K205 Ecotoxicity Prediction using 3D-descriptors

Yoshihiro OMAYU[1], Hiroaki SHIRAISHI[2], Kousuke OKAMOTO[1], Shinnosuke HIDAKA[1], Junichi TAKAHARA[1], Norihito KAWASHITA[1,4], Tatsuya TAKAGI[1,3,4*]

1) *Graduate school of Pharmaceutical Sciences, Osaka University,* 1-6 Yamadaoka, Suita 565-0871 JAPAN
2) *National Institute for Environmental Studies,*16-2 Onogawa, Tsukuba, 305-8506, JAPAN
3) *Research Institute for Microbial Diseases, Osaka University*, 3-1 Yamadaoka, Suita 565-0871 JAPAN
4) *Research Collaboration Center on Emerging and Re-emerging Infections*, 3-1 Yamadaoka, Suita 565-0871 JAPAN

ttakagi@phs.osaka-u.ac.jp

1. Introduction

Today, the use of [Quantitative] structure-activity relationships ([Q]SARs) for predicting (eco)toxicity is attracting attention in the fields of toxicology, because we can cut down on time and cost of experiment. Although ECOSAR [1], TIMES [2] and KATE [3] predict daphnia acute toxicity by using one or two traditional descriptors, such as logPow (Ecosar TIMES and Kate) and LUMO (lowest unoccupied molecular orbital, TIMES), there is room for improvement of QSAR in the accuracy of prediction and applicability for various kinds of chemicals.

In this study, using some descriptors including 3D descriptors which have been used recently, we expected more information of chemicals can reflect in QSAR model. Therefore we tried to develop more reliable QSAR model using 3D descriptors.

2. Method

2.1. Dataset

We used the results of ecotoxicity test (Oct. 2005) which published by the Ministry of the Environment [4]. An initial screening was done in order to avoid electrolyte, inorganic substances, and chemicals with unspecified value. Then, the dataset contains 265 chemicals. And they are classified into three categories (aromatic amines, aliphatic amines, others). We tried to predict acute toxicity towards daphnia (EC_{50} [mg/L]). Acute toxicity value were convert to molar concentration and negative of the logarithm of EC_{50} [M].

2.2. Analysis

Each chemical structure was checked and downloaded from Nikkaji Web site and ChemIDplus Web site, then saved as an MDL mol file [5, 6]. Using MOE2006.08 (CCG Inc.), descriptors were calculated after energy minimization. (3D : 42, 2D : 183). There is no experimental data about logarithm of octanol/water-partition coefficients (log*P*), thus, we used calculated log*P* by MOE2006.08.

And in this study, using dummy variables, it is possible to develop one QSAR model without classification of chemicals. The statistical analyses were carried out

using multiple linear regression (MLR) and partial least squares regression (PLSR) using R 2.5.1 with package pls.

The QSAR model was evaluated by leave-one-out cross-validation R^2 (Q^2). And, we count the number of chemicals that their residuals are less 10% of range of experimental EC_{50} (logarithm). (min -3.19, MAX 3.89)

3. Results and Discussion

As an example, we show the results of analysis using 3D descriptors and log*P*.

Table.1 shows the results of MLR and PLSR. Figure.1 shows correlation of experimental and predicted toxicity.

3.1 MLR

Descriptors were decreased from 45 to 29, using stepwise selection, and then, 10 descriptors were selected by significance of coefficient.

3.2 PLSR

We tried two way, using all descriptor, and some descriptors selected by step-up procedure method. Figure shows the prediction using 20 descriptors selected by step-up.

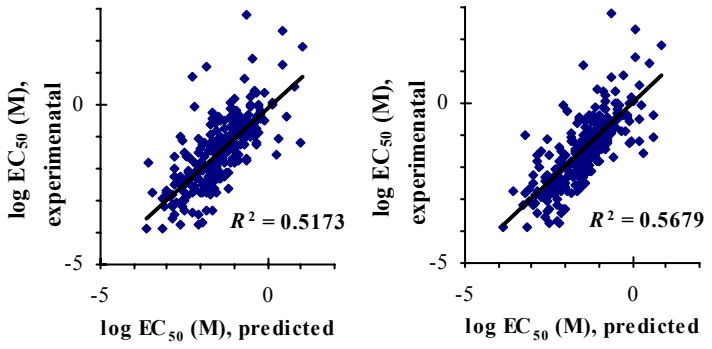

Fig.1 Correlation of experimental and predicted toxicity (right : MLR, left : PLSR)

3.3 Discussion

The QSAR models using 3D descriptors show higher performance than simple linear model using log*P*. The best model in them is PLSR using 20 descriptors selected by step-up. In addition, the residuals of 192 chemicals are less 10% of range of experimental EC_{50}, 72.5% of chemicals can predict their toxicity.

Using dummy variables of structure information, it is possible to predict toxicity enough, without developing separate QSAR models.

Table. 1 The results of analysis

	Descriptors [number of descriptors]	Q^2	N ($\|e^*\|$<0.7) (%)	N ($\|e^*\|$>2) (%)
MLR	Stepwise + decreasing [10]	0.516	185 (70.0)	6 (2.3)
PLSR	All [45]	0.446	180 (68.0)	9 (3.4)
	Step-up [20]	0.568	192 (72.5)	6 (2.3)
LM	descriptor "log*P*"	0.423	173 (65.3)	7 (2.6)

* $e = y_i - \hat{y}_i$, *residual*

References

1. http://www.epa.gov/oppt/newchems/tools/21ecosar.htm
2. http://www.oasis-lmc.org/?section=software&swid=4
3. http://kate.nies.go.jp/index.html
4. http://www.env.go.jp/chemi/sesaku/02.pdf (update , Mar 2008)
5. http://chem.sis.nlm.nih.gov/chemidplus/
6. http://nikkajiweb.jst.go.jp/nikkaji_web/pages/top.html