QSAR models for organometallic compounds and salts

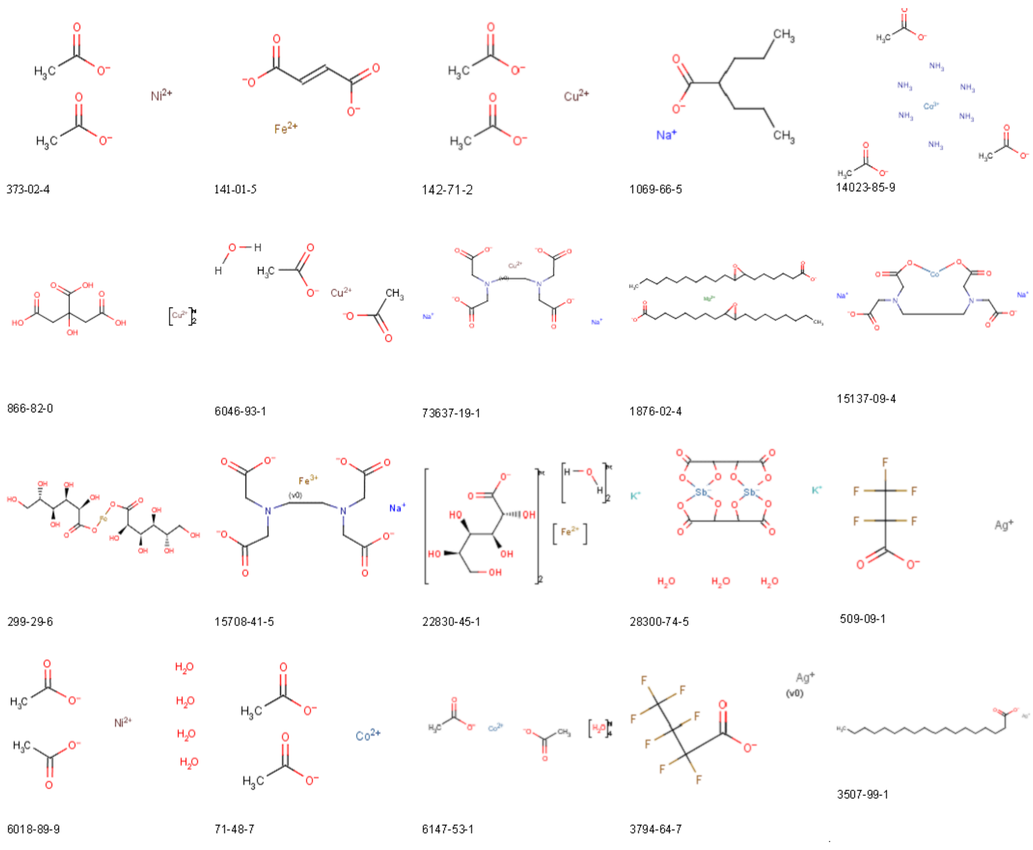
A.A. Toropov, A.P. Toropova and E. Benfenati

Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy

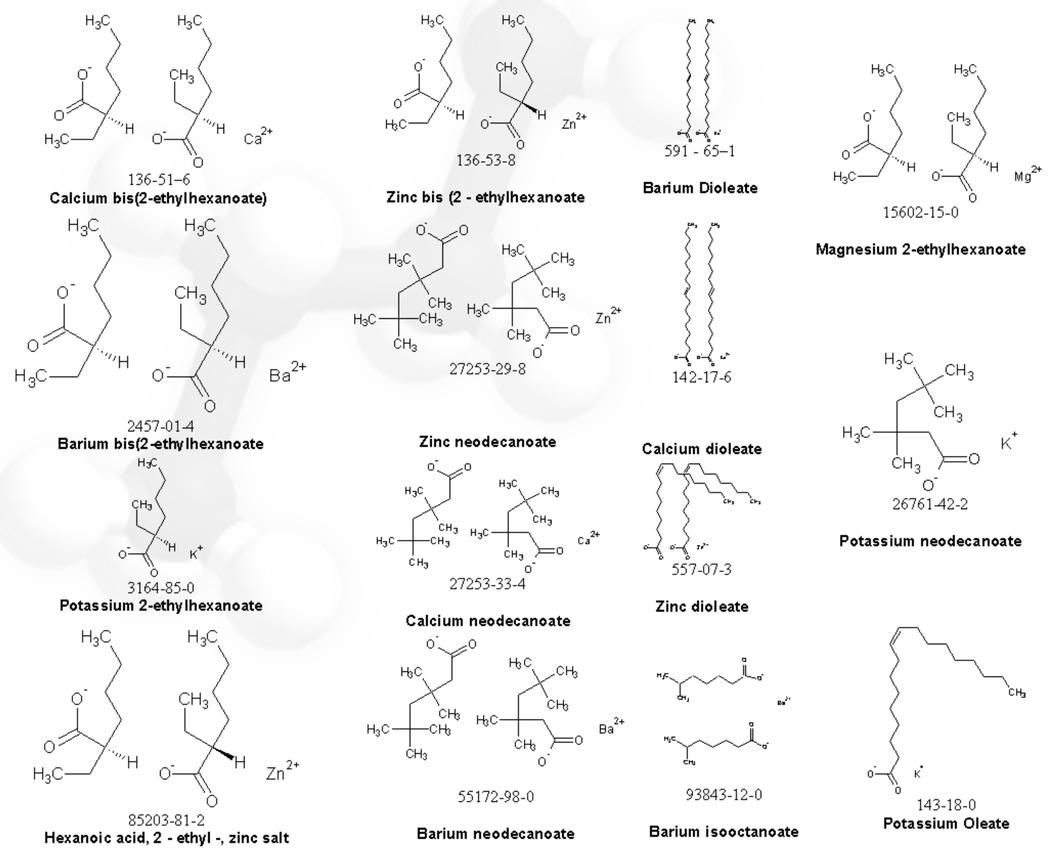
Abstract

Typically QSAR models are done on neutral compounds, and salts are transformed into the neutral form. Quite often modeling metals is difficult. Thus, the number of models on salts and organometallic compounds is very scarse. Here SMILES-based optimal descriptors have been used to model octanol/water partition coefficient and toxicity toward rats of organometallic compounds. This approach has been checked with three splits into the training and test sets. Models with very similar and satisfactory statistical quality for both the training and test sets have been obtained.

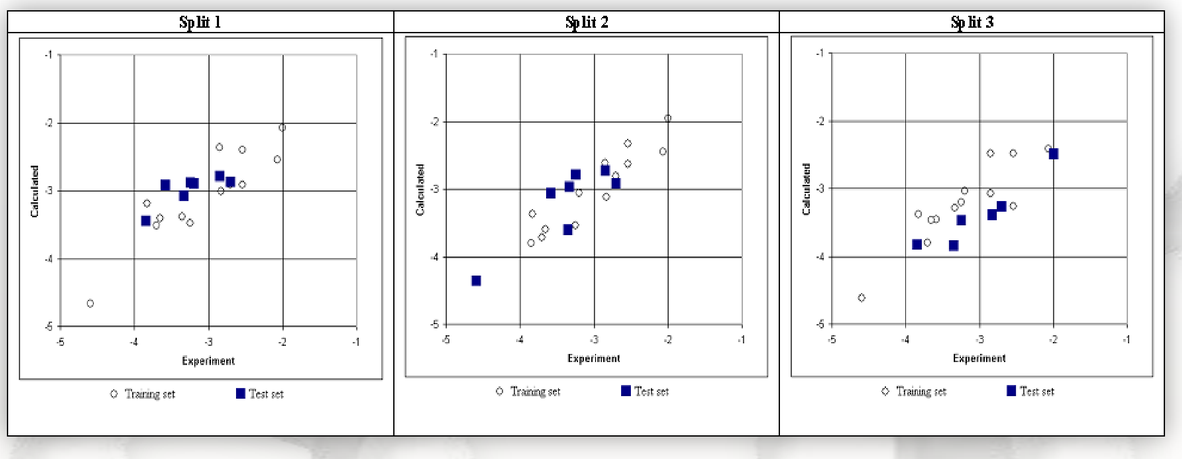
Substances and CAS numbers taken from US National Medicinal Library (TOXNET, 2008) database used to constuct the model



Substances for which experimental toxicity data is not available



Model of toxicity



Method

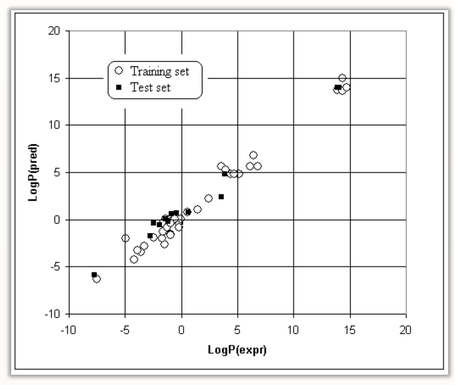
The optimal SMILES-based descriptors[1-3] are calculated as a mathematical function of SMILES-attributes:

DCW(Threshold) = F(SMILES)

The Threshold (0,1,2,3,…) are used to define rare attributes which should be blocked (in order to avoid overtraining). The **F(SMILES)** is a summation or multiplication of the correlation weights for not rare attributes. The numerical data on the correlation weights are calculated by the Monte Carlo method. The correlation weights are provider of the maximum of correlation coefficient between the DCW and toxicity (pLD50). The model has been checked with an external test set. The statistical characteristics for these models are shown in Table. In addition this approach gave a good model for octanol/water partition coefficient of these Compounds (logP).

Model of logP

Threshold	Nact	Training set n	Training set r^2	Training set s	Training set F	Test set n	Test set r^2	Test set s	Test set F
split 1									
1	28	13	0.9776	0.112	484	7	0.1801	0.499	1
2	20	13	0.8532	0.286	64	7	0.4692	0.378	4
3	**16**	**13**	**0.8004**	**0.333**	**44**	**7**	**0.6111**	**0.380**	**8**
4	13	13	0.8172	0.319	49	7	0.4698	0.404	4
5	10	13	0.7957	0.337	43	7	0.5318	0.378	6
split 2									
1	28	13	0.9956	0.041	2802	7	0.7450	0.578	15
2	**19**	**13**	**0.8578**	**0.239**	**67**	**7**	**0.7727**	**0.346**	**17**
3	14	13	0.7909	0.290	42	7	0.7281	0.500	13
4	11	13	0.7952	0.287	43	7	0.7447	0.721	15
5	9	13	0.7785	0.299	39	7	0.7252	0.781	13
split 3									
1	32	13	0.9764	0.102	460	7	0.4283	0.786	4
2	20	13	0.8899	0.221	89	7	0.7388	0.696	14
3	**13**	**13**	**0.7928**	**0.304**	**42**	**7**	**0.9137**	**0.496**	**56**
4	10	13	0.7738	0.317	38	7	0.8386	0.487	26
5	9	13	0.7651	0.323	36	7	0.8087	0.484	21



References

[1] A. Toropov, E. Benfenati Bioorg. Med. Chem. 16 (2008) 4801–4809

[2] A.A. Toropov, A.P. Toropova, E. Benfenati Chem. Phys. Lett. 461 (2008) 343–347

[3] A.A. Toropov, A.P. Toropova and E. Benfenati Chem. Biol. Drug Des. 73 (2009) 301–312

Conclusions

Optimal SMILES-based descriptors are a tool for the QSAR models of inorganic compounds for logP and toxicity toward rats (pLD50). There is a reproducibility of these models in series of the Monte Carlo optimization and for three splits into the training and the test sets. This is the first example of QSAR models of salts and organometallic compounds.

Acknowledgements

The authors thank the *Marie Curie Fellowships*, through the contract MIF1-CT-2006-039036 - CHEMPREDICT, the EC funded project CAESAR (contract SSPI-022674) and FEDERCHIMICA for financial support

FEDERCHIMICA CONFINDUSTRIA

CAESAR ChemPredict