HelmholtzZentrum münchen
German Research Center for Environmental Health

Comparison of 30 log *P* Calculation Methods on Public and Proprietary Datasets with more than 96,000 Compounds

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

1-Helmholtz Zentrum München, Neuherberg, Germany; 2-Pfizer Global R & D, Chesterfield, USA 3-Nycomed GmbH, Konstanz, Germany; 4-Heinrich-Heine-Universität, Düsseldorf, Germany

Zürich, 9 February 2009

HELMHOLTZ ASSOCIATION

Log *P* programs used in benchmarking (Programs used for in-house datasets are given in bold)

Name	Provider	URL/E-mail
AB/LogP v. 2.0	Pharma Algorithms, Lithuania/Canada	http://www.ap-algorithms.com
ABSOLV, LSER	Pharma Algorithms, Lithuania/Canada	http://www.ap-algorithms.com
ACD/logP v. 11	Advanced Chemistry Development, USA	http://www.acdlabs.com
ALOGP (DragonX 1.4)	Talete SRL, Milano, Italy	http://www.talete.mi.it
ALOGP98	Accelrys Software Inc., USA	http://www.accelrys.com
ALOGPS v. 2.1	Virtual Computational Chemistry Laboratory, Germany	http://www.vcclab.org
CLIP	University of Geneva, Switzerland	pierre-alain.carrupt@pharm.unige.ch
CLOGP v. 4.3 (v. 5.0)	BioByte Inc., USA	http://www.biobyte.com
COSMOFrag v. 2.3	COSMOlogic GmbH & Co. KG, Germany	http://www.cosmologic.de
CSlogP v. 2.2.0.0	ChemSilico LLC, USA	http://www.chemsilico.com
GBLOGP	Max Totrov, USA	max@molsoft.com
HINT	EduSoft, LC, USA	http://www.edusoft-lc.com
KowWIN v. 1.67	Syracuse Inc., USA	http://www.syrres.com
LSER UFZ	Helmholtz Centre for Environm. Research UFZ, Germany	ralph.kuehne@ufz.de
MiLogP v. 2.2	Molinspiration Chemoinformatics, Slovak Republic	http://www.molinspiration.com
MLOGP (DragonX 1.4)	Talete SRL, Milano, Italy	http://www.talete.mi.it
MLOGP(S+), ADMET 2.3	Simulations Plus, Inc., USA	http://www.simulations-plus.com
MolLogP	MolSoft LLC, USA	http://www.molsoft.com
NC+NHET	Virtual Computational Chemistry Laboratory, Germany	http://www.vcclab.org
OsirisP	Actelion, Switzerland	http://www.actelion.com
QikProp v. 3.0	Schrödinger, LLC, USA	http://www.schrodinger.com
QLOGP	University of Miami, USA	PBuchwald@med.miami.edu
QuantlogP	Quantum Pharmaceuticals, Russia	http://q-pharm.com
S+logP, ADMET 2.3	Simulations Plus, Inc., USA	http://www.simulations-plus.com
SLIPPER-2002	Institute of Physiol. Active Compounds, Russia	http://camd.ipac.ac.ru
SPARC	University of Georgia, USA	http://ibmlc2.chem.uga.edu
TLOGP	Upstream Solutions, Switzerland	http://www.upstream.ch
VEGA	University of Milan, Italy	http://www.ddl.unimi.it
VLOGP	TOPKAT, Accelrys Software Inc., USA	http://www.accelrys.com
XLOGP2	Inst. of Physical Chemistry, Peking University, China	ftp://ftp2.ipc.pku.edu.cn
XLOGP3	Institute of Organic Chemistry, Shanghai, China	http://sioc-ccbg.ac.cn

Other analyzed models

NC+NHET equation (developed using Pfizer set):

$$\log P = 1.46(\pm 0.02) + 0.11(\pm 0.001)\ NC - 0.11(\pm 0.002)\ NHET$$

$$N=95809,\ RMSE=1.04,\ R^2=0.2$$

NC is number of carbons and **NHET** is number of heteroatoms

Consensus logP: average value of several best logP models

AAM: Arithmetic Average Model - the most basic model, i.e. an average value of all logP values

Benchmarking data

Public dataset:

N=266 molecules[1]

N=233 Star set (supported with experimental values from CLOGP v5.0 program)

N=43 Non-Star set (no experimental logP values in CLOGP v5.0)

[1]Provided by A. Avdeef, "*Absorption and drug development. Solubility, permeability and charge state"*, ed. Hoboken, NJ: Wiley–Interscience, 2003.

Industrial datasets:

N=95809 (Pfizer Inc.)[2]

N=882 (Nycomed GmbH)[3]

[2]logP and logD (pH 7.4) measurements

[3]logP measurements only

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
OsiricP	0.71	II	59	26	16	0.94	I	42	26	33
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

$$\log P = 1.244(CX)^{0.6} - 1.017(NO)^{0.9} + 0.406PRX - 0.145(UB)^{0.8} +$$

With four parameters I can fit an elephant and with five I can make him wiggle his trunk. Enrico Fermi (according to F. Dyson, Nature, 2004)

Benchmarking of logP methods for in-house data of Pfizer & Nycomed

Large number of methods could not perform better than AAM model

Best results are calculated using Consensus logP model

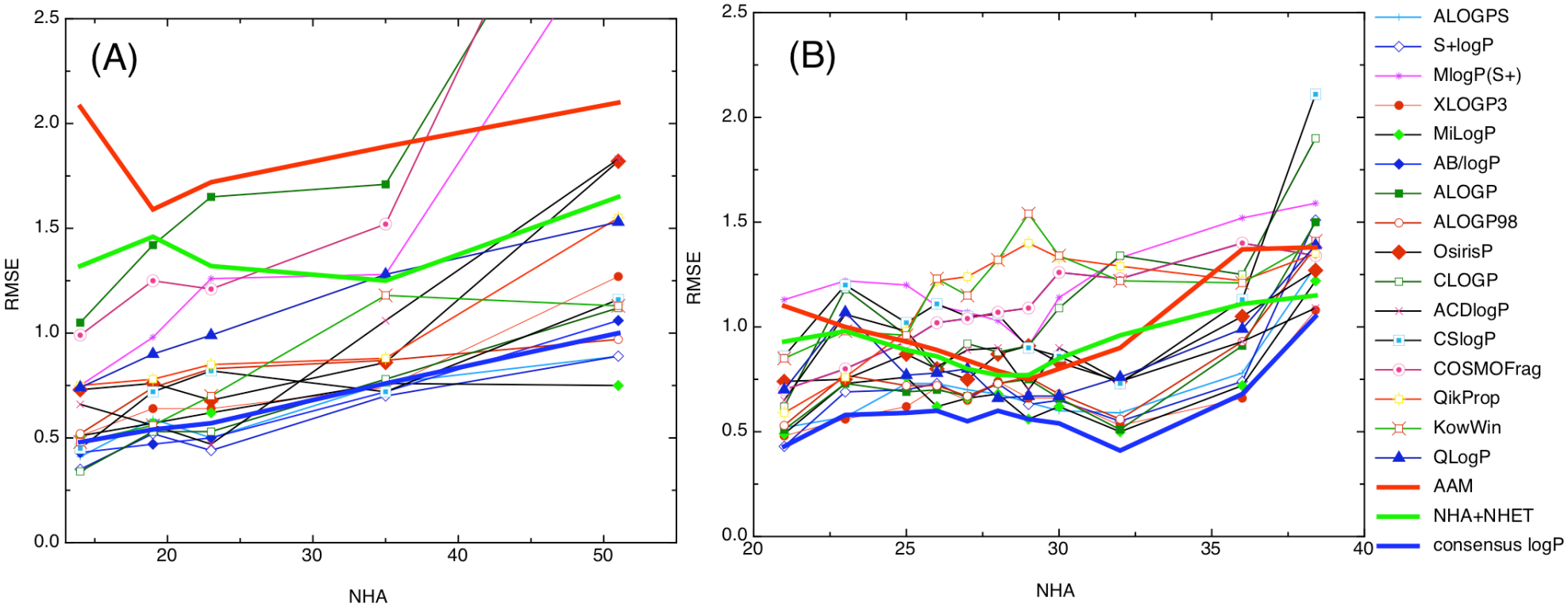
Different MlogP implementations demonstrate again very different performance for both sets

Performance of algorithms for *in-house* datasets

Method	Pfizer set (*N* = 95 809)							Nycomed set (*N* = 882)				
	RMSE	Failed[1]	rank	% in error range <0.5	0.5-1	>1	*RMSE*, zwitterions excluded[2]	*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

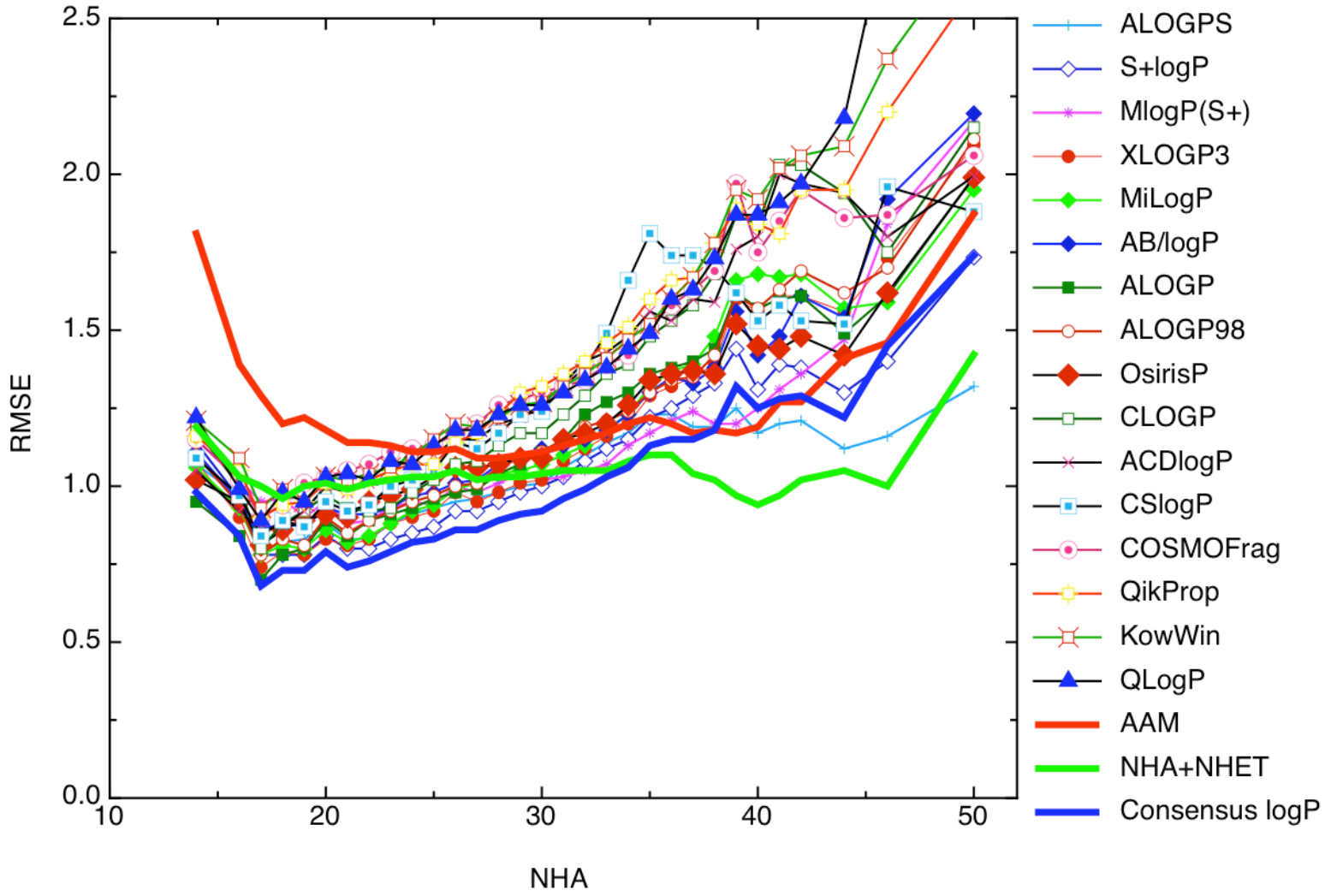
[1]Nr of molecu les with ca lculations failures due to errors or crash of programs. All methods predicted all molecules for the Nycomed dataset. [2]*RMSE* calculated after excluding of 769 zwitterionic compounds from the Pfizer dataset. [3]Most molecules failed by COSMOFrag are zwitterions.

Method performance as a function of the Number of Non Hydrogen Atoms (NHA) for the public (A) and Nycomed (B) dataset



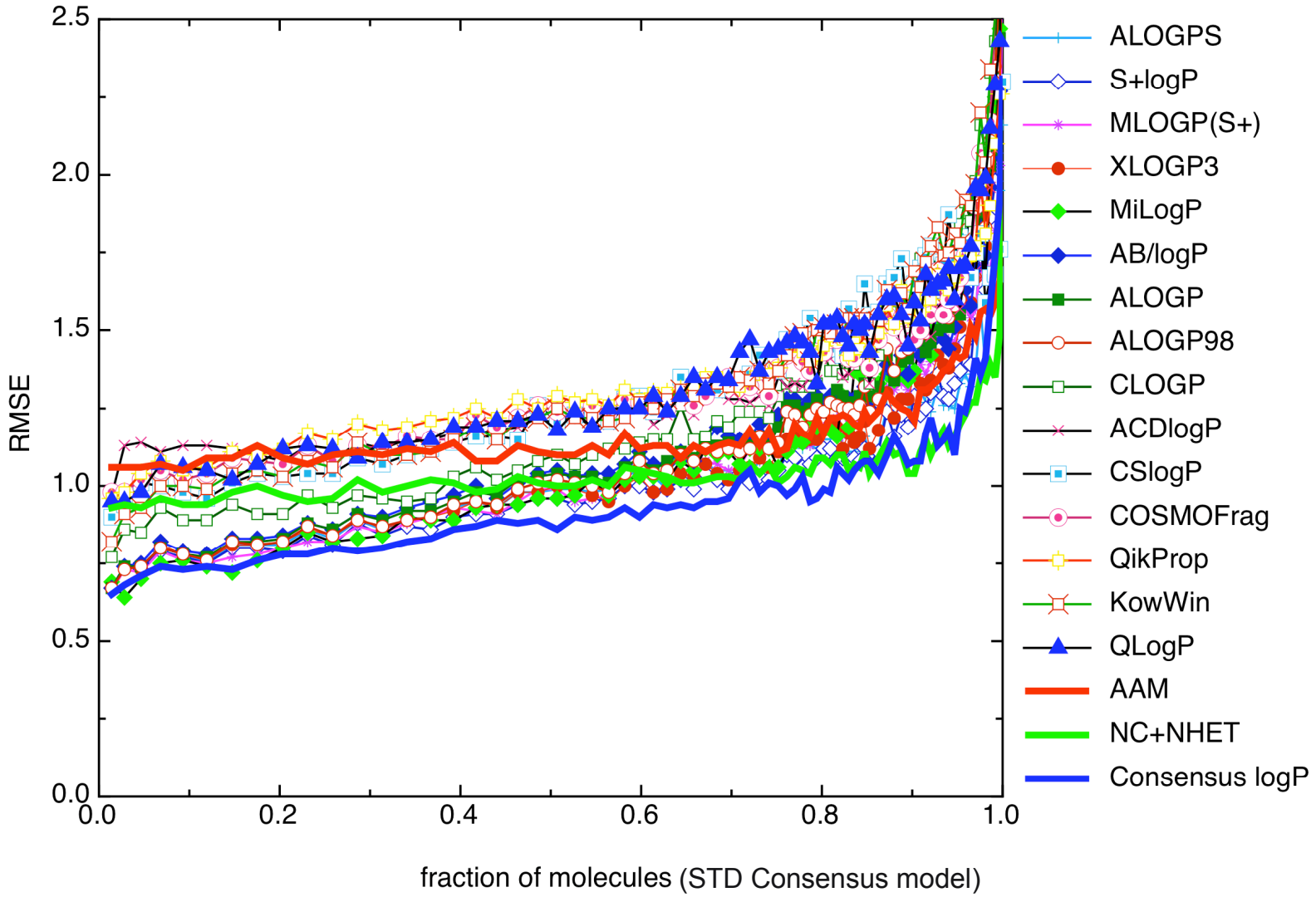
Mannhold, R. et al, *J. Pharm. Sci.*, 2009, in press.

Method performance as a function of the Non Hydrogen Atoms (NHA) for the Pfizer dataset

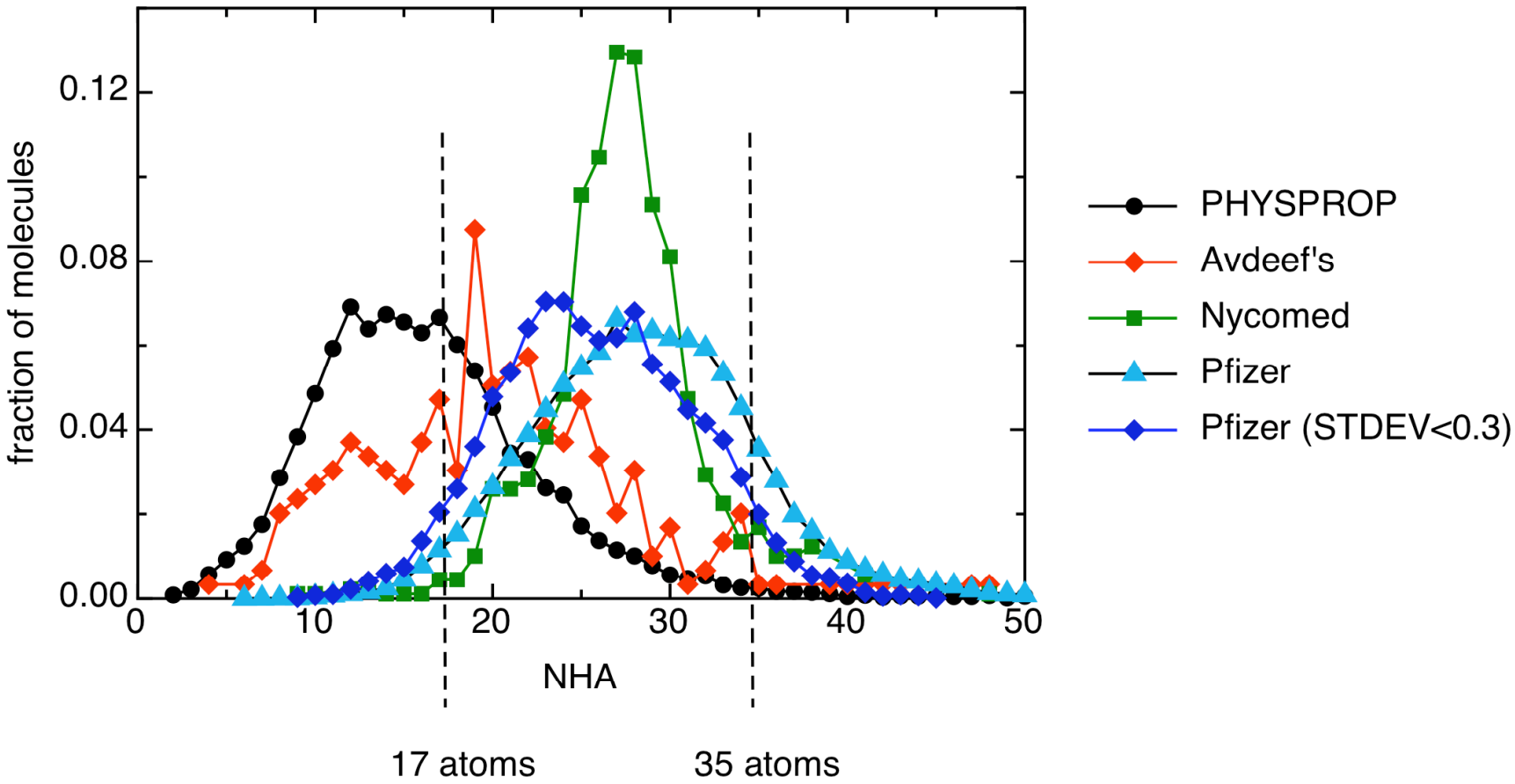


Mannhold, R. et al, *J. Pharm. Sci.*, 2009, in press.

Reliable *versus* non-reliable predictions



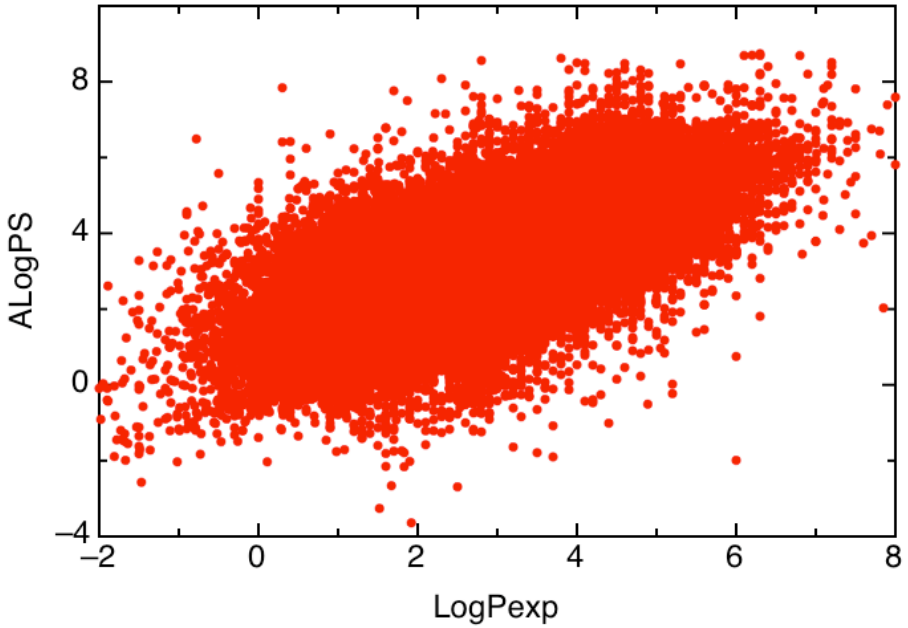
Size-related distribution of molecules in the analyzed datasets



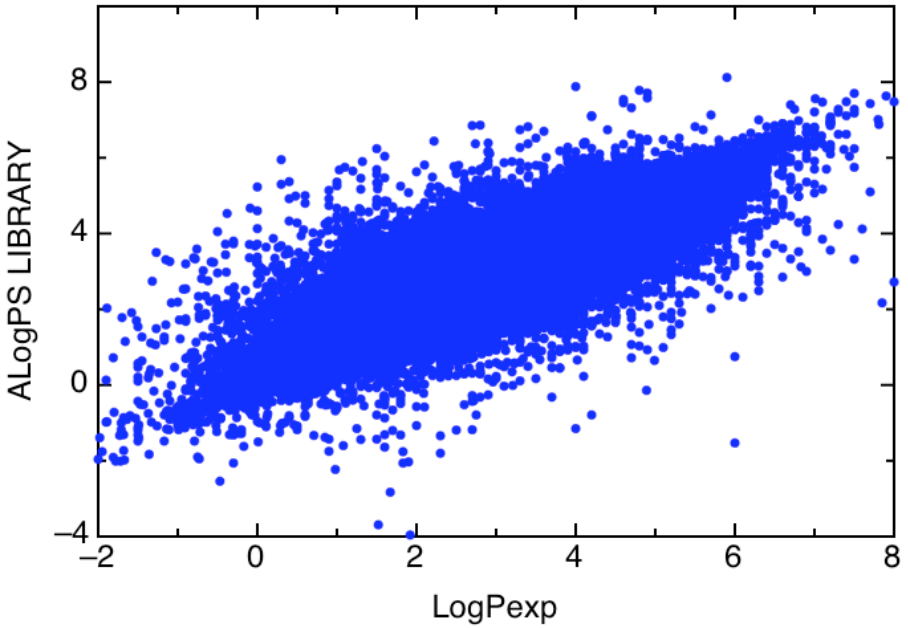
Mannhold, R. et al, *J. Pharm. Sci.*, 2009, in press.

User-training feature dramatically improves prediction accuracy, *N*=95809

Blind prediction



ALOGPS LIBRARY



RMSE=1.02 → RMSE=0.59

in less than 30 minutes of calculations on a notebook!

Conclusions

- ✓ Despite log P calculation is commonly viewed as simple, our analysis showed low prediction accuracy for most of the existing calculation methods.
- ✓ The different implementations of the same algorithm calculated very different performances
- ✓ Most methods produced reliable results for the public dataset, but for the in-house datasets only a few were superior to AAM.
- ✓ Despite low average accuracy of prediction, the confidence of prediction and the analysis of molecular size serve to distinguish reliable from non-reliable predictions.
- ✓ When prediction accuracy is low, measuring log P is highly recommended. The new values allow to improve prediction power via a user-training option without redevelopment of models

You can find references & pre-prints at http://www.vcclab.org site

Ideen zünden!

Acknowledgement

A. Klamt (COSMOlogic GmbH & Co. KG, Germany)
M. Slimak (Molinspiration Chemoinformatics, Slovak Republic)
A. Petrauskas (Pharma Algorithms, Canada)
T. Sanders (Actelion, Switzerland)
R. Todeschini (Universita di Milano – Bicocca, Italy)
L. Lai (Peking University, P. R. China)
R. Wang (Shanghai Institute of Organic Chemistry, P. R. China)

P. Buchwald (University of Miami, USA)
P.-A. Carrupt (University of Geneva, Switzerland)
R. Fraczkiewicz (SimulationsPlus, Inc. USA)
D. Haney (EduSoft LC, USA)
P. Japertas (Pharma Algorithms, Canada)
R. Kühne (Helmholtz Centre for Environmental Research, Germany)
R. Hoffmann (Accelrys Software Inc., France)
H. McClelland (Accelrys Software Inc., USA)
M. Medlin (BioByte Inc)
O. Raevsky (Institute of Physiologically Active Compounds, Russia)
A. Saravanaraj (University of Georgia, USA)
M. Totrov and **R. Abagyan** (MolSoft LLC, USA)
G. Vistoli (University of Milano)
J. Votano (ChemSilico LLC, USA)

Go-Bio BMBF
FP6 INTAS (http://www.vcclab.org)