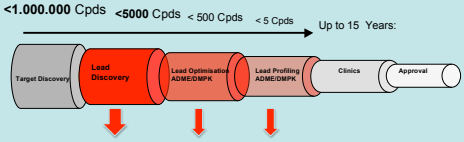
HelmholtzZentrum münchen
Deutsches Forschungszentrum für Gesundheit und Umwelt

Ideen zünden!

Development of novel ADME/T models and estimation of the accuracy of predictions for drug discovery

Igor V. Tetko, Iurii Sushko, Robert Körner, Anil Kumar Pandey, Sergii Novotarskyi, Institute of Bioinformatics & Systems Biology, Neuherberg

Pharma R&D: Cost and Productivity issues



In vitro and in vivo ADME/T property determination:
Millions of screens for solubility, stability, absorption, metabolism, transport, reactive products, drug interactions, etc.

Preclinics Costs: > $300m PER COMPOUND to reach approval

ADME market is projected to reach $4.4 billion by 2012 ($1.5 billon in 2003)

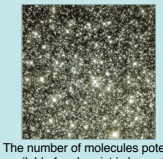
Despite an increasing investment in drug discovery over last 15 years, there has been no corresponding increase in the number of approved drugs. One of the bottlenecks causing the low success rate is the failure of drugs in the pre-clinical and clinical studies because of their unfavorable ADME/T (Absorption, Distribution, Metabolism, Excretion and Toxicology) properties. The development of methods that can predict as early as possible a failure of compounds at the late stages of drug testing are very important. The rapid progress of combinatorial chemistry approaches and the possibility to virtually screen millions of compounds enormously raised massive interest in the computational prediction of these properties as the only way to analyze such enormous collections of molecules.

We are developing novel software suites for ADME/T properties predictions. These software tools provide self-learning features to develop local models for "in-house" data and estimate the prediction accuracy. On this poster we highlight some of our recent studies based on applications of local models and we present novel tools to assess the prediction accuracy of different methods.

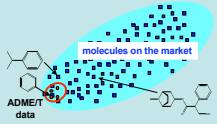
Reasons for failure in drug development

➢60% of drug failures are due to absorption, distribution, metabolism, excretion and toxicology (ADME/T) problems. Selection of compounds with most favorable properties can prevent failures on the later stages of drug development.

Computational challenges



The number of molecules potentially available for chemist is larger than the number of stars in a galaxy.

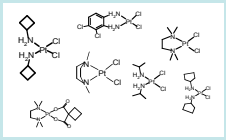


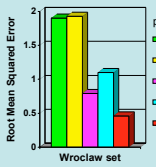
Typical ADME/T datasets have 10^2-10^5 molecules, while prediction of 10^7-10^{24} could be required. **Thus predictions *cannot be* performed with the same accuracy for all possible molecules.**

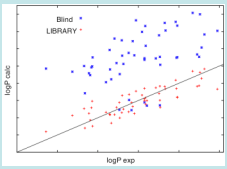
We are developing methods which calculate accurate models and estimate the quality of predictions for target scaffolds of molecules

Prediction of lipophilicity of Pt-complexes[1]

Platinum containing compounds are promising antitumor agents. Their activity is to a large extent determined by their lipophilicity (logP) but existing programs cannot satisfactory predict this property for platinum complexes.







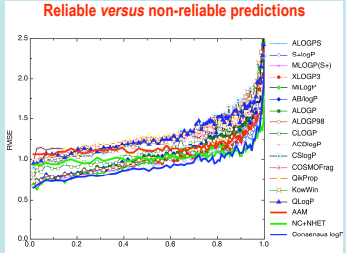
The use of our program in local mode achieved a reasonable accuracy of predictions for a set of 12 new complexes compared to experimentally determined values.

Are LogP Calculators Accurate? Benchmarking on 96 000 Compounds[2]

Performance of algorithms for *in-house* datasets

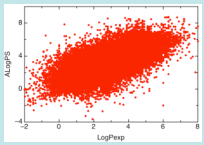
Method	Pfizer set (N = 95 809) RMSE	Failed	rank	% in error range <0.5	0.5-1	>1	RMSE partitions excluded	Nycomed set (N = 882) RMSE	rank	% in error range <0.5	0.5-1	>1
Consensus log *P*	0.95		I	48	29	24	0.94	0.58	I	61	32	7
ALOGPS	1.02		I	41	30	29	1.01	0.68	I	51	34	15
S+logP	1.02		I	44	29	27	1.00	0.69	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	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

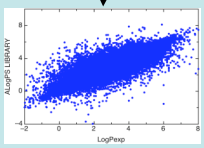
Reliable *versus* non-reliable predictions



The *RMSE* of methods for Pfizer data as function of the fraction of molecules sorted along increasing *StdDev* values. Each point (at least 500 molecules) averages errors of methods with the same (or very similar) *StdDev* values. The first 10% of molecules have a *StdDev* < 0.3.

Blind prediction,[3] RMSE=1.02





LIBRARY prediction , RMSE=0.59

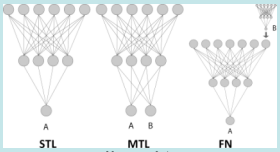
ALOGPS model not only calculated the high accuracy of predictions in blind mode but could be further used to increase it as shown above.

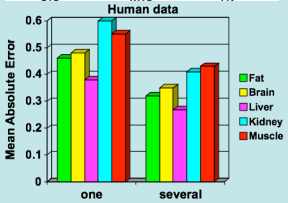
Multi-task learning[4]

Problem:
prediction of tissue-air partition coefficients
• small datasets 30-100 molecules (human & rat data)

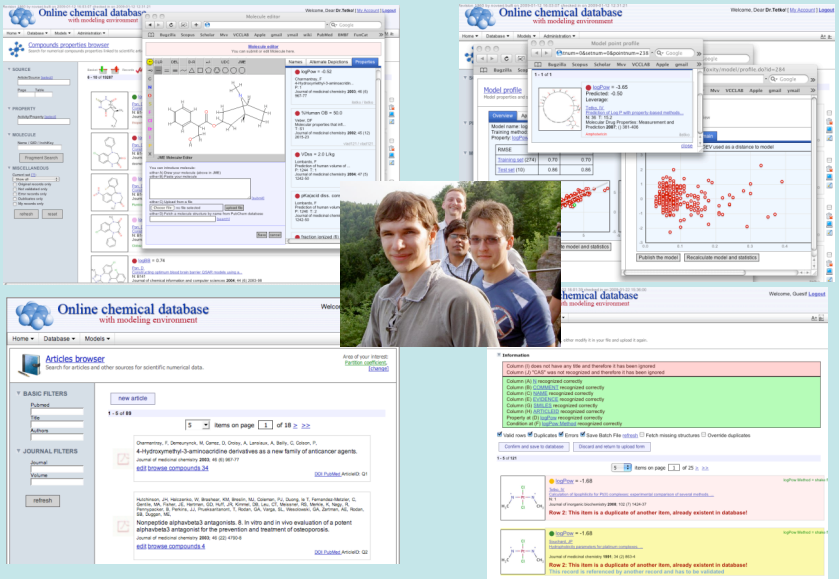
Results:
simultaneous prediction of several properties increased the accuracy of models

Perspectives:
Prediction of complex ADME/T properties with small number of data





Software developments



Work progress statistics: ~40,000 lines of code, ~ 20,000 records; ~ 200 articles; ~200 properties.

References

1) Tetko et al. *J. Inorg Biochem.*, **2008**, 102(7):1424-37; 2) Mannhold et al, *J. Pharm. Sci*, **2009**, in press; 3) Tetko et al. *QSAR Comb. Sci.*, **2009**, in press; 4) Varnek et al, *J. Chem. Inf. Comput. Sci.*, **2009**, in press.

Acknowledgement

Dr. S. Scheek (Ascenion)
Dr. C. Höfer (DMPKORE)
Prof. G. Wess (HMGU)
Prof. H.W. Mewes (HMGU)
Dr. G. Poda (Pfizer)
Dr. P. Bruneau (AstraZeneca)
Dr. C. Ostermann (Nycomed GmbH)
Prof. R. Mannhold (University of Düsseldorf)
+ many other colleagues & co-authors

Supported by BMBF GO-Bio Grant 0313883

HELMHOLTZ GEMEINSCHAFT