Prediction of Reaction Rates of VOCs with NO₃ Radicals

Michael Schindler, Sebastian Schmidt
Bayer AG Crop Science Division, Environmental Safety, Alfred-Nobel-Str. 50, D-40789 Monheim, Germany
E-mail contact: sebastian.schmidt1@bayer.com

Abstract

A QSAR for the prediction of rate constants for the degradation of volatile organic compounds by nitrate radicals is developed using the Partial Least Squares technique. The QSAR is based on experimental data published in the literature for 260 compounds. They are modeled by a set of calculated descriptors from standard descriptor generation tools and from quantum chemistry.

A diversity-based data set of 99 compounds turned out to be the optimum choice with regard to simplicity and performance. The final QSAR model is characterized by $r^2=0.831$ (fit) and $q^2=0.823$ (prediction), and by an $r_{pred}=0.862$ for the n=155 external validation set. The QSAR needs 3 latent variables. The most important descriptors for the QSAR are the ionization potential, obtained from density functional theory, and the energy of the highest occupied molecular orbital, which are modulated by fingerprints indicating the presence of specific molecular fragments like functional groups or ring systems.

Results

QSAR models for the prediction of rate constants ($k_{NO_3}$) for the reaction of nitrate radicals with VOCs have been developed [1] by using the PLS approach [2]. They are based on compound selections guided by diversity considerations. The optimum QSAR uses 99 out of 260 compounds as the work set, needs three latent variables (LVs), is characterized by $r^2=0.831$ (fit), $q^2=0.823$ (prediction), and by an $r_{pred}=0.862$ for the validation set of n=155 compounds and has a larger applicability domain (AD) than previous models [3].

Graph 1. Experimental versus predicted rate constants for the full data set from the final 3-LV-PLS model

The most important descriptors for the QSAR are the ionization potential (IP) and the energy of the highest occupied molecular orbital (HOMO) which are modulated by fingerprints indicating the presence of specific molecular fragments. The effect of insufficiently differentiating between phenolic- and non-phenolic aromatic compounds is shown in Graph 2.

Graph 2: Comparison of predicted rate constants by two QSAR models with (3 LV) and without (2 LV) phenolic fingerprints

The validity of a QSAR is affected by high-leverage data points. Care has to be taken to avoid so-called ‘bad high-leverage points’ (Graph 3). However, ‘good high-leverage points’ can stabilize a model.

Graph 3: Williams plot of the 2-LV- and the final 3-LV-QSAR model

Dotted and dashed lines denote ±3σ and ±2.5σ. Aromatic and phenolic compound clusters are designated by dot-dashed and dashed ellipses; work- and test-set compounds are shown in blue and red

Conclusions

A robust QSAR model for the prediction of reaction rates of VOCs with nitrate radicals has been developed, extending the range of applicability to hydroxybenzenes. Before using the QSAR for predictions, it is important to check whether the new compounds belong to the AD of the QSAR by using the statistical membership probability criteria and the descriptor value ranges. The best validation of a QSAR is the comparison of predictions with new experimental data. This is shown below for methoxophenols [4], demonstrating the reliability of the present QSAR.

Table 1. External validation by comparison with new experiments

<table>
<thead>
<tr>
<th>Methoxophenol</th>
<th>$k_{NO_3}$ (mp)</th>
<th>T320</th>
<th>HOMO [eV]</th>
<th>IP [eV]</th>
<th>$-\log(k_{NO_3})$</th>
<th>experiment</th>
<th>prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-mp</td>
<td>1.375E-10</td>
<td>1</td>
<td>-8.64</td>
<td>5.81</td>
<td>9.06</td>
<td>9.52</td>
<td></td>
</tr>
<tr>
<td>3-mp</td>
<td>1.15E-11</td>
<td>1</td>
<td>-8.94</td>
<td>6.08</td>
<td>10.94</td>
<td>10.23</td>
<td></td>
</tr>
<tr>
<td>2,6-dmp</td>
<td>1.584E-10</td>
<td>1</td>
<td>-8.61</td>
<td>5.87</td>
<td>9.80</td>
<td>9.59</td>
<td></td>
</tr>
<tr>
<td>2-mp</td>
<td>2.69E-11</td>
<td>1</td>
<td>-8.69</td>
<td>5.97</td>
<td>10.57</td>
<td>9.82</td>
<td></td>
</tr>
<tr>
<td>2-n-4-mp</td>
<td>8.41E-11</td>
<td>1</td>
<td>-8.61</td>
<td>5.84</td>
<td>10.08</td>
<td>9.55</td>
<td></td>
</tr>
</tbody>
</table>

References

[1] M.Schindler, A QSAR for the prediction of rate constants for the reaction of VOCs with nitrate radicals, Chemosphere 154 (2016) 23-33