QSPR prediction of physico-chemical properties and degradation of PBDEs

Ester Papa, Simona Kovarich, and Paola Gramatica
QSAR Research Unit in environmental chemistry and ecotoxicology
DBSF - University of Insubria, Varese - Italy

e-mail: paola.gramatica@uninsubria.it
ester.papa@uninsubria.it
INTRODUCTION – why PBDEs?

Polybrominated Diphenyl Ethers
209 congeners

- Class of brominated flame retardants added to plastics, polyurethane foams, textiles, electronic equipment.
- Have been marketed in 3 formulations:
  - “penta”, “octa”, “deca” – different % of congeners
- Levels in the environment and humans increased since they came into use
- Ban of “penta” and “octa” formulations since 2004
INTRODUCTION – why PBDEs?

Background knowledge about PBDEs:

- Low water solubility < 1 µg/kg
- High LogKow > 5
- Instability of the highly brominated congeners
- Persistence of medium brominated congeners
- Liver toxicity, thyroid toxicity, developmental toxicity
- Most of the data are available for a few congeners only (i.e. 28, 47, 99, 100, 138, 153, 154, 183, 209); insufficient data is available for the other PBDEs

There is the need to extend knowledge about properties and ecotoxicological data for a better understanding of PBDEs behaviour and related risks

This topic is included in the FP7- EU Project CADASTER under negotiation.
AIMS of the STUDY

- Overview of the data for physico-chemical properties of PBDEs available in literature, and compilation of the experimental data set.

- Development of QSPR models (linear regression) for all available endpoints with strong external validation.

- To compare our results with other existing QSPR models.
## Materials and Methods

### 1 - Dataset

<table>
<thead>
<tr>
<th>Property</th>
<th>n° of available exp.data (→modelled)</th>
<th>Bibliography</th>
<th>Comparison with other <em>ad hoc</em> models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half life&lt;sub&gt;photol.&lt;/sub&gt;</td>
<td>15</td>
<td>Eriksson et al. (2004)</td>
<td>not available</td>
</tr>
<tr>
<td>Logk&lt;sub&gt;hydrol.&lt;/sub&gt;</td>
<td>7</td>
<td>Rahm et al. (2005)</td>
<td>not available</td>
</tr>
<tr>
<td>Half life&lt;sub&gt;hydrol.&lt;/sub&gt;</td>
<td>7</td>
<td>Rahm et al. (2005)</td>
<td>not available</td>
</tr>
</tbody>
</table>
Materials and Methods

2 – Theoretical Molecular Descriptors

• Over 1600 mono-, bi- and tri-dimensional theoretical molecular descriptors were calculated for the studied compounds using DRAGON
• Constant and near to constant variables were eliminated
• To check for and eliminate pair correlations
• Four quantum-chemical descriptors were calculated using Hyperchem-MOPAC.
• The final number of variables used to build the QSPR models was 620
Materials and Methods

3 – QSPR – Modelling and validation

- Linear Regression QSPRs developed for all the endpoints (models including 1 or 2 variables – All Subset variable selection method).
- Internal validation: to verify the robustness of the model based on the training set only.
- External validation: to verify the real predictivity of the models using chemicals that are not included for the development of the model (training set).

External validation was performed by splitting the data into training and prediction sets using the Random by Response method (50% as prediction set).
Materials and Methods

4 – QSPR – Applicability Domain (verified for 209 PBDEs)

• The approach based on the leverage distance was applied for the analysis of the structural domain of the models (check for unreliable predictions).
• Standardised residuals were calculated to identify outliers for the response (residuals > 2.5σ).
• Williams Plot

To check the Applicability Domain is particularly important since these models are based on a small amount of experimental data:

RISK OF EXTRAPOLATION TO UNRELIABLE PREDICTIONS!
## RESULTS

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Obj. training</th>
<th>Descriptors</th>
<th>$R^2%$</th>
<th>$Q^2%$</th>
<th>$Q^2_{\text{EXT}}%$ (rand50%-30%*)</th>
<th>AD% (209 PBDEs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>logH</td>
<td>7</td>
<td>BEHe7</td>
<td>96.87</td>
<td>93.34</td>
<td>90.84*</td>
<td>64.7</td>
</tr>
<tr>
<td>MP</td>
<td>26</td>
<td>X2A</td>
<td>84.56</td>
<td>82.24</td>
<td>88.55</td>
<td>97.61</td>
</tr>
<tr>
<td>logPL</td>
<td>34</td>
<td>T(O…Br)</td>
<td>98.63</td>
<td>98.45</td>
<td>98.62</td>
<td>91.38</td>
</tr>
<tr>
<td>logS</td>
<td>12</td>
<td>Mor23m</td>
<td>91.8</td>
<td>88.55</td>
<td>85.04</td>
<td>95.69</td>
</tr>
<tr>
<td>LogKoa</td>
<td>30</td>
<td>T(O…Br)</td>
<td>97.37</td>
<td>96.78</td>
<td>95.17</td>
<td>92.34</td>
</tr>
<tr>
<td>LogKow</td>
<td>20</td>
<td>T(O…Br)</td>
<td>96.44</td>
<td>95.63</td>
<td>91.6</td>
<td>96.65</td>
</tr>
<tr>
<td>Logk\text{}photol.</td>
<td>15</td>
<td>MW</td>
<td>94.91</td>
<td>93.83</td>
<td>92.93</td>
<td>92.82</td>
</tr>
<tr>
<td>Logk\text{}hydrol.</td>
<td>7</td>
<td>HATS2p</td>
<td>91.19</td>
<td>85.05</td>
<td>98.8*</td>
<td>73.68</td>
</tr>
<tr>
<td>Half-Life\text{}photol.</td>
<td>15</td>
<td>T(O…Br)</td>
<td>94.39</td>
<td>92.66</td>
<td>90.97</td>
<td>86.6</td>
</tr>
<tr>
<td>Half-Life\text{}hydrol.</td>
<td>7</td>
<td>PW3</td>
<td>96.22</td>
<td>92.07</td>
<td>97.15*</td>
<td>88.99</td>
</tr>
</tbody>
</table>

Focus on the following aspects of interest:

VALIDATION

DOMAIN

COMPARISON
Model for Log Koa

LogKoa = 6.658 + 0.222 T(O...Br)

Are the predictions in the structural domain?

Experimental range of LogKoa: 7.34 (mono-BDE) – 11.96 (hepta-BDE)
Comparison with existing models

Predicted and Experimental data for 30 PBDEs

<table>
<thead>
<tr>
<th>Author</th>
<th>Method</th>
<th>n° obj.</th>
<th>n° vars</th>
<th>R²%</th>
<th>Q²%</th>
<th>Q² EXT % (rand50%)</th>
<th>RMSE (30 obj.)</th>
<th>RMSE 25obj</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papa et al. (2008)</td>
<td>MLR</td>
<td>30</td>
<td>1</td>
<td>97.37</td>
<td>96.68</td>
<td>95.17</td>
<td>0.25</td>
<td>0.47</td>
</tr>
<tr>
<td>Xu et al. (2007)</td>
<td>MLR</td>
<td>22</td>
<td>2</td>
<td>97.61</td>
<td>97.25</td>
<td>-</td>
<td>0.31</td>
<td>0.23</td>
</tr>
<tr>
<td>Chen et al. (2003)</td>
<td>PLS</td>
<td>13</td>
<td>10</td>
<td>98.13</td>
<td>97.59</td>
<td>-</td>
<td>-</td>
<td>0.21</td>
</tr>
</tbody>
</table>
Comparison with existing models

Predictions for 209 PBDEs

\[ \Delta \text{log units} \]

\[ \text{n° bromine increase} = \Delta \text{increase} \]

\[ \Delta \text{log} \]

\[ \text{average } \Delta \text{ (|YPapa- YKoaWIN|)} \]

\[ \text{average } \Delta \text{ (|YPapa-YXu|)} \]

\[ \text{Y}_{\text{Papa}} = \text{Predictions by our model (range Log Koa: 7.32 – 15.09)} \]

\[ \text{Y}_{\text{Episuite}} = \text{Predictions by KoaWIN (\Delta \text{max } = 3.33 \text{ log units; range Log Koa: 6.81-18.23)} \]

\[ \text{Y}_{\text{Xu}} = \text{Predictions by Xu et al. (2007) (\Delta \text{max } = 1.06 \text{ log units; range Log Koa: 7.4-15.73)} \]
Model for Log Kow

\[ \text{LogKow} = 3.675 + 0.162 \text{T(O...Br)} \]

Experimental range of LogKow: 5.03 (di-BDE) – 8.62 (octa-BDE)

Predictions for nona- and decaPBDE fall outside the AD
Comparison with existing models

Predicted and Experimental data for 20 PBDEs

Experimental range of LogKow: 5.03 (di-BDE) – 8.62 (octa-BDE)
Comparison with existing models

Predictions for 209 PBDEs

\[ \Delta \] log units

\[ n^\circ \text{ bromine increase} = \Delta \text{ increase} \]

\[ Y_{\text{Papa}} = \text{Pred. by our model} \text{ (range of LogKow: 4.2 – 9.8)} \]

\[ Y_{\text{KowWIN}} = \text{Pred. by KowWIN} \text{ (} \Delta_{\text{max}} = 2.27 \text{ log units; range of LogKow: 4.1 – 12.1)} \]

\[ Y_{\text{MlogP}} = \text{Pred. by MLogP} \text{ (} \Delta_{\text{max}} = 2.45 \text{ log units; range of LogKow: 4.1–7.4)} \]

\[ Y_{\text{AlogP}} = \text{Pred. by ALogP} \text{ (} \Delta_{\text{max}} = 1.15 \text{ log units; range of LogKow: 4.1 –10.9)} \]
Conclusions (1)

• New predictive models were developed *ad-hoc* for several physico-chemical properties and degradation parameters of the 209 PBDE congeners, according to the OECD principles for QSAR.
  \[ \text{LogKow, MP, Water Solubility, Half-life}_{\text{photo}}, \text{Half-life}_{\text{hydro}} : \text{modeled for the first time for PBDEs (no comparison in literature).} \]
  \[ \text{LogKoa, Henry Law constant, P}_L, k_{\text{photo}} : \text{our statistical results are equal to, or better than, those of the existing ad hoc models}. \]

• Advantages of our models: simplicity (1 descriptor only), strong external validation provided, AD is determined for all the 209 PBDE congeners.

• The accuracy of *ad hoc* models is higher than “non specific” models (i.e. KoaWIN, KowWIN, MlogP, AlogP).
Conclusions (2)

- Check AD! ~ 90% of the predictions from our models fell into their AD (reliable predictions).

  Exceptions are: models developed for LogH and for Logk\textsubscript{hydrol} (covered domain respectively: 65% and 73%; mono, di and tri PBDEs outside the AD - unreliable predictions)

- There is a need for more experimental data:
  1) to increase the AD for our models.
  2) to define a clearer mechanistic interpretation of these QSPRs, by clarifying the relationships among the Br position and the properties/reactivities. (See also POSTER POP24175E).
Our QSPRs could be used to fill data gaps according to the new REACH regulation, they can facilitate the screening and prioritisation of chemicals as well as help the search for alternative/safer chemicals (coming soon CADASTER EU Project - FP7)
THANK YOU!