Bull Environ Contam Toxicol (2009) 83:520–524
DOI 10.1007/s00128-009-9811-8

Mechanism-Based Quantitative Structure–Activity Relationships on Toxicity of Selected Herbicides to *Chlorella vulgaris* and *Raphidocelis subcapitata*

Guanghui Ding · Xue Li · Fan Zhang · Jingwen Chen · Liping Huang · Xianliang Qiao

Received: 29 December 2008 / Accepted: 22 June 2009 / Published online: 7 July 2009
© Springer Science+Business Media, LLC 2009

Abstract Four quantitative structure-activity relationships were developed for toxicity of selected photosynthesis (PHS) inhibitors and acetolactate synthase (ALS) inhibitors to *Chlorella Vulgaris* and *Raphidocelis subcapitata* using a mechanism-based approach. These models have good fitness and predictive ability. The potential of electron transfer, intermolecular interactions with weak electron-transfer, and intermolecular dispersive interactions between PHS inhibitors and the active site of action are key factors influencing the toxicity of these PHS inhibitors. Intermolecular weak electron-transfer interactions and intermolecular dispersive interactions mainly determine the toxicity of these ALS inhibitors. Sulfonyl is an important functional group governing the toxicity of ALS inhibitors investigated.

Keywords Herbicides · Toxicity · QSARs · Green algae

Herbicides provide an effective and economical means of weed control, and have been widely used in agriculture, landscape turf management, gardening, and hard surfaces maintenance (Gerecke et al. 2002; Kempenaar and Spijker 2004). Through a variety of mechanisms, including overspray, drift, surface runoff and leaching, they may enter various aquatic ecosystems (Gerecke et al. 2002; Leu et al. 2004). Adverse effects of these herbicides on non-target organisms of aquatic ecosystems, especially algae, are of special concern, and have to be investigated to assess their environmental risk. Thus, many researches have focused on the toxicity of herbicides to algae (Ma et al. 2002, 2006; Junghans et al. 2003). Among these studies, Ma et al. (2002, 2006) extensively studied 96-h acute toxicity of 40 herbicides to *Chlorella vulgaris* and *Raphidocelis subcapitata*.

As experimental determination of toxicity is costly and time-consuming, it is desirable to develop mathematical predictive relationships to theoretically quantify toxicity. Quantitative structure-activity relationship (QSAR) provides a convenient tool for toxicity evaluation and prediction, and can also give some insight into the mechanism of toxic actions (Schultz et al. 2003a; Eriksson et al. 2003). Two approaches, an analog-based approach and a mechanism-based approach, can be used for QSAR modeling. The former develops QSAR models for a series of homologous or congeneric compounds. However, structural similarity does not necessarily imply ecotoxicological similarity, and inclusion of compounds having different toxicity mechanisms may deteriorate the quality of QSAR models (Mekapati and Hansch 2002). Therefore it is appropriate to combine chemicals by their toxicity mechanisms, instead of chemical classes, to develop high-quality QSARs (Escher and Hermens 2002; Schultz et al. 2003a, b). The purpose of this study is to investigate toxicity of selected herbicides to green algae *Chlorella Vulgaris* and *Raphidocelis subcapitata* using the mechanism-based QSAR approach.

G. Ding · X. Li · F. Zhang · J. Chen (✉) · L. Huang · X. Qiao
Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), Department of Environmental Science and Technology, Dalian University of Technology, Linggong Road 2, 116024 Dalian, People's Republic of China
e-mail: jwchen@dlut.edu.cn

G. Ding
e-mail: guanghuid2001@yahoo.com.cn

G. Ding
College of Environmental Science and Engineering, Dalian Maritime University, Linghai Road 1, 116026 Dalian, People's Republic of China

Materials and Methods

The 96-h toxicity data of 40 herbicides on *Chlorella vulgaris* and *Raphidocelis subcapitata* were taken from Ma et al. (2002, 2006). These 40 herbicides have different modes of toxic action. Thus, only two groups of the herbicides could be employed for QSAR establishment for they have comparatively large data sets to warrant regression analyses. These herbicides are 9 photosynthesis (PHS) inhibitors and 9 acetolactate synthase (ALS) inhibitors with toxicity data on *Chlorella vulgaris*, and 11 PHS inhibitors and 9 ALS inhibitors with toxicity data on *Raphidocelis subcapitata*. As paraquat (a PHS inhibitor) and bispyribac-sodium (an ALS inhibitor) are ionic compounds, which are different from the other herbicides, they were excluded from QSAR analyses. Thus the following QSAR studies focused on toxicity of selected PHS inhibitors and ALS inhibitors to *Chlorella Vulgaris* and *Raphidocelis subcapitata*, the CAS numbers and EC_{50} values of which are listed in Table 1. The EC_{50} values were converted to logarithmic form before model development.

Generally, chemical toxicity is the result of two key steps: partitioning of toxicant into/through the biological membrane, and interactions of the toxicant with the site of action. Thus it is important to select suitable molecular structural descriptors to characterize these processes in QSAR studies.

The partitioning processes of toxicants can be described by the *n*-octanol/water partition coefficient (K_{OW}). log K_{OW} values of these herbicides were estimated using the software KOWWIN (Version 1.67, EPI Suite, US-EPA). The interactions of the toxicant with the sites of action could be described by 17 molecular structural descriptors, the definitions of which are listed in Table 2. Among the 17 descriptors, 13 quantum-chemical descriptors were computed using PM3 Hamiltonian algorithm of MOPAC (2000), and the other 4 molecular structural descriptors were obtained by software CS Chem3D Ultra (Version 6.0, www.camsoft.com). For ALS inhibitors, two special descriptors were considered to characterize their special structures, which were the net atomic charges of the sulfur atom (q_S^+) and the average net atomic charges of oxygen atoms on sulfonyl (q_O^-). These two descriptors were also calculated by CS Chem3D Ultra.

When too many molecular structural descriptors were included in QSAR models, multicollinearity among descriptors may occur, and classical multiple linear regression will not provide predictive and robust models. Thus partial least squares (PLS) regression was adopted, for this method can analyze data with strongly collinear, noisy and numerous predictor variables. In this study, the PLS analyses were carried out by Simca-S (Version 6.0, Umetri AB & Erisoft AB). Simca-S employs "cross validation" to determine the number of PLS components (A). In the end, a statistic Q^2_{cum} is provided for a PLS model, which denotes the cumulative variance of the dependent variable explained by the extracted PLS components, and is a good measurement of the predictive power and robustness of the model. When Q^2_{cum} of a model is larger than 0.5, the model is considered to be predictive and robust.

If insignificant descriptors are included in a PLS model, the prediction ability and robustness of the model may decrease, and the interpretation of the model becomes difficult. It is therefore necessary to eliminate redundant descriptors and identify important descriptors. The following variable selection procedure was adopted: At first, correlation analyses were performed between log EC_{50} and all the predictor variables to find the most significant variable. A simple linear regression between log EC_{50} and the most significant variable were established. Then one other variable was added and a PLS model was built. This step was repeated until every remaining variable had been added once and only once. Q^2_{cum} values of this series of models were compared and the model with the highest Q^2_{cum} was selected to enter the next step. The variable-addition and

Table 1 The CAS numbers and log EC_{50} values of selected herbicides

No.	Herbicides	CAS number	EC_{50} (mol/L)[a]	EC_{50} (mol/L)[b]
Photosynthesis (PHS) inhibitors				
1	Ametryne	834-12-8		5.45×10^{-8}
2	Atrazine	1912-24-9	1.92×10^{-6}	5.60×10^{-7}
3	Bromoxynil	1689-84-5	2.83×10^{-4}	2.27×10^{-5}
4	Chlorotoluron	15545-48-9	1.19×10^{-7}	4.00×10^{-9}
5	Cyanazine	21725-46-2	5.35×10^{-7}	2.48×10^{-7}
6	Diuron	330-54-1	1.84×10^{-8}	3.01×10^{-9}
7	Isoproturon	34123-59-6	1.09×10^{-7}	6.58×10^{-8}
8	Methabenzthiazuron	18691-97-9		9.46×10^{-8}
9	Prometryne	7287-19-6	2.22×10^{-7}	4.81×10^{-8}
10	Simazine	122-34-9	1.08×10^{-5}	3.73×10^{-6}
Acetolactate synthase (ALS) inhibitors				
11	Bensulfuron-methyl	83055-99-6	4.33×10^{-5}	3.31×10^{-5}
12	Chlorimuron-ethyl	90982-32-4	4.64×10^{-5}	1.33×10^{-5}
13	Cyclosulfamuron	136849-15-5	7.03×10^{-7}	9.59×10^{-7}
14	Ethametsulfuron	111353-84-5	1.64×10^{-4}	7.36×10^{-5}
15	Flumetsulam	98967-40-9	3.28×10^{-5}	7.34×10^{-6}
16	Metsulfuron-methyl	74223-64-6		6.39×10^{-5}
17	Nicosulfuron	111991-09-4	1.06×10^{-5}	3.49×10^{-6}
18	Pyrazonsulfuron-ethyl	93697-74-6	4.54×10^{-5}	2.68×10^{-5}
19	Tribenuron	106040-48-6	9.70×10^{-5}	

[a] EC_{50} of herbicides to *Chlorella Vulgaris*

[b] EC_{50} of herbicides to *Raphidocelis subcapitata*

Table 2 Definitions of the molecular structural descriptors

No.	Descriptors	Descriptions
1	Mw	Molecular weight (atomic mass units)
2	α	Average molecular polarizability (atomic units)
3	μ	Dipole moment (Debye)
4	ΔH_f	Standard heat of formation (kJ mol^{-1})
5	TE	Total energy (electron Volts, eV)
6	EE	Electronic energy (eV)
7	CCR	Core-core repulsion energy (eV)
8	E_{HOMO}	The energy of the highest occupied molecular orbital (eV)
9	E_{LUMO}	The energy of the lowest unoccupied molecular orbital (eV)
10	q_C^-	The most negative net atomic charges on a carbon atom (atomic charge unit, a.c.u.)
11	q_H^+	The most positive net atomic charges on a hydrogen atom (a.c.u.)
12	q_N^-	The most negative net atomic charges on a nitrogen atom (a.c.u.)
13	q_{max}^+	The most positive net atomic charges on a atom (a.c.u.)
14	CAA	Connolly accessible area ($Å^2$)
15	CMA	Connolly molecular area ($Å^2$)
16	$CSEV$	Connolly solvent-excluded volume ($Å^3$)
17	Ov	Ovality is the ratio of the molecular surface area to the minimum surface area

model-building processes were repeated until the number of PLS components was bigger than $n/4$ or all predictor variables had been included. The model with the highest Q^2_{cum} was selected as the optimal model from all the models obtained.

Results and Discussion

Previous studies indicated that narcotic toxicity could be explained by log K_{OW}. Therefore simple regression equations between log EC_{50} and log K_{OW} were analyzed.

For toxicity of 8 PHS inhibitors to *Chlorella vulgaris*:

$$\log EC_{50} = -7.917 + 6.346 \times 10^{-1} \log K_{OW} \quad (1)$$
$$n = 8,\ r = 0.221,\ SE = 1.401,\ F = 0.307,\ p = 0.599$$

where r is the correlation coefficient between observed and fitted values, SE is the standard error, F is the F statistic, and p is probability that r equals zero.

For toxicity of 10 PHS inhibitors to *Raphidocelis subcapitata*:

$$\log EC_{50} = -7.873 + 3.494 \times 10^{-1} \log K_{OW} \quad (2)$$
$$n = 10,\ r = 0.128,\ SE = 1.265,\ F = 0.133,\ p = 0.725.$$

For toxicity of 8 ALS inhibitors to *Chlorella vulgaris*:

$$\log EC_{50} = -4.634 + 6.562 \times 10^{-2} \log K_{OW} \quad (3)$$
$$n = 8,\ r = 0.099,\ SE = 0.787,\ F = 0.060,\ p = 0.815.$$

For toxicity of 8 ALS inhibitors to *Raphidocelis subcapitata*:

$$\log EC_{50} = -5.011 + 1.255 \times 10^{-1} \log K_{OW} \quad (4)$$
$$n = 8,\ r = 0.218,\ SE = 0.687,\ F = 0.300,\ p = 0.603.$$

However, F-test showed that log EC_{50} did not significantly correlate with log K_{OW} for these regression equations. Thus, further statistical analyses were performed. Following the aforementioned variable selection procedure, optimal QSAR models were obtained as follows.

PHS inhibitors on *Chlorella vulgaris*:

$$\log EC_{50} = -1.119 \times 10 + 3.904 \times 10^{-4} Mw + 5.154 \times 10^{-3} TE - 1.750 E_{HOMO} - 7.164 \times 10^{-1} E_{LUMO} - 3.966 q_N^- \quad (5)$$
$$n = 8,\ A = 2,\ R^2_{X(cum)} = 0.633,\ R^2_{Y(cum)} = 0.979,\ Q^2_{(cum)} = 0.911,\ r = 0.989,\ SE = 0.299$$

where A is the number of extracted PLS components, $R^2_{X(cum)}$ and $R^2_{Y(cum)}$ stand for the cumulative sum of squares of all the predictor variables and dependent variable explained by all extracted components, respectively.

PHS inhibitors on *Raphidocelis subcapitata*:

$$\log EC_{50} = -3.899 - 1.564 \times 10^{-3} Mw + 4.464 \times 10^{-3} TE - 1.075 E_{HOMO} - 8.131 \times 10^{-1} E_{LUMO} - 1.183 \times 10 q_{max}^+ \quad (6)$$
$$n = 10,\ A = 2,\ R^2_{X(cum)} = 0.623,\ R^2_{Y(cum)} = 0.934,\ Q^2_{cum} = 0.847,\ r = 0.966,\ SE = 0.351$$

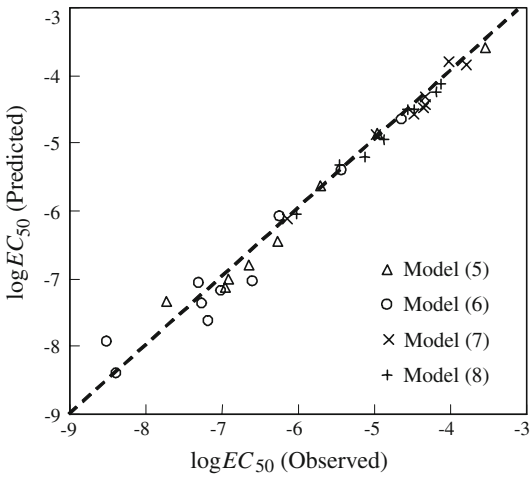

Fig. 1 Plot of predicted versus observed log EC_{50} values for 4 QSAR models

ALS inhibitors on *Chlorella vulgaris*:

$$\begin{aligned}\log EC_{50} = &-3.357\times 10-1.266\times 10^{-3}Mw\\ &-1.591\times 10^{-1}\mu+1.722\times 10^{-5}EE\\ &-1.954\times 10^{-5}CCR-6.524\\ &\times 10^{-1}E_{\mathrm{HOMO}}-3.119q_C^-\\ &-3.007\times 10q_{\mathrm{O}}^- - 7.618\times 10^{-4}CSEV \qquad (7)\end{aligned}$$

$$\mathrm{n}=8,\ A=2,\ R^2_{X(\mathrm{cum})}=0.786,\ R^2_{Y(\mathrm{cum})}=0.977,$$
$$Q^2_{\mathrm{cum}}=0.921,\ r=0.988,\ SE=0.132.$$

ALS inhibitors on *Raphidocelis subcapitata*:

$$\begin{aligned}\log EC_{50} = &-3.733\times 10-7.740\times 10^{-3}\alpha-9.436\\ &\times 10^{-1}E_{\mathrm{HOMO}}-2.794q_{\mathrm{C}}^-\\ &-2.263q_{\mathrm{S}}^+-3.548\times 10q_{\mathrm{O}}^- \qquad (8)\end{aligned}$$

$$\mathrm{n}=8,\ A=2,\ R^2_{X(\mathrm{cum})}=0.624,\ R^2_{Y(\mathrm{cum})}=0.987,$$
$$Q^2_{\mathrm{cum}}=0.887,\ r=0.993,\ SE=0.088.$$

From Fig. 1, it can be seen that the predicted log EC_{50} values agree well with the observed values for these 4 QSAR models. Furthermore, all Q^2_{cum} values are far higher than 0.5, indicating high predictability and robustness of these models. Considering the paucity of algae toxicity data for many herbicides, the difficulty or high expenditures involved in experimental determinations, these models could serve as a first and fast approximation of toxicity for the relevant herbicides with the same mode of toxic action.

For each QSAR model, 2 PLS components were extracted. From PLS weights shown in Fig. 2, one can see how predictor variables and response variable combine in PLS components, and how they relate to each other.

For toxicity of the PHS inhibitors, model (5) and (6) are similar, with only one different molecular descriptor. The first PLS components of the two models are mainly related with E_{HOMO} and E_{LUMO}, which correlate with the global readiness of a molecule to donate or accept electron charge (Wang et al. 2004). It is known that the PHS inhibitors are inhibitors of photosynthesis II (PSII), which displace the Q_B plastoquinone from its binding site on the D1 protein, and then block the electron transfer chain of PSII (Gramatica et al. 2001; Fufezan et al. 2002). So E_{HOMO} and E_{LUMO} characterize the potential of electron transfer of the herbicide molecules in governing the toxicity.

The second PLS component of model (5) is mainly related with q_{N}^-, while for model (6) is q_{max}^+ and *Mw*. q_{N}^- and q_{max}^+ relate with intermolecular interactions with weak electron-transfer, and may characterize hydrogen bonding between the PHS inhibitors and the active site of action. *Mw* has some important influence on the second PLS component of model (6). *Mw* is correlated with molecular size, and therefore with intermolecular dispersive interactions or steric hindrance. As *Mw* is in negative correlation with log EC_{50}, it primarily describes the influence of intermolecular dispersive interactions between PHS inhibitors and the active site of action on the toxicity. For model (5), *Mw* has certain influence on both PLS components. However, it reveals different effects on the toxicity via these two PLS components. In model (5), these effects were summarized, and *Mw* takes positive correlation with log EC_{50}. As the coefficient of *Mw* in model (5) is only 3.904×10^{-4}, the influence of *Mw* on the toxicity is too small to attend. In addition, *TE* has important influence on both PLS components for model (5) and (6). As *TE* is significantly negative correlated with molecular size, it could also characterize intermolecular dispersive interactions or steric hindrance. In both models, *TE* is in positive correlation with log EC_{50}, so it mainly describe intermolecular dispersive interactions between PHS inhibitors and the active site of action, which works together with the above influencing factors.

For toxicity of the ALS inhibitors, model (7) and model (8) have some common descriptors. E_{HOMO}, μ, q_{C}^- and q_{O}^- are the most important descriptors for model (7), and E_{HOMO}, q_{O}^-, q_{S}^+ and q_{C}^- for model (8). E_{HOMO}, $q^-q_{\mathrm{S}}^+$ and q_{C}^- are all related to intermolecular interactions with weak electron-transfer including hydrogen bond, and μ relates to electrostatic interactions. It is known that the ALS inhibitors inhibit the activity of the acetolactate synthase enzyme and thereby block the biosynthesis of the branched-chain amino acids valine, leucine and isoleucine (Simpson et al. 1995). Therefore these variables characterize the influence of intermolecular weak electron-transfer interactions. q_{O}^- was included in both models, however, q_{S}^+ only in model (8). This means that sulfonyl is an important functional group for toxicity of the ALS inhibitors, especially to *Raphidocelis subcapitata*.

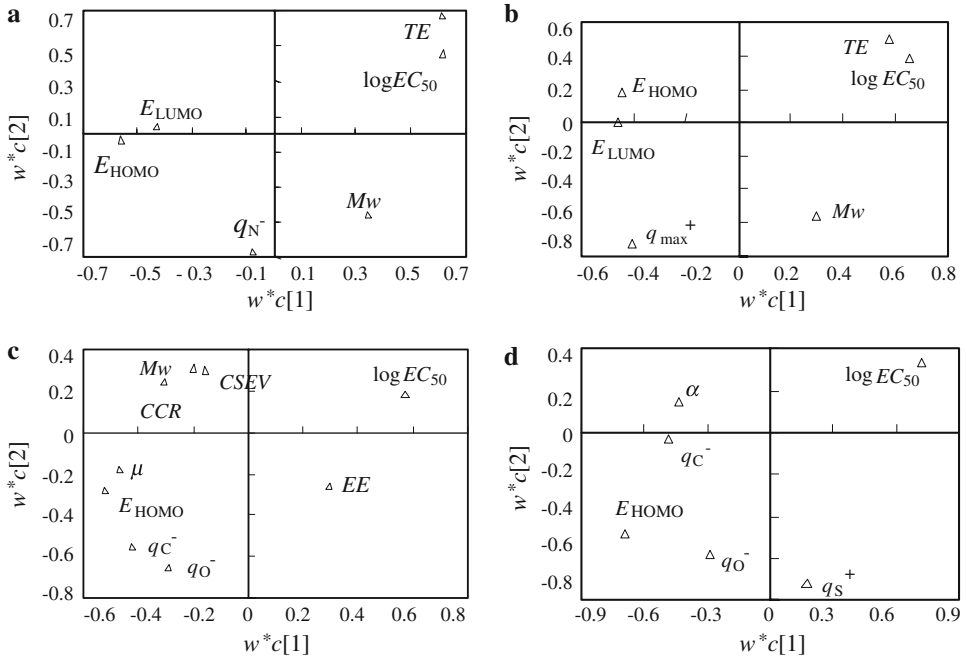

Fig. 2 Plots of PLS weights: **a** for model (5), **b** for model (6), **c** for model (7), **d** for model (8)

In addition, *Mw*, *EE*, *CCR* and *CSEV* have certain contribution to both PLS components for model (7), while α has some contribution to both PLS components for model (8). These descriptors are inter-correlated, and may represent intermolecular dispersive interactions or steric hindrance. From their coefficients, it can be determined that they may mainly describe the influence of intermolecular dispersive interactions on toxicity of these herbicides.

Acknowledgments The study was supported by the National Basic Research Program of China (973 Project 2006CB403302) and National Natural Science Foundation of P. R. China (No. 20337020).

References

Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Persp 111:1361–1375

Escher BI, Hermens JLM (2002) Modes of action in ecotoxicology: their role in body burdens, species sensitivity, QSARs, and mixture effects. Environ Sci Technol 36:4201–4217

Fufezan C, Rutherford AW, Krieger-Liszkay A (2002) Singlet oxygen production in herbicide-treated photosystem II. FEBS Lett 532:407–410

Gerecke AC, Schärer M, Singer HP, Müller SR, Schwarzenbach RP, Sägesser M, Ochsenbein U, Popow G (2002) Sources of pesticides in surface waters in Switzerland: pesticide load through waste water treatment plants–current situation and reduction potential. Chemosphere 48:307–315

Gramatica P, Vighi M, Consolaro F, Todeschini R, Finizio A, Faust M (2001) QSAR approach for the selection of congeneric compounds with a similar toxicological mode of action. Chemosphere 42:873–883

Junghans M, Backhaus T, Faust M, Scholze M, Grimme LH (2003) Toxicity of sulfonylurea herbicides to the green alga *Scenedesmus vacuolatus*: predictability of combination effects. Bull Environ Contam Toxicol 71:585–593

Kempenaar C, Spijker JH (2004) Weed control on hard surfaces in The Netherlands. Pest Manag Sci 60:595–599

Leu C, Singer H, Stamm C, Müller SR, Schwarzenbach RP (2004) Simultaneous assessment of sources, processes, and factors influencing herbicide losses to surface waters in a small agricultural catchment. Environ Sci Technol 38:3827–3834

Ma JY, Xu LG, Wang SF, Zheng RQ, Jin SH, Huang SQ, Huang YJ (2002) Toxicity of 40 herbicides to the green alga *Chlorella vulgaris*. Ecotox Environ Safe 51:128–132

Ma JY, Wang SF, Wang PW, Ma LJ, Chen XL, Xu RF (2006) Toxicity assessment of 40 herbicides to the green alga *Raphidocelis subcapitat*. Ecotox Environ Safe 63:456–462

Mekapati SB, Hansch C (2002) On the parametrization of the toxicity of organic chemicals to *Tetrahymena pyriformis*. The problem of establishing a uniform activity. J Chem Inf Comput Sci 42:956–961

Schultz TW, Cronin MTD, Walker JD, Aptula AO (2003a) Quantitative structure-activity relationships (QSARs) in toxicology: a historical perspective. J Mol Struct (Theochem) 622:1–22

Schultz TW, Cronin MTD, Netzeva TI (2003b) The present status of QSAR in toxicology. J Mol Struct (Theochem) 622:23–38

Simpson DM, Stoller EW, Wax LM (1995) An in vivo acetolactate synthase assay. Weed Technol 9:17–22

Wang ZY, Zhai ZC, Wang LS (2004) Quantitative structure-activity relationship of toxicity of alkyl(1-phenylsulfonyl) cycloalkanecarboxylates using MLSER model and ab initio. QSAR Comb Sci 24:211–217