Exploring Predictive QSAR Models Using Quantum Topological Molecular Similarity (QTMS) Descriptors for Toxicity of Nitroaromatics to Saccharomyces cerevisiae

ARTICLE *in* QSAR & COMBINATORIAL SCIENCE · AUGUST 2008

Impact Factor: 1.55 · DOI: 10.1002/qsar.200810028

2 AUTHORS:

Kunal Roy
Jadavpur University
290 PUBLICATIONS **4,885** CITATIONS
SEE PROFILE

Paul L A Popelier
The University of Manchester
187 PUBLICATIONS **6,953** CITATIONS
SEE PROFILE

QSAR & Combinatorial Science QCS

Exploring Predictive QSAR Models Using Quantum Topological Molecular Similarity (QTMS) Descriptors for Toxicity of Nitroaromatics to *Saccharomyces cerevisiae*

Kunal Roy and Paul L. A Popelier***

Manchester Interdisciplinary Biocenter (MIB), 131 Princess Street, Manchester M1 7DN, UK, E-mail: kroy@pharma.jdvu.ac.in; pla@manchester.ac.uk, Fax: +91 33 2837 1078; +44-161-3065201, Tel.: +91 9831594140; +44-161-3064511

Keywords: External validation, Nitroaromatics, QSAR, QTMS, Toxicity

Received: March 08, 2008; Accepted: May 06, 2008

DOI: 10.1002/qsar.200810028

Abstract

In view of the widespread industrial use of nitroaromatics and their consequent ecotoxicological hazard potential, we constructed predictive Quantitative Structure – Activity Relationship (QSAR) models for the toxicity of nitroaromatics to the ecologically important species *Saccharomyces cerevisiae*. We used Quantum Topological Molecular Similarity (QTMS) descriptors along with electrophilicity index (E_{LUMO}) and lipid water partition coefficient (log K_{ow}) as predictor variables. The QTMS descriptors were calculated at B3LYP/6-311 + G(2d,p) level of theory. QTMS descriptors were employed to complement the deficiency of E_{LUMO} in setting up predictive QSAR models from the view point of *external* validation. The dataset was divided into a training set (18 compounds) and test set (six compounds) in a ratio of three to one. Partial Least Square (PLS) models were developed based on the training set compounds. The predictive capacity of the models was assessed by the test compounds. The models were also validated by a randomisation test and leave-one-seventh-out crossvalidation test. The results suggest that Bond Critical Point (BCP) descriptors can develop predictive QSAR models for nitroaromatic toxicity to *Saccharomyces cerevisiae* when used along with E_{LUMO} and log K_{ow}. The diagnostic potential of QTMS descriptors could also reveal the importance of the nitro group for nitroaromatic toxicity.

1 Introduction

The global production of chemicals increases on a daily basis without a match in proper documentation in terms of environmental safety. Of the tens of thousands of current commercial chemical products, only very few have been extensively characterised and evaluated for their safety and potential toxicity [1 – 3]. The production of these chemicals is increasing rapidly but detailed data on them (as required for approval by chemical regulatory authorities) are not sufficiently available. Furthermore, test data may be inadequate, inappropriate or incomplete for many chemical substances. Human and material resources are lacking in order to obtain even basic experimental information on environmental fate and effects for all these chemicals [4]. An attractive alternative is the development of a methodology that enables predictions of effects to be made directly from chemical structure. The chemistry approach to predictive toxicology relies on Quantitative Structure – Activity Relationship (QSAR) modelling to predict biological activity from chemical structure [5 – 8]. Such approaches have proven capabilities when applied to well-defined toxicity end points or regions of chemical space. The U.S. Environmental Protection Agency uses QSARs to estimate the toxicity of existing and new chemicals, while the current European regulation is based on experimental data, employing QSARs only for specific and limited purposes such as data evaluation or the provision of additional evidence for conducting long-term tests [9]. To minimise the existing gaps in toxicity data, the European Union parliament recently adopted a new chemical control system called Registration, Evaluation and Authorisation of CHemicals (REACH) [2]. One of the aims of REACH is to improve the protection of human health and the environment by requiring industry to provide in-

** Currently on leave from Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India.

formation on toxicity of the chemicals manufactured or distributed by them. Under the new REACH [10] system to be introduced in Europe, *in vitro* techniques and QSAR models will most likely gain importance in the regulatory decision-making process [9].

Nitroaromatics are representative of electrophilic toxicants causing narcosis, mutagenicity and carcinogenicity [11]. These compounds form an important class of industrial chemicals produced in substantial volumes and with diverse use [12]. They are widely used as solvents, intermediates for the synthesis of explosives, dyestuffs, urethan polymers and other plastics as well as for the synthesis of anilines, insecticides, herbicides and pharmaceuticals [13]. Due to the many origins and magnitude of uses, nitro-substituted aromatic compounds are widespread in ecosystems and consequently have a high potential for environmental pollution. It is therefore useful to develop theoretical models to predict toxicity of nitroaromatics against different ecologically relevant organisms. Yeasts are found widely distributed in nature, playing important roles in many ecosystems and their cellular structure resembles that of higher organisms [14]. Yeasts are easy to maintain and cultivate under controlled conditions, avoiding the problems of variability found with more complex organisms [12]. These advantages make them a good alternative method for toxicity assessment. Wang *et al.* [12] developed QSAR for the toxicity of nitroaromatics to the yeast *S. cerevisiae* using the electrophilicity index (E_{LUMO}) and the octanol–water partition coefficient ($\log K_{ow}$) as predictor variables. The ultimate target of any QSAR modelling is that the model should be robust enough to make accurate and reliable predictions of the property of interest of new compounds. Hence, QSAR models, which are developed from training set compounds should be validated using new chemical entities for checking the predictive capacity of the developed models. Validation strategies check the reliability of models for their possible application on a new set of data, and confidence of prediction can thus be assessed [15, 16]. In the present paper, we further examine the predictability of the indices used by Wang *et al*. in their paper [12], which we took as a data source. Here we evaluate Quantum Topological Molecular Similarity (QTMS) indices for developing predictive models for the nitroaromatic toxicity when used along with E_{LUMO}. As QTMS does not encode for hydrophobicity we have retained log - K_{ow} in all our models. QTMS descriptors are known to model well electronic properties and activities for which the electronic factor is important [17–31]. These descriptors have been used in this communication to complement the deficiency of E_{LUMO} in obtaining predictive QSAR models from the view point of external validation.

2 Materials and Methods

The toxicity of nitroaromatics against *S. cerevisiae* [12] is adopted as the model dataset for the present QSAR study. QTMS descriptors were used as the predictor variables along with E_{LUMO} and $\log K_{ow}$, whose values were taken from the source paper [12]. We mention here that the E_{LUMO} values taken from Ref. [12] were calculated at PM3 level while QTMS descriptors reported in this communication were calculated at B3LYP/6-311+G(2d,p) level as Bond Critical Point (BCP) descriptors at this level gave better predictive models and more consistent results compared to those calculated at lower levels of theory. Table 1 lists the 24 compounds, divided in training set and test set, along with the values of E_{LUMO} and $\log K_{ow}$. The details of QTMS descriptors can be found in the previous publications [17–31]. In summary, QTMS descriptors focus on BCPs, which occur when the gradient of the electron density ($\nabla\rho=0$) vanishes at some point between two bonded nuclei. The electron density at a BCP can be related to bond order *via* an exponential relationship. At a BCP, the Hessian of ρ has two negative eigenvalues ($\lambda_1<\lambda_2<0$) and one positive one ($\lambda_3>0$). Eigenvalues express local curvature of ρ in a point: negative eigenvalues are curvatures perpendicular to the bond, while the positive eigenvalue measures the curvature along the bond. The sum of the eigenvalues is the Laplacian, denoted by $\nabla^2\rho$, which is a measure of how much ρ is concentrated ($\nabla^2\rho<0$) or depleted ($\nabla^2\rho>0$) in a point. The descriptors ρ_b and λ_3 can be interpreted as measures of σ character whilst $\lambda_1+\lambda_2$ measures the degree of π character [32]. Another measure of π character for homopolar bond is ellipticity which is defined as $\varepsilon=\lambda_1/\lambda_2-1$. In the QTMS bond descriptor vector, there are two more components, the kinetic energy density $K(\mathbf{r})$ and a more classical kinetic energy $G(\mathbf{r})$, as defined earlier [33]. Additionally, the equilibrium bond length (R_e) has also been used as one of the descriptors along with other QTMS descriptors.

Firstly, an estimated geometry was obtained using the program GaussView [34], which was then passed on to the *ab initio* program GAUSSIAN03 [35]. The BCP descriptors were calculated at B3LYP/6-311+G(2d,p) level using a local version of the program MORPHY98 [36], which locates the BCPs using an automatic and robust algorithm [37]. The BCP descriptors of nine common bonds of the nitroaromatic compounds (six C–C aromatic bonds, one C–N bond and two N–O bonds) were considered as variables for the statistical model development. The numbering scheme of the nitrobenzene skeleton is shown in Figure 1. Finally, the program SIMCA [38] was used for Partial Least Squares (PLSs) analysis of the dataset. PLS is a generalisation of regression, which can handle data with strongly correlated and/or noisy or numerous independent variables [39]. It gives a reduced solution, which is statistically more robust than Multiple Linear Regression (MLR). The linear PLS model finds 'new variables' [La-

Table 1. Observed and calculated inhibition toxicity of nitroaromatic compounds to yeast *S. cerevisiae*.

Sl. No.	Compounds	log K_{ow}[a]	E_{LUMO}[a]	Toxicity to *S. cerevisiae* (log 1/*C*)			
				Obs.[a]	Calculated Models		
					1	7	9
Training set							
2	4-Chloronitrobenzene	2.58	−1.352	1.65	1.70	1.63	1.73
3	3-Chloronitrobenzene	2.58	−1.305	1.65	1.68	1.61	1.65
4	2-Chloronitrobenzene	2.58	−1.180	1.64	1.53	1.65	1.71
5	4-Bromonitrobenzene	2.58	−1.486	2.13	1.78	1.71	1.81
6	2,4-Dinitrobromobenzene	2.70	−1.859	2.47	2.43	2.49	2.36
7	2,4-Dinitrotoluene	2.18	−1.713	2.06	2.02	2.11	1.78
8	Nitrobenzene	1.86	−1.212	1.01	1.19	1.20	1.17
9	2,4-Dinitrochlorobenzene	1.98	−1.755	2.02	2.03	2.16	1.95
13	3-Nitrotoluene	2.53	−1.222	1.52	1.55	1.44	1.45
14	2-Nitrotoluene	2.53	−1.180	1.29	1.41	1.48	1.41
15	3-Chloro-4-fluoronitrobenzene	2.71	−1.400	1.56	1.80	1.79	1.91
16	*p*-Dinitrobenzene	1.84	−4.237	3.23	3.31	3.25	3.28
17	*o*-Dinitrobenzene	1.84	−3.170	2.82	2.80	2.73	2.88
18	*m*-Dinitrobenzene	1.84	−1.547	1.45	1.39	1.43	1.56
20	4-Nitro-2-chloroaniline	1.58	−1.303	1.42	1.39	1.26	1.37
21	4-Nitroaniline	1.39	−1.202	0.96	1.09	1.00	1.11
23	2-Nitroaniline	1.37	−1.068	1.08	0.99	1.12	0.99
24	3-Nitrobenzoic acid	1.83	−1.422	1.52	1.39	1.42	1.37
Test set							
1	3,4-Dichloronitrobenzene	3.29	−1.442	2.20	2.27	2.08	2.18
10	2,6-Dinitrotoluene	2.28	−1.532	1.61	1.67	1.75	1.62
11	2,3-Dinitrotoluene	1.98	−1.500	1.97	1.64	2.09	1.66
12	4-Nitrotoluene	2.53	−1.233	1.50	1.64	1.45	1.55
19	4-Nitrophenol	1.92	−1.155	1.24	1.26	1.24	1.33
22	3-Nitroaniline	1.37	−1.220	0.88	0.92	1.08	0.82

[a] Taken from Ref. [12].

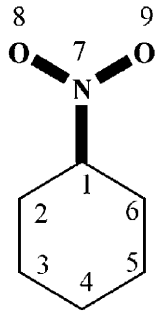

Figure 1. Numbering of atoms in the nitrobenzene skeleton.

tent Variables (LV) or independent scores] that are linear combinations of the original variables. To avoid overfitting, a strict test for the significance of each consecutive LV is necessary and no new LVs are added when they become non-significant. Crossvalidation is a practical and reliable method of testing this significance [40]. PLS models were developed for each class of descriptors, *i.e.* ρ, $\nabla^2\rho$, ε, λ, *K*, *G* and equilibrium bond lengths. Note that λ_1, λ_2 and λ_3 are clubbed together into the class of λ. There are $3 \times 9 = 27$ descriptors in the class of λ while in other classes there are only nine. At the outset, models were tried with all available descriptors, but subsequently, descriptors with smaller Variable Importance for the Projection (VIP) values were gradually deleted until a model with the best Q^2 (leave-one-seven-out crossvalidation) was obtained. Then, using important descriptors appearing in the PLS equations of different descriptor classes, the PLS model for the combined set of descriptors was developed.

For the division of the dataset into training and test sets, the compounds were ranked according to the toxicity values and every fifth compound was assigned to the test set starting with the first compound. All the developed models were crossvalidated by leave-one-seventh-out crossvalidation as default in SIMCA. The model quality was characterised by a coefficient of variation (R^2) and leave-one-seventh-out crossvalidation Q^2. In order to evaluate the prediction potential of the models, the quantity R^2_{pred} was calculated as

$$R^2_{pred} = 1 - \frac{\sum (Y_{pred(Test)} - Y_{(Test)})^2}{\sum (Y_{(Test)} - \overline{Y}_{training})^2}$$

where $Y_{pred(Test)}$ and $Y_{(Test)}$ denote, respectively, predicted and observed toxicity values of the test set compounds.

The symbol $\overline{Y}_{\text{training}}$ refers to the mean toxicity value of the training set.

We checked the criteria for external validation as recommended by Golbraikh and Tropsha [41]. These authors suggested that in addition to a high value of crossvalidated R^2 (Q^2), the correlation coefficient r between the predicted and observed activities of compounds from an external test set should be close to 1. Either the squared correlation coefficient (through the origin) between predicted and observed activities, r_0^2, or the coefficient between observed and predicted activities, $r_0'^2$, should be close to r^2. Ideally this is true for both correlation coefficients. Furthermore, at least one slope of regression lines (k or k') through the origin should be close to 1. Models are considered acceptable, if they satisfy the following three conditions: (i) $Q^2 > 0.5$, (ii) $r^2 > 0.6$ and (iii) r_0^2 or $r_0'^2$ is close to r^2, such that $(r^2 - r_0^2)/r^2$ or $(r^2 - r_0'^2)/r^2 < 0.1$ and $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$. Moreover we have checked the value of a modified r^2, denoted by r_m^2 [42, 43], which is defined by

$$r_m^2 = r^2(1 - \sqrt{r^2 - r_0^2})$$

where r_0^2 represents the squared correlation coefficient between the observed and predicted values of the test set compounds when the intercept is set to 0. Note that r^2 is always larger than r_0^2. In case of good external prediction, predicted values will be very close to observed activity values. So, the r^2 value will be very near to the r_0^2 value. In the best possible case, r_m^2 will be equal to r^2.

The final models were also subjected to a randomisation test. In this test, the toxicity data (Y) are randomly permuted keeping the descriptor matrix intact, followed by a PLS run. Each randomisation and subsequent PLS analysis generates a new set of R^2 and Q^2 values, which are plotted against the correlation coefficient between the original Y values and the permuted Y values. The intercepts for the R^2 and Q^2 lines in this plot are a measure of the overfit. A model is considered valid if $R_{\text{int}}^2 < 0.4$ and $Q_{\text{int}}^2 < 0.05$.

3 Results and Discussion

A reference PLS model was generated using E_{LUMO} and $\log K_{\text{ow}}$ as predictor variables. Models 1–9 were developed using BCP descriptors along with E_{LUMO} and $\log K_{\text{ow}}$. These models were compared to the reference model (Table 2). The reference model with one LV showed an R^2 value of 0.867 and leave-one-seventh-out crossvalidation R^2 (Q^2) value of 0.875 for the training set. However, to consider predictability, we are more interested in *external* predictability parameters. The predicted R^2 value of the reference model for the test set is moderately good (0.728). However, the r^2 value differs significantly from the r_0^2 value leading to a significantly lower r_m^2 value (0.508). This points towards a significant difference between the observed and predicted values (reference model) of the test set compounds. We strive to obtain models with better external predictability by using BCP descriptors along with E_{LUMO} and $\log K_{\text{ow}}$.

Model 1 with five distance descriptors along with $\log K_{\text{ow}}$ and E_{LUMO} gave a four-LV model with a predicted R^2 value of 0.897. The r^2 and r_0^2 values of this model being the same, it is evident that the predicted values of the test set compounds are very near to the corresponding observed values. Model 2 with ρ descriptors (along with E_{LUMO} and $\log K_{\text{ow}}$) gave a four-LV model with excellent external validation characteristics. The R_{pred}^2 value of model 2 is 0.967 while the r_m^2 value is 0.936. Models 3–6 (derived from $\nabla^2\rho$, ε, λ and K descriptors, respectively, along with E_{LUMO} and $\log K_{\text{ow}}$) do not give very promising external validation characteristics although the r_m^2 value in each case is better than that of the reference model. Model 7, which is derived from five G descriptors, E_{LUMO} and $\log K_{\text{ow}}$, is a three-LV model showing an R_{pred}^2 value of 0.932 and r_m^2 value of 0.849. Model 8 was developed, by including the four most important BCP descriptors appearing in each of the models 1–7 along with E_{LUMO} and $\log K_{\text{ow}}$. Unfortunately the external validation characteristics turned out to be only comparable to those of the reference model.

Table 2. Comparative analysis of QSAR models for toxicity of nitroaromatics to *S. cerevisiae*.

Model no.	Type of descriptors in addition to E_{LUMO} and $\log K_{\text{ow}}$	No. of descriptors	LV[a]	R^2	Q^2	R_{pred}^2	r^2	r_0^2	r_m^2
9	**All (merged descriptors)**	**8**	**4**	**0.929**	**0.837**	**0.915**	**0.910**	**0.910**	**0.910**
8	All	7	4	0.953	0.910	0.668	0.928	0.719	0.504
7	***G***	**7**	**3**	**0.940**	**0.880**	**0.932**	**0.938**	**0.929**	**0.849**
6	K_b	10	6	0.968	0.886	0.771	0.938	0.806	0.597
5	λ	8	3	0.915	0.848	0.780	0.823	0.743	0.590
4	ε	7	4	0.934	0.845	0.939	0.972	0.939	0.795
3	$\nabla^2\rho$	10	6	0.971	0.889	0.784	0.952	0.849	0.646
2	**ρ**	**11**	**4**	**0.946**	**0.878**	**0.971**	**0.967**	**0.966**	**0.936**
1	Distance	7	4	0.953	0.902	0.897	0.880	0.880	0.880
Reference	Nil	2	1	0.867	0.875	0.728	0.847	0.687	0.508

Bold models indicate the best four models according to r_m^2 values.
[a]Number of latent variables.
[b]λ_1, λ_2 and λ_3.

Models 1, 2 and 7 showed very good external validation results, which are significantly better than that of the reference model derived without BCP descriptors. This suggests that BCP descriptors have helped to increase external predictability (or true predictability) of the models. Furthermore, ρ descriptors (model 2) yield the best model suggesting the importance of electron density at BCPs for predictability. For all three models (1, 2 and 7), descriptors for the BCP of one of the two N–O bonds appear as the second most significant descriptors (E_{LUMO} being the most significant in all the cases) according to the VIP values (see Figure 2). This indicates the importance of the nitro group (shown with bold lines in Figure 1) for the toxicity of nitroaromatics. The results reveal the diagnostic potential of QTMS descriptors in revealing important fragments of molecules contributing to the response property. To further check the importance of the nitro group, BCP descriptors of a particular class (such as ρ, $\nabla^2\rho$, *etc.*) of the two N–O bonds and the C–N(nitro) bond were clubbed together to form seven new descriptors from which model 9 with four LVs was constructed. Model 9 showed excellent external predictability, though marginally inferior to model 2. Model 9 was developed only from descriptors for the three bonds involving the nitro group. This reconfirms the importance of the nitro group for the toxicity.

The calculated values of nitroaromatic toxicity to *S. cerevisiae* according to models 1, 2, 7 and 9 are given in Table 1. Selected models (1, 2, 7 and 9) were subjected to randomisation test with 100 permutations (default is 20) in each case. For all the models tested, the Q^2_{int} values are less than 0.05 and the R^2_{int} values are less than 0.4, except for model 2 (Table 3). Varying the number of permutations does not change this conclusion. Hence, model 2 must be discarded. Only models 1, 7 and 9 are *not* obtained by chance.

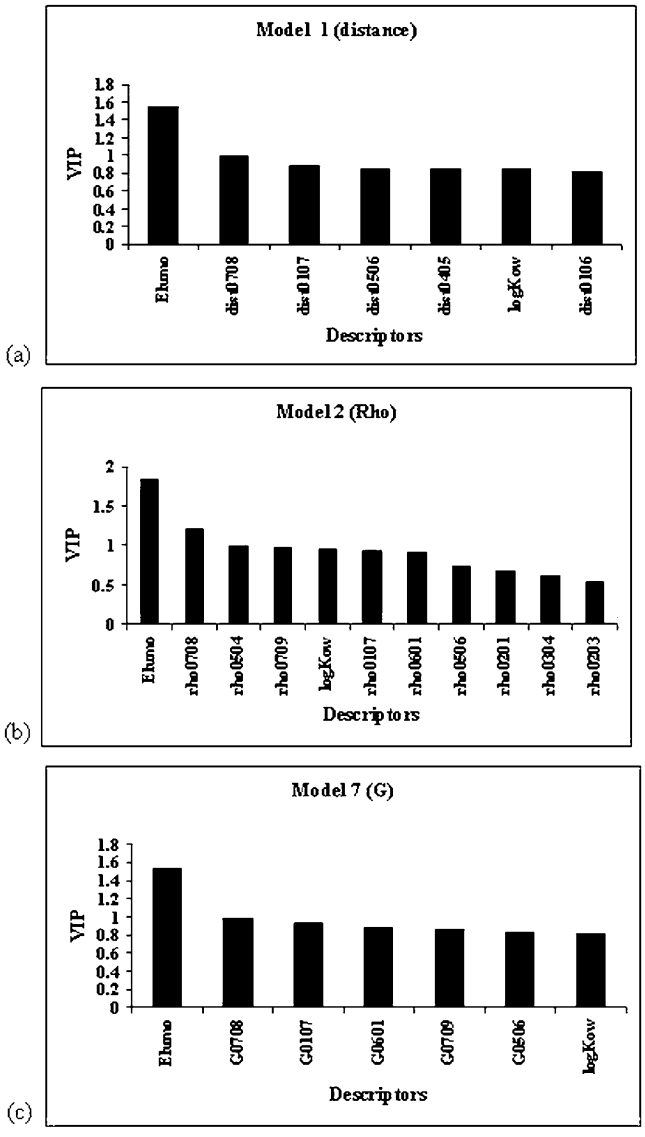

Figure 2. VIP plots for (a) model 1, (b) model 2 and (c) model 7.

Table 3. List of R^2_{int} and Q^2_{int} values from randomisation test of selected models for toxicity of nitroaromatics to *S. cerevisiae*.

Model No.	R^2_{int}	Q^2_{int}
1	0.28	–0.53
2	0.48	–0.46
7	0.23	–0.38
9	0.15	–0.51

To further check the predictability of the models, we have applied the criteria recommended by Golbraikh and Tropsha [41] and the results are shown in Table 4. It can be seen that the reference model does not fulfil one of the recommended criteria: both $(r^2 - r_0^2)/r^2$ and $(r^2 - r_0'^2)/r^2$ values are more than 0.1. However, the models developed from BCP descriptors along with E_{LUMO} and log K_{ow} fulfilled all the criteria. Models 1, 2 and 9 show either $(r^2 - r_0^2)/r^2$ or $(r^2 - r_0'^2)/r^2$ vanishing, which indicates good predictability. Thus, BCP descriptors increase external predictability of models for nitroaromatic toxicity to *S. cerevisiae* when used along with electrophilicity index and lipophilicity. Models developed excluding BCP descriptors (reference model) show lower external predictivity.

4 Conclusions

The results of the present study suggest that BCP descriptors can develop predictive QSAR models for nitroaromatic toxicity to *S. cerevisiae* when used along with electrophilicity index and partition coefficient. Furthermore, the importance of the nitro group for toxicity of nitroaro-

Table 4. External validation characteristics of different models for toxicity of nitroaromatics to *S. cerevisiae* according to Golbraikh and Tropsha [41].

Statistical parameters		Model number									
Sl. No.	Parameters	Reference	1	2	3	4	5	6	7	8	9
1	r^2	0.847	0.880	0.967	0.952	0.972	0.823	0.938	0.938	0.928	0.910
2	r_0^2	0.687	0.880	0.966	0.849	0.939	0.743	0.806	0.929	0.719	0.910
3	$r_0'^2$	−0.115	0.866	0.967	0.908	0.907	0.36	0.888	0.907	0.857	0.901
4	$(r^2 - r_0^2)/r^2$	0.189	0.000	0.001	0.108	0.034	0.097	0.141	0.010	0.225	0.000
5	$(r^2 - r_0'^2)/r^2$	1.136	0.016	0.000	0.046	0.067	0.563	0.053	0.033	0.077	0.010
6	Minimum of 4 and 5	**0.189**	**0.000**	**0.000**	0.046	0.034	0.097	0.053	0.010	0.077	**0.000**
7	k	1.021	0.999	0.996	0.92	0.972	1.011	0.931	0.976	0.917	1.026
8	k'	0.957	0.992	1.002	1.075	1.024	0.971	1.058	1.019	1.067	0.968

matics could also be identified from the QSAR models with QTMS descriptors.

Acknowledgement

We thank the Commonwealth Scholarship Commission in the UK for their financial support towards an Academic Staff Fellowship for one of us (K. R.).

References

[1] S. S. Young, S. J. Profeta, R. J. Unwalla, *Chemom. Intell. Lab. Syst.* **1997**, *37*, 115 – 124.
[2] E. Zvinavashe, H. van den Berg, A. E. Soffers, J. Vervoort, A. Freidig, A. J. Murk, I. M. Rietjens, *Chem. Res. Toxicol.* **2008**, *21*, 739 – 745.
[3] A. Bassan, A. P. Worth, *QSAR Comb. Sci.* **2008**, *27*, 6 – 20.
[4] J. M. Bermudez-Saldana, L. Escuder-Gilabert, M. J. Medina-Hernandez, R. M. Villanueva-Camanas, S. Sagrado, *J. Chromatogr. B* **2005**, *814*, 115 – 125.
[5] M. Davoren, A. M. Fogarty, *Ecotoxicol. Environ. Saf.* **2005**, *60*, 203 – 212.
[6] K. L. Kaiser, *SAR QSAR Environ. Res.* **2007**, *18*, 3 – 20.
[7] J. Tunkel, K. Mayo, C. Austin, A. Hickerson, P. Howard, *Environ. Sci. Technol.* **2005**, *39*, 2188 – 2199.
[8] E. Benfenati, *Chem. Cent. J.* **2007**, 1 – 32, Volume 1.
[9] P. C. von der Ohe, R. Kuhne, R. U. Ebert, R. Altenburger, M. Liess, G. Schuurmann, *Chem. Res. Toxicol.* **2005**, *18*, 536 – 555.
[10] http://europa.eu.int/comm/enterprise/reach/overview.htm.
[11] X.-F. Yan, H.-M. Xiao, X.-D. Gong, X.-H. Ju, *Chemosphere* **2005**, *59*, 467 – 471.
[12] X. Wang, C. Yin, L. Wang, *Chemosphere* **2002**, *46*, 1045 – 1051.
[13] G. Booth, Nitro Compounds, Aromatic, in: M. Pelc (Ed.), *Ullmann's Encyclopaedia of Industrial Chemistry*, VCH, Weinheim **1991**, 411 – 455
[14] J. Ahlers, M. Benzing, A. Gies, W. Pauli, E. Rösick, *Chemosphere* **1988**, *17*, 1603 – 1615.
[15] P. P. Roy, J. T. Leonard, K. Roy, *Chemom. Intell. Lab. Syst.* **2008**, *90*, 31 – 42.
[16] K. Roy, J. T. Leonard, *QSAR Comb. Sci.* **2006**, *25*, 235 – 251.
[17] P. L. A. Popelier, *J. Phys. Chem. A*, **1999**, *103*, 2883 – 2890.
[18] S. E. O'Brien, P. L. A. Popelier, *Can. J. Chem.* **1999**, *77*, 28 – 36.
[19] S. E. O'Brien, P. L. A. Popelier, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 764 – 775.
[20] P. L. A. Popelier, U. A. Chaudry, P. J. Smith, *J. Chem. Soc., Perkin Trans. 2*, **2002**, 1231 – 1237.
[21] S. E. O'Brien, P. L. A. Popelier, *J. Chem. Soc., Perkin Trans. 2*, **2002**, 478 – 483.
[22] U. A. Chaudry, P. L. A. Popelier, *J. Phys. Chem. A* **2003**, *107*, 4578 – 4582.
[23] P. J. Smith, P. L. A. Popelier, *J. Comput. Aided Mol. Des.* **2004**, *18*, 135 – 143.
[24] P. L. A. Popelier, P. J. Smith, U. A. Chaudry, *J. Comput. Aided Mol. Des.* **2004**, *18*, 709 – 718.
[25] P. J. Smith, P. L. A. Popelier, *Org. Biomol. Chem.* **2005**, *3*, 3399 – 3407.
[26] R. J. Loader, N. Singh, P. J. O'Malley, P. L. A. Popelier, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1249 – 1254.
[27] N. Singh, R. J. Loader, P. J. O'Malley, P. L. A. Popelier, *J. Phys. Chem. A* **2006**, *110*, 6498 – 6503.
[28] P. L. A. Popelier, P. J. Smith, *Eur. J. Med. Chem.* **2006**, *41*, 862 – 873.
[29] K. Roy, P. L. A. Popelier, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2604 – 2609.
[30] A. Mohajeri, B. Hemmateenejad, A. Mehdipour, R. Miri, *J. Mol. Graph. Mod.* **2008**, *26*, 1057 – 1065.
[31] A. Mohajeri, M. H. Dinpajooh, *J. Mol. Struct. (THEOCHEM)* **2008**, *855*, 1 – 5.
[32] S. T. Howard, O. Lamarche, *J. Phys. Org. Chem.* **2003**, *16*, 133 – 141.
[33] R. F. W. Bader, H. J. T. Preston, *Int. J. Quantum Chem.* **1969**, *3*, 327 – 347.
[34] GaussView3.0 Semichem Inc., Gaussian Inc., Pittsburgh, PA, USA **2003**.
[35] GAUSSIAN03, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. J. Montgomery, J. T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Ba-

boul, S. Clifford, J. Cioslowski, B. B.Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, In Gaussian, Inc., Pittsburgh PA **2003**.
[36] MORPHY98, a program written by P. L. A. Popelier with a contribution from R. G. A. Bone, UMIST, Manchester, England, EU **1998**.
[37] P. L. A. Popelier, *Chem. Phys. Lett.* **1994**, *228*, 160 – 164.
[38] UMETRICS *SIMCA-P 10.0*, info@umetrics.com: www.umetrics.com, Umeå, Sweden **2002**.
[39] S. Wold, M. Sjostrom, L. Eriksson, *Chemo. Intell. Lab. Syst.* **2001**, *58*, 109 – 130.
[40] S. Wold, in: H. van de Waterbeemd (Ed.), *Chemometric Methods in Molecular Design*, VCH, Weinheim **1995**, pp. 195 – 218.
[41] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Mod.* **2002**, *20*, 269 – 276.
[42] P. P. Roy, K. Roy, *QSAR Comb. Sci.* **2008**, *27*, 302 – 313.
[43] K. Roy, *Expert Opin. Drug Discov.* **2007**, *2*, 1567 – 1577.