Contributed data, collaboration and experience with ionization models – Part 1. Bayer Pharmaceuticals

7th of November 2023

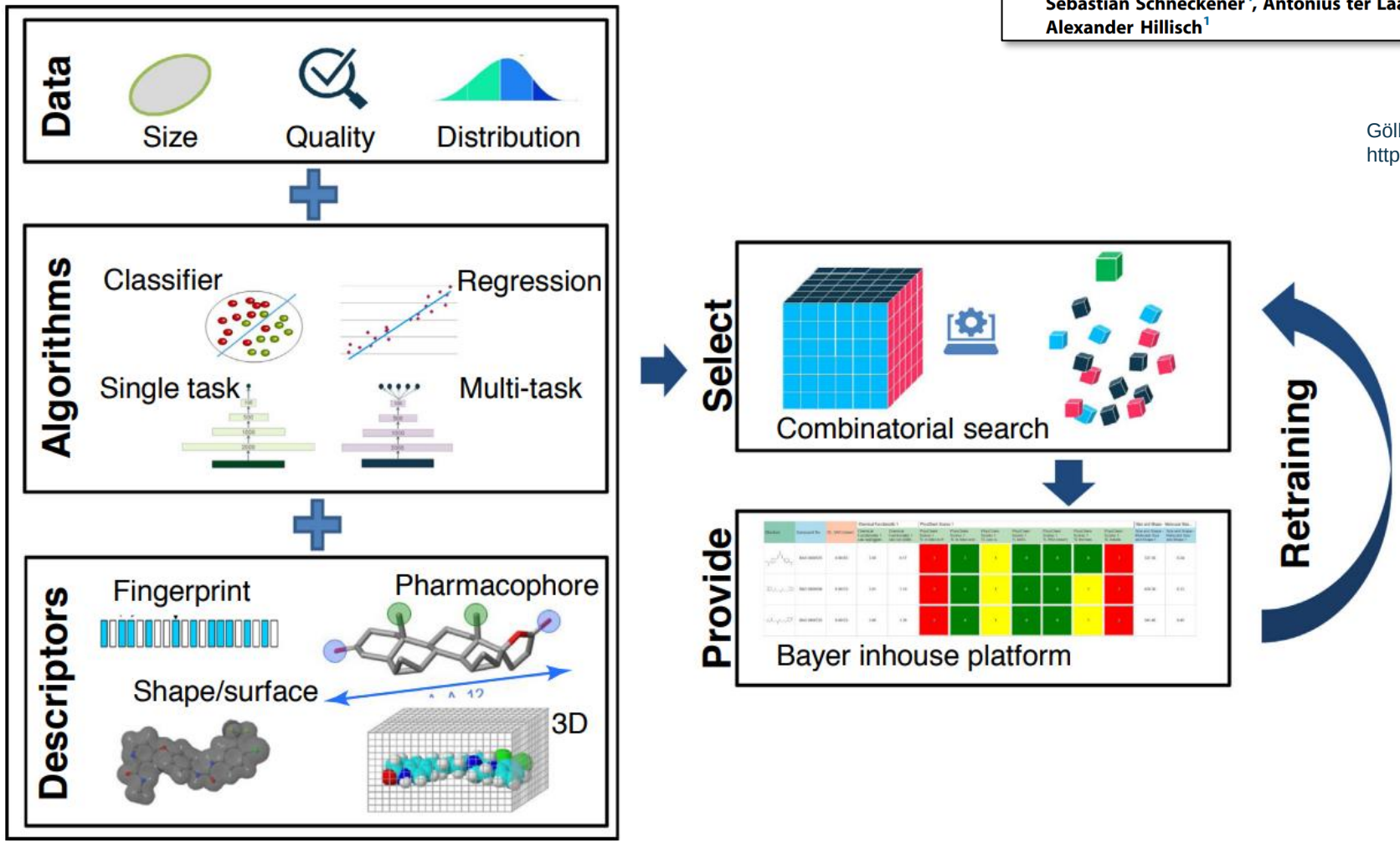
Dr. Antonius ter Laak
Dr. Judith Günther

Comp. Mol. Design Berlin
Bayer Pharma

ADMETpredictor in Bayer’s in silico Platform

Bayer Pharma in silico Platform

The Machine Learning Triade of Pharma



Reviews • INFORMATICS

Bayer's *in silico* ADMET platform: a journey of machine learning over the past two decades

Andreas H. Göller[1], Lara Kuhnke[2], Floriane Montanari[3], Anne Bonin[1], Sebastian Schneckener[4], Antonius ter Laak[2], Jörg Wichard[5], Mario Lobell[1] and Alexander Hillisch[1]

Göller et al. 2020, Drug Discorey Today
https://doi.org/10.1016/j.drudis.2020.07.001

Bayer Pharma in silico Platform

The Machine Learning Triade of Pharma

Göller et al. 2020, Drug Discovery Today
https://doi.org/10.1016/j.drudis.2020.07.001

Insufficient quality	First approach	Medium model	Good model	Robust model

Endpoint		Model type	Data set size	2005	2009	2014	2019	Retraining
Absorption	Caco-2 permeation	C (N)	>10 000			RF	SVR	Weekly
	Caco-2 efflux	C (N)	>10 000			RF	SVR	Weekly
	Bioavailability (rat)	C	~2000				RF	On demand
Distribution	Human serum albumin	N	>30 000			PLS	MTNN	On demand
	Fraction unbound	N	>1000			PLS	MTNN	On demand
Metabolism	Microsomal stability (hum)	C (N)	>10 000			RF	RF	Weekly
	Microsomal stability (mouse)	C (N)	>10 000			RF	RF	Weekly
	Microsomal stability (rat)	C (N)	>10 000			RF	RF	Weekly
	Hepatocyte stability (rat)	C (N)	>30 000			RF	RF	Weekly
Toxicity	hERG inhibition	C	>10 000			RF	SVM	Weekly
	Ames mutagenicity	C	>10 000			RF	RF	On demand
	CYP inhibition isoforms	C	>10 000			RF	RF	On demand
	Phospholipidosis	C	<1000			SVM	SVM	On demand
	Structure filter tool	Score	n.a.	-	-	-	-	On demand
PhysChem	Solubility (DMSO)	N	>30 ,000			PLS	MTNN	On demand
	Solubility (Powder)	N	<10 000				MTNN	On demand
	logD @ pH 7.5	N	>70 000			PLS	MTNN	On demand
	Membrane affinity	N	<10 000			PLS	MTNN	On demand
	pKa	N	>10 000			ANN	ANN	On demand
	Oral PhysChem score	Score	n.a.	-	-	-	-	On demand
	i.v. PhysChem score	Score	n.a.	-	-	-	-	On demand

Göller et al. 2020, Drug
https://doi.org/10.1016/j

How it all started

Data sharing in SimPlus Bayer Pharma collaboration in 2012 boosted performance in pKa prediction

Best of Both Worlds: Combining Pharma Data and State of the Art Modeling Technology To Improve *in Silico* pK_a Prediction

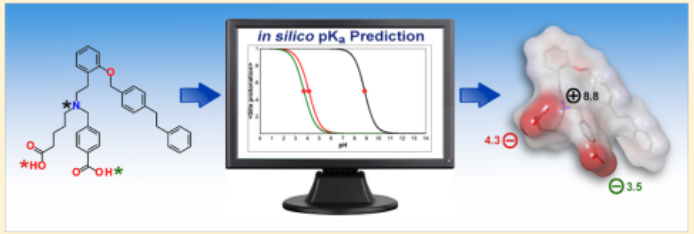
Robert Fraczkiewicz,*,† Mario Lobell,*,‡ Andreas H. Göller,‡ Ursula Krenz,‡ Rolf Schoenneis,‡ Robert D. Clark,† and Alexander Hillisch‡

†Simulations Plus, Inc. 42505 10th Street West, Lancaster, California 93534, United States

‡Global Drug Discovery, Bayer Pharma AG, Wuppertal, Germany

Supporting Information

ABSTRACT: In a unique collaboration between a software company and a pharmaceutical company, we were able to develop a new *in silico* pK_a prediction tool with outstanding prediction quality. An existing pK_a prediction method from Simulations Plus based on artificial neural network ensembles (ANNE), microstates analysis, and literature data was retrained with a large homogeneous data set of drug-like molecules from Bayer. The new model was thus built with curated sets of ~14,000 literature pK_a values (~11,000 compounds, representing literature chemical space) and ~19,500 pK_a values experimentally determined at Bayer Pharma (~16,000 compounds, representing industry chemical space). Model validation was performed with several test sets consisting of a total of ~31,000 new pK_a values measured at Bayer. For the largest and most difficult test set with >16,000 pK_a values that were not used for training, the original model achieved a mean absolute error (MAE) of 0.72, root-mean-square error (RMSE) of 0.94, and squared correlation coefficient (R^2) of 0.87. The new model achieves significantly improved prediction statistics, with MAE = 0.50, RMSE = 0.67, and R^2 = 0.93. It is commercially available as part of the Simulations Plus ADMET Predictor release 7.0. Good predictions are only of value when delivered effectively to those who can use them. The new pK_a prediction model has been integrated into Pipeline Pilot and the PharmacophorInformatics (PIx) platform used by scientists at Bayer Pharma. Different output formats allow customized application by medicinal chemists, physical chemists, and computational chemists.



Training set:

	compounds	pKa values
▪ Literature	~11.000	~14.000
▪ Bayer Pharma	~16.000	~19.500

* R. Fraczkiewicz,, M. Lobell, et al.: *J. Chem. Inf. Model*. **2015**, *55*, 389−397
DOI: 10.1021/ci500585w

World‘s best in class pK_a prediction tool

* R. Fraczkiewicz,, M. Lobell, et al.: *J. Chem. Inf. Model*. **2015**, *55*, 389−397
# https://github.com/MobleyLab/SAMPL6/tree/master/physical_properties/pKa/analysis/analysis_of_typeI_predictions

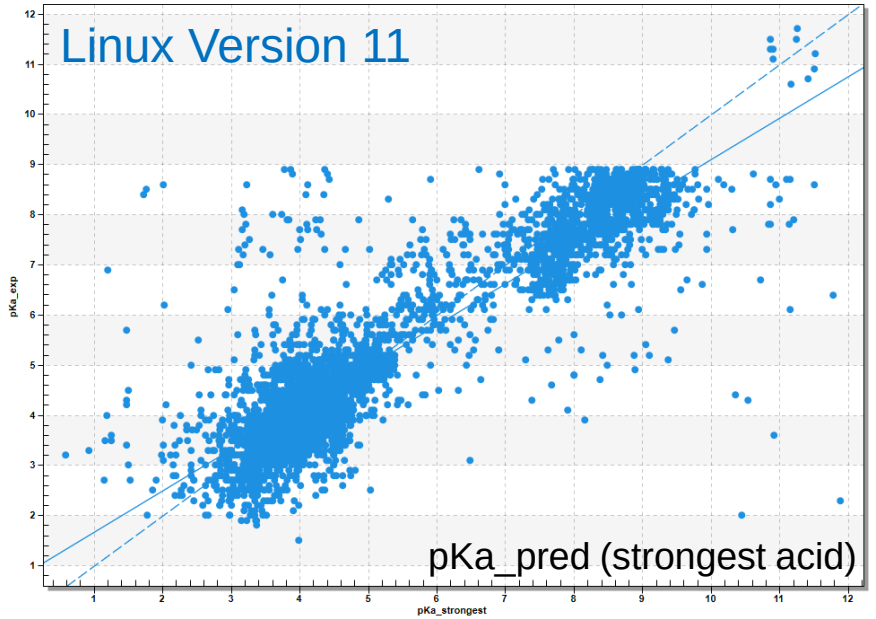
- In 2012 we joined forces with the company Simulations Plus with the aim to jointly develop the world‘s best in class pK_a prediction tool
- In 2013 the new pK_a prediction tool was rolled out at Bayer and became also available world-wide via its integration into the ADMET Predictor software sold and distributed by Simulations Plus
- The tool showed superb predictivity in our internal validation with ~13K new compounds (MUE=0.50, R^2=0.93)*
- In May 2018 the pK_a tool won the SAMPL6 pK_a prediction challenge as best of 32 participants#
- pK_a prediction can help to select and prioritize the right compounds for synthesis
- This is even more so the case for ionizable groups for which pK_a transitions cannot be detected experimentally

Validation of pKa prediction at Bayer Pharma

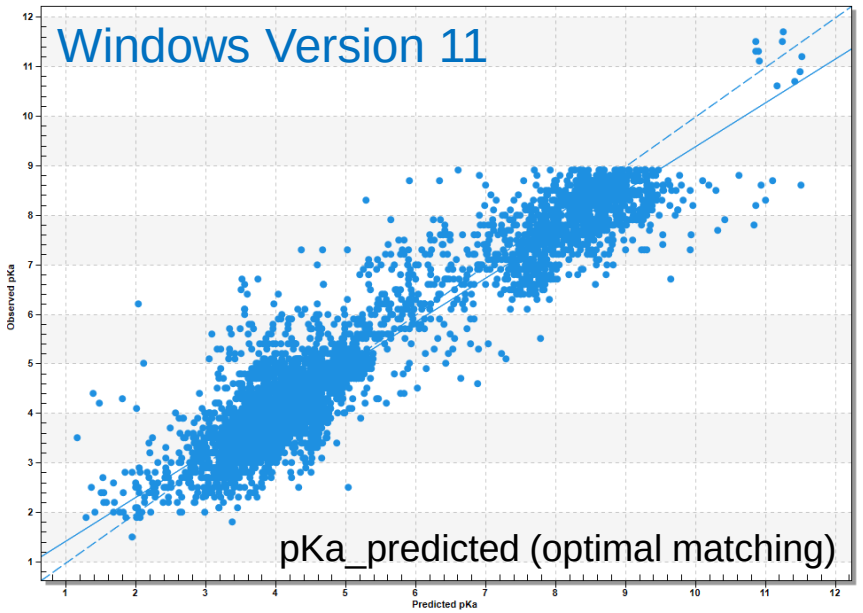
* R. Fraczkiewicz,, M. Lobell, et al.: *J. Chem. Inf. Model*. **2015**, *55*, 389−397
DOI: 10.1021/ci500585w

Test Set 2
Acids only

R^2 = 0,803; Linear Regression RMSE = 0,827 and MAE = 0,545
Slope = 0,826; Intercept = 0,845
N = 4397; RMSE = 0,902; MAE = 0,575

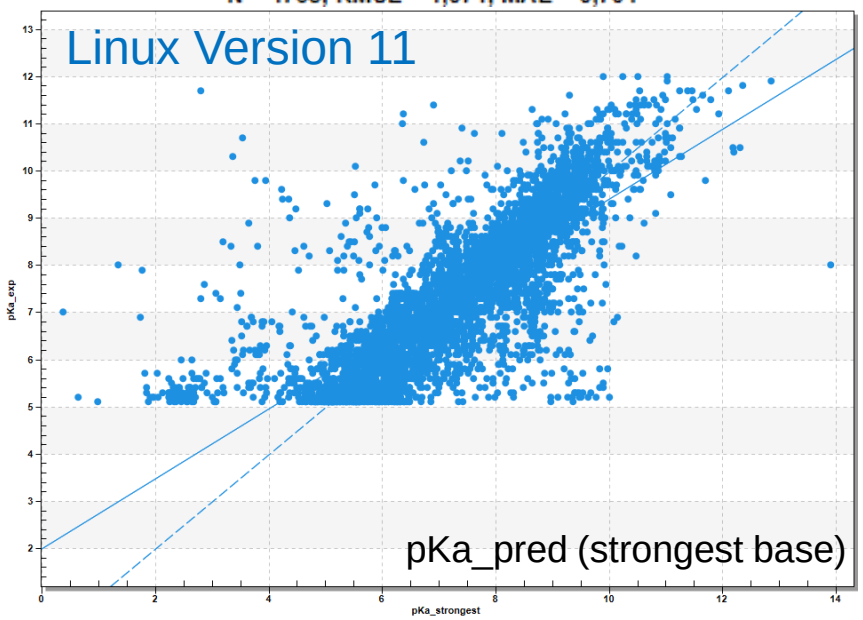


R^2 = 0,895; Linear Regression RMSE = 0,603 and MAE = 0,442
Slope = 0,885; Intercept = 0,541
N = 4402; RMSE = 0,649; MAE = 0,471

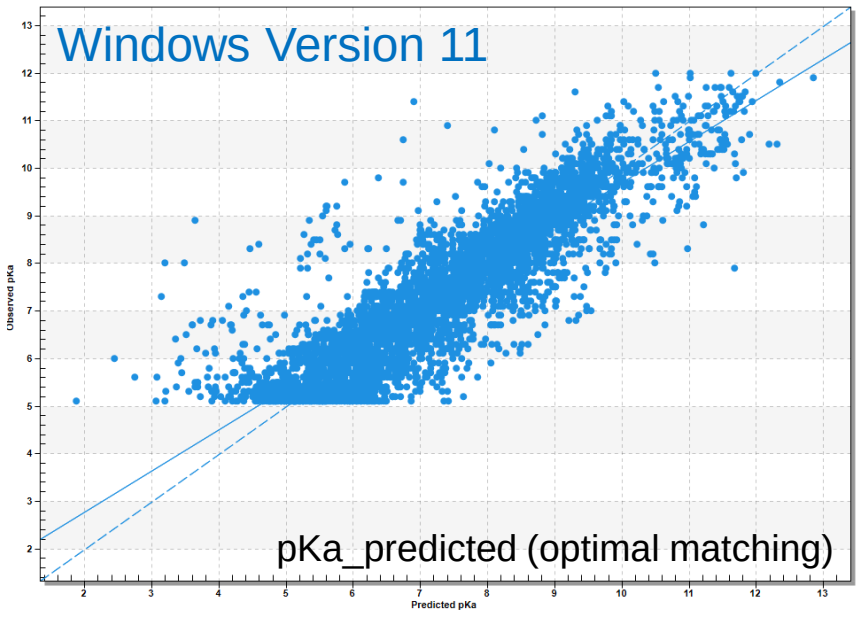


Test Set 2
Bases only

R^2 = 0,610; Linear Regression RMSE = 0,980 and MAE = 0,722
Slope = 0,740; Intercept = 2,005
N = 4765; RMSE = 1,074; MAE = 0,704



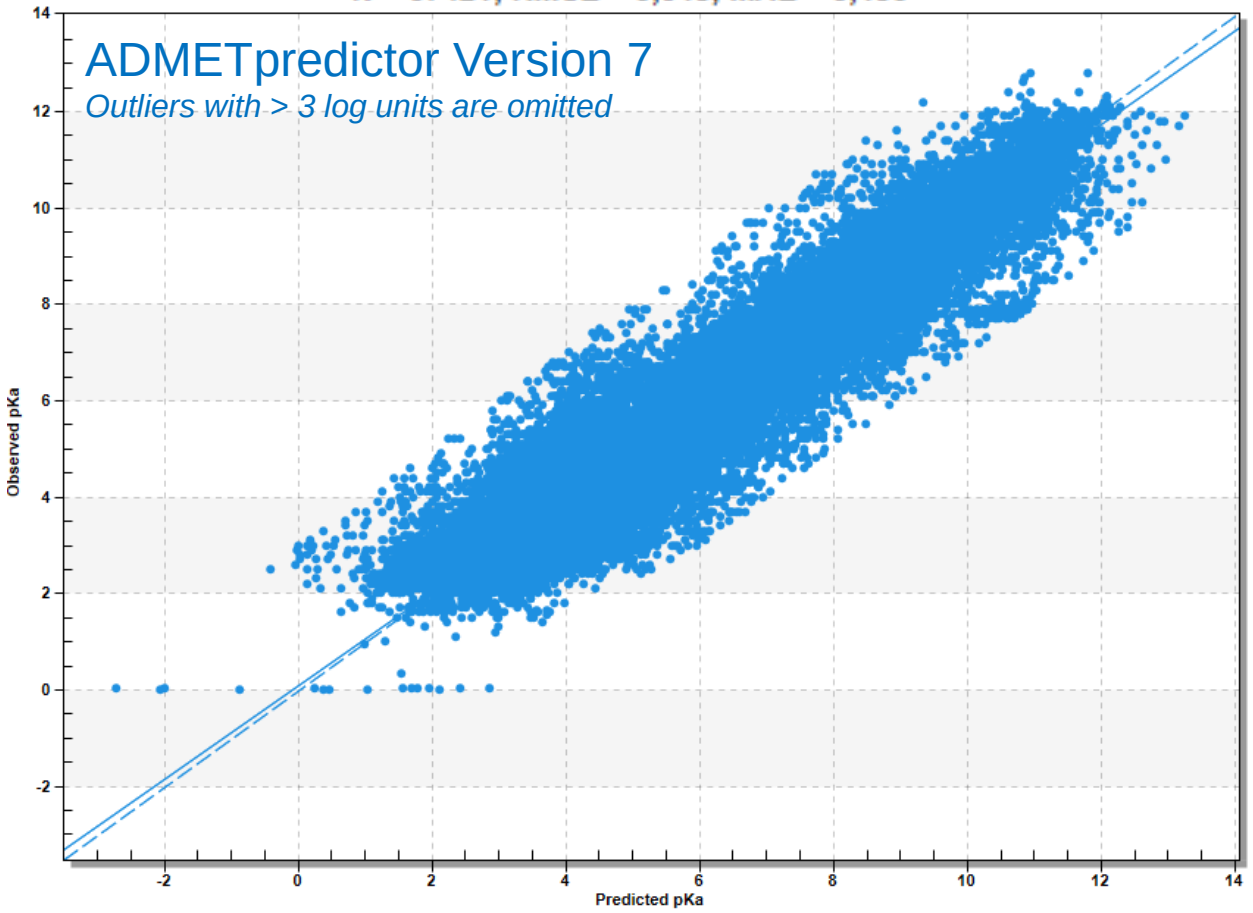
R^2 = 0,787; Linear Regression RMSE = 0,725 and MAE = 0,542
Slope = 0,866; Intercept = 1,042
N = 4766; RMSE = 0,758; MAE = 0,546



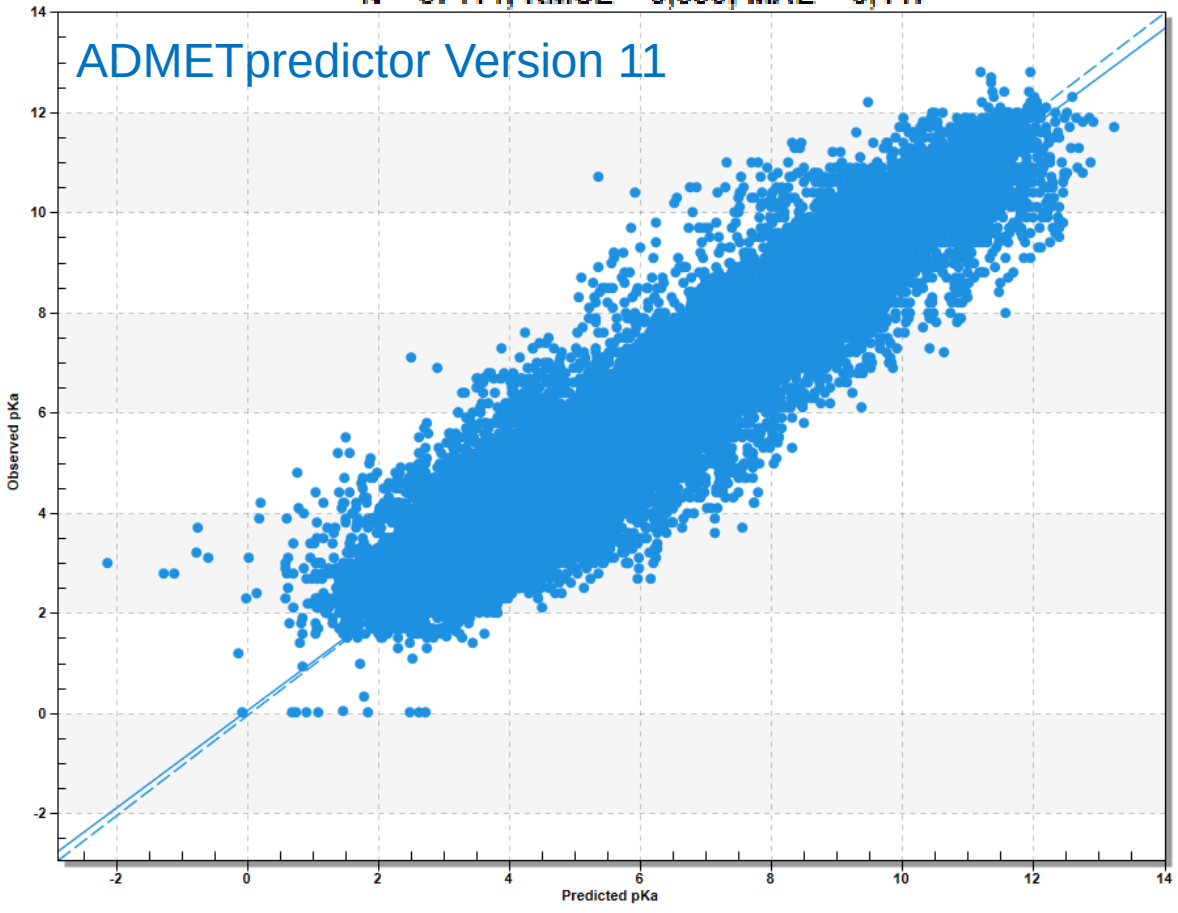
SimPlus Validation with largest external test set (N=67421)

All Bayer Pharma pKa values except training sets 1-4

R^2 = 0,942; Linear Regression RMSE = 0,639 and MAE = 0,458
Slope = 0,971; Intercept = 0,088
N = 67421; RMSE = 0,648; MAE = 0,456



R^2 = 0,945; Linear Regression RMSE = 0,620 and MAE = 0,446
Slope = 0,971; Intercept = 0,086
N = 67414; RMSE = 0,630; MAE = 0,447

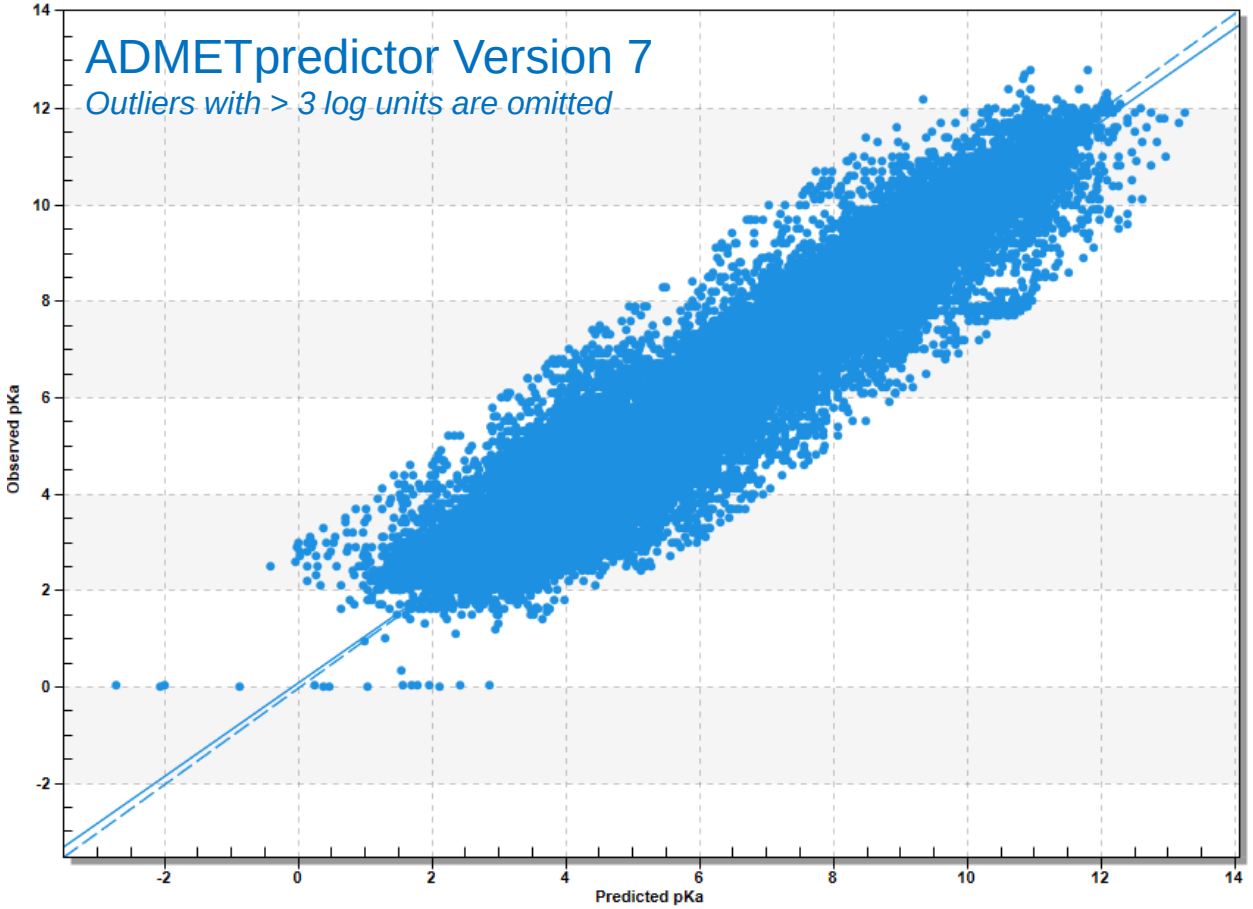


*) 510 pKa values were predicted > 3 log units in V10.4 (‚Version 7') and are taken out of the validation

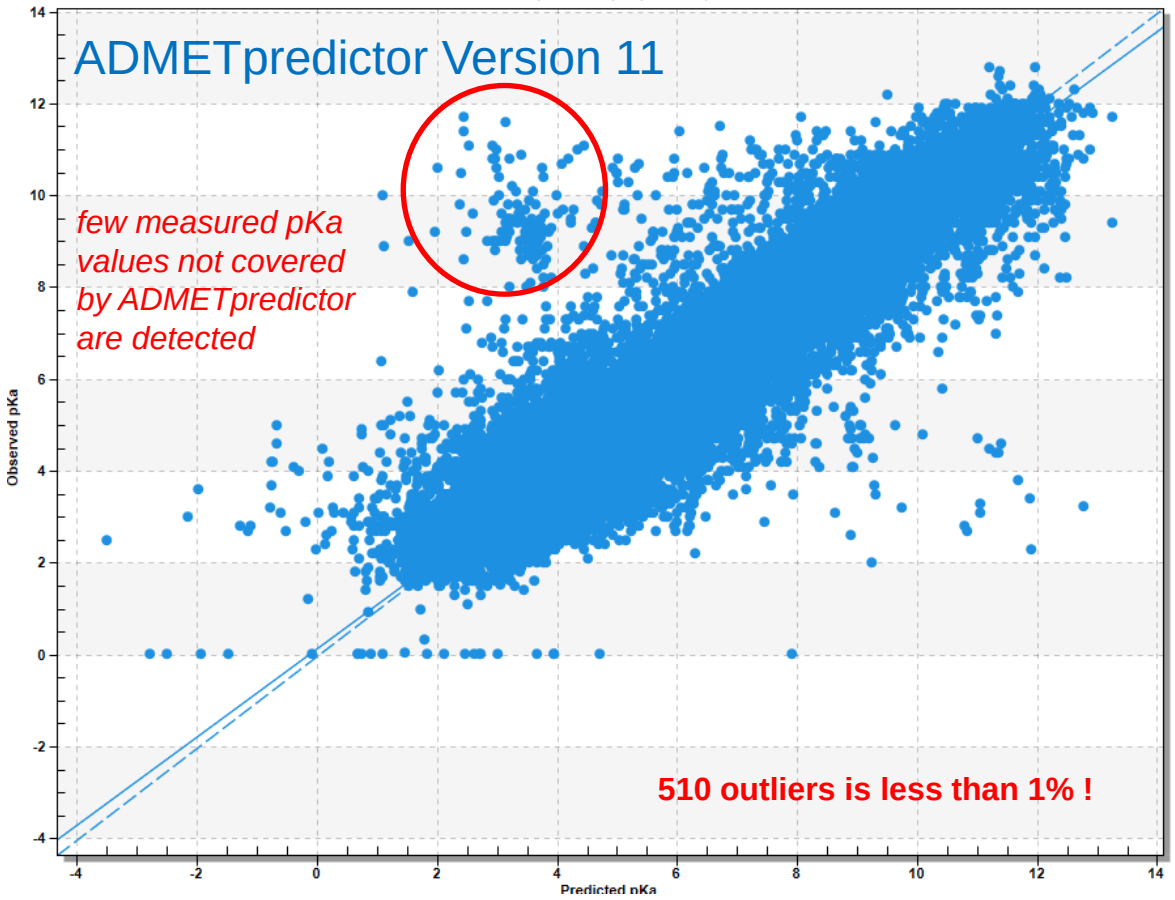
SimPlus Validation with largest external test set (N=67927)

All Bayer Pharma pKa values except training sets 1-4

R^2 = 0,942; Linear Regression RMSE = 0,639 and MAE = 0,458
Slope = 0,971; Intercept = 0,088
N = 67421; RMSE = 0,648; MAE = 0,456



R^2 = 0,925; Linear Regression RMSE = 0,727 and MAE = 0,476
Slope = 0,960; Intercept = 0,160
N = 67927; RMSE = 0,738; MAE = 0,474



*) 510 pKa values were predicted > 3 log units in V10.4 (‚Version 7') and are taken out of the validation

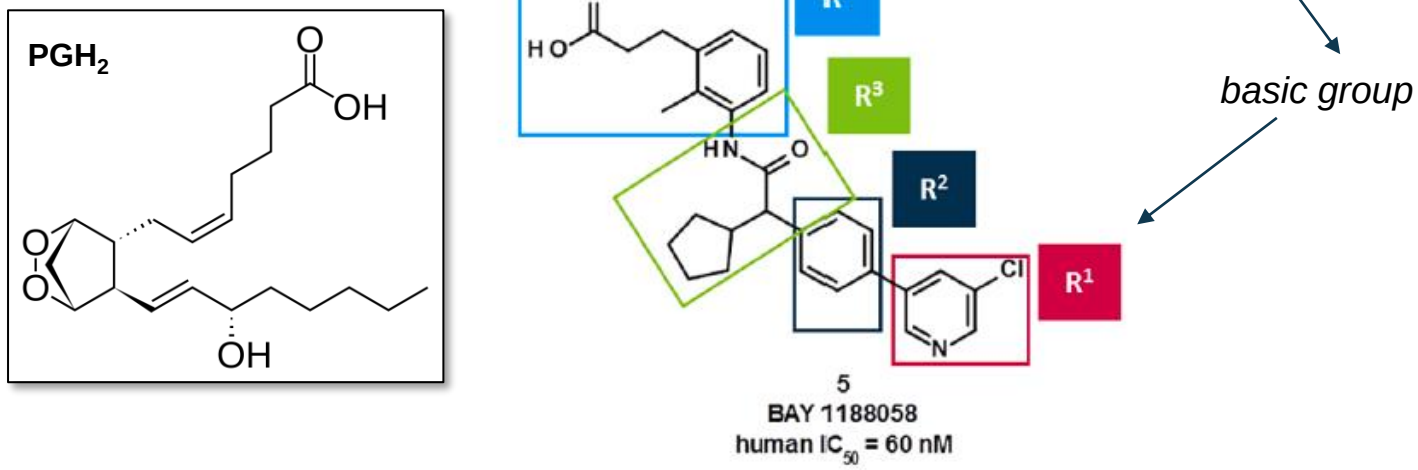
Validation of pKa prediction at Bayer Pharma

- Windows version of ADMETpredictor has an optimal matching algorithm, Linux not
- ADMETpredictor Version 7 removed outliers (>3 log units)
- ADMETpredictor Version 11 has the advantage that it has no need to remove outliers
- In addition, Version 11 appears to predict Bayer‘s pKa values marginally better than Version 7
- ADMETpredictor Version 11 has a great prediction quality on a huge external test set of N=67.000
- Bayer Pharma chemical space is very well covered (few exceptions, work to be continued)

Example 1: mPGES-1

Example pKa prediction in Drug Discovery

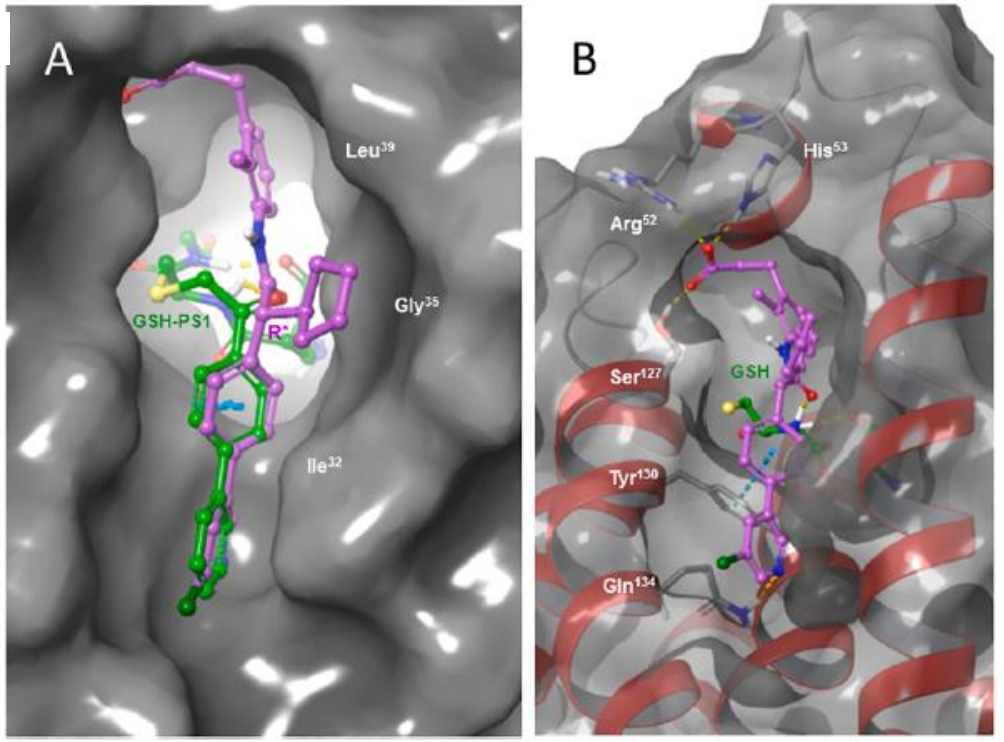
Optimisation of mPGES-1 inhibitor



Contents lists available at ScienceDirect
Bioorganic & Medicinal Chemistry Letters
journal homepage: www.elsevier.com/locate/bmcl

Discovery and optimization of pyridyl-cycloalkyl-carboxylic acids as inhibitors of microsomal prostaglandin E synthase-1 for the treatment of endometriosis

Marcus Koppitz[a,*], Nico Bräuer[a], Antonius Ter Laak[a], Horst Irlbacher[a], Andrea Rotgeri[a], Anne-Marie Coelho[b], Daryl Walter[c], Andreas Steinmeyer[a], Thomas M. Zollner[a], Michaele Peters[a], Jens Nagel[a]



Microsomal prostglandin E synthase inhibitors

→ mPGES1 = Transmembrane enzyme

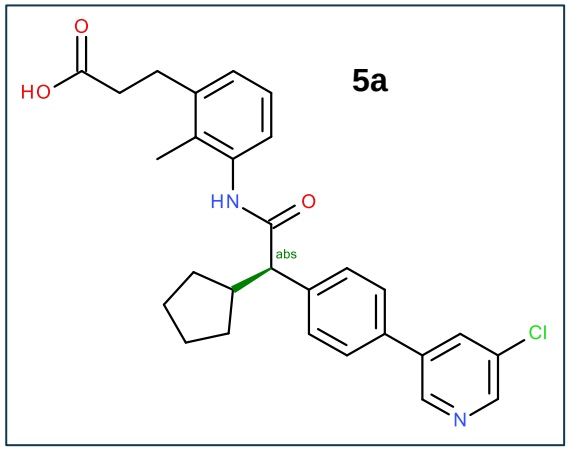
→ Competitive inhibitors for acidic natural substrate Prostaglandin H2

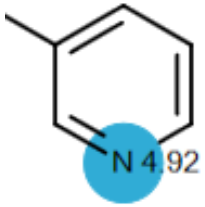
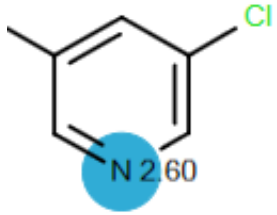
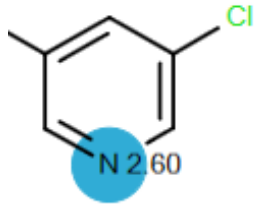
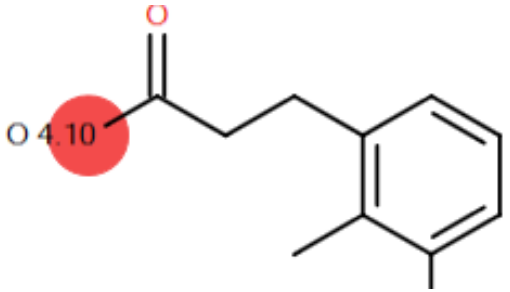
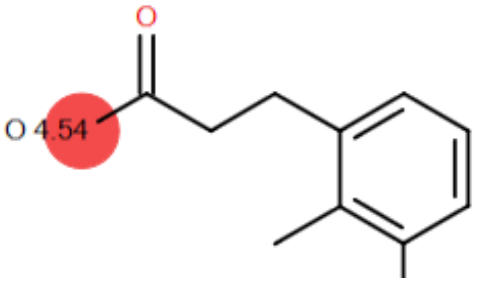
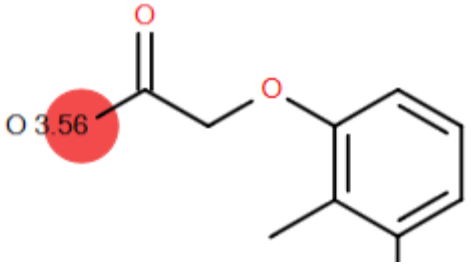
→ Variation of R1, R2, R3, R4 decribed (Koppitz *et al.*, 2019)

Koppitz *et al.* 2019
https://doi.org/10.1016/j.bmcl.2019.07.007

Example pKa prediction in Drug Discovery

Optimisation of mPGES-1 inhibitor



	4a		5a		28a
R1		chlorine in pyridyl reduces the basicity			
R4				ether oxygen increases the acidity	

pKa values predicted with latest ADMETpredictor V.11

Koppitz *et al.* 2019
https://doi.org/10.1016/j.bmcl.2019.07.007

Example pKa prediction in Drug Discovery

Optimisation of mPGES-1 inhibitor

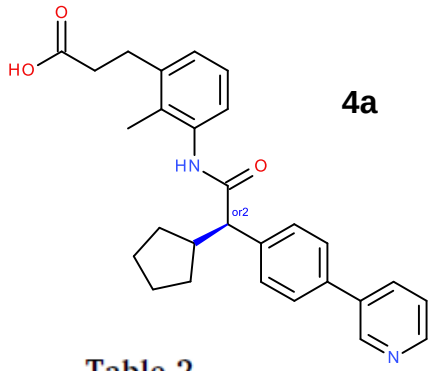
Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

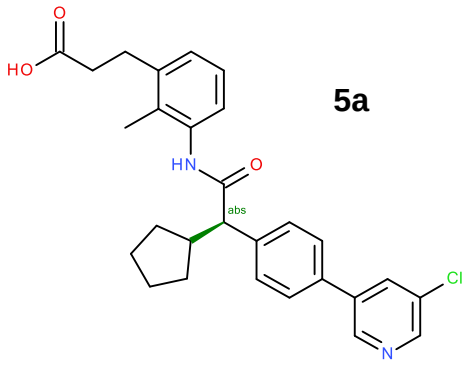
journal homepage: www.elsevier.com/locate/bmcl

Discovery and optimization of pyridyl-cycloalkyl-carboxylic acids as inhibitors of microsomal prostaglandin E synthase-1 for the treatment of endometriosis

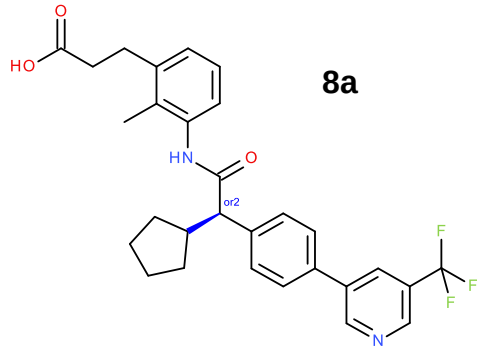
Marcus Koppitz[a,*], Nico Bräuer[a], Antonius Ter Laak[a], Horst Irlbacher[a], Andrea Rotgeri[a], Anne-Marie Coelho[b], Daryl Walter[c], Andreas Steinmeyer[a], Thomas M. Zollner[a], Michaele Peters[a], Jens Nagel[a]



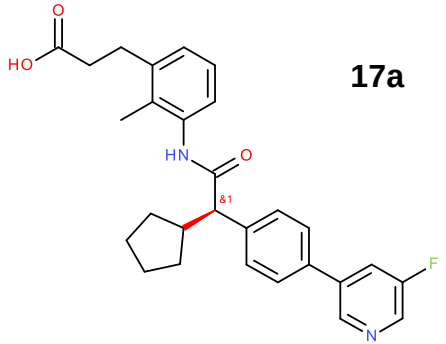
4a



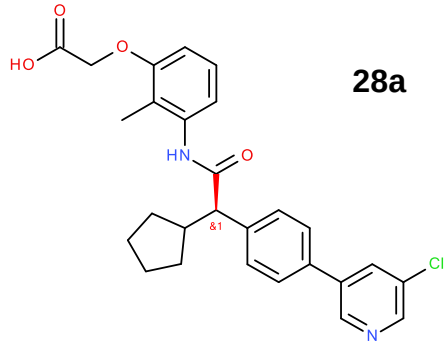
5a



8a



17a



28a

Table 2

PK properties for selected active (-) enantiomers.

Cmpd	h PTGESIC$_{50}$ [nM]	E_H rat hepatocytes	Caco-2 Papp A-B Mari [nm/s]	Caco-2 efflux ratio	Delta pKa calc.[20]	F [%] male rat
4a	270	0.15	36	12	0.8	–
5a	35	0.02	69	2.9	1.3	98
8a	20	0.07	80	1.2	2.6	73
17a	18	0.24	59	1.4	1.7	72
28a	14	0.01	2.2	69	0.7	8

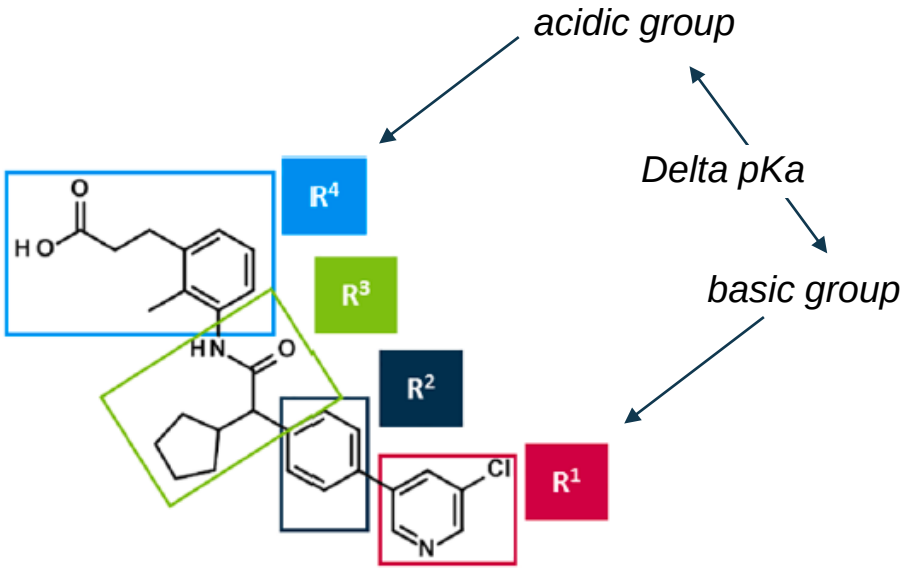
- Trend in this project: large ‚delta pKa calc' reduces efflux and increases bioavailability F%

Koppitz *et al.* 2019
https://doi.org/10.1016/j.bmcl.2019.07.007

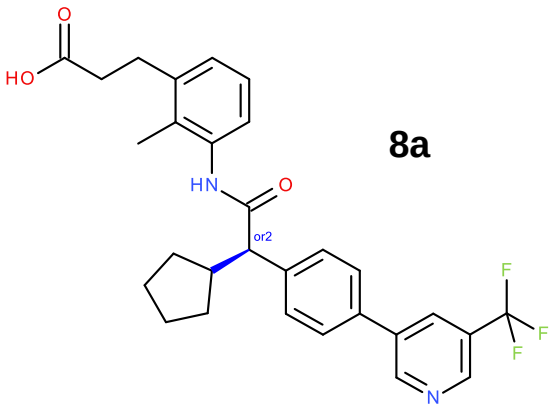
Example pKa prediction in Drug Discovery

Optimisation of mPGES-1 inhibitor

Koppitz *et al.* 2019
https://doi.org/10.1016/j.bmcl.2019.07.007

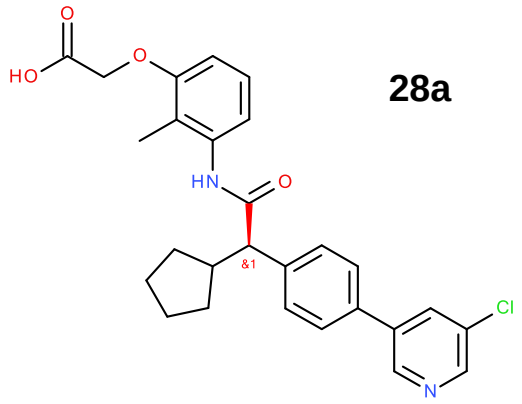


5
BAY 1188058
human IC_{50} = 60 nM



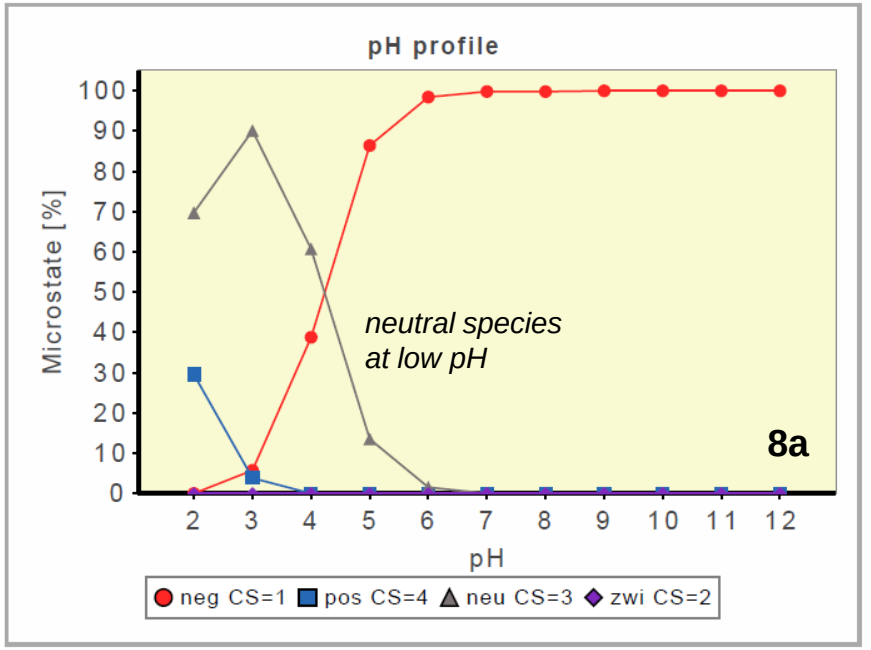
8a

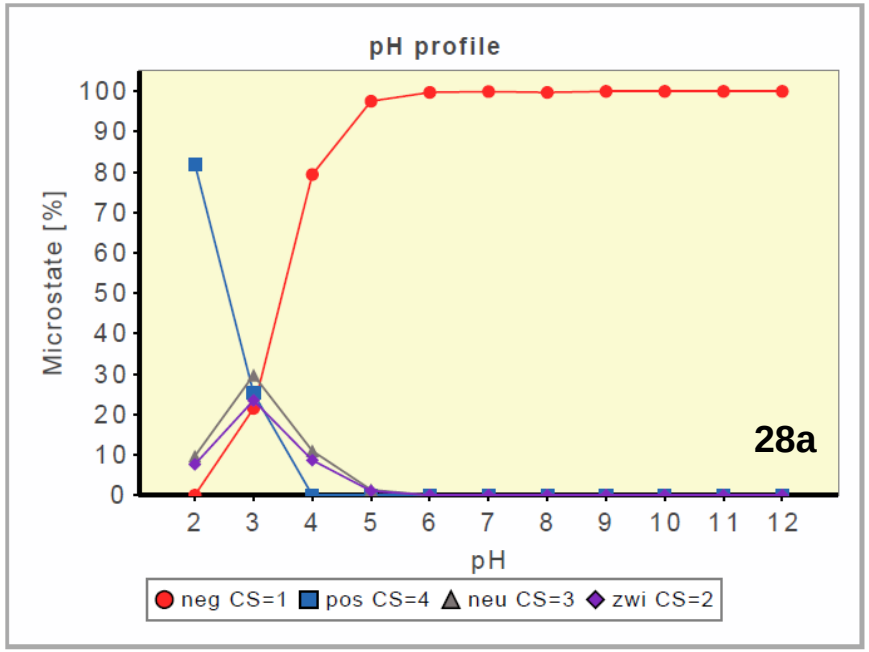
Delta pKa calc = 2.6



28a

Delta pKa calc = 0.7

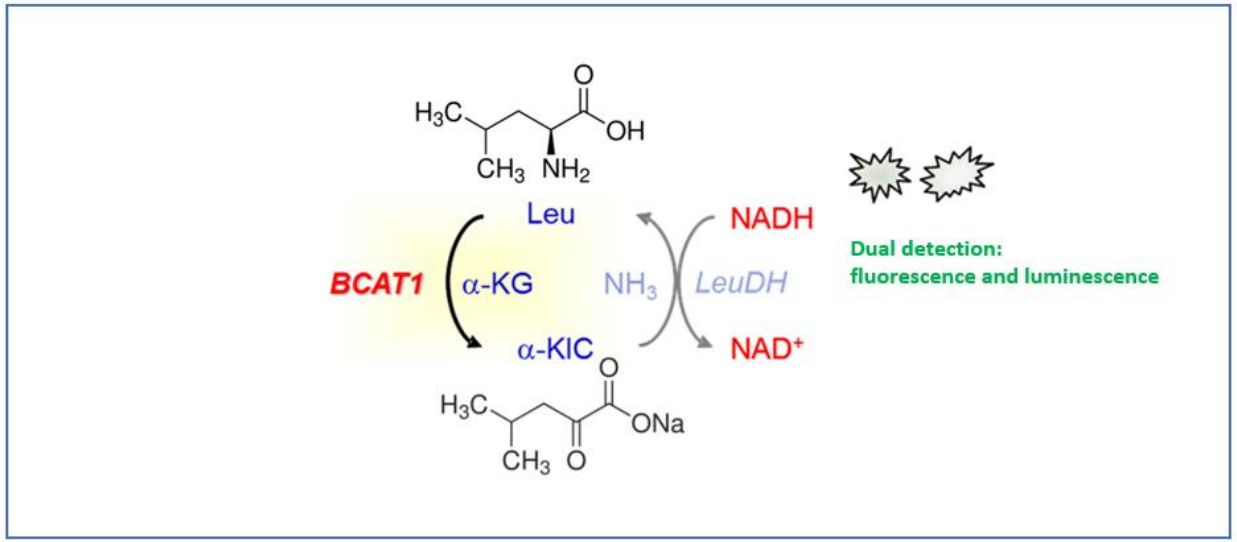


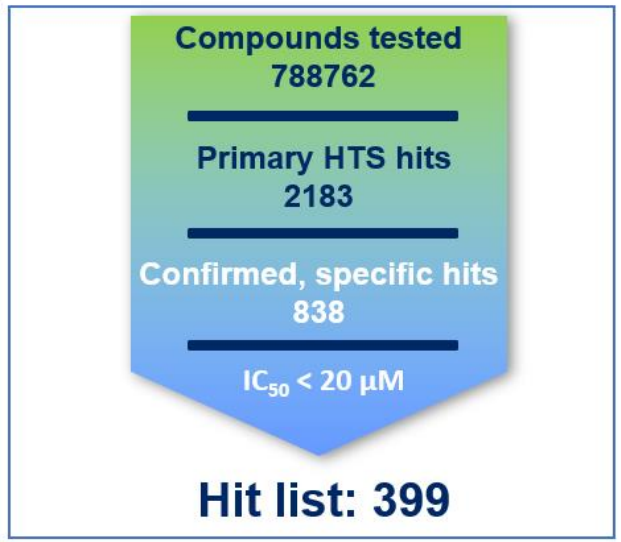


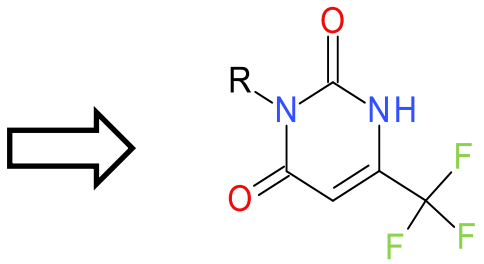
Example 2: BCAT

The BCAT1 pyrimidinedione series – using a CropScience core

WO 9531440 A1, 1994



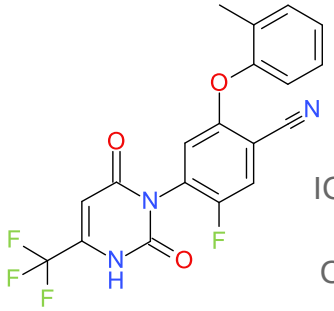




- Branched Chain Amino acids transferases (BCATs) catalyse the catabolism of the essential amino acids **valine, leucin and isoleucine**
- BCAT's preference for binding negatively charged compounds reflected in structures of obtained hit series
- **Anticipated challenge in hit progression: balancing acidity (as driver for target potency) with suitable PK properties**

→ Cluster prioritization favoured a structural series with acidity easily tunable by chemistry

The development of probe BAY-069



HTS hit

IC_{50} (BCAT1) = 554 nM

pK_a = 4.4

CaCo efflux ratio = 6.2

P_{app} A-B = 30 nm/s

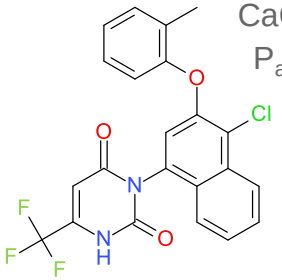
Final probe for OpenScience

IC_{50} (BCAT1) = 31 nM

pK_a = 5.7

CaCo efflux ratio = 0.5

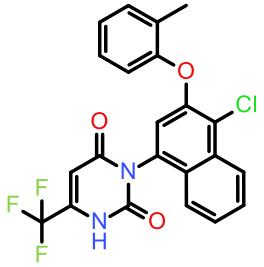
P_{app} A-B = 252 nm/s



Target-related criteria

- **Potent**
  - *BCAT1* IC_{50} = 31 nM
  - *BCAT2* IC_{50} = 153 nM
- **Cellularly active**
  - *MDA-MB-231* IC_{50} = 875 nM
  - *U-87-MG* IC_{50} = 358 nM
- **Selective vs. *in house* protease and transaminase panel**
- **Target engagement shown** *by X-ray crystallo-graphy*

BCAT1/2



BAY-069

Chemical matter criteria

- ***Tractable Chemical Matter*** Stringent SAR established
- ***Inactive control:*** BAY-771
- ***Chemical & Plasma*** Stable
- ***Standard formulations for in vitro***

SCAN ME

https://doi.org/10.1021/acs.jmedchem.2c00441

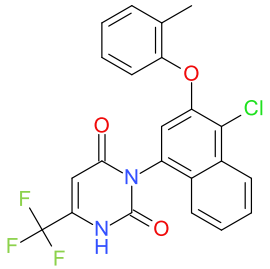
Successful usage in BCAT1 project (2016/17)

// Dealing with different tautomeric states inherent to chemical series

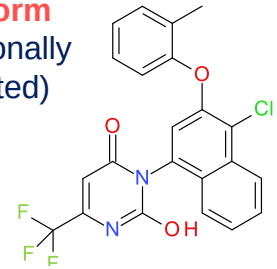
SimPlus tool recognizes core acidity as well as substituent positions that modulate pK_a
→ trends correctly reproduced by tool on chemist‘s desk

Output of pK_a predictor depends on input tautomer

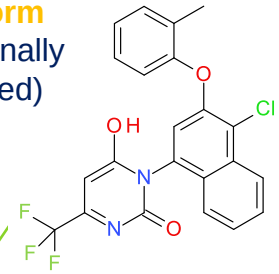
Keto form (used in warehouse registration)



Enole1 form (computationally enumerated)

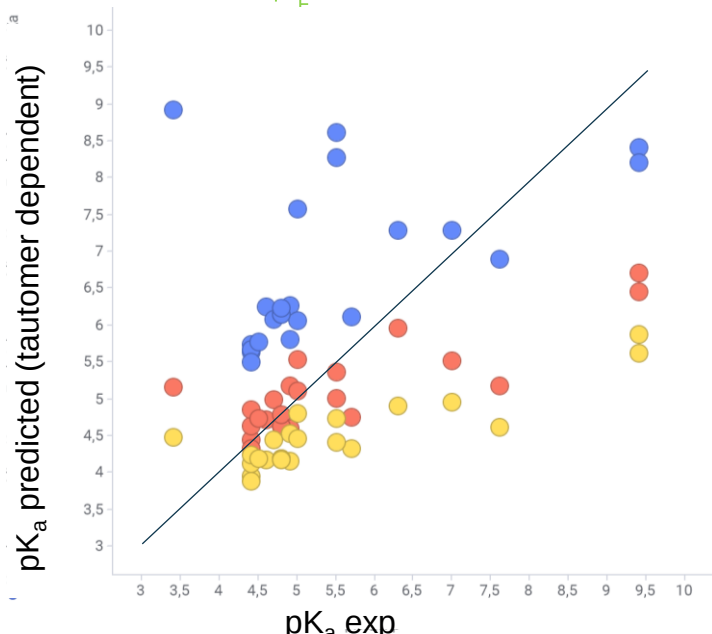


Enole2 form (computationally enumerated)



Highest correlation coefficient R^2=0.76 obtained with enole 2 form

Average $\Delta(pK_a)$ between matching tautomer forms: 1.5



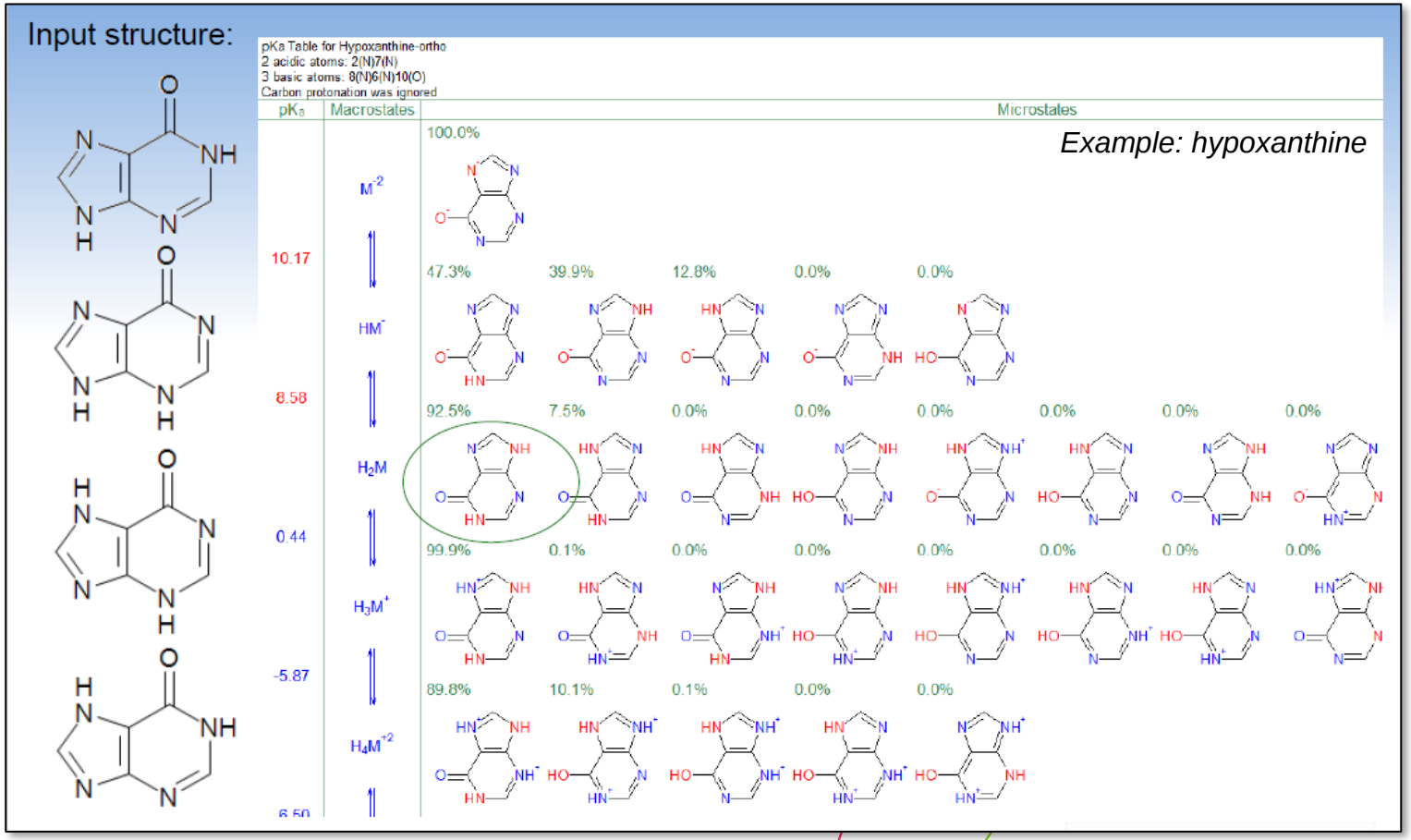
Critical optimization parameters (1): pK_a

// Progress in pK_a prediction: implementation aspects

Tautomer-independent pK_a prediction tool → average $\Delta(pK_a)$ between matching tautomer forms = 0.0

under development @SimPlus

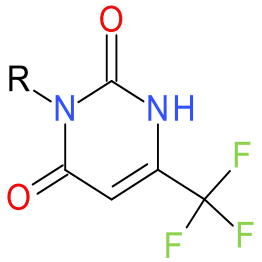
(Robert Fraczkiewicz)



BCAT1 series revisited with latest pK_a predictor

// Benefit from training on larger chemical space

// As a „side effect“ of training version 11 on a much broader chemical space, the dependency of the predicted pKa on the input tautomer is decreased in the BCAT1 pyrimidine dione series



Name	Δ(pK$_a$) tautomers
ADMETpred 9_7	1,5
ADMETpred 11	1,0

// Explanation: since the expanded training set did not contain any 2,4-dione substituted pyrimidines in enolic form, the model (using atom-based descriptors!) likely learnt indirectly from other heterocycles with keto-enol-tautomerism

// Wish for a long term solution: tautomer-independent prediction tool (development by SimulationPlus)

Thank you for your attention!

Acknowledgements

Dr. Judith Günther
Dr. Anke Hackl
Anne Bonin
Dr. Andreas Göller
Dr. Mario Lobell