InSilicoTox Project

Deliverable 3 (D3)

(Q)SARs for the Prediction of Reactive Toxicity

April 2010

European Union Sixth Framework Programme Marie Curie Actions
Project No.: **42328**
Project Acronym: **InSilicoTox**
Project Full Name: ***In Silico*** **Tools to Reduce the Use of Animals in Toxicity Testing for Bioreactive Chemicals**

Deliverable Prepared by

Johannes Schwöbel, Mark Cronin, Judith Madden
School of Pharmacy and Chemistry
Liverpool John Moores University
Byrom Street
Liverpool
L3 3AF
England

Important Information

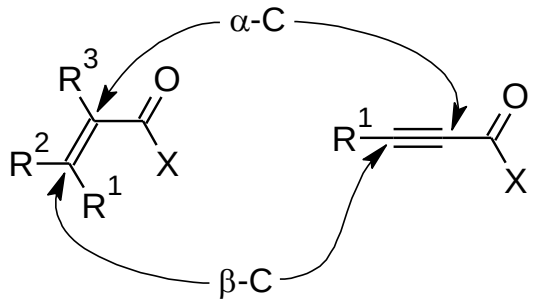
D3. (Q)SARs for the Prediction of Reactive Toxicology

1. Summary

Electrophilic reactivity is an important issue to assist in the understanding and description of interactions between xenobiotic toxicants and biochemical macromolecules.[1] An increasing part of developing alternatives for the toxicological assessment of chemicals, including *in chemico* (experimental measurement of reactivity) and *in silico* (computational) approaches, involves the understanding of reactive toxicity. It has long been understood that compounds capable of covalent interactions with biological macromolecules are liable to elicit toxic effects.[2]

α,β-Unsaturated compounds acting via Michael type reactions are considered to be particularly reactive.[3] They are capable of forming irreversible bonds with biological macromolecules, such as proteins or DNA molecules.[4] This covalent reactivity is known to result in a number of toxic effects, including evaluated acute aquatic toxicity above baseline narcosis. Compounds with an α,β-unsaturated carbonyl or carboxyl fragment are very important industrial substances, for example in the manufacture of polymers, textiles, or auxiliary materials in medicine.[5, 6]

Types of chemical fragments associated with this mechanism define the Michael acceptor domain, which is a very important chemical domain in reactive toxicity. Michael acceptors are characterised as having carbon-carbon double (C=C) or triple (C≡C) bonds with electron withdrawing substituents (see Figure 1).



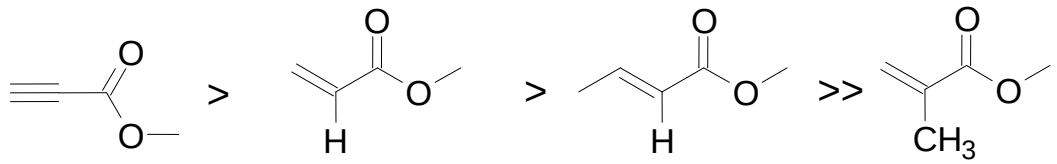
R^1, R^2 = aliphatic (R) or aromatic groups
R^3 = -H, -R
X = -H, -R, -O-R, -NH-R

Figure 1. *Michael addition domain: α,β-unsaturated carbonyl or carboxyl groups with double or triple bonds, including cyclic systems (in addition, electron-withdrawing groups for "Michael type" electrophiles might include pyridino heterocyclic groups and, not considered in this deliverable, –CN, –SO_2R, –NO_2)*[7].

The electron withdrawing group results in a polarisable electron density at the π-bond, where the β-carbon atom is positively polarised and becomes the preferred site of an attack with soft nucleophiles, for example glutathione.

Various studies have shown good quantitative relationships between reactivity of chemicals acting as Michael type acceptors and acute aquatic toxicity to *Tetrahymena pyriformis.*[8] For example, the differential toxicity of methacrylate, acrylate and crotonate can be explained by their varying reactivity, although the structures are very similar: methacrylate is poorly reactive, while acrylate is very reactive, and crotonate lies somewhere in between (structures shown in Figure 2).[9]

There is current interest to use *in silico* techniques to predict potential *in chemico* reactivity.[10] This is an attractive approach, especially where experimental measurements are not possible or restricted, for example because of low solubility or high volatility, if required for prioritisation of chemical inventories.



1 **2** **3** **4**

Figure 2. *Structures of methyl propiolate (**1**), methyl acrylate (**2**), methyl crotonate (**3**) and methyl methacrylate (**4**) in order of decreasing reactivity.*

A quantitative *in silico* model has been constructed as a measure for reactivity of Michael addition and potential toxicity. The model developed herein considers both electronic and steric effects, while the descriptors are based on easily computed ground state properties, performed by quantum chemical calculations.

2. Material and Methods

2.1 Data set

From the chemical reactivity database (as described in deliverable D5, referring to various literature sources), 50% inhibition growth concentration to *Tetrahymena pyriformis* data (IGC_{50}) have been collected for α,β-unsaturated compounds. 70 Michael type acceptor compounds are included in the dataset, which contains α,β-unsaturated compounds with a double bond or a triple bond, cyclic systems, and with or without substituents at the α-carbon. The compound classes included are aldehydes, ketones, amides, and esters.

2.2 Computational Details

The geometries of all molecules have been optimised at the density functional theory B3LYP/6-31G** level, using Gaussian 03.[11] For all compounds frequency analysis was carried out to confirm that correct ground-state geometries had been obtained. Natural Bond Orbital (NBO) analysis methods were applied by means of

the NBO 3.1 implementation in Gaussian 03.[12, 13] For the calculation of the local electrophilicity we used an in-house program (which will be implemented into the InSilicoTox Toolbox Software, as described in deliverable D5). Parameters for steric hindrance were calculated by the freely available MSMS program,[14] using published Connolly radii.[15] Multilinear regression analysis was performed using the Minitab 15.1 statistical software.

3. Model for the Reactive Toxicity of Compounds Acting by Michael Addition

The aim of the method described is to predict Michael addition reactivity for acute aquatic toxicity to *Tetrahymena pyriformis*, based on the ground state properties of a single molecule. To estimate reactive toxicity in a physically reliable way, one should take into account the following considerations:

In the chemical reaction between an α,β-unsaturated electrophilic compound and soft cellular amino acid sites, it is probable that multiple steps are involved, all of which have an influence on the toxicity initiated by Michael addition reactions.

One important step in the reaction is the formation of a peptide or protein adduct complex. This peptide binds to the β-carbon atom of an α,β-unsaturated compound. Therefore, a sound reactivity descriptor is needed which models the capability to accept electrons at the β-carbon atom reliably. As a site specific reactivity descriptor is desired here for the prediction of reactivity, a local counterpart of the electrophilicity index is applied, which has been described in Deliverable D2 already. We have defined the local electrophilicity index $\omega_{EQ}(r)$ at an atomic reaction site *r* in the following way:

$$\omega_{EQ}(r) = \frac{EQ_{occ}(r)^2 + 2EQ_{occ}(r)EQ_{vac}(r) + EQ_{vac}(r)^2}{4\left(EQ_{vac}(r) - EQ_{occ}(r)\right)} \qquad (1)$$

In this equation, $EQ_{occ}(r)$ is the effective (charge limited) electron donor energy, and $EQ_{vac}(r)$ the analogous electron acceptor energy. Both functions are site specific, located at the β-carbon. These, and related, descriptors were originally invented by

Klamt (1993).[16] They were recently reintroduced by Schüürmann and co-workers, in order to predict the rate constants of indirect photolysis by •OH radicals,[17, 18] and the affinity to form hydrogen bonds.[19-21] Furthermore, we define a maximum electrophilicity index ω_{max} which accounts for the observation that after a certain point higher (calculated) electrophilicity indices do not reflect higher (experimental) rate constants.

Next, we have assumed that resonance stabilisation of the (nominally) carbanionic intermediate plays an important role in the reaction pathway, in the case of α,β-unsaturated aldehydes and ketones accompanied by a keto-enol-tautomerism. In all likelihood, the carbanionic intermediate or the enolic form are reactive intermediates while forming the peptide-electrophile-complex. Moreover, this reaction should involve the rate limiting step. Therefore, a second descriptor is needed to describe the energetic of this intermediate.

The resonance stabilisation of the carbanionic adduct intermediate is difficult to describe from the ground state properties alone. For that reason, we have made the following assumption: The easier it is to switch from the notation on the left (3a) of the transition state to that on the right (3b) in Figure 3, the better should be the resonance stabilisation and the higher the reaction rate.

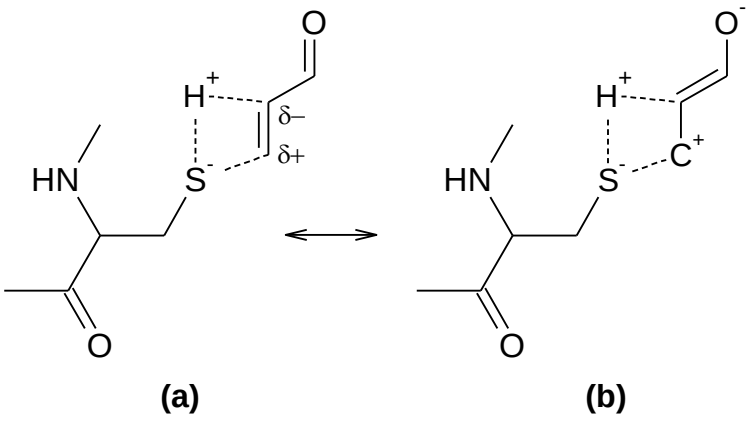


Figure 3. *Resonance stabilisation of the Michael type addition transition state and participation of the α-carbon atom in the transition state structure (solvent mediated).*

In the notation on the left (3a) there is, formally, a single bond between the ipso carbon atom (the carbonyl or carboxyl C atom) and the α-carbon atom. The energy level E_σ(C-C) of this σ-bond can be calculated by Weinhold's natural bond orbital (NBO) analysis.[12, 13] In the notation on the right (3b), this bond turns into a double bond. As is known, π-bonds always have a higher energy level than σ-bonds. Now the approximation is that the closer the original σ-bond comes to the π-bond character, in the calculations assumed by a higher σ-bond NBO-energy level, the easier it should be to switch from the notation on the left (3a) into the resonance stabilising notation on the right (3b). The last step, the reaction from any enolic intermediate to the final keto product, should be very fast and not rate limiting, so this step can be neglected in the modelling.

Electronic effects will have, almost certainly, the main influence on the reaction rates. Nevertheless, steric hindrance could slow down the reaction significantly, so we have considered steric hindrance as an additional term in modelling rate constants. This could be another reason why molecules with a highly accessible triple bond react significantly faster with glutathione (GSH), but molecules with an α-substituent remarkably slower compared to analogous molecules[8] (e.g. methyl methacrylate: rate constant log k_{GSH} = –1.14 [M^{-1} min^{-1}] < methyl acrylate: log k_{GSH} = 1.06 < methyl propiolate: log k_{GSH} = 2.07; the structures are shown in Figure 1). The main reason remains the higher (or lower) polarised – and thus higher (or less, respectively) reactive – β-carbon atom, but to this end we cannot explain this phenomenon by electronic effects alone.

If the steric effect of the α-substituent plays a major role, it is involved in the transition state of the rate determining step, e.g. by solvent mediated stabilising of the transition state. Indeed, further calculations confirm the involvement of the α-carbon atom in the transition state (currently prepared as a publication), shown in Figure 3.

One way to quantify steric hindrance is to calculate site specific solvent accessible surface areas σ_{SAS} (SAS, in percent accessibility of a particular atom). The higher the accessible surface area, the lower is the steric hindrance and, therefore, the

higher is the reaction rate. We take both the local accessible surface area at the α-carbon as well as at the β-carbon atom into account.

The combined reactive toxicity (*RT*) model is shown in the following equation:

$$\log (IGC_{50,RT}^{-1}) = RT = a\ \omega_{EQ}(C_\beta) + b\ E_\sigma + c \log \sigma_{SAS}(C_\beta) + d \log \sigma_{SAS}(C_\alpha) + e \qquad (2)$$

The equation estimates 50% inhibition growth concentration (IGC_{50}) to *Tetrahymena pyriformis* after 40h by the following descriptors: the local electrophilicity index ω_{EQ} at the β-carbon as a reactivity descriptor, the energy level E_σ of the σ-bond between the α-carbon and the ipso-carbon as an indication of resonance stabilisation of the carbanionic intermediate, and the steric accessibility σ_{SAS} at both the α-carbon and the β-carbon atoms. *a*, *b*, *c*, *d* and *e* are regression coefficients.

Hydrophobicity (logarithm of the octanol-water partition coefficient, log *P*) is related to baseline toxicity, often connected with a narcotic mode of action. Ellison *et al* developed the following equation for baseline toxicity (*BT*) for *Tetrahymena pyriformis*:[23]

$$\log (IGC_{50,BT}^{-1}) = BT = 0.78 \log P - 2.01 \qquad (3)$$

The overall toxicity has been defined as the maximum value of equations (2) and (3):

$$\log (IGC_{50}^{-1}) = \text{maximum}(RT; BT) \qquad (4)$$

If a particular compound has no significant reactive toxicity, then it will produce a minimum of baseline toxicity for acute aquatic toxicity.

4. Results and Discussion

4.1 Model Calibration

The Michael addition reactive toxicity model has been applied and tested for all 70 potential Michael acceptor compounds. Because the descriptor ω_{EQ} contains two internal non-linear parameters (namely the limiting charge amount Q_{occ} of the descriptor EQ_{occ} for occupied MOs, and the equivalent limiting charge amount Q_{vac} of the descriptor EQ_{vac} for virtual MOs), they were obtained by an in-house iterative root-mean-square error (*rms*) minimisation algorithm, comparing experimental and model values in each iteration step. Based on those optimised parameters, a multilinear calibration was performed using the Minitab statistical software.

At first, equation (2) was calibrated for 44 compounds, where excess toxicity (= log $IGC_{50,exp}^{-1}$ – log $IGC_{50,BT}^{-1}$) is greater than 0.7. In principle, Michael acceptor compounds are considered to be highly reactive. However, at least 37% of the toxicity values in our dataset can be explained by baseline toxicity: equation (3) applies to 26 compounds with excess toxicity values lower than 0.7.

The overall prediction quality is good. The correlation between predicted and experimental values is shown in Figure 4. The statistical parameters of the prediction quality are: number of compounds n = 70; squared correlation coefficient r^2 = 0.87; predictive squared correlation coefficient q^2 = 0.86; root mean square error *rms* = 0.34; bias = –0.04.

It should be noted that calculation of excess toxicity is based on experimental toxicity values, which are needed to distinguish between reactive and narcotic level compounds. A way to overcome this problem is the use of a minimum toxicity, defined by equation (4). At first, the toxicity of a compound is defined by its baseline toxicity; if the reactive toxicity value is higher than its baseline toxicity, then the reactive toxicity value is taken, instead. This approach is very straightforward. Moreover, the statistical parameters using equation (4) are very similar to the combination of equations (2) and (3): n = 70; r^2 = 0.86; q^2 = 0.85; *rms* = 0.35; bias =

+0.06. Therefore, equation (4) is appropriate to predict toxicity to *Tetrahymena pyriformis*.

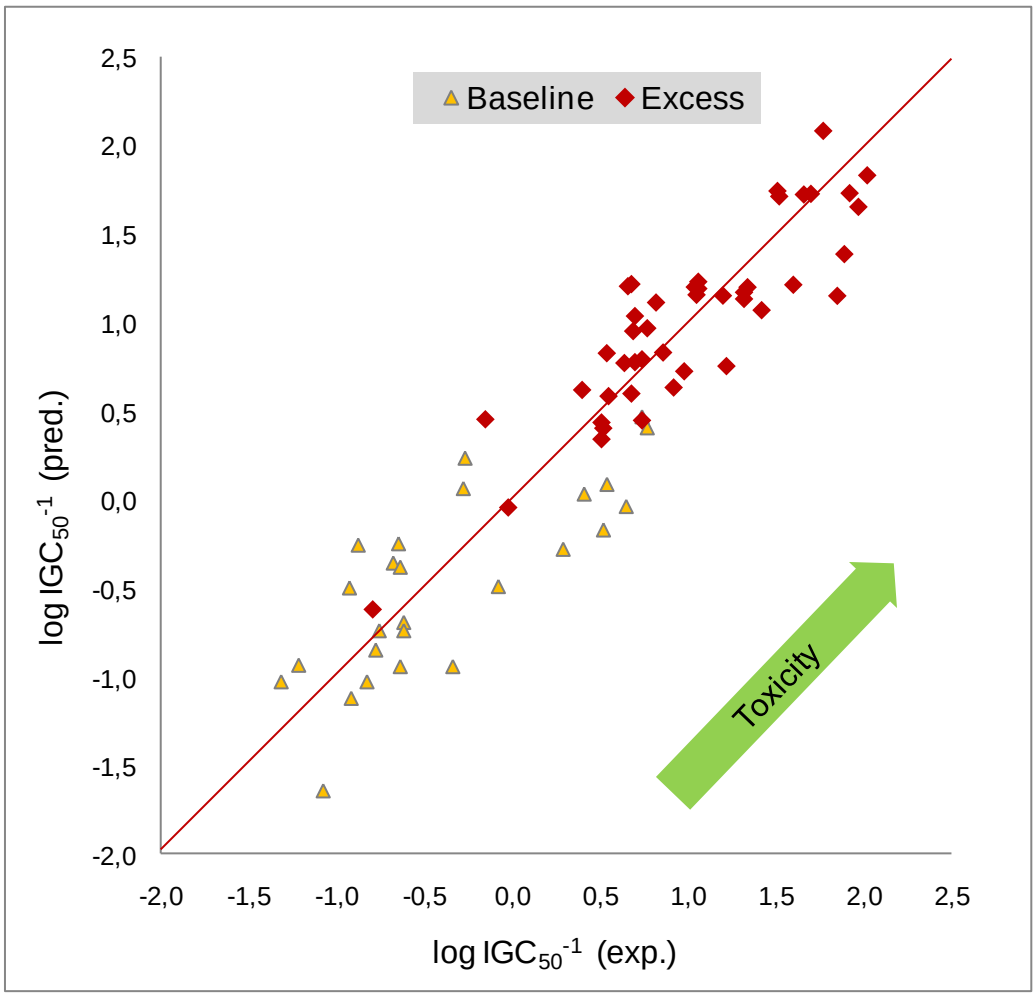


Figure 4. *Predicted and experimental IGC_{50} values for acute aquatic toxicity to Tetrahymena pyriformis: Eq. (2) was used for reactive compounds (excess toxicity ≥ 0.7), and Eq. (3) for narcotic level compounds (excess toxicity < 0.7, defined as the difference between experimental [log $IGC_{50,exp}^{-1}$] and baseline toxicity values [log $IGC_{50,BT}^{-1}$]).*

4.2 Non-reactive and reactive compounds

It is an important issue in toxicology to distinguish qualitatively between reactive and non-reactive compounds. For this reason, narcotic compounds are defined by

predicted toxicity values (equation 4), which are lower than 0.7 logarithmic units above baseline toxicity, (equation 3). All remaining compounds are defined as having excess toxicity.

Applying this definition, both sensitivity (the ratio between the number of correctly predicted compounds and the actual number of experimental compounds in a category) and predictivity (the ratio between the number of correctly predicted compounds and the number of all predicted values in a category) are good. For baseline toxicity, sensitivity is 85% and predictivity 81%. For excess toxicity, the values are even better, with a sensitivity of 89% and a predictivity of 91%.

Therefore, the toxicity model of potential Michael acceptor compounds, is both robust and valid.

5. Conclusions

We have introduced a model to predict reactive toxicity of α,β-unsaturated compounds, in terms of acute aquatic toxicity to *Tetrahymena pyriformis*. For the first time a quantum chemical overall model for various compound classes involved in Michael addition has been shown to work in a quantitative manner. The model has good statistical performance and is mechanistically interpretable. Even more, the model is based on easily computed ground state properties, which are less computational demanding than, for example, quantum chemical transition state calculations.

This knowledge might help in the reactivity or toxicity profiling of electrophilic compounds. It can support mechanistic grouping to order compounds according to reactivity and also confirm, that a compound is reactive. It should be noted in this context, that whilst there is no global model likely for reactivity, the use of local models will deepen confidence into the prediction of reactive peptide-binding or DNA-binding for mechanistically developed groupings and categories.

6. References

1. Aptula, A. O.; Roberts, D. W., Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: General principles and application to reactive toxicity. *Chem. Res. Toxicol.* **2006,** *19*, 1097-1105.
2. Schultz, T. W.; Carlson, R. E.; Cronin, M. T. D.; Hermens, J. L. M.; Johnson, R.; O'Brien, P. J.; Roberts, D. W.; Siraki, A.; Wallace, K. D.; Veith, G. D., A conceptual framework for predicting toxicity of reactive chemicals: Models for soft electrophilicity. *SAR QSAR Environ. Res.* **2006,** *17*, 413-428.
3. Schultz, T. W.; Yarbrough, J. W.; Hunter, R. S.; Aptula, A. O., Verification of the structural alerts for Michael acceptors. *Chem. Res. Toxicol.* **2007,** *20*, 1359-1363.
4. Koleva, Y. K.; Madden, J. C.; Cronin, M. T. D., Formation of categories from structure-activity relationships to allow read-across for risk assessment: Toxicity of α,β-unsaturated carbonyl compounds. *Chem. Res. Toxicol.* **2008,** *21*, 2300-2313.
5. Tanii, H.; Hashimoto, K., Structure-toxicity relationship of acrylates and methacrylates. *Toxicol. Lett.* **1982,** *11*, 125-129.
6. Potter, D. W.; Tran, T.-B., Rates of ethyl acrylate binding to glutathione and protein *Toxicol. Lett.* **1992,** *62*, 275-285.
7. Roberts, D. W.; Patlewicz, G.; Kern, P. S.; Gerberick, F.; Kimber, I.; Dearman, R. J.; Ryan, C. A.; Basketter, D. A.; Aptula, A. O., Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization. *Chem. Res. Toxicol.* **2007,** *20*, 1019-1030.
8. Böhme, A.; Thaens, D.; Paschke, A.; Schüürmann, G., Kinetic glutathione chemoassay to quantify thiol reactivity of organic electrophiles - Application to α,β-unsaturated ketones, acrylates, and propiolates. *Chem. Res. Toxicol.* **2009,** *22*, 742-750.
9. Yarbrough, J. W.; Schultz, T. W., Abiotic sulfhydryl reactivity: a predictor of aquatic toxicity for carbonyl-containing α,β-unsaturated compounds. *Chem. Res. Toxicol.* **2007,** *20*, 558-562.
10. Cronin, M. T. D.; Bajot, F.; Enoch, S. J.; Madden, J. C.; Roberts, D. W.; Schwöbel, J., The in chemico–in silico interface: Challenges for integrating experimental and computational chemistry to identify toxicity. *ATLA* **2009,** *37*, 513-521.
11. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pompelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*, Gaussian, Inc.: Pittsburgh, PA, USA, 2003.
12. Reed, A. E.; Weinstock, R. B.; Weinhold, F., Natural population analysis. *J. Chem. Phys.* **1985,** *83*, 735-746.
13. Foster, J. P.; Weinhold, F., Natural hybrid orbitals. *J. Am. Chem. Soc.* **1980,** *102*, 7211-7218.
14. Sanner, M. *MSMS*, Molecular Graphics Laboratory: La Jolla, CA, USA, 1996.

15. Connolly, M. L., The molecular surface package. *J. Mol. Graphics* **1992,** *11*, 139-141.
16. Klamt, A., Estimation of gas-phase hydroxyl radical rate constants of organic compounds from molecular orbital calculations. *Chemosphere* **1993,** *26*, 1273-1289.
17. Böhnhardt, A.; Kühne, R.; Ebert, R.-U.; Schüürmann, G., Indirect photolysis of organic compounds: prediction of OH reaction rate constants through molecular orbital calculations. *J. Phys. Chem. A* **2008,** *112*, 11391-11399.
18. Böhnhardt, A.; Kühne, R.; Ebert, R.-U.; Schüürmann, G., Predicting rate constants of OH radical reactions with organic substances: advances for oxygenated organics through a molecular orbital HF/6-31G** approach *Theor. Chem. Acc.* **2010,** *in press*.
19. Schwöbel, J.; Ebert, R.-U.; Kühne, R.; Schüürmann, G., Modeling the H bond donor strength of -OH, -NH, and -CH sites by local molecular parameters. *J. Comput. Chem.* **2009,** *30*, 1454-1464.
20. Schwöbel, J.; Ebert, R.-U.; Kühne, R.; Schüürmann, G., Prediction of the intrinsic hydrogen bond acceptor strength of chemical substances from molecular structure. *J. Phys. Chem. A* **2009,** *113*, 10104-10112.
21. Schwöbel, J.; Ebert, R.-U.; Kühne, R.; Schüürmann, G., Prediction of the Intrinsic Hydrogen Bond Acceptor Strength of Organic Compounds by Local Molecular Parameters. *J. Chem. Inf. Model.* **2009,** *49*, 956-962.
22. Aizman, A.; Contreras, R.; Pérez, P., Relationship between local electrophilicity and rate coefficients for the hydrolysis of carbenium ions. *Tetrahedron* **2005,** *61*, 889-895.
23. Ellison, C. M.; Cronin, M. T. D.; Madden, J. C.; Schultz, T. W., Definition of the structural domain of the baseline non-polar narcosis model for *Tetrahymena pyriformis*. *SAR QSAR Environ. Res.* **2008,** *19*, 751-783.