Delivering on the promise of AI-driven drug discovery with ADMET Predictor® 10 Background and application examples

Marvin Waldman
Sr. Research Fellow
30-Sep-2020

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Motivation

- In-house project to develop novel anti-malarial compounds as proof of concept to use in-silico models to design compounds

Journal of Computer-Aided Molecular Design
https://doi.org/10.1007/s10822-020-00333-x

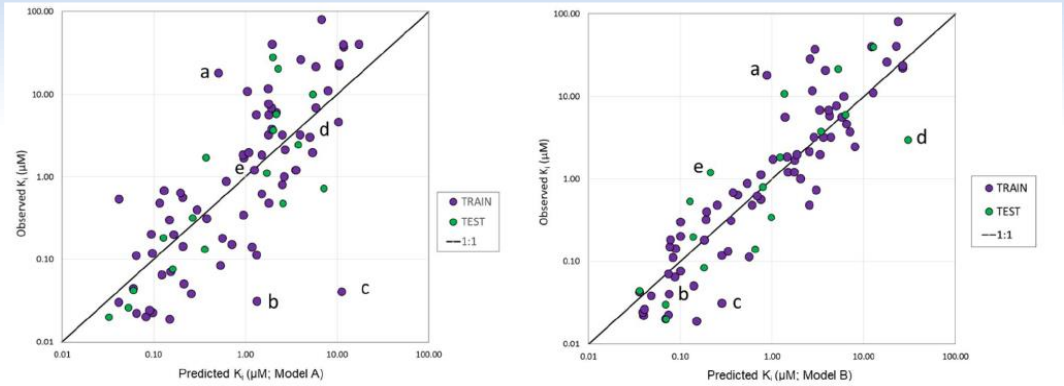
Design and tests of prospective property predictions for novel antimalarial 2-aminopropylaminoquinolones

Robert D. Clark[1] · Denise N. Morris[2] · Gary Chinigo[3,6] · Michael S. Lawless[1] · Jacques Prudhomme[4] · Karine G. Le Roch[4] · Maria José Lafuente[5] · Santiago Ferrer[5] · Francisco Javier Gamo[5] · Robert Gadwood[3] · Walter S. Woltosz[1]

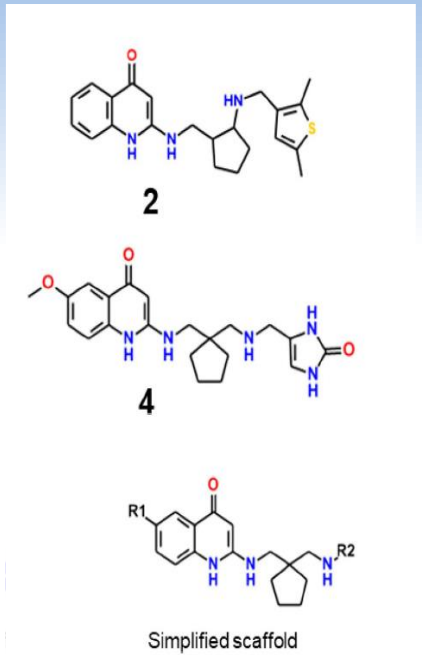
Received: 9 April 2020 / Accepted: 21 July 2020

Compound Design Protocol

Activity Models



Analog Generation

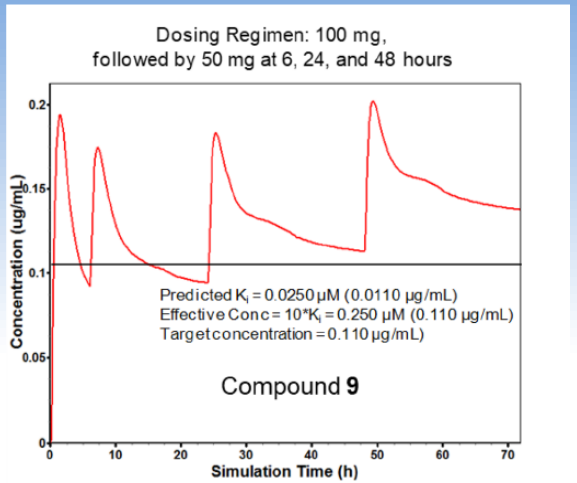


SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Compound Design Protocol

Compound	8	9
Name	SLP0005	SLP0003
S+Sw (μg/mL)	1.8	0.32
obsd. solubility	33	0.76
S+logP	4.7	5.05
obsd. logP	4.2	4.4
S+pK_{a1}	5.24	5.24
obsd. pK_{a1}	4.95	4.80
S+pK_{a2}	8.01	7.63
obsd. pK_{a2}	8.33	7.36
S+$logD_{6.8}$	3.44	4.46
obsd. $logD_{6.8}$	2.68	3.76

Compound	Pred. *Pf*DHODH	XC_{50} (μM)[a,b]	
	K_i (μM)	3d7(−)	Dd2(+)
12a	0.049	10.0	46
12b	0.051	1.61	6.4
11a	0.023	0.55	2.3
11c	0.037	0.37	1.78
8	0.037	0.30	1.47
9	0.025	0.106	0.21
11b	0.038	0.037	0.24
CID 44534046[c]	0.112	0.89	4.6
CID 44535189[c]	0.077	0.85	8.6



Predict/Measure Properties

Conclusion: “It seems likely that further iterations and in vivo characterization would be productive.”

Can this be automated?

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Background

- History of Computer-Based ("*De Novo*") Drug Design
  - Early 1990's
    - Structure (Receptor) Based
      - Ludi, MCSS/Hook, Sprout
  - Late 1990's – Rule of 5, ADMET, Drug Design is multi-objective
    - "It ain't just activity anymore"
  - Early to late 2000's
    - Multi-objective ligand and structure-based design
      - Typically multi-objective "combined" into a single function
      - EA-Inventor, Muse, in-house Pharma programs
  - In the last decade
    - Pareto-based optimization
    - Deep Learning Generative algorithms

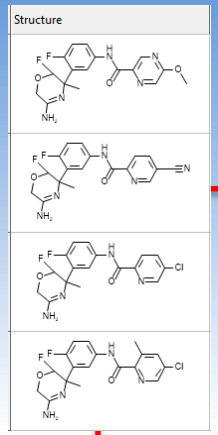
Generating Virtual Molecules

- 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.
  - Library of known synthetic reactions and building blocks
    - Chemical Diversity/Novelty?
    - SMIRKS-based
  - Deep Learning
    - Generative algorithms based on SMILES or graph representations

Molecule Selection/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 Workflow



Initialize with K randomly generated analogs using chemical transforms starting from initial seed molecules

Evaluate properties:
ADMET_Risk
SynthDiff
Activity(s)
HTPK

Prune molecules using Pareto optimal layers

Generate M more analogs using chemical transforms and randomly selected molecules from current population

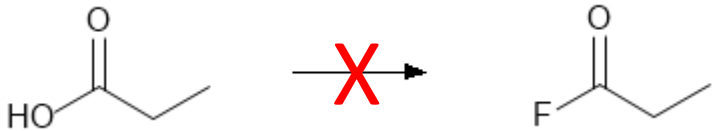
Structure	ADMET_Risk	S+logP	S+logD	S+Peff	S+Sw
	0.961	2.309	2.272	2.231	0.349
	0.930	2.705	2.667	2.687	0.023
	1.558	3.569	3.533	3.379	0.026
	1.110	3.793	3.748	3.462	0.022

Repeat N times

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Generating Analogs

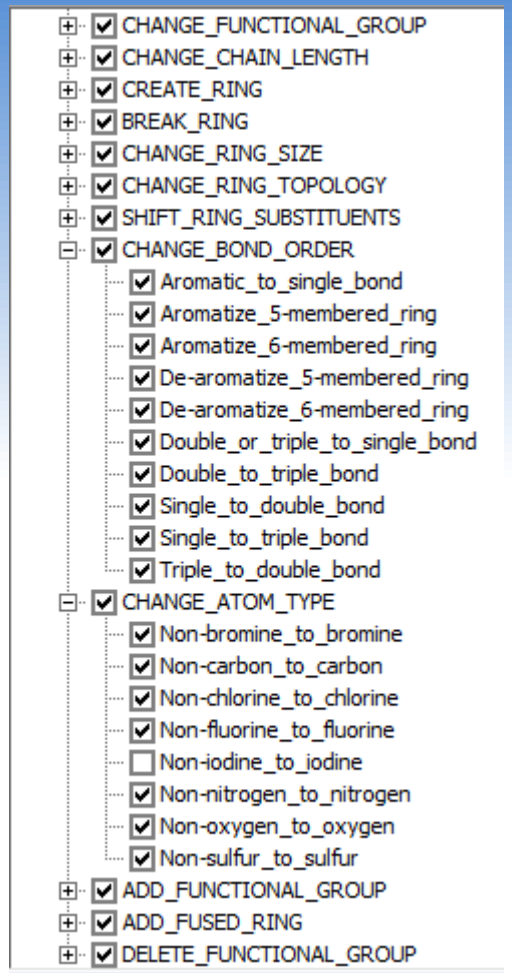
- Uses a library of chemically “intelligent” SMIRKS transforms
  - Example: Non-fluorine_to_fluorine
    - Simple version: [!#9:1]>>[#9:1]
    - Problem (Need to avoid)

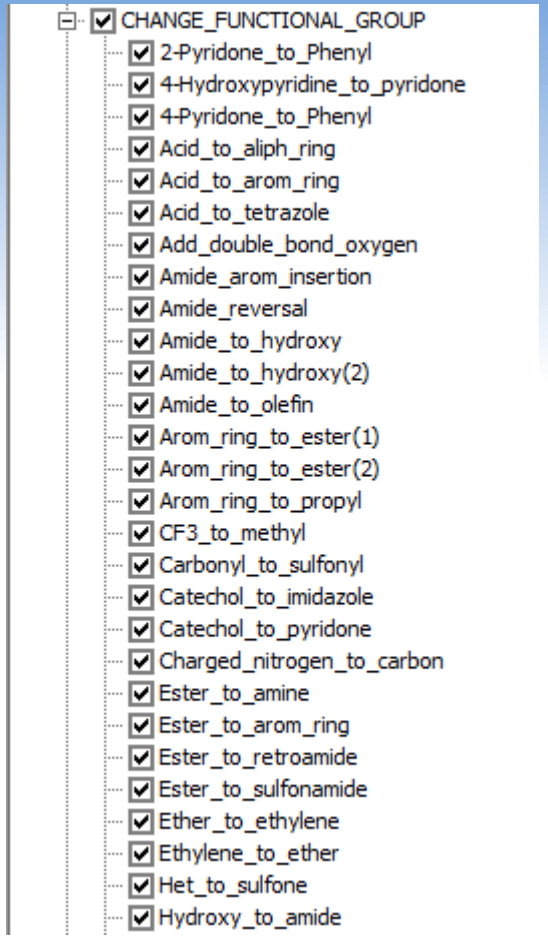


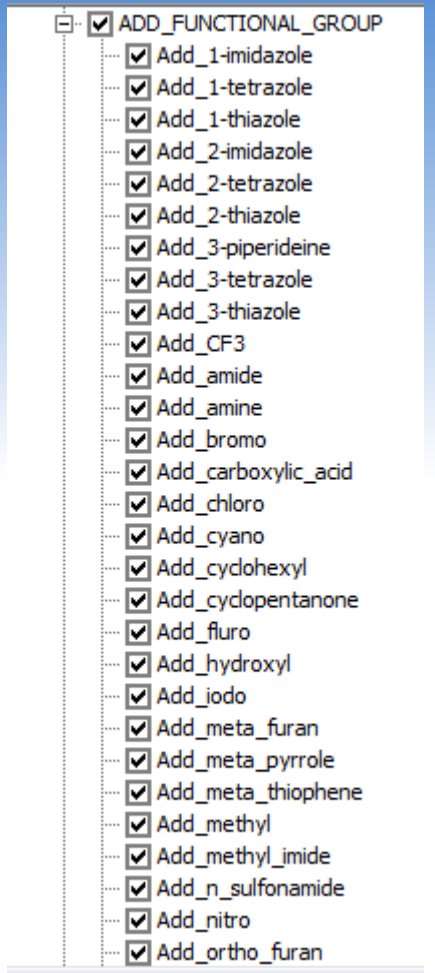
Highly reactive acid halide

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

Transforms

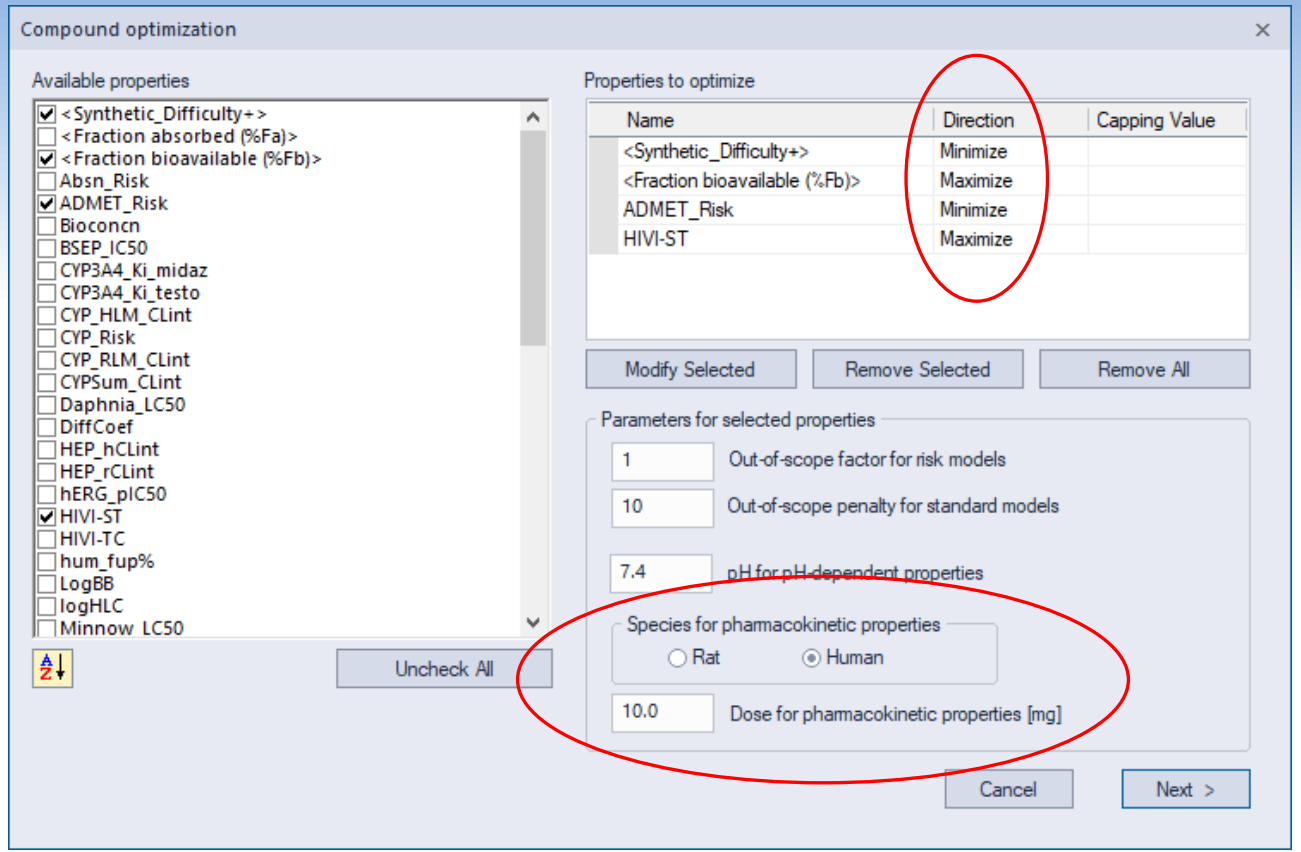






~50 Built-in models
%Fa, %Fb
Synthetic Difficulty+
User Models

Properties

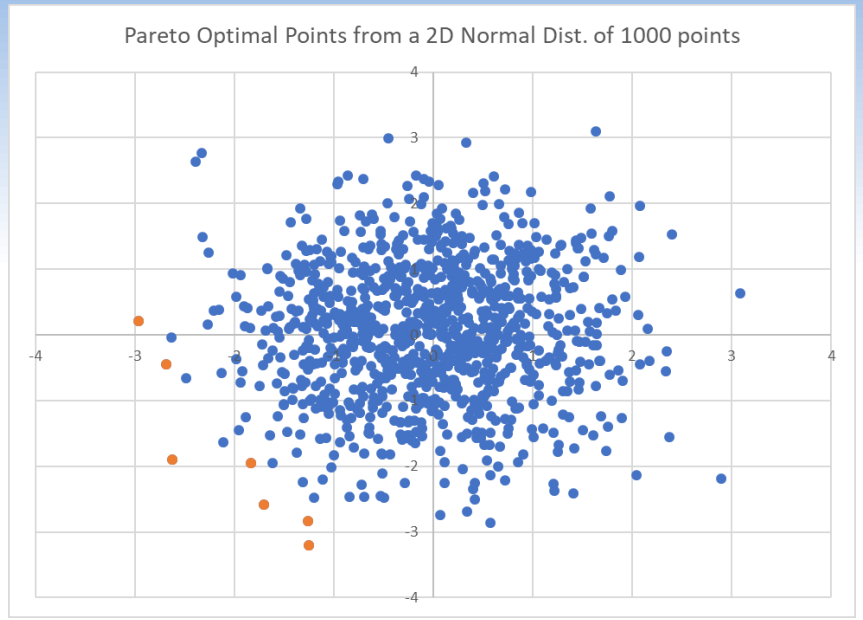


SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

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)

Pareto Selection



- Pareto Results
  - 1000 normally distributed points
    - 2-dim. : 7 Pareto optimal points
    - 5-dim. : ~100 Pareto optimal points

Too Many objectives leads to too many molecules!

Most of them are good in just one or a few properties.

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

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

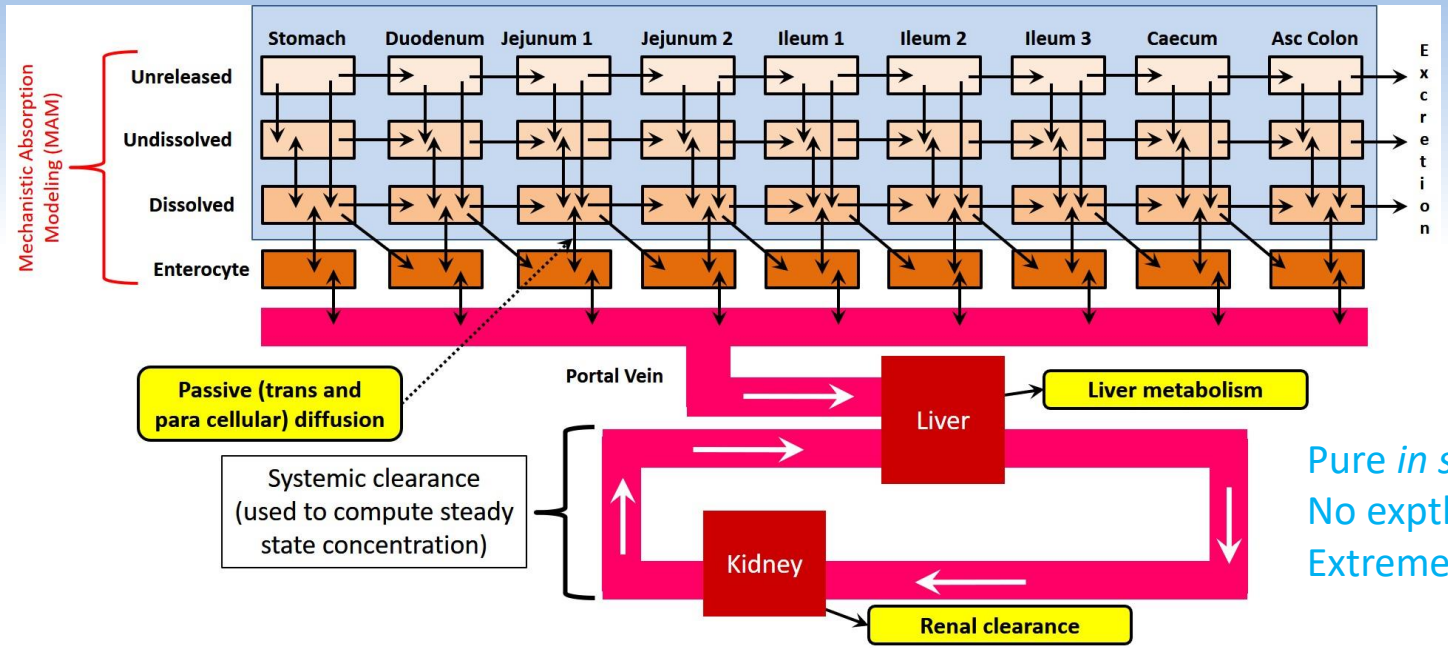
Toxicity

Metabolism

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

PBPK Simulation: Methodology

ACAT™ Model* + Compartmental Model

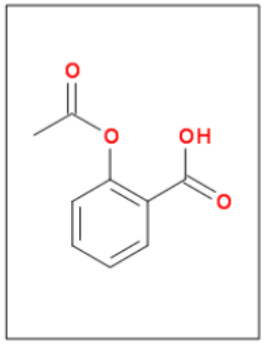


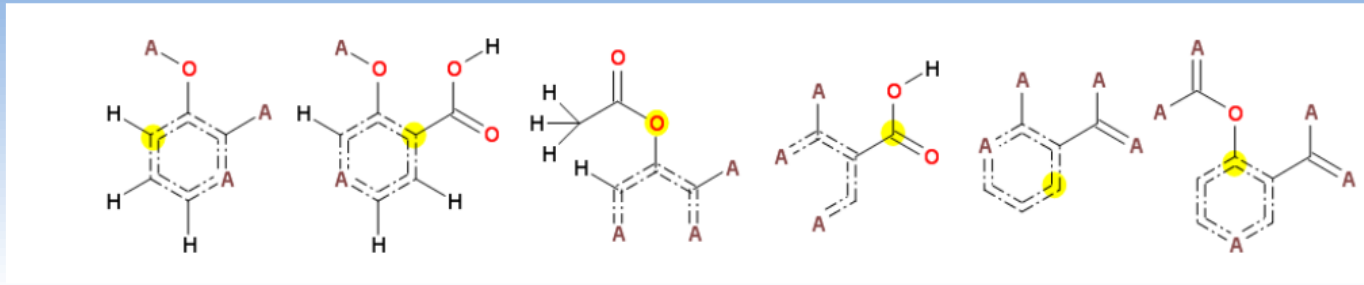
Pure *in silico* simulation
No exptl properties required
Extremely rapid, multi-threaded

* Advanced Compartmental Absorption and Transit Model

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Synthetic Accessibility/Difficulty





Score = fragmentScore – complexityPenalty

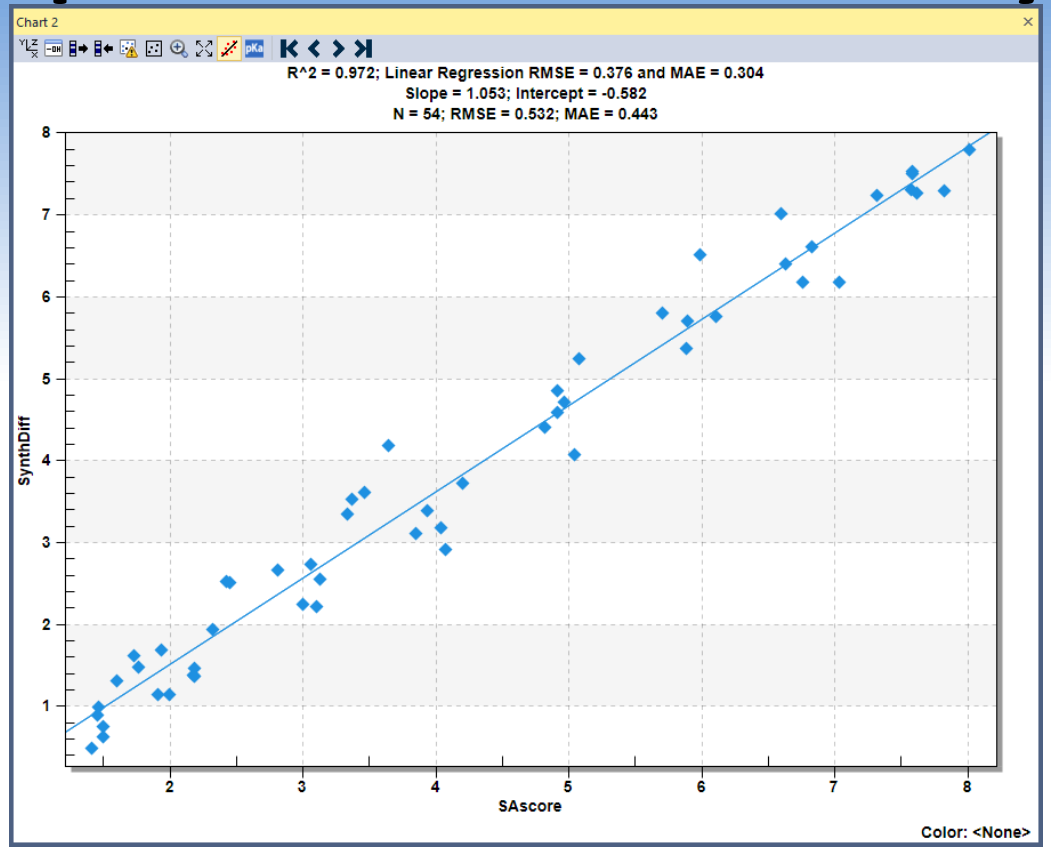
Fragment frequencies

Heavy Atoms
Macrocycles
Stereocenters
Spiro centers
Bridges

	SA Ertl	Synth Diff
Training	~1 million	~47 million
Outer Layer	Any	aromatic vs. aliphatic
Complexity	Same	Same
Range	1-10	0-10

Ertl and Shuffenhauer, J Cheminf, 2009, doi:10.1186/1758-2946-1-8

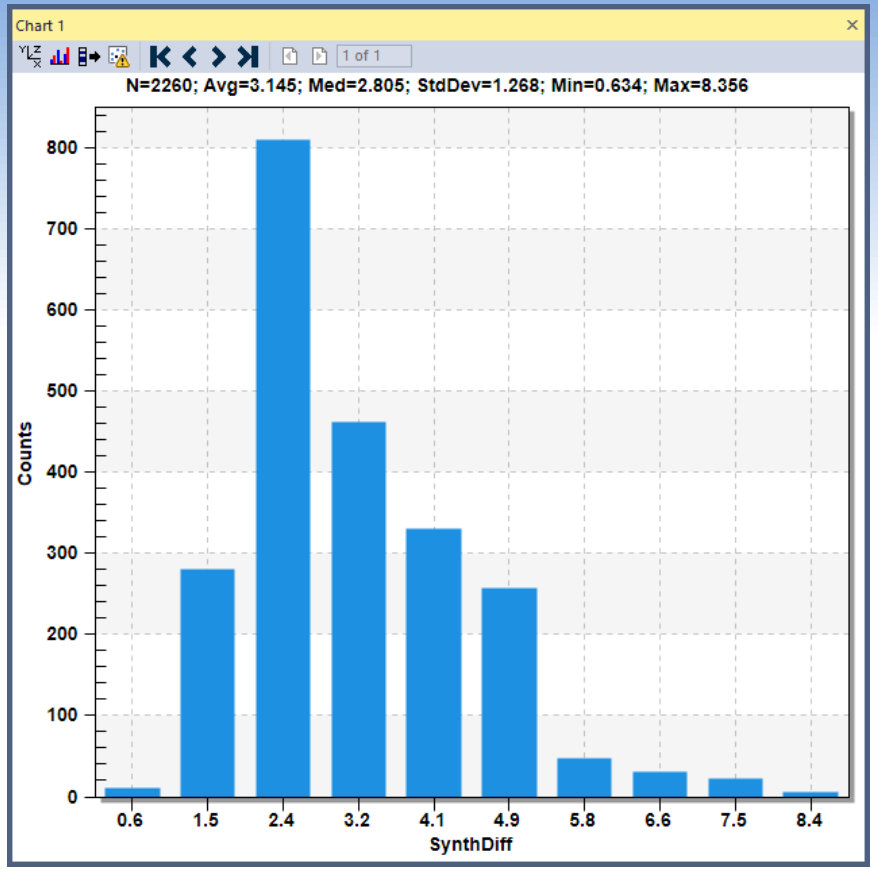
Comparison: SAScore vs. SythDiff



Ertl and Shuffenhauer, J Cheminf, 2009, doi:10.1186/1758-2946-1-8
Li et al., J Cheminf, 2018, doi:10.1186/s13321-018-0287-6

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Distribution of SynthDiff Scores



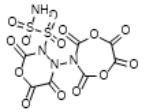
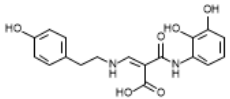
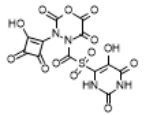
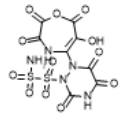
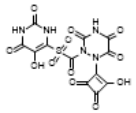
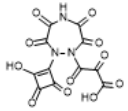
WDI Focused Set of 2260 compounds

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

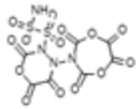
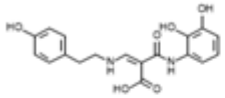
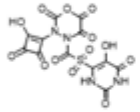
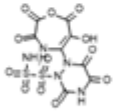
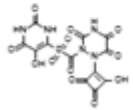
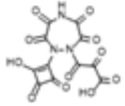
SynthDiff+

- SynthDiff was designed to “score” real molecules
- Performance on virtual molecules sometimes too “optimistic”
  - Add additional penalties based on drug-likeness filters
    - Brenk et al., ChemMedChem, 3, 435 (2008)
    - Rishton, Drug Disc. Today, 2, 382 (1997)
  - Maximum penalty of 4
  - Augmented version preferred over adding another Pareto objective

Early Results

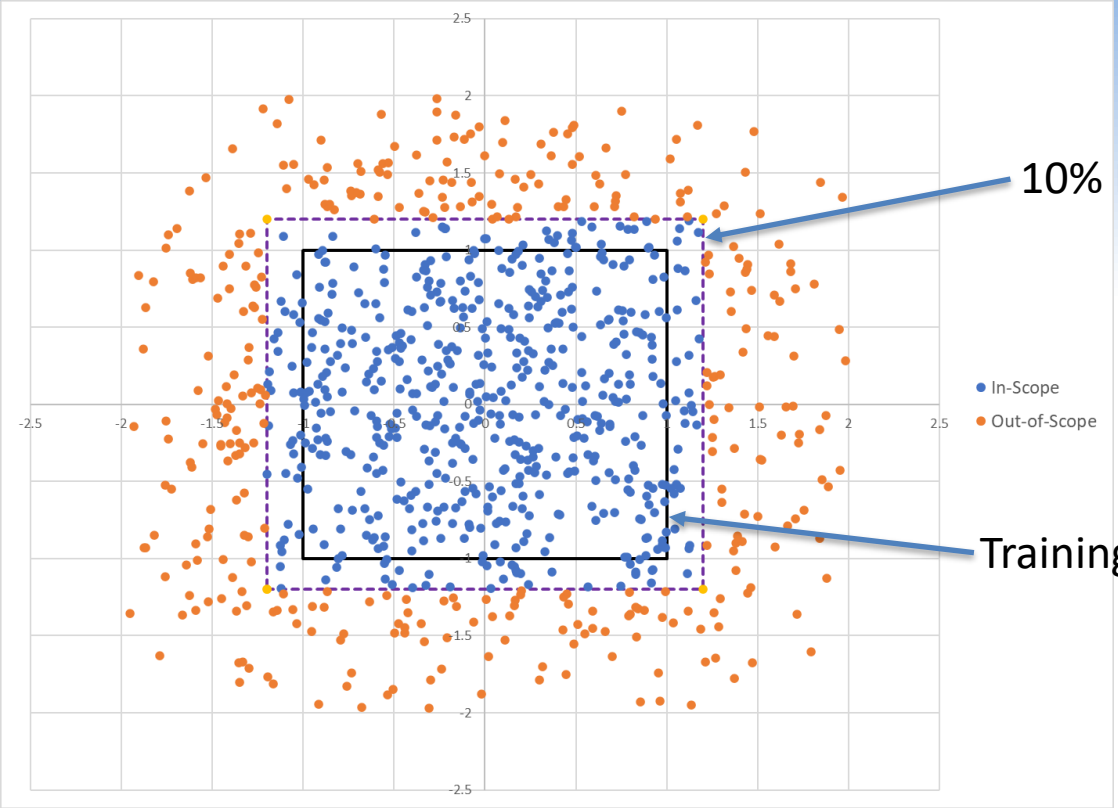
	Structure	OBJ_HIVI-ST	ADMET_Risk	SynthDiff
1		12.799	3.000	4.256
2		6.525	3.956	2.313
3		12.248	3.500	4.181
4		12.163	6.000	4.132
5		11.952	4.000	4.082
6		11.051	1.694	3.674

Out-of-Scope Predictions

	Structure	OBJ_HIVI-ST	ADMET_Risk	SynthDiff	HIVI-ST
1		12.799	3.000	4.256	12.799
2		6.525	3.956	2.313	6.525
3		12.248	3.500	4.181	12.248
4		12.163	6.000	4.132	12.163
5		11.952	4.000	4.082	11.952
6		11.051	1.694	3.674	11.051

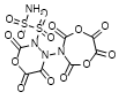
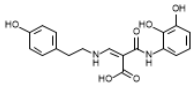
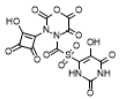
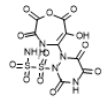
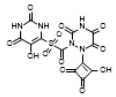
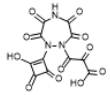
SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Applicability Domain



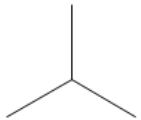
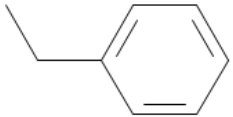
SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

After Applying “Penalties”

	Structure	OBJ_HIVI-ST	ADMET_Risk	SynthDiff	HIVI-ST	HIVI-ST+	ADMET_Risk+	SynthDiff+
1		12.799	3.000	4.256	12.799	2.799	14.000	8.253
2		6.525	3.956	2.313	6.525	-3.475	11.956	6.312
3		12.248	3.500	4.181	12.248	2.248	13.500	7.897
4		12.163	6.000	4.132	12.163	2.163	15.000	7.848
5		11.952	4.000	4.082	11.952	1.952	15.000	6.848
6		11.051	1.694	3.674	11.051	1.051	13.404	7.674

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

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

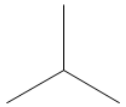
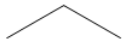
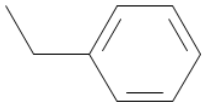
Modify Selected | Remove Selected | Remove All

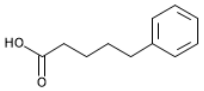
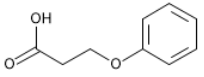
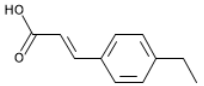
Trivially Simple Molecules:
Very easy to make
Very good in one objective

Assigning a capping value tends to filter out such molecules.
The capping value is assigned as the result when the actual result is “better”, because this value is “good enough”.

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

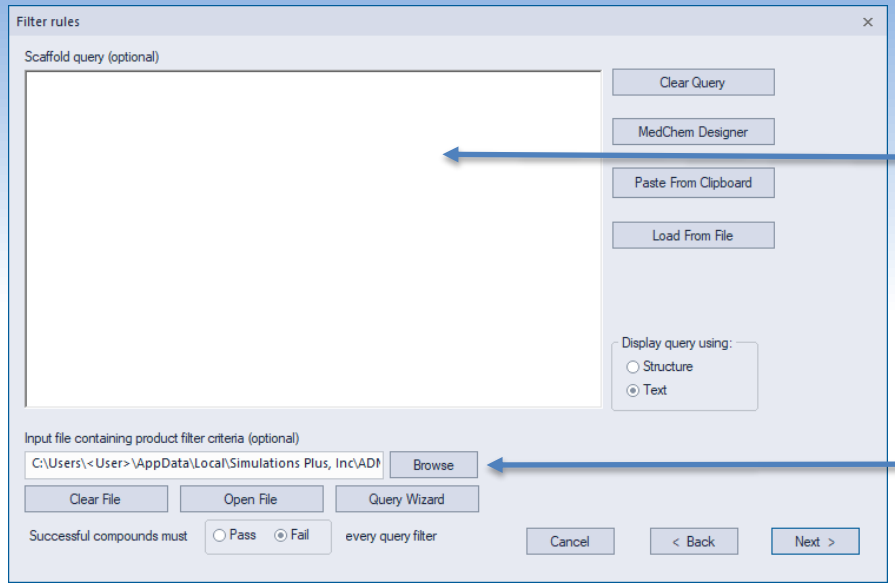
Applying Capping Values: Example

	Structure	OBJ_HIVI-ST	ADMET_Risk	SynthDiffCap
1		3.645	1.000	2.500
2		3.676	1.000	2.500
3		3.846	1.000	2.500
4		3.506	0.431	2.500
5		3.173	1.000	2.500
6		3.250	1.000	2.500

	Structure	OBJ_HIVI-ST	ADMET_Risk	SynthDiff+	SynthDiffCap
1		4.524	0.000	1.880	2.500
2		4.228	0.000	1.907	2.500
3		4.701	0.000	2.356	2.500

Molecules on right “dominate” molecules on left after applying capping to SynthDiff

Restricting Chemical Space



User Definable Scaffold

- Drawn using MedChem Designer
- Specified via SMARTS

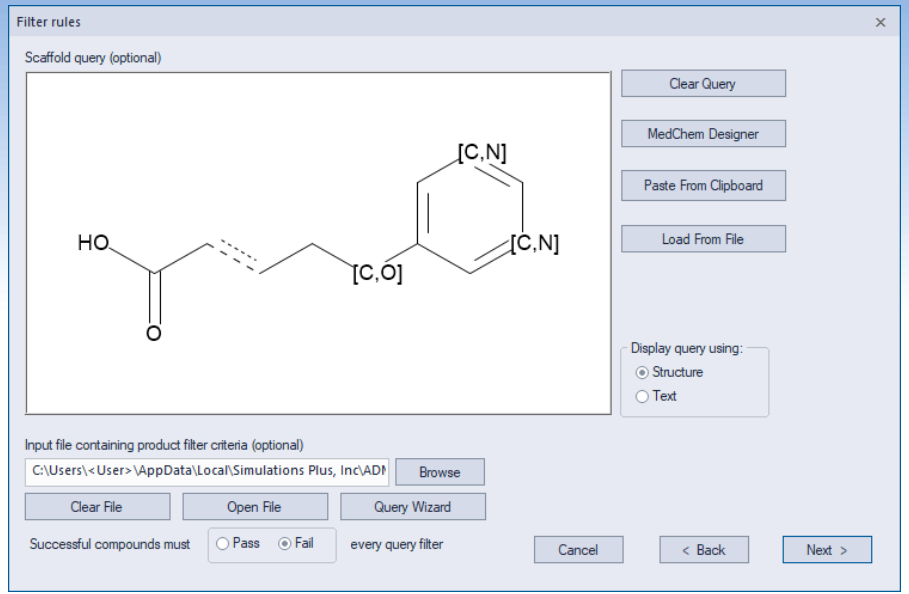
Filtering Criteria

- Default file supplied
- Can be modified/replaced or omitted
- SMARTS patterns or more general queries

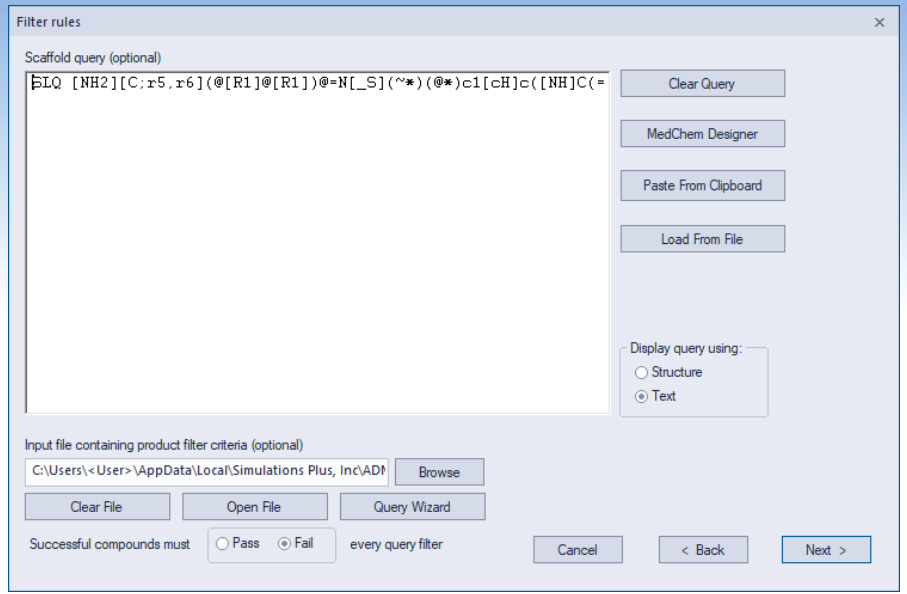
  SLQ [#6]C(=O)N U >= 4 SCORE 1

  NPQ AtomCount >= 66 SCORE 1

Scaffold Query Examples



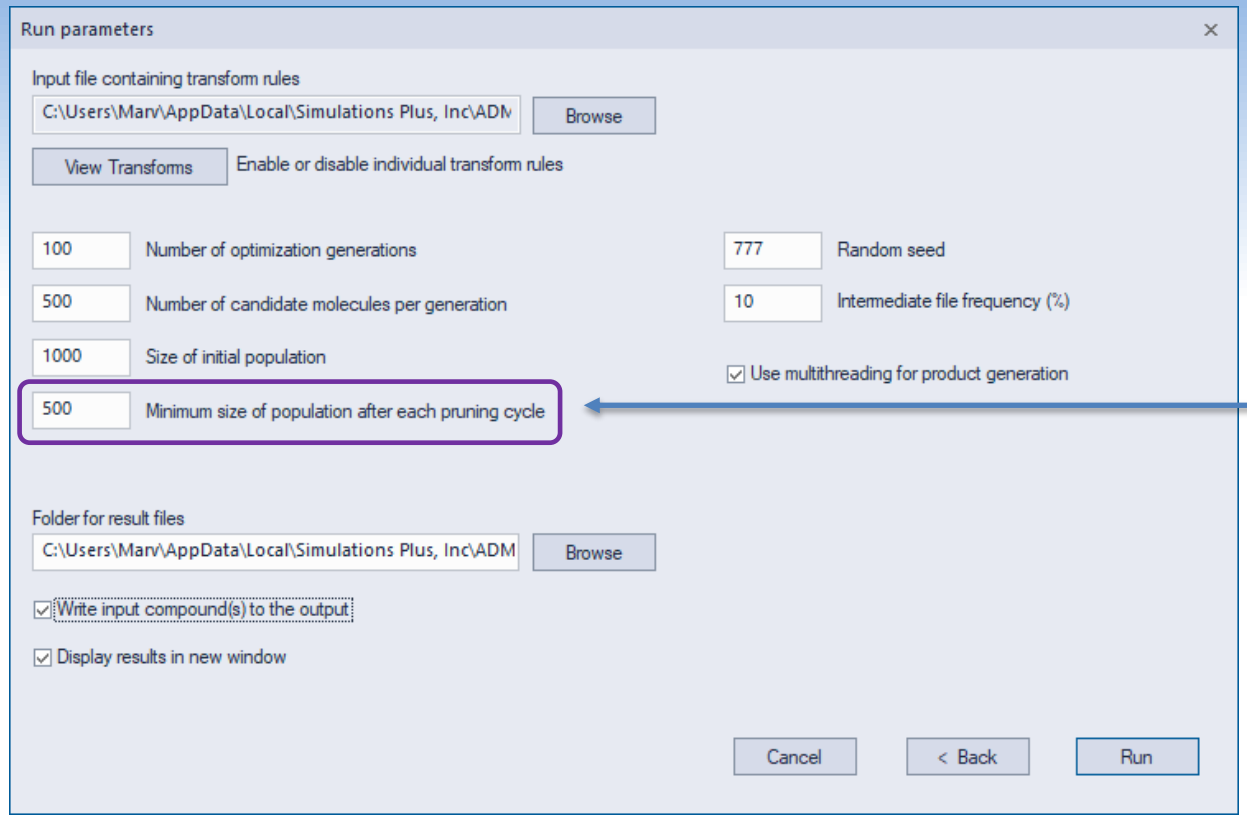
Structure-Based scaffold



SMARTS-Based scaffold

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

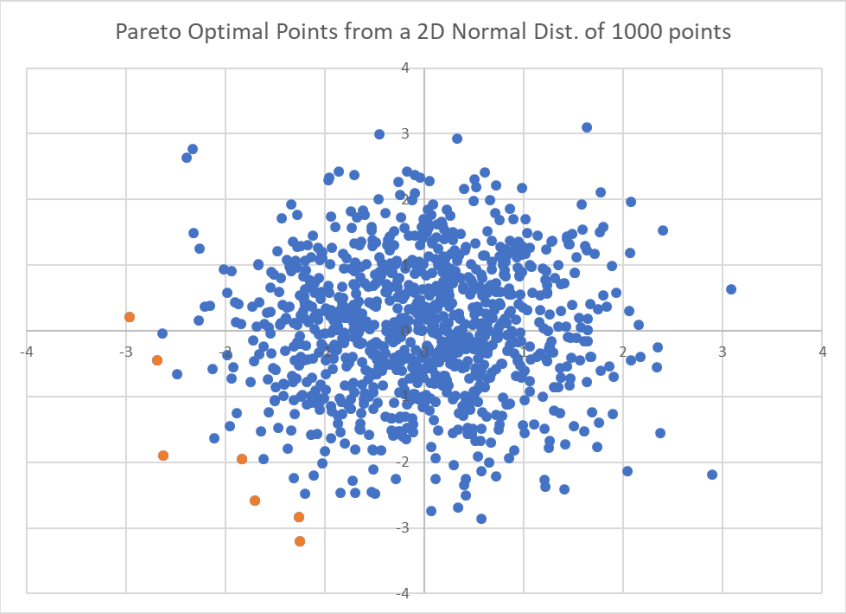
Run Parameters



Minimum Size?

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

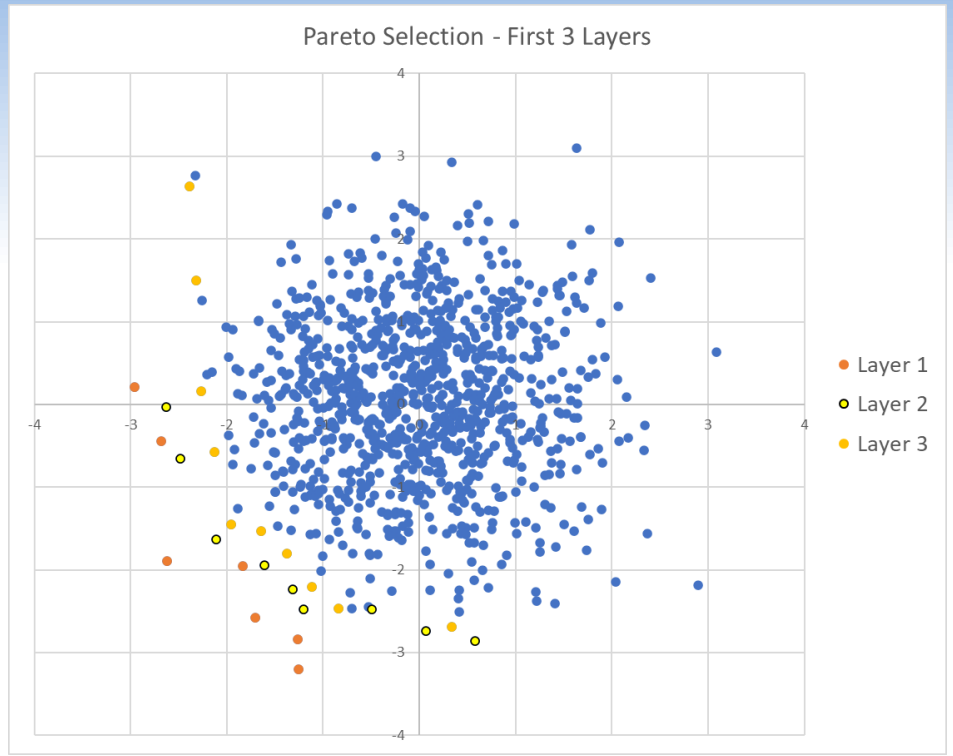
Pareto Selection



- Pareto Results
  - 1000 normally distributed points
    - 2-dim. : 7 Pareto optimal points

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Pareto Selection – Multiple Layers



If initial Pareto-based selection results in too few molecules, repeat the selection using the remaining candidate molecules => Pareto Layers.

Enforced after running 50% of generations. Avoids “bad” molecules entering the population too early.

Summary/Key Points

- Highly automated customizable protocol commercially available
- ~150 chemically intelligent transforms
  - Customizable, user-controllable
- Property objectives
  - ~50 built-in property models available including ADMET Risk to combine them
  - HTPK simulations
  - User models from ADMET Modeler™
- Many options for controlling/limiting chemical space to avoid “chemical nonsense”
  - User-specified scaffold definition
  - Chemical filters – built-in or user-specified
  - Out-of-scope penalties
  - Augmented synthetic difficulty
- Performance!
  - Up to ~10 million molecules evaluated per 24 hours on i7-8 core laptop running 7-8 threads

Evaluation

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Sept. 22, 2020 12:30 UTC

Simulations Plus Partners with Large Pharmaceutical Company to Validate AI-Driven Drug Design Capabilities in ADMET Predictor®

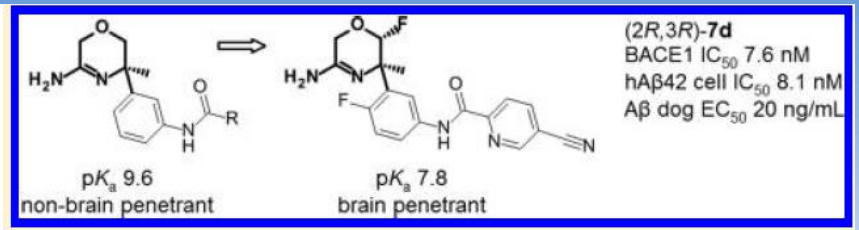
New AIDD Module applied to active therapeutic program to provide lead optimization support

LANCASTER, Calif.--(BUSINESS WIRE)-- Simulations Plus, Inc. (Nasdaq: SLP), the leading provider of modeling and simulation solutions for the pharmaceutical, biotechnology, chemicals, and consumer goods industries, today announced that it entered into a collaborative research agreement with a large pharmaceutical company to evaluate the new AIDD Module in the recently launched version of ADMET Predictor®.

Working Example/Demo

- Design BACE1 inhibitors starting from a known active “seed”
  - BACE1 inhibition has been investigated as a potential target for treating Alzheimer’s disease
- Use activity model based on 370 known BACE1 inhibitors
- Augment ADMET Risk with an additional Risk model blood-brain barrier penetration
- Specify scaffold and filtering criteria based on literature “pharmacophore”
- Illustrate scaffold hopping to find alternative scaffolds and compounds that were previously synthesized/tested

Journal of Medicinal Chemistry



1,4-Oxazine β-Secretase 1 (BACE1) Inhibitors: From Hit Generation to Orally Bioavailable Brain Penetrant Leads

Frederik J. R. Rombouts,*[,†] Gary Tresadern,[#] Oscar Delgado,[‡] Carolina Martínez-Lamenca,[†] Michiel Van Gool,[‡] Aránzazu García-Molina,[‡] Sergio A. Alonso de Diego,[‡] Daniel Oehlrich,[†] Hana Prokopcova,[†] José Manuel Alonso,[#] Nigel Austin,[∥] Herman Borghys,[∥] Sven Van Brandt,[†] Michel Surkyn,[†] Michel De Cleyn,[†] Ann Vos,[∥] Richard Alexander,[⊥] Gregor Macdonald,[†] Dieder Moechars,[§] Harrie Gijsen,[†] and Andrés A. Trabanco*[,‡]

[†]Neuroscience Medicinal Chemistry, Janssen Research & Development, Janssen Pharmaceutica NV, Turnhoutseweg 30, B-2340 Beerse, Belgium

[‡]Neuroscience Medicinal Chemistry, Janssen Research & Development, Janssen−Cilag SA, C/Jarama 75A, 45007 Toledo, Spain

[§]Neuroscience Biology, Janssen Research & Development, Janssen Pharmaceutica NV, Turnhoutseweg 30, B-2340 Beerse, Belgium

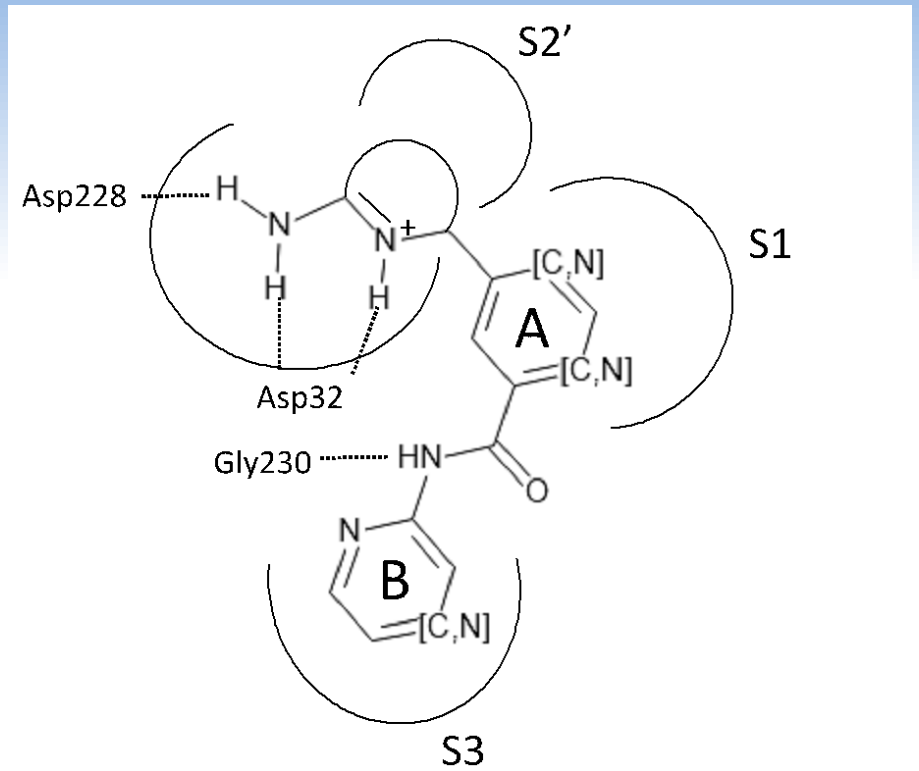
[∥]Discovery Sciences, Janssen Research & Development, Janssen Pharmaceutica NV, Turnhoutseweg 30, B-2340 Beerse, Belgium

[⊥]Biologics Research, Janssen Research & Development, 1400 McKean Road, Spring House, Pennsylvania 19477, United States

[#]Discovery Sciences, Janssen Research & Development, Janssen−Cilag SA, C/Jarama 75A, 45007 Toledo, Spain

DOI: 10.1021/acs.jmedchem.5b01101
J. Med. Chem. 2015, 58, 8216−8235

BACE1 Pharmacophore/Scaffold

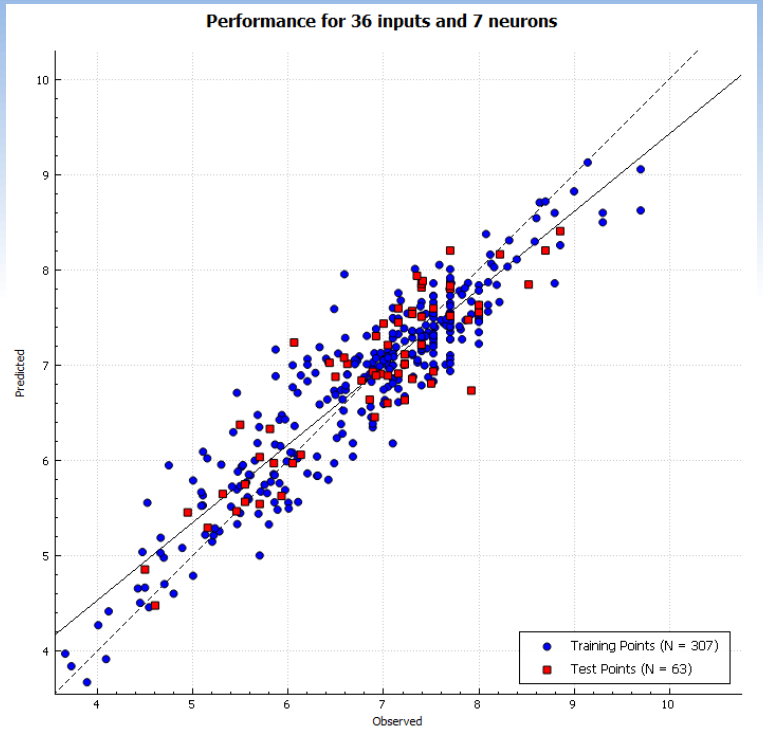


SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Customized Risk Rule for BBB

Identifier	Weight	Rule logic
BBB	1.0	BBB_Filter = Low OR LogBB < -2{-1}
Pgp	1.0	Pgp_Substr = Yes
Fcat	1.0	FCation > 0.5{0.975}

BACE1 Activity Model



ALL: Slope=0.816 Intercept=1.259 QSqd=0.851 RMSE=0.418 MAE=0.326 RMSU=0.424

TRAIN: Slope=0.817 Intercept=1.250 QSqd=0.857 RMSE=0.418 MAE=0.324 RMSU=0.423

TEST: Slope=0.812 Intercept=1.322 QSqd=0.812 RMSE=0.420 MAE=0.338 RMSU=0.430

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Acknowledgments

David Miller
Robert D. Clark
Michael Lawless

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

For more information, visit our website at:

www.simulations-plus.com

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS