A Surface-Integral Model for $\log P_{ow}$ and a Local Hydrophobicity

Dr. Bernd Beck

Department of Lead Discovery, Boehringer-Ingelheim Pharma GmbH & Co. KG, 88397 Biberach (Germany)

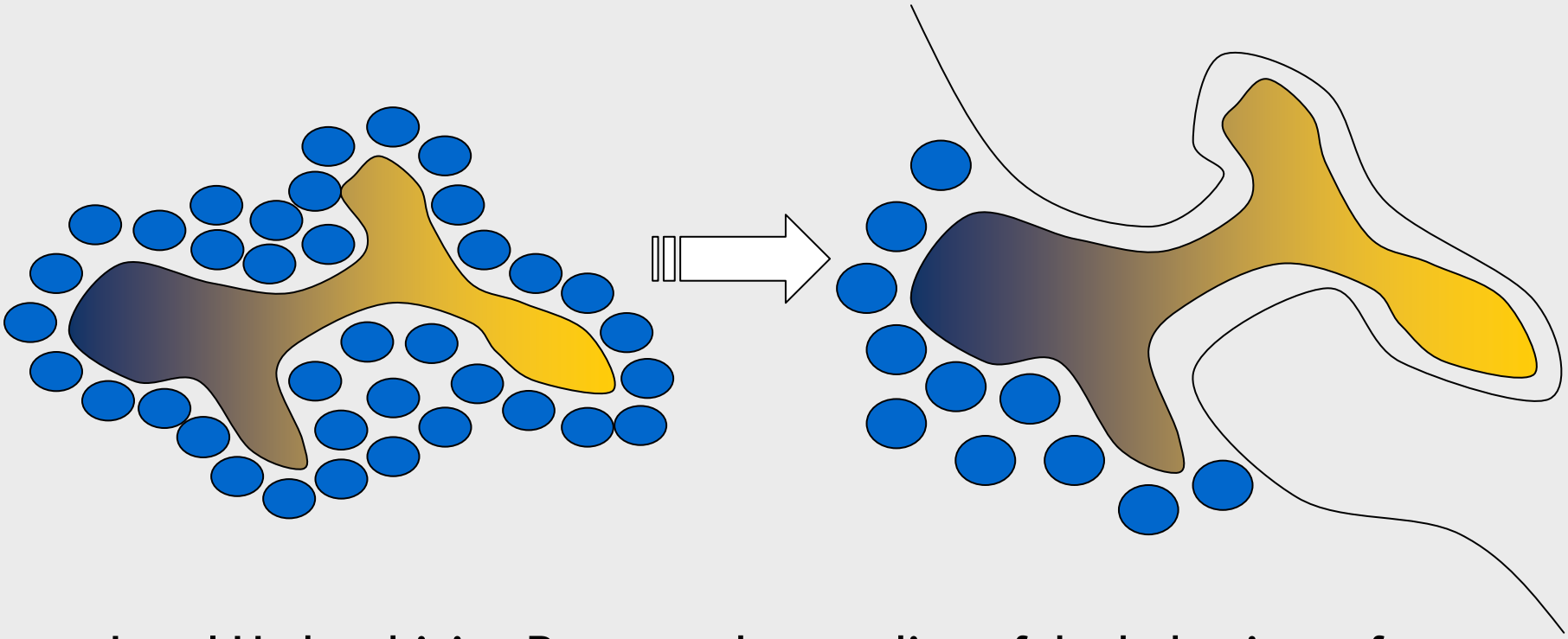
Boehringer Ingelheim

Overview

- Introduction
- Calculation & Dataset
- Surface Integral Models
- Results
- Examples of Hydrophobic Surfaces
- Summary and Outlook

Introduction – logP revisited

- $logP_{ow}$ is an important descriptor in many QSAR approaches



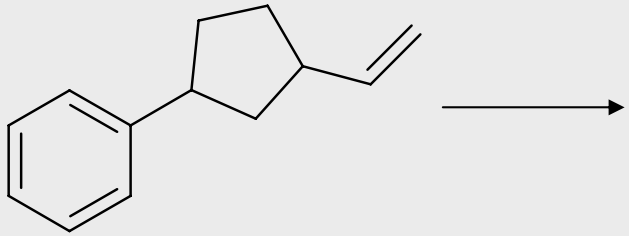
- Local Hydrophicity: Better understanding of the behaviour of a molecule (visualization)
- Integral: Partition Coefficients

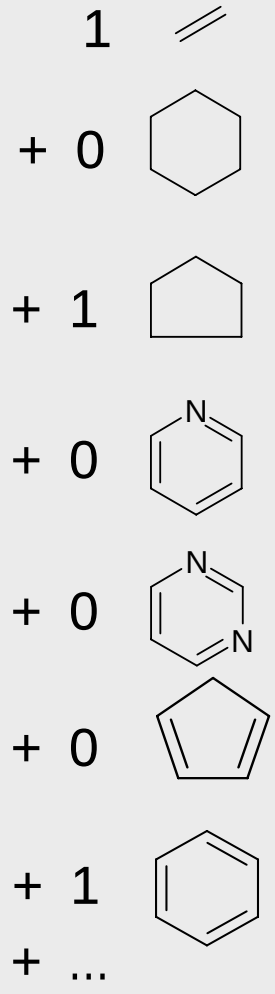
Introduction

Locality:
Substructural Fragments (e.g. ClogP)

Rekker[1]

Hansch & Leo[2]

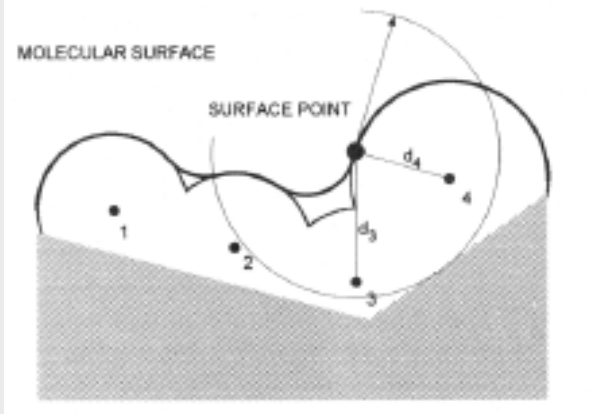




[1] Nys, G.C.; Rekker, R.F. *Chim Ther* **1973**, *8*, 521
[2] Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; WILEY: New York, 1979

Introduction

Map local hydrophobicity to the surface:



[3]

MolFESD approach 1 (1994, based on Ghose-Crippen Atom Types)[3]

MolFESD approach 2 (2003, based on MM-interactions with nonpolar probe on a GRID field)[4]

[4]

[3] Pixner, P.; Heiden, W.; Merx, H.; Moeckel, G.; Möller, A.; Brickmann, J. *J Chem Inf Comput Sci* **1994**, *34*, 1309-1319
[4] Jäger, R.; Kast, S.M.; Brickmann, J. *J Chem Inf Comp Sci* **2003**, *43*, 237-247

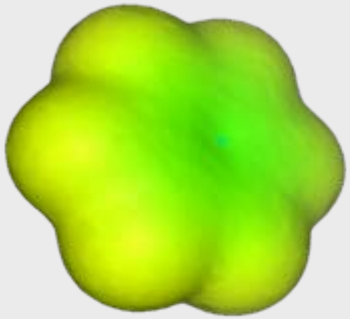
Calculation Procedure

- CORINA (2D → 3D)
- VAMP (AM1, Geometry Optimization)
- ParaSurf (Surface @ 0.003eA^{-3} ;surface, surface property calculation)
  - MEP, IEL, EAL, HARD, POL, FN surfaces
- Calculation time around 10 to 20 seconds

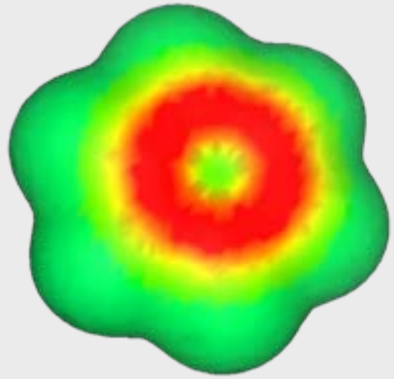
Calculated Local Properties

Boehringer Ingelheim

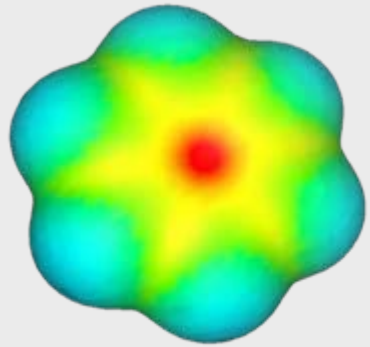
Our approach: Local Hydrophobicity as a function of QM Surface Properties



Molecular Electrostatic Potential[5]



Local Electron Affinity[7]



Local Ionization Energy[6]

+ Local Polarizability,[7] local Hardness,[7] Field normal to the surface[8]

[5] Politzer, P.; Truhlar, D.G. **1981**
[6] Sjoberg, P.; Murray, J.S.; Brinck, T.; Politzer, P. *Can J Chem* **1990**, *68*, 1440-1443
[7] Ehresmann, B.; de Groot, M.J.; Alex, A.; Clark, T. *J Chem Inf Comp Sci* **2004**, *43*, 658-668
[8] Clark, T. unpublished yet

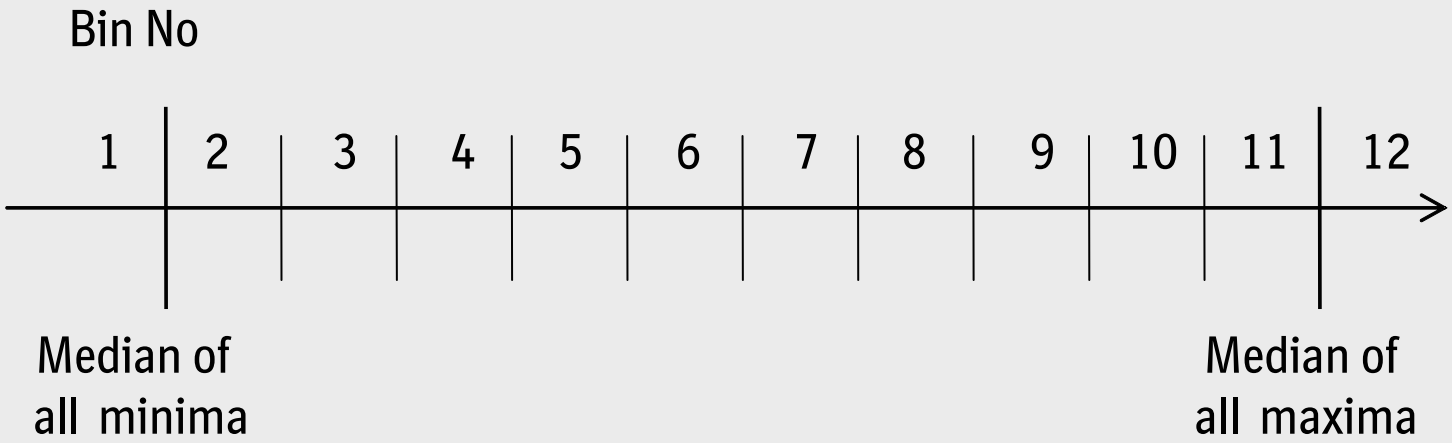
Dataset for Fitting

- Sangster Lab LOGKOW database
  - Dec. 2008: 37,783 values for 23,479 compounds
- Quality check by Sangster Lab: Suggested good quality values: ~11,500
- Structural check, only H,C,N,O,F,P,S,Cl,Br,I compounds
- Removal of zwitterions and permanently charged compounds

→ 10,812 compounds, 12.5% randomly selected as validation set

- 768 Boehringer Ingelheim Compounds as additional validation set

Surface Integral Models using: Binning



Surface Integral Models: Fitting SIM Models

Boehringer Ingelheim

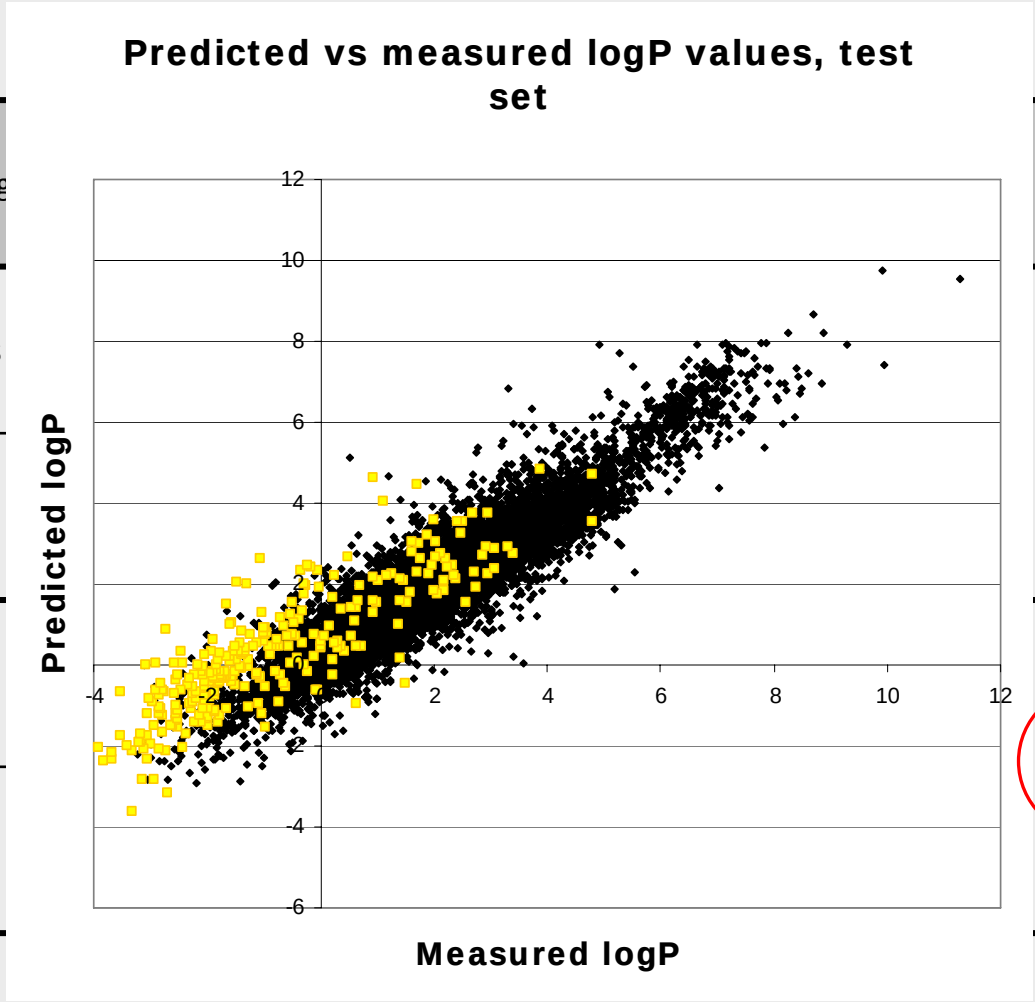
- Bagged stepwise multiple linear regression:
  - 9,462 collected and inspected $logP_{OW}$ values from James Sangsters LOGKOW database (http://logkow.cisti.nrc.ca)
  - 126 (Power Series) [9] or 252 (Binning) descriptors available
  - 50 bagging samples – 50 different models
  - 75% training set randomly chosen per sample
  - Test Set Performance evaluated after complete stepping
- Final regression contains $\frac{\sum coefficients}{50}$
- Most important descriptors: IE_L, EA_L, local hardness and MEP

[9] Ehresmann, B.; de Groot, M.J.; Clark, T. *J Chem Inf Comp Sci* **2005**, *45*, 1053-1060

Results: Integral of the local Hydrophobicity - logP

Boehringer Ingelheim

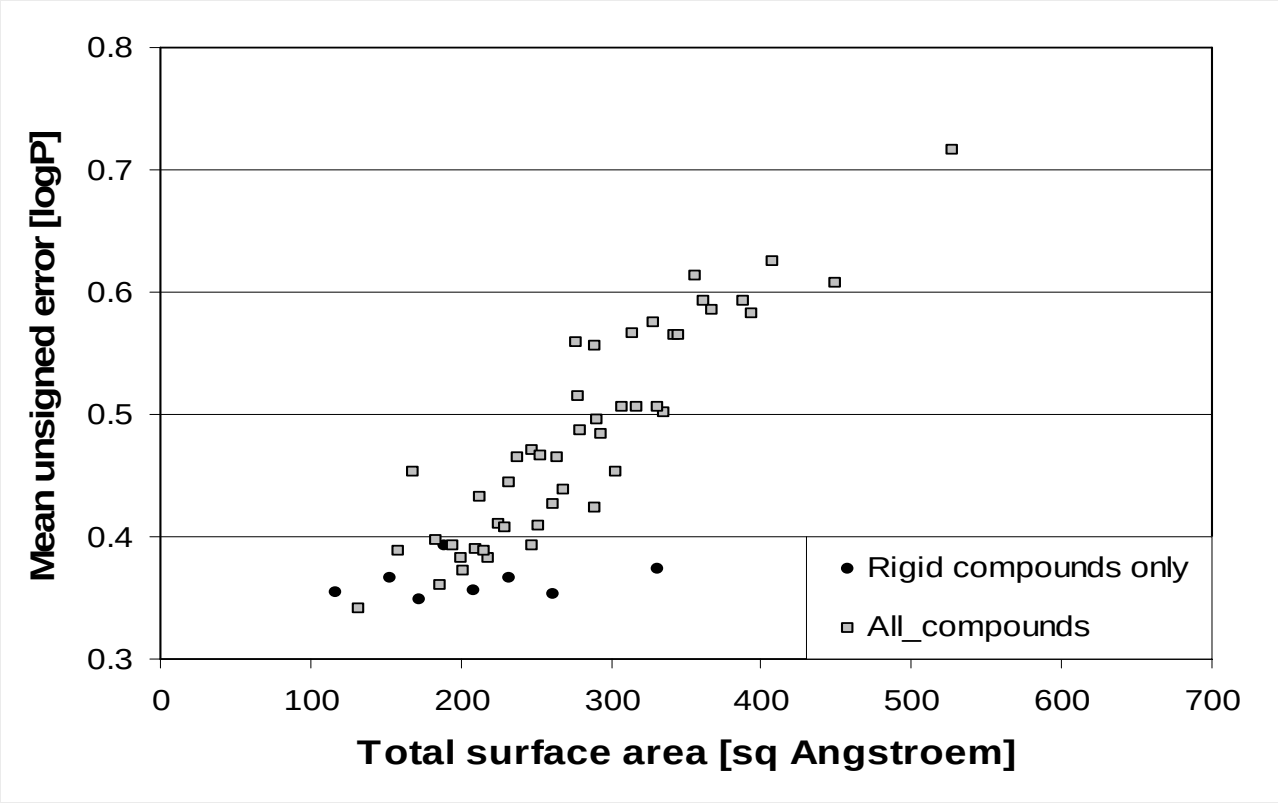
Modeling	Constant
Power series	0.58
Bagging Test Set	
Binning	0.12
Bagging Test Set	



Results: Error distribution – Rigid compounds only

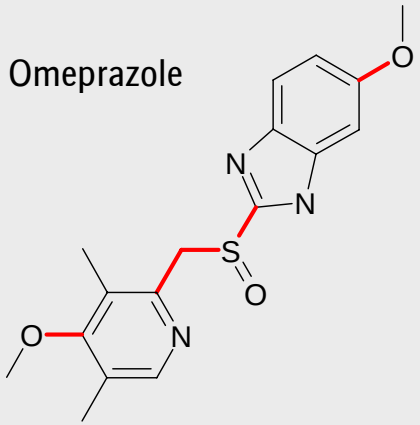
Boehringer Ingelheim

Average mean unsigned error versus TotalSurface area, comparison of compounds with and without rotatable bonds

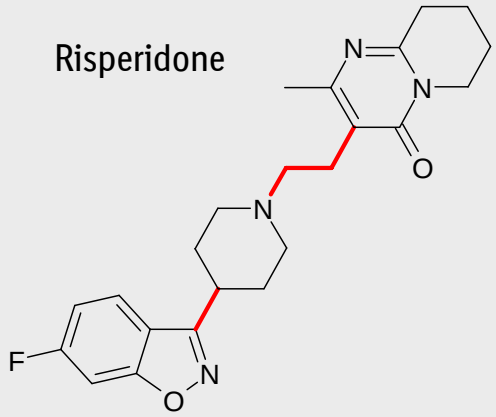


Results: Improve prediction by conformational scanning

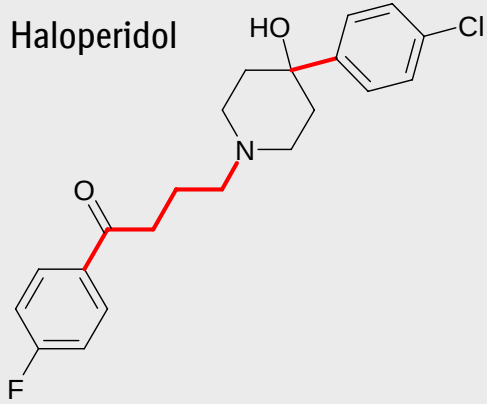
Omeprazole



Risperidone



Haloperidol



Compound	Experiment	Model	CORINA conformation	Boltzmann weighted	
				Gas-phase	SCRF
Omeprazole	2.23 2.38	Total Dataset	2.94	2.97	2.65
		Single Conf.	2.92	2.68	**2.51**
Risperidone	3.04	Total Dataset	2.48	3.20	3.14
		Single Conf.	2.78	3.17	**3.03**
Haloperidol	4.30	Total Dataset	4.52	4.23	4.19
		Single Conf.	4.17	4.20	4.18

Results: Validation Set Performance

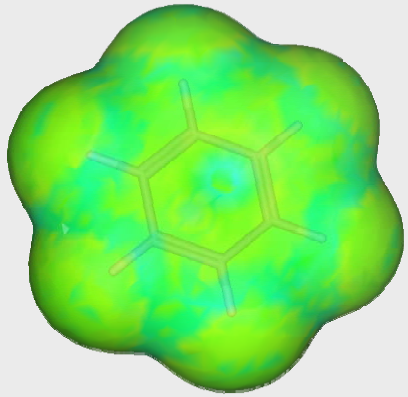
Boehringer Ingelheim

Model	Public Validation Set (1352 compounds)			Boehringer Ingelheim Validation Set (767 compounds)		
	MUE	RMSE	R^2	MUE	RMSE	R^2
SIM-logP	0.48	0.64	0.86	0.85	1.10	0.53
ACDlabs logP	0.26	0.45	0.94	1.03	1.36	0.47
ClogP	0.31	0.52	0.92	0.86	1.14	0.59
SlogP	0.53	0.68	0.85	0.92	1.19	0.51
logP_o/w	0.53	0.77	0.82	1.00	1.28	0.49
AlogP	0.62	0.86	0.79	0.97	1.24	0.50

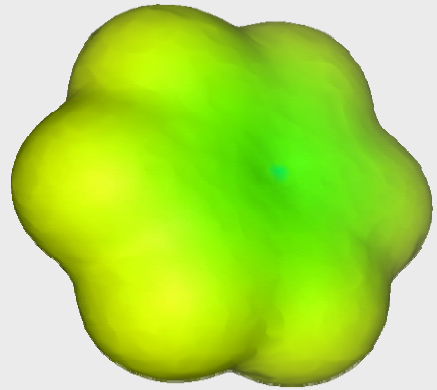
Results: Surface Integral Models for $logP_{OW}$ for Benzene

Boehringer Ingelheim

Local Hydro-phobicity



Local Molecular Electrostatic Potential (MEP)

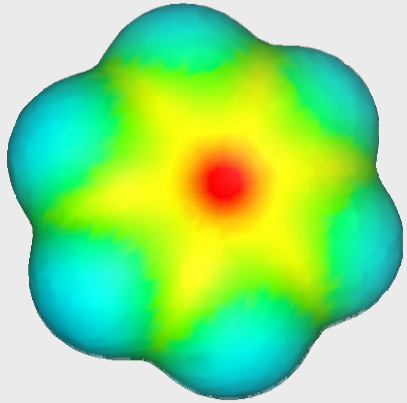


Benzene

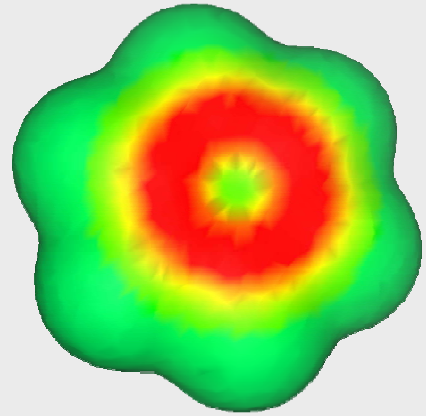
$logP_{ow}(exp) = 1.56 - 2.39$

$logP_{ow}(calc) = 2.19$

Local Ionization Energy (IEL)



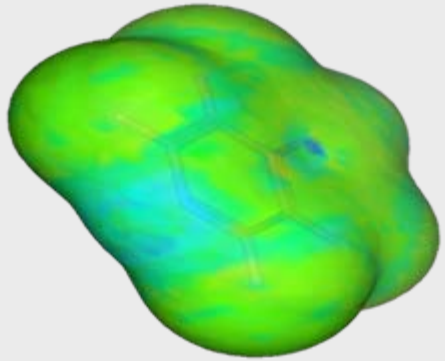
Local Electron Affinity (EAL)



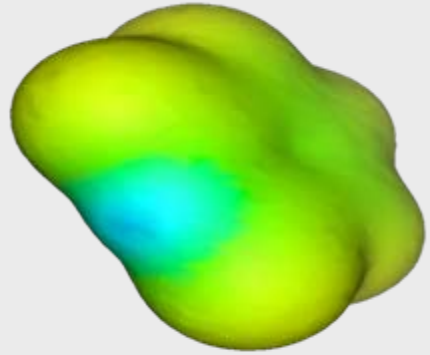
Results: Surface Integral Models for $logP_{OW}$ for Pyridine

Boehringer Ingelheim

Local Hydro-phobicity



Local Molecular Electrostatic Potential (MEP)

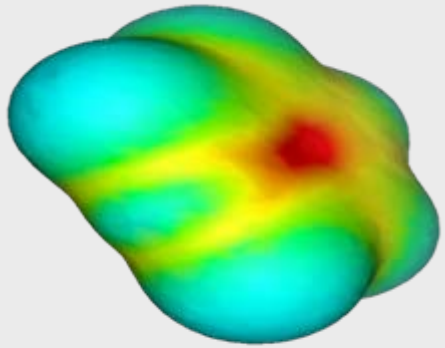


Pyridine

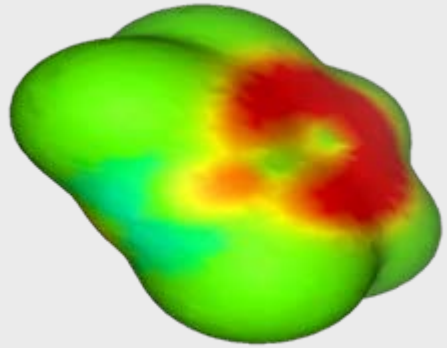
$logP_{ow}$(exp) = 0.44 – 1.09

$logP_{ow}$(calc) = 0.98

Local Ionization Energy (IEL)



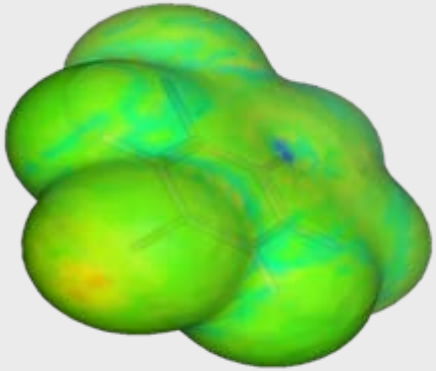
Local Electron Affinity (EAL)



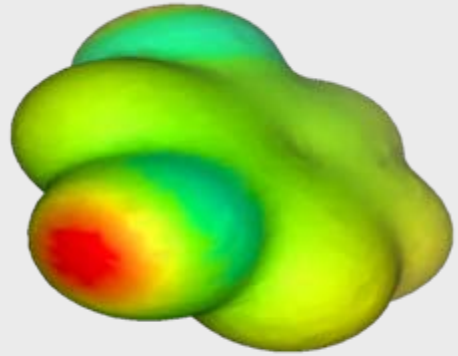
Results: Surface Integral Models for $logP_{OW}$ for 1,3 Dichlorobenzene

Boehringer Ingelheim

Local Hydro-phobicity



Local Molecular Electrostatic Potential (MEP)

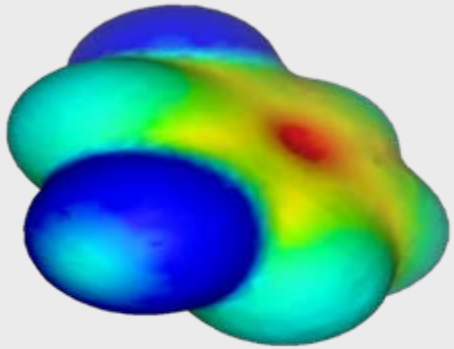


1,3 Dichlorobenzene

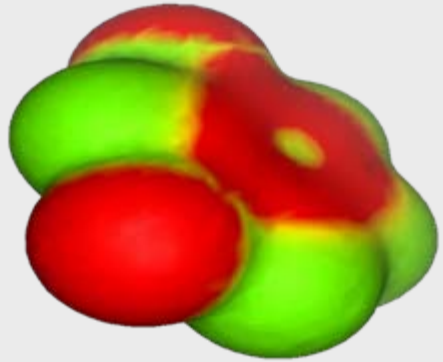
$logP_{ow}$(exp) = 3.38 – 3.62

$logP_{ow}$(calc) = 3.41

Local Ionization Energy (IEL)



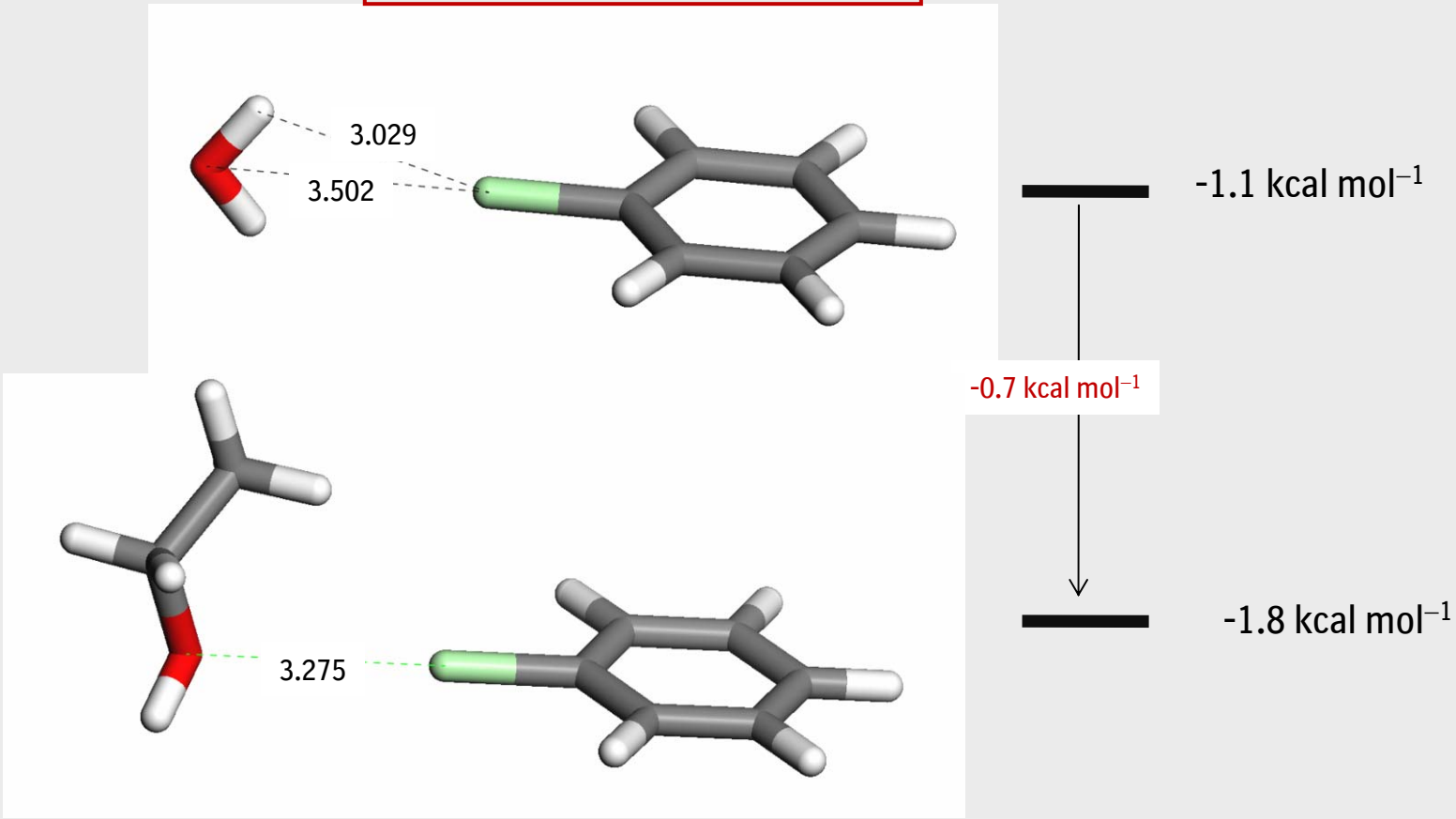
Local Electron Affinity (EAL)



MP2/aug-cc-PVTZ: Halogen Bond

Boehringer Ingelheim

C_{ipso}-Cl-O constrained to 180°



Summary and Outlook

- Local hydrophobicity defined; Integral = $logP_{ow}$
- Prediction performance comparable or better than state-of-the-art commercial models
- Prediction performance benefit from conformational sampling
- All calculations including QM ~10 – 20s for a single conformer
- Visualization

- Systematic improvements possible:
  - Choice of Hamiltonian (e.g. AM1*,PM6)
  - Conformational Sampling
  - Use of Solvation model
- SIM models for different liquids / states:
  - Hexane/Water, Chloroform/Water, …
  - Zwitterions, charged species (→ $logD_x$)

Acknowledgments

Dr. Christian Kramer
Prof. Timothy Clark
Dr. Matthias Hennemann

This work is published in J. Chem. Inf. Model

Boehringer Ingelheim

Thank you for your attention