THIRD HUMBOLDT CONFERENCE ON COMPUTATIONAL CHEMISTRY

Varna, Bulgaria
24 - 28 June 2006

THIRD HUMBOLDT CONFERENCE ON COMPUTATIONAL CHEMISTRY

Editors:
Hristiyan A. Aleksandrov
Galina P. Petrova
Georgi N. Vayssilov
Tzonka Mineva
Rositca Nikolova

The picture at the cover is produced by Miroslav Rangelov with MolRan visualization program.

ISBN-10: 954-323-199-0
ISBN-13: 978-954-323-199-7
Avangard Prima
Sofia, 2006

Modeling the Octanol-Water Partition Coefficient of Substituted Phenols by the Use of Structure Information

Lorentz Jäntschi[1], Sorana-Daniela Bolboacă[2]

[1]Technical University of Cluj-Napoca, Romania, E-mail: lori@academicdirect.org
[2]"Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca, Romania, E-mail: sorana@sorana.academicdirect.ro

Aim: The paper presents the abilities in estimation and prediction of octanol-water partition coefficient of some para-substituted phenols through integration of complex structure information by the use of an original molecular descriptors family on structure-property relationships.

Material and Method: The measured octanol-water partition coefficient of thirty para-substituted phenols, express in logarithmical scale was taken from a previous study [1]. All para-substituted phenols were used in order to generate and calculate the molecular descriptors family [2]. Based on the generated descriptors, the structure-property relationships models were build. The obtained models (bi-, and four-varied) were validated through the assessment of the cross-validation leave-one-out score, and analyzed through the squared correlation coefficient and the model stability. The comparison between the bi and four-varied models was performed by applying the Steiger's Z test.

Results: The characteristics of the bi-varied model are: the squared correlation coefficient (r^2) equal with 0.8943, the leave-one-out scores ($r^2_{cv\text{-}loo}$) equal with 0.8659, and the model stability (r^2- $r^2_{cv\text{-}loo}$) equal with 0.02838. The multiple linear regression analysis shows that best results are obtained in four-varied model ($\hat{Y}= 8.69\cdot10^{-2}+5.56\cdot10^{-3}\cdot isDDkGg-0.42\cdot IMmrKQg+9.41\cdot10^{-3}\cdot lPMDKQg-7.8\cdot10^{-2}\cdot lFMMKQg$): $r^2 = 0.9781$, proving to be a stable model ($r^2_{cv\text{-}loo} = 0.01$) and having best prediction ability ($r^2_{cv\text{-}loo}= 0.9680$). The correlation coefficient obtained by the four-varied model is statistically significant higher comparing with the correlation coefficient obtained by the bi-varied model (Steiger's Z parameter = 4.3501, p-value = $6.81\cdot10^{-6}$).

Conclusion: The four-varied model indicates that the octanol-water partition coefficient of studied para-substituted phenols is like to be of geometry nature, is strongly dependent on the partial change of compounds and group electronegativity and it is in relation with elastic force.

[1] T.W. Schultz, Relative toxicity of para-substituted phenols: Log K_{OW} and pKa-dependent structure-activity relationships. Bull. Environ. Contam. Toxicol., 1987, 38, 994-9.
[2] L. Jäntschi, Molecular Descriptors Family on Structure Activity Relationships 1. The review of Methodology, Leonardo Electronic Journal of Practices and Technologies, 2005, 6, 76-98.

Modeling the Octanol-Water Partition Coefficient of Substituted Phenols by the Use of Structure Information

Lorentz JÄNTSCHI[*], Sorana-Daniela BOLBOACĂ[**]

[*]Technical University of Cluj-Napoca, Romania, http://lori.academicdirect.org

[**]"Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca, Romania, http://sorana.academicdirect.ro

Aim

The poster presents the abilities in estimation and prediction of octanol-water partition coefficient of some para-substituted phenols through integration of complex structure information by the use of an original molecular descriptors family on structure-property relationships.

Materials and Methods

The measured octanol-water partition coefficient of thirty para-substituted phenols, express in logarithmical scale was taken from a previous study [1]. All para-substituted phenols were used in order to generate and calculate the molecular descriptors family [2]. Based on the generated descriptors, the structure-property relationships models were build. The obtained models (bi-, and four-varied) were validated through the assessment of the cross-validation leave-one-out score, and analyzed through the squared correlation coefficient and the model stability. The comparison between the bi and four-varied models was performed by applying the Steiger's Z test.

Results

Two MDF-SAR models, one bi- and one four-varied, proved to have abilities in estimation and prediction of octanol-water partition coefficient for para-substituted phenols. The MDF-SAR models are:

- Bi-varied: $Y_{2v} = 1.07 + 3.38 \cdot 10^{-3} \cdot isDDkGg - 0.40 \cdot lMmrKQg$ (1)
- Four-varied: $Y_{4v} = 8.69 \cdot 10^{-2} + 5.56 \cdot 10^{-3} \cdot isDDkGg - 4.16 \cdot 10^{-1} \cdot lMmrKQg + 9.41 \cdot 10^{-3} \cdot lPMDKQg - 7.80 \cdot 10^{-2} \cdot lFMMKQg$ (2)

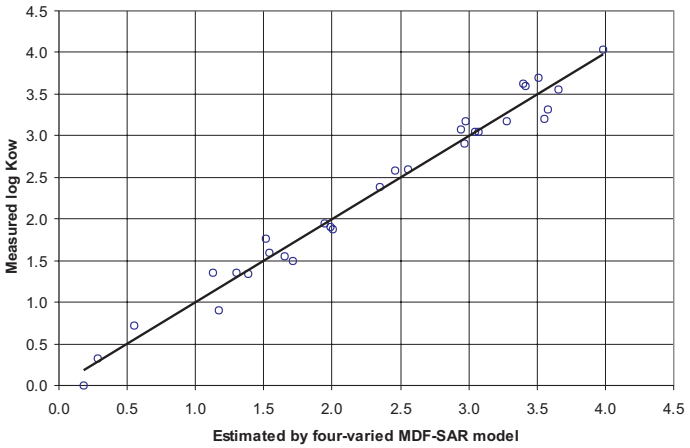
Statistical characteristics of bi- and four-varied MDF-SAR models are in Table 1 The plot of the estimated log K_{ow} by four-varied MDF-SAR model versus measured log K_{ow} is in Figure 1.

Table 1. Statistical characteristics of the MDF-SAR models

Parameter	Value	
	Bi -varied (n = 30, v = 2)	Four -varied (n = 30, v = 4)
r (correlation coefficient)	0.9457	0.9890
CI$_r$ [lower, upper] (95% confidence intervals)	[0.8897, 0.9740]	[0.9767, 0.9948]
r^2 (squared correlation coefficient)	0.8943	0.9781
r^2_{adj} (adjusted correlation coefficient)	0.8865	0.9745
s_{ser} (standard error)	0.3671	0.1739
F_{ser} (Fisher parameter of regression model)	114	279
p_{ser} (F significance)	< 0.0001	< 0.0001
r^2_{cvloo} (cross -validation loo squared correlation coefficient)	0.8660	0.9680
s_{loo} (standard error of cross -validation loo analysis)	0.4139	0.2101
F_{pred} (Fisher p arameter in loo analysis)	87	189
p_{pred} (%) (F significance in loo analysis)	< 0.0001	< 0.0001
$r^2 - r^2_{cvloo}$ (measure of model stability)	0.028 4	0.0100
r^2(*isDDkGg* , *lMmrKQg*)*	0.0760	0.0760
r^2(*isDDkGg* , *lPMDKQg*)*	n.a.	0.3346
r^2(*isDDkGg* , *lFMMKQg*)*	n.a.	0.0083
r^2(*lMmrKQg* , *lPMDKQg*)*	n.a.	0.0232
r^2(*lMmrKQg* , *lFMMKQg*)*	n.a.	0.1624
r^2(*lPMDKQg* , *lFMMKQg*)*	n.a.	0.6214

n = number of compounds, v = number of descriptor, * =squared correlation coefficient between descriptors , n.a. = not applicable

Figure 1. Estimated versus measured log Kow



The correlation coefficient obtained by the four-varied model is statistically significant higher comparing with the correlation coefficient obtained by the bi-varied model (Steiger's Z parameter = 4.3501, p-value = $6.81 \cdot 10^{-6}$).

Validation of the four-varied MDF-SAR model was performed by spilling the set of para-substituted phenols in training and test sets. The characteristics of the regression models and theirs performances in training and test sets are in Table 6. There were included into table 2.

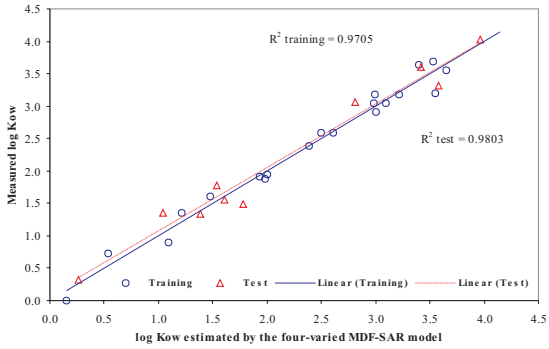
The abilities of four-varied MDF-SAR model in estimation and prediction obtained in training versus test analysis, when the number of compounds in training set was equal with 2/3 from the total number of compounds is in Figure 2.

Table 2. Statistical characteristics of models in training versus test analysis

Training sets				Test se ts				$Z_{tr\ ts}$
No_{tr}	r_{tr}	95%CI r_{tr}	F_{tr}	No_{ts}	r_{ts}	95%CI r_{ts}	F_{ts}	
15	0.9958	[0.9870, 0.9986]	299*	15	0.9808	[0.9416, 0.9937]	52*	1.871*
16	0.9891	[0.9679, 0.9963]	124*	14	0.9855	[0.9534, 0.9955]	35*	1.197
17	0.9858	[0.9600, 0.9949]	104*	13	0.9880	[0.9591, 0.9965]	80*	0.630
18	0.9877	[0.9665, 0.9955]	130*	12	0.9875	[0.9545, 0.9965]	43*	0.019
19	0.9922	[0.9793, 0.9970]	221*	11	0.9807	[0.9249, 0.9951]	29*	1.053
20	0.9901	[0.9745, 0.9961]	186*	10	0.9851	[0.9360, 0.9965]	39*	0.458
21	0.9918	[0.9794, 0.9967]	241*	9	0.9834	[0.9203, 0.9966]	20**	0.753
22	0.9856	[0.9649, 0.9941]	145*	8	0.9921	[0.9552, 0.9986]	34**	0.600
23	0.9881	[0.9716, 0.9950]	186*	7	0.9882	[0.9191, 0.9983]	20**	0.007

* = $p < 0.0001$,** = $0.0001 < p < 0.05$, = $p > 0.05$

Figure 2. Prediction and estimation abilities of four-varied MDF-SAR model in training versus test analysis



Conclusions

Octanol-water partition coefficient of para-substituted phenols, being a property of chemicals, is proved to be related with complex information obtained from the compounds structure, being strongly related with compounds geometry, theirs partial change and elastic force, in relationship with group electronegativity and inverse related with the property potential.

The goodnes-of-fit of the four-varied MDF-SAR model and internal validation results sustain that the model is stable and valid. Future studies on new external para-substituted phenols are necessary in order to assess the four-varied model robustness and predictivity.

Acknowledgement

Research was partly supported by UEFISCSU Romania through project ET36/2005.

References

[1] T.W. SCHULTZ, Relative toxicity of para-substituted phenols: Log K_{OW} and pKa-dependent structure-activity relationships. Bull. Environ. Contam. Toxicol., 1987, 38, 994-9.

[2] L. JÄNTSCHI, Molecular Descriptors Family on Structure Activity Relationships 1. The review of Methodology, Leonardo Electronic Journal of Practices and Technologies, 2005, 6, 76-98.