### List of Tables

<table>
<thead>
<tr>
<th>Table Sr.No.</th>
<th>Description</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>List of various quantum-chemical descriptors available in the literature and proposed in the present work.</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Internal and external statistical validation parameters employed for assessing the robustness and external predictivity of the QSAR models</td>
<td>11</td>
</tr>
<tr>
<td>2.1a</td>
<td>List of nitrated-PAHs with their mutagenicity expressed as Log TA98 (in log revertants/nmol).</td>
<td>30</td>
</tr>
<tr>
<td>2.1b</td>
<td>Quantum-mechanical descriptor (all in a.u.) calculated with PM6, RM1, HF/6-311G(d,p) and DFT/B3LYP/6-311G(d,p) level of the theory for different nitrated PAHs.</td>
<td>31</td>
</tr>
<tr>
<td>2.2</td>
<td>Optimum value of internal and external validation parameters of QSAR models developed using PM6, RM1, HF, DFT and electron-correlation (CORR) based descriptors</td>
<td>35</td>
</tr>
<tr>
<td>2.3</td>
<td>Range of various internal validation parameters of QSAR models obtained using different types of splitting</td>
<td>37</td>
</tr>
<tr>
<td>2.4</td>
<td>Range of predicted mutagenic activity, log TA98 (using model VII in different splitting methods), of inactive compounds, and those with unusual mutagenicity</td>
<td>44</td>
</tr>
<tr>
<td>3.1</td>
<td>Optimum values of the internal and external validation parameters of QSAR models, for the TA100 mutagenicity of nitrated-AHs and PAHs, developed with descriptors computed using PM6, PM7, RM1 methods, and with electron-correlation contribution (CORR) of the descriptor (also see the supporting information Tables S3.1-S3.14 of Annexure II)</td>
<td>57</td>
</tr>
<tr>
<td>3.2</td>
<td>Comparison of internal and external validation parameters of QSAR models based on electron-correlation energy ((E_{\text{CORR}})) and correlation contribution of electrophilicity index ((\omega_{\text{CORR}})), in different splitting methods. The best CORR based model is presented in bold</td>
<td>58</td>
</tr>
<tr>
<td>3.3</td>
<td>Comparison of internal and external validation parameters of QSAR model VII based on electron correlation-energy ((E_{\text{CORR}})) and correlation contribution of electrophilicity index ((\omega_{\text{CORR}})), with the best known models in the literature for the TA100 mutagenicity of nitrated-AHs and PAHs.(^{20,23}) It should be noted that the compared models, developed in this work and literature, are based on exactly the same data set (see supporting information Tables S3.15 and S3.16 of Annexure II)</td>
<td>59</td>
</tr>
<tr>
<td>4.1</td>
<td>Comparison of the exchange (X)-only, exchange-correlation (X+C) and correlation (CORR) quantum-mechanical methods in terms of the quantum-mechanical exchange and correlation interactions</td>
<td>70</td>
</tr>
<tr>
<td>4.2</td>
<td>Comparison of the key internal and external validation parameters for the models based on the total energy ((E)) and energy of the HOMO ((E_{\text{HOMO}})) computed with the Exchange (X) only, Exchange + Correlation (X+C) methods, and with the CORR method incorporating mainly the effect of electron-correlation in the descriptors, for modeling the TA100 mutagenicity of nitrated-PAHs (models for the TA98 mutagenicity are provided in the supporting information Table S4.5 of Annexure III)</td>
<td>75</td>
</tr>
<tr>
<td>4.3</td>
<td>Same as Table 4.2 but for the models based on the total electronic energy ((E)) and energy of the LUMO ((E_{\text{LUMO}})), (models for the TA98 mutagenicity are provided in the supporting information Table S4.6 of Annexure III)</td>
<td>76</td>
</tr>
</tbody>
</table>
Table Sr.No. | Description                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Page No.
--- | ---                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | ---
4.4 | Same as Table 4.2 but for the models based on the total electronic energy ($E$) and absolute electronegativity ($\chi$), (models for the TA98 mutagenicity are provided in the supporting information Table S4.7 of Annexure III)                                                                                     | 77
4.5 | Same as Table 4.2 but for the models based on the total energy ($E$) and chemical hardness ($\eta$), (models for the TA98 mutagenicity are provided in the supporting information Table S4.8 of Annexure III)                                                                                                                             | 78
4.6 | Same as Table 4.2 but for the models based on the total electronic energy ($E$) and electrophilicity index ($\omega$), (models for the TA98 mutagenicity are provided in the supporting information Table S4.9 of Annexure III)                                                                                              | 79
4.7 | Same as Table 4.2 but for the consensus models based on the best descriptors and the methods observed in the Tables 4.2-4.6 for the TA100 mutagenicity, and in the supporting information Tables S4.5-S4.9 for the TA98 mutagenicity of Annexure III                                                                                       | 90
5.1 | Comparison of key internal and external validation parameters of the QSAR models based on different quantum-chemical molecular, quantum-mechanically computed thermodynamic descriptors along with physicochemical, electrostatic and topological descriptors, with model known in literature (Ref. 9) for the acute toxicity of chemicals towards the *Daphnia magna*                                                                 | 106
5.2 | Important internal and external validation parameters of the four-descriptors based QSAR models developed with the quantum-chemical descriptors computed with PM7, HF, DFT methods and with the CORR descriptors along with the physicochemical, electrostatic and topological descriptors employed by Moosus and Maran (Ref. 9). The model indicated in the bold are among the best models observed. | 111
5.3 | Important internal and external validation parameters of the three-descriptors based QSAR models developed with the quantum chemical descriptors computed with PM7, HF, DFT methods and with the CORR descriptors along with descriptors employed by Moosus and Maran (Ref. 9). The models indicated in bold are among the best models observed.            | 114
6.1 | Key internal and external validation parameters of the one-descriptors QSAR models developed with quantum-chemical descriptors computed with the PM7, HF, DFT and CORR methods employing 30% ordered response splitting                                                                                                                      | 127
6.2 | Key internal and external validation parameters of the two-descriptors QSAR models developed with the quantum-chemical descriptors computed with PM7, HF, DFT and CORR methods employing 30% ordered response splitting.                                                                                                                  | 130
6.3 | Key internal and external validation parameters of the two-descriptors based QSAR models developed with combination of correlation energy descriptor ($E_{\text{CORR}}$) with whole descriptors computed with PM7, HF, DFT methods employing 30% ordered response splitting.                                                                 | 132
6.4 | Comparison of internal and external validation parameters of QSAR model VII based on electron-correlation energy ($E_{\text{CORR}}$) and energy computed with PM7 ($E_{\text{PM7}}$) and HF ($E_{\text{HF}}$) methods, with the model known in the literature for inhibitory growth concentration (Log $IGC_{50}^{-1}$) of alkyl- and halogen-substituted nitrobenzenes. | 135
### List of Figures

<table>
<thead>
<tr>
<th>Figure Sr.No.</th>
<th>Description</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Workflow of a typical QSAR strategy (for details, see section 1.4).</td>
<td>3</td>
</tr>
<tr>
<td>3.1</td>
<td>Scatter plot for model VII based on $E_{\text{CORR}}$ and $\omega_{\text{CORR}}$. Open (yellow) circles represents training set compounds while filled (blue) circles represents prediction set compounds.</td>
<td>61</td>
</tr>
<tr>
<td>3.2</td>
<td>Williams plot for model VII based on $E_{\text{CORR}}$ and $\omega_{\text{CORR}}$. Warning leverage $h^* = 3.00$. Encircled value represents ID number of the compounds (see supporting information Table S3.1 in Annexure II).</td>
<td>62</td>
</tr>
<tr>
<td>4.1</td>
<td>Comparison of the mean absolute error ($MAE$) and the root mean square error ($RMSE$) in the external (EXT) predictivity of various models, for the TA100 mutagenicity of nitrated-PAHs, based on the total energy ($E$), energy of the HOMO and the LUMO, absolute electronegativity ($\chi$), chemical hardness ($\eta$) and electrophilicity index ($\omega$) computed through exchange-only methods (HFX, B88), exchange-correlation (XC) methods (HFX+LYP, BLYP, B3LYP, M06, M06-L, M06-2X), and also based on the descriptors incorporating mainly the effect of electron correlation (CORR) from the respective XC methods (for the TA98 mutagenicity models, see supporting information Figure S4.1 provided in Annexure III).</td>
<td>80</td>
</tr>
<tr>
<td>4.2</td>
<td>(a) Williams plot: Standardized residuals versus leverage ($h$), of the TA100 mutagenicity model based on the total energy and the energy of the HOMO computed using DFT/HFX method. The training and prediction set chemicals, represented with open (yellow) and filled (blue) circles, respectively, are obtained using 30% random splitting method. The encircled values represent the ID number of the compounds, provided in the supporting information Table S4.1 of Annexure III (for other best models, see supporting information Figure S4.2 of Annexure III). The vertical (solid) line indicates the warning leverage ($h^*$). (b) Scatter plot: Experimental versus predicted $\log_{10}$TA100 mutagenicity using model as specified in (a), (for other best models, see the supporting information Figure S4.3 of Annexure III).</td>
<td>82</td>
</tr>
<tr>
<td>4.3</td>
<td>Same as Figure 4.2, but for the TA100 mutagenicity model based on $E_{\text{CORR}}$ (B3LYP) and $\omega_{\text{CORR}}$ (B3LYP) descriptors, incorporating mainly the effect of electron-correlation (CORR) in the total-energy ($E$) and the electrophilicity index ($\omega$), computed using the DFT employing B3LYP hybrid XC functional (see also supporting information Figures S4.2 and S4.3 of Annexure III).</td>
<td>87</td>
</tr>
<tr>
<td>5.1</td>
<td>Various structural functionalities of the data set comprising 252 chemicals (A) linear hydrocarbons, (B) benzene &amp; substituted benzene, (C) chemicals with –OH group, (D) chemicals with -NO$_2$ group, (E) chemicals with -NH$_2$ group, (F) chemicals containing Cl atom (G) ring system of C &amp; H with hetero atom (X=N,O,S), (H) ring system of C &amp; H without hetero atom, (I) chemicals with =C=O group, (J) chemicals with =C=S group, (K) chemicals containing P atom, (L) chemicals with R-O-R’ group. (R represents alkyl group and Ar represents an aryl group).</td>
<td>98</td>
</tr>
<tr>
<td>5.2</td>
<td>Williams plot of the standardized residuals vs leverage ($h$) for the model based on $E_{\text{HOMO}}$, log $P_{\text{mix}}$, WNSA-1, $^1BIC$ given by Moosus and Maran (Ref.9) (entry 1 in Table 4.1). Training and prediction set chemicals are represented with open (yellow) and filled (blue) circles, respectively. Encircled number represents the ID number of chemicals detected as structural and response outliers (for details, refer to supporting information Table S5.1 of Annexure III). The vertical (solid) line indicates warning leverage $h^*$, whereas the horizontal (dashed) line specifies the standardized residual value of 2.5.</td>
<td>107</td>
</tr>
</tbody>
</table>
5.3 Williams plot of the standardized residuals vs leverage \((h)\) for model II(a-d) (without excluding outliers), based on \(E_{\text{HOMO}}, \log P_{\text{mix}}, \text{WNSA-1}, ^{1}BIC\) with descriptor \(E_{\text{HOMO}}\) computed using (a) PM7 method (b) HF method (c) DFT method and (d) CORR method. Training and prediction set chemicals, using activity sampling splitting, are represented with open (yellow) and filled (blue) circles, respectively. Encircled number represents the ID number of chemicals detected as structural and response outliers (for details, refer to Supporting Information Table S5.1 of Annexure III). The vertical (solid) line indicates warning leverage \(h^*\), whereas the horizontal (dashed) line specifies standardized residual value of 3.0.

5.4 Williams plot of the standardized residuals vs leverage \((h)\) for the models II(a-d) (listed in Table 4.2), based on \(E_{\text{HOMO}}, \log P_{\text{mix}}, \text{WNSA-1}, ^{1}BIC\), after excluding the outliers which are depicted in Figure 4.3, with descriptor \(E_{\text{HOMO}}\) computed using (a) PM7 method (b) HF method (c) DFT method and (d) CORR method. Training and prediction set chemicals, using ordered response 50% splitting, are represented with open (yellow) and filled (blue) circles, respectively. Encircled number represents the ID number of chemical with high leverage value (for details, refer to supporting information Table S4.1 of Annexure III). The vertical (solid) line indicates warning leverage \(h^*\), whereas the horizontal (dashed) line specifies the standardized residual value of 3.0.

5.5 Scatter plot of the experimental vs predicted \(\text{Log}_{10}LC_{50}-48h\) for the models II(a-d) (listed in Table 5.2), based on \(E_{\text{HOMO}}, \log P_{\text{mix}}, \text{WNSA-1}, ^{1}BIC\) with \(E_{\text{HOMO}}\) descriptor computed using (a) PM7 method (b) HF method (c) DFT method and (d) CORR method. Training and prediction set chemicals using activity sampling splitting are represented with open (yellow) and filled (blue) circles, respectively

6.1 (a) Williams plot: Standardized residuals versus leverage \((h)\), of the \(\text{Log}_{10}IGC_{50}-1\) for the model based on the \(E_{\text{CORR}}\). The training and prediction set chemicals, represented with open (yellow) and filled (blue) circles, respectively, are obtained with activity sampling ordered response splitting method. The encircled values represent the ID number of the compounds, provided in the supporting information Table S6.2 of Annexure V.

(b) Scatter plot: Experimental versus predicted \(\text{Log}_{10}IGC_{50}-1\) using model as specified in (a).

6.2 Influence of the addition of descriptor on the internal reliability of the QSAR models, developed with the CORR and PM7 methods.

6.3 (a) Williams plot: Standardized residuals versus leverage \((h)\), of the \(\text{Log}_{10}IGC_{50}-1\) for the model based on the \(E_{\text{CORR}}, \omega_{\text{CORR}}\). The training and prediction set chemicals, represented with open (yellow) and filled (blue) circles, respectively, are obtained with activity sampling ordered response splitting method. The encircled values represent the ID number of the compounds, provided in the supporting information Table S6.2.

(b) Scatter plot: Experimental versus predicted \(\text{Log}_{10}IGC_{50}-1\) using model as specified in (a).
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM</td>
<td>Average consensus model</td>
</tr>
<tr>
<td>ADMET</td>
<td>Absorption, distribution, metabolism, excretion and toxicity</td>
</tr>
<tr>
<td>AHs</td>
<td>Aromatic hydrocarbons</td>
</tr>
<tr>
<td>AM1</td>
<td>Austin model 1</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial neural network</td>
</tr>
<tr>
<td>(^1)BIC</td>
<td>Average bonding information content</td>
</tr>
<tr>
<td>BLYP</td>
<td>Becke, Lee-Yang-Parr</td>
</tr>
<tr>
<td>B3LYP</td>
<td>Becke, 3 parameter, Lee-Yang-Parr</td>
</tr>
<tr>
<td>B88</td>
<td>Becke exchange</td>
</tr>
<tr>
<td>CA-II</td>
<td>Carbonic anhydrase – II</td>
</tr>
<tr>
<td>CC</td>
<td>Coupled cluster</td>
</tr>
<tr>
<td>CCC</td>
<td>Concordance correlation coefficient</td>
</tr>
<tr>
<td>CI</td>
<td>Configuration interaction</td>
</tr>
<tr>
<td>CIC1</td>
<td>Complementary information content</td>
</tr>
<tr>
<td>CORR</td>
<td>Electron-correlation contribution</td>
</tr>
<tr>
<td>CV</td>
<td>Cross validation</td>
</tr>
<tr>
<td>(D_{\text{CORR}})</td>
<td>Descriptors incorporating the effects of electron – correlation.</td>
</tr>
<tr>
<td>DFT</td>
<td>Density functional theory</td>
</tr>
<tr>
<td>EA</td>
<td>Electron affinity</td>
</tr>
<tr>
<td>(E_{\text{CORR}})</td>
<td>Correlation energy</td>
</tr>
<tr>
<td>(E_{\text{CORR}}[\rho])</td>
<td>Correlation energy-density functional</td>
</tr>
<tr>
<td>(E_{x}[\rho])</td>
<td>Exchange energy-density functional</td>
</tr>
<tr>
<td>(E_{xc}[\rho])</td>
<td>Exchange-correlation energy-density functional</td>
</tr>
<tr>
<td>EXT</td>
<td>External</td>
</tr>
<tr>
<td>FCI</td>
<td>Full configuration interaction</td>
</tr>
<tr>
<td>GA-VS</td>
<td>Genetic algorithm for the variable selection</td>
</tr>
<tr>
<td>GGA</td>
<td>Generalized gradient approximation</td>
</tr>
<tr>
<td>HF</td>
<td>Hartree-Fock</td>
</tr>
<tr>
<td>HFX</td>
<td>Hartree-Fock exchange</td>
</tr>
<tr>
<td>HOMO</td>
<td>Highest occupied molecular orbital</td>
</tr>
<tr>
<td>IP</td>
<td>Ionization potential</td>
</tr>
<tr>
<td>(k)-NN</td>
<td>(k)-nearest neighbor</td>
</tr>
<tr>
<td>KS</td>
<td>Kohn and Sham</td>
</tr>
<tr>
<td>LDA</td>
<td>Local density approximation</td>
</tr>
<tr>
<td>LMO</td>
<td>Leave many out</td>
</tr>
<tr>
<td>Log(P)</td>
<td>Octanol/water partition coefficient</td>
</tr>
<tr>
<td>LOO</td>
<td>Leave one out</td>
</tr>
<tr>
<td>LSD</td>
<td>Local spin density</td>
</tr>
<tr>
<td>LUMO</td>
<td>Lowest unoccupied molecular orbital</td>
</tr>
<tr>
<td>LYP</td>
<td>Lee-Yang-Parr</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean absolute error</td>
</tr>
<tr>
<td>MBPT</td>
<td>Many body perturbation theory</td>
</tr>
<tr>
<td>MINDO</td>
<td>Modified intermediate neglect of differential overlap</td>
</tr>
<tr>
<td>MCSCF</td>
<td>Multiconfigurational self-consistent field</td>
</tr>
<tr>
<td>MLR</td>
<td>Multiple linear regression</td>
</tr>
<tr>
<td>MOPAC</td>
<td>Molecular orbital package</td>
</tr>
<tr>
<td>MP2</td>
<td>Second order Møller-Plesset perturbation theory</td>
</tr>
<tr>
<td>MP3</td>
<td>Third order Møller-Plesset perturbation theory</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PAHs</td>
<td>Polyaromatic hydrocarbons</td>
</tr>
<tr>
<td>PBE</td>
<td>Perdew-Burke-Ernzerhof</td>
</tr>
<tr>
<td>PCA</td>
<td>Principle component analysis</td>
</tr>
<tr>
<td>PCDDs</td>
<td>polychlorinated-dibenzo-(p)-dioxins</td>
</tr>
<tr>
<td>PCDFs</td>
<td>polychlorinated-dibenzo-furans</td>
</tr>
<tr>
<td>PM3</td>
<td>Parameterized model number 3</td>
</tr>
<tr>
<td>PM6</td>
<td>Parameterized model number 6</td>
</tr>
<tr>
<td>PM7</td>
<td>Parameterized model number 7</td>
</tr>
<tr>
<td>PRESS</td>
<td>Predicted error sum of squares</td>
</tr>
<tr>
<td>PW2</td>
<td>Path/Walk-2 Randic shape index</td>
</tr>
<tr>
<td>QSAR</td>
<td>Quantitative structure-activity relationship</td>
</tr>
<tr>
<td>QSPR</td>
<td>Quantitative structure-property relationship</td>
</tr>
<tr>
<td>RM1</td>
<td>Recife model 1</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root mean square error</td>
</tr>
<tr>
<td>SE</td>
<td>Semi-empirical method</td>
</tr>
<tr>
<td>SOM</td>
<td>Self organizing map</td>
</tr>
<tr>
<td>TR</td>
<td>Training</td>
</tr>
<tr>
<td>TSS</td>
<td>Total sum of squares</td>
</tr>
<tr>
<td>VS</td>
<td>Voorhis and Scuseria</td>
</tr>
<tr>
<td>VWN</td>
<td>Vosko, Wilk, and Nusair</td>
</tr>
<tr>
<td>XC</td>
<td>Exchange-correlation</td>
</tr>
<tr>
<td>WCM</td>
<td>Weight consensus model</td>
</tr>
<tr>
<td>WNSA-1</td>
<td>weighted partial negative surface area</td>
</tr>
</tbody>
</table>
List of Publications

