HelmholtzZentrum münchen
German Research Center for Environmental Health

Large Scale Evaluation of log *P* Prediction Methods: Local Corrections Compensate Insufficient Accuracy and Eliminate the Need of Testing Every Other Compound

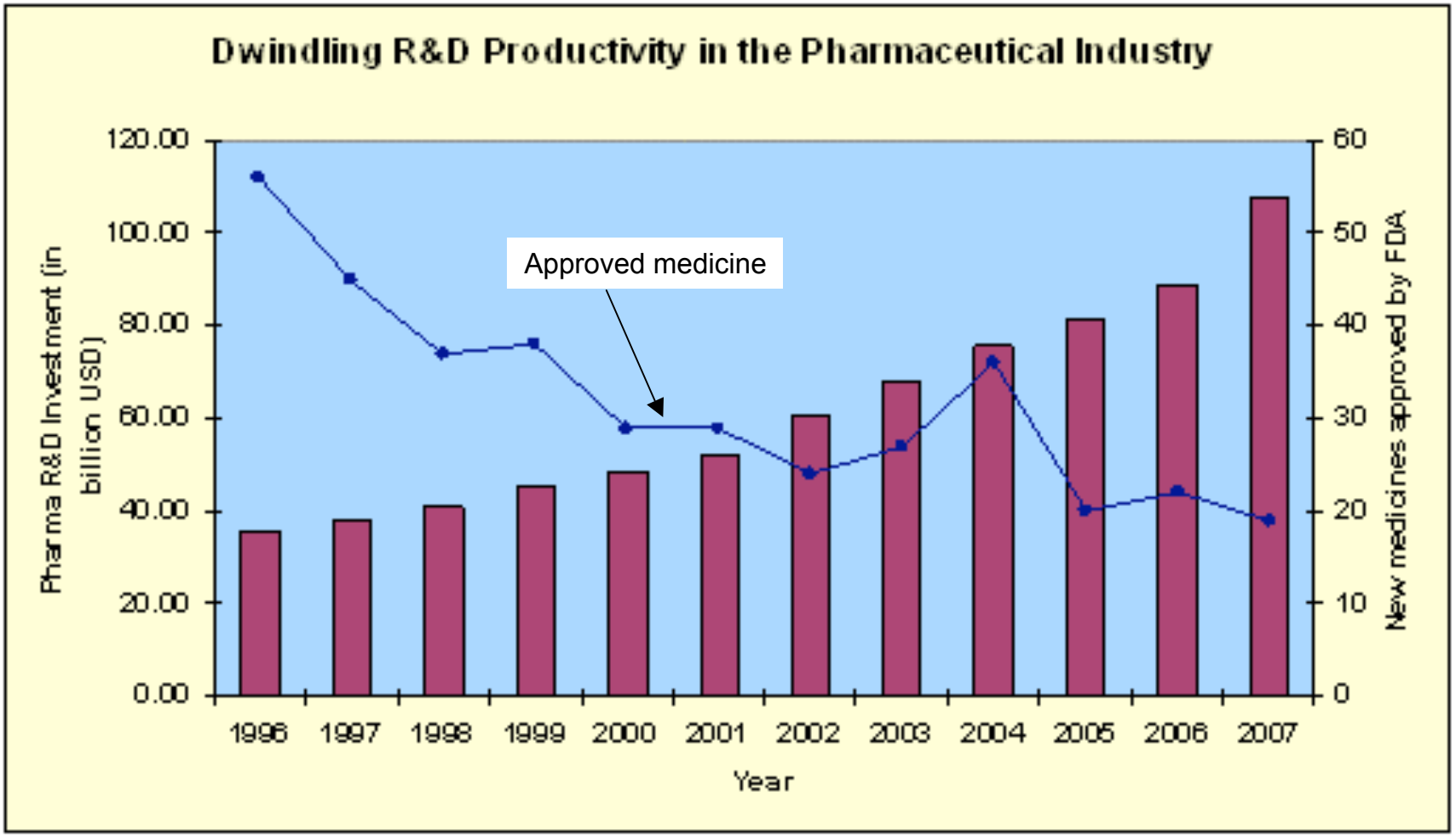
Gennadiy Poda,[1] Claude Ostermann,[2] Raimund Mannhold,[3] Joseph McDonald,[1] Igor V. Tetko[4]

1-Structural & Computational Chemistry, Pfizer Global R & D, Chesterfield, MO 63017;
2-Nycomed GmbH, Germany;
3-Heinrich-Heine-Universität, Düsseldorf, Germany;
4-Helmholtz Zentrum München - German Research Center for Environmental Health (GmbH), Institute of Bioinformatics & Systems Biology

ACS National Meeting, Washington DC, 18 August 2009

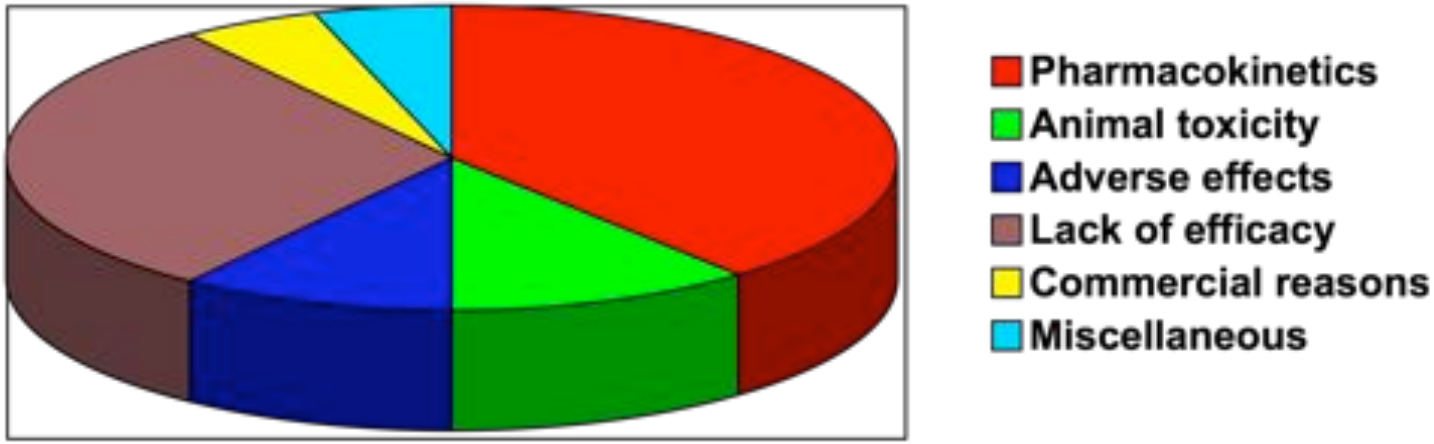
HELMHOLTZ ASSOCIATION

Declining R&D productivity in the pharmaceutical industry



http://www.frost.com/prod/servlet/market-insight-top.pag?docid=128394740

Pharma R&D Cost and productivity: Reasons for compound failure



TOP four reasons are connected to compound
Absorption, **D**istribution, **M**etabolism and **E**xcretion, all of which may contribute to lack of efficacy and **T**oxicity: **ADME/T** issues

Solutions: *in vitro* tests? *In vivo* animal tests? Costs... Time...

What is about *in silico*??? Not accurate enough? Or incorrectly interpreted...

"One can not embrace the unembraceable."

Kozma Prutkov

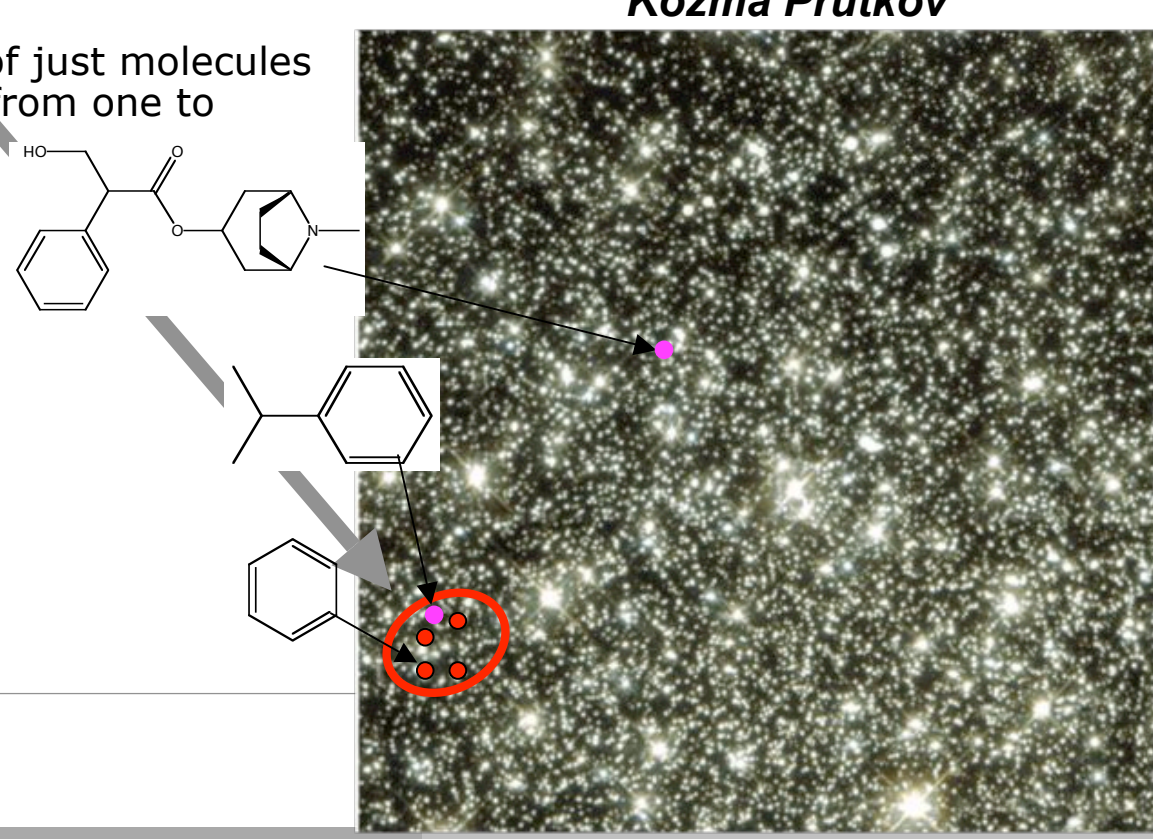
Possible: 10^{60} - 10^{100} molecules theoretically exist
(> 10^{80} atoms in the Universe)

Achievable: 10^{20} - 10^{24} can be synthesized now
by companies (weight of the Moon is ca 10^{23} kg)

Available: $2*10^{7}$ molecules are on the market

Measured: 10^{2} - 10^{4} molecules with ADME/T data

<u>Problem:</u> To predict ADME/T properties of just molecules on the market we must extrapolate data from one to 1,000 - 100,000 molecules!



There is a need for methods which can estimate the accuracy of predictions!

Existing Dogma

- Prediction of physico-chemical properties, in particular **log P**, is simple
- There is no need to measure them
- We have enough number of good computational methods
- **Is this true?**

Performance of algorithms for the public dataset

Method	Star set (*N* = 223)					Non-Star set (*N* = 43)				
			% within error range					% within error range		
	RMSE	rank	<0.5	0.5-1	>1	*RMSE*	**rank**	<0.5	0.5-1	>1
AB/LogP	0.41	I	84	12	4	1.00	I	42	23	35
S+logP	0.45	I	76	22	3	0.87	I	40	35	26
ACD/logP	0.50	I	75	17	7	1.00	I	44	33	23
Consensus log *P*	0.50	I	74	18	8	0.80	I	47	28	26
CLOGP	0.52	II	74	20	6	0.91	I	47	28	26
VLOGP OPS	0.52	II	64	21	7	1.07	I	33	28	26
ALOGPS	0.53	II	71	23	6	0.82	I	42	30	28
MiLogP	0.57	II	69	22	9	0.86	I	49	30	21
XLOGP	0.62	II	60	30	10	0.89	I	47	23	30
KowWIN	0.64	II	68	21	11	1.05	I	40	30	30
CSlogP	0.65	II	66	22	12	0.93	I	58	19	23
ALOGP (Dragon)	0.69	II	60	25	16	0.92	I	28	40	33
MolLogP	0.69	II	61	25	14	0.93	I	40	35	26
ALOGP98	0.70	II	61	26	13	1.00	I	30	37	33
OsirisP	0.71	II	59	26	16	0.94	I	42	26	33
VLOGP	0.72	II	65	22	14	1.13	I	40	28	33
TLOGP	0.74	II	67	16	13	1.12	I	30	37	30
ABSOLV	0.75	II	53	30	17	1.02	I	49	28	23
QikProp	0.77	II	53	30	17	1.24	II	40	26	35
QuantlogP	0.80	II	47	30	22	1.17	II	35	26	40
SLIPPER-2002	0.80	II	62	22	15	1.16	II	35	23	42
COSMOFrag	0.84	II	48	26	19	1.23	II	26	40	33
XLOGP2	0.87	II	57	22	20	1.16	II	35	23	42
QLOGP	0.96	II	48	26	25	1.42	II	21	26	53
VEGA	1.04	II	47	27	26	1.24	II	28	30	42
CLIP	1.05	II	41	25	30	1.54	III	33	9	49
LSER	1.07	II	44	26	30	1.26	II	35	16	49
MLOGP (Sim+)	1.26	II	38	30	33	1.56	III	26	28	47
NC+NHET	1.35	III	29	26	45	1.71	III	19	16	65
SPARC	1.36	III	45	22	32	1.70	III	28	21	49
MLOGP(Dragon)	1.52	III	39	26	35	2.45	III	23	30	47
LSER UFZ	1.60	III	36	23	41	2.79	III	19	12	67
AAM	**1.62**	**III**	**22**	**24**	**53**	**2.10**	**III**	**19**	**28**	**53**
VLOGP-NOPS	1.76	III	1	1	7	1.39	III	7	0	7
HINT	1.80	III	34	22	44	2.72	III	30	5	65
GBLOGP	1.98	III	32	26	42	1.75	III	19	16	65

Mannhold, Poda, Ostermann, Tetko, *J. Pharm. Sci.*, 2009, 98(3), 861-893

Background models

Arithmetic Average Model (AAM):
mean log*P*, used as model predicting the same value for all dataset molecules

Rank III: models with root mean squared errors (*RMSE*) close to or larger than that of AAM, **i.e. models are non-predictive**

Rank I: methods with *RMSE* identical or close to AB/LogP and ALOGPS

Rank II: remaining models

NC+NHET: log P = 1.46 + 0.11 (NC - NHET)
N=95 809, RMSE=1.04, R^2=0.2

Consensus logP: average of predicted log*P* from all rank I and II methods

Performance of algorithms for the public dataset

Method	Star set (*N* = 223)					Non-Star set (*N* = 43)				
			% within error range					% within error range		
	RMSE	rank	<0.5	0.5-1	>1	*RMSE*	**rank**	<0.5	0.5-1	>1
AB/LogP	0.41	I	84	12	4	1.00	I	42	23	35
S+logP	0.45	I	76	22	3	0.87	I	40	35	26
ACD/logP	0.50	I	75	17	7	1.00	I	44	33	23
Consensus log *P*	0.50	I	74	18	8	0.80	I	47	28	26
CLOGP	0.52	II	74	20	6	0.91	I	47	28	26
VLOGP OPS	0.52	II	64	21	7	1.07	I	33	28	26
ALOGPS	0.53	II	71	23	6	0.82	I	42	30	28
MiLogP	0.57	II	69	22	9	0.86	I	49	30	21
XLOGP	0.62	II	60	30	10	0.89	I	47	23	30
KowWIN	0.64	II	68	21	11	1.05	I	40	30	30
CSlogP	0.65	II	66	22	12	0.93	I	58	19	23
ALOGP (Dragon)	0.69	II	60	25	16	0.92	I	28	40	33
MolLogP	0.69	II	61	25	14	0.93	I	40	35	26
ALOGP98	0.70	II	61	26	13	1.00	I	30	37	33
OsirisP	0.71	II	59	26	16	0.94	I	42	26	33
VLOGP	0.72	II	65	22	14	1.13	I	40	28	33
TLOGP	0.74	II	67	16	13	1.12	I	30	37	30
ABSOLV	0.75	II	53	30	17	1.02	I	49	28	23
QikProp	0.77	II	53	30	17	1.24	II	40	26	35
QuantlogP	0.80	II	47	30	22	1.17	II	35	26	40
SLIPPER-2002	0.80	II	62	22	15	1.16	II	35	23	42
COSMOFrag	0.84	II	48	26	19	1.23	II	26	40	33
XLOGP2	0.87	II	57	22	20	1.16	II	35	23	42
QLOGP	0.96	II	48	26	25	1.42	II	21	26	53
VEGA	1.04	II	47	27	26	1.24	II	28	30	42
CLIP	1.05	II	41	25	30	1.54	III	33	9	49
LSER	1.07	II	44	26	30	1.26	II	35	16	49
MLOGP (Sim+)	1.26	II	38	30	33	1.56	III	26	28	47
NC+NHET	1.35	III	29	26	45	1.71	III	19	16	65
SPARC	1.36	III	45	22	32	1.70	III	28	21	49
MLOGP(Dragon)	1.52	III	39	26	35	2.45	III	23	30	47
LSER UFZ	1.60	III	36	23	41	2.79	III	19	12	67
AAM	**1.62**	**III**	**22**	**24**	**53**	**2.10**	**III**	**19**	**28**	**53**
VLOGP-NOPS	1.76	III	1	1	7	1.39	III	7	0	7
HINT	1.80	III	34	22	44	2.72	III	30	5	65
GBLOGP	1.98	III	32	26	42	1.75	III	19	16	65

Mannhold, Poda, Ostermann, Tetko, *J. Pharm. Sci.*, 2009, 98(3), 861-893

Performance of algorithms for *in-house* datasets

Method	Pfizer set (*N* = 95 809) RMSE	Failed[1]	rank	% in error range <0.5	0.5-1	>1	RMSE, zwitterions excluded[2]	Nycomed set (*N* = 882) RMSE	rank	% in error range <0.5	0.5-1	>1
Consensus log *P*	0.95		I	48	29	24	0.94	0.58	I	61	32	7
ALOGPS	1.02		I	41	30	29	1.01	0.68	I	51	34	15
S+logP	1.02		I	44	29	27	1.00	0.69	I	58	27	15
NC+NHET	1.04		II	38	30	32	1.04	0.88	III	42	32	26
MLOGP(S+)	1.05		II	40	29	31	1.05	1.17	III	32	26	41
XLOGP3	1.07		II	43	28	29	1.06	0.65	I	55	34	12
MiLogP	1.10	27	II	41	28	30	1.09	0.67	I	60	26	14
AB/LogP	1.12	24	II	39	29	33	1.11	0.88	III	45	28	27
ALOGP	1.12		II	39	29	32	1.12	0.72	II	52	33	15
ALOGP98	1.12		II	40	28	32	1.10	0.73	II	52	31	17
OsirisP	1.13	6	II	39	28	33	1.12	0.85	II	43	33	24
AAM	**1.16**		III	**33**	**29**	**38**	**1.16**	**0.94**	III	**42**	**31**	**27**
CLOGP	1.23		III	37	28	35	1.21	1.01	III	46	28	22
ACD/logP	1.28		III	35	27	38	1.28	0.87	III	46	34	21
CSlogP	1.29	20	III	37	27	36	1.28	1.06	III	38	29	33
COSMOFrag	1.30	1088[3]	III	32	27	40	1.30	1.06	III	29	31	40
QikProp	1.32	103	III	31	26	43	1.32	1.17	III	27	24	49
KowWIN	1.32	16	III	33	26	41	1.31	1.20	III	29	27	44
QLogP	1.33	24	III	34	27	39	1.32	0.80	II	50	33	17
XLOGP2	1.80		III	15	17	68	1.80	0.94	III	39	31	29
MLOGP(Dragon)	2.03		III	34	24	42	2.03	0.90	III	45	30	25

[1]Nr of molecules with calculations failures due to errors or crash of programs. All methods predicted all molecules for the Nycomed dataset. [2]*RMSE* calculated after excluding of 769 zwitterionic compounds from the Pfizer dataset. [3]Most molecules failed by COSMOFrag are zwitterions.

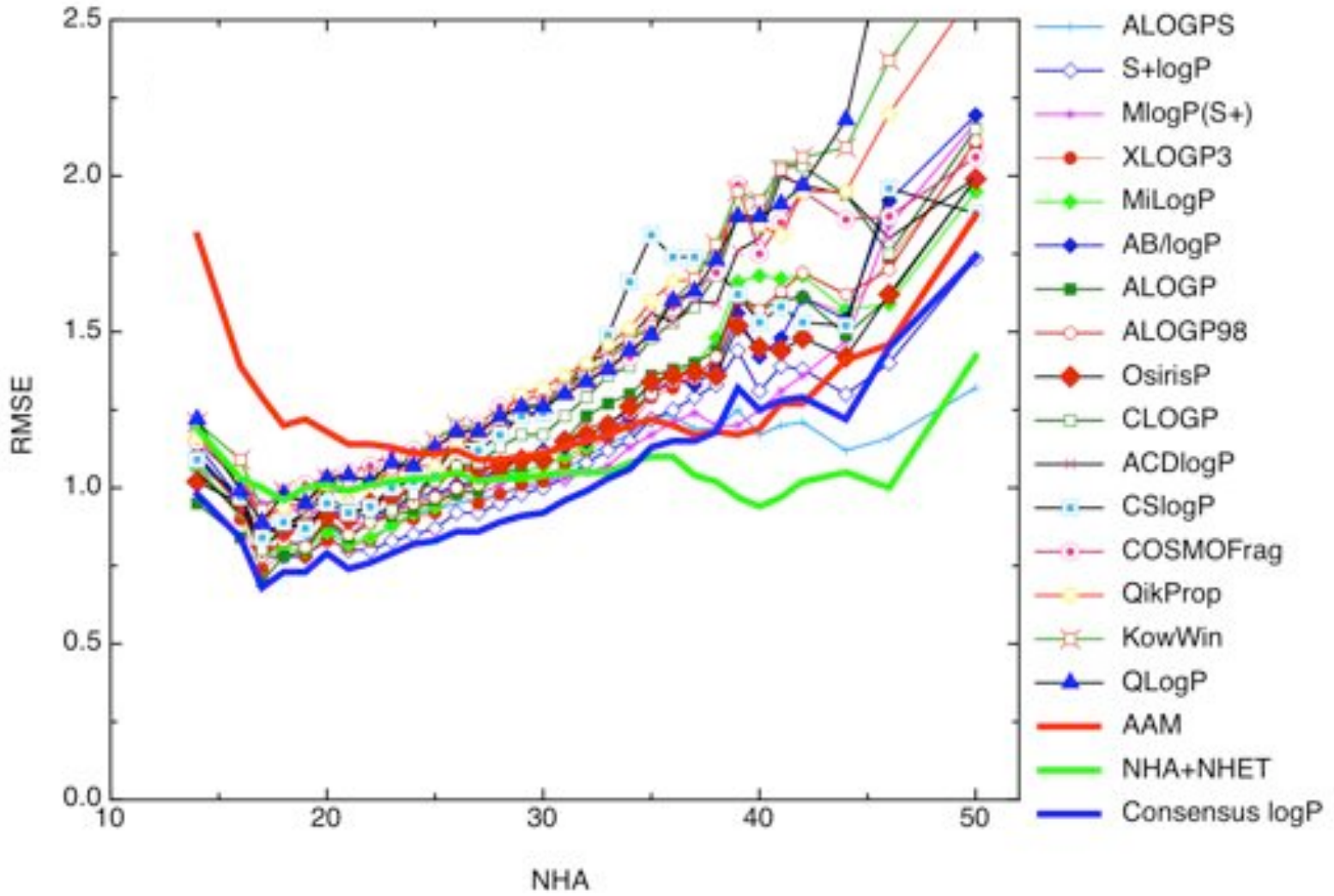
Different MlogP implementations demonstrate different performances for both sets

Mannhold, Poda, Ostermann, Tetko, *J. Pharm. Sci.*, 2009, 98(3), 861-893

Highlighted Recipes

- Identification and distinction of accurate and inaccurate predictions for global model
- Development of focused (local) models
- Estimation of the accuracy of predictions
- Conclusions

Methods performances for the Pfizer dataset



Mannhold, Poda, Ostermann, Tetko, *J. Pharm. Sci.*, 2009, 98(3), 861-893

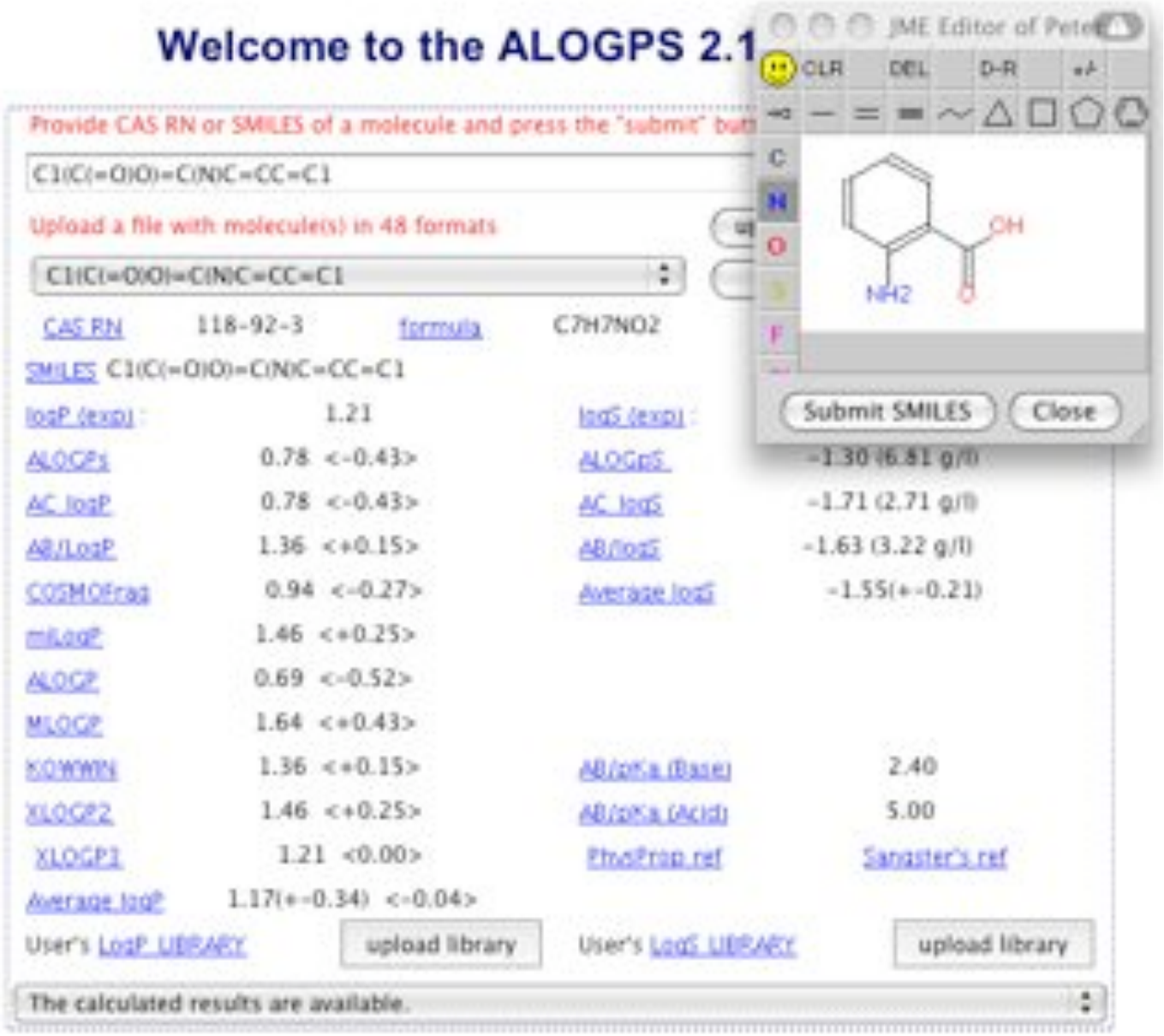
The model does not work for my data…

Is it possible to improve the model by incorporating new measurements?

ALOGPS 2.1

- LogP: 75 variables,
  12908 molecules,
  RMSE=0.35,
  MAE=0.26

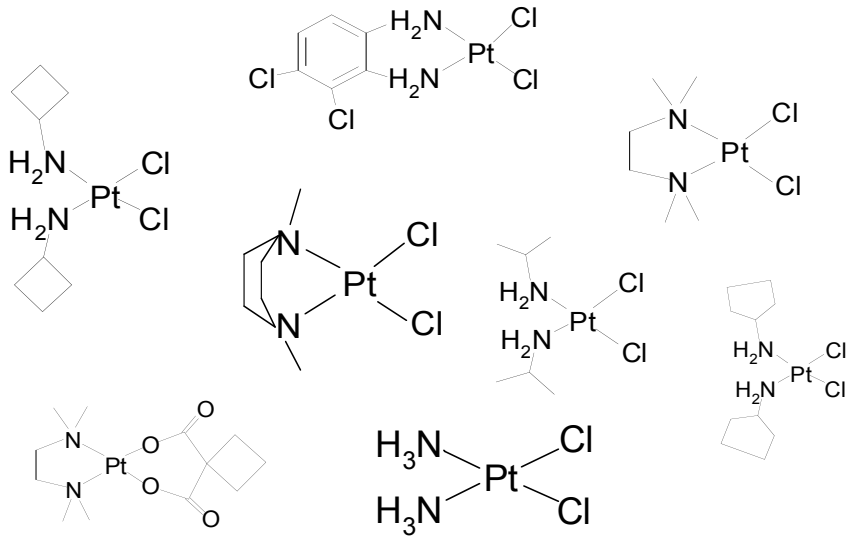
- LogS: 33 variables,
  1291 molecules,
  RMSE=0.49,
  MAE=0.35

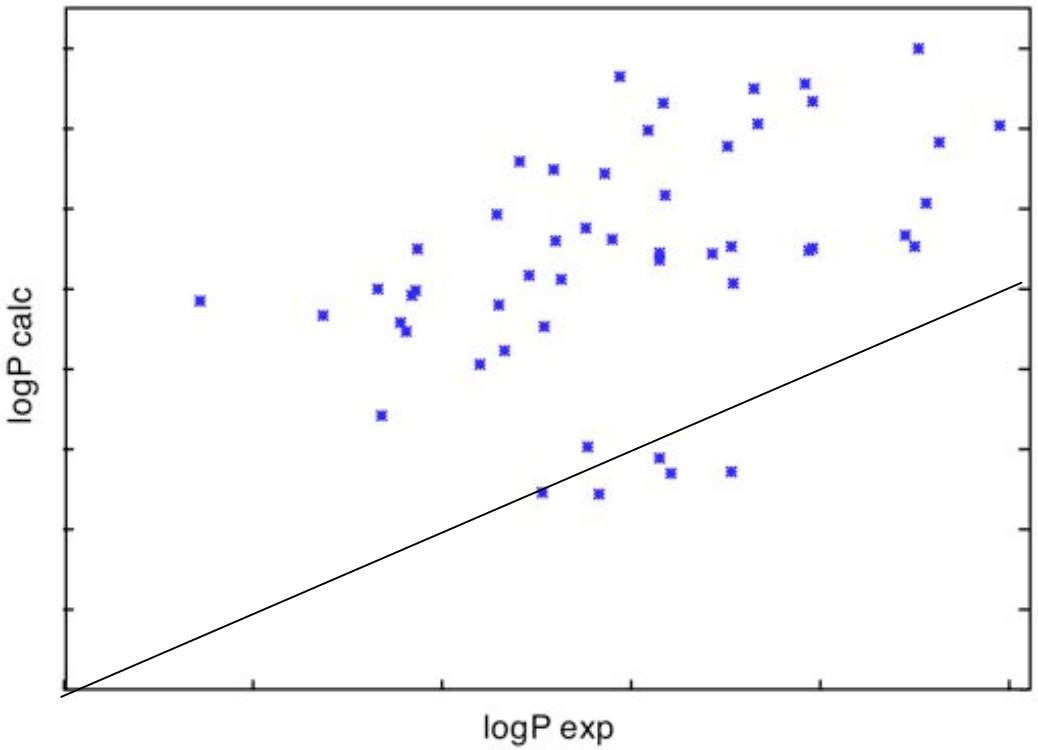


Tetko et al, J. Comput. Aided Mol. Des. 2005, 19, 453-463.

Tetko & Tanchuk, *J. Chem. Info. Comput. Sci.*, 2002, 42, 1136-1145.

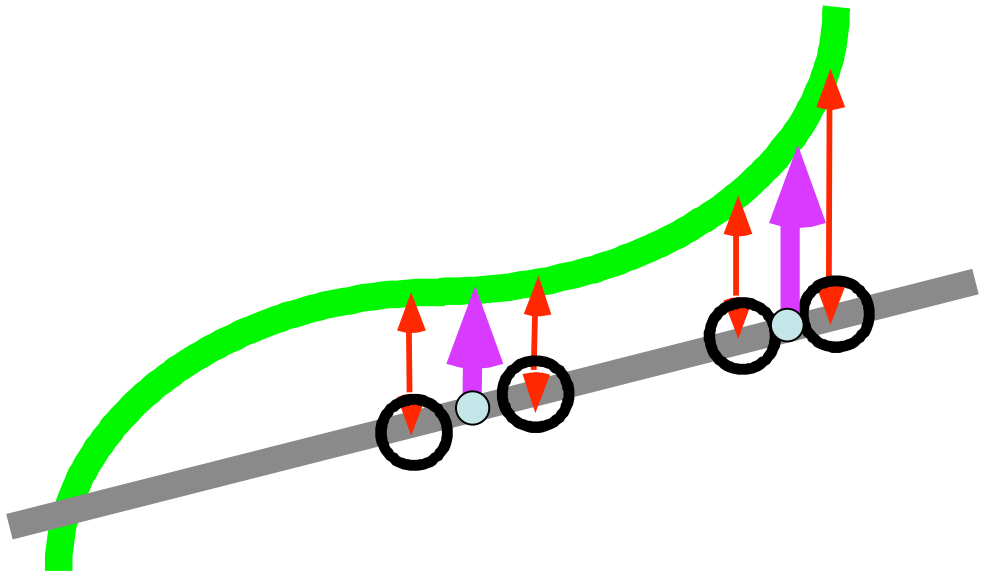
Local models: Instant learning of logP for Pt(II) molecules



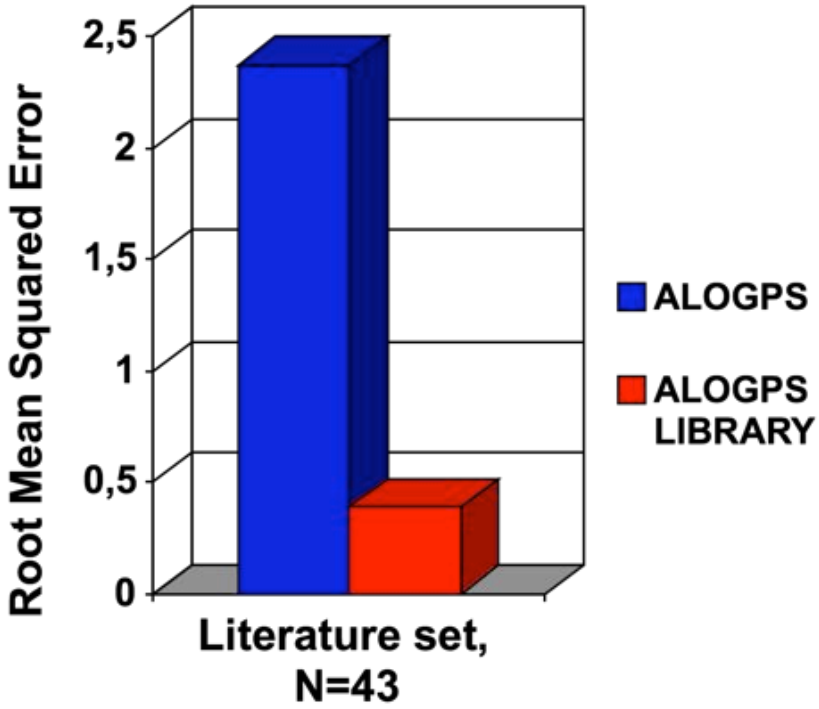


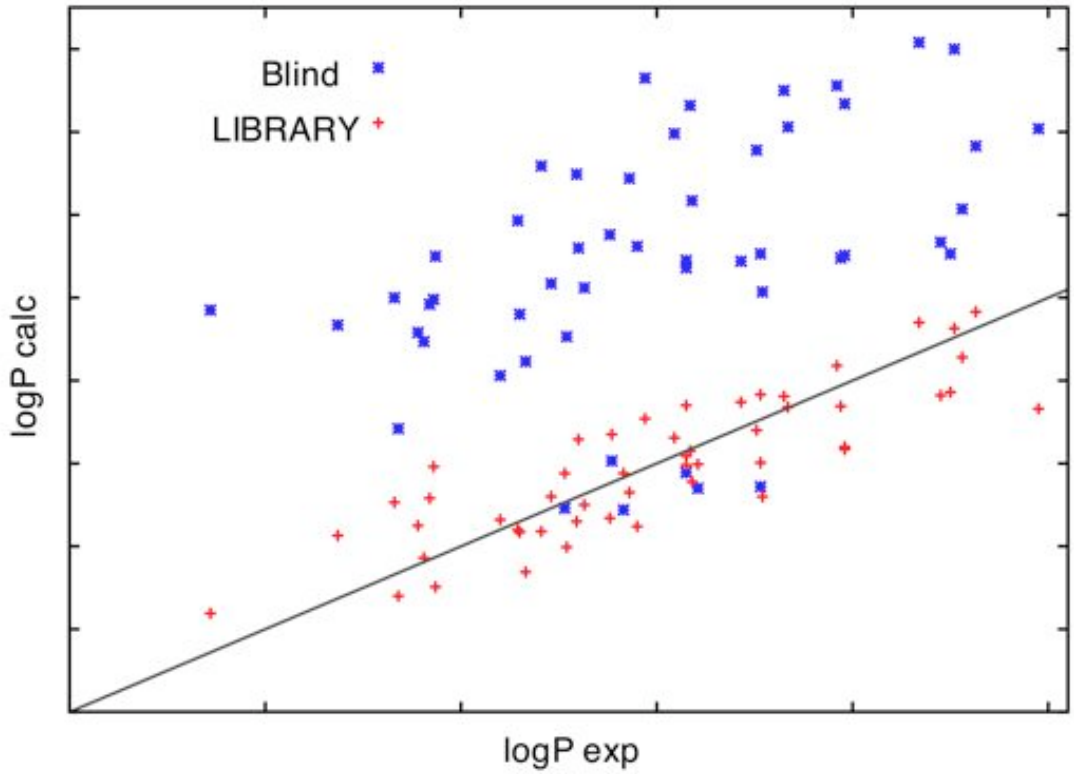
Prediction of new classes of compounds can be extremely difficult as exemplified by an absence of correlations between predicted and experimental values using the ALOGPS program.

Local correction of a model based on nearest neighbors



Local models: Instant learning by knowledge transfer





The use of LIBRARY mode (local correction of the global model) dramatically (5 times!) decreased logP errors,

Tetko et al, J. Inorg. Biochem, 2008, 102, 1424-37.

Performance of algorithms for *in-house* datasets

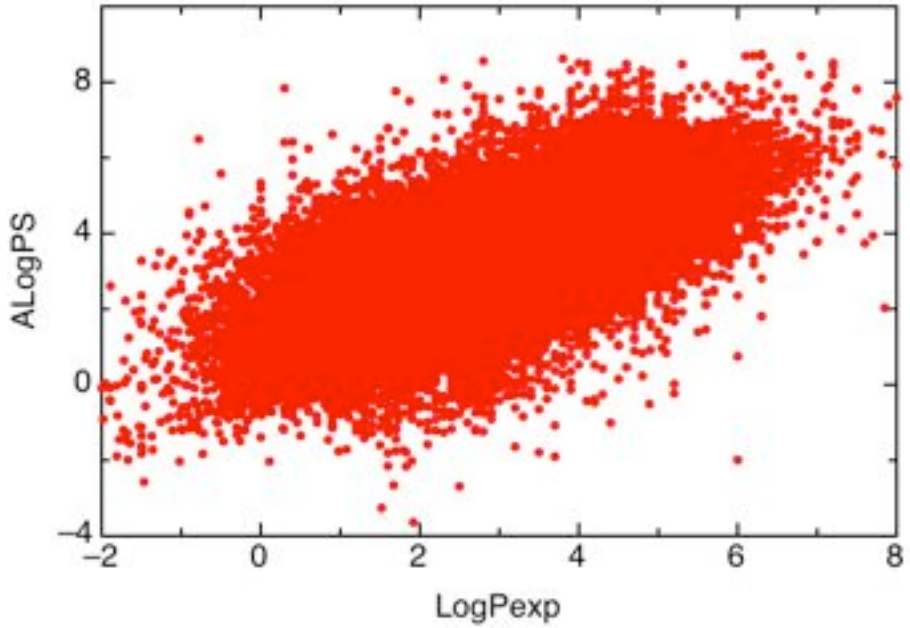
Method	Pfizer set (*N* = 95 809)							Nycomed set (*N* = 882)				
	RMSE	Failed[1]	rank	% in error range			*RMSE*, *zwitterions excluded*[2]	*RMSE*	rank	% in error range		
				<0.5	0.5-1	>1				<0.5	0.5-1	>1
Consensus log *P*	0.95		I	48	29	24	0.94	0.58	I	61	32	7
ALOGPS	1.02		I	41	30	29	1.01	0.68	I	51	34	15
S+logP	1.02		I	44	29	27	1.00	0.69	I	58	27	15
NC+NHET	1.04		II	38	30	32	1.04	0.88	III	42	32	26
MLOGP(S+)	1.05		II	40	29	31	1.05	1.17	III	32	26	41
XLOGP3	1.07		II	43	28	29	1.06	0.65	I	55	34	12
MiLogP	1.10	27	II	41	28	30	1.09	0.67	I	60	26	14
AB/LogP	1.12	24	II	39	29	33	1.11	0.88	III	45	28	27
ALOGP	1.12		II	39	29	32	1.12	0.72	II	52	33	15
ALOGP98	1.12		II	40	28	32	1.10	0.73	II	52	31	17
OsirisP	1.13	6	II	39	28	33	1.12	0.85	II	43	33	24
AAM	**1.16**		**III**	**33**	**29**	**38**	**1.16**	**0.94**	**III**	**42**	**31**	**27**
CLOGP	1.23		III	37	28	35	1.21	1.01	III	46	28	22
ACD/logP	1.28		III	35	27	38	1.28	0.87	III	46	34	21
CSlogP	1.29	20	III	37	27	36	1.28	1.06	III	38	29	33
COSMOFrag	1.30	1088[3]	III	32	27	40	1.30	1.06	III	29	31	40
QikProp	1.32	103	III	31	26	43	1.32	1.17	III	27	24	49
KowWIN	1.32	16	III	33	26	41	1.31	1.20	III	29	27	44
QLogP	1.33	24	III	34	27	39	1.32	0.80	II	50	33	17
XLOGP2	1.80		III	15	17	68	1.80	0.94	III	39	31	29
MLOGP(Dragon)	2.03		III	34	24	42	2.03	0.90	III	45	30	25

[1]Nr of molecu les with ca lculations failures due to errors or crash of programs. All methods predicted all molecules for the Nycomed dataset. [2]*RMSE* calculated after excluding of 769 zwitterionic compounds from the Pfizer dataset. [3]Most molecules failed by COSMOFrag are zwitterions.

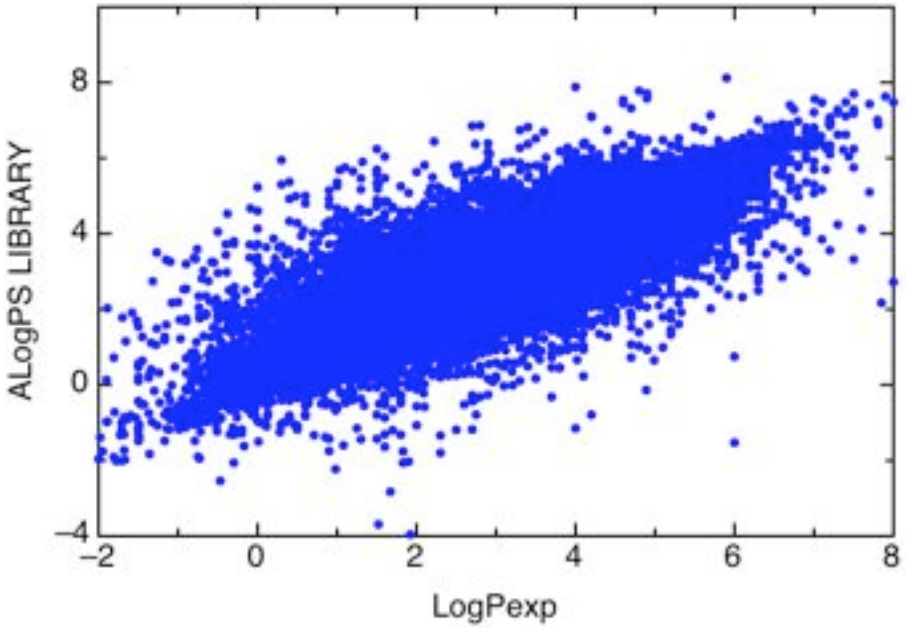
Mannhold, Poda, Ostermann, Tetko, *J. Pharm. Sci.*, 2009, 98(3), 861-893.

Local models: Instant learning of in-house data (Pfizer Inc.), *N*=95809

ALOGPS Blind prediction



ALOGPS LIBRARY

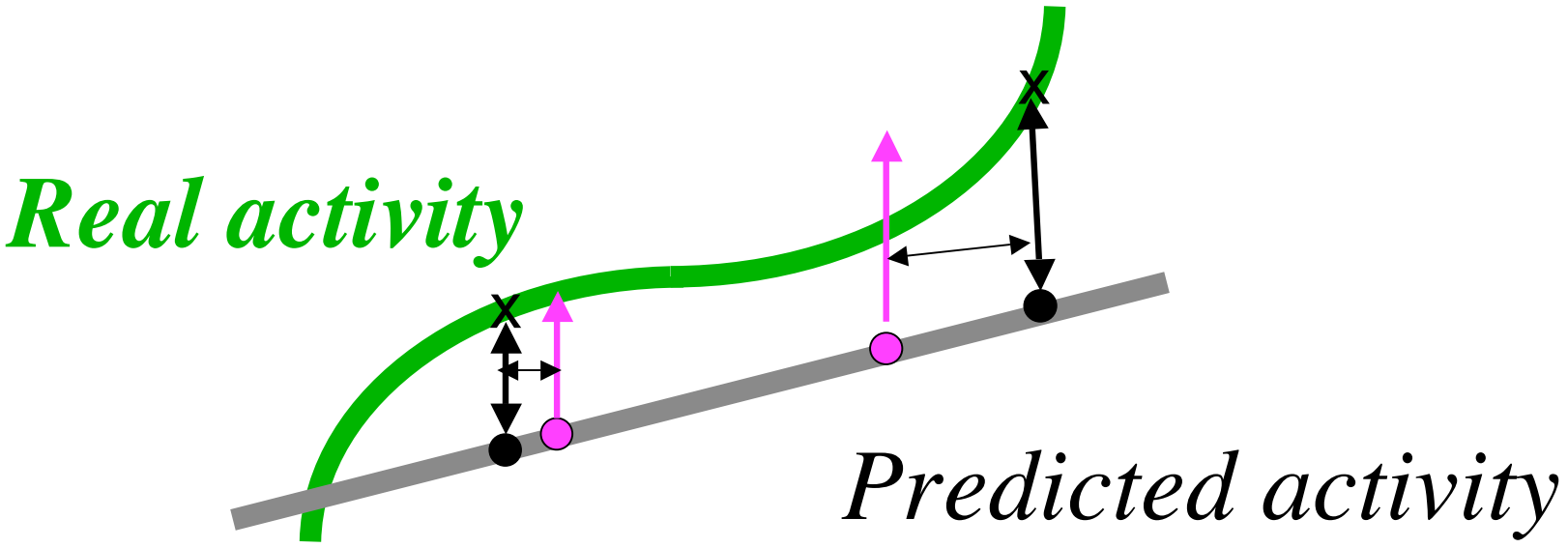


RMSE=1.02 → RMSE=0.59

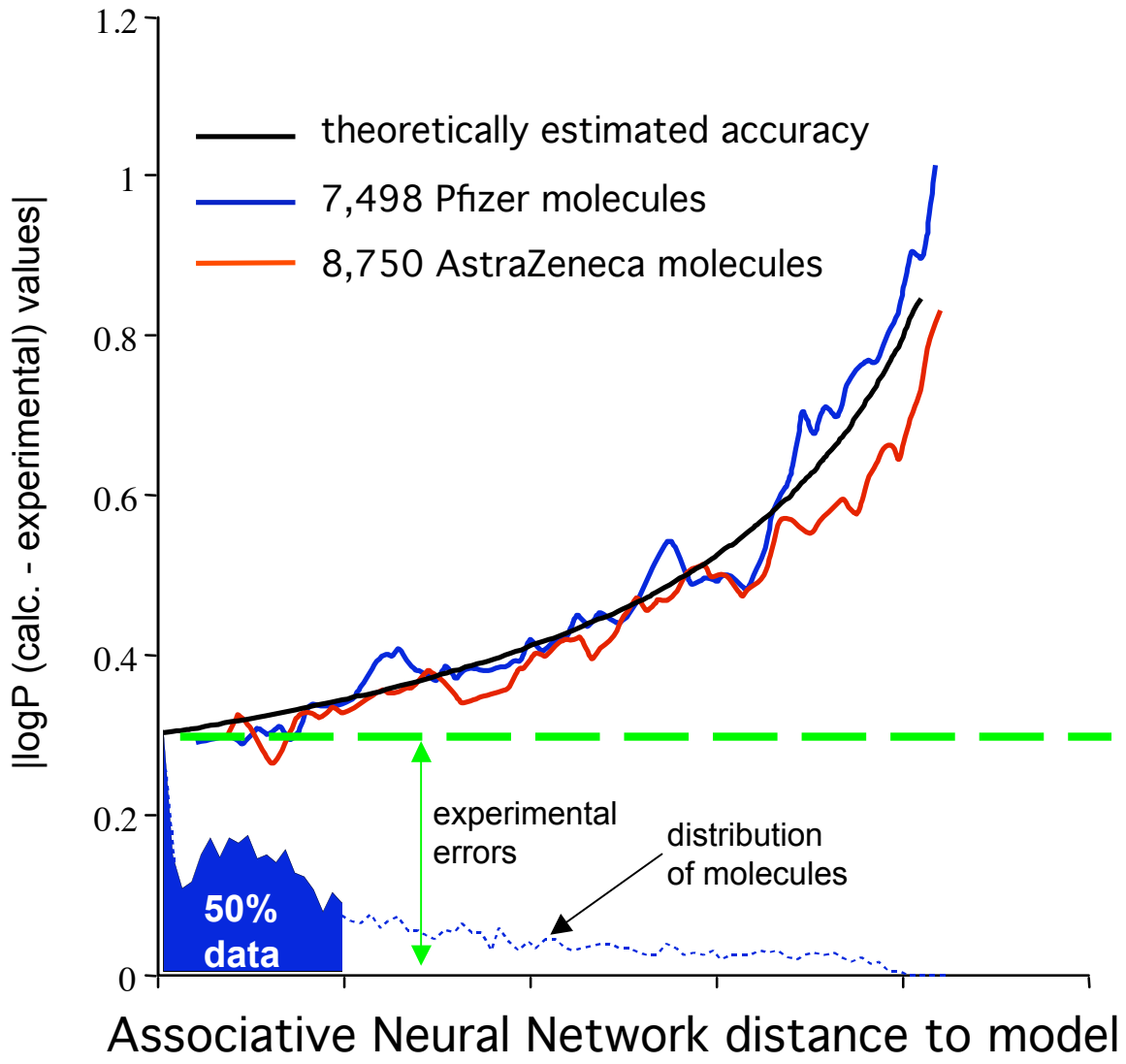
in less than 30 minutes of calculations on a notebook!

Is it possible to save costs by skipping measurements of some compounds and be satisfied with the calculated values?

Estimation of the model accuracy by the distance to nearest neighbors

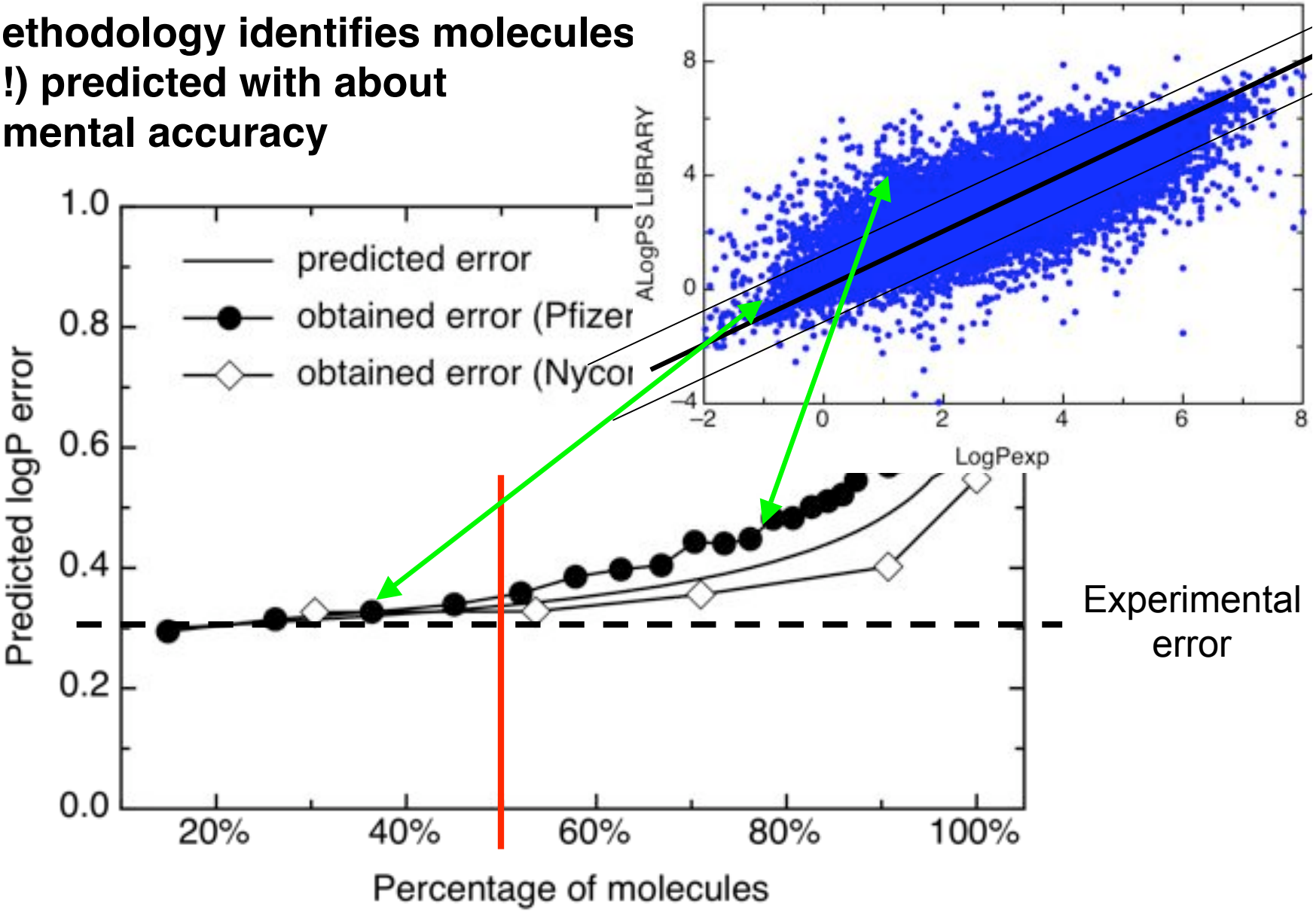


Local model: Accuracy of logP predictions



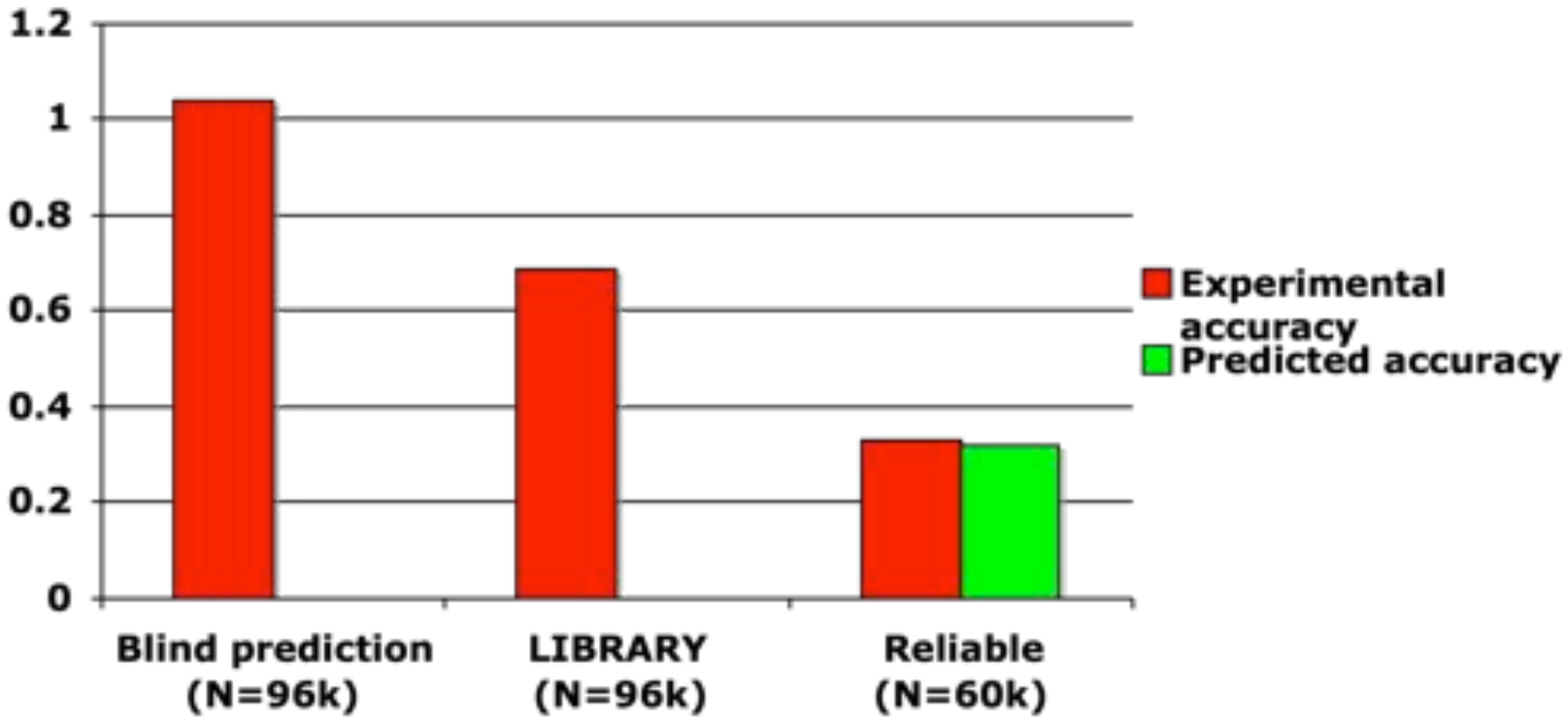
Tetko, Bruneau, Mewes, Rohrer, Poda, DDT, 2006, 11, 700-7.

Our methodology identifies molecules (>50%!) predicted with about experimental accuracy



Tetko, et al, Chemistry & Biodiversity, in press, 2009.

Improving accuracy with proposed methodology



CAse studies on the development and application of in-silico techniques for environmental hazard and risk assessment

Home

Search this site:
Search

Related topics

- Open Positions
- People
- Publications
- Related Projects
- Links
- Contact

Latest news

- Challenge on www.CADASTER.eu

Challenge

Organizers

This challenge is organized by ICANN'09: International Conference on Artificial Neural Networks, European Neural Network Society (ENNS) and CADASTER project.

Goals of this study

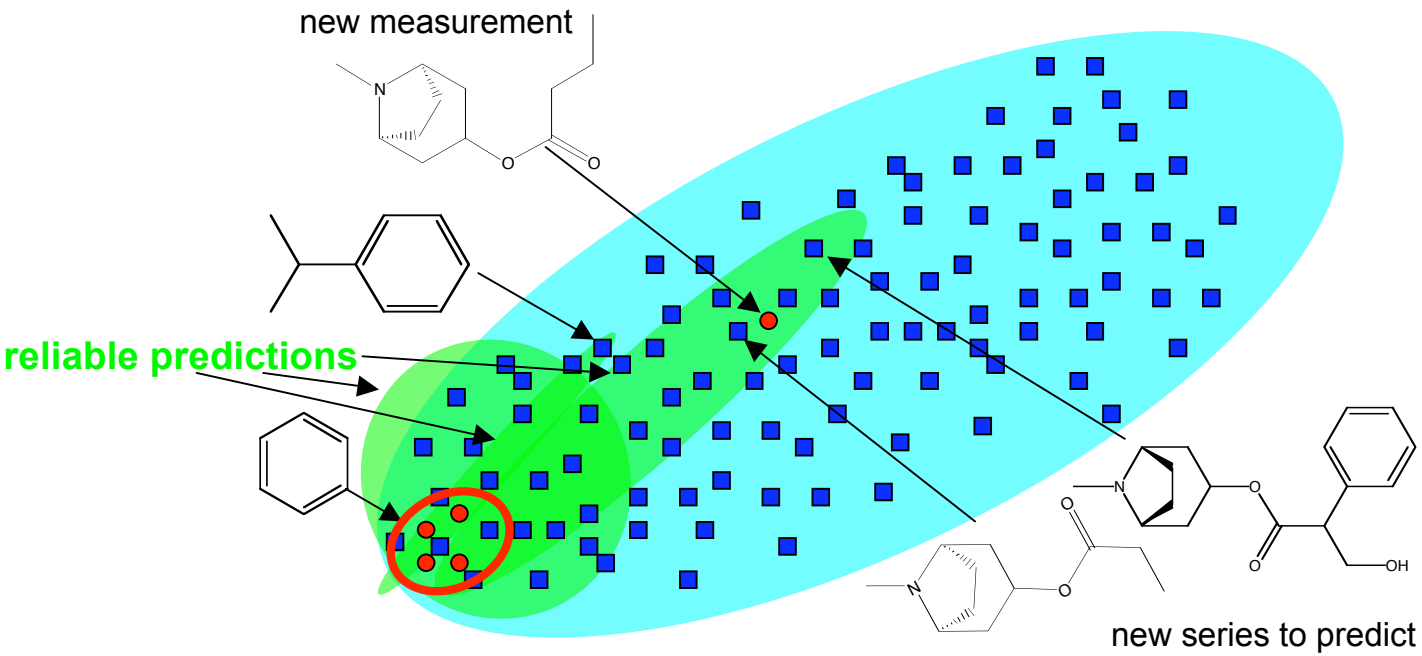
- Develop in silico models to predict environmental toxicity of molecules against T. pyriformis using data from [1].
- Estimate the accuracy of prediction of new compounds. Further information can be found here.

Important key dates

- **May 20 2009** All data for the challenge are available.
- **June 1** The submission of results is open.
- **August 31** The submission of results is closed.
- **September 14-17 2009** The winner will be announced at ICANN'09 conference in September.

The winner will be identified according to the criteria defined below and (s)he will receive a prize. It is expected that the winner as well as other participants will submit articles describing their methodological approaches for publication in a peer-reviewed journal (under discussion). Information on how you can participate can be found here.
Grand prize for the competition-winners is 1.000 € !

Challenges and solutions



Our methodology allows confident navigation in a defined molecular space.

- ✓ It can be used to develop targeted (local) models covering specific series.
- ✓ It can be used to reliably estimate which compounds can/can't be reliably predicted.
- ✓ It can be used to guide experimental design and to minimize costs of new measurements.

Acknowledgements

My group

Mr I. Sushko
Mr S. Novotarskyi
Mr A.K. Pandey
Mr R. Körner
Mr S. Brandmaier
Mrs F. Ruggiu
Dr M. Rupp

Visiting Scientists

Dr. V. Kovalishyn
Dr. V. Prokopenko
Prof. J. Emmersen

preprints?
presentation?
google “tetko”

Funding

GO-Bio BMBF http://qspr.eu
Germany-Ukraine grant UKR 08/006
DFG TE 380/1-1

FP7 MC ITN ECO
FP7 CADASTER http://www.cadaster.ue
FP6 INTAS VCCLAB http://www.vcclab.org