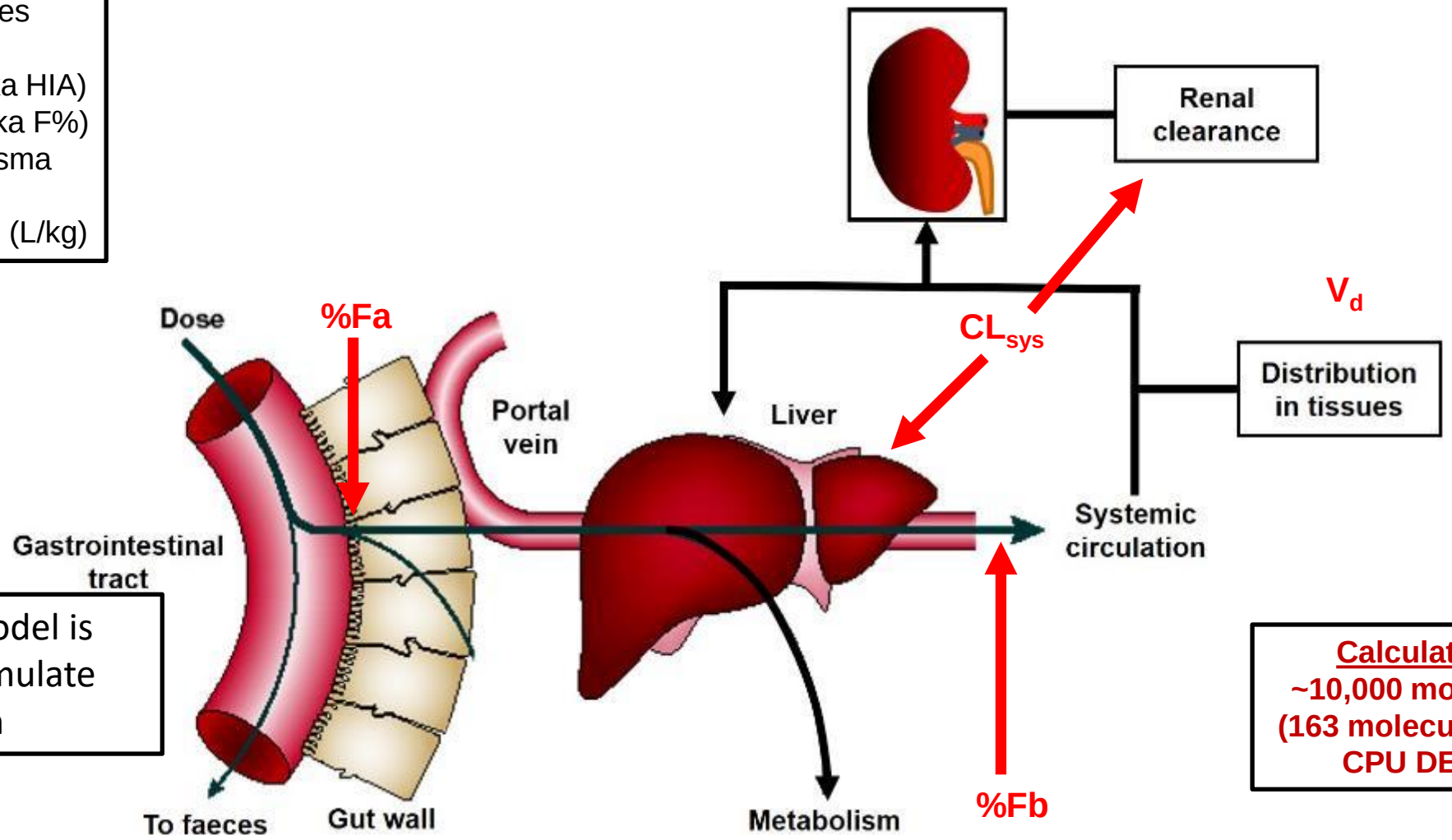
[1]Ertl and Shuffenhauer, *J. Cheminfo,* **2009**, 1, 8.
[2]Implemented in ADMET Predictor



SimulationsPlus
Cognigen | DILIsym Services | Lixoft

HTPK Simulation Module

Predicts:
- Rat and human species
- IR tablet and IV bolus
- Percent absorbed (aka HIA)
- Oral bioavailability (aka F%)
- Dose required for plasma concentration
- Volume of distribution (L/kg)



ACAT™ model is used to simulate absorption

Calculation Speed
~10,000 molecule in 60 s (163 molecules per s) on 8 CPU DELL laptop

Gut clearance and active transport (efflux/influx) are **not** considered

* Modified from van de Waterbeemd, H, and Gifford, E. Nat. Rev. Drug Disc. 2003, 2:192-204

SimulationsPlus
Cognigen | DILIsym Services | Lixoft

ADMET Risk™

Identifier	Weight	Rule logic
Size	1.0	MWt > 450{500} OR N_Atoms > 30{35} OR MolVol > 470{520} OR N_Bonds > 35{40}
RotB	1.0	N_FrRotB > 8{10}
HBD	1.0	HBDH > 3{5} AND HBDch > 1.5{2.0}
HBA	1.0	HBA > 7{10} AND HBAch < -6.0{-5.0}
ch	1.0	NPA_ABSQ > 19{21} OR T_PSA > 120{140}
Kow	1.0	S+logP > 4.5{5.0} OR S+logD > 3.5{4.0} OR MlogP > 3.5{4.0}
Peff	1.0	S+Peff < 0.40{0.60} OR S+MDCK < 10{25}
Sw	1.0	S+Sw < 0.005{0.010}
fu	1.0	hum_fup% < 4{6}
Vd	1.0	Vd > 4{5}
hERG	1.0	hERG_Filter = Yes AND hERG_pIC50 > 5.5{6.0}
rat	1.0	Rat_Acute < 200{300}
Xr	1.0	Rat_TD50 < 2.5{5.0}
Xm	1.0	Mouse_TD50 < 25{40}
HEPX	1.0	Ser_AST = Elevated AND Ser_ALT = Elevated AND Ser_LDH = Elevated
MUT	1.0	MUT_Risk > 1
1A2	1.0	(CYP1A2_CLint > 20{40} AND CYP1A2_Substr = Yes)
2C9	1.0	(CYP2C9_CLint > 10{20} AND CYP2C9_Substr = Yes)
2C19	1.0	(CYP2C19_CLint > 10{20} AND CYP2C19_Substr = Yes)
2D6	1.0	(CYP2D6_CLint > 10{20} AND CYP2D6_Substr = Yes)
3A4	1.0	(CYP3A4_CLint > 20{50} AND CYP3A4_HLM_CLint > 30{75} AND CYP3A4_Substr = Yes)
CL	1.0	CYP_HLM_CLint > 90{150} OR HEP_hCLint > 60{90}

Absorption

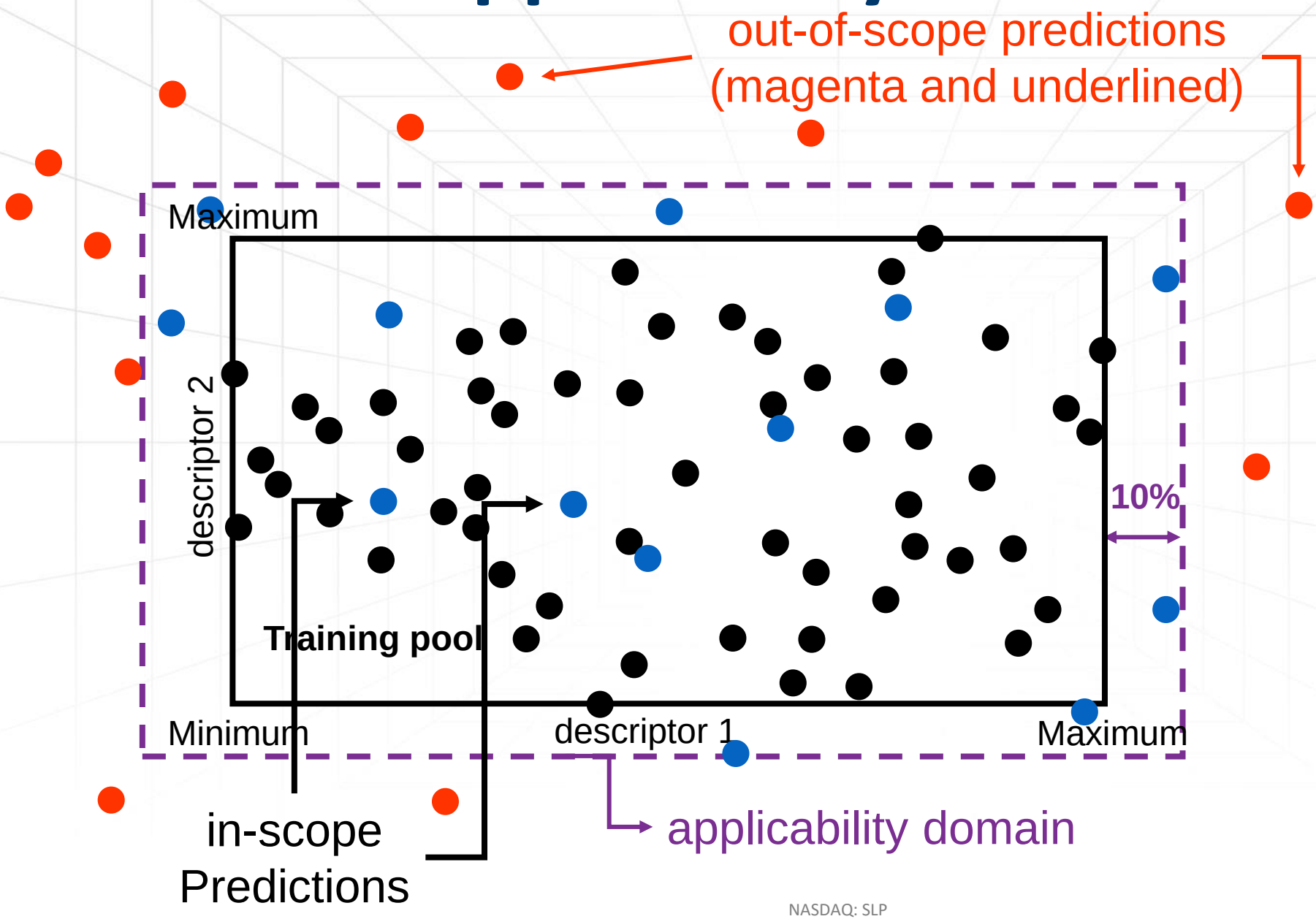
Distribution

Toxicity

Metabolism

SimulationsPlus
Cognigen | DILIsym Services | Lixoft

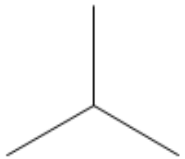
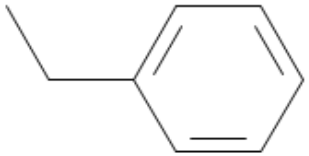
Applicability Domains



In AIDD a penalty factor can be applied to avoid out of scope predictions from being on the Pareto front.

SimulationsPlus
Cognigen | DILIsym Services | Lixoft

Flip Side of the Coin: Capping Values

	Structure	OBJ_HIVI-ST	ADMET_Risk	SynthDiff
1		3.645	1.000	0.000
2		3.676	1.000	0.000
3		3.846	1.000	0.000
4		3.506	0.431	0.131
5		3.173	1.000	0.243
6		3.250	1.000	0.384

Properties to optimize

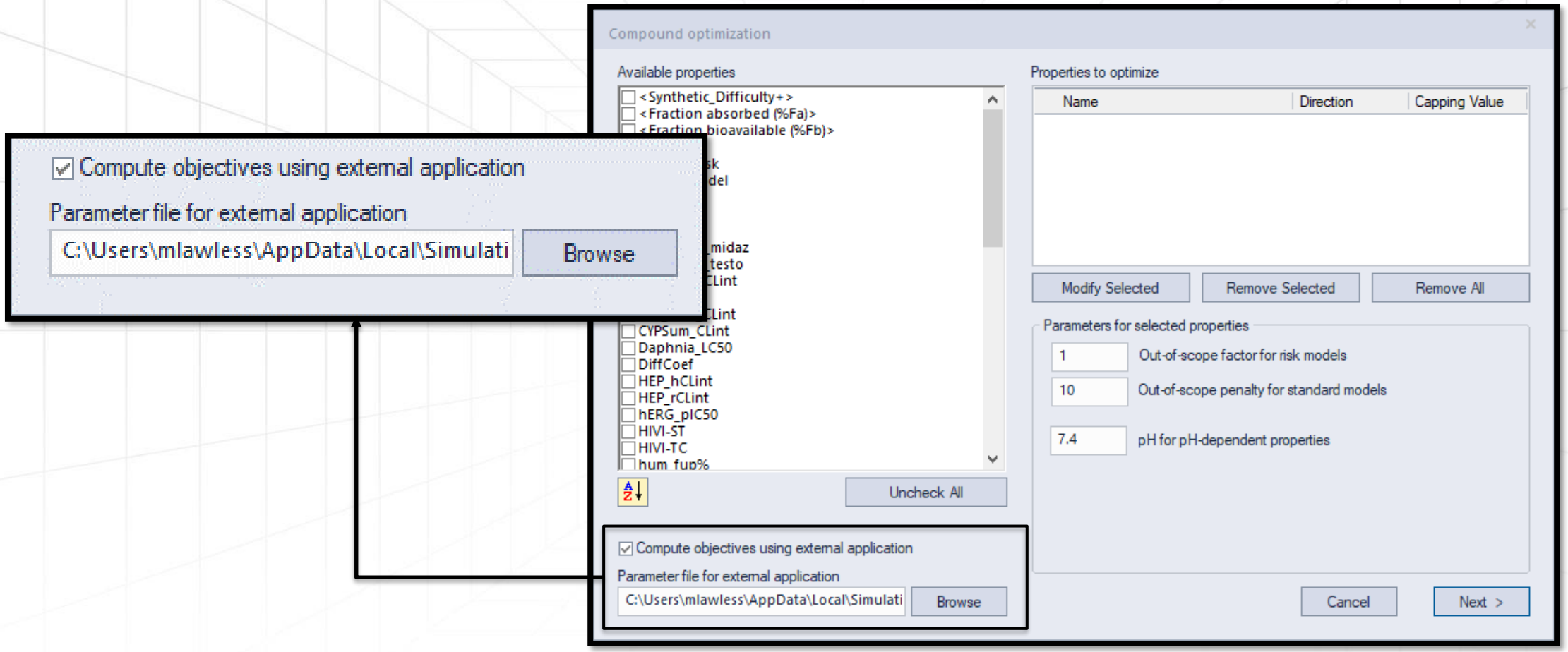
Name	Direction	Capping Value
<Synthetic_Difficulty+>	Minimize	2.5
ADMET_Risk	Minimize	
<Fraction bioavailable (%Fb)>	Maximize	90

Modify Selected | Remove Selected | Remove All

Trivially Simple Molecules:
Very easy to make
Very good in one objective

- Capping value is assigned as the result when the actual value is “better”
  - Value is “good enough”
- Capping value tends to filter out trivial molecules
- Doesn’t encourage the algorithms to go into “simplistic” compounds

AIDD External Application File



External File Parameters

filePathAlias – directory name

- Defines [PATH] variable
- Can be set to AIDD output directory

commandLine – script or executable program

- BAT, shell script, or executable

structureFile – name of file to contain AIDD generated structures

- SDF or SMILES file

InputAttributes – comma separated list of AP attributes, e.g., S+logP, to be written to structure file

outputFile – name of file containing objective function value(s)

- Created by external program

objectivesList – names of objective functions from external program

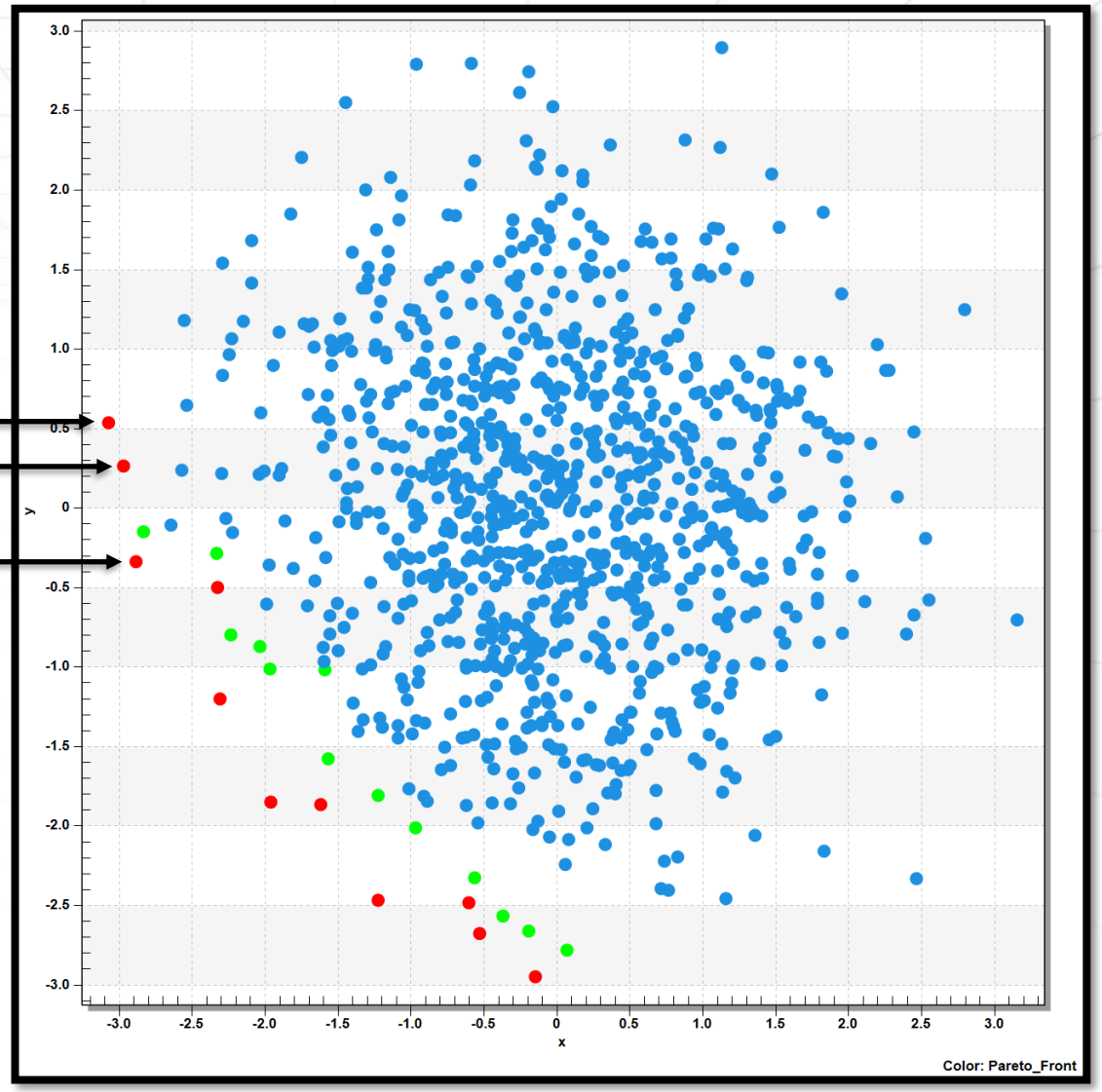
objectiveMaximize – 0 if minimum is preferred and 1 if maximum is preferred

Pareto Optimal Selection

Point 1 – best y value

Point 2 – better x value than point 1

Point 3 – better x value than points 1 and 2



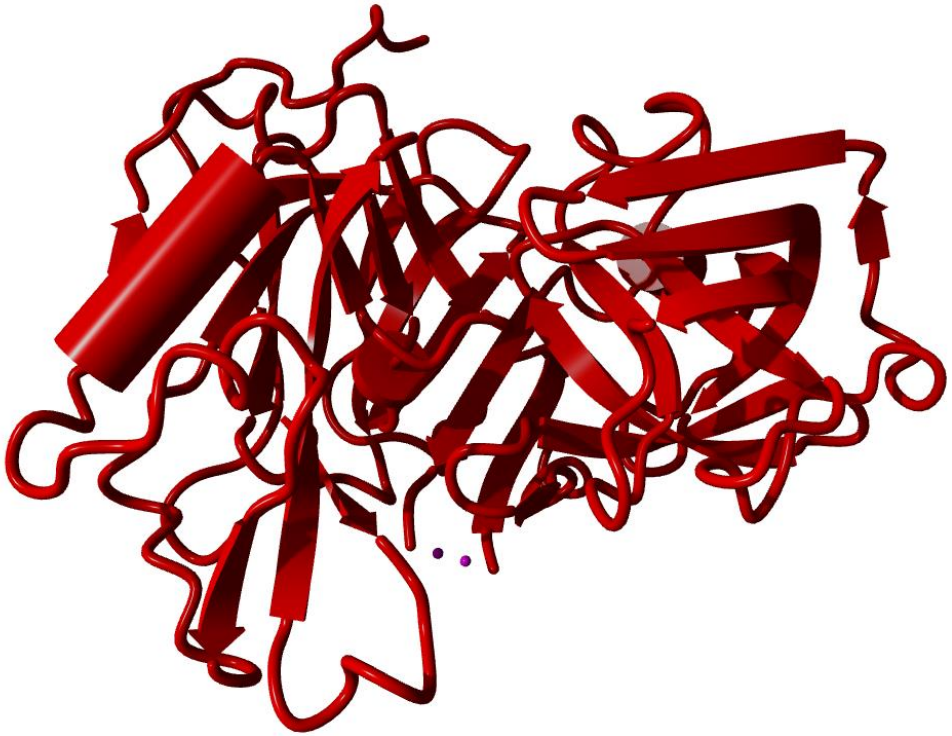
SimulationsPlus
Cognigen | DILIsym Services | Lixoft

Compromise Solution

- Multi-objective, but combine related objectives (e.g. ADMET) into one : ADMET_Risk™
- Typical/Recommended use: 4-5 objectives
  - ADMET_Risk
  - Synthetic Difficulty
  - 1-2 activity models (e.g., activity and selectivity)
  - Good PK (e.g, bioavailability)

BACE1 Inhibitors

- BACE1 (β-secretase 1) cleaves the amyloid precursor protein producing a soluble peptide C99, which is the rate-limiting step in amyloid β formation. It is a target for Alzheimer's disease (AD).
- "Despite BACE1 inhibitor clinical trials conducted so far being discontinued for futility or safety reasons, BACE1 remains a well-validated therapeutic target for AD. A safe and efficacious compound with high substrate selectivity as well as a more accurate dose regimen, patient population, and disease stage may yet be found."*

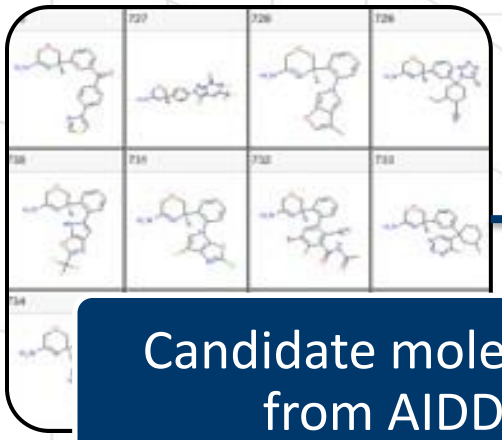


*Hampel, et al. Biol Psychiartry. **2021**, Apr 15;89(8):745-756.

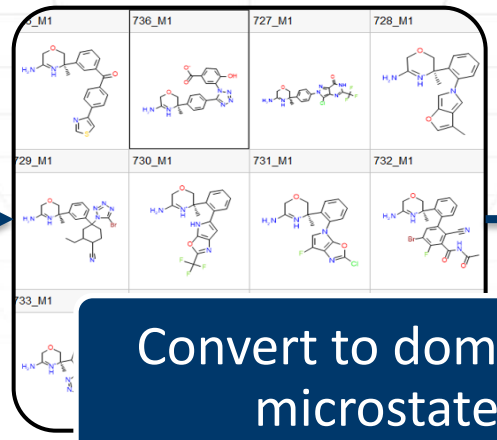
Acknowledgements

- AutoDock – Ravindranath PA, Forli S, Goodsell DS, Olson AJ, and Sanner MF. AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility, *PLoS Computational Biology* **2015**, 11(12): e1004586.
- YASARA – Krieger E and Vriend G. YASARA View-molecular graphics for all devices-from smartphones to workstations, *Bioinformatics*, **2014**, 30(20), 2981-2982.
- Babel - O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
- 6FGY – Veenstra, SJ. et al. Discovery of amino-1,4-oxazines as potent BACE-1 inhibitors. *Bioorg Med Chem Lett,* **2018**, 2195.
- Rombouts FJR et al. 1,4-Oxazine β-Secretase 1 (BACE1) Inhibitors: From Hit Generation to Orally Bioavailable Brain Penetrant Leads, *J Med Chem* **2015**, 58, 8216.

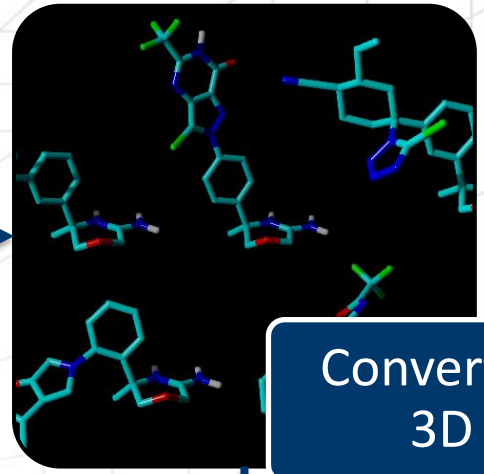
Workflow for AutoDock AIDD Example



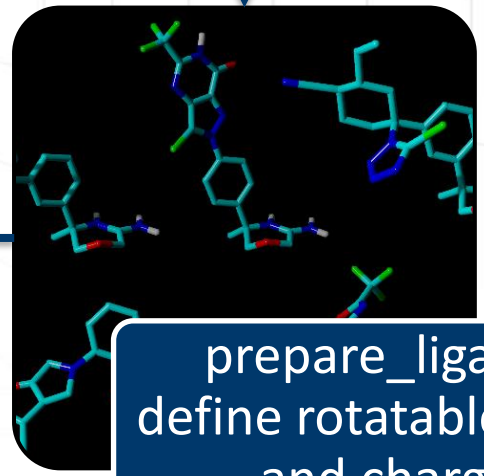
Candidate molecules from AIDD



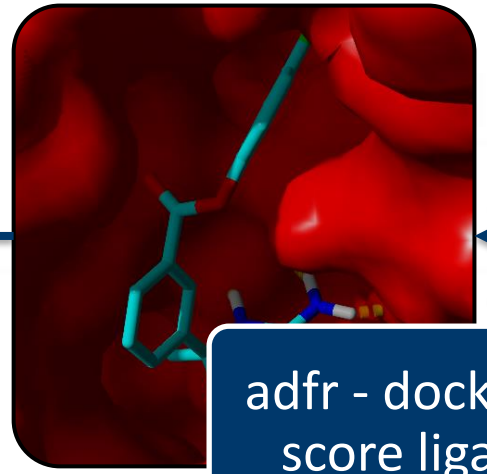
Convert to dominant microstate



Convert to 3D



prepare_ligand – define rotatable bonds and charges



adfr - dock and score ligand

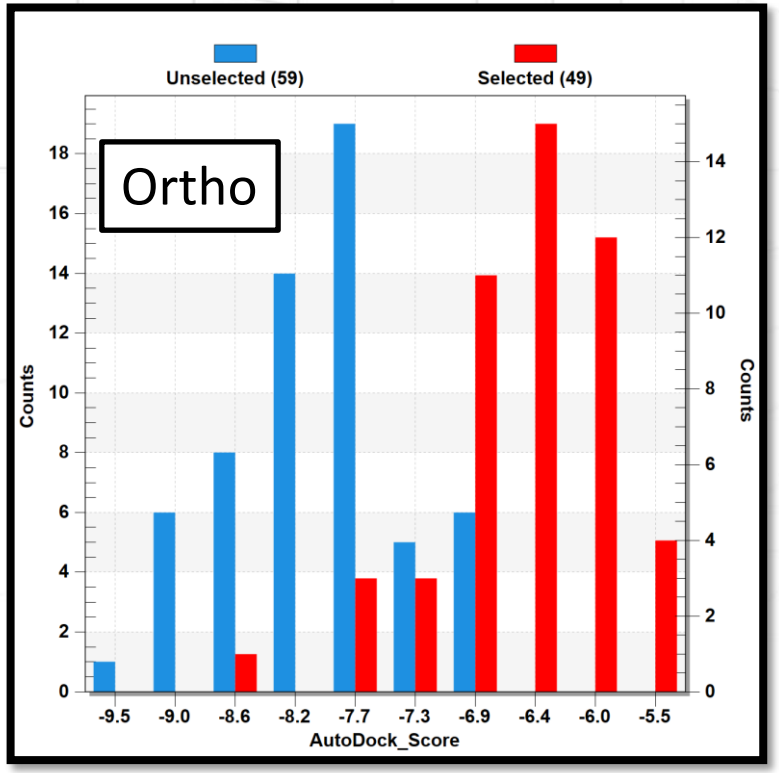
Return results to AIDD

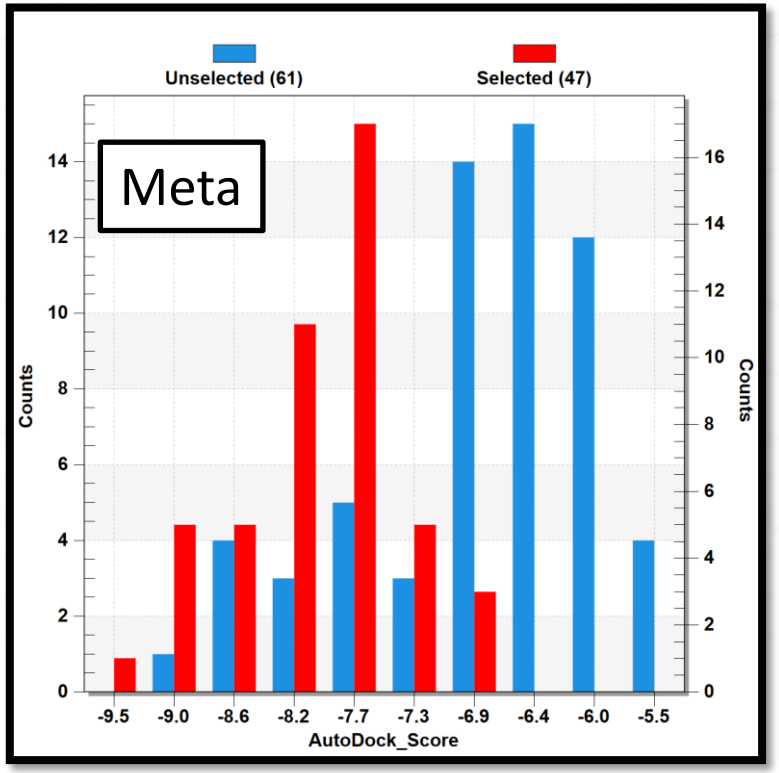
SimulationsPlus
Cognigen | DILIsym Services | Lixoft

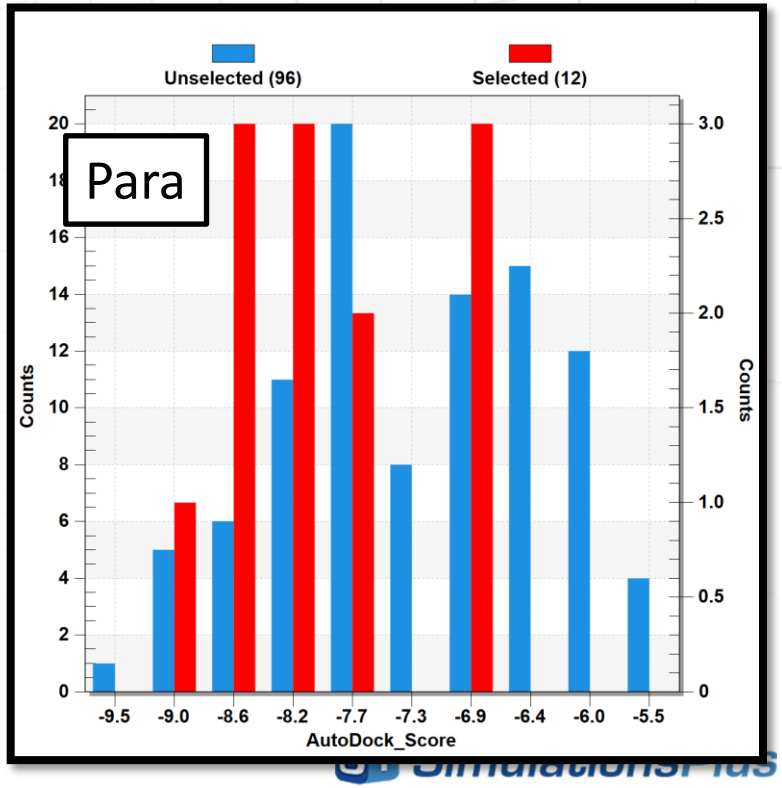
Docking Score Statistics

Phenyl substituent	Number	Best	Worst	Ave
Ortho	49	-8.5	-5.5	-6.5
Meta	47	-9.5	-7.0	-8.0
Para	12	-9.0	-6.8	-7.9

Meta substituents have best docking scores followed by molecules with substituents in the para position







Cognigen | DILIsym Services | Lixoft

AI-driven Lead Discovery

- Computer based *de novo* drug design background
- The AIDD Module in ADMET Predictor®
- Using chemically intelligent SMIRKS transformations to generate molecules
- Synthetic feasibility assessment
- Incorporating pharmacokinetic and ADMET liability predictions
- Penalizing out of scope predictions
- Using capping values when objective function value is “good enough”
- Using external programs to compute objective functions
- Selecting Pareto optimal compounds